



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Química Biológica

**Metabolismo del arsénico en microorganismos
extremófilos de los Andes centrales**

Tesis presentada para optar por el título de Doctor de la
Universidad de Buenos Aires en el área Química Biológica

Lic. Federico Agustín Vignale

Director de tesis:	Dra. María Eugenia Farías
Codirector de tesis:	Dr. Adrián Gustavo Turjanski
Consejero de Estudios:	Dra. Nancy Irene López
Lugar de trabajo:	IQUIBICEN-UBA y PROIMI (CONICET)

Buenos Aires, 14 de Diciembre de 2022

Metabolismo del arsénico en microorganismos extremófilos de los Andes centrales

Los salares y volcanes de los Andes centrales de Sudamérica albergan diversos ecosistemas microbianos, tales como biopelículas, tapetes microbianos y microbialitos. Los registros fósiles indican que los microbialitos surgieron hace más de 3.700 millones de años y fueron la principal evidencia de vida en la Tierra durante los siguientes 2.000 millones de años. Debido a las condiciones ambientales extremas de los Andes centrales (alta radiación, baja tasa de precipitación, alta tasa de evaporación, alta salinidad, alta concentración de arsénico, baja presión de oxígeno, fuertes vientos y una gran amplitud térmica), el estudio de los ecosistemas microbianos andinos podría ayudarnos a comprender los límites de la vida en la Tierra y potencialmente en sus orígenes.

El presente trabajo tiene como objetivo comprender los mecanismos moleculares que permiten a los microorganismos extremófilos de los Andes centrales resistir a las elevadas concentraciones ambientales de arsénico y, eventualmente, obtener energía metabólica a partir de su oxidorreducción. En la primera parte del trabajo, se llevó a cabo una prospección de ecosistemas microbianos en tres salares y un volcán de los Andes centrales. En el salar de Antofalla se recolectaron muestras de tapetes microbianos de la laguna Pozo Bravo y de una pequeña laguna o *puquio de campo* naranja; en el volcán Galán se recolectaron biopelículas de la laguna Diamante; en el salar Socompa se tomaron muestras de microbialitos de la laguna Socompa; mientras que en el salar de Atacama se muestrearon tapetes microbianos de las lagunas Tebenquiche y Brava. Junto a los muestreos se realizó un análisis fisicoquímico de cada ambiente. Posteriormente se realizó una secuenciación masiva del ADN total de las muestras para llevar a cabo estudios metagenómicos. Los resultados revelaron que la laguna Diamante es el

ambiente más extremo de los estudiados, presentando la mayor alcalinidad, salinidad, y concentración de arsénico en agua. Por tal motivo, la composición microbiana de la biopelícula de Diamante difiere considerablemente a la del resto de los ecosistemas microbianos estudiados. La biopelícula de Diamante está dominada por haloarqueas, mientras que los ecosistemas de Brava, Tebenquiche, *puquio* de campo naranja, Pozo Bravo, y Socompa están dominados por bacterias de diferentes gremios. Tras analizar la abundancia relativa de los genes del metabolismo del arsénico en dichos ecosistemas, se pudo determinar que la mayoría de los microorganismos andinos presentan mecanismos moleculares de resistencia al arsénico. Los principales mecanismos consisten en (1) la reducción citoplasmática del arsenato, mediada por la arsenato reductasa citoplasmática ArsC, seguido de la expulsión del arsenito resultante por la bomba Acr3, y (2) la metilación citoplasmática del arsenito, mediada por la arsenito metiltransferasa ArsM, seguido de la expulsión del metilarsenito resultante por la bomba ArsP. Además de resistir al arsénico, algunos microorganismos andinos pueden oxidar el arsenito para obtener energía utilizando la enzima arsenito oxidasa AioAB, y/o reducir el arsenato para crecer en anaerobiosis utilizando la enzima arsenato reductasa ArrAB. La abundancia de los genes que codifican las respectivas enzimas resultó ser mayor en la biopelícula de Diamante que en el resto de los ecosistemas, lo que resultó interesante debido a que aún no hay evidencias de estos procesos metabólicos en haloarqueas.

Con el fin de profundizar la comprensión del metabolismo del arsénico en las haloarqueas, la segunda parte del trabajo consistió en el aislamiento de una haloarquea de Diamante (*Halorubrum* sp. DM2) y su caracterización por medio de análisis genómicos y ensayos de crecimiento. Los correspondientes análisis revelaron que *Halorubrum* sp. DM2 presenta la mayoría de los genes del metabolismo del arsénico (genes *ars*, *aio* y *arr*), resiste altas concentraciones de arsénico, y es capaz de crecer como quimioorganótrofo

oxidando compuestos orgánicos y respirando arsenato. Al identificarse una mutante *Halorubrum* sp. DM2 $\Delta arrA$ que no es capaz de crecer en anaerobiosis respirando arsenato, se pudo corroborar que *Halorubrum* sp. DM2 utiliza la enzima arsenato reductasa ArrAB para respirar arsenato.

De esta manera, el trabajo presentado en esta tesis da a conocer los principales procesos metabólicos del arsénico presentes en los microorganismos extremófilos de los Andes centrales, y extiende el conocimiento que se tiene de los mismos, con especial foco en las haloarqueas.

Palabras claves: Andes centrales, biopelículas, tapetes microbianos, microbialitos, metagenómica, genómica, arsénico

Arsenic metabolism in extremophilic microorganisms from the central Andes

The Central Andes of South America are rich in salt flats and volcanoes which harbor diverse microbial ecosystems, such as biofilms, microbial mats, and microbialites. Geological records indicate that microbialites first appeared more than 3.7 billion years ago and were the main evidence of life on Earth for the next 2 billion years. Due to the extreme environmental conditions of the central Andes (high radiation, low precipitation rate, high evaporation rate, high salinity, high arsenic concentration, low oxygen pressure, strong winds and wide daily range in temperatures), the study of the Andean microbial ecosystems could help us understand the limits of life on Earth, as well as advance knowledge in the evolution of early life.

The aim of this work is to understand the arsenic resistance and bioenergetic mechanisms in the extremophilic microorganisms from the central Andes. In the first part of the work, a search for microbial ecosystems was carried out in three salt flats and one volcano from the central Andes. In Antofalla salt flat, microbial mat samples were collected from Pozo Bravo lake and a small orange lagoon or *puquio de campo*; in Galán volcano, biofilm samples were collected from Diamante lake; in Socompa salt flat, microbialite samples were collected from Socompa lake; while in Atacama salt flat, microbial mats were collected from Tebenquiche and Brava lakes. Along with the sampling, a physicochemical analysis of each environment was performed. Afterwards, high-throughput sequencing of the total DNA extracted from the samples was carried out to perform metagenomic studies. The results revealed that Diamante lake is the most extreme environment among those studied, presenting the highest alkalinity, salinity, and arsenic concentration in water. For this reason, the microbial composition of the Diamante biofilm differs considerably from that of the rest of the microbial ecosystems studied. The

Diamante biofilm is dominated by haloarchaea, while Brava, Tebenquiche, *puquio* de campo naranja, Pozo Bravo, and Socompa ecosystems are dominated by bacteria of different guilds. The relative abundance of arsenic metabolism genes in the microbial ecosystems revealed that most of the Andean microorganisms present arsenic resistance mechanisms. The main mechanisms consist of (1) cytoplasmic arsenate reduction, mediated by the cytoplasmic arsenate reductase ArsC, followed by the extrusion of the resulting arsenite by the Acr3 pump, and (2) cytoplasmic arsenite methylation, mediated by the arsenite methyltransferase ArsM, followed by the extrusion of the resulting methylarsenite by the ArsP pump. In addition to these detoxification mechanisms, some Andean microorganisms have mechanisms to obtain energy through the oxidation of arsenite by arsenite oxidase AioAB, and the reduction of arsenate by arsenate reductase ArrAB. The abundance of the genes encoding the respective enzymes resulted to be higher in the Diamante biofilm than in the rest of the ecosystems studied. This was interesting because there is no previous evidence of these metabolic processes in haloarchaea.

To deepen the understanding of the arsenic metabolism in haloarchaea, the second part of the work consisted in the isolation of an haloarchaea from Diamante lake (*Halorubrum* sp. DM2) and its characterization through genomic analyses and growth assays. The corresponding analyses revealed that *Halorubrum* sp. DM2 presents most of the arsenic metabolism genes (*ars*, *aio*, and *arr* genes), is resistant to high arsenic concentrations, and is able to grow as a chemoorganotroph oxidizing organic compounds and reducing arsenate. The identification of a mutant *Halorubrum* sp. DM2 $\Delta arrA$, which is not able to grow in anaerobiosis with arsenate, demonstrated that *Halorubrum* sp. DM2 employs the enzyme arsenate reductase ArrAB to reduce arsenate.

Summing up, the work presented in this thesis sheds light on the arsenic metabolisms present in the extremophilic microorganisms from the central Andes, with special focus on haloarchaea.

Keywords: Central Andes, biofilms, microbial mats, microbialites, metagenomics, genomics, arsenic

"El papel de lo infinitamente pequeño en la naturaleza es infinitamente grande"

Louis Pasteur (1822-1895)

ÍNDICE

PARTE 1.....	1
INTRODUCCIÓN.....	1
Cordillera de los Andes.....	1
Andes centrales.....	3
Ecosistemas microbianos de los Andes centrales.....	4
Interés astrobiológico de los ecosistemas microbianos andinos.....	12
Metabolismo del arsénico en organismos procariotas.....	13
Técnicas metagenómicas para el estudio de ecosistemas microbianos.....	20
OBJETIVOS.....	31
Objetivo general.....	31
Objetivos específicos.....	31
MATERIALES Y MÉTODOS.....	32
Descripción del área de estudio.....	32
Toma de muestras y parámetros ambientales.....	38
Extracción de ADN total y secuenciación.....	39
Control de calidad y preprocesamiento de las lecturas metagenómicas.....	40
Ensamblado de las lecturas metagenómicas.....	42
Reconstrucción de genomas a partir de los metagenomas ensamblados...	43
Anotación estructural y funcional automática.....	44
Análisis de genes implicados en el metabolismo del arsénico.....	44

Análisis taxonómico y funcional de las lecturas metagenómicas.....	46
Análisis de diversidad.....	46
RESULTADOS.....	48
Condiciones ambientales de las lagunas altoandinas.....	48
Composición taxonómica de los ecosistemas microbianos andinos.....	50
Diversidad de los ecosistemas microbianos andinos.....	58
Ensamblado y anotación de los metagenomas.....	59
Reconstrucción de genomas.....	65
Potencial metabólico de los ecosistemas microbianos andinos.....	68
Abundancia de los genes del metabolismo del arsénico.....	72
Organización genómica de los genes del metabolismo del arsénico.....	76
CONCLUSIONES.....	84
PARTE 2.....	86
INTRODUCCIÓN.....	86
Oxidación y reducción del arsénico en el dominio <i>Archaea</i>	86
OBJETIVOS.....	87
Objetivos específicos.....	87
MATERIALES Y MÉTODOS.....	88
Aislamiento de haloarqueas del <i>biofilm</i> rojo de la laguna Diamante.....	88
Extracción y secuenciación del ADN genómico de <i>Halorubrum</i> sp. DM2....	90
Ensamblado de las lecturas genómicas de <i>Halorubrum</i> sp. DM2.....	90
Anotación estructural y funcional del genoma de <i>Halorubrum</i> sp. DM2.....	92

Análisis de similitud genómica de <i>Halorubrum</i> sp. DM2.....	92
Análisis filogenético de <i>Halorubrum</i> sp. DM2.....	93
Análisis de genes del metabolismo del arsénico en <i>Halorubrum</i> sp. DM2..	93
Ensayos de crecimiento de <i>Halorubrum</i> sp. DM2.....	94
Obtención de una mutante <i>Halorubrum</i> sp. DM2 Δ arrA.....	97
RESULTADOS.....	99
Ensamblado y anotación del genoma de <i>Halorubrum</i> sp. DM2.....	99
Análisis de similitud genómica de <i>Halorubrum</i> sp. DM2.....	102
Análisis filogenético de <i>Halorubrum</i> sp. DM2.....	103
Análisis de genes del metabolismo del arsénico en <i>Halorubrum</i> sp. DM2...	104
Ensayos de crecimiento de <i>Halorubrum</i> sp. DM2.....	109
CONCLUSIONES.....	116
CONCLUSIONES FINALES Y PERSPECTIVAS FUTURAS.....	117
PUBLICACIONES.....	119
AGRADECIMIENTOS.....	121
BIBLIOGRAFÍA.....	128
APÉNDICE.....	159

PARTE 1

INTRODUCCIÓN

Cordillera de los Andes

La cordillera de los Andes es un cordón montañoso, localizado en la región occidental de Sudamérica, que se extiende desde la costa del Caribe hasta el extremo sur del continente atravesando siete países: Colombia, Venezuela, Ecuador, Bolivia, Perú, Chile y Argentina (Figura 1). Es la cadena de montañas más extensa del mundo y la segunda en altura, por detrás del Himalaya, con una longitud aproximada de 7.000 km y una elevación máxima de 6.960 m. Su ancho varía entre 200 y 700 km, y ocupa una superficie aproximada de 3.300.000 km² (Zeil, 1979).

Su origen se remonta a fines del Cretácico tardío, momento en el cual se terminó de fragmentar el supercontinente de Pangea provocando la separación de África y Sudamérica. Esta separación produjo la apertura del océano Atlántico y, al mismo tiempo, el nacimiento de la cordillera de los Andes debido al cabalgamiento de la placa continental Sudamericana por encima de las placas oceánicas (placa del Caribe, placa de Cocos, placa de Nazca, placa Antártica y placa Escocesa). Este proceso, denominado subducción, se ha mantenido a lo largo del tiempo provocando el crecimiento de la cordillera hasta llegar al relieve que observamos hoy en día (Zeil, 1979). Este proceso, además, es el principal responsable de la actividad sísmica y volcánica que caracteriza a la región. La cordillera alberga más de un centenar de volcanes activos y forma parte del Cinturón de Fuego del Pacífico (Stern, 2004). Como consecuencia de dicha actividad volcánica, es uno de los

principales yacimientos minerales (hierro, cobre, zinc, plata, oro y plomo) del mundo (Wallace & Hall-Wallace, 2003).



Figura 1. Mapa de Sudamérica en donde se detallan la cordillera de los Andes y sus tres grandes regiones (Andes del norte, Andes centrales y Andes del sur). A su vez, se detallan las regiones que constituyen los Andes centrales.

Teniendo en cuenta la relación de las placas oceánicas con el continente, la cordillera de los Andes puede dividirse en tres grandes regiones (Gansser, 1973) (Figura 1):

- 1) Andes del norte o septentrionales: En esta región se encuentran los Andes venezolanos, colombianos y ecuatorianos. Este segmento es el resultado de la compleja interacción entre las placas oceánicas del Caribe, de Cocos y de Nazca, y la placa continental Sudamericana.
- 2) Andes centrales: En esta región se encuentran los Andes peruanos, bolivianos y gran parte de los argentino-chilenos. Este segmento resulta de la interacción de la placa oceánica de Nazca con la placa continental Sudamericana.
- 3) Andes del sur o australes: En esta región se encuentra la otra parte de los Andes argentino-chilenos. Este segmento es el resultado de la interacción entre las placas oceánicas de Nazca y Antártica, y la placa continental Sudamericana.

Andes centrales

Los Andes centrales se dividen en dos ramales: la Cordillera Occidental y la Cordillera Oriental. Los materiales erosionados de estas cordilleras, y los expulsados por los volcanes, rellenaron la depresión central dando lugar a una extensa meseta conocida como el Altiplano en Bolivia, Perú y Chile, y la Puna en Argentina (Figura 1). La meseta del Altiplano-Puna tiene una elevación promedio de 3.700 m (Isacks, 1988) y cubre un área de 300 km de ancho por 1.800 km de largo (Allmendinger et al., 1997). Debido a estas características, es la segunda meseta más alta y extensa del mundo, superada únicamente por la meseta del Tíbet en Asia oriental. El segmento Altiplano-Puna está constituido por numerosas cuencas endorreicas, es decir cuencas cerradas sin drenaje exterior, que varían en forma, tamaño, relleno y

marco geológico. Dichas cuencas acumulan aguas saturadas en sales (boratos, cloruros, nitratos, sulfatos, etc.) que forman lagunas y lagos hipersalinos. Debido al clima árido extremo, dichos cuerpos de agua se evaporan dejando las sales precipitadas. Con el pasar del tiempo, las sales se acumulan y forman salares (R. N. Alonso & Rojas, 2020). Existen numerosos salares en la meseta del Altiplano-Puna, siendo los más extensos el Salar de Uyuni en Bolivia, el Salar de Arizaro en Salta, Argentina, y el Salar de Antofalla en Catamarca, Argentina. Si bien el Salar de Atacama en Chile no forma parte del segmento Altiplano-Puna, es uno de los salares más extensos de los Andes centrales (H. Alonso & Risacher, 1996).

Ecosistemas microbianos de los Andes centrales

Los salares y volcanes de los Andes centrales albergan numerosos ecosistemas microbianos denominados ecosistemas microbianos andinos (AMEs, *Andean microbial ecosystems*). Los AMEs son complejas asociaciones de microorganismos con minerales, principalmente halita (NaCl), yeso ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$), calcita (CaCO_3), y aragonita (CaCO_3). Dichos ecosistemas pueden clasificarse, en base a la organización de la comunidad y al grado de mineralización, en cuatro grupos: *biofilms* o biopelículas (1), tapetes microbianos (2), microbialitos (3) y endoevaporitas (4) (Farías, 2020) (Figura 2).

Los *biofilms* o biopelículas (1) son comunidades microbianas adheridas a una superficie y embebidas en una matriz de sustancias poliméricas extracelulares (EPS, *extracellular polymeric substances*) (López et al., 2010). La matriz es normalmente una mezcla de polisacáridos, proteínas y ácidos nucleicos que unen las células entre sí, protegiéndolas de entornos adversos y facilitando su supervivencia. La formación de *biofilms* permite a las células vivir en estrecha asociación unas con otras. Cuando las células están muy

próximas entre sí, tienen más oportunidades para intercambiar nutrientes y material genético (Jain et al., 2007).

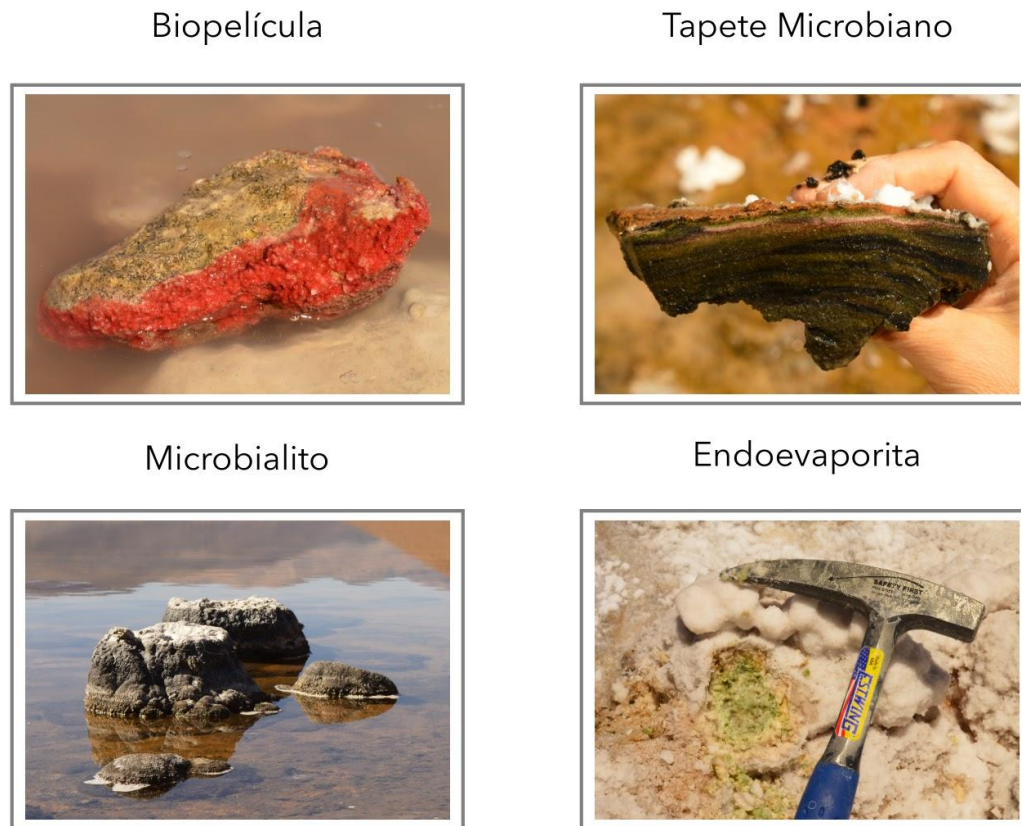


Figura 2. Clasificación de los ecosistemas microbianos andinos (AMEs) en base a la organización de la comunidad y al grado de mineralización.

Los tapetes microbianos (2) son *biofilms* estratificados que pueden tener un grosor de varios centímetros, incluso decímetros. Las capas de los tapetes microbianos están compuestas por especies de diferentes gremios microbianos cuya actividad está gobernada por la disponibilidad de luz, oxígeno y otros recursos (Prieto-Barajas et al., 2018). La combinación de metabolismo microbiano y transporte de nutrientes controlado por difusión ocasiona un acusado gradiente de concentraciones de los diferentes nutrientes y metabolitos microbianos, de manera que en el tapete se crean nichos especiales a diversos intervalos de profundidad. De todas formas, la

estructura química y biológica de un tapete microbiano puede cambiar drásticamente en un período de veinticuatro horas (llamado ciclo circadiano) como consecuencia del cambio en la intensidad de la luz (Madigan et al., 2014). Durante el día existe una intensa producción de oxígeno en la capa superficial fótica y una activa reducción de sulfato en las regiones inferiores. Cerca de la zona en la que el oxígeno y el sulfuro de hidrógeno empiezan a mezclarse, la intensa actividad metabólica de las bacterias del azufre fotótrofas y quimiolitótrofas consume estos sustratos rápidamente en distancias verticales muy cortas. Estos gradientes desaparecen por la noche, cuando todo el tapete se vuelve anóxico y se acumula el sulfuro de hidrógeno. Algunos organismos del tapete aprovechan su motilidad y siguen el desplazamiento de los gradientes químicos.

Los tapetes microbianos han existido hace más de 3.500 Ma (millones de años), sin embargo, la evolución de los metazoos depredadores y la competencia con los macrófitos (plantas acuáticas) provocó su declive hace 1.000 Ma. Hoy en día, estos ecosistemas se desarrollan únicamente en ambientes acuáticos en los que las adversidades ambientales restringen la depredación y la competencia, condiciones que se encuentran sobre todo en hábitats hipersalinos o geotérmicos como los de los Andes centrales (Farías, 2020).

Bajo ciertas condiciones, los *biofilms* y tapetes microbianos litifican formando microbialitos. Los microbialitos (3) son estructuras organosedimentarias formadas por comunidades microbianas bentónicas, que atrapan y fijan detritos sedimentarios y/o constituyen el lugar donde ocurre la precipitación mineral (Burne & Moore, 1987). Los microbialitos se pueden clasificar como estromatolitos (R. E. Riding & Awramik, 2000), trombolitos (Aitken, 1967; Kennard & James, 1986; Shapiro, 2000), dendrolitos (Suosaari et al., 2018) y leirolitos (Braga et al., 1995) de acuerdo a sus mesoestructuras laminadas, moteadas, dendríticas (o *shrubby*) y masivas, respectivamente (Grey &

Awramik, 2020) (Figura 3). En general, los microbialitos son de composición carbonática, aunque los hay de composición silíceea e incluso de evaporitas (R. Riding, 2000). Los diferentes tipos de microbialitos son el resultado de diferentes tipos de *biofilms* y tapetes microbianos (R. Riding, 1991), y de los cambios en las condiciones ambientales (Feldmann & McKenzie, 1998; Moore & Burne, 1994).

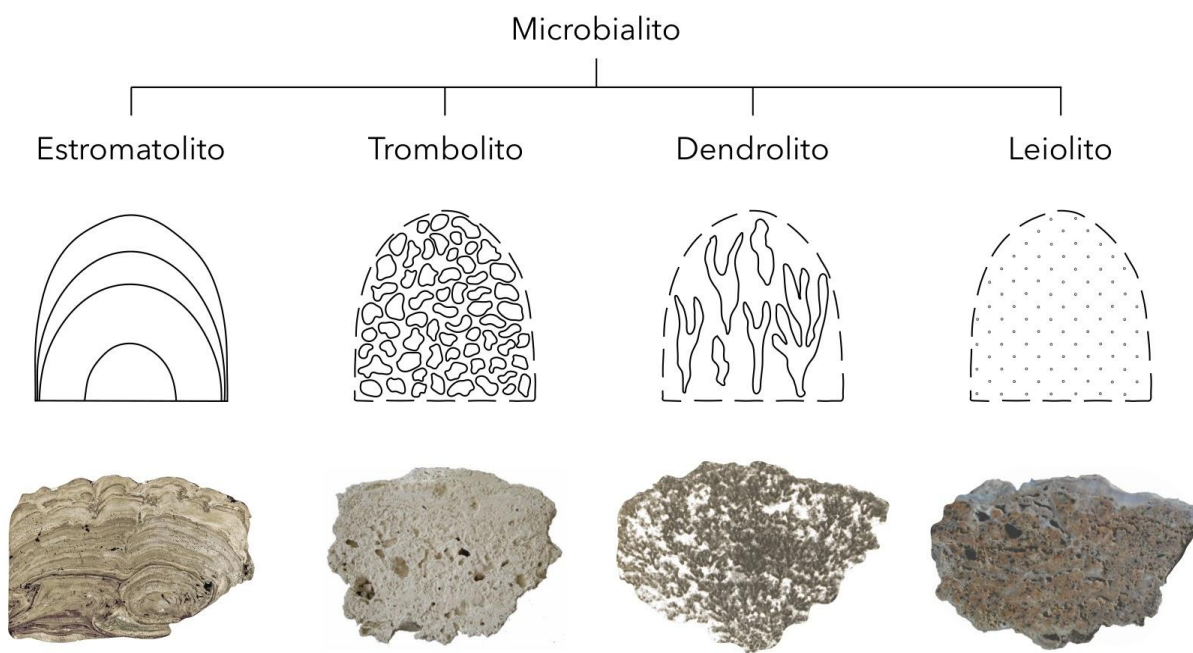


Figura 3. Clasificación de los microbialitos de acuerdo a su mesoestructura.

Los microbialitos representan las primeras formas de vida en la Tierra y se extienden en el registro geológico desde el Precámbrico (hace ~3.500 Ma) hasta el presente. A lo largo de la historia de la Tierra, experimentaron períodos de expansión (entre 2.800 y 1.000 Ma atrás) y declive (2.000, 1.000 y 675 Ma atrás) (R. Riding, 2000). En la actualidad, los microbialitos modernos se encuentran distribuidos en una diversidad de ambientes marinos y continentales (ambientes lacustres hipersalinos, salobres y de agua dulce, lechos de ríos salobres y frescos, y fuentes termales, entre otros). Los microbialitos modernos más famosos y mejor estudiados son los de Hamelin

Pool, Shark Bay, Australia occidental (Babilonia et al., 2018; Burne & Johnson, 2012; Burns et al., 2004; Logan, 1961; Pamela Reid et al., 2003; Suosaari et al., 2018), los de Bahamas (M. S. Andres & Pamela Reid, 2006; Dupraz et al., 2004; Foster et al., 2009; Hubert et al., 2018; Planavsky & Ginsburg, 2009; Reid et al., 1995), los del Lago Clifton, Australia occidental (Burne & Moore, 1993; Moore & Burne, 1994; Warden et al., 2016), y los del Gran Lago Salado, Utah, Estados Unidos (Lindsay et al., 2017, 2020; Pace et al., 2016), entre otros. Su estudio permitió comprender, en cierta medida, los complejos procesos biológicos y sedimentológicos que operan en la construcción de los distintos tipos de microbialitos (Grey & Awramik, 2020).

Las endoevaporitas (4) son comunidades microbianas endolíticas asociadas a yeso y halita, donde los minerales proveen protección contra la desecación y la radiación UV-B, pero permiten el paso de la radiación fotosintéticamente activa (PAR, *photosynthetically active radiation*, 400 a 700 nm), que es importante para el desarrollo de los fotoautótrofos. Estos microorganismos producen pigmentos fotosintéticos y biomoléculas fotoprotectoras (clorofila *a* (Chl*a*), bacterioclorofila *a* (BChl*a*) y sus respectivas feofitinas) que son clave para mantener el equilibrio entre la captación de luz y la protección contra un exceso de radiación. Los minerales que componen las endoevaporitas preservan estas biomoléculas, potenciando aún más el crecimiento de los microorganismos fotoautótrofos en estos ecosistemas (Farías, 2020; Vignale, Lencina, et al., 2021).

Desde hace más de una década, se vienen realizando exploraciones en los salares y volcanes de los Andes centrales para registrar y estudiar los AMEs (Albarracín et al., 2015; Carrizo et al., 2022; dC Rubin et al., 2017; Demergasso et al., 2003, 2010; Dorador et al., 2008, 2013; Farías, 2020; Farías et al., 2009, 2014; Farias et al., 2017; Fernandez et al., 2016; Fleming & Prufert-Bebout, 2010; Haferburg et al., 2017; Kurth et al., 2017, 2021; Perez et al., 2020; Rascovan et al., 2016; Rasuk et al., 2014, 2016; Stivaletta et al.,

2011; Thiel et al., 2010; Toneatti et al., 2017; Valero-Garcés et al., 2001; Vignale, Kurth, et al., 2021; Vignale, Lencina, et al., 2021; Villafañe et al., 2021). Estas exploraciones permitieron identificar *biofilms*, tapetes microbianos, microbialitos y endoevaporitas en los diversos ambientes prospectados, tal como se detalla en la Tabla 1 y en la Figura 4.

Tabla 1. Ecosistemas microbianos andinos reportados en los Andes centrales.

País	Región	Laguna	Salar o volcán asociado	AMEs
Bolivia	Uyuni	S. de Uyuni		B.
	Nor LÍpez	L. Verde	S. de la L. Verde	B., T.M. y M.
		L. Blanca	S. de la L. Blanca	T.M. y M.
Chile	Región II	L. de Huasco	S. de Huasco	T.M.
		L. Coposa	S. de Coposa	B. y T.M.
		<i>Puquios</i> de Llamará	S. de Llamará	T.M. y E.
		L. Verde	S. de Carcote	T.M. y E.
		L. Turquesa	S. de Ascotán	T.M. y E.
		Campo de Géiseres del Tatio	Pastos Grandes y Complejo Guacha Caldera	T.M. y M.
		L. de la Piedra	S. de Atacama	E.
		L. Cejar		T.M. y E.
		L. Tebenquiche		T.M. y E.
		L. Chaxa		T.M. y M.
		L. Barros Negros		T.M. y E.
		Aguas de Quelana		T.M.

		L. Salada		T.M.
		L. Brava		T.M. y M.
		L. Pujsa	S. de Pujsa	B.
		S. Aguas Calientes		T.M.
		L. Lejía	S. de Lejía	T.M.
		L. Punta Negra	S. de Punta Negra	T.M.
		L. Agua Amarga	S. Agua Amarga	E.
Argentina	Jujuy	L. Vilama	V. Vilama	T.M.
	Salta	Ojos de Mar de Tolar Grande	S. de Tolar Grande	E.
		F. de Socompa	S. y V. Socompa	B.
		L. Socompa		M. y E.
	Catamarca	L. Verde	S. de Antofalla	T.M. y E.
		L. Pozo Bravo		T.M. y M.
		<i>Puquio</i> de Campo Naranja		T.M. y M.
		Las Quinoas		T.M. y M.
		L. El Peinado	V. El Peinado	T.M. y M.
		L. Carachipampa	S. y V. Carachipampa	T.M. y M.
		F. de Diamante	Caldera del V. Galán	T.M.
		L. Diamante		B. y M.
		Río Salado	Complejo L. Verde	M.
		L. Negra		T.M. y M.

L., Laguna; V., Volcán; S., Salar; F., Fumarola; B., *Biofilms*; T.M., Tapetes microbianos; M., Microbialitos; E., Endoevaporitas.



Figura 4. Salares y lagunas de los Andes centrales en donde se reportaron ecosistemas microbianos andinos (AMEs). A, Salar de Uyuni; B, Salar de

Coposa; C, *Puquios* de Lllamará; D, Laguna Verde del Salar de Carcote; E, Laguna Cejar; F, Laguna Tebenquiche; G, Laguna Chaxa; H, Laguna Barros Negros; I, Laguna Salada; J, Laguna Brava; K, Salar Aguas Calientes; L, Laguna Agua Amarga; M, Laguna Vilama; N, Ojos de Mar de Tolar Grande; O, Laguna Socompa; P, Laguna Pozo Bravo; Q, *Puquio* de Campo Naranja; R, Las Quinoas; S, Laguna El Peinado; T, Laguna Carachipampa; U, Laguna Diamante.

Interés astrobiológico de los ecosistemas microbianos andinos

Debido a las características climáticas extremas (baja presión de oxígeno por la elevada altitud, intensa radiación solar, alta tasa de evaporación, gran amplitud térmica estacional y diaria, condiciones hipersalinas de los cuerpos de agua, fuertes vientos, y elevada actividad volcánica) los ambientes de los Andes centrales son escenarios ecológicamente interesantes para entender los límites de la vida (Carrizo et al., 2022). Por lo tanto, el estudio de los microorganismos extremófilos que constituyen los AMEs podría ayudarnos a comprender los mecanismos moleculares que permiten su resistencia a la desecación, a la intensa radiación solar y a la elevada concentración de elementos de origen volcánico como el arsénico (As). Algunos de estos mecanismos moleculares podrían haber surgido en los albores de la vida, hace más de 3.500 Ma, debido a la falta de una capa de ozono en la atmósfera terrestre primitiva que protegiera de la radiación UV-B, y a la gran actividad volcánica de dicho período que contribuyó a la especiación de las aguas. En este sentido, el estudio de los AMEs permitiría mejorar nuestra comprensión sobre la evolución de la vida primitiva en la Tierra. Por otro lado, las condiciones ambientales únicas de las lagunas y lagos de altura de los Andes centrales muestran cierta analogía con los paleolagos de Marte del período Noachiano, hace unos 3.500 Ma (Cabrol, 2001; Wharton et al., 1995).

Por lo tanto, el estudio de los AMEs también podría aportar información sobre la vida en otros planetas (Sánchez-García et al., 2022).

Metabolismo del arsénico en organismos procariotas

El arsénico es un componente natural de la corteza terrestre y un constituyente de muchos minerales, sin embargo, resulta tóxico para la mayoría de los organismos (Nordstrom, 2002). Dependiendo de las condiciones fisicoquímicas y de las actividades biológicas que se lleven a cabo, el arsénico puede encontrarse en diferentes estados de oxidación y metilación, mostrando diversos niveles de toxicidad y biodisponibilidad (Rahman & Hassler, 2014). La toxicidad del arsénico varía de la siguiente manera: metilarsenito [MAs(III)] > arsenito [As(III)] > arsenato [As(V)] > metilarsenato [MAs(V)]. Las formas más comunes son el arsenito y el arsenato inorgánico. La toxicidad del arsenato se debe a que funciona como análogo del fosfato e interfiere en todos los procesos en los que interviene este compuesto. Mientras que la toxicidad del arsenito está dada por su capacidad de unirse a cisteínas y a oxidar glutatión, interfiriendo con el funcionamiento de muchas enzimas y aumentando la cantidad de radicales libres en la célula. Por este motivo, la toxicidad del arsenito es mayor que la del arsenato (Murphy & Saltikov, 2009).

A lo largo de la evolución, los organismos procariotas han desarrollado diversos mecanismos moleculares para detoxificarse del arsénico, e incluso, obtener energía metabólica a partir de su oxidorreducción (Páez-Espino et al., 2009; Silver & Phung, 2005; Slyemi & Bonnefoy, 2012; Stolz et al., 2006; Tsai et al., 2009; Yoshinaga et al., 2011) (Figura 5). La mayoría de estos procesos, que involucran diversos genes (Tabla 2), afectan la movilización del arsénico en el ambiente. Por consiguiente, los organismos procariotas juegan un papel clave en el ciclo biogeoquímico del arsénico (Gadd, 2010).

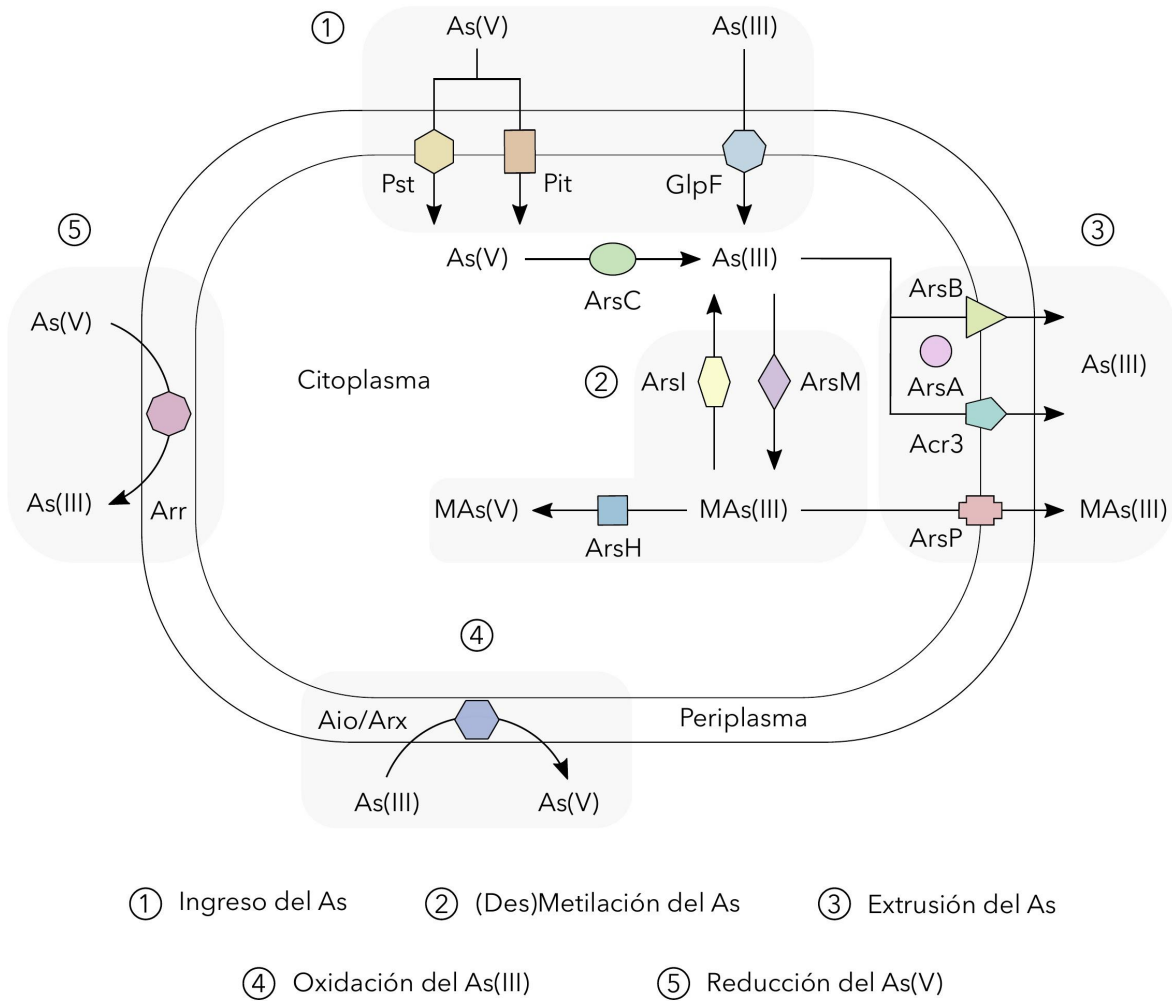


Figura 5. Mecanismos de resistencia y utilización del arsénico en organismos procariotas.

Normalmente, el arsénico entra a las células procariotas en forma de arsenato a través de transportadores de fosfato específicos (Pst, *phosphate-specific transporters*) e inespecíficos (Pit, *non-specific inorganic phosphate transporters*), y en forma de arsenito a través de acuagliceroporinas (GlpF, *glycerol facilitator*). En el citoplasma, el arsénico se une al represor transcripcional ArsR, induciendo un cambio conformacional en su estructura. Este cambio lo libera de su sitio de unión al ADN, permitiendo la transcripción de los operones *ars* (Rawle et al., 2021). Como respuesta, se produce un aumento significativo en los niveles de ArsA, ArsB/Acr3, ArsC y

ArsD, que constituyen la respuesta básica de resistencia al arsénico. La enzima arsenato reductasa citoplasmática (ArsC) reduce el arsenato a arsenito (Gladysheva et al., 1994; Liu & Rosen, 1997), el cual luego es expulsado del citoplasma por las bombas ArsB y Acr3 con gasto de ATP (Bröer et al., 1993; Meng et al., 2004; Yang et al., 2012). ArsA funciona como ATPasa, y se une a ArsB y Acr3 para mejorar su eficiencia de transporte (Rosen et al., 1999). Mientras que ArsD funciona como segundo represor transcripcional de los operones *ars*, previniendo la sobreexpresión de las bombas ArsB y Acr3 (Y. Chen & Rosen, 1997).

Además de este mecanismo básico, ampliamente distribuido en bacterias y arqueas (Mukhopadhyay & Rosen, 2002; Stolz et al., 2006), existen otros procesos que permiten a las células procariotas detoxificarse del arsénico. Por ejemplo, el arsenito puede ser metilado a metilarsenito [MAs(III)] por la enzima arsenito metiltransferasa (ArsM) (Ajees & Rosen, 2015; Qin et al., 2006), para luego ser expulsado por la bomba ArsP o ser oxidado a metilarsenato [MAs(V)] por la enzima metilarsenito oxidasa (ArsH) (J. Chen, Bhattacharjee, et al., 2015; J. Chen, Madegowda, et al., 2015). Otro proceso consiste en la conversión del metilarsenito a su forma inorgánica menos tóxica, el arsenito, a través de una liasa que rompe el enlace C-As (ArsI) (Yoshinaga & Rosen, 2014). Debido a que el arsenito genera estrés oxidativo, algunos organismos procariotas además presentan un sistema tiorredoxina (ArsTX) que ayuda a mantener el estado redox balanceado (Achour-Rokbani et al., 2010).

A lo largo de las últimas tres décadas, se han ido descubriendo organismos procariotas capaces no solo de resistir al arsénico sino además de utilizarlo en procesos bioenergéticos. Algunos de estos organismos, como las bacterias *Chrysiogenes arsenatis* (Macy et al., 1996) y *Shewanella* sp. ANA-3 (Saltikov et al., 2003), o las arqueas hipertermófilas *Pyrobaculum arsenaticum* y *Pyrobaculum aerophilum* (Huber et al., 2000), pueden respirar

anaeróbicamente arsenato utilizando la enzima arsenato reductasa ArrAB. Mientras que otros, como las bacterias *Rhizobium* sp. NT-26 (Santini et al., 2000; Santini & vanden Hoven, 2004), *Azoarcus* sp. DA01 (Rhine et al., 2007), *Dechloromonas* sp. ECC1-pb1 (Sun et al., 2010) y *Alkalilimnicola ehrlichii* MLHE-1 (Zargar et al., 2010), pueden obtener energía de la oxidación del arsenito mediante el uso de arsenito oxidasas (AioAB y ArxAB). Aunque no todas las arsenito oxidasas son utilizadas para obtener energía. Algunas, como la del heterótrofo *Herminiimonas arsenicoxydans* (Muller et al., 2007), sirven como mecanismo de detoxificación.

Actualmente se postula que el arsenito podría haber sido la fuente de energía utilizada por las primeras formas de vida (Sforna et al., 2014; van Lis et al., 2013). Los organismos primitivos habrían surgido en ambientes hidrotermales o al menos en ambientes cercanos a zonas con actividad volcánica, donde las emanaciones volcánicas eran ricas en arsénico. Lo más probable es que estos organismos fueran quimiolitioautótrofos, es decir, que obtuvieran energía de la oxidación de compuestos inorgánicos, y carbono de la fijación del dióxido de carbono. De todas las enzimas utilizadas para obtención de energía por oxidación de compuestos inorgánicos, las arsenito oxidasas (Aio) se encuentran tanto en bacterias como en arqueas y son las únicas que por análisis filogenéticos han mostrado un origen evolutivo anterior al último ancestro común universal (LUCA, *last universal common ancestor*). Se cree entonces que en la Era Arcaica (hace 3.800-2.500 Ma), la oxidación de arsenito por Aio habría estado acoplada a la respiración anaeróbica de los óxidos de nitrógeno. Sólo después de la divergencia de *Bacteria* y *Archaea*, Aio habría sido reclutada a vías metabólicas más recientes como la fotosíntesis. En la Era Proterozoica (hace 2500-600 Ma), después de la aparición del oxígeno atmosférico, la actividad de Aio se habría asociado a la respiración aeróbica. La oxidación de la Tierra habría conducido tanto a la acumulación de arsenato como a la aparición de quinonas de alto potencial.

Por lo tanto, una enzima originalmente empleada en la reducción de compuestos de azufre y relacionada con la actual enzima polisulfuro reductasa (Psr), habría evolucionado hasta convertirse en la enzima Arr reductora de arsenato. Simultáneamente, Arx se habría originado invirtiendo la reacción catalítica de Arr/Psr para oxidar arsenito, dando lugar a la reducción de un conjunto de ubiquinonas. A pesar de que se han encontrado y caracterizado decenas de secuencias de Aio en bacterias, solo unas pocas han sido secuenciadas en arqueas y hasta la fecha no se ha demostrado funcionalidad en ninguna de ellas. Por este motivo, la hipótesis del origen y relevancia ancestral de Aio sigue siendo poco robusta (van Lis et al., 2013).

Tabla 2. Genes procariotas implicados en el metabolismo del arsénico.

Gen	Proteína codificante	Función
<i>arsA</i>	ATPasa	Hidroliza ATP para proveer de energía a ArsB y Acr3
<i>arsB</i>	As(III) permeasa	Extruye del citoplasma el As(III)
<i>acr3</i>	As(III) permeasa	Extruye del citoplasma el As(III)
<i>arsC</i>	As(V) reductasa citoplasmática	Reduce el As(V) a As(III) en el citoplasma
<i>arsD</i>	Represor transcripcional	Regula negativamente la transcripción de los operones <i>ars</i> (regula su expresión máxima)
<i>arsR</i>	Represor transcripcional	Regula negativamente la transcripción de los operones <i>ars</i> y <i>arr</i> (regula su expresión basal)
<i>arsH</i>	MAs(III) oxidasa	Oxida el MAs(III) a MAs(V)
<i>arsM</i>	As(III) metiltransferasa	Metila el As(III)
<i>arsI</i>	C-As liasa	Desmetila el MAs(III)
<i>arsN</i>	Acetiltransferasa	Se desconoce su función exacta
<i>arsP</i>	MAs(III) permeasa	Extruye del citoplasma el MAs(III)
<i>arsT</i>	Tiorredoxina reductasa	Ayuda a mantener el estado redox
<i>arsX</i>	Tiorredoxina	Ayuda a mantener el estado redox

<i>arsO</i>	Monooxigenasa	Se desconoce su función exacta
<i>aioA</i>	As(III) oxidasa, subunidad grande	Oxida en el periplasma el As(III) a As(V) para la respiración de oxígeno, nitrato o clorato
<i>aioB</i>	As(III) oxidasa, subunidad pequeña	Oxida en el periplasma el As(III) a As(V) para la respiración de oxígeno, nitrato o clorato
<i>aioC</i>	Citrocromo c	Participa en la transferencia de electrones
<i>aioD</i>	Sintetasa del cofactor de molibdeno	Sintetiza el cofactor de molibdeno (MoCo) para AioA
<i>aioX</i>	Proteína de unión a As(III) periplasmática	Regula positivamente la transcripción del operón <i>aio</i>
<i>aioS</i>	Sistema regulador de dos componentes (sensor quinasa)	Regula positivamente la transcripción del operón <i>aio</i>
<i>aioR</i>	Sistema regulador de dos componentes (regulador de la respuesta)	Regula positivamente la transcripción del operón <i>aio</i>
<i>aioF</i>	Activador transcripcional	Regula positivamente la transcripción del operón <i>aio</i>
<i>arrA</i>	As(V) reductasa respiratoria, subunidad grande	Reduce en el periplasma el As(V) a As(III) para la respiración de arsénico
<i>arrB</i>	As(V) reductasa respiratoria, subunidad pequeña	Reduce en el periplasma el As(V) a As(III) para la respiración de arsénico
<i>arrC</i>	Subunidad de membrana	Participa en la transferencia de electrones
<i>arrD</i>	Chaperona	Participa en el plegado de ArrAB
<i>arxA</i>	As(III) oxidasa anaeróbica, subunidad grande	Oxida el As(III) a As(V) para la respiración de nitrato o para la transferencia de electrones en la fotosíntesis anoxigénica
<i>arxB</i>	As(III) oxidasa anaeróbica, subunidad pequeña	Oxida el As(III) a As(V) para la respiración de nitrato o para la transferencia de electrones en la fotosíntesis anoxigénica
<i>arxB'</i>	Proteína de Fe-S	Se desconoce su función exacta

<i>arxC</i>	Subunidad de membrana	Participa en la transferencia de electrones
<i>arxD</i>	Chaperona	Participa en el plegado de ArxAB
<i>arxX</i>	Proteína de unión a As(III) periplasmática	Regula positivamente la transcripción del operón <i>arx</i>
<i>arxS</i>	Sistema regulador de dos componentes (sensor quinasa)	Regula positivamente la transcripción del operón <i>arx</i>
<i>arxR</i>	Sistema regulador de dos componentes (regulador de la respuesta)	Regula positivamente la transcripción del operón <i>arx</i>
<i>pit</i>	Transportador de fosfato inorgánico de baja afinidad	Transporta fosfato (o arsenato) inorgánico al interior de la célula
<i>pstS</i>	Transportador de fosfato inorgánico de alta afinidad (subunidad de unión)	Transporta fosfato (o arsenato) inorgánico al interior de la célula
<i>pstC</i>	Transportador de fosfato inorgánico de alta afinidad (subunidad permeasa)	Transporta fosfato (o arsenato) inorgánico al interior de la célula
<i>pstA</i>	Transportador de fosfato inorgánico de alta afinidad (subunidad de la membrana interna)	Transporta fosfato (o arsenato) inorgánico al interior de la célula
<i>pstB</i>	Transportador de fosfato inorgánico de alta afinidad (subunidad de unión a ATP)	Transporta fosfato (o arsenato) inorgánico al interior de la célula

As(III), Arsenito; As(V), Arsenato; MAs(III), Metilarsenito; MAs(V), Metilarsenato.

Los ambientes de los Andes centrales presentan altas concentraciones de arsénico de origen volcánico (hasta ~340 mg/L). Recientes estudios genómicos y metagenómicos (Kurth et al., 2017; Rascovan et al., 2016) han

demostrado que las bacterias y arqueas de los AMEs presentan genes *ars*, *aio* y *arr* implicados en el metabolismo del arsénico. Sin embargo, aún no se tiene información sobre su abundancia, organización genómica, filogenia y expresión, entre otras cosas. Por este motivo, nuevos estudios ómicos de los AMEs nos ayudarían a mejorar nuestra comprensión sobre los mecanismos moleculares que permiten a los organismos procariotas desarrollarse a elevadas concentraciones de arsénico. Algunos de estos mecanismos podrían haber sido ancestrales, por lo que estos estudios ómicos también nos permitirían hipotetizar sobre la evolución de la vida primitiva en la Tierra.

Técnicas metagenómicas para el estudio de ecosistemas microbianos

Tradicionalmente, las comunidades microbianas ambientales han sido estudiadas a partir del aislamiento de diferentes organismos y su caracterización fenotípica y genética. Mediante esta metodología se descubrió y caracterizó la gran mayoría de los microorganismos que conocemos hasta el momento. Sin embargo, este enfoque tiene una importante limitación ya que la mayoría de los microorganismos ambientales (90-99 %) no han podido ser cultivados en el laboratorio (Rappé & Giovannoni, 2003; Whitman et al., 1998). Esto se debe en parte a que los mismos, por lo general, no se encuentran en forma libre e independiente del resto, sino que forman consorcios donde se relacionan íntimamente entre sí y con su microambiente, lo cual impide separar unos de otros. Por otro lado, resulta muy difícil, sin conocimiento previo de sus capacidades y requerimientos metabólicos, diseñar estrategias para el aislamiento y medios de cultivo para su crecimiento en el laboratorio.

Una manera de estudiar a ese gran porcentaje de organismos no cultivables es a través de la metagenómica. Esta técnica permite estudiar el conjunto de genomas de una comunidad a partir de una extracción de ADN total, sin la

necesidad del cultivo (Kunin et al., 2008; Thomas et al., 2012; Wooley et al., 2010). La misma puede ser utilizada para conocer la identidad taxonómica de las especies, así como su capacidad metabólica. Con las tecnologías de secuenciación de primera generación (p.ej., secuenciación de Sanger) esta tarea resultaba difícil. En primer lugar, el ADN total era fragmentado al azar. Luego, estos fragmentos eran clonados en plásmidos de gran tamaño como fósidos, cósmidos y cromosomas artificiales (BACs, *bacterial artificial chromosomes*) que crecían en bibliotecas monoclonales para obtener suficiente material genético para secuenciar. A partir de aquí, el ADN era secuenciado (Slatko et al., 2011). Con la llegada de las tecnologías de secuenciación de nueva generación (NGS, *Next generation sequencing*), la metagenómica se volvió más accesible. En primer lugar, ninguna de estas tecnologías requiere que las secuencias sean clonadas previamente, sino que permiten secuenciar muestras de ADN total en forma directa. En segundo lugar, debido a su mayor rendimiento y menor costo, es posible abarcar una mayor proporción de la diversidad microbiana en una muestra de lo que jamás ha sido posible, permitiendo además trabajar con múltiples muestras ambientales a la vez. Una de las limitaciones de estas tecnologías es el tamaño de los fragmentos, que son más cortos comparados con las tecnologías de secuenciación anterior. De todas formas, desde su introducción, las tecnologías de secuenciación de nueva generación han revolucionado el campo de la microbiología. Nuevos grupos de organismos han sido descubiertos, incluso a niveles tan generales como filo, pese a que nunca han sido cultivados (Cardenas & Tiedje, 2008; Lasken, 2012; Rinke et al., 2013). Y se ha determinado la diversidad metabólica y el funcionamiento ecológico de numerosas comunidades microbianas (Bertin et al., 2011; Woyke et al., 2006; Wrighton et al., 2012). Como consecuencia de estos enormes avances, se han generado iniciativas de gran envergadura para comenzar a estudiar y caracterizar los microbiomas de toda clase de ambientes alrededor del mundo. Ejemplos de estas iniciativas son: *Global*

Ocean Sampling (Kannan et al., 2007; Yooseph et al., 2007), *Human Microbiome Project* (Turnbaugh et al., 2007), *Earth Microbiome Project* (Gilbert et al., 2014), MetaSUB (MetaSUB International Consortium, 2016), y *Tara Oceans* (Sunagawa et al., 2020). El objetivo último de estos grandes proyectos es que los datos recolectados permitan formular y poner a prueba nuevas teorías ecológicas y evolutivas.

Existen dos tipos de enfoques que pueden ser utilizados en metagenómica: secuenciar únicamente un gen o región acotada (*amplicon sequencing*) o secuenciar todos los genomas completos al azar (*shotgun sequencing*). El primer enfoque utiliza un gen marcador, típicamente el que codifica para la subunidad 16S del ácido ribonucleico (ARN) ribosomal. Debido a su tasa de mutación relativamente constante entre linajes, puede ser utilizado como indicador taxonómico para determinar la estructura de una comunidad microbiana. Esto se realiza normalmente a través de la secuenciación de amplicones de una o más de sus nueve regiones hipervariables (V1-V9), siendo V3 y V4 las más usadas (Huse et al., 2008). El segundo enfoque, por otro lado, permite el estudio del rol funcional de los taxones presentes en la comunidad, además de sus características genómicas. De esta manera se pueden establecer relaciones funcionales entre los diferentes miembros microbianos y comprender tanto su presencia como las interacciones metabólicas que ocurren (Kunin et al., 2008; Thomas et al., 2012). Un análisis metagenómico por *shotgun sequencing* normalmente consta de los siguientes pasos (Kunin et al., 2008; Thomas et al., 2012) (Figura 6):

1) Muestreo y extracción de ADN total: Teniendo en cuenta la pregunta que se quiere responder, se podrá determinar el tipo y número de muestras a tomar. La toma de muestras normalmente se realiza con pinzas, espátulas, tubos, frascos, y/o bolsas previamente esterilizadas para evitar contaminaciones que afecten la interpretación de los datos. A su vez, las muestras recolectadas se almacenan en frío (4°C) y en soluciones especiales (p.ej., RNeasy®) para

preservar el material genético hasta el momento de su extracción. Existen diversos métodos para extraer el ADN total dependiendo del tipo de muestra (suelo, agua, fluidos humanos, etc.). Independientemente del método elegido, el ADN total extraído debe ser representativo de todas las células presentes en la muestra y estar presente en buena cantidad y calidad para su secuenciación.

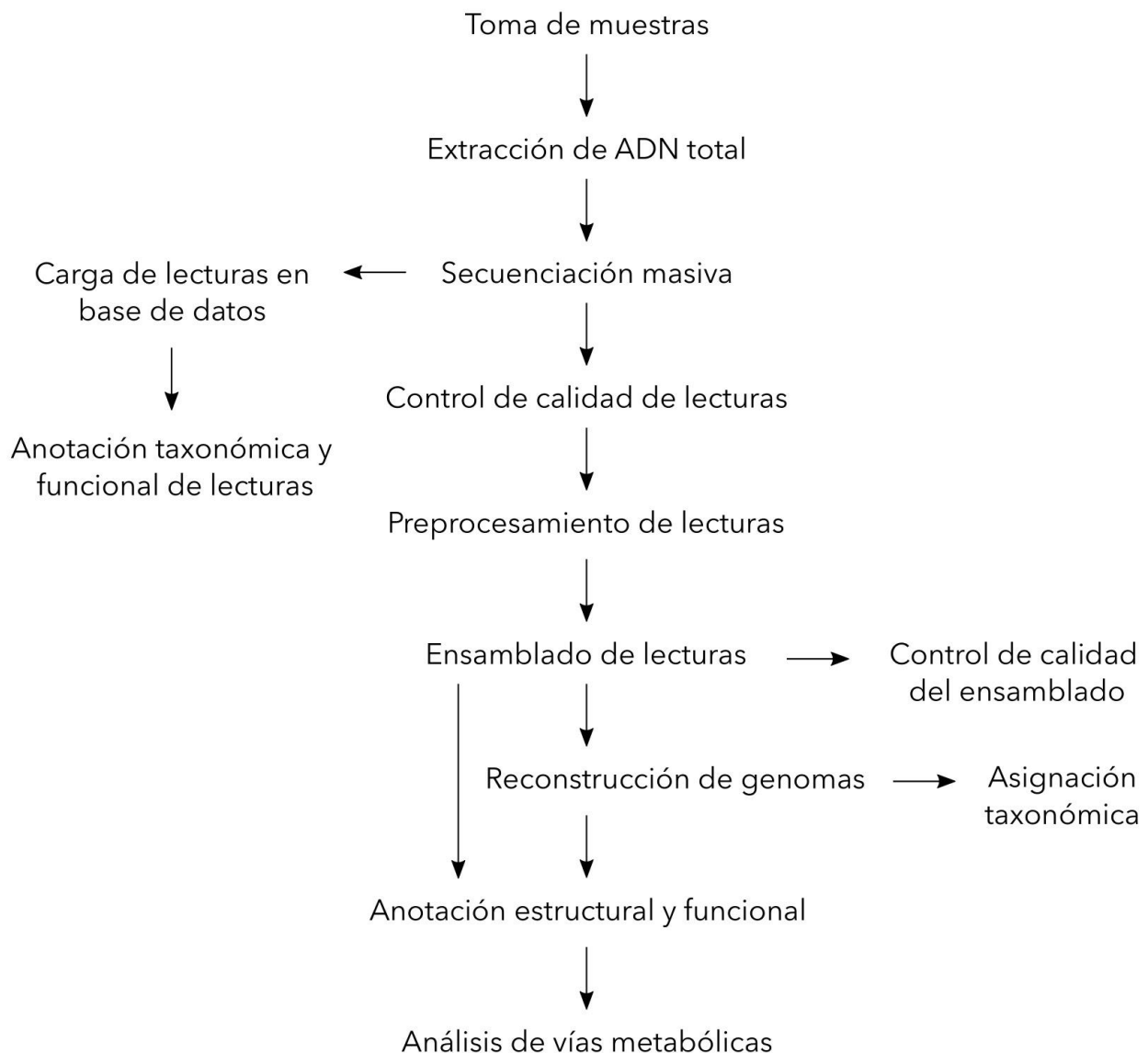


Figura 6. Flujo de trabajo de un análisis metagenómico por *shotgun sequencing*.

2) Secuenciación: La secuenciación del ADN total extraído se puede llevar a cabo a través de diferentes plataformas de secuenciación como Illumina, PacBio, Ion Torrent y SOLiD, entre otras. A pesar de que cada plataforma utiliza distintos enfoques y principios de funcionamiento, todas mantienen varios puntos en común:

- a) El ADN se fragmenta al azar y los fragmentos se unen a adaptadores específicos.
- b) Los fragmentos se amplifican por anclaje de sus adaptadores a una superficie sólida (microesferas, placas de secuenciación, etc).
- c) La detección de las bases ocurre en forma simultánea y masiva, a través de millones de reacciones paralelas.
- d) Las bases de cada fragmento de ADN son identificadas por señales emitidas (lumínicas, químicas, etc.) durante la síntesis del fragmento a partir de una hebra molde.
- e) Las lecturas generadas son cortas. A pesar de que se han logrado avances en este sentido, la relación señal/ruido es un factor limitante para la obtención de lecturas más largas.

3) Control de calidad y preprocesamiento de lecturas: Las tecnologías NGS pueden generar un número masivo de lecturas en un solo experimento. Sin embargo, dependiendo de la tecnología, se generan diferentes tipos y cantidades de errores en las lecturas. Por lo tanto, realizar un control de calidad y preprocesamiento de las lecturas crudas (*raw reads*) es necesario para no afectar la interpretación de los análisis subsiguientes. La herramienta online FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) permite de forma rápida y sencilla llevar a cabo un control de calidad. En el control de calidad se determina número, largo, % GC, calidad por base, calidad promedio, duplicaciones, número de bases ambiguas, y secuencias

adaptadoras de las lecturas. Luego, con la información obtenida, se puede realizar un preprocesamiento de las lecturas empleando herramientas como PRINSEQ (Schmieder & Edwards, 2011) o Trimmomatic (Bolger et al., 2014). En el preprocesamiento, dependiendo de las necesidades, se pueden remover secuencias adaptadoras, secuencias contaminantes, lecturas de baja calidad y/o de corta longitud, y extremos de baja calidad de las lecturas.

4) Ensamblado de lecturas: Luego del preprocesamiento, las lecturas son ensambladas para obtener *contigs* (secuencias consenso de múltiples lecturas superpuestas). El ensamblado puede ser *de novo* o comparativo. El ensamblado *de novo* no utiliza secuencias de genomas como referencia, mientras que el ensamblado comparativo sí utiliza como referencia las secuencias de genomas pre-existentes (Thomas et al., 2012). Los ensamblados *de novo* son computacionalmente más complicados y requieren herramientas bioinformáticas que optimicen el tiempo computacional. Los algoritmos para ensamblado *de novo* se pueden agrupar en 3 categorías, según estén basados en algoritmos *greedy*, *Overlap-Layout-Consensus* (OLC), o grafos de *De Bruijn* (DBG):

a) El algoritmo *greedy* busca siempre el mejor solapamiento entre lecturas. Cada vez que encuentra un par adecuado, une las lecturas correspondientes en una nueva secuencia más larga, y esta es incluida otra vez en el espacio de búsqueda. El proceso continúa iterativamente con todas las lecturas y fragmentos generados, hasta que no existan más uniones posibles. La información usada es, por lo tanto, local y no se tienen en cuenta las relaciones globales entre lecturas. Algunas herramientas de ensamblado que usan este método son SSAKE (Warren et al., 2007) y VCAKE (Jeck et al., 2007).

b) El algoritmo OLC utiliza un grafo de solapamientos, en el cual los nodos están constituidos por las lecturas y los arcos con sus pesos representan el

solapamiento existente entre dos lecturas. Esta estructura permite la aplicación de algoritmos sobre grafos que toman en cuenta las relaciones globales entre las lecturas. El proceso puede dividirse en tres etapas:

(1) En la primera etapa se realiza el cálculo de solapamientos entre todas las secuencias en forma similar al método *greedy*, por lo que puede ser paralelizado de igual manera.

(2) En la segunda etapa se genera un grafo de solapamientos, en el que cada lectura estará conectada con todas las lecturas con quienes tenga un nivel de solapamiento mínimo predefinido, generando múltiples recorridos posibles del grafo a partir de los cuales generar la secuencia final.

(3) A partir del grafo original se analizan los diversos grupos de lecturas solapantes y se determina el camino que atraviesa todos los nodos solo una vez para cada grupo (camino euleriano). Cada camino obtenido determina una secuencia consenso que da lugar a un *contig* final. Esta parte del proceso también se puede realizar en paralelo ya que cada grupo que da lugar a un *contig* es independiente de los otros, aunque esto depende del número de subgrafos que se generen en la etapa previa. Idealmente, desearíamos tener un solo subgrafo conteniendo todas las lecturas.

Algunas herramientas de ensamblado que usan este método son Celera Assembler (Miller et al., 2008), SGA (Simpson & Durbin, 2012) y Newbler (Margulies et al., 2005).

c) En un grafo de *De Bruijn* los nodos representan subsecuencias de longitud fija k (k -mers) y los arcos conectan nodos que se superponen por $k-1$ letras. El grafo modela las relaciones entre todas las subsecuencias de longitud k que forman parte de las lecturas de entrada. De esta manera, las lecturas no están representadas directamente por el grafo, sino que están implícitas en los caminos entre los nodos del grafo. Esta forma de representar las lecturas tiene varias ventajas. Una es la escalabilidad: en un grafo cuyos nodos son

lecturas, el tamaño del grafo es proporcional al número de nodos, con lo que se vuelve inmanejable para secuenciaciones de gran profundidad y lecturas cortas. La representación mediante *k-mers* logra que el grafo crezca muy poco o nada con el aumento de la cantidad de lecturas, lo que aumenta es la cantidad de veces que un *k-mer* dado aparece en las lecturas. También tiene la ventaja de que ya no es necesario realizar una comparación de todo contra todo, cada lectura se incorpora al grafo con las conexiones correspondientes, y luego solo se comparan entre sí las lecturas conectadas. En contrapartida, tiene altos requerimientos de memoria porque es necesario trabajar con el grafo completo. Las implementaciones varían mucho, pero en general, los programas que usan esta metodología suelen seguir un esquema común:

(1) Previamente a la construcción del grafo suele aplicarse algún método de corrección de errores. Como el algoritmo se basa en *matches* exactos, esta etapa es crucial para lograr ensamblados de alta calidad y para reducir la complejidad de la estructura de grafo generada. Durante la generación y simplificación del grafo también pueden aplicarse correcciones de errores, basadas en la calidad de las bases y los solapamientos entre nodos.

(2) En la siguiente etapa se genera el grafo de *De Bruijn*. Cada lectura es fragmentada en sus *k-mers* constituyentes, según un tamaño *k* definido previamente. La elección de *k* es clave para los resultados obtenidos y el desempeño del algoritmo, e implica un compromiso entre especificidad y sensibilidad. Cuanto mayor sea el tamaño *k*, más sencillo será el grafo generado y menos conexiones espurias habrá, pero también se perderán conexiones válidas entre lecturas con un solapamiento insuficiente o en regiones de baja cobertura, generando mayor cantidad de *gaps* en el ensamblado.

(3) Una vez generado el grafo, se usa la información de las lecturas para eliminar las zonas no consistentes y se aplican diversas simplificaciones y

modificaciones: se unen nodos redundantes, se eliminan burbujas de baja cobertura y se seleccionan los mejores caminos, hasta conformar un grafo irreducible.

(4) Finalmente, se obtienen los *contigs* finales a partir de cada subgrafo independiente, recorriendo los nodos en forma similar al método de OLC.

La mayoría de las herramientas de ensamblado usan este método, como SOAPdenovo2 (Luo et al., 2012) y SPAdes (Bankevich et al., 2012).

5) Control de calidad del ensamblado: Tras la obtención de los *contigs*, se determinan distintos parámetros para evaluar la calidad del ensamblado:

- a) Largo máximo de *contig* (*Maximum contig length*) [Kb]: Si bien es deseable que este valor sea alto, no necesariamente indica una mejor eficiencia en el ensamblado.
- b) Largo total del ensamblado (*Total assembly length*) [Mb]: Cuanto mayor es el largo total, mejor es el ensamblado, porque significa que incorporó mayor información.
- c) Número de *contigs* (*Contig count*) [Número]: Si los *contigs* son cortos, lo más probable es que tenga muchos, mientras que si son largos lo más probable es que tenga pocos.
- d) N50 [Kb]: Es la longitud de *contig* tal que la suma de *contigs* de igual o mayor tamaño representa el 50 % de todo el ensamblado. Cuanto mayor es el N50, mejor es el ensamblado.
- e) Porcentaje de mapeo de las lecturas (*Read mapping rate*) [%]: Es el porcentaje de lecturas que fueron incorporadas en el ensamblado. Cuanto mayor es el porcentaje, mejor es el ensamblado.
- f) Eficiencia de ensamblado (*Assembly performance*): Es el N50 multiplicado por el porcentaje de mapeo de las lecturas. Cuanto mayor es este valor, mejor es el ensamblado. Este estadístico permite

comparar los ensamblados obtenidos a partir de diferentes herramientas.

La herramienta QUAST (Gurevich et al., 2013) permite determinar los primeros cuatro estadísticos, mientras que la herramienta Bowtie2 (Langmead & Salzberg, 2012) permite determinar el porcentaje de mapeo de las lecturas.

6) Reconstrucción de genomas (*Binning*): Los *contigs* obtenidos del ensamblado de las lecturas pueden juntarse según su frecuencia de tetranucleótidos, % GC o su clasificación taxonómica, en grupos (*bins*) que representan un genoma individual (MAG, *Metagenome-assembled genome*) (Setubal, 2021). La reconstrucción de genomas completos (o *drafts* de alta calidad) de bacterias y arqueas permite analizar el potencial funcional de cada taxón, revelando la versatilidad metabólica de los microorganismos, los mecanismos de adaptación ecológica y las interacciones tróficas. La herramienta MetaWRAP (Uritskiy et al., 2018) es una de las más utilizadas para la reconstrucción de MAGs.

7) Anotación: La anotación de los *contigs* consiste en la identificación de los genes y sus funciones:

- a) Identificación de genes (Anotación estructural): Para secuencias de origen procariota, los genes codificantes de proteínas se pueden predecir fácilmente utilizando herramientas como FragGeneScan (Rho et al., 2010) o Prodigal (Hyatt et al., 2010). Mientras que los genes de ARN se pueden identificar empleando herramientas como tRNAscan-SE (Lowe & Eddy, 1997), Rnammer (Lagesen et al., 2007) o Infernal (Nawrocki & Eddy, 2013).
- b) Asignación de funciones a los genes (Anotación funcional): Una vez que se han identificado los genes, se les puede asignar una función. Para ello se utilizan herramientas, como InterProScan (Jones et al., 2014), que comparan los genes identificados con secuencias en bases

de datos. Existen diferentes bases de datos que pueden ser utilizadas. La de NCBI (<http://www.ncbi.nlm.nih.gov/>) es quizás la mas extensa y completa de todas, pero dado que prácticamente no existe un control de cómo anotan los usuarios las secuencias almacenadas, el uso de la misma puede traer como consecuencia una alta tasa de error en el proceso de anotación. Otras bases de datos muy populares son SEED (Overbeek et al., 2005), KEGG (Kanehisa et al., 2012), COG/KOG (Tatusov et al., 1997), eggNOG (Jensen et al., 2008), Pfam (Bateman et al., 2004) y TIGRFAMs (Haft et al., 2003), de las cuales las tres primeras tienen la ventaja extra de ser curadas a mano y estar organizadas en categorías y a diferentes niveles jerárquicos, permitiendo ordenar las secuencias anotadas según funciones o rutas metabólicas.

Los diferentes pasos de un análisis metagenómico por *shotgun sequencing* pueden llevarse a cabo de manera separada. Sin embargo, en la actualidad, existen diversos programas que realizan varios pasos del *pipeline* metagenómico. Entre los más usados se encuentran MG-RAST (Meyer et al., 2008) y MGnify (Mitchell et al., 2020). Ambos funcionan como repositorio de datos y *pipeline* de análisis a la vez. MG-RAST permite realizar control de calidad de las lecturas, predicción de genes, anotación funcional y anotación taxonómica. Mientras que MGnify permite realizar ensamblado de las lecturas, predicción de genes, anotación funcional y anotación taxonómica.

OBJETIVOS

Objetivo general

El objetivo general de la tesis es comprender los mecanismos moleculares que permiten a los microorganismos extremófilos de los Andes centrales resistir a las elevadas concentraciones ambientales de arsénico y, eventualmente, obtener energía metabólica a partir de su oxidorreducción.

Objetivos específicos

Los objetivos específicos de la primera parte de la tesis son:

1. Caracterizar, a partir del análisis metagenómico, la composición taxonómica y el potencial metabólico de diferentes comunidades microbianas de los Andes centrales.
2. Identificar los genes implicados en el metabolismo del arsénico, comparar su abundancia relativa entre las diferentes comunidades microbianas, y determinar su organización genómica.

MATERIALES Y MÉTODOS

Descripción del área de estudio

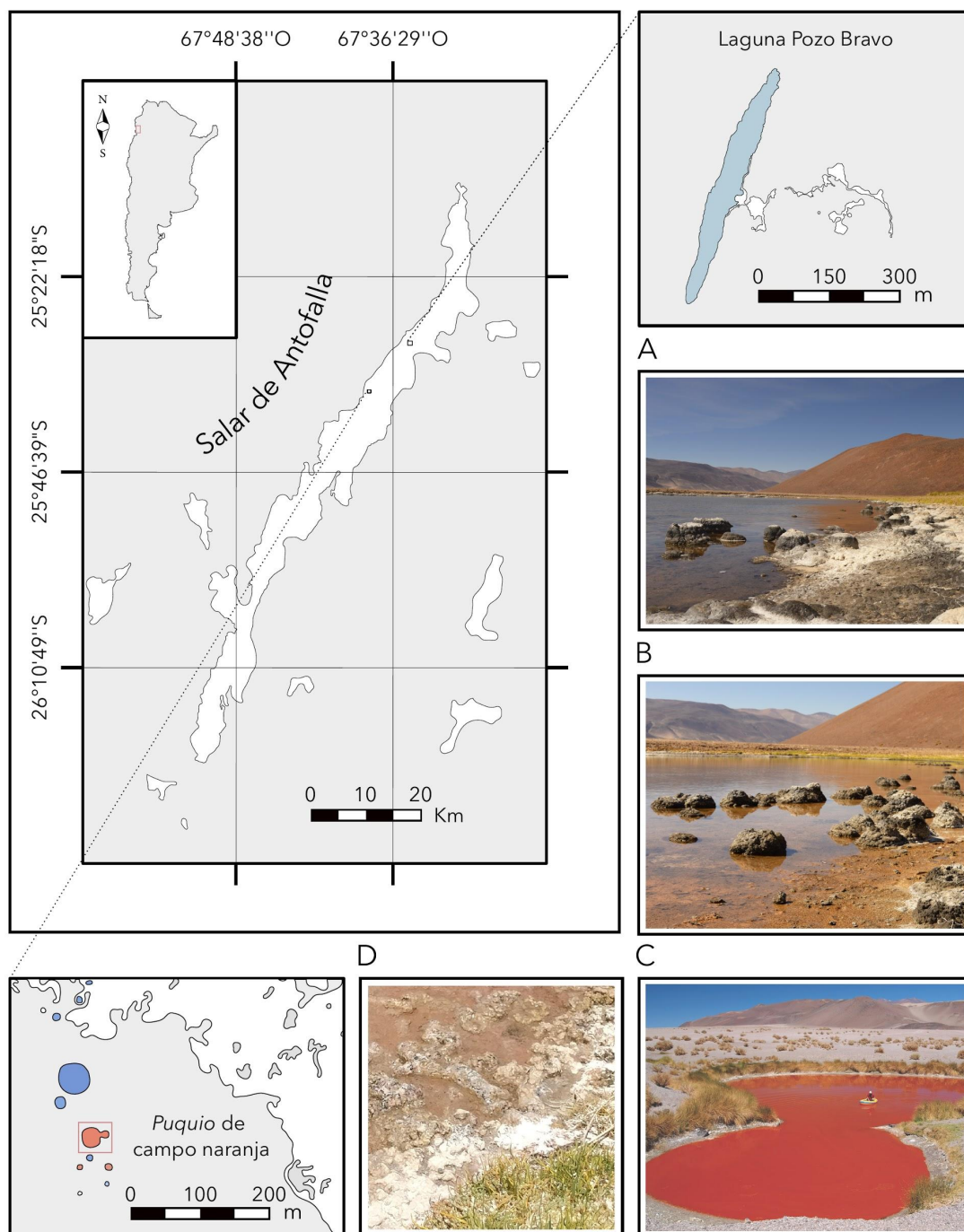


Figura 7. Mapa del salar de Antofalla (Puna argentina) en donde se detallan las regiones de la laguna Pozo Bravo y los *puquios* de campo. A, Vista

general de la laguna Pozo Bravo y su arrecife de microbialitos; B, Vista del margen suroeste de la laguna Pozo Bravo, donde los microbialitos y tapetes microbianos están expuestos a variaciones de la columna de agua; C, Vista general del *puquio* de campo naranja; D, Vista de la orilla del *puquio* de campo naranja, donde se desarrollan tapetes microbianos y microbialitos.

El salar de Antofalla es un salar de aproximadamente 500 km² situado a 3.340 m s.n.m. (sobre el nivel del mar), al sur de la Puna argentina en la provincia de Catamarca (Figura 7). En el norte del salar de Antofalla (25° 30' 48.55" S, 67° 34' 37.18" W) se encuentra la laguna Pozo Bravo. Esta es una laguna mediana, de 500 m de largo por 52 m de ancho, que alberga un arrecife de microbialitos potencialmente análogo a las primeras bioestructuras de la Tierra (Figura 7 A y B). Los microbialitos que conforman el arrecife tienen una altura de entre 0,05 y 0,40 m, y presentan una amplia gama de morfologías externas. La laguna Pozo Bravo además alberga tapetes microbianos que se extienden por casi todo el sustrato y cubren algunos de los microbialitos cuando el nivel del agua alcanza su superficie. Su mayor desarrollo se observa en el margen suroeste de la laguna, donde los microbialitos y tapetes microbianos están expuestos a variaciones de la columna de agua y, por ende, a períodos de sequía y humedad. Por este motivo, en esta zona se intercalan tapetes microbianos blandos y sus homólogos litificados (Lencina et al., 2021). Por otro lado, en el centro del salar de Antofalla (25° 36' 48.09" S, 67° 40' 20.81" W) se encuentra un conjunto de pequeñas lagunas, conocidas localmente como *puquios* u ojos de campo, sujetas a continuos procesos de desecación que aumentan progresivamente su salinidad (Carrizo et al., 2022). *Puquios* es la versión en español del término Quechua *Pukyu*, que significa manantiales (Vilcabana Sánchez et al., 2007). Los *puquios* u ojos de campo albergan numerosos tapetes microbianos y microbialitos en sus márgenes (Figura 7 C y D).

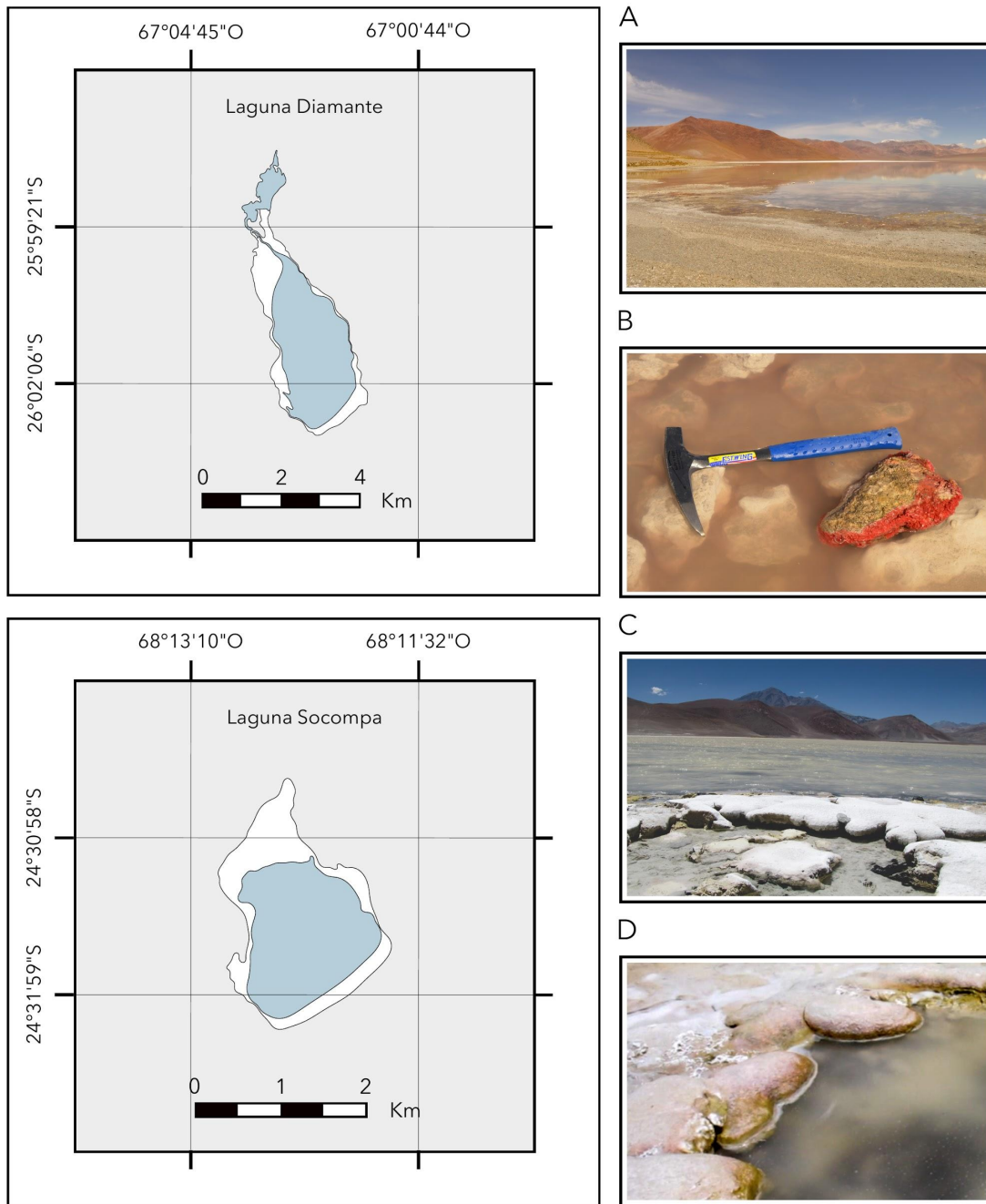


Figura 8. Mapas de la laguna Diamante (volcán Galán) y la laguna Socompa (salar Socompa en volcán Socompa). A, Vista general de la laguna Diamante; B, Vista del *biofilm* rojo asociado a *Gaylussita* que se desarrolla, durante los meses cálidos del año, en la parte inferior de la rocas volcánicas de la laguna Diamante; C, Vista general de la laguna Socompa y sus microbialitos; D, Vista de la orilla de la laguna Socompa, donde se desarrollan los microbialitos.

Al este del salar de Antofalla, en la provincia de Catamarca, se encuentra el volcán Galán. Este es uno de los pocos supervolcanes que existen en la Tierra, con una caldera de 40 km de diámetro y una altura máxima de 5.912 m s.n.m. Entre el Mioceno tardío y el Plioceno (hace 7 y 4 Ma) generó numerosas erupciones de gran volumen que dieron lugar a la deposición de ignimbritas (Ruggieri et al., 2010; Soler et al., 2007; Sparks et al., 1985). En la actualidad, la caldera del volcán Galán alberga una laguna de aproximadamente 5 km de largo por 2 km de ancho (26° 0' 49.75" S, 67° 2' 10.08" W) que se conoce como laguna Diamante (Farías, 2020) (Figura 8 A). El ingreso de agua a la laguna se lleva a cabo por diversas fuentes hidrotermales. El agua que ingresa es alcalina y presenta baja salinidad. Sin embargo, el pH y la conductividad del agua aumentan notablemente en la laguna debido a la constante evaporación. En la laguna suelen precipitar minerales inestables en condiciones normales y por lo tanto poco frecuentes, como la Gaylussita ($\text{Na}_2\text{Ca}(\text{CO}_3)_2 \cdot 5\text{H}_2\text{O}$). De hecho, en el fondo de rocas volcánicas, se forman *biofilms* rojos asociados a Gaylussita (Rascovan et al., 2016) (Figura 8 B). La formación de estos *biofilms* es dependiente de la temperatura, dado que solo fueron observados en el sector de la laguna donde desembocan los afluentes hidrotermales y el agua es más cálida (28°C respecto a los 11°C en promedio de la laguna). Por otro lado, los *biofilms* rojos asociados a gaylussita solo pueden ser encontrados en verano, cuando el nivel del agua es menor debido a la evaporación, y la concentración de sales y metales pesados es mayor. Mientras que en invierno, cuando el agua alcanza los 4°C de temperatura y el nivel del agua es mayor, no se observa la formación de estas comunidades microbianas. Además de *biofilms* rojos, en la laguna Diamante se han reportado microbialitos que exhiben una mesoestructura moteada (Sancho-Tomás et al., 2020).

Al norte de la Puna argentina, en la provincia de Salta, se encuentra el volcán Socompa (24° 23' 45" S, 68° 14' 44.42" W) con una altura máxima de 6.051 m

s.n.m. En la ladera del volcán, se extiende un salar del mismo nombre que alberga una laguna de aproximadamente 2 km de largo por 1,5 km de ancho, conocida como laguna Socompa ($24^{\circ} 31' 38.4''$ S, $68^{\circ} 12' 23.7''$ W) (Figura 8 C). Numerosos microbialitos se desarrollan al sur de la laguna, donde desembocan diversos afluentes hidrotermales. Dichos microbialitos presentan una mesoestructura laminada, y forman biohermas amplias y redondeadas. Las secciones verticales muestran una estratificación visualmente clara. La superficie expuesta continuamente al aire está cubierta por una costra blanca-rosada, mientras que la superficie expuesta intermitentemente al agua es verde-amarilla (Figura 8 D). Por debajo de esta cubierta se encuentra una capa de color verde oscuro de 0,5-1,5 mm de espesor, que se desvanece gradualmente y desaparece a profundidades de 3-5 mm. Las partes más profundas se caracterizan por la alternancia de capas de color marrón claro (5-20 mm de espesor) y marrón oscuro (0,5-1,5 mm de espesor). En las capas más oscuras se encuentran ocasionalmente manchas negras, presumiblemente signo de una deposición de ceniza volcánica (Farías et al., 2013).

El salar de Atacama, ubicado al norte de Chile, es uno de los salares más extensos (3.000 km^2) del planeta (H. Alonso & Risacher, 1996) (Figura 9). Al norte del salar ($23^{\circ} 08' 07.6''$ S, $68^{\circ} 15' 31.2''$ W), se encuentra la laguna Tebenquiche de aproximadamente 1,5 km de largo por 3 km de ancho. Esta laguna alberga endoevaporitas y tapetes microbianos (Farías et al., 2014; Fernandez et al., 2016) (Figura 9 A). Por otro lado, al sur del salar se encuentra la laguna Brava, de aproximadamente 500 m de largo por 1.500 m de ancho ($23^{\circ} 43' 43.5''$ S, $68^{\circ} 14' 56.0''$ W) (Figura 9 B). Esta laguna no alberga endoevaporitas, pero sí presenta tapetes microbianos y microbialitos de tipo estromatolito y oncoide (estromatolitos esféricos no adheridos a un sustrato) (Vignale, Kurth, et al., 2021) (Figura 9 C y D).

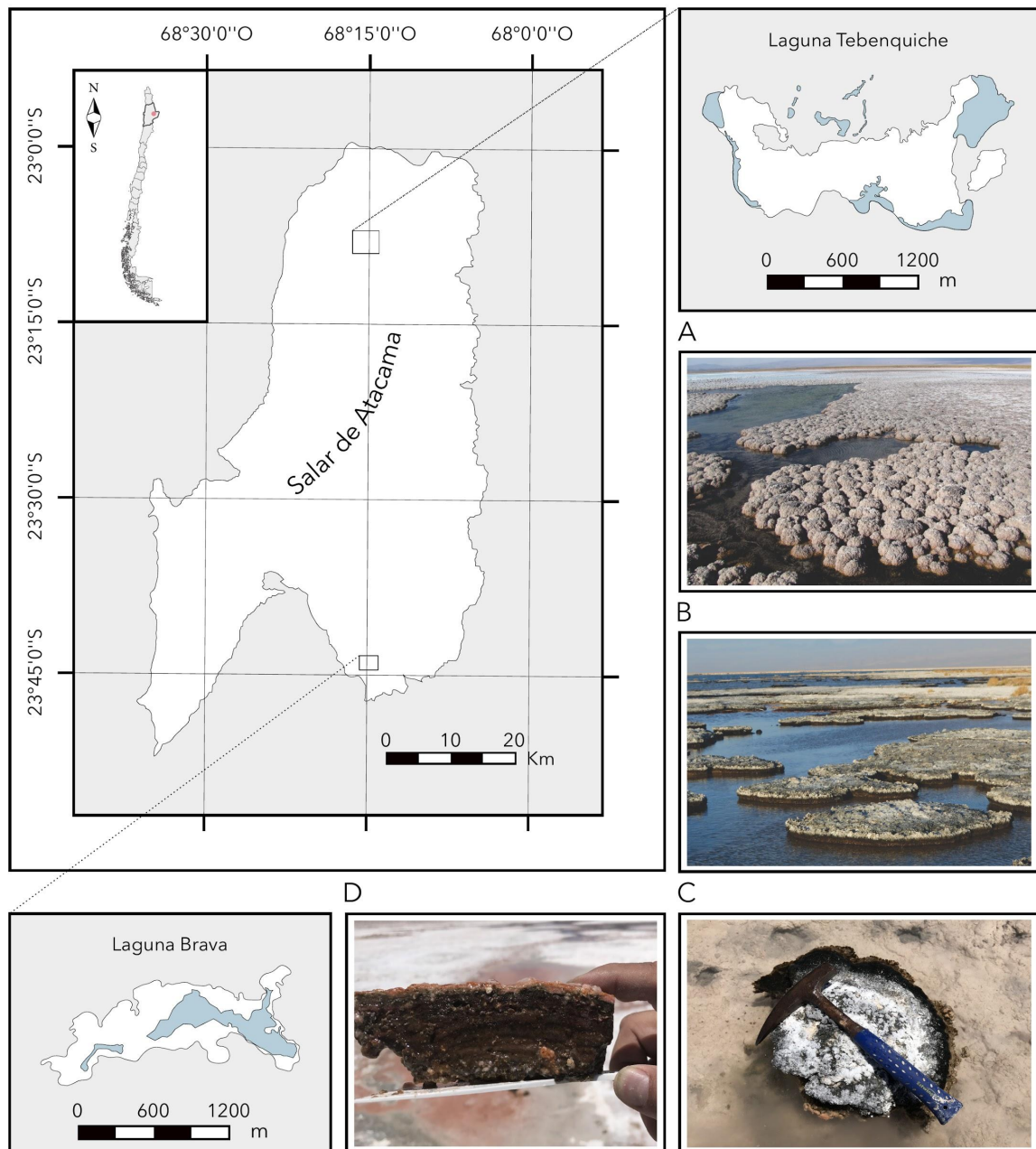


Figura 9. Mapa del salar de Atacama (Chile) en donde se detallan las regiones de la laguna Tebenquiche y la laguna Brava. A, Vista general de la laguna Tebenquiche y sus endoevaporitas; B, Vista general de la laguna Brava y sus microbialitos; C, Vista de un microbialito de la laguna Brava; D, Vista de un tapete microbiano de la laguna Brava.

Toma de muestras y parámetros ambientales

En primer lugar se llevó a cabo la toma de muestras de diferentes ecosistemas microbianos andinos. Debido a las condiciones climáticas particulares de los Andes centrales, el muestreo se realizó entre los meses cálidos de octubre y abril de diferentes años. En el salar de Antofalla, se tomaron muestras de tapetes microbianos de la laguna Pozo Bravo y de un *puquio* de campo naranja. En la caldera del volcán Galán, se recolectaron *biofilms* rojos asociados a gaylussita de la laguna Diamante. En el salar Socompa, se tomaron muestras de estromatolitos de la laguna Socompa. Mientras que en el salar de Atacama, se muestrearon tapetes microbianos de las lagunas Tebenquiche y Brava.

La toma de muestras se llevó a cabo empleando pinzas, espátulas, tubos, frascos, y/o bolsas previamente esterilizadas mediante autoclave a 121°C. Las muestras recolectadas se almacenaron en RNAlater® a 4°C y en oscuridad, para preservar el material genético hasta el momento de su extracción. Junto a los muestreos se determinaron parámetros geográficos y fisicoquímicos de cada ambiente. La latitud, longitud y altitud de los sitios de muestreo se determinaron utilizando un dispositivo GPS Garmin® eTrex 20x. El pH, la temperatura y la conductividad eléctrica del agua se determinaron *in situ* empleando una sonda multiparamétrica Hanna HI 9811-5. Mientras que el análisis químico del agua se llevó a cabo en la Estación Experimental Agroindustrial Obispo Colombres, San Miguel de Tucumán, Argentina. Para ello se empleó espectrometría de masa con plasma acoplado inductivamente (ICP-MS).

Extracción de ADN total y secuenciación

Debido a que las muestras de ecosistemas microbianos andinos se tomaron en diferentes años, la metodología de extracción de ADN total y secuenciación difirió levemente para cada una de ellas.

La extracción de ADN total, de dos muestras de tapete microbiano de la laguna Brava y dos muestras de tapete microbiano de la laguna Tebenquiche, se llevó a cabo utilizando el *PowerBiofilm™ DNA Isolation Kit* (MO BIO Laboratories, Inc.) de acuerdo con las instrucciones del fabricante. En INDEAR (Rosario, Argentina) se realizó la secuenciación masiva (*shotgun*) de las cuatro muestras de ADN a través de la plataforma de secuenciación Illumina HiSeq 1500. Las lecturas *paired end* crudas (*raw sequence data*) se depositaron en el repositorio de secuencias nucleotídicas ENA (*European Nucleotide Archive*) bajo el número de acceso ERP107533.

Para la extracción de ADN total de una muestra de estromatolito de la laguna Socompa se siguieron diversos pasos. En primer lugar, la muestra fue liofilizada y homogeneizada. El polvo obtenido (0,5 g) se lavó tres veces añadiendo 5 ml de una solución de NaCl al 10 % (p/v), incubando a 40°C durante 15 minutos para disolver las sustancias poliméricas extracelulares, y centrifugando a 8.200 g durante 15 minutos. El *pellet* de células obtenido se resuspendió en 5 ml de *buffer* de lisis (Tris-HCl, pH 8, 100 mM; NaCl 1,5 M; EDTA, pH 8, 100 mM; Na₃PO₄, pH 8, 100 mM; CTAB 1 %). Luego, las células se congelaron a -120°C y se descongelaron en un baño de agua a 65°C. Este proceso se repitió tres veces. Posteriormente, la mezcla se incubó con lisozima (1 mg ml⁻¹) durante 30 min a 37°C, y se le adicionó proteinasa K (0,1 mg ml⁻¹) y SDS (dodecilsulfato sódico) al 3 % (p/v) final. Tras una incubación *overnight* a 60°C, se centrifugó por 10 min a 10.000 g y se colectó el sobrenadante. Las extracciones orgánicas se hicieron primero con un volumen de fenol:cloroformo:alcohol isoamílico (25:24:1) y luego con un

volumen de cloroformo:alcohol isoamílico (24:1). El ADN (fase acuosa) se precipitó con isopropanol frío (0,7 vol.), se incubó durante 2 horas a -20°C y se lavó tres veces con etanol frío al 70 % (v/v), recogiendo el ADN entre cada lavado mediante centrifugación durante 15 min a 15.000 g y 4°C. Por último, el ADN se secó en aire estéril y se resuspendió en agua Milli-Q® libre de ADNasa. La secuenciación masiva (*shotgun*) de la muestra de ADN se realizó a través de la plataforma de secuenciación Illumina Genome Analyzer (GAIIx). Las lecturas *paired end* crudas (*raw sequence data*) se depositaron en el repositorio de secuencias nucleotídicas NCBI SRA (*NCBI Sequence Read Archive*) bajo el número de acceso SRP072938.

La extracción de ADN total, de una muestra de tapete microbiano de la laguna Pozo Bravo, una muestra de tapete microbiano de un *puquio* de campo naranja, y tres muestras de *biofilm* asociado a gaylussita de la laguna Diamante, se llevó a cabo utilizando el *FastDNA® SPIN Kit for Soil* (MP Biomedicals, United States) de acuerdo a las especificaciones del fabricante. La secuenciación masiva (*shotgun*) de las cinco muestras de ADN se realizó a través de la plataforma de secuenciación Illumina MiSeq. Las lecturas *paired end* crudas (*raw sequence data*) se depositaron en el repositorio de secuencias nucleotídicas ENA bajo los números de acceso ERP113390 (Pozo Bravo), ERP113392 (*Puquio* de campo naranja) y ERP113536 (Diamante).

Control de calidad y preprocesamiento de las lecturas metagenómicas

Empleando la herramienta *online* FastQC v.0.11.5 (*Babraham Bioinformatics*), se llevó a cabo un control de calidad de las lecturas metagenómicas obtenidas. En dicho control de calidad se determinó para cada muestra: número total de lecturas (*total sequences*), % GC de lecturas, calidad por base de lecturas (*per base sequence quality*), calidad promedio de lecturas (*per sequence quality scores*), contenido de bases de lecturas (*per base*

sequence content), contenido de bases ambiguas (N) de lecturas (*per base N content*), distribución del largo de lecturas (*sequence length distribution*), duplicaciones de lecturas, y contenido de secuencias adaptadoras de las lecturas (*adapter content*). Posteriormente, teniendo en cuenta la información obtenida en el control de calidad, se realizó un preprocesamiento de las lecturas empleando la herramienta Trimmomatic v.0.36 (Bolger et al., 2014). Los parámetros utilizados para las lecturas metagenómicas de Brava y Tebenquiche fueron los siguientes: ILLUMINACLIP = TruSeq3-PE-2.fa:2:30:5, SLIDINGWINDOW = 4:20, AVGQUAL = 28, MINLEN = 50. De esta manera, se removieron secuencias adaptadoras, lecturas con calidad promedio menor a Q28, lecturas con un largo menor a 50 bases, y extremos de lecturas con calidad promedio (cada cuatro bases) menor a Q20. Los parámetros utilizados para las lecturas metagenómicas de Socompa fueron los siguientes: ILLUMINACLIP = TruSeq2-PE.fa:2:30:5, SLIDINGWINDOW = 4:20, AVGQUAL = 15, y MINLEN = 36. De esta forma, se removieron secuencias adaptadoras, lecturas con calidad promedio menor a Q15, lecturas con un largo menor a 36 bases, y extremos de lecturas con calidad promedio (cada cuatro bases) menor a Q20. Mientras que los parámetros utilizados para las lecturas metagenómicas de Pozo Bravo, *Puquio* de campo naranja, y Diamante fueron: ILLUMINACLIP = TruSeq2-PE.fa:2:30:5, SLIDINGWINDOW = 4:20, AVGQUAL = 20, y MINLEN = 50. Así se removieron secuencias adaptadoras, lecturas con calidad promedio menor a Q20, lecturas con un largo menor a 50 bases, y extremos de lecturas con calidad promedio (cada cuatro bases) menor a Q20.

Ensamblado de las lecturas metagenómicas

Luego del preprocesamiento, las lecturas metagenómicas fueron ensambladas *de novo* para obtener *contigs*. El ensamblado *de novo* se llevó a cabo empleando el *pipeline* metaSPAdes (Nurk et al., 2017) de la herramienta SPAdes v.3.11.0 (Bankevich et al., 2012), que utiliza grafos de *De Bruijn* y un enfoque iterativo *multi-k-mer* para optimizar el proceso. SPAdes no reduce el tamaño del conjunto de datos sustituyendo las lecturas por *contigs* preensamblados en cada iteración de la construcción del grafo de *De Bruijn*, sino que utiliza la información completa de las lecturas junto con los *contigs* preensamblados en cada paso. De esta forma tiene en cuenta las pequeñas inserciones o deleciones (*indels*), que ocurren con mayor frecuencia cuando el ensamblado se hace con *k-mers* cortos. Para una mejor resolución de las repeticiones, SPAdes implementa grafos de *De Bruijn* emparejados. A diferencia de la mayoría de los ensambladores, que utilizan la información de los extremos emparejados para los pasos de simplificación después de la construcción estándar del grafo de *De Bruijn*, SPAdes incorpora directamente esta información en el grafo utilizando *k-bimers*. Estos son conjuntos de *k-mers* derivados de pares de lecturas y separados por un valor de distancia estimado. A continuación, SPAdes corrige y ajusta iterativamente la estimación de la distancia de cada *k-bimer*, teniendo así en cuenta la distribución no uniforme de la longitud de inserción de la mayoría de las bibliotecas de secuenciación masiva. El rango de *k-mer* para la construcción iterativa del grafo de *De Bruijn* se determinó automáticamente en función de la longitud de la lectura y el tipo de datos de la secuencia.

Una vez finalizado el ensamblado, se procedió a evaluar la calidad del mismo. Utilizando el *pipeline* MetaQUAST (Mikheenko et al., 2016) de la herramienta QUAST v.4.6 (Gurevich et al., 2013), se determinó el largo máximo de *contig* (*maximum contig length*), el largo total del ensamblado (*total assembly*

length), el número de *contigs* (*contig count*) y el N50. Posteriormente, empleando la herramienta Bowtie2 v.2.3.2 (Langmead & Salzberg, 2012), se determinó el porcentaje de mapeo de las lecturas sobre los *contigs* obtenidos (*read mapping rate*). La eficiencia de ensamblado se obtuvo multiplicando el N50 por el porcentaje de mapeo de las lecturas.

Reconstrucción de genomas a partir de los metagenomas ensamblados

La reconstrucción de genomas (MAGs) a partir de los *contigs* se llevó a cabo empleando la herramienta MetaWRAP v.1.1 (Uritskiy et al., 2018). Esta herramienta presenta un enfoque híbrido, ya que combina tres *software*: metaBAT (Kang et al., 2015), CONCOCT (Alneberg et al., 2013), y MaxBin2 (Y.-W. Wu et al., 2016). Además, una vez que obtiene el conjunto inicial de MAGs, extrae las lecturas que pertenecen a cada MAG y las reensambla para mejorar la completitud y reducir la contaminación. Tras la reconstrucción de los MAGs, se llevó a cabo un control de calidad de los mismos utilizando la herramienta CheckM v.1.0.13 (Parks et al., 2015). Este control de calidad permitió determinar su completitud y contaminación, utilizando un conjunto de genes que son ubicuos y de copia única. Luego, las métricas obtenidas fueron visualizadas en R v.3.5.3 (Ihaka & Gentleman, 1996). Utilizando las métricas del control de calidad, se llevó a cabo la desreplicación de los MAGs con la herramienta dRep v.1.4.3 (Olm et al., 2017). Esta herramienta identifica grupos de MAGs con alta similitud (ANI, *Average Nucleotide Identity* > 95%) y luego selecciona el MAG representativo de cada grupo. De esta manera, se filtran MAGs que son redundantes. Por último, se realizó la asignación taxonómica de los MAGs por comparación contra distintas bases de datos de referencia, como RefSeq (Enero de 2018) (Haft et al., 2018) y GenBank (Septiembre de 2018) (Benson et al., 2018). Para ello se utilizó el *pipeline* mashdiff.sh(<https://github.com/Finn-Lab/MGS-gut/blob/master/pipelines/mashdiff.sh>).

Anotación estructural y funcional automática

La predicción de los genes codificantes de proteínas (anotación estructural) en los metagenomas ensamblados y MAGs, se realizó con la herramienta Prodigal (*PROkaryotic DYnamic programming Gene-finding ALgorithm*) v.2.6.3 (Hyatt et al., 2010). La gran ventaja de Prodigal es que se trata de un algoritmo de aprendizaje automático no supervisado. Es decir, no necesita que se le proporcione ningún dato de entrenamiento o conjunto de datos, sino que aprende automáticamente las propiedades del archivo genómico dado, incluyendo la presencia de sitios de unión al ribosoma (RBS, *Ribosome-binding sites*), y el uso de codones. Una vez identificados los genes codificantes de proteínas, se llevó a cabo la caracterización funcional de los mismos (anotación funcional) empleando la herramienta InterProScan v.5.34-73.0 (Jones et al., 2014). InterProScan es un software que permite comparar secuencias desconocidas de genes (nucleotídicas y/o aminoacídicas) con las secuencias en las bases de datos de InterPro (Mitchell et al., 2019), y así asignar funciones. Como resultado se obtienen distintos archivos de salida (TSV, XML y GFF3) con la información de la anotación funcional. Luego de la caracterización funcional de los genes, se llevó a cabo un análisis de las vías metabólicas presentes en los MAGs. Para ello, se empleó la herramienta Genome Properties v.2.0 (Richardson et al., 2019), que utiliza los archivos de salida en formato TSV de InterProScan como archivos de entrada.

Análisis de genes implicados en el metabolismo del arsénico

Debido a que los genes que participan en el metabolismo del arsénico se encuentran poco caracterizados, la identificación de los mismos se llevó a cabo de manera diferencial utilizando la herramienta HMMER v.3.1b2 (Finn et

al., 2011) que emplea perfiles de modelos ocultos de Markov (HMMs, *Hidden Markov Models*). La construcción de dichos perfiles, para cada uno de los genes del metabolismo del arsénico (Tabla 2), se realizó de la siguiente manera:

- 1) Se descargó una secuencia aminoacídica de referencia del gen de interés de una base de datos curada, como Swiss-Prot (Boeckmann et al., 2003).
- 2) Empleando dicha secuencia como *query*, se realizó una búsqueda de secuencias homólogas lejanas con el *software* PSI-BLAST (*Position-Specific Iterated BLAST*) (Altschul et al., 1997). Para ello, se utilizó la base de datos de proteínas no redundantes (nr).
- 3) Una vez obtenidas las secuencias homólogas lejanas, se eliminaron las secuencias redundantes (secuencias con más del 90 % de similitud) empleando la herramienta cd-hit v.4.6.8 (W. Li & Godzik, 2006).
- 4) Con la herramienta Clustal-Omega v.1.2.4 (Sievers et al., 2011) se realizó un alineamiento múltiple de las secuencias homólogas lejanas no redundantes.
- 5) Finalmente, utilizando el alineamiento múltiple, se construyó el perfil HMM con la herramienta HMMER v.3.1b2 (Finn et al., 2011).

Una vez identificados los genes del metabolismo del arsénico, se determinó su abundancia en cada metagenoma. Primero, se mapearon las lecturas metagenómicas contra los genes identificados utilizando la herramienta Bowtie2 v.2.3.2 (Langmead & Salzberg, 2012). Luego, con el número de lecturas mapeadas se calculó la abundancia de cada gen, normalizada por la longitud del gen y el tamaño del metagenoma (cuantificación en FPKM, Fragmentos por kilobase por millón de lecturas mapeadas):

$$\text{Abundancia del gen}_i \text{ (FPKM)}_i = \frac{(\text{N}^\circ \text{ de lecturas mapeadas} * 10e3 * 10e6)}{(\text{N}^\circ \text{ total de lecturas} * \text{largo del gen}_i \text{ [pb]})}$$

Por último, se realizó un estudio de la organización genómica de los genes del metabolismo del arsénico. Para ello, se colocaron las coordenadas de dichos genes en archivos CSV. Luego, se importaron los archivos a R v.3.5.3 (Ihaka & Gentleman, 1996) y se visualizaron con los paquetes *gggenes* y *ggplot2*.

Análisis taxonómico y funcional de las lecturas metagenómicas

Si bien la mayoría de los análisis metagenómicos se realizan a partir de los *contigs*, también resulta útil analizar las lecturas, ya que pueden contener información que se perdió en el proceso de ensamblado. Teniendo en cuenta esto, se llevó a cabo un análisis funcional y taxonómico de las lecturas con el *pipeline* metagenómico MGnify v.4.1 (Mitchell et al., 2020). Este pipeline, por un lado, identifica los genes de la subunidad pequeña (SSU, *small subunit*) del ácido ribonucleico ribosomal y los clasifica taxonómicamente usando MAPseq (Matias Rodrigues et al., 2017) y SILVA (Quast et al., 2013). Por otro lado, identifica los genes codificantes de proteínas con FragGeneScan (Rho et al., 2010), y los clasifica funcionalmente usando la base de datos InterPro (Mitchell et al., 2019) y los términos de Gene Ontology (GO *terms*) (The Gene Ontology Consortium, 2019). El conjunto de anotaciones taxonómicas y funcionales puede visualizarse en la página web de MGnify a través de una serie de gráficos y tablas. Aunque los datos también pueden ser descargados para subsiguientes análisis.

Análisis de diversidad

El análisis de los datos taxonómicos se llevó a cabo en R v.3.5.3 (Ihaka & Gentleman, 1996) con el paquete phyloseq (McMurdie & Holmes, 2013). Por un lado, se filtraron las unidades taxonómicas operacionales (OTUs,

Operational Taxonomic Units) con una abundancia media menor a $1e^{-5}$, se generó un *subset* con los datos de *Bacteria* y *Archaea*, se transformaron los datos a abundancias relativas, y se graficaron las abundancias relativas para los diferentes niveles taxonómicos.

Por otro lado, luego de realizar una rarefacción, se determinó el índice de diversidad Shannon-Wiener (Shannon, 1948) para cada una de las muestras:

$$H' = -\sum[(p_i) * \log(p_i)],$$

donde:

H' = Índice de diversidad Shannon-Wiener

$p_i = (n/N)$ = Proporción de individuos de la especie i respecto al total de individuos

n_i = Número de individuos de la especie i

N = Número total de individuos

Cabe destacar que, cuanto mayor es el índice de Shannon-Wiener, mayor es la diversidad.

RESULTADOS

Condiciones ambientales de las lagunas altoandinas

Tabla 3. Parámetros geográficos y fisicoquímicos de las lagunas argentinas.

	Catamarca			Salta
	Laguna Diamante	<i>Puquio</i> de campo naranja	Laguna Pozo Bravo	Laguna Socompa
Latitud	26°00'49,75"	25°36'48,09"	25°30'48,55"	24°31'38,4"
Longitud	67°02'10,08"	67°40'20,81"	67°34'37,18"	68°12'23,7"
Altitud [m]	4.590	3.335	3.333	3.570
pH	11,0	8,1	8,0	8,6
Temperatura [°C]	20,0	27,3	23,0	22,0
Conductividad [mS cm ⁻¹]	210,0	110,0	149,0	115,0
Cloro [mg L ⁻¹]	125.000	46.100	75.200	52.234
Sodio [mg L ⁻¹]	51.200	23.600	43.100	37.113
Sulfato [mg L ⁻¹]	9.890	1.090	1.900	31.847
Potasio [mg L ⁻¹]	8.900	2.732	1.100	5.298
Calcio [mg L ⁻¹]	1.274	1.785	1.590	2.383
Magnesio [mg L ⁻¹]	919	7.568	1.240	4.090
Arsénico [mg L ⁻¹]	340	6,9	0,9	18,5

En base al análisis que se llevó a cabo de los parámetros ambientales (Tablas 3 y 4), se puede decir que todas las lagunas altoandinas estudiadas presentan condiciones extremas. Por un lado, todas se encuentran a más de 2.000 m s.n.m., y por ende presentan baja presión de oxígeno. Esta presión sería menor a 16 Kilopascales (KPa) de acuerdo a la relación que existe entre la altitud y la presión parcial de oxígeno (Peacock, 1998). Por otro lado, todas presentan un pH mayor o igual a 8, una concentración de sales mayor a 50.000 mg L⁻¹ (35.000 mg L⁻¹ es el contenido medio de sal del mar), y una

conductividad eléctrica mayor a 70 mS cm^{-1} , por lo que pueden ser clasificadas como alcalinas e hipersalinas (Kawabata et al., 1997; Nthunya et al., 2018). Otra característica que comparten todas las lagunas es la elevada concentración de arsénico de origen volcánico ($0,9 - 340 \text{ mg L}^{-1}$). Los cuerpos de agua que presentan las mayores concentraciones naturales de arsénico del mundo son los lagos Searles (225 mg L^{-1}), Mono (15 mg L^{-1}), Big Salt ($1,5 \text{ mg L}^{-1}$), y Walker (1 mg L^{-1}) en Estados Unidos de América (Oremland et al., 2004), y las fuentes hidrotermales de Wairakei, Nueva Zelanda ($4,5 \text{ mg L}^{-1}$), de la caldera Uzon, Kamchatka, Rusia ($2-30 \text{ mg L}^{-1}$), y de Tongonan, Filipinas ($20-34 \text{ mg L}^{-1}$) (Bundschuh & Maity, 2015), entre otros. Dado que todas las lagunas altoandinas estudiadas presentan niveles de arsénico similares o mayores a los de estos lagos y fuentes hidrotermales, se podrían incluir dentro del grupo de cuerpos de agua con mayor concentración de arsénico del mundo.

Tabla 4. Parámetros geográficos y fisicoquímicos de las lagunas chilenas.

	Región II	
	Laguna Brava	Laguna Tebenquiche
Latitud	23°43'43,5"	23°08'07,6"
Longitud	68°14'56,0"	68°15'31,2"
Altitud [m]	2.300	2.340
pH	8,0	8,6
Temperatura [°C]	24,6	21,5
Conductividad [mS cm^{-1}]	107,3	138,7
Cloro [mg L^{-1}]	43.057	64.290
Sodio [mg L^{-1}]	17.810	40.640
Sulfato [mg L^{-1}]	5.625	20.388
Potasio [mg L^{-1}]	6.072	5.121
Calcio [mg L^{-1}]	812	1.164
Magnesio [mg L^{-1}]	2.568	2.705
Arsénico [mg L^{-1}]	10,0	2,3

Si bien todas las lagunas altoandinas estudiadas presentan condiciones ambientales extremas, la laguna Diamante es sin duda la más extrema de todas. La altitud, el pH, la concentración de sales y el nivel de arsénico son considerablemente mayores en la laguna Diamante que en el resto de las lagunas estudiadas. De hecho, hasta donde se sabe, la laguna Diamante sería la laguna con la mayor concentración natural de arsénico del mundo.

Composición taxonómica de los ecosistemas microbianos andinos

El análisis de los genes de la subunidad pequeña del ácido ribonucleico ribosomal, llevado a cabo a partir de las lecturas metagenómicas, permitió determinar y comparar la estructura taxonómica de los diferentes ecosistemas microbianos andinos (Figuras 10 y 11). El *biofilm* rojo de la laguna Diamante está dominado principalmente por arqueas del filo *Euryarchaeota* (77,4-91,2 %), aunque también presenta una pequeña proporción de bacterias pertenecientes a los filos *Pseudomonadota* (0,8-12,0 %), *Bacillota* (0,2-1,1 %), y *Bacteroidota* (0,0-0,3 %). Las arqueas del filo *Euryarchaeota* presentes en el *biofilm* pertenecen en su totalidad a la clase *Halobacteria*. Dentro de esta clase, se encuentran miembros de los órdenes *Halobacteriales* (15,5-20,5 %) y *Haloferacales* (43,7-50,7 %). En el orden *Halobacteriales*, el género dominante es *Natronomonas* (5,9-8,6 %), mientras que en el orden *Haloferacales* los géneros dominantes son *Halorubrum* (22,6-26,3 %), *Halonotius* (1,6-8,6 %), y *Halohasta* (1,5-1,7 %). Las pocas bacterias del filo *Pseudomonadota* presentes en el *biofilm* pertenecen casi en su totalidad a la clase *Gammaproteobacteria*. Dentro de esta clase, los órdenes más abundantes son *Oceanospirillales* (0,0-4,3 %) y *Chromatiales* (0,3-0,6 %). El género dominante en *Oceanospirillales* es *Halomonas* (0,0-2,3 %), mientras que los géneros dominantes en *Chromatiales* son *Halorhodospira* (0,0-0,4 %) y *Aquisalimonas* (0,1-0,2 %).

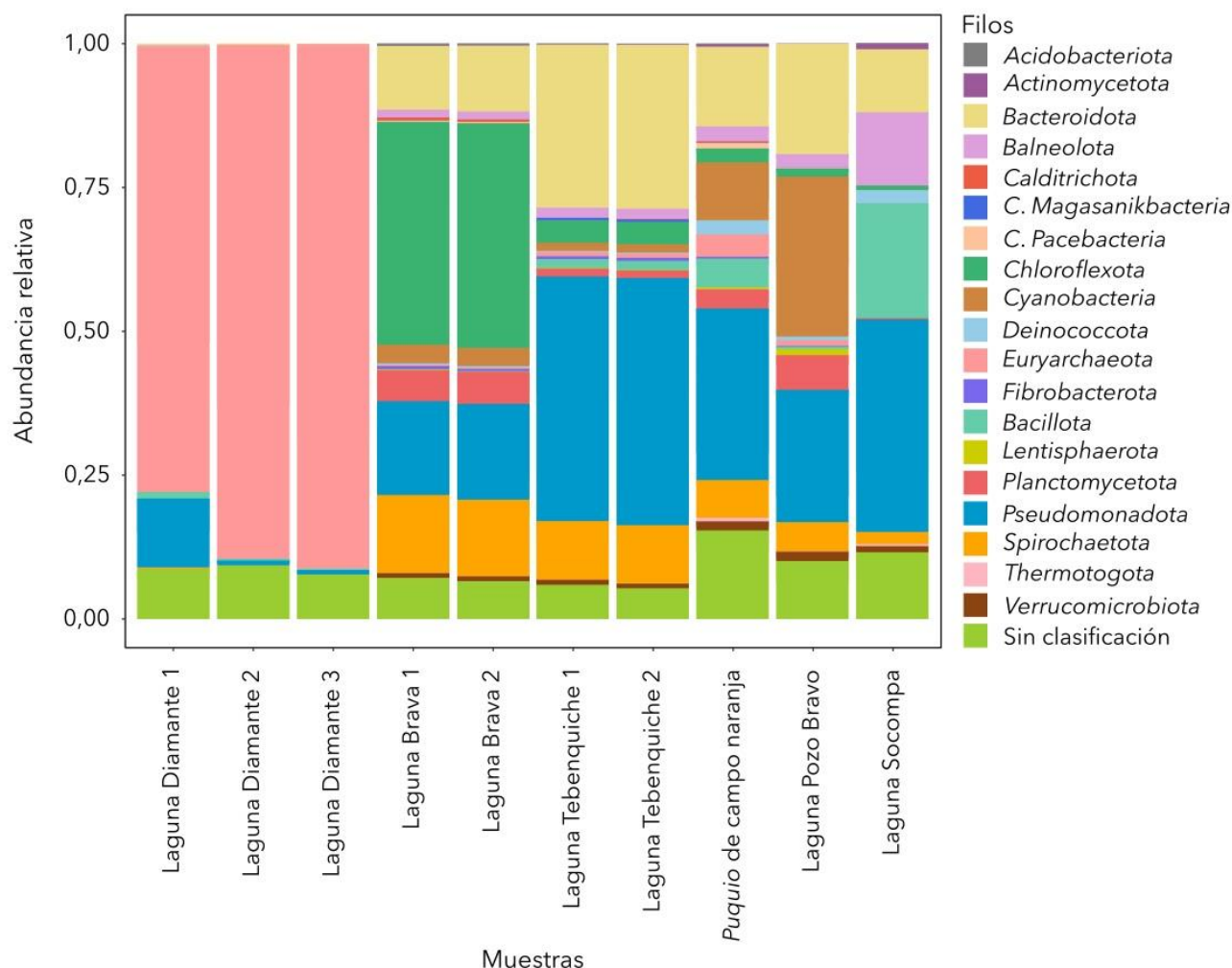


Figura 10. Composición taxonómica procariota, a nivel de filo, de los ecosistemas microbianos andinos estudiados.

El tapete microbiano de la laguna Brava está dominado por bacterias de los filos *Chloroflexota* (39,0 %), *Pseudomonadota* (16,5 %), *Spirochaetota* (13,5 %) y *Bacteroidota* (11,3 %). En el filo *Chloroflexota*, las clases más abundantes son *Anaerolineae* (25,6 %) y *Chloroflexia* (8,5 %). En el filo *Pseudomonadota*, abundan las clases *Alphaproteobacteria* (6,7 %), *Deltaproteobacteria* (5,0 %), y *Gammaproteobacteria* (4,3 %). En la primera clase dominan los órdenes *Rhodobacterales* (2,6%), *Rhizobiales* (1,8 %), y *Rhodospirillales* (0,4 %); en la segunda clase los órdenes *Desulfobacterales* (2,8 %) y *Myxococcales* (1,7 %); y en la tercera clase los órdenes

Cellvibrionales (0,5 %), *Chromatiales* (0,2 %), y *Alteromonadales* (0,1 %). Los géneros predominantes de estos órdenes son *Hyphomonas* (0,2 %), *Dichotomicrobium* (1,6 %), *Rhodovibrio* (0,05 %), *Desulfatiglans* (0,5 %), *Sandaracinus* (0,08 %), *Pseudohalaea* (0,3 %), *Wenzhouxiangella* (0,1 %), e *Idiomarina* (0,08 %), respectivamente. El filo *Spirochaetota* está compuesto principalmente por miembros de la clase *Spirochaetia*, orden *Spirochaetales* (12,7 %), y dentro de este orden, el género más abundante es *Sediminispirochaeta* (1,6 %). El filo *Bacteroidota* se encuentra dominado por las clases *Bacteroidia* (5,1 %) y *Saprospira* (2,3 %), y su género más abundante es *Phaeodactylibacter* (1,8 %). En menor proporción, el tapete microbiano de la laguna Brava presenta bacterias de los filos *Planctomycetota* (5,5 %), *Cyanobacteria* (3,2 %), y *Balneolota* (1,4 %).

El tapete microbiano de la laguna Tebenquiche está compuesto principalmente por bacterias de los filos *Pseudomonadota* (42,7 %) y *Bacteroidota* (28,4 %). Dentro del filo *Pseudomonadota*, las clases más abundantes son *Alphaproteobacteria* (28,5 %), *Deltaproteobacteria* (8,1 %), y *Gammaproteobacteria* (4,9 %). En la primera clase, dominan los órdenes *Rhodospirillales* (3,5 %), *Rhodobacterales* (2,4 %), y *Rhizobiales* (0,6 %), en la segunda clase los órdenes *Desulfobacterales* (5,7 %), *Desulfovibrionales* (0,7 %), y *Myxococcales* (1,1 %), y en la tercera clase los órdenes *Chromatiales* (0,8 %) y *Cellvibrionales* (0,5 %). Los géneros predominantes de estos órdenes son *Rhodovibrio* (1,5 %), *Roseovarius* (0,1 %), *Dichotomicrobium* (0,5 %), *Desulfatiglans* (0,4 %), *Desulfovermiculus* (0,4 %), *Phaselicystis* (0,05 %), *Wenzhouxiangella* (0,2 %), y *Pseudohalaea* (0,2 %). En el filo *Bacteroidota*, las clases más abundantes son *Bacteroidia* (7,9 %) y *Saprospira* (1,7 %), y dentro de estas clases, los géneros más abundantes son *Anaerophaga* (1,1 %) y *Phaeodactylibacter* (1,5 %), respectivamente. De todos modos, el género más abundante dentro del filo es *Salinibacter* (9,6 %). En menor proporción, el tapete microbiano de la laguna Tebenquiche

presenta bacterias de los filos *Spirochaetota* (10,1 %), *Chloroflexota* (3,4 %), *Balneolota* (1,8 %), *Bacillota* (1,5 %), *Planctomycetota* (1,4 %), *Cyanobacteria* (1,4 %), y *Verrucomicrobiota* (1,0 %). Las clases dominantes de estos filos son *Spirochaetia* (10,0 %), *Anaerolineae* (3,7 %), *Balneolia* (1,8 %), *Clostridia* (1,5 %), *Phycisphaerae* (0,6 %), *Oscillatoriothymiceae* (0,4 %), y *Opiritutae* (0,7 %), respectivamente.

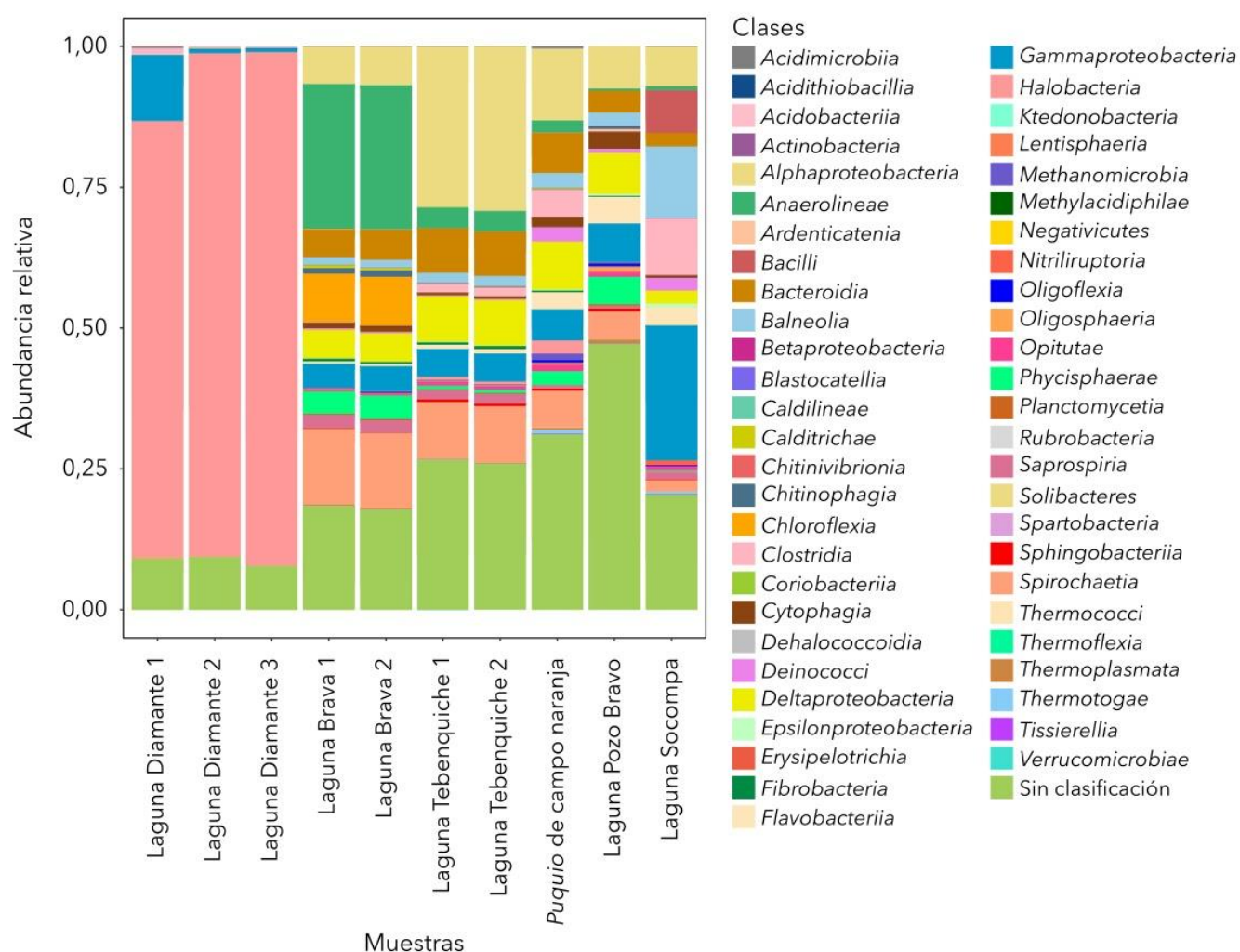


Figura 11. Composición taxonómica procariota, a nivel de clase, de los ecosistemas microbianos andinos estudiados.

El tapete microbiano del *puquio* de campo naranja está dominado por bacterias de los filos *Pseudomonadota* (29,8 %), *Bacteroidota* (13,9 %) y *Cyanobacteria* (10,1 %). Dentro del filo *Pseudomonadota*, las clases más

abundantes son *Alphaproteobacteria* (12,8 %), *Deltaproteobacteria* (8,5 %), y *Gammaproteobacteria* (5,6 %). La clase *Alphaproteobacteria* está compuesta principalmente por la orden *Rhodospirillales* (6,0 %), cuyo género más abundante es *Rhodovibrio* (4,7 %). La clase *Deltaproteobacteria* presenta dos órdenes dominantes, *Desulfobacterales* (4,4 %) y *Desulfovibrionales* (3,1 %). Los géneros en mayor proporción dentro de estos órdenes son *Desulfotignum* (0,4 %) y *Desulfovermiculus* (1,9 %), respectivamente. Mientras que la clase *Gammaproteobacteria* está compuesta principalmente por la orden *Chromatiales* (2,5 %), cuyo género más abundante es *Halochromatium* (1,0 %). El filo *Bacteroidota* está dominado por las clases *Bacteroidia* (7,1 %) y *Flavobacteriia* (3,0 %), dentro de las cuales abundan los géneros *Tangfeifania* (0,5 %) y *Psychroflexus* (1,5 %), respectivamente. En el filo *Cyanobacteria* abunda el orden *Oscillatoriales* (6,2 %), donde predomina el género *Coleofasciculus* (3,7 %). En menor proporción, el tapete microbiano del *puquio* de campo naranja presenta bacterias de los filos *Spirochaetota* (6,5 %), *Bacillota* (5,0 %), *Euryarchaeota* (3,9 %), *Planctomycetota* (3,3 %), *Balneolota* (2,6 %), *Deinococcota* (2,5 %), *Chloroflexota* (2,4 %), y *Verrucomicrobiota* (1,7 %). Las clases dominantes de estos filos son *Spirochaetia* (6,5 %), *Clostridia* (4,8 %), *Halobacteria* (2,3 %), *Phycisphaerae* (2,4 %), *Balneolia* (2,6 %), *Deinococci* (2,5 %), *Anaerolineae* (2,1 %), y *Opitutae* (1,2 %), respectivamente.

El tapete microbiano de la laguna Pozo Bravo está dominado por bacterias de los filos *Cyanobacteria* (27,8 %), *Pseudomonadota* (23,0 %), y *Bacteroidota* (19,1 %). La mayoría de las cianobacterias pertenecen a la orden *Nostocales* (12,4 %), aunque también se encuentran miembros de las órdenes *Oscillatoriales* (3,3 %) y *Chroococcales* (1,3 %). En el filo *Pseudomonadota* dominan las clases *Alphaproteobacteria* (7,5 %), *Deltaproteobacteria* (7,3 %), y *Gammaproteobacteria* (6,9 %). La clase *Alphaproteobacteria* está representada principalmente por los órdenes *Rhodobacterales* (2,7 %) y

Rhodospirillales (1,5 %), la clase *Deltaproteobacteria* por los órdenes *Desulfobacterales* (5,3 %) y *Desulfovibrionales* (0,7 %), y la clase *Gammaproteobacteria* por el orden *Chromatiales* (1,1 %). Los géneros predominantes dentro de estos órdenes son *Roseovarius* (0,3 %), *Rhodovibrio* (0,6 %), *Desulfopila* (0,4 %), *Desulfovermiculus* (0,5 %), y *Thioalkalivibrio* (0,4 %), respectivamente. En el filo *Bacteroidota*, las clases más abundantes son *Flavobacteriia* (4,8 %), *Bacteroidia* (3,9 %), y *Cytophagia* (3,0 %). En menor proporción, el tapete microbiano de la laguna Pozo Bravo presenta bacterias de los filos *Planctomycetota* (6,0 %), *Spirochaetota* (5,0 %), *Balneolota* (2,3 %), *Verrucomicrobiota* (1,7 %), *Chloroflexota* (1,4 %), y *Lentisphaerota* (1,1 %). Dichos filos están representados principalmente por las clases *Phycisphaerae* (4,9 %), *Spirochaetia* (4,9 %), *Balneolia* (2,3 %), *Opitutae* (0,9 %), *Anaerolineae* (0,4 %), y *Oligosphaeria* (1 %), respectivamente.

Por último, el estromatolito de la laguna Socompa está dominado por bacterias de los filos *Pseudomonadota* (36,9 %), *Bacillota* (19,9 %), *Balneolota* (12,7 %), y *Bacteroidota* (10,9 %). En el filo *Pseudomonadota*, abundan las clases *Gammaproteobacteria* (23,9 %) y *Alphaproteobacteria* (7,0 %). La clase *Gammaproteobacteria* está compuesta principalmente por los órdenes *Alteromonadales* (7,6 %) y *Oceanospirillales* (4,4 %), en los cuales predominan los géneros *Idiomarina* (5,2 %) y *Halomonas* (2,6 %), respectivamente. Mientras que la clase *Alphaproteobacteria* está constituida principalmente por el orden *Rhodobacterales* (5,8 %), en el cual predomina el género *Roseovarius* (1,5 %). En el filo *Bacillota*, abundan las clases *Clostridia* (10,1 %) y *Bacilli* (7,6 %). La primera clase está compuesta mayoritariamente por el orden *Clostridiales* (8,7 %), en el cual predomina el género *Fusibacter* (4,5 %). Mientras que la segunda clase está constituida principalmente por el orden *Lactobacillales* (4,2 %), en el cual predomina el género *Alkalibacterium* (2,0 %). El filo *Balneolota* está compuesto en su totalidad por la clase

Balneolia, orden *Balneolales* (12,7 %), y los géneros predominantes son *Rhodohalobacter* (0,8 %) y *Gracilimonas* (0,5 %). El filo *Bacteroidota*, por otro lado, está dominado por la clase *Flavobacteriia* (3,2 %). Esta clase está constituida en su totalidad por el orden *Flavobacteriales*, en el cual predomina el género *Psychroflexus* (1,9 %). En menor proporción, el estromatolito presenta bacterias de los filos *Deinococcota* (2,3 %), *Spirochaetota* (2,1 %), y *Verrucomicrobiota* (1,2 %). Las clases más abundantes de dichos filos son *Deinococci* (2,3 %), *Spirochaetia* (2,1 %), y *Opitutae* (0,8 %), respectivamente.

Al comparar la composición taxonómica de los diferentes ecosistemas microbianos andinos, se pudo determinar que la comunidad microbiana del *biofilm* rojo de la laguna Diamante es completamente diferente a las comunidades del resto de los ecosistemas microbianos estudiados. El *biofilm* rojo de Diamante está dominado principalmente por haloarqueas, mientras que los ecosistemas de Brava, Tebenquiche, *puquio* de campo naranja, Pozo Bravo, y Socompa están dominados por bacterias de diferentes gremios. Como se mencionó anteriormente, la altitud, el pH, la concentración de sales y el nivel de arsénico son considerablemente mayores en la laguna Diamante que en el resto de las lagunas estudiadas. Estas condiciones pueden ser toleradas por microorganismos poliextremófilos como las haloarqueas (Soppa, 2006), pero no por la mayoría de las bacterias, lo que explicaría la diferencia observada.

Por otro lado, el análisis de la composición taxonómica de los diferentes ecosistemas microbianos andinos permitió determinar los gremios de microorganismos presentes en cada ecosistema. Los organismos fotoautótrofos oxigénicos (filo *Cyanobacteria*) fueron identificados en todos los ecosistemas, excepto en el *biofilm* rojo de la laguna Diamante. Dichos organismos obtienen energía a través de la fotosíntesis oxigénica y fijan el dióxido de carbono atmosférico a través del ciclo reductivo de las pentosas

(ciclo de Calvin–Benson–Bassham). También suelen ser responsables de la fijación del nitrógeno atmosférico (Whitton & Potts, 2012). Si bien abundan en los tapetes microbianos del *puquio* de campo naranja y la laguna Pozo Bravo, en el resto de los ecosistemas se encuentran en baja proporción. La baja abundancia de organismos fotoautótrofos oxigénicos también fue reportada en otros ecosistemas microbianos estudiados de los Andes centrales, lo que apoya la idea de que no son los principales productores primarios en estos ecosistemas (Vignale, Lencina, et al., 2021). Los principales productores primarios en los ecosistemas microbianos andinos serían los organismos fotoautótrofos anoxigénicos y los organismos quimiolitotótrofos, los cuales abundan en los ecosistemas de Brava, Tebenquiche, *puquio* de campo naranja, Pozo Bravo y Socompa. Entre los organismos fotoautótrofos anoxigénicos se encuentran las bacterias púrpura del azufre (géneros *Halochromatium*, *Halorhodospira*, *Thiohalocapsa*) y las bacterias verdes no del azufre (clase *Chloroflexia*). Dichos organismos obtienen energía a través de la fotosíntesis anoxigénica, y fijan el dióxido de carbono atmosférico a través del ciclo reductivo de las pentosas y el ciclo del 3-hidroxipropionato, respectivamente (Bar-Even et al., 2012; Berg, 2011). Entre los organismos quimiolitotótrofos se encuentran las bacterias sulfato reductoras (géneros *Desulfotignum*, *Desulfovermiculus*, *Desulfococcus*, *Desulfobacterium*) y las bacterias oxidadoras de azufre (géneros *Thioalkalivibrio*, *Thiohalospira*). Estos organismos obtienen energía a través de la oxidación de compuestos inorgánicos, y fijan el dióxido de carbono atmosférico a través de la vía reductora de la acetil-coenzima A (vía de Wood-Ljungdahl) y el ciclo reductivo de las pentosas, respectivamente (Bar-Even et al., 2012; Berg, 2011). En los ecosistemas microbianos andinos también abundan los organismos heterótrofos, que consumen los compuestos orgánicos exudados por los organismos autótrofos. Entre los organismos heterótrofos se encuentran las bacterias púrpura no del azufre (géneros *Rhodovibrio*, *Phaeovibrio*, *Roseospira*, *Rubribacterium*), las bacterias fotosintéticas anoxigénicas

aeróbicas (géneros *Roseovarius*, *Rubrimonas*), y las arqueas y bacterias quimioheterótrofas (miembros de los filos *Euryarchaeota*, *Pseudomonadota*, *Bacteroidota*, *Spirochaetota*, *Chloroflexota*, *Bacillota*, *Planctomycetota*, *Balneolota*, *Verrucomicrobiota*, *Lentisphaerota*, *Deinococcota*).

Diversidad de los ecosistemas microbianos andinos

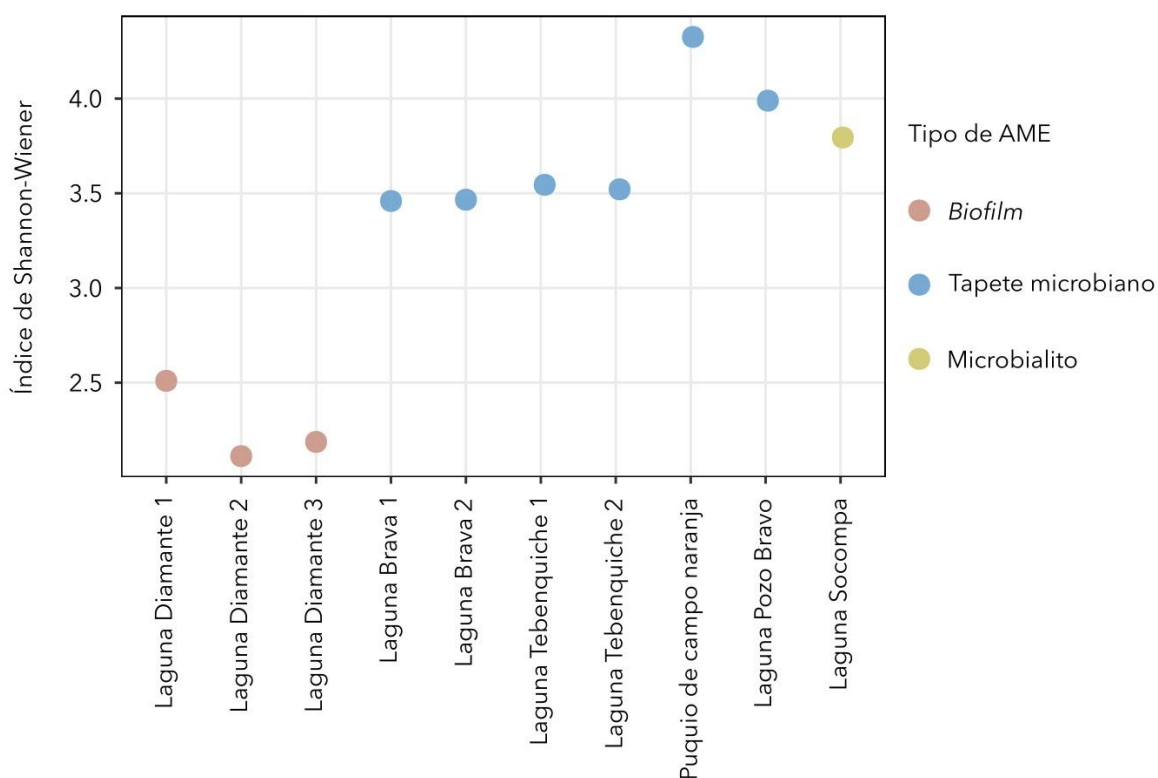


Figura 12. Diversidad alfa, medida a partir del índice de Shannon-Wiener, de los ecosistemas microbianos andinos estudiados.

El cálculo del índice de Shannon-Wiener para cada una de las muestras permitió determinar la diversidad alfa en los distintos ecosistemas microbianos estudiados. El ecosistema que presenta el valor más bajo de diversidad alfa ($H' = 2,11-2,50$) es el *biofilm* rojo de la laguna Diamante. Esto se debe a que dicho ecosistema está dominado por unas pocas especies de

haloarqueas capaces de resistir las condiciones extremas de la laguna Diamante; condiciones que exceden las capacidades metabólicas de la mayoría de las especies bacterianas. El resto de los ecosistemas microbianos estudiados presentan valores relativamente altos ($H' = 3,46-4,33$) de diversidad alfa (Baliton et al., 2020). Aun así, no alcanzan los valores que normalmente se reportan para los ecosistemas microbianos de hábitats no extremos ($H' > 4$), como los suelos forestales o agrícolas (Jiao et al., 2019; Wei et al., 2018). Si bien los tapetes microbianos y microbialitos presentan diferentes nichos ecológicos (zonas óxicas, anóxicas, fóticas, afóticas, de alta radiación UV-B, de baja radiación UV-B, alcalinas, ácidas, etc.) en los cuales pueden desarrollarse diversos gremios de microorganismos, la presión de selección ejercida por las condiciones ambientales de los Andes centrales limita considerablemente el número de especies (Farías, 2020).

Ensamblado y anotación de los metagenomas

A partir del control de calidad de las lecturas metagenómicas, se pudieron determinar parámetros como la calidad por base, la calidad promedio, el contenido de bases ambiguas, el contenido de secuencias adaptadoras, y el contenido de GC. La calidad de las lecturas metagenómicas disminuye en las últimas bases, como es habitual en la mayoría de las lecturas obtenidas a partir de plataformas de secuenciación Illumina (Tan et al., 2019). En las lecturas *forward* de Diamante (~300 pb) y Socompa (~120 pb), la calidad cae por debajo de Q20 aproximadamente en las posiciones 230 y 80, respectivamente. Mientras que, en las lecturas *reverse* de Diamante, Brava (~150 pb), Tebenquiche (~150 pb), *puquio* de campo naranja (~250 pb), Pozo Bravo (~250 pb), y Socompa, la calidad cae por debajo de Q20 alrededor de las posiciones 180, 100, 100, 210, 210 y 70, respectivamente. Por este motivo, en el preprocesamiento de las lecturas metagenómicas se removieron

los extremos con calidad promedio (cada cuatro bases) menor a Q20. La mayoría de las lecturas de Brava y Tebenquiche presentan una calidad promedio superior a Q30, mientras que la mayor parte de las lecturas de Diamante, *puquio* de campo naranja, y Pozo Bravo tienen una calidad promedio superior a Q20. Si bien la calidad promedio de las lecturas de Socompa es generalmente superior a Q20, hay un gran número de lecturas con valores inferiores a Q20. Teniendo en cuenta esto, en el preprocesamiento de las lecturas metagenómicas se removieron lecturas con calidad promedio menor a Q28 en Brava y Tebenquiche; menor a Q20 en Diamante, *puquio* de campo naranja, y Pozo Bravo; y menor a Q15 en Socompa. La mayor parte de las lecturas metagenómicas no contienen bases ambiguas ni secuencias adaptadoras, aún así, la remoción de secuencias adaptadoras durante el preprocesamiento se llevó a cabo en todas ellas. Por último, resulta interesante mencionar que el contenido GC de las lecturas varía entre 51 y 62 % (Tablas 5, 6 y 7). El contenido GC de los genomas procariotas puede variar de 17 a 75 % (Brocchieri, 2014), por ende, se puede decir que los porcentajes que presentan las lecturas de los ecosistemas microbianos andinos son relativamente elevados. Esto podría estar relacionado parcialmente con la alta radiación UV-B a la que están expuestos los microorganismos andinos, ya que un alto contenido de GC promueve la reparación del ADN dañado por la radiación (Brocchieri, 2014).

Luego de realizar el control de calidad y preprocesamiento de las lecturas metagenómicas, las mismas fueron ensambladas *de novo* para obtener *contigs*. Con el fin de obtener el mejor ensamblado de las lecturas posible, se evaluaron tres *software* de ensamblado diferentes: SPAdes, MEGAHIT (D. Li et al., 2015) e IDBA-UD (Peng et al., 2012). Para comparar la eficacia de dichos *software*, se realizó un control de calidad de los ensamblados obtenidos en el que se determinaron parámetros como el largo total, el número de *contigs*, el largo máximo de *contig*, el N50, el porcentaje de mapeo

de las lecturas, y la eficiencia de ensamblado. Los ensamblados obtenidos a partir de la herramienta SPAdes presentaron los valores más altos de largo total, largo máximo de *contig*, N50, porcentaje de mapeo de las lecturas, y eficiencia de ensamblado. Por tal motivo, la herramienta SPAdes fue la utilizada para ensamblar todos los metagenomas de los ecosistemas microbianos andinos, como se especificó en la sección materiales y métodos. Resulta interesante además mencionar que el largo total, el número de *contigs*, y el largo máximo de *contig* presentaron cierta relación con el número de lecturas (Tablas 5, 6 y 7). Los metagenomas de Diamante, que tienen el menor número de lecturas (alrededor de 5 y 6 millones), presentaron los valores más bajos de largo total, número de *contigs*, y largo máximo de *contig*. Mientras que los metagenomas de Brava y Tebenquiche, que tienen el mayor número de lecturas (alrededor de 100 y 140 millones), presentaron los valores más altos de estos parámetros. Por el contrario, el N50 y el porcentaje de mapeo de las lecturas no presentaron relación con el número de lecturas ni con el largo de las mismas. Como el porcentaje de mapeo de las lecturas fue mayor a 60 % en todos los ensamblados, y cercano a 80 % en la mayoría, se puede decir que la mayor parte de las lecturas metagenómicas fueron ensambladas con éxito.

Tras obtener los metagenomas ensamblados, se llevó a cabo la anotación estructural de los mismos. El número de marcos de lectura abiertos (ORFs, *open reading frames*) identificados en los distintos metagenomas ensamblados varió de acuerdo al largo total de los mismos (Tablas 5, 6 y 7). En los metagenomas de Diamante, que presentan el menor largo total, se identificaron alrededor de 200.000 ORFs. En los metagenomas de *puquio* de campo naranja, Pozo Bravo, y Socompa, que presentan un largo total intermedio, se identificaron más de 1.000.000 de ORFs. Mientras que en los metagenomas de Brava y Tebenquiche, que presentan el mayor largo total, se identificaron cerca de 4.000.000 de ORFs.

Tabla 5. Estadísticas de los metagenomas de la laguna Diamante.

Estadísticas	Laguna Diamante 1	Laguna Diamante 2	Laguna Diamante 3
Secuenciación masiva			
Plataforma	Illumina MiSeq	Illumina MiSeq	Illumina MiSeq
Tipo de lecturas	<i>paired end</i>	<i>paired end</i>	<i>paired end</i>
N° de acceso	ERR3132462	ERR3132463	ERR3132464
N° de lecturas	2.440.421 (x2)	3.054.366 (x2)	2.750.479 (x2)
% GC de lecturas	61	61	62
Largo de lecturas (pb)	35-301	35-301	35-301
Anotación de las lecturas metagenómicas			
N° de CDSs identificadas	3.388.415	4.140.049	3.744.850
N° de CDSs anotadas funcionalmente	1.582.828	1.920.100	1.722.020
N° de OTUs procariotas	61	35	33
Ensamblado de las lecturas metagenómicas			
Largo total (Mb) (<i>contigs</i> ≥ 500 pb)	85,06	65,4	67,6
N° de <i>contigs</i>	79.376	57.038	60.299
Largo máximo de <i>contig</i> (kb)	73,2	50,2	44,8
N50 (pb)	1.138	1.297	1.262
% de mapeo de las lecturas	78,22	82,32	80,09
Eficiencia de ensamblado	89.014,36	106.769,04	101.073,58
Anotación de los <i>contigs</i>			
N° de ORFs	268.252	204.997	221.681

CDSs (*coding sequences*), Secuencias codificantes; ORFs (*open reading frames*), Marcos de lectura abiertos.

Tabla 6. Estadísticas de los metagenomas de las lagunas Brava y Tebenquiche.

Estadísticas	Laguna Brava 1	Laguna Brava 2	Laguna Tebenquiche 1	Laguna Tebenquiche 2
Secuenciación masiva				
Plataforma	Illumina HiSeq 1500	Illumina HiSeq 1500	Illumina HiSeq 1500	Illumina HiSeq 1500
Tipo de lecturas	<i>paired end</i>	<i>paired end</i>	<i>paired end</i>	<i>paired end</i>
N° de acceso	ERR2407577	ERR2407578	ERR2437276	ERR2437277
N° de lecturas	52.248.302 (x2)	53.164.322 (x2)	71.152.673 (x2)	72.289.598 (x2)
% GC de lecturas	56	56	60	60
Largo de lecturas (pb)	151	151	151	151
Anotación de las lecturas metagenómicas				
N° de CDSs identificadas	68.797.099	63.335.009	85.716.863	77.688.595
N° de CDSs anotadas funcionalmente	24.807.311	25.713.862	33.690.575	34.210.280
N° de OTUs procariotas	443	419	547	492
Ensamblado de las lecturas metagenómicas				
Largo total (Mb) (<i>contigs</i> ≥ 500 pb)	831,8	887,0	654,7	688,7
N° de <i>contigs</i>	582.037	623.888	442.433	466.215
Largo máximo de <i>contig</i> (kb)	241,3	287,0	298,4	268,4
N50 (pb)	1.869	1.839	2.007	2.005
% de mapeo de las lecturas	78,41	87,85	74,30	83,12
Eficiencia de ensamblado	146.548,29	161.556,15	149.120,1	166.655,6
Anotación de los <i>contigs</i>				
N° de ORFs	3.935.685	4.325.317	3.564.011	3.905.801

CDSs (*coding sequences*), Secuencias codificantes; ORFs (*open reading frames*), Marcos de lectura abiertos.

Tabla 7. Estadísticas de los metagenomas del *puquio* de campo naranja y las lagunas Pozo Bravo y Socompa.

Estadísticas	<i>Puquio</i> de campo naranja	Laguna Pozo Bravo	Laguna Socompa
Secuenciación masiva			
Plataforma	Illumina MiSeq	Illumina MiSeq	Illumina Genome Analyzer (GAIIx)
Tipo de lecturas	<i>paired end</i>	<i>paired end</i>	<i>paired end</i>
N° de acceso	ERR3083899	ERR3082345	SRR3341855
N° de lecturas	7.798.852 (x2)	8.232.105 (x2)	47.304.781 (x2)
% GC de lecturas	51	52	58
Largo de lecturas (pb)	35-251	35-251	120
Anotación de las lecturas metagenómicas			
N° de CDSs identificadas	7.456327	8.367.672	35.249.197
N° de CDSs anotadas funcionalmente	3.364.377	3.741.958	11.826.628
N° de OTUs procariotas	333	286	451
Ensamblado de las lecturas metagenómicas			
Largo total (Mb) (<i>contigs</i> ≥ 500 pb)	224,8	233,2	280,3
N° de <i>contigs</i>	243.817	237.161	210.218
Largo máximo de <i>contig</i> (kb)	54,9	83,2	217,1
N50 (pb)	890	955	1.613
% de mapeo de las lecturas	63,25	67,78	61,83
Eficiencia de ensamblado	56.292,5	64.729,9	99.731,79
Anotación de los <i>contigs</i>			
N° de ORFs	1.079.965	1.072.274	1.371.377

CDSs (*coding sequences*), Secuencias codificantes; ORFs (*open reading frames*), Marcos de lectura abiertos.

Reconstrucción de genomas

Tabla 8. Número de genomas reconstruidos a partir de los metagenomas.

	N° de MAGs	N° de MAGs luego de la 1° desreplicación	N° de MAGs luego de la 2° desreplicación
Laguna Diamante 1	3	3	5
Laguna Diamante 2	3	2	
Laguna Diamante 3	4	3	
Laguna Brava 1	80	76	83
Laguna Brava 2	86	81	
Laguna Tebenquiche 1	59	52	57
Laguna Tebenquiche 2	69	56	
<i>Puquio</i> de campo naranja	13	8	8
Laguna Pozo Bravo	13	13	13
Laguna Socompa	27	23	23
Total	357	317	189

A partir de los diferentes metagenomas ensamblados, se pudieron reconstruir 357 genomas (MAGs) en total (Tabla 8). Luego de realizar un control de calidad de dichos MAGs, se pudo determinar que los mismos presentan altos porcentajes de completitud (mayores a 50 %) y bajos porcentajes de contaminación (menores a 5 %) (Tabla A1 en Apéndice). Por tal motivo, todos los MAGs reconstruidos tienen una calidad media o alta (Figura 13). Con el fin de eliminar la redundancia de los MAGs provenientes de un mismo metagenoma, posteriormente se realizó una desreplicación a nivel de especie entre los MAGs reconstruidos a partir de un mismo metagenoma (primera desreplicación). Como consecuencia, el número total de MAGs se redujo a 317 (Tabla 8).

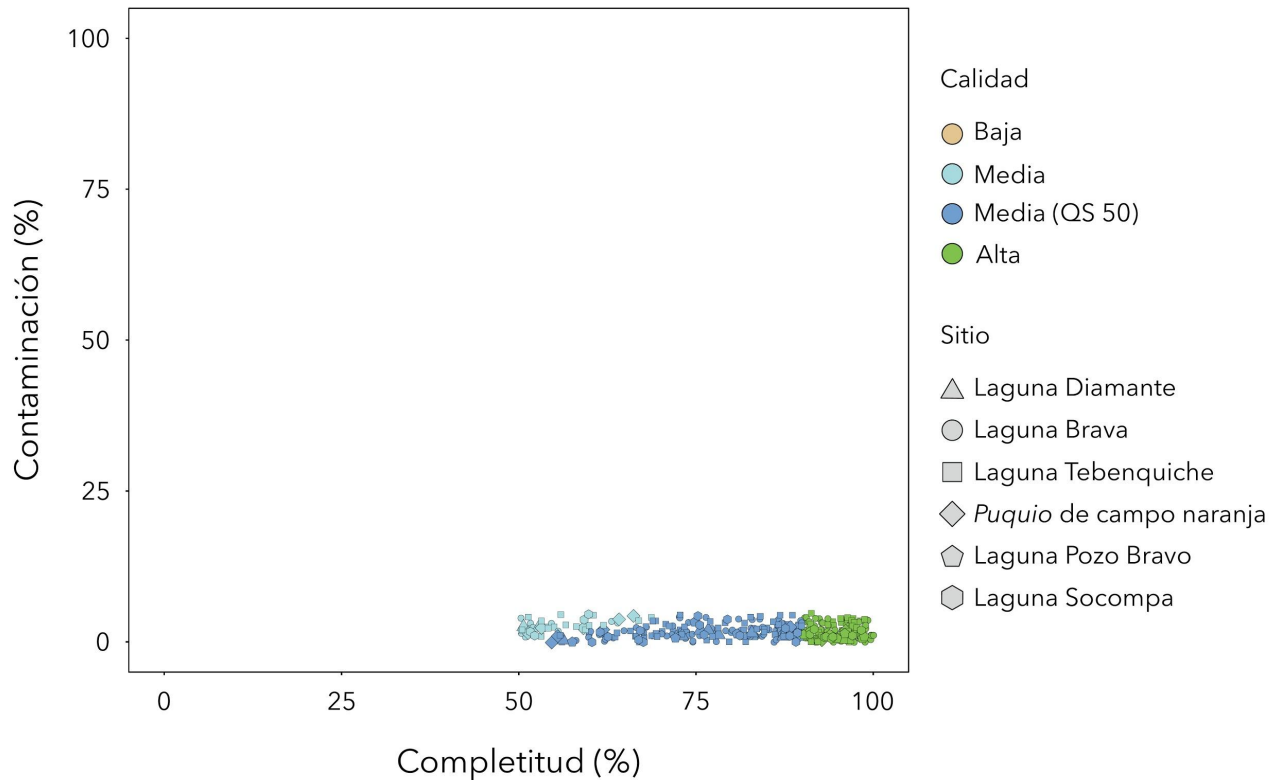


Figura 13. Control de calidad de los MAGs reconstruidos, basado en los porcentajes de completitud y contaminación.

Luego se procedió a eliminar la redundancia de los MAGs provenientes de una misma laguna. Para ello, se realizó una desreplicación a nivel de especie entre los MAGs pertenecientes a una misma laguna (segunda desreplicación). Como resultado, el número total de MAGs se redujo a 189 (Tabla 8). Por último, con el fin de simplificar la anotación taxonómica y funcional, se procedió a filtrar los MAGs redundantes entre lagunas. Para ello, se llevó a cabo una desreplicación a nivel de especie entre los 189 MAGs remanentes (tercera desreplicación). Como consecuencia, el número final de MAGs disminuyó a 152. De los 189 MAGs, 69 comparten la misma identidad con uno o varios MAGs provenientes de otras lagunas (Figura 14). De estos 69 MAGs, la mayoría (46) pertenecen a las lagunas Brava y Tebenquiche, que se encuentran a poca distancia en el salar de Atacama. Los 120 MAGs

restantes, de los 189 MAGs originales, son únicos de cada laguna: 5 pertenecen a la laguna Diamante, 56 a la laguna Brava, 29 a la laguna Tebenquiche, 5 al *puquio* de campo naranja, 9 a la laguna Pozo Bravo, y 16 a la laguna Socompa. A pesar de que todas las lagunas estudiadas presentan condiciones ambientales extremas, dichas condiciones varían entre las lagunas, incluso entre las que se encuentran próximas. Esta variabilidad en las condiciones ambientales genera una variabilidad en la composición taxonómica de los AMEs, tal como se demostró anteriormente (Figuras 10 y 11), lo cual explicaría el elevado número de MAGs únicos en cada laguna.

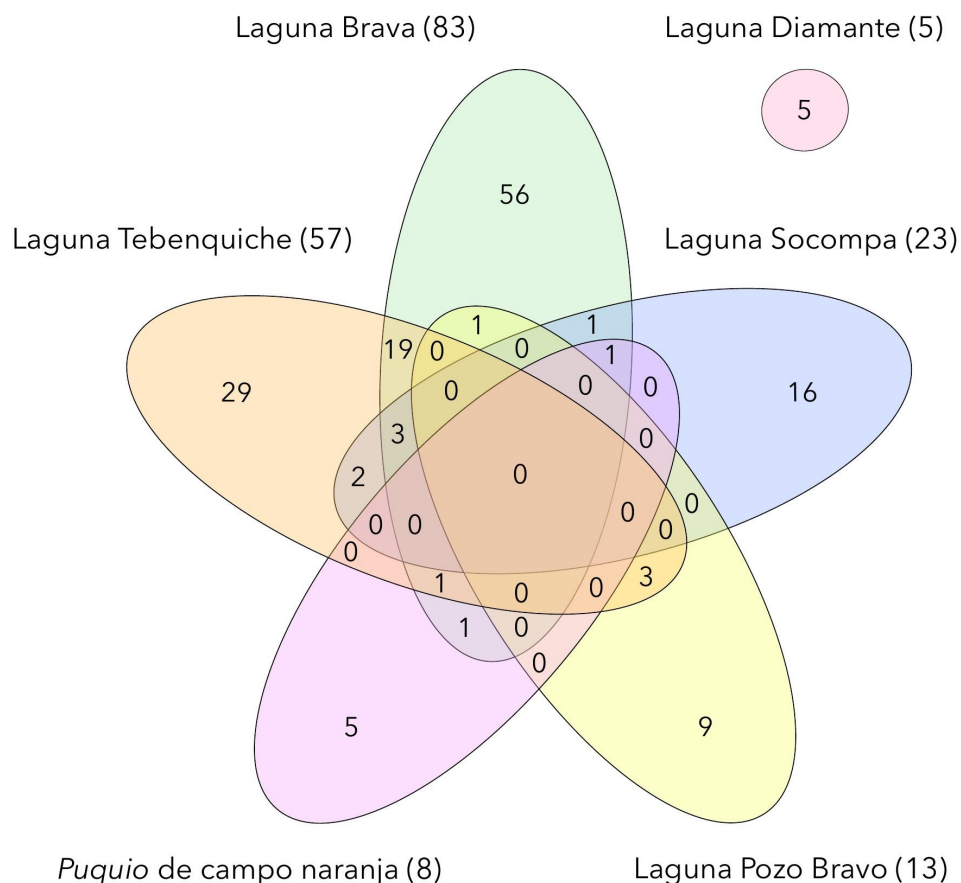


Figura 14. Los diferentes ecosistemas microbianos andinos estudiados presentan un elevado número de MAGs únicos.

La asignación taxonómica de los 152 MAGs finales (Tabla A2 en Apéndice) permitió determinar que los mismos pertenecen a los filos *Pseudomonadota* (30), *Bacteroidota* (29), *Chloroflexota* (25), *Spirochaetota* (21), *Planctomycetota* (8), *Cyanobacteria* (7), *Bacillota* (5), *Euryarchaeota* (5), *Verrucomicrobiota* (3), *Actinomycetota* (2), y *Deinococcota* (1). A pesar de que no se logró identificar el filo de 16 MAGs, se pudo determinar que 3 de ellos pertenecen al dominio *Archaea* y 13 al dominio *Bacteria*. La máxima asignación taxonómica alcanzada fue a nivel de género en 14 MAGs. Los géneros identificados fueron *Roseovarius*, *Haliea*, *Pseudoalteromonas*, *Halomonas*, y *Rhodothermus*. Debido a que ninguno de los 152 MAGs finales comparte identidad a nivel de especie con los genomas en las bases de datos, se puede decir que los mismos corresponden a genomas nunca antes caracterizados. De hecho, es muy probable que la mayoría de estos genomas pertenezcan a microorganismos no cultivables en el laboratorio. Por ende, la obtención de estos MAGs podría ayudar a comprender la “materia oscura microbiana” (*Microbial dark matter* en inglés) de los ecosistemas microbianos andinos.

Potencial metabólico de los ecosistemas microbianos andinos

El análisis funcional de los 152 MAGs finales (Figuras 15 y 16) corroboró lo hipotetizado a partir de la composición taxonómica de los ecosistemas microbianos andinos. Teniendo en cuenta las propiedades genómicas (GenProp0033: *TCA cycle*, GenProp0691: *Glycolysis*, GenProp1637: *Aerobic respiration I (cytochrome c)*, entre otras) identificadas en los MAGs D.1.1, D.2.1, D.3.2, y D.3.4, se puede decir que las haloarqueas que dominan el *biofilm* rojo de la laguna Diamante son organismos quimioorganoheterótrofos aeróbicos. El MAG D.1.3 fue el único genoma bacteriano (orden *Oceanospirillales*) reconstruido en este ecosistema microbiano, y sus

Heatmap visualization showing the presence (blue) or absence (green) of 1000 metabolic pathways across 1000 bacterial species. The species are grouped into 10 taxonomic levels on the left: Actinomycetota, Cyanobacteria, Chloroflexota, Bacillota, Pseudomonadota, Spirochaetota, Verrucomicrobiota, Bacteroidota, Planctomycetota, Euryarchaeota, Deinococcota, Archaea sin clasificación, Bacteria sin clasificación, O, Laguna Diamante, Laguna Brava, Laguna Tebenquiche, Puquio de campo naranja, Laguna Pozo Bravo, Laguna Socompa, ausencia, Si, Parcial, No, egoría, and Metabolismo Energético. The pathways are listed on the right, grouped into 10 categories: Glyoxylate shunt, TCA cycle, Pentose phosphate cycle, Nitrogen fixation, Nitrogen metabolism, Nitrogen fixation, Nitrogen metabolism, Nitrogen fixation, Nitrogen metabolism, and Nitrogen metabolism. The heatmap is color-coded: blue for presence, green for absence, and grey for unknown.

A diferencia del *biofilm* rojo de la laguna Diamante, el resto de los ecosistemas microbianos andinos está dominado por bacterias de diferentes

gremios, lo cual corrobora lo hipotetizado en base a la composición taxonómica. En primer lugar, MAGs de organismos fotoautótrofos oxigénicos, pertenecientes al filo *Cyanobacteria*, fueron identificados en los ecosistemas de Brava (MAGs B.1.79 y B.2.47), Tebenquiche (MAG T.1.35), *puquio* de campo naranja (MAG P.9) y Pozo Bravo (MAGs PB.5, PB.12 y PB.13). Las propiedades genómicas de dichos MAGs no solo indicaron que son capaces de producir oxígeno a partir de la fotosíntesis oxigénica (GenProp0660: *Photosystem I* y GenProp0661: *Photosystem II*), sino además indicaron que son capaces de fijar el nitrógeno atmosférico (GenProp0029: *Nitrogen fixation*). De todos modos, su número es bajo comparado al número de MAGs pertenecientes a organismos fotoautótrofos anoxigénicos y quimiolitotótrofos. Por ende, los organismos fotoautótrofos oxigénicos podrían no ser los principales productores primarios en los tapetes microbianos y microbialitos de los Andes centrales. Numerosos MAGs, que pertenecen a bacterias verdes no del azufre, fueron identificados en el tapete microbiano de la laguna Brava (MAGs B.2.15, B.1.34, B.1.59, B.2.50 y B.2.67). Como indican sus propiedades genómicas (GenProp0662: *Photosynthetic reaction center, alphaproteobacterial type*), estos organismos obtienen energía a través de la fotosíntesis anoxigénica. Por otro lado, diversos MAGs pertenecientes a bacterias púrpura del azufre se identificaron en los ecosistemas de Tebenquiche (MAGs T.1.39, T.2.8 y T.2.62), *puquio* de campo naranja (MAG P.12) y Pozo Bravo (MAG PB.7). Al igual que las bacterias verdes no del azufre, las bacterias púrpura del azufre obtienen energía a través de la fotosíntesis anoxigénica. Así como se identificaron MAGs de organismos fotoautótrofos anoxigénicos, también se encontraron diversos MAGs que podrían pertenecer a organismos quimiolitotótrofos. Entre ellos se destacan los MAGs de bacterias sulfato reductoras. Los mismos se identificaron en los ecosistemas de Brava (MAGs B.1.42 y B.2.3), Tebenquiche (MAGs T.2.13, T.2.40 y T.2.69), *puquio* de campo naranja (MAG P.6), Pozo Bravo (MAG PB.6) y Socompa (MAG S.23). Estos organismos

tienen la capacidad de crecer en anaerobiosis utilizando el sulfato como aceptor de electrones alternativo al oxígeno (GenProp0155: *Dissimilatory sulfate reduction*). Por último, cabe destacar que la gran mayoría de los MAGs reconstruidos de los filos *Pseudomonadota*, *Bacteroidota*, *Spirochaetota*, *Bacillota*, *Planctomycetota*, *Verrucomicrobiota* y *Deinococcota*, pertenecen a organismos heterótrofos que consumen los compuestos orgánicos exudados por los organismos autótrofos. Entre ellos, se destacan los MAGs de bacterias púrpura no del azufre (MAG T.1.45) y bacterias fotosintéticas anoxigénicas aeróbicas (MAGs S.5 y S.28), que como indican sus propiedades genómicas (GenProp0662: *Photosynthetic reaction center, alphaproteobacterial type*) llevan a cabo la fotosíntesis anoxigénica oxidando compuestos orgánicos.

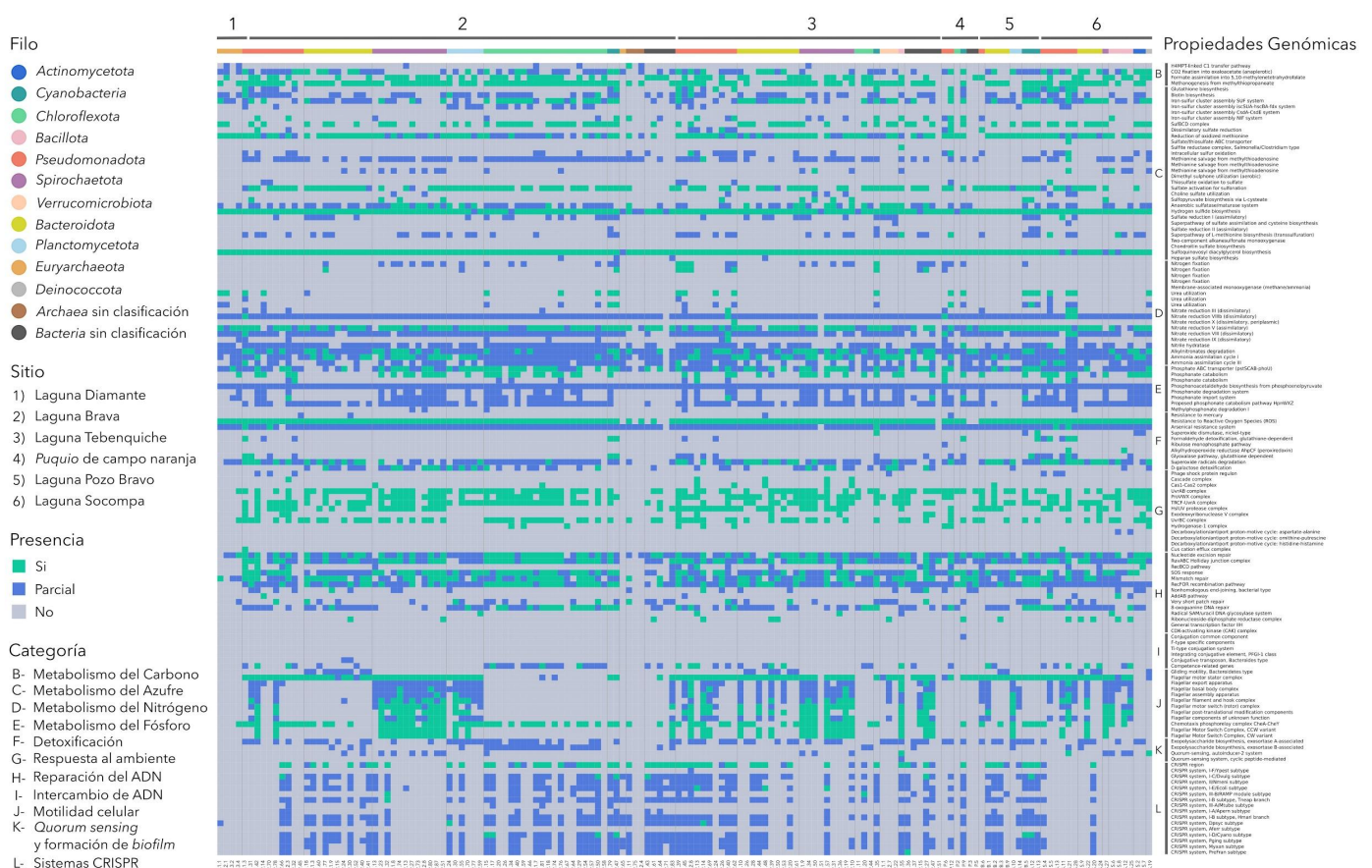


Figura 16. Propiedades genómicas (categorías B-L) de los MAGs reconstruidos a partir de los diferentes metagenomas ensamblados.

Con respecto al metabolismo del arsénico, se ha determinado que 148 de los 152 MAGs finales presentan genes *ars* de resistencia al arsénico (GenProp0474: *Arsenical resistance system*). De estos 148 MAGs, 10 a su vez contienen genes *aio* (MAGs D.1.1, D.2.1, B.2.32, T.1.45, T.2.8, T.2.14, T.2.62, P.12, S.11 y S.19). Por ende, estos organismos además de presentar mecanismos moleculares de resistencia al arsénico, tendrían la capacidad de oxidar el arsenito para obtener energía (GenProp0473: *Arsenite oxidase*). Resulta interesante que los MAGs que contienen genes *aio* no pertenecen a un único grupo taxonómico, sino a varios. Entre estos grupos se encuentran los filos *Euryarchaeota* (familia *Halobacteriaceae*), *Pseudomonadota* (órdenes *Alteromonadales*, *Chromatiales*, *Myxococcales*, y *Xanthomonadales*), y *Deinococcota* (familia *Deinococcaceae*). Teniendo en cuenta estos resultados, se puede decir que la mayoría de los microorganismos que componen los ecosistemas microbianos andinos tienen la capacidad de resistir a las concentraciones ambientales de arsénico, y que la capacidad de obtener energía a partir del arsénico podría estar presente tanto en bacterias como en arqueas de diferentes grupos taxonómicos.

Abundancia de los genes del metabolismo del arsénico

Con el fin de comprender mejor el metabolismo del arsénico en los microorganismos de los Andes centrales, se llevó a cabo un profundo análisis de todos los genes implicados en la resistencia al arsénico y obtención de energía a partir de su oxidorreducción. La identificación de dichos genes y la estimación de su abundancia (expresada como FPKM) en cada ecosistema estudiado (Figura 17), permitió determinar los mecanismos moleculares presentes en cada uno de ellos.

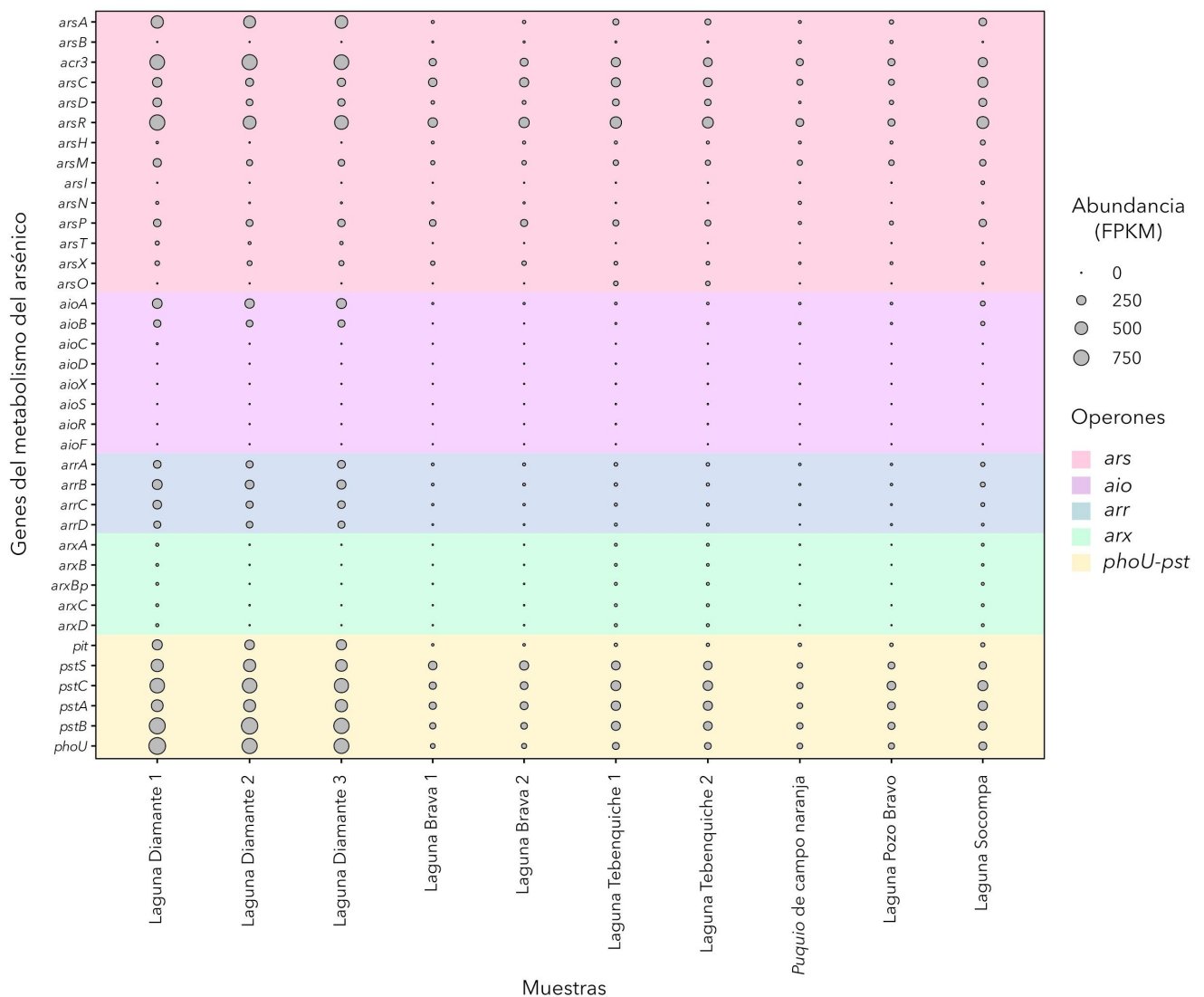


Figura 17. Abundancia de los genes del metabolismo del arsénico, expresada como FPKM, en los distintos metagenomas de los ecosistemas microbianos andinos.

En todos los ecosistemas andinos se identificaron genes *pit* y *pst* implicados en el ingreso del arsenato a las células. La abundancia de los genes *pit* fue menor que la de los genes *pst* en todos los ecosistemas. Por ende, se podría pensar que la participación de los transportadores de fosfato inorgánico de baja afinidad (Pit), en el ingreso del arsenato a las células, es menor que la de los transportadores de fosfato inorgánico de alta afinidad (Pst). El gen *arsC* también fue identificado, en similar abundancia, en todos los ecosistemas

andinos. Por lo tanto, en la mayoría de los microorganismos andinos, el ingreso del arsenato a las células estaría acompañado por su reducción a arsenito a través de la arsenato reductasa citoplasmática (ArsC). En cuanto a los genes involucrados en la extrusión del arsenito resultante, se ha observado que la abundancia del gen *acr3* es considerablemente mayor que la del gen *arsB* en todos los ecosistemas estudiados. Por consiguiente, la arsenito permeasa Acr3 sería la principal responsable de la extrusión celular del arsenito en los microorganismos andinos, y no la arsenito permeasa ArsB. Esto resulta interesante debido a que, a diferencia de lo observado, la mayoría de los microorganismos aislados de ambientes con altas concentraciones de arsénico contienen ArsB. Se desconoce la razón por la cual los microorganismos andinos presentan en mayor proporción Acr3 que ArsB, por lo que resultaría interesante investigarlo. Por otro lado, se ha observado la presencia de los genes *arsM* y *arsP*, en similar abundancia, en todos los ecosistemas estudiados. Con lo cual, alternativamente a la extrusión del arsenito, los microorganismos andinos tendrían la capacidad de metilar el arsenito a través de la enzima arsenito metiltransferasa (ArsM), para luego expulsar el metilarsenito por la bomba ArsP. Si bien se han identificado genes *arsH* en todos los ecosistemas, su abundancia es menor que la de los genes *arsM* y *arsP*, por lo que se podría pensar que la extrusión del metilarsenito por la bomba ArsP sería un mecanismo de detoxificación más utilizado que la oxidación del metilarsenito por la enzima metilarsenito oxidasa (ArsH). Como el gen *arsI* no fue identificado en muchos de los ecosistemas, y en los ecosistemas donde sí fue identificado presenta baja abundancia, se puede decir que la conversión del metilarsenito a arsenito por medio de la liasa ArsI no representa un mecanismo de detoxificación ampliamente utilizado por los microorganismos andinos. Debido a que el arsenito genera estrés oxidativo, algunos microorganismos aislados de ambientes con altas concentraciones de arsénico presentan un sistema tiorredoxina (ArsTX) que ayuda a mantener el estado redox balanceado. Ciertos genes *arsT* y *arsX* fueron identificados en

los ecosistemas estudiados, pero su abundancia no es alta comparada con la del resto de los genes del metabolismo del arsénico. Por lo tanto, si bien este mecanismo podría ser utilizado por algunos microorganismos andinos, no representaría un mecanismo de resistencia al arsénico ampliamente distribuido entre los microorganismos andinos. De igual forma, algunos genes *arsO* y *arsN* han sido identificados en los ecosistemas estudiados, pero su abundancia es baja comparada con la de otros genes del metabolismo del arsénico. Por ende, aunque no es clara la función de estos genes en el metabolismo del arsénico, lo más probable es que no constituyan mecanismos de resistencia claves para los microorganismos andinos. Por último, cabe mencionar que la gran abundancia de genes *arsR* y *arsD* observada en todos los ecosistemas estudiados indicaría que los represores transcripcionales ArsR y ArsD regulan la transcripción de varios genes *ars* en los microorganismos andinos.

Con respecto a los genes implicados en la obtención de energía a través de la oxidación del arsenito, se han identificado genes *aio* y *arx* en todos los ecosistemas estudiados, excepto en el tapete microbiano de la laguna Pozo Bravo donde no se ha podido identificar genes *arx*. Como la abundancia de los genes *aio* es mayor que la de los genes *arx*, lo más probable es que la oxidación del arsenito sea llevada a cabo principalmente por AioAB, en lugar de ArxAB, en los microorganismos andinos. Esto era esperable debido a que los genes *arx* han sido identificados únicamente en el filo *Pseudomonadota*, mientras que los genes *aio* se han encontrado en diferentes grupos de bacterias y arqueas (filos *Pseudomonadota*, *Deinococcota*, *Bacillota*, *Crenarchaeota*, y *Euryarchaeota*). A su vez, cabe mencionar que la abundancia de los genes *aio* es considerablemente mayor en el *biofilm* rojo de la laguna Diamante que en el resto de los ecosistemas microbianos andinos. Esto resulta interesante debido a que el *biofilm* rojo de la laguna Diamante está dominado por haloarqueas, sin embargo la obtención de energía a partir

de la oxidación de arsenito aún no ha sido demostrada en el dominio *Archaea* (J. Andres & Bertin, 2016). Por último, con respecto a la respiración de arsenato, se han identificado genes *arr* en todos los ecosistemas estudiados. Esto sugiere que diversos microorganismos andinos tendrían la capacidad de crecer en anaerobiosis o microaerobiosis utilizando el arsenato como aceptor de electrones alternativo al oxígeno. Cabe mencionar que, al igual que los genes *aio*, la abundancia de los genes *arr* es considerablemente mayor en el *biofilm* rojo de la laguna Diamante que en el resto de los ecosistemas microbianos andinos. Si bien se ha demostrado la capacidad de ciertas arqueas del filo *Crenarchaeota* (*Pyrobaculum arsenaticum* y *Pyrobaculum aerophilum*) de respirar arsenato en lugar de oxígeno, no se ha demostrado dicha capacidad metabólica en ninguna arquea del filo *Euryarchaeota*. Por ende, el estudio de las haloarqueas de la laguna Diamante resulta interesante para determinar efectivamente si la respiración de arsenato es un proceso metabólico presente también en los miembros del filo *Euryarchaeota*.

Organización genómica de los genes del metabolismo del arsénico

Con el fin de comprender mejor los mecanismos moleculares involucrados en la resistencia al arsénico y obtención de energía a partir de su oxidorreducción, se llevó a cabo un análisis de la organización de los genes del metabolismo del arsénico. A partir de dicho análisis, se pudo determinar que los mismos se agrupan de manera diversa en los genomas de los microorganismos andinos (Figuras 18, 19, 20, 21 y 22). Los genes *pstS*, *pstC*, *pstA* y *pstB*, involucrados en el ingreso del arsenato a las células, se encuentran agrupados normalmente con el gen *arsC*, involucrado en la reducción citoplasmática del arsenato (Figura 18). Esto sugiere que ambos procesos metabólicos están íntimamente ligados, lo cual tiene sentido, ya que conforman la primera respuesta celular a las elevadas concentraciones ambientales de arsénico.



Figura 18. Organización de los genes *ars* en los metagenomas ensamblados de los ecosistemas microbianos andinos, parte 1. L.D., Laguna Diamante; L.B., Laguna Brava; L.T., Laguna Tebenquiche; P.C.N., *Puquio* de campo naranja; L.P.B., Laguna Pozo Bravo; L.S., Laguna Socompa.

El gen *acr3*, involucrado en la extrusión celular del arsenito, también suele estar agrupado con el gen *arsC* (Figura 19). Como la reducción citoplasmática del arsenato está a su vez ligada a la extrusión del arsenito resultante, y ambos procesos metabólicos constituyen el principal mecanismo de detoxificación del arsénico en los microorganismos andinos, es de esperar que dichos genes se encuentren juntos y estén regulados transcripcionalmente de la misma manera.

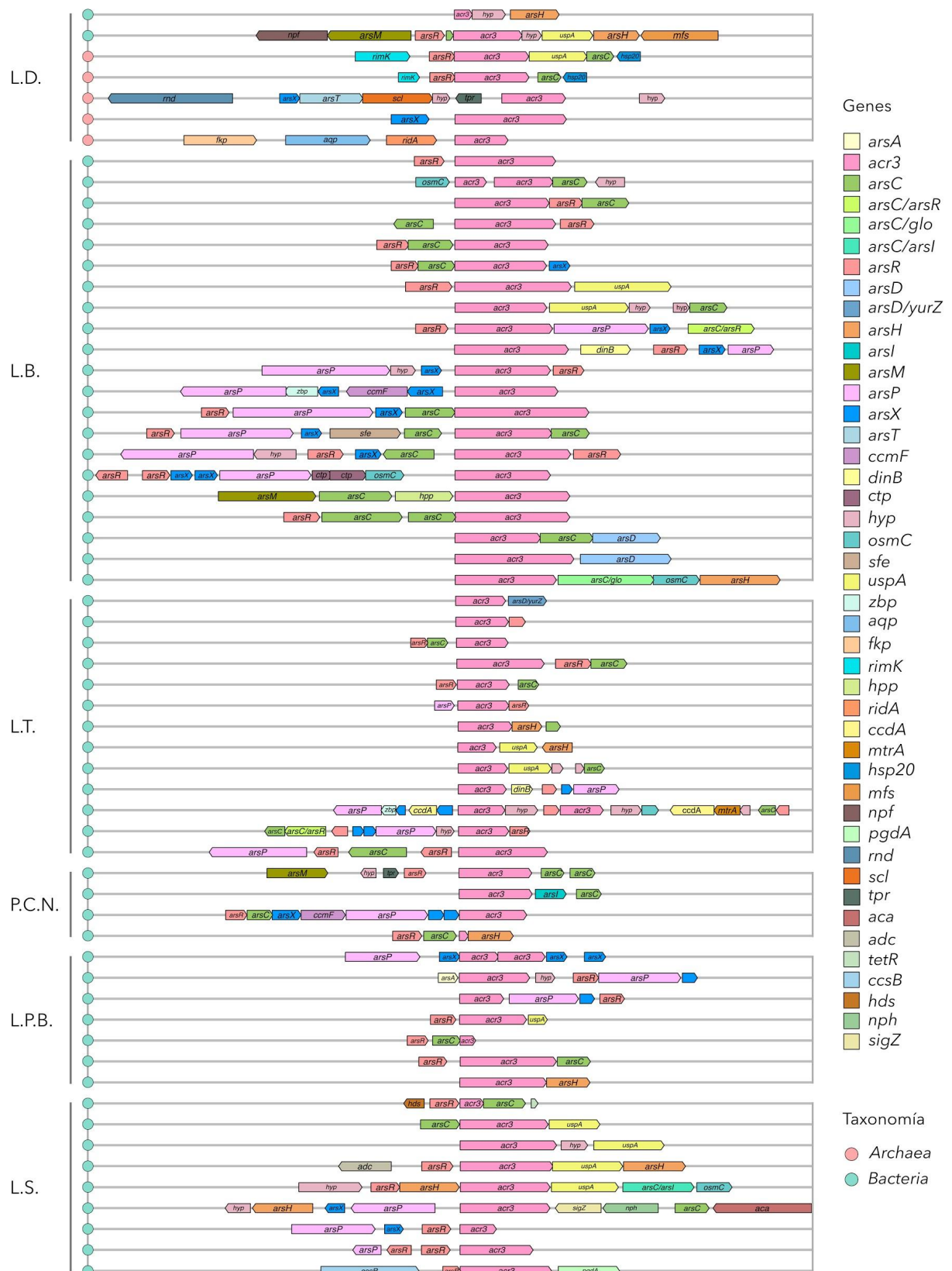


Figura 19. Organización de los genes *ars* en los metagenomas ensamblados de los ecosistemas microbianos andinos, parte 2. L.D., Laguna Diamante;

L.B., Laguna Brava; L.T., Laguna Tebenquiche; P.C.N., *Puquio* de campo naranja; L.P.B., Laguna Pozo Bravo; L.S., Laguna Socompa.

En muchas regiones donde el gen *acr3* se agrupa con el gen *arsC*, también suele estar presente el gen *arsR* (Figura 19). Esto podría indicar que el represor ArsR regula de manera conjunta la transcripción de *acr3* y *arsC*.

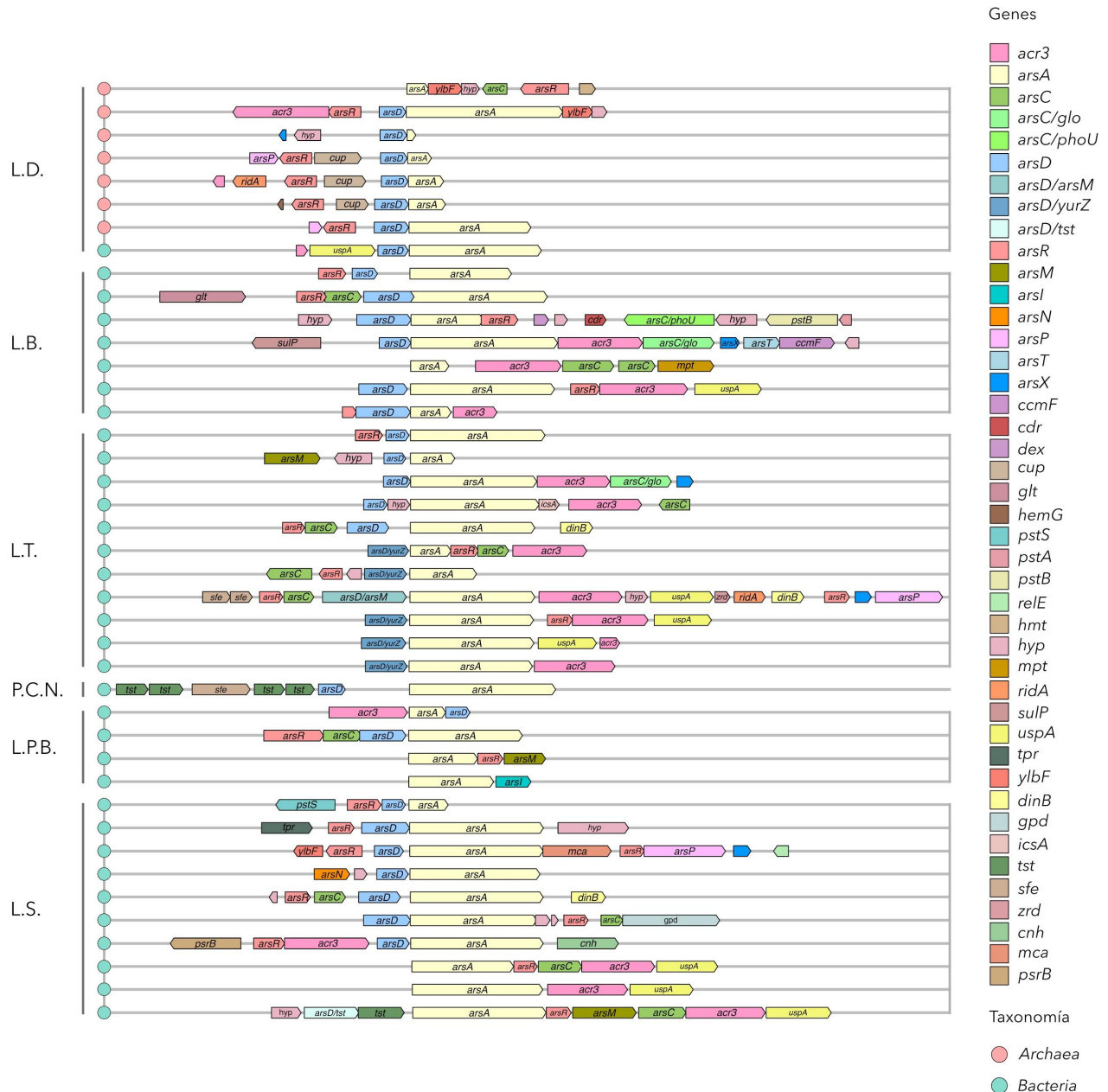


Figura 20. Organización de los genes *ars* en los metagenomas ensamblados de los ecosistemas microbianos andinos, parte 3. L.D., Laguna Diamante;

L.B., Laguna Brava; L.T., Laguna Tebenquiche; P.C.N., *Puquio* de campo naranja; L.P.B., Laguna Pozo Bravo; L.S., Laguna Socompa.

Sin embargo, en muchas otras regiones donde el gen *acr3* se agrupa, o no, con el gen *arsC* (Figura 20), se suelen encontrar genes *arsR* y *arsD*. Por ende, en dichas regiones, es probable que tanto ArsR como ArsD regulen la transcripción de Acr3. A su vez, se ha observado que el gen *arsA* suele estar agrupado con el gen *acr3* (Figura 20). Si bien esto era de esperar, debido a que ArsA funciona como ATPasa y se une a Acr3 para mejorar su eficiencia de transporte, resulta una novedad ya que los genomas de microorganismos aislados de ambientes con altas concentraciones de arsénico presentan normalmente el gen *arsA* agrupado con el gen *arsB* y no con el gen *acr3*.



Figura 21. Organización de los genes *ars* en los metagenomas ensamblados de los ecosistemas microbianos andinos, parte 4. L.D., Laguna Diamante;

L.B., Laguna Brava; L.T., Laguna Tebenquiche; P.C.N., *Puquio* de campo naranja; L.P.B., Laguna Pozo Bravo; L.S., Laguna Socompa.

Otros genes que suelen agruparse con *acr3* son *arsM*, *arsP* y *arsH* (Figuras 18, 19, 20 y 21). A pesar de que dichos genes están involucrados en mecanismos de detoxificación distintos al del gen *acr3*, su agrupamiento podría indicar que presentan la misma regulación transcripcional. De hecho, es común encontrar el gen *arsR* junto a los genes *arsM*, *arsP* o *arsH* (Figura 21). Por lo tanto, es muy probable que ArsR regule la transcripción de dichos genes en los microorganismos andinos. Los genes *arsM* y *arsP* también suelen encontrarse en cercanía del gen *arsC* (Figuras 18, 19, 20 y 21). Como la metilación del arsenito por la enzima arsenito metiltransferasa (ArsM) y la extrusión del metilarsenito por la bomba ArsP están ligadas a la reducción citoplasmática del arsenato llevada a cabo por ArsC, resulta esperable que dichos genes se encuentren juntos y estén regulados transcripcionalmente de la misma manera.

En cuanto a los genes bioenergéticos (Figura 22), se ha observado que *aioA* y *aioB* por lo general se agrupan solos. En pocos casos, los genes *aioA* y *aioB* se agrupan con los genes *aioX*, *aioS* y *aioR*, o *aioF*. Por lo tanto, en pocos microorganismos andinos la transcripción de AioAB estaría regulada por los factores de transcripción AioX, AioS y AioR, o AioF. Al igual que los genes *aioA* y *aioB*, los genes *arxB'*, *arxA*, *arxB*, *arxC* y *arxD* se suelen agrupar solos y no con los genes que codifican para los factores de transcripción ArxX, ArxS y ArxR. De hecho, dichos genes no han sido identificados en ningún ecosistema estudiado, por lo que la expresión de ArxAB no estaría regulada por ArxX, ArxS y ArxR en los microorganismos andinos. Si bien no hay reportes de que los genes *arr* puedan estar regulados transcripcionalmente por un sistema de dos componentes, se han identificado posibles genes *arrX*, *arrS* y *arrR* en el metagenoma de la laguna Socompa.

Esta observación es novedosa, pero deberá ser investigada con más detalle para ser confirmada.

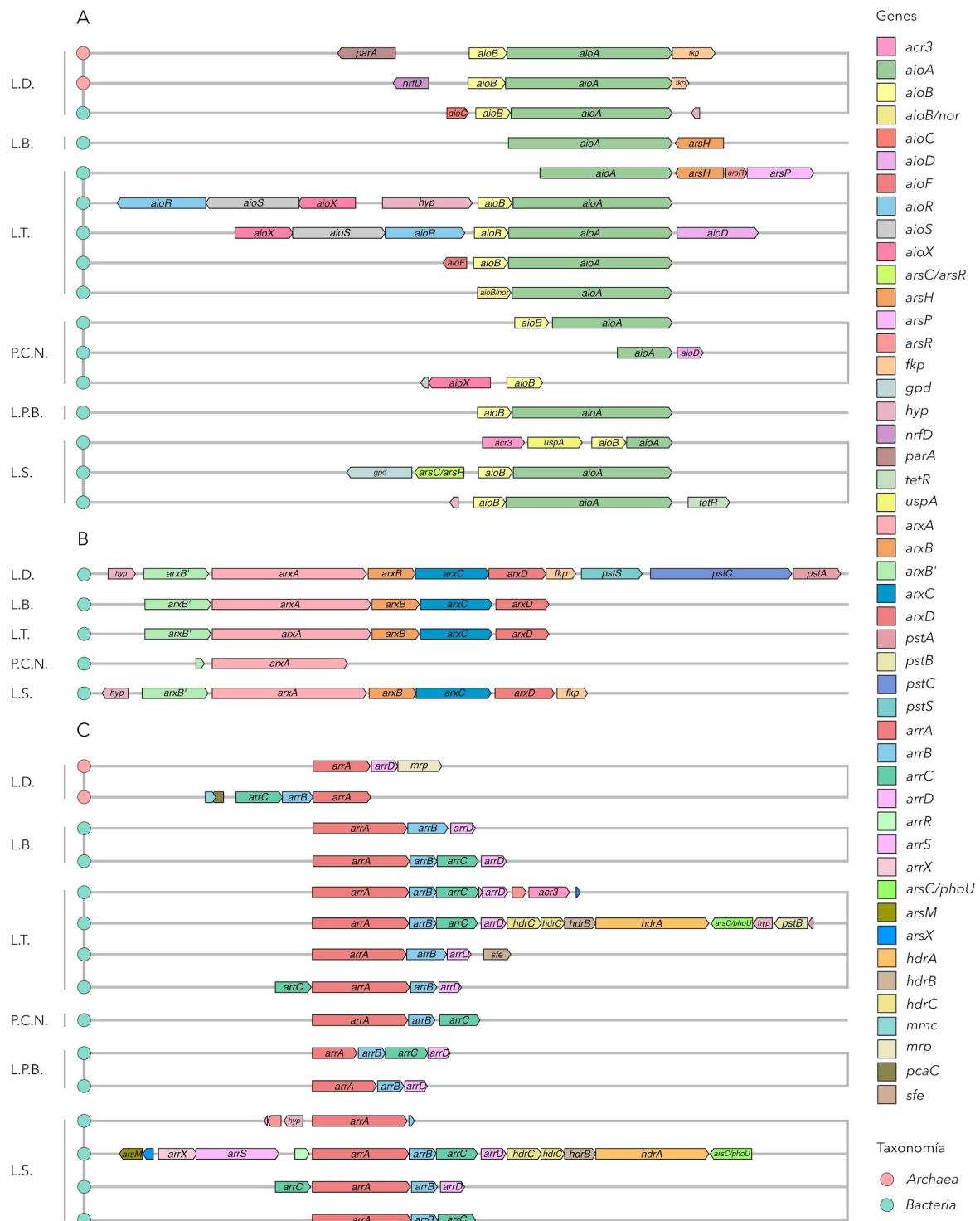


Figura 22. Organización de los genes *aio*, *arx* y *arr* en los metagenomas ensamblados de los ecosistemas microbianos andinos. L.D., Laguna

Diamante; L.B., Laguna Brava; L.T., Laguna Tebenquiche; P.C.N., *Puquio* de campo naranja; L.P.B., Laguna Pozo Bravo; L.S., Laguna Socompa.

Por otro lado, el análisis de los genes implicados en el metabolismo del arsénico ha revelado que muchos de ellos se encuentran fusionados. Como estas fusiones génicas han sido observadas en la mayoría de los ecosistemas microbianos andinos, lo más probable es que sean funcionales y estén conservadas evolutivamente. Entre las fusiones génicas identificadas se encuentran *arsC-arsR*, *arsC-phoU*, *arsC-arsI*, *arsC-arsH*, *arsD-arsM*, y *arsR-arsM*, entre otras. La fusión génica *arsC-arsR* ha sido reportada previamente en genomas de bacterias aisladas de ambientes con altas concentraciones de arsénico (J. Chen et al., 2022). Como el represor transcripcional ArsR suele unirse al arsenito y no al arsenato, se ha sugerido que la fusión de ArsC con ArsR permite a ArsR sensor el arsenato al canalizar el producto de ArsC hacia el sitio de unión de ArsR. Teniendo en cuenta esto, se podrían sugerir funciones semejantes para las fusiones ArsR-ArsM, ArsD-ArsM, y ArsC-PhoU. La fusión de ArsM con ArsR permitiría a ArsR sensor el metilarsenito al canalizar el producto de ArsM hacia el sitio de unión de ArsR. De la misma manera, la fusión de ArsM con ArsD permitiría a ArsD sensor el metilarsenito. Mientras que la fusión de ArsC con PhoU permitiría a PhoU sensor el arsenato en lugar del fosfato. De esta manera, se vincularía la transcripción de los genes *pst* con la concentración de arsénico. Las fusiones de ArsC con ArsI e ArsH, por otro lado, permitirían acoplar la reducción citoplasmática del arsenato con la desmetilación del metilarsenito y la oxidación del metilarsenito, respectivamente. Por algún motivo que se desconoce, el acoplamiento de estos procesos metabólicos traería una ventaja adaptativa frente a la detoxificación del arsénico, ya que dichas fusiones se encuentran ampliamente conservadas en la mayoría de los ecosistemas microbianos andinos.

CONCLUSIONES

A partir de los resultados obtenidos en la primera parte de la tesis, se puede concluir que:

- I. Todas las lagunas altoandinas estudiadas presentan condiciones ambientales extremas (aguas alcalinas, hipersalinas, con altos niveles de arsénico y bajos niveles de oxígeno). Sin embargo, de todas ellas, la laguna Diamante es la que presenta las condiciones ambientales más extremas.
- II. Como las condiciones ambientales de la laguna Diamante son más extremas que las del resto de las lagunas estudiadas, el *biofilm* rojo de Diamante está dominado por haloarqueas quimioheterótrofas, mientras que los ecosistemas microbianos de Brava, Tebenquiche, *Puquio* de campo naranja, Pozo Bravo y Socompa están dominados por bacterias de diferentes gremios (cianobacterias, bacterias púrpura del azufre, bacterias verdes no del azufre, bacterias sulfato reductoras, bacterias oxidadoras de azufre, bacterias púrpura no del azufre, bacterias fotosintéticas anoxigénicas aeróbicas, y bacterias quimioheterótrofas, entre otras). Por consiguiente, la diversidad alfa en el *biofilm* rojo de Diamante es menor que la diversidad alfa en el resto de los ecosistemas microbianos estudiados.
- III. La mayoría de los microorganismos de los Andes centrales presentan mecanismos moleculares de resistencia al arsénico. Si bien se identificaron diversos mecanismos, los principales son:

- Reducción citoplasmática del arsenato, mediada por ArsC, seguido de la expulsión del arsenito resultante por la bomba Acr3.
 - Metilación citoplasmática del arsenito, mediada por ArsM, seguido de la expulsión del metilarsenito resultante por la bomba ArsP. Eventualmente, el metilarsenito puede ser oxidado a metilarsenato por ArsH.
- IV. Algunos microorganismos de los Andes centrales, además de resistir al arsénico, tienen la capacidad de obtener energía metabólica a partir de su oxidorreducción:
- La obtención de energía, por oxidación del arsenito, es llevada a cabo principalmente por AioAB en los microorganismos andinos. Como la abundancia de genes *aio* en el *biofilm* rojo de la laguna Diamante es alta, las haloarqueas de Diamante podrían oxidar el arsenito, utilizando AioAB, para obtener energía o detoxificarse. De ser así, se demostraría un nuevo metabolismo en el dominio *Archaea*.
 - La respiración de arsenato, llevada a cabo por ArrAB, permite a algunos microorganismos andinos crecer en anaerobiosis. Como la abundancia de genes *arr* en el *biofilm* rojo de la laguna Diamante es alta, las haloarqueas de Diamante podrían reducir el arsenato, utilizando ArrAB, para oxidar compuestos orgánicos en anaerobiosis. De ser así, se demostraría un nuevo metabolismo en el filo *Euryarchaeota*.
- V. Los genes del metabolismo del arsénico se arreglan de diversas maneras en los genomas de los microorganismos andinos.

PARTE 2

INTRODUCCIÓN

Oxidación y reducción del arsénico en el dominio *Archaea*

Dentro del dominio *Bacteria*, se ha demostrado la capacidad de ciertos organismos de utilizar las enzimas arsenito oxidasas (AioAB y ArxAB) para obtener energía de la oxidación del arsenito. Entre estos organismos se encuentran *Rhizobium* sp. NT-26 (Santini et al., 2000; Santini & vanden Hoven, 2004), *Azoarcus* sp. DA01 (Rhine et al., 2007), *Dechloromonas* sp. ECC1-pb1 (Sun et al., 2010) y *Alkalilimnicola ehrlichii* MLHE-1 (Zargar et al., 2010). A su vez, se ha demostrado que ciertos organismos, como *Chrysiogenes arsenatis* (Macy et al., 1996) y *Shewanella* sp. ANA-3 (Saltikov et al., 2003), utilizan la enzima arsenato reductasa (ArrAB) para respirar arsenato en lugar de oxígeno.

Sin embargo, hasta la fecha de escritura de este trabajo, no se ha demostrado la oxidación del arsenito mediante arsenito oxidasas (AioAB o ArxAB) en ningún organismo del dominio *Archaea*. Únicamente se ha demostrado que dos arqueas hipertermófilas del filo Crenarchaeota, *Pyrobaculum arsenaticum* y *Pyrobaculum aerophilum* (Huber et al., 2000), pueden respirar arsenato utilizando la enzima arsenato reductasa (ArrAB). Por lo tanto, si se prueba que las haloarqueas de la laguna Diamante pueden oxidar el arsenito utilizando AioAB, se demostraría un nuevo metabolismo en el dominio *Archaea*. Mientras que, si se prueba que las haloarqueas de la laguna Diamante pueden reducir el arsenato utilizando ArrAB, se demostraría un nuevo metabolismo en el filo *Euryarchaeota*.

OBJETIVOS

Objetivos específicos

Los objetivos específicos de la segunda parte de la tesis son:

1. Aislar una haloarquea, resistente a altas concentraciones de arsénico, del *biofilm* rojo de la laguna Diamante.
2. Caracterizar, a partir del análisis genómico, la haloarquea aislada:
 - a. Identificar genes implicados en el metabolismo del arsénico y determinar su organización dentro del genoma.
 - b. Comparar el genoma de la haloarquea aislada con el de otras haloarqueas que poseen también genes implicados en el metabolismo del arsénico.
3. Determinar si la haloarquea aislada es capaz de crecer como quimiolitótrofo oxidando el arsenito y respirando oxígeno, y/o como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato.

MATERIALES Y MÉTODOS

Aislamiento de haloarqueas del *biofilm* rojo de la laguna Diamante



Figura 23. Toma de muestras del *biofilm* rojo de la laguna Diamante para el aislamiento de haloarqueas.

Para aislar haloarqueas del *biofilm* rojo de la laguna Diamante, en primer lugar se llevó a cabo un enriquecimiento selectivo. Para ello, se tomaron 2 gr de *biofilm* (almacenado a 4°C luego de su recolección (Figura 23)) y se colocaron en 20 ml de medio WJK líquido con $\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$ 20 mM y NaAsO_2 1 mM. El medio WJK es un medio rico para haloarqueas que contiene, por litro de agua desionizada, 252 gr de NaCl ; 0,5 gr de $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$; 0,5 gr de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$; 1 gr de KNO_3 ; 5,84 gr de KCl ; 7,5 gr de peptona (Oxoid); 1,5 gr de extracto de levadura (Oxoid); y 1 ml de solución

traza (HCl 25% 10 ml L⁻¹; FeCl₃ 1,5 g L⁻¹; ZnCl₂ 70 mg L⁻¹; MnCl₂·4H₂O 100 mg L⁻¹; H₃BO₃ 6 mg L⁻¹; CoCl₂·2H₂O 190 mg L⁻¹; CuCl₂·2H₂O 2 mg L⁻¹; NiCl₂·6H₂O 24 mg L⁻¹; Na₂MoO₄·2H₂O 36 mg L⁻¹). El preparado se incubó durante 30 días a 37°C bajo agitación (120 RPM). Transcurrida la incubación, se tomaron 100 µl de cultivo y se plaquearon sobre 20 ml de medio WJK sólido (1,45% de agar) con Na₂HAsO₄·7H₂O 20 mM y NaAsO₂ 1 mM. La placa preparada se incubó aeróbicamente durante 12 días a 37°C. Como resultado se obtuvieron cinco colonias de haloarqueas resistentes a altas concentraciones de arsénico. Posteriormente, las colonias de haloarqueas fueron identificadas taxonómicamente mediante amplificación y secuenciación (*3730xl DNA Analyzer; Macrogen, Inc.*) del gen ARN ribosomal 16S utilizando los *primers* 344 F y 915 R. Al comparar las secuencias obtenidas con las secuencias en la base de datos SILVA (Quast et al., 2013), se pudo determinar que todas las colonias analizadas pertenecían a haloarqueas del género *Halorubrum*. Finalmente, con el fin de seleccionar la haloarquea con mayor resistencia al arsénico, se empleó la técnica de goteo en placa. Para ello, se crecieron las cinco haloarqueas aisladas (DM1-DM5) a 37°C bajo agitación (120 RPM) en 50 ml de medio WJK líquido hasta alcanzar una densidad óptica (DO600) igual a 0,7. Luego se tomaron 10 µl de cinco diluciones seriadas (10⁻¹-10⁻⁶) de cada cultivo, y se plaquearon sobre 20 ml de medio CDM sólido (1,45% de agar) con Na₂HAsO₄·7H₂O 50, 150 y 250 mM por un lado, y con NaAsO₂ 1, 3 y 5 mM por otro lado. El medio CDM (Kauri et al., 1990) es un medio mínimo para haloarqueas que contiene, por litro de agua desionizada, 250 gr de NaCl; 1 gr de MgCl₂·6H₂O; 5 gr de K₂SO₄; 267 mg de NH₄Cl; 174 mg de K₂HPO₄; 367,5 mg de CaCl₂; 0,8 mg de tiamina; 0,1 mg de biotina; 5 gr de piruvato de sodio; 20 ml de Tris-HCl, pH 8, 1M; y 1 ml de solución traza (FeSO₄·7H₂O 2,3 gr L⁻¹; ZnSO₄·7H₂O 440 mg L⁻¹; MnCl₂·4H₂O 360 mg L⁻¹; CuSO₄·5H₂O 50 mg L⁻¹; Na₂MoO₄·2H₂O 0,5 mg L⁻¹). Todas las placas se incubaron aeróbicamente durante 15 días a 37°C. Los resultados demostraron que todas las haloarqueas aisladas son capaces de crecer en medio CDM con la

concentración más alta de As(V) evaluada (250 mM). Sin embargo, únicamente *Halorubrum* sp. DM2 es capaz de crecer en medio CDM con la concentración más alta de As(III) evaluada (5 mM). Por lo tanto, como esta haloarquea es la que presenta la mayor resistencia al arsénico, se la seleccionó para proceder con los subsiguientes análisis. El aislamiento de haloarqueas fue un trabajo que se realizó en conjunto con la Lic. Mariana Noelia Soria y el Dr. Omar Federico Ordóñez en PROIMI-CONICET.

Extracción y secuenciación del ADN genómico de *Halorubrum* sp. DM2

Para extraer el ADN genómico de *Halorubrum* sp. DM2, primero se creció la haloarquea en 50 ml de medio WJK durante 4 días a 37°C bajo agitación (180 RPM). Transcurridos los 4 días, se centrifugó el cultivo (6.500 xg por 10 min a 4°C) y se descartó el sobrenadante. Con el *pellet* de células remanente, se procedió a extraer el ADN genómico utilizando el *ZymoBIOMICS™ DNA Miniprep Kit* (Zymo Research) de acuerdo con las instrucciones del fabricante. La secuenciación del ADN genómico extraído se realizó a través de la plataforma de secuenciación Illumina MiSeq. Las lecturas *paired end* crudas (*raw sequence data*) se depositaron en el repositorio de secuencias nucleotídicas ENA bajo el número de acceso ERP113393.

Ensamblado de las lecturas genómicas de *Halorubrum* sp. DM2

En primer lugar, empleando la herramienta *online* FastQC v.0.11.5 (Babraham Bioinformatics), se llevó a cabo un control de calidad de las lecturas genómicas obtenidas. En dicho control de calidad se determinó el número total de lecturas (*total sequences*), % GC de lecturas, calidad por base de lecturas (*per base sequence quality*), calidad promedio de lecturas (*per sequence quality scores*), contenido de bases de lecturas (*per base sequence*

content), contenido de bases ambiguas (N) de lecturas (*per base N content*), distribución del largo de lecturas (*sequence length distribution*), duplicaciones de lecturas, y contenido de secuencias adaptadoras de las lecturas (*adapter content*). Posteriormente, teniendo en cuenta la información obtenida en el control de calidad, se realizó un preprocesamiento de las lecturas empleando la herramienta Trimmomatic v.0.36 (Bolger et al., 2014). Los parámetros utilizados para el preprocesamiento de las lecturas fueron los siguientes: ILLUMINACLIP = NexteraPE-PE.fa:2:30:5, SLIDINGWINDOW = 4:20, AVGQUAL = 20, MINLEN = 50. De esta manera, se removieron secuencias adaptadoras, lecturas con calidad promedio menor a Q20, lecturas con un largo menor a 50 bases, y extremos de lecturas con calidad promedio (cada cuatro bases) menor a Q20.

Luego del preprocesamiento, las lecturas genómicas fueron ensambladas *de novo* para obtener *contigs*. El ensamblado *de novo* se llevó a cabo empleando la herramienta SPAdes v.3.11.0 (Bankevich et al., 2012). Una vez finalizado el ensamblado, se evaluó la calidad del mismo utilizando la herramienta QUAST v.4.6 (Gurevich et al., 2013). En dicho control de calidad se determinó el largo máximo de *contig* (*maximum contig length*), el largo total del ensamblado (*total assembly length*), el número de *contigs* (*contig count*) y el N50. Posteriormente, empleando la herramienta Bowtie2 v.2.3.2 (Langmead & Salzberg, 2012), se determinó el porcentaje de mapeo de las lecturas sobre los *contigs* obtenidos (*read mapping rate*). La eficiencia de ensamblado se obtuvo multiplicando el N50 por el porcentaje de mapeo de las lecturas.

Finalmente, utilizando la herramienta Multi-CSAR (*Multiple reference-based Contig Scaffold using Algebraic Rearrangements*) (K.-T. Chen et al., 2018), se compararon los *contigs* obtenidos con múltiples genomas de referencia del género *Halorubrum*. De esta manera, se ordenaron y orientaron los *contigs* en un único *scaffold*.

Anotación estructural y funcional del genoma de *Halorubrum* sp. DM2

La predicción de los genes en el genoma ensamblado (anotación estructural) se realizó con la herramienta Prodigal (*PROkaryotic DYnamic programming Gene-finding ALgorithm*) v.2.6.3 (Hyatt et al., 2010) y el servidor RAST (Aziz et al., 2008). Una vez identificados los genes, se llevó a cabo la caracterización funcional de los mismos (anotación funcional) empleando la herramienta InterProScan v.5.34-73.0 (Jones et al., 2014) y el servidor RAST (Aziz et al., 2008). La identificación de los genes implicados en el metabolismo del arsénico y el estudio de su organización genómica, se llevó a cabo como se describió en los materiales y métodos de la primera parte del trabajo de tesis. Por último, se obtuvo una visualización circular del genoma anotado utilizando la herramienta Circleator v.1.0.1 (Crabtree et al., 2014).

Análisis de similitud genómica de *Halorubrum* sp. DM2

Para determinar si *Halorubrum* sp. DM2 es una especie no caracterizada dentro del género *Halorubrum*, se obtuvieron los valores de ANIb (Identidad promedio nucleotídica usando BLAST), DDH (Hibridación ADN-ADN) y *Mash distance* (Distancia genómica usando Mash) entre el genoma de dicha especie y los genomas disponibles (Mayo de 2019) del género *Halorubrum* en la base de datos GenBank (Benson et al., 2018). La herramienta *online* JSpeciesWS (Richter et al., 2016) se utilizó para obtener los valores de ANIb y DDH, mientras que la herramienta Mash v2.1.1 (Ondov et al., 2016) se utilizó para obtener los valores de *Mash distance*. Los valores de ANIb, DDH y *Mash distance* también se obtuvieron entre el genoma de *Halorubrum* sp. DM2 y los MAGs D.1.1, D.2.1, D.3.2, y D.3.4 para determinar si alguno de los MAGs reconstruidos de los metagenomas de Diamante pertenece a *Halorubrum* sp. DM2.

Análisis filogenético de *Halorubrum* sp. DM2

Posteriormente se llevó a cabo un análisis filogenético de *Halorubrum* sp. DM2 basado en el gen ARN ribosomal 16S. Dicho análisis se realizó con la herramienta MEGA X v.10.1 (Kumar et al., 2018) empleando el método de máxima verosimilitud basado en el modelo de Tamura-Nei. Las secuencias del gen ARN ribosomal 16S de las especies conocidas del género *Halorubrum*, que se utilizaron en este análisis, se descargaron de la base de datos SILVA (Quast et al., 2013) en mayo de 2019. Secuencias del gen ARN ribosomal 16S de ciertas especies conocidas del género *Haloferax* también se descargaron de esta base de datos y se incorporaron en el análisis para funcionar como grupo externo (*outgroup* en inglés) y enraizar el árbol filogenético.

Análisis de genes del metabolismo del arsénico en *Halorubrum* sp. DM2

Para comprender mejor los mecanismos de resistencia al arsénico y obtención de energía a partir de su oxidorreducción en el dominio *Archaea*, se analizó la abundancia y organización genómica de los genes del metabolismo del arsénico en *Halorubrum* sp. DM2 y otras 15 arqueas que presentan genes *aio* y/o *arr* [*Halorubrum* sp. AJ67 (GCA_000513435.1), *Halorubrum* sp. BC156, *Halorubrum tebenquichense* (GCA_000337415.1), *Halorubrum kocurii* (GCA_000337355.1), *Halobiforma lacisalsi* (GCA_000226975.3), *Sulfolobus tokodaii* (GCA_000011205.1), *Sulfolobus metallicus* (GCA_009729515.1), *Acidianus hospitalis* (GCA_000213215.1), *Acidianus manzaensis* (GCA_002116695.1), *Metallosphaera cuprina* (GCA_000204925.1), *Aeropyrum camini* (GCA_000591035.1), *Aeropyrum pernix* (GCA_000011125.1), *Pyrobaculum oguniense* (GCA_000247545.1), *Pyrobaculum aerophilum* (GCA_000007225.1), y *Pyrobaculum arsenaticum*

(GCA_000016385.1)]. Los genomas ensamblados de estas arqueas se descargaron de la base de datos GenBank (Benson et al., 2018), excepto el genoma de *Halorubrum* sp. BC156 que fue secuenciado y ensamblado previamente en el laboratorio de la Dra. María Eugenia Farías (Benson et al., 2018). La identificación de los genes implicados en el metabolismo del arsénico y el estudio de su organización genómica, en estos genomas descargados, se llevó a cabo como se describió en los materiales y métodos de la primera parte del trabajo de tesis.

Ensayos de crecimiento de *Halorubrum* sp. DM2

Con el fin de caracterizar el metabolismo energético de *Halorubrum* sp. DM2, se llevaron a cabo los siguientes ensayos de crecimiento:

1. Crecimiento en aerobiosis, oscuridad y medio rico (WJK)
2. Crecimiento en aerobiosis, oscuridad y medio mínimo (CDM) con:
 - A. Arsenito [NaAsO_2] 1mM
 - B. Arsenato [$\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$] 20 mM
 - C. Sin arsenito ni arsenato (control)
3. Crecimiento en aerobiosis, oscuridad y medio mínimo (CDM) modificado (se reemplazó el piruvato por bicarbonato de sodio 0,2 % m/v) con:
 - A. Arsenito [NaAsO_2] 1mM
 - B. Arsenato [$\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$] 20 mM
 - C. Sin arsenito ni arsenato (control)
4. Crecimiento en anaerobiosis, oscuridad y medio mínimo (CDM) con:
 - A. Arsenito [NaAsO_2] 1mM
 - B. Arsenito [NaAsO_2] 1mM y sulfato [Na_2SO_4] 20 mM
 - C. Arsenito [NaAsO_2] 1mM y nitrato [NaNO_3] 20 mM

- D. Arsenato [$\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$] 20 mM
- E. Arsenato [$\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$] 20 mM y sulfato [Na_2SO_4] 20 mM
- F. Arsenato [$\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$] 20 mM y nitrato [NaNO_3] 20 mM
- G. Sulfato [Na_2SO_4] 20 mM
- H. Nitrato [NaNO_3] 20 mM
- I. Sin arsenito, arsenato, sulfato, ni nitrato (control)

El ensayo de crecimiento en aerobiosis, oscuridad y medio rico (1) se llevó a cabo por triplicado utilizando frascos *Erlenmeyer* de vidrio (100 ml) estériles. Primero se inocularon 500 μl de un cultivo en fase exponencial ($\text{DO}_{600} \approx 1$) de *Halorubrum* sp. DM2 (crecido en medio WJK) en 50 ml de medio WJK. Luego se incubó el cultivo durante 7 días en oscuridad a 37°C bajo agitación (180 RPM). Alícuotas de 500 μl se tomaron cada 24 horas para medir la densidad óptica (DO_{600}) y así poder determinar el crecimiento del cultivo a lo largo del tiempo.

El ensayo de crecimiento en aerobiosis, oscuridad y medio mínimo (2) también se llevó a cabo por triplicado utilizando frascos *Erlenmeyer* de vidrio (100 ml) estériles. En primer lugar se inocularon 500 μl de un cultivo en fase exponencial ($\text{DO}_{600} \approx 1$) de *Halorubrum* sp. DM2 (crecido en medio WJK) en 50 ml de medio CDM. El cultivo se incubó a 37°C bajo agitación (180 RPM) hasta alcanzar una densidad óptica (DO_{600}) igual a 0,7. Luego se tomaron 500 μl del cultivo, se inocularon nuevamente en 50 ml de medio CDM, y el preparado se incubó a 37°C bajo agitación (180 RPM) hasta alcanzar una densidad óptica (DO_{600}) igual a 0,7. Finalmente se inocularon 500 μl del cultivo en 50 ml de medio CDM (A) con arsenito 1 mM, (B) con arsenato 20 mM y (C) sin arsenito ni arsenato (control). Los cultivos se incubaron durante 7 días en oscuridad a 37°C bajo agitación (180 RPM). Alícuotas de 500 μl se tomaron cada 24 horas para medir la densidad óptica (DO_{600}) y así poder determinar el crecimiento de los cultivos a lo largo del tiempo. El ensayo de

crecimiento en aerobiosis, oscuridad y medio mínimo modificado (3) se llevó a cabo de la misma manera que el ensayo (2). Lo único que se modificó fue el medio mínimo, que en lugar de tener piruvato tiene bicarbonato de sodio 0,2 % m/v.

A diferencia de los ensayos anteriores, el ensayo de crecimiento en anaerobiosis, oscuridad y medio mínimo (4) se llevó a cabo por triplicado utilizando frascos de septos de vidrio de boca estrecha con tapón de apertura (60 ml) estériles. Primero se inocularon 500 µl de un cultivo en fase exponencial ($DO_{600} \approx 1$) de *Halorubrum* sp. DM2 (crecido en medio WJK) en 50 ml de medio CDM. El cultivo se incubó a 37°C bajo agitación (180 RPM) hasta alcanzar una densidad óptica (DO_{600}) igual a 0,7. Luego se tomaron 500 µl del cultivo, se inocularon nuevamente en 50 ml de medio CDM, y el preparado se incubó a 37°C bajo agitación (180 RPM) hasta alcanzar una densidad óptica (DO_{600}) igual a 0,7. Finalmente, utilizando una jeringa de 1 ml estéril, se inocularon 500 µl del cultivo en frascos de septos conteniendo 50 ml de medio CDM (A) con arsenito 1 mM, (B) con arsenito 1 mM y sulfato 20 mM, (C) con arsenito 1 mM y nitrato 20 mM, (D) con arsenato 20 mM, (E) con arsenato 20 mM y sulfato 20 mM, (F) con arsenato 20 mM y nitrato 20 mM, (G) con sulfato 20 mM, (H) con nitrato 20 mM y (I) sin arsenito, arsenato, sulfato ni nitrato (control). Los cultivos se burbujearon con nitrógeno gaseoso (N_2) durante 30 minutos para desplazar el oxígeno disuelto y luego se incubaron durante 12 días en oscuridad a 37°C. Alícuotas de 500 µl se tomaron cada 3 días, utilizando una jeringa de 1 ml estéril, para medir la densidad óptica (DO_{600}) y así poder determinar el crecimiento de los cultivos a lo largo del tiempo.

Cabe destacar que todos los ensayos de crecimiento se realizaron en conjunto con la Lic. Mariana Noelia Soria y la Dra. Tatiana Mariel Stepanenko.

Obtención de una mutante *Halorubrum* sp. DM2 $\Delta arrA$

Para corroborar que *Halorubrum* sp. DM2 es capaz de crecer como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato, se procedió a obtener una mutante *Halorubrum* sp. DM2 $\Delta arrA$. En un principio, se intentó obtener la mutante siguiendo un protocolo diseñado originalmente para las haloarqueas modelo *Haloferax volcanii* y *Halorubrum lacusprofundi* (Allers et al., 2004; Bitan-Banin et al., 2003; Liao et al., 2016). El mismo consiste en transformar la haloarquea con un plásmido pTA131 que contiene el gen de resistencia a pravastatina (*hmgA*) flanqueado por los extremos 3' y 5' del gen *arrA*. Una vez transformada la haloarquea, se la crece en un medio sólido con pravastatina para seleccionar las células que hayan incorporado el gen *hmgA* por recombinación homóloga del gen *arrA* cromosomal con los extremos 3' y 5' del mismo gen que flanquean a *hmgA* en el plásmido pTA131. Dichas células, al haber realizado la recombinación homóloga de manera exitosa, serán mutantes $\Delta arrA$. De todas formas, a pesar de numerosos intentos, no se logró obtener una mutante *Halorubrum* sp. DM2 $\Delta arrA$ a partir de esta metodología. Por lo tanto, se procedió a obtener una mutante *Halorubrum* sp. DM2 $\Delta arrA$ mediante numerosos pasajes de *Halorubrum* sp. DM2 en medio WJK en aerobiosis (37°C, 180 RPM) sin arsénico. Al eliminar la presión de selección (alta concentración de arsénico), se buscó que el gen *arrA* mutara y perdiera su función. Para identificar una mutante *Halorubrum* sp. DM2 $\Delta arrA$ en los sucesivos pasajes, se realizaron ensayos de crecimiento en anaerobiosis, oscuridad y medio mínimo CDM con arsenato 20 mM (12 días a 37°C). De esta manera, se identificó una posible mutante *Halorubrum* sp. DM2 $\Delta arrA$ capaz de crecer en aerobiosis en medio WJK o CDM, pero incapaz de crecer en anaerobiosis en medio CDM con arsenato 20 mM. Para confirmar la obtención de dicha mutante, se extrajo su ADN genómico de la misma manera en que se extrajo el ADN genómico de la cepa salvaje (*wildtype* en inglés). Luego se secuenció el ADN genómico

extraído a través de la plataforma de secuenciación Oxford Nanopore. Las lecturas genómicas obtenidas fueron ensambladas para obtener *contigs* utilizando la herramienta Canu v.2.2. (Koren et al., 2017). A partir de dichos *contigs* se identificaron los genes codificantes de proteínas empleando la herramienta Prodigal (*PRO*karyotic *DY*namically *programmed* *Gene-finding AL*gorithm) v.2.6.3 (Hyatt et al., 2010). Al analizar los genes del metabolismo del arsénico en la cepa mutante, se pudo determinar que el gen *arrA* se encuentra mutado ya que presenta un codón de terminación prematuro. De esta manera se corroboró la obtención de una mutante *Halorubrum* sp. DM2 $\Delta arrA$.

RESULTADOS

Ensamblado y anotación del genoma de *Halorubrum* sp. DM2

A partir del control de calidad de las lecturas genómicas, se determinaron parámetros como la calidad por base, la calidad promedio, el contenido de bases ambiguas, el contenido de secuencias adaptadoras, y el contenido de GC. La calidad de las lecturas genómicas disminuye en las últimas bases, como es habitual en la mayoría de las lecturas obtenidas a partir de plataformas de secuenciación Illumina (Tan et al., 2019). En las lecturas *forward*, la calidad cae por debajo de Q20 aproximadamente en la posición 240. Mientras que en las lecturas *reverse*, la calidad cae por debajo de Q20 alrededor de la posición 170. Por esta razón, en el preprocesamiento de las lecturas genómicas se removieron los extremos con calidad promedio (cada cuatro bases) menor a Q20. La mayoría de las lecturas genómicas presentan una calidad promedio superior a Q20. Teniendo en cuenta esto, en el preprocesamiento de las lecturas genómicas se removieron lecturas con calidad promedio menor a Q20. La mayor parte de las lecturas genómicas no contienen bases ambiguas ni secuencias adaptadoras, aún así, se removieron posibles secuencias adaptadoras durante el preprocesamiento. Como el contenido de GC fue homogéneo en las lecturas genómicas, se pudo determinar que el genoma no presenta contaminaciones.

Luego de realizar el control de calidad y preprocesamiento de las lecturas genómicas, las mismas fueron ensambladas *de novo* con la herramienta SPAdes para obtener *contigs*. Para evaluar la calidad del ensamblado, se determinaron parámetros como el largo total, el número de *contigs*, el largo máximo de *contig*, el N50, el porcentaje de mapeo de las lecturas, y la eficiencia de ensamblado (Tabla 9). Como el porcentaje de mapeo de las lecturas fue mayor a 99 %, se puede decir que prácticamente la totalidad de

las lecturas genómicas fueron ensambladas con éxito. A su vez, como el largo total (3,52 Mb) es semejante al largo de los genomas del género *Halorubrum* en las bases de datos, se podría decir que el genoma ensamblado de *Halorubrum* sp. DM2 está completo.

Tabla 9. Estadísticas del genoma de *Halorubrum* sp. DM2.

Estadísticas	<i>Halorubrum</i> sp. DM2
Secuenciación	
Plataforma	Illumina MiSeq
Tipo de lecturas	<i>paired end</i>
N° de acceso	ERP113393
% GC de lecturas	65
Largo de lecturas (pb)	35-301
Ensamblado de las lecturas genómicas	
Largo total (Mb) (<i>contigs</i> ≥ 500 pb)	3,52
N° de <i>contigs</i>	144
Largo máximo de <i>contig</i> (kb)	482,06
N50 (pb)	206.816
% de mapeo de las lecturas	99,10
Eficiencia de ensamblado	20.495,47
Anotación de los <i>contigs</i>	
N° de ORFs identificados	3.605

ORFs (*open reading frames*), Marcos de lectura abiertos.

La anotación del genoma ensamblado permitió identificar 3.605 genes codificantes de proteínas, 48 genes ARN de transferencia y 6 genes ARN ribosomales (Figura 24). Estos valores también son semejantes a los valores reportados en los genomas anotados del género *Halorubrum*, lo cual corrobora que el genoma ensamblado de *Halorubrum* sp. DM2 está completo.

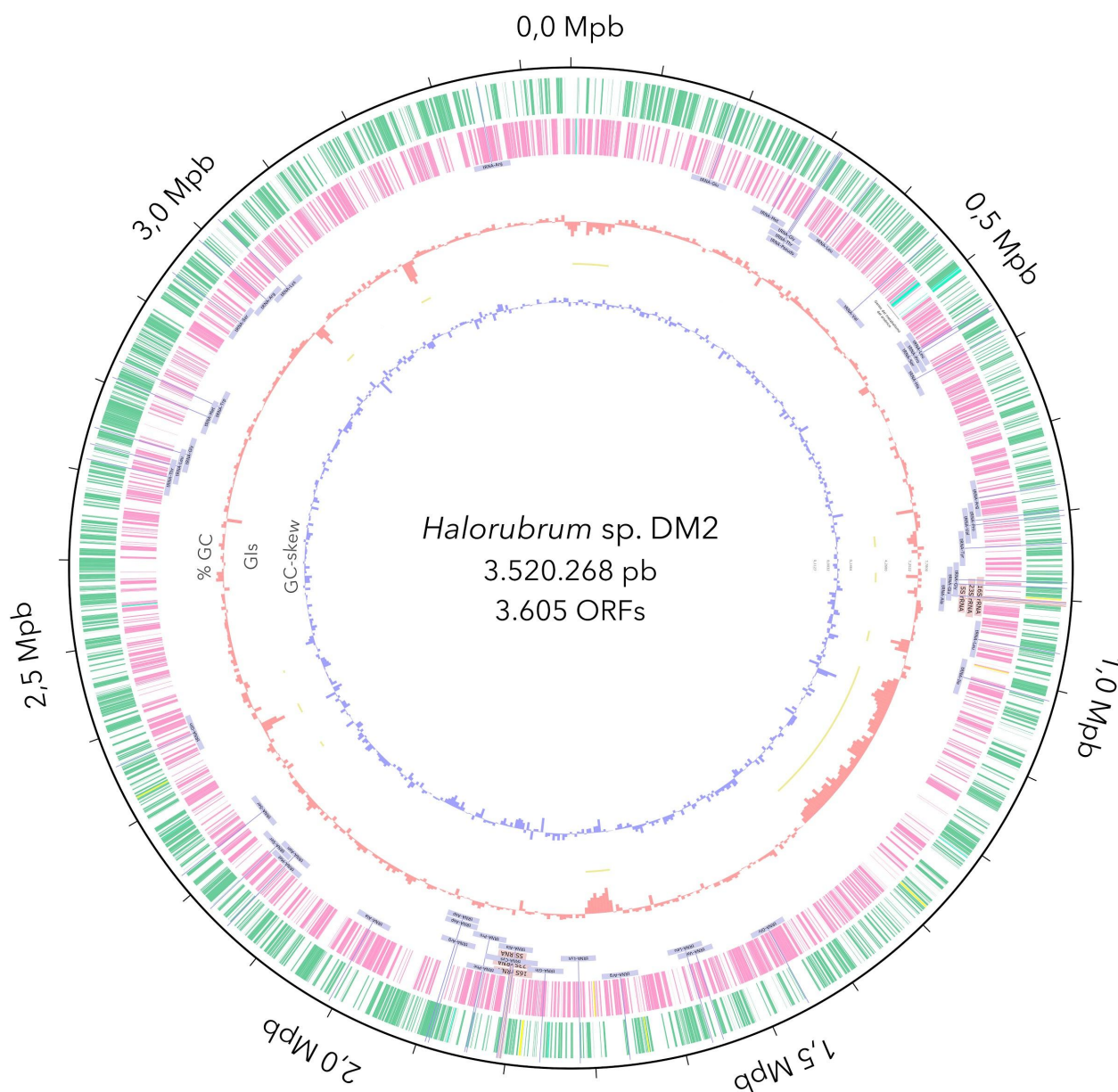


Figura 24. Visualización circular del genoma de *Halorubrum* sp. DM2. El primer anillo de color verde representa los genes “sentido”, el segundo anillo de color rosa representa los genes “antisentido”, el tercer anillo de color rojo grafica el %GC a lo largo del genoma, el cuarto anillo de color amarillo señala las regiones con posibles elementos móviles, y el quinto anillo de color azul indica el *GC skew* a lo largo del genoma. Adicionalmente, se encuentran etiquetas de color azul y rojo que señalan la ubicación de los genes ARN de transferencia y ARN ribosomales, respectivamente.

Análisis de similitud genómica de *Halorubrum* sp. DM2

Para determinar si *Halorubrum* sp. DM2 es una especie no caracterizada dentro del género *Halorubrum*, se analizaron los valores obtenidos de ANIb (Identidad promedio nucleotídica usando BLAST), DDH (Hibridación ADN-ADN) y *Mash distance* (Distancia genómica usando Mash) entre el genoma ensamblado de *Halorubrum* sp. DM2 y los genomas disponibles (Mayo de 2019) del género *Halorubrum* en la base de datos GenBank (Benson et al., 2018) (Tabla A3 en Apéndice). La comparación del genoma ensamblado de *Halorubrum* sp. DM2 con el genoma de *Halorubrum* sp. AJ67 fue la única que presentó un valor de ANIb > 95 %, un valor de DDH > 80 % y un valor de *Mash distance* < 0,05 %. Por lo tanto, se puede decir que *Halorubrum* sp. DM2 y *Halorubrum* sp. AJ67 son distintas cepas de una misma especie. *Halorubrum* sp. AJ67 también fue aislada de una laguna en la Puna de Catamarca (Burguener et al., 2014), lo cual podría significar que esta especie es una de las pocas especies del género *Halorubrum* capaces de ser cultivadas y aisladas de las lagunas altoandinas. Al analizar los valores obtenidos de ANIb, DDH y *Mash distance* entre el genoma ensamblado de *Halorubrum* sp. DM2 y los MAGs D.1.1, D.2.1, D.3.2, y D.3.4 (Tabla A4 en Apéndice), se pudo determinar que ninguno de los genomas reconstruidos a partir de los metagenomas de Diamante corresponde al genoma de *Halorubrum* sp. DM2. Esto indicaría que *Halorubrum* sp. DM2 no es una haloarquea representativa del *biofilm* rojo de la laguna Diamante, y que dicho ecosistema microbiano está dominado por haloarqueas no cultivables (o difíciles de cultivar) en el laboratorio.

Análisis filogenético de *Halorubrum* sp. DM2

Escala: 0,01 —

Género

■ *Halorubrum*
■ *Haloferax*

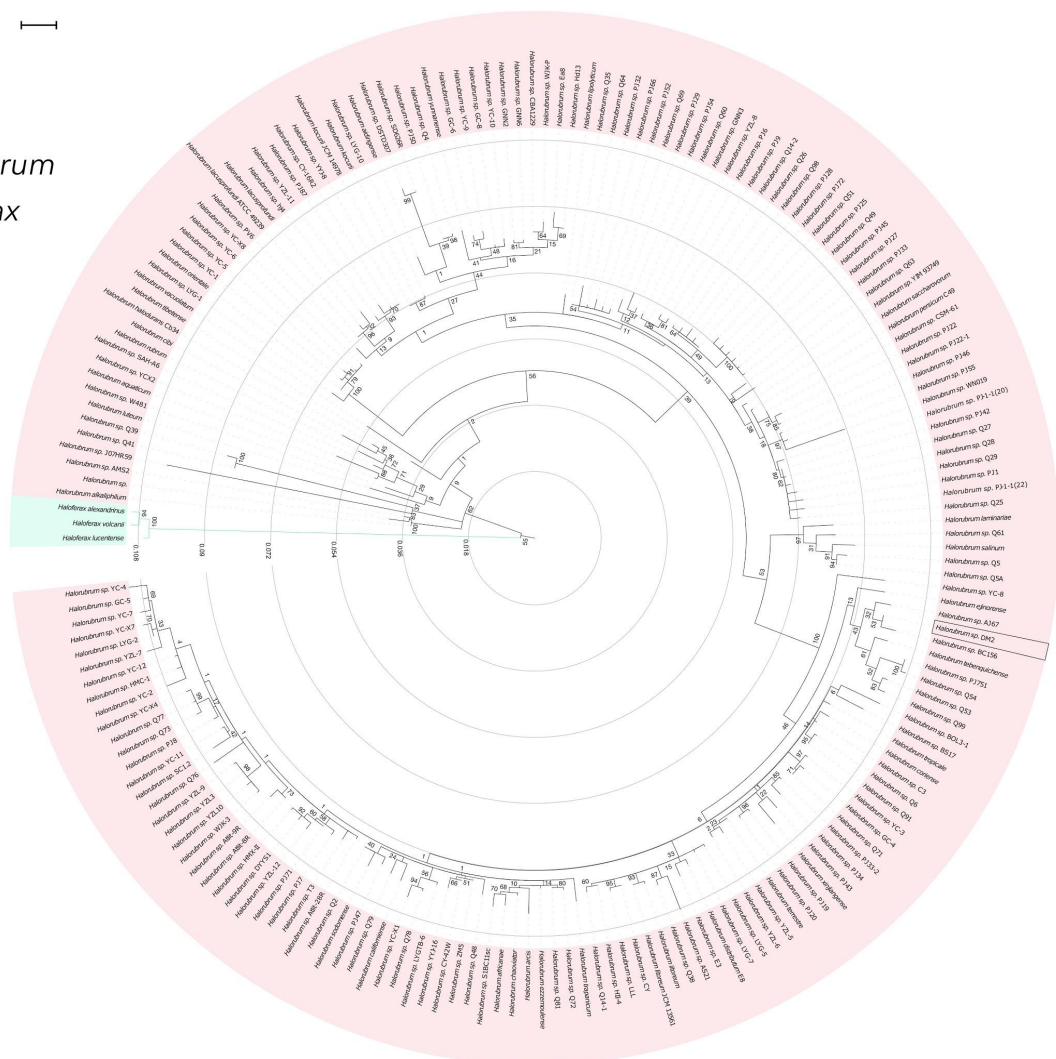


Figura 25. Árbol filogenético del género *Halorubrum* basado en el gen ARN ribosomal 16S. Los miembros del género *Halorubrum* se encuentran resaltados en rosa, mientras que los miembros del género *Haloferax* (grupo externo) se encuentran resaltados en verde.

El análisis filogenético de *Halorubrum* sp. DM2, basado en el gen ARN ribosomal 16S (Figura 25), permitió determinar que la haloarquea aislada del *biofilm* rojo de la laguna Diamante está más relacionada evolutivamente con *Halorubrum* sp. AJ67 y *Halorubrum* sp. BC156 que con otras haloarqueas del

género *Halorubrum* aisladas de lagunas altoandinas (p.ej., *Halorubrum tebenquichense* y *Halorubrum* sp. BOL3-1). Este resultado era de esperar, por un lado, debido a que *Halorubrum* sp. AJ67 y *Halorubrum* sp. DM2 son dos cepas diferentes de una misma especie, y por otro lado, debido a que *Halorubrum* sp. BC156 es una haloarquea que también fue aislada del *biofilm* rojo de la laguna Diamante. Lo sorprendente es que *Halorubrum* sp. DM2 esté más relacionada evolutivamente con *Halorubrum* sp. BC156 que con *Halorubrum* sp. AJ67. Sin embargo, discrepancias como esta son comunes entre los análisis genómicos y los análisis basados en un único gen marcador (Paul et al., 2019).

Análisis de genes del metabolismo del arsénico en *Halorubrum* sp. DM2

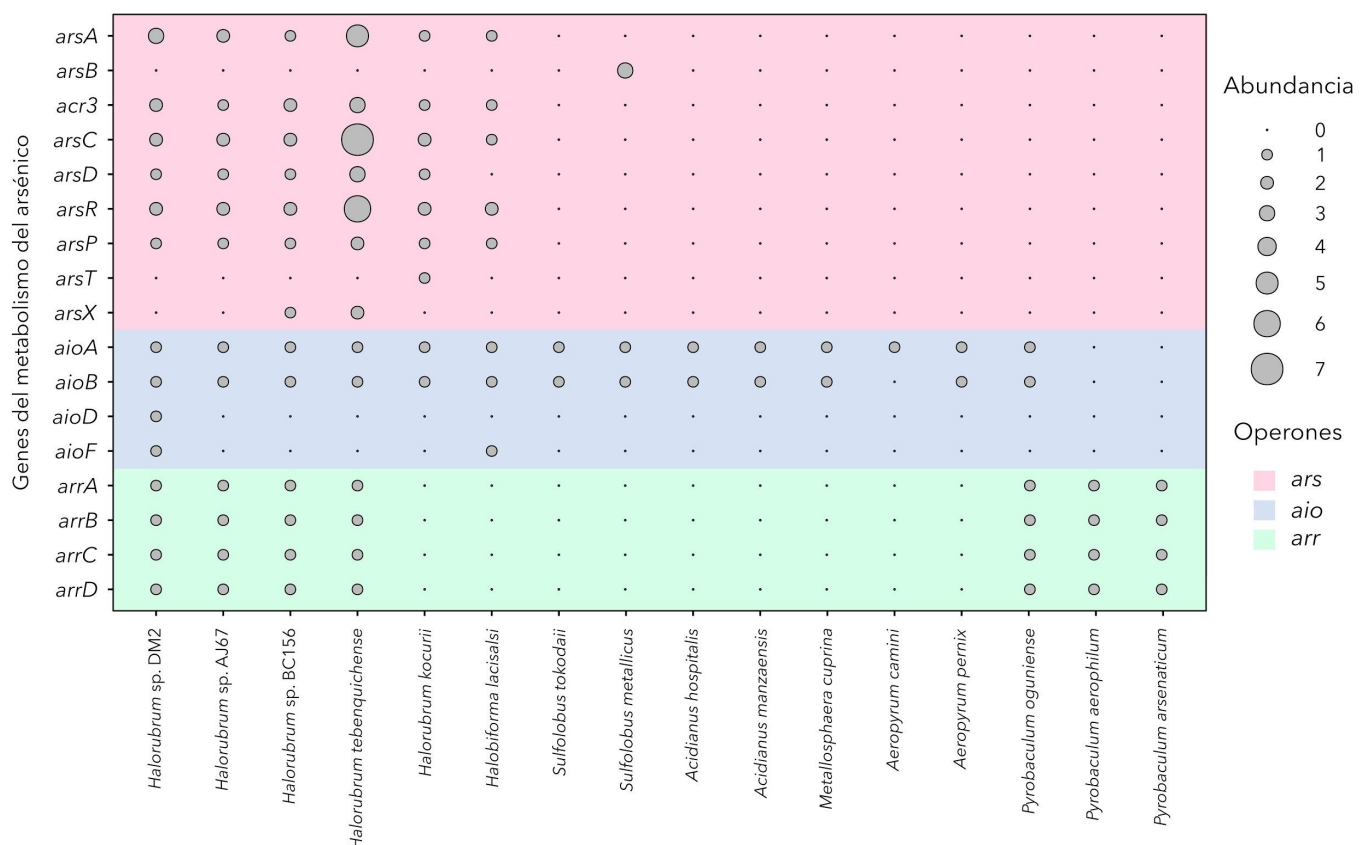


Figura 26. Abundancia de los genes del metabolismo del arsénico en *Halorubrum* sp. DM2 y otras 15 arqueas que presentan genes *aio* y/o *arr*.

Con el fin de comprender mejor el metabolismo del arsénico en el dominio *Archaea*, se llevó a cabo un profundo análisis de los genes implicados en la resistencia al arsénico y obtención de energía a partir de su oxidorreducción en *Halorubrum* sp. DM2 y otras 15 arqueas que presentan genes *aio* y/o *arr*. La identificación de dichos genes y la estimación de su abundancia en cada genoma (Figura 26) permitió determinar los mecanismos moleculares que estarían presentes en las arqueas estudiadas.

En los genomas de las haloarqueas (*Halorubrum* sp. DM2, *Halorubrum* sp. AJ67, *Halorubrum* sp. BC156, *Halorubrum tebenquichense*, *Halorubrum kocurii*, y *Halobiforma lacisalsi*) se identificaron numerosos genes *ars*. En base a la presencia y abundancia de dichos genes en los genomas, se pudo determinar que el principal mecanismo de resistencia al arsénico en las haloarqueas estudiadas consistiría en la reducción citoplasmática del arsenato, mediada por ArsC, seguido de la extrusión celular del arsenito resultante por la bomba Acr3. Como no se identificaron genes *arsB* en ninguna de las haloarqueas, la bomba ArsB no participaría en la extrusión celular del arsenito, y por ende, la ATPasa ArsA mejoraría la eficiencia de transporte de Acr3 y no de ArsB. Si bien se han identificado genes *arsP* en los genomas de las haloarqueas estudiadas, no se ha podido identificar ningún gen *arsM*. Con lo cual, es incierto si las haloarqueas estudiadas tienen la capacidad de metilar el arsenito, a través de una enzima arsenito metiltransferasa (ArsM), para luego expulsar el metilarsenito por la bomba ArsP. Si efectivamente presentan este mecanismo de detoxificación, lo más probable es que su enzima arsenito metiltransferasa no sea homóloga a la enzima arsenito metiltransferasa caracterizada en otros organismos. Tampoco se han identificado genes *arsH*, *arsI*, *arsN* o *arsO* en los genomas de las haloarqueas. Por ende, tanto la oxidación del metilarsenito a metilarsenato como la conversión del metilarsenito a arsenito no serían mecanismos de detoxificación utilizados por las haloarqueas estudiadas. Por último, cabe

mencionar que se han identificado varios genes *arsR* y *arsD* en todas las haloarqueas, lo que indicaría que los represores transcripcionales ArsR y ArsD regulan la transcripción de los genes *ars* en las haloarqueas estudiadas. En los genomas de las arqueas termófilas (*Sulfolobus tokodaii*, *Sulfolobus metallicus*, *Acidianus hospitalis*, *Acidianus manzaensis*, *Metallosphaera cuprina*, *Aeropyrum camini*, *Aeropyrum pernix*, *Pyrobaculum oguniense*, *Pyrobaculum aerophilum*, y *Pyrobaculum arsenaticum*) no se pudieron identificar genes *ars*. Si bien el genoma de *Sulfolobus metallicus* presenta 3 copias del gen *arsB*, la ausencia de otros genes *ars* pone en duda la funcionalidad de dichos genes *arsB*. Con lo cual, las arqueas termófilas estudiadas no presentarían ningún mecanismo conocido de detoxificación del arsénico.

Con respecto a los genes implicados en la obtención de energía a través de la oxidación del arsenito (genes *aio*), se han observado ciertas diferencias entre las haloarqueas y las arqueas termófilas que presentan genes *aioA* y *aioB*. Las arqueas termófilas (excluyendo a *Pyrobaculum aerophilum* y *Pyrobaculum arsenaticum*) únicamente presentan genes *aioA* y *aioB*. Mientras que ciertas haloarqueas (*Halorubrum* sp. DM2 y *Halobiforma lacisalsi*), además de presentar genes *aioA* y *aioB*, presentan genes *aioD* y/o *aioF*. Lo que tienen en común todas las arqueas que presentan genes *aioA* y *aioB*, es que ninguna tiene genes *aioC*, *aioX*, *aioS* o *aioR*. Por lo tanto, en base a estas observaciones, se puede decir que: (1) la transcripción de la arsenito oxidasa AioAB en las arqueas no estaría regulada por los factores de transcripción AioX, AioS y AioR; (2) únicamente en *Halorubrum* sp. DM2 y *Halobiforma lacisalsi*, la transcripción de la arsenito oxidasa AioAB estaría regulada por el factor de transcripción AioF; (3) como ninguna arquea presenta el citocromo AioC, se desconoce quién actúa como aceptor de electrones en la oxidación del arsenito, y abre la posibilidad de que el aceptor final de electrones sea uno alternativo al oxígeno (p.ej., nitrato o sulfato) (Y.-F.

Wu et al., 2021). Con respecto a los genes implicados en la respiración de arsenato (genes *arr*), se ha observado que todas las arqueas que presentan genes *arrA* y *arrB*, además presentan genes *arrC* y *arrD*. Con lo cual, el operón *arr* estaría completo en dichas arqueas. A pesar de que las arqueas estudiadas presentan genes *aio* y/o *arr*, no se ha demostrado aún si pueden obtener energía a través de la oxidación del arsenito o si pueden respirar arsenato en lugar de oxígeno. Únicamente se ha demostrado que *Pyrobaculum aerophilum* y *Pyrobaculum arsenaticum* pueden crecer en anaerobiosis respirando arsenato. Con lo cual, se deberán realizar diferentes ensayos de crecimiento (con y sin arsénico, en aerobiosis y anaerobiosis) para demostrar que el resto de las arqueas estudiadas pueden oxidar el arsenito y/o reducir el arsenato.

Al analizar la organización genómica de los genes del metabolismo del arsénico en las arqueas estudiadas (Figura 27), se pudieron observar ciertas diferencias entre las haloarqueas y las arqueas termófilas. En las haloarqueas, los genes *aio* se agrupan normalmente con genes *acr3*, lo cual tiene sentido ya que la bomba Acr3 está involucrada en el transporte del arsenito desde el citoplasma hasta el periplasma, donde se encuentra la arsenito oxidasa AioAB. A su vez, en las haloarqueas, los genes *aio* se encuentran agrupados con genes *aqp* y *fkp* que codifican acuaporinas y chaperonas, respectivamente. Con lo cual, es probable que dichas acuaporinas estén involucradas en el transporte celular del arsenito, y que las respectivas chaperonas colaboren en el plegado de AioAB. Por el contrario, en las arqueas termófilas, los genes *aio* se agrupan normalmente con genes *sox*, *qox*, y *cytB* que codifican sulfocianinas, quinol oxidasas, y citocromos b6, respectivamente. Por lo tanto, en las arqueas termófilas, es probable que el oxígeno sea el aceptor final de electrones en la oxidación del arsenito.

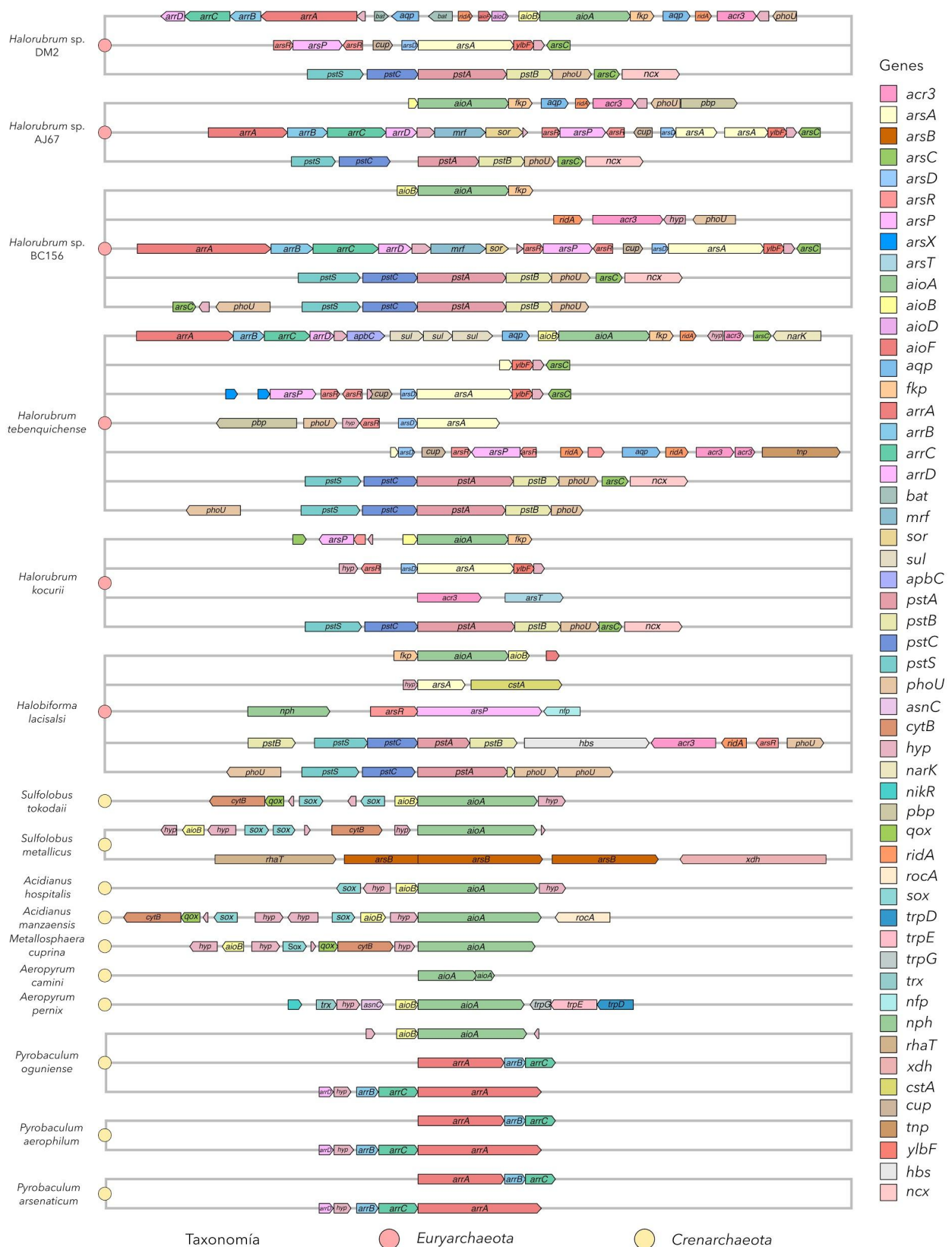


Figura 27. Organización de los genes *ars*, *aio*, y *arr* en *Halorubrum* sp. DM2 y otras 15 arqueas que presentan genes *aio* y/o *arr*.

En cuanto a los genes *arr*, no se podría decir si su organización es similar o no entre haloarqueas y arqueas termófilas ya que los genes *arr* podrían estar organizados de dos maneras diferentes en las arqueas termófilas *Pyrobaculum oguniense*, *Pyrobaculum aerophilum* y *Pyrobaculum arsenaticum*. Esto es debido a que aún no se sabe con exactitud cuales son los genes *arrA* y *arrB* responsables de la respiración del arsenato en dichas arqueas (J. Andres & Bertin, 2016). De todos modos, en el esquema de la Figura 27, se han representado las dos posibles organizaciones de los genes *arr* en *Pyrobaculum oguniense*, *Pyrobaculum aerophilum* y *Pyrobaculum arsenaticum*. Por último cabe mencionar que, en las haloarqueas, el gen *arsC* se encuentra agrupado con los genes *pstS*, *pstC*, *pstA* y *pstB*. Esto indicaría que el ingreso del arsenato a través de los transportadores Pst está íntimamente ligado a la reducción citoplasmática del arsenato por ArsC.

Ensayos de crecimiento de *Halorubrum* sp. DM2

A partir del análisis genómico de *Halorubrum* sp. DM2, se pudo determinar que la haloarquea aislada del *biofilm* rojo de la laguna Diamante presenta genes *aio* y *arr*. Por lo tanto, dicha haloarquea podría crecer como quimiolitótrofo oxidando arsenito y respirando oxígeno, y/o como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato. Con el objetivo de evaluar dichas capacidades metabólicas en *Halorubrum* sp. DM2, se llevaron a cabo diferentes ensayos de crecimiento.

En primer lugar se evaluó la capacidad de *Halorubrum* sp. DM2 de crecer como quimiolitóautótrofo oxidando arsenito y respirando oxígeno (Figura 28). Para ello se realizó un ensayo de crecimiento en aerobiosis, oscuridad y medio CDM modificado (se reemplazó el piruvato por bicarbonato de sodio 0,2 % m/v) (A) con arsenito 1 mM, (B) con arsenato 20 mM y (C) sin arsenito ni arsenato (control). Como no se observó crecimiento en ninguna de las condiciones evaluadas, se pudo concluir que *Halorubrum* sp. DM2 no es

capaz de crecer como quimiolitotótrofo oxidando arsenito y respirando oxígeno.

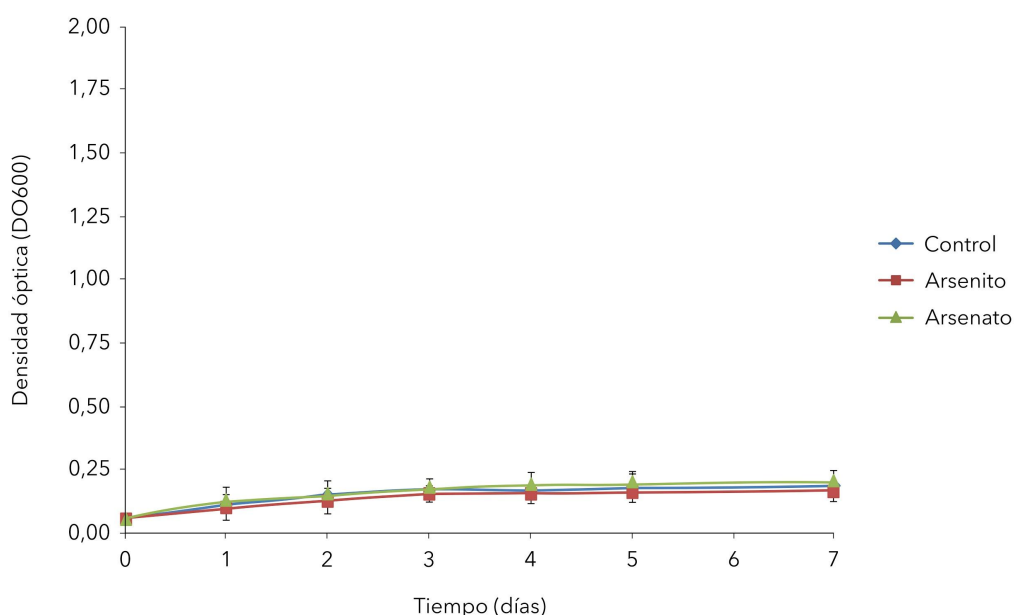


Figura 28. Curvas de crecimiento de *Halorubrum* sp. DM2 en aerobiosis, oscuridad y medio mínimo (CDM) bicarbonato con arsenito 1 mM, con arsenato 20 mM y sin arsenito ni arsenato (control).

En segundo lugar se evaluó la capacidad de *Halorubrum* sp. DM2 de crecer como quimiolitoheterótrofo oxidando el arsenito y respirando oxígeno (Figura 29 B). Para ello se realizó un ensayo de crecimiento en aerobiosis, oscuridad y medio CDM (A) con arsenito 1 mM, (B) con arsenato 20 mM y (C) sin arsenito ni arsenato (control). Como el crecimiento de la haloarquea sin arsénico (control) fue mayor que el crecimiento con arsénico (condiciones B y C), se pudo concluir que *Halorubrum* sp. DM2 no es capaz de crecer como quimiolitoheterótrofo oxidando el arsenito y respirando oxígeno. De todas formas, este resultado no descarta la posibilidad de que la haloarquea aislada oxide arsenito a través de la enzima arsenito oxidasa AioAB. Al igual que *Herminiimonas arsenicoxydans* (Muller et al., 2007), *Halorubrum* sp. DM2 podría oxidar el arsenito como mecanismo de detoxificación.

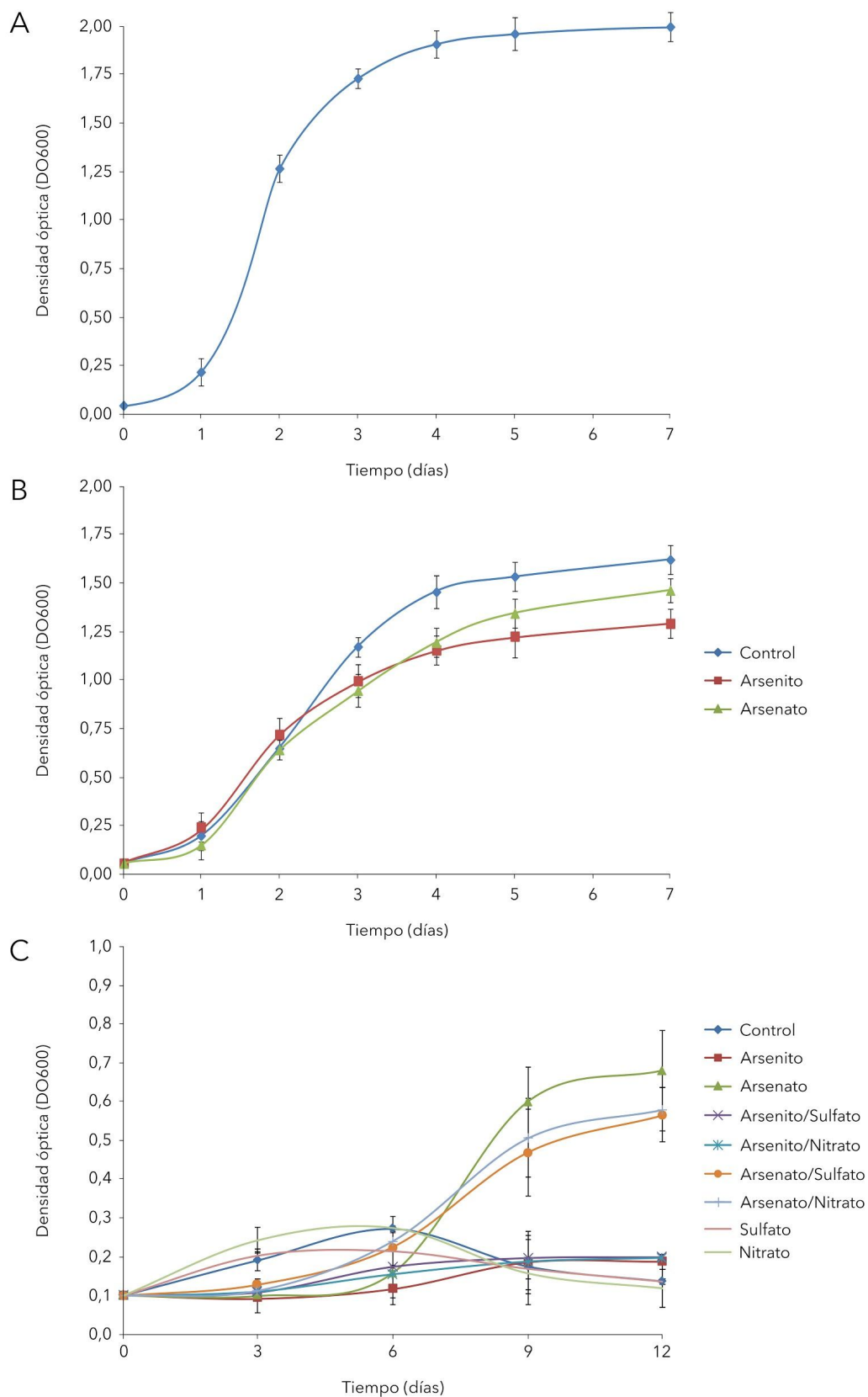


Figura 29. Curvas de crecimiento de *Halorubrum* sp. DM2. A, Crecimiento en aerobiosis, oscuridad y medio rico (WJK); B, Crecimiento en aerobiosis,

oscuridad y medio mínimo (CDM) con arsenito 1 mM, con arsenato 20 mM y sin arsenito ni arsenato (control); C, Crecimiento en anaerobiosis, oscuridad y medio mínimo (CDM) con arsenito 1 mM, con arsenito 1 mM y sulfato 20 mM, con arsenito 1 mM y nitrato 20 mM, con arsenato 20 mM, con arsenato 20 mM y sulfato 20 mM, con arsenato 20 mM y nitrato 20 mM, con sulfato 20 mM, con nitrato 20 mM, y sin arsenito, arsenato, sulfato ni nitrato (control).



Figura 30. *Halorubrum* sp. DM2 tiene la capacidad de crecer en anaerobiosis, oscuridad y medio mínimo (CDM) con arsenato 20 mM, con arsenato 20 mM y sulfato 20 mM, y con arsenato 20 mM y nitrato 20 mM.

Por último se evaluó la capacidad de *Halorubrum* sp. DM2 de crecer como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato (Figuras 29 C y 30). Para ello se realizó un ensayo de crecimiento en anaerobiosis, oscuridad y medio CDM (A) con arsenito 1 mM, (B) con arsenito 1 mM y sulfato 20 mM, (C) con arsenito 1 mM y nitrato 20 mM, (D) con arsenato 20 mM, (E) con arsenato 20 mM y sulfato 20 mM, (F) con arsenato 20 mM y nitrato 20 mM, (G) con sulfato 20 mM, (H) con nitrato 20 mM y (I) sin arsenito, arsenato, sulfato ni nitrato (control). Al observarse crecimiento de la haloarquea únicamente en las condiciones con arsenato (condiciones D, E, y F), se pudo concluir que *Halorubrum* sp. DM2 es capaz de crecer como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato. Ahora bien, para corroborar que *Halorubrum* sp. DM2 puede respirar arsenato utilizando la arsenato reductasa ArrAB, se procedió a obtener una mutante *Halorubrum* sp. DM2 $\Delta arrA$. Dicha mutante se obtuvo mediante numerosos pasajes de *Halorubrum* sp. DM2 en medio WJK en aerobiosis (37°C, 180 RPM) sin arsénico. Al eliminar la presión de selección (alta concentración de arsénico), se buscó que el gen *arrA* mutara y perdiera su función. Para identificar una mutante *Halorubrum* sp. DM2 $\Delta arrA$ en los sucesivos pasajes, se realizaron ensayos de crecimiento en anaerobiosis, oscuridad y medio mínimo (CDM) con arsenato 20 mM (12 días a 37°C). De esta manera, se identificó una posible mutante *Halorubrum* sp. DM2 $\Delta arrA$ capaz de crecer en aerobiosis en medio WJK o CDM, pero incapaz de crecer en anaerobiosis en medio CDM con arsenato 20 mM (Figura 31). Al analizar el genoma de la posible mutante se confirmó que el gen *arrA* se encuentra mutado, ya que presenta un codón de terminación prematuro (Figura A1 en Apéndice). Por lo tanto, como la mutante obtenida presenta el gen *arrA* mutado y no es capaz de crecer en anaerobiosis respirando arsenato, se puede concluir que *Halorubrum* sp. DM2 utiliza la enzima arsenato reductasa ArrAB para respirar arsenato.

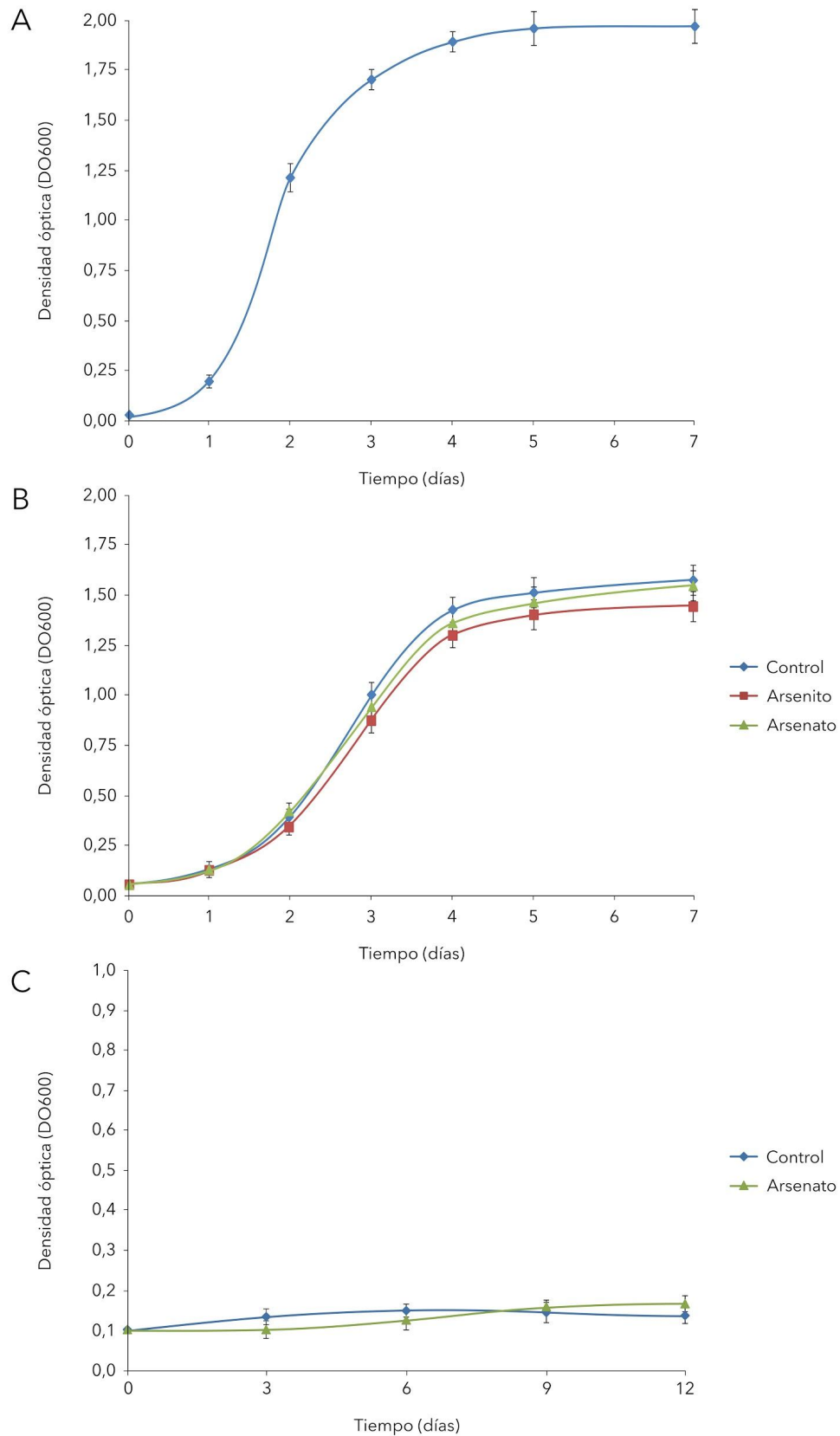


Figura 31. Curvas de crecimiento de *Halorubrum* sp. DM2 $\Delta arrA$. A, Crecimiento en aerobiosis, oscuridad y medio rico (WJK); B, Crecimiento en

aerobiosis, oscuridad y medio mínimo (CDM) con arsenito 1 mM, con arsenato 20 mM y sin arsenito ni arsenato (control); C, Crecimiento en anaerobiosis, oscuridad y medio mínimo (CDM) con arsenito 1 mM, con arsenito 1 mM y sulfato 20 mM, con arsenito 1 mM y nitrato 20 mM, con arsenato 20 mM, con arsenato 20 mM y sulfato 20 mM, con arsenato 20 mM y nitrato 20 mM, con sulfato 20 mM, con nitrato 20 mM, y sin arsenito, arsenato, sulfato ni nitrato (control).

CONCLUSIONES

A partir de los resultados obtenidos en la segunda parte de la tesis, se puede concluir que:

- I. De todas las haloarqueas aisladas del *biofilm* rojo de la laguna Diamante, *Halorubrum* sp. DM2 es la que presenta la mayor resistencia al arsénico.
- II. La capacidad de *Halorubrum* sp. DM2 de resistir altas concentraciones de arsénico se debe en parte a que presenta genes *ars* implicados en la detoxificación del arsénico.
- III. El principal mecanismo de detoxificación del arsénico en *Halorubrum* sp. DM2 consistiría en la reducción citoplasmática del arsenato, mediada por ArsC, seguido de la expulsión del arsenito resultante por la bomba Acr3. Este mecanismo de detoxificación del arsénico también sería utilizado por otras haloarqueas que presentan genes *aio* y/o *arr*.
- IV. *Halorubrum* sp. DM2, además de resistir al arsénico, tiene la capacidad de utilizar el arsenato como aceptor final de electrones en la cadena respiratoria. De esta manera, puede crecer como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato.
- V. *Halorubrum* sp. DM2 utiliza la enzima arsenato reductasa ArrAB para respirar arsenato.

CONCLUSIONES FINALES Y PERSPECTIVAS FUTURAS

Los salares y volcanes de los Andes centrales presentan lagunas alcalinas e hipersalinas con elevadas concentraciones de arsénico de origen volcánico. Por lo tanto, los ecosistemas microbianos que se desarrollan en dichas lagunas (*biofilms*, tapetes microbianos, y microbialitos, entre otros) están compuestos por microorganismos que presentan diversos mecanismos moleculares de resistencia al arsénico. Los principales mecanismos de resistencia al arsénico en dichos microorganismos consistirían en: (1) la reducción citoplasmática del arsenato, mediada por ArsC, seguido de la expulsión del arsenito resultante por la bomba Acr3, y (2) la metilación citoplasmática del arsenito, mediada por ArsM, seguido de la expulsión del metilarsenito resultante por la bomba ArsP. Eventualmente, el metilarsenito podría ser oxidado a metilarsenato por ArsH. Algunos microorganismos de los Andes centrales, además de resistir al arsénico, tendrían la capacidad de obtener energía metabólica a partir de su oxidorreducción. Ciertos microorganismos podrían oxidar el arsenito, utilizando las enzimas AioAB o ArxAB, para obtener energía metabólica. Mientras que otros podrían reducir el arsenato, utilizando la enzima ArrAB, para crecer en anaerobiosis. Dicho es el caso de la haloarquea *Halorubrum* sp. DM2 aislada del *biofilm* rojo de la laguna Diamante. Al probarse que dicha haloarquea puede crecer como quimioorganótrofo oxidando compuestos orgánicos y respirando arsenato, se demostró un nuevo metabolismo en el filo *Euryarchaeota*. De todas formas, para terminar de comprender el metabolismo del arsénico en *Halorubrum* sp. DM2, se deberá realizar una caracterización bioquímica de las enzimas AioAB y ArrAB. Para ello, se deberán expresar dichas enzimas en un organismo heterólogo, y purificar utilizando diferentes técnicas cromatográficas. Luego, con las enzimas purificadas, se podrá evaluar su actividad. Si se prueba que la enzima AioAB oxida el arsenito, no solo se demostraría un nuevo

metabolismo en el filo *Euryarchaeota*, sino un nuevo metabolismo en el dominio *Archaea*.

PUBLICACIONES

Publicaciones relacionadas con este trabajo de tesis doctoral:

- **Vignale, F.A.**, Lencina, A.I., Stepanenko, T.M., Soria, M.N., Saona, L.A., Kurth, D., Guzmán, D., Foster, J., Poiré, D.G., Villafañe, P.G., Albarracín, V.H., Contreras M. & Farías, M.E. (2021). Lithifying and Non-Lithifying Microbial Ecosystems in the Wetlands and Salt Flats of the Central Andes. *Microbial Ecology*, 83, 1–17. DOI:10.1007/s00248-021-01725-8
- **Vignale, F.A.**, Kurth, D., Lencina, A.I., Poiré, D.G., Chihuailaf, E., Muñoz-Herrera, N., Novoa, F., Contreras, M., Turjanski, G.A., Farías, M.E. (2021). Geobiology of Andean Microbial Ecosystems Discovered in Salar de Atacama, Chile. *Frontiers in Microbiology*, 12, 762076. DOI: 10.3389/fmicb.2021.762076
- Carrizo, D., **Vignale, F.A.**, Sánchez-García, L., Farías, M.E. (2022). Ecological variability based on lipid biomarkers in astrobiologically interesting wetlands from the Argentinian central Andes. *FEMS Microbiology Ecology*, 98(5). DOI: 10.1093/femsec/fiac049

Publicaciones no relacionadas con este trabajo de tesis doctoral:

- **Vignale, F.A.**, Bernal, R.D., Pardo, A.M., Almasqué, F.J., Ibarra, J.G., Fernández Do Porto, D., Turjanski, A.G., López, N.I., Menendez Helman, R.J., Raiger Iustman, L.J. (2022). Spatial and Seasonal Variations in the Bacterial Community of an Anthropogenic Impacted Urban Stream. *Microbial Ecology*. DOI: 10.1007/s00248-022-02055-z

- Barone, F., Dorr, F., Marasco, L.E., Mildiner, S., Patop, I.L., Sosa, S., Vattino L.G., **Vignale, F.A.**, Altszyler, E., Basanta, B., Carlotto, N., Gasulla, J., Giménez, M., Grande, A., Nieto Moreno, N., Bonomi, H.R., Nadra, A.D. (2017). Design and evaluation of an incoherent feed-forward loop for an arsenic biosensor based on standard iGEM parts. *Synthetic Biology*, 2(1), ysx006. DOI: 10.1093/synbio/ysx006

AGRADECIMIENTOS

La manera en que los científicos nos comunicamos es a través de documentos cortos, concisos y objetivos, para que la transmisión de la información sea eficiente. Pero esta forma de comunicarnos deja a un lado un aspecto importante de la actividad científica, aquel relacionado con la subjetividad y los sentimientos. Eso no suele ser considerado relevante, o al menos, no existen ámbitos donde se puedan expresar esas cosas. Ahora bien, ¿no son nuestros sentimientos los que nos motivan para hacer ciencia? Afortunadamente, la tesis de Doctorado, además de ser un documento científico es un espacio donde uno puede compartir lo que siente y agradecer a aquellas personas e instituciones que, de una forma u otra, formaron parte del camino.

Así que, aquí vamos!

A mi vieja, la persona más importante de mi vida. No sería quien soy de no haber sido por ella. Siempre estuvo a mi lado, cuidándome, enseñándome, dándome lo poco que tenía para que no me faltara nada. Mucha gente dice que nos parecemos, y eso me llena de orgullo, porque no conozco a nadie que tenga la fortaleza que ella tuvo para criar a dos hijos sola de la manera admirable en que lo hizo. Ma, siempre voy a estar agradecido por lo que me diste y por lo que me das. Pero sobre todo, porque no me dejaste solo en el momento que más te necesitaba. No me diste la espalda, al contrario, me abrazaste y me dijiste que siempre me ibas a amar y aceptar como soy. Yo también te voy a amar por siempre, y algún día espero que estés tan orgullosa de mí como yo estoy de vos.

A mi viejo, que lo amo y lo extraño todos los días. Solo pude compartir trece años de mi vida con él, pero aún hoy recuerdo su forma de ser, alegre y divertida. Gracias a él comprendí que la felicidad es una forma de ver la vida, una actitud, y no una meta. Hay que hacer valer cada día, preocuparse menos por lo que dirán, compararse menos con otros, y hacer lo que a uno lo haga feliz. Porque al final de todo, cuando ya no haya más días por vivir, lo único que nos quedará son nuestros recuerdos.

A mi hermano Nico, una de las personas más buenas y honestas que conozco. Al igual que mi vieja, siempre estuvo a mi lado cuidándome, enseñándome, apoyándome en todos mis proyectos y siendo mi modelo a seguir. Estoy tan orgulloso de tener un hermano como él, porque no solo es mi hermano, es mi mejor amigo. Vivimos la mayor parte de nuestras vidas juntos y espero que sigamos unidos por siempre.

A mis abuelos, tito e tita. Su amor fue siempre incondicional, y gracias a ellos tuve una infancia hermosa. Mi abuela era la artista de la familia, y con ella aprendí a dibujar y pintar. Ella me enseñó la importancia de los detalles, tanto en el arte como en la vida en general, y por ello estoy muy agradecido. Hoy en día, las figuras e ilustraciones científicas que realizo no serían iguales sin sus enseñanzas. Mi abuelo, por otro lado, era el intelectual de la familia. Siempre admiré su forma de trabajar, prolija y ordenada, la cual me transmitió con su ejemplo, y por ello también estoy muy agradecido. Porque hoy en día, probablemente no tendría la capacidad de organizar proyectos científicos y llevarlos adelante de manera ordenada, de no haber seguido su ejemplo.

A mis primos, Mati y Fede. Ellos perdieron a su madre, mi tía, demasiado pronto en la vida, pero igual pudieron salir adelante y ser dos personas

ejemplares. Admiro su fortaleza y les agradezco por haberme apoyado siempre en mis proyectos y en mis sueños. Sin importar a dónde nos lleven nuestros caminos, siempre estaremos juntos.

A mi directora de tesis, María Eugenia Farías. Mariu es una persona única y su personalidad la llevó a realizar grandes cosas. Al mismo tiempo, y sin la ayuda de nadie, llevó adelante un grupo de laboratorio, crió a sus tres hijos, y defendió los ambientes de los Andes centrales de los intereses privados que los ponían en peligro. Sin lugar a duda su trabajo fue admirable, y me enorgullece haber podido ser parte de ello. Toda la vida soñé con explorar, descubrir lugares nuevos e investigarlos. Gracias a ella pude cumplir mi sueño, conocí y estudié ambientes únicos en el mundo. En el futuro espero continuar su legado, y haré lo posible para conservar y estudiar los ambientes andinos de la manera admirable en que lo hizo Mariu.

A mi director de tesis, Adrián Turjanski. Adri fue una pieza fundamental en mi formación como científico. Además de guiarme y aconsejarme en el camino de la bioinformática, me enseñó a llevar adelante una vida académica. Él era mi cable a tierra, con él aprendí a focalizarme en lo importante y a no preocuparme por cuestiones irrelevantes. Estoy seguro que muchas de sus enseñanzas me serán útiles en mi futuro académico, y se lo agradezco enormemente.

To my supervisor in EMBL-EBI, Rob Finn. He taught me most of the things I know about metagenomics, and guided me on my first steps in science. Along with Alex Mitchell and Lorna Richardson, he has made my research secondment at EMBL-EBI an experience that I will never forget.

A mi supervisora en EMBL-Hamburg, María García-Alai. María confió en mí desde el momento en que nos conocimos, y gracias a esa confianza que tuvo en mí, pude llevar adelante una hermosa y fructífera estadía de investigación en EMBL-Hamburg. De todos modos, mi camino académico con María recién comienza, y estoy seguro de que juntos vamos a realizar grandes descubrimientos.

A mi amigo y compañero de laboratorio, Lucas Defelipe. Sin lugar a duda, la realización de esta tesis doctoral no habría sido posible sin su ayuda. Lucas debe ser la persona más generosa y trabajadora con la que me crucé en la vida. Él es de esas personas que saben todo y tienen una solución para todo. Cuando necesitaba ayuda para instalar o correr un programa, él me lo solucionaba; cuando necesitaba ayuda para resolver una cuestión burocrática de la universidad, él me lo solucionaba; cuando necesitaba llorar, reír, o simplemente despejarme, él estaba ahí. Por ende, no me va a alcanzar la vida para devolverle todo los favores que me hizo, pero al menos quiero que sepa que estoy muy agradecido y feliz por haber compartido cinco años de momentos inolvidables junto a él. Y espero que nuestros caminos sigan juntos por muchos años más.

A mi amigo y compañero de laboratorio, Osvaldo Burastero. Con Osva compartí mis seis años de carrera de grado y mis cinco años de carrera de doctorado. Prácticamente es un hermano, lo conozco como a nadie y estoy agradecido a la vida por haberme cruzado con él. Osva debe ser la única persona que tiene tiempo para estudiar chino, alemán, practicar *functional*, Jiu-Jitsu, yoga, y además hacer ciencia. Su forma de administrar el tiempo es envidiable, pero más envidiable es su forma de ser, amena y calma. A lo largo

del doctorado estuvo a mi lado para ayudarme, aconsejarme o simplemente compartir lindos momentos, y le agradezco enormemente eso.

A mis amigos y compañeros de laboratorio en IQUIBICEN, Juan, Elias, Mode, Chelo, Cata, JB, Cordo, Sol, Jordi, Fede, Franco, Eze, Claudio y Maria. Hacer ciencia a su lado no podría haber sido más agradable y divertido, y les agradezco por todos los momentos que compartimos juntos.

A mis amigos y compañeros de laboratorio en PROIMI, Tati, Maru, Agus, Lucho, Vir, Pato, Omar, Flor y Gretel. Disfruté enormemente las campañas científicas a la Puna que realicé junto a ellos, y les agradezco por todo lo que me enseñaron y ayudaron durante estos años de doctorado.

To my friends and colleagues in EMBL-Hamburg, Javi, George, Katha, Stephan, Angélica, David, Janina, Niklas, Lea, Vir, Rita, Florian, Max, Katharina, Lenka, Tania, Gabi, Panos, Iuliia and Claudia. They have been very kind to me and have helped me whenever I needed it during my stay in Germany. I will always be grateful to them.

A mis amigas Maru, Juli Kessler, Euge y Juli Martínez. Somos amigas desde el momento en que nos conocimos en el colegio y nunca más nos separamos. Son parte de mi familia, las quiero como si fueran mis hermanas, y les agradezco por haberme apoyado y acompañado siempre. Toda mi vida estaré agradecido con ellas por aceptarme y quererme como soy, y estoy seguro de que nuestros caminos seguirán unidos por siempre.

A mis amigas Guilli, Mari y Cata, las churris. Nos conocimos en el CBC, y desde aquel momento estamos unidos. Ellas hicieron que mis días de

cursada en la Facultad fueran inolvidables. Tengo numerosos recuerdos y anécdotas con ellas que nunca olvidaré. Toda la vida estaré agradecido por el cariño y la ayuda que me dieron por años, y estoy seguro que permaneceremos juntos por siempre.

A mis amigos Bruno, Mati y Lupi. Sin ellos, mis días en la Facultad hubieran sido muy diferentes. Ellos me han acompañado, ayudado y apoyado en todos mis proyectos y siempre estaré agradecido por ello.

A mis amigas Caro y Lau. Con ellas compartí momentos en la Facultad inolvidables, y les agradezco por haberme acompañado y ayudado siempre que lo necesité.

A Marcela y José. En mis días de cursada en la Facultad se convirtieron en mis padres postizos porque me brindaban ayuda o me guiaban en todo lo que necesitaba. Nunca voy a olvidar nuestros días en el laboratorio, tomando mates, riéndonos, y pensando en posibles proyectos. Les agradezco por todo lo que me dieron, siempre los voy a recordar con una sonrisa.

A mi hermana postiza Cami y a mi madre postiza Sil. Con ellas compartí todos los veranos de mi infancia y adolescencia, de los cuales tengo recuerdos hermosos. Nunca olvidaré toda la ayuda que me brindaron, y estoy seguro de que nuestras vidas continuarán entrelazadas por siempre.

A mis amigos Luqui y Ceci. Son mis amigos de la vida, nos conocemos desde el jardín de infantes. Compartimos la infancia, la adolescencia y ahora, la

adultez. Siempre estuvieron a mi lado, en los buenos y en los malos momentos, y eso nunca lo olvidaré.

Al Instituto San Román, y en especial, a Teresa, Sergio, José y Pancho. Gracias a ellos tuve una extraordinaria educación primaria y secundaria. Me inculcaron valores humanos que nunca olvidaré y estoy orgulloso de haber podido formar parte de la familia sanromanense.

A la Universidad de Buenos Aires, y en especial, a la Facultad de Ciencias Exactas y Naturales. En la Facultad pasé los momentos más lindos de mi vida. Es un lugar donde siempre me encontré cómodo, feliz, donde pude ser quien era porque sabía que no me iban a juzgar o criticar, donde pude aprender, charlar con gente de temas que me apasionaban, y sobre todo hacer ciencia, que es lo que más me gusta. Por siempre estaré agradecido a la Facultad por haberme educado y brindado un segundo hogar.

Al CONICET, por haber financiado mi carrera de doctorado. Estoy muy orgulloso de haber formado parte de esta institución que lleva adelante la ciencia argentina. Lamentablemente, en Argentina, la ciencia no es una política de Estado, y por esta razón ha sido olvidada y dejada de lado en numerosas ocasiones. Espero y deseo que algún día el pueblo argentino se dé cuenta de que un país sin ciencia no progresa, y que el dinero destinado a educación no es un gasto, sino una inversión.

BIBLIOGRAFÍA

- Achour-Rokbani, A., Cordi, A., Poupin, P., Bauda, P., & Billard, P. (2010). Characterization of the *ars* gene cluster from extremely arsenic-resistant *Microbacterium* sp. strain A33. *Applied and Environmental Microbiology*, 76(3), 948–955.
- Aitken, J. D. (1967). Classification and environmental significance of cryptalgal limestones and dolomites, with illustrations from the Cambrian and Ordovician of southwestern Alberta. *Journal of Sedimentary Research*, 37(4), 1163–1178.
- Ajees, A. A., & Rosen, B. P. (2015). As(III) S-adenosylmethionine methyltransferases and other arsenic binding proteins. *Geomicrobiology Journal*, 32(7), 570–576.
- Albarracín, V. H., Kurth, D., Ordoñez, O. F., Belfiore, C., Luccini, E., Salum, G. M., Piacentini, R. D., & Farías, M. E. (2015). High-Up: A Remote Reservoir of Microbial Extremophiles in Central Andean Wetlands. *Frontiers in Microbiology*, 6, 1404.
- Allers, T., Ngo, H.-P., Mevarech, M., & Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Applied and Environmental Microbiology*, 70(2), 943–953.
- Allmendinger, R. W., Jordan, T. E., Kay, S. M., & Isacks, B. L. (1997). The evolution of the Altiplano-Puna plateau of the Central Andes. *Annual Review of Earth and Planetary Sciences*, 25(1), 139–174.

- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Loman, N. J., Andersson, A. F., & Quince, C. (2013). *CONCOCT: Clustering cONtigs on COverage and ComposiTiOn*. <https://doi.org/10.48550/ARXIV.1312.4038>
- Alonso, H., & Risacher, F. (1996). Geoquímica del Salar de Atacama, parte 1: origen de los componentes y balance salino. *Andean Geology*, 23(2), 113–122.
- Alonso, R. N., & Rojas, W. (2020). Origin and Evolution of the Central Andes: Deserts, Salars, Lakes, and Volcanoes. In M. E. Farías (Ed.), *Microbial Ecosystems in Central Andes Extreme Environments: Biofilms, Microbial Mats, Microbialites and Endoevaporites* (pp. 3–19). Springer International Publishing.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.
- Andres, J., & Bertin, P. N. (2016). The microbial genomics of arsenic. *FEMS Microbiology Reviews*, 40(2), 299–322.
- Andres, M. S., & Pamela Reid, R. (2006). Growth morphologies of modern marine stromatolites: A case study from Highborne Cay, Bahamas. *Sedimentary Geology*, 185(3-4), 319–328.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., ... Zagnitko, O. (2008). The RAST Server: rapid annotations using

- subsystems technology. *BMC Genomics*, 9, 75.
- Babilonia, J., Conesa, A., Casaburi, G., Pereira, C., Louyakis, A. S., Reid, R. P., & Foster, J. S. (2018). Comparative Metagenomics Provides Insight Into the Ecosystem Functioning of the Shark Bay Stromatolites, Western Australia. *Frontiers in Microbiology*, 9, 1359.
- Baliton, R., Landicho, L., Cabahug, R. E., Paelmo, R. F., Laruan, K., Rodriguez, R., Visco, R. G., & Castillo, A. K. A. (2020). Ecological services of agroforestry systems in selected upland farming communities in the Philippines. *Biodiversitas Journal of Biological Diversity*, 21(2). <https://doi.org/10.13057/biodiv/d210237>
- 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: A Journal of Computational Molecular Cell Biology*, 19(5), 455–477.
- Bar-Even, A., Noor, E., & Milo, R. (2012). A survey of carbon fixation pathways through a quantitative lens. *Journal of Experimental Botany*, 63(6), 2325–2342.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C., & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Research*, 32(Database issue), D138–D141.
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D.,

- & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, 46(D1), D41–D47.
- Berg, I. A. (2011). Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Applied and Environmental Microbiology*, 77(6), 1925–1936.
- Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-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.
- Bitan-Banin, G., Ortenberg, R., & Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *Journal of Bacteriology*, 185(3), 772–778.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Research*, 31(1), 365–370.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Braga, J. C., Martin, J. M., & Riding, R. (1995). Controls on microbial dome fabric development along a carbonate-siliciclastic shelf-basin transect, Miocene, SE Spain. *Palaios*, 10(4), 347.

- Brocchieri, L. (2014). The GC content of bacterial genomes. *Journal of Phylogenetics & Evolutionary Biology*, 02(01). <https://doi.org/10.4172/2329-9002.1000e108>
- Bröer, S., Ji, G., Bröer, A., & Silver, S. (1993). Arsenic efflux governed by the arsenic resistance determinant of *Staphylococcus aureus* plasmid pI258. *Journal of Bacteriology*, 175(11), 3480–3485.
- Bundschuh, J., & Maity, J. P. (2015). Geothermal arsenic: Occurrence, mobility and environmental implications. *Renewable & Sustainable Energy Rev.*, 42, 1214–1222.
- Burguener, G. F., Maldonado, M. J., Revale, S., Fernández Do Porto, D., Rascován, N., Vázquez, M., Farías, M. E., Marti, M. A., & Turjanski, A. G. (2014). Draft Genome Sequence of the Polyextremophilic *Halorubrum* sp. Strain AJ67, Isolated from Hyperarsenic Lakes in the Argentinian Puna. *Genome Announcements*, 2(1). <https://doi.org/10.1128/genomeA.01096-13>
- Burne, R. V., & Johnson, K. (2012). Sea-level variation and the zonation of microbialites in Hamelin Pool, Shark Bay, Western Australia. *Marine & Freshwater Research*, 63(11), 994.
- Burne, R. V., & Moore, L. S. (1987). Microbialites; organosedimentary deposits of benthic microbial communities. *Palaios*, 2(3), 241–254.
- Burne, R. V., & Moore, L. S. (1993). Microatoll microbialites of Lake Clifton, Western Australia: Morphological analogues of *Cryptozoön proliferum* Hall, the first formally-named stromatolite. *Facies*, 29(1), 149–168.
- Burns, B. P., Goh, F., Allen, M., & Neilan, B. A. (2004). Microbial diversity of extant

- stromatolites in the hypersaline marine environment of Shark Bay, Australia. In *Environmental Microbiology* (Vol. 6, Issue 10, pp. 1096–1101).
<https://doi.org/10.1111/j.1462-2920.2004.00651.x>
- Cabrol, N. (2001). The evolution of lacustrine environments on mars: Is mars only hydrologically dormant? *Icarus*, 149(2), 291–328.
- Cardenas, E., & Tiedje, J. M. (2008). New tools for discovering and characterizing microbial diversity. *Current Opinion in Biotechnology*, 19(6), 544–549.
- Carrizo, D., Vignale, F. A., Sánchez-García, L., & Fariás, M. E. (2022). Ecological variability based on lipid biomarkers in astrobiologically interesting wetlands from the Argentinian central Andes. *FEMS Microbiology Ecology*, 98(5).
<https://doi.org/10.1093/femsec/fiac049>
- Chen, J., Bhattacharjee, H., & Rosen, B. P. (2015). ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. *Molecular Microbiology*, 96(5), 1042–1052.
- Chen, J., Galván, A. E., Nadar, V. S., Yoshinaga, M., & Rosen, B. P. (2022). An ArsRC fusion protein enhances arsenate sensing and detoxification. *Environmental Microbiology*, 24(4), 1977–1987.
- Chen, J., Madegowda, M., Bhattacharjee, H., & Rosen, B. P. (2015). ArsP: a methylarsenite efflux permease. *Molecular Microbiology*, 98(4), 625–635.
- Chen, K.-T., Shen, H.-T., & Lu, C. L. (2018). Multi-CSAR: a multiple reference-based contig scaffolder using algebraic rearrangements. *BMC Systems Biology*,

12(Suppl 9), 139.

Chen, Y., & Rosen, B. P. (1997). Metalloregulatory properties of the ArsD repressor.

The Journal of Biological Chemistry, 272(22), 14257–14262.

Crabtree, J., Agrawal, S., Mahurkar, A., Myers, G. S., Rasko, D. A., & White, O.

(2014). Circleator: flexible circular visualization of genome-associated data with

BioPerl and SVG. *Bioinformatics*, 30(21), 3125–3127.

dC Rubin, S. S., Marín, I., Gómez, M. J., Morales, E. A., Zekker, I., San Martín-Uriz,

P., Rodríguez, N., & Amils, R. (2017). Prokaryotic diversity and community composition in the Salar de Uyuni, a large scale, chaotropic salt flat.

Environmental Microbiology, 19(9), 3745–3754.

Demergasso, C., Chong, G., Galleguillos, P., Escudero, L., MARTÍNEZ-ALONSO,

M., & Esteve, I. (2003). Tapetes microbianos del Salar de Llamará, norte de Chile. *Revista Chilena de Historia Natural*, 76(3).

<https://doi.org/10.4067/s0716-078x2003000300012>

Demergasso, C., Dorador, C., Meneses, D., Blamey, J., Cabrol, N., Escudero, L., &

Chong, G. (2010). Prokaryotic diversity pattern in high-altitude ecosystems of the Chilean Altiplano. In *Journal of Geophysical Research: Biogeosciences* (Vol.

115, Issue G2). <https://doi.org/10.1029/2008jg000836>

Dorador, C., Busekow, A., Vila, I., Imhoff, J. F., & Witzel, K.-P. (2008). Molecular

analysis of enrichment cultures of ammonia oxidizers from the Salar de Huasco, a high altitude saline wetland in northern Chile. *Extremophiles: Life under*

Extreme Conditions, 12(3), 405–414.

- Dorador, C., Vila, I., Witzel, K.-P., & Imhoff, J. F. (2013). Bacterial and archaeal diversity in high altitude wetlands of the Chilean Altiplano. *Fundamental and Applied Limnology: Official Journal of the International Association of Theoretical and Applied Limnology*, 182(2), 135–159.
- Dupraz, C., Visscher, P. T., Baumgartner, L. K., & Reid, R. P. (2004). Microbe-mineral interactions: early carbonate precipitation in a hypersaline lake (Eleuthera Island, Bahamas). *Sedimentology*, 51(4), 745–765.
- Farías, M. E. (Ed.). (2020). *Microbial Ecosystems in Central Andes Extreme Environments: Biofilms, Microbial Mats, Microbialites and Endoevaporites*. Springer, Cham.
- Farías, M. E., Contreras, M., Rasuk, M. C., Kurth, D., Flores, M. R., Poiré, D. G., Novoa, F., & Visscher, P. T. (2014). Characterization of bacterial diversity associated with microbial mats, gypsum evaporites and carbonate microbialites in thalassic wetlands: Tebenquiche and La Brava, Salar de Atacama, Chile. *Extremophiles: Life under Extreme Conditions*, 18(2), 311–329.
- Farías, M. E., Fernández-Zenoff, V., Flores, R., Ordóñez, O., & Estévez, C. (2009). Impact of solar radiation on bacterioplankton in Laguna Vilama, a hypersaline Andean lake (4650 m): SOLAR RADIATION IN BACTERIOPLANKTON. *Journal of Geophysical Research*, 114(G2).
<https://doi.org/10.1029/2008jg000784>
- Farías, M. E., Rascovan, N., Toneatti, D. M., Albarracín, V. H., Flores, M. R., Poiré, D. G., Collavino, M. M., Aguilar, O. M., Vazquez, M. P., & Polerecky, L. (2013).

- The discovery of stromatolites developing at 3570 m above sea level in a high-altitude volcanic lake Socompa, Argentinean Andes. *PloS One*, 8(1), e53497.
- Farias, M. E., Rasuk, M. C., Gallagher, K. L., Contreras, M., Kurth, D., Fernandez, A. B., Poiré, D., Novoa, F., & Visscher, P. T. (2017). Prokaryotic diversity and biogeochemical characteristics of benthic microbial ecosystems at La Brava, a hypersaline lake at Salar de Atacama, Chile. *PloS One*, 12(11), e0186867.
- Feldmann, M., & McKenzie, J. A. (1998). Stromatolite-thrombolite associations in a modern environment, lee stocking island, Bahamas. *Palaaios*, 13(2), 201.
- Fernandez, A. B., Rasuk, M. C., Visscher, P. T., Contreras, M., Novoa, F., Poire, D. G., Patterson, M. M., Ventosa, A., & Farias, M. E. (2016). Microbial Diversity in Sediment Ecosystems (Evaporites Domes, Microbial Mats, and Crusts) of Hypersaline Laguna Tebenquiche, Salar de Atacama, Chile. *Frontiers in Microbiology*, 7, 1284.
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(Web Server issue), W29–W37.
- Fleming, E. D., & Prufert-Bebout, L. (2010). Characterization of cyanobacterial communities from high-elevation lakes in the Bolivian Andes: HIGH-ELEVATION ANDES CYANOBACTERIA. *Journal of Geophysical Research*, 115(G2). <https://doi.org/10.1029/2008jg000817>
- Foster, J. S., Green, S. J., Ahrendt, S. R., Golubic, S., Reid, R. P., Hetherington, K. L.,

- & Bebout, L. (2009). Molecular and morphological characterization of cyanobacterial diversity in the stromatolites of Highborne Cay, Bahamas. *The ISME Journal*, 3(5), 573–587.
- Gadd, G. M. (2010). Metals, minerals and microbes: geomicrobiology and bioremediation. *Microbiology*, 156(Pt 3), 609–643.
- Gansser, A. (1973). Facts and theories on the Andes. *Journal of the Geological Society*, 129(2), 93–131.
- Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biology*, 12, 69.
- Gladysheva, T. B., Oden, K. L., & Rosen, B. P. (1994). Properties of the arsenate reductase of plasmid R773. *Biochemistry*, 33(23), 7288–7293.
- Grey, K., & Awramik, S. (2020). *Handbook for the study and description of microbialites*. Geological Survey of Western Australia Bulletin 147.
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075.
- Haferburg, G., Gröning, J. A. D., Schmidt, N., Kummer, N.-A., Erquicia, J. C., & Schlömann, M. (2017). Microbial diversity of the hypersaline and lithium-rich Salar de Uyuni, Bolivia. *Microbiological Research*, 199, 19–28.
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., ... Pruitt, K. D. (2018). RefSeq: an update on prokaryotic genome

- annotation and curation. *Nucleic Acids Research*, 46(D1), D851–D860.
- Haft, D. H., Selengut, J. D., & White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Research*, 31(1), 371–373.
- Huber, R., Sacher, M., Vollmann, A., Huber, H., & Rose, D. (2000). Respiration of arsenate and selenate by hyperthermophilic archaea. *Systematic and Applied Microbiology*, 23(3), 305–314.
- Hubert, H. L., Rankey, E. C., & Omelon, C. (2018). Organic matter, textures, and pore attributes of hypersaline lacustrine microbial deposits (Holocene, Bahamas). *Journal of Sedimentary Research*, 88(7), 827–849.
- Huse, S. M., Dethlefsen, L., Huber, J. A., Mark Welch, D., Relman, D. A., & Sogin, M. L. (2008). Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics*, 4(11), e1000255.
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics: A Joint Publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, 5(3), 299–314.
- Isacks, B. L. (1988). Uplift of the Central Andean Plateau and bending of the Bolivian Orocline. *Journal of Geophysical Research*, 93(B4), 3211.
- Jain, A., Gupta, Y., Agrawal, R., Khare, P., & Jain, S. K. (2007). Biofilms--a microbial

- life perspective: a critical review. *Critical Reviews in Therapeutic Drug Carrier Systems*, 24(5), 393–443.
- Jeck, W. R., Reinhardt, J. A., Baltrus, D. A., Hickenbotham, M. T., Magrini, V., Mardis, E. R., Dangl, J. L., & Jones, C. D. (2007). Extending assembly of short DNA sequences to handle error. *Bioinformatics* , 23(21), 2942–2944.
- Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Research*, 36(Database issue), D250–D254.
- Jiao, S., Xu, Y., Zhang, J., Hao, X., & Lu, Y. (2019). Core Microbiota in Agricultural Soils and Their Potential Associations with Nutrient Cycling. *mSystems*, 4(2). <https://doi.org/10.1128/mSystems.00313-18>
- 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.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40(Database issue), D109–D114.
- Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3, e1165.

- Kannan, N., Taylor, S. S., Zhai, Y., Venter, J. C., & Manning, G. (2007). Structural and functional diversity of the microbial kinome. *PLoS Biology*, 5(3), e17.
- Kauri, T., Wallace, R., & Kushner, D. J. (1990). Nutrition of the halophilic archaeobacterium, *Haloferax volcanii*. *Systematic and Applied Microbiology*, 13(1), 14–18.
- Kawabata, Y., Nakahara, H., Katayama, Y., & Ishida, N. (1997). The phytoplankton of some saline lakes in Central Asia. *International Journal of Salt Lake Research*, 6(1), 5–16.
- Kennard, J. M., & James, N. P. (1986). Thrombolites and stromatolites: Two distinct types of microbial structures. *Palaios*, 1(5), 492.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive -mer weighting and repeat separation. *Genome Research*, 27(5), 722–736.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549.
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews: MMBR*, 72(4), 557–578, Table of Contents.
- Kurth, D., Amadio, A., Ordoñez, O. F., Albarracín, V. H., Gärtner, W., & Farías, M. E. (2017). Arsenic metabolism in high altitude modern stromatolites revealed by metagenomic analysis. *Scientific Reports*, 7(1), 1024.

- Kurth, D., Elias, D., Rasuk, M. C., Contreras, M., & Farías, M. E. (2021). Carbon fixation and rhodopsin systems in microbial mats from hypersaline lakes Brava and Tebenquiche, Salar de Atacama, Chile. *PloS One*, *16*(2), e0246656.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.
- Lasken, R. S. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nature Reviews. Microbiology*, *10*(9), 631–640.
- Lencina, A. I., Soria, M. N., Gomez, F. J., Gérard, E., & Farias, M. E. (2021). Composite microbialites: Thrombolite, dendrolite, and stromatolite associations in a modern environment, Pozo Bravo lake, Salar de Antofalla, Catamarca Puna, Argentina. *Journal of Sedimentary Research*, *91*(12), 1305–1330.
- Liao, Y., Williams, T. J., Walsh, J. C., Ji, M., Poljak, A., Curmi, P. M. G., Duggin, I. G., & Cavicchioli, R. (2016). Developing a genetic manipulation system for the Antarctic archaeon, *Halorubrum lacusprofundi*: investigating acetamidase gene function. *Scientific Reports*, *6*, 34639.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676.
- Lindsay, M. R., Anderson, C., Fox, N., Scofield, G., Allen, J., Anderson, E., Bueter,

- L., Poudel, S., Sutherland, K., Munson-McGee, J. H., Van Nostrand, J. D., Zhou, J., Spear, J. R., Baxter, B. K., Lageson, D. R., & Boyd, E. S. (2017). Microbialite response to an anthropogenic salinity gradient in Great Salt Lake, Utah. *Geobiology*, 15(1), 131–145.
- Lindsay, M. R., Dunham, E. C., & Boyd, E. S. (2020). Microbialites of great Salt Lake. In *Great Salt Lake Biology* (pp. 87–118). Springer International Publishing.
- Liu, J., & Rosen, B. P. (1997). Ligand interactions of the ArsC arsenate reductase. *The Journal of Biological Chemistry*, 272(34), 21084–21089.
- Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659.
- Logan, B. W. (1961). Cryptozoon and associate stromatolites from the Recent, Shark Bay, Western Australia. *Journal of Geology*, 69, 517–533.
- López, D., Vlamakis, H., & Kolter, R. (2010). Biofilms. *Cold Spring Harbor Perspectives in Biology*, 2(7), a000398.
- Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955–964.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1), 18.

- Macy, J. M., Nunan, K., Hagen, K. D., Dixon, D. R., Harbour, P. J., Cahill, M., & Sly, L. I. (1996). *Chrysiogenes arsenatis* gen. nov., sp. nov., a new arsenate-respiring bacterium isolated from gold mine wastewater. *International Journal of Systematic Bacteriology*, 46(4), 1153–1157.
- Madigan, M. T., Martinko, J. M., Bender, K. S., Buckley, D. H., & Stahl, D. A. (2014). *Brock Biology of Microorganisms*. Benjamin Cummings.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376–380.
- Matias Rodrigues, J. F., Schmidt, T. S. B., Tackmann, J., & von Mering, C. (2017). MAPseq: highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics*, 33(23), 3808–3810.
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One*, 8(4), e61217.
- Meng, Y.-L., Liu, Z., & Rosen, B. P. (2004). As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. *The Journal of Biological Chemistry*, 279(18), 18334–18341.
- MetaSUB International Consortium. (2016). The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural

- meeting report. *Microbiome*, 4(1), 24.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., & Edwards, R. A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9, 386.
- Mikheenko, A., Saveliev, V., & Gurevich, A. (2016). MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics*, 32(7), 1088–1090.
- Miller, J. R., Delcher, A. L., Koren, S., Venter, E., Walenz, B. P., Brownley, A., Johnson, J., Li, K., Mobarry, C., & Sutton, G. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics*, 24(24), 2818–2824.
- Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2020). MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research*, 48(D1), D570–D578.
- 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. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, 47(D1), D351–D360.

- Moore, L. S., & Burne, R. V. (1994). The modern thrombolites of lake Clifton, western Australia. In *Phanerozoic Stromatolites II* (pp. 3–29). Springer Netherlands.
- Mukhopadhyay, R., & Rosen, B. P. (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environmental Health Perspectives*, 110 Suppl 5, 745–748.
- Muller, D., Médigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E., Bonnefoy, V., Krin, E., Arsène-Ploetze, F., Carapito, C., Chandler, M., Cournoyer, B., Cruveiller, S., Dossat, C., Duval, S., Heymann, M., Leize, E., Lieutaud, A., Lièvremon, D., ... Bertin, P. N. (2007). A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genetics*, 3(4), e53.
- Murphy, J. N., & Saltikov, C. W. (2009). The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. strain ANA-3. *Journal of Bacteriology*, 191(21), 6722–6731.
- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29(22), 2933–2935.
- Nordstrom, D. K. (2002). Public health. Worldwide occurrences of arsenic in ground water. *Science*, 296(5576), 2143–2145.
- Nthunya, L., Maifadi, S., Mamba, B., Verliefde, A., & Mhlanga, S. (2018). Spectroscopic determination of water salinity in brackish surface water in Nandoni Dam, at Vhembe District, Limpopo Province, South Africa. *WATER*, 10(8), 990.
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a

- new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834.
- Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), 2864–2868.
- Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17(1), 132.
- Oremland, R. S., Stolz, J. F., & Hollibaugh, J. T. (2004). The microbial arsenic cycle in Mono Lake, California. *FEMS Microbiology Ecology*, 48(1), 15–27.
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., ... Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17), 5691–5702.
- Pace, A., Bourillot, R., Bouton, A., Vennin, E., Galaup, S., Bundeleva, I., Patrier, P., Dupraz, C., Thomazo, C., Sansjofre, P., Yokoyama, Y., Franceschi, M., Anguy, Y., Pigot, L., Virgone, A., & Visscher, P. T. (2016). Microbial and diagenetic steps leading to the mineralisation of Great Salt Lake microbialites. *Scientific Reports*, 6, 31495.
- Páez-Espino, D., Tamames, J., de Lorenzo, V., & Cánovas, D. (2009). Microbial responses to environmental arsenic. *Biometals: An International Journal on the*

- Role of Metal Ions in Biology, Biochemistry, and Medicine*, 22(1), 117–130.
- Pamela Reid, R., James, N. P., Macintyre, I. G., Dupraz, C. P., & Burne, R. V. (2003). Shark Bay stromatolites: Microfabrics and reinterpretation of origins. *Facies*, 49(1). <https://doi.org/10.1007/s10347-003-0036-8>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055.
- Paul, B., Dixit, G., Murali, T. S., & Satyamoorthy, K. (2019). Genome-based taxonomic classification. *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*, 62(2), 45–52.
- Peacock, A. J. (1998). ABC of oxygen: oxygen at high altitude. *BMJ*, 317(7165), 1063–1066.
- Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420–1428.
- Perez, M. F., Kurth, D., Farías, M. E., Soria, M. N., Castillo Villamizar, G. A., Poehlein, A., Daniel, R., & Dib, J. R. (2020). First Report on the Plasmidome From a High-Altitude Lake of the Andean Puna. *Frontiers in Microbiology*, 11, 1343.
- Planavsky, N., & Ginsburg, R. N. (2009). TAPHONOMY OF MODERN MARINE BAHAMIAN MICROBIALITES. *Palaios*, 24(1), 5–17.
- Prieto-Barajas, C. M., Valencia-Cantero, E., & Santoyo, G. (2018). Microbial mat

- ecosystems: Structure types, functional diversity, and biotechnological application. *Electronic Journal of Biotechnology: EJB*, 31, 48–56.
- Qin, J., Rosen, B. P., Zhang, Y., Wang, G., Franke, S., & Rensing, C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), 2075–2080.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue), D590–D596.
- Rahman, M. A., & Hassler, C. (2014). Is arsenic biotransformation a detoxification mechanism for microorganisms? *Aquatic Toxicology*, 146, 212–219.
- Rappé, M. S., & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Review of Microbiology*, 57, 369–394.
- Rascovan, N., Maldonado, J., Vazquez, M. P., & Eugenia Farías, M. (2016). Metagenomic study of red biofilms from Diamante Lake reveals ancient arsenic bioenergetics in haloarchaea. *The ISME Journal*, 10(2), 299–309.
- Rasuk, M. C., Fernández, A. B., Kurth, D., Contreras, M., Novoa, F., Poiré, D., & Farías, M. E. (2016). Bacterial Diversity in Microbial Mats and Sediments from the Atacama Desert. *Microbial Ecology*, 71(1), 44–56.
- Rasuk, M. C., Kurth, D., Flores, M. R., Contreras, M., Novoa, F., Poire, D., & Farias, M. E. (2014). Microbial characterization of microbial ecosystems associated to

- evaporites domes of gypsum in Salar de Llamara in Atacama desert. *Microbial Ecology*, 68(3), 483–494.
- Rawle, R., Saley, T. C., Kang, Y.-S., Wang, Q., Walk, S., Bothner, B., & McDermott, T. R. (2021). Introducing the ArsR-Regulated Arsenic Stimulon. *Frontiers in Microbiology*, 12, 630562.
- Reid, R. P., Macintyre, I. G., Browne, K. M., Steneck, R. S., & Miller, T. (1995). Modern marine stromatolites in the Exuma Cays, Bahamas: Uncommonly common. *Facies*, 33(1), 1–17.
- Rhine, E. D., Ní Chadhain, S. M., Zylstra, G. J., & Young, L. Y. (2007). The arsenite oxidase genes (aroAB) in novel chemoautotrophic arsenite oxidizers. *Biochemical and Biophysical Research Communications*, 354(3), 662–667.
- Rho, M., Tang, H., & Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research*, 38(20), e191.
- Richardson, L. J., Rawlings, N. D., Salazar, G. A., Almeida, A., Haft, D. R., Ducq, G., Sutton, G. G., & Finn, R. D. (2019). Genome properties in 2019: a new companion database to InterPro for the inference of complete functional attributes. *Nucleic Acids Research*, 47(D1), D564–D572.
- Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., & Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, 32(6), 929–931.
- Riding, R. (1991). Classification of Microbial Carbonates. In R. Riding (Ed.), *Calcareous Algae and Stromatolites* (pp. 21–51). Springer Berlin Heidelberg.

- Riding, R. (2000). Microbial carbonates: the geological record of calcified bacterial-algal mats and biofilms. *Sedimentology*, 47, 179–214.
- Riding, R. E., & Awramik, S. M. (Eds.). (2000). *Microbial Sediments*. Springer, Berlin, Heidelberg.
- Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., ... Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459), 431–437.
- Rosen, B. P., Bhattacharjee, H., Zhou, T., & Walmsley, A. R. (1999). Mechanism of the ArsA ATPase. *Biochimica et Biophysica Acta, Biomembranes*, 1461(2), 207–215.
- Ruggieri, F., Saavedra, J., Fernandez-Turiel, J. L., Gimeno, D., & Garcia-Valles, M. (2010). Environmental geochemistry of ancient volcanic ashes. *Journal of Hazardous Materials*, 183(1-3), 353–365.
- Saltikov, C. W., Cifuentes, A., Venkateswaran, K., & Newman, D. K. (2003). The ars detoxification system is advantageous but not required for As(V) respiration by the genetically tractable *Shewanella* species strain ANA-3. *Applied and Environmental Microbiology*, 69(5), 2800–2809.
- Sánchez-García, L., Carrizo, D., Vignale, F. A., & Farías, M. E. (2022). *Molecular and Isotopic impact of desiccation in high-altitude wetlands analogous to Martian paleolakes*. <https://doi.org/10.5194/epsc2022-690>

- Sancho-Tomás, M., Somogyi, A., Medjoubi, K., Bergamaschi, A., Visscher, P. T., van Driessche, A. E. S., Gérard, E., Farias, M. E., Contreras, M., & Philippot, P. (2020). Geochemical evidence for arsenic cycling in living microbialites of a High Altitude Andean Lake (Laguna Diamante, Argentina). *Chemical Geology*, 549, 119681.
- Santini, J. M., Sly, L. I., Schnagl, R. D., & Macy, J. M. (2000). A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Applied and Environmental Microbiology*, 66(1), 92–97.
- Santini, J. M., & vanden Hoven, R. N. (2004). Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. *Journal of Bacteriology*, 186(6), 1614–1619.
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863–864.
- Setubal, J. C. (2021). Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophysical Reviews*, 13(6), 905–909.
- 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.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423.

- Shapiro, R. S. (2000). A comment on the systematic confusion of thrombolites. *Palaaios*, 15(2), 166.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539.
- Silver, S., & Phung, L. T. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Applied and Environmental Microbiology*, 71(2), 599–608.
- Simpson, J. T., & Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. *Genome Research*, 22(3), 549–556.
- Slatko, B. E., Kieleczawa, J., Ju, J., Gardner, A. F., Hendrickson, C. L., & Ausubel, F. M. (2011). “First generation” automated DNA sequencing technology. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, Chapter 7, Unit7.2.
- Slyemi, D., & Bonnefoy, V. (2012). How prokaryotes deal with arsenic(†). *Environmental Microbiology Reports*, 4(6), 571–586.
- Soler, M. M., Caffè, P. J., Coira, B. L., Onoe, A. T., & Kay, S. M. (2007). Geology of the Vilama caldera: A new interpretation of a large-scale explosive event in the Central Andean plateau during the Upper Miocene. *Journal of Volcanology and Geothermal Research*, 164(1-2), 27–53.
- Soppa, J. (2006). From genomes to function: haloarchaea as model organisms.

- Microbiology*, 152(3), 585–590.
- Sparks, R. S. J., Francis, P. W., Hamer, R. D., Pankhurst, R. J., O’Callaghan, L. O., Thorpe, R. S., & Page, R. (1985). Ignimbrites of the Cerro Galan caldera, NW Argentina. *Journal of Volcanology and Geothermal Research*, 24(3-4), 205–248.
- Stern, C. R. (2004). Active Andean volcanism: its geologic and tectonic setting. *Revista Geologica de Chile*, 31(2).
<https://doi.org/10.4067/s0716-02082004000200001>
- Stivaletta, N., Barbieri, R., Cevenini, F., & López-García, P. (2011). Physicochemical Conditions and Microbial Diversity Associated with the Evaporite Deposits in the Laguna de la Piedra (Salar de Atacama, Chile). *Geomicrobiology Journal*, 28(1), 83–95.
- Stolz, J. F., Basu, P., Santini, J. M., & Oremland, R. S. (2006). Arsenic and selenium in microbial metabolism. *Annual Review of Microbiology*, 60, 107–130.
- Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Tara Oceans Coordinators, Eveillard, D., Gorsky, G., Guidi, L., Iudicone, D., Karsenti, E., Lombard, F., Ogata, H., Pesant, S., Sullivan, M. B., Wincker, P., & de Vargas, C. (2020). Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews. Microbiology*, 18(8), 428–445.
- Sun, W., Sierra-Alvarez, R., Milner, L., & Field, J. A. (2010). Anaerobic oxidation of arsenite linked to chlorate reduction. *Applied and Environmental Microbiology*, 76(20), 6804–6811.
- Suosaari, E. P., Awramik, S. M., Reid, R. P., Stolz, J. F., & Grey, K. (2018). Living

- dendrolitic microbial mats in Hamelin pool, shark bay, western Australia. *Geosciences*, 8(6), 212.
- Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, 9(1), 2856.
- Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, 278(5338), 631–637.
- The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1), D330–D338.
- Thiel, V., Tank, M., Neulinger, S. C., Gehrmann, L., Dorador, C., & Imhoff, J. F. (2010). Unique communities of anoxygenic phototrophic bacteria in saline lakes of Salar de Atacama (Chile): evidence for a new phylogenetic lineage of phototrophic Gammaproteobacteria from pufLM gene analyses. *FEMS Microbiology Ecology*, 74(3), 510–522.
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3.
- Toneatti, D. M., Albarracín, V. H., Flores, M. R., Polerecky, L., & Fariás, M. E. (2017). Stratified Bacterial Diversity along Physico-chemical Gradients in High-Altitude Modern Stromatolites. *Frontiers in Microbiology*, 8, 646.
- Tsai, S.-L., Singh, S., & Chen, W. (2009). Arsenic metabolism by microbes in nature and the impact on arsenic remediation. *Current Opinion in Biotechnology*, 20(6), 659–667.

- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164), 804–810.
- Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, 6(1), 158.
- Valero-Garcés, B. L., Arenas, C., & Delgado-Huertas, A. (2001). Depositional environments of Quaternary lacustrine travertines and stromatolites from high-altitude Andean lakes, northwestern Argentina. *Canadian Journal of Earth Sciences*, 38(8), 1263–1283.
- van Lis, R., Nitschke, W., Duval, S., & Schoepp-Cothenet, B. (2013). Arsenics as bioenergetic substrates. *Biochimica et Biophysica Acta*, 1827(2), 176–188.
- Vignale, F. A., Kurth, D., Lencina, A. I., Poiré, D. G., Chihuailaf, E., Muñoz-Herrera, N. C., Novoa, F., Contreras, M., Turjanski, A. G., & Farías, M. E. (2021). Geobiology of Andean Microbial Ecosystems Discovered in Salar de Atacama, Chile. *Frontiers in Microbiology*, 12, 762076.
- Vignale, F. A., Lencina, A. I., Stepanenko, T. M., Soria, M. N., Saona, L. A., Kurth, D., Guzmán, D., Foster, J. S., Poiré, D. G., Villafañe, P. G., Albarracín, V. H., Contreras, M., & Farías, M. E. (2021). Lithifying and Non-Lithifying Microbial Ecosystems in the Wetlands and Salt Flats of the Central Andes. *Microbial Ecology*. <https://doi.org/10.1007/s00248-021-01725-8>
- Vilcabana Sánchez, V., Colestock de Shaver, G., Shaver Ashby, A., Van der Walt de Groenewald, R., & Sánchez Bernilla, R. (2007). *Guía de lengua quechua para*

- Castellano—hablantes*. <http://repositorio.cultura.gob.pe/handle/CULTURA/402>
- Villafañe, P. G., Lencina, A. I., Soria, M., Saona, L. A., Gómez, F. J., Alonso, G. E., & Farias, M. E. (2021). LAS QUÍNOAS ONCOIDS: A NEW DEPOSIT OF MICROBIALITES IN THE SALAR DE ANTOFALLA (CATAMARCA, ARGENTINA). *Andean Geology*, 48(2).
<https://doi.org/10.5027/andgeoV48n2-3292>
- Wallace, T. C., & Hall-Wallace, M. K. (2003). Minerals of the Andes: Emeralds, gold, and silver from the sky. *Rocks & Minerals*, 78(1), 12–38.
- Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., & Foster, J. S. (2016). Characterization of Microbial Mat Microbiomes in the Modern Thrombolite Ecosystem of Lake Clifton, Western Australia Using Shotgun Metagenomics. *Frontiers in Microbiology*, 7, 1064.
- Warren, R. L., Sutton, G. G., Jones, S. J. M., & Holt, R. A. (2007). Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*, 23(4), 500–501.
- Wei, H., Peng, C., Yang, B., Song, H., Li, Q., Jiang, L., Wei, G., Wang, K., Wang, H., Liu, S., Liu, X., Chen, D., Li, Y., & Wang, M. (2018). Contrasting Soil Bacterial Community, Diversity, and Function in Two Forests in China. *Frontiers in Microbiology*, 9, 1693.
- Wharton, R. A., Jr, Crosby, J. M., McKay, C. P., & Rice, J. W., Jr. (1995). Paleolakes on Mars. *Journal of Paleolimnology*, 13, 267–283.
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen

- majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), 6578–6583.
- Whitton, B. A., & Potts, M. (2012). Introduction to the Cyanobacteria. In *Ecology of Cyanobacteria II* (pp. 1–13). Springer Netherlands.
- Wooley, J. C., Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS Computational Biology*, 6(2), e1000667.
- Woyke, T., Teeling, H., Ivanova, N. N., Huntemann, M., Richter, M., Gloeckner, F. O., Boffelli, D., Anderson, I. J., Barry, K. W., Shapiro, H. J., Szeto, E., Kyrpides, N. C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E. M., & Dubilier, N. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature*, 443(7114), 950–955.
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., Hettich, R. L., Lipton, M. S., Williams, K. H., Long, P. E., & Banfield, J. F. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*, 337(6102), 1661–1665.
- Wu, Y.-F., Chai, C.-W., Li, Y.-N., Chen, J., Yuan, Y., Hu, G., Rosen, B. P., & Zhang, J. (2021). Anaerobic as(III) oxidation coupled with nitrate reduction and attenuation of dissolved arsenic by *Noviherbaspirillum* species. *ACS Earth & Space Chemistry*, 5(8), 2115–2123.
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.

- Bioinformatics* , 32(4), 605–607.
- Yang, H.-C., Fu, H.-L., Lin, Y.-F., & Rosen, B. P. (2012). Pathways of arsenic uptake and efflux. *Current Topics in Membranes*, 69, 325–358.
- Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., Eisen, J. A., Heidelberg, K. B., Manning, G., Li, W., Jaroszewski, L., Cieplak, P., Miller, C. S., Li, H., Mashiyama, S. T., Joachimiak, M. P., van Belle, C., Chandonia, J.-M., Soergel, D. A., ... Venter, J. C. (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biology*, 5(3), e16.
- Yoshinaga, M., Cai, Y., & Rosen, B. P. (2011). Demethylation of methylarsonic acid by a microbial community. *Environmental Microbiology*, 13(5), 1205–1215.
- Yoshinaga, M., & Rosen, B. P. (2014). A C·As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 111(21), 7701–7706.
- Zargar, K., Hoeft, S., Oremland, R., & Saltikov, C. W. (2010). Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. *Journal of Bacteriology*, 192(14), 3755–3762.
- Zeil, W. (1979). *The Andes: A Geological Review*. Gebrüder Borntraeger.

APÉNDICE

Tabla A1. Estadísticas de los MAGs.

MAG	Compleitud	Contaminación	Heterogeneidad	MAGs con igual identidad
D.1.1	87,36	1,19	40	-
D.1.3	95,9	0,65	0	-
D.2.1	73,03	1,28	33,33	D.3.1
D.3.2	56,23	0,65	0	-
D.3.4	82,85	1,72	33,33	D.1.2,D.2.2
B.1.3	86,49	1,36	0	B.2.19
B.1.7	84,34	2,37	25	B.2.44
B.1.8	90,53	4,05	0	-
B.1.9	89,77	3,53	12,5	B.2.72,T.1.58,T.2.21
B.1.11	90,7	3,27	0	B.2.48
B.1.13	63,31	0,66	0	B.2.10
B.1.15	100	1,13	0	B.2.85
B.1.17	84,32	2,36	25	B.2.20
B.1.18	98,66	0,54	0	B.2.81
B.1.20	91,82	0	0	B.2.23
B.1.23	99,6	0,4	0	B.2.84
B.1.24	95,45	0,1	100	B.2.37
B.1.26	97,27	1,82	0	B.2.42
B.1.30	91,75	0,64	0	B.2.68,P.1
B.1.31	89,41	1,68	50	B.2.53
B.1.32	59,91	1,23	0	-
B.1.34	97,8	1,89	50	B.2.9
B.1.40	97,27	1,92	33,33	B.2.21
B.1.42	96,77	2,9	20	B.2.6,T.1.27,T.2.55
B.1.43	94,32	0,16	100	B.2.46
B.1.45	93,76	1,33	0	B.2.63,T.1.41,T.2.54

B.1.55	97,16	1,14	0	B.2.86
B.1.59	96,23	0,19	0	B.2.5
B.1.60	73,64	1,74	66,67	B.2.2
B.1.61	74,53	0	0	B.2.31
B.1.64	97,8	1,1	0	B.2.89
B.1.66	97,27	0	0	B.2.35
B.1.70	87,5	1,24	100	B.2.79
B.1.74	96	2	0	B.2.36,T.1.16,T.2.66
B.1.77	97,94	3,33	0	B.2.43,T.1.46,T.2.11
B.1.79	98,1	0,54	0	B.2.25
B.1.81	87,27	3,03	0	B.2.61
B.2.3	96,77	1,1	66,67	B.1.72,PB.8
B.2.4	62,19	0,47	0	-
B.2.8	80,97	3,83	14,29	-
B.2.13	86,27	0,38	0	B.1.56
B.2.14	68,79	1,04	75	B.1.1
B.2.15	71,69	2,69	0	B.1.29
B.2.18	94,55	2	0	B.1.57
B.2.24	79,15	1,88	14,29	B.1.21
B.2.27	92,67	3,2	0	B.1.25,T.2.32
B.2.29	89,09	2,08	0	B.1.54
B.2.30	87,33	1,79	22,22	B.1.46,T.1.15,T.2.9
B.2.32	84,5	1,27	0	B.1.41,S.1
B.2.40	92,55	1,09	0	B.1.62
B.2.41	73,94	1,08	25	-
B.2.45	86,44	2,83	46,15	B.1.10
B.2.47	97,04	1	0	B.1.50,T.1.47,T.2.4,P.13
B.2.50	99,69	1,13	0	B.1.44
B.2.51	99,2	3,6	0	B.1.2,T.1.19,T.2.33
B.2.52	87,03	1,41	33,33	B.1.6,T.1.8,T.2.68

B.2.54	96,82	1,82	0	B.1.27
B.2.57	82,36	1,17	0	B.1.19
B.2.58	70,14	1,36	0	B.1.36
B.2.59	77,81	0,1	0	B.1.47
B.2.64	81,26	1,92	0	B.1.38
B.2.65	59,83	0,8	100	-
B.2.66	80,09	1,97	0	B.1.71
B.2.67	79,92	1,6	0	B.1.48
B.2.70	73,58	1,81	13,33	B.1.22
B.2.71	62,06	0,16	0	B.1.51
B.2.73	58,04	0,27	0	-
B.2.74	92,73	0	0	B.1.16
B.2.75	85,98	0	0	B.1.52
B.2.76	89,82	2	33,33	B.1.33
B.2.77	88,53	2,27	0	B.1.12
B.2.78	74,42	1,31	0	B.1.5
B.2.80	98,8	0	0	B.1.80
B.2.83	77,19	2,27	0	B.1.49
B.2.87	92,8	3,2	0	B.1.53,T.1.18,T.2.56
T.1.1	67,68	0,71	0	T.2.49
T.1.9	91,6	2,27	0	B.1.28,B.2.11,T.2.57
T.1.10	76,53	0,28	0	T.2.16
T.1.11	72,77	1,98	50	T.2.25
T.1.20	78,18	1,36	0	B.2.22,T.2.30
T.1.26	96,19	2,62	0	T.2.48
T.1.28	96,83	1,88	0	T.2.17
T.1.34	96,59	0	0	T.2.35
T.1.35	88,86	1	60	T.2.28,PB.4
T.1.36	82,02	0	0	T.2.46
T.1.37	86,97	1,95	14,29	T.2.1

T.1.39	96,11	0,51	0	T.2.3,PB.7
T.1.44	83,3	0,85	0	T.2.2
T.1.45	70,25	1,74	80	T.2.5
T.1.48	77,68	1,4	60	T.2.18
T.1.50	89,47	2,8	0	B.1.68,B.2.16,T.2.43
T.1.51	56,12	0	0	-
T.1.55	98,23	1,33	0	T.2.42
T.1.57	96,52	1,52	0	B.1.75,B.2.33,T.2.63
T.2.7	90,2	0,68	0	T.1.42
T.2.8	96,21	1,7	28,57	T.1.43
T.2.10	81,93	3,07	20	T.1.38
T.2.13	92,33	1,62	40	T.1.7
T.2.14	95,48	3,87	0	B.1.14,B.2.39,T.1.21
T.2.15	55,91	1,1	0	-
T.2.19	98,48	1,14	0	T.1.52
T.2.20	91,08	2,02	50	B.1.39,B.2.34,T.1.31
T.2.22	96,96	0,14	0	T.1.3,S.16
T.2.24	67,47	1,64	42,86	T.1.12
T.2.26	86,97	4,38	0	T.1.49
T.2.27	97,74	2,3	0	T.1.59
T.2.31	68,97	3,45	25	-
T.2.36	69,62	0,85	50	B.2.49,T.1.2
T.2.39	95,9	1,37	0	B.1.37,B.2.12,T.1.32
T.2.40	97,42	0,99	33,33	B.1.69,B.2.28,T.1.33,S.26
T.2.44	69,45	1,36	50	B.1.63,B.2.55,T.1.13,S.3
T.2.47	92,83	1,08	0	B.1.58,B.2.38,T.1.30
T.2.50	92,29	0	0	T.1.40
T.2.51	66,67	0,93	0	T.1.25
T.2.53	82,77	4,03	14,29	-
T.2.61	97,04	3,49	11,11	T.1.6

T.2.62	73,05	1,24	33,33	-
T.2.69	83,65	2,1	37,5	T.1.4
P.2	55,54	0,99	0	-
P.3	54,68	0	0	-
P.5	92,76	0,37	100	-
P.6	61,96	2,11	15,38	B.2.26,B.1.67,S.23
P.9	89,63	1,26	28,57	-
P.12	96,12	1,94	36,36	-
PB.1	80,79	1,34	50	-
PB.2	93,9	1,16	0	-
PB.3	90,2	1,2	66,67	-
PB.5	98,72	0,85	0	-
PB.6	74,8	3,3	28,57	T.2.12,T.1.22
PB.9	89,07	1,11	66,67	-
PB.10	94,89	0,57	0	-
PB.12	98,51	0,58	0	-
PB.13	89,73	0,89	0	-
PB.14	57,53	0	0	-
S.2	89,08	0	0	-
S.4	79,43	3,77	8.7	-
S.5	93,35	0,9	50	-
S.6	91,08	1,28	66,67	-
S.7	94,3	0,85	0	-
S.8	61,49	1,66	0	-
S.9	94,09	0,54	0	-
S.11	60,34	0	0	-
S.12	62,6	0,88	0	-
S.13	89,32	1,53	12,5	-
S.19	91,67	2,12	20	-
S.20	88,24	1,01	0	-

S.21	67,59	0	0	-
S.22	98,31	1,98	0	T.1.24,B.1.35,B.2.69,T.2.6
S.24	93,11	0,82	0	T.1.14,T.2.23
S.25	75,27	4,33	28,57	-
S.27	87,6	2,31	40	-
S.28	81,19	1,09	42,86	-

D.1, Laguna Diamante 1; D.2, Laguna Diamante 2; D.3, Laguna Diamante 3; B.1, Laguna Brava 1; B.2, Laguna Brava 2; T.1, Laguna Tebenquiche 1; T.2, Laguna Tebenquiche 2; P, *Puquio* de campo naranja; PB, Laguna Pozo Bravo; S, Laguna Socompa.

Tabla A2. Clasificación taxonómica de los MAGs.

MAG	Dominio	Filo	Mejor clasificación
D.1.1	<i>Archaea</i>	<i>Euryarchaeota</i>	p: <i>Euryarchaeota</i>
D.1.3	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Oceanospirillales</i>
D.2.1	<i>Archaea</i>	<i>Euryarchaeota</i>	f: <i>Halobacteriaceae</i>
D.3.2	<i>Archaea</i>	<i>Euryarchaeota</i>	f: <i>Halobacteriaceae</i>
D.3.4	<i>Archaea</i>	<i>Euryarchaeota</i>	f: <i>Halobacteriaceae</i>
B.1.3	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
B.1.7	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.8	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.9	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Flammeovirgaceae</i>
B.1.11	<i>Archaea</i>	-	d: <i>Archaea</i>
B.1.13	<i>Bacteria</i>	<i>Spirochaetota</i>	o: <i>Spirochaetales</i>
B.1.15	<i>Bacteria</i>	<i>Chloroflexota</i>	f: <i>Chloroflexaceae</i>
B.1.17	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.18	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
B.1.20	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.23	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.1.24	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.1.26	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.30	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.1.31	<i>Bacteria</i>	<i>Pseudomonadota</i>	p: <i>Pseudomonadota</i>
B.1.32	<i>Bacteria</i>	<i>Spirochaetota</i>	o: <i>Spirochaetales</i>
B.1.34	<i>Bacteria</i>	<i>Chloroflexota</i>	f: <i>Chloroflexaceae</i>
B.1.40	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.42	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
B.1.43	<i>Bacteria</i>	-	d: <i>Bacteria</i>
B.1.45	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Rhodospirillales</i>
B.1.55	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.1.59	<i>Bacteria</i>	<i>Chloroflexota</i>	f: <i>Chloroflexaceae</i>

B.1.60	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
B.1.61	<i>Bacteria</i>	-	d: <i>Bacteria</i>
B.1.64	<i>Bacteria</i>	-	d: <i>Bacteria</i>
B.1.66	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.1.70	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.1.74	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.1.77	<i>Bacteria</i>	<i>Bacteroidota</i>	o: <i>Bacteroidales</i>
B.1.79	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
B.1.81	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.3	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
B.2.4	<i>Archaea</i>	-	d: <i>Archaea</i>
B.2.8	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.13	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.14	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Hyphomicrobiaceae</i>
B.2.15	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Saprospiraceae</i>
B.2.18	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.24	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Flammeovirgaceae</i>
B.2.27	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.29	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.30	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Flammeovirgaceae</i>
B.2.32	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Xanthomonadales</i>
B.2.40	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
B.2.41	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Porphyromonadaceae</i>
B.2.45	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Xanthomonadales</i>
B.2.47	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
B.2.50	<i>Bacteria</i>	<i>Chloroflexota</i>	f: <i>Chloroflexaceae</i>
B.2.51	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.52	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
B.2.54	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.57	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>

B.2.58	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.59	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.2.64	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.65	<i>Archaea</i>	<i>Euryarchaeota</i>	p: <i>Euryarchaeota</i>
B.2.66	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Myxococcales</i>
B.2.67	<i>Bacteria</i>	<i>Chloroflexota</i>	f: <i>Chloroflexaceae</i>
B.2.70	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Hyphomonadaceae</i>
B.2.71	<i>Bacteria</i>	-	d: <i>Bacteria</i>
B.2.73	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.74	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.75	<i>Archaea</i>	-	d: <i>Archaea</i>
B.2.76	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
B.2.77	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
B.2.78	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Myxococcales</i>
B.2.80	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
B.2.83	<i>Bacteria</i>	-	d: <i>Bacteria</i>
B.2.87	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.1.1	<i>Bacteria</i>	<i>Verrucomicrobiota</i>	o: <i>Opitutales</i>
T.1.9	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.1.10	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
T.1.11	<i>Bacteria</i>	<i>Chloroflexota</i>	c: <i>Dehalococcoidetes</i>
T.1.20	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
T.1.26	<i>Bacteria</i>	<i>Bacteroidota</i>	o: <i>Bacteroidales</i>
T.1.28	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Porphyromonadaceae</i>
T.1.34	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.1.35	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
T.1.36	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.1.37	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.1.39	<i>Bacteria</i>	<i>Pseudomonadota</i>	c: <i>Gammaproteobacteria</i>
T.1.44	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>

T.1.45	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Rhodospirillaceae</i>
T.1.48	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
T.1.50	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.1.51	<i>Bacteria</i>	<i>Spirochaetota</i>	o: <i>Spirochaetales</i>
T.1.55	<i>Bacteria</i>	<i>Bacillota</i>	f: <i>Halobacteroidaceae</i>
T.1.57	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.2.7	<i>Bacteria</i>	<i>Verrucomicrobiota</i>	o: <i>Opitutales</i>
T.2.8	<i>Bacteria</i>	<i>Pseudomonadota</i>	c: <i>Gammaproteobacteria</i>
T.2.10	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.2.13	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
T.2.14	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Myxococcales</i>
T.2.15	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.2.19	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.2.20	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Saprospiraceae</i>
T.2.22	<i>Bacteria</i>	<i>Verrucomicrobiota</i>	o: <i>Opitutales</i>
T.2.24	<i>Bacteria</i>	<i>Pseudomonadota</i>	g: <i>Haliea</i>
T.2.26	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Myxococcales</i>
T.2.27	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.2.31	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
T.2.36	<i>Bacteria</i>	<i>Spirochaetota</i>	o: <i>Spirochaetales</i>
T.2.39	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
T.2.40	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
T.2.44	<i>Bacteria</i>	<i>Chloroflexota</i>	p: <i>Chloroflexota</i>
T.2.47	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.2.50	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Cryomorphaceae</i>
T.2.51	<i>Bacteria</i>	-	d: <i>Bacteria</i>
T.2.53	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
T.2.61	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Porphyromonadaceae</i>
T.2.62	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Chromatiaceae</i>
T.2.69	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfohalobiaceae</i>

P.2	<i>Bacteria</i>	<i>Chloroflexota</i>	c: <i>Dehalococcoidetes</i>
P.3	<i>Bacteria</i>	-	d: <i>Bacteria</i>
P.5	<i>Bacteria</i>	-	d: <i>Bacteria</i>
P.6	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
P.9	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
P.12	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Chromatiaceae</i>
PB.1	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
PB.2	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>
PB.3	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Cytophagaceae</i>
PB.5	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
PB.6	<i>Bacteria</i>	<i>Pseudomonadota</i>	f: <i>Desulfobacteraceae</i>
PB.9	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
PB.10	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
PB.12	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
PB.13	<i>Bacteria</i>	<i>Cyanobacteria</i>	p: <i>Cyanobacteria</i>
PB.14	<i>Bacteria</i>	<i>Planctomycetota</i>	f: <i>Planctomycetaceae</i>
S.2	<i>Bacteria</i>	<i>Actinomycetota</i>	c: <i>Actinomycetota</i>
S.4	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Xanthomonadales</i>
S.5	<i>Bacteria</i>	<i>Pseudomonadota</i>	g: <i>Roseovarius</i>
S.6	<i>Bacteria</i>	<i>Bacillota</i>	o: <i>Clostridiales</i>
S.7	<i>Bacteria</i>	<i>Actinomycetota</i>	c: <i>Actinomycetota</i>
S.8	<i>Bacteria</i>	<i>Bacillota</i>	f: <i>Bacillaceae</i>
S.9	<i>Bacteria</i>	<i>Bacteroidota</i>	o: <i>Flavobacteriales</i>
S.11	<i>Bacteria</i>	<i>Pseudomonadota</i>	g: <i>Pseudoalteromonas</i>
S.12	<i>Bacteria</i>	<i>Bacillota</i>	f: <i>Bacillaceae</i>
S.13	<i>Bacteria</i>	<i>Pseudomonadota</i>	o: <i>Xanthomonadales</i>
S.19	<i>Bacteria</i>	<i>Deinococcota</i>	f: <i>Deinococcaceae</i>
S.20	<i>Bacteria</i>	<i>Bacteroidota</i>	f: <i>Saprospiraceae</i>
S.21	<i>Bacteria</i>	<i>Pseudomonadota</i>	g: <i>Halomonas</i>
S.22	<i>Bacteria</i>	<i>Bacteroidota</i>	g: <i>Rhodothermus</i>

S.24	<i>Bacteria</i>	<i>Bacteroidota</i>	p: <i>Bacteroidota</i>
S.25	<i>Bacteria</i>	<i>Bacillota</i>	f: <i>Carnobacteriaceae</i>
S.27	<i>Bacteria</i>	<i>Spirochaetota</i>	f: <i>Spirochaetaceae</i>
S.28	<i>Bacteria</i>	<i>Pseudomonadota</i>	g: <i>Roseovarius</i>

D.1, Laguna Diamante 1; D.2, Laguna Diamante 2; D.3, Laguna Diamante 3; B.1, Laguna Brava 1; B.2, Laguna Brava 2; T.1, Laguna Tebenquiche 1; T.2, Laguna Tebenquiche 2; P, *Puquio* de campo naranja; PB, Laguna Pozo Bravo; S, Laguna Socompa; p, Filo; c, Clase; o, Orden; f, Familia; g, Género.

Tabla A3. Comparación entre el genoma de *Halorubrum* sp. DM2 y los genomas disponibles (Mayo de 2019) del género *Halorubrum*.

	ANIb (%)	DDH (%)	Mash distance	P-value	Matching hashes
	<i>Halorubrum</i> sp. DM2				
<i>Halorubrum distributum</i> JCM 9100	86,34-85,92	31,50	0,102268	8,03E-299	62/1000
<i>Halorubrum distributum</i> JCM 10118	86,18-85,92	31,50	0,102268	7,93E-299	62/1000
<i>Halorubrum distributum</i> E8	86,27-86,03	34,10	0,119909	1,31E-199	42/1000
<i>Halorubrum lacusprofundi</i> ATCC 49239	81,57-81,48	31,40	0,138572	2,42E-125	28/1000
<i>Halorubrum lacusprofundi</i> HLS1	81,29-81,41	30,30	0,138572	4,60E-126	28/1000
<i>Halorubrum lacusprofundi</i> DL18	81,04-81,38	25,40	0,142009	9,55E-116	26/1000
<i>Halorubrum ezzemoulense</i> Fb21	84,86-85,25	30,60	0,112887	8,64E-237	49/1000
<i>Halorubrum ezzemoulense</i> DSM 17463	85,17-85,40	30,60	0,117785	4,64E-205	44/1000
<i>Halorubrum ezzemoulense</i> Ga2p	85,35-85,37	30,50	0,113823	3,89E-226	48/1000
<i>Halorubrum ezzemoulense</i> LD3	84,92-85,40	30,60	0,115759	6,48E-215	46/1000
<i>Halorubrum ezzemoulense</i> LG1	84,49-85,38	30,90	0,113823	6,15E-223	48/1000
<i>Halorubrum ezzemoulense</i> Ga36	85,00-85,36	30,40	0,112887	3,07E-231	49/1000
<i>Halorubrum ezzemoulense</i> G37	84,93-85,37	30,60	0,115759	2,45E-215	46/1000

<i>Halorubrum ezzemoulense</i> Ec15	85,39-85,44	30,70	0,115759	3,90E-216	46/1000
<i>Halorubrum tebenquichense</i> DSM 14210	90,15-90,07	40,40	0,0690061	0	133/1000
<i>Halorubrum litoreum</i> JCM 13561	86,66-86,08	31,60	0,106047	2,14E-273	57/1000
<i>Halorubrum kocurii</i> JCM 14978	81,85-81,75	25,80	0,129604	6,66E-155	34/1000
<i>Halorubrum hochstenium</i> ATCC 700873	90,03-89,50	39,20	0,0740698	0	118/1000
<i>Halorubrum coriense</i> DSM 10284	85,27-85,30	30,90	0,107658	3,53E-261	55/1000
<i>Halorubrum californiensis</i> DSM 19288	85,78-85,60	30,80	0,105264	1,40E-276	58/1000
<i>Halorubrum arcis</i> JCM 13916	86,24-86,10	31,60	0,104495	7,59E-283	59/1000
<i>Halorubrum aidingense</i> JCM 13560	81,28-81,04	25,10	0,156619	4,80E-83	19/1000
<i>Halorubrum saccharovorum</i> DSM 1137	82,16-81,92	26,00	0,130979	2,34E-150	33/1000
<i>Halorubrum saccharovorum</i> H3	82,12-82,13	26,50	0,132399	1,03E-145	32/1000
<i>Halorubrum terrestre</i> JCM 10247	86,08-85,93	31,60	0,102997	2,61E-293	61/1000
<i>Halorubrum lipolyticum</i> DSM 21995	82,17-82,02	26,30	0,126974	2,97E-165	36/1000
<i>Halorubrum trapanicum</i> CBA1232	86,67-86,06	31,60	0,0974869	0	69/1000

<i>Halorubrum chaoviator</i> DSM 19316	84,99- 85,42	30,70	0,112887	2,68E-230	49/1000
<i>Halorubrum aethiopicum</i> SAH-A6	80,27- 80,10	24,40	0,147707	8,90E-102	23/1000
<i>Halorubrum persicum</i> C49	81,63- 81,96	25,90	0,13538	2,73E-135	30/1000
<i>Halorubrum halodurans</i> Cb34	79,97- 79,93	24,40	0,159146	4,61E-78	18/1000
<i>Halorubrum vacuolatum</i> DSM 8800	76,57- 76,66	22,70	0,182269	8,43E-46	11/1000
<i>Halorubrum tropicale</i> 5	85,62- 85,37	30,70	0,107658	1,97E-261	55/1000
<i>Halorubrum sodomense</i> RD26	86,47- 85,64	31,00	0,100846	0	64/1000
<i>Halorubrum halophilum</i> B8	81,73- 82,21	26,40	0,132399	6,33E-145	32/1000
<i>Halorubrum sp.</i> PV6	81,29- 80,82	24,50	0,149777	5,68E-98	22/1000
<i>Halorubrum sp.</i> BOL3-1	87,30- 87,56	35,00	0,0898429	0	82/1000
<i>Halorubrum sp.</i> T3	86,25- 85,93	31,20	0,109331	1,69E-252	53/1000
<i>Halorubrum sp.</i> BV1	80,84- 80,50	24,90	0,154223	2,49E-88	20/1000
<i>Halorubrum sp.</i> lb24	81,51- 81,00	25,20	0,145727	3,38E-107	24/1000
<i>Halorubrum sp.</i> WN019	82,66- 82,53	26,40	0,133864	2,50E-140	31/1000
<i>Halorubrum sp.</i> 48-1-W	79,46- 79,66	24,40	0,161822	4,21E-73	17/1000
<i>Halorubrum sp.</i> Atlit-28R	85,50- 85,40	30,80	0,108487	1,82E-256	54/1000

<i>Halorubrum</i> sp. Atlit-9R	86,03- 85,77	31,20	0,102997	2,04E-291	61/1000
<i>Halorubrum</i> sp. Atlit-26R	85,79- 85,93	31,60	0,102268	6,68E-296	62/1000
<i>Halorubrum</i> sp. Atlit-8R	86,15- 85,78	31,30	0,102997	1,38E-293	61/1000
<i>Halorubrum</i> sp. CBA1229	81,80- 81,86	25,80	0,130979	7,43E-150	33/1000
<i>Halorubrum</i> sp. CSM-61	82,22- 82,39	26,30	0,13538	4,72E-135	30/1000
<i>Halorubrum</i> sp. AJ67	96,73- 96,99	83,50	0,0316813	0	346/1000
<i>Halorubrum</i> sp. SD626R	82,21- 81,77	26,00	0,132399	8,70E-146	32/1000
<i>Halorubrum</i> sp. SD612	86,64- 85,72	31,10	0,109331	5,53E-252	53/1000
<i>Halorubrum</i> sp. SD683	85,43- 85,41	30,70	0,11676	1,92E-211	45/1000
<i>Halorubrum</i> sp. Hd13	81,71- 81,79	25,60	0,140258	2,58E-120	27/1000
<i>Halorubrum</i> sp. Eb13	82,03- 81,91	25,60	0,138572	1,89E-125	28/1000
<i>Halorubrum</i> sp. Ea8	81,69- 81,80	25,70	0,142009	1,38E-115	26/1000
<i>Halorubrum</i> sp. Ea1	81,54- 81,74	25,60	0,140258	3,77E-120	27/1000
<i>Halorubrum</i> sp. C191	84,79- 85,31	30,60	0,11197	7,16E-235	50/1000
<i>Halorubrum</i> sp. ZC67	81,93- 82,14	26,10	0,136948	2,59E-130	29/1000
<i>Halorubrum</i> sp. J07HR59	70,82- 70,81	21,50	1	1	0/1000
<i>Halorubrum</i> sp. E3	84,47- 84,85	32,20	0,132399	4,69E-144	32/1000
<i>Halorubrum</i> sp. C3	85,64- 85,80	33,00	0,106845	1,80E-267	56/1000

<i>Halorubrum</i> sp. PL-Br10_E2g5	79,24- 78,49	24,20	0,182269	2,95E-48	11/1000
<i>Halorubrum</i> sp. B1Sed10_22 B1	77,55- 76,54	22,70	0,182269	2,97E-49	11/1000
<i>Halorubrum</i> sp. T1Sed10_15 T1	76,77- 76,53	22,30	0,182269	1,20E-46	11/1000
<i>Halorubrum</i> sp. BC156	92,97- 93,52	55	0,0482105	0	222/1000

Tabla A4. Comparación entre el genoma de *Halorubrum* sp. DM2 y los MAGs reconstruidos de los metagenomas de Diamante.

	ANIb (%)	DDH (%)	<i>Mash distance</i>	<i>P-value</i>	<i>Matching hashes</i>
	<i>Halorubrum</i> sp. DM2				
D.1.1	67,94-67,38	17,40	1	1	0/1000
D.2.1	73,23-72,19	20,20	0,263022	2,96818E-08	2/1000
D.3.2	71,17-70,66	18,70	0,263022	2,74244E-08	2/1000
D.3.4	72,33-71,99	19,20	1	1	0/1000

Figura A1. En *Halorubrum* sp. DM2 $\Delta arrA$, la subunidad grande de la enzima arsenato reductasa (ArrA) está truncada en el extremo C terminal ya que el gen *arrA* presenta un codón de terminación prematuro. En la figura se muestra el alineamiento entre la secuencia aminoacídica de ArrA salvaje y la secuencia aminoacídica de ArrA mutada.