Meta-omics-aided isolation of an elusive anaerobic arsenic-methylating soil bacterium

Karen Viacava[1,2], Jiangtao Qiao[1], Andrew Janowczyk[3], Suresh Poudel[4], Nicolas Jacquemin[5], Karin Lederballe Meibom[1], Him K. Shrestha[4,6], Matthew C. Reid[7], Robert L. Hettich[4], Rizlan Bernier-Latmani[1*].

[1] Environmental Microbiology Laboratory, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

[2] Soil Science Group, Institute of Geography, University of Bern, Bern, Switzerland.

[3] Bioinformatics Core Facility, Swiss Institute of Bioinformatics, Lausanne, Switzerland.

[4] BioSciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States.

[5] Translational Bioinformatics and Statistics, Department of Oncology, Université de Lausanne, Lausanne, Switzerland.

[6] Genome Science and Technology Graduate School, University of Tennessee, Knoxville, TN, United States.

[7] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, United States.

*Corresponding author. E-mail address rizlan.bernier-latmani@epfl.ch. Postal address EPFL ENAC IIE EML CH A1 375 (Bâtiment CH) Station 6 CH-1015 Lausanne Switzerland. Tel. +41 21 69 35001.

Competing Interests Statement: The authors declare there are no competing financial interests in relation to the work described. The work was funded by the Swiss National Science Foundation (SNSF) grant 310030_176146-1 and the SNSF NCCR Microbiomes (grant # 51NF40_180575). The work at ORNL was conducted under the Plant-Microbe Interface Science Focus Area, as supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program.

Key words: metagenomics, metaproteomics, metatranscriptomics, *Paraclostridium*, arsenic metabolism.

Abstract

Soil microbiomes harbour unparalleled functional and phylogenetic diversity. However, extracting isolates with a targeted function from complex microbiomes is not straightforward, particularly if the associated phenotype does not lend itself to high-throughput screening. Here, we tackle the methylation of arsenic (As) in anoxic soils. As methylation was proposed to be catalysed by sulphate-reducing bacteria. However, to date, there are no available anaerobic isolates capable of As methylation, whether sulphate-reducing or otherwise. The isolation of such a microorganism has been thwarted by the fact that the anaerobic bacteria harbouring a functional arsenite *S*-adenosylmethionine methyltransferase (ArsM) tested to date did not methylate As in pure culture. Additionally, fortuitous As methylation can result from the release of non-specific methyltransferases upon lysis. Thus, we combined metagenomics, metatranscriptomics, and metaproteomics to identify the microorganisms actively methylating As in anoxic soil-derived microbial cultures. Based on the metagenome-assembled genomes of microorganisms expressing ArsM, we isolated *Paraclostridium sp.* strain EML, which was confirmed to actively methylate As anaerobically. This work is an example of the application of meta-omics to the isolation of elusive microorganisms.

Introduction

Soil microbiomes represent a rich source of novel metabolisms and taxa [1–4]. However, isolating microorganisms from them to study specific functions can be challenging, and even more so in cases for which the phenotype is not identifiable with high-throughput methods [5, 6]. An example of challenging microorganisms to isolate are anaerobic As-methylating strains. Arsenic methylation, catalysed by arsenite (As(III)) *S*-adenosylmethionine methyltransferase (ArsM, in prokaryotes), entails the binding of one to three methyl group(s) to the As atom [7]. At present, there are no available microorganisms capable of anaerobic As methylation. This is because, adding to the constraints associated with maintaining an anoxic environment [8], there is no assay for ArsM activity that can be adapted for high-throughput assessment, despite recent endeavours [9]. Arsenic methylation occurs in anoxic, flooded rice paddy soils, is mediated by soil microorganisms [10], and results in the accumulation of methylated As in rice grains [11]. The bioaccumulation of methylated As in rice grains is considerably more efficient than that of inorganic As [12, 13].

The gene encoding ArsM (*arsM*) has been identified in phylogenetically diverse soil microorganisms [14–17]. Anaerobic As methylation is expected to produce a toxic trivalent monomethylated As species (MMAs(III)). The function of this transformation is hypothesized to be microbial warfare, by which the As-methylating organism inhibits microbial competitors via the production of MMAs(III) [18, 19]. If that is confirmed, it is conceivable that As methylation may not occur in pure cultures but only in microbial communities, triggered by metabolites produced by the microbiota. Alternatively, *arsM*-harbouring microorganisms that express As(III) efflux pump(s), the major pathway of As resistance within bacteria [20], may not methylate As due to the efficient removal of As(III) from the cytoplasm, which is the location of ArsM [21, 22]. This effect could be direct, i.e., insufficient substrate concentration, or indirect, i.e., the intracellular As(III) concentration is too low to induce *arsM* expression. Either occurrence (microbial warfare or rapid As(III) efflux) would render the isolation of pure cultures of As-methylating anaerobes very challenging using standard approaches. The

74 latter hypothesis is supported by recent work showing the lack of As methylation by
75 anaerobic pure cultures harbouring functional ArsM enzymes [22].
76 An additional complexity is evidence for the fortuitous methylation of As upon cell lysis and
77 the release of methyltransferases. This fortuitous activity was suggested for the methanogen
78 *Methanosarcina mazei,* for which As methylation was initiated only when cell viability
79 decreased [22], and by the *in vitro* methylation of various metals, including As, by MtaA, a
80 methyltransferase involved in methanogenesis [23]. Thus, As methylation activity in cultures
81 incubated beyond the exponential phase may simply be an experimental artefact [22].
82 Finally, the detection of methylated As requires relatively complex analytical tools (high
83 pressure liquid chromatography coupled to inductively-coupled plasma mass spectrometry,
84 HPLC-ICP-MS) that do not lend themselves readily to high-throughput screening of a large
85 number of colonies [9]. As a result of these challenges, there are no anaerobic
86 microorganisms available known to actively methylate As despite many efforts to identify
87 them. In one instance, researchers had identified a Gram-positive sulphate-reducing
88 bacterium (SRB) [24] that was reported to methylate As but this isolate is no longer
89 available, precluding further investigation.
90 Thus, this study aimed to conclusively identify an active anaerobic As methylator in soil-
91 derived microbial cultures using a multi-omics approach. The experimental strategy was to
92 build Metagenome-Assembled Genomes (MAGs) from metagenomic data and to identify the
93 subset of MAGs harbouring the gene *arsM* that also expressed the *arsM* RNA transcript
94 (metatranscriptomics) and/or the enzyme ArsM (metaproteomics). Based on the genetic
95 information from the target MAG, an isolation strategy was devised that allowed the recovery
96 of a pure culture, later confirmed to be a novel anoxic As-methylating strain.
97
98 **Materials and methods**
99 *Rice paddy soil microbiomes*
100 The soil-derived cultures consisted of two anaerobic microbial enrichments derived from a
101 Vietnamese rice paddy soil and described in Reid *et al*. [25]. The microbiota from the first

soil-derived microbiome was grown in ¼ strength tryptic soy broth (TSB) medium (7.5 g l^{-1} TSB), used previously to enrich As-methylating microbes from a lake sediment [26], and henceforth referred to as the TSB culture. The medium for the second soil-derived microbiome, in addition to ¼ strength TSB, included electron acceptors and two additional carbon sources to simultaneously allow the growth of nitrate-, iron-, and sulphate-reducers, as well as fermenters and methanogens (EA medium: 5 mM $NaNO_3$, 5 mM Na_2SO_4, 5 mM ferric citrate, 0.2 g l^{-1} yeast extract (Oxoid, Hampshire, UK) and 1 g l^{-1} cellobiose, pH 7). This enrichment will be referred to as the EA culture. Both media were boiled, cooled down under 100% N_2 gas and 50 ml of medium were dispensed into 100-ml serum bottles. The bottle headspace was flushed with 100% N_2 gas prior to autoclaving. All culture manipulations were carried out using N_2-flushed syringes and needles. Cultures were grown at 30°C. Growth was quantified using optical density at 600 nm (OD_{600}).

Arsenic methylation assays

Pre-cultures from each enrichment were started from -80°C glycerol stocks. The EA culture started from the glycerol stock was transferred only after a dark precipitate, presumably iron sulphide resulting from sulphate reduction, was formed. The first experimental set-up consisted of bottles containing medium amended with As(III) as $NaAsO_2$ (+As condition) pre-inoculation or unamended (no-As control). For this set-up, cell pellets were sampled for DNA sequencing and proteome characterization during the stationary phase, and for RNA sequencing at the mid-exponential growth phase (see Figures S1, S2 and S3 in Supplementary Information (SI)). In a second experimental set-up, cultures were grown in unamended (no As(III) added) medium and As(III) was added at the mid-exponential growth phase. For this set-up, cell pellets were sampled before (no-As control) and 30 min after As amendment (+As condition) and were used solely for a second transcriptomic analysis. Triplicate biological experiments were performed for each condition (no-As, +As) and per soil-derived enrichment and were used for DNA and RNA sequencing and metaproteome characterization. Sampling for soluble As species, determination of As speciation, and total

As concentration are described in SI.

DNA sequencing and metagenomic analysis

DNA was extracted from the pellet (10 min, 4,500 g) of 4 ml of culture using the DNeasy Power Soil Kit (Qiagen, Hilden, Germany) homogenizing with a Precellys 24 Tissue Homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France) (6,500 rpm for 10 s, repeated 3x with 10 s pause intervals). Metagenomic sequencing was performed by the Genomics Platform of the University of Geneva, Switzerland (iGE3) on a HiSeq 4000 (Illumina, San Diego, CA, US). Libraries were multiplexed and prepared using 100-base reads with paired ends according to the Nextera DNA Flex Library Preparation Kit protocol (Illumina). The quality of sequence reads was assessed with FastQC [27] and duplicated reads eliminated by FastUniq [28]. Reads from all biological replicates within the same experimental condition were assembled into contigs using MegaHit [29]. The contig abundance was determined by aligning the sequencing reads from each biological replicate back to the assembled contigs using Kallisto [30, 31]. The abundance for each gene was considered equivalent to the abundance of the contig in which it was encoded. Gene abundance is reported as 'transcripts per million' (TPM), referred to as TPM-DNA when used for gene abundance. TPM includes normalization for gene length and read sequencing depth [32]. Prodigal was used for the prediction of protein-coding genes [33], generating protein sequence libraries for each culture (EA, TSB) and condition (no-As control, +As condition). The annotation server GhostKOALA [34] was used to assign a KEGG Orthology (KO) database number to each protein-coding gene to identify its encoded function and taxonomic category. The 16S small subunit (SSU) rRNA sequences were identified in the contigs and their taxonomy assigned by Metaxa2 [35]. The relative abundance of the 16S SSU rRNA sequences identified in each of the four metagenomes was quantified using the Kallisto-calculated contig abundance. Contigs with length >2,000 bp were clustered into bins based on composition and coverage using CONCOCT [36], MetaBAT2 [37] and MaxBin 2.0 [38]. The final bin set was obtained by using the Bin_refinement module from MetaWRAP

158 [39]. Completeness, contamination, strain heterogeneity and community (%) in contigs for
159 each bin were calculated using CheckM [40]. Matching bins between the no-As and +As
160 metagenomes, and between the EA +As and TSB +As metagenomes were identified by
161 pairwise comparison of the predicted genomes using dRep [41]. Bins with an average
162 nucleotide identity >95% were considered identical genomes.
163
164 *RNA sequencing and metatranscriptomic analysis*
165 Each culture (5 ml) was harvested at mid-exponential phase for metatranscriptomic analysis.
166 The cells were lysed and the RNA purified using the RNeasy Mini Kit following the
167 manufacturer's instructions (RNAprotect Bacteria, Qiagen). The purified RNA was DNase-I
168 treated (Promega, Madison, WI, US) (1h, 37 °C) and cleaned using the RNeasy Mini Kit a
169 second time. Ribosomal RNA (rRNA) depletion (kit QIAseq FastSelect −5S/16S/23S,
170 Qiagen), library preparation using single-end 100 bases reads (TrueSeq Stranded mRNA,
171 Illumina) and RNA sequencing (on a HiSeq 4000) were performed by the iGE3 Platform.
172 Reads were quality-assessed by FastQC, trimmed by Trimmomatic [42], post-sequencing
173 rRNA-depleted by SortMeRNA [43] and aligned to their corresponding protein sequence
174 library by Bowtie2 [44]. The program featureCounts [45] was employed to count the number
175 of RNA reads aligned to the Prodigal-predicted protein-coding genes. The raw counts were
176 used to calculate the TPM, referred as TPM-RNA when employed for transcript abundance.
177 Finally, to assess RNA expression changes in the +As condition relative to the no-As
178 condition, a differential abundance analysis was performed using the DESeq2 package [46]
179 using the protein sequence libraries from the +As condition to align the RNA reads. A gene
180 was considered to have a significant difference in transcription when the absolute $\log_2$ fold
181 change was ≥1 (i.e., 0.5 ≥fold change ≥2) and the adjusted *q* value ≤0.05.
182
183 *Metaproteome characterization and metaproteomic analysis*
184 The metaproteome analysis was performed at Oak Ridge National Laboratory (Oak Ridge,
185 TN, US). Biomass pellets from 100 ml of culture were washed with 100 mM NH_4HCO_3 buffer

(ABC) (pH 8.0), re-suspended in lysis buffer (4% sodium dodecyl sulphate, 100 mM Tris-HCl, pH 8.0) and disrupted by bead-beating. Lysate proteins were reduced with 5 mM dithiothreitol (30 min, 37 °C), alkylated with 15 mM iodoacetamide (30 min in the dark, room temperature) and isolated by a chloroform-methanol extraction. Extracted proteins were solubilized in 4% sodium deoxycholate (SDC) in ABC and the concentration estimated with a Nanodrop (Thermo Fisher Scientific, Waltham, MA, US). Sequencing-grade trypsin (Promega) at a 1:75 enzyme:protein ratio (w/w) was used to digest the proteins and formic acid (1% final concentration) was used to precipitate the SDC and collect tryptic peptides. Aliquots of 12 µg of peptides were analysed by 2D LC-MS/MS consisting of a Vanquish UHPLC connected to a Q Exactive Plus MS (Thermo Fisher Scientific). Spectral data were collected using MudPIT (multidimensional protein identification technology) as described previously [47, 48]. Peptides were separated in three steps (35, 100, and 500 mM ammonium acetate eluent) with organic gradients after each step. Eluted peptides were measured and sequenced by data-dependent acquisition using previously described parameters [49].

Protein databases were created for the +As experimental condition (EA +As and TSB +As) from the corresponding protein sequence libraries generated by Prodigal. The MS/MS spectra raw files were processed in Proteome Discoverer version 2.4 (Thermo Fisher Scientific) with MS Amanda 2.0 [50] and Percolator [51]. Spectral data were searched against the protein database of the corresponding culture (i.e., EA or TSB). The following parameters were used in the search algorithm MS-Amanda 2.0 to derive tryptic peptides: MS1 tolerance = 5 ppm; MS2 tolerance = 0.02 Da; missed cleavages = 2; carbamidomethyl (C, + 57.021 Da) as static modification; and oxidation (M, + 15.995 Da) as dynamic modifications. The false discovery rate (FDR) threshold was set to 1% for strict FDR and 5% for relaxed FDR at the peptide-spectrum matched (PSM), peptide, and protein levels. FDR-controlled peptides were then quantified according to the chromatographic area-under-the-curve and mapped to their respective proteins. Areas were summed to estimate protein-level abundance.

For differential abundance analysis of proteins, the spectral data from the no-As control, EA no As and TSB no As, were searched against the EA +As and TSB +As protein databases, respectively. All the above-described parameters were maintained. The proteins with at least one peptide detected were exported from Proteome Discoverer. Protein data matrix from EA no As and EA +As were merged and TSB no As and TSB +As were merged. Protein abundance values were $\log_2$ transformed, LOESS-normalized among biological replicates and mean-centred across all conditions using the software InfernoRDN [52]. Stochastic sampling of the proteins was filtered by removing the proteins without abundance value in at least two of the biological triplicates in at least one condition (no-As control or +As condition). Remaining missing data were imputed by random numbers drawn from a normal distribution (width = 0.3 and downshift = 2.8 using the Perseus software http://www.perseus-framework.org) [53]. The differentially abundant proteins were identified by Student t-test method with adjusted *q* value ≤0.05. Proteins were further filtered using the absolute $\log_2$ fold change ≥1.

The isolation of the *Paraclostridium sp.* strain EML is described in SI.

Results

Arsenic methylation by soil-derived microbiomes

The first experimental set-up yielded samples for the metagenome, metaproteome and one of the metatranscriptomes (labeled metatranscriptome G for 'growth in the presence of As') (Figures S1, S2 and S3). The second set-up, assessing the microbiota's short-term response to As(III), provided sample for the second metatranscriptome (labeled metatranscriptome R for 'response to arsenic addition') (Figures S1-A and S2-A). Both EA and TSB cultures exhibited As methylation, reaching an efficiency of As(III) transformation of 27.7% and 19.5%, respectively (Figures S1 and S2).

Microbiota composition

The taxonomic classification of 16S SSU rRNA sequences show that, although eukaryotic

DNA was also identified, the main fraction of the communities was bacterial (>89.0 ±0.8% for EA cultures and >98.5 ±0.3% for TSB cultures, relative abundance) and was distributed amongst eight operational taxonomic units (OTUs) at the order level (Figure 1 and Tables S1-S4). Statistically significant changes (unpaired Student t-test and no significant difference considered when *p* value > 0.05) in the OTUs relative abundances, +As condition versus no-As control, are described in SI and summarized in Tables S5 and S6.

MAG selection

The contigs from the four metagenomes, EA (+As, no-As control) and TSB (+As, no-As control), were clustered separately into bins. High-quality (≥90% completeness and ≤5% contamination) bins were designated as MAGs [54]. For the +As condition, the parsing process led to a total of 36 MAGs (Table 1). Additionally, matching bins were sought in the bins from the no-As control cultures (Tables S7 and S8). Only one of the 36 MAGs in the +As condition was left unpaired (TSB MAG 8).

For each MAG, a lineage was assigned by CheckM, based on lineage-specific marker genes [40]. The MAGs identified belonged to the phyla: *Firmicutes* (orders *Clostridiales*, *Selenomonadales* and *Lactobacillales*, and the genus *Clostridium*), *Proteobacteria* (*Enterobacteriaceae* family and *Deltaproteobacteria* class) and *Bacteroidetes* (order *Bacteroidales*). Fifteen MAGs presented non-zero strain heterogeneity (Table 1), an index of the phylogenetic relatedness of binned contigs based on the amino acid identity of the encoded proteins. For ten MAGs, the value is ≥50%, suggesting some phylogenetic relation with the contaminating strains. Five MAGs had heterogeneity values ≤33.33%, suggesting contamination with microorganisms that are not closely related. In the remaining 21 MAGs, the strain heterogeneity is 0%, i.e., no strain heterogeneity or no contamination (Tables S7 and S8).

Changes in the relative abundance of MAGs (no-As control vs. +As condition), relatedness of the +As EA and TSB microbial communities, along with the presence, transcription and translation of genes encoding key enzymes from major metabolic pathways of each MAG in

the +As condition are included in SI.

Arsenic resistance genes

The metagenomic libraries from the +As condition of the EA and TSB cultures were mined for arsenic resistance (*ars*) genes and their encoded proteins (the pipeline is described in SI). A total of 309 and 282 genes were annotated as *ars* genes in the EA and TSB +As metagenomic libraries, respectively (Tables S9 and S10). Of those, 255 and 226 were considered correctly annotated as *ars* genes based on BLAST and HMMER (refer to SI for pipeline), and 225 and 147 had above-threshold DNA abundances, respectively (Figure 2) (refer to SI for abundance threshold values). Individual abundance values of *ars* genes, transcripts and proteins in the +As condition and the no-As control and their transcript and protein relative abundance values in the +As condition vs. the no-As control for each MAG group from the EA and TSB cultures are available in Tables S11 and S12, respectively.

The *ars* genes encode proteins involved in the detoxification of As oxyanions: *arsB* and *acr3,* encoding As(III)-efflux systems; *arsA,* encoding the ATPase energizing the efflux of As(III) and As(III) chaperone; *arsD*, encoding a weak *ars* operon repressor [55]; *arsC1* and *arsC2,* encoding As(V) reductases coupling As reduction to the oxidation of glutaredoxin or thioredoxin, respectively; and *arsR* genes encoding As(III)-regulated repressors (ArsR1, ArsR2, and ArsR3) classified based on the location of the As(III)-binding cysteine residues [56–58].

The most common *ars* genes in EA and TSB culture metagenomes were *arsR*, *arsC,* and *arsP* (Figure 2). The first two genes are part of the canonical *ars* operon *arsRBC* [59], whilst *arsP*, encoding a recently discovered membrane transporter, has been found to be widely distributed in bacterial genomes [20]. Most of the surveyed *arsP* genes, 57% in EA and 50% in TSB, are encoded in putative *ars* operons, represented by *ars* genes contiguously encoded in the same contig (Tables S11 and S12), supporting their As-related function and correct annotation. The next most abundant genes were those responsible for As(III) efflux (*arsB*, *acr3,* and *arsA*), typically found in organisms living in reducing environments in

association with *arsC* [16, 60]. Finally, *arsM* and the two genes, *arsI* and *arsH*, encoding MMAs(III)-resistance mechanisms, were the least recurrent genes in the metagenomes. The results of gene and protein relative expression vs. the no-As control of the *ars* genes involved in the metabolism of inorganic As in the MAGs are described in SI.

Arsenic-methylating MAGs

The *arsM* gene can be expressed at similar, or slightly different levels in the absence or presence of As(III) in some organisms [61, 62], but expressed at significantly higher levels in the presence of As(III) in others [63–66]. Thus, we sought to identify *arsM* genes transcribed and ArsM proteins showing increased expression in the +As condition relative to the no-As control (Figure 3) but also those simply exhibiting expression, not necessarily increased relative to the control (Figure S4).

Sixteen phylogenetically distinct *arsM* genes were identified in the EA +As metagenome, but increased transcriptome reads or peptides (relative to the no-As control) were only detected for three genes (Figure 3). The first is an *arsM* in *Clostridiales* EA MAG 8 classified by GhostKOALA as belonging to *Paeniclostridium sordellii* (EA MAG 8, *arsM*-1, psor type strain, in Table S11). The second was found in *Clostridium* EA MAG 9, also detected in the metaproteome, and the GhostKOALA taxonomic classification of the corresponding gene (EA MAG 9, *arsM-1* in Table S11) revealed that it was attributed to the unclassified species *Ruminococcaceae bacterium* CPB6 (Figure 3, Table S11) [67]. Finally, the third *arsM* was obtained from transcriptomic data but not clustered in any EA MAG (EA unbinned, *arsM-5* in Table S11) and likewise classified as pertaining to *Paeniclostridium sordellii*.

In the TSB +As metagenome, nine distinct *arsM* genes were identified but none were detected in the metatranscriptome and only one exhibited increased expression in the metaproteome (Figure 3). It corresponds to an *arsM* gene from MAG 9 (TSB MAG 9, *arsM-2* in Table S12). The expressed ArsM protein was assigned by GhostKOALA to a *Clostridiales* strain: *Clostridium botulinum* (*cby* type strain) (TSB MAG 9, *arsM-2*) (Figure 3, Table S12). Finally, there was one *arsM* expressed in the TSB +As metaproteome but with no increased

326 expression relative to the no-As control, it was classified as *Ruminococcaceae bacterium*
327 CPB6 (TSB MAG 9, *arsM-1*) (Figure S4), the same organism identified in the EA culture (EA
328 MAG 9, *arsM-1*).
329 In addition to evidence for active As methylation, there was evidence for active detoxification
330 of methylated arsenic. Indeed, the metagenome included genes encoding proteins involved
331 in the metabolism of methylated As such as *arsH*, *arsI, arsP,* and *arsR4* (Figures 2 and 3).
332 These genes encode proteins involved in the detoxification of methylated arsenic like
333 MMAs(III) and roxarsone: the oxidase ArsH, responsible for the oxidation of trivalent
334 methylated As to the less toxic pentavalent form [68]; the demethylase ArsI that removes
335 methyl groups from the As atom [69]; and the transmembrane transporter ArsP, thought to
336 efflux methylated As [70]. The *arsR4* gene encodes an atypical MMAs(III)-responsive ArsR
337 repressor, containing only two conserved cysteine residues [71]. The *Enterobacteriaceae*
338 TSB MAG 14 exhibited activity of the oxygen-dependent ArsH protein [68] (Figure 3). An
339 *arsR4*, shown to induce expression of *arsP* in the presence of MMAs(III) [71], had increased
340 transcription along with an *arsP* encoded in the same contig in the *Selenomonadales* TSB
341 MAG 19 (Figure 3, Table S12). Both gene transcripts were <5 TPM-RNA (Table S12) and
342 thus, were not considered as transcribed in Figure S4. Finally, an ArsI protein, taxonomically
343 related to class *Clostridia* ([*Eubacterium*] *rectale*), was expressed but encoded in an
344 unbinned gene from the EA culture (Figure 3, Table S12).
345
346 *Isolation of an arsenic-methylating anaerobic microorganism*
347 Based on the analysis of the active metabolic activity from the EA MAG 8, expressing an
348 ArsM (Figure S5), an appropriate selective medium was identified for its isolation. We
349 utilized the fact that this MAG harbours and expresses the anaerobic assimilatory sulphite
350 reductase encoded by the *asrABC* operon which is responsible for the NADH-dependent
351 reduction of sulphite to sulphide [72–74] in sulphite-reducing *Clostridia* (SRC). From the nine
352 Clostridia MAGs, only two expressed this capability in the EA microbiome (Figure S5). Thus,
353 the isolation relied on growing the EA culture on agar medium selective for the SRC

phenotype. In TSC agar, designed for the enumeration of *Clostridium perfringens* in food [75], the colonies from SRC are black, as the ammonium ferric citrate forms iron sulphide during sulphite reduction. Additionally, D-cycloserine acts a selective agent for the isolation of *Clostridia* strains [76] while inhibiting facultative anaerobes [75]. Finally, the bromocresol purple contained in the agar allows the identification of sucrose fermenters, resulting in a change of colour from purple to yellow. As none of the genes involved in sucrose transport or hydrolysis were binned in EA *Clostridiales* MAG 8 (Figure S5), only non-sucrose fermenting black colonies were considered. Those colonies were selected and using a colony PCR screen specifically targeting the *arsM* gene of EA MAG 8, we isolated a *Clostridiales* strain encoding the gene of the expressed ArsM in the EA MAG 8 (protein id k119_30669_28, Table S11) (Figure S6).

The isolate consists of non-sucrose-fermenting, rod-shaped and spore-forming bacteria forming convex and circular black colonies on TSC agar (Figures S7 and S8). The BLAST (NCBI) search of the 16S rRNA sequence gives >99% identity to *Paraclostridium* strains (Table S13). On the basis of the 16S rRNA sequence, we assign the following name to the bacterium: "*Paraclostridium* species str. EML". Strain EML was tested for As methylation under anaerobic conditions with 25 µM As(III). The growth of strain EML was hindered by As(III) (Figure 4-A) and starting from ~4 hours, the isolate transformed As(III) to monomethylated soluble As representing 48.3±1.5% of the soluble arsenic in the culture after 83 h (panels B and C from Figure 4). A fraction (14.7±0.6 µM) of the arsenic was found associated with biomass almost exclusively as inorganic As (Figure 4-D).

Discussion

Our results demonstrate the successful translation of multi-omic information to a specific strategy for targeted microbial isolation. The metagenomes from the anaerobic soil-derived cultures identified the potential for As methylation in microorganisms from diverse taxa. While there were a large number of *ars* genes in the metagenomes, only a small proportion

was transcribed or translated in the presence of As when compared to the no-As control (Figure 2). This contrast was particularly evident for the gene responsible for As methylation, *arsM*. The post-genomic approaches of community gene and protein expression in TSB as in EA soil-derived microbiomes clearly pointed to the active As-methylating role of various fermenting bacteria from the order *Clostridiales*. This information paved the way for the identification of As-methylating microorganisms and the successful isolation of an anaerobic As methylator.

The TSB and EA media were chosen to selectively enrich for putative As methylators from the microbial soil community based on the study from Bright *et al.*, in which lake sediments enriched in TSB medium, either sulphate-amended or unamended, were shown to have greater As methylation rates than in iron- or manganese-reducing TSB cultures [26]. The selected media caused a great shift in the original soil microbial diversity [25] along with the loss of putative As-methylating microorganisms. Nonetheless, the As-methylating TSB and EA soil-derived cultures offered the opportunity to study active As methylation from paddy-soil microbiota in an environment that is less complex than soil but that remains environmentally relevant. In contrast to soil slurries, the absence of soil minerals in the soil-derived cultures facilitated the detection of soluble methylarsenicals and the extraction of DNA, RNA and proteins. The multi-omic approach made it possible to identify putative microorganisms driving As methylation and their metabolism. Targeting a specific *arsM* gene rather than the synthesis of methylarsenicals greatly accelerated colony screening, as colony PCR could be employed instead of analytical detection by HPLC-ICP-MS.

Had only the metagenomic approach been implemented, the data would have pointed to SRB MAGs as putative As methylators, as they harboured the most abundant *arsM* genes (Figure 5). Indeed, SRB have been proposed as drivers of As methylation in rice paddy soils based on the correlation in the abundance of *arsM* and dissimilatory sulphite reductase (*dsr*) genes [77] and RNA transcripts [78], and a decrease in As methylation by the addition chemical inhibitors of dissimilatory sulphate reduction (DSR) [77, 78]. Additionally, the use of degenerate primers for *arsM* amplification may underestimate *arsM* phylogenetic diversity, a

409 drawback overcome by metagenomic and metatranscriptomic sequencing. In the present
410 findings, the SRB *Deltaproteobacteria* MAGs, although actively reducing sulphate (Figures
411 S5 and S9), did not exhibit As-methylating activity as their *arsM* genes were neither
412 transcribed nor translated (Figure 5). *Desulfovibrio* MAGs were metabolically active in both
413 cultures, but amongst all their encoded *ars* genes, only an *arsR3* exhibited increased
414 expression in the presence of As(III), providing strong evidence for their lack of involvement
415 in As methylation in the TSB and EA cultures.
416 Previous work had identified another As-methylating *Clostridiales* strain, *Clostridium* sp.
417 BXM [24], that performed fermentation and DSR but that is no longer available. The sole
418 attribution of As-methylating activity to fermenting Firmicutes in that work, along with the
419 isolation of the present sulphite-reducing fermenter, point to a key role for fermenting
420 *Clostridiales* microorganisms harbouring sulphur-related metabolism in As methylation.
421 Other studies have reported an increase in As methylation efficiency after the amendment of
422 sulphate [79] or organic matter to soil [15, 80–82], or after the increase in dissolved organic
423 carbon in soil [83]. The positive impact of sulphate amendment on As methylation was
424 interpreted as pointing to the role of SRB in As methylation [79]. Here, we offer an
425 alternative explanation, supported by examples of organic amendments enhancing As
426 methylation [15, 80–82]. The sulphate amendment could have indirectly increased the
427 availability of short-chain fatty acids through DSR, providing fermentable substrates. Thus,
428 we propose that direct or indirect organic amendments result in the enrichment of fermenting
429 communities, and consequently, in an increase in As methylation.
430 It was previously proposed that the As-methylating activity of anaerobic microorganisms may
431 be limited by efficient efflux of intracellular As(III) [22], or that it may function as a defensive
432 response against nutrient competition [18]. Indeed, the identification of MAGs exhibiting a
433 detoxification response to methylarsenicals supports the hypothesis of the role of
434 monomethylated As as an arsenic-bearing antibiotic. Although the expression of ArsI and
435 ArsH, catalysing oxygen-dependent MMAs(III)-resistance mechanisms (Figure 3), is difficult
436 to reconcile with anoxic conditions, it is conceivable that these proteins are capable of

437 additional functions in the absence of O_2. Up until now, the lack of available anaerobic
438 microbial isolates able to methylate As *in vitro* [22] precluded the investigation of the
439 hypotheses raised above. This work represents the first study applying a combination of
440 three meta-omic techniques in order to characterize As metabolism in microbial communities
441 and to perform meta-omics-aided isolation of a microorganism [84][85]. The successful
442 isolation of *Paraclostridium* sp. EML is part of the "new era of omics information-guided
443 microbial cultivation technology" described by Gutleben *et al*. [84] and represents a
444 milestone to obtain novel targeted microbial isolates from the environment and to elucidate
445 the controls on anaerobic As methylation.
446 Further work is needed to elucidate why ArsM expression was restricted to members of
447 *Clostridiales* fermenters and did not occur in other organisms harbouring *arsM* genes. The
448 availability of As-methylating anaerobes will allow investigation of why the *arsM* gene
449 evolved under an anoxic atmosphere [86], of the controls on the production of toxic
450 methylated As species in flooded rice paddies, and the development of microbially-mediated
451 remediation technologies for As-contaminated soils via the synthesis of volatile
452 methylarsenicals [87, 88].
453

454 **Acknowledgements**

455 The work was funded by the Swiss National Science Foundation (SNSF) grant
456 310030_176146-1 and the SNSF NCCR Microbiomes (grant # 51NF40_180575). The work
457 at ORNL was conducted under the Plant-Microbe Interface Science Focus Area, as
458 supported by the U.S. Department of Energy, Office of Science, Office of Biological and
459 Environmental Research, Genomic Science Program.
460

461 **Data availability**

462 Metagenomic and metatranscriptomic raw sequencing reads are available at the National
463 Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA), BioProject
464 PRJNA714492 (post publication). Data from the meta-omic analyses and source data from

figures are available in Zenodo data repository (10.5281/zenodo.4605527, post publication).

Conflict of interest

The authors declare no conflict of interest.

Figures Legends

Figure 1. Operational taxonomic units (OTUs) at order and genus level (with > 1% relative abundance at genus level) identified from 16S SSU rRNA sequences from soil-derived cultures. Abbreviations: EA no As: EA culture no-As control EA +As: EA culture +As condition, TSB no As: TSB culture no-As control TSB +As: TSB culture +As condition. OTUs at the order level are indicated in bold in the legend. Plotted values are the average relative abundance and together with SD values and Student t test results are available in Tables S5 and S6.

Figure 2. Number of *ars* genes, encoded in the +As condition cultures and with increased expression in metatranscriptomes/metaproteome relative to no-As controls. Number of *ars* genes encoded in metagenomes and with increased expression in metatranscriptomes, R or G, or metaproteomes and the non-redundant overlap between genes with increased expression in metatranscriptomes and/or metaproteomes from +As condition EA (left panels) and TSB (right panels) cultures. Bar length and numbers on the right side of the bars correspond to the number of genes per *ars* gene category. Bold numbers on the lower left corner of each panel correspond to the sum of all *ars* genes per category.

Figure 3. Distribution of *ars* genes involved in methylated arsenic metabolism encoded in MAGs from the +As condition and differentially expressed in metatranscriptomes/metaproteome relative to the no-As EA control. Each numbered box represents an *ars* gene. The number in each box corresponds to the “Numbering”

493 column in Tables S11 and S12 where individual gene abundance and fold change values
494 can be found. Com. (%): community (%) as defined in caption from Table 1.
495
496 **Figure 4. Isolate "*Paraclostridium* sp. EML":** (A) growth as OD_{600} with 25 µM As(III) and
497 without, (B) proportion of soluble arsenic species in filtered medium containing 25 µM As(III),
498 (C) concentration of arsenic species soluble in filtered medium containing 25 µM As(III)
499 (solid lines) and biomass-bound (dashed lines) and (D) proportion of biomass-bound arsenic
500 species. Data points and bars represent the mean value and error bars, plus and minus one
501 standard deviation. Individual values for each measurement and biological replicate are
502 available in Tables S23 and S24.
503
504 **Figure 5. Gene abundance of *arsM* genes in MAGs from the +As condition cultures.**
505 Coloured bars correspond to *arsM* genes with increased expression in the metaproteome
506 (blue-green) or in the metatranscriptome G (purple) from +As condition relative to the no-As
507 control in EA (left panel) and TSB (right panel) cultures. The taxonomic classification shown
508 on the right side of the error bars for selected *arsM* genes corresponds to the individual gene
509 classification assigned by GhostKOALA - "Genus" column in Tables S11 and S12. Columns
510 with matching symbols on the right side of the error bars, correspond to matching *arsM*
511 genes between the cultures. Individual gene abundance can be found in Tables S11 and
512 S12. Numbers inside bar and bar length represent mean and error bars one standard
513 deviation.
514
515 **Table Legend**
516 **Table 1. Metagenome assembled genomes (MAGs) from EA (upper Table A) and TSB**
517 **(lower Table B) cultures in the +As condition.** Marker lineage: taxonomic rank set by
518 CheckM. Completeness and contamination (%): estimated completeness and contamination
519 of genome as determined by CheckM from the presence/absence of single-copy marker
520 genes and the expected colocalization of these genes. Strain heterogeneity: index between

0 and 100 where a value of 0 means no strain heterogeneity, high values suggest the majority of reported contamination is from closely related organisms (i.e., potentially the same species) and low values suggest the majority of contamination is from phylogenetically diverse sources. Proportion of binned proteins assigned to MAG (%): number of protein-coding genes assigned to the MAG divided by the total number of protein-coding genes binned. Community (%): sum of the number of reads mapped to the contigs in each MAG divided by the total number of reads mapped to all contigs including the unbinned contigs, and normalized to MAG size, assuming an average genome size for all unbinned populations.

References

1. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. *Proc Natl Acad Sci* 1998; **95**: 6578–6583.

2. Bardgett RD, Van Der Putten WH. Belowground biodiversity and ecosystem functioning. *Nature* 2014; **515**: 505–511.

3. Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, *et al*. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 2007; **1**: 283–290.

4. Daniel R. The soil metagenome - A rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 2004; **15**: 199–204.

5. Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, *et al*. Challenges of unculturable bacteria: environmental perspectives. *Rev Environ Sci Biotechnol* 2020; **19**: 1–22.

6. Houle D, Govindaraju DR, Omholt S. Phenomics: The next challenge. *Nat Rev Genet* 2010; **11**: 855–866.

7. Ajees AA, Marapakala K, Packianathan C, Sankaran B, Rosen BP. Structure of an As(III) *S*-adenosylmethionine methyltransferase: Insights into the mechanism of arsenic biotransformation. *Biochemistry* 2012; **51**: 5476–5485.

8. Mori K, Kamagata Y. The challenges of studying the anaerobic microbial world. *Microbes Environ* 2014; **29**: 335–337.

9. Dong H, Xu W, Pillai JK, Packianathan C, Rosen BP. High-throughput screening-compatible assays of As(III) *S*-adenosylmethionine methyltransferase activity. *Anal Biochem* 2015; **480**: 67–73.

10. Lomax C, Liu WJ, Wu L, Xue K, Xiong J, Zhou J, et al. Methylated arsenic species in plants originate from soil microorganisms. *New Phytol* 2012; **193**: 665–672.

11. Zhao FJ, Zhu YG, Meharg AA. Methylated arsenic species in rice: Geographical variation, origin, and uptake mechanisms. *Environ Sci Technol* 2013; **47**: 3957–3966.

12. Zheng MZ, Li G, Sun GX, Shim H, Cai C. Differential toxicity and accumulation of

inorganic and methylated arsenic in rice. *Plant Soil* 2013; **365**: 227–238.

13. Abedin MJ, Feldmann J, Meharg AA. Uptake kinetics of arsenic species in rice plants. *Plant Physiol* 2002; **128**: 1120–1128.

14. Dunivin TK, Yeh SY, Shade A. A global survey of arsenic-related genes in soil microbiomes. *BMC Biol* 2019; **17**: 45.

15. Jia Y, Huang H, Zhong M, Wang F-H, Zhang L-M, Zhu Y-G. Microbial arsenic methylation in soil and rice rhizosphere. *Environ Sci Technol* 2013; **47**: 3141–3148.

16. Xiao KQ, Li LG, Ma LP, Zhang SY, Bao P, Zhang T, *et al*. Metagenomic analysis revealed highly diverse microbial arsenic metabolism genes in paddy soils with low-arsenic contents. *Environ Pollut* 2016; **211**: 1–8.

17. Zhang SY, Su JQ, Sun GX, Yang Y, Zhao Y, Ding J, *et al*. Land scale biogeography of arsenic biotransformation genes in estuarine wetland. *Environ Microbiol* 2017; **19**: 2468–2482.

18. Chen J, Rosen BP. The arsenic methylation cycle: How microbial communities adapted methylarsenicals for use as weapons in the continuing war for dominance. *Front Environ Sci* 2020; **8**: 43.

19. Ping Li Y, Ben Fekih I, Chi Fru E, Moraleda-Munoz A, Li X, Rosen BP, *et al*. Antimicrobial activity of metals and metalloids. *Annu Rev Microbiol* 2021; **75**: 175–197.

20. Yang Y, Wu S, Lilley RM, Zhang R. The diversity of membrane transporters encoded in bacterial arsenic-resistance operons. *PeerJ* 2015; **3**: e943.

21. Yang P, Ke C, Zhao C, Kuang Q, Liu B, Xue X, *et al*. ArsM-mediated arsenite volatilization is limited by efflux catalyzed by As efflux transporters. *Chemosphere* 2020; **239**: 124822.

22. Viacava K, Meibom KL, Ortega D, Dyer S, Gelb A, Falquet L, *et al*. Variability in arsenic methylation efficiency across aerobic and anaerobic microorganisms. *Environ Sci Technol* 2020; **54**: 14343–14351.

23. Thomas F, Diaz-Bone RA, Wuerfel O, Huber B, Weidenbach K, Schmitz RA, *et al*.

588 Connection between multimetal(loid) methylation In methanoarchaea and central
589 intermediates of methanogenesis. *Appl Environ Microbiol* 2011; **77**: 8669–8675.
590 24. Wang PP, Bao P, Sun GX. Identification and catalytic residues of the arsenite
591 methyltransferase from a sulfate-reducing bacterium, *Clostridium sp*. BXM. *FEMS*
592 *Microbiol Lett* 2015; **362**: 1–8.
593 25. Reid MC, Maillard J, Bagnoud A, Falquet L, Le Vo P, Bernier-Latmani R. Arsenic
594 methylation dynamics in a rice paddy soil anaerobic enrichment culture. *Environ Sci*
595 *Technol* 2017; **51**: 10546–10554.
596 26. Bright DA, Brock S, Reimer KJ, Cullen WR, Hewitt GM, Jafaar J. Methylation of
597 arsenic by anaerobic microbial consortia isolated from lake sediment. *Appl*
598 *Organomet Chem* 1994; **8**: 415–422.
599 27. Andrews S. FastQC: A quality control tool for high throughput sequence data [Online].
600 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. .
601 28. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, *et al*. FastUniq: A fast *de novo*
602 duplicates removal tool for paired short reads. *PLoS One* 2012; **7**: e52249.
603 29. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node
604 solution for large and complex metagenomics assembly via succinct de Bruijn graph.
605 *Bioinformatics* 2015; **31**: 1674–1676.
606 30. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq
607 quantification. *Nat Biotechnol* 2016; **34**: 525–527.
608 31. Schaeffer L, Pimentel H, Bray N, Melsted P, Pachter L. Pseudoalignment for
609 metagenomic read assignment. *Bioinformatics* 2017; **33**: 2082–2088.
610 32. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq
611 data: RPKM measure is inconsistent among samples. *Theory Biosci* 2012; **131**: 281–
612 285.
613 33. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal:
614 Prokaryotic gene recognition and translation initiation site identification. *BMC*
615 *Bioinformatics* 2010; **11**: 119.

34. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016; **428**: 726–731.

35. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, *et al*. Metaxa2: Improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour* 2015; **15**: 1403–1414.

36. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, *et al*. Binning metagenomic contigs by coverage and composition. *Nat Methods* 2014; **11**: 1144–1146.

37. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, *et al*. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; e7359.

38. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016; **32**: 605–607.

39. Uritskiy G V., Diruggiero J, Taylor J. MetaWRAP - A flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 2018; **6**: 158.

40. 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**: 1043–1055.

41. 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**: 2864–2868.

42. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

43. Kopylova E, Noé L, Touzet H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 2012; **28**: 3211–3217.

44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*

2012; **9**: 357–359.

45. Liao Y, Smyth GK, Shi W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**: 923–930.

46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; **15**: 550.

47. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001; **19**: 242–247.

48. McDonald WH, Ohi R, Miyamoto DT, Mitchison TJ, Yates JR. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: Single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom* 2002; **219**: 245–251.

49. Clarkson SM, Giannone RJ, Kridelbaugh DM, Elkins JG, Guss AM, Michenera JK. Construction and optimization of a heterologous pathway for protocatechuate catabolism in Escherichia coli enables bioconversion of model aromatic compounds. *Appl Environ Microbiol* 2017; **83**.

50. Dorfer V, Pichler P, Stranzl T, Stadlmann J, Taus T, Winkler S, *et al*. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res* 2014; **13**: 3679–3684.

51. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 2007; **4**: 923–925.

52. Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, *et al*. DAnTE: A statistical tool for quantitative analysis of -omics data. *Bioinformatics* 2008; **24**: 1556–1558.

53. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, *et al*. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 2016; **13**: 731–740.

54. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK,

et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.

55. Lin Y-F, Walmsley AR, Rosen BP. An arsenic metallochaperone for an arsenic detoxification pump. *PNAS* 2006; **103**: 15617–15622.

56. Shi W, Wu J, Rosen BP. Identification of a putative metal binding site in a new family of metalloregulatory proteins. *J Biol Chem* 1994; **269**: 19826–19829.

57. Murphy JN, Saltikov CW. The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella sp.* strain ANA-3. *J Bacteriol* 2009; **191**: 6722–6731.

58. Santha S, Pandaranayaka EPJ, Rosen BP, Thiyagarajan S. Purification, crystallization and preliminary X-ray diffraction studies of the arsenic repressor ArsR from *Corynebacterium glutamicum*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2011; **67**: 1616–1618.

59. Fekih I Ben, Zhang C, Li YP, Zhao Y, Alwathnani HA, Saquib Q, *et al*. Distribution of arsenic resistance genes in prokaryotes. *Front Microbiol* 2018; **9**: 2473.

60. Cai L, Yu K, Yang Y, Chen BW, Li XD, Zhang T. Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Appl Microbiol Biotechnol* 2013; **97**: 9579–9588.

61. Huang K, Xu Y, Packianathan C, Gao F, Chen C, Zhang J, *et al*. Arsenic methylation by a novel ArsM As(III) S-adenosylmethionine methyltransferase that requires only two conserved cysteine residues. *Mol Microbiol* 2018; **107**: 265–276.

62. Zhang J, Cao T, Tang Z, Shen Q, Rosen BP, Zhao FJ. Arsenic methylation and volatilization by arsenite *S*-adenosylmethionine methyltransferase in *Pseudomonas alcaligenes* NBRC14159. *Appl Environ Microbiol* 2015; **81**: 2852–2860.

63. Huang K, Chen C, Zhang J, Tang Z, Shen Q, Rosen BP, *et al*. Efficient arsenic methylation and volatilization mediated by a novel bacterium from an arsenic-contaminated paddy soil. *Environ Sci Technol* 2016; **50**: 6389–6396.

64. Yin XX, Chen J, Qin J, Sun GX, Rosen BP, Zhu YG. Biotransformation and volatilization of arsenic by three photosynthetic cyanobacteria. *Plant Physiol* 2011; **156**: 1631–1638.

65. Zhao C, Zhang Y, Chan Z, Chen S, Yang S. Insights into arsenic multi-operons expression and resistance mechanisms in *Rhodopseudomonas palustris* CGA009. *Front Microbiol* 2015; **6**: 986.

66. Wang G, Kennedy SP, Fasiludeen S, Rensing C, DasSarma S. Arsenic resistance in Halobacterium sp. strain NRC-1 examined by using an improved gene knockout system. *J Bacteriol* 2004; **186**: 3187–3194.

67. Tao Y, Zhu X, Wang H, Wang Y, Li X, Jin H, *et al*. Complete genome sequence of *Ruminococcaceae bacterium* CPB6: A newly isolated culture for efficient n-caproic acid production from lactate. *J Biotechnol* 2017; **259**: 91–94.

68. Chen J, Bhattacharjee H, Rosen BP. ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. *Mol Microbiol* 2015; **96**: 1042–1052.

69. Yoshinaga M, Rosen BP. A C-As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. *PNAS* 2014; **111**: 7701–7706.

70. Chen J, Madegowda M, Bhattacharjee H, Rosen BP. ArsP: A methylarsenite efflux permease. *Mol Microbiol* 2015; **98**: 625–635.

71. Chen J, Nadar VS, Rosen BP. A novel MAs(III)-selective ArsR transcriptional repressor. *Mol Microbiol* 2017; **106**: 469–478.

72. Doyle CJ, O'Toole PW, Cotter PD. Genomic characterization of sulphite reducing bacteria isolated from the dairy production chain. *Front Microbiol* 2018; **9**: 1507.

73. Czyzewski BK, Wang DN. Identification and characterization of a bacterial hydrosulphide ion channel. *Nature* 2012; **483**: 494–497.

74. Huang CJ, Barrett EL. Sequence analysis and expression of the *Salmonella typhimurium asr* operon encoding production of hydrogen sulfide from sulfite. *J*

Bacteriol 1991; **173**: 1544–1553.

75. Harmon SM, Kautter DA, Peeler JT. Improved medium for enumeration of *Clostridium perfringens*. *Appl Microbiol* 1971; **22**: 688–692.

76. George WL, Sutter VL, Finegold SM. Toxigenicity and antimicrobial susceptibility of *Clostridium difficile*, a cause of antimicrobial agent-associated colitis. *Curr Microbiol* 1978; **1**: 55–58.

77. Wang M, Tang Z, Chen XP, Wang X, Zhou WX, Tang Z, *et al*. Water management impacts the soil microbial communities and total arsenic and methylated arsenicals in rice grains. *Environ Pollut* 2019; **247**: 736–744.

78. Chen C, Li L, Huang K, Zhang J, Xie WY, Lu Y, *et al*. Sulfate-reducing bacteria and methanogens are involved in arsenic methylation and demethylation in paddy soils. *ISME J* 2019; **13**: 2523–2535.

79. Zeng X, Jiang Y, Fan X, Chao S, Yang Y, Liu J, *et al*. Effects of sulfate application on inhibiting accumulation and alleviating toxicity of arsenic in Panax notoginseng grown in arsenic-polluted soil. *Water Air Soil Pollut* 2016; **227**.

80. Mestrot A, Feldmann J, Krupp EM, Hossain MS, Roman-Ross G, Meharg AA. Field fluxes and speciation of arsines emanating from soils. *Environ Sci Technol* 2011; **45**: 1798–1804.

81. 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**: 2163–8.

82. Seyfferth AL, Morris AH, Gill R, Kearns KA, Mann JN, Paukett M, *et al*. Soil incorporation of silica-rich rice husk decreases inorganic arsenic in rice grain. *J Agric Food Chem* 2016; **64**: 3760–3766.

83. Zhao FJ, Harris E, Yan J, Ma J, Wu L, Liu W, *et al*. Arsenic methylation in soils and its relationship with microbial *arsM* abundance and diversity, and As speciation in rice. *Environ Sci Technol* 2013; 7147–7154.

84. Gutleben J, Chaib De Mares M, van Elsas JD, Smidt H, Overmann J, Sipkema D. The

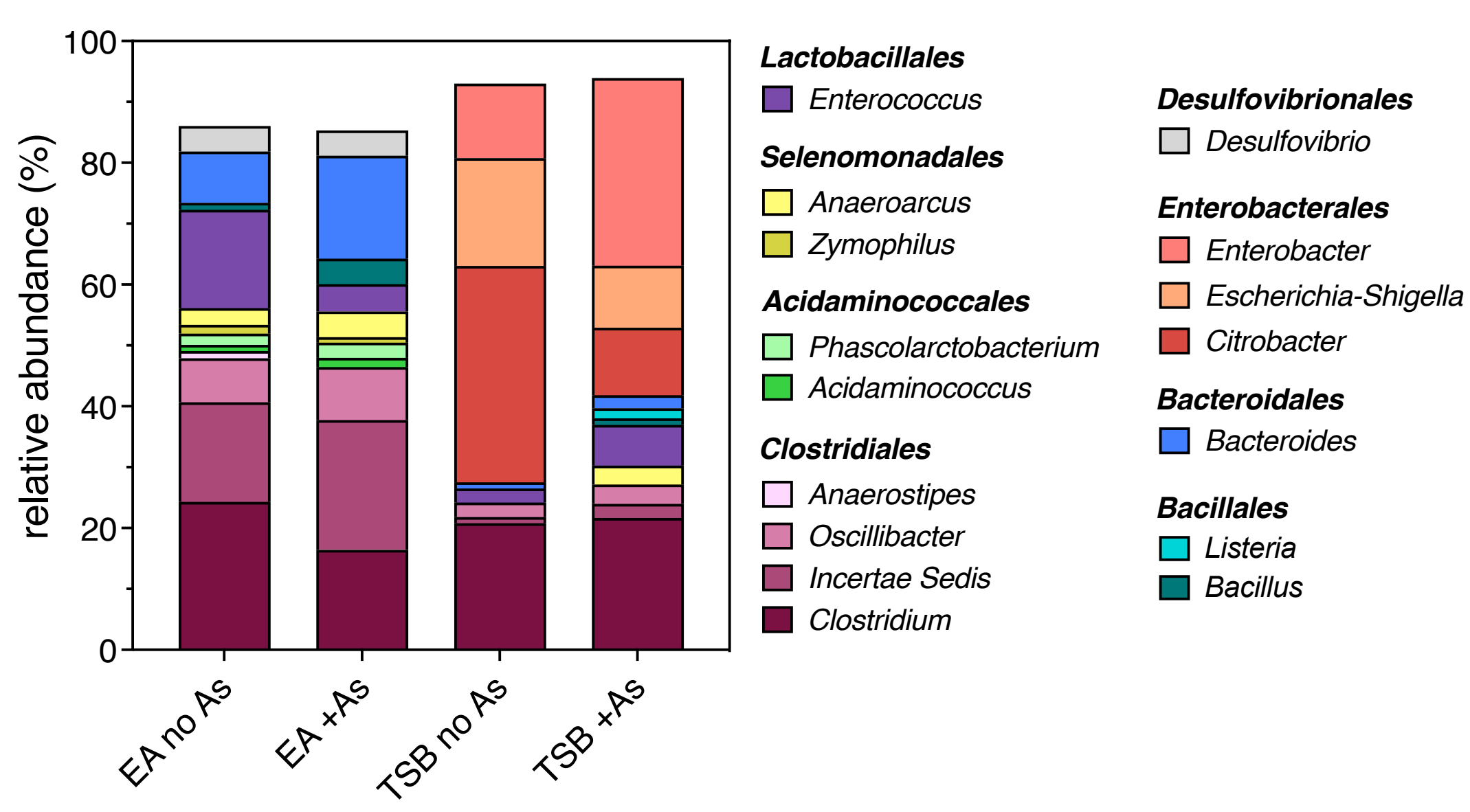
multi-omics promise in context: From sequence to microbial isolate. *Crit Rev Microbiol* 2018; **44**: 212–229.

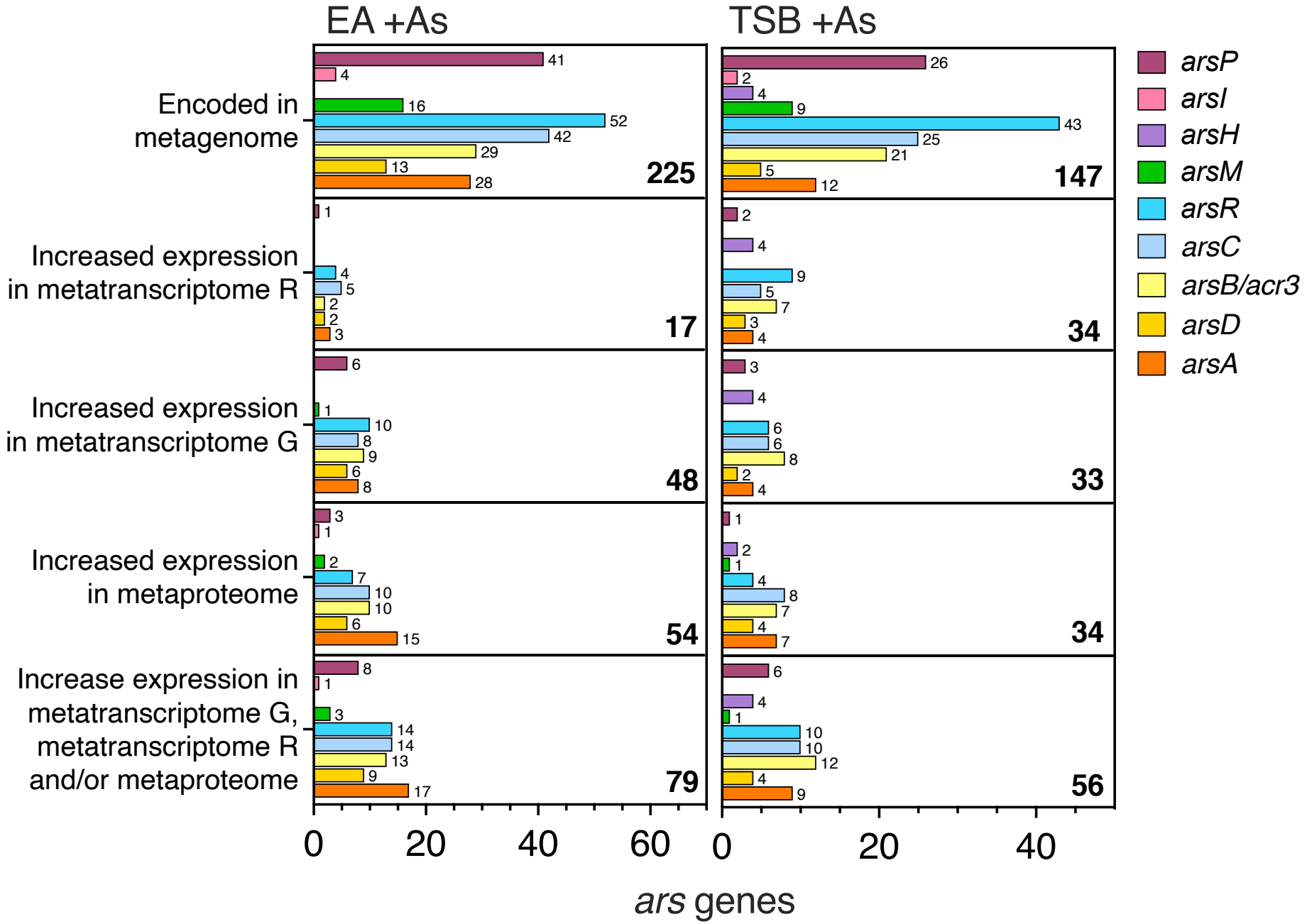
85. Chandran H, Meena M, Sharma K. Microbial biodiversity and bioremediation assessment through omics approaches. *Front Environ Chem* 2020; **1**: 9.

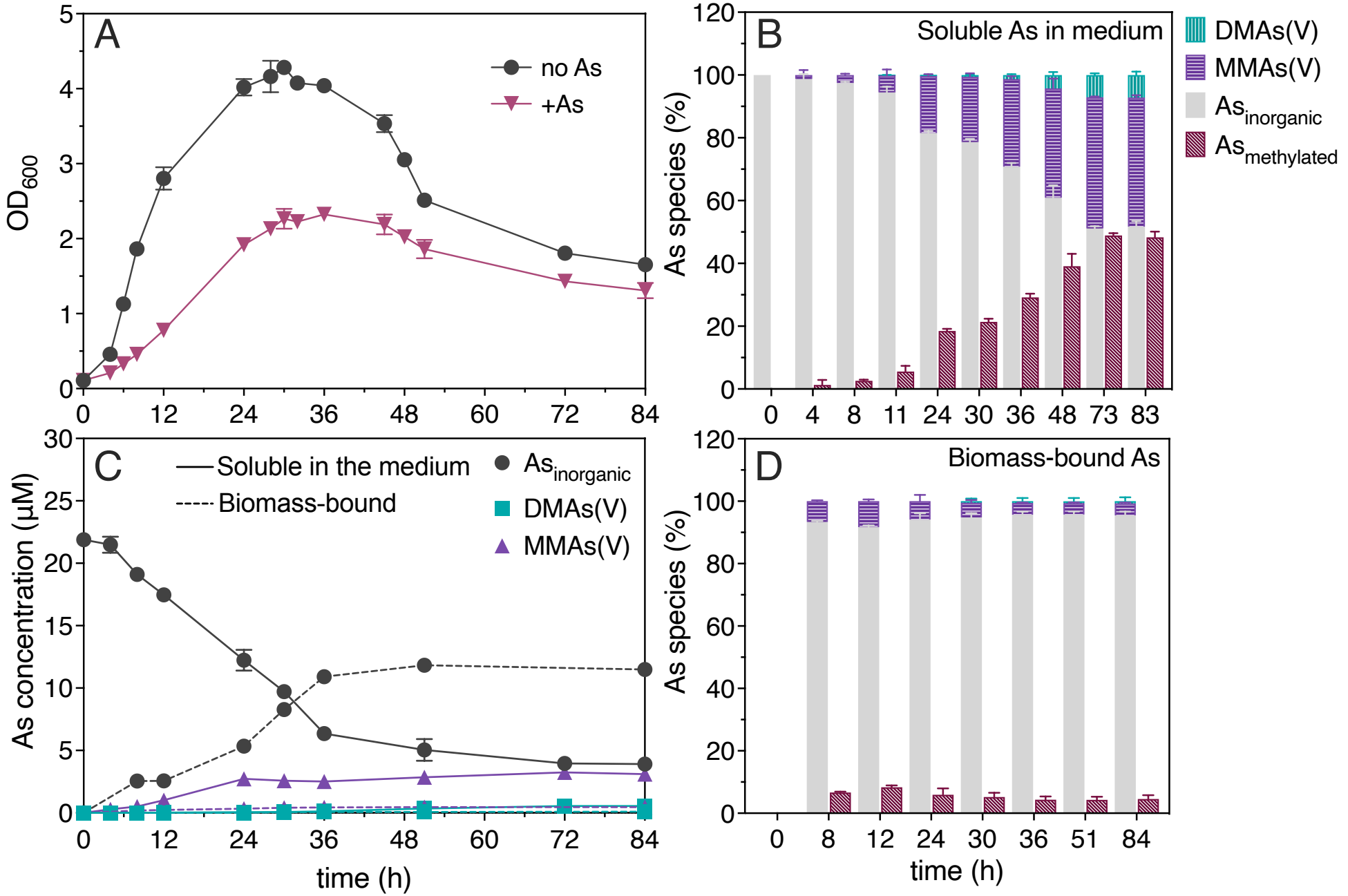
86. Chen S-C, Sun G-X, Yan Y, Konstantinidis KT, Zhang S-Y, Deng Y, *et al*. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *PNAS* 2020; **117**: 10414–10421.

87. Liu S, Zhang F, Chen J, Sun G. Arsenic removal from contaminated soil via biovolatilization by genetically engineered bacteria under laboratory conditions. *J Environ Sci* 2011; **23**: 1544–1550.

88. Chen J, Sun GX, Wang XX, Lorenzo V De, Rosen BP, Zhu YG. Volatilization of arsenic from polluted soil by Pseudomonas putida engineered for expression of the *arsM* arsenic(III) *S*-adenosine methyltransferase gene. *Environ Sci Technol* 2014; **48**: 10337–10344.


relative abundance (%)
100
80
60
40
20
0
EA no As
EA +As
TSB no As
TSB +As
Lactobacillales
Enterococcus
Selenomonadales
Anaeroarcus
Zymophilus
Acidaminococcales
Phascolarctobacterium
Acidaminococcus
Clostridiales
Anaerostipes
Oscillibacter
Incertae Sedis
Clostridium
Desulfovibrionales
Desulfovibrio
Enterobacterales
Enterobacter
Escherichia-Shigella
Citrobacter
Bacteroidales
Bacteroides
Bacillales
Listeria
Bacillus


EA +As
TSB +As
Encoded in metagenome
Increased expression in metatranscriptome R
Increased expression in metatranscriptome G
Increased expression in metaproteome
Increase expression in metatranscriptome G, metatranscriptome R and/or metaproteome
225
147
17
34
48
33
54
34
79
56
0
20
40
60
0
20
40
ars genes
arsP
arsI
arsH
arsM
arsR
arsC
arsB/acr3
arsD
arsA


A
no As
+As
OD600
B
Soluble As in medium
As species (%)
DMAs(V)
MMAs(V)
Asinorganic
Asmethylated
C
Soluble in the medium
Biomass-bound
Asinorganic
DMAs(V)
MMAs(V)
As concentration (μM)
time (h)
D
Biomass-bound As
As species (%)
time (h)

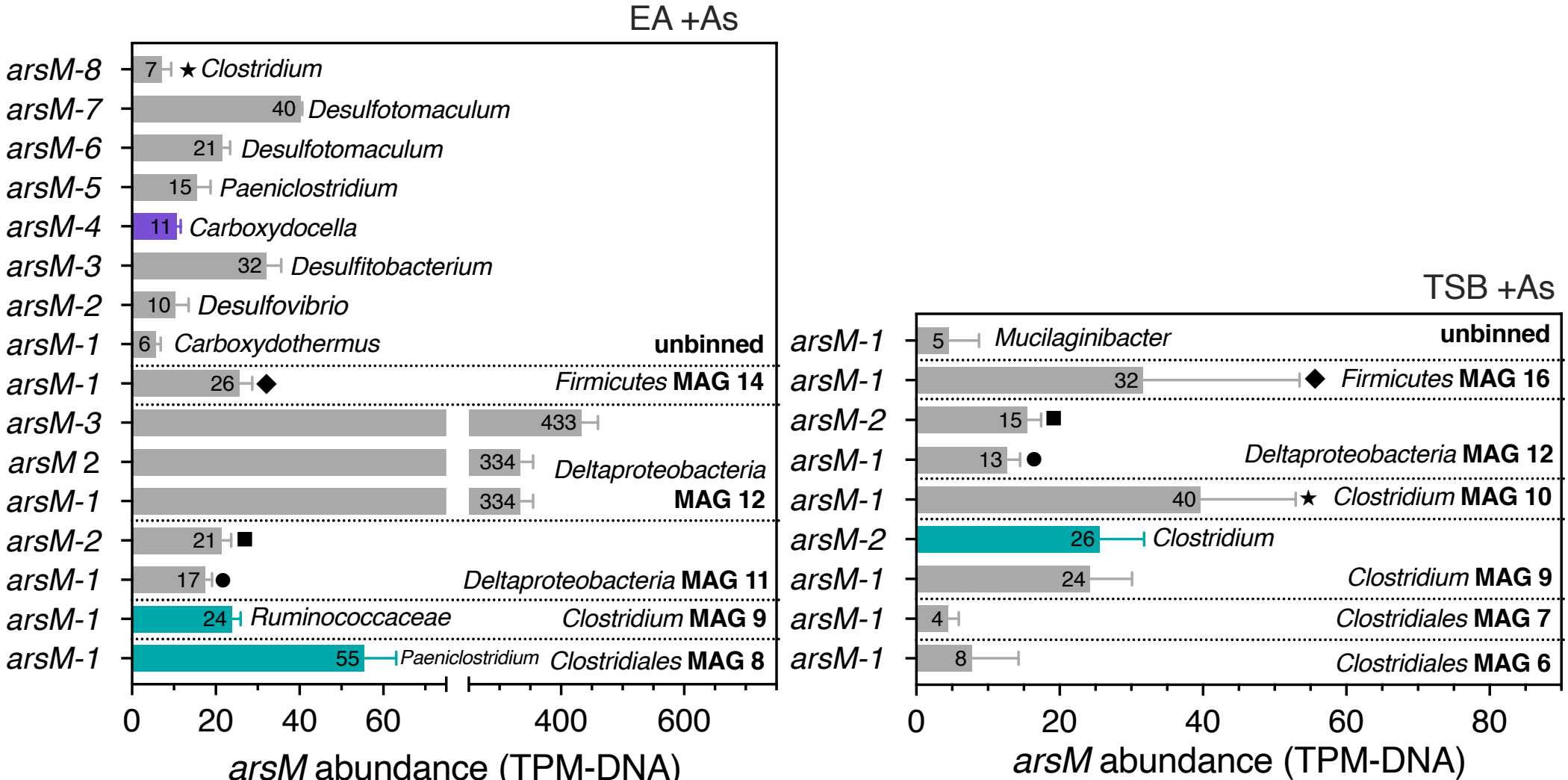

EA +As
arsM-8
arsM-7
arsM-6
arsM-5
arsM-4
arsM-3
arsM-2
arsM-1
arsM-1
arsM-3
arsM 2
arsM-1
arsM-2
arsM-1
arsM-1
arsM-1
7
★ Clostridium
40
Desulfotomaculum
21
Desulfotomaculum
15
Paeniclostridium
11
Carboxydocella
32
Desulfitobacterium
10
Desulfovibrio
6
Carboxydothermus
unbinned
26
◆
Firmicutes MAG 14
433
334
334
Deltaproteobacteria MAG 12
21
■
17
●
Deltaproteobacteria MAG 11
24
Ruminococcaceae
Clostridium MAG 9
55
Paeniclostridium
Clostridiales MAG 8
0
20
40
60
400
600
arsM abundance (TPM-DNA)
TSB +As
arsM-1
arsM-1
arsM-2
arsM-1
arsM-1
arsM-2
arsM-1
arsM-1
arsM-1
5
Mucilaginibacter
unbinned
32
◆
Firmicutes MAG 16
15
■
13
●
Deltaproteobacteria MAG 12
40
★
Clostridium MAG 10
26
Clostridium
24
Clostridium MAG 9
4
Clostridiales MAG 7
8
Clostridiales MAG 6
0
20
40
60
80
arsM abundance (TPM-DNA)

Table 1. Metagenome assembled genomes (MAGs) from EA (upper Table A) and TSB (lower Table B) cultures in the +As condition. Marker lineage: taxonomic rank set by CheckM. Completeness and contamination (%): estimated completeness and contamination of genome as determined by CheckM from the presence/absence of single-copy marker genes and the expected colocalization of these genes. Strain heterogeneity: index between 0 and 100 where a value of 0 means no strain heterogeneity, high values suggest the majority of reported contamination is from closely related organisms (i.e., potentially the same species) and low values suggest the majority of contamination is from phylogenetically diverse sources. Proportion of binned proteins assigned to MAG (%): number of protein-coding genes assigned to the MAG divided by the total number of protein-coding genes binned. Community (%): sum of the number of reads mapped to the contigs in each MAG divided by the total number of reads mapped to all contigs including the unbinned contigs, and normalized to MAG size, assuming an average genome size for all unbinned populations.

A

MAG	Bin	Marker lineage*	Completeness (%)	Contamination (%)	Strain heterogeneity (%)	Genome size (Mbp)	Community (%)	GC content	ArsM-encoding genes	Binner §
1	36	*Bacteroidales* (o)	98.5	0.4	0.0	3.8	7.26 ±0.32	39.1	0	A
2	15	*Clostridiales* (o)	98.7	0.0	0.0	2.2	0.52 ±0.06	58.4	0	B
3	21	*Clostridiales* (o)	95.2	0.0	0.0	4.3	5.73 ±0.24	28.5	0	C
4	4	*Clostridiales* (o)	90.7	0.0	0.0	2.1	0.30 ±0.03	57.5	0	A
5	24	*Clostridiales* (o)	97.8	0.3	0.0	2.0	25.21 ±0.23	43.2	0	A
6	9	*Clostridiales* (o)	100.0	1.3	50.0	3.2	12.68 ±0.68	54.9	0	A
7	35	*Clostridiales* (o)	98.0	3.3	55.6	5.3	0.84 ±0.21	44.0	0	A
8	31	*Clostridiales* (o)	97.9	3.5	0.0	3.8	1.48 ±0.12	28.2	1	A
9	20	*Clostridium* (g)	97.2	2.2	0.0	3.4	0.70 ±0.06	30.1	1	C
10	18	*Clostridium* (g)	96.5	2.9	16.7	4.0	0.27 ±0.05	30.0	0	A
11	11	*Deltaproteobacteria* (o)	99.2	0.7	100.0	3.4	0.63 ±0.05	57.4	2	BC
12	33	*Deltaproteobacteria* (o)	100.0	1.2	0.0	3.3	12.71 ±0.49	57.8	3	BC
13	28	*Firmicutes* (p)	99.9	0.0	0.0	2.5	1.02 ±0.10	47.2	0	B
14	27	*Firmicutes* (p)	91.9	3.3	92.3	3.1	0.84 ±0.07	49.2	1	BC
15	8	*Lactobacillales* (o)	99.6	0.0	0.0	2.7	3.30 ±0.61	36.8	0	A
16	1	*Lactobacillales* (o)	99.3	4.6	0.0	4.1	0.99 ±0.05	39.1	0	BC
17	16	*Selenomonadales* (o)	100.0	1.5	0.0	2.2	2.60 ±0.24	41.3	0	C

* (p) phylum, (o) order, or (g) genus.

§ A,B and C refer to MetaBAT 2, MaxBin 2.0 and CONCOCT respectively.

B

MAG	Bin	Marker lineage*	Completeness (%)	Contamination (%)	Strain heterogeneity (%)	Genome size (Mbp)	Community (%)	GC content	N. of ArsM-encoding genes	Binner §
1	12	*Clostridiales* (o)	100.0	0.0	0.0	3.1	0.41 ±0.03	54.8	0	B
2	9	*Clostridiales* (o)	98.9	0.0	0.0	4.7	1.26 ±0.24	28.4	0	C
3	39	*Clostridiales* (o)	98.0	0.3	0.0	2.0	2.36 ±0.51	43.2	0	A
4	4	*Clostridiales* (o)	99.3	0.7	100.0	2.7	4.77 ±3.14	56.1	0	A
5	16	*Clostridiales* (o)	98.7	0.9	0.0	2.8	0.35 ±0.21	35.7	0	A
6	19	*Clostridiales* (o)	99.2	1.1	0.0	3.5	0.39 ±0.28	31.2	1	BC
7	1	*Clostridiales* (o)	98.7	1.3	50.0	2.6	0.25 ±0.07	56.1	1	A
8	15	*Clostridiales* (o)	97.3	2.5	16.7	2.7	0.17 ±0.02	60.5	0	C
9	28	*Clostridium* (g)	99.3	5.5	23.1	5.6	1.42 ±0.30	30.1	2	A
10	27	*Clostridium* (g)	98.6	6.9	0.0	4.6	2.10 ±0.60	32.3	1	A
11	32	*Deltaproteobacteria* (o)	94.8	0.0	0.0	3.1	0.21 ±0.06	59.3	0	BC
12	38	*Deltaproteobacteria* (o)	98.3	1.8	50.0	3.4	0.81 ±0.08	57.6	2	BC
13	10	*Enterobacteriaceae* (f)	96.6	0.7	33.3	4.3	0.39 ±0.08	52.8	0	B
14	42	*Enterobacteriaceae* (f)	95.7	2.1	12.5	5.1	6.77 ±0.35	56.3	0	BC
15	31	*Firmicutes* (p)	99.9	0.0	0.0	2.4	0.37 ±0.06	47.6	0	A
16	33	*Firmicutes* (p)	100.0	0.6	0.0	3.2	1.97 ±1.09	49.1	1	BC
17	7	*Lactobacillales* (o)	99.6	0.0	0.0	2.9	2.64 ±0.52	36.5	0	C
18	5	*Lactobacillales* (o)	98.9	4.2	0.0	4.1	1.94 ±0.81	39.1	0	AB
19	36	*Selenomonadales* (o)	100.0	1.5	0.0	2.3	0.81 ±0.11	41.1	0	A

* (p) phylum, (o) order, or (g) genus.

§ A,B and C refer to MetaBAT 2, MaxBin 2.0 and CONCOCT, respectively.