EPFL

Thèse n° 7638

Methylation of arsenic by single-species and soil-derived microbial cultures

Présentée le 29 juin 2020

à la Faculté de l'environnement naturel, architectural et construit
Laboratoire de microbiologie environnementale
Programme doctoral en génie civil et environnement

pour l'obtention du grade de Docteur ès Sciences

par

Karen Elda VIACAVA ROMO

Acceptée sur proposition du jury

Dr S. Joost, président du jury
Prof. R. Bernier-Latmani, Prof. A. Mestrot, directeurs de thèse
Prof. L. Winkel, rapporteuse
Prof. S. Mitri, rapporteuse
Dr J. Maillard, rapporteur

École polytechnique fédérale de Lausanne

2020

"Oh not because happiness exists,
that too-hasty profit snatched from approaching loss.
But because truly being here is so much
because everything here apparently needs us,
this fleeting world, which is in some strange way
keeps calling to us."

— From the Ninth Duino Elegy, Reiner Maria Rilke.

To that 16-yearl-old school girl, on that summer day...

Acknowledgements

As every event in life, this thesis was made possible by innumerable circumstances and people impossible to be individually named, nevertheless, I would like to acknowledge some of the people whose role has been evident to me.

First of all, my gratitude goes to my supervisor Rizlan, one of the most sensible and straightforward people I have had the chance to meet: thank you for giving me the opportunity to be part of your team, to benefit from your guidance in the world of biogeochemistry, and for your understanding and support, to this day. Thanks to Karin, for always being willing to thoroughly read my manuscripts' drafts just one more time, your smart questions, your help with experiments and friendship. My thanks to all the current and previous members of the Environmental Microbiology Lab, specially: Matthew, for flattening the ground where I got to build my project; Manon, for being such fun and willing to listen in the times I needed; Solenne, for putting up with more than one kombucha explosion and being the sweetest officemate one could ask for; Leia, for your open laughter and lab assistance during difficult moments; Shannon, for your hard work and friendship; Margaux, for sharing with me your artistic creativity and delicious cooking; Zezhen, for your friendship, being so kind and just great to be around and Niels, for being the open-minded extroverted lad you are and never giving up on the quiet me to go for beers.

During this time, I was lucky to sporadically be part of the Soil Science Group at the University of Bern thanks to my co-supervisor Adrien. Thank you, Adrien, for having spent so much time in the lab with me during the first days, teaching me how an analytical chemist works, and after, for always being there when I needed something despite your growing responsibilities. Thanks also to Patrick, for being the most unwound, uncomplicated, reassuring person to work with; to Lorenz, for his everlasting good humor and uplifting laughter and to my North Americans girls: Jaime and Amy, for being such awesome fierce women, caring so much about others, thank you for your friendship.

Un grand merci à Sylvain, Karin and Felippe for our shared ICP-MS "parcours", you made me feel welcome in your team. I would also like to thank the great scientists without whom this work would not exist: Prof. Nigel P. Minton, David Ortega, Bob Hettich, Suresh Poudel and specially, Andrew Janowczyk.

Gracias mamá, gracias papá, por apoyarme en todas las etapas y cambios en mi vida. Sofia, gracias hermana por tu amor incondicional. Oscar, thank you for being my partner and refuge for four years (and thoroughly photo documenting it!) and for being the patient, loving person that you are. Finally, my love, I know between us "thank you" becomes redundant, so you will have to put up with it because it is the acknowledgement part: thank you for the nourishment and the washing up.

Renens, June 13th 2020 K.V.

Abstract

Arsenic (As) is simultaneously a ubiquitous and a toxic element. The toxicity of As depends on its chemical speciation. For instance, inorganic arsenic in its trivalent state, arsenite, which predominates under anoxic conditions, is more mobile and toxic than the pentavalent arsenate, found under oxygenated conditions and typically adsorbed to metal oxides. Additionally, arsenic is subject to bio−transformations catalyzed by living organisms, microorganisms being the major contributors. The combination of all changes in As speciation constitutes the biogeochemical cycle of arsenic in the environment.

The ecosystems in primordial Earth were devoid of oxygen, exposing primitive life to the highly mobile and toxic arsenite. Traces of this early contact with harmful As concentrations is imprinted throughout the tree of life in the widespread persistence of protective mechanisms, encoded by arsenic resistance genes. For example, efflux of arsenite by transmembrane transporters remains the most widespread microbial detoxification mechanism. Arsenic bio-transformations include changes in its redox state as well as the synthesis of methylated As (methylarsenicals). The latter reaction is catalyzed by the enzyme arsenite S−adenosylmethionine methyltransferase (ArsM in prokaryotes). ArsM attaches one or several methyl groups to inorganic As, generating organic As in the trivalent (highly toxic species) and pentavalent (relatively innocuous species) forms. Methylated As species can be soluble or volatile and are precursors for more complex arseno−organic molecules.

Worldwide concern about exposure to As has been raised due to the high levels of As in groundwater in South-East Asia, putting millions of people at potential risk. In those regions, groundwater is used for consumption and for cultivation of the regional staple food: rice. Rice plants are grown in paddy fields alternating between dry and flooded conditions. The flooded (anoxic) conditions are conducive to arsenite mobilization and uptake by rice roots, resulting in the presence of inorganic and methylated As in rice grains. Due to this mobilization of arsenic from soil into the human food chain, the paddy rice system is of particular importance in the As cycle.

Compared to inorganic As, methylarsenicals are more easily accumulated in rice grains and are implicated in rice plant disorders resulting in sterility. However, methylated species are not synthesized by the plant but by soil microorganisms. To date, mainly aerobic microbial

species able to methylate As as an apparent detoxification strategy, have been isolated from rice paddies. Thus, despite the fact that methylarsenical synthesis appears to be enhanced during anoxic conditions, the drivers and the physiological role of As methylation remain unknown in the absence of oxygen.

The work described in this thesis focuses on the identification of active As−methylating microorganisms as single species and as members of anaerobic microbiomes from rice paddy soils.

The first part of the work is centered around single microbial species encoding the ArsM protein. Seven strains (five bacterial strains and two methanogenic archaea), belonging to genera identified in paddy soils, were evaluated for active arsenite methylation. The results show that, while aerobic bacterial strains were sensitive to increasing As concentrations, they efficiently methylated As. Conversely, anaerobic bacterial strains were resistant to increasing As concentrations but methylate As poorly. The methanogenic archaea, although resistant to As, had similar methylation efficiencies as aerobic bacteria but those likely resulted from the release of methyltransferases due to cell lysis. Suspecting that As detoxification was occurring through As efflux and outcompeting As methylation under anaerobic conditions, we proceeded with the deletion of the only gene annotated as an arsenite transmembrane transporter in one of the anaerobic strains (*Clostridium pasteurianum*). The efflux of arsenite was disrupted in the mutant leading to higher intracellular As concentrations, higher *arsM* gene transcription, and increase in the As methylation efficiency. Based on the results, we hypothesize that under anoxic conditions, the efficient arsenite efflux systems from anaerobic microorganisms might preclude efficient As methylation.

In the second part of the work, meta−omic approaches (metagenomics, metatranscriptomics and metaproteomics) were used to identify active As methylators in two microbial communities, derived from rice paddy soil previously shown to methylate arsenite. The metagenomic data was used to reconstruct the microbial genomes present in the community as metagenome−assembled genomes (MAG). The metabolic capacity for each MAG was identified by the functional annotation of genes while active metabolisms were deciphered from mRNA transcripts (metatranscriptomic data) and protein expression (metaproteomic data). Identification of key enzymes from different metabolic pathways evidenced active dissimilatory sulfate reduction, denitrification, dissimilatory nitrate reduction, lactic acid, mixed acid, and propionic acid fermentation. Annotation and expression of arsenic resistance genes pointed to fermenting microorganisms as the main drivers of As methylation in both microbiomes. This work is a contribution to the understanding of the linkages between microbial diversity and arsenic bio-transformation, key factors to predicting which pathways will predominate and hence As chemical speciation.

Key words: Arsenic, arsenic methylation, arsenite methyltransferase, ArsM, Acr3, *Clostridium pasteurianum*, soil enrichments, rice paddy soil, metagenomics, metaproteomics, metatranscriptomics, *Paeniclostridium*, Clostridiales, fermenters.

Résumé

L'arsenic (As) est un élément chimique à la fois répandue et toxique. La toxicité de l'As dépend de sa forme chimique et de son degré d'oxydation. Par exemple, l'arsenic inorganique dans son état d'oxydation trivalent, aussi appelé arsénite et qui prédomine en conditions anoxiques, est plus mobile et plus toxique que l'arséniate pentavalent, que l'on trouve en conditions oxygénées et généralement adsorbé à des oxides de métal. Par ailleurs, l'arsenic fait l'objet de transformations biologiques catalysées par, en particulier, des microorganismes. L'ensemble des transformations modifiant les spéciations de l'arsenic constituent le cycle bio−géologique de l'arsenic dans l'environnement.

Sur la Terre primordiale, les écosystèmes étaient dépourvus d'oxygène, et la vie primitive était donc exposée à la forme mobile et toxique d'arsénite. Les traces de ce contact primitif avec des concentrations nocives d'arsenic ont été conservées dans l'arbre de la vie, et se manifestent par une persistance notable des mécanismes de protection, encodés par des gènes de résistance à l'arsenic. Par exemple, l'efflux de l'arsénite par les transporteurs transmembranaires est le mécanisme de détoxification le plus répandu chez les microorganismes. Les biotransformations de l'arsenic se manifestent par des changements de degré d'oxydation, mais également par la synthèse de composés d'arsenic méthylés. Cette dernière réaction est catalysée par l'enzyme arsenite S−adénosylméthionine méthyltransférase (ArsM chez les procaryotes). La protéine ArsM greffe un ou plusieurs groupements méthyles aux composés d'arsenic inorganiques, pour former des composés d'arsenic organiques trivalents (formes très toxique) et pentavalents (formes relativement inoffensives). Les composés méthylés de l'arsenic peuvent être solubles ou volatiles, et ce sont des précurseurs pour la synthèse de molécules organiques d'arsenic plus complexes.

Les niveaux d'exposition à l'arsenic font l'objet d'une préoccupation mondiale, en particulier à cause des taux élevés d'arsenic détectés dans les eaux souterraines du Sud-Est de l'Asie, qui représentent une potentielle menace pour des millions de personnes. Dans ces régions, l'eau souterraine alimente la consommation en eau potable, mais aussi les cultures de riz. Les plants de riz sont cultivés dans les rizières, immergées puis asséchées de façon répétée. Lorsque les rizières sont immergées (conditions anoxiques) l'arsénite est mobilisé et peut être ainsi assimilé par les racines des plants de riz, ce qui résulte en la présence de formes inorganiques et méthylées d'arsenic dans les grains de riz. En raison de la migration de l'arsenic du sol vers

la chaîne alimentaire humaine, le rôle des rizières est clé dans le cycle de l'arsenic.

Bien plus que les composés inorganiques de l'arsenic, les formes méthylées s'accumulent plus facilement dans les grains de riz et sont responsables de disfonctionnements observés dans la plante du riz, tels que la stérilité. Toutefois, les formes méthylées de l'arsenic ne sont pas synthétisées directement par la plante mais par les microorganismes du sol. Aujourd'hui, ce sont principalement les souches microbiennes aérobies, utilisant la méthylation de l'arsenic comme stratégie de détoxification, qui ont été isolées des sols de rizières. En revanche, en conditions anoxiques, bien que la synthèse de formes méthylées d'arsenic semble amplifiée, les microorganismes qui en sont responsables et son rôle restent inconnus à ce jour.

La recherche présentée dans cette thèse est focalisée sur l'identification de microorganismes capables de méthyler activement l'arsenic en tant qu'espèce isolée ou au sein de communautés microbiennes anaérobiques provenant du sol de rizières.

La première partie de ce travail est centrée sur l'étude d'espèces microbiennes isolées qui possèdent le gène codant pour la protéine ArsM. Sept souches (cinq souches bactériennes et deux archaea méthanogènes), appartenant aux genres identifiés dans les sols de rizières, ont activement méthylé l'arsenic. Les résultats montrent que les souches bactériennes aérobies sont sensibles à des concentrations croissantes d'arsenic, et par conséquent méthylent efficacement l'arsenic. A l'inverse, les souches bactériennes anaérobiques sont résistantes à des concentrations croissantes en arsenic, mais méthylent peu. Les archaea méthanogènes, bien que résistantes à l'arsenic, ont une efficacité de méthylation similaire à celle observée pour les bactéries aérobies, mais probablement liée à la libération de méthyltransférases après la lyse des cellules. En faisant l'hypothèse que le mécanisme de détoxification de l'arsenic a lieu par efflux plutôt que par méthylation en conditions anaérobiques, nous avons supprimé le gène codant pour un transporteur transmembranaire d'arsénite dans l'une des souches bactériennes anaérobies (*Clostridium pasteurianum*). La perturbation du mécanisme d'efflux de l'arsénite engendré dans la souche modifiée a entraîné l'accumulation intracellulaire d'arsenic, accompagnée d'une transcription plus importante du gène *arsM*, ainsi que d'une augmentation de l'efficacité de la méthylation de l'arsenic. Ces résultats nous ont conduit à supposer qu'en conditions anoxiques, le système d'efflux utilisé par les microorganismes anaérobies empêche la méthylation efficace de l'arsenic.

Dans un second temps, des approches méta−omiques (métagenomique, métatranscriptomique et métaprotéomique) ont été employées pour identifier les espèces microbiennes capables de méthyler l'arsenic dans deux communautés microbiennes de sols de rizières, méthylant activement l'arsenic. Les données de métagénomique ont permis de reconstruire les génomes microbiens présents dans la communauté en tant que génomes assemblés de métagénomes (MAG). Les fonctions métaboliques de chacun des MAGs ont été identifiées par annotation fonctionnelle des gènes, et parallèlement les mécanismes actifs ont été déterminés à partir de données de transcription de l'ARNm (métatranscriptomes) et de l'expression des protéines (métaprotéomes). L'identification d'enzymes clés intervenant dans la réduction des

sulfates, la dénitrification, la réduction des nitrates, et la fermentation des acides lactiques, d'acides mixtes et de l'acide propionique a confirmé que ces voies métaboliques sont actives. L'annotation et l'expression de la résistance à l'arsenic s'accordent à désigner les microorganismes fermentaires comme principaux acteurs de la méthylation de l'arsenic dans les deux microbiomes étudiés.

Ce travail apporte une contribution à la compréhension des liens existants entre la diversité microbienne et la biotransformation de l'arsenic, facteurs clés pour prédire quelles sont voies métaboliques prédominantes et les formes d'arsenic qui en découlent.

Mots clefs : Arsenic, méthylation de l'arsenic, arsenite méthyltransférase, ArsM, Acr3, *Clostridium pasteurianum*, enrichissement du sol, rizières, métagénomique, métaprotéomique, métatranscriptomique, *Paeniclostridium*, Clostridiales, microorganismes fermentaires.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The element Arsenic

Arsenic (As) is a chemical element, atomic number 33, classified in Group 5 of the Periodic Table. It is considered to be a metalloid with mixed metals and nonmetals properties. As all natural elements, it is found throughout the globe in soil, water and air. Natural arsenic consists of a single stable isotope, ^{75}As, and four oxidation states: pentavalent such as arsenate As(V) and trivalent such as arsenite As(III), being the most common, and elemental As(0) and reduced arsenide (As(-III)) being rarer (Cullen & Reimer, 1989).

The name arsenic is derived from the Persian word *zarnikh*, "gold−colored" later adopted by Greek as *arsenikon*, meaning valiant, virile or potent. It was the name used for orpiment, the deep yellow−colored arsenic sulfide mineral (As_2S_3) known to react readily. Arsenic sulfides, such as orpiment or realgar, and oxides were already known in ancient civilizations for medical purposes (Cullen, 2008c). Due to being frequently used to assassinate members of ruling classes and to its potency, As became known as the "Poison of Kings" or the "King of Poisons" (Mudhoo et al., 2011).

Still to date, in popular culture, As is best known as a poison, even though it is an ubiquitous element and the environmental implications of its toxicity have made it the object of scientific study.

1.2 Sources of human exposure to Arsenic

1.2.1 Arsenic in groundwater

The concentration of arsenic in groundwater varies greatly worldwide, from 0.5 to 5,000 $\mu g\ l^{-1}$ (1−10 $\mu g\ l^{-1}$ for unpolluted surface water) (Singh et al., 2015). The World Health Organization (WHO) guideline for maximum As level in drinking water is 10 $\mu g\ l^{-1}$ (World Health Organization, 2018). The major natural sources of As in the hydrosphere are: geothermal

activity, As−bearing minerals and the reduction of iron and aluminum oxyhydroxides/oxides. Areas with high As levels in the American continent are concentrated around locations with magmatic activity (geothermal waters, volcanic ash, sulfide deposits), for example, the well−known Yellowstone National Park, USA. Hot−spring waters, containing soluble As(III) from magma, transport the element to surface waters where it is chemically or microbially oxidized to As(V) and precipitated, becoming part of the river sediments. The second main source is ores deposits. There about 150 species of As−enriched minerals, but only three are considered to be arsenic ores: arsenopyrite (FeAsS), realgar (AsS or As_4S_4) and orpiment (As_2S_3) (Mudhoo et al., 2011). Mining for arsenic is rare. However, As−bearing sulfide minerals can be part of mine tailings and sludge from mining activity targeting more economically valuable metals. As part of the mining waste, arsenic is released into the hydrosphere via the oxidation of the As−bearing sulfide minerals. Finally, the third and most widely recognized mechanism for mobilization of the metalloid into groundwater is the reductive dissolution of iron−bearing oxyhydroxides/oxides. Iron oxyhydroxides/oxides strongly absorb arsenic and are stable under oxic and weakly acidic to alkaline conditions. Arsenic is released under reducing and alkaline conditions and/or by microbial dissimilatory reduction of ferric iron (Fe(III)) or As(V) (Masuda, 2018). This mechanism of As mobilization causes the contamination of subsurface waters (Oremland & Stolz, 2005).

Groundwater contaminated with As is a problem that affects many populations worldwide but the most severely affected areas are in South−East Asia (Figure 1.1). In this region, the source of As is from the reduction of As−enriched sediments derived from the Himalayas and carried downstream by the river systems (Goodbred & Kuehl, 2000). Millions of people in India (70−80 million people), Bangladesh (28−60 million), Pakistan (47−60 million), Vietnam (about 10 million), China (2−5 million), Nepal (more than 3 million), Myanmar (about 3.4 million) and Cambodia (0.32−2.4 million) rely on As−contaminated ground water (Uppal et al., 2019) for consumption and for crop (mainly rice) irrigation. The concern of As contamination in the above−mentioned Asian countries was raised since the late 1980s when inhabitants were diagnosed with arsenicosis, an illness resulting from chronic exposure to As, and the disease linked to the high contents of As in their groundwater (Mandal & Suzuki, 2002).

1.2.2 Arsenic in rice

When As−contaminated water is used for irrigation of aerated soils such as for wheat, corn and most vegetables, the metalloid is likely to be present as As(V), which, as mentioned, is readily adsorbed onto iron and aluminum oxyhydroxides/oxides, thus becoming unavailable to plants. In the case of anaerobic or seasonally−flooded soils, like rice paddy fields, the metalloid is present as As(III), which is highly mobile and readily available to plants. As rice plants deliver oxygen to their roots, the oxygen that leaks out oxidizes ferrous iron (Fe(II)), generating an iron oxide plaque on rice roots. In the thick roots, the iron plaque absorbs As, limiting its transport towards the plant. However, in the young, fine roots, such as the root hairs responsible for overall nutrient uptake, no plaque is formed (Seyfferth et al., 2010). In the exposed roots, As(V)

Figure 1.1 – Distribution of As−contaminated groundwater in South−East Asia. Fig. 1 from (Brammer & Ravenscroft, 2009).

competes with phosphate for uptake while As(III) is taken up along with water. The plant, in turn, accumulates the assimilated As into rice grains to concentrations approximately tenfold greater than those encountered in wheat or barley (0.11−0.34 μg g^{-1} vs. 0.02−0.04 μg g^{-1}) (Williams et al., 2007). Many rice−exporting countries are impacted by As contamination of water or soil. Additionally, more than half of the world's population consumes rice as its staple food. Consequently, rice is the main dietary source of As for populations unaffected by As−contaminated groundwater (Meharg & Zhao, 2012). The recommended limit of As in rice by the Codex Alimentarius Commission is 0.2 mg kg^{-1} (polished rice) and 0.35 mg kg^{-1} (unpolished/husked rice). The threshold values in rice are for inorganic As (iAs); however As can also be present in an organic form, mainly the As methylated form (see Section 1.5) dimethylarsinic acid (DMAs(V)). The ratio of DMAs(V) to iAs varies widely as a function of geographic location. In Asia, the concentrations of total As and iAs in rice have been shown to correlate, while in Europe and the USA, the percent contribution of DMAs(V) increases with increasing total As (Zhao et al., 2013).

1.2.3 Arsenic in soil

Arsenic constitutes 0.0001% of the Earth's crust. The average As concentration in soil is estimated to be 7.2 mg kg^{-1} and, in uncontaminated soils, varies between 0.1 to 40 mg kg^{-1} (Singh et al., 2015). In contrast with other substances, for which International threshold values

and standards are similar, International guidelines for safe levels of As in soil vary widely, from 0.039 to 40 mg kg^{-1}, in the US and from 5 (Finland) to 150 mg kg^{-1} (Japan) in other countries (Teaf et al., 2010). In addition to the natural weathering processes of As−bearing minerals, soils with high As concentrations also result from its use in various anthropogenic activities.

Although banned from residential use, currently, the main use of As by humans is chromated copper arsenate (CCA), used as wood preservative for outdoor use (https://www.epa.gov/ingredients-used-pesticide-products/chromated-arsenicals-cca). In the recent past, its use was also widespread in poultry and swine feed in the USA and 14 other countries. Until 2011, arsenic was an animal food additive in the form of the drug roxarsone (derivative of phenylarsonic acid) (https://www.nytimes.com/2011/06/09/business/09arsenic.html), marketed as 3−Nitro, which kills intestinal parasites, increases weight gain and makes meat pinker. Most of the drug was excreted by the animal and the use of chicken and pig manure for fertilization of agricultural lands resulted in roxarsone being released into the environment (Rutherford et al., 2003). The employment of As for the synthesis of insecticides and herbicides (as cacodylic acid, another name for DMAs(V)) is also in decline (Bencko & Foong, 2017), except for the case of monosodium methyl arsonate (or methanearsonate) (MSMA), which is still used in golf courses and cotton fields (https://www.epa.gov/ingredients-used-pesticide-products/monosodium-methanearsonate-msma-organic-arsenical). Some other applications of the metalloid, with a much lower environmental impact, include its use: as a semiconductor and doping agent in electronic devices (Z. Liu et al., 2005), for the production of diodes (LED) in combination with gallium (Grund et al., 2011), as a lead alloy for bullets and batteries (Bagshaw, 1995), or for leukemia therapy in a drug called Trisenox (P. Zhang, 2017). Other anthropogenic sources of As, not derived from its direct use, include phosphate fertilizers (reported to contain up to 31 mg kg^{-1} of As)(Jayasumana et al., 2015), and as a byproduct of mining and smelting (main source of As−containing raw materials) (Cullen, 2008b).

1.3 Arsenic toxicity and effect in human health

As shown in the previous section, the environmental presence of As in any form is or can become a global public health issue due to its toxicity. Arsenic toxicity is determined by its speciation. By comparing the dose that results in 50% lethality (LD_{50}) (Table 1.1), it is evident that inorganic As species are more toxic than pentavalent methylated species but less toxic than the reduced (trivalent) methylarsenicals monomethylarsonous acid (MMAs(III)) and dimethylarsinous acid (DMAs(III)) (Table 1.2). Furthermore, the latter species have been found to be genotoxic (Mass et al., 2001), damaging DNA by the formation of reactive oxygen species (ROS) during their oxidation (Nesnow et al., 2002) and causing chromosomal mutations (Vasken Aposhian et al., 2004). Furthermore, when human cells (from liver, lung, skin and bladder) were exposed to trivalent arsenic−bearing species, MMAs(III) had the highest cytotoxicity and DMAs(III) was either more or as toxic as As(III), depending on the cell type. Also, it is important to note that MMAs(III) has been detected in urine from people

chronically exposed to As (Aposhian et al., 2000).

Table 1.1 – Lethal dose, 50% (LD_{50}) in mice from different arsenic species.

As species	LD_{50} in mouse (mg kg^{-1})	Study
Arsenite (As(III))	14	(Prohaska & Stingeder, 2005)
Arsenate (As(V))	20	(Prohaska & Stingeder, 2005)
Monomethylarsonic acid (MMAs(V))	1,800	(Kaise et al., 1989)
Dimethylarsinic acid (DMAs(V))	1,200−2,600	(Kaise et al., 1989; Prohaska & Stingeder, 2005)
Trimethylarsine oxide (TMAsO)	10,600	(Kaise et al., 1989)
Trimethylarsine ($(CH_3)_3As$)	7,870	(Yamaguchi et al., 2011)
Arsenobetaine	10,000	(Cullen, 2008a)

Table 1.2 – Lethal dose, 50% (LD_{50}) in human hepatocytes from different arsenic species.

As species	LD_{50} in human hepatocytes (mg kg^{-1})	Study
Arsenite (As(III))	9.8	(Dopp et al., 2008)
Arsenate (As(V))	37.5	(Dopp et al., 2008)
Monomethylarsonous acid (MMAs(III))	1.5	(Dopp et al., 2008)
Dimethylarsinous acid (DMAs(III))	0.9−2,600	(Dopp et al., 2008)

Both chronic and acute toxicity have been connected to exposure to As. Chronic exposure to As causes skin lessons and liver fibrosis. Among the most common diseases encountered in areas severely impacted by As are black−foot disease (severe pain in extremities resulting in gangrene), hyperpigmentation and hyperkeratosis (Hettick et al., 2015). The metalloid has been classified by the International Agency for Research on Cancer (IARC) in the Group 1 of carcinogens, *i.e.*, causally associated with cancer in humans (skin, bladder, kidney, liver, lung, and prostate) (McCarty et al., 2011). Symptoms of acute toxicity include cramps, nausea, vomiting, diarrhea, hypotension, pulmonary edema and heart failure. Death by acute poisoning is caused by cardiovascular collapse and hypovolemic shock (loss of one−fifth of blood body or fluid supply making it impossible for the heart to function). The LD_{50} of As trioxide for humans is 70−300 mg (Mochizuki, 2019).

1.4 The biogeochemical cycle of Arsenic

In spite of the detrimental effects of arsenic exposure has on human health, life has evolved under the pressure of this ubiquitous toxic element and many organisms are able to detoxify from it. Bio−transformation, along with chemical processes, gives shape to the biogeochemical cycle of arsenic (Figure 1.2). Amongst all the living organisms forming part of this cycle, microbes are by far its most important and therefore the focus of this section.

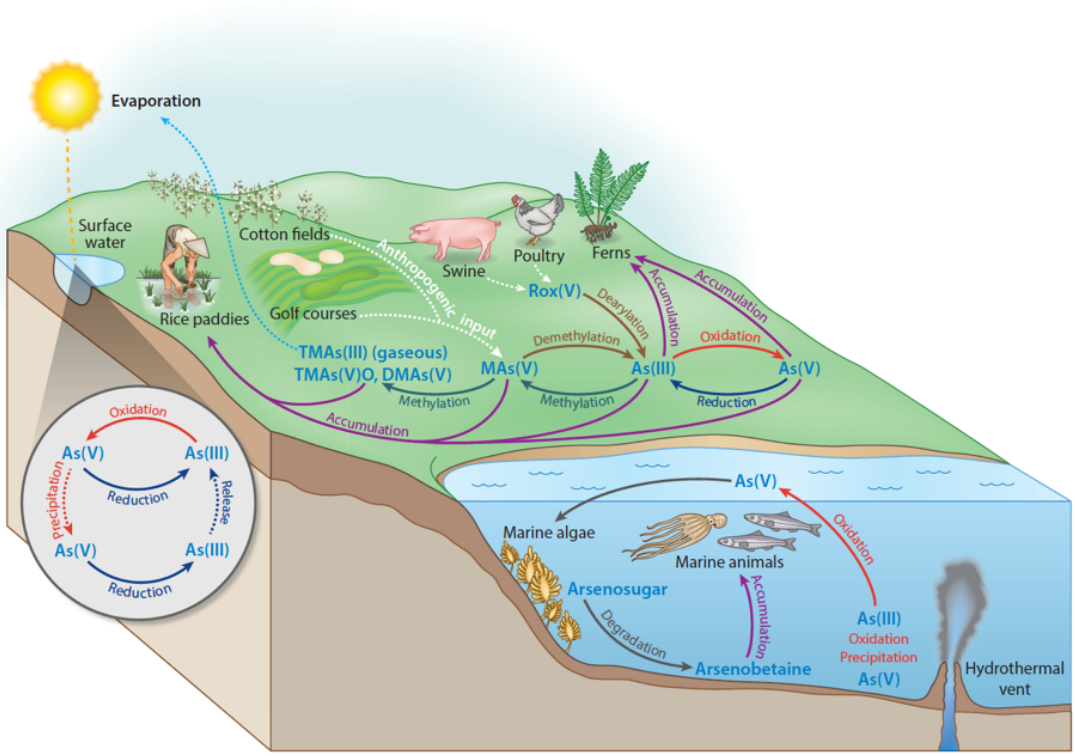


Figure 1.2 – Arsenic biogeochemical cycle. Fig. 4 from (Y.-G. Zhu et al., 2014).

Under anaerobic conditions, As−bearing sulfide minerals are very stable and insoluble in water. These minerals are especially common in low−temperature hydrothermal deposits. If they come into contact with oxidants (such as air or oxyhydroxides/oxides), oxidizing microorganisms, or rising temperatures, they readily decompose releasing soluble and mobile As(III) which can be biologically or chemically oxidized to As(V) (red arrows in Figure 1.2). The adsorption of As(V) and As(III) on iron and aluminum oxyhydroxides/oxides is pH−dependent. Furthermore, phosphate adsorbs strongly onto iron oxides and competes with As for surface sites. In the pH range 6−9, typical of natural environments, As(III) adsorbs to a similar extent as As(V), however phosphate prevents the sorption of As(III) much more efficiently than that of As(V) (Dixit & Hering, 2003). Besides, As(V) is reduced to As(III) under suboxic conditions (oxidation−redox potential <100 mV at pH 7) (blue arrows in Figure 1.2), and, as a result, As(III) is the main aqueous species in flooded paddy soils or anaerobic sediments.

Before the evolution of oxygenic photosynthesis, the ancient atmosphere was composed of methane, carbon dioxide and nitrogen (Hohmann-Marriott & Blankenship, 2011). Thus, early life was anaerobic, providing an environment favorable for As(III) release into aqueous environments. Consequently, Archaean ancestors would have been exposed to an environment rich in As(III). At present, vestiges of this environmental pressure are retained in the genome of nearly every living organism as arsenic−resistance genes (*ars*) or genes that encode for catabolism associated with the oxidation of As(III) and the reduction of As(V) (*arr* genes). A list of these As−related genes, the function of the proteins they encode, and the process in which they intervene are presented in Table 1.3 and a schematic diagram of the catalyzed cellular processes in Figure 1.3.

Table 1.3 – Genes involved in arsenic resistance and metabolism.

Process	Gene	Description
Resistance	*arsR*	Transcriptional repressor
	arsB/acr3	As(III) efflux system
	arsC	As(V) reductase
	arsA	Anion−stimulated ATPase
	arsD	As(III) chaperone
	arsH	MMAs(III) oxidase
	arsP	MMAs(III) permease
	arsJ	1As3PGA permease
	arsI	As−C lyase
	arsK	As(III) and MMAs(III) efflux system
Resistance or metabolism	*arsM*	As(III) methyltransferase
	arsS	Protein involved in arsenosugar synthesis
	aioA	As(III) oxidase large subunit
	aioB	As(III) oxidase small subunit
	aoiR/arxR	Transcriptional regulator
	aoiS/arxS	Histidine kinase sensor
	arxA	As(III) oxidase molybdopterin subunit
	arxB	As(III) oxidase iron−sulfur subunit
	arxC	Membrane anchor
Metabolism	*arrA*	As(V) reductase large subunit
	arrB	As(V) reductase small subunit
	arrC	Membrane−bound reductase subunit
	arrD	As(V) chaperone
	arrS	Histidine kinase sensor
	arrR	Transcriptional regulator

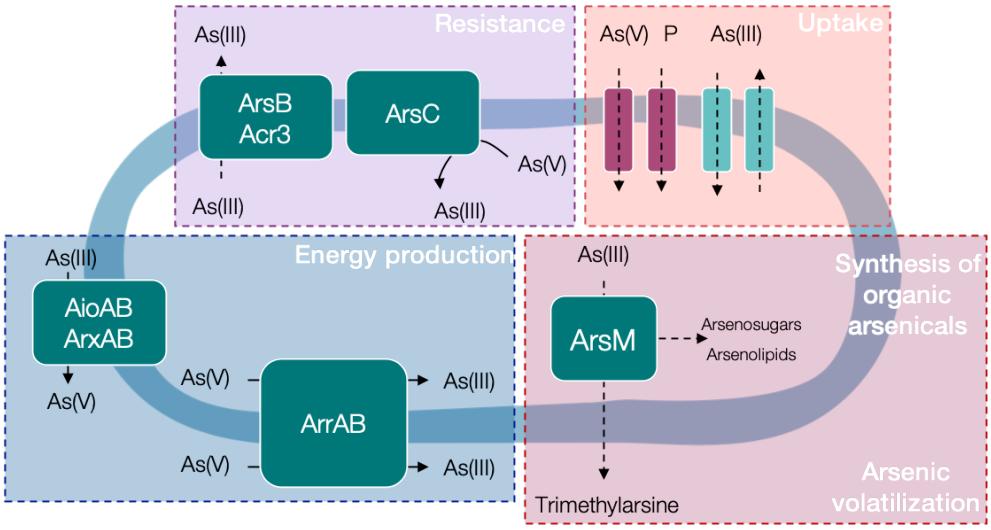


Figure 1.3 – Some microbial cellular processes involving arsenic.

1.4.1 Arsenic efflux

As(III) enters the intracellular space through passive transport via aquaglyceroporin channels. Once inside the cell, it inactivates enzymes by binding to cysteine residues, preventing disulfide bond formation (Bhattacharjee et al., 2008). Perhaps one of the most straightforward mechanisms of detoxification for As(III) is to pump it out of the intracellular space. This is achieved by two types of As(III)−efflux systems, ArsB and Acr3.

ArsB is an antiporter found mostly in bacteria, and exchanges As(III) efflux for the infflux of a proton. The extrusion of As(III) can be either coupled to the electrochemical proton gradient or be ATP−dependent. In the second case, ArsB binds to ArsA, a As(III)−translocating ATPase. The ArsAB pump can work with a chaperone, ArsD, which binds cytosolic As(III) and transfers it to the ArsA subunit of the efflux pump (Garbinski et al., 2019).

Acr3 is more widespread than ArsB, and is also found in yeast and archaea. It is a member of the bile/arsenite/riboflavin transporter superfamily (BART) and consists of a uniporter that extrudes As(III) driven by the membrane potential. Thus, Acr3 and ArsB are key encoded proteins in the *ars* operons. Expression of the *ars* genes is controlled by ArsR, a As(III)−responsive transcriptional repressor (Andres & Bertin, 2016).

There is only one system known for direct extrusion of As(V). It consists of an enzyme (ArsJ) capable of extruding 1−arseno−3−phosphoglycerate (1As3PGA) to the extracellular space. Once outside the cell, 1As3PGA splits into As(V) and 3−phosphoglycerate (Antizar-Ladislao, 2010; H. L. Chen et al., 2016). The gene *arsJ* is localized along side the glyceraldehyde−3−phosphate dehydrogenase that synthesizes the 1As3PGA.

1.4.2 Arsenite oxidation

In early ecosystems, the abundance of As(III) and the possibility to harvest its electrons for bioenergetics would have been a driving force for the evolution of As(III) oxidases. The first discovered As(III) oxidase was AioAB (also referred to as AoxAB), which is composed of two different subunits: a large subunit (AioA) containing a molybdopterin and an iron−sulfur cluster and a small subunit (AioB) containing an Rieske−type iron−sulfur cluster. The first AioAB dimers studied belonged to bacteria in thermophilic habitats rich in As(III), which are often cited as modern analogs to where life might have originated (Lebrun et al., 2003). Several bacterial strains have been shown to couple As(III) oxidation to the reduction of electron acceptors such as nitrate or selenate (Fisher & Hollibaugh, 2008; Oremland et al., 2002). Accompanying the *aioAB* genes, there is a transcriptional regulator (*aioR*) and a histidine kinase sensor (*aioS*) that work together in a two−component signal transduction system (Sardiwal et al., 2010).

There is a second As(III) oxidase alternative to AioAB called Arx which is able to work in both directions (oxidation−reduction) (Van Lis et al., 2013). In addition to the molybdopterin-containing ArxA, the *arx* operon includes a gene encoding a protein with an iron−sulfur cluster (*arxB*), a membrane anchor (*arxC*), a histidine kinase sensor (*arxS*) and a transcriptional regulator (*arxR*). Oxidation of As(III) by ArxA has been found to be coupled to nitrate reduction and anoxygenic photosynthesis (Ospino et al., 2019).

In addition to an energy−generating process, As(III) oxidation can also be purely an arsenic−resistance pathway due to the decrease in toxicity resulting from oxidation of As(III) to As(V).

1.4.3 Arsenate reduction

The As(V) chemical structure is analogous to that of phosphate. Thus, As(V) is able to enter the intracellular space via phosphate transporters (Y. Yan et al., 2017). In early Earth, oxidation of As(III) would have produced As(V), a potential sink for electrons. However, after the Great Oxidation Event, when the oxygen concentration in the atmosphere rose (Holland, 2006), As(V) became more abundant as a poison and organisms had to diversify their existing As(III) detoxification pathways to include the oxidized species. The latter explains the apparent paradox of As(V) reduction being part of the arsenic detoxification pathway given that this process generates the more toxic species, As(III), in the process of detoxifying As(V). Once As(V) is reduced to As(III), it can be promptly pumped out, as previously described.

Cytoplasmic As(V) reduction to As (III) is catalyzed by the As(V) reductase ArsC. There are two families of ArsC proteins. One family couples As(V) reduction to the oxidation of glutaredoxin (Grx), a small enzyme chemically reduced by the oxidation of glutathione (GSH). The second family makes use of internal cysteine residues generating a disulfide bond in the protein. Subsequently, thioredoxin (Trx) re−reduces the cysteine residues. ArsC−encoding genes form part of the *ars* operon.

As(V) reduction can also be catalyzed by dissimilatory As(V)−respiring microorganisms (DARPs). The reaction is catalyzed by the respiratory As(V) reductase, a dimer consisting of a large catalytic subunit (ArrA) and a small subunit (ArrB). The dissimilatory respiration operon (*arr* genes) also includes genes encoding a membrane−bound reductase subunit (*arrC*), a As(V) chaperone (*arrD*), a histidine kinase sensor (*arrS*) and a transcriptional regulator (*arrR*) (Van Lis et al., 2013).

1.4.4 Arsenic thiolation

In sulfide−rich geothermal environments, thioarsenates (H_3AsSnO_{4-n}) are the dominant As species. Given that they contain sulfide and As(V), they can act both as electron donors and acceptors. However they decompose easily if exposed to chemical oxidizers, high pH or sulfur−oxidizing bacteria (Ospino et al., 2019).

Along with being oxidized, reduced and thiolated, As can be incorporated into organic compounds, denominated organoarsenicals. Because the present work focuses on the initial step of this transformation, As methylation, the topic will be developed in a separate section.

1.5 Arsenic methylation

1.5.1 Pathways and enzymatic mechanism

Methylated As species have been detected as products of inorganic As transformation by bacteria, archaea (Meyer et al., 2008), fungi (Bentley & Chasteen, 2002), algae (Miyashita et al., 2012), cyanobacteria (Ye et al., 2012), animals (Lin et al., 2002), and humans (Aposhian et al., 2000). The methylation of As consists of the transfer of up to three methyl groups (CH_3) to As(III) in a cascade of reactions (dark green arrows in Figure 1.2). The reaction is catalyzed by the enzyme As(III) S−adenosylmethionine (SAM) methyltransferase, usually denominated ArsM in prokaryotes and As3MT in eukaryotes. The products of the reaction are shown in Figure 1.4. Trimethylarsine, the end product of the reaction (red arrows in Figure 1.5), is volatile and thus, As methylation can lead to the mobilization of As from soil and groundwater into the air.

The first prokaryotic ArsM heterologously expressed was from *Rhodopseudomonas palustris* (Qin et al., 2006). Its was expressed in an arsenic−hypersensitive strain of *Escherichia coli* named AW3110(DE3), whose *ars* operon (*arsRBC*) had been deleted (Carlin et al., 1995). The expression of ArsM in the As−sensitive strain conferred As(III) resistance to the strain and SAM was confirmed as the methyl donor for As(III). This finding supported the hypothesis that As methylation serves to detoxify As(III).

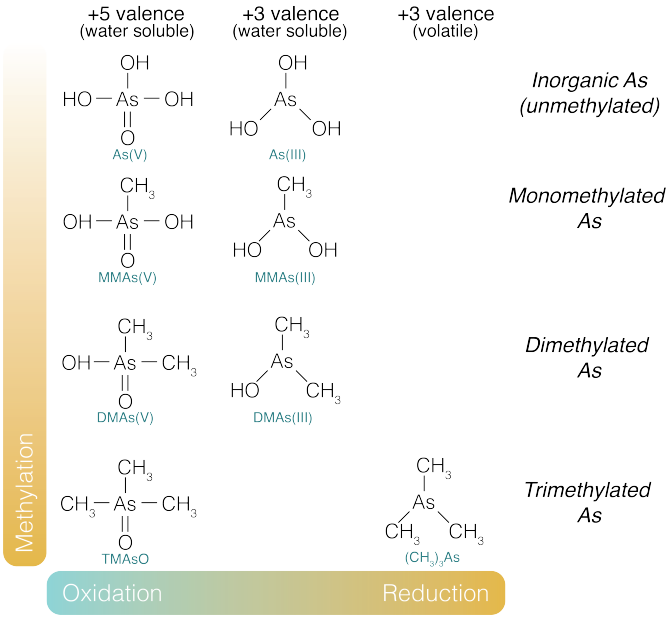


Figure 1.4 – Methylated arsenic species.

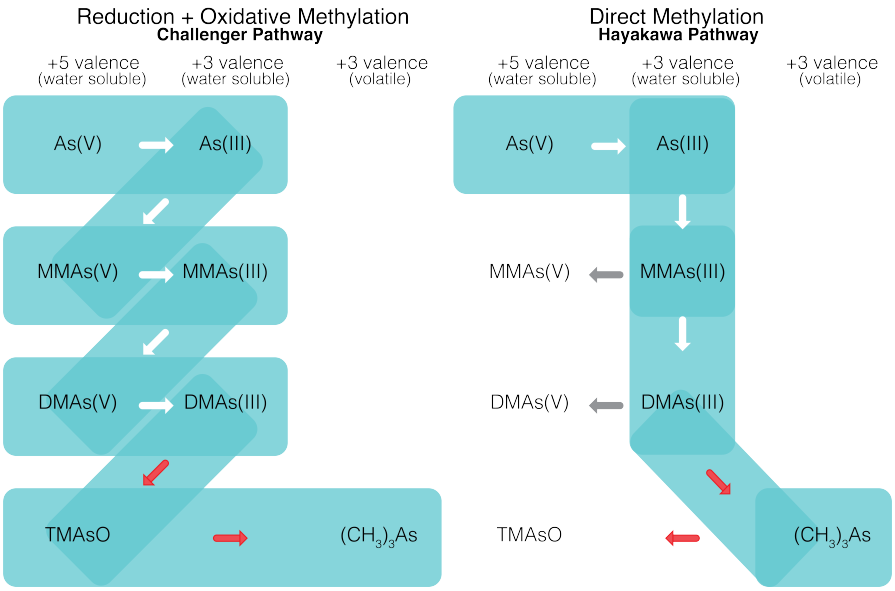


Figure 1.5 – Proposed arsenic methylation pathways.

The first As methylation pathway was proposed by Challenger, 1951 (left panel in Figure 1.5). In this pathway As(III) undergoes a series of oxidative methylation steps, with the pentavalent forms MMAs(V), DMAs(V) and TMAsO as intermediate products which are reduced to their trivalent forms and, in turn, become the substrates for the next step. In a second proposed pathway (right panel in Figure 1.5), Hayakawa et al., 2005, suggest that the substrates are the trivalent form complexed with glutathione, $As(GS)_3$, $MAs(GS)_2$ and DMAs(GS). In this scenario,

methylation occurs without a change in oxidation state of the substrate and oxidation to pentavalent forms would takes place post−reaction.

Most of the studies to elucidate the molecular mechanisms of As methylation have been performed using the eukaryotic ArsM from the red algae *Cyanidioschyzon sp.* 5508 (CsArsM) (Packianathan et al., 2018). In this model, four cysteine residues, found in most known ArsM, corresponding to Cys44, Cys72, Cys174 and Cys224 (Figure 1.6) are involved in the reaction. Although CmArsM methylates As(III) to a trimethylarsine, the steps of the catalysis proposed so far are for the synthesis of DMAs(V) and MMAs(V).

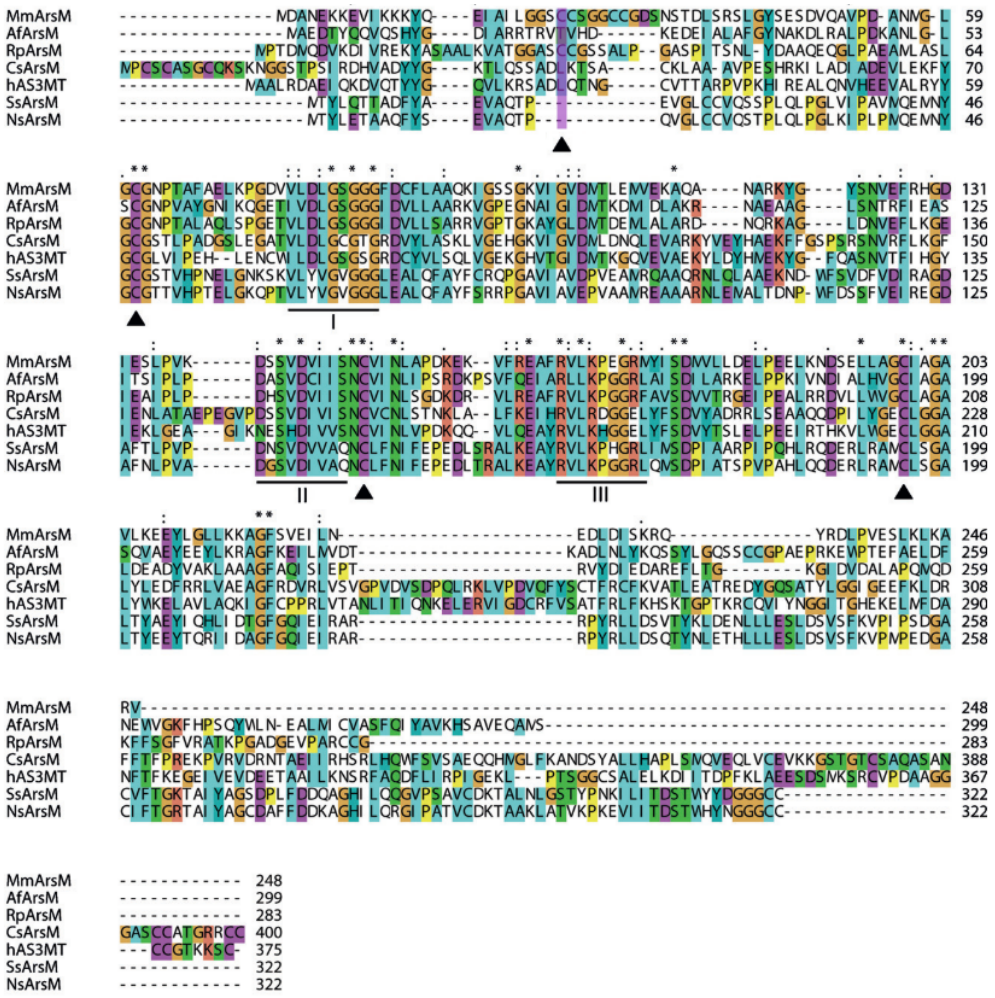


Figure 1.6 – Multi−sequence alignment of ArsM from different organisms. The three conserved motifs are underlined and labeled I, II, and III. The conserved cysteine residues are indicated by solid triangles and correspond to the Cys44, Cys72, Cys174 and Cys224 in *Cyanidioschyzon sp.* 5508. The sequences from seven species are: AfArsM (*Aspergillus fumigatus* A1163), CsArsM (*Cyanidioschyzon sp.* 5508), hAS3MT (*Homo sapiens*); MmArsM (*Methanosarcina mazei* Gö1), NsArsM (*Nostoc sp.* PCC 7120), RpArsM (*Rhodopseudomonas palustris* CGA009), and SsArsM (*Synechocystis sp.* PCC 6803). Modified Fig. 1 from (Ye et al., 2012).

The steps of the catalysis of DMAs(V) formation, as proposed in (Marapakala et al., 2015; Packianathan et al., 2018), are shown in Figure 1.7:

1. As(III), complexed with glutathione (As(GS)3), binds to Cys44, Cys174 and Cys224.
2. The methyl group of SAM is attacked by the arsenic ion pair.
3. As(III) is oxidatively methylated to form MMAs(V).
4. MMAs(V) is reduced to enzyme−bound MMAs(III), bound to Cys174 and Cys224, with electrons from Cys72 that generates a disulfide bond with Cys44.
5. The Cys 44−Cys72 disulfide bond is re−reduced by thioredoxin.
6. MMAs(III) is oxidatively methylated to form DMAs(V).
7. DMAs(V) is reduced to DMAs(III), bound to Cys224, with electrons from Cys174 that generates a disulfide bond with Cys72.
8. The Cys72−Cys174 disulfide bond is re−reduced by thioredoxin and DMAs(III), which is weakly bound, is released.

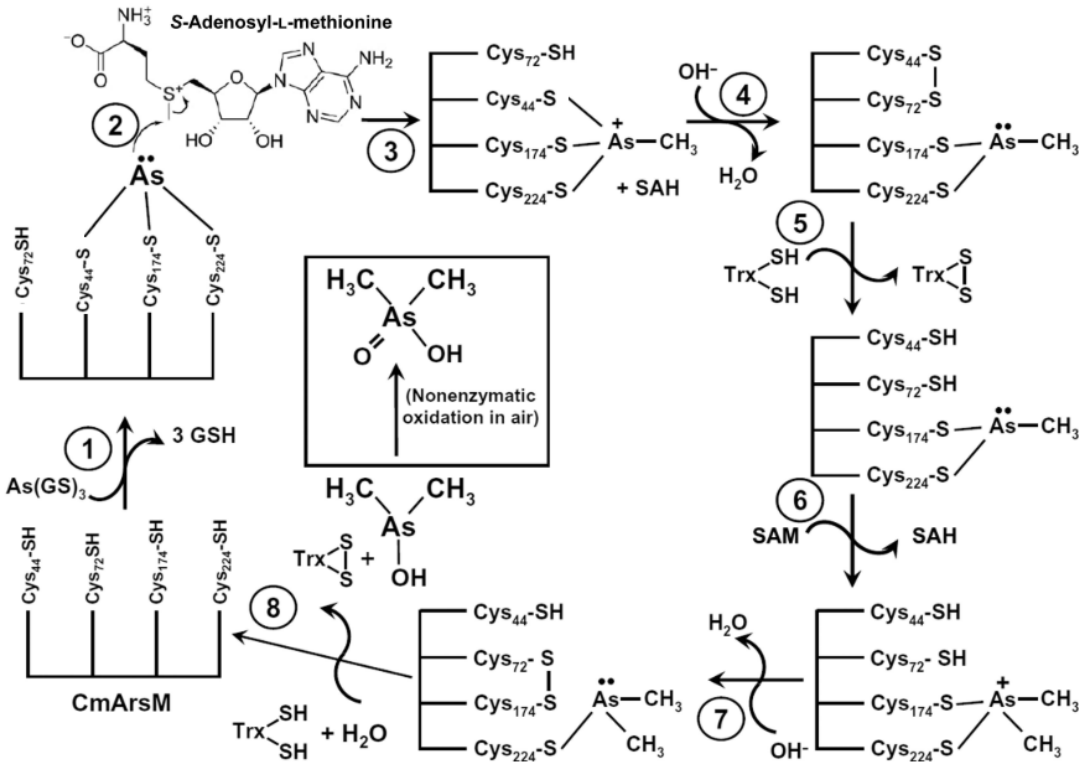


Figure 1.7 – Proposed mechanism for ArsM enzymatic reaction. Numbering of cysteines corresponds to *Cyanidioschyzon sp.* 5508. (Modified Fig. 7 from (Marapakala et al., 2015)).

In this model, As(III) complexed with glutathione, $As(GS)_3$ is hypothesized to be the substrate for CmArsM because it has been shown to bind more rapidly to the enzyme than free As(III) (Marapakala et al., 2012). Thus, the proposed overall reaction is an oxidative methylation but the pentavalent species are transient and enzyme−bound, substrates and products are trivalent and the oxidized species are formed post−reaction as side products.

Though the proposed mechanism is based on CmArsM, the amino acid sequences of ArsMs are highly conserved and there are three motifs in the enzyme (denoted by I, II and III in Figure 1.6) apparently preserved in ArsMs from all species (Ye et al., 2012). Mutation in any of the four cysteine residues from CmArsM prevented the methylation of As(III) to MMAs(III), but a mutation in Cys44 or Cys72 allowed the methylation of MMAs(III) to DMAs(III). Hence, Cys44 and Cys72 are involved in the first methylation step. Single−residue mutagenesis experiments in ArsMs from *Clostridium sp.* BXM and *Methanosarcina acetivorans* C2A, encoding cysteine residues equivalent to Cys72, Cys174 and Cys224, have shown similar results. All three protein mutants, where cysteine residues were replaced, lost their As(III) methylation capacity while mutations in Cys174 and Cys224 did not prevent the methylation of MMAs(III) (P. P. Wang et al., 2015; P. P. Wang et al., 2014). Thus, three cysteine residues have been preserved in all known ArsMs making possible the methylation of As(III) (C. Chen et al., 2017). However, the role and evolution of these four preserved cysteine residues needs further research.

In the ArsM from *Bacillus sp.* CX−1, encoding the last three CmArsM preserved residues, only residues equivalent to Cys174 and Cys224 were needed for the methylation of As(III) and MMAs(III) to DMAs(III) (K. Huang et al., 2018). While one of the four ArsMs encoded in *Aspergillus fumigatus* Af293, although presenting three of the four Cyst residues, could only methylate MMAs(III) to DMAs(III) but not As(III). When Cys72 was replaced, MMAs(III) was methylated whilst when Cys174 and Cys224 residues were replaced the enzyme's methylation capability was lost (C. Chen et al., 2017).

C. Chen et al., 2017, hypothesized that a two−cysteine ArsM, with only Cys174 and Cys224, could have been the ancestor of three− and four−cysteine enzymes. The appearance of a third−cysteine by evolutionary mutation would have had small repercussions, from what is observed in *Aspergillus fumigatus*, but the four−cysteine would have had a beneficial effect making As(III) more efficient. A second hypothesis is that three−cysteine ArMs are predecessors of four−cysteine ArsMs as the result of loss−of−function mutation.

1.5.2 Function of arsenic methylation

Efforts to elucidate the catalytic mechanism of eukaryotic ArsM are important given the significant difference between the toxicity of the oxidized and reduced methylarsenicals when compared to the inorganic species (see Tables 1.1 and 1.2). The current order of toxicity is MMAs(III)> DMAs(III)>As(III)>As(V)> MMAs(V)>DMAs(V). If the more toxic trivalent species are the main products of As methylation intracellularly, the health risk associated with human exposure to inorganic As would be significantly enhanced. As mentioned earlier, trivalent

methylarsenicals have been found in human urine but were oxidized once exposed to air (Le et al., 2000).

This disparity in the toxicity between of pentavalent and trivalent methylated As species also rises a major question as to the function of ArsM in prokaryotes. The reaction could lead to As detoxification via (i) the direct synthesis of the pentavalent forms, (ii) the rapid oxidation of the trivalent forms in oxic environments, or (iii) the removal of intracellular As via volatilization as trimethylarsine. Thus, methylation seems to be a viable detoxification strategy for aerobic microorganisms, which indeed represent the majority and the most efficient As methylators of the single species isolated from the environment (Table 1.4).

In the case of anoxic environments, trivalent methylarsenicals could persist and negatively impact microorganisms exposed to them. Indeed, MMAs(III) has been proposed as a toxin capable of killing off the competitors of anaerobic As methylators (J. Li et al., 2016), if they lack MMAs(III) detoxification strategies. According to a phylogenomic study of *arsM*'s evolutionary history (S. C. Chen et al., 2017), ArsM evolved within the anoxic environment of the early Earth. The hypothesis states that ArsM would have evolved to synthesize highly toxic MMAs(III). Once the atmosphere became oxidizing, As(III) methyltransferase−encoding genes would have been maintained as a competitive advantage against organisms sensitive to trivalent methylarsenicals.

Supporting this theory is the discovery of soil bacteria able to detoxify MMAs(III). The strains *Streptomyces sp.* MD1 and *Bacillus sp.* MD1, isolated from golf course soils (Yoshinaga et al., 2011; Yoshinaga & Rosen, 2014), exhibited MMAs(III)−demethylating activity. The responsible enzyme was characterized in *Bacillus sp.* MD1 and named ArsI. The ArsI enzyme could also demethylate the reduced form of the drug roxarsone. When *Streptomyces sp.* MD1 was grown in co−culture with a soil bacterium *Burkholderia sp.* MR1, which has shown the capability to reduce MMAs(V) to MMAs(III) under aerobic conditions, it could also demethylate MMAs(V) and monosodium methyl arsonate (MSMA).

ArsH, is an MMAs(III) oxidase, found in *Pseudomonas putida* (J. Chen, Bhattacharjee, et al., 2015) and *Serratia meliloti* (H. C. Yang et al., 2005). Heterologous expression of ArsH in *E. coli* AW3110(DE3) renders the strain resistant to MMAs(III), and to reduced roxarsone and phenylarsenic. The *arsH* gene is commonly found in *ars* operons (S. C. Chen et al., 2017).

A third and last mechanism found so far for MMAs(III) detoxification are the membrane transporters ArsP and ArsK. The enzyme ArsP, a MMAs(III) efflux permease, it was first identified as a resistance mechanism for oxidized roxarsone in *Campylobacter jejuni* (Shen et al., 2014). When expressed in *E. coli* AW3110(DE3), it provided resistance to MMAs(III) and to reduced roxarsone but the bacterium remained sensitive to inorganic As(III) (J. Chen, Madegowda, et al., 2015). The ArsK transporter, identified in *Agrobacterium tumefaciens* GW4 (Shi et al., 2018), provided resistance to and its expression was induced by As(III), Sb(III), reduced roxarsone and MMAs(III). Its heterologous expression did not provide resistance to As(V) nor DMAs(III).

Table 1.4 – Known As−methylating single−species and community cultures.

Oxygen requirement	Strain or culture type	% methylation (maximum)	Major products	Incubation time	Study
Anaerobe	Enrichment culture from lake sediment	12%	DMAs(V)	35d	(Bright et al., 1994)
Anaerobe	*Clostridium sp.* BXM	10%	MMAs(V)	20d	(P. P. Wang et al., 2015)
Anaerobe	*Methanosarcina acetivorans* C2A	13%	MMAs(V)	20d	(P. P. Wang et al., 2014)
Anaerobe	Mixed culture enriched from paddy soil (EA)	63%	DMAs(V)	4d	(Reid et al., 2017)
Aerobe	*Pseudomonas alcaligenes*	80%	DMAs(V)	4d	(J. Zhang et al., 2015)
Aerobe	*Streptomyces sp.*	67%	MMAs(V), DMAs(V)	12d	(Kuramata et al., 2015)
Aerobe	*Arsenicibacter rosenii* SM−1	99%	$(CH_3)_3As$	24h	(K. Huang et al., 2016)
Aerobe	*Pseudomonas putida* (engineered strain with *R. palustris arsM*)	97%	DMAs(V)	16h	(J. Chen et al., 2013)
Aerobe	*Bacillus sp.* CX−1	99%	(DMAs(V)	24h	(K. Huang et al., 2018)
Facultative	*Shewanella oneidensis* MR−1 (do not encode an *arsM* gene)	27.9%	(DMAs(V)	8d	(J. Wang et al., 2016)

It is worth noticing that all the mentioned representative ArsI, ArsH, ArsP and ArsK enzymes were identified in studies with aerobic strains. ArsI and ArsH are monooxygenases requiring oxygen, suggesting they most likely evolved after the appearance of oxygen in the atmosphere. Homologues to *arsH* have been found to be scarce, although existent, in anaerobes coexisting

with *arsM* (S. C. Chen et al., 2017) whilst to date, all organisms encoding *arsI* are aerobic bacteria (Yoshinaga & Rosen, 2014). However, the facts that *arsH* and *arsI* are found only in *ars* operons and present high levels of activity when MMAs(III) is used as substrate, suggest the genes must have arisen as a response to reduced organoarsenicals. The hypothesis is that the production of reduced methylated As was the driving force. Because these highly toxic species are more stable under anoxic conditions, it is possible that alternate pathways to ArsI and ArsH exist in anaerobic strains. Finally, homologues to *arsK* have been found in facultative anaerobes (Garbinski et al., 2019) and *arsP* has been found in anaerobic microorganisms coexisting with *arsM* (S. C. Chen et al., 2017).

If production of MMAs(III) were a defense mechanism for anoxic As methylators, it is reasonable to think that, in addition to As(III), inter-species regulation of the *arsM* gene might be needed to trigger As methylation. The latter could explain the lack of anaerobic As−methylating representatives isolated from nature (Table1.4).

The *arsM* gene can be found as part of *ars* operons whose expression is generally controlled by ArsR and induced in the presence of As(III). However, it also can be encoded outside of it, as is the case of *Bacillus sp.* CX−1. In the *Bacillus sp.* CX−1's study, the authors interpret the lack of change in *arsM* gene expression with increasing As(III) concentrations, and the fact that *arsM* is not adjacent to the *ars* operon, as meaning that the gene must be constitutively expressed and not controlled by the transcriptional repressor ArsR (K. Huang et al., 2018). The latter interpretation may be flawed, as it is well established that the regulatory gene from an operon need not be in the same location on the chromosome (Lawrence, 2002). As an example, the *arsBHC* operon from *Synechocystis sp.* strain PCC6803 is regulated by an *arsR* encoded outside the operon (López-Maury et al., 2003). It is unclear at the moment how *arsM* expression is regulated. In some of the few organisms it was examined, *arsM* expression was induced by As(III) (K. Huang et al., 2016; Yin et al., 2011) but in others there was no or little change of expression (K. Huang et al., 2018; J. Zhang et al., 2015).

Finally, As methylation also plays a role in the synthesis of other organoarsenicals in cyanobacteria. After deletion of the *arsM* gene from the cyanobacterium *Synechocystis sp.* PCC 6803, which produces two species of arsenosugars, the Δ*arsM* mutant was only able to synthesize the arsenosugars if provided with DMAs(V) but not MMAs(V) or As(III) (X.-M. Xue et al., 2017), indicating DMAs(V) as the immediate precursor. Recently, a putative radical SAM enzyme encoded adjacent to *arsM* forming an *ars* operon, ArsS, was discovered and shown to be involved in the arsenosugar synthesis by *Synechocystis sp.* PCC 6803. It is worth noticing that the *ars* operon encoding *arsM* and *arsS* was not induced by As(III) (X. M. Xue et al., 2019). Furthermore, the *Synechocystis sp.* PCC 6803 Δ*arsM* mutant lost 3−35% of its lipid−soluble arsenic content relative to wild−type and no arsenolipids could be detected in the mutant (X. M. Xue et al., 2014).

1.5.3 Arsenic-methylating anaerobic microorganisms

The interest in arsenic methylation dates to the mid 1800s, after the fatal intoxication of people allegedly through inhalation of trimethylarsine, a product of the As methylation of the arsenic in wallpaper pigments by fungi. In the last decade, interest has been amplified by the detection of methylarsenicals in rice. The methylated As in rice originates from the microbes in the soil and rhizosphere. In fact, no As(III) methyltransferases have been identified in higher plants (Lomax et al., 2012; Ye et al., 2012). This fact and the unresolved function of ArsM under anoxic conditions has made rice paddy soils an environmentally important and representative habitat in which to study the microbial drivers of As methylation.

Only two anaerobic pure cultures have been reported to methylate in the literature (Table 1.4). These are a sulfate−reducing bacterium (SRB), *Clostridium sp.* BXM (P. P. Wang et al., 2015), no longer available for further research (personal communication with the authors of the publication); and a methanoarchaeon, *Methanosarcina acetivorans* C2A, reported to methylate after 20 days (P. P. Wang et al., 2014).

Michalke et al., 2000 have reported that other pure cultures of anaerobic bacteria are capable of As methylation, based on the detection of volatile trimethylarsine. The study included *Methanobacterium formicicum*, *Desulfovibrio vulgaris* DSMZ 644, *Desulfovibrio gigas* DSMZ 496 and *Clostridium collagenovorans* DSMZ 3089. The results reported the strains were able to volatilize <0.3% of the initial 0.05−0.5 mM As(V). However, the genome of the *D. vulgaris*, *D. gigas* and *C. collagenovorans* strains do not encode genes annotated as *arsM* nor aligning to *arsM* sequences encoding active ArsM proteins (our unpublished data). Therefore, the observed methylation may not be ArsM−mediated. For example, SRB are able to use methylcobalamin to methylate mercury (Choi & Bartha, 1993) and arsenic can be chemically methylated by methylcobalamin (Zakharyan & Aposhian, 1999). Furthermore, methanogens have been observed to volatilize metals in low quantities (as methylated metal species) but this process was deemed fortuitous, due to the multiplicity of their methyl transfer reactions rather than deliberate metal−specific enzymatic processes (Meyer et al., 2008; Thomas et al., 2011).

The use of degenerate *arsM* primers has been the most common approach to finding As-methylating microorganisms in rice paddies (C. Chen et al., 2019; Jia et al., 2013; Reid et al., 2017; M. Wang et al., 2019; S. Y. Zhang et al., 2015; Zhao et al., 2013). Most of these studies have shown that the *arsM* gene in rice paddy soils is abundant and belongs to taxonomically diverse microorganisms with no specific species or metabolism identified (Jia et al., 2013; S. Y. Zhang et al., 2015; Zhao et al., 2013). In two studies, SRB are strongly suggested to be drivers of anoxic As methylation based on the correlation between the *arsM* and *dsrB* (dissimilatory sulfite reductase b−subunit) copy numbers, and on the decrease in methylation after addition of molybdate and monofluorophosphate (MFP), chemical inhibitors of sulfate reduction (C. Chen et al., 2019; M. Wang et al., 2019).

Finally, there is only a single study using metagenomic analysis to look at the phylogeny of

arsenic resistance genes in rice paddy soils (Xiao et al., 2016) where the most abundant *arsM* gene sequence belonged to the Actinobacterium *Candidatus* Solibacter usitatus Ellin6076 (YP_825656).

1.6 Thesis outline

Understanding the dynamics of microbial As transformation in rice paddy soils is key in the efforts to mitigate the threat arsenic in rice represents to public health. Unraveling the role of As methylation in soil microorganisms could provide us with the knowledge needed to apply this singular reaction as a mitigation strategy for high arsenic levels in rice.

The present work results from efforts to identify the As methylation drivers in the environment. The research is divided in two parts. The first part consists of the thorough study of seven different microbial strains as possible As methylators. All of the strains, harboring varied metabolisms, encode the *arsM* gene.

The microorganisms were tested for their resistance to As(III), based on their growth in the presence of arsenic, and for their ability to methylate the metalloid. A clear negative correlation between the efficiency of As methylation and As(III) resistance was observed. This finding sheds light into the role of efficient As(III)−efflux systems may play a role in the dynamics of anaerobic As methylation. To confirm the influence of the efflux of As(III) out of the intracellular space, a mutant lacking the As(III)−efflux protein was constructed, providing further insight into the question.

The second part of the research focuses on microbial communities derived from rice paddy soils. This is because microbes exert their influence on the As biogeochemical cycle in nature, as part of a complex environmental community. In this part, we have made use of meta−omic techniques, specifically metagenomics, metatranscriptomics and metaproteomics. We provide insights into the As methylation dynamics of an environmentally−representative system, moving beyond the interferences encountered in complex environmental systems such as soil.

2 Variability in arsenic methylation efficiency across aerobic and anaerobic microorganisms

Karen Viacava[1], Karin Lederballe Meibom[1], David Ortega[3], Shannon Dyer[1], Arnaud Gelb[4], Leia Falquet[1], Nigel P. Minton[3], Adrien Mestrot[2], Rizlan Bernier-Latmani[1].

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

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

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

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

Authors' contributions. KV conducted the experimental work and laboratory analyses. KLM conducted the cloning of the *arsM* genes in *E. coli* and the RT−qPCR of *C. pasteurianum*'s *arsM* and *acr3*. SD and LF assisted with experimental work. AG contributed to the planning of flow cytometry analyses and acquired the fluorescence microscopy pictures. DO constructed and complemented the *C. pasteurianum* Δ*acr3* mutant under the supervision of NM. AM contributed to project conception and development. RB-L conceived and supervised the study. KV wrote the manuscript with the help and advice of KLM and RB-L.

This is a preprint version of the manuscript to be submitted for publication.

2.1 Abstract

Microbially−mediated methylation of arsenic (As) plays an important role in the As biogeochemical cycle, particularly in rice paddy soils where methylated As generated by soil microorganisms is translocated into rice grains. The presence of arsenite (As(III)) methyltransferase genes (*arsM*) in soil microbes has been used as an indication of their capacity for As methylation. Here, we evaluate the ability of seven microorganisms encoding the *arsM* gene to methylate As. Whilst the arsM genes from all seven species encoded active enzymes, only the aerobic species were efficient methylators, likely as a detoxification strategy. The anaerobic microorganisms presented high resistance to As exposure presumably due to their efficient As(III) efflux but methylated As poorly, except methanoarchea, for whom efficient As methylation was seemingly an artifact of membrane disruption. Deletion of an efflux pump (*acr3*) in one of the anaerobes, *Clostridium pasteurianum*, rendered the strain sensitive to As and capable of more efficient As methylation. These results show that encoding a functional ArsM enzyme does not guarantee a microorganism will effectively drive As methylation in the presence of the metalloid. Furthermore, we propose an inverse relationship between efficient microbial As efflux and its methylation, because the former prevents the intracellular accumulation of As.

2.2 Introduction

Early life evolved in waters rich in reduced arsenic (Kulp, 2014; Oremland et al., 2004). In modern times, nearly every organism, prokaryotes and eukaryotes alike, possesses strategies to tolerate or utilize this ubiquitous toxic element (Bhattacharjee & Rosen, 2007). Arsenic detoxification and its harnessing as an electron donor or acceptor often result in its chemical transformation. Microbial As transformations include arsenate (As(V)) reduction and arsenite (As(III)) oxidation (Mukhopadhyay & Rosen, 2002; Van Lis et al., 2013) as well as As(III) methylation by S−adenosyl methionine methyltransferases (ArsM in prokaryotes) (Ajees & Rosen, 2015). Additionally, As(III) detoxification can occur via its efflux by membrane transporters (ArsB and Acr3) (Garbinski et al., 2019). Taken together, these microbial transformations contribute to the arsenic biogeochemical cycle (Y. G. Zhu et al., 2017).

Arsenic cycling is of particular interest in rice paddy soils because the rice plant is an efficient accumulator of As when compared to other cereals (M. Z. Zheng et al., 2013). Arsenic is found in rice grains in varying ratios of inorganic arsenic (iAs) and dimethylarsenate (DMAs(V)) (Zhao et al., 2013), the latter derived from arsenic methylation by paddy soil microorganisms (Lomax et al., 2012). Arsenic methylation is promoted by the anoxic conditions prevalent in flooded rice paddy soils (Mestrot et al., 2009; X. Y. Xu et al., 2008). Methylarsenicals are toxic to rice plants causing sterility (straight−head disease) and a decreased yield (W. Yan et al., 2005). Moreover, volatile methylarsines can be produced via As methylation, causing potential airborne exposure of humans to As (Mestrot et al., 2013). The function of As methylation remains poorly understood. Arsenic detoxification has been proposed as a possible function

partly because As methyltransferase genes are often found in operons providing As resistance (*ars* operons) or associated with the arsenic regulator *arsR* (Y. G. Zhu et al., 2017). However, their chemical products are trivalent methylated arsenicals (Packianathan et al., 2018) which are potent genotoxins (Mass et al., 2001) and mammalian cell toxins (Styblo et al., 2000). Thus, the transformation actually increases the toxicity of As and its carcinogenicity. On the other hand, the toxic trivalent methylated As species can be oxidized to the less toxic pentavalent methylated species (Cullen & Reimer, 1989). An alternative rationale for As methylation has been proposed by which the increase in As toxicity associated with methylation may be a strategy to kill off microbial competitors. Indeed, the recent identification of enzymes providing resistance to monomethylarsenite (MMAs(III)), reflecting a possible adaptation to this type of microbial chemical warfare, provides support for this theory (J. Chen et al., 2019; J. Li et al., 2016).

Efforts to identify key microbial mediators of As methylation have involved metagenomic (Xiao et al., 2016) as well as gene amplification approaches (Jia et al., 2013; Reid et al., 2017; M. Wang et al., 2019). These studies have shown that the phylogenetic diversity of *arsM* genes from paddy soil microbial communities is extensive. To pinpoint active As methylators, the studies correlated the abundance of *arsM*−harboring microorganisms to As content in soil or other soil geochemical characteristics (S. Y. Zhang et al., 2015). Unfortunately, to date, the organisms actively methylating As in anoxic soils remain unknown. In fact, despite the fact that the *arsM* gene is relatively commonly found in soil microorganisms, the isolation of active As methylators from soil has only succeeded for aerobic bacteria (K. Huang et al., 2016; K. Huang et al., 2018; Kuramata et al., 2015). This is notwithstanding that As methylation is greater under anoxic conditions (Mestrot et al., 2009; X. Y. Xu et al., 2008) and the fact that great emphasis has been put on isolating As−methylating sulfate−reducing bacteria (Bright et al., 1994; C. Chen et al., 2019). To our knowledge, only one As−methylating anaerobic strain has been isolated from paddy soils, *Clostridium sp.* BXM (P. P. Wang et al., 2015) and it appears to have been lost (Dr. Peng Wao and Prof. Guo-Xin Sun, pers. commun.). With the aim of identifying microbes likely to drive As methylation in the paddy soil environment, this study presents a systematic investigation of As methylation by pure cultures of *arsM*−harboring microbial species. The species considered vary in their metabolism and their phylogeny and were selected to represent some of the diversity of *arsM*−harboring soil microbes. We find that while ArsM was functional in all microorganisms studied, only the aerobes carried out efficient methylation. Furthermore, we show that the As(III) efflux pump found in *Clostridium pasteurianum*, and by extension in the other anaerobes, precludes ArsM activity by limiting As(III) accumulation intracellularly. Thus, the findings in this study bring novel insights into the conditions needed for active As methylation in *arsM*−harboring microorganisms.

2.3 Materials and methods

2.3.1 Cultures and growth conditions

Five of the seven strains tested were purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ, Germany) and grown in their recommended liquid medium and incubation temperatures: *Geobacter metallireducens* GS−15 (DSM 7210), *Anaeromusa acidaminophila* DSM 3853, *Streptomyces vietnamensis* DSM 41927, *Methanosarcina mazei* Gö1 (DSM 3647) and *Methanosarcina acetivorans* C2A (DSM 2834). *Arsenicibacter rosenii* SM−1 was kindly provided by Prof. Fang-Jie Zhao. *Clostridium pasteurianum* DSM 525 was part of the strain collection from the Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC) in Nottingham, UK. The latter two strains were grown in R2A liquid medium (0.5 g l^{-1} yeast extract, 0.5 g l^{-1} tryptone, 0.5 g l^{-1} casamino acid, 0.5 g l^{-1} D-glucose, 0.5 g l^{-1} starch, 0.3 g l^{-1} K_2HPO_4, 0.05 g l^{-1} $MgSO_4 \cdot 7H_2O$, 0.3 g l^{-1} sodium pyruvate) and 2xYTG liquid medium (16 g l^{-1} tryptone (Oxoid Ltd), 10 g l-1 yeast extract (Oxoid Ltd), 5 g l^{-1} NaCl and 5 g l^{-1} D−glucose, pH adjusted to 6.2) respectively, and incubated at 37 °C. The strain *E. coli* AW3110−(DE3) (Δ*arsRBC*) was kindly provided by Prof. Barry Rosen and grown at 37 °C in Luria−Bertani (LB) broth. For anaerobic strains, the medium was brought to a boil, cooled down to room temperature under gas flow (100% N_2 or 80%:20% N_2:CO_2) and, in 100 mL, were dispensed into 200−ml serum bottles (100 ml of medium per bottle) under the same gas atmosphere. The bottles were sealed with butyl rubber stoppers and aluminum crimps, and autoclaved at 121 °C for 30 min. Prior to autoclaving, the headspace of the serum bottles containing anaerobic media was flushed with the recommended gas. For anaerobic strains, all culture manipulations were carried out either under anoxic conditions in an anaerobic chamber (COY Laboratory Products, MI, USA) or in sealed serum bottles and using thoroughly N_2−flushed syringes and needles.

2.3.2 Arsenic methylation by pure strains

The As methylation assays consisted of time−course experiments, one per strain. The growth media were amended with 2, 5, 10 or 50 μM As(III) as sodium arsenite and a no−As(III) control was included (three replicates per strain). Bottles were inoculated with 1% inoculum (v/v) of a pre−grown exponential phase culture. Samples were obtained for soluble and volatile As as well as microbial growth. Growth was quantified using optical density at 600 nm (OD_{600}) or with the Pierce BCA Protein Assay Kit (Thermo Scientific, MA, USA) after cell lysis (lysis buffer: 0.1% triton X−100, 0.1% SDS, 10 mM EDTA and 1 mM TrisHCl) at 95 °C for 10 min with vortexing every 3 min.

2.3.3 Dissolved and volatile As species sampling

Soluble As samples were filtered (0.22 μm), oxidized by adding 10% (v/v) of 30% H_2O_2 (w/v) and preserved in 1% HNO_3 (v/v) (>69%, Trace Select, Honeywell Fluka Analytical, Switzerland)

at 4°C. To sample for volatile As in aerobic strains, a continuous flow of 0.22 μm−filtered air was pumped into the medium via a syringe (100 l h^{-1}) (Eheim air pump 100, Germany) and the air outflow was passed through a chemo−trap (sorbent tube containing silica gel (SKC Ltd, UK) impregnated with 1% $AgNO_3$ (w/v)) (Mestrot et al., 2009). For anaerobic strains, the bottle head−space was flushed with 100% N_2 at 100 kPa (1 bar) for 10 min at the end of the experiment while the outflow was passed through a chemo−trap. The chemo−trap sorbent tubes were carefully cracked open and the silica gel digested in 4.8 ml 1% HNO_3 and 200 μml 35% H_2O_2 in 50−ml centrifuge tubes with loose caps inside a Microwave Accelerated Reaction System (CEM, Matthews, NC, USA). The heating program was: initial temperature 55 °C (ramp 5 min, hold 10 min), middle temperature 75°C (ramp 5 min, hold 10 min) and final temperature 95°C (ramp 5 min, hold 30 min). While the digestion extracts 80.1 to 95.6% of the trapped As and preserves the As-C bond, volatile arsines are oxidized to their corresponding pentavalent ions (Mestrot et al., 2009). Silica gel extracts were filtered with 0.22 μm and stored at 4°C. For each experiment a set of 5 chemo−traps, $AgNO_3$−impregnated but unused, were digested and analyzed as background As values in total As analyses.

2.3.4 Arsenic speciation analysis

Arsenic speciation was obtained using either an ICP−MS 7700x coupled to an HPLC 1260 Infinity or an ICP−QQQ 8900 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}, 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). 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). The As speciation analysis allowed discrimination between inorganic As (detected as As(V): sum of remaining As(III) plus As(V) being produced) and tri−, di− or monomethylated arsenicals (as TMAsO, DMAs(V) or MMAs(V)) but not between the redox state of the methylarsenicals originally produced by the strains. Total As concentrations were measured using the same ICP−MS instruments in stand−alone mode.

2.3.5 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 A.1). The ArsM proteins were aligned against ArsM enzymes proven to be active (Appendix A, Figure A.1) (Multiple Sequence Alignment tool, EMBL−EBI, (Madeira et al., 2019)). 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 A.2). 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 above.

2.3.6 Construction and complementation of Δ*acr3* mutant of *C. pasteurianum*

To create the in−frame deletion of *acr3* (annotated as *arsB*) in *C. pasteurianum* (Table A.3), the RiboCas system was used as described (Cañadas et al., 2019). The CRISPR guide tool of the online software Benchling was used to generate a DNA fragment encoding the single guide (sg) RNA (5'−ATTAGGCCCACCATAAAAGG−3') targeting *acr3* (Sajith Wickramasekara & Li, 2012). A sgRNA fragment was generated using primer 5AA as described and digested using *Sal*I and *Aat*II (Cañadas et al., 2019). Using genomic DNA from wildtype *C. pasteurianum* (WT), left and right regions of homology were cloned consisting of 567bp upstream and 768bp downstream, including the start and stop codon of *acr3* using the primer pairs 1AA, 2AA and 3AA, 4AA (Table A.4). The two fragments were joined together using splicing by overlap extension and digested using *Aat*II and *Asc*I. The vector pRECas1 was digested using *Sal*I and *Asc*I and the sgRNA−encoding fragment and knockout cassette were ligated in a three−way ligation to generate the vector pRECas1_acr3.

To create the complementation of the Δ*acr3* mutant, restoration of uracil autotrophy was used for selection of integration of *acr3* at the *pyrE* locus (Heap et al., 2012). A 1,211bp DNA fragment encoding the *acr3* gene and upstream intergenic region was generated by PCR using the primer pair 7AA and 8AA and digested with *Not*I and *Nhe*I. The resulting digested fragment was ligated into the complementation vector pMTL_KS_12 digested by the same restriction endonucleases to generate the plasmid pMTL_KS_12_acr3 (Schwarz et al., 2017). To create the strains Δ*acr3* and *pyrE*::*acr3*, the methylated vector pRECas1_acr3 was transformed into *C. pasteurianum* WT and *C. pasteurianum* Δ*pyrE* in parallel. The methods used to isolate the mutants Δ*acr3* and Δ*pyrE*::Δ*acr3* are described in Appendix A, subsection A.1.2 and in (Cañadas et al., 2019). The method to assess the methylation of As by Δ*acr3* and Δ*pyrE*::Δ*acr3* are described in Appendix A, subsection A.1.3.

2.3.7 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 GoScriptTM ReverseTranscription System (Promega, WI, USA) and random hexamers as primers. 200 nM of species−specific primers targeting *arsM*, *acr3*, *gyrA* (encoding gyrase subunit A) and *rho* (transcription termination factor Rho), the last two as the reference genes for relative quantification (Table A.5), 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−containing samples vs. no−As controls were calculated by the relative expression software tool (REST©) (Pfaffl et al., 2002) included in the Mic qPCR Cycler software.

2.4 Results and discussion

2.4.1 Arsenic resistance and arsenic methylation by pure strains and overexpressed ArsMs

In order to probe the potential of individual soil microorganisms to catalyze As methylation, we selected species: (a) harboring *arsM*; (b) phylogenetically and metabolically diverse, and; (c) belonging to genera identified in paddy soils. We selected two aerobic heterotrophs: a known As methylator, *Arsenicibacter rosenii* SM−1 (K. Huang et al., 2016), and *Streptomyces vietnamensis* DSM 41927, because this genus is predominant within the Actinobacteria in rice paddies (Ishizawa et al., 1969; Reid et al., 2017) and a *Streptomyces* strain has been shown to methylate As (Kuramata et al., 2015). We also selected five anaerobes, an iron−reducing organism, *Geobacter metallireducens* GS−15 (DSM 7210), two fermenters, one saccharolytic and one peptolytic, *Clostridium pasteurianum* DSM 525 and *Anaeromusa acidaminophila* DSM 3853, and two methanogens, *Methanosarcina mazei* Gö1 (DSM 3647) and *Methanosarcina acetivorans* C2A (DSM 2834) because methanoarchaea have been studied as As methylators (P. P. Wang et al., 2014; Wuerfel et al., 2012). Thus, to probe for active arsenic methylation in these pure cultures, cells were exposed to varying concentrations of As(III). Monitoring their growth allowed us to evaluate their sensitivity to the toxicant. The observed growth of all the anaerobes, *G. metallireducens*, *A. acidaminophila*, *C. pasteurianum*, *M. mazei* and *M. acetivorans*, was not significantly impacted by any of the As concentrations tested (Figure 2.1). In contrast, the growth of the two aerobes varied depending on the As concentration used (Figure 2.1). Growth of *A. rosenii* with 2 μM As(III) was faster than that of the no−As control, not impacted with 5 μM As(III), and inhibited with 10 and 50 μM As(III). Growth of *S. vietnamensis* was also affected by As(III), starting at a concentration of 5 μM As(III).

In *G. metallireducens*, *A. acidaminophila*, and *C. pasteurianum* experiments, the As concentration in the medium remained constant (Appendix A, Figure A.2 shows no As volatilization) and As remained mainly in its inorganic form (Figure 2.2). For *G. metallireducens* and *C. pasteurianum*, for which we report <0.06% methylation efficiency, only the experiments with 50 μM As(III) presented concentrations of methylated arsenicals which were above the limit of

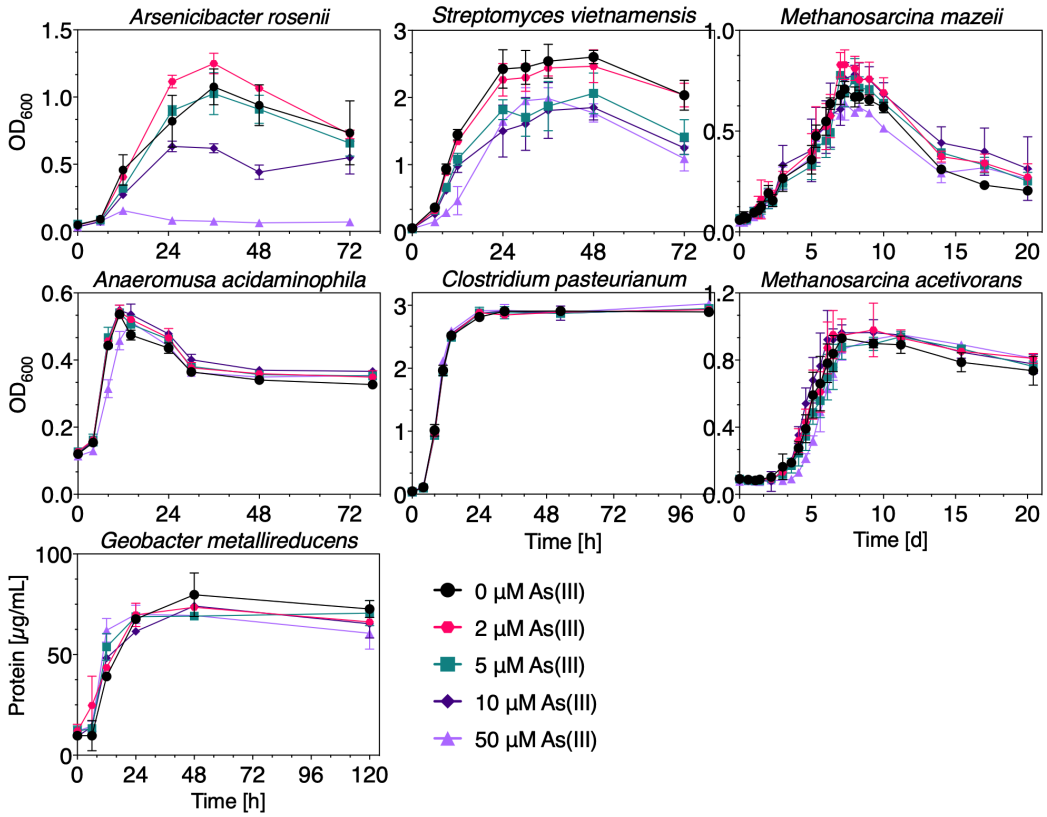


Figure 2.1 – Growth curves (OD_{600} or protein for *G. metallireducens*) as a function of time for each individual species grown in the presence of varying initial As(III) concentrations concentrations (0, 2, 5, 10 and 50 µM As(III)). Data points represent the mean value and error bars, plus and minus one standard deviation.

detection (LOD ~0.007 µM As, 1:1 signal to noise ratio) but below the limit of quantification (LOQ ~0.02 µM As, 1:3 signal to noise ratio) (Appendix A, Figure A.3). Conversely, *A. rosenii* efficiently converted As(III) into organic arsenicals (~10% with 50 µM As(III) and ~60–80% with 2–10 µM As(III)) to soluble and volatile species as previously reported (K. Huang et al., 2016) (Figures 2.2 and in Appendix A, Figure A.4). In the same fashion but to a lesser extent, *S. vietnamensis* transformed some As(III) (~5–10%) mainly to TMAsO (Figure 2.2 and in Appendix A, Figure A.5). For both bacterial species, methylated arsenic was detected before 24 hours, *i.e.*, the conversion took place during the exponential growth phase. Amongst the anaerobes, the two methanogens, *M. mazei* and *M. acetivorans*, were the only ones to exhibit As methylation to an extent comparable to the aerobes. They methylated up to ~5% and ~15% of the initial As(III), respectively (Figure 2.2). In these species, methylation did not start until the stationary growth phase was reached (Figures 2.1 and 2.2). In addition to aqueous arsenic, measurements of volatile arsenic captured in chemo–traps revealed the volatilization of As for *A. rosenii*, *S. vietnamensis*, *M. mazei* and *M. acetivorans*, but As speciation could only be quantitatively determined in *A. rosenii* cultures. Volatilization started from 12h and increased until

the last sampling point at 72h. The main species was trimethylarsine ($(CH_3)_3As$), detected as TMAsO after chemo−trap digestion for all As(III) concentrations, except with 50 μM where dimethylarsine ($(CH_3)_2AsH$), detected as DMAs(V), was predominant (Appendix A, Figure A.4).

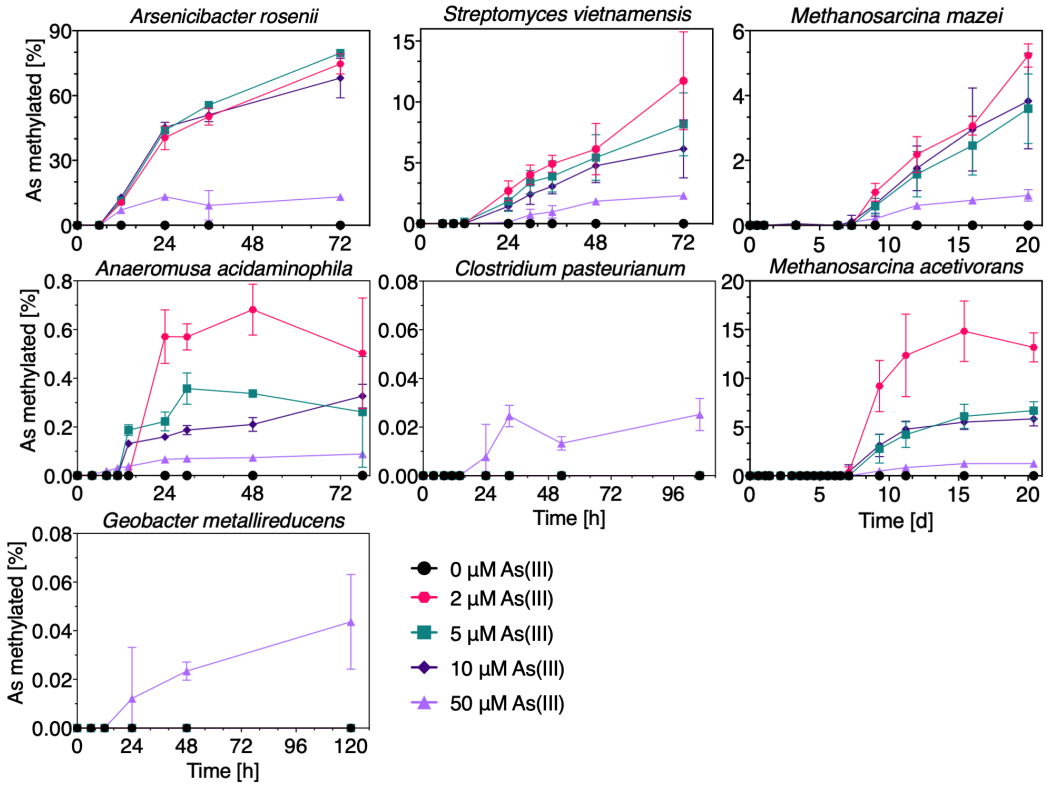


Figure 2.2 – Percentage of methylated arsenic produced (sum of mono−, di− and tri−methylated soluble arsenic species in filtered medium and volatile arsenic species chemo−trapped during medium flushing, relative to total As present) as a function of time for each individual species grown in the presence of varying initial As(III) concentrations (0, 2, 5, 10, and 50 μM). Data points represent the mean value and error bars, plus and minus one standard deviation. Note the variable scale on the y−axis, depending on the species.

To confirm the seven putative ArsM proteins encoded in the selected organisms are indeed functional As(III) methyltransferases, each of them was expressed in the *E. coli* strain AW3110−(DE3) (devoid of any arsenic resistance genes) and probed for As resistance and methylation. The results were unambiguous. As resistance was restored in all cases (Appendix A, Figure A.6) and all the proteins were confirmed to be functional ArsMs (Appendix A, Figure A.7). DMAs(V) (11−84%) and TMAsO (2−71%) were the main products from all microbial ArsMs (Appendix A, Figure A.7). The control, the construct with an empty plasmid, did not exhibit any evidence of methylation.

Arsenic methylation under anoxic conditions was only present in significant amounts in the two archaeal strains. In contrast to the more efficient aerobic As−methylating bacterial

counterparts, it started during the stationary rather than the exponential phase and the main product was monomethylated As rather than trimethylated As (Appendix A, Figures A.8 and A.9). Based on these observations, we propose that, under the present conditions, methanogens may not be actively methylating As during growth. Instead, it is conceivable that during the stationary phase, cells whose membrane becomes compromised, release methyltransferases involved in methanogenesis, *e.g.*, MtaA (Thomas et al., 2011), therefore catalyzing the reaction in the medium rather than within intact viable cells. To assess for membrane damage in *M. mazei*, cells were stained with SYBR Green I (SG), which can be taken up through an intact membrane, as well as with propidium iodide (PI), which cannot. Flow cytometry analysis (material and methods in Appendix A, subsection A.1.4) identified cells with intact membranes (stained with SG only) and cells with compromised membranes (stained with both PI and SG) (Appendix A, Figure A.10). Results for 10 and 50 μM As(III) show that, once the stationary phase starts (11 days), both the proportion of cells stained with both PI and SG and the concentration of methylated As increase (between 16 and 34 days) (Figure 2.3 and in Appendix A, Figure A.11). The increase of double stained cells (PI/SG) at mid−exponential phase (the 5−day sample) could be attributed to changes in the membrane during cell division which disappear at the end of the exponential phase (11−day sample).

Previous work has identified *M. acetivorans* as an As−methylating strain (P. P. Wang et al., 2014) based on activity after 20 days of growth, which is well within the stationary phase. In essence, our results confirm previous findings but offer an alternative interpretation of As methylation by methanogens. The findings here suggest that As methylation by these methanogenic organisms is not an active metabolism carried out by viable cells but rather a fortuitous event taking place due to the release of methyltransferases (ArsM or metanogenesis-related) from dead cells during the stationary phase.

In summary, the results thus far show that, despite encoding at least one confirmed functional ArsM, many of the strains were not able to methylate As efficiently. Therefore, it is likely that the abundance, phylogenetic affiliation, and diversity of *arsM* amplicons reveal little with respect to As methylation activity. Moreover, all the anaerobic organisms we probed were either quite inefficient in methylating As or were likely to methylate only as a result of membrane disruption. Arsenite induces the expression of *ars* operons (Murphy & Saltikov, 2009; Prabaharan et al., 2019; Yu et al., 2015). Thus, a major question emerges from these findings: why doesn't efficient methylation take place in the presence of As(III) in anaerobic microbial species harboring an *arsM* gene that encodes a functional protein?

2.4.2 Relationship between arsenic resistance/sensitivity and arsenic methylation

Methanogens and anaerobic bacteria presented resistance to 50 μM As(III). At the same time, anaerobic bacteria did not exhibit significant active As methylation activity. Their resistance to relatively high concentrations of As suggests the use of mechanism(s) other than methylation to facilitate As detoxification. Under reducing conditions, anaerobes typically

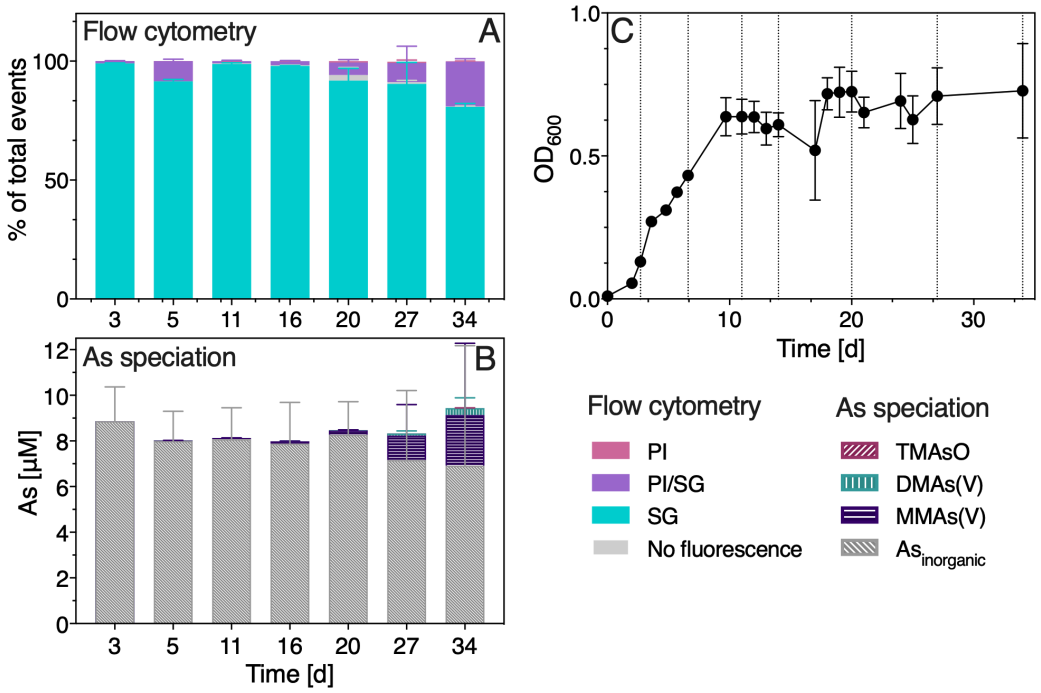


Figure 2.3 – Assessment of membrane integrity for a *M. mazei* culture grown in the presence of 10 µM As(III) as the initial concentration: (A) Relative abundance of four populations (no fluorescence, double-stained (SG/PI) and single-stained: PI or SG) measured by flow cytometry after SG and PI staining 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. A similar dataset is available for 50 µM As(III) as the initial concentration A.11

need to efficiently detoxify the highly toxic and mobile trivalent inorganic arsenic species, As(III). Most of them accomplish this through efficient As(III) efflux systems (Y.-G. Zhu et al., 2014).

Six out of the seven microbial species considered in the present study harbor genes corresponding either to *acr3* or *arsB* As(III) transporters in their *ars* operons (Appendix A, Table A.3). The only exception being *A. rosenii*, the most efficient As−methylating pure culture known to date. We hypothesized that the presence of an efficient As(III)−efflux pump may preclude As(III) methylation, due to insufficient accumulation of the methylation substrate As(III) (Packianathan et al., 2018) in the cytoplasm where ArsM is expected to be active. Hence, all anaerobes, which exhibited very high resistance to As(III), may present highly efficient As(III)−efflux pumps as an adaptation to anoxic environments. In contrast, *S. vietnamensis*, an aerobic bacterium, may harbor a less efficient pump, and thus present lower resistance to As(III) than the anaerobes in question but more resistance than *A. rosenii*, due to the absence of an efflux pump in the latter. This hypothesis is also consistent with the lower efficiency of

methylation of *S. vietnamensis* relative to *A. rosenii*. The intracellular As(III) concentration is likely lower in *S. vietnamensis*, resulting in less substrate for methylation.

Thus, we propose a model by which there is a competition for intracellular As(III) between efflux pumps and methyltransferases. The more efficient the pumping of As(III), the less likely for the methyltransferases to be either expressed or active. In order to provide direct evidence for this model, we constructed a knockout of the sole known As(III) efflux pump (*acr3*) in *C. pasteurianum*.

2.4.3 Phenotype of *acr3* deletion in *Clostridium pasteurianum*: efficient As(III) efflux precludes arsenic methylation.

As described above, the wild−type (WT) strain of *Clostridium pasteurianum* exhibited high resistance to As because comparison of growth curves in the presence of varying concentrations of As (up to 100 μM) evidenced no observable differences (Figures 2.1 and 2.4). In contrast, the addition of 100 μM As(III) hindered the growth of the Δ*acr3* mutant compared to the no−As control (Figure 2.4). The resistance was restored in the complementation strain Δ*acr3 pyrE*::*acr3* (Appendix A, Figure 2.4−A). This finding confirms the role of *acr3* in As(III) detoxification via efflux.

Importantly, the expression of the *arsM* gene (normalized to *gyrA* and *rho* expression) was ~150−fold greater in the Δ*acr3* mutant in the presence of As(III) relative to the no−As control (Appendix A, Figure A.12). The expression of the *gyrA* and *rho* genes (and others candidate reference genes assessed) seemed to be markedly lower in the Δ*acr3* mutant in the presence of As(III), possibly as a result of the noticeable change in growth. Thus, we also calculated the expression ratio without normalization and found a ~14−fold increase in *arsM* expression, confirming the upregulation. In the WT, *acr3* is upregulated (~33−fold), an indication of As(III) efflux taking place as the detoxification strategy. However, in contrast to the Δ*acr3* mutant, no significant upregulation of *arsM* was observed (~1.2 to 1.5−fold increase) in the WT in the presence of As(III) (Appendix A, Figure A.12). This suggests that the disruption of As(III) efflux results in greater expression of *arsM*, which in turn, means that *arsM* expression may be limited by the intracellular concentration of As(III) in the WT strain. This conclusion is further bolstered by measurements of the intracellular concentration of As(III).

Cell pellets from cultures exposed to As(III) were collected, lysed and the As concentration and speciation measured. As concentrations in the lysate from Δ*acr3* were larger than those in WT (Figure 2.4−C) or Δ*acr3* complementation strain lysates (Appendix A, Figure 2.4−C), by a factor of ~1.5−fold at 20 hours, ~3−fold after 37 hours, and ~2−fold after 44 hours and 70 hours. This is direct evidence for the greater intracellular accumulation of As in the mutant, which is consistent with the disruption of As(III) efflux. Furthermore, methylated arsenic species were detected in lysates from the Δ*acr3* mutant, in the form of mono− and dimethylated species (Figure 2.4−C), as well as in soluble form in the medium in concentrations higher than in the WT, not only as mono− but also as di− and trimethylated arsenic (Figures 2.4−B).

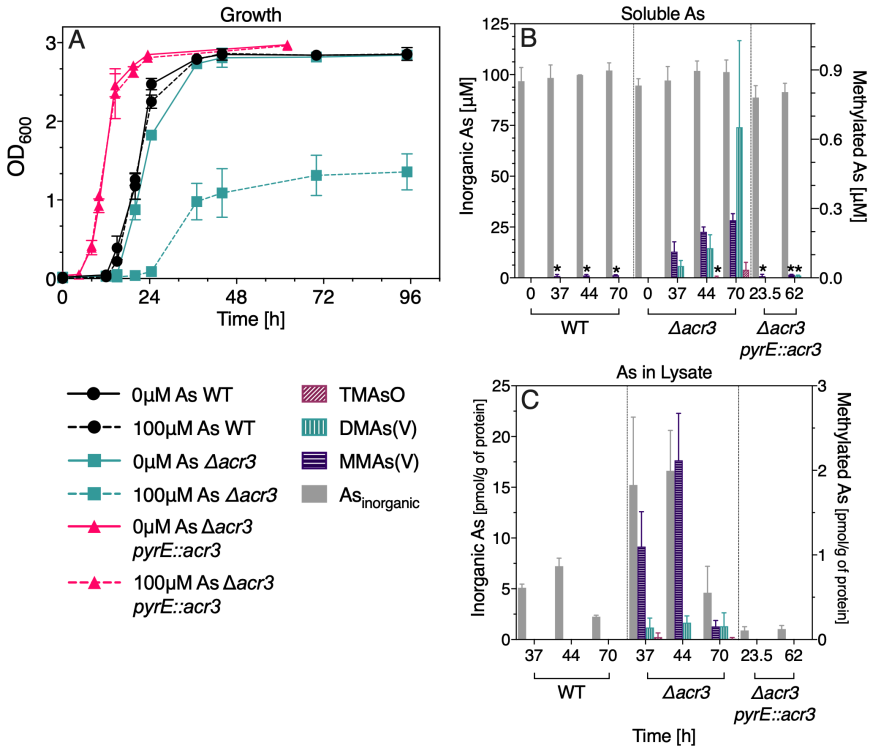


Figure 2.4 – Cultures of *C. pasteurianum* WT (black circles), Δ*acr3* mutant strain (teal squares), and the complementation strain *pyrE*::*acr3* (pink triangles): (A) growth as OD_{600} with 100 µM As(III) (dashed lines) or no As(III) (solid lines), (B) soluble arsenic species in filtered medium, and (C) arsenic species found in pellet lysate (normalized to biomass). Inorganic arsenic is plotted on the left y−axis and methylated arsenic on the right y−axis. Samples marked with an asterisk (*) are below LOQ but above LOD. Data points and bars represent the mean value and error bars, plus and minus one standard deviation.

Thus, the deletion of *acr3* resulted in an 100−fold increase in As methylation, in agreement with the upregulation of the *arsM* gene in the Δ*acr3* mutant vs. the WT strain in response to As(III). This finding provides strong evidence for our model stating that in *arsM*−harboring microorganisms, the presence of an efficient As(III) membrane transporter precludes efficient As methylation. Further supporting this finding, As(III) efflux has been found to be the limiting factor for As volatilization in recombinant strains co−expressing *arsB*/*acr3* and *arsM* genes (P. Yang et al., 2020).

2.4.4 Multiple roles of arsenic methylation in aerobes versus anaerobes

It is apparent that, in the case of aerobic strains, ArsM plays a role in the detoxification of inorganic As. Once cells methylate As, the trivalent and more toxic methylarsenicals will be readily oxidized to their less harmful pentavalent form. *Pseudomonas alcaligenes NBRC14159* represents another example of an aerobic bacterium efficiently methylating As(III) (>80%)

and lacking an As(III)−efflux protein (J. Zhang et al., 2015). More specifically in *A. rosenii*, it is known that the presence of arsenic is sufficient to activate the transcription of the *arsM* gene (K. Huang et al., 2016). This is consistent with the proposed model, in which a sufficient accumulation of As intracellularly activates *arsM* expression and allows the corresponding protein to carry out methylation.

In contrast, for WT *C. pasteurianum*, it appears that the amount of intracellular arsenic, not sufficient to increase *arsM* transcription compared to the no As control, leads to very limited synthesis of monomethyl arsenic. Thus, the efficient efflux system of this organism precludes ArsM activity, consistent with the proposed model. Furthermore, despite a disrupted As(III) efflux system, the *C. pasteurianum* Δ*acr3* strain continues to methylate As to a limited extent in comparison to aerobic organisms. There are several possible explanations for this fact. One is that an additional efflux system is present in this organism, limiting the accumulation of As intracellularly to suboptimal levels for ArsM catalysis. The second is that, for anaerobes, the role of As methylation is distinct from detoxification, and may be related to the proposed interspecies chemical warfare (J. Chen et al., 2019). In which case, efficient As methylation may be triggered by a signal other than intracellular As(III), perhaps one produced by other members of the community. Alternatively, given the high toxicity of MMAs(III) (Petrick et al., 2001), which is the only product in WT anaerobic strains, anaerobic As methylators only generate small amounts of methylated arsenic, which are sufficient for microbial warfare.

Based on our model, we hypothesize more generally that organisms harboring an efficient As(III) efflux system will not drive As methylation in the environment. The fact that all the anaerobes we tested are resistant to As but unable to methylate it efficiently, is not a coincidence. It reflects the reality that microbes living in anoxic environments are commonly exposed to As(III) and may have evolved more robust As(III) resistance than aerobes, typically exposed to the less toxic As(V).

The implication is that ArsM is not necessary for detoxification in anaerobes with robust efflux systems. However, that suggestion calls into question the persistence of the *arsM* gene in the chromosomes of anaerobic microorganisms harboring As(III)−efflux pumps. It is typically assumed that genes that do not confer an evolutionary advantage on their host organism are mutated out of existence. This is clearly not the case, as all the *arsM* genes considered were functional. The logical conclusion of these findings raises the possibility that ArsM serves a primary function other than detoxification in anaerobes. Proving this conjecture will require deconvoluting the dynamics of anoxic microbial communities in order to identify the signals or conditions unleashing the production of methylarsenicals.

3 Active arsenic-methylating bacteria in a rice paddy soil microbiome

Karen Viacava[1], Andrew Janowczyk[2], Suresh Poudel[3], Matthew C. Reid[4], Léonard Jequier[5], Him Shrestha[3], Karin Lederballe Meibom[1], Robert L. Hettich[3], Rizlan Bernier-Latmani[1].

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

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

[3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States.

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

[5]Translational bioinformatics and statistics (TBS), Department of Oncology, Université de Lausanne, Lausanne, Switzerland.

Authors' contributions. KV conducted the experimental work and laboratory analyses. AJ conducted all bioinformatic analyses with assistance of LJ. MR provided the soil microbiomes. Metaproteomic analysis were performed by SP, assisted by HS and supervised by RH. RB-L conceived and supervised the study. KV and RB-L guided the bioinformatic analyses. KV interpreted the bioinformatic analyses output and wrote the manuscript with the advice of KLM and RB-L.

This is a preprint version of the manuscript to be submitted for publication.

3.1 Abstract

Rice paddy soils are key environments in the biogeochemical cycle of arsenic (As), especially for the microbial formation of methylated arsenic species (methylarsenicals). The methylation of As is enhanced in paddy environments, leading to the accumulation of organic As in rice grains and As mobilization into the atmosphere by volatilization. To date, little is known about microbial drivers of As methylation or about the physiological role of this bio−transformation under anoxic conditions. The gene encoding the arsenite S−adenosylmethionine methyltransferase (*arsM* in prokaryotes) has been identified in phylogenetically diverse soil microorganisms and found to be widely distributed in rice paddy soils. The abundance of *arsM* genes has been used to implicate specific microbial functional groups (*e.g.*, sulfate−reducing bacteria) in As methylation. While the *arsM* gene can be a good indicator of the potential role of individual organisms in methylation, mRNA and protein expression allow more robust insight into the key players in As methylation. Here, we combine three meta−omics techniques (metagenomics, metatranscriptomics and metaproteomics) to identify drivers of As methylation in soil−derived microbial cultures. The findings point to members of the order Clostridiales as being the active As methylators in this system.

3.2 Introduction

Despite being naturally−occurring and extensively distributed in the environment, arsenic (As) is toxic to life (Singh et al., 2015). In the last two decades, the biogeochemical cycle of As in rice paddy soils has been the subject of extensive research due to the accumulation of soil As in rice grains (Zhao et al., 2013) and to the fact that diet has been identified as a significant As source for humans (Meharg & Zhao, 2012). In rice as in soil, arsenic can be present in inorganic forms but also as organic species in the way of methylarsenicals (monomethyl, dimethyl, or trimethyl species). Dimethylarsinic acid (DMAs(V)), the dominant organic As species present in rice, is phytotoxic to the rice plant and causes straight-head disease, which affects rice production (M. Z. Zheng et al., 2013). Inorganic As species in rice include either the pentavalent As, arsenate (As(V)) or the trivalent As, arsenite (As(III)) form (H. L. Chen et al., 2016).

Methylated As is synthesized by soil microorganisms from inorganic As, which is geogenic in origin (Lomax et al., 2012). In prokaryotes, the As methylation cascade is catalyzed by the As(III) S−adenosylmethionine (SAM) methyltransferase (ArsM). Methylation of As also leads to As volatilization from paddy systems and has been proposed as an option to remediate high As concentrations in soil, water and rice (Plewniak et al., 2018). In the presence of oxygen, As methylation may lead to the detoxification of inorganic As because the pentavalent methylarsenicals, produced under those conditions, are less toxic than inorganic As(V). In contrast, their trivalent counterparts have been shown to be more toxic than inorganic As(III) (Akter et al., 2006; Watanabe & Hirano, 2013). Therefore, for conditions under which trivalent methylarsenicals are expected, *i.e.*, anoxic conditions, As methylation effectively increases

As toxicity. Hence, As methylation by anaerobic prokaryotes, particularly the production of monomethylarsonous acid (MMAs(III)), has been proposed as a possible mechanism to thwart microbial competitors (J. Li et al., 2016). Given the discrepancy in the toxicity of oxidized and reduced methylarsenicals, doubts on a common role for microbial As methylation in the presence and absence of oxygen persist and have become the subject of much debate (Rahman & Hassler, 2014).

The gene encoding the arsenite S−adenosylmethionine methyltransferase (*arsM*) has been identified in phylogenetically diverse soil microorganisms and found to be abundant in rice paddy soils (Dunivin et al., 2019; Jia et al., 2013; Xiao et al., 2016). To date, most of the As−methylating microorganisms isolated from soil are aerobic bacteria (K. Huang et al., 2016; K. Huang et al., 2018; Kuramata et al., 2015). However, it has been reported that flooding rice paddy soil led not only to increased As methylation (Bright et al., 1994; Mestrot, Feldmann, et al., 2011; Zhao et al., 2013) and the associated increase in As volatilization from soil (Mestrot et al., 2009), but also generated higher DMAs(V) concentrations in rice grains (X. Y. Xu et al., 2008). The latter suggests that anaerobic microbes may play an important role in As methylation. Yet, little is known about the microorganisms that catalyze As methylation under anoxic conditions. Thus, the identity of the microorganisms involved in anaerobic As methylation is a key knowledge gap that may enable the elucidation of the physiological role of As methylation and the assessment of its potential as a bioremediation strategy.

Efforts to identify microbial mediators of As methylation in rice paddy soils have consisted mainly of sequencing of *arsM* amplicons produced using degenerate primers (Reid et al., 2017; Xiao et al., 2016; S. Y. Zhang et al., 2015; Zhao et al., 2013). Sulfate−reducing bacteria (SRB) have been proposed to be active As methylators (C. Chen et al., 2019; M. Wang et al., 2019) based on the positive correlation between the abundance of the *arsM* gene and that of the dissimilatory sulfite reductase gene (*dsr*) and on the detrimental effect of the addition of chemical inhibitors of dissimilatory sulfate reduction on As methylation.

Due to the phylogenetic diversity of *arsM* genes, degenerate primers likely underestimate its abundance and may overlook microorganisms playing an important role in As methylation (J. Zhang et al., 2015). This drawback is overcome by metagenomic DNA sequencing, yet there is only one study, thus far, probing *arsM* taxonomic classification within a metagenomic dataset (Xiao et al., 2016). Furthermore, although *arsM*−harboring microorganisms have the potential for As methylation, the presence of this gene in the genome might not necessarily translate into As−methylating activity as was recently reported for pure cultures (Viacava et al., n.d.).

This study aims to identify active As methylators in microbial cultures derived from a Vietnamese rice paddy by relying on a range of meta−omics techniques (metagenomics, metatranscriptomics and metaproteomics) to demonstrate expression of the *arsM* gene and the ArsM protein in specific members of the microbial community. Thus, these approaches allow the microbiome−wide study of As methylation and the deconvolution of the activity of individual microorganisms.

3.3 Materials and methods

3.3.1 Anaerobic soil-derived cultures

The soil−derived cultures consist of two distinct anaerobic microbial communities derived from a Vietnamese rice paddy soil grown by Reid et al., 2017. The first soil−derived microbiome was grown in medium enriched in electron acceptors (EA medium) (5 mM $NaNO_3$, 5 mM Na_2SO_4, 5 mM ferric citrate, 0.2 g l^{-1} yeast extract (Oxoid Ltd), 1 g l^{-1} cellobiose and 7.5 g l^{-1} tryptic soy broth (TSB), pH 7) and will be hereafter referred to as the EA culture. The soil−derived microbiome grown in TSB medium (7.5 g l^{-1} TSB) will be referred to as the TSB culture. Both media were boiled, cooled down to room temperature under 100% N_2 gas flow and dispensed into 200−ml serum bottles. The bottles were sealed with butyl rubber stoppers and aluminium crimps and their headspace flushed with 100% N_2 gas prior to autoclaving. All culture manipulations were carried using thoroughly N2−flushed syringes and needles. Both cultures were grown at 30 °C. Growth was quantified using optical density at 600 nm (OD_{600}).

3.3.2 Arsenic methylation assays

Pre−cultures were started from -80°C glycerol stocks and the EA culture was transferred at least twice before being used as a preculture. The latter was performed because only after a second transfer did dark precipitates, presumably iron sulfides, form in the EA culture, suggesting that ferric iron or sulfate reduction metabolisms recovered only after two transfers post−freezing. The first experimental set−up consisted of the inoculation of bottles containing medium amended with 25 μM arsenite (As(III)) as $NaAsO_2$ (+As condition) or unamended (no−As condition). Cell pellets were sampled during the stationary phase (~44 h) for DNA sequencing and for proteome characterization and at the mid−exponential growth phase for RNA sequencing. In a second experimental set−up, cultures were grown in medium without As(III) and, at the mid−exponential growth phase, As(III) was added to a final concentration of 25 μM As(III). Cell pellets were sampled before (no−As sample) and 30 min after As spiking (+As sample) and were used only for RNA analysis. Sampling for soluble As species consisted in filtering 1 ml of culture through a 0.22-μm filter, oxidation by adding 10% (v/v) of 30% H_2O_2 (w/v) and 1:1 dilution with 2% HNO_3 (v/v) (>69%, Trace Select, Honeywell Fluka Analytical, Switzerland). All experiments were performed with three biological replicates.

3.3.3 Arsenic speciation analysis

Arsenic speciation was obtained using an Agilent ICP−QQQ 8900 coupled to an HPLC 1260 Infinity II (Agilent Technologies, Santa Clara, CA, US) with an anion exchange PRP X−100 HPLC column (150 × 4.1 mm I.D., 10 μm, Hamilton, Reno, NV, US) (pump: 1 ml min^{-1}, autosampler: 4°C and column compartment: 20 °C). The eluent consisted of 6.66 mM NH_4NO_3 and 6.66 mM $(NH_4)_2HPO_4$ (pH 6.2) and the As standards were: TMAsO as trimethyl arsine oxide (Argus Chemicals Srl., Italy), DMAs(V) as sodium dimethylarsinate (ABCR, Germany),

MMAs(V) as monomethylarsonic acid (Chemservice, PA, US) and As(V) as $Na_2HAsO_4 \cdot 7H_2O$ (Sigma−Aldrich, MO, USA). Because As samples were oxidized prior to analysis, As speciation analysis discriminated between inorganic As (detected as As(V): sum of As(III) (oxidized) plus As(V) in solution) and tri−, di− or monomethylated arsenicals but did not allow the identification of the redox state of the methylarsenicals produced by the strains. Thus, we refer to DMAs(V) and MMAs(V) in the text because they are the species measured, but they could correspond to trivalent species prior to oxidation. Total As concentrations were measured using the same ICP−QQQ instrument in stand−alone mode.

3.3.4 DNA sequencing and metagenomic analysis

DNA was extracted from 4 ml of culture, pelleted (10 min, 4,500 g), using the DNeasy® Power® Soil Kit (Qiagen, Germany) with the modification of homogenizing with a Precellys 24 Tissue Homogenizer (Bertin Instruments, France) (6,500 rpm for 10 s, repeated 3× with 10 s pause intervals). Metagenomic sequencing was performed by the Genomics Platform of the University of Geneva, Switzerland (iGE3) on an Illumina HiSeq 4000 (Illumina, 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 (Andrews, 2020) and duplicated reads eliminated by FastUniq (H. Xu et al., 2012). Reads from all biological replicates within the same experimental condition were assembled into contiguous sequences (contigs) using MegaHit (D. Li et al., 2015). The contig abundance was determined by aligning the sequencing reads from each biological replicate back to the assembled contigs using Kallisto (Bray et al., 2016), a program for quantifying abundance of transcripts that relies on a pseudoalignment method shown to be highly applicable for quantifying reads in metagenomic data (Schaeffer et al., 2017). 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), henceforth referred as TPM−DNA when used for gene abundance, a measure of gene abundance where reads are normalized for gene length and sequencing depth (Wagner et al., 2012). Prodigal was used for the prediction of protein−coding genes (Hyatt et al., 2010). The annotation server GhostKOALA was used to assign a KEGG Orthology (KO) database number to each gene to identify its encoded function and taxonomic category (Kanehisa et al., 2016). Gene function annotation by TIGRFAMs was performed using WebMGA (Wu et al., 2011). The small subunit (SSU; 16S) rRNA sequences were identified directly in the contigs and their taxonomy assigned by Metaxa2 (Bengtsson-Palme et al., 2015). Contigs with length>2,000 bp were clustered into bins based on composition and coverage using CONCOCT (Alneberg et al., 2014). Completeness, contamination, strain heterogeneity and % community in contigs for each bin were calculated using CheckM (Parks et al., 2015).

3.3.5 RNA sequencing and metatranscriptomic analysis

A culture (5 mL) was harvested at mid−exponential phase for metatranscriptomic analysis. The cells were lysed and the RNA purified using the RNeasy Mini Kit following the manufacturer's instructions (RNAprotect® Bacteria, Qiagen, Germany). The purified RNA was DNase−I treated (Promega, WI, US) (1h, 37 °C) and cleaned using the RNeasy Mini Kit a second time. Ribosomal RNA (rRNA) depletion (kit QIAseq FastSelect −5S/16S/23S), library preparation using single−end 100 bases reads (TrueSeq Stranded mRNA) and RNA sequencing (on an Illumina HiSeq 4000) were performed by the iGE3 Platform. Reads were quality−assessed by FastQC, trimmed by Trimmomatic (Bolger et al., 2014), post−sequencing rRNA depleted by SortMeRNA (Kopylova et al., 2012) and aligned to their corresponding metagenome by Bowtie2 (Langmead & Salzberg, 2012). Finally, featureCounts (Liao et al., 2014) was employed to count the number of RNA reads aligned to the Prodigal−predicted protein−coding genes, the raw counts were used to calculate the TPM, henceforth referred as TPM−RNA when employed for transcript abundance.

3.3.6 Metaproteome characterization and metaproteomic analysis

The metaproteome analysis was performed at Oak Ridge National Laboratory (Oak Ridge, 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% SDS, 100 mM Tris−HCl, 10 mM dithiothreitol, 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 buffer and the concentration estimated with a Nanodrop (Thermo Fisher Scientific, Waltham, MA, US). Sequencing−grade trypsin (Promega, WI, US) 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 remaining SDC and collect tryptic peptides (peptide products of trypsin digestion). Aliquots of 12 μg of peptides were analyzed by 2D LC−MS/MS consisting of a Vanquish UHPLC connected to a Q Exactive Plus MS (Thermo Fisher Scientific, Waltham, MA, US). Spectral data were collected using MudPIT (multidimensional protein identification technology) as described previously (McDonald et al., 2002; Washburn et al., 2001). 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 with parameters according to (Clarkson et al., 2017).

Four protein databases, one per experimental condition (EA +As, EA no As, TSB +As, TSB no As), were created from the protein translations of the protein−encoding genes generated by Prodigal. The MS/MS spectra raw files were processed in Proteome Discoverer version 2.4 with MS Amanda 2.0 and Percolator. Spectral data were searched against the protein database of their corresponding condition. 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.

3.4 Results and discussion

3.4.1 Arsenic methylation by soil-derived microbiomes

Two experimental set−ups were used to probe the As−methylating cultures. The first compared the soil−derived microbial communities grown in the presence or absence of As(III). Samples for the metagenome, metaproteome and one of the metatranscriptomes (labeled G for "growth in the presence of As") were obtained from this set−up (Appendix B, Figures B.1-A and B.2-A). Growth in no−As controls is reported in Appendix B, Figure B.3. In the second set−up, the intention was to capture the short−term response of the community to As(III). Thus, the cultures were grown to mid−exponential phase in the absence of As, sampled, amended with As(III), and sampled a second time after 30 minutes. Samples obtained from this set−up were used for the second type of metatranscriptome (labeled R for "response to arsenic addition") (purple lines in Appendix B, Figures B.1-A and B.2-A). A decrease in the total soluble As (Appendix B, Figures B.1-B and B.2-B) was observed in most cases and could be due either to volatilization or intracellular accumulation of As (not assessed). Soluble methylarsenicals were detected in all experiments, with MMAs(V) and DMAs(V) as the main species (in that order, panels C and D from Figures B.1 and B.2 in Appendix B). The methylation efficiency varied with the highest methylation efficiency being 27.7% of the total initial As(III) for the EA culture and 19.5% for the TSB culture. Previously, the EA culture had been reported to convert 63% of the inorganic arsenic into mainly DMAs(V) and a small proportion to MMAs(V) (Reid et al., 2017). The higher rate in conversion of inorganic As in the TSB metagenome experiment, in comparison to the proteome experiment, may be due to the lower substrate concentration used (~27.5 μM vs. 15 μM). Lower As concentrations were observed to lead to higher As methylation efficiencies (Hettick et al., 2015). Compared to previous studies of As−methylating anaerobic cultures, enrichments from lake sediments (12%) (Bright et al., 1994), or As−methylating single−strain cultures, *Clostridium sp.* BXM (10%) (P. P. Wang et al., 2015) and the archaeon *Methanosarcina acetivorans* C2A (10-13%) (P. P. Wang et al., 2014), these EA and TSB cultures remain the most efficient As−methylating anoxic microbial cultures reported.

3.4.2 Microbiome composition

The taxonomic classification of identified 16S small subunit (SSU) rRNA sequences in the metagenomes showed that the main fraction of the communities is bacterial (>89.0 ±0.8% for EA culture and >98.5 ±0.3% for TSB culture, relative abundance) and is distributed amongst six operational taxonomic units (OTUs) at the order level: Clostridiales, Bacillales, Lactobacillales, Bacteroidales, Enterobacterales and Desulfovibrionales (Figure 3.1). In the no−As condition for the EA culture, Firmicutes (76.0 ±0.4%) was the most abundant phylum, particularly members of the order Clostridiales (56.6 ±1.4%) and Lactobacillales (16.9 ±1.2%). When the community was grown in the presence of arsenic, Clostridiales remained at the same proportion (56.8 ±1.7%) while Lactobacillales decreased by two thirds (4.7 ±0.4%) and members of the order Bacteroidales, from the phylum Bacteroidetes, doubled from 8.4 ±0.5% (no−As control) to 16.9 ±0.4%.

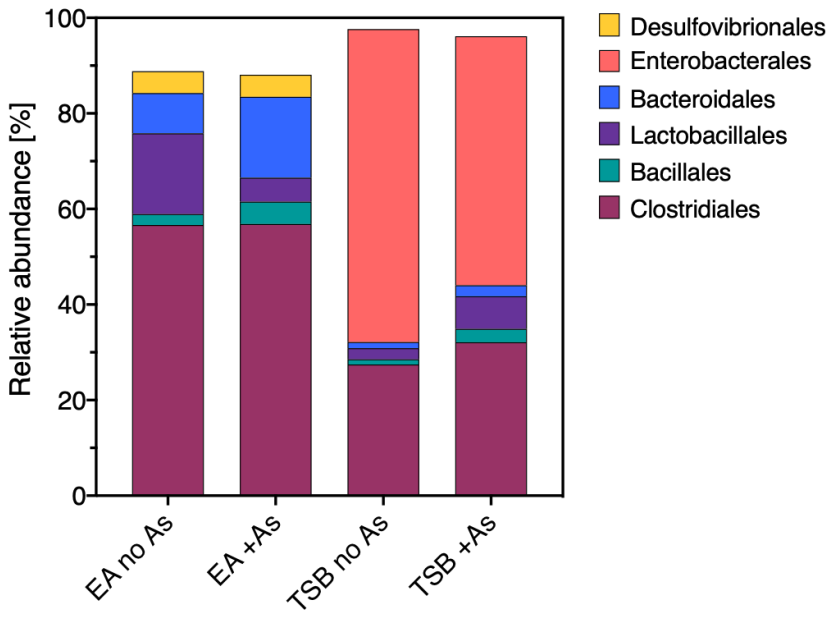


Figure 3.1 – Abundant operational taxonomic units (OTUs) at order level (> 1% relative abundance) identified from rRNA sequences present in metagenomes from soil−derived cultures. OTUs belonged to the phyla; Firmicutes (Clostridiales, Bacillales, Lactobacillales), Bacteroidetes (Bacteroidales) and Proteobacteria (Desulfovibrionales, Enterobacterales).

The community in the TSB no−As control was predominately from the phylum Proteobacteria, from the order Enterobacterales (65.5 ±6.4%), followed by Firmicutes: Clostridiales (27.5 ±5.7%), Lactobacillales (2.3 ±0.1%) and Bacillales (1.1 ±0.4%). The TSB community was altered upon exposure to As and exhibited a decrease in Enterobacterales (52.1 ±6.3%) and an increase in the rest of the taxonomic orders.

At the genus level (Appendix B, Figure B.4), the order Clostridiales includes primarily contri-

butions from members of *Clostridium* for both EA and TSB. In addition, for EA, there was a large contribution from organisms for which there is no attribution at the genus level (*Incertae Sedis*). In addition, *Oscillibacter* species are also identified for EA. In EA, in the Metaxa2 output, there were genera from the order Selenomonadales, which were misclassified as Clostridiales at the order level (and thus Selenomonadales does not appear in Figure 3.1), including the genera *Acidanomococcus*, *Zymophilus*, *Anaeroarcus* and *Phascolarctobacterium*. Desulfovibrionales consisted entirely of the *Desulfovibrio* genus. In the TSB culture, the most abundant genera are *Citrobacter* and *Enterobacter*, in the absence and presence of arsenic respectively, both from the dominant Enterobacterales order.

3.4.3 Building metagenome-assembled genomes (MAGs)

Contigs from the four metagenomes generated, EA +As, EA no As, TSB +As and TSB no As, were clustered into bins. High quality (≥95% completeness and ≤5% contamination) and good quality (≥75% completeness and ≤12% contamination) bins were designated Metagenome−Assembled Genomes (MAGs). Since the focus of this study is active As methylation, bins encoding *arsM* genes were also retained as MAGs even if they did not match the high or good quality criteria, and only MAGs from the +As condition libraries were further investigated for key metabolic and arsenic resistance (*ars*) genes.

The parsing process led to a total of 36 MAGs: 16 from EA including 11 high−quality, one good−quality and four low−quality but *arsM*−encoding MAGs (Table 3.1), and 20 from TSB, including 14 high−quality, four good−quality and two low−quality but *arsM*−encoding MAGs (Table 3.2). For each MAG, a marker lineage was assigned by CheckM, based on the lineage−specific marker genes used to estimate genome completeness and contamination (Parks et al., 2015).

Furthermore, a supplementary taxonomic assignment was obtained for each MAG and indicated in the "Taxonomy" column in Tables 3.1 and 3.2. The taxonomic assignment was obtained by considering the predominant OTUs at the genus level from the GhostKOALA taxonomic annotation for all putative protein−encoding genes, the marker lineage (both detailed in Appendix B, Tables B.1 and B.2), and the taxonomic classification of the SSU 16S rRNA sequences within the MAG (Appendix B, Tables B.3 and B.4). Bins classified under the same MAG group from no−As and +As condition are shown in Appendix B, Tables B.5 and B.6.

The MAGs belonged mainly to the phylum Firmicutes (order Clostridiales: the genera *Clostridium*, *Paeniclostridium* and *Oscillibacter*, and the family *Ruminococcaceae*; order Selenomonadales: the genera *Pelosinus* and *Phascolarctobacterium*; order Lactobacillales: the genus *Enterococcus*). Proteobacteria MAGs included the order Enterobacterales (the genera *Citrobacter* and *Raoultella*) and the order Desulfovribrionales. Finally, one MAG from the order Bacteroidetes (genus *Bacteroides*) was present in both EA and TSB cultures. The relative abundance of each MAG in the community was extracted using the 'Profile' command in CheckM (reported as % Community in Tables 3.1 and 3.2). This percentage is calculated by dividing the

Table 3.1 – Metagenome−assembled genomes (MAGs) from EA culture, +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 marker genes and the expected colocalization of these genes. **Strain heterogeneity**: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from phylogenetically diverse sources. **% 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. High−quality MAGs are denoted by bolded numbers, good−quality MAGs by bolded and italicized numbers.

MAG	Taxonomy*	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of *arsM* genes	% of binned proteins assigned to MAG	% Community
1	*Bacteroides* (g)	Bacteroidales (UID2657)	98.88	0.74	0	3.81	41	0	3.3%	5.82
2	*Ruminococcaceae* (f)	Clostridiales (UID1212)	97.82	0.34	0	2.16	62	0	2.3%	19.91
3	Clostridiales (o)	Clostridiales (UID1212)	94.90	0.67	100	2.51	67	0	2.7%	0.45
4	Clostridiales (o)	root (UID1)	83.33	170.83	3.97	10.09	258	2	10.9%	0.52
5	*Clostridium* (g)	Clostridiales (UID1375)	98.39	0	0	3.96	336	0	4.1%	5.05
6	*Clostridium* (g)	Bacteria (UID203)	91.38	82.76	12.5	8.90	266	1	9.7%	0.19
7	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	100	0	0	3.25	18	2	3.1%	10.19
8	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	99.23	0.68	100	3.42	35	2	3.2%	0.50
9	*Enterococcus* (g)	Lactobacillales (UID544)	99.63	0	0	2.74	32	0	2.8%	2.65
10	*Enterococcus* (g)	Lactobacillales (UID544)	99.31	4.62	0	4.18	105	0	4.6%	0.80
11	*Oscillibacter* (g)	root (UID1)	100	190.62	43.61	8.33	197	1	8.8%	0.36
12	*Oscillibacter* (g)	Clostridiales (UID1212)	98.66	1.34	50	2.92	174	0	3.2%	10.81
13	*Paeniclostridum* (g)	Clostridiales (UID1120)	95.07	1.41	0	3.51	50	1	3.9%	1.24
14	*Pelosinus* (g)	root (UID1)	60.42	56.25	100	6.45	395	3	6.9%	0.95
15	*Phascolarctobacterium*	Selenomonadales	99.98	2.1	0	2.26	26	0	2.4%	2.09
16	Selenomonadales (o)	Firmicutes (UID1022)	99.91	0	0	2.78	50	0	3.0%	0.81

* (o) order, (f) family or (g) genus.

Table 3.2 – Metagenome−assembled genomes (MAGs) from TSB culture, +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 marker genes and the expected colocalization of these genes. **Strain heterogeneity**: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from phylogenetically diverse sources. **% 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. High−quality MAGs are denoted by bolded numbers, good−quality MAGs by bolded and italicized numbers.

MAG	Taxonomy*	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of *arsM* genes	% of binned proteins assigned to MAG	% Community
1	*Bacteroides* (o)	Bacteroidales (UID2657)	96.38	10.3	83.72	2.89	516	0	4.3%	7.41
2	*Citrobacter* (g)	Enterobacteriaceae (UID5103)	89.24	3.57	70.97	2.84	755	0	5.5%	23.90
3	Clostridiales (o)	Clostridiales (UID1212)	95.21	2.52	0	2.80	277	0	3.0%	0.18
4	Clostridiales (o)	Clostridiales (UID1212)	100	0	0	2.17	13	0	3.3%	0.44
5	Clostridiales (o)	Clostridiales (UID1212)	98.66	0.89	0	2.32	24	0	3.2%	0.36
6	*Clostridium* (g)	Clostridiales (UID1375)	99.19	1.08	0	3.02	23	1	3.9%	0.40
7	*Clostridium* (g)	Clostridiales (UID1375)	98.12	0	0	4.43	306	0	5.1%	1.33
8	*Clostridium* (g)	root (UID1)	100	104.17	32.76	3.34	147	3	10.9%	1.88
9	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	94.04	0	0	3.24	247	1	3.5%	0.22
10	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	98.47	1.81	50	3.43	62	2	3.4%	0.86
11	Enterobacterale s (o)	Enterobacteriaceae (UID5054)	98.76	1.83	15	5.55	74	0	4.8%	0.52
12	*Enterococcus* (g)	Lactobacillales (UID544)	99.31	4.62	0	3.53	105	0	4.8%	2.07
13	*Enterococcus* (g)	Lactobacillales (UID544)	99.63	0	0	4.67	48	0	3.2%	2.85
14	*Oscillibacter* (g)	Bacteria (UID203)	98.28	85.13	74.68	3.09	551	1	6.4%	0.21
15	*Oscillibacter* (g)	Clostridiales (UID1212)	98.99	0.67	100	4.35	152	0	3.3%	4.75
16	*Pelosinus* (g)	Firmicutes (UID1022)	100	0.63	0	4.55	35	1	3.5%	2.05
17	*Phascolarctobacterium* (g)	Selenomonadales (UID1024)	99.98	1.5	0	4.19	15	0	2.6%	0.87
18	*Raoultella* (g)	Enterobacteriaceae (UID5103)	97.73	12.00	39.18	6.60	391	0	7.3%	6.52
19	*Ruminococcaceae* (f)	Clostridiales (UID1212)	97.99	1.01	0	2.67	36	0	2.4%	2.49
20	Selenomonadales (o)	Firmicutes (UID1022)	99.91	0	0	10.27	132	0	3.3%	0.36

* (o) order, (f) family or (g) genus.

sum of the number of reads mapped to the contigs in each MAG by the total number of reads mapped to all contigs (including the unbinned contigs), adjusted for the size of the MAG (and assuming an average genome size for the unbinned fraction). The MAGs with >5% relative abundance in the EA community are *Ruminococcaceae* (19.91%), *Desulfovibrio* (MAG 7, 10.19%), *Oscillibacter* (MAG 12, 10.81%), *Bacteroides* (5.82%) and *Clostridium* (MAG 5, 5.05%); and in the TSB community, they are *Citrobacter* (23.90%), *Bacteroides* (7.41%) and *Raoultella* (6.52%).

3.4.4 MAG metabolic pathways

The presence, transcription and translation of genes encoding key enzymes from different metabolic pathways were assessed for each MAG and shown in Figures 3.2 and 3.3, where the MAGs were arranged according to their taxonomic order. The *arsM*−encoding MAGs with contamination >12% are indicated in red given they likely combine genes from distinct organisms.

The *Desulfovibrio* MAGs expressed the pathway for dissimilatory sulfate reduction (DSR) in both cultures, and the dissimilatory reduction of nitrate to ammonia (DNRA) in the EA culture, presumably due to the presence of nitrate in the medium. The proteome of *Desulfovibrio*, in the TSB culture, expressed key enzymes in the acetyl−CoA (Wood−Ljungdahl) pathway, an indication of either CO_2 reduction or acetate oxidation. In the proteome, evidence of the expression of an alcohol dehydrogenase enzyme (YiaY), in both the EA and TSB cultures, suggested alcohol oxidation. The *Desulfovibrio* MAGs in EA encoded the decaheme outer membrane protein (MtrB), suggesting their capability to respire extracellular electron acceptors such as ferric iron. Three MAGs, *Citrobacter* (the most abundant community member), Enterobacterales, and *Raoultella*, all pertaining to the order Enterobacterales and detected only in the TSB culture, displayed similar active metabolisms. The three MAGs encoded and expressed genes for sulfur disproportionation, denitrification and carbon fixation (via the reductive pentose pathway (rPP)). In addition, all three expressed the Entner−Doudoroff (ED) pathway (a variant of the glycolysis pathway), the glyoxylate pathway (isocitrate lyase and malate synthase) in combination with the tricarboxylic acid cycle (TCA), for the metabolism of organic acids under anoxic conditions and reduction/oxidation of pyruvate to acetate (PoxB) and formate (PFL).

The two *Enterococcus* MAGs (order Lactobacillales) presented evidence of lactic acid fermentation, the ED pathway and the PP cycle. The Enterobacterales and the *Enterococcus* proteomes had clear indication of the conversion of pyruvate to acetyl−CoA (pyruvate metabolism), a key step of fermentation. One of the two *Enterococcus* MAGs proteomes provided evidence for denitrification in the EA culture. The *Bacteroidetes* MAG in the TSB culture was the only one presenting evidence for propionyl−CoA carboxylase expression, suggesting active propionic acid fermentation, for which lactic acid is a substrate. In EA, the Selenomonadales MAG included some proteomic evidence for DNRA and mRNA for lactic acid fermentation.

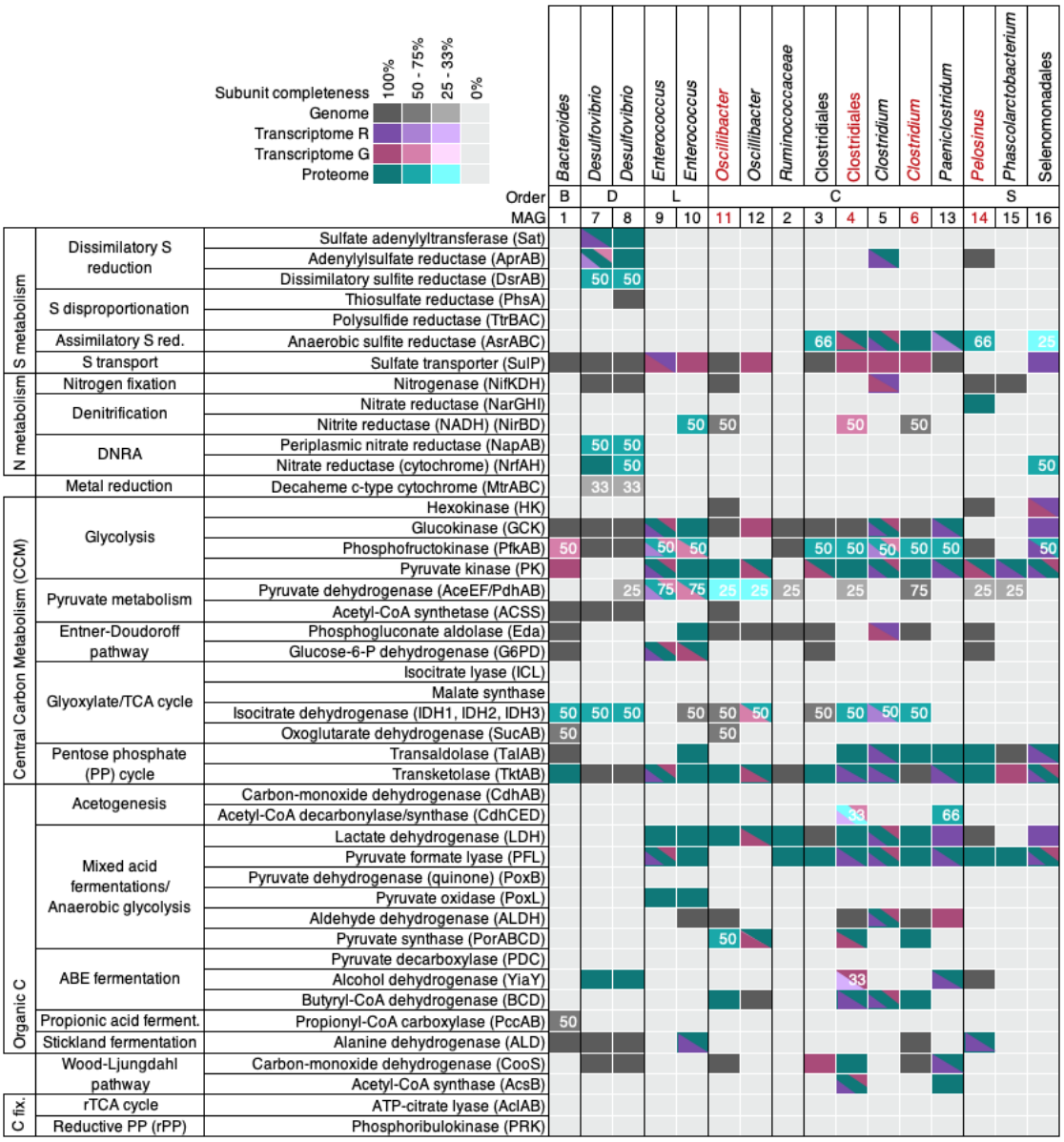


Figure 3.2 – Key enzymes from metabolic pathways in MAGs from the EA culture. Subunit completeness correspond to the fraction of gene encoding subunits present (genome) or transcribed (transcriptome) or to the fraction of protein subunits expressed (proteome). MAGs indicated in red have >12% contamination (and harbor arsM genes). Order corresponds to the OTUs at order level: (B) Bacteroidales, (D) Desulfovibrionales, (L) Lactobacillales, (C) Clostridiales and (S) Selenomonadales. Abbreviations: dissimilatory nitrate reduction to ammonia (DNRA), central carbon metabolism (CCM), organic carbon metabolism (Org. C), carbon fixation (C. fix.), tricarboxylic acid cycle (TCA), reductive TCA cycle (rTCA) and acetone–butanol–ethanol (ABE) fermentation. Number in boxes refer to percentage. Colored boxes with no numbers correspond to a 100% presence/expression. Multi–colored boxes represents presence in more than one analysis. Refer to Appendix B, Table B.7 for further metabolic analysis.

Figure 3.3 – Key enzymes from metabolic pathways in MAGs from the TSB culture. Subunit completeness correspond to the fraction of gene encoding subunits present (genome) or transcribed (transcriptome) or to the fraction of protein subunits expressed (proteome). MAGs indicated in red have >12% contamination (and harbor arsM genes). Order corresponds to the OTUs at order level: (B) Bacteroidales, (D) Desulfovibrionales, (L) Lactobacillales, (C) Clostridiales and (S) Selenomonadales. Abbreviations: dissimilatory nitrate reduction to ammonia (DNRA), central carbon metabolism (CCM), organic carbon metabolism (Org. C), carbon fixation (C. fix.), tricarboxylic acid cycle (TCA), reductive TCA cycle (rTCA) and acetone−butanol−ethanol (ABE) fermentation. Number in boxes refer to percentage. Colored boxes with no numbers correspond to a 100% presence/expression. Multi−colored boxes represents presence in more than one analysis. Refer to Appendix B, Table B.8 for further metabolic analysis.

The MAGs belonging to the Clostridiales order in both cultures displayed various fermentation metabolisms: lactic acid fermentation (*Oscillibacter*, *Ruminococcaceae*, *Clostridium*), production of acetate and molecular hydrogen via the pyruvate synthase cluster (*porABCD*) (*Oscillicibacter*, *Clostridium*), acetogenesis (*Paeniclostridium*, *Clostridium*), pyruvate oxidation to formate (*Clostridium*, *Paeniclostridium*, *Ruminococcaceae*), acetone−butanol−ethanol (ABE) fermentation (*Oscillicibacter*, *Clostridium* and *Paeniclostridium*) and amino acid fermentation (*Clostridium*). Their expression of the TCA cycle enzymes indicates the use of organic acids as a source of carbon. All of the MAGs harbored genes, and some expressed the proteins, necessary for sulfur transport and assimilation and the PP cycle. More information about the metabolic potential of MAGs is available in Appendix B, Tables B.7 and B.8.

3.4.5 Arsenic metabolism

As the major focus of this work is on arsenic methylation, arsenic metabolism was considered separately and in detail. In addition to *arsM*, arsenic−related genes (*ars* genes) encode proteins that include As(III)−efflux systems (*arsB*, *acr3*) which often require an ATPase (*arsA*). As(III) efflux can be preceded by As(V) reduction coupled to the oxidation of glutaredoxin (Grx) or thioredoxin (Trx), both of which are catalyzed by an arsenate reductase (*arsC*) (Andres & Bertin, 2016; Y. G. Zhu et al., 2017; Y.-G. Zhu et al., 2014). The expression of *ars*−resistance genes is regulated by the ArsR repressor, which binds to the operon's promoter region, impeding transcription. ArsR is only released upon binding of intracellular As(III), causing a conformational change via interaction with three of its cysteine residues (Prabaharan et al., 2019). Thus, we looked at genes with increased expression, relative to the no−As control, in transcriptomes and proteomes to investigate the *ars* gene involved in As metabolism.

All the *ars* genes present in the EA and TSB metagenomes have been numbered in Figures 3.4 and 3.5 and color−coded according to their increased expression, relative to the no−As condition, in the transcriptomes and/or in the proteomes. The inset tables from Figures 3.4 and 3.5 show the total number of *ars* genes found in the metagenome and/or with increased expression. The criterion for robust presence of genes within the metagenomes was to select genes encoded in contigs with >5 estimated counts (number of mapped unique reads, Kallisto output), in two of the three biological replicates (data not shown).

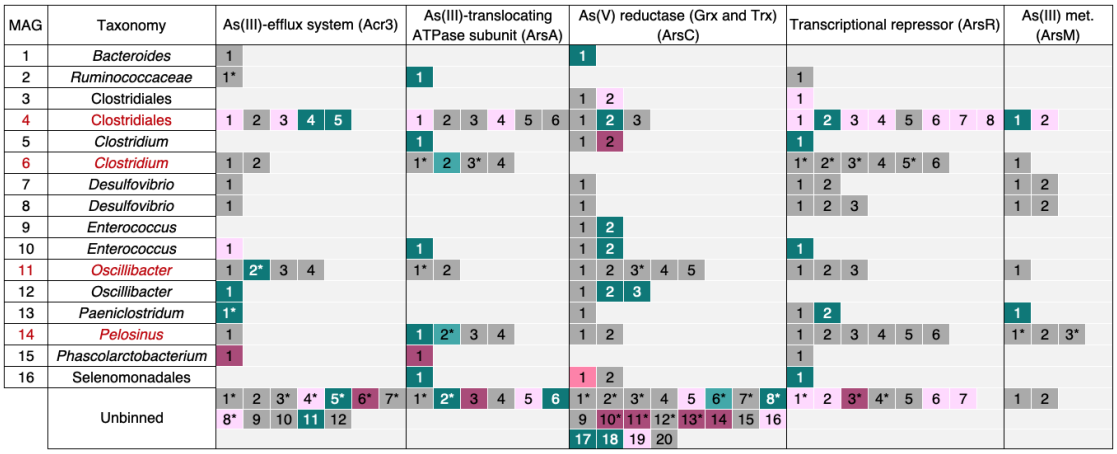
The criteria for induction or increased expression was that the number of TPM−RNA and/or the protein abundance were two−fold greater than those in the no−As control (or no detection in the no−As control versus expression in the +As condition) in at least two of three biological replicates, for transcriptome R, or in the average of the three biological replicates for transcriptome G and proteome (considering only genes with transcripts or detected peptides in at least two of the three biological replicates). The gene expression relative to the no−As control in transcriptomes and the proteome for all identified *ars* genes can be found in Appendix B, Tables B.9 and B.10. Individual values of gene abundance in the genome (TPM−DNA), gene expression in transcriptomes (TPM−RNA) and protein abundance in the proteome are listed

in Appendix B, Table B.11 and B.12 for all identified *ars* genes in the +As condition.

To find the equivalent gene between no−As and +As condition metagenomes, the binned and unbinned *ars* genes from each metagenomic library were clustered using CD-HIT-2D (W. Li & Godzik, 2006) with an 80% identity threshold (95% of proteins were ≥97% identical, see Match identity column in Appendix B Tables B.9 and B.10). Unbinned *ars* genes, representing ~30% of the total number of *ars* genes in both +As metagenomes, are encoded in contigs that could not be clustered during the binning process or that belonged to low quality bins that did not include the *arsM* gene.

According to the Kallisto output, *ars* genes presented higher abundance (TPM−DNA) in the EA culture metagenome than in that of the TSB culture (Appendix B, Figure B.5). The most abundant *ars* genes in EA and TSB culture metagenomes were the genes encoding the As(V) reductase *arsC*, and the transcriptional repressor *arsR* (Appendix B, Figure B.5). Both genes are part of the cluster *arsRBC* considered to be the canonical *ars* operon (Fekih et al., 2018). In addition, *acr3* was also highly abundant in the EA metagenome (Appendix B, Figure B.5). Both *acr3* and *arsB* encode As(III)−efflux systems. However, genes annotated as *arsB* were only found in contigs from the TSB metagenome (Figure B.5, numbers in italics in the first column); the rest of the As(III)−efflux−system genes were classified as *acr3*. The ATPase (ArsA), which is generally found associated with ArsB, has been proposed to interact with other transporters, including Acr3 (Castillo & Saier, 2010) and was the fourth most abundant gene (Appendix B, Figure B.5). Finally, *arsM* was the least abundant gene as observed in previous studies of paddy soils (Xiao et al., 2016; S. Y. Zhang et al., 2015). All of the MAGs encoded at least one *ars* gene and, as expected, MAGs with a high contamination level (>12%) included more *ars* genes than what has been reported for a single genome (Figures 3.4 and 3.5).

Whilst encoding a gene is an indication of the organism's metabolic potential, its transcription and expression provide more robust evidence of its active role in microbial metabolism. Transcriptome G corresponds to mRNA sampled at mid−exponential growth phase from the community growing under continuous As(III) pressure. However, given that degradation of mRNA is a means of regulation of gene expression in prokaryotes (Arraiano, 1993), and that mRNA inventories rapidly respond to shifts in the environment, an additional experimental approach was used for mRNA sampling. The mRNA was sampled from the community shortly (30 minutes) after its exposure to As(III). This corresponds to transcriptome R. A similar number of *ars* genes with increased transcription in both transcriptomes was found for the TSB culture, but, for the EA culture, three as many genes had increased RNA reads in transcriptome G as in transcriptome R (see inset tables in Figures 3.4 and 3.5).

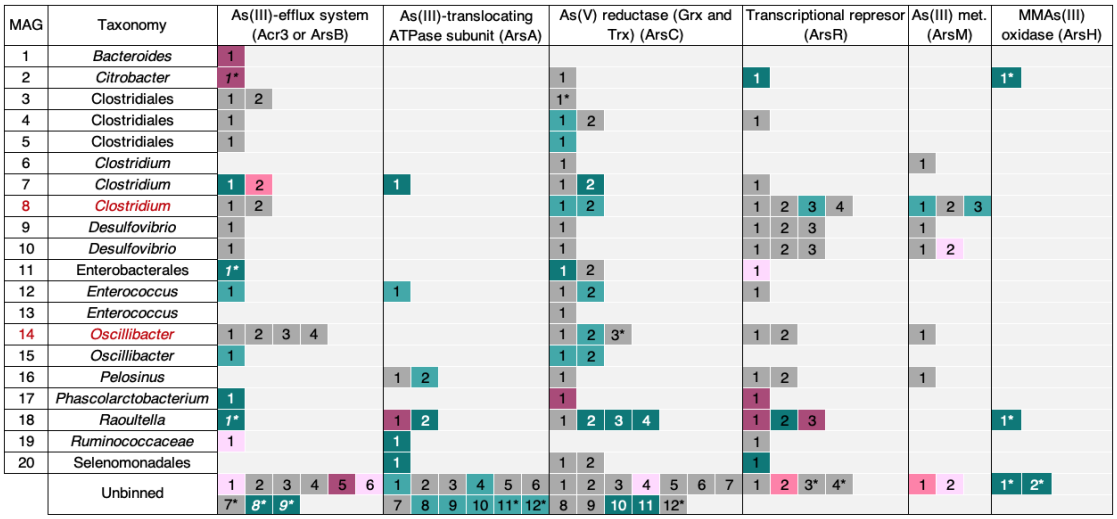


	Acr3	ArsA	ArsC	ArsR	ArsM	Total
# in Genome	32	26	48	43	14	163
# in Transcriptome R	4	4	7	4	1	20
# in Transcriptome G	14	9	18	16	2	59
# in Proteome	7	9	10	5	2	33
# Total expressed	14	14	20	17	3	68

Figure 3.4 – Presence and expression of arsenic resistance (*ars*) genes in EA culture. Each box represents an *ars* gene and the specific gene number can be trace in Appendix B, Table B.9 for increased expression in transcriptomes and proteome and in Appendix B, Table B.11 for abundance in the genome (TPM−DNA), transcriptomes (TPM−RNA) and the proteome (protein abundance). # in Genome: total number of *ars* genes with TPM−DNA in genome. # in Transcriptome R/G: total number of *ars* genes with TPM−RNA in transcriptome R/G. # in Proteome: total number of *ars* genes expressed in proteome. # Total expressed: total number non−redundant *ars* genes with TPM−RNA in the transcriptomes, or expressed in proteome. MAGs indicated in red have >12% contamination.

Most of the *ars* genes encoding proteins induced in the proteome were also found to be induced in the transcriptomes. Three of the 33 EA culture *ars* genes induced in the proteome did not present any transcripts (*i.e.*, they were present in genome and had increased expression in the proteome only). This number was higher for TSB *ars* genes, where almost half (22 out of 46) of the proteins with increased expression did not have corresponding transcripts in the transcriptomes. This might be a consequence from the half−life mismatch of mRNA and proteins: while mRNA are short−lived, changes in protein inventories are slow (Moran et al., 2013).

In EA culture MAGs, with the exception of the two *Desulfovibrio*, there was evidence of increased expression of ArsA, Acr3 and/or ArsC in every MAG, suggesting that the efflux of intracellular As(III) is the predominant detoxification strategy in this microbiome. The TSB culture metagenome included the *arsH* gene, encoding an enzyme related to NADPH:FMN oxidoreductases. This enzyme is one of the four known MMAs(III) resistance genes (*arsH*, *arsK*, *arsI* and *arsP*) (J. Li et al., 2016). Every *arsH*−encoding contig also contained an *arsB* gene (Appendix B, Table B.10).



	Efflux	ArsA	ArsC	ArsR	ArsM	ArsH	Total
# in Genome	31	20	42	29	11	4	137
# in Transcriptome R	9	4	6	7	1	4	31
# in Transcriptome G	11	4	8	7	2	4	36
# in Proteome	8	13	15	4	2	4	46
# Total expressed	15	14	17	9	5	4	66

Figure 3.5 – Presence and expression of arsenic resistance (*ars*) genes in TSB culture. Each box represents an *ars* gene and the specific gene number can be trace in Appendix B, Table B.10 for increased expression in transcriptomes and proteome and in Appendix B, Table B.12 for abundance in the genome (TPM−DNA), transcriptomes (TPM−RNA) and the proteome (protein abundance). # in Genome: total number of *ars* genes with TPM−DNA in genome. # in Transcriptome R/G: total number of *ars* genes with TPM−RNA in transcriptome R/G. # in Proteom: total number of *ars* genes expressed in proteome. # Total expressed: total number non−redundant *ars* genes with TPM−RNA in the transcriptomes, or expressed in proteome. MAGs indicated in red have >12% contamination.

The Enterobacterales MAGs *Citrobacter* and *Raoultella*, were the only ones where contigs containing *arsH* were binned and both exhibited evidence of increased expression in the proteome. The Enterobacterales expression of ArsH might come as a response to environmental MMAs(III) produced by As−methylating soil microbes. MMAs(III) has been shown to induce expression of *arsP*, which encodes another MMAs(III) detoxifying protein, through an atypical ArsR repressor containing only two conserved cysteine residues (J. Chen et al., 2017). *Citrobacter* and *Raoultella* (formerly designated *Klebsiella*) (Drancourt et al., 2001) are genera of facultative anaerobic bacteria (Guentzel, 1996), consistent with ArsH proposed to detoxify MMAs(III) by oxidation to MMAs(V) coupled to the reduction of oxygen to water (J. Chen, Bhattacharjee, et al., 2015). Based on GhostKOALA taxonomic classification, the two unbinned *arsH* genes cluster with the two ArsB proteins identified in the proteome and pertain to the class *Enterobacteria*.

3.4.6 Active arsenic methylation

The aim of the present study was to identify the microorganisms catalysing As methylation in an anaerobic microbial community that had previously demonstrated high As methylation efficiency. Given that *arsM* genes have been reported as constitutively expressed in the absence of As(III) (K. Huang et al., 2018; J. Zhang et al., 2015), and therefore might not be induced by As(III), we sought all *arsM* genes and ArsM proteins expressed, both induced and not induced relative to the no−As control.

In the EA culture metagenome, 14 phylogenetically distinct *arsM* genes (encoding As(III) methyltransferases) were identified (Figure 3.4). Transcriptomic RNA reads were only detected for 3 out of the 14 *arsM* genes (Figure 3.4). These three genes belonged to contigs that were binned in MAGs 4 (MAG_4_*arsM*_#1 and MAG_4_*arsM*_#2) and 13, both identified as belonging to the order Clostridiales based on GhostKOALA (Figure 3.4). The proteomic results corresponded remarkably well to the transcriptomic findings and reported increased expression for two of the three genes identified above: one in MAG 4 (MAG_4_*arsM* #1) and one in MAG 13 (MAG 13_*arsM*) (Figure 3.4). The second *arsM* in MAG 4 (MAG_4_*arsM*_#2) only exhibited transcriptomic evidence for induction. Probing the GhostKOALA taxonomic classification of the *arsM* genes revealed that two of three *arsMs* were attributed to *Paeniclostridium sordellii* (MAG_4_*arsM*_#2 and MAG_13_*arsM*) and one to *Ruminococcaceae bacterium* CPB6 (MAG_4_*arsM*_#1), unclassified species also referred as *Clostridium bacterium* CPB6 (Subhraveti et al., 2018) (Figure 3.6). Thus, in EA, all three *arsM* genes showing evidence of their involvement in active As methylation pertain to fermenting microorganisms from the order Clostridiales. For one organism, multiple lines of evidence pointed to activity (transcriptome, proteome) and to attribution to the genus *Paeniclostridium* (GhostKOALA taxonomic attribution of MAG (Appendix B, Table B.1) and *arsM* gene and SSU 16S rRNA attribution of MAG, classified as [*Clostridium*] *sordellii* (Appendix B, Table B.3). In the TSB metagenome, 11 distinct *arsM* genes were identified (Figure 3.5). However, there was no overlap in the detection of RNA transcripts of *arsM* genes from three genes and the detection of peptides from two ArsM proteins (Figure 3.5). The two ArsMs detected in the proteome corresponded to *arsM* genes from MAG 8 (MAG_8_*arsM*_#1 and MAG_8_*arsM*_#3) and assigned by GhostKOALA to two Clostridiales strains: *Ruminococcaceae bacterium* CPB6 (MAG_8_*arsM*_#1 and *Clostridium botulinum* (MAG_8_*arsM*_#3) (Figure 3.6). Two of the three *arsM* genes detected in the transcriptomes were unbinned (NB_*arsM*_#1 and NB_*arsM*_#2) and their phylogenetic classification pointed to the Bacteroidetes strains *Mucilaginibacter gotjawali* (NB_*arsM*_#1) and *Niastella sp.* (NB_*arsM*_#2) and the third *arsM* (MAG_10_*arsM*) corresponded to the Betaproteobacterium *Oxalobacter formigenes* (perhaps inaccurately assigned to TSB MAG 10, identified as *Desulfovibrio*). Thus, in TSB, As methylation appears to be catalyzed by some fermenting Clostridiales along with potentially fermenting organisms from the phylum Bacteroidetes and either an SRB from the genus *Desulfovibrio* or an organism from the genus *Oxalobacter*.

Thus, the combination of metagenome, metatranscriptome, and metaproteome analyses

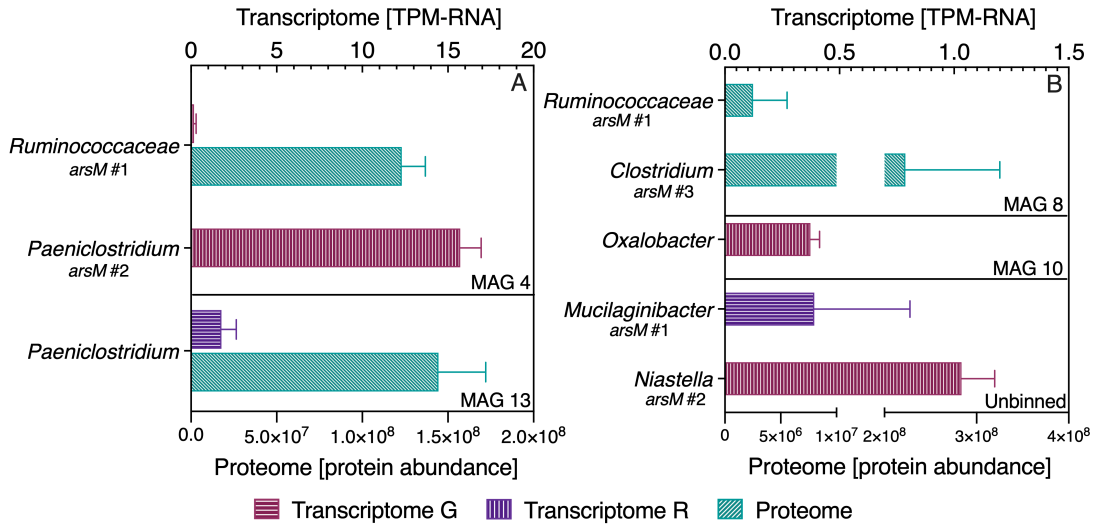


Figure 3.6 – Phylogeny and expression in +As condition of ArsM proteins in transcriptomes and proteomes. ArsMs from EA culture on panel A, ArsMs from TSB culture on panel B. Bar lengths represent mean and error bars, one standard deviation. The *arsM* number corresponds to the ones in Figures 3.4 and 3.5.

point to a role for fermenting microorganisms in As−methylation in paddy soil−derived microbial communities. Previous work had identified an As−methylating Clostridiales strain, *Clostridium sp.* BXM (P. P. Wang et al., 2015), that performed fermentation and DSR. However, very unfortunately, that strain has been lost (Dr. Peng Wao and Prof. Guo-Xin Sun, pers. commun.). The *Paeniclostridium* and *Clostridium* MAGs expressing ArsM did not harbour the potential for DSR (Figures 3.2 and 3.3) but did appear to catalyze lactic acid and mixed−acid fermentations.

As mentioned above, SRB have been proposed as drivers of As methylation based on a correlation in *arsM* gene abundance to that of *dsrB* and the decrease in As methylation by the addition of molybdate and monofluorophosphate (MFP), both chemical DSR inhibitors (Kim & Liesack, 2015; M. Wang et al., 2019). The present findings suggest that, in an efficient As−methylating community, metabolically active *Desulfovibrio* strains did not exhibit significant As−methylation activity (Figure 3.2), as none of their *ars* genes were transcribed nor translated in the EA culture. This result is particularly striking in light of the fact that the most abundant *arsM* genes in the metagenomic libraries were from *Desulfovibrio* (EA MAG 7) (Appendix B, Figures B.6 and B.7). In comparison, *arsM* genes from *Paeniclostridium* and *Ruminococcaceae* that were detected in the proteome and/or in the transcriptome were at least 8−times less abundant. Thus, it is very likely that, had *arsM* genes from *Desulfovibrio* species been expressed, they would have been detected in our analysis. This observation suggests the robustness of the conclusion that, in this system, fermenters are the key As−methylating organisms.

This is consistent with the findings of Reid et al., 2017, who showed that methylation efficiency

in the EA culture could be effectively inhibited (up to 90%) by decreasing the concentration of tryptic soy broth in the medium, effectively limiting the amount of fermentable substrate. Moreover, amendment with organic matter, like rice straw (Jia et al., 2013), has been shown to increase As methylation efficiency, suggesting a correlation of organic substrate addition, enriching fermenting communities, to As methylation.

However, the same is not true for the TSB culture. An *arsM* gene attributed to the *Desulfovibrio* MAG (MAG 10) was transcribed in the TSB culture (Figure 3.5). Although the ArsM protein was taxonomically classified by GhostKOALA as being close to *Oxalobacter*, a Betaproteobacterium (Figure 3.6), most of the other genes (78%) encoded in the same MAG were attributed to Deltaproteobacteria (Appendix B, Table B.2). Additionally, amino acid sequence alignment using BLAST in NCBI gives "Methyltransferase domain−containing protein (fragment) [uncultured Deltaproteobacterium]" as the top alignment. Thus, the attribution of this *arsM* gene is ambiguous and it is not possible to conclusively identify the organism to which it belongs. At this stage, we can only state that SRB may play a role in As methylation.

What is triggering anoxic As methylation in the microbiomes? If intracellular As(III) is the trigger, an efficient efflux of As(III) has been observed to preclude As methylation (Viacava et al., n.d.) and therefore absence of or lack of expression of Acr3/ArsB might trigger As methylation. Putative *ars* operons, represented by other *ars* genes encoded contiguously and in the same contig as the expressed *arsMs* are listed in Appendix B, Figure B.8. Interestingly, the two most highly expressed *arsM* genes or ArsM proteins (highest number of transcripts or the highest protein abundance) correspond to putative operons lacking genes annotated as either *acr3* or *arsB*. These are *arsM* assigned to *Paeniclostridium* (MAG_4_*arsM*_#2 in EA, contig k119_253) and to Clostridium (MAG_8_*arsM*_#3 in TSB, contig k119_8737). Although both MAGs contain additional contigs that encode As(III)−efflux systems, these efflux pumps are either not expressed or expressed at low levels. In the case of MAG_8_*arsM* #3 in TSB, the second contig (k119_19293) harbors an *acr3* gene that is not detected in the transcriptome nor are corresponding peptides observed in the proteome (Appendix B, Figure B.8). This suggests that the organism corresponding to this MAG may not express efflux system. Furthermore, in EA, MAG 4 includes a contig (k119_6159) that harbours the gene *acr3*, which is expressed in the proteome (Appendix B, Figure B.8). However, Acr3 in this contig is expressed to a lesser extent when compared to ArsM (Appendix B, Figure B.9). However, this is only speculation because both MAGs have a contamination >12% and their clustered contigs may come from more than one microorganism. Furthermore, in the high−quality EA MAG 13 classified as *Paeniclostridium*, the transporter Acr3 was more abundantly expressed compared to ArsM (Appendix B, Figure B.9) and in the only MAG encoding an *arsM* and no As(III)−efflux system, high−quality TSB MAG 6, the *arsM* was not expressed nor translated.

Another hypothesis is that nutrient competition could trigger As methylation. In this scenario, the synthesis of reduced methylarsenicals would represent a metabolic advantage resulting in the inhibition of competitors sensitive to those compounds. It was noted above that the majority of the MAGs expressed key enzymes for lactic and mixed−acid fermentation. Thus,

there are many community members depending on the same organic substrates. A possible function of As(III) methylation may be to inhibit other fermenters competing for the same substrate by unleashing MMAs(III) and its genotoxic effect. Thus, As methylation may serve a microbial warfare function. Clostridiales, the main As−methylating strains in the microbiomes, may be producing the toxin as a poison for strains unable to detoxify this compound. The expression of the MMAs(III)−resistance protein ArsH by Enterobacterales in the TSB culture provides support for this hypothesis. Although anoxic conditions do not favour oxidation of MMAs(III) by the ArsH enzyme.

The EA and TSB soil−derived cultures offered the opportunity to study active As methylation from paddy−soil microbiomes in an environment less complex than soil but still relevant. Fermenting bacteria, from the order Clostridiales, presented mRNA *arsM* transcripts and expressed ArsM, pinpointing them as active drivers in As−methylation in the two soil−derived cultures. The latter might be a consequence of lacking an efficient As(III) efflux system or a defensive response against nutrient competitors. As further indication of the pathogenic effect of MMAs(III), Enterobacterales strains, the most abundant community in the TSB enrichment, expressed ArsH. While *Desulfovibrio* MAGs were metabolically active in both cultures, only one *arsM* from all their encoded *ars* genes exhibited some expression as RNA transcripts. Further work is needed to elucidate why Clostridiales fermenters expressed ArsM. Specifically, the high−quality EA MAG expressing ArsM classified as *Paeniclostridium sp.*, could be further interrogated for metabolic characteristics, facilitating its isolation. This soil microbiome study, which combines metagenomics, metatranscriptomics and metaproteomics, is a significant contribution to the comprehensive elucidation of methylarsenical cycling.

4 Future development and concluding remarks

4.1 Recapitulation

The main objective of this work was to identify active microbial As methylators under anoxic conditions. In the first part of the research project, active As methylation was investigated using single−species, *arsM*−harboring and metabolically−diverse microorganisms. The results showed that all tested anaerobic bacteria were resistant to As but unable to efficiently methylate it, regardless of their functional encoded ArsMs, while the aerobic bacterial strains were sensitive to increasing As concentrations but were efficiently methylating the metalloid. Because the *Clostridium pasteurianum* Δ*acr3* mutant was constructed, it was possible to establish an inverse relationship between efficient efflux of intracellular As(III) and its methylation. This finding provided an explanation for the lack of efficient As methylation in As−resistant bacteria, which was observed in the study. Given that the ratio of methylated to total As detected in the *C. pasteurianum* Δ*acr3* mutant was much lower than the ones reported for aerobic strains, we propose a distinct role for As methylation in aerobic and anaerobic microorganisms. For aerobic As methylators, it is likely that the main function is that of As detoxification. In contrast, this study raises the possibility that the *arsM* gene in anaerobes represents an evolutionary advantage against competitors, as proposed by the research of Rosen and coworkers (J. Chen et al., 2019). Furthermore, the findings also suggested that As methylation observed in methanogens is not ArsM−catalyzed but a fortuitous consequence of cell lysis, releasing methyltransferases involved in methanogenesis and/or ArsM, if the latter was expressed during growth.

In the second part, anoxic arsenic methylators were sought in As−methylating anaerobic cultures derived from a rice paddy soil from the deltaic region of the Mekong river. This was done using three meta−omics techniques: metagenomics, metatranscriptomics and metaproteomics. A total of 36 metagenome−assembled genomes (MAGs) were reconstructed, phylogenetically classified, and probed for key metabolic and arsenic resistance genes. The four ArsMs expressed in the metaproteome belonged to Clostridiales fermenters, a strong indication of their role as As methylators. The ArsM−expressing MAGs were mostly incomplete

or contaminated, except for one high−quality MAG classified as a *Paeniclostridium*. An exciting finding was the expression of ArsH (MMAS(III)−oxidase) encoded in Enterobacterales MAGs, the most abundant community members of their soil−derived culture.

In the next sections, future avenues to unravel the role of anoxic As methylation, an outlook in the promising potential from meta−omics techniques in As research and a small review of a bioremediation technique based on As methylation are presented.

4.2 Role of arsenic methylation under anoxic conditions, future avenues

Very simply stated, As(V) is the prevalent As species in oxic environments and is strongly adsorbed into iron and aluminum oxides. Conversely, As(III) is prevalent in anoxic environments and, because it does not sorb as efficiently, it is mobile.

Rice paddy soils are defined by their frequent transition between wet and dry states, *i.e.*, between states of aerobiosis and anaerobiosis (S.-Y. Zhang et al., 2017). Soluble DMAs(V) and MMAs(V) and volatile trimethylarsine are commonly detected in rice paddy soils and flooded soil porewater (Mestrot, Feldmann, et al., 2011; S. Y. Zhang et al., 2015). Continuous flooding has been shown to increase the concentration of DMAs(V) in rice grains as compared to oxic conditions (P. P. Wang et al., 2015). DMAs(V) can represent >50% of total As in rice from USA (Zhao et al., 2013) and although it is less toxic than inorganic As, it remains a toxic substance. DMAs(V) was extensively used during the Vietnam war as the herbicide 'Agent Blue' (a combination of dimethylarsenic acid and its salt) to defoliate and kill plants providing food or cover (Bencko & Foong, 2017).

It can be said that the patterns of As methylation in rice paddy soils are a strong indication that anaerobic microbes may be drivers of this bio−transformation. Based on the results presented in this thesis, considering the taxonomy, distribution and abundance of the *arsM* gene within anoxic microbial communities provides evidence of the potential for As−methylating activity in specific taxa, but no information on the actual activity of the associated organisms.

As detailed in the previous chapters, there are two possible, not mutually exclusive, explanations proposed for the presence of As methylation under anoxic conditions:

1. The limiting factor for As methylation is the intracellular As(III) level that remains low due to efficient As(III)−efflux pumps in anaerobic prokaryotes. In this case, *arsM*−encoding microorganisms lacking As(III) transmembrane transporters should be targeted as putative active As methylators.

2. The persistence of the *arsM* gene in the genomes of anoxic prokaryotes confers an evolutionary advantage, underpinned by the production of toxic trivalent methylarsenicals. In this scenario, we can hypothesize that efficient anoxic As methylation will take place

in microbial communities rather than in pure cultures.

The first scenario was partially confirmed by the investigations performed with the *Clostridium pasteurianum* Δ*acr3* mutant. This mutant, lacking the only identifiable As(III) efflux pump in the strain, exhibited higher intracellular As concentrations as well as higher concentrations of soluble methylated As. However, methylation efficiency was much lower than that presented by aerobic strains or soil−derived cultures. One possibility, in the case of the *C. pasteurianum* Δ*acr3* mutant, is that there is another, as−of−yet unidentified, efflux system available for the As(III) extrusion. Indeed, the gene *arsA*, encoding the ATPase energizing ArsB, encoded in a *arsRDA* operon, different from the *arsRacr3* operon where the *acr3* gene was deleted, showed higher transcription levels in the presence of As(III) in the Δ*acr3* mutant than in the wildtype strain (WT) (623−fold vs. 18−fold normalized to *gyrA*, our unpublished data). Most probably, the increment in transcription is a consequence of the higher intracellular As concentration, however, ArsA is able to function with distinct and nonhomologous transporters making As(III) efflux more efficient (Castillo & Saier, 2010). Thus, it is conceivable that the extrusion of As(III) was still ongoing in the mutant, through a separate transmembrane transporter whose arsenic−resistance role has not yet been discovered. In order to tackle this question, the mRNA pool of transmembrane transporters from *C. pasteurianum* Δ*acr3* mutant vs. WT in the presence of As(III) could be compared to target putative novel As(III) transporters. A second option would be to generate Δ*arsB* and Δ*acr3* mutants of other *arsM*-encoding anaerobic strains, with available protocols for gene deletion, like *Geobacter metallireducens*, to enquire if higher As methylation efficiencies could be obtained.

As further explained in the next section, the meta−omics techniques could be a good strategy to identify microbial strains lacking known As(III)−efflux systems. In the metagenomic libraries, microbial members encoding *arsM* but lacking *arsB* or *acr3* could be identified and metabolically characterized to facilitate their isolation. Subsequently, these organisms without an *arsB* or *acr3* would be characterized for As methylation efficiency.

In the second scenario, we propose that *arsM* could only be expressed if the anaerobic prokaryote senses the presence of competitors. In this case, interspecies cell−to−cell communication would be needed and *arsM* expression would be likely linked to a second trigger factor, in addition to the presence of As(III).

In an attempt to test whether interspecies cell-to-cell communication plays a role in triggering arsenic methylation in *C. pasteurianum*, we tested whether spent medium from the EA and TSB microbial communities would enhance As methylation by the WT culture. The communities were grown in the absence of As, and their spent medium was filter−sterilized under anoxic conditions and used to grow *C. pasteurianum* in the presence of As. No increase in methylation was observed. However, there was an increase in the transcription of *arsM* with spent EA medium with As as compared to the control grown in fresh medium with As (43.6−fold increase normalized to expression of *gyrA, rho* and *rpsJ* or 3.6−fold with no normalization, our unpublished data). Other experiments testing cell-to-cell communication as a

trigger for As methylation could include testing other species with the spent medium as well as constructing a synthetic microbial community that includes only one organism harboring *arsM*. Finally, the isolation of active As methylators from the studied microbiomes expressing Acr3 or ArsB (*e.g.* the more complete *Paeniclostridium* strain) and their characterization for As methylation efficiency would also allow to test the hypothesis of the community being required for the expression of ArsM.

4.3 Outlook of using meta-omics techniques in studies of the As biogeochemical cycle

The meta-omics techniques, metagenomics, metatranscriptomics, metaprotemics and metabolomics 'global study of small molecules or metabolites in a particular physiological state of a community' (O'Malley, 2013), could allow a thorough understanding of the interactions of microorganisms with arsenic at the level of both single species and microbial communities.

Metagenomics provide taxonomic and functional profiles of the microbial community in a niche. Metatranscriptomics, the level of gene transcription in the community, represent a sensitive window to the immediate response of microbes to environmental conditions (such as arsenic concentrations or the presence of specific arsenic species). While the mRNA inventories are highly responsive and sensitive to shifts in conditions, the changes in protein levels are slower. Metaproteomes, may not reflect the immediate microbial response of the community but they show how microbes shape the environment because proteins, contrary to mRNA transcripts, are long−lived. Finally, metametabolomics provide a global overview of biochemical reactions involved in metabolism. In contrast with the other three techniques, its subject of study, metabolites, are highly variable in their chemical structures, properties and dynamic nature, limiting its application especially in complex environments such as soil.

Metagenomics, the first approach to emerge, has already shed light on the evolutionary history of arsenic bioenergetics (Rascovan et al., 2016) and allowed the discovery of an arsenic resistance gene (*arsN*, a putative As acetyltransferase) (Chauhan et al., 2008). Conversely, there is only one published study using metatranscriptomics to interrogate microbial As metabolism. Yet, it does not tackle arsenic−induced transcriptomic changes but rather, the abundance and distribution of expressed *ars* genes (Cai et al., 2013). A metaproteomics pipeline was published to study the changes in the gut microbiome taxonomy caused by arsenic ingestion (C.-W. Liu et al., 2018). Finally, an As−related metabolomics study was performed using human urine, to associate methylarsenicals and diabetes (Spratlen et al., 2019) but there are no studies of microbial communities published until now.

Metagenome characterization of microbial community members can highlight metabolic characteristics of microorganisms of interest, in this case, As−methylators, which could, in turn, be leveraged for their isolation. However, it may not be possible to isolate some microorganisms due to their nutrient requirements, extremely slow growth, obligate symbiosis

or because they belong to low−abundance species. In such cases, the meta−omics techniques would be the only way to learn about their metabolism, the enzymes they encode, and their role in biogeochemical cycles.

The meta−omics data gathered for this project is multidimensional. The analysis presented focused on the search for active As methylators by the identification of ArsM−expressing MAGs. However, further data mining could yield additional information:

1. Further characterization of the MAGs to find specific metabolic characteristics that make it possible to isolate the ArsM−expressing MAGs, specifically the more complete *Paeniclostridium* strain, and the ArsH−expressing Enterobacterales strains.
2. The protein sequences obtained from the metagenomic libraries could be queried further for the more−recently discovered arsenic−related proteins such as ArsP, ArsI, ArsJ (organosenicals permease extruding arseno−phosphoglycerate) or the proteins encoded in the operon *arsEFG* (providing resistance to aromatic arsenicals).
3. The *ars* operons could be reconstructed to look for genes of unknown function (novel *ars* genes).
4. The impact of As on other metabolic processes could be probed based on mRNA transcription data from transcriptome R.

4.4 Application of arsenic methylation as a bioremediation technique

Bioremediation consists in the acceleration of metabolic processes already taking place in nature by which microorganisms alter the chemical speciation of toxic metals and organic molecules rendering them less mobile or toxic (Antizar-Ladislao, 2010). *In situ* biologically-based remediation techniques are typically cheaper, safer, and more sustainable than conventional remediation techniques, which often rely on excavation and disposal of contaminated soils. Natural biological processes can be accelerated in two ways: i) by adding amendments (substrates, electron donors and acceptors, etc.) to stimulate indigenous microbial populations, called biostimulation, and/or ii) by injecting microbial cultures capable of catalyzing the desired chemical reaction, called bioaugmentation (biological augmentation).

Since arsenic can only change its chemical speciation, but not be degraded, it will always be present in the environment. However, bioremediation techniques directed to render As less toxic can be advantageous. These technologies could benefit millions of people around the globe currently depending on soil and groundwater with high As content for drinking and agriculture and millions of consumers of rice exported from countries with substantial As content in soil and groundwater.

Some examples of studies exploring the microbial potential for bioremediation of inorganic As include: As(III) oxidation to As(V) in contaminated aquifers by using aerobic As(III) oxidizers

combined with the coprecipitation of As(V) metal oxides like activated alumina and ferric chloride (Andrianisa et al., 2008; Ike et al., 2008) or As(V) reduction to As(III) by dissimilatory As reducers in combination with As(III)−sulfide precipitation and surface sorption onto iron sulfides (Upadhyaya et al., 2010).

The *in−situ* methylation of arsenic has been considered as a possible bioremediation strategy for As removal from soils and groundwater by means of volatilization (P. Wang et al., 2014). Trimethylarsine ($(CH_3)_3As$), the final product of the consecutive As methylation (Figure 1.5, Chapter 1), is volatile at atmospheric pressure. Other volatile species include the inorganic arsine (AsH_3) and the organic mono− and dimethyl arsines (CH_3AsH_2 and $(CH_3)_2AsH$, respectively) (Mestrot, Feldmann, et al., 2011). Bioaugmentation by highly−efficient, transgenic As−volatilizing organisms has received special attention. The *arsM* gene from *Rhodopseudomonas palustris* (RPArsM) has been cloned into *Sphingomonas desiccabilis*, *Bacillus idriensis* (S. Liu et al., 2011) and in *Pseudomonas putida* (P. Chen et al., 2017) yet only small improvements in As biovolatilization from soils were observed in comparison to the control soils. Meng et al., 2011, engineered a transgenic plant rice expressing RPArsM able to produce volatile arsenicals. Although a 10−fold increment in As volatilization was observed, compared to the wildtype plant, it corresponded to only 0.06% of total As in plant and no significant difference regarding total As content was observed between the control and the transgenic rice.

Despite the possibility to engineer successful genetically−modified As volatilizers, release of any transgenic technology into the environment is not desirable given the risk it may pose to the natural ecosystems. Besides, the fact that trimethylarsine is the dominant volatile species from field measurements (Mestrot, Feldmann, et al., 2011) is an indication of the already existing metabolic potential in the soil microbiomes.

Bioaugmentation by natural strains has only been studied employing aerobic strains. Arsenic-methylating fungi (Edvantoro et al., 2004) increased arsine production rates in field contaminated soils after 5 months and the efficient As methylator *Arsenicibacter rosenii* SM−1 boosted the production of trimethylarsine in soil slurry (K. Huang et al., 2016). However flooded paddy soils are seasonally oxygen depleted, limiting the application of aerobic strains.

The most effective way to enhance As volatilization so far has been the use of organic matter (OM) amendments such as clover, dried distillers grain (H. Huang et al., 2012), rice straw (P. Chen et al., 2017; Jia et al., 2013; Mestrot, Feldmann, et al., 2011) and cattle manure (Edvantoro et al., 2004; Mestrot, Feldmann, et al., 2011). Decomposition of organic amendments is a source of low molecular weight organic acids supporting the growth of microbes (R. L. Zheng et al., 2013). Negatively−charged functional groups from organic matter can desorb As(V) from iron oxides, making As more bioavailable (R. L. Zheng et al., 2013). The underlying mechanisms of biostimulation of As methylation by OM amendments requires further investigation.

Herein lies the future application of the research presented in this study. Identifying the metabolism of efficient As methylators, that leads to increased As volatilization, would allow either their isolation, for bioaugmentation, or the identification of the appropriate environ-

mental conditions, for their biostimulation.

However, arsines produced by As volatilization would have to be managed. Trimethylarsine has been reported as both highly toxic (Andrewes et al., 2003) and harmless (Cullen & Bentley, 2005). While monomethylarsine and dimethylarsine are considered genotoxic (Andrewes et al., 2003) and inorganic AsH_3, detected during the use of OM amendments (H. Huang et al., 2012), is lethal at low concentrations (Pakulska & Czerczak, 2006). In addition, the half−life of arsines in the presence of UV light has been calculated to be 7−8 days which allows them to be transported away from the remediated site (Mestrot, Merle, et al., 2011). Therefore, a system to recover the volatilized As, like temporary greenhouses where the air could be filtered through sorbent materials, such as activated carbon or alumina, would have to be engineered.

4.5 Concluding remarks

After one and a half centuries of research on arsenic methylation, considering as a starting point the research of Bartolomeo Gosio on fungi volatilizing As from wallpaper, the complexity of this process does not seem to abate. Arsenic methylation by prokaryotes was explained as a response to arsenic toxicity when first discovered but now we know that it may not be the case for all metabolisms; in anoxic systems, it could provide a competitive advantage.

Microorganisms have colonized nearly every ecological niche and their metabolisms are ancient and endlessly diverse. As more genomes are being sequenced, more genes involved in arsenic metabolism are being found. The more meta−omics techniques will be employed, the clearer and faster the connections between microbial community composition and the As biogeochemical cycle will become. All of this will make forecasting arsenic fate and health risks more tangible. Such approaches should pave the way for the utilization of microbial strategies to cope with arsenic toxicity to design new, efficient and environmentally−sound remediation techniques, providing safe drinking water and food. I believe that the understanding of fundamental mechanisms in arsenic cycling can and eventually will yield fruitful insights to handling this toxic and ubiquitous metalloid.

A Appendix Chapter 2

A.1 Methods

A.1.1 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 (Grosse-Honebrink, 2017). 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* (Grosse-Honebrink, 2017; Schwarz et al., 2017). *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 (O'Brien & Morris, 2009). 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 (Schwarz et al., 2017). Methylated vectors were then isolated and electroporated into *C. pasteurianum* as previously described (Pyne et al., 2013).

A.1.2 Isolation of the Δ*acr3* and Δ*pyrE* Δ*acr3* mutants

Transformed *C. pasteurianum* was plated on RCM agar supplemented with chloramphenicol. Positive colonies were confirmed for plasmid presence by colony PCR and positive clones were then grown up in liquid 2xYPG for 4h and patch plated onto RCM agar with chloramphenicol

or with chloramphenicol containing the inducer theophylline. Colonies that grew on RCM agar with chloramphenicol 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 (A.4). 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 (Pyne et al., 2013). The complementation vector pMTL_K_12_acr3 was then transformed into the Δ*pyrE* Δ*acr3* mutant and plated onto RCM agar supplemented with chloramphenicol. 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. 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_K_12_acr3 by patch plating on RCM agar with and without chloramphenicol. A PCR fragment of the *pyrE* locus was sequenced to confirm successful restoration of the *pyrE* allele and integration of *acr3*.

A.1.3 Arsenic methylation by *C. 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 was sampled as explained above, but volatile As was not assayed. To quantify intracellular As, a pellet from a 2−ml culture was collected (10 min, 16,873 g, 20 °C), washed twice with 1 ml 1x PBS buffer (10x solution, Fisher BioReagents, NH, US), re−suspended in 200 μl 1x 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.

A.1.4 Membrane-integrity assessment for *M. 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 1x 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^{TM} 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 555nm excitation wavelength was used in combination with a LP 560nm emission filter.

Table A.1 – Accession numbers for ArsM proteins and *arsM* genes. **M. mazei* Gö1 has two putative *arsM* genes (proteins: WP_011032612.1 and WP_011034171.1, genes: *MM_2243* and *MM_0661*). We overexpressed only gene *MM_2243*.

Strain	Protein	Gene
Geobacter metallireducens GS15	WP_004511671.1	*Gmet_2791*
Anaeromusa acidaminophila DSM 3853	WP_018703741.1	*C508_RS0111605*
Clostridium pasteurianum DSM 525	WP_004455181.1	*CLPA_c05000*
Streptomyces vietnamensis DSM 41927	WP_041130008.1	*SVTN_17945*
Arsenicibacter rosenii SM1	ANN44297.1=WP_071503661.10	*KU641426.1*
Methanosarcina mazei Gö1*	WP_011032612.1	*MM_2243*
Methanosarcina acetivorans C2A	WP_011023683.1=AAM07134	*MA_RS19700*

Table A.2 – Species−specific primers for *arsM* gene amplifications.

Oligonucleotide	Sequence (5'->3')
Gmet_2791_F	CTTTAAGAAGGAGATATACCATGGACAAGCAAAGGAATG
Gmet_2791_R	CAGTGGTGGTGGTGGTGGTGCGCTGGTTTTATTGCTTCGATAG
KU641426_F	CTTTAAGAAGGAGATATACCATGCAAACTGACGAACAAC
KU641426_R	CAGTGGTGGTGGTGGTGGTGCGCACAACAACCACTACCCGG
CLPA_c05000_F	CTTTAAGAAGGAGATATACCATGAAAAATAATATTAAGGGACAAG
CLPA_c05000_R	CAGTGGTGGTGGTGGTGGTGCGCATGCCAAGAGCTCTGTTTATATG
MM_2243_F	CTTTAAGAAGGAGATATACCATGGATGCGAATGAAAAAAAG
MM_2243_R	CAGTGGTGGTGGTGGTGGTGCGCCACCCGGGCTTTTAATTTC
MA_RS19700_F	CTTTAAGAAGGAGATATACCATGGATGCCGCTGAAAAAAAAG
MA_RS19700_R	CAGTGGTGGTGGTGGTGGTGCGCTACCCAGGCCTTTAATTTGAGAC
C508_RS0111605_F	CTTTAAGAAGGAGATATACCATGAATAAGATCAGGCAAAC
C508_RS0111605_R	CAGTGGTGGTGGTGGTGGTGCGCTTTTGAAGAAGGTTTACTTGC
SVTN_17945_F	CTTTAAGAAGGAGATATACCATGAGCGAGCAGTCCACCG
SVTN_17945_R	CAGTGGTGGTGGTGGTGGTGCGCTCGGGGAGACGCCGATTC

Table A.3 – Accession numbers for As(III)−efflux proteins and genes.

Strain	Protein	Gene	ArsB / Acr3
Geobacter metallireducens GS15	WP_004512216.1	*GMET_0520*	Acr3
Anaeromusa acidaminophila DSM 3853	C508_RS0112495	*CLPA_4880*	ArsB
Clostridium pasteurianum DSM 525	WP_004455168.1 (annotated as ArsB)	*CLPA_4880*	Acr3
Streptomyces vietnamensis DSM 41927	WP_041134619.1	*SVTN_36250*	ArsB
Methanosarcina mazei Gö1	WP_011032693.1	*MM_0730*	Acr3
Methanosarcina acetivorans C2A	AAM07296.1	*MA_3945*	Acr3

Table A.4 – List of primers used in preparing mutants of *C. pasteurianum*.

Primer name	Primer Sequence	Function
1AA	GGCCGACGTCAGTTTATTGAAGGAGGAAACAACAGT	LHA_forward
2AA	CTATATATTTTACATAATGATACCTACCTTTCTTTATTAAATTGAAAA	LHA_reverse
3AA	GGTATCATTATGTAAAATATATAGGTATATGATTATAGATATCCAACTT	RHA_forward
4AA	TGCAGGCGCGCCAGTCTCTCTATAAGAGTAACCTGAA	RHA_reverse
5AA	GCATGTCGACATTAGGCCCACCATAAAAGGGTTTTAGAGCTAGAAATAGCAA	Guide fragment forward
	GTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	
7AA	CGGCGCGGCCGCAGTAGAAGAAAAGGAATTTAACCTATGTTAAATATATG	Complementation_forward
8AA	CGACGCTAGCTTAATTTTCCTTTTTTAAACCAAAGTATTTCCTTC	Complementation_reverse
2AB	GTTAGTGGTGCAGTAGAAATATTGGTAG	Screening_forward
3AB	TATTGAGCCTTTAGCCCCAAGC	Screening_reverse
9Q	GCAGTGCAGGTTGGAACA	*pyrE* screening_forward
2R	ACCACCACTGATGAGCTGTT	*pyrE* screening_reverse

Table A.5 – Primers for RT−qPCR.

Oligo	Sequence (5'->3')
ArsM-Cpast-F-RT	AGCAATCGCTGATGTTGTGGTTCT
ArsM-Cpast-R-RT	CACCAGCAATGCATCCAACCCA
ACR3-Cpast-F-RT	TGTGTGGAACAGCTTAGCAGATGG
ACR3-Cpast-R-RT	ACCTGCATCCCCTGAAGTCCT
GyrA-Cpast-F-RT	TGGCGTTAGACTTACAAATGGTGA
GyrA-Cpast-R-RT	ACCCCAGATGCATTTCTTCCAA
Rho-Cpast-F-RT	GTTGAACAGGGACAGGATGTGGT
Rho-Cpast-F-RT	GTTGAACAGGGACAGGATGTGGT
Rho-Cpast-R-RT	AAGCCCCTGGATCAAGCCCT

```
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      DMTPEMITKSRRNADKANFGNVDFRLGELENLPVADGIVDVIISNCVINLSPEKAKVFSE
WP_011034171.1      DMTLEMVEKAQANARKYGYSNVEFRHGDIESLPVKDSSVDVIISNCVINLAPDKEKVFRE
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 A.1 – 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 (Ajees et al., 2012)

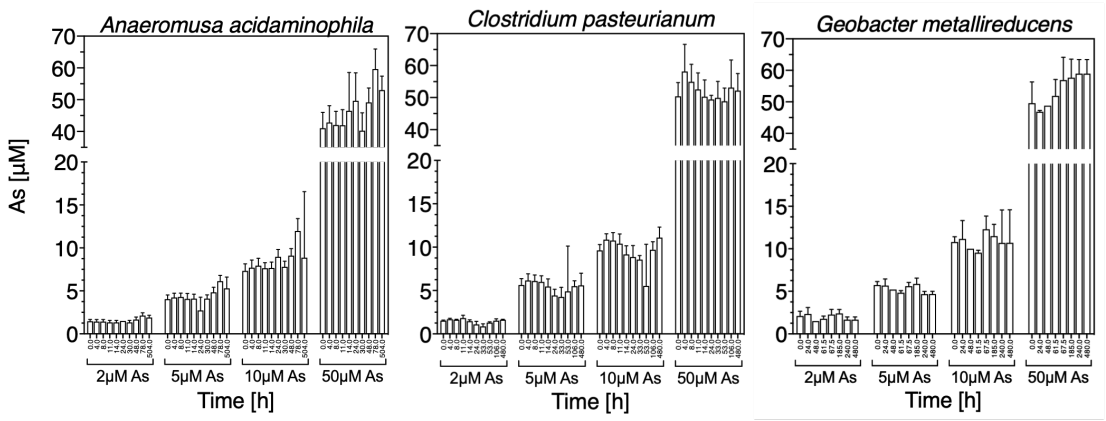


Figure A.2 – Total soluble arsenic concentration in filtered medium for each anaerobic bacterium culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 µM). Sampling times (x−axis) *Anaeromusa acidaminophila*: 0, 4, 8, 11, 14, 24, 30, 48, 78 and 504 hours. *Clostridium pasteurianum*: 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.

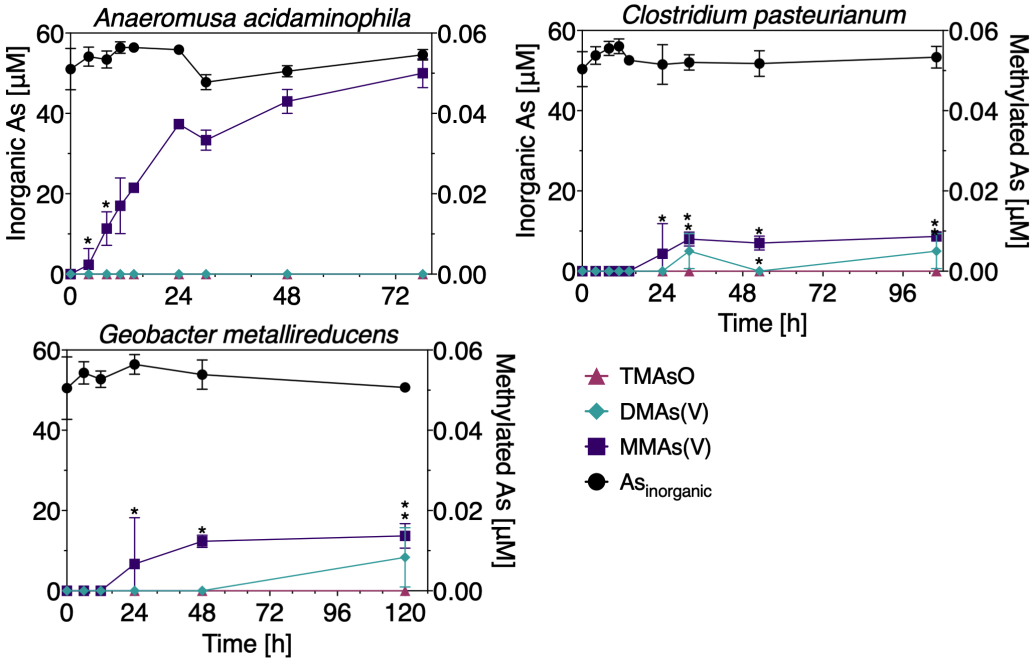


Figure A.3 – Concentration of soluble arsenic species in filtered medium from anaerobic bacterial strains cultures grown with 50 µM As(III) as the initial concentration. Data points represent the mean value and error bars, plus and minus one standard deviation. Time points marked with an asterisk are below LOQ but above LOD.

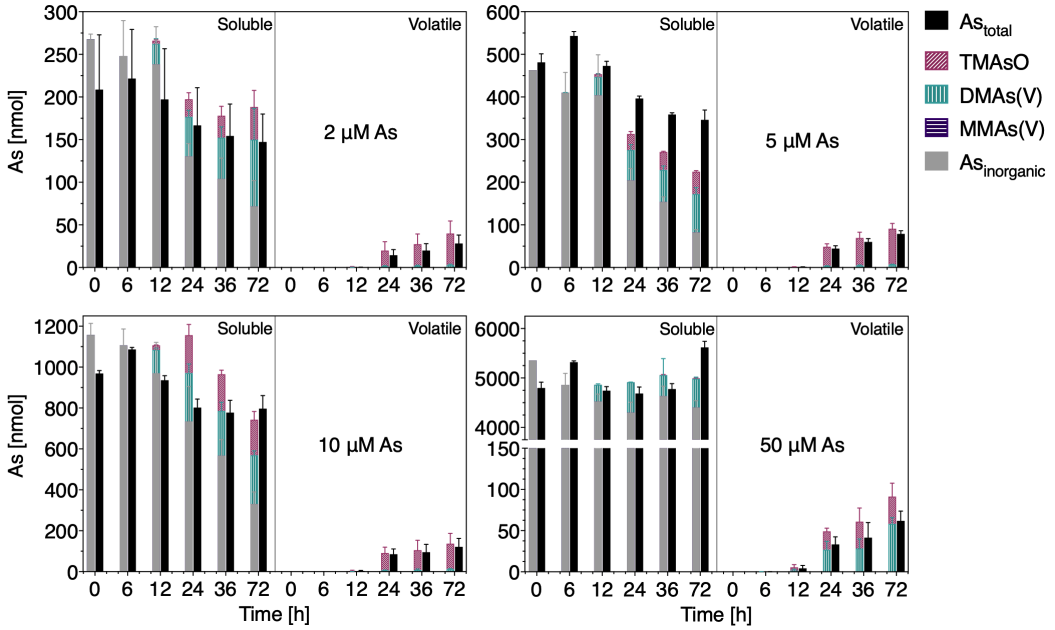


Figure A.4 – Concentration of soluble arsenic species in filtered medium (left panels) and volatile arsenic species chemo−trapped during medium flushing (right panels) from an *A. rosenii* culture grown in the presence of varying initial As(III) concentrations (2, 5, 10, and 50 µM). Bar height represents the mean value and error bars plus one standard deviation.

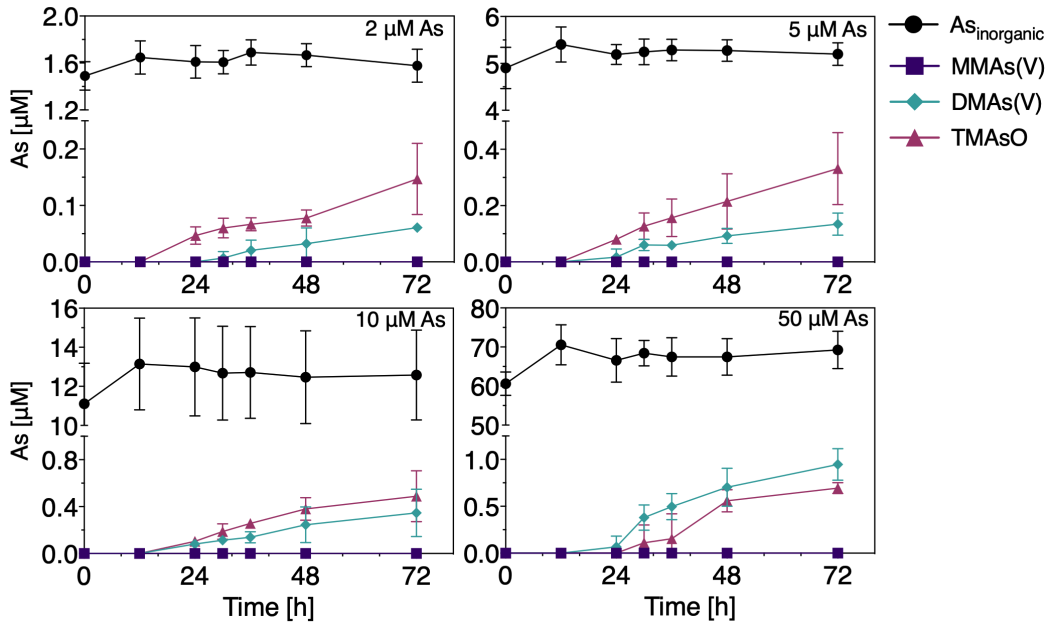


Figure A.5 – Concentration of soluble arsenic species in filtered medium from a *S. vietnamensis* culture grown in the presence of varying initial As(III) concentrations (2, 5, 10, and 50 µM). Data points represent the mean value and error bars, plus and minus one standard deviation.

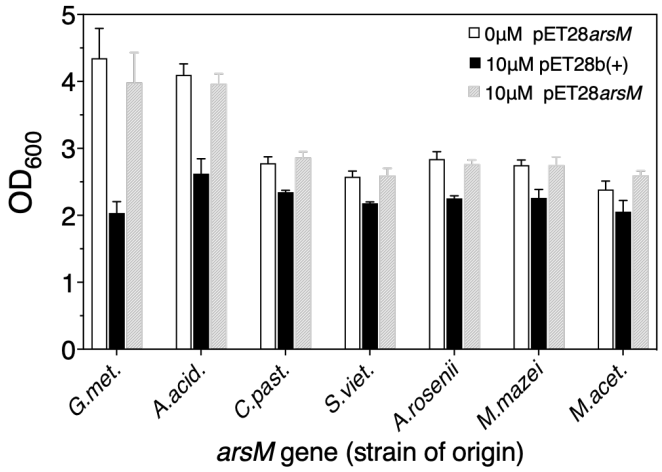


Figure A.6 – Growth as $OD_{6}00$ at the end of the exponential phase in a culture of *E. coli* AW3110−(DE3) expressing ArsM from the various species grown with 10−15 µM As(III) as the initial concentration (grey bars) or no As(III) (0 µM As(III)) (empty bars) and controls with empty plasmid pET28b(+) 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.

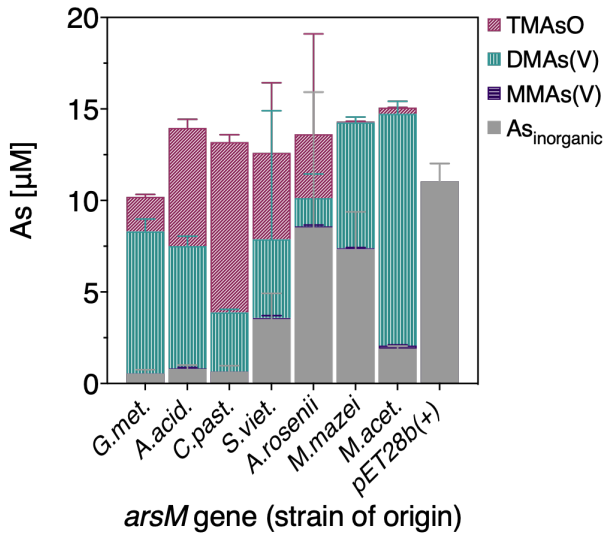


Figure A.7 – Concentration of soluble arsenic species in filtered medium after 72h in a culture of *E. coli* AW3110−(DE3) expressing ArsM from the various species or bearing the empty pET28b(+) plasmid. Bars represent mean and error bars plus one standard deviation. *G. met.= Geobacter metallireducens*; *A. acid.= Anaeromusa acidaminophila*; *C. past.= Clostridium pasteurianum*; *S. viet.= Streptomyces vietnamensis*; *A. rosenii= Arsenicibacter rosenii*; *M. mazei= Methanosarcina mazei*; *M. acet.= Methanosarcina acetivorans*.

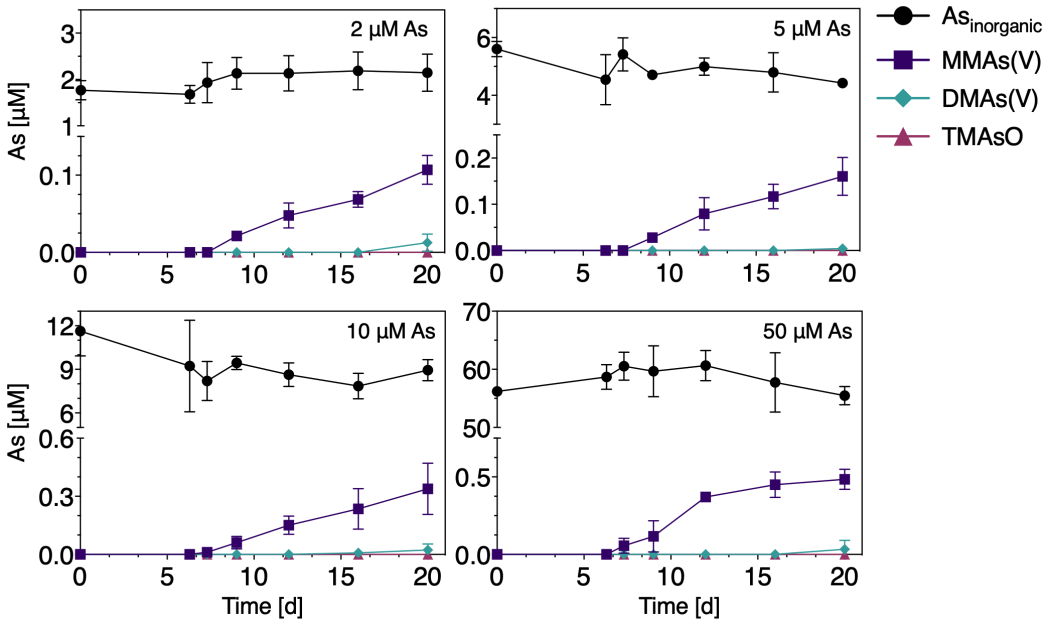


Figure A.8 – Concentration of soluble arsenic species in filtered medium from a *M. mazei* culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 µM). Data points represent the mean value and error bars, plus and minus one standard deviation. Time points marked with an asterisk are below LOQ but above LOD.

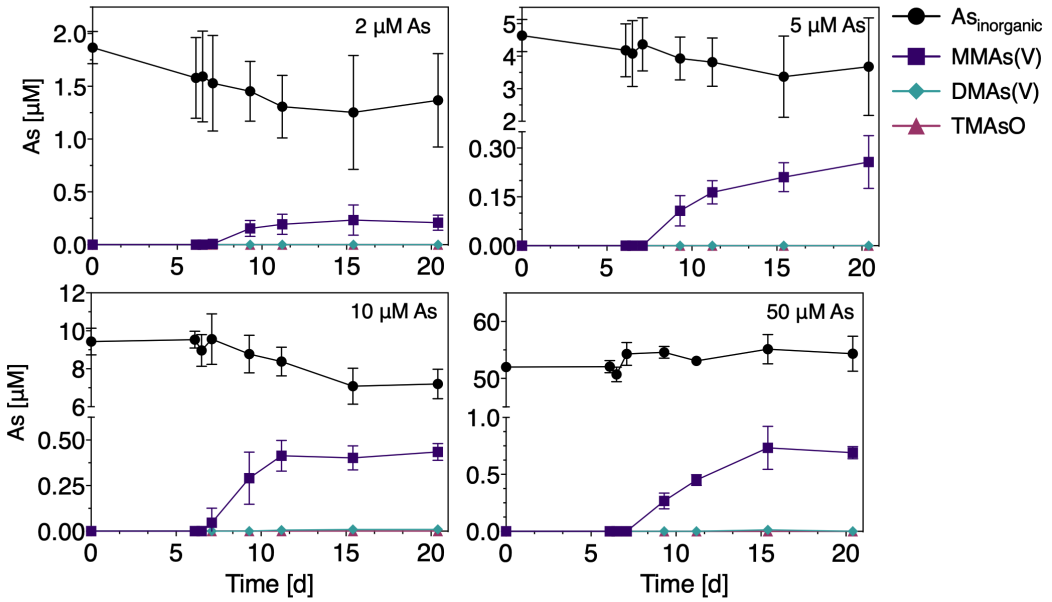


Figure A.9 – Concentration of soluble arsenic species in filtered medium from a *M. acetivorans* culture grown with varying initial As(III) concentrations (2, 5, 10, and 50 µM). Data points represent the mean value and error bars, plus and minus one standard deviation. Time points marked with an asterisk are below LOQ but above LOD.

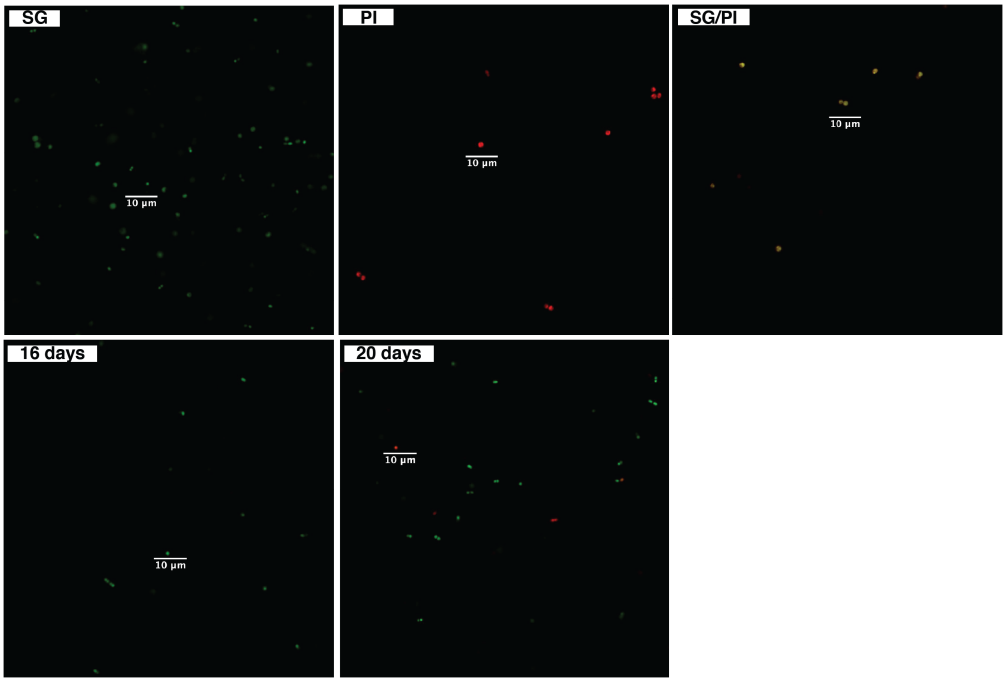


Figure A.10 – Fluorescence microscopy pictures of single−stained SG control (SG), single−stained PI control (PI), double−stained control (SG/PI), 16−days sample (16 days) and 20−day sample (20 days) of 10 µM As(III) *M. mazei* culture grown with 10 µM As(III) as initial concentration.

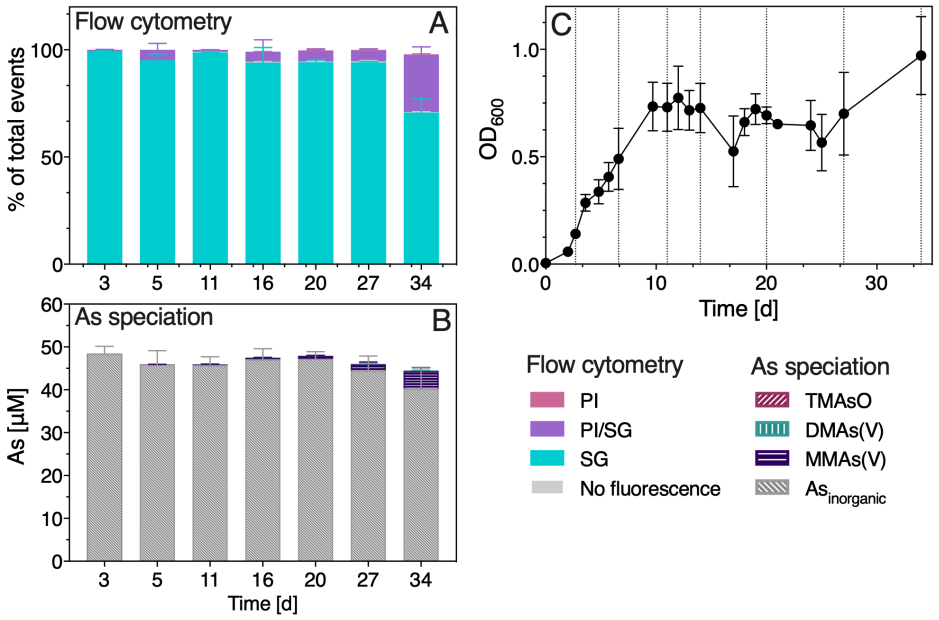


Figure A.11 – Assessment of membrane integrity for a *M. mazei* culture grown in the presence of 10 µM As(III) as the initial concentration: (A) Relative abundance of four populations (no fluorescence, double-stained (SG/PI) and single-stained: PI or SG) measured by flow cytometry after SG and PI staining 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.

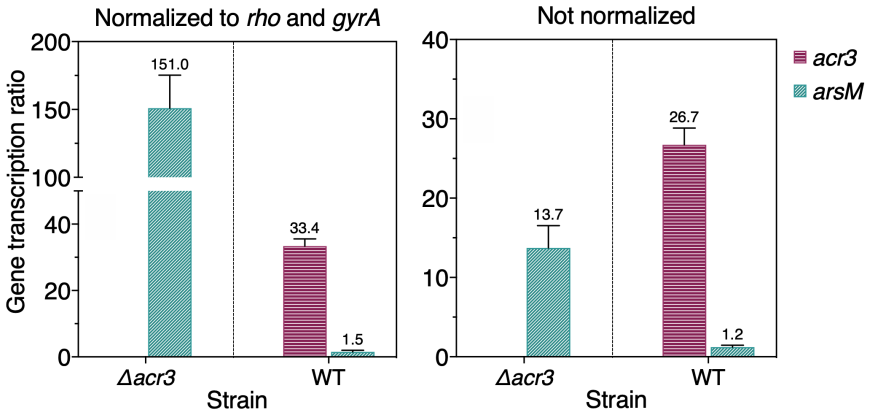


Figure A.12 – Expression of *arsM* and *acr3* in *C. pasteurianum* WT and Δ*acr3* mutant at mid−exponential phase (20 h for *C. pasteurianum* WT and no−arsenic control *C. pasteurianum* Δ*acr3* and 37 h for *C. pasteurianum* Δ*acr3* with 100 µM As(III)). The gene expression ratio represents cultures with 100µM As(III) relative to cultures without As(III) for each strain. Left panel: gene expression normalized to *gyrA* and *rho*. Right panel: gene expression without any normalization. Bars represent mean and error bars plus one standard deviation.

B Appendix Chapter 3

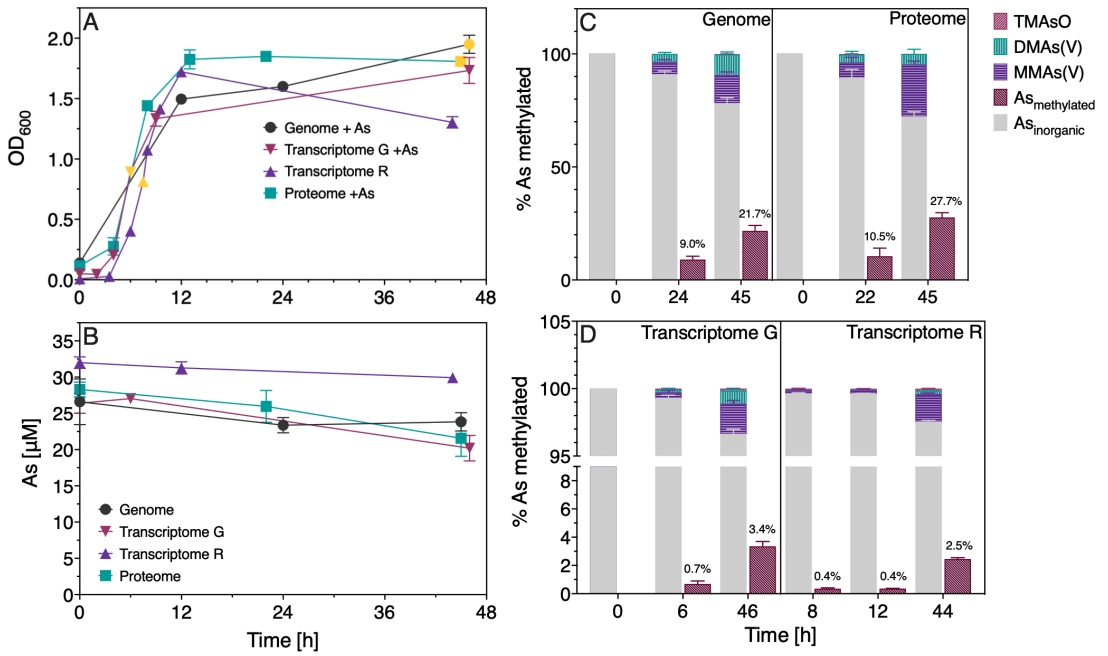


Figure B.1 – Arsenic methylation in EA culture. A. Growth curves from experiments. Yellow symbols correspond to sampling times for meta−omics analysis. B. Total soluble arsenic. C. Percentage of arsenic species in genome and proteome experiments. D. Percentage of arsenic species in transcriptome growth experiment (G) and response experiment (R). Points and bar heights represent mean and error bars one standard deviation.

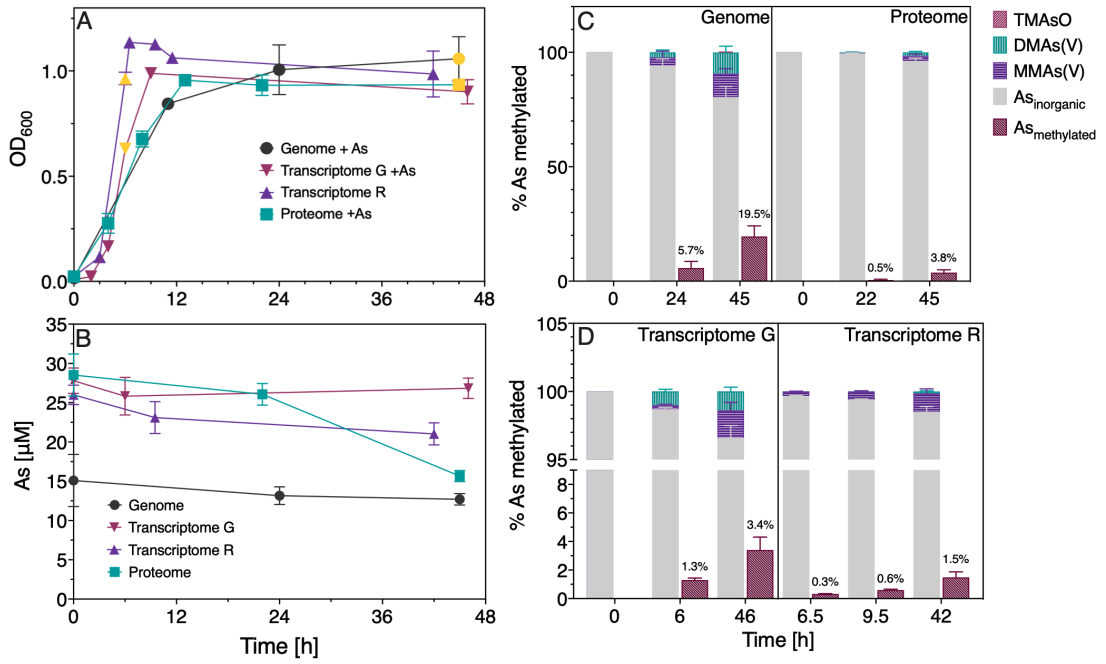


Figure B.2 – Arsenic methylation in TSB culture. A. Growth curves from experiments. Yellow symbols correspond to sampling times for meta−omics analysis. B. Total soluble arsenic. C. Percentage of arsenic species in genome and proteome experiments. D. Percentage of arsenic species in transcriptome growth experiment (G) and response experiment (R). Points and bar heights represent mean and error bars one standard deviation.

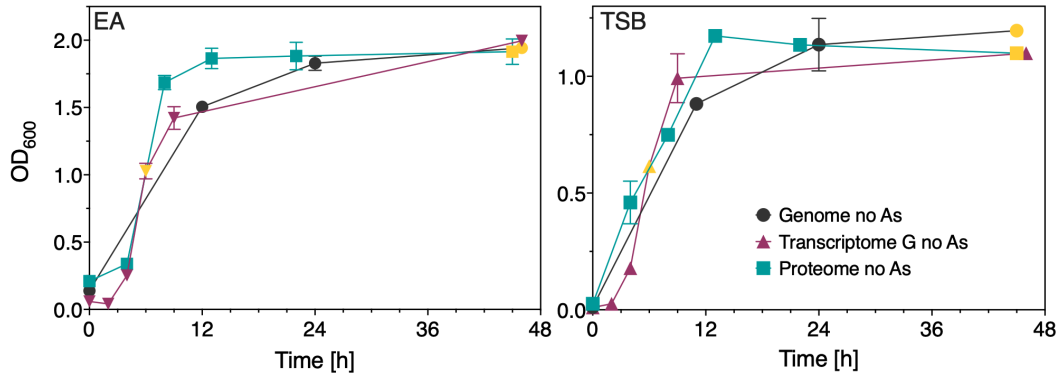


Figure B.3 – Growth in no−As controls. EA culture (left panel) and TSB culture (right panel). Yellow symbols correspond to sampling times for meta−omics analysis. Points represent mean and error bars one standard deviation.

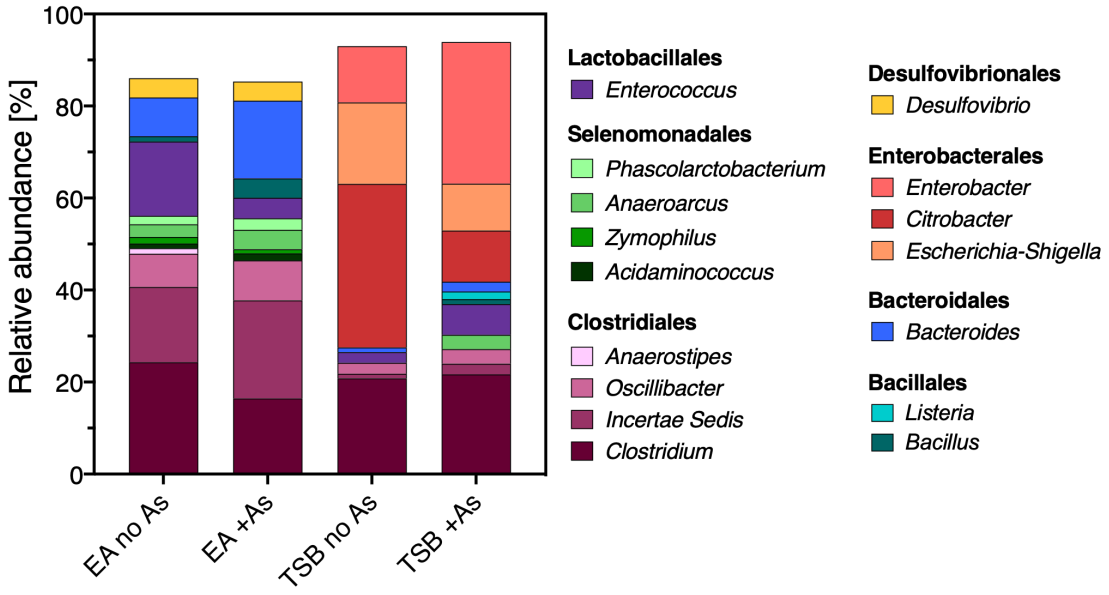


Figure B.4 – Abundant operational taxonomic units (OTUs) at genus level (> 1% relative abundance) identified from SSU rRNA sequences from soil–derived cultures.

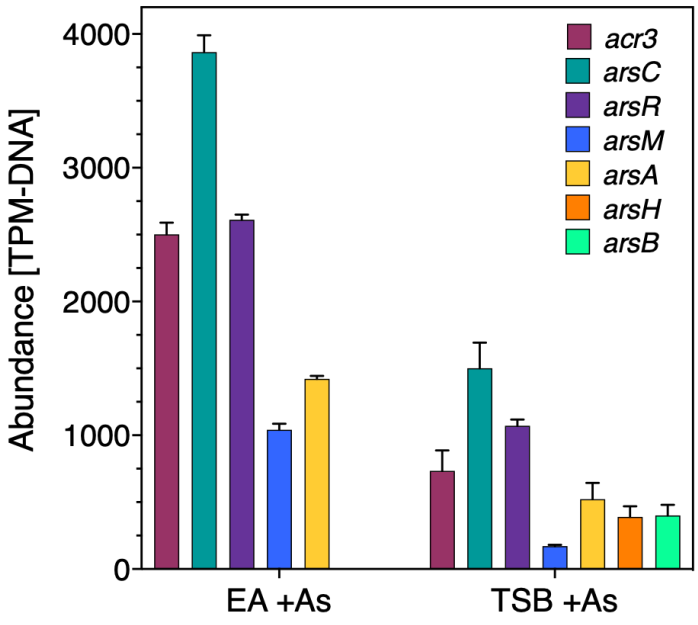


Figure B.5 – Abundance of *ars* genes in the metagenome of soil–derived cultures +As condition. The abundance corresponds to the normalized value of the number of reads mapped to each gene in transcripts per million (TPM–DNA). The bar height represents mean and error bars, one standard deviation.

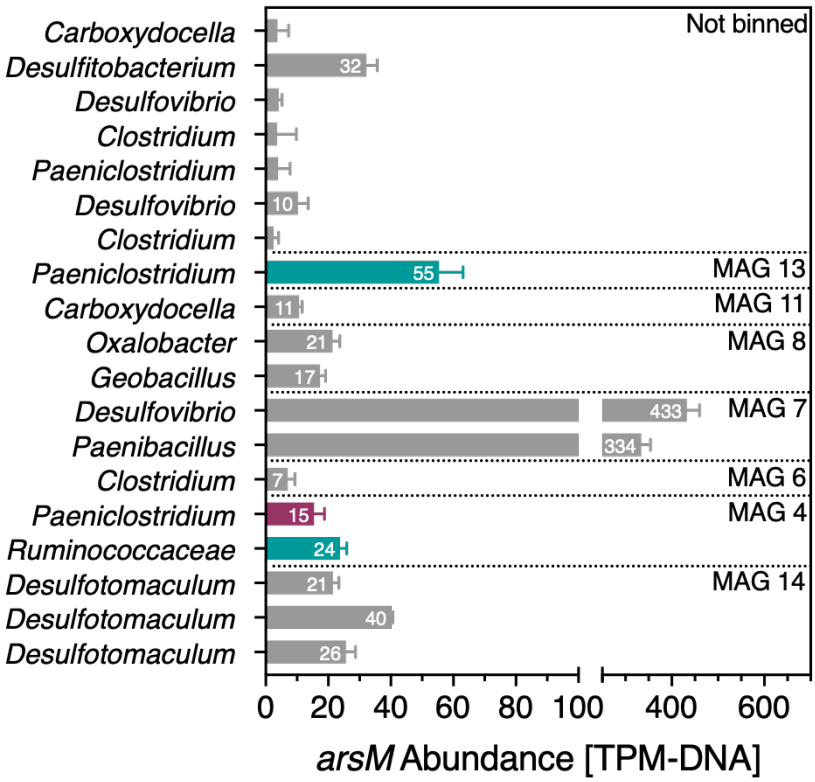


Figure B.6 – Taxonomic classification and abundance of *arsM* genes in EA culture. Taxonomic gene classification by GhostKOALA. Colored bars correspond to *arsM* genes expressed in the proteome (green) and in the transcriptome (red). Bar lengths represent mean and error bars one standard deviation.

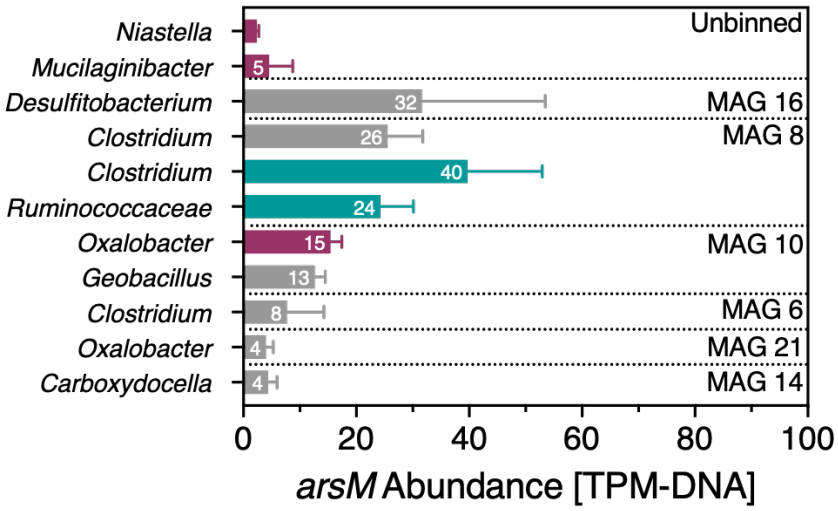


Figure B.7 – Taxonomic classification and abundance of *arsM* genes in TSB culture. Taxonomic gene classification by GhostKOALA. Colored bars correspond to *arsM* genes expressed in the proteome (green) and in the transcriptome (red). Bar lengths represent mean and error bars one standard deviation.

Cult.	MAG	ArsM Taxonomy	Contig	Putative *ars* operon	# genes in contig
EA	4	*Ruminococcaceae*	k119_6159	*arsM* *arsR* *acr3* *arsC* *arsR*	368
		Paeniclostridium	k119_253	*arsR* *arsR* *arsM*	32
	13	*Paeniclostridium*	k119_30669	*arsR* *acr3* *arsM*	164
TSB	8	*Ruminococcaceae*	k119_19293	*arsR* *arsC* *acr3* *arsM*	10
		Clostridium	k119_8737	*arsM* *arsR*	50
	10	*Oxalobacter*	k119_15818	*acr3* *arsC* *arsM* *arsR*	54
	NB	*Mucilaginibacter*	k119_10596	*arsM*	1
		Niastella	k119_29944	*arsM*	1

Transcriptome G and R
Genome and Proteome
Genome, Transcriptome and Proteome
Genome
Transcriptome G
Transcriptome R

Figure B.8 – Putative *ars* operons encoding an expressed *arsM*. NB stands for 'not binned'.

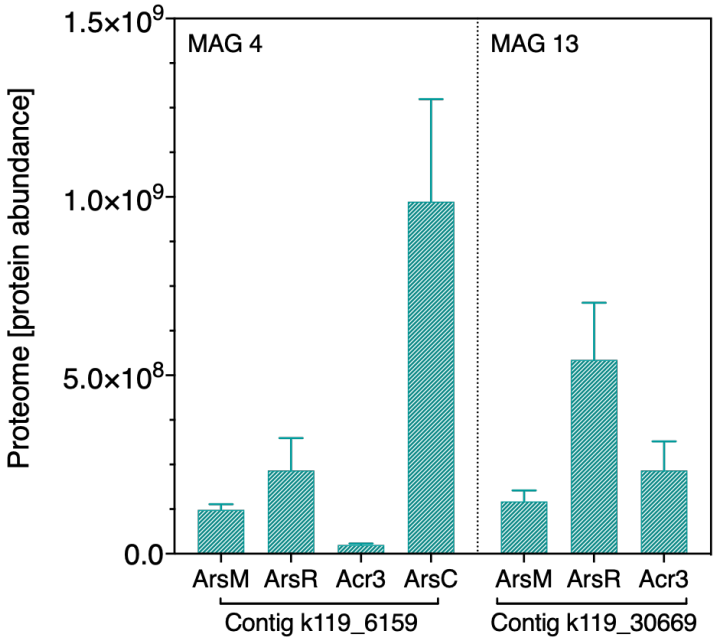


Figure B.9 – Protein abundance of putative *ars* operons encoding an expressed ArsM in EA culture. Bar heights represents mean and error bars, one standard deviation.

Table B.1 – Predominant OTUs in MAGs from EA culture according to GhostKOALA gene taxonomic classification. Percentages correspond to the number of genes within the corresponding OTU classification divided by the total number of genes in the corresponding MAG. The OTUs used to assign the supplementary taxonomic assignment under the 'Taxonomy' column are indicated in bold. In red, MAGs >12% contamination.

				GhostKOALA taxonomic classification					
MAG Group	Taxonomy*	Marker lineage	No of contigs	Phylum-Class	%	Genus or family	%	Genus 2.	%
1	*Bacteroides* (g)	Bacteroidales (UID2657)	41	Bacteroidetes	91%	***Bacteroides***	**68%**	*Parabacteroides*	3%
2	*Ruminococcaceae* (f)	Clostridiales (UID1212)	62	Firmicutes - Clostridia	81%	***Ruminococcaceae***	**43%**	*Oscillibacter*	5%
3	Clostridiales (o)	**Clostridiales (UID1212)**	67	Firmicutes - Clostridia	76%	*Flavonifractor*	12%	*Intestinimonas*	11%
4	Clostridiales (o)	root (UID1)	258	Firmicutes - Clostridia	82%	***Clostridium***	**40%**	***Clostridioides***	**12%**
5	*Clostridium* (g)	Clostridiales (UID1375)	336	Firmicutes - Clostridia	86%	***Clostridium***	**63%**	*Cellulosilyticum*	3%
6	*Clostridium* (g)	Bacteria (UID203)	266	Firmicutes - Clostridia	77%	***Clostridium***	**50%**	*Geosporobacter*	3%
7	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	18	Deltaproteobacteria	77%	***Desulfovibrio***	**66%**	*Desulfomicrobium*	2%
8	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	35	Deltaproteobacteria	78%	***Desulfovibrio***	**66%**	*Desulfomicrobium*	2%
9	*Enterococcus* (g)	Lactobacillales (UID544)	32	Firmicutes - Bacilli	94%	***Enterococcus***	**73%**	*Tetragenococcus*	7%
10	*Enterococcus* (g)	Lactobacillales (UID544)	105	Firmicutes - Bacilli	82%	***Enterococcus***	**53%**	*Tetragenococcus*	7%
11	*Oscillibacter* (g)	root (UID1)	197	Firmicutes - Clostridia	85%	***Oscillibacter***	**60%**	*Flavonifractor*	2%
12	*Oscillibacter* (g)	Clostridiales (UID1212)	174	Firmicutes - Clostridia	86%	***Oscillibacter***	**67%**	*Lachnoclostridium*	3%
13	*Paeniclostridum* (g)	Clostridiales (UID1120)	50	Firmicutes - Clostridia	86%	***Paeniclostridium***	**68%**	*Clostridium*	8%
14	*Pelosinus* (g)	root (UID1)	395	Firmicutes - Others	55%	***Pelosinus***	**48%**	*Selenomonas*	3%
15	*Phascolarctobacterium* (g)	Selenomonadales (UID1024)	26	Firmicutes - Others	70%	***Phascolarctobacterium***	**53%**	*Pelosinus*	7%
16	Selenomonadales (o)	Firmicutes (UID1022)	50	Firmicutes - Others	67%	***Selenomonas***	**33%**	***Pelosinus***	**19%**

* (o) order, (f) family or (g) genus.

Table B.2 – Predominant OTUs in MAGs from TSB culture according to GhostKOALA gene taxonomic classification. Percentages correspond to the number of genes under the corresponding OTU classification divided by the total number of genes in the corresponding MAG. In bold the OTUs used to assign the supplementary taxonomic assignment under the 'Taxonomy' column. In red MAGs >12% contamination.

				GhostKOALA taxonomic classification (relative abundance)					
MAG Group	Taxonomy*	Marker lineage	No of contigs	Phylum-Class	%	Genus or family	%	Genus 2.	%
1	*Bacteroides* (o)	Bacteroidales (UID2657)	516	Bacteroidetes	90%	***Bacteroides***	**65%**	*Parabacteroides*	32%
2	*Citrobacter* (g)	Enterobacteriaceae (UID5103)	755	Gammaproteobacteria - Enterobacteria	97%	***Citrobacter***	**53%**	*Salmonella*	12%
3	Clostridiales (o)	**Clostridiales (UID1212)**	277	Firmicutes - Clostridia	80%	*[Clostridium] cellulosi*	31%	*Ethanoligenens*	20%
4	Clostridiales (o)	**Clostridiales (UID1212)**	13	Firmicutes - Clostridia	72%	*Faecalibacterium*	17%	*Lachnoclostridium*	5%
5	Clostridiales (o)	**Clostridiales (UID1212)**	24	Firmicutes - Clostridia	78%	*Anaerotignum*	31%	*Clostridium*	6%
6	*Clostridium* (g)	Clostridiales (UID1375)	23	Firmicutes - Clostridia	78%	***Clostridium***	**45%**	*Paeniclostridium*	4%
7	*Clostridium* (g)	Clostridiales (UID1375)	306	Firmicutes - Clostridia	84%	***Clostridium***	**61%**	*Cellulosilyticum*	2%
8	*Clostridium* (g)	root (UID1)	147	Firmicutes - Clostridia	78%	***Clostridium***	**53%**	*Geosporobacter*	2%
9	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	247	Deltaproteobacteria	73%	***Desulfovibrio***	**61%**	*Desulfomicrobium*	2%
10	*Desulfovibrio* (g)	Deltaproteobacteria (UID3218)	62	Deltaproteobacteria	78%	***Desulfovibrio***	**65%**	*Desulfomicrobium*	2%
11	Enterobacterale s (o)	Enterobacteriaceae (UID5054)	74	Gammaproteobacteria - Enterobacteria	96%	***Cronobacter***	**18%**	***Citrobacter***	**8%**
12	*Enterococcus* (g)	Lactobacillales (UID544)	105	Firmicutes - Bacilli	82%	***Enterococcus***	**53%**	*Tetragenococcus*	7%
13	*Enterococcus* (g)	Lactobacillales (UID544)	48	Firmicutes - Bacilli	93%	***Enterococcus***	**73%**	*Tetragenococcus*	7%
14	*Oscillibacter* (g)	Bacteria (UID203)	551	Firmicutes - Clostridia	81%	***Oscillibacter***	**48%**	*Flavonifractor*	4%
15	*Oscillibacter* (g)	Clostridiales (UID1212)	152	Firmicutes - Clostridia	88%	***Oscillibacter***	**70%**	*Lachnoclostridium*	2%
16	*Pelosinus* (g)	Firmicutes (UID1022)	35	Firmicutes - Others	59%	***Pelosinus***	**52%**	*Selenomonas*	3%
17	*Phascolarctobacterium* (g)	Selenomonadales (UID1024)	15	Firmicutes - Others	67%	***Phascolarctobacterium***	**51%**	*Pelosinus*	7%
18	*Raoultella* (g)	Enterobacteriaceae (UID5103)	391	Gammaproteobacteria - Enterobacteria	97%	***Raoultella***	**73%**	*Citrobacter*	9%
19	*Ruminococcaceae* (f)	Clostridiales (UID1212)	36	Firmicutes - Clostridia	81%	***Ruminococcaceae***	**43%**	*Oscillibacter*	5%
20	Selenomonadales (o)	Firmicutes (UID1022)	132	Firmicutes - Others	66%	***Selenomonas***	**33%**	***Pelosinus***	**18%**

* (o) order, (f) family or (g) genus.

Table B.3 – Metaxa2 taxonomic assignment of contigs encoding SSU 16S rRNA sequences from MAGs of the EA culture. Taxa from each contig correspond to the sequence with the longest alignment from the top 10 hits from Metaxa2 output. Contigs marked with an asterisk (*) had one alignment as output. Abbreviations: Ident.: % of identity to the assigned taxonomy. Length: number of base pairs of the sequence alignment. In bold, the taxa used to assign the supplementary taxonomic assignment under the 'Taxonomy' column. In red MAGs >12% contamination.

MAG Group	Taxonomy	# Contig with 16S	Ident. (%)	Length	Phylum	Class	Order	Family	Genus	Species
3	Clostridiales	1*	88	75	Firmicutes	Bacilli	Bacillales	*Staphylococcaceae*	*Staphylococcus*	*Staphylococcus aureus subsp. aureus S0385*
4	Clostridiales	5	100	107	Firmicutes	Clostridia	**Clostridiales**	*Clostridiaceae*	*Clostridium*	*Clostridium carboxidivorans P7*
			94.3	193						*Clostridium sp. 6-39*
			94.52	73						*Clostridium botulinum NCTC 2916*
			94.44	72						*Clostridium botulinum NCTC 2916*
			94.69	207						*Clostridium botulinum NCTC 2916*
5	*Clostridium*	3	95.83	120	Firmicutes	Clostridia	Clostridiales	*Clostridiaceae*	***Clostridium***	*Clostridium botulinum NCTC 2916*
			93.75	48						*Clostridium botulinum NCTC 2916*
			96.25	80						*Clostridium botulinum NCTC 2916*
6	*Clostridium*	6	97.37	76	Firmicutes	Clostridia	Clostridiales	*Clostridiaceae*	***Clostridium***	*Clostridium botulinum NCTC 2916*
			96.04	101						*Clostridium sp. M-43*
			100	50						*Clostridium sardiniense*
			97.37	76						*Clostridium botulinum NCTC 2916*
			100	50						*Clostridium sardiniense*
			100	28	Firmicutes	Bacilli	Lactobacillales	*Leuconostocaceae*	*Leuconostoc*	*Leuconostoc sp. C2*
9	*Enterococcus*	1	99.83	598	Firmicutes	Bacilli	Lactobacillales	*Enterococcaceae*	***Enterococcus***	*Enterococcus hirae*
11	*Oscillibacter*	2	90.93	364	Firmicutes	Clostridia	Clostridiales	*Ruminococcaceae*	*Oscillospira*	*Oscillospira guilliermondii*
			90.42	167						*Oscillospira guilliermondii*
13	*Paeniclostridum*	5	100	44	Firmicutes	Clostridia	Clostridiales	*Peptostreptococcaceae*	*Incertae Sedis*	*[Clostridium] sordellii*
			100	40						*[Clostridium] sordellii*
			100	33						*[Clostridium] sordellii*
			100	33						*[Clostridium] sordellii*
			94.44	72					*Clostridium*	*Clostridium botulinum NCTC 2916*
14	*Pelosinus*	1	98.08	52	Actinobacteria	Actinobacteria	Corynebacteriales	*Halobacteroidaceae*	*Amycolicicoccus*	*Amycolicicoccus subflavus DQS3-9A1*
16	Selenomonadales	1	100	55	Actinobacteria	Actinobacteria	Propionibacteriales	*Propionibacteriaceae*	*Propionibacterium*	*Propionibacterium acnes SK137*

Table B.4 – Metaxa2 taxonomic assignment of contigs encoding SSU 16S rRNA sequences from MAGs of the TSB culture. Taxa from each contig correspond to the sequence with the longest alignment from the top 10 hits from Metaxa2 output. Contigs marked with an asterisk (*) had one alignment as output. Abbreviations: Ident.: % of identity to the assigned taxonomy. Length: number of base pairs of the sequence alignment. In bold, the taxa used to assign the supplementary taxonomic assignment under the 'Taxonomy' column. In red MAGs >12% contamination.

MAG Group	Taxonomy	# Contig with 16S	Ident. (%)	Length	Phylum	Class	Order	Family	Genus	Species
1	Bacteroides	1	93.18	1524	Bacteroidetes	Bacteroidia	Bacteroidales	*Bacteroidaceae*	***Bacteroides***	*Bacteroides caccae*
5	Clostridiales	2	93.2	103	Firmicutes	Clostridia	**Clostridiales**	*Lachnospiraceae*	*Cellulosilyticum*	*Clostridium lentocellum DSM 5427*
			93.75	112					*Incertae Sedis*	*Clostridium colinum*
6	*Clostridium*	1	97.25	109	Firmicutes	Clostridia	Clostridiales	*Clostridiaceae*	***Clostridium***	*Clostridium perfringens ATCC 13124*
7	*Clostridium*	3	95.83	120	Firmicutes	Clostridia	Clostridiales	*Clostridiaceae*	***Clostridium***	*Clostridium botulinum NCTC 2916*
			93.88	49						*Clostridium botulinum NCTC 2916*
			96.25	80						*Clostridium botulinum NCTC 2916*
8	*Clostridium*	7	100	52	Firmicutes	Clostridia	Clostridiales	*Clostridiaceae*	***Clostridium***	*Clostridium sardiniense*
			97.37	76						*Clostridium acetobutylicum DSM*
			95.45	88						*Clostridium sp. M-43*
			96.04	101						*Clostridium sp. M-43*
			100	51						*Clostridium sardiniense*
			94.12	51						*Clostridium botulinum NCTC 2916*
			98.85	87						*Clostridium carboxidivorans P7*
9	*Desulfovibrio*	1	100	98	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	*Desulfovibrionaceae*	***Desulfovibrio***	*Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774*
11	Enterobacterales	2	100	52	Proteobacteria	Gammaproteobacteria	**Enterobacterales**	*Enterobacteriaceae*	*Escherichia-Shigella*	*Salmonella enterica subsp. enterica serovar Matopeni*
			100	60						*Salmonella enterica subsp. enterica serovar Matopeni*
12	*Enterococcus*	2	100	32	Firmicutes	Bacilli	Lactobacillales	*Enterococcaceae*	***Enterococcus***	*Enterococcus sulfureus*
			100	32						*Enterococcus sulfureus*
13	*Enterococcus*	1	97.83	46	Firmicutes	Bacilli	Lactobacillales	*Enterococcaceae*	***Enterococcus***	*Enterococcus sulfureus*
16	*Pelosinus*	1*	100	59	Firmicutes	Clostridia	Halanaerobiales	*Halobacteroidaceae*	*Halobacteroides*	*Halobacteroides halobius*
17	*Phascolarctobacterium*	1	100	64	Firmicutes	Clostridia	Halanaerobiales	*Halobacteroidaceae*	*Halobacteroides*	*Halobacteroides halobius*
18	*Raoultella*	1*	77.82	257	Proteobacteria	Gammaproteobacteria	Pasteurellales	*Pasteurellaceae*	*Haemophilus*	*Haemophilus influenzae*
20	Selenomonadales	1*	86.67	75	Firmicutes	Bacilli	Bacillales	*Staphylococcaceae*	*Staphylococcus*	*Staphylococcus aureus subsp. aureus S0385*

Table B.5 – part 1 of 2. Metagenome–assembled genomes (MAGs) from EA culture, +As condition and no As control. 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 marker genes and the expected colocalization of these genes. Strain heterogeneity: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from more phylogenetically diverse sources. % 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.

MAG group	Taxonomy*	Condition	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of ArsMs	% of binned proteins assigned to MAG
1	*Bacteroides* (g)	As	Bacteroidales (UID2657)	98.88	0.74	0.00	3.81	41	0.0	3.3%
		no As	Bacteroidales (UID2657)	98.88	2.97	11.11	3.86	57	0.0	3.4%
2	*Ruminococcaceae* (f)	As	Clostridiales (UID1212)	97.82	0.34	0.00	2.16	62	0.0	2.3%
		no As	Clostridiales (UID1212)	97.82	0.34	0.00	2.16	56	0.0	2.4%
3	Clostridiales (o)	As	Clostridiales (UID1212)	94.9	0.67	100.00	2.51	67	0.0	2.7%
		no As	Clostridiales (UID1212)	98.85	0	0.00	2.54	41	0.0	2.8%
4	Clostridiales (o)	As	root (UID1)	83.33	170.83	3.97	10.09	258	2.0	10.9%
5		As	Clostridiales (UID1375)	98.39	0	0.00	3.96	336	0.0	4.1%
	Clostridium (g)	no As	Clostridiales (UID1375)	97.58	0	0.00	4.30	318	0.0	4.6%
6	*Clostridium* (g)	As	Bacteria (UID203)	91.38	82.76	12.50	8.90	266	1.0	9.7%
7	*Desulfovibrio* (g)	As	Deltaproteobacteria (UID3218)	100	0	0.00	3.25	18	2.0	3.1%
		no As	Deltaproteobacteria (UID3218)	90.83	1.23	75.00	3.05	46	1.0	2.9%
8	*Desulfovibrio* (g)	As	Deltaproteobacteria (UID3218)	99.23	0.68	100.00	3.42	35	2.0	3.2%
		no As	Deltaproteobacteria (UID3218)	100	7.74	64.71	3.46	27	2.0	3.3%
9	*Enterococcus* (g)	As	Lactobacillales (UID544)	99.63	0	0.00	2.74	32	0.0	2.8%
		no As	Lactobacillales (UID544)	99.63	0	0.00	2.73	32	0.0	2.8%
10	*Enterococcus* (g)	As	Lactobacillales (UID544)	99.31	4.62	0.00	4.18	105	0.0	4.6%
		no As	Lactobacillales (UID544)	99.31	4.62	0.00	4.19	108	0.0	4.7%
11	*Oscillibacter* (g)	As	root (UID1)	100	190.62	43.61	8.33	197	1.0	8.8%
		no As	root (UID1)	100	194	46.21	8.15	299	2.0	8.8%
12	*Oscillibacter* (g)	As	Clostridiales (UID1212)	98.66	1.34	50.00	2.92	174	0.0	3.2%
		no As	Clostridiales (UID1212)	97.99	1.46	25.00	3.09	197	0.0	3.4%
13	*Paeniclostridum* (g)	As	Clostridiales (UID1120)	95.07	1.41	0.00	3.51	50	1.0	3.9%

Table B.5 - part 2 of 2. Metagenome−assembled genomes (MAGs) from EA culture, +As condition and no As control. 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 marker genes and the expected colocalization of these genes. Strain heterogeneity: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from more phylogenetically diverse sources. % 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.

MAG group	Taxonomy*	Condition	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of ArsMs	% of binned proteins assigned to MAG
14	*Pelosinus* (g)	As	root (UID1)	60.42	56.25	100.00	6.45	395	3.0	6.9%
		no As	Firmicutes (UID1022)	93.9	1.96	60.00	3.46	225	1.0	3.7%
		no As	Firmicutes (UID1022)	86.94	0.95	50.00	3.03	183	2.0	3.3%
15	*Phascolarctobacterium* (g)	As	Selenomonadales (UID1024)	99.98	2.1	0.00	2.26	26	0.0	2.4%
		no As	Selenomonadales (UID1024)	99.98	2.1	8.33	2.25	19	0.0	2.4%
16	Selenomonadales (o)	As	Firmicutes (UID1022)	99.91	0	0.00	2.78	50	0.0	3.0%
		no As	Firmicutes (UID1022)	99.91	0	0.00	2.85	53	0.0	3.1%

* (o) order, (f) family or (g) genus.

Table B.6 – part 1 of 2. Metagenome assembled genomes (MAGs) from TSB culture, +As condition and no As control. 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 marker genes and the expected colocalization of these genes. Strain heterogeneity: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from more phylogenetically diverse sources. % 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.

MAG group	Taxonomy*	Condition	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of ArsMs	% of binned proteins assigned to MAG
1	*Bacteroides* (g)	As	Bacteroidales (UID2657)	96.38	10.3	83.72	2.89	516	0.0	4.3%
2	*Citrobacter* (g)	As	Enterobacteriaceae (UID5103)	89.24	3.57	70.97	2.84	755	0.0	5.5%
		no As	Enterobacteriaceae (UID5103)	99.38	0.99	25.00	4.95	88	0.0	5.3%
3	Clostridiales (o)	As	Clostridiales (UID1212)	95.21	2.52	0.00	2.80	277	0.0	3.0%
4	Clostridiales (o)	As	Clostridiales (UID1212)	100	0	0.00	2.17	13	0.0	3.3%
		no As	Clostridiales (UID1212)	96.41	0.68	0.00	2.95	173	0.0	3.2%
5	Clostridiales (o)	As	Clostridiales (UID1212)	98.66	0.89	0.00	2.32	24	0.0	3.2%
6	*Clostridium* (g)	As	Clostridiales (UID1375)	99.19	1.08	0.00	3.02	23	0.0	3.9%
		no As	Clostridiales (UID1375)	99.19	1.08	0.00	3.53	25	1.0	3.8%
7	*Clostridium* (g)	As	Clostridiales (UID1375)	98.12	0	0.00	4.43	306	0.0	5.1%
		no As	Clostridiales (UID1375)	88.71	0.27	100.00	4.58	242	0.0	4.9%
8	*Clostridium* (g)	As	root (UID1)	100	104.17	32.76	3.34	147	1.0	10.9%
		no As	root (UID1)	100	125	28.18	10.40	150	3.0	11.0%
9	*Desulfovibrio* (g)	As	Deltaproteobacteria (UID3218)	94.04	0	0.00	3.24	247	1.0	3.5%
		no As	Bacteria (UID203)	57.63	7.02	80.00	3.05	798	2.0	3.6%
10	*Desulfovibrio* (g)	As	Deltaproteobacteria (UID3218)	98.47	1.81	50.00	3.43	62	2.0	3.4%
		no As	Deltaproteobacteria (UID3218)	99.17	0.18	50.00	3.42	72	2.0	3.4%
11	Enterobacterales (o)	As	Enterobacteriaceae (UID5054)	98.76	1.83	15.00	5.55	74	1.0	4.8%
		no As	Enterobacteriaceae (UID5054)	98.26	1.51	5.88	4.38	66	0.0	4.7%
12	*Enterococcus* (g)	As	Lactobacillales (UID544)	99.31	4.62	0.00	3.53	105	1.0	4.8%
		no As	Lactobacillales (UID544)	98.94	4.62	0.00	4.19	123	0.0	4.8%
13	*Enterococcus* (g)	As	Lactobacillales (UID544)	99.63	0	0.00	4.67	48	0.0	3.2%
		no As	Lactobacillales (UID544)	99.63	0	0.00	2.86	42	0.0	3.2%

Table B.6 - part 2 of 2. Metagenome assembled genomes (MAGs) from TSB culture, +As condition and no As control. 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 marker genes and the expected colocalization of these genes. Strain heterogeneity: high values suggest the majority of reported contamination is from closely related organisms (*i.e.*, potentially the same species), while low values suggest the majority of contamination is from more phylogenetically diverse sources. % 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.

MAG group	Taxonomy*	Condition	Marker lineage	Completeness (%)	Contamination (%)	Strain heterogeneity	Genome size (Mbp)	# of contigs	# of ArsMs	% of binned proteins assigned to MAG
14	*Oscillibacter* (g)	As	Bacteria (UID203)	98.28	85.13	74.68	3.09	551	0.0	6.4%
		no As	Clostridiales (UID1212)	99.06	0	0.00	3.14	37	0.0	3.4%
15	*Oscillibacter* (g)	As	Clostridiales (UID1212)	98.99	0.67	100.00	4.35	152	0.0	3.3%
		no As	Clostridiales (UID1212)	98.35	1.01	100.00	3.03	142	0.0	3.3%
16	*Pelosinus* (g)	As	Firmicutes (UID1022)	100	0.63	0.00	4.55	35	0.0	3.5%
		no As	Firmicutes (UID1022)	100	0.63	0.00	3.23	33	1.0	3.5%
17	*Phascolarctobacterium* (g)	As	Selenomonadales (UID1024)	99.98	1.5	0.00	4.19	15	0.0	2.6%
18	*Raoultella* (g)	As	Enterobacteriaceae (UID5103)	97.73	12.01	39.18	6.60	391	0.0	7.3%
		no As	Enterobacteriaceae (UID5103)	96.81	4.13	26.92	5.74	196	0.0	6.2%
19	*Ruminococcaceae* (f)	no As	Clostridiales (UID1212)	97.99	1.01	0.00	2.14	36	0.0	2.4%
		As	Clostridiales (UID1212)	97.99	1.01	0.00	2.67	277	0.0	2.4%
20	Selenomonadales (o)	As	Firmicutes (UID1022)	99.91	0	0.00	10.27	132	3.0	3.3%
		no As	Bacteria (UID203)	82.76	1.72	0.00	2.55	155	0.0	2.9%

* (o) order, (f) family or (g) genus.

Table B.7 – part 1 of 5. Metabolic pathway completeness (%) from MAGs in culture EA metagenome. Predicted functions are represented on a scale from 0 to 100 denoting the fraction of completeness of a pathway or function within a MAG. Analysis done using KEGGDecoder.

	Bacteroides	*Desulfovibrio*	*Desulfovibrio*	*Enterococcus*	*Enterococcus*	*Oscillibacter*	*Oscillibacter*	*Ruminococcaceae*	*Clostridiales*	*Clostridia*	*Clostridium*	*Clostridium*	*Paeniclostridium*	*Pelosinus*	*Phascolarctobacterium*	*Selenomonadales*
Bin	7	1	14	26	2	8	23	3	28	24	13	0	19	4	15	20
MAG	1	7	8	9	10	11	12	2	3	4	5	6	13	14	15	16
glycolysis	89%	89%	89%	89%	89%	78%	67%	89%	89%	89%	89%	89%	89%	78%	67%	78%
gluconeogenesis	0%	89%	89%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
TCA Cycle	88%	50%	50%	0%	38%	50%	50%	12%	38%	62%	25%	38%	25%	62%	25%	50%
NAD(P)H-quinone oxidoreductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NADH-quinone oxidoreductase	63%	70%	77%	0%	0%	21%	7%	7%	21%	21%	14%	14%	7%	77%	14%	84%
Na-NADH-ubiquinone oxidoreductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F-type ATPase	100%	88%	88%	100%	100%	88%	88%	0%	100%	100%	75%	100%	100%	100%	88%	100%
V-type ATPase	0%	0%	0%	77%	77%	88%	88%	77%	0%	77%	77%	77%	0%	0%	66%	0%
Cytochrome c oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ubiquinol-cytochrome c reductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome o ubiquinol oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome aa3-600 menaquinol oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome c oxidase, cbb3-type	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome bd complex	100%	100%	100%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	100%
RuBisCo	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
CBB Cycle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
rTCA Cycle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Wood-Ljungdahl	0%	17%	17%	0%	0%	17%	0%	0%	17%	83%	0%	17%	67%	0%	0%	0%
3-Hydroxypropionate Bicycle	12%	6%	6%	6%	12%	12%	12%	12%	12%	12%	12%	12%	12%	12%	6%	6%
4-Hydroxybutyrate/3-hydroxypropionate	10%	0%	0%	20%	20%	30%	20%	20%	30%	40%	20%	30%	20%	10%	20%	20%
pectinesterase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
diacetylchitobiose deacetylase	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	100%	0%	0%	0%
glucoamylase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D-galacturonate epimerase	100%	100%	100%	0%	0%	0%	0%	0%	0%	100%	0%	100%	100%	0%	100%	0%
exo-poly-alpha-galacturonosidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
oligogalacturonide lyase	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
cellulase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
exopolygalacturonase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.7 - part 2 of 5. Metabolic pathway completeness (%) from MAGs in culture EA metagenome.

MAG	1	7	8	9	10	11	12	2	3	4	5	6	13	14	15	16
chitinase	100%	0%	0%	100%	0%	0%	0%	0%	0%	100%	0%	0%	100%	0%	0%	0%
basic endochitinase B	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
bifunctional chitinase/lysozyme	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
beta-N-acetylhexosaminidase	0%	100%	100%	0%	100%	100%	100%	100%	0%	0%	0%	100%	100%	100%	0%	0%
D-galacturonate isomerase	100%	0%	0%	0%	0%	100%	0%	0%	100%	0%	100%	0%	0%	0%	0%	0%
alpha-amylase	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%
beta-glucosidase	100%	0%	0%	100%	0%	100%	0%	0%	0%	100%	0%	100%	0%	0%	0%	0%
pullulanase	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%
ammonia oxidation (amo/pmmo)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydroxylamine oxidation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrite oxidation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
dissim nitrate reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
DNRA	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
nitrite reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitric oxide reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrous-oxide reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrogen fixation	0%	99%	99%	0%	0%	33%	0%	0%	0%	0%	99%	0%	0%	99%	99%	33%
hydrazine dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydrazine synthase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dissimilatory sulfate < > APS	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dissimilatory sulfite < > APS	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	100%	0%	0%
dissimilatory sulfite < > sulfide	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
thiosulfate oxidation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alt thiosulfate oxidation tsdA	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alt thiosulfate oxidation doxAD	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur reductase sreABC	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
thiosulfate/polysulfide reductase	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfhydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur disproportionation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur dioxygenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfite dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfite dehydrogenase (quinone)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfide oxidation	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur assimilation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DMSP demethylation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.7 - part 3 of 5.Metabolic pathway completeness (%) from MAGs in culture EA metagenome.

MAG	1	7	8	9	10	11	12	2	3	4	5	6	13	14	15	16
DMS dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DMSO reductase	0%	66%	66%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	66%	0%	0%
NiFe hydrogenase	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
ferredoxin hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
membrane-bound hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydrogen:quinone oxidoreductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NAD-reducing hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NADP-reducing hydrogenase	0%	0%	0%	0%	0%	25%	0%	0%	0%	75%	25%	100%	50%	50%	50%	25%
NiFe hydrogenase Hyd-1	0%	99%	99%	0%	0%	0%	0%	0%	0%	66%	0%	66%	0%	99%	0%	99%
thiamin biosynthesis	36%	73%	73%	45%	36%	45%	36%	45%	73%	55%	73%	73%	73%	73%	45%	64%
riboflavin biosynthesis	100%	100%	100%	100%	100%	0%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%
cobalamin biosynthesis	88%	62%	62%	12%	75%	62%	0%	25%	75%	88%	88%	88%	62%	88%	75%	75%
transporter: vitamin B12	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
transporter: thiamin	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
transporter: urea	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
transporter: phosphonate	0%	33%	0%	0%	99%	99%	33%	0%	0%	0%	99%	0%	0%	33%	0%	33%
transporter: phosphate	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Flagellum	4%	78%	78%	0%	0%	83%	74%	0%	74%	78%	57%	74%	78%	87%	0%	87%
Chemotaxis	12%	88%	88%	0%	0%	88%	88%	0%	88%	88%	88%	88%	75%	88%	0%	88%
Methanogenesis via methanol	0%	0%	0%	0%	0%	0%	0%	0%	0%	33%	0%	33%	33%	33%	0%	0%
Methanogenesis via acetate	0%	0%	0%	0%	0%	0%	0%	0%	0%	66%	0%	0%	66%	0%	0%	0%
Methanogenesis via dimethylsulfide, methanethiol, methylpropanoate	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via methylamine	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via trimethylamine	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via dimethylamine	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via CO2	0%	5%	5%	0%	0%	0%	0%	0%	0%	5%	0%	5%	5%	0%	5%	0%
Coenzyme B/Coenzyme M regeneration	0%	60%	60%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	60%	0%
Coenzyme M reduction to methane	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Soluble methane monooxygenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
methanol dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alcohol oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dimethylamine/trimethylamine dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Photosystem II	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Photosystem I	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome b6/f complex	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.7 - part 4 of 5. Metabolic pathway completeness (%) from MAGs in culture EA metagenome.

MAG	1	7	8	9	10	11	12	2	3	4	5	6	13	14	15	16
anoxygenic type-II reaction center	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
anoxygenic type-I reaction center	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Retinal biosynthesis	25%	25%	25%	75%	25%	25%	25%	25%	25%	50%	25%	50%	50%	25%	25%	25%
Entner-Doudoroff Pathway	75%	25%	25%	50%	75%	25%	25%	25%	75%	0%	50%	25%	0%	75%	0%	0%
Mixed acid: Lactate	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%
Mixed acid: Formate	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Mixed acid: Formate to CO2 & H2	0%	33%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	17%	17%
Mixed acid: Acetate	0%	50%	50%	100%	100%	0%	0%	0%	0%	0%	50%	0%	0%	0%	0%	50%
Mixed acid: Ethanol, Acetate to Acetylaldehyde	0%	0%	0%	0%	100%	100%	0%	0%	0%	100%	100%	100%	100%	0%	0%	0%
Mixed acid: Ethanol, Acetyl-CoA to Acetylaldehyde (reversible)	0%	0%	0%	100%	100%	100%	0%	0%	0%	100%	100%	100%	100%	100%	0%	0%
Mixed acid: Ethanol, Acetylaldehyde to Ethanol	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Mixed acid: PEP to Succinate via OAA, malate & fumarate	40%	38%	38%	6%	31%	38%	6%	38%	13%	13%	13%	44%	38%	52%	13%	52%
Naphthalene degradation to salicylate	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Biofilm PGA Synthesis protein	0%	0%	0%	0%	0%	0%	0%	0%	0%	25%	0%	0%	25%	50%	0%	0%
Colanic acid and Biofilm transcriptional regulator	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Biofilm regulator BssS	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Colanic acid and Biofilm protein A	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Curli fimbriae biosynthesis	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Adhesion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Competence-related core components	14%	14%	14%	49%	49%	21%	21%	21%	14%	28%	21%	28%	21%	28%	21%	21%
Competence-related related components	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Competence factors	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Glyoxylate shunt	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Anaplerotic genes	50%	25%	25%	25%	50%	50%	0%	25%	25%	25%	25%	50%	50%	25%	25%	25%
Sulfolipid biosynthesis	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
C-P lyase cleavage PhnJ	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
CP-lyase complex	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
CP-lyase operon	18%	0%	0%	0%	0%	0%	0%	0%	0%	0%	63%	9%	9%	0%	0%	0%
Type I Secretion	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	66%	0%	66%
Type III Secretion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Type II Secretion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	0%	8%	8%
Type IV Secretion	0%	0%	0%	8%	0%	8%	8%	8%	8%	0%	0%	8%	0%	0%	0%	0%
Type VI Secretion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Sec-SRP	72%	75%	83%	0%	0%	0%	0%	81%	75%	66%	66%	66%	0%	58%	75%	83%
Twin Arginine Targeting	50%	75%	75%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	0%	50%

Table B.7 - part 5 of 5. Metabolic pathway completeness (%) from MAGs in culture EA metagenome.

MAG	1	7	8	9	10	11	12	2	3	4	5	6	13	14	15	16
Type Vabc Secretion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Serine pathway/formaldehyde assimilation	40%	40%	40%	20%	30%	30%	20%	20%	20%	30%	30%	20%	20%	40%	20%	30%
Arsenic reduction	50%	75%	75%	25%	100%	100%	50%	75%	50%	100%	75%	100%	75%	100%	75%	75%
Cobalt transporter CbiMQ	0%	100%	100%	0%	100%	100%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%
Cobalt transporter CbtA	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cobalt transporter CorA	100%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Nickel ABC-type substrate-binding NikA	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Copper transporter CopA	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Ferrous iron transporter FeoB	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Ferric iron ABC-type substrate-binding AfuA	0%	100%	100%	0%	100%	100%	100%	0%	100%	100%	100%	100%	0%	100%	100%	100%
Fe-Mn transporter MntH	100%	100%	100%	100%	100%	0%	0%	0%	0%	100%	0%	100%	0%	100%	100%	100%
histidine	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%
arginine	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%
lysine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
serine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
threonine	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%
asparagine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%
glutamine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
cysteine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
glycine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
proline	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
alanine	0%	0%	0%	100%	100%	0%	0%	0%	0%	100%	0%	100%	100%	0%	0%	0%
valine	83%	83%	83%	33%	83%	83%	17%	83%	83%	66%	83%	83%	0%	83%	66%	83%
methionine	100%	100%	100%	0%	100%	100%	0%	100%	100%	100%	100%	100%	0%	100%	100%	100%
phenylalanine	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	100%	100%	100%	100%
isoleucine	83%	83%	83%	33%	83%	83%	17%	83%	83%	66%	83%	83%	0%	83%	66%	83%
leucine	100%	100%	100%	25%	100%	100%	25%	100%	100%	100%	100%	100%	0%	100%	75%	100%
tryptophan	100%	100%	100%	0%	100%	100%	100%	50%	100%	50%	100%	50%	0%	100%	100%	100%
tyrosine	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	100%	100%	100%	100%
aspartate	100%	100%	100%	0%	0%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%
glutamate	100%	100%	100%	0%	0%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%

Table B.8 – part 1 of 5. Metabolic pathway completeness (%) from MAGs in culture TSB metagenome. Predicted functions are represented on a scale from 0 to 100 denoting the fraction of completeness of a pathway or function within a MAG. Analysis done using KEGGDecoder.

	Bacteroides	*Desulfovibrio*	*Desulfovibrio*	*Citrobacter*	*Enterobacterales*	*Raoultella*	*Enterococcus*	*Enterococcus*	*Oscillibacter*	*Oscillibacter*	*Ruminococcaceae*	*Clostridiales*	*Clostridiales*	*Clostridiales*	*Clostridium*	*Clostridium*	*Clostridium*	*Phascolarctobacterium*	*Pelosinus*	*Selenomonadales*
MAG	1	9	10	2	11	18	12	13	14	15	19	3	4	5	6	7	8	17	16	20
Bin	9	3	25	6	26	29	15	27	0	30	32	21	28	5	17	20	35	14	36	4
glycolysis	89%	78%	89%	89%	89%	89%	89%	89%	78%	67%	89%	89%	89%	78%	89%	89%	89%	67%	78%	78%
gluconeogenesis	0%	78%	89%	89%	89%	89%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
TCA Cycle	75%	62%	50%	75%	88%	88%	38%	0%	50%	50%	12%	50%	50%	38%	25%	25%	38%	25%	62%	50%
NAD(P)H-quinone oxidoreductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NADH-quinone oxidoreductase	63%	70%	77%	84%	84%	84%	0%	0%	21%	21%	7%	21%	0%	21%	14%	21%	21%	14%	77%	84%
Na-NADH-ubiquinone oxidoreductase	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
F-type ATPase	100%	75%	88%	100%	50%	100%	100%	100%	100%	0%	0%	0%	100%	0%	100%	75%	100%	88%	100%	100%
V-type ATPase	0%	0%	0%	0%	0%	0%	77%	77%	88%	88%	77%	55%	77%	77%	77%	77%	0%	66%	0%	0%
Cytochrome c oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Ubiquinol-cytochrome c reductase	0%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome o ubiquinol oxidase	0%	0%	0%	75%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome aa3-600 menaquinol oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome c oxidase, cbb3-type	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome bd complex	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%
RuBisCo	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
CBB Cycle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
rTCA Cycle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Wood-Ljungdahl	0%	0%	17%	0%	0%	0%	0%	0%	0%	17%	0%	0%	0%	0%	0%	0%	67%	0%	0%	0%
3-Hydroxypropionate Bicycle	12%	6%	6%	24%	24%	24%	12%	6%	12%	12%	12%	12%	12%	6%	12%	12%	12%	6%	12%	6%
4-Hydroxybutyrate/3-hydroxypropionate	10%	0%	0%	20%	20%	20%	20%	20%	30%	30%	20%	10%	10%	20%	30%	20%	20%	20%	30%	20%
pectinesterase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
diacetylchitobiose deacetylase	0%	0%	0%	100%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%
glucoamylase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D-galacturonate epimerase	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%
exo-poly-alpha-galacturonosidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
oligogalacturonide lyase	100%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
cellulase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
exopolygalacturonase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.8 - part 2 of 5. Metabolic pathway completeness (%) from MAGs in culture TSB metagenome.

MAG	1	9	10	2	11	18	12	13	14	15	19	3	4	5	6	7	8	17	16	20
chitinase	100%	0%	0%	100%	0%	100%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
basic endochitinase B	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
bifunctional chitinase/lysozyme	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
beta-N-acetylhexosaminidase	0%	100%	100%	100%	100%	0%	100%	0%	100%	100%	100%	0%	100%	0%	100%	0%	100%	0%	100%	0%
D-galacturonate isomerase	100%	0%	0%	100%	100%	100%	0%	0%	100%	0%	0%	100%	100%	0%	0%	100%	0%	0%	0%	0%
alpha-amylase	0%	0%	0%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%	0%
beta-glucosidase	100%	0%	0%	100%	100%	100%	0%	100%	100%	0%	0%	100%	100%	0%	100%	100%	100%	0%	0%	0%
pullulanase	100%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%	0%
ammonia oxidation (amo/pmmo)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydroxylamine oxidation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrite oxidation	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dissim nitrate reduction	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DNRA	0%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%
nitrite reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitric oxide reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrous-oxide reduction	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
nitrogen fixation	0%	99%	33%	0%	0%	0%	0%	0%	0%	0%	0%	99%	0%	0%	0%	99%	99%	99%	99%	0%
hydrazine dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydrazine synthase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dissimilatory sulfate < > APS	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dissimilatory sulfite < > APS	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
dissimilatory sulfite < > sulfide	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
thiosulfate oxidation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alt thiosulfate oxidation tsdA	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alt thiosulfate oxidation doxAD	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur reductase sreABC	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
thiosulfate/polysulfide reductase	0%	0%	33%	66%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfhydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur disproportionation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur dioxygenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfite dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfite dehydrogenase (quinone)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfide oxidation	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
sulfur assimilation	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DMSP demethylation	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.8 - part 3 of 5. Metabolic pathway completeness (%) from MAGs in culture TSB metagenome.

MAG	1	9	10	2	11	18	12	13	14	15	19	3	4	5	6	7	8	17	16	20
DMS dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
DMSO reductase	0%	0%	66%	99%	99%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NiFe hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
ferredoxin hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
membrane-bound hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
hydrogen:quinone oxidoreductase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NAD-reducing hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NADP-reducing hydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	75%	75%	0%	25%	25%	75%	50%	50%	25%
NiFe hydrogenase Hyd-1	0%	99%	99%	99%	66%	0%	0%	0%	0%	0%	0%	0%	0%	0%	66%	0%	66%	0%	99%	99%
thiamin biosynthesis	27%	73%	73%	55%	82%	82%	36%	45%	45%	45%	45%	55%	45%	45%	55%	73%	73%	45%	73%	64%
riboflavin biosynthesis	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	100%	0%	100%	100%	100%	100%	100%
cobalamin biosynthesis	88%	62%	62%	88%	88%	88%	75%	12%	62%	0%	25%	25%	12%	88%	88%	88%	88%	75%	88%	75%
transporter: vitamin B12	0%	0%	0%	99%	99%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
transporter: thiamin	0%	0%	0%	99%	99%	99%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
transporter: urea	0%	100%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
transporter: phosphonate	0%	33%	0%	99%	0%	99%	99%	0%	99%	33%	0%	0%	99%	0%	0%	99%	0%	0%	33%	33%
transporter: phosphate	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Flagellum	4%	78%	78%	100%	100%	17%	0%	0%	74%	70%	0%	74%	70%	78%	74%	61%	74%	0%	87%	87%
Chemotaxis	12%	88%	88%	88%	75%	38%	0%	0%	88%	88%	0%	75%	75%	88%	50%	88%	88%	0%	88%	88%
Methanogenesis via methanol	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	33%	0%	0%	0%
Methanogenesis via acetate	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via methylamine, methanethiol, methylpropanoate	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via methylamine	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via trimethylamine	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via dimethylamine	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Methanogenesis via CO2	0%	0%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	5%	0%	0%
Coenzyme B/Coenzyme M regeneration	0%	60%	60%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	60%	0%	0%
Coenzyme M reduction to methane	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Soluble methane monooxygenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
methanol dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
alcohol oxidase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
dimethylamine/trimethylamine dehydrogenase	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Photosystem II	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Photosystem I	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cytochrome b6/f complex	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table B.8 - part 4 of 5. Metabolic pathway completeness (%) from MAGs in culture TSB metagenome.

MAG	1	9	10	2	11	18	12	13	14	15	19	3	4	5	6	7	8	17	16	20
anoxygenic type-II reaction center	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
anoxygenic type-I reaction center	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Retinal biosynthesis	25%	25%	25%	0%	100%	0%	25%	75%	25%	25%	25%	25%	25%	25%	50%	50%	50%	25%	25%	25%
Entner-Doudoroff Pathway	75%	25%	25%	100%	100%	100%	75%	50%	25%	25%	25%	50%	25%	0%	25%	50%	25%	0%	75%	0%
Mixed acid: Lactate	0%	0%	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
Mixed acid: Formate	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%
Mixed acid: Formate to CO2 & H2	0%	50%	33%	50%	50%	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	17%	0%	17%
Mixed acid: Acetate	0%	50%	50%	100%	100%	100%	100%	100%	0%	0%	0%	0%	50%	0%	50%	50%	0%	0%	0%	50%
Mixed acid: Ethanol, Acetate to Acetylaldehyde	0%	0%	0%	100%	100%	100%	100%	0%	100%	0%	0%	100%	100%	0%	100%	100%	100%	0%	0%	0%
Mixed acid: Ethanol, Acetyl-CoA to Acetylaldehyde (reversible)	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	0%	100%	0%
Mixed acid: Ethanol, Acetylaldehyde to Ethanol	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Mixed acid: PEP to Succinate via OAA, malate & fumarate	40%	38%	38%	88%	71%	88%	31%	6%	38%	31%	38%	38%	44%	38%	19%	13%	44%	13%	52%	52%
Naphthalene degradation to salicylate	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Biofilm PGA Synthesis protein	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	25%	0%	0%	50%	0%
Colanic acid and Biofilm transcriptional regulator	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Biofilm regulator BssS	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Colanic acid and Biofilm protein A	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Curli fimbriae biosynthesis	0%	0%	0%	99%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Adhesion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Competence-related core components	14%	14%	14%	14%	14%	14%	49%	42%	21%	21%	21%	21%	21%	21%	21%	21%	28%	21%	21%	21%
Competence-related related components	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Competence factors	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Glyoxylate shunt	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Anaplerotic genes	50%	25%	25%	75%	75%	75%	50%	25%	50%	50%	25%	25%	25%	50%	0%	25%	50%	25%	25%	25%
Sulfolipid biosynthesis	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
C-P lyase cleavage PhnJ	0%	0%	0%	100%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
CP-lyase complex	0%	0%	0%	100%	25%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
CP-lyase operon	18%	0%	0%	99%	18%	99%	0%	0%	0%	0%	0%	9%	0%	9%	0%	63%	9%	0%	0%	0%
Type I Secretion	33%	0%	0%	33%	33%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	66%
Type III Secretion	0%	0%	0%	53%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Type II Secretion	0%	0%	0%	0%	0%	92%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	0%	15%
Type IV Secretion	0%	0%	0%	0%	75%	8%	0%	8%	8%	8%	8%	8%	8%	8%	0%	0%	0%	0%	0%	0%
Type VI Secretion	0%	0%	0%	67%	67%	67%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Sec-SRP	72%	58%	75%	91%	83%	83%	0%	0%	0%	0%	72%	83%	0%	75%	81%	66%	66%	75%	75%	83%
Twin Arginine Targeting	50%	75%	75%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	50%	50%

Table B.8 - part 5 of 5. Metabolic pathway completeness (%) from MAGs in culture TSB metagenome.

MAG	1	9	10	2	11	18	12	13	14	15	19	3	4	5	6	7	8	17	16	20
Type Vabc Secretion	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Serine pathway/formaldehyde assimilation	40%	40%	40%	40%	40%	40%	30%	20%	30%	10%	20%	20%	30%	20%	20%	30%	20%	20%	40%	30%
Arsenic reduction	25%	75%	75%	75%	75%	100%	100%	25%	75%	50%	75%	50%	75%	50%	25%	100%	75%	75%	75%	75%
Cobalt transporter CbiMQ	0%	100%	100%	100%	0%	0%	100%	0%	0%	100%	0%	100%	0%	100%	100%	100%	100%	100%	100%	100%
Cobalt transporter CbtA	0%	0%	0%	100%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Cobalt transporter CorA	100%	0%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Nickel ABC-type substrate-binding NikA	0%	0%	0%	100%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%
Copper transporter CopA	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Ferrous iron transporter FeoB	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Ferric iron ABC-type substrate-binding AfuA	0%	100%	100%	100%	100%	100%	100%	0%	100%	0%	0%	0%	100%	0%	100%	100%	100%	100%	100%	100%
Fe-Mn transporter MntH	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%	0%	100%	100%	100%	100%	100%	100%
histidine	0%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%	100%	100%
arginine	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%
lysine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
serine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
threonine	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
asparagine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	0%	0%	100%
glutamine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
cysteine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
glycine	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
proline	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%
alanine	0%	0%	0%	100%	100%	100%	100%	0%	0%	0%	0%	0%	0%	100%	0%	0%	100%	0%	0%	0%
valine	83%	83%	83%	100%	100%	100%	83%	33%	83%	33%	17%	17%	83%	0%	0%	83%	66%	66%	83%	83%
methionine	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
phenylalanine	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%
isoleucine	83%	83%	83%	100%	100%	100%	83%	33%	83%	33%	83%	17%	83%	0%	0%	83%	66%	66%	100%	100%
leucine	100%	100%	100%	100%	100%	100%	100%	25%	100%	50%	100%	50%	100%	100%	0%	100%	100%	75%	100%	100%
tryptophan	50%	100%	100%	100%	100%	100%	0%	0%	100%	100%	50%	50%	50%	100%	0%	100%	100%	100%	100%	100%
tyrosine	0%	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	100%
aspartate	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%
glutamate	100%	100%	100%	100%	100%	100%	0%	0%	100%	100%	0%	0%	0%	100%	100%	100%	100%	100%	100%	100%

Table B.9 – part 1 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from EA culture. The values correspond to gene/protein abundance in the +As condition divided by the gene/protein abundance in the no−As control (values in green) or, when no expression was reported in the no−As condition for that gene/protein, the abundance values from the +As condition (values in black) in transcripts per million (TPM−RNA) for transcriptomes and in protein abundance for proteome. Data shown for the three biological replicates for transcriptome R and average from the three biological replicates fro transcriptome G and proteome. Contig: contig number assinged by Megahit. Protein id: id number for the protein−encoding genes found by Prodigal. Match identity: percentage of identity between clustered proteins from +As condition vs. no−As control by CD-HIT-2D. Annotation: protein annotation according to GhostKOALA. An asterisk (*) in protein annotation means no matching protein was clustered by CD-HIT-2D from the no−As control. N. column: corresponds to the gene number in Figure 4. NB in Bin column means the gene was not clustered in any bin. ArsM proteins are in bold.

							Transcriptome R			Transcriptome G	Proteome
MAG	Bin	Contig	Protein id	Match identity	Annotation	N.	rep1	rep2	rep3	AVG	AVG
1	7	k119_11963	k119_11963_422	100%	ArsC	1	0	0	0	0.1	44110773.3
1	7	k119_11963	k119_11963_423	100%	Acr3	1	0	0	0	0	0
2	3	k119_7198	k119_7198_17	100%	ArsR	1	0	0	0	0	0
2	3	k119_7198	k119_7198_14	100%	ArsA	1	0.1	0.1	0.0	0	2811257644
2	3	k119_7198	k119_7198_16	no match	Acr3*	1	0	0	0	0	0
3	28	k119_6934	k119_6934_5	100%	ArsR	1	0	0	0	0.3	0
3	28	k119_6934	k119_6934_2	100%	ArsC	2	0	0	0	0.4	0
3	28	k119_6748	k119_6748_20	100%	ArsC	1	0	0	0	0	0
4	24	k119_8323	k119_8323_82	100%	ArsA	2	0	0	0	0	0
4	24	k119_8323	k119_8323_84	100%	ArsA	3	0	0	0	0	0
4	24	k119_6159	k119_6159_77	100%	ArsR	2	0	0	0	0.1	233511106.7
4	24	k119_6159	k119_6159_80	100%	ArsR	3	0	0	0	0	0
4	**24**	**k119_6159**	**k119_6159_75**	**100%**	**ArsM**	**1**	**0.0**	**0.0**	**0.0**	**0.2**	**123085395.3**
4	24	k119_6159	k119_6159_79	100%	ArsC	2	0	0	0	0.2	1010620344
4	24	k119_6159	k119_6159_78	100%	Acr3	4	0.2	0	0.1	0.6	24360044.67
4	24	k119_32423	k119_32423_101	100%	ArsA	6	0	0	0	0	0
4	24	k119_32423	k119_32423_99	100%	ArsA	5	0	0	0	0	0
4	24	k119_253	k119_253_29	100%	ArsR	6	0	0	0	2.8	0
4	24	k119_253	k119_253_30	98%	ArsR	1	0	0	0	3.5	0
4	**24**	**k119_253**	**k119_253_32**	**99%**	**ArsM**	**2**	**0**	**0**	**0**	**5.0**	**0**
4	24	k119_20241	k119_20241_17	100%	ArsR	4	0	0	0	2.2	0
4	24	k119_20241	k119_20241_26	100%	ArsR	8	0	0	0	7.4	0
4	24	k119_20241	k119_20241_27	100%	ArsR	7	0	0	0	1.9	0
4	24	k119_20241	k119_20241_22	100%	ArsA	1	0	0	0	16.2	3672369.333
4	24	k119_20241	k119_20241_24	100%	Acr3	5	0	0	0	14.9	234906949.3
4	24	k119_15640	k119_15640_32	100%	ArsR	5	0	0	0	0	0
4	24	k119_15640	k119_15640_33	100%	ArsC	3	0	0	0	0	0
4	24	k119_15640	k119_15640_34	100%	Acr3	2	0	0	0	0	0
4	24	k119_1470	k119_1470_3	99%	ArsC	1	0	0	0	1.2	0
4	24	k119_14467	k119_14467_5	100%	ArsA	4	0	0	0	9.1	0
4	24	k119_14467	k119_14467_2	98%	Acr3	3	0	0	0	9.7	0
4	24	k119_14467	k119_14467_3	91%	Acr3	1	0	0	0.1	20.1	0
5	13	k119_6620	k119_6620_13	100%	ArsR	1	6.7	6.5	4.8	1.9	151359136
5	13	k119_5563	k119_5563_1	100%	ArsC	2	34.5	49.0	61.4	6.0	0
5	13	k119_5563	k119_5563_2	99%	ArsA	1	137.4	178.3	165.7	19.9	390.4
5	13	k119_4365	k119_4365_11	100%	ArsC	1	1.1	1.3	1.3	0.9	1.5
6	**0**	**k119_3176**	**k119_3176_11**	**85%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
6	0	k119_31663	k119_31663_43	no match	ArsR*	1	0	0	0	0	0
6	0	k119_31663	k119_31663_46	no match	ArsR*	2	0	0	0	0	0
6	0	k119_31663	k119_31663_48	no match	ArsR*	3	0	0	0	0	0
6	0	k119_31663	k119_31663_52	82%	ArsR	4	0	0	0	0	0
6	0	k119_31663	k119_31663_50	no match	ArsA*	1	0	0	0	0	0
6	0	k119_31663	k119_31663_44	90%	Acr3	1	0	0	0	0	0
6	0	k119_29035	k119_29035_253	no match	ArsR*	5	0	0	0	0	0
6	0	k119_29035	k119_29035_255	100%	ArsA	2	0	0	0	0.5	1.42739E+11
6	0	k119_29035	k119_29035_257	100%	Acr3	2	0	0	0	0	0
6	0	k119_25456	k119_25456_39	99%	ArsR	6	0	0	0	0	0
6	0	k119_16333	k119_16333_131	no match	ArsA*	3	0	0	0	0	0
6	0	k119_16333	k119_16333_130	100%	ArsA	4	0	0	0	0	0
7	**1**	**k119_6403**	**k119_6403_76**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
7	1	k119_30549	k119_30549_130	100%	ArsR	1	0	0	0	0	0
7	1	k119_3002	k119_3002_290	100%	ArsR	2	0	0	0	0	0
7	**1**	**k119_14048**	**k119_14048_169**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**0**	**0**
7	1	k119_14048	k119_14048_170	100%	ArsC	1	0	0	0	0	0

Table B.9 - part 2 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from EA culture.

							Transcriptome R			Transcriptome G	Proteome
MAG	Bin	Contig	Protein id	Match identity	Annotation	N.	rep1	rep2	rep3	AVG	AVG
7	1	k119_14048	k119_14048_171	100%	Acr3	1	0	0	0	0	0
8	14	k119_9963	k119_9963_273	100%	ArsR	1	0	0	0	0	0
8	14	k119_31112	k119_31112_112	100%	ArsR	2	0	0	0	0	0
8	**14**	**k119_31112**	**k119_31112_113**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**0**	**0**
8	14	k119_31112	k119_31112_114	100%	ArsC	1	0	0	0	0	0
8	14	k119_31112	k119_31112_115	100%	Acr3	1	0	0	0	0	0
8	14	k119_10316	k119_10316_160	100%	ArsR	3	0	0	0	0	0
8	**14**	**k119_10316**	**k119_10316_34**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0.3**
9	26	k119_6608	k119_6608_203	100%	ArsC	1	1.2	1.0	1.8	0.9	0.6
9	26	k119_6608	k119_6608_217	100%	ArsC	2	1.3	1.1	1.0	1.9	3.2
10	2	k119_8470	k119_8470_156	100%	ArsC	1	0	0	0	0.3	0
10	2	k119_36161	k119_36161_56	100%	ArsR	1	0	0	0	3.5	15722144.7
10	2	k119_36161	k119_36161_52	100%	ArsC	2	0	0	0	0.3	1350929639.0
10	2	k119_36161	k119_36161_54	100%	ArsA	1	0	0	0	19.4	503.9
10	2	k119_36161	k119_36161_53	100%	Acr3	1	0	0	0	0.7	0
11	8	k119_8741	k119_8741_12	100%	ArsR	1	0	0	0	0	0
11	8	k119_8741	k119_8741_7	100%	ArsC	1	0	0	0	0	0
11	8	k119_8741	k119_8741_8	no match	ArsA*	1	0	0	0	0	0
11	8	k119_8741	k119_8741_10	100%	Acr3	1	0	0	0	0	0
11	8	k119_5776	k119_5776_29	100%	ArsC	2	0	0	0	0	0
11	**8**	**k119_25138**	**k119_25138_7**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
11	8	k119_20026	k119_20026_10	no match	ArsC*	3	0	0	0	0	0
11	8	k119_20026	k119_20026_11	no match	Acr3*	2	0	0	0	0.1	971334741.3
11	8	k119_19663	k119_19663_28	100%	ArsR	2	0	0	0	0	0
11	8	k119_19663	k119_19663_24	100%	ArsC	4	0	0	0	0	0
11	8	k119_19663	k119_19663_25	100%	ArsA	2	0	0	0	0	0
11	8	k119_19663	k119_19663_27	100%	Acr3	3	0	0	0	0	0
11	8	k119_18016	k119_18016_14	100%	ArsR	3	0	0	0	0	0
11	8	k119_16849	k119_16849_72	100%	ArsC	5	0	0	0	0	0
11	8	k119_16849	k119_16849_74	100%	Acr3	4	0	0	0	0	0
12	23	k119_29568	k119_29568_3	100%	ArsC	2	0	0	0	74.5	4379976986.3
12	23	k119_29568	k119_29568_4	100%	Acr3	1	0	0	0	49.5	971334741.3
12	23	k119_19604	k119_19604_9	100%	ArsC	3	0	0	0	0.7	458990600
12	23	k119_13744	k119_13744_35	99%	ArsC	1	0	0	0	1.4	0
13	19	k119_3553	k119_3553_229	100%	ArsR	1	0	0	0.5	0	0
13	19	k119_30669	k119_30669_25	100%	ArsR	2	0.7	0.6	0.7	0	544214332
13	**19**	**k119_30669**	**k119_30669_28**	**100%**	**ArsM**	**1**	**2.4**	**2.2**	**0.8**	**0**	**144453609.3**
13	19	k119_30669	k119_30669_27	no match	Acr3*	1	1.6	2.1	4.0	0	234906949.3
13	19	k119_16281	k119_16281_97	100%	ArsC	1	1.0	1.0	0.7	0	0
14	4	k119_828	k119_828_10	100%	ArsR	2	0.0	0.0	0.0	0.1	0
14	4	k119_828	k119_828_4	100%	ArsR	1	0.0	0.0	0.0	0	0
14	4	k119_828	k119_828_8	100%	ArsA	1	0.0	0.0	0.0	0	5620412971.7
14	**4**	**k119_35396**	**k119_35396_11**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**0**	**0**
14	4	k119_1963	k119_1963_12	100%	ArsR	4	0	0	0	0	0
14	4	k119_1963	k119_1963_5	100%	ArsR	3	0	0	0	0	0
14	4	k119_1963	k119_1963_10	no match	ArsA*	2	0	0.1	0	0	7164242970.3
14	**4**	**k119_16473**	**k119_16473_64**	**no match**	**ArsM***	**1**	**0**	**0**	**0**	**0**	**0**
14	4	k119_1596	k119_1596_10	100%	ArsR	6	0	0	0	0	0
14	4	k119_1596	k119_1596_30	100%	ArsR	5	0	0	0	0	0
14	4	k119_1596	k119_1596_28	100%	ArsA	3	0	0	0	0	0
14	4	k119_1596	k119_1596_12	100%	Acr3	1	0	0	0	0	0
14	4	k119_15184	k119_15184_2	100%	ArsC	1	0	0	0	0	0
14	4	k119_1367	k119_1367_3	100%	ArsA	4	0	0	0	0	0
14	**4**	**k119_11341**	**k119_11341_1**	**no match**	**ArsM***	**3**	**0**	**0**	**0**	**0**	**0**

Table B.9 - part 3 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from EA culture.

							Transcriptome R			Transcriptome G	Proteome
MAG	**Bin**	**Contig**	**Protein id**	**Match identity**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**AVG**	**AVG**
14	4	k119_11115	k119_11115_59	100%	ArsC	2	0	0	0	0	0
15	15	k119_13704	k119_13704_197	100%	ArsR	1	0	0	0	0	0
15	15	k119_13704	k119_13704_200	100%	ArsC	1	0.3	0.3	0.7	0.2	0
15	15	k119_13704	k119_13704_199	100%	Acr3	1	0.5	0.3	0.5	0	0
16	20	k119_31022	k119_31022_78	100%	ArsR	1	78.2	49.2	88.1	16.6	100786538.7
16	20	k119_31022	k119_31022_80	100%	ArsA	1	86.0	190.1	135.2	119.2	1115.6
16	20	k119_28101	k119_28101_15	100%	ArsC	1	1.3	14.1	4.0	0	0
16	20	k119_1223	k119_1223_28	100%	ArsC	2	0.8	2.6	1.0	0	0
	NB	k119_9933	k119_9933_27	100%	ArsC	4	0	0	0	1.1	0
	NB	k119_9898	k119_9898_5	81%	ArsC	20	0	0	0	0	0
	NB	k119_3840	k119_3840_2	no match	ArsC*	6	0	0	0	0	133855805.3
	NB	k119_3840	k119_3840_1	no match	Acr3*	4	0	0	0	0.6	0
	NB	k119_3548	k119_3548_3	no match	ArsC*	2	0	0	0	0	0
	NB	k119_34958	k119_34958_15	no match	ArsR*	1	0	0	0	0	0
	NB	k119_34958	k119_34958_18	100%	ArsC	16	0	0	0	0.4	0
	NB	k119_34671	k119_34671_35	100%	ArsR	6	0	0	0	4.4	0
	NB	k119_34671	k119_34671_39	99%	ArsC	17	0	0	0	7.1	133855805.3
	NB	k119_33744	k119_33744_1	no match	ArsC*	13	763.4	926.3	1090.5	37.9	0
	NB	k119_32701	k119_32701_5	no match	ArsC*	7	0	0	0	0	0
	NB	k119_30826	k119_30826_1	93%	ArsC	5	0	0	0	2.3	0
	NB	k119_30063	k119_30063_5	no match	ArsA*	1	0	0	0	0	0
	NB	k119_30063	k119_30063_3	100%	ArsA	5	0	0	0	0	0
	NB	k119_30061	k119_30061_2	no match	ArsC*	3	0	0	0	0	0
	NB	k119_28281	k119_28281_4	100%	ArsR	2	0	0	0	5.9	0
	NB	k119_28281	k119_28281_8	100%	ArsC	18	0	0	0	0.9	605712277.3
	NB	k119_28281	k119_28281_7	no match	ArsA*	2	0	0	0	0.9	3058809747
	NB	k119_28281	k119_28281_5	100%	Acr3	11	0	0	0	0.9	162178346.7
	NB	k119_22168	k119_22168_2	no match	ArsR*	4	0	0	0	0	0
	NB	k119_21124	k119_21124_1	no match	ArsR*	3	287.4	301.0	212.4	5.7	0
	NB	**k119_2076**	**k119_2076_2**	**99%**	**ArsM**	**2**	**0**	**0**	**0**	**0**	**0**
	NB	k119_2055	k119_2055_3	100%	ArsC	14	1747.9	805.6	600.2	148.2	0
	NB	k119_2055	k119_2055_2	no match	Acr3*	6	745.0	880.1	924.8	48.6	0
	NB	k119_20281	k119_20281_1	100%	ArsR	5	0	0	0	0	0
	NB	**k119_20281**	**k119_20281_2**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
	NB	k119_20281	k119_20281_3	100%	ArsC	15	0	0	0	0	0
	NB	k119_20281	k119_20281_4	no match	Acr3*	1	0	0	0	0	0
	NB	k119_18877	k119_18877_1	no match	Acr3*	8	0	0	0	0.6	0
	NB	k119_18459	k119_18459_2	no match	ArsC*	12	0	0	0	0	0
	NB	k119_15186	k119_15186_2	no match	ArsC*	11	202.9	215.8	237.8	12.4	0
	NB	k119_14937	k119_14937_3	100%	ArsA	4	0	0	0	0	0
	NB	k119_14937	k119_14937_1	no match	Acr3*	7	0	0	0	0	0
	NB	k119_13766	k119_13766_3	100%	Acr3	9	0	1.3	0	1.8	0
	NB	k119_13677	k119_13677_3	100%	ArsC	19	0	0	0	2.2	0
	NB	k119_12459	k119_12459_7	100%	ArsR	7	0	0	0	0.9	0
	NB	k119_12459	k119_12459_2	no match	ArsC*	8	0	0	0	0	768215805.6
	NB	k119_12459	k119_12459_3	100%	ArsA	6	0	0	0	1.6	3158878099.0
	NB	k119_12459	k119_12459_6	no match	Acr3*	5	0	0	0	0.5	162178346.7
	NB	k119_12047	k119_12047_6	no match	Acr3*	3	0	0	0	0	0
	NB	k119_12047	k119_12047_4	98%	Acr3	10	0	0	0	0	0
	NB	k119_12047	k119_12047_5	96%	Acr3	2	0	0	0	0	0
	NB	k119_11234	k119_11234_15	100%	ArsC	9	0	0	0	0	0
	NB	k119_11234	k119_11234_14	100%	Acr3	12	0	0	0	0	0
	NB	k119_10507	k119_10507_2	no match	ArsC*	10	147.6	160.1	183.4	3.5	0
	NB	k119_10507	k119_10507_1	100%	ArsA	3	20.1	26.4	23.1	6.6	0
	NB	k119_10148	k119_10148_27	no match	ArsC*	1	0	0	0	0	0

Table B.10 – part 1 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from TSB culture. The values correspond to gene/protein abundance in the +As condition divided by the gene/protein abundance in the no−As control (values in green) or, when no expression was reported in the no−As condition for that gene/protein, the abundance values from the +As condition (values in black) in transcripts per million (TPM−RNA) for transcriptomes and in protein abundance for proteome. Data shown for the three biological replicates for transcriptome R and average from the three biological replicates fro transcriptome G and proteome. Contig: contig number assinged by Megahit. Protein id: id number for the protein−encoding genes found by Prodigal. Match identity: percentage of identity between clustered proteins from +As condition vs. no−As control by CD-HIT-2D. Annotation: protein annotation according to GhostKOALA. An asterisk (*) in protein annotation means no matching protein was clustered by CD-HIT-2D from the no−As control. N. column: corresponds to the gene number in Figure 4. NB in Bin column means the gene was not clustered in any bin. ArsM proteins are in bold.

							Transcriptome R			Transcriptome G	Proteome
MAG	Bin	Contig	Protein id	Match identity	Annotation	N.	rep1	rep2	rep3	AVG	AVG
1	9	k119_7837	k119_7837_1	100%	Acr3	1	0.139	0.154	0.503	6.0	0
2	6	k119_12528	k119_12528_12	100%	ArsR	1	5.3	7.8	11.4	14.5	246021986.0
2	6	k119_12528	k119_12528_11	no match	ArsH*	1	219.5	295.1	341.4	254.4	26399449203.0
2	6	k119_13639	k119_13639_3	100%	ArsC	1	0.5	0.4	0.6	1.1	0
2	6	k119_12528	k119_12528_13	no match	ArsB*	1	60.81	78.37	89.92	66.0	0
3	21	k119_11309	k119_11309_2	no match	ArsC*	1	0	0	0	0	0
3	21	k119_11309	k119_11309_1	100%	Acr3	1	0	0	0	0	0
3	21	k119_3422	k119_3422_1	100%	Acr3	2	0	0	0	0	0
4	28	k119_3434	k119_3434_358	100%	ArsR	1	0	0	0	0	0
4	28	k119_3434	k119_3434_49	100%	ArsC	1	0	0	0	0	0
4	28	k119_3434	k119_3434_362	100%	ArsC	2	0	0	0	0.1	2.5
4	28	k119_19999	k119_19999_324	100%	Acr3	1	0	0	0	0	0
5	5	k119_9868	k119_9868_36	100%	ArsC	1	0	0	0	0	2.5
5	5	k119_9868	k119_9868_38	84%	Acr3	1	0	0	0	0	0
6	**17**	**k119_11069**	**k119_11069_212**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
6	17	k119_29213	k119_29213_24	100%	ArsC	1	0	0	0	0	0
7	20	k119_14056	k119_14056_9	97%	ArsR	1	0.02	0.01	0.00	0	0
7	20	k119_15390	k119_15390_5	83%	ArsC	2	315.5	245.6	275.4	0	21.0
7	20	k119_601	k119_601_127	100%	ArsC	1	1.2	0.7	0.4	0	0
7	20	k119_15390	k119_15390_4	100%	ArsA	1	74.4	49.6	49.8	0	140.1
7	20	k119_15390	k119_15390_6	100%	Acr3	1	141.0	126.7	31.1	0	1034821738.7
7	20	k119_7477	k119_7477_3	100%	Acr3	2	6.4	4.4	8.4	0	0
8	35	k119_19293	k119_19293_1	100%	ArsR	1	0	0	0	0	0
8	35	k119_27112	k119_27112_365	100%	ArsR	3	0	0	0	0	1
8	35	k119_27112	k119_27112_362	100%	ArsR	2	0	0	0	0	0
8	35	k119_8737	k119_8737_69	100%	ArsR	4	0	0	0	0	0
8	**35**	**k119_19293**	**k119_19293_5**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**2518946.5**
8	**35**	**k119_30244**	**k119_30244_221**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**0**	**0**
8	**35**	**k119_8737**	**k119_8737_68**	**100%**	**ArsM**	**3**	**0**	**0**	**0**	**0**	**222768773.8**
8	35	k119_19293	k119_19293_2	100%	ArsC	1	0	0	0	0	2.5
8	35	k119_27112	k119_27112_363	86%	ArsC	2	0	0	0	0	2.5
8	35	k119_19293	k119_19293_3	89%	Acr3	1	0	0	0	0	0
8	35	k119_27112	k119_27112_364	100%	Acr3	2	0	0	0	0	0
9	3	k119_17050	k119_17050_30	100%	ArsR	1	0	0	0	0	0
9	3	k119_30803	k119_30803_17	100%	ArsR	2	0	0	0	0	0
9	3	k119_4945	k119_4945_26	100%	ArsR	3	0	0	0	0	0
9	**3**	**k119_17050**	**k119_17050_29**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
9	3	k119_17050	k119_17050_28	83%	ArsC	1	0	0	0	0	6158750.7
9	3	k119_17050	k119_17050_27	97%	Acr3	1	0	0	0	0.03	0
10	25	k119_15797	k119_15818_38	100%	ArsR	1	0	0	0	0	0
10	25	k119_30629	k119_30629_18	100%	ArsR	2	0	0	0	0	0
10	25	k119_33283	k119_33283_62	100%	ArsR	3	0	0	0	0.1	0
10	**25**	**k119_15057**	**k119_15057_34**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
10	**25**	**k119_15818**	**k119_15818_37**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**0.4**	**0**
10	25	k119_15818	k119_15818_36	100%	ArsC	1	0	0	0	0	6158750.7
10	25	k119_15818	k119_15818_35	95%	Acr3	1	0	0	0	0.1	0
11	26	k119_4666	k119_4666_37	100%	ArsR	1	0	0	0	3.0	0
11	26	k119_4666	k119_4666_39	100%	ArsC	1	0	0	0	11.9	1514992017.7
11	26	k119_5491	k119_5491_11	100%	ArsC	2	0	0	0	1.2	0
11	26	k119_4666	k119_4666_38	no match	ArsB*	1	0	0	0	2.5	38215765.7
12	15	k119_30292	k119_30292_56	100%	ArsR	1	0	0	0	0	0
12	15	k119_20383	k119_20383_128	100%	ArsC	1	0	0	0	0	0
12	15	k119_30292	k119_30292_52	100%	ArsC	2	0	0	0	0	1424675691.1
12	15	k119_30292	k119_30292_54	100%	ArsA	1	0	0	0	0	101.4

Table B.10 - part 2 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from TSB culture.

							Transcriptome R			Transcriptome G	Proteome
MAG	Bin	Contig	Protein id	Match identity	Annotation	N.	rep1	rep2	rep3	AVG	AVG
12	15	k119_30292	k119_30292_53	100%	Acr3	1	0	0	0	0	57221944
13	27	k119_32990	k119_32990_69	100%	ArsC	1	2.9	1.4	1.5	0.7	0
14	0	k119_1034	k119_1034_7	100%	ArsR	1	0	0	0	0	0
14	0	k119_12891	k119_12891_5	100%	ArsR	2	0	0	0	0	0
14	**0**	**k119_26245**	**k119_26245_5**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
14	0	k119_27340	k119_27340_1	no match	ArsC*	3	0	0	0	0	0
14	0	k119_23682	k119_23682_21	100%	ArsC	2	0	0	0	0	116557353.6
14	0	k119_30081	k119_30081_12	100%	ArsC	1	0	0	0	0	0
14	0	k119_1034	k119_1034_9	100%	Acr3	1	0	0	0	0	0
14	0	k119_12891	k119_12891_6	99%	Acr3	3	0	0	0	0	0
14	0	k119_12891	k119_12891_3	100%	Acr3	2	0	0	0	0	0
14	0	k119_23682	k119_23682_23	84%	Acr3	4	0	0	0	0	0
15	30	k119_2647	k119_2647_6	100%	ArsC	1	0	0	0	0	2.5
15	30	k119_9504	k119_9504_27	100%	ArsC	2	0	0	0	0	10.6
15	30	k119_9504	k119_9504_26	81%	Acr3	1	0	0	0	0	829964696.0
16	36	k119_3430	k119_3430_28	100%	ArsR	2	0	0	0	0	0
16	36	k119_3430	k119_3430_22	100%	ArsR	1	0	0	0	0	0
16	**36**	**k119_3430**	**k119_3430_32**	**100%**	**ArsM**	**1**	**0**	**0**	**0**	**0**	**0**
16	36	k119_10810	k119_10810_102	100%	ArsC	1	0	0	0	0	0
16	36	k119_10810	k119_10810_115	100%	ArsA	1	0	0	0	0	0
16	36	k119_3430	k119_3430_26	100%	ArsA	2	0	0	0	0	3634112731
17	14	k119_33769	k119_33769_88	100%	ArsR	1	3.907	4.328	2.441	2.0	0
17	14	k119_33769	k119_33769_85	100%	ArsC	1	1.593	6.7	4.7	3.8	0
17	14	k119_33769	k119_33769_86	100%	Acr3	1	0.829	1.033	0.935	2.6	9287051.3
18	29	k119_15797	k119_15797_19	100%	ArsR	1	25.63	16.23	27.16	81.8	5142968.7
18	29	k119_4071	k119_4071_3	100%	ArsR	2	17.33	11.67	11.96	134.8	325375616.3
18	29	k119_4205	k119_4205_129	100%	ArsR	3	7.818	7.154	9.179	342.5	5912919.3
18	29	k119_4205	k119_4205_128	no match	ArsH*	1	4.839	8.818	11.55	244.0	1402367323.7
18	29	k119_1079	k119_1079_16	100%	ArsC	1	0	0.671	0.547	0.8	0
18	29	k119_15797	k119_15797_15	82%	ArsC	2	64.8	53.62	65.86	664.5	9420302646
18	29	k119_4071	k119_4071_5	80%	ArsC	3	25.1	17.31	15.94	129.2	2595981184.9
18	29	k119_4205	k119_4205_131	100%	ArsC	4	24.6	9	25.3	64.9	720818943.1
18	29	k119_4205	k119_4205_130	no match	ArsB*	1	2.11	3.645	3.121	65.1	38215765.67
18	29	k119_15797	k119_15797_17	100%	ArsA	2	184.2	114.5	90.7	277.7	11.5
18	29	k119_15797	k119_15797_13	100%	ArsA	1	1.839	11.1	25.3	25.0	0
19	32	k119_19214	k119_19214_17	100%	ArsR	1	0	0	0	0	0
19	32	k119_19214	k119_19214_14	100%	ArsA	1	0.123	0	0	1.1	1473047760.3
19	32	k119_19214	k119_19214_16	100%	Acr3	1	0	0	0	0.4	0
20	4	k119_12019	k119_12019_77	100%	ArsR	1	154.5	247.3	348.4	3.6	246519001.7
20	4	k119_11336	k119_11336_29	100%	ArsC	1	1.068	0.344	1.402	0	0
20	4	k119_9746	k119_9746_30	100%	ArsC	2	0.3	3.1	0.5	0	0
20	4	k119_12019	k119_12019_79	100%	ArsA	1	269.4	439.4	250.5	29.2	4172943067.2
	NB	k119_24571	k119_24571_3	no match	ArsR*	3	0	0	0	0	0
	NB	k119_31332	k119_31332_2	no match	ArsR*	4	0	0	0	0	0
	NB	k119_18612	k119_18612_2	100%	ArsR	1	0	0	0	0	0
	NB	k119_5901	k119_5901_1	99%	ArsR	2	32.43	16.75	15.52	0	0
	NB	**k119_10596**	**k119_10596_1**	**100%**	**ArsM**	**1**	**0**	**0.831**	**0.338**	**0.5**	**0**
	NB	**k119_29944**	**k119_29944_1**	**100%**	**ArsM**	**2**	**0**	**0**	**0**	**1.0**	**0**
	NB	k119_15797	k119_15797_12	no match	ArsH*	1	3.82	4.57	6.066	37.0	1384059107
	NB	k119_4071	k119_4071_2	no match	ArsH*	2	4.683	9.337	9.577	53.2	6739884920
	NB	k119_18176	k119_18176_2	no match	ArsC*	12	0	0	0	0	0
	NB	k119_11372	k119_11372_2	100%	ArsC	1	0	0	0	0	0
	NB	k119_14073	k119_14073_1	99%	ArsC	3	0	0	0	0	0
	NB	k119_14299	k119_14299_2	100%	ArsC	4	0.236	0	0	2.2	0
	NB	k119_17909	k119_17909_2	82%	ArsC	7	0	0	0	0	0

Table B.10 - part 3 of 3. Abundance and relative expression (vs. no As control) values of *ars* genes from TSB culture.

							Transcriptome R			Transcriptome G	Proteome
MAG	Bin	Contig	Protein id	Match identity	Annotation	N.	rep1	rep2	rep3	AVG	AVG
	NB	k119_19823	k119_19823_4	100%	ArsC	5	0	0	0	0	0
	NB	k119_21777	k119_21777_1	100%	ArsC	8	0	0	0	0	0
	NB	k119_27192	k119_27192_1	100%	ArsC	9	0	0	0	0	0
	NB	k119_30429	k119_30429_2	100%	ArsC	2	0	0	0	0	0
	NB	k119_4811	k119_4811_2	100%	ArsC	10	0	0.133	0	0.8	666546261.4
	NB	k119_7031	k119_7031_5	84%	ArsC	6	0	0	0	0	1.7
	NB	k119_8477	k119_8477_2	100%	ArsC	11	0.12	0.133	0.108	62.4	476591579.5
	NB	k119_15797	k119_15797_16	no match	ArsB*	8	9.05	12.37	16.1	144.8	192784845.8
	NB	k119_4071	k119_4071_4	no match	ArsB*	9	1.692	4.497	4.197	24.2	306547766.5
	NB	k119_22971	k119_22971_2	no match	ArsA*	12	0	0	0	0	699720426.0
	NB	k119_24571	k119_24571_6	no match	ArsA*	11	0	0	0	0	699720426.0
	NB	k119_12429	k119_12429_1	100%	ArsA	2	0	0	0	0	0
	NB	k119_13181	k119_13181_1	93%	ArsA	3	0	0	0	0	0
	NB	k119_15852	k119_15852_1	98%	ArsA	4	0	0	0	0	960489234.0
	NB	k119_19823	k119_19823_2	100%	ArsA	1	0	0	0	0	2192396006.5
	NB	k119_20090	k119_20090_1	88%	ArsA	5	0	0	0	0	0
	NB	k119_21122	k119_21122_1	96%	ArsA	6	0	0	0	0.2	0
	NB	k119_26176	k119_26176_1	92%	ArsA	7	0	0	0	0	0
	NB	k119_29234	k119_29234_1	88%	ArsA	8	0	0	0	0	1202822396.5
	NB	k119_6791	k119_6791_3	87%	ArsA	9	0	0	0	0	699720426.0
	NB	k119_9020	k119_9020_1	99%	ArsA	10	0	0	0	0	1445926656.0
	NB	k119_14845	k119_14845_1	no match	Acr3*	7	0	0	0	0	0
	NB	k119_13996	k119_13996_2	100%	Acr3	4	0	0	0	0	0
	NB	k119_14299	k119_14299_1	100%	Acr3	1	0	0	0	1.5	0
	NB	k119_19823	k119_19823_3	100%	Acr3	2	0	0	0	0	0
	NB	k119_23125	k119_23125_1	86%	Acr3	5	0	0.141	0.229	0.1	0
	NB	k119_24571	k119_24571_4	100%	Acr3	3	0	0	0	0.04	0
	NB	k119_8477	k119_8477_1	100%	Acr3	6	0	0	0	8.5	0

Table B.11 – part 1 of 4. Abundance values of *ars* genes from EA culture, +As condition. Values are in transcripts per million for genome (TPM−DNA) and transcriptomes (TPM−RNA) and protein abundance for proteome. Contig: contig number assinged by Megahit. Protein id: id number for the protein−encoding genes found by Prodigal. Annotation: protein annotation according to GhostKOALA. An asterisk (*) in protein annotation means no matching protein was clustered by CD-HIT-2D from the no−As control. N. column: corresponds to the gene number in Figure 4. NB in Bin column means the gene was not clustered in any bin. ArsM proteins are in bold. Data shown for the three biological replicates.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
1	7	k119_11963	k119_11963_422	ArsC	1	214.4	225.1	242.3	0	0	0	0	0.1	0.1	36728236.0	28776740.0	66827344.0
1	7	k119_11963	k119_11963_423	Acr3	1	214.4	225.1	242.3	0	0	0	0.2	0	0	0	0	0
2	3	k119_7198	k119_7198_17	ArsR	1	732.7	741.7	754.7	0	0	0	0	0	0	0	0	0
2	3	k119_7198	k119_7198_14	ArsA	1	732.7	741.7	754.7	0.1	0.1	0	0	0	0	3013451592.0	2067061379.0	3353259962.0
2	3	k119_7198	k119_7198_16	Acr3*	1	732.7	741.7	754.7	0	0	0	0	0	0	0	0	0
3	28	k119_6934	k119_6934_5	ArsR	1	18.7	20.4	14.8	0	0	0	0.3	0.3	0.3	0	0	0
3	28	k119_6934	k119_6934_2	ArsC	2	18.7	20.4	14.8	0	0	0	0.4	0.5	0.4	0	0	0
3	28	k119_6748	k119_6748_20	ArsC	1	16.7	18.3	13.3	0	0	0	0	0	0	0	0	0
4	24	k119_8323	k119_8323_82	ArsA	2	22.4	23.9	19.8	0	0	0	0	0	0	0	0	0
4	24	k119_8323	k119_8323_84	ArsA	3	22.4	23.9	19.8	0	0	0	0	0	0	0	0	0
4	24	k119_6159	k119_6159_77	ArsR	2	24.6	25.3	21.3	0	0	0	0	0.2	0.2	153631560.0	218310176.0	328591584.0
4	24	k119_6159	k119_6159_80	ArsR	3	24.6	25.3	21.3	0	0	0	0.2	0.8	0	0	0	0
4	**24**	**k119_6159**	**k119_6159_75**	**ArsM**	**1**	**24.6**	**25.3**	**21.3**	**0**	**0**	**0**	**0.3**	**0.1**	**0.1**	**117143376.0**	**113314260.0**	**138798550.0**
4	24	k119_6159	k119_6159_79	ArsC	2	24.6	25.3	21.3	0	0	0	0.2	0.2	0.2	1128726691.0	663299606.0	1239834734.0
4	24	k119_6159	k119_6159_78	Acr3	4	24.6	25.3	21.3	0.2	0	0.1	0.7	1.0	0.2	20520962.0	23820546.0	28738626.0
4	24	k119_32423	k119_32423_101	ArsA	6	26.4	24.1	24.6	0	0	0	0	0	0	0	0	0
4	24	k119_32423	k119_32423_99	ArsA	5	26.4	24.1	24.6	0	0	0	0	0	0	0	0	0
4	24	k119_253	k119_253_29	ArsR	6	18.6	15.7	11.9	0	0	0	23.3	14.7	17.8	0	0	0
4	24	k119_253	k119_253_30	ArsR	1	18.6	15.7	11.9	0	0	0	16.4	14.4	17.9	0	0	0
4	**24**	**k119_253**	**k119_253_32**	**ArsM**	**2**	**18.6**	**15.7**	**11.9**	**0**	**0**	**0**	**16.3**	**16.5**	**14.3**	**0**	**0**	**0**
4	24	k119_20241	k119_20241_17	ArsR	4	24.1	19.9	17.4	0	0	0	65.9	53.8	48.9	0	0	0
4	24	k119_20241	k119_20241_26	ArsR	8	24.1	19.9	17.4	0	0	0	123.8	114.6	114.2	0	0	0
4	24	k119_20241	k119_20241_27	ArsR	7	24.1	19.9	17.4	0	0	0	14.0	13.3	11.8	0	0	0
4	24	k119_20241	k119_20241_22	ArsA	1	24.1	19.9	17.4	0	0	0	394.7	356.2	333.1	11017108.0	0	0
4	24	k119_20241	k119_20241_24	Acr3	5	24.1	19.9	17.4	0	0	0	103.3	91.3	89.8	143020016.0	289034560.0	272666272.0
4	24	k119_15640	k119_15640_32	ArsR	5	22.2	21.3	20.3	0	0	0	0	0	0	0	0	0
4	24	k119_15640	k119_15640_33	ArsC	3	22.2	21.3	20.3	0	0	0	0	0	0	0	0	0
4	24	k119_15640	k119_15640_34	Acr3	2	22.2	21.3	20.3	0	0	0	0	0.1	0	0	0	0
4	24	k119_1470	k119_1470_3	ArsC	1	16.2	16.9	11.4	0	0	0	46.9	41.0	40.9	0	0	0
4	24	k119_14467	k119_14467_5	ArsA	4	18.3	14.5	10.7	0	0	0	48.5	42.3	41.4	11017108.0	0	0
4	24	k119_14467	k119_14467_2	Acr3	3	18.3	14.5	10.7	0	0	0	30.0	27.8	28.3	0	0	0
4	24	k119_14467	k119_14467_3	Acr3	1	18.3	14.5	10.7	0	0	0.1	20.5	21.5	18.5	0	0	0
5	13	k119_6620	k119_6620_13	ArsR	1	186.0	200.1	202.1	713.0	766.3	556.7	9.6	12.0	13.9	222933952.0	0	231143456.0
5	13	k119_5563	k119_5563_1	ArsC	2	175.9	172.9	183.2	457.4	484.9	599.6	19.2	15.6	15.1	0	0	0
5	13	k119_5563	k119_5563_2	ArsA	1	175.9	172.9	183.2	1462.6	1724.7	1841.8	50.3	57.9	54.3	50579317370.5	41342747371.1	60417638994.9
5	13	k119_4365	k119_4365_11	ArsC	1	207.7	229.3	232.6	39.4	45.7	57.5	4.4	4.7	5.6	51626812.0	31116488.0	51923644.0

Table B.11 - part 2 of 4. Abundance values of *ars* genes from EA culture, +As condition.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
6	0	k119_31663	k119_31663_43	ArsR*	1	6.5	7.0	3.6	0	0	0	0	0	0	0	0	0
6	0	k119_31663	k119_31663_46	ArsR*	2	6.5	7.0	3.6	0	0	0	0	0	0	0	0	0
6	0	k119_31663	k119_31663_48	ArsR*	3	6.5	7.0	3.6	0	0	0	0	0	0	0	0	0
6	0	k119_31663	k119_31663_52	ArsR	4	6.5	7.0	3.6	0	0	0	0	0	0	0	0	0
6	0	k119_31663	k119_31663_50	ArsA*	1	6.5	7.0	3.6	0	0	0	0	0	0	0	0	0
6	0	k119_31663	k119_31663_44	Acr3	1	6.5	7.0	3.6	0	0	0	0.1	0	0	0	0	0
6	0	k119_29035	k119_29035_253	ArsR*	5	9.9	8.3	6.0	0	0	0	0	0	0	0	0	0
6	0	k119_29035	k119_29035_255	ArsA	2	9.9	8.3	6.0	0	0	0	0	0	0.0	4433547220.0	4897816519.0	5967478096.9
6	0	k119_29035	k119_29035_257	Acr3	2	9.9	8.3	6.0	0	0	0	0	0	0	0	0	0
6	0	k119_25456	k119_25456_39	ArsR	6	8.6	6.6	4.8	0	0	0	0	0	0	0	0	0
6	0	k119_16333	k119_16333_131	ArsA*	3	12.8	10.7	7.6	0	0	0	0	0	0	0	0	0
6	0	k119_16333	k119_16333_130	ArsA	4	12.8	10.7	7.6	0	0	0	0	0	0	0	0	0
7	**1**	**k119_6403**	**k119_6403_76**	**ArsM**	**1**	**342.5**	**310.3**	**348.7**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
7	1	k119_30549	k119_30549_130	ArsR	1	360.7	328.6	367.9	0	0	0	0	0	0	0	0	0
7	1	k119_3002	k119_3002_290	ArsR	2	424.3	384.1	429.2	0	0	0	0	0	0	0	0	0
7	**1**	**k119_14048**	**k119_14048_169**	**ArsM**	**2**	**448.5**	**401.1**	**448.5**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
7	1	k119_14048	k119_14048_170	ArsC	1	448.5	401.1	448.5	0	0	0	0	0	0	0	0	0
7	1	k119_14048	k119_14048_171	Acr3	1	448.5	401.1	448.5	0	0	0	0	0	0	0	0	0
8	14	k119_9963	k119_9963_273	ArsR	1	20.4	18.4	22.6	0	0	0	0	0	0	0	0	0
8	14	k119_31112	k119_31112_112	ArsR	2	21.2	19.2	23.8	0	0	0	0	0	0	0	0	0
8	**14**	**k119_31112**	**k119_31112_113**	**ArsM**	**2**	**21.2**	**19.2**	**23.8**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
8	14	k119_31112	k119_31112_114	ArsC	1	21.2	19.2	23.8	0	0	0	0	0	0	0	0	0
8	14	k119_31112	k119_31112_115	Acr3	1	21.2	19.2	23.8	0	0	0	0	0	0	0	0	0
8	14	k119_10316	k119_10316_160	ArsR	3	17.1	15.7	19.2	0	0	0	0	0	0	0	0	0
8	**14**	**k119_10316**	**k119_10316_34**	**ArsM**	**1**	**17.1**	**15.7**	**19.2**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
9	26	k119_6608	k119_6608_203	ArsC	1	71.4	116.7	104.3	36.6	32.1	49.9	51.3	51.9	47.4	171302452.0	189874348.0	74023072.0
9	26	k119_6608	k119_6608_217	ArsC	2	71.4	116.7	104.3	7.7	9.0	7.6	7.8	10.7	9.1	431127944.0	264660688.0	471082832.0
10	2	k119_8470	k119_8470_156	ArsC	1	30.5	28.1	32.3	0	0	0	0.2	0.0	0.2	0	0	0
10	2	k119_36161	k119_36161_56	ArsR	1	30.1	28.2	32.9	0	0	0	0.2	0.3	0.2	27815848.0	5132310.0	14218276.0
10	2	k119_36161	k119_36161_52	ArsC	2	30.1	28.2	32.9	0	0	0	0.4	0	0.6	1806791984.0	902413487.0	1343583446.0
10	2	k119_36161	k119_36161_54	ArsA	1	30.1	28.2	32.9	0	0	0	1.5	1.0	2.0	12660321962.0	5920731963.4	7562266479.6
10	2	k119_36161	k119_36161_53	Acr3	1	30.1	28.2	32.9	0	0	0	0.5	0.7	0.9	0	0	0
11	8	k119_8741	k119_8741_12	ArsR	1	15.0	18.4	13.8	0	0	0	0	0	0	0	0	0
11	8	k119_8741	k119_8741_7	ArsC	1	15.0	18.4	13.8	0	0	0	0	0	0	0	0	0
11	8	k119_8741	k119_8741_8	ArsA*	1	15.0	18.4	13.8	0	0	0	0	0	0	0	0	0
11	8	k119_8741	k119_8741_10	Acr3	1	15.0	18.4	13.8	0	0	0	0	0	0	0	0	0
11	8	k119_5776	k119_5776_29	ArsC	2	12.7	10.9	9.7	0	0	0	0	0	0	0	0	0
11	**8**	**k119_25138**	**k119_25138_7**	**ArsM**	**1**	**11.2**	**9.5**	**11.2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
11	8	k119_20026	k119_20026_10	ArsC*	3	8.7	6.4	6.4	0	0	0	0	0	0	0	0	0
11	8	k119_20026	k119_20026_11	Acr3*	2	8.7	6.4	6.4	0	0	0	0.1	0	0.1	684853824.0	841604288.0	1387546112.0
11	8	k119_19663	k119_19663_28	ArsR	2	10.5	8.5	9.6	0	0	0	0	0	0	0	0	0
11	8	k119_19663	k119_19663_24	ArsC	4	10.5	8.5	9.6	0	0	0	0	0	0	0	0	0
11	8	k119_19663	k119_19663_25	ArsA	2	10.5	8.5	9.6	0	0	0	0	0	0	0	0	0

Table B.11 - part 3 of 4. Abundance values of *ars* genes from EA culture, +As condition.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
11	8	k119_19663	k119_19663_27	Acr3	3	10.5	8.5	9.6	0	0	0	0	0	0	0	0	0
11	8	k119_18016	k119_18016_14	ArsR	3	17.8	21.4	16.1	0	0	0	0	0	0	0	0	0
11	8	k119_16849	k119_16849_72	ArsC	5	12.1	10.1	12.3	0	0	0	0	0	0	0	0	0
11	8	k119_16849	k119_16849_74	Acr3	4	12.1	10.1	12.3	0	0	0	0	0	0	0	0	0
12	23	k119_29568	k119_29568_3	ArsC	2	454.4	394.6	462.6	0	0	0	8.1	11.3	14.2	4211806182.0	2277218968.0	6650905809.0
12	23	k119_29568	k119_29568_4	Acr3	1	454.4	394.6	462.6	0	0	0	13.6	14.0	16.7	684853824.0	841604288.0	1387546112.0
12	23	k119_19604	k119_19604_9	ArsC	3	467.2	409.1	471.6	0	0	0	0.9	1.0	0.1	523701964.0	453957420.0	399312416.0
12	23	k119_13744	k119_13744_35	ArsC	1	467.0	413.1	471.3	0	0	0	0.2	0	0.1	0	0	0
13	19	k119_3553	k119_3553_229	ArsR	1	49.2	43.1	42.5	0	0	0.3	0	0	0	0	0	0
13	19	k119_30669	k119_30669_25	ArsR	2	64.3	51.4	50.6	0.7	0.6	0.7	0	0	0	701002372.0	549939840.0	381700784.0
13	**19**	**k119_30669**	**k119_30669_28**	**ArsM**	**1**	**64.3**	**51.4**	**50.6**	**2.4**	**2.2**	**0.8**	**0**	**0**	**0**	**122617608.0**	**135220438.0**	**175522782.0**
13	19	k119_30669	k119_30669_27	Acr3*	1	64.3	51.4	50.6	1.6	2.1	4.0	0	0	0	143020016.0	289034560.0	272666272.0
13	19	k119_16281	k119_16281_97	ArsC	1	49.5	43.3	42.8	3.9	2.8	3.2	0	0	0	0	0	0
14	4	k119_828	k119_828_10	ArsR	2	23.3	25.7	22.9	0	0	0	0	0	0.2	0	0	0
14	4	k119_828	k119_828_4	ArsR	1	23.3	25.7	22.9	0	0	0	0	0	0	0	0	0
14	4	k119_828	k119_828_8	ArsA	1	23.3	25.7	22.9	0	0	0	0	0.1	0	6094545454.0	3932648553.0	6834044908.0
14	**4**	**k119_35396**	**k119_35396_11**	**ArsM**	**2**	**40.0**	**40.7**	**39.9**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	4	k119_1963	k119_1963_12	ArsR	4	41.9	44.0	41.3	0	0	0	0	0	0	0	0	0
14	4	k119_1963	k119_1963_5	ArsR	3	41.9	44.0	41.3	0	0	0	0	0	0	0	0	0
14	4	k119_1963	k119_1963_10	ArsA*	2	41.9	44.0	41.3	0	0.1	0	0	0	0	7425039486.0	4858285821.0	9209403604.0
14	**4**	**k119_16473**	**k119_16473_64**	**ArsM***	**1**	**24.3**	**29.1**	**23.4**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	4	k119_1596	k119_1596_10	ArsR	6	24.3	26.9	21.4	0	0	0	0	0	0	0	0	0
14	4	k119_1596	k119_1596_30	ArsR	5	24.3	26.9	21.4	0	0	0	0	0	0	0	0	0
14	4	k119_1596	k119_1596_28	ArsA	3	24.3	26.9	21.4	0	0	0	0	0	0	0	0	0
14	4	k119_1596	k119_1596_12	Acr3	1	24.3	26.9	21.4	0	0	0	0	0	0	0	0	0
14	4	k119_15184	k119_15184_2	ArsC	1	28.1	32.7	26.1	0	0	0	0	0	0	0	0	0
14	4	k119_1367	k119_1367_3	ArsA	4	29.0	29.1	26.6	0	0	0	0	0	0	0	0	0
14	**4**	**k119_11341**	**k119_11341_1**	**ArsM***	**3**	**19.7**	**23.6**	**21.1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
14	4	k119_11115	k119_11115_59	ArsC	2	48.1	49.3	47.6	0	0	0	0	0	0	0	0	0
15	15	k119_13704	k119_13704_197	ArsR	1	101.5	79.1	96.3	0	0	0	0	0	0	0	0	0
15	15	k119_13704	k119_13704_200	ArsC	1	101.5	79.1	96.3	0.3	0.3	0.7	0.2	0	0.4	0	0	0
15	15	k119_13704	k119_13704_199	Acr3	1	101.5	79.1	96.3	0.5	0.3	0.5	0.1	0.1	0	0	0	0
16	20	k119_31022	k119_31022_78	ArsR	1	37.3	29.8	30.1	23.2	24.1	23.7	2.8	0.6	1.8	76456080.0	72120864.0	153782672.0
16	20	k119_31022	k119_31022_80	ArsA	1	37.3	29.8	30.1	34.5	26.9	21.1	1.4	1.0	2.7	6123088015.5	4214232778.4	7629433184.0
16	20	k119_28101	k119_28101_15	ArsC	1	37.9	30.4	31.0	0.7	3.3	1.0	0	0	0	0	0	0
16	20	k119_1223	k119_1223_28	ArsC	2	37.5	29.0	28.9	0.6	0.5	0.2	0.2	0	0	0	0	0
	NB	k119_9933	k119_9933_27	ArsC	4	10.2	22.6	11.5	0	0	0	0.9	0.3	0	0	0	0
	NB	k119_9898	k119_9898_5	ArsC	20	87.7	79.0	81.0	0	0	0	0	0	0	0	0	0
	NB	k119_3840	k119_3840_2	ArsC*	6	3.5	3.7	4.0	0	0	0	0	0.2	0	0	123636104	277931312
	NB	k119_3840	k119_3840_1	Acr3*	4	3.5	3.7	4.0	0	0	0	0.2	0.9	0.7	0	0	0
	NB	k119_3548	k119_3548_3	ArsC*	2	6.3	6.7	5.1	0	0	0	0	0	0	0	0	0
	NB	k119_34958	k119_34958_15	ArsR*	1	4.9	4.4	5.6	0	0	0	0.6	0.3	0	0	0	0

Table B.11 - part 4 of 4. Abundance values of *ars* genes from EA culture, +As condition.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
	NB	k119_34958	k119_34958_18	ArsC	11	4.9	4.4	5.6	0	0	0	0.5	0.3	0.3	0	0	0
	NB	k119_34671	k119_34671_35	ArsR	6	15.0	14.5	14.9	0	0	0	4.0	2.9	6.2	0	0	0
	NB	k119_34671	k119_34671_39	ArsC	12	15.0	14.5	14.9	0	0	0	6.0	6.5	8.7	0	123636104	277931312
	NB	k119_33744	k119_33744_1	ArsC*	13	14.4	38.7	36.6	763.4	926.3	1090.5	35.2	32.2	46.2	0	0	0
	NB	k119_32701	k119_32701_5	ArsC*	7	6.0	11.9	7.6	0	0	0	0	0	0	0	0	0
	NB	k119_30826	k119_30826_1	ArsC	5	147.4	135.1	172.0	0	0	0	0.1	0.5	0	0	0	0
	NB	k119_30063	k119_30063_5	ArsA*	1	7.8	10.0	8.6	0	0	0	0	0	0	0	0	0
	NB	k119_30063	k119_30063_3	ArsA	5	7.8	10.0	8.6	0	0	0	0	0	0	0	0	0
	NB	k119_30061	k119_30061_2	ArsC*	3	7.3	6.9	2.6	0	0	0	0	0	0	0	0	0
	NB	k119_28281	k119_28281_4	ArsR	2	25.8	40.7	34.8	0	0	0	1.1	0.6	0.5	0	0	0
	NB	k119_28281	k119_28281_8	ArsC	18	25.8	40.7	34.8	0	0	0	0.4	1.2	1.1	1028548032	788588800	0
	NB	k119_28281	k119_28281_7	ArsA*	2	25.8	40.7	34.8	0	0	0	1.2	0.8	0.8	3324732044	2125691183	3726006014
	NB	k119_28281	k119_28281_5	Acr3	16	25.8	40.7	34.8	0	0	0	0.6	0.9	1.1	148083424	161309744	177141872
	NB	k119_22168	k119_22168_2	ArsR*	4	7.5	6.0	8.1	0	0	0	0	0	0	0	0	0
	NB	k119_21124	k119_21124_1	ArsR*	3	73.1	78.9	79.7	287.4	301.0	212.4	6.1	4.5	6.4	0	0	0
	NB	**k119_2076**	**k119_2076_2**	**ArsM**	**2**	**33.1**	**35.0**	**28.2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	NB	k119_2055	k119_2055_3	ArsC	9	107.9	104.5	123.4	1260.8	1436.4	1571.1	211.1	191.2	99.0	0	0	0
	NB	k119_2055	k119_2055_2	Acr3*	6	107.9	104.5	123.4	745.0	880.1	924.8	47.1	45.3	53.4	0	0	0
	NB	k119_20281	k119_20281_1	ArsR	5	3.6	5.4	3.3	0	0	0	0	0	0	0	0	0
	NB	**k119_20281**	**k119_20281_2**	**ArsM**	**1**	**3.6**	**5.4**	**3.3**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	NB	k119_20281	k119_20281_3	ArsC	10	3.6	5.4	3.3	0	0	0	0	0	0	0	0	0
	NB	k119_20281	k119_20281_4	Acr3*	1	3.6	5.4	3.3	0	0	0	0	0	0	0	0	0
	NB	k119_18877	k119_18877_1	Acr3*	8	3.8	4.5	9.1	0	0	0	0.8	0.3	0.6	0	0	0
	NB	k119_18459	k119_18459_2	ArsC*	12	11.2	14.8	7.2	0	0	0	0	0	0	0	0	0
	NB	k119_15186	k119_15186_2	ArsC*	11	58.8	85.3	93.8	202.9	215.8	237.8	10.9	12.0	14.3	0	0	0
	NB	k119_14937	k119_14937_3	ArsA	4	6.6	0.8	3.8	0	0	0	0	0	0	0	0	0
	NB	k119_14937	k119_14937_1	Acr3*	7	6.6	0.8	3.8	0	0	0	0	0	0	0	0	0
	NB	k119_13766	k119_13766_3	Acr3	14	12.5	8.5	5.6	0	0.1	0	0	0.3	0.1	0	0	0
	NB	k119_13677	k119_13677_3	ArsC	19	18.6	32.3	23.4	0	0	0	0.4	0.0	2.7	0	0	0
	NB	k119_12459	k119_12459_7	ArsR	7	6.8	13.9	6.9	0	0	0	0.5	1.2	1.1	0	0	0
	NB	k119_12459	k119_12459_2	ArsC*	8	6.8	13.9	6.9	0	0	0	0.2	0.5	0	1055036445	926977021.5	322633950.5
	NB	k119_12459	k119_12459_3	ArsA	6	6.8	13.9	6.9	0	0	0	1.7	1.3	1.7	3423519980	2191957243	3861157074
	NB	k119_12459	k119_12459_6	Acr3*	5	6.8	13.9	6.9	0	0	0	0.6	0.8	0.2	148083424	161309744	177141872
	NB	k119_12047	k119_12047_6	Acr3*	3	3.7	3.5	4.3	0	0	0	0	0	0	0	0	0
	NB	k119_12047	k119_12047_4	Acr3	15	3.7	3.5	4.3	0	0	0	0	0	0	0	0	0
	NB	k119_12047	k119_12047_5	Acr3	2	3.7	3.5	4.3	0	0	0	0	0	0	0	0	0
	NB	k119_11234	k119_11234_15	ArsC	9	100.0	87.7	92.3	0	0	0	0	0	0	0	0	0
	NB	k119_11234	k119_11234_14	Acr3	17	100.0	87.7	92.3	0	0	0	0	0	0	0	0	0
	NB	k119_10507	k119_10507_2	ArsC*	10	55.3	78.7	66.3	147.6	160.1	183.4	5.4	2.0	3.0	0	0	0
	NB	k119_10507	k119_10507_1	ArsA	3	55.3	78.7	66.3	214.0	254.9	257.0	7.3	8.6	4.8	0	0	0
	NB	k119_10148	k119_10148_27	ArsC*	1	24.5	44.5	35.0	0	0	0	0	0	0	0	0	0

Table B.12 – part 1 of 4. Abundance values of *ars* genes from TSB culture, +As condition. Values are in transcripts per million for genome (TPM−DNA) and transcriptomes (TPM−RNA) and protein abundance for proteome. Contig: contig number assinged by Megahit. Protein id: id number for the protein−encoding genes found by Prodigal. Annotation: protein annotation according to GhostKOALA. An asterisk (*) in protein annotation means no matching protein was clustered by CD-HIT-2D from the no−As control. N. column: corresponds to the gene number in Figure 4. NB in Bin column means the gene was not clustered in any bin. ArsM proteins are in bold. Data shown for the three biological replicates.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
1	9	k119_7837	k119_7837_1	Acr3	1	135.5	159.5	87.7	0.139346	0.154	0.503	4.09065	5.17127	8.69302	0	0	0
2	6	k119_12528	k119_12528_12	ArsR	1	240.4	237.7	381.2	41.80775	81.7	78.2	70.9	73.0	80.9	49300460.88	52300962	636464535
2	6	k119_12528	k119_12528_11	ArsH*	1	240.4	237.7	381.2	219.4648	295.1	341.4	242.531	234.605	286.168	33777216762	32453747931	12967382917
2	6	k119_13639	k119_13639_3	ArsC	1	369.2	355.6	509.8	35.46011	35.92	41.56	99.1	95.6	75.6	0	0	0
2	6	k119_12528	k119_12528_13	ArsB*	1	240.4	237.7	381.2	60.81431	78.37	89.92	63.1344	64.4378	70.341	0	0	0
3	21	k119_11309	k119_11309_2	ArsC*	1	3.4	2.6	2.5	0	0	0	0	0	0	0	0	0
3	21	k119_11309	k119_11309_1	Acr3	1	3.4	2.6	2.5	0	0	0	0	0	0	0	0	0
3	21	k119_3422	k119_3422_1	Acr3	2	3.1	3.4	2.0	0	0	0	0	0	0	0	0	0
4	28	k119_3434	k119_3434_358	ArsR	1	8.1	9.4	8.3	0	0	0	0	0	0	0	0	0
4	28	k119_3434	k119_3434_49	ArsC	1	8.1	9.4	8.3	0	0	0	0	0	0	0	0	0
4	28	k119_3434	k119_3434_362	ArsC	2	8.1	9.4	8.3	0	0	0	0	0	0.34299	0.0	345635152.0	4036908.8
4	28	k119_19999	k119_19999_324	Acr3	1	7.5	8.5	7.6	0	0	0	0	0	0	0	0	0
5	5	k119_9868	k119_9868_36	ArsC	1	7.1	7.9	0.8	0	0	0	0	0	0	0.0	345635152.0	4036908.8
5	5	k119_9868	k119_9868_38	Acr3	1	7.1	7.9	0.8	0	0	0	0	0	0	0	0	0
6	**17**	**k119_11069**	**k119_11069_212**	**ArsM**	**1**	**9.6**	**13.1**	**0.4**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
6	17	k119_29213	k119_29213_24	ArsC	1	6.9	11.2	0.4	0	0	0	0	0	0	0	0	0
7	20	k119_14056	k119_14056_9	ArsR	1	28.3	22.6	16.7	0.343107	0.38	0	0	0	0	0	0	0
7	20	k119_15390	k119_15390_5	ArsC	2	28.9	24.9	16.4	315.4704	245.6	275.4	0	0	0	744561135.0	1044839960.0	1111166504.8
7	20	k119_601	k119_601_127	ArsC	1	32.2	27.9	19.0	13.16821	8.335	4.526	0	0	0	0	0	0
7	20	k119_15390	k119_15390_4	ArsA	1	28.9	24.9	16.4	421.4601	262.2	260.3	0	0	0	15177292771.1	11650375826.5	15447200982.4
7	20	k119_15390	k119_15390_6	Acr3	1	28.9	24.9	16.4	13.71644	13.12	12.47	0	0	0	833627296	683245664	1587592256
7	20	k119_7477	k119_7477_3	Acr3	2	12.7	9.9	7.6	1.833748	2.031	2.482	0	0	0	0	0	0
8	35	k119_19293	k119_19293_1	ArsR	1	18.4	24.4	30.1	0	0	0	0	0	0	0	0	0
8	35	k119_27112	k119_27112_365	ArsR	3	21.6	41.7	43.5	0	0	0	0	0	0	1	1	1
8	35	k119_27112	k119_27112_362	ArsR	2	21.6	41.7	43.5	0	0	0	0	0	0	0	0	0
8	35	k119_8737	k119_8737_69	ArsR	4	19.0	26.3	31.4	0	0	0	0	0	0	0	0	0
8	**35**	**k119_19293**	**k119_19293_5**	**ArsM**	**1**	**18.4**	**24.4**	**30.1**	0	0	0	**0**	**0**	**0**	**5917922**	**1638917.375**	**0**
8	**35**	**k119_30244**	**k119_30244_221**	**ArsM**	**2**	**24.5**	**45.7**	**48.9**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
8	**35**	**k119_8737**	**k119_8737_68**	**ArsM**	**3**	**19.0**	**26.3**	**31.4**	0	0	0	**0**	**0**	**0**	**255680482.8**	**107974884**	**304650954.6**
8	35	k119_19293	k119_19293_2	ArsC	1	18.4	24.4	30.1	0	0	0	0	0	0	0.0	345635152.0	4036908.8
8	35	k119_27112	k119_27112_363	ArsC	2	21.6	41.7	43.5	0	0	0	0	0	0	0.0	345635152.0	4036908.8
8	35	k119_19293	k119_19293_3	Acr3	1	18.4	24.4	30.1	0	0	0	0	0	0	0	0	0
8	35	k119_27112	k119_27112_364	Acr3	2	21.6	41.7	43.5	0	0	0	0	0	0	0	0	0
9	3	k119_17050	k119_17050_30	ArsR	1	5.0	4.5	2.5	0	0	0	0	0	0	0	0	0
9	3	k119_30803	k119_30803_17	ArsR	2	4.8	4.3	2.8	0	0	0	0	0	0	0	0	0
9	3	k119_4945	k119_4945_26	ArsR	3	4.7	4.1	2.4	0	0	0	0	0	0	0	0	0
9	**3**	**k119_17050**	**k119_17050_29**	**ArsM**	**1**	**5.0**	**4.5**	**2.5**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
9	3	k119_17050	k119_17050_28	ArsC	1	5.0	4.5	2.5	0	0	0	0	0	0	0	0	18476252

Table B.12 - part 2 of 4. Abundance values of *ars* genes from TSB culture, +As condition.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
9	3	k119_17050	k119_17050_27	Acr3	1	5.0	4.5	2.5	0	0	0	0	0.10062	0	0	0	0
10	25	k119_15797	k119_15818_38	ArsR	1	92.0	86.8	60.8	0	0	0	0	0	0	0	0	0
10	25	k119_30629	k119_30629_18	ArsR	2	17.2	13.2	14.6	0	0	0	0	0	0	0	0	0
10	25	k119_33283	k119_33283_62	ArsR	3	15.4	11.7	13.2	0	0	0	0	0	0.39576	0	0	0
10	**25**	**k119_15057**	**k119_15057_34**	**ArsM**	**1**	**14.5**	**10.9**	**12.6**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
10	**25**	**k119_15818**	**k119_15818_37**	**ArsM**	**2**	**98.6**	**100.8**	**93.5**	0	0	0	**0.40854**	**0.38055**	**0.33074**	**0**	**0**	**0**
10	25	k119_15818	k119_15818_36	ArsC	1	17.7	13.9	14.7	0	0	0	0	0	0	0	0	18476252
10	25	k119_15818	k119_15818_35	Acr3	1	17.7	13.9	14.7	0	0	0	0.32406	0	0	0	0	0
11	26	k119_4666	k119_4666_37	ArsR	1	7.5	5.8	9.7	0	0	0	3.20725	3.31948	2.59647	0	0	0
11	26	k119_4666	k119_4666_39	ArsC	1	7.5	5.8	9.7	0	0	0	10.1326	16.6415	9.06645	678981717	1123927120	2742067216
11	26	k119_5491	k119_5491_11	ArsC	2	6.2	5.0	8.2	0	0	0	0.6	0.29599	0.4	0	0	0
11	26	k119_4666	k119_4666_38	ArsB*	1	7.5	5.8	9.7	0	0	0	3.02201	1.73866	2.59041	9645447	31508162	73493688
12	15	k119_30292	k119_30292_56	ArsR	1	42.8	24.9	15.8	0	0	0	0	0	0	0	0	0
12	15	k119_20383	k119_20383_128	ArsC	1	58.1	32.5	19.7	0	0	0	0	0	0	0	0	0
12	15	k119_30292	k119_30292_52	ArsC	2	42.8	24.9	15.8	0	0	0	0	0	0	1068039700	1307011274	1898976099
12	15	k119_30292	k119_30292_54	ArsA	1	42.8	24.9	15.8	0	0	0	0	0	0	7392998484.0	7421770091.0	3743273753.7
12	15	k119_30292	k119_30292_53	Acr3	1	42.8	24.9	15.8	0	0	0	0	0	0	39000948	41082988	91581896
13	27	k119_32990	k119_32990_69	ArsC	1	55.3	32.5	50.8	13.44749	9.818	7.997	25.0	24.8	21.4	0	0	0
14	0	k119_1034	k119_1034_7	ArsR	1	2.6	2.8	2.7	0	0	0	0	0	0	0	0	0
14	0	k119_12891	k119_12891_5	ArsR	2	2.7	1.2	2.1	0	0	0	0	0	0	0	0	0
14	**0**	**k119_26245**	**k119_26245_5**	**ArsM**	**1**	**5.4**	**5.1**	**2.6**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
14	0	k119_27340	k119_27340_1	ArsC*	3	3.0	3.1	3.8	0	0	0	0	0	0	0	0	0
14	0	k119_23682	k119_23682_21	ArsC	2	5.6	5.0	2.4	0	0	0	0	0	0	0	345635152	4036908.75
14	0	k119_30081	k119_30081_12	ArsC	1	2.3	2.5	2.9	0	0	0	0	0	0	0	0	0
14	0	k119_1034	k119_1034_9	Acr3	1	2.6	2.8	2.7	0	0	0	0	0	0	0	0	0
14	0	k119_12891	k119_12891_6	Acr3	3	2.7	1.2	2.1	0	0	0	0	0	0	0	0	0
14	0	k119_12891	k119_12891_3	Acr3	2	2.7	1.2	2.1	0	0	0	0	0	0	0	0	0
14	0	k119_23682	k119_23682_23	Acr3	4	5.6	5.0	2.4	0	0	0	0	0	0	0	0	0
15	30	k119_2647	k119_2647_6	ArsC	1	135.9	146.9	8.3	0	0	0	0	0	0	0.0	345635152.0	4036908.8
15	30	k119_9504	k119_9504_27	ArsC	2	98.4	107.4	5.6	0	0	0	0	0	0	734768256.0	1104357648.0	349866824.8
15	30	k119_9504	k119_9504_26	Acr3	1	98.4	107.4	5.6	0	0	0	0	0	0	1221521664	987113600	281258824
16	36	k119_3430	k119_3430_28	ArsR	2	55.8	25.7	13.4	0	0	0	0	0	0	0	0	0
16	36	k119_3430	k119_3430_22	ArsR	1	55.8	25.7	13.4	0	0	0	0	0	0	0	0	0
16	**36**	**k119_3430**	**k119_3430_32**	**ArsM**	**1**	**55.8**	**25.7**	**13.4**	0	0	0	**0**	**0**	**0**	**0**	**0**	**0**
16	36	k119_10810	k119_10810_102	ArsC	1	67.9	30.6	17.1	0	0	0	0	0	0	0	0	0
16	36	k119_10810	k119_10810_115	ArsA	1	67.9	30.6	17.1	0	0	0	0	0	0	0	0	0
16	36	k119_3430	k119_3430_26	ArsA	2	55.8	25.7	13.4	0	0	0	0	0	0	4403864280	3709710861	2788763051
17	14	k119_33769	k119_33769_88	ArsR	1	19.3	19.8	14.7	3.90745	4.328	2.441	2.52104	2.34833	1.14802	0	0	0
17	14	k119_33769	k119_33769_85	ArsC	1	19.3	19.8	14.7	1.592818	1.764	1.198	3.61824	4.14814	3.71781	0	0	0
17	14	k119_33769	k119_33769_86	Acr3	1	19.3	19.8	14.7	0.828931	1.033	0.935	2.93313	2.32742	2.50646	13521992	9829666	4509496
18	29	k119_15797	k119_15797_19	ArsR	1	92.0	86.8	60.8	14.92076	25.14	24.4	232.4	289.3	138.1	0	0	15428906
18	29	k119_4071	k119_4071_3	ArsR	2	69.9	67.6	46.8	17.33492	11.67	11.96	105.5	134.1	71.4	142415272.9	151349210	682362366
18	29	k119_4205	k119_4205_129	ArsR	3	98.6	100.8	93.5	7.817709	7.154	9.507	197.4	169.625	163.145	0	0	17738758

Table B.12 - part 3 of 4. Abundance values of *ars* genes from TSB culture, +As condition.

						Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
18	29	k119_4205	k119_4205_128	ArsH*	1	98.6	100.8	93.5	4.838839	8.818	11.55	251.041	234.605	246.224	473003296	561663447	3172435228
18	29	k119_1079	k119_1079_16	ArsC	1	121.4	118.3	107.9	0	0.671	0.547	22.6	21.0	22.0	0	0	0
18	29	k119_15797	k119_15797_15	ArsC	2	92.0	86.8	60.8	64.7989	53.62	65.86	700.0	823.7	469.9	5200324976	7486884378	15573698586
18	29	k119_4071	k119_4071_5	ArsC	3	69.9	67.6	46.8	25.09997	17.31	15.94	122.179	174.34	90.9769	1174010696	1660568361	4953364498
18	29	k119_4205	k119_4205_131	ArsC	4	98.6	100.8	93.5	5.890809	9.362	12.48	225.0	231.4	200.9	800789713.5	445091655.1	916575460.8
18	29	k119_4205	k119_4205_130	ArsB*	1	98.6	100.8	93.5	2.109591	3.645	3.121	63.964	63.5376	67.8977	9645447	31508162	73493688
18	29	k119_15797	k119_15797_17	ArsA	2	92.0	86.8	60.8	42.97062	42.63	54.39	464.685	599.921	347.594	3746024686	5190362923	2775593150
18	29	k119_15797	k119_15797_13	ArsA	1	92.0	86.8	60.8	1.838899	2.263	4.977	19.7	24.1	18.5	0	0	0
19	32	k119_19214	k119_19214_17	ArsR	1	54.6	33.1	36.2	0	0	0	0	0	0	0	0	0
19	32	k119_19214	k119_19214_14	ArsA	1	54.6	33.1	36.2	0.123077	0	0	0.51615	0.90148	1.802	2077541384	1779374532	562227364.9
19	32	k119_19214	k119_19214_16	Acr3	1	54.6	33.1	36.2	0	0	0	0.10894	0.30444	0.79378	0	0	0
20	4	k119_12019	k119_12019_77	ArsR	1	8.4	6.8	5.7	372.3834	396.5	324.1	2.02464	2.20026	6.55627	305389713.5	265157848	169009443.5
20	4	k119_11336	k119_11336_29	ArsC	1	9.3	7.1	6.0	0.621698	0.344	1.402	0	0	0	0	0	0
20	4	k119_9746	k119_9746_30	ArsC	2	7.6	6.4	5.3	1.31773	4.087	2.853	0	0	0	0	0	0
20	4	k119_12019	k119_12019_79	ArsA	1	8.4	6.8	5.7	282.251	353.8	345	1.7	1.9	6.6	5382006955	3916159701	3220662545
	NB	k119_24571	k119_24571_3	ArsR*	3	2.8	3.4	2.2	0	0	0	0.0	0.0	0.0	0	0	0
	NB	k119_31332	k119_31332_2	ArsR*	4	2.6	3.1	3.0	0	0	0	0.0	0.0	0.0	0	0	0
	NB	k119_18612	k119_18612_2	ArsR	1	4.4	3.7	1.9	0	0	0	0	0	0	0	0	0
	NB	k119_5901	k119_5901_1	ArsR	2	14.8	9.8	7.1	552.0591	515.4	389.7	0	0	0	0	0	0
	NB	**k119_10596**	**k119_10596_1**	**ArsM**	**1**	**0.0**	**5.4**	**8.2**	0	0.831	0.338	**0**	**0**	**1.5724**	**0**	**0**	**0**
	NB	**k119_29944**	**k119_29944_1**	**ArsM**	**2**	**2.4**	**2.7**	**2.1**	0	0	0	**0.9298**	**0.96971**	**1.19782**	**0**	**0**	**0**
	NB	k119_15797	k119_15797_12	ArsH*	1	92.0	86.8	60.8	3.820309	4.57	6.066	45.965	30.1557	34.9264	442735664	902823263	2806618394
	NB	k119_4071	k119_4071_2	ArsH*	2	69.9	67.6	46.8	4.682748	9.337	9.577	62.5003	42.7265	54.4958	6197012908	7490883503	6531758348
	NB	k119_18176	k119_18176_2	ArsC*	12	1.9	4.2	2.5	0	0	0	0	0	0	0	0	0
	NB	k119_11372	k119_11372_2	ArsC	1	1.5	3.1	2.3	0	0	0	0	0	0	0	0	0
	NB	k119_14073	k119_14073_1	ArsC	3	3.3	6.4	2.5	0	0	0	0	0	0	0	0	0
	NB	k119_14299	k119_14299_2	ArsC	4	1.0	2.3	2.5	0.236165	0	0	2.53703	1.95437	1.98081	0	0	0
	NB	k119_17909	k119_17909_2	ArsC	7	3.6	4.1	2.1	0	0	0	0	0	0	0	0	0
	NB	k119_19823	k119_19823_4	ArsC	5	2.1	4.7	2.3	0	0	0	0	0	0	0	0	0
	NB	k119_21777	k119_21777_1	ArsC	8	2.8	4.7	4.9	0	0	0	0	0	0	0	0	0
	NB	k119_27192	k119_27192_1	ArsC	9	1.3	2.3	3.3	0	0	0	0	0	0	0	0	0
	NB	k119_30429	k119_30429_2	ArsC	2	1.4	3.4	1.7	0	0	0	0	0	0	0	0	0
	NB	k119_4811	k119_4811_2	ArsC	10	107.6	113.3	63.5	0	0.133	0	0.58611	1.37536	0.50338	766466488	560810347.1	672361949
	NB	k119_7031	k119_7031_5	ArsC	6	4.3	3.4	3.0	0	0	0	0	0	0	0.0	345635152.0	4036908.8
	NB	k119_8477	k119_8477_2	ArsC	11	55.3	66.9	45.2	0.120031	0.133	0.108	6.2128	7.5	6.9	506854673	369630563	553289502.5
	NB	k119_15797	k119_15797_16	ArsB*	8	92.0	86.8	60.8	9.050045	12.37	16.1	193.534	98.3145	142.502	131172892	96781594.5	350400050.9
	NB	k119_4071	k119_4071_4	ArsB*	9	69.9	67.6	46.8	1.691597	4.497	4.197	31.9666	14.6449	26.0707	230461438	181904174.5	507277686.9
	NB	k119_22971	k119_22971_2	ArsA*	12	1.1	3.1	2.3	0	0	0	0	0	0	943800448	810032048	345328782
	NB	k119_24571	k119_24571_6	ArsA*	11	2.8	3.4	2.2	0	0	0	0	0	0	943800448	810032048	345328782
	NB	k119_12429	k119_12429_1	ArsA	2	1.7	0.0	0.0	0	0	0	0	0	0	0	0	0
	NB	k119_13181	k119_13181_1	ArsA	3	1.7	0.0	5.8	0	0	0	0	0	0	0	0	0
	NB	k119_15852	k119_15852_1	ArsA	4	5.2	3.2	1.6	0	0	0	0	0	0	1396086592	940082152	545298958

Table B.12 - part 4 of 4. Abundance values of *ars* genes from TSB culture, +As condition.

					Genome			Transcriptome R			Transcriptome G			Proteome		
MAG	**Bin**	**Contig**	**Protein id**	**Annotation**	**N.**	**rep1**	**rep2**	**rep3**	**rep1**	**rep2**	**rep3**	rep1	rep2	rep3	rep1	rep2	rep3
	NB	k119_19823	k119_19823_2	ArsA	1	2.1	4.7	2.3	0	0	0	0	0	0	2978261096	2449234509	1149692414
	NB	k119_20090	k119_20090_1	ArsA	5	6.8	11.6	0.0	0	0	0	0	0	0	0	0	0
	NB	k119_21122	k119_21122_1	ArsA	6	12.1	12.3	2.1	0	0	0	0	0.73498	0	0	0	0
	NB	k119_26176	k119_26176_1	ArsA	7	0.0	3.1	0.0	0	0	0	0	0	0	0	0	0
	NB	k119_29234	k119_29234_1	ArsA	8	6.0	3.4	10.5	0	0	0	0	0	0	1782193256	1355532317	470741616.5
	NB	k119_6791	k119_6791_3	ArsA	9	3.8	2.7	3.4	0	0	0	0	0	0	943800448	810032048	345328782
	NB	k119_9020	k119_9020_1	ArsA	10	3.2	7.7	5.1	0	0	0	0	0	0	2008254040	1777201061	552324866.9
	NB	k119_14845	k119_14845_1	Acr3*	7	4.9	3.3	1.4	0	0	0	0	0	0		0	0
	NB	k119_13996	k119_13996_2	Acr3	4	0.0	11.6	8.8	0	0	0	0	0	0		0	0
	NB	k119_14299	k119_14299_1	Acr3	1	1.0	2.3	2.5	0	0	0	1.85852	1.21143	1.55183		0	0
	NB	k119_19823	k119_19823_3	Acr3	2	2.1	4.7	2.3	0	0	0	0	0	0		0	0
	NB	k119_23125	k119_23125_1	Acr3	5	0.7	3.1	2.2	0	0.141	0.229	0.12419	0	0.26665		0	0
	NB	k119_24571	k119_24571_4	Acr3	3	2.8	3.4	2.2	0	0	0	0	0.13043	0		0	0
	NB	k119_8477	k119_8477_1	Acr3	6	55.3	66.9	45.2	0	0	0	7.64633	12.219	5.56072	0	0	0

Bibliography

Ajees, A. A., Marapakala, K., Packianathan, C., Sankaran, B., & Rosen, B. P. (2012). Structure of an As(III) S -adenosylmethionine Methyltransferase: Insights Into The Mechanism of Arsenic Biotransformation. *Biochemistry*, *51*(27), arXiv NIHMS150003, 5476–5485. https://doi.org/10.1021/bi3004632

Ajees, A. A., & Rosen, B. P. (2015). As(III) S-Adenosylmethionine Methyltransferases and Other Arsenic Binding Proteins. *Geomicrobiology Journal*, *32*(7), 570–576. https://doi.org/10.1080/01490451.2014.908983

Akter, K. F., Owens, G., Davey, D. E., & Naidu, R. (2006). *Arsenic speciation and toxicity in biological systems* (Vol. 184). https://doi.org/10.1007/0-387-27565-7_3

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146. https://doi.org/10.1038/nmeth.3103

Andres, E., & Bertin, P. N. (2016). The microbial genomics of arsenic. *FEMS Microbiology Reviews*, *050*(40), 299–322. https://doi.org/10.1093/femsre/fuv050

Andrewes, P., Kitchin, K. T., & Wallace, K. (2003). Dimethylarsine and trimethylarsine are potent genotoxins in vitro. *Chemical Research in Toxicology*, *16*(8), 994–1003. https://doi.org/10.1021/tx034063h

Andrews, S. (2020). FastQC v0.11.9. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Andrianisa, H. A., Ito, A., Sasaki, A., Aizawa, J., & Umita, T. (2008). Biotransformation of arsenic species by activated sludge and removal of bio-oxidised arsenate from wastewater by coagulation with ferric chloride. *Water Research*, *42*(19), 4809–4817. https://doi.org/10.1016/j.watres.2008.08.027

Antizar-Ladislao, B. (2010). Bioremediation: Working with bacteria. *Elements*, *6*(6), 389–394. https://doi.org/10.2113/gselements.6.6.389

Aposhian, H. V., Gurzau, E. S., Le, X. C., Gurzau, A., Healy, S. M., Lu, X., Ma, M., Yip, L., Zakharyan, R. A., Maiorino, R. M., Dart, R. C., Tircus, M. G., Gonzalez-Ramirez, D., Morgan, D. L., Avram, D., & Aposhian, M. M. (2000). Occurrence of monomethylarsonous acid in urine of humans exposed to inorganic arsenic. *Chemical Research in Toxicology*, *13*(8), 693–697. https://doi.org/10.1021/tx000114o

Arraiano, C. M. (1993). Post-transcriptional control of gene expression: bacterial mRNA degradation. Kluwer Academic Publishers. https://doi.org/10.1007/BF00328030

Bagshaw, N. E. (1995). Lead alloys: past, present and future. *Journal of Power Sources*, *53*, 25–30.

Bencko, V., & Foong, F. Y. L. (2017). The history of arsenical pesticides and health risks related to the use of Agent Blue. *Annals of Agricultural and Environmental Medicine*, *24*(2), 312–316. https://doi.org/10.26444/aaem/74715

Bengtsson-Palme, J., Hartmann, M., Eriksson, K. M., Pal, C., Thorell, K., Larsson, D. G. J., & Nilsson, R. H. (2015). Metaxa2: Improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Molecular Ecology Resources*, *15*(6), 1403–1414. https://doi.org/10.1111/1755-0998.12399

Bentley, R., & Chasteen, T. G. (2002). Microbial methylation of metalloids: arsenic, antimony, and bismuth. *Microbiology and molecular biology reviews : MMBR*, *66*(2), 250–71. https://doi.org/10.1128/MMBR.66.2.250-271.2002

Bhattacharjee, H., Mukhopadhyay, R., Thiyagarajan, S., & Rosen, B. P. (2008). Aquaglyceroporins: Ancient channels for metalloids. *Journal of Biology*, *7*(9), 33. https://doi.org/10.1186/jbiol91

Bhattacharjee, H., & Rosen, B. P. (2007). Arsenic Metabolism in Prokaryotic and Eukaryotic Microbes. *Molecular Microbiology of Heavy Metals*, *6*, 371–406. https://doi.org/10.1007/7171_2006_086

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brammer, H., & Ravenscroft, P. (2009). Arsenic in groundwater: A threat to sustainable agriculture in South and South-east Asia. *Environment International*, *35*(3), 647–654. https://doi.org/10.1016/j.envint.2008.10.004

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Bright, D. A., Brock, S., Reimer, K. J., Cullen, W. R., Hewitt, G. M., & Jafaar, J. (1994). Methylation of arsenic by anaerobic microbial consortia isolated from lake sediment. *Applied Organometallic Chemistry*, *8*(4), 415–422. https://doi.org/10.1002/aoc.590080416

Cai, L., Yu, K., Yang, Y., Chen, B. W., Li, X. D., & Zhang, T. (2013). Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology*, *97*(21), 9579–9588. https://doi.org/10.1007/s00253-012-4678-8

Cañadas, I. C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., & Minton, N. P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for Clostridium. *ACS Synthetic Biology*. https://doi.org/10.1021/acssynbio.9b00075

Carlin, A., Shi, W., Dey, S., & Rosen, B. P. (1995). The ars operon of Escherichia coli confers arsenical and antimonial resistance. *Journal of Bacteriology*, *177*(4), 981–986.

Castillo, R., & Saier, M. H. (2010). Functional promiscuity of homologues of the bacterial ArsA ATPases. *International Journal of Microbiology*, *2010*. https://doi.org/10.1155/2010/187373

Challenger, F. (1951). Biological methylation. *Adv. Enzymol. Relat. Areas Mol. Biol.*, *12*(139), 396–416. https://doi.org/10.1039/qr9550900255

Chauhan, N. S., Ranjan, R., Purohit, H. J., Kalia, V. C., & Sharma, R. (2008). Identification of genes conferring arsenic resistance to Escherichia coli from an effluent treatment plant sludge metagenomic library. *FEMS Microbiol. Ecol*, (67), 130–139. https://doi.org/10.1111/j.1574-6941.2008.00613.x

Chen, C., Huang, K., Xie, W.-Y., Chen, S., Tang, Z., & Zhao, F.-J. (2017). Microbial processes mediating the evolution of methylarsine gases from dimethylarsenate in paddy soils. *Environmental Science & Technology*, acs.est.7b04791. https://doi.org/10.1021/acs.est.7b04791

Chen, C., Li, L., Huang, K., Zhang, J., Xie, W. Y., Lu, Y., Dong, X., & Zhao, F. J. (2019). Sulfate-reducing bacteria and methanogens are involved in arsenic methylation and demethylation in paddy soils. *ISME Journal*, 1–13. https://doi.org/10.1038/s41396-019-0451-7

Chen, H. L., Lee, C. C., Huang, W. J., Huang, H. T., Wu, Y. C., Hsu, Y. C., & Kao, Y. T. (2016). Arsenic speciation in rice and risk assessment of inorganic arsenic in Taiwan population. *Environmental Science and Pollution Research*, *23*(5), 4481–4488. https://doi.org/10.1007/s11356-015-5623-z

Chen, J., Bhattacharjee, H., & Rosen, B. P. (2015). ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. *Molecular Microbiology*, *96*(5), 1042–1052. https://doi.org/10.1111/mmi.12988

Chen, J., Madegowda, M., Bhattacharjee, H., & Rosen, B. P. (2015). ArsP: A methylarsenite efflux permease. *Molecular Microbiology*, *98*(4), 625–635. https://doi.org/10.1111/mmi.13145

Chen, J., Qin, J., Zhu, Y. G., de Lorenzo, V., & Rosen, B. P. (2013). Engineering the soil bacterium Pseudomonas putida for arsenic methylation. *Applied and Environmental Microbiology*, *79*(14), 4493–4495. https://doi.org/10.1128/AEM.01133-13

Chen, J., Sarkarai Nadar, V., & Rosen, B. P. (2017). A novel MAs(III)-selective ArsR transcriptional repressor HHS Public Access. *Mol Microbiol*, *106*(3), 469–478. https://doi.org/10.1111/mmi.13826

Chen, J., Yoshinaga, M., & Rosen, B. P. (2019). The antibiotic action of methylarsenite is an emergent property of microbial communities. *Molecular Microbiology*, *111*(2), 487–494. https://doi.org/10.1111/mmi.14169

Chen, P., Li, J., Wang, H. Y., Zheng, R. L., & Sun, G. X. (2017). Evaluation of bioaugmentation and biostimulation on arsenic remediation in soil through biovolatilization. *Environmental Science and Pollution Research*, *24*(27), 21739–21749. https://doi.org/10.1007/s11356-017-9816-5

Chen, S. C., Sun, G. X., Rosen, B. P., Zhang, S. Y., Deng, Y., Zhu, B. K., Rensing, C., & Zhu, Y. G. (2017). Recurrent horizontal transfer of arsenite methyltransferase genes facilitated

adaptation of life to arsenic. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-08313-2

Choi, S. C., & Bartha, R. (1993). Cobalamin-mediated mercury methylation by Desulfovibrio desulfuricans LS. *Applied and environmental microbiology*, *59*(1), 290–5. http://www.ncbi.nlm.nih.gov/pubmed/8439155%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC202093

Clarkson, S. M., Giannone, R. J., Kridelbaugh, D. M., Elkins, J. G., Guss, A. M., & Michenera, J. K. (2017). Construction and optimization of a heterologous pathway for protocatechuate catabolism in escherichia coli enables bioconversion of model aromatic compounds. *Applied and Environmental Microbiology*, *83*(18). https://doi.org/10.1128/AEM.01313-17

Cullen, W. R. (2008a). Arsenic and Crime : The Law of Intended Consequences, In *Is arsen. an aphrodisiac? sociochemistry an elem.*

Cullen, W. R. (2008b). Arsenic Where You Least Expect It, In *Is arsenic an aphrodisiac? the sociochemistry of an element.*

Cullen, W. R. (2008c). Medicinal Arsenic : Toxic Arsenic, In *Is arsenic an aphrodisiac? the sociochemistry of an element.*

Cullen, W. R., & Bentley, R. (2005). The toxicity of trimethylarsine: an urban myth. *J Environ Monit*, *7*, 11–15. https://doi.org/10.1039/b413752n

Cullen, W. R., & Reimer, K. J. (1989). Arsenic speciation in the environment. *Chemical Reviews*, *89*(4), 713–764. https://doi.org/10.1021/cr00094a002

Dixit, S., & Hering, J. G. (2003). Comparison of arsenic(V) and arsenic(III) sorption onto iron oxide minerals: Implications for arsenic mobility. *Environmental Science and Technology*, *37*(18), 4182–4189. https://doi.org/10.1021/es030309t

Dopp, E., Von Recklinghausen, U., Hartmann, L. M., Stueckradt, I., Pollok, I., Rabieh, S., Hao, L., Nussler, A., Katier, C., Hirner, A. V., & Rettenmeier, A. W. (2008). Subcellular distribution of inorganic and methylated arsenic compounds in human urothelial cells and human hepatocytes. *Drug Metabolism and Disposition*, *36*(5), 971–979. https://doi.org/10.1124/dmd.107.019034

Drancourt, M., Bollet, C., Carta, A., & Rousselier, P. (2001). Phylogenetic analyses of Klebsiella species delineate Klebsiella and Raoultella gen. nov., with description of Raoultella ornithinolytica comb. nov., Raoultella terrigena comb. nov. and Raoultella planticola comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*(3), 925–932. https://doi.org/10.1099/00207713-51-3-925

Dunivin, T. K., Yeh, S. Y., & Shade, A. (2019). A global survey of arsenic-related genes in soil microbiomes. *BMC Biology*, *17*(1). https://doi.org/10.1186/s12915-019-0661-5

Edvantoro, B. B., Naidu, R., Megharaj, M., Merrington, G., & Singleton, I. (2004). Microbial formation of volatile arsenic in cattle dip site soils contaminated with arsenic and DDT. *Applied Soil Ecology*, *25*(3), 207–217. https://doi.org/10.1016/j.apsoil.2003.09.006

Fekih, I. B., Zhang, C., Li, Y. P., Zhao, Y., Alwathnani, H. A., Saquib, Q., Rensing, C., & Cervantes, C. (2018). Distribution of arsenic resistance genes in prokaryotes. *Frontiers in Microbiology*, *9*, 2473. https://doi.org/10.3389/fmicb.2018.02473

Fisher, J. C., & Hollibaugh, J. T. (2008). Selenate-dependent anaerobic arsenite oxidation by a bacterium from Mono Lake, California. *Appl. Environ. Microbiol.*, *74*(9), 2588–2594. https://doi.org/10.1128/AEM.01995-07

Garbinski, L. D., Rosen, B. P., & Chen, J. (2019). Pathways of arsenic uptake and efflux. *Environment International*, *126*, 585–597. https://doi.org/10.1016/J.ENVINT.2019.02.058

Goodbred, S. L., & Kuehl, S. A. (2000). The significance of large sediment supply, active tectonism, and eustasy on margin sequence development: Late Quaternary stratigraphy and evolution of the Ganges-Brahmaputra delta. *Sedimentary Geology*, *133*(3-4), 227–248. https://doi.org/10.1016/S0037-0738(00)00041-5

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

Grund, S. C., Hanusch, K., & Wolf, H. U. (2011). Arsenic and Arsenic Compounds. *Ullmann's encyclopedia*, *100 C*, 41–93. https://doi.org/10.1002/14356007.a03

Guentzel, M. N. (1996). Escherichia, Klebsiella, Enterobacter, Serratia, Citrobacter, and Proteus (4th), In *Medical microbiology* (4th). University of Texas Medical Branch at Galveston. http://www.ncbi.nlm.nih.gov/pubmed/21413290

Hayakawa, T., Kobayashi, Y., Cui, X., & Hirano, S. (2005). A new metabolic pathway of arsenite: Arsenic-glutathione complexes are substrates for human arsenic methyltransferase Cyt19. *Archives of Toxicology*, *79*(4), 183–191. https://doi.org/10.1007/s00204-004-0620-x

Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K., & Minton, N. P. (2012). Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research*, *40*(8). https://doi.org/10.1093/nar/gkr1321

Hettick, B. E., Cañas-Carrell, J. E., French, A. D., & Klein, D. M. (2015). Arsenic: A Review of the Element's Toxicity, Plant Interactions, and Potential Methods of Remediation. American Chemical Society. https://doi.org/10.1021/acs.jafc.5b02487

Hohmann-Marriott, M. F., & Blankenship, R. E. (2011). Evolution of Photosynthesis. *Annual Review of Plant Biology*, *62*(1), 515–548. https://doi.org/10.1146/annurev-arplant-042110-103811

Holland, H. D. (2006). The oxygenation of the atmosphere and oceans. Royal Society. https://doi.org/10.1098/rstb.2006.1838

Huang, H., Jia, Y., Sun, G.-X., & Zhu, Y.-G. (2012). Arsenic speciation and volatilization from flooded paddy soils amended with different organic matters. *Environmental science & technology*, *46*(4), 2163–8. https://doi.org/10.1021/es203635s

Huang, K., Chen, C., Zhang, J., Tang, Z., Shen, Q., Rosen, B. P., & Zhao, F.-J. (2016). Efficient Arsenic Methylation and Volatilization Mediated by a Novel Bacterium from an Arsenic-Contaminated Paddy Soil. *Environmental Science & Technology*, *50*, 6389–6396. https://doi.org/10.1021/acs.est.6b01974

Huang, K., Xu, Y., Packianathan, C., Gao, F., Chen, C., Zhang, J., Shen, Q., Rosen, B. P., & Zhao, F. J. (2018). Arsenic methylation by a novel ArsM As(III) S-adenosylmethionine

methyltransferase that requires only two conserved cysteine residues. *Molecular Microbiology*, *107*(2), 265–276. https://doi.org/10.1111/mmi.13882

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Ike, M., Miyazaki, T., Yamamoto, N., Sei, K., & Soda, S. (2008). Removal of arsenic from groundwater by arsenite-oxidizing bacteria. *Water Science and Technology*, *58*(5), 1095–1100. https://doi.org/10.2166/wst.2008.462

Ishizawa, S., Araragi, M., & Suzuki, T. (1969). Actinomycete flora Of Japanese soils. *Soil Science and Plant Nutrition*, *15*(3), 104–112. https://doi.org/10.1080/00380768.1969.10432788

Jayasumana, C., Fonseka, S., Fernando, A., Jayalath, K., Amarasinghe, M., Siribaddana, S., Gunatilake, S., & Paranagama, P. (2015). Phosphate fertilizer is a main source of arsenic in areas affected with chronic kidney disease of unknown etiology in Sri Lanka. *SpringerPlus*, *4*(1), 1–8. https://doi.org/10.1186/s40064-015-0868-z

Jia, Y., Huang, H., Zhong, M., Wang, F.-H., Zhang, L.-M., & Zhu, Y.-G. (2013). Microbial Arsenic Methylation in Soil and Rice Rhizosphere. *Environmental Science & Technology*, *47*, 3141–3148. https://doi.org/10.1021/es303649v

Kaise, T., Yamauchi, H., Horiguchi, Y., Tani, T., Watanabe, S., Hirayama, T., & Fukui, S. (1989). A comparative study on acute toxicity of methylarsonic acid, dimethylarsinic acid and trimethylarsine oxide in mice. *Applied Organometallic Chemistry*, *3*(3), 273–277. https://doi.org/10.1002/aoc.590030311

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kim, Y., & Liesack, W. (2015). Differential Assemblage of Functional Units in Paddy Soil Microbiomes (A. Moustafa, Ed.). *PLOS ONE*, *10*(4), e0122221. https://doi.org/10.1371/journal.pone.0122221

Kopylova, E., Noé, L., & Lè Ne Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kulp, T. R. (2014). Early earth: Arsenic and primordial life. *Nature Geoscience*, *7*(11), 785–786. https://doi.org/10.1038/ngeo2275

Kuramata, M., Sakakibara, F., Kataoka, R., Abe, T., Asano, M., Baba, K., Takagi, K., & Ishikawa, S. (2015). Arsenic Biotransformation By Streptomyces sp. Isolated From Rice Rhizosphere. *Environmental Microbiology*, *17*(6), 1897–1909. https://doi.org/10.1111/1462-2920.12572

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lawrence, J. G. (2002). Shared strategies in gene organization among prokaryotes and eukaryotes. Cell Press. https://doi.org/10.1016/S0092-8674(02)00900-5

Le, X. C., Lu, X., Ma, M., Cullen, W. R., Aposhian, H. V., & Zheng, B. (2000). Speciation of key arsenic metabolic intermediates in human urine. *Analytical Chemistry*, *72*(21), 5172–5177. https://doi.org/10.1021/ac000527u

Lebrun, E., Brugna, M., Baymann, F., Muller, D., Lett, M.-C., & Nitschke, W. (2003). Arsenite Oxidase, an Ancient Bioenergetic Enzyme. *Mol. Biol. Evol.*, *20*(5), 686–693. https://doi.org/10.1093/molbev/msg071

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, J., Pawitwar, S. S., & Rosen, B. P. (2016). The organoarsenical Biocycle And The Primordial Antibiotic Methylarsenite. *Metallomics*, *8*(10), 1047–1055. https://doi.org/10.1039/c6mt00168h

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Liao, Y., Smyth, G. K., & Shi, W. (2014). Sequence analysis featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, S., Shi, Q., Brent Nix, F., Styblo, M., Beck, M. A., Herbin-Davis, K. M., Hall, L. L., Simeonsson, J. B., & Thomas, D. J. (2002). A novel S-adenosyl-L-methionine:arsenic(III) methyltransferase from rat liver cytosol. *Journal of Biological Chemistry*, *277*(13), 10795–10803. https://doi.org/10.1074/jbc.M110246200

Liu, C.-W., Chi, L., Tu, P., Xue, J., Ru, H., & Lu, K. (2018). Isobaric Labeling Quantitative Metaproteomics for the Study of Gut Microbiome Response to Arsenic. https://doi.org/10.1021/acs.jproteome.8b00666

Liu, S., Zhang, F., Chen, J., & Sun, G. (2011). Arsenic removal from contaminated soil via biovolatilization by genetically engineered bacteria under laboratory conditions. *Journal of Environmental Sciences*, *23*(9), 1544–1550. https://doi.org/10.1016/S1001-0742(10)60570-0

Liu, Z., Osamura, M., Ootsuka, T., Wang, S., Fukuzawa, Y., Suzuki, Y., Kuroda, R., Mise, T., Otogawa, N., Nakayama, Y., Tanoue, H., & Makita, Y. (2005). Doping of β-FeSi2 films with boron and arsenic by sputtering and its application for optoelectronic devices, In *Optical materials*, Elsevier. https://doi.org/10.1016/j.optmat.2004.08.040

Lomax, C., Liu, W. J., Wu, L., Xue, K., Xiong, J., Zhou, J., McGrath, S. P., Meharg, A. A., Miller, A. J., & Zhao, F. J. (2012). Methylated arsenic species in plants originate from soil microorganisms. *New Phytologist*, *193*(3), 665–672. https://doi.org/10.1111/j.1469-8137.2011.03956.x

López-Maury, L., Florencio, F. J., & Reyes, J. C. (2003). Arsenic sensing and resistance system in the cyanobacterium Synechocystis sp. Strain PCC 6803. *Journal of Bacteriology*, *185*(18), 5363–5371. https://doi.org/10.1128/JB.185.18.5363-5371.2003

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. (2019). The EMBL-EBI search and sequence

analysis tools APIs in 2019. *Nucleic Acids Research*, (1). https://doi.org/10.1093/nar/gkz189

Mandal, B. K., & Suzuki, K. T. (2002). Arsenic round the world: A review. *Talanta*, *58*(1), 201–235. https://doi.org/10.1016/S0039-9140(02)00268-0

Marapakala, K., Packianathan, C., Ajees, A. A., Dheeman, D. S., Sankaran, B., Kandavelu, P., & Rosen, B. P. (2015). A disulfide-bond cascade mechanism for arsenic(III) S-adenosylmethionine methyltransferase. *Acta Crystallographica Section D: Biological Crystallography*, *71*(3), 505–515. https://doi.org/10.1107/S1399004714027552

Marapakala, K., Qin, J., & Rosen, B. P. (2012). Identification of catalytic residues in the As(III) S-adenosylmethionine methyltransferase. *Biochemistry*, *51*(5), 944–951. https://doi.org/10.1021/bi201500c

Mass, M. J., Tennant, A., Roop, B. C., Cullen, W. R., Styblo, M., Thomas, D. J., & Kligerman, A. D. (2001). Methylated Trivalent Arsenic Species are Genotoxic. *Chemical Research in Toxicology*, *14*(4), 355–361. https://doi.org/10.1021/tx000251l

Masuda, H. (2018). Arsenic cycling in the Earth's crust and hydrosphere: interaction between naturally occurring arsenic and human activities. *Progress in Earth and Planetary Science*, *5*(1). https://doi.org/10.1186/s40645-018-0224-3

McCarty, K. M., Hanh, H. T., & Kim, K. W. (2011). Arsenic geochemistry and human health in South East Asia. *Reviews on Environmental Health*, *26*(1), 71–78. https://doi.org/10.1515/REVEH.2011.010

McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J., & Yates, J. R. (2002). 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. *International Journal of Mass Spectrometry*, *219*(1), 245–251. https://doi.org/10.1016/S1387-3806(02)00563-8

Meharg, A. A., & Zhao, F.-J. (2012). *Arsenic & Rice*. Springer. https://link.springer.com/content/pdf/10.1007%7B%5C%%7D2F978-94-007-2947-6.pdf

Meng, X.-Y., Qin, J., Wang, L.-H., Duan, G.-L., Sun, G.-X., Wu, H.-L., Chu, C.-C., Ling, H.-Q., Rosen, B. P., & Zhu, Y.-G. (2011). Arsenic biotransformation and volatilization in transgenic rice. *New Phytologist*, *191*(1), 49–56. https://doi.org/10.1111/j.1469-8137.2011.03743.x

Mestrot, A., Feldmann, J., Krupp, E. M., Hossain, M. S., Roman-Ross, G., & Meharg, A. A. (2011). Field fluxes and speciation of arsines emanating from soils. *Environmental Science and Technology*, *45*(5), 1798–1804. https://doi.org/10.1021/es103463d

Mestrot, A., Merle, J. K., Broglia, A., Feldmann, J., & Krupp, E. M. (2011). Atmospheric stability of arsine and methylarsines. *Environmental Science and Technology*, *45*(9), 4010–4015. https://doi.org/10.1021/es2004649

Mestrot, A., Planer-Friedrich, B., & Feldmann, J. (2013). Biovolatilisation: A Poorly Studied Pathway Of The Arsenic Biogeochemical Cycle. *Environmental science. Processes & impacts*, *15*(9), 1639–1651. https://doi.org/10.1039/c3em00105a

Mestrot, A., Uroic, M. K., Plantevin, T., Islam, M. R., Krupp, E. M., Feldmann, J., & Meharg, A. A. (2009). Quantitative and Qualitative Trapping of Arsines Deployed to Assess Loss

of Volatile Arsenic from Paddy Soil. *Environmental Science and Technology*, *43*(21), 8270–8275.

Meyer, J., Michalke, K., Kouril, T., & Hensel, R. (2008). Volatilisation of metals and metalloids: An inherent feature of methanoarchaea? *Systematic and Applied Microbiology*, *31*(2), 81–87. https://doi.org/10.1016/j.syapm.2008.02.001

Michalke, K., Wickenheiser, E. B., Mehring, M., Hirner, A. V., & Hensel, R. (2000). Production of volatile derivatives of metal(loid)s by microflora involved in anaerobic digestion of sewage sludge. *Applied and Environmental Microbiology*, *66*(7), 2791–2796. https://doi.org/10.1128/AEM.66.7.2791-2796.2000

Miyashita, S. I., Fujiwara, S., Tsuzuki, M., & Kaise, T. (2012). Cyanobacteria produce arsenosugars. *Environmental Chemistry*, *9*(5), 474–484. https://doi.org/10.1071/EN12061

Mochizuki, H. (2019). Arsenic neurotoxicity in humans. *International Journal of Molecular Sciences*, *20*(14). https://doi.org/10.3390/ijms20143418

Moran, M. A., Satinsky, B., Gifford, S. M., Luo, H., Rivers, A., Chan, L. K., Meng, J., Durham, B. P., Shen, C., Varaljay, V. A., Smith, C. B., Yager, P. L., & Hopkinson, B. M. (2013). Sizing up metatranscriptomics. Nature Publishing Group. https://doi.org/10.1038/ismej.2012.94

Mudhoo, A., Sharma, S. K., Garg, V. K., & Tseng, C. H. (2011). Arsenic: An overview of applications, health, and environmental concerns and removal processes. *Critical Reviews in Environmental Science and Technology*, *41*(5), 435–519. https://doi.org/10.1080/10643380902945771

Mukhopadhyay, R., & Rosen, B. P. (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environmental Health Perspectives*, *110*(5), 745–748. https://doi.org/10.1289/ehp.02110s5745

Murphy, J. N., & Saltikov, C. W. (2009). The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of Shewanella sp. strain ANA-3. *Journal of Bacteriology*, *191*(21), 6722–6731. https://doi.org/10.1128/JB.00801-09

Nesnow, S., Roop, B. C., Lambert, G., Kadiiska, M., Mason, R. P., Cullen, W. R., & Mass, M. J. (2002). DNA damage induced by methylated trivalent arsenicals is mediated by reactive oxygen species. *Chemical Research in Toxicology*, *15*(12), 1627–1634. https://doi.org/10.1021/tx025598y

O'Brien, R. W., & Morris, J. G. (2009). Oxygen And The Growth And Metabolism of Clostridium Acetobutylicum. *Journal of General Microbiology*, *68*(3), 307–318. https://doi.org/10.1099/00221287-68-3-307

O'Malley, M. A. (2013). Metametabolomics, In *Encyclopedia of systems biology*. Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_903

Oremland, R. S., Hoeft, S. E., Santini, J. M., Bano, N., Hollibaugh, R. A., & Hollibaugh, J. T. (2002). Anaerobic oxidation of arsenite in Mono Lake water and by a facultative, arsenite-oxidizing chemoautotroph, strain MLHE-1. *Appl. Environ. Microbiol.*, *68*(10), 4795–4802. https://doi.org/10.1128/AEM.68.10.4795-4802.2002

Oremland, R. S., & Stolz, J. F. (2005). Arsenic, microbes and contaminated aquifers. *Trends in Microbiology*, *13*(2), 45–49. https://doi.org/10.1016/j.tim.2004.12.002

Oremland, R. S., Stolz, J. F., & Hollibaugh, J. T. (2004). The microbial arsenic cycle in Mono Lake, California. *FEMS Microbiology Ecology*, *48*(1), 15–27. https://doi.org/10.1016/j.femsec.2003.12.016

Ospino, M. C., Kojima, H., & Fukui, M. (2019). Arsenite Oxidation by a Newly Isolated Betaproteobacterium Possessing arx Genes and Diversity of the arx Gene Cluster in Bacterial Genomes. *Frontiers in Microbiology*, *10*(MAY), 1210. https://doi.org/10.3389/fmicb.2019.01210

Packianathan, C., Kandavelu, P., & Rosen, B. P. (2018). The Structure of an As(III) S-Adenosylmethionine Methyltransferase with 3-Coordinately Bound As(III) Depicts the First Step in Catalysis. *Biochemistry*, *57*, 4083–4092. https://doi.org/10.1021/acs.biochem.8b00457

Pakulska, D., & Czerczak, S. (2006). Hazardous effects of arsine: A short review. *International Journal of Occupational Medicine and Environmental Health*, *19*(1), 36–44. https://doi.org/10.2478/v10001-006-0003-z

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Petrick, J. S., Jagadish, B., Mash, E. A., & Aposhian, H. V. (2001). Monomethylarsonous acid (MMAIII) and arsenite: LD50in hamsters and in vitro inhibition of pyruvate dehydrogenase. *Chemical Research in Toxicology*, *14*(6), 651–656. https://doi.org/10.1021/tx000264z

Pfaffl, M. W., Horgan, G. W., & Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research*, *30*(9), e36.

Plewniak, F., Crognale, S., Rossetti, S., & Bertin, P. N. (2018). A genomic outlook on bioremediation: The case of arsenic removal. *Frontiers in Microbiology*, *9*(820), 1–8. https://doi.org/10.3389/fmicb.2018.00820

Prabaharan, C., Kandavelu, P., Packianathan, C., Rosen, B. P., & Thiyagarajan, S. (2019). Structures of two ArsR As(III)-responsive transcriptional repressors: Implications for the mechanism of derepression. *Journal of Structural Biology*. https://doi.org/10.1016/j.jsb.2019.05.009

Prohaska, T., & Stingeder, G. (2005). Speciation of Arsenic: Arsenic and Arsenic Species in Environment and Human Nutrition. *Handbook of Elemental Speciation II - Species in the Environment, Food, Medicine and Occupational Health*, *3*, 69–85. https://doi.org/10.1002/0470856009.ch2d(i)

Pyne, M. E., Moo-Young, M., Chung, D. A., & Chou, C. P. (2013). Development Of An Electrotransformation Protocol For Genetic Manipulation Of Clostridium Pasteurianum. *Biotechnology for Biofuels*, *6*(1). https://doi.org/10.1186/1754-6834-6-50

Qin, J., Rosen, B. P., Zhang, Y., Wang, G., Franke, S., & Rensing, C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine

methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2075–80. https://doi.org/10.1073/pnas.0506836103

Rahman, M. A., & Hassler, C. (2014). Is arsenic biotransformation a detoxification mechanism for microorganisms? *Aquatic Toxicology*, *146*, 212–219. https://doi.org/10.1016/j.aquatox.2013.11.009

Rascovan, N., Maldonado, J., Vazquez, M. P., & Eugenia Farías, M. (2016). Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. *ISME Journal*, *10*(2), 299–309. https://doi.org/10.1038/ismej.2015.109

Reid, M. C., Maillard, J., Bagnoud, A., Falquet, L., Le Vo, P., & Bernier-Latmani, R. (2017). Arsenic Methylation Dynamics in a Rice Paddy Soil Anaerobic Enrichment Culture. *Environmental Science & Technology*, *51*, 10546–10554. https://doi.org/10.1021/acs.est.7b02970

Rutherford, D. W., Bednar, A. J., Garbarino, J. R., Needham, J., Staver, K. W., & Wershaw, R. L. (2003). Environmental Fate of Roxarsone in Poultry Litter. Part II. Mobility of Arsenic in Soils Amended with Poultry Litter. https://doi.org/10.1021/ES026222+

Sajith Wickramasekara, A. S., & Li, C. (2012). Benchling: The Life Sciences R&D Cloud. https://www.benchling.com/

Sardiwal, S., Santini, J. M., Osborne, T. H., & Djordjevic, S. (2010). Characterization of a two-component signal transduction system that controls arsenite oxidation in the chemolithoautotroph NT-26. John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1574-6968.2010.02121.x

Schaeffer, L., Pimentel, H., Bray, N., Melsted, P., & Pachter, L. (2017). Pseudoalignment for metagenomic read assignment. *Bioinformatics*, *33*(14), arXiv 1510.07371, 2082–2088. https://doi.org/10.1093/bioinformatics/btx106

Schwarz, K. M., Grosse-Honebrink, A., Derecka, K., Rotta, C., Zhang, Y., & Minton, N. P. (2017). Towards Improved Butanol Production Through Targeted Genetic Modification of Clostridium Pasteurianum. *Metabolic Engineering*, *40*, 124–137. https://doi.org/10.1016/j.ymben.2017.01.009

Seyfferth, A. L., Webb, S. M., Andrews, J. C., & Fendorf, S. (2010). Arsenic localization, speciation, and co-occurrence with iron on rice (Oryza sativa L) roots having variable Fe coatings. *Environmental Science and Technology*, *44*(21), 8108–8113. https://doi.org/10.1021/es101139z

Shen, Z., Luangtongkum, T., Qiang, Z., Jeon, B., Wang, L., & Zhang, Q. (2014). Identification of a novel membrane transporter mediating resistance to organic arsenic in Campylobacter jejuni. *Antimicrobial Agents and Chemotherapy*, *58*(4), 2021–2029. https://doi.org/10.1128/AAC.02137-13

Shi, K., Li, C., Rensing, C., Dai, X., Fan, X., & Wang, G. (2018). Efflux transporter ArsK is responsible for bacterial resistance to arsenite, antimonite, trivalent roxarsone, and methylarsenite. *Applied and Environmental Microbiology*, *84*(24), 1842–1860. https://doi.org/10.1128/AEM.01842-18

Singh, R., Singh, S., Parihar, P., Singh, V. P., & Prasad, S. M. (2015). Arsenic contamination, consequences and remediation techniques: A review. *Ecotoxicology and Environmental Safety*, *112*, 247–270. https://doi.org/10.1016/j.ecoenv.2014.10.009

Spratlen, M. J., Grau-Perez, M., Umans, J. G., Yracheta, J., Best, L. G., Francesconi, K., Goessler, W., Bottiglieri, T., Gamble, M. V., Cole, S. A., Zhao, J., & Navas-Acien, A. (2019). Targeted metabolomics to understand the association between arsenic metabolism and diabetes-related outcomes: Preliminary evidence from the Strong Heart Family Study. *Environmental Research*, *168*, 146–157. https://doi.org/10.1016/j.envres.2018.09.034

Styblo, M., Del Razo, L. M., Vega, L., Germolec, D. R., LeCluyse, E. L., Hamilton, G. A., Reed, W., Wang, C., Cullen, W. R., & Thomas, D. J. (2000). Comparative toxicity of trivalent and pentavalent inorganic and methylated arsenicals in rat and human cells. *Archives of Toxicology*, *74*(6), 289–299. https://doi.org/10.1007/s002040000134

Subhraveti, P., Ong, Q., Keseler, I., Kothari, A., Caspi, R., & Karp, P. D. (2018). Summary of Clostridium sp. CPB-6 CPB6, version 23.5. Retrieved March 30, 2020, from https://biocyc.org/organism-summary?object=GCF%7B%5C_%7D002119605

Teaf, C. M., Covert, D. J., Teaf, P. A., Page, E., & Starks, M. J. (2010). Arsenic Cleanup Criteria for Soils in the US and Abroad: Comparing Guidelines and Understanding Inconsistencies. *Proceedings of the Annual International Conference on Soils, Sediments, Water and Energy*, *15*(January), 94–102. http://scholarworks.umass.edu/soilsproceedings%7B%5C%%7D5Cnhttp://scholarworks.umass.edu/soilsproceedings

Thomas, F., Diaz-Bone, R. A., Wuerfel, O., Huber, B., Weidenbach, K., Schmitz, R. A., & Hensel, R. (2011). Connection Between Multimetal(loid) Methylation In Methanoarchaea And Central Intermediates Of Methanogenesis. *Applied and Environmental Microbiology*, *77*(24), 8669–8675. https://doi.org/10.1128/AEM.06406-11

Upadhyaya, G., Jackson, J., Clancy, T. M., Hyun, S. P., Brown, J., Hayes, K. F., & Raskin, L. (2010). Simultaneous removal of nitrate and arsenic from drinking water sources utilizing a fixed-bed bioreactor system. *Water Research*, *44*(17), 4958–4969. https://doi.org/10.1016/j.watres.2010.07.037

Uppal, J. S., Zheng, Q., & Le, X. C. (2019). Arsenic in drinking water—recent examples and updates from Southeast Asia. *Current Opinion in Environmental Science and Health*, *7*, 126–135. https://doi.org/10.1016/j.coesh.2019.01.004

Van Lis, R., Nitschke, W., Duval, S., & Schoepp-Cothenet, B. (2013). Arsenics as bioenergetic substrates. *Biochimica et Biophysica Acta - Bioenergetics*, *1827*(2), 176–188. https://doi.org/10.1016/j.bbabio.2012.08.007

Vasken Aposhian, H., Zakharyan, R. A., Avram, M. D., Sampayo-Reyes, A., & Wollenberg, M. L. (2004). A review of the enzymology of arsenic metabolism and a new potential role of hydrogen peroxide in the detoxication of the trivalent arsenic species. *Toxicology and Applied Pharmacology*, *198*(3), 327–335. https://doi.org/10.1016/j.taap.2003.10.027

Viacava, K., Meibom, K. L., Ortega, D., Dyer, S., Gelb, A., Falquet, L., Minton, N. P., Mestrot, A., & Bernier-Latmani, R. (n.d.). Arsenic methylation in arsM-harboring microorganisms. *Unpublished manuscript*.

Wagner, G. P., Kin, K., & Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, *131*(4), 281–285. https://doi.org/10.1007/s12064-012-0162-3

Wang, J., Wu, M., Lu, G., & Si, Y. (2016). Biotransformation and biomethylation of arsenic by Shewanella oneidensis MR-1. *Chemosphere*, *145*, 329–335. https://doi.org/10.1016/j.chemosphere.2015.11.107

Wang, M., Tang, Z., Chen, X. P., Wang, X., Zhou, W. X., Tang, Z., Zhang, J., & Zhao, F. J. (2019). Water management impacts the soil microbial communities and total arsenic and methylated arsenicals in rice grains. *Environmental Pollution*, *247*, 736–744. https://doi.org/10.1016/j.envpol.2019.01.043

Wang, P. P., Bao, P., & Sun, G. X. (2015). Identification And Catalytic Residues Of The Arsenite Methyltransferase From a Sulfate-Reducing Bacterium, Clostridium sp. BXM. *FEMS Microbiology Letters*, *362*(1), 1–8. https://doi.org/10.1093/femsle/fnu003

Wang, P. P., Sun, G. X., & Zhu, Y. G. (2014). Identification And Characterization Of Arsenite Methyltransferase From An Archaeon, Methanosarcina Acetivorans C2A. *Environmental Science and Technology*, *48*(21), arXiv 1011.1669, 12706–12713. https://doi.org/10.1021/es503869k

Wang, P., Sun, G., Jia, Y., Meharg, A. A., & Zhu, Y. (2014). A review on completing arsenic biogeochemical cycle: Microbial volatilization of arsines in environment. *Journal of Environmental Sciences (China)*, *26*(2), 371–381. https://doi.org/10.1016/S1001-0742(13)60432-5

Washburn, M. P., Wolters, D., & Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, *19*(3), 242–247. https://doi.org/10.1038/85686

Watanabe, T., & Hirano, S. (2013). Metabolism of arsenic and its toxicological relevance. *Archives of Toxicology*, *87*(6), 969–979. https://doi.org/10.1007/s00204-012-0904-5

Williams, P. N., Villada, A., Deacon, C., Raab, A., Figuerola, J., Green, A. J., Feldmann, J., & Meharg, A. A. (2007). Greatly enhanced arsenic shoot assimilation in rice leads to elevated grain levels compared to wheat and barley. *Environmental Science and Technology*, *41*(19), 6854–6859. https://doi.org/10.1021/es070627i

World Health Organization. (2018). Arsenic. Retrieved March 7, 2020, from https://www.who.int/news-room/fact-sheets/detail/arsenic

Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, *12*(1), 444. https://doi.org/10.1186/1471-2164-12-444

Wuerfel, O., Thomas, F., Schulte, M. S., Hensel, R., & Diaz-Bone, R. A. (2012). Mechanism of multi-metal(loid) methylation and hydride generation by methylcobalamin and cob(I)alamin: A side reaction of methanogenesis. *Applied Organometallic Chemistry*, *26*(2), 94–101. https://doi.org/10.1002/aoc.2821

Xiao, K. Q., Li, L. G., Ma, L. P., Zhang, S. Y., Bao, P., Zhang, T., & Zhu, Y. G. (2016). Metagenomic analysis revealed highly diverse microbial arsenic metabolism genes in paddy soils

with low-arsenic contents. *Environmental Pollution*, *211*, 1–8. https://doi.org/10.1016/j.envpol.2015.12.023

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., & Chen, S. (2012). FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads. *PLoS ONE*, *7*(12), e52249. https://doi.org/10.1371/journal.pone.0052249

Xu, X. Y., McGrath, S. P., Meharg, A. A., & Zhao, F. J. (2008). Growing Rice Aerobically Markedly Decreases Arsenic Accumulation. *Environmental Science and Technology*, *42*(15), 5574–5579. https://doi.org/10.1021/es800324u

Xue, X. M., Raber, G., Foster, S., Chen, S. C., Francesconi, K. A., & Zhu, Y. G. (2014). Biosynthesis of arsenolipids by the cyanobacterium Synechocystis sp. PCC 6803. *Environmental Chemistry*, *11*(5), 506–513. https://doi.org/10.1071/EN14069

Xue, X. M., Ye, J., Raber, G., Rosen, B. P., Francesconi, K., Xiong, C., Zhu, Z., Rensing, C., & Zhu, Y. G. (2019). Identification of Steps in the Pathway of Arsenosugar Biosynthesis. *Environmental Science and Technology*, *53*(2), 634–641. https://doi.org/10.1021/acs.est.8b04389

Xue, X.-M., Ye, J., Raber, G., Francesconi, K. A., Li, G., Gao, H., Yan, Y., Rensing, C., & Zhu, Y.-G. (2017). Arsenic Methyltransferase is Involved in Arsenosugar Biosynthesis by Providing DMA. https://doi.org/10.1021/acs.est.6b04952

Yamaguchi, N., Nakamura, T., Dong, D., Takahashi, Y., Amachi, S., & Makino, T. (2011). Arsenic release from flooded paddy soils is influenced by speciation, Eh, pH, and iron dissolution. *Chemosphere*, *83*(7), 925–932. https://doi.org/10.1016/j.chemosphere.2011.02.044

Yan, W., Dilday, R. H., Tai, T. H., Gibbons, J. W., McNew, R. W., & Rutger, J. N. (2005). Differential response of rice germplasm to straighthead induced by arsenic. *Crop Science*, *45*(4), 1223–1228. https://doi.org/10.2135/cropsci2004.0348

Yan, Y., Ding, K., Yu, X. W., Ye, J., & Xue, X. M. (2017). Ability of Periplasmic Phosphate Binding Proteins from Synechocystis sp. PCC 6803 to Discriminate Phosphate Against Arsenate. *Water, Air, and Soil Pollution*, *228*(4), 1–9. https://doi.org/10.1007/s11270-017-3334-4

Yang, H. C., Cheng, J., Finan, T. M., Rosen, B. P., & Bhattacharjee, H. (2005). Novel pathway for arsenic detoxification in the legume symbiont Sinorhizobium meliloti. *Journal of Bacteriology*, *187*(20), 6991–6997. https://doi.org/10.1128/JB.187.20.6991-6997.2005

Yang, P., Ke, C., Zhao, C., Kuang, Q., Liu, B., Xue, X., Rensing, C., & Yang, S. (2020). ArsM-mediated arsenite volatilization is limited by efflux catalyzed by As efflux transporters. *Chemosphere*, *239*, 124822. https://doi.org/10.1016/j.chemosphere.2019.124822

Ye, J., Rensing, C., Rosen, B. P., & Zhu, Y. G. (2012). Arsenic biomethylation by photosynthetic organisms. https://doi.org/10.1016/j.tplants.2011.12.003

Yin, X. X., Chen, J., Qin, J., Sun, G. X., Rosen, B. P., & Zhu, Y. G. (2011). Biotransformation and volatilization of arsenic by three photosynthetic cyanobacteria. *Plant Physiology*, *156*(3), 1631–1638. https://doi.org/10.1104/pp.111.178947

Yoshinaga, M., Cai, Y., & Rosen, B. P. (2011). Demethylation of methylarsonic acid by a microbial community. *Environmental Microbiology*, *13*(5), arXiv NIHMS150003, 1205–1215. https://doi.org/10.1111/j.1462-2920.2010.02420.x

Yoshinaga, M., & Rosen, B. P. (2014). A C-As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(21), 7701–6. https://doi.org/10.1073/pnas.1403057111

Yu, X., Zheng, W., Bhat, S., Aquilina, J. A., & Zhang, R. (2015). Transcriptional and posttranscriptional regulation of Bacillus sp. CDB3 arsenic-resistance operon ars1. *PeerJ*, *2015*(9). https://doi.org/10.7717/peerj.1230

Zakharyan, R. A., & Aposhian, H. V. (1999). Arsenite methylation by methylvitamin B12 and glutathione does not require an enzyme. *Toxicology and Applied Pharmacology*, *154*(3), 287–291. https://doi.org/10.1006/taap.1998.8587

Zhang, J., Cao, T., Tang, Z., Shen, Q., Rosen, B. P., & Zhao, F. J. (2015). Arsenic methylation and volatilization by arsenite s-adenosylmethionine methyltransferase in Pseudomonas alcaligenes NBRC14159. *Applied and Environmental Microbiology*, *81*(8), 2852–2860. https://doi.org/10.1128/AEM.03804-14

Zhang, P. (2017). On arsenic trioxide in the clinical treatment of acute promyelocytic leukemia. Elsevier Ltd. https://doi.org/10.1016/j.lrr.2017.03.001

Zhang, S. Y., Zhao, F. J., Sun, G. X., Su, J. Q., Yang, X. R., Li, H., & Zhu, Y. G. (2015). Diversity and abundance of arsenic biotransformation genes in paddy soils from Southern China. *Environmental Science & Technology*, *49*(7), 4138–4146. https://doi.org/10.1021/acs.est.5b00028

Zhang, S.-Y., Williams, P. N., Luo, J., & Zhu, Y.-G. (2017). Microbial mediated arsenic biotransformation in wetlands. *Front. Environ. Sci. Eng.*, *11*(1), 1–11. https://doi.org/10.1007/s11783-017-0893-y

Zhao, F. J., Zhu, Y. G., & Meharg, A. A. (2013). Methylated arsenic species in rice: Geographical variation, origin, and uptake mechanisms. *Environmental Science and Technology*, *47*(9), 3957–3966. https://doi.org/10.1021/es304295n

Zheng, M. Z., Li, G., Sun, G. X., Shim, H., & Cai, C. (2013). Differential toxicity and accumulation of inorganic and methylated arsenic in rice. *Plant and Soil*, *365*(1-2), 227–238. https://doi.org/10.1007/s11104-012-1376-3

Zheng, R. L., Sun, G. X., & Zhu, Y. G. (2013). Effects of microbial processes on the fate of arsenic in paddy soil. Springer. https://doi.org/10.1007/s11434-012-5489-0

Zhu, Y. G., Xue, X. M., Kappler, A., Rosen, B. P., & Meharg, A. A. (2017). Linking Genes to Microbial Biogeochemical Cycling: Lessons from Arsenic. *Environmental Science and Technology*, *51*(13), 7326–7339. https://doi.org/10.1021/acs.est.7b00689

Zhu, Y.-G., Yoshinaga, M., Zhao, F.-J., & Rosen, B. P. (2014). Earth Abides Arsenic Biotransformations. *Annual Review of Earth and Planetary Sciences*, *42*(1), 443–467. https://doi.org/10.1146/annurev-earth-060313-054942

Curriculum Vitae

Karen Elda Viacava Romo was born in Morelia, Mexico, on the 12th of January 1988. In 2010, she completed her Engineer's degree in Biochemical Engineering with a minor in Environmental Engineering at the Morelia Institute of Technology. For her Bachelor thesis she worked on the optimization of hydrogen production by a lab-scale UASB reactor at the Laboratory of advance processes for wastewater treatment (LIPATA) at the National Autonomous University of Mexico (UNAM) supported by a scholarship awarded by the Mexican Academy of Science (AMC). Immediately after she was offered a job at the Faculty of Basic Sciences from her university where she taught Mathematics and Probability to first- and second-year bachelor students for a year.

In 2011, she was granted a scholarship from the Mexican National Council of Science and Technology (CONACyT) in cooperation with the German Academic Exchange Service (DAAD) to continue her studies at the Biotechnology Center (BIOTEC) of Dresden University of Technology (TU Dresden). In 2013, she obtained her MSc degree there in Molecular bioengineering. She completed her master thesis at the Institute of Resource Ecology from the Helmholtz-Zentrum Dresden-Rossendorf (HZDR) studying the metal sorption efficiency on the cell wall components from gram-positive bacteria isolated from a uranium-mining site.

In 2014, she moved to Basel to the Institute of Ecopreneurship at the University of Applied Sciences and Arts Northwestern Switzerland (FHNW) working on the use of selenium nanoparticles as a food additive for cattle suffering of selenium deficiency.

Since February 2016, she has been working on her PhD-thesis at the Environmental Microbiology Laboratory (EML). She has been under the supervision of Prof. Rizlan Bernier-Latmani, at the Swiss Federal Institute of Technology (EPFL), and the co-supervision of Prof. Adrien Mestrot at the Group of Soil Science from the University of Bern. The research focussed on the methylation of arsenic by soil microorganisms, the results of which are described in this thesis.

Conferences attended

- 14^{th} Swiss Geoscience Meeting. November 2016, Geneva, Switzerland. **Poster** in the session *Environmental biochemistry of trace elements.*
- 14^{th} International Conference on the Biogeochemistry of Trace Elements (ICOBTE). July 2017, Zurich, Switzerland. **Short oral presentation** in the session *Interactions of organisms with TE: from molecular mechanisms to ecological effects.*
- 28^{th} Goldschmidt. August 2018. **Oral presentation** in the session *Biogeochemistry of Oxyanion-Forming Elements: The Good, the Bad and the Ugly.*
- 16^{th} Swiss Geoscience Meeting. November 2018. Bern, Switzerland. **Oral presentation** in the session *Environmental biochemistry of trace elements.*
- 15^{th} ICOBTE. May 2019, Nanjing, China. **Oral presentation** in the session *Methylation and demethylation of trace elements.*
- 29^{th} Goldschmidt. August 2019. **Oral presentation** in the session *Trace elements speciation: novel methodologies and insights into transformations influencing their global biogeochemical cycle.*

Publications

- **Viacava K**, Meibom KL, Ortega D, Dyer S, Gelb A, Falquet L, Minton NP, Mestrot A, Bernier-Latmani R. *Variability in arsenic methylation efficiency across aerobic and anaerobic microorganisms.* Unpublished Manuscript.
- **Viacava K**, Janowczyk A, Poudel S, Reid MC, Jecquier L, Shrestha H, Hettich RL, Bernier-Latmani, R. *Active arsenic-methylating bacteria in a rice paddy soil microbiome.* Unpublished Manuscript.
- Asta MP, Wang Y, Frutschi M, **Viacava K**, Loreggian L, Le Pape P, Le Vo P, Fernández AM, Morin G, Bernier-Latmani R. *Microbially Mediated Release of As from Mekong Delta Peat Sediments.* Environ Sci Technol. 2019; 53:10208-10217. DOI:10.1021/acs.est.9b02887.
- Suhr M, Unger N, **Viacava KE**, Günther TJ, Raff J, Pollmann K. *Investigation of metal sorption behavior of Slp1 from* Lysinibacillus sphaericus *JG-B53: a combined study using QCM-D, ICP-MS and AFM.* BioMetals. 2014; 27(6). DOI:10.1007/s10534-014-9794-8