Revised version submitted Friday, July 10, 2021

Identification of the biosynthetic gene cluster for the organoarsenical antibiotic arsinothricin

Adriana E. Galván[1,orcid.org/0000-0002-2472-0438], Ngozi P. Paul[1,orcid.org/0000-0003-0423-195X], Jian Chen[1], Kunie Yoshinaga-Sakurai[1], Sagar M. Utturkar[2], Barry P. Rosen[1,orcid.org/0000-0002-5230-4271,] and Masafumi Yoshinaga[1*,orcid.org/0000-0002-7243-1761]

[1]*Department of Cellular Biology and Pharmacology, Florida International University, Herbert Wertheim College of Medicine, Miami, U.S.A.*

[2]*Purdue University Center for Cancer Research, Purdue University, West Lafayette, U.S.A.*

***Corresponding author**: Masafumi Yoshinaga (myoshina@fiu.edu), Department of Cellular Biology and Pharmacology, Florida International University, Herbert Wertheim College of Medicine, 11200 SW 8th Street, AHC1 419G, Miami, Florida, U. S. A. 33199, Tel: (+1) 305-348-1489, Fax: (+1) 305-348-0651

Abbreviations: AST, arsinothricin (2-amino-4-(hydroxymethylarsinoyl)butanoate); AST-OH, hydroxyarsinothricin (2-amino-4-(dihydroxyarsinoyl)butanoate); ACP, 3-amino-3-carboxypropyl; MAs(III), methylarsenite; MAs(V), methylarsenate; DMAs(III), dimethylarsenite; DMAs(V), dimethylarsenate

Keywords: Arsinothricin, organoarsenical antibiotic, biosynthetic gene cluster, *Burkholderia gladioli* GSRB05

ABSTRACT:

The soil bacterium *Burkholderia gladioli* GSRB05 produces the natural compound arsinothricin (2-amino-4-(hydroxymethylarsinoyl) butanoate (AST) that has been demonstrated to be a broad-spectrum antibiotic. To identify the genes responsible for AST biosynthesis, a draft genome sequence of *B. gladioli* GSRB05 was constructed. Three genes, *arsQML*, in an arsenic resistance operon, were found to be a biosynthetic gene cluster responsible for synthesis of AST and its precursor, hydroxyarsinothricin (2-amino-4-(dihydroxyarsinoyl) butanoate (AST-OH). The *arsL* gene product is a non-canonical radical S-adenosylmethionine (SAM) enzyme that is predicted to transfer the 3-amino-3-carboxypropyl (ACP) group from SAM to the arsenic atom in inorganic arsenite, forming AST-OH, which is methylated by the *arsM* gene product, a SAM methyltransferase, to produce AST. Finally, the *arsQ* gene product is an efflux permease that extrudes AST from the cells, a common final step in antibiotic-producing bacteria. Elucidation of the biosynthetic gene cluster for this novel arsenic-containing antibiotic adds an important new tool for continuation of the antibiotic era.

Importance

Antimicrobial resistance is an emerging global public health crisis, calling for urgent development of novel potent antibiotics. We propose that arsinothricin and related arsenic-containing compounds may be the progenitors of a new class of antibiotics to extend our antibiotic era. Here we report identification of the biosynthetic gene cluster for arsinothricin and demonstrate that only three genes, two of which are novel, are required for the biosynthesis and transport of arsinothricin, in contrast to the phosphonate counterpart, phosphinothricin, which requires over 20 genes. Our discoveries will provide insight for the development of more of effective organoarsenical antibiotics and illustrate the previously unknown complexity of the arsenic biogeochemical cycle, as well as bring new perspective to environmental arsenic biochemistry.

INTRODUCTION

Arsenic is one of the most ubiquitous environmental toxic substances. Bacteria have taken advantage of its prevalence to evolve mechanisms that give them competitive advantages over bacterial competitors. These include pathways for arsenic respiration for energy generation,[1-3] methylation of inorganic arsenic to increase its potency as an antimicrobial[4, 5] and even incorporation into arsenolipids for sparing phosphate under nutrient limiting conditions.[6, 7] A newly-recognized adaptation is the synthesis of a novel arsenic-containing antibiotic.

Antibiotic resistance is a global health challenge, and new antibiotics are urgently needed. Recently the rice rhizosphere bacterium *Burkholderia gladioli* GSRB05 was shown to produce the broad-spectrum arsenic-containing antibiotic arsinothricin (2-amino-4-hydroxymethylarsinoyl) butanoate, AST).[8] AST effectively inhibits growth of the World Health Organization (WHO) priority pathogens such as carbapenem-resistant *Enterobacter cloacae* (CRE) but has low cytotoxicity in human monocytes.[9] It also inhibits growth of *Mycobacterium bovis* BCG, which causes tuberculosis in animals and humans and is closely related to *M. tuberculosis* (MTB), the major causative agent of human tuberculosis (TB). Tuberculosis is classified by the WHO as a global health emergency[10] and has called for the development of new and innovative antibiotics against MTB (https://www.who.int/activities/tackling-the-drug-resistant-tb-crisis). AST is a pentavalent organoarsenical and a non-proteinogenic amino acid analog of glutamate. It inhibits glutamine synthetase (GS), most likely because of its chemical similarity to the acyl-phosphate intermediate, γ-glutamyl phosphate, in the GS catalytic cycle.[9] GS is essential for nitrogen metabolism in MTB, and inhibitors of MTB GS are actively sought after as drugs against TB.[11]

B. gladioli GSRB05 is the only known AST producer that has been shown to produce both pentavalent hydroxyarsinothricin (2-amino-4-(dihydroxyarsinoyl)butanoate, AST-OH) and AST from As(III), with a possible precursor-product relationship.[8] Here we identified the biosynthetic gene cluster (BGC) for AST production from the genome of *B. gladioli* GSRB05. Three genes, *arsQ*, *arsM* and *arsL*, are organized in an arsenic resistance (*ars*) operon. By

way of comparison, the phosphonate analog of AST is the *Streptomyces* antibiotic phosphinothricin, which is used commercially as the herbicide glufosinate has a very complicated biosynthetic pathway.[12] The BGC from *Streptomyces viridochromogenes* consists of 24 genes (PMC535184),[13] so a three-gene BGC for arsinothricin production is surprisingly small.

Identification of the AST BGC makes substantial contributions in two important areas, treatment of infectious diseases and radical SAM chemistry. First, the BGC is an uncomplicated pathway with only three steps required for the synthesis of AST. This arsenic-containing compound may be the founding member of a new class of antibiotics, adding to our arsenal of weapons against multidrug-resistant pathogens. Second, the radical SAM enzyme BgArsL catalyzes the key enzymatic reaction in AST biosynthesis. Radical SAM enzymes form the largest enzyme superfamily.[14] Most members catalyze the transfer of a 5′-deoxyadenosyl radical to the substrate or function as methyltransferases using a methylene fragment from SAM.[15] In contrast, BgArsL is a non-canonical radical SAM enzyme that transfers the 3-amino-3-carboxypropyl moiety to As(III), forming AST-OH, a unique radical SAM reaction.

METHODS AND MATERIALS

Strains, plasmid, media, and growth conditions

Strains and plasmids used in this study are described in Table S1. *B. gladioli* GSRB05 and *E. coli* cultures were grown aerobically overnight with shaking in lysogeny broth (LB) medium[16] at 30 °C or 37 °C, respectively. M9 medium[16] was supplemented with 0.2 % glucose, 0.1 mM $CaCl_2$ and 1 mM $MgSO_4$. For resistance assays antibiotics were supplemented at the following final concentration: 50 µg/ml streptomycin (Sm); 25 µg/ml chloramphenicol (Cm); 100 µg/ml ampicillin (Ap), as indicated.

AST production

The protocol for AST production was modified from the original method.[17] A single colony of *B. gladioli* GRSB05 was inoculated in 10 ml of LB medium and grown overnight. The culture

was 100-fold diluted into 1 l of fresh LB and grown again to an A_{600nm} of 1. Next, the cells were harvested and transferred into the same volume of ST 10^{-1} medium[18] supplemented with 20 µM As(III) and 0.2 % glucose. The culture was incubated until As(III) was completely transformed into AST. AST was chromatographically purified as described[8] with a yield of approximately 5 mg AST/l of culture medium and was free of AST-OH.

The time course of As(III) conversion into AST-OH and AST by *B. gladioli* GSRB05 was performed in a 50 ml culture of ST10^{-1} medium containing 3 µM As(III). The same procedure was used to determine As(III) transformation by *E. coli* Top10 bearing vector plasmids pUC118, pUC*arsL-orf1-4* or pUC*arsML-orf1-4* as for *B. gladioli* GRSB05, but for *E. coli* BL21 bearing plasmids pETDuet-1, pETDuet-1*arsL*, pETDuet-1*arsM* or pETDuet-1*arsLM*, the pellet was directly transferred from the overnight LB culture into ST10^{-1} medium supplemented with 1 µM As(III) and 0.4 % glycerol, following which cultures were incubated for 36 h. For both *E. coli* Top10 and BL21 basal expression of genes in plasmids pUC118 and pETDuet-1 was sufficient to observe the activity of the genes.

Analysis of arsenic species

Arsenic species in the supernatant were analyzed by high-performance liquid chromatography (HPLC) inductively coupled plasma mass spectrometry (ICP-MS). For sample preparation, 0.5 ml of culture was collected and centrifuged at 13,000 rpm for 2 m at 4 °C. Supernatants were filtered with 3 kDa Amicon Ultra Centrifugal Filters (MilliporeSigma, St. Louis, MO, USA) for 10 m. A mobile phase of 3 mM malonic acid, 5% of methanol, and tetrabutylammonium hydroxide to reach a pH of 5.9 was used to elute a C18 reverse-phase column at a flow rate of 1.0 ml m^{-1}.

Construction of the draft genome sequence of *B. gladioli* GSRB05

Genome sequencing of *B. gladioli* GSRB05 was performed using the Illumina NextSeq platform at Center for Genome Technology, University of Miami, Miller School of Medicine (Miami, FL, USA). Sequence data was comprised of 7.8 million paired-end (2 X 150) reads. Quality trimming and filtering were performed using the TrimGalore (version 0.6.4) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) tool to remove adapter

sequences, base pairs with quality score <30 and reads shorter than 50 bp. Quality trimmed reads were assembled using SPAdes (version 3.13.0)[19] and ABySS (version 2.1.5)[20] at different KMER, and optimal assembly was selected as described previously.[21] Genome annotation was accomplished using the NCBI Prokaryotic Genome Annotation Pipeline,[22] and predictions of *arsM* orthologous genes were performed via BLAST[23] and OrthoMCL[24] sequence analysis against the predicted proteins.

Cloning and expression

Genomic DNA was extracted from 3 ml of a fresh overnight culture of *B. gladioli* GSRB05 using the E.Z.N.A.® Bacterial DNA Kit (Omega Bio-tek Inc., Norcross, GA, USA) The gene cluster was sequentially amplified by polymerase chain reaction (PCR) with *PfuTurbo* high-fidelity DNA polymerase (Agilent Technologies Inc., Santa Clara, CA, USA), and the entire cluster or groups of genes were cloned into plasmid pUC118 between the *Kpn*I and *Xba*I restriction sites. To construct pUC*arsL-orf1-4* the cloning was carried out in two steps, while for pUC*arsML-orf1-4* an additional step was necessary (Figure S1) to amplify fragments of no more than 3.2 kb to avoiding incorrect base insertion. Primers were designed with unique restriction sites to serially construct the final plasmids. The amplified products were gel purified, digested with the appropriate restriction enzymes and inserted into vector plasmid pUC118 with the first gene in frame with the *lacZα* gene of the vector. The *arsL* and *arsML* genes were ordered from GenScript® (GenScript, Piscataway, NJ, USA) and cloned into pETDuet-1 (MilliporeSigma) and *arsQ* in pTrcHisA (Thermo Fisher Scientific Inc., Waltham, MA, USA) to construct pETDuet-1*arsL*, pETDuet-1*arsML* and pTrc*arsQ*, respectively. Each step of cloning was verified by sequencing the fragments. The complete list of the oligonucleotides used in this study is given in Table S1.

Site-directed mutagenesis

The primers for site-directed mutagenesis (Table S2) were designed using the online QuikChange® Primer Design Program (https://www.agilent.com/store/primerDesignProgram.jsp). A stop codon was individually inserted into each of the five genes in plasmid pUC*arsL-orf1-4* by substitution of one

nucleotide at the beginning of the sequence. Unmutated plasmid pUC*arsL-orf1-4* was removed from the reaction by digesting the methylated DNA with restriction enzyme *Dpn*I (New England Biolabs, Ipswich, MA, USA). Each mutated plasmid was transformed into competent cells of *E. coli* TOP10, purified, and the presence of the mutation verified by sequencing.

mRNA extraction, reverse transcription and quantitative real-time PCR

An RNeasy Mini Kit (QIAGEN, Valencia, CA, USA) was used to isolate total RNA from a 3 ml culture of *B. gladioli* GSRB05 that had been cultured with or without exposure to 3 μM As(III) in ST 10^{-1} medium for 13 h. The purity and concentrations of RNA were determined from the A_{260nm} using a Synergy™ H4 Hybrid Microplate Reader (BioTek Instruments, Inc. Winooski, VT, USA). RNA integrity was verified by electrophoresis (data not shown). Reverse transcription-polymerase chain reaction (RT-PCR) was performed to synthesize complementary DNA (cDNA) using a Verso cDNA Synthesis Kit (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. Primer sets for real-time qPCR of target genes are listed in Table S2. 1 μl of each of purified RT-PCR product corresponding to 50 ng of total RNA was amplified in a 10 μl reaction mixture containing 0.5 μM of each primer set and 5 μl of iQSYBR Green supermix (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Real-time qPCR assays were carried out using a Realplex2 PCR instrument (Eppendorf, Hamburg, Germany) with the following cycle steps: initial denaturation was carried out for 2 m at 94 °C, followed by 40 cycles of 15 s at 94 °C for denaturation, then 30 s at 50 °C for annealing and 1 m 30 s at 72 °C for extension of *arsQ*, *arsM and arsL.* For fragments containing sequences from *arsM* to *orf1* (*arsM – orf1*) or from *arsL* to *orf2* (*arsL – orf2*), the reaction condition was changed to 30 s at 58 °C for annealing and 3 m 10 s at 72 °C for extension. All data were normalized to the amount of 16S rRNA. $2^{-\Delta Ct}$ was calculated to compare the expression level of each gene.

Assays of *arsQ* function

To examine whether the *arsQ* gene could confer AST resistance, cells of *E. coli* AW3110 harboring plasmid pTrcHis2A*arsQ* were grown overnight in LB medium, washed and

suspended in 0.9 % NaCl. The washed cells were then inoculated into M9 medium at A_{600nm} of 0.03 and incubated for 24 h in the presence of the indicated concentrations of AST.

To determine the function of ArsQ in AST transport, the arsQ gene was cloned into vector plasmid pTrcHisA. Cells of *E. coli* AW3110 expressing plasmid pTrcHisA*arsQ* or vector plasmid pTrcHisA only were grown to A_{600nm} = 2 at 37 °C with aeration in LB medium. The cells were harvested and suspended in buffer A (75 mM HEPES-KOH, pH 7.5, 0.15 M KCl and 1 mM $MgSO_4$) at A_{600nm} = 4. To initiate the transport reaction, 20 μM AST was added to 1 ml of cell suspension. Portions (0.1 ml) from the cell suspension were withdrawn at the indicated times, filtered through nitrocellulose filters (0.2 μm pore diameter; EMD Millipore, Billerica, MA) and washed twice at room temperature with 5 ml of buffer A. The filters were digested with 0.3 ml of concentrated HNO_3 (68–70%) overnight at room temperature. The dissolved filters were incubated for 10 min at 70 °C, allowed to cool to room temperature and diluted with HPLC-grade water (MilliporeSigma) to produce a final HNO_3 concentration of 2%. Arsenic was quantified by ICP-MS. Standard solutions were made in the range of 1–50 ppb in 2% nitric acid using arsenic standards (Ultra Scientific, N. Kingstown, RI, USA). Protein content was determined using a Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific Inc.).

Everted membrane vesicles and transport assays were performed as described.[25] Transport assays were performed in buffer A containing 0.25 M sucrose. The reaction mixture contained 1 mg/ml membrane proteins, 40 μM AST and 5 mM NADH in a final volume of 0.6 ml of the same buffer. Portions (0.1 ml) were withdrawn at the indicated times, filtered through 0.2 μm pore size nitrocellulose filters and washed twice with 5 ml of the same buffer. Arsenic content was determined by ICP-MS.

Data availability

The draft genome sequence for *B. gladioli* GSRB05 is deposited at NCBI under the accession number JAGSIB000000000. Raw sequence reads are deposited at NCBI under the bioproject accession number PRJNA722678.

AST BGC distribution

The prevalence of AST BGC was analyzed in representative organisms. GenBank accession numbers of the following bacterial genomes are given in parentheses. *B. gladioli* GSBR05 (JAGSIB000000000) is compared with putative orthologs from *Burkholderia oklahomensis* (NZ_UFUH01000001), *Burkholderia cepacia* (NZ_CADEUO010000007), *Pseudomonas aeruginosa* (NZ_CACPET010000007), *Pseudomonas fluorescens* (NZ_LVEJ01000018), *Pseudomonas amygdali* (NZ_LGLI01000031), *Actibacterium* sp. (NZ_JAFEUL010000009) and *Rhodobacter* sp. (NVUP01000011). Multiple alignments of the sequences of ArsM (Figure S3), ArsL (Figure S4) and ArsQ (Figure S5) orthologs were performed using T-Coffee[26] and the BoxShade version 3.21 server (https://embnet.vital-it.ch/software/BOX_form.html)

RESULTS

Identification of the AST biosynthetic gene cluster

In this study, a time course of AST biosynthesis by *B. gladioli* GSRB05 was conducted in cells grown in ST 10^{-1} medium (Figure 1). After a lag period of approximately 10 h, both trivalent and pentavalent AST-OH and pentavalent AST were produced. After approximately 15 h, all of the As(III) was consumed, the trivalent AST-OH peak decreases, and both the AST and the pentavalent AST-OH peaks increase correspondingly. We interpret this result as As(III) conversion into trivalent AST-OH in the first step, which is then either methylated to AST in a second step or oxidized to pentavalent AST-OH. After 24 h, both the pentavalent and trivalent AST-OH peaks decrease, and the AST peak increases resiprocally, suggesting that the strain is able to reduce pentavalent AST-OH to the trivalent form to produce more AST. It is not clear if the product is trivalent or pentavalent AST because in our hands trivalent AST oxidizes in air too rapidly to be isolated. Trivalent AST-OH also oxidizes in air but less rapidly than AST.

To identify the AST BGC knowing that methylation is involved, we made the assumption that a *B. gladioli* GSRB05 enzyme would be related to other known methylating enzymes. Arsenic methylation catalyzed by the ArsM As(III) *S*-adenosylmethionine methyltransferase is a common reaction in arsenic metabolism.[27] Bacterial and algal ArsM enzymes methylate inorganic As(III) up to three times to produce methylarsenite (MAs(III)),

dimethylarsenite (DMAs(III)) and trimethylarsenite (TMAs(III)). These are rapidly oxidized non-enzymatically in air to the pentavalent species, which are neither substrates nor products of the enzyme-catalyzed reaction. We predicted that a *B. gladioli* GSRB05 ArsM ortholog is involved in the methylation of trivalent AST-OH, and that its gene would be in the AST BGC. For that reason, a draft sequence of the *B. gladioli* GSRB05 genome was constructed (Figure S2). As anticipated, an *arsM* sequence was identified next to a cluster of genes related to arsenic metabolism but divergently oriented (Figure 2A). These genes located on the opposite DNA strand are *arsR*,[28] *pitA*,[29] *aioAB*,[30] encoding a putative ArsR As(III)-responsive transcriptional repressor, PitA, a low-affinity inorganic phosphate/arsenate transporter, and the AioAB arsenite oxidase that oxidizes As(III) to As(V), respectively. The predicted products of the genes upstream and downstream of *arsM* were searched against the Protein Data Bank (PDB; http://blast.ncbi.nlm.nih.gov) using the BLASTP program. The gene immediately upstream of *arsM* is termed *arsQ* and encodes putative membrane transporter annotated as related to the GntP family of gluconate permeases.[31] The gene immediately downstream of *arsM* encodes a putative radical SAM protein and is termed *arsL.* The *arsQML* cluster is found in the genomes of other Proteobacterial genomes but is not widespread in the bacterial kingdom (Figure 2B). Their association suggests that these three genes have a related function. The next four genes were termed *orf1-4* and are annotated as encoding two class I SAM-dependent methyltransferases, a cytochrome P450-like protein and an α/β hydrolase. The four genes are not adjacent to *arsQML* in other bacterial genomes, and, as described below, these four genes appear to be unrelated to AST biosynthesis.

The *arsM* gene product (BgArsM) is a 378 amino acid residue protein with a predicted mass of 41.71 kDa (accession number BurGSRB05_34905). Most ArsM and related animal AS3MT enzymes have four conserved cysteine residues that are required for catalysis.[32, 33] Multiple sequence alignment of BgArsM and orthologs shows that BgArsM has four cysteine residues at positions 28, 54, 181 and 233 (Figure S3). The N-terminus of BgArsM does not align well with other ArsMs, but Cys28 is in approximately the position expected for the first conserved cysteine, and the other three align well with the remaining three conserved cysteine

residues. The *arsL* gene product (BgArsL) is a 428 amino acid residue protein with a predicted mass of 47.5 kDa (accession number BurGSRB05_34900). Multiple sequence alignment with other putative ArsL orthologs shows that BgArsL has three conserved cysteine residues at positions 194, 198 and 201 (Figure S4). It is annotated as a radical SAM enzyme with these three cysteine residues forming a CX_3CX_2C motif that is found in more than 90% of members of the radical SAM superfamily. In the cyanobacterium *Synechocystis* sp. PCC 6803, a radical SAM enzyme, SsArsS, has been shown to function with SsArsM to catalyze the initial steps in arsenosugar biosynthesis.[34] It was therefore reasonable to consider that BgArsM and BgArsL might function together in AST biosynthesis. The *arsQML* genes plus the four downstream *orfs* were cloned and transformed into *E. coli* Top10. However, the transformants grew poorly. Considering the heterologous expression of membrane proteins, in general, might be toxic for the host cells, *arsQ* was eliminated from the construct, leaving *arsML* and the four *orfs*. Cells of *E. coli* Top10 with the remaining six genes grew well, indicating that *arsQ* was responsible for growth retardation. With the addition of 1μM As(III), cells expressing the six genes produced AST and smaller amounts of AST-OH (Figure 3A). Cells expressing an *arsL-orf1-4* construct lacking *arsM* produced only AST-OH and not AST. These results indicate that one or more of the five genes downstream of *arsM* are involved in AST-OH biosynthesis, followed by methylation to AST by BgArsM.

BgArsM and BgArsL are both required for AST biosynthesis

To determine which gene(s) is/are required for AST-OH production, a mutation producing a stop codon was introduced into *arsL* and each of the other four genes in the *arsL-orf1-4* construct. Cells expressing each mutant were incubated with As(III), and the culture medium was analyzed by HPLC-ICP-MS. Stop codons in the downstream *orf1-4* genes had no effect, and only the point mutation in the *arsL* gene lost the ability to produce AST-OH (Figure 3B). Thus only *arsL* is required for AST-OH production. These results strongly suggest that the biosynthetic pathway of AST is composed of only two reactions catalyzed by BgArsM and

BgArsL. To confirm this hypothesis we expressed *arsL* alone and the *arsML* genes together in the pETDuet-1 system (Figure 4). Cells of *E. coli* BL21 expressing *arsL* alone produced AST-OH but not AST. In contrast, AST production was observed when *arsL* and *arsM* were co-expressed.

The *arsQML* genes comprise an As(III)-inducible *ars* operon

To examine whether the *arsQ, arsL* and *arsM* genes comprise an As(III)-responsive *ars* operon, total RNA was obtained from cells of *B. gladioli* GSRB05 with or without As(III), and cDNA was subsequently synthesized by reverse transcription polymerase chain reaction (RT-PCR). From the amount of RNA detected in quantitative real-time PCR (RT-qPCR) analysis, *arsQ*, *arsM* and *arsL* increased approximately 8.5-, 7.2- and 11.9-fold, respectively, following induction with 3 μM As(III) compared with RNA from uninduced cells (Figure 5). The results demonstrate that *arsQML* genes comprise an *ars* operon. As(III) responsiveness suggests that the operon is controlled by an ArsR repressor, likely the product of the upstream *arsR* gene. When the reverse primers were designed to anneal a cDNA region including either *orf1* or *orf2,* almost no amplification was observed, indicating that the downstream four *orfs* are not part of the *arsQML* operon.

ArsQ is an AST efflux permease

Antibiotic BGCs frequently have genes for efflux of the antibiotic. This serves dual purposes of removing the active antibiotic from the cytosol, thus protecting the producing organism, and exporting the compound into the medium, where it can exert its antibiotic action against other bacteria. The *arsQ* gene encodes a putative membrane protein of 408 amino acid residues with a predicted mass of 42.9 kDa (accession number BurGSRB05_34910). ArsQ orthologs are found primarily in Proteobacteria and are annotated as gluconate permeases. AST and gluconate are similar in size and charge, so a reasonable inference is that the BgArsQ functions as an AST efflux permease. A multiple sequence alignment with the 100 most closely related proteins shows variation in the N-terminal 90 residues, but the remaining 318 residues

329 are highly conserved (Figure S5). The *arsQ* gene was cloned into vector pTrcHisA and
330 transformed into the As(III) hypersensitive strain of *E. coli* AW3110.[35]

331 Addition of the inducer isopropyl β-D-1-thiogalactopyranoside produced growth
332 inhibition, so the cells were grown without inducer, and the effect of leaky expression of *arsQ*
333 on AST inhibition assayed (Figure 6A). Cells expressing *arsQ* were clearly more resistant to
334 AST than those with vector only. The resistance was not dramatic, probably due to low levels
335 of expression of BgArsQ. Next, the effect of *arsQ* expression on accumulation of AST was
336 examined. Cells expressing BgArsQ accumulated significantly less AST than the control
337 (Figure 6B). Uptake into everted vesicles reflects effux from cells,[36, 37] so everted membrane
338 vesicles were prepared from those cells, and energy-driven uptake of AST into those vesicles
339 was assayed. Cells expressing BgAST took up significantly more AST than the control (Figure
340 6C). These results demonstrate that expression *arsQ* increases resistance by active efflux of
341 AST.

342

343 **DISCUSSION**

344 Identification of the biosynthetic gene cluster for arsinothricin is an essential step in elucidation
345 of the antimicrobial action of this novel arsenic-containing antibiotic. With only three genes,
346 the BGC is surprisingly small, especially compared with the BGC for the phosphonate
347 counterpart, phosphinothricin, which includes 24 genes. The AST BGC has a rather narrow
348 phylogenetic distribution. The *arsQML* gene cluster is found primarily in members of the
349 *Proteobacteria* phylum such as Alpha class (*Rhodobacter*), Beta class (*Burkholderia*) and
350 Gamma class (*Pseudomonas*). Most BgArsQ orthologs are found in Proteobacteria, while
351 orthologs of BgArsM and BgArsL are more widely distributed. However, it is not clear if those
352 are involved in AST biosynthesis. Near other *arsL* genes are also genes for other putative
353 permeases that are unrelated to ArsQ. It is not known if their substrates include AST, but
354 alternate AST transporters would not be surprising. This is reminiscent of the existence of
355 multiple permeases for inorganic and organic arsenicals such as ArsB, Acr3, ArsP, ArsK and
356 ArsJ.[38] We propose that AST is synthesized in a three-step pathway: 1) AST-OH synthesis by

BgArsL, 2) methylation of AST-OH to AST by BgArsM, and 3) export of AST by BgArsQ (Figure 7).

BgArsL can be predicted to be a radical SAM enzyme, a member of the largest enzyme superfamily. It has three cysteine residues, Cys194, Cys198 and Cys201 that are highly conserved in radical SAM enzymes and form a $[4Fe\text{-}4S]^{+}$ cluster for reductive cleavage of SAM. Since BgArsL has only three conserved cysteine residues, it likely has a single $[4Fe\text{-}4S]^{+}$ cluster. The majority of radical SAM enzymes produce a 5′-deoxyadenosyl radical that perform a wide variety of organic chemical biotransformations.[15] AST-OH is an amino acid with arsenate replacing the γ-carboxyl group of glutamic acid, so it is reasonable to propose that BgArsL catalyzes addition of the 3-amino-3-carboxypropyl (ACP) group of SAM to As(III), forming trivalent AST-OH, as shown in Figure 7. In this respect, BgArsL is a novel non-canonical radical SAM enzyme that forms an ACP radical rather than the more common 5′-deoxyadenosyl radical. This reaction would produce trivalent AST-OH. The only other non-canonical radical SAM enzyme that transfers ACP is the diphthamide biosynthetic enzyme Dph2, which adds ACP to a histidine residue in translation elongation factor 2.[39] Rather than catalyzing protein modification like Dph2, ArsL catalyzes the unique chemistry of C-As bond formation. Enzymes with $[4Fe\text{-}4S]^{+}$ clusters are typically very oxygen sensitive, usually requiring anaerobic purification and/or reconstitution of the iron-sulfur cofactor.[40] Future directions of study with BgArsL will be designed for the challenging characterization of the enzyme *in vitro.*

BgArsM is a typical As(III) SAM methyltransferase. ArsM substrates include inorganic arsenite, methylarsenicals and aromatic arsenicals. A mechanistic feature of this group of arsenic methyltransferases is that both the substrates and products are trivalent, and the pentavalent species are produced nonenzymatically by oxidation, usually with atmospheric O_2.[27] Cells of *E. coli* heterologously expressing the *arsL* gene transform As(III) into AST-OH and cells expressing both *arsL* and *arsM* genes prodcuce ATS. Thus, we propose that in the second step, trivalent AST-OH is methylated by BgArsM to form the reduced form of AST, which would rapidly oxidize, yielding the pentavalent antibiotic AST. Future tests of this

proposal will involve purification of BgArsM for enzymatic assays and development of analytical methods to detect the trivalent form of AST.

Antibiotic producers frequently have transporters that transport the active antibiotic from the cells, both for self-protection and to inhibit growth of competitors.[41] In the third and final step, AST is exported from the cells by the efflux permease BgArsQ. The results of transport assays with cells of *E. coli* heterologously expressing *arsQ* clearly demonstrate that BgArsQ reduces the intracellular concentration of AST, and in complementary assays with everted membrane vesicles, there is energy-dependent uptake of AST. It is not known whether BgArsQ substrates are the trivalent, pentavalent or both, or whether AST-OH and AST are both substrates, or even if gluconate is a substrate, due to apparent toxicity caused by *arsQ* expression. Future experiments will be designed to produce sufficient BgArsQ for biochemical characterization.

ACKNOWLEDGMENTS: This work was supported by NSF BIO/MCB grant 1817962 to M.Y. and NIH grants R35GM136211 and R01GM55425 and R01 ES023779 to B.P.R. and the National Natural Science Foundation of China grant 41967023 to J.C. The authors state that they have no competing interests.

REFERENCES

1. Krafft, T.; Macy, J. M., Purification and characterization of the respiratory arsenate reductase of Chrysiogenes arsenatis. *Eur. J. Biochem.* **1998,** *255*, (3), 647-53.

2. Ellis, P. J.; Conrads, T.; Hille, R.; Kuhn, P., Crystal Structure of the 100 kDa Arsenite Oxidase from Alcaligenes faecalis in Two Crystal Forms at 1.64 Å and 2.03 Å. *Structure* **2001,** *9*, (2), 125-132.

3. Zargar, K.; Hoeft, S.; Oremland, R.; Saltikov, C. W., Identification of a Novel Arsenite Oxidase Gene, arxA, in the Haloalkaliphilic, Arsenite-Oxidizing Bacterium Alkalilimnicola ehrlichii Strain MLHE-1. *Journal of Bacteriology* **2010,** *192*, (14), 3755-3762.

4. Qin, J.; Rosen, B. P.; Zhang, Y.; Wang, G.; Franke, S.; Rensing, C., Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, (7), 2075-80.

5. Chen, J.; Yoshinaga, M.; Rosen, B. P., The antibiotic action of methylarsenite is an emergent property of microbial communities. *Mol Microbiol* **2019,** *111*, (2), 487-494.

6. Devalla, S.; Feldmann, J., Determination of lipid-soluble arsenic species in seaweed-eating sheep from Orkney. *Applied Organometallic Chemistry* **2003,** *17*, (12), 906-912.

7. Amayo, K. O.; Petursdottir, A.; Newcombe, C.; Gunnlaugsdottir, H.; Raab, A.; Krupp, E. M.; Feldmann, J. R., Identification and Quantification of Arsenolipids Using Reversed-Phase HPLC Coupled Simultaneously to High-Resolution ICPMS and High-Resolution Electrospray MS without Species-Specific Standards. *Analytical Chemistry* **2011,** *83*, (9), 3589-3595.

8. Kuramata, M.; Sakakibara, F.; Kataoka, R.; Yamazaki, K.; Baba, K.; Ishizaka, M.; Hiradate, S.; Kamo, T.; Ishikawa, S., Arsinothricin, a novel organoarsenic species produced by a rice rhizosphere bacterium. *Environ Chem* **2016,** *13*, 723-731.

9. Nadar, V. S.; Chen, J.; Dheeman, D. S.; Galván, A. E.; Yoshinaga-Sakurai, K.; Kandavelu, P.; Sankaran, B.; Kuramata, M.; Ishikawa, S.; Rosen, B. P.; Yoshinaga, M., Arsinothricin, an arsenic-containing non-proteinogenic amino acid analog of glutamate, is a broad-spectrum antibiotic. *Commun Biol* **2019,** *2*, 131.

10. Gygli, S. M.; Borrell, S.; Trauner, A.; Gagneux, S., Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. *FEMS Microbiology Reviews* **2017,** *41*, (3), 354-373.

11. Berlicki, Ł.; Obojska, A.; Forlani, G.; Kafarski, P., Design, Synthesis, and Activity of Analogues of Phosphinothricin as Inhibitors of Glutamine Synthetase. *Journal of Medicinal Chemistry* **2005,** *48*, (20), 6340-6349.

12. Schinko, E.; Schad, K.; Eys, S.; Keller, U.; Wohlleben, W., Phosphinothricin-tripeptide biosynthesis: An original version of bacterial secondary metabolism? *Phytochemistry* **2009,** *70*, (15-16), 1787-1800.

13. Blodgett, J. A.; Zhang, J. K.; Yu, X.; Metcalf, W. W., Conserved biosynthetic pathways for phosalacine, bialaphos and newly discovered phosphonic acid natural products. *The Journal of Antibiotics* **2016,** *69*, (1), 15-25.

14. Wang, S. C.; Frey, P. A., S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem Sci* **2007,** *32*, (3), 101-10.

15. Roach, P. L., Radicals from S-adenosylmethionine and their application to biosynthesis. *Curr Opin Chem Biol* **2011,** *15*, (2), 267-75.

16. Sambrook, J.; Fritsch, E. F.; Maniatis, T., *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory: New York., 1989.

17. Kuramata, M.; Sakakibara, F.; Kataoka, R.; Yamazaki, K.; Baba, K.; Ishizaka, M.; Hiradate, S.; Kamo, T.; Ishikawa, S., Arsinothricin, a novel organoarsenic species produced by a rice rhizosphere bacterium. *Environmental Chemistry* **2016,** *13*, (4), 723.

18. Maki, T.; Hasegawa, H.; Watarai, H.; Ueda, K., Classification for dimethylarsenate-decomposing bacteria using a restrict fragment length polymorphism analysis of 16S rRNA genes. *Anal Sci* **2004,** *20*, (1), 61-8.

19. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A., SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **2012,** *19*, (5), 455-477.

20. Simpson, J. T.; Wong, K.; Jackman, S. D.; Schein, J. E.; Jones, S. J. M.; Birol, I., ABySS: A parallel assembler for short read sequence data. *Genome Research* **2009,** *19*, (6), 1117-1123.

21. Utturkar, S. M.; Klingeman, D. M.; Land, M. L.; Schadt, C. W.; Doktycz, M. J.; Pelletier, D. A.; Brown, S. D., Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics* **2014,** *30*, (19), 2709-2716.

22. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E. P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K. D.; Borodovsky, M.; Ostell, J., NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* **2016,** *44*, (14), 6614-24.

23. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L., BLAST+: architecture and applications. *BMC Bioinformatics* **2009,** *10*, (1), 421.

24. Li, L., OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research* **2003,** *13*, (9), 2178-2189.

25. Villadangos, A. F.; Fu, H. L.; Gil, J. A.; Messens, J.; Rosen, B. P.; Mateos, L. M., Efflux permease CgAcr3-1 of *Corynebacterium glutamicum* is an arsenite-specific antiporter. *J. Biol. Chem.* **2012,** *287*, (1), 723-35.

26. Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000,** *302*, (1), 205-17.

27. Chen, J.; Rosen, B. P., The Arsenic Methylation Cycle: How Microbial Communities Adapted Methylarsenicals for Use as Weapons in the Continuing War for Dominance. *Frontiers in Environmental Science* **2020,** *8*.

28. Wu, J.; Rosen, B. P., The ArsR protein is a trans-acting regulatory protein. *Mol. Microbiol.* **1991,** *5*, (6), 1331-6.

29. Willsky, G. R.; Malamy, M. H., Characterization of two genetically separable inorganic phosphate transport systems in *Escherichia coli*. *J. Bacteriol.* **1980,** *144*, (1), 356-65.

30. Ellis, P. J.; Conrads, T.; Hille, R.; Kuhn, P., Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 A and 2.03 A. *Structure* **2001,** *9*, (2), 125-32.

31. Klemm, P.; Tong, S.; Nielsen, H.; Conway, T., The gntP gene of Escherichia coli involved in gluconate uptake. *Journal of bacteriology* **1996,** *178*, (1), 61-67.

491 32. Dheeman, D. S.; Packianathan, C.; Pillai, J. K.; Rosen, B. P., Pathway of human
492 AS3MT arsenic methylation. *Chem Res Toxicol* **2014,** *27*, (11), 1979-89.
493 33. Marapakala, K.; Packianathan, C.; Ajees, A. A.; Dheeman, D. S.; Sankaran, B.;
494 Kandavelu, P.; Rosen, B. P., A disulfide-bond cascade mechanism for As(III) S-
495 adenosylmethionine methyltransferase. *Acta Crystallogr. D. Biol. Crystallogr.* **2015,** *71*, (Pt 3),
496 505-15.
497 34. Xue, X. M.; Ye, J.; Raber, G.; Rosen, B. P.; Francesconi, K.; Xiong, C.; Zhu, Z.;
498 Rensing, C.; Zhu, Y. G., Identification of Steps in the Pathway of Arsenosugar Biosynthesis.
499 *Environ Sci Technol* **2019,** *53*, (2), 634-641.
500 35. Carlin, A.; Shi, W.; Dey, S.; Rosen, B. P., The ars operon of Escherichia coli confers
501 arsenical and antimonial resistance. *J Bacteriol* **1995,** *177*, (4), 981-6.
502 36. Rosen, B. P.; Tsuchiya, T., Preparation of everted membrane vesicles from
503 *Escherichia coli* for the measurement of calcium transport. *Methods Enzymol.* **1979,** *56*, 233-
504 41.
505 37. Chen, J.; Madegowda, M.; Bhattacharjee, H.; Rosen, B. P., ArsP: a methylarsenite
506 efflux permease. *Mol. Microbiol.* **2015,** *98*, (4), 625-635.
507 38. Garbinski, L. D.; Rosen, B. P.; Chen, J., Pathways of arsenic uptake and efflux. *Environ*
508 *Int* **2019,** *126*, 585-597.
509 39. Zhang, Y.; Zhu, X.; Torelli, A. T.; Lee, M.; Dzikovski, B.; Koralewski, R. M.; Wang, E.;
510 Freed, J.; Krebs, C.; Ealick, S. E.; Lin, H., Diphthamide biosynthesis requires an organic
511 radical generated by an iron–sulphur enzyme. *Nature* **2010,** *465*, (7300), 891-896.
512 40. Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M., Radical S-
513 adenosylmethionine enzymes. *Chem Rev* **2014,** *114*, (8), 4229-317.
514 41. Li, X.-Z.; Plésiat, P.; Nikaido, H., The Challenge of Efflux-Mediated Antibiotic
515 Resistance in Gram-Negative Bacteria. *Clinical Microbiology Reviews* **2015,** *28*, (2), 337-418.
516
517

518 **Figure legends**

519

520 **Figure 1**. **Time course of AST biosynthesis by *B. gladioli* GSRB05.** Production of AST-
521 OH and AST by cells of *B. gladioli* GSRB05 was assayed over a 24-h period. Cells were
522 grown in LB medium until late-log phase and transferred to $ST10^{-1}$ for up to 24 h in the
523 presence of 3 μM As(III). Samples were speciated by C18 reverse phase HPLC, and arsenic
524 determined by ICP-MS at the indicated times. After a lag period of approximately 10 h, AST-
525 OH and AST synthesis was observed. During the remaining time, the amount of AST
526 increased, and As(III) decreased correspondingly. Production of AST-OH leveled off after
527 approximately 14-15 h, indicating that it is an intermediate in the pathway of AST biosynthesis
528 from As(III). Bottom line: standards of arsenic compounds. As(III)T-OH, trivalent
529 hydroarsinothricin; As(III), arsenite; MAs(III), methylarsenite; DMAs(V), dimethylarsenate;
530 AST, arsinothricin; AST-OH, hydroxyarsinothricin; MAs(V), methylarsenate; As(V), arsenate.

531

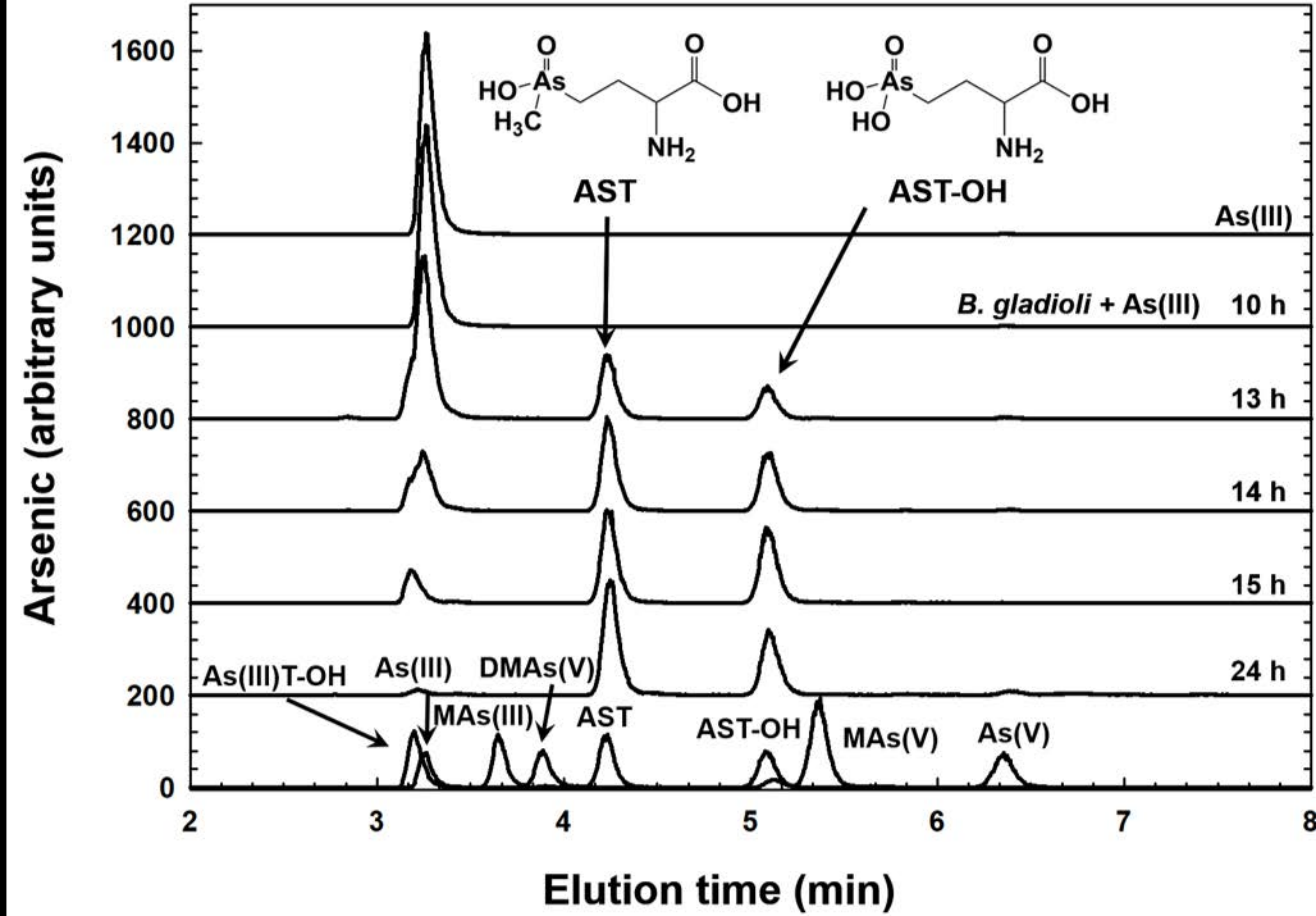
532 **Figure 2**. **The *B. gladioli* GSRB05 AST biosynthetic gene cluster. A**. *B. gladioli* GSRB05
533 genome (accession number JAGSIB000000000) contains a cluster of *ars* genes including
534 *arsQML*. **B**. Phylogenetic distribution of the genes of the AST BCG. GenBank accession
535 numbers are given in METHODS AND MATERIALS.

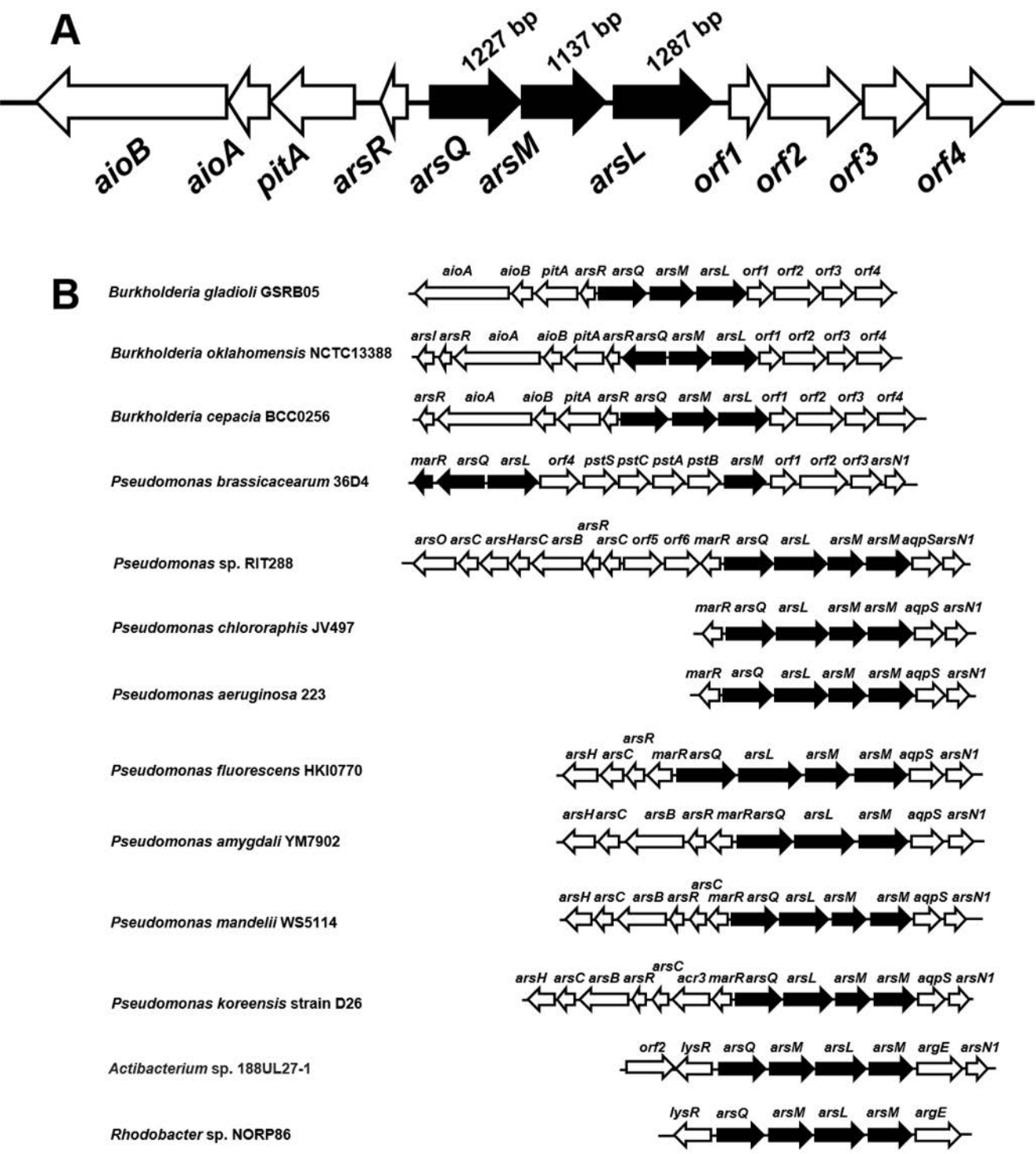
536

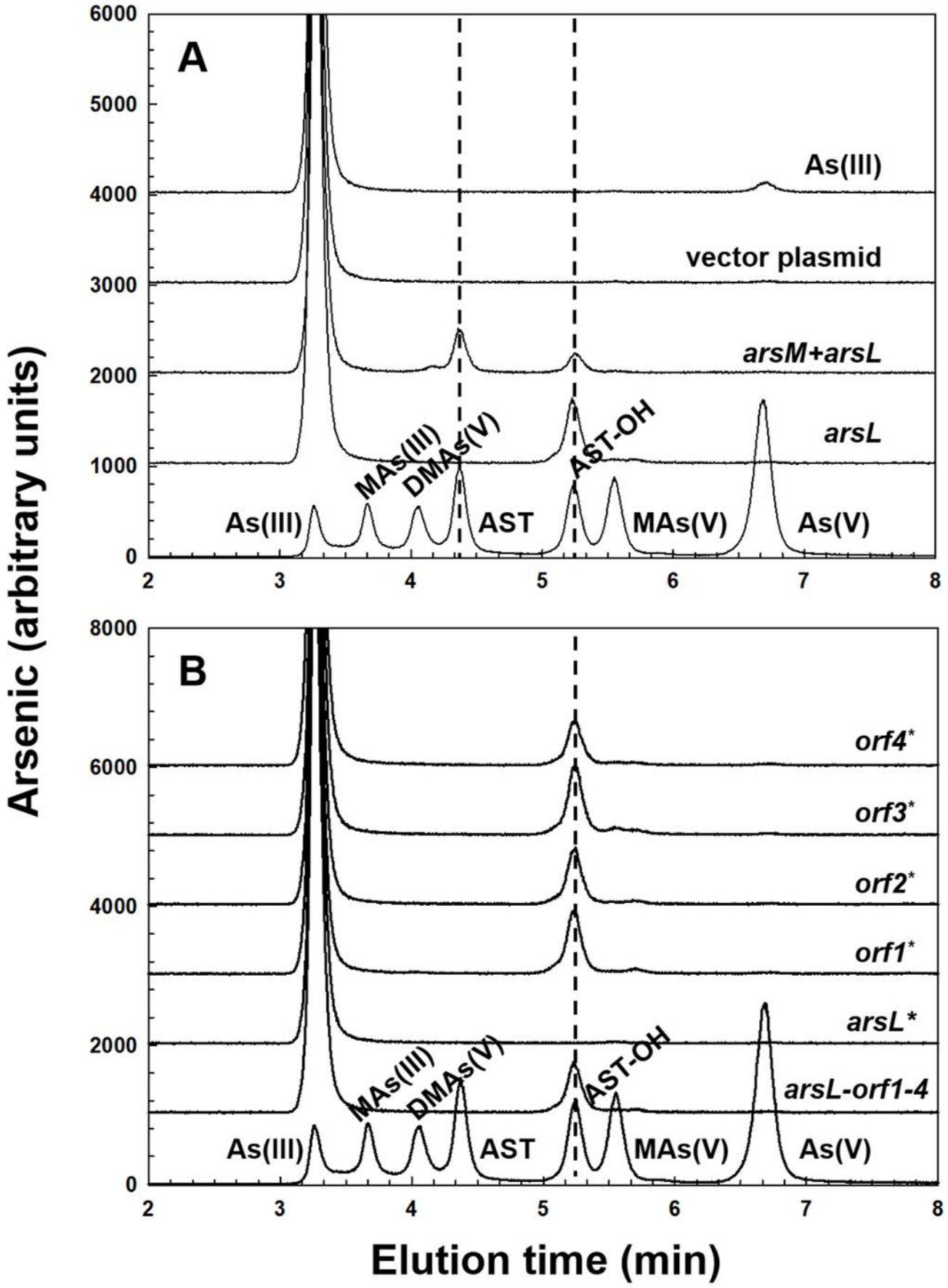
537 **Figure 3**. **AST-OH and AST biosynthesis in cells bearing *arsML-orf1-4*.** As(III)
538 biotransformation was assayed with selected genes from the AST BGC from *B. gladioli*
539 GSRB05. **A.** Cultures of *E. coli* Top10 expressing *arsML-orf1-4* and/or *arsL-orf1-4* in plasmid
540 pUC118. From the bottom the lanes are arsenic standards, cells with 1 μM As(III) bearing
541 pUC118*arsL-orf-1-4,* pUC118*arsML-orf1-4*, vector plasmid pUC118 and medium with only 1
542 μM As(III). **B**. Stop codons(*) were individually introduced into the genes in plasmid
543 pUC118*arsL-orf1-4*. From the bottom the lanes are arsenic standards, cells expressing wild
544 type *arsL-orf1-4* genes and stop codons in each individual gene. Cultures were grown in LB

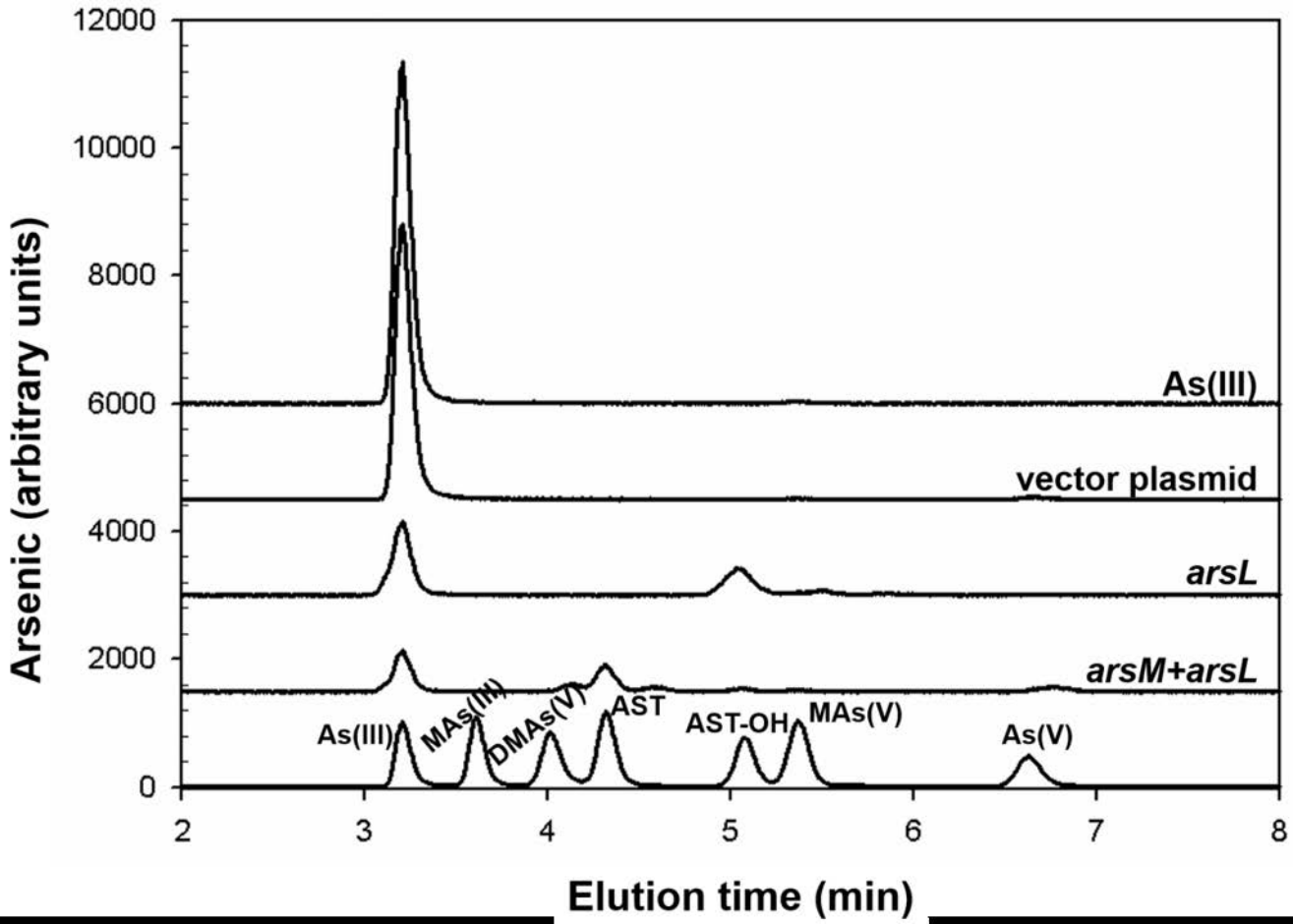
545 medium to late-log phase and transferred to ST 10^{-1} with 1μM As(III) and incubated for 36 h.
546 Arsenic-containing species were determined by HPLC-ICP-MS.
547
548 **Figure 4**. **The *arsL* and *arsM* gene products are sufficient to catalyze sequential steps**
549 **in the biosynthesis of AST.** Production of AST-OH or AST was assayed in cultures of *E. coli*
550 BL21 with 1 μM As(III), as described in the legend to Figure 3. From the bottom the lanes are
551 arsenic standards, cells bearing pETDuet-1*arsLM,* pETDuet-1*arsL*, vector plasmid pETDuet-
552 1and medium with only 1 μM As(III).
553
554 **Figure 5. The BGC is the *arsQML* operon**. **A.** The BGC and upstream and downstream
555 genes are illustrated with the size of each gene and intergenic regions given in nucleotide
556 base pairs. **B.** The indicated forward (F) and reverse (R) primers were used to amplify
557 transcriptional products by RT-PCR. **C.** Analysis of transcription of the *arsQ* (primers 1F-1R),
558 *arsM* (primers 2F-2R) *or arsL* (primers 3F-3R) genes individually and co-transcription of *arsM-*
559 *orf1* (primers 2F-4R) and *arsL-orf2* (primers 3F-5R) genes. *B. gladioli* GSRB05 was grown in
560 LB medium until late-log phase and then transferred to ST10^{-1} medium. Cultures were
561 incubated for 13 h in the presence (black bars) or absence (shaded bars) of 3 μM As(III).
562 cDNA was synthesized by RT-PCR from total RNA obtained from cultures of *B. gladioli*
563 GSRB05 . Data are the mean ± SE (n = 4).
564
565 **Figure 6. BgArsQ is an AST efflux permease**. **A**. Expression of *arsQ* confers resistance to
566 AST. Overnight cultures of *E. coli* AW3110 bearing either pTrcHisA*arsQ* (shaded bars) or
567 vector plasmid pTrcHisA (black bars) were diluted 100-fold into fresh M9 medium containing
568 the indicated concentrations of AST. A_{600nm} was measured after 24 h of growth at 37 °C. **B**.
569 Uptake of AST (40 μM) was assayed in cells of *E. coli* AW3110 bearing either pTrcHisA*arsQ*
570 (●) or vector plasmid pTrcHisA (o). **C**. Uptake of AST 20 μM was assayed in everted vesicles
571 prepared from cells of *E. coli* AW3110 expressing pTrcHisA*arsQ* (●) or vector plasmid

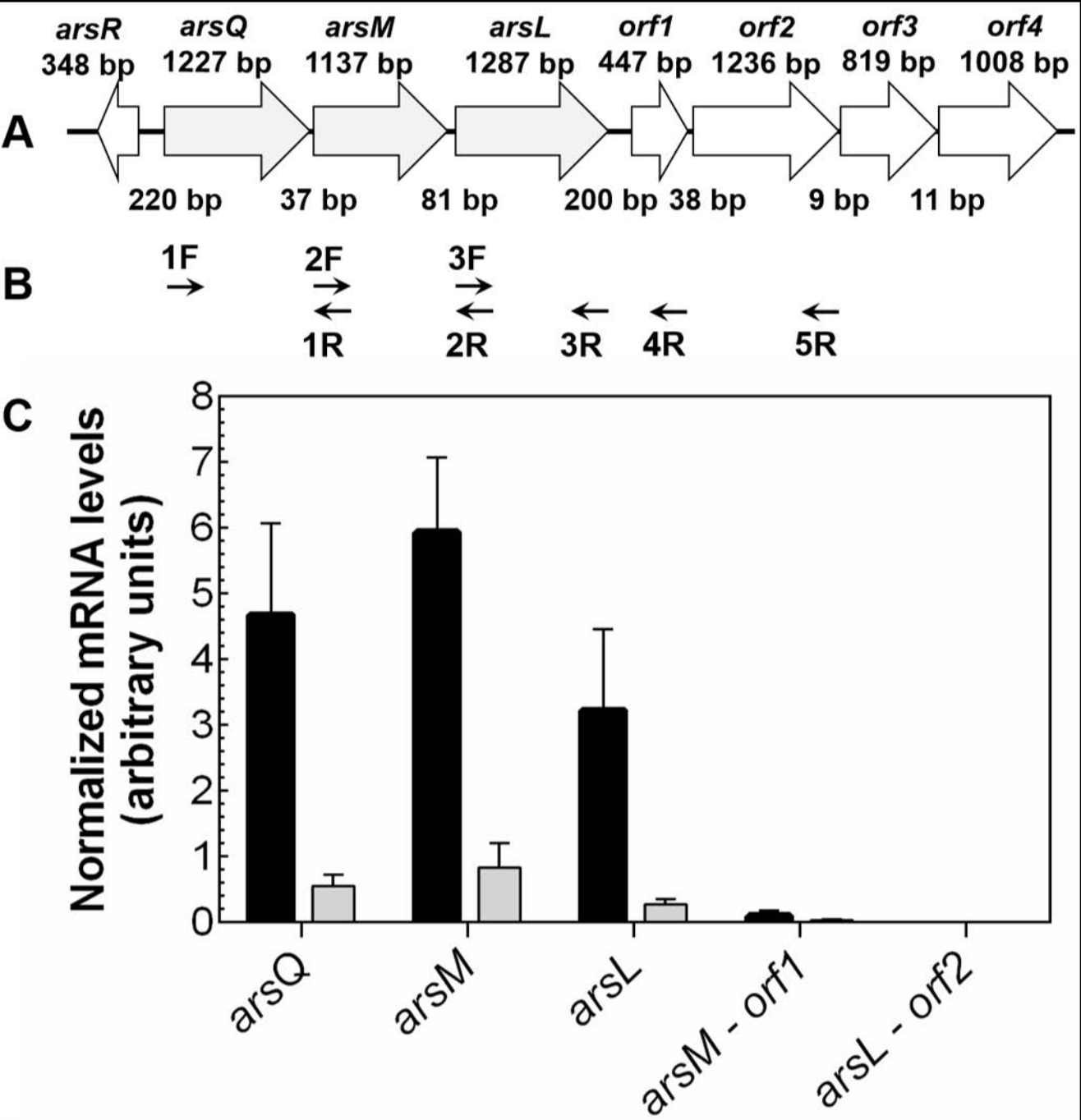
572 pTrcHisA (o) with 5 mM NADH as an energy source, as described in METHODS AND
573 MATERIALS. Data are the mean ± SE (n = 3).
574
575 **Figure 7. Proposed pathway of AST biosynthesis.** AST biosynthesis by *B. gladioli* GSRB05
576 is composed of two steps. In the first step, the non-canonical radical SAM enzyme BgArsL
577 cleaves the Cγ bond of SAM, forming a 3-amino-3-carboxypropyl (ACP) radical that creates a
578 C-As bond with As(III), producing trivalent AST-OH. This is different from most radical SAM
579 enzymes that from 5′-deoxyadenosyl radical by reductive cleavage of SAM at the 5' position
580 of the adenosine moiety. In the second step, BgArsM uses a second molecule of SAM to
581 methylate trivalent AST-OH, generating trivalent AST, which spontaneously oxidizes
582 nonenzymatically to the antibiotic pentavalent AST.
583

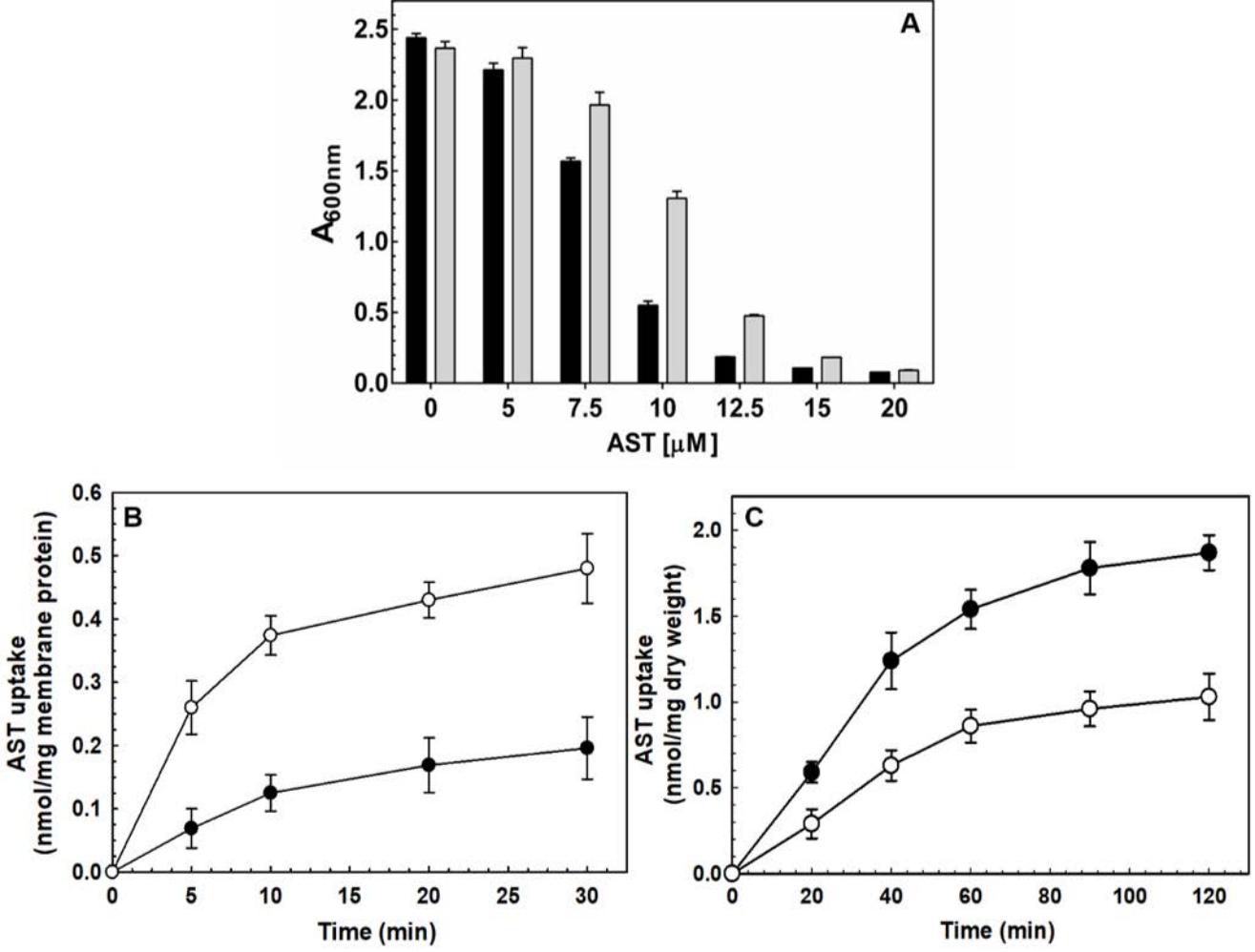

Arsenic (arbitrary units)
Elution time (min)
AST
AST-OH
As(III)
B. gladioli + As(III) 10 h
13 h
14 h
15 h
24 h
As(III)T-OH
As(III)
DMAs(V)
MAs(III)
AST
AST-OH
MAs(V)
As(V)

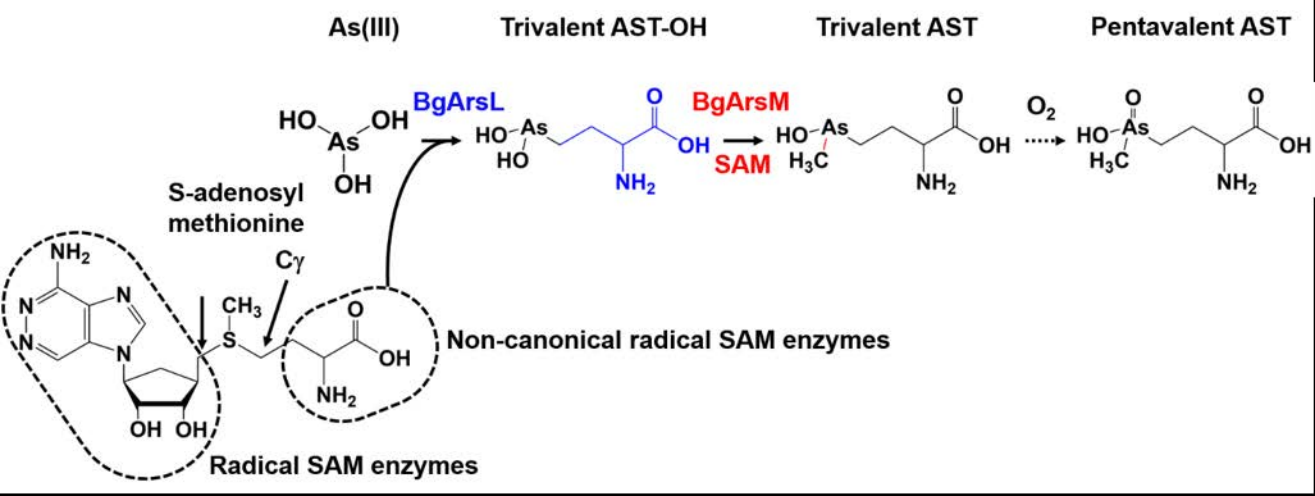

A
1227 bp
1137 bp
1287 bp
aioB
aioA
pitA
arsR
arsQ
arsM
arsL
orf1
orf2
orf3
orf4
B
Burkholderia gladioli GSRB05
aioA aioB pitA arsR arsQ arsM arsL orf1 orf2 orf3 orf4
Burkholderia oklahomensis NCTC13388
arsI arsR aioA aioB pitA arsR arsQ arsM arsL orf1 orf2 orf3 orf4
Burkholderia cepacia BCC0256
arsR aioA aioB pitA arsR arsQ arsM arsL orf1 orf2 orf3 orf4
Pseudomonas brassicacearum 36D4
marR arsQ arsL orf4 pstS pstC pstA pstB arsM orf1 orf2 orf3 arsN1
Pseudomonas sp. RIT288
arsR
arsO arsC arsH arsC arsB arsC orf5 orf6 marR arsQ arsL arsM arsM aqpS arsN1
Pseudomonas chlororaphis JV497
marR arsQ arsL arsM arsM aqpS arsN1
Pseudomonas aeruginosa 223
marR arsQ arsL arsM arsM aqpS arsN1
Pseudomonas fluorescens HKI0770
arsR
arsH arsC marR arsQ arsL arsM arsM aqpS arsN1
Pseudomonas amygdali YM7902
arsH arsC arsB arsR marR arsQ arsL arsM aqpS arsN1
Pseudomonas mandelii WS5114
arsC
arsH arsC arsB arsR marR arsQ arsL arsM arsM aqpS arsN1
Pseudomonas koreensis strain D26
arsC
arsH arsC arsB arsR acr3 marR arsQ arsL arsM arsM aqpS arsN1
Actibacterium sp. 188UL27-1
orf2 lysR arsQ arsM arsL arsM argE arsN1
Rhodobacter sp. NORP86
lysR arsQ arsM arsL arsM argE


A
As(III)
vector plasmid
arsM+arsL
arsL
As(III)
MAs(III)
DMAs(V)
AST
AST-OH
MAs(V)
As(V)
B
orf4*
orf3*
orf2*
orf1*
arsL*
arsL-orf1-4
Arsenic (arbitrary units)
Elution time (min)


12000
10000
8000
6000
4000
2000
0
Arsenic (arbitrary units)
As(III)
vector plasmid
arsL
arsM+arsL
As(III)
MAs(III)
DMAs(V)
AST
AST-OH
MAs(V)
As(V)
2
3
4
5
6
7
8
Elution time (min)


A
arsR
348 bp
arsQ
1227 bp
arsM
1137 bp
arsL
1287 bp
orf1
447 bp
orf2
1236 bp
orf3
819 bp
orf4
1008 bp
220 bp
37 bp
81 bp
200 bp
38 bp
9 bp
11 bp
B
1F
2F
3F
1R
2R
3R
4R
5R
C
Normalized mRNA levels
(arbitrary units)
0
1
2
3
4
5
6
7
8
arsQ
arsM
arsL
arsM - orf1
arsL - orf2


A
A_{600nm}
2.5
2.0
1.5
1.0
0.5
0.0
0
5
7.5
10
12.5
15
20
AST [μM]
B
AST uptake
(nmol/mg membrane protein)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
Time (min)
C
AST uptake
(nmol/mg dry weight)
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80
100
120
Time (min)


As(III)
Trivalent AST-OH
Trivalent AST
Pentavalent AST
BgArsL
BgArsM
SAM
O_2
S-adenosyl methionine
Cγ
Non-canonical radical SAM enzymes
Radical SAM enzymes

Submitted to Microbiology Spectrum June 9, 2021

Identification of the biosynthetic gene cluster for the organoarsenical antibiotic arsinothricin

Adriana E. Galván[1,orcid.org/0000-0002-2472-0438], Ngozi P. Paul[1,orcid.org/0000-0003-0423-195X], Jian Chen[1], Kunie Yoshinaga-Sakurai[1], Sagar M. Utturkar[2], Barry P. Rosen[1,orcid.org/0000-0002-5230-4271,] and Masafumi Yoshinaga[1*,orcid.org/0000-0002-7243-1761]

[1]*Department of Cellular Biology and Pharmacology, Florida International University, Herbert Wertheim College of Medicine, Miami, U.S.A.*

[2]*Purdue University Center for Cancer Research, Purdue University, West Lafayette, U.S.A.*

*__Corresponding author__: Masafumi Yoshinaga (myoshina@fiu.edu), Department of Cellular Biology and Pharmacology, Florida International University, Herbert Wertheim College of Medicine, 11200 SW 8th Street, AHC1 419G, Miami, Florida, U. S. A. 33199, Tel: (+1) 305-348-1489, Fax: (+1) 305-348-0651

Abbreviations: AST, arsinothricin (2-amino-4-(hydroxymethylarsinoyl) butanoate); AST-OH, hydroxyarsinothricin (2-amino-4-(dihydroxyarsinoyl) butanoate; ACP, 3-amino-3-carboxypropyl; MAs(III), methylarsenite; MAs(V), methylarsenate; DMAs(III), dimethylarsenite; DMAs(V), dimethylarsenate

Keywords: Arsinothricin, organoarsenical antibiotic, biosynthetic gene cluster, *Burkholderia gladioli* GSRB05

Supplemental information

Figure S1. Diagram of the sequential cloning of the AST BGC genes. DNA fragments of no more than 3.2 kb were sequentially amplified by PCR with the indicated restriction sites. The PCR products were inserted into plasmid pUC118. The first amplified PCR fragment included *orf2-4* cloned between the *Kpn*I and *Xba*I sites. In the second step *arsL* and *orf1* were amplified and cloned between the *Kpn*I and *Pac*I sites, resulting in plasmid pUC*arsL-orf1-4*. For the construction of pUC*arsML-orf1-4* or pUC*arsQML-orf1-4*, a third fragment including *arsM* or *arsQ-arsM* genes, respectively, was cloned between the *Kpn*I and *Nhe*I sites.

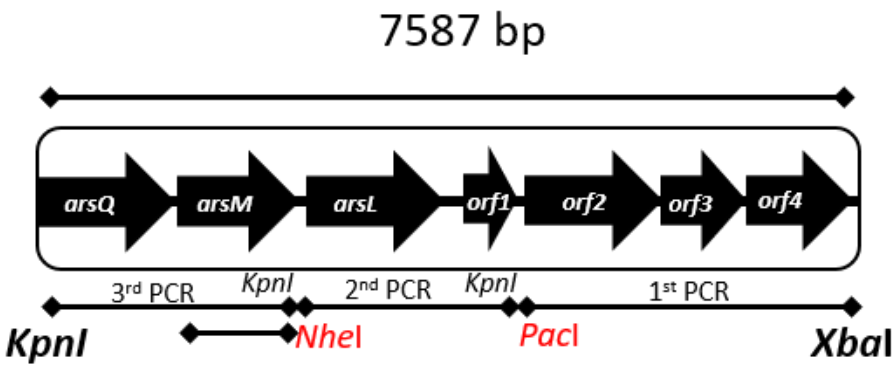


Figure S2. Genome sequencing of *B. gladioli* GSRB05

After quality trimming and filtering, >64% of reads were retained that represent >160X genome coverage for *B. gladioli* GSRB05. Optimal genome assembly obtained through SPAdes was comprised of 171 contigs in 9.2 MB with N50 contigs size 132 KB and average contig length 54 KB.

Sample_ID	file_name	Total Reads	Read Counts after quality filtering	Percentage of Reads after quality filtering
Burkholderia GSRB05	Bgladio_1.fastq	7,879,185	5,083,214	64.51%
	Bgladio_2.fastq	7,879,185	5,083,214	64.51%

Tool	KMER	Contigs	Minimum Conting Length	Maximum Conting Length	Average Contig Length	Contig N50	Assembly Size
SPAdes version 3.13.0	21,33,55,77,99,127	171	512	445,577	54,067	132,483	9,245,390
AByss version 2.1.5	31	522	507	149,544	17,564	39,307	9,168,461
	41	387	501	196,023	23,919	52,853	9,256,639
	51	364	509	198,283	25,481	53,516	9,275,193
	61	343	508	196,909	27,148	57,530	9,311,829
	71	343	505	198,297	27,150	58,494	9,312,489
	81	350	500	216,645	26,806	51,115	9,381,985
	91	408	504	150,616	22,918	45,268	9,350,690
	101	584	500	95,090	15,859	30,938	9,261,652
	111	1,316	500	62,372	6,935	14,061	9,126,425
	121	3,793	500	33,284	2,147	3,484	8,144,987

Figure S3. Multiple alignment of BgArsM orthologs (accession numbers in parentheses). The protein sequence of BgArsM identified from *B. gladioli* GSBR05 (BurGSRB05_34905) is compared with putative orthologs from *Burkholderia oklahomensis* (WP_010121994.1), *Burkholderia cepacia* (WP_059666519.1), *Pseudomonas aeruginosa* (WP_174517314.1), *Pseudomonas fluorescens* (WP_064119070.1), *Pseudomonas amygdali* (WP_054068373.1), *Actibacterium* sp. (WP_204414387.1) and *Rhodobacter* sp. (PCJ07719.1). The four conserved cysteines are indicated (*). Identities are shaded in black, and conservative replacements are shaded in grey.

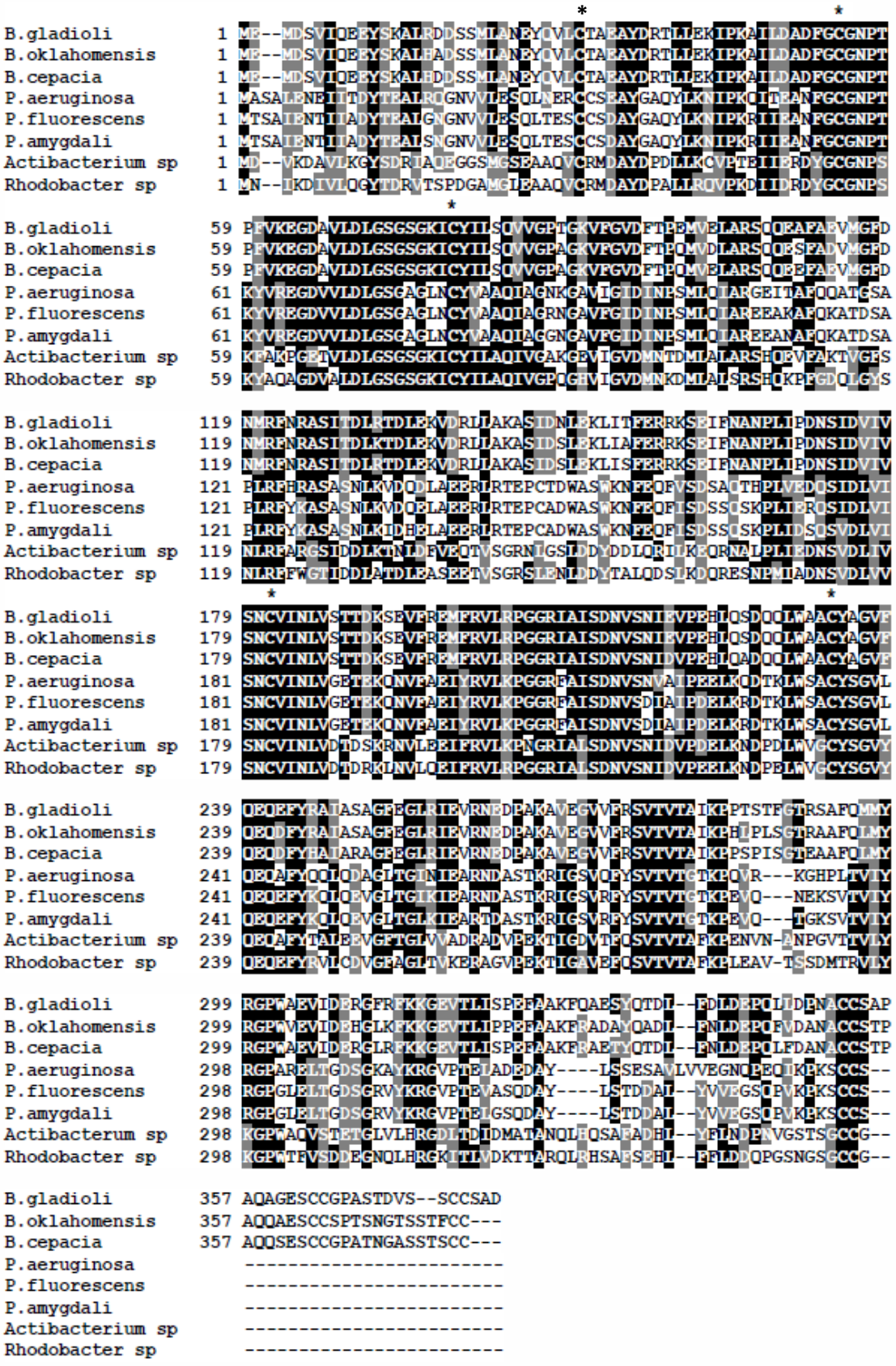


Figure S4. Multiple alignment of BgArsL orthologs (accession numbers in parentheses). The protein sequence of BgArsL identified from *B. gladioli* GSBR05 (BurGSRB05_34900) is compared with putattive orthologs from *Burkholderia oklahomensis* (WP_038802160.1), *Burkholderia cepacia* (WP_059666518.1), *Pseudomonas aeruginosa* (WP_174517312.1), *Pseudomonas fluorescens* (WP_064119068.1), *Pseudomonas amygdali* (WP_054068372.1), *Actibacterium* sp. (WP_204414396.1) and *Rhodobacter* sp. (PCJ07718.1). Identities are shaded in black, and conservative replacements are shaded in grey. The cysteines of the conserved CX_3CX_2C motif are identified (*).

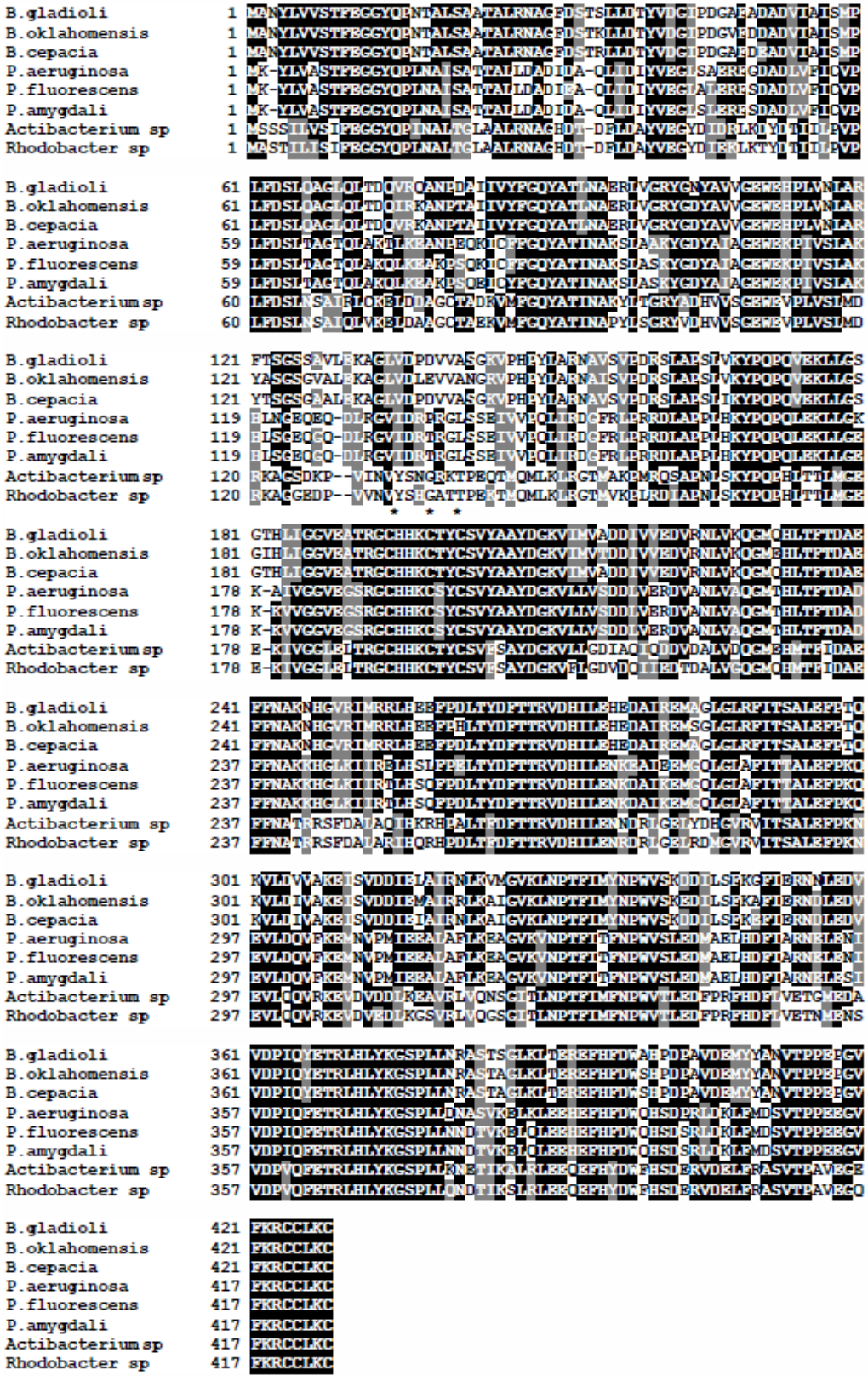


Figure S5. Multiple alignment of BgArsQ orthologs (accession numbers in parentheses). The protein sequence of BgArsQ identified from *B. gladioli* GSBR05 (BurGSRB05_34910) is compared with putative orthologs from *Burkholderia oklahomensis* (WP_010121992.1), *Burkholderia cepacia* (WP_081062335.1), *Pseudomonas aeruginosa* (WP_174517311.1), *Pseudomonas fluorescens* (WP_064119067.1), *Pseudomonas amygdali* (WP_054068371.1), *Actibacterium* sp. (WP_204414385.1) and *Rhodobacter* sp. (PCJ07720.1). Identities are shaded in black, and conservative replacements are shaded in grey.

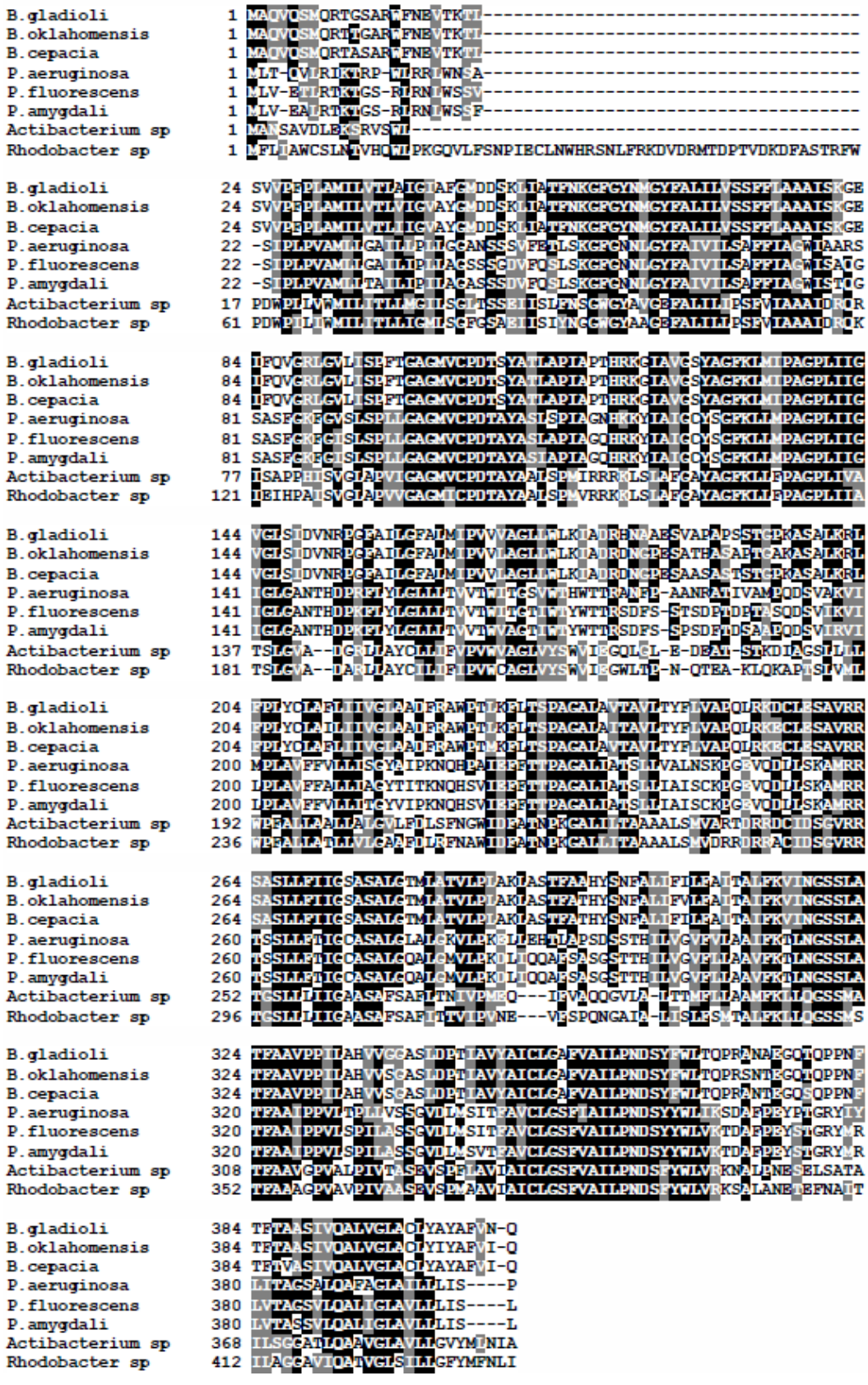


Table S1. Bacterial strains and plasmids

	Genotype/Description	References
Strain		
B. gladioli GSRB05	Soil isolated. AST producer	
E. coli Top10	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L(Str^R) *end*A1 *nup*G. Sm^r	Invitrogen
E. coli BL21 (DE3)	F⁻ *omp*T *hsd*S_B (r_B⁻, m_B⁻) *gal dcm* (DE3)	Novagen
E. coli AW3110	*ars*::*cam* F- IN(*rrnD-rrnE*). Cm^r	(1)
Plasmid		
pUC118	*E. coli* cloning and expression using *lac* promoter. Ap^r	Takara Bio USA
pUC*arsL-orf1-4*	PCR amplified *arsL* and orf1-4 genes inserted into the *Kpn*I and *Xba*I sites of pUC118.	This study
pUC*arsML-orf1-4*	PCR amplified *arsM-arsL* and orf1-4 genes inserted into the *Kpn*I and *Xba*I sites of pUC118	This study
pUC*arsQML-orf1-4*	PCR amplified *arsQ-arsM-arsL* and orf1-4 genes inserted into the *Kpn*I and *Xba*I sites of pUC118	This study
pETDuet-1	*E. coli* expression vector. Ap^r	Novagen
pETDuet-1a*rsL*	*arsM* gene cloned in MCSI	This study
pETDuet1a*rsML*	*arsM* gene cloned in MCSI and *arsL* gene cloned in MCSII	This study
pTrcHisA	*E. coli* expression vector. Ap^r	Thermo Fisher Scientific Inc
pTrcHisA2*arsQ*	*arsQ* genes inserted into the *Nco*I and *Sal*I sites of pTrcHisA.	This study

Abbreviations: Sm^r, streptomycin resistant. Cm^r, chloramphenicol resistant. Ap^r, ampicillin resistant.

1. Carlin, A.; Shi, W.; Dey, S.; Rosen, B. P., The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. J. Bacteriol. 1995, 177, (4), 981-6.

Table S2. Primers

Clonning rimers	Sequence
Bg_ars_XbaI_Rv	5’ ACCACCACCTCTAGATGGGGCATAGCGATAG 3’
Bg_ars_KpnI PacI_Fw	5’ GGTGGTGGTACCTTAATTAAGTGATAGCACGTCCGTCATTG 3’
Bg_ars_PacI_Rv	5’ ACCACCTTAATTAAGTGGCGCATGCCGGAGG 3’
Bg_ars_KpnI_NheI _Fw	5’ GGTGGTGGTACCGCTAGCTGAATAGATCAACAGACGACTTCAAG 3’
Bg_ars_NheI_Rv	5’ ACCACCGCTAGCATCAGCCGAGCAGCAG 3’
Bg_arsL_KpnI_Fw	5’ GGTGGTGGTACCTGGCCCAGGTTCAATC 3’
Bg_arsM_KpnI_Fw	5’ GGTGGTGGTACCAATGGAAATGGATTCTGTTATTC 3’
Red: unique site introduced for the sequentially cloning	

Site-mutant primers	
arsL-G31T-Fwd	5’ CTGATATCCACCTTAAAAAGTGGAAACGACTAGATAGTTGG 3’
arsL-G31T-Rev	5’ CCAACTATCTAGTCGTTTCCACTTTTTAAGGTGGATATCAG 3’
orf1-C55T-Fwd	5’ CATCGAGCATTTCCTACACGGGGTCTATCGC 3’
orf1-C55T-Rev	5’ GCGATAGACCCCGTGTAGGAAATGCTCGATG 3’
orf2 -G28T-Fwd	5’ GCGACCGGTTATTCCTTGGCGAGTGCTGC 3’
orf2 -G28T-Rev	5’ GCAGCACTCGCCAAGGAATAACCGGTCGC 3’
orf3-G31T-Fwd	5’ CGGTATAGAAGAGCTAAACCCCGTCGTTAGTTCT 3’
orf3-G31T-Rev	5’ AGAACTAACGACGGGGTTTAGCTCTTCTATACCG 3’
orf4-C28T-Fwd	5’ CGCGTCGGGCTAGGCGCTGCTGA 3’
orf4-C28T-Rev	5’ TCAGCAGCGCCTAGCCCGACGCG 3’

Primers used in *ars* operon identification by RT-qPCR	
1F	5’ ATGGCCCAGGTTCAATCGATG 3’
1R	5’ CCTGAATAACAGAATCCATTTCC 3’
2F	5’ ATGGAAATGGATTCTGTTATTCAG 3’
2R	5’ GTGGAAACGACTAGATAGTTG 3’
3F	5’ ATGGCCAACTATCTAGTCGTTTC 3’
3R	5’ TCAGCATTTGAGGCAGCATCG 3’
4R	5’ GAAAACGCCGATGCCGTAC 3’
5R	5’ AGAGTGAAATCGTCATCGATTG 3’
27F-16S	5’ AGAGTTTGATCTGGCTCAG 3’
1492R-16S	5’ GGTTACCTTGTTACGACTT 3’