Supplementary Information (SI) for the research article: Variability in Arsenic Methylation Efficiency across Aerobic and Anaerobic Microorganisms.

Karen Viacava †, Karin Lederballe Meibom †, David Ortega §, Shannon Dyer †, Arnaud Gelb ¦, Leia Falquet †, Nigel P. Minton §, Adrien Mestrot ‡, Rizlan Bernier-Latmani †.

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

‡ Institute of Geography, Soil Science Group, University of Bern, Bern, Switzerland.

§ Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Centre for Biomolecular Sciences, University of Nottingham, Nottingham, United Kingdom.

¦ Laboratory for Environmental Biotechnology, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

This SI contains part of Materials and Methods and **17 supplemental figures**. The **26 supplemental tables** can be found in the SI XLSX file.

SI Materials and Methods

Total Arsenic and Arsenic Speciation Analysis.

Arsenic speciation was obtained using either an ICP-MS 7700x coupled to an HPLC 1260 Infinity or an ICP-QQQ 8900, respective instrument settings in Table S19, coupled to an HPLC 1260 Infinity II (Agilent Technologies, CA, US) with an anion exchange PRP X-100 HPLC column (150 × 4.1 mm I.D., 10 μm, Hamilton, NV, US) (pump: 1ml min^{-1}, injection volume: 15 μl, autosampler: 4°C and column compartment: 20 °C). The eluent was 6.66 mM NH_4NO_3 and 6.66 mM $(NH_4)_2HPO_4$ (pH 6.2 adjusted with HNO_3). The As standards consisted of TMAsO as trimethyl arsine oxide (Argus Chemicals Srl., Italy), DMAs(V) as sodium dimethylarsinate (ABCR, Germany), MMAs(V) as monomethylarsonic acid (Chemservice, PA, USA) and As(V) as $Na_2HAsO_4 \cdot 7H_2O$ (Sigma-Aldrich, MO, USA). As mentioned in the manuscript, the samples were oxidized prior to analysis, due to the co-elution of As(III) and TMAsO species. The remaining As(III) was quantified in its oxidized form as As(V). In this way, As speciation analysis allowed the discrimination of inorganic As (detected as As(V): sum of remaining As(III) plus As(V) produced during incubations) from tri-, di- or monomethylated arsenicals (as TMAsO, DMAs(V) or MMAs(V)) but not the identification of the original redox state of the methylarsenicals as produced by the strains. Total As concentrations were measured using the same ICP-MS instruments in stand-alone mode. To avoid overloading the HPLC column, the preserved soluble As samples were diluted (1:5) in 1% HNO_3 prior to As speciation analysis by HPLC-ICP-MS.

Cloning the *arsM* Genes and Gene Expression in *E. coli* AW3110 (DE3).

The NCBI database was probed for putative *arsM* genes for each strain (Table S20). The ArsM proteins were aligned against ArsM enzymes proven to be active (Figure S16) (Multiple Sequence

Alignment tool, EMBL-EBI, [1]). Genomic DNA from each strain was used as a template in PCR using Phusion polymerase (New England Biolabs, MA, USA) to amplify the genes encoding *arsM* using primers gene_F and gene_R (Table S21). Gibson assembly master mix (New England Biolabs) was used to clone the *arsM* genes into pET28b(+) digested with *Xho*I and *Nco*I. The corresponding constructs contain the native gene followed by a sequence encoding a C-terminal tag of Ala-6xHis. All constructs were verified by Sanger sequencing prior to transformation into *E. coli* AW3110 (DE3). The activity of the *arsM* genes was tested in 10-15 μM As(III)-amended LB medium (containing 50 μg ml[-1] kanamycin and 0.5 mM isopropyl-β-D- thiogalactopyranoside (IPTG)) inoculated with 1% inoculum (v/v) of pre-grown overnight culture (in 50 μg ml[-1] kanamycin). The cultures final volume was 100 ml. Controls with the empty vector pET28b(+) or no-As(III) were included. All conditions were run in triplicate. Soluble and volatile As produced were sampled and analyzed using the methods described in the manuscript.

Growth Conditions of *Clostridium pasteurianum* H0D0R4, Strain Used for Genetic Modification

The strain used for genetic modification of *C. pasteurianum* is a hypertransformable variant based on the type strain *C. pasteurianum* DSM 525, designated *H0D0R4* [2]. The *pyrE* truncated mutant of *H0D0R4* Δ*pyrE* was kindly provided by Dr. Grosse-Honebrink and was generated in the same manner as the strain H1 Δ*pyrE* [34]. *C. pasteurianum* was cultured in Don Whitley A95TG Anaerobic Workstation (Don Whitley, UK), maintained with a gas mixture of (10%:80%:10% CO_2:N_2:H_2, 1 atm). It was grown on RCM agar (Oxoid Ltd) or in liquid 2xYPG (16 g L[-1] veggie peptone (Novagen, Merck), 10 g L[-1] veggie yeast extract (Novagen, Merck) and 5 g L[-1] NaCl) supplemented as required with 7.5 μg mL[-1] thiamphenicol or 40 μg mL[-1] uracil. To select for

integration, restoration of uracil autotrophy was used by growing *C. pasteurianum* on clostridial basal media agar with 5% (w/v) glucose [5]. Vectors used in genetic modifications of *C. pasteurianum* were cloned in *E. coli* (DH5α, New England Biolabs UK Ltd, UK) and grown aerobically in LB broth or agar at 30°C for cloning purposes and at 37°C for overnight cultures. Broth and agar were supplemented with 25 µg mL[-1] chloramphenicol and 100 µg ml[-1] kanamycin as required. The same vectors were further propagated in an *E.coli* TOP10 strain for the purpose of *in vivo dam*+ and *dcm*+ methylation as outlined in the previous work [4]. Methylated vectors were then isolated and electroporated into *C. pasteurianum* as previously described [6].

Isolation of the Δ*acr3* and Δ*pyrE* Δ*acr3* Mutants.

Transformed *C. pasteurianum* was plated on RCM agar supplemented with thiamphenicol. Positive colonies were confirmed for plasmid presence by colony PCR and positive clones were then grown in liquid 2xYPG for 4h and patch plated onto RCM agar with thiamphenicol or with thiamphenicol containing the inducer theophylline. Colonies that grew on RCM agar with thiamphenicol and theophylline were re-streaked on the same agar to single colony and screened for successful deletion of *acr3* by colony PCR using the primers 2AB and 3AB (Table S2). Clones showing the required deletion were re-streaked to single colony on RCM agar without antibiotic or inducer and checked for plasmid curing by patch plating before being stored at -80°C in 10% DMSO for further use. An amplified PCR fragment of the *acr3* locus was sequenced to confirm in-frame, marker-less deletion of the *acr3* gene.

Competent cells of the mutant Δ*pyrE* Δ*acr3* were then prepared as described [6]. The complementation vector pMTL_KS_12_acr3 was then transformed into the Δ*pyrE* Δ*acr3* mutant and plated onto RCM agar supplemented with thiamphenicol. Clones were screened for plasmid

presence by colony PCR and then streaked onto clostridia basal media (CBM) agar. Growing colonies were then re-streaked on the same media agar and screened for *pyrE* repair and *acr3* integration using colony PCR with the primer pair 9Q and 2R (Table S2). Successful integration of *acr3* at the *pyrE* locus resulted from uracil autotrophy being restored and generated the complementation strain Δ*acr3 pyrE::acr3*. Complementation mutants were then cured of the plasmid pMTL_KS_12_acr3 by patch plating on RCM agar with and without thiamphenicol. A PCR fragment of the *pyrE* locus was sequenced to confirm successful restoration of the *pyrE* allele and integration of *acr3*.

Arsenic Methylation by *Clostridium pasteurianum* Δ*acr3*.

The 2xYTG liquid medium was amended with 100 μM As(III) along with a no-As(III) control, inoculated with 1% inoculum (v/v) of pre-grown overnight culture from *C. pasteurianum* Δ*acr3*, parental *C. pasteurianum* WT, or *C. pasteurianum* Δ*acr3* complementation strain *pyrE::acr3*. All conditions were set in triplicates. Soluble As species in the medium were sampled as explained in the main text, volatile As was not assayed. To quantify intracellular As species, a pellet from a 2-ml culture was collected (10 min, 16,873 g, 20 °C), washed twice with 1 ml 1× PBS buffer (10x solution, Fisher BioReagents, NH, US), re-suspended in 200 μl 1× PBS and digested with 2 mg ml^{-1} lysozyme (AppliChem, Germany) at 37 °C for 1 h. The lysate was centrifuged (5 min, 16,873 g, 20 °C), oxidized by adding 10% of 35% H_2O_2 and heating to 95 °C for 1 h and preserved in 1% HNO_3 at 4 °C for As speciation analysis. The lysate was centrifuged and the supernatant diluted (1:20) in 1% HNO_3 prior to As speciation analysis by HPLC-ICP-MS. The biomass concentration in the lysate was quantified with the Pierce BCA Protein Assay Kit.

Transcription of *arsM* in *Clostridium pasteurianum* WT and Δ*acr3*.

Quantitative reverse transcription PCR (qRT-PCR) was performed on RNA derived from three replicate cultures (four replicates per biological sample) at mid-exponential phase, OD_{600} ~1.5. The cells were lysed and total RNA purified on a RNeasy column using the RNeasy Mini Kit according to the manufacturer's instructions, treated with DNase I (Promega, WI, USA) for 1h at 37 °C and cleaned (RNeasy clean-up protocol, Qiagen, Germany). Reverse transcription of RNA was performed using the GoScript™ ReverseTranscription System (Promega, WI, USA) and random hexamers as primers. 200 nM of species-specific primers targeting *arsM*, *acr3*, *gyrA* (encoding gyrase subunit A) or *rho* (transcription termination factor Rho), the last two as the reference genes for relative quantification (Table S22), were used to amplify the cDNA template in 10 μl volume of 1× SensiFast SYBR Mix (BioLine, UK) in a Mic qPCR Cycler (Bio Molecular Systems, Australia) (95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 5 s and annealing/elongation at 60 °C for 20s). PCR control reactions were performed using RNA in the absence of reverse transcriptase and using no template. The gene expression ratios between As(III)-containing samples vs. no-As(III) controls were calculated by the relative expression software tool (REST©) [7] included in the Mic qPCR Cycler software.

Membrane-integrity Assessment for *Methanosarcina mazei* Cells Using Flow Cytometry

M. mazei Gö1 cells were grown as described above with 10 or 50 μM As(III). At each sampling time, 1 ml of culture was centrifuged (10 min, 10,000 g, 20 °C), the pellet re-suspended in 1× PBS buffer (10x solution, Fisher BioReagents, NH, US), thoroughly mixed with the staining mixture of 100 μl of 100× SYBR Green I (SG) (Invitrogen), and 133 μl of 1.5 mM propidium iodide (PI)

(Sigma) to a final volume of 1 ml (10× SG and 200 μM PI final concentrations) and incubated for 15 min at room temperature.

Flow cytometry measurements were performed with a 5-laser LSRII SORP flow cytometer (Becton, Dickinson and Company, NJ, USA). SG was excited by the Blue laser (488 nm) and detected using a 530/30 band pass filter. PI was excited by the YG laser (561 nm) and detected using a 610/20 band pass filter. 30'000 events per sample were analyzed into four populations (no fluorescence, SG, SG/PI, or PI). Cells could be assigned to the membrane-compromised population, based on the gating of double-stained and single-stained controls of glutaraldehyde-fixed and ethanol-permeabilized cells. Cytometric data were acquired and analyzed using BD FACSDiva™ software v. 8.0.1 (BD Biosciences, CA, USA).

Visualization with confocal microscopy of cells treated and stained in the same way as for the cytometric measurements was carried out using Zeiss LSM 700 in the upright configuration equipped with a Plan-Apochromat 63x/1.40 oil immersion objective. The cells were immobilized on the glass slides with 1% low melting agarose (Carl Roth, Germany). The image acquisitions were performed with a zoom factor of 1 and a pixel size of 0.1 μm. The pinhole size was set at 1.0 Airy unit for the PI channel leading to an optical section of 47 μm; the same section thickness was used for the SG channel. Images were acquired sequentially, for the SG, a 488 nm excitation wavelength was used and the signal measured after a LP 490nm emission filter, for the PI, a 555 nm excitation wavelength was used in combination with a LP 560nm emission filter.
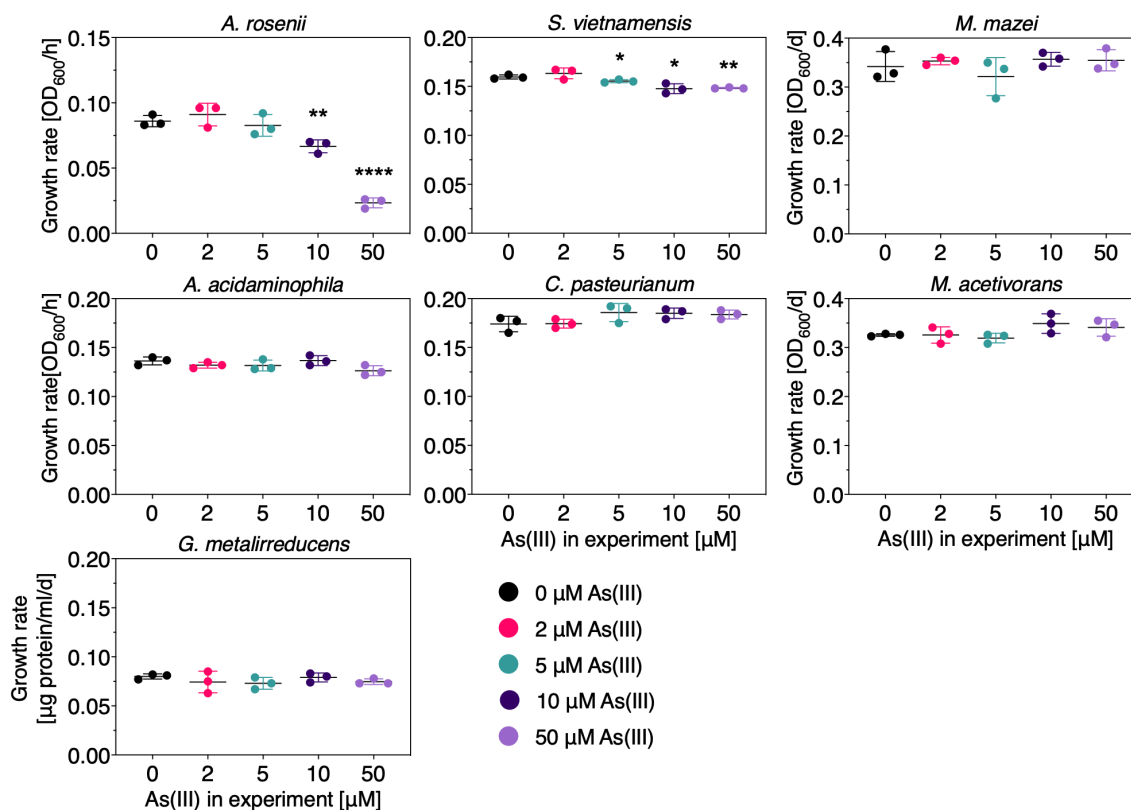
SI Figures

Figure S1. Growth rate between time 0h and the end of exponential growth phase (*Arsenicibacter rosenii*: 36h*, Streptomyces vietnamensis*: 24h, *Methanosarcina mazei*: 7.1d, *Anaeromusa acidaminophila*: 11h, *Clostridium pasteurianum*: 24h, *Methanosarcina acetivorans*: 7.3d, *Geobacter metallireducens:* 24h) for each individual species grown in the presence of varying initial As(III) concentrations (0, 2, 5, 10 and 50 µM As(III)). Black horizontal line represents the mean value and coloured error bars, plus and minus one standard deviation. Samples with one or more asterisk (*) are significantly different from samples in 0 µM As(III) (p value <0.05, See p-value summary in Table S5). Growth rate values and t-test results listed in Tables S4 and S5.
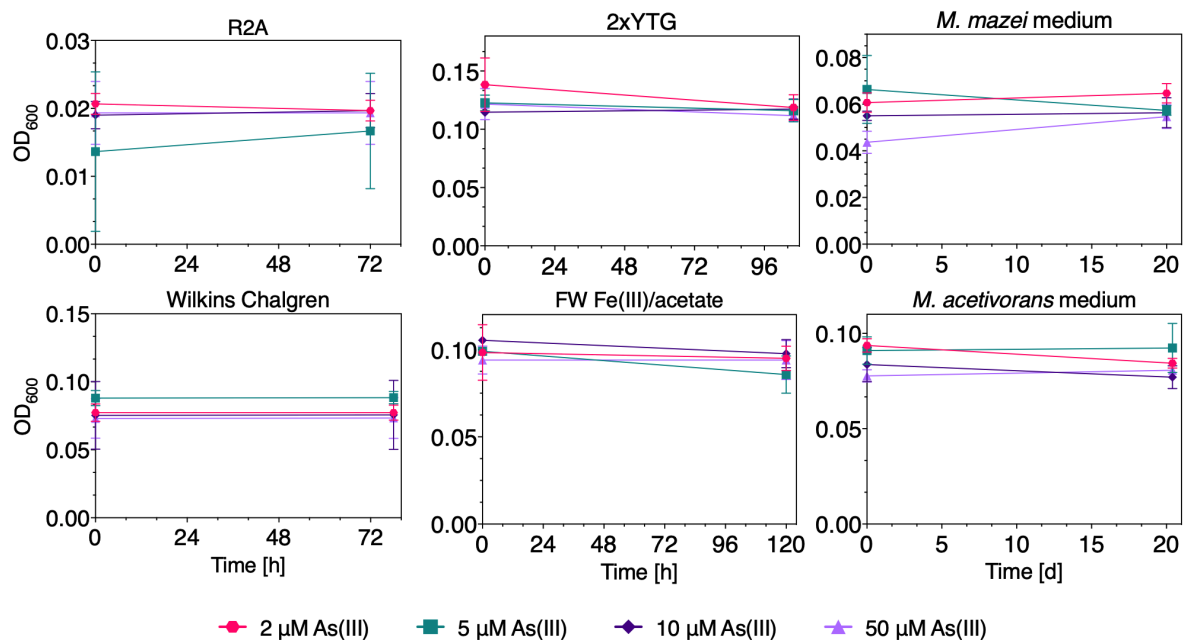
Figure S2. Abiotic control growth curves (OD$_{600}$) as a function of time for each type of medium used with varying initial As(III) concentrations (0, 2, 5, 10 and 50 μM As(III)). R2A medium used for *A. rosenii* and *S. vietnamensis;* 2xYTG used for *C. pasteurianum*; Wilkins Chalgren broth used for *A. acidaminophila* and fresh water (FW) Fe(III)/acetate medium used for *G. metallireducens*. Data points represent the mean value and error bars, plus and minus one standard deviation. Individual values for each biological replicate are listed in Table S6.
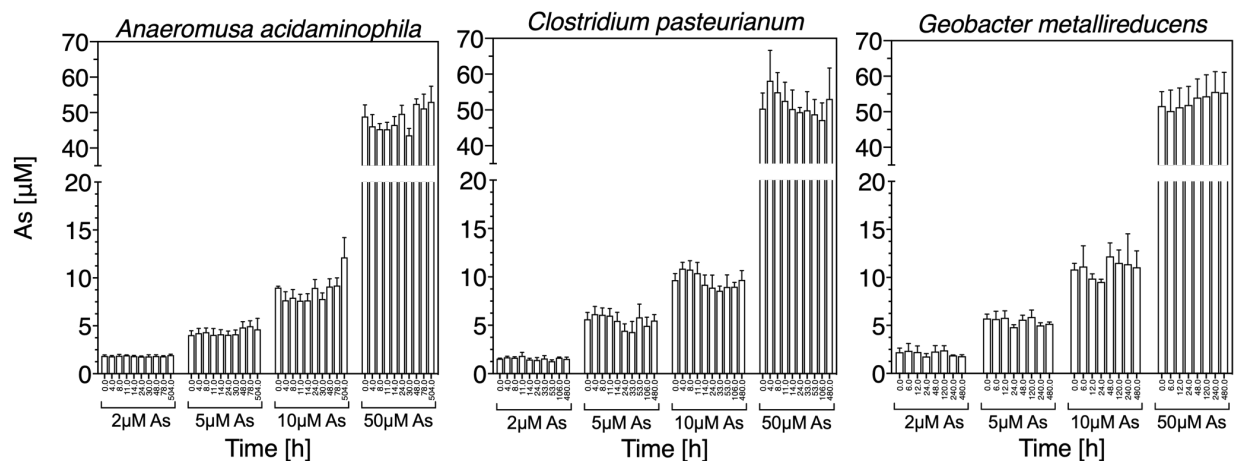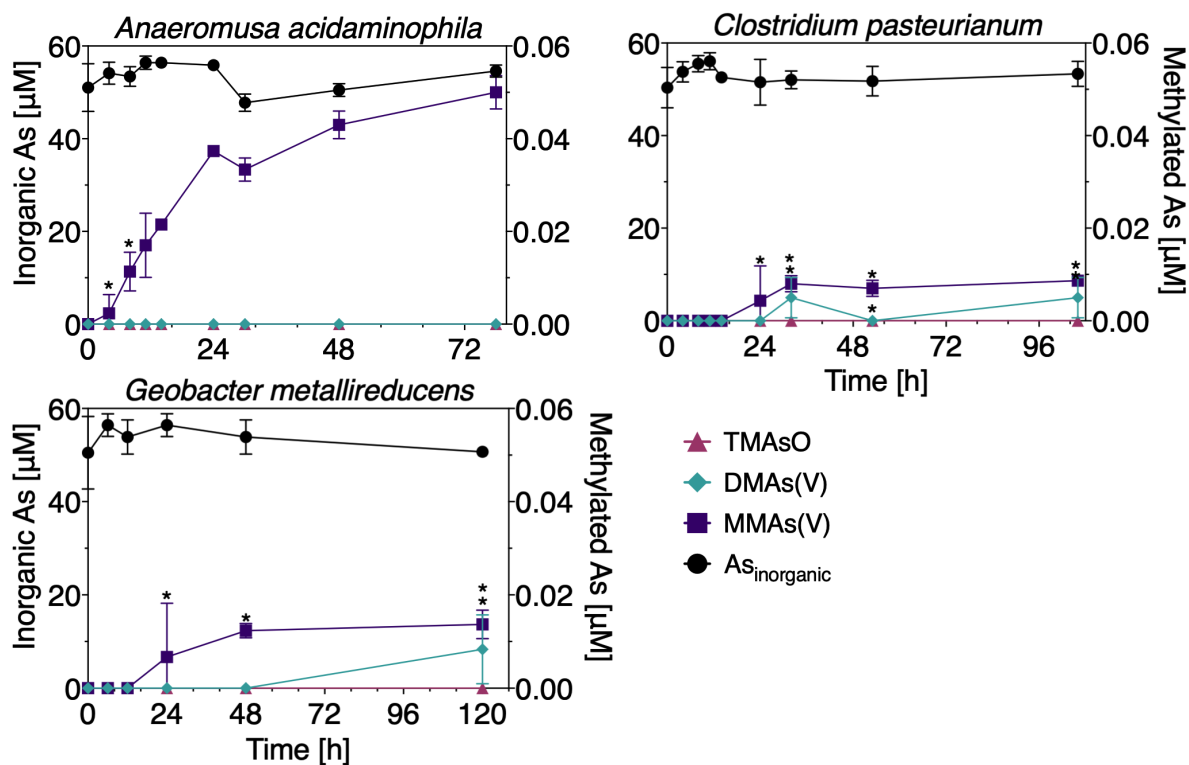
Figure S3. Total soluble arsenic concentration in filtered medium for each anaerobic bacterium culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 μM). Sampling times (x-axis) *Anaeromusa acidaminophila*: 0, 4, 8, 11, 14, 24, 30, 48, 78 and 504 hours. *Clostridium pasteurianum*: 0, 4, 8, 11, 14, 24, 33, 53, 106 and 480 hours. *Geobacter metallireducens*: 0, 6, 12, 24, 48, 120 and 480 hours. Bar height represents the mean value and error bars plus one standard deviation. Values for each biological replicate are listed in Tables S7-S9.

Figure S4. Concentration of soluble arsenic species in filtered medium from anaerobic bacterial strains cultures grown with 50 μM As(III) as the initial concentration. Inorganic arsenic is plotted on the left y-axis and methylated arsenic on the right y-axis. Data points represent the mean value and error bars, plus and minus one standard deviation. Time points marked with an asterisk (*) are below LOQ but above LOD. Values for each biological replicate are listed in Tables S7-S9.
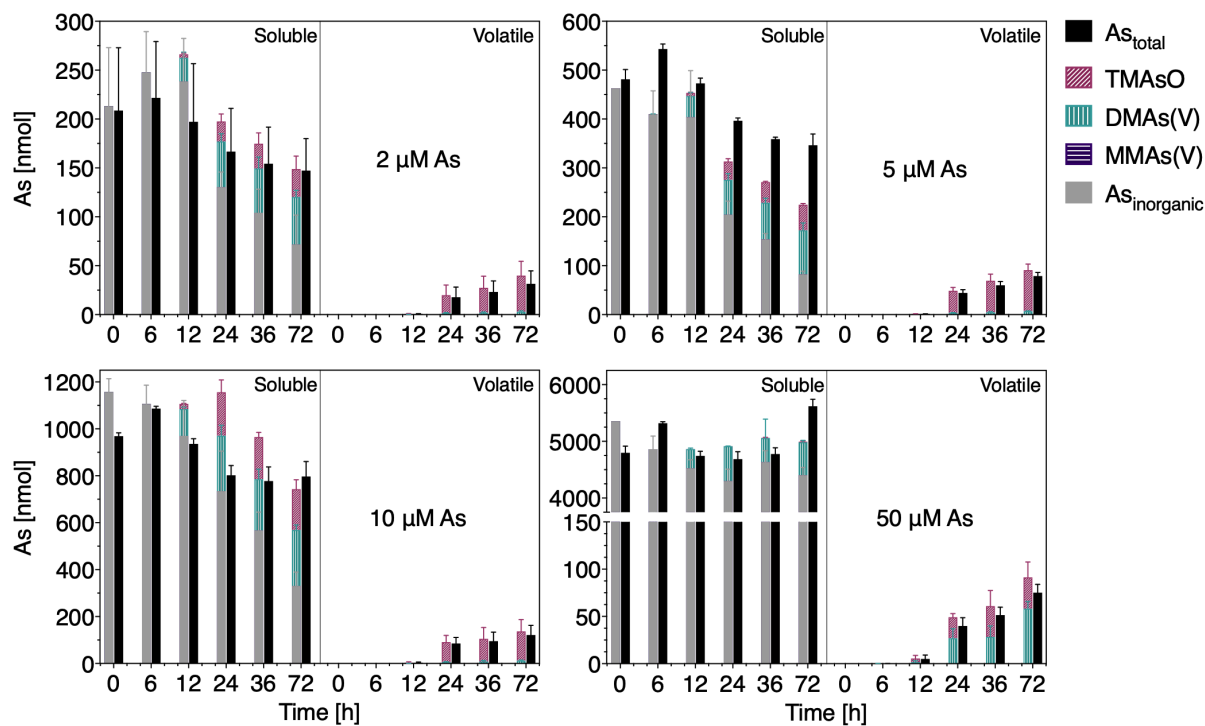
Figure S5. Amount of soluble arsenic species in filtered medium (left panels) and volatile arsenic species chemo-trapped during medium flushing (right panels) from an *A. rosenii* culture grown in the presence of varying initial As(III) concentrations (2, 5, 10, and 50 µM). Bar height represents the mean value and error bars plus one standard deviation. Values for each biological replicate are listed in Table S10.
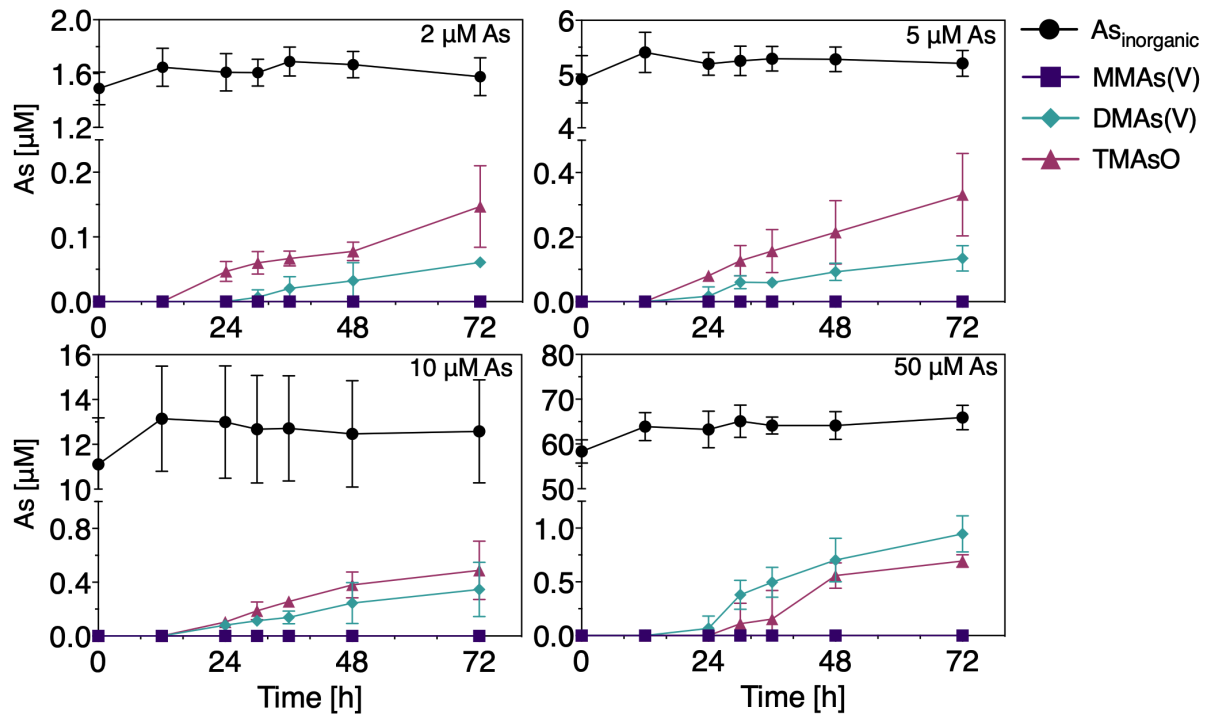
Figure S6. Concentration of soluble arsenic species in filtered medium from a *S. vietnamensis* culture grown in the presence of varying initial As(III) concentrations (2, 5, 10, and 50 μM). Data points represent the mean value and error bars, plus and minus one standard deviation. Values for each biological replicate are listed in Table S11.
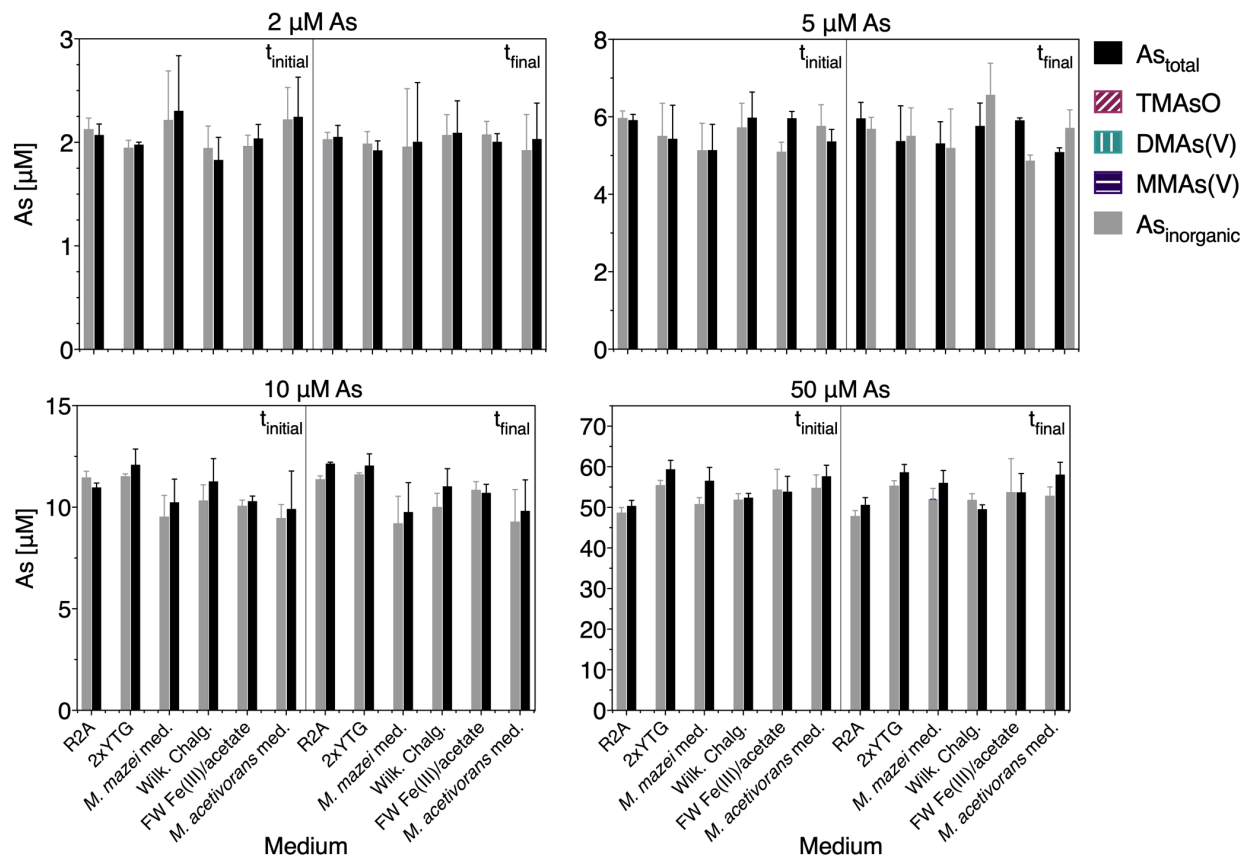
Figure S7. Concentration of soluble arsenic species in filtered media at time 0h (left panels) and at: R2A, 72h; 2xYTG, 106h; *M. mazei* medium, 20d; Wilkins Chalgren broth, 78h; FW Fe(III)/acetate, 120h and *M. acetivorans* medium, 20d (right panels). Bar height represents the mean value and error bars plus one standard deviation. Values for each replicate are listed in Tables S7-S13.
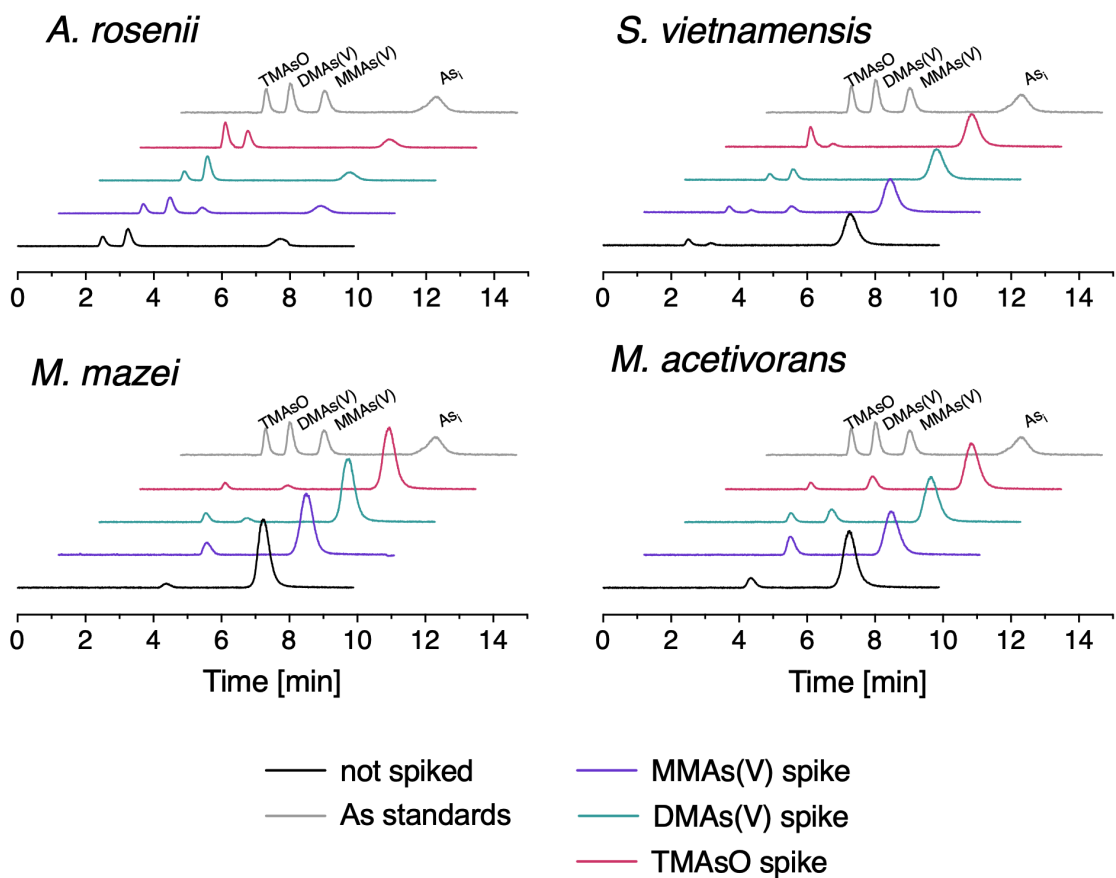
Figure S8. HPLC-ICP-MS spectra from samples: *A. rosenii* (2 μM As(III) after 72h), *S. vietnamensis* (2 μM As(III) after 72h), *M. mazei* (2 μM As(III) after 20d) and *M. acetivorans* (2 μM As(III) after 20d). Individual values are listed in Table S23.

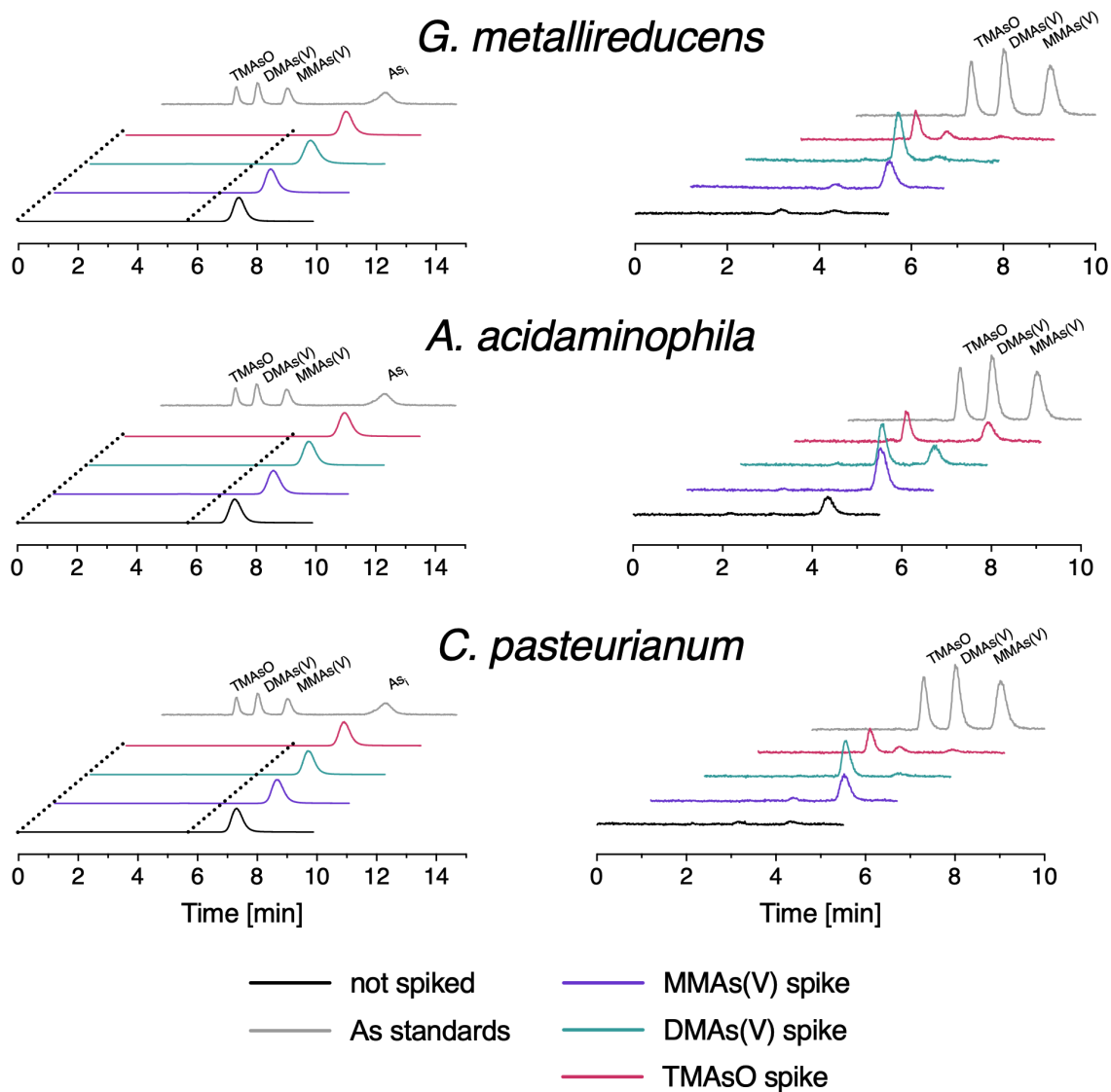**Figure S9.** HPLC-ICP-MS spectra from samples: *G. metallireducens* (50 μM As(III) after 120h), *A. acidaminophila* (50 μM As(III) after 78h) and *C. pasteurianum* (50 μM As(III) after 106h). Full spectra are on left panels and spectra from time 0 to 5.5 min (indicated with perpendicular dashed lines on left panels) on right panels. Individual values are listed in Table S23.
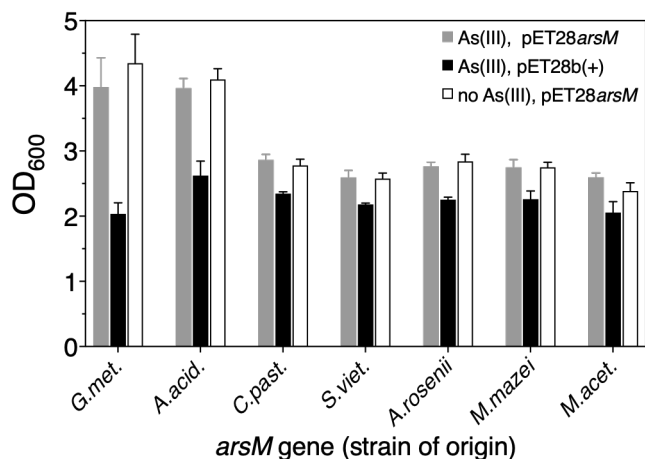
Figure S10. Growth as OD_{600} at the end of the exponential phase in a culture of *E. coli* AW3110 (DE3) expressing ArsM from the various species grown with 10-15 μM As(III) as the initial concentration (grey bars) or no As(III) (0 μM As(III)) (empty bars) and controls with empty plasmid pET28b(+) grown with 10-15 μM As(III) as the initial concentration (black bars). Bar height represents the mean value and error bars plus one standard deviation. Values for each biological replicate are listed in Table S24.
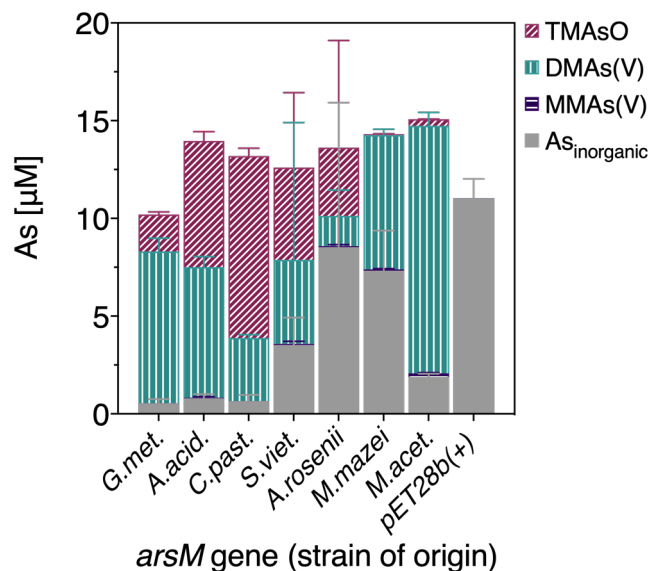
Figure S11. Concentration of soluble arsenic species in filtered medium after 72h in a culture of *E. coli* AW3110-(DE3) expressing ArsM from the various species or bearing the empty pET28b(+) plasmid and grown with 10-15 μM As(III) as the initial concentration. Bar height represents the mean value and error bars plus one standard deviation. G. met.= *Geobacter metallireducens*; A. acid.= *Anaeromusa acidaminophila*; C. past.= *Clostridium pasteurianum*; S. viet.= *Streptomyces vietnamensis*; A. rosenii= *Arsenicibacter rosenii*; M. mazei= *Methanosarcina mazei*; M. acet.= *Methanosarcina acetivorans*. Values for each biological replicate are listed in Table S25.
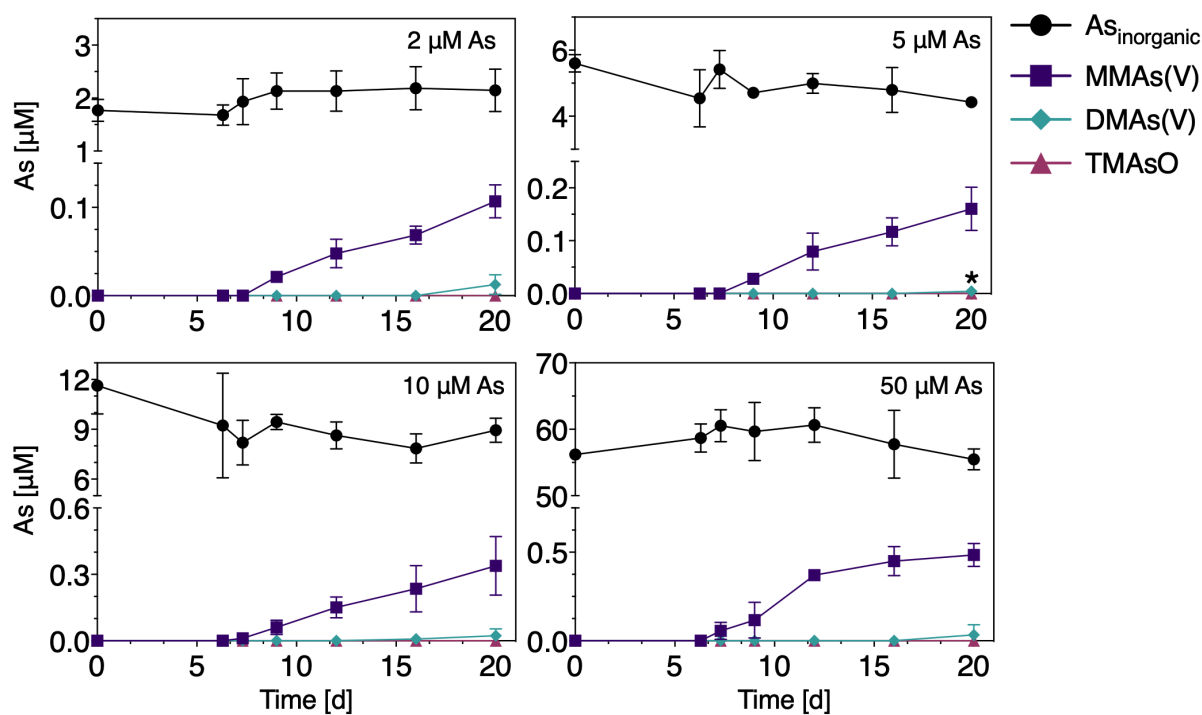
Figure S12. Concentration of soluble arsenic species in filtered medium from a *M. mazei* culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 μM). Data points represent the mean value and error bars, plus and minus one standard deviation. Time points marked with an asterisk are below LOQ but above LOD. Values for each biological replicate are listed in Table S12.
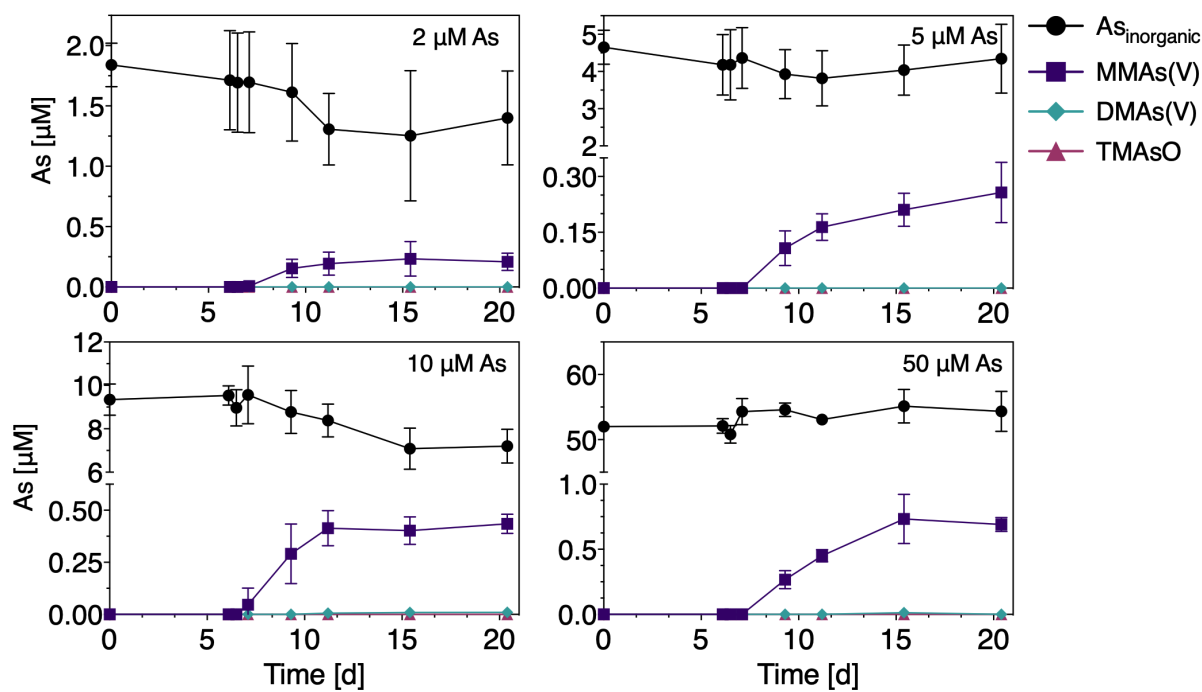
Figure S13. Concentration of soluble arsenic species in filtered medium from a *M. acetivorans* culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 μM). Data points represent the mean value and error bars, plus and minus one standard deviation. Values for each biological replicate are listed in Table S13.
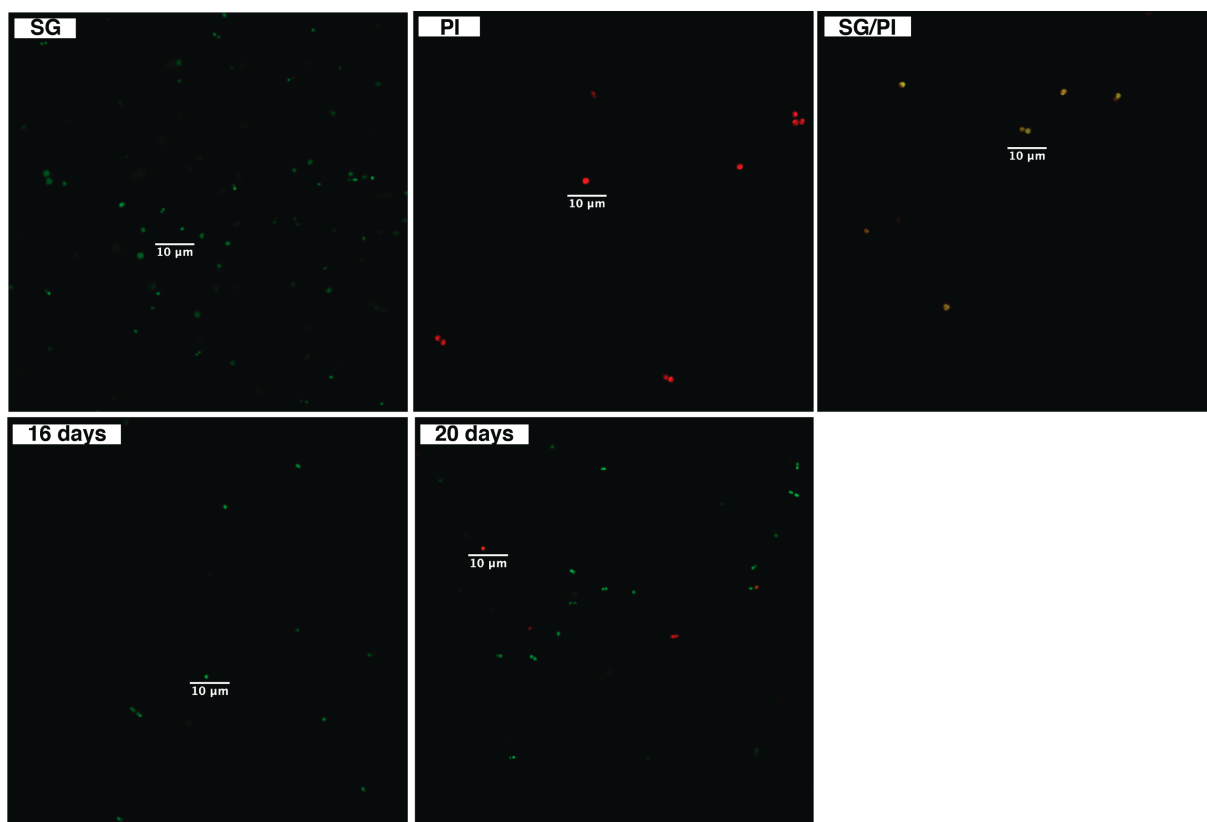
Figure S14. Fluorescence microscopy pictures of single-stained SG control (SG), single-stained PI control (PI), double-stained control (SG/PI), 16-days sample (16 days) and 20-day sample (20 days) of a *M. mazei* culture grown with 10 μM As(III) as initial concentration.
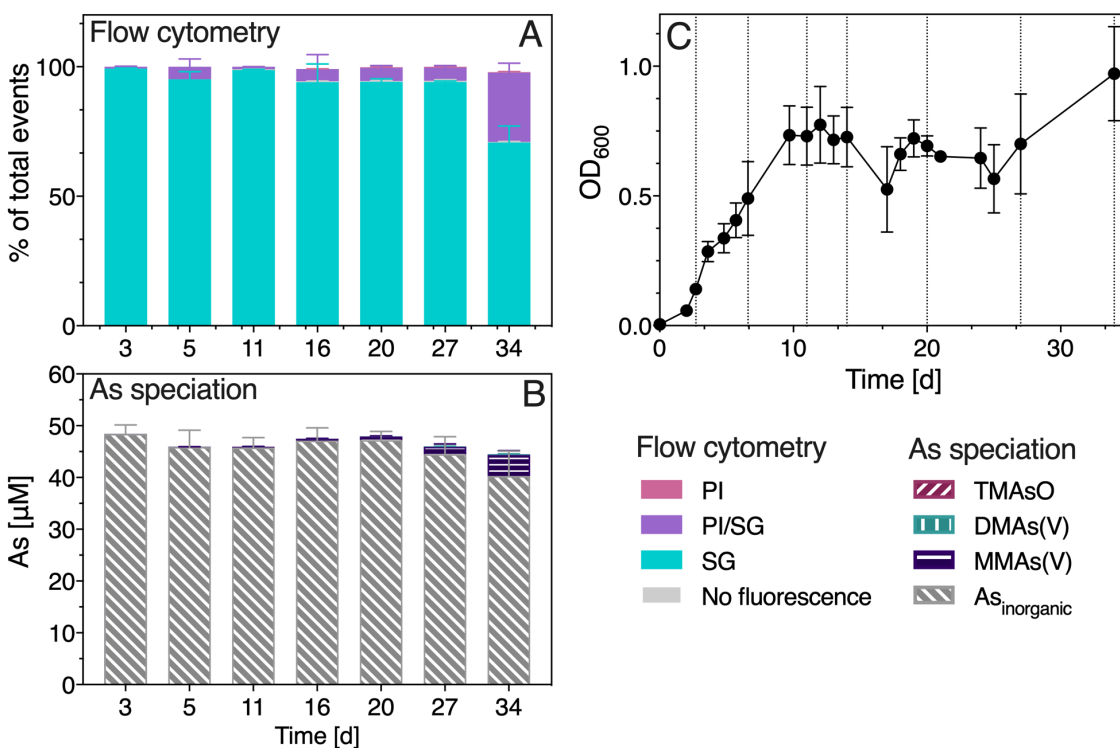
Figure S15. Assessment of membrane integrity for a *M. mazei* culture grown with 50 μM As(III) as the initial concentration: (A) Relative abundance of four populations (no fluorescence, double-stained (SG/PI) and single-stained: PI or SG) measured by flow cytometry after SG and PI staining of the cells at various time points along the growth curve; (B) concentration of soluble arsenic species in filtered medium; and (C) growth as OD_{600} with sampling points indicated as vertical lines. Bars in panels A and B represent the mean value and error bars plus one standard deviation. Individual values for each measurement and biological replicate are listed in Tables S14-S16.
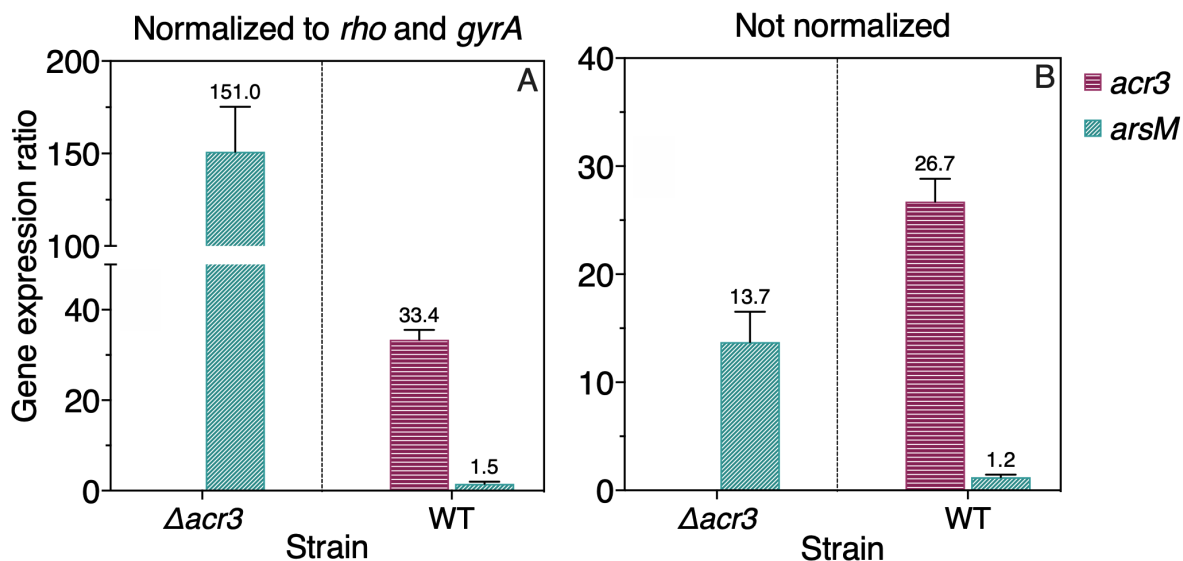
Figure S16. Expression of *arsM* and *acr3* in *C. pasteurianum* WT and Δ*acr3* mutant at mid-exponential phase. The gene expression ratio represents transcription in cultures grown with 100 μM As(III) relative to cultures grown without As(III) for each strain: (A) gene expression normalized to *gyrA* and *rho* expression and (B) gene expression without any normalization. Mid-exponential phase corresponds to 20 h for *C. pasteurianum* WT and no-As(III) control *C. pasteurianum* Δ*acr3*, and 37 h for *C. pasteurianum* Δ*acr3* grown with 100 μM As(III). Bars represent the mean value and error bars plus one standard deviation. Expression values are listed in Table S26.

```
ANN44297.1         --MQTDEQLKAIVREKYTAIAGQPAEQNAASC-CGSGPVSCCSPADVPIMAD-------D
WP_004455181.1     ----MKNNIKGQVKAYYGGIAKKVNAETKVTCGCGS---SCCGD-----DAAINSNL-YT
WP_041130008.1     -MSEQSTDLRETVRRRYAAAA----VQV---T-EGGT--ACCGPQAVEVDENFGSVL-YA
WP_047308040.1     -MPTDVQDVKDIVREKYAGAA----LKVA--T-GGGA--SCCGSSVLPGASPITSNL-YD
AIM18906.1         -----MDNIREGVRQKYAFAI----ANRGQGC-CGSP--GCCSDGLSDAADPITGNL-YD
WP_018703741.1     -----MNKIRQTVKKKYAEA-----IQSRTGCGANSS--SCCGGTGG--TSQITSNL-YS
WP_004511671.1     MDKQRNEEIRGAVRENYGKVA----VSGGAGCGCSSS--SCCGTPNGATTEDISLGLGYS
WP_011034171.1     ---MDANEKKEVIKKKYQEIA----ILGGS-C-CSG---GCCGDSN---STDLSRSLGYS
WP_011023683.1     ---MDAAEKKEVIKKKYQEIA----TLGGS-C-CSGG--GCCGDLS---AADLSRSLGYS
                            . .  :.   *                  ..   .**.

ANN44297.1         YTQLDGYVADA-DLALGCGLPTKFALIKEGDTVIDLGSGAGNDCFVARHETGPTGKVIGI
WP_004455181.1     KDYIEGLPEEAINASLGCANPIVLANPQKGEVVLDLGSGGGIDVFISSKYVGESGKVYGL
WP_041130008.1     ADERETLPAEAVAASLGCGNPTAVADLNEGERVLDLGSGGGIDVLLSARRVGPTGRAYGL
WP_047308040.1     AAQEQGLPAEAMLASLGCGNPTALAQLSPGEIVLDLGSGGGIDVLLSARRVGPTGKAYGL
AIM18906.1         ESDLQGLDPELIANSFGCGNPTALMNLNLGEVVLDLGSGSGLDVLLSAKRVGPTGKAYGL
WP_018703741.1     HLEVSELPEDILASSFGCGNPTALTELHEGETVLDLGSGAGLDVLLSAKRVGPTGKAYGL
WP_004511671.1     GEDVAAVPEGA-NLGLGCGNPQAIASLQLGETVLDLGSGGGFDCFLAARAVGHTGHVIGV
WP_011034171.1     ESDVQAVPDA--NMGLGCGNPTAFAELKPGDVVLDLGSGGGFDCFLAAQKIGSSGKVIGV
WP_011023683.1     EADVQAVPDA--NLGLGCGNPTAFAELKPGDIVLDLGSGAGFDSFLAAQRVGSLGKVIGV
                     .:**. *   .        *: *:*****.* * ::: .  *  *.. *:

ANN44297.1         DFTEAMIQRARTNAEVRGFNNVEFRQGDIEQMPVSDNVADVIVSNCVLNLVPNKKNVFAE
WP_004455181.1     DMTDEMLELANKNKDKMEVKNVEFIKGYIEDIPLKNESVDVITSNCVINLCESKEDALKE
WP_041130008.1     DMTEEMLALALANAARAGATNVEFLKGTIEAIPLPASTIDVVISNCVINLSTDKPAVFAE
WP_047308040.1     DMTDEMLALARDNQRKAGLDNVEFLKGEIEAIPLPDASVDVIISNCVINLSGDKDRVLRE
AIM18906.1         DMTDEMLAVAKENQRKSGIENAEFLKGHIEEIPLAAKSIDVIISNCVINLSGDKDKVLKE
WP_018703741.1     DMTDEMLEVAEKNRELSGLTNAEFLKGHIEEIPLPNSTVDVIISNCVINLSGDKDKVLKE
WP_004511671.1     DMTPEMITKSRRNADKANFGNVDFRLGELENLPVGIDVGIVDVIISNCVINLSPEKAKVFSE
WP_011034171.1     DMTLEMVEKAQANARKYGYSNVEFRHGDIESLPVKDSSVDVIINCVINLAPDKEKVFRE
WP_011023683.1     DMTQEMVKKAQDNARKYGYSNVEFRQGDIEALPLDDRSVDVIISNCVINLAPDKEKVFRE
                   *:*  *:  :  *       *.:*  * :* :*:    **: ****:**  .*  .: *

ANN44297.1         IFRVLKPGGHFSISDIVLLGDLPANLQQAAEMYAGCVSGAIQRADYLRLIDEAGFTAVSL
WP_004455181.1     AYRVLKNGGRLAIADVVVLKDIPEDIKKSVEMWVGCIAGALEVNEYKKILENVGFKDIEI
WP_041130008.1     TYRVLKPGGRIGVSDVVADDTLTPEQRAERGDHVGCIAGALSFTEYRNGLEAAGFTDIAI
WP_047308040.1     AFRVLKPGGRFAVSDVVTRGEIPDALRRDVLLWVGCLAGALDEADYVARLEAAGFAQISI
AIM18906.1         AYRVLKPQGRFAVSDIVIKRPLPEKIRDNILAWAGCIAGAMTEEEYRGKLSRAGFENISL
WP_018703741.1     AFRVLKPEGRFAVSDIVILRELPEKIRKNLLAWAGCISGAMTISEYQNKLLKAGFEKVDV
WP_004511671.1     SFRVLKSGGRLAISDVVATAEMPGDIKNNMAFHTGCIAGASSIEEIESMLERTGFVNIRI
WP_011034171.1     AFRVLKPEGRMYISDMVLLDELPEELKNDSELLAGCIAGAVLKEEYLGLLKKAGFSVEIL
WP_011023683.1     AFRVLKPGGRMYVSDMVLLEDLPEDLKNDCDLLAGCVAGALLKEEYLGLLKKAGFSFKIL
                    :****  *.: ::*:*    :.   .      .**::**   :    :  .**    :

ANN44297.1         QKEKAIDLPDDILQHYLSADEIAAFRDSGTGIQSITVYAEKPGGPQSGAAQSPKPKRQLE
WP_004455181.1     TPVNIYT-KEIIEDIAKQKNLEDVYSKIDSELLD-------------GAFAGAHVKAYKQ
WP_041130008.1     TPTHPVA--DGMHSAVVRAVKPAAGACTPSSDACCEVGTCCGGGVCCGAGACCAPAEGGP
WP_047308040.1     EPTRVYDIEDAREFLAGKGIDVDALAPKMQ----------------GKFFSGFVRATKP
AIM18906.1         QVTREYNLEDPSLRGMLEDLTDGEIKEFQ-----------------GAMVSCFIRAAKP
WP_018703741.1     EITQTYE-FSEIASDIYSGLTTDEQANLE-----------------NSIASAFIRASKP
WP_004511671.1     NPKTE---SRAFIRDWMPGSKIEDYV----------------------VSATIEAIKP
WP_011034171.1     NEDLDIS-KRQYRDLPVESLKLKARV----------------------------------
WP_011023683.1     AEDSDVS-KRQYEGLPVESLKLKAWV----------------------------------

ANN44297.1         SLAGTSAENCCTPGSGCC
WP_004455181.1     SSWH--------------
WP_041130008.1     SPASESASPR--------
WP_047308040.1     DGNSSAGCCI--------
AIM18906.1         A-----------------
WP_018703741.1     SSK---------------
```

Figure S17. Alignment of ArsM proteins from selected species vs. ArsM proteins from *Clostridium sp*. BXM (AIM18906.1) and *Rhodopseudomonas palustris* (WP_047308040.1).

Grayed out the cysteine residues equivalent to cysteines Cys44, Cys72, Cys174 and Cys224 in *Cyanidioschyzon* sp. 5508 [8]. (Multiple Sequence Alignment tool, EMBL-EBI, [1]).

References

(1)   Madeira, F.; Mi Park, Y.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47* (W1), W636–W641.

(2)   Grosse-Honebrink, A.; Schwarz, K. M.; Wang, H.; Minton, N. P.; Zhang, Y. Improving Gene Transfer In Clostridium Pasteurianum Through The Isolation Of Rare Hypertransformable Variants. *Anaerobe* **2017**, *48*, 203–205.

(3)   Grosse-Honebrink, A. Thesis: Forward and Reverse Genetics in Industrially Important Clostridia, University of Nottingham, 2017.

(4)   Schwarz, K. M.; Grosse-Honebrink, A.; Derecka, K.; Rotta, C.; Zhang, Y.; Minton, N. P. Towards Improved Butanol Production Through Targeted Genetic Modification of Clostridium Pasteurianum. *Metab. Eng.* **2017**, *40*, 124–137.

(5)   O'Brien, R. W.; Morris, J. G. Oxygen And The Growth And Metabolism of Clostridium Acetobutylicum. *J. Gen. Microbiol.* **2009**, *68* (3), 307–318.

(6)   Pyne, M. E.; Moo-Young, M.; Chung, D. A.; Chou, C. P. Development Of An Electrotransformation Protocol For Genetic Manipulation Of Clostridium Pasteurianum. *Biotechnol. Biofuels* **2013**, *6* (1).

(7)   Pfaffl, M. W.; Horgan, G. W.; Dempfle, L. Relative Expression Software Tool (REST) for Group-Wise Comparison and Statistical Analysis of Relative Expression Results in Real-Time PCR. *Nucleic Acids Res*. **2002**, *30* (9), e36.

(8)   Ajees, A. A.; Marapakala, K.; Packianathan, C.; Sankaran, B.; Rosen, B. P. Structure of an As(III) S-Adenosylmethionine Methyltransferase: Insights Into The Mechanism of Arsenic Biotransformation. *Biochemistry* **2012**, *51* (27), 5476–5485.