



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA

BIOLOGÍA EXPERIMENTAL

**DIVERSIDAD GENÓMICA DE BACTERIAS TERMOACIDÓFILAS DE SEDIMENTOS DEL
CAMPO GEOTÉRMICO DE LOS AZUFRES**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

HERMES HERNÁN BOLÍVAR TORRES

TUTOR PRINCIPAL: DR. LUIS EDUARDO SERVÍN GARCIDUEÑAS

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA. UNAM

COMITÉ TUTOR: DRA. YUNUEN TAPIA TORRES

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA. UNAM

DRA. MARIA ESPERANZA MARTÍNEZ ROMERO

CENTRO DE CIENCIAS GENÓMICAS. UNAM

MORELIA, MICHOACÁN, OCTUBRE 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA

BIOLOGÍA EXPERIMENTAL

**DIVERSIDAD GENÓMICA DE BACTERIAS TERMOACIDÓFILAS DE SEDIMENTOS DEL
CAMPO GEOTÉRMICO DE LOS AZUFRES**

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

HERMES HERNÁN BOLÍVAR TORRES

TUTOR PRINCIPAL: DR. LUIS EDUARDO SERVÍN GARCIDUEÑAS

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA. UNAM

COMITÉ TUTOR: DRA. YUNUEN TAPIA TORRES

ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA. UNAM

DRA. MARIA ESPERANZA MARTÍNEZ ROMERO

CENTRO DE CIENCIAS GENÓMICAS. UNAM

MORELIA, MICHOACÁN 2021



COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
ENTIDAD ESCUELA NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA
OFICIO CPCB/955/2021
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 08 de marzo de 2021 se aprobó el siguiente jurado para el examen de grado de MAESTRO EN CIENCIAS BIOLÓGICAS en el campo de conocimiento de Biología Experimental del estudiante BOLÍVAR TORRES HERMES HERNÁN con número de cuenta 519491953 con la tesis titulada "Diversidad genómica de bacterias termoacidófilas de sedimentos del campo geotérmico de Los Azufres", realizada bajo la dirección del DR. LUIS EDUARDO SERVÍN GARCIDUEÑAS quedando integrado de la siguiente manera:

Presidente: DRA. LUISA ISAURA FALCÓN ÁLVAREZ
Vocal: DRA. PATRICIA VÉLEZ AGUILAR
Vocal: DRA. LILIA MONTOYA LORENZANA
Vocal: DRA. ROCÍO JETZABEL ALCÁNTARA HERNÁNDEZ
Secretario: DRA. MARÍA ESPERANZA MARTÍNEZ ROMERO

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, Cd. Mx., a 04 de octubre de 2021

COORDINADOR DEL PROGRAMA

DR. ADOLFO GERARDO NAVARRO SIGÜENZA



Agradecimientos institucionales

Al **Posgrado de Ciencias Biológicas UNAM** por brindarme la oportunidad de desarrollar mis estudios de maestría en la Escuela Nacional de Estudios Superiores Unidad Morelia, por medio de la cual se gestó la elaboración de este trabajo.

Al **Consejo Nacional de Ciencia y Tecnología (CONACYT)** por haberme otorgado la beca N° 964652, con la cual pude realizar mis estudios de maestría, ya que sin este apoyo económico hubiese sido imposible poder dedicar mis esfuerzos exclusivamente al desarrollo de mi posgrado y a la elaboración de este trabajo.

Al **Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) de la UNAM** que por medio de los proyectos con clave PAPIIT IA210617 y clave PAPIIT IA208019, permitieron realizar esta investigación

A mi tutor el Dr. Luis Eduardo Servín Garcidueñas, por su constante disposición a apoyarme durante todo el proceso de mi maestría y de la elaboración de este trabajo.

A la Dra. Yunuen Tapia Torres y a la Dra. María Esperanza Martínez Romero, quienes junto con el Dr. Servín, formaron parte de mi Comité Tutor y estuvieron siempre dispuestas a la constante evaluación y seguimiento en la elaboración de esta tesis.

Agradecimientos personales

Agradezco de primera mano a mi tutor el Dr. Luis Eduardo Servín Garcidueñas, por su permanente disposición de asesorarme y de ser una mano amiga durante mi estancia en la UNAM y en México.

Al Laboratorio Nacional de Análisis y Síntesis Ecológica (LANASE) por permitirme el uso de sus herramientas computacionales para realizar los análisis de la tesis consignada en este texto.

A los miembros del jurado por su disposición para revisar y evaluar el trabajo aquí presentado.

A mi familia, mis colegas y amigos tanto en México como en Colombia, que siempre me han apoyado, tanto en el desarrollo de la maestría como de este trabajo y creyeron en mi capacidad para sacar este estudio adelante. A pesar de la distancia y las circunstancias adversas que se presentaron durante el transcurso de estos dos años.

ÍNDICE

RESUMEN	1
ABSTRACT	3
1. INTRODUCCIÓN.....	4
1.1 Antecedentes	6
1.2 Justificación	8
2. HIPÓTESIS.....	9
3. OBJETIVOS	9
3.1 General	9
3.2 Específicos	9
4. METODOLOGÍA.....	10
4.1 Obtención de muestras ambientales	10
4.2 Purificación y secuenciación de ADN metagenómico	11
4.3 Ensamble y anotación de metagenomas	11
4.4 Análisis de diversidad taxonómica de los metagenomas con base en genes ribosomales 16S rARN.	12
4.5 Recuperación de Genomas	12
4.6 Anotación de genomas e identificación taxonómica.....	13
5. RESULTADOS	14
5.1 Secuenciación de metagenomas.....	14
5.2 Ensamble de metagenomas.....	14
5.3 Fumarolas de Los Azufres (FLA)	15
5.3.1 Estructura y diversidad.....	15
5.3.2 Diversidad funcional.....	17
5.3.3 Diversidad taxonómica	25
5.3.4 Recuperación de genomas	28
5.3.4.1 Análisis genómico comparativo	28
‘ <i>Candidatus Acidibrevibacterium</i> sp. FLA01’	28
‘ <i>Candidatus Aciditerrimonas</i> sp. FLA02’	31
5.3.4.2 Construcción de genomas a partir de metagenomas (MAG) via binning	34
5.4 Manantiales de Maritaro (MM).....	36
5.4.1 Estructura y diversidad.....	36
5.4.2 Diversidad funcional.....	38

5.4.3	Diversidad taxonómica	48
5.4.4	Recuperación de genomas	51
5.4.4.1	Análisis genómico comparativo	51
	‘ <i>Candidatus Acidobacterium</i> sp. MM01’	52
	‘ <i>Candidatus Metallibacterium</i> sp. MM02’	55
5.4.4.2	Construcción de genomas a partir de metagenomas (MAG) vía binning	58
6.	DISCUSIÓN.....	60
6.1	Fumarolas de los Azufres (FLA).....	60
6.1.1	Diversidad funcional.....	62
6.1.2	Genomas recuperados	65
	‘ <i>Candidatus Acidibrevibacterium</i> sp. FLA01’	65
	‘ <i>Candidatus Aciditerrimonas</i> sp. FLA02’	67
6.2	Manantiales de Maritaro (MM).....	68
6.2.1	Diversidad funcional.....	69
6.2.2	Genomas recuperados	71
	‘ <i>Candidatus Acidobacterium</i> sp. MM01’	71
	‘ <i>Candidatus Metallibacterium</i> sp. MM02’	72
6.3	Consideraciones derivadas de los metagenomas de FLA y MM.....	74
7.	CONCLUSIONES	77
8.	PERSPECTIVAS.....	78
9.	REFERENCIAS BIBLIOGRÁFICAS	78
10.	ANEXOS	91
10.1	Otros grupos de organismos detectados en los metagenomas.....	91
10.2	Ensamblaje de Metagenomas.....	92
10.3	Estructura y diversidad de los metagenomas	93
10.4	Recuperación de genomas	93
10.5	Scripts usados en el estudio	100

RESUMEN

Los microorganismos termófilos conocidos por vivir en ambientes con altas temperaturas presentan una amplia diversificación debido a la estrecha relación con los ambientes en los que se encuentran. Siendo un ejemplo son los ambientes geotérmicos, como es el campo geotérmico Los Azufres. Debido a las limitaciones para poder cultivarlos en laboratorio, los estudios metagenómicos han sido una de las principales herramientas para comprender la evolución e interacciones de estos microorganismos como también para comprender el origen de la vida en la Tierra. En el desarrollo de este estudio, estas herramientas fueron aplicadas para realizar un análisis de la diversidad de bacterias y la recuperación de genomas bacterianos de diversos ambientes en el campo geotérmico de Los Azufres. Para ello, se realizaron muestreos de sedimentos en fumarolas cercanas al spa de Los Azufres (FLA) y los manantiales hidrotermales de Maritaro (MM). Se usaron técnicas de secuenciación masiva de ADN ambiental y análisis bioinformáticos con el fin de recuperar genomas de bacterias extremófilas poco estudiadas presentes en este campo geotérmico. Como resultado, se identificaron varios géneros de bacterias relacionados con los filos *Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, *Acidobacteria*, y *Nitrospirae* tanto para FLA como para MM. Adicionalmente, se detectaron comunidades de arqueas y eucariontes previamente reportadas en FLA y comunidades de eucariontes poco conocidas en MM. Utilizando genómica comparativa, se recuperaron dos genomas relacionados con los géneros *Acidibrevibacterium* y *Aciditerrimonas* en FLA y dos genomas relacionados con los géneros *Acidobacterium* y *Metallibacterium* en MM. Además, se reconstruyeron 38 genomas a partir de metagenomas (MAGs) para FLA y 44 MAGs para MM. A pesar de las diferencias a nivel taxonómico, se encontraron grandes similitudes a nivel funcional entre FLA y MM, donde se hallaron adaptaciones de resistencia al choque térmico, metales pesados, biosíntesis y resistencia a diferentes compuestos antimicrobianos. Esto nos indica que las capacidades funcionales se han conservado dentro de distintas comunidades bacterianas de Los Azufres. Este trabajo provee información para comprender mejor las

adaptaciones de las bacterias que residen en ambientes extremos, posiblemente análogos a las condiciones de la Tierra primitiva.

ABSTRACT

Thermophile microorganisms occurring in high temperature environments, there are highly diverse because of their close relationship with the environment where they inhabit, such as is the case of the geothermal environments, being an example Los Azufres geothermal field. Due to culture-dependent methods limitations, metagenomic analysis represents a main tool to understand microbial evolution and interactions, as well as to understand the origin of life on Earth. In this study metagenomic approaches were employed for a several analyses of bacteria and to recover bacterial genomes from different environments within Los Azufres geothermal field. Sediment samples were obtained from the inside of steam vents nearby to Los Azufres Spa (FLA) and hot springs in the Maritaro area (MM). Massive environmental DNA sequencing techniques and bioinformatic analysis were used to recover genomes of minimally studied extremophilic bacteria present in this geothermal field. As a result, several bacterial genera related to the phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes*, *Acidobacteria*, and *Nitrospirae* were identified for both FLA and MM. Besides, previously reported archaea and eukaryotic communities and little-known eukaryotic communities were detected in FLA and MM, respectively. Using comparative genomics, we recovered two genomes related to the genera *Acidibrevibacterium* and *Aciditerrimonas* in FLA and two genomes related to the genera *Acidobacterium* and *Metallibacterium* in MM. In addition, we reconstructed 38 MAGs for FLA and 44 MAGs for MM from the obtained metagenomes. Despite differences at the taxonomic level, great similarities were found at the functional level between FLA and MM, where adaptations of resistance to heat shock, heavy metals, biosynthesis and resistance to different antimicrobial compounds were found. This could indicate that functional capacities have been conserved within different bacterial communities of Los Azufres. This work provides information to understand better the adaptations of bacteria that reside in extreme environments, which may be analogous to the conditions found in the primitive Earth.

1. INTRODUCCIÓN

Los microorganismos termófilos son aquellos que requieren de altas temperaturas para vivir (desde 50°C a 80°C) en esta categoría se encuentran también los hipertermófilos, los cuales llegan a vivir en temperaturas desde 80°C hasta 121°C. Además de esta característica, estos organismos han desarrollado una resistencia a otro tipo de condiciones, tales como valores de pH ácidos y alcalinos extremos y tolerancia a elementos tóxicos como metales pesados. Estos microorganismos pueden ser encontrados en algunos sitios volcánicos donde se presentan manifestaciones geotérmicas, incluyendo; manantiales termales, géiseres, fumarolas y volcanes de fango. Dichas manifestaciones presentan generalmente condiciones extremas de acidez y altas temperaturas (Kawarabayasi, 2013; V. Kumar & Satyanarayana, 2013; Merino et al., 2019; Satyanarayana et al., 2005; Urbietta et al., 2015).

Al ser considerados como ambientes extremos, el estudio de su microbiota nativa por medio de técnicas dependientes de cultivo es limitado para estimar la diversidad real presente en estos lugares (Handelsman, 2005; Moroenyane & Yergeau, 2019; Panda et al., 2018). Una solución a este problema han sido los estudios metagenómicos acoplados a la secuenciación masiva, los cuales permiten la secuenciación de grandes cantidades de ADN a partir de muestras ambientales. Este enfoque no depende de la realización de cultivos de manera convencional, por lo que su implementación ha permitido una interpretación más amplia de la diversidad microbiana presente en estos ambientes (Escobar-Zepeda et al., 2015; Kunin et al., 2008; Olson et al., 2018). Ejemplos de este tipo de estudios fueron los realizados por Baker y Banfield, (2003) en sedimentos de drenajes ácidos de minas, en el cual se recuperaron genomas relacionados con los géneros *Leptospirillum* y *Ferroplasma*. Esto abrió las puertas a estudios más profundos de la diversidad de ambientes considerados recónditos, en los cuales se asumía la existencia de una diversidad microbiana escasa. De esta manera, los estudios metagenómicos han permitido describir funciones y adaptaciones de comunidades microbianas en ambientes generalmente

hostiles, como es el caso del estudio de Guazzaroni et al., (2013) en el cual se identificaron genes de resistencia a la acidez en comunidades microbianas provenientes de Rio Tinto en España. El trabajo de Lau et al., (2014) en ambientes subterráneos profundos en Sudáfrica, se hallaron genes involucrados en el ciclo del nitrógeno (*nif*, *nar*) replanteando la relevancia de las comunidades microbianas presentes en estos ambientes en los ciclos biogeoquímicos. La investigación de Klatt et al., (2011) en muestras de tapetes microbianos provenientes de afluentes hidrotermales localizados en el Parque Nacional Yellowstone, lograron recuperar genomas de microorganismos cloro-fotótrofos anoxigénicos pertenecientes al filum *Chloroflexi*.

Los ambientes geotérmicos no han estado exentos de las limitaciones relacionadas con las técnicas dependientes de cultivo. Por lo que los primeros análisis de diversidad microbiana en manifestaciones geotérmicas utilizaron métodos independientes de cultivo, en los que se secuenciaron y se analizaron genes de subunidades ribosomales directamente de muestras de ADN ambiental y se identificaron microorganismos poco conocidos pertenecientes a los dominios Bacteria y Arquea (Huber et al., 2000). De acuerdo con Strazzulli et al. (2017) el uso de tecnologías de secuenciación masiva han permitido acceder a la composición genética de las comunidades microbianas y su función, otorgando una descripción filogenética más profunda de la que ofrecen otras tecnologías como la amplificación de marcadores moleculares como el gen ribosomal 16S rARN. Con el paulatino desarrollo y maduración de las tecnologías de secuenciación masiva, ha sido posible el incremento del estudio del microbioma en este tipo de ecosistemas en diferentes partes del mundo tales como: el Parque Nacional Yellowstone en EEUU, zonas geotérmicas en Islandia, la península de Kamchatka en Rusia, el valle Furnas localizado en el archipiélago de Azores, los yacimientos termales de Sungai Klah en Malasia, la zona termal de Tenchong en China, los campos de Phlegraean en Italia, el parque volcánico de Lassen en EEUU, las

ventilas termales en aguas someras de Tor Caladra y la isla Vulcano en Italia (Antranikian et al., 2017; Patwardhan et al., 2021) y la isla Kueishantao en Taiwan (Tang et al., 2018).

En México, se encuentra el Eje Volcánico TransMexicano, una cordillera rocosa que une las sierras Madre Occidental y Oriental y que se extiende desde el Estado de Veracruz hasta el Estado de Nayarit. Dentro de esta extensión orográfica, existen diferentes lugares con manifestaciones termales entre las que están: el campo Los Géiseres ubicado en el valle de Querétaro (González-Guzmán et al., 2019), el campo geotérmico de Los Humeros localizado en el estado de Puebla (Carrasco-Núñez et al., 2017), los manantiales termales ubicados en Araró (Prieto-Barajas, Alcaraz, et al., 2018) La Escalera, Agua Caliente-Tzitzio (Jácome-Paz, 2019) y el campo geotérmico de Los Azufres, ubicados en el Estado de Michoacán, México.

El campo geotérmico de Los Azufres posee una gran cantidad de manifestaciones geotérmicas de origen natural. Principalmente se encuentran manantiales termales y fumarolas que pueden alcanzar temperaturas mayores a los 90 °C y un pH menor a 4.0 (Arriaga Suarez et al., 2000), en los que se acumulan sedimentos que como se hablará más adelante, y las cuales albergan comunidades microbianas. De acuerdo con (Damer & Deamer, 2020) los sitios los lugares con alta actividad volcánica como Los Azufres son similares a los encontrados en los primeros años de la tierra, debido a la presencia de fuentes hidrotermales con altas temperaturas y valores de pH entre 2 y 5.

1.1 Antecedentes

Se han realizado estudios microbiológicos en Los Azufres a través de microscopía electrónica de barrido o mediante pruebas bioquímicas para determinar el tipo de microorganismos presentes sobre metales corroídos y dentro de tuberías empleadas para la generación de energía geotérmica. En estos estudios se evidenció la presencia de arqueas y bacterias (Alfaro-Cuevas-Villanueva et al., 2006; Castorena et al.,

2006; Navarrette-Bedolla et al., 1999; Torres-Sanchez et al., 1996; Valdez Salas et al., 2000). Años más tarde, Brito et al. (2014) analizó la diversidad microbiana presente en muestras de lodos ácidos de Los Azufres. Brito et al. (2014) obtuvieron y analizaron secuencias de genes ribosomales 16S rARN y se detectaron poblaciones de los géneros *Acidithiobacillus*, y *Desulfurella*, así como los fila *Cyanobacteria*, *Firmicutes*, *Acidobacteria*, *Thermotogae* junto a secuencias relacionadas con *Proteobacteria*.

Recientemente, se han elaborado investigaciones de diversidad microbiana utilizando enfoques metagenómicos de manantiales termales de Los Azufres. Derivado de estos estudios, se logró obtener el genoma consenso casi completo de una población de arqueas no caracterizada y relacionada filogenéticamente con arqueas pertenecientes al orden *Sulfolobales* (Servín-Garcidueñas & Martínez-Romero, 2014). En estos trabajos se secuenciaron de muestras de ADN ambiental sobrepasando las limitantes inherentes a la amplificación de genes ribosomales y de cultivo microbiano. Fue por medio de este estudio metagenómico que se logró la detección de bacterias poco estudiadas y pertenecientes a linajes filogenéticamente novedosos en manifestaciones termales de Los Azufres.

El estudio de Servín-Garcidueñas y Martínez-Romero (2014) también analizó la diversidad microbiana de muestras de sedimentos circundantes a fumarolas del mismo campo geotérmico. En las muestras de sedimentos se detectaron comunidades microbianas integradas por abundantes poblaciones de bacterias. Recientemente, también se reportó la obtención de genomas y la versatilidad metabólica de poblaciones poco estudiadas de arqueas acidófilas a partir de datos metagenómicos (Chen et al., 2018). Entre las principales se encuentran genomas pertenecientes al género *Cuniculiplasma* (Marín-Paredes & Servín-Garcidueñas, 2020).

Como antecedente directo a este estudio, en el Laboratorio de Microbiómica de la ENES-Morelia se han logrado identificar bacterias termoacidófilas a partir de tres metagenomas de sedimentos termales del campo geotérmico de Los Azufres. En estos metagenomas se han detectado poblaciones de bacterias

pertenecientes a los géneros *Leptospirillum*, *Alicyclobacillus* y *Sulfobacillus* mediante la asignación taxonómica de genes ribosomales. Estos géneros bacterianos son comunes en sitios termales con temperatura elevada y valores de pH bajo. Sin embargo, y de forma relevante, las poblaciones de bacterias más abundantes pertenecieron a grupos filogenéticos que no han sido descritos, que han sido poco estudiados, o que cuentan con información genómica limitada.

1.2 Justificación

A la fecha, no se han obtenido ni analizado genomas parciales o completos de bacterias de sedimentos superficiales provenientes de fumarolas y manantiales termales del campo geotérmico de Los Azufres. Los análisis metagenómicos que se han realizado se han enfocado en analizar la diversidad taxonómica y el potencial genético de arqueas, a pesar del conocimiento previo de la existencia de comunidades bacterianas en este ambiente.

Las comunidades bacterianas procedentes de ambientes hidrotermales pueden ofrecer un gran potencial biotecnológico, especialmente en ramas como la industria farmacéutica y la biorremediación, ya que pueden ser fuente de obtención de nuevos compuestos antimicrobianos y tienen la capacidad de asimilar o secuestrar elementos tóxicos para otros organismos (Prieto-Barajas, Valencia-Cantero, et al., 2018; Ranawat & Rawat, 2018; Tulasi Satyanarayana et al., 2013). También se ha propuesto su uso en procesos de fermentación y degradación de materia orgánica a altas temperaturas (G.-Q. Chen & Jiang, 2018).

Adicionalmente, en los últimos años, el estudio de este tipo de ambientes ha sido de gran interés para la astrobiología, puesto que se considera que, en los primeros años de la formación de la Tierra, pudo existir una amplia cantidad de ambientes geotérmicos tanto terrestres como marinos, en los cuales pudo llevarse a cabo el origen de los primeros organismos vivos. Razón por la cual, se considera que los

ambientes geotérmicos terrestres actuales pueden ser un vestigio de cómo fueron las primeras actividades bióticas en la tierra (Cavalazzi et al., 2018; Damer & Deamer, 2020). Además, tras los hallazgos recientes de rastros de actividad geotérmica en la edad temprana de Marte (Ruff et al., 2020), se ha propuesto el uso de estos ambientes como posibles análogos para el descubrimiento de biofirmas que ayuden a determinar si hubo actividad biológica en los ambientes geotérmicos del planeta rojo durante sus primeros años de formación (Cockell et al., 2019; Rodríguez & Van Bergen, 2016).

2. HIPÓTESIS

Bajo las condiciones ambientales adversas propias de los sedimentos termales de Los Azufres, las comunidades bacterianas presentes poseen atributos funcionales potenciales de resistencia a condiciones ambientales extremas; como temperaturas mayores e iguales a 40 °C, valores de pH inferiores a 5, tolerancia a metales pesados y mecanismos alternativos de obtención de energía, lo que favorece la presencia y abundancia de comunidades bacterianas endémicas en sedimentos termales.

3. OBJETIVOS

3.1 General

Analizar la diversidad bacteriana presente en sedimentos termales colectados de áreas poco exploradas del campo geotérmico de Los Azufres, así como su metabolismo potencial a través de enfoques genómicos y metagenómicos.

3.2 Específicos

1. Realizar el ensamble y la anotación de las secuencias genómicas y metagenómicas con énfasis en la obtención de genomas bacterianos.
2. Realizar un perfil taxonómico de las comunidades bacterianas presentes en las muestras de sedimentos termales.

3. Determinar y analizar el metabolismo potencial contenido en las secuencias genómicas y metagenómicas.
4. Identificar genes bacterianos que sirvan como marcadores clave para determinar posibles mecanismos de resistencia a las condiciones extremas que presentan los sedimentos.
5. Analizar y comparar ensambles de genomas obtenidos a partir de las secuencias genómicas y metagenómicas de los sedimentos termales contra genomas de referencia utilizando genómica comparativa.

4. METODOLOGÍA

4.1 Obtención de muestras ambientales

Se colectaron dos muestras de sedimentos termales superficiales que se acumulan sobre rocas alrededor de fumarolas del campo geotérmico Los Azufres. En la zona del balneario de Los Azufres 19.7816° N 100.6577° W (56°C , pH 3.8) (**Figura 1**) (FLA) y en los Manantiales termales de Maritaro 19.8221° N 100.6681° W (43°C , pH 4.5) (**Figura 2**) (MM). Los sedimentos se tomaron frotando cajas Petri estériles sobre las rocas.

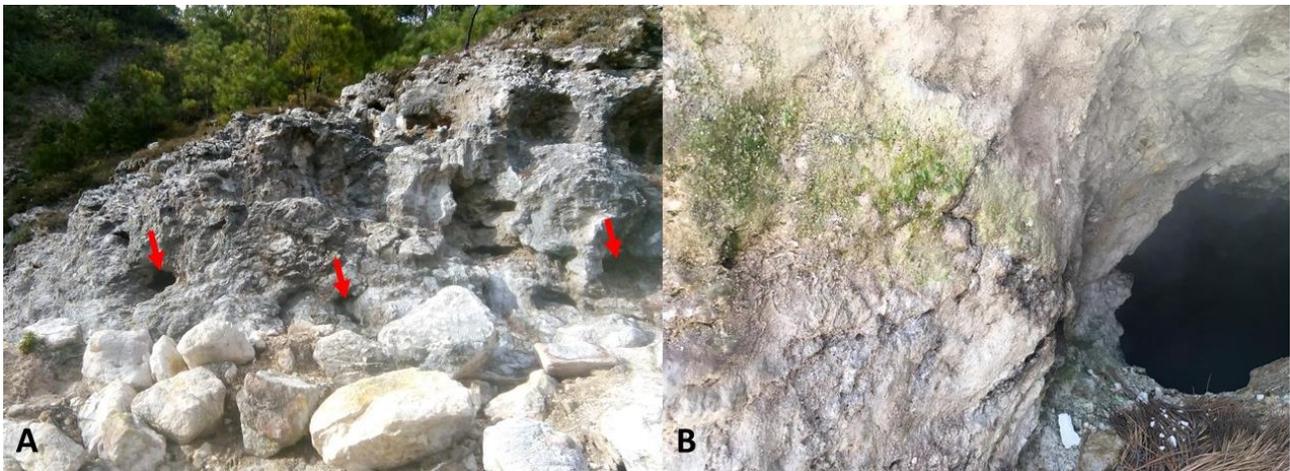


Figura 1. Fumarolas cercanas al balneario de Los Azufres (FLA). A) Fumarolas activas ubicadas en la zona de muestreo (Flechas rojas) B) Sedimentos microbianos presentes en la salida de una de las fumarolas.

4.2 Purificación y secuenciación de ADN metagenómico

Se extrajo y purificó ADN mediante el kit comercial *PowerMax Soil Kit* (QIAGEN, Carlsbad, CA) siguiendo las instrucciones del fabricante. Las muestras de ADN ambiental se secuenciaron en la compañía MacroGen Inc., en Corea del Sur utilizando la técnica de secuenciación masiva de Illumina HiSeq 2500 para obtener lecturas pareadas de 250 pares de bases.



Figura 2. Manantiales termales de Maritaro (MM). A) Arroyo termal que corre en la zona. B y C) sedimentos microbianos encontrados en zonas aledañas al arroyo los cuales fueron colectados para el presente estudio.

4.3 Ensamble y anotación de metagenomas

Las lecturas pareadas de cada metagenoma fueron evaluadas usando FastQC v. 0.11.8 (Brown et al., 2017). Posteriormente se filtraron las lecturas por calidad usando Trimmomatic v. 0.39 (Bolger et al.,

2014). El ensamble de las lecturas metagenómicas en *contigs* se realizó con SPAdes v.3.8.2 (Bankevich et al., 2012). Posterior a ello, se realizó una comparación de los diferentes *contigs* ensamblados a través del programa QUAST v.4.4 (Gurevich et al., 2013). Por último, la anotación de los *contigs* metagenómicos se realizó en las plataformas en línea de MG-RAST v.4.0.3 (Meyer et al., 2008) e IMG/ER (I. M. A. Chen et al., 2017).

4.4 Análisis de diversidad taxonómica de los metagenomas con base en genes ribosomales 16S rARN

Los *contigs* metagenómicos resultantes del ensamblaje de las lecturas fueron procesados con la versión en línea de comandos de RNAmmer v.1.2 (S. Wu et al., 2011). Para identificar, extraer y curar las secuencias de genes ribosomales 16S rARN de bacterias. Por último, se procedió a la identificación taxonómica de las secuencias ribosomales 16S rARN obtenidas de los *contigs* metagenómicos por medio de la plataforma de BLASTN en GenBank usando la base de datos de la colección de nucleótidos no redundante nr/nt y por medio de la plataforma SILVAngs (Quast et al., 2012).

4.5 Recuperación de Genomas

Tras la identificación taxonómica, se llevaron a cabo dos vías para la recuperación de genomas: Primeramente se hizo la búsqueda de genomas de referencia en GenBank a partir de genes ribosomales 16S rARN teniendo en cuenta su alto porcentaje de identidad y la cantidad de genomas reportados en esta base de datos. Se seleccionaron los genomas de *Acidibrevibacterium fodinaquatile* G 45-3 con número de acceso: GCA003352165, *Aciditerrimonas ferrireducens* JCM15389 con número de acceso: GCA001311945, para el metagenoma de FLA y *Acidobacterium capsulatum* ATCC51196 con número de acceso: GCF000022565, *Metallibacterium scheffleri* DKE6 con número de acceso: GCF002077135, para el metagenoma de MM. Posteriormente, se llevaron a cabo alineamientos usando el programa Mauve v.2015.02.13 (Darling et al., 2004) con el fin de determinar la similitud de los *contigs* con los

genomas de referencia, los genomas que tuvieron una alta similitud con los *contigs* del metagenoma se usaron para realizar mapeos con las lecturas del metagenoma usando Bowtie2 v.2.3.5 (Langmead et al., 2009) las lecturas mapeadas fueron extraídas usando Samtools v.1.7 (Li et al., 2009). Finalmente, las lecturas fueron nuevamente ensambladas usando SPAdes y se corroboraron por genómica comparativa a partir de alineamientos en Mauve.

Segundo, por medio de la recuperación de genomas ensamblados a partir de metagenomas (Metagenome Assembled-Genome MAG) usando programas como Maxbin v.2.2 (Y.-W. Wu et al., 2015), MetaBAT v.2.12.1 (Kang et al., 2015) y CONCOCT v. 1.1.0 (Alneberg et al., 2014), los MAGs recuperados de los tres programas fueron filtrados usando DASTool v.1.1.2 (Sieber et al., 2018). Se examinó su completitud y porcentaje de contaminación con CheckM v.1.0.13 (Parks et al., 2015). Finalmente se realizó un perfil taxonómico de los MAGs a través de la herramienta CAT/BAT v.5.1.2 (Von Meijenfeldt et al., 2019). La cual usa DIAMOND (Buchfink et al., 2014) y Prodigal (Hyatt et al., 2010) como herramientas de predicción y búsqueda de genes para así ejecutar su algoritmo correctamente.

4.6 Anotación de genomas e identificación taxonómica

La anotación de los genomas recuperados se realizó a partir de la plataforma RAST (Aziz et al., 2008; Brettin et al., 2015; Overbeek et al., 2014). Las anotaciones fueron visualizadas por medio del programa ARTEMIS v.16.0.17 (Carver et al., 2011). La determinación de similitud de los genomas recuperados con su respectivo genoma de referencia se obtuvo por medio del uso de diferentes métricos de identidad de secuencia tales como; el cálculo de promedio de identidad de nucleótido (Average Nucleotide Identity ANI) y por hibridación digital de ADN-ADN (*in silico* DNA-DNA hybridization DDH) por medio de las plataformas JspeciesWS (Richter et al., 2016), y Genome-to-Genome Distance Calculator (GGDC) (Meier-Kolthoff et al., 2013). Los estudios filogenómicos se realizaron con el uso de la

herramienta PhyloPhlAn v.3.0 (Segata et al., 2013), la visualización de los árboles se hizo por medio del uso del software MEGAX v. 10.1.8 (S. Kumar et al., 2018).

5. RESULTADOS

5.1 Secuenciación de metagenomas

se obtuvieron dos metagenomas del procesamiento de las muestras ambientales de sedimentos termales de Los Azufres (**Tabla 1**). Teniendo en cuenta la información procedente del laboratorio de secuenciación y la generación de las lecturas, es posible determinar que en general un alto porcentaje de las lecturas poseen una alta calidad ($Q \Rightarrow 30$).

Tabla 1. Características de los metagenomas obtenidos de los sitios muestreados de los azufres.

Muestra	Bases totales	Lecturas totales	%GC	%AT	%Q20	%Q30
Fumarolas Los Azufres (FLA)	21,579,531,228	85,974,228	52.547	47.45	96.375	93.135
Manantiales de Maritaro (MM)	22,618,252,560	90,112,560	58.933	41.07	95.393	91.010

5.2 Ensamble de metagenomas

De los tres ensamblajes realizados para el metagenoma de FLA (**Tabla 2**), el ensamblaje hecho en SPAdes con los k-meros 21, 33, 55, 77, 99, fue el seleccionado de acuerdo con sus valores de número de *contigs*, secuencia y valor de N50. El ensamblaje seleccionado contiene 177,623 *contigs*, que representan 215,215,420 pb de secuencia. El valor de N50 del metagenoma es de 6,519 pb. Por otra parte, para el metagenoma de la muestra MM, de los 3 ensamblajes realizados (**Tabla 2**, **Tabla S4**) se seleccionó el que fue elaborado con SPAdes con los k-meros 21, 33, 55, 77, 99 debido a sus valores de número de *contigs*, número total de secuencia ensamblado y de N50. Este ensamblaje posee un total de 391,625 *contigs* que representan 486,892,674 pb de secuencia y cuyo valor de N50 es de 5,421 pb.

5.3 Fumarolas de Los Azufres (FLA)

5.3.1 Estructura y diversidad

El metagenoma de FLA está compuesto por bacterias, seguido de eucariotas y arqueas (**Figura 3A**). Los filos de bacterias más abundantes en el metagenoma corresponden a *Actinobacteria*, *Proteobacteria*, y *Firmicutes*, junto con la abundancia de filos de arqueas correspondientes a *Euryarchaeota* y *Crenarchaeota* (**Figura 3B**).

Tabla 2. Valores derivados de los ensamblajes de secuencias de los metagenomas provenientes de FLA y MM con el software de ensamblaje SPAdes, los valores fueron obtenidos al evaluar los *contigs* con QUAST. En rojo se resaltan los ensamblajes seleccionados.

Estadísticos	<i>Contigs</i> FLAK21_99	<i>Contigs</i> FLAK33_111	<i>Contigs</i> FLAK55_127	<i>Contigs</i> MMK21_99	<i>Contigs</i> MMK33_111	<i>Contigs</i> MMK55_127
K-meros	21, 33, 55, 77, 99	33, 55, 77, 99, 111	55, 77, 99, 111, 127	21, 33, 55, 77, 99	33, 55, 77, 99, 111	55, 77, 99, 111, 127
# <i>contigs</i> (>= 0 bp)	177,623	166,817	146,374	391,625	363,915	311,327
Longitud total (>= 0 bp)	215,215,420	211,084,537	199,739,366	486,892,674	474,025,919	440,966,402
N50	6,519	6,373	6,740	5,421	5,259	5,263
GC (%)	55.95	56.03	56.09	55.31	55.46	55.76

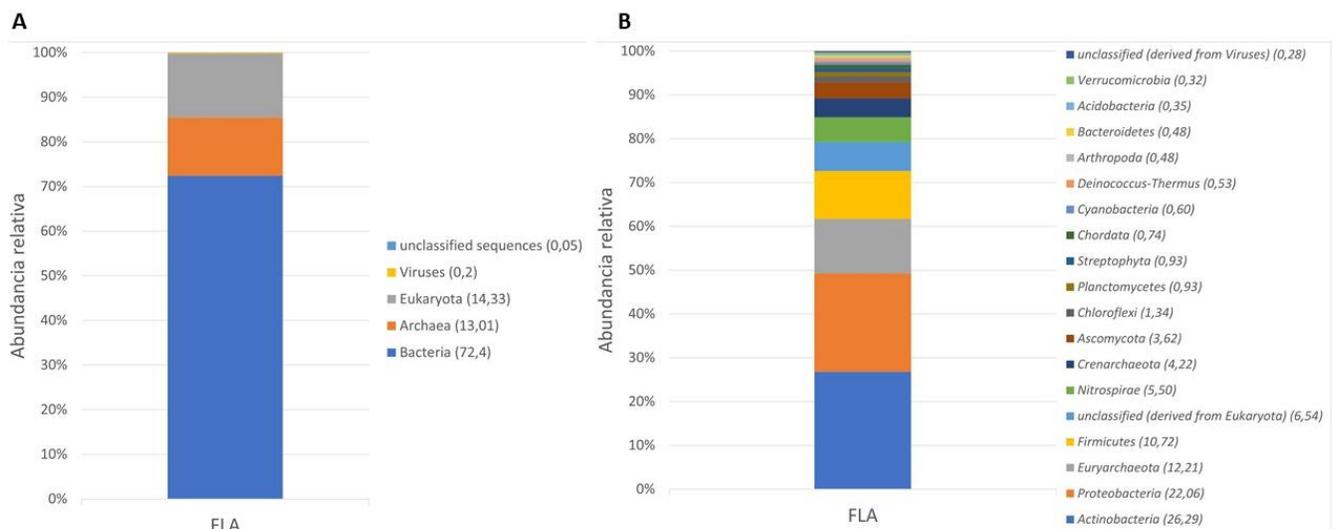


Figura 3. Proporción de grupos taxonómicos encontrados en el ensamblaje del metagenoma de FLA, obtenidos de la anotación en la plataforma MG-RAST usando el algoritmo de contigLCA (*LastCommonAncestor*).

A nivel de género, el análisis de MG-RAST muestra una abundancia de *Acidithiobacillus* (8.48 %), *Leptospirillum* (8.46%), *Mycobacterium* (4.39%), *Acidimicrobium* (3.9%), *Conexibacter* (2.04%), *Streptomyces* (2.94%), *Frankia* (2.0%), y *Alicyclobacillus* (1.69%), Mientras que los géneros más abundantes de arqueas están relacionados con *Ferroplasma* (5.89%), *Thermoplasma* (3.7%) y *Sulfolobus* (3.31%) (Figura 4).

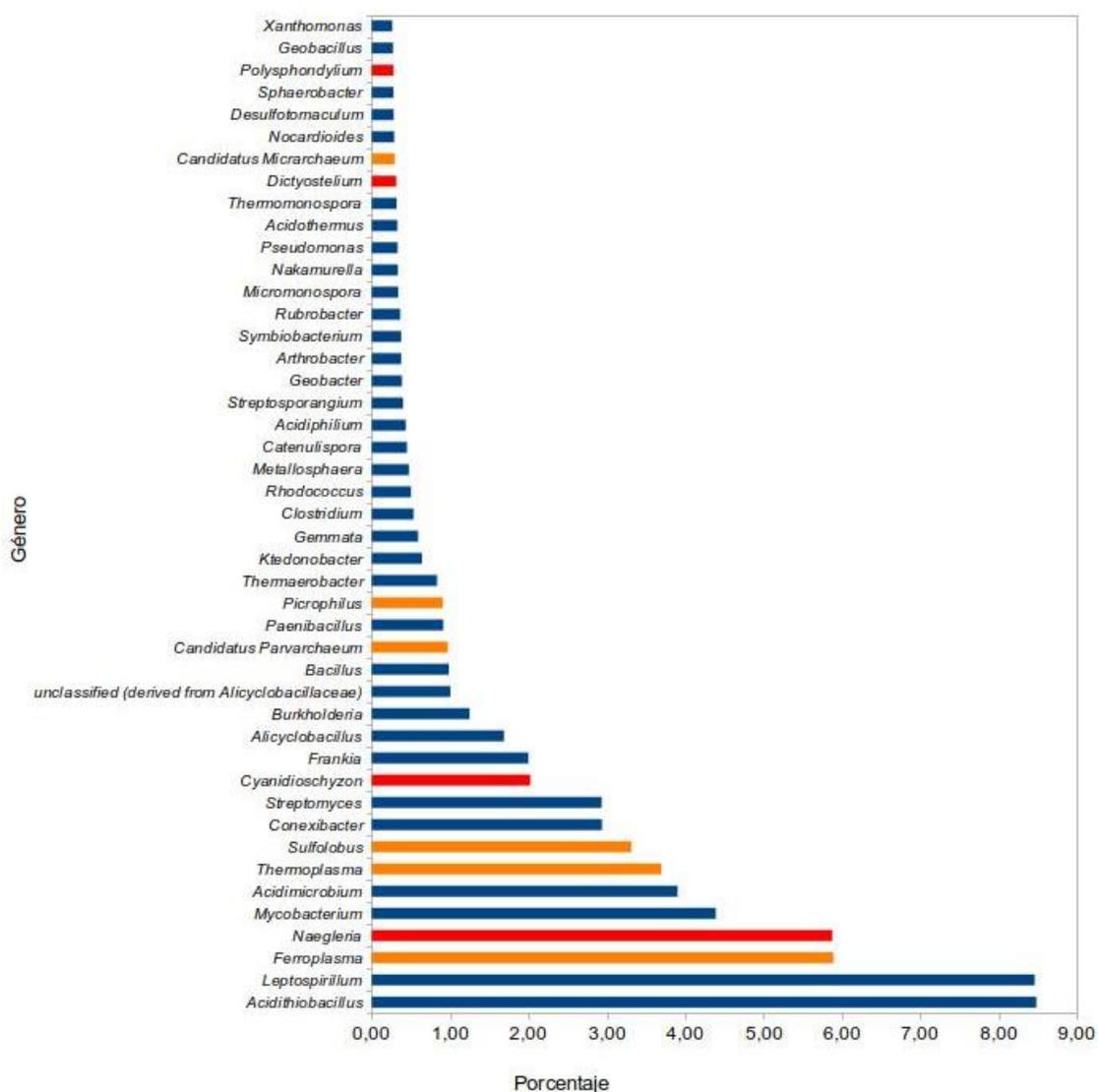


Figura 4. Abundancias de géneros de microorganismos encontrados en el metagenoma de FLA, resultado de la anotación en la plataforma de MG-RAST usando el algoritmo de contigLCA (*LastCommonancestor*). En azul se señalan los microorganismos pertenecientes al dominio Bacteria, en naranja el dominio Arquea y en rojo el dominio Eucariota.

5.3.2 Diversidad funcional

En cuanto a diversidad funcional, se encontró que el metagenoma de FLA posee un total de 316,278 genes codificantes para proteínas, de los cuales 173,778 tienen una anotación funcional conocida (54,3%). 3724 genes de ARN y 939 CRISPRs (**Tabla 3**). Dentro del total de genes codificantes fue posible hallar 165267 genes ortólogos los cuales se distribuyen en las funciones mencionadas en la **Tabla 4**.

Tabla 3. Características del metagenoma recuperado de las fumarolas de Los Azufres (FLA), obtenidos a partir de la anotación en la plataforma IMG.

Característica	N° de fragmentos de genes
CRISPR	939
Genes ARN	3,724
Genes rARN	327
5S rARN	69
16S rARN	97
18S rARN	13
23S rARN	124
28S rARN	24
Genes tARN	3,136
Genes codificantes de proteínas	316,278
Genes con nombre de producto	173,778

Tabla 4. Abundancia relativa de categorías de genes ortólogos (COG) encontrados en los metagenomas de Los Azufres (FLA) enumerados por número de genes y porcentaje frente al total del metagenoma.

Categoría COG	N° Genes	Porcentaje
Transporte y metabolismo de aminoácidos	18,473	9.50%
Transporte y metabolismo de carbohidratos	13,001	6.68%
Control del ciclo celular, división celular, partición cromosómica	2,728	1.40%
Motilidad celular	2,696	1.39%
Pared celular/membrana/biogénesis de envoltura	10,279	5.28%
Estructura y dinámica de la cromatina	148	0.08%
Coenzimas de transporte y metabolismo	9,948	5.11%
Citoesqueleto	202	0.10%
Mecanismos de defensa	5,535	2.85%

Tabla 4. Abundancia relativa de categorías de genes ortólogos (COG) encontrados en los metagenomas de Los Azufres (FLA) enumerados por número de genes y porcentaje frente al total del metagenoma (continuación).

Categoría COG	N° Genes	Porcentaje
Producción y conversión de energía	12,704	6.53%
Estructuras extracelulares	974	0.50%
Función desconocida	7,951	4.09%
Función general	21,686	11.15%
Transporte y metabolismo de iones inorgánicos	10,021	5.15%
Trafico extracelular, secreción y transporte vesicular	2,787	1.43%
Transporte y metabolismo de lípidos	8,939	4.60%
Mobiloma: profagos, transposones	6,604	3.39%
Transporte y metabolismo de nucleótidos	4,912	2.52%
Modificación postranscripcional, recambio proteico, chaperonas	7,619	3.92%
Procesamiento y modificación de ARN	112	0.06%
Replicación recombinación y reparación	10,027	5.15%
Biosíntesis de metabolitos secundarios, transporte y catabolismo	5,553	2.85%
Mecanismos de transducción de señal	7,223	3.71%
Transcripción	10,956	5.63%
Traducción, estructura ribosomal y biogénesis	13,459	6.92%

Se observó que la mayoría de los genes detectados (**Tabla 4**), están concentrados en determinadas taxa de microorganismos. En el caso de FLA la mayoría de los grupos de genes están concentrados en el grupo *Euryarcheota*, seguido de *Proteobacteria* y *Nitrospirae*, y en menor medida en los grupos *Firmicutes*, *Actinobacteria*, *Crenarcheota* y *Ascomycota* (**Figura 5**).

Al observar de manera detallada los resultados de la anotación, se encontraron genes clave para ciclos biogeoquímicos dentro de las funciones metabólicas de la comunidad (**Tabla 5**). Dentro de los principales se destacan genes relacionados con la fijación de carbono (asociados a las rutas de CBB y RTCA), genes involucrados en el metabolismo del nitrógeno que intervienen en procesos de desnitrificación, principalmente en la respiración de nitrato y nitrito. Para el metabolismo del fósforo fue posible encontrar diversos transportadores de fosfatos, genes encargados en la asimilación y en la

regulación ante la escasez de fósforo. También se detectaron genes relacionados con el metabolismo del azufre, donde se detectaron genes involucrados en la asimilación de sulfatos y con el sistema SOX y genes involucrados en el uso de sustratos inorgánicos como hierro.

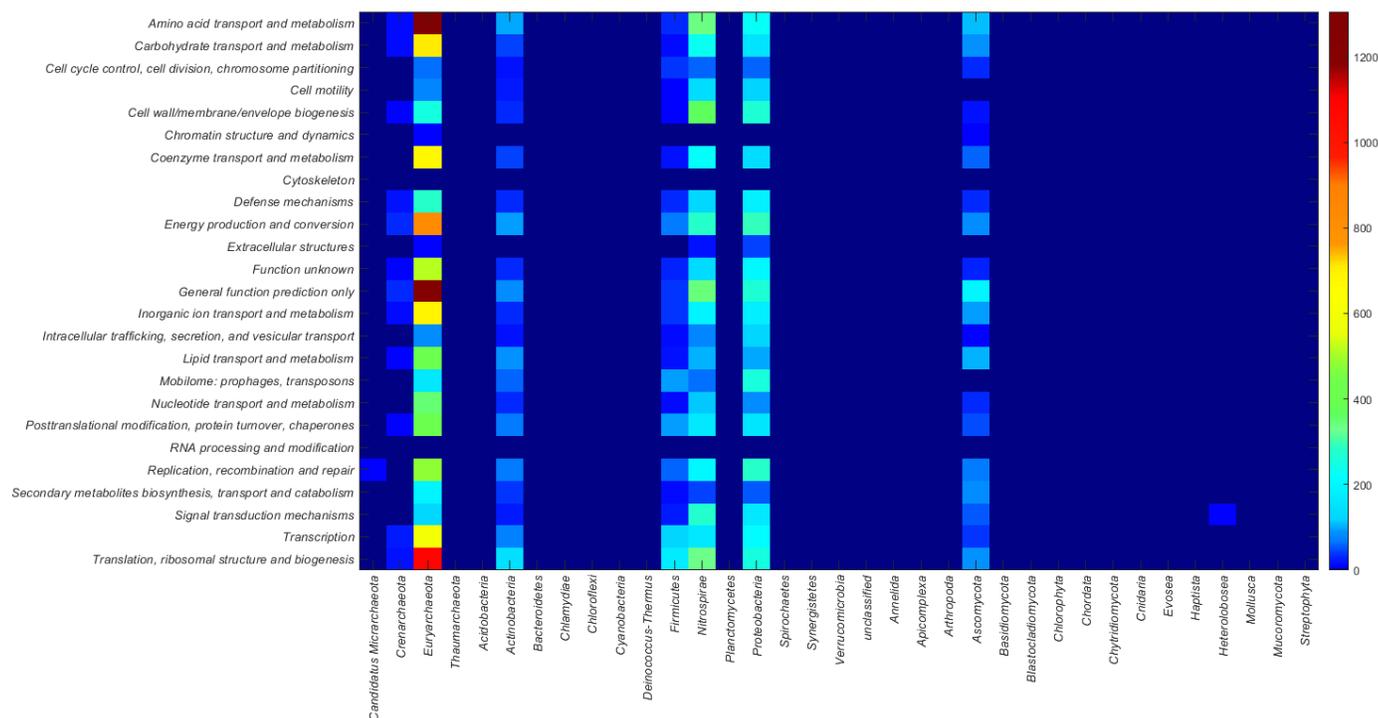


Figura 5. Mapa de calor que representa de forma general la diversidad funcional de metagenoma de FLA, las barras indican la cantidad de genes ortólogos identificados a través de la plataforma de IMG.

Tabla 5. Lista de genes relacionados con procesos metabólicos en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO).

Metabolismo	KO ID	Nombre	Producto	N° genes
Azufre	K17222	soxA	sulfur-oxidizing protein SoxA [EC:2.8.5.2]	5
	K17223	soxX	sulfur-oxidizing protein SoxX [EC:2.8.5.2]	6
	K17224	soxB	sulfur-oxidizing protein SoxB [EC:3.1.6.20]	7
	K17226	soxY	sulfur-oxidizing protein SoxY	5
	K17227	soxZ	sulfur-oxidizing protein SoxZ	5
	K00380	cysJ	sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2]	8
	K00381	cysI	sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2]	10
	K00390	cysH	phosphoadenosine phosphosulfate reductase [EC:1.8.4.8]	49

Tabla 5. Lista de genes relacionados con procesos metabólicos en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO) (continuación).

Metabolismo	KO ID	Nombre	Producto	N° genes
Azufre	K00860	<i>cysC</i>	<i>adenylylsulfate kinase</i> [EC:2.7.1.25]	9
	K00955	<i>cysNC</i>	<i>bifunctional enzyme CysN/CysC</i> [EC:2.7.7.4 2.7.1.25]	12
	K00956	<i>cysN</i>	<i>sulfate adenylyltransferase subunit 1</i> [EC:2.7.7.4]	3
	K00957	<i>cysD</i>	<i>sulfate adenylyltransferase subunit 2</i> [EC:2.7.7.4]	24
	K02045	<i>cysA</i>	<i>sulfate transport system ATP-binding protein</i> [EC:3.6.3.25]	1
	K02046	<i>cysU</i>	<i>sulfate transport system permease protein</i>	2
	K02047	<i>cysW</i>	<i>sulfate transport system permease protein</i>	1
	K02048	<i>cysP, sbp</i>	<i>sulfate transport system substrate-binding protein</i>	1
	K13811	PAPSS	<i>3'-phosphoadenosine 5'-phosphosulfate synthase</i> [EC:2.7.7.4 2.7.1.25]	1
	K00958	<i>sat, met3</i>	<i>sulfate adenylyltransferase</i> [EC:2.7.7.4]	58
	K00392	<i>sir</i>	<i>sulfite reductase (ferredoxin)</i> [EC:1.8.7.1]	20
	K00394	<i>aprA</i>	<i>adenylylsulfate reductase, subunit A</i> [EC:1.8.99.2]	11
	K00395	<i>aprA</i>	<i>adenylylsulfate reductase, subunit A</i> [EC:1.8.99.2]	7
Carbono (Fijación)	K01601	<i>rbcL</i>	<i>ribulose-bisphosphate carboxylase large chain</i> [EC:4.1.1.39]	33
	K01602	<i>rbcS</i>	<i>ribulose-bisphosphate carboxylase small chain</i> [EC:4.1.1.39]	19
	K00174	<i>korA, oorA, oforA</i>	<i>2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha</i> [EC:1.2.7.3 1.2.7.11]	102
	K00175	<i>korB, oorB, oforB</i>	<i>2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit beta</i> [EC:1.2.7.3 1.2.7.11]	83
	K00176	<i>korD, oorD</i>	<i>2-oxoglutarate ferredoxin oxidoreductase subunit delta</i> [EC:1.2.7.3]	1
	K00177	<i>korC, oorC</i>	<i>2-oxoglutarate ferredoxin oxidoreductase subunit gamma</i> [EC:1.2.7.3]	2
Nitrógeno	K00367	<i>narB</i>	<i>ferredoxin-nitrate reductase</i> [EC:1.7.7.2]	1
	K00370	<i>narG</i>	<i>nitrate reductase alpha subunit</i> [EC:1.7.99.4]	29
	K00371	<i>narH</i>	<i>nitrate reductase beta subunit</i> [EC:1.7.99.4]	19
	K00372	<i>nasA</i>	<i>assimilatory nitrate reductase catalytic subunit</i> [EC:1.7.99.4]	8
	K00373	<i>narJ</i>	<i>nitrate reductase delta subunit</i>	11
	K00374	<i>narI</i>	<i>nitrate reductase gamma subunit</i> [EC:1.7.99.4]	16
	K07684	<i>narL</i>	<i>two-component system, NarL family, nitrate/nitrite response regulator NarL</i>	2
	K10534	NR	<i>nitrate reductase (NAD(P)H)</i> [EC:1.7.1.1 1.7.1.2 1.7.1.3]	2
	K15576	<i>nrtA, nasF, cynA</i>	<i>nitrate/nitrite transport system substrate-binding protein</i>	6
	K00362	<i>nirB</i>	<i>nitrite reductase (NADH) large subunit</i> [EC:1.7.1.15]	7
	K00363	<i>nirD</i>	<i>nitrite reductase (NADH) small subunit</i> [EC:1.7.1.15]	3
	K00366	<i>nirA</i>	<i>ferredoxin-nitrite reductase</i> [EC:1.7.7.1]	10
	K00368	<i>nirK</i>	<i>nitrite reductase (NO-forming)</i> [EC:1.7.2.1]	6
	K02575	NRT, <i>narK, nrtP, nasA</i>	<i>MFS transporter, NNP family, nitrate/nitrite transporter</i>	10

Tabla 5. Lista de genes relacionados con procesos metabólicos en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO) (continuación).

Metabolismo	KO ID	Nombre	Producto	N° genes
Nitrógeno	K17877	NIT-6	<i>nitrite reductase</i> (NAD(P)H) [EC:1.7.1.4]	1
	K00491	<i>nos</i>	<i>nitric-oxide synthase, bacterial</i> [EC:1.14.13.165]	3
	K02164	<i>norE</i>	<i>nitric oxide reductase NorE protein</i>	1
	K04561	<i>norB</i>	<i>nitric oxide reductase subunit B</i> [EC:1.7.2.5]	4
	K04748	<i>norQ</i>	<i>nitric oxide reductase NorQ protein</i>	14
	K05916	<i>hmp, YHB1</i>	<i>nitric oxide dioxygenase</i> [EC:1.14.12.17]	5
	K13125	NOSIP	<i>nitric oxide synthase-interacting protein</i>	1
	K13771	<i>nsrR</i>	<i>Rrf2 family transcriptional regulator, nitric oxide-sensitive transcriptional repressor</i>	6
	K00376	<i>nosZ</i>	<i>nitrous-oxide reductase</i> [EC:1.7.2.4]	1
	K07218	<i>nosD</i>	<i>nitrous oxidase accessory protein</i>	1
	K19339	<i>nosR</i>	<i>NosR/NirI family transcriptional regulator, nitrous oxide reductase regulator</i>	1
Fósforo	K01077	E3.1.3.1, <i>phoA, phoB</i>	<i>alkaline phosphatase</i> [EC:3.1.3.1]	3
	K01078	E3.1.3.2	<i>acid phosphatase</i> [EC:3.1.3.2]	9
	K01113	<i>phoD</i>	<i>alkaline phosphatase D</i> [EC:3.1.3.1]	3
	K01507	<i>ppa</i>	<i>inorganic pyrophosphatase</i> [EC:3.6.1.1]	63
	K02036	<i>pstB</i>	<i>phosphate transport system ATP-binding protein</i> [EC:3.6.3.27]	30
	K02037	<i>pstC</i>	<i>phosphate transport system permease protein</i>	34
	K02038	<i>pstA</i>	<i>phosphate transport system permease protein</i>	33
	K02039	<i>phoU</i>	<i>phosphate transport system protein</i>	40
Fósforo	K02040	<i>pstS</i>	<i>phosphate transport system substrate-binding protein</i>	26
	K02445	<i>glpT</i>	<i>MFS transporter, OPA family, glycerol-3-phosphate transporter</i>	3
	K03306	TC.PIT	<i>inorganic phosphate transporter, PiT family</i>	51
	K08176	PHO84	<i>MFS transporter, PHS family, inorganic phosphate transporter</i>	44
	K14430	PHO87_91	<i>phosphate transporter</i>	1
	K06217	<i>phoH, phoL</i>	<i>phosphate starvation-inducible protein PhoH and related proteins</i>	46
	K06653	PHO81	<i>CDK inhibitor PHO81</i>	1
	K07175	<i>phoH2</i>	<i>PhoH-like ATPase</i>	14
	K07636	<i>phoR</i>	<i>two-component system, OmpR family, phosphate regulon sensor histidine kinase PhoR</i> [EC:2.7.13.3]	29
	K07637	<i>phoQ</i>	<i>two-component system, OmpR family, sensor histidine kinase PhoQ</i> [EC:2.7.13.3]	1
	K07657	<i>phoB</i>	<i>two-component system, OmpR family, phosphate regulon response regulator PhoB</i>	13
	K07658	<i>phoB1, phoP</i>	<i>two-component system, OmpR family, alkaline phosphatase synthesis response regulator PhoP</i>	10
	K07660	<i>phoP</i>	<i>two-component system, OmpR family, response regulator PhoP</i>	1
	K15640	<i>phoE</i>	<i>uncharacterized phosphatase</i>	1

Tabla 5. Lista de genes relacionados con procesos metabólicos en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO) (continuación).

Metabolismo	KO ID	Nombre	Producto	Nº genes
Hierro	K18683	<i>rus</i>	<i>rusticyanin</i>	1
	K02274	<i>coxA</i>	<i>cytochrome c oxidase subunit I</i> [EC:1.9.3.1]	80
	K02275	<i>coxB</i>	<i>cytochrome c oxidase subunit II</i> [EC:1.9.3.1]	56
	K02276	<i>coxC</i>	<i>cytochrome c oxidase subunit III</i> [EC:1.9.3.1]	21
	K04755	<i>fdx</i>	<i>ferredoxin, 2Fe-2S</i>	9
	K05524	<i>fdxA</i>	<i>ferredoxin</i>	13

Adicionalmente, se encontraron genes involucrados en diversos sistemas de resistencia a metales pesados, entre los que se hallaron genes relacionados con transportadores Ni/Cd, transportadores catiónicos Co/Zn/Cd, transportadores Zn/Cd, genes de resistencia a Ni, As, Cr, Cd y Hg, y genes de resistencia a múltiples metales (**Tabla 6**).

Tabla 6. Lista de genes relacionados con resistencia a metales pesados presentes en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO).

	KO ID	Nombre	Producto	Nº genes
Arsénico	K00537	ARSC1, <i>arsC</i>	<i>arsenate reductase</i> [EC:1.20.4.1]	3
	K03741	ARSC2, <i>arsC</i>	<i>arsenate reductase</i> [EC:1.20.4.1]	15
	K03893	<i>arsB</i>	<i>arsenical pump membrane protein</i>	92
	K11811	<i>arsH</i>	<i>arsenical resistance protein ArsH</i>	5
	K08356	<i>aoxB</i>	<i>arsenite oxidase large subunit</i> [EC:1.20.9.1]	1
	K08355	<i>aoxA</i>	<i>arsenite oxidase small subunit</i> [EC:1.20.9.1]	4
	K03325	TC.ACR3	<i>arsenite transporter, ACR3 family</i>	18
	K01551	<i>arsA</i> , ASNA1	<i>arsenite-transporting ATPase</i> [EC:3.6.3.16]	33
Cromo	K07240	<i>chrA</i>	<i>chromate transporter</i>	23
	K19784	<i>chrR</i>	<i>chromate reductase</i>	27
Mercurio	K00221	E4.99.1.2	<i>alkylmercury lyase</i> [EC:4.99.1.2]	5
	K00520	<i>merA</i>	<i>mercuric reductase</i> [EC:1.16.1.1]	102
	K08363	<i>merT</i>	<i>mercuric ion transport protein</i>	2
	K08364	<i>merP</i>	<i>periplasmic mercuric ion binding protein</i>	2
	K08365	<i>merR</i>	<i>MerR family transcriptional regulator, mercuric resistance operon regulatory protein</i>	9
	K19058	<i>merC</i>	<i>mercuric ion transport protein</i>	7

Tabla 6. Lista de genes relacionados con resistencia a metales pesados presentes en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación)

	KO ID	Nombre	Producto	N° genes
Níquel	K00518	<i>sodN</i>	<i>nickel superoxide dismutase</i> [EC:1.15.1.1]	13
	K02031	ABC.PE.A	<i>peptide/nickel transport system ATP-binding protein</i>	119
	K02032	ABC.PE.A1	<i>peptide/nickel transport system ATP-binding protein</i>	180
	K02033	ABC.PE.P	<i>peptide/nickel transport system permease protein</i>	347
	K02034	ABC.PE.P1	<i>peptide/nickel transport system permease protein</i>	312
	K02035	ABC.PE.S	<i>peptide/nickel transport system substrate-binding protein</i>	703
	K04651	<i>hypA, hybF</i>	<i>hydrogenase nickel incorporation protein HypA/HybF</i>	11
	K04652	<i>hypB</i>	<i>hydrogenase nickel incorporation protein HypB</i>	4
	K07241	<i>nixA</i>	<i>high-affinity nickel-transport protein</i>	9
K07722	<i>nikR</i>	<i>CopG family transcriptional regulator, nickel-responsive regulator</i>	13	
Multi-metal	K02006	<i>cbiO</i>	<i>cobalt/nickel transport system ATP-binding protein</i>	1
	K02008	<i>cbiQ</i>	<i>cobalt/nickel transport system permease protein</i>	1
	K06189	<i>corC</i>	<i>magnesium and cobalt transporter</i>	7
	K08970	<i>rcnA</i>	<i>nickel/cobalt exporter</i>	1
	K15725	<i>czcC</i>	<i>outer membrane protein, cobalt-zinc-cadmium efflux system</i>	1
	K15726	<i>czcA</i>	<i>cobalt-zinc-cadmium resistance protein CzcA</i>	22
	K15727	<i>czcB</i>	<i>membrane fusion protein, cobalt-zinc-cadmium efflux system</i>	2
	K16264	<i>czcD, zitB</i>	<i>cobalt-zinc-cadmium efflux system protein</i>	45
	K16267	<i>zipB</i>	<i>zinc and cadmium transporter</i>	1

También se detectó una amplia gama de genes que participan en la tolerancia al choque térmico, genes de adaptaciones a la acidez e incluso genes involucrados en funciones de comunicación y desarrollo de comunidades bacterianas tales como *quorum sensing*, formación de biopelículas y una amplia cantidad de transportadores ABC (**Tabla 7**).

Por último, se identificaron diversos genes involucrados en la síntesis y defensa de compuestos antimicrobianos tales como: la síntesis de estreptomicina y cefalosporina, tetraciclinas, beta-lactamasas, vancomicina, cloranfenicol y mecanismos de multi-resistencia a antibióticos (**Tabla 8**).

Tabla 7. Lista de genes relacionados con el choque térmico, la formación de biopelículas y de comunicación bacteriana en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO).

	KO ID	Nombre	Producto	N° genes
Choque térmico	K03705	<i>hrcA</i>	<i>heat-inducible transcriptional repressor</i>	48
	K03708	<i>ctsR</i>	<i>transcriptional regulator of stress and heat shock response</i>	16
	K03799	<i>htpX</i>	<i>heat shock protein HtpX</i> [EC:3.4.24.-]	54
	K04762	<i>hsIR</i>	<i>ribosome-associated heat shock protein Hsp15</i>	3
	K09419	HSFF	<i>heat shock transcription factor, other eukaryote</i>	1
	K09487	HSP90B, TRA1	<i>heat shock protein 90kDa beta</i>	2
	K09489	HSPA4	<i>heat shock 70kDa protein 4</i>	4
	K09490	HSPA5, BIP	<i>heat shock 70kDa protein 5</i> [EC:3.6.4.10]	6
	K13640	<i>hspR</i>	<i>MerR family transcriptional regulator, heat shock protein HspR</i>	17
	K19369	HSPB11	<i>heat shock protein beta-11</i>	1
	K19765	HSBP1	<i>heat shock factor-binding protein 1</i>	1
	K03704	<i>cspA</i>	<i>cold shock protein (beta-ribbon, CspA family)</i>	43
Resistencia a acidez	K01580	<i>gadB</i> , <i>gadA</i> , GAD	<i>glutamate decarboxylase</i> [EC:4.1.1.15]	57
	K01584	<i>adiA</i>	<i>arginine decarboxylase</i> [EC:4.1.1.19]	3
Formación de biopelículas	K11931	<i>pgaB</i>	<i>biofilm PGA synthesis lipoprotein PgaB</i> [EC:3.5.1.-]	1
	K11936	<i>pgaC</i>	<i>biofilm PGA synthesis N-glycosyltransferase PgaC</i> [EC:2.4.1.-]	3
	K19336	<i>bdcA</i>	<i>cyclic-di-GMP-binding biofilm dispersal mediator protein</i>	1
	K19449	<i>sinR</i>	<i>XRE family transcriptional regulator, master regulator for biofilm formation</i>	1
<i>Quorum sensing</i>	K07782	<i>sdiA</i>	<i>LuxR family transcriptional regulator, quorum-sensing system regulator SdiA</i>	4
	K07173	<i>luxS</i>	<i>S-ribosylhomocysteine lyase</i> [EC:4.4.1.21]	1
	K01243	<i>pfs</i>	<i>adenosylhomocysteine nucleosidase</i> [EC:3.2.2.9]	12
	K10439	<i>rbsB</i>	<i>ribose transport system substrate-binding protein</i>	87
Transportadores ABC	K01990	ABC-2.A	<i>ABC-2 type transport system ATP-binding protein</i>	636
	K01992	ABC-2.P	<i>ABC-2 type transport system permease protein</i>	426
	K02003	ABC.CD.A	<i>putative ABC transport system ATP-binding protein</i>	94
	K02004	ABC.CD.P	<i>putative ABC transport system permease protein</i>	108
	K02021	ABC.MR	<i>putative ABC transport system ATP-binding protein</i>	3
	K02069	ABC.X2.P	<i>putative ABC transport system permease protein</i>	3
	K05832	ABC.X4.P	<i>putative ABC transport system permease protein</i>	2

Tabla 8. Lista de genes relacionados con procesos de síntesis y resistencia a antibióticos presentes en FLA obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG Orthology (KO).

	KO ID	Nombre	Producto	Nº genes
Síntesis de cefalosporina	K01060	<i>cah</i>	<i>cephalosporin-C deacetylase</i> [EC:3.1.1.41]	8
Resistencia a cloranfenicol	K07552	<i>bcr</i>	<i>MFS transporter, DHA1 family, bicyclomycin/chloramphenicol resistance protein</i>	3
	K18554	<i>cpt</i>	<i>chloramphenicol 3-O phosphotransferase</i> [EC:2.7.1.-]	3
	K19646	<i>fexA_B</i>	<i>MFS transporter, DHA2 family, florfenicol/chloramphenicol resistance protein</i>	1
Síntesis de estreptomicina	K04343	<i>strB</i>	<i>streptomycin 6-kinase</i> [EC:2.7.1.72]	6
Resistencia a beta-lactamasas	K07576	K07576	<i>metallo-beta-lactamase family protein</i>	7
	K08218	<i>ampG</i>	<i>MFS transporter, PAT family, beta-lactamase induction signal transducer AmpG</i>	8
	K17836	<i>penP</i>	<i>beta-lactamase class A</i> [EC:3.5.2.6]	20
Multi resistencia	K05595	<i>marC</i>	<i>multiple antibiotic resistance protein</i>	51
Resistencia a tetraciclina	K08151	<i>tetA</i>	<i>MFS transporter, DHA1 family, tetracycline resistance protein</i>	3
	K08168	<i>tetB</i>	<i>MFS transporter, DHA2 family, metal-tetracycline-proton antiporter</i>	1
	K18220	<i>tetM, tetO</i>	<i>ribosomal protection tetracycline resistance protein</i>	2
Resistencia a Vancomicina	K03862	<i>vanA</i>	<i>vanillate O-demethylase monooxygenase subunit</i> [EC:1.14.13.82]	13
	K03863	<i>vanB</i>	<i>vanillate O-demethylase ferredoxin subunit</i> [EC:1.14.13.82]	1
	K07260	<i>vanY</i>	<i>D-alanyl-D-alanine carboxypeptidase</i> [EC:3.4.16.4]	2
	K08641	<i>vanX</i>	<i>D-alanyl-D-alanine dipeptidase</i> [EC:3.4.13.22]	3
	K15739	<i>vanB, vanA, vanD</i>	<i>D-alanine---(R)-lactate ligase</i> [EC:6.1.2.1]	1
	K18347	<i>vanH</i>	<i>D-specific alpha-keto acid dehydrogenase</i> [EC:1.1.1.-]	1

5.3.3 Diversidad taxonómica

Se obtuvieron 29 secuencias de genes ribosomales 16S rARN del metagenoma de FLA. De estas secuencias, 17 corresponden a bacterias (**Tabla 9**), siete corresponden a arqueas (**Tabla S1**), y cinco corresponden a organelos de eucariotas (**Tabla S2**). Estos resultados coinciden con lo inicialmente encontrado en el proceso de anotación (**Figura 4**) y con el análisis taxonómico realizado en SILVAngs (**Figura 6**).

Tabla 9. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a bacterias encontradas en el metagenoma de FLA realizadas en la plataforma de BLASTN de NCBI, para las secuencias cuyo mejor hit fue una secuencia ambiental, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida.

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% Cobertura
1527	<i>Firmicutes</i>	<i>Bacilli</i>	<i>Uncultured bacterium clone</i> NEC03080	AY911447	99.21	82
			<i>Alicyclobacillus contaminans</i> 3-A191	NR041475	88.9	98
1535	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Sulfobacillus acidophilus</i> TPY	CP002901	91.74	100
1530	<i>Proteobacteria</i>	γ - <i>proteobacteria</i>	<i>Uncultured gamma proteobacterium clone</i> METASED12	KJ907765	100	100
			<i>Ectothiorhodospira mobilis</i> DSM 237	NR125567	90.15	100
1504	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Acidithiomicrobium</i> sp. P1	GQ225720	98.38	99
			<i>Acidimicrobium ferrooxidans</i> DSM 10331	NR074390	94.88	98
1531	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Sulfobacillus acidophilus</i> DSM 10332	NR074758	91.83	100
1485	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Acidibrevibacterium fodinaquatile</i> G45-3	CP029176	100	100
1493	<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Uncultured actinobacterium clone</i> METASED13	KJ907766	100	100
			<i>Spirillospora tritici</i> SJ 21	NR169368	91.92	100
1535	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Sulfobacillus acidophilus</i> TPY	CP002901	91.74	100
1514	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Uncultured Acidimicrobiaceae bacterium clone</i> METASED10	KJ907763	100	100
			<i>Aciditerrimonas ferrireducens</i> JCM 15389	NR112972	94.81	97
1514	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Uncultured Aciditerrimonas</i> sp. clone METASED20	KJ907773	100	100
			<i>Aciditerrimonas ferrireducens</i> JCM 15389	NR112972	99.93	97
1380	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured bacterium clone</i> DSJB13	DQ499175	96.82	100
			<i>Sulfobacillus acidophilus</i> DSM 10332	NR040943	91.8	100
1528	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured Clostridia bacterium clone</i> Tik_17	AM749763	91.28	89
			<i>Thermaerobacter marianensis</i> DSM 12885	NR074944	85.25	100
1389	<i>Firmicutes</i>	<i>Bacilli</i>	<i>Uncultured Alicyclobacillus</i> sp. clone METASED16	KJ907769	99.93	99
			<i>Alicyclobacillus disulfidooxidans</i> DSM 12064	NR040944	99.6	99
969	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Uncultured Ferrimicrobium</i> sp. clone METASED14	KJ907767	100	100
			<i>Ferrimicrobium acidiphilum</i> T23	NR041798	98.56	100

Tabla 9. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a bacterias encontradas en el metagenoma de FLA realizadas en la plataforma de BLASTN de NCBI, para las secuencias cuyo mejor hit fue una secuencia ambiental, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida (continuación).

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% Cobertura
1548	<i>Nitrospirae</i>	<i>Nitrospira</i>	<i>Leptospirillum ferriphilum</i> ZJ	MN093805	99.94	100
1526	<i>Proteobacteria</i>	<i>Acidithiobacillia</i>	<i>Acidithiobacillus caldus</i> MTH-04	CP043926	99.93	100
949	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured Firmicutes bacterium clone RT10C_1B_37</i>	EF441917	94.58	93
			<i>Sulfobacillus acidophilus</i> DSM 10332	NR074758	88.45	99



Figura 6. Análisis taxonómico de las secuencias de 16S rARN derivadas del metagenoma de FLA realizado en la plataforma SILVAngs. El gráfico fue generado en la plataforma KRONA.

5.3.4 Recuperación de genomas

5.3.4.1 Análisis genómico comparativo

Teniendo en cuenta los resultados recuperados de la anotación de genes ribosomales 16S rARN y por medio de genómica comparativa, fue posible recuperar dos genomas a partir de este metagenoma, estos corresponden a microorganismos posiblemente pertenecientes a los géneros *Acidibrevibacterium* y *Aciditerrimonas*. Los cuales fueron denominados como ‘*Candidatus Acidibrevibacterium* sp. FLA01’ y ‘*Candidatus Aciditerrimonas* sp. FLA02’, respectivamente.

‘*Candidatus Acidibrevibacterium* sp. FLA01’

Es un genoma de 3.8 Mb, con un porcentaje de GC de 65.8% y está compuesto por 625 *contigs* y cuyo valor de N50 es de 58,699 pb (**Tabla S5**). El alineamiento con el genoma de referencia muestra regiones ampliamente conservadas a lo largo del genoma y una alta variabilidad con los genes ubicados en los plásmidos presentes en el genoma de referencia (**Figura 7**).

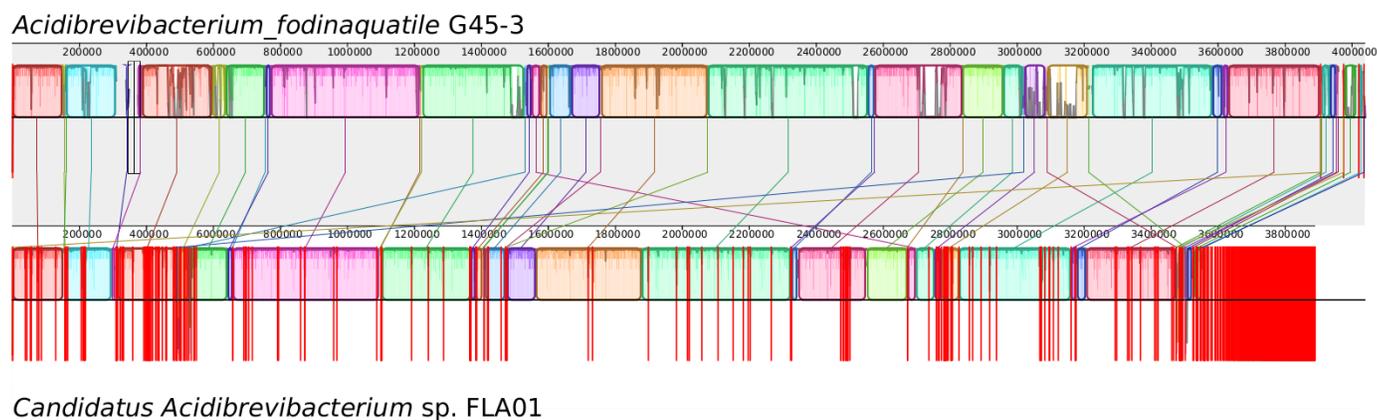


Figura 7. Alineamiento genómico desarrollado en MAUVE entre los genomas de *Acidibrevibacterium fodinaquatile* G45-3 (parte superior) y de ‘*Ca. Acidibrevibacterium* sp. FLA01’ (parte inferior). Las regiones conservadas se muestran en los bloques coloreados.

Del proceso de anotación (**Tabla 11**) se estableció que este genoma tiene 4,203 secuencias codificantes distribuidas en 290 subsistemas, 72 ARNs y ningún CRISPR. El genoma revela que esta bacteria posee tolerancia a diversos tipos de estrés, como el choque térmico (*grpE*) estrés osmótico (*ompA*), oxidativo (*ahpC*, *crp*, *ohr*, *sod*, *rbr*), radicales libres de monóxido de nitrógeno (*nnrS*, *nsrR*) y radicales de azufre (*trxR*, *bcp*, *ahpC*). Posee diversos mecanismos de defensa y resistencia a metales pesados tales como; cromo (*chrA*), cobalto-zinc-cadmio (*czcD*, *cusB/czsB*), cobre (*cutE*, *corC*, *cflA*, *ccm*, *copG*) mercurio (*merA*, *mir*) y zinc (*zraR*). La presencia de componentes de resistencia a fluoroquinolonas (*gyrA*, *gyrB*) y beta-lactamasas (*phnP*), le confiere la resistencia a una amplia gama de compuestos antimicrobianos.

A nivel metabólico, la presencia de un complejo de fotosistema II y de un carboxisoma le confiere un metabolismo probablemente quimioautótrofo, con capacidad para fijar CO₂ e incorporar hierro a través de hemo oxigenasas (*hemO*), el cual podría ser usado por ferredoxina reductasas (FR) como aceptor final de electrones. Adicionalmente, a nivel del metabolismo del nitrógeno se encontraron genes relacionados con la hidrólisis de cianato (*cynS*) y la asimilación de amonio (*glnABED*, *amt*). En cuanto al metabolismo del fósforo, se detectó la incorporación de fosfatos a partir de pirofosfatos (*ppa*) y polifosfatos (*ppx*, *ppk*, *ppgK*), transportadores y mecanismos de regulación (*phoURBQ*) y escasez de fósforo (*phoH*). Finalmente, para el metabolismo del azufre se encontraron genes ligados a su adquisición a partir de glutatión (*ggt*) mostrando las posibles formas por las cuales se estaría incorporando estos macronutrientes.

A nivel taxonómico, tanto los análisis de ANIs como *in silico* DDH (**Tabla 10**) muestran que el genoma recuperado está altamente relacionado con el genoma de referencia. Los análisis filogenómicos muestran que en efecto este genoma tiene una relación muy cercana al género *Acidibrevibacterium*, comparándolo tanto con el genoma de referencia existente para este género como con otros genomas pertenecientes a la

familia *Acetobacteraceae* (**Figura 10**). Y se trata del primer genoma de este microorganismo reportado en el campo geotérmico de Los Azufres.

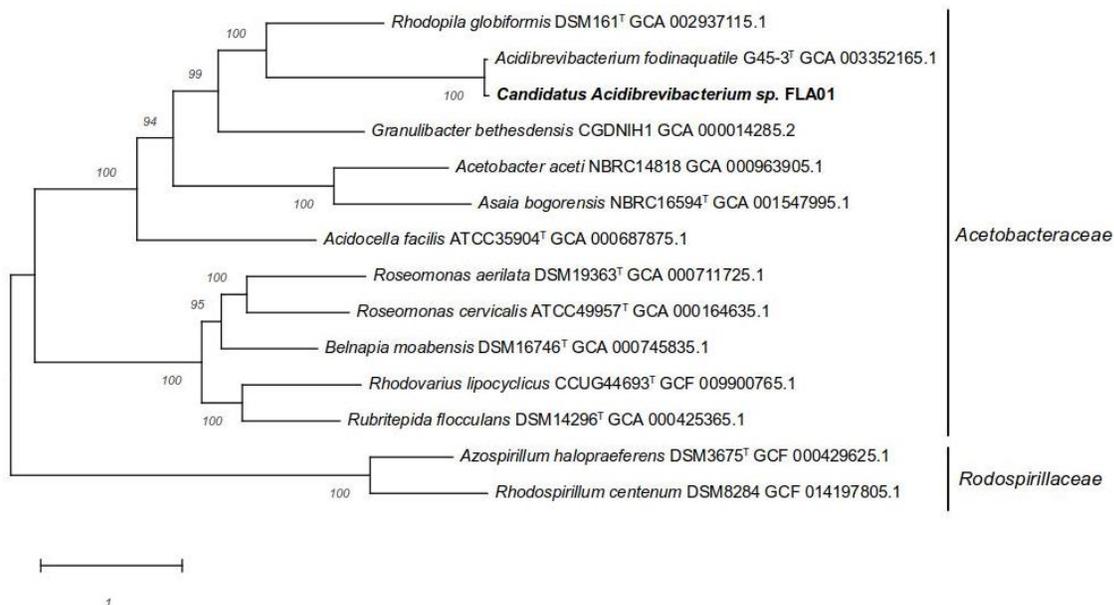


Figura 8. Análisis filogenómico en el cual se representa la relación evolutiva predicha entre ‘*Ca. Acidibrevibacterium sp. FLA01*’ y diferentes secuencias de integrantes de la familia *Acetobacteraceae*. El árbol fue reconstruido con PhyloPhlAn utilizando 390 alineamientos de marcadores moleculares, PhyloPhlAn realiza alineamientos individuales de cada set de proteínas recuperados de los genomas introducidos correspondientes al orden *Rodospiriales*. Así, PhyloPhlAn concatena las posiciones más discriminativas en cada alineamiento de proteínas en una única secuencia larga para reconstruir un árbol filogenético usando FastTree. La posición de ‘*Ca. Acidibrevibacterium sp. FLA01*’ se resalta en negrita. Los números de accesos se encuentran seguidos a cada uno de los nombres de los genomas. Los números en los soportes de ramas representan los valores de probabilidades (basado en 1000 remuestreos). La barra de escala en la parte inferior izquierda representa el número estimado de sustituciones de aminoácidos por sitio por unidad de longitud de rama.

Tabla 10. Métricas de identidad de secuencia utilizados para la identificación del genoma de ‘*Ca. Acidibrevibacterium sp. FLA01*’.

	‘ <i>Ca. Acidibrevibacterium sp. FLA01</i> ’	<i>Acidibrevibacterium fodinaquatile</i> G 45-3
Número de <i>contigs</i>	625	4 (1 cromosoma, 3 plásmidos)
Tamaño del genoma	3.88 MB	4.03 MB
ANIb	100	98.39
ANIm	100	98.69
<i>In silico</i> DDH	100	87.3
GC%	65.8	65.9

Tabla 11. Subsistemas detectados en los genomas recuperados de ‘*Ca. Acidibrevibacterium* sp. FLA01’ y ‘*Ca. Aciditerrimonas* sp. FLA02’ por medio de análisis de anotación en RAST.

Subsistema	‘ <i>Ca. Acidibrevibacterium</i> sp. FLA01’	‘ <i>Ca. Aciditerrimonas</i> sp. FLA02’
Cofactores, vitaminas, grupos prostéticos, pigmentos	137	64
Pared celular y cápsula	29	22
Virulencia, enfermedad y defensa	59	72
Metabolismo del K	6	10
Fotosíntesis	8	0
Misceláneos	30	15
Fagos, profagos, elementos transponibles, plásmidos	13	0
Transporte de membrana	77	30
Metabolismo y adquisición de Fe	1	0
Metabolismo ARN	36	75
Nucleósidos y nucleótidos	75	67
Metabolismo de las proteínas	194	162
Regulación y señalización celular	19	12
Metabolismo secundario	4	0
Metabolismo ADN	99	30
Ácidos grasos, lípidos, e isoprenoides	97	94
Metabolismo del N	15	26
Dormancia y esporulación	1	0
Respiración	117	126
Respuesta a estrés	50	16
Metabolismo de compuestos aromáticos	38	12
Aminoácidos y derivados	249	172
Metabolismo del S	8	10
Metabolismo del P	24	39
Carbohidratos	247	194

‘*Candidatus Aciditerrimonas* sp. FLA02’

Se compone de un genoma con un tamaño de 2.6 Mb, con un contenido de GC del 71.5% dividido en 1,802 *contigs* (Tabla S6). Al consultar el único genoma de referencia disponible del género *Aciditerrimonas* (*Aciditerrimonas ferrireducens* JCM15389), se observó que éste no se encontraba completamente finalizado. Razón por la cual, al realizar los alineamientos con el genoma recuperado a

partir del mapeo de lecturas, se obtuvo una alta fragmentación (**Figura 9**) y una mayor longitud del genoma recuperado de 2.6 Mb contra los 1.2 Mb del genoma de referencia.

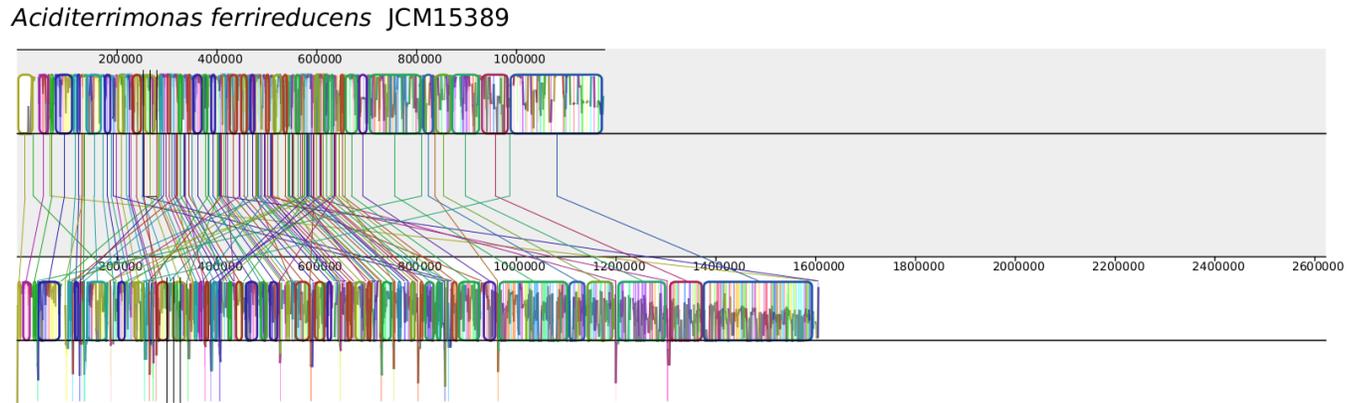


Figura 9. Alineamiento genómico desarrollado en MAUVE entre los genomas de *Aciditerrimonas ferrireducens* JCM15389 (parte superior) y de ‘*Ca. Aciditerrimonas* sp. FLA02’ (parte inferior). Las regiones conservadas se muestran en los bloques coloreados.

De acuerdo con los datos procedentes de la anotación (**Tabla 8**), se observa que el genoma de ‘*Ca. Aciditerrimonas* sp. FLA02’ posee 2,879 secuencias codificantes, 68 ARNs y ningún CRISPR. Dentro del genoma se detectaron genes relacionados con condiciones de resistencia a diversos tipos de adversidades ambientales como el estrés oxidativo (*rbr*, *sod*, *ahpC*, *ahr*), radicales de azufre (*trxR*, *bcp*), mecanismos de defensa a sustancias antimicrobianas como fluoroquinolonas (*gyrAB*), y resistencia a metales pesados, principalmente cobre (*CIA*, *ccm*, *cutE*, *corC*) y arsénico (*arsABC*, *acr3*).

En cuanto a su metabolismo, se encontraron genes involucrados en el metabolismo del nitrógeno que le confieren capacidad de asimilar amonio (*amt*, *glnAB*) y de hidrolizar cianato (*cynS*) los cuales revelan su participación en la incorporación de compuestos nitrogenados. Con respecto al metabolismo del fósforo, fue posible encontrar genes relacionados con la escasez y regulación de fosfatos (*phoHRB*) incorporación de pirofosfatos (*ppa*) y polifosfatos (*ppgK*, *ppk*). Adicionalmente se hallaron genes dentro

del metabolismo del carbono relacionados con la fijación de CO₂ (*rbcAIS*, *pcpB*, *cbbR*). Por último, se detectó la capacidad de usar arseniato como aceptor final de electrones (*arsR*).

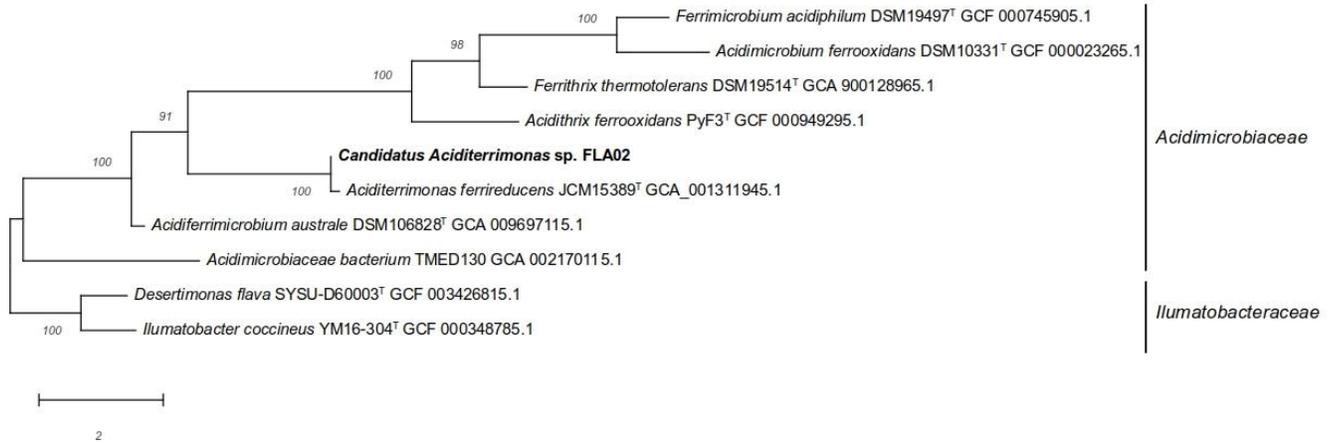


Figura 10. Análisis filogenómico en el cual se representa la relación evolutiva predicha entre ‘*Ca. Aciditerrimonas sp. FLA02*’ y diferentes secuencias de integrantes de la familia *Acidimicrobiaceae*. El árbol fue reconstruido con PhyloPhlAn utilizando 357 alineamientos de marcadores moleculares, PhyloPhlAn realiza alineamientos individuales de cada set de proteínas recuperados de los genomas introducidos correspondientes al orden *Acidimicrobiales*. Así, PhyloPhlAn concatena las posiciones más discriminativas en cada alineamiento de proteínas en una única secuencia larga para reconstruir un árbol filogenético usando FastTree. La posición de ‘*Ca. Aciditerrimonas sp. FLA02*’ se resalta en negrita. Los números de accesos se encuentran seguidos a cada uno de los genomas. Los números en los soportes de ramas representan los valores de probabilidades (basado en 1000 remuestros). La barra de escala en la parte inferior izquierda representa el número estimado de sustituciones de aminoácidos por sitio por unidad de longitud de rama.

La relación del genoma recuperado con el genoma de referencia también se ve soportada por medio del uso de métricas como los ANI e *in silico* DDH los cuales revelaron una alta semejanza entre el genoma de ‘*Ca. Aciditerrimonas sp. FLA02*’ y *Aciditerrimonas ferrireducens* JCM15389 (**Tabla 12**). Dicha relación se encontró también al realizar los análisis filogenómicos (**Figura 12**), por lo que existe una alta probabilidad de que se trate quizá de otro tipo de microorganismo del género *Aciditerrimonas*.

Tabla 12. Métricas de identidad de secuencia utilizadas para la identificación del genoma de ‘*Ca. Aciditerrimonas* sp.FLA02’

	‘ <i>Candidatus Aciditerrimonas</i> sp. FLA02’	<i>Aciditerrmonas ferrireducens</i> JCM15389
Número de <i>contigs</i>	1,802	1,004
Tamaño del genoma	2.62 MB	1.17MB
ANIb	100	98.3
ANIm	100	98.5
<i>In silico</i> DDH	100	72.9
GC%	71.5	70.1

5.3.4.2 Construcción de genomas a partir de metagenomas (MAG) via binning

A partir del proceso de *binning*, fue posible recuperar 38 MAGs, de los cuales 22 corresponden a genomas relacionados con bacterias, 15 a arqueas y 2 a eucariontes. Dentro de los genomas reconstruidos, se puede apreciar que dos de ellos están repetidos debido a que se lograron reconstruir independientemente del método de *binning*.

Por otra parte, de todos los MAGs recuperados, solo 23 poseen una completitud mayor del 90% y 16 del 95 %, siendo el MAG MaxBin_FLA.019 el que posee la mayor completitud, siendo éste del 99% de acuerdo con los marcadores que usa CheckM (**Figura S3A**). De acuerdo con el perfil taxonómico (**Tabla 13**) dicho MAG, corresponde al género *Acidibrevibacterium*. Gran parte de los genomas reconstruidos tienen relación con los filos *Actinobacteria*, *Proteobacteria*, *Firmicutes* y *Euryarcheota*.

Tabla 13. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de FLA vía binning. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud igual o mayor al 90%. (n.d) Rangos taxonómicos no determinados.

# bin	Dominio	Phylum	Clase	Orden	Familia	Genero	Especie
CONCOCT_FLA15.fa	Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Leptospirillum	n.d.
CONCOCT_FLA30.fa	Bacteria	Proteobacteria	Acidithiobacillia	Acidithiobacillales	Acidithiobacillaceae	Acidithiobacillus	n.d.
CONCOCT_FLA48.fa	Bacteria	Actinobacteria	Acidimicrobia	Acidimicrobiales	Acidimicrobiaceae	Ferrimicrobium	Ferrimicrobium acidiphilum
CONCOCT_FLA51.fa	Bacteria	Actinobacteria	Actinobacteria	n.d.	n.d.	n.d.	n.d.

Tabla 13. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de FLA vía binning. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud igual o mayor al 90%. (n.d) Rangos taxonómicos no determinados (continuación).

CONCOCT _FLA55.fa	Archaea	'Candidatus Marsarchaeot a'	'Candidatus Marsarchae ota_group 2'	n.d.	n.d.	n.d.	n.d.
CONCOCT _FLA67.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.
CONCOCT _FLA71.fa	Archaea	Crenarchaeot a	Thermoprot ei	Sulfolobales	Sulfolobaceae	Acidianus	Acidianus brierleyi
CONCOCT _FLA76_su b.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.
CONCOCT _FLA80.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.
CONCOCT _FLA6.fa	Bacteria	Proteobacteri a	n.d.	n.d.	n.d.	n.d.	n.d.
MaxBin_FL A.001_sub.f a	Archaea	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
MaxBin_FL A.002_sub.f a	Eukaryota	Rhodophyta	Bangiophyc eae	Cyanidiales	Cyanidiaceae	unclassified Cyanidiaceae	Cyanidiaceae sp. MX-AZ01
MaxBin_FL A.002_sub.f a	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	n.d.	n.d.	n.d.
MaxBin_FL A.004_sub.f a	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	n.d.	n.d.	n.d.
MaxBin_FL A.009_sub.f a	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales:	unclassified Thermoplasmat ales	n.d.	n.d.
MaxBin_FL A.012.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.
MaxBin_FL A.014_sub.f a	Archaea	'Candidatus Parvarchaeota '	'Candidatus Parvarchaeu m'	'Candidatus Parvarchaeu m acidophilus'	'Candidatus Parvarchaeum acidophilus ARMAN-5'	n.d.	n.d.
MaxBin_FL A.019.fa	Bacteria	Proteobacteri a	Alphaproteo bacteria	Rhodospirill ales	Acetobacterace ae	Acidibrevibacteri um	Acidibrevibacter ium fodinaquatile
MetaBAT_ FLA.10.fa	Bacteria	Actinobacteri a	Acidimicrob iia	Acidimicrob iales	Acidimicrobiac eae	Ferrimicrobium	Ferrimicrobium acidiphilum
MetaBAT_ FLA.11.fa	Bacteria	Firmicutes	n.d.	n.d.	n.d.	n.d.	n.d.
MetaBAT_ FLA.15.fa	Bacteria	Proteobacteri a	Acidithioba cillia	Acidithioba cillales	Acidithiobacilla ceae	Acidithiobacillus	Acidithiobacillus caldus
MetaBAT_ FLA.16.fa	Bacteria	Actinobacteri a	n.d.	n.d.	n.d.	n.d.	n.d.
MetaBAT_ FLA.19.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	unclassified Thermoplasmat ales	unclassified Thermoplasmat ales (miscellaneous)	Thermoplasmat ales archaeon E- plasma:
MetaBAT_ FLA.21.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.
MetaBAT_ FLA.22.fa	Bacteria	Firmicutes	Bacilli	Bacillales	unclassified Bacillales	Acidibacillus	n.d.
MetaBAT_ FLA.27.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.

Tabla 13. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de FLA vía binning. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud igual o mayor al 90%. (n.d) Rangos taxonómicos no determinados (continuación).

# bin	Dominio	Phylum	Clase	Orden	Familia	Genero	Especie
MetaBAT_ FLA.28.fa	Eukaryota	Rhodophyta	Bangiophyc eae	Cyanidiales	Cyanidiaceae	Cyanidioschyzon	Cyanidioschyzon merolae
MetaBAT_ FLA.30.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	Ferroplasmace ae	Acidiplasma	n.d.
MetaBAT_ FLA.33.fa	Bacteria	Firmicutes	Bacilli	Bacillales	Alicyclobacillac eae	Alicyclobacillus	Alicyclobacillus montanus
MetaBAT_ FLA.38.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	unclassified Thermoplasmat ales	unclassified Thermoplasmat ales (miscellaneous)	Thermoplasmat ales archaeon A- plasma
MetaBAT_ FLA.39.fa	Bacteria	Actinobacteri a	n.d.	unclassified Actinobacte ria	n.d.	n.d.	Actinobacteria bacterium 15TR583:
MetaBAT_ FLA.42.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	unclassified Thermoplasmat ales	unclassified Thermoplasmat ales	Thermoplasmat ales archaeon I- plasma
MetaBAT_ FLA.43.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	unclassified Thermoplasmat ales	unclassified Thermoplasmat ales (miscellaneous)	Thermoplasmat ales archaeon A- plasma
MetaBAT_ FLA.44.fa	Archaea	environmental samples	uncultured archaeon	n.d.	n.d.	n.d.	n.d.
MetaBAT_ FLA.48.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	n.d.	n.d.	n.d.
MetaBAT_ FLA.50.fa	Archaea	Euryarchaeot a	Thermoplas mata	Thermoplas matales	n.d.	n.d.	n.d.
MetaBAT_ FLA.53.fa	Bacteria	Firmicutes	Clostridia (class)	Clostridiale s (order)	Clostridiales Family XVII. Incertae Sedis	Sulfobacillus	Sulfobacillus acidophilus
MetaBAT_ FLA.8.fa	Bacteria	Firmicutes	Clostridia	Clostridiale s	Clostridiales Family XVII. Incertae Sedis	Sulfobacillus	n.d.
MetaBAT_ FLA.9.fa	Bacteria	Actinobacteri a	Actinobacte ria	n.d.	n.d.	n.d.	n.d.

5.4 Manantiales de Maritaro (MM)

5.4.1 Estructura y diversidad

Al igual que en FLA, en este metagenoma, el dominio Bacteria es el más abundante con el 80.86%, seguido de los eucariotas con el 18.56% y apenas aparecen rastros de arqueas y virus (**Figura 11A**). Los filos de bacterias más abundantes corresponden a *Proteobacteria*, *Firmicutes*, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes* y *Nitrospirae*. Con una gran abundancia de eucariotas del filo *Chlorophyta* y de grupos sin identificar (**Figura 11B**).

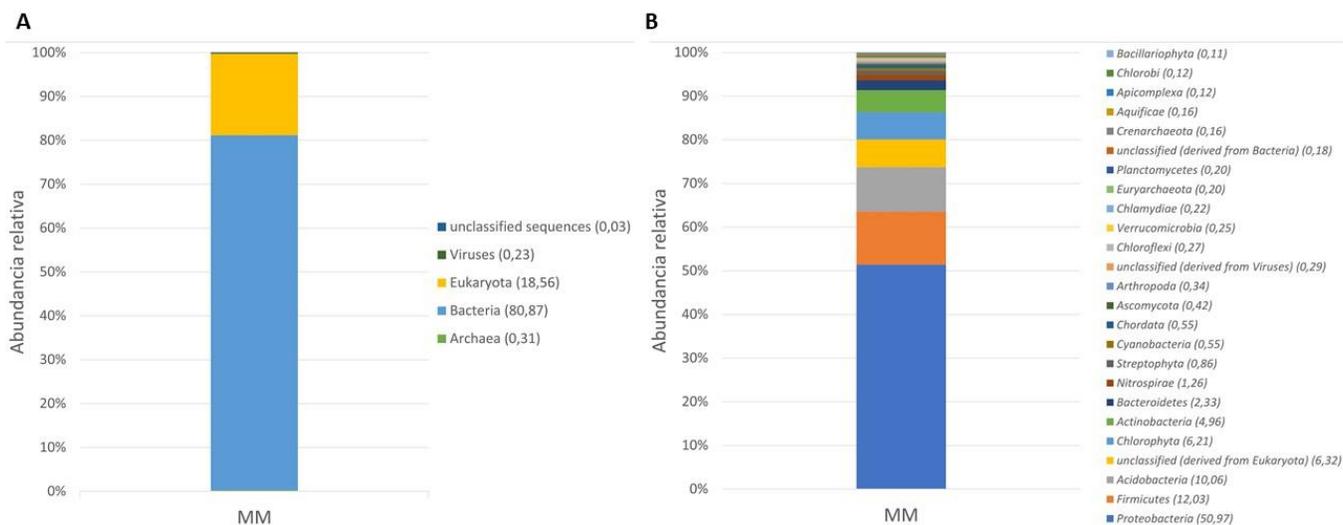


Figura 11. Proporción de grupos de microorganismos encontrados a partir de número de *contigs* presentes en el metagenoma de MM, obtenidos del análisis en la plataforma MG-RAST usando el algoritmo de contigLCA (*LastCommonancestor*).

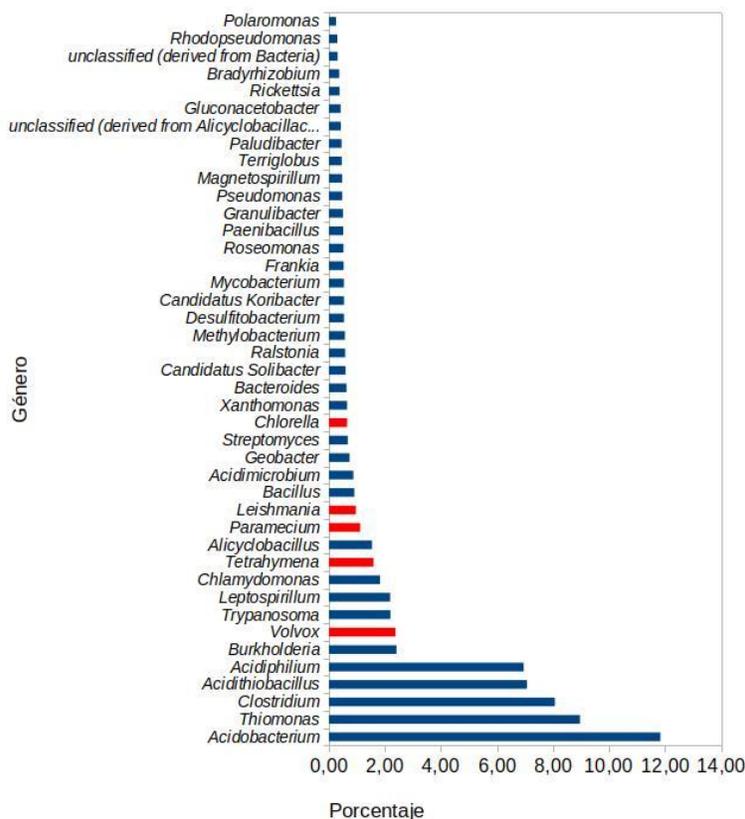


Figura 12. Abundancias de géneros de microorganismos encontrados en el metagenoma de MM, resultado del análisis en la plataforma de MG-RAST usando el algoritmo de contigLCA (*LastCommonancestor*).

A nivel de género, se detectaron bacterias relacionadas con *Acidobacterium* (11.83%), *Thiomonas* (8.96%), *Clostridium* (8.07%), *Acidithiobacillus* (7.07%), *Acidiphilium* (6.95%), *Burkholderia* (2.41%), *Leptospirillum* (2.18%) y *Alicyclobacillus* (1.54%) (**Figura 12**).

5.4.2 Diversidad funcional

Para este metagenoma (MM), fue posible encontrar un total de 766,315 genes codificantes para proteínas de los cuales 346,243 poseen un nombre asignado (44.8%), 415 CRISPR y 6,466 genes ARN de los cuales 667 son genes rARN (**Tabla 14**). Dentro de los genes codificantes fue posible hallar un total de 324,092 genes ortólogos, los cuales se distribuyeron en las categorías mencionadas en la **Tabla 15**.

Tabla 14. Características de los dos metagenoma procedente de las muestras de los manantiales de Maritaro (MM), obtenidos a partir de la anotación en la plataforma IMG.

Característica	Número de fragmentos de genes
CRISPR	415
Genes ARN	6,466
Genes rARN	667
5s rARN	146
16S rARN	166
18S rARN	58
23S rARN	237
28S rARN	60
Genes tARN	5,269
Genes codificantes de proteínas	766,315
Genes con nombre de producto	346,243

Tabla 15. Abundancia relativa de categorías de genes ortólogos (COG) encontrados en los metagenomas de Maritaro (MM) enumerados por número de genes y porcentaje frente al total del metagenoma.

Categoría COG	N° Genes	porcentaje
Transporte y metabolismo de aminoácidos	30,134	8%
Transporte y metabolismo de carbohidratos	22,042	5.85%
Control del ciclo celular, división celular, particionamiento cromosómico	5,591	1.48%
Motilidad celular	8,477	2.25%
Pared celular/membrana/biogénesis de envoltura	23,481	6.23%
Estructura y dinámica de la cromatina	341	0.09%
Coenzimas de transporte y metabolismo	16,803	4.46%
Citoesqueleto	367	0.10%
Mecanismos de defensa	11,081	2.94%
Producción y conversión de energía	24,378	6.47%
Estructuras extracelulares	3,076	0.82%
Función desconocida	17,586	4.67%
Función general	36,634	9.72%
Transporte y metabolismo de iones inorgánicos	19,725	5.24%
Trafico extracelular, secreción y transporte vesicular	7,939	2.11%
Transporte y metabolismo de lípidos	16,350	4.34%
Mobiloma: profagos, transposones	9,572	2.54%
Transporte y metabolismo de nucleótidos	8,379	2.22%
Modificación postranscripcional, recambio proteico, chaperonas	16,889	4.48%
Procesamiento y modificación de ARN	259	0.07%
Replicación recombinación y reparación	18,672	4.96%
Biosíntesis de metabolitos secundarios, transporte y catabolismo	9,681	2.57%
Mecanismos de transducción de señal	23,550	6.25%
Transcripción	21,959	5.83%
Traducción, estructura ribosomal y biogénesis	23,773	6.31%

En cuanto a la distribución de los genes presentados en la **Tabla 15**, se observa que la mayoría de estos se encuentran concentrados dentro del filum *Proteobacteria*, seguido de *Acidobacteria*, *Firmicutes* y *Chlorophyta*. Con una proporción mucho más reducida en los filos *Nitrospirae*, *Actinobacteria* y *Bacterioidetes* (**Figura 13**).

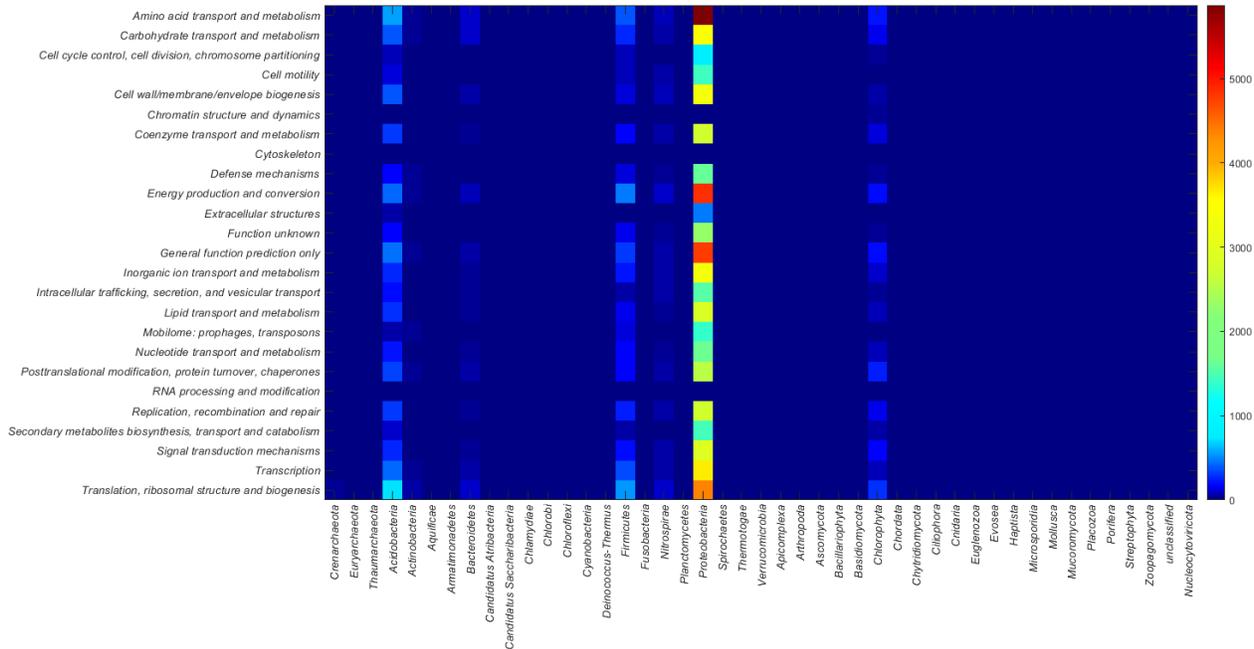


Figura 13. Representación diversidad funcional del metagenoma MM, las barras indican la cantidad de genes ortólogos identificados para cada comunidad de microorganismos en el metagenoma hallados a través de la plataforma IMG/MER.

Observando a detalle las categorías mencionadas en la **Tabla 15**, se identificaron diversos genes involucrados en funciones metabólicas relacionadas con ciclos biogeoquímicos, como es el caso de la fijación de carbono (relacionados con rutas de CBB y rTCA). Del mismo modo, para el metabolismo del nitrógeno se hallaron genes relacionados con la fijación de nitrógeno y la desnitrificación. En el metabolismo del fósforo se detectaron genes involucrados en la asimilación y en la regulación ante la escasez de fósforo y transportadores de fosfatos. Por último, también se hallaron registros de genes involucrados en la asimilación y el ciclo del azufre y uso de sustratos inorgánicos como hierro (**Tabla 16**).

Tabla 16. Lista de genes relacionados con procesos metabólicos en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO).

	KO ID	Nombre	Producto	N° genes
Azufre	K17222	<i>soxA</i>	<i>sulfur-oxidizing protein SoxA</i> [EC:2.8.5.2]	25
	K17223	<i>soxX</i>	<i>sulfur-oxidizing protein SoxX</i> [EC:2.8.5.2]	22
	K17224	<i>soxB</i>	<i>sulfur-oxidizing protein SoxB</i> [EC:3.1.6.20]	31

Tabla 16. Lista de genes relacionados con procesos metabólicos en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	Nº genes
Azufre	K17225	<i>soxC</i>	<i>sulfane dehydrogenase subunit SoxC</i>	26
	K17226	<i>soxY</i>	<i>sulfur-oxidizing protein SoxY</i>	32
	K17227	<i>soxZ</i>	<i>sulfur-oxidizing protein SoxZ</i>	21
	K00958	<i>sat, met3</i>	<i>sulfate adenylyltransferase</i> [EC:2.7.7.4]	57
	K00392	<i>sir</i>	<i>sulfite reductase (ferredoxin)</i> [EC:1.8.7.1]	58
	K13811	PAPSS	<i>3'-phosphoadenosine 5'-phosphosulfate synthase</i> [EC:2.7.7.4 2.7.1.25]	1
	K00394	<i>aprA</i>	<i>adenylylsulfate reductase, subunit A</i> [EC:1.8.99.2]	16
	K00395	<i>aprB</i>	<i>adenylylsulfate reductase, subunit B</i> [EC:1.8.99.2]	13
	K00380	<i>cysJ</i>	<i>sulfite reductase (NADPH) flavoprotein alpha-component</i> [EC:1.8.1.2]	44
	K00381	<i>cysI</i>	<i>sulfite reductase (NADPH) hemoprotein beta-component</i> [EC:1.8.1.2]	58
	K00390	<i>cysH</i>	<i>phosphoadenosine phosphosulfate reductase</i> [EC:1.8.4.8]	73
	K00640	<i>cysE</i>	<i>serine O-acetyltransferase</i> [EC:2.3.1.30]	108
	K00860	<i>cysC</i>	<i>adenylylsulfate kinase</i> [EC:2.7.1.25]	17
	K00956	<i>cysN</i>	<i>sulfate adenylyltransferase subunit 1</i> [EC:2.7.7.4]	17
	K00957	<i>cysD</i>	<i>sulfate adenylyltransferase subunit 2</i> [EC:2.7.7.4]	105
	K02045	<i>cysA</i>	<i>sulfate transport system ATP-binding protein</i> [EC:3.6.3.25]	6
	K02046	<i>cysU</i>	<i>sulfate transport system permease protein</i>	5
	K02047	<i>cysW</i>	<i>sulfate transport system permease protein</i>	4
	K02048	<i>cysP, sbp</i>	<i>sulfate transport system substrate-binding protein</i>	2
	K11180	<i>dsrA</i>	<i>sulfite reductase alpha subunit</i> [EC:1.8.99.3 1.8.99.5]	10
K11181	<i>dsrB</i>	<i>sulfite reductase beta subunit</i> [EC:1.8.99.3 1.8.99.5]	10	
Carbono (Fijación)	K01601	<i>rbcL</i>	<i>ribulose-bisphosphate carboxylase large chain</i> [EC:4.1.1.39]	96
	K01602	<i>rbcS</i>	<i>ribulose-bisphosphate carboxylase small chain</i> [EC:4.1.1.39]	36
	K15230	<i>aclA</i>	<i>ATP-citrate lyase alpha-subunit</i> [EC:2.3.3.8]	1
	K15231	<i>aclB</i>	<i>ATP-citrate lyase beta-subunit</i> [EC:2.3.3.8]	2
	K00174	<i>korA, oorA, oforA</i>	<i>2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit alpha</i> [EC:1.2.7.3 1.2.7.11]	95
	K00175	<i>korB, oorB, oforB</i>	<i>2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase subunit beta</i> [EC:1.2.7.3 1.2.7.11]	95
	K00176	<i>korD, oorD</i>	<i>2-oxoglutarate ferredoxin oxidoreductase subunit delta</i> [EC:1.2.7.3]	14
	K00177	<i>korC, oorC</i>	<i>2-oxoglutarate ferredoxin oxidoreductase subunit gamma</i> [EC:1.2.7.3]	21
	K00169	<i>porA</i>	<i>pyruvate ferredoxin oxidoreductase alpha subunit</i> [EC:1.2.7.1]	26
	K00170	<i>porB</i>	<i>pyruvate ferredoxin oxidoreductase beta subunit</i> [EC:1.2.7.1]	19
	K00171	<i>porD</i>	<i>pyruvate ferredoxin oxidoreductase delta subunit</i> [EC:1.2.7.1]	3
	K00172	<i>porG</i>	<i>pyruvate ferredoxin oxidoreductase gamma subunit</i> [EC:1.2.7.1]	18

Tabla 16. Lista de genes relacionados con procesos metabólicos en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	Nº genes
Nitrógeno	K02586	<i>nifD</i>	<i>nitrogenase molybdenum-iron protein alpha chain</i> [EC:1.18.6.1]	16
	K02587	<i>nifE</i>	<i>nitrogenase molybdenum-cofactor synthesis protein NifE</i>	10
	K02588	<i>nifH</i>	<i>nitrogenase iron protein NifH</i> [EC:1.18.6.1]	14
	K02591	<i>nifK</i>	<i>nitrogenase molybdenum-iron protein beta chain</i> [EC:1.18.6.1]	12
	K02592	<i>nifN</i>	<i>nitrogenase molybdenum-iron protein NifN</i>	13
	K02595	<i>nifW</i>	<i>nitrogenase-stabilizing/protective protein</i>	5
	K00370	<i>narG</i>	<i>nitrate reductase alpha subunit</i> [EC:1.7.99.4]	65
	K00371	<i>narH</i>	<i>nitrate reductase beta subunit</i> [EC:1.7.99.4]	37
	K00372	<i>nasA</i>	<i>assimilatory nitrate reductase catalytic subunit</i> [EC:1.7.99.4]	32
	K00373	<i>narJ</i>	<i>nitrate reductase delta subunit</i>	31
	K00374	<i>narI</i>	<i>nitrate reductase gamma subunit</i> [EC:1.7.99.4]	32
	K02575	NRT, <i>narK</i> , <i>nrP</i> , <i>nasA</i>	<i>MFS transporter, NNP family, nitrate/nitrite transporter</i>	77
	K07673	<i>narX</i>	<i>two-component system, NarL family, nitrate/nitrite sensor histidine kinase NarX</i> [EC:2.7.13.3]	23
	K07674	<i>narQ</i>	<i>two-component system, NarL family, nitrate/nitrite sensor histidine kinase NarQ</i> [EC:2.7.13.3]	2
	K07684	<i>narL</i>	<i>two-component system, NarL family, nitrate/nitrite response regulator NarL</i>	24
	K07685	<i>narP</i>	<i>two-component system, NarL family, nitrate/nitrite response regulator NarP</i>	2
	K10534	NR	<i>nitrate reductase (NAD(P)H)</i> [EC:1.7.1.1 1.7.1.2 1.7.1.3]	17
	K15576	<i>nrtA</i> , <i>nasF</i> , <i>cynA</i>	<i>nitrate/nitrite transport system substrate-binding protein</i>	18
	K15577	<i>nrtB</i> , <i>nasE</i> , <i>cynB</i>	<i>nitrate/nitrite transport system permease protein</i>	6
	K15578	<i>nrtC</i> , <i>nasD</i>	<i>nitrate/nitrite transport system ATP-binding protein</i> [EC:3.6.3.-]	7
	K00362	<i>nirB</i>	<i>nitrite reductase (NADH) large subunit</i> [EC:1.7.1.15]	40
	K00363	<i>nirD</i>	<i>nitrite reductase (NADH) small subunit</i> [EC:1.7.1.15]	21
	K00366	<i>nirA</i>	<i>ferredoxin-nitrite reductase</i> [EC:1.7.7.1]	19
	K00368	<i>nirK</i>	<i>nitrite reductase (NO-forming)</i> [EC:1.7.2.1]	16
	K02598	<i>nirC</i>	<i>nitrite transporter NirC</i>	1
	K03385	<i>nrfA</i>	<i>nitrite reductase (cytochrome c-552)</i> [EC:1.7.2.2]	1
	K15876	<i>nrfH</i>	<i>cytochrome c nitrite reductase small subunit</i>	1
	K02164	<i>norE</i>	<i>nitric oxide reductase NorE protein</i>	4
	K02305	<i>norC</i>	<i>nitric oxide reductase subunit C</i>	9
	K02448	<i>norD</i>	<i>nitric oxide reductase NorD protein</i>	5
	K04561	<i>norB</i>	<i>nitric oxide reductase subunit B</i> [EC:1.7.2.5]	66
	K04748	<i>norQ</i>	<i>nitric oxide reductase NorQ protein</i>	61
	K05916	<i>hmp</i> , YHB1	<i>nitric oxide dioxygenase</i> [EC:1.14.12.17]	25
	K13125	NOSIP	<i>nitric oxide synthase-interacting protein</i>	3

Tabla 16. Lista de genes relacionados con procesos metabólicos en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	N° genes
Nitrógeno	K13771	<i>nsrR</i>	<i>Rrf2 family transcriptional regulator, nitric oxide-sensitive transcriptional repressor</i>	58
	K00376	<i>nosZ</i>	<i>nitrous-oxide reductase</i> [EC:1.7.2.4]	13
	K00491	<i>nos</i>	<i>nitric-oxide synthase, bacterial</i> [EC:1.14.13.165]	2
	K07218	<i>nosD</i>	<i>nitrous oxidase accessory protein</i>	4
	K13125	NOSIP	<i>nitric oxide synthase-interacting protein</i>	3
	K19339	<i>nosR</i>	<i>NosR/NirI family transcriptional regulator, nitrous oxide reductase regulator</i>	4
Fósforo	K01077	E3.1.3.1, <i>phoA, phoB</i>	<i>alkaline phosphatase</i> [EC:3.1.3.1]	14
	K01078	E3.1.3.2	<i>acid phosphatase</i> [EC:3.1.3.2]	25
	K01113	<i>phoD</i>	<i>alkaline phosphatase D</i> [EC:3.1.3.1]	19
	K01507	<i>ppa</i>	<i>inorganic pyrophosphatase</i> [EC:3.6.1.1]	121
	K02039	<i>phoU</i>	<i>phosphate transport system protein</i>	114
	K06217	<i>phoH, phoL</i>	<i>phosphate starvation-inducible protein PhoH and related proteins</i>	95
	K07175	<i>phoH2</i>	<i>PhoH-like ATPase</i>	50
	K07636	<i>phoR</i>	<i>two-component system, OmpR family, phosphate regulon sensor histidine kinase PhoR</i> [EC:2.7.13.3]	92
	K07637	<i>phoQ</i>	<i>two-component system, OmpR family, sensor histidine kinase PhoQ</i> [EC:2.7.13.3]	12
	K07657	<i>phoB</i>	<i>two-component system, OmpR family, phosphate regulon response regulator PhoB</i>	80
	K07658	<i>phoB1, phoP</i>	<i>two-component system, OmpR family, alkaline phosphatase synthesis response regulator PhoP</i>	16
	K07660	<i>phoP</i>	<i>two-component system, OmpR family, response regulator PhoP</i>	19
	K08176	PHO84	<i>MFS transporter, PHS family, inorganic phosphate transporter</i>	5
	K09474	<i>phoN</i>	<i>acid phosphatase (class A)</i> [EC:3.1.3.2]	1
	K11929	<i>phoE</i>	<i>outer membrane pore protein E</i>	1
	K14430	PHO87_91	<i>phosphate transporter</i>	3
	K15640	K15640, <i>phoE</i>	<i>uncharacterized phosphatase</i>	1
	K02036	<i>pstB</i>	<i>phosphate transport system ATP-binding protein</i> [EC:3.6.3.27]	120
	K02037	<i>pstC</i>	<i>phosphate transport system permease protein</i>	110
	K02038	<i>pstA</i>	<i>phosphate transport system permease protein</i>	107
	K02039	<i>phoU</i>	<i>phosphate transport system protein</i>	114
	K02040	<i>pstS</i>	<i>phosphate transport system substrate-binding protein</i>	154
	K03306	TC.PIT	<i>inorganic phosphate transporter, PiT family</i>	63
	K16322	<i>pit</i>	<i>low-affinity inorganic phosphate transporter</i>	2

Tabla 16. Lista de genes relacionados con procesos metabólicos en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	Nº genes
Hierro	K18683	<i>rus</i>	<i>rusticyanin</i>	3
	K02274	<i>coxA</i>	<i>cytochrome c oxidase subunit I</i> [EC:1.9.3.1]	172
	K02275	<i>coxB</i>	<i>cytochrome c oxidase subunit II</i> [EC:1.9.3.1]	122
	K02276	<i>coxC</i>	<i>cytochrome c oxidase subunit III</i> [EC:1.9.3.1]	85
	K02277	<i>coxD</i>	<i>cytochrome c oxidase subunit IV</i> [EC:1.9.3.1]	2
	K04755	<i>fdx</i>	<i>ferredoxin, 2Fe-2S</i>	78
	K05524	<i>fdxA</i>	<i>ferredoxin</i>	75

Se halló también una amplia diversidad de genes involucrados en la tolerancia a los cambios abruptos de la temperatura, la resistencia a la acidez y genes relacionados con procesos de comunicación bacteriana, formación de biopelículas y una amplia cantidad de transportadores ABC (**Tabla 17**).

Tabla 17. Lista de genes relacionados con adaptaciones a resistencia a cambios de temperatura, acidez, mecanismos de comunicación bacteriana y transporte presentes en MM obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO).

	KO ID	Nombre	Producto	Nº genes
resistencia a altas temperaturas	K03668	<i>hslJ</i>	<i>heat shock protein HslJ</i>	1
	K03705	<i>hrcA</i>	<i>heat-inducible transcriptional repressor</i>	97
	K03708	<i>ctsR</i>	<i>transcriptional regulator of stress and heat shock response</i>	9
	K03799	<i>htpX</i>	<i>heat shock protein HtpX</i> [EC:3.4.24.-]	96
	K04762	<i>hslR</i>	<i>ribosome-associated heat shock protein Hsp15</i>	40
	K09487	HSP90B, TRA1	<i>heat shock protein 90kDa beta</i>	19
	K09489	HSPA4	<i>heat shock 70kDa protein 4</i>	13
	K09490	HSPA5, BIP	<i>heat shock 70kDa protein 5</i> [EC:3.6.4.10]	31
	K11940	<i>hspQ</i>	<i>heat shock protein HspQ</i>	1
	K13640	<i>hspR</i>	<i>MerR family transcriptional regulator, heat shock protein HspR</i>	25
	K19369	HSPB11	<i>heat shock protein beta-11</i>	1
K19765	HSBP1	<i>heat shock factor-binding protein 1</i>	1	
resistencia a bajas temperaturas	K03704	<i>cspA</i>	<i>cold shock protein (beta-ribbon, CspA family)</i>	218
Tolerancia a acidez	K01584	<i>adiA</i>	<i>arginine decarboxylase</i> [EC:4.1.1.19]	42
	K01585	<i>speA</i>	<i>arginine decarboxylase</i> [EC:4.1.1.19]	24
	K01580	<i>gadB, gadA, GAD</i>	<i>glutamate decarboxylase</i> [EC:4.1.1.15]	37

Tabla 17. Lista de genes relacionados con adaptaciones a resistencia a cambios de temperatura, acidez, mecanismos de comunicación bacteriana y transporte presentes en MM obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	N° genes
formación de biopelículas	K11931	<i>pgaB</i>	<i>biofilm PGA synthesis lipoprotein PgaB</i> [EC:3.5.1.-]	4
	K11936	<i>pgaC</i>	<i>biofilm PGA synthesis N-glycosyltransferase PgaC</i> [EC:2.4.1.-]	7
	K11937	<i>pgaD</i>	<i>biofilm PGA synthesis protein PgaD</i>	1
	K12148	<i>bssS</i>	<i>biofilm regulator BssS</i>	1
	K13654	<i>mcbR</i>	<i>GntR family transcriptional regulator, colanic acid and biofilm gene transcriptional regulator</i>	1
	K19334	<i>tabA</i>	<i>biofilm protein TabA</i>	1
	K19336	<i>bdcA</i>	<i>cyclic-di-GMP-binding biofilm dispersal mediator protein</i>	2
	K19449	<i>sinR</i>	<i>XRE family transcriptional regulator, master regulator for biofilm formation</i>	2
	K19688	<i>bssR</i>	<i>biofilm regulator BssR</i>	1
<i>quorum sensing</i>	K07782	<i>sdiA</i>	<i>LuxR family transcriptional regulator, quorum-sensing system regulator SdiA</i>	4
	K13651	<i>mqsR</i>	<i>motility quorum-sensing regulator / GCU-specific mRNA interferase toxin</i>	12
	K18304	<i>lasR</i>	<i>LuxR family transcriptional regulator, quorum-sensing system regulator LasR</i>	1
	K19731	<i>cciR</i>	<i>LuxR family transcriptional regulator, quorum-sensing system regulator CciR</i>	1
	K01243	<i>pfs</i>	<i>adenosylhomocysteine nucleosidase</i> [EC:3.2.2.9]	55
	K10439	<i>rbsB</i>	<i>ribose transport system substrate-binding protein</i>	165
Transportadores ABC	K01989	ABC.X4.S	<i>putative ABC transport system substrate-binding protein</i>	10
	K01990	ABC-2.A	<i>ABC-2 type transport system ATP-binding protein</i>	664
	K01992	ABC-2.P	<i>ABC-2 type transport system permease protein</i>	618
	K02003	ABC.CD.A	<i>putative ABC transport system ATP-binding protein</i>	255
	K02004	ABC.CD.P	<i>putative ABC transport system permease protein</i>	534
	K02068	ABC.X2.A	<i>putative ABC transport system ATP-binding protein</i>	13
	K02069	ABC.X2.P	<i>putative ABC transport system permease protein</i>	41
	K05832	ABC.X4.P	<i>putative ABC transport system permease protein</i>	4
	K05833	ABC.X4.A	<i>putative ABC transport system ATP-binding protein</i>	3
	K11636	<i>yxdM</i>	<i>putative ABC transport system permease protein</i>	1

Por otro lado, se hallaron genes involucrados en sistemas de defensa como la síntesis de compuestos antimicrobianos tales como; estreptomicina y cefalosporina. Además de mecanismos de resistencia a diferentes antibióticos, por ejemplo; tetraciclinas, beta-lactamasas, cloranfenicol, vancomicina y mecanismos de multi-resistencia (**Tabla 18**).

Tabla 18. Lista de genes relacionados con síntesis y mecanismos de resistencia de componentes antimicrobianos presentes en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO).

	KO ID	Nombre	Producto	Nº genes
beta lactamasas	K01467	<i>ampC</i>	<i>beta-lactamase class C</i> [EC:3.5.2.6]	13
	K07576	K07576	<i>metallo-beta-lactamase family protein</i>	57
	K08218	<i>ampG</i>	<i>MFS transporter, PAT family, beta-lactamase induction signal transducer AmpG</i>	114
	K17836	<i>penP</i>	<i>beta-lactamase class A</i> [EC:3.5.2.6]	17
	K17837	<i>bla2, blm, ccrA, blaB</i>	<i>metallo-beta-lactamase class B</i> [EC:3.5.2.6]	5
	K17838	<i>oxa</i>	<i>beta-lactamase class D</i> [EC:3.5.2.6]	1
cefalosporina	K01060	<i>cah</i>	<i>cephalosporin-C deacetylase</i> [EC:3.1.1.41]	10
cloranfenicol	K05786	<i>rarD</i>	<i>chloramphenicol-sensitive protein RarD</i>	17
	K07552	<i>bcr</i>	<i>MFS transporter, DHA1 family, bicyclomycin/chloramphenicol resistance protein</i>	83
	K08160	<i>mdfA, cmr</i>	<i>MFS transporter, DHA1 family, multidrug/chloramphenicol efflux transport protein</i>	3
	K18554	<i>cpt</i>	<i>chloramphenicol 3-O phosphotransferase</i> [EC:2.7.1.-]	5
	K19271	<i>catA</i>	<i>chloramphenicol O-acetyltransferase type A</i> [EC:2.3.1.28]	1
estreptomicina	K04343	<i>strB</i>	<i>streptomycin 6-kinase</i> [EC:2.7.1.72]	3
tetraciclinas	K08151	<i>tetA</i>	<i>MFS transporter, DHA1 family, tetracycline resistance protein</i>	22
	K18214	<i>tetP_A, tet40</i>	<i>MFS transporter, DHA3 family, tetracycline resistance protein</i>	1
	K18220	<i>tetM, tetO</i>	<i>ribosomal protection tetracycline resistance protein</i>	2
	K18476	<i>tetR</i>	<i>TetR/AcrR family transcriptional regulator, tetracycline repressor protein</i>	1
vancomicina	K03862	<i>vanA</i>	<i>vanillate O-demethylase monooxygenase subunit</i> [EC:1.14.13.82]	26
	K03863	<i>vanB</i>	<i>vanillate O-demethylase ferredoxin subunit</i> [EC:1.14.13.82]	36
	K07260	<i>vanY</i>	<i>D-alanyl-D-alanine carboxypeptidase</i> [EC:3.4.16.4]	3
	K08641	<i>vanX</i>	<i>D-alanyl-D-alanine dipeptidase</i> [EC:3.4.13.22]	8
	K11475	<i>vanR</i>	<i>GntR family transcriptional regulator, vanillate catabolism transcriptional regulator</i>	11
	K18346	<i>vanW</i>	<i>vancomycin resistance protein VanW</i>	4
multi resistencia	K03712	<i>marR</i>	<i>MarR family transcriptional regulator, multiple antibiotic resistance protein MarR</i>	1
	K05595	<i>marC</i>	<i>multiple antibiotic resistance protein</i>	94

Finalmente, y al igual que en FLA, se encontraron genes relacionados con la resistencia y tolerancia a metales pesados tales como; transportadores cobalto/níquel, transportadores catiónicos cobalto/zinc/cadmio y transportadores zinc/cadmio. Junto con genes que codifican a proteínas de

resistencia a níquel, arsénico, cromo, cadmio y mercurio y mecanismos de resistencia multi metal (Tabla 19).

Tabla 19. Lista de genes relacionados con adaptaciones a resistencia a metales pesados presentes en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO).

	KO ID	Nombre	Producto	N° genes
Arsénico	K00537	ARSC1, <i>arsC</i>	<i>arsenate reductase</i> [EC:1.20.4.1]	52
	K03741	ARSC2, <i>arsC</i>	<i>arsenate reductase</i> [EC:1.20.4.1]	122
	K03893	<i>arsB</i>	<i>arsenical pump membrane protein</i>	97
	K11811	<i>arsH</i>	<i>arsenical resistance protein ArsH</i>	46
	K08356	<i>aoxB</i>	<i>arsenite oxidase large subunit</i> [EC:1.20.9.1]	22
	K08355	<i>aoxA</i>	<i>arsenite oxidase small subunit</i> [EC:1.20.9.1]	13
	K03325	TC.ACR3	<i>arsenite transporter, ACR3 family</i>	142
	K01551	<i>arsA</i> , ASNA1	<i>arsenite-transporting ATPase</i> [EC:3.6.3.16]	108
Cromo	K07240	<i>chrA</i>	<i>chromate transporter</i>	134
	K19784	<i>chrR</i>	<i>chromate reductase</i>	42
Mercurio	K00221	E4.99.1.2	<i>alkylmercury lyase</i> [EC:4.99.1.2]	6
	K00520	<i>merA</i>	<i>mercuric reductase</i> [EC:1.16.1.1]	134
	K08363	<i>merT</i>	<i>mercuric ion transport protein</i>	62
	K08364	<i>merP</i>	<i>periplasmic mercuric ion binding protein</i>	41
	K08365	<i>merR</i>	<i>MerR family transcriptional regulator, mercuric resistance operon regulatory protein</i>	111
	K19057	<i>merD</i>	<i>MerR family transcriptional regulator, mercuric resistance operon regulatory protein</i>	5
	K19058	<i>merC</i>	<i>mercuric ion transport protein</i>	12
	K19059	<i>merE</i>	<i>mercuric ion transport protein</i>	9
Níquel	K00518	<i>sodN</i>	<i>nickel superoxide dismutase</i> [EC:1.15.1.1]	10
	K02031	ABC.PE.A	<i>peptide/nickel transport system ATP-binding protein</i>	97
	K02032	ABC.PE.A1	<i>peptide/nickel transport system ATP-binding protein</i>	92
	K02033	ABC.PE.P	<i>peptide/nickel transport system permease protein</i>	199
	K02034	ABC.PE.P1	<i>peptide/nickel transport system permease protein</i>	177
	K02035	ABC.PE.S	<i>peptide/nickel transport system substrate-binding protein</i>	316
	K04651	<i>hypA</i> , <i>hybF</i>	<i>hydrogenase nickel incorporation protein HypA/HybF</i>	48
	K04652	<i>hypB</i>	<i>hydrogenase nickel incorporation protein HypB</i>	49
	K07241	<i>nixA</i>	<i>high-affinity nickel-transport protein</i>	112
	K07722	<i>nikR</i>	<i>CopG family transcriptional regulator, nickel-responsive regulator</i>	11

Tabla 19. Lista de genes relacionados con adaptaciones a resistencia a metales pesados presentes en MM, obtenida a partir del proceso de anotación en IMG utilizando las bases de datos de KEGG orthology (KO) (continuación).

	KO ID	Nombre	Producto	N° genes
Multi-Metal	K02006	<i>cbiO</i>	<i>cobalt/nickel transport system ATP-binding protein</i>	13
	K02007	<i>cbiM</i>	<i>cobalt/nickel transport system permease protein</i>	13
	K02008	<i>cbiQ</i>	<i>cobalt/nickel transport system permease protein</i>	13
	K02009	<i>cbiN</i>	<i>cobalt/nickel transport protein</i>	5
	K06189	<i>corC</i>	<i>magnesium and cobalt transporter</i>	40
	K15725	<i>czcC</i>	<i>outer membrane protein, cobalt-zinc-cadmium efflux system</i>	28
	K15726	<i>czcA</i>	<i>cobalt-zinc-cadmium resistance protein CzcA</i>	185
	K15727	<i>czcB</i>	<i>membrane fusion protein, cobalt-zinc-cadmium efflux system</i>	60
	K16264	<i>czcD, zitB</i>	<i>cobalt-zinc-cadmium efflux system protein</i>	117
	K16267	<i>zipB</i>	<i>zinc and cadmium transporter</i>	17

5.4.3 Diversidad taxonómica

Contrario a FLA, en el metagenoma de MM se encontró una mayor cantidad de secuencias de genes ribosomales 16S rARN con 40 registros, de los cuales, 32 corresponden únicamente a bacterias (**Tabla 20**) y ocho a organelos de eucariontes (**Tabla S3**). Aunque, de igual forma como sucedió con FLA, los resultados del análisis en BLAST coincidieron con lo hallado en el proceso de anotación (**Figura 7**) y parcialmente con el análisis taxonómico en SILVAngs (**Figura 14**).

Tabla 20. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a bacterias encontradas en el metagenoma de MM realizadas en la plataforma de BLASTN de NCBI. Para las secuencias cuyo mejor hit fue una secuencia desconocida, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida.

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
1161	<i>Proteobacteria</i>	<i>γ-proteobacteria</i>	<i>Metallibacterium scheffleri</i> S30H22	MT605224	99.66	100
1544	<i>Firmicutes</i>	<i>Negativicutes</i>	<i>Uncultured bacterium clone</i> RABS_B58	HQ660794	94.02	98
			<i>Pelosinus defluvii</i> SHI-1	NR109398	93.18	99
1488	<i>Acidobacteria</i>	<i>Acidobacteriia</i>	<i>Acidobacterium capsulatum</i> ATCC 51196	NR074106	99.33	100
1517	<i>Proteobacteria</i>	<i>β-proteobacteria</i>	<i>Uncultured bacterium clone</i> IN_MF_1830	MT362442	97.37	100
			<i>Thiomonas intermedia</i> K12	CP002021	96.05	100

Tabla 20. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a bacterias encontradas en el metagenoma de MM realizadas en la plataforma de BLASTN de NCBI. Para las secuencias cuyo mejor hit fue una secuencia desconocida, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida (continuación).

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
1250	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Uncultured bacterium clone CEM_g6</i>	JQ815670	99.43	98
			<i>Acidimicrobium ferrooxidans</i> DSM 10331	NR074390	94.25	100
1234	<i>Acidobacteria</i>	<i>Acidobacteriia</i>	<i>Acidobacterium capsulatum</i> ATCC 51196	NR074106	99.27	100
1230	<i>Firmicutes</i>	<i>Bacilli</i>	<i>Uncultured Alicyclobacillaceae bacterium clone S3BC3</i>	GU229856	99.15	95
			<i>Alicyclobacillus pomorum</i> 3A	NR024801	92.34	99
1479	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Acidiphilium multivorum</i> AIU301	AP012035	96.89	100
1243	<i>Proteobacteria</i>	<i>Acidithiobacillia</i>	<i>Acidithiobacillus thiooxidans</i> ATCC 19377	CP045571	98.79	100
1524	<i>Proteobacteria</i>	γ - <i>proteobacteria</i>	<i>Uncultured bacterium clone G51-19</i>	DQ480485	99.4	98
			<i>Acidibacter ferrireducens</i> MCF85	NR126260	97.65	80
1473	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium clone STPCL-2</i>	HM124389	97.49	99
			<i>Acidocella aromatica</i> LGS-3	AF253413	97.73	95
1475	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium clone STPCL-2</i>	HM124389	97.15	99
			<i>Acidocella aluminiidurans</i> AL46	NR112716	96.95	99
1550	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured bacterium clone O117206E10</i>	FN865191	99.89	58
			<i>Thermoanaerobacter wiegelii</i> Rt8.B1	NR075059	76.3	55
1530	<i>Proteobacteria</i>	γ - <i>proteobacteria</i>	<i>Mangrovibacter</i> sp. AB6	JX188076	99.67	100
			<i>Mangrovibacter yixingensis</i> TULL-A	NR_136780	99.93	95
1152	<i>Proteobacteria</i>	β - <i>proteobacteria</i>	<i>Thiomonas cuprina</i> YE4-D1-1-CH	FN870344	98.87	100
1480	<i>Actinobacteria</i>	<i>Acidimicrobiia</i>	<i>Uncultured bacterium, clone Central-Bottom-cDNA_clone56</i>	HE604016	98.85	100
			<i>Aciditerrimonas ferrireducens</i> JCM 15389	NR112972	89.46	98
1481	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium clone RS-G27</i>	KC541487	97.3	99
			<i>Acidiphilium iwatense</i>	AB561883	94.82	100
1519		<i>Phycisphaerae</i>	<i>Uncultured bacterium clone NSCom-bac_h10</i>	KC619551	98.73	98
			<i>Tepidisphaera mucosa</i> 2842	NR148582	79.92	100
1481	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Acidocella aluminiidurans</i> AL46	NR112716	94.87	99

Tabla 20. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a bacterias encontradas en el metagenoma de MM realizadas en la plataforma de BLASTN de NCBI. Para las secuencias cuyo mejor hit fue una secuencia desconocida, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida (continuación).

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
1556	<i>Proteobacteria</i>	δ - <i>proteobacteria</i>	<i>Uncultured proteobacterium clone pCOF 65.7 H11</i>	EU156151	99.15	98
			<i>Desulfurella acetivorans</i> A63	NR121723	98.27	100
1472	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium clone BT1028</i>	JF800770	96.08	100
			<i>Phaeospirillum molischianum</i> DSM 120	NR104755	95.93	99
1202	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Rickettsiaceae endosymbiont of Carteria cerasiformis rrs, nlpD2</i>	LC004724	99.25	100
			<i>Rickettsia heilongjiangensis</i> 054	NR074469	94.52	100
1498	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured bacterium clone 500-69</i>	FR687470	98.78	98
			<i>Clostridium magnum</i> MM5	GU129927	98.78	96
964	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Uncultured bacterium clone 500-69</i>	FR687470	97.31	100
			<i>Clostridium magnum</i> OZK30	KT696502	97.11	100
1556	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Desulfosporosinus orientis</i> DSM 765	CP003108	96.41	100
1470	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium, clone IRON_SNOW_CB_C1_P2</i>	FR667789	99.72	98
			<i>Rhodoplanes cryptolactis</i>	AB087718	90.72	100
1490	<i>Proteobacteria</i>	α - <i>proteobacteria</i>	<i>Uncultured bacterium clone EV818SWSAP21</i>	DQ337066	96.73	98
			<i>Trojanella thessalonices</i>	AF069496	86.42	97
1526	<i>Proteobacteria</i>	β - <i>proteobacteria</i>	<i>Uncultured bacterium clone G5</i>	DQ480481	96.68	98
			<i>Ferritrophicum radicola</i> CCJ	DQ386263	95.14	99
1083	<i>Nitrospirae</i>	<i>Nitrospirales</i>	<i>Uncultured bacterium clone placa1_b9</i>	EF446274	99.91	100
			<i>Leptospirillum ferrooxidans</i> Sy	AF356839	99.63	100
1512	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Uncultured bacterium clone 24c04</i>	EF515423	96.99	92
			<i>Microbacter margulisiae</i> ADRI	NR126316	94.9	98
1544	<i>Firmicutes</i>	<i>Negativicutes</i>	<i>Uncultured bacterium clone RABS_B58</i>	HQ660794	94.02	99
			<i>Pelosinus defluvii</i> SHI-1	NR109398	93.18	99
1054	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridium acidisoli</i> CK74	AJ237756	96.88	100

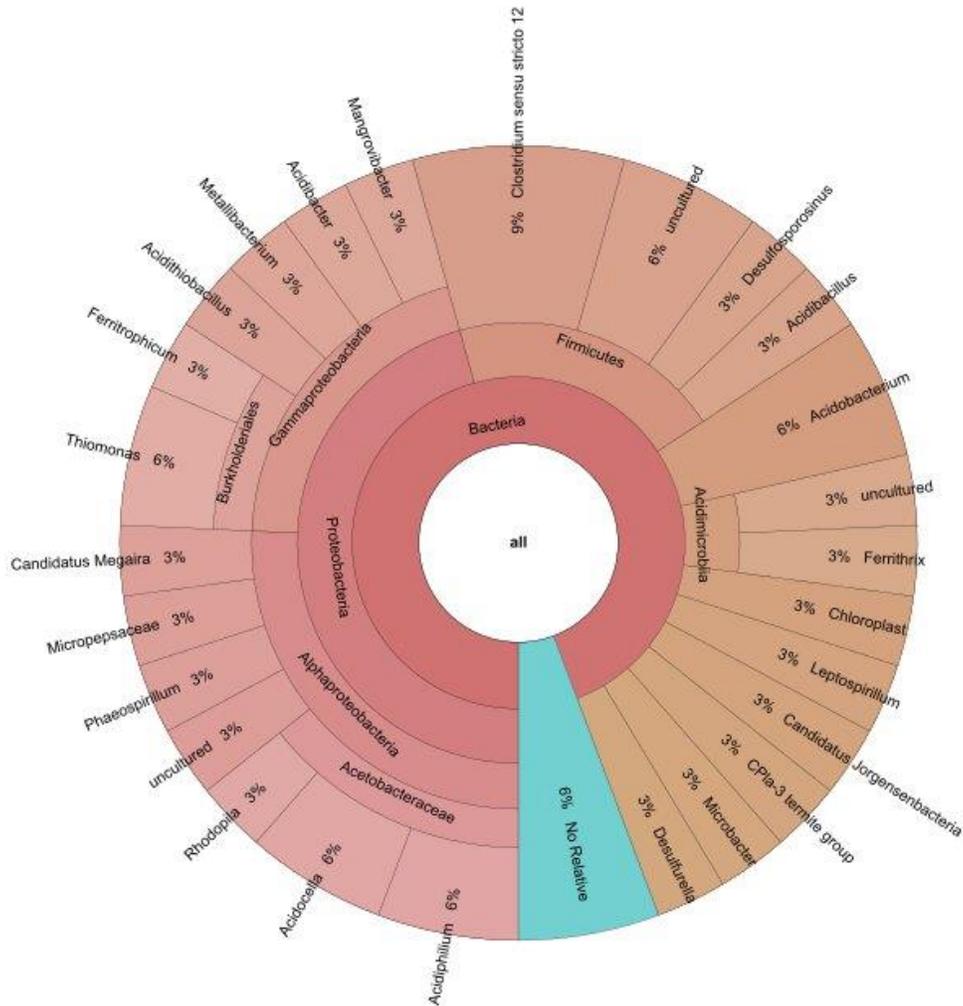


Figura 14. Análisis taxonómico de las secuencias de 16S rARN derivadas del metagenoma de FLA realizado en la plataforma SILVAngs. El grafico presentado fue generado automáticamente para ser mostrado a través de la plataforma KRONA.

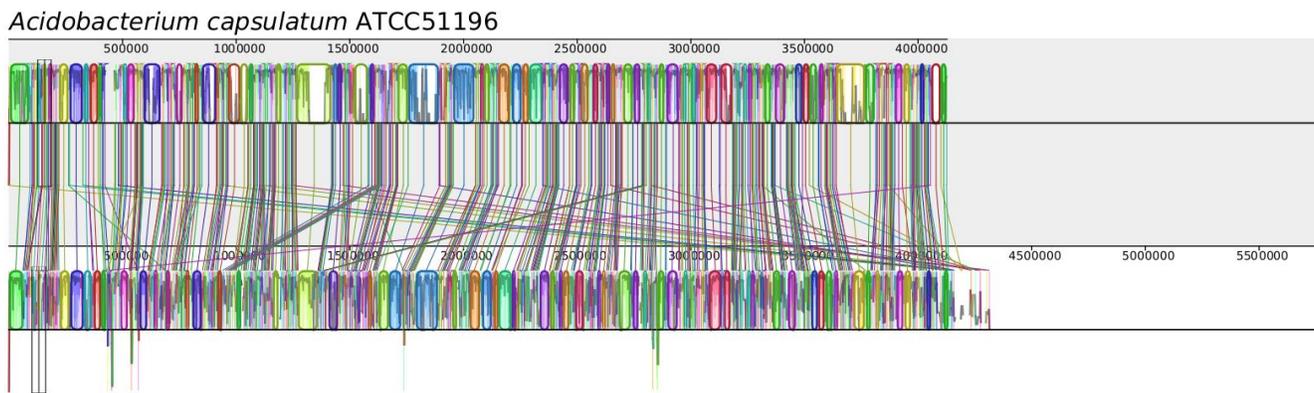
5.4.4 Recuperación de genomas

5.4.4.1 Análisis genómico comparativo

De forma similar al caso de FLA, en MM fue posible recuperar genomas parciales de dos bacterias a partir de genomas de referencia, usando como base los resultados de la anotación taxonómica de genes ribosomales 16S rARN. Para este caso fue posible recuperar un genoma parcial de un microorganismo cercano al género *Acidobacterium* y otro genoma parcial cercano al género *Metallibacterium* denominados; *Candidatus Acidobacterium* sp. MM01 y *Candidatus Metallibacterium* sp. MM02 respectivamente.

'Candidatus Acidobacterium sp. MM01'

Es un genoma parcial de 5.7 Mb distribuidos en 2,119 *contigs* con un valor de N50 de 6,753 pb (**Tabla S7**), el alineamiento con el genoma de referencia muestra un mayor tamaño del genoma recuperado, además de una elevada fragmentación. Sin embargo, se alcanzan a denotar algunas regiones conservadas sin inversiones y algunas transposiciones (**Figura 15**).



Candidatus Acidobacterium sp. MM01

Figura 15. Alineamiento global desarrollado en MAUVE entre el genoma de *Acidobacterium capsulatum* ATCC51196 (parte superior) y *'Ca. Acidobacterium sp. MM01'* (parte inferior). Las regiones conservadas se muestran en los bloques coloreados.

La anotación de este genoma (**Tabla 22**) muestra que posee 6,409 secuencias codificantes distribuidas en 240 subsistemas, 78 ARNs y ningún CRISPR, la información derivada revela que al igual que los recuperados en FLA. Esta bacteria posee adaptaciones a condiciones ambientales adversas ya mencionadas tales como el choque térmico (*grpE*, *hspR*), el estrés osmótico (*ompA*) y oxidativo (*ahpC*, *ohr*, *sod*), tolerancia a altas concentraciones de monóxido de nitrógeno (*nnsR*, *nsrR*) y radicales libres derivados de azufre (*bcp*). Mecanismos de defensa y resistencia a metales como mercurio (*merAT*, *mir*) cobre (*copAB*, *cutACE*, *corC*) cobalto-zinc-cadmio (*czcD*, *cusB/czsB*), resistencia a antibióticos betalactámicos (*phnP*) y fluoroquinolonas (*gyrAB*).

Por otro lado, en el espectro metabólico, este organismo posee genes relacionados con la asimilación de amonio (*glnADB*, *amt*). De la misma forma se detectó transportadores y genes relacionados con la incorporación de pirofosfato (*ppa*) polifosfatos (*ppx*, *ppk*, *ppgK*), regulación y transporte de fosfatos (*phoURB*) y escasez de Fósforo (*phoH*).

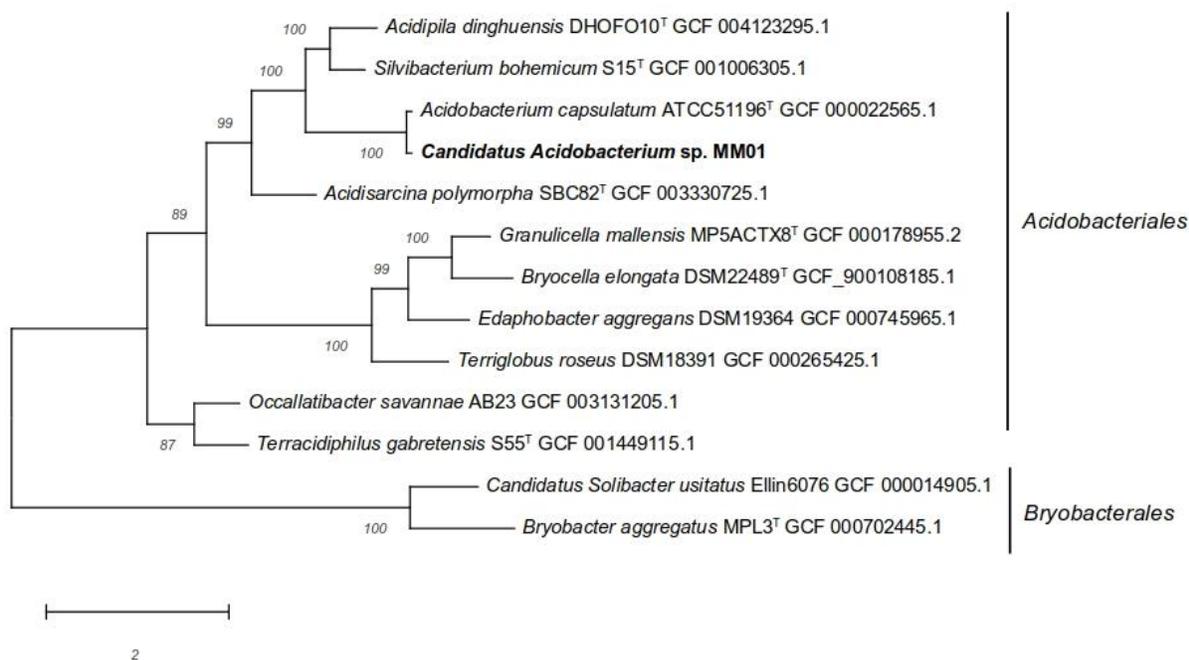


Figura 16. Análisis filogenómico en el cual se representa la relación evolutiva predicha entre ‘*Ca. Acidobacterium sp. MM01*’ y diferentes secuencias de integrantes del orden *Acidobacteriales*. El árbol fue reconstruido con PhyloPhlAn utilizando 376 alineamientos de marcadores moleculares, PhyloPhlAn realiza alineamientos individuales de cada set de proteínas recuperados de los genomas introducidos correspondientes a la clase *Acidobacteriia*. Así, PhyloPhlAn concatena las posiciones más discriminativas en cada alineamiento de proteínas en una única secuencia larga para reconstruir un árbol filogenético usando FastTree. La posición de ‘*Ca. Acidobacterium sp. MM01*’ se resalta en negrita. Los números de accesos se encuentran seguidos a cada uno de los genomas. Los números en los soportes de ramas representan los valores de probabilidades (basado en 1000 remuestreos). La barra de escala en la parte inferior izquierda representa el número estimado de sustituciones de aminoácidos por sitio por unidad de longitud de rama.

A nivel taxonómico, si bien este genoma presenta una clara relación con el genoma de referencia a nivel filogenómico (**Figura 16**), al observar los métricos de ANI y de *in silico* DDH los resultados son más controversiales (**Tabla 21**), debido a que los valores de estos métricos nos están indicando que probablemente no se trate de la misma especie de bacteria.

Tabla 21. Métricas de identidad de secuencia utilizados para la identificación del genoma de ‘*Ca. Acidobacterium* sp. MM01’

	‘ <i>Ca. Acidobacterium</i> sp. MM01’	<i>Acidobacterium capsulatum</i> ATCC51196
Número de <i>contigs</i>	2,119	1 cromosoma
Tamaño del genoma	5.75 Mb	4.13 Mb
ANIb	100	90.41
ANIm	100	90.94
<i>In silico</i> DDH	100	39.6
GC%	61.1	60.5

Tabla 22. Subsistemas detectados en los genomas recuperados de ‘*Candidatus Acidobacterium* sp. MM01’ y ‘*Candidatus Metallibacterium* sp. MM02’ por medio de análisis de anotación en RAST.

Subsistema	‘ <i>Ca. Acidobacterium</i> sp. MM01’	‘ <i>Ca. Metallibacterium</i> sp. MM02’
Cofactores, vitaminas, grupos prostéticos, pigmentos	138	246
Pared celular y cápsula	61	56
Virulencia, enfermedad y defensa	175	85
Metabolismo del K	15	15
Fotosíntesis	8	0
Misceláneos	2	20
Fagos, profagos, elementos transponibles, plásmidos	0	1
Transporte de membrana	127	30
Metabolismo y adquisición de Fe	1	0
Metabolismo ARN	147	75
Nucleósidos y nucleótidos	161	67
Metabolismo de las proteínas	494	162
Motilidad y Quimiotaxis	24	
Regulación y señalización celular	21	12
Metabolismo secundario	14	0
Metabolismo ADN	203	30
Ácidos grasos, lípidos, e isoprenoides	65	94
Metabolismo del N	19	26
Dormancia y esporulación	2	0
Respiración	189	126
Respuesta a estrés	22	16
Metabolismo de compuestos aromáticos	24	12
Aminoácidos y derivados	389	172

Tabla 22. Subsistemas detectados en los genomas recuperados de ‘*Candidatus Acidobacterium* sp. MM01’ y ‘*Candidatus Metallibacterium* sp. MM02’ por medio de análisis de anotación en RAST (continuación).

Subsistema	‘ <i>Ca. Acidobacterium</i> sp. MM01’	‘ <i>Ca. Metallibacterium</i> sp. MM02’
Metabolismo del S	19	10
Metabolismo del P	50	39
Carbohidratos	266	194

‘*Candidatus Metallibacterium* sp. MM02’

Su genoma está compuesto por 5.6 Mb distribuidos en 2,062 *contigs*, con un valor de N50 de 6,309 pb (**Tabla S8**). De igual forma que ‘*Ca. Acidobacterium* sp. MM01’, el genoma recuperado posee una mayor longitud que el genoma de referencia, además de encontrarse altamente fragmentado. En contraste, el alineamiento muestra algunas regiones relativamente conservadas en las partes medias del genoma, con altos rastros de variabilidad en los extremos (**Figura 17**).

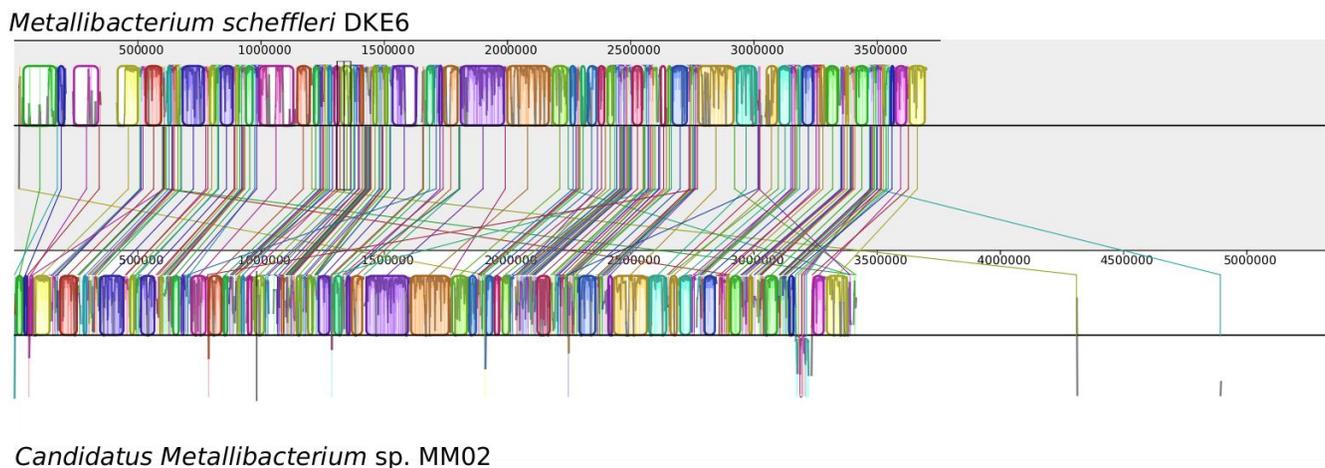


Figura 17. Alineamiento genómico desarrollado en MAUVE entre los genomas de *Metallibacterium scheffleri* DKE6 (parte superior) y de ‘*Ca. Metallibacterium* sp. MM02’ (parte inferior). Las regiones conservadas se muestran en los bloques coloreados.

La anotación de este genoma (**Tabla 22**) muestra que posee 6,266 secuencias codificantes distribuidas en 274 subsistemas, 113 ARNs y ningún CRISPR, al revisar detalladamente los subsistemas se encontró que esta bacteria muestra tolerancia a diversos tipos de estrés ambiental como el choque térmico (*grpE*) El estrés oxidativo, tanto por radicales de oxígeno (*ahpC*, *sodB*, *rdr*, *ahr*, *oxyR*, *fur*) concentraciones elevadas de monóxido de nitrógeno (*nrsR*) y radicales de azufre (*trxR*, *bcp*, *ahpC*), estrés osmótico (*ompA*, *ndvB*), Mecanismos de defensa y resistencia a metales pesados como: mercurio por medio de la presencia de mercurio reductasas (*merAT*, *mir*), el cobre a través de genes relacionados con la homeostasis y tolerancia (*copAB*, *cutA*, *corC*), la existencia de complejos de resistencia cobalto-zinc-cadmio (*czcAD*, *cusB/czsB*, *czrR*,) y girasas con resistencia a fluoroquinolonas (*gyrAB*).

En cuanto a su metabolismo, fue posible detectar genes relacionados con la escasez de carbono (*csrA*, *sipP*, *sspAB*), y la presencia de un carboxisoma que permite la fijación de CO₂ (*rbcS*, *rbcL*, *cbbR*). Genes involucrados en la adquisición de hierro (*tonB*) y su uso como aceptor final de electrones por medio de ferredoxina reductasas (FR) y hetero disulfuro reductasas (*hdrB*). Para el metabolismo del nitrógeno se observa que hay genes relacionados con la asimilación de amonio (*glnADE*, *gltAD*, *amt*) y de nitrificación (*nirK*, *nor*, *nnrS*). Por último, en cuanto al metabolismo del fósforo, se detectaron variedad de transportadores y elementos de regulación para la adquisición de fosfatos (*phoBRU*, *alkP*) y con la escasez de fosfatos (*phoH*) y adquisición de pirofosfatos (*ppa*) y polifosfatos (*ppx*, *ppk*, *gppA*).

En cuanto a su clasificación, si bien se denota una alta relación entre el genoma de referencia y el genoma de ‘*Ca. Metallibacterium* sp. MM02’ (**Figura 18**), se observó un patrón similar al mencionado en el genoma anterior, donde los resultados de los métricos ANI e *in silico* DDH arrojan resultados los cuales indican que estos dos genomas poseen diferencias entre ellos para considerar que no se trata de la misma especie de bacteria (**Tabla 23**).

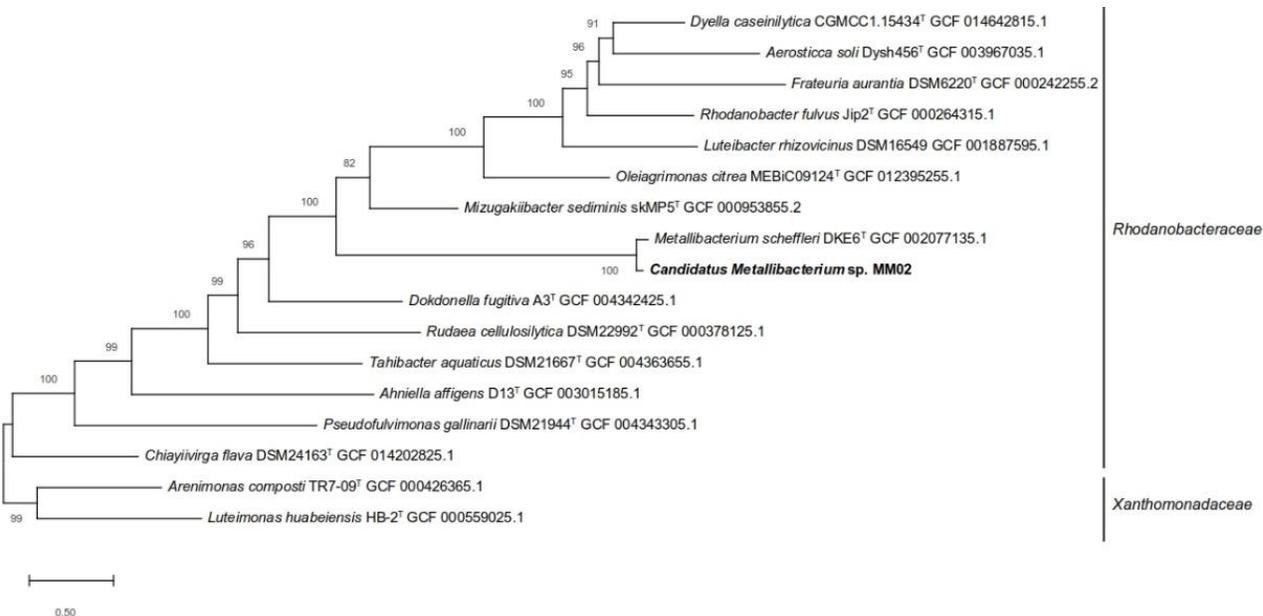


Figura 18. Análisis filogenómico en el cual se representa la relación evolutiva predicha entre ‘*Ca. Metallibacterium sp. MM02*’ y diferentes secuencias de integrantes de la familia *Rhodanobacteraceae*. El árbol fue reconstruido con PhyloPhlAn utilizando 378 alineamientos de marcadores moleculares, PhyloPhlAn realiza alineamientos individuales de cada set de proteínas recuperados de los genomas introducidos correspondientes al orden *Xanthomonadales*. Así, PhyloPhlAn concatena las posiciones más discriminativas en cada alineamiento de proteínas en una única secuencia larga para reconstruir un árbol filogenético usando FastTree. La posición de ‘*Ca. Metallibacterium sp. MM02*’ se resalta en negrita. Los números de accesos se encuentran seguidos a cada uno de los genomas. Los números en los soportes de ramas representan los valores de probabilidades (basado en 1000 remuestreos). La barra de escala en la parte inferior izquierda representa el número estimado de sustituciones de aminoácidos por sitio por unidad de longitud de rama.

Tabla 23. Métricas de identidad de secuencia utilizados para la identificación del genoma de ‘*Ca. Metallibacterium sp. MM02*’.

	<i>Ca. Metallibacterium sp. MM02</i>	<i>Metallibacterium scheffleri</i> DKE6
Número de <i>contigs</i>	2,062	5
Tamaño del genoma	5.37	3.76
ANiB	100	95.88
ANIm	100	96.3
dDDH	100	58.5
GC%	66.8	65.8

5.4.4.2 Construcción de genomas a partir de metagenomas (MAG) vía binning

A diferencia de los MAGs reconstruidos en el metagenoma de FLA, en MM la cantidad fue mayor (44 MAGs), todos correspondientes al Dominio Bacteria (**Tabla 24**). Entre los MAGs recuperados se encuentran principalmente el filum *Proteobacteria*, y en menor medida *Firmicutes* y *Actinobacteria*. De todos los MAGs recuperados, 13 poseen una completitud superior al 95%, siendo el MaxBin_MM.005 (**Tabla 24**) el MAG que posee un 100% de completitud con respecto a los marcadores de CheckM (**Figura S3B**), de acuerdo con el perfil taxonómico, este MAG posee relación con el género *Acidisoma*.

Tabla 24. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de MM vía *binning*. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud del 90%. (n.d) Rangos taxonómicos no determinados.

# bin	Dominio	Phylum	Clase	Orden	Familia	Género	Especie
CONCOCT_MM.100.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rickettsiales	unclassified Rickettsiales	n.d	n.d
CONCOCT_MM.116.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidibrevibacterium	Acidibrevibacterium fodinaquatile
CONCOCT_MM.118.fa	Bacteria	Firmicutes	Negativicutes	Selenomonadales	Sporomusaceae	n.d	n.d
CONCOCT_MM.126_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	n.d	n.d
CONCOCT_MM.133.fa	Bacteria	Actinobacteria	Acidimicrobiia	Acidimicrobiales	Acidimicrobiaceae	Ferrithrix	Ferrithrix thermotolerans
CONCOCT_MM.138.fa	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	Desulfosporosinus	
CONCOCT_MM.142.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	unclassified Rhodospirillales	unclassified Rhodospirillales (miscellaneous)	Rhodospirillales bacterium 20-64-7
CONCOCT_MM.142.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidocella	n.d
CONCOCT_MM.15.fa	Bacteria	Actinobacteria	Acidimicrobiia	Acidimicrobiales	Acidimicrobiaceae	Ferrithrix	Ferrithrix thermotolerans*
CONCOCT_MM.17.fa	Bacteria	Firmicutes	Clostridia	Thermoanaerobacterales	Thermodesulfobiaceae	Thermodesulfobium	n.d
CONCOCT_MM.24.fa	Bacteria	Bacteria candidate phyla	Patescibacteria group	Parcubacteria group	n.d	n.d	n.d
CONCOCT_MM.30_sub.fa	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfurellales	Desulfurellaceae	Desulfurella	Desulfurella amilsii
CONCOCT_MM.34_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidiphilium	n.d
CONCOCT_MM.53.fa	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Metallibacterium	Metallibacterium scheffleri
CONCOCT_MM.56_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidiphilium	n.d

Tabla 24. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de MM vía *binning*. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud del 90%. (n.d) Rangos taxonómicos no determinados (continuación).

# bin	Dominio	Phylum	Clase	Orden	Familia	Género	Especie
CONCOCT_MM.62_sub.fa	Bacteria	Proteobacteria	Acidithiobacillia	Acidithiobacillales	Acidithiobacillaceae	Acidithiobacillus*	n.d
CONCOCT_MM.83.fa	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified Burkholderiales	Thiomonas	Thiomonas sp. FB-6
CONCOCT_MM.84_sub.fa	Bacteria	Firmicutes	Bacilli	Bacillales	n.d	n.d	n.d
MaxBin_MM.004_sub.fa	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified Burkholderiales	Thiomonas	n.d
MaxBin_MM.005.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidisoma	unclassified Acidisoma
MaxBin_MM.009_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidiphilium	unclassified Acidiphilium
MaxBin_MM.025_sub.fa	Bacteria	Firmicutes	Bacilli	Bacillales	unclassified Bacillales	Acidibacillus	n.d
MaxBin_MM.036_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	n.d	n.d	n.d	n.d
MaxBin_MM.052_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	n.d	n.d	n.d	n.d
MaxBin_MM.061_sub.fa	Bacteria	Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified Burkholderiales	Thiomonas	n.d
MaxBin_MM.067.fa	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Mangrovibacter	n.d
MetaBAT_M M.12_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	n.d	n.d	n.d
MetaBAT_M M.15.fa	Bacteria	Actinobacteria	Actinobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.16_sub.fa	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	n.d	n.d
MetaBAT_M M.22_sub.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidocella	unclassified Acidocella
MetaBAT_M M.25.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidiphilium	Acidiphilium sp. 20-67-58
MetaBAT_M M.26.fa	Bacteria	Actinobacteria	unclassified Actinobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.38.fa	Bacteria	Proteobacteria	Gammaproteobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.39_sub.fa	Bacteria	Proteobacteria	Gammaproteobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.40.fa	Bacteria	Actinobacteria	Acidimicrobiia	Acidimicrobiales	Acidimicrobiaceae	Acidithrix	Acidithrix ferrooxidans*
MetaBAT_M M.42.fa	Bacteria	Acidobacteria	Acidobacteriia	Acidobacteriales	Acidobacteriaceae	Acidobacterium	Acidobacterium capsulatum
MetaBAT_M M.43.fa	Bacteria	Proteobacteria	Betaproteobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.44.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidocella	Acidocella aminolytica
MetaBAT_M M.52.fa	Bacteria	PVC group	n.d	n.d	n.d	n.d	n.d
MetaBAT_M M.58.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidocella	n.d
MetaBAT_M M.62.fa	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodospirillales	n.d	n.d	n.d
MetaBAT_M M.68_sub.fa	Bacteria	Acidobacteria	Acidobacteriia	Acidobacteriales	Acidobacteriaceae	n.d	n.d

Tabla 24. Perfil taxonómico de los MAGs recuperados a partir de reconstrucción de genomas del metagenoma de MM vía *binning*. Los nombres en rojo corresponden a MAGs con un porcentaje de completitud del 90%. (n.d) Rangos taxonómicos no determinados (continuación).

# bin	Dominio	Phylum	Clase	Orden	Familia	Género	Especie
MetaBAT_M M.71.fa	Bacteria	Actinobacteria	Actinobacteria	n.d	n.d	n.d	n.d
MetaBAT_M M.79.fa	Bacteria	Acidobacteria	Acidobacteriia	Acidobacteriales	Acidobacteriaceae	Acidobacterium	Acidobacterium capsulatum
MetaBAT_M M.83.fa	Bacteria	Proteobacteria	Deltaproteobacteria	Desulfurellales	Desulfurellaceae	Desulfurella	Desulfurella amilsii

6. DISCUSIÓN

6.1 Fumarolas de los Azufres (FLA)

A partir de los resultados presentados, se encontraron filotipos presentes dentro del metagenoma de FLA que fueron reportados en estudios previos Servín Garcidueñas y Martínez Romero (2015). Donde se detectaron integrantes de los géneros *Leptospirillum*, *Aciditerrimonas*, *Acidithiobacillus*, *Ferrimicrobium* y *Sufobacillus* (**Tabla 9**). El estudio de Servín Garcidueñas y Martínez Romero, (2015) detectó previamente poblaciones de gammaproteobacterias y alfaproteobacterias no descritas, las cuales han sido encontradas en este estudio (**Tabla 9, Figura 6**). Aunque, con la actualización de las bases de datos, se ha podido corroborar con mayor precisión el taxa al cual pertenecen algunos de los registros de estas poblaciones, e incluso, se pudo recuperar un genoma, tal como es el caso de ‘*Ca. Acidibrevibacterium* sp. FLA01’, del cual se hablará con mayor detalle más adelante. Además, se detectó otra *gammaproteobacteria* cuya relación más cercana es con el género *Ectothiorhodospira* (**Tabla 9**).

Adicionalmente, en el trabajo de Servín Garcidueñas y Martínez Romero (2015) se han detectado poblaciones de actinobacterias que están relacionadas filogenéticamente con los géneros *Ferrimicrobium*, *Aciditerrimonas*, y *Conexibacter*. En este estudio fue posible detectar la presencia de estos géneros (**Figura 4, Tabla 9**) y recuperar de manera parcial un genoma correspondiente al género *Aciditerrimonas*. Por otro lado, se detectaron otras poblaciones con relación cercana a los géneros

Acidithiobacillum y *Spirillospora*. Logrando expandir la identificación de microorganismos de esta comunidad dentro de los sedimentos presentes en las fumarolas de los Azufres. El género *Acidithiobacillum* ha sido hallado previamente en ambientes geotérmicos (Norris et al., 2011), por lo que no es extraordinario su hallazgo en un lugar como las fumarolas de Los Azufres. Caso contrario ocurre con el género *Spirillospora* el cual se ha encontrado en otro tipo de ambientes extremos, por ejemplo, cavernas (Nakaew et al., 2009). Teniendo en cuenta que el filum *Actinobacteria* es el más abundante dentro de este metagenoma (**Figura 3**) y uno de los más abundantes en diversos ambientes de acuerdo con su versatilidad y capacidad de producir metabolitos secundarios (Shivlata & Satyanarayana, 2015) su presencia dentro de las fumarolas de los azufres puede ser determinante en la estructura de las comunidades microbianas en FLA.

El trabajo de Servín Garcidueñas y Martínez Romero (2015) reportó algunos integrantes del filo *Firmicutes*, los cuales también se encontraron en este estudio, tales como poblaciones de los géneros *Sulfobacillus* y *Allicyclobacillus* (**Figura 4, Figura 6, Tabla 9**). Sin embargo, en este estudio fue posible hallar registros cercanos con el género *Thermoaerobacter*, el cual no ha sido reportado por estudios previos en la zona. En contraste, en investigaciones de otros ambientes geotérmicos en México sí se han encontrado registros de este género, tal es el caso de las fumarolas presentes en inmediaciones del volcán Parícutín (Medrano-Santillana et al., 2017; Mora-González et al., 2020).

A pesar del uso de los filtros de RNAmmer, se detectó la presencia de Euryarchaeotas y Crenoarhaeotas en los resultados de identificación taxonómica (**Figura 6, Tabla S1**), siendo *Euryarchaeota* la más abundante. Aunque las arqueas no son parte del objetivo principal de este estudio, los resultados presentes en la Tabla S1 coinciden con trabajos previos realizados por (Marín-Paredes & Servín-Garcidueñas, 2020; Servín-Garcidueñas & Martínez-Romero, 2014) en los cuales, se describe con

mayor detalle la presencia y recuperación de genomas de arqueas del orden *Sulfolobales* presentes en los sedimentos de muestras provenientes de fumarolas de Los Azufres.

Tanto la anotación, los análisis de genes ribosomales como la reconstrucción de MAGs muestran la dominancia de comunidades de actinobacterias y proteobacterias, las cuales han sido detectadas en ambientes similares a los presentes en FLA, tal es el caso de las fumarolas presentes en el Volcán Parícutín en México (Souza-Brito et al., 2019) y en Vulcano, Italia (Fagorzi et al., 2019). Sin embargo, Mora-González et al. (2020) encontraron una alta dominancia de OTUs pertenecientes al filo *Firmicutes* y en menor medida a *Proteobacteria* en muestras de suelo provenientes de fumarolas ubicadas en el campo geotérmico de Los Humeros. Además, Wall et al. (2015) detectaron una gran abundancia de OTUs correspondientes a los filos *Chloroflexi* y *Cyanobacteria* detectados también en esta investigación, pero en abundancias menores. Aunque, es importante agregar que gran parte de los OTUs encontrados en dicho estudio no pudieron ser identificados. De modo que, es difícil establecer un patrón taxonómico estático dentro de las comunidades microbianas que habitan los sedimentos de las fumarolas, aun así, los estudios publicados de comunidades microbianas a lo largo del Eje Volcánico Transmexicano, parecen mostrar tendencias que podrían explicar cómo se distribuyen estos grupos taxonómicos, pero se requieren más estudios en otros ambientes similares dentro de esta región para poder corroborar dicho patrón.

6.1.1 Diversidad funcional

De acuerdo con los resultados de la anotación, un gran porcentaje de los genes presentes dentro de FLA son desconocidos (**Tabla 3**), esto pudo deberse en parte a las limitaciones propias de las bases de datos de IMG o de secuencias incompletas dentro del ensamblaje las cuales no pudieron ser identificadas por los algoritmos de anotación, por lo que no es extraño encontrar porciones de genes no identificados. Sin embargo, esto abre las puertas a explorar con mayor detalle dichas secuencias en estudios futuros, con el

fin de detectar genes novedosos. Dentro de los genes detectados, se pudieron catalogar funciones de relevancia tanto en el metabolismo central de las comunidades que habitan en FLA como en adaptaciones clave para su supervivencia en este ambiente (**Tablas 5-8**).

En primer lugar, cabe destacar el hallazgo de genes involucrados en la obtención de nutrientes fundamentales como el fósforo, donde se encontraron genes que codifican para la creación de transportadores y elementos de regulación que se activan ante la escasez de fósforo, lo que estaría reflejando que en estos sedimentos la adquisición de fósforo es limitada, por consiguiente, las comunidades bacterianas han desarrollado adaptaciones para poder hacer frente a la limitación de fósforo. Por otra parte, se observó que en el metabolismo del carbono que se está presentando tanto fenómenos de fijación de carbono y otros procesos como la incorporación de elementos inorgánicos como el hierro (**Tabla 5**). Procesos como la fijación de carbono, así como la incorporación de elementos inorgánicos suelen ser típicos de microorganismos que habitan en este tipo de ambientes (Hügler et al., 2010; Ilbert & Bonnefoy, 2013; Zhou et al., 2015).

En segundo lugar, se identificaron diversos genes relacionados con la síntesis y defensa ante componentes antimicrobianos, a la fecha, los estudios realizados en fumarolas presentes en zonas continentales no han reportado genes o rutas metabólicas involucradas en la síntesis y mecanismos de resistencia a componentes antimicrobianos. Contrario a ello, si se ha reportado la presencia de este tipo de componentes dentro de las comunidades microbianas presentes en fumarolas de sistemas hidrotermales marinos, tanto someros (Iqbalsyah et al., 2019), como profundos (Tortorella et al., 2018; Ziko et al., 2019). Allí se han encontrado sustancias con actividad antimicrobiana producidas por actino bacterias (similar a lo presentado en FLA) y hongos. Sin embargo, se trata de compuestos antimicrobianos diferentes a los encontrados en FLA. La lista de genes presentados en la Tabla 8, es una evidencia de que dichos compuestos son producidos por las comunidades bacterianas que habitan los

sedimentos de las fumarolas y por tanto pueden ser una fuente de nuevos antibióticos. Adicionalmente, la presencia de esta serie de componentes podría ser un indicio de actividad dinámica dentro de las comunidades de microorganismos, donde podrían estar ocurriendo fenómenos de competencia en respuesta al espacio o por la búsqueda de recursos limitados.

En tercer lugar, se detectaron genes responsables en la adaptación a cambios extremos de temperatura, tolerancia a la acidez junto a mecanismos de comunicación, interacción y transporte como lo son; la formación de biopelículas, una gran cantidad de transportadores ABC y comunicación bacteriana a través de *quorum sensing* (**Tabla 7**). Las adaptaciones ante el choque térmico y la acidez han sido uno de los más estudiados en comunidades bacterianas procedentes de ambientes geotermales, debido a la gran importancia que han tenido para el desarrollo de aplicaciones biotecnológicas (Mehta et al., 2016). Por otro lado, los transportadores ABC son fundamentales para la obtención de nutrientes, minerales y secreción de componentes antimicrobianos, por lo que una alta cantidad de estos últimos permitiría a los microorganismos poder incorporar rápidamente nutrientes necesarios o responder a perturbaciones a su entorno. Los mecanismos de comunicación bacteriana, aunque son ampliamente conocidos en bacterias ambientales y de suelo, aun suele ser desconocidos en microorganismos provenientes de ambientes termales (Kaur et al., 2018). Por lo que la información recopilada de ambientes como FLA son una evidencia de la presencia de estos mecanismos en las comunidades microbianas que habitan en este tipo de ambientes y por tanto, abren una puerta para aprender cómo interactúan las bacterias a la hora de establecer una organización incluso con otros microorganismos presentes en los sedimentos, tales como arqueas y eucariontes. La formación de biopelículas ha sido habitual en ambientes termales (Roy et al., 2020), por lo que no es de extrañar la presencia de genes involucrados en este proceso en FLA, la formación de estas estructuras contribuye a un manejo eficiente de los recursos presentes en el ambiente y a aumentar la supervivencia de las comunidades bacterianas en ambientes extremos.

Finalmente, se evidenció la presencia de genes relacionados con la tolerancia y resistencia a metales pesados (**Tabla 6**), la cual es fundamental para la supervivencia de las comunidades bacterianas en ambientes geotérmicos. El hallazgo de arseniato reductasas, por ejemplo, no solo les permite convertir los iones de arsénico en componentes más estables y por tanto menos tóxicos, sino que a su vez son aprovechados como aceptores de electrones, fundamentales en procesos metabólicos, la presencia de transportadores les permite movilizar estos elementos evitando su toxicidad. La existencia de transportadores de metales permite mantener las concentraciones de estos elementos en cantidades que no representen niveles de toxicidad (Ranawat & Rawat, 2018).

6.1.2 Genomas recuperados

‘*Candidatus Acidibrevibacterium* sp. FLA01’

Para este momento, ‘*Ca. Acidibrevibacterium* sp. FLA01’ es el segundo genoma reportado para el género *Acidibrevibacterium*. El primer reporte de un microorganismo de este género corresponde a *Acidibrevibacterium fodinaquatile* G45-3^T, recuperado de una muestra extraída de un drenaje de residuos ácidos en una mina en China (Muhadesi et al., 2019). A excepción de este estudio, no existen artículos que reporten esta bacteria en otras partes, tampoco se han descrito características adicionales a las reportadas en su genoma tipo.

Tanto ‘*Ca. Acidibrevibacterium* sp. FLA01’ como el genoma de referencia citado poseen un contenido de GC similar y un tamaño cercano entre ellos (**Tabla 10, Figura 7**). Además de tener una relación filogenética muy cercana (**Tabla 10, Figura 8**). Entre las 4,203 secuencias codificantes su mayoría se presentan en el metabolismo de los carbohidratos y de los aminoácidos, seguido por el metabolismo de las proteínas y la producción de cofactores, vitaminas y demás compuestos (**Tabla 11**). Para este genoma fue interesante el hallazgo de un fotosistema 2 el cual le confiere la capacidad de fijar CO₂, cabe recordar que es común encontrar este tipo de sistemas en proteobacterias (Badger & Bek, 2008) dicho

sistema también se encuentra incorporado a *A. fodinaquatile* G45-3^T. Otro gen de interés es la presencia de una hemo oxigenasa, la cual suele ser típica de organismos eucariontes, aunque encontrada hace algunos años en bacterias principalmente patógenas, parece ser una forma alternativa de obtener hierro a partir de grupos hemo, principalmente de fuente eucariota (Frankenberg-Dinkel, 2004). En este genoma ha sido posible observar la presencia de transportadores y elementos de regulación que promueven la incorporación de fósforo, lo que parece una adaptación a su escasez de este elemento esencial dentro de este ambiente, en cuanto al nitrógeno, la asimilación de este elemento a partir de cianato le permitiría a ‘*Ca. Acidibrevibacterium* sp. FLA01’ tener fuentes alternas de nitrógeno.

En cuanto a la tolerancia al estrés ambiental, ‘*Ca. Acidibrevibacterium* sp. FLA01’ parece tener una serie de mecanismos que le confieren la capacidad de resistir los cambios de temperatura, el estrés oxidativo y osmótico, condiciones que suelen ser fuertemente variables en los alrededores de las bocas de las fumarolas, donde las emanaciones de vapor son azarosas y abruptas, ocasionando cambios repentinos de temperatura y humedad. La presencia de genes que le confieren resistencia a antibióticos betalactámicos le permitiría hacer frente a competidores productores de sustancias antimicrobianas. Por otra parte, la presencia de transportadores y genes de resistencia a metales pesados, le permite convertir estos elementos en formas más fáciles de manejar y por consiguiente reducir su toxicidad.

Finalmente, teniendo en cuenta la información procedente de la anotación del genoma de ‘*Ca. Acidibrevibacterium* sp. FLA01’, se puede concluir que se trata de una bacteria que posee una gran versatilidad para tolerar diversas presiones ambientales, como las que se presentan en las bocas de las fumarolas de los Azufres.

‘*Candidatus Aciditerrimonas* sp. FLA02’

El genoma de ‘*Ca. Aciditerrimonas* sp. FLA02’, es el segundo genoma reportado de este género, si bien, en el trabajo realizado por Servín Garcidueñas y Martínez Romero (2015), reportó amplicones relacionados con este género procedentes del mismo sector de FLA. No había sido posible recuperar un genoma parcial o total de un microorganismo cercano a este género hasta la fecha. La primera descripción del género *Aciditerrimonas* fue realizada por Itoh et al. (2011), de muestras extraídas de solfataras. En dicho estudio, se presentó el genoma de *Aciditerrimonas ferrireducens* JCM 15389^T una bacteria ligeramente acidotermófila y aerobia. Estas características se asemejan a las del genoma recuperado en este estudio teniendo en cuenta la información recopilada de la anotación.

‘*Ca. Aciditerrimonas* sp. FLA02’, posee un contenido de GC ligeramente superior al genoma de referencia y un tamaño mucho mayor (**Tabla 12, Figura 9**). A pesar de estas diferencias, su relación filogenética es cercana a *Aciditerrimonas ferrireducens* JCM 15389^T (**Tabla 12, Figura 10**). Uno de los aspectos a destacar en este genoma es la presencia de genes que le permite usar iones de arsénico como aceptor final de electrones. A la fecha se conoce la alta toxicidad de representa el Arsénico en una gran cantidad de organismos vivos, cada vez es mayor el reporte de la incorporación de iones de As en el metabolismo de bacterias (Kumari & Jagadevan, 2016).

Por otra parte, Itoh et al. (2011) reportaron que los microorganismos del género *Aciditerrimonas* se caracterizan por ser termoacidófilos lo cual coincide con las características presentes en FLA. Además, poseen la capacidad de reducir iones férricos. Sin embargo, para el caso de ‘*Ca. Aciditerrimonas* sp. FLA02’ no se detectaron genes involucrados en este proceso (**Tabla 11**). A pesar de esta discrepancia, fundamental en la descripción de este género, es posible que dichas rutas y genes no estén caracterizados en las bases de datos, o se requiere realizar mejoras de calidad del genoma recuperado en este estudio con el fin de lograr anotar genes y reconstruir rutas metabólicas esenciales en estos microorganismos.

Se han reportado bacterias relacionadas con este género en otros lugares con características volcánicas como los tubos volcánicos de las Islas Galápagos en Ecuador (Miller et al., 2016), la presencia de bacterias del género *Aciditerrimonas* en estos lugares es una posible señal de su capacidad de realizar procesos de biomineralización tal como lo mencionan (Miller et al., 2016). Por lo que se ha tratado de explotar las características de este y otros microorganismos en procesos industriales (Cheng et al., 2018; Shiers et al., 2016).

6.2 Manantiales de Maritaro (MM)

Dentro de este metagenoma, fue posible encontrar comunidades que no se hallaron en FLA tales como *Acidobacteria*, *Planctomycetes* y *Bacteroidetes* (**Tabla 20, Figura 14**). A pesar de que también se encontraron los mismos filos presentes en FLA (*Proteobacteria*, *Firmicutes* y *Actinobacteria*), estos varían en su composición. Por ejemplo; en el caso del filo *Proteobacteria*, fue posible tener registros de poblaciones correspondientes a Deltaproteobacterias con integrantes del género *Desulfurella* y Betaproteobacterias de los géneros *Thiomonas* y *Ferritrophicum*, la presencia de estos microorganismos ha sido previamente reportada en otros estudios de ambientes hidrotermales cercanos como es el caso del estudio de los manantiales de Araró (Prieto-Barajas, Alcaraz, et al., 2018), y en otros ambientes hidrotermales presentes en el país como en el estado de Puebla (Castelán-Sánchez et al., 2020). Tanto en MM como en los estudios previamente mencionados, se pudo determinar un claro dominio del filo *Proteobacteria*, seguido de los filos *Firmicutes* y *Actinobacteria* (**Tabla 12, Figura 14**), este patrón fue encontrado también en manantiales hidrotermales presentes en los Himalayas (Najar et al., 2020; Roy et al., 2020) y en los manantiales hidrotermales de Sungaiklah en Malasia (Chan et al., 2015). A pesar de esta gran similitud, tanto los resultados de la reconstrucción e identificación de genes rARN 16S (**Tabla 20, Figura 14**) como la reconstrucción de MAGs vía *binnig* (**Tabla 24**), muestra que la mayoría de las

bacterias presentes en este ambiente se trata de microorganismos no identificados y posiblemente endémicos de esta zona del campo geotérmico de Los Azufres.

6.2.1 Diversidad funcional

La información recopilada a partir de la anotación (**Tablas 15-19**) muestra que dentro de este ambiente se están llevando procesos metabólicos clave, tales como; la fijación y asimilación de nitrógeno. Estas rutas fueron halladas también por (Prieto-Barajas, Alcaraz, et al., 2018) en metagenomas provenientes de tapetes microbianos presentes en manantiales hidrotermales localizados en Araró, cerca de Los Azufres. De otra mano, al igual que en FLA, la presencia de transportadores de fosfatos mostraría que existen microorganismos dentro de la comunidad encargados de la asimilación de fósforo inorgánico, que junto con los diferentes elementos de regulación detectados, evidencian la adaptación ante la escasez de este elemento, vital para el desarrollo de actividades biológicas esenciales, la presencia de estos mecanismos de regulación también fue encontrada en manantiales termales en India (Badhai et al., 2015). En MM también se detectaron genes involucrados en la asimilación de sulfatos, fijación de carbono y mecanismos de incorporación y oxidación de hierro (**Tabla 16**) algunos de estos detectados también por (Prieto-Barajas, Alcaraz, et al., 2018) en Araró. Por lo que es evidente la importancia de este tipo de rutas en ambientes hidrotermales, así como lo han sido en los diferentes ecosistemas de la tierra.

La presencia tanto de mecanismos de síntesis, como de elementos de resistencia a sustancias antimicrobianas (**Tabla 18**), parece mostrar que existen interacciones de competencia entre las comunidades de microorganismos, en las que probablemente, algunos microorganismos han desarrollado mecanismos para hacer frente a la producción de estreptomicinas, vancomicinas, penicilinas y cefalosporinas procedentes de microorganismos competidores, entre los cuales estarían las actinobacterias, conocidas por producir metabolitos secundarios con actividad antimicrobiana (Tulasi

Satyanarayana et al., 2013; Shivilata & Satyanarayana, 2015) entre las que están antibióticos betalactámicos y cefalosporinas (Aharonowitz et al., 1992; Schmitt et al., 2004). La existencia tanto de mecanismos de síntesis como de resistencia a componentes antimicrobianos abre las puertas a la posibilidad de encontrar nuevos antibióticos. En estudios previos (Jardine et al., 2019; Mahajan & Balachandran, 2017) se propone que en los manantiales hidrotermales habitan comunidades microbianas desconocidas en las cuales existe la posibilidad de recuperar nuevos tipos de antibióticos. El caso de MM no es una excepción al presentar comunidades conocidas por su producción de sustancias antimicrobianas como los son el filo *Actinobacteria*, por lo que los microorganismos provenientes de MM podrían tener el potencial para encontrar nuevos tipos de antibióticos.

Al igual que en FLA, en MM fue posible hallar genes involucrados en la resistencia a las altas temperaturas, tolerancia la acidez, una amplia cantidad de transportadores tipo ABC, la formación de biopelículas, y mecanismos de comunicación bacteriana como el *quorum sensing* (**Tabla 17**). A diferencia de ambientes como las fumarolas, en manantiales hidrotermales estas adaptaciones han sido más estudiadas y descritas (Prieto-Barajas, Valencia-Cantero, et al., 2018; Ranawat & Rawat, 2017). Sin embargo, el conocimiento acerca de fenómenos como el *quorum sensing* en comunidades microbianas procedentes de estos ambientes es escaso (Kaur et al., 2018).

Por último, la presencia tanto de transportadores, como de elementos de resistencia a metales pesados (**Tabla 19**) es una respuesta adaptativa a un ambiente en el cual, estos elementos siempre se han encontrado presentes a concentraciones que por lo general pueden afectar a otros organismos vivos, en algunos casos este tipo de elementos proporcionan alternativas metabólicas como es el caso particular del ion arseniato, usado como aceptor de electrones (Ranawat & Rawat, 2018). Y no sólo su presencia, sino que su biodisponibilidad por las condiciones ácidas del medio, vuelven a los metales un factor ambiental de potencial toxicidad a los microorganismos. Consecuentemente, varios de estos mecanismos

de resistencia han sido encontrados en otros ambientes hidrotermales continentales (Badhai et al., 2015; Jardine et al., 2019), así como mecanismos de empleo como aceptores y donares de electrones en la producción de energía.

6.2.2 Genomas recuperados

‘*Candidatus Acidobacterium* sp. MM01’

Este genoma, contrario a los demás recuperados en este estudio, es el que posee más reportes en GenBank con un total de 3 registros correspondientes al género *Acidobacterium*. Adicionalmente se encuentran reportados 12 registros de ensamblajes de este género sin una identificación clara. Sin embargo, no hay registros previos de microorganismos de este género en los estudios previamente realizados en Los Azufres ni en otros estudios realizados en otros ambientes geotermales de la zona.

El genoma de ‘*Ca. Acidobacterium* sp. MM01’ posee un contenido de GC ligeramente superior al del genoma de referencia (**Tabla 21**). Adicionalmente las métricas de identidad de ANI e *in silico* DDH muestran una alta discrepancia entre ‘*Ca. Acidobacterium* sp. MM01’ y *A. capsulatum*, por lo que podría tratarse de una especie nueva de este género, teniendo en cuenta que se requiere un valor de 95% de ANI y un valor mayor del 70% de *in silico* DDH para considerar dos genomas como procedentes de la misma especie (Chun et al., 2018). A pesar de estas discrepancias, ‘*Ca. Acidobacterium* sp. MM01’ y *A. capsulatum* presentan una gran cercanía a nivel filogenómico (**Figura 16**) lo que quizás podría indicar que se trata de un microorganismo del mismo género.

Los reportes de este género son mayores y por lo tanto más antiguos, el primero de ellos es derivado de muestras de aguas residuales de minas en Japón (Kishimoto et al., 1991). Los organismos de este género se caracterizan por ser quimiorganotrofos, lo cual coincide con estos estudios teniendo en cuenta como respaldo la anotación funcional de ‘*Ca. Acidobacterium* sp. MM01’ (**Tabla 22**). La gran mayoría de reportes se ha realizado de drenajes ácidos de minas (Carlier et al., 2020; Hedrich & Schippers, 2020) y

lugares extremadamente ácidos como Río Tinto (Sánchez-Andrea et al., 2011), donde los genes de resistencia metales pesados como los encontrados en el genoma de ‘*Ca. Acidobacterium* sp. MM01’, les confieren la capacidad de sobrevivir en estos ambientes. Algunos integrantes de este género han sido hallados también en campos geotermales (Akimov et al., 2013), por ejemplo *Acidobacterium ailaau* (Myers & King, 2016). Esto no es de extrañar, puesto que el grupo de las acidobacterias es uno de los más diversos y extendidos, por tanto posee integrantes de su comunidad en la gran mayoría de ambientes terrestres, en especial aquellos con valores de pHs ácidos (Kalam et al., 2020), como es el caso del campo geotérmico de Los Azufres. Por lo que se ha sugerido su uso en procesos de restauración de suelos impactados por la actividad minera (Roth et al., 2019).

‘*Candidatus Metallibacterium* sp. MM02’

Este es el tercer reporte de un genoma perteneciente al género *Metallibacterium* en las bases de datos de GenBank. Los dos genomas ya reportados pertenecen a *Metallibacterium scheffleri*, genoma de referencia usado en este estudio. A pesar de que ‘*Ca. Metallibacterium* sp. MM02’ posee una estrecha cercanía con *M. scheffleri* (**Figura 18**), las diferencias de contenido de GC, tamaños de genoma junto con los resultados de las métricas de ANI e *in silico* DDH, indican que puede tratarse de un microorganismo de una especie diferente, teniendo en cuenta los valores previamente mencionados (95% ANI, 70% *in silico* DDH) (Chun et al., 2018) (**Tabla 23**).

El primer reporte de un microorganismo de género *Metallibacterium* fue realizado por Ziegler et al. (2013) de una muestra proveniente de una mina de pirita localizada en Alemania. Después de la descripción de esta bacteria, se han encontrado registros de microorganismos relacionados con este género en diversas minas, siendo unos de los más abundantes en estos ambientes (Bartsch et al., 2017; Ettamimi et al., 2019; Kadnikov et al., 2019; Lukhele et al., 2019) e incluso, se les ha encontrado en ambientes altamente ácidos como Río Tinto (Bartsch et al., 2017) en cavernas ricas en ácido sulfúrico

(D'Angeli et al., 2019) y en ambientes hidrotermales en Rumania (Coman et al., 2013). Por lo que no ha sido extraño encontrar microorganismos de este grupo taxonómico en MM.

Teniendo en cuenta estas consideraciones, es posible observar que los microorganismos pertenecientes al género *Metallibacterium* son altamente versátiles (Bartsch et al., 2017) y 'Ca *Metallibacterium* sp. MM02' no es la excepción. Los resultados de la anotación funcional muestran que se trata de un microorganismo que posiblemente posee la capacidad de incorporar Fe y reducir Fe, esto es consistente con la descripción inicial que realizó (Ziegler et al., 2013) de *M. scheffleri* DKE6. Sin embargo, (Bartsch et al., 2017) no encontró evidencias de dicha capacidad tras realizar una anotación taxonómica de este microorganismo con la misma plataforma usada en este estudio. La anotación de 'Ca. *Metallibacterium* sp. MM02' también muestra la capacidad de fijar CO₂, algo que tampoco es claro en *M. scheffleri*, al cual se atribuye su principal fuente de energía a partir de la degradación de proteínas, aunque también fuesen detectados dichos genes por medio de procesos de anotación (Bartsch et al., 2017).

El hallazgo de genes ribosomales relacionados con el género *Metallibacterium* en diversos ambientes de actividad minera, lo han puesto como un probable candidato para procesos de biorremediación de aguas residuales de actividades industriales y mineras (H. Chen et al., 2020; Liang et al., 2020; Liu et al., 2019; Sun et al., 2019; Xia et al., 2019; Yang et al., 2020). El genoma de 'Ca. *Metallibacterium* sp. MM02', es por tanto una contribución adicional que puede ayudar a futuras investigaciones relacionadas con este grupo de microorganismos y por tanto incorporarlo a actividades aplicadas a la restauración de ambientes degradados o perturbados por la actividad minera, especialmente la relacionada con la extracción de metales.

6.3 Consideraciones derivadas de los metagenomas de FLA y MM

Los datos presentados en este documento revelan que no solamente ha sido posible hallar nuevos registros de bacterias en la zona de estudio, especialmente en MM. Adicionalmente, es evidente la variación de la composición taxonómica entre los dos sitios de muestreo, esto debido principalmente al tipo de hábitat en el cual se han desarrollado los sedimentos microbianos. Por lo que fue posible ver un patrón de ciertas poblaciones bacterianas que parecen estar presentes en ambientes hidrotermales como es el caso de MM y las cuales no están en sedimentos circundantes de fumarolas como sucede con FLA. Por ejemplo, otros estudios realizados en respiraderos de vapor volcánico muestran un patrón taxonómico similar al presentado en FLA, debido a la alta de dominancia de grupos taxonómicos como *Actinobacteria* y *Proteobacteria* (Benson et al., 2011; Fagorzi et al., 2019; Souza-Brito et al., 2019). En contraste, las poblaciones bacterianas presentes en MM son similares a las presentes en otros ambientes hidrotermales continentales (Delgado-Serrano et al., 2014; Prieto-Barajas, Alcaraz, et al., 2018) Debido a una clara dominancia de grupos tales como; *Proteobacteria*, *Firmicutes* y *Actinobacteria*, incluso, al realizar las comparaciones en bases de datos en GenBank, fue posible observar que algunos de los hits que tuvieron las secuencias de genes rARN 16S reconstruidas de MM poseen una alta similitud con secuencias ribosomales encontradas en Rio Tinto, España (Amaral-Zettler et al., 2011; García-Moyano et al., 2007).

Si bien, existen discrepancias a nivel taxonómico, a nivel funcional se pueden observar importantes coincidencias en adaptaciones como la resistencia a ciertas condiciones de estrés tales como el choque térmico, la tolerancia/resistencia la acidez, la producción de compuestos antimicrobianos y la resistencia a metales pesados. Este tipo de fenómenos suelen encontrarse también en otros ambientes similares localizados tanto en el Eje Volcánico Transmexicano (Prieto-Barajas, Alcaraz, et al., 2018; Souza-Brito et al., 2019) como en otros ambientes geotérmicos del mundo (Chan et al., 2015; Najar et al., 2020). Sin

embargo, hay evidentes discrepancias funcionales, especialmente en algunos ciclos biogeoquímicos como el nitrógeno, donde en MM se detectaron genes relacionados con la fijación de nitrógeno (*nif*) (**Tabla 16**), los cuales no fueron detectados en FLA (**Tabla 5**), a pesar de que se detectó la presencia de integrantes del filo *Nitrospirae*, *Firmicutes* y *Proteobacteria*, los cuales tienen la capacidad de fijar nitrógeno atmosférico, pero son mejor conocidos por su capacidad nitrificante, por lo que no se deben descartar completamente que se estén llevando a cabo estos procesos en FLA. Por otra parte, tanto en FLA como en MM, fue posible detectar la presencia de transportadores de fosfato, y elementos de regulación relacionados con la escasez de Fósforo. Lo que estaría indicando la posibilidad de una presión derivada de la falta de este elemento en ambos ambientes. Aunque se requiere de análisis fisicoquímicos en las zonas de muestreo para determinar las concentraciones reales de fósforo, estudios hechos en solfataras ubicadas de Los Azufres (Servín Garcidueñas & Martínez Romero, 2015), mostraron que en efecto, las concentraciones de este elemento en el lugar son bajas (<1.13 mg/L).

Ahora bien, al observar la distribución entre los diferentes genes ortólogos y su distribución taxonómica tanto para FLA como para MM (**Figuras 5, Figura 13**), se puede ver que solo determinados grupos taxonómicos tendrían la capacidad de realizar gran parte de las funciones necesarias para la supervivencia en FLA y en MM. Para el caso de FLA, estos microorganismos serían *Euryarchaeota* y en menor medida *Nitrospirae*, *Proteobacteria*, *Firmicutes* y *Actinobacteria*. En MM serían *Proteobacteria* y *Acidobacteria*. Este patrón es consistente con lo reportado por Badhai et al. (2015) y Chan et al. (2015). Adicionalmente Iacono et al. (2020) encontró un fenómeno similar, aunque con un dominio de arqueas del filo *Crenarchaeota* y bacterias del filo *Aquificae*. Sin embargo, los grupos taxonómicos con una menor abundancia podrían estar ejerciendo funciones esenciales dentro de las comunidades microbianas (Kurm et al., 2017). Por ejemplo, responder ante condiciones de estrés o cambios de estacionalidad (Sánchez-Pérez et al., 2020), fenómenos que actualmente ocurren en Los Azufres.

Teniendo en cuenta la información presentada previamente, es importante resaltar que si bien, existe una discrepancia a nivel taxonómico de las comunidades bacterianas que habitan en FLA con respecto a las que están establecidas en MM, estas comunidades han logrado resolver su supervivencia en estos ambientes por medio de adaptaciones similares, lo que podría indicar que se trata de adaptaciones que se han conservado a lo largo de la evolución de los diferentes ambientes localizados en el campo geotérmico de Los Azufres o que existe aún un alto intercambio genético entre los microorganismos allí presentes. Se podría sugerir, que ambos tapetes microbianos analizados tienen diversidad taxonómica diferente, pero comparten aspectos de su diversidad funcional.

Finalmente, la información taxonómica y funcional recopilada en este estudio se ha podido observar que dentro de las comunidades microbianas presentes en Los Azufres se han encontrado genes involucrados en procesos con características primitivas tales como el ciclo de los ácidos tricarbónicos reductivo (rTCA) y el metabolismo del azufre (**Tabla 5, Tabla 16**). Damer y Deamer (2020) han propuesto que los ambientes hidrotermales continentales son un punto donde posiblemente pudieron llevarse a cabo las primeras manifestaciones químicas que permitieron la aparición de seres vivos en la Tierra. Por otra parte, Kitadai et al. (2019), Muchowska et al. (2017) y Hickman-Lewis et al. (2020) han propuesto que algunos de los metales pesados considerados como tóxicos en la actualidad tales como; cobalto, cromo, cobre, níquel y zinc, que junto con otros metales como el hierro pudieron servir como intermediarios de reacciones de síntesis de compuestos orgánicos y hubiesen podido propiciar la aparición de los primeros ciclos vitales como el ciclo de los ácidos tricarbónicos reductivo (rTCA). Si esto fuese cierto, surge la pregunta de si la actual resistencia a ciertos metales pesados como los encontrados en este trabajo (**Tabla 6, Tabla 19**), es en realidad un vestigio del papel de estos elementos en un metabolismo primitivo menos desarrollado, lo cual explicaría el origen de estas adaptaciones las cuales son comunes

actualmente en ambientes como el campo geotérmico de los azufres, como otros ambientes donde existe una alta concentración de estos metales.

7. CONCLUSIONES

A partir del estudio aquí presentado, es posible afirmar que dentro de Los Azufres encontramos una diversidad bacteriana poco conocida la cual varía dependiendo del tipo de manifestación termal como lo son las fumarolas (FLA) y los manantiales termales (MM). La diversidad previamente conocida de manera parcial en el primer punto de muestreo se pudo estudiar con mayor profundidad gracias al uso de herramientas de secuenciación masiva y de análisis bioinformáticos. Este estudio Permitió identificar nuevos integrantes de las comunidades presentes en FLA y en MM, adicionalmente arrojó patrones en los cuales es evidente la dominancia de ciertas poblaciones de procariotas presentes en manantiales hidrotermales, los cuales no se presentan en sedimentos que habitan las bocas de las fumarolas. Los ensamblajes metagenómicos también permitieron la reconstrucción de genomas vía *binnig* y por medio de genómica comparativa de los cuales algunos aun poseen una identificación desconocida o poco clara. Finalmente, el análisis funcional permitió realizar una reconstrucción del metabolismo tanto de las comunidades presentes en el metagenoma como de los genomas recuperados por medio de genómica comparativa. La información proporcionada de este análisis muestra una relación clara con la adaptación a condiciones ambientales adversas tales como el choque térmico y la resistencia a los metales pesados. Además de proporcionar información de posibles interacciones de competencia y/o de defensa por medio de la síntesis o producción de mecanismos de resistencia de componentes antimicrobianos. Este conjunto de adaptaciones muestra una relación estrecha entre las comunidades bacterianas y los hábitats presentes dentro de Los Azufres. Independientemente de que FLA y MM posean una composición taxonómica distinta, su diversidad funcional similar es el indicio de un origen ancestral común de las comunidades bacterianas presentes en el campo geotérmico de Los Azufres al ser esenciales para su

supervivencia. O bien, es una evidencia de las presiones selectivas que han sufrido dichas comunidades en este ambiente.

8. PERSPECTIVAS

El presente estudio ha mostrado la existencia de una diversidad bacteriana aún desconocida. Por lo que es recomendable continuar trabajando en procesos de curación de los genomas que se obtuvieron vía *Binning*, Adicionalmente, sería importante realizar la secuenciación de genomas de bacterias en cultivo con el fin de facilitar la identificación de bacterias procedentes tanto de FLA como de MM, por medio del uso de tecnologías de nueva generación como Nanopore o PacBio.

En cuanto a nivel funcional, se espera que los hallazgos presentados aquí se puedan complementar por medio de análisis fisicoquímicos de las muestras trabajadas, y si es posible de análisis transcriptómicos y proteómicos. Lo anterior permitirá profundizar en las funciones que los microorganismos están realizando en este ambiente, en especial en aquellas que aún son desconocidas.

9. REFERENCIAS BIBLIOGRÁFICAS

- Aharonowitz, Y., Cohen, G., & Martin, J. F. (1992). PENICILLIN AND CEPHALOSPORIN BIOSYNTHETIC GENES: Structure, Organization, Regulation, and Evolution. *Annual Review of Microbiology*, 46(1), 461–495. <https://doi.org/10.1146/annurev.mi.46.100192.002333>
- Akimov, V. N., Podosokorskaya, O. A., Shlyapnikov, M. G., & Gal'chenko, V. F. (2013). Dominant phylotypes in the 16S rRNA gene clone libraries from bacterial mats of the Uzon caldera (Kamchatka, Russia) hydrothermal springs. *Microbiology (Russian Federation)*, 82(6), 721–727. <https://doi.org/10.1134/S0026261713060027>
- Alfaro-Cuevas-Villanueva, R., Cortes-Martinez, R., García-Díaz, J. J., Galvan-Martinez, R., & Torres-Sanchez, R. (2006). Microbiologically influenced corrosion of steels by thermophilic and mesophilic bacteria. *Materials and Corrosion*, 57(7), 543–548. <https://doi.org/10.1002/maco.200503948>
- Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11(11), 1144–1146. <https://doi.org/10.1038/nmeth.3103>
- Amaral-Zettler, L. A., Zettler, E. R., Theroux, S. M., Palacios, C., Aguilera, A., & Amils, R. (2011). Microbial community structure across the tree of life in the extreme Río Tinto. *ISME Journal*, 5(1), 42–50. <https://doi.org/10.1038/ismej.2010.101>

- Antranikian, G., Suleiman, M., Schäfers, C., Adams, M. W. W., Bartolucci, S., Blamey, J. M., Birkeland, N. K., Bonch-Osmolovskaya, E., da Costa, M. S., Cowan, D., Danson, M., Forterre, P., Kelly, R., Ishino, Y., Littlechild, J., Moracci, M., Noll, K., Oshima, T., Robb, F., ... Stetter, K. O. (2017). Diversity of bacteria and archaea from two shallow marine hydrothermal vents from Vulcano Island. *Extremophiles*, 21(4), 733–742. <https://doi.org/10.1007/s00792-017-0938-y>
- Arriaga Suarez, M. C., Tello, M. del R., & Samaniego V, F. (2000). Geochemical evolution of the Los Azufres, Mexico, geothermal reservoir. *Proceedings World Geothermal Congress*, 2227–2233.
- Aziz, R. K., Bartels, D., Best, 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, 1–15. <https://doi.org/10.1186/1471-2164-9-75>
- Badger, M. R., & Bek, E. J. (2008). Multiple Rubisco forms in proteobacteria: Their functional significance in relation to CO₂ acquisition by the CBB cycle. *Journal of Experimental Botany*, 59(7), 1525–1541. <https://doi.org/10.1093/jxb/erm297>
- Badhai, J., Ghosh, T. S., & Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Frontiers in Microbiology*, 6(OCT), 1–15. <https://doi.org/10.3389/fmicb.2015.01166>
- Baker, B. J., & Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiology Ecology*, 44(2), 139–152. [https://doi.org/10.1016/S0168-6496\(03\)00028-X](https://doi.org/10.1016/S0168-6496(03)00028-X)
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bartsch, S., Gensch, A., Stephan, S., Doetsch, A., & Gescher, J. (2017). Metallibacterium scheffleri: genomic data reveal a versatile metabolism. *FEMS Microbiology Ecology*, 93(3), 1–10. <https://doi.org/10.1093/femsec/fix011>
- Benson, C. A., Bizzoco, R. W., Lipson, D. A., & Kelley, S. T. (2011). Microbial diversity in nonsulfur, sulfur and iron geothermal steam vents. *FEMS Microbiology Ecology*, 76(1), 74–88. <https://doi.org/10.1111/j.1574-6941.2011.01047.x>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5. <https://doi.org/10.1038/srep08365>
- Brito, E. M. S., Villegas-Negrete, N., Sotelo-González, I. A., Caretta, C. A., Goñi-Urriza, M., Gassie, C., Hakil, F., Colin, Y., Duran, R., Gutiérrez-Corona, F., Piñón-Castillo, H. A., Cuevas-Rodríguez,

- G., Malm, O., Torres, J. P. M., Fahy, A., Reyna-López, G. E., & Guyoneaud, R. (2014). Microbial diversity in Los Azufres geothermal field (Michoacán, Mexico) and isolation of representative sulfate and sulfur reducers. *Extremophiles*, *18*(2), 385–398. <https://doi.org/10.1007/s00792-013-0624-7>
- Brown, J., Pirrung, M., & Mccue, L. A. (2017). FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, *33*(19), 3137–3139. <https://doi.org/10.1093/bioinformatics/btx373>
- Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Carlier, J. D., Ettamimi, S., Cox, C. J., Hammani, K., Ghazal, H., & Costa, M. C. (2020). Prokaryotic diversity in stream sediments affected by acid mine drainage. *Extremophiles*, *24*(6), 809–819. <https://doi.org/10.1007/s00792-020-01196-8>
- Carrasco-Núñez, G., López-Martínez, M., Hernández, J., & Vargas, V. (2017). Subsurface stratigraphy and its correlation with the surficial geology at Los Humeros geothermal field, eastern Trans-Mexican Volcanic Belt. *Geothermics*, *67*(May), 1–17. <https://doi.org/10.1016/j.geothermics.2017.01.001>
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2011). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, *28*(4), 464–469. <https://doi.org/10.1093/bioinformatics/btr703>
- Castelán-Sánchez, H. G., Meza-Rodríguez, P. M., Carrillo, E., Ríos-Vázquez, D. I., Liñan-Torres, A., Batista-García, R. A., Pérez-Rueda, E., Rojas-Ruíz, N. E., & Dávila-Ramos, S. (2020). The microbial composition in circumneutral thermal springs from chignahuapan, puebla, mexico reveals the presence of particular sulfur-oxidizing bacterial and viral communities. *Microorganisms*, *8*(11), 1–26. <https://doi.org/10.3390/microorganisms8111677>
- Castorena, G., Mugica, V., Le Borgne, S., Acuña, M. E., Bustos-Jaimes, I., & Aburto, J. (2006). Carbazole biodegradation in gas oil/water biphasic media by a new isolated bacterium Burkholderia sp. strain IMP5GC. *Journal of Applied Microbiology*, *100*(4), 739–745. <https://doi.org/10.1111/j.1365-2672.2005.02799.x>
- Cavalazzi, B., Glamoclija, M., Brack, A., Westall, F., Orosei, R., & Cady, S. L. (2018). Astrobiology, the Emergence of Life, and Planetary Exploration. In A. P. Rossi & S. van Gasselt (Eds.), *Planetary Geology* (pp. 347–367). Springer International Publishing. https://doi.org/10.1007/978-3-319-65179-8_14
- Chan, C. S., Chan, K. G., Tay, Y. L., Chua, Y. H., & Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Frontiers in Microbiology*, *6*(MAR), 1–15. <https://doi.org/10.3389/fmicb.2015.00177>
- Chen, G.-Q., & Jiang, X.-R. (2018). Next generation industrial biotechnology based on extremophilic bacteria. *Current Opinion in Biotechnology*, *50*, 94–100.
- Chen, H., Xiao, T., Ning, Z., Li, Q., Xiao, E., Liu, Y., Xiao, Q., Lan, X., Ma, L., & Lu, F. (2020). In-situ remediation of acid mine drainage from abandoned coal mine by filed pilot-scale passive treatment system: Performance and response of microbial communities to low pH and elevated Fe.

Bioresource Technology, 317(June), 123985. <https://doi.org/10.1016/j.biortech.2020.123985>

- Chen, I. M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Andersen, E., Huntemann, M., Varghese, N., Hadjithomas, M., Tennessen, K., Nielsen, T., Ivanova, N. N., & Kyrpides, N. C. (2017). IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Research*, 45(D1), D507–D516. <https://doi.org/10.1093/nar/gkw929>
- Chen, L. X., Méndez-García, C., Dombrowski, N., Servín-Garcidueñas, L. E., Eloë-Fadrosh, E. A., Fang, B. Z., Luo, Z. H., Tan, S., Zhi, X. Y., Hua, Z. S., Martínez-Romero, E., Woyke, T., Huang, L. N., Sánchez, J., Peláez, A. I., Ferrer, M., Baker, B. J., & Shu, W. S. (2018). Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME Journal*, 12(3), 756–775. <https://doi.org/10.1038/s41396-017-0002-z>
- Cheng, H., Asakura, Y., Kanda, K., Fukui, R., Kawano, Y., Okugawa, Y., Tashiro, Y., & Sakai, K. (2018). Dynamic bacterial community changes in the autothermal thermophilic aerobic digestion process with cell lysis activities, shaking and temperature increase. *Journal of Bioscience and Bioengineering*, 126(2), 196–204. <https://doi.org/https://doi.org/10.1016/j.jbiosc.2018.02.012>
- Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahall, D. R., da Costa, M. S., Rooney, A. P., Yi, H., Xu, X. W., De Meyer, S., & Trujillo, M. E. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 68(1), 461–466. <https://doi.org/10.1099/ijsem.0.002516>
- Cockell, C. S., Harrison, J. P., Stevens, A. H., Payler, S. J., Hughes, S. S., Kobs Nawotniak, S. E., Brady, A. L., Elphic, R. C., Haberle, C. W., Sehlke, A., Beaton, K. H., Abercromby, A. F. J., Schwendner, P., Wadsworth, J., Landenmark, H., Cane, R., Dickinson, A. W., Nicholson, N., Perera, L., & Lim, D. S. S. (2019). A Low-Diversity Microbiota Inhabits Extreme Terrestrial Basaltic Terrains and Their Fumaroles: Implications for the Exploration of Mars. *Astrobiology*, 19(3), 284–299. <https://doi.org/10.1089/ast.2018.1870>
- Coman, C., Drugă, B., Hegedus, A., Sicora, C., & Dragoș, N. (2013). Archaeal and bacterial diversity in two hot spring microbial mats from a geothermal region in Romania. *Extremophiles*, 17(3), 523–534. <https://doi.org/10.1007/s00792-013-0537-5>
- D'Angeli, I. M., Ghezzi, D., Leuko, S., Firrincieli, A., Parise, M., Fiorucci, A., Vigna, B., Adesso, R., Baldantoni, D., Carbone, C., Miller, A. Z., Jurado, V., Saiz-Jimenez, C., De Waele, J., & Cappelletti, M. (2019). Geomicrobiology of a seawater-influenced active sulfuric acid cave. In *PLoS ONE* (Vol. 14, Issue 8). <https://doi.org/10.1371/journal.pone.0220706>
- Damer, B., & Deamer, D. (2020). The hot spring hypothesis for an origin of life. *Astrobiology*, 20(4), 429–452. <https://doi.org/10.1089/ast.2019.2045>
- Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14(7), 1394–1403.
- Delgado-Serrano, L., López, G., Bohorquez, L. C., Bustos, J. R., Rubiano, C., Osorio-Forero, C., Junca, H., Baena, S., & Zambrano, M. M. (2014). Neotropical Andes hot springs harbor diverse and distinct planktonic microbial communities. *FEMS Microbiology Ecology*, 89(1), 56–66. <https://doi.org/10.1111/1574-6941.12333>

- Escobar-Zepeda, A., De León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6(DEC), 1–15. <https://doi.org/10.3389/fgene.2015.00348>
- Ettamimi, S., Carlier, J. D., Cox, C. J., Elamine, Y., Hammani, K., Ghazal, H., & Costa, M. C. (2019). A meta-taxonomic investigation of the prokaryotic diversity of water bodies impacted by acid mine drainage from the São Domingos mine in southern Portugal. *Extremophiles*, 23(6), 821–834. <https://doi.org/10.1007/s00792-019-01136-1>
- Fagorzi, C., Duca, S. Del, Venturi, S., Chiellini, C., Bacci, G., Fani, R., & Tassi, F. (2019). Bacterial communities from extreme environments: Vulcano Island. *Diversity*, 11(8), 1–17. <https://doi.org/10.3390/d11080140>
- Frankenberg-Dinkel, N. (2004). Forum Review Bacterial Heme Oxygenases. *Antioxidants and Redox Signaling*, 6(5), 825–834.
- García-Moyano, A., González-Toril, E., Aguilera, A., & Amils, R. (2007). Prokaryotic community composition and ecology of floating macroscopic filaments from an extreme acidic environment, Río Tinto (SW, Spain). *Systematic and Applied Microbiology*, 30(8), 601–614. <https://doi.org/10.1016/j.syapm.2007.08.002>
- González-Guzmán, R., Inguaggiato, C., Peiffer, L., Weber, B., & Kretzschmar, T. (2019). Fault-controlled geothermal fluids of the northern Trans-Mexican Volcanic Belt: A geochemical and isotopic study of the Los Geysers field (Valley of Queretaro, Mexico). *Journal of Volcanology and Geothermal Research*, 388(November). <https://doi.org/10.1016/j.jvolgeores.2019.106681>
- Guazzaroni, M. E., Morgante, V., Mirete, S., & González-Pastor, J. E. (2013). Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environmental Microbiology*, 15(4), 1088–1102. <https://doi.org/10.1111/1462-2920.12021>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Handelsman, J. (2005). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 69(1), 195–195. <https://doi.org/10.1128/mmbr.69.1.195.2005>
- Hedrich, S., & Schippers, A. (2020). Distribution of acidophilic microorganisms in natural and man-made acidic environments. *Current Issues in Molecular Biology*, 40, 25–48. <https://doi.org/10.21775/CIMB.040.025>
- Hickman-Lewis, K., Cavalazzi, B., Sorieul, S., Gautret, P., Foucher, F., Whitehouse, M. J., Jeon, H., Georgelin, T., Cockell, C. S., & Westall, F. (2020). Metallomics in deep time and the influence of ocean chemistry on the metabolic landscapes of Earth's earliest ecosystems. *Scientific Reports*, 10(1), 1–16. <https://doi.org/10.1038/s41598-020-61774-w>
- Huber, R., Huber, H., & Stetter, K. O. (2000). Towards the ecology of hyperthermophiles: Biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiology Reviews*, 24(5), 615–623. [https://doi.org/10.1016/S0168-6445\(00\)00049-8](https://doi.org/10.1016/S0168-6445(00)00049-8)
- Hügler, M., Gärtner, A., & Imhoff, J. F. (2010). Functional genes as markers for sulfur cycling and CO₂

fixation in microbial communities of hydrothermal vents of the Logatchev field. *FEMS Microbiology Ecology*, 73(3), 526–537. <https://doi.org/10.1111/j.1574-6941.2010.00919.x>

- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- Iacono, R., Cobucci-Ponzano, B., de Lise, F., Curci, N., Maurelli, L., Moracci, M., & Strazzulli, A. (2020). Spatial Metagenomics of Three Geothermal Sites in Pisciarelli Hot Spring Focusing on the Biochemical Resources of the Microbial Consortia. *Molecules*, 25(17). <https://doi.org/10.3390/molecules25174023>
- Ilbert, M., & Bonnefoy, V. (2013). Insight into the evolution of the iron oxidation pathways. *Biochimica et Biophysica Acta - Bioenergetics*, 1827(2), 161–175. <https://doi.org/10.1016/j.bbabi.2012.10.001>
- Iqbalsyah, T. M., Nisak, C., Sartika, F., Saidi, N., & Febriani, F. (2019). Identification of Antibiotic in Ethyl Acetate Fraction Produced by A Local Isolate PLS 80 Isolated from Shallow Sea Fumaroles. *IOP Conference Series: Earth and Environmental Science*, 276(1). <https://doi.org/10.1088/1755-1315/276/1/012006>
- Itoh, T., Yamanoi, K., Kudo, T., Ohkuma, M., & Takashina, T. (2011). Aciditerrimonas ferrireducens gen. nov., sp. nov., an iron-reducing thermoacidophilic actinobacterium isolated from a solfataric field. *International Journal of Systematic and Evolutionary Microbiology*, 61(6), 1281–1285. <https://doi.org/10.1099/ijs.0.023044-0>
- Jácome-Paz, M. P. (2019). Two new geothermal prospects in the Mexican Volcanic Belt: La Escalera and Agua Caliente – Tzitzio geothermal springs, Michoacán, México. *Geothermics*, 80(March 2018), 44–55. <https://doi.org/10.1016/j.geothermics.2019.02.004>
- Jardine, J., Mavumengwana, V., & Ubomba-Jaswa, E. (2019). Antibiotic resistance and heavy metal tolerance in cultured bacteria from hot springs as indicators of environmental intrinsic resistance and tolerance levels. *Environmental Pollution*, 249, 696–702. <https://doi.org/10.1016/j.envpol.2019.03.059>
- Kadnikov, V. V., Gruzdev, E. V., Ivasenko, D. A., Beletsky, A. V., Mardanov, A. V., Danilova, E. V., Karnachuk, O. V., & Ravin, N. V. (2019). Selection of a Microbial Community in the Course of Formation of Acid Mine Drainage. *Microbiology (Russian Federation)*, 88(3), 292–299. <https://doi.org/10.1134/S0026261719030056>
- Kalam, S., Basu, A., Ahmad, I., Sayyed, R. Z., El-Enshasy, H. A., Dailin, D. J., & Suriani, N. L. (2020). Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. *Frontiers in Microbiology*, 11(October). <https://doi.org/10.3389/fmicb.2020.580024>
- Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 2015(8), 1–15. <https://doi.org/10.7717/peerj.1165>
- Kaur, A., Capalash, N., & Sharma, P. (2018). Quorum sensing in thermophiles: Prevalence of autoinducer-2 system. *BMC Microbiology*, 18(1), 1–16. <https://doi.org/10.1186/s12866-018-1204-x>
- Kawarabayasi, Y. (2013). Acido- and Thermophilic Microorganisms: Their Features, and the Identification of Novel Enzymes or Pathways. In J. Seckbach, A. Oren, & H. Stan-Lotter (Eds.),

Polyextremophiles: Life Under Multiple Forms of Stress (pp. 297–313). Springer Netherlands. https://doi.org/10.1007/978-94-007-6488-0_12

- Kishimoto, N., Kosako, Y., & Tano, T. (1991). *Acidobacterium capsulatum* gen. nov., sp. nov.: An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Current Microbiology*, 22(1), 1–7. <https://doi.org/10.1007/BF02106205>
- Kitadai, N., Nakamura, R., Yamamoto, M., Takai, K., Yoshida, N., & Oono, Y. (2019). Metals likely promoted protometabolism in early ocean alkaline hydrothermal systems. *Science Advances*, 5(6), 1–10. <https://doi.org/10.1126/sciadv.aav7848>
- Klatt, C. G., Wood, J. M., Rusch, D. B., Bateson, M. M., Hamamura, N., Heidelberg, J. F., Grossman, A. R., Bhaya, D., Cohan, F. M., Köhl, M., Bryant, D. A., & Ward, D. M. (2011). Community ecology of hot spring cyanobacterial mats: Predominant populations and their functional potential. *ISME Journal*, 5(8), 1262–1278. <https://doi.org/10.1038/ismej.2011.73>
- 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. <https://doi.org/10.1093/molbev/msy096>
- Kumar, V., & Satyanarayana, T. (2013). Thermoalkaliphilic Microbes. In J. Seckbach, A. Oren, & H. Stan-Lotter (Eds.), *Polyextremophiles: Life Under Multiple Forms of Stress* (pp. 271–296). Springer Netherlands. https://doi.org/10.1007/978-94-007-6488-0_11
- Kumari, N., & Jagadevan, S. (2016). Genetic identification of arsenate reductase and arsenite oxidase in redox transformations carried out by arsenic metabolising prokaryotes – A comprehensive review. *Chemosphere*, 163, 400–412. <https://doi.org/10.1016/j.chemosphere.2016.08.044>
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., & Hugenholtz, P. (2008). A Bioinformatician's Guide to Metagenomics. *Microbiology and Molecular Biology Reviews*, 72(4), 557–578. <https://doi.org/10.1128/mubr.00009-08>
- Kurm, V., Van Der Putten, W. H., De Boer, W., Naus-Wiezer, S., & Gera Hol, W. H. (2017). Low abundant soil bacteria can be metabolically versatile and fast growing. *Ecology*, 98(2), 555–564. <https://doi.org/10.1002/ecy.1670>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3). <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lau, M. C. Y., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S., van Heerden, E., Kieft, T. L., & Onstott, T. C. (2014). of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*, 5(OCT), 1–17. <https://doi.org/10.3389/fmicb.2014.00531>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liang, Z., Wang, J., Zhang, Y., Han, C., Ma, S., Chen, J., Li, G., & An, T. (2020). Removal of volatile organic compounds (VOCs) emitted from a textile dyeing wastewater treatment plant and the

- attenuation of respiratory health risks using a pilot-scale biofilter. *Journal of Cleaner Production*, 253, 120019. <https://doi.org/10.1016/j.jclepro.2020.120019>
- Liu, J., Yin, M., Zhang, W., Tsang, D. C. W., Wei, X., Zhou, Y., Xiao, T., Wang, J., Dong, X., Sun, Y., Chen, Y., Li, H., & Hou, L. (2019). Response of microbial communities and interactions to thallium in contaminated sediments near a pyrite mining area. *Environmental Pollution*, 248, 916–928. <https://doi.org/10.1016/j.envpol.2019.02.089>
- Lukhele, T., Selvarajan, R., Nyoni, H., Mamba, B. B., & Msagati, T. A. M. (2019). Diversity and functional profile of bacterial communities at Lancaster acid mine drainage dam, South Africa as revealed by 16S rRNA gene high-throughput sequencing analysis. *Extremophiles*, 23(6), 719–734. <https://doi.org/10.1007/s00792-019-01130-7>
- Mahajan, G. B., & Balachandran, L. (2017). Sources of antibiotics: Hot springs. In *Biochemical Pharmacology* (Vol. 134). <https://doi.org/10.1016/j.bcp.2016.11.021>
- Marín-Paredes, R., & Servín-Garcidueñas, L. E. (2020). *Estudio metagenómico de arqueas termoacidófilas de sedimentos del campo geotérmico de los azufres*. [Tesis de maestría, Universidad Nacional Autónoma de México]. <http://132.248.9.195/ptd2020/agosto/0802640/Index.html>
- Medrano-Santillana, M., Souza-Brito, E. M., Duran, R., Gutierrez-Corona, F., & Reyna-López, G. E. (2017). Bacterial diversity in fumarole environments of the Parícutín volcano, Michoacán (Mexico). *Extremophiles*, 21(3), 499–511. <https://doi.org/10.1007/s00792-017-0920-8>
- Mehta, R., Singhal, P., Singh, H., Damle, D., & Sharma, A. K. (2016). Insight into thermophiles and their wide-spectrum applications. *3 Biotech*, 6(1), 1–9. <https://doi.org/10.1007/s13205-016-0368-z>
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14(1), 60. <https://doi.org/10.1186/1471-2105-14-60>
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., & Giovannelli, D. (2019). Living at the extremes: Extremophiles and the limits of life in a planetary context. *Frontiers in Microbiology*, 10(MAR). <https://doi.org/10.3389/fmicb.2019.00780>
- 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, 1–8. <https://doi.org/10.1186/1471-2105-9-386>
- Miller, A. Z., Garcia-Sanchez, A. M., Pereira, M. F. ~C., Gazquez, F., Calaforra, J. M., Forti, P., Toulkeridis, T., Mart\`inez-Fr\`ias, J., & Saiz-Jimenez, C. (2016). Biomineralization and biosignatures of coralloid-type speleothems from lava tubes of Galapagos Islands: evidences on the fossil record of prokaryotes. *EGU General Assembly Conference Abstracts*, EPSC2016-16574.
- Mora-González, P. C., Espinosa-Luna, G., Ramírez-Higuera, A., Peña-Montes, C., Valerio-Alfaro, G., Sánchez-Flores, A., Quintana-Castro, R., Sánchez-Otero, M. G., & Oliart-Ros, R. M. (2020). Metagenomic Approach to Bacterial Diversity and Lipolytic Enzymes' Genes from a Steam Soil of Los Humeros Geothermal Field (Puebla, México). *Geomicrobiology Journal*, 0(0), 1–11. <https://doi.org/10.1080/01490451.2020.1852452>

- Moroenyane, I., & Yergeau, É. (2019). 7. Techniques and approaches to quantify microbial diversity in extreme environments. *Advanced Techniques for Studying Microorganisms in Extreme Environments*, 151–166. <https://doi.org/10.1515/9783110525786-007>
- Muchowska, K. B., Varma, S. J., Chevallot-Beroux, E., Lethuillier-Karl, L., Li, G., & Moran, J. (2017). Metals promote sequences of the reverse Krebs cycle. *Nature Ecology and Evolution*, 1(11), 1716–1721. <https://doi.org/10.1038/s41559-017-0311-7>
- Muhadesi, J. B., Huang, Y., Wang, B. J., Jiang, C. Y., & Liu, S. J. (2019). Acidibrevibacterium fodinaquatile gen. Nov., sp. nov., isolated from acidic mine drainage. *International Journal of Systematic and Evolutionary Microbiology*, 69(10), 3248–3255. <https://doi.org/10.1099/ijsem.0.003618>
- Myers, M. R., & King, G. M. (2016). Isolation and characterization of Acidobacterium ailaui sp. Nov., a novel member of Acidobacteria subdivision 1, from a geothermally heated Hawaiian microbial mat. *International Journal of Systematic and Evolutionary Microbiology*, 66(12), 5328–5335. <https://doi.org/10.1099/ijsem.0.001516>
- Najar, I. N., Sherpa, M. T., Das, S., Das, S., & Thakur, N. (2020). Diversity analysis and metagenomic insights into the antibiotic resistance and metal resistances among Himalayan hot spring bacteriobiome-insinuating inherent environmental baseline levels of antibiotic and metal tolerance. *Journal of Global Antimicrobial Resistance*, 21, 342–352. <https://doi.org/10.1016/j.jgar.2020.03.026>
- Nakaew, N., Pathom-aree, W., & Lumyong, S. (2009). First Record of the Isolation, Identification and Biological Activity of a New Strain of Spirillospora albida from Thai Cave Soil. *Actinomycetologica*, 23(1), 1–7. <https://doi.org/10.3209/saj.saj230102>
- Navarrette-Bedolla, M., Ballesteros-Almanza, M. L., Sanchez-Yanez, J. M., Valdez-Salas, B., & Hernandez-Duque, G. (1999). Biocorrosion in a geothermal power plant. *Materials Performance*, 38(4).
- Norris, P. R., Davis-Belmar, C. S., Brown, C. F., & Calvo-Bado, L. A. (2011). Autotrophic, sulfur-oxidizing actinobacteria in acidic environments. *Extremophiles*, 15(2), 155–163. <https://doi.org/10.1007/s00792-011-0358-3>
- Olson, N. D., Treangen, T. J., Hill, C. M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., & Pop, M. (2018). Metagenomic assembly through the lens of validation: Recent advances in assessing and improving the quality of genomes assembled from metagenomes. *Briefings in Bioinformatics*, 20(4), 1140–1150. <https://doi.org/10.1093/bib/bbx098>
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1), 206–214. <https://doi.org/10.1093/nar/gkt1226>
- Panda, A. K., Bisht, S. S., De Mandal, S., & Kumar, N. S. (2018). Microbial Diversity of Thermophiles Through the Lens of Next Generation Sequencing. In *Microbial Diversity in the Genomic Era*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-814849-5.00013-7>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM:

- Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Patwardhan, S., Smedile, F., Giovannelli, D., & Vetriani, C. (2021). Metaproteogenomic Profiling of Chemosynthetic Microbial Biofilms Reveals Metabolic Flexibility During Colonization of a Shallow-Water Gas Vent. *Frontiers in Microbiology*, 12(April), 1–17. <https://doi.org/10.3389/fmicb.2021.638300>
- Prieto-Barajas, C. M., Alcaraz, L. D., Valencia-Cantero, E., & Santoyo, G. (2018). Life in Hot Spring Microbial Mats Located in the Trans-Mexican Volcanic Belt: A 16S/18S rRNA Gene and Metagenomic Analysis. *Geomicrobiology Journal*, 35(8), 704–712. <https://doi.org/10.1080/01490451.2018.1454555>
- Prieto-Barajas, C. M., Valencia-Cantero, E., & Santoyo, G. (2018). Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. In *Electronic Journal of Biotechnology* (Vol. 31). <https://doi.org/10.1016/j.ejbt.2017.11.001>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Ranawat, P., & Rawat, S. (2017). Stress response physiology of thermophiles. *Archives of Microbiology*, 199(3), 391–414. <https://doi.org/10.1007/s00203-016-1331-4>
- Ranawat, P., & Rawat, S. (2018). Metal-tolerant thermophiles: metals as electron donors and acceptors, toxicity, tolerance and industrial applications. *Environmental Science and Pollution Research*, 25(5), 4105–4133. <https://doi.org/10.1007/s11356-017-0869-2>
- 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. <https://doi.org/10.1093/bioinformatics/btv681>
- Rodríguez, A., & Van Bergen, M. J. (2016). Volcanic hydrothermal systems as potential analogues of Martian sulphate-rich terrains. *Geologie En Mijnbouw/Netherlands Journal of Geosciences*, 95(2), 153–169. <https://doi.org/10.1017/njg.2015.12>
- Roth, H., Gallo, S., Badger, P., & Hillwig, M. (2019). Changes in microbial communities of a passive coal mine drainage bioremediation system. *Canadian Journal of Microbiology*, 65(10), 775–782. <https://doi.org/10.1139/cjm-2018-0612>
- Roy, C., Rameez, M. J., Haldar, P. K., Peketi, A., Mondal, N., Bakshi, U., Mapder, T., Pyne, P., Fernandes, S., Bhattacharya, S., Roy, R., Mandal, S., O’Neill, W. K., Mazumdar, A., Mukhopadhyay, S. K., Mukherjee, A., Chakraborty, R., Hallsworth, J. E., & Ghosh, W. (2020). Microbiome and ecology of a hot spring-microbialite system on the Trans-Himalayan Plateau. *Scientific Reports*, 10(1), 1–22. <https://doi.org/10.1038/s41598-020-62797-z>
- Ruff, S. W., Campbell, K. A., Van Kranendonk, M. J., Rice, M. S., & Farmer, J. D. (2020). The case for ancient hot springs in gusev crater, mars. *Astrobiology*, 20(4), 475–499. <https://doi.org/10.1089/ast.2019.2044>
- Sánchez-Andrea, I., Rodríguez, N., Amils, R., & Sanz, J. L. (2011). Microbial diversity in anaerobic sediments at Río Tinto, a naturally acidic environment with a high heavy metal content. *Applied*

and Environmental Microbiology, 77(17), 6085–6093. <https://doi.org/10.1128/AEM.00654-11>

- Sánchez-Pérez, J., Villar, J. D., Alvarez-López, N., Águila, B., Buenrostro, J., Chino-Palomo, L. J., Navarro-Miranda, M., Cifuentes, J. F., Cruz-Cruz, A. G., Vega-Baray, B., Viladomat, M., Kalambokidis, M., Eguiarte, L. E., & Souza, V. (2020). The Importance of the Rare Biosphere for Astrobiological Studies and the Diversification and Resilience of Life on Earth. In V. Souza, A. Segura, & J. S. Foster (Eds.), *Astrobiology and Cuatro Ciénegas Basin as an Analog of Early Earth* (pp. 135–148). Springer International Publishing. https://doi.org/10.1007/978-3-030-46087-7_6
- Satyanarayana, Tulasi, Kawarabayasi, Y., & Littlechild, J. (2013). Thermophilic microbes in environmental and industrial biotechnology: Biotechnology of thermophiles. In *Thermophilic Microbes in Environmental and Industrial Biotechnology: Biotechnology of Thermophiles*. <https://doi.org/10.1007/978-94-007-5899-5>
- Satyanarayana, Tulsi, Raghukumar, C., & Shivaji, S. (2005). *Extremophilic microbes: Diversity and perspectives*.
- Schmitt, E. K., Hoff, B., & Kück, U. (2004). Regulation of cephalosporin biosynthesis. *Advances in Biochemical Engineering/Biotechnology*, 88(May 2014), 1–43. <https://doi.org/10.1007/b99256>
- Segata, N., Börnigen, D., Morgan, X. C., & Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*, 4. <https://doi.org/10.1038/ncomms3304>
- Servín-Garcidueñas, L. E., & Martínez-Romero, E. (2014). Draft genome sequence of the Sulfolobales archaeon AZ1, obtained through metagenomic analysis of a Mexican hot spring. *Genome Announcements*, 2(2), 1–2. <https://doi.org/10.1128/genomeA.00164-14>
- Servín Garcidueñas, L. E., & Martínez Romero, M. E. (2015). *Estudio metagenómico de microorganismos extremófilos del campo geotérmico de los azufres y diversidad genómica de simbioses de phaseolus e insectos nativos de México*. [Tesis de Doctorado, Universidad Nacional Autónoma de México]. <http://132.248.9.195/ptd2015/agosto/0734923/Index.html>
- Shiers, D. W., Collinson, D. M., & Watling, H. R. (2016). Life in heaps: a review of microbial responses to variable acidity in sulfide mineral bioleaching heaps for metal extraction. *Research in Microbiology*, 167(7), 576–586. <https://doi.org/https://doi.org/10.1016/j.resmic.2016.05.007>
- Shivlata, L., & Satyanarayana, T. (2015). Thermophilic and alkaliphilic Actinobacteria: Biology and potential applications. *Frontiers in Microbiology*, 6(SEP), 1–29. <https://doi.org/10.3389/fmicb.2015.01014>
- Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843. <https://doi.org/10.1038/s41564-018-0171-1>
- Souza-Brito, E. M., Romero-Núñez, V. M., Caretta, C. A., Bertin, P., Valerdi-Negreros, J. C., Guyoneaud, R., & Goñi-Urriza, M. (2019). The bacterial diversity on steam vents from Parícutín and Sapichu volcanoes. *Extremophiles*, 23(2), 249–263. <https://doi.org/10.1007/s00792-019-01078-8>
- Strazzulli, A., Fusco, S., Cobucci-Ponzano, B., Moracci, M., & Contursi, P. (2017). Metagenomics of microbial and viral life in terrestrial geothermal environments. *Reviews in Environmental Science*

and Biotechnology, 16(3), 425–454. <https://doi.org/10.1007/s11157-017-9435-0>

- Sun, S., Jia, T., Chen, K., Peng, Y., & Zhang, L. (2019). Simultaneous removal of hydrogen sulfide and volatile organic sulfur compounds in off-gas mixture from a wastewater treatment plant using a two-stage bio-trickling filter system. *Frontiers of Environmental Science and Engineering*, 13(4). <https://doi.org/10.1007/s11783-019-1148-x>
- Tang, K., Zhang, Y., Lin, D., Han, Y., Chen, C. T. A., Wang, D., Lin, Y. S., Sun, J., Zheng, Q., & Jiao, N. (2018). Cultivation-independent and cultivation-dependent analysis of microbes in the shallow-sea hydrothermal system off Kueishantao Island, Taiwan: Unmasking heterotrophic bacterial diversity and functional capacity. *Frontiers in Microbiology*, 9(FEB), 1–15. <https://doi.org/10.3389/fmicb.2018.00279>
- Torres-Sanchez, R., Magaña-Vazquez, A., Sanchez-Yañez, J. M., & Martinez Gomez, L. (1996). High temperature microbial corrosion in the condenser of a geothermal electric power unit. *NACE - International Corrosion Conference Series, 1996-March*(3).
- Tortorella, E., Tedesco, P., Esposito, F. P., January, G. G., Fani, R., Jaspars, M., & De Pascale, D. (2018). Antibiotics from deep-sea microorganisms: Current discoveries and perspectives. *Marine Drugs*, 16(10), 1–16. <https://doi.org/10.3390/md16100355>
- Urbieto, M. S., Donati, E. R., Chan, K. G., Shahar, S., Sin, L. L., & Goh, K. M. (2015). Thermophiles in the genomic era: Biodiversity, science, and applications. *Biotechnology Advances*, 33(6), 633–647. <https://doi.org/10.1016/j.biotechadv.2015.04.007>
- Valdez Salas, B., Schorr Wiener, M., Rioseco de la Peña, L., & Navarrete Bedolla, M. (2000). Deterioration of materials in geothermal fields in Mexico. *Materials and Corrosion*, 51(10), 698–704. [https://doi.org/10.1002/1521-4176\(200010\)51:10<698::aid-maco698>3.3.co;2-7](https://doi.org/10.1002/1521-4176(200010)51:10<698::aid-maco698>3.3.co;2-7)
- Von Meijenfildt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., & Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology*, 20(1), 1–14. <https://doi.org/10.1186/s13059-019-1817-x>
- Wall, K., Cornell, J., Bizzoco, R. W., & Kelley, S. T. (2015). Biodiversity hot spot on a hot spot: Novel extremophile diversity in Hawaiian fumaroles. *MicrobiologyOpen*, 4(2), 267–281. <https://doi.org/10.1002/mbo3.236>
- Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: A customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 12(Cd). <https://doi.org/10.1186/1471-2164-12-444>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2015). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Xia, G., Zhou, X., Hu, J., Sun, Z., Yao, J., Chen, D., & Wang, J. (2019). Simultaneous removal of carbon disulfide and hydrogen sulfide from viscose fibre waste gas with a biotrickling filter in pilot scale. *Journal of Cleaner Production*, 230, 21–28. <https://doi.org/10.1016/j.jclepro.2019.05.097>
- Yang, N., Wang, C., & Han, M. F. (2020). Gel-encapsulated microorganisms used as a strategy to rapidly recover biofilters after starvation interruption. *Journal of Environmental Management*, 261(February), 110237. <https://doi.org/10.1016/j.jenvman.2020.110237>

- Zhou, D., Peng, T., Zhou, H., Liu, X., Gu, G., Chen, M., Qiu, G., & Zeng, W. (2015). Expression of Critical Sulfur- and Iron-Oxidation Genes and the Community Dynamics During Bioleaching of Chalcopyrite Concentrate by Moderate Thermophiles. *Current Microbiology*, *71*(1), 62–69. <https://doi.org/10.1007/s00284-015-0817-7>
- Ziegler, S., Waidner, B., Itoh, T., Schumann, P., Spring, S., & Gescher, J. (2013). *Metallibacterium scheffleri* gen. nov., sp. nov., an alkalinizing gammaproteobacterium isolated from an acidic biofilm. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART4), 1499–1504. <https://doi.org/10.1099/ijms.0.042986-0>
- Ziko, L., Adel, M., Malash, M. N., & Siam, R. (2019). Insights into red sea brine pool specialized metabolism gene clusters encoding potential metabolites for biotechnological applications and extremophile survival. *Marine Drugs*, *17*(5). <https://doi.org/10.3390/md17050273>

10. ANEXOS

10.1 Otros grupos de organismos detectados en los metagenomas

Tabla S1. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a arqueas encontradas en el metagenoma de FLA, realizadas en la plataforma de BLASTN de NCBI, para las secuencias cuyo mejor hit fue una secuencia ambiental, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida.

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
1469	<i>Euryarchaeota</i>	<i>Thermoplasmata</i>	<i>Uncultured Thermoplasmatales archaeon clone METASED07</i>	KJ907760	100	100
			<i>Thermoplasma acidophilum</i> 122-1B2	NR_028235	84.67	100
2533	<i>Euryarchaeota</i>	<i>Methanococci</i>	<i>Uncultured archaeon clone B133706F09</i>	FN863102	99.88	100
			<i>Methanococcus aeolicus</i> Nankai-3	NR_074182	74.57	49
1465	<i>Euryarchaeota</i>	<i>Thermoplasmata</i>	<i>Uncultured Thermoplasmatales archaeon clone METASED02</i>	KJ907755	97.34	100
			<i>Cuniculiplasma divulgatum</i> S5	NR_144620	92.57	100
1465	<i>Euryarchaeota</i>	<i>Thermoplasmata</i>	<i>Cuniculiplasma divulgatum</i> S5	NR_144620	99.66	100
1483	<i>Euryarchaeota</i>	<i>Thermoplasmata</i>	<i>Uncultured DPANN archaeon clone Parys_Mountain_Mia_clone_7</i>	MH463114	91.41	94
			<i>Cuniculiplasma divulgatum</i> S5	NR_144620	73.94	71
1485	<i>Crenarchaeota</i>	<i>Thermoprotei</i>	<i>Uncultured thermal soil archaeon clone YNPFFA4</i>	AF391990	97.22	94
			<i>Thermogladius calderae</i> 1633	NR_148751	84.62	99
1997	<i>Euryarchaeota</i>	<i>Halobacteria</i>	<i>Uncultured euryarchaeote clone METASED09</i>	KJ907762	99.75	100
			<i>Halostagnicola alkaliphila</i> 167-74	NR_113016	74.65	58

Tabla S2. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a organelos de eucariontes encontradas en el metagenoma de FLA, realizadas en la plataforma de BLASTN de NCBI, para las secuencias cuyo mejor hit fue una secuencia ambiental, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida.

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
1554	<i>Heterolobosea</i>	<i>Tetramitia</i>	<i>Naegleria lovaniensis</i> 7372	AY376155	76.74	31
1534	<i>Discosea</i>	<i>n.d</i>	<i>Acanthamoeba</i> sp. JPH19 mitochondrial	AB795713	80.24	99
2567	<i>Rhodophyta</i>	<i>Bangiophyceae</i>	<i>Cyanidiaceae</i> sp. MX-AZ01 mitochondrion	KJ569774	99.96	100
1423	<i>Rhodophyta</i>	<i>Bangiophyceae</i>	<i>Uncultured Cyanidiaceae clone METASED23 chloroplast</i>	KJ907776	100	100
3388	<i>Heterolobosea</i>	<i>Tetramitia</i>	<i>Uncultured organism clone MB0515c3</i>	AY897693	87.42	25
			<i>Naegleria gruberi</i> mitochondrial DNA	AF288092	76.88	29

Tabla S3. Alineamientos de secuencias de genes ribosomales 16S rARN correspondientes a organelos de eucariontes encontradas en el metagenoma de MM, realizadas en la plataforma de BLASTN de NCBI, para las secuencias cuyo mejor hit fue una secuencia ambiental, se agregó en la parte inferior el hit más cercano con una secuencia de referencia conocida.

Longitud (pb)	Filo	Clase	GenBank Hit	Numero de acceso	% identidad	% cobertura
4948	<i>Stramenopiles</i>	<i>Hyphochytriomycetes</i>	<i>Hyphochytrium catenoides</i> mitochondrion	LT578416	74.37	20
1964	<i>Viridiplantae</i>	<i>Chlorophyta</i>	<i>Trebouxiophyceae</i> sp. MX-AZ01 mitochondrion	JX315601	100	94
1514	<i>Viridiplantae</i>	<i>Chlorophyta</i>	<i>Trebouxiophyceae</i> sp. MX-AZ01 chloroplast	JX402620	99.87	100
1663	<i>Discosea</i>	<i>Flabellinia</i>	Uncultured bacterium clone B234806A03	FN863845	98.6	51
			<i>Paravannella minima</i> mitochondrion	NC040955	74.8	42
1577	<i>Discosea</i>	<i>n.d</i>	Uncultured bacterium clone NBDTU13	FJ529933	81.5	92
			<i>Jakoba bahamiensis</i> ATCC50695 mitochondrion	KC353354	78.28	78
3874	<i>Stramenopiles</i>	<i>Oomycota</i>	Uncultured bacterium 051011_S1_W_T_SDP_079	FJ524933	87.56	22
			<i>Pythium insidiosum</i> mitochondrial DNA	AP014838	75.2	19
1531	<i>Discosea</i>	<i>Flabellinia</i>	Uncultured bacterium clone O218706A04	FN866364	98.28	57
			<i>Paravannella minima</i> mitochondrion	NC040955	73.6	72
1579	<i>Discosea</i>	<i>n.d</i>	Uncultured bacterium clone O138106A04	FN865852	100	54
			<i>Jakoba bahamiensis</i> ATCC 50695 mitochondrion	KC353354	77.12	81

10.2 Ensamblaje de Metagenomas

Primero se realizó la revisión de calidad de los metagenomas con FastQC, con los parámetros predeterminados de software. Posteriormente, los metagenomas fueron filtrados con Trimmomatic con el fin de descartar *contigs* con valores de calidad inferiores a Q30 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:30 MINLEN:36) para el metagenoma de FLA y Q33 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:33 MINLEN:36) para el metagenoma de MM. Se elaboró el ensamblaje de las secuencias con SPAdes usando el parámetro (-meta) recomendado para ensamblajes metagenómicos (Tabla 1, Tabla S4).

Tabla S4. Valores derivados de los ensamblajes de secuencias de los metagenomas provenientes de FLA y MM con el software de ensamblaje SPAdes, los valores fueron obtenidos al evaluar los *contigs* con QUAST. En rojo se señala los ensamblajes seleccionados de acuerdo con sus características. (Continuación).

Estadísticos	<i>Contigs</i> FLA K21_99	<i>Contigs</i> FLAK33_111	<i>Contigs</i> FLAK55_127	<i>Contigs</i> MMK21_99	<i>Contigs</i> MMK33_111	<i>Contigs</i> MMK55_127
K-meros	21, 33, 55, 77, 99	33, 55, 77, 99, 111	55, 77, 99, 111, 127	21, 33, 55, 77, 99	33, 55, 77, 99, 111	55, 77, 99, 111, 127
# <i>contigs</i>	86317	86675	83353	196668	197452	188996
# <i>contigs</i> (>= 0 bp)	177623	166817	146374	391625	363915	311327
# <i>contigs</i> (>= 1000 bp)	31429	31269	29731	73453	73037	70458
# <i>contigs</i> (>= 5000 bp)	4946	4954	4883	11361	10997	10523
# <i>contigs</i> (>= 10000 bp)	2702	2683	2640	5695	5513	5224
# <i>contigs</i> (>= 25000 bp)	994	998	948	2065	2010	1899
# <i>contigs</i> (>= 50000 bp)	364	339	331	745	731	730
<i>Contig</i> más largo	401414	502202	372776	835358	689788	689788
Longitud total	183650729	182405891	176350784	411511324	407837617	390528220
Longitud total (>= 0 bp)	215215420	211084537	199739366	486892674	474025919	440966402
Longitud total (>= 1000 bp)	146345918	144792772	140081994	328678301	324220141	310759309
Longitud total (>= 5000 bp)	97489100	96541710	94530407	210448137	206654974	198165047
Longitud total (>= 10000 bp)	81754437	80646989	78755184	171192594	168505562	161331326
Longitud total (>= 25000 bp)	55353549	54524926	52440737	115059190	114027592	109663946
Longitud total (>= 50000 bp)	33458734	31726093	31065623	69519489	69838184	69029908
N50	6519	6373	6740	5421	5259	5263
N75	1233	1212	1217	1256	1230	1227
L50	3948	4006	3787	10461	10464	9958
L75	23646	23986	22613	55496	56438	54283
GC (%)	55.95	56.03	56.09	55.31	55.46	55.76

10.3 Estructura y diversidad de los metagenomas

El proceso de filtrado y curaduría de las secuencias ribosomales 16S se realizó por medio del programa RNAmmer con el uso del siguiente comando: `rnammer -S bac -m ssu -gff -f`. Posteriormente, la identificación taxonómica en BLASTN versión en línea se llevó a cabo con los parámetros predeterminados usando como base de datos las secuencias nr/nt almacenadas en GenBank.

10.4 Recuperación de genomas

Tanto para la elaboración de *bins*, como la recuperación con genomas de referencia, se realizaron mapeos usando Bowtie 2. Para ello, se crearon archivos de indexación con el comando `Bowtie2-build`, paso seguido se llevó a cabo el proceso de mapeo con el comando `Bowtie2`, para crear los archivos de cobertura para la elaboración de *bins* para MetaBAT y CONCOCT. Los archivos de mapeo en formato SAM, fueron convertidos en archivos BAM a través del software SAMTOOLS con el comando `samtools view -bS` y ordenados con el comando `samtools sort`. Finalmente, para obtener el archivo con la información de la cobertura para MetaBAT se usó un paquete de este programa llamado `jgi_summarize_bam_contig_depths`, mientras que para CONCOCT se usó el comando `concoct_coverage_table.py`.

Para la elaboración de los MAG, se corrieron los programas Maxbin, MetaBAT y CONCOCT con su configuración predeterminada. Acto seguido, se filtraron usando DASTool, lo cual requirió convertir la información de los MAG en formatos tabulares con el paquete `Fasta_to_Scaffolds2Bin.sh`, después de obtener los archivos tabulados, se procedió a realizar el proceso de filtrado con DASTool con los parámetros predeterminados, los resultados del análisis se pueden observar en la **Figura S1** para FLA y en la **Figura S2** para MM. Posteriormente los MAG de cada metagenoma fueron analizados para observar completitud y contaminación con CheckM con el comando `lineage_wf` para poder obtener la información de porcentaje de contaminación y completitud presentados en la **Figura S3**.

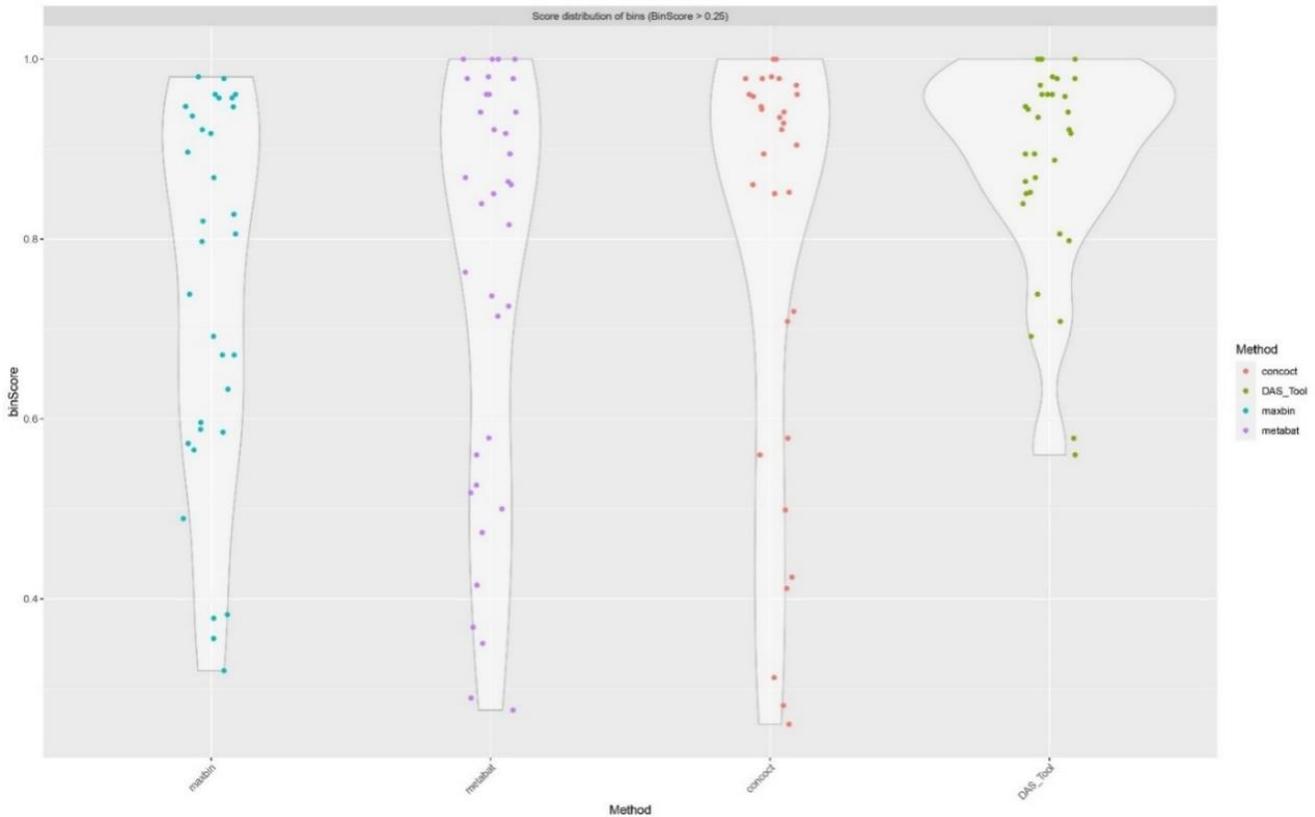


Figura S1. Filtración por calidad de *bins* recuperados de las lecturas ensambladas del metagenoma de FLA con el uso de DASTool, los *bins* con un puntaje mayor a 0,6 fueron los seleccionados por ser de mayor calidad entre el total de los *bins* elaborados con cada uno de los programas (Maxbin, MetaBAT y CONCOCT).

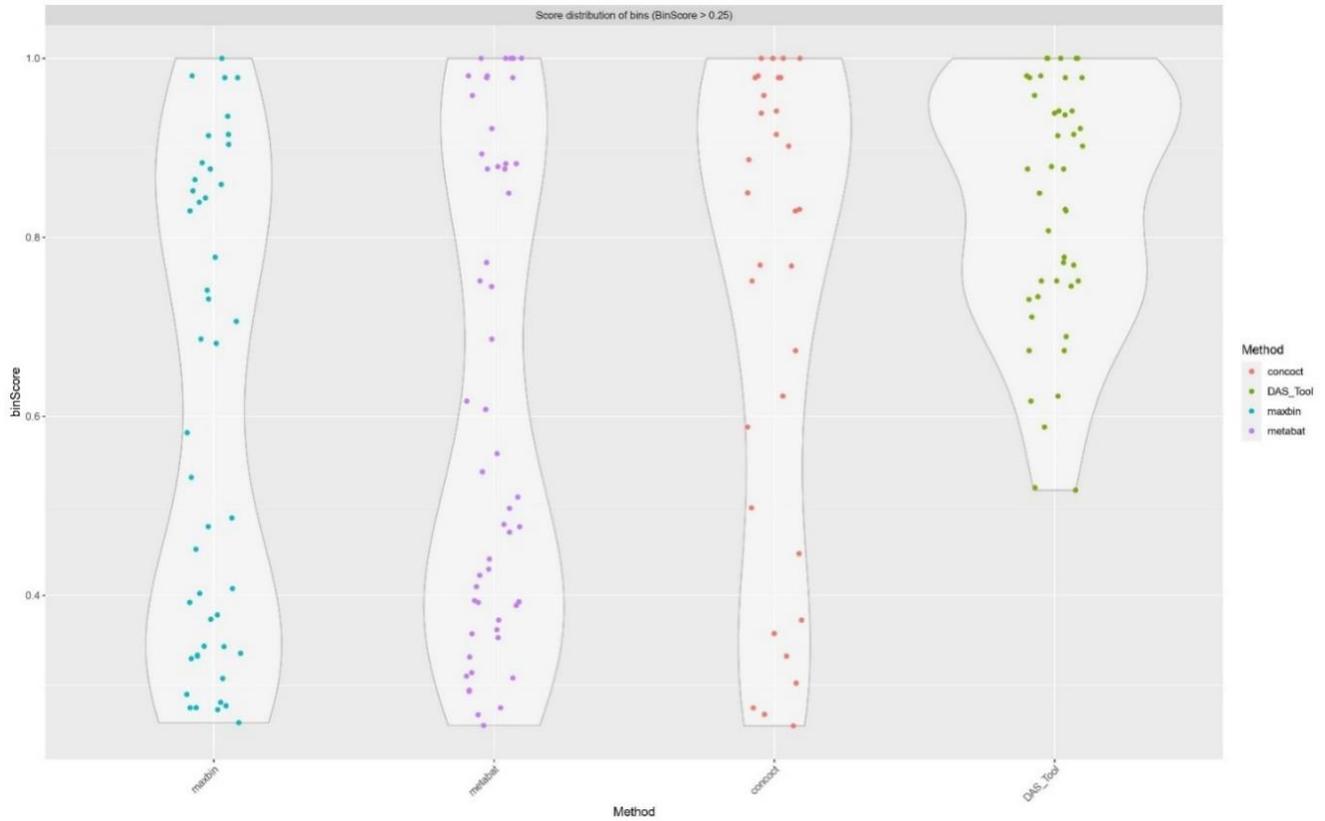


Figura S2. Filtración por calidad de *bins* recuperados de las lecturas ensambladas del metagenoma de MM con el uso de DASTool, los *bins* con un puntaje mayor a 0,6 fueron los seleccionados por ser de mayor calidad entre el total de los *bins* elaborados con cada uno de los programas (Maxbin, MetaBAT y CONCOCT).

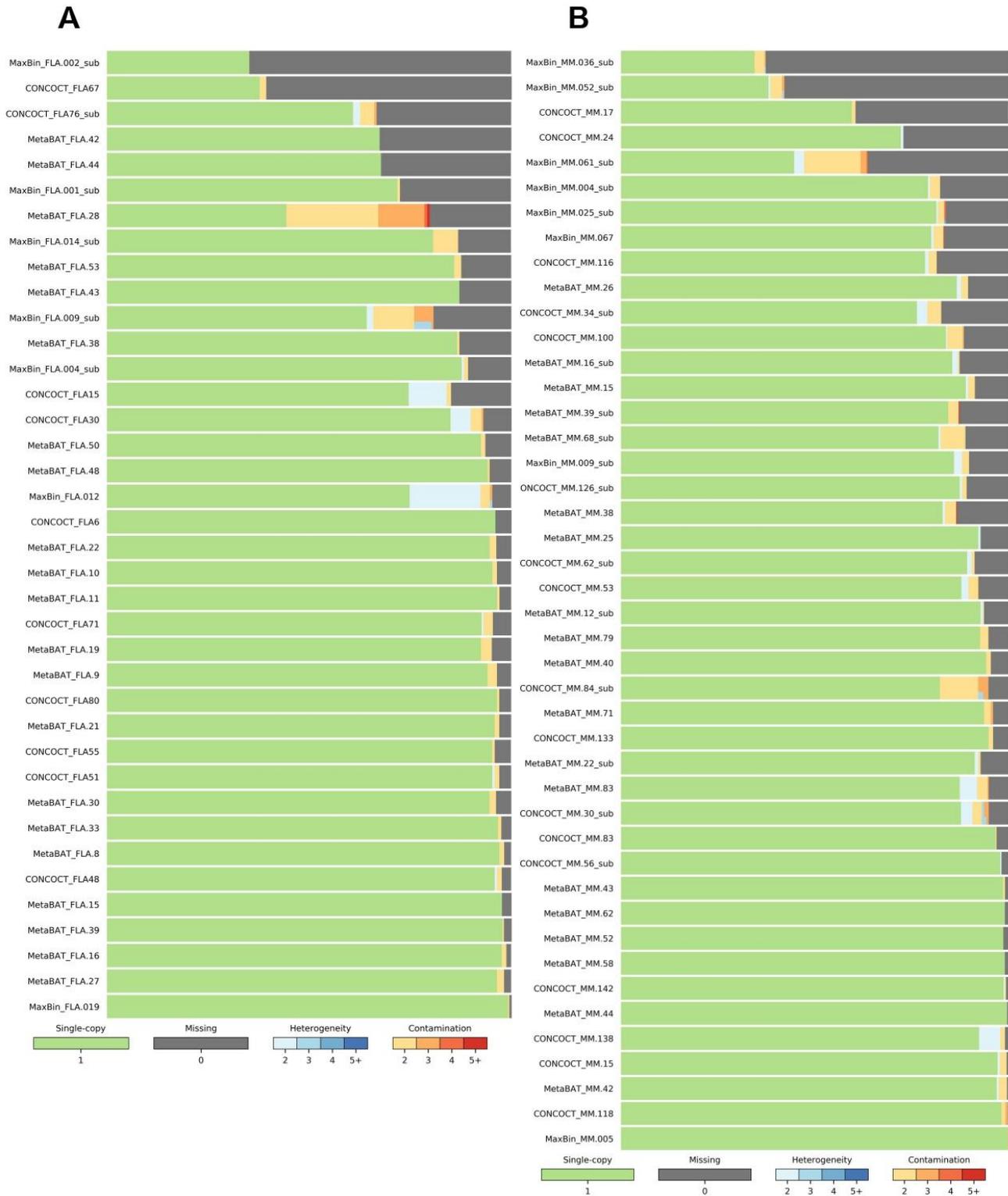


Figura S3. Representación de la integridad de los MAGs recuperados del ensamblaje del metagenoma FLA (A), y para el metagenoma de MM (B).

Para la recuperación de genomas a partir de genomas de referencia, se descargaron los genomas de *Acidibrevibacterium fodinaquatile* G 45-3 con número de acceso: GCA003352165, *Aciditerrimonas ferrireducens* JCM15389 con número de acceso: GCA001311945, para el metagenoma de FLA y *Acidobacterium capsulatum* ATCC51196 con número de acceso: GCF000022565, *Metallibacterium scheffleri* DKE6 con número de acceso: GCF002077135, para el metagenoma de MM. Los cuales fueron mapeados contra las lecturas de los metagenomas correspondientes usando Bowtie2 como se describió arriba, posteriormente se extrajo la información de las lecturas mapeadas con el uso de SAMTOOLS por medio del siguiente comando `samtools view -S -F4`. Luego, los ID de las lecturas mapeadas se obtuvieron a través del uso de comandos de UNIX `cut -f1 | sort | uniq`, para finalmente recuperar las lecturas con el uso de la herramienta seqtk disponible en (<https://github.com/lh3/seqtk>). Las lecturas se ensamblaron en SPAdes usando el parámetro `--only-assembler`, con los K-meros (21, 33, 55, 77,99), (33, 55, 77, 99, 111) y (55, 77, 99, 111, 127).

Tabla S5. Parámetros y resultados de los ensamblajes de ‘*Candidatus Acidibrevibacterium* sp. FLA01’ realizados en SPAdes. En rojo se señala el ensamblaje seleccionado de acuerdo con sus características.

Estadísticos	‘Ca. <i>Acidibrevibacterium</i> sp. FLA01’ K21-99	‘Ca. <i>Acidibrevibacterium</i> sp. FLA01’ K33-111	‘Ca. <i>Acidibrevibacterium</i> sp. FLA01’ k55-127
# contigs	369	394	1251
# contigs (>= 0 bp)	609	625	1594
# contigs (>= 1000 bp)	259	276	613
# contigs (>= 5000 bp)	98	100	199
# contigs (>= 10000 bp)	67	61	107
# contigs (>= 25000 bp)	36	39	20
# contigs (>= 50000 bp)	18	18	6
Contig más largo	295112	243461	127122
Total lectura	3759483	3807382	4154289
Total lectura (>= 0 bp)	3839563	3888630	4275341
Total lectura (>= 1000 bp)	3680110	3724640	3735335
Total lectura (>= 5000 bp)	3352216	3364327	2830499
Total lectura (>= 10000 bp)	3130107	3086421	2190428
Total lectura (>= 25000 bp)	2621178	2733031	933510
Total lectura (>= 50000 bp)	1979655	1950856	474240
N50	53782	58699	10542
N75	20141	20565	3358
L50	17	18	97
L75	45	45	269
GC (%)	65.79	65.78	65.82

Tabla S6. Parámetros y resultados de los ensamblajes de ‘*Candidatus Aciditerrimonas FLA02*’ realizados en SPAdes. En rojo se señala el ensamblaje seleccionado de acuerdo con sus características.

Estadísticos	‘<i>Ca. Aciditerrimonas sp. FLA02</i>’_K21-99	‘<i>Ca. Aciditerrimonas sp. FLA02</i>’_K33-111	‘<i>Ca. Aciditerrimonas sp. FLA02</i>’_K55-127
# contigs	1430	1570	1676
# contigs (>= 0 bp)	1802	1923	1997
# contigs (>= 1000 bp)	874	914	938
# contigs (>= 5000 bp)	66	64	62
# contigs (>= 10000 bp)	5	6	5
# contigs (>= 25000 bp)	0	0	0
# contigs (>= 50000 bp)	0	0	0
Contig más largo	14875	14875	14875
Total lectura	2470963	2590535	2692830
Total lectura (>= 0 bp)	2621913	2736851	2828225
Total lectura (>= 1000 bp)	2073118	2124147	2169730
Total lectura (>= 5000 bp)	449085	436560	429431
Total lectura (>= 10000 bp)	60720	72581	62534
Total lectura (>= 25000 bp)	0	0	0
Total lectura (>= 50000 bp)	0	0	0
N50	2253	2129	2054
N75	1305	1248	1209
L50	313	345	367
L75	683	753	802
GC (%)	71.5	71.47	71.42

Tabla S7. Parámetros y resultados de los ensamblajes de *Candidatus Acidobacterium* sp. MM01 realizados en SPAdes. En rojo se señala el ensamblaje seleccionado de acuerdo con sus características.

Estadísticos	'Ca. Acidobacterium sp. MM01' K21-99	'Ca. Acidobacterium sp. MM01' K33-111	'Ca. Acidobacterium sp. MM01' K55-127
# contigs	1672	1704	1791
# contigs (>= 0 bp)	2119	2114	2139
# contigs (>= 1000 bp)	1112	1120	1151
# contigs (>= 5000 bp)	246	242	251
# contigs (>= 10000 bp)	107	112	113
# contigs (>= 25000 bp)	26	26	24
# contigs (>= 50000 bp)	3	3	3
Contig más largo	105539	99908	99908
Total lectura	5571726	5601321	5612048
Total lectura (>= 0 bp)	5755161	5773341	5760919
Total lectura (>= 1000 bp)	5164126	5177647	5147757
Total lectura (>= 5000 bp)	3298803	3284728	3234080
Total lectura (>= 10000 bp)	2340540	2398386	2300618
Total lectura (>= 25000 bp)	1042814	1020081	910389
Total lectura (>= 50000 bp)	218865	213234	208509
N50	7036	7004	6598
N75	2606	2590	2452
L50	160	161	177
L75	493	502	536
GC (%)	61.11	61.1	61.09

Tabla S8. Parámetros y resultados de los ensamblajes de ‘*Candidatus Metallibacterium* sp. MM02’ realizados en SPAdes. En rojo se señala el ensamblaje seleccionado de acuerdo con sus características.

Estadísticos	‘<i>Ca. Metallibacterium</i> sp. MM02’_K21-99	‘<i>Ca. Metallibacterium</i> sp. MM02’_K33-111	‘<i>Ca. Metallibacterium</i> sp. MM02’_K55-127
# contigs	1684	1751	1840
# contigs (>= 0 bp)	2062	2120	2181
# contigs (>= 1000 bp)	1092	1101	1097
# contigs (>= 5000 bp)	192	197	184
# contigs (>= 10000 bp)	76	72	79
# contigs (>= 25000 bp)	30	34	33
# contigs (>= 50000 bp)	10	9	12
Contig más largo	94472	94484	88420
Total lectura	5220289	5325355	5417405
Total lectura (>= 0 bp)	5373112	5477461	5562156
Total lectura (>= 1000 bp)	4794911	4859578	4890598
Total lectura (>= 5000 bp)	2907651	2970008	2996795
Total lectura (>= 10000 bp)	2122339	2129444	2293947
Total lectura (>= 25000 bp)	1433976	1550292	1581259
Total lectura (>= 50000 bp)	713189	639287	836824
N50	6309	6116	6268
N75	2250	2175	2077
L50	140	142	133
L75	498	510	522
GC (%)	66.8	66.81	66.8

10.5 Scripts usados en el estudio

FastQC

Fastqc (Archivos de entrada)

Trimmomatic

```
trimmomatic-0.39.jar PE Archivo de entrada_1.fastq Archivo de
entrada_2.fastq Archivo de salida_1_P.fastq Archivo de salida
_1_UP.fastq Archivo de salida_2_P.fastq Archivo de salida
_2_UP.fastq ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:10 TRAILING:10
SLIDINGWINDOW:4:30 MINLEN:36
```

SPAdes

Metagenomas

```
spades.py -k K-meros --meta -1 Archivo_1.fastq -2 Archivo_2.fastq -o
SPADES(K-mero inicial y final)
```

genomas

```
spades.py -k K-meros --only-assembler -1 Archivo_1.fastq -2
Archivo_2.fastq -o SPADES(K-mero inicial y final)
```

Bowtie2

Generación de archivos de indexación

```
bowtie2-build contigs.fasta contings.btindex
```

Mapeo de lecturas.

```
bowtie2 --threads 25 -x contings.btindex -q -1 B_1.fastq -2 B_2.fastq
-S archivo de mapeo.sam
```

Conversión a archivo BAM

```
Samtools view -bS -o archivo de mapeo.bam archivo de mapeo.sam
```

Ordenación de archivo BAM

```
Samtools sort archivo de mapeo.bam -o archivo de mapeo.sorted.bam
```

Maxbin

```
run_MaxBin.pl -contig contigs.fasta -reads Archivo de
lecturas_1.fastq -reads2 Archivo de lecturas_2.fastq -out prefijo de
archivo de salida
```

MetaBat

Elaboración de tablas de profundidad de lecturas

```
jgi_summarize_bam_contig_depths --outputDepth depth.txt --
pairedContigs FLA_paired.txt Archivo de mapeo.sorted.bam
```

Elaboración de MAG

```
metabat -i contigs.fasta -a depth.txt -o Directorio
```

CONCOCT

fragmentación de *contigs*

```
cut_up_fasta.py contigs.fasta -c 10000 -o 0 --merge_last -b
contigs_10K.bed > contigs_10K.fa
```

Elaboración de tabla de cobertura

```
concoct_coverage_table.py contigs_10K.bed archivo de
mapeo.sorted.bam>tabla de cobertura.tsv
```

Elaboración de *bins*

```
concoct --composition_file contigs_FLA10K.fa --coverage_file
coverage_table.tsv -b concoct_output/
merge_cutup_clustering.py concoct_output/clustering_gt1000.csv >
concoct_output/clustering_merged.csv
```

Extracción de *bins*

```
extract_fasta_bins.py
contigs.fastaconcoct_output/clustering_merged.csv --
output_pathconcoct_output/fasta_bins
```

DASTool

Elaboración de tablas de Scaffold para MaxBin

```
Fasta_to_Scaffolds2Bin.sh -i ../MaxBin -e fasta >
maxbin.scaffolds2bin.tsv
```

Elaboración de tablas de Scaffold para MetaBAT

```
Fasta_to_Scaffolds2Bin.sh -i ../MetaBAT -e fa >
metabat.scaffolds2bin.tsv
```

Elaboración de tablas de Scaffold para CONCOCT

```
perl -pe "s/,/\tconcoct./g;" concoct_output/clustering_gt1000.csv >
concoct.scaffolds2bin.tsv
```

Ejecución de DasTOOL

```
DAS_Tool -i
maxbin.scaffolds2bin.tsv,metabat.scaffolds2bin.tsv,concoct.scaffolds2
bin.tsv -l maxbin,metabat,concoct -c ../CONCOCT/contigs.fasta -t 2 --
write_bins 1 -o directorio de almacenamiento de MAGs
```

CheckM

Ejecución típica de CheckM

```
Checkm lineage_wf -t 20 -x fa directorio de almacenamiento de MAGs/
CheckM/
```

Elaboración de gráficos

```
Checkm bin_qa_plot --image_type png -x fa CheckM/
DASTool_Bins_DASTool_bins/ plots/
```

RNAmmer

```
rnammer -S bac -m ssu -gff -f (Archivo de salida) < (Contigs del
metagenoma)
```

Mauve

`Mauve.jar org.gel.mauve.contigs.ContigOrderer -output (Carpeta de salida) -ref (Genoma de referencia) -draft (Genoma borrador)`

CAT/BAT

Clasificación de MAGs

`CAT bins -b Carpeta contenedora de MAGs -d base de datos de CAT/BAT -t Base de datos taxonomica de CAT/BAT -o prefijo de archivos de salida`

Asignación de nombres al perfil taxonómico.

`CAT add_names -i Archivo de clasificación -o prefijo archivo de salida -t Base de datos taxonomica de CAT/BAT`

Phyloplan

`Phylophan.py (Directorio de genomas a analizar)`