Microbiome

RESEARCH

Open Access

Various microbial taxa couple arsenic transformation to nitrogen and carbon cycling in paddy soils

Xin-Di Zhao[1], Zi-Yu Gao[1], Jingjing Peng[2], Konstantinos T. Konstantinidis[3] and Si-Yu Zhang[1*]

Abstract

Background Arsenic (As) metabolism pathways and their coupling to nitrogen (N) and carbon (C) cycling contribute to elemental biogeochemical cycling. However, how whole-microbial communities respond to As stress and which taxa are the predominant As-transforming bacteria or archaea in situ remains unclear. Hence, by constructing and applying ROCker profiles to precisely detect and quantify As oxidation (*aioA*, *arxA*) and reduction (*arrA*, *arsC1*, *arsC2*) genes in short-read metagenomic and metatranscriptomic datasets, we investigated the dominant microbial communities involved in arsenite (As(III)) oxidation and arsenate (As(V)) reduction and revealed their potential pathways for coupling As with N and C in situ in rice paddies.

Results Five ROCker models were constructed to quantify the abundance and transcriptional activity of short-read sequences encoding As oxidation (*aioA* and *arxA*) and reduction (*arrA*, *arsC1*, *arsC2*) genes in paddy soils. Our results revealed that the sub-communities carrying the *aioA* and *arsC2* genes were predominantly responsible for As(III) oxidation and As(V) reduction, respectively. Moreover, a newly identified As(III) oxidation gene, *arxA*, was detected in genomes assigned to various phyla and showed significantly increased transcriptional activity with increasing soil pH, indicating its important role in As(III) oxidation in alkaline soils. The significant correlation of the transcriptional activities of *aioA* with the *narG* and *nirK* denitrification genes, of *arxA* with the *napA* and *nirS* denitrification genes and of *arrA*/*arsC2* with the *pmoA* and *mcrA* genes implied the coupling of As(III) oxidation with denitrification and As(V) reduction with methane oxidation. Various microbial taxa including *Burkholderiales*, *Desulfatiglandales,* and *Hyphomicrobiales* (formerly *Rhizobiales*) are involved in the coupling of As with N and C metabolism processes. Moreover, these correlated As and N/C genes often co-occur in the same genome and exhibit greater transcriptional activity in paddy soils with As contamination than in those without contamination.

Conclusions Our results revealed the comprehensive detection and typing of short-read sequences associated with As oxidation and reduction genes via custom-built ROCker models, and shed light on the various microbial taxa involved in the coupling of As and N and C metabolism in situ in paddy soils. The contribution of the *arxA* sub-communities to the coupling of As(III) oxidation with nitrate reduction and the *arsC* sub-communities to the coupling of As(V) reduction with methane oxidation expands our knowledge of the interrelationships among As, N, and C cycling in paddy soils.

Keywords Arsenite oxidation genes, Arsenate reduction genes, Denitrification genes, Dissimilatory nitrate reduction genes, Methane cycling genes, Paddy soil, ROCker model

*Correspondence:
Si-Yu Zhang
syzhang@des.ecnu.edu.cn
Full list of author information is available at the end of the article

Background

The coupling of arsenic (As) oxidation and reduction with nitrogen (N) and carbon (C) metabolism plays a critical role at the global level by influencing their biogeochemical cycles and ecosystem health. This coupling facilitates the transformation and mobility of As in various environments, particularly in paddy fields, where anaerobic conditions predominate. Owing to the anaerobic conditions in paddy fields during rice cultivation, As is dominated by more mobilizable and bioavailable inorganic As species, i.e., arsenite (As(III)) and arsenate (As(V)) [1]. Microbial-mediated As oxidation and reduction, which include respiratory As(III) oxidation, respiratory As(V), and detoxification As(V) reduction, are important for the coupling of As, C, and N metabolism [2–6].

Microbial respiratory As(III) oxidation is catalyzed by either of two periplasmic soluble As(III) oxidases, AioA and ArxA [7]. As(III)-oxidizing bacteria carrying *aioA* genes have been isolated from different environments and are widespread among microorganisms, including members of *Pseudomonadota*, *Bacteroidota*, and *Actinomycetota* [8]. ArxA is a clade that is distinct from AioA in the DMSO reductase family of enzymes but has greater similarity to respiratory As(V) reductase (ArrA) [9]. The verified *arxA*-carrying bacteria are mainly isolated from soda lakes and belong to *Gammaproteobacteria* [9–11]. To date, only one *arxA*-carrying strain, i.e., the *Noviherbaspirillum* strain HC18, which also carries the *aioA* gene, has been identified in paddy fields [12]. For As(V) reduction, both the respiratory As(V) reduction mediated by *arrA* genes and the detoxification As(V) reduction mediated by *arsC* genes are important As metabolism pathways for bacteria and archaea [7, 13]. Although the *arrA* gene is restricted to bacteria and archaea [14, 15], the *arsC* gene evolved before the Great Oxidation Event and is widespread in bacteria, archaea, and eukaryotes [16, 17]. In bacteria and archaea, ArsC reductases are classified into the glutaredoxin family (ArsC1) and low-molecular-weight phosphatases (ArsC2) [17–19]. Genes responsible for As oxidation and reduction, including the *aioA*, *arrA*, and *arsC* genes, are widely distributed in paddy soils [20], and microbially mediated As oxidation and reduction are closely associated with the biogeochemical processes of nonmetal nitrogen (N) and carbon (C) [7].

The coupling processes of As(III) oxidation with denitrification and dissimilatory nitrate (NO_3^-) reduction to ammonium (NH_4^+; DNRA) have been reported to be mediated by *aioA* sub-communities, including *Azoarcus*, *Rhodanobacter*, *Pseudomonas*, *Burkholderiales*, *Aromatoleum*, *Paenibacillus*, *Microvirga*, *Herbaspirillum*, *Bradyrhizobium*, and *Azospirillum*-related bacteria [5, 21–23]. In flooded paddy soils, amendment with NO_3^- effectively enhanced As(III) oxidation and increased the relative abundance of *aioA* genes [2, 3, 24]. As(III)-oxidizing isolates from paddy soils carrying both *aioA* and denitrification genes are capable of transferring electrons from As(III) oxidation to denitrification, suggesting the coupling of As(III) oxidation and NO_3^- reduction [2, 25, 26]. Moreover, *Burkholderiales* containing both the *aioA* and *nrfA* genes were recently identified in As-contaminated paddy soil, indicating their important role in the coupling process of DNRA and As(III) oxidation [23]. However, whether the *arxA* sub-communities, which are also capable of As(III) oxidation, contribute to the coupling of As(III) oxidation with N metabolism has rarely been explored.

With respect to As(V) reduction, respiratory As(V) reduction mediated by *arrA* genes has been reported to be coupled with aerobic and anaerobic methane (CH_4) oxidation in wetlands and paddy soils [4, 6, 27, 28]. Aerobic methanotrophs in cooperation with *Burkholderiaceae* are implicated in the coupling of As(V) reduction to aerobic CH_4 oxidation [6]. Moreover, anaerobic methane-oxidizing archaea (ANMEs) have been shown to involve in the coupling of As(V) reduction with anaerobic CH_4 oxidation, either independently or in combination with As(V)-reducing bacteria harboring *arrA* genes [4]. However, these studies seldom investigated the detoxification of As(V) mediated by *arsC* genes, which constitute a significant proportion of As(V) reduction communities in wetland environments [20, 29]. In addition, previous studies showing the coupling of As oxidation and reduction with N and C metabolism are primarily based on cultivation experiments with high-level amendments of either NO_3^- or CH_4 [2–4, 6], which facilitate the coupling processes. Considering the complex environmental factors contributing to microbial community composition in situ in rice paddies, the extent to which these As oxidation and reduction processes are coupled with N and C metabolism in situ and the microbial communities involved in As oxidation and reduction and the coupled process remain unclear.

Owing to the varied levels of sequence similarity among As oxidation and reduction genes, standard analyses based on similarity searches are limited in defining robust cut-off values for the detection of As oxidation and reduction genes in short-read datasets. A recent bioinformatic tool, ROCker, can effectively overcome this limitation and has shown more accurate performance than traditional fixed cut-offs (e.g., *e*-value of 1e − 5) [30, 31]. In this study, we developed customized As oxidation and reduction gene ROCker models to quantify the short-read sequences carrying As oxidation and reduction genes in metagenomic and metatranscriptomic datasets. The coupling of As oxidation and reduction reactions with N and C metabolism was further evaluated at the whole-community level (total microbial communities) to elucidate how various microbial taxa couple As transformation to N and C cycling in situ in paddy soils. Additionally, microbial genomes were binned from the metagenomic datasets, and the transcriptional activity of genes responsible for the coupling process to N- and C-transformed genes under different As gradients was investigated. We aimed to answer the following questions: (1) what are the dominant microbial communities responsible for As oxidation and reduction reactions in paddy soils? (2) Which of the key As oxidation and reduction sub-communities (functional communities carrying either of the As redox genes) are coupled with nitrogen and carbon metabolism? (3) Does the level of As contamination affect the coupling of As oxidation and reduction and nitrogen and carbon metabolism?

Materials and methods

Paddy soil sample collection and analysis

From July 2022 to June 2023, a total of 36 samples, three replicates from the same field at each site, were collected from 12 distinct paddy fields in South China with different As levels, including 8 As-contaminated (total As $\geq$ 15 mg kg^{-1}, abbreviated as AsContam) sites and 4 As-noncontaminated (total As $<$ 15 mg kg^{-1}, abbreviated as NoContam) sites (Table S1). The surface soil (0 – 20 cm) was sampled with an ethanol-sterilized shovel, stored in sterile plastic bags, and transported to the laboratory on ice. The soils used for physiochemistry analysis were stored in a refrigerator at 4 °C, whereas the other samples were stored at $-$80 °C for microbiological analysis. For geochemical analysis, the soil pH, WH_2O (%) (soil moisture content), total concentrations of carbon (TC), nitrogen (TN), phosphorus (TP), sulfur (TS), iron (Fe), As, lead (Pb), cadmium (Cd), chromium (Cr), antimony (Sb), copper (Cu), organic matter (OM), NH_4^+ and NO_3^- were measured in the laboratory (Table S2). As levels were classified according to the National Soil Environmental Quality Standard of China (15 mg kg^{-1} in GB-15168–1995). The detailed methods are listed in the supplementary material.

Soil DNA and RNA extraction and metagenomic and metatranscriptomic sequencing

Total DNA and RNA were extracted from 2 g of well-mixed soil using the RNeasy PowerSoil Total RNA Kit (Qiagen) and RNeasy PowerSoil DNA Elution Kit (Qiagen), following the instructions of the manufacturers. DNA and RNA concentrations were measured via a NanoDrop 2000 spectrophotometer (Thermo Scientific, MA, USA). The purity of the DNA and RNA was checked via gel electrophoresis and an Agilent 5300 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), respectively. Paired-end DNA and cDNA libraries (2 × 150 bp) were constructed via NEXTFLEX Rapid DNA-Seq (Bio Scientific, Austin, TX, USA) and Illumina® Stranded mRNA Prep, Ligation (Illumina, San Diego, CA, USA), respectively, following the manufacturer's guidelines. The DNA and cDNA libraries were sequenced on a NovaSeq 6000 sequencer (Illumina, San Diego, CA, USA) via the paired-end sequencing method at Majorbio Bio-Pharm Technology Company in Shanghai, China, and 36 metagenomic and 36 metatranscriptomic datasets were generated, respectively.

Trimming, rRNA removal, assembly, and taxonomic classification of metagenomic and metatranscriptomic data

Quality control of the raw paired-end reads from the metagenomic (metaG) and metatranscriptomic (metaT) data was carried out with multitrim (v1.2.6) (https://github.com/Kgerhardt/multitrim) with the default settings. The coverage and diversity of these metagenomes were evaluated via Nonpareil (v3.303) [32]. For the metatranscriptomes, SortMeRNA (v4.3.4) [33] with the option "--paired_in" was used to eliminate rRNA. The basic information (e.g., results of trimming, nonpareil coverage, and assembly) of the metagenomic and metatranscriptomic sequencing is listed in Tables S3 and S4. The clean reads from the metagenomics and metatranscriptomics were classified at the phylum and genus levels using Kraken 2 (v2.0.7) [34] with the setting of "--confidence 0.05". The genome equivalent for each metagenomic sample was estimated by MicrobeCensus (v1.1.1) [35] with default parameters.

Building the aioA, arxA, arrA, arsC1, and arsC2 gene databases and ROCker models

The workflow of the As oxidation and reduction gene database and ROCker model construction and evaluation are detailed in the supplementary material. Briefly, there are 3 main steps for database and ROCker model building: (i) collection of target (positive) and nontarget (negative) reference sequences: for the *aioA*, *arxA*, *arrA*, *arsC1*, and *arsC2* genes, a total of 167, 41, 151, 490 and 719 positive reference protein sequences, respectively, were obtained from the literature [15, 17, 36] and the UniRef90 database [37] after checking for the protein sequence length, characteristic amino acids, protein motifs and phylogenetic relationships to the experimentally verified references. These positive references were also used as As oxidation and reduction gene databases for BLAST-based searches. For ROCker building, a list of negative references (i.e., nontarget references) that were evolutionarily related in the phylogenetic tree but had completely different functions were further screened. (ii) ROCker model construction and evaluation: ROCker models for the five As oxidation and reduction genes were developed on the basis of positive and negative references (Table S5 and Fig. S1). The approach used for ROCker building and evaluation is described in the supplementary material. (iii) Comparison of ROCker models with DIAMOND BLASTx search versus different fixed cut-offs: The relative abundance of As oxidation and reduction genes detected by ROCker was compared with that detected by DIAMOND (v2.0.14.152) [38] BLASTx search with a minimum cut-off for a match of amino acid identities of 70%, 80%, and 90% (abbreviated as id70, id80, and id90, respectively), along with a minimum alignment length of 25 amino acids and a maximum *e*-value of $1e-5$.

Estimation of As oxidation and reduction and nitrogen and carbon metabolism gene abundance and activity on the basis of metagenomic and metatranscriptomic reads

The functional genes involved in As, C and N metabolism were annotated using DIAMOND BLASTx (with the option "--sensitive") against the custom As oxidation and reduction gene databases, MCycDB [39], NCycDB [40] and the denitrification gene ROCker databases [*narG*, *nirK*, *norB,* and *nosZ*] (http://enve-omics.ce.gatech.edu/rocker/models). The best matches were subsequently sorted using the script BlastTab.best_hit_sorted.pl (http://enve-omics.ce.gatech.edu/enveomics/docs?t=BlastTab.best_hit_sorted.pl) [41] and then filtered either by ROCker or by a minimum cut-off for a match of 80% identity, alignment length of 25 amino acids and *e*-value of $1e-5$. For metagenomics and metatranscriptomics, the relative abundance of target genes was normalized by RPKG [i.e., copies per cell, RPKG = (the number of reads mapped to gene)/(gene length in kb)/(genome equivalents)] and RPKM [i.e., reads per kb per millions of reads, RPKM = (the number of reads mapping to gene)/(gene length in kb)/(the total number of reads after removal of rRNA reads)], respectively [42]. The transcriptional activity of the genes was estimated as the relative abundance of genes in the metaG (RPKM) normalized to the relative abundance of the same gene identified in the metaT (RPKM) (i.e., metaT/metaG RPKM).

Phylogenetic placement of As metabolism gene-harboring reads

Five As oxidation and reduction protein trees were constructed with amino acid sequences from positive protein sequences for building the ROCker models. These protein sequences were first aligned via the program MAFFT (v7.490) [43] with the "--auto" option. Phylogenetic trees were subsequently constructed with the program RAxML (v8.2.12) [44] (-f a, -N 100, -m PROTGAMMAWAG). The AioA, ArxA, ArrA, ArsC1, and ArsC2 trees were used as reference trees for the phylogenetic placement of metagenomic and metatranscriptomic reads. Specifically, the identified As oxidation and reduction protein-containing reads were added to the corresponding nucleotide reference alignments using MAFFT [43] with the parameter "--addfragments". The nucleotide and corresponding protein reference used the same ID during alignment. These reads were placed in the corresponding phylogenetic tree of different As oxidation and reduction gene trees via RAxML (-f v, -G 0.2) [44] and then visualized by iTOL [45] after the resulting jplace file was processed with Jplace.to_Itol.rb (http://enve-omics.ce.gatech.edu/enveomics/docs?t=Jplace.to_iToL.rb) [41].

Binning and pathway annotation of the metagenomic assembled genomes

Metagenomic contigs ≥ 500 bp in length were selected for coassembly (i.e., contigs from 3 replicated samples were caught together for assembly) via MEGAHIT (v1.2.9) [46]. Metagenome-assembled genomes (MAGs) were recovered from contigs with lengths ≥ 1 kb using MetaWRAP (v1.2.1) [47] and dereplicated by dRep (v3.3.0) [48] with an ANI > 95%. The completeness and contamination of the recovered MAGs were evaluated by CheckM (v1.1.3) [49]. Only the MAGs with completeness ≥ 50% and contamination ≤ 10% were considered for further analysis. Taxonomic classification and phylogenetic analysis of the obtained MAGs were conducted with the inferred module of GTDB-Tk [50]. TvBOT (v2.5.0) [51] was used to visualize the phylogenetic tree. The relative abundance of MAG in different samples was estimated by coverage per genome equivalent [CPG = (sum of the sequencing depth of contigs in interested MAG)/(number of contigs in interested MAG)/(genome equivalent)] and is detailed in the supplementary material. Open reading frames (ORFs) were predicted from dereplicated MAGs using prodigal (v2.6.3) [52] with the "-p meta" option. The functional genes involved in As oxidation and reduction and C and N metabolism were annotated via DIAMOND BLASTp against the custom As genes ROCker database, MCycDB [39] and NCycDB [40] after filtering by a minimum cut-off for a match of 35% identity, *e*-value of $1e-5$ and query length coverage of 70%.

Statistical analyses and data visualization

The trends in As concentration with respect to the relative abundance of As oxidation and reduction genes were analyzed by linear regression based on the basis of the Pearson correlation coefficients and visualized using the ggplot2 (v3.5.1) [53] package. The significance of the differences in soil geochemical variables and functional gene abundance and activity was calculated using two-sided Wilcoxon rank-sum tests at $p < 0.05$ and visualized via the ggsignif (v0.6.4) [54] package. With the DESeq2 (v1.44.0) [55] package, the differentially abundant metagenomic and metatranscriptomic taxa between As-contaminated and As-noncontaminated samples were determined and visualized by the pheatmap (v1.0.12) package (https://github.com/raivokolde/pheatmap). The potential correlations among As and N metabolism gene abundance and activity at the DNA and RNA levels were estimated by Spearman correlation analyses. To assess the relationships between the soil properties and As gene communities, a partial Mantel test was performed via the linkET (v 0.0.7.4) package (https://github.com/Hy4m/linkET). The data and codes for making maps of China, Hunan, Zhejiang, Yunnan, and Guangdong Provinces were obtained from https://github.com/EasyChart/Beautiful-Visualization-with-R [56]. All of the analyses and plots were performed in R version 4.1.3.

Results

Comparison of As oxidation and reduction ROCker models and similarity searches with DIAMOND BLASTx

In total, five ROCker models for As(III) oxidation (*aioA* and *arxA*) and As(V) reduction (*arrA*, *arsC1*, and *arsC2*) genes were constructed. The sensitivity, specificity, and accuracy, which indicate the performance of the ROCker models, ranged from 92.04 to 99.85%, 99.59 to 100%, and 99.50 to 99.99%, respectively, for the 150-bp ROCker models (Fig. 1, Fig. S2). Compared with the DIAMOND BLASTx search with minimum identities of 80% and 90%, ROCker revealed a relatively greater abundance of As oxidation and reduction genes than did the similarity search, whereas compared with the minimum identity of 70%, ROCker identified a relatively lower abundance of As oxidation and reduction genes (Fig. 1, Fig. S2). The results obtained by ROCker and the commonly suggested threshold (minimum amino acid identity of 90%) were further validated by placing the AioA-, ArxA-, ArrA-, ArsC1-, and ArsC2-carrying reads in the metagenomes

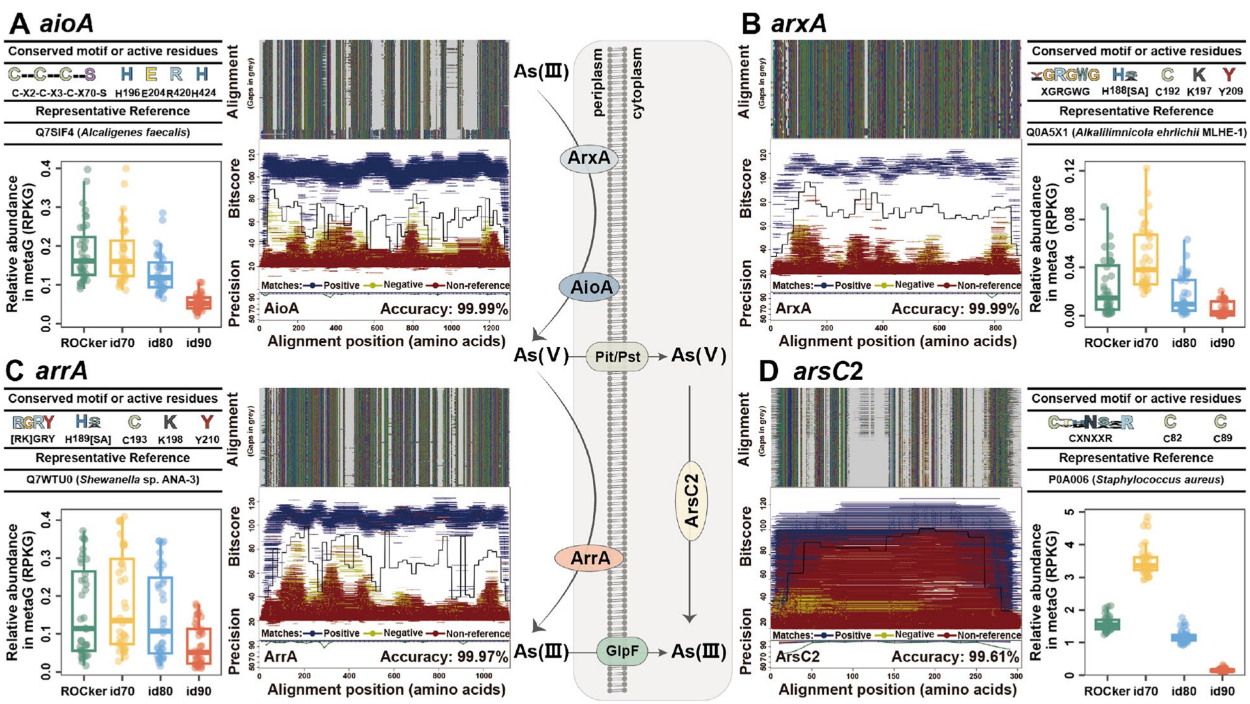

Fig. 1 150-bp ROCker models for **A** *aioA*, **B** *arxA*, **C** *arrA*, and **D** *arsC2* and comparison of ROCker models versus different fixed cut-offs: id70, id80, and id90. Three parts of the ROCker plots show: (TOP) Positive reference sequence alignments with amino acids in different colors and gaps in light grey are displayed in the top panel. (Middle) Bit scores (*y*-axis) of the hits (i.e., the matches from simulated 150-bp read datasets) from positive references (blue), negative references (yellow), and nontarget sequences (red). The solid black traversing line represents the calculated ROCker best bit score thresholds for consecutive windows of variable length. (Bottom) Summary statistics on the performance of each window are based on true/false positives (TP and FP, respectively) and true/false negatives (TN and FN, respectively). The accuracy (i.e., [TP + TN]/[TP + FP + TN + FN]) of ArxA and AioA was 99.99%, that of ArsC2 was 99.61%, and that of ArrA was 99.97%. Boxplot components: centerline, median values; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers

(DNA reads) on their representative phylogenetic clade (Fig. S3). Overall, ROCker and a similarity search with a minimum identity of 90% (id90) identified the exact sequence variant carried by the reads. However, the ratio of ROCker/id90 reads assigned to the same reference clade was relatively greater for all five As oxidation and reduction genes, especially for ArsC1 and ArsC2. Specifically, a greater percentage of ROCker/id90 reads were observed for approximately 76.05% to 97.95% of the total clades for the AioA, ArxA, ArrA, and ArsC proteins. Moreover, a wider range of clades was detected by ROCker than by a minimum identity of 90% for the five As genes, suggesting that the more comprehensive detection and typing of short-read sequences carrying As oxidation and reduction genes was achieved by ROCker.

Overview of soil properties and metagenomic and metatranscriptomic datasets

In total, 36 paddy soil samples were collected from 12 distinct paddy fields located in Hunan (CD_BY, CS_LH, CS_YCP, CZ_LT, HN_HH, HN_LY, HY_SKS, ZJJ_DYG, and ZJJ_SH), Yunnan (YN_MG), Zhejiang (ZJ_SY) and Guangdong (GD_SG) Provinces (Fig. S4 and Table S1). The pH of these paddy soils ranged from 4.6 to 8.0, with 22 of the samples having a soil pH < 6.5 (acid) and 14 of the samples having a soil pH ≥ 6.5 (neutral/alkaline; 9 out of the 14 samples acquired pH > 7.5; Table S2). These soil samples were classified into As-contaminated and As-noncontaminated groups according to the National Soil Environmental Quality Standard of China (15 mg kg^{-1} in GB-15168–1995), with significant ($p < 0.001$; Fig. S5) variations in the total As concentrations (average of 37.22 vs. 8.21 mg kg^{-1}). The total concentrations of Pb and Cu, as well as the WH_2O (%) were also significantly greater in As-contaminated paddy soils than in As-noncontaminated paddy soils ($p < 0.001$; Fig. S5). However, the total concentrations of Sb, Cd, Cr, NH_4^+, NO_3^-, OM, Fe, C, N, P, and S and the pH did not significantly differ between the As-contaminated and As-noncontaminated paddy soils (Fig. S5, Table 1). For the metagenomic and metatranscriptomic datasets, an average of 12.1 Gbp and 18.7 Gbp of clean data were acquired after trimming and/or removing the rRNA reads from each sample, respectively (Tables S3 and S4).

Variation in As oxidation and reduction gene abundance and transcriptional activity in paddy soils with or without As contamination

As(III) oxidation in these paddy soils was predominantly mediated by the *aioA* genes rather than by the *arxA* genes, as evidenced by the significantly ($p < 0.001$) greater abundance (0.18 vs. 0.02 RPKG; Fig. S6A) and transcriptional activity observed for the *aioA* genes than for the *arxA* genes (34.71 vs. 9.85; Fig. S6B). For the *aioA* genes, both the relative gene abundance and transcriptional activity were significantly ($p < 0.01$) greater in As-contaminated paddy soils than in those without As contamination (Fig. 2A). Moreover, the abundance and transcriptional activity of the *aioA* genes significantly ($p < 0.05$) correlated with the As concentration in the paddy soils (Fig. 2A). For the *arxA* genes, no significant differences were detected in terms of gene abundance or transcriptional activity between As-contaminated and As-noncontaminated soils.

In terms of As(V) reduction, the sub-communities carrying the *arsC2* gene were dominant in this process and presented significantly ($p < 0.001$) greater gene abundance (1.57 RPKG) and transcriptional activity (90.74) than did both the *arrA* (0.16 RPKG for gene abundance and 29.26 for transcriptional activity) and *arsC1* (0.39 RPKG for gene abundance and 39.80 for transcriptional activity) genes (Fig. S6). The abundance of the *arsC2* gene was significantly ($p < 0.01$) greater in As-contaminated soils than in As-noncontaminated soils and was significantly ($p < 0.05$) correlated with the As concentration in paddy soils (Fig. 2B). For the abundance and transcriptional activity of the *arrA* and *arsC1* genes, no significant differences were detected between the As-contaminated and As-noncontaminated soils. Nevertheless, the transcriptional activity of all the respiratory and detoxification As(V) reduction genes was significantly correlated with the As concentration in the paddy soils (Fig. 2B, Fig. S7). Consistently, most of the bacterial communities (metagenomic/metatranscritomic) with significantly increased abundance in As-contaminated paddy soils

Table 1 Soil properties of As-contaminated and As-noncontaminated paddy fields

As level	pH	WH_2O(%)	mg kg^{-1}			g kg^{-1}						
			As	Pb	Cu	NH_4^+-N	NO_3^--N	OM	TC	TN	TP	TS
AsContam	6.2±0.2 a	35.7±0.8 a	37.2±4.3 a	112±21 a	34.5±2.1 a	10.4±2.2 a	16.6±3.4 a	35.9±3 a	27.0±1.6 a	2.2±0.1a	0.8±0.1a	0.3±0.03 a
NoContam	6.5±0.3 a	29.6±2.4 b	8.2±0.4 b	30.8±1.4 b	26.1±2.3 b	3.7±1.7 a	15.7±6.1 a	43.0±5.4 a	30.7±1.6 a	2.1±0.1a	0.7±0.1a	0.4±0.01 a

Mean ± standard error of samples from As-contaminated and As-noncontaminated paddy fields. Different letters indicate significant differences between groups (Wilcoxon rank-sum tests, $p < 0.05$)

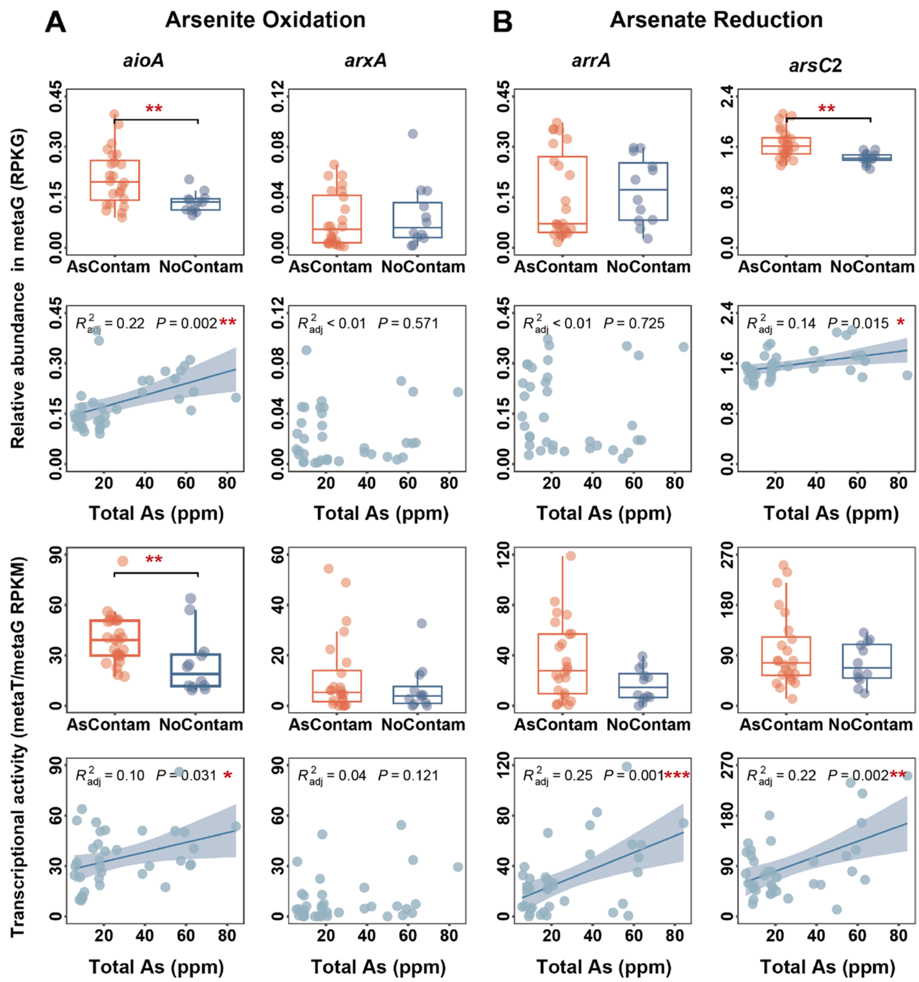

Fig. 2 Relative abundance/activity of As oxidation and reduction genes in paddy soils. Relative abundance and transcriptional activity of the **A** *aioA*, *arxA*, **B** *arrA* and *arsC2* genes in two paddy soil types and their correlation with As concentration. Each point represents an individual sample. The solid line indicates the linear regression of the abundance/activity of As genes and As concentrations. The light blue areas indicate the 95% confidence intervals. The relative abundance of As oxidation and reduction genes in the metagenomes and transcriptional activity were evaluated by the RPKG and metaT/metaG RPKM as described in the "Materials and methods" section. The significance values for two-sided Wilcoxon rank-sum tests and Pearson correlations are denoted as follows: (*, $p < 0.05$; **, $0.001 < p < 0.01$; ***, $p < 0.001$)

(DESeq; adjusted $p < 0.05$) were significantly ($p < 0.05$) positively correlated with the *aioA* and *arsC2* genes but negatively correlated with the *arrA* and *arxA* genes in terms of relative abundance ($p < 0.05$; Fig. S8).

Community-wide response of As(III)-oxidizing and As(V)-reducing bacteria and archaea to As contamination in paddy soils

Overall, approximately 61.43–97.37% of the total clades of the phylogenetic tree of the AioA, ArxA, ArrA, ArsC1, and ArsC2 proteins recruited reads from metatranscriptomic datasets, indicating that bacteria and archaea involved in the As oxidation and reduction response to As contamination represented community-wide activity rather than that of several specific species (Fig. 3, Fig. S9). Among the classified As oxidation and reduction bacteria and archaea with transcriptional activity, *Pseudomonadota* was the predominant phylum in both the *aioA* and *arxA* sub-communities, accounting for 89.09% and 97.33%, respectively, of the total identified reads (Fig. 3A, B, Table S6). At the genus level, *Thiobacillus* (4.79%), *Cupriavidus* (4.01%) and *Mesorhizobium* (3.48%) represented the dominant *aioA* sub-communities in As-contaminated soils, whereas *Methylocystis* (6.39%), *Rubrivivax* (4.05%), and *Burkholderia* (3.36%) were the dominant *aioA* sub-communities in As-noncontaminated soils (Fig. S10A). *arxA* sub-communities were dominated by *Aromatoleum* (12.94% vs. 14.20%), *Tepidimonas* (12.64% vs. 12.94%), and *Magnetospirillum*

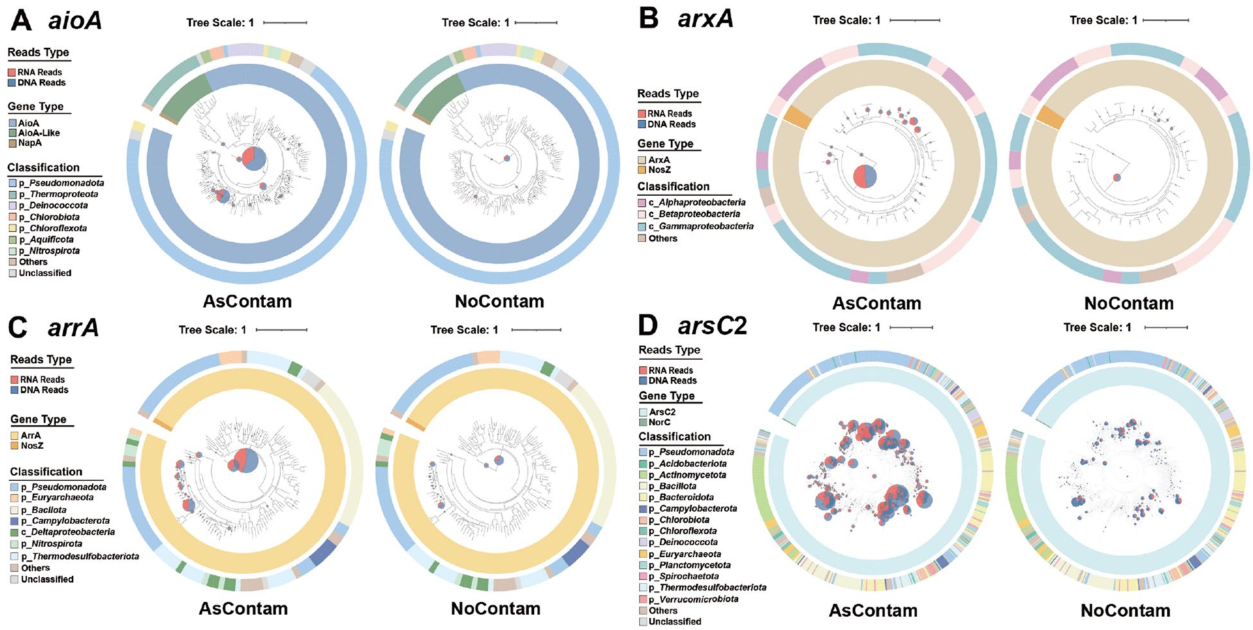

Fig. 3 Phylogenetic placement of metagenomic and metatranscriptomic reads showing the community-wide response of As(III)-oxidizing and As(V)-reducing bacteria and archaea to As contamination. Phylogenetic placement of As oxidation and reduction genes, i.e., **A** *aioA*, **B** *arxA*, **C** *arrA*, and **D** *arsC2*, which carry reads identified in metagenomes (DNA reads) and metatranscriptomes (RNA reads) from As-contaminated and As-noncontaminated paddy soils. Here, the metagenomes and metatranscriptomes from the same As level (i.e., 24 As-contaminated samples and 12 As-noncontaminated samples) were combined. The pie size indicates the number of DNA (in blue) and RNA (in red) reads, and the ratio indicates the transcriptional activity for each clade (i.e., the fraction of RNA versus DNA reads assigned to the taxa with the same thresholds for a match). The clades are colored on the outside on the basis of the taxon type, e.g., *Pseudomonadota* (blue) and *Chloroflexota* (yellow)

(12.29% vs. 10.23%) in both As-contaminated and As-noncontaminated paddy soils (Fig. S10B).

The respiratory As(V) reduction (*arrA*) sub-communities were dominated by *Pseudomonadota* (44.84%), *Thermodesulfobacteriota* (14.99%), *Deltaproteobacteria* (12.14%), and *Euryarchaeota* (8.31%) (Fig. 3C, Table S6). Specifically, *arrA* sub-communities were dominated by *Propionivibrio* (12.23% vs. 16.14%) and *Candidatus Methanoperedens* (5.55% vs. 4.72%) in both As-contaminated and As-noncontaminated paddy soils at the genus level (Fig. S10C). For bacteria and archaea involved in the detoxification of As(V), the *arsC1* sub-communities were dominated by *Pseudomonadota* (67.04%) and *Actinomycetota* (22.21%; Fig. S9, Table S6), whereas the *arsC2* sub-communities had a wider microbial distribution, including *Pseudomonadota* (30.52%), *Actinomycetota* (9.55%), *Thermodesulfobacteriota* (8.67%), *Bacteroidota* (8.10%), and *Planctomycetota* (6.93%) at the phylum level (Fig. 3D, Table S6). At the genus level, the *arsC1* sub-communities were dominated by *Defluviicoccus* (13.55% vs. 8.67%), *Conexibacter* (6.78% vs. 7.22%), *Corynebacterium* (6.57% vs. 4.07%), and *Methylocystis* (4.25% vs. 4.72%) in both As-contaminated and As-noncontaminated paddy soils (Fig. S10D). The *arsC2* sub-communities were predominantly associated with *Gallionella* (2.12%), *Anaerolinea* (1.71%), and *Isosphaera* (1.51%) in As-contaminated soils, and with *Spiribacter* (1.99%), *Anaerolinea* (1.67%), and *Isosphaera* (1.58%) in As-noncontaminated soils (Fig. S10E).

Moreover, a greater ratio of RNA/DNA reads was detected in As-contaminated paddy soils than in As-noncontaminated paddy soils, suggesting that As oxidation and reduction genes are more highly expressed under As stress. Among the various taxa affiliated with *aioA* sub-communities, *Aquificota* (1.66) and *Deinococcota* (1.51) presented greater transcriptional activity of *aioA* genes than did *Pseudomonadota* (0.62; Fig. S11A). *Betaproteobacteria* (1.19) presented the highest transcriptional activity of the *arxA* genes, followed by *Alphaproteobacteria* (0.96) and *Gammaproteobacteria* (0.82), but the observed differences in transcriptional activity across the soil pH gradients were minor (Fig. S11B). For *arrA* sub-communities, greater transcriptional activity of *arrA* genes was observed in *Deltaproteobacteria* (1.51) and *Pseudomonadota* (1.36) than in *Thermodesulfobacteriota* (0.89) and *Euryarchaeota* (0.74; Fig. S11C). In addition, the highest transcriptional activity of the *arsC1* and *arsC2* genes was detected in *Mucoromycota* (2.13) and *Bacillota* (2.75), respectively (Fig. S11D, E).

Environmental factors affecting the abundance and transcriptional activity of As(III)-oxidizing and As(V)-reducing bacteria and archaea

The concentrations of NH_4^+-N and heavy metal(loid)s, including As, Sb, and Pb, were significantly ($p < 0.05$, 0.01, or 0.001) positively correlated with the relative abundances of the *aioA*, *arsC2*, and/or *arsC1* genes, which are responsible for As detoxification (Fig. 4A, Fig. S12A). WH_2O (%) was significantly ($p < 0.01$, 0.001, or 0.05) positively correlated with the transcriptional activity of *aioA*, *arrA*, and *arsC2* genes, indicating greater expression levels of As oxidation and reduction genes under flooded conditions in paddy soils (Fig. 4A). The soil pH was significantly ($p < 0.001$) negatively correlated with the relative abundances of the *aioA*, *arsC2*, and *arsC1* genes but positively correlated with the relative abundances of the *arxA* and *arrA* genes (Fig. 4A, Figs. S12A, S13). Nevertheless, for the transcriptional activity of As oxidation and reduction genes, soil pH was significantly ($p < 0.001$ or 0.05) positively correlated with the *arxA* and *arsC2* genes (Fig. 4A, Fig. S13B, E). Specifically, in paddy soil with pH < 6.5 (acid soils), the relative abundances of the *aioA*, *arsC1*, and *arsC2* genes were significantly ($p < 0.01$ or 0.001) greater than those in paddy soil with pH ≥ 6.5 (neutral/alkaline soils; Fig. S14A, D, E), whereas the relative abundance of the *arrA* genes was significantly ($p < 0.001$) lower than that in neutral/alkaline paddy soils (Fig. S14C). For the *arxA* genes, both the relative gene abundance and transcriptional activity were significantly ($p < 0.001$) greater in paddy soil with pH ≥ 6.5 than in soil with pH < 6.5 (Fig. S14B). Moreover, the bacterial community compositions of the *aioA*, *arxA*, *arrA*, *arsC1*, and *arsC2* sub-communities also significantly differed between the two pH groups, as revealed by the NMDS plot of the Bray–Curtis metric based on the relative

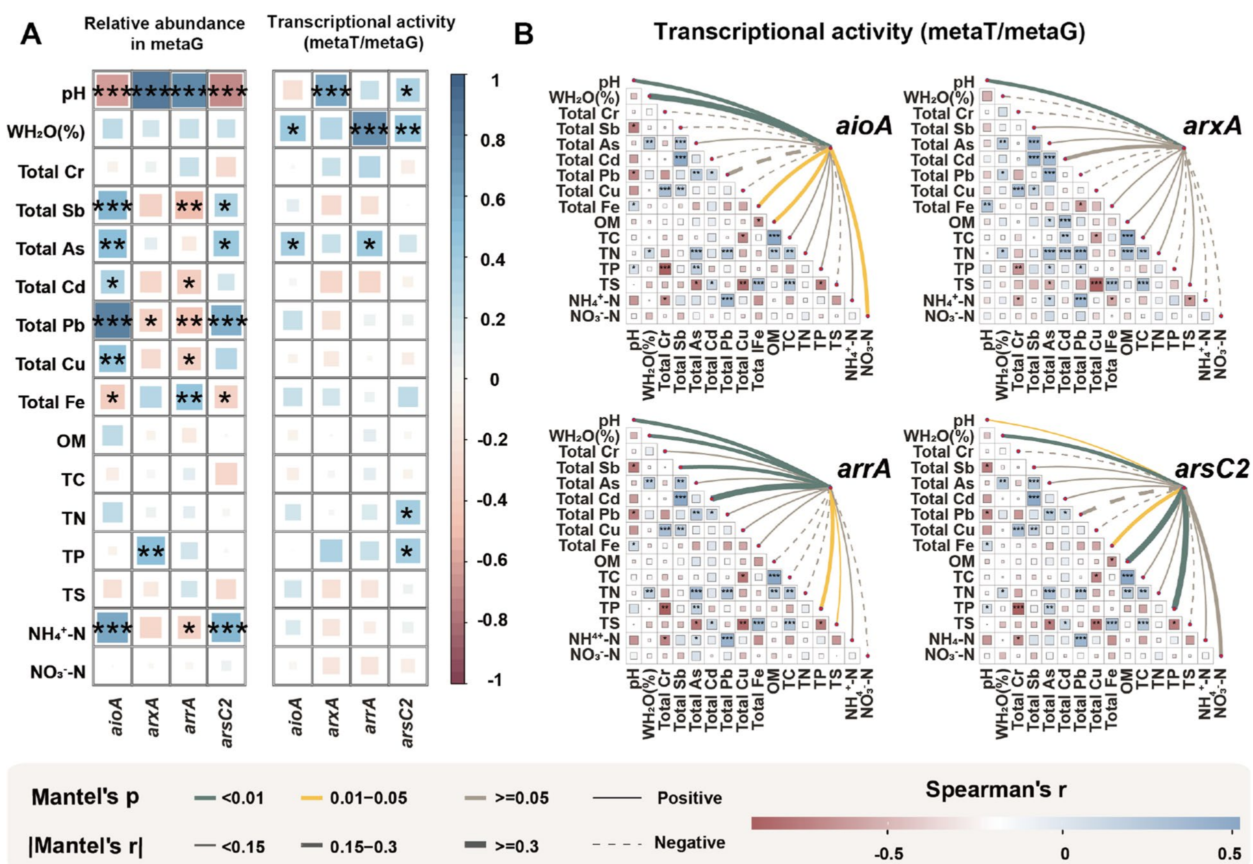

Fig. 4 Environmental factors affecting the abundance/activity of As oxidation and reduction genes and their community compositions. **A** Spearman correlation of the concentrations of environmental factors and the abundance/activity of the *aioA*, *arxA*, *arrA*, and *arsC2* genes. **B** Potential drivers of the As gene community structure (Bray–Curtis dissimilarity) analyzed by the partial Mantel test. Edge width and color correspond to Mantel's absolute *r* value and Mantel's *p* value (green, $p < 0.01$; yellow, $0.01 \le p < 0.05$; gray, $p \ge 0.05$), respectively. The solid and dashed lines indicate positive and negative correlations, respectively. Pairwise Spearman's correlations of these variables are shown with a color gradient. Asterisks (*) represent *p* values (*, $p < 0.05$; **, $0.001 < p < 0.01$; ***, $p < 0.001$)

abundance of these sub-communities (Fig. S15). Consistently, the Mantel test revealed that pH significantly ($p < 0.01$ or 0.05) contributed to the variance in the transcriptionally active *aioA*, *arxA*, *arrA*, *arsC1*, and *arsC2* sub-community compositions (Fig. 4B). Other environmental factors, including WH_2O (%) and the concentrations of OM, TP, NO_3^--N, Fe, and Sb also significantly ($p < 0.01$ or 0.05) contributed to the variance in their community compositions (Fig. 4B, Fig. S15).

Coupling of As oxidation and reduction reactions with nitrogen and carbon metabolism

To evaluate the potential coupling of As oxidation and reduction reactions and N and C metabolism at the community level, the correlations of the relative abundance of read-based annotated As oxidation and reduction genes with denitrification genes (*napA*, *narG*, *nirK*, *nirS*, *norB*, *nosZ*), dissimilatory nitrate reduction to NH_4^+ (DNRA) genes (*nrfA*), anaerobic methane (CH_4) production and oxidation genes (*mcrA*), and aerobic CH_4 oxidation genes (*pmoA*) were investigated. Potential coupling of As(III) oxidation with denitrification and DNRA was revealed, as confirmed by the significant ($p < 0.001$, 0.01, or 0.05) positive correlations between the relative abundances of the *aioA* or *arxA* genes and denitrification genes (*napA*/*narG*, *nirK*/*nirS*, and *norB*) and with the *nrfA* gene (Fig. 5A). Nevertheless, at the RNA level, although the transcriptional activity of the *aioA* gene was significantly ($p < 0.05$) positively correlated with that of the *narG* (NO_3^- to NO_2^-) and *nirK* (NO_2^- to NO) genes, the transcriptional activity of the *arxA* gene was significantly ($p < 0.001$ or 0.01 or 0.05) positively correlated with that of the *narG*/*napA* (NO_3^- to NO_2^-) and *nirS* (NO_2^- to NO) genes (Fig. 5A). For the coupling process of As(III) oxidation with DNRA, both *aioA* and *arxA* were significantly ($p < 0.05$ and $p < 0.001$, respectively) positively correlated with the *nrfA* gene (NO_2^- to NH_4^+; Fig. 5A).

Respiratory As(V) reduction mediated by *arrA* genes was revealed to be coupled with CH_4 oxidation, as *arrA* was shown to be significantly ($p < 0.001$ or 0.01) positively correlated with the *pmoA* and *mcrA* genes at both the DNA (genetic abundance) and RNA (transcriptional activity) levels (Fig. 5B). For the detoxification As(V) reduction genes, although the abundances of the *arsC1* and *arsC2* genes were not significantly positively correlated with the abundances of the *pmoA* and *mcrA* genes at the DNA level, the transcriptional activity of the *arsC2* gene was significantly ($p < 0.01$) positively correlated with

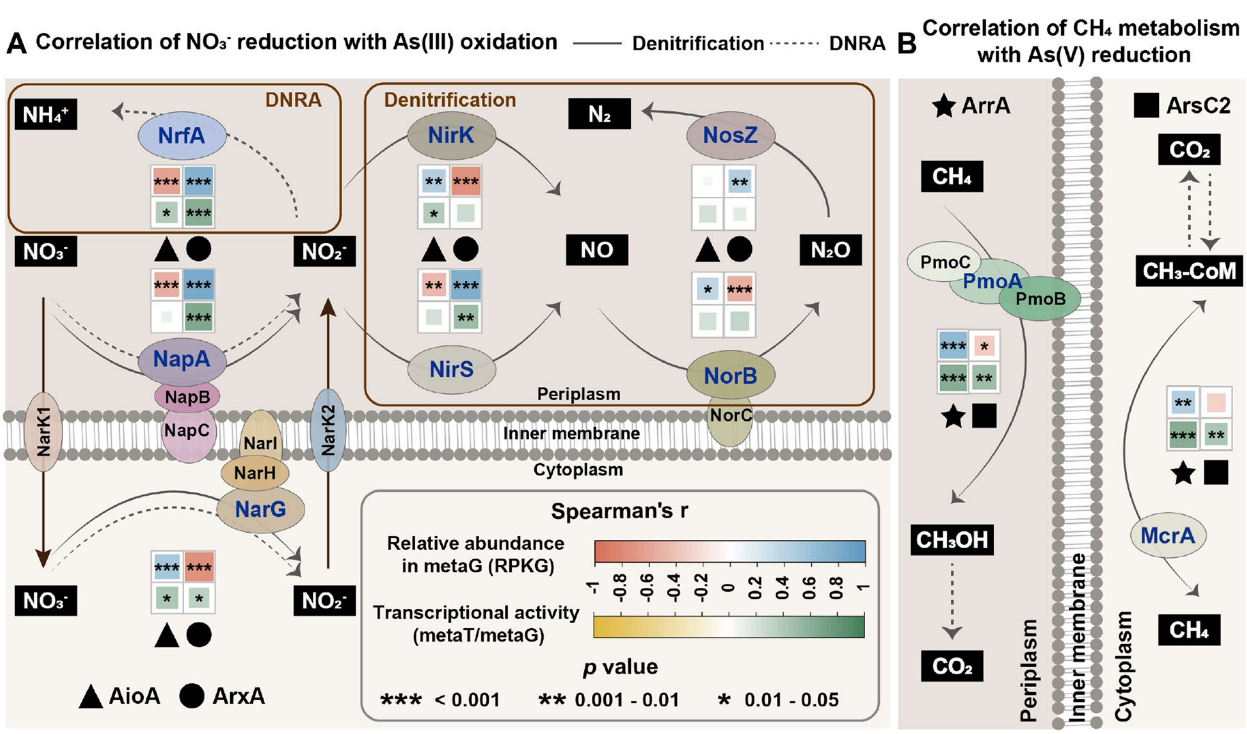

Fig. 5 Potential coupling metabolism of the N and C with As oxidation and reduction. **A** Spearman correlation of the abundance/activity of the *aioA*/*arxA* genes with those of the *narG*/*napA*/*nirK*/*nirS*/*norB*/*nosZ*/*nrfA* genes. **B** Spearman correlation of the abundance/activity of *arrA*/*arsC2* genes with *pmoA*/*mcrA* genes. Pairwise Spearman's correlations of these variables are shown with a color gradient, i.e., red-blue for abundance at the DNA level and yellow-green for transcriptional activity at the RNA level. Asterisks (*) represent *p* values (*, $p < 0.05$; **, $0.001 < p < 0.01$; ***, $p < 0.001$)

the transcriptional activity of the *pmoA* and *mcrA* genes at the RNA level (Fig. 5B and Fig. S16B). According to the phylogenetic tree of McrA (ANME and methanotrophic), 14 ANME-McrA and 18 methanogenic McrA-encoding sequences were classified from the metagenomic assembled contigs (Fig. S17). Metatranscriptomic analysis revealed that the relative abundance of the transcribed ANME-*mcrA* genes (average of 8.68 RPKM) was greater than that of methanogenic *mcrA* genes (average of 7.97 RPKM; Wilcoxon rank-sum tests, $p = 0.055$). Moreover, both the relative abundances of the transcribed *arrA* and *arsC2* genes were significantly ($p < 0.001$ and $p < 0.01$, respectively) positively correlated with those of the ANME-*mcrA* genes (Fig. S18A), whereas no significant correlation was detected between those of the methanogenic *mcrA* and As(V) reduction genes (Fig. S18B).

Identification of metagenome-assembled genomes (MAGs) capable of coupling As oxidation and reduction with nitrogen and carbon metabolism

A total of 179 MAGs (170 bacteria and 9 archaea) with completeness > 50% and contamination < 10% were binned from the metagenomic datasets and were dominated by *Pseudomonadota* (22.16% of the total identified MAGs), *Acidobacteriota* (17.37%), *Desulfobacterota* (13.17%) and *Chloroflexota* (12.57%; Table S7). Approximately 93.30% (167/179) of the MAGs presented at least one functional gene responsible for As(III) oxidation, As(V) reduction, denitrification, or DNRA (Table S8). Only 20 MAGs containing the As(III) oxidation gene (either the *aioA* or *arxA* gene), as well as at least one denitrification or DNRA gene, were identified (Fig. 6A, Table S9). These MAGs were classified into the phyla *Pseudomonadota* (6 *Burkholderiales* and 1 *UBA6522*), *Desulfobacterota* (2 *Desulfatiglandales*, 2 *UBA9968* and 1 *SM23-61*), *Acidobacteriota* (1 *Acidobacteriales*, 1 *UBA5066*, and 1 *RBC074*), *Chloroflexota* (1 *UBA4142* and 1 *Ktedonobacterales*), *Nitrospirota* (1 *Thermodesulfovibrionales*), *Methylomirabilota* (1 *Rokubacteriales*), and *Gemmatimonadota* (1 *Gemmatimonadales*; Fig. 6A, Table S9).

Genes involved in As(III) oxidation and denitrification or DNRA were co-transcribed in 17 out of the 20 MAGs, indicating the active coupling metabolism of As(III) oxidation and denitrification or DNRA. Furthermore, eight MAGs presented increased transcriptional activity of both the As(III) oxidation and denitrification or DNRA genes in As-contaminated soils compared with As-noncontaminated paddy soils (Fig. 6B), demonstrating that As contamination could facilitate coupled As(III) oxidation and denitrification or DNRA in paddy soil microbiomes. Among the eight MAGs, the marker gene of DNRA (*nrfA*) was only co-transcribed with the *arxA* gene in two MAGs that belong to *Desulfobacterota*. However, denitrification genes co-transcribed with either the *arxA* or *aioA* genes or both were detected. Specifically, co-transcription of the *arxA* and denitrification genes was detected in three MAGs assigned to *Pseudomonadota* and *Desulfobacterota*, whereas co-transcription of the *aioA* and denitrification genes was detected in three MAGs distributed in *Gemmatimonadota*, *Desulfobacterota*, and *Chloroflexota*. Moreover, although ZJJ_SH_bin10 exhibited increased transcriptional activity of both *arxA* and *nrfA* in As-contaminated soils compared with As-noncontaminated paddy soils, denitrification genes presented the opposite trend. There was also one MAG carrying both the *aioA* and *arxA* genes, which were co-transcribed with denitrification genes (Fig. 6B).

In addition, two MAGs containing both the As(V) reduction genes (*arsC1*/*arsC2* gene) and the aerobic CH_4 oxidation gene (*pmoA*/*pmoB* gene) were also identified in *Pseudomonadota* (CS_YCP_bin3 and HY_SKS_bin21; Table S8). Moreover, for HY_SKS_bin21, *arsC1*/*arsC2*, and *pmoB* co-transcribed with increased expression levels in As-contaminated paddy soils compared with in soils without contamination (Fig. 6C, Table S10). However, the coexistence of *arrA* and C metabolism genes was not detected in the MAGs identified herein, possibly due to the high complexity of the datasets.

Discussion

While similarity searches are commonly used to annotate short-read sequences carrying genes in metagenomic datasets, it is still challenging to determine the optimal cut-off threshold for searches, especially considering the variances in the similarities of As redox genes and their variants. Moreover, owing to the continuous release of genomes and metagenomes carrying novel As gene variants, such as ArxA (a newly found As oxidation enzyme) [9], these curated databases are not updated in a timely fashion. Although polymerase chain reaction (PCR)-based analyses have been widely used to investigate As redox genes in paddy soils [57], they rely on DNA/cDNA template quality, primer immaturity, and reaction conditions [58], and limit the detection of various As genes in different environments. The five customized ROCker models reported here in this study offer a new approach to overcome these limitations by identifying position-specific, most discriminant bit score thresholds in sliding windows along the sequence of the target protein sequence via the receiver operating characteristic (ROC) curve [30, 31] to detect and quantify As redox genes. Specifically, more comprehensive detection and typing of short-read sequences carrying the *aioA*, *arxA*, *arrA*, *arsC1*, and *arsC2* genes was achieved by ROCker, as indicated by the wider range of clades detected by ROCker

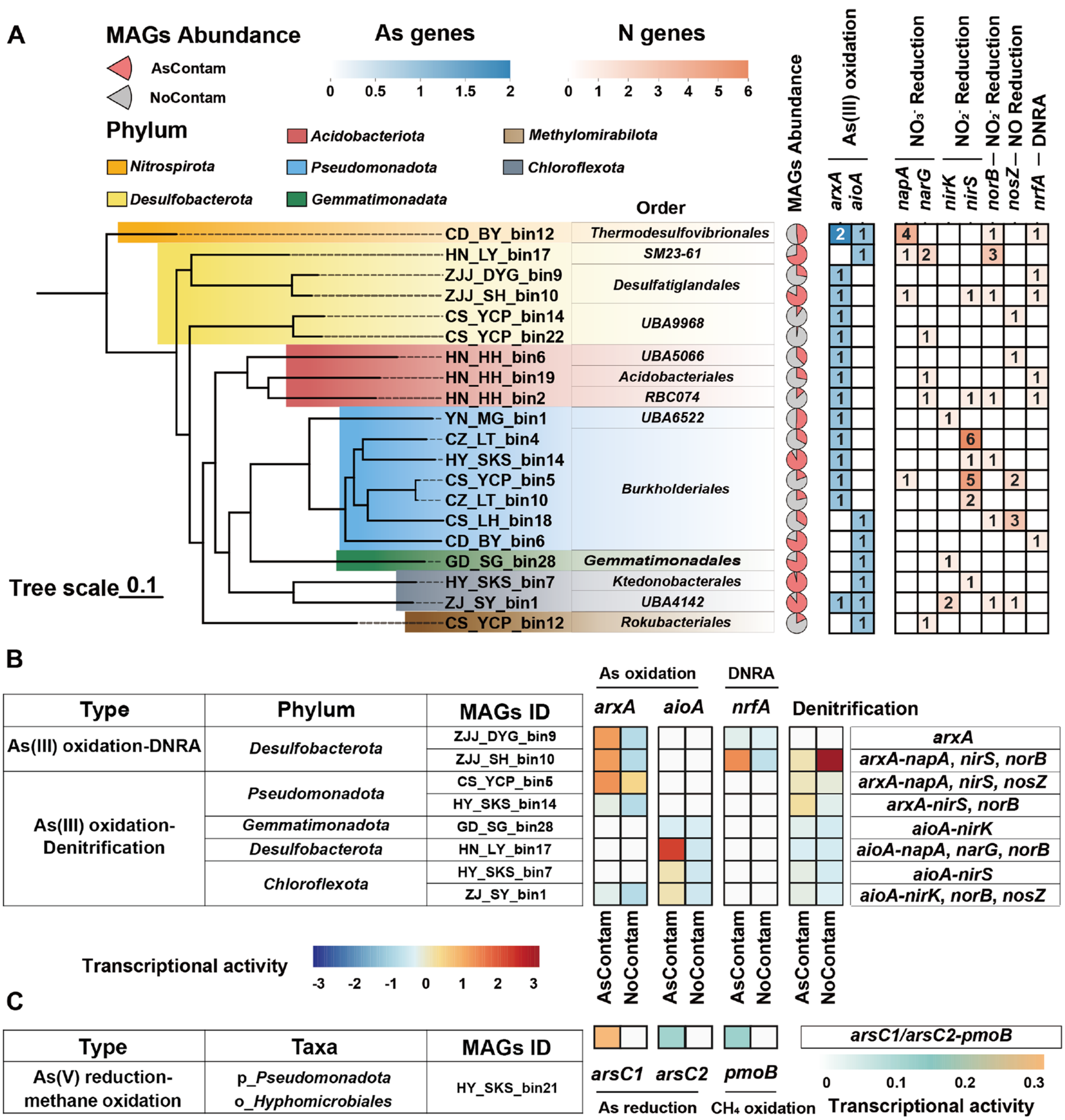

Fig. 6 Metagenome-assembled genomes (MAGs) and annotation of related As and N genes and their transcriptional activity. **A** Phylogenetic distribution and numbers of functional genes involved in As(III) oxidation, denitrification, and DNRA. The clades are colored on the basis of the phylum classification, and the identified functional genes are numbered with a color gradient. The pie charts indicate the ratio of the abundance of the corresponding MAG in the As-contaminated (red) and As-noncontaminated (gray) paddy soils. **B** Heatmap showing the transcriptional activity of functional genes involved in As(III) oxidation, denitrification, and DNRA in As-contaminated and As-noncontaminated paddy sites. The expression values are scaled within the individual functional genes and colored with a gradient. **C** Heatmap showing the transcriptional activity of functional genes involved in As(V) reduction and CH_4 oxidation in As-contaminated and As-noncontaminated paddy sites. The expression values are shown with a color gradient

than by a fixed threshold of id90 (Fig. S3). Various As redox sub-communities were revealed to have greater relative abundances and transcriptional activities in paddy soils with As contamination than in those without As contamination.

As(III) oxidation in paddy soils is mediated by both the *arxA* and *aioA* sub-communities, as revealed by the widespread presence of the *aioA* and *arxA* genes at the DNA and RNA levels in paddy soils either with or without As contamination (Fig. 2A, Fig. S6). Compared with those of the *arxA* gene, the relative abundance and transcriptional activity of the *aioA* gene increased 8-fold and 3.95-fold, respectively, suggesting the dominant role of the *aioA* sub-communities in As(III) oxidation. Nevertheless, the contribution of the *arxA* sub-communities to As(III) oxidation should not be neglected because of their relatively greater transcriptional activities than their relative abundances. In contrast to the *aioA* genes, which have been reported in various paddy soils and are diverse in terms of being carried by multiple lineages of both bacteria and archaea, including *Pseudomonadota*, *Chloroflexota*, *Thermoproteota* (Fig. 3A), *Bacteroidota* and *Actinomycetota* [8, 36], the *arxA* sub-communities have only recently been reported. With the first discovery of *arxA*-carrying bacteria in hypersaline soda lakes characterized by high temperature, As, pH and salinity, various *arxA*-carrying bacteria belonging to *Gammaproteobacteria* were isolated from alkaline saline lakes, including *Alkalilimnicola ehrlichii* MLHE-1 [9], three photosynthetic bacteria *Ectothiorhodospira* strains [11], *Halomonas* sp. strain ANAO-440 [59], *Thioalkalivibrio jannaschii* ALM2^T, and *Thioalkalivibrio thiocyanoxidans* ARh2^T [10]. In other environments, on the basis of a qPCR approach, the presence of the *arxA* genes was previously reported to be associated with *Betaproteobacteria* [12, 60, 61] and was confirmed herein via read-based annotation of the *arxA* genes via the ROCker model (Fig. 3B). In addition, we further identified *arxA*-harboring MAGs associated with the phyla *Nitrospirota*, *Chloroflexota*, *Desulfobacterota*, *Acidobacteriota* and *Actinobacteriota* (Fig. 6, Table S9), indicating that *arxA* genes might have more diverse taxonomic lineages. Moreover, we demonstrated that these *arxA* sub-communities were transcriptionally active in paddy soils (Fig. 3B), suggesting that they also contribute to As(III) oxidation together with the *aioA* sub-communities. The transcriptional activity of the *arxA* genes in paddy soils was affected by soil pH rather than by As content, as revealed by the significant positive correlation between the transcriptional activity of the *arxA* genes and soil pH (Fig. 4A, Fig. S13). Specifically, in neutral/alkaline soils with pH $\geq$ 6.5 ($\sim$ 64% of the total soil samples acquired pH > 7.5), both the relative abundance and transcriptional activities of the *arxA* genes were significantly greater than those in acidic soils (pH < 6.5), corroborating the wide presence of *arxA*-carrying bacteria in soda lakes characterized by high pH [14]. Additionally, our results suggested that the *arxA* sub-communities could also play important roles in As (III) oxidation, especially in alkaline paddy soils.

Coupling of *aioA*-mediated respiratory As(III) oxidation with NO_3^- reduction has been reported in various strains isolated from paddy soils [2, 12, 25, 26]. In addition to the *aioA* genes, the relative abundance of the *arxA* genes was also significantly positively correlated with the relative abundance of the genes responsible for denitrification and DNRA at both the DNA and RNA levels (Fig. 5A). Consistently, the concentrations of NH_4^+-N, NO_3^--N, and TN, which possibly resulted from the substantial application of NO_3^- or NH_4^+ fertilizers in paddy fields during cultivation [24], significantly affected the gene abundance and activity of As(III) oxidation communities (Fig. 4), suggesting that the addition of NO_3^- or NH_4^+ could promote the microbial conversion of As(III) to As(V) and NO_3^- reduction [2, 3, 5, 24]. Considering that a wide range of *aioA* and *arxA* clades were detected with transcriptional activity (Fig. 3A, B) and that MAGs assigned to different phyla, such as *Pseudomonadota*, *Desulfobacterota*, and *Chloroflexota*, carry the *aioA*, *arxA*, and denitrification/DNRA genes, the coupling of As(III) oxidation with denitrification or DNRA is more likely to occur at the whole-community level than at the individual taxon level. For the genes responsible for the first step of denitrification (NO_3^- to NO_2^-), although the coexistence of the *aioA*/*arxA* genes with either the *napA* or *narG* genes was detected in the MAGs (Fig. 6A), a significant positive correlation of gene transcriptional activity was observed only for *aioA* with *narG* and for *arxA* with *napA* or *narG* genes (Fig. 5A). For the genes responsible for the second step of denitrification (NO_2^- to NO), although the transcriptional activity of the *aioA* gene tended to correlate with that of the *nirK* genes, the *arxA* gene was significantly positively correlated with only the *nirS* gene (Fig. 5A). The coupling of As(III) oxidation and denitrification can be accomplished by a single microorganism; for example, a bacterial strain belonging to the genera *Acidovorax* and *Paracoccus* can use NO_3^- as an electron acceptor to oxidize As(III) to As(V) [2, 25, 26]. Additionally, it can be completed by the collaboration of NO_3^--reducing and As(III)-oxidizing bacteria, such as *Pseudogulbenkiania* and *Azoarcus*, respectively [21, 62]. Similarly, the classes *Alphaproteobacteria* and *Betaproteobacteria*, to which these genera belong, were also observed as the dominant *aioA*/*arxA* sub-communities in these paddy soils (Fig. 3, Table S6).

Moreover, As contamination in paddy soils could facilitate this coupling process, as four MAGs (GD_SG_bin

28, HN_LY_bin17, HY_SKS_bin7, and ZJ_SY_bin1) carrying denitrification and *aioA* genes presented greater transcriptional activity in As-contaminated paddy soils than in As-noncontaminated paddy soils (Fig. 6B). MAGs carrying *arxA* genes and denitrifying genes were identified in *Pseudomonadota* (Fig. 6A), and the *arxA* and denitrification (*napA*, *nirS*, *norB*/*nosZ*) genes in these two MAGs were found to have simultaneous greater transcriptional activity in As-contaminated paddy soils than in As-noncontaminated paddy soils (Fig. 6B). Under anaerobic NO_3^--reducing conditions, *Alkalilimnicola ehrlichii* MLHE-1 [63], *Azoarcus* sp. DAO1 [64], and *Noviherbaspirillum* sp. HC18 [12], which contains *arxA* genes, as well as *Paracoccus* sp. SY [26], *Acidovorax* sp. ST3 [2], *Ensifer* sp. ST2 and *Paracoccus* sp. QY30 [25], which contain *aioA* genes, could use NO_3^- as a terminal electron acceptor when oxidizing As(III) to As(V), corroborating our results here showing upregulated transcriptional activity for both *aioA*/*arxA* and denitrification genes. To date, the possible coupling of DNRA and As(III) oxidation mediated by *arxA* genes has been reported to be synergistic between *Citrobacter* and *Desulfobulbus* isolates [65]. Nevertheless, the coexistence of the *arxA* gene and the DNRA marker gene *nrfA* was observed in the present study, and these genes were identified mainly in *Desulfobacterota*, *Acidobacteriota*, and *Nitrospirota* (Fig. 6A). Both the *arxA* and *nrfA* genes were upregulated in *Desulfatiglandales* under As stress (Fig. 6B, Table S10), suggesting a new pathway involving the coupling of DNRA with As(III) oxidation mediated by the *arxA* gene.

With respect to As(V) reduction, the *arrA*, *arsC1*, and *arsC2* genes were pervasively expressed in paddy soils (Fig. 2B, Figs. S6, S7). Among them, the *arsC2* sub-communities were dominant in As(V) reduction in paddy soils, especially considering the greater abundance of the *arsC2* genes at both the DNA and RNA levels, which was 9.88-fold (DNA), 8.45-fold (RNA) and 4.07-fold (DNA), and 2.84-fold (RNA) greater than that of the *arrA* and *arsC1* genes, respectively. Moreover, more MAGs were annotated to the *arsC2* genes than to the *arrA* and *arsC1* genes (Table S8). Consistently, previous studies have shown that the abundance of *arrA* genes is much lower than that of *arsC* genes in paddy fields [20, 66]. Soil pH significantly ($p < 0.05$) contributed to the variance in the *arsC1* and *arsC2* microbial community compositions (Fig. 4B, Fig. S12) as well as the total abundance of *arrA* genes, which has also been confirmed in previous studies [67, 68]. Significant (Adonis $p < 0.01$) variances in As(V)-reducing bacterial community (*arrA*, *arsC1*, and *arsC2*) compositions were revealed between the acidic (pH < 6.5) and neutral/alkaline soils (pH ≥ 6.5; Fig. S15C, D, E). Additionally, the effects of other heavy metals, including Pb, on As oxidation and/or reduction genes should not be neglected. However, their contents in the soil are significantly correlated with only the relative abundance of As oxidation and/or reduction genes, possibly because various metal(loid)s, including Pb and As, have similar and closely related contamination sources of origin [69, 70]. Moreover, the presence of Pb in soils has been reported to increase the relative abundance of As-metabolizing microorganisms [71]. Nevertheless, the total concentrations of metal(loid)s are not always proportional to their bioavailability, which is influenced by a range of environmental factors, such as pH, redox potential, electrical conductivity, particle size, organic matter content, and other potential variables [72, 73]. Thus, quantifying the concentrations of inorganic As species, i.e., As(III) and As(V), is crucial in future studies to assess the impact of bioavailable As forms on the abundance and activity of As oxidation and reduction genes more thoroughly.

The coupling of respiratory As(V) reduction with aerobic and anaerobic CH_4 oxidation has been previously reported [4, 6]. Specifically, under anaerobic conditions in flooded paddy fields, the coupling of As(V) reduction with anaerobic CH_4 oxidation is possibly mediated by the collaboration of *arrA* sub-communities and ANMEs [4], which was corroborated by the significant ($p < 0.01$) positive correlation between the relative abundances of the transcribed *arrA* and ANME-*mcrA* genes (Fig. S18A). In addition, the coupling of CH_4 oxidation and As(V) reduction can also be independently completed by *Methanoperedenaceae* (formerly known as ANME-2d) [4]. Under aerobic conditions, the coupling of As(V) reduction and CH_4 oxidation is accomplished by the collaboration of *arrA* sub-communities (e.g., *Burkholderiaceae*) and methanotrophs [6] and was also demonstrated by the significant ($p < 0.001$) positive correlation identified in the transcriptional activity of the *arrA* and *pmoA* genes (Fig. 5B). Soil acidification strongly inhibited the coupling of CH_4 oxidation with respiratory As(V) reduction, but this inhibitory effect only occurred in the short term and diminished over the long term [27]. The significant ($p < 0.01$) positive correlation between the transcriptional activity of the *arsC2* and ANME-*mcrA*/*pmoA* genes also suggested the potential role of *arsC2* sub-communities in the coupling of As(V) reduction with both anaerobic and aerobic CH_4 oxidation (Fig. 5B, Fig. S18). In addition, the co-occurrence of *arsC1* and *arsC2* with *pmoB* in the MAG (HN_SKS_bin21, *Hyphomicrobiales*, formerly known as *Rhizobiales*), whose expression was upregulated in As-contaminated paddy soils compared with that in noncontaminated paddy soils (Fig. 6C), further suggested a new pathway for coupling As(V) reduction mediated by the *arsC* gene with methane oxidation. Nevertheless, further cultivation experiments should be carried out to confirm whether the *arsC* sub-communities also play important roles in the coupling process of As(V) reduction with methane oxidation.

Conclusions

In this study, we constructed five ROCker models to quantify the abundance and transcriptional activity of short-read sequences encoding As oxidation (*aioA* and *arxA*) and reduction (*arrA*, *arsC1*, *arsC2*) genes in paddy soils. Our results showed that the *aioA* and *arsC2* sub-communities were predominantly responsible for As oxidation and reduction, respectively. Notably, in paddy soils with relatively high pH values (pH ≥ 6.5), the *arxA* sub-communities also play a significant role in As oxidation. Various bacteria associated with *Burkholderiales* and *Desulfatiglandales* carry the *aioA*, *arxA*, and denitrification/DNRA genes and are the predominant bacteria responsible for the coupling of As(III) oxidation with denitrification or DNRA. In addition to the previously reported CH_4 oxidation with As(V) reduction mediated by *arrA* sub-communities, we further revealed the considerable contribution of *arsC2* sub-communities (*Hyphomicrobiales,* formerly known as *Rhizobiales*) to the oxidation of CH_4. Moreover, As contamination could enhance the coupling of As oxidation and reduction with N and C metabolism, as indicated by the increased transcriptional activity of the relevant genes in MAGs in As-contaminated soils. Overall, these results expand our current knowledge of the various microbial taxa dominant in As(III) oxidation and As(V) reduction and reveal their potential pathways for coupling As with N and C in paddy soils.

Abbreviations

As	Arsenic
As(III)	Arsenite
As(V)	Arsenate
NO_3^-	Nitrate
NH_4^+	Ammonium
DNRA	Dissimilatory nitrate reduction to ammonium
AsContam	As-contaminated
NoContam	As-noncontaminated
WH_2O (%)	Soil moisture content, i.e., WH_2O (%) = the mass of moisture present in the soil sample/the total mass of the soil sample
TC	Total concentration of carbon
TN	Total concentration of nitrogen
TP	Total concentration of phosphorus
TS	Total concentration of sulfur
Total Fe	Total concentration of iron
Total As	Total concentration of arsenic
Total Pb	Total concentration of lead
Total Cd	Total concentration of cadmium
Total Cr	Total concentration of chromium
Total Sb	Total concentration of antimony
Total Cu	Total concentration of copper
OM	Organic matter
metaG	Metagenomic
metaT	Metatranscriptomic
id70	BLASTx search with a minimum cut-off for a match of amino acid identities of 70%, along with a minimum alignment length of 25 amino acids and a maximum *e*-value of 1e − 5
id80	BLASTx search with a minimum cut-off for a match of amino acid identities of 80%, along with a minimum alignment length of 25 amino acids and a maximum *e*-value of 1e − 5
id90	BLASTx search with a minimum cut-off for a match of amino acid identities of 90%, along with a minimum alignment length of 25 amino acids and a maximum *e*-value of 1e − 5
MAGs	Metagenome-assembled genomes
ORFs	Open reading frames
CH_4	Methane

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01952-4.

Additional file 1: Fig. S1. Phylogenetic relationships of the verified As oxidation and reduction genes with verified nontarget proteins (gray) and other related proteins from the UniRef90 database. Fig. S2 150-bp ROCker models for *arsC1* and comparison of the ROCker model versus different fixed cut-offs: id70, id80 and id90. Fig. S3 Phylogenetic placement of five As(III) oxidation and As(V) reduction gene reads identified by ROCker and id90. Fig. S4 Geographic location of the sampling sites in South China. Fig. S5 Comparison of soil properties between As-contaminated and As-noncontaminated paddy soils. Fig. S6 Comparison of As oxidation and reduction genes at the DNA and RNA levels. Fig. S7 Relative abundance and transcriptional activity of *arsC1* genes in As-contaminated versus As-noncontaminated paddy soils and their correlation with As levels. Fig. S8 Comparison of the microbial community composition in As-contaminated and As-noncontaminated paddy soils. Fig. S9 Phylogenetic placement of metagenomic and metatranscriptomic reads showing the community-wide response to As contamination. Fig. S10 Taxonomic classification of the transcriptionally active (A) *aioA*, (B) *arxA*, (C) *arrA*, (D) *arsC1*, and (E) *arsC2* sub-communities at the genus level in As-contaminated and As-noncontaminated paddy soils. Fig. S11 Gene ratios between phyla affiliated with (A) *aioA*, (B) *arxA*, (C) *arrA*, (D) *arsC1*, and (E) *arsC2*. Fig. S12 Environmental factors affecting the abundance/activity of *arsC1* genes and their community compositions. Fig. S13 Pearson correlation between soil pH and the abundance/activity of the (A) *aioA*, (B) *arxA*, (C) *arrA*, (D) *arsC1*, and (E) *arsC2* genes. Fig. S14 Boxplot of the abundance/activity of the (A) *aioA*, (B) *arxA*, (C) *arrA*, (D) *arsC1*, and (E) *arsC2* genes in acidic (pH < 6.5) versus alkaline (pH ≥ 6.5) paddy soils. Fig. S15 Nonmetric multidimensional scaling (NMDS) ordination plots showing the β diversity of the (A) *aioA*, (B) *arxA*, (C) *arrA*, (D) *arsC1* and (E) *arsC2* sub-communities at different pH values. Fig. S16 Potential coupling metabolism of N and C genes with As oxidation and reduction genes. Fig. S17 Phylogenetic tree of ANME-McrA and methanogenic McrA based on amino acid sequences. Fig. S18 Pearson correlation between the abundance of transcribed ANME-*mcrA* (A), methanogenic *mcrA* (B) and As reduction genes.

Additional file 2: Table S1. Information on the samples included in this study. Table S2. Physicochemical parameters of the paddy soil. Table S3. Metagenomic dataset statistics. Table S4. Metatranscriptomic dataset statistics. Table S5. UniProt IDs of selected positive and negative references for As oxidation and reduction gene ROCker construction. Table S6. As oxidation and reduction genes-carrying reads identified in metatranscriptomes from As-contaminated and As-noncontaminated paddy soils. Table S7. Taxa information and relative abundance of 179 MAGs with completeness > 50% and contamination < 10%. Table S8. The functional genes involved in As(III) oxidation, As(V) reduction, denitrification, DNRA and methane oxidation identified in the MAGs. Table S9. Twenty MAGs contained the As(III) oxidation gene (either the *aioA* or *arxA* gene), as well as at least one denitrification or DNRA gene. Table S10. The transcriptional activity of functional genes involved in As(III) oxidation, As(V) reduction, denitrification, DNRA and methane oxidation in paddy soils.

Acknowledgements
Not applicable.

Authors' contributions
SYZ and XDZ designed the study. XDZ performed the sampling and bioinformatic analysis and constructed the *aioA*, *arxA*, and *arrA* ROCker models. ZYG constructed the *arsC1* and *arsC2* ROCker models. XDZ, ZYG, and SYZ wrote the manuscript. JJP and KTK performed the conceptualization, manuscript reviewing, and language modification. All the authors discussed the results and contributed to the final manuscript.

Funding
This study is financially supported by the Natural Science Foundation of China (Grant No. 42107135).

Data availability
The raw sequencing data for the metagenomic datasets are deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject of PRJNA1068274. The metatranscriptomic datasets are deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject of PRJNA1068685.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China. [2]State Key Laboratory of Nutrient Use and Management, College of Resources and Environmental Sciences, Key Laboratory of Plant-Soil Interactions, National Academy of Agriculture Green Development, China Agricultural University, Beijing, China. [3]School of Civil & Environmental Engineering and School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA.

Received: 11 April 2024 Accepted: 16 October 2024
Published online: 14 November 2024

References

1. Takahashi Y, Minamikawa R, Hattori KH, Kurishima K, Kihou N, Yuita K. Arsenic behavior in paddy fields during the cycle of flooded and non-flooded periods. Environ Sci Technol. 2004;38(4):1038–44.
2. Zhang J, Zhao SC, Xu Y, Zhou WX, Huang K, Tang Z, et al. Nitrate stimulates anaerobic microbial arsenite oxidation in paddy soils. Environ Sci Technol. 2017;51(8):4377–86.
3. Lin ZJ, Wang X, Wu X, Liu DH, Yin YL, Zhang Y, et al. Nitrate reduced arsenic redox transformation and transfer in flooded paddy soil-rice system. Environ Pollut. 2018;243:1015–25.
4. Shi LD, Guo T, Lv PL, Niu ZF, Zhou YJ, Tang XJ, et al. Coupled anaerobic methane oxidation and reductive arsenic mobilization in wetland soils. Nat Geosci. 2020;13(12):799–805.
5. Zhang MM, Li Z, Häggblom MM, Young L, He ZJ, Li FB, et al. Characterization of nitrate-dependent As(III)-oxidizing communities in arsenic-contaminated soil and investigation of their metabolic potentials by the combination of DNA-stable isotope probing and metagenomics. Environ Sci Technol. 2020;54(12):7366–77.
6. Shi LD, Zhou YJ, Tang XJ, Kappler A, Chistoserdova L, Zhu LZ, et al. Coupled aerobic methane oxidation and arsenate reduction contributes to soil-arsenic mobilization in agricultural fields. Environ Sci Technol. 2022;56(16):11845–56.
7. Zhu YG, Xue XM, Kappler A, Rosen BP, Meharg AA. Linking genes to microbial biogeochemical cycling: lessons from arsenic. Environ Sci Technol. 2017;51(13):7326–39.
8. Yamamura S, Amachi S. Microbiology of inorganic arsenic: from metabolism to bioremediation. J Biosci Bioeng. 2014;118(1):1–9.
9. Zargar K, Hoeft S, Oremland R, Saltikov CW. Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J Bacteriol. 2010;192(14):3755–62.
10. Ahn AC, Cavalca L, Colombo M, Schuurmans JM, Sorokin DY, Muyzer G. Transcriptomic analysis of two *Thioalkalivibrio* species under arsenite stress revealed a potential candidate gene for an alternative arsenite oxidation pathway. Front Microbiol. 2019;10:1514.
11. Hernandez-Maldonado J, Sanchez-Sedillo B, Stoneburner B, Boren A, Miller L, McCann S, et al. The genetic basis of anoxygenic photosynthetic arsenite oxidation. Environ Microbiol. 2017;19(1):130–41.
12. Wu YF, Chai CW, Li YN, Chen J, Yuan Y, Hu G, et al. Anaerobic As(III) oxidation coupled with nitrate reduction and attenuation of dissolved arsenic by *Noviherbaspirillum* species. ACS Earth Space Chem. 2021;5(8):2115–23.
13. Zhu YG, Yoshinaga M, Zhao FJ, Rosen BP. Earth abides arsenic biotransformations. Annu Rev Earth Planet Sci. 2014;42:443–67.
14. Edwardson CF, Hollibaugh JT. Metatranscriptomic analysis of prokaryotic communities active in sulfur and arsenic cycling in Mono Lake, California, USA. ISME J. 2017;11(10):2195–208.
15. Glasser NR, Oyala PH, Osborne TH, Santini JM, Newman DK. Structural and mechanistic analysis of the arsenate respiratory reductase provides insight into environmental arsenic transformations. Proc Natl Acad Sci U S A. 2018;115(37):E8614–23.
16. Rosen BP. Biochemistry of arsenic detoxification. FEBS Lett. 2002;529(1):86–92.
17. Chen SC, Sun GX, Yan Y, Konstantinidis KT, Zhang SY, Deng Y, et al. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. Proc Natl Acad Sci U S A. 2020;117(19):10414–21.
18. Ji G, Silver S. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. Proc Natl Acad Sci U S A. 1992;89(20):9474–8.
19. Oden KL, Gladysheva TB, Rosen BP. Arsenate reduction mediated by the plasmid-encoded ArsC protein is coupled to glutathione. Mol Microbiol. 1994;12(2):301–6.
20. Zhang SY, Zhao FJ, Sun GX, Su JQ, Yang XR, Li H, et al. Diversity and abundance of arsenic biotransformation genes in paddy soils from Southern China. Environ Sci Technol. 2015;49(7):4138–46.
21. Li XM, Qiao JT, Li S, Häggblom MM, Li FB, Hu M. Bacterial communities and functional genes stimulated during anaerobic arsenite oxidation and nitrate reduction in a paddy soil. Environ Sci Technol. 2020;54(4):2172–81.
22. Feng M, Du YH, Li XM, Li FB, Qiao JT, Chen GN, et al. Insight into universality and characteristics of nitrate reduction coupled with arsenic oxidation in different paddy soils. Sci Total Environ. 2023;866: 161342.
23. Gan XL, Hu HQ, Fu QL, Zhu J. Nitrate reduction coupling with As(III) oxidation in neutral As-contaminated paddy soil preserves nitrogen, reduces NO emissions and alleviates as toxicity. Sci Total Environ. 2024;912:169360.
24. Chen GN, Du YH, Fang LP, Wang XQ, Liu CP, Yu HY, et al. Distinct arsenic uptake feature in rice reveals the importance of N fertilization strategies. Sci Total Environ. 2023;854: 158801.
25. Zhang J, Chai CW, ThomasArrigo LK, Zhao SC, Kretzschmar R, Zhao FJ. Nitrite accumulation is required for microbial anaerobic iron oxidation, but not for arsenite oxidation, in two heterotrophic denitrifiers. Environ Sci Technol. 2020;54(7):4036–45.
26. Zhang J, Zhou WX, Liu BB, He J, Shen QR, Zhao FJ. Anaerobic arsenite oxidation by an autotrophic arsenite-oxidizing bacterium from an arsenic-contaminated paddy soil. Environ Sci Technol. 2015;49(10):5956–64.
27. Yuan ZF, Zhou YJ, Zou LA, Chen Z, Gustave W, Duan DC, et al. pH dependence of arsenic speciation in paddy soils: the role of distinct methanotrophs. Environ Pollut. 2023;318:120880.
28. Zhou Y, Guo T, Gustave W, Yuan Z, Yang J, Chen D, et al. Anaerobic methane oxidation coupled to arsenate reduction in paddy soils: insights from laboratory and field studies. Chemosphere. 2023;311(Pt 2): 137055.
29. Zhang SY, Su JQ, Sun GX, Yang YF, Zhao Y, Ding JJ, et al. Land scale biogeography of arsenic biotransformation genes in estuarine wetland. Environ Microbiol. 2017;19(6):2468–82.

30. Orellana LH, Rodriguez-R LM, Konstantinidis KT. ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. Nucleic Acids Res. 2017;45(3):e14.
31. Zhang SY, Suttner B, Rodriguez-R LM, Orellana LH, Conrad RE, Liu F, et al. ROCker models for reliable detection and typing of short-read sequences carrying β-Lactamase genes. mSystems. 2022;7(3):e0128121.
32. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems. 2018;3(3):e00039-18.
33. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211–7.
34. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.
35. Nayfach S, Pollard KS. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biol. 2015;16(1):51.
36. Saunders JK, Fuchsman CA, Mckay C, Rocap G. Complete arsenic-based respiratory cycle in the marine microbial communities of pelagic oxygen-deficient zones. Proc Natl Acad Sci U S A. 2019;116(20):9925–30.
37. Suzek BE, Wang YQ, Huang HZ, McGarvey PB, Wu CH, and Consortium U. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31(6):926–32.
38. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–8.
39. Qian L, Yu XL, Zhou JY, Gu H, Ding JJ, Peng YS, et al. MCycDB: a curated database for comprehensively profiling methane cycling processes of environmental microbiomes. Mol Ecol Resour. 2022;22(5):1803–23.
40. Tu QC, Lin L, Cheng L, Deng Y, He ZL. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. Bioinformatics. 2019;35(6):1040–8.
41. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints. 2016;4:e1900v1.
42. Zhang SY, Xiao X, Chen SC, Zhu YG, Sun GX, Konstantinidis KT. High arsenic levels increase activity rather than diversity or abundance of arsenic metabolism genes in paddy soils. Appl Environ Microbiol. 2021;87(20):e0138321.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
44. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
45. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
46. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3–11.
47. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6(1):158.
48. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017;11(12):2864–8.
49. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
50. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38(23):5315–6.
51. Xie JM, Chen YR, Cai GJ, Cai RL, Hu Z, Wang H. Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51(W1):W587–92.
52. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
53. Villanueva RAM, Chen ZJ. ggplot2: Elegant graphics for data analysis, 2nd edition. Measurement. 2019;17(3):160–7.
54. Ahlmann-Eltze C, Patil I. ggsignif: R package for displaying significance brackets for 'ggplot2'. 2021. https://github.com/YY-SONG0718/ggsignif/.
55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
56. Zhang J. Beautiful data visualization of R- how to make professional charts. Beijing: Publishing House of Electronics Industry; 2019.
57. Zhao Y, Su JQ, Ye J, Rensing C, Tardif S, Zhu YG, et al. AsChip: a high-throughput qPCR chip for comprehensive profiling of genes linked to microbial cycling of arsenic. Environ Sci Technol. 2019;53(2):798–807.
58. Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol. 2009;67(1):6–20.
59. Hamamura N, Damdinsuren N, Nakajima N, Yamamura S. Draft genome sequence of the anaerobic arsenite-oxidizing Halomonas sp. strain ANAO-440, isolated from an alkaline saline lake in Khovsgol, Mongolia. Microbiol Resour Announc. 2021;10(42):e0089921.
60. Durante-Rodríguez G, Fernández-Llamoses H, Alonso-Fernandes E, Fernández-Muñiz MN, Muñoz-Olives R, Diaz E, et al. ArxA from Azoarcus sp. CIB, an anaerobic arsenite oxidase from an obligate heterotrophic and mesophilic bacterium. Front Microbiol. 2019;10:1699.
61. Ospino MC, Kojima H, Fukui M. Arsenite oxidation by a newly isolated *Betaproteobacterium* possessing *arx* genes and diversity of the *arx* gene cluster in bacterial genomes. Front Microbiol. 2019;10:1210.
62. Li S, Chen JJ, Ma J, Lu YL. Functional microbial communities involved in As(III) oxidation coupled with nitrate reduction in a paddy soil. Water Air Soil Pollut. 2023;234(9):598.
63. Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, et al. *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int J Syst Evol Microbiol. 2007;57:504–12.
64. Rhine ED, Phelps CD, Young LY. Anaerobic arsenite oxidation by novel denitrifying isolates. Environ Microbiol. 2006;8(5):899–908.
65. Wu Z, Chen Z, Wang H, Liu H, Wei Z. Arsenic removal in flue gas through anaerobic denitrification and sulfate reduction cocoupled arsenic oxidation. Chemosphere. 2023;337:139350.
66. Wang N, Xue XM, Juhasz AL, Chang ZZ, Li HB. Biochar increases arsenic release from an anaerobic paddy soil due to enhanced microbial reduction of iron and arsenic. Environ Pollut. 2017;220:514–22.
67. Zhang C, Xiao X, Zhao Y, Zhou JZ, Sun B, Liang YT. Patterns of microbial arsenic detoxification genes in low-arsenic continental paddy soils. Environ Res. 2021;201:111584.
68. Bei Q, Yang TT, Ren CY, Guan EX, Dai YC, Shu DT, et al. Soil pH determines arsenic-related functional gene and bacterial diversity in natural forests on the Taibai Mountain. Environ Res. 2023;220:115181.
69. Li SZ, Zhao B, Jin M, Hu L, Zhong H, He ZG. A comprehensive survey on the horizontal and vertical distribution of heavy metals and microorganisms in soils of a Pb/Zn smelter. J Hazard Mater. 2020;400:123255.
70. Gui H, Yang QC, Lu XY, Wang HL, Gu QB, Martín JD. Spatial distribution, contamination characteristics and ecological-health risk assessment of toxic heavy metals in soils near a smelting area. Environ Res. 2023;222:115328.
71. Li XH, Liu XX, Cao N, Fang SJ, Yu CH. Adaptation mechanisms of arsenic metabolism genes and their host microorganisms in soils with different arsenic contamination levels around abandoned gold tailings. Environ Pollut. 2021;291:117994.
72. Wang CC, Zhang QC, Kang SG, Li MY, Zhang MY, Xu WM, et al. Heavy metal(loid)s in agricultural soil from main grain production regions of China: bioaccessibility and health risks to humans. Sci Total Environ. 2023;858:159819.
73. Wieczorek J, Baran A, Bubak A. Mobility, bioaccumulation in plants, and risk assessment of metals in soils. Sci Total Environ. 2023;882:163574.

Publisher's Note