Phylogeny and diversity of genes for poorly characterized type of arsenite oxidase involved in anaerobic arsenic oxidation

Melody Christine Cabrera Ospino

Doctoral thesis

2018

Graduate School of Environmental Science

Hokkaido University

CONTENTS

Abstract

The biological arsenic cycle is mainly governed by microbial processes of interconversion of two main inorganic forms of arsenic: arsenite, and arsenate, which are proceeded by oxidation and reduction, respectively. Arsenite is more abundant in anoxic environments than oxic environments and thus anaerobic arsenite oxidation may be an important process in the arsenic cycle. Anaerobic arsenite oxidation have been described in bacteria having an enzyme coded by the recently discovered *arx* genes. The *arx* genes were first reported in 2010, and have been primarily found in anaerobic arsenite-oxidizing bacteria within the class *Gammaproteobacteria*, isolated from alkaline, saline arsenic-rich environments. Finding of an *arx*-like gene cluster in the *Betaproteobacteria Sulfuricella denitrificans* skB26^{T}, isolated from a freshwater lake, prompted us to ask about the distribution and phylogenetic diversity of these genes. However, until now, there are only a limited numbers of *arx*-harboring cultivated strains available.

To contribute to further understand the phylogeny and diversity of *arx* genes, the following studies were conducted. The first analysis was aimed to explore the *arxA* gene in three samples from non-alkaline saline environments. The exploration was made by clone library approach, analyzing fragments of the *arxA* gene amplified by PCR. For this purpose, a PCR primer set was newly designed based on the sequences available in the public database. Partial *arxA*-like gene fragments were recovered in all samples, and they were phylogenetically distinct from those of *Alkalilimnicola ehrlichii* MLHE-1^{T}, *Ectothiorhodospira* and *Thioalkalivibrio* species isolated from soda lakes (high-saline environments). The majority of the *arxA*-like gene fragments formed a unique cluster within the ArxA clade including the sequence of *Sulfuricella denitrificans* skB26^{T}. The remaining sequences were clustered with the ArxA sequence *Azoarcus* sp. CIB isolated from

diesel-fuel contaminated aquifer in Switzerland. These results show evidence of a wider distribution of *arxA*-like sequences in non-extreme environments.

The second analysis was planned to explore the taxonomic and environmental distribution of *arx* gene cluster using prokaryotic genome database. All the genomes harboring a homologous *arx* gene cluster were affiliated to the phylum *Proteobacteria*, and the majority belonged to the classes *Gammaproteobacteria* and *Betaproteobacteria*. The *arx* gene cluster is formed by the *arxB'ABCD* genes, encoding for the putative arsenite oxidase were homogeneous in almost all the genomes regarding both their genetic content and organization; and the regulatory genes, *arxXSR* predicted to control the expression of the enzyme, were absence in closely related species of the order *Ocenanospirillales*. It was also found that the *arxA* gene sequences in some genomes hold a long or short insertion at the same position. In the phylogenetic tree of ArxA, sequences with the insertion exclusively belonged to the cluster which includes *Sulfuricella denitrificans* skB26^{T}. A clone library exploration of *arxA* sequences with long insert confirmed the presence of diverse *arxA* gene fragments with insertion in two environmental samples. These results show evidence of the genetic diversity of *arxA* genes and conservation of *arxB'ABCD* gene cluster within the *Proteobacteria*.

Lastly, a novel arsenite-oxidizing betaproteobacterium strain M52 was isolated from a hot spring microbial mat. Strain M52 can transform arsenite to arsenate in the presence of nitrate or low concentration of oxygen. Detection of *nap* genes encoding nitrate reductase and *ccoNOQP* and *cydAB* genes coding for two microaerophilic-related cytochrome oxidases may suggest the use of nitrate and oxygen as electron acceptors. Lack of *aio* genetic signature and presence of a complete *arx* gene cluster are suggestive of an *arx*-mediated arsenite oxidation. The *arxA* gene of this strain has the long insertion, as is the case with other two strains isolated from the same microbial mat.

The nearest relatives of strain M52 are *Georgfuchsia toluolica* G5G6 and *Denitratisoma oestradiolicum* AcBE2-1 with 94% of similarity, based on comparison of nearly full length (~1400 bp) 16S rRNA gene sequences. The distinct phylogenetic identity suggests that it represent a novel genus within the family *Sterolibacteriaceae*. Further study is needed to genetically validate the arsenite oxidation by the *arx* genes, and confirm the chemical species used as electron acceptor by strain M52.

Chapter 1
General Introduction

1.1 Introduction

The arsenic (As) is a natural ocurring and widespread metalloid, present in different environments, but usually associated with hydrothermal water and minerals (1). The terrestial occurrence of arsenic is around 1.5 to 3 mg/Kg, and it can be concentrated in various types of metal iron and sulfide deposits, one of the most common the arsenopyrite (2). The release of arsenic to the surface can be through the various natural process, such as weathering of the minerals, hydrothermal and volcanos emissions; but also anthropogenic, like the use of arsenical pesticides, mining, among others (3). Therefore, some of these conditions can lead to the enrichment of arsenic in the geothermal fluids, being usually found in geothermal surface manifestation, like hot springs, fumaroles, geyser. (4). However, arsenic is a hazardous chemical, and the excessive intake may cause health deteriorations, named arsenicosis, that include skin disorder, internal cancers, diseases of the blood vessels and reproductive disorder (5).

Thee most prevalent arsenic chemical species in the terrestrial environment are the inorganic forms, arsenite ($As(OH)_3$ and H_2AsO^{3-}, referred to as As(III), or As^{3+}) and arsenate (H_2AsO^{4-} and $HAsO_4^{2-}$, referred to as As(V), or As^{5+}) (5). Arsenite is very soluble, and thus more bioavailable and toxic of the inorganic arsenic pair. Despite this, prokaryotes have evolved to develop arsenic resistant mechanisms, and some others can use them as a source of energy (3), by reducing the arsenate or oxidizing the arsenite, interconverting the arsenic between the 5+ and 3+ valence state (arsenotrophy). These “arsenotrophic” bacteria have been found present in As-rich environments, but also presumably were distributed widely on the early earth (6).

The arsenic microbial transformation plays an essential role in the biochemical arsenic cycle by modifying the arsenic speciation and mobility (7). For example, arsenate respiring

bacteria can contribute to the release of As(III) from sediments or arsenate-containing minerals (8–11). And in other cases, this arsenite can serve as an electron donor for many microbes. In these processes, the study of microbial arsenite oxidation becomes even crucial as a strategy for removal of the arsenic in solution, since the arsenate, as the product of the arsenite oxidation, can be effectively removed by adsorption (12).

The first arsenite oxidase was purified from the *Betaproteobacterium Alcaligenes faecalis*, and all the aerobic arsenite-oxidizing bacteria discovered to date involves the arsenite oxidase code by the *aioAB* genes (13). However, the arsenite and thioarsenites are the dominant arsenic species in the anoxic or low redox aquatic systems (14). And regardless of this, AioA has only been linked to aerobic respiration, and there is no direct evidence showing the involvement of AioA in anaerobic arsenite oxidation (13).

Recently a new clade of arsenite oxidase, named Arx, within the DMSO reductase family was described to be involved in the arsenite oxidation under anaerobic conditions (15). The *arxA* was identified and genetically validated in anaerobic respiration of nitrate by a haloalkaliphilic arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* MLHE-1^T (16). Arx has also been genetically demonstrated to be responsible during the anoxygenic photosynthesis of *Ectothiorhodospira* sp. strain BLS-9 while growing on arsenite (arsenophototrophy) (17). *Ectothiorhodospira* sp. PHS-1 which was the first reported “arsenophototroph” also holds an *arxA*-like gene (15, 18, 19). It is note worthy that all of them belong to the class *Gammaproteobacteria* and the family *Ectothiorhodospiraceae*. Later it was reported the presence of an *arx*-like gene cluster within the plasmid of the *Betaproteobacterium Sulfuricella denitrificans* skB26^T (20), isolated from freshwater in Lake Mizugaki (21). It was the first described Arx other than the class *Gammaproteobacteria*, although the bacterium is not capable

of oxidizing de arsenite in the laboratory experiment. Until now, a couple of strains harboring an *arxA*-like gene with arsenite-oxidizing activity have been isolated.

Previously, *arxA*-like genes were known to be widely distribute in arsenic-rich alkaline saline Mono Lake, and Hot Creek sediment in California, alkaline microbial mat in Yellowstone National Park (15), and alkaline saline lake in Khovsgol, Mongolia (22). Therefore, it is highly possible that *arxA*-like genes are widely distributed in non-extreme environments.

This thesis aimed to explore more the ecological distribution and phylogenetic diversity of the *arx* genes, involved in anaerobic arsenite oxidation. The diversity and distribution of *arxA* gene were analyzed in three low-salt environments, a freshwater lake, hot spring water and microbial mat, by using a newly designed PCR primer set (chapter 2). The phylogenetic and genetic diversity of *arx*-like gene cluster were analyzed in available prokaryotic genomes, retrieved from Genbank database (chapter 3). Moreover, the isolation of an arsenite-oxidizing betaproteobacterium harboring an *arx*-like gene cluster was described (chapter 4). This thesis contributes to the growing knowledge of *arx* genetic diversity in non-extreme environments.

1.2 Reference

1. **Henke KR**. 2009 Arsenic in Natural Environments.

2. **Mandal BK and Suzuki KT**. 2002 Arsenic round the world: A review. Talanta **58**:201–235.

3. **Slyemi D and Bonnefoy V**. 2012 How prokaryotes deal with arsenic. Environ. Microbiol. Rep. **4**:571–586.

4. **Bundschuh J and Maity JP**. 2015 Geothermal arsenic: Occurrence, mobility and environmental implications. Renew. Sustain. Energy Rev. **42**:1214–1222.

5. **Shankar S, Shanker U, Shikha, Shankar S, Shanker U and Shikha**. 2014 Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation, Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation. Sci. World J. **2014**, **2014**:e304524.

6. **Kulp TR**. 2014 Early earth: Arsenic and primordial life. Nat. Geosci. **7**:785–786.

7. **Oremland RS, Saltikov CW, Wolfe-Simon F and Stolz JF**. 2009 Arsenic in the evolution of earth and extraterrestrial ecosystems. Geomicrobiol. J. **26**:522–536.

8. **Newman DK, Beveridge TJ and Morel FMM**. 1997 Precipitation of arsenic trisulfide by *Desulfotomaculum auripigmentum*. Appl. Environ. Microbiol. **63**:2022–2028.

9. **Newman DK, Ahmann D and Morel FM**. 1998 A brief review of micobial arsenate respiration. **15**:255–268.

10. **Macy JM, Santini JM, Pauling B V, O'Neill a H and Sly LI**. 2000 Two new arsenate/sulfate-reducing bacteria: mechanisms of arsenate reduction. Arch. Microbiol. **173**:49–57.

11. **Zobrist J, Dowdle PR, Davis JA and Oremland RS**. 2000 Mobilization of arsenite by dissimilatory reduction of adsorbed arsenate. Environ. Sci. Technol. **34**:4747–4753.

12. **Páez-Espino D, Tamames J, De Lorenzo V and Cánovas D**. 2009 Microbial responses to environmental arsenic. BioMetals **22**:117–130.

13. **Zhu Y-G, Yoshinaga M, Zhao F-J and Rosen BP**. 2014 Earth Abides Arsenic Biotransformations. Annu. Rev. Earth Planet. Sci. **42**:443–467.

14. **Wilkin RT, Wallschläger D and Ford RG**. 2003 Speciation of arsenic in sulfidic waters. Geochem. Trans. **4**:1.

15. **Zargar K, Conrad A, Bernick DL, Lowe TM, Stolc V, Hoeft S, Oremland RS, Stolz J and Saltikov CW**. 2012 ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. Environ. Microbiol. **14**:1635–1645.

16. **Zargar K, Hoeft S, Oremland R and Saltikov CW**. 2010 Identification of a novel arsenite oxidase gene, arxA, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J. Bacteriol. **192**:3755–3762.

17. **Hernandez-Maldonado J, Sanchez-Sedillo B, Stoneburner B, Boren A, Miller L, Mccann S, Rosen M, Oremland RS and Saltikov CW**. 2017 The genetic basis of anoxygenic photosynthetic arsenite oxidation. Environ. Microbiol. **19**:130–141.

18. **Kulp TR, Hoeft SE, Asao M, Madigan MT, Hollibaugh JT, Fisher JC, Stolz JF, Culbertson CW, Miller LG and Oremland RS**. 2008 Arsenic(III) Fuels Anoxygenic Photosynthesis in Hot Spring Biofilms from Mono Lake, California. Science (80-.). **1**:967–970.

19. **Hoeft SE, Kulp TR, Han S, Lanoil B and Oremland RS**. 2010 Coupled arsenotrophy in a hot spring photosynthetic biofilm at mono lake, california. Appl. Environ. Microbiol. **76**:4633–4639.

20. **Watanabe T, Kojima H and Fukui M**. 2014 Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: Genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst. Appl. Microbiol. **37**:387–395.

21. **Kojima H and Fukui M**. 2010 *Sulfuricella denitrificans* gen. nov., sp. nov., a sulfur-oxidizing autotroph isolated from a freshwater lake. Int. J. Syst. Evol. Microbiol. **60**:2862–2866.

22. **Hamamura N, Itai T, Liu Y, Reysenbach AL, Damdinsuren N and Inskeep WP**. 2014 Identification of anaerobic arsenite-oxidizing and arsenate-reducing bacteria associated with an alkaline saline lake in Khovsgol, Mongolia. Environ. Microbiol. Rep. **6**:476–482.

Chapter 2

Diversity of anaerobic arsenite-oxidizing bacteria in low-salt environments analyzed with a newly developed PCR-based method.

Ospino MC, Kojima H, Watanabe T, Iwata T and Fukui M

Limnology 19:177-183 (2018).

2.1 Introduction

Arsenic is a ubiquitous element in the environment. It is the 20th most abundant element on the earth's crust, with a terrestrial abundance of *ca* 1.5 – 3.0 mg/Kg (1). The majority of the arsenic is present in the environment as inorganic forms, pentavalent arsenate and trivalent arsenite. These chemical species are very toxic to life, because they both have detrimental effects on cellular mechanism (2). Despite the high toxicity, some prokaryotes can take advantage of these inorganic arsenic compounds to support their growth, by reducing arsenate or oxidizing arsenite. The respiratory arsenate reduction can generate energy, and this dissimilatory reduction of arsenate is mediated by an arsenate reductase referred to as Arr. Some other prokaryotes can oxidize arsenite either via two types of arsenite oxidase, Aio or Arx. The catalytic subunits of these three enzymes, i.e. ArrA, AioA and ArxA, belong to the dimethyl sulfide (DMSO) reductase family (3).

The Arx is known as enzyme involved in anaerobic arsenite oxidation which can be coupled with nitrate respiration or photosynthetic inorganic carbon fixation. The name ArxA was first used to refer to the arsenite oxidase found in *Alkalilimnicola ehrlichii* MLHE-1^{T}, which is more closely related to ArrA rather than AioA (4). This enzyme was originally recognized as Arr with bidirectional activity, which was expressed when the strain was chemoautotrophically grown under nitrate-reducing arsenite-oxidizing conditions (5). By gene disruption experiments, it was demonstrated that *arxA* gene is essential for nitrate-dependent oxidation of arsenite by the strain MLHE-1 (3, 6). The involvement of Arx in phototrophic arsenite oxidation was first suggested by gene expression experiments with *Ectothiorhodospira* sp. PHS-1 (7). The presence of the *arxA* gene was further shown in some related phototrophic

strains capable of anaerobic arsenite oxidation (8), and its indispensable role in the light-dependent arsenite oxidation was confirmed by disrupting this gene in one of these strains (9).

In order to explore diversity and distribution of microbes involved in the arsenic cycle, some PCR primers specific to genes for the ArrA, AioA and ArxA have been developed and applied. With these primers, some studies revealed community structure of arsenic-metabolizing microbes which have the *arrA* or *aioA* genes (10–15). As for the *arxA* gene, clones of *arxA* gene have been recovered environments with a specific primer pair, arxA_Deg_F_B and arxA_Deg_R_B (3, 9, 16). These primers were designed based on the Arx sequences of only three strains belonging to the family *Ectothiorhodopiraceae* in the class *Gammaproteobacteria* (3). Therefore, it is likely that *arx*A genes of phylogenetically distant organisms are not covered by this primer pair. After these primers were reported, more diverse *arxA* gene sequences have been available in the public databases. Majority of them were found in the genomes of halophilic and alkaliphilic bacteria, mainly phototrophic or chemolithotrophic sulfur oxidizers. Some others are originated from bacteria phylogenetically distant from the family *Ectothiorhodopiraceae*, such as *Sulfuricella denitrificans* skB26^{T} belonging to the class *Betaproteobacteria* (17, 18). In fact, the *arxA* gene sequences of some strains have mismatches with the sequences of arxA_Deg_F_B and arxA_Deg_R_B (Table 2-1).

Considering that arsenite is more abundant in anoxic environments (2), anaerobic oxidation by Arx must be an important process in the arsenic cycle. The purpose of this study was to provide effective method to explore diversity and distribution of the *arxA* gene in the environment. For that purpose, a PCR primer set was newly designed and then modified to facilitate rapid cloning analysis. Further, effectiveness of the developed method was

demonstrated by analyzing the *arxA* gene diversity in samples of lake water, spring water, and hot spring microbial mat, characterized by low salinity and nearly neutral pH.

2.2 Materials and Methods

2.2.1 Primer design

PCR primers were designed on the basis sequences available in the GenBank database at the beginning of this study. The sequences were retrieved from the genome of the following strains: *Alkalilimnicola ehrilichii* MLHE-1, *Ectothiorhodospira* sp. PHS-1, *Halorhodospira halophile* SL1, *Halomonas* sp. A3H3, *Thioalkalivibrio nitratireducens* DMS 14787, *Azoarcus* sp. CIB, *Halomonas* sp. BC04, *Thioalkalivibrio* sp. ALM2T, *Halomonas boliviensis* LC1, *Nitrincola lacisaponensis* 4CA, *Thioalkalivibrio* sp. AKL19, *Thioalkalivibrio sulfidiphilus* ALJ17, *Thioalkalivibrio thiocyanodenitrificans* ARhD1, *Thioalkalivibrio* sp. ARh3, *Thioalkalivibrio* sp. ALMg11, *Thioalkalivibrio* sp. ALSr1, *Thioalkalivibrio* sp. ALR17-21, *Thioalkalivibrio* sp. AKL17, *Magnetospirillum magnetotacticum* MS-1 and *Sulfuricella denitrificans* skB26^{T}. The gene sequences were translated to amino acids, and aligned with Dialign-Pfam (Al Ait et al. 2013). From conserved regions manually identified, arxA_deg_194F (5'-GTCCGGCNCGYTGCGGNAT-3') and arxA_deg_490R (5'-CCCCAKCCRCGVCCGTA-3') were designed.

2.2.2 PCR amplification of arxA gene from reference strains

Specificity of the designed primer pair was tested with bacterial strains which have the *arx*A gene. The genomic DNA of following strains were purchased from DSMZ (Germany); *Halorhodospira halophile* (DSM-244), *Thioalkalivibrio nitratireducens* (DSM-14787),

Halomonas boliviensis (DSM-15516), *Nitrincola lacisaponensis* (DSM-16316), *Thioalkalivibrio thiocyanodenitrificans* (DSM-16954), and *Alkalilimnicola ehrilichii* MLHE-1 (DSM-17681). The genomic DNA of *Sulfuricella denitrificans* skB26^T, available in the laboratory, was also used for the experiments. As a control strain without the *arxA* gene, *Sulfuritalea hydrogenivorans* sk43H was also tested. This strain was selected because it has the *arrA* gene (18). PCR reactions were performed in 25-μL volume reaction mixture, containing 0.5 μmol l^{-1} of each primer, 1x Ex Taq Buffer (Takara, Shiga, Japan), 0.2 mmol l^{-1} dNTPs (Takara), 0.625 U of Ex Taq (Takara), 3% DMSO and template DNA solution. PCR conditions were optimized to reduce nonspecific amplification, and following touchdown program was finally selected: initial denaturalization at 94°C for 3 min; 14 cycles of 94°C for 30 s, annealing temperature (reduced by 1°C in each cycle from 69°C to 56°C) for 30 s, and 72°C for 30 s; 20 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s.

2.2.3 Physicochemical and microbiological characterization of environmental samples

Three environmental samples were obtained from three sites in Japan, characterized by relatively high arsenic concentration in water (Table 2-2). Two water samples were collected from Lake Mizugaki and Masutomi hot spring, both situated in Yamanashi Prefecture, Japan. Lake Mizugaki is an artificial lake where arsenic has been detected by governmental water monitoring and isolation source of *Sulfuricella denitrificans* skB26^T was obtained. The sample analyzed in this study (hereafter referred to as Mizugaki) was also used in previous studies, and obtained from 35 m water depth. Masutomi hot spring is located upstream of Lake Mizugaki. The sample of this spring (hereafter referred to as Masutomi), obtained from a storage tank, is identical to that used in a previous study (19). The other sample of microbial mat, named

Jozankei, was obtained from Jozankei hot spring situated in Hokkaido, Japan (42° 57' 53" N 141° 09' 47" E) on May 13, 2016. The dark green microbial mat sample was kept in sterile plastic tubes completely filled with hot spring water, transferred to the laboratory on ice, and used for DNA extraction within the 24 hours.

For the water samples of Mizugaki and Masutomi, pH, conductivity and temperature were measured on site by using portable meters. Concentrations of chloride and sulfate were measured with an ion chromatograph, and total arsenic concentration was determined with an inductively coupled plasma optical emission spectroscopy. For Jozankei, spring water running on the mat was analyzed in the same ways for the other samples of water. Some physicochemical data have already been presented in the previous studies, as indicated in Table 2-2.

From the water samples, DNA was extracted as described previously (Kojima et al. 2009). As for Jozankei, DNA was prepared by using Wizard® Genomic DNA Purification Kit (Promega, USA). Initially, approximately 0.5 g (wet weight) of the microbial mat was placed in 0.15 ml of Nuclei Lysis buffer in the kit. The mat matrix was mechanically destructured in the buffer with a sterile pestle. The homogenized product was subdivided into aliquots (approximately 50 μL), and subjected to the procedures described in the manufacturer's instruction.

In order to characterize bacterial communities of the samples, clone libraries of the 16S rRNA gene and *apr*A gene were built. The 16S rRNA gene was analyzed with a universal primer pair to analyze whole bacterial communities. The *apr*A gene, which encodes adenylylsulfate reductase, has been found in the genome of many *arxA*-carrying bacteria. The analyses of these genes were performed as described previously in the paper by Kojima et al (20) and Watanabe

et al (21), with the DNA samples of Masutomi and Jozankei. As for the Mizugaki, sequences of these genes have already been published in the previous study by Kojima et al. (20), and sequences in the public database were retrieved.

2.2.4 Clone library analysis of arxA genes in the environmental samples

From the DNA samples of three sites, fragments of the *arxA* gene were amplified with the primers arxA_deg_194F and modified version of the reverse primer, GT-arxA_deg_490R (5'-gtCCCCAKCCRCGVCCGTA-3'). The modified primer has two extra nucleotides at 5'-end (shown in lower-case letters) to improve efficiency of TA cloning. After electrophoresis on 1.5% (w/v) agarose gel, PCR products of the expected size were purified with QIAquick gel extraction kit (Qiagen, Valencia, CA). The purified products were ligated into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA), and the vectors were transformed into TOP10 cells (Invitrogen), according to the protocol described by the manufacturer. The resulting colonies were subjected to direct colony PCR using the vector primers, and the resulting PCR products of the expected were sequenced.

The cloned sequences of the *arxA* gene were translated into amino acids sequences using MEGA software package version 6 (22). The deduced amino acid sequences were aligned using ClustalW (http://www.genome.jp/tools/clustalw/). A pairwise distance matrix was calculated based on a Poisson model on the MEGA. The calculated distance matrices were input into Mothur (23), and grouped into operational taxonomic units (OTUs) at a 97% cutoff level. The most abundant sequence in each OTU was selected as representative sequence for further analysis. The sequences were aligned with references sequences, by using ClustalW. The resulting alignment was manually trimmed, and then uploaded to PHYML 3.0 online (24) to

infer a maximum-likelihood tree. The generated tree was statistically evaluated by Approximate Likelihood-Ratio Test for Branches (aLRT) (25).

2.2.5 Nucleotide sequence accession numbers

The nucleotide sequences obtained in this study have been deposited under the accession numbers LC275986-LC276219.

2.3 Results

2.3.1 Design and evaluation of the primers for *arx*A gene

In this study, new primer pair was designed on the basis of the sequences 20 strains. During period of this study, further *arxA* gene-like sequences became available in the public database (Table 2-3). By constructed phylogenetic tree of the DMSO reductase family molybdenum enzymes, it was confirmed that proteins coded by these sequences form a robust cluster, distinct from ArrA, AioA and other related proteins (Table 2-1). This clade contains ArxA of the strains MLHE-1 and PHS-1. Further, proteins of this clade share a conserved sequence motif in a molybdenum-binding site (Table 2-4), which is thought to be indispensable for the function of ArxA (3). Considering these observations, it is reasonable to regard these sequences as the *arxA* gene. The primers designed in this study have some mismatches with these sequences of *arxA*, but the numbers of mismatches are generally smaller than those of the previously reported primer pair (Table 2-1, Table 2-3).

The designed primer pair, arxA_deg_194F and arxA_deg_490R, was tested with bacterial strains known to have the *arx*A gene. With the touch down-PCR program, products of the

expected size (around 330 bp) were obtained from all *arxA*-carrying strains, except for *Halomonas boliviensis* LC1, whose *arxA* gene has four mismatches with the primer arxA_deg_194F (Table 2-3). With the same strains, the modified version of the reverse primer, GT-arxA_deg_490R was also used in combination with arxA_deg_194F. This modified reverse primer also gave PCR products of the expected size (Fig. 2-1).

2.3.2 Characteristics of the environmental samples

Physicochemical characteristics of the environmental samples are summarized in Table 2-2. The whole bacterial community structure was analyzed by constructing clone libraries of the 16S rRNA gene. The clones were grouped into OTUs based on 97% similarity and their taxonomic affiliations were identified by RDP classifier (confidence threshold = 80%). As shown in Table 2-5, the phylum *Proteobacteria* was the dominant in the water samples (Mizugaki and Masutomi), and phyla *Cyanobacteria* and *Bacteroidetes* were dominant in Jozankei. Among the OTUs, 46 OTUs could be classified toward the genus level, but none of the genera identified were known to include strains with the *arxA* gene (Table 2-6). The samples were also analyzed with the *aprA* gene, which has been widely used as a marker gene to detect sulfur oxidizers and sulfate reducers. The clones of *aprA* gene were grouped into 32 OTUs, and 17 OTUs showed high sequence similarities to sulfur oxidizers (Table 2-7). Although many *arxA* gene-carrying strains also have the *aprA* gene (Fig. 2-2), none of the detected *aprA* OTUs was closely related to that of the organisms known to have the *arxA* gene (Table 2-7).

2.3.3 Detection of diverse *arx*A sequences in the environmental samples

With the new primers for *arxA* gene, PCR products were obtained from the environmental samples. In the earlier stage of this study, PCR products were obtained with the primer pair arxA_deg_194F and arxA_deg_490R. Despite repeated trials, these products could not be analyzed with TA cloning. To facilitate more rapid analysis, GT-arxA_deg_490R was applied for further investigations of the environmental samples. The PCR products obtained with the modified reverse primer were successfully analyzed by subsequent TA cloning. In total, 146 clones were obtained from the three libraries and grouped into 11 OTUs. The phylogenetic analysis revealed that all these OTUs fall within cluster of ArxA (Fig. 2-2). This result indicates that all clones of the expected size in the libraries were partial sequences of the *arxA* gene.

In the phylogenetic tree, ArxA sequences formed some distinct clusters (Fig. 2). One of the clusters, named “cluster 1”, was formed by the sequences which were available prior to the present study, found in the gammaproteobacterial haloalkaliphilic strains. Majority of the strains used for the primer design belonged to this cluster, but no sequence of this group was obtained in this study. Another group named “cluster 2” included majority of OTUs obtained in this study, along with *Sulfuricella denitrificans* skB26^{T}. This group also encompassed ArxA sequences found in the genome of uncultured bacteria, including two gammaproteobacteria from a groundwater samples (26) and two betaproteobacteria from CO_2-driven geyser eruptions (27). Two OTUs obtained in this study were phylogenetically distinct from the other OTUs, and they formed “cluster 3”, along with *Azoarcus* sp. CIB isolated from a diesel fuel-contaminated aquifer (28). This cluster included a ArxA sequence coded in the genome of an uncultured bacterium belonging to *Gallionellales*, reconstructed from metagenomic data recovered from the groundwater sample (26).

2.4 Discussion

As apparent in Fig. 2-2, phylogeny of the *arxA* gene is not consistent with 16S rRNA gene-based phylogeny. As a significant example, *arx* gene cluster of *Sulfuricella denitrificans* skB26^T is located in a plasmid (29), and its closest relative has no *arx* genes in its genome (30). In this situation, it is almost impossible to predict presence of the *arx* genes from the 16S rRNA gene sequence, as is true with the *aio* and *arr* genes. In other words, *arx* gene-targeted analysis is the only way to explore diversity and distribution of organisms which have this notable gene.

With the primers newly designed in this study, various *arx*A gene sequences were successfully obtained (Fig. 2-2). Although nonspecific amplification was observed in some cases, non-target PCR products could be easily excluded just by size separation on agarose gels. These results indicate that specificity of the designed primer pair is practically sufficient. For the reason mentioned above, there is no way to make connections between the OTUs of *arxA* gene and those of the other genes. However, the sequences of 16S rRNA and *aprA* provided some information about the samples analyzed. For instance, results of these genes suggested that all DNA samples contained aerobic and anaerobic bacteria. They also indicated that all three samples had specific bacterial community structures distinct from the others. This dissimilarity might have been reflected in occurrence pattern of *arxA* OTUs, in which majority of the OTUs were detected in only one of the samples (Fig .2-2).

The *arxA* sequences detected in this study were phylogenetically distinct from those of alkaliphilic strains, belonging to cluster 1. They are more closely related to *Sulfuricella denitrificans* skB26^T (cluster 2) or *Azoarcus* sp. CIB (cluster 3), both isolated from samples characterized by neutral pH (31, 32). They are also related to *arxA* of uncultured bacteria

recovered from groundwater and geyser eruptions, characterized by nearly neutral pH (26, 27). These phylogenetic proximity might have reflected similarity in physicochemical characteristics of the sample analyzed, low salinity and nearly neutral pH (Table 2-2). On the other hand, they are phylogenetically distant from cluster 1, encompassing the majority of previously known ArxA. The cluster 1 exclusively consists of alkaliphiles and halophiles, and may represent only a specific lineage of the enzyme advantageous for organisms living in harsh habitats with high pH and salinity.

In this study, the environmental samples were analyzed with the primer GT-arxA_deg_490R which has extra bases to enhance efficiency of TA cloning. The additional bases potentially cause PCR bias which may affect the composition of the clone libraries obtained, because this modification was made without any consideration for the gene sequences (Table 2-3). Therefore, use of the primer arxA_deg_490R without additional GT may be more suitable in case that such bias should be minimalized to describe accurate community structure. In PCR-based analyses targeting multiple sequences, however, PCR bias cannot be fully excluded anyway. On the other hand, TA cloning is a very convenient and powerful tool to analyze large numbers of samples promptly. Considering these facts, the combination of PCR with GT-arxA_deg_490R and following TA cloning still has certain advantage over the other approaches.

2.5 References

1. **Mandal BK and Suzuki KT**. 2002 Arsenic round the world: A review. Talanta **58**:201–235.

2. **Shankar S, Shanker U, Shikha, Shankar S, Shanker U and Shikha**. 2014 Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation, Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation. Sci. World J. **2014**, **2014**:e304524.

3. **Zargar K, Conrad A, Bernick DL, Lowe TM, Stolc V, Hoeft S, Oremland RS, Stolz J and Saltikov CW**. 2012 ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. Environ. Microbiol. **14**:1635–1645.

4. **Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM and Oremland RS**. 2007 *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int. J. Syst. Evol. Microbiol. **57**:504–512.

5. **Richey C, Chovanec P, Hoeft SE, Oremland RS, Basu P and Stolz JF**. 2009 Respiratory arsenate reductase as a bidirectional enzyme. Biochem. Biophys. Res. Commun. **382**:298–302.

6. **Zargar K, Hoeft S, Oremland R and Saltikov CW**. 2010 Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J. Bacteriol. **192**:3755–3762.

7. **Kulp TR, Hoeft SE, Asao M, Madigan MT, Hollibaugh JT, Fisher JC, Stolz JF, Culbertson CW, Miller LG and Oremland RS**. 2008 Arsenic(III) Fuels Anoxygenic Photosynthesis in Hot Spring Biofilms from Mono Lake, California. Science (80-.). **1**:967–970.

8. **Hoeft McCann S, Boren A, Hernandez-Maldonado J, Stoneburner B, Saltikov C, Stolz J and Oremland R**. 2017 Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of *Ectothiorhodospira* from Mono Lake, California and Big Soda Lake, Nevada. Life **7**:1–14.

9. **Hernandez-Maldonado J, Sanchez-Sedillo B, Stoneburner B, Boren A, Miller L, Mccann S, Rosen M, Oremland RS and Saltikov CW**. 2017 The genetic basis of anoxygenic photosynthetic arsenite oxidation. Environ. Microbiol. **19**:130–141.

10. **Hollibaugh JT, Budinoff C, Hollibaugh R a., Ransom B and Bano N**. 2006 Sulfide oxidation coupled to arsenate reduction by a diverse microbial community in a soda lake. Appl. Environ. Microbiol. **72**:2043–2049.

11. **Hamamura N, Macur RE, Korf S, Ackerman G, Taylor WP, Kozubal M, Reysenbach AL and Inskeep WP**. 2009 Linking microbial oxidation of arsenic with detection and phylogenetic analysis of arsenite oxidase genes in diverse geothermal environments. Environ. Microbiol. **11**:421–431.

12. **Hamamura N, Fukushima K and Itai T**. 2013 Identification of antimony- and arsenic-oxidizing bacteria associated with antimony mine tailing. Microbes Env. **28**:257–263.

13. **Meyer-Dombard DR, Amend JP and Osburn MR**. 2013 Microbial diversity and potential for arsenic and iron biogeochemical cycling at an arsenic rich, shallow-sea hydrothermal vent (Tutum Bay, Papua New Guinea). Chem. Geol. **348**:37–47.

14. **Costa PS, Scholte LLS, Reis MP, Chaves A V., Oliveira PL, Itabayana LB, Suhadolnik MLS, Barbosa FAR, Chartone-Souza E and Nascimento AMA**. 2014 Bacteria and genes involved in arsenic speciation in sediment impacted by long-term gold mining. PLoS One **9**:1–12.

15. **Switzer Blum J, Hoeft McCann S, Bennett S, Miller LG, Stolz JR, Stoneburner B, Saltikov C and Oremland RS**. 2016 A microbial arsenic cycle in sediments of an acidic mine impoundment: Herman Pit, Clear Lake, California. Geomicrobiol. J. **33**:677–689.

16. **Hamamura N, Itai T, Liu Y, Reysenbach AL, Damdinsuren N and Inskeep WP**. 2014 Identification of anaerobic arsenite-oxidizing and arsenate-reducing bacteria associated with an alkaline saline lake in Khovsgol, Mongolia. Environ. Microbiol. Rep. **6**:476–482.

17. **Kojima H and Fukui M**. 2010 *Sulfuricella denitrificans* gen. nov., sp. nov., a sulfur-oxidizing autotroph isolated from a freshwater lake. Int. J. Syst. Evol. Microbiol. **60**:2862–2866.

18. **Watanabe T, Kojima H and Fukui M**. 2014 Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: Genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst. Appl. Microbiol. **37**:387–395.

19. **Watanabe T, Kojima H and Fukui M**. 2016 *Sulfuriferula thiophila* sp. Nov., a chemolithoautotrophic sulfur-oxidizing bacterium, and correction of the name sulfuriferula plumbophilus Watanabe, Kojima and Fukui 2015 to sulfuriferula plumbiphila corrig. Int. J. Syst. Evol. Microbiol. **66**:2041–2045.

20. **Kojima H, Watanabe T, Iwata T and Fukui M**. 2014 Identification of major planktonic sulfur oxidizers in stratified freshwater lake. PLoS One **9**:1–7.

21. **Watanabe T, Kojima H, Takano Y and Fukui M**. 2013 Diversity of sulfur-cycle prokaryotes in freshwater lake sediments investigated using *aprA* as the functional marker gene. Syst. Appl. Microbiol. **36**:436–443.

22. **Tamura K, Stecher G, Peterson D, Filipski A and Kumar S**. 2013 MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30**:2725–2729.

23. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF**. 2009 Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75**:7537–7541.

24. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O**. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

25. **Anisimova M and Gascuel O**. 2006 Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol **55**:539–552.

26. **Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS and Banfield JF**. 2016 Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. **7**:1–11.

27. **Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, Hug LA, Burstein D, Emerson JB, Thomas BC and Banfield JF**. 2017 Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. Environ. Microbiol. **19**:459–474.

28. **Martin-Moldes Z, Zamarro MT, del Cerro C, Valencia A, Gomez MJ, Arcas A, Udaondo Z, Garcia JL, Nogales J, Carmona M and Diaz E**. 2015 Whole-genome

analysis of *Azoarcus* sp. strain CIB provides genetic insights to its different lifestyles and predicts novel metabolic features. Syst. Appl. Microbiol. **38**:462–471.

29. **Watanabe T, Kojima H and Fukui M**. 2014 Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: Genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst. Appl. Microbiol. 2014: 10.1016/j.syapm.2014.05.010.

30. **Watanabe T, Kojima H and Fukui M**. 2015 Draft Genome Sequence of a Sulfur-Oxidizing Autotroph, Sulfuricella sp. Strain T08, Isolated from a Freshwater Lake. Genome Announc. **3**:8–9.

31. **Tsutsumi M, Iwata T, Kojima H and Fukui M**. 2011 Spatiotemporal variations in an assemblage of closely related planktonic aerobic methanotrophs. Freshw. Biol. **56**:342–351.

32. **Fernández-Llamosas H, Prandoni N, Fernández-Pascual M, Fajardo S, Morcillo C, Diaz E and Carmona M**. 2014 *Azoarcus* sp. CIB, an anaerobic biodegrader of aromatic compounds shows an endophytic lifestyle. PLoS One **9**:1–11.

Table 2-1. Mismatches between the primers arxA_Deg_F_B and arxA_Deg_R_B and target arxA gene sequences. Matching bases are shown as "-". The primers were designed on the basis of gene sequences from organisms shown in bold

Organism	arxA_Deg_F_B																									arxA_Deg_R_B																								
	5'												→												3'	3'													←												5'
	C	C	A	T	C	W	S	C	T	G	G	R	A	C	G	A	G	G	C	C	Y	T	S	G	S	A	A	G	G	C	G	G	G	G	A	T	G	T	T	G	W	T	G
Sulfuricella denitrificans skB26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	G	G	-
***Alkalilimnicola ehrlichii* MLHE-1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
***Halorhodospira halophila* SL1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
***Ectothiorhodospira* sp. PHS-1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Nitrincola lacisaponensis 4CA	-	-	-	-	-	-	-	A	-	-	-	-	-	T	-	-	A	-	-	G	-	-	T	-	T	-	-	-	-	-	T	-	-	-	-	-	T	-	-	T	-	-	T
Ectothiorhodospira sp. BSL-9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio thiocyanodenitrificans ARhD 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	A	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	T
Halomonas boliviensis LC1	-	-	-	-	-	-	-	T	-	-	-	-	-	T	-	-	-	-	-	A	-	-	A	-	T	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	T
Halomonas chromatireducens AGD 8-3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Halomonas sp. BC04	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Halomonas sp. A3H3	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio nitratireducens DMS 14787	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sulfidiphilus ALJ17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sp. AKL19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	A	T	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sp. ALMg11	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	A	T	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sp. ARh3	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G
Thioalkalivibrio sp. ALSr1	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	T	-	-	-	-	-	-	-
Thioalkalivibrio sp. ALR17-21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sp. AKL17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
Thioalkalivibrio sp. ALM2T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thiocapsa sp. KS1	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-
Azoarcus sp. CIB	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	C	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-
Magnetospirillum magnetotacticum MS-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	G	-	-	-	-	-	-	-	A	-	-	G	-	-	G	C	T	G	-	-	-	-	-	-	-
Rhodospirillales bacterium RIFCSPLOWO2 12 FULL 67 15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-
Candidatus Muproteobacteria bacterium RBG 16 65 34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	G	-	-	G	-	-	-	-	-	-	-	-	-
Gallionellales bacterium RIFCSPLOWO2 12 FULL 59 22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	A	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 58	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	A	-	-	G	-	-	-	A	-	-	-	A	-	-	-	-	-	-	G	C	T	-	-	-	-	-	-	-
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 13	-	-	-	-	-	-	-	G	-	-	-	-	-	T	-	-	-	-	G	-	-	-	A	-	-	-	A	T	-	-	-	-	A	-	-	-	-	-	-	-	C	-	-
Gammaproteobacteria bacterium RIFOXYD12_FULL_61_37	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	C	T	T	-	-	G	C	-	-
Gammaproteobacteria bacterium RIFOXYA12_FULL_61_12	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	C	T	T	-	-	G	C	-	-
Deltaproteobacteria bacterium GWC2_42_11	-	-	-	-	-	-	-	A	-	-	-	-	-	T	-	-	-	-	G	-	-	-	T	-	-	-	-	-	A	-	A	-	-	A	-	T	A	-	-	-	-	-	T
Betaproteobacteria bacterium CG2_30_68_42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	C	-	-	-	-	-	-	T	-	-	-	-	T	C	-	C	-	G	G	-
Betaproteobacteria bacterium CG2_30_59_46	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	A	-	-	-	-	-	T	-	-	-	-	T	-	-	C	-	G	G	-

.

Table 2-2 Physicochemical characteristics and number of clones in *arxA*, 16S rRNA, and *aprA* genes clone libraries of samples.

Sample	Temperature (°C)	pH	Cl^- (mM)	SO_4^{2-} (mM)	Conductivity (S/m)	Total As (mg/L)	Number of clones		
							arxA	16S	*aprA*
Mizugaki	5.1*	6.5*	1.6*	0.04*	0.04*	nd†	43	96*	54*
Masutomi	28.3*	6.3*	26.2	1.3	0.97*	3.26	61	65	44
Jozankei	42.6*	7.8*	40.2	0.9	0.56	3.20	42	60	51

* Data presented in the references, Kojima *et al.* 2009, 2014, 2017 and Watanabe *et al.* 2016.

† nd = Not determined.

Table 2-3. Mismatches between the primers designed in this study and target genes. The primers were designed on the basis of sequences from bolded organisms. Matching bases are shown as "-", and mismatches resulting from additional two bases (shown in lower-case letter) of GT-arxA_deg_490R are shaded.

Organism	arxA_deg_194F																			(GT)-arxA_deg_490R																			
	5'																		3'	3'																		5'	
	G	T	C	C	G	G	G	C	N	C	G	Y	T	G	C	G	G	N	A	T	A	T	G	C	C	V	G	C	R	C	C	K	A	C	C	C	C	t	g
***Sulfuricella denitrificans* skB26**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
***Alkalilimnicola ehrlichii* MLHE-1**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	t
***Halorhodospira halophila* SL1**	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	g
***Ectothiorhodospira* sp. PHS-1**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
***Nitrincola lacisaponensis* 4CA**	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Ectothiorhodospira sp. BSL-9	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	t
***Thioalkalivibrio thiocyanodenitrificans* ARhD 1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	t
***Halomonas boliviensis* LC1**	-	C	-	-	C	-	-	-	-	-	-	G	-	-	T	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	g
Halomonas chromatireducens AGD 8-3	-	C	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	g
***Halomonas* sp. BC04**	-	C	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	g	g
***Halomonas* sp. A3H3**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	g
***Thioalkalivibrio nitratireducens* DMS 14787**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
***Thioalkalivibrio sulfidiphilus* ALJ17**	-	T	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	c	t
***Thioalkalivibrio* sp. AKL19**	-	C	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
***Thioalkalivibrio* sp. ALMg11**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	-
***Thioalkalivibrio* sp. ARh3**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	-
***Thioalkalivibrio* sp. ALSr1**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	-
***Thioalkalivibrio* sp. ALR17-21**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-
***Thioalkalivibrio* sp. AKL17**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	-
***Thioalkalivibrio* sp. ALM2T**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Thiocapsa sp. KS1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	c	g
***Azoarcus* sp. CIB**	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Magnetospirillum magnetotacticum MS-1	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	t	a
Rhodospirillales bacterium RIFCSPLOWO2 12 FULL 67 15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Candidatus Muproteobacteria bacterium RBG 16 65 34	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Gallionellales bacterium RIFCSPLOWO2 12 FULL 59 22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 58	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 13	-	C	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	g	g
Gammaproteobacteria bacterium RIFOXYD12_FULL_61_37	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Gammaproteobacteria bacterium RIFOXYA12_FULL_61_12	-	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g
Deltaproteobacteria bacterium GWC2_42_11	-	-	-	-	C	-	G	-	A	-	-	G	-	-	T	-	-	-	-	-	T	-	-	-	-	A	-	G	-	-	-	-	-	-	-	-	-	t	g
Betaproteobacteria bacterium CG2_30_68_42	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	A	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	t
Betaproteobacteria bacterium CG2_30_59_46	-	C	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	c	g

Table 2-4 Amino acid sequences of Mo binding site in ArxA. Conserved amino acids are shaded in gray. For comparison, corresponding positions of ArrA, AioA and PsrA (thiosulfate reductase) are also shown.

Organism	Protein	Amino acid sequence									
Alkalilimnicola ehrlichii	ArxA	G	H	A	S	M	C	A	E	G	S
Ectothiorhodospira sp. PHS-1	ArxA	G	H	S	S	M	C	S	D	G	S
Thioalkalivibrio sulfidiphilus	ArxA	G	H	A	S	I	C	A	E	G	S
Thioalkalivibrio thiocyanodeni	ArxA	G	H	A	S	M	C	A	E	G	S
Halomonas boliviensis	ArxA	G	H	S	S	M	C	A	D	G	S
Halomonas sp. A3H3	ArxA	G	H	S	S	M	C	A	D	G	S
Halomonas sp. BC04	ArxA	G	H	A	S	M	C	A	E	G	S
Nitrincola lacisaponensis	ArxA	G	H	A	S	M	C	A	E	G	S
Ectothiorhodospira sp. BSL-9	ArxA	G	H	A	S	I	C	A	E	G	S
Thioalkalivibrio sp. AKL17	ArxA	G	H	A	S	I	C	A	E	G	S
Thioalkalivibrio sp. ALR17-21	ArxA	G	H	A	S	I	C	A	E	G	S
Thioalkalivibrio sp. AKL19	ArxA	G	H	A	S	M	C	A	E	G	S
Thioalkalivibrio sp. ALMg11	ArxA	G	H	A	S	M	C	A	E	G	S
Halorhodospira halophila	ArxA	G	H	S	S	M	C	S	D	G	S
Azoarcus sp. CIB	ArxA	G	H	S	S	M	C	S	D	G	S
Sulfuricella denitrificans	ArxA	G	H	S	S	M	C	S	D	G	S
Halomonas chromatireducens	ArxA	G	H	A	S	M	C	A	E	G	S
Thiocapsa sp. KS1	ArxA	G	H	S	S	M	C	S	D	G	S
Rhodospirillales bacterium RIFCSPLOWO2 12 FULL 67 15	ArxA	G	H	S	S	M	C	S	D	G	S
Candidatus Muproteobacteria bacterium RBG 16 65 34	ArxA	G	H	S	S	I	C	S	D	G	S
Gallionellales bacterium RIFCSPLOWO2 12 FULL 59 22	ArxA	G	H	S	S	I	C	S	D	G	S
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 58	ArxA	G	H	S	S	I	C	A	D	A	S
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 13	ArxA	G	H	S	A	T	C	A	D	A	S
Gammaproteobacteria bacterium RIFOXYD12 FULL 61 37	ArxA	N	H	S	S	M	C	S	D	A	S
Gammaproteobacteria bacterium RIFOXYA12 FULL 61 12	ArxA	N	H	S	S	M	C	S	D	A	S
Magnetospirillum magnetotactica	ArxA	G	H	S	S	L	C	S	D	A	S
Deltaproteobacteria bacterium GWC2 42 11	ArxA	G	H	A	S	M	C	S	E	G	S
Bacillus selenitireducens	ArrA	S	H	S	S	I	C	A	E	S	E
Shewanella sp. ANA-3	ArrA	S	H	S	S	V	C	A	E	A	H
Halarsenatibacter silvermanii	ArrA	S	H	S	S	I	C	A	E	A	E
Alkaliphilus oremlandii	ArrA	S	H	S	S	I	C	A	E	A	E
Chrysiogenes arsenatis	ArrA	S	H	S	A	I	C	A	E	V	E
Sulfuritalea hydrogenivorans	ArrA	S	H	S	A	I	C	A	E	A	E
Wolinella succinogenes	PsrA	G	H	E	S	T	C	P	L	A	Y
Salmonella enterica	PsrA	T	H	A	S	T	C	P	A	G	K
Agrobacterium tumefaciens	AioA	I	H	N	R	P	A	Y	N	S	E

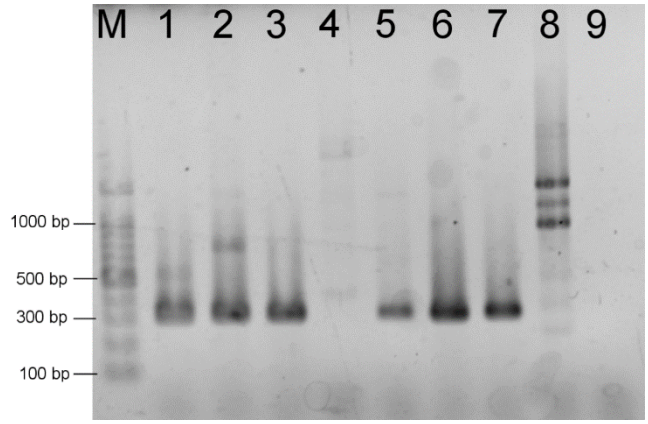

Figure 2-1. Gel image of PCR products obtained with the primer pair arxA_deg_194F and GT-arxA_deg_490R. Lanes: M, 100 bp DNA ladder as a size marker; 1, *Halorhodospira halophile* SL1; 2, *Thioalkalivibrio thiocyanodenitrificans* ARhD1; 3, *Thioalkalivibrio nitratireducens* DMS 14787; 4, *Halomonas boliviensis* LC1; 5, *Nitrincola lacisaponensis* 4CA; 6, *Alkalilimnicola ehrlichii* MLHE-1^{T}; 7, *Sulfuricella denitrificans* skB26^{T}; 8, *Sulfuritalea hydrogenivorans* sk43H (strain without *arxA* gene); 9, negative control (no template DNA).

Table 2-5 Number of clones and OTUs belonging to major phyla, detected in the clone libraries of the 16S rRNA gene. The numbers are shown in the order as clones/OTUs. Detailed taxonimc affiliations of all OTUs are provided in supplementary Table 2-6.

-

	Total	*Proteobacteria*	*Cyanobacteria*	*Bacteroidetes*	*Chloroflexi*
Mizugaki	96/31	78/20	0/0	6/3	0/0
Masutomi	65/30	45/17	0/0	1/1	2/2
Jozankei	60/34	8/6	23/11	17/8	6/3

Table 2-6. Taxonomic affiliation of OTUs detected in clone libraries of the 16S rRNA gene. OTU names are prefixed with the corresponding clone libraries. Taxa of the lowest level identified by the RDP Classifier (confidence threshold of 80%) are indicated.

OTU	No .of clone	Taxa	OTU	No .of clone	Taxa	OTU	No .of clone	Taxa
Mizugaki_01	14	*Rhodoferax*	Masutomi_01	8	*Gallionella*	Jozankei_01	6	*Cyanobacteria*
Mizugaki_02	12	*Sulfuritalea*	Masutomi_02	6	*Sulfuricellaceae*	Jozankei_02	6	*Cytophagales*
Mizugaki_03	10	*Gallionella*	Masutomi_03	5	*Desulfocapsa*	Jozankei_03	6	GpIV
Mizugaki_04	6	*Desulfatirhabdium*	Masutomi_04	5	Unidentified	Jozankei_04	4	*Chloroflexia*
Mizugaki_05	6	*Methylobacter*	Masutomi_05	3	Unidentified	Jozankei_05	4	*Terrimonas*
Mizugaki_06	5	*Sideroxydans*	Masutomi_06	3	*Gallionella*	Jozankei_06	2	*Azospira*
Mizugaki_07	4	*Polymucleobacter*	Masutomi_07	3	*Sulfuricurvum*	Jozankei_07	2	*Cyanobacteria*
Mizugaki_08	4	*Ferruginibacter*	Masutomi_08	3	*Gallionella*	Jozankei_08	2	*Rhodocyclaceae*
Mizugaki_09	3	*Desulfurivibrio*	Masutomi_09	3	*Sideroxydans*	Jozankei_09	2	*Saprospiraceae*
Mizugaki_10	3	*Sulfurimonas*	Masutomi_10	3	*Syntrophaceae*	Jozankei_10	2	*Cyanobacteria*
Mizugaki_11	3	*Gemmatimonas*	Masutomi_11	2	*Desulfurivibrio*	Jozankei_11	1	*Dechloromonas*
Mizugaki_12	3	*Methylotenera*	Masutomi_12	2	*Rhodoferax*	Jozankei_12	1	*Gammaproteobacteria*
Mizugaki_13	2	*Geothrix*	Masutomi_13	2	Unidentified	Jozankei_13	1	*Cyanobacteria*
Mizugaki_14	2	*Gallionella*	Masutomi_14	1	Unidentified	Jozankei_14	1	*Anaerolineaceae*
Mizugaki_15	2	Unidentified	Masutomi_15	1	*Betaproteobacteria*	Jozankei_15	1	GpXIII
Mizugaki_16	2	*Ferribacterium*	Masutomi_16	1	*Deltaproteobacteria*	Jozankei_16	1	*Cytophagales*
Mizugaki_17	1	*Desulfobulbaceae*	Masutomi_17	1	*Caldisericum*	Jozankei_17	1	*Cyanobacteria*
Mizugaki_18	1	*Legionella*	Masutomi_18	1	*Geobacter*	Jozankei_18	1	*Rhodobacteraceae*
Mizugaki_19	1	*Methyloparacoccus*	Masutomi_19	1	Unidentified	Jozankei_19	1	*Bacteroidetes*
Mizugaki_20	1	Unidentified	Masutomi_20	1	*Bacteroidetes*	Jozankei_20	1	*Verrucomicrobia*
Mizugaki_21	1	*Methylococcaceae*	Masutomi_21	1	*Verrucomicrobia*	Jozankei_21	1	*Flavobacteriales*
Mizugaki_22	1	*Spartobacteria_genera_incertae_sedis*	Masutomi_22	1	*Methylobacter*	Jozankei_22	1	*Acidobacteria* _Gp4
Mizugaki_23	1	*Sterolibacterium*	Masutomi_23	1	*Gammaproteobacteria*	Jozankei_23	1	*Armatimonadetes* _gp5
Mizugaki_24	1	*Victivallis*	Masutomi_24	1	*Aridibacter*	Jozankei_24	1	*Rehaibacterium*
Mizugaki_25	1	Unidentified	Masutomi_25	1	*Sulfurimonas*	Jozankei_25	1	*Verrucomicrobiaceae*
Mizugaki_26	1	*Sulfuritalea*	Masutomi_26	1	*Anaerolineaceae*	Jozankei_26	1	*Armatimonadetes*
Mizugaki_27	1	*Oligosphaera*	Masutomi_27	1	Unidentified	Jozankei_27	1	*Cyanobacteria*
Mizugaki_28	1	*Geobacter*	Masutomi_28	1	*Methylococcaceae*	Jozankei_28	1	*Bacteroidetes*
Mizugaki_29	1	*Bacteroidetes*	Masutomi_29	1	*Anaerolineaceae*	Jozankei_29	1	*Opitutus*
Mizugaki_30	1	*Methylobacter*	Masutomi_30	1	*Ignavibacterium*	Jozankei_30	1	GpIV
Mizugaki_31	1	*Prolixibacter*				Jozankei_31	1	*Cyanobacteria*
						Jozankei_32	1	*Cytophagaceae*
						Jozankei_33	1	*Cyanobacteria*
						Jozankei_34	1	*Anaerolineaceae*

Table 2-7. OTUs of the *aprA* gene detected in the clone libraries and their closest cultured relatives.

OUT	No of clones in each library			Closest cultured relative		
	Mizugaki	Masutomi	Jozankei	Species name	Identity (%)	Function
apr_OTU01	0	0	20	*Thermodesulfobacterium commune*	93	Sulfate reduction
apr_OTU02	8	7	0	*Desulfobulbus elongatus*	94	Sulfate reduction
apr_OTU04	14	0	0	*Desulfatitalea tepidiphila*	90	Sulfate reduction
apr_OTU08	8	0	0	*Desulfocapsa thiozymogenes*	93	Sulfate reduction
apr_OTU09	0	7	0	*Thermodesulfovibrio* sp. (ex DSM 12270)	72	Sulfate reduction
apr_OTU10	0	4	0	*Desulfobulbus elongatus*	92	Sulfate reduction
apr_OTU12	0	0	4	*Desulfocaldus* sp. Hobo	97	Sulfate reduction
apr_OTU20	0	0	2	*Desulfatirhabdium butyrativorans*	97	Sulfate reduction
apr_OTU22	0	0	1	*Desulfovibrio termitidis*	97	Sulfate reduction
apr_OTU24	1	0	0	*Desulfobacterium indolicum*	95	Sulfate reduction
apr_OTU25	1	0	0	*Desulfovibrio fructosivorans*	81	Sulfate reduction
apr_OTU27	1	0	0	*Desulfobacterium catecholicum*	92	Sulfate reduction
apr_OTU29	1	0	0	*Desulfobulbus elongatus*	93	Sulfate reduction
apr_OTU30	1	0	0	*Desulfobulbus rhabdoformis*	92	Sulfate reduction
apr_OTU31	1	0	0	*Thermodesulfovibrio* sp. (ex DSM 12270)	79	Sulfate reduction
apr_OTU03	0	0	14	*Sulfuriferula multivorans*	92	Sulfur oxidation
apr_OTU05	0	8	0	*Candidatus* Pelagibacter ubique	84	Sulfur oxidation
apr_OTU06	0	8	0	*Sulfuriferula multivorans*	94	Sulfur oxidation
apr_OTU07	8	0	0	*Sulfuritalea hydrogenivorans*	98	Sulfur oxidation
apr_OTU11	0	0	4	*Sulfuricaulis limicola*	94	Sulfur oxidation
apr_OTU13	0	4	0	*Thiothrix* sp. 12730	97	Sulfur oxidation
apr_OTU14	0	4	0	*Sulfuricaulis limicola*	94	Sulfur oxidation
apr_OTU15	0	0	4	*Sulfuriferula multivorans*	94	Sulfur oxidation
apr_OTU16	3	0	0	*Sulfuriferula multivorans*	87	Sulfur oxidation
apr_OTU17	2	0	0	*Sulfuriferula multivorans*	87	Sulfur oxidation
apr_OTU18	0	2	0	*Sulfuriferula multivorans*	91	Sulfur oxidation
apr_OTU19	2	0	0	*Sulfuriferula multivorans*	90	Sulfur oxidation
apr_OTU21	1	0	0	*Sulfuricaulis limicola*	97	Sulfur oxidation
apr_OTU23	0	0	1	*Sulfuricaulis limicola*	91	Sulfur oxidation
apr_OTU26	1	0	0	*Sulfuricaulis limicola*	81	Sulfur oxidation
apr_OTU28	1	0	0	*Sulfuriferula multivorans*	87	Sulfur oxidation

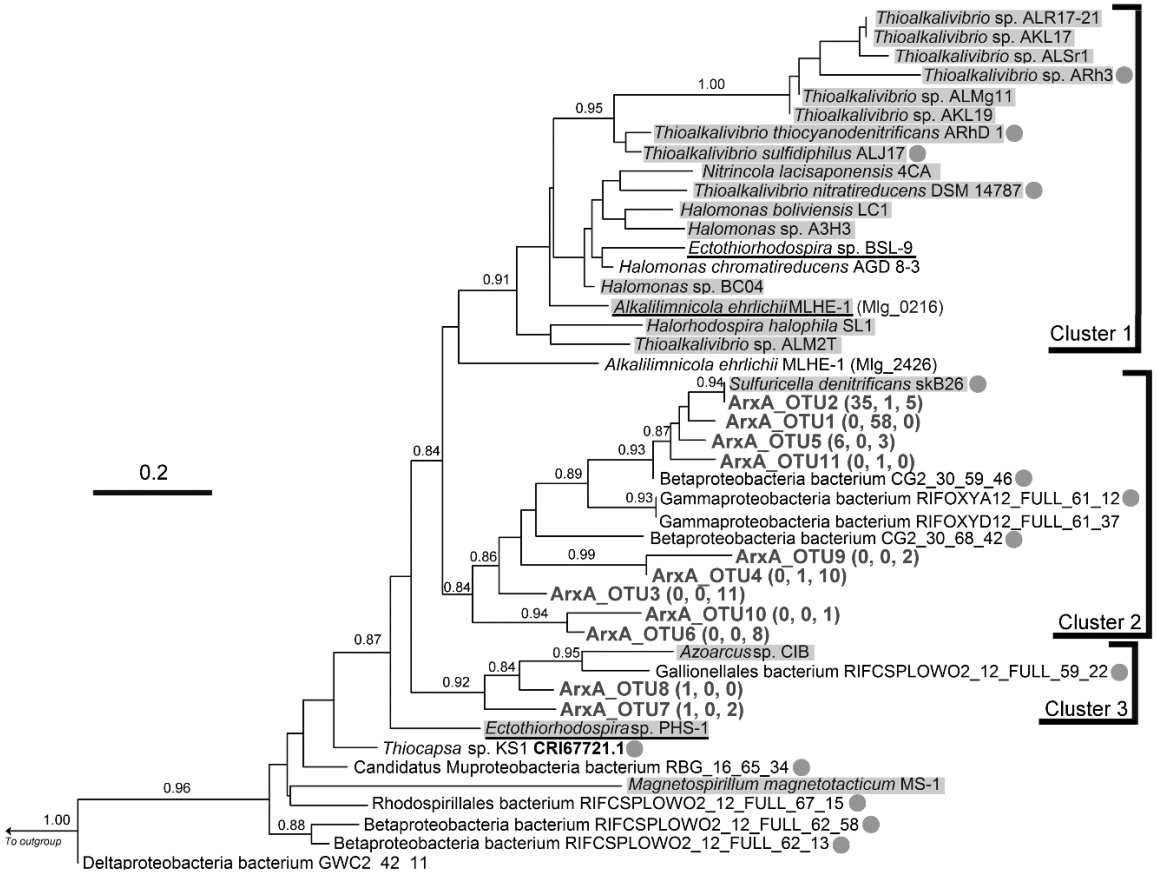

Figure 2-2. Phylogenetic tree of proteins encoded by the *arxA* genes. The sequences used to design new primers are show in shaded boxes. For the strains labeled with underline, involvement of *arxA* in anaerobic arsenic oxidation has been experimentally demonstrated. The OTUs obtained in this study are show with numbers in parentheses, which represent number of clones in the libraries of Mizugaki, Masutomi, and Jozankei, respectively. Gray circles indicate organisms which have the *aprA* gene. The aLRT branch support larger than 0.8 are indicated above the branches. Two archaeal proteins from *Candidatus* Methanoperedens, included in the analysis as outgroup, are not displayed in this figure.

Chapter 3

Phylogeny and genetic diversity of *arx*A gene and *arx* operon-like gene cluster.

3.1 Introduction

Arsenic, despite its toxic nature, it is known to be used as an energy substrate by microorganisms. Mechanism of microbial transformation of arsenic in the environment includes arsenate As(V) reduction (use as electron acceptor), and arsenite As(III) oxidation (use as an electron donor), in an energy-generating process, defined as "arsenothrophy" (1). The interconversion of arsenate and arsenite is catalyzed by the arsenate reductase ARR, and two phylogenetically distant arsenite oxidases, AIO and ARX (2). The latter correspond to a new group of arsenite oxidases enzyme first identified in the chemoautotrophic strain *Alkalilimnicola ehrlichii* MLHE-1^T.

A. ehrlichii MLHE-1^T was isolated from the anoxic bottom water of Mono Lake, California, able to grow with As(III) as electron donor and nitrate as its electron acceptor (3). Unsuccessful amplification, and absence of an *aio* cluster after inspection of the genome suggest the presence of another enzyme responsible for the arsenite oxidation activity (4). Proteomic analysis, mutation and gene expression experiments led to uncovering a new type of arsenite oxidase enzyme, named ARX, belonging to the DMSO reductase family, and phylogenetically related to the ArrA arsenate reductase rather than the AioA arsenite oxidase (4–6).

Another exciting discovery was the presence of an *arxA* gene in the genome of *Ectothiorhodospira* sp PHS-1 (2), isolated from a photosynthetic arsenite-oxidizing red-pigmented biofilm, found within a hot spring of Paoha Island in Mono Lake, CA (7). The genome of *Ectothiorhodospira* sp. PHS-1^T lacked of *aio* genes while harboring homologous *arx* genes as in *A. ehrlichii* MLHE-1^T, suggesting an arsenite oxidation activity mediated by the Arx enzyme (2). *Ectothiorhodospira* sp. strain BSL-9, isolated from Big Soda Lake, NV,

was capable to oxidize the arsenite coupled to anoxigenic photosynthesis (photoarsenotrophy) (8). A couple more *arx* arsenite-oxidizing organism have been reported, one of them, *Desulfotomaculum* sp. TC-1, isolated from a sulfidic hot spring in China (9).

Additionally, *arxA*-like gene fragments have been identified in alkaline –saline places (2, 10) and non-saline environments sites through clone library (11), also within metagenomics sequence data (8); and even, there has been detected *arx* transcripts as well (8, 12). These reports support the notion of a wide distribution of *arx* genes, and may also indicate active *arx* arsenite-oxidizing microbes.

Initial inspection of the genomes of *arx*-containing strains, *A. ehrlichii* MLHE-1^T and *Ectothiorhodospira* sp. PHS-1 revealed the presence of a set of genes predicted to encode for the structural components of the putative ARX arsenite oxidase, *arxB2ABCD*; and regulatory proteins *arxXRS* (2). Disruption of the *arxA* gene in the strain *A. ehrlichii* MLHE-1^T help to verify the requirement of the *arxA* gene in the chemoautotrophic arsenite oxidation by the MLHE-1^T strain (6). Likewise, the lack of the expression of *arxB2ABCD* genes and inhibition of the arsenite oxidation activity after arsenite induction in the *Ectothiorhodospira* sp. BSL-9 *arxA*-disrupted mutant provided the genetic evidence for photoarsenotrophy (8).

The continuous growth of the public databases are a useful source of information and can contribute to study the presence, diversity, and organization of *arx* homologous genes. In this sense, the purpose of this work was to explore the taxonomic and environmental distribution of *arx* homologous gene cluster in prokaryotic genomes. The sequences analysis center on the putative functional coding genes *arxAB*, and the sequences variations found in *arxA*. Then, it was compared the *arxA* genes cluster with other molybdoenzymes, and the results were used to analyze their phylogenetical relationship within the DMSO reductase family. The present work

highlights the genomic occurrence of *arx*-like gene cluster within the prokaryotic group, and the environmental distribution of two novel sequence variants of *arxA*-type arsenite oxidase.

3.2 Materials and Methods

3.2.1 Detection and recovery of homologous *arx* gene cluster

The translated sequences of the *arxXRSB'ABCD* gene cluster of *Alkalilimnicola ehrlichi* MLHE-1^T was used as a chosen model to search for homologous protein in the NCBI-NR database (downloaded in January of 2018), by using Diamond v 0.9.14 (13). A gene was considered to be a homolog if it had an amino acid identity of ≥ 30%, a spanning alignment of ≥60% of the length of the query, and an e-value ≤1e^{-3}. Detection of homologous sequences belonging to the same organism genome was grouped and used for subsequent analysis.

Good quality assembled genomes without protein annotation were downloaded from the NCBI genomes repository. It was considered good quality genomes those consisting of contigs ≤ 300, have an N50 ≥20 kb and containing ≤10 kb of ambiguous base pairs. Protein coding regions were established for each genome using Prodigal v2.6.3 in normal mode (with default parameters) (14), and the predicted proteins were used as a custom database to identify homologous sequences by using Diamond v 0.9.14 as it was explained before.

Additionally, it was included in the analysis the available genomes of three isolated strains from Jozankei Hot spring: *Sulfuritortus calidifontis* J1A^T (15), strain J5B isolated by Profesor Kojima, and strain M52, since partial *arxA* gene were amplified or detected in their respective genomes (11). Protein coding regions for each genome were predicted using Prodigal v2.6.3,

and comparison to the *arx* genes of *A. ehrlichii* MLHE-1^{T} was also performed as described before.

For those genomes lacking of a particular gene belonging to the arx gene cluster, the specific genome was manually analyzed to confirm its absence, or on the contrary its presence.

3.2.2 Phylogenetic distribution of the genomes harboring *arx*-like gene cluster

Several of the assemblies lacked 16S rRNA gene; therefore the phylogenetic analysis was performed using ten ribosomal proteins (RP) consistently present in all the genomes (L2 –L4, L13, L23, S2, S4, S9, S10, and S11). The ribosomal proteins have been widely used for phylogenetic analyses since are steadily conserved in all three domains (16–18). Detail information about each RP and amino acid sequences used in the alignment can be found in table 3-1. Each RP was independently aligned using MAFFT (19) with default parameters. The poorly aligned regions were removed using the Gblocks v0.91b program (20), allowing for a gap if it was located in less than 50% of the sequences. Individuals RP alignments were concatenated using MEGA version 6 (21). The resulting sequence alignment contained a total of around 1461 aligned positions in the final dataset. Maximum-likelihood and neighbor-joining phylogeny were carried out by the online program PhyML (v3.0)(22), with LG substitution model; and MEGA 6 with p-distance model(21), respectively. The gamma shape parameter was estimated directly from the data. Support values were calculated using the aLRT test (SH-Like) (23), and 1000 bootstrap, for ML and NJ, respectively.

3.2.3 Sequence and Phylogenetic Analysis of *arxA* and *arxB* homologous genes

The analysis was performed on the Phylogeny.fr platform (24) and comprised the following steps. The homologous translated sequences from *arxA* and *arxB* gene dataset were aligned separately with MUSCLE online program with the find diagonal parameter enabled (25). Ambiguous and poor aligned regions were removed using the Gblocks v0.91b tolerating gaps if it were located within an appropriate block in less than 50% of the sequences. The trimmed alignments were used to infer the ArxA and ArxB phylogenetic tree by the neighbor-joining algorithm with the distance criterion in MEGA 6, and maximum-likelihood algorithm with LG + G + I substitution model using the online program PhyML (v3.0). The gamma shape parameter was estimated directly from the data. Bootstrap and the aLRT test (SH-Like) were used to calculate the support values depending on the method.

A co-phylogenetic tree was made with the previously inferred ArxA and ArxB individuals trees by the *mirrortree* approach (26, 27). Complementary evidence of evolutionary protein–protein interactions was assessed by the linear correlation between the distance matrices of the aforementioned phylogenetic trees, as it was proposed by the *mirrortree* method (28). Identification of the putative protein domain in the deduces amino acid sequences of the genes were searched by Pfam (29) (Batch sequence search, https://pfam.xfam.org/). The conserved residues in the multiple sequences alignments were detected and visualized by Jalview Version 2 (30). Detections of the presence and location of twin-arginine signal peptide cleavage sites were made with TatP 1.0 server online (31). Additional *arx* gene trees were inferred separately using their predicted amino acid sequences using the same parameters mentioned before.

Additionally, a separated phylogenetic analysis was made between the predicted protein of the *arxA* gene sequences, and the subunit A of some molybdenum enzymes from other clades within the DMSO reductase family. The protein multi-sequence alignment was trimmed and

gaps were allowed within an appropriate block by Gblocks v0.91b. The reconstruction of the evolutionary tree was made with the distance criterion for tree searching using the BioNJ algorithm and 1000 bootstraps replicates (32), employing FastME 2.0 online software with default parameters (33).

The final trees were visualized and annotated using the package ggtree (version 1.10.4) (34) and phytools (version 0.6-44) (35) within R software (version 3.4.3).

3.2.4. Clone library construction and phylogenetic analysis of *arxA* sequences with insert

A set of degenerate primer was developed using as a template the *arxA*-sequence variant with a long insertion of the following organism: *Sulfuricella denitrificans* skB26^{T}, *Sulfuritortus calidifontis* J1A^{T}, *Gammaproteobacteria* RIFOXYD12 FULL_61_37, *Gammaproteobacteria* RIFOXYA12 FULL_61_12, and *Betaproteobacteria* CG2_30_59_46. The binding primer location within the *arxA* sequence took into account the conserved amino acid regions flanking the insertion. The degenerate primers were arxA_G2_F (AARCGTACCAAYCCSAAVAAGG) and arxA_G2_R (GTTCTTGGCGTAGTCRTCCAT). The PCR reaction was performed in 25 µl volume reaction mixtures, containing 0.5 µmol l−1 of each primer, 1× Ex Taq Buffer (Takara, Shiga, Japan), 0.2 mmol l−1 dNTPs (Takara), 0.625 U of Ex Taq (Takara), 3% DMSO and template DNA solution. PCR conditions were carried out as follow: 94°C for 3 min; 34 cycles of 94°C for 45s, annealing temperature at 55°C for 45s, extension at 72°C for 1.5 min, and then a final extension step at 72°C for 7 min. The primer specificity was confirmed by amplifying the partial *arxA* sequences of *Sulfuricella denitrificans* skB26^{T}, the first reported strain holding an insertion within the *arxA* gene (36); and *Sulfuritortus calidifontis* J1A^{T}.

Two DNA samples were used to look for this type of *arxA*-like sequence with insertion. One water sample obtained at 40 m depth in Lake Mizugaki (Yamanashi Prefecture, Japan), and a dark green microbial mat sample recovered in Jozankei Hot spring situated in Hokkaido. DNA extraction procedure for Mizugaki sample is described more in detail by Watanabe et al. (37), and for Jozankei sample can be found in chapter 2.

The PCR product was run on a 1.1% agarose gel. Bands corresponding to the expected product size (~ 1200 bp of sequence with insert; and ~ 900 bp sequence without insert) were cut, and the DNA was extracted from the gel with QIAquick Gel Extraction Spin Kit (Qiagen, Valencia, CA). The amplicons were cloned to pCR2.1-TOPO vector, and transformed into TOP10 cells (TOPO TA Cloning Kit, Invitrogen Carlsbad, CA), according to the manufacturer's instructions. Colonies carrying the expected size insert were selected and sequenced.

The partial *arxA* sequences of the positive clones were translated to amino acid in MEGA 6, and aligned using MAFFT program with default parameters. A pairwise distance matrix was calculated based on Poisson model using MEGA, employed to define operational taxonomic units (OTUs) at a cutoff value of 0.02 in Mothur (38). The most abundant sequence of each OTU was selected as a representative sequence.

Lastly, the OTUs and the putative ArxA sequences of previous steps were aligned and used to make two phylogenetic trees, with and without the partial *arxA* sequences clones reported in earlier studies by Zargar et al. (2) and Hamamura et al. (10). The alignment was carried out using MUSCLE online program and trimmed manually, excluding gaps. The distance criterion was used to construct a Neighbor-joining tree with the BioNJ algorithm with 1000 bootstraps replicates (32), employing FastME 2.0 online software with default parameters (33). The

arsenate reductase ArrA amino acid sequences of *Alkaliphilus oremlandii* OhILAs and *Halarsenatibacter silvermanii* were used as the outgroup.

3.3 Results

3.3.1 The homologous *arx* gene cluster are mainly detected in proteobacterias organism

A total of 201 sequences met the criteria for homology to the *arxA* gene sequence of *A. ehrlichii* MLHE-1^T with 30% of identity or more, at least 60% alignment coverage, and e-value less than 0.001. From this group, 46 genomes were selected to continue the analysis due to the detection of homologous sequences in the majority, if not all the *arx* genes cluster (*arxXSRB'ABCD*), with identity above 44%, coverage around 99.8%, and e-value below 9.2e-203 (Table 3-2). Manually revision of each bacterial organism genomes allow us to confirm the presence of the *arx* homologous genes, and in some cases identify other that may not have been detected due to low similarity with that of *A. ehrlichii* MLHE-1^T. The revision also led us to exclude from the analysis six assemblies, four of them having identical sequences to other two genomes kept as representatives (Table 3-3), another one due to the localization of genes in different scaffolds, and the last one because of an assembly gap within the *arxA* homologous gene. At this point, 40 assemblies, corresponding to different prokaryotic organisms, were used to continue the analysis (Table 3-3).

Unannotated genomes were also used in the study, considering it may also constitute an excellent source of information to explore for *arx* gene homologous cluster. At the time of the analysis, the unannotated assemblies from the bacterial subdirectory comprehend around 11000 genomes. Hereupon, 5450 good quality assembled genomes were downloaded which

met the criteria of contigs below 300, N50 above 20 kb and less than 10 kb of ambiguous base pairs. From this group, only three genomes were found to possess an *arxA* homologous sequence above 44% of identity, with a coverage of 99.8% and e-value lower than 8.40E-204. The results also revealed the presence of other homologous sequences to the rest of *arx* gene cluster, similar to those of *A. ehrlichii* MLHE-1^T (Table 3-2). A complete set of *arx* homologous genes were also detected in the genomes of *Sulfuritortus calidifontis* J1A^T, strains J5B and M52, isolated from Jozankei, with an amino acid identity above 50%, and e-value lower than 2.6 E-290 (Table 3-2).

At the end of the analysis, 46 complete assemblies (41 belonging to bacteria and 5 to archaea) were found to have *arx*-like gene cluster, and description of the organisms are presented in Table 3-3. These 46 sequences included *Ectothiorodospira* sp. PHS-1 and *Halorhodospira halophila* SL1, previously used to define the ArxA clade (2), as well as *Ectothiorodospira* sp. BSL-9, first genetically validated ARX photoasernotroph (8, 39).

All the bacteria were associated to the phylum *Proteobacteria* (Fig. 3-1). The majority of the proteobacteria organism were members of the class *Gammaproteobacteria*, most of them (15 assemblies) from the order *Chromatiales,* family *Ectothiorhodospiraceae*. This group includes mostly but not exclusively organism related to the genus *Thioalkalivibrio*; as well as *Alkalilimnicola ehrlichii* MLHE-1^T, the first reported *arx* containing strain, and the also reported *arx* harboring strains *Halorhodospira halophile* SL1 and *Ectothiorhodospira* sp. strains PSH-1 and BSL9. Six other assemblies were affiliated to the order *Oceanospirillales*, including members of the genus *Halomonas* and *Marinospirillum* (Fig. 3-1).

The majority of the genomes of gammaproteobacterial organisms were derived from an isolated strain, mostly from hypersaline, alkaline environments known as Soda lakes dispersed

thorough the world (Table 3-3). Only two genomes were derived from a microbial metagenome from a groundwater sample (USA) (Table 3-3).

The class *Betaproteobacteria* was the second group with more number of assemblies harboring a homologous sequences to the *arx* gene cluster. The genomes of the betaproteobacterial organisms were derived from isolated strains, and also from microbial metagenomes, recovered in various distinct samples sites, e.g., lake, groundwater, geyser, aquifer and hot spring, some of which reported low pH (Fig 3-1, Table 3-3).

The remaining bacterial assemblies belonged to two organisms affiliated to the class *Alphaproteobacteria*, one to the class *Deltaproteobacteria*, and one representative of a new lineage of proposed name *Candidatus* Muproteobacteria (Fig 3-1).

The remaining five assemblies belonged to a novel archaeal lineage of proposed name *Candidatus Methanoperedens*, within the phylum *Euryarchaeota*, order *Methanosarcinales.* The majority of these assemblies were derived of metagenomics data obtained from anaerobic bioreactors performing methane oxidation coupled to nitrate reductions in different laboratories (Table 3-3).

3.3.2 Conservation and organization of the *arxB'ABCD* and *arxXSR*-like gene cluster

Representation of the *arx* gene cluster and gene arrangement for each genome are shown in the figure 3-2. The majority of the assemblies contain a complete set of *arx*-like genes, and 17 of them were identical to the *arx* operon arrangement found in *A. ehrlichii* MLHE-1^{T}.

The complete homologous genes coding for the structural components of the putative ARX arsenite oxidase, *arxB'ABCD*, were present in 34 of the 46 assemblies. Two genomes appear to lack of *arxB'* and *arxD*-like genes while harboring a partial cluster of homologous *arxABC*

sequences. The *arxB'*-like gene was absence in all the archaea organisms affiliated as *Candidatus Methanoperedens*. On the contrary, the absence of only an *arxD*-like gene, while having a homologous to the *arxB'* gene was only observed in one assembly. Overall, at least the *arxABC*-like genes, coding for the putative molybdopterin oxidoreductase (subunit A), ferredoxin type protein (subunit B), and a quinol-oxidizing membrane protein subunit C), respectively, seems to be steadily conserved among all the genomes.

In the majority of the assemblies the gene placement within the homologous *arx* operon were also consistent with that of *A. ehrlichii* MLHE-1^{T}. The *arxB'* –like gene in the majority of cases was located immediately upstream and in the same direction to the *arxABCD* gene; except for *Candidatus* Muproteobacteria bacterium and strain M52, placed downstream. Additionally, when the genome harbors the *arxD*-like gene, it was positioned in all cases, downstream to the *arxB'ABC* genes, and on the same orientation (Fig. 3-2).

The *arx* operon of *A. ehrlichii* MLHE-1^{T} also included the accesories *arxXSR* gene, encoding for putative regulatory proteins. The *arxXSR*-like gene cluster was present in the majority of the organism assemblies. Particularly, these regulatory coding homologous genes were missing in the genomes of all the representatives of the order *Oceanospirillales.* Other genomes that didn't include the *arxXSR*-like genes were *Halorhodospira halophile* SL1, *Candidatus* Muproteobacteria bacterium, Rhodospirillales bacterium and *Candidatus* Methanoperedens sp. BLZ2 (Fig. 3-2).

Whether the *arxXSR* genes were present in the homologous *arx* operon, it appears in the majority of cases to be placed upstream and in the opposite direction to the *arxB'ABCD* gene. Two exceptions were in the genomes of *Thioalkalivibrio* sp. ALMg11 and *Thioalkalivibrio* sp. AKL19, where it was localized downstream and on the same orientation to the ARX structural

coding genes. In case of the archaeal organisms, the regulatory genes were present upstream and in the same orientation to the homologous *arxABCD* genes (Fig. 3-2).

3.3.3 Comparative sequence analysis of *arxA* and *arxB*-like gene

Currently, ArxA and ArxB of *A. erlichii* MLHE-1^T constituted the experimentally demonstrated subunits necessary for arsenite oxidation activity after enzymatic assays and proteomic analyses (5). Comparison between the *arxA* and *arxB*-like translated sequences found in the assemblies to that of *A. ehrlichii* MLHE-1^T indicate shared characteristics.

The putative *arxA* gene product is predicted to code for the catalytic subunit. The examined multiple alignments of the putative ArxA protein revealed a molybdopterin cofactor domain. Two additional motifs were detected, an iron-sulfur [4Fe-4S] motif (C-X_2-C-X_3-C-X_{27-33}-C) starting around residues 53 to 61, and a twin-arginine signal sequences found at the N-terminal region (-R-R-X-F-L-K/Q), consistent also with the general characteristics of the ArxA sequence from *A. ehrlichii* MLHE-1^T. The TAT signal motif region was followed by a hydrophobic and weakly charged region, rich in alanine, valine, and glycine amino acids until the predicted signal peptide cleavage site designed as AXA in the majority of the sequences (Fig 3-3).

The suggested catalytic binding pocket sequences, GHSSMCSDG in *Ectothiorhodospira* sp. PHS-1^T and GHASMCAEG in *A. ehrlichii* MLHE-1^T (2), were also identified in several of the sequences, and some others exhibit a slightly modified version (G/N)-H-(A/S)-S-(I/M)-C-(A/S)-(D/E)-(A/G). A conserved cysteine residue predicted to coordinate to the molybdenum was observed in almost all the evaluated putative ArxA sequences, except in Deltaproteobacteria bacterium (Fig 3-3).

Analysis of the translated *arxB*-like gene sequences detected multiple iron-sulfur FeS-harboring clusters consistent with the beta subunit of the typical molybdoenzymes, and similarly to that of *A. ehrlichii* MLHE-1^T. At least three iron-sulfur [4Fe-4S] clusters coordinated by C-X_2-C-X_2-C-X_3-C; C-X_2-C-X_4-C-X_3 –CP and C-X_2-C-X_2 -C-X_3 -CP (2) were detected. In addition, a four [4Fe-4S] binding motif could also be identified based on the multi sequences alignment analysis and Pfam search, similar to what it is proposed for the beta subunit of ARR and DMSO reductases (40), with the only exception of *Candidatus* Methanoperedens sp. BLZ1 (Fig 3-4).

The ArxA arsenite oxidase was confirmed as a new clade within the molybdenum-containing enzymes family by Zargar et al. (2012)(2). The homologous *arxA*-like sequences found in the previous steps were subject to phylogenetic analysis with other oxidoreductases (DMSO) enzymes to verify their belonging to the ArxA clade (Fig. 3-5).

The phylogenetic tree reveals that only 41 of the 46 ArxA-like sequences were placed with *Alkalilimnicola ehrlichii* MLHE-1^T within the ArxA clade. The remaining five sequences corresponding to those of the *Candidatus Methanoperedens* organisms, annotated as formate dehydrogenase on the NCBI database, formed a separated cluster suggesting a diverging event from the ArxA clade. Moreover, they are not located within the formate dehydrogenase branches aggrupation in the phylogenetic tree, but forming a cluster positioned between the ArrA and ArxA clades. Further biochemical evidence is necessary to draw more conclusions about the *Candidatus* Methanoperedens cluster, and whether the predicted protein belong to the arsenic metabolism, or not.

Considering that *Candidatus* Methanopederens sequences were located outside the ArxA clade, they will not be acknowledged from this point forward as putative ArxA sequences, and instead will be treated as an outgroup in the following analysis.

3.3.4 Comparative phylogenetic analysis of ArxA and ArxB

Maximum-likelihood (ML), and neighbor-joining (NJ) phylogenies were reconstructed to gain insight into the evolutionary history of ArxA and ArxB (Fig. 3-6 and Fig. 3-7).

The correspondence between ArxA and ArxB phylogeny was evaluated by constructing a co-phylogenetic plot, employing the ML inference tree of each protein. The topology of both phylogenetic trees was very similar, and two main monophyletic groups were identified (fig 3-8.). The monophyletic groups were named clade I and clade II for efficiently description, and both exhibit strong support values in each phylogenetic trees (clade I support values: arxA= 97.2, ArxB = 94; clade II support values: arxA= 91, ArxB = 90). Additionally, within each clades, a similar pattern of branching was also identified (Fig 3-8).

Concerning clade I, three highly supported clusters were formed. Cluster 1 mainly comprised gammaproteobacterial haloalkaliphilic organisms, the majority of which were isolated from alkaline and hypersaline places, including the *Alkalilimnocola ehrlichii* MLHE-1^{T}. This phyletic group was consistent with the cluster 1 of the ArxA phylogenetic tree on chapter 2. The two remaining clusters were formed by sequences from organism recovered in places with more varied environmental conditions (Table 3-3). One of them, include the photoarsenotroph *Ectothiorhodospira* sp. PHS-1^{T}, grouped alongside *Thiocapsa* sp. KS1^{T}, and a *Rhodospirillales* bacterium, the three of them being related to energy production by anoxygenic photosynthesis. The last cluster within clade I (equivalent to cluster 3 of ArxA

phylogeny in chapter 2) was formed by the sequences of *Azoarcus* species and a *Gallionellales* bacterium (41).

Clade II is equivalent to cluster 2 in the ArxA phylogenetic tree from chapter 2; and after the inclusion of the Jozankei isolated strains sequences, it was formed two highly-supported group. One corresponded to the phyletic group enclosing *Sulfuricella denitrificans* skB26^T, and other formed by ArxA sequences of the Jozankei isolated strains, *Sulfuritortus calidifontis* J1A^T, strain J5B and M52 (Fig 3-8). Additionally, one of the particularities of this clade II is that contain eight *arxA* sequences with a long insertion of approximately 106 amino acid, localized between position 376 and 472 of the multiple sequences alignment (Fig. 3-9).

The phyletic group in that resides *Sulfuricella denitrificans* skB26^T, first reported strain harboring an insert (36), is similar in conformation to the cluster 2 in chapter two. This phyletic group contains ArxA sequences also harboring the long insertion but found in uncultured bacteria genomes of two gammaproteobacteria from groundwater sample (USA)(41) and two betaproteobacteria derived from CO2 driven Geyser (USA)(42) (Fig. 3-6, Table 3-3).

Another phyletic group is formed by the ArxA sequences with a long insertion of the Jozankei isolated strain, *Sulfuritortus calidifontis* J1A^T, strain J5B and M52, and by sequences not harboring this feature of *Rhodocyclales* bacterium and *Betaproteobacteria* bacterium, from CO2 driven Geyser (USA) metagenome.

From the co-phylogenetic tree, it is evident a correlated evolutionary history among the putative ArxA and ArxB proteins. Additional to the topology evaluation, the mirror tree method also compares the distance matrices. Pearson's correlation coefficient was used to evaluate the

intensity of co-evolution between putative proteins A and B. The distance matrices showed a high correlation (r = 0.8756, p = 0.001), usually observed in interacting proteins.

3.3.6 Diversity of *arxA* sequences harboring long insertion

Two samples, a water of 40 m depth of Lake Mizugaki and a microbial mat obtained in Jozankei hot springs were used to evaluate the presence and diversity of *arxA*-type sequence variants with long insertion. The expected band length of sequences with insertion (~1200 bp) was only detected in the Jozankei sample, obtaining 12 single clones clustered in 5 OTUs, called JZK-1200 clone library.

Amplified PCR product of approximately ~900 bp band length, corresponding to *arxA* sequence variants without insertion, were obtained and examined in both samples. In the Mizugaki water sample was recovered a total 41 sequences, grouped in 7 OTUs, called MZG900 clone library. In contrast, 33 single clones were detected from the Jozankei biomat sample, and clustered in 5 OTUs, named JZK900 clone library.

The closest cultured relatives to each OTU was determined according to BLASTP (Table 3-4). All OTUs from JZK-1200 clone library were related to *Sulfuricella denitrificans* skB26^{T} (58 to 60% of amino acid identity). In contrast, the clone's libraries made by OTUs not harboring the insertion were, in the majority of the cases, strongly related to *Azoarcus* sp. CIB (78 to 81% amino acid identity). Other OTUs of partial *arxA* sequence without insertion had low similarity to *Thioalkalivibrio* sp. ALM2T, *Halomonas* sp. BC04, *Halomonas chromatireducens* AGD 8-3 and *Azoarcus tolulyticus* ATCC 51758 (Table 3-4), between 60 to 68% amino acid identity.

The OTUs and the complete ArxA sequences of the genomes from previous steps were used to reconstruct a neighbor-joining phylogenetic tree using the distance criterion. It was excluded the gaps within the multi-sequence alignment, which resulted in 303 final positions for reconstruction of the tree (fig. 3-10).

The majority of the OTUs were related to the ArxA sequences of the organism genomes that fell within the so-called clade II of previous phylogenies (equivalent to cluster 2 in chapter 2) (fig. 2-2, fig. 3-8 and fig. 3-10). Interestingly, four of the five OTUs from JZK-1200 clone library (those harboring the insertion), formed a strongly supported group within the clade II with the ArxA sequences of the Jozankei isolated strain, *Sulfuritortus calidifontis* J1A^T, strain J5B and M52 (Bootstrap = 100), also containing this feature. This group also include three OTUs of sequences without insertion, MZG900-Otu2, MZG900-Otu7, and JZK900-Otu3, and two ArxA from uncultured bacteria genomes (fig. 3-10).

In the clade II was also placed five other OTUs. One of them, the JZK900-Otu5 was closely related to a recently submitted ArxA protein sequence from a *Rhodocyclaceae* bacterium to the Genbank Database. In contrast, the evolutionary relationship of the remaining 4 OTUs (MZG900-Otu3, JZK900-Otu1, JZK900-Otu2, and JZK-1200-Otu1) was not determined because of a low bootstrap value of their respective node. A closer look into the amino acid multi-sequence alignment showed that all the complete and partial ArxA sequences within the clade II that did not contain the long insertion, presented a small sequence of 12 amino acids between positions 336 and 348 (Fig 3-10).

Four of the five OTUs located outside the clade II, MZG900-Otu1, MZG900-Otu6, MZG900-Otu5, and JZK900-Otu4 were closely related to *Azoarcus* species (Bootstrap = 99.7) (Fig. 3-10), and present amino acid sequence similarity between 78% to 81% identity to the

ArxA of *Azoarcus* sp. CIB (Table 3-4). In contrast, MZG900-Otu4 was similar to that of *Azoarcus tolulyticus* ATCC 51758 (72% identity), and somewhat related to the previous *Azoarcus* species group based on the node bootstrap (Bootstrap = 77.6) (Fig 3-10).

An additional phylogenetic analysis was made including partial ArxA clones sequences of previous studies. Some of the clones were obtained in California (USA) from a red microbial mat in Paoha Island and water from Mono Lake Beach (alkaline-saline lake - MLBX), and a riverbed sediment sample of the Hot Creek area of hot springs (HC) (2). Other clones were recovered from an alkaline-saline lake (HJ) in Mongolia (10). The final dataset consisted of 92 amino acid positions in the multi-sequence alignment used for the reconstructions of the phylogenetic tree (fig. 3-11).

The majority of the clones recovered from the alkaline-saline places (Mono Lake and soda lake in Mongolia) were closely related to the ArxA sequences of mainly haloalkaliphilic gammaproteobacterial organism including *A. ehrlichii* MLHE-1^T (Bootstrap = 87.8, representing the cluster 1 in the previous ArxA phylogenies). Only two clones obtained in Mono Lake Beach and one from the Mongolia Soda Lake were outside of the cluster 1, and in contrast, were closely related to *Ectothiorhodospira* sp PHS-1 ^{T} (fig. 3-11).

More importantly, all the clones recovered from the Hot Creek riverbed sediment (HC) were closely related to the ArxA sequences of the Jozankei isolated strain *Sulfuritortus calidifontis* J1A^T, strain J5B and M52 (Bootstrap = 99.9) located within clade II (Fig. 3-11).

3.4 Discussion

In the present study homologous genes to the *arxABC* of *A. ehrlichii* MLHE-1^T were detected in 46 genomes analyzed (supplementary table 2, Figure 1). These genes are predicted

to code for the functional elements of the arsenite oxidase; *arxA*, molybdopterin oxidoreductase; *arxB*, iron-sulfur [4Fe-4S] containing protein; and *arxC*, quinol-oxidizing membrane protein (2). Putative functional elements of Aio arsenite oxidase and Arr arsenate reductase were also constant within the *aio* gene cluster (*aioBA* genes) (43) and *arr* operon (*arrAB* genes) (44) of many organism. The occurrence of the subunits AB coding genes in many enzymes is expected since putative functional elements tend to be more conserved than less functional ones (45).

There is a lack of studies involving the *arxC*, which codes for a putative membrane-integrated protein of the PsrC/NrfD family, and suggested to inject or extract electrons to or from the quinones/quinol (46). This family devoid of known prosthetic groups, but it is present in a variety of molybdoenzymes genetic clusters, like in polysulfide reductase (PsrC), tetrathionate reductase (TtrC), selenite reductase and DMSO reductase (DmsC) (46, 47).

The *arxB'* and *arxD* genes, also included in the *arx* gene cluster of *A. ehrlichii* MLHE-1^{T}, were present in the majority of the genomes, with only a few exceptions. The *arxB'* gene appear to be exclusively part of the *arx* gene clusters, since no homologous sequences had been described in the *aio* or *arr* operons. The predicted protein sequences of the *arxB'* gene contain at least two motifs of iron-sulfur cluster, similar to the translated *arxB* gene (Fig. 3-12). Expression of the *arxB'* gene was detected during the arsenite-induction and arsenite oxidation in *Ectothiorhodospira* sp. BSL-9 (8), and also in the southern basin of Mono Lake (12). Currently, the role of the *arxB'* gene in the arsenite oxidation activity has not been demonstrated; however, the study showed high conservation in the majority of the *arx*-like gene cluster (37 assemblies), except in all of the members of the archaea domain (Fig. 3-2).

The *arxD* gene was also very conserved among the *arx*-containing genomes (43 of the 46 assemblies) (Fig. 3-2), and on the contrary, the respective homologous sequences in the *aio* or *arr* operon (*arr*D) present a considerable variability from one strain to another (43)(44). The *arxD* gene, predicted to code a TorD-like protein, has been extensively studied. The TorD-like protein is required for the last step of molybdopterin biosynthesis, and also in the insertion of the cofactor into the catalytic site of the enzyme (48). Phylogenetic analysis of the translated *arxD* homologous sequences of the genomes showed a similar topology to the phylogeny of ArxA and ArxB (Fig.3-8 and Fig. 3-13). The ArxA, ArxB, and ArxD phylogeny resemblance are expected since the evolution of the chaperones appear to be profoundly influenced by their molybdopartner, to adapt to their related enzyme and its specific activity and, assembly (48, 49).

On the other hand, the majority of the *arx* containing genomes (31 assemblies) possesses conserved homologous sequences to the *arxXSR* gene cluster, predicted to encode putative regulatory proteins (Fig 3-2). Similar genes were reported in the *aio* gene cluster (*aioXSR*) of some microbes, and in less fashion in the *arr* operon (2, 43, 44). Some notorious finding was the absence of these *arxXSR* homologous genes in all the organism affiliated to the order *Oceanospirillales* (genus *Halomonas* and *Marinospirillum*) (Fig 3-2). Conrad and col (2013) also found a lack of these regulatory proteins coding genes in the genome of *Halomonas* sp. strain BSL-1, another arsenite-oxidizing *arxA* containing bacteria (50).

The homologous *aioS* and *aioR* genes coding for a sensor kinase, and a response regulator, respectively. Both have been extensively studied, uncovering their importance on the arsenite recognition and regulation in the expression of *aioAB* genes (51–53). The homologous *aioX* constitute a third regulatory component coding for a putative periplasmic As(III) binding

protein, also essential in the *aioAB* expression (54). The AioX was recently heterologously expressed in an *E.coli*, as well as its homologous proteins ArxX and the ArrX, from the *arx* and *arr* operon, respectively. Interestingly, the three of them showed a substrate binding specificity for the arsenic oxyanions(55).

As it was mention before, 46 genomes were found to harbor at least three homologous genes to the *arxABC* genes of *A. ehrlichii* MLHE-1^{T}, yet, only 41 *arxA*-like sequences fell within the ARX arsenite oxidase clade (Fig. 3-5). The remaining *arxA*-like sequences, all of them affiliated to the *Candidatus Methanoperedens*, formed a distant cluster, placed between the ARR arsenate reductase and ARX arsenite oxidase clade, maybe due to an earlier divergent event. It is needed more study to verify the biochemical function of these *Candidatus Methanoperedens* homologous genes, and whether or not they are related to the arsenic metabolism.

The ArxA and ArxB phylogenetic trees of the 41 *Proteobacteria* organism members of the Arx Clade have a highly correlated evolutionary history evidenced in similar branching patterns. (Fig. 3-8). This tree resemblance is often observed in interacting proteins, like that of enzymes formed by multiple subunits, as a result of the compensatory mutation to maintain the stability and function of the catalyst over the course of evolution (26, 56).

In the same way, the cluster formation in ArxAB co-phylogenetic tree is similar to the observed in the ArxA phylogeny of chapter 2. One noticeable cluster, is perhaps the monophyletic group mainly composed by the organisms affiliated to the *Gammaproteobacteria* class, the majority of them recovered from extreme habitats with high salinity and pH conditions. The adaptive evolution may explain the distinction of this particular phyletic group, probably to maintain the “fitness” of the proteins under those extreme conditions.

Ectothiorhodospira sp. PHS-1, a *Gammaproteobacteria* organism isolated from a saline-alkaline site (Red biofilm in Paoha Island, CA, USA) correspond to one of the exceptions to the abovementioned particular phyletic group. Instead, PHS-1 strain was closely related to *Thiocapsa* sp. KS1, isolated from sewage sludge of the municipal treatment plant of Konstanz, Germany (57), and Rhodospirillales bacterium RIFCSPLOWO2_12_FULL_67_15, an *Alphaproteobacteria*. Also, the PHS-1 strain was distantly related to *Ectothiorhodospira* sp. BSL-9, despite being part of the same genus and oxidize the arsenite through anaerobic photosynthesis (Fig. 3-8, Fig 3-14), suggesting the occurrence of a horizontal gene transfer (HGT).

Similarly, the sequences of *Betaproteobacteria* organisms were placed in separated clusters within the phylogenetic tree also indicating HGT and different evolutionary history, or perhaps arisen under different conditions that may have resulted in distantly related phylotypes (Fig. 3-8).

Horizontal gene transfer is a process usually responsible for the evolutionary history of other arsenic-related genes, like the arsenic resistance *ars* gene, and in few cases in *aio* and *arr* gene cluster (44). Although not exhaustive, the results obtained in the present study suggests that HGT may have contributed to the evolution of the *arx* gene cluster as well. For example, some *arx*-like gene clusters were located in genetic mobile elements, like in the plasmid of *Sulfuricella denitrificans* skB26^{T} (36) and *Halomonas* sp. A3H3 (58); within a genomic island in *Azoarcus* sp. CIB (59), or next to a transposase gene in strain M52. Additionally, two ArxA sequences of *Gammaproteobacteria* were found to be closely related to those of *Betaproteobacteria* organisms like *Sulfuricella denitrificans* skB26^{T}.

One of the particularities in *Sulfuricella denitrificans* skB26^{T} is the presence of a long insertion within the *arxA* sequence (approx. 106 amino acid), a feature described for the first time in 2014 by Watanabe, T., Kojima H., and Fukui, M. (41). In the present chapter, we described other *arxA* sequences harboring this insertion in the genomes of isolated strains and metagenomics studies, affiliated mostly to the *Betaproteobacteria* class. A phylogenetic analysis was made to evaluate the distribution and diversity of the *arxA* sequences with and without insertion. In doing this, a small insertion of 12 amino acids was detected in some *arxA* sequences not harboring the long one.

The results of the phylogenetic analysis indicate that those *arxA* gene sequences holding an insert even if is long or short, belong exclusively to the clade II (also named cluster 2 in chapter 2). It seems that these insertions are an evolutionary feature that probably arisen from an earlier divergent event that led to the formation of this separate group, and may perhaps be used to distinguish them from the rest of *arxA* genes sequences located in distantly related clusters (Fig. 3-10).

After inclusion of the recovered clones of partial *arxA* sequences of previous studies to the ArxA phylogenetic tree along with the OTUs and complete ArxA sequences, it was evident that those clones recovered from Hot Creek area of hot spring in California (CA, USA) (2), were closely related the ArxA sequences of *Sulfuritortus calidifontis* J1A^{T}, strain J5B and, M52 within the clade II (bootstrap = 99.9, Fig. 3-12). Therefore, it may be possible that the Hot Creek clone sequences correspond to *arxA* gene variants with insert.

The Jozankei and Hot Creek hot springs area shared some similarities in the physicochemical condition. The geothermal waters from a magmatic fluid generate both hot springs (60, 61). Both contain arsenic, although the concentration was higher in Jozankei than

Hot Creek waters (3.20 mg/L - ~40 μM; 10–18 μM, respectively). The chloride content was also higher in Jozankei (40 mM) to that found in Hot Creek (0.6 to 0.8 mM) (2, 11, 60).

Previously, it was reported the arsenite oxidation activity in isolated strains and stream waters of Hot Creek (61, 62); however, at the moment, there has not been designated a connection between the *arxA* gene to the arsenite oxidation activity in the place.

3.5 Conclusion

As it was shown, 41 *arx* homologous gene clusters were found to be phylogenetically diverse but limited for the moment to *Proteobacteria* organisms, and associated with a varied type of environments, some of which were arsenic-rich places. Overall, the *arxB'ABCD* genes, coding for the putative structural and functional proteins, were highly conserved among the genomes, and in case of *arxXSR* genes coding for the putative regulatory proteins, were absent especially in closely related organism affiliated to the order *Oceanospirillales* (genus *Halomonas*, *Marinospirillum*, and *Nitrincola*).

Additionally, the ArxA amino acid sequences analysis showed the taxonomical and environmental diversity of two novel sequence variants containing a long or a short insertion, first addressed as an exciting feature in *Sulfuricella denitrificans* skB26^{T} elsewhere. This study represents the first description of a group of novel *arxA* sequences variants that collectively represents an extensive, large group of unexplored *arxA*-type of genes. We also report the presence of a complete ArxA sequences variant with long insertion in three cultured Arx-harboring betaproteobacterial strains, isolated from a microbial mat obtained in the area of Jozankei hot spring (Hokkaido, Japan). Based on an ArxAB consensus phylogeny the ArxA

sequence holding an insert constituted a phyletic group distantly related to those of without insert.

Finally, we propose a wider environmental distribution of this type of *arxA* sequences variants due to the strong close relationship of the AxrA clones obtained in Hot Creek riverbed sediment in Mono Lake (CA, USA) to the sequences with insertion of the OTUs and Jozankei isolated strains recovered in Japanese samples sites.

References

1. **Oremland RS, Saltikov CW, Wolfe-Simon F and Stolz JF**. 2009 Arsenic in the Evolution of Earth and Extraterrestrial Ecosystems. Geomicrobiol. J. **26**:522–536.

2. **Zargar K, Conrad A, Bernick DL, Lowe TM, Stolc V, Hoeft S, Oremland RS, Stolz J and Saltikov CW**. 2012 ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. Environ. Microbiol. **14**:1635–1645.

3. **Oremland RS, Hoeft SE, Santini JM, Bano N, Hollibaugh R a and Hollibaugh JT**. 2002 Anaerobic Oxidation of Arsenite in Mono Lake Water and by a Facultative, Arsenite-Oxidizing Chemoautotroph, Strain MLHE-1. Society **68**:4795–4802.

4. **Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM and Oremland RS**. 2007 *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int. J. Syst. Evol. Microbiol. **57**:504–512.

5. **Richey C, Chovanec P, Hoeft SE, Oremland RS, Basu P and Stolz JF**. 2009 Respiratory arsenate reductase as a bidirectional enzyme. Biochem. Biophys. Res. Commun. **382**:298–302.

6. **Zargar K, Hoeft S, Oremland R and Saltikov CW**. 2010 Identification of a novel arsenite oxidase gene, arxA, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J. Bacteriol. **192**:3755–3762.

7. **Kulp TR, Hoeft SE, Asao M, Madigan MT, Hollibaugh JT, Fisher JC, Stolz JF, Culbertson CW, Miller LG and Oremland RS**. 2008 Arsenic(III) Fuels Anoxygenic Photosynthesis in Hot Spring Biofilms from Mono Lake, California. Science (80-.). **1**:967–970.

8. **Hernandez-Maldonado J, Sanchez-Sedillo B, Stoneburner B, Boren A, Miller L, Mccann S, Rosen M, Oremland RS and Saltikov CW**. 2017 The genetic basis of anoxygenic photosynthetic arsenite oxidation. Environ. Microbiol. **19**:130–141.

9. **Wu G, Huang L, Jiang H, Peng Y, Guo W, Chen Z, She W, Guo Q and Dong H**. 2017 Thioarsenate formation coupled with anaerobic arsenite oxidation by a sulfate-reducing bacterium isolated from a hot spring. Front. Microbiol. **8**:1–9.

10. **Hamamura N, Itai T, Liu Y, Reysenbach AL, Damdinsuren N and Inskeep WP**. 2014 Identification of anaerobic arsenite-oxidizing and arsenate-reducing bacteria associated with an alkaline saline lake in Khovsgol, Mongolia. Environ. Microbiol. Rep. **6**:476–482.

11. **Ospino MC, Kojima H, Watanabe T, Iwata T and Fukui M**. 2018 Diversity of anaerobic arsenite-oxidizing bacteria in low-salt environments analyzed with a newly developed PCR-based method. Limnology **19**:177–183.

12. **Edwardson CF and Hollibaugh JT**. 2017 Metatranscriptomic analysis of prokaryotic communities active in sulfur and arsenic cycling in Mono Lake, California, USA. ISME J. **11**:2195–2208.

13. **Buchfink B, Xie C and Huson DH**. 2014 Fast and sensitive protein alignment using DIAMOND. Nat. Methods **12**:59–60.

14. **Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW and Hauser LJ**. 2010 Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:.

15. **Kojima H, Watanabe M and Fukui M**. 2017 *Sulfuritortus calidifontis* gen . nov ., sp . nov ., a novel sulfur oxidizer isolated from a hot spring microbial mat. Int. J. Syst. Evol. Microbiol. 2017: 10.1099/ijsem.0.001813.

16. **Lecompte O, Ripp R, Thierry JC, Moras D and Poch O**. 2002 Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. Nucleic Acids Res. **30**:5382–5390.

17. **Yutin N, Puigbo P, Koonin E V. and Wolf YI**. 2012 Phylogenomics of Prokaryotic Ribosomal Proteins. PLoS One **7**:e36972.

18. **Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC and Banfield JF**. 2016 A new view of the tree of life. Nat. Microbiol. **1**:1–6.

19. **Katoh K, Rozewicki J and Yamada KD**. 2017 MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 2017: 10.1093/bib/bbx108.

20. **Castresana J**. 2000 Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. **17**:540–552.

21. **Tamura K, Stecher G, Peterson D, Filipski A and Kumar S**. 2013 MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30**:2725–2729.

22. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O**. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

23. **Anisimova M and Gascuel O**. 2006 Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol **55**:539–552.

24. **Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM and Gascuel O**. 2008 Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36**:465–469.

25. **Edgar RC**. 2004 MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**:1792–1797.

26. **Pazos F and Valencia A**. 2001 Similarity of phylogenetic trees as indicator of protein–protein interaction. Protein Eng. Des. Sel. **14**:609–614.

27. **De Juan D, Pazos F and Valencia A**. 2013 Emerging methods in protein co-evolution. Nat. Rev. Genet. **14**:249–261.

28. **Ochoa D and Pazos F**. 2010 Studying the co-evolution of protein families with the Mirrortree web server. Bioinformatics **26**:1370–1371.

29. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J and Punta M**. 2014 Pfam: The protein families database. Nucleic Acids Res. **42**:222–230.

30. **Waterhouse AM, Procter JB, Martin DMA, Clamp M and Barton GJ**. 2009 Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics **25**:1189–1191.

31. **Bendtsen JD, Nielsen H, Widdick D, Palmer T and Brunak S**. 2005 Prediction of twin-arginine signal peptides. BMC Bioinformatics **6**:1–9.

32. **Gascuel O**. 1997 BIONJ: An Improved Version of the NJ Algorithm Based on a Simple Model of Sequence Data. Mol. Biol. Evol. **14**:685–695.

33. **Lefort V, Desper R and Gascuel O**. 2015 FastME 2 . 0 : A Comprehensive , Accurate , and Fast Distance-Based Phylogeny Inference Program. Mol. Biol. Evol. **32**:2798–2800.

34. **Yu G, Smith DK, Zhu H, Guan Y and Lam TTY**. 2017 Ggtree: an R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data. Methods Ecol. Evol. **8**:28–36.

35. **Revell LJ**. 2012 phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. **3**:217–223.

36. **Watanabe T, Kojima H and Fukui M**. 2014 Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: Genetic insights into the sulfur oxidation pathway of betaproteobacteria. Syst. Appl. Microbiol. 2014: 10.1016/j.syapm.2014.05.010.

37. **Watanabe T, Miura A, Iwata T, Kojima H and Fukui M**. 2017 Dominance of *Sulfuritalea* species in nitrate-depleted water of a stratified freshwater lake and arsenate respiration ability within the genus. Environ. Microbiol. Rep. **9**:522–527.

38. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF**. 2009 Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75**:7537–7541.

39. **Hoeft McCann S, Boren A, Hernandez-Maldonado J, Stoneburner B, Saltikov C, Stolz J and Oremland R**. 2017 Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of *Ectothiorhodospira* from Mono Lake, California and Big Soda Lake, Nevada. Life **7**:1–14.

40. **Saltikov CW and Newman DK**. 2003 Genetic identification of a respiratory arsenate reductase. Proc. Natl. Acad. Sci. U. S. A. **100**:10983–8.

41. **Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS and Banfield JF**. 2016 Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. **7**:1–11.

42. **Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, Hug LA, Burstein D, Emerson JB, Thomas BC and Banfield JF**. 2017 Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. Environ. Microbiol. **19**:459–474.

43. **Van Lis R, Nitschke W, Duval S and Schoepp-Cothenet B**. 2013 Arsenics as bioenergetic substrates. Biochim. Biophys. Acta - Bioenerg. **1827**:176–188.

44. **Andres J and Bertin PN**. 2016 The microbial genomics of arsenic. FEMS Microbiol. Rev. **39**:fuv050.

45. **Levy ED and Teichmann S**. 2013 Structural, evolutionary, and assembly principles of protein oligomerization. 1st ed., Elsevier Inc.

46. **Grimaldi S, Schoepp-cothenet B, Ceccaldi P, Guigliarelli B and Magalon A**. 2013 The prokaryotic Mo / W- bis PGD enzymes family : A catalytic workhorse in bioenergetic. Biochim. Biophys. Acta **1827**:1048–1085.

47. **Simon J and Kern M**. 2008 Quinone-reactive proteins devoid of haem *b* form widespread membrane-bound electron transport modules in bacterial respiration. Biochem. Soc. Trans. **36**:1011–1016.

48. **Genest O, Méjean V and Iobbi-Nivol C**. 2009 Multiple roles of TorD-like chaperones in the biogenesis of molybdoenzymes. FEMS Microbiol. Lett. **297**:1–9.

49. **Bay DC, Chan CS and Turner RJ**. 2015 NarJ subfamily system specific chaperone diversity and evolution is directed by respiratory enzyme associations Genome evolution and evolutionary systems biology. BMC Evol. Biol. **15**:1–18.

50. **Conrad A**. 2013 The *arx* anaerobic arsenite-oxidization pathway is conserved in Halomonas and Ectothiorhodospira strains isolated from Big Soda Lake, nevada. 2013.

51. **Kashyap DR, Botero LM, Franck WL, Daniel J, Mcdermott TR and Hassett DJ**. 2006 Complex Regulation of Arsenite Oxidation in Agrobacterium tumefaciens Complex Regulation of Arsenite Oxidation in *Agrobacterium tumefaciens*. J. Bacteriol. **188**:1081–1088.

52. **Koechler S, Cleiss-Arnold J, Proux C, Sismeiro O, Dillies M-A, Goulhen-Chollet F, Hommais F, Lievremont D, Arsene-Ploetze F, Coppee J-Y, Bertin PN, Mead M, Rosen B, Smith A, Lingas E, Rahman M, Muller D, Simeonova D, Riegel P, et al.** 2010 Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. BMC Microbiol. **10**:53.

53. **Sardiwal S, Santini JM, Osborne TH and Djordjevic S**. 2010 Characterization of a two-component signal transduction system that controls arsenite oxidation in the chemolithoautotroph NT-26. FEMS Microbiol. Lett. **313**:20–28.

54. **Liu G, Liu M, Kim EH, Maaty WS, Bothner B, Lei B, Rensing C, Wang G and McDermott TR**. 2012 A periplasmic arsenite-binding protein involved in regulating arsenite oxidation. Environ. Microbiol. **14**:1624–1634.

55. **Badilla C, Osborne TH, Cole A, Watson C, Djordjevic S and Santini JM**. 2018 A new family of periplasmic-binding proteins that sense arsenic oxyanions. Sci. Rep. **8**:1–12.

56. **Goh CS, Bogan AA, Joachimiak M, Walther D and Cohen FE**. 2000 Co-evolution of proteins with their interaction partners. J. Mol. Biol. **299**:283–293.

57. **Schott J, Griffin BM and Schink B**. 2010 Anaerobic phototrophic nitrite oxidation by *Thiocapsa* sp. strain KS1 and *Rhodopseudomonas* sp. strain LQ17. Microbiology **156**:2428–2437.

58. **Koechler S, Plewniak F, Barbe V, Battaglia-Brunet F, Jost B, Joulian C, Philipps M, Vicaire S, Vincent S, Ye T and Bertin PN**. 2013 Genome Sequence of *Halomonas* sp. Strain A3H3, Isolated from Arsenic-Rich Marine Sediments. Genome Announc. **1**:e00819-13.

59. **Martin-Moldes Z, Zamarro MT, del Cerro C, Valencia A, Gomez MJ, Arcas A, Udaondo Z, Garcia JL, Nogales J, Carmona M and Diaz E**. 2015 Whole-genome analysis of *Azoarcus* sp. strain CIB provides genetic insights to its different lifestyles and predicts novel metabolic features. Syst. Appl. Microbiol. **38**:462–471.

60. **Kubota Y, Sakakibara M and Sano S**. 2010 Arsenic cocentration of hot spring and river waters at Jyozankei hot spring area in Sapporo city, western Hokkaido, Japan. Mem. Fac. Sci. Ehime Univ. **16**:7–13.

61. **Wilkie JA and Hering JG**. 1998 Rapid oxidation of geothermal arsenic(III) in streamwaters of the eastern Sierra Nevada. Environ. Sci. Technol. **32**:657–662.

62. **Salmassi TM, Walker JJ, Newman DK, Leadbetter JR, Pace NR and Hering JG**. 2006 Community and cultivation analysis of arsenite oxidizing biofilms at Hot Creek. Environ. Microbiol. **8**:50–59.

63. **Ban N, Beckmann R, Cate JHD, Dinman JD, Dragon F, Ellis SR, Lafontaine DLJ, Lindahl L, Liljas A, Lipton JM, McAlear MA, Moore PB, Noller HF, Ortega J, Panse**

VG, Ramakrishnan V, Spahn CMT, Steitz TA, Tchorzewski M, et al. 2014 A new system for naming ribosomal proteins. Curr. Opin. Struct. Biol. **24**:165–169.

64. **Oremland RS, Saltikov CW, Stolz JF and Hollibaugh JT**. 2017 Autotrophic microbial arsenotrophy in arsenic-rich soda lakes. FEMS Microbiol. Lett. **364**:.

65. **Emerson JB, Thomas BC, Alvarez W, Banfield JF and Near- TH**. 2016 Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla. Environ. Microbiol. **18**:1686–1703.

Table 3-1. Set of 10 Ribosomal protein used for the tree inference.

Ribosomal protein name	alternative name †	gene name in bacteria	Taxonomic range † *	Marker ID (Pfam)	Description	Length of block in alignment (aa)
50S ribosomal protein L2	uL2	rplB	B A E	PF00181	Ribosomal Proteins L2, RNA binding domain	269
50S ribosomal protein L3	uL3	rplC	B A E	PF00297	Ribosomal protein L3	142
50S ribosomal protein L4	uL4	rplD	B A E	PF00573	Ribosomal protein L4/L1 family	159
50S ribosomal protein L13	uL13	rplM	B A E	PF00572	Ribosomal protein L13	133
50S ribosomal protein L23	uL23	rplW	B A E	PF00276	Ribosomal protein L23	80
30S ribosomal protein s2	uS2	rpsB	B A E	PF00164	Ribosomal protein S12/S23	184
30S ribosomal protein s4	uS4	rpsD	B A E	PF00163	Ribosomal protein S4/S9 N-terminal domain	137
30S ribosomal protein s9	uS9	rpsI	B A E	PF00380	Ribosomal protein S9/S16	105
30S ribosomal protein s10	uS10	rpsJ	B A E	PF00338	Ribosomal protein S10p/S20e	89
30S ribosomal protein s11	uS11	rpsK	B A E	PF00411	Ribosomal protein S11	120

† Reference: Ban, N. et al., 2014 (63)

* Abbreviation: B: bacteria, A: archaea, E: eukaryotes.

Many of the genomes lack the 16S rRNA, therefore was not used for the phylogeny reconstruction of the species trees.

Table 3-2. DIAMOND results for the *arx* sequences of 46 genomes homologous to the *arxXSRB'ABCD* gene cluster of *A. ehrlichii*

Organism name	arxA				arxB				arxC				arxD			
	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number
Ectothiorhodospira sp. BSL-9	0	78.3	99.6	WP_063465591	1.3E-133	85.9	96.8	WP_063465592	3.6E-80	41.8	93.5	WP_063465593	1.8E-44	40.8	92.7	WP_082829968
Thioalkalivibrio nitratireducens DSM 14787	0	74.5	99.6	WP_015257777	1.2E-137	87.6	97.2	WP_043740181	2.6E-83	42.9	94.5	WP_015257775	2.2E-45	41.4	94.7	WP_083499345
Halomonas sp. A3H3	0	68.7	99.6	WP_022524225	2.9E-136	88.4	96.8	WP_053857286	2.4E-137	61.9	99.2	WP_022524227	1E-34	39.7	96.7	WP_022524228
Halomonas chromatireducens AGD 8-3	0	73.7	99.6	AMD01645	4.9E-136	88.4	96.8	AMD01646	3.1E-148	68.4	99	AMD01647	2.2E-29	35.5	96.3	AMD01648
Halomonas boliviensis LC1	0	69.9	99.6	WP_040480658	4.4E-137	88	96.8	WP_040481000	1.3E-135	63.3	96	WP_050804846	1.7E-21	33.1	90.3	WP_007113188
Thioalkalivibrio denitrificans ALJD	0	72.1	99.6	WP_077279016	9.9E-137	87.6	97.2	WP_077279015	3.6E-157	73	100	WP_077279014	2.3E-31	39	71.3	WP_077279013
Halomonas sp. BC04	0	73.1	99.6	WP_043513786	1.3E-136	88.8	96.8	WP_043513788	2.1E-149	71.7	94.2	WP_081785949	3.5E-27	34.6	90	WP_043513790
Nitrincola lacisaponensis 4CA	0	69.7	99.6	WP_036549261	9.6E-132	85.9	96.8	WP_036549263	1.9E-150	70.6	97	WP_036549176	4E-31	33.2	100	WP_051632834
Thioalkalivibrio thiocyanodenitrificans ARhD 1	0	73.4	99.6	WP_018231946	3.2E-135	87.6	96.8	WP_018231947	1E-79	43.2	94.2	WP_018231948	6.9E-44	40	92.7	WP_083908660
Thioalkalivibrio sulfidiphilus ALJ17	0	73.3	99.6	WP_026289649	4.6E-134	87.1	96.8	WP_018952917	3.9E-87	45.8	95	WP_018952918	6.1E-40	39	94.7	WP_018952919
Marinospirillum celere DSM 18438	0	71.3	99.6	WP_091965171	3.1E-130	83.4	96.8	WP_091965207	5.4E-153	69.5	100	WP_091965173	4E-31	33.4	96.7	WP_091965175
Thioalkalivibrio sp. ALSr1	0	67	99.9	WP_019643167	3E-133	84	97.6	WP_019643166	4.2E-121	56.6	97.5	WP_019643165	1.4E-52	42.1	98	WP_019643164
Marinospirillum alkaliphilum DSM 21637	0	73.4	93.8	WP_072327058	2E-129	83.5	96.8	WP_072327057	1.6E-149	69.2	98	WP_072327056	3E-31	34	90.3	WP_072327055
Thioalkalivibrio sp. ALMg11	0	68.4	99.9	WP_018950173	6.6E-133	83.5	97.6	WP_018950172	8.7E-119	56.8	94	WP_018950171	8.8E-47	40.7	95.3	WP_018950170
Thioalkalivibrio sp. AKL19	0	68.9	99.9	WP_024327966	1.9E-132	83.1	97.6	WP_024327967	2.1E-117	56.8	94	WP_024327968	2.6E-43	39.9	93	WP_024327969
Thioalkalivibrio sp. ARh3	0	70.8	97.6	WP_018864019	3.1E-130	82.3	97.6	WP_018864020	4.5E-123	58.7	97.7	WP_018864021	5.3E-52	43.1	98.7	WP_018864022
Halorhodospira halophila SL1	0	68.1	99.6	WP_011813170	6.9E-122	79.3	97.2	WP_041595000	2.5E-41	30.2	93.7	WP_011813172	1.7E-34	44.4	67.7	WP_011813173
Thioalkalivibrio sp. ALM2T	0	68.1	99.6	WP_019593327	2.8E-123	79.6	96.4	WP_019593328	2.3E-55	36.1	92.9	WP_019593329	2.2E-26	33.5	98.7	WP_081622651
Thioalkalivibrio sp. ALR17-21	0	66.7	99.9	WP_024329954	3E-163	83	100	WP_018946173	1E-153	57	90	WP_018946172	3.7E-53	42.8	98.7	WP_024329955
Thioalkalivibrio sp. AKL17	0	66.5	100	WP_018946174	3E-163	83	90	WP_018946173	1E-153	57	98.7	WP_018946172	4.8E-53	42.8	88	WP_018946171
Thiocapsa sp. KS1	0	61.8	99.6	WP_093186008	1.4E-111	74.6	95.6	WP_093186012	**4.3E-41**	**36.6**	**72.8**	**WP_093186014**	6.8E-23	41.3	72	WP_093186016
Candidatus Muproteobacteria bacterium RBG 16 65 34	0	61.4	100	OGI47522	1.5E-113	76.8	95.2	OGI47523	3E-71	40.8	96.2	OGI47524	3.6E-24	38.1	69.7	OGI47525
Ectothiorhodospira sp. PHS-1	0	62	99.6	WP_008932021	2.7E-110	72.9	95.6	WP_008932022	9.5E-49	33.2	91.4	WP_008932023	1.3E-18	32.6	93.3	WP_083838795
Azoarcus sp. CIB	1.1E-299	60.1	99.6	WP_050415005	4.3E-108	69.7	96.8	WP_050415004	1.9E-73	41.3	95.5	WP_050415003	3.8E-18	32.4	95.7	WP_050415002
Azoarcus tolulyticus ATCC 51758	1.7E-305	61.1	99.6	WP_076602992	2.8E-107	69.3	96.8	WP_076602993	5.5E-73	42.1	94.2	WP_076602994	1E-26	32.4	95.7	WP_076602995
Rhodospirillales bacterium RIFCSPLOWO2 12 FULL 67 15	0	63.3	99.4	OHC82914	3.7E-107	69.5	98	OHC82915	**1.8E-63**	**42.4**	**86.4**	**OHC82916**				
Sulfuricella denitrificans skB26^{T}	1E-278	50.7	99.5	WP_009207744	2.7E-102	67.6	95.6	WP_041674340	5.1E-87	48.9	93.7	WP_009207742	4.5E-27	33.1	93.7	WP_021035847
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 58	0	60.2	99.5	OGA55942	1.9E-103	70.1	95.6	OGA55947	2.7E-80	46.6	94.2	OGA55946	1.7E-26	39.9	64.3	OGA55941
Gallionellales bacterium RIFCSPLOWO2 02 FULL 59 110	4.4E-290	58.3	99.6	OGS99566	1.5E-100	66.5	96.8	OGS99567	6.1E-80	44.8	94.7	OGS99568	1.9E-25	32.1	94	OGS99569
Gammaproteobacteria bacterium RIFOXYD12 FULL 61 37	1.5E-282	51.5	98.9	OGT88201	3.3E-100	66.8	95.6	OGT88200	1.4E-89	49.1	92.4	OGT88244	1.2E-27	34.9	93.7	OGT88199
Gammaproteobacteria bacterium RIFOXYA12 FULL 61 12	2.6E-282	51.5	98.9	OGT93884	3.3E-100	66.8	95.6	OGT93883	1.4E-89	49.1	92.4	OGT93896	1.2E-27	34.9	93.7	OGT93882
Betaproteobacteria bacterium CG2 30 59 46	1.1E-277	50.6	99.5	OIP17342	3.3E-100	66.8	95.6	OIP17341	6.3E-77	43.5	92.2	OIP17355	2E-27	36.5	93.7	OIP17340
Magnetospirillum magnetotacticum MS-1	1.9E-301	59.2	98.8	KIM00421	1.4E-103	66.1	97.2	KIM00420	3.7E-69	40.5	93.7	KIM00419				
Betaproteobacteria bacterium CG2 30 68 42	1.9E-285	56	99.4	OIP08971	1.2E-94	63.6	95.2	OIP08972	8.5E-90	48.5	94	OIP09018	4.1E-28	35	92.7	OIP08973
Rhodocyclales bacterium CG 4 10 14 3 um filter 68 10	1.5E-285	56.1	99.4	PJA58270	1.2E-94	63.6	95.2	PJA58269	8.5E-90	48.5	94	PJA58268	4E-23	39	86	PJA58267
Deltaproteobacteria bacterium GWC2 42 11	5.6E-301	59.4	98.9	OGP31207	4.7E-70	51.3	92.4	OGP31219	2.3E-26	28.9	90.7	OGP31208				
Hydrogenophilales bacterium CG18 big fil WC 8 21 14 2 50 58 12	6.3E-77	43.5	92.2	PIQ11014	2.7E-19	39.5	64.3	PIQ11015	3.1E-37	32.9	84.3	PIQ11016	5.4E-65	44.5	94.6	PIQ11017
Candidatus Methanoperedens nitroreducens ANME-2d	6.1E-215	46.2	99.8	KCZ73329	1.1E-45	37.7	94.4	KCZ73328	1E-26	28.6	89.9	KCZ73327	1.40E-07	27.5	45	KCZ73326
Candidatus Methanoperedens sp. BLZ1	9.2E-203	44.3	99.8	KPQ43588	1E-28	37	85	KPQ43587	1E-29	28	93	KPQ43586	2E-11	27	45	KPQ43585
Candidatus Methanoperedens sp. BLZ2	9.2E-203	44.3	99.8	WP_097300993	3E-43	38	94	WP_097301035	6.2E-24	28.1	92.4	WP_097300994	9E-10	27	45	WP_097300995
Rhodocyclaceae UBA4043	3.7E-276	55.1	99.5		2.2E-104	65	97.6		7E-92	48.7	94.7		1.5E-17	26	93.3	
Methanoperedens sp. UBA453	5.4E-219	46.2	99.8		3.6E-46	37.7	94.4		3.5E-27	28.6	89.9		0.000000047	27.5	45	
Candidatus Methanoperedens sp. Ru	8.1E-207	44.3	99.8		2.8E-46	38	94.4		5.5E-28	28.1	92.4		0.000000062	27	45	
Sulfuritortus calidifontis J1A^{T}	8.4E-297	53	99.5		1.4E-98	61.6	97.6		2.8E-85	48.2	89.9		1.9E-28	32.7	91	
Strain J5B	2.6E-290	53	99.6		4.7E-102	64.1	97.6		9.5E-89	46.6	92.4		1.3E-26	33.6	89.7	
Strain M52	4E-293	53.4	99.6		2.3E-103	64.9	97.6		1.4E-94	49.2	89.2		3.3E-24	32.3	88	

Cont. Table 3-2. DIAMOND results for the *arx* sequences of 46 genomes homologous to the *arxXSRB'ABCD* gene cluster from *A. ehrlichii*

Organism name	arxB2				arxX				arxS				arxR			
	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number	e value	identity (%)	coverage	Accesion number
Ectothiorhodospira sp. BSL-9	1.9E-47	38.5	83.4	WP_082829967	8.1E-69	53.2	89.6	WP_082829966	5.1E-121	47.5	99	WP_082830080	1.8E-136	58.4	98.6	WP_063465588
Thioalkalivibrio nitratireducens DSM 14787	6.2E-38	33.5	90	WP_015257778	6.2E-38	33.5	90	WP_015257780	6.2E-38	33.5	90	WP_015257781	6.2E-38	33.5	90	WP_015257782
Halomonas sp. A3H3	1E-21	30	93	WP_022524224												
Halomonas chromatireducens AGD 8-3	7E-13	29	95	AMD01644												
Halomonas boliviensis LC1	2.4E-13	25.8	85	WP_007113184												
Thioalkalivibrio denitrificans ALJD	7.9E-41	33.7	91.8	WP_077279017	6.4E-74	52	94	WP_077279018	1.3E-116	46.5	98.6	WP_077279019	1.7E-139	59.7	98.4	WP_077279020
Halomonas sp. BC04	3.2E-10	28.2	79	WP_081785948												
Nitrincola lacisaponensis 4CA	7E-18	26	93	WP_036549174												
Thioalkalivibrio thiocyanodenitrificans ARhD 1	2.1E-54	38.2	84.1	WP_083908659	7.1E-73	52.6	91	WP_018231944	2.8E-111	44.4	98.6	WP_018231943	2.2E-139	60	99.1	WP_018231942
Thioalkalivibrio sulfidiphilus ALJ17	8.1E-54	37.4	90.7	WP_081617226	2.6E-67	51.7	90.6	WP_018952914	1.9E-120	47.9	98.4	WP_026289648	4.5E-140	58.7	99.1	WP_018952912
Marinospirillum celere DSM 18438	4.1E-13	27.6	85.7	WP_091965170												
Thioalkalivibrio sp. ALSr1	1.5E-44	35.2	91.6	WP_019643168	1.7E-58	46.2	83.9	WP_019643171	4E-97	42.4	96.7	WP_019643172	2.8E-126	54.5	99.5	WP_019643173
Marinospirillum alkaliphilum DSM 21637	0.000053	26.2	71.7	WP_072327059												
Thioalkalivibrio sp. ALMg11	1.3E-56	38.5	94.4	WP_081617373	2.5E-70	53.6	88	WP_018950169	1.1E-110	45.8	98.4	WP_026289326	2.9E-139	60.2	98.6	WP_018950167
Thioalkalivibrio sp. AKL19	4.9E-51	37	89.3	WP_024327965	4E-68	52.4	89.3	WP_024327970	5.3E-118	46.2	98.4	WP_024327971	1.1E-143	60.8	99.3	WP_024327972
Thioalkalivibrio sp. ARh3	5.6E-47	35.5	92.3	WP_018864018	2.8E-69	51.8	92.3	WP_018864016	1.4E-115	45.8	98.4	WP_018864015	8.5E-139	59.4	98.9	WP_018864014
Halorhodospira halophila SL1	8.1E-62	41.4	85	WP_011813169												
Thioalkalivibrio sp. ALM2T	1.5E-60	39.3	86.9	WP_019593326	2.8E-69	54.9	89.3	WP_019593325	1.3E-116	47.1	97.3	WP_081622650	1.5E-140	60.6	98.4	WP_019593323
Thioalkalivibrio sp. ALR17-21	2E-76	43	88	WP_018946175	1E-70	45	91	WP_018946178	1.1E-104	42.2	98.4	WP_024329952	1.5E-127	55.3	99.8	WP_024329951
Thioalkalivibrio sp. AKL17	2E-76	43	91	WP_018946175	1E-70	45	98.4	WP_018946178	2E-104	43.3	99.8	WP_018946179	3.9E-128	55.5	99.8	WP_018946180
Thiocapsa sp. KS1	6E-41	52	74	WP_093186005	3.9E-71	50	95	WP_093186003	3.6E-82	40.5	96.5	WP_093186000	8.3E-110	52.3	98.4	WP_093185998
Candidatus Muproteobacteria bacterium RBG 16 65 34	5.4E-50	36.9	85.3	OGI47526												
Ectothiorhodospira sp. PHS-1	7.4E-23	28.2	87.1	WP_083838794	2.9E-66	50.4	85.6	WP_008932019	6.3E-95	42.7	97.5	WP_083838793	4.4E-111	49	99.5	WP_040405412
Azoarcus sp. CIB	1.2E-12	27.5	82.7	WP_050415006	8.7E-71	53.4	84.6	WP_083446883	3.3E-96	42.2	96.9	WP_050415007	1.3E-131	56.5	98.9	WP_050415008
Azoarcus tolulyticus ATCC 51758	6.5E-11	29.8	62.4	WP_076602991	7.3E-70	53.4	84.6	WP_084205107	2.3E-97	42.4	96.9	WP_084205117	4.5E-132	56.5	98.9	WP_076602989
Rhodospirillales bacterium RIFCSPLOWO2 12 FULL 67 15																
Sulfuricella denitrificans skB26^{T}	2.3E-40	33.4	83.6	WP_021035848	3.5E-64	45.9	92.3	WP_009207746	1.1E-99	42.4	97.1	WP_051338868	6.9E-133	58.6	99.3	WP_009207748
Betaproteobacteria bacterium RIFCSPLOWO2 12 FULL 62 58	1.8E-37	33.6	87.4	OGA55943	2.8E-54	35.9	87.9							0.21		OGA55939
Gallionellales bacterium RIFCSPLOWO2 02 FULL 59 110	2.8E-54	35.9	87.9	OGS99565	2E-67	49.6	86	OGS99592	2.6E-104	43.5	97.7	OGS99564	2E-135	57.7	98.6	OGS99563
Gammaproteobacteria bacterium RIFOXYD12 FULL 61 37					8.1E-69	50.6	87.3	OGT88203	5.7E-96	41.5	96.9	OGT88204	3.9E-128	56.8	98.9	OGT88205
Gammaproteobacteria bacterium RIFOXYA12 FULL 61 12	4.2E-34	37	65.2	OGT93885	8.1E-69	50.6	87.3	OGT93886	5.7E-96	41.5	96.9	OGT93887	3.9E-128	56.8	98.9	OGT93888
Betaproteobacteria bacterium CG2 30 59 46	3.1E-37	32.9	84.3	OIP17343	5.4E-65	44.5	94.6	OIP17356	4.8E-95	41.8	97.1	OIP17344	2.9E-131	58.4	99.3	OIP17345
Magnetospirillum magnetotacticum MS-1	3E-40	33.7	82.9	KIM00422	9E-60	44.1	90.3	KIM00423	2.4E-54	31.2	96.7	KIM00424	1E-69	0.38	0.97	KIM00425
Betaproteobacteria bacterium CG2 30 68 42	2E-28	34	65.2	OIP08970	2.4E-60	46.9	85.6	OIP09017	7.4E-104	45.8	96.7	OIP08969	2.2E-134	57.3	99.8	OIP08968
Rhodocyclales bacterium CG 4 10 14 3 um filter 68 10	9E-29	34	65.2	PJA58271	2.1E-61	45.6	91.6	PJA58284	8E-135	46	96	PJA57415	7.4E-135	57.6	99.8	PJA58273
Deltaproteobacteria bacterium GWC2 42 11																
Hydrogenophilales bacterium CG18 big fil WC 8 21 14 2 50 58 12	**0**	**0**	**0**	**PIQ11013**	0	0	0	PIQ11027	0	0	0	PIQ11012	0	0	0	PIQ11011
Candidatus Methanoperedens nitroreducens ANME-2d					7.2E-49	38.2	85.6	KCZ73332	6.7E-20	31.8	46.7	KCZ73331	6E-18	33	55	KCZ73330
Candidatus Methanoperedens sp. BLZ1					1E-61	38	84	KPQ43592	1E-20	27	94	KPQ43591	3E-14	28	45	KPQ43590
Candidatus Methanoperedens sp. BLZ2																
Rhodocyclaceae UBA4043	5.7E-31	29.8	89.3		8.5E-66	45.5	92.6		2.8E-103	43.5	98		3.2E-138	60.4	98.4	
Methanoperedens sp. UBA453					3E-16	33	55		2.5E-49	38.2	85.6		2.3E-20	31.8	46.7	
Candidatus Methanoperedens sp. Ru					7.2E-49	36.9	87		8.2E-18	27.5	46.7		5E-14	0.33	0.55	
Sulfuritortus calidifontis J1A^{T}					2.4E-68	48.2	86		3E-105	44.1	97.1		2.4E-141	60.1	98.6	
Strain J5B	6.7E-32	33.3	68.5		1.8E-68	46.1	84.9		3.4E-96	42.2	98		2.9E-139	57.9	100	
Strain M52	1.9E-52	35.4	82.2		3.4E-61	41.7	92.3		3E-83	37.2	97.1		9.4E-132	53.5	99.5	

Bold letters correspond to partial genes.

Table 3-3. Characteristic of the 46 genomes containing an *arxXSRB'ABCD* –like gene cluster similar to that of *A. ehrlichii.*

Organism name	Size (Mb)	GC%	Proteins	Scaffolds	Reference	Location of isolation/ sample	Temperature sample (°C)	pH sample	Arsenic concentration	oxygen requirement
Halorhodospira halophila SL1	2.67845	68	2402	1	Raymond , J.C. Sistrom, W.R. 1967; Challacombe, JF et al. 2013	salt-encrusted mud on hypersaline, Summer lake (OR) USA	NA	NA	NA	obligate anaerob
Sulfuricella denitrificans skB26^{T}	3.21721	56.1162	3016	2	Kojima, H & Fukui, M. 2010; Watanabe, T et al. 2014	water from Lake Mizugaki. Japan	5	6.5	NA	microaerophilic
Ectothiorhodospira sp. PHS-1	2.94321	63.7	2616	114	Kulp, T et al. 2008; Zargar, K et al. 2012	Red pigmented biofilm on rock surfaces in a hot pool Paoha Island hot spring (CA). USA	43	9.4	Arsenite 92 µM; arsenate 9.7 µM	obligate anaerob
Magnetospirillum magnetotacticum MS-1	4.52393	63.6	4077	36	Blakemore, R. P. et al. 1979; Maratea, D. Blakemore, R P. 1981; Smalley, M et al. 2015	Muds of a freshwater swamp	NA	NA	NA	microaerophilic
Alkalilimnicola ehrlichii MLHE-1	3.27594	67.5	2858	1	Oremland, R et al. 2002; Hoeft, S. et al. 2007; Zargar, K et al. 2010	Anoxic bottom water of Mono Lake (CA). USA	20	9.8	200 µM	Facultative anaerob
Halomonas boliviensis LC1	4.1334	54.7	3685	48	Quillaguamán, J et al. 2004; Balderrama-Subieta, A & Quillaguamán, J. 2013	soil sample arouNA Laguna Colorada (red-coloured hypersaline lake). Bolivia	NA	NA	NA	heterotrophs. aerobic
Thiocapsa sp. KS1	10.3826	57.2	4432	43	Schott, J et al. 2010	sewage sludge of the municipal sewage treatment plant at Konstanz, Germany.	NA	neutral	NA	facultative anaerob
Marinospirillum alkaliphilum DSM 21637	3.13755	54.6	2854	22	Zhang, W et al. 2002	Haoji soda lake, Mongolia. China	NA	9.5	NA	aerobic
Marinospirillum celere DSM 18438	3.08356	50.4	2829	12	Namsaraev, Z et al. 2009	combined water– sediment slurry sample taken from a hot spring on Paoha island on Mono Lake (CA). USA	40	9.3	200 µM**	aerobic
Thioalkalivibrio nitratireducens DSM 14787	4.00235	66.5	3532	1	Sorokin, D. et al. 2003	Sediment from hypersaline Lake Fazda, soda lake in the Wadi Natrun. Egypt	NA	10	NA	facultative anaerob
Azoarcus sp. CIB	5.25703	65.8	4603	1	Fernández-Llamosas, H et al. 2014; Martin-Moldes, Z et al. 2015;	From a DSMZ 12184 culture, supposed to be an Azoarcus sp. strain M3, isolated from a diesel fuel-contaminated aquifer at Menziken. Switzerland	NA	NA	NA	facultative anaerob
Ectothiorhodospira sp. BSL-9	3.55008	63	3103	1	Hoeft McCann, S et al. 2017; Hernandez-maldonado, J. 2016	water collected in the shallow swamp region of Big Soda Lake	NA	9.7	25 µM	anaerob
Halomonas sp. BC04	5.9211	62.4	4212	433	https://www.ncbi.nlm.nih.gov/bioproject/221128	Mono Lake, (CA) USA	NA	NA	200 µM**	aerobic
Azoarcus tolulyticus ATCC 51758	5.11039	66.4	4539	36	Zhou, J et al. 1995	Petroleum-contaminated freshwater aquifer sediment, 24 to 25 m deep (MI). USA	NA	NA	NA	facultative anaerob
Thioalkalivibrio thiocyanodenitrificans ARhD 1	3.74665	64.8	3450	3	Sorokin, D et al. 2004; Berben, T et al. 2015	Two mixed sediment samples from the soda lakes in Wadi Natrun (Egypt) and Kulunda steppe (Siberia, Russia)	NA	9.2 - 10.5	NA	facultative anaerob
Thioalkalivibrio sp. ARh3	2.7882	66.2	2626	15	Sorokin, D et al. 2001; Sorokin, D et al. 2002; Sorokin, D et al. 2011	Soda Lake in Kenya, Africa	NA	9.5 - 11.0	NA	facultative anaerob
Thioalkalivibrio sp. ALSr1	2.92881	67.5	2664	25	Sorokin, D et al. 2011	Searles Lake, (CA) USA	NA	9.8*	4000 µM*	facultative anaerob
Thioalkalivibrio sp. ALR17-21	2.99896	67.4	2756	65	Sorokin, D et al. 2011	Kulunda Steppe, Altai, Russia	NA	9.3 - 10.6	NA	facultative anaerob
Thioalkalivibrio sp. ALMg11	3.03779	66	2832	18	Sorokin, D et al. 2011	North-eastern Mongolia	NA	9.2 - 10.5	NA	facultative anaerob
Thioalkalivibrio sp. ALM2T	2.92411	65.9	2668	26	Sorokin, D et al. 2013	Mono Lake, (CA) USA	NA	9.7	200 µM**	facultative anaerob
Thioalkalivibrio sp. AKL17	2.80322	67.6	2581	18	Sorokin, D et al. 2011	Kulunda Steppe, Altai, Russia	NA	9.3 - 10.6	NA	facultative anaerob

Cont. Table 3-3. Characteristic of the 46 genomes containing an *arxXSRB'ABCD* –like gene cluster similar to that of *A. ehrlichii.*

Organism name	Size (Mb)	GC%	Proteins	Scaffolds	Reference	Location of isolation/ sample	Temperature sample (°C)	pH sample	Arsenic	oxygen requirement
Thioalkalivibrio denitrificans ALJD	3.63935	64.6	3295	167	Sorokin, D et al. 2001	Soda Lake in Kenya, Africa	NA	9.5 - 11.0	NA	facultative anaerob
Thioalkalivibrio sulfidiphilus ALJ17	3.05302	65.2	2834	69	Sorokin, D et al. 2012	Soda Lake in Kenya, Africa	NA	9.5 - 11.0	NA	facultative anaerob
Thioalkalivibrio sp. AKL19	2.88491	66.8	2639	18	Sorokin, D et al. 2011	Kulunda Steppe, Altai, Russia	NA	9.3 - 10.6	NA	facultative anaerob
Halomonas chromatireducens AGD 8-3	3.97365	62.8	3509	1	Shapovalova, A et al. 2009; Gotovtsev, P et al. 2016	mixed soil sample of soda solonchak from the Kulunda steppe (Altai krai) Russia	NA	alkaline	NA	facultative anaerob
Halomonas sp. A3H3	5.64388	54.8016	10291	33	Koechler, S et al. 2013	harbor sediments of l'Estaque in the south of France	NA	NA	653 µg/liter in sediment interstitial water and 165 mg/kg in solid phase	facultative anaerob
Nitrincola lacisaponensis 4CA	3.41213	52.1	3055	43	Dimitriu, P t al. 2005	Soap Lake in Grant County (WA) USA	NA	9.8	NA	facultative anaerob
Candidatus Muproteobacteria bacterium RBG_16_65_34†	2.03354	64.8	1828	109	Anantharaman, K. et al. 2016	sediment. suboxic/anoxic aquifer adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.11	NA	NA
Rhodospirillales bacterium RIFCSPLOWO2_12_FULL_67_15†	2.14886	65.7	1751	290	Anantharaman, K. et al. 2016	Groundwater adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.28	NA	NA
Betaproteobacteria bacterium RIFCSPLOWO2_12_FULL_62_58†	5.87674	61.8	5370	277	Anantharaman, K. et al. 2016	Groundwater adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.28	NA	NA
Gallionellales bacterium RIFCSPLOWO2_02_FULL_59_110†‡	2.80931	59	2496	112	Anantharaman, K. et al. 2016	Groundwater adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.28	NA	NA
Gammaproteobacteria bacterium RIFOXYD12_FULL_61_37†	3.46975	61.3	3093	57	Anantharaman, K. et al. 2016	Grounwater O2 injection, adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.39	NA	NA
Gammaproteobacteria bacterium RIFOXYA12_FULL_61_12†	3.05181	61.1	2617	218	Anantharaman, K. et al. 2016	Grounwater O2 injection, adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.67	NA	NA
Betaproteobacteria bacterium CG2_30_59_46†	2.29362	58.8	2077	425	Probst, A. et al, 2017	water from Crystal Geyser system in (UT) USA	17°C***	circumneutral***	NA	NA
Betaproteobacteria bacterium CG2_30_68_42†	3.09877	68.2	2853	95	Probst, A. et al, 2017	water from Crystal Geyser system in (UT) USA	17°C***	circumneutral***	NA	NA
Rhodocyclales bacterium CG_4_9_14_3_um_filter_68_10†	2.96182	68.1	2791	185	Probst, A. et al, 2017	water from Crystal Geyser system in (UT) USA	17°C***	circumneutral***	NA	NA
Deltaproteobacteria bacterium GWC2_42_11†	1.75104	42	1598	73	Anantharaman, K. et al. 2016	Grounwater acetate injection adjacent to the Colorado River, near Rifle, (CO) USA	NA	7.39	NA	NA
Hydrogenophilales bacterium CG18_big_fil_WC_8_21_14_2_50_58_12†‡	2.90516	58.1	2783	314	Probst, A. et al, 2017	water from Crystal Geyser system in (UT) USA	17°C***	circumneutral***	NA	NA
Candidatus Methanoperedens nitroreducens ANME-2d	3.20339	43.2	3254	10	Haroon, M et al. 2013	laboratory scale bioreactor enriched with freshwater sediment and anaerobic wastewater sludge in Brisbane, Australia	NA	NA	NA	Anaerobic
Candidatus Methanoperedens sp. BLZ2†	3.73558	40.3	3790	85	Raghoebarsing, A et al. 2006; Berger, S et al. 2017	anoxic bioreactor inoculated with the effluent of the enrichment with sediment of the Twentekanaal. The Netherlands	NA	NA	NA	NA
Candidatus Methanoperedens sp. BLZ1†	3.73847	40.2	4528	514	Raghoebarsing, A et al. 2006	anoxic bioreactor inoculated with the effluent of the enrichment with sediment of the Twentekanaal. The Netherlands	NA	NA	NA	NA

Cont. Table 3-3. Characteristic of the 46 genomes containing an *arxXSRB'ABCD* –like gene cluster similar to that of *A. ehrlichii.*

Organism name	Size (Mb)	GC%	Proteins	Scaffolds	Reference	Location of isolation/ sample	Temperature sample (°C)	pH sample	Arsenic	oxygen requirement	N50 (bp)
Candidatus Methanoperedens sp. UBA453	3.08605	43.3	**3348**	72	Parks, D et al. 2017	Metagenomic sequencing of bioreactor performing anaerobic oxidation of methane coupled to nitrate reduction.	NA	NA	NA	NA	47995
Rhodocyclaceae bacterium UBA4043	2.8851	67.8	**2828**	218	Parks, D et al. 2017	Metagenomic sequencing of coal cuttings from Coal bed Methane well site	NA	NA	NA	NA	21308
Candidatus Methanoperedens sp. Ruh	3.59264	40.4	**3983**	146	BioSample: SAMN07483404	Bioreactor mimicking brackish/coastal conditions	NA	NA	NA	NA	35785
Sulfuritortus calidifontis J1A^{T}	**2.720636**	**65.5**	**2699**	1	Kojima, H. et al. 2017	Microbial Mat in Jozankei Hot spring. Japan	43 ¶	7.8 ¶	**3.20 mg/L**	facultative anaerob	circular genome
Strain J5B	**2.868032**	**62.4**	**2557**	3	not published	Microbial Mat in Jozankei Hot spring. Japan	43¶	7.8¶	**3.20 mg/L**	-	3 contig
Strain M52	**2.72384**	**63.6**	**2719**	**19**	not published	Microbial Mat in Jozankei Hot spring. Japan	43 ¶	7.8 ¶	**3.20 mg/L**	facultative anaerob, microerophilic	**549140**

* Arsenic concentration published in Oremland, R. et al. 2017. (64)

** Arsenic concentration published in Oremland, R et al. 2002.(3)

*** pH and temperature published in Emerson, J et al. 2016.(65)

¶ pH and temperature published in Kojima et al., 2017.(15)

†Correspond to genomes derived from metagenomes.

‡Representative genome of three identical assemblies.

Abbreviation. NA, not available.

Bold letters correspond to data reported in this study.

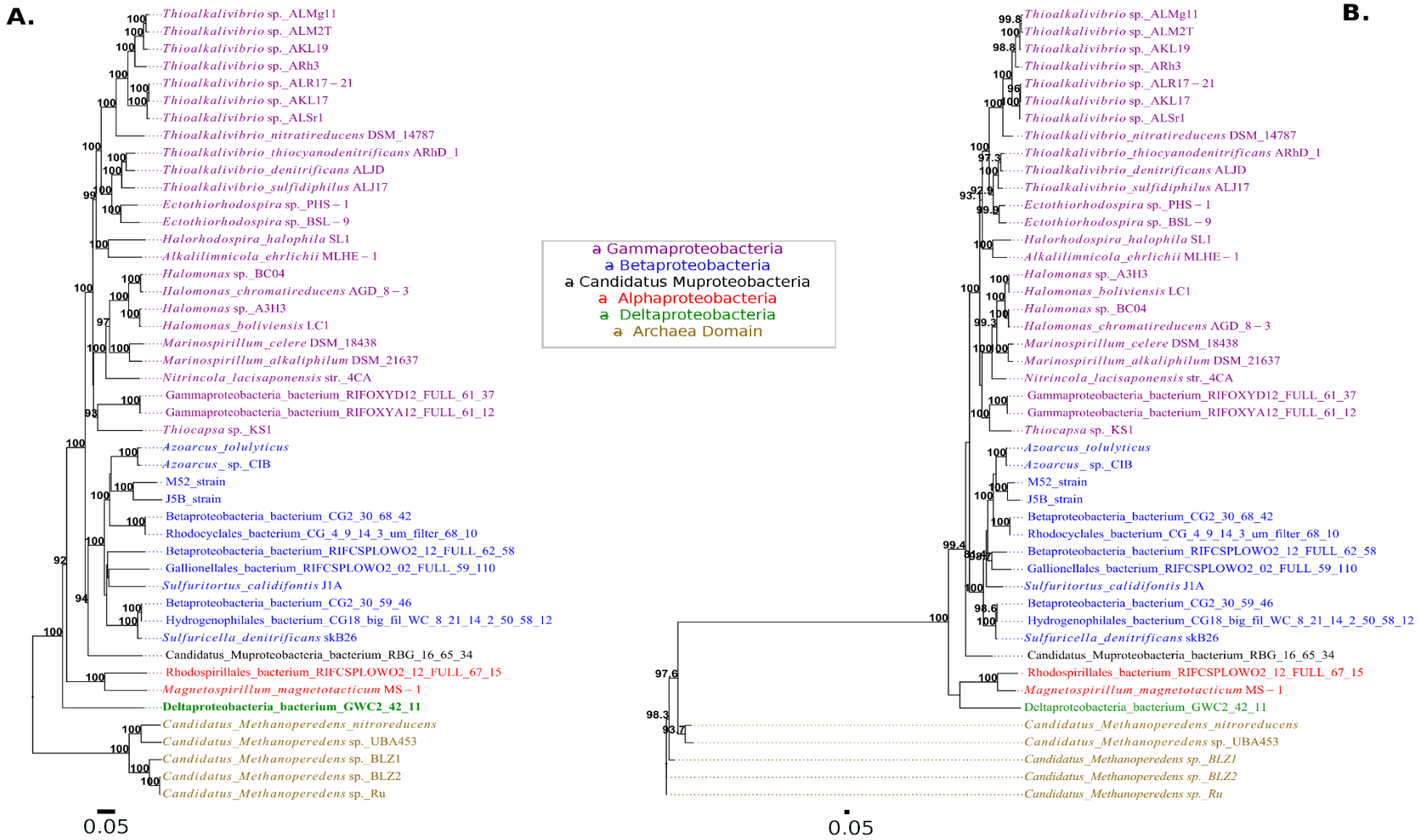

Figure 3-1. Phylogenetic tree based on the concatenated amino acid sequences encoded by 10 ribosomal proteins described in table 3-1. **A**. Neighbor Joining phylogeny with p-distance model. **B.** Maximum likelihood phylogeny with LG substitution model. Bar = number of substitutions per sequence position; Support values greater than 70% are indicated in bold.

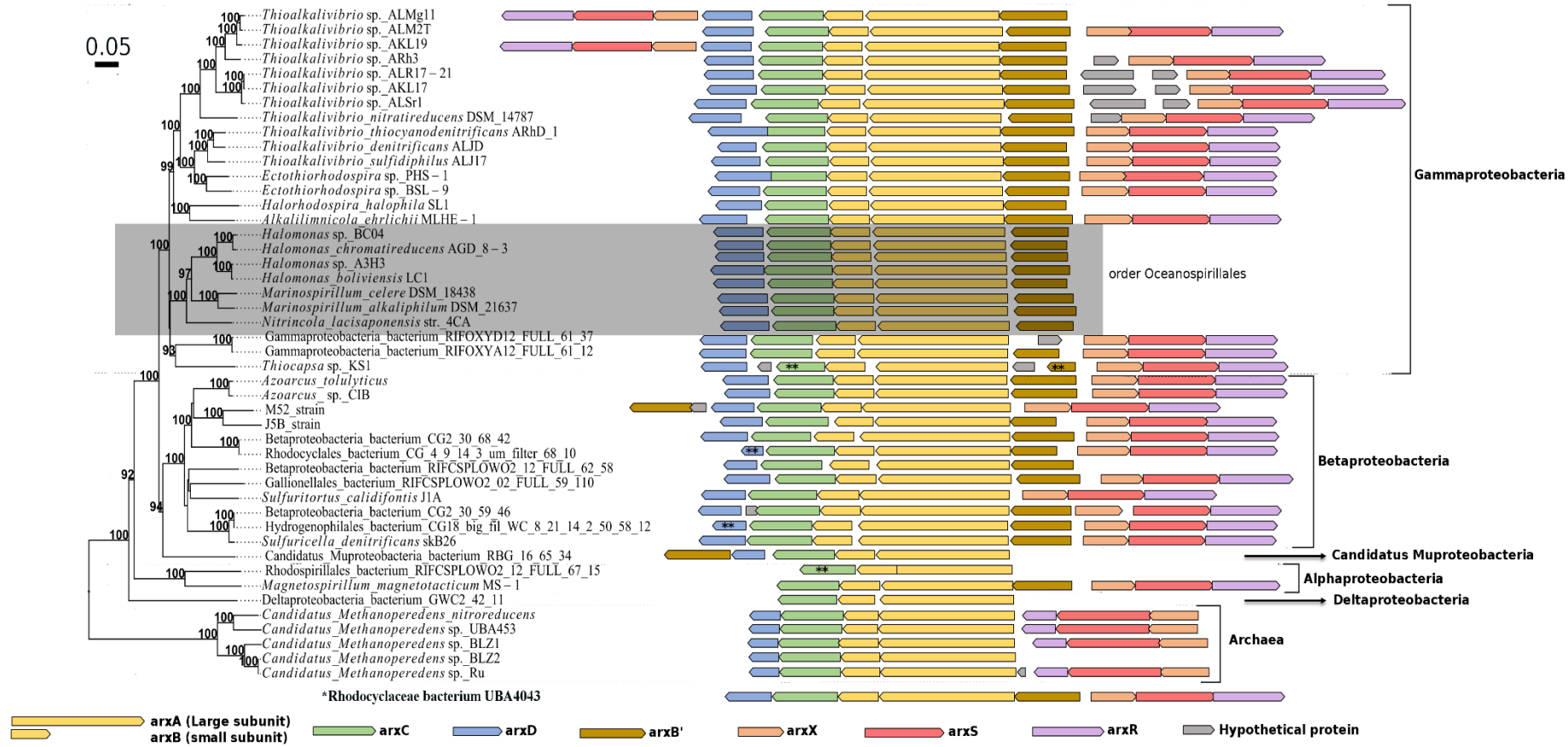

Figure 3-2. Phylogenetic tree based on the concatenated amino acid sequences encoded by 10 ribosomal proteins. *arx* homologous genes arrangement is shown in front of the organism name. Tree = Neighbor Joining with p-distance model; bar = number of substitutions per sequence position; Support values greater than 70% are indicated in bold. Double asterisk (**) represent protein annotated as partial protein. Rhodocyclaceae bacterium UBA4043 was not included in the tree due to absence of the majority of ribosomal proteins and 16S rRNA gene. Order Oceanospirillales representatives and correspondent *arx* gene cluster are highlighted in gray box.

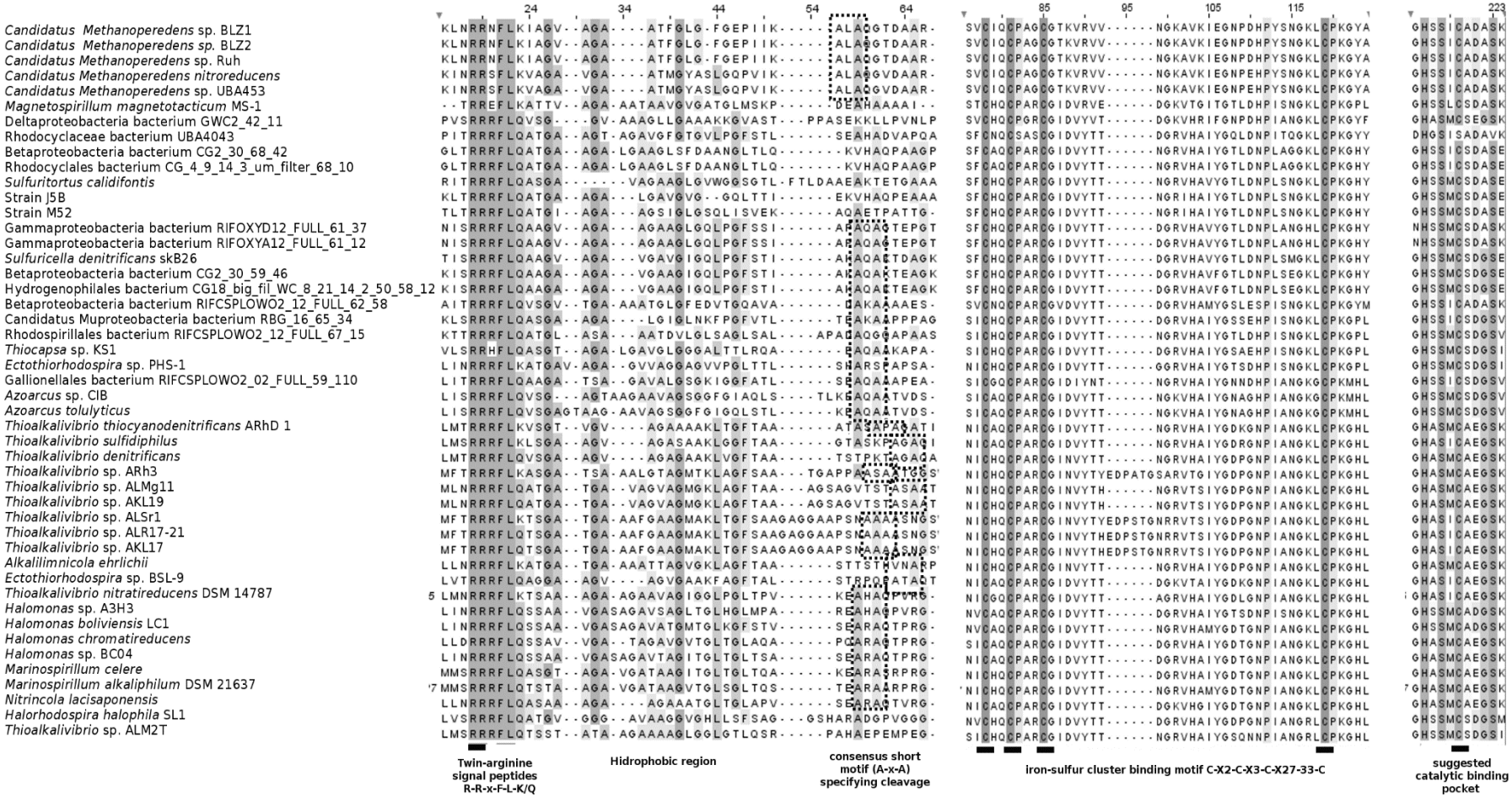

Figure 3-3. The conserved motif of the putative subunit A of the arsenite oxidase ArxAB. The conserved cysteine residues predicted to coordinate to the Fe in the iron-sulfur binding motif, and the conserved cysteine residue in the suggested catalytic binding domain predicted to coordinate to the molybdenum are underlined. Number in the upper part of the figure represents the position in the amino acid multi-sequence alignment. Note: Rhodocyclaceae bacterium UBA4043 possess a serine instead of a cysteine residue in the suggested catalytic binding pocket.

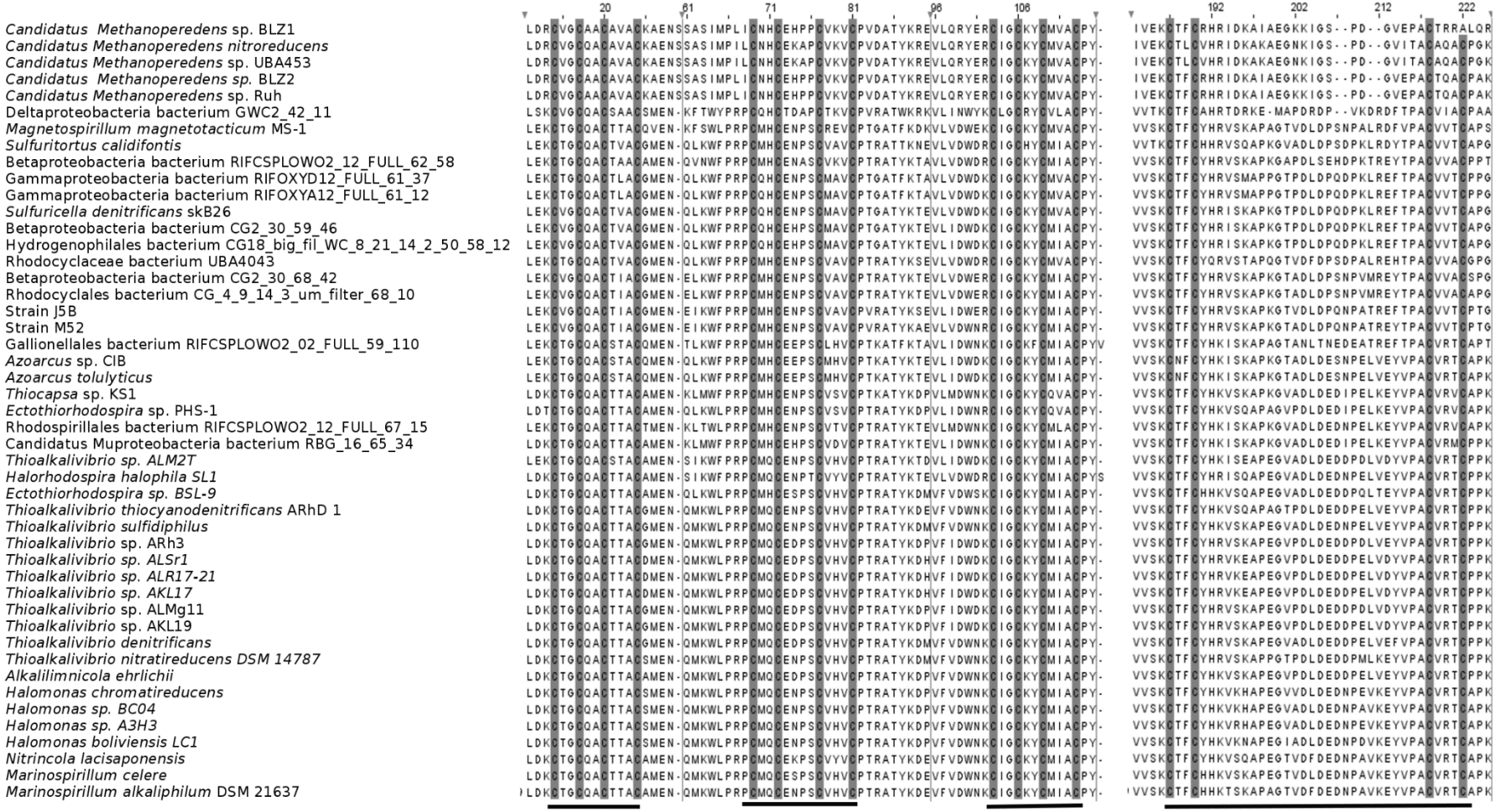

Figure 3-4. The conserved motif of the putative subunit B of the arsenite oxidase ArxAB. An iron-sulfur motif is underlined. The conserved cysteine of each iron-sulfur binding motif predicted to coordinate to the Fe are shaded in gray. Note: absence of one cysteine residue of Candidatus Methanoperedens sp. BLZ1. Number in the upper part of the figure represents the position in the amino acid multi-sequence alignment.

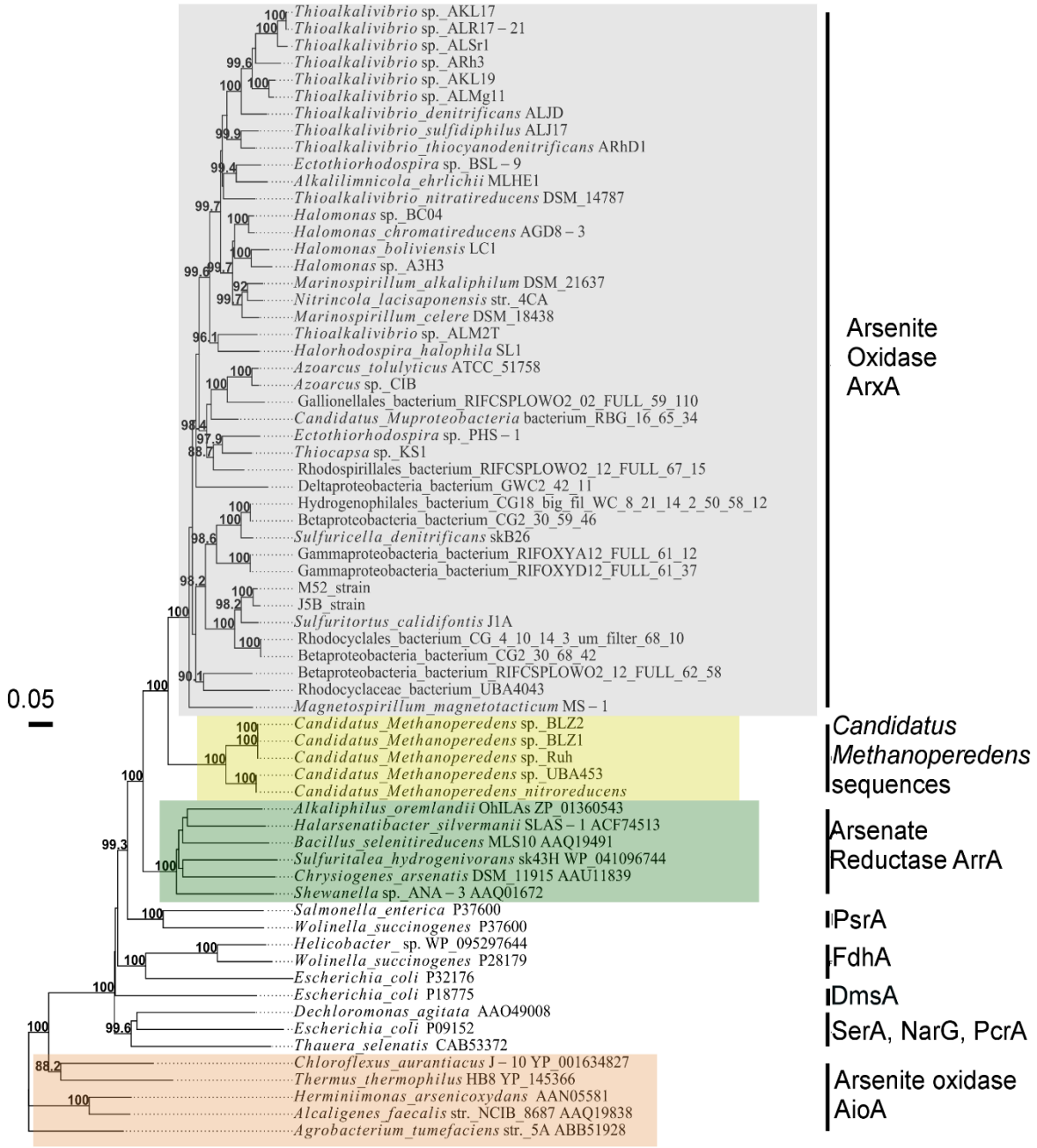

Figure 3-5. Phylogenetic tree based on the ArxA-type arsenite oxidase, and the putative subunit A of others molybdoenzymes of the DMSO reductase family. Abbreviation: PrsA, polysulfide reductase. FdhA, formate dehydrogenase. DmsA, dimethyl sulfoxide reductase. SerA, selenite reductase. NarG, respiratory nitrate reductase. PcrA, perchlorate reductase. Gaps were allowed in the multi-sequence alignment. The phylogenetic inference was neighbor-joining for tree reconstruction with distance criterion. The tree was rooted on Arsenite oxidase AioA. Bootstrap values (1000 resampling) above 70% are shown.

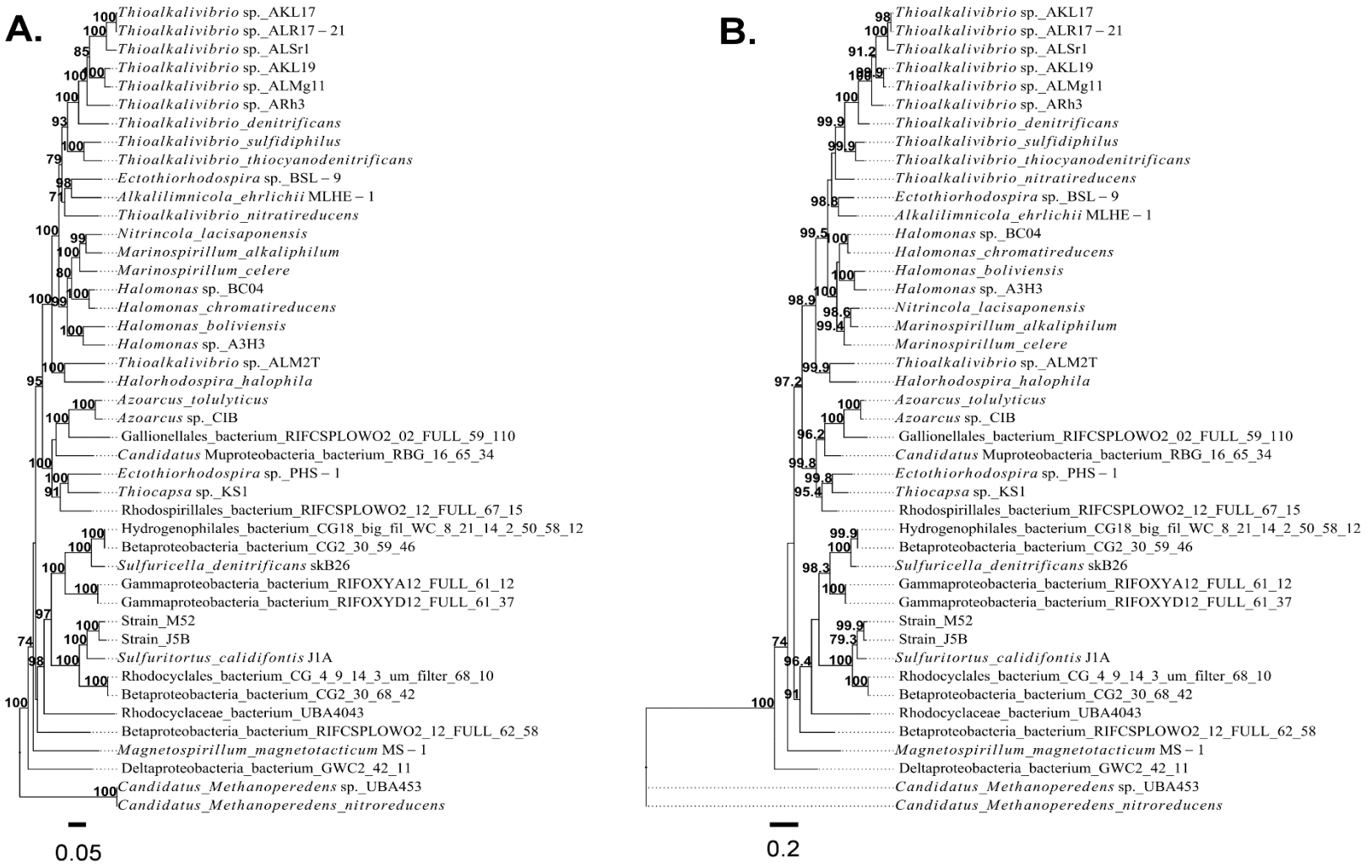

Figure 3-6. Phylogenetic tree based on ArxA-type arsenite oxidase. **A**. Neighbor Joining phylogeny with p-distance model. **B**. Maximum likelihood phylogeny with LG substitution model. Support values above 70 are shown. *Candidatus Methanoperedens* sequences were used as outgroup.

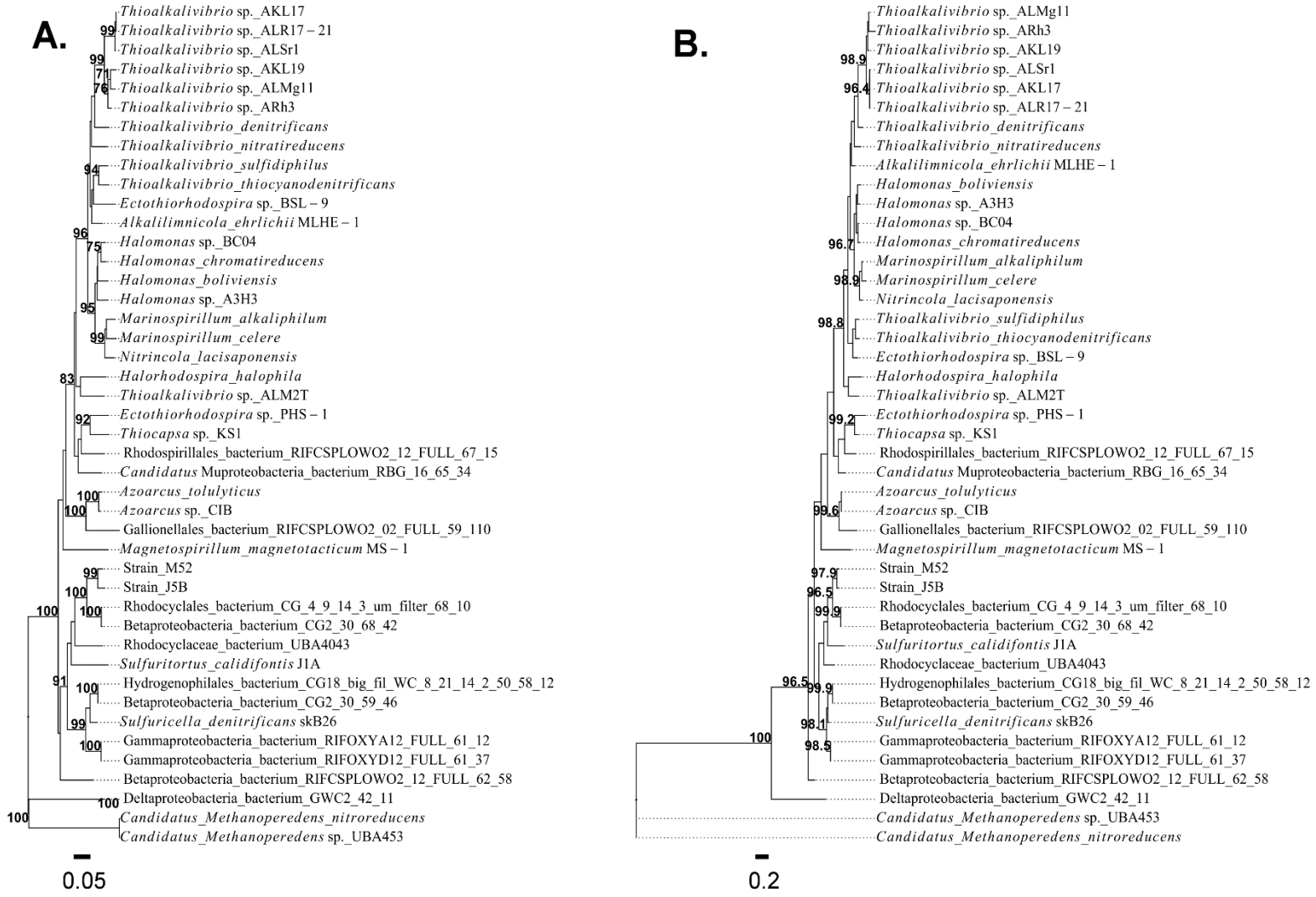

Figure 3-7. Phylogenetic tree based on ArxB-type arsenite oxidase. **A**. Neighbor Joining phylogeny with p-distance model. **B**. Maximum likelihood phylogeny with LG substitution model. Support values above 70 are shown. *Candidatus Methanoperedens* sequences were used as outgroup.

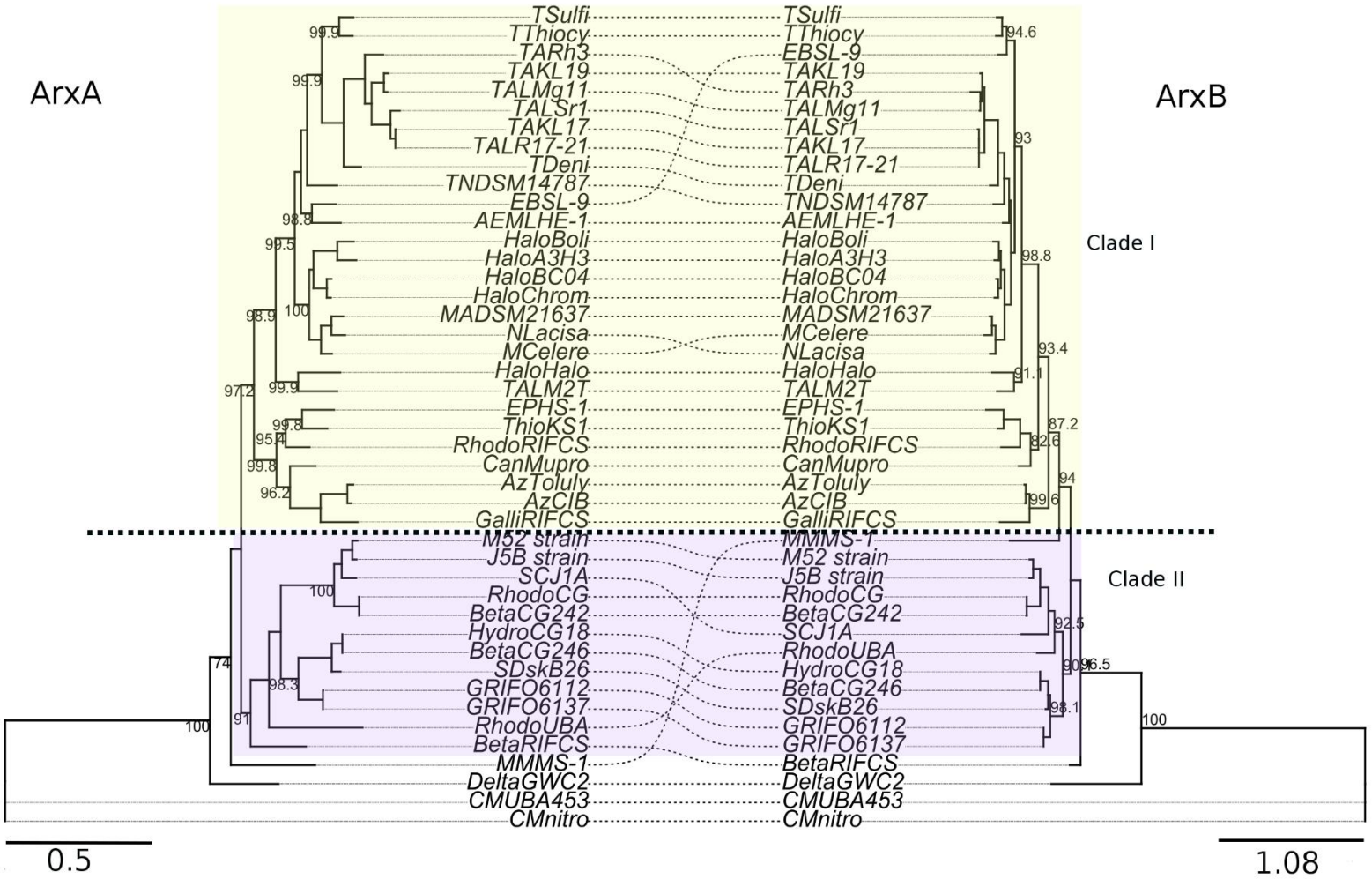

Figure 3-8. Co-phylogenetic tree based on the aminoacid sequences of ArxA-type and ArxB-type arsenite oxidase. PhyML Maximun-likelihood inferred tree with the LG substitution model. It was rooted on *Candidatus Methanoperedens* sequences. Support values above 70 are shown. Abbreviations referring to strain designations are present in Table 3-4.

Table 3-4. Assignation of strain abbreviation used in the Figure 3-8.

Organism name	Abbreviation
Alkalilimnicola ehrlichii MLHE-1	AEMLHE-1
Ectothiorhodospira sp. BSL-9	EBSL-9
Thioalkalivibrio nitratireducens DSM 14787	TNDSM14787
Halomonas sp. A3H3	HaloA3H3
Halomonas chromatireducens AGD 8-3	HaloChrom
Halomonas boliviensis LC1	HaloBoli
Thioalkalivibrio denitrificans ALJD	TDeni
Halomonas sp. BC04	HaloBC04
Nitrincola lacisaponensis 4CA	NLacisa
Thioalkalivibrio thiocyanodenitrificans ARhD 1	TThiocy
Thioalkalivibrio sulfidiphilus ALJ17	TSulfi
Marinospirillum celere DSM 18438	MCelere
Thioalkalivibrio sp. ALSr1	TALSr1
Marinospirillum alkaliphilum	MADSM21637
Thioalkalivibrio sp. ALMg11	TALMg11
Thioalkalivibrio sp. AKL19	TAKL19
Thioalkalivibrio sp. ARh3	TARh3
Halorhodospira halophila	HaloHalo
Thioalkalivibrio sp. ALM2T	TALM2T
Thioalkalivibrio sp. ALR17-21	TALR17-21
Thioalkalivibrio sp. AKL17	TAKL17
Thiocapsa sp. KS1	ThioKS1
Candidatus Muproteobacteria bacterium RBG_16_65_34	CanMupro
Ectothiorhodospira sp. PHS-1	EPHS-1
Azoarcus sp. CIB	AzCIB
Azoarcus tolulyticus	AzToluly
Rhodospirillales bacterium RIFCSPLOWO2_12_FULL_67_15	RhodoRIFCS
Sulfuricella denitrificans skB26^{T}	SDskB26
Betaproteobacteria bacterium RIFCSPLOWO2_12_FULL_62_58	BetaRIFCS
Gallionellales bacterium RIFCSPLOWO2_02_FULL_59_110	GalliRIFCS
Gammaproteobacteria bacterium RIFOXYD12_FULL_61_37	GRIFO6137
Gammaproteobacteria bacterium RIFOXYA12_FULL_61_12	GRIFO6112
Betaproteobacteria bacterium CG2_30_59_46	BetaCG246
Magnetospirillum magnetotacticum MS-1	MMMS-1
Betaproteobacteria bacterium CG2_30_68_42	BetaCG242
Rhodocyclales bacterium CG_4_9_14_3_um_filter_68_10	RhodoCG
Deltaproteobacteria bacterium GWC2_42_11	DeltaGWC2
Hydrogenophilales bacterium CG18_big_fil_WC_8_21_14_2_50_58_12	HydroCG18
Candidatus Methanoperedens nitroreducens	CMnitro
Candidatus Methanoperedens sp. BLZ1	CMBLZ1
Candidatus Methanoperedens sp. BLZ2	CMBLZ2
Strain J5B	J5B strain
Sulfuritortus calidifontis J1A^{T}	SCJ1A
Candidatus Methanoperedens sp. Ruh	CMRuh
Rhodocyclaceae bacterium UBA4043	RhodoUBA
Candidatus Methanoperedens sp. UBA453	CMUBA453
Strain M52	M52 strain

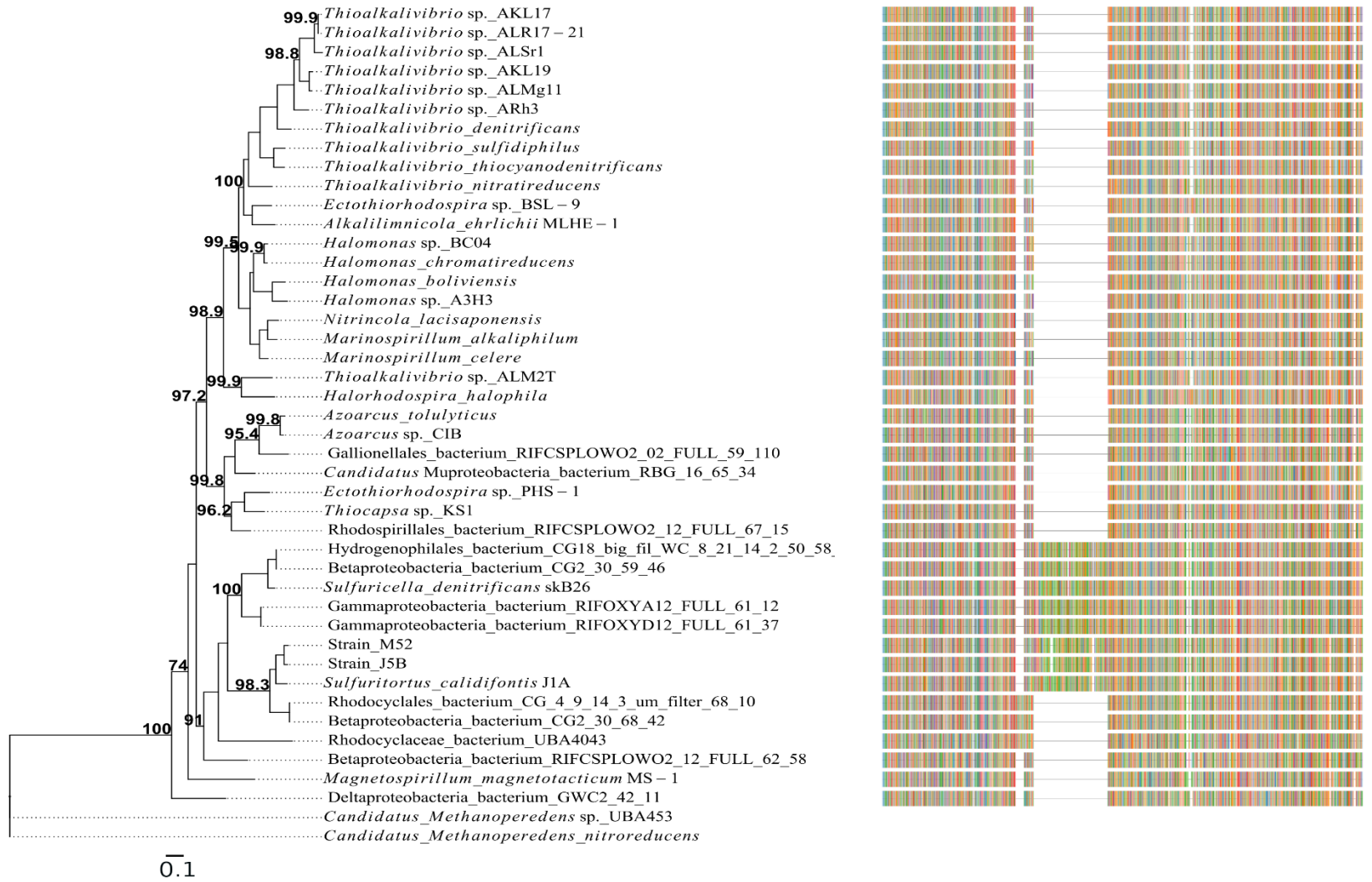

Figure 3-9. Phylogenetic analysis and multiple sequence alignment based on ArxA amino acid sequences. Tree: Maximum likelihood phylogeny with LG substitution model. Support values above 70 are shown.

Table 3-4. Phylogenetic affiliation of each OTU of the three clones libraries, based on the closest cultured relatives, determined according to the results of BLASTP search.

Clone library	OTU name	# clones	closest cultured relatives	Ident. %	QC	Accession number
JZK-1200	OTU1	7	*Sulfuricella denitrificans* skB26^T	58%	100%	WP_009207744
	OTU2	2	*Sulfuricella denitrificans* skB26^T	60%	100%	WP_009207744
	OTU3	1	*Sulfuricella denitrificans* skB26^T	59%	100%	WP_009207744
	OTU4	1	*Sulfuricella denitrificans* skB26^T	59%	100%	WP_009207744
	OTU5	1	*Sulfuricella denitrificans* skB26^T	59%	99%	WP_009207744
JZK900	OTU1	13	*Thioalkalivibrio* sp. ALM2^T	66%	100%	WP_019593327
	OTU2	10	*Halomonas* sp. BC04	68%	100%	WP_043513786
	OTU3	8	*Ectothiorhodospira* sp. BSL-9	60%	100%	WP_102041243
	OTU4	1	*Azoarcus* sp. CIB	81%	99%	WP_050415005
	OTU5	1	*Thioalkalivibrio* sp. ALM2T	65%	100%	WP_019593327
MZG900	OTU1	20	*Azoarcus* sp. CIB	80%	99%	WP_050415005
	OTU2	10	*Thioalkalivibrio* sp. ALM2T	62%	100%	WP_019593327
	OTU3	7	*Halomonas chromatireducens* AGD 8-3	62%	100%	WP_066449856
	OTU4	2	*Azoarcus tolulyticus* ATCC 51758	72%	100%	WP_076602992
	OTU5	1	*Azoarcus* sp. CIB	78%	100%	WP_050415005
	OTU6	1	*Azoarcus* sp. CIB	79%	99%	WP_050415005
	OTU7	1	*Thioalkalivibrio* sp. ALM2T	61%	100%	WP_019593327

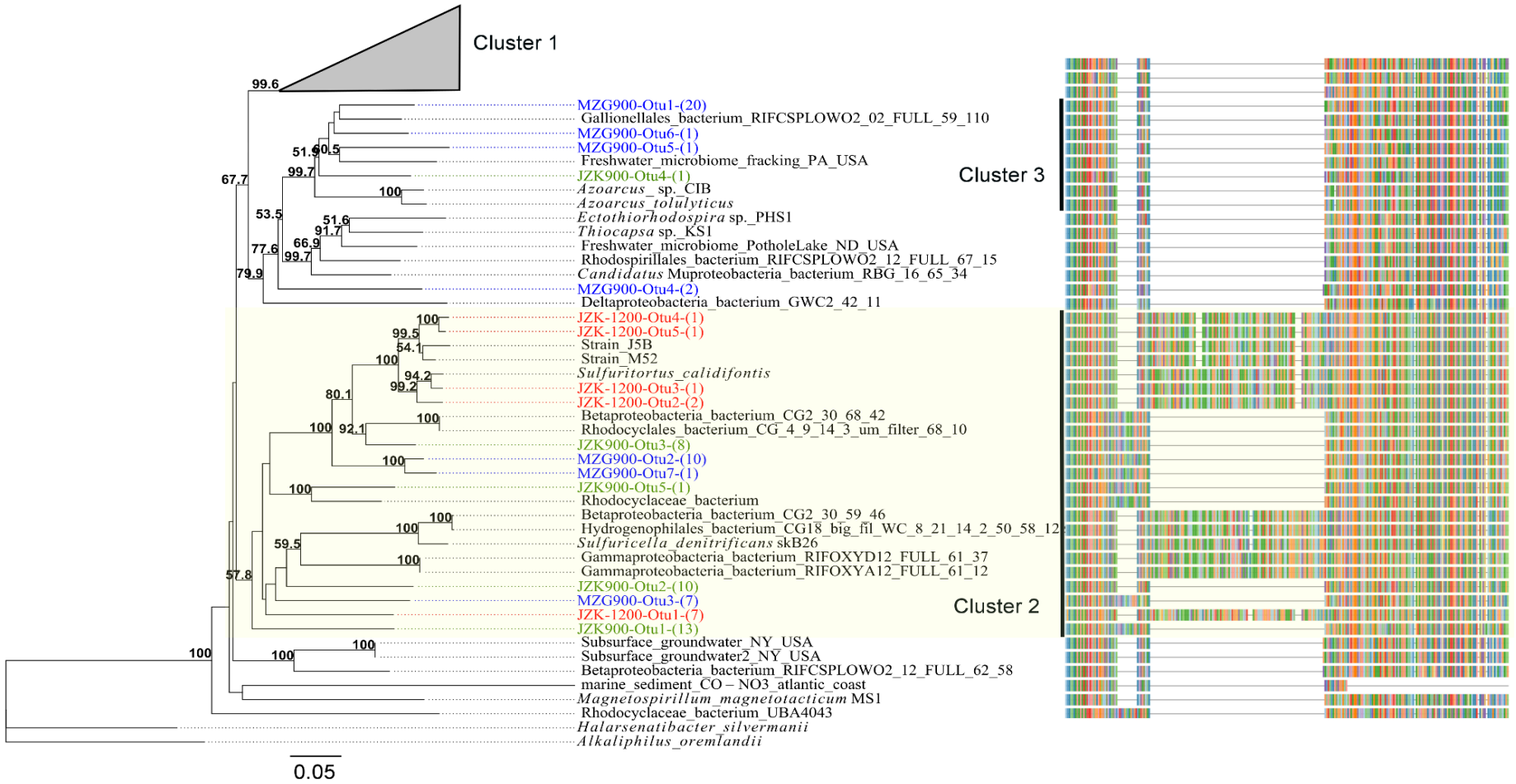

Figure 3-10. Expanded version of the phylogenetic analysis and multiple protein sequence alignment of ArxA representation, including OTUs from clone libraries, and complete ArxA sequences obtained from the genome analysis. Bootstrap values (1000 resampling) above 50% are shown. Shaded box corresponds to the OTUs and sequences within cluster 2, equivalent to the group of the same name in the ArxA phylogeny of chapter 2.

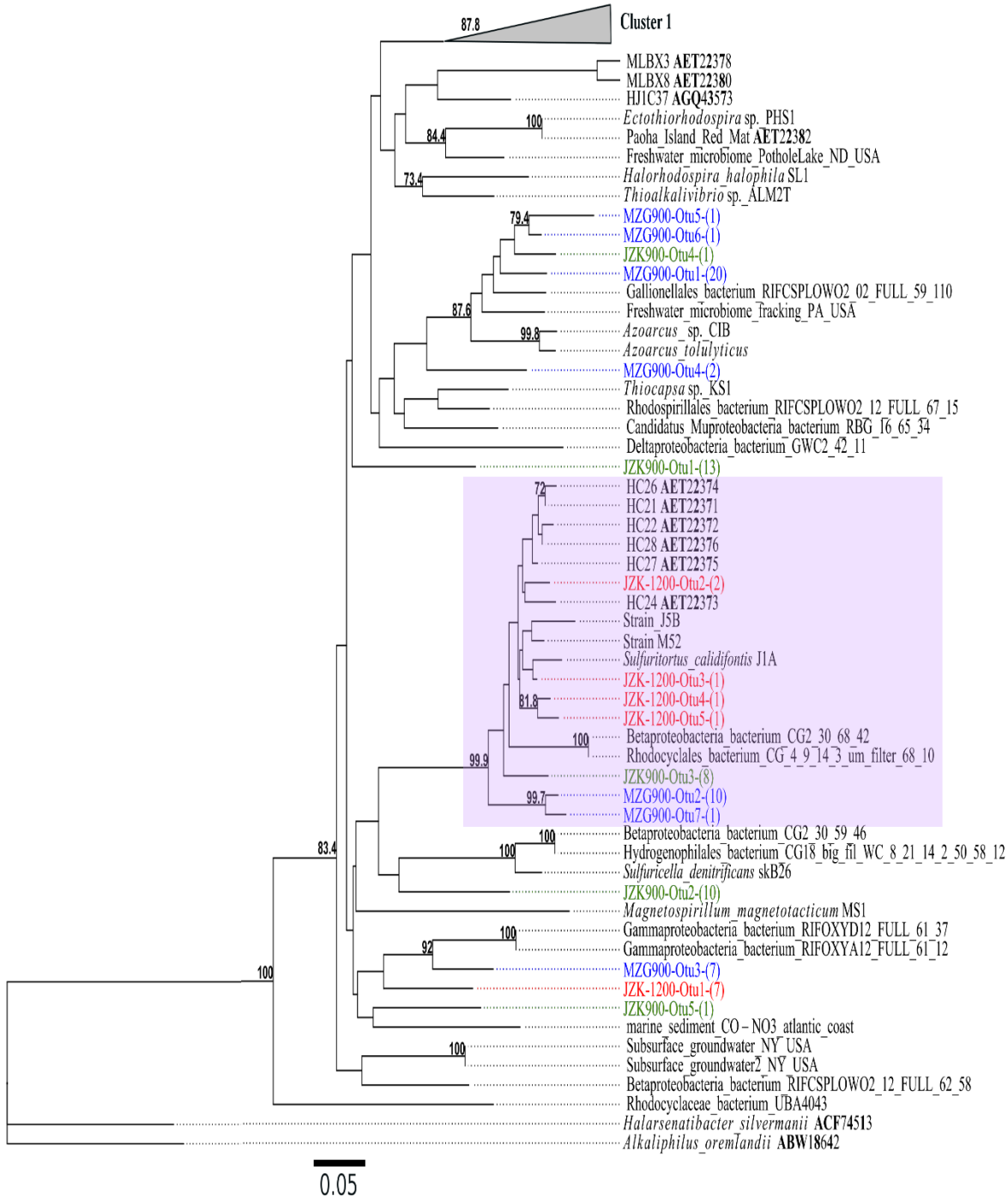

Figure 3-11. Phylogenetic tree based on ArxA sequences including OTUs sequences with and without insertion and partial ArxA clones of previous studies. Final dataset contains 92 amino acid positions for tree reconstruction. Tree: Neighbor-joining with the p-distance model. Rooted on: Arsenate reductase ArrA. Bootstrap values (1000 resampling) above 70% are shown. Red OTUs: Jozankei sequences holding an insertion (approx.106 aa). Green and blue OTUs: Jozankei and Mizugaki clones respectively, of sequences without a long insertion. Purple shaded box: strongly supported group comprising Jozankei isolated strains, OTUs with long and short insert and clones from Hot Creek Area of hot spring in California (USA).

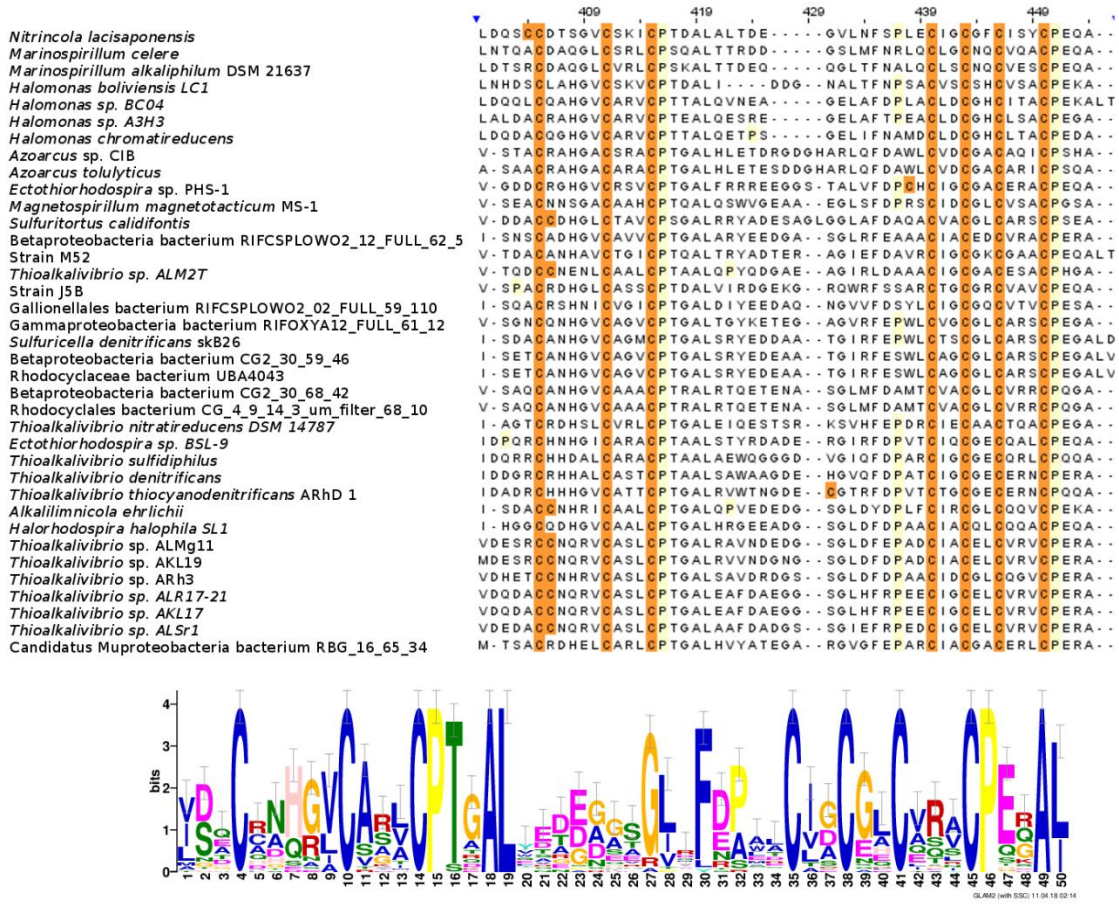

Figure 3-12. Conserved motif of the translated amino acid sequences of *arxB*'. Conserved cysteine are shaded in orange in the multi-sequence alignment. Sequences logo of the consensus sequences of the iron-sulfur motif C-X_5-C-X_3 –CP and C-X_2-C-X_2-C-X_3 –CP. The first motifs appears to have either lost one cysteine affecting its binding property, or bind to a 3Fe-4S centre instead of a 4Fe-4S centre.

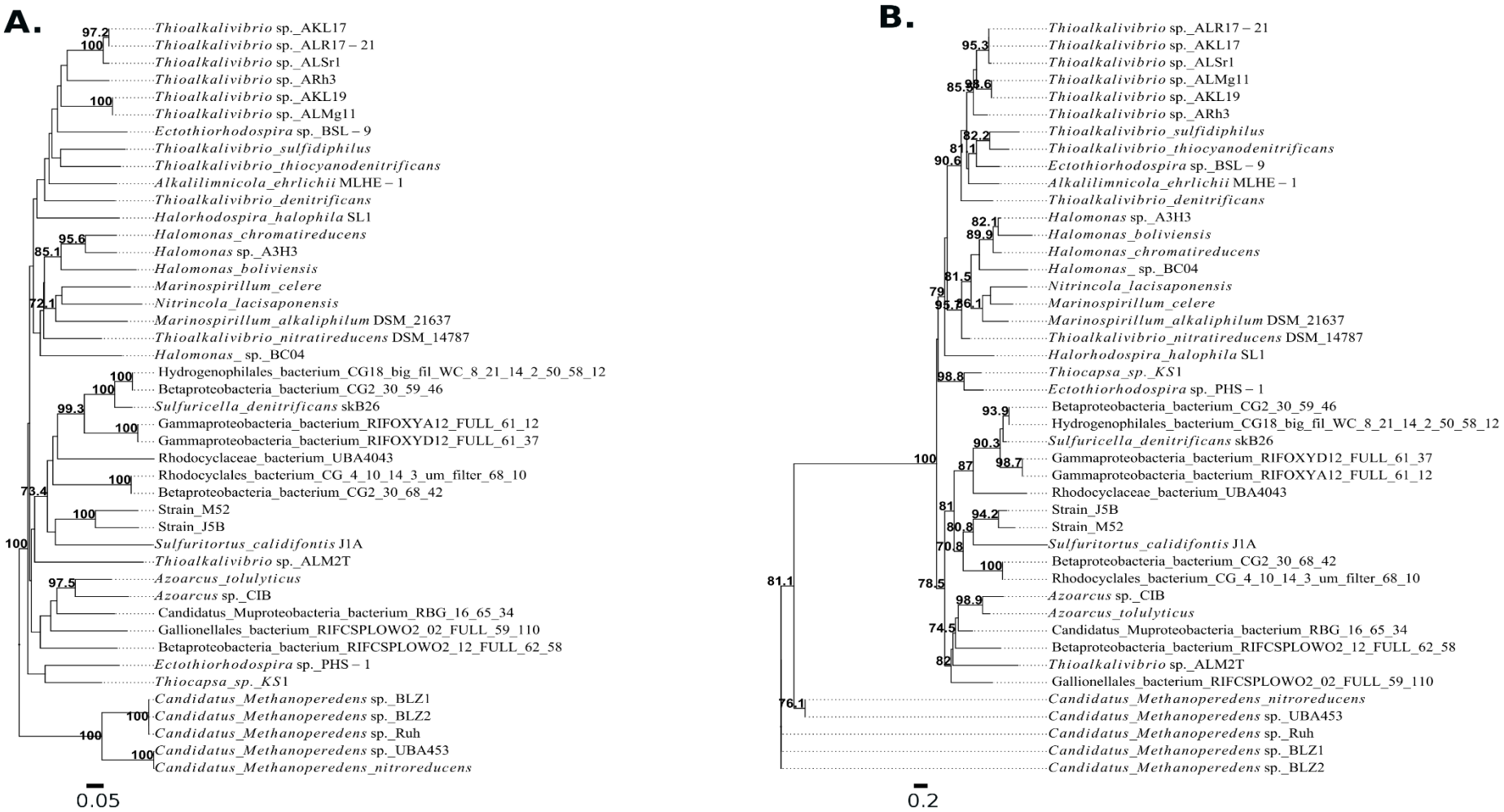

Figure 3-13. Phylogenetic tree based on the amino acid translation of *arx*D-like gene. **A**. Neighbor Joining phylogeny with p-distance model. **B**. Maximum likelihood phylogeny with LG substitution model. Support values above 70 are shown. *Candidatus Methanoperedens* sequences were used as outgroup.

Chapter 4

Arsenite oxidation by a newly isolated bacterium possessing *arx* genes.

4.1 Introduction

Despite the toxicity of the arsenic inorganic forms, arsenite and arsenate, the prokaryotes have evolved to use them as a bioenergetics substance, with enzymes that catalyze both redox reactions, arsenite oxidation, and arsenate reduction. The arsenite oxidation is the main studied process of both biochemical reactions, with most of the described arsenite-oxidizing bacteria as being heterotrophic (1). The first chemolithoautotrophically arsenite-oxidizing bacterium reported was the *Pseudomonas arsenitoxidans* in 1981, able to grow with oxygen as an electron acceptor, and carbon dioxide as the carbon source (2).

Two enzymes are involved in the arsenite oxidation coded by the *aioAB* and the *arxAB* genes (3). The *aioAB* genes were the first discovered of both arsenite oxidase coding genes, extensively reported and mainly related to aerobic conditions (4, 5). Evidence of anaerobic arsenite-oxidizing bacteria through AioAB is still limited (6). In contrast, the *arxAB* genes have been linked to anaerobic arsenite oxidation, detected first in the chemolithoautotrophically strain *Alkalilimnicola ehrlichii* MLHE-1, using arsenite as the electron acceptor (2002)(7, 8)(9, 10). To this date, a few more *arx* arsenite-oxidizing microbes had been isolated, all of which are related to anoxic conditions.

Arsenite dependent anoxygenic phototrophic growth was observed in *Ectothiorhodospira* sp. PHS-1, isolated from a red pigmented biofilm on Paoha Island in Mono Lake, California (11). Strain PHS-1 constituted the first suggested photoarsenotroph to employ an *arx*-like arsenite oxidase, due to the lack of *aio* genetic signature in the genome, and *arxA* gene expression after arsenite induction (10, 11). The disruption of the *arxA*-like gene in the photoarsenotroph *Ectothiorhodospira* sp. BSL-9, isolated from Big Soda Lake (NV, USA) confirmed the ArxAB as the central arsenite oxidase in the photoarsenotroph bacteria (12).

Photoarsenotrophy was also observed in *Ectothiorhodospira* strain MLW-1, isolated from the north shore of Mono Lake (CA, USA), with genetic studies suggesting to be an *arx* mediated As(III) oxidation after PCR detection of a partial *arxA* sequence, and absence of *aio* gene (13). Three more related species *Ectothiorhodospira shaposhnikovii* DSM 2111, *Ectothiorhodospira shaposhnikovii* DSM 243T, and *Halorhodospira halophila* DSM 244, also harboring an *arx*-like gene, were able to oxidize the arsenite, but didn't grow with As(III) as the sole electron donor, needing the addition of sulfide (13).

Recently, a novel strain *Desulfotomaculum* sp. TC-1 was isolated from a sulfidic hot spring in China, capable of oxidase the arsenite coupled to sulfate reduction, having the thioarsenates, as the main reduced product instead of arsenate (14). *Desulfotomaculum* sp. TC-1 represented the first isolated strain of the phylum *Firmicutes* to holds an *arx* gene, and therefore the *arx* arsenite oxidation is likely to be more phylogenetically diverse.

In this chapter it was reported the enrichment procedure and isolation of a bacterium that could oxidize arsenite under micro-aerobic conditions or in presence of nitrate. This strain holds an *arx*-like gene cluster but lacks of *aio* genes after inspection in the genome.

4.2 Materials and methods

4.2.1 Study area

Jozankei Hot spring, Hokkaido, Japan, was chosen to pursue the isolation of an Arx containing arsenite-oxidizing bacterium, because of previously reported arsenic presence in the Toyohira River and Jozankei hot spring area (15, 16); and detection of partial *arxA* genes reported comprehensively in chapter 2.

4.2.2 Sampling description

A dark green microbial mat was recovered from an inclined concrete wall with overflowing hot spring water located in Jozankei, Hokkaido Japan on May 13, 2016. The details of sampling procedure and physicochemical characterization of the site were described more in detail by Kojima et al. (17) and can be found in chapter 2 as well.

4.2.3 Culture media

A synthetic basal medium described previously (18) was used for enrichment and arsenite oxidation experiments. This basal medium was made by combining three separate solutions (A, B and C) with the following composition: Solution A (pH 7.0) contained (per L) 80.1 mg H_3BO_3 (Sigma-Aldrich), 341.0 mg NaCl, 100.0 mg NH_4Cl, 54.7 mg KCl, 6.7 mg KH_2PO_4, 3.0 mg $MgCl_2 \bullet 6\ H_2O$ (Wako Pure Chemical, Japan), 2.6 g HEPES to control pH (Dojindo, Japan), 10 mL Wolfe's vitamin solution, and 10 mL Wolfe's modified mineral (19) solution. Solution B contained 55 g/L $NaHCO_3$ (Wako), and solution C contained 5.3 g/L $CaCl_2 \bullet 2\ H_2O$ (Wako). An aliquot of 10 mL of each solution (B and C) were added to 1 L of solution A.

4.2.4 Enrichment procedure and isolation of strain

A volume of 20 mL of solution A was aliquoted in 70 mL serum bottle vial, and sealed with rubber stoppers. Then it was bubbled with N_2 gas for 15 min to create anoxic conditions, and sterilized by autoclave. After that, the remaining solutions B and C were added aseptically to each bottle. The basal medium was further supplemented with 0.5 mM or 1 mM of arsenite

in the form of $NaAsO_2$. Two different electron acceptor were tested separately during the enrichment step: 5 mM nitrate as $NaNO_3$, and 3 mM chlorate as $NaClO_3$.

A piece of the green microbial mat was aseptically added to the 20 mL enrichment liquid media. Then, the headspace was purged with N_2 gas for 5 min to maintain the anaerobic conditions. The cultures were incubated at 32°C. After growth was visually confirmed and arsenate was detected in the enrichment culture, an aliquot of 800 µL was transferred to a fresh media. During the second transfer, the arsenite/chlorate grown culture was used to inoculate two different arsenite/chlorate medium: one named CEYE, to which was added 1 mg/L of yeast extract (YE); and other named CE, with no addition of YE. During second enrichment, the CEYE culture was further supplemented with thiosulfate (2 mM), to improve bacterial growth. Schematic representation of the enrichment procedure is described in the figure 4-1.

Arsenite oxidation was monitored through arsenate production, measured every 2 to 3 weeks using ion chromatography with conductivity detection (DIONEX ICS-1500). The procedure to measure arsenate was carried out as follow: approximately 500 µL was aseptically removed from the culture, preserving the anoxic conditions in the media. The cells were removed by passing the sample through nylon syringe filter of pore size 0.22 µm and diameter 4 mm (Starlab Scientific, China). The sample was stored in a freezer of -30°C, until the measurement. The sample was subject to a dilution factor of 10 with Milli-Q water before injection by the AS50 Autosampler. The eluent consisted in 1X of a Na_2CO_3-$NaHCO_3$ buffer solution. The flow rate was 1.5 mL/min. The retention time for arsenate was around 9 min, and the total time for each sample run was 15 min. A calibration curve was established with a minimum of three points from a prepared arsenate standard, and the lower limit of detection was 0.01 mM.

Screening of a partial *arxA* gene was made using previous developed degenerate primers sets, and under the same PCR conditions according to the method described in chapter 2. The microbial community of the enrichment cultures was monitored and compared after every transfer by denaturing gradient gel electrophoresis (DGGE) analysis of partial 16S rRNA gene, amplified using the primer set GC341F/907R.

The DGGE method is described as follow: fifty milliliters of PCR master mix was poured into a tube containing the pellet of around 500 µL of enrichment culture after it was centrifuged at 14.000 for 5 min. PCR master mix were prepared by combining: 0.5 µmol/L of each primer, 1x Ex Taq Buffer (Takara, Shiga, Japan), 0.2 mmol/L dNTPs (Takara), 0.625 U of Ex Taq (Takara). PCR condition consisted in an initial denaturation step at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 30 sec, annealing at 45°C for 30 sec, and elongation at 72°C for 45 sec, and a final extension step at 72°C for 10min. DGGE was performed using the Bio-Rad D-code system (CA, USA). Major DGGE bands were excised from the gels and re-amplified by PCR using the same primers as described above but applying 19 cycles, and half the amount of Ex Taq. Then, the PCR product was purified and sequenced by ABI Prism 3130 Genetic Analyzer using BigDye^{TM} Terminator Cycle Sequencing 3.1 (Applied Biosystems, Foster City, CA) with both primer. The obtained partial 16S rRNA gene sequences were subject to a BLASTn search (http://www.ncbi.nlm.nih.gov/BLAST/), to determinate the phylogenetic affiliation by the closest cultured relative nucleotide similarity.

Following the fifth transfer, a pure culture was isolated from the enrichment culture positive for the partial *arxA* gene, and arsenate production. The pure strain was obtained by repeated agar shake dilution at 32˚C. The method and medium were identical to those used for isolation of *Sulfuritortus calidifontis* J1A^{T}, and detailed description of the procedure can be

found in the paper by Kojima et al. (20). Strain M52 was also grown with thiosulfate and nitrate in the basal media described before.

4.2.5 Genome sequencing of M52 strain

Genomic DNA was extracted using Wizard® Genomic DNA purification kit (Promega, Madison, WI, USA). The resulting genomic DNA was sequenced using Illumina HiSeq 2500, generating a 100 bp paired-end library with a total of 147,290,232 reads. The reads were assembled using Edena (v3) software (21). From these it was obtained 19 contigs.

The genome assembly was used for the *arx* like gene cluster analysis in chapter 3. The genome contigs were annotated using The SEED and RAST: Rapid Annotation using Subsystem Technology (22). The annotation overview of the genome was made in the SEED Viewer version 2.0 (online). To confirm the presence or absence of *aioA* gene, the predicted proteins from strain M52 genome was compare with 292 complete AioA protein sequences obtained from Genbank database. Diamond v0.9.14.115 was used to find the homologous gene if it had an amino acid identity of > 60%, a spanning alignment of > 60% of the length of the query, and an e-value < 1e-3. Those proteins in strain M52 meeting the criteria were used as query sequence in a BLASTp search against the non-redundant protein sequences (nr) database, excluding uncultured/environmental sample sequences, to confirm the identification of the protein.

4.2.6 Phylogenetic analysis

The 16 rRNA gene was first amplified using the universal primer 27F and 1492R (23). The resulting PCR product was purified with QIAquick PCR Purification Kit (QIAGEN,

Germany), and used as a template for sequencing. The following seven 16S rDNA sequencing primers were used in the sequencing reactions: 27F, 533F, 533R, 926F, 907R, 953F and 1492R.

The RDP classifier (https://rdp.cme.msu.edu/classifier/classifier.jsp) and the SINA aligner and classification system (https://www.arb-silva.de/aligner/) was used to assign a taxonomic identity to the nearly full-length 16S rRNA sequence. Ten related sequences (nearest neighbors) to the 16S rRNA gene of strain M52 were downloaded from SILVAweb interface (24), with a minimum identity of 95%. Additionally, the best ten hits based on BLASTn results, excluding uncultured/environmental sample sequences, were retrieved as well. All the sequences obtained were used to reconstruct a phylogenetic tree. Sequence alignment was performed with MAFFT version 7. Gaps and poorly aligned regions were excluded using Gblocks v0.91b, resulting in 1416 nucleotides positions in the final dataset. The tree topology was inferred by distance criterion using the neighbor-joining on MEGA 6; and maximum-likelihood algorithm with the GTR substitution model in PhyML online program (v3.0). The accuracy of the branches was determined from 1000 replicates, and the aLRT test (SH-Like), within each program, respectively.

4.2.7 Verification of arsenite transformation by the isolated strain

All batch cultures were made in 70 ml serum bottles filled with 20 mL of synthetic basal medium supplemented with 1 mg/L of yeast extract as a carbon source, 0.5 mM of molybdate (Na_2MoO_4), and amended with 1 mM of arsenite in the form of $NaAsO_2$. Control without inoculum was made to examine any abiotic mediated arsenite oxidation under anoxic conditions.

Preliminary experiments demonstrated that chlorate was not being reduced. Inspection of the genome revealed the absence of *clrABDC* genes coding for the chlorate reductase, and *cld*

genes, coding for chlorite dismutases, necessary for the metabolism of chlorate to molecular oxygen. Therefore, chlorate was not used in the assays as an electron acceptor.

The arsenite oxidation experiments were also done to evaluate different electron acceptors, according to the findings made in the genome and the results of each analysis. Thus, the experiments were executed in different time periods of the study. Each of the experiments is mentioned below in the order of implementation:

1) Anaerobic condition experiment: to verify the arsenite oxidation without the addition of electron acceptor. Two concentration of NaCl (1% and 0.2%) were also tested, since chloride was found to be around 40 mM (0.25%) in Jozankei Hot spring water. The experiment was conducted in triplicate.

2) Aerobic culture experiment: to examine the oxygen as electron acceptor. It was employed the same supplemented liquid media with 0.2% NaCl and 2 mg/L YE. The experiment was conducted in triplicate.

3) Anaerobic condition, with nitrate as electron acceptor experiment: In this experiment ~10 mM of nitrate as $NaNO_3$ was added. The experiment was made due to the detection of periplasmic nitrate reductase (*nap* genes) in the strain M52 genome. The experiment was conducted in duplicate.

4) Microaerophilic condition experiment: to confirm the use of oxygen as electron acceptor under micro-oxic conditions. The micro-aerobic conditions were prepared as follows: the anoxic basal media was prepared under anaerobic conditions (nitrogen in the headspace) with all the supplementation as it was explained before. To the prepared media, it was incorporated oxygen by passing an air volume through a polyethersulfone (PES) sterile syringe filter of 0.22 μM (Starlab Scientific, China). The amount of air added to the media was based

on the total volume of the headspace in the bottle (50 mL = 100%). Thus, 5% corresponded to 2.5 mL, and 10% to 5 mL of air included to the bottle. It was also evaluated no addition of air along with the other experiments, as a control (0% of air - anaerobic). The experiment was conducted in duplicate.

Arsenite oxidation was monitored through arsenate production, every 5 to 7 days, using ion chromatography with conductivity detection (ICS-1500), as it was explained before Also, arsenite and arsenate were measured by HPLC procedure as it was described previously by Watanabe et al. (25).

For the inoculation, it was transferred 800 µL from previously grown culture on arsenite to each experiment.

4.3 Results

4.3.1 Enrichment cultures and isolation of strain.

Enrichment cultures were conducted to isolate a possible *arx* arsenite-oxidizing bacterium after the presence of partial *arxA*-like genes were detected in Jozankei sample. Nitrate NO^{3-} and chlorate ClO^{3-} were tested separately as potential electron acceptors taking into account past reports involving nitrate reduction by the *arx*-containing strain A. ehrlichii MLHE-1 (7, 8); and chlorate associated to *aio*-dependent arsenite oxidation pathway (26).

The first and second enrichment cultures were subject to an extended period of incubation since As(III) generate low growth, evident by weak turbidity. Subsequent additions of ~0.5 mM arsenite, electron acceptor and $NaHCO_3$ were made starting from the second enrichment culture to help increase the cell density, as it was reported to work for *A. ehrlichii* MLHE-1 (7). The

additional supplementation also helped to exert more selective pressure on arsenite-oxidizing bacteria as well (Fig. 4-1).

Arsenate was detected in all the cultures after the prolonged incubation, but only higher concentration was observed in the CEYE and CE cultures, with the final concentration of arsenate been about the same as the sum of the arsenite added (Table 4-1). DGGE revealed the presence of only one major band in the arsenite/nitrate enrichment culture related to *Shewanella*, in contrast, several significant bands were part of the arsenite/chlorate communities. Notably, CEYE and CE cultures shared the same microbial composition: *Ignavibacterium album*, and *Georgfuchsia toluolica* strain G5G6 related sequences (86-99% and 96% nucleotide identity, respectively) (Fig. 4-2, Table 4-2). It was decided to continue working exclusively with the CEYE and CE cultures during the study, based on the high arsenate detection (Table 4-1).

After that, CEYE was further supplemented with thiosulfate (2 mM), considering recent reports revealed three anaerobic sulfur and arsenite-oxidizing bacteria showed growth only in the presence of sulfide (13). Additionally, it was found similarity between the partial 16S rRNA gene of one of the significant DGGE band in the CEYE and CE community to a sulfur oxidizer strain isolated by Professor Kojima, strain J5B, prompting the question about the improvement of growth by the inclusion of this additional electron donor.

Measurement of arsenate in CEYE enrichment culture after addition of thiosulfate seemed to have delayed the conversion to arsenate (table 4-1), yet it appears to have increased the bacterial abundance, evidenced after observation under the microscope (Fig 4-3). Regardless of this, the CEYE + $S_2O_3^-$ and CE microbial community remained constant, maintaining as significant bands those related to *Ignavibacterium album* and *Georgfuchsia toluolica* strain

G5G6 (Fig. 4-2). Moreover, it was detected an identical partial *arxA* sequence in both of these enrichment cultures (Fig. 4-4).

After the fifth enrichments, CE enrichment culture E4 was used to isolate a strain named M52. This culture was selected for the active arsenite conversion and the detection of a partial *arxA*-like sequence. Strain M52 was very similar to a strain J5B, a sulfur-oxidizing bacteria (SOB) isolated by Professor Kojima. Therefore, it was used the same procedure of isolation as it was implemented for *Sulfuritortus calidifontis* $J1A^T$, also a SOB recovered from Jozankei, described in detail by Kojima et al. (17). The optimum temperature of growth was 45°C, hereupon used in all the subsequent experiments. M52 strain also showed chemoautotrophic growth with thiosulfate as an electron donor and $NO3^-$ as the electron acceptor. The growth of strain M52 on thiosulfate was more abundant than that with arsenite, evidenced by high turbidity.

4.3.2 Phylogenetic analysis

Strain M52 is a member of the class *Betaproteobacteria*, family *Rhodocyclaceae,* based on the results of the SINA Aligner and RDP classifier analysis of the almost complete 16S rRNA sequence. BLAST results indicated the closest cultured relatives were *Georgfuchsia toluolica* strain $G5G6^T$ and *Denitratisoma oestradiolicum* strain AcBE2-1 ^{T}, with 94% of sequence identity, both belonging to the family *Sterolibacteriaceae*. Cultured and uncultured closest relatives were used for the reconstruction of a phylogenetic tree, including the 16S rRNA gene of strain J5B, provided by Professor Kojima. In the basis of the reconstructed evolutionary tree, J5B and M52 strain are placed together within the family *Sterolibacteriaceae* (Fig. 4-5). As

a side note, the genera of the *Sterolibacteriaceae* were formerly assigned to the order *Rhodocyclales* but recently transferred to the *Nitrosomonadales* (27).

It is worth to notice that strain M52 and J5B may represent a novel genus, due to the formation of a strongly supported cluster distantly related to the representative species of the other genera within the family *Sterolibacteriaceae*. However, in-depth taxonomical classification of the strain J5B and M52 is outside of the scope of the present work.

4.3.3 M52 strain genomic exploration in relation to arsenic metabolism

A complete *arx*-like operon *arxRSXABCDB'* coding for an arsenite oxidase was detected in the genome of strain M52. The identity and similarity of each gene can be found in table 4-3. The closest cultured relatives to all the *arx* putative proteins correspond to organism affiliated to class *Betaproteobacteria*, including *Sulfuricella denitrificans* skB26^{T} and *Rhodocyclaceae* bacterium. The *arxB'* gene of strain M52 is positioned downstream to the *arxD* gene and in the same orientation, contrary to the majority of the *arx* operon gene arrangement which is located upstream the *arxABCD* gene cluster. More details about the *arx*-like operon of strain M52 can be found in chapter 3.

It is worth to notice the presence of a gene coding for a mobile element protein placed upstream the *arxRSX* that has 92% identity to the IS110 family transposase of *Thiomonas* sp. FB-Cd (BLASTp). Additionally, the same coting where the *arx* operon is localized, include three more genes coding for different transposases, and one integrase. These mobile genetic elements can be involved in horizontal gene transfer (HGT), and may be responsible for DNA exchanges between some species.

The genes involved in the detoxification-based on arsenate reduction mechanism were also identified. Although not complete, the minimum genes necessary for the regulation and function were present, the couple *arsRC* genes, with the following function: *arsR*, controlling the expression of the arsenic-resistant operon, and *arsC*, coding for the arsenate reductase that converts arsenate to arsenite. It was also recognized the *acr3* gene, coding for one of the two types of the arsenite-specific efflux pump, which expel arsenite outside the cell.

No other arsenic metabolism-related genes were detected. For instance, the *aioAB* gene coding for the AIO arsenite oxidase, the *arrAB* gene operon coding for the respiratory arsenate reductase, and the *arsM* coding for the methylation of arsenic species, all of which were absent in the strain M52 genome.

On the other hand, inspection of the respiratory pathways revealed the presence of only two terminal oxidases. The *ccoNOQP* and *ccoGHIS* gene cluster coding for the *cbb3* type cytochrome *c* oxidase; and the *cydAB* genes coding for the cytochrome *bd*. Both cytochrome oxidases are usually reported been expressed under low tension of oxygen, thus indicative of microaerophilic respiration (28, 29). Cytochrome *bc1* complex, which serves as a coupling site for many of these respiration pathways, was also detected in the genome, including several cytochrome *c* centers. The genes necessaries for the ubiquinone, as well as menaquinone, which work as a substrate for the terminal acceptors, were also present; however, some genes coding the biosynthesis pathways enzymes were missing.

Under anaerobic conditions, genes involved in respiration via nitrification/denitrification mechanism were detected but incomplete. Strain M52 lack of all the genes necessary for nitrate reduction respiration via NarGHJ. But, contain all of the genes of the periplasmic nitrate reductase (NapABC enzyme), which also catalyzes the transformation of nitrate to nitrite. There

was an absence of some genes essential in the nitrite reductase *nir* operon, although, the genes coding for nitric oxide and nitrous oxide reductase, involved in the final steps of denitrification, were complete.

The molybdopterin cofactor and the iron-sulfur cluster biosynthesis genes were found complete. These genes are required for the synthesis and insertion of the molybdopterin cofactors and in general the assembly of the molybdoenzymes. RuBisCO and Calvin-Benson genes were present suggestive of autotrophic CO_2 fixation.

4.3.2 Verification of arsenite transformation by strain M52.

Strain M52 was tested for its ability to transform arsenite to arsenate. In the initials experiment, conversion of As(III) to arsenate were detected without the reduction of chlorate. Inspection of the genome revealed the absence of the genes involved in the chlorate metabolism, for which it was decided to exclude the chlorate in the future experiments.

First, an experiment was made with 1 mM of arsenite and 1% of NaCl, without the addition of an electron acceptor, and performed in triplicate. In this experiment, it was detected arsenate only in one of the three replicates (in replicate No. 3). As a result, after 12 days of incubation, the replicate No. 3 had around 1.007 mM of arsenate; that increased to 1.546 mM after 19 days (Fig. 4-6, Table 4-4). Around ~0.5mM of arsenite was added to the replicate No. 3 on day 24, and after another 18 days of incubation (day 42), arsenate values went up to 1.715 mM. The remaining cultures replicate No. 1 and No. 2, on the contrary, reduced the initial concentration of arsenate product of the carryover from the inoculum, until no detection on day 42 (Fig. 4-6, Table 4-4).

After visual inspection of the cultures, it was notable a difference in the rubber stopper used to plug up the serum bottle (Fig. 4-7). In the case of replicate No. 3 was used a much thinner type of stopper, which could have allowed low oxygen diffusion to the culture, generating microaerophilic conditions. A triplicate experiment was prepared using cultures with 1 mM arsenite, 0.2% NaCl, devoid of an electron acceptor, and sealed with the same rubber stopper as the replicate No. 3 to confirm that the type of stopper used influenced the outcome of the previous experiment. After six days of incubation, an average of 1.07 mM of arsenate was detected in all three replicates (Table 4-4, Fig.4-6). After this, ~0.5mM of arsenite was added to all three cultures, and the next six days the average arsenate concentration increased to 1.458 mM (Fig. 4-6, Table 4-4). As a result, strain M52 seemed to be able to oxidize the arsenite, probably by using oxygen as the electron acceptor that might have been diffused to the media.

To confirm if strain M52 was using oxygen as the electron acceptor, a triplicate batch experiment was made under aerobic conditions, with 1mM arsenite, 0.2% NaCl and 2 mg/L of YE. After 6 days of incubation, no significate concentration of arsenate was detected; instead the arsenate (~0.09 mM) product of the carryover from the inoculum was being reduced (~0.06 mM), and after 12 days it decreased further to ~0.02 mM (Fig 4-6, table 4-4). Strain M52 have a detoxification-based arsenate reductase *arsC*, which might have been the responsible for the reduction of the arsenate carryover. The absence of arsenate production did not explain the arsenite oxidation results of the first experiments.

Inspection of the genome led to detect the presence of *nap* nitrate reductase, and two terminal oxidases specific for micro-aerobic conditions, cytochromes *cbb3* and *bd* oxidases. Duplicates experiments were made to test the ability of the strain M52 to use nitrate (~ 8.5 mM),

or low volume of oxygen (referring as the volume of air in the total volume of the headspace, in percentage) as an electron acceptor. The concentration of arsenate was measured after 6 days of incubation.

Strain M52 seemed to oxidase the arsenite in the presence of nitrate under anaerobic condition, and after 6 days was detected around ~ 0.79 mM of arsenate (Fig 4.8-E). High concentration of arsenate was also found when strain M52 grew in micro-aerobic condition with 5% and 10% of the air in the headspace, producing around ~0.96 mM and ~0.95 mM, respectively (Fig 4.8-C and Fig4.8-D, table 4-4, table 4-5). In contrast, in the aerobic and anaerobic cultures, no arsenite oxidation was detected after 6 days of incubation (Fig4.8-A and B, table 4-4, table 4-5), but only the arsenate carryover (table 4-4).

4.4 Discussion

4.4.1 Micro-aerobic and nitrate arsenite oxidation by M52 strain.

Strain M52 was isolated from Jozankei hot spring microbial mat. It was one of the major band present in the arsenite/chlorate enrichment cultures, closely related to *Georgfuchsia toluolica* strain G5G6, also member of the family *Sterolibacteriaceae* (Fig. 4-2 and Fig 4-3). The 16S rRNA phylogenetic tree revealed that strain M52 was closely related to sequences from uncultured bacteria recovered from 1.7 Km soil depth in Northam platinum mine (South Africa) (Fig. 4-5). Physicochemical characteristics from the sampling site in the platinum mine showed high temperature (48°C), pH of 8.3, and was nearly hypoxic (24 μM O_2) (30). Similar conditions were found in the Jozankei microbial mat hot spring water, with a temperature of 42.6°C and slightly similar pH (7.8).

Verification of the arsenite oxidation suggested with strong evidence that strain M52 was the responsible for the arsenate production, rather than any other physicochemical condition; since no detection of As(V) was obtained in the abiotic control during all measured time points (fig. 4-5).

Lack of *aio* genetic signature, and a confirmation of a complete *arx* gene cluster in the genome of strain M52 are suggestive of an *arx* -arsenite oxidation. One interestingly results was perhaps the potential use of oxygen as electron acceptor, considering that currently *arx* has only been related to anaerobic respiration using either nitrate or sulfate (8, 14) and during anaerobic photosynthesis (31).

Arsenite-oxidizing bacteria were first described in 1918 (32), and the majority are heterotrophic mesophilic aerobes *aio*-harboring organism (3, 33). No aerobic *arx* dependent pathway had been reported yet. Strain M52, however, was able to oxidase the arsenite under 5% and 10% of air (low volume of oxygen), but not under full aerobic condition (Fig 4-6, Table 4-4). Inspection of the genome revealed only the presence of cytochrome *cbb*3 and *bd* oxidases, which expression are strongly induced in cells growing at a low tension of oxygen.

Strain M52 was isolated from a microbial mat, which are self-sustain and vertically stratified communities, developed in many cases in extreme environments. Microbial mat has a dynamic physicochemical gradient, and specifically, the oxygen profile shows decreasing concentration with depth (34–36). Additionally, biological process and high temperature might also influence the dissolved O_2, and thereby affect the microbial nutrition, and bacterial composition along the vertical column (37).

High-affinity oxidases genes may enable strain M52 to colonize low levels of O_2 niches in the microbial mat, and use it as electron acceptor with the arsenite as a substrate, which is

the abundant form of arsenic in low redox condition. Additionally, the presence of other respiratory genes in strain M52, like the periplasmic nitrate reductase *nap* genes, could also confer a survival advantage in anoxic zones, by probably using nitrate as an alternative final oxidizing agent in the electron flow pathway.

Arsenite oxidation coupled to nitrate or oxygen reduction is energetically favorable since the reduction potential (E_0') of the arsenate/arsenite couple (+ 60 mV) is lower than the more positive pair NO_3^-/NO_2^- (+ 430 mV) (7) or $½O_2/H_2O$ (+ 820 mV) (38). It thus, might have free energy yields with oxygen of around $\Delta G^{0'}$ = -169.84 kJ; and with nitrate of approximately $\Delta G^{0'}$ = -87.16 kJ (calculated with the Nernst equation: $\Delta G^0=-nF\Delta E_0$', where n is the number of electrons transferred, F is the Faraday constant = 96.5 kJ/V, and the ΔE_0' were calculated from the E_0' values mentioned above (39)). However, no difference in turbidity was visible when growing with arsenite in nitrate or oxygen supplemented cultures.

On the contrary, higher bacterial density evidenced as high turbidity was observed when strain M52 growth with thiosulfate and nitrate, as the electron donor and the electron acceptor, respectively. The substantially lower redox potential of thiosulfate ($S_2O_3^-/HS^- + SO_3^{2-} = -400$ mV) (38), make it a better source of electron donor as compared to the inorganic arsenic pair; thus a higher amount of energy may be released in the net reaction, that can be used to grow by strain M52.

To this date no experimental redox data is available for the electron flow with the Arx enzyme; nonetheless, evidence suggests the electrons resulting from the oxidation of arsenite would reduce liposoluble quinones, specifically ubiquinone (UQ), as it was reported for *Halorhodospira halophila* (40). Furthermore, *Alkalilimnicola ehrlichii* MLHE-1, the first reported *arx* arsenite-oxidizing bacterium, uses as a single quinone the UQ, identified in the

genome and by biochemical analysis (3). In strain M52 was detected both ubiquinone and menaquinone biosynthesis pathways.

As for the terminal oxidases, strain M52 contain the high-affinity O_2 cytochromes, *cbb*3 and *bd* complex, both reducing O_2 to H_2O as the final step of the electron transport chain. If strain M52 happen to use the cytochrome *cbb*3, the electron flows from the quinone to the cytochrome *bc*1 complex, and then to a *c* type cytochrome, which serves as the primary electron donor for this terminal oxidase (28). In contrast, cytochromes *bd* complex, also named cytochrome *d* ubiquinol oxidase, accept electrons directly from reduced ubiquinone (ubiquinol) (41)

In line with the limited information available for Arx and the genome inspection of strain M52, the metabolic pathway in a putative micro-aerobic chain involving the high affinity oxidases might feature the UQ as the electron transfer molecule.

Overall, the flexibility of strain M52 to use arsenite the under micro-aerobic condition and maybe coupled to nitrate respiration could grant it adaptation to small environmental changes in oxygen concentration, and redox potential in deepest parts of the microbial mat. Besides, the arsenate release can also be used as a potential electron acceptor by other organisms present in the community.

4.4.2 Biotechnological implication of arsenite oxidation under micro-aerobic conditions

The capability of strain M52 to oxidize the arsenite under anaerobic and micro-oxic condition could be explored for bioremediation of contaminated places of low redox potential like groundwater.

Many countries around the world suffer from arsenic contamination of groundwater, and the most several cases are known to be in Bangladesh and West Bengal (42, 43), but also in Taiwan (44), China (45) and Pakistan (46), with an arsenic concentration above > 50 μg/L.

The deposits of arsenic-bearing sediment are the most critical source of contamination in the groundwater in the majority of these countries (47). Abiotic and biotic conditions led to the release of the inorganic arsenic to the aquifers, either through mobilization by anaerobic metal-reducing bacteria (48), or changes in the redox potential, with anoxic conditions dominated by elevated dissolved As(III) concentration (47).

Many physicochemical removal strategies required the oxidation of arsenite to arsenate as the most common step, since the latter can be adsorbed more freely onto solid surfaces than As(III), and thus, can be immobilized for a satisfactory As removal (49–51). Many chemicals used for arsenic oxidation include iron, chlorine, hydrogen peroxide and permanganate, some of which leave unwanted byproducts in the water that will ultimately be used for drinking or irrigation (49, 52). Therefore, the use of biological oxidation is deemed as promising, for its low cost, eco-friendly, and generation of low natural disturbance (53).

Some organism used in the biological oxidation of arsenite strategies include iron-oxidizing (52), or sulfate reducing microorganism (54), that often induce a change of pH and biogeochemical transformation of other surrounding metals. In this sense, the oxidation of arsenite under low levels of oxygen could be promising during the bioremediation of groundwater, with seemingly less perturbation of the surround physicochemical conditions, although, not other parameters were measured in the experiment with strain M52.

Strain M52 is not the only microaerophilic arsenite-oxidizing bacteria, in the literature *Hydrogenobaculum* strain H55 isolated from an acid-sulfate-chloride geothermal spring in

Yellowstone National Park, is also able to oxidize the arsenite under micro-aerobic levels of oxygen (~1% O_2) with CO_2 and H_2 as the primary carbon and energy source, respectively (55). However, no gene responsible for the activity was reported in the study. Therefore, the present work may contribute to raising the questions about the different alternatives metabolic pathways that may be used by the Arx-arsenite-oxidizing bacteria and its prospective application.

4.5 Reference

1. **Santini JM, Sly LI, Wen A, Comrie D, Wulf-Durand P De and Macy JM**. 2002 New Arsenite-Oxidizing Bacteria Isolated from Australian Gold Mining Environments--Phylogenetic Relationships. Geomicrobiol. J. **19**:67–76.
2. **Ilyaletdinov AN and Abdrashitova SA**. 1981 Autotrophic arsenic oxidation by a *Pseudomonas arsenitoxidans* culture. Mikrobiologiia **50**:197–204.
3. **Van Lis R, Nitschke W, Duval S and Schoepp-Cothenet B**. 2013 Arsenics as bioenergetic substrates. Biochim. Biophys. Acta - Bioenerg. **1827**:176–188.
4. **Rhine ED, Ní Chadhain SM, Zylstra GJ and Young LY**. 2007 The arsenite oxidase genes (*aroAB*) in novel chemoautotrophic arsenite oxidizers. Biochem. Biophys. Res. Commun. **354**:662–667.
5. **Inskeep WP, Macur RE, Hamamura N, Warelow TP, Ward SA and Santini JM**. 2007 Detection, diversity and expression of aerobic bacterial arsenite oxidase genes. Environ. Microbiol. **9**:934–943.
6. **Zhu Y-G, Yoshinaga M, Zhao F-J and Rosen BP**. 2014 Earth Abides Arsenic Biotransformations. Annu. Rev. Earth Planet. Sci. **42**:443–467.
7. **Oremland RS, Hoeft SE, Santini JM, Bano N, Hollibaugh R a and Hollibaugh JT**. 2002 Anaerobic Oxidation of Arsenite in Mono Lake Water and by a Facultative, Arsenite-Oxidizing Chemoautotroph, Strain MLHE-1. Society **68**:4795–4802.
8. **Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM and Oremland RS**. 2007 *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic

gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int. J. Syst. Evol. Microbiol. **57**:504–512.

9. **Zargar K, Hoeft S, Oremland R and Saltikov CW**. 2010 Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. J. Bacteriol. **192**:3755–3762.

10. **Zargar K, Conrad A, Bernick DL, Lowe TM, Stolc V, Hoeft S, Oremland RS, Stolz J and Saltikov CW**. 2012 ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. Environ. Microbiol. **14**:1635–1645.

11. **Kulp TR, Hoeft SE, Asao M, Madigan MT, Hollibaugh JT, Fisher JC, Stolz JF, Culbertson CW, Miller LG and Oremland RS**. 2008 Arsenic(III) Fuels Anoxygenic Photosynthesis in Hot Spring Biofilms from Mono Lake, California. Science (80-.). **321**:967–970.

12. **Hernandez-Maldonado J, Sanchez-Sedillo B, Stoneburner B, Boren A, Miller L, Mccann S, Rosen M, Oremland RS and Saltikov CW**. 2017 The genetic basis of anoxygenic photosynthetic arsenite oxidation. Environ. Microbiol. **19**:130–141.

13. **Hoeft McCann S, Boren A, Hernandez-Maldonado J, Stoneburner B, Saltikov C, Stolz J and Oremland R**. 2017 Arsenite as an Electron Donor for Anoxygenic Photosynthesis: Description of Three Strains of Ectothiorhodospira from Mono Lake, California and Big Soda Lake, Nevada. Life **7**:1–14.

14. **Wu G, Huang L, Jiang H, Peng Y, Guo W, Chen Z, She W, Guo Q and Dong H**. 2017 Thioarsenate formation coupled with anaerobic arsenite oxidation by a sulfate-reducing bacterium isolated from a hot spring. Front. Microbiol. **8**:1–9.

15. **Tatsumi K, Jin K and Tachibana H**. 2006 Dynamic state of Arsenic and Basin management in Toyohira River. J. Japan Soc. Water Environ. **29**:671–677.

16. **Kubota Y, Sakakibara M and Sano S**. 2010 Arsenic cocentration of hot spring and river waters at Jyozankei hot spring area in Sapporo city, western Hokkaido, Japan. Mem. Fac. Sci. Ehime Univ. **16**:7–13.

17. **Kojima H, Watanabe M and Fukui M**. 2017 *Sulfuritortus calidifontis* gen. nov., sp. nov., a sulfur oxidizer isolated from a hot spring microbial mat. Int. J. Syst. Evol. Microbiol. **67**:1355–1358.

18. **Connon SA, Koski AK, Neal AL, Wood SA and Magnuson TS**. 2008 Ecophysiology and geochemistry of microbial arsenic oxidation within a high arsenic, circumneutral hot spring system of the Alvord Desert. FEMS Microbiol. Ecol. **64**:117–128.

19. **Wolin EA, Wolin MJ and Wolfe RS**. 1963 Formation of methane by bacterial extracts. J. Biol. Chem. **238**:2882–2886.

20. **Kojima H and Fukui M**. 2011 *Sulfuritalea hydrogenivorans* gen. nov., sp. nov., a facultative autotroph isolated from a freshwater lake. Int. J. Syst. Evol. Microbiol. **61**:1651–1655.

21. **Hernandez D, François P, Farinelli L, Dohm JC, Lottaz C, Borodina T, Butler J, Maccallum I, Kleber M, Zerbino DR, Birney E, Østerås M and Schrenzel J**. 2008 De novo bacterial genome sequencing : Millions of very short reads assembled on a desktop computer De novo bacterial genome sequencing : Millions of very short reads assembled on a desktop computer. 2008: 10.1101/gr.072033.107.

22. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F and Stevens R**. 2014 The SEED

and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. **42**:206–214.

23. **Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML and Pace NR**. 1985 Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. **82**:6955–9.

24. **Pruesse E, Peplies J and Glöckner FO**. 2012 SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics **28**:1823–1829.

25. **Watanabe T, Miura A, Iwata T, Kojima H and Fukui M**. 2017 Dominance of *Sulfuritalea* species in nitrate-depleted water of a stratified freshwater lake and arsenate respiration ability within the genus. Environ. Microbiol. Rep. **9**:522–527.

26. **Sun W, Sierra-Alvarez R, Milner L and Field JA**. 2010 Anaerobic oxidation of arsenite linked to chlorate reduction. Appl. Environ. Microbiol. **76**:6804–6811.

27. **Boden R, Hutt LP and Rae AW**. 2017 Reclassification of *Thiobacillus aquaesulis* (Wood & Kelly, 1995) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of *Thiobacillus* (Beijerinck, 1904) from the *Hydrogenophilales* to the *Nitrosomonadales*, proposal of *Hydrogenophilalia* class. nov. withi. Int. J. Syst. Evol. Microbiol. **67**:1191–1205.

28. **Pitcher RS and Watmough NJ**. 2004 The bacterial cytochrome *cbb3* oxidases. Biochim. Biophys. Acta - Bioenerg. **1655**:388–399.

29. **Lu P, Heineke MH, Koul A, Andries K, Cook GM, Lill H, Van Spanning R and Bald D**. 2015 The cytochrome *bd*-type quinol oxidase is important for survival of *Mycobacterium smegmatis* under peroxide and antibiotic-induced stress. Sci. Rep. **5**:1–10.

30. **Borgonie G, García-Moyano A, Litthauer D, Bert W, Bester A, Van Heerden E, Möller C, Erasmus M and Onstott TC**. 2011 Nematoda from the terrestrial deep subsurface of South Africa. Nature **474**:79–82.

31. **Hoeft SE, Kulp TR, Han S, Lanoil B and Oremland RS**. 2010 Coupled arsenotrophy in a hot spring photosynthetic biofilm at mono lake, california. Appl. Environ. Microbiol. **76**:4633–4639.

32. **Green HH**. 1918 Description of a bacterium which oxidizes arsenite to arsenate, and of one which reduces arsenate to arsenite, isolated from a cattle-dipping tank. S. Afr. J. Sci. **14**:465–467.

33. **Kumari N and Jagadevan S**. 2016 Genetic identification of arsenate reductase and arsenite oxidase in redox transformations carried out by arsenic metabolising prokaryotes – A comprehensive review. Chemosphere **163**:400–412.

34. **Revsbech NP, Trampe E, Lichtenberg M, Ward DM and Kühl M**. 2016 In situ hydrogen dynamics in a hot spring microbial mat during a diel cycle. Appl. Environ. Microbiol. **82**:4209–4217.

35. **Wong HL, Smith DL, Visscher PT and Burns BP**. 2015 Niche differentiation of bacterial communities at a millimeter scale in Shark Bay microbial mats. Sci. Rep. **5**:15607.

36. **Epping EHG, Khalili A and Thar R**. 1999 Photosynthesis and the dynamics of oxygen consumption in a microbial mat as calculated from transient oxygen microprofiles. Limnol. Oceanogr. **44**:1936–1948.

37. **Prieto-Barajas CM, Valencia-Cantero E and Santoyo G**. 2018 Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. Electron. J. Biotechnol. **31**:48–56.

38. **Grimaldi S, Schoepp-cothenet B, Ceccaldi P, Guigliarelli B and Magalon A**. 2013 The prokaryotic Mo / W- bis PGD enzymes family : A catalytic workhorse in bioenergetic. Biochim. Biophys. Acta **1827**:1048–1085.

39. **Madigan MT, Bender KS, Buckley DH, Sattley WM and Stahl DA**. 2019 Brock biology of microorganisms. 15 Ed., Pearson Education, Upper Saddle River, NJ.

40. **Warelow TP, Oke M, Schoepp-Cothenet B, Dahl JU, Bruselat N, Sivalingam GN, Leimkuhler S, Thalassinos K, Kappler U, Naismith JH and Santini JM**. 2013 The Respiratory Arsenite Oxidase: Structure and the Role of Residues Surrounding the Rieske Cluster. PLoS One **8**:.

41. **Borisov VB, Gennis RB, Hemp J and Verkhovsky MI**. 2011 The cytochrome *bd* respiratory oxygen reductases. Biochim. Biophys. Acta - Bioenerg. **1807**:1398–1413.

42. **Majumder A, Bhattacharyya K, Bhattacharyya S and Kole SC**. 2013 Arsenic-tolerant, arsenite-oxidising bacterial strains in the contaminated soils of West Bengal, India. Sci. Total Environ. **463**–**464**:1006–1014.

43. **Argos M, Kalra T, Rathouz PJ, Chen Y, Pierce B, Parvez F, Islam T, Ahmed A, Rakibuz-Zaman M, Hasan R, Sarwar G, Slavkovich V, Van Geen A, Graziano J and Ahsan H**. 2010 Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): A prospective cohort study. Lancet **376**:252–258.

44. **Liang C-P, Wang S-W, Kao Y-H and Chen J-S**. 2016 Health risk assessment of groundwater arsenic pollution in southern Taiwan. Environ. Geochem. Health **38**:1271–1281.

45. **Pi K, Wang Y, Xie X, Liu Y, Ma T and Su C**. 2016 Multilevel hydrogeochemical monitoring of spatial distribution of arsenic: A case study at Datong Basin, northern China. J. Geochemical Explor. **161**:16–26.

46. **Tabassum RA, Shahid M, Dumat C, Niazi NK, Khalid S, Shah NS, Imran M and Khalid S**. 2018 Health risk assessment of drinking arsenic-containing groundwater in Hasilpur, Pakistan: effect of sampling area, depth, and source. Environ. Sci. Pollut. Res. 2018: 10.1007/s11356-018-1276-z.

47. **Schaefer M V., Guo X, Gan Y, Benner SG, Griffin AM, Gorski CA, Wang Y and Fendorf S**. 2017 Redox controls on arsenic enrichment and release from aquifer sediments in central Yangtze River Basin. Geochim. Cosmochim. Acta **204**:104–119.

48. **Islam FS, Gault AG, Boothman C, Polya DA, Chamok JM, Chatterjee D and Lloyd JR**. 2004 Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature **430**:68–71.

49. **Shankar S, Shanker U, Shikha, Shankar S, Shanker U and Shikha**. 2014 Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation, Arsenic Contamination of Groundwater: A Review of Sources, Prevalence, Health Risks, and Strategies for Mitigation. Sci. World J. **2014**, **2014**:e304524.

50. **Kim MJ and Nriagu J**. 2000 Oxidation of arsenite in groundwater using ozone and oxygen. Sci. Total Environ. **247**:71–79.

51. **Sun W, Sierra-Alvarez R, Fernandez N, Sanz JL, Amils R, Legatzki A, Maier RM and Field JA**. 2009 Molecular characterization and in situ quantification of anoxic arsenite-oxidizing denitrifying enrichment cultures. FEMS Microbiol. Ecol. **68**:72–85.

52. **Zouboulis AI and Katsoyiannis IA**. 2005 Recent advances in the bioremediation of arsenic-contaminated groundwaters. Environ. Int. **31**:213–219.

53. **Pal A and Paknikar KM**. 2011 Bioremediation of Arsenic from Contaminated Water. In Microorganisms in Environmental Management. pp. 477–523.

54. **Maguffin SC and Qusheng J**. 2018 Testing biostimulated sulfate reduction as a strategy of arsenic remediation in iron-rich aquifers. Chem. Geol. **In Press**:

55. **Donahoe-Christiansen J, D'Imperio S, Jackson CR, Inskeep WP and McDermott TR**. 2004 Arsenite-oxidizing *Hydrogenobaculum* strain Isolated from an acid-sulfate-chloride geothermal spring in Yellowstone National Park. Appl. Environ. Microbiol. **70**:1865–1868.

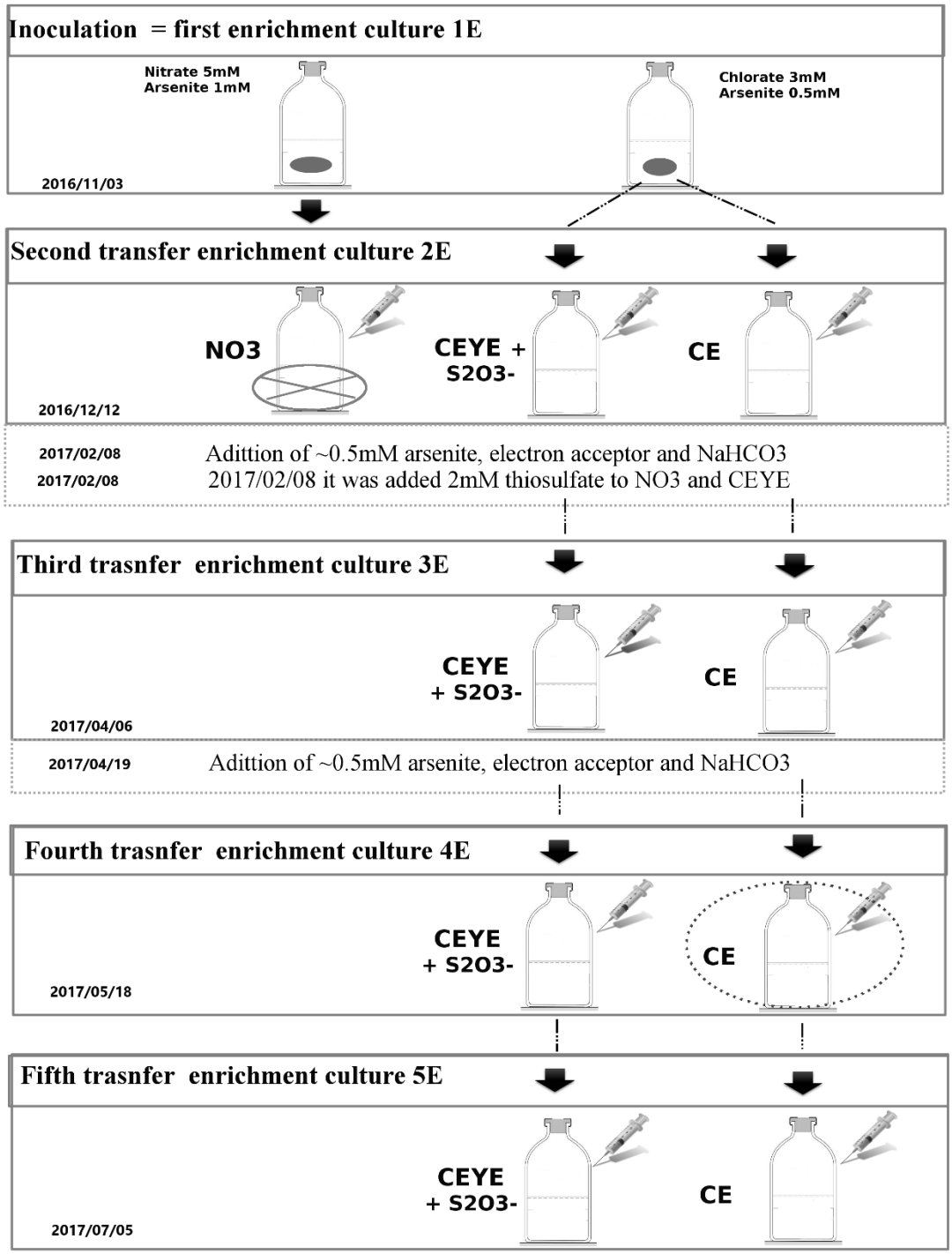

Figure 4-1. Enrichment procedure. CEYE: correspond to arsenite/chlorate enrichment culture, with added yeast extract. CE: correspond to arsenite/chlorate enrichment culture, without added yeast extract. The culture encircles by dots lines was used for the isolation of strain M52. Enrichment culture of arsenite/nitrate was not continued after the second transfer due to low arsenite transformation. Arsenate concentration of each enrichment culture can be seen in Table 4-1. The microbial community of the CEYE and CE enrichment cultures are described in Figure 4-2 and Table 4-2.

Table 4-1. Approximate arsenate concentration detected in the enrichment cultures.

Enrichment culture		Arsenate concentration (mM)					
name	culture content	2E	3E 13 days incubation	3E 42 days incubation	3E 63 days incubation	4E	5E
NO3	As^{3+}/NO_3-	0.3					
CEYE	As^{3+}/ClO_3-/Yeast Extract	3.71	1.39	0.98†	1.36†	0.11†	n.d.
CE	As^{3+}/ClO_3-	2.96	2.09	2.4	2.57	2.29	1.7

Abbreviations: 2E, second transfer. 3E, third transfer. 4E, fourth transfer. 5E, fifth transfer.

† After addition of thiosulfate 2 mM.

n.d. undetermined.

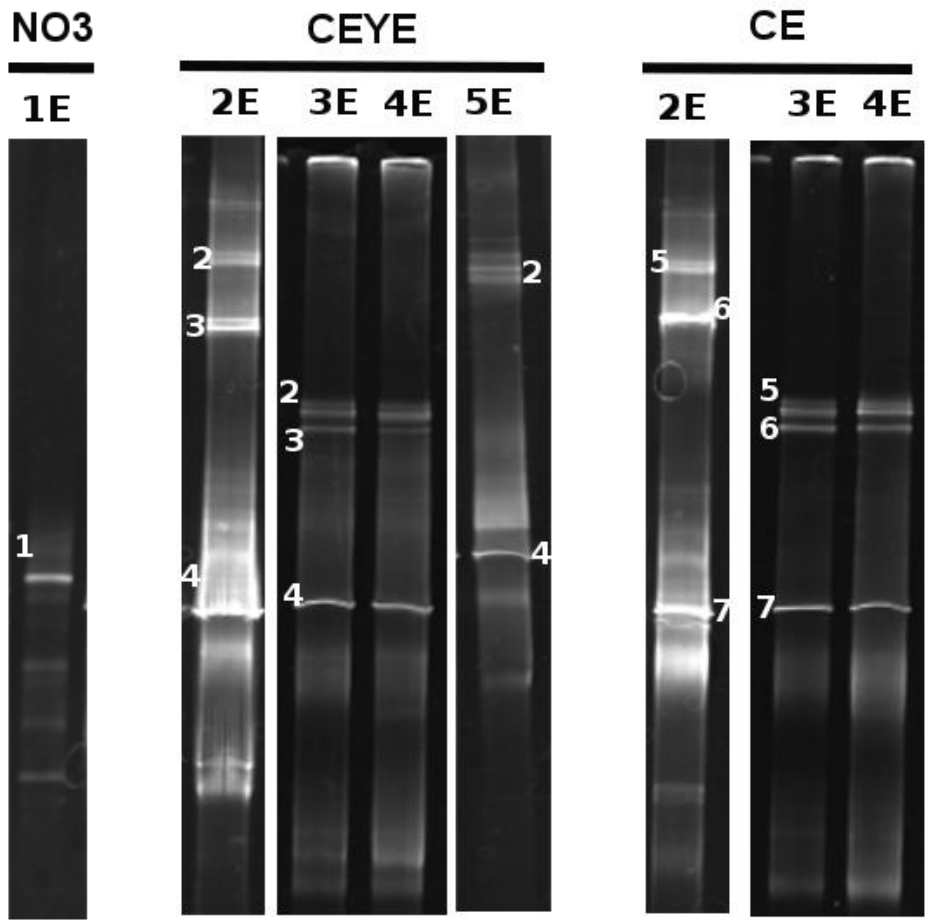

Figure 4-2 Image of the 16S rRNA gene fragments DGGE gel comparing the community of the enrichment cultures until the fourth to fifth transfers. Abbreviations: 2E, second transfer. 3E, third transfer. 4E, fourth transfer. 5E, fifth transfer. NO3, enrichment culture with As^{3+}/NO_3^-. CEYE, enrichment culture with As^{3+}/ClO_3^-/Yeast Extract. CE, enrichment culture with As^{3+}/ClO_3^-. Numbers represent the major bands selected for sequencing of 16S rRNA gene fragments from DGGE gels. Equal numbers represent the same band within CEYE and CE enrichment cultures. Results of the community composition and affiliation of the major bands can be observed in table 4-2.

Table 4-2 Community composition of the enrichment cultures based on the phylogenetic affiliation of the major bands obtained from DGGE gels to the closest cultured relatives determined according to the results of BLASTn search.

Enrichment culture		number of band	Closest cultured relative	QC	E value	Ident.	related phylum
name	culture content						
NO3	As^{3+}/NO_3^-	band 1	*Shewanella* sp. strain C17	100%	0	99%	*Proteobacteria*
CEYE	As^{3+}/ClO_3^-/Yeast Extract	band 2	Bacterium YC-LK-LKJ2	98%	3.00E-146	86%	*Ignavibacteriae*
		band 3	*Ignavibacterium album* strain JCM 16511	100%	0	99%	
		band 4	*Georgfuchsia toluolica* strain G5G6	100%	0	96%	*Proteobacteria*
CE	As^{3+}/ClO_3^-	band 5	Bacterium YC-LK-LKJ2	99%	5.00E-149	87%	*Ignavibacteriae*
		band 6	*Ignavibacterium album* strain JCM 16511	100%	3.00E-149	99%	
		band 7	*Georgfuchsia toluolica* strain G5G6	100%	0	96%	*Proteobacteria*

Abbreviation: QC, query coverage. Ident., identity percentage.

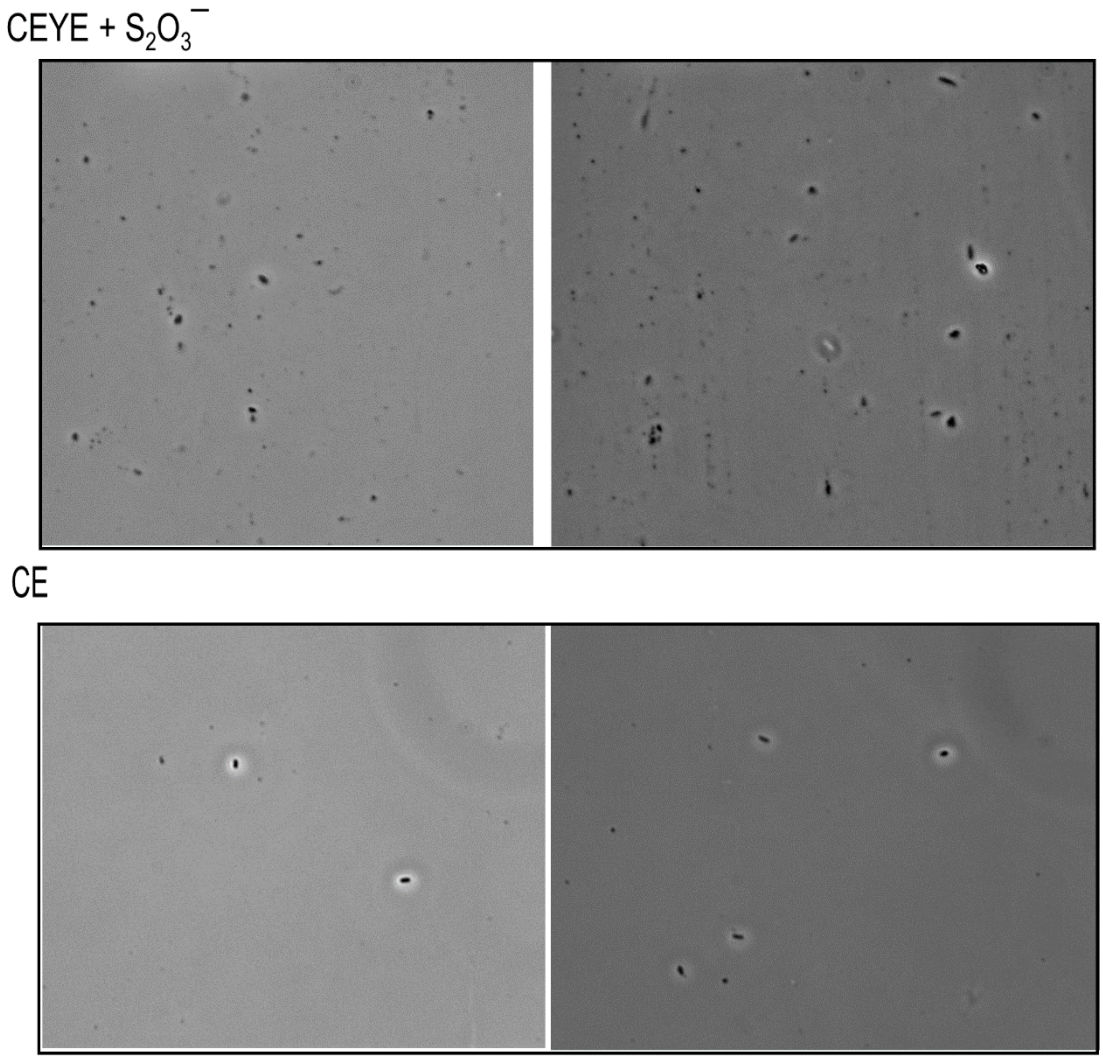

Figure 4-3. Observation under microscope of the enrichment cultures with and without addition of thiosulfate. Abbreviation: CEYE, enrichment culture with As^{3+}/ClO_3^-/Yeast Extract. CE, enrichment culture with As^{3+}/ClO_3^-.

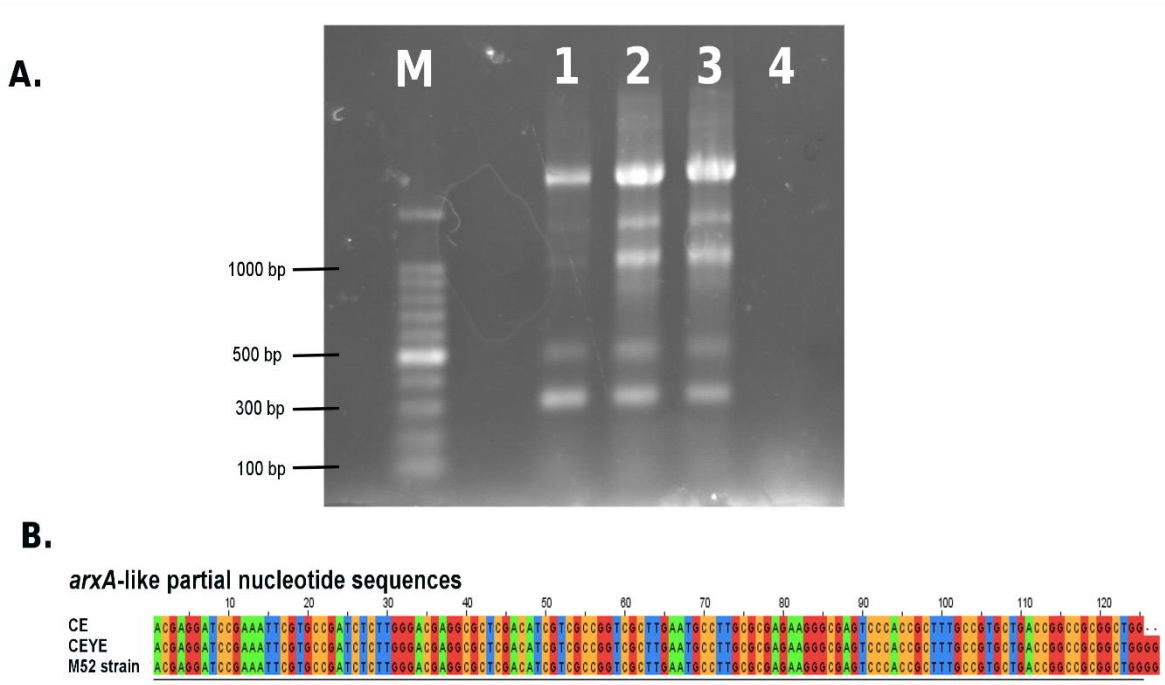

Figure 4-4. **A**. Detection of a partial *arx*A –like sequence in the CEYE, enrichment culture with As^{3+}/ClO_3^-/Yeast Extract; and CE, enrichment culture with As^{3+}/ClO_3^-. Lanes: 1.) 100 bp DNA ladder as a size marker; 2.) CE fourth enrichment culture E4; 3.) CEYE fourth enrichment culture E4; 4.) CE fourth enrichment culture E5; 5.) negative control, water as the template.

B. Comparison of the nucleotide sequences of the 300 bp amplified PCR product and the *arxA* –like sequence in strain M52, isolated from CE enrichment culture E4. The alignment is colored taking into account the nucleotide residue.

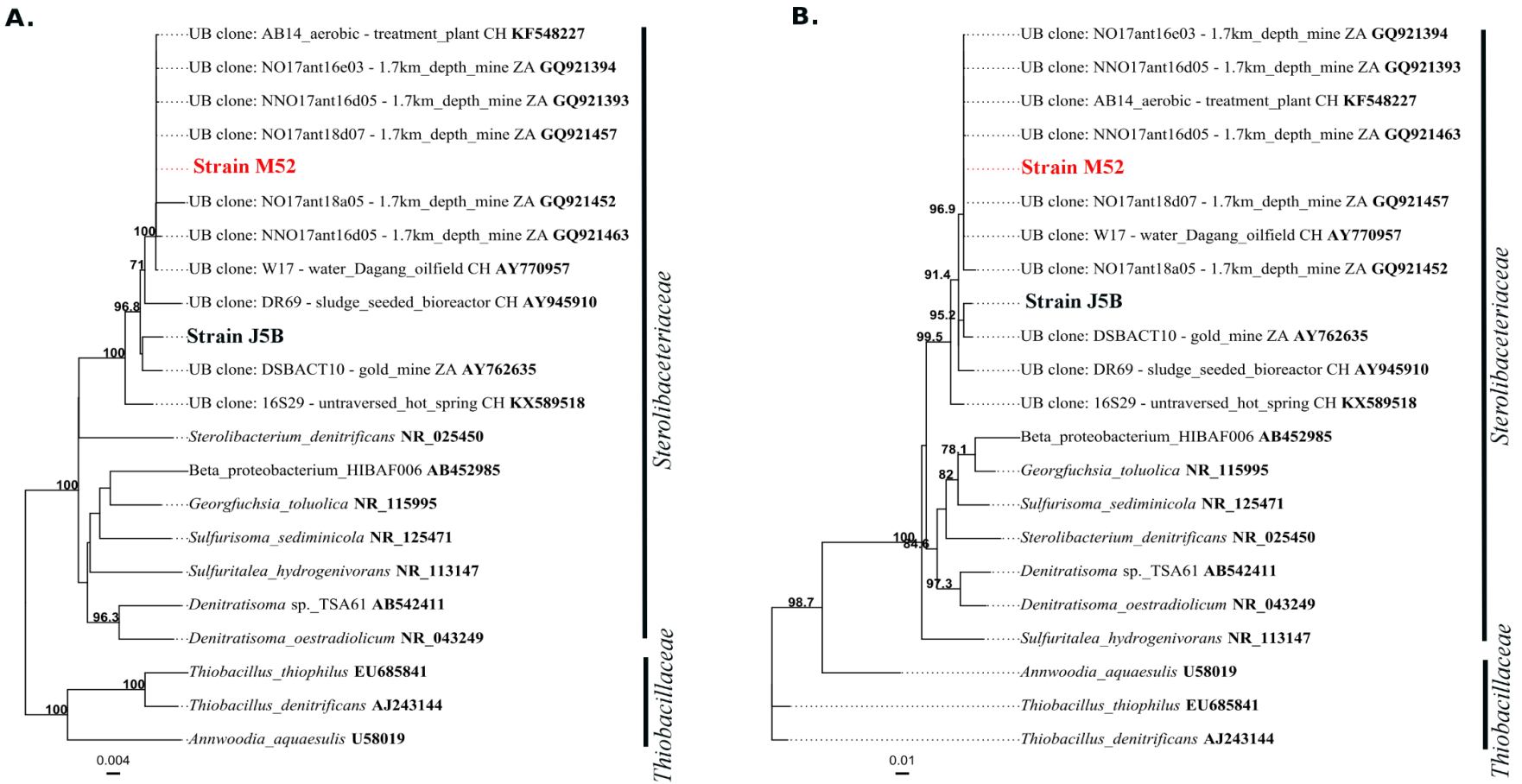

Figure 4-5. Phylogenetic tree based on 16S rRNA gene including uncultured and cultured closest relatives. **A**. Neighbor-Joining phylogeny with the p-distance model. **B**. Maximum likelihood phylogeny with LG substitution model. Support values above 70% are shown. Species from the family *Thiobacillaceae* were used as an outgroup. ZA: correspond to South Africa. CH: correspond to China. The accession number is in bold. UB clone means uncultured bacteria clone.

Table 4-3. Phylogenetic affiliation based on the closest cultured relatives of the *arx*-like operon genes in strain M52, determined according to the results of BLASTp search.

Putative gene	Putative protein and function	closest cultured relatives	Identity %	QC	Accession number
arx B'	4Fe-4S ferredoxin protein	*Sulfuricella denitrificans* skB26^{T}	41%	93%	WP_021035848
arx D	cytoplasmic chaperone TorD family protein	*Azoarcus tolulyticus*	37%	95%	WP_076602995
arx C	Polysulfide reductase (NrfD) protein	*Rhodocyclaceae* bacterium	69%	91%	WP_102043118
arx B	4Fe-4S ferredoxin protein	*Rhodocyclaceae* bacterium	82%	97%	WP_102041239
arx A	molybdopterin oxidoreductase protein	*Sulfuricella denitrificans* skB26^{T}	62%	99%	WP_009207744
arx X	ABC transporter substrate-binding protein	*Sulfuricella denitrificans* skB26^{T}	59%	97%	WP_041096742
arx S	two-component sensor histidine kinase protein	*Rhodocyclaceae* bacterium	48%	99%	WP_102041242
arx R	sigma-54-dependent Fis family transcriptional regulator protein	*Rhodocyclaceae* bacterium	63%	98%	WP_102041243

Abbreviation: QC, query coverage.

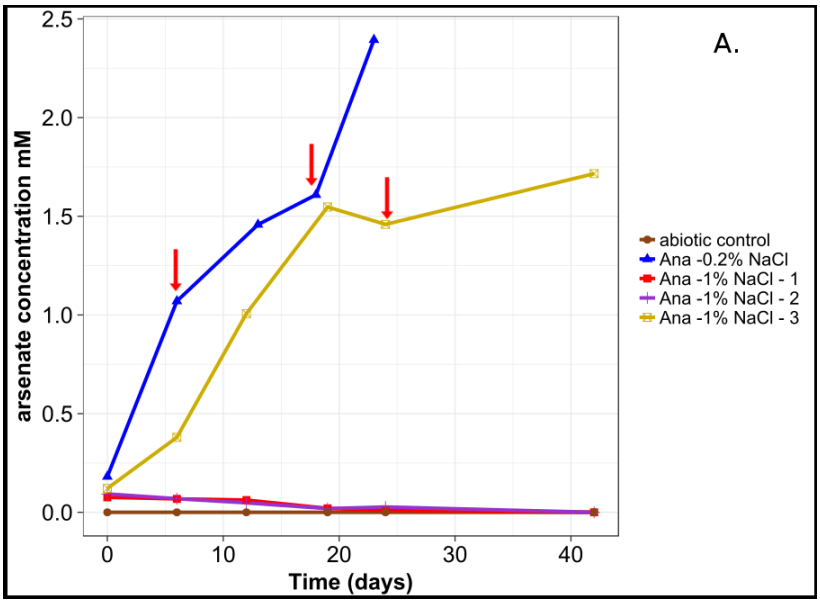

Figure 4-6. Line plot for the arsenite transformation experiments measured every 6-7 days. Results of each replicate in the anaerobic culture experiment made with 1% NaCl are shown in separated lines: red line, replicate No. 1; purple line, replicate No. 2; and, yellow line, replicate No. 3. The results from the anaerobic experiment with 0.2% NaCl are shown as the average arsenate concentration of the three replicate in the blue line. Abiotic control is shown in brown line. The standard deviation can be seen in table 4-4. The red arrow indicates the addition of ~0.5mM arsenite, NaHCO3, and 1 mg/L YE.

Figure 4-7. Image of the arsenite oxidation anaerobic experiment cultures made with 1% NaCl and inoculated with strain M52, and rubber stopper used to cap the serum bottles of replicate 1, 2 and 3.

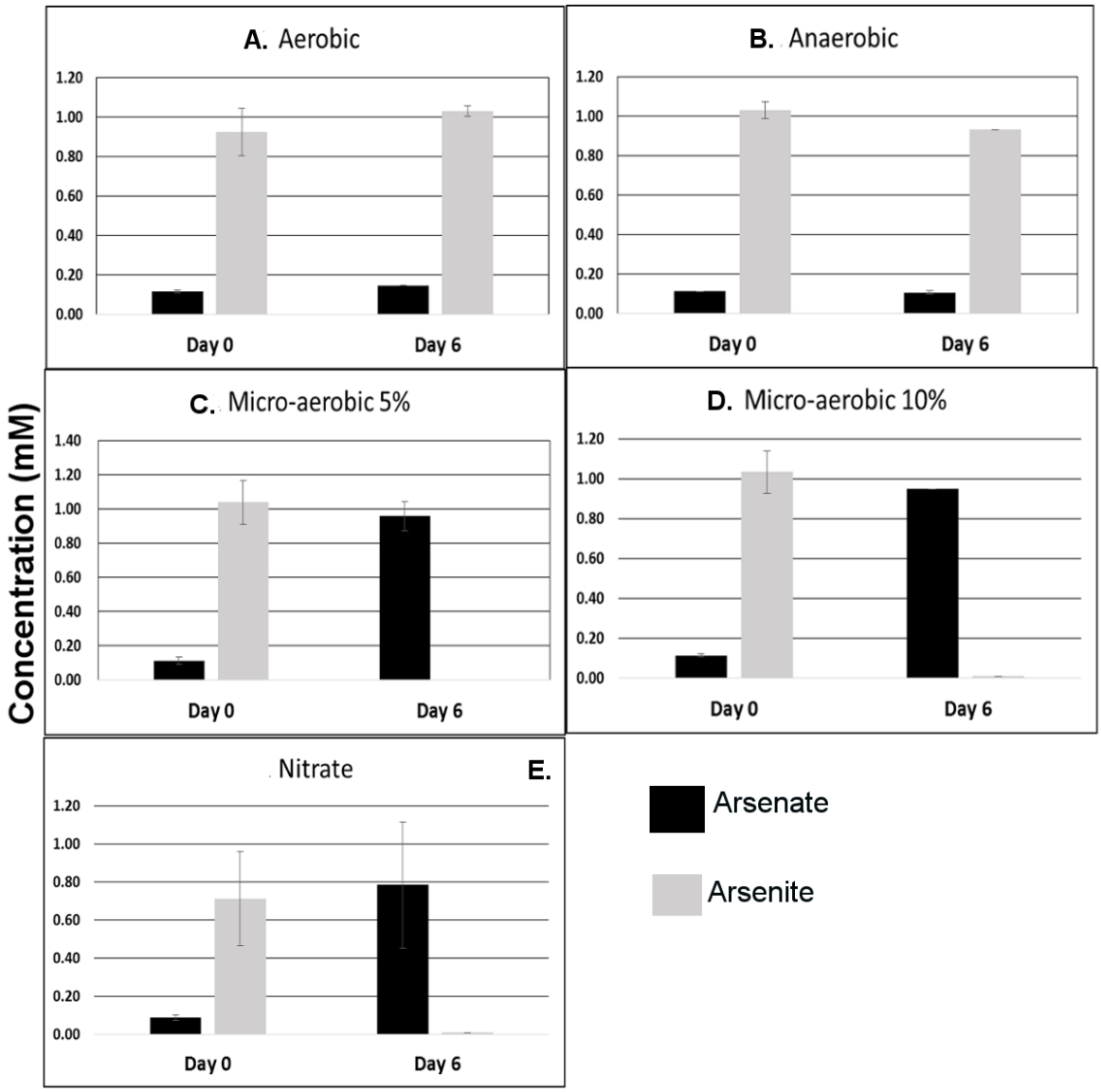

Figure 4-8. Average arsenate and arsenite concentration on day 0 and day 6 of incubation, under aerobic, micro-oxic, and anaerobic conditions (with and without nitrate as an electron acceptor). A) Aerobic, air in the headspace. B) Anaerobic, nitrogen in the headspace, without an electron acceptor. C) Micro-aerobic 5%, N_2/Air 95: 5, % in headspace. D) Micro-aerobic 10%, N_2/Air 90: 10, % in headspace. G) Anaerobic + Nitrate, culture supplemented with ~10 mM of NO_3^-.

Table 4-4. Arsenate and arsenite concentration measured in every time point for each experiment.

Experiment		Replicate	Arsenate concentration (mM)					
name	culture content		Day 0	Day 6	Day 12	Day 19	Day 24	Day 42
Anaerobic 1%NaCl	1 mM As^{3+}, 1% NaCl, 1 mg/L YE	1	0.08	0.15	0.27	0.17	0.05	0.00
		2	0.09	0.08	0.38	0.07	0.07	0.00
		3	0.12	0.45	1.01	1.55	1.46	1.72
Anaerobic 0.2%NaCl -thinner rubber stopper	1 mM As^{3+}, 0.2% NaCl, 1 mg/L YE	1	0.23	1.24	1.77	1.91	2.55	
		2	0.14	0.59	0.89	1.32	2.38	
		3	0.17	1.38	1.71	1.59	2.25	
		Average	0.18	1.07	1.46	1.61	2.39	
Aerobic	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + Air in the head space	1	0.08	0.04	0.03			
		2	0.10	0.12	0.01			
		3	0.09	0.03	0.02			
		Average	0.09	0.06	0.02			
Nitrate	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + NO_3-	1	0.1	1.02				
		2	0.08	0.55				
		Average	0.09	0.79				
Anaerobic	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE	1	0.11	0.1				
		2	0.11	0.11				
		Average	0.11	0.11				
Micro-aerobic 5% Oxygen	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + 5% Oxygen	1	0.1	1.02				
		2	0.13	0.9				
		Average	0.12	0.96				
Micro-aerobic 10% Oxygen	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + 10% Oxygen	1	0.12	0.95				
		2	0.11	0.95				
		Average	0.12	0.95				
Abiotic control	1 mM As^{3+}, 1% NaCl - No inoculation	1	nd	nd	nd	nd	nd	nd
		2	nd	nd	nd	nd	nd	nd
		3	nd	nd	nd	nd	nd	nd

Abbreviation: nd, not detected. Detection limit of method = 0.01 mM arsenate.

Table 4-5. Arsenite concentration measured in every time point for each experiment.

Experiment		Replicate	Arsenite concentration (mM)	
name	culture content		Day 0	Day 6
Aerobic	1 mM As3+, 0.2% NaCl, 2 mg/L YE + Air in the head space	1	1.11	1.02
		2	0.94	1.06
		3	0.91	1.00
		Average	0.93	1.03
Nitrate	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + NO_3-	1	0.89	ud
		2	0.54	ud
		Average	0.72	-
Anaerobic	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE	1	1	0.93
		2	1.06	0.93
		Average	1.03	0.93
Micro-aerobic 5% Oxygen	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + 5% Oxygen	1	0.95	0
		2	1.13	0
		Average	1.04	0
Micro-aerobic 10% Oxygen	1 mM As^{3+}, 0.2% NaCl, 2 mg/L YE + 10% Oxygen	1	1.11	0.01
		2	0.96	0.01
		Average	1.04	0.01

Abbreviation: ud, undetermined.

Chapter 5

General Discussion

5.1 Discussion

The focal point in chapter 2 and 3 was the ecological and phylogenetical distribution of *arxA* sequences on three non-alkaline saline environments by a molecular approach based on two newly designed primer set. The phylogenetic analyses of the clone libraries established the presence of three persistent and highly supported clusters that were named cluster 1, cluster 2 and cluster 3 (Fig. 2-2 and Fig 3-10). The OTUs obtained in both clones' libraries fell exclusively within cluster 2 and 3, which also led to reveal the marked distinction of cluster 1, as a specific lineage of *arxA* in organism recovered from saline-alkaline environments. Therefore, only the analysis on cluster 2 and cluster 3 will be described in the general discussion.

The cluster 2 was the most abundant and diverse, including OTUs from all the sample analyzed. It was characterized for being formed by *arxA* sequences holding a long or short insertion, as it was described in chapter 3. *Sulfuricella denitrificans* skB26^{T} and also, three more isolated strains from Jozankei microbial mat, *Sulfuritortus calidifontis* J1A^{T}, strain J5B, and strain M52 formed a branch within this cluster, and correspond the only cultivated representatives.

The cluster 2 are formed by partial and complete *arxA* sequences recovered from a diversity of non-saline alkaline places, mainly from sites with oxygen-limiting conditions, including: hot spring microbial mat, deep water samples from a lake (1), CO2 driven geyser (2), groundwater sample (3) and Riverbend sediment in a hot spring area (4). However, to this date, the only report of an anaerobic arsenite oxidation mediated by the Arx- arsenite oxidase came from an arsenic-rich soda lake, Mono Lake in California (5).

Strain M52, isolated from a microbial mat obtained in Jozankei, and also a branch within cluster 2, was found to be capable of anaerobic arsenite oxidation with nitrate as an electron

acceptor. Also, it was experimentally demonstrated the arsenite oxidation under low concentration of O_2 (5% and 10% air in the headspace of culture bottle), but not under fully aerobic conditions (chapter 4). The genome of strain M52 contains homologous to high-affinity terminal oxidases, cytochrome *cbb3*, and cytochrome *bd* oxidases, involved in respiration under low tension of oxygen (6, 7). These oxidases have extremely low Km values for oxygen (7 nM in the case of *cbb3* oxidase, and 3 – 8 nM in the case of *bd*-type oxidases), which facilitate bacteria colonize semi-anoxic environments (8, 9).

Alkalilimnicola ehrlichii MLHE-1, the first isolated *arx*-harboring bacterium, is a facultative anaerobe also incapable of oxidizing the arsenite under fully aerobic conditions as it was reported by Hoeft et al. (10). However, inspection of the genome revealed the presence of the high-affinity terminal oxidases, cytochrome *cbb3*, and cytochrome *bd* oxidases; later detected in the genome of other *arx*-containing arsenite-oxidizing bacteria as well (Table 5-1). It would be interesting to verify if these cytochromes oxidases are expressed during the arsenite oxidation on these strains, and if these oxidases would allow them to use oxygen as an electron acceptor. Regardless of this, in light of the results obtained in this study, it is proposed that the *arx* arsenite oxidation may also be involved in aerobic respiration under micro-oxic conditions.

Only a few *arx* harboring arsenite-oxidizing bacteria have been obtained in pure cultures, predominantly affiliated to class *Gammaproteobacteria*, and whose *arxA* gene belongs to cluster 1 (table 5.2). Strain M52, however, is affiliated to the class *Betaproteobacteria*, and despite other bacteria of this class were found holding an *arx*-like gene in chapter 3, to our best knowledge, strain M52 is the only experimentally demonstrated for being capable of oxidizing the arsenite.

Another distinct feature in strain M52 from the rest of the cultured *arx* arsenite-oxidizing bacteria is the long insertion within the ArxA sequence formed by 99-amino acid between positions 319 and 417. This unique characteristic is exclusively present in sequences that belong to cluster 2. Thus, strain M52 would also become the first cultured arsenite-oxidizing bacterium with an *arxA* gene with a long insertion.

On the other hand, the cluster 3 was dominated instead by OTUs obtained from Lake Mizugaki samples. Notably, the PCR primer designed in chapter 3 allows us to detect more unique clones in this site, than the one used in chapter 2. One possible explanation would be the difference in the *arxA* diversity within the water column of Lake Mizugaki, which may have influenced the outcome since it was use the DNA of two distinct water depth samples. However, both results reflect that there is a considerable proportion of the *arxA* sequences closely related to that of *Azoarcus* species (cluster 3) in Lake Mizugaki.

Previously it was demonstrated the presence of genes involved in dissimilatory arsenate reduction (*arrA*) from the same 40-meter depth water DNA sample of Lake Mizugaki used in this study (chapter 3). The *arrA* sequences were related to *Sulfuritalea hydrogenivorans* sk43H^{T} (91-99% amino acid identities) (11), a bacterium formerly isolated from Lake Mizugaki (12) experimentally demonstrated to be capable of growth via arsenate respiration (11). This sample was reported to be hypoxic nitrate-depleted, with a total arsenic concentration of 1.5 μM (11). For the *Sulfuritalea* species survive in this conditions using the arsenate as an electron acceptor, it would be needed the arsenate to be replenished. Thus, under anaerobic conditions, the arsenite-oxidizing bacteria may play an essential role at 40 meters in Lake Mizugaki, to support the flow of arsenate on a simple arsenic cycle, probably by the microbial transformation between arsenite and arsenate.

In conclusion, this study revealed the presence of many as-yet uncultured *arx* harboring bacteria involved in probably anaerobic/microaerobic arsenite oxidation by using newly designed PCR primers pairs targeting the *arxA* gene. The results obtained may contribute to the knowledge of the genetic diversity of the *arx* gene in non-alkaline and low-salinity environments. Also, a novel betaproteobacterium strain M52 capable of oxidizing the arsenite under micro-aerobic conditions was isolated. The strain M52 may be helpful to further research on arsenite oxidation biological process, and also increase the spectrum of possible alternatives metabolic pathways.

5.2 Reference

1. **Kojima H, Watanabe T, Iwata T and Fukui M**. 2014 Identification of major planktonic sulfur oxidizers in stratified freshwater lake. PLoS One **9**:1–7.

2. **Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, Hug LA, Burstein D, Emerson JB, Thomas BC and Banfield JF**. 2017 Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations. Environ. Microbiol. **19**:459–474.

3. **Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS and Banfield JF**. 2016 Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. **7**:1–11.

4. **Zargar K, Conrad A, Bernick DL, Lowe TM, Stolc V, Hoeft S, Oremland RS, Stolz J and Saltikov CW**. 2012 ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. Environ. Microbiol. **14**:1635–1645.

5. **Oremland RS, Saltikov CW, Stolz JF and Hollibaugh JT**. 2017 Autotrophic microbial arsenotrophy in arsenic-rich soda lakes. FEMS Microbiol. Lett. **364**:.

6. **Pitcher RS and Watmough NJ**. 2004 The bacterial cytochrome cbb3 oxidases. Biochim. Biophys. Acta - Bioenerg. **1655**:388–399.

7. **Borisov VB, Gennis RB, Hemp J and Verkhovsky MI**. 2011 The cytochrome bd respiratory oxygen reductases. Biochim. Biophys. Acta - Bioenerg. **1807**:1398–1413.

8. **Huang Y, Reimann J, Singh LMR and Ädelroth P**. 2010 Substrate binding and the catalytic reactions in cbb3-type oxidases: The lipid membrane modulates ligand binding. Biochim. Biophys. Acta - Bioenerg. **1797**:724–731.

9. **Morris RL and Schmidt TM**. 2013 Shallow breathing: Bacterial life at low O 2. Nat. Rev. Microbiol. **11**:205–212.

10. **Hoeft SE, Blum JS, Stolz JF, Tabita FR, Witte B, King GM, Santini JM and Oremland RS**. 2007 Alkalilimnicola ehrlichii sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. Int. J. Syst. Evol. Microbiol. **57**:504–512.

11. **Watanabe T, Miura A, Iwata T, Kojima H and Fukui M**. 2017 Dominance of Sulfuritalea species in nitrate-depleted water of a stratified freshwater lake and arsenate respiration ability within the genus. Environ. Microbiol. Rep. **9**:522–527.

12. **Kojima H and Fukui M**. 2011 Sulfuritalea hydrogenivorans gen. nov., sp. nov., a facultative autotroph isolated from a freshwater lake. Int. J. Syst. Evol. Microbiol. **61**:1651–1655.

Table 5-3. Oxygen requirement comparisons between strain M52 and known *arxA* gene-containing arsenite-oxidiing strains with available genomes, and presence of respiratory oxygen reductases genes expressed under low O_2-concentration.

Strain name	Oxygen requirement	cytochrome *c* oxidases - cbb3-type (*ccoNOPQ*)	Accesion number	cytochromes *bd* (*cydAB*)	Accesion number
Alkalilimnicola ehrlichii MLHE-1	Facultative anaerob	+	WP_011629620-WP_011629619-WP_011629617-WP_011629618	+	WP_011629781-WP_011629780
Ectothiorhodospira sp. PHS-1	anaerobe	+	WP_008931898-WP_008931899-WP_008931901-WP_008931900	+	WP_040405020-WP_008931270
Ectothiorhodospira sp. BSL-9	anaerobe/ aerotolerant	+	WP_063463760-WP_063465967-WP_063463758-WP_063463759	+	WP_063465447-WP_063465446
Halorhodospira halophila SL1/DSM 244	anaerobe	(CooP partial protein)	WP_011812935	+	WP_011815205-WP_011815204
Strain M52	Microerophilic	+	n.s.	+	n.s.

Table 5-2. Physiological, phylogenetic and phenotypic comparisons between strain M52 and known *arxA* gene-containing strain.

strain name	affiliation	Isolation source	arsenite oxidation	Reference[¶]
Alkalilimnicola ehrlichii MLHE-1	*Gammaproteobacteria*	Water column of Mono Lake, CA, United States	Anaerobically on 10 mM nitrate	Oremland R et al. 2002; Hoeft S. et al. 2007; Zargar et al. 2010, 2012
Ectothiorhodospira sp. PHS-1	*Gammaproteobacteria*	Red biofilm associated to a hot spring in Paoha Island of Mono Lake, CA, United States	Anaerobic phototrophic growth	Kulp T. et al. 2008; Hoeft S. et al. 2010
Halomonas sp. ANAO-440[†‡]	*Gammaproteobacteria*	An alkaline saline lake in Mongolia	Anaerobically on 10 mM nitrate and 0.5% (wt/vol) Luria broth	Hamamura N.et al. 2014
Ectothiorhodospira sp. BSL-9	*Gammaproteobacteria*	Big Soda Lake, NV, United States	Anaerobic phototrophic growth	Hoeft S. et al. 2017; Hernandez-Maldonado J et al. 2017
Halorhodospira halophila SL1/DSM 244	*Gammaproteobacteria*	Summer Lake, OR, United States	Anaerobic phototrophic growth with succinate (6 mM) and sulfide (0.5 mM)	Challacombe J et al. 2013; Hernandez-Maldonado J et al. 2017
Ectothiorhodospira sp. MLW-1[†‡]	*Gammaproteobacteria*	Sediment north shore of Mono Lake, Ca, United State	Anaerobic phototrophic growth	Hoeft S. et al. 2017
Desulfotomaculum sp. TC-1[†]	*Firmicutes*	Sediment from a hot spring of Tengchong, China	Anaerobically with sulfate and lactate	Wu, G et al., 2017
Strain M52	*Betaproteobacteria*	A microbial mat exposed to overflowing hot spring water, HKD, Japan	Anaerobically with nitrate or microaerphilic condition	The present study

[†] Not available genome

[‡] Not available *arxA* gene sequence

[¶] Reference with evidenced arsenite oxidation

Acknowledgements

I would like to express my most profound thank to Dr. Manabu Fukui for giving me the opportunity to come to Japan and be part of his laboratory, his support throughout the process of the scholarship, and also his sincere advice, guidance, and encouragement during all my doctoral course. Also, my gratitude to Dr. Kojima Hisaya for all the discussions, for the support of my work and guidance, for his patience and assistance that helped me get results of better quality. I am also grateful to Dr. Mia Terashima, Dr. Tomohiro Watanabe, Dr. Aya Miura, Dr. Watanabe Miho, Dr. Umezawa Kazuhiro and Dr. Hideaki Ogata for their support and remarks that make me overcome numerous obstacles I have been facing through my research.

I would like to thank my past and present fellow labmates for their feedback, cooperation and of course friendship. Also, I would like to express my gratitude to the staff Tokizawa Riho, Shinohara Arisa for their aid in many technical situations, and their never-ending kindness.

I also would like to thank my friends in Colombia, and the ones I made along the way in Japan, for their constant encouragements, their funny comments that alleviate many difficult times. And last but not least, all my big family, for accepting nothing less than excellence from me, for supporting me spiritually throughout my doctoral course and especially during the writing of this thesis. And finally to the love of my live, Santiago Ruiz, which presence make this journey the best one I have ever lived. I dedicate this work to him