PhD thesis

Microbial populations inhabiting extreme environments affected by severe arsenic and mercury pollution

Poblaciones microbianas en ambientes extremos contaminados con arsénico y mercurio

Alexander A. Prosenkov

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

Functional Biology
PhD program in Molecular and Cellular Biology

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: POBLACIONES MICROBIANAS EN AMBIENTES EXTREMOS CONTAMINADOS CON ARSÉNICO Y MERCURIO.	Inglés: MICROBIAL POPULATIONS INHABITING EXTREME ENVIRONMENTS AFFECTED BY SEVERE ARSENIC AND MERCURY POLLUTION.

2.- Autor	
Nombre: ALEXANDER PROSENKOV	DNI/Pasaporte/NIE:
Programa de Doctorado: BIOLOGÍA MOLECULAR Y CELULAR	
Órgano responsable: CENTRO INTERNACIONAL DE POSTGRADO	

RESUMEN (en español)

Previamente al abandono de la minería de mercurio, la provincia montañosa de Asturias, localizada en el norte de España, se encontraba entre las regiones productoras de mercurio más importantes del mundo. El alto contenido de arsénico de los minerales de mercurio locales, unido a la extracción y producción a gran escala de este metal, provocaron importantes contaminaciones en una serie de antiguos distritos mineros, siendo el sitio de extracción y fundición denominado “El Terronal”, localizado cerca de la ciudad de Mieres, el más afectado. En este emplazamiento, grandes cúmulos de desechos ricos en arsénico y mercurio altamente tóxicos contribuyeron a la contaminación del suelo y el agua debido a la lixiviación por lluvias y dispersión por viento de óxidos de arsénico. Adicionalmente a esta contaminación, la presencia de altos niveles de contaminación por metales y metaloides puede conducir a una mayor incidencia de transferencia horizontal de genes de resistencia a estos elementos dentro de las comunidades microbianas, aspecto que también puede resultar en la transferencia conjunta con genes de resistencia a los antibióticos, convirtiendo el sitio contaminado en un reservorio de genes de resistencia a antibióticos. No obstante, desde un ángulo más positivo, evolucionar en condiciones tan extremas podría dar lugar a microorganismos con propiedades útiles para diversas aplicaciones biotecnológicas.

En este trabajo se lleva a cabo un estudio de la microbiota, enfocada en las poblaciones microbianas y cepas bacterianas individuales, presente en los suelos y sedimentos de aguas subterráneas altamente contaminados del emplazamiento “El Terronal”, así como también aquellos microorganismos presentes en los ambientes extremos, tóxicos y oligotróficos correspondientes a los cúmulos de desechos.

Para caracterizar las comunidades procarióticas, fúngicas y SAR en los suelos, sedimentos de aguas subterráneas y tres tipos diferentes de desechos (hollín rico en arsénico, residuos de condensadores de mercurio y polvo de chimenea), se llevó a cabo la secuenciación de amplicones de genes de ARNr 16S y ARNr 18S utilizando la tecnología de secuenciación por síntesis Illumina, con un análisis posterior de diversidad y taxonomía, así como también la reconstrucción metagenómica basada en el ARNr 16S. Los resultados revelaron la presencia de comunidades microbianas muy diversas en los suelos contaminados, semejantes a las encontradas en las muestras de polvo de chimenea, encontrándose menos diversidad en estas últimas. Las muestras de hollín ricas en arsénico mostraron una diversidad significativamente menor y estaban colonizadas principalmente por el filo Proteobacteria. La muestra de residuos de condensadores de mercurio, que representa el más extremo de los entornos muestreados, se encontraba conformada principalmente por el dominio Archaea. En contraste con las comunidades procariotas bien diferenciadas, las poblaciones de hongos y SAR de las muestras de hollín ricos en arsénico y de residuos de condensadores de mercurio eran muy similares. Los sedimentos de aguas subterráneas albergaban diversas poblaciones microbianas con poca similitud con las muestras de suelo y cúmulos de desechos.

El estudio de las cepas bacterianas individuales se centró en las bacterias cultivables. Se identificaron 59 cepas aisladas mediante la secuenciación de sus genes de ARNr 16S y se caracterizaron con respecto a su resistencia a metales y metaloides (mediante técnicas de cultivo y biología molecular), su resistencia a antibióticos y según la capacidad para degradar y emulsionar hidrocarburos. Todas las cepas mostraron cierto grado de resistencia al arsénico y al mercurio, indicando la mayoría de ellas niveles muy altos de resistencia. El mecanismo de resistencia comúnmente empleado por estas bacterias fue la expulsión de As (III) fuera de la célula. Varias cepas manifestaron altos niveles de resistencia a los antibióticos, pero no se detectó correlación con la resistencia al arsénico o al mercurio. Varias cepas mostraron capacidad para degradar hidrocarburos o crearon emulsiones de hidrocarburos estables.

Se seleccionaron diecisiete cepas que mostraban una alta resistencia a los metales y metaloides, alta resistencia a antibióticos, capacidad para degradar o emulsionar hidrocarburos o una combinación de esos rasgos, y se secuenciaron sus genomas mediante secuenciación shotgun Illumina. Los genomas resultantes se ensamblaron, anotaron y se utilizaron para la determinación de genes de resistencia al arsénico, mercurio y antibióticos. La expulsión de arsenito fue el método de detoxificación de arsénico más común. En un gran número de cepas se detectó el operón mer, implicado en la resistencia al mercurio. Adicionalmente, se encontraron diversos genes muy similares (pero no idénticos) a los genes conocidos de resistencia a los antibióticos, lo que sugiere un peligro potencial de transferencia horizontal de estos genes a bacterias clínicamente relevantes.

RESUMEN (en Inglés)

Before abandonment of mercury mining, the mountainous region of Asturias in the north of Spain was among the most important mercury-producing provinces in the world. High arsenic content of local mercury ores and large scale of mercury mining and production led to the extensive contamination in several former mining districts, with El Terronal mining & smelting site near the town of Mieres being the most affected. Several large heaps of highly toxic arsenic- and mercury-rich waste contributed to the soil and water pollution both due to leaching by rainwater and dispersal of arsenic oxide particles by wind. In addition to this danger, high levels of metal(loid) pollution can lead to increased incidence of horizontal transfer of metal(loid) resistance genes within microbial communities, which could result in conjoined transfer of antibiotic resistance genes as well, converting contaminated site into a reservoir of antibiotic resistance genes. From a more positive angle, evolving in such extreme conditions could produce microorganisms with properties that could be of use in biotechnological applications.

This work studies microbial populations as a whole as well as individual bacterial strains inhabiting severely contaminated soils and groundwater sediments at the El Terronal site, as well as extreme, highly toxic & nutrient-poor environments of the waste heaps.

To characterise prokaryotic, as well as fungal and SAR communities in soils, groundwater sediments and three different types of waste (arsenic-rich soot, stupp from the mercury condensers, and flue dust), 16S rRNA and 18S rRNA amplicon sequencing was performed using Illumina sequencing-by-synthesis technology, with subsequent analysis of diversity and taxonomy, as well as 16S rRNA-based metagenomic reconstruction. Obtained data revealed highly diverse microbial communities in contaminated soils, similar but less diverse communities in the flue dust. Arsenic-rich soot samples were much less diverse, and populated mostly by *Proteobacteria*. Stupp sample, representing the most extreme of the environments found on the site, was populated mostly by *Archaea*. In contrast to well-differentiated prokaryotic communities, fungal and SAR populations of the arsenic-rich soot and stupp were very similar. Groundwater sediments housed diverse microbial populations with little similarity to the soil and waste samples.

Study of individual microorganisms was focused on cultivable *Bacteria*. Fifty nine isolated strains were identified by sequencing their 16S rRNA genes, and characterised in respect to their metal(loid) resistance (using both cultivation and molecular biology techniques), antibiotic resistance, and capacity to degrade and emulsify hydrocarbons. All strains have shown some degree of arsenic and mercury resistance, with most strains showing very high levels of resistance. As(III) efflux was the most commonly employed resistance mechanism. A number of strains had high levels of antibiotic resistance, but it had no correlation with arsenic or mercury

Universidad de Oviedo
Universidá d'Uviéu
University of Oviedo

resistance. Several strains displayed capacity to degrade hydrocarbons or created stable hydrocarbon emulsions.

Seventeen strains displaying either high metal(loid) resistance, high resistance to antibiotics, capacity to degrade or emulsify hydrocarbons, or a combination of those traits were selected, and their genomes were sequenced using Illumina shotgun sequencing. Genomes were assembled and annotated, and were used for determination of arsenic, mercury and antibiotic resistance genes. Arsenite efflux was the most common arsenic detoxification method, and mer operon providing mercury resistance was detected in many strains. A number of genes highly similar (but not identical) to the known antibiotic resistance genes were found, suggesting potential danger of horizontal gene transfer into clinically relevant bacteria.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN BIOLOGÍA MOLECULAR Y CELULAR----------

Tutor

María Rosario Rodicio Rodicio

Supervisors

Ana Isabel Peláez Andrés

José Luis Rodríguez Gallego

Contents

1. Introduction **1**
- 1.1. Arsenic . . . 1
- 1.2. Mercury . . . 5
- 1.3. Antibiotic resistance and its co-occurrence with metal(loid) resistance . . . 7

2. El Terronal mining site **9**
- 2.1. Geography and climatic conditions . . . 9
- 2.2. History of the mining site . . . 10
- 2.3. Geology and hydrology of the site . . . 13
- 2.4. Industrial process . . . 15
- 2.5. Characterisation of the site . . . 18
  - 2.5.1. Physico-chemical and mineralogical caracterisation of the site . . . 18
  - 2.5.2. Soil and groundwater contamination . . . 20
- 2.6. Previous remediation work on the site . . . 22

3. Aims and objectives **25**

4. Materials and methods **29**
- 4.1. Sample collection and chemical characterisation . . . 29
- 4.2. DNA extraction from environmental samples, construction and Illumina sequencing of rRNA gene libraries . . . 31
  - 4.2.1. Nucleic acid extraction from environmental samples . . . 31
  - 4.2.2. Construction of 16S rRNA and 18S rRNA gene libraries . . . 31
  - 4.2.3. MiSeq Illumina sequencing . . . 32
- 4.3. Isolation and characterisation of bacteria . . . 33
  - 4.3.1. Isolation and cultivation of bacteria . . . 33
  - 4.3.2. Characterisation of arsenic and mercury resistance of isolated strains . . . 33
  - 4.3.3. Antimicrobial susceptibility testing . . . 34
  - 4.3.4. Hydrocarbon degradation and emulsification . . . 35
  - 4.3.5. Bacterial gDNA extraction . . . 37
  - 4.3.6. Amplification of 16S rRNA genes for identification of isolated strains . . . 37
  - 4.3.7. Screening for arsenic and mercury resistance genes . . . 39
  - 4.3.8. Whole-genome sequencing of isolated strains . . . 39
- 4.4. Data processing . . . 39
  - 4.4.1. Processing of 16S and 18S Illumina-generated data . . . 39

4.4.2. Genome assembly and annotation . . . 42
4.4.3. Genome visualisation and comparisons . . . 43

5. Results **45**
5.1. Chemical analysis of the samples . . . 45
5.2. 16S and 18S rRNA gene metagenomic studies . . . 47
5.2.1. Coverage and general statistics . . . 47
5.2.2. Alpha-diversity and sequencing depth . . . 48
5.2.3. Beta-diversity . . . 55
5.2.4. Taxonomic analysis . . . 57
5.2.5. Functional metagenome prediction . . . 64
5.3. Seeding and isolation of bacteria . . . 69
5.4. Identification of isolated strains . . . 70
5.5. Arsenic resistance of isolated strains . . . 79
5.6. Mercury resistance of isolated strains . . . 83
5.7. Antibiotic resistance and production of antimicrobial compounds . . . 84
5.8. Degradation of organic compounds . . . 86
5.9. Emulsification of organic compounds . . . 89
5.10. Selection of strains for genome sequencing . . . 91
5.11. Genome analysis . . . 93
5.11.1. Genome assemblies . . . 93
5.11.2. Synteny . . . 95
5.11.3. Arsenic and mercury resistance genes . . . 101
5.11.4. Detection and analysis of antibiotic resistance genes . . . 115

6. Discussion **119**

7. Conclusions **127**

A. Annex: Materials and Methods **137**

B. Annex: Results **139**

Bibliography **155**

1. Introduction

1.1. Arsenic

Chemistry and occurrence

Arsenic is a naturally occurring metalloid from the group 15 (nitrogen group) of the Periodic Table. Its atomic number is 33; it is a monoisotopic element. In its pure form, it can exist as various allotropes, of which grey arsenic is the most common. In this form, it is a brittle, crystalline solid of silvery-grey colour with a slight green tint. Its atomic weight is 74.92; its density is 5.727 g/cm^3. Its sublimation point resides at 887K (615℃).

In nature, it exists in one of the four oxidation states: pentavalent arsenate, trivalent arsenite, elemental arsenic As^0, and arsines (such as arsine H_3As). Elemental arsenic is fairly insoluble, while solubility of other chemical forms of arsenic varies depending on pH and overall chemistry of the environment, with As(V) being the most stable in oxidising environments, while As(III) compounds can be very stable under reducing conditions.

It is the 20th most abundant element in the Earth's crust (0.00021% by weight). Despite this relative scarcity, it is a major constituent of no less than 245 mineral species. The most common of them is arsenopyrite (FeAsS); however, appreciable amounts of arsenic concentrate in other sulphide minerals such as pyrite (FeS), galena (PbS), covellite (CuS) and acanthite (Ag_2S) , as well as in clays and coals. Most of arsenic, however, is dispersed in the environment. In the clean open-ocean water, concentration ranges from 0.5 to 3 µg/L, with mean of about 1.7 µg/L (Neff, 2002). In the soil and groundwater, arsenic concentration vary depending on the mineral composition of the underlying bedrock, presence of sulphide ore deposits in the area, and degree of soil erosion. As an example, arsenic concentrations in European soils follow a dichotomy that corresponds almost exactly to the maximum extent of the last glaciation, with less eroded soils in the north of the continent having 2 to 4 times less arsenic (Tarvainen et al., 2013).

Toxicity and detoxification

Toxicity to arsenic is universal, affecting all living beings from bacteria to plants to humans. Inorganic forms of arsenic are considered to be more toxic than organic forms[1], with arsenites being more toxic than As(V) compounds. Arsenic toxicity arises from

[1] Arsine H_3As is by far the most dangerous to humans due to its gaseous form and induction of haemolysis; however, it is chemically unstable, and thus extremely rare in nature.

capacity of arsenite to bind to sulfhydryl groups of sulfhydryl-containing enzymes, disrupting biochemical machinery of the cell (primarily of the pyruvate oxidation pathway and tricarboxylic acid cycle). Another toxicity mechanism involves arsenate similarity to phosphorus, which it can replace in many enzymatic reactions, the most prominent being cell respiration where substitution by (less stable) arsenate leads to rapid dissolution of ATP (Rossman, 2007). Due to being able to bind to the DNA , arsenic has a pronounced carcinogenic effect as well.

No credible[2] reports of organisms utilising arsenic for any of the vital processes such as DNA replication or protein synthesis have been published to date. On the other hand, energy use of arsenic such as arsenite oxidation and dissimilar (respiratory) arsenate reduction is well-established (Oremland & Stolz, 2000; Mukhopadhyay et al., 2002). It is thought that physiological activities of microorganisms inhabiting anoxic environments of the ancient Earth, where As(III) was, most likely, the predominant form of arsenic, revolved around arsenic metabolism (Kulp, 2014; Sforna et al., 2014). Arsenite removal from the cell was necessary for survival, while arsenite oxidation served as an energy source and lead to production of arsenate which, in turn, enabled evolution of the arsenate-breathing bacteria. While the advent of photosynthesis and massive shift towards oxygen-rich oxidising environments decreased the importance of arsenic metabolism, it still exists and contributes to geochemical cycling of arsenic on the global scale (Huang, 2014). It has been postulated that most prokaryotes contain arsenic resistance genes and participate in biogeochemical cycling of arsenic; the relative contribution of each arsenic detoxification and transformation system, however, remains poorly understood (Zhu et al., 2014).

Major detoxification and metabolic mechanisms employed by prokaryotic organisms include As(III) expulsion, As(III) oxidation, As(V) reduction, arsenic methylation and demethylation, degradation of organoarsenic compounds and their expulsion from the cell (Crognale et al., 2017). Main prokaryotic genes responsible for those processes are summarised in the Table 1.

The most common detoxification pathway employs arsenate reductase ArsC converting As(V) into As(III) with its subsequent expulsion by ArsAB ATP-driven arsenite pump or (the more common) ACR3 arsenite pump. While As(III) is more mobile and more toxic, As(V) is significantly more abundant under current environmental conditions, and can easily enter cells via phosphate uptake systems, necessitating its removal (Zhu et al., 2017). Another detoxification pathway common in prokaryotes involves converting arsenic into gaseous trimethylarsine by repeated methylation (P. Wang et al., 2014). Some bacteria were capable of weaponising this process in order to gain competitive advantage in the manner similar to production of antibiotics, in turn prompting a defensive response by expressing organoarsenical oxidase gene *arsH* (J. Chen & Rosen, 2020). Graphical summary of various metabolic and detoxification pathways is given in the Figure 1.

[2]Basturea et al., 2012; Reaves et al., 2012.

Table 1.: Most common prokaryotic genes involved in arsenic metabolism and detoxification. Adapted from Dunivin et al. (2019) and Zhu et al. (2017).

Gene	Encoded protein	Function
arsB	As(III) pump	Expulsion of As(III) from the cell
arsA	As(III)-pump ATPase	Catalytic subunit of oxyanion-translocating ATPase
arsD	Arsenical metallochaperone	Transfer of arsenite ions to ArsA
arsC	Cytoplasmic As(V) reductase	Reduction of arsenate
arsP	Trivalent organoarsenicals efflux system	Expulsion of trivalent organoarsenicals from the cell
arsT	Putative thioredoxin reductase	Unknown
arsH	Organoarsenical oxidase	Oxidation of trivalent methylated and aromatic arsenicals
arrA	As(V) respiratory reductase (large subunit)	Reduction of arsenate
arrB	As(V) respiratory reductase (small subunit)	
aioA	As(III) oxidase (large subunit)	Oxidation of arsenite
aioB	As(III) oxidase (small subunit)	
arxA	As(III) oxidase	
acr3	As(III) permease	Expulsion of As(III) from the cell
arsM	As(III) S-adenosylmethionine methyltransferases	Methylation of arsenic
arsR	Arsenic-responsive repressor	Regulation of *ars* operon expression
arsJ	Organoarsenical efflux permease	As(V) expulsion using binding to organic molecules

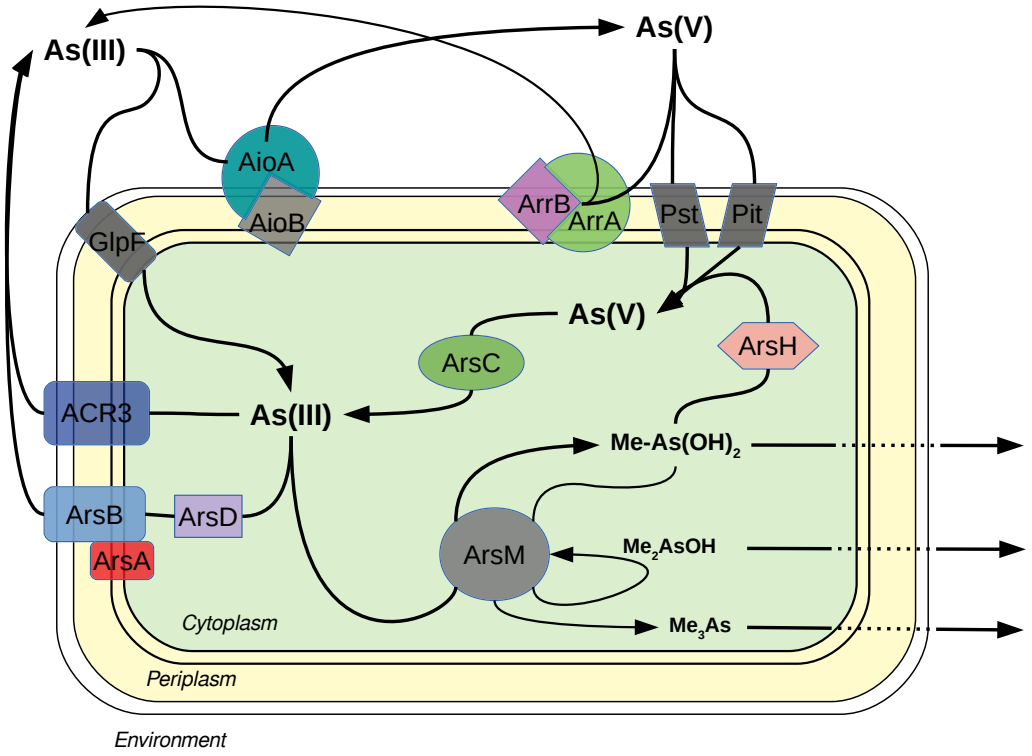


Figure 1.: Schematic of major arsenic metabolism and detoxification mechanisms. Based on Páez-Espino et al. (2009), with additional information from Zhu et al. (2017) and Crognale et al. (2017).

Arsenic sources and biogeochemical cycle

Geogenic sources like geothermal springs, volcanic eruptions and erosion of arsenic-bearing minerals all play an important role in arsenic release into the environment, while arsenic sinks are usually associated with aquatic biogenic activity leading to accretion or deposition of arsenic in marine and freshwater sediments, as well as purely geological processes of ore deposit formation (Masuda, 2018). Human activities over the last two centuries, however, managed to significantly increase the amount of arsenic released annually (Han et al., 2003). Most significant anthropogenic sources include metal ore mining and smelting and burning of fossil fuels (especially coal), followed by its use in wood preservation (copper-chromium-arsenic-impregnated wood), glass manufacturing, use as pesticide or insecticide, leaching from landfills, as well as emissions from chemical industry and semiconductor manufacturing (Zevenhoven et al., 2007; W.-Q. Chen et al., 2016). Schematic view of arsenic biogeochemical cycle can be found in the Figure 2.

Arsenic contamination effects are fairly local, but in cases of arsenic contamination of groundwater resulting from arsenic-rich mineral composition of the area they can affect wide swathes of land as is the case on the Indian peninsula (Shahid et al., 2018; Hasan et al., 2019; Shaji et al., 2020).

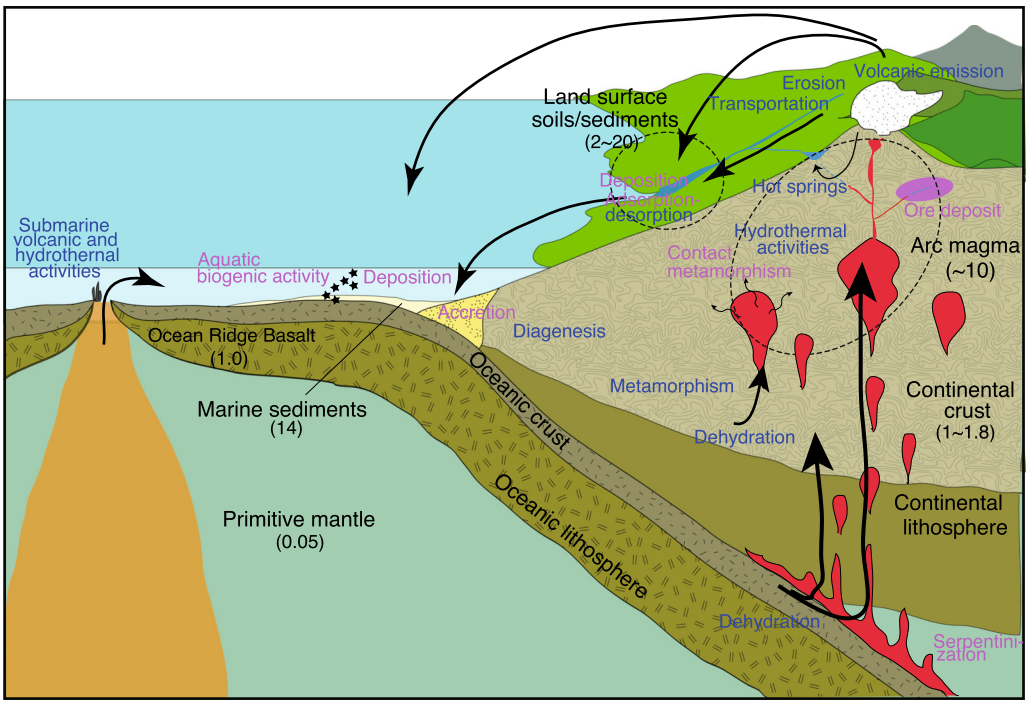


Figure 2.: Natural biogeochemical cycle of arsenic. Digits in parenthesis give average concentrations of arsenic in each geologic body. Purple-coloured letters indicate arsenic sinks; blue-coloured letters denote arsenic release. Adapted from Masuda (2018).

1.2. Mercury

Chemistry and occurrence

Mercury is a naturally occurring transition metal from the group 15 (zinc group) of the periodic table. Its atomic number is 80; it is found in nature as a mixture of seven stable isotopes. A silvery white, heavy, easily evaporating liquid metal, it is the only metal to be in the liquid state under standard conditions[3]. Its melting point resides at 234.31K (- 38.9℃), its boiling point - at 629.88K (356.73℃). Its density is 13.534 g/cm^3. Mercury occurs in elemental form (Hg^0), as well as in the form of various Hg(I) and Hg(II) compounds. The most common mercury mineral is cinnabar (HgS).

Toxicity and detoxification

Mercury is the most toxic of heavy metals due to its high affinity for sulfhydril ligands in amino acids; upon binding, it leads to the loss of protein function. However, in terms of toxicology, an even greater danger comes from methylated mercury due to its higher toxicity and very high bioaccumulation rate.

It is thought that mercury detoxification system appeared long before separation between *Bacteria* and *Archaea*, arising in primitive cells clinging to geothermal vents on

[3]By the most strict definition, standard conditions for temperature and pressure refer to the temperature of 273.15K (0℃) and absolute pressure of 100 kPa (1 bar, ~1 atm).

the ocean floor . Thus, it can be considered one of the most ancient detoxification systems on the planet. Over billions of years of evolution, it changed and adapted, gradually expanding its scope and increasing efficiency (Boyd & Barkay, 2012).

In its most basic form, *mer* operon requires mercury(II) reductase *merA* that is responsible for converting Hg^{2+} ions into volatile Hg^{0}, and regulatory gene *merR*. Alkylmercury lyase is encoded by *merB* gene; it evolved to deal with organomercuric compounds by cleaving the mercury-carbon bond, releasing Hg^{2+} ion that can be reduced by mercury reductase. Genes *merF*, *merH*, *merTP*, *merC* and *merE* encode Hg^{2+} carrier proteins; they all belong to mercury transporter superfamily and catalyse Hg^{2+} uptake for its subsequent reduction to elemental mercury.

Mercury sources and biogeochemical cycle

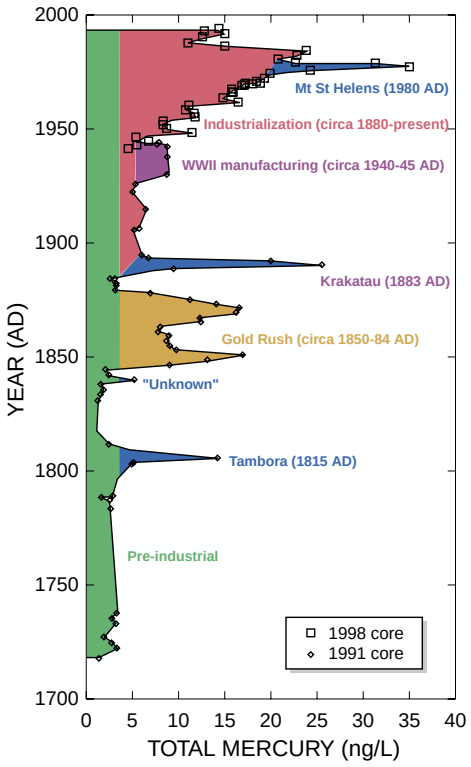


Figure 3.: Atmospheric mercury deposition rates over the last three centuries. Pre-industrial deposition rate (4 ng/L, in green) conservatively extrapolated to present time. Adapted from Krabbenhoft and Schuster (2002).

Unlike arsenic, where negative effects are mostly local, and biogeochemical cycle functions on geological timescales, high mobility of mercury (in elemental form as well as when moving up the food chain as methylmercury) creates a truly global biogeochemical cycle regardless of the place of emission. (Obrist et al., 2018).

While significant release of mercury usually occurs during volcanic eruptions, human use of mercury in metallurgy, chemical industry, electronics, daylight fluorescent lamps and medical applications more than tripled yearly mercury emissions since the beginning of the industrialisation (Schuster et al., 2002). However, a series of massive methylmercury poisoning incidents led to phasing out of mercury and mercury-based products in the industry, culminating in the eventual signing of Minamata Convention on Mercury in 2013 that codified the phase-out as international law. Currently, the majority of the mercury emissions come from coal-powered power plants and artisanal gold mining. Released into the environment in elemental form, mercury is then redistributed by atmospheric circulation.

Eventually, it is oxidised into Hg(II) and deposited into the surface water, although the process can be very slow. Once in the water, mercury gets methylated by waterborne

bacteria and archaea, and starts moving up the marine food chain. Due to accumulating in fish, it can cause health issues in populations subsisting on the most affected fish species (Višnjevec et al., 2014; Jewett & Duffy, 2007). A schematic view of the mercury cycle can be found in Figure 4.

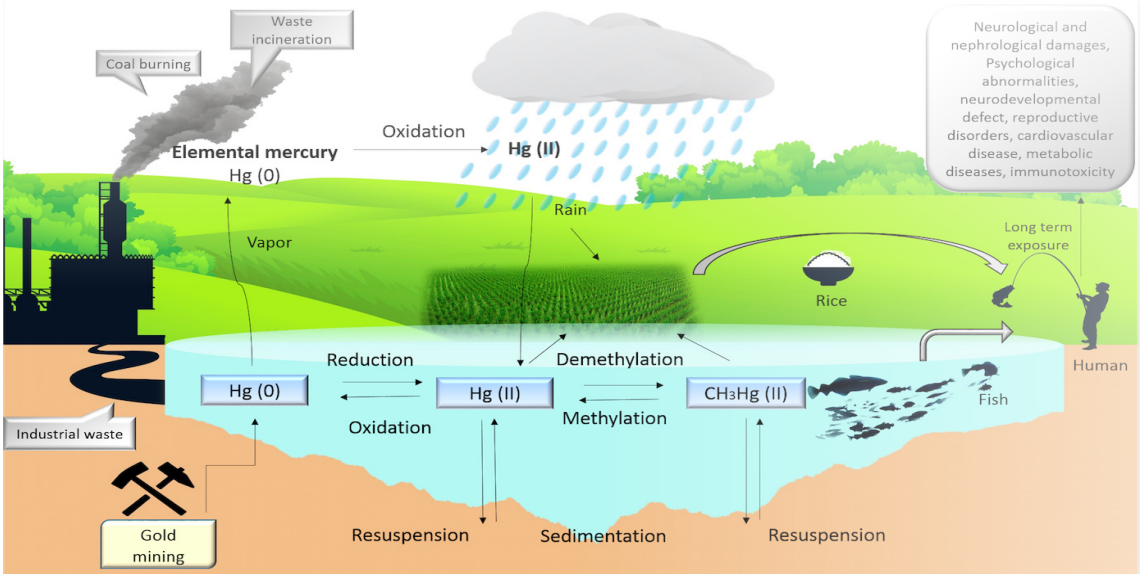


Figure 4.: Biogeochemical cycle of mercury. Adapted from Nurfatini and Syahir (2018).

1.3. Antibiotic resistance and its co-occurrence with metal(loid) resistance

Production of antimicrobial agents is widespread in nature, with majority of known antibacterial compounds coming from *Streptomyces* bacteria and fungi. Their ecological and biological roles were mostly disregarded, with research being focused firmly on their clinical applications. However, with virtually unrestricted use of antibiotics in healthcare, farming and everyday life, accelerated acquisition of antibiotic resistance genes (ARGs) by both environmental and pathogenic bacteria started to pose an immediate danger for human health (Ventola, 2015). Understanding and predicting emergence of antibiotic resistance mechanisms thus became a priority, leading to better understanding of their and occurrence in nature and function (Sengupta et al., 2013). 'Natural' roles of antibiotics in sub-inhibitory concentrations can vary from suppressing other bacteria by antibiotic-producing strain in order to gain competitive advantage, to various forms of signalling, including transcription regulation and quorum sensing, with antibiotic resistance evolving to counteract it (Aminov, 2009; Sengupta et al., 2013).

Co-occurrence of antibiotic and metal(loid) resistance

Emplacements subjected to heavy metal and metalloid contamination have a potential to influence antibiotic resistance patterns in environmental bacteria. This connection is multi-faceted and often indirect, with co-occurrence between antibiotic and metal(loid) resistance generally falling into one of four categories (Baker-Austin et al., 2006; Imran et al., 2019):

- Co-selection of resistance to metall(oids) and antibiotics due to location of their respective genes in the same mobile genetic element;
- Cross-resistance due to multi-drug resistance pumps not discriminating between metal and metalloid ions and antimicrobial molecules;
- Co-regulation, when generic anti-stress responses such as removal of reactive oxygen forms and cell repair triggered by the metall(oids) increases resistance to antimicrobial agents;
- Biofilm formation leading to trapping of harmful substances in the biofilm matrix and mutual help between biofilm-forming bacteria;

In practical (clinical) terms, the first two mechanisms are the most important when considering acquisition of new antibiotic resistance genes by pathogenic bacteria, and deserve to be given some consideration among other efforts to slow down the spread of antibiotic resistance.

2. El Terronal mining site

2.1. Geography and climatic conditions

El Terronal is a mercury mining and metallurgy site in the Principality of Asturias, Northern Spain. It is located 2 km north-east of the town of Mieres, and about 20 km south-east of the provincial capital of Oviedo. It sits on the northern bank of the San Tirso river, 1.5 km away from its confluence with the Caudal river (Figure 5). The area is sparsely populated: as of 2019, only 37.959 people lived in the municipality of Mieres, with majority of them inhabiting the town of Mieres itself (INE, 2019).

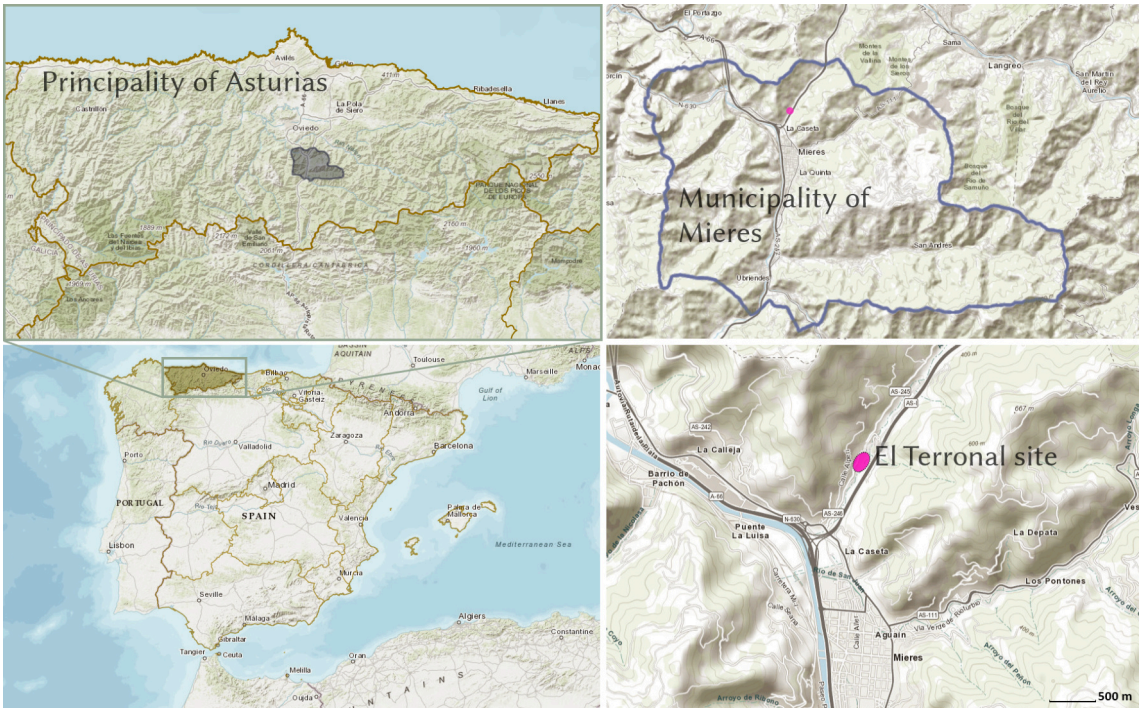


Figure 5.: Geographical location of the study site. Retrieved from ArcGIS web server using World Topo Map and Municipal Border layers. Map sources: INE, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, Intermap, USGS, METI/NASA.

Unlike many other regions of Spain, Asturias is characterised by humid, oceanic climate with mild temperatures (yearly average of 12.5 ℃) and low seasonal temperature variations. Local temperature variation is mostly altitude-dependent, standing at about

0.5 ℃ decrease in average temperature for every 100 metres of altitude. Precipitation frequently comes in the form of heavy rain, with average yearly rainfall reaching 1000-1200 mm. With potential evaporation estimated as 600 mm, average effective annual precipitation stands at around 400 mm (Méndez Pazos, 2013). Average humidity is about 80%.

2.2. History of the mining site

Mercury extraction at the El Terronal site had a long and turbulent history. Small-scale mercury extraction was carried out in the Roman times, limiting itself to shallow, easily reachable deposits . Prospecting of cinnabar deposits started in the beginning of the XIX century, and large-scale industrial exploitation of the site begun in 1844, when mining rights to La Esperanza mine and El Terronal site were conceded to *Sociedad Minera La Fraternidad*, which was acquired by the *El Porvenir* company two years later. This acquisition resulted in a large investment into mining works and creation of industrial facilities, with four wood-fired Idria-type furnaces installed at the El Terronal site by 1850, and two more - at the nearby La Peña mercury mine (Gutiérrez, 2001).

The need to increase efficiency resulted in a range of domestically produced innovations. Starting from 1852, furnaces were converted to burning coal as fuel, greatly improving fuel economy. In 1880s, additional furnaces were added: one multi-channel Livermore furnace modified for roasting cinnabar, three locally designed muffle-type Rodríguez furnaces, as well as one conical Gascue-Rodríguez furnace for roasting lump ore (Figure 6).

Figure 6.: El Terronal at the turn of the century. Left: View at the industrial installations in the year 1893. Right: a closer view of the furnaces, 1910. . It is of note that furnaces were located on the Southern bank of the San Tirso river. In the modern configuration of the site this space was used exclusively for storing waste (see 2.4). Photographs retrieved from El Blog de ‘Acebedo.

Locally designed furnaces turned out to be both more efficient and better equipped to deal with arsenic-rich ores than Idria-type furnaces, showing lower loss of mercury

vapour and higher fuel efficiency. An original process of extracting mercury from arsenic-rich pyrometallurgic waste was also developed. This waste could contain up to 14% of Hg^0, as well as small amounts of mercury (II) sulphate. In order to extract this residual mercury, the waste was first mixed with lime and washed by copious amount of water, and then, after adding coke and clay serving as binding agents, formed into briquettes and roasted in the furnaces similarly to the lumps of cinnabar (Rothwell, 1895). In the 1890s, coke gradually replaced coal as main furnace fuel, further increasing efficiency and reducing the amount of waste produced in all furnaces (Gutiérrez, 2001).

Eventually, increasing competition with overseas mercury producers, limited capacity of the metallurgic plant, and volatility of the mercury market resulted in low profitability of the El Terronal plant. Installations entered a series of ownership changes: in 1907 all mining rights were sold to the Anglo-Spanish *The Porvenir Mercury Mines Ltd.*, with shareholders of the old *El Porvenir* company obtaining the majority of shares in it. The company very soon rebranded itself as *The Oviedo Mercury Mines Ltd.*, and proceeded with acquisition of other mercury and coal mines in the region and elsewhere in Spain. However, neither consolidation of mercury mining works, nor injection of capital were able to return the plant to profitability. In 1910 all of its assets and mining rights, including El Terronal site, were put on sale, and in 1912 they were bought by *Sociedad Herrero y Compañia* that continued mercury mining and extraction till at least 1920 (Gutiérrez, 2001). History of the site is not clear in the decade that follows, but if any work had been done, it was put to a halt in 1934 due to the revolution and subsequent Spanish Civil War.

Activities in El Terronal were not resumed till 1947 when the site was bought by *Astur-Belga de Minas SA.* New galleries connecting Esperanza mine at El Terronal with nearby La Peña and La Unión mines were excavated, giving access to more cinnabar deposits. Mercury distillation plant was completely rebuilt with modern technology, including two oil-fired rotary furnaces, one multi-hearth Herreshoff kiln, and a modern muffle furnace, mechanised equipment and two high-capacity mercury condensers and waste treatment plants (Figure 7), resulting in a period of previously-unseen productivity and profitability (Méndez Pazos, 2013). El Terronal thus become the second-largest mercury producing site in Spain and the 8th largest in the world, with annual production of up to 15000 flasks[1]. Besides industrial installations, two large buildings for housing workers and their families were also erected just outside the premises.

Ten years later, a sharp drop in demand due to environmental concerns and overabundant supply made El Terronal unprofitable again. In 1973, all activities were halted, and the site was mothballed and eventually abandoned. Attempts to revive mining activity in the area during the decades that followed were not successful; the site was used for irregular disposal of waste coming from other industries (mainly from

[1] *Flask* is a common unit for measuring and pricing mercury. Descending from vessels that stored three Castilian *arroba*s of mercury, with the adoption of the metric system the flask was standardised at 34.5 kilograms of mercury, which corresponded to the volume of 2.549 litres. It was made from cast iron, with thick walls and a 3/4-inch threaded iron plug (Myers, 1951).

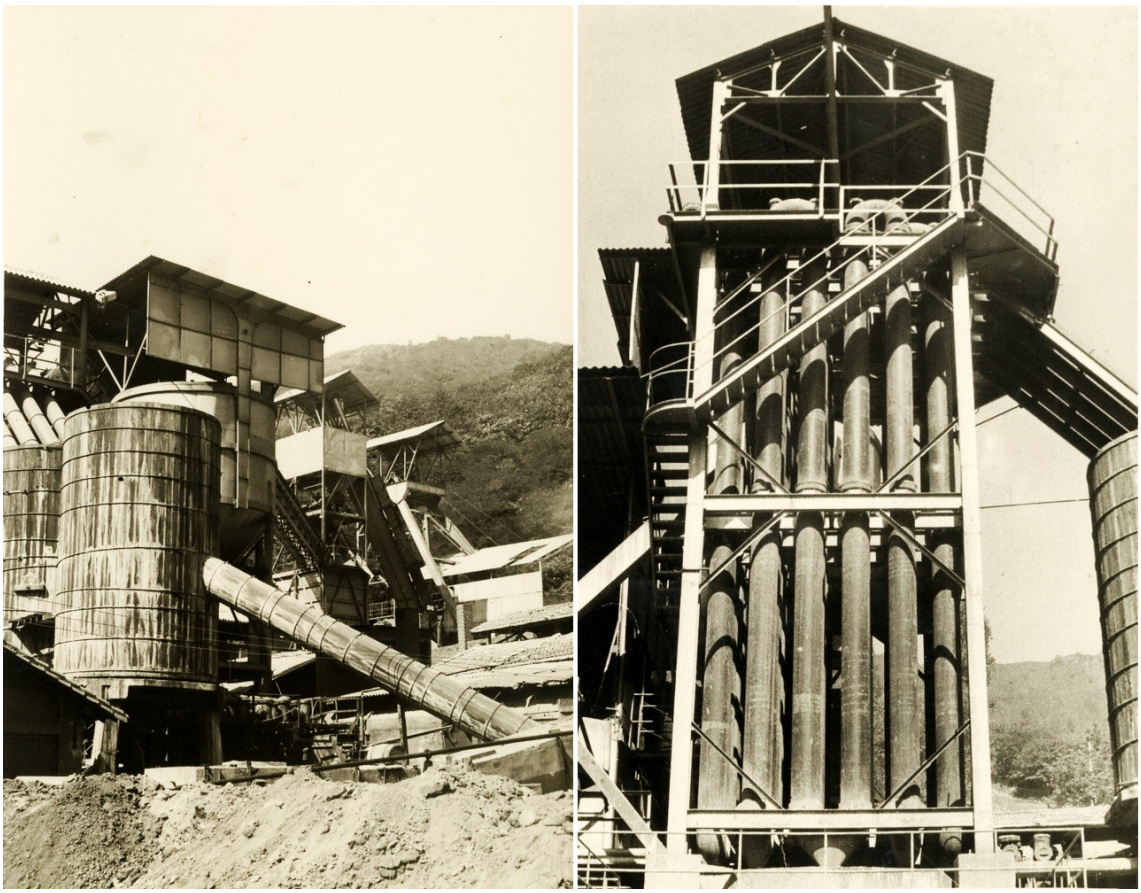


Figure 7.: El Terronal in the 1960-1970s. Left: General view of the installations. Right: Close-up view of one of the mercury condensers. Retrieved from MTI Minas Asturias (2013).

coal mining and machinery workshops of Mieres). In 1999, concession to *Astur-Belga de Minas* expired, leaving unclear the ownership of the site (Cabal & Claverol, 2006).

Over time, installations fell into disrepair. Presence of large, highly toxic spoil heaps open to weathering and leaching resulted in heavy arsenic and mercury pollution at and around the site, as well as heavy contamination of the river water and groundwater (See subsection 2.5.2). In 2002, construction of the AS-I motorway begun in the area, some years later resulting in the largest spoil heap being enclosed in a concrete protective cover under the motorway. While designed, at least in part, to limit the mobilisation and release of arsenic into the environment, those measures were shown to be insufficient (Ordóñez et al., 2013), most likely due to the continued presence of smaller waste heaps open to the elements. Those smaller heaps were finally removed in 2019.

2.3. Geology and hydrology of the site

The study site is located in the north-western part of the Asturias Carboniferous Central Basin, one of the principal subdivisions of the Cantabrian Zone of the Iberian Massif (Lotze, 1945). The area is comprised mainly of siliceous breccias, sandstone, lutites, limestone and coal (Westphalian age). It has approximately NW-SE direction, anticlinal structure, and is limited by La Carrera and La Peña faults from South-East and North-West, respectively (Figure 8).

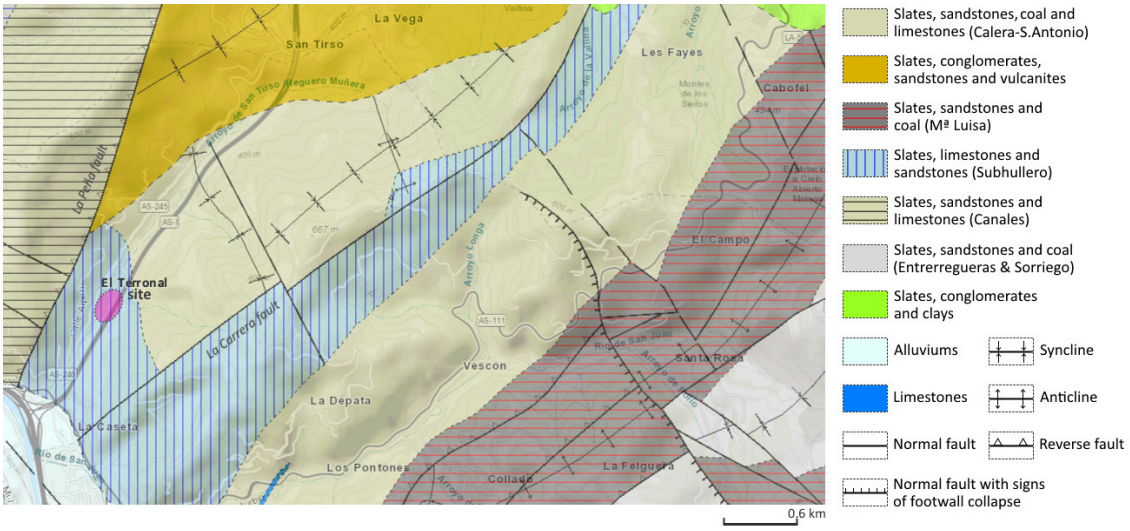


Figure 8.: Geological map of the site. Retrieved from IGME ArcGIS web server using IGME MAGNA 50 and World Topo Map layers. Map sources: INE, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, Intermap, USGS, METI/NASA.

Mercury and arsenic ores are found in mineralisations associated mainly with breccias and, to a lesser degree, with sandstones. Mercury occurs mostly as cinnabar (HgS) and metacinnabar, although elemental mercury and guadalcazarite (zinc-containing metacinnabar) can also be found. Arsenic is present mainly as realgar (AsS) and arsenic-rich pyrites (Cabal & Claverol, 2010). Other sulphide minerals in mineralisations include iron sulphides pyrite, marcasite and melnikovite, sphalerite (an iron-containing zinc sulphide), and sulphides of antimony and lead in the form of stibnite and galena, respectively (Cabal et al., 1991). Such mineralisations are epigenetic in origin, occur largely at and around the fault lines, and were formed by circulation of low-temperature hydrothermal solutions through the fractures that correspond to the continental rifting during the Permian period (García, 1981).

Hydrology of the site is complex, with both surface and groundwater flows significantly modified by industrial activities. On the surface, the study site is located in the San Tirso river drainage basin that covers about 7 km^2. Average slope is 21.57°, with total height difference of 476.53 metres. The river stretches for 4.5 km, with average slope of the channel of 9.17°.

Under heavy rains, the river becomes a torrential channel (Méndez Pazos, 2013). Man-made structures such as derelict chimneys and pipes around industrial installations created preferential channels for rainwater, forcing it to pass through abandoned installations, picking up contaminants on the way (Figure 9).

Figure 9.: Rainwater flow through the abandoned industrial installations and the brownfield, February of 2015.

Before extensive mining works, subsurface hydrology was mainly defined by underlying geological structure, with breccias and sandstones serving as aquifers, and clay lutites and coal beds as aquicludes and aquitards, respectively. Groundwater flow was mostly directed to the San Tirso river. However, excavation of multiple mining and ventilation shafts, as well as mining galleries, including two galleries connecting El Terronal with the lower-altitude La Peña shaft, had a significant impact on permeability, infiltration and groundwater flows. Currently, significant amount of groundwater is redirected towards La Peña drift, where it discharges into sewage systems and river Caudal (González-Fernández et al., 2018). Both surface and groundwater flows, as well as geological structure of the area are represented in the Figure 10.

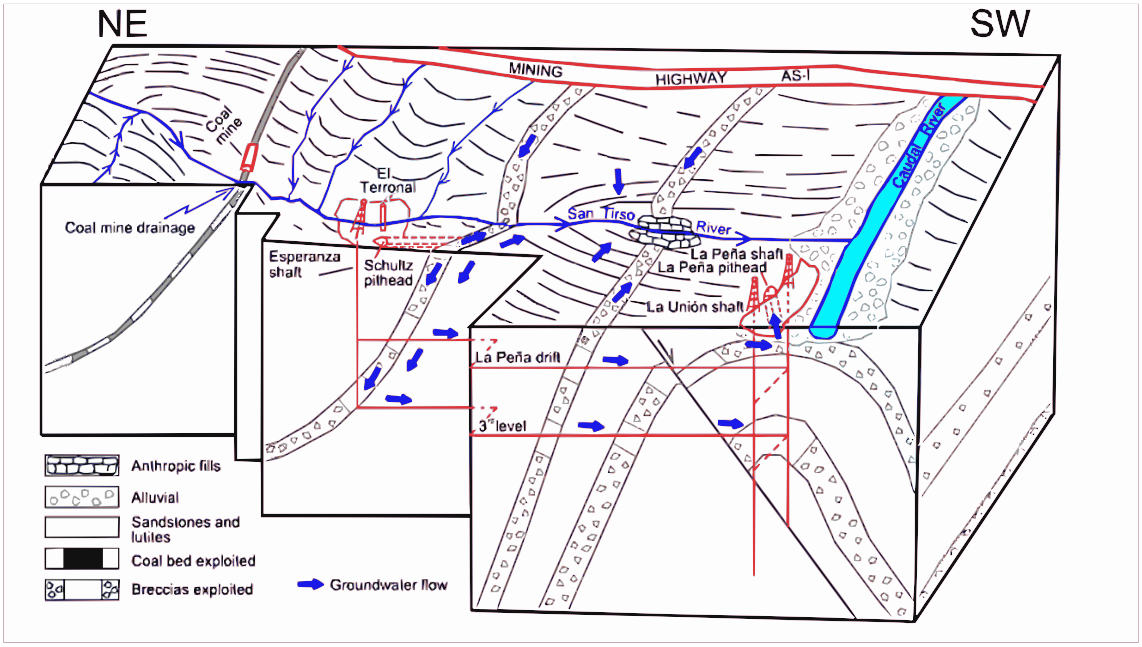


Figure 10.: Geological structures and hydrogeological system of the San Tirso river valley (unscaled schematic vertical section). Adapted from González-Fernández et al. 2018.

2.4. Industrial process

Since the beginning of the industrial-scale mercury mining El Terronal used a pyrometallurgic process to extract mercury from the ore. The main difference was that due to the high arsenic content of the ore, arsenic had to be condensed before mercury. Specialised arsenic condensers were used, helped by the fact that arsenic trioxide vapours begin to condense at higher temperature than mercury (Duschak & Schuette, 1925).

Cinnabar ore was mined in the Esperanza and La Peña mines, and transported to the Esperanza mine headframe through the connecting tunnels. Once there, ore was brought to the surface, sorted according to quality, grounded into smaller fragments and fed into the furnaces, where it was mixed with fuel oil and burned. Two semi-independent production lines were active:

1. First rotary furnace was connected to an arsenic condenser. From the documents, it is not clear whether the muffle furnace was connected to it as well; likely, it had its own arsenic condenser that was located in the same building. Mercury vapours were then passed through a water-cooled condenser, with liquid mercury condensing and being stored in four wooden tanks;
2. Herreshoff kiln and second rotary furnace had a larger baffle-type arsenic condenser and a water-cooled mercury condenser; mercury was condensed and stored in two large wooden tanks.

Remaining gases were vented through a common smokestack. A schematic view of the installations can be found on Figure 11.

Figure 11.: Schematic view of the former location of known industrial installations and auxiliary buildings, overlaid with a recent satellite image. (1) Esperanza mine headframe. (2) Draw-works machinery. (3) Electric facilities. (4) Flue duct. (5) Offices and laboratories. (6) Dining room and lockers room. (7) Ore grinding and sorting station. (8) Furnace fuel storage tanks. (9) Herreshoff kiln. (10) Rotary furnaces. (11) Muffle-type furnace. (12) Arsenic condensers. (13) Mercury condensers. (14) Waste treatment areas. (15) Mercury and mercury soot condensation tanks. (16) Workshop and carpentry area. (17) Storage buildings. (18) Bridges across San Tirso river. (19) Large spoil heap, currently buried under the AS-I motorway. Satellite imagery obtained via ArcGIS web server using World Imagery map layer and upscaled using ISR (Cardinale et al., 2018). Map sources: Esri Community Maps Contributors, Dirección General de Catastro, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, METI/NASA, USGS.

With the end of all activities, several types of waste remained inside industrial installations. As buildings on the site fell further and further into disrepair, those wastes were exposed to the elements and eventually formed several distinct areas:

- Derelict arsenic condensers contain arsenic-rich soot formed by condensation of arsenic oxides and other impurities. Two large and one small waste heaps are present;

- Two areas where stupp (residue accumulating in the mercury condensers) was spoil-heaped. During the normal operation, it was either disposed of, or, after chemical treatment not dissimilar to the process described in 2.2 sth sel'ida 11, was roasted in the furnaces again. However, due to the closure of the plant and dismantlement of the mercury condensers, accumulated stupp was left in place;
- Structures belonging to the smokestack contain large amounts of flue dust;
- Large brownfield that served as a main waste disposal area during the functioning of the plant; it housed both gangue and metallurgic waste. It is currently partially encapsulated under the AS-I motorway.

In addition to the brownfield housing wastes produced by the ore smelting, a small area around the remains of the fuel storage tank is contaminated with fuel oil, and allochtonous waste such as metal chips from machining workshops can be found in two large areas of the site. Location of waste disposed of within the boundaries of the site can be seen on Figure 12.

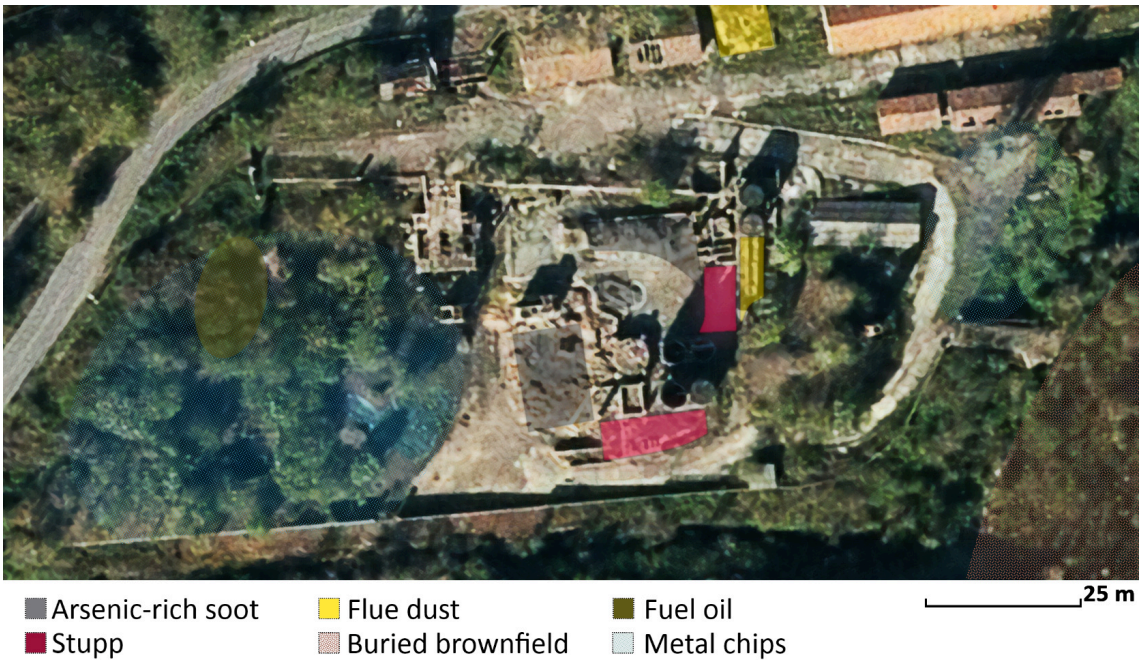


Figure 12.: Areas of interest overlaid with a satellite photograph. Satellite imagery obtained via ArcGIS web server using World Imagery map layer and upscaled using ISR (Cardinale et al., 2018). Map sources: Esri Community Maps Contributors, Dirección General de Catastro, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, METI/NASA, USGS.

2.5. Characterisation of the site

2.5.1. Physico-chemical and mineralogical caracterisation of the site

Extensive chemical and mineralogical characterisation of the site was performed previously by Gallego et al. (2015). In their work, samples from the brownfield were characterised by acid digestion and ICP-MS , as well as X-ray diffraction and microscopy. Additionally, bioavailability and mobility of the contaminants were assessed by the Tessier approach[2], and organic contaminants were studied by liquid and gas chromatography and mass spectrometry.

According to their findings, all pyrometallurgic waste fell into four distinct categories: calcines, arsenic-rich soot, stupp and flue dust.

Calcines Calcines are calcination residues produced in the furnaces. Periodically stored at various locations around the site, most of them were eventually transported to the main waste disposal area, but small accumulations can still be found elsewhere. They are comprised mainly of quartz and Al-Si oxides. Concentrations of arsenic and mercury were high in absolute terms (in the range of 1000 and 50 mg/kg), but low relative to other types of waste. Arsenic was occurring as As(V), and mercury was occurring as inorganic species. Their bioavailability was low; however, significant mobile fractions of arsenic (35.3%) and mercury (49.3%) were found by Tessier approach.

Arsenic-rich soot Produced by arsenic oxide condensation in the arsenic condensers, it represented most of the waste by volume, forming large spoil heaps. Arsenolite (As_2O_3) was found to be the most abundant mineral, followed by claudetite (a monoclinic dimorph of As_2O_3). Scorodite ($FeAsO_4 \cdot 2H_2O$), quartz and kaolinite were also present. Arsenic concentration was correspondingly high, exceeding 65%. Concentration of mercury was above 1%; it was represented by inorganic mercury species. Bioavailable fraction represented 14.1% of total arsenic, and mobile fraction represented 48.6%, as found by Tessier approach. Due to having such a large concentration of arsenic, significant leaching also occurred in alkaline (1M NaOH) solutions. Bioavailability of mercury was rather low at just 0.9%, and mobile fraction represented 8.5% of all mercury. However, when taking exceedingly high mercury concentration into account, it is clear that release of mercury into environment in absolute terms would still be extremely high.

Stupp Consisting mainly of quartz and arsenolite, it also contained up to 3% mercury, observable as Hg^0 droplets, as well as illite and kaolinite. Arsenic concentration could

[2]Tessier, A., Campbell, P. G. C. & Bisson, M. (1979). Sequential extraction procedure for the speciation of particulate trace metals. *Analytical Chemistry*, *51*(7), 844–851. https://doi.org/10/dxfk6n.

exceed 55%. Bioavailability and mobility of arsenic and mercury, as determined by the Tessier approach, was slightly higher than for arsenic-rich soot. Bioavailable fractions represented 14.7% of total arsenic, and 1.1% of mercury, while mobile fractions represented 55.5% of arsenic and 7.8% of mercury. Given that concentration of mercury is three times higher than in arsenic-rich soot, release of mercury into the environment would be even higher in absolute terms than for soot. Leaching of arsenic also occurred at alkaline and neutral pH.

Flue dust Composed mainly of quartz, it had significant admixture of calomel ($HgCl_2$), as well as presence of kaolinite, biotite and gypsum. Arsenic concentrations, while being an order of magnitude lower than in soot and stupp, was still extremely high at 1.5%. Mercury concentration was at 0.87%. Arsenic was present as various As(V) species; unlike soot and stupp, it was not bioavailable and suffered very little leaching, with mobile fraction representing just 0.2% of the total arsenic content. Bioavailability of mercury, in contrast, was higher than in soot and stupp, standing at 3.5%, while mobile fraction represented 13.6% of total mercury. Given high mercury concentration, release of mercury into the environment would be very similar to stupp.

Organic contaminants

Pyrometallurgic process, as implemented in El Terronal, required a direct contact of ore with fuel oil, resulting in production of Polycyclic Aromatic Hydrocarbons (PAHs) and their deposition in metallurgic waste. They were present in soot, stupp and flue dust, with flue dust having significantly higher (~7.3x) concentration of them. Out of 16 compounds detected, phenanthrene, fluoranthene, pyrene, chrysene and benzo(k) fluoranthene were the most abundant (Table 2). Increased concentrations of PAHs in the flue dust is mainly due to the lower condensation temperature of most PAHs compared to mercury, favouring their condensation at the end of the pyrometallurgic process. Apart from PAHs, phenylmercury propionate ($C_9H_{10}HgO_2$) was also found. It is a rare, highly soluble and toxic substance that was used an insecticide in the past (European Chemicals Agency, 2011). It was likely formed by combination of mercury and hydrocarbons in the gaseous phase as they passed through the arsenic and mercury condensers.

Other waste

The same work analysed machining waste and fuel oil in other areas of El Terronal. With regards to the fuel oil, absence of linear alkanes and predominance of saturated hydrocarbons and PAHs was consistent with a long process of weathering. Machining waste had secondary arsenic and mercury contamination; arsenic was present as arsenolite that originated in the arsenic condensers. It was suggested that especially fine size of arsenolite particles (<50 μm in diameter) created by condensation process helps it to spread easily via suspension-deposition mechanism, allowing it to affect wider area.

Table 2.: PAH concentrations in representative samples of the solid wastes. Concentrations in ppb. Adapted from Gallego et al. (2015).

Compounds	Calcines	As-rich soot	Stupp	Flue dust
Naphthalene	10.92	22.06	87.89	123.09
Acenaphthylene	4.27	21.53	8.64	21.52
Acenaphthene	9.99	33.2	9.9	318.53
Fluorene	9.76	39.87	20.3	342.79
Phenanthrene	104.71	438.01	1048.1	3134.74
Anthracene	7.76	43.29	34.63	137.78
Fluoranthene	112.53	440.97	335.27	2689.25
Pyrene	56.68	292.35	165.94	1770.93
Benzo(a) anthracene	17.17	120.31	16.49	747.21
Chrysene	51.56	164.89	66.02	938.23
Benzo(k) fluoranthene	45.72	42.7	28.24	1141.65
Benzo(b) fluoranthene	17.91	18.51	9.93	45.97
Benzo(a) pyrene	11.49	21.8	3.38	633.98
Indene(1,2,3-c,d) pyrene	12.06	65.83	6.6	446.6
Dibenzo(a,h) anthracene	2.38	15.3	2.67	96.54
Benzo(g,h,i) perylene	6.19	59.3	5.44	437.99

2.5.2. Soil and groundwater contamination

Presence of large, open spoil heaps containing arsenic- and mercury-rich waste that is both soluble and easily transportable by wind resulted in severe pollution affecting soil and groundwater at El Terronal, as well as surface water downstream of the site. First study of soil and water contamination was performed by Loredo et al. (1999), detecting high concentrations of arsenic and mercury in soil and surface water, as well as in plants growing in the area, including grasses used for grazing of livestock and food crops grown by local populations. Concentrations of arsenic and mercury in soils were highly correlated; highest levels of arsenic and mercury were detected in and around the spoil heaps, with concentrations progressively decreasing with distance.

Subsequent study found that contamination of soil and groundwater was even more extensive, with soil arsenic concentrations as high as 9116 mg/kg (including concentrations of up to 1400 mg/kg found right at the outskirts of Mieres), and arsenic levels in the water downstream of the site reaching 7.9 mg/L (Loredo et al., 2003).

Additional studies were performed to monitor the situation and assess the effectiveness of the waste encapsulation after the AS-I motorway (Ordóñez et al., 2013). Their work showed that encapsulation of the biggest landfills did not significantly decrease surface waters contamination downstream of the site; arsenic concentrations as high as 6 mg/L were found. This suggested that waste remaining in the derelict installations was contributing significantly to the water pollution.

Most recently, a detailed study of soil, surface water and groundwater in the area was performed by González-Fernández et al. (2018). Similarly to the previous studies, concentrations of arsenic were the highest around the spoil heaps in the mining site itself, decreasing with distance (Figure 13). Mercury concentration was highly correlated to that of arsenic. In most of samples taken, those concentrations significantly exceeded safe levels established by the government of Principality of Asturias (BOPA, 2014).

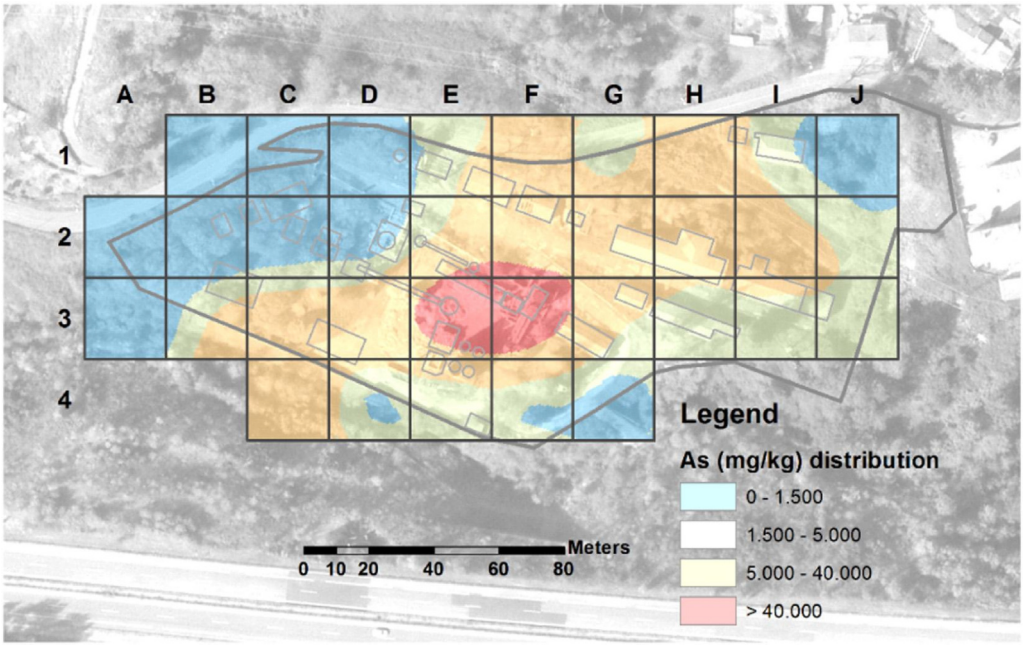


Figure 13.: As (mg/kg) distribution in soils at and around the mining site based on an ordinary kriging with logarithmic transformation and no trend elimination. Adapted from González-Fernández et al. (2018)

Mercury pollution of surface water and groundwater was limited to the El Terronal itself, with San Tirso river having mercury levels below detection limit downstream of the site. Arsenic concentrations downstream of the mining site were similar to the previous studies, exceeding the safe limit of <50 µg/L (BOE, 2015) by orders of magnitude. Measurements taken downstream of the confluence of the San Tirso river with river Caudal have shown levels of arsenic exceeding those limits as well.

San Tirso river sediments downstream of El Terronal were severely affected as well, with 40-fold increase in arsenic and almost 100-fold increase in mercury concentrations compared to the sediments upstream of the site. Sediment quality in Caudal river was not affected.

2.6. Previous remediation work on the site

Given the clear and immediate danger posed by arsenic and mercury, as well as additional contaminants such as PAHs, as well as wide spread of contamination, remediation of this site became a necessity. Most of the work concentrated on the removal of waste or its encapsulation, and was performed by construction crews during the building of the AS-I motorway or by the companies that inherited the ownership of the site. Additionally, two pilot-scale remediation projects were carried out by the Technology, Biotechnology and Environmental Geochemistry research group of the University of Oviedo under LIFE I+DARTS European project (LIFE11 ENV/ES/000547): bacteria-augmented phytoextraction of arsenic with native birch species (Mesa, 2017; Mesa et al., 2017) and arsenic immobilisation by zero-valent iron nanoparticles (Gil-Díaz et al., 2019). Location of the pilot remediation plots are shown on Figure 14.

Bacteria-augmented phytoextraction was shown to be effective in mobilising arsenic from soil and stroing it in the plants' roots, as well as transporting it into the leaves. This opens the possibility that it would be possible, over large timescales, to drain arsenic from the soil and eventually store it in a secure landfill.

Application of zero-valent iron nanoparticles significantly reduced mobility of both arsenic and mercury, achieving up to 51% mobility reduction for arsenic, and 93% mobility reduction for mercury[3]. On a plot with higher degree of contamination results were less pronounced despite repeated application of nanoparticles, reducing arsenic and mercury mobilisation by 35 and 54%, respectively. Immobilisation of contaminants was stable over the course of 32 months. Overall, this strategy was deemed effective, especially for less severely polluted soils.

[3]TCLP and Tessier approaches were used; results of the Tessier approach are given.

Figure 14.: Location of remediation plots within the site overlaid with a satellite image. Green: phytoremediation; blue: soil stabilisation with zero valent iron nanoparticles. Satellite imagery obtained via ArcGIS web server using World Imagery map layer and upscaled using ISR (Cardinale et al., 2018). Map sources: Esri Community Maps Contributors, Dirección General de Catastro, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, METI/NASA, USGS.

3. Aims and objectives

The rationale

El Terronal site presents a series of environmental, technological, biological, administrative and even legal challenges. Extreme degree of arsenic and mercury contamination make land reclamation very difficult and costly, yet its closeness to population centres and on-going release of contaminants (mainly arsenic) require an intensive and concentrated remediation effort. But even if some form of remediation takes place, and arsenic release stops or diminishes to levels we would consider acceptable, the site would still pose a potential biological threat. After more than a century of industrialised mercury mining and almost four decades of abandonment, microorganisms that adapted to survive in the brownfield may have also acquired some traits (such as novel or uncommon antibiotic resistance genes) that would be highly undesirable should they find their way into pathogenic microorganisms.

On the other hand, those highly localised, arsenic- and mercury-rich, nutrient-poor, PAH-contaminated waste can be considered extreme environments. They represent a unique opportunity to study microbial adaptation to such harsh conditions, and to gain more insights on survival of cellular life under extreme environmental pressures. And, similarly to how such an environment may produce organisms with potential to endanger human health, they can also produce organisms with a combination of traits that would make them useful for biotechnological applications.

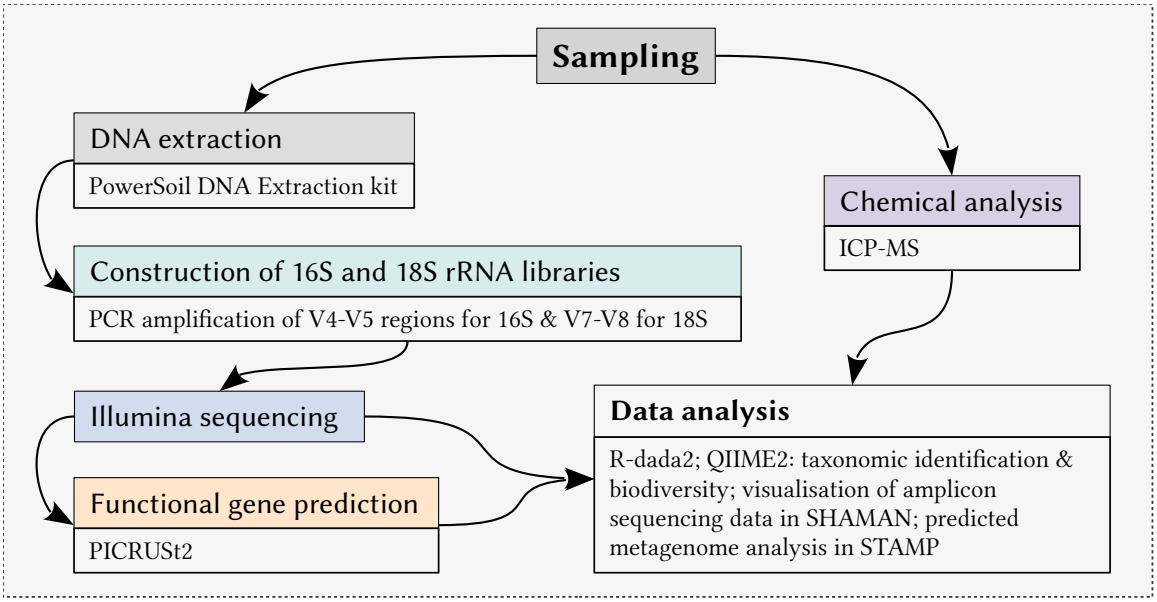
Apart from that, El Terronal is, essentially, both a paradigm and the most extreme known example of an industrial site where mercury was extracted from arsenic-rich ores. Pyrometallurgic process employed for such ores is bound to generate sizeable amounts of highly concentrated, very pure arsenic oxide waste. Thus, even a small-scale mercury mining operation can become a major source of arsenic pollution, regardless of whether the waste was stored compactly on-site or disposed of in remote landfills. While most mercury production sites around the world worked with ores with low arsenic content, a number of places (both in Spain and abroad) had to contend with arsenic-rich ores and contamination they entail.

Thus, any insights gained from studying El Terronal may come useful when dealing with contamination in similar mercury production sites, and potentially - with contamination in and around mining operations dedicated to production of arsenic.

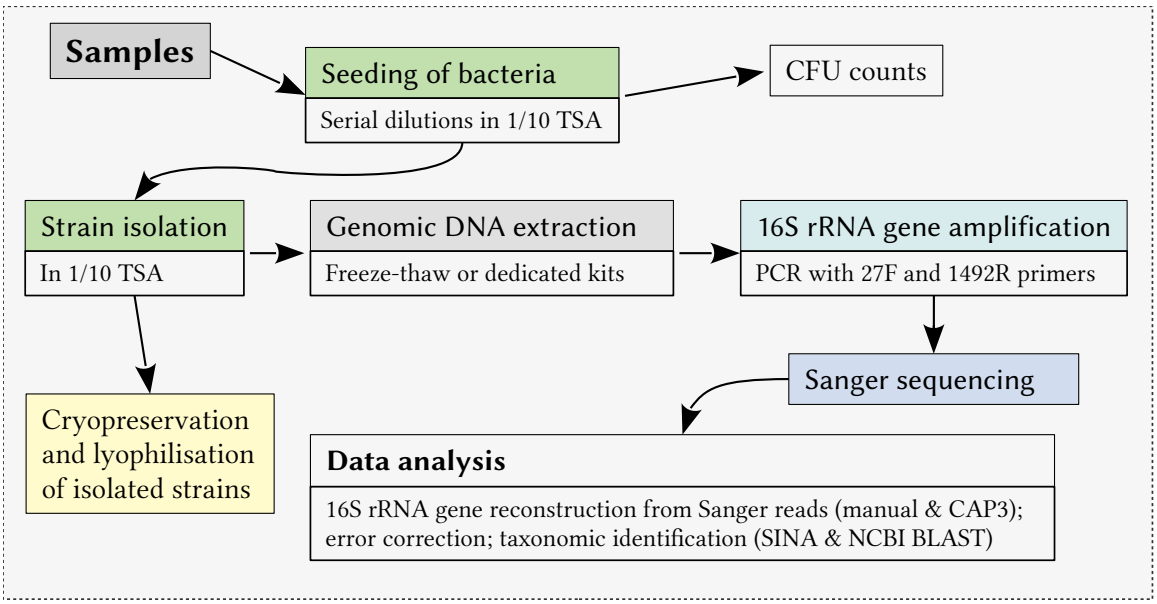
Objectives

The general purpose of this work is to study microbial communities inhabiting arsenic- and mercury-rich brownfield as well as contaminated soils and groundwater sediments of the El Terronal industrial site. To achieve that, the following objectives (and methodological approaches) were established:

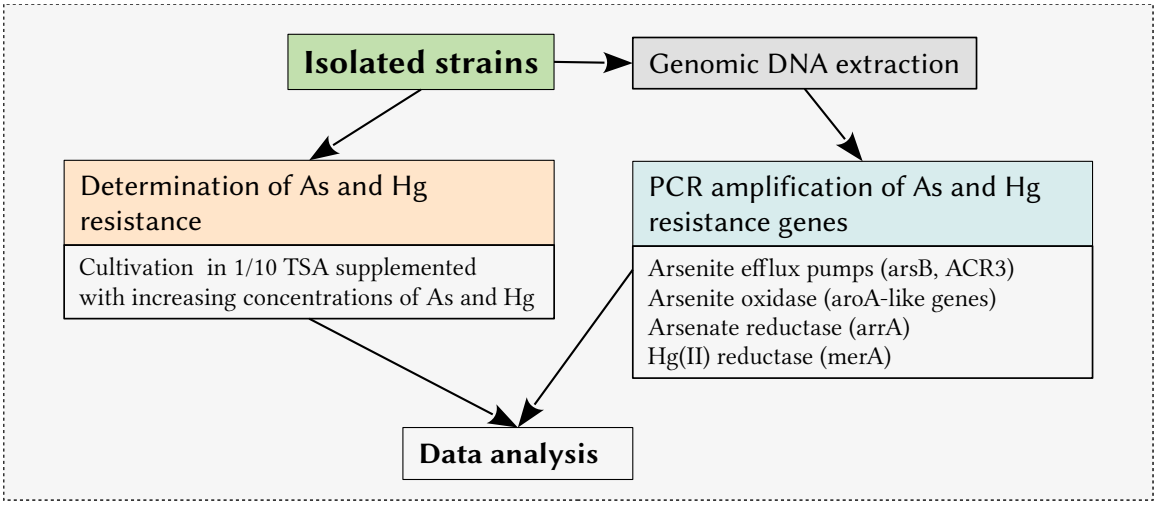
- *Study of bacterial, archaean and fungal communities inhabiting waste heaps, groundwater sediments and soils using metagenomic approach:*



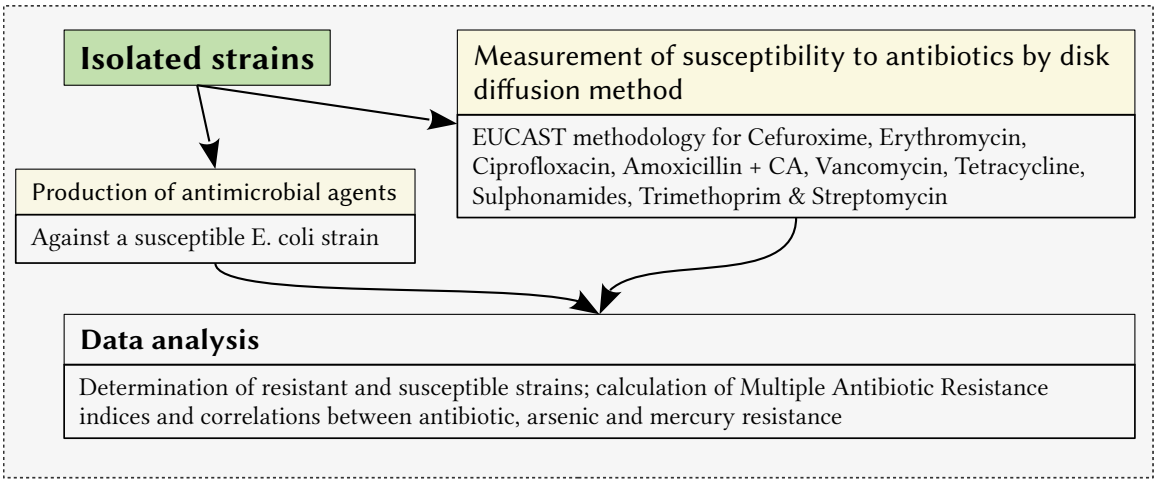
- *Isolation, cultivation and identification of cultivable bacteria inhabiting those environments*:



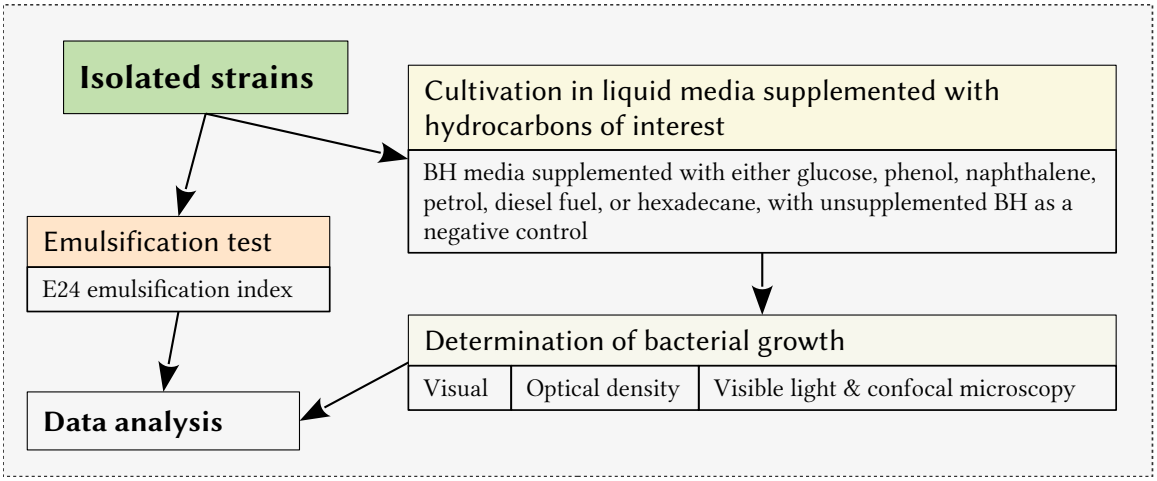
- *Assessment of the degree of arsenic and mercury resistance, and determination of their mechanisms among isolated bacterial strains:*



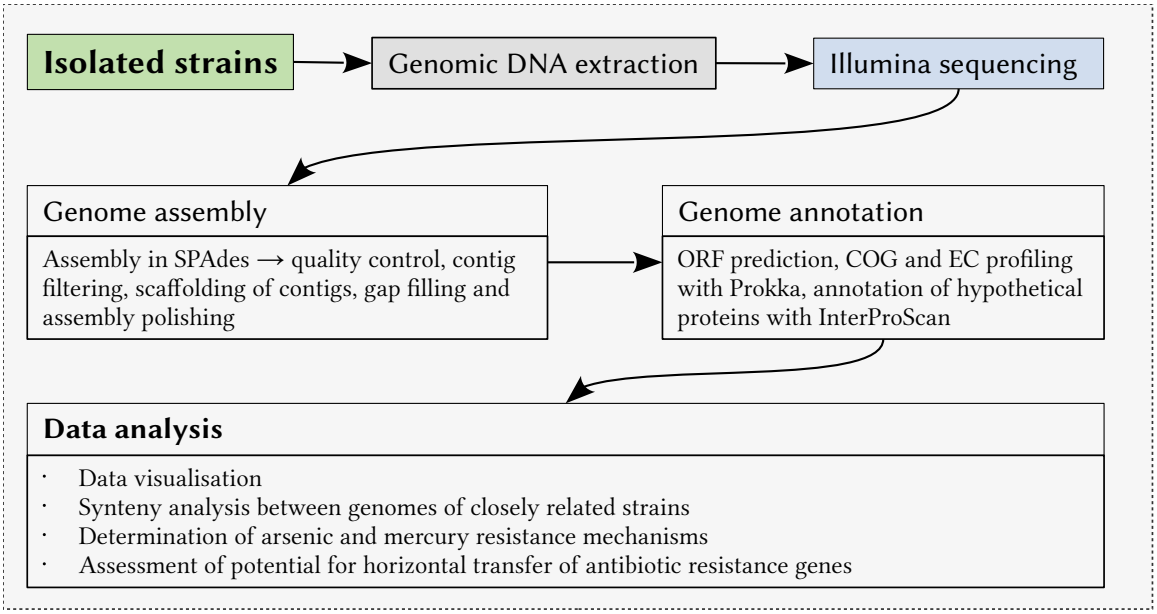
- *Determination of resistance of isolated strains to antibiotics and its correlation with their resistance to arsenic and mercury:*



- *Evaluation of the capacity of isolated strains to degrade hydrocarbons and / or emulsify hydrocarbons:*



Those objectives were supported by the whole-genome sequencing and annotation of strains selected on the basis of their arsenic and mercury resistance, resistance to antibiotics, capacity to degrade and/or emulsify hydrocarbons, or combination of those factors, as depicted below:



4. Materials and methods

4.1. Sample collection and chemical characterisation

Samples of soil from three locations around the mining site and three different types of waste were taken using sterile instruments. Soil samples were taken in the winter of 2015 from the topsoil (sample A), mixed and tilled soil from the phytoremediation plot (B), and soil that formed naturally over time on top of the heavily contaminated river bank (C). They can be seen on the Figure 15.

Figure 15.: Photographs of the soil sampling sites.

Four waste samples (Figure 16) were representative of arsenic-rich soot (samples F and FS), stupp (D), and flue dust (E). Sample FS was taken in the summer 2016; the rest of the samples were taken in the winter of 2015.

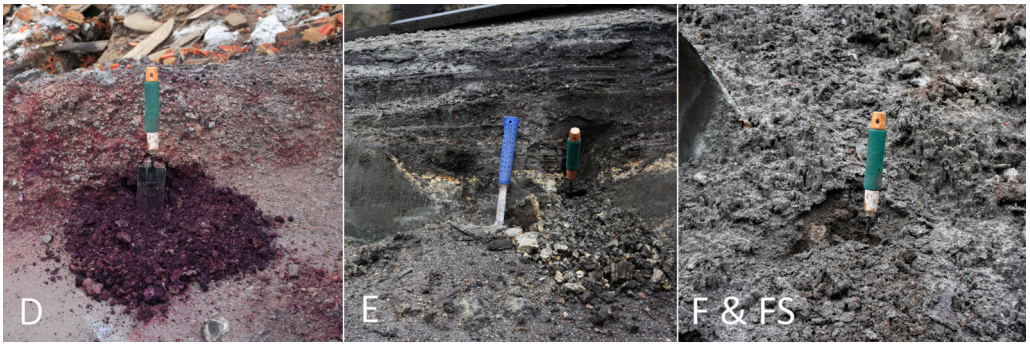


Figure 16.: Photographs of the samples of wastes

Groundwater and groundwater sediment samples were taken in 2016 from two groundwater monitoring wells located at the premises: one south-east of the bioremediation plot (WB and SB for groundwater and sediment, respectively; water table at the depth of 3.5 metres), and another near the river bank (SR for sediment and WR for groundwater; water table at the depth of 2.5 metres).

Sampling locations can be seen on Figure 17. All samples were taken in duplicates.

Figure 17.: Sampling locations overlaid with a satellite photograph of the site. Satellite imagery obtained via ArcGIS web server using World Imagery map layer and upscaled using ISR (Cardinale et al., 2018). Map sources: Esri Community Maps Contributors, Dirección General de Catastro, CC-BY 4.0 scne.es, Esri, HERE, Garmin, INCREMENT P, METI/NASA, USGS.

Sampled materials were placed into sterile 50 mL plastic tubes with screw-on caps (Labbox, Spain) and processed within two days of sampling. To determine concentrations of arsenic and mercury, representative subsamples of soils, waste and groundwater sediments were dried at room temperature to minimise the loss of mercury to evaporation, and ground to the particle size of less than 100 µm. 0.250 g representative subsamples were leached by means of an 'Aqua regia' digestion ($HCl + HNO_3$, 3:1) in an Anton Paar 3000 microwave. The samples were diluted up to 50 mL with ultrapure water and passed through 0.45 µm PTFE filters prior to analysis. As and Hg were quantified by Inductively Coupled Plasma Mass Spectrometry (ICP-MS 7700, Agilent Technologies, California, USA) using IDA (Isotopic Dilution Analysis) with a spike solution from ISC Science, Spain. High-purity standards (Charleston, SC, USA) for calibration and a Certified Reference Material (soil, ERM-CC018) were used.

4.2. DNA extraction from environmental samples, construction and Illumina sequencing of rRNA gene libraries

4.2.1. Nucleic acid extraction from environmental samples

DNA was extracted from 0.25 g of sample using PowerSoil DNA Extraction kit (formerly of MoBio, USA; currently sold as DNeasy PowerSoil Kit, Quiagen, USA) according to the manufacturer's instructions. In most cases, two DNA extraction resulted in enough DNA for downstream analysis. DNA from each sample replica was pooled together and quantified using Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, USA). For samples with very low DNA yield (D, F & FS), as well as groundwater sediment samples (SR and SB), additional DNA extractions were performed, and instead of pooling DNA on the per-replica basis all obtained DNA was pooled together and concentrated using DyNA Vap centrifugal evaporator (Labnet International, USA). This resulted in absence of technical replicates for those samples. Extracted DNA was stored at -20℃.

Only solid samples were used for DNA extraction.

4.2.2. Construction of 16S rRNA and 18S rRNA gene libraries

16S rRNA gene libraries

Amplicons for Illumina MiSeq sequencing were generated from V4-V5 regions of the Bacterial and Archaeal 16S rRNA genes by PCR, using primers 515F (Y. Wang & Qian, 2009; Parada et al., 2016) and 928R (Y. Wang & Qian, 2009). Primers were modified with Illumina overhang adapters (Table 3).

PCRs were performed on a Thermo Fisher Scientific Verity thermal cycler using AmpliTaq Gold 360 polymerase (Thermo Fisher Scientific, USA). Reaction conditions were as follows: initial denaturation at 95℃ for 10 minutes, 25 cycles of 95℃ for 30 seconds, 52℃ for 30 seconds, and 72℃ for 30 seconds, followed by the final elongation of 7 minutes at 72℃. Lack of contamination was ensured by running a blank PCR (with water as template) together with the samples.

For each sample or sample replica, three PCRs were performed; results were pooled and visualised on an agarose gel with ExactLadder DNA PreMix 2 log ladder (Ozyme, France). For samples with low DNA concentration (D, F and DS) two additional amplifications were performed, with subsequent DNA concentration.

18S rRNA gene libraries

18S rRNA gene libraries for Fungi were generated by PCR with FF390 and FR1 primers that target V7-V8 hypervariable regions of the 18S rRNA gene. Originally developed

for DGGE (Vainio & Hantula, 2000), those primers have since been extensively tested and validated for other applications, such as qPCR and high-throughput sequencing (Prévost-Bouré et al., 2011; Banos et al., 2018). This primer pair shows high co-amplification of 18S rRNA genes belonging to the SAR supergroup which requires the use of blocking oligos when working in SAR-rich environments such as marine habitats (Prévost-Bouré et al., 2011). However, no such oligos were employed as most samples were from a terrestrial domain, and it was decided that biodiversity of SAR microorganisms required a study as well. Primers were modified by addition of Illumina overhang adaptors similarly to the 16S primers; PCRs conditions were the same as for 16S rRNA gene amplification, as were negative controls.

For each sample, three PCRs were performed; results were pooled and visualised on an agarose gel with ExactLadder DNA PreMix 2 log ladder (Ozyme, France).

Table 3.: Primers used for the construction of 16S and 18S rRNA gene libraries.

Primer	Illumina adapter (5' to 3')		Primer sequence (5' to 3')
ill-515F	CTTTCCCTACACGACGCTCTTCCGATCT	-	GTGYCAGCMGCCGCGGTA
ill-928R	GGAGTTCAGACGTGTGCTCTTCCGATCT	-	CCCCGYCAATTCMTTTRAGT
ill-FF390	CTTTCCCTACACGACGCTCTTCCGATCT	-	CGATAACGAACGAGACCT
ill-FR1	GGAGTTCAGACGTGTGCTCTTCCGATCT	-	AICCATTCAATCGGTAIT

4.2.3. MiSeq Illumina sequencing

Amplicons were sent to INRAE (Toulouse, France) for indexing and Illumina MiSeq sequencing. Single multiplexing was performed using a home made 6 bp index, which was added to the 3' end of the amplicon during an additional PCR with 12 cycles[1]. Absence of contamination was checked with a negative control, with water serving as template. The resulting PCR products were purified using AMPure XP beads (Beckman Coulter, United Kingdom) and loaded onto the Illumina MiSeq cartridge and sequenced using MiSeq v3 kit forced into the paired-end, 2x250 bp mode. The quality of the sequencing run was checked using PhiX control as well as by including several internal quality standards (mock communities) in the run. Data was generated as a set of compressed .fastq files containing demultiplexed paired-end reads.

[1] Primers 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC-3' and 5'-CAAGCAGAAGACGGCATACGAGAT-index-GTGACTGGAGTTCAGACGTGT-3' were used.

4.3. Isolation and characterisation of bacteria

4.3.1. Isolation and cultivation of bacteria

Microorganisms from the samples were resuspended into a 0.1M sodium pyrophosphate ($Na_4P_2O_7$) solution (Merck, Germany) using Vortex FB15013 shaker (Fisher Scientific, USA), and seeded on agar plates with 1/10 TSA medium (Merck, Germany) using serial dilutions method. Five grams (wet weight) of the sample and 15 mL of $Na_4P_2O_7$ were used; suspension was shaken for 10 minutes at maximum speed and was allowed to settle for 5 minutes, after which 1 mL of supernatant was taken and used to make serial dilutions from 0 up to 10^{-7}. Groundwater samples were seeded directly.

Agar plates were cultivated at 28℃ for two days. Counts of CFUs were taken, and morphological diversity of the colonies was noted. Morphologically distinct, well-separated individual colonies were picked with a sterile seeding hook and streaked on agar plates. All isolation procedures were performed in duplicates.

Pure cultures were maintained by periodic re-seeding on new agar plates. Long-term storage of the original bacterial stock was achieved by freezing bacteria in 10% skim milk (Oxoid, UK) and storing them at -70℃, as well as freeze-drying in 20% skim milk and 1.7% trehalose solution.

4.3.2. Characterisation of arsenic and mercury resistance of isolated strains

All isolated strains were seeded with a sterile toothpick on 1/10 TSA plates with addition of increasing concentrations (Table 4) of either sodium arsenite (VWR, USA), sodium arsenate (Sigma-Aldrich, USA) or mercury(II) chloride (Sigma-Aldrich, USA), and incubated for 24 hours at 28℃, followed by 48 hours at room temperature.

Table 4.: Concentrations of $NaAsO_2$, Na_2HAsO_4, and $HgCl_2$ used for characterisation of As(III), As(V) and Hg(II) resistance.

Contaminant	Concentrations					
As(III)	2 mM	5 mM	10 mM	15 mM	20 mM	25 mM
As(V)	5 mM	20 mM	50 mM	100 mM	200 mM	300 mM
Hg(II)	25 µM	50 µM	100 µM	200 µM	300 µM	400 µM

Plates were inspected visually for signs of bacterial growth. Highest concentrations of contaminants each strain was able to tolerate were recorded.

4.3.3. Antimicrobial susceptibility testing

Antimicrobial susceptibility testing was performed according to the EUCAST disk diffusion method (Matuschek et al., 2014) using a modified version 6.0 of EUCAST disc diffusion method protocol (EUCAST, 2015). Briefly, a single colony was used to prepare a bacterial suspension to the McFarland 0.5 turbidity standard (corresponding approximately to 1-2 $*10^8$ CFU/mL for *E. coli*). Using sterile cotton swabs, Mueller-Hinton agar plates (Oxoid, UK) were innoculated by swabbing in three directions, spreading the inoculum evenly over the entire agar surface. Antimicrobial disks were placed on the surface of agar. Plates were then incubated at 35℃ for 16-20 hours. Strains not showing any growth were additionally incubated for 24 hours at room temperature.

Growth inhibition zones were read visually and measured with a ruler, with plates held against a dark background. EUCAST's *Breakpoint tables for interpretation of MICs and zone diameters v. 8.0* were used to determine whether a given strain was resistant. Clinical and Laboratory Standards Institute's guidelines (Cockerill, 2013) were used to determine antimicrobial susceptibility of microorganisms that were not present in EUCAST tables.

Antimicrobial agents used for the study are listed in the Table 5. *Escherichia coli* strain ATCC 25922 (wild-type, susceptible) was used for quality control.

Table 5.: Antimicrobial agents used for antimicrobial susceptibility testing.

Antimicrobial agent	Concentration	Supplier
Cefuroxime	30 µg	Bio-Rad, UK
Erythromycin	15 µg	
Ciprofloxacin	5 µg	
Amoxicillin / Clavulanic Acid	20 / 10 µg	
Vancomycin	30 µg	
Tetracycline	30 µg	
Compound Sulphonamides	300 µg	Oxoid, UK
Trimethoprim	5 µg	
Streptomycin	10 µg	

Production of antimicrobial agents

All isolates were streaked in two long, parallel streaks on one side of Mueller-Hinton agar plates. After incubation for 16-20 hours at 35℃ (followed by 24 hours at room temperature for strains not showing any growth) wild-type, susceptible *Escherichia coli* strain ATCC 25922 was streaked on the same agar plate perpendicularly to the first streaks. Plates were incubated for 16-20 hours at 35℃, and visually checked for growth inhibition of *E. coli*.

Statistical analysis

Statistical analysis of antibiotic and metal(loid) resistance data was performed in LibreOffice Calc and R. MAR indices were calculated as described by Krumperman (1983).

For individual strains,

$$MAR_{isolate} = \frac{\text{Number of antibiotics strain is resistant to}}{\text{Number of antibiotics tested}}$$

For entire samples,

$$MAR_{sample} = \frac{\text{Aggregate antibiotic resistance score}}{(\text{Number of antibiotics tested}) \times (\text{Number of isolates in the sample})}$$

Necessary adjustments for strains with natural antibiotic resistance (such as resistance to vancomycin in Gram-negative bacteria) were made. Relationship between antibiotic resistance and resistance to arsenic and mercury was assessed by calculating Spearman's rank correlation coefficient.

4.3.4. Hydrocarbon degradation and emulsification

Hydrocarbon degradation

Isolates were seeded into 6 mL of Bushnell-Haas (BH) media (Sigma-Aldrich, USA) supplemented with different hydrocarbons. Inoculated 1/10 TSB media and BH media supplementeed with 1 g/L of glucose (VWR, USA) were used as positive controls; unsupplemented BH media was innoculated to provide negative control. Test tubes with isolates were incubated in an inclined position in Zhicheng ZHWY-100B incubator shaker (Shanghai ZHICHENG Analytical Instrument Manufacturing, China) at 120 rpm and 28℃ for two weeks. Growth was measured as an increment in OD compared to the OD of negative control for the given strain. Strains not providing meaningful OD measurement, such as these forming aggregates or growing along the wall of the test

Table 6.: Hydrocarbon degradation testing.

Hydrocarbon		Concentration	Supplier
Solid	Phenol	0.1 g/L	Fluka Analytical, Switzerland
	Naphthalene	2.5 g/L, partially dissolved	Fluka Analytical, Switzerland
Liquid	Petrol	2%	Repsol, Spain
	Diesel	0.7%	Repsol, Spain
	Hexadecane	0.7%	Fluka Analytical, Switzerland

tube were accessed visually. Hydrocarbons used for testing and their concentrations are presented in Table 6.

Strains showing increase in optical density or other visible signs of growth were visualised using Nikon Eclipse E200 visible light microscope (Nikon Corporation, Japan). Additionally, a vital staining was performed using LIVE/DEAD BacLight Bacterial Viability Kit (ThermoFisher Scientific, USA) containing membrane-permeant nucleic acid stain SYTO9 and membrane-impermeant Propidium Iodide dye. Results were studied on the Leica TCS-SP2-AOBS confocal microscope (Leica Microsystems, Switzerland) at Scientific and Technical Services department of the University of Oviedo. Microphotographs were visualised and edited using ImageJ (National Institutes of Health, USA) with additional plug-ins.

Statistical analysis of bacterial growth and hydrocarbon degradation data was performed in LibreOffice Calc.

Emulsification test

To determine production of surface-active agents, emulsification measurement was performed according to Cooper and Goldenberg (1987), with modifications. Briefly, isolated strains were cultivated in BH medium supplemented with hexadecane at 1% (as emulsification trigger) and glucose to the final concentration of 5 g/L as a carbon source. Zhicheng ZHWY-100B incubator shaker (Shanghai ZHICHENG Analytical Instrument Manufacturing, China) was used; bacteria were incubated at 120 rpm and 28℃ for 5 days. Subsequently, cell-free aqueous solution was obtained by centrifugation (30 min at 9000 g); it was separated from the cell pellet, and a volume of it was mixed with an equal volume of either xylene, hexadecane or octane. Test tubes with mixtures were shaken vigorously, and volume of interphase between aqueous and hydrocarbon solutions was measured 24 hours after shaking. Emulsification index (E_{24}) was calculated as

$$E_{24} = \frac{Emulsion\ height}{Total\ height} \times 100\%$$

where $Emulsion\ height$ is the height of the emulsified interphase zone in the test tube, and $Total\ height$ is the total height of the test tube.

4.3.5. Bacterial gDNA extraction

Genomic DNA was extracted using either GeneMATRIX Bacterial & Yeast Genomic DNA Purification Kit (EURx, Poland) or by the freeze-thaw method followed by DNA precipitation in isopropanol. DNA extraction kits were used according to manufacturers' instructions. Extraction by the freeze-thaw method was performed by resuspending an individual colony in 500 microlitres of PBS , and subjecting it to three cycles of freezing (for 5 minutes in ethanol cooled to -70℃) and thawing (5 minutes in a hot water bath at 90℃), with mixing by tube inversion before each freeze cycle. Resulting suspension was centrifuged at 12000 g, and DNA-containing supernatant was collected into a clean 1.5 mL Eppendorf tube. A volume of isopropanol equal to the supernatant volume and 5M NaCl to the final concentration of 0.5M were added; resulting solution was mixed by inverting the tube several times, and incubated at room temperature for 2 hours. Isopropanol was then discarded, the DNA pellet at the bottom of the tube was washed in 70% ice-cold ethanol, and allowed to dry completely at room temperature. DNA pellet was then re-suspended in 100 µL of Tris-HCl buffer. All reagents used were provided by Sigma-Aldrich, USA. All extracted DNA was stored at -20℃.

4.3.6. Amplification of 16S rRNA genes for identification of isolated strains

Amplification of 16S rRNA genes was done by PCR using Bio-Rad MJ Mini thermal cycler (Bio-Rad, USA). OptiTaq PCR Master Mix (2x) kit (EURx, Poland) together with primers 27F and 1492R (Lane, 1991) was used, producing an amplicon in the 1450-1500 bp range. PCR program was as follows: 5 minutes at 95℃, 25 cycles of 20 seconds at 93℃, 30 seconds at 58℃ and 1.5 minutes at 72℃, followed by 7 minutes at 72℃. A blank PCR reaction was used as a negative control.

Success of the amplification and absence of non-specific amplification or contamination were checked by electrophoresis in a 1% agarose gel with λDNA/ PstI molecular weight ladder (Fermentas Life Sciences, Lithuania).

Up to three amplicons from 50 µL reactions for each bacterial strain were pooled together and purified with the help of Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare, USA) according to the manufacturer's instructions.

Purified amplicons were stored at -20℃ till further use.

Table 7.: Primers used for 16S rRNA gene sequencing.

Primer	Sequence (5' to 3')	Reference
27F	AGAGTTTGATCMTGGCTCAG	Lane (1991)
341F	CCTACGGGAGGCAGCAG	Muyzer et al. (1993)
515F	GTGCCAGCMGCCGCGGTAA	Y. Wang and Qian (2009) and Parada et al. (2016)
518R	ATTACCGCGGCTGCTGG	Muyzer et al. (1993)
907R	CCGTCAATTCMTTTGAGTTT	Lane (1991)
1492R	TACCTTGTTACGACTT	Lane (1991)

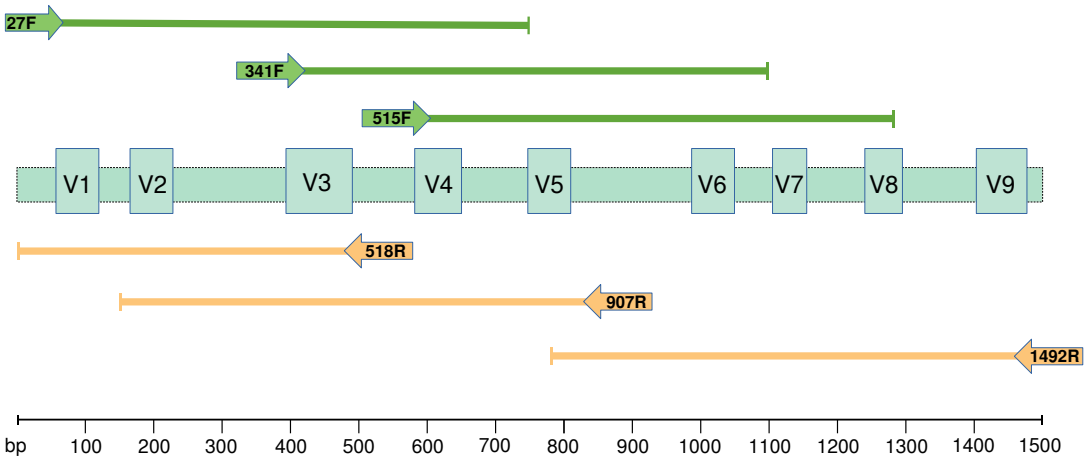


Figure 18.: Target regions for 16S Sanger sequencing and generated reads.

Sanger sequencing was performed on the ABI PRISM 3130xl Genetic Analyzer in the Sequencing laboratory of the Scientific and Technical Services department of the University of Oviedo. Data was received as chromatograms (.ab1 files) and imported into Unipro UGENE bioinformatics software (Okonechnikov et al., 2012), where they were aligned and checked against each other. Consensus sequences spanning most of the amplicon length were exported and additionally screened for errors in conserved regions by using BLAST against NCBI RefSeq and nt databases (Altschul et al., 1990).

Taxonomic identification was performed using SINA (Pruesse et al., 2012) on-line tool for the SILVA rRNA databases (Quast et al., 2012), as well as NCBI Blast (Altschul et al., 1990) against RefSeq 16S rRNA database for species-level affiliation. Identity cut-off value was 95% for genus identification (SINA default setting), and defined as 99% for species-level affiliation as per Stackebrandt and Ebers (2006) and Kim et al. (2014), further revised upwards by Edgar (2018).

4.3.7. Screening for arsenic and mercury resistance genes

Genes responsible for arsenite efflux (*arsB*, *acr3*), arsenite oxidation (*aio*-like genes), arsenate reduction (*arrA*) and mercury(II) reduction (*merA*) were amplified by PCR (Table 33). Either OptiTaq PCR Master Mix (2x) kit (EURx, Poland) or Speedy Supreme NZYTaq 2x Green Master Mix kit (NZYTech, Portugal) were used; PCR programs were adjusted for 10x higher amplification speed of Speedy Supreme NZYTaq whenever it was utilised. Results were visualised on 1% agarose gel with λDNA/ PstI molecular weight ladder (Fermentas Life Sciences, Lithuania). A blank PCR reaction was used as a negative control.

Photographs of the gels were taken using Kodak Gel Logic 200 imaging system (Eastman Kodak, USA).

4.3.8. Whole-genome sequencing of isolated strains

Strains were selected based on a combination of traits such as resistance to concentrations of arsenite exceeding 15 mM and/or resistance to mercury at concentrations higher than 300 μM, low number of amplified arsenic resistance genes combined with high arsenic resistance, as well as multiple antibiotic resistance and capacity to degrade hydrocarbons.

Genomic DNA of selected strains was extracted using NZY Microbial gDNA Isolation kit (NZYTech, Portugal), and sent to Eurofins Genomics (Ebersberg, Germany) for Illumina sequencing. Results were received as .fastq files containing paired-end 150 bp reads.

4.4. Data processing

4.4.1. Processing of 16S and 18S Illumina-generated data

Files in .fastq format containing Illumina reads were imported into R (R Core R Core Team, 2020). Removal of non-biological sequences, quality filtering, trimming, demultiplication, merging of paired-end reads, removal of chimeric sequences and generation of the Amplicon Sequence Variants[2] (ASV) table was done using dada2 package (Callahan et al., 2016). Primers were removed according to their length. For 16S rRNA gene libraries, forward reads were trimmed to 240 nucleotides, while reverse reads were trimmed to 200 nucleotides. For 18S rRNA gene libraries, full-length forward reads were used, while reverse reads were trimmed to 230 nucleotides. In all cases, quality filtering was set to 5, and PhiX detection and removal was enabled. To increase recovery rate of ASVs with very low copy numbers, the core dada2 algorithm was run

[2]Amplicon Sequence Variants are 100% identity-level OTUs obtained by de-noising algorithms such as dada2 and UNOISE, while traditional OTUs are obtained via clustering of 95-97% identity sequences.

with sample pooling enabled (`pool = TRUE` option). Resulting ASV tables were then imported into QIIME2 (Bolyen et al., 2019) for further analysis.

Rarefaction and alpha diversity

To ascertain representativeness of the data, as well as generate correct alpha diversity indices of all samples, rarefaction curves were obtained by measuring alpha diversity on progressively-larger data subsets from the samples using QIIME2's built-in tools. Diversity indices used for rarefaction curves included Chao1 (Chao, 1984) index, Faith's Phylogenetic Diversity index (Faith, 1992), number of observed features (DeSantis et al., 2006), Lladser's point estimate (Lladser et al., 2011), Robbins' estimator (Robbins, 1968) and Good's coverage of counts (Good, 1953), as well as Shannon (Shannon, 1948) and Simpson (Simpson, 1949) indices. When representative level of rarefaction was achieved, samples were rarefied to that depth, and in addition to the aforementioned indices, ACE (Chao & Lee, 1992) index, effective number of species/probability of intra-or interspecific encounter metric (Chase & Knight, 2013) and Pielou's evenness (Pielou, 1966) were measured.

Taxonomic classification

Taxonomic classification was performed using SILVA database. To improve accuracy of taxonomic classification (Werner et al., 2011), built-in classifier training tools in QIIME2 were used to generate a custom-made 99% sequence identity Bayesian classifier based on 16S SSU rRNA SILVA128 majority taxonomy dataset (Yilmaz et al., 2013). Classifier was trained on V4-V5 hypervariable regions flanked by 515F and 928R primers (Table 3) with *scikit-learn* version 0.19.1. For 18S SSU rRNA, a full-length classifier based on SILVA128 dataset was used. Additional taxonomic classification was obtained by aligning ASVs against NCBI nt database with the help of *q2-brocc* plug-in for QIIME2 (Dollive et al., 2012).

Resulting taxonomy tables were manually compared and curated: sequence variants representing *Chloroplast* and *Mitochondria* as well as non-identified sequences were removed, ASVs that were not classified similarly in both SILVA and NCBI nt databases were manually checked against SILVA and RefSeq databases using corresponding on-line tools, and ASVs not represented in the SILVA128 database were replaced with classification entries obtained from the NCBI nt database and manually reformatted into SILVA format.

Phylogenetic trees were built by inserting target sequences into SILVA128-based reference tree. *q2-fragment-insertion* plug-in (Janssen et al., 2018) implementing SEPP (Mirarab et al., 2011) approach to insertion was used. Reference tree supplied with the plug-in was built with RAxML version 8.2.3 (Stamatakis, 2014) upon SILVA 128 dataset aligned at 99% sequence identity.

Defining the place of isolated strains within the community

For better understanding of the place of cultivable organisms in the total structure of the microbial community, *in silico* PCR was performed on 16S rRNA gene sequences from isolated strains using UGENE. Primer sequences were analagous to those used for generating 16S rRNA gene libraries for Illumina sequences. Obtained ASV analogues were compared to the Illumina-generated data.

Beta diversity

Variance-Adjusted Weighted UniFrac (Chang et al., 2011) matrices were calculated using QIIME2's built-in tools and used for generating PCoA plots. Statistical significance of the differences between samples and sample groups was determined by analysis of similarities (Clarke, 1993); ADONIS (Anderson, 2001) test was used to explore their relationship with environmental factors and variables.

Data visualisation

ASV and taxonomy tables, as well as information about the samples were exported from QIIME2 using built-in export functionality and uploaded to SHAMAN website (Volant et al., 2020). PcoA data was visualised as interactive Emperor plots using QIIME2's built-in tools (Vázquez-Baeza et al., 2013; Vázquez-Baeza et al., 2017).

Predictive functional profiling of microbial communities

Metagenomic profiles were predicted using PICRUSt2 tool (Douglas et al., 2020). Profiles for Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologues, Clusters of Orthologous Groups (COG), as well as Enzyme Commission (EC) numbers were predicted both for the microbial communities and for individual ASVs (`--strat_out` flag enabled). EC numbers were subsequently re-grouped into metabolic pathways according to the MetaCyc pathway database (Caspi et al., 2019). To estimate the accuracy of predicted data, PICRUSt2-generated COG profiles for several ASVs were imported into STAMP (Parks et al., 2014) and compared to their counterparts among sequenced cultivable bacteria (see Section 4.4.2) by means of G-test with Yates' correction (as individual genomes), and White's non-parametric t-test (White et al., 2009) as a group. Šidák correction was subsequently applied to p-values. Metagenomic data (metagenomic profiles for individual samples and groups of samples) was analysed and visualised in STAMP as well.

4.4.2. Genome assembly and annotation

De novo genome assembly

Assembly Pre-assembly steps included Illumina adapter trimming with Skewer (Jiang et al., 2014), and generating extended reads from overlapping paired-end reads with FLASH (Magoč & Salzberg, 2011). Resulting reads (forward, reverse, and extended), as well as 16S rRNA gene sequences from Sanger sequencing were used for *de novo* assembly with SPAdes version 3.13.1 (Bankevich et al., 2012). Assembly was run with sequencing error correction and repeat resolution (default behaviour unless `--only-assembler` flag is passed), as well as mismatch correction (`--careful` flag) enabled; k-mer lengths to be used in the assembly process were set to 21, 33, 55, 77, 99 and 121 nucleotides.

Quality control and contig filtering Genome assemblies were visualised in Bandage (Wick et al., 2015) for assembly quality control; additionally, quality checks were performed using QUAST (Gurevich et al., 2013). To assess possible contamination, all contigs produced by SPAdes were imported into UGENE and taxonomically classified by Kraken (Wood & Salzberg, 2014), CLARK (Ounit et al., 2015) and DIAMOND (Buchfink et al., 2014), with subsequent consensus classification generated with WEVOTE (Metwally et al., 2016). Taxonomic classification, as well as contigs and coverage information produced by SPAdes were imported into BlobTools (Laetsch & Blaxter, 2017) for visualisation and filtering, where short contigs that showed highly dissimilar coverage, GC skew and taxonomic identity were filtered out.

Scaffolding and assembly improvement Filtered contigs were scaffolded using ScaffMatch (Mandric & Zelikovsky, 2015) with default parameters. Information on insert size of the library and standard deviation of insert size required by the scaffolder were extracted from SPAdes' assembly log. Gaps in the scaffolds were filled using GapFiller (Boetzer & Pirovano, 2012) using a minimal overlap length of 65 bp. Finally, assembled genomes were checked and corrected with ntEdit (Warren et al., 2019).

Genome annotation

Assembled genomes were annotated using Prokka (Seemann, 2014) running with default parameters, generating COG- and EC-based annotations. Additionally, all putative protein sequences that were not identified by Prokka were exported as .faa files using Seqtk (H. Li, 2012) and searched against CATH-Gene3D (Lewis et al., 2017), CDD (Marchler-Bauer et al., 2016), HAMAP (Pedruzzi et al., 2014), PANTHER (Mi et al., 2016), Pfam (El-Gebali et al., 2018), PIRSF (Nikolskaya et al., 2006), PRINTS (Attwood et al., 2012), ProSite profiles (Sigrist et al., 2012), SFLD (Akiva et al., 2013), SMART (Letunic et al., 2014; Letunic & Bork, 2017), SUPERFAMILY (Oates et al., 2014) and TIGRFAMs (Haft et al., 2013) databases with InterProScan 5 (Jones et al., 2014; Mitchell et al., 2018).

Search for integrative and conjugative elements was performed using ICEfinder front-end for the ICEberg 2.0 database (M. Liu et al., 2018). Screening of ICEs and genome assemblies for clinically relevant antibiotic resistance genes was performed by Resistance Gene Identifier online tool (Alcock et al., 2019). Perfect, Strict and Loose hits were permitted for integrative and conjugative elements; only perfect and strict hits were permitted for whole genomes.

Plasmid assembly and annotation

In addition to the main genome, assembly of plasmids was also attempted. To achieve that, SPAdes was run with the `--plasmid` flag enabled; other settings were the same as for the main genome assembly. Quality of the assemblies was checked in Bandage; in case of partial (non-circular) contigs, additional scaffolding was attempted using SSPACE (Boetzer et al., 2010) running with default settings. Gaps in contigs were filled with GapFiller, and resulting contigs were checked and corrected with ntEdit similarly to the main genome. Annotation was performed in InterProScan.

4.4.3. Genome visualisation and comparisons

Annotated genomes and putative plasmids were visualised and analysed in JBrowse (Buels et al., 2016). Synteny between genomes of strains belonging to the same genus was inferred with Sibelia (Minkin et al., 2013) using minimal synteny block size of 2500 nucleotides. Circular visualisation of genomic data was created in Circos (Krzywinski et al., 2009).

5. Results

5.1. Chemical analysis of the samples

Due to predominance of As and Hg contamination in the area, chemical characterisation of the samples in this work was focused solely on measuring arsenic and mercury concentrations (Table 8). Arsenic-rich metallurgic waste contained around 55% of arsenic and almost 3% of mercury. Stupp samples had lower concentration of arsenic (around 12%), but much higher concentration of mercury, almost reaching 6.6%. Flue dust was about 2% arsenic and 0.77% mercury.

Table 8.: Arsenic and mercury concentration in solid samples as measured by ICP-MS. Each value is a rounded average of three measurements.

Sample	As, $mg \cdot kg^{-1}$	Hg, $mg \cdot kg^{-1}$	Sample description
A	2630	840	Topsoil
B	2635	155	Soil (phytoremediation plot)
C	5390	1130	Soil (river bank)
D	119000	65880	Stupp
E	20115	7705	Flue dust
F	552400	28200	As-rich soot
FS	552430	28280	As-rich soot
SR	30000	800	Groundwater sediment (river bank)
SB	8000	600	Groundwater sediment (remediation plot)

Soil samples have shown high degree of arsenic and mercury contamination as well, without reaching the levels observed in the waste. Soil recovered at and near the phytoremediation plot (sample series A and B) was contaminated to a lesser degree compared to the soil from the river bank. Additionally, while samples A and B had almost identical concentrations of arsenic, mercury concentration was significantly lower in the sample taken from the phytoremediation plot. In all likelihood, this has occurred due to the increased mercury loss to evaporation as a result of tilling the soil. Groundwater sediment samples have shown significant arsenic contamination, with sediment from the well near the river containing as much as 3% of arsenic. Concentrations of mercury

were similar to the soil levels. Incidence of arsenic and mercury in solid samples was correlated ($\rho = 0.79, p = 0.003$), which is expected for contamination originating from the same source.

In line with previous studies (Gallego et al., 2015; González-Fernández et al., 2018), concentrations of As and Hg in solid samples were above regulatory limits. According to local regulations[1], maximum permitted soil level of arsenic on a land destined for industrial use was set at 200 mg/kg, and for residential use – at 40 mg/kg, while maximum allowed mercury levels were 100 mg/kg and 10 mg/kg for industrial and residential use, respectively. In the study site, soil samples significantly exceeded those limits (Figure 19), while As and Hg concentrations in wastes and groundwater sediments exceeded those limits by orders of magnitude.

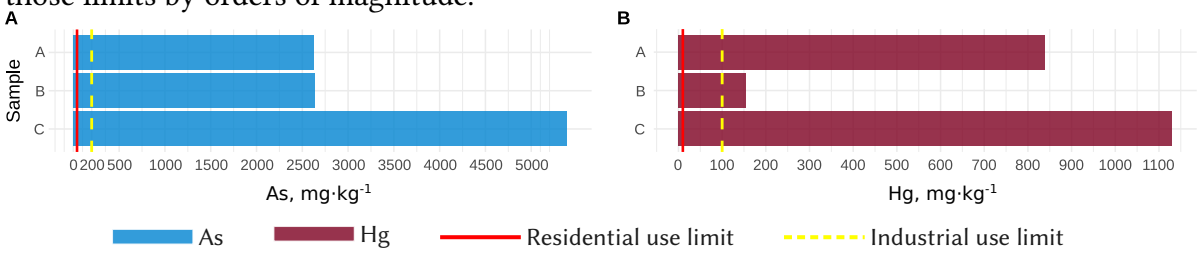


Figure 19.: A: Arsenic concentration in soil samples (A, B & C). B: Mercury concentration in soil samples (A, B & C). Baselines represent maximum permitted levels for each contaminant.

Groundwater was also affected by the arsenic and mercury contamination (Table 9). In both locations, contamination levels significantly exceeded regulatory limits[2] of 50 µg/L for arsenic and 0.07 µg/L for mercury.

Table 9.: Arsenic and mercury levels in groundwater. Both samples were taken in May, 2016.

Sample	Arsenic, $mg \cdot L^{-1}$	Mercury, $mg \cdot L^{-1}$	Sample description
WR	200000	7	Groundwater (river bank)
WB	100	0.5	Groundwater (remediation plot)

[1]BOPA, 2014.
[2]BOE, 2015.

5.2. 16S and 18S rRNA gene metagenomic studies

5.2.1. Coverage and general statistics

Solid samples (soils A, B & C; flue dust E; stupp D, arsenic-rich soot F & FS, as well as groundwater sediments SR and SB) were used for construction of 16S and 18S rRNA libraries. After merging Illumina reads and removal of non-biological, plastid and mitochondrial sequences, 4067 unique amplicon sequence variants (ASVs) belonging to Bacteria and Archaea were generated from sequenced 16S libraries. Similar procedure yielded 1081 unique ASVs of Fungi, Chromista and Protozoa from 18S libraries (Table 10).

Table 10.: General data on Illumina sequencing results for 16S and 18S rRNA gene libraries.

Library	Input reads[3]	ASV count				No. of unique ASVs
		Initial	Chimera removal	Taxa filtering	Coverage	
16S	470516	225798	208016	205787	0.91	4067
18S	476082	356616	324746	321719	0.90	1081

For 16S rRNA gene libraries, sequence counts had significant inter-sample differences, ranging from 6598 to 46635 sequences, with mean frequency of 15829.8. On the per-ASV level, count frequency ranged from 1 to 12915, with mean frequency of 50.6 counts/ASV. Frequency per sample differences were lower in 18S rRNA gene libraries, with sequence count ranging from 29037 to 44663 sequences, with mean frequency of 35746.5; per-ASV frequency ranged from 1 to 64577, with mean frequency of 297.6 counts per sequence (Table 11).

[3] $N_{reads} = N_{forward\ reads} = N_{reverse\ reads}$

Table 11.: ASV counts for each sample in 16S and 18S rRNA gene libraries. For 16S libraries, in case of availability of replicas, ASV count for each sample replica is given.

Sample	Replica	Sequence count		Sample description
		16S	18S	
A	A1	7004	30420	Topsoil
	A2	7906		
B	B1	13196	29037	Soil (phytoremediation plot)
	B2	11664		
C	C1	6598	29575	Soil (river bank)
	C2	9034		
D	D	11375	39474	Stupp
E	E1	12208	30483	Flue dust
	E2	20187		
F	F	46635	43130	As-rich soot (winter)
FS	FS	31602	44663	As-rich soot (summer)
SR	SR	11189	43324	Groundwater sediment (river bank)
SB	SB	17189	31613	Groundwater sediment (remediation plot)

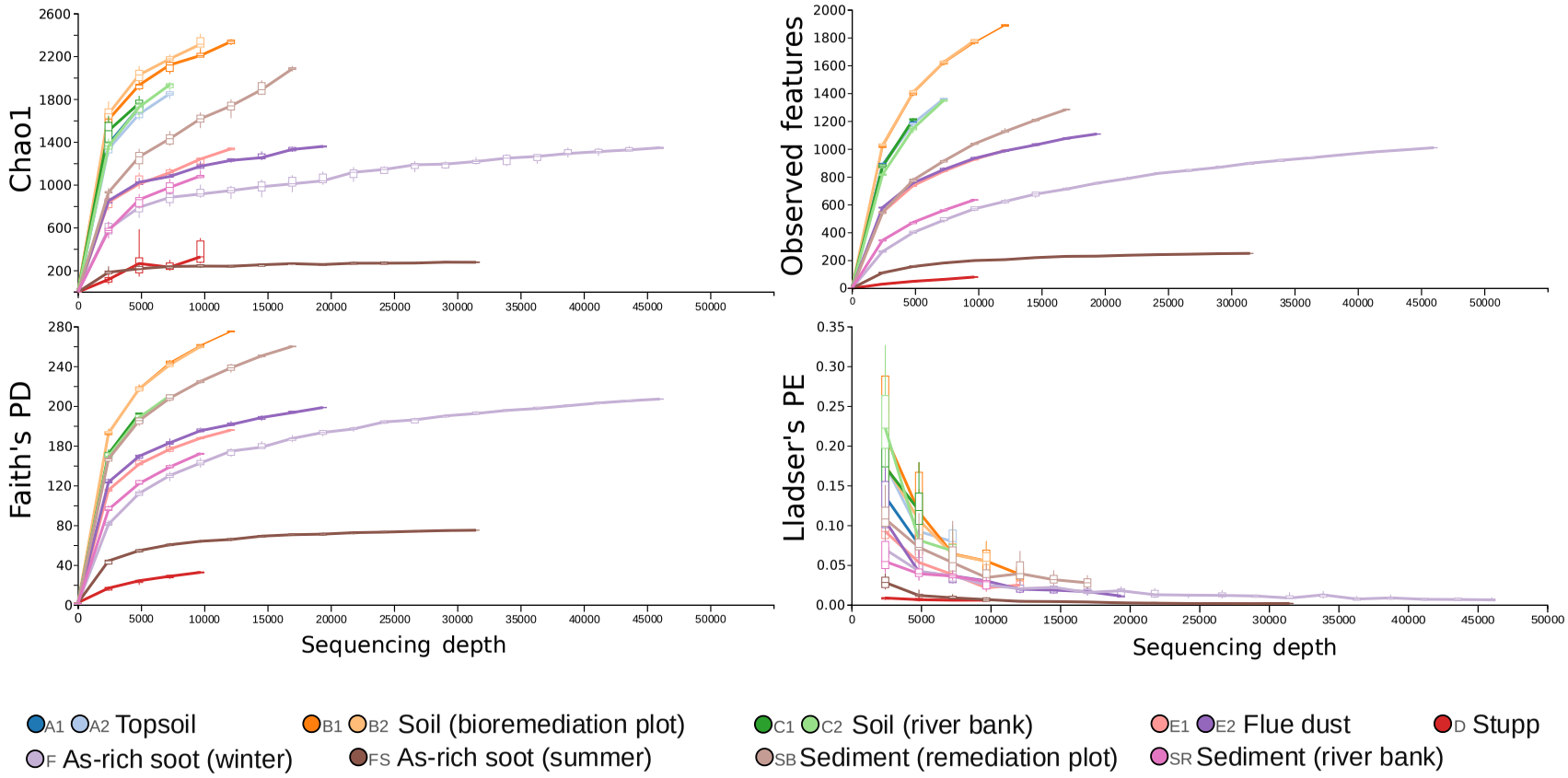
5.2.2. Alpha-diversity and sequencing depth

Bacteria and Archaea

Rarefaction curves for Chao1 index, Faith's Phylogenetic Diversity, number of Observed ASVs and Lladser's Point Estimate have shown that achieved sequencing depth, while representative of the bacterial community, was not sufficient to cover all of biodiversity in some of the samples (Figure 20). In Chao1, Faith's PD and Observed ASVs metrics, rarefaction curves for soil samples (A, B & C) and groundwater sediment sample SB never fully flatten out, suggesting that some of the less-abundant species in those samples were not sequenced. Lladser's Point Estimate[4] also indicated that species with relative abundances below (on average) $9.6 \cdot 10^{-2}$, $4.79 \cdot 10^{-2}$ and $5.56 \cdot 10^{-2}$ for soil samples A, B and C, respectively, were left out due to insufficient sequencing depth. Rarefaction curves based on Simpson's[5] and Shannon-Wiener alpha-diversity indices, Good's coverage of counts and Robbins' estimator fared very poorly in comparison with other metrics, and failed to identify insufficient sequencing depth (Figure 67).

[4]Lladser's Point Estimate predicts how much of the environment contains upsampled taxa.
[5]Here and throughout the work, Simpson's index refers to the 'inverse' Simpson $(1 - D)$

Figure 20.: Bacteria and Archaea. Rarefaction curves for Chao1 index, Faith's PD, Observed Features and LLadser's PE.



In general, sequencing depth and community profiles were considered representative, even thought some of the rarer ASVs were not recovered. On the other hand, while rarefying data steadily becomes less and less acceptable (McMurdie & Holmes, 2014; Willis, 2019), extremely low biodiversity of microbial communities from samples with high sequencing depth (D, F & FS) meant that they could be rarefied to the depth of soil samples without any significant loss of data, simplifying downstream analysis.

Chao1 index estimates species richness by using abundances of rare species (singletons and doubletons); ACE uses abundance data for species with less than 10 individuals and incidence data for the rest of the community; they both indicated that alpha-diversity in samples of stupp (D), soot (F, FS), flue dust (E) and groundwater sediments (SR & SB) was significantly lower than in the soil samples (Figure 21a, 21b).

ENS_{PIE} calculates Effective Number of Species / Probability of intra-or-interspecific encounter metrics; it is, primarily, an evenness metric that is considered to be independent of the sample size. This metrics differed from the rest: while soil samples A and B had high level of diversity, in the soil sample C (river bank) it was not nearly as high. Significantly lower levels of diversity were observed in the samples of waste (D, E, F & FS) and sediments (SR & SB) indicating uneven microbial populations dominated by a relatively low number of distinct organisms relative to the soil samples (Figure 21c).

Faith's PD uses phylogenetic distances between ASVs to measure diversity of the community. It was in broad agreement with Chao1 and ACE indices, but indicated very low diversity in the stupp sample, suggesting that microbial community there consisted mostly of closely-related organisms (Figure 21d).

Observed features (also called observed species or observed OTUs) is a simple incidence metrics counting the number of unique features observed in the sample. Similarly to Faith's PD, it indicated very low biodiversity in the stupp (D), with all Bacteria and Archaea in the community profile belonging to just 66 distinct ASVs (Figure 21e).

Shannon-Wiener index (H') takes both abundance and evenness of the species into account without leaning too heavily onto any of them. Differences indicated by it were similar to ACE , Chao1 and Faith's PD indices, although suggested diversity estimates were higher (Figure 21g).

Pielou's evenness (J) is a measure of relative evenness of species; it is derived from Shannon-Weiner index. Simpson's index measures relative abundance of different species. Both indices (but Pielou's J especially) are sensitive to sample size (Mccune & Grace, 2002); they identified lower diversity in stupp (D) and soot (F) samples, but were otherwise not very indicative of the differences between samples (Figures 21f, 21h).

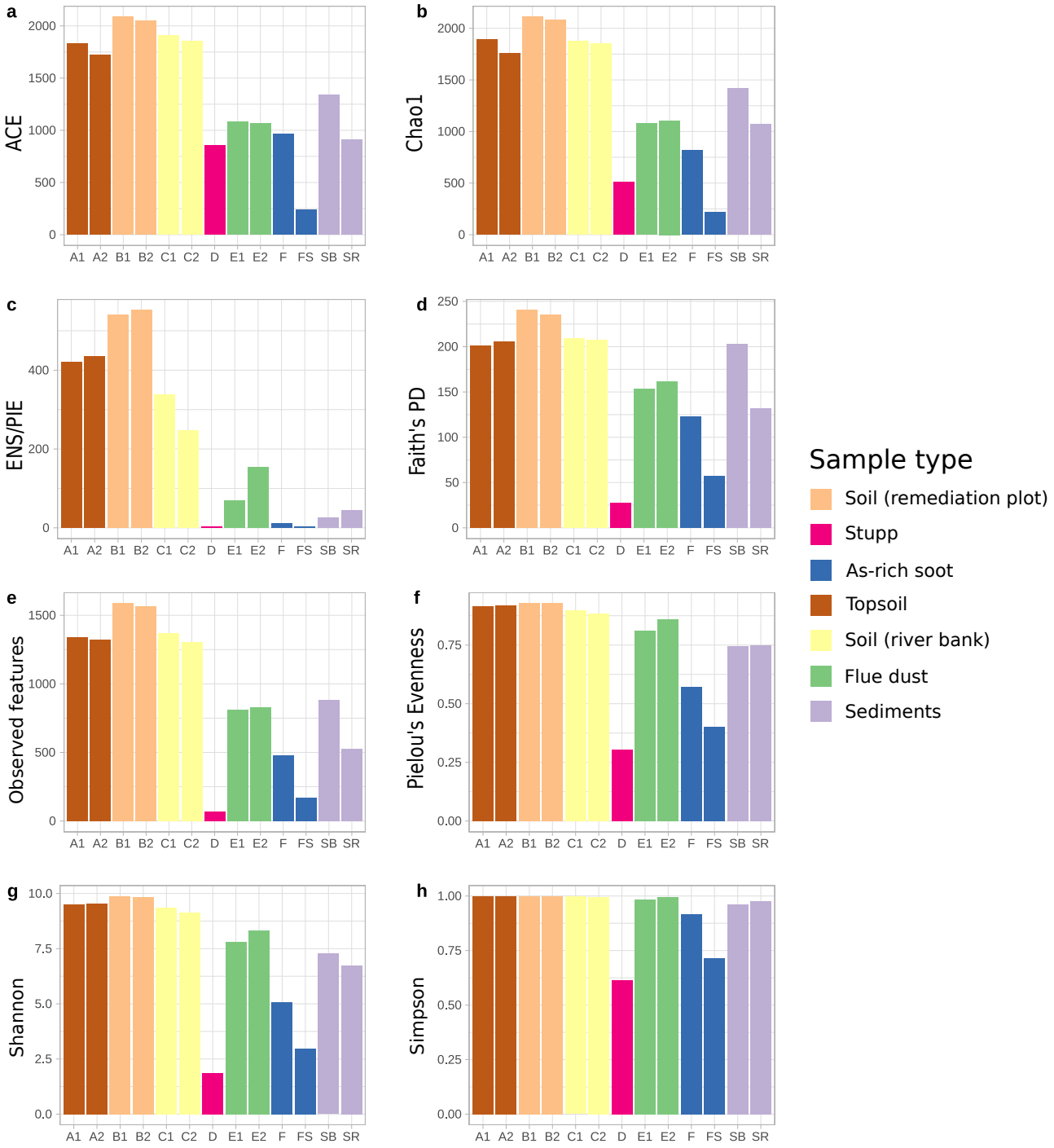


Figure 21.: Alpha-diversity indices of Prokaryotic communities. a: Abundance-based Coverage Estimator; b: Chao1 index; c: effective number of species / probability of intra- or interspecific encounter; d: Faith's Phylogenetic Diversity; e: observed features; f: Pielou's evenness; g: Shannon-Weiner index; h: Simpson's index.

Fungi and SAR

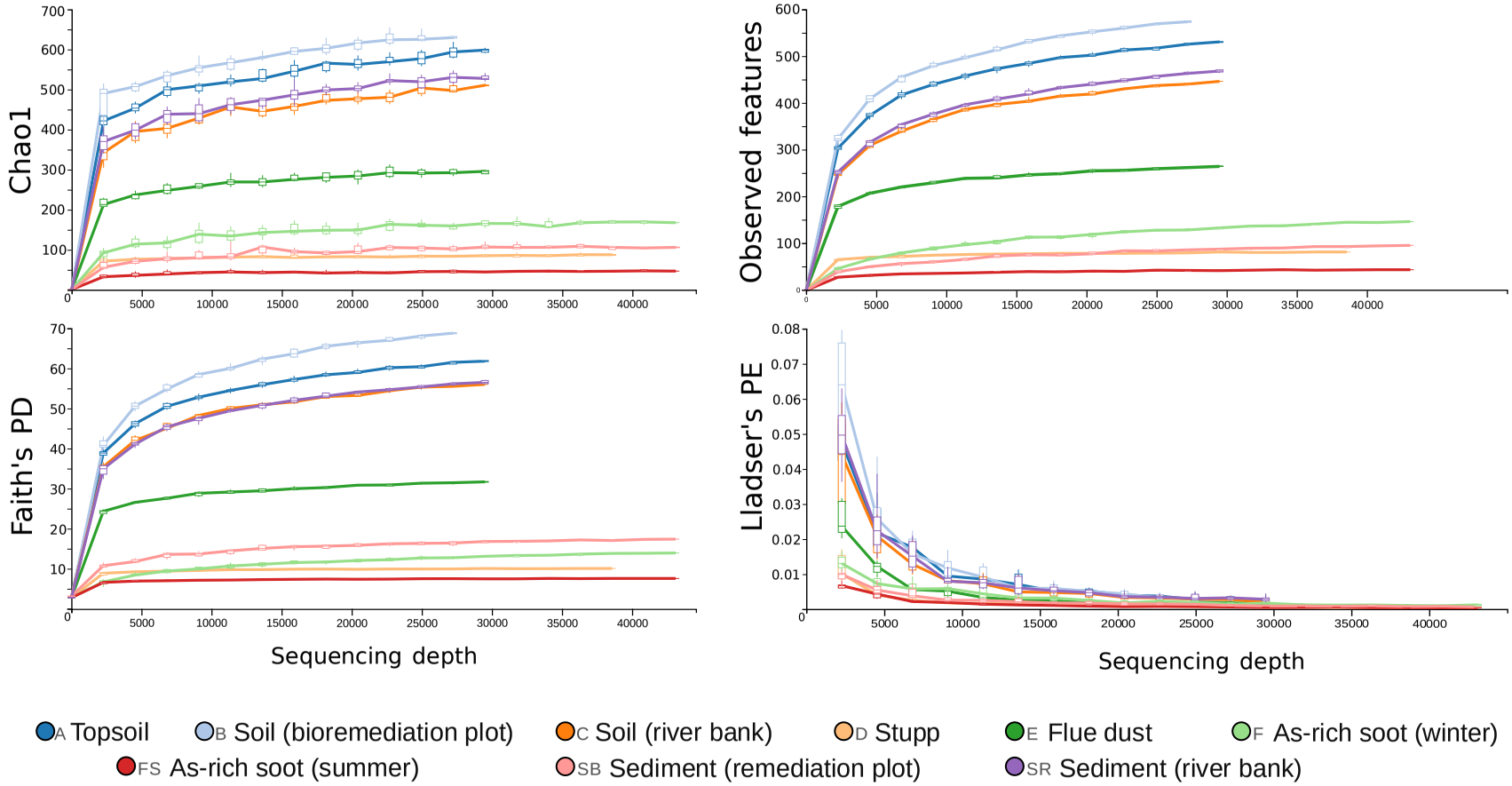
In contrast to 16S rRNA gene libraries, smaller number of unique sequences (1081, compared to 4067 for 16S libraries) allowed sufficient sequencing depths to almost completely flatten out rarefaction curves for all samples (Chao1,Faith's PD and Observed features metrics). Lladser's point estimate indicated that species with abundances as low as $10^{-3} - 10^{-4}$ were within sequencing depth (Figure 22).

ACE, Chao1, Faith's PD and Observed features metrics indicated much lower biodiversity in samples of stump (D), soot (F, FS) and groundwater sediments from the river bank well (SR) compared to samples of soil (A, B &C) and the other groundwater sediments sample (SB). Alpha-diversity of Fungi and SAR in flue dust (E) was higher, but did not reach soil diversity levels (Figures 23a, b, d & e).

Shannon-Wiener index displayed a similar pattern (Figure 23g), with higher alpha-diversity values for all samples, but especially for flue dust (E). In contrast, ENS_{PIE} metrics (23c) indicated that diversity level in river bank soil (C) and groundwater sediments (SB) was actually lower than in other soil samples (A&B), and comparable to the flue dust (E), suggesting an uneven community structure where a relatively small number of organisms dominate the sample.

Pielou's J and Simpson's index suffered from large sample sizes as with Bacterial and Archaeal community profiles, albeit to a smaller degree. They indicated lower biodiversity in soot (F, FS), stupp (D) and river bank well sediments (SR) compared to soils (A, B & C).

Figure 22.: Fungi and SAR. Rarefaction curves for Chao1 index, Faith's PD, Observed Features and LLadser's PE.



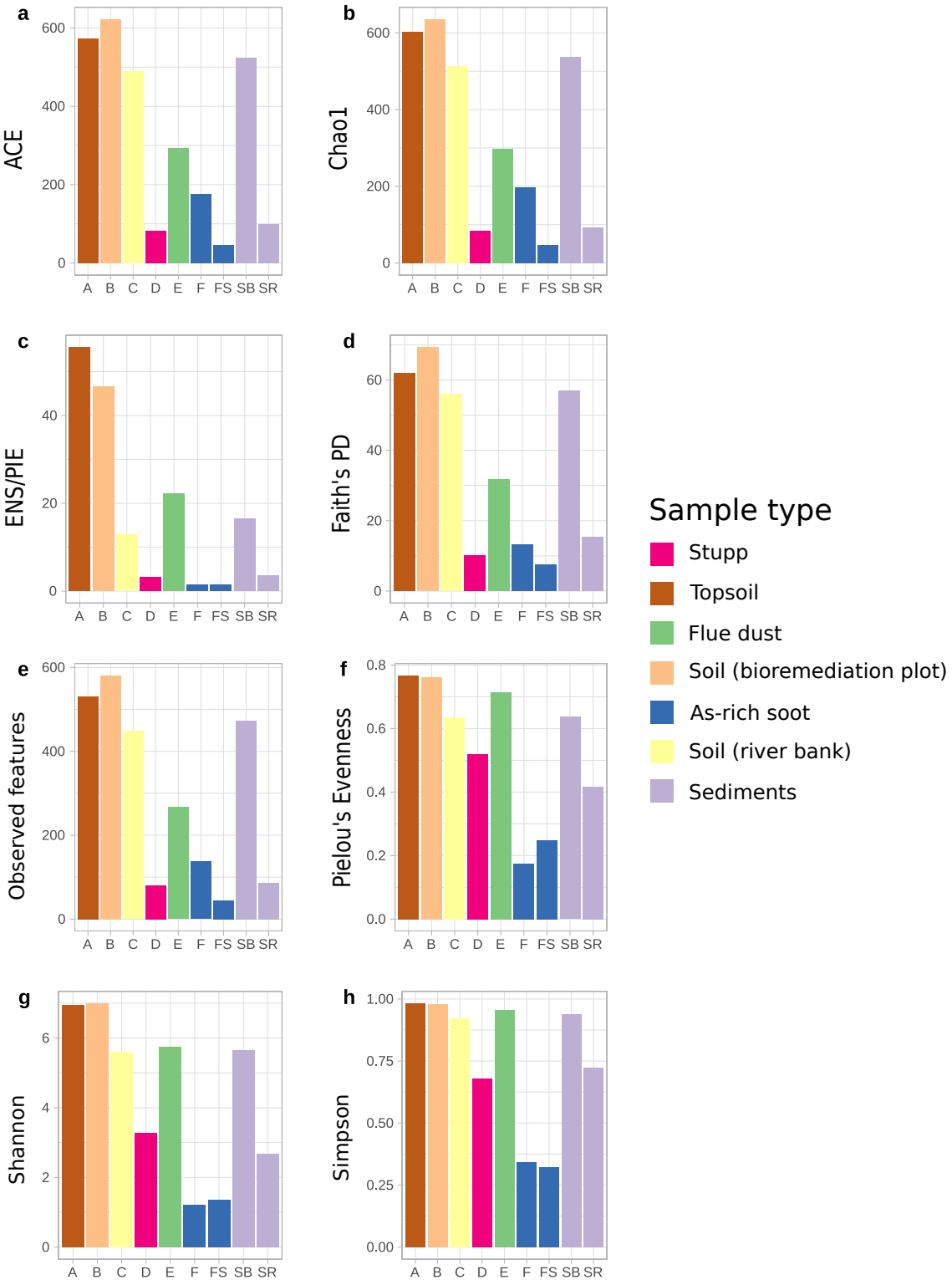


Figure 23.: Alpha-diversity indices for Fungal and SAR communities. a: Abundance-based Coverage Estimator; b: Chao1 index; c: effective number of species / probability of intra- or interspecific encounter; d: Faith's Phylogenetic Diversity; e: observed features; f: Pielou's evenness; g: Shannon-Weaver index; h: Simpson's index of diversity.

5.2.3. Beta-diversity

Bacteria and Archaea

Significant differences between microbial communities from different samples were observed (ANOSIM $R = 0.639$, $p = 0.001$). On a PCoA plot, the biggest observed differences were between stupp (D) and the rest of the samples. Topsoil, soil from phytoremediation plot and soil from the river bank (sample series A, B & C) were grouped together; distances between sediment samples (SR & SB) were comparable to distances to other samples, and samples of soot taken at different times (F & FS), while not completely dissimilar, still had significant differences between each other. Flue dust samples (E), while having some similarities to soil (A, B & C), soot (FS) and sediment (SR), were distinct enough to form their own grouping (Figure 24).

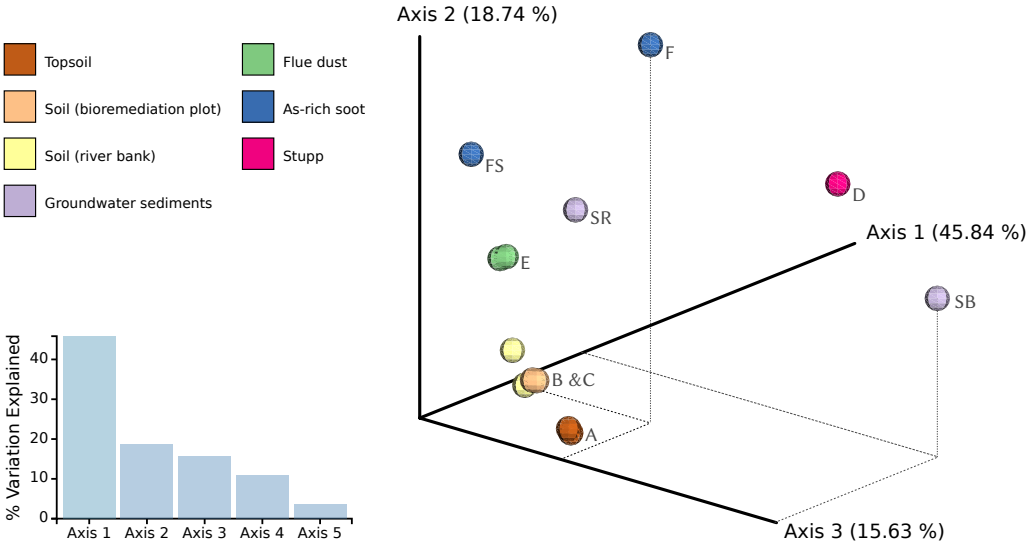


Figure 24.: Principal coordinates analysis (PCoA) 3D plot of variance-adjusted weighted UniFrac distance matrix generated from rarefied taxon abundances for Bacterial and Archaeal communities of each sample, and depicting patterns of Prokaryotic beta-diversity. Points that are closer together are more similar; the numbers in parentheses represent percentage of variation explained by each axis. A, B & C: soil samples; E: flue dust samples; D: stupp sample; F: soot sample taken in the winter; FS: soot sample taken in the summer; SR: groundwater sediments from the well on the river bank; SB: groundwater sediments from the well near the phytoremediation plot. Additional barplot illustrates contribution of each principle coordinate axis to variation between samples and includes axes not shown on the PCoA plot.

According to the ADONIS test , increase in mercury and arsenic concentrations across different samples explained 35.4% of variation in UniFrac distances between samples for mercury ($R^2 = 0.3542$, $p = 0.008$) and 17.9% for arsenic ($R^2 = 0.1794$, $p = 0.03$). Differences in sample type could explain 84% of variation ($R^2 = 0.8399$, $p = 0.0001$).

Fungi and SAR

Differences between samples grouped into soils (A, B & C), sediments (SR & SB), flue dust (E) and high-grade waste (D, F & FS) were significant (ANOSIM $R = 0.813,\ p = 0.007$). On a PCoA plot, soil samples (A, B & C), flue dust (E) and sediments (SR & SB) formed a pattern similar to that of Prokaryotic communities, with soils grouped together, flue dust being somewhat similar to those in soils, and sediments (SR & SB) being dissimilar to both other samples and each other. In stark contrast to the Prokaryotic communities, fungal and SAR communities in samples of stupp (D) and arsenic-rich soot (F, FS) were very similar to each other (Figure 25).

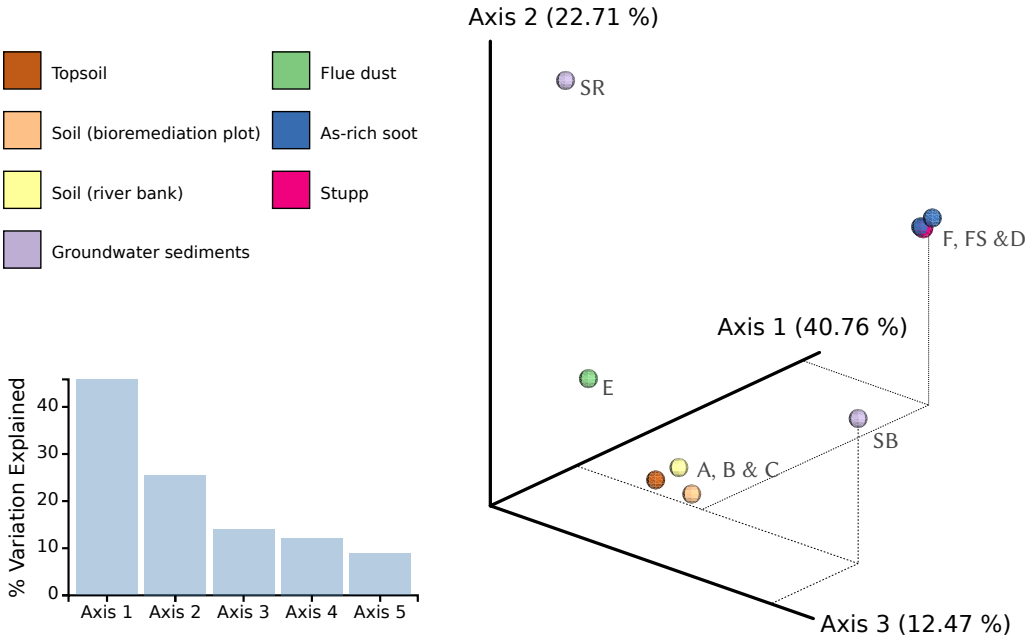


Figure 25.: Principal coordinates analysis (PCoA) 3D plot of variance-adjusted weighted UniFrac distance matrix generated from rarefied taxon abundances for Fungi and SAR communities of each sample, and depicting patterns of their beta-diversity. Points that are closer together are more similar; the numbers in parentheses represent percentage of variation explained by each axis. A, B & C: soil samples; E: flue dust sample; D: stupp sample; F: soot sample taken in the winter; FS: soot sample taken in the summer; SR: groundwater sediments from the well on the river bank; SB: groundwater sediments from the well near the phytoremediation plot. Additional barplot illustrates contribution of each principle coordinate axis to variation between samples and includes axes not shown on the PCoA plot.

Higher arsenic concentration in the flue dust (E), stupp (D), soot (F, FS) and sediment (SR, SB) samples compared to the soils (A, B &C) could explain 23% of variation in UniFrac distances (ADONIS test; $R^2 = 0.2338,\ p = 0.045$). Separate from it, increasing mercury levels across different samples could explain 25% of variation ($R^2 = 0.2539,\ p = 0.044$). Differences in sample type, however, explained 71% of inter-group UniFrac distance variation with higher degree of confidence ($R^2 = 0.7125,\ p = 0.004$).

5.2.4. Taxonomic analysis

Bacteria and Archaea

Annotation depth was high, with majority of ASVs receiving genus-level annotation (Table 12).

Table 12.: Annotation depth for Bacterial and Archaeal communities.

Taxonomic level	Domain	Phylum	Class	Order	Family	Genus
% of annotated ASVs	100	99.95	99.66	91.81	87.19	75.81

Prokaryotic communities in different soil samples had very similar taxonomic composition. Majority of organisms belonged to the phyla Proteobacteria, Actinobacteria and Acidobacteria. Bacteroidetes and mostly-uncultured Chloroflexi, Planctomycetes, Gemmatimonadetes and Nitrospirae were also present. *Archaea* were represented by Thaumarchaeota. Members of the phylum Firmicutes were a minority, albeit a large one. Flue dust (E) communities were similar in composition, but had higher proportion of Proteobacteria at the expense of other phyla, as well as higher proportion of Verrucomicrobia and Cyanobacteria. *Archaea* were a minority there. Groundwater sediments (SR & SB), while also dominated by Proteobacteria, differed significantly from the soils and between themselves. Phylum-level composition of bacterial communities in the groundwater sediment samples SB and SR was somewhat similar to soil samples, but had a much higher proportion of Thaumarchaeota and Gemmatimonadetes, a small minority of Ignavibacteriae in sample SB, and a large amount of Elusimicrobia and Ignavibacteriae in sample SR.

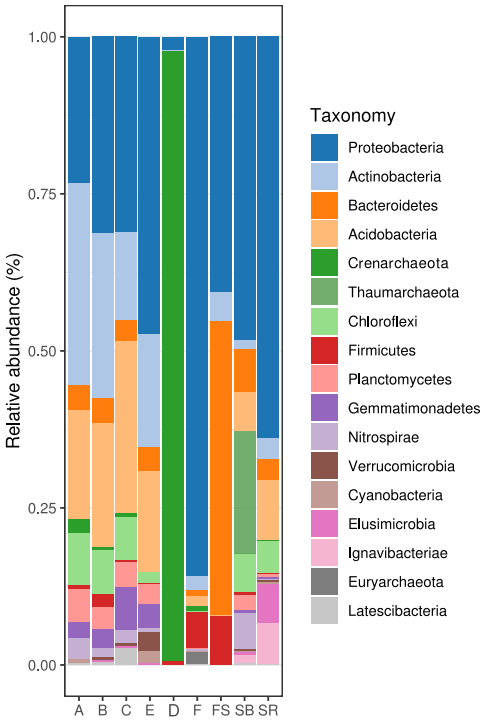


Figure 26.: Barplot of the relative abundance of *Bacteria* and *Archaea* at the phylum level. Replicates are merged.

Samples of stupp (D) and arsenic-rich soot (F & FS) were the most unusual. Stupp sample community consisted almost entirely of *Archaea* of the phylum Thaumarchaeota, with minor bacterial populations, while sample F was dominated by Proteobacteria with minor communities of Firmicutes and Euryarchaeota. Microbial community of the

sample FS was comprised by Proteobacteria and Bacteroidetes, with some Firmicutes and Actinobacteria (Figure 26).

On the lower taxonomic levels, similarities between soil samples (A, B & C) and the flue dust samples (E) persisted (Figures 27 and 28).

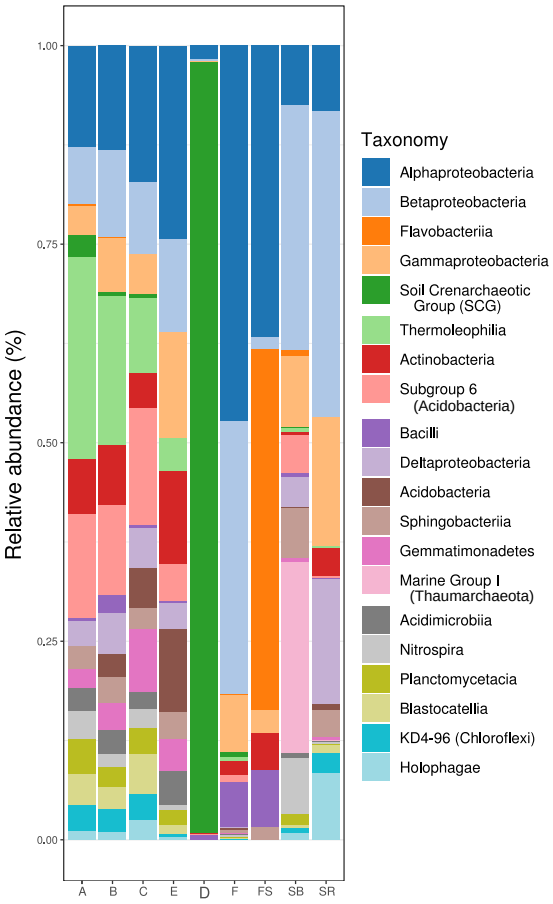


Figure 27.: Relative abundance of *Bacteria* and *Archaea* at the class level. 20 most abundant classes are shown; replicates of the samples are merged.

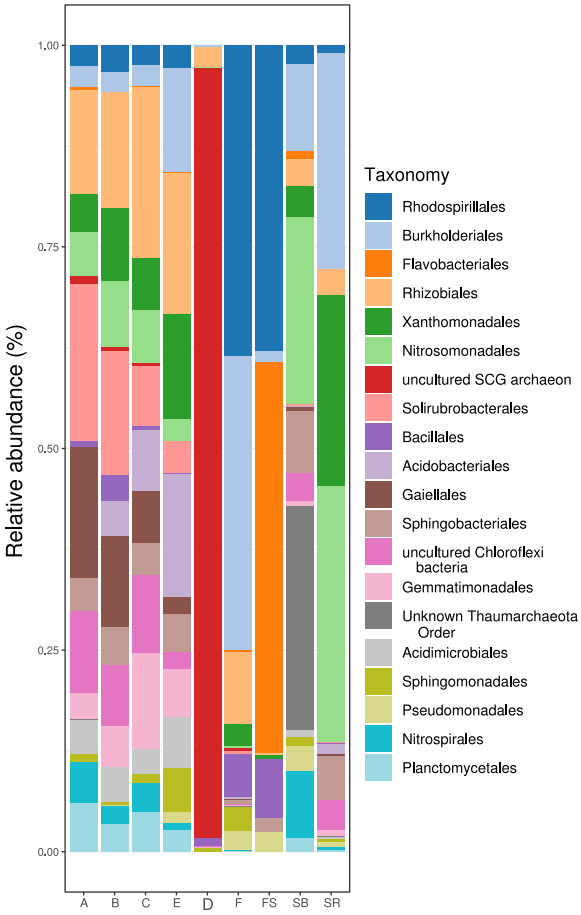


Figure 28.: Relative abundance of *Bacteria* and *Archaea* at the order level. 20 most abundant orders are shown; replicates of the samples are merged.

Bacterial communities there consisted of:

- Class Alphaproteobacteria, with members of the orders Rhizobiales, Rhodospirillales, Sphingomonadales, Rhodobacteriales and Caulobacteriales being the most abundant;
- Betaproteobacteria which consisted mainly of Nitrosomonadales and Burkholderiales, as well as unidentified uncultured bacteria;
- Gammaproteobacteria with Xanthomonadales being the most abundant,

and Deltaproteobacteria which was comprised mainly of Myxococcales and Desulforellales;

- Phylum Actinobacteria that was represented by the members of the orders Solirubrobacteriales and Gaiellales of the class Thermoleophilia, various members of the class Actinobacteria, and order Acidimicrobiales of the class Acidimicrobiia;
- Acidobacteria, with majority of them belonging to Acidobacteria Subgroup 6 together, Acidobacteriales and Blastocatellales;
- Firmicutes, represented mostly by the Bacilli;
- Phylum Bacteroidetes, with Sphingobacteriales and Cytophagales of classes Sphingobacteriia and Cytophagia, respectively, being the most abundant;
- Various members of Nitrospira, Planctomyceta, Gemmatimonadetes, Chloroflexi and other (mostly) uncultured bacteria;
- *Archaea* belonging to the Soil Crenarchaeotic Group.

Groundwater sediments samples differed significantly from each other and the rest of the samples. While both had large proportion of Burkholderiales, Nitrosomonadales and Sphingobacteriales, sample SR also had a large proportion of Thaumarchaeota of the Marine Group I and Nitrospirales, while sample SB had higher proportion of Xanthomonadales and Burkholderiales. On the ASV level, there was very little similarity, with samples SB and SR sharing just 26 ASVs with counts $\geqq 10$ (belonging mostly to Proteobacteria) and only 3 ASVs with counts $\geqq 75$. Two of those ASVs belonged to the Gallionellaceae family of the order Nitrosomonadales and one - to the family Comamonadaceae of the the order Burkholderiales.

In stupp (D), just seven ASVs belonging to the Soil Crenarchaeotic Group (SCG) archaeons of the phylum Crenarchaeota[6] made up the majority of prokaryotic organisms in the sample.

Soot sample taken in the winter (F) was populated mainly by Betaproteobacteria of the class Burkholderiales and Alphaproteobacteria of the classes Rhodospirales, Rhizobiales and Sphingomonadales. There were large minority populations of Gammaproteobacteria (represented chiefly by Xanthomonadales and Pseudomonadales) and Bacillales.

Microbial communities in the soot sample taken in the summer (FS) were dominated by the orders Rhodospirillales and Flavobacteriales of classes Alpha- and Betaproteobacteria, respectively. Other bacteria in the sample included Bacillales, Sphingomonadales, Pseudomonadales and Burkholderiales.

[6] Alternativele classifications: phylum *Thaumarchaeota* or superphylum TACK.

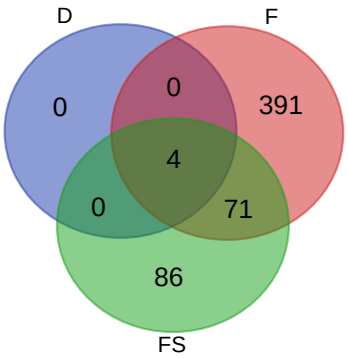


Figure 29.: Venn diagram showing the distribution of ASVs shared between stupp (D) and As-rich soot sampled in the winter (F) and summer (FS). Only ASVs with counts $\geqq$ 10 were included.

There was a significant overlap between populations in the sample FS and sample F, with 75 shared ASVs after filtering out low-abundance (count $\leqq$ 10) features (Figure 29). Most abundant of the shared ASVs belonged to genera *Bacillus*, *Acidocella*, *Burkholderia*, and unknown Comamonadaceae bacteria. Four ASVs were shared between samples F, FS and D: one belonged to *Stenotrophomonas sp.* (of family Xanthomonadaceae, order Xanthomonadales, Gammaproteobacteria), two ASVs belonged to the genus *Bacillus*, and one – to archaeaon of the Soil Crenarchaeotic Group.

Fungi, SAR (and Protozoa)

Annotation depth, compared to the 16S rRNA data, was much lower, with only half of ASVs receiving Class-level annotations (Table 13).

Table 13.: Annotation depth for Fungal, SAR and Protozoan communities.

Taxonomic level	Kingdom	Subkingdom	Infrakingdom	Phylum	Subphylum	Class	Order	Family	Genus
% of annotated ASVs	100	66.79	66.05	65.77	51.53	51.34	37.56	26.68	15.26

On the kingdom and subkingdom levels, soils (A, B & C), flue dust (E) and groundwater sediments from the phytoremediation plot well (SB) were populated predominantly by Fungi of subkingdoms Dikarya (Neocmycota) and Eomycota. Chromista (SAR) and Protozoa of the Sarcomastigota subkingdom were also present. In the groundwater sediment sample from the river bank well (SR) *Fungi* were a minority, while *Chromista* predominated. In contrast, high-grade waste samples (D, F and FS) were populated almost exclusively by Fungi of the subkingdom Neomycota (Figure 30). SAR microorganisms were predominantly Rhizaria; Halvaria was a large minority in the groundwater sediment sample SB; Alveolata was a minor clade in all samples.

On the phylum level, similarity between soil samples (A, B & C) and flue dust (E) persisted: fungal populations were represented mostly by Ascomycota, Basidiomycota

and Zygomycota; Rhizaria (of the SAR) were represented by the phylum Cercozoa, and Sarcomastigota (of Protozoa) – by Choanozoa. There were some differences as well, with more-contaminated river bank soil (C) and flue dust (E) samples almost lacking populations of Glomeromycota and Chytridiomycota present in the other soil samples (A & B), and having higher relative abundance of Zygomycota fungi. Populations in stupp (D) and soot (F, FS) were comprised by Ascomycota and Basidiomycota fungi and minority population of Choanozoa. Sediment sample from the phytoremediation plot well (SB) had abundant fungal populations including Ascomycota, Basidiomycota, Chytridiomycota and Zygomycota, as well as members of the phylum Bigyra of the infrakingdom Halvaria (SAR) and Choanozoa protists, while populations in the sediment sample from the river bank well (SR) was comprised mainly of Cercozoa and Choanozoa (Figure 31).

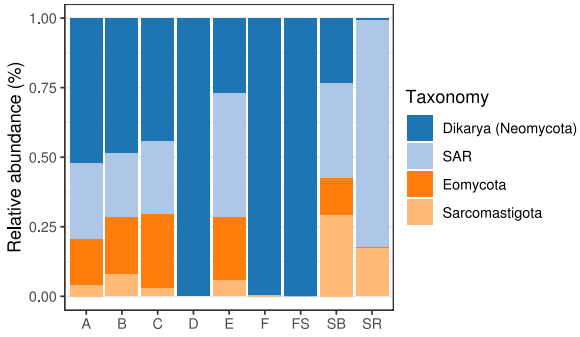


Figure 30.: Relative abundances of *Fungi*, *SAR* and *Protozoa* at the subkingdom level.

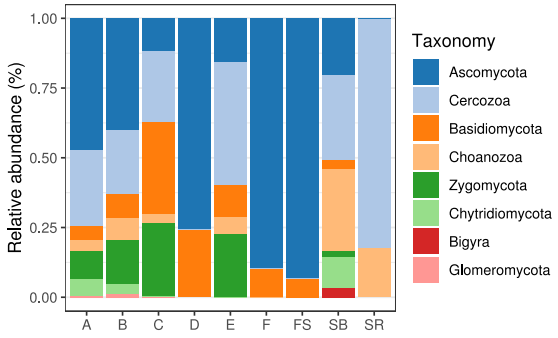


Figure 31.: Relative abundances of *Fungi*, *SAR* and *Protozoa* at the phylum level.

Taxonomic identification and relative abundances comparisons on the lower taxonomic levels were much less reliable as almost half of the ASVs below phylum level remained unassigned. Among identified fungi in soils (A, B & C) Dothideomycetes, Eurotiomycetes, Sordariomycetes, Leotiomycetes and Pezizomycetes (of the phylum Ascomycetes), Agaricomycetes (of the phylum Basidiomycota) and Mortierellales (of Zygomycota) were the most abundant (Figure 32). Flue dust (E) had more fungi belonging to subphylum Entomophthoromycotina (of the phylum Zygomycota) and Ustilagnomycetes (of Basidiomycota), and smaller populations of Leotiomycetes and Pezizomycetes compared to the soil samples. Predominant populations of fungi in stupp (D) and soot (F, FS) samples consisted of Dothideomycetes of the phylum Ascomycetes, and Entomophthoromycotina (of Zygomycota) and Ustilagnomycetes (of Basidiomycota). In the soil, flue dust and sediment samples Dothideomycetes were almost equally divided into non-identified Dothideomycetes fungi, Pleosporales, Botryosphaeriales, Capnoidales and Venturiales; in contrast, in stupp (D) and soot (F, FS) samples non-identified Dothideomycetes ASVs that constituted a very small minority in

the rest of the samples were much more abundant than the rest of Dothideomycetes.

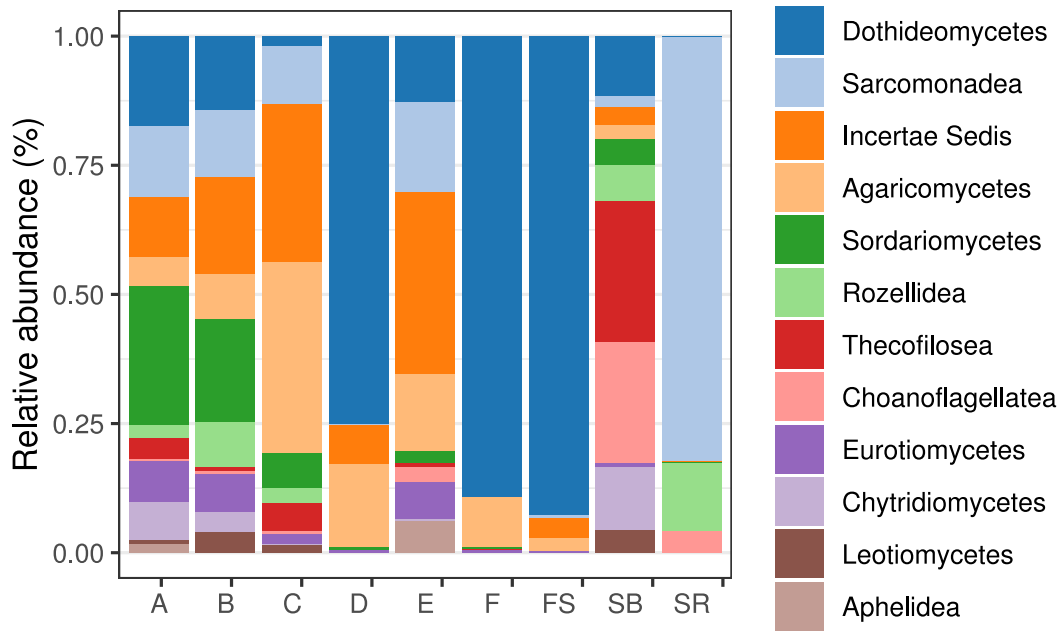


Figure 32.: Relative abundances of identified *Fungi*, *SAR* and *Protozoa* at the class level (most abundant features).

SAR populations in all samples were represented mostly by the members of the subphylum Monadofilosa; and while in soil samples, flue dust and sediment sample SR majority of them belonged to the class Sarcomonadea (*Heteromita* and *Cercomonas* of the orders Glissomonadida and Cercomonadida, respectively), in the sediment sample SB they belonged mostly to the genus *Rhogostoma* of the family Rhizaspididae, order Cryomonadida, class Thecofilosea.

Most abundant protozoans in the soil samples (A, B &C) and groundwater sediments (SR, SB) belonged to *Craspedifa* of the class Choanoflagellatea, subphylum Choanofila and *Rozellida* of the class Rozellidea, subphylum Paramycia (of phylum Choanozoa both). In contrast to that, in the flue dust (E), Aphelidida of the class Aphelidea, subphylum Paramycia was more abundant.

5.2.5. Functional metagenome prediction

Three metagenomic profiles containing genes and their predicted abundances for each ASV were generated:

- Enzyme Commission numbers (EC), with 2330 predicted entries;
- Clusters of Orthologous Groups (COG), with a total of 4449 predicted COGs;
- Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology, with 7569 KEGG orthologues.

EC numbers were used only for prediction of MetaCyc pathways. COG profiles were used for validation of PICRUSt2 predictions; analysis of incidence and abundance of arsenic and mercury resistance genes was carried out on the KEGG Orthology (KO) dataset.

Validation of PICTUSt2 predictions

Validation of genomic predictions was carried out by comparing COG profiles generated by PICRUSt2 to COG-annotated genomes of thirteen strains isolated from the El Terronal site that had V4-V5 16S rRNA gene sequences matching their corresponding entries in the ASV table (See Section 5.3 and Section 5.11 for details on strain isolation and genome annotation). As a group, comparison of abundances of COGs using White's non-parametric t-test with Šidák correction of p-values indicated that only COG1145 (4Fe-4S ferredoxin, iron-sulfur binding domain protein) was differentially abundant ($p = 0.02867$). On the level of individual genomes, statistically significant differences between genome predictions and sequenced genomes were low in number as well (Table 14).

Table 14.: Comparison of sequenced genomes to their PICRUSt-predicted counterparts using G-test with Yates' correction and Šidák correction of p-values. Strain: a name of the isolated strain used for comparison. Diff. abundant genes: nuber of differentially abundant genes ($p < 0.05$) for each predicted vs. sequenced comparison.

Strain	C19	C22	C110	F12	F14	F21	SB21	SB22	SR1	SR3	WB11	WB17	WR7
Diff. abundant genes	1	5	2	0	0	0	2	0	0	0	0	2	2

Analysis of predicted metagenomes

Principal Component Analysis has of predicted abundances of MetaCyc pathways indicated clear differences in metabolism of prokaryotic communities inhabiting different environments at El Terronal. Soil samples (sample series A, B & C) were predicted to house prokaryotic communities with very similar metabolic profiles. Flue dust samples (E1 & E2) were somewhat similar to the soils, and also had higher inter-sample variation. Predicted metabolic pathways of populations from two soot samples (winter: F & summer: FS) were substantially different from both other samples and each other. Interestingly, while PCoA and analysis of taxonomy clearly demonstrated differences in microbial communities inhabiting two sediment samples (SR & SB), EC-based prediction of MetaCyc pathways suggested that overall metabolism of their respective prokaryotic communities was fairly similar (Figure 33). Predicted metabolic pathways of *Archaea*-dominated populations of the stupp sample (D) differed substantially from the rest of the samples (not shown on the PCA chart); however, due to uncertainty in the veracity of PICRUSt2 predictions for *Archaea* stemming from much smaller number of available genomes to draw predictions from compared to *Bacteria*, it was excluded from the analysis.

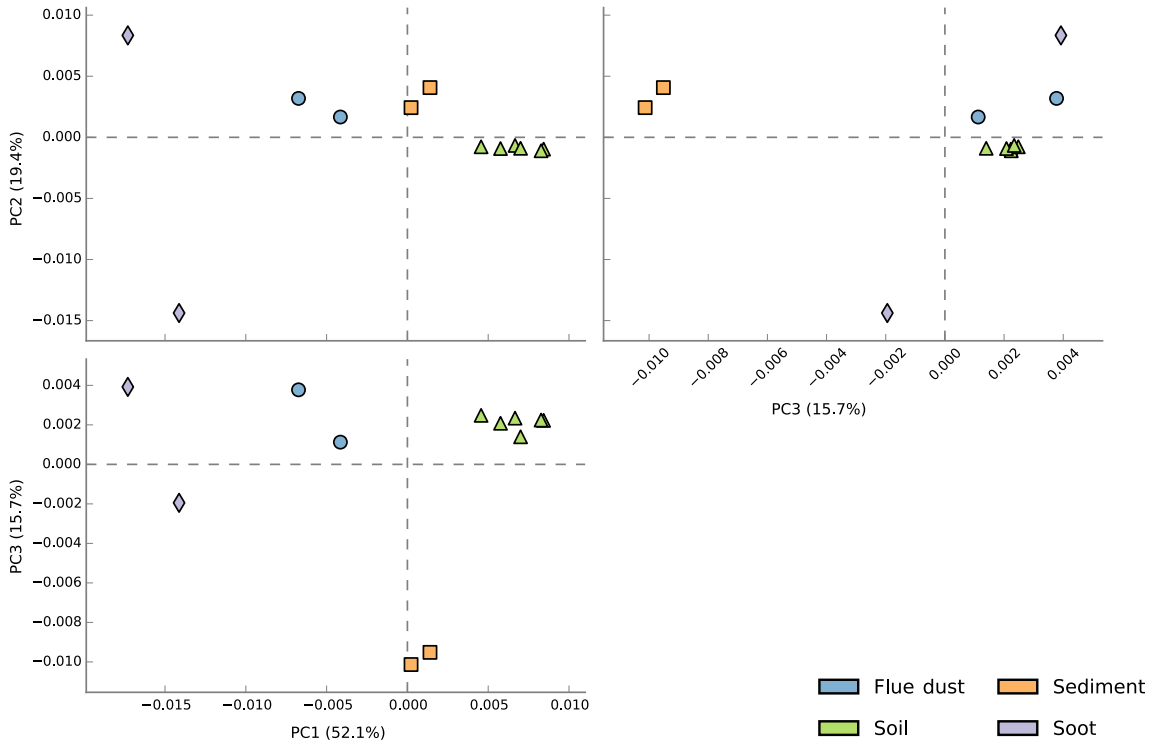


Figure 33.: Principal Component analysis of MetaCyC pathways predictions generated by PICRUSt2 from rarefied taxon abundances for bacterial and archaeal communities. Points that are closer together are more similar. Soil samples include A1, A2, B1, B2, & C1, C2; flue dust samples include E1 and E2; soot samples are F & FS; groundwater sediments include samples SB and SR. Stupp sample (D) is excluded from the analysis.

To better understand the differences between prokaryotic populations of sediment samples, their predicted metabolic profiles were compared directly. Most significant differences in pathway abundances included higher abundance of flavin biosynthesis pathway II and phosphopantothenate biosynthesis III (typical of *Archaea*; supported by the much higher proportion of *Thaumarchaeota* in the community), as well as higher abundance of pyridoxal 5'-phosphate biosynthesis I pathway and CMP-legionaminate biosynthesis I pathway. Prokaryotic communities in the sediment sample SR had higher proportion of genes encoding UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis superpathway typical for Gram-negative bacteria (Figure 34).

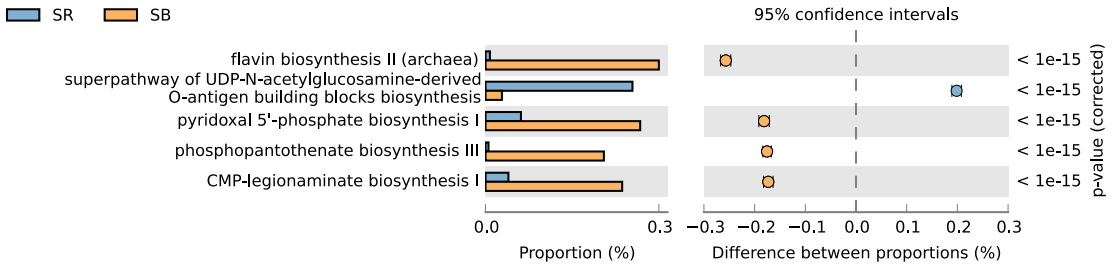


Figure 34.: Main differences in predicted MetaCyc pathway abundances for groundwater sediment samples SR and SB.

Analysis of predicted arsenic and mercury resistance genes

Predicted abundances of KEGG orthologues corresponding to arsenite efflux pumps *acr3* and *arsB* (K03325 and K03893, respectively), arsenite pump ATPase *arsA* (K01551), three subtypes of arsenate reductase *arsC* (K00537, K03741 and K18701), organoarsenical oxidase *arsH* (K11811), small and large subunits of arsenite oxidase *aio* (K08355 and K08356, respectively) and *arsR* regulator gene (K03892) were compared between four groups: soils (sample series A, B & C), flue dust (sample series E), groundwater sediment (SR & SB) and arsenic-rich soot (F & FS). Stupp sample (D) was excluded due to uncertainty in the veracity of PICRUSt2 predictions for *Archaea* and lack of replicates necessary for ANOVA.

All arsenic resistance genes represented only a small proportion of predicted genes (typically between 0.1 and 0.5%; up to 0.11% for regulatory gene *arsR*). Four genes were differentially abundant ($p < 0.05$). Arsenite pump *arsB* was predicted to be more abundant in flue dust (E) compared to soils (A, B & C) and sediments (SR & SB), and even more abundant in soot samples F & FS (Figure 35A) despite phylogenetic differences in their prokaryotic communities, while more common *acr3* pump was not differentially abundant. Arsenate reductase *arsC* and organoarsenical oxidase *arsH* were significantly more abundant in soot (F & FS) as well, with flue dust (E), sediments (SR & SB) and soils (A, B & C) showing progressively less predicted abundance (Figures 35B and 35C). Predicted abundance of regulator/repressor genes of *arsR* type, in contrast, was significantly lower in soot (F & FS) compared to other samples, with soil samples (A, B & C) having the highest abundance, followed by the flue dust (E) and groundwater sediment samples SR & SB (Figure 35D).

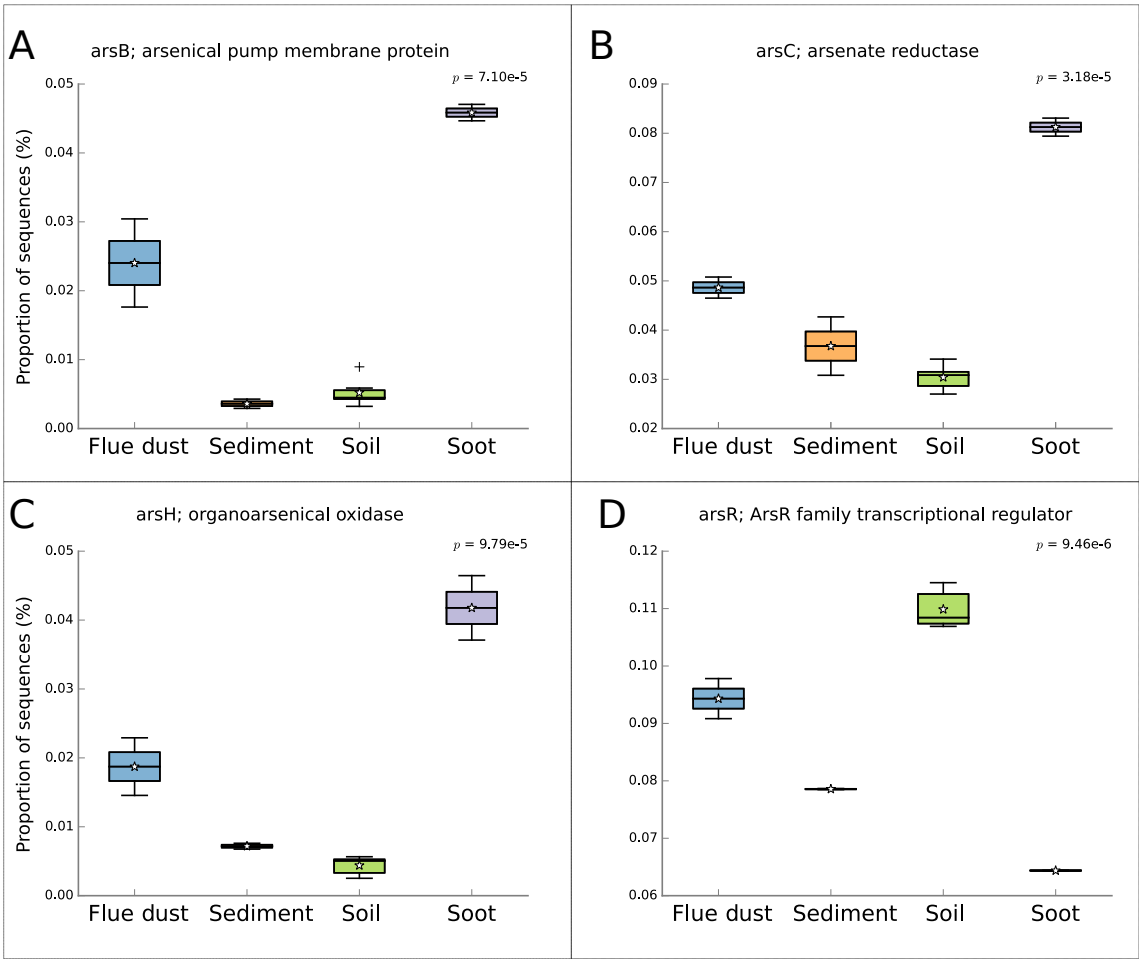


Figure 35.: Predicted arsenic resistance genes differentially abundant between sample groups of soil (samples A1&2; B1&2; C1&2), flue dust (E1&2), groundwater sediments (SR&SB) and arsenic-rich soot (F&FS): arsenite pump *arsB* (A), arsenate oxidase *arsC* (B), organoarsenical oxidase *arsH* (C) and *arsR*-type regulator/repressor gene (D). Significance of differences between groups established by ANOVA.

Predicted abundances of KEGG orthologues corresponding to mercury(II) oxidase merA (K00520), alkylmercury lyase merB (K00221), mercuric ion carriers (transport proteins) merE, merP and merT (K19059, K08364 and K08363, respectively), as well as regulatory proteins merR (K08365) and merD (K19057) were compared between sample groups. Unlike predictions for arsenic resistance genes, soot samples taken at different time (F & FS) have shown significant inter-sample variation exceeding inter-group differences. Comparison was thus made for three remaining groups: soils (sample series A, B & C), flue dust (sample series E) and groundwater sediment (SR & SB). Stupp sample (D) was excluded due to uncertainty in the veracity of PICRUSt2 predictions for *Archaea* and lack of replicates necessary for ANOVA similarly to the previous analysis.

Proportion of predicted mercury resistance genes relative to the rest of the predicted metagenomes was very low (typically less than 0.01%). Four genes were predicted to

be differentially abundant across all three groups ($p < 0.05$): alkylmercury lyase *merB* (Figure 36A), mercuric ion transport protein *merC* (Figure 36B), mercuric ion transport protein *merE* (Figure 36C), and mercuric ion transport protein *merT* (Figure 36D). In all cases, predicted abundances were significantly higher in flue dust (E) compared to soils (sample series A, B & C) and groundwater sediments (SR & SB).

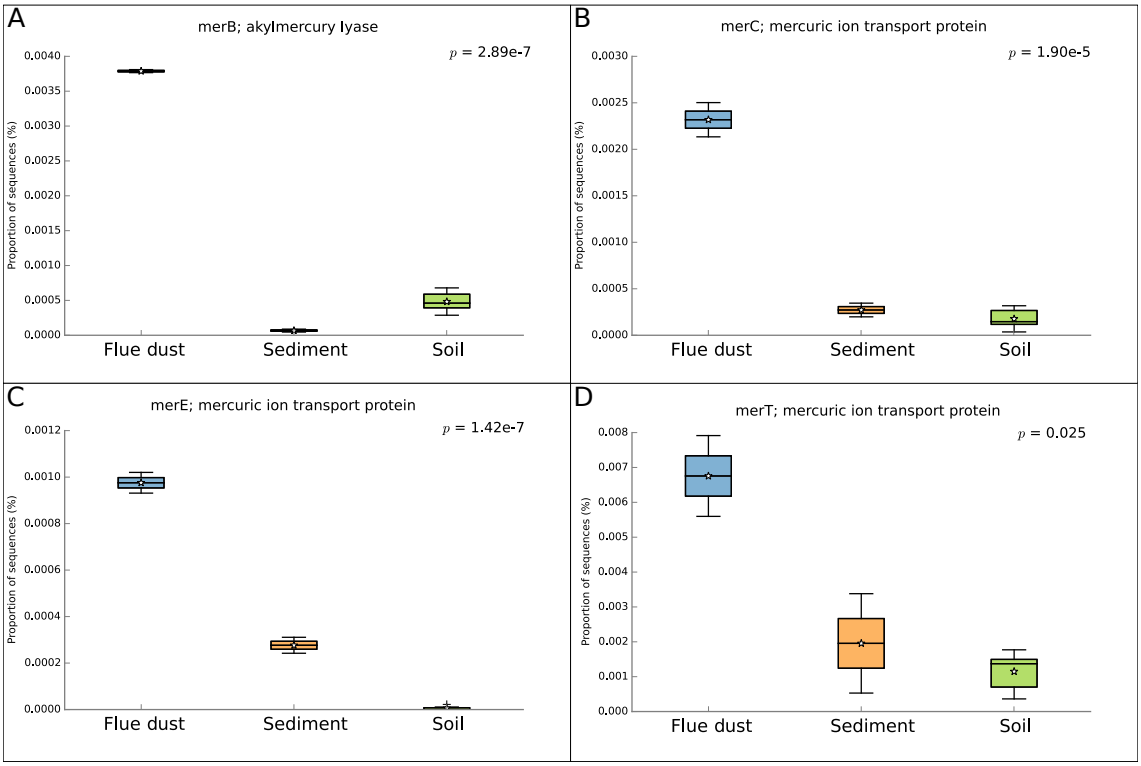


Figure 36.: Predicted mercury resistance genes differentially abundant between sample groups of soil (samples A1&2; B1&2; C1&2), flue dust (E1&2) and groundwater sediments (SR&SB): alkylmercury lyase *merB* (A), mercuric ion transport protein *merC* (B), mercuric ion transport protein *merE* (C), and mercuric ion transport protein *merT* (D). Significance of differences between groups established by ANOVA.

5.3. Seeding and isolation of bacteria

Samples have shown significant differences in microbial populations when seeded on TSA medium. By CFU count, samples of soil (A, B and C) had the highest abundance of cultivable bacteria, as well as the highest degree of apparent diversity as reflected by the number of morphologically distinct colonies selected for isolation and cultivation[7]. CFU count obtained from the samples of soot (F) and flue dust (E) was somewhat lower, and apparent biodiversity was lower than in the soil as well. Groundwater and groundwater sediments had even lower CFU counts, but somewhat higher diversity when compared to soot and flue dust. Additional arsenic-rich soot sample taken in the summer (FS) had only 6 individual colonies appearing in the Petri dish seeded from non-diluted sample extract, suggesting that lower humidity and rainfall in the summer was detrimental to the bacterial growth. Stupp sample had no bacteria capable of growing on TSA medium (Table 15). Isolation was attempted for a total of 64 strains; however, only 59 of them were capable of sustained growth once re-seeded on a fresh TSA plate. Due to isolation and cultivation conditions (TSA medium, 28℃), all isolated strains belonged to chemoorganotroph mesophilic bacteria.

Table 15.: Viable counts (colony-forming units, CFU), and number of isolated and viable strains for each sample. CFU count is an average of three seedings.

Sample	CFUs/g	Cultured strains	Description
A	$9.3 \cdot 10^5$	7	Topsoil
B	$1.07 \cdot 10^6$	4	Soil (phytoremediation plot)
C	$9.68 \cdot 10^5$	10	Soil (river bank)
D	0	0	Stupp
E	$1 \cdot 10^5$	6	Flue dust
F	$3.63 \cdot 10^5$	6	As-rich soot (winter)
FS	Out of countable range	5	As-rich soot (summer)
SR	$7.2 \cdot 10^2$	8	Groundwater sediment (river bank)
SB	$1.6 \cdot 10^4$	3	Groundwater sediment (remediation plot)
WR	$1.29 \cdot 10^2$	4	Groundwater (river bank)
WB	$4.8 \cdot 10^2$	6	Groundwater (remediation plot)

[7]For sample F, three out of six isolated strains had no apparent morphological differences; their isolation and cultivation were attempted regardless of that.

CFU count appeared to decrease with increasing arsenic and mercury concentrations, but it was not statistically significant. This suggests that habitability of those environments was dependent not only on cytotoxic effects of arsenic and mercury, but also on other factors such as aforementioned humidity, chemical composition and availability of nutrients.

5.4. Identification of isolated strains

Isolated strains and their taxonomic affiliation are represented in the Table 16. Additionally, correspondence of taxonomic ranks between 16S rRNA gene-based SILVA128 and whole genome-based SILVA138 databases for isolated strains can be found in 34. The most-closely related species for strains for which clear species-level affiliation (identity above 99% with single corresponding species) could be established are given in the Table 17.

Table 16.: Taxonomic affiliation of isolated bacterial strains: genus-level identification against SILVA128 database (95% genus identity cut-off as per SINA default settings).

Strain	Taxonomy: Phylum; Class; Order; Family; Genus	Identity, %
	A (Topsoil)	
A11	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	100
A12	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	99.64
A13	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*	98.79
A14	Actinobacteria; Actinobacteria; Actinomycetales; Micrococcaceae; *Arthrobacter*	94.83
A15	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	98.56
A18	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.6
A21	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.72
	B (Soil, mixed and tilled)	
B11	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	99.44
B13	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	99.64
B14	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	99.71
B15	Actinobacteria; Actinobacteria; Micromonosporales; Micromonosporaceae; *Micromonospora*	100

Table 16.: Taxonomic affiliation of isolated bacterial strains: genus-level identification against SILVA128 database (95% genus identity cut-off as per SINA default settings).

Strain	Taxonomy: Phylum; Class; Order; Family; Genus	Identity, %
	C (Soil of the river bank)	
C11	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	99.52
C12	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	96.7
C14	Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Pedobacter*	97.89
C17	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	98.78
C18	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	99.51
C19	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	99.4
C110	Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Mucilaginibacter*	97.2
C111	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	98.58
C22	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	100
C23	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.72
	E (Flue dust)	
E11	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Novosphingobium*	99.71
E12	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.92
E13	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	99.64
E14	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	97.55
E15	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.86
E16	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	98.78
	F (Arsenic-rich soot, winter)	

Table 16.: Taxonomic affiliation of isolated bacterial strains: genus-level identification against SILVA128 database (95% genus identity cut-off as per SINA default settings).

Strain	Taxonomy: Phylum; Class; Order; Family; Genus	Identity, %
F11	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.92
F12	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	100
F14	Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus	99.52
F21	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.92
F22	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.69
F24	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	99.77
	FS (Arsenic-rich soot, summer)	
FDC1	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.3
FDC2	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.44
FS31	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	99.83
FS32	Actinobacteria; Actinobacteria; Propionibacteriales; Nocardioidaceae; *Nocardioides*	98.91
FS34	Actinobacteria; Actinobacteria; Corynebacteriales; Mycobacteriaceae; *Mycolicibacterium*	98.67
	SR (Groundwater sediment, river bank)	
SR1	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.29
SR3	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.41
SR4	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.81
SR5	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	99.61
SR71	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	97.75
SR72	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	99.57
SR73	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	99.5
SR74	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*	98.29
	SB (Groundwater sediment, remediation plot)	

Table 16.: Taxonomic affiliation of isolated bacterial strains: genus-level identification against SILVA128 database (95% genus identity cut-off as per SINA default settings).

Strain	Taxonomy: Phylum; Class; Order; Family; Genus	Identity, %
SB1	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Delftia*	100
SB21	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	100
SB22	Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Chryseobacterium*	99.1
	WR (Groundwater, river bank)	
WR2	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	99.57
WR5	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	99.37
WR6	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	97.71
WR7	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	99.93
	WB (Groundwater, remediation plot)	
WB1	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	97.73
WB11	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	100
WB15	Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Flavobacterium*	99.53
WB17	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	99.22
WB21	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	98.88
WB22	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Sphingobium*	99.29

Table 17.: Identification of bacterial strains against NCBI RefSeq database. Only strains with clear species-level affiliation (99% identity cut-off against single corresponding species with standing in the nomenclature) are given.

Strain	Most closely-related species (NCBI RefSeq)	Identity, %	Reference
A11	*Bacillus mycoides* NBRC 101228	99,72	NR_113990.1
A14	*Arthrobacter pascens* DSM 20545	99,85	NR_026191.1
B11	*Streptomyces canus* CSSP527	99,72	NR_043347.1
B15	*Micromonospora tulbaghiae* TVU1	100	NR_116241.1
C17	*Arthrobacter globiformis* JCM 1332	99,14	NR_112192.1
C18	*Variovorax ginsengisoli* Gsoil 3165	99,51	NR_112562.1
C19	*Arthrobacter globiformis* JCM 1332	99,72	NR_112192.1
C22	*Variovorax boronicumulans* NBRC 103145 BAM-48	99,93	NR_041588.1
E11	*Novosphingobium fuchskuhlense* FNE08-7	99,85	NR_118270.1
E15	*Paraburkholderia fungorum* LMG 16225	99,86	NR_025058.1
E16	*Arthrobacter subterraneus* CH7	98,99	NR_043546.1
F12	*Paraburkholderia fungorum* LMG 16225	100	NR_025058.1

Members of the phylum Proteobacteria were the most numerous (21 strains), followed closely by Actinobacteria (19) and Firmicutes (15 strains). Bacteroidetes were a minority, with only 4 strains (Figure 37).

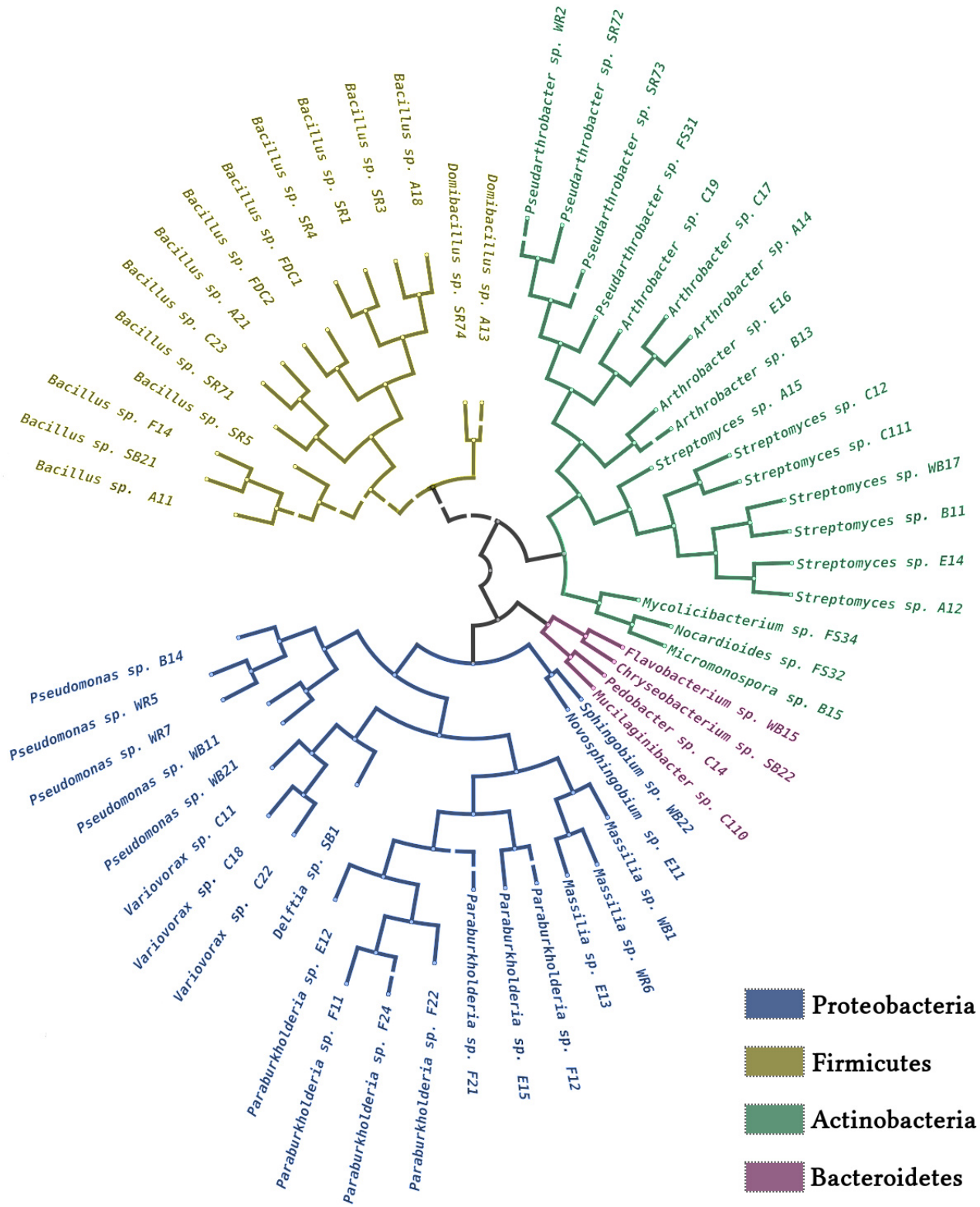


Figure 37.: Cladogram derived from the analysis of 16S rRNA gene sequences of 59 cultured strains isolated from the sample, with their taxonomic affiliation.

Most of Proteobacteria belonged to the class Betaproteobacteria[8] and order Burkholderiales (14 strains) as well as order Pseudomonadales of Gammaproteobacteria (5 strains). Two remaining strains belonged to the order Sphingomonadales of the class Alphaproteobacteria. Phylum Firmicutes was represented exclusively by the order Bacillales of the class Bacilli. Majority of the strains from the phylum Actinobacteria belonged to orders Streptomycetales (7 strains), Actinomycetales (5 strains) and Micrococcales (4 strains) of the class Actinobacteria; remaining three strains belonged to orders Corynebacteriales, Micromonosporales and Propionibacteriales of the same class. In the phylum Bacteroidetes, orders Flavobacteriales and Sphingobacteriales belonging to classes Flavobacteriia and Sphingobacteriia had two strains each.

On the per-sample basis, isolated strains varied in number and taxonomic affiliation. On phylum level (Figure 38), soil samples (A,B and C) and soot sample taken in the summer (FS) contained a sizeable amount of strains belonging to the phylum Actinobacteria, while majority of the strains isolated from flue dust (E), arsenic-rich soot (F) and groundwater (WB and WR) belonged to Proteobacteria. Several strains isolated from sediment sample SR, soil sample A, and arsenic-rich soot sample taken in the summer (FS) belonged to the phylum Firmicutes.

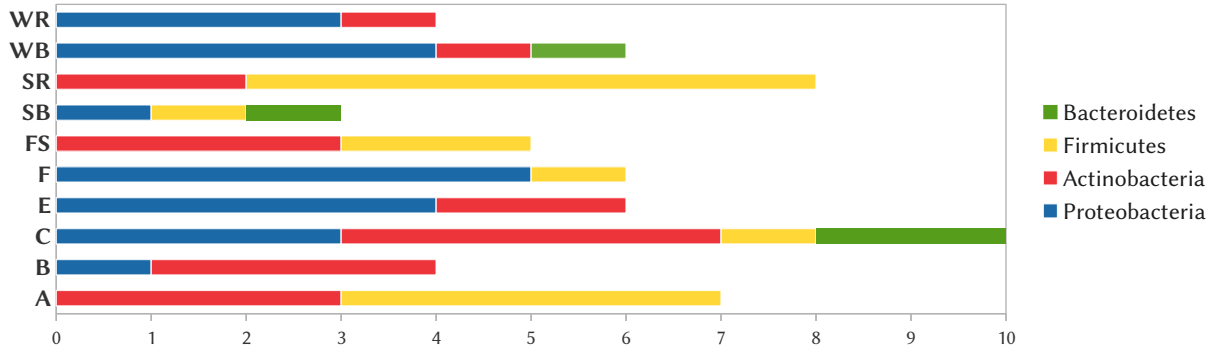


Figure 38.: Taxonomic distribution (phylum level) of bacterial strains isolated from El Terronal samples. Affiliation using 97% identity cut-off against SILVA128 database.

When only solid samples taken in the winter (soil samples A, B and C; flue dust sample E and soot sample F) were considered, proportion of Proteobacteria seemed to increase with increasing contamination levels. Statistical validation of this relationship, however, was extremely tenuous, as sample number was low, and there were pronounced outlier effects caused by orders-of-magnitude differences in As and Hg concentrations between some of the samples. With groundwater samples and soot sample taken under different climatic conditions taken into account, no such correlations appeared. For example, while arsenic-rich soot sample taken in the winter (F) was dominated by Proteobacteria, sampling from the same brownfield in the summer (FS) produced only Actinobacteria and Firmicutes. On the other hand, cultivable microbial communities of significantly less

[8]Betaproteobacteria were merged into Gammaproteobacteria as order Burkholderiales since the adoption of the Genome Taxonomy Database by SILVA in the version 138 (Parks et al., 2018; Parks et al., 2020).

polluted groundwater samples (WB and WR) were dominated by Proteobacteria. This again suggests high degree of dependence on factors other than arsenic and mercury concentration in the environment.

On the order level (Figure 39), the picture was a bit more nuanced: Proteobacteria strains isolated from groundwater belonged mostly to the order Pseudomonadales, while those isolated from riverbank soil, flue dust and soot belonged to the order Burkholderiales. There were no statistically significant correlations between increasing pollution levels and order-level phylogeny of the isolated strains.

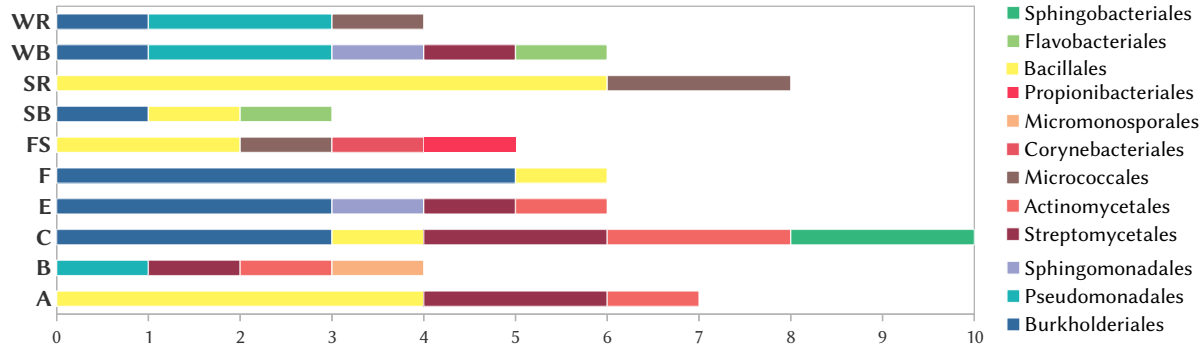


Figure 39.: Taxonomic distribution (order level) of bacterial strains isolated from El Terronal samples. Affiliation using 97% identity cut-off against SILVA128 database.

Comparison with metagenomic data

Illumina sequencing of V4-V5 hypervariable regions of 16S rRNA gene generates sequences ~400 bp long. This results in lower resolution compared Sanger sequencing or high-throughput, long read sequencing of the full-length 16S rRNA gene (Callahan et al., 2019). Therefore, when sequences analogous to Illumina-generated ASVs (hereinafter: the ASV analogues) were obtained from 16S rRNA gene sequences of cultivable strains, resolution was effectively lowered. It was not uncommon to see several morphologically and phylogenetically distinct strains (or even putative species of the same genus) grouped into the same ASV analogue; 59 strains were thus reduced to just 35 ASV analogues (Table 35). Some of the cultured strains did not have corresponding sequences in the Illumina data, with most of them coming from high-diversity samples such as soils and sediments. This is concurrent with alpha-diversity rarefaction curves suggesting lost diversity due to insufficient sequencing depth (see subsection 5.2.2).

Isolated strains represented no more[9] than 0.39% of all ASVs, and up to ~1% of total ASV count in (high-diversity) samples of soil (A, B & C), as well as flue dust (E) and groundwater sediments (SR & SB). This suggested that their place and role in the overall microbial community was relatively minor. Situation was different in arsenic-rich soot samples: in the sample F, ASV analogues of isolated strains corresponded to 0.59% of

[9]Culturability yield calculated allowing for complete representation of ASVs by isolated strains; however, sharing the ASV with an unknown number of uncultured strains appears more likely.

all ASVs, but up to 11.95% of total counts, contributed mostly by the *Paraburkholderia* strains F11, F12, F21 and F24, suggesting their importance in the functioning of the microbial community. In contrast, bacteria isolated from the soot sample FS, while coinciding with ~1.2% of all ASVs, could represent no more than 0.14% of total counts (Table 18).

Table 18.: Place of isolated strains within Prokaryotic community (culturability yield). Replicas merged into composite samples.

	A	B	C	D	E	F	FS	SR	SB
% of ASVs	0.18	0.08	0.15	0	0.28	0.59	1.19	0.15	0.39
% of counts	0.56	0.74	0.95	0	0.99	11.95	0.14	0.13	0.31

5.5. Arsenic resistance of isolated strains

Cultivation in TSA media supplemented by increasing concentrations of arsenic demonstrated impressive rates of As(III) and As(V) resistance across all samples, while amplification of genes responsible for As(III) oxidation (*aio*-like arsenite oxidase genes), As(III) efflux (*acr3* and *arsB* arsenite pumps), and As(V) reduction (*arrA* respiratory reductase) revealed their widespread presence (Table 19).

Table 19.: Arsenic resistance of cultivable strains as determined by cultivation in TSA medium supplemented with As(III) and As(V), and PCR amplification of As resistance genes.

Strain	As(III) resistance			As(V) resistance	
	As(III), mM	Oxidation	Efflux	As(V), mM	Reduction (respiratory)
A (Topsoil)					
Bacillus sp. A11	10			300	
Streptomyces sp. A12	2		+	300	
Domibacillus sp. A13	10	+	+	200	
Arthrobacter sp. A14	10		+	300	
Streptomyces sp. A15	5		+	300	
Bacillus sp. A18	10		+	300	
Bacillus sp. A21	5			20	
B (Soil, mixed and tilled)					
Streptomyces sp. B11	0			200	
Arthrobacter sp. B13	10	+	+	300	
Pseudomonas sp. B14	15		+	300	+
Micromonospora sp B15	5		+	300	
C (Soil of the river bank)					
Variovorax sp. C11	0	+	+	200	
Streptomyces sp. C12	0		+	200	
Pedobacter sp. C14	10			200	
Arthrobacter sp. C17	10		+	300	
Variovorax sp. C18	10		+	50	
Arthrobacter sp. C19	20		+	300	
Mucilaginibacter sp. C110	0		+	100	
Streptomyces sp. C111	2		+	300	

Table 19.: Arsenic resistance of cultivable strains as determined by cultivation in TSA medium supplemented with As(III) and As(V), and PCR amplification of As resistance genes.

Strain	As(III) resistance			As(V) resistance	
	As(III), mM	Oxidation	Efflux	As(V), mM	Reduction (respiratory)
Variovorax sp. C22	25	+	+	300	
Bacillus sp. C23	25		+	20	
E (Flue dust)					
Novosphingobium sp. E11	20		+	100	
Paraburkholderia sp. E12	5	+	+	50	+
Massilia sp. E13	5		+	20	+
Streptomyces sp. E14	2			300	
Paraburkholderia sp. E15	5	+	+	20	
Arthrobacter sp. E16	25	+	+	300	
F (Arsenic-rich soot, winter)					
Paraburkholderia sp. F11	5	+	+	50	
Paraburkholderia sp. F12	10	+	+	100	
Bacillus sp. F14	25		+	100	
Paraburkholderia sp. F21	20	+	+	50	+
Paraburkholderia sp. F22	10		+	100	
Paraburkholderia sp. F24	20	+	+	100	+
FS (Arsenic-rich soot, summer)					
Bacillus sp. FDC1	25		+	300	
Bacillus sp. FDC2	25		+	300	
Pseudarthrobacter sp. FS31	2		+	300	
Nocardioides sp. FS32	25		+	200	
Mycolicibacterium sp. FS34	25		+	200	
SR (Groundwater sediment, river bank)					
Bacillus sp. SR1	20		+	200	
Bacillus sp. SR3	20		+	200	
Bacillus sp. SR4	5			20	
Bacillus sp. SR5	2	+		5	
Bacillus sp. SR71	25	+	+	200	

Table 19.: Arsenic resistance of cultivable strains as determined by cultivation in TSA medium supplemented with As(III) and As(V), and PCR amplification of As resistance genes.

Strain	As(III) resistance			As(V) resistance	
	As(III), mM	Oxidation	Efflux	As(V), mM	Reduction (respiratory)
Pseudarthrobacter sp. SR72	10		+	300	
Pseudarthrobacter sp. SR73	10		+	300	
Domibacillus sp. SR74	2			200	
SB (Groundwater sediment, remediation plot)					
Delftia sp. SB1	20		+	300	
Bacillus sp. SB21	10			200	
Chryseobacterium sp. SB22	10	+	+	300	
WR (Groundwater, river bank)					
Pseudarthrobacter sp. WR2	20		+	300	+
Pseudomonas sp. WR5	20	+	+	300	+
Massilia sp. WR6	5	+	+	200	
Pseudomonas sp. WR7	10	+	+	300	
WB (Groundwater, remediation plot)					
Massilia sp. WB1	0	+		20	+
Pseudomonas sp. WB11	5	+		300	+
Flavobacterium sp. WB15	2	+	+	200	
Streptomyces sp. WB17	2		+	300	
Pseudomonas sp. WB21	2			0	
Sphingobium sp. WB22	5		+	300	

Resistance to As(V) was almost universal, with only one strain isolated from groundwater (*Pseudomonas sp.* WB21) not capable of growth in presence of arsenate. 47 strains out of 59 (79.6%) resisted sodium arsenate concentrations in excess of 100 mM, of which 27 strains (45.8%) and 14 strains (23.7%) were capable of tolerating arsenate concentrations of 300 mM and 200 mM, respectively (Figure 40). Distribution of resistant strains within samples was even, apart from strains from arsenic-rich soot (F) having lower maximum arsenate resistance levels.

Resistance to As(III) followed a different pattern, with 18 hyper-resistant strains (30.5%) growing on media supplemented with either 20 or 25 mM of sodium arsenite, and 26 strains (44%) tolerating either 5 or 10 mM of arsenite. Five strains (8.47%) were not capable of growth in the presence of sodium arsenite at all (Figure 41).

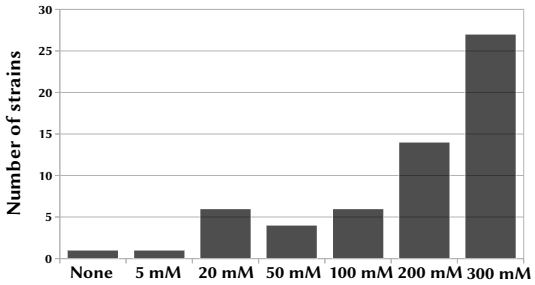


Figure 40.: Resistance to As(V).

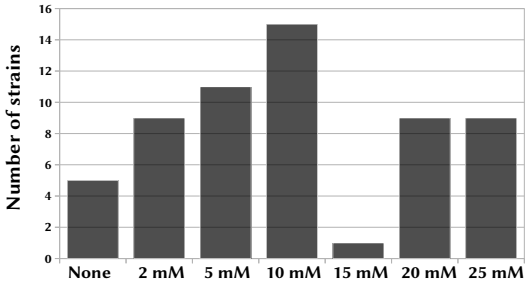


Figure 41.: Resistance to As(III).

Most-contaminated samples such as arsenic-rich soot (F & FS), riverside soil (C), riverside well groundwater (WR) and sediment (SR) contributed more strains with arsenite resistance in excess of 20 mM. Strains isolated from relatively-clean groundwater from phytoremediation plot sounding well and soil had lower resistance levels.

Out of four families of arsenic resistance genes amplified by PCR, *acr3* arsenite efflux pumps were by far the most common, being found in 38 strains (64%). Next most-common gene families were *aio*-like arsenite oxidases (20; 34% of all strains) and *arsB* arsenite efflux pumps (22 strains and 37%). Respiratory arsenate reductase genes were the least common, and appeared in only 9 strains (15%). When it comes to occurrence of more than one arsenic resistance gene, 17 strains (29%) carried both arsenite oxidase and arsenite efflux pumps, and 4 of them (~7%) had arsenate reductase genes as well.

For several strains with high degree of arsenic resistance there was no amplification of arsenic resistance genes. This indicates that arsenic detoxification was encoded either by (a) more distant orthologues of known genes that were not amplified with selected primer sets or (b) other arsenic resistance genes not covered by the selected primer sets.

5.6. Mercury resistance of isolated strains

Majority of cultivable strains demonstrated confident growth in TSA medium supplemented with mercury chloride (See Table 36). 14 (24%) and 7 (12%) strains could tolerate 400 and 300 μM of $HgCl_2$, respectively, while 22 strains (37%) were capable of growth on the medium supplemented with 50-100 μM of $HgCl_2$. Only 3 strains (5%) were not capable of growing on TSA medium supplemented with mercury (Figure 42).

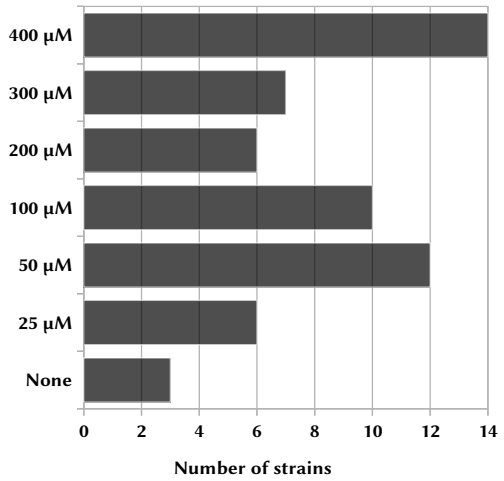


Figure 42.: Resistance to $HgCl_2$ among cultivable strains.

Table 20.: Strains for which *merA* was successfully amplified with primer set A1s-n.F and A5-n.R.

Strain	$HgCl_2$, μM
Massilia sp. E13	25
Paraburkholderia sp. E15	50
Variovorax sp. C22	100
Paraburkholderia sp. F12	300
Pseudarthrobacter sp. WR2	400
Massilia sp. WR6	400
Massilia sp. WB1	200

Amplification of the *merA* gene fragment by PCR was successful in only 7 strains (Table 20). It is highly likely that *merA* was present in more strains, but was not amplifiable by the selected primer set.

Strains with very high resistance (300-400 μM of $HgCl_2$) were isolated mainly from two arsenic-rich soot samples (F & FS), as well as groundwater taken from the river bank well (WR). All three strains that were not capable of growth in TSA medium supplemented with mercury originated from the soil samples (A, B and C).

5.7. Antibiotic resistance and production of antimicrobial compounds

Testing by the disk diffusion method revealed that many isolated strains were resistant to at least one antibiotic (Table 37). Most common resistances were those to trimetoprim (23 strains, 39% of all strains) and cefuroxime (16 strains, 27%). Resistance to vancomycin was the least common, with only one resistant strain; however, 25 strains that belonged to the phyla Proteobacteria and Bacteroidetes were naturally resistant to it due to lacking a cell wall. Out of wide-spectre antibiotics, resistance to tetracycline was the least common, with only 3 resistant strains (Figure 43).

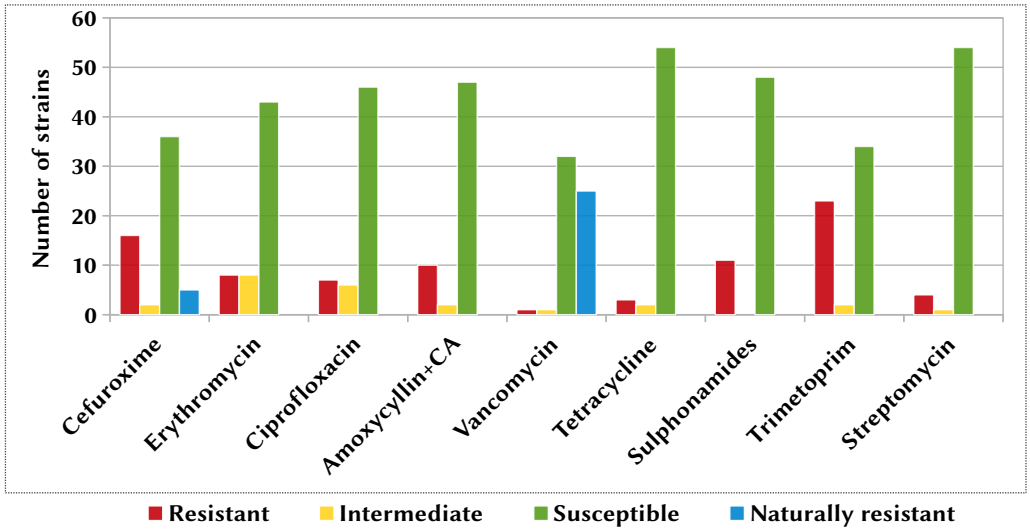


Figure 43.: Prevalence of antibiotic resistance and susceptibility among cultivable strains.

Resistance to multiple antibiotics was also common, with 37 strains having resistance to more than one antibiotic. MAR index for individual strains (MAR_{strain}) was ranging from 0 to 0.75 (mean 0.17, $\sigma = 0.195$). 22 strains had MAR_{strain} index above 0.2, and 3 strains (C110, C22 and SB22) had very high index of 0.75.

There was no statistically significant ($\rho \geqq 0.85, p < 0.05$) correlation between resistance to any antibiotic and resistance to arsenic or mercury. Multiple antibiotic resistance did not correlate with resistance to arsenic and mercury, either, as there were similar number of cases of high antibiotic and metal(loid) resistance, high antibiotic and low metal(loid) resistance, and low antibiotic and high metal(loid) resistance (Figure 44).

There was considerable variation in MAR_{sample} index (Table 21), but no correlation with arsenic or mercury concentration, nor with the sample type.

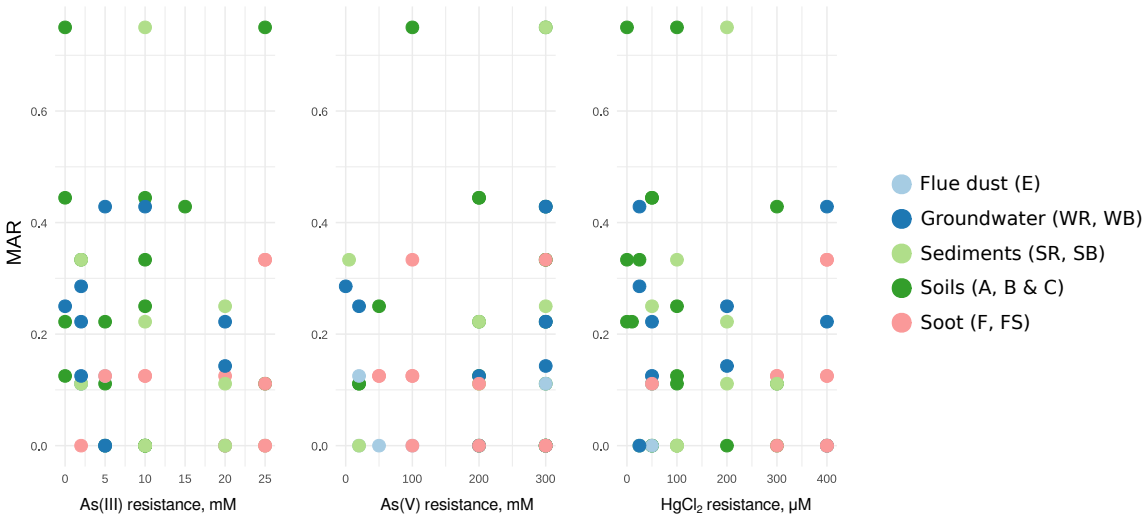


Figure 44.: Scatter plots of resistance to As(III), As(V) and $HgCl_2$against MAR_{strain} index showing lack of relationship between them. Each point represents a strain; their colours indicate their sample of origin.

Table 21.: MAR index of samples.

Sample	Sample description	MAR index
A	Topsoil	0.175
B	Soil (phytoremediation plot)	0.139
C	Soil (river bank)	0.256
E	Flue dust	0.037
F	As-rich soot (winter)	0.13
FS	As-rich soot (summer)	0.069
SR	Groundwater sediment (river bank)	0.167
SB	Groundwater sediment (remediation plot)	0.37
WR	Groundwater (river bank)	0.185
WB	Groundwater (remediation plot)	0.089

In parallel to the antibiotic resistance testing, evaluation of production of substances with antimicrobial activity by isolated strains was carried out. Majority of strains were not affecting the growth of succeptible *E. coli* strain. Three strains, however, had a very mild growth inhibition effect on susceptible *Escherichia coli* strain ATCC 25922 during co-cultivation on agar plates: *Streptomyces sp.* A12, *Streptomyces sp.* E14 and *Paraburkholderia sp.* E16.

5.8. Degradation of organic compounds

Twenty-eight strains were capable of at least limited growth utilising at least one of the tested hydrocarbons (diesel, hexadecane, naphthalene or phenol) as its sole source of carbon. Of them, 22 strains were capable of growing on diesel, 9 strains could grow on hexadecane, 7 strains were capable of using naphthalene, and 6 strains could utilise phenol. (Table 22).

Majority of strains were growing in suspension; eight strains were growing in aggregations with diesel fuel as the sole carbon source; of those strains, four (B11, WR6, WB1 and WB17) were growing in aggregations with hexadecane as sole carbon source as well. Those strains belonged either to *Streptomyces* or *Massilia*, with the latter forming loose, easily breakable lumps. Three strains (A21, B13 and SR74) were not capable of growth in liquid TSB medium at all, while 17 strains could not grow with glucose as sole carbon source; some of them, however, were capable of growth on diesel or hexadecane.

Table 22.: Strains capable of utilising diesel, hexadecane and phenol as their sole carbon source.

Strain	Diesel	Hexadecane	Naphthalene	Phenol
Streptomyces sp. A12	+	-	-	-
Arthrobacter sp. A14	-	-	+	+
Streptomyces sp. A15	+	-	-	-
Streptomyces sp. B11	+	+	-	-
Streptomyces sp. C12	-	-	-	-
Pedobacter sp. C14	-	+	-	-
Arthrobacter sp. C17	-	-	-	+
Variovorax sp. C18	-	-	-	-
Arthrobacter sp. C19	+	-	+	+
Streptomyces sp. C111	+	-	-	-
Variovorax sp. C22	+	-	-	+
Paraburkholderia sp. E12	+	-	-	-
Streptomyces sp. E14	+	-	-	-
Paraburkholderia sp. E15	+	-	-	-
Paraburkholderia sp. F12	+	-	-	-
Paraburkholderia sp. F21	+	-	-	-
Paraburkholderia sp. F24	+	-	+	+
Mycolicibacterium sp. FS34	+	-	-	-
Bacillus sp. SR3	-	-	-	+
Bacillus sp. SR4	+	-	-	-
Pseudarthrobacter sp. SR72	+	+	+	-
Pseudarthrobacter sp. SR73	+	+	+	-
Chryseobacterium sp. SB22	+	-	-	-

Table 22.: Strains capable of utilising diesel, hexadecane and phenol as their sole carbon source.

Strain	Diesel	Hexadecane	Naphthalene	Phenol
Pseudarthrobacter sp. WR2	+	+	+	-
Massilia sp. WR6	+	+	-	-
Massilia sp. WB1	+	+	-	-
Pseudomonas sp. WB11	+	+	-	-
Streptomyces sp. WB17	+	+	-	-
Sphingobium sp. WB22	-	-	+	-

Growth in suspension usually resulted in breaking down large hydrocarbon droplets into smaller ones (Figures 45A, 45B), sometimes down to their complete dissolution (Figure 45C). Growth in aggregates occurred mainly in strains capable of forming hyphae; hydrocarbon droplets become trapped between them (Figure 45D). Growth with naphthalene or phenol as a sole source of carbon occurred in suspension and was very weak.

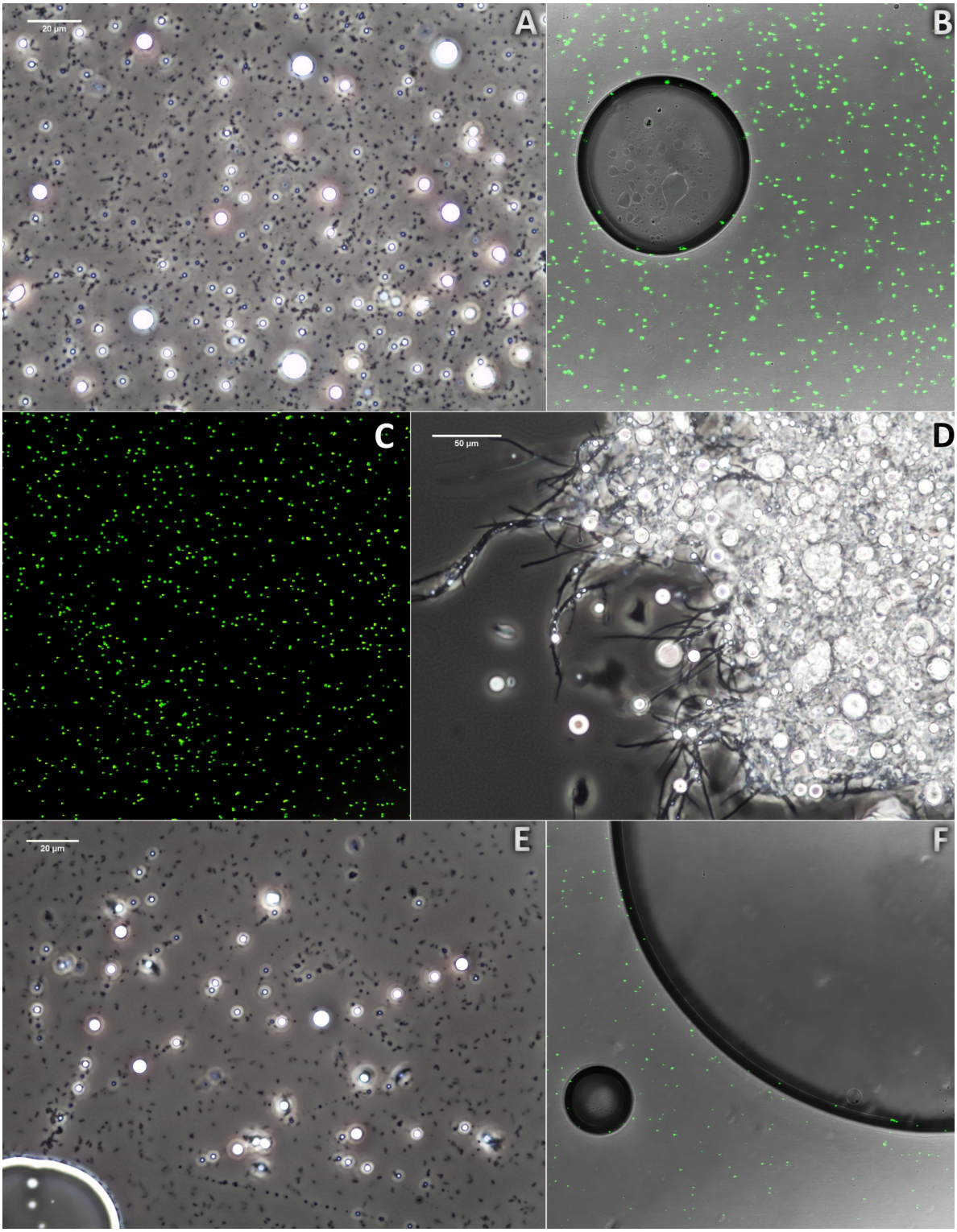


Figure 45.: Examples of strains growing in BH medium with either diesel or hexadecane as their sole carbon source. Microphotographs of visible light microscopy (A,D a& E) and confocal microscopy with vital staining by LIVE/DEAD BacLight Bacterial Viability Kit (B, C &F). A: *Paraburkholderia sp. E12*, diesel. B: *Pseudarthrobacter sp. SR72*, diesel. C: *Pedobacter sp. C14*, hexadecane. D: *Streptomyces sp. C111*, diesel. E: *Pseudomonas sp. B14*, diesel. F: *Pseudarthrobacter sp. WR2*, hexadecane.

5.9. Emulsification of organic compounds

Emulsification index allows to assess surfactant production by bacteria, as well as surfactant efficiency at emulsifying hydrocarbons and stability of produced emulsions over time. Out of 59 tested strains, 14 were not capable of growth in BH-glucose-hexadecane emulsification-inducing media and 9 strains did not produce an emulsion. The rest of strains (36 strains, 61%) were capable of emulsifying at least one of the tested hydrocarbons (Table 23).

Table 23.: E_{24} index of cultivable strains measured for hexadecane, xylene and octane. Strains that were not capable of growth in the emulsification-inducing medium not shown. Emulsions stable over the course of 3 weeks are marked with an asterisk.

Strain	E_{24}		
	Hexadecane	Octane	Xylene
A (Topsoil)			
Streptomyces sp. A12	50*	0	0
Domibacillus sp. A13	10	0	20
Arthrobacter sp. A14	20	15	20
Streptomyces sp. A15	0	0	0
Bacillus sp. A21	0	5	0
B (Soil, mixed and tilled)			
Streptomyces sp. B11	60*	5	10
Pseudomonas sp. B14	5	45	5
Micromonospora sp. B15	0	0	0
C (Soil of the river bank)			
Variovorax sp. C11	0	0	0
Streptomyces sp. C12	0	0	0
Arthrobacter sp. C17	0	40	35
Variovorax sp. C18	0	0	0
Arthrobacter sp. C19	25	25	20
Mucilaginibacter sp. C110	0	15	10
Streptomyces sp. C111	10	0	0
Variovorax sp. C22	0	0	15
Bacillus sp. C23	40*	15	20
E (Flue dust)			
Paraburkholderia sp. E12	10	15	35

Table 23.: E_{24} index of cultivable strains measured for hexadecane, xylene and octane. Strains that were not capable of growth in the emulsification-inducing medium not shown. Emulsions stable over the course of 3 weeks are marked with an asterisk.

Strain	E_{24}		
	Hexadecane	Octane	Xylene
Massilia sp. E13	55*	20*	40
Streptomyces sp. E14	50*	5	25
Paraburkholderia sp. E15	50*	20	20*
Arthrobacter sp. E16	0	0	0
F (Arsenic-rich soot, winter)			
Paraburkholderia sp. F11	20	5	0
Paraburkholderia sp. F12	10	5	25
Paraburkholderia sp. F21	0	10	10
Paraburkholderia sp. F22	0	30*	20*
Paraburkholderia sp. F24	20	25	20
FS (Arsenic-rich soot, summer)			
Pseudarthrobacter sp. FS31	30*	0	35*
Nocardioides sp. FS32	0	0	0
Mycolicibacterium sp. FS34	0	0	0
SR (Groundwater sediment, river bank)			
Bacillus sp. SR1	0	0	10
Bacillus sp. SR3	0	15	10
Bacillus sp. SR4	0	5	15
Bacillus sp. SR5	0	0	0
Pseudarthrobacter sp. SR72	0	0	25
Pseudarthrobacter sp. SR73	0	10	30
Domibacillus sp. SR74	50*	5	0
SB (Groundwater sediment, remediation plot)			
Delftia sp. SB1	10	5	10
WR (Groundwater, river bank)			
Pseudarthrobacter sp. WR2	0	10	5
Massilia sp. WR6	5	15	15
Pseudomonas sp. WR7	20	35	25
WB (Groundwater, remediation plot)			
Massilia sp. WB1	60*	0	55*

Table 23.: E_{24} index of cultivable strains measured for hexadecane, xylene and octane. Strains that were not capable of growth in the emulsification-inducing medium not shown. Emulsions stable over the course of 3 weeks are marked with an asterisk.

Strain	E_{24}		
	Hexadecane	Octane	Xylene
Pseudomonas sp. WB11	15	20	25
Flavobacterium sp. WB15	0	10	0
Streptomyces sp. WB17	0	0	0
Sphingobium sp. WB22	25	20	20

23 strains (39%) were capable of emulsifying hexadecane; 30 strains (51%) – octane, and 28 strains (32%) – xylene. Of all tested strains, 29 (49%) could emulsify more than one hydrocarbon, and 16 strains (27%) were able to emulsify all three of them. There was no statistically significant correlation between capacity to emulsify different hydrocarbons.

Differences between samples were not statistically significant; however, flue dust sample (E) had the highest number of emulsions stable over the course of three weeks (five emulsions of two hydrocarbons produced by three strains) compared to the rest of the samples ($1.1,\ \sigma = 0.78$).

5.10. Selection of strains for genome sequencing

Strains for Illumina whole genome sequencing were selected based on the following characteristics:

- High As(III), As(V) and/or mercury resistance levels, preferably in combination, and with few or no arsenic/mercury resistance genes amplified by PCR;
- Multiple antibiotic resistance (high MAR_{strain}) with or without high resistance to As and Hg;
- Other properties of interest, such as capacity to emulsify and / or degrade hydrocarbons.

Intersection of those characteristics was favoured in strain selection. A total of 17 strains was selected, as detailed below:

***Pseudomonas sp.* B14** High resistance to As(III), As(V) & $HgCl_2$; high MAR_{strain};

***Arthrobacter sp.* C19** High resistance to As(III) and As(V); hydrocarbon degradation and emulsification;

***Mucilaginibacter sp.* C110** Very high MAR_{strain};

***Variovorax sp.* C22** High resistance to As(III), As(V) and $HgCl_2$; very high MAR_{strain};

***Paraburkholderia sp.* F12** Medium-to-high As(III) and As(V) resistance; high $HgCl_2$ resistance;

Bacillus sp. F14 Very high As(III) and $HgCl_2$ resistance and high MAR_{strain};

***Paraburkholderia sp.* F21** Very high As(III) and $HgCl_2$ resistance;

***Nocardioides sp.* FS32** Very high As(III), As(V) and $HgCl_2$ resistance; high MAR_{strain};

***Bacillus sp.* SR1** Very high As(III), As(V) and $HgCl_2$ resistance;

***Bacillus sp.* SR3** Very high As(III), As(V) and $HgCl_2$ resistance;

***Delftia sp.* SB1** High As(III) and As(V) resistance and high MAR_{strain};

***Bacillus sp.* SB21** Medium-to-high As(III), As(V) and $HgCl_2$ resistance & no amplified As nor Hg resistance genes;

***Chryseobacterium sp.* SB22** High As(V) and $HgCl_2$ resistance; very high MAR_{strain};

***Pseudomonas sp.* WR7** Medium-to-high As(III), As(V) resistance and very high $HgCl_2$ resistance; high MAR_{strain};

***Massilia sp.* WB1** High $HgCl_2$ resistance and high MAR_{strain}; hydrocarbon degradation and emulsification;

***Pseudomonas sp.* WB11** High As(V) resistance; high MAR_{strain};

***Streptomyces sp.* WB17** High As(V) resistance;

5.11. Genome analysis

5.11.1. Genome assemblies

Most genomes were assembled into 20-30 contigs each; genomes of strains *Variovorax sp. C22*, *Mucilaginibacter sp. C110*, *Delftia sp. SB1*, *Chryseobacterium sp. SB22* and *Pseudomonas sp. WB11* were assembled into ~10 contigs or less. Genomes of strains *Paraburkholderia sp. F12*, *Paraburkholderia sp. F21* and *Bacillus sp. SB21*, in comparison, were more fragmented. The largest genomes belonged to the two *Paraburkholderia* strains (F12 and F21), with ~ 8.5 million base pairs each. *Pseudomonas sp.* WB11 had the shortest assembly, at just over 4.1 million base pairs. Average length was 6.457 million bp, σ =1.272 million bp (Table 24).

Table 24.: General information on genome assemblies: number of contigs, total length of the assembly, N50 and L50 statistics, number of mismatches (total and per 100 nucleotides), and GC content.

Strain	№ of contings	total length, bp	N50	L50	№ of N's	N's per 100 kbp	GC (%)
Pseudomonas sp. B14	28	5689426	389938	4	9	0.16	60.81
Arthrobacter sp. C19	24	5335847	316244	5	0	0	65.86
Variovorax sp. C22	11	6946589	6625774	1	10	0.13	42.38
Mucilaginibacter sp. C110	2	7566577	1406912	2	1	0.01	67.82
Paraburkholderia sp. F12	67	8505473	265539	11	9	0.11	62.01
Bacillus sp. F14	22	5364626	566519	4	0	0	35.02
Paraburkholderia sp. F21	40	8520700	626833	5	37	0.43	62.24
Nocardioides sp. FS32	17	5254779	864123	3	1	0.21	71.38
Delftia sp. SB1	6	6517827	2069324	2	24	0.37	66.81
Bacillus sp. SB21	83	6296586	1650640	2	160	2.54	35.1
Chryseobacterium sp. SB22	5	5208856	3669315	1	1	0.02	36.02
Bacillus sp. SR1	27	6605176	1010851	3	2	0.03	36.88
Bacillus sp. SR3	17	5464308	832481	3	2	0.04	40.46
Massilia sp. WB1	27	7448912	1007917	4	10	0.13	65.69
Pseudomonas sp. WB11	8	4103612	646001	3	1	0.02	60.86
Streptomyces sp. WB17	22	7874767	1266493	3	27	0.34	72.89
Pseudomonas sp. WR7	22	7064299	596478	5	4	0.06	59.12

N50, L50 and number of uncalled bases describe assembly quality. N50 is defined as the length (in bp) at which contigs of at least that length together account for at least 50% of the total assembly length, and L50 is the number of such contigs. Both are measures of genome fragmentation, with higher N50 and lower L50 indicating less fragmented assemblies with longer contigs. Number of mismatches (№ of N's) refers to the total number of uncalled bases (N's) in the assembly, while N's per 100 kbp indicates an average number of uncalled bases per 100 000 nucleotides of the assembly; high number of uncalled bases is indicative of low assembly quality, especially in cases where excessive scaffolding my artificially inflate N50 numbers.

Generally, assembly quality was high. N50 for genome assemblies of all strains except *Paraburkholderia sp. F12* was above 300 thousands bp; coverage of at least half of the total assembly length was achieved in 5 contigs or less. For *Paraburkholderia sp. F12*, to cover more than half of the assembled genome 11 contigs were required; its N50 number was ~250 thousands bp. The number of uncalled bases for this strain, however, was low (9 N's; 0.11 N's per 100 kbp), indicating that, despite higher fragmentation, the assembly quality was high. For the rest of the strains, the relative number of uncalled bases was below 0.5 per 100 000 bp for all strains except *Bacillus sp. SB21,* where 160 uncalled bases (107 of them – in a large scaffolding gap) contributed to the ratio of 2.54 uncalled bases per 100 000 bp (Table 24).

Complete or partial assembly of putative plasmids was successful for strains *Variovorax sp. C22*, *Paraburkholderia sp. F12* and *F21*, *Nocardioides sp. FS32*, *Bacillus sp. SB21*, *Bacillus sp. SR1* and *Streptomyces sp. WB17* (Table 25).

Table 25.: General information on plasmid assemblies: number of contigs and their total length, and number of circular sequences.

Strain	№ of contings	№ of circular sequences	total length, bp
Variovorax sp. C22	2	2	97260
Paraburkholderia sp. F12	2	0	302381
Paraburkholderia sp. F21	5	0	309973
Nocardioides sp. FS32	1	1	38182
Bacillus sp. SB21	4	0	66755
Bacillus sp. SR1	9	4	520585
Streptomyces sp. WB17	2	0	213317

5.11.2. Synteny

Synteny blocks are non-overlapping, highly conserved segments any given genome can be decomposed into. In the context of comparative microbial genomics, the number of shared synteny blocks between two or more genomes, and proportion of each genome represented by them are indicative of how related to each other the organisms are.

Among sequenced strains, four belonged to the genus *Bacillus*, two belonged to *Paraburkholderia*, and three more – to the *Pseudomonas* species. For each group of strains, due to their morphological similarity, closeness of their 16S rRNA gene sequences, and other similarities such as similar metal(loid) or antibiotic resistance, it was deemed important to access their similarities on the genomic level.

Bacillus strains

This group included *Bacillus spp.* strains F14, SR1, SR3 and SB21. Strains SR1 and SR3, have shown no synteny to each other nor to other sequenced *Bacillus* strains. In contrast, strains F14 and SB21 demonstrated high degree of similarity, with ~90% of their genomes covered by the shared synteny blocks (Table 26).

Multiplicity is the maximum number of times a synteny block can be encountered in the input genomes; encountering a given synteny block more than once in a (in the complete, circularised genome could indicate a duplication event. In fragmented genomes consisting of a set of contigs, interpretation is more difficult since there can be no clear indication whether a synteny block is encountered more than once due to duplication event, or simply belongs to the same part of the chromosome (or even a set of plasmids) that could not be scaffolded properly. In case of *Bacillus* strains F14 and SB21, strain SB21 had one block that was encountered twice, and one that was encountered thrice throughout the genome (Table 27). Analysis of the translated amino acid sequences of the genes from those synteny blocks indicated their viral origin.

Table 26.: Genome coverage by synteny blocks for *Bacillus* strains F14 and SB21.

Strain	Coverage
Bacillus sp. F14	96.29%
Bacillus sp. SB21	83.66%

Table 27.: Synteny block statistics for *Bacillus* strains F14 and SB21. Multiplicity is the number of times a block is encountered. Coverage is given for all input contigs together.

Multiplicity	№ of blocks	Coverage
2	65	89.39%
3	1	0.09%
All	66	89.47%

Graphical representation of the synteny block comparison between the two strains can be found in the Figure 46.

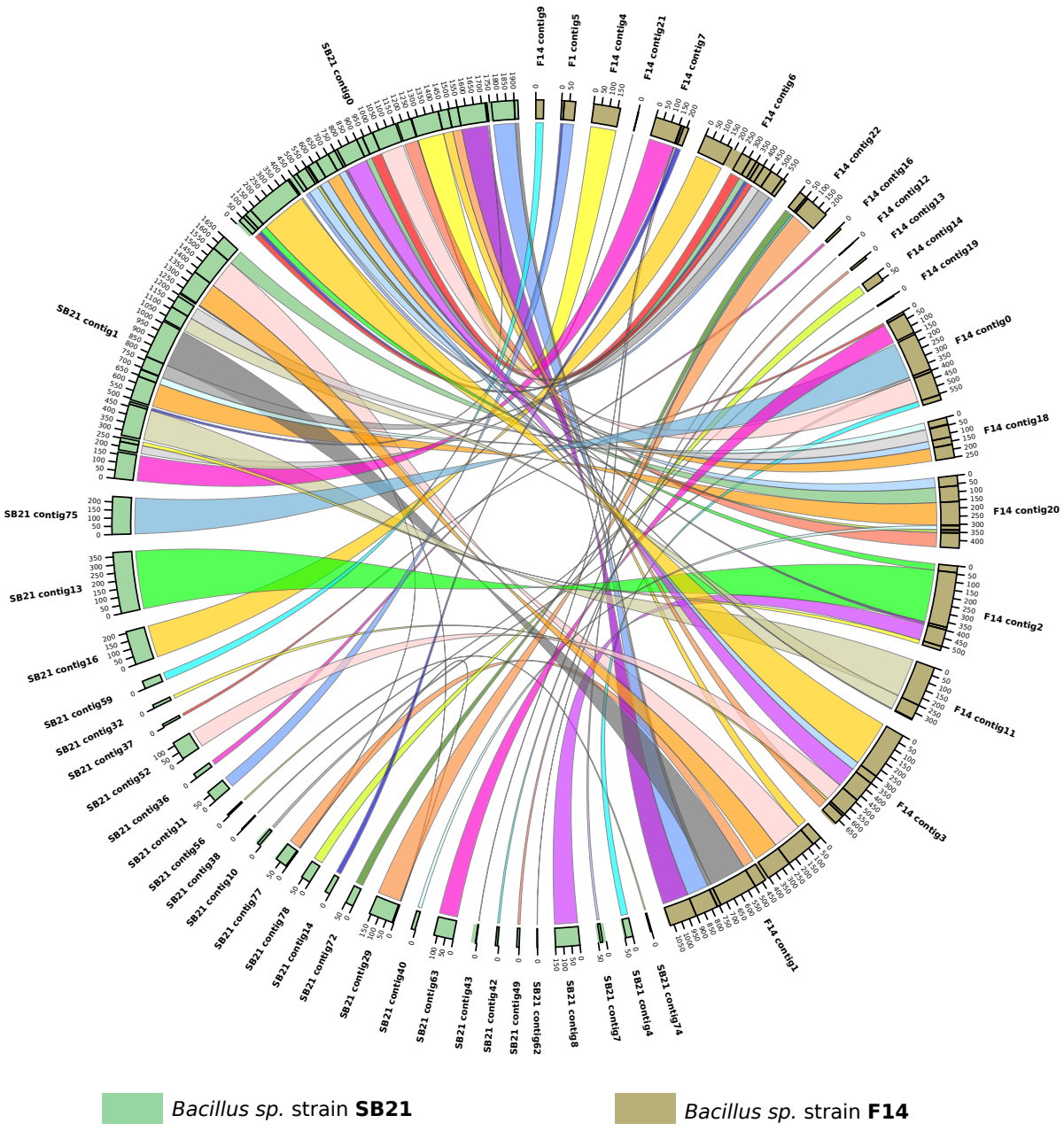


Figure 46.: Synteny blocks comparison between the genomes of *Bacillus* strains SB12 and F14. Contigs without shared synteny blocks are not shown. Colours are arbitrary.

Paraburkholderia strains

This group contained two *Paraburkholderia* strains: F12 and F21. They had no morphological differences, but demonstrated slightly different levels of tolerance towards arsenic and mercury as well as differences in their antibiotic resistance. On the level of genomes, about 20% of them belonged to the shared synteny blocks.

In the strain F12, four synteny blocks were encountered more than once; in F21 there were two duplicate blocks (Table 29). In the genomes of both strains they appeared to be true duplications, and carried bacterial genes coding for both known and uncharacterised proteins.

Table 28.: Synteny block statistics for *Paraburkholderia* strains F12 and F21.

Strain	coverage
Paraburkholderia sp. F12	21.05%
Paraburkholderia sp. F21	19.35%

Table 29.: Synteny block statistics for *Paraburkholderia* strains F12 and F21. Multiplicity is the number of times a block is encountered. Coverage is given for all input contigs together.

Multiplicity	№ of blocks	Coverage
2	53	20.03%
3	1	0.09%
4	1	0.09%
All	55	20.20%

Graphical representation of the synteny block comparison between the two strains can be found in the Figure 47.

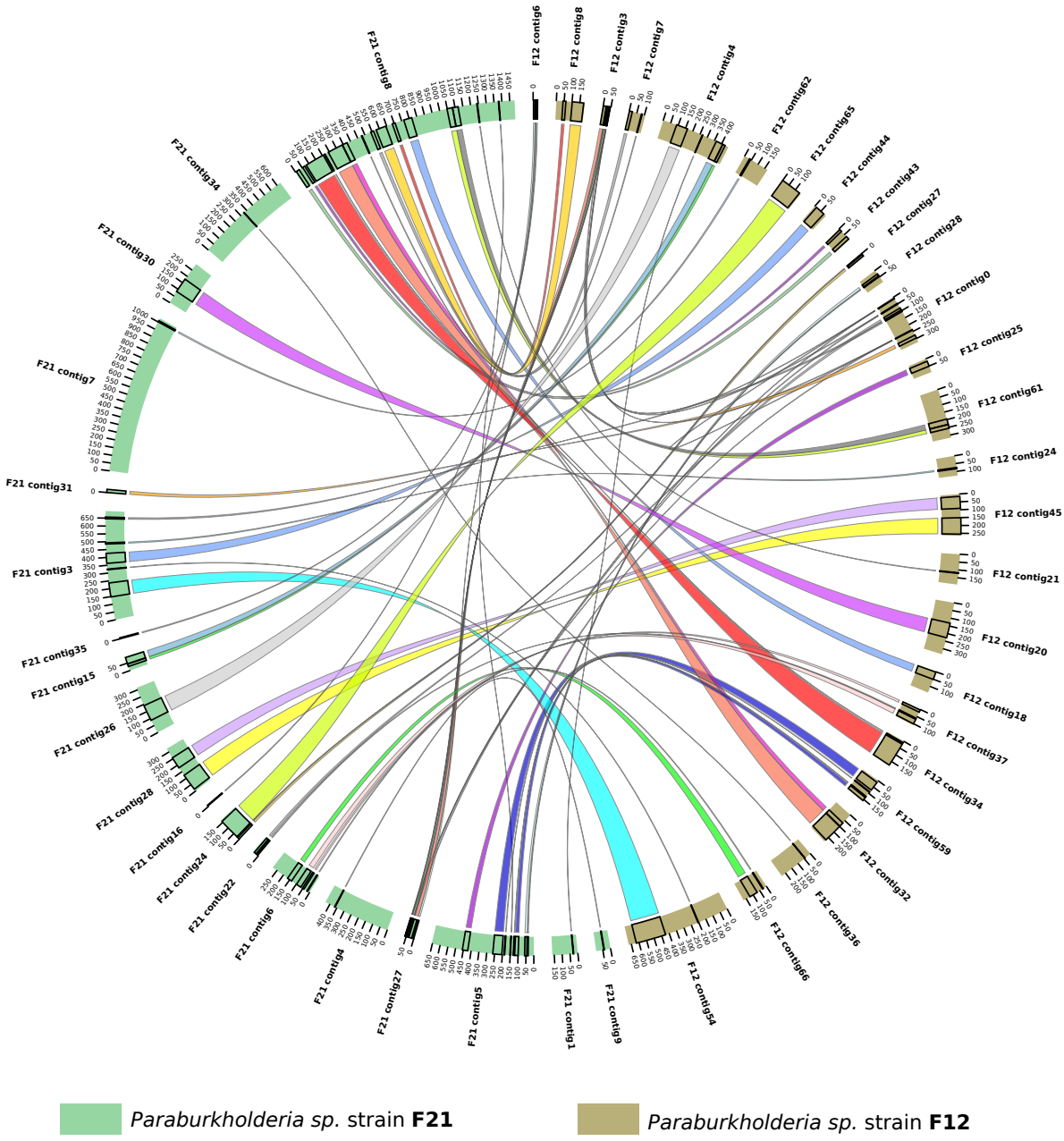


Figure 47.: Synteny block comparison between *Paraburkholderia* strains F12 and F21. Contigs without shared synteny blocks are not shown. Colours are arbitrary.

Pseudomonas strains

This comparison group consisted of *Pseudomonas spp.*, strains B14, WB11 and WR7. Their genomes had very little in common, with shared synteny blocks occupying ~2% of the genome for strains B14 and WR7, and around 0.5% - for WB11 (Table 30).

Most synteny blocks were shared between two genomes, and only two synteny blocks were shared between all three of them (Table 31). One of those blocks contained genes encoding ATP synthase; the other one contained ribosomal genes.

Table 30.: Synteny block coverage for *Pseudomonas* strains B14, WB11 and WR7.

Strain	Coverage
Pseudomonas sp. B14	2.23%
Pseudomonas sp. WB11	0.55%
Pseudomonas sp. WR7	1.97%

Table 31.: Synteny block statistics for *Pseudomonas* strains B14, WB11 and WR7. Multiplicity is the number of times a block is encountered. Coverage is given for all input contigs together.

Multiplicity	№ of blocks	Coverage
2	13	1.31%
3	2	0.40%
All	15	1.71%

Graphical representation of the synteny block comparison between the two strains can be found in the Figure 48.

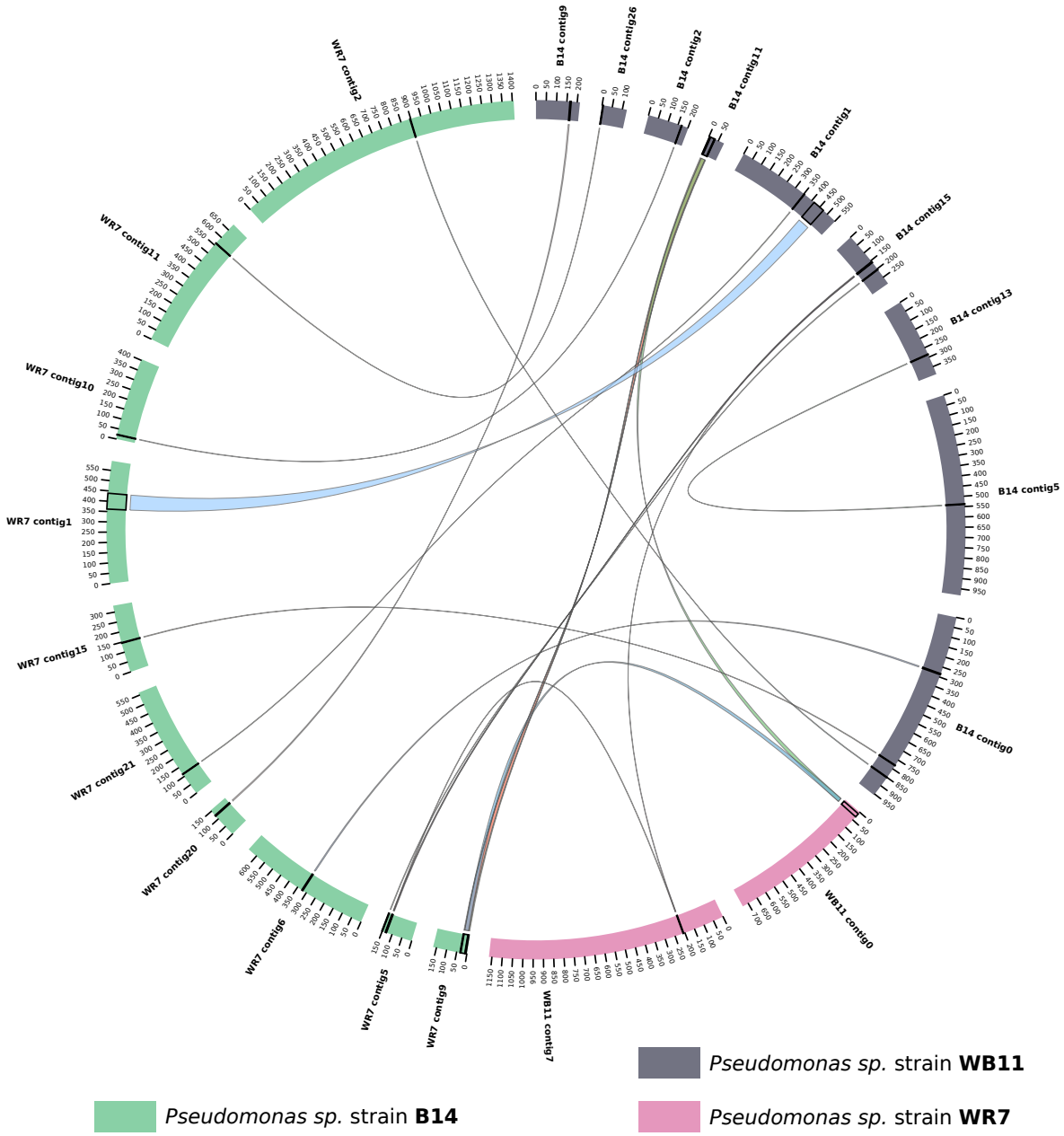


Figure 48.: Synteny block comparison between *Pseudomonas* strains B14, WB11 and WR7. Contigs without shared synteny blocks are not shown. Colours are arbitrary.

5.11.3. Arsenic and mercury resistance genes

Strains isolated from soil samples (A, B and C)

Pseudomonas sp. B14 It is a fairly resistant strain, capable of growth in TSA medium supplemented with either 15 mM of sodium arsenite, 300 mM of sodium arsenate, or 300 µM of mercury chloride. It possessed a number of metal(loid) resistance genes, including arsenate reductase (*arsC*), arsenite efflux pump (*arsB*), and organoarsenical oxidase (*arsH*). They were organised in two distinct clusters (Figure 49). Apart from that, a solitary *acr3* (arsenic efflux pump) gene was found. No mercury resistance genes apart from *merR* expression regulators/repressors were detected; this does not rule out their presence in the chromosome entirely, but suggests they might reside in a plasmid instead.

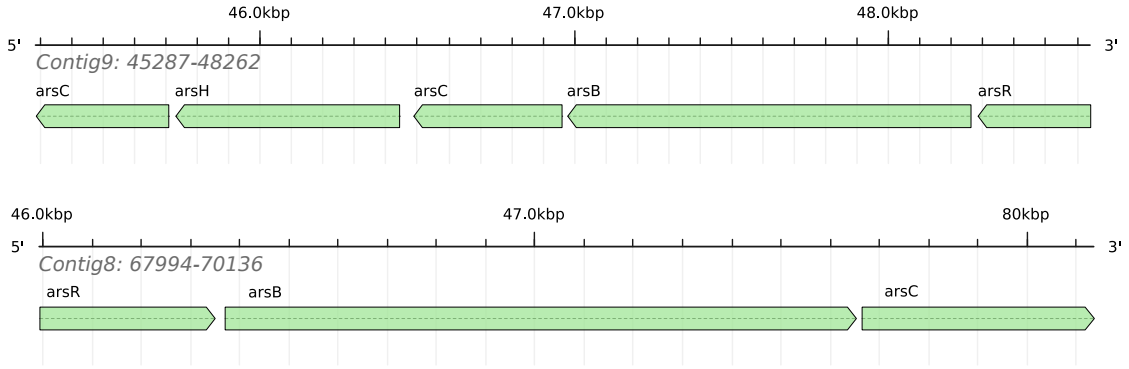


Figure 49.: Clusters of arsenic resistance genes in the assembled genome of *Pseudomonas sp. B14.*

Variovorax sp. C22 Being one of the most resistant to arsenic among isolated strains, C22 possessed four arsenic resistance gene clusters and a large cluster of mercury and arsenic resistance genes that were placed in an active integrone, downstream of the integrase gene (Figure 50). An additional arsenate reductase (*arsC*) gene was located outside of the clusters. Arsenic resistance determinants were represented by arsenate reductase (*arsC*), arsenite efflux pump (*arsB*), organoarsenical oxidase (*arsH*) and *acr3* arsenic efflux pump, as well arsenite oxidases *aioA* and *aioB*. In two clusters, arsenic resistance genes were located downstream of the cadmium-induced *cadI* genes. Mercury resistance genes formed a fairly typical *mer* operon that included mercury(II) reductase gene *merA*, mercury transporter *merC*, and mercuric transport genes *merP* and *merT*. Putative plasmid assemblies contained no arsenic or mercury resistance genes.

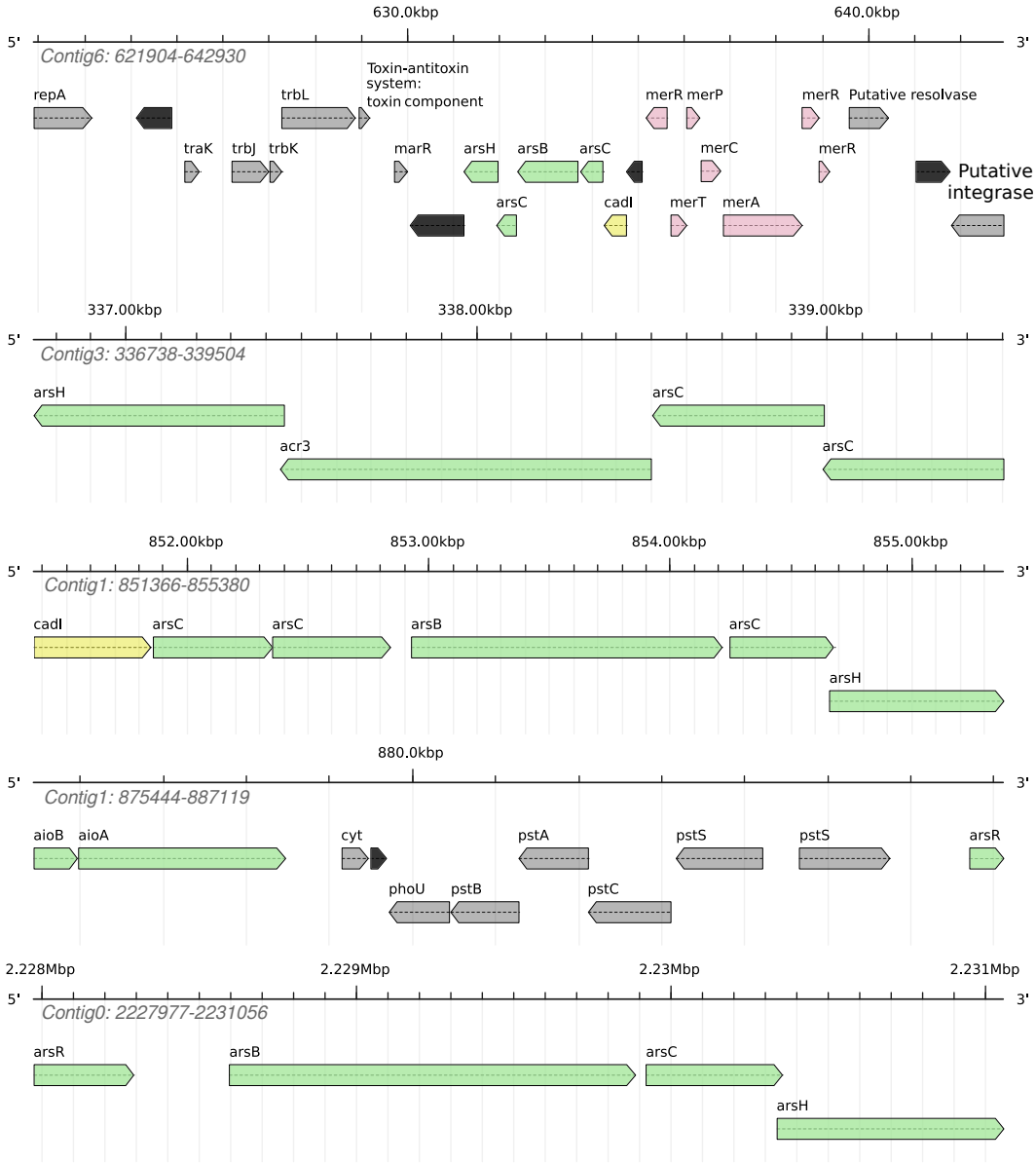


Figure 50.: Clusters of arsenic and mercury resistance genes in the assembled genome of *Variovorax sp. C22*. Green: arsenic resistance genes; pink: mercury resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Mucilaginibacter sp. C110 This strain has shown no resistance to arsenite and mercury in the cultivation experiments. Amplification of arsenic resistance genes by PCR, however, strongly suggested the presence of arsenite efflux pumps of the *ACR3* family. Genome analysis revealed the presence of two arsenic resistance gene clusters containing arsenate reductase (*arsC*) and arsenic efflux pump *acr3* (Figure 51). Separate

arsenite efflux pumps of both *arsB* and *acr3* types, and two *merT* mercury transport system genes were also present.

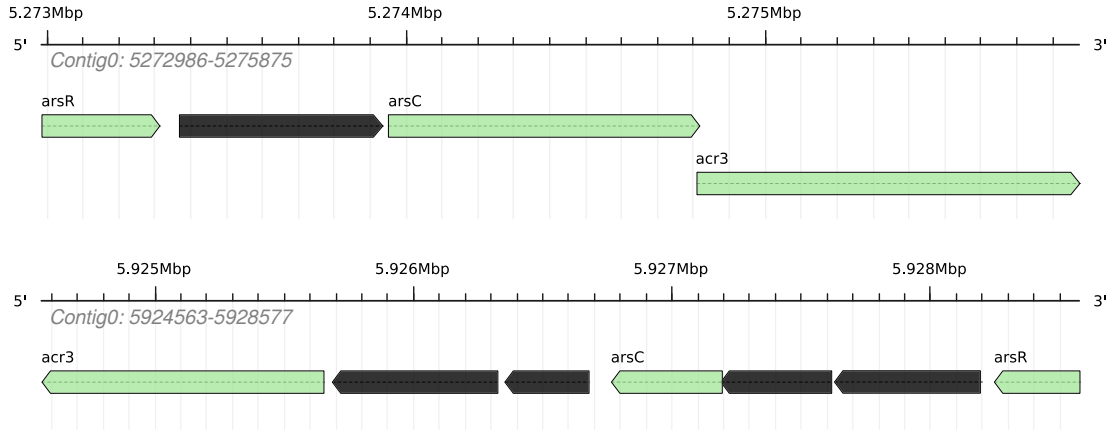


Figure 51.: Clusters of arsenic resistance genes in the assembled genome of *Mucilaginibacter sp. C110*. Green: arsenic resistance genes; black: genes with unknown function.

Arthrobacter sp. C19 Another strain with high arsenic resistance; its resistance to mercury, however, was rather modest in comparison to most other isolated strains, as it was capable of tolerating only 50 µM of mercury chloride.

Arsenic resistance determinants were represented mainly by *arsC* (arsenate reductase) genes, located two large clusters throughout the chromosome (Figure 52). A large gene encoding a cation efflux domain was found in one of the clusters; its role and function, however, remain unknown. One of the clusters had thioredoxin reductase gene *trxB* that can be beneficial in managing arsenic-caused oxidative stress. One additional *arsC* gene and an arsenite pump (of the *arsB* type) were located separately. For mercury resistance, similarly to the strain B14, only *merR* mercury resistance regulators were found.

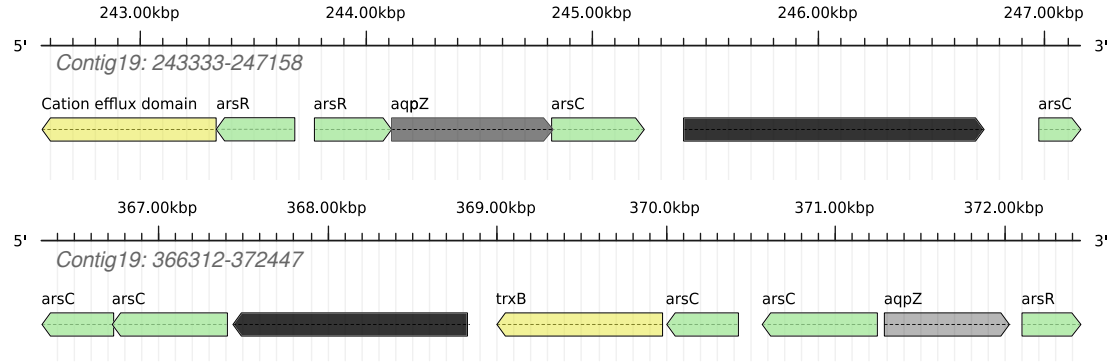


Figure 52.: Clusters of arsenic resistance genes in the assembled genome of *Arthrobacter sp. C19*. Green: arsenic resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Strains isolated from soot samples (F and FS)

Nocardioides sp. FS32 Found it the arsenic-rich soot, this strain had one of the highest metal(loid) resistances of all isolated strains, showing confident growth on TSA medium supplemented with either 25 mM of arsenite, 200 mM of arsenate, or 400 µM of mercury chloride. Its resistance machinery was distributed in two clusters (one for arsenic resistance genes, and combining arsenic and mercury resistance genes). The mixed-genes cluster contained an *arsRDABC* operon (regulatory/repressor gene *arsR*; metallochaperone *arsD*; arsenical pump-driving ATPase gene *arsA*; arsenite efflux pump *arsB* and arsenate reductase *arsC*), and a small *mer* operon (*merR* regulator/repressor; mercury(II) reductase gene *merA* and alkylmercury lyase *merB*), while the other cluster contained *arsC* genes and *acr3* arsenite efflux pump (Figure 53). Three copies of *arsA* ATPase, two *acr3* arsenite efflux pump genes, two arsenate reductase (*arsC*) genes, as well as four copies of *merA* mercury reductase and two copies of *merB* alkylmercury lyase were present outside of the clusters. No arsenic or mercury resistance genes were discovered in the putative plasmid.

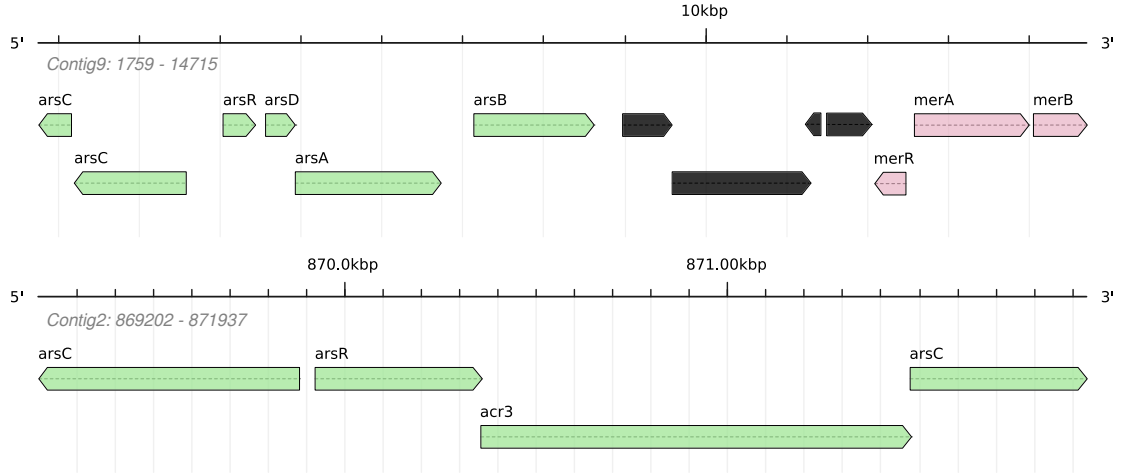


Figure 53.: Clusters of arsenic resistance genes in the assembled genome of *Nocardioides sp. FS32*. Green: arsenic resistance genes; pink: mercury resistance genes; black: genes with unknown function.

Bacillus sp. F14 Despite ranking among the most resistant of isolated strains in regard to arsenite (over 25 mM) and mercury (over 400 µM), detected arsenic and mercury detoxification systems were few in number. ATPase *arsA*, *acr3* and *arsB* arsenite efflux pumps, and arsenite methyltransferase-like protein genes were distributed around the chromosome. There was a small cluster of arsenic resistance genes containing *arsC* arsenate reductase and *acr3* arsenite efflux pump under regulation from the *arsR* regulator/repressor gene. With regards to mercury resistance, there were four merR regulatory genes located in different parts of the genome, with one of them adjustment to several hypothetical proteins that contained transport protein domains, but no clear indications of a *mer* operon. This, however, does not exclude its presence in the non-sequenced regions of the chromosome or its location in the plasmids.

Paraburkholderia sp. F12 Isolated from arsenic-rich soot, this strain was expected to have high metal(loid) resistance. In cultivation experiments, it has shown high resistance to mercury, tolerating over 300 μM of $HgCl_2$. Its resistance to arsenic, while still high, was lower than in some other strains, at over 10 mM of sodium arsenite, and 100 mM of sodium arsenate. On the genetic level, this strain possessed a large set of mercury resistance determinants, with four large clusters of *mer* genes, as well as two clusters of arsenic resistance genes and two clusters having both As an Hg resistance genes, several regulation/repression *merR* genes distributed throughout the chromosome, and an additional *arsB* (arsenite efflux pump) gene.

One of the clusters of mercury resistance genes found next to the arsenic resistance genes (Figure 54) had conventional *mer* operon layout, containing mercury reductase gene (*merA*), alkylmercury lyase *merB*, mercuric transporters *merC*, *merP* and *merT*, as well as broad-spectrum mercury transporter gene *merE*, but *mer*-specific regulatory genes. The rest of the mercury resistance gene clusters were highly irregular, containing non-related genes, genes with unknown function, genes encoding for heavy-metal-associated domains, or putative mercury transporters; there also was one instance of *garB* gene that is related to oxidative stress (Figure 55).

Genetic determinants of arsenic resistance contained arsenate reductase (*arsC*), arsenite efflux pumps (*arsB* and *acr3*), organoarsenical oxidase (*arsH*), as well arsenite oxidases *aioA* and *aioB*, metallochaperone *arsD* and regulatory genes *arsR*. Cadmium-induced *cadI* gene was also present.

Paraburkholderia sp. F21 Another strain isolated from arsenic-rich stupp, it could tolerate 20 mM of sodium arsenite and 400 μM of mercury chloride when cultivated on TSA medium. Its resistance to arsenate was lower than most other strains, at just 50 mM. On the genetic level, it possessed no fewer than five clusters of arsenic resistance genes and one large *mer* operon (Figure 56). Additionally, *arsC* (arsenate reductase) gene and arsenical pump-driving ATPase gene *arsA* together with metallochaperone *arsD* were located elsewhere on the chromosome.

Similarly to *Paraburkholderia* strain F12 (see above), mercury resistance genes cluster had seemingly non-related genes and genes encoding proteins with unknown function. Two genes encoding proteins with heavy-metal-associated domains were also present. Similarly to *Paraburkholderia* strain F12, arsenic resistance was provided by arsenate reductase (*arsC*), arsenite efflux pumps (*arsB* and *acr3*), organoarsenical oxidase (*arsH*), as well arsenite oxidases *aioA* and *aioB*, metallochaperone *arsD* and regulatory genes *arsR*. Cadmium-induced *cadI* regulatory genes were present among arsenic resistance genes as well.

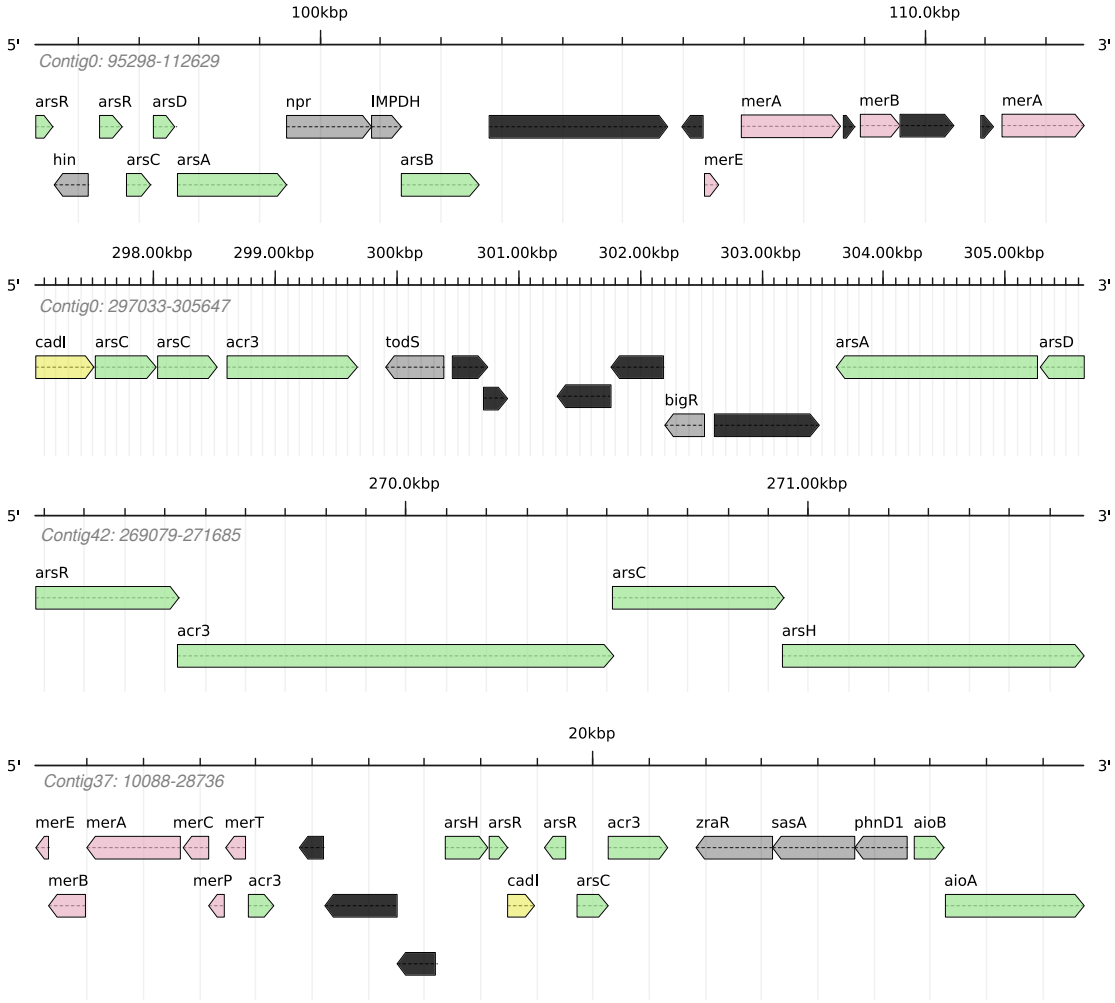


Figure 54.: Clusters of arsenic resistance genes in the assembled genome of *Paraburkholderia sp. F12.* Green: arsenic resistance genes; pink: mercury resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

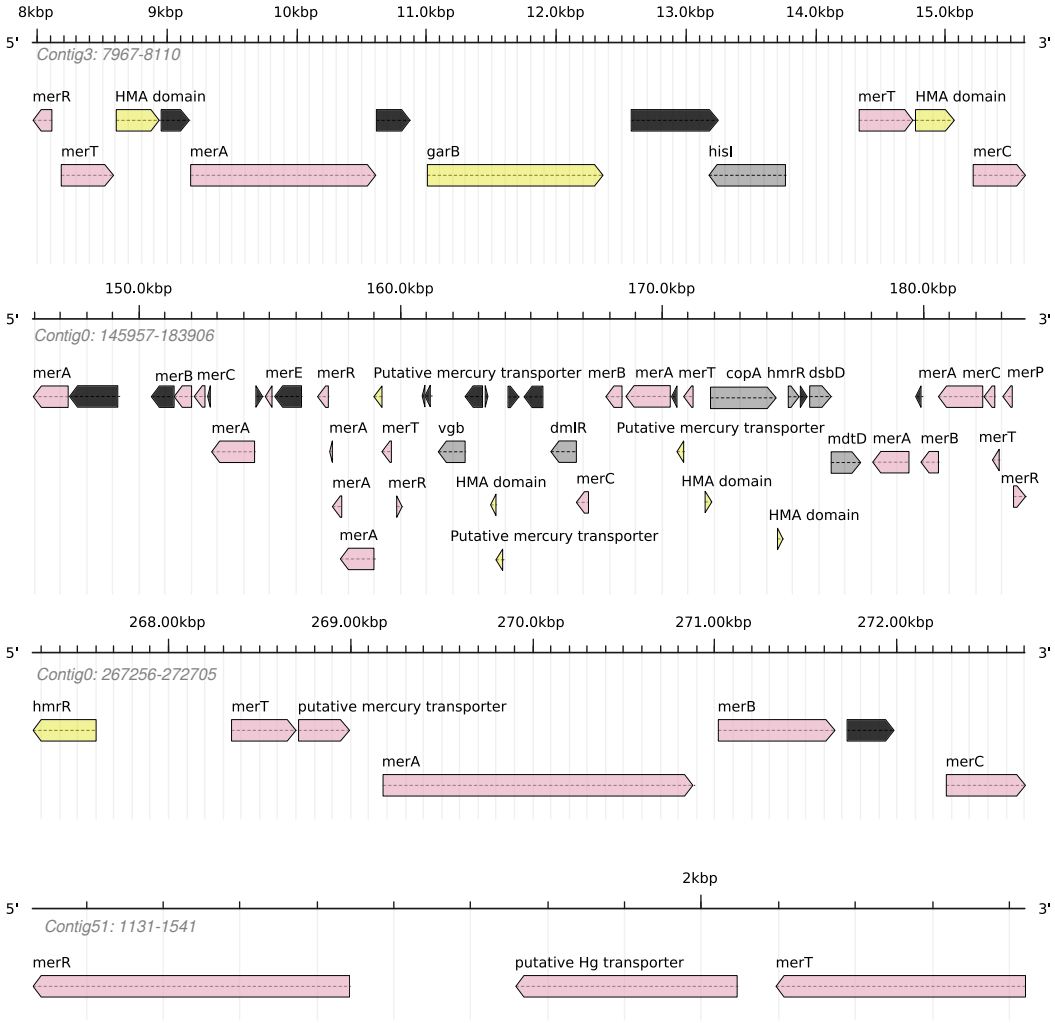


Figure 55.: Clusters of mercury resistance genes in the assembled genome of *Paraburkholderia sp. F12*. Pink: mercury resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

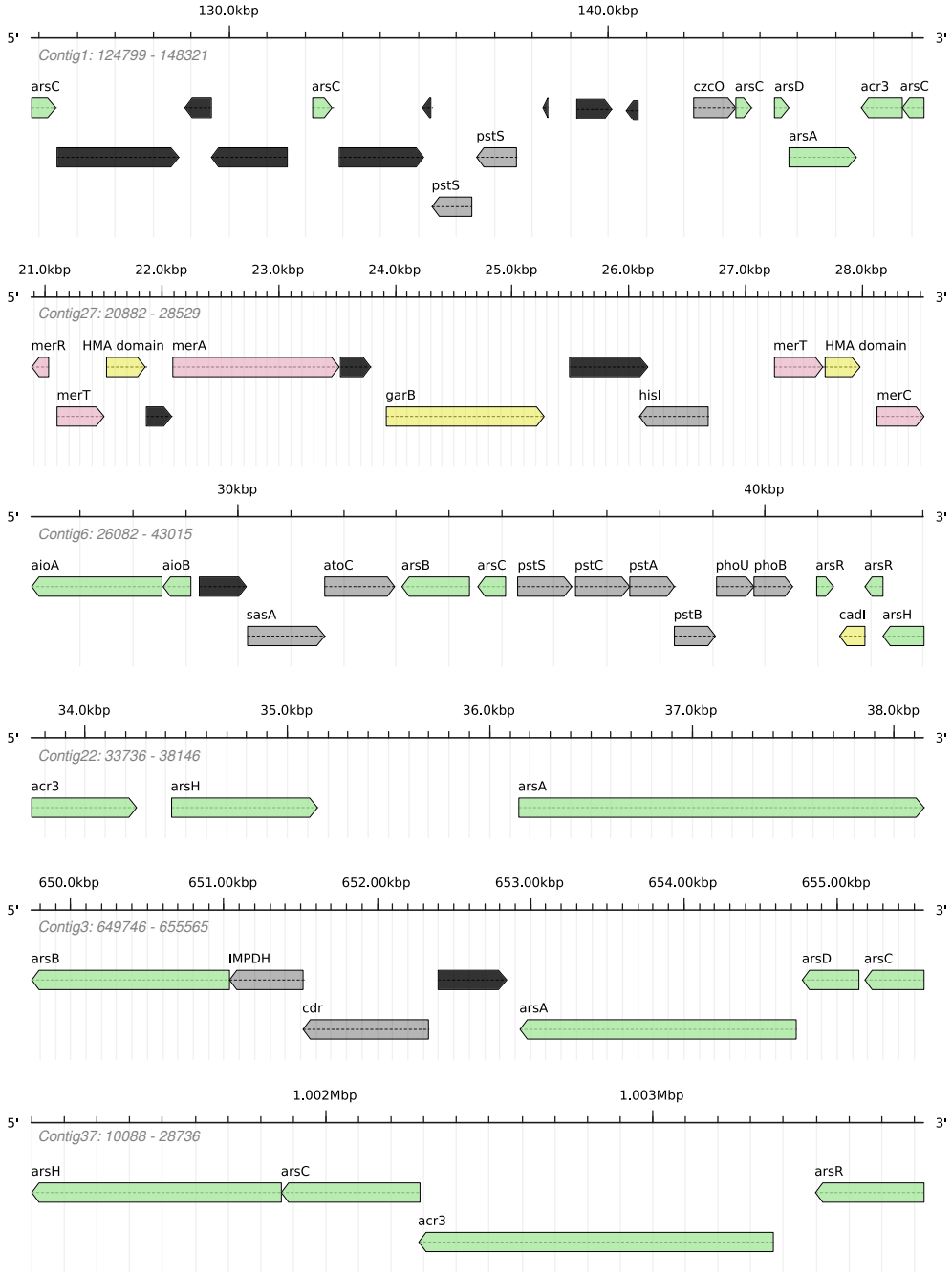


Figure 56.: Clusters of arsenic resistance genes in the assembled genome of *Paraburkholderia sp. F12*. Green: arsenic resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Strains isolated from groundwater (WB & WR) and groundwater sediments (SB & SR)

Massilia sp. WB1 One of the strains that showed no resistance to As(III), it was capable of growth in TSA medium supplemented with 20 mM of As(V). Its tolerance to mercruy was low as well, at just over 50 µM. Nonetheless, three clusters of arsenic resistance genes and one fairly small *mer* operon (Figure 57). Arsenic resistance gene clusters contained arsenate reductase *(arsC)* genes, regulatory genes *arsR*, and could contain organoarsenical oxidase gene *arsH.* All of them were either intersected by, or were located near an aquaporinZ gene. The rest of the genome contained arsenate reductase *arsC*, ATPase *arsA*, and arsenite efflux pump *acr3.*

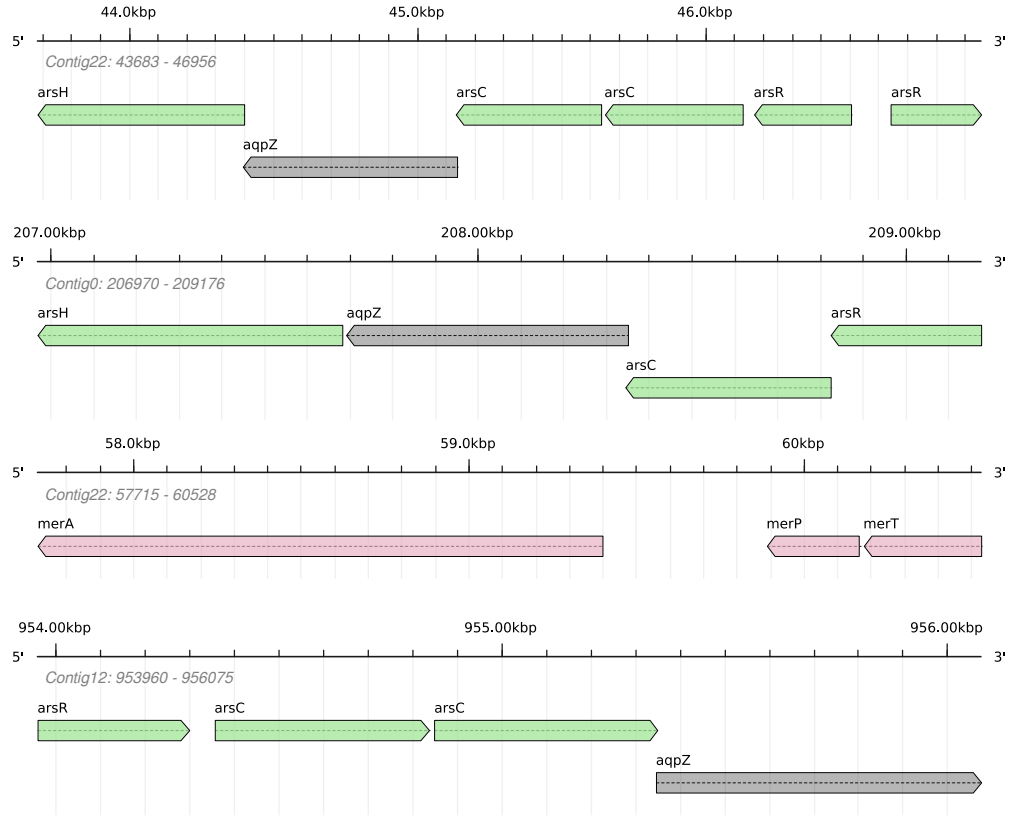


Figure 57.: Clusters of arsenic resistance genes in the assembled genome of *Massilia sp. WB1.* Green: arsenic resistance genes; pink: mercury resistance genes; grey: genes not directly related to metal(loid) resistance.

Pseudomonas sp. WB11 Another strain with low (compared to most other isolated strains) resistance to arsenite and mercury, it had high resistance to As(V), tolerating over 300 mM of sodium arsenate during cultivation. Its arsenic resistance genes were located in two short clusters containing arsenate reductase (*arsC*) genes, arsenite efflux pump gene *acr3*, organoarsenical oxidase gene *arsH* and arsenite methyltransferase gene *arsM* (Figure 58). Multiple *merR*-like regulator/repressor genes

were detected throughout the chromosome, but no genes originating from the *mer* operon were found.

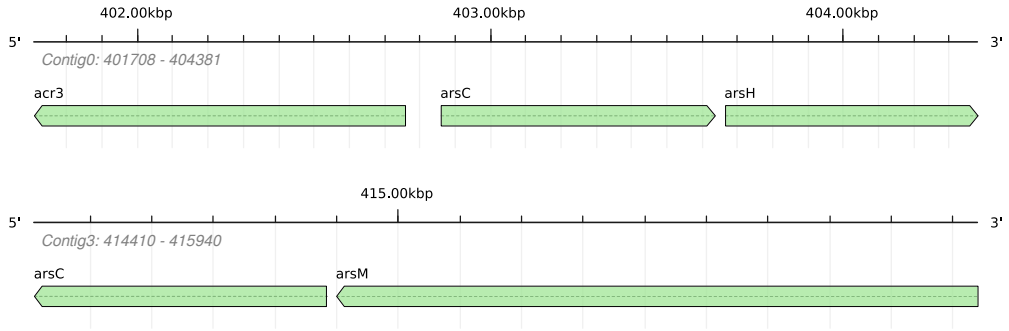


Figure 58.: Clusters of arsenic resistance genes in the assembled genome of *Pseudomonas sp. WB11.* Green: arsenic resistance genes.

Pseudomonas sp. WR7 This strain had high resistance to arsenate (tolerating over 300 mM of sodium arsenate) and mercury (growth in TSA medium supplemented with 400 µM of mercury chloride). It had a well-defined *mer* operon that included mercuric transport protein genes *merT* and *merP*, mercury transport protein *merC*, mercury reductase gene *merA*, regulatory gene *merD* and broad-spectrum mercury transporter gene *merE*, as well as one protein of unknown function. Three clusters of arsenic resistance genes (with two of them located near each other) contained arsenate reductase *arsC* genes, organoarsenical oxidase gene *arsH* and arsenite efflux pump gene *acr3*, regulatory genes *arsR* and arsenite oxidase genes *aio* (Figure 59). Additional arsenite efflux pump genes *arsB* and *acr3* were located elsewhere on the chromosome.

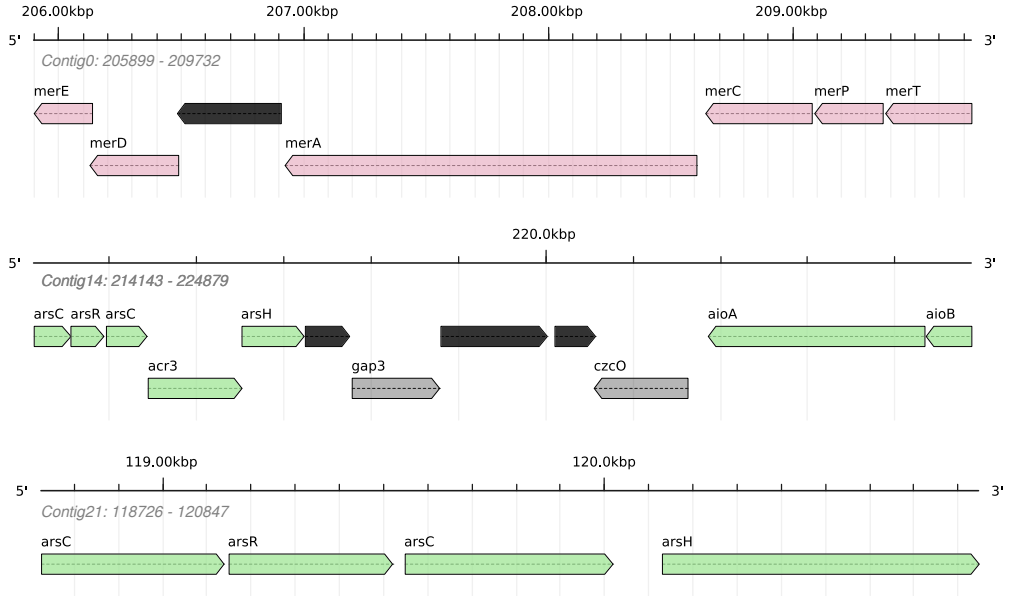


Figure 59.: Clusters of arsenic resistance genes in the assembled genome of *Pseudomonas sp. WR7.* Green: arsenic resistance genes; pink: mercury resistance genes; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Streptomyces sp. WB17 Demonstrating low resistance to arsenite and mercury (tolerating 2 mM and 50 µM of sodium arsenite and mercury chloride, respectively), it had very high resistance to sodium arsenate (over 300 mM). It had a number of arsenic and mercury resistance genes, with most of them forming several large gene clusters. Two arsenic resistance gene clusters had embedded thioredoxin reductase genes *trxA* and *trxB*; their activation could be helpful in dealing with arsenic-induced oxidative stress (Figure 60). In addition, several copies of *arsA* ATPase, *arsB* and *ACR3* arsenite efflux pumps, *arsC* arsenate reductase and *merA* mercury(II) reductase were distributed throughout the genome. No arsenic or mercury resistance genes were recovered from the plasmid assemblies.

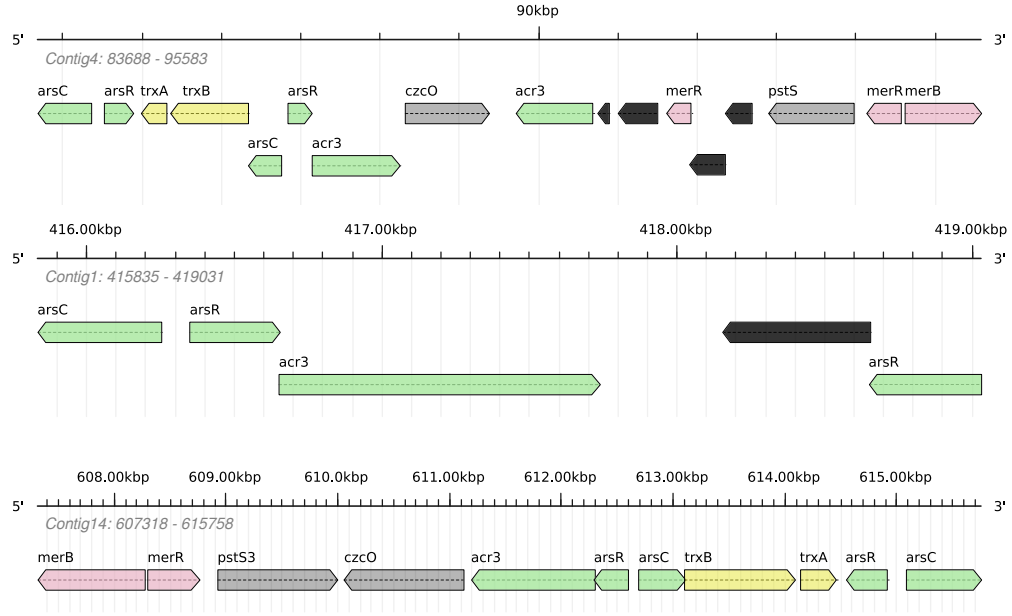


Figure 60.: Clusters of arsenic resistance genes in the assembled genome of *Streptomyces sp. WB17*. Green: arsenic resistance genes; pink: mercury resistance genes; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Bacillus sp. SR1 Tolerating 20 mM of sodium arsenite and 400 µM of mercury chloride, this strain had one cluster with arsenic resistance genes (Figure 61), as well *arsB* (arsenite efflux pump) and *arsC* (arsenate reductase) genes (two copies each) distributed throughout the chromosome. Interestingly, in addition to the *arsR* regulatory/repressor genes, *aseR* regulatory genes and *cadI* cadmium-induced genes were also present. Arsenate reductase (*arsC*) gene was found in one of the fragments of a putative plasmid. No mercury resistance genes from the *mer* operon were found neither in the chromosome nor in the partially assembled plasmids; their existence, however, can not be completely ruled out.

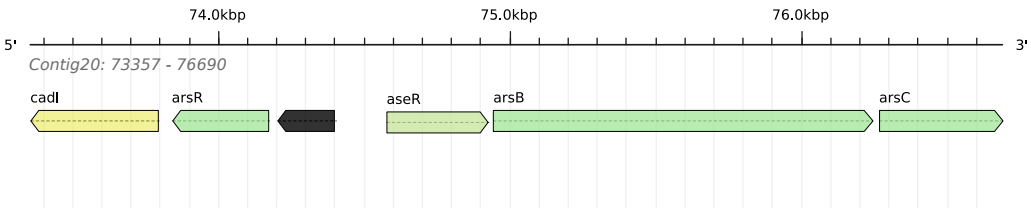


Figure 61.: Cluster of arsenic resistance genes in the assembled genome of *Bacillus sp. SR1*. Green: arsenic resistance genes (*aseR* regulatory genes are marked by a different shade of green); yellow: genes with possible role in metal(loid) resistance; black: genes with unknown function.

Bacillus sp. SR3 A strain highly resistant to both arsenic and mercury, it had one cluster of arsenic resistance genes, as well as one large cluster having both arsenic and mercury resistance genes (Figure 62). Arsenic resistance determinants included regulatory/repressor gene *arsR*; metallochaperone *arsD*; arsenical pump-driving ATPase gene *arsA*; arsenite efflux pump *arsB* and arsenate reductase *arsC*, as well as arsenic efflux pump *acr3*. Similarly to SR1 (see above) and unlike other sequenced strains, clusters of resistance genes included *aseR* regulation genes. Mercury resistance genes formed a *mer* operon containing mercury reductase gene (*merA*), alkylmercury lyase (*merB*), regulatory/repressor gene *merR* and broad-spectrum mercury transporter gene *merE*, and included two genes of unknown function, one of which was profiled to have s heavy-metal-associated domain. The operon was adjustent to the oxidative stress-related thioredoxin reductase gene *trxB*.

Two additional copies of *arsB* arsenite efflux pump, and one *acr3* arsenite efflux pump, as well as several regulatory genes (*merR* and *arsR*) were also present at various locations throughout the chromosome.

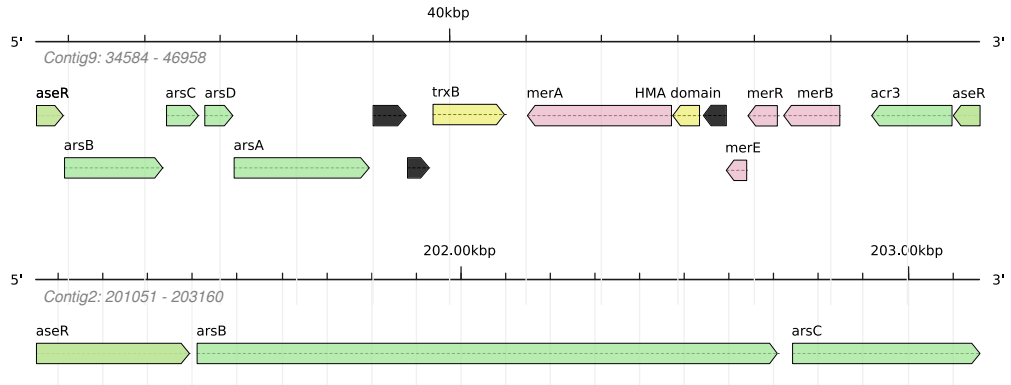


Figure 62.: Clusters of arsenic resistance genes in the assembled genome of *Bacillus sp. SR3*. Green: arsenic resistance genes (*aseR* regulatory genes are marked by a different shade of green); pink: mercury resistance genes; yellow: genes with possible role in metal(loid) resistance; grey: genes not directly related to metal(loid) resistance; black: genes with unknown function.

Delftia sp. SB1 This strain had high resistance to both forms of arsenic, tolerating over 20 mM of arsenite and 300 mM of arsenate. Its mercury resistance was lower than most other isolated strains, at just 50 µM of mercury chloride. Two clusters containing incomplete *ars* operons were found, as well as one *acr3* gene located elsewhere on the chromosome. Genes from the *mer* operon were not found, although several *merR* regulatory genes were present (Figure 63).

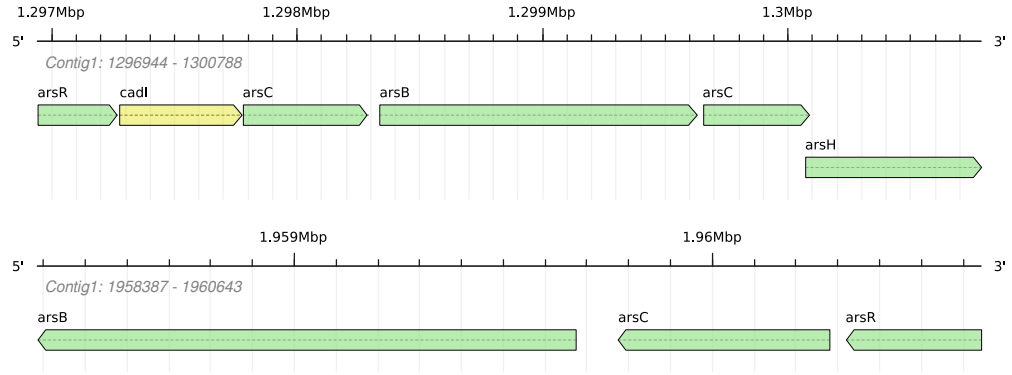


Figure 63.: Clusters of arsenic resistance genes in the assembled genome of *Delftia sp. SB1.* Green: arsenic resistance genes; yellow: genes with possible role in metal(loid) resistance.

Bacillus sp. SB21 With average (compared to other isolated strains) levels of resistance to both forms of arsenic and mercury, the genome of this strain contained *arsA* ATPase, *arsB* and *acr3* arsenite efflux pump, arsenite methyltransferase-like protein, and several regulatory *arsR* genes located in different parts of the genome. In addition, a small cluster containing *arsC* arsenate reductase and *acr3* arsenite efflux pump under control of the *arsR* regulator/repressor gene was discovered (Figure 64). A number of *merR*-like regulatory genes were found in different contigs, but no *mer*-operon-derived genes were discovered. While high fragmentation of the assembled genome could have obfuscated some of the resistance genes, their amount and location in the chromosome corresponds very well to the *Bacillus* strain F14 which has very similar genome that was less fragmented upon assembly. No arsenic or mercury resistance genes were recovered from the assembled fragments of putative plasmids.

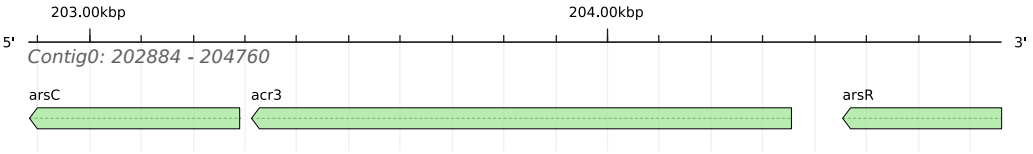


Figure 64.: Cluster of arsenic resistance genes in the assembled genome of *Bacillus sp. SB21.* Green: arsenic resistance genes.

Chryseobacterium sp. SB22 Despite having high levels of resistance, metal(loid) resistance genes were low in number. An arsenate reductase gene *arsC*, two copies of *acr3* arsenite efflux pump and *merC* mercury transporter, as well as several regulatory genes (*arsR* and *merR*) were distributed throughout the genome. Additionally, a hypothetical protein with Hg-scavenging domain was located adjustment to one of the regulatory genes, but its exact role and function remain unknown.

5.11.4. Detection and analysis of antibiotic resistance genes

Genome-wide search for antibiotic resistance genes

Analysis of genome assemblies with RGI revealed that 13 out of 17 sequenced strains had functional antibiotic resistance genes. There were 38 detected genes; none of them had a perfect sequence similarity to the known resistance genes. Most of them (27 genes; 71%) had high similarity with resistance-nodulation-cell division antibiotic efflux pump *adeF*, and were predicted to be effective against fluoroquinolone and tetracycline antibiotics. Other resistance genes included *abaQ* major facilitator superfamily antibiotic efflux pump (2 genes), *fosA* and *fosB* fosfomycin thiol transferases (3 genes), two different beta-lactamase genes, macrolide phosphotransferase gene (*mphL*), glycopeptide resistance cluster, as well as *rsmA* antibiotic efflux pump and *soxR* system (Table 32).

Table 32.: Clinically relevant antibiotic resistance genes.

Strain	Gene identifier	RGI hit	Identity, %	Drug classes	Resistance mechanism
B14	LAHEKKBB_02830	*adeF*	66.19	Fluoroquinolone; Tetracycline	Efflux
	LAHEKKBB_03118	*adeF*	43.85	Fluoroquinolone; Tetracycline	Efflux
	LAHEKKBB_03328	*abaQ*	72.1	Fluoroquinolone	Efflux
	LAHEKKBB_04448	*adeF*	41.6	Fluoroquinolone; Tetracycline	Efflux
C110	ACBIKEBC_00657	*adeF*	42.08	Fluoroquinolone; Tetracycline	Efflux
C22	NDGNHEOH_02553	*adeF*	42.78	Fluoroquinolone; Tetracycline	Efflux
	NDGNHEOH_05201	*adeF*	70.22	Fluoroquinolone; Tetracycline	Efflux
F12	AHAGGIMI_01325	*adeF*	80.55	Fluoroquinolone; Tetracycline	Efflux
	AHAGGIMI_06343	*adeF*	42.71	Fluoroquinolone; Tetracycline	Efflux
F14	OHECGCDO_02857	*fosB*	99.28	Fosfomycin	Inactivation
F21	OJCGJKFE_04304	*adeF*	44.62	Fluoroquinolone; Tetracycline	Efflux

Table 32.: Clinically relevant antibiotic resistance genes.

Strain	Gene identifier	RGI hit	Identity, %	Drug classes	Resistance mechanism
SB1	OJCGJKFE_05485	*adeF*	80.55	Fluoroquinolone; Tetracycline	Efflux
	OJCGJKFE_06651	*adeF*	42.07	Fluoroquinolone; Tetracycline	Efflux
	KACNDILH_00007	*adeF*	42.59	Fluoroquinolone; Tetracycline	Efflux
	KACNDILH_03938	*adeF*	57.97	Fluoroquinolone; Tetracycline	Efflux
	KACNDILH_05226	*adeF*	70.61	Fluoroquinolone; Tetracycline	Efflux
SB21	BMMGMKON_02280	*fosB*	99.28	Fosfomycin	Inactivation
	BMMGMKON_03156	*BcI*	96.08	Cephalosporin; Penam	Inactivation
	BMMGMKON_04654	*mphL*	89.26	Macrolide	Inactivation
SB22	DFIBHECD_01639	*adeF*	41.31	Fluoroquinolone; Tetracycline	Efflux
	DFIBHECD_02955	*CGB-1*	93.78	Carbapenem; Cephalosporin; Penam	Inactivation
	DFIBHECD_04429	*adeF*	41.2	Fluoroquinolone; Tetracycline	Efflux
SR3	BOLFLJFK_02486	*vanRF*	93.51	Glycopeptide	Target Alteration
WB1	PCAHBMBF_00249	*adeF*	42.68	Fluoroquinolone; Tetracycline	Efflux
	PCAHBMBF_01010	*adeF*	66.76	Fluoroquinolone; Tetracycline	Efflux
	PCAHBMBF_02227	*adeF*	48.95	Fluoroquinolone; Tetracycline	Efflux
	PCAHBMBF_05322	*adeF*	69.09	Fluoroquinolone; Tetracycline	Efflux
	PCAHBMBF_05325	*adeF*	68.21	Fluoroquinolone; Tetracycline	Efflux

Table 32.: Clinically relevant antibiotic resistance genes.

Strain	Gene identifier	RGI hit	Identity, %	Drug classes	Resistance mechanism
WB11	LDNCCFPP_01266	*adeF*	47.5	Fluoroquinolone; Tetracycline	Efflux
	LDNCCFPP_01736	*rsmA*	93.44	Fluoroquinolone; Diaminopyrimidine; Phenicol	Efflux
	LDNCCFPP_02023	*adeF*	41.84	Fluoroquinolone; Tetracycline	Efflux
WR7	MPOPHJGJ_01289	*abaQ*	72.81	Fluoroquinolone	Efflux
	MPOPHJGJ_02974	*soxR*	67.61	Fluoroquinolone; Cephalosporin; Glycylcycline; Penam; Tetracycline; Acridine Dye; Rifamycin; Phenicol; Triclosan	Target alteration; Efflux
	MPOPHJGJ_04149	*fosA*	70.59	Fosfomycin	Inactivation
	MPOPHJGJ_04184	*adeF*	41.68	Fluoroquinolone; Tetracycline	Efflux
	MPOPHJGJ_05285	*adeF*	42.99	Fluoroquinolone; Tetracycline	Efflux
	MPOPHJGJ_05360	*adeF*	42.9	Fluoroquinolone; Tetracycline	Efflux
	MPOPHJGJ_05505	*adeF*	66.03	Fluoroquinolone; Tetracycline	Efflux

Search for antibiotic resistance genes in integrative and conjugative elements

Analysis of 18 ICE sequences from 13 strains[10] with RGI has shown that none of the sequenced strains had genes with high similarity to the known antibiotic resistance genes (no Perfect or Strict hits, in RGI's terminology). Functional state, affinity and selectivity towards antibiotics, and efficiency of resistance mechanisms encoded by those genes at coping with clinically employed concentrations of antimicrobial agents thus remains unknown.

[10]C19, C22, F12 (2 ICEs), F14, F21 (2 ICEs), FS32 (3 ICEs), SB1, SB21, SR1, SR3, WB1, WB11, WB17 92 ICEs). Other sequenced strains either had no ICEs, or had ICEs of unknown type.

A total of 102 genes with some similarity to the known antibiotic resistance genes (Loose hits, in RGI's terminology) were identified.

Most of them (73 genes; 71.5%) were predicted to act through antibiotic efflux; 17 genes would act by target alteration, further 7 - by reducing permeability to antibiotic molecules, and 6 by antibiotic inactivation. Target replacement and target protection functions were a minority (Figure 65).

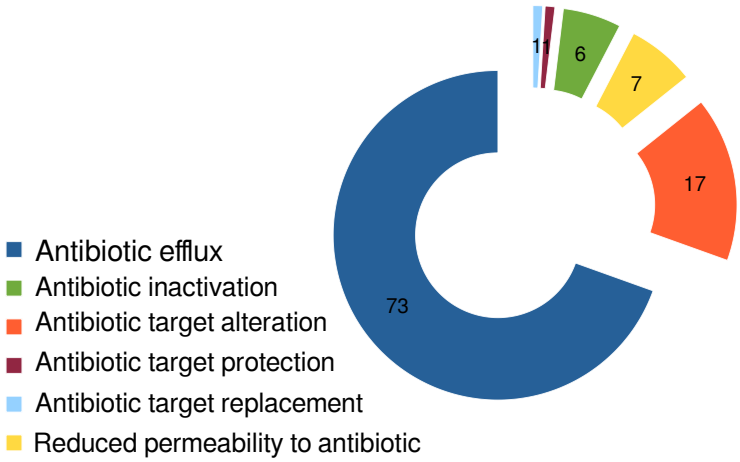


Figure 65.: Actuation mechanisms of antibiotic resistance genes identified in the ICEs.

Acting against tetracycline, fluoroquinolone, macrolides, penams (such as β-lactam antibiotics) was predominant; action against cephalosporins, phenicols, cephamycin, penem and carbapenem antibiotics„ monobactam and peptide antibiotics was also common. Genes similar to those providing resistance against polyamine, ethionamide, nucleosides, sulfonamides, fusidic acid, pleuromutilin, oxazolidinone, streptogramin, lincosamide and antibacterial free fatty acids were the least common (Figure 66).

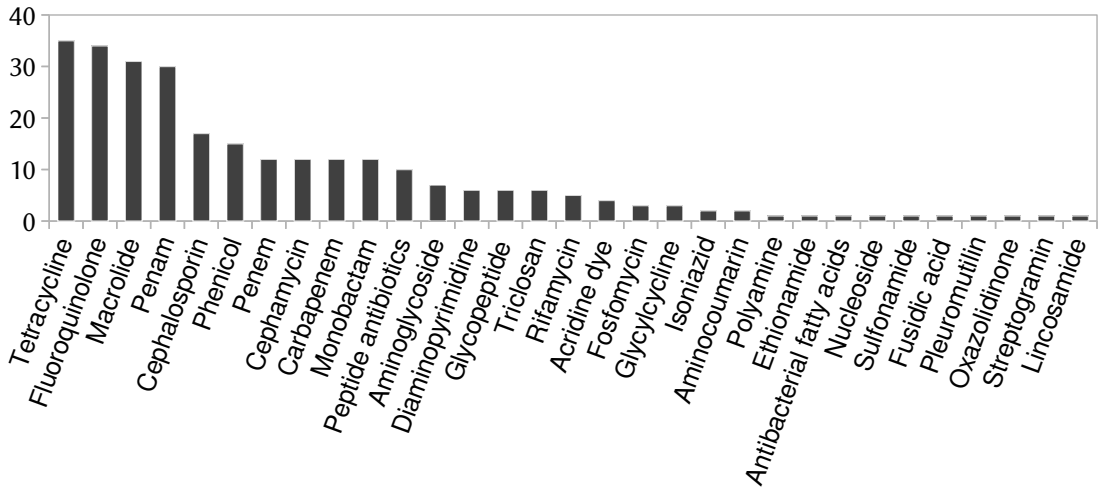


Figure 66.: Spectrum of antibiotic resistance genes identified in the ICEs of resistant strains.

6. Discussion

Contamination levels

Arsenic and mercury concentrations in soil (samples A, B & C), soot (F & FS) and flue dust (E) were very similar to that obtained previously by Gallego et al. (2015). Unlike their results, recovered samples of stupp (D) had lower arsenic and higher mercury content. Such a discrepancy, most likely, is a result of a somewhat different sampling strategy (recovering stupp from deeper, less weathered layers, and avoiding soot inclusions). In line with all previous studies (Loredo et al., 1999; Ordóñez et al., 2013; Gallego et al., 2015; González-Fernández et al., 2018), it should once again be stressed that El Terronal mining site continues to be a source of severe arsenic and mercury pollution.

Biodiversity and microbial populations of the site

When analysing different components of diversity for both prokaryotic and fungal (and SAR) communities, *richness* in soil samples (A, B & C) was shown to be on the similar level regardless of the differences in arsenic and mercury concentration. This is broadly comparable to the data from arsenic-contaminated soils (Gu et al., 2017; Simmler et al., 2019). Estimations of *evenness*, however, painted a more complex picture: similarly to the study by Gu et al. (2017), no appreciable differences in *evenness* were detected using the same evenness metrics (Pielou's J); in fact, evenness levels of soils and significantly less diverse flue dust and sediments were similar to the soil levels, a likely artefact resulting from dependence on the sample size (Mccune & Grace, 2002). Study by Simmler et al. (2019) used a different evenness metrics (^{1}E) that was better suited for metagenomic datasets (Lucas et al., 2017); somewhat counter-intuitively, it indicated higher evenness in more contaminated soils, which could be explained by environmental pressures disproportionally affecting predominant members of the community. However, it should be noted that similarly to Pielou's J, ^{1}E index is not completely free of the sampling effects as it is derived from the exponent of the Shannon index and species richness. In our study, ENS_{PIE} metrics was used to estimate evenness; it is based on calculating Effective Number of Species from the inverse Simpson index ($1 - D$), and thus is mostly independent of the sample size, unless sampled organisms are not spatially random (Chase & Knight, 2013); this problem is completely avoided in metagenomic datasets, although it should be taken into account at the time of sampling. This metrics revealed significant differences between samples, with more contaminated samples (such as soil samples C and flue dust samples E which had a similar community structure) being more uneven. This suggests that increasing arsenic and mercury contamination levels in soil did not radically alter the structure of microbial community by precluding

significant number of organisms from surviving, but instead simply favoured some organisms over the others, most likely depending on their resistance to arsenic and mercury. Richness of microbial communities in flue dust (E) and groundwater sediments (SB & SR), while somewhat lower than in soils, was still high; however, their structure was very uneven compared to the soil samples. In contrast, high-grade metallurgical waste (arsenic-rich soot F & FS, and stupp D) had considerably less diverse communities heavily dominated by a very small number of strains.

Taxonomically, prokaryotic microorganisms found in contaminated soils (A, B & C) and flue dust (E) were very similar to those found by other authors in arsenic-polluted soils (Gu et al., 2017; Wu et al., 2016; Narendrula-Kotha & Nkongolo, 2017) and mining waste undergoing natural attenuation (Bertin et al., 2011; J.-l. Liu et al., 2019). Similarly to those studies, Proteobacteria, Actinobacteria and Acidobacteria were the most abundant phyla, followed by Bacteroidetes, Chloroflexi, Planctomycetes, Gemmatimonadetes, Nitrospirae and Firmicutes. On higher taxonomic levels, this community structure was also similar to the soil contaminated by other metals and metalloids such as abandoned non-ferrous metal(loid) tailings(J.-l. Liu et al., 2019) and soils from a chromite mine area (Pradhan et al., 2020). Fungal communities in the soil were represented by Ascomycota, Basidiomycota and Zygomycota, similarly to the work of Narendrula-Kotha and Nkongolo (2017) performed on arsenic-polluted soils in nickel and copper mining areas in Canada.

Sediment samples taken from two different locations differed both from the soil and mining waste samples and between each other. At higher taxonomic levels, Proteobacteria, Bacteriodetes, Acidobacteria and Chloroflexi were among the most abundant phyla, while minority groups differed a lot; this is broadly comparable to the results from the study of arsenic-contaminated freshwater sediments, with some of them showing significant changes in community structure of less abundant bacteria depending on location or season (Halter et al., 2011; Cavalca et al., 2019; W. Li et al., 2020). Unexpectedly, functional metagenomic predictions demonstrated that metabolism of microbial communities in two groundwater sediment samples from El Terronal was very similar, with most of the differences explained by much higher proportion of *Archaea* in the sample SB (with higher abundance of genes responsible for Archaea-specific flavin and phosphopantothenate biosynthesis) and higher proportion of Gram-negative Proteobacteria in the sample SR (UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis). Predictions for higher abundance of genes responsible for pyridoxal 5'-phosphate biosynthesis I pathway and CMP-legionaminate biosynthesis I pathway in the sample SB were attributed to Subgroup 6 of Acidobacteria and (predominantly) *Mariprofundus spp.* of Zetaproteobacteria, respectively. Their biological significance remains unclear.

Samples of arsenic-rich soot and stupp were distinct from soils. Remarkably, *Archaea* dominated in prokaryotic communities inhabiting stupp waste heap (sample D). This sets it apart from the rest of the samples mostly populated by *Bacteria*. It suggests that stupp is even less hospitable to life than (already extremely hostile) arsenic-rich soot (samples F & FS), with its higher mercury content and presence of PAHs

and organomercuric compounds (Gallego et al., 2015) preventing most bacteria from colonising it. Conversely, fungal communities of both soot and stupp waste heaps were phylogenetically similar, and were dominated by just a few species of Dothideomycetes fungi.

Samples taken from the same arsenic-rich soot waste heap at different times (winter: sample F; summer: sample FS) had significant differences in composition of their prokaryotic communities, with sample F being more diverse, and dominated by Rhodospirillales and Burkholderiales of Proteobacteria, while sample FS was divided very evenly between Rhodospirillales *(*of Proteobacteria*)* and Flavobacteriales *(*of the phylum Bacteroidetes). Both samples, however, shared a significant proportion of ASVs belonging mostly to Bacillales, Sphingobacteriales and Pseudomonadales. This hints at presence of a core community with strong seasonal changes; however, comparing more samples over time would be required to prove that with any degree of certainty. Additionally, four ASVs were shared between all high-grade waste samples (F, FS & D). Of those ASVs, one belonged to *Archaea* and two belonged to *Bacillu*s, a genus that is known to have hyper-resistant species (Niane et al., 2019; Aguilar et al., 2020). The remaining ASV belonged to the genus *Stenotrophomonas*; one of the members of this genus was recently isolated from an abandoned arsenic mine and has proven to be exceptionally resistant to arsenic (Bermanec et al., 2021).

As was evidenced by beta-diversity studies of both 16S and 18S rRNA genes and predicted KO metagenomes, differences in composition of microbial communities correlated mainly with sample type. Individual contribution of increasing arsenic and mercury concentrations, while not completely insignificant, could not fully explain those differences. This is similar to conclusions of other studies employing metagenomic methods to study microbial communities in arsenic-contaminated soils, such as soils contaminated by irrigation (Bangladesh and China) and arable soil with high levels of geogenic arsenic from the United Kingdom(Gu et al., 2017) and floodplain soils contaminated with mining waste (Simmler et al., 2019).

Isolated strains and comparison with metagenomic data

As mentioned in the section 5.4, isolation was attempted for 64 strains, but only 59 of them were once isolated and re-seeded on a fresh agar plate. There are several possible reasons for those strains dying off, such as being able to grow only in the presence of 'helper' strains due to reliance on their metabolic cycle or signalling, or cultivation media not being wholly suitable for their growth (Vartoukian et al., 2010; Stewart, 2012). Majority of isolated bacteria belonged to either Proteobacteria, Firmicutes or Actinobacteria; a small minority belonged to Bacteroidetes. Representatives of the genera *Bacillus*, *Pseudomonas*, Arthrobacter and *Pseudoarthrobacter* were isolated from soil and sediments as well as mining waste; this is similar to arsenic- or mercury-contaminated mining sites around the world such as arsenic-contaminated soils near an abandoned tungsten-tin mining area in Salamanca province, Spain (Valverde et al., 2011), mercury-contaminated aquatic sediments near artisanal small-scale gold mining activities in Senegal (Niane et al., 2019) and arsenic-contaminated soils at a gold

mining site in Brazil (Aguilar et al., 2020). Unlike the aforementioned studies, a high proportion of *Streptomyces* bacteria was also isolated. In our study, proportion of isolated Proteobacteria in soil and (somewhat similar) flue dust was increasing with higher arsenic and mercury concentrations; this runs contrary to the findings by Valverde et al., where Firmicutes were more abundant among strains isolated from more contaminated soils. Such differences can be attributed to isolation bias (relatively rich Nutrient Agar medium supplemented with arsenic was used in their study; a diluted (1:10) Triptic Soy Agar was used in our work).

Study of only cultivable bacteria, however, does not reveal the true biodiversity of the microbial communities (Razumov, 1932; Staley & Konopka, 1985; Pace et al., 1986). Despite recent advances of using nutrient-poor media and high-throughtput cultivation methods such as dilution-to-extinction technique (Connon & Giovannoni, 2002; Bartelme et al., 2020), using 'helper' strains (Ohno et al., 2000; Ge et al., 2016), signal molecules extracted from them (Nichols et al., 2008) or systems simulating helper strain action (Guzman et al., 2019), as well as directly mimicking natural bacterial habitats (Kaeberlein, 2002; Chaudhary et al., 2019) and using microfluidic devices (Ma et al., 2014; Watterson et al., 2020), it is now widely accepted that majority of bacteria in any given environment remains uncultured (Pedrós-Alió & Manrubia, 2016; Lloyd et al., 2018). Therefore, a proper study of microbial diversity should always involve metagenomic methods.

While cultivation of most bacteria in the sample, even if possible, would be both time-consuming and impractical, even obtaining a set of cultured strains that can represent (in the broadest of terms) the overall diversity of the microbial community, remains mostly out of reach as well. In our study, high amount of isolated Proteobacteria and Actinobacteria corresponded to the biggest taxonomic groups revealed by the metagenomic approach, especially for river bank soil (C), flue dust (E) and arsenic-rich soot sample F, while Firmicutes strains that were a minority in all samples except arsenic-rich soot samples (F and FS) were isolated in a disproportionally large amount. Meanwhile, strains from large taxonomic groups such as Acidobacteria, Chloroflexi, Planctomycetes, Gemmatimonadetes and Nitrospirae that, when taken together, represented over 50% of all bacteria in soil samples, remained unisolated.

Arsenic resistance

Most of the isolated strains could tolerate very high levels of both As(III) and As(V), as well as mercury chloride. The most common arsenic detoxification mechanism as detected by PCR was arsenite efflux, delivered mostly by As(III) efflux pumps of the ACR3 family. Remarkably, almost half of isolated strains had positive amplification of respiratory arsenic oxidase *aio*-like genes, indicating capacity to utilise arsenite as an electron donor for respiratory processes. Arsenite oxidase genes were found in the strains belonging to all phyla, but majority of them (13 out of 20) belonged to the Proteobacteria. On the other hand, dissimilatory arsenate reductase (*arrA*) genes were found only in Proteobacteria. Of those, *Paraburkholderia* strains E12, F21 and F24, *Pseudomonas* strains WB11 and WR5, as well as *Massilia* strain WB1 had both *aio*-like

and *arrA* genes, suggesting their capacity to switch from arsenite oxidation to arsenate reduction depending on the environmental conditions.

Bacterial genome sequencing allows direct analysis of genetic determinants of arsenic, mercury and antibiotic resistance. To achieve this goal, genomes of 17 strains were sequenced (see section 5.10). Their analysis revealed a number of fairly conventional *ars* operons, as well as *acr3* and *aio* genes. Most sequenced strains had multiple copies of ars operon, as well as several isolated arsenic detoxification genes (most often - arsenite efflux pump gene *acr3* and arsenate reductase *arsC*) elsewhere on the chromosome. As(III) detoxification was performed mostly by actively transporting it out of the cell via ACR3 or ArsB arsenite pumps. As(V) was reduced to As(III) by arsenite reductase ArsC and expelled from the cell using the same route. In addition, *arsH* genes encoding organoarsenical oxidase were also present in most sequenced genomes. In contrast, *arsM* gene encoding As(III) S-adenosylmethionine methyltransferase was found only once, in the *Pseudomonas sp.* WB11 isolated from groundwater with the lowest arsenic content. Detection of dissimilatory arsenate reductase (*arrA*) from genome sequences alone can be difficult due to its similarity with other genes encoding 4Fe-4S ferrodoxin-type domains. Regulation of *ars* operon expression was performed via *arsR* regulation/repression genes; however, in *Bacillus spp.* SR1 and SR3, ars operon flanked by *aseR* genes similarly to some previously described *Bacillus* strains (Moore et al., 2005), with *arsR*-like genes located elsewhere on the chromosome. Presence of *cadI* cadmium-activated gene at or near *ars* operons was also quite common; while this association was described previously (Hotter et al., 2001), its biological significance remains poorly studied.

None of the isolated strains were fully susceptible to all forms of arsenic, since strains that were not resistance to arsenite such as *Mucilaginibacter sp.* C110 and *Massilia sp.* WB1 were capable of growing on TSA media supplemented with arsenate, and possessed numerous arsenic resistance genes in their genomes, suggesting that their detoxification systems were primarily aimed at As(V) reduction and removal.

On the microbial community level, functional predictions of arsenic detoxification and metabolism genes have shown a clear trend towards higher abundance of detoxification genes (arsenite efflux pump *arsB*, arsenate reductase *arsC*, and organoarsenical oxidase *arsH*) in communities inhabiting environments with higher levels of arsenic concentration. On the other hand, functional predictions for *arsR*-type regulatory/suppressor genes followed an inverse trend, with predicted metagenomes of communities inhabiting arsenic-rich soot (samples F and FS) having the lowest abundance of *arsR*-type genes, suggesting a shift towards constitutive expression of the arsenic resistance genes. This can be partially corroborated by the absence of the regulatory genes in some of the *ars* operons detected in the sequenced genomes of strains *Variovorax sp.* C22 and *Paraburkholderia sp.* F21), similarly to some of the previously described hyper-resistant strains (Koechler et al., 2015; Puopolo et al., 2020). However, in order to determine constitutive expression with a high degree of certainty a further transcriptomic and proteomic analysis is required.

Mercury resistance

Majority of isolated strains possessed high degree of mercury resistance as identified by their ability to grow on TSA medium supplemented with mercury chloride. Identification of mercury resistance determinants by PCR amplification of mercury(II) reductase gene *merA* yielded very few positive results, most likely due to low affinity of the used primer set to some of the types of the *merA* genes found in prokaryotes (Oregaard & Sørensen, 2007).

Genome sequencing was more successful, identifying either entire *mer* operons or *mer* operon-derived genes in 10 out of 17 sequenced strains. Mercury reductase gene *merA* was the most common; alkylmercury lyase *gene merB* was present in 5 out of 17 sequenced strains. Mercury reductase genes were not found in seven strains, of which four (*Pseudomonas sp.* B14 and *Bacillus spp.* F14, SR1 and SB21) had very high mercury resistance. One possible explanation is that *mer* operon in those strains is present, but was residing in the part of the genome that was not successfully sequenced, or in a plasmid. Another possible explanation is presence of very divergent mercury resistance genes. In many cases (such as *Paraburkholderia* spp. F12 and F21) *mer* operons contained genes with unknown function that could be a part of the mercury detoxification system as well. Further study of those genes is required.

On the community level, functional metagenome predictions has shown a simple trend of higher abundance of mercury resistance genes in environments with higher mercury concentration. This comparison was performed only for soils (A, B & C), sediments (SR & SB) and flue dust (E), as samples of arsenic-rich soot (F and FS) had significant inter-sample differences, and veracity of metagenomic predictions for the stupp sample (D) could not be verified due to the lack of cultured strains.

Antibiotic resistance and its relationship with metal(loid) resistance

In our study, high number of multi-resistant strains was detected, with 22 strains having MAR above the (somewhat arbitrary) cut-off of 0.2 denoting high-risk strains Krumperman (1983), Blasco et al. (2008) and Meng et al. (2020). While biological necessity of resisting multiple antibiotics remains unclear, one possible explanation is the presence of high proportion of antibiotic-producing *Streptomyces* bacteria in the soils that are both capable of resisting the compounds they produce, and force other bacteria to adopt various strategies of antibiotic resistance Laskaris et al. (2010) and de Lima Procópio et al. (2012). This coincides with two of the cultured *Streptomyces* strains (A12 and E14) being capable to inhibit growth of a susceptible *E. coli* strain.

While co-selection of antibiotic and heavy metal and metalloid resistance is well-documented (Baker-Austin et al., 2006), many of those observations come from environments that were affected by both heavy metal / metalloid and antibiotic pollution such as wastewater (Di Cesare et al., 2016), livestock farms and agricultural lands (Hölzel et al., 2012; Ji et al., 2012; Mazhar et al., 2021). When it comes to environments that had little or no prior contact with antibiotics currently in use, results obtained from heavy metal-polluted soils (Safari Sinegani & Younessi, 2017) and mercury-contaminated gold

mining area (Yan et al., 2020) were inconclusive, while similar study of arsenic-rich hot springs (Jardine et al., 2019) yielded negative results. In our work, cultivation experiments indicated a lack of relationship - linear or otherwise - between resistance to arsenic and mercury and resistance to antibiotics as well.

A more in-depth look revealed that in the strains with sequenced genomes, none of the putative antibiotic genes were of the type known to be present in pathogenic bacteria. Given that many strains were multi-resistant, it makes them a potential sources of novel antibiotic resistance genes. In this case, antibiotic resistance genes residing in integrative and conjugative elements (ICEs) tend to be of higher concern as they are much more likely to be horizontally transmitted into other strains, and eventually - into clinically relevant bacteria - both pathogenic and those causing opportunistic infections. Being located in the same ICE with arsenic and/or mercury resistance genes could easily result in conjoined horizontal transfer and very fast propagation of antibiotic resistance under environmental pressures (Andrea et al., 2013). In sequenced strains, most of identified antibiotic resistance genes did not reside close to the arsenic or mercury resistance gene clusters or individual genes, and were located outside of the identified ICEs, making their conjoined horizontal transfer within identified ICEs unlikely. However, it does not completely rule out presence of undetected ICEs or transfer of such genes via plasmids. This conclusion mirrors that from a work by Pal et al. (2015) which studied occurence of metal(loid) and antibiotic resistance genes using 2666 chromosomes and 4582 plasmids from publicly available bacterial genomes and revealed limited potential of metals to promote horizontal gene transfer of antibiotic resistance genes, while giving plenty of opportunities to select for antibiotic-resistant bacteria due to widespread presence of metal resistance genes in the bacterial chromosomes.

From a strictly clinical viewpoint, however, it does not matter if most bacteria inhabiting a given area are not capable of co-transfer of antibiotic and metal(loid) resistance, as it requires just one horizontal transfer event from a strain where such co-selection occurred to insert a novel antibiotic resistance gene into pathogenic bacteria. With those considerations in mind, microbiota of highly contaminated environments of El Terronal should be treated as potential source of novel antibiotic resistance genes.

Strains with biotechnological potential

Several strains held some promise with regards to their biotechnological potential.

For bioremediation of environments with co-occurrence of metal(loid) and hydrocarbon contamination, an organism combining high metal(loid) resistance, capacity to degrade hydrocarbons, and absence of pronounced antibiotic resistance to minimise the risk of horizontal gene transfers would be desirable. Strains *Arthrobacter sp.* C19 and *Paraburkholderia sp.* F24 had high arsenite and arsenate resistance, were capable of rapidly degrading diesel, and to some extent - phenol and naphthalene, and were not resistant to any of the tested antibiotics. Strain *Pseudarthrobacter sp.* SR72 degraded diesel and hexadecane, and had no resistance to any of the tested antibiotics either.

Additionally, strains *Streptomyces sp.* B11 and *Massilia sp.* WB1 were very strong

hydrocarbon emulsifiers, producing emulsions that were stable over the course of several weeks, with potential to use tensoactive compounds produced by them in bioremediation or industrial processes.

7. Conclusions

Conclusions

El Terronal mining and metallurgy site is a heterogeneous brownfield, with conditions ranging from soils with arsenic and mercury concentrations several times above permitted levels, to the extreme environments of waste heaps consisting mostly of arsenic oxides, with significant admixture of mercury. Microbial communities inhabiting those environments varied accordingly.

In relatively mild conditions offered by polluted soils, prokaryotic communities were the most diverse and consisted of Proteobacteria, Actinobacteria, Acidobacteria, Chloroflexi, Planctomycetes, Gemmatimonadetes, Bacteroidetes and Nitrospirae, *Archaea* from the Soil Crenarchaeotic Group as well as many minor bacterial phyla, while fungal and SAR communities were represented by Eomycota and Neomycota fungi and Cercozoa, respectively. With increasing contamination levels, evenness of the microbial communities was reduced, but the overall community structure remained the same.

Prokaryotic communities of the mercury-contaminated flue dust from a derelict chimney, surprisingly, were fairly similar to soils. Housing a less diverse community dominated by Proteobacteria, Actinobacteria and Acidobacteria, it also contained sizeable populations of Gemmatimonadetes, Planctomycetes and Bacteroidetes, as well as Cyanobacteria and Verrucomicrobia. Fungal and SAR communities were similar to those in soils.

Two samples of arsenic-rich soot taken at different times, and one sample of stupp (a type of solid mercury condenser residue) represented the most extreme environments encountered on the site. Prokaryotic communities of all three samples had considerable differences, with stupp sample dominated by *Archaea* of the Soil Crenarchaeotic Group, while two arsenic-rich soot samples - one taken in the winter, and another in the summer - were dominated by Proteobacteria and Firmicutes, and Proteobacteria, Bacteroidetes and Firmicutes, respectively. Alpha-diversity indices of those samples were extremely low. Arsenic-rich soot samples shared most of their low- and medium-abundance inhabitants, while only four organisms were shared between soot and stupp samples. In contrast, fungal communities of those samples were very similar, and dominated by Neomycota fungi.

Two groundwater sediments from different parts of the site were taken; their

communities differed both from the rest of the samples and between themselves. Sediment sample taken from a sounding well located on the heavily polluted river bank (SR) had low biodiversity, was dominated by Proteobacteria, and had sizeable minorities of Acidobacteria, Chloroflexi, Bacteroidetes, Actinobacteria, Elusimicrobia and Ignavibacteriae; its SAR population consisted of Cercozoa. Additionally, Choanozoa protists were present, while fungi were mostly absent. Second sediment sample, taken from a sounding well located further in-land (SB) was dominated by Proteobacteria and Thaumarchaeota, with smaller populations of Chloroflexi, Bacteroidetes, Acidobacteria and Nitrospirae. Functional metagenome predictions indicated that, despite those differences, overall metabolism in those samples was very similar. Fungi were represented by Neomycota and Eomycota; and SAR - by Cercozoa and Halvaria. A large population of Choanozoa protists was also present.

Fifty-nine strains were isolated from the soil, flue dust, soot and groundwater sediment samples. No bacteria were isolated from the stupp sample. Isolated strains belonged to the phyla Proteobacteria (21 strains), Actinobacteria (19 strains), Firmicutes (15 strains) and Bacteroidetes (4 strains). They represented less than 1% of the total diversity (average 0.334; $\sigma = 0.366$), and were not taxonomically representative of the overall microbial community. Proteobacteria strains belonged mostly to the genera Paraburkholderia and Pseudomonas; Firmicutes belonged predominantly to the genus *Bacillus*, while Actinobacteria belonged mainly to the genera *Arthrobacter*, *Pseudoarthrobacter* and *Streptomyces*.

Majority of isolated strains were capable of growth on TSA medium supplemented with either As(III), As(V) or mercury(II) chloride. Level of resistance was almost universally high, with 47 strains out of 59 (79.6%) tolerating As(V) concentrations in excess of 100 mM, 18 strains (30.5%) growing on media supplemented with either 20 or 25 mM of As(III), and 21 strains (36%) tolerating $HgCl_2$ concentrations in excess of 300 µM. Instances of very high resistance to all three contaminants were also common.

On a genetic level, presence of arsenic resistance genes was universal, even in low-tolerance strains. As(III) pumps of the *acr3* family were the most common detoxification mechanism; subsequent genome sequencing demonstrated that it was usually coupled with As(V) reductase *arsC*. As(III) oxidase genes *aio* were found in members of all phyla, but were prevalent in Proteobacteria. Dissimilatory arsenate reductase (*arrA*) genes were found only in Proteobacteria. Two *Pseudomonas spp.,* two *Paraburkholderia spp.* and one *Massilia sp.* had both *aio* and *arrA* genes, suggesting that they were capable of both arsenite and arsenate respiration, depending on the prevalent conditions in the environment. Genome sequencing also revealed that organoarsenical oxidase genes *arsH* were very common, being present in most strains with sequenced genomes.

For mercury resistance, amplification of mercury(II) reductase gene *merA* by PCR was not very successful, with only 7 strains giving positive results. Genome sequencing fared somewhat better, allowing to identify mercury(II) reductase in 10 out of 17 strains. Genome sequencing also revealed that *mer* operons tended to include genes coding for non-identified proteins, or proteins with heavy-metal-associated domains. Genomes of four strains did not include any mercury resistance genes despite having very high levels of resistance, suggesting that mer operon was either not sequenced, or resided in a plasmid.

Many isolated strains were resistant to at least one antibiotic. Most common resistances were to trimetoprim and cefuroxime. Resistance to multiple antibiotics was also common, with 37 strains having resistance to more than one antibiotic. Thirteen out of seventeen sequenced strains had functional antibiotic resistance genes. None of the detected genes had perfect sequence similarity to the known resistance genes currently circulating in pathogenic bacteria. Most of them had similarity with resistance-nodulation-cell division antibiotic efflux pump *adeF*, and were predicted to be effective against fluoroquinolone and tetracycline.

Search for antibiotic resistance genes in integrative and conjugative elements revealed 102 potential antibiotic resistance genes. None of the identified integrative and conjugative elements had arsenic or mercury resistance genes. However, this does not account for plasmids, transposons and prophages, nor for unidentified integrative and conjugative elements, suggesting that all antibiotic-resistant strains isolated from El Terronal should be considered as a potential sources of novel antibiotic resistance genes.

Strains *Arthrobacter sp.* C19, *Paraburkholderia sp.* F24 and *Pseudarthrobacter sp.* SR72 were selected as potentially useful for bioremediation of As/Hg and hydrocarbon co-contaminated soils due to their capacity to degrade hydrocarbons, high degree of arsenic and mercury resistance, and low resistance to antibiotics.

Strains *Streptomyces sp.* B11 and *Massilia sp.* WB1 were capable of creating strong and stable hydrocarbon emulsions, making them potentially interesting for bioremediation or industrial applications.

Conclusiones

El sitio minero y metalúrgico de El Terronal presenta una contaminación heterogénea, con condiciones que van desde suelos con concentraciones de arsénico y mercurio varias veces por encima de los niveles permitidos, hasta ambientes extremos causados por los cúmulos de desechos en los que se encuentran, principalmente, óxidos de arsénico y una cantidad considerable de mercurio. Las comunidades microbianas que habitan esos entornos variaron en consecuencia a este entorno.

En las zonas en las que la contaminación del suelo era relativamente baja, las comunidades procariotas fueron las más diversas y consistieron en Proteobacteria, Actinobacteria, Acidobacteria, Chloroflexi, Planctomycetes, Gemmatimonadetes, Bacteroidetes y Nitrospirae, *Archaea* del *Soil Crenarchaeotic Group*, así como muchos filos bacterianos menores. Las comunidades fúngicas y SAR estuvieron representadas por los hongos Eomycota y Neomycota y por el grupo Cercozoa, respectivamente. Sin embargo, con el aumento de los niveles de contaminación, se redujo la uniformidad de las comunidades microbianas, aunque la estructura comunitaria general siguió siendo la misma.

Sorprendentemente, las comunidades procariotas del polvo de chimenea contaminado con mercurio eran bastante similares a las comunidades presentes en los suelos muestreados, aunque menos diversas. Estas estaban dominadas por Proteobacteria, Actinobacteria y Acidobacteria, conteniendo también poblaciones considerables de Gemmatimonadetes, Planctomycetes y Bacteroidetes, así como Cyanobacteria y Verrucomicrobia. Las comunidades de hongos y SAR fueron similares a las encontradas en los suelos.

Las dos muestras de hollín rico en arsénico tomadas en diferentes momentos y la muestra de residuo sólido del condensador de mercurio representaron los ambientes más extremos encontrados en el sitio. Las comunidades procarióticas de las tres muestras reflejaron diferencias considerables; mientras que la muestra tomada del condensador de mercurio se encontraba dominada por *Archaea* del *Soil Crenarchaeotic Group*, las dos muestras de hollín ricas en arsénico, tomadas en una invierno y otra en verano, estuvieron dominadas por los filos Proteobacteria y Firmicutes, y Proteobacteria, Bacteroidetes y Firmicutes, respectivamente. Los índices de diversidad alfa de esas muestras fueron extremadamente bajos. Las muestras de hollín ricas en arsénico compartieron la mayoría de sus microorganismos de abundancia baja y media, mientras que solamente cuatro microorganismos se encontraron simultáneamente en las muestras de hollín y la muestra de residuo sólido del condensador de mercurio. No obstante, las comunidades de hongos de estas muestras fueron muy similares, ambas dominadas por hongos Neomycota.

En las dos muestras tomadas de los sedimentos de agua subterránea en diferentes partes

del sitio, se encontró que las comunidades microbianas diferían tanto del resto de las muestras como entre sí. La muestra de sedimento tomada en un pozo de sondeo ubicado en la ribera del río altamente contaminado (SR) presentaba baja biodiversidad y estaba dominada por el filo Proteobacteria, con importantes minorías de los filos Acidobacteria, Chloroflexi, Bacteroidetes, Actinobacteria, Elusimicrobia e Ignavibacteriae; su población SAR consistió en el grupo Cercozoa, aunque también se encontraron protistas del filo Choanozoa. Los hongos estaban en su mayoría ausentes. La segunda muestra de sedimento, tomada de un pozo de sondeo ubicado tierra adentro (SB) estuvo dominada por los filos Proteobacteria y Thaumarchaeota, con poblaciones más pequeñas de los filos Chloroflexi, Bacteroidetes, Acidobacteria y Nitrospirae. Las predicciones del metagenoma funcional de las muestras SR y SB indicaron que, a pesar de las diferencias taxonómicas entre ellas, el metabolismo general de cada una fue muy similar entre sí. Los hongos estuvieron representados por los filos Neomycota y Eomycota; y SAR por los grupos Cercozoa y Halvaria, con una gran población de protistas del filo Choanozoa.

En total se aislaron 59 cepas de las muestras de suelo, polvo de combustión, hollín y sedimentos de agua subterránea. No se aislaron bacterias de la muestra de residuo sólido de condensador de mercurio. Las cepas aisladas pertenecían a los filos Proteobacteria (21 cepas), Actinobacteria (19 cepas), Firmicutes (15 cepas) y Bacteroidetes (4 cepas). Estos filos representaban menos del 1% de la diversidad total (promedio 0.334; $\sigma = 0.366$) y, por tanto, no fueron taxonómicamente representativos de la comunidad microbiana general. Las cepas de proteobacterias pertenecían principalmente a los géneros *Paraburkholderia* y *Pseudomonas*; las cepas del filo Firmicutes pertenecían predominantemente al género *Bacillus*, mientras que las del filo Actinobacteria pertenecían principalmente a los géneros *Arthrobacter*, *Pseudoarthrobacter* y *Streptomyces*.

La mayoría de las cepas aisladas fueron capaces de crecer en medio TSA suplementado con As(III), As(V) o cloruro de mercurio(II). El nivel de resistencia fue casi universalmente alto, con 47 cepas de 59 (79,6%) que toleraban concentraciones de As(V) superiores a 100 mM, 18 cepas (30,5%) capaces de crecer en sustratos suplementados con 20 o 25 mM de As(III) y 21 cepas (36%) que toleraban concentraciones de HgCl2 superiores a 300 µM. También fueron comunes los casos de resistencia extremadamente alta a los tres contaminantes.

A nivel genético, la presencia de genes de resistencia al arsénico fue universal, incluso en cepas de baja tolerancia. Las bombas As(III) de la familia *acr3* fueron el mecanismo de desintoxicación más común; la secuenciación posterior del genoma demostró que este mecanismo se encontraba normalmente acoplado a genes *arsC*, que codifican para reductasas de As(V). También se encontraron genes *aio*, que codifican para oxidasas de As(III), en miembros de todos los filos, siendo frecuentes en el filo Proteobacteria. Los genes disimilatorios de la arseniato reductasa (*arrA*) se encontraron solo en el filo Proteobacteria. Dos microorganismos del género *Pseudomonas spp.*, dos *Paraburkholderia spp.* y una *Massilia sp.* presentaban genes *aio* y *arrA*, lo que sugiere

que estos organismos eran capaces de utilizar en la respiración celular tanto arsenito como arsenato, dependiendo de las condiciones predominantes en el medio ambiente. La secuenciación de los genomas también reveló que los genes de oxidasa organoarsénica (*arsH*) eran muy comunes y estaban presentes en la mayoría de las cepas con genomas secuenciados.

En cuanto a la identificación de genes de resistencia al mercurio, la amplificación por PCR del gen *merA*, que codifica para una reductasa de mercurio (II), no obtuvo mucho éxito, con solo 7 cepas que dieron resultados positivos. La secuenciación del genoma fue algo mejor, lo que permitió identificar el gen *merA* en 10 de 17 cepas. Esta secuenciación del genoma también reveló que los operones mer tendían a incluir genes que codifican proteínas no identificadas o proteínas con dominios asociados a metales pesados. Los genomas de cuatro cepas no incluían ningún gen de resistencia al mercurio a pesar de tener niveles muy altos de resistencia, lo que sugiere que el operón mer o bien no se secuenció o residía en un plásmido.

Un gran número de las cepas aisladas fueron resistentes al menos a un antibiótico. Las resistencias más comunes fueron a trimetoprim y cefuroxima. La resistencia a múltiples antibióticos también fue común, con 37 cepas que tenían resistencia a más de un antibiótico. Trece de las diecisiete cepas secuenciadas tenían genes funcionales de resistencia a los antibióticos. Ninguno de los genes detectados tenía una similitud de secuencia perfecta con los genes de resistencia conocidos que circulan actualmente en bacterias patógenas. La mayoría de ellos tenían similitud con genes codificantes de bombas de eflujo de antibiótico de división celular de nodulación de resistencia (*adeF*), los cuales serían efectivos contra fluoroquinolona y tetraciclina, según las predicciones llevadas a cabo.

La búsqueda de genes de resistencia a antibióticos en elementos integradores y conjugativos reveló 102 genes de resistencia a antibióticos potenciales. Ninguno de los elementos integradores y conjugativos identificados tenía genes de resistencia al arsénico o al mercurio. Sin embargo, esto no tiene en cuenta los plásmidos, transposones y profagos, ni elementos integradores y conjugativos no identificados, lo que sugiere que todas las cepas resistentes a antibióticos aisladas de El Terronal deben considerarse como fuentes potenciales de nuevos genes de resistencia a antibióticos.

Las cepas *Arthrobacter sp.* C19, *Paraburkholderia sp.* F24 y *Pseudarthrobacter sp.* SR72 fueron seleccionadas como potencialmente útiles para la biorremediación de suelos co-contaminados con As/Hg e hidrocarburos debido a su capacidad para degradar hidrocarburos, alto grado de resistencia al arsénico y al mercurio, y baja resistencia a los antibióticos.

Las cepas *Streptomyces sp.* B11 y *Massilia sp.* WB1 presentaban la capacidad de

crear emulsiones de hidrocarburos fuertes y estables, lo que las hace potencialmente interesantes para aplicaciones industriales o de biorremediación.

Future studies

Present study could readily serve as a foundation for a series of further studies, each offering a more in-depth look at either the microbial communities inhabiting El Terronal mining site, or at some of the isolated strains. Study of isolated strains capitalises on the benefit of having sequenced genomes for many of them, and vast amounts of data they can provide.

One approach involves a more profound investigation of the arsenic and mercury resistance operons (*ars* and *mer*, respectively) found in the genomes of the sequenced strains. Some of those operons contained genes that were not identified nor fully characterised characterised, but were suspected of playing some role in the metall(oid) resistance either due to encoding Heavy-Metal-Associated (HMA) domains, or simply due to their position inside the operon. DNA sequences of such genes as well as corresponding translated amino acid sequences can be extracted and compared between themselves and to the databases, and 3D structure of proteins they encode can be predicted, giving some insight into their function. Finally, those genes can be cloned, either by themselves or together with their operons, into suitable bacterial strains to study their function, as well as to produce, isolate and characterise the proteins they encode, giving us better understanding of the arsenic and mercury resistance mechanisms employed by bacteria.

Another opportunity is a study of differences between strains with sequenced genomes isolated from El Terronal and their closest relatives from other environments using a comparative genomics approach.

Possibility of the horizontal transfer of antibiotic resistance genes together with those giving resistance to arsenic and mercury can be studied in more detail as well. While searching and annotating mobile genetic elements not covered by the present study (for example, by identification and annotation of prophage sequences by PHASTER tool (Zhou et al., 2011; Arndt et al., 2016), or looking for transposons by locating transposase genes within areas flanked by the inverted repeats) is a preferred strategy that leverages high-quality genome assemblies produced in the present study, isolating, sequencing and annotating plasmid DNA can be a viable approach as well. Additionally, horizontal gene transfer can be inferred experimentally, by using either extracted and purified plasmid DNA or direct cell-to-cell contact, both in the culture media and in simulated natural systems (Ray & Nielsen, 2005).

Removal of the large proportion of waste from the site was achieved as a result of the recent (2018-2019) decontamination effort by the company that inherited the site. While designed to reduce leaching of arsenic and mercury into the environment, the positive effects of it on the surrounding area, and especially on the surface waters of the San Tirso river downstream of the site, are yet to be evaluated. Combined with such (purely geochemical) evaluation, an assessment of the state of the microbial communities inhabiting soils and groundwater sediments of the site can be performed either by the community fingerprinting methods such as Terminal Restriction Fragment Length Polymorphism (W. T. Liu et al., 1997), Denaturating Gradient Gel Electrophoresis

(Muyzer & Smalla, 1998) or Automated Ribosomal Intergenic Spacer Analysis (Fisher & Triplett, 1999; Slabbert et al., 2010), or, if the required funding becomes available, by the 16S rRNA gene amplicon sequencing similarly to the present study.

A. Annex: Materials and Methods

Table 33.: Primers used for amplification of arsenic and mercury resistance genes

Target	Name	Sequence (5' to 3')	Cycling conditions (*Taq* polymerase)	amplicon length	Reference
arrA	arrA-CVF1 arrA-CVR1	CACAGCGCCATCTGCGCCGA CCGACGAACT CCYTGYTCCA	3 min at 93℃, 35 cycles of 94℃ for 30 s, 60℃ for 30 s, 72℃ for 30 s, and 72℃ for 7 min	330 bp	Mirza et al. (2017)
aio-like (1)	aroA-F1 aroA-R1	GTSGGBTGYGGMTAYCABGYCTA TTGTASGCBGGNCGRTTRTGRAT	4 min at 94℃, 35 cycles of 95℃ for 45 s,50℃ - 0.5℃/cycle till 46℃ for 45 s, 72℃ for 50 s, and 72℃ for 5 min	~500 bp	Inskeep et al. (2007)
aio-like (2)	aroA-F2 aroA-R2	GTCGGYTGYGGMTAYCAYGYYTA YTCDGARTTGTAGGCYGGBCG	3 min at 93℃, 40 cycles of 93℃ for 60 s, 50℃ for 1.5 min, 72℃ for 60 s, and 72℃ for 5 min		
acr3 (1)	aacr1F aacr2R	GGCCTGATCGTNATGATGTAYCC GCGATGGCCAGCTCRAARTTRTT	10 min at 94℃, 40 cycles of 95℃ for 45 s, 57℃ - 0.5℃/cycle till 52℃ for 45 s, 72℃ for 50 s, and 72℃ for 5 min		Achour et al. (2007)
acr3 (2)	dacr5F dacr4R	TGATCTGGGTCATGATCTTCCCVATGMTGVT CGGCCACGGCCAGYTCRAARAARTT	10 min at 94℃, 35 cycles of 94℃ for 45 s, 52℃ for 45 s, 72℃ for 50 s, and 72℃ for 5 min	~750 bp	Cai et al. (2009)
arsB	AarsB1F AarsB1R	GAACATCGTCTGGAAYGCNAC GTACACCACCAGRTACATNCC	5 min at 94℃, 35 cycles of 94℃ for 45 s, 55℃ for 45 s, 72℃ for 50 s, and 72℃ for 10 min		Fahy et al. (2015)
merA	A1s-n.F A5-n.R	TCCGCAAGTNGCVACBGTNGG ACCATCGTCAGRTARGGRAAVA	5 min at 94℃, 40 cycles of 94℃ for 30 s, 54℃ for 30 s, 72℃ for 60 s, and 72℃ for 7 min	285 bp[1]	Chadhain et al. (2006)

[1] Exibits frequent non-specific amplification at higher molecular weight

B. Annex: Results

Table 34.: Taxonomic identification of isolated strains according to SILVA128 and SILVA138 databases.

Strain	Species	Taxonomy (SILVA 128): Phylum; Class; Order; Family; Genus	Taxonomy (SILVA138): Phylum; Class; Order; Family; Genus
A11	*Bacillus sp.*	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
A12	*Streptomyces sp.*	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
A13	*Domibacillus sp.*	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*
A14	*Arthrobacter sp.*	Actinobacteria; Actinobacteria; Actinomycetales; Micrococcaceae; *Arthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*
A15	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
A18	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
A21	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
B11	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
B13	*Arthrobacter sp.*	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*
B14	*Pseudomonas sp.*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*
B15	*Micromonospora sp.*	Actinobacteria; Actinobacteria; Micromonosporales; Micromonosporaceae; *Micromonospora*	Actinobacteriota; Actinobacteria; Micromonosporales; Micromonosporaceae; *Micromonospora*
C11	*Variovorax sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*
C12	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
C14	*Pedobacter sp.*	Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Pedobacter*	Bacteroidota; Bacteroidia; Sphingobacteriales; Sphingobacteriaceae; *Pedobacter*
C17	*Arthrobacter sp.*	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*

Table 34.: Taxonomic identification of isolated strains according to SILVA128 and SILVA138 databases.

Strain	Species	Taxonomy (SILVA 128): Phylum; Class; Order; Family; Genus	Taxonomy (SILVA138): Phylum; Class; Order; Family; Genus
C18	*Variovorax sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*
C19	*Arthrobacter sp.*	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*
C110	*Mucilaginibacter sp.*	Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Mucilaginibacter*	Bacteroidota; Bacteroidia; Sphingobacteriales; Sphingobacteriaceae; *Mucilaginibacter*
C111	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
C22	*Variovorax sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*
C23	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
E11	*Novosphingobium sp.*	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Novosphingobium*	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Novosphingobium*
E12	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
E13	*Massilia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*
E14	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
E15	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
E16	*Arthrobacter sp.*	Actinobacteria; Actinobacteria; Actinomycetales; Actinomycetales; *Arthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*
F11	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*

Table 34.: Taxonomic identification of isolated strains according to SILVA128 and SILVA138 databases.

Strain	Species	Taxonomy (SILVA 128): Phylum; Class; Order; Family; Genus	Taxonomy (SILVA138): Phylum; Class; Order; Family; Genus
F12	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
F14	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus
F21	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
F22	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
F24	*Paraburkholderia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Paraburkholderia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Caballeronia-Paraburkholderia*
FDC1	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
FDC2	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
FS31	*Pseudarthrobacter sp.*	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*
FS32	*Nocardioides sp.*	Actinobacteria; Actinobacteria; Propionibacteriales; Nocardioidaceae; *Nocardioides*	Actinobacteriota; Actinobacteria; Propionibacteriales; Nocardioidaceae; *Nocardioides*
FS34	*Mycolicibacterium sp.*	Actinobacteria; Actinobacteria; Corynebacteriales; Mycobacteriaceae; *Mycolicibacterium*	Actinobacteriota; Actinobacteria; Corynebacteriales; Mycobacteriaceae; *Mycobacterium*
SR1	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
SR3	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
SR4	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
SR5	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
SR71	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
SR72	*Pseudarthrobacter sp.*	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*
SR73	*Pseudarthrobacter sp.*	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*

Table 34.: Taxonomic identification of isolated strains according to SILVA128 and SILVA138 databases.

Strain	Species	Taxonomy (SILVA 128): Phylum; Class; Order; Family; Genus	Taxonomy (SILVA138): Phylum; Class; Order; Family; Genus
SR74	*Domibacillus sp.*	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*	Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*
SB1	*Delftia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Delftia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Comamonadaceae; *Delftia*
SB21	*Bacillus sp.*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*
WR2	*Pseudarthrobacter sp.*	Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	Actinobacteriota; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*
WR5	*Pseudomonas sp.*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*
WR6	*Massilia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*
WR7	*Pseudomonas sp.*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*
SB22	*Chryseobacterium sp.*	Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Chryseobacterium*	Bacteroidota; Bacteroidia; Flavobacteriales; Weeksellaceae; *Chryseobacterium*
WB1	*Massilia sp.*	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	Proteobacteria; Gammaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*
WB11	*Pseudomonas sp.*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*
WB15	*Flavobacterium sp.*	Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Flavobacterium*	Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; *Flavobacterium*
WB17	*Streptomyces sp.*	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	Actinobacteriota; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*
WB21	*Pseudomonas sp.*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*
WB22	*Sphingobium sp.*	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Sphingobium*	Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Sphingobium*

Figure 67.: Bacteria and Archaea (16S libraries). Rarefaction curves for Good's counts coverage, Robbins' estimator as well as Shannon and Simpson alpha-diversity indices.

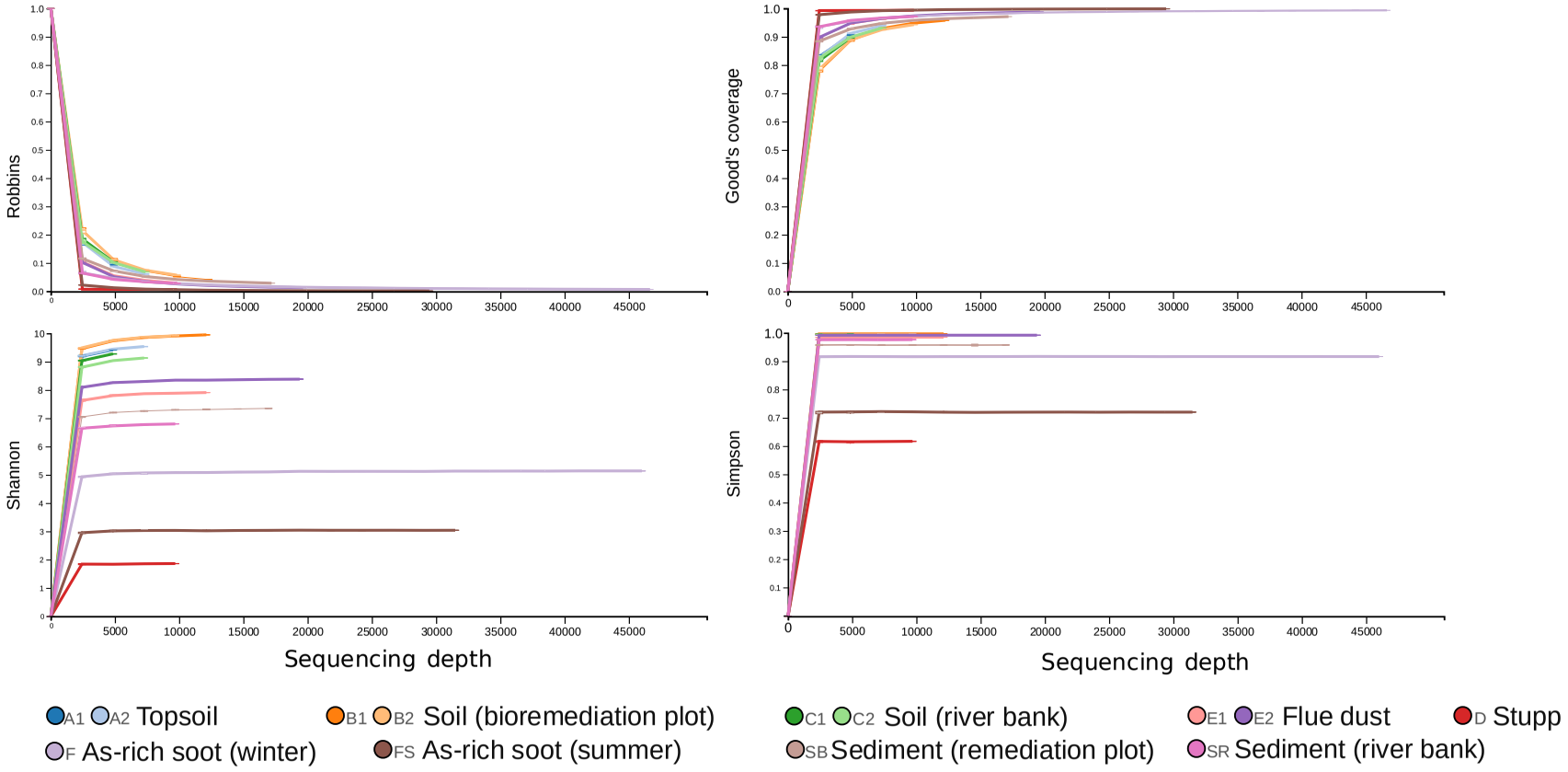


Table 35.: ASV analogues obtained from cultivable strains.

ASV analogue	Taxonomy (SILVA128): Phylum; Class; Order; Family; Genus	Corresponding strains
ASVa01	Bacteria; Actinobacteria; Actinobacteria; Micromonosporales; Micromonosporaceae; *Micromonospora*	B15
ASVa02	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	WB11, WB21
ASVa03	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	WR5, WR7
ASVa04	Bacteria; Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Mucilaginibacter*	C110
ASVa05	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Novosphingobium*	E11
ASVa06	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Sphingobium*	WB22
ASVa07	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Chryseobacterium*	SB22
ASVa08	Bacteria; Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Pseudarthrobacter*	A14, C17, C19, FS31, WR2
ASVa09	Bacteria; Actinobacteria; Actinobacteria; Micrococcales; Micrococcaceae; *Arthrobacter*	B13, E16

Table 35.: ASV analogues obtained from cultivable strains.

ASV analogue	Taxonomy (SILVA128): Phylum; Class; Order; Family; Genus	Corresponding strains
ASVa10	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	C111
ASVa11	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	A12, E14, WB17
ASVa12	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	B11
ASVa13	Bacteria; Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	C12
ASVa14	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Paraburkholderia*	F12
ASVa15	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; *Burkholderia-Paraburkholderia*	E12, F11, F21, F22, F24
ASVa16	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; *Comamonadaceae*	C11
ASVa17	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; *Comamonadaceae*	C18, C22
ASVa18	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	WR6
ASVa19	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	E13
ASVa20	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	SR1, SR4

Table 35.: ASV analogues obtained from cultivable strains.

ASV analogue	Taxonomy (SILVA128): Phylum; Class; Order; Family; Genus	Corresponding strains
ASVa21	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	A11, F14, SR5, SB21
ASVa22	Bacteria; Firmicutes; Bacilli; Bacillales; Planococcaceae; *Domibacillus*; uncultured *Domibacillus* bacterium	A13, SR74
ASVa23	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; *Bacillus*	FDC1, FDC2
ASVa24	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; *Bacillus licheniformis*	A21
ASVa25	Bacteria; Firmicutes; Bacilli; Bacillales; Bacillaceae; Bacillus; *Bacillus simplex*	A18, SR3
ASVa26	Bacteria; Firmicutes; Bacilli; Bacillales; *Planococcaceae*	SR71
ASVa27	Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; *Pseudomonas*	B14
ASVa28	Bacteroidetes; Flavobacteriia; Flavobacteriales; Flavobacteriaceae; *Flavobacterium*	WB15
ASVa29	Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae; *Pedobacter*	C14
ASVa30	Actinobacteria; Actinobacteria; Streptomycetales; Streptomycetaceae; *Streptomyces*	A15
ASVa31	Actinobacteria; Actinobacteria; Corynebacteriales; Mycobacteriaceae; *Mycolicibacterium*	FS34

Table 35.: ASV analogues obtained from cultivable strains.

ASV analogue	Taxonomy (SILVA128): Phylum; Class; Order; Family; Genus	Corresponding strains
ASVa32	Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; *Massilia*	WB1
ASVa33	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Delftia*	SB1
ASVa34	Actinobacteria; Actinobacteria; Propionibacteriales; Nocardioidaceae; *Nocardioides*	FS32
ASVa35	Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; *Variovorax*	C23

Table 36.: Mercury resistance of cultivable strains

Strain	$HgCl_2$, μM	Reduction
	$HgCl_2$ resistance	
A (Topsoil)		
Bacillus sp A11	0	
Streptomyces sp. A12	50	
Domibacillus sp. A13	50	
Arthrobacter sp. A14	50	
Streptomyces sp. A15	10	
Bacillus sp. A18	100	
Bacillus sp. A21	100	
B (Soil, mixed and tilled)		
Streptomyces sp. B11	0	
Arthrobacter sp. B13	200	
Pseudomonas sp. B14	300	
Micromonospora sp. B15	25	
C (Soil of the river bank)		
Variovorax sp. C11	100	
Streptomyces sp. C12	50	
Pedobacter sp. C14	300	
Arthrobacter sp. C17	400	
Variovorax sp. C18	100	
Arthrobacter sp. C19	50	
Mucilaginibacter sp. C110	0	
Streptomyces sp. C111	25	
Variovorax sp. C22	100	+
Bacillus sp. C23	300	
E (Flue dust)		
Novosphingobium sp. E11	400	
Paraburkholderia sp. E12	400	
Massilia sp. E13	25	+
Streptomyces sp. E14	50	
Paraburkholderia sp. E15	50	+

Table 36.: Mercury resistance of cultivable strains

Strain	$HgCl_2$, μM	Reduction
Arthrobacter sp. E16	50	
F (Arsenic-rich soot, winter)		
Paraburkholderia sp. F11	400	
Paraburkholderia sp. F12	300	+
Bacillus sp. F14	400	
Paraburkholderia sp. F21	400	
Paraburkholderia sp. F22	300	
Paraburkholderia sp. F24	400	
FS (Arsenic-rich soot, summer)		
Bacillus sp. FDC1	400	
Bacillus sp. FDC2	50	
Pseudarthrobacter sp. FS31	300	
Nocardioides sp. FS32	400	
Mycolicibacterium sp. FS34	400	
SR (Groundwater sediment, river bank)		
Bacillus sp. SR1	400	
Bacillus sp. SR3	300	
Bacillus sp. SR4	100	
Bacillus sp. SR5	100	
Bacillus sp. SR71	100	
Pseudarthrobacter sp. SR72	100	
Pseudarthrobacter sp. SR73	100	
Domibacillus sp. SR74	200	
SB (Groundwater sediment, remediation plot)		
Delftia sp. SB1	50	
Bacillus sp. SB21	200	
Chryseobacterium sp. SB22	200	
WR (Groundwater, river bank)		
Pseudarthrobacter sp. WR2	400	+
Pseudomonas sp. WR5	200	
Massilia sp. WR6	400	+
Pseudomonas sp. WR7	400	

Table 36.: Mercury resistance of cultivable strains

Strain	$HgCl_2$, µM	Reduction
WB (Groundwater, remediation plot)		
Massilia sp. WB1	200	
Pseudomonas sp. WB11	25	+
Flavobacterium sp. WB15	50	
Streptomyces sp. WB17	50	
Pseudomonas sp. WB21	25	
Sphingobium sp. WB22	25	

Table 37.: Antibiotic resistance of cultivable strains. R is resistant, I is intermediate, and S is susceptible. MAR is multiple antibiotic resistance index.

Strain	Cefuroxime	Erythromycin	Ciprofloxacin	Amoxycyllin+CA	Vancomycin	Tetracycline	Sulphonamides	Trimetoprim	Streptomycin	MAR_{strain}
A (Topsoil)										
Bacillus sp. A11	R	S	S	R	S	S	S	R	S	0.33
Streptomyces sp. A12	S	I	S	S	S	S	S	R	S	0.11
Domibacillus sp. A13	I	R	S	R	R	S	S	R	S	0.44
Arthrobacter sp. A14	S	S	S	S	S	S	S	S	S	0
Streptomyces sp. A15	R	S	S	S	S	S	S	R	S	0.22
Bacillus sp. A18	S	S	S	S	S	S	S	S	S	0
Bacillus sp. A21	R	S	S	S	S	S	S	S	S	0.11
B (Soil, mixed and tilled)										
Streptomyces sp. B11	S	S	S	S	S	R	S	R	S	0.22
Arthrobacter sp. B13	S	S	S	S	S	S	S	I	S	0
Pseudomonas sp. B14	NR	R	S	R	NR	S	S	R	S	0.43
Micromonospora sp. B15	S	S	S	S	S	S	S	S	S	0
C (Soil of the river bank)										
Variovorax sp. C11	S	I	I	S	NR	S	S	R	S	0.13
Streptomyces sp. C12	R	S	I	R	S	S	R	R	S	0.44
Pedobacter sp. C14	S	S	S	S	NR	S	S	S	S	0
Arthrobacter sp. C17	S	S	I	S	S	S	S	S	S	0
Variovorax sp. C18	S	S	S	S	NR	S	S	R	R	0.25
Arthrobacter sp. C19	S	S	I	S	S	S	S	S	S	0
Mucilaginibacter sp. C110	R	S	R	R	NR	R	R	R	I	0.75
Streptomyces sp. C111	R	S	S	R	S	I	S	R	S	0.33
Variovorax sp. C22	R	R	R	R	NR	S	S	R	R	0.75
Bacillus sp. C23	R	S	S	S	S	S	S	S	S	0.11
E (Flue dust)										
Novosphingobium sp. E11	S	S	S	S	NR	S	S	S	S	0

Table 37.: Antibiotic resistance of cultivable strains. R is resistant, I is intermediate, and S is susceptible. MAR is multiple antibiotic resistance index.

Strain	Cefuroxime	Erythromycin	Ciprofloxacin	Amoxycyllin+CA	Vancomycin	Tetracycline	Sulphonamides	Trimetoprim	Streptomycin	MAR_{strain}
Paraburkholderia sp. E12	S	S	S	S	NR	S	S	S	S	0
Massilia sp. E13	S	S	S	S	NR	S	S	S	S	0
Streptomyces sp. E14	S	S	S	S	S	S	S	R	S	0.11
Paraburkholderia sp. E15	S	S	S	S	NR	S	R	S	S	0.13
Arthrobacter sp. E16	S	S	S	S	S	S	S	S	S	0
F (Arsenic-rich soot, winter)										
Paraburkholderia sp. F11	S	I	S	S	NR	S	R	S	S	0.13
Paraburkholderia sp. F12	S	I	R	S	NR	S	S	S	S	0.13
Bacillus sp. F14	R	S	S	R	S	S	S	R	S	0.33
Paraburkholderia sp. F21	S	S	S	S	NR	S	R	S	S	0.13
Paraburkholderia sp. F22	S	I	S	S	NR	S	R	S	S	0.13
Paraburkholderia sp. F24	S	S	S	S	NR	S	S	S	S	0
FS (Arsenic-rich soot, summer)										
Bacillus sp. FDC1	S	S	S	S	S	S	S	S	S	0
Bacillus sp. FDC2	S	I	R	S	S	S	S	S	S	0.11
Pseudarthrobacter sp. FS31	S	S	I	S	S	S	S	I	S	0
Nocardioides sp. FS32	R	S	S	S	S	S	R	R	S	0.33
Mycolicibacterium sp. FS34	S	S	S	S	S	S	S	S	S	0
SR (Groundwater sediment, river bank)										
Bacillus sp. SR1	S	S	S	S	S	S	S	S	S	0
Bacillus sp. SR3	R	I	S	S	S	S	S	S	S	0.11
Bacillus sp. SR4	S	S	S	S	S	S	S	S	S	0
Bacillus sp. SR5	R	S	R	I	S	S	S	R	S	0.33
Bacillus sp. SR71	S	S	S	S	I	S	S	S	S	0
Pseudarthrobacter sp. SR72	S	S	S	S	S	S	S	S	S	0
Pseudarthrobacter sp. SR73	S	S	S	S	S	S	S	S	S	0

Table 37.: Antibiotic resistance of cultivable strains. R is resistant, I is intermediate, and S is susceptible. MAR is multiple antibiotic resistance index.

Strain	Cefuroxime	Erythromycin	Ciprofloxacin	Amoxycyllin+CA	Vancomycin	Tetracycline	Sulphonamides	Trimetoprim	Streptomycin	MAR_{strain}
Domibacillus sp. SR74	S	S	R	S	S	S	S	S	S	0.11
SB (Groundwater sediment, remediation plot)										
Delftia sp. SB1	S	R	S	I	NR	S	S	S	R	0.25
Bacillus sp. SB21	R	S	S	S	S	S	S	R	S	0.22
Chryseobacterium sp. SB22	R	R	I	R	NR	R	R	S	R	0.75
WR (Groundwater, river bank)										
Pseudarthrobacter sp. WR2	I	S	S	S	S	S	R	R	S	0.22
Pseudomonas sp. WR5	NR	S	S	S	NR	S	S	R	S	0.14
Massilia sp. WR6	S	S	S	S	NR	S	S	S	S	0
Pseudomonas sp. WR7	NR	R	S	R	NR	S	S	R	S	0.43
WB (Groundwater, remediation plot)										
Massilia sp. WB1	R	R	S	S	NR	S	S	S	S	0.25
Pseudomonas sp. WB11	NR	R	S	S	NR	S	R	R	S	0.43
Flavobacterium sp. WB15	R	S	S	S	NR	S	S	S	S	0.13
Streptomyces sp. WB17	S	I	R	S	S	I	S	R	S	0.22
Pseudomonas sp. WB21	NR	S	S	S	NR	S	R	R	S	0.29
Sphingobium sp. WB22	S	S	S	S	NR	S	S	S	S	0

Bibliography

Achour, A. R., Bauda, P. & Billard, P. (2007). Diversity of arsenite transporter genes from arsenic-resistant soil bacteria. *Research in Microbiology*, *158*(2), 128–137. https://doi.org/10/cbd2m5

Aguilar, N. C., Faria, M. C. S., Pedron, T., Batista, B. L., Mesquita, J. P., Bomfeti, C. A. & Rodrigues, J. L. (2020). Isolation and characterization of bacteria from a brazilian gold mining area with a capacity of arsenic bioaccumulation. *Chemosphere*, *240*, 124871. http://www.sciencedirect.com/science/article/pii/S0045653519321101

Akiva, E., Brown, S., Almonacid, D. E., Barber, A. E., Custer, A. F., Hicks, M. A., Huang, C. C., Lauck, F., Mashiyama, S. T., Meng, E. C., Mischel, D., Morris, J. H., Ojha, S., Schnoes, A. M., Stryke, D., Yunes, J. M., Ferrin, T. E., Holliday, G. L. & Babbitt, P. C. (2013). The structure–function linkage database. *Nucleic Acids Research*, *42*(D1), D521–D530. https://doi.org/10/gf5cgq

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., … McArthur, A. G. (2019). Card 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*. https://doi.org/10/ggckg6

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10/cnsjsz

Aminov, R. I. (2009). The role of antibiotics and antibiotic resistance in nature. *Environmental Microbiology*, *11*(12), 2970–2988. https://doi.org/https://doi.org/10.1111/j.1462-2920.2009.01972.x

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. https://doi.org/10/cdwcns

Andrea, W., Ball, M., Botello, W. & Yarzábal, L. (2013). Horizontal transfer of heavy metal and antibiotic-resistance markers between indigenous bacteria, colonizing mercury contaminated tailing ponds in southern venezuela, and human pathogens. *Revista de la Sociedad Venezolana de Microbiología*.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., Romá-Mateo, C., Theodosiou, A. & Mitchell, A. L. (2012). The prints database:

A fine-grained protein sequence annotation and analysis resource - its status in 2012 [bas019]. *Database*, *2012*. https://doi.org/10/ghkdc4

Baker-Austin, C., Wright, M. S., Stepanauskas, R. & McArthur, J. V. (2006). Co-selection of antibiotic and metal resistance. *Trends in Microbiology*, *14*(4), 176–182. https://doi.org/10/fvkg6d

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. & Pevzner, P. A. (2012). Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.

Banos, S., Lentendu, G., Kopf, A., Wubet, T., Glöckner, F. O. & Reich, M. (2018). A comprehensive fungi-specific 18s rrna gene sequence primer toolkit suited for diverse research issues and sequencing platforms. *BMC Microbiology*, *18*(1). https://doi.org/10/ghkdc8

Bartelme, R. P., Custer, J. M., Dupont, C. L., Espinoza, J. L., Torralba, M., Khalili, B. & Carini, P. (2020). Influence of substrate concentration on the culturability of heterotrophic soil microbes isolated by high-throughput dilution-to-extinction cultivation (S. G. Tringe, Ed.). *mSphere*, *5*(1). https://doi.org/10.1128/mSphere.00024-20

Basturea, G. N., Harris, T. K. & Deutscher, M. P. (2012). Growth of a bacterium that apparently uses arsenic instead of phosphorus is a consequence of massive ribosome breakdown. *Journal of Biological Chemistry*, *287*(34), 28816–28819. https://doi.org/10/f39hx2

Bermanec, V., Paradžik, T., Kazazić, S. P., Venter, C., Hrenović, J., Vujaklija, D., Duran, R., Boev, I. & Boev, B. (2021). Novel arsenic hyper-resistant bacteria from an extreme environment, crven dol mine, allchar, north macedonia. *Journal of Hazardous Materials*, *402*, 123437. https://doi.org/10/ghkdcj

Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsene-Ploetze, F., Gallien, S., Lauga, B., Casiot, C., Calteau, A., Vallenet, D., Bonnefoy, V., Bruneel, O., Chane-Woon-Ming, B., Cleiss-Arnold, J., Duran, R., Elbaz-Poulichet, F., Fonknechten, N., Giloteaux, L., Halter, D., … Le Paslier, D. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *The ISME Journal*, *5*(11), 1735–1747. https://doi.org/10.1038/ismej.2011.51

Blasco, M. D., Esteve, C. & Alcaide, E. (2008). Multiresistant waterborne pathogens isolated from water reservoirs and cooling systems. *Journal of Applied Microbiology*, *105*(2), 469–475. https://doi.org/10.1111/j.1365-2672.2008.03765.x

BOE. (2015, September 12). Real decreto 817/2015, de 11 de septiembre, por el que se establecen los criterios de seguimiento y evaluación del estado de las aguas superficiales y las normas de calidad ambiental.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. (2010). Scaffolding pre-assembled contigs using sspace. *Bioinformatics*, *27*(4), 578–579. https://doi.org/10/bmrh33

Boetzer, M. & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, *13*(6), R56. https://doi.org/10/gfkh6d

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10/gf5292

BOPA. (2014, April 21). Resolución de 20 de marzo de 2014, de la consejería de fomento, ordenación del territorio y medio ambiente,por la que se establecen los niveles genéricos de referencia para metales pesados en suelos del principado de asturias. https://sedemovil.asturias.es/bopa/2014/04/21/2014-06617.pdf

Boyd, E. S. & Barkay, T. (2012). The mercury resistance operon: From an origin in a geothermal environment to an efficient detoxification machine. *Frontiers in microbiology*, *3*(23087676), 349–349. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466566/

Buchfink, B., Xie, C. & Huson, D. H. (2014). Fast and sensitive protein alignment using diamond. *Nature Methods*, *12*(1), 59–60. https://doi.org/10/gftzcs

Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., Goodstein, D. M., Elsik, C. G., Lewis, S. E., Stein, L. & Holmes, I. H. (2016). Jbrowse: A dynamic web platform for genome visualization and analysis. *Genome Biology*, *17*(1). https://doi.org/10/ghkdct

Cabal, C. L. & Claverol, M. G. (2010). *Riquezas geológicas de asturias.*

Cabal, C. L. & Claverol, M. G. (2006). *La minería del mercurio en asturias: Rasgos históricos.*

Cabal, C. L., García, E. M., Iglesias, J. G. & Claverol, M. G. (1991). Mineralizaciones de hg-as-sb en el borde occidental de la cuenca carbonífera central de asturias y su relación con la tectónica: El yacimiento de el terronal-la peña. *Boletín de la Sociedad Española de Mineralogía*, *14*, 161–170.

Cai, L., Liu, G., Rensing, C. & Wang, G. (2009). Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. *BMC Microbiology*, *9*(1), 4. https://doi.org/10/dt9snz

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. (2016). Dada2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10/f84fxp

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K. & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, *47*(18), e103–e103. https://doi.org/10/gg4nkf

Cardinale, F., Zubin, J. & Tran, D. (2018). *Image super-resolution (ISR).* https://github.com/idealo/image-super-resolution

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P. & Karp, P. D. (2019). The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research*, *48*(D1), D445–D453. https://doi.org/10.1093/nar/gkz862

Cavalca, L., Zecchin, S., Zaccheo, P., Abbas, B., Rotiroti, M., Bonomi, T. & Muyzer, G. (2019). Exploring biodiversity and arsenic metabolism of microbiota inhabiting arsenic-rich groundwaters in northern italy. *Frontiers in Microbiology*, *10*. https://doi.org/10/ghkdcv

Chadhain, S. M. N., Schaefer, J. K., Crane, S., Zylstra, G. J. & Barkay, T. (2006). Analysis of mercuric reductase (mera) gene diversity in an anaerobic mercury-contaminated sediment enrichment. *Environmental Microbiology*, *8*(10), 1746–1752. https://doi.org/10/bmvf5m

Chang, Q., Luan, Y. & Sun, F. (2011). Variance adjusted weighted unifrac: A powerful beta diversity measure for comparing communities based on phylogeny. *BMC Bioinformatics*, *12*(1), 118. https://doi.org/10/fp8wdb

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, *11*(4), 265–270. http://www.jstor.org/stable/4615964

Chao, A. & Lee, S.-M. (1992). Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, *87*(417), 210–217. https://doi.org/10/ghkdch

Chase, J. M. & Knight, T. M. (2013). Scale-dependent effect sizes of ecological drivers on biodiversity: Why standardised sampling is not enough. *Ecology Letters*, *16*, 17–26. https://doi.org/10/f4xrj9

Chaudhary, D. K., Khulan, A. & Kim, J. (2019). Development of a novel cultivation technique for uncultured soil bacteria. *Scientific Reports*, *9*(1), 6666. https://doi.org/10.1038/s41598-019-43182-x

Chen, J. & Rosen, B. P. (2020). The arsenic methylation cycle: How microbial communities adapted methylarsenicals for use as weapons in the continuing war for dominance. *Frontiers in Environmental Science*, *8*, 43. https://doi.org/10/ghkdcw

Chen, W.-Q., Shi, Y.-L., Wu, S.-L. & Zhu, Y.-G. (2016). Anthropogenic arsenic cycles: A research framework and features. *Journal of Cleaner Production*, *139*, 328–336. http://www.sciencedirect.com/science/article/pii/S0959652616311945

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, *18*(1), 117–143. https://doi.org/10/d7x85z

Cockerill, F. (2013). *Performance standards for antimicrobial susceptibility testing : Twenty-third informational supplement; m100 - s23*. Clinical and Laboratory Standards Institute. Wayne, PA, CLSI.

Connon, S. A. & Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology*, *68*(8), 3878–3885. https://doi.org/10/d9362c

Cooper, D. G. & Goldenberg, B. G. (1987). Surface-active agents from two bacillus species. *Applied and Environmental Microbiology*, *53*(2), 224–229. https://doi.org/10/ghkddj

Crognale, S., Amalfitano, S., Casentini, B., Fazi, S., Petruccioli, M. & Rossetti, S. (2017). Arsenic-related microorganisms in groundwater: A review on distribution, metabolic activities and potential use in arsenic removal processes. *Reviews in*

Environmental Science and Bio/Technology, *16*(4), 647–665. https://doi.org/10.1007/s11157-017-9448-8

de Lima Procópio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L. & de Araújo, J. M. (2012). Antibiotics produced by streptomyces. *The Brazilian Journal of Infectious Diseases*, *16*(5), 466–471. http://www.sciencedirect.com/science/article/pii/S1413867012001341

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. & Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rrna gene database and workbench compatible with arb. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10/ch85b4

Di Cesare, A., Eckert, E. M., D'Urso, S., Bertoni, R., Gillan, D. C., Wattiez, R. & Corno, G. (2016). Co-occurrence of integrase 1, antibiotic and heavy metal resistance genes in municipal wastewater treatment plants. *Water Research*, *94*, 208–214. http://www.sciencedirect.com/science/article/pii/S0043135416301087

Dollive, S., Peterfreund, G. L., Sherrill-Mix, S., Bittinger, K., Sinha, R., Hoffmann, C., Nabel, C. S., Hill, D. A., Artis, D., Bachman, M. A., Custers-Allen, R., Grunberg, S., Wu, G. D., Lewis, J. D. & Bushman, F. D. (2012). A tool kit for quantifying eukaryotic rrna gene sequences from human microbiome samples. *Genome Biology*. https://doi.org/10/ghkdcs

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C. & Langille, M. G. I. (2020). Picrust2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10/gg5v22

Dunivin, T. K., Yeh, S. Y. & Shade, A. (2019). A global survey of arsenic-related genes in soil microbiomes. *BMC Biology*, *17*(1). https://doi.org/10/ghkdc9

Duschak, L. H. & Schuette, C. N. (1925). *The metallurgy of quicksilver* (research rep.). U.S. Bureau of Mines. Retrieved June 20, 2019, from https://digital.library.unt.edu/ark:/67531/metadc12405/

Edgar, R. C. (2018). Updating the 97% identity threshold for 16s ribosomal RNA OTUs (A. Valencia, Ed.). *Bioinformatics*, *34*(14), 2371–2375. https://doi.org/10/gdtj9r

El Blog de Acebedo. (2017). *Los efectos de la calcinación del cinabrio en el barrio de la peña*. https://elblogdeacebedo.blogspot.com/2017/01/los-efectos-de-la-calcinacion-del.html

El Blog de Acebedo. (2019). *La mina de el tarronal (mieres-asturias)*. https://elblogdeacebedo.blogspot.com/2019/06/la-mina-de-el-tarronal-mieres-asturias.html

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E. & Finn, R. D. (2018). The pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10/gfhx7r

EUCAST. (2015). *Antimicrobial susceptibility testing: Eucast disk diffusion method*. Version 5.0. European Committee on Antimicrobial Susceptibility Testing.

European Chemicals Agency. (2011). Background document to rac and seac opinions on five phenylmercury compounds. *ECHA (European Chemicals Agency)*. https://echa.europa.eu/documents/10162/18c9392a-1f07-4eee-a358-e27ff9e2f172

Fahy, A., Giloteaux, L., Bertin, P., Paslier, D. L., Médigue, C., Weissenbach, J., Duran, R. & Lauga, B. (2015). 16S rRNA and As-related functional diversity: Contrasting fingerprints in arsenic-rich sediments from an acid mine drainage. *Microbial Ecology*, *70*(1), 154–167. https://doi.org/10/ghkdcf

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10/dqq86n

Fisher, M. M. & Triplett, E. W. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology*, *65*(10), 4630–4636. https://doi.org/10.1128/AEM.65.10.4630-4636.1999

Gallego, J. R., Esquinas, N., Rodríguez-Valdés, E., Menéndez-Aguado, J. M. & Sierra, C. (2015). Comprehensive waste characterization and organic pollution co-occurrence in a hg and as mining and metallurgy brownfield. *Journal of Hazardous Materials*, *300*, 561–571. https://doi.org/10/f74jm7

García, E. M. (1981). Tectónica y mineralizaciones pérmicas en la cordillera cantábrica oriental (noroeste de españa). *Cadernos do Laboratorio Xeolóxico de Laxe*, *2*(2), 263–274.

Ge, Z., Girguis, P. R. & Buie, C. R. (2016). Nanoporous microscale microbial incubators. *Lab on a Chip*, *16*, 480–488. https://doi.org/10.1039/C5LC00978B

Gil-Díaz, M., Rodríguez-Valdés, E., Alonso, J., Baragaño, D., Gallego, J. R. & Lobo, M. C. (2019). Nanoremediation and long-term monitoring of brownfield soil highly polluted with as and hg. *Science of The Total Environment*, *675*, 165–175. https://doi.org/10/ghkdb6

González-Fernández, B., Rodríguez-Valdés, E., Boente, C., Menéndez-Casares, E., Fernández-Braña, A. & Gallego, J. R. (2018). Long-term ongoing impact of arsenic contamination on the environmental compartments of a former mining-metallurgy area. *Science of The Total Environment*, *610-611*, 820–830. https://doi.org/10/ghkdb5

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, *40*(3-4), 237–264. https://doi.org/10/fw7dvw

Gu, Y., D. Van Nostrand, J., Wu, L., He, Z., Qin, Y., Zhao, F.-J. & Zhou, J. (2017). Bacterial community and arsenic functional genes diversity in arsenic contaminated soils from different geographic locations. *PLOS ONE*, *12*(5), 1–18. https://doi.org/10.1371/journal.pone.0176696

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013). Quast: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10/f4sdst

Gutiérrez, M. F. F. Innovación tecnológica y desarrollo económico: La metalurgia del mercurio en mieres, asturias, siglos xix – xx. el ejemplo de la sociedad especial minera el porvenir. In: *Vii congreso de la asociación de historia económica*. 2001.

Guzman, J. J. L., Sousa, D. Z. & Angenent, L. T. (2019). Development of a bioelectrochemical system as a tool to enrich h2-producing syntrophic bacteria. *Frontiers in Microbiology*, *10*, 110. https://doi.org/10.3389/fmicb.2019.00110

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K. & Beck, E. (2013). Tigrfams and genome properties in 2013. *Nucleic Acids Research*, *41*.

Halter, D., Cordi, A., Gribaldo, S., Gallien, S., Goulhen-Chollet, F., Heinrich-Salmeron, A., Carapito, C., Pagnout, C., Montaut, D., Seby, F., Van Dorsselaer, A., Schaeffer, C., Bertin, P. N., Bauda, P. & Arsene-Ploetze, F. (2011). Taxonomic and functional prokaryote diversity in mildly arsenic-contaminated sediments. *Research in Microbiology*, *162*(9), 877–887. http://www.sciencedirect.com/science/article/pii/S0923250811001069

Han, F. X., Su, Y., Monts, D. L., Plodinec, M. J., Banin, A. & Triplett, G. E. (2003). Assessment of global industrial-age anthropogenic arsenic contamination. *Naturwissenschaften*, *90*(9), 395–401. https://doi.org/10.1007/s00114-003-0451-2

Hasan, M. K., Shahriar, A. & Jim, K. U. (2019). Water pollution in bangladesh and its impact on public health. *Heliyon*, *5*(8), e02145. http://www.sciencedirect.com/science/article/pii/S2405844019358050

Hölzel, C., Müller, C., Harms, K., Dr. Mikolajewski, S., Schäfer, S., Schwaiger, K. & Bauer, J. (2012). Heavy metals in liquid pig manure in light of bacterial antimicrobial resistance. *Environmental research*, *113*, 21–7. https://doi.org/10/fx47xx

Hotter, G. S., Wilson, T. & Collins, D. M. (2001). Identification of a cadmium-induced gene in mycobacterium bovis and mycobacterium tuberculosis. *FEMS Microbiology Letters*, *200*(2), 151–155. https://doi.org/10.1111/j.1574-6968.2001.tb10707.x

Huang, J.-H. (2014). Impact of microorganisms on arsenic biogeochemistry: A review. *Water, Air, & Soil Pollution*, *225*(2). https://doi.org/10/ghkdb4

Imran, M., Das, K. R. & Naik, M. M. (2019). Co-selection of multi-antibiotic resistance in bacterial pathogens in metal and microplastic contaminated environments: An emerging health threat. *Chemosphere*, *215*, 846–857. https://doi.org/https://doi.org/10.1016/j.chemosphere.2018.10.114

INE. (2019). *Población del padrón continuo por unidad poblacional*. https://www.ine.es/index.htm

Inskeep, W. P., Macur, R. E., Hamamura, N., Warelow, T. P., Ward, S. A. & Santini, J. M. (2007). Detection, diversity and expression of aerobic bacterial arsenite oxidase genes. *Environmental Microbiology*, *9*(4), 934–943. https://doi.org/10/c46dp5

Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., Winker, K., Kado, D. M., Orwoll, E., Manary, M., Mirarab, S. & Knight, R. (2018). Phylogenetic placement of exact amplicon sequences improves associations with clinical information (N. Chia, Ed.). *mSystems*, *3*(3). https://doi.org/10/ghkdcc

Jardine, J., Mavumengwana, V. & Ubomba-Jaswa, E. (2019). Antibiotic resistance and heavy metal tolerance in cultured bacteria from hot springs as indicators of environmental intrinsic resistance and tolerance levels. *Environmental Pollution*, *249*, 696–702. http://www.sciencedirect.com/science/article/pii/S0269749118334584

Jewett, S. C. & Duffy, L. K. (2007). Mercury in fishes of alaska, with emphasis on subsistence species. *Science of The Total Environment*, *387*(1), 3–27. http://www.sciencedirect.com/science/article/pii/S0048969707007929

Ji, X., Shen, Q., Liu, F., Ma, J., Xu, G., Wang, Y. & Wu, M. (2012). Antibiotic resistance gene abundances associated with antibiotics and heavy metals in animal manures and agricultural soils adjacent to feedlots in shanghai; china. *Journal of Hazardous Materials*, *235-236*, 178–185. http://www.sciencedirect.com/science/article/pii/S0304389412007716

Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1). https://doi.org/10/gb8wj4

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. & Hunter, S. (2014). Interproscan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10/f53532

Kaeberlein, T. (2002). Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science*, *296*(5570), 1127–1129. https://doi.org/10/dcbfxr

Kim, M., Oh, H.-S., Park, S.-C. & Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16s rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt_2), 346–351. https://doi.org/10/ggbwtz

Koechler, S., Arsene-Ploetze, F., Brochier-Armanet, C., Goulhen-Chollet, F., Heinrich-Salmeron, A., Jost, B., Lievremont, D., Philipps, M., Plewniak, F., Bertin, P. N. & Lett, M.-C. (2015). Constitutive arsenite oxidase expression detected in arsenic-hypertolerant pseudomonas xanthomarina s11. *Research in Microbiology*, *166*(3), 205–214. http://www.sciencedirect.com/science/article/pii/S092325081500039X

Krabbenhoft, D. & Schuster, P. (2002). *Glacial ice cores reveal a record of natural and anthropogenic atmospheric mercury deposition for the last 270 years: Usgs fact sheet fs-051-02* (research rep.). U.S. Geological Survey. https://toxics.usgs.gov/pubs/FS-051-02/pdf/fs-051-02.pdf

Krumperman, P. H. (1983). Multiple antibiotic resistance indexing of escherichia coli to identify high-risk sources of fecal contamination of foods. *Applied and Environmental Microbiology*, *46*(1), 165–170. https://doi.org/10/ghbshk

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10/dwmn6z

Kulp, T. R. (2014). Arsenic and primordial life. *Nature Geoscience*, *7*(11), 785–786. https://doi.org/10/ghkdcb

Laetsch, D. R. & Blaxter, M. L. (2017). Blobtools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10/gffk7z

Lane, D. J. (1991). 16S/23S rRNA sequencing. In E. Stackebrandt & M. Goodfellow (Eds.), *Nucleic acid techniques in bacterial systematics* (pp. 115–175). John Wiley & Sons.

Laskaris, P., Tolba, S., Calvo-Bado, L. & Wellington, L. (2010). Coevolution of antibiotic production and counter-resistance in soil bacteria. *Environmental Microbiology*, *12*(7), 2048–2048. https://doi.org/10.1111/j.1462-2920.2010.02262.x

Letunic, I. & Bork, P. (2017). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, *46*(D1), D493–D496. https://doi.org/10/gcwht8

Letunic, I., Doerks, T. & Bork, P. (2014). Smart: Recent updates, new developments and status in 2015. *Nucleic Acids Research*, *43*(D1), D257–D260. https://doi.org/10/f64qr5

Lewis, T. E., Sillitoe, I., Dawson, N., Lam, S. D., Clarke, T., Lee, D., Orengo, C. & Lees, J. (2017). Gene3d: Extensive prediction of globular domains in proteins. *Nucleic Acids Research*, *46*(D1), D435–D439. https://doi.org/10/gfzrt7

Li, H. (2012). *Seqtk: A fast and lightweight tool for processing sequences in the fasta or fastq format*. https://github.com/lh3/seqtk

Li, W., Liu, J. & Hudson-Edwards, K. A. (2020). Seasonal variations in arsenic mobility and bacterial diversity: The case study of huangshui creek, shimen realgar mine, hunan province, china. *Science of The Total Environment*, *749*, 142353. http://www.sciencedirect.com/science/article/pii/S0048969720358824

Liu, J.-l., Yao, J., Wang, F., Min, N., Gu, J.-h., Li, Z.-f., Sunahara, G., Duran, R., Solevic-Knudsen, T., Hudson-Edwards, K. A. & Alakangas, L. (2019). Bacterial diversity in typical abandoned multi-contaminated nonferrous metal(loid) tailings during natural attenuation. *Environmental Pollution*, *247*, 98–107. http://www.sciencedirect.com/science/article/pii/S0269749118332081

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., Tai, C., Deng, Z. & Ou, H.-Y. (2018). ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research*, *47*(D1), D660–D665. https://doi.org/10/ggpzfz

Liu, W. T., Marsh, T. L., Cheng, H. & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16s rrna. *Applied and Environmental Microbiology*, *63*(11), 4516–4522. https://aem.asm.org/content/63/11/4516

Lladser, M. E., Gouet, R. & Reeder, J. (2011). Extrapolation of urn models via poissonization: Accurate measurements of the microbial unknown (D. Zhu, Ed.). *PLoS ONE*, *6*(6), e21105. https://doi.org/10/fvg6jd

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. & Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate earth microbiomes (J. D. Neufeld, Ed.). *mSystems*, *3*(5). https://doi.org/10/gd8kgx

Loredo, J., Ordóñez, A., Baldo, C. & García-Iglesias, J. (2003). Arsenic mobilization from waste piles of the el terronal mine,asturias, spain. *Geochemistry: Exploration, Environment, Analysis*, *3*(3), 229–237. https://doi.org/10/cx7kfj

Loredo, J., Ordóñez, A., Gallego, J. R., Baldo, C. & García-Iglesias, J. (1999). Geochemical characterisation of mercury mining spoil heaps in the area of mieres (asturias, northern spain). *Journal of Geochemical Exploration*, *67*(1-3), 377–390. https://doi.org/10/bqsk3f

Lotze, F. (1945). Zur gliederung der varisziden der iberischen meseta. *Geotektonische Forschungen*, *6*, 78–92.

Lucas, R., Groeneveld, J., Harms, H., Johst, K., Frank, K. & Kleinsteuber, S. (2017). A critical evaluation of ecological indices for the comparative analysis of microbial communities based on molecular datasets. *FEMS Microbiology Ecology*, *93*(1). https://doi.org/10.1093/femsec/fiw209

Ma, L., Kim, J., Hatzenpichler, R., Karymov, M. A., Hubert, N., Hanan, I. M., Chang, E. B. & Ismagilov, R. F. (2014). Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in human microbiome project's most wanted taxa. *Proceedings of the National Academy of Sciences*, *111*(27), 9768–9773. https://doi.org/10.1073/pnas.1404753111

Magoč, T. & Salzberg, S. L. (2011). Flash: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://doi.org/10/fbhm2j

Mandric, I. & Zelikovsky, A. (2015). Scaffmatch: Scaffolding algorithm based on maximum weight matching. *Bioinformatics*, *31*(16), 2632–2638. https://doi.org/10/f7nn4t

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., … Bryant, S. H. (2016). Cdd/sparcle: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, *45*(D1), D200–D203. https://doi.org/10/f9v2z6

Masuda, H. (2018). Arsenic cycling in the earth's crust and hydrosphere: Interaction between naturally occurring arsenic and human activities. *Progress in Earth and Planetary Science*, *5*(1), 68. https://doi.org/10.1186/s40645-018-0224-3

Matuschek, E., Brown, D. F. J. & Kahlmeter, G. (2014). Development of the eucast disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. *Clinical Microbiology and Infection*, *20*(4), O255–O266. https://doi.org/10/f5xwb8

Mazhar, S. H., Li, X., Rashid, A., Su, J., Xu, J., Brejnrod, A. D., Su, J.-Q., Wu, Y., Zhu, Y.-G., Zhou, S. G., Feng, R. & Rensing, C. (2021). Co-selection of antibiotic resistance genes, and mobile genetic elements in the presence of heavy metals in poultry farm environments. *Science of The Total Environment*, *755*, 142702. http://www.sciencedirect.com/science/article/pii/S0048969720362318

Mccune, B. & Grace, J. (2002, January). *Analysis of ecological communities. mjm software design, gleneden beach, or.*

McMurdie, P. J. & Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible (A. C. McHardy, Ed.). *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10/f54g5k

Méndez Pazos, A. (2013). *Peligrosidad asociada a la escombrera "el terronal" a través del análisis del relieve y de la solubilidad de metales pesados mediante extracciones secuenciales* (Master's thesis). University of Oviedo. http://hdl.handle.net/10651/22904

Meng, L., Liu, H., Lan, T., Dong, L., Hu, H., Zhao, S., Zhang, Y., Zheng, N. & Wang, J. (2020). Antibiotic resistance patterns of pseudomonas spp. isolated from raw milk revealed by whole genome sequencing. *Frontiers in Microbiology*, *11*, 1005. https://www.frontiersin.org/article/10.3389/fmicb.2020.01005

Mesa, V. (2017, April 19). *Biodiversidad microbiana en drenaje ácido de minas y sueldos contaminados con metales y sus aplicaciones biotecnológicas* (Doctoral dissertation). Universidad de Oviedo. http : / / hdl . handle . net / 10651/43798

Mesa, V., Navazas, A., González-Gil, R., González, A., Weyens, N., Lauga, B., Gallego, J. L. R., Sánchez, J. & Peláez, A. I. (2017). Use of endophytic and rhizosphere bacteria to improve phytoremediation of arsenic-contaminated industrial soils by autochthonous betula celtiberica (H. L. Drake, Ed.). *Applied and Environmental Microbiology*, *83*(8). https://doi.org/10/ghkddk

Metwally, A. A., Dai, Y., Finn, P. W. & Perkins, D. L. (2016). Wevote: Weighted voting taxonomic identification method of microbial sequences. *PLOS ONE*, *11*(9), 1–18. https://doi.org/10/f9rpc3

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. & Thomas, P. D. (2016). Panther version 11: Expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, *45*(D1), D183–D189. https://doi.org/10/f9v8bb

Minkin, I., Patel, A., Kolmogorov, M., Vyahhi, N. & Pham, S. (2013). Sibelia: A scalable and comprehensive synteny block generation tool for closely related microbial genomes. *Lecture Notes in Computer Science*, 215–229. https://doi.org/10/ghkdcg

Mirarab, S., Nguyen, N. & Warnow, T. (2011). Sepp: Saté-enabled phylogenetic placement. *Biocomputing 2012*. https://doi.org/10/bmtzc4

Mirza, B. S., Sorensen, D. L., Dupont, R. R. & McLean, J. E. (2017). New arsenate reductase gene (arra) pcr primers for diversity assessment and quantification in environmental samples (G. Voordouw, Ed.). *Applied and Environmental Microbiology*, *83*(4). https://doi.org/10/f9nx3g

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2018). Interpro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351–D360. https://doi.org/10/gfkd6r

Moore, C. M., Gaballa, A., Hui, M., Ye, R. W. & Helmann, J. D. (2005). Genetic and physiological responses of bacillus subtilis to metal ion stress. *Molecular Microbiology*, *57*(1), 27–40. https://doi.org/10.1111/j.1365-2958.2005.04642.x

MTI Minas Asturias. (2013). *Pozo esperanza*. https://mti-minas-asturias.blogspot.com/2013/04/pozo-esperanza.html

Mukhopadhyay, R., Rosen, B. P., Phung, L. T. & Silver, S. (2002). Microbial arsenic: From geocycles to genes and enzymes. *FEMS Microbiology Reviews*, *26*(3), 311–325. https://doi.org/10/b7fs2p

Muyzer, G., De Waal, E. C. & Uitierlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16s rrna. *Applied and Environmental Microbiology*, *59*(3), 695–700. https://doi.org/10/ghkddh

Muyzer, G. & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (dgge) and temperature gradient gel electrophoresis (tgge) in microbial ecology. *Antonie van Leeuwenhoek*, *73*(1), 127–141. https://doi.org/10.1023/A:1000669317571

Myers, D. K. (1951). History of the mercury flask. *Journal of Chemical Education*, *28*(3), 127. https://doi.org/10/fd99z9

Narendrula-Kotha, R. & Nkongolo, K. K. (2017). Bacterial and fungal community structure and diversity in a mining region under long-term metal exposure revealed by metagenomics sequencing. *Ecological Genetics and Genomics*, *2*, 13–24. http://www.sciencedirect.com/science/article/pii/S2405985416300234

Neff, J. (2002, December). Arsenic in the ocean. *Bioaccumulation in marine organisms* (pp. 57–78). Elsevier. https://doi.org/10/d883sh

Niane, B., Devarajan, N., Poté, J. & Moritz, R. (2019). Quantification and characterization of mercury resistant bacteria in sediments contaminated by artisanal small-scale gold mining activities, kedougou region, senegal. *Journal of Geochemical Exploration*, *205*, 106353. http://www.sciencedirect.com/science/article/pii/S0375674218305806

Nichols, D., Lewis, K., Orjala, J., Mo, S., Ortenberg, R., O'Connor, P., Zhao, C., Vouros, P., Kaeberlein, T. & Epstein, S. S. (2008). Short peptide induces an uncultivable microorganism to grow in vitro. *Applied and Environmental Microbiology*, *74*(15), 4889–4897. https://doi.org/10/c7nhqp

Nikolskaya, A. N., Arighi, C. N., Huang, H., Barker, W. C. & Wu, C. H. (2006). Pirsf family classification system for protein functional and evolutionary analysis. *Evolutionary Bioinformatics*, *2*, 117693430600200. https://doi.org/10/ghkdcx

Nurfatini, B. & Syahir, A. (2018). Recent advances in mercury detection; towards enabling a sensitive and rapid point-of-check measurement. *Journal of Toxicology and Risk Assessment*, *4*.

Oates, M. E., Stahlhacke, J., Vavoulis, D. V., Smithers, B., Rackham, O. J. L., Sardar, A. J., Zaucha, J., Thurlby, N., Fang, H. & Gough, J. (2014). The superfamily 1.75 database in 2014: A doubling of data. *Nucleic Acids Research*, *43*(D1), D227–D233. https://doi.org/10/f64rnf

Obrist, D., Kirk, J. L., Zhang, L., Sunderland, E. M., Jiskra, M. & Selin, N. E. (2018). A review of global environmental mercury processes in response to human and natural perturbations: Changes of emissions, climate, and land use. *Ambio*, *47*(2), 116–140. https://doi.org/10/gcz3mv

Ohno, M., Shiratori, H., Park, M. J., Saitoh, Y., Kumon, Y., Yamashita, N., Hirata, A., Nishida, H., Ueda, K. & Beppu, T. (2000). Symbiobacterium thermophilum gen. nov., sp. nov., a symbiotic thermophile that depends on co-culture with a bacillus strain for growth. *International Journal of Systematic and Evolutionary Microbiology*, *50*(5), 1829–1832. https://doi.org/10/ghkdcq

Okonechnikov, K., Golosova, O. & Fursov, M. (2012). Unipro ugene: A unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166–1167. https://doi.org/10/ggbwmq

Ordóñez, A., Silva, V., Galán, P., Loredo, J. & Rucandio, I. (2013). Arsenic input into the catchment of the river caudal (northwestern spain) from abandoned hg mining works: Effect on water quality. *Environmental Geochemistry and Health*, *36*(2), 271–284. https://doi.org/10/f5zn9v

Oregaard, G. & Sørensen, S. J. (2007). High diversity of bacterial mercuric reductase genes from surface and sub-surface floodplain soil (oak ridge, usa). *The ISME Journal*, *1*(5), 453–467. https://doi.org/10.1038/ismej.2007.56

Oremland, R. S. & Stolz, J. (2000). Dissimilatory reduction of selenate and arsenate in nature. *Environmental microbe-metal interactions* (pp. 199–224). John Wiley & Sons. https://doi.org/10/ghkdc3

Ounit, R., Wanamaker, S., Close, T. J. & Lonardi, S. (2015). Clark: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics*, *16*(1). https://doi.org/10/gb3h2t

Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. *Advances in microbial ecology* (pp. 1–55). Springer US. https://doi.org/10/ghkddn

Páez-Espino, D., Tamames, J., de Lorenzo, V. & Cánovas, D. (2009). Microbial responses to environmental arsenic. *BioMetals*, *22*(1), 117–130. https://doi.org/10.1007/s10534-008-9195-y

Pal, C., Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. (2015). Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-2153-5

Parada, A. E., Needham, D. M. & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rrna primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10/f8n8hp

Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J. & Hugenholtz, P. (2020). A complete domain-to-species taxonomy for bacteria and archaea. *Nature Biotechnology*, *38*(9), 1079–1086. https://doi.org/10/ggtbk2

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004. https://doi.org/10/gffzgd

Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, *30*(21), 3123–3124. https://doi.org/10/f6nvf4

Pedrós-Alió, C. & Manrubia, S. (2016). The vast unknown microbial biosphere. *Proceedings of the National Academy of Sciences*, *113*(24), 6585–6587. https://doi.org/10/ghkdb7

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuche, B. A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I. & Bridge, A. (2014). Hamap in 2015: Updates to the protein family classification and annotation system. *Nucleic Acids Research*, *43*(D1), D1064–D1070. https://doi.org/10/f64sgw

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, *13*, 131–144. https://doi.org/10/fqtpwp

Pradhan, S. K., Singh, N. R., Kumar, U., Mishra, S. R., Perumal, R. C., Benny, J. & Thatoi, H. (2020). Illumina miseq based assessment of bacterial community structure and diversity along the heavy metal concentration gradient in sukinda chromite mine area soils, india. *Ecological Genetics and Genomics*, *15*, 100054. http://www.sciencedirect.com/science/article/pii/S2405985420300033

Prévost-Bouré, N. C., Christen, R., Dequiedt, S., Mougel, C., Lelievre, M., Jolivet, C., Shahbazkia, H. R., Guillou, L., Arrouays, D. & Ranjard, L. (2011). Validation and application of a pcr primer set to quantify fungal communities in the soil environment by real-time quantitative pcr (J.-H. Yu, Ed.). *PLoS ONE*, *6*(9), e24166. https://doi.org/10/b7bn8v

Pruesse, E., Peplies, J. & Glöckner, F. O. (2012). Sina: Accurate high-throughput multiple sequence alignment of ribosomal rna genes. *Bioinformatics*, *28*(14), 1823–1829. https://doi.org/10/gb5gxk

Puopolo, R., Gallo, G., Mormone, A., Limauro, D., Contursi, P., Piochi, M., Bartolucci, S. & Fiorentino, G. (2020). Identification of a new heavy-metal-resistant strain of geobacillus stearothermophilus isolated from a hydrothermally active volcanic area in southern italy. *International Journal of Environmental Research and Public Health*, *17*(8), 2678. https://doi.org/10.3390/ijerph17082678

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. (2012). The silva ribosomal rna gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10/gfb6mr

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Ray, J. L. & Nielsen, K. M. (2005). Experimental methods for assaying natural transformation and inferring horizontal gene transfer, 491–520. https://doi.org/10.1016/s0076-6879(05)95026-x

Razumov, A. S. (1932). The direct method of calculation of bacteria in water: Comparison with the koch method. *Mikrobiologiya*, *2*, 131–146.

Reaves, M. L., Sinha, S., Rabinowitz, J. D., Kruglyak, L. & Redfield, R. J. (2012). Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science*, *337*(6093), 470–473. https://doi.org/10/rmx

Robbins, H. E. (1968). Estimating the total probability of the unobserved outcomes of an experiment. *Ann. Math. Statist.*, *39*(1), 256–257. https://doi.org/10/cp82c6

Rossman, T. (2007). Arsenic. In W. Rom & S. Markiwitz (Eds.), *Environmental and occupational medicine* (4th ed., pp. 1006–1017). Wolters Kluwer-Lippincott Williams & Wilkins.

Rothwell, R. P. (Ed.). (1895). Treatment of quicksilver ores in the asturias, Spain: Its statistics, technology and trade in the united states and other countries. In *The mineral industry* (pp. 524–530). The Scientific Publishing Company.

Safari Sinegani, A. A. & Younessi, N. (2017). Antibiotic resistance of bacteria isolated from heavy metal-polluted soils with different land uses. *Journal of Global Antimicrobial Resistance*, *10*, 247–255. https://doi.org/10/ghkdck

Schuster, P., Krabbenhoft, D., Naftz, D., Cecil, L., Dewild, J., Susong, D., Green, J. & Abbott, M. (2002). Atmospheric mercury deposition during the last 270 years: A glacial ice core record of natural and anthropogenic sources. *Environmental science & technology*, *36*, 2303–10. https://doi.org/10/bzx8gk

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10/f6b3k4

Sengupta, S., Chattopadhyay, M. & Grossart, H.-P. (2013). The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in Microbiology*, *4*, 47. https://doi.org/10.3389/fmicb.2013.00047

Sforna, M. C., Philippot, P., Somogyi, A., van Zuilen, M. A., Medjoubi, K., Schoepp-Cothenet, B., Nitschke, W. & Visscher, P. T. (2014). Evidence for arsenic metabolism and cycling by microorganisms 2.7 billion years ago. *Nature Geoscience*, *7*(11), 811–815. https://doi.org/10/f6p56f

Shahid, M., Niazi, N. K., Dumat, C., Naidu, R., Khalid, S., Rahman, M. M. & Bibi, I. (2018). A meta-analysis of the distribution, sources and health risks of arsenic-contaminated groundwater in pakistan. *Environmental Pollution*, *242*, 307–319. http://www.sciencedirect.com/science/article/pii/S0269749118317925

Shaji, E., Santosh, M., Sarath, K. V., Prakash, P., Deepchand, V. & Divya, B. V. (2020). Arsenic contamination of groundwater: A global synopsis with focus on the indian peninsula. *Geoscience Frontiers*. http://www.sciencedirect.com/science/article/pii/S1674987120302115

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10/b39t

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. & Xenarios, I. (2012). New and continuing developments at prosite. *Nucleic Acids Research*, *41*(D1), D344–D347. https://doi.org/10/ggbwr7

Simmler, M., Christl, I. & Kretzschmar, R. (2019). Effect of extreme metal(loid) concentrations on prokaryotic community structure in floodplain soils contaminated with mine waste. *Applied Soil Ecology*, *144*, 182–195. http://www.sciencedirect.com/science/article/pii/S0929139318315117

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688–688. https://doi.org/10/c6c7h8

Slabbert, E., Heerden, C. & Jacobs, K. (2010). Optimisation of automated ribosomal intergenic spacer analysis for the estimation of microbial diversity in fynbos soil. *South African Journal of Science*, *106*. https://doi.org/10.4102/sajs.v106i7/8.329

Stackebrandt, E. & Ebers, J. (2006). Taxonomic parameters revisited: Tarnished gold standards. *Microbiology Today*, *8*, 6–9.

Staley, J. T. & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*(1), 321–346. https://doi.org/10/d6v9pq

Stamatakis, A. (2014). Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10/f536fd

Stewart, E. J. (2012). Growing unculturable bacteria. *Journal of Bacteriology*, *194*(16), 4151–4160. https://doi.org/10/f3468z

Tarvainen, T., Albanese, S., Birke, M., Poňavič, M., Reiman, C., Andersson, M., Arnoldussen, A., Baritz, R., Batista, M., Bel-lan, A., Cicchella, D., Demetriades, A., De Vivo, B., Dinelli, E., De Vos, W., Ďuriš, M., Dusza-Dobek, A., Eggen, O., Eklund, M. & Zomeni, Z. (2013). Arsenic in agricultural and grazing land soils of europe. *Applied Geochemistry*, *28*, 2–10. https://doi.org/10/f4kfx3

Tessier, A., Campbell, P. G. C. & Bisson, M. (1979). Sequential extraction procedure for the speciation of particulate trace metals. *Analytical Chemistry*, *51*(7), 844–851. https://doi.org/10/dxfk6n

Vainio, E. J. & Hantula, J. (2000). Direct analysis of wood-inhabiting fungi using denaturing gradient gel electrophoresis of amplified ribosomal dna. *Mycological Research*, *104*(8), 927–936. https://doi.org/10/d4d3n9

Valverde, A., González-Tirante, M., Medina-Sierra, M., Santa-Regina, I., García-Sánchez, A. & Igual, J. M. (2011). Diversity and community structure of culturable arsenic-resistant bacteria across a soil arsenic gradient at an abandoned tungsten-tin mining area. *Chemosphere*, *85*(1), 129–134. http : / / www . sciencedirect.com/science/article/pii/S0045653511006680

Vartoukian, S. R., Palmer, R. M. & Wade, W. G. (2010). Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, *309*(1), 1–7. https://doi.org/10/dkrnfp

Vázquez-Baeza, Y., Gonzalez, A., Smarr, L., McDonald, D., Morton, J. T., Navas-Molina, J. A. & Knight, R. (2017). Bringing the dynamic microbiome to life with animations. *Cell Host & Microbe*, *21*(1), 7–10. https://doi.org/10/ghkdc5

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. (2013). Emperor: A tool for visualizing high-throughput microbial community data [2047-217X-2-16]. *GigaScience*, *2*(1). https://doi.org/10/gftjdb

Ventola, C. L. (2015). The antibiotic resistance crisis: Part 1: Causes and threats. *P & T : a peer-reviewed journal for formulary management*, *40*(25859123), 277–283. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4378521/

Višnjevec, A. M., Kocman, D. & Horvat, M. (2014). Human mercury exposure and effects in europe. *Environmental Toxicology and Chemistry*, *33*(6), 1259–1270. https://doi.org/10/ghkdcz

Volant, S., Lechat, P., Woringer, P., Motreff, L., Campagne, P., Malabat, C., Kennedy, S. & Ghozlane, A. (2020). Shaman: A user-friendly website for metataxonomic analysis from raw reads to statistical analysis. *BMC Bioinformatics*, *21*(1). https://doi.org/10/ghkddb

Wang, P., Sun, G., Jia, Y., Meharg, A. A. & Zhu, Y. (2014). A review on completing arsenic biogeochemical cycle: Microbial volatilization of arsines in environment. *Journal of Environmental Sciences*, *26*(2), 371–381. https://doi.org/10/f5tnrz

Wang, Y. & Qian, P.-Y. (2009). Conservative fragments in bacterial 16s rrna genes and primer design for 16s ribosomal dna amplicons in metagenomic studies (D. Field, Ed.). *PLoS ONE*, *4*(10), e7401. https://doi.org/10/dwr678

Warren, R. L., Coombe, L., Mohamadi, H., Zhang, J., Jaquish, B., Isabel, N., Jones, S. J. M., Bousquet, J., Bohlmann, J. & Birol, I. (2019). ntEdit: Scalable genome sequence polishing (B. Berger, Ed.). *Bioinformatics*, *35*(21), 4430–4432. https://doi.org/10/ggfncc

Watterson, W. J., Tanyeri, M., Watson, A. R., Cham, C. M., Shan, Y., Chang, E. B., Eren, A. M. & Tay, S. (2020). Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes (V. S. Cooper, G. Storz, V. S. Cooper, B. Momeni & H. Jung Kim, Eds.). *eLife*, *9*, e56998. https://doi.org/10.7554/eLife.56998

Werner, J. J., Koren, O., Hugenholtz, P., DeSantis, T. Z., Walters, W. A., Caporaso, J. G., Angenent, L. T., Knight, R. & Ley, R. E. (2011). Impact of training sets on classification of high-throughput bacterial 16s rrna gene surveys. *The ISME Journal*, *6*(1), 94–103. https://doi.org/10/dvfbp4

White, J. R., Nagarajan, N. & Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLOS Computational Biology*, *5*(4), 1–11. https://doi.org/10/cwfkbs

Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*(20), 3350–3352. https://doi.org/10/gb5g64

Willis, A. D. (2019). Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*, *10*. https://doi.org/10/ggbn3x

Wood, D. E. & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46. https://doi.org/10/gfkndk

Wu, F., Wang, J.-T., Yang, J., Li, J. & Zheng, Y.-M. (2016). Does arsenic play an important role in the soil microbial community around a typical arsenic mining area? *Environmental Pollution*, *213*, 949–956. http://www.sciencedirect.com/science/article/pii/S026974911630238X

Yan, C., Wang, F., Liu, H., Liu, H., Pu, S., Lin, F., Geng, H., Ma, S., Zhang, Y., Tian, Z., Chen, H., Zhou, B. & Yuan, R. (2020). Deciphering the toxic effects of metals in gold mining area: Microbial community tolerance mechanism and change of antibiotic resistance genes. *Environmental Research*, *189*, 109869. http://www.sciencedirect.com/science/article/pii/S0013935120307647

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W. & Glöckner, F. O. (2013). The silva and all-species living tree project (ltp) taxonomic frameworks. *Nucleic Acids Research*, *42*(D1), D643–D648. https://doi.org/10/f5sdx4

Zevenhoven, R., Mukherjee, A. B. & Bhattacharya, P. (2007). Arsenic flows in the environment of the european union: A synoptic review. *Arsenic in soil and groundwater environment* (pp. 527–547). Elsevier. https://doi.org/10/bxrz5b

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, *39*(suppl 2), W347–W352. https://doi.org/10.1093/nar/gkr485

Zhu, Y.-G., Xue, X.-M., Kappler, A., Rosen, B. P. & Meharg, A. A. (2017). Linking genes to microbial biogeochemical cycling: Lessons from arsenic. *Environmental Science & Technology*, *51*(13), 7326–7339. https://doi.org/10/ghkdc2

Zhu, Y.-G., Yoshinaga, M., Zhao, F.-J. & Rosen, B. P. (2014). Earth abides arsenic biotransformations. *Annual Review of Earth and Planetary Sciences*, *42*(1), 443–467. https://doi.org/10/ghkddm

Nomenclature

ACE	Abundance-based Coverage Estimator
ADONIS	Permutational multivariate analysis of variance using distance matrices
ANOSIM	Analysis of similarities
ANOVA	Analysis of variance
ARG	Antibiotic resistance genes
As	Arsenic
ASV	Amplicon Sequence Variant
ATP	Adenosine triphosphate
CFU	Colony Forming Units
COG	Cluster of Orthologous Groups
DNA	Deoxyribonucleic acid
EC	Enzyme Comission
gDNA	genomic DNA
Hg	Mercury
HMA	Heavy Metal-Associated domain
ICE	Integrative Conjugative Element
ICP-MS	Inductively Coupled Plasma - Mass Spectrometry
IDA	Isotopic Dilution Analysis
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthology
MAR	Multiple Antibiotic Resistance index

OTU	Operational Taxonomic Unit
PAH	Polycyclic Aromatic Hydrocarbon
PBS	Phosphate Buffered Saline
PCoA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
ppb	parts per billion
ppm	parts per million
PTFE	Polytetrafluoroethylene
rRNA	ribosomal RNA
SAR	Sarcodia - Alveolata - Rhizaria
TSA	Triptic Soy Agar
TSB	Triptic Soy Broth