SCIENTIFIC REPORTS
natureresearch

OPEN

Arsenic mobilization in a high arsenic groundwater revealed by metagenomic and Geochip analyses

Zhou Jiang[1,2], Ping Li[1], Yanhong Wang[1], Han Liu[1], Dazhun Wei[1], Changguo Yuan[1] & Helin Wang[1]

Microbial metabolisms of arsenic, iron, sulfur, nitrogen and organic matter play important roles in arsenic mobilization in aquifer. In this study, microbial community composition and functional potentials in a high arsenic groundwater were investigated using integrated techniques of RNA- and DNA-based 16S rRNA gene sequencing, metagenomic sequencing and functional gene arrays. 16S rRNA gene sequencing showed the sample was dominated by members of *Proteobacteria* (62.3–75.2%), such as genera of *Simplicispira* (5.7–6.7%), *Pseudomonas* (3.3–5.7%), *Ferribacterium* (1.6–4.4%), *Solimonas* (1.8–3.2%), *Geobacter* (0.8–2.2%) and *Sediminibacterium* (0.6–2.4%). Functional potential analyses indicated that organics degradation, assimilatory sulfate reduction, As-resistant pathway, iron reduction, ammonification, nitrogen fixation, denitrification and dissimilatory nitrate reduction to ammonia were prevalent. The composition and function of microbial community and reconstructed genome bins suggest that high level of arsenite in the groundwater may be attributed to arsenate release from iron oxides reductive dissolution by the iron-reducing bacteria, and subsequent arsenate reduction by ammonia-producing bacteria featuring *ars* operon. This study highlights the relationship between biogeochemical cycling of arsenic and nitrogen in groundwater, which potentially occur in other aquifers with high levels of ammonia and arsenic.

Arsenic (As) is a toxic metalloid and is widespread in the environment. Elevated As in groundwater endangers the health of hundreds of millions of people worldwide, especially in Southern and Southeastern Asia[1,2]. The widespread problem of high levels of As in groundwaters prompts intensive studies from multi-disciplinary research fields, such as geochemistry, hydrology, mineralogy, microbiology and ecology, to understand the hydro-biogeochemical behaviors of As in global aquifer systems[3]. To date, three main mechanisms have been proposed for As mobilization in groundwater: (1) the reductive dissolution of As-bearing iron (Fe) oxides under reducing conditions[4,5], (2) the oxidative dissolution of As-bearing pyrite and arsenopyrite under oxidizing conditions[6,7], and (3) the competitive adsorption between As and phosphate/bicarbonate under alkaline conditions[8,9].

The first mechanism commonly prevails in the late Pleistocene-Holocene aquifer systems with the sedimentation of fresh organic matter, such as inland basins from West Bengal, Bangladesh, Nepal, Cambodia, Vietnam and China, where degradation of organic matter fuels microorganisms to reduce and dissolve As-bearing iron oxides[4,10,11]. Among those microbes, iron-reducing and sulfate-reducing bacteria are known to play important roles in As mobilization. For example, dissimilatory iron-reducing bacteria, such as *Geobacter* sp. and *Shewanella* sp., could directly utilize organic carbon as electron sources to reduce As-bearing iron oxides, leading to As release or re-immobilization by forming secondary iron minerals, such as vivianite and magnetite[12–14]. Sulfide generated by sulfate-reducing bacteria in aquifers abiotically reduces As-bearing iron oxides to mobilize As[15–17]. If derived ferrous iron is sufficiently available in groundwater, sulfide could form iron-sulfide minerals to further sequester As through co-precipitation and/or adsorption[18–20]. However, the reductive dissolution mechanism only accounts for the initial trigger of arsenate release from iron oxides, but not for the transformation from arsenate to arsenite in reducing aquifers. Previous studies based on *arrA* gene coding respiratory arsenate reductase large subunit highlight the role of dissimilatory arsenate-respiring bacteria on arsenate transformation[13,21,22]. In contrast to sporadic arsenate respiring organisms, As detoxifying organisms harboring arsenic-resistance (*ars*)

[1]State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, 430074, P.R. China. [2]School of Environmental Studies, China University of Geosciences, Wuhan, 430074, P.R. China. Correspondence and requests for materials should be addressed to Z.J. (email: jiangzhou@cug.edu.cn) or P.L. (email: pli@cug.edu.cn).

operon are much more prevalent among prokaryotes and well known to contribute to arsenate reduction[23–25]. In addition, it was reported that microbial oxidation and methylation of arsenite occur in some groundwater[26,27]. Therefore, to comprehensively understand As transformation processes in groundwater systems, those functional genes of As reduction, oxidation and methylation should be evaluated together using some high-throughput molecular technologies, such as metagenome and Geochip[28]. Besides, the integrated metabolic potential of metagenome and Geochip enables As metabolism to link to other corresponding elements cycle, such as C, N and S, and further elucidate their relationship and role in As mobilization. With the complementary beneficial characteristics including new genes discovery, low detection limits, high reproducibility, and potential for quantification, metagenome and Geochip could be combined to reveal microbial functional potentials in specific ecosystems of interest.

High levels of As in groundwater systems is often accompanied by high ammonium concentrations, such as Hetao basin, China[29], Bengal delta plain and Ganges River floodplain, Bangladesh[30,31], Red River floodplain, Vietnam[32], and Mekong Delta, Cambodia[33]. It was reported that ammonium in groundwaters reflects degradation of organic matter and/or inputs of anthropogenic activities, such as fertilizers and manure, and only act on the formation of reducing conditions to favor the reductive dissolution of As-bearing Fe oxides[32,34,35]. Recently, isotopic evidence of nitrogen sources in As-contaminated groundwater suggested the prevalence of nitrification, denitrification, ammonification and dissimilatory nitrate reduction to ammonia (DNRA)[34,36,37]. These results infer that some nitrogen metabolic processes possibly link to As mobilization in groundwater systems. After all, none of the above iron-reducing, sulfate-reducing and arsenate-reducing bacteria could reasonably explain the positive correlation between As and ammonium concentrations in groundwater.

The objective of this study is therefore to further understand the microbial mobilization and transformation processes of As in groundwater, especially the role of nitrogen metabolism on As mobilization. To this end, we used the integrated techniques of RNA- and DNA-based 16S rRNA gene sequencing, metagenomic sequencing and functional gene array (Geochip 4.0) to assess microbial community composition and functional potential in a high As groundwater. This groundwater is located in Hetao Basin, Inner Mongolia, China, a representative endemic arsenicosis area with an affected population of more than two hundred thousand[38]. Our results indicated that some ammonium-producing bacteria with *ars* operon is likely to mediate the transformation from arsenate to arsenite in groundwater after it is released from the reductive dissolution of As-bearing iron oxides by iron-reducing bacteria.

Material and Methods

Field sampling. The hand-pumped tube well (N40.96683°, E107.00517°) from resident Dingshan Li in Shahai village, Hangjinhouqi County, western Hetao Basin, was selected for this study because of high levels of As detected and arseniasis family history (two members died of long-term drinking groundwater). The sample herein was abbreviated LDS. Groundwater biomass were collected using both normal filtration, with a 0.22 μm × 142 mm (pore size × diameter) nitrocellulose membrane filter (Merck Millipore, Billerica, MA, USA), and tangential flow filtration (TFF; Prep/Scale Spiral Cartridge with 30 kDa molecular weight cutoff, Millipore, Billerica, MA, USA)[39]. The filtered water was collected into 100 mL acid-washed polypropylene bottle (to measure As species), brown glass bottle (to measure dissolved organic carbon, DOC) and serum vial with butyl rubber septum stoppered and aluminum seal (to measure methane). The filtered membrane was cut into two equal parts and placed into 15 mL sterile polypropylene tubes with/without prefilled Ambion RNA *later* Solution (ThermoFisher Scientific, Waltham, WA, USA) (RNA and DNA extraction for 16S rRNA gene sequencing) respectively. Concentrated cell suspension by TFF was packed into 500 mL sterile nuclease-free plastic bottles (DNA extraction for metagenomic sequencing and Geochip analyses). The groundwater volumes of normal filtration and TFF is approximately 10 L and 2000 L respectively. All samples were transported to the laboratory on dry ice and then stored at −80 °C until further analyses.

Geochemical measurements. Geochemical parameters including temperature, pH, electrical conductivity and oxidation-reduction potential (ORP) were measured *in situ* using a portable multi meter (HQ40D, Hach, Loveland, CO, USA). Concentrations of sulfide, sulfate, ammonium, nitrite, nitrate, phosphate, ferrous iron and total iron in groundwater were determined at the sampling with a portable colorimeter (DR890, Hach, Loveland, CO, USA) according to the manufacture's procedures. In the laboratory, four As species of arsenite, arsenate, monomethylarsenate (MMA) and dimethylarsonate (DMA) were performed using liquid chromatography combined to hydride generation atomic fluorescence spectrometry (AFS-9780, Haiguang, Beijing, China). Dissolved organic carbon (DOC) concentrations were determined using a TOC analyzer (Vario MICRO cube, Elementar, Langenselbold, Germany). Soluble methane concentrations were analyzed by gas chromatography (Thermo Trace Ultra) at the Third Institute of Oceanography, State Oceanic Administration of China. All samples were run in triplicate and then averaged.

DNA and RNA extraction and cDNA synthesis. Cells in TFF concentrates were pelleted by centrifugation at 6000 rcf for 10 min after defrosting. Genomic DNA was extracted from the pellet and filter using MoBio PowerSoil DNA Isolation Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions. Total RNA was extracted from the filter preserved in Ambion RNA*later* Solution using RNeasy Mini Kit (Qiagen, Valencia, CA, USA). An Ambion Tubro DNA-free Kit (ThermoFisher Scientific, Waltham, WA, USA) was used to digest residual DNA. To verify complete removal of genomic DNA in total RNA, we utilized total RNA as a template to amplify 16S rRNA gene with general primer 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) and performed electrophoresis detection. DNA and RNA concentrations were measured by NanoDrop One (Thermo Fisher Scientific, Waltham, MA, USA).

Reverse transcription-PCR (RT-PCR) amplification of 16S rRNA was achieved using Invitrogen SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase (ThermoFisher Scientific, Waltham, WA, USA). Specifically, it was carried out in a 50 μL reaction containing 25 μL 2 × Reaction buffer, 0.2 μM of both sense and anti-sense primers (515F and 806R), 10–15 ng template and 2 μL SuperScript III RT/Platinum Taq Mix using the following program: cDNA synthesis at 55 °C for 30 min; initial denaturation at 94 °C for 3 min; 35 cycles of 94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s; final extension at 72 °C for 10 min. The synthesized cDNA was purified with the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA).

16S rRNA gene amplification, sequencing and preprocessing. The V4 region of 16S rRNA gene was amplified from both DNA and synthesized cDNA using the primer 515F and 806R combined with Illumina adapter sequences, a pad and a linker, as well as barcodes on the reverse primers. PCR amplification was carried out in a 25 μL reaction buffer containing 12.5 μL GoTaq colorless Master Mix (Promega, Madison, WI, USA), 0.5 μM of both forward and reverse primers, 10–15 ng template using the following program: initial denaturation at 94 °C for 3 min, followed by 35 cycles of 94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s, and then a final extension at 72 °C for 10 min. Reactions were performed in triplicate and pooled.

Amplicons were verified by agarose gel electrophoresis and quantified with PicoGreen using a FLUOstar Optima microplate reader (BMG Labtech, Jena, Thuringia, Germany). 500 ng of amplicon from each sample were combined together and purified using MoBio UltraClean PCR Clean-Up Kit (Qiagen, Valencia, CA, USA) and then re-quantified with PicoGreen. Sample library was prepared according to the MiSeq Reagent Kit Preparation Guide (Illumina, San Diego, CA, USA). Briefly, the sample was denatured by mixing 10 μL 2 nM combined amplicon and 10 μL 0.2 mol NaOH and incubated for 5 min at room temperature. Denatured sample was diluted to 7 pM using HT1 buffer and mixed with 7 pM PhiX control (final PhiX control concentration, 10%). A total of 600 μL sample mixture, together with customized sequencing primers for forward, reverse, and index reads, were loaded into the corresponding wells on the cartridge of a 300-cycle MiSeq Reagent Kit v2 and run on an Illumina MiSeq system (Illumina, San Diego, CA, USA). RNA- and DNA-based 16S rRNA gene sequencing raw data were deposited to the Short Read Archive database at NCBI (Accession number: PRJNA493661).

Raw sequencing data with perfect matches to barcodes were split to sample libraries. Forward and reverse reads with at least 30 bp overlap and less than 25% mismatches were joined using Fast Length Adjustment of SHort reads (FLASH)[40]. The joined sequences were trimmed using Btrim with a QC threshold of greater than 30 over a 5 bp window size and a minimum length of 140 bp[41]. After deleting sequence containing N base and trimming by sequence length of <251 and >255 bp, the sequences were clustered to operational taxonomic units (OTUs) by Uparse at a similarity level of 97%[42]. Singletons in generated OTU tables were removed for downstream analyses. The Ribosomal Database Project (RDP) classifier was used to assign the taxonomy of representative sequences at a minimal 50% confidence[43]. Samples were rarefed at 55841 sequences per sample. The above steps were performed through a Galaxy-based pipeline compiled by Ye Deng's lab in Research Center for Eco-Environmental Sciences, China Academy of Sciences, China (http://mem.rcees.ac.cn:8080).

Metagenomic sequencing, preprocessing and genome binning. DNA was randomly sheared into small pieces of ~350 bp using Covaris M220 Focused-ultrasonicator (ThermoFisher Scientific, Waltham, WA, USA). Sample library was prepared according to the instruction manual of NEB Next Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA). Briefly, it consists of fragmented DNA end-repaired, adaptor ligation, size selection and cleanup of adaptor-ligated DNA, PCR enrichment of adaptor-ligated DNA (5 cycles) and cleanup of PCR reaction. The library concentration was determined using Qubit 3.0 Fluorometer (ThermoFisher Scientific, Waltham, WA, USA) and sequenced on an Illumina HiSeq X Ten platform at Guangdong Magigene Biotechnology Co. Ltd (Guangdong, China). ~13 Gbp (2 × 150 bp) reads were generated. Metagenomic sequencing raw data were deposited to the Short Read Archive database at NCBI (Accession number: PRJNA493663).

The raw sequencing data was trimmed to produce clean data using Trimmomatic (Version 0.36)[44] and *de novo* assembled using MEGAHIT(Version 1.0.6)[45]. The assembled scaffolds were interrupted from N base site and the <500 bp scafftigs were removed. MetaGeneMark(Version 3.38) was used with default parameters to predict the open reading frame (ORF) of scafftigs and those ORFs with a length of <90 bp were omitted[46]. Redundant gene catalogue(Unigenes) was obtained using CD-HIT (Version:4.7) by grouping sequences with an identity of >95% and coverage of >90% in clusters and assigning the longest sequence as representative of the cluster[47]. The clean data was aligned to the Unigenes using BBMap (http://jgi.doe.gov/data-and-tools/bbtools/) to obtain the number of mapped reads of each gene[48]. Based on the number of mapped reads and length of gene, each gene abundance was computed and utilized for subsequent statistical analysis. DIAMOND software was used to blast the Unigenes to the sequences of bacteria, fungi, archaea and viruses extracted from NCBI non-redundant (NCBI-nr) database with the e value threshold of 1e-10[49]. The best hit was chosen to be the taxonomic unit of Unigenes based on the results of lowest common ancestor (LCA) algorithm compiled in MEGAN software[50]. Functional annotation was performed by blasting Unigenes to the functional database including NCBI-nr database, KEGG database (http://www.kegg.jp/kegg/) and eggNOG database (http://eggnogdb.embl.de/) using DIAMOND software with the e value threshold of 1e-10. For each gene's blast result, the best hit with lowest e value and Blast coverage ratio between query and reference sequences of >40 was chosen for subsequent analysis. In addition, the coverage information of the scaffolds with lengths of >5 kb and tetranucleotide frequency (TNF) were used to perform the genome binning by using MetaBAT with a parameter of verysensitive[51]. The completeness, contamination and strain heterogeneity of genomic bins were then estimated by using CheckM[52]. Finally, three genome bins belonging to *Pseudomonas* and *Acidovorax* were retained for further analysis.
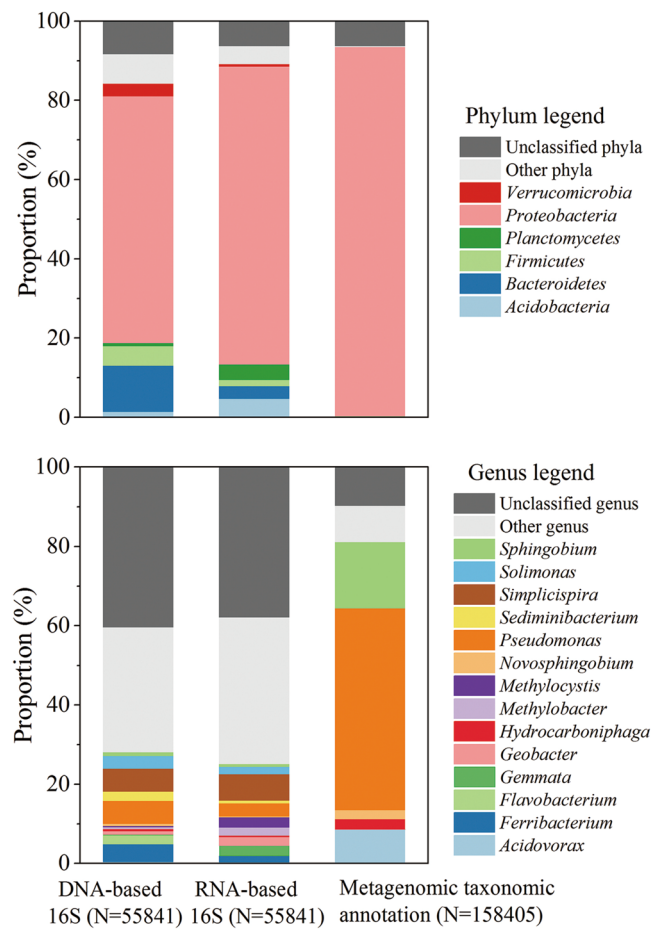
Figure 1. Microbial community composition and structure in sample LDS revealed by RNA- and DNA-based 16S rRNA gene sequencing and metagenomic taxonomic annotation. Upper diagram: at the level of phylum; Down diagram: at the level of genus. Only the phyla and genera with abundances of >2% are displayed. N in parentheses indicates the number of reads (16S rRNA gene) or Unigenes (metagenomic).

Geochip 4.0 analysis. The GeoChip 4.0, containing 82074 oligonucleotide probes targeting 141995 genes in 410 gene families involving various element cycle, was used to analyze the functional potential of the sample. The experimental procedures referred to Tu *et al.*[53] for details. Briefly, the purified DNA was labeled with Cy-3 using random primers and the Klenow fragment of DNA polymerase I and then mixed with sample tracking control (NimbleGen, Madison, WI, USA). With an addition of hybridization buffer, the sample was hybridized at 42 °C for approximately 16 h on a MAUI hybridization station (BioMicro, Salt Lake City, UT, USA). After the hybridization, the array was scanned with a NimbleGen MS200 Microarray Scanner (Roche NimbleGen, Madison, WI, USA) at full laser power. Signal intensities were measured based on scanned images. Poor quality spots with a signal-to-noise ratio of less than 2.0 were removed and the positive signals were normalized before statistical analysis. The sample was run in triplicate and then averaged. GeoChip 4.0 hybridization data process was performed through a pipeline compiled by the Institute for Environmental Genomics, University of Oklahoma, USA (http://ieg.ou.edu/microarray).

Results and Discussion

Geochemistry. Sample LDS is under a slightly alkaline reducing condition with a total As concentration of 1.28 mg/L (Table S1). No MMA and DMA were detected and the ratio of arsenite to total As was 0.77, suggesting that inorganic As reduction prevails. Given high concentrations of DOC (8.58 mg/L) and ferrous iron (0.15 mg/L) and total iron (0.66 mg/L), the reductive dissolution of As-bearing iron oxides seems to contribute to As release in the groundwater, similar to previously reported results[54,55]. Sample LDS has extremely low concentrations of sulfide and sulfate (below detection limit), generally excluding the role of sulfur on As mobilization or sequestration[19,56,57]. High concentrations of ammonium (1.02 mg/L) and nitrate (0.79 mg/L) in sample LDS suggest the presence of some nitrogen metabolic processes, such as denitrification or DNRA[34,36].

Taxonomic profiling of microbial community. The As mobilization and transformation in groundwater systems are mediated by a variety of microorganisms[58]. An integrated analysis of 16S rRNA gene sequencing based on both RNA and DNA and metagenomic taxonomic annotations was used to investigate the composition and structure of microbial community in sample LDS (Figure 1 and Table S2). As the representative of active
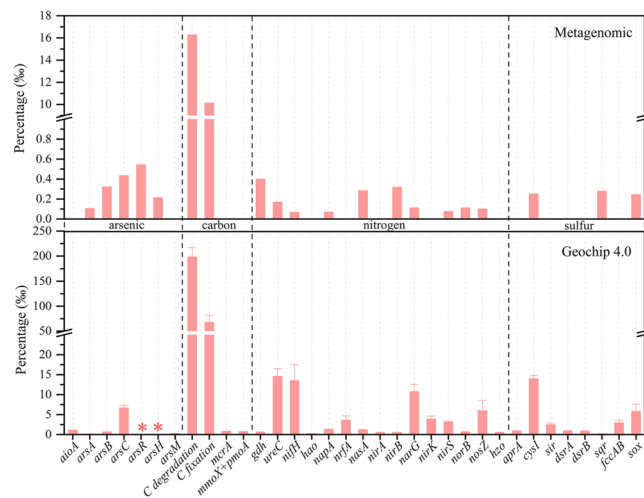
Figure 2. Relative abundances of arsenic, carbon, nitrogen and sulfur-metabolizing genes detected from metagenomic annotations in KEGG database (Upper diagram) and Geochip 4 (Down diagram). The genes with no detection by both metagenomic and Geochip 4, such as *arrA* and *amo*, are not displayed in this figure. Gene abbreviations refer to Table S4 for details. Gene marked by an asterisk represents no corresponding probe and error bar indicates standard deviation among triplicates in Geochip 4.0.

and present microbial population in sample LDS respectively, RNA- and DNA-based 16S rRNA gene sequencing showed similar results. *Proteobacteria* predominates in sample LDS, with a proportion of 62.3–75.2%. The following phyla are *Bacteroidetes* (3.2–11.6%), *Firmicutes* (1.6–4.9%), *Actinobacteria* (1.4–4.7%), *Planctomycetes* (0.8–3.9%) and *Verrucomicrobia* (0.6–3.2%). At the level of genus, the majority (37.8–40.3%) is unclassified, with the rest consist of *Simplicispira* (5.7–6.7%), *Pseudomonas* (3.3–5.7%), *Ferribacterium* (1.6–4.4%), *Solimonas* (1.8–3.2%), *Sediminibacterium* (0.6–2.4%), *Geobacter* (0.8–2.2%) and numerous rare genera. In contrast to the high sensitivity of 16S rRNA genes, metagenomics sequencing without deep coverage usually detect principal species in complex microbial communities[59,60]. As we found in this study, dominant phylum *Proteobacteria* has a proportion of up to 93.2% in metagenomic taxonomic annotation, in which major genera are mainly composed of *Pseudomonas* (50.9%), *Sphingobium* (16.7%), *Acidovorax* (8.6%), *Hydrogenophaga* (2.6%) and *Novosphingobium* (2.2%).

Typical iron-reducing bacteria detected by 16S rRNA gene sequencing based on both RNA and DNA, such as *Geobacter* and *Ferribacterium*, suggest that iron reduction probably occurs in sample LDS. These iron-reducing bacteria may utilize organic matter as electron donors to reduce As-bearing iron oxides and release absorbed As[5,61,62]. High concentrations of As, iron and DOC detected in sample LDS are consistent with the suggested roles of iron-reducing bacteria (Table S1). Except for reductive dissolution, As-bearing iron oxides might dissociate a small fraction of As into solution due to the competitive adsorption of phosphate in the groundwater (Table S1)[5,8,9]. As the most abundant genus in the sample LDS, *Simplicispira* likely involve nitrate reduction as the strains of the genus *Simplicispira* isolated from freshwater, soil, and activated sludge were reported to be capable of denitrification[63]. Many members of *Pseudomonas* and *Sphingobium* have demonstrated to be able to degrade aromatic compounds, and thus those derived intermediates may further serve as electron donors for themselves or other microorganisms to reduce iron, nitrate and arsenate[64,65].

Functional potential of microbial community. Metagenomic functional annotation and Geochip 4.0 were used to assess microbial functional potentials in sample LDS. Close to 11 Gbp clean data from metagenomic were assembled de novo to scaftigs and further derived 240574 protein-coding genes (Table S3). Among them, 104064 genes were successfully annotated to orthologue numbers in KEGG database. The results from metagenomic and Geochip 4.0 are similar regarding the relative abundances of As, carbon, nitrogen and sulfur-metabolizing genes (Figure 2 and Table S4). High abundances of genes in *ars* cluster (*arsRABCH*) and absence of *arrA* infer that arsenate reduction in sample LDS may be mediated by As-resistant organisms harboring *ars* operon[66]. Arsenite oxidation and methylation are much less prevalent, revealed by near absent *aioA* and *arsM*. These results coincide with high arsenite concentration found and no detectable MMA and DMA in sample LDS. Carbon compounds degradation genes are significantly higher than carbon fixation genes, suggesting the consumption of DOC as energy and electron sources. Detection of *mcrA*, *mmoX* and *pmoA* by Geochip 4.0 indicate that some methanogens and methanotrophs inhabit sample LDS and regulate the methane concentration together (204.8 μg/L) (Table S1). Methanogens detected by sequence- and group-specific probes in Geochip 4.0 are composed of uncultured archaea (60.7%), *Methanoculleus marisnigri* JR1 (10.1%), *Methanocorpusculum labreanum* (8.4%) and *Methanococcus aeolicus Nankai*-3 (6.7%) (Table S5), which is similar to our previous study on *mcrA* libraries in groundwater samples from Hetao basin, Inner Mongolia[67,68]. Regarding nitrogen cycle, the dominant processes in sample LDS mainly include denitrification (presence of a whole set of *narG*, *nirK*, *nirS*, *norB* and *nosZ*), nitrogen fixation (*nifH*) and ammonification (*gdh* and *ureC*). Detections of *napA*, *nrfA*, *nasA* and *nirAB* imply that DNRA and assimilatory nitrate

Pathway	bin 1	bin 2	bin 3
Carbon			
Glycolysis	√	√	√
Citrate cycle (TCA cycle)	√	√	√
The reductive acetyl-CoA pathway	partial	partial	partial
Methanogenesis	×	×	×
Methane oxidation	×	×	×
Nitrogen			
Assimilatory nitrate reduction	partial	×	partial
Dissimilatory nitrate reduction	√	partial	√
Denitrification	√	×	√
Nitrogen fixation	√	×	×
Ammonification	√	√	√
Nitrification	partial	×	partial
Anammox	×	×	×
Sulfur			
Assimilatory sulfate reduction	√	√	√
Dissimilatory sulfate reduction	×	×	×
Sulfur oxidation	×	×	×
Sox System	partial	partial	partial
Arsenic			
Arsenic resistant pathway	√	√	√
Dissimilatory arsenate reduction	×	×	×
Arsenite oxidation	×	×	×
Arsenite methylation	×	×	×

Table 1. Overview of carbon, nitrogen, sulfur and As-metabolizing pathways in reconstructed genome bins from sample LDS.

reductions to ammonia may occur. The near absence of *amo*, *hao* and *hzo* suggest inert nitrification and anammox activities. This is also supported by no detection of these characteristic anammox bacteria (genera of *Candidatus Kuenenia*, *Candidatus Brocadia*, *Candidatus Jettenia*, *Candidatus Anammoxoglobus* and *Candidatus Scalindua*) in RNA- and DNA-based 16S rRNA gene sequencing[69]. High levels of ammonium in sample LDS is thus attributed to multiple production pathways of ammonium as well as lack of appreciable ammonium-consuming process. In contrast to low levels of *aprA*, *dsrA* and *dsrB*, high abundances of *cysI* and *sqr* indicate that assimilatory sulfate reduction rather than dissimilatory sulfate reduction dominate in sulfur-deficient sample LDS. The transformation from thiosulfate to sulfate probably occurs, revealed by the presence of *sox*.

Metagenomic and Geochip 4.0 analyses display different results in terms of what microorganism possess above functional genes (Table S5). For example, metagenomic analyses indicate that the majority of *arsC*-harboring bacteria are affiliated with *Pseudomonas*, followed by *Sphingobium*, *Sphingopyxis* and *Acidovorax*. Differently, Geochip 4.0 analyses show the presence of *arsC* gene in *Delftia acidovorans* SPH-1, *Maricaulis maris* MCS10, *Aspergillus fumigatus* Af293, *Shewanella putrefaciens* CN-32, *Shewanella* sp. ANA-3 and *Pseudomonas putida* GB-1. This disparity should be attributed to the different nature of those technologies[70]. It is known that Geochip requires a *priori* sequence information to synthesize probes before hybridization and these probes enable Geochip to identify both abundant and rare species by theirs sensitivity and specificity[70,71]. On the contrary, metagenome without a high sequencing depth only detect those dominant species in complex microbial community, as we found in taxonomic profiling results in this study. Therefore, As-, N- and S-metabolizing functional genes in metagenome are similar to the members from the dominant genera *Pseudomonas*, *Sphingobium* and *Acidovorax*, whereas them in Geochip is more taxonomically diverse (Table S5). It should be noted that the majority of metagenomic ORFs in sample LDS is functionally unknown (Figure S1 and Table S3), suggesting the novelty of sample LDS and highlighting that the relative abundance of functional genes may underestimated in metagenome. Geochip results cannot go beyond the detection capability of the probes fabricated on the array and thus may overestimate relative abundance of corresponding genes. Exactly for those reasons, the relative abundance of functional genes found in metagenomic analysis is much lower than that in Geochip 4.0 (Figure 2).

Binning of metagenomic. Three high-quality genomic bins belonging to *Pseudomonas stutzeri*, *Pseudomonas* and *Acidovorax* were derived from metagenomic data in sample LDS (Table S6). CheckM result showed their completeness of 96.5–99.3% and contamination of <1.1%. The genome sizes of bins range from 4.5 Mbp to 5.5 Mbp. The number of protein-coding genes is 4238, 5084 and 5092 respectively. These three bins encode all genes involved in glycolysis and citrate cycle, but lack key genes in reverse citrate cycle (rTCA cycle), reductive acetyl-CoA pathway and reductive pentose phosphate pathway (Table 1 and Figure 3), supporting the idea that these microorganisms
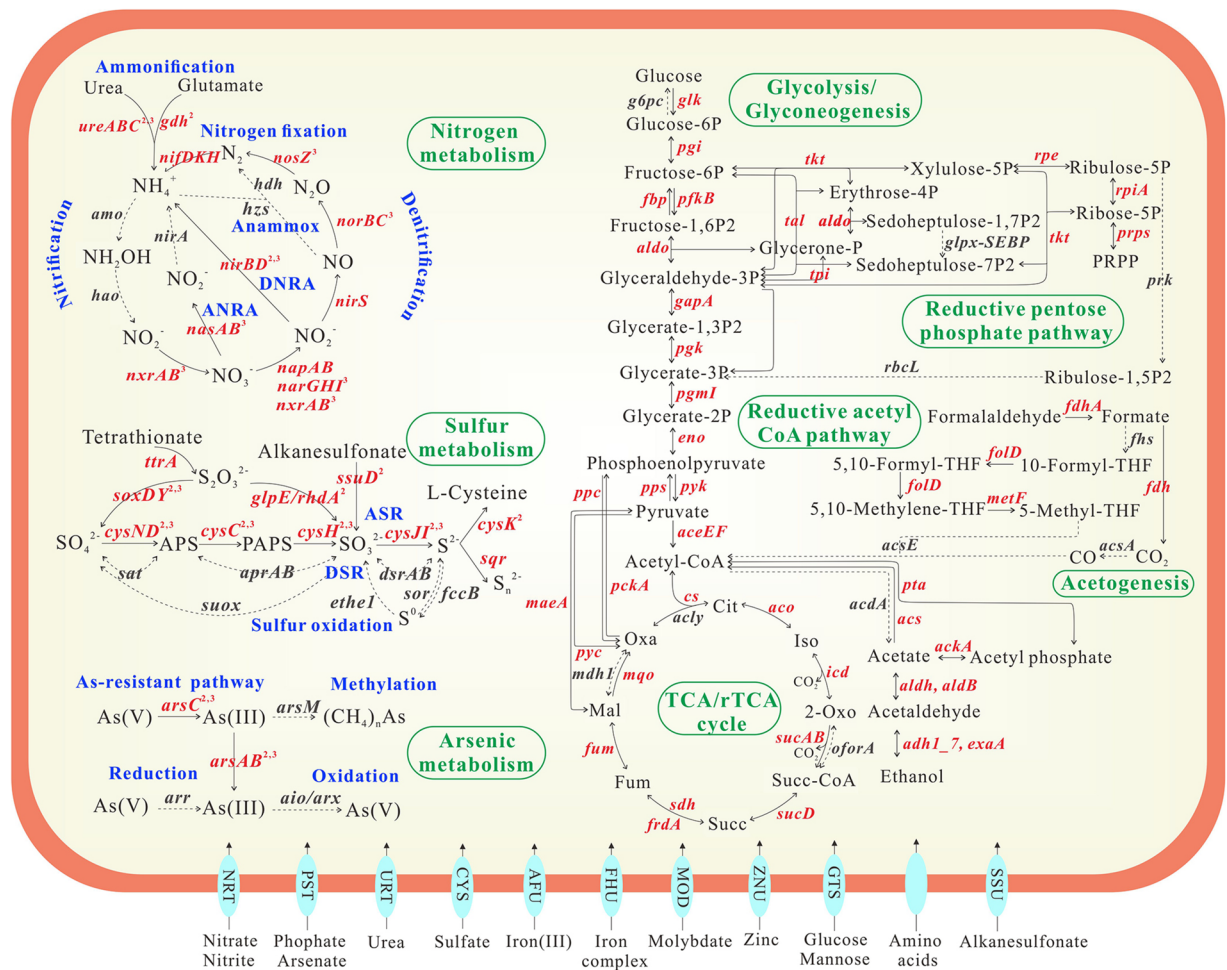
Figure 3. Overview of metabolic potentials in the genome bin 1. Genes related to nitrogen, arsenic, sulfur and carbon metabolisms and membrane transporters are shown. Genes in red font are detected and their corresponding reactions are highlighted by solid arrows. Genes in grey font are not detected and their corresponding reaction are displayed by dotted arrows. Gene abbreviations refer to Table S4 for details. Nitrogen-, arsenic- and sulfur-related genes marked by the superscripts of 2 and 3 indicate their presence in the genome of bin 2 and bin 3. Due to the same detection of genes from carbon metabolism and membrane transporters in all bins, the superscripts of 2 and 3 in these genes are omitted. ANRA: Assimilatory nitrate reduction to ammonia; DNRA: Dissimilatory nitrate reduction to ammonia; ASR: Assimilatory sulfate reduction; DSR: Dissimilatory sulfate reduction.

utilize organic carbon as energy and electron sources. To cope with As toxicity and sulfur deficiency in sample LDS, they all possess the pathway of As-resistant, assimilatory sulfate reduction and thiosulfate metabolism. Bin 1, *Pseudomonas stutzeri*, participate in other metabolic processes of nitrogen cycle except for anammox and the first step of nitrification. Bin 2, *Pseudomonas*, only perform ammonification and partial DNRA. Bin 3, *Acidovorax*, has capabilities of denitrification, DNRA, ammonification and the second step of nitrification. Dissimilatory arsenate reduction, arsenite oxidation, arsenite methylation, methane oxidation, methanogenesis, anammox, dissimilatory sulfate reduction, and sulfur oxidation were all absent in these bins, which is basically consistent with metagenomic and Geochip 4.0 functional analysis results (Figures 2 and 3).

Given that *Pseudomonas* dominate in microbial community and As-resistant bacteria in sample LDS (Table S5 and Figure 1), metabolic potentials of *Pseudomonas stutzeri* bin were further analyzed (Figure 3). The results show that *Pseudomonas stutzeri* could intake a variety of substrates, such as nitrate, nitrite, urea, phosphate, arsenate, sulfate, glucose and some amino acids, as well as iron, molybdate and zinc, some general cofactors of enzymes. These transporters of substrates provide a precipitate for intracellular metabolisms. The pathways of DNRA, denitrification, nitrogen fixation, ammonification, assimilatory sulfate reduction, glycolysis, citrate cycle and As-resistance are all found in the genome of *Pseudomonas stutzeri* bin. It was reported that *Pseudomonas stutzeri* is a well-known heterotrophic denitrifying bacterium, and some strains simultaneously possess the capacities of denitrification, nitrogen fixation and DNRA[64,72]. We thus propose that *Pseudomonas stutzeri* utilizes organic matter as energy and electron source to produce ammonium as well as reduce cytoplastic arsenate and pump-out arsenite to alleviate the stress of As, which leads to the accumulation of arsenite and ammonium in sample LDS. The co-existence of carbon-, nitrogen- and As-related corresponding genes in the genome of bin 2 and bin 3 supports this assumption
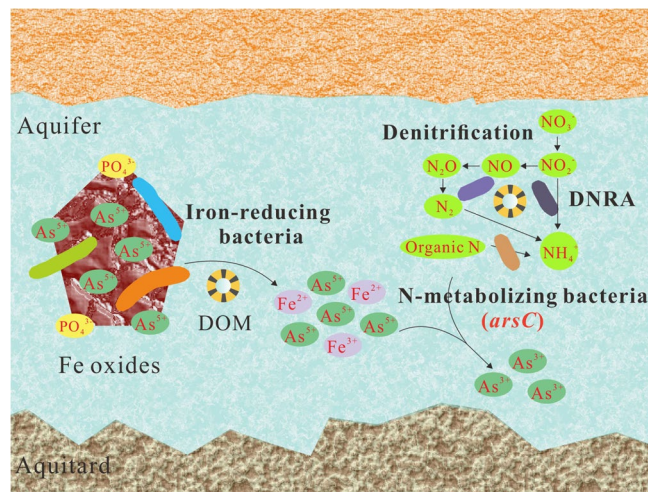
Figure 4. A schematic description of As mobilization mediated by microorganisms in groundwater sample LDS. OM refers to organic matter. Iron-reducing bacteria refer to *Geobacter*, *Ferribacterium* and other microorganism which are able to utilize iron as electron acceptor. N-metabolizing bacteria mainly refer to N-transforming bacteria featuring *ars* operon, such as *Pseudomonas stutzeri*.

(Table 1). In addition, a total of 11 complete genomes of *Pseudomonas stutzeri* strains retrieved from NCBI bacterial genome database also demonstrate that the As-resistant pathway is concomitant with denitrification, DNRA and ammonification (Table S7). Therefore, these ammonium-producing bacteria featuring *ars* operon may be main contributors to the transformation of arsenate to highly mobile arsenite in sample LDS.

Mechanism of As mobilization. The reductive dissolution of iron oxides is the main mechanism of As release in young groundwater systems with reducing condition[3,73]. Based on the microbial compositional and functional results from this study, we propose the pathways for arsenate release and reduction in sample LDS (Fig. 4). In this study, the iron-reducing bacteria, such as *Geobacter* and *Ferribacterium*, possibly initiate arsenate release by reductive dissolution of As-bearing iron oxides[2,13]. The competitive adsorption of phosphate may also contribute to a small fraction of dissociated arsenate[5]. The desorbed arsenate is subsequently reduced to highly mobile arsenite by some nitrogen-metabolizing bacteria featuring *ars* operon, such as *Pseudomonas stutzeri*[58,64]. Organic matter provide carbon and electron sources to fuel above biogeochemical processes[11,74]. As a consequence, high levels of ammonium, ferrous iron and arsenite were present in groundwater sample LDS (Table S1).

References

1. Rodríguezlado, L. *et al*. Groundwater arsenic contamination throughout China. *Science* **341**, 866–868 (2013).
2. Fendorf, S., Michael, H. A. & van Geen, A. Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. *Science* **328**, 1123–1127 (2010).
3. Wang, Y., Pi, K., Fendorf, S., Deng, Y. & Xie, X. Sedimentogenesis and hydrobiogeochemistry of high arsenic Late Pleistocene-Holocene aquifer systems. *Earth-Sci. Rev.* **189**, 79–98 (2019).
4. Guo, H., Wen, D., Liu, Z., Jia, Y. & Guo, Q. A review of high arsenic groundwater in Mainland and Taiwan, China: Distribution, characteristics and geochemical processes. *Appl. Geochem.* **41**, 196–217 (2014).
5. Michael, H. A. Geochemistry. An arsenic forecast for China. *Science* **341**, 852–853 (2013).
6. Neil, C. W., Yang, Y. J., Schupp, D. & Jun, Y. S. Water chemistry impacts on arsenic mobilization from arsenopyrite dissolution and secondary mineral precipitation: implications for managed aquifer recharge. *Environ. Sci. Technol.* **48**, 4395–4405 (2014).
7. Appleyard, S. J., Angeloni, J. & Watkins, R. Arsenic-rich groundwater in an urban area experiencing drought and increasing population density, Perth, Australia. *Appl. Geochem.* **21**, 83–97 (2006).
8. Biswas, A. *et al*. Role of competing ions in the mobilization of arsenic in groundwater of Bengal Basin: Insight from surface complexation modeling. *Water Res.* **55**, 30–39 (2014).
9. Kanematsu, M., Young, T. M., Fukushi, K., Green, P. G. & Darby, J. L. Arsenic(III, V) adsorption on a goethite-based adsorbent in the presence of major co-existing ions: Modeling competitive adsorption consistent with spectroscopic and molecular evidence. *Geochim. Cosmochim. Acta* **106**, 404–428 (2013).
10. Mailloux, B. J. *et al*. Advection of surface-derived organic carbon fuels microbial reduction in Bangladesh groundwater. *Proc. Natl. Acad. Sci. USA* **110**, 5331–5335 (2013).
11. Postma, D. Groundwater arsenic concentrations in Vietnam controlled by sediment age. *Nat. Geosci.* **5**, 656–661 (2012).
12. Muehe, E. M. *et al*. Arsenic(V) incorporation in vivianite during microbial reduction of arsenic(V)-bearing biogenic Fe(III) (oxyhydr)oxides. *Environ. Sci. Technol.* **50**, 2281–2291 (2016).
13. Mirza, B. S. *et al*. Arsenic(V) reduction in relation to Iron(III) transformation and molecular characterization of the structural and functional microbial community in sediments of a basin-fill aquifer in Northern Utah. *Appl. Environ. Microbiol.* **80**, 3198–3208 (2014).
14. Kocar, B. D., Herbel, M. J., Tufano, K. J. & Fendorf, S. Contrasting effects of dissimilatory iron (III) and arsenic (V) reduction on arsenic retention and transport. *Environ. Sci. Technol.* **40**, 6715–6721 (2006).
15. Guo, H. *et al*. Sulfur cycling-related biogeochemical processes of arsenic mobilization in the western Hetao Basin, China: Evidence from multiple isotope approaches. *Environ. Sci. Technol.* **50**, 12650–12659 (2016).
16. Sun, J., Quicksall, A. N., Chillrud, S. N., Mailloux, B. J. & Bostick, B. C. Arsenic mobilization from sediments in microcosms under sulfate reduction. *Chemosphere* **153**, 254–261 (2016).

17. Xie, X. *et al.* Multiple isotope (O, S and C) approach elucidates the enrichment of arsenic in the groundwater from the Datong Basin, northern China. *J. Hydrol.* **498**, 103–112 (2013).
18. Pi, K. *et al.* Remediation of arsenic-contaminated groundwater by *in-situ* stimulating biogenic precipitation of iron sulfides. *Water Res.* **109**, 337–346 (2017).
19. Burton, E. D., Johnston, S. G. & Kocar, B. D. Arsenic mobility during flooding of contaminated soil: the effect of microbial sulfate reduction. *Environ. Sci. Technol.* **48**, 13660–13667 (2014).
20. Kirk, M. F., Roden, E. E., Crossey, L. J., Brealey, A. J. & Spilde, M. N. Experimental analysis of arsenic precipitation during microbial sulfate and iron reduction in model aquifer sediment reactors. *Geochim. Cosmochim. Acta* **74**, 2538–2555 (2010).
21. Lear, G., Song, B., Gault, A. G., Polya, D. A. & Lloyd, J. R. Molecular analysis of arsenate-reducing bacteria within Cambodian sediments following amendment with acetate. *Appl. Environ. Microbiol.* **73**, 1041–1048 (2007).
22. Lu, X. *et al.* Molecular characterization of the total bacteria and dissimilatory arsenate-reducing bacteria in core sediments of the Jianghan Plain, central China. *Geomicrobiol. J.* **34**, 467–479 (2017).
23. Paul, D., Kazy, S. K., Gupta, A. K., Pal, T. & Sar, P. Diversity, metabolic properties and arsenic mobilization potential of indigenous bacteria in arsenic contaminated groundwater of West Bengal, India. *PLoS One* **10**, e0118735 (2015).
24. Li, X., Zhang, L. & Wang, G. Genomic evidence reveals the extreme diversity and wide distribution of the arsenic-related genes in Burkholderiales. *PLoS One* **9**, e92236 (2014).
25. Rosen, B. P. Biochemistry of arsenic detoxification. *FEBS Lett.* **529**, 86–92 (2002).
26. Maguffin, S. C., Kirk, M. F., Daigle, A. R., Hinkle, S. R. & Jin, Q. Substantial contribution of biomethylation to aquifer arsenic cycling. *Nat. Geosci.* **8**, 290–293 (2015).
27. Liao, V. H. *et al.* Arsenite-oxidizing and arsenate-reducing bacteria associated with arsenic-rich groundwater in Taiwan. *J. Contam. Hydrol.* **123**, 20–29 (2011).
28. Kurth, D. *et al.* Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Sci. Rep.* **7**, 1024 (2017).
29. Li, P. *et al.* Analysis of the functional gene structure and metabolic potential of microbial community in high arsenic groundwater. *Water Res.* **123**, 268–276 (2017).
30. Kurosawa, K. *et al.* Variation in arsenic concentration relative to ammonium nitrogen and oxidation reduction potential in surface and groundwater. *Commun. Soil Sci. Plant Anal.* **39**, 1467–1475 (2008).
31. Anawar, H. M. *et al.* Geochemical occurrence of arsenic in groundwater of Bangladesh: sources and mobilization processes. *J. Geochem. Explor.* **77**, 109–131 (2003).
32. Postma, D. *et al.* Arsenic in groundwater of the Red River floodplain, Vietnam: Controlling geochemical processes and reactive transport modeling. *Geochim. Cosmochim. Acta* **71**, 5054–5071 (2007).
33. Berg, M. *et al.* Magnitude of arsenic pollution in the Mekong and Red River Deltas—Cambodia and Vietnam. *Sci. Total Environ.* **372**, 413–425 (2007).
34. Weng, T. N., Liu, C. W., Kao, Y. H. & Hsiao, S. S. Isotopic evidence of nitrogen sources and nitrogen transformation in arsenic-contaminated groundwater. *Sci. Total Environ.* **578**, 167–185 (2017).
35. Jiao, J. J. *et al.* Abnormally high ammonium of natural origin in a coastal aquifer-aquitard system in the Pearl River Delta, China. *Environ. Sci. Technol.* **44**, 7470–7475 (2010).
36. Fukada, T., Hiscock, K. M., Dennis, P. F. & Grischek, T. A dual isotope approach to identify denitrification in groundwater at a river-bank infiltration site. *Water Res.* **37**, 3070–3078 (2003).
37. Kim, H., Kaown, D., Mayer, B., Lee, J. Y. & Lee, K. K. Combining pyrosequencing and isotopic approaches to assess denitrification in a hyporheic zone. *Sci. Total Environ.* **631–632**, 755–764 (2018).
38. Lin, N. F., Tang, J. & Bian, J. M. Characteristics of environmental geochemistry in the arseniasis area of the Inner Mongolia of China. *Environ. Geochem. Health* **24**, 249–259 (2002).
39. Cole, J. K. *et al.* Sediment microbial communities in Great Boiling Spring are controlled by temperature and distinct from water communities. *ISME J.* **7**, 718–729 (2013).
40. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
41. Kong, Y. Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152–153 (2011).
42. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
43. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
45. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
46. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, e132 (2010).
47. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
48. Hua, Z. S. *et al.* Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* **9**, 2832 (2018).
49. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
50. Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).
51. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
52. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
53. Tu, Q. *et al.* GeoChip 4: a functional gene array-based high throughput environmental technology for microbial community analysis. *Mol. Ecol. Resour.* **14**, 914–928 (2015).
54. Li, P. *et al.* Microbial community in high arsenic shallow groundwater aquifers in Hetao Basin of Inner Mongolia, China. *PLoS One* **10**, e0125844 (2015).
55. Guo, H., Yang, S., Tang, X., Li, Y. & Shen, Z. Groundwater geochemistry and its implications for arsenic mobilization in shallow aquifers of the Hetao Basin, Inner Mongolia. *Sci. Total Environ.* **393**, 131–144 (2008).
56. Wang, Y. *et al.* Coupled iron, sulfur and carbon isotope evidences for arsenic enrichment in groundwater. *J. Hydrol.* **519**, 414–422 (2014).
57. Buschmann, J. & Berg, M. Impact of sulfate reduction on the scale of arsenic contamination in groundwater of the Mekong, Bengal and Red River deltas. *Appl. Geochem.* **24**, 1278–1286 (2009).
58. Crognale, S. *et al.* Arsenic-related microorganisms in groundwater: a review on distribution, metabolic activities and potential use in arsenic removal processes. *Rev. Environ. Sci. Bio/Technol.* **16**, 647–665 (2017).

59. Tessler, M. *et al.* Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* **7**, 6589 (2017).
60. Tamames, J., de la Pena, S. & de Lorenzo, V. COVER: a priori estimation of coverage for metagenomic sequencing. *Environ. Microbiol. Rep.* **4**, 335–341 (2012).
61. Héry, M. *et al.* Molecular and cultivation-dependent analysis of metal-reducing bacteria implicated in arsenic mobilisation in south-east asian aquifers. *Appl. Geochem.* **23**, 3215–3223 (2008).
62. Islam, F. S. *et al.* Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature* **430**, 68–71 (2004).
63. Grabovich, M. *et al.* Proposal of *Giesbergeria voronezhensis* gen. nov., sp. nov. and *G. kuznetsovii* sp. nov. and reclassification of [*Aquaspirillum*] *anulus*, [*A.*] *sinuosum* and [*A.*] *giesbergeri* as *Giesbergeria anulus* comb. nov., *G. sinuosa* comb. nov. and *G. giesbergeri* comb. nov., and [*Aquaspirillum*] *metamorphum* and [*A.*] *psychrophilum* as *Simplicispira metamorpha* gen. nov., comb. nov. and *S. psychrophila* comb. nov. *Int. J. Syst. Evol. Microbiol.* **56**, 569–576 (2006).
64. Lalucat, J., Bennasar, A., Bosch, R., Garcia-Valdes, E. & Palleroni, N. J. Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* **70**, 510–547 (2006).
65. Révész, F. *et al.* *Sphingobium aquiterrae* sp. nov., a toluene, meta-and para-xylene-degrading bacterium isolated from petroleum hydrocarbon-contaminated groundwater. *Int. J. Syst. Evol. Microbiol.* **68**, 2807–2812 (2018).
66. Andres, J. & Bertin, P. N. The microbial genomics of arsenic. *FEMS Microbiol. Rev.* **40**, 299–322 (2016).
67. Wang, Y. *et al.* Abundance and diversity of methanogens: Potential role in high arsenic groundwater in Hetao Plain of Inner Mongolia, China. *Sci. Total Environ.* **515–516**, 153–161 (2015).
68. Wang, S., Zhu, G., Peng, Y., Jetten, M. S. & Yin, C. Anammox bacterial abundance, activity, and contribution in riparian sediments of the Pearl River estuary. *Environ. Sci. Technol.* **46**, 8834–8842 (2012).
69. Oshiki, M., Satoh, H. & Okabe, S. Ecology and physiology of anaerobic ammonium oxidizing bacteria. *Environ. Microbiol.* **18**, 2784–2796 (2016).
70. Zhou, J. *et al.* High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *MBio* **6**, e02288–14 (2015).
71. Roh, S. W., Abell, G. C., Kim, K. H., Nam, Y. D. & Bae, J. W. Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends Biotechnol.* **28**, 291–299 (2010).
72. Yan, Y. *et al.* Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. *Proc. Natl. Acad. Sci. USA* **105**, 7564–7569 (2008).
73. Stuckey, J. W. *et al.* Arsenic release metabolically limited to permanently water-saturated soil in Mekong Delta. *Nat. Geosci.* **9**, 70–76 (2015).
74. Ghosh, D., Routh, J. & Bhadury, P. Characterization and microbial utilization of dissolved lipid organic fraction in arsenic impacted aquifers (India). *J. Hydrol.* **527**, 221–233 (2015).

Acknowledgements

Author Contributions

Z.J. and P.L. conceived and designed the experiments. C.Y. and H.W. contributed to field sampling. H.L. and D.W. conducted the experiment. Z.J. and Y.W. analyzed sequencing data. Z.J. wrote the draft of the manuscript. P.L. contributed to the revisions of this manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-49365-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.