Environmental Science & Technology

pubs.acs.org/est

Article

Lignin and Peptide Promote the Abundance and Activity of Arsenic Methylation Microbes in Paddy Soils

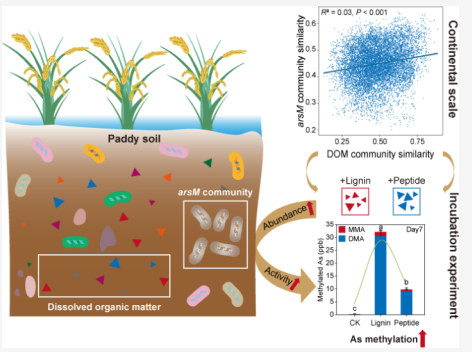
Si-Yu Zhang, Zi-Teng Liu, Xin-Di Zhao, Zi-Yu Gao, Ouyuan Jiang, Jie Li, Xiaoming Li, Andreas Kappler, Jianming Xu, and Xianjin Tang*

Cite This: *Environ. Sci. Technol.* 2025, 59, 2541−2553

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Rice physiological straighthead disease is induced by microbially mediated arsenic methylation and usually regionally distributed in paddy soils. However, the biogeochemical mechanism underlying the geographic distribution of microbial communities harboring methylating genes (*arsM*) remains unclear. Herein, we revealed significant (p = 0.001) differences in the *arsM* communities in different regions of Chinese paddy soils at the continental scale. Moreover, a positive correlation between the diversity of *arsM* communities and the chemodiversity of soil dissolved organic matter (DOM) was revealed. Among the different DOM components, lignin- and peptide-like DOM are the most important DOM components impacting the abundance and diversity of *arsM* communities. Metatranscriptomic analyses of 18 selected paddy soil samples revealed that the expression of the *arsM* gene increased with increasing soil lignin and peptide contents. Compared with the control, the addition of lignin and peptide significantly (p < 0.05) increased the methylated As concentration in the incubated paddy soils. Communities harboring *arsM* genes belonging to the phyla *Chloroflexota*, *Verrucomicrobiota*, *Deltaproteobacteria*, *Thermodesulfobacteriota*, and *Ignavibacteriota* mostly dominated in paddy soils with relatively high lignin and peptide contents. This study highlights the correlation between the diversity of DOM and *arsM* communities in paddy soils and provides mechanistic information for soil arsenic contamination control and sustainable rice production.



KEYWORDS: *paddy soils, soil dissolved organic matter, arsenic methylation, arsM communities*

INTRODUCTION

Arsenic (As) contamination in paddy fields represents a critical environmental issue that threatens sustainable agriculture and human health.[1−3] The dominant As species in paddy soils include inorganic As (iAs, including arsenite (As(III)) and arsenate (As(V))), methylated arsenic (monomethylarsenate (MMA), dimethylarsenate (DMA), and trimethylarsenate (TMAO)), among which As(III) and DMA are the main As species reported in rice grains.[4−6] The overaccumulation of DMA can cause outbreaks of rice straight-head disease, resulting in spikelet sterility and yield losses.[7−9] Since straight-head disease is prevalent in many rice-growing regions,[8] it is important to understand how As methylation occurs in paddies and what drives such occurrence.

Microbes can biomethylate inorganic As into methylated As species, which is one of the most important As detoxification processes.[10,11] Microbially mediated As methylation is catalyzed by As(III) S-adenosylmethionine methyltransferase (ArsM), an enzyme encoded by the *arsM* gene.[12] The abundant organic matter (e.g., root exudates, plant residue) of paddies as well as the added organic matter (e.g., manure, straw returned as biochar) supply labile carbon to promote the activity of indigenous microbial communities potentially involved in As(V) reduction and As release, increasing the concentration of As(III) as a methylation substrate and the abundance of *arsM*-harboring bacteria in paddy soils.[13,14] Dissolved organic matter (DOM) represents an important fraction of organic matter. DOM influences the diversity, structure, and functioning of microbial communities due to the diverse preferences of microbes for DOM consumption.[15] The distribution patterns of DOM in paddy soils are correlated with the taxonomic profile and metabolic potential of the rice paddy microbiome.[16] Studies have shown that a relatively high DOM content is responsible for the increase in *arsM*-harboring microbes.[17] Furthermore, evidence from microcosm experiments revealed that the complexity of organic carbon molecules is associated with As methylation in paddy soils.[18] However, it is not clear whether the variances in microbial communities involved in As methylation are associated with

Received: October 9, 2024
Revised: January 18, 2025
Accepted: January 21, 2025
Published: January 31, 2025

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

changes in soil DOM molecular structure at large regional scales. As such, it is necessary to expand the experimental application to paddy fields and to verify the hypothesis from local to global scales.

In this study, we used Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) combined with metagenomic sequencing to reveal the associations between As methylation microbes and DOM molecular structure in paddy soils at the continental level. The objectives of this study were to (1) elucidate the diversity of As(III) methylation microbes in Chinese paddy soils at the continental scale, (2) explore the correlation between As(III) methylation microbial communities and different DOM molecular compositions in paddy soils, and (3) determine how different DOM molecular compositions affect the abundance and diversity of As(III) methylation microbes.

MATERIALS AND METHODS

Site Description and Sample Collection. From 2022 to 2023, a total of 141 soil samples were collected from paddy fields distributed in the main rice-growing regions in China, including North China (Heilongjiang, Jilin, Liaoning and Tianjin Provinces, n = 21), East China (Shandong, Jiangsu, Anhui, Shanghai, Zhejiang and Fujian Provinces, n = 42), Central China (Henan, Hubei, Hunan and Jiangxi Provinces, n = 42) and South China (Guangdong, Guangxi, Guizhou and Yunnan Provinces, n = 36). Details of the locations of these paddy fields are provided in the Supporting Information. Three replicates from the same field at each site were collected from the soil surface (0−20 cm) after rice harvest. Specifically, for each paddy field, we randomly selected three sampling plots and utilized a five-point sampling method for soil sample collection. For each of the sampling plots, soil samples were collected at the center of the selected sampling plot (the midpoint of the diagonal of the plot) and at each of the four corners (equidistant from the center on the diagonal). These five collected soil samples were subsequently mixed at the same weight. The soil samples were placed in sterile plastic bags and preserved in the cooler boxes with ice to maintain a cold temperature, and then transported as soon as possible to the laboratory on ice (within 2 days) for subsequent chemical analyses and DNA/RNA extraction immediately. Soil properties, including pH and the total concentrations of organic carbon (TOC), nitrogen (TN), phosphorus (TP), and sulfur (TS) were determined following standard methods.[19] Heavy metal(oid)s, including arsenic (As), lead (Pb), cadmium (Cd), chromium (Cr), and copper (Cu) were measured by ICP-MS, as described previously.[20] Details of the underlying methods are provided in the Supporting Information.

Soil DNA/RNA Extraction, Metagenomic, Metatranscriptomic Sequencing and Quantitative PCR (qPCR) Analysis. In total, 141 soil DNA and RNA samples were extracted from 0.5 and 2 g of soil using an ALFA-SEQ Advanced Soil DNA Kit (Findrop, Guangzhou) and an RNeasy Mini Kit (Qiagen, Germany), respectively, according to the manufacturer's instructions. The purity and concentration of the extracted DNA were assessed using 1% agarose gels and a Qubit 3.0 (Thermo Fisher Scientific, Waltham, USA), respectively. The DNA extract was fragmented to an average size of approximately 400 bp using Covaris M220 (Gene Company Limited, China) for paired-end library construction. A paired-end library (2 × 150 bp) was constructed using NEXTFLEX Rapid DNA-Seq (Bioo Scientific, Austin, TX, USA) and sequenced on an NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China). The purity integrity of the RNA was checked via 1% agarose gel electrophoresis (5 V/cm, 15 min) to make sure it is qualified for further analyses. The quality and concentration of the extracted RNA were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). The extracted RNA was reverse transcribed to cDNA using the HiScript II first Strand cDNA Synthesis Kit and then used for quantitative PCR (qPCR) analysis of the *arsM* genes using a Bioer Quant Gene 9600 Plus (FQD-96C, Hangzhou, China). The details of the primer sequences, procedures, and systems used for qPCR amplification are provided in the Supporting Information. Among the 141 paddy soils, 1−2 sampling sites from each region, i.e., North China (SD_TC), East China (ZJ_SY, ZJ_QJ), Central China (HN_HH), and South China (GD_SG, YN_MG), for a total of six sampling sites (18 soils with triplicate samples for each site) were further selected for metatranscriptomic sequencing at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China).

Trimming, Assembly and Binning of Metagenomic And/Or Metatranscriptomic Data Sets. The raw reads were trimmed of adaptors, and low-quality reads (length <50 bp or with a quality value <20) were removed by multitrim (https://github.com/KGerhardt/multitrim) for both the metagenomic and metatranscriptomic data sets. SortMeRNA v4.3.4[21] was used to eliminate rRNA with the option "--paired_in" for metatranscriptomic data sets. Clean reads in metagenomic data sets were then used to estimate genome equivalents for each sample using MicrobeCensus v1.1.0.[22] The clean reads were assembled using MEGAHIT v1.2.9[23] (parameters: --min-contig-len 500). Clean reads/contigs from three replicate samples from one site were mixed into one sample to recover more and higher quality metagenome-assembled genomes (MAGs) via MetaWRAP v1.2.1.[24] The raw MAGs were subsequently refined using MetaWRAP's'-Bin_refinement' model. The completeness and contamination of the refined MAGs were evaluated using CheckM v1.1.3.[25] MAGs that met the threshold criteria of a minimum completeness of 90% and a maximum contamination of 5% were selected for subsequent analysis. The MAGs were subsequently dereplicated using dRep v3.3.0[26] (parameters: -sa 0.95 -nc 0.30). The open reading frames (ORFs) of the MAGs were predicted using Prodigal v2.6.3[27] (parameters: -p meta). The taxonomic classification and functional annotation of the MAGs are detailed in the Supporting Information.

Taxonomic Classification and Functional Annotation of the Metagenomic and/or Metatranscriptomic Data Sets. Taxonomic classification of the metagenomic data sets was based on the clean reads using Kraken2 v2.0.7[28] (parameter: --confidence 0.05), and Bracken v2.7[29] was used to estimate species abundance (parameter: -t 10). The clean reads were searched against the Swiss-Prot database (UniProt, downloaded in Sep. 2022) using Diamond BLASTx (v2.0.14.152).[30] Only the best matches with e-values <1 × 10^{-5}, amino acid identities >80%, and alignment lengths >25 aa were retained, and the carbon, nitrogen, phosphorus, and sulfur metabolism genes were subsequently functionally annotated against the SEED database using the large category.[31] Annotation of the *arsM* gene for the clean reads was performed via Diamond BLASTx against the in-home

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

constructed ArsM database (detailed in the Supporting Information) with a minimum e-value cutoff of 1×10^{-5}, an amino acid identity of 80%, and an alignment length of 25 aa. The taxonomic classification of the *arsM*-carrying reads was determined based on the taxonomic information on the reference sequences in the in-home constructed ArsM database. Annotation of the *arsM* gene in the ORFs of MAGs was performed via Diamond BLASTp against the in-home constructed ArsM database with a minimum e-value cutoff of 1×10^{-5}, an amino acid identity of 60%, and query length coverage of 70%. The relative abundances of *arsM* and other functional genes in the metagenomic data sets were normalized as [(reads mapped to gene)/(gene length in kb)/(genome equivalents)], which indicates the fraction of total cells encoding the gene of interest, i.e., copies per cell. The relative abundance of *arsM* genes in the 18 metatranscriptomic data sets was normalized as RPKM [(reads mapped to gene)/(gene length in kb)/(total millions of reads)].[32]

Dissolved Organic Matter (DOM) Extraction, Characterization, and Analyses. DOM was extracted at a soil-to-water mass ratio of 1:10, concentrated using an Agilent Bond Elut PPL (500 mg, 6 mL) column and dissolved in 1 mL of a 1:1 methanol−water mixture as previously described.[33] The molecular composition of DOM was analyzed using a SolariX-type FT-ICR MS (Bruker Corporation) instrument equipped with an electrospray ionization (ESI) source. Details of the extraction and detection methods for DOM are provided in the Supporting Information. The relative abundance of each molecular formula (MF) in the sample was calculated by dividing the peak intensity of each MF by the sum of the peak intensities of all MFs in the sample. The'vegan' package in R was used to calculate the alpha diversity (Shannon, Simpson, and Chao1 indices) of the DOM. The MFs were treated as individual species, and their relative abundances in the samples were considered equivalent to the species abundance. The MFs of DOM were assigned to molecular compound classes based on the recommended classification criteria (data-based chemical class regions for Van Krevelen diagrams): amino sugars-like (H/C = 1.62−2.35, O/C = 0.56−0.95), carbohydrates-like (H/C = 1.53−2.20, O/C = 0.56−1.23), lignin-like (H/C = 0.86−1.34, O/C = 0.21−0.44), lipids-like (H/C = 1.34−2.18, O/C = 0.01−0.35), peptides-like (H/C = 1.33−1.84, O/C = 0.17−0.48), and tannins-like (H/C = 0.70−1.01, O/C = 0.16−0.84).[34]

SEM and Random Forest Analyses for the Estimation of Important Environmental Factors. Partial least-squares structural equation modeling (PLS-SEM) was used to explore the effects of multiple variables on *arsM* richness via the R package'plspm'. The soil variables included TOC and TN. The abundances of carbon, nitrogen, phosphorus, and sulfur metabolism genes were used as CNPS metabolism variables. The final model selection was based on Cronbach's alpha (C.alpha), Dillon-Goldstein's rho (DG. rho), loading, average variance extracted (AVE), and goodness of fit (GoF), which evaluate the model's reliability and goodness of fit. The random forest model using the R package'randomForest' with ntree = 226 was applied to predict the relative importance scores of the factors involved in the PLS-SEM. The confidence level of each factor was subsequently calculated using the R package'rfPermute' with nrep = 1000.

Lignin, Peptide and Tannin Addition to the Paddy Soils. Considering the geographic location (South and North China), total As concentration (contaminated and non-contaminated) and relative abundance of *arsM* genes (relatively high and low) of the paddy soils, soils collected from Tancheng, Shandong Province (SD_TC: North China; As levels <10 mg kg^{-1}; transcribed *arsM* abundance: average of 4.19×10^{-2}) and Maguan, Yunnan Province (YN_MG: South China; As levels: 15−30 mg kg^{-1}; transcribed *arsM* abundance: average of 5.78×10^{-3}) were further selected to validate the effects of lignin and peptide on As methylation mediated by *arsM* communities. According to the correlation of the soil content of different DOM components with the relative abundance of *arsM* genes, lignin (CAS number of 8068-05-1), peptide (CAS number 158563-45-2) and tannin (CAS number 1401-55-4) were selected to study their contribution to As methylation efficiency. The soils were incubated under anaerobic conditions. Twenty-five grams of soil (<2 mm) was placed in a 120 mL serum bottle, and 50 mL of deionized water was added. After 7 days of preincubation, in combination with the commonly suggested straw application amount (1%) and previous studies of As methylation affected by different carbon sources,[18] 0.125 g of lignin, peptide or tannin was added and mixed with 25 g of soil in serum bottles, amounting to dissolved organic matter inputs of 0.5% of the soil weight. N_2 was injected for 30 min to maintain anaerobic conditions, and the bottles were then sealed with rubber stoppers and aluminum crimp seals. Each treatment was repeated in triplicate, and the samples were incubated at 25 °C in the dark. Porewater samples were collected by destructive sampling at days 0, 3, 7, and 14 of incubation. The porewater samples were acidified with concentrated HCl (100:1, v:v) to prevent As precipitation and transformation and filtered through 0.22 μm membrane filters.[8] The As species in the porewater were quantified immediately (within 24 h) after destructive sampling via high performance liquid chromatography (HPLC 1260II, Agilent, CA, USA) inductively coupled plasma mass spectrometry (ICP-MS 7850, Agilent, CA, USA). The soil samples were collected on day 7, and the soil microbial DNA and RNA were extracted as described. qPCR of the *arsM* gene (at the RNA level) was performed, and the V3 and V4 regions of the bacterial 16S rRNA gene were amplified and sequenced on a NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA) at Guangdong Magigene Biotechnology Co., Ltd. (Guangzhou, China). Details are provided in the Supporting Information.

Statistical Analyses. The alpha diversity (Shannon and Richness) of the microbial communities and *arsM* genes was calculated by the R package'vegan'.[35] The beta diversity of the *arsM* communities was assessed by Bray−Curtis similarity using the R packages'vegan' and'dplyr'.[36] The distance decay of the *arsM* gene and DOM molecules was calculated by the R package'geosphere'.[37] Spearman correlation analyses were performed with the R packages'ggcorrplot'[38] and'corrplot'.[39] Comparisons between groups were analyzed using the Wilcoxon test (R package'stats'). The R package'linkET'[40] was used for the Mantel test. The'ggmap'[41] R package was used to construct a map, and the phylogenetic tree was visualized using tvBOT v2.5.0 55.[30] The Sankey graph was constructed with the R packages'tidyverse'[42] and'networkD3'.[43] The remaining of the plots were plotted by the R package'ggplot2'.[44]

RESULTS

Overview of Soil Properties and DOM Components in China Paddy Soils. In total, 141 paddy soil samples were

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

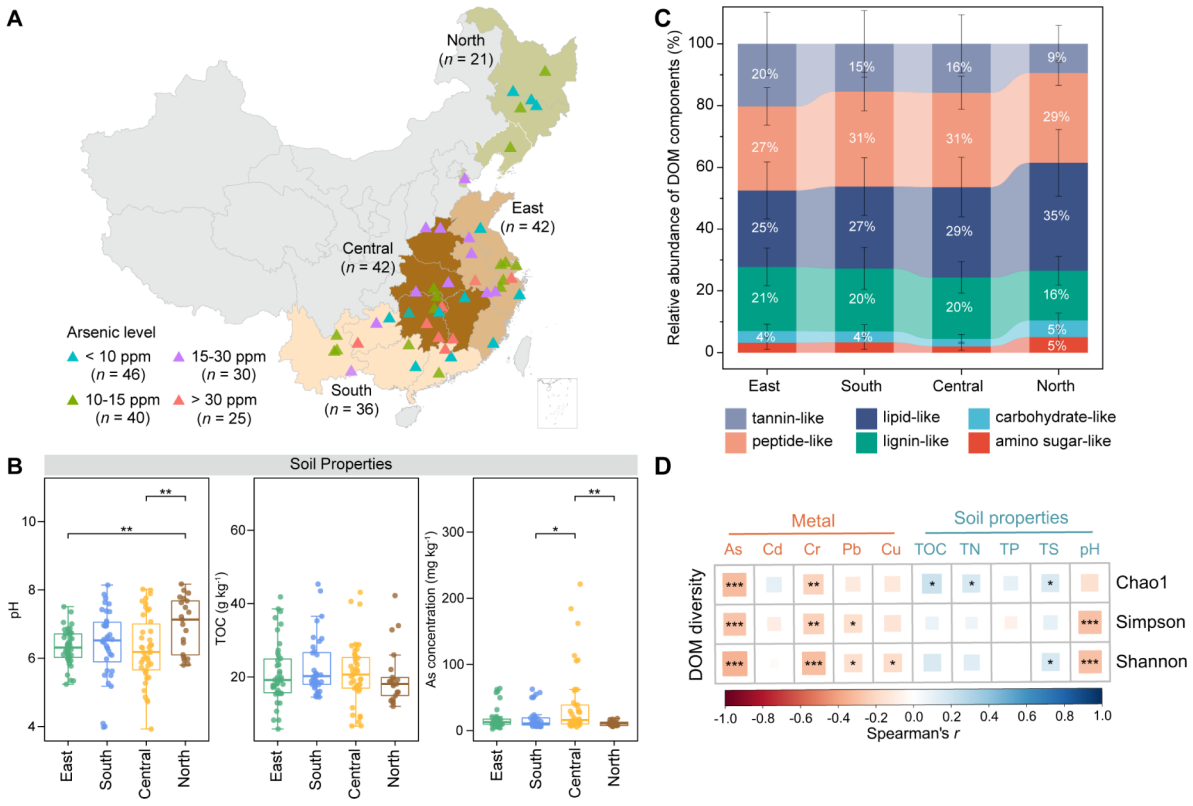


Figure 1. Sampling sites, soil properties and characterization of DOM components. (A) Distribution of the sampling sites. Each triangle represents a sampling site (3 replicates), and the color indicates the As concentration. The colors of the map blocks represent different regions, including North China, East China, Central China, and South China. (B) Soil properties (pH, TOC, and As concentration) of the samples. The boxes indicate the 25th to 75th percentiles (with the median as a horizontal line), and the whiskers represent the maximum and minimum values. All significance values were determined by two-sided Wilcoxon tests (*: $p < 0.05$, **: $p < 0.01$). (C) Relative abundances of DOM components in East China ($n = 42$), South China ($n = 36$), Central China ($n = 42$), and North China ($n = 21$). Each error bar corresponds to the standard deviation (SD), and the data are shown as the mean ± SD D Spearman correlation between environmental factors (soil properties and metal concentrations) and the alpha diversity of DOM molecules.

collected from northern ($n = 21$), eastern ($n = 42$), central ($n = 42$) and southern ($n = 36$) China at the continental scale (Figure 1A). Among the characterized soil properties, the soil pH and As content varied significantly ($p < 0.05$) among the different districts (Figure 1B). The soil pH in these paddy soils ranged from 3.9 to 8.2, and the As content ranged from 2.3 to 221.3 mg kg^{-1} (Table S1). The total content of organic carbon (TOC) ranged from 5.81 to 45.29 g kg^{-1}, and no significant difference was detected among the different districts ($p > 0.05$; Figure 1B). Other heavy metals, including Cd, Cr, Pb and Cu ranged from 0−6.87, 9.07−197.69, 0.14−330.77 and 5.31−193.18 mg kg^{-1}, respectively. The highest concentrations of these heavy metals were observed in the central districts (Figure S1). In the metagenomic data sets, an average of 84,829,684 clean reads were obtained after trimming each sample, and an average coverage of 53% was achieved based on the read redundancy value estimated by Nonpareil v3.303[45] for each data set (Table S2).

The soil DOM in the different districts was dominated by peptide-, lipid-, lignin-, and tannin-like components, which accounted for an average of 29.4%, 28.9%, 19.3%, and 15.2%, respectively, of the total DOM (Figures 1C and S2A), and were affected mostly by the soil pH, TOC, TN, and TS (Supporting Information). Spearman's correlation of the content of soil heavy metal (loid)s with the chemodiversity of soil DOM (Figure 1D) indicated that As in paddy soils was the most important factor that was significantly ($p < 0.001$) negatively correlated with both the richness (Chao1) and the Shannon and Simpson diversity indices of soil DOM. Similarly, the DOM compositions also exhibited significant ($p < 0.01$) differences in paddy soils among different As levels based on Bray−Curtis dissimilarity (<10, 10−15, 15−30 and >30 mg kg^{-1}; Figure S2B).

Relative Abundance of Total and Transcriptional Active *arsM* Communities and Their Correlations with DOM Contents. At the DNA level, the relative abundance of *arsM* genes ranged from 2.12 to 4.07 copies per cell in paddy soils at the continental scale. A significantly greater abundance of the *arsM* gene was detected in paddy soils in southern (average of 3.27 copies per cell) and central (average of 3.20 copies per cell) China than in eastern (average of 3.04 copies per cell, $p < 0.01$) and northern (average of 2.86 copies per cell, $p < 0.001$) China. The relative abundance of the expressed *arsM* gene (at the RNA level) was also greater in paddy soils collected from southern and central China than in those collected from eastern and northern China (Figure 2A). The relative abundance of the *arsM* gene at both the DNA (Figure 2B) and RNA levels (Figure 2C) was significantly ($p < 0.05$) positively correlated with the soil content of lignin- and peptide-like DOM components, accounting for 39.2% and 26.0% of the total significant correlation of DOM molecules with *arsM* genes in terms of relative abundance at the DNA level, respectively, and 32.2% and 28.1% of the total significant correlation of DOM molecules with *arsM* genes in terms of

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

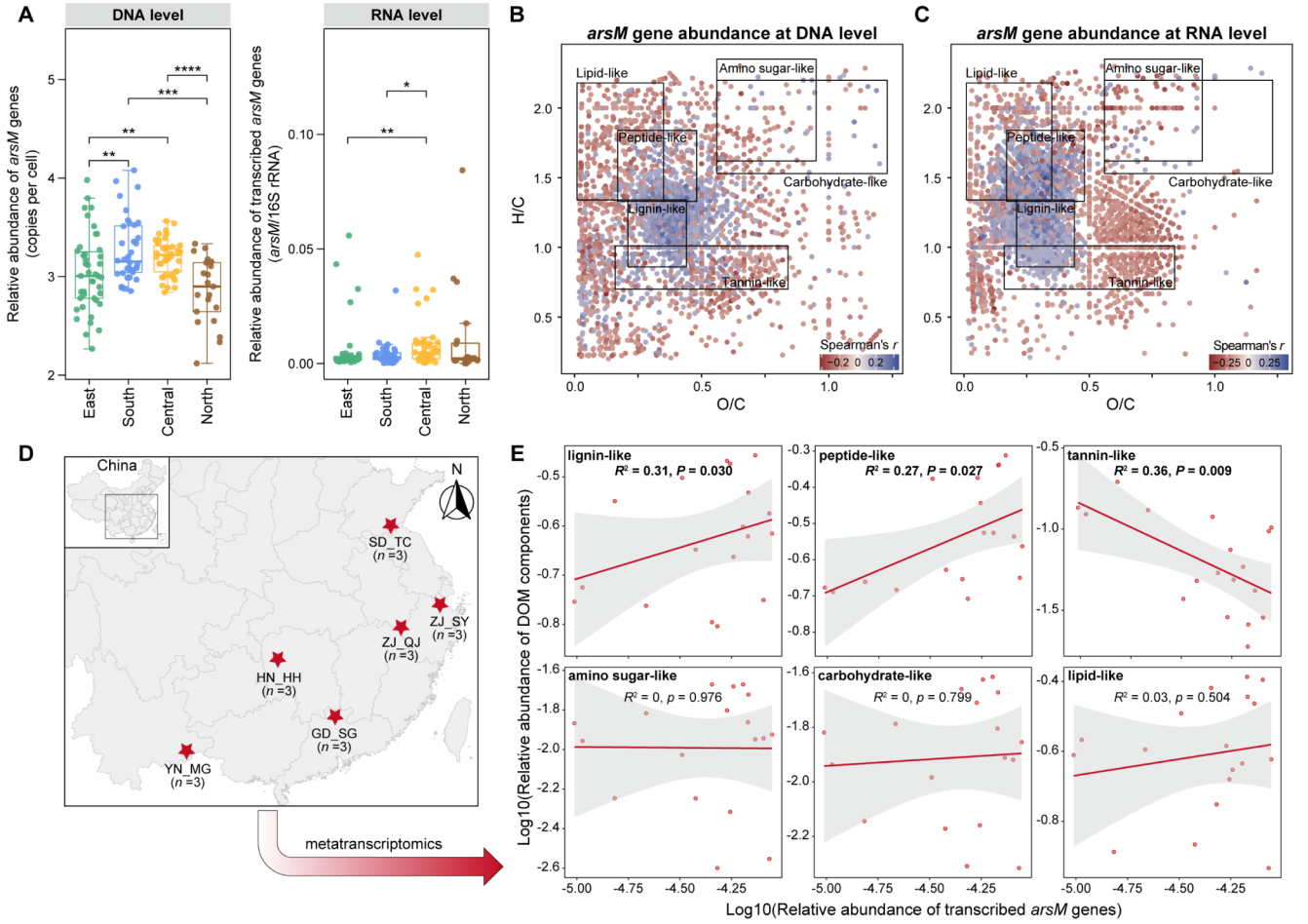


Figure 2. Relative abundances of *arsM* genes at the DNA and RNA levels and their correlation with DOM content. (A) Relative abundance of *arsM* genes at the DNA level (based on metagenomic analyses) and RNA level (the ratio of *arsM* gene transcripts to 16S rRNA). The boxes indicate the 25th to 75th percentiles (with the median as a horizontal line), and the whiskers represent the maximum and minimum values. All significance values were determined by two-sided Wilcoxon tests (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$). The distribution of DOM molecules significantly correlated (Spearman correlation, $p < 0.05$) with the relative abundance of *arsM* genes at the DNA (B) and RNA (C) levels. (D) Sites of the paddy soils (18 samples) used for metatranscriptomic sequencing. (E) Linear regression analysis of the relative abundance of different DOM components against the relative abundance of the transcribed *arsM* genes.

relative abundance at the RNA level, respectively (Figure S3). To further validate the effects of lignin- and peptide-like DOM components on the *arsM* communities, a total of 18 paddy soil samples were collected from Shandong, Zhejiang, Hunan, Guangdong and Yunan Provinces and subjected to metatranscriptomic analyses (Figure 2D and Table S3). The relative abundance of transcribed *arsM* genes ranged from 2.37×10^{-5} to 7.65×10^{-5} RPKM (Table S4). A significant ($p < 0.05$) positive correlation was observed between the relative abundance of transcribed *arsM* genes and the soil content of lignin-like ($R^2 = 0.31$) and peptide-like DOM components ($R^2 = 0.27$), whereas the soil content of tannin-like DOM components was significantly ($p < 0.01$) negatively correlated with the relative abundance of transcribed *arsM* genes ($R^2 = 0.36$). For the other dominant DOM structures, such as amino sugar-, carbohydrate- and lipid-like components, no significant correlation was revealed between their contents and the relative abundance of transcribed *arsM* genes ($p > 0.05$; Figure 2E).

Diversity of *ArsM* Communities and Their Correlations with Soil DOM Components. A nonmetric multidimensional scaling (NMDS) plot of *arsM* communities based on Bray−Curtis dissimilarity revealed significant differences in *arsM* communities between paddy soils from different districts (Figure 3A; PERMANOVA, $p < 0.001$). Moreover, the similarity of *arsM* communities was negatively correlated with geographic distance, indicating a significant distance-decay pattern for As(III) methylation microbial community similarity ($p < 0.001$; Figure S4). Soil properties, including TOC and pH, are the most important factors influencing the abundance and diversity of *arsM* genes in paddy soils (Figure S5). According to the taxonomic classification based on the short-reads, bacteria were the predominant microbes in these soils, accounting for 97.4−97.7% of the total microbes (Figure S6). Eighty bacterial MAGs and 10 archaeal MAGs were binned from the metagenomic data sets with completeness ≥90% and contamination ≤5% and were mostly assigned to *Pseudomonadota*, *Acidobacteriota*, *Chloroflexota*, *Actinobacteriota* and *Bacteroidota* for bacteria and *Nitrososphaerota* for archaea. Approximately 67% of these bacterial MAGs carried *arsM* genes according to a relatively strict cutoff of amino acid identity greater than 60% and query length coverage greater than 70% (Figure S7). At the class level, these 80 bacterial and 10 archaeal MAGs were assigned mostly to *Anaerolineae*,

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

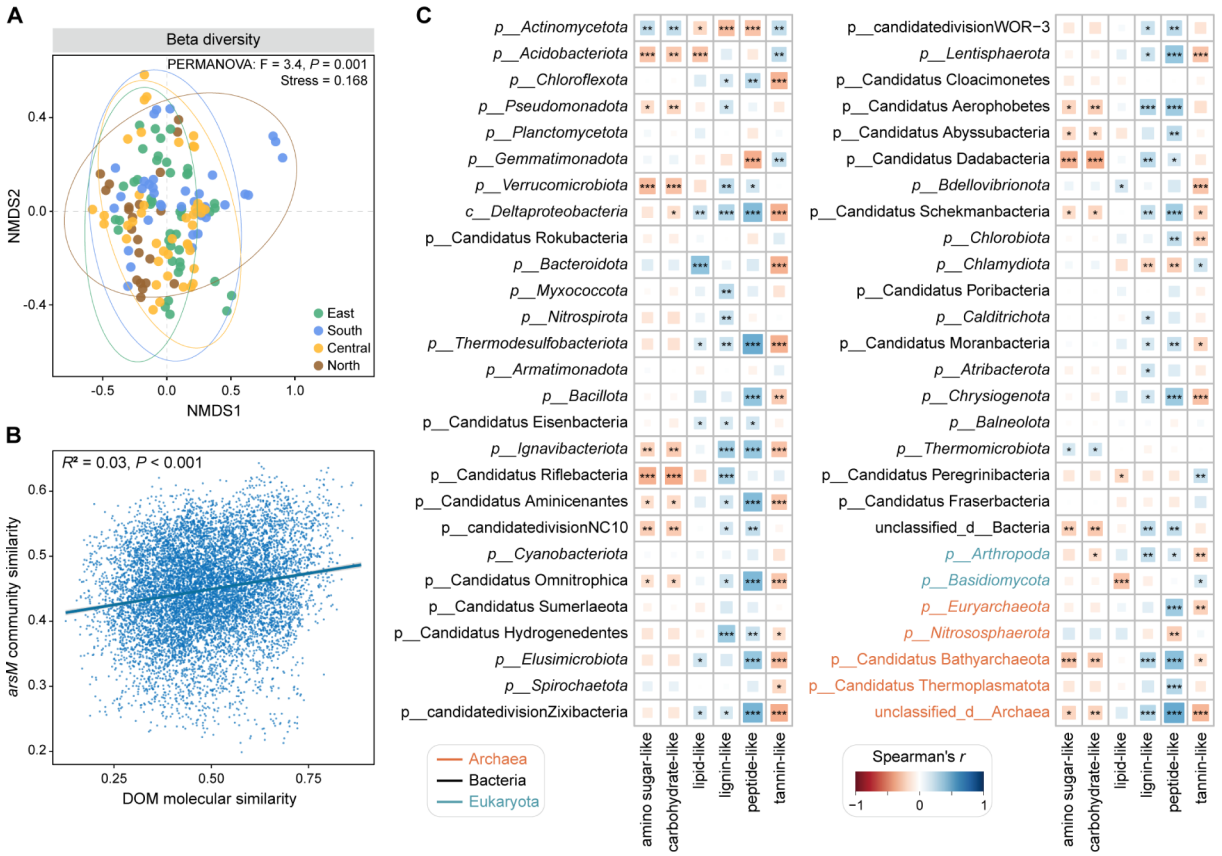


Figure 3. Diversity of *arsM* communities and their correlation with DOM components. (A) The beta diversity of the *arsM* genes. Nonmetric multidimensional scaling (NMDS) analysis was applied based on the Bray−Curtis distance (PERMANOVA test, $F = 3.4$, $p = 0.001$). The green, blue, yellow and brown colors indicate samples from East China ($n = 42$), South China ($n = 36$), Central China ($n = 42$), and North China ($n = 21$), respectively. (B) Spearman correlation of the similarity of *arsM* communities and DOM molecules. Each point represents a pair of sites. (C) Spearman correlation between microbes (at the phylum level) carrying the *arsM* gene and DOM components. The orange, black and blue colors represent *Archaea*, *Bacteria* and *Eukaryota*, respectively.

Alphaproteobacteria, *Thermoleophilia*, *Gemmatimonadetes*, *Gammaproteobacteria*, *Nitrospiria*, *Myxococcia*, *Phycisphaerae*, *Dehalococcoidia*, *Acidimicrobiia*, and *Bacteroidia* (Figure S8A), which accounted for approximately 29.4−32.6% of the total *arsM* community (Figure S8B). In paddy soils, *arsM* communities in the phyla *Chloroflexota* and *Acidobacteriota* were more abundant in southern China, whereas *Chitinophagaceae* (phylum *Bacteroidota*), *Planctomycetota* and *Gemmatimonadales* (phylum *Gemmatimonadota*) had the greatest relative abundances in central China. Other *arsM*-carrying MAGs assigned to *Actinobacteriota*, *Desulfobacterota*, *Mycococcota*, *Nitrospirota*, and *Pseudomonadota* were evenly distributed in paddy soils from different districts in China (Figure S8C).

A significant ($p < 0.001$) positive correlation between the similarity of *arsM* communities and DOM compositions was further observed in paddy soils across China, albeit with a relatively low correlation ($R^2 = 0.03$; Figure 3B). With respect to the significant correlation of the DOM components with the *arsM* communities, lignin-, peptide-, and lipid-like components were the predominant DOM components that exhibited a significant ($p < 0.05$) positive correlation with the variances in the *arsM* communities (PCoA1 axis value; Figure S9), accounting for 80.1% of the total significant positive correlations observed between the *arsM* communities and the DOM components. Specifically, the relative abundance of lignin- and/or peptide-like DOM components in paddy soils was significantly ($p < 0.05$) positively correlated with the abundances of members of the phyla *Pseudomonadota*, *Deltaproteobacteria*, *Bacillota*, *Chloroflexota*, *Nitrospirota*, *Thermodesulfobacteriota*, *Verrucomicrobiota*, and a variety of uncultured bacteria identified in environmental samples that have not been classified and named candidates (Figure 3C). The different DOMs, mostly lignin- and peptide-like components, were also significantly ($p < 0.05$) positively correlated with the relative abundance of the archaeal *arsM* communities, which included *Euryarchaeota*, *Nitrososphaerota*, candidates and unclassified archaea. Specifically, the significant ($p < 0.05$) correlation pattern of the DOM components with the *arsM* communities was detected mostly in paddy soils with As levels ranging from 15 to 30 mg kg^{-1} rather than in those with As levels higher than 30 or lower than 10 mg kg^{-1}. Moreover, the correlation patterns of the soil DOM structure with different *arsM* communities varied in paddy soils with different As levels (Figure S10).

Variables That Affect *ArsM* Gene Abundance and Diversity. Among the variables, microbial diversity, relative gene abundance of N, C and S metabolism, soil pH, the content of As, and the contents of lignin- and peptide-like DOM components were the most important factors contributing to the variances in *arsM* community abundance and diversity, as indicated by random forest (RF) analyses (Figures 4A and S11). The partial least-squares structural equation model (PLS-SEM; goodness of fit of 0.46) was further used to describe the direct and/or indirect interaction pathways among

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

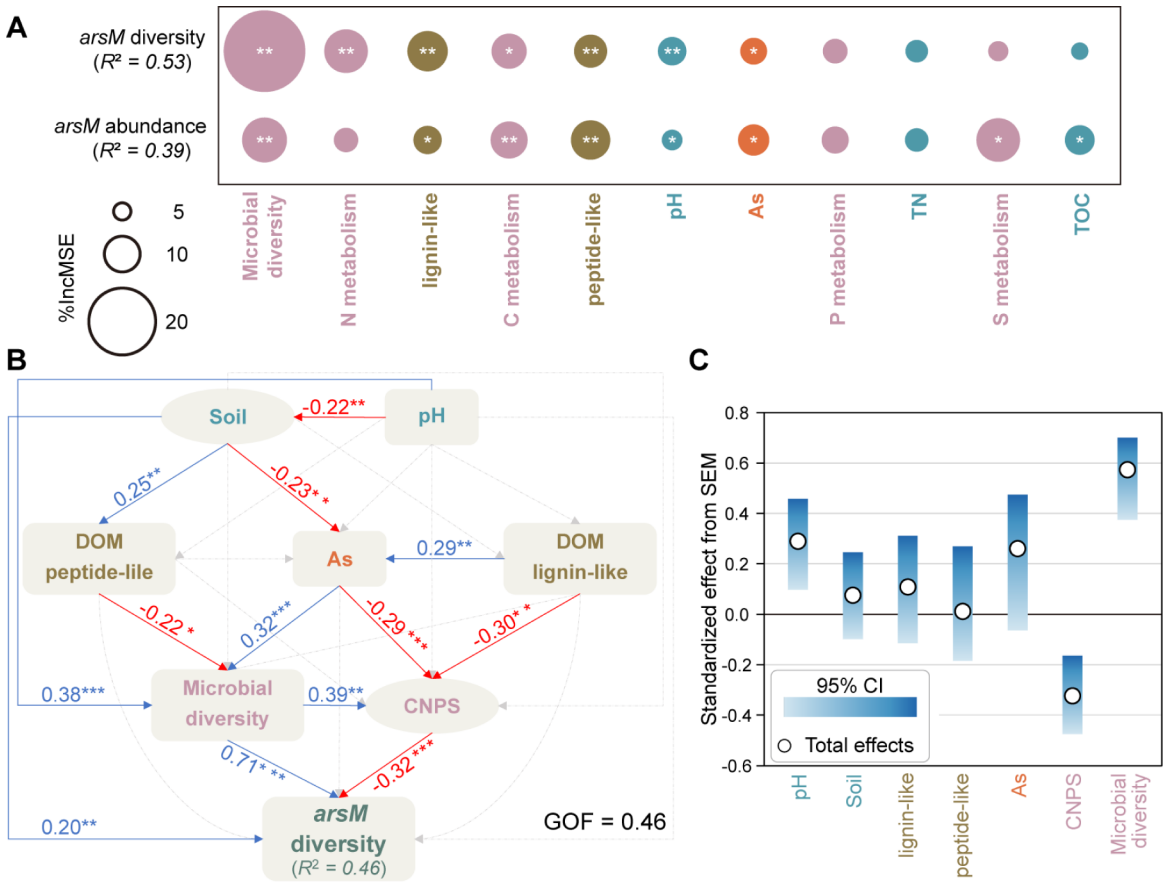


Figure 4. Random forest and partial least-squares-structural equation modeling (PLS-SEM) analyses of the important factors contributing to *arsM* gene diversity in paddy soils. (A) Mean predictor importance of environmental factors (pH, TOC, total nitrogen (TN), and As concentration), DOM components (relative abundance of lignin- and peptide-like DOM components), microbial diversity (richness index), and elemental metabolism genes (relative abundance of carbon, nitrogen, phosphorus, and sulfur metabolism genes) on the richness and abundance of *arsM* genes based on random forest modeling (permutation test, $n = 1000$, *: $p < 0.05$; **: $p < 0.01$). (B) PLS-SEM illustrating the direct and indirect effects on the richness of *arsM* genes. Blue, red, and gray indicate significantly positive, significantly negative, and nonsignificant effects, respectively (*: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$). Path coefficients and coefficients of determination (R^2) were calculated after bootstrapping ($n = 100$). Different colors represent different types of variables. (C) 95% confidence intervals (CIs) for the standardized effects of different variables on *arsM* richness from PLS-SEM. Inc.MSE: increase in the mean squared error; GOF: goodness of fit.

the important variables affecting the diversity of *arsM* communities, which explained 46% of the total variance in *arsM* communities in Chinese paddy soils (Figure 4B). Overall, microbial diversity and elemental metabolism gene abundance were the factors that contributed the most to *arsM* community diversity, with the greatest total positive and negative effects, respectively (Figure 4C). Soil properties (TOC and TN) and microbial community diversity directly impact the diversity of *arsM* communities, with significant ($p < 0.05$) positive path effects of 0.20 and 0.71, respectively. However, the relative abundance of elements in the metabolic communities (C, N, S and P) significantly ($p < 0.05$) negatively affected the diversity of the *arsM* communities, with a path effect of −0.32. The soil pH and As content contributed to the variance in the *arsM* communities mostly through indirect effects on either the microbial community diversity or the functional gene abundances. Among the DOM components, the soil content of lignin-like components impacted the *arsM* community diversity by directly impacting the relative abundance of elemental metabolism communities (path coefficient of −0.30), while the peptide-like component content impacted the *arsM* community diversity by directly impacting the whole microbial community diversity (path coefficient of −0.22). The lignin-like component content also significantly ($p < 0.05$) directly affected the soil As content (path coefficient of 0.29) and had an overall greater effect on *arsM* diversity than did the peptide content (Figure 4C).

Lignin and Peptide Increased the Abundance of Transcribed *ArsM* Genes and Enhanced as Methylation in Paddy Soils. Paddy soil samples collected from Tancheng (SD_TC) and Maguan (YN_MG) were selected for the incubation experiment with the addition of lignin and peptide (Figure S12). Overall, the concentration of methylated As, including MMA and DMA, increased during the incubation time from day 0 to day 7 (Figure 5A and Table S5) and gradually decreased after incubation for 7 days (Figure S13), possibly due to methylation to volatile trimethylarsine (TMA) or demethylation to inorganic As.[46,47] Addition of lignin and peptide increased the concentration of methylated As than that in the control group without DOM addition (below the detection limit of ICP-MS; less than 0.490 $\mu g\ kg^{-1}$). For sample SD_TC, the concentration of methylated As in the lignin treatment group was significantly (two-sided *t* test, $p < 0.05$) greater than that in the peptide addition group on both day 3 (25.3 vs 2.9 $\mu g\ kg^{-1}$, respectively) and day 7 (32.2 vs 9.8 $\mu g\ kg^{-1}$, respectively; Figure 5A). For sample YN_MG, overall,

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

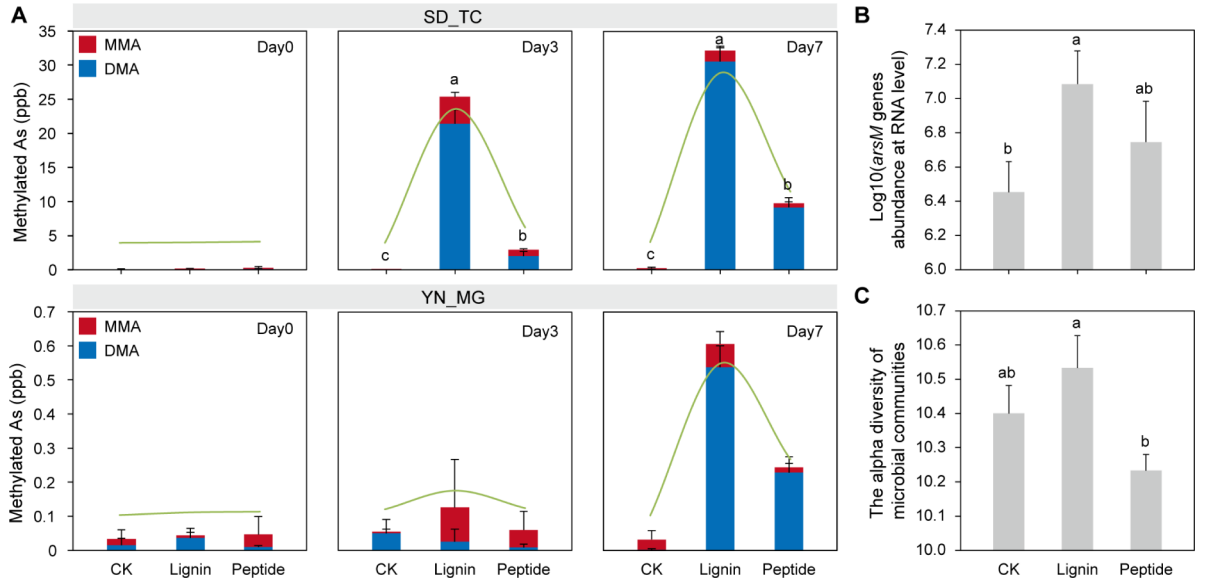


Figure 5. Validation of the effects of lignin and peptide on As(III) methylation via incubation experiments. (A) Concentration of methylated As in media incubated with soils from SD_TC and YN_MG with lignin and peptide addition. (B) Relative abundance of *arsM* genes in SD_TC soils at the RNA level (copies per μL) with lignin and peptide addition. (C) Diversity of microbial communities in SD_TC soils with the addition of DOM components (lignin and peptide) and without DOM addition (CK). Each error bar corresponds to the standard deviation (SD), and the data are shown as the mean ± SD. Significant differences between groups are indicated by different lower-case letters (two-sided *t* test, $p < 0.05$).

the concentration of the methylated As species was below the detection limit of ICP-MS during the incubation time from day 0 to day 3 (Table S5). On day 7, an increase in methylated As was detected in the lignin treatment group (0.61 μg kg^{-1}, respectively) compared with the peptide and control treatment groups (below detection limit). To investigate how lignin and peptide affect the transcriptional activity of *arsM* genes and the microbial communities in the soils, we selected the soil samples collected on day 7, which presented the highest As methylation efficiency, for further analyses. In addition, considering the relatively high efficiency of As methylation, the SD_TC sample was further selected for qPCR analysis of the abundance of the *arsM* gene (at the RNA level), and the 16S rRNA gene was sequenced to estimate the microbial diversity. An increase in the abundance of the transcribed *arsM* genes was detected in the treatment with lignin (1.34×10^7 copies per μL) and peptide (6.32×10^6 copies per μL) addition compared with the control (3.06×10^6 copies per μL), and a significant (two-sided *t* test, $p < 0.05$) difference was observed in the group with lignin addition but not with peptide addition (Figure 5B). Compared with the control treatment, the addition of lignin also increased microbial diversity, although no significant difference was detected (two-sided *t* test, $p > 0.05$; Figure 5C). Nevertheless, in the group with peptide addition, the microbial diversity was significantly lower than that in the control group (two-sided *t* test, $p < 0.05$; Figure 5C).

DISCUSSION

Great Potential for as Methylation by *ArsM* Communities in Paddy Soils in China. *arsM* genes were prevalent and associated with transcriptional activity in 141 paddy soils across northern, eastern, central, and southern China (Figure 2), suggesting that microbially mediated As methylation has great potential in paddy soils at the continental scale. The potential for As methylation to be acquired by diverse *arsM* communities in paddy soils is the key factor underlying the accumulation of DMA in paddy soils[48] and thus induces straight-head disease in rice plants.[48,49] The significant ($p < 0.001$) differences in the As methylation microbial community compositions in different districts of China (Figure 3A) and the distance-decay pattern in the similarity of the *arsM* communities (Figure S4) indicated that As methylation in paddy soils was dominated by different microbial communities across China. A previous study based on metatranscriptomic analyses also suggested that the *Actinobacteriota*, *Bacteroidota*, *Acidobacteriota* and *Chloroflexota* phyla identified herein exhibited increased expression levels of *arsM* genes in paddy soils associated with rice straight-head disease.[48] In addition, since the soil As content ranged from background (<15 mg kg^{-1}) to contaminated levels (>15 mg kg^{-1}), our results showing the wide presence of *arsM* genes with transcriptional activity further support that microbial-mediated As methylation could occur even in paddy soils with background As levels[50,51] and thus explain why this physiological disorder of rice is also detected in Shandong and Jiangsu Provinces, where As contamination has seldom been reported.[48] Nevertheless, the content of different As species, especially the methylated As in the soil porewater, should be analyzed in situ in paddy soils by using Rhizon porewater samplers to better understand the contribution of transcriptional active *arsM* communities to methylated As accumulation in such area.

Soil DOM Structures Contributed to the Variance in the Abundance and Diversity of the *ArsM* Communities. The TOC content was confirmed to be an important factor affecting the abundance and diversity of the *arsM* gene in the present study (Figure S5) and elsewhere, as previously reported,[52,53] suggesting the impact of soil organic carbon on *arsM* communities. Moreover, we demonstrated that the DOM structure determined the abundance and diversity of *arsM* communities in paddy soils at the continental scale in China (Figures 2 and 3). Considering that microbial taxonomic diversity is correlated with DOM molecular diversity[16,54] and that *arsM* genes appear to be widely present in microbial

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

genomes,[55] it is reasonable that DOM molecular compositions also exhibit a significant ($p < 0.05$) positive correlation with the similarity of *arsM* community compositions (Figure 3A). Consistently, the DOM composition has been shown to be associated with the relative abundances of microbial metabolic pathways,[16] demonstrating their potential correlation with the functional microbial communities reported herein. Soil pH, which has been established as a key factor controlling DOM profiles,[54,56] is also correlated with the diversity of *arsM* communities in paddy soils.[6,57,58]

Among the various DOM structures, the soil contents of lignin- and peptide-like components in paddy soils significantly contributed to the abundance and diversity of the *arsM* communities ($p < 0.05$; Figure 4). Similarly, the selective utilization of organic carbon molecules has been reported to significantly increase As methylation efficiency and increase methylated As in paddy soils collected from Guizhou, Zhejiang and Hunan Provinces.[18] Specifically, DOM with different O/C ratios could be directly or indirectly involved in the methionine or tetrahydrofolate metabolic pathway to promote S-adenosylmethionine synthesis in As methylating microorganisms.[18] Our results better reflect the geographic distribution of *arsM* communities and their correlations with various DOM molecules in paddy soils at the continental scale in China. The significant positive correlation of the content of lignin- and peptide-like DOM components with the abundance of *arsM* communities ($p < 0.05$; Figure 3C) was mainly revealed in the phyla *Chloroflexota*, *Verrucomicrobiota*, *Deltaproteobacteria*, and *Thermodesulfobacteriota*, which have been demonstrated to be predominant taxa in *arsM* communities.[48,51,59] Other candidates of uncultured taxa, such as *Aminicenantes*, *Omnitrophica*, *Hydrogenedentes*, *Zixibacteria*, *Aerophobetes*, *Dadabacteria*, *Schekmanbacteria* in Bacteria, and *Bathyarchaeota* in Archaea, together with the previously reported phyla of *Ignavibacteriota,* also exhibited significantly ($p < 0.05$) positive correlations in their abundances with the contents of lignin- and peptide-like DOM components. The phylum *Ignavibacteriota* is ubiquitously abundant in paddy soils but is seldom reported to be involved in As methylation.[60−62] Based on a targeted metagenomic/transcriptomic approach, *Ignavibacteria* genomes were recovered from straw-amended paddy soils in Italy and showed transcriptional activity for more than half of the carbohydrate-active enzymes (CAZymes), suggesting that these bacteria contribute to the decomposition of complex polymers.[62] We suspected that these microbes constituted a considerable part of the *arsM* community and were determined by the DOM composition in paddy soils, although further in situ investigations should be carried out to verify their contribution to As methylation in paddy soils. Moreover, the proposed elemental ratio bounds for determining different DOM components are not definitive regions, which is highlighted by the overlap between the different chemical classes.[34] Therefore, for more precise characterization of different DOM components and estimation of their effects on *arsM* communities, additional analyses of DOM components, such as biomarker extraction experiments for lignin and peptide, are also desirable in future studies. In addition, while van Krevelen diagrams are a widely used visualization technique for DOM data, other visualization approaches should also be considered in further studies, such as Kendrick mass defect plots to identify formulas that are related by carboxylic and methylene groups, or carbon nominal oxidation state against carbon number to distinguish between oligomerization, fragmentation, and functionalization of DOM components.[34]

Mechanism of How Lignin and Peptide Impact as Methylation. The presence of lignin- and peptide-like DOM components in paddy soils could affect As methylation efficiency by affecting the abundance of the *arsM* gene as well as the diversity of the *arsM* community (Figure 4A). Similarly, lignin- and peptide-like DOM components in paddy soils increased the abundance of transcriptionally active *arsM* communities, as shown by the significant positive correlation of the abundance of the transcribed *arsM* genes with lignin and peptide contents in paddy soils in situ based on metatranscriptomic analyses of the 18 selected paddy soils ($p < 0.05$; Figure 2E). Owing to the limitation of paddy field soil preservation during transportation, the possible degradation of soil RNA might result in possible variances in the transcribed *arsM* gene abundances among different samples. However, this effect could be neglected, especially considering that the soil RNA was checked for integrity before it is used for further analyses. Consistently, further incubation experiments, in which the soil RNA was extracted immediately after incubation, also revealed a consistently increased abundance of transcribed *arsM* genes in the DOM treatment group (lignin and peptide) compared with the control group (Figure 5B). Moreover, the addition of lignin and peptide significantly increased the concentration of methylated As on day 7 in SD_TC soils compared with those without DOM addition ($p < 0.05$; Figure 5A and Table S5), confirming that the As methylation efficiency increased when lignin and peptide were applied. During incubation, we also observed a decrease in the soil Eh after the addition of lignin and peptide for the incubation of both SD_TC (averages of 81.6 and 109.3 mV, respectively) and YN_MG (averages of 86.7 and 123.9 mV, respectively) soils compared with the control treatment (average of 275.6 and 454.2 mV for SD_TC and YN_MG, respectively), especially on day 7 (Figure S14). The decrease in soil Eh could also contribute to the increased transcriptional activity of *arsM* communities. Previous study has demonstrated that the small variation in soil redox potential can cause a large variation in the abundance of As methylating microorganisms.[48] In addition, the increased concentration of As(III) during incubation in the treatment with lignin and peptide addition (Table S5) could also provide more abundant substrates for microbially mediated As methylation and thus result in increased As methylation efficiency.

Specifically, lignin had a greater impact on As methylation than did the peptide, as confirmed by continental-scale analyses and incubation experiments. At the continental scale, a greater percentage of the significant ($p < 0.05$) positive correlations were revealed between the abundance of the *arsM* gene and lignin than between the abundance of the *arsM* gene and peptide (39.2% vs 26.0%; Figure S3). Additionally, PLS-SEM analyses revealed that the total effect of lignin on *arsM* community diversity was greater than that of peptide on *arsM* community diversity (0.1083 vs 0.0109; Figure 4). Similarly, the methylated As content in the lignin-treated group was significantly greater than that in the peptide-treated group ($p < 0.05$; Figure 5A), indicating that lignin facilitated As methylation more strongly than did peptides. The greater contribution of As methylation by lignin could be attributed to the fact that lignin not only significantly increased the abundance of the *arsM* gene but also increased the diversity of the microbial community ($p < 0.05$; Figure 5B,C). However,

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

no significant difference in the abundance of transcribed *arsM* genes was detected between the treatment with peptide addition and the control treatment. Moreover, in contrast to the increase in the diversity of microbial communities, the addition of peptides significantly decreased the diversity of microbes compared with that resulting from the addition of lignin ($p < 0.05$; Figure 5C). Considering that the diversity of microbial communities is the most essential factor contributing to the diversity of *arsM* communities (Figure 4C), it is reasonable that the negative impact of peptides on microbial community diversity would undermine their ability to facilitate As methylation. While the addition of lignin increased the relative abundances of *Clostridia* and *Anaerolineae* by approximately 9.3% and 2.2%, the addition of peptide increased the relative abundances of *Desulfuromonadia* and *Gammaproteobacteria* by approximately 4.2% and 2.5%, respectively, compared with those in the control treatment (Figure S15). Whereas these taxa have been demonstrated to have As methylation ability,[63−65] the results herein further suggested that these taxa were associated with different As methylation efficiencies.

Furthermore, the effect of tannin, which was negatively correlated with *arsM* genes in terms of relative abundance, was also evaluated. Overall, no significant increase in As methylation efficiency was observed in the soils with tannin addition during the incubation time for either the SD_TC or YN_MG soils (Figure S13), confirming the various effects of different DOM components on As methylation in paddy soils. Nevertheless, further studies regarding the effects of other DOM components, such as amino acids, lipids and carbohydrates, on As methylation should also be conducted to elucidate the mechanism by which different DOM components affect As methylation. In addition, owing to the limitation of the DOM detected by FT-ICR-MS, which captures only DOM with molecular weights ranging from 117.05−799.39 Da, the possible contribution of high-molecular-weight DOM to As methylation would be neglected. Further studies should also investigate the potential effects of DOM with molecular weights greater than 800 Da on *arsM* communities. Various advanced DOM detection technologies such as nuclear magnetic resonance spectroscopy (NMR) and gas chromatography−mass spectrometry (GC−MS) should also be applied for more comprehensive understanding of the effect of DOM with molecular weights on *arsM* communities.

Environmental Implications. The occurrence of straight-head disease in rice, which is mostly induced by microbially mediated As methylation, is prevalent in many rice-growing regions in China and other Asian countries. From the regional investigation of paddy soils in China, we revealed a positive correlation between the diversity of paddy soil DOM and *arsM* communities. Among the different DOM components, the contents of lignin- and peptide-like DOM are the most important factors affecting the abundance, activity, and diversity of *arsM* communities. The positive correlation between the soil content of lignin- and peptide-like DOM components and *arsM* gene abundance was further demonstrated by the increased abundance of the transcribed *arsM* genes and As methylation efficiency in paddy soils incubated with lignin and peptide addition. This study highlights the correlation between the chemodiversity of soil DOM components and the diversity of *arsM* communities in paddy soils and reveals the mechanism by which lignin and peptide facilitate As methylation in paddy soils. Considering that organic matter fertilization, such as manure application and straw return, is commonly applied in agricultural practices, the possible promotion of As methylation due to various DOM components should not be neglected.

■ ASSOCIATED CONTENT

Data Availability Statement

The raw sequencing data of 141 metagenomic and 18 metatranscriptomic data sets were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject of PRJNA1068274 and PRJNA1068685, respectively.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.4c10809.

Figures: soil properties; DOM distribution and composition; number or proportion of the DOM molecules that significantly positively correlated with *arsM* genes; linear regression of the distance-decay relationships of the *arsM* genes; environmental factors affecting *arsM* gene abundance and diversity; relative abundance of different microbial communities; taxonomic classification, phylogenetic tree of MAGs and *arsM* communities; correlation of DOM molecules with *arsM* gene communities; mean predictor importance of factors based on random forest modeling; incubation experimental setup, soil pH, Eh, As concentrations on incubation days 0, 3, 7 and 14; soil microbial taxonomic compositions with the addition of DOM components (lignin and peptide) on day 7; Tables: basic information on the sampling sites and the metagenomic and metatranscriptomic data sets; detailed information on relative abundance of transcribed *arsM* genes; the concentration of methylated As in the incubation experiments (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Xianjin Tang − *Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China;* orcid.org/0000-0002-7515-2047; Email: xianjin@zju.edu.cn

Authors

Si-Yu Zhang − *School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China;* orcid.org/0000-0001-6701-5790

Zi-Teng Liu − *School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China;* orcid.org/0000-0002-3583-4238

Xin-Di Zhao − *School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China*

Zi-Yu Gao − *School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241, China*

Ouyuan Jiang − *Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China*

Jie Li − *Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, College of Environmental and*

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

Resource Sciences, Zhejiang University, Hangzhou 310058, China

Xiaoming Li − *Key Laboratory of Drinking Water Science and Technology, Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China; University of Chinese Academy of Sciences, Beijing 101408, China*

Andreas Kappler − *Department of Geosciences, University of Tübingen, Tübingen 72076, Germany;* orcid.org/0000-0002-3558-9500

Jianming Xu − *Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China;* orcid.org/0000-0002-2954-9764

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c10809

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study is financially supported by the National Natural Science Foundation of China (Grant Nos. 42122048 and 42107135).

■ REFERENCES

(1) Liao, X. Y.; Chen, T. B.; Xie, H.; Liu, Y. R. Soil As contamination and its risk assessment in areas near the industrial districts of Chenzhou City, Southern China. *Environ. Int.* **2005**, *31* (6), 791−798.

(2) Williams, P. N.; Villada, A.; Deacon, C.; Raab, A.; Figuerola, J.; Green, A. J.; Feldmann, J.; Meharg, A. A. Greatly enhanced arsenic shoot assimilation in rice leads to elevated grain levels compared to wheat and barley. *Environ. Sci. Technol.* **2007**, *41* (19), 6854−6859.

(3) Zhu, Y. G.; Sun, G. X.; Lei, M.; Teng, M.; Liu, Y. X.; Chen, N. C.; Wang, L. H.; Carey, A. M.; Deacon, C.; Raab, A.; Meharg, A. A.; Williams, P. N. High percentage inorganic arsenic content of mining impacted and nonimpacted Chinese rice. *Environ. Sci. Technol.* **2008**, *42* (13), 5008−5013.

(4) Yuan, C. G.; Jiang, G. B.; He, B. Evaluation of the extraction methods for arsenic speciation in rice straw, Oryza sativa L., and analysis by HPLC-HG-AFS. *J. Anal. At. Spectrom.* **2005**, *20* (2), 103−110.

(5) Meharg, A. A.; Williams, P. N.; Adomako, E.; Lawgali, Y. Y.; Deacon, C.; Villada, A.; Cambell, R. C. J.; Sun, G.; Zhu, Y. G.; Feldmann, J.; Raab, A.; Zhao, F. J.; Islam, R.; Hossain, S.; Yanai, J. Geographical variation in total and inorganic arsenic content of polished (white) rice. *Environ. Sci. Technol.* **2009**, *43* (5), 1612−1617.

(6) Zhao, F. J.; Harris, E.; Yan, J.; Ma, J. C.; Wu, L. Y.; Liu, W. J.; McGrath, S. P.; Zhou, J. Z.; Zhu, Y. G. Arsenic Methylation in Soils and Its Relationship with Microbial arsM Abundance and Diversity, and As Speciation in Rice. *Environ. Sci. Technol.* **2013**, *47* (13), 7147−7154.

(7) Yan, W.; Dilday, R. H.; Tai, T. H.; Gibbons, J. W.; McNew, R. W.; Rutger, J. N. Differential response of rice germplasm to straighthead induced by arsenic. *Crop Sci.* **2005**, *45* (4), 1223−1228.

(8) Chen, C.; Li, L. Y.; Huang, K.; Zhang, J.; Xie, W. Y.; Lu, Y. H.; Dong, X. Z.; Zhao, F. J. Sulfate-reducing bacteria and methanogens are involved in arsenic methylation and demethylation in paddy soils. *ISME J.* **2019**, *13* (10), 2523−2535.

(9) Norton, G. J.; Pinson, S. R. M.; Alexander, J.; Mckay, S.; Hansen, H.; Duan, G. L.; Islam, M. R.; Islam, S.; Stroud, J. L.; Zhao, F. J.; et al. Variation in grain arsenic assessed in a diverse panel of rice (Oryza sativa) grown in multiple sites. *New Phytol.* **2012**, *193* (3), 650−664.

(10) Qin, J.; Rosen, B. P.; Zhang, Y.; Wang, G. J.; Franke, S.; Rensing, C. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (7), 2075−2080.

(11) Lomax, C.; Liu, W. J.; Wu, L. Y.; Xue, K.; Xiong, J. B.; Zhou, J. Z.; McGrath, S. P.; Meharg, A. A.; Miller, A. J.; Zhao, F. J. Methylated arsenic species in plants originate from soil microorganisms. *New Phytol.* **2012**, *193* (3), 665−672.

(12) Zhu, Y. G.; Xue, X. M.; Kappler, A.; Rosen, B. P.; Meharg, A. A. Linking genes to microbial biogeochemical cycling: Lessons from arsenic. *Environ. Sci. Technol.* **2017**, *51* (13), 7326−7339.

(13) Kujawinski, E. B. The impact of microbial metabolism on marine dissolved organic matter. *Annu. Rev. Mar Sci.* **2011**, *3*, 567−599.

(14) Huang, H.; Jia, Y.; Sun, G. X.; Zhu, Y. G. Arsenic speciation and volatilization from flooded paddy soils amended with different organic matters. *Environ. Sci. Technol.* **2012**, *46* (4), 2163−2168.

(15) Ng, E. L.; Patti, A. F.; Rose, M. T.; Schefe, C. R.; Wilkinson, K.; Smernik, R. J.; Cavagnaro, T. R. Does the chemical nature of soil carbon drive the structure and functioning of soil microbial communities? *Soil Biol. Biochem.* **2014**, *70*, 54−61.

(16) Li, H. Y.; Wang, H.; Wang, H. T.; Xin, P. Y.; Xu, X. H.; Ma, Y.; Liu, W. P.; Teng, C. Y.; Jiang, C. L.; Lou, L. P.; Arnold, W.; Cralle, L.; Zhu, Y. G.; Chu, J. F.; Gilbert, J. A.; Zhang, Z. J. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. *Microbiome* **2018**, *6* (1), 187.

(17) Yan, M. M.; Zeng, X. B.; Wang, J.; Meharg, A. A.; Meharg, C.; Tang, X. J.; Zhang, L. L.; Bai, L. Y.; Zhang, J. Z.; Su, S. M. Dissolved organic matter differentially influences arsenic methylation and volatilization in paddy soils. *J. Hazard. Mater.* **2020**, *388*, 121795.

(18) Li, L. Y.; Zhang, N.; Zhang, Y.; Zhang, Y. H.; Shi, Q.; Sun, Y. F.; Huo, L. J.; Zeng, X. B.; Su, S. M. Selective utilization of organic carbon molecules promotes arsenic methylation by increasing methyltransferase activity in arsM-harboring microbes of paddy soils. *Soil Biol. Biochem.* **2024**, *190*, 109305.

(19) Sparks, D. L.; Page, A. L.; Helmke, P. A.; Loeppert, R. H.; Soltanpour, P. N.; Tabatabai, M. A.; Johnston, C. T.; Sumner, M. E. *Methods of Soil Analysis, Part 3: Chemical Methods*; Wiley: Madison, WI, 1996.

(20) Shi, L. D.; Guo, T.; Lv, P. L.; Niu, Z. F.; Zhou, Y. J.; Tang, X. J.; Zheng, P.; Zhu, L. Z.; Zhu, Y. G.; Kappler, A.; Zhao, H. P. Coupled anaerobic methane oxidation and reductive arsenic mobilization in wetland soils. *Nat. Geosci.* **2020**, *13* (12), 799−805.

(21) Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28* (24), 3211−3217.

(22) Nayfach, S.; Pollard, K. S. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **2015**, *16* (1), 51.

(23) Li, D. H.; Luo, R. B.; Liu, C. M.; Leung, C. M.; Ting, H. F.; Sadakane, K.; Yamashita, H.; Lam, T. W. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **2016**, *102*, 3−11.

(24) Uritskiy, G. V.; DiRuggiero, J.; Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **2018**, *6* (1), 158.

(25) 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.* **2015**, *25* (7), 1043−1055.

(26) Olm, M. R.; Brown, C. T.; Brooks, B.; Banfield, J. F. dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **2017**, *11* (12), 2864−2868.

(27) Hyatt, D.; Chen, G. L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* **2010**, *11*, 119.

(28) Wood, D. E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **2019**, *20* (1), 257.

(29) Lu, J.; Breitwieser, F. P.; Thielen, P.; Salzberg, S. L. Bracken: Estimating species abundance in metagenomics data. *PeerJ. Comput. Sci.* **2017**, *3*, No. e104.
(30) Buchfink, B.; Reuter, K.; Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18* (4), 366−368.
(31) Overbeek, R.; Olson, R.; Pusch, G. D.; Olsen, G. J.; Davis, J. J.; Disz, T.; Edwards, R. A.; Gerdes, S.; Parrello, B.; Shukla, M.; Vonstein, V.; Wattam, A. R.; Xia, F. F.; Stevens, R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **2014**, *42* (D1), D206−D214.
(32) Zhang, S. -Y.; Xiao, X.; Chen, S. -C.; Zhu, Y. -G.; Sun, G. X.; Konstantinidis, K. -T. High Arsenic Levels Increase Activity Rather than Diversity or Abundance of Arsenic Metabolism Genes in Paddy Soils. *Appl. Environ. Microbiol.* **2021**, *87* (20), No. e01383-21.
(33) Li, P. F.; Wu, M.; Li, T.; Dumbrell, A. J.; Saleem, M.; Kuang, L.; Luan, L.; Wang, S.; Li, Z. P.; Jiang, J. D. Molecular weight of dissolved organic matter determines its interactions with microbes and its assembly processes in soils. *Soil Biol. Biochem.* **2023**, *184*, 109117.
(34) Laszakovits, J. R.; MacKay, A. A. Data-based chemical class regions for van krevelen diagrams. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (1), 198−202.
(35) Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg Sci.* **2003**, *14* (6), 927−930.
(36) Wickham, H.; François, R.; Henry, L.; Müller, K.; Vaughan, D. *dplyr: A Grammar of Data Manipulation: R package version 1.1.4.*, The Comprehensive R Archive Network, 2023.
(37) Hijmans, R., *Geosphere: Spherical Trigonometry: R package version R 1.5−18*. The Comprehensive R Archive Network. 2022.
(38) Kassambara, A., *ggcorrplot: Visualization of a Correlation Matrix using'ggplot2': R package version 0.1.4.1*. The Comprehensive R Archive Network, 2023.
(39) Wei, T. Y.; Simko, V., *R package'corrplot': Visualization of a Correlation Matrix: R package version 0.92*. The Comprehensive R Archive Network, 2021.
(40) Huang, H. Y., *linkET: Everything is Linkable: R package version 0.0.7.4*. The Comprehensive R Archive Network, 2021.
(41) Kahle, D.; Wickham, H. ggmap: Spatial Visualization with ggplot2. *R J.* **2013**, *5* (1), 144−161.
(42) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L. D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, T. L.; Miller, E.; Bache, S. M.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, D. P.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4* (43), 1686.
(43) Allaire, J.; Gandrud, C.; Russell, K.; Yetman, C., *networkD3: d3 JavaScript Network Graphs from R package version 0.4*. The Comprehensive R Archive Network, 2017.
(44) Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. Roy Stat Soc. A* **2011**, *174*, 245−245.
(45) Rodriguez-R, L. M.; Gunturu, S.; Tiedje, J. M.; Cole, J. R.; Konstantinidis, K. T. Nonpareil 3: Fast estimation of metagenomic coverage and sequence diversity. *Msystems* **2018**, *3* (3), No. e0003918.
(46) Chen, C.; Shen, Y.; Li, Y. H.; Zhang, W. W.; Zhao, F. J. Demethylation of the Antibiotic Methylarsenite is Coupled to Denitrification in Anoxic Paddy Soil. *Environ. Sci. Technol.* **2021**, *55* (22), 15484−15494.
(47) Chen, C.; Huang, K.; Xie, W. Y.; Chen, S. H.; Tang, Z.; Zhao, F. J. Microbial Processes Mediating the Evolution of Methylarsine Gases from Dimethylarsenate in Paddy Soils. *Environ. Sci. Technol.* **2017**, *51* (22), 13190−13198.
(48) Gao, A. X.; Chen, C.; Gao, Z. Y.; Zhai, Z. Q.; Wang, P.; Zhang, S. Y.; Zhao, F. J. Soil redox status governs within-field spatial variation in microbial arsenic methylation and rice straighthead disease. *ISME J.* **2024**, *18* (1), wrae057.
(49) Chen, C.; Yu, Y.; Wang, Y. J.; Gao, A. X.; Yang, B. Y.; Tang, Z.; Zhao, F. J. Reduction of dimethylarsenate to highly toxic dimethylarsenite in paddy soil and rice plants. *Environ. Sci. Technol.* **2023**, *57* (1), 822−830.
(50) Qiao, J. T.; Liu, J. Y.; Palomo, A.; Bostick, B. C.; Phan, K.; Zheng, Y.; Li, F. B. Prevalence of methylated arsenic and microbial arsenic methylation genes in paddy soils of the Mekong Delta. *Environ. Sci. Technol.* **2023**, *57* (26), 9675−9682.
(51) Liu, J.; Pei, R.; Liu, R. Z.; Jing, C. Y.; Liu, W. J. Arsenic methylation and microbial communities in paddy soils under alternating anoxic and oxic conditions. *J. Environ. Sci.* **2025**, *148*, 468−475.
(52) Wang, Y. H.; Li, P.; Jiang, Z.; Liu, H.; Wei, D. Z.; Wang, H. L.; Wang, Y. X. Diversity and abundance of arsenic methylating microorganisms in high arsenic groundwater from Hetao Plain of Inner Mongolia, China. *Ecotoxicology* **2018**, *27* (8), 1047−1057.
(53) Guo, T.; Li, L. G.; Zhai, W. W.; Xu, B. L.; Yin, X. L.; He, Y.; Xu, J. M.; Zhang, T.; Tang, X. J. Distribution of arsenic and its biotransformation genes in sediments from the East China Sea. *Environ. Pollut.* **2019**, *253*, 949−958.
(54) Li, X. M.; Chen, Q. L.; He, C.; Shi, Q.; Chen, S. C.; Reid, B. J.; Zhu, Y. G.; Sun, G. X. Organic carbon amendments affect the chemodiversity of soil dissolved organic matter and its associations with soil microbial communities. *Environ. Sci. Technol.* **2019**, *53* (1), 50−59.
(55) Chen, S. C.; Sun, G. X.; Yan, Y.; Konstantinidis, K. T.; Zhang, S. Y.; Deng, Y.; Li, X. M.; Cui, H. L.; Musat, F.; Popp, D.; Rosen, B. P.; Zhu, Y. G. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (19), 10414−10421.
(56) Seifert, A. G.; Roth, V. N.; Dittmar, T.; Gleixner, G.; Breuer, L.; Houska, T.; Marxsen, J. Comparing molecular composition of dissolved organic matter in soil and stream water: Influence of land use and chemical characteristics. *Sci. Total Environ.* **2016**, *571*, 142−152.
(57) Zhang, S. Y.; Zhao, F. J.; Sun, G. X.; Su, J. Q.; Yang, X. R.; Li, H.; Zhu, Y. G. Diversity and Abundance of Arsenic Biotransformation Genes in Paddy Soils from Southern China. *Environ. Sci. Technol.* **2015**, *49* (7), 4138−4146.
(58) Bei, Q.; Yang, T. T.; Ren, C. Y.; Guan, E. X.; Dai, Y. C.; Shu, D. T.; He, W. X.; Tian, H. X.; Wei, G. H. Soil pH determines arsenic-related functional gene and bacterial diversity in natural forests on the Taibai Mountain. *Environ. Res.* **2023**, *220*, 115181.
(59) Zhang, S. Y.; Su, J. Q.; Sun, G. X.; Yang, Y. F.; Zhao, Y.; Ding, J. J.; Chen, Y. S.; Shen, Y.; Zhu, G. B.; Rensing, C.; Zhu, Y. G. Land scale biogeography of arsenic biotransformation genes in estuarine wetland. *Environ. Microbiol.* **2017**, *19* (6), 2468−2482.
(60) Das, S.; Chou, M. L.; Jean, J. S.; Liu, C. C.; Yang, H. J. Water management impacts on arsenic behavior and rhizosphere bacterial communities and activities in a rice agro-ecosystem. *Sci. Total Environ.* **2016**, *542*, 642−652.
(61) Li, W. T.; Chen, X. F.; Liu, M.; Kuzyakov, Y.; Jiang, C. Y.; Wu, M.; Li, Z. P. Shifts in microbial communities with increasing soil fertility across a chronosequence of paddy cultivation in subtropical China. *Appl. Soil Ecol.* **2017**, *120*, 153−159.
(62) Bei, Q. C.; Peng, J. J.; Liesack, W. Shedding light on the functional role of the Ignavibacteria in Italian rice field soil: A metagenomic/transcriptomic analysis. *Soil Biol. Biochem.* **2021**, *163*, 108444.
(63) Wang, P. P.; Bao, P.; Sun, G. X. Identification and catalytic residues of the arsenite methyltransferase from a sulfate-reducing bacterium, Clostridium sp. BXM. *FEMS Microbiol. Lett.* **2015**, *362* (1), 1−8.
(64) Zhang, J.; Cao, T. T.; Tang, Z.; Shen, Q. R.; Rosen, B. P.; Zhao, F. J. Arsenic methylation and volatilization by arsenite S-adenosylmethionine methyltransferase in Pseudomonas alcaligenes NBRC14159. *Appl. Environ. Microbiol.* **2015**, *81* (8), 2852−2860.
(65) Viacava, K.; Qiao, J. T.; Janowczyk, A.; Poudel, S.; Jacquemin, N.; Meibom, K. L.; Shrestha, H. K.; Reid, M. C.; Hettich, R. L.; Bernier-Latmani, R. Meta-omics-aided isolation of an elusive

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553

anaerobic arsenic-methylating soil bacterium. *ISME J.* **2022**, *16* (7), 1740−1749.

https://doi.org/10.1021/acs.est.4c10809
Environ. Sci. Technol. 2025, 59, 2541−2553