

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

ANÁLISIS METAGENÓMICO DE FUENTES TERMALES DE COSTA RICA:
POTENCIAL FUENTE DE ENZIMAS HIDROLÍTICAS TERMOESTABLES.

Tesis sometida a la consideración de la Comisión del Programa de Estudios de
Posgrado en Ciencias Biomédicas para optar al grado y título de Maestría
Académica en Bioinformática y Biología de Sistemas

LAURA BRENES GUILLÉN

Ciudad Universitaria Rodrigo Facio, Costa Rica
2021

Dedicatoria

A Enrique, mi papá, mi mamá, mi hermana y a los perris que me acompañaron todo este tiempo.

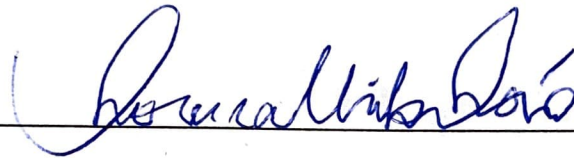
Agradecimientos

A Dios, mi familia, compañeros del CIBCM, a mi tutora Lorena, y lectores César y Rebeca por sus comentarios, sugerencias y recomendaciones. A Daniela, Kaylen y Daniel por su ayuda. A todos los profesores que de una u otra forma estuvieron en el camino y contribuyeron a mi formación profesional.

“Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Ciencias Biomédicas de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Bioinformática y Biología de Sistemas”



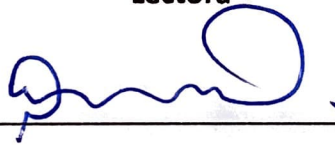
M.S.c. Marielos Mora López
Representante del Decana
Sistema de Estudios de Posgrado



M.Sc. Lorena Uribe Lorío
Profesora Guía



PhD. Rebeca Campos Sánchez
Lectora



PhD. César Rodríguez Sánchez
Lectora



PhD. Christopher Vaglio Cedeño
Representante

Programa de Posgrado en Ciencias Biomédicas



Laura Brenes Guillén
Sustentante

Tabla de contenido	
Dedicatoria	ii
Agradecimientos.....	ii
Hoja de aprobación	iii
Resumen.....	vi
Lista de cuadros	vii
Lista de figuras	viii
Cuadro de abreviaciones.....	x
Introducción.....	1
Antecedentes	6
Caracterización de comunidades de bacterias y arqueas en fuentes termales...	6
Metagenómica <i>shotgun</i> para el estudio de comunidades microbianas.....	7
Ensamblaje y anotación de (meta)genomas	11
Enzimas hidrolíticas termoestables.....	16
Peptidasas	18
Lipasas	20
Amilasas	22
Enzimas involucradas en la remoción de metales pesados y agroquímicos .	23
Justificación.....	26
Objetivos	27
Objetivo General	27
Objetivos Específicos.....	27
Materiales y Métodos	28
Características de las muestras	28
Análisis Bioinformático	35
Análisis de calidad de los datos de secuenciación	37
Ensamblaje, alineamiento y anotación de metagenomas	37
Perfil taxonómico y funcional de las comunidades microbianas	38
Genomas obtenidos a partir de (meta)genomas (MAGs)	41
Creación de base de datos.....	41
Mapeo y abundancia de variantes alélicas	42
Mapeo de genes de resistencia a metales pesados y agroquímicos en tapetes microbianos de fuentes termales.	43
Resultados y discusión.....	44

Análisis de calidad de las secuencias	44
Ensamblaje de los metagenomas	46
Genomas obtenidos a partir de (meta)genomas	49
Comparación del perfil taxonómico de las diferentes fuentes termales	55
Fuentes termales de Costa Rica.....	58
Porcelana, Chile	59
Mound y Bijah, PNY.....	60
Obsidian Pool, PNY	61
Liberty Cap, Mammoth springs, PNY.....	62
Isla Raoult, Nueva Zelanda.....	62
Perfil funcional de las comunidades de las fuentes termales de Costa Rica, PNY, Chile y Nueva Zelanda.....	63
Identificación y clasificación de las enzimas hidrolíticas	65
Peptidasas	73
Proteasas de serina termofílicas.....	73
Metalopeptidasa (Metaloproteínas)	78
Lipasas	92
Esterasas.....	97
Amilasas	102
Genes involucrados en la remoción de metales y agroquímicos encontrados en los tapetes microbianos	105
Arsénico.....	107
Cobre.....	111
Plomo.....	117
Hierro.....	118
Aluminio.....	120
Paraquat	121
Atrazina.....	125
Análisis de correlación entre los genes relacionados con la remoción de agroquímicos y los principales filos encontrados en los tapetes microbianos.	126
Conclusiones.....	130
Recomendaciones.....	132
Referencias	133
Anexos	190

Resumen

En el presente estudio independiente de cultivo, se investigó a nivel taxonómico y funcional las comunidades microbianas de seis metagenomas de fuentes termales cercanas a la Cordillera de Tilarán, volcán Miravalles, volcán Rincón de la Vieja y zona sur, y datos de 13 metagenomas de fuentes termales del Parque Nacional Yellowstone, Nueva Zelanda y Chile. Se obtuvieron en total 229 Gbp de secuencias cortas obtenidas con la plataforma Illumina. Los resultados de la comparación del perfil taxonómico muestran que los tapetes microbianos de Chile y Costa Rica tienen 60% similitud, junto a las muestras de Nueva Zelanda que provienen de agua filtrada, siendo Proteobacteria, Firmicutes y en algunas muestras Cyanobacteria y Chloroflexi los filos más abundantes. Se creó una base de datos de potenciales enzimas hidrolíticas tales como peptidasas, amilasas, lipasas y esterases, así como enzimas involucradas en la remoción de metales pesados y agroquímicos. La base de datos tiene más de 48 000 secuencias, muchas de las cuales podrían ser novedosas y objeto de estudios posteriores. La mayoría de las secuencias tienen un porcentaje de identidad entre el 70% y el 90% con las de microorganismos que toleran altas temperaturas, lo que sugiere que podrían tener funciones similares. Muchas de las enzimas encontradas dentro de estas categorías tienen una función *housekeeping*, por lo que son muy importantes para el mantenimiento de las funciones básicas de las células, sin embargo, podrían participar en los procesos de degradación e hidrólisis de herbicidas y diversos xenobióticos. En los MAGs obtenidos se encontraron genes que participan en el transporte, acumulación y oxidación-reducción de arsénico, cobre, plomo, hierro, aluminio y los herbicidas paraquat y atrazina, estos resultados sugieren la presencia de diversos mecanismos celulares, que les confieren a estos microorganismos resistencia/tolerancia a alta concentraciones de estos metales o por el contrario tienen mecanismos para poder obtenerlos cuando las concentraciones son muy bajas. Este estudio proporcionó un primer acercamiento diversidad funcional de los microorganismos que habitan estos sitios, las secuencias encontradas en este estudio podrían utilizarse para futuras investigaciones en el área de la biorremediación y la geomicrobiología.

Lista de cuadros

Cuadro 1. Localización geográfica, temperatura, pH de muestras utilizadas en este proyecto.....	28
Cuadro 2. Estadísticas de los datos crudos de secuenciación de las muestras de las fuentes termales de Costa Rica	30
Cuadro 3. Características generales de cada muestra utilizada en el análisis.....	31
Cuadro 4. Categorías funcionales según COG.	40
Cuadro 5. Esquema de clasificación numérica para enzimas, basado en las reacciones químicas que catalizan.....	42
Cuadro 6. Resultados de análisis de calidad de las muestras antes y después del trimming.....	46
Cuadro 7. Resultados de MetaQUASTt obtenido a partir de los ensamblajes con MEGAHIT	47
Cuadro 8. Resultados de MetaQUAST obtenido a partir de los ensamblajes con MetaSPAdes	48
Cuadro 9. Resumen de los MAGs encontrados en los tapetes microbianos.	51
Cuadro 10. Resultados generales de la anotación de los metagenomas utilizando Prokka.	66
Cuadro 11. Número total de variantes/alelos en cada metagenoma según la clasificación de las enzimas (bajo la categoría EC3).	68
Cuadro 12. Resultados de UniProt-Blast para las variantes de metaloproteínas más abundantes de las variantes ftsH.....	81
Cuadro 13. Resultados de Uniprot-Blast para las variantes de metaloproteínas más abundantes de las variantes de los genes mmpA y rip.....	85
Cuadro 14. Resultados de UniProt-Blast de las variantes de metaloproteínas más abundantes de los genes pmA.	87
Cuadro 15. Resultados de UniProt-Blast para las variantes de metaloproteínas .	89
Cuadro 16. Resultados de UniProt-Blast para las variantes de lipasas	96
Cuadro 17. Resultados de Uniprot-Blast para las variantes de esterases	102
Cuadro 18. Resultados de UniProt-Blast para las variantes de amilasas.....	104
Cuadro 19. Genes que participan en la degradación de atrazina encontrados en las fuentes termales.	126

Lista de figuras

Figura 1. Descripción general del análisis bioinformático para el perfil microbiano del gen 16S ARNr y la metagenómica shotgun _____	10
Figura 2. Clasificación de peptidasas según su especificidad y tipo de catálisis que realizan _____	19
Figura 3. Distribución de peptidasas encontradas en bacterias y/o arqueas en clanes y familias según MEROPS _____	20
Figura 4. Árbol filogenético de las lipasas y esterases clasificadas en 19 subfamilias _____	221
Figura 5. Fotografías de fuentes termales costarricenses _____	33
Figura 6. Fotografías de fuentes termales de Yellowstone _____	34
Figura 7. Diagrama general de la metodología de análisis bioinformático _____	38
Figura 8. Cantidad de MAGs aislados de tapetes microbianos de seis fuentes termales según el filo al que pertenecen _____	50
Figura 9. Diagrama de calor (Primer 7) que muestra las relaciones globales entre grupos de muestras _____	58
Figura 10. Ordenación de escala multidimensional no métrica (NMDS) de las comunidades bacterianas a nivel taxonómico de filo de las fuentes termales _____	58
Figura 11. Mapa de calor de la abundancia estandarizada de cada una de las variantes encontradas en los tapetes microbianos _____	71
Figura 12. Abundancia de las enzimas clasificadas como serina proteasa en las fuentes termales. _____	75
Figura 13. Alineamiento de secuencias clasificadas como serina proteasa encontradas en los tapetes microbianos de las fuentes termales costarricenses _____	77
Figura 14. Abundancia de las diferentes variantes probablemente asociadas a metalopeptidasas en los tapetes microbianos de las fuentes termales costarricenses _____	80
Figura 15. Alineamiento de secuencias clasificadas como FtsH encontradas en los tapetes microbianos de las fuentes termales costarricenses _____	84
Figura 16. Árbol filogenético de las secuencias de lipasas y esterases encontradas en los tapetes microbianos. _____	93

Figura 17. G ráfico de sombras de la abundancia relativa de cada uno de los genes involucrados en la tolerancia a metales encontrados en los tapetes microbianos.	107
Figura 18. Representación de los operones arsRDABC y aioXSRBA en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio.	109
Figura 19. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al arsénico	110
Figura 20. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al cobre	112
Figura 21. Representación de los operones copBAC y copRMS en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio	115
Figura 22. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al plomo	118
Figura 23. Representación de los operones chrBACF y adeABC en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio.	124
Figura 24. Resultados de correlación de Spearman entre los genes relacionados con la remoción/degradación de paraquat y atrazina	127

Cuadro de abreviaciones

ARNr	Ácido Ribonucleico
arCOG	Archaeal clusters of orthologous genes
ATP	Adenosín Trifosfato
ADN	Ácido Desoxirribonucleico
CAzy	<i>The Carbohydrate-Active EnZymes database</i>
COG	Clúster de Grupos Ortólogos
CDS	Región codificante
CDD	<i>Conserved Domains Database</i>
dBg	<i>Bruijn graph</i>
dNTPs	Desoxirribonucleótidos trifosfatos
FISH	Hibridación in situ de fluorescencia
GO	<i>Gene Ontology</i>
HMM	Modelos ocultos de Markov
ITS	Regiones Espaciadoras Transcritas
KEGG	Enciclopedia de Kyoto de Genes y Genomas
KO	<i>KEGG Orthology</i>
LCA	<i>Last Common Ancestor</i>
MAG	Genomas obtenidos a partir de metagenomas
MDR	<i>MultiDrug Resistance</i>
MATE	<i>Multidrug And Toxin Extrusion</i>
SMR	<i>Small Multidrug Resistance</i>
MPs	Metalopeptidasas
MGRast	<i>Metagenomics Rapid Annotation using Subsystem Technology</i>
MFS	<i>The major facilitator superfamily</i>
NMDS	<i>Non-metric multidimensional scaling</i>
NGS	Secuenciación de Nueva Generación
Pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
PNY	Parque Nacional Yellowstone
Pfam	<i>protein families database</i>
PRMER7	<i>Plymouth Routines in Multivariate Ecological Research</i>
RAM	<i>Random Access Memory</i>
RND	<i>Resistance-nodulation-division</i>
SBS	<i>Sequencing by synthesis</i>
SRA	<i>Sequence Read Archive</i>
SRH	<i>Second Region of Homology</i>
SIMPROF	<i>The similarity profile routine</i>
SIMPER	<i>Similarity percentage análisis</i>
UNIREF	<i>UniProt Reference Clusters</i>
UniProt	<i>Universal Protein database</i>



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, _____, con cédula de identidad _____, en mi condición de autor del TFG titulado _____

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. **SI** **NO** *

***En caso de la negativa favor indicar el tiempo de restricción: _____ año (s).**

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

Introducción

En el pasado y el presente, la búsqueda de vida requiere de una sólida comprensión del origen y evolución de los organismos en el planeta Tierra. Los procariotas como las arqueas y bacterias son posiblemente los únicos organismos que fueron capaces de tolerar las condiciones ambientales y atmosféricas de la tierra de hace 3.8 mil millones de años, tales como falta de oxígeno, alta concentración de gases (metano, dióxido de carbono, nitrógeno, hidrogeno y monóxido de carbono), alta radiación y elevadas temperaturas (Javaux 2006), y persistir hasta el día de hoy. Los ambientes extremos actuales son una representación de algunas de las condiciones de la Tierra hace millones de años, por lo que estudiar los microorganismos que allí se encuentran resulta de gran interés. Un organismo que es capaz de sobrevivir en un ambiente extremo se considera extremófilo. Los extremófilos pueden tolerar condiciones extremas físicas como la temperatura, la radiación, la presión y concentración de nutrientes, y extremos geoquímicos como por ejemplo la desecación, salinidad, pH, especies de oxígeno o potencial redox (Rothschild y Mancinelli, 2001). Según Kay (1980) un organismo extremófilo debe ser capaz de desarrollar un mecanismo para limitar, reducir o eliminar los efectos ambientales en su metabolismo o simplemente desarrollar estrategias funcionales para “vivir” con ese factor.

Los estudios de ambientes extremos, como lo son las fuentes termales, han generado mucho interés en la comunidad científica desde 1960, con las investigaciones en el Parque Nacional Yellowstone realizadas por el microbiólogo Thomas Brock (Brock 2001), y los análisis elementales en este sitio realizados por Allen and Day en 1935 y Schoen en 1969 (Brock y Mosser, 1975), los cuales se consideran pioneros. La caracterización microbiológica, geológica y química de las fuentes termales ha sido objeto de estudio alrededor del mundo. La temperatura es uno de los principales factores ambientales que controlan las funciones de los organismos, reproducción y el crecimiento (Kay, 1980). Las fuentes termales se caracterizan por tener temperaturas elevadas (45°C-100°C), pueden ser alcalinas

o ácidas, y generalmente tienen una alta concentración de iones (Rothschild y Mancinelli, 2001, Djokic et al., 2017).

La metagenómica ha permitido realizar estudios de comunidades de ambientes extremos a detalle, no solo a nivel taxonómico sino también a nivel metabólico y funcional (López-López et al., 2013, Cowan et al., 2015, van der Walt et al., 2017). Los estudios para la identificación taxonómica son un paso fundamental para describir la distribución de los miembros de la comunidad, ya que brindan información acerca de la abundancia de los diferentes microorganismos (Jiang y Takacs-Vesbach, 2017, Thiel et al., 2016). La secuenciación de alto rendimiento o también llamada Secuenciación de Nueva Generación (NGS) ha permitido ampliar los estudios de comunidades microbianas mediante la secuenciación de marcadores filogenéticos como los genes para el ARNr 16S y las regiones ITS (regiones espaciadoras transcritas), lo cual permite realizar estudios de estructura. Por otro lado, la metagenómica permite realizar una caracterización de sus potencialidades a nivel funcional (Boteva y Kambourova 2018). Esta herramienta ha generado grandes avances en el estudio de las características metabólicas, fisiológicas y ecológicas de las comunidades termotolerantes (Hiraoka et al., 2016 y Saxena et al., 2017). Las secuencias de los metagenomas ensamblados de las comunidades microbianas han revelado la existencia de genes que le confieren a diferentes organismos la capacidad adaptarse o tolerar condiciones extremas. Más específicamente para esta investigación, se han realizado estudios para la búsqueda y caracterización de genes que codifican para enzimas hidrolíticas termoestables (Saxena et al., 2017, Sahay et al., 2017), las cuales juegan un rol importante dentro del metabolismo de los microorganismos. Además, se han realizado investigaciones sobre la dinámica y vías metabólicas, procesos celulares y biogeoquímicos de los microorganismos que conforman las comunidades bacterianas y de arqueas (Mehetre et al., 2016a, Mehetre et al., 2018). Todas estas investigaciones ofrecen una amplia información sobre posibles mecanismos o herramientas que pueden utilizarse en procesos industriales o biotecnológicos, y que pueden servir para solucionar problemas ambientales y de salud pública como la contaminación (Saxena et al., 2017, Bao et al., 2017).

La contaminación ambiental por compuestos tóxicos es una gran amenaza para el medio ambiente y la salud humana a nivel mundial (Singh y Naidu, 2012), por lo tanto, es de gran importancia conocer las vías metabólicas que tienen los microorganismos relacionados con la eliminación de sustancias contaminantes. El estudio e identificación de las vías metabólicas involucradas en la degradación de xenobióticos por parte de microorganismos podría tener un gran impacto en la biorremediación de compuestos contaminantes (Bao et al., 2017). En este sentido, los microorganismos aislados de ambientes extremos como las fuentes termales son buenos candidatos para la remoción de ciertos compuestos contaminantes, ya que la mayoría poseen diferentes estrategias metabólicas para remover y/o tolerar sustancias tóxicas, que resultan importantes para la biorremediación (Segretin et al., 2016). Cabe destacar las investigaciones que involucran bacterias termófilas en la remoción de metales pesados como hierro, manganeso, cromo y plomo (Vargas et al., 1998, Kashefi y Lovley 2000, Masaki et al., 2015).

La búsqueda de enzimas con uso potencial en procesos de biorremediación en estos ambientes ha sido discutida previamente en la literatura, debido a la estabilidad de estas, inherente aún en condiciones de temperatura y pH extremos (Demirjian et al., 2001, de Miguel Bouzas et al., 2006), y en presencia de algunos solventes y detergentes (Bhalla et al., 2013). Los avances en el estudio de estas enzimas han sido posibles gracias al aislamiento de microorganismos termófilos de diferentes ambientes y su posterior extracción y purificación (Burrows, 1973, Antranikian et al., 1987, Groboillot, 1994, Bhushan y Hoondal 1998, Bauer et al., 1999, Kohilu et al., 2001).

En Costa Rica se han realizado estudios de las comunidades microbianas y de organismos eucariotas de fuentes termales que se caracterizan por tener temperaturas desde los 35 °C, rangos de pH de 3 a 9, y en su mayoría poseen una alta concentración de iones (Sittenfeld et al., 2002, Sittenfeld et al., 2004, Fisinger et al., 2008, Morales et al., 2008, Caldwell et al., 2010, Hernández-Ascencio, 2012, Hynek et al., 2018, Arce-Rodríguez et al., 2019, Uribe-Lorío et al., 2019). El objetivo principal de estos proyectos ha sido el aislamiento, identificación molecular y análisis de estructura de las comunidades de bacterias, arqueas y eucariotas presentes. Muchas de las fuentes termales estudiadas poseen orígenes

geotectónicos distintos debido a su asociación con volcanes y/o fallas tectónicas (Alvarado y Vargas, 2017), aunque también difieren en el grado de mineralización, y en las concentraciones iones como hidrocarbonato, sulfuro, sulfato, cloruro, nitrato, amonio, entre otros (Alvarado y Vargas, 2007, Alvarado y Vargas, 2017, Uribe-Lorío et al., 2019).

Según la información consultada en artículos científicos y libros disponibles en internet en el presente año tanto en inglés como en español (utilizando palabras clave como “metagenómica”, “Costa Rica”, “enzimas”, entre otras) no se ha llevado a cabo hasta el momento investigaciones sobre la búsqueda de enzimas hidrolíticas en estos ambientes de nuestro país y en la región centroamericana utilizando secuenciación masiva paralela, por lo que resulta de un estudio novedoso. La identificación de secuencias de enzimas encontradas en ambientes con altas temperaturas es muy importante para generar información en ciencia básica y biotecnología, ya que son recursos que todavía no han sido descritos, y que por ende aumentan nuestro conocimiento de las comunidades microbianas y las características metabólicas que poseen estos organismos.

Es por lo que en este proyecto se plantea la utilización de metagenómica para i) caracterizar parcialmente a nivel metabólico y funcional las comunidades de microorganismos de tres fuentes termales de Costa Rica, y ii) buscar en ellas secuencias de enzimas hidrolíticas que están posiblemente relacionadas con el metabolismo, degradación y remoción de xenobióticos como metales pesados y agroquímicos.

Dentro de los principales resultados encontrados destacan que las comunidades microbianas de los seis metagenomas de tapetes microbianos de fuentes termales de Costa Rica son similares a otras alrededor del mundo, sin embargo incluyen algunos microorganismos únicos. Se identificaron secuencias completas de proteasas, lipasas, amilasas, esterases e hidrolasas, categorías que según diversos estudios incluyen enzimas capaces de hidrolizar compuestos como los agroquímicos. La mayoría de las variantes/alelos encontrados son similares (80% -95% identidad) a las de microorganismos que toleran altas temperaturas, por lo que estas enzimas muy probablemente sean termotolerantes, sin embargo, en la mayoría de los casos, el porcentaje de identidad es menor al 95%, lo que sugiere

que estas comunidades constituyen un novedoso reservorio de potenciales enzimas de posible aplicación en el largo plazo en la degradación o remoción de xenobióticos como agroquímicos. La búsqueda de genes que posiblemente participan en la remoción de metales pesados en los 84 MAGs (Genomas obtenidos a partir de metagenomas) reconstruïdos permiti3 asociar los mecanismos de remoci3n tanto de xenobi3ticos como de metales pesados con los diferentes grupos de microorganismos presentes.

Antecedentes

Caracterización de comunidades de bacterias y arqueas en fuentes termales

Los microorganismos son parte fundamental de la dinámica de los ecosistemas del planeta, sin embargo, a pesar de su importancia se desconocen muchas de sus funciones ecológicas y papeles dentro de una comunidad (Konopka et al., 2015, Saini et al., 2017). Las comunidades microbianas están típicamente compuestas por especies dominantes y algunos grupos taxonómicos poco comunes o desconocidos (Sogin et al., 2006). Estos organismos poco abundantes pueden tener importantes roles dentro de los ciclos biogeoquímicos y los flujos metabólicos del carbono, nitrógeno, azufre y fósforo (Musat et al., 2008, Wrighton et al., 2012).

Las comunidades microbianas de ambientes extremos, específicamente los consorcios termofílicos, han sido ampliamente estudiadas y caracterizadas en algunas regiones del mundo (Amin et al., 2017, Saini et al., 2017, Oliverio et al., 2018), como Canadá, Nueva Zelanda, Estados Unidos, Japón, India y Malasia (Saini et al., 2010). Según Dick y Shock (2013) la estructura microbiana de las fuentes termales depende de gradientes de temperatura y pH, así como de interacciones ecológicas, la química de las rocas y las concentraciones de compuestos inorgánicos (Dick y Shock, 2013). Se ha encontrado que las comunidades microbianas de fuentes termales a lo largo de Costa Rica con temperaturas entre los 37 °C y 63 °C se rigen por las variables fisicoquímicas, tales como pH, temperatura, concentración de gases, cationes y aniones (Uribe-Lorío et al., 2019), por lo que la estructura microbiana al igual que en otras fuentes alrededor del mundo, va a depender principalmente de las características ambientales.

Dentro de los grupos de microorganismos más abundantes en estos ambientes se encuentran las bacterias y arqueas (Bowen De León., 2013, Huang et al., 2013). Las fuentes termales poseen condiciones muy específicas que pueden favorecer el crecimiento de bacterias de los filos Acidobacteria, Actinobacteria y Proteobacteria, así como de otros filos que se encuentran en baja abundancia. En fuentes con temperaturas menores a los 70 °C, filos como Cyanobacteria y Chloroflexi están entre los principales grupos de bacterias autótrofas que se pueden encontrar (Miller

y Castenholz, 2000), siendo Chloroflexi, un conjunto de microorganismos fotoheterótrofos capaces de tolerar un rango de temperaturas muy amplio (Ward et al., 2018).

Entre las técnicas clásicas que se utilizan para describir y estudiar las comunidades de microorganismos se encuentran los microarreglos y análisis de *Hibridación in situ* de fluorescencia (FISH), reacción en cadena de la polimerasa (PCR), así como técnicas caídas en desuso como clonaje de ADN en fósidos y cósmidos (Kublanov et al., 2009, Kumar et al., 2015), y que han sido progresivamente sustituidas por la secuenciación masiva de un fragmento de ADN como lo es el gen que codifica el ARNr 16S, y la secuenciación *shotgun*. Las técnicas de NGS permiten generar información de la estructura de la comunidad microbiana mediante la amplificación masiva de un marcador filogenético, así como características metabólicas y funcionales de las comunidades microbianas (Sharpton et al., 2014).

Según Lozupone y Knight (2006), en un estudio en donde se comparan 202 muestras de diferentes ambientes existe una amplia variación filogenética y de diversidad de especies, siendo los sedimentos y los suelos los más diversos filogenéticamente. En el caso específico de las fuentes termales, la mayoría de las investigaciones señalan que el aumento de la temperatura afecta la diversidad de microorganismos (Tank et al., 2017, Pagaling et al., 2012, Ross et al., 2012), siendo las fuentes con temperaturas mayores a los 70 °C las menos diversas. No obstante, se ha demostrado que sitios calientes como las ventilas térmicas y las aguas calientes terrestres son fuente de alelos nuevos (Zhang et al., 2008, Auguet et al., 2010, Kumar et al., 2017, Campbell et al., 2017).

Metagenómica *shotgun* para el estudio de comunidades microbianas

El término metagenómica fue utilizado por primera vez por Jo Handelsman (Handelsman et al., 1998) y se refiere al análisis genómico de ADN extraído directamente a partir de una muestra ambiental. Los enfoques de metagenómica históricos consisten en el clonaje y la secuenciación por el método de Sanger de librerías de plásmidos y fósidos (Ghai et al., 2010). No obstante, la

implementación de estos métodos consume más tiempo y requiere una concentración mayor de ADN, además de que son técnicas económicamente más costosas. La secuenciación de alto rendimiento permite examinar miles de organismos en paralelo y obtener de forma exhaustiva la mayoría de los genes, proporcionando así información de la biodiversidad y las características funcionales de las comunidades microbianas (Ranjan et al., 2016, Tessler et al., 2017). Con los avances NGS se desarrolló una visión mucho más detallada de las comunidades microbianas, utilizando secuenciación de amplicones para la identificación taxonómica de las comunidades y la secuenciación *shotgun*, la cual permite caracterizar a nivel funcional y metabólico los microorganismos (Figura 1). El término secuenciación *shotgun* fue acuñado en 1979 por Rodger Staden quien propuso que la secuenciación *shotgun*, era en la que se utilizaban vectores bacterianos para clonar fragmentos aleatorios de una molécula de ADN larga, Luego, los fragmentos se secuencian en paralelo y las lecturas se ensamblan utilizando sus superposiciones, lo que permite secuenciar genomas más grandes (Staden 1979), posteriormente, propone en 1980 el término *contig*, para describir los datos obtenidos después del ensamblaje de las lecturas de secuenciación *shotgun*. En el contexto de la publicación “un contig es un conjunto de lecturas de gel que se relacionan entre sí por superposición de sus secuencias” (Staden, 1980). Estos conceptos aplican desde entonces para diferentes áreas de la biología molecular, tipos de secuenciación y análisis bioinformáticos (Giani et al., 2019).

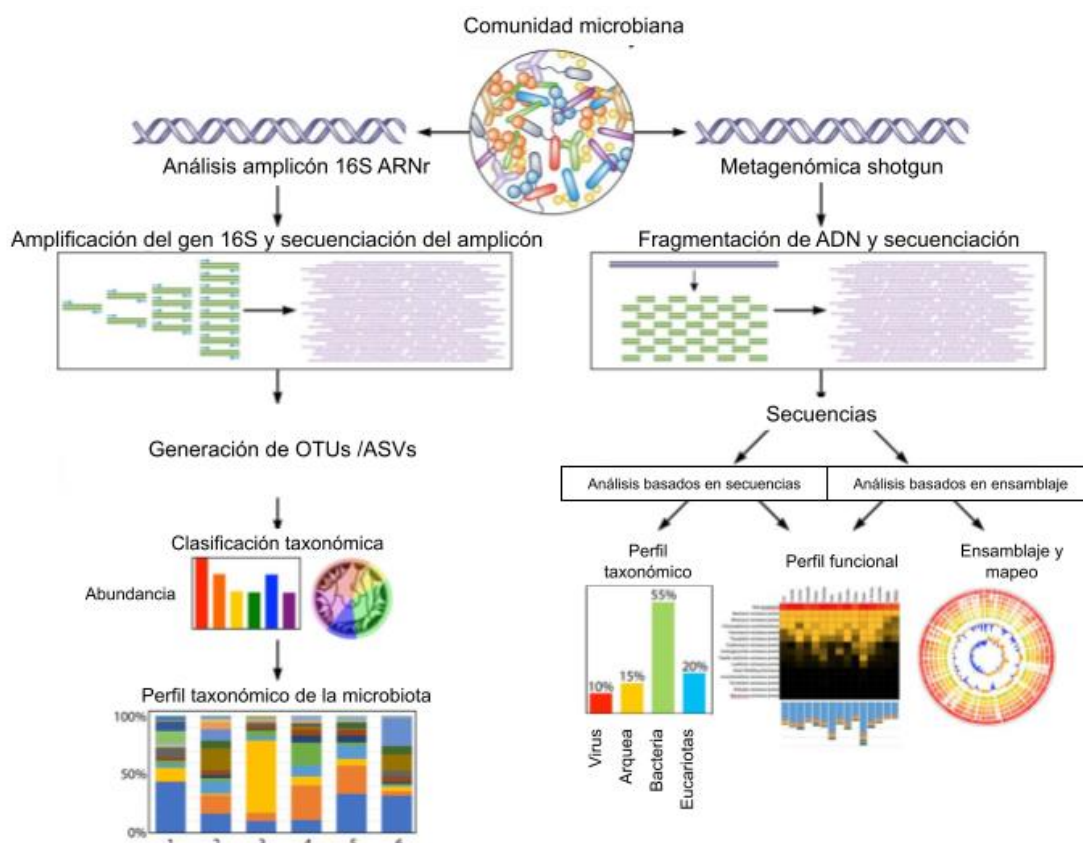


Figura 1. Descripción general del análisis bioinformático para el perfil microbiano del gen 16S ARNr y la metagenómica shotgun. Modificado de Milani et al., 2017.

Dentro de las principales plataformas para secuenciar metagenomas se encuentra la de la casa comercial Illumina Inc, la cual ofrece algunas opciones para la preparación de librerías, la secuenciación y el análisis e interpretación de los datos. Esta genera una gran cantidad de información y alta precisión (tasa de error de 0.1-1%) en comparación con otras plataformas, como Ion Torrent (~ 1 Gb, 1,5%) y Pacific Biosciences (100 Mb, 10-15% tasa de error) (Quail et al., 2012, Buermans y den Dunnen, 2014, Laehnemann et al., 2016, Roumpeka et al., 2017, Driscoll et al., 2017). Es importante considerar la química que se utiliza y el equipo, ya que estas características cambian según las combinaciones.

La tecnología de Illumina, que es la más utilizada, es capaz de generar secuencias cortas de entre 100 pb, 300 pb, o más, y miles de millones de secuencias para una única muestra (Breitwieser et al., 2017, van der Walt et al., 2017). Esta plataforma utiliza el método de Secuenciación por Síntesis (SBS). En este tipo de técnica se le adiciona adaptadores en los extremos a los fragmentos de ADN por secuenciar,

que consisten en secuencias cortas de nucleótidos que permiten la unión por complementariedad a la celda de flujo, y así poder ser enriquecidos mediante la amplificación tipo puente. Los adaptadores contienen: i) secuencias que permiten que la biblioteca se una y genere *cluster* en el *flow cell* (secuencias P5 y P7), ii) secuencias de sitios de unión del primer para iniciar la secuenciación (Rd1 SP y Rd2 SP) y iii) *index* (*index* 1 y, cuando corresponda, *index* 2), que son identificadores de muestra que permiten la multiplexación / agrupación de múltiples muestras en una única secuenciación. La formación de estos grupos o *clusters* permite amplificar la señal generada por nucleótidos fluorescentes incorporados en la cadena creciente, que es detectada por una cámara durante un proceso cíclico (Roumpeka et al., 2017). Los 2'-desoxinucleósido-trifosfatos o dNTPs se incorporan a la cadena de ADN en crecimiento, donde al menos uno de los cuatro nucleótidos (A, C, T, G) tienen una etiqueta fluorescente diferente que sirve para identificar la base y actuar como terminador reversible, esto con el fin de evitar múltiples eventos de extensión.

En circunstancias ideales, todas las bases dentro de un grupo se amplificarán en la misma fase, por lo que los nucleótidos deben agregarse en las cadenas de cada clúster al mismo tiempo, de lo contrario la calidad de la señal irá disminuyendo hacia los extremos de las lecturas (Buermans y Den Dunnen. 2014). La precisión en cada fase ayudará a disminuir la cantidad de falsos positivos y falsos negativos. La señal de fluorescencia es detectada por una cámara, lo cual permite identificar los nucleótidos que conforman cada una de las secuencias (Roumpeka et al., 2017). Como resultado final se pueden obtener secuencias *single end* (un único extremo, hacia una única dirección) o *paired end* (extremos pareados, ambas direcciones) según el *kit* utilizado.

Con la secuenciación *single end* el inserto de la biblioteca se lee solo desde el extremo P5. Con la secuenciación *paired end*, se genera una segunda lectura desde el otro extremo del inserto (extremo P7). La secuenciación *paired end* incrementa la cobertura de secuenciación, este tipo de estrategia se utiliza generalmente para la secuenciación del genoma o del exoma completo, y en algunos casos de secuenciación del transcriptoma o metagenómica. La secuenciación de un solo extremo es el método de elección para la secuenciación

de ARN, por ejemplo, para el análisis diferencial de la expresión génica. Los extremos pareados ofrecen datos de secuencias de alta calidad debido a que se conoce la distancia entre cada lectura, mejora la alineación y el ensamblaje, además de que permiten la detección precisa de variantes estructurales, fusiones de genes e isoformas. Por otro lado, la utilización de secuencias hacia una única dirección permite generar grandes volúmenes de datos de alta calidad, de forma rápida y económica.

Por otro lado, la minería de alelos/genes es uno de estos enfoques novedosos que permite identificar alelos nuevos de organismos diferentes, algunas estrategias que se utilizaron en este estudio son el mapeo y el alineamiento. El mapeo consiste en alinear secuencias de nucleótidos cortas (100 pb – 300 pb) contra secuencias de nucleótidos largas (por ejemplo, *contigs*), esto ayuda a determinar el origen de la secuencia. En el alineamiento se tienen secuencias de nucleótidos o proteínas que se quiere alinear entre sí, lo que permite identificar similitudes y/o diferencias entre estas.

Ensamblaje y anotación de (meta)genomas

El ensamblaje de los (meta)genomas es un paso importante para caracterizar la ecología y la fisiología de los microorganismos ambientales. Se han desarrollado una serie de algoritmos y herramientas que permiten generar secuencias de mayor tamaño llamados *contigs* a partir de secuencias cortas (100 pb, 300 pb o más). Estos *contigs* pueden unirse en secuencias más largas y generar *scaffolds*, que se espera contengan al menos secuencias completas de genes. La mayoría de los algoritmos trabajan con *k-mers*, que son secuencias de oligonucleótidos de tamaño k , también llamados *l-tuples* o *n-grams* (Dubinkina et al., 2016). Según estos autores, este método es más simple y rápido que el análisis basado en referencias y alineamientos, ya que los *k-mers* son la representación comprimida de las secuencias.

Existen al menos tres enfoques para el ensamblaje de secuencias: el paradigma de Greedy, la superposición (OLC) y los grafos de Bruijn, siendo este último el más utilizado al considerarse mejor para ensamblar secuencias cortas (< 100 pb) (Li et

al., 2012, Nagarajan y Pop 2013). Los algoritmos OLC y grafos de Bruijn utilizan grafos, que son un conjunto de objetos llamados vértices (o nodos) unidos por aristas (*edges*) que pueden representar relaciones binarias entre elementos de un conjunto. Típicamente, un grafo se representa mediante una serie de puntos (los vértices) conectados por líneas (las aristas). Estos permiten generar posibles soluciones al ensamblaje mediante conexiones de tal manera que las secuencias se vinculan para formar secuencias contiguas (*contigs*). El enfoque OLC fue muy exitoso para ensamblar secuencias generadas a partir de la tecnología de Sanger (> 200 pb). Algunos de los ensambladores que utilizan este algoritmo son CAP3 (Huang y Madan, 1999), Celera (Myers et al., 2000), Arachne (Batzoglou et al., 2002), PCAP (Huang y Yang, 2005) y Newbler (Margulies et al., 2005). La mayoría de los ensambladores mencionados están en desuso.

Los grafos de Bruijn es un algoritmo propuesto por Ramana y Waterman en 1995 (Idury y Waterman, 1995), el cual utiliza *k-mers* para inferir la secuencia de ADN ensamblada. Existen dos tipos de métodos que utilizan los grafos de Bruijn, los sencillos en los que se utiliza un solo tamaño de *k-mer* como solución del ensamblaje y los múltiples, los cuales combinan los resultados de diferente tamaño de *k-mers* para generar el ensamblaje, lo que mejora el resultado de este (Pell et al., 2012). El primer ensamblador que usó este algoritmo fue EULER, en el año 2001 (Pevzner y Waterman, 2001), y actualmente la mayoría de los ensambladores utilizados en análisis de metagenomas utilizan modificaciones de este, tales como la eliminación de lecturas erróneas para múltiples valores de *k-mers* o alineamiento de otras secuencias, corrección de llamado de bases erróneas basado en la frecuencia de *k-mer* o el valor de calidad de cada base, partición del grafo, detección de *contigs* mal ensamblados, entre otros (Miller et al., 2010, Vollmers et al., 2017, Sczyrba et al., 2017, van der Walt et al., 2017, Forouzan et al., 2018). En el Cuadro I se mencionan algunas de las características de los ensambladores Metavelvet (Namiki et al., 2012), MetaSPAdes (Nurk et al., 2017) y MEGAHIT (Li et al., 2015), todos estos ensambladores de *novο*, o sea que no utilizan un genoma de referencia como molde para ensamblar los fragmentos de ADN y generar secuencias más largas. Debido a que no se utiliza un genoma de referencia con el cual se pueda comparar, el rendimiento del ensamblaje de *novο* debe considerar

qué tanto del total de secuencias se incluye dentro en el ensamblaje y qué tanta información del genoma se puede encontrar (Baker 2015).

MEGAHIT es actualmente el ensamblador de *novu* más eficiente en términos de memoria, ya que para ensamblar metagenomas se necesita menor uso de memoria RAM (Memoria de Acceso Aleatorio) (Olson *et al.*, 2017). La memoria se reduce al eliminar *k-mers* por debajo de un umbral de frecuencia definido. Este enfoque minimiza el impacto negativo de los errores de secuencia en el ensamblaje. Para conservar los *k-mers* de organismos de baja abundancia, distinguiéndolos de los errores, MEGAHIT reconsidera a los *k-mers* eliminados en las regiones de baja cobertura en el ensamblaje.

MetaSPAdes es una versión específica para metagenomas del ensamblador SPAdes (Nurk *et al.*, 2017). Una de las principales innovaciones de este ensamblador es el uso de información de extremos pareados durante el proceso de ensamblaje en lugar de hacerlo posteriormente. Esta información se incorpora en el gráfico utilizando un par de *k-mers* separados por una distancia estimada (Olson *et al.*, 2017). En MetaSPAdes las micro-variaciones entre posibles cepas muy similares se combinan para formar secuencias de consenso de alta calidad, lo que genera una mejor representación de cada “especie”.

Existen algunos factores que afectan directamente el desempeño de los ensambladores que utilizan grafos de Bruijn, como los errores de secuenciación, las regiones repetidas, la cobertura de secuenciación y la secuenciación de cepas muy relacionadas genéticamente. Las regiones repetidas de ADN pueden encontrarse en el mismo organismo o estar compartidas entre distintos organismos. Las repeticiones crean un mayor número de conexiones (*edges*) incrementando el número de posibles soluciones, lo que ocasiona ambigüedad en el ensamblaje.

Por otro lado, el proceso de anotación consiste en la delimitación de características relevantes en una secuencia del genoma o metagenoma (Richardson y Watson, 2012). Dentro de las herramientas que existen para la predicción de genes bacterianos y de arqueas se encuentran los paquetes Glimmer (Delcher *et al.*, 2007), GenemarkHMM (Lukashin y Borodovsky, 1998), Easygene (Larsen y Krogh, 2003) y Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm, Hyatt *et al.*, 2010). Este último algoritmo ha sido ampliamente utilizado en estudios

de metagenomas (Jungbluth et al., 2017, Tamames y Puente-Sánchez 2018, Tully et al., 2018). En 2018, se desarrolló METAWRAP el cual incluye numerosos módulos para el análisis de MAGs, incluida la asignación de taxonomía, estimación de abundancia, anotación funcional y visualización de datos de secuenciación tipo *shotgun* (Uritskiy et al., 2018).

Existen otros algoritmos de predicción de genes *ab initio* los cuales utilizan propiedades de las regiones codificantes como características y una amplia variedad de técnicas matemáticas como redes neuronales para encontrar las regiones codificantes (Do y Choi 2006). Por otro lado, algunos de los algoritmos de anotación más implementados utilizan un conjunto de secuencias de referencia para entrenar un modelo, y luego utiliza ese modelo para predecir las regiones de codificación del genoma de interés. Algunos de estos algoritmos se han implementado en programas de predicción de genes de código abierto, como Prodigal, el cual identifica CDS (región codificante), sin asignar alguna función.

Para la predicción de los marcos abiertos de lectura (ORF, por sus siglas en inglés) y de ARNs, la herramienta Prodigal es entrenada previamente con secuencias. En su modo básico este programa toma las secuencias que se le proporcionan, las estudia, aprende sus propiedades y luego es capaz de predecir los genes basándose en esas propiedades. Dentro de las propiedades necesarias se encuentran los codones de iniciación (ATG, GTG, TTG), el sitio de unión ribosomal (RBS), contenido de GC, estadísticas de codificación de hexámeros y otra información necesaria para construir un perfil de entrenamiento completo (Hyatt et al., 2010). El algoritmo tiene que determinar automáticamente un conjunto de genes "reales" putativos sobre los cuales entrenarse. Por otro lado, existe el modo anónimo, en donde se utiliza un archivo de entrenamiento precalculado, y predice los genes basándose en los mejores resultados. También existe el modo de entrenamiento en donde se aplica el modo normal, pero Prodigal guarda la información de entrenamiento para usarla en futuros análisis. Todas esas variantes lo que buscan es mejorar la predicción, el reconocimiento del sitio del inicio de la transducción y la reducción de falsos positivos.

Las proteínas están generalmente compuestas por una o más regiones funcionales, comúnmente denominadas dominios. Las diferentes combinaciones de dominios

dan lugar a la diversidad de proteínas que se encuentran en la naturaleza. Por lo tanto, la identificación de los dominios dentro de las proteínas puede proporcionar información sobre su función. Posterior a la identificación de genes, se debe realizar la búsqueda de homología entre las secuencias y las bases de datos que sirven para los análisis de funciones como Pfam (colección de secuencias de familias de proteínas, Finn et al., 2016), COG (Clúster de Grupos Ortólogos, Tatusov et al., 2000) y KEGG (Enciclopedia de Kyoto de Genes y Genomas). Esta búsqueda se puede realizar utilizando la herramienta DIAMOND (Buchfink et al., 2015) o BLASTX (Altschul et al., 1990), lo que permite posteriormente asignar una clasificación funcional a cada gen y buscar variantes alélicas, que resulta de utilidad para la descripción de enzimas y vías metabólicas.

La base de datos Pfam es una gran colección de dominios y familias de proteínas, cada una representada por múltiples alineaciones de secuencias y modelos ocultos de Markov (HMM) (El-Gebali et al., 2019). La versión 34.0 (marzo 2021) que es la más reciente contiene un total de 19179 familias proteicas las cuales se basan en datos de secuencias de UniProtKB, NCBI y en secuencias de proyectos seleccionados de metagenómica (Mistry et al., 2020). Esta base tiene 420933 secuencias de 199 arqueas, 20190161 secuencias de 5337 bacterias, 20073593 secuencias de 1009 eucariotas y 370423 secuencias de 7459 virus. Por otro lado, la base COG, es un intento de clasificación filogenética de las proteínas codificadas en genomas completos de bacterias, arqueas y eucariotas (Tatusov et al., 1997, Galperin et al., 2015, Makarova et al., 2015). Se basa en una comparación exhaustiva de todas las secuencias de proteínas, lo que ha generado una asignación de genes ortólogos (misma función en especies diferentes) y parálogos (homólogos en una misma especie, función distinta). Además, incluye una cuidadosa curación manual de los COGs, lo que evita errores de anotación y sobrepredicción (Galperin et al., 2015). La base de datos COG se creó en 1997 y la actualización más reciente es de enero del 2021. La última versión, incluye genomas completos de 1187 bacterias y 122 arqueas, y 4877 COG, para un total de 3236575 genes únicos mapeados en 3455867 loci genómicos, lo que representa el 73,5% de las 4401819 proteínas. Los planes futuros incluyen una mayor expansión de la colección COG agregando COG de arqueas (arCOG), dividiendo

los COG que contienen múltiples parálogos y el refinamiento continuo de las anotaciones COG (Galperin et al., 2021).

Por otra parte, la base de datos KEGG es utilizada como referencia para la interpretación biológica de secuencias genómicas y otros datos generados a partir de NGS. Esta base se ha desarrollado para comprender la conservación y la variación de genes y genomas (Kanehisa et al., 2019). En particular, el sistema KO (KEGG Orthology) para genes ortólogos funcionales se ha desarrollado para representar características conservadas de genes y proteínas. La base de conocimientos de referencia de los mapas de ruta de KEGG dibujados como redes de nodos KO representa las características conservadas de procesos celulares, lo que permite la reconstrucción automática de las vías metabólicas y los genes involucrados. KEGG se ha convertido, según Kanehisa y colaboradores (2019), en un recurso ampliamente utilizado para la interpretación biológica de diferentes tipos de datos como genomas, transcriptomas, metabolomas y metagenomas. La implementación de todas las herramientas de análisis de metagenomas arriba mencionadas han permitido generar mucha información acerca de la diversidad y metabolismo de los microorganismos en diferentes ambientes.

Dentro de los principales estudios de metagenómica funcional se encuentra el análisis de 20 metagenomas de sitios calientes y ácidos del Parque Nacional Yellowstone (Inskeep et al., 2013), 7 metagenomas provenientes de fuentes termales de la India (Saxena et al., 2017), además de otros estudios realizados en Taiwán (Lin et al., 2015), India (Mehetre et al., 2016b, Mangrola et al., 2018), Japón (Sato et al., 2017) y España (López-López et al., 2015). Todos estos estudios se han enfocado en la caracterización funcional de las comunidades y la asociación de éstas con variables ambientales en cada sitio, así como a la elucidación de relaciones ecológicas en los consorcios microbianos.

Enzimas hidrolíticas termoestables

Una enzima o proteína se considera termoestable cuando tiene una temperatura de desnaturalización mayor a 55 °C, o una vida media larga a alta temperatura (Turner et al., 2007). Las enzimas hidrolíticas aceleran las reacciones en las que un compuesto se rompe en componentes más simples. La adaptación de

biomoléculas a condiciones extremas está relacionada con la estabilidad y la flexibilidad de optimizar los estados funcionales de las proteínas (Cuadro II). La adaptación de proteínas de organismos extremófilos a ambientes extremos ha sido ampliamente estudiada (Ueno et al., 2016).

Las enzimas termoestables se aíslan principalmente de organismos termófilos y poseen un gran número de aplicaciones debido a su estabilidad general inherente (Demirjian et al., 2001). Los avances en el estudio de estas enzimas han sido posibles gracias al aislamiento de microorganismos termófilos de diferentes ambientes y su posterior extracción y purificación (Burrows, 1973, Antranikian et al., 1987, Groboillot, 1994, Bhushan y Hoondal 1998, Bauer et al., 1999, Kohilu et al., 2001). Los organismos de ambientes termales han demostrado ser fuente de enzimas con actividad hidrolítica como proteasas, lipasas, glicosidasas, amilasas e hidrolasas principalmente (Aanniz et al., 2015, Dalmaso et al., 2015, De Castro et al., 2016, Mohammad et al., 2017, Sahay et al., 2017, Baltazi et al., 2017). Muchas de estas enzimas juegan un rol importante en las comunidades, facilitando la degradación de nutrientes y permitiendo el flujo de energía entre autótrofos y heterotrófos. Es importante mencionar que las enzimas más representativas involucradas en la biorremediación incluyen hidrolasas, esterases, proteasas y lipasas, las cuales han mostrado un potencial para la degradación de diversos compuestos como polímeros, hidrocarburos aromáticos, compuestos halogenados, tintes, detergentes, compuestos agroquímicos (Bhandari, et al., 2021).

Peptidasas

En la literatura existe ambigüedad con respecto a los términos relacionados con las enzimas que degradan péptidos o proteínas. El concepto de "enzima proteolítica", "proteasa", "proteínasa" y "peptidasa" a menudo son utilizados como sinónimos. El término "proteolítico" fue utilizado por primera vez en 1877 por Michael Foster (Rawlings, 2013) y aplicado a la enzima pepsina. La "proteólisis" fue documentada por primera vez en 1880 (Roberts 1880), cuando se realizaron estudios de la digestión en el estómago. Vines acuñó el término "proteasa" en 1903 (Vines 1903), posteriormente "proteínasa" se usó como sinónimo de "proteasa" (Fischer 1907). Finalmente, "peptidasa" se utilizó por primera vez en 1918 (Petersen y Short 1918), y en 1923 se distinguió una peptidasa de una verdadera proteasa: una proteasa degradada proteínas, mientras que una peptidasa hidroliza solo péptidos o "peptonas". Estas distinciones no son aceptables en la actualidad, ya que se consideran términos similares (Cerdá-Costa y Gomis-Rüth, 2014). En este trabajo utilizaremos el término peptidasa para referirnos a las proteasas, proteínasas y peptidasas. Las peptidasas se pueden clasificar según su especificidad en endopeptidasas y exopeptidasas o por el tipo de catálisis que realizan (Fig. 2) También se pueden clasificar en peptidasas extracelulares o intracelulares. Existen clasificaciones por homología y evolución. Una clasificación ampliamente aceptada es la propuesta en MEROPS (https://www.ebi.ac.uk/merops/cgi-bin/clan_index?type=P), la cual clasifica conjuntos homólogos de peptidasas e inhibidores de proteínas jerárquicamente en especies, familias y clanes de proteínas en función de la similitud de secuencia, las distancias evolutivas y el tipo de residuo que utilizan para la catálisis de los sustratos peptídicos (versión más reciente v12.3). Ésta es similar a la encontrada en UniProt (<https://www.uniprot.org/>). Muchas de estas familias de peptidasas se encuentran ampliamente distribuidas en los diferentes filos de bacterias, sin embargo, algunas se pueden encontrar solamente en arqueas o bacterias (Fig. 3).

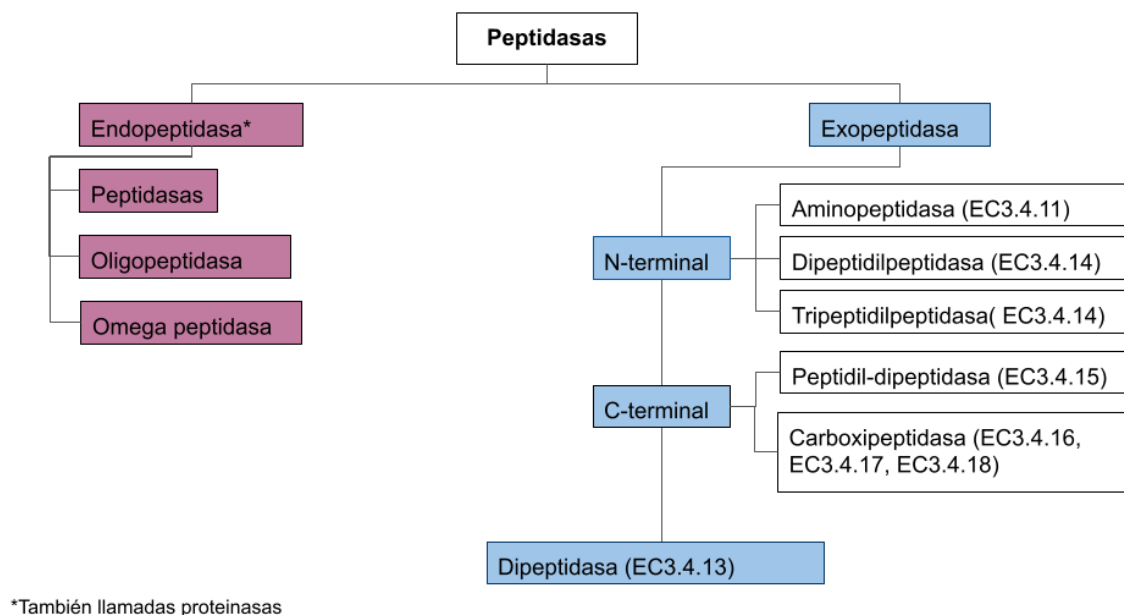


Figura 2. Clasificación de peptidasas según su especificidad y tipo de catálisis que realizan Fuente: Elaboración propia

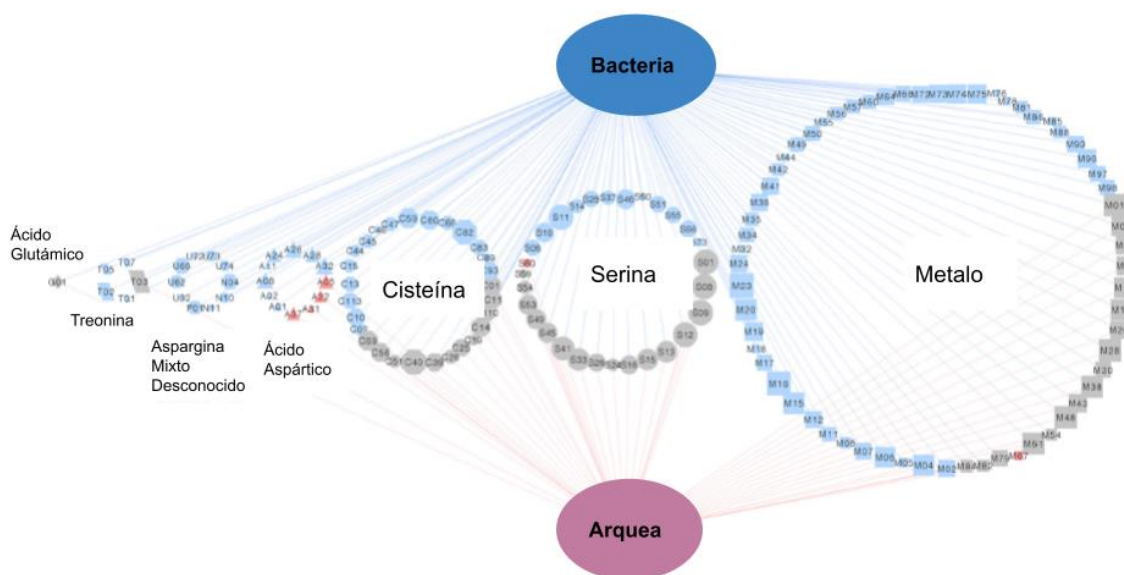


Figura 3. Distribución de peptidasas encontradas en bacterias y/o arqueas en clanes y familias según MEROPS. Modificado de Nguyen et al., 2019.

Algunos estudios señalan que las metalopeptidasas y las peptidasas de serina son las más abundantes en bacterias y arqueas (Nguyen et al., 2019). Las peptidasas aspárticas son en su mayoría codificadas por hongos, las metalopeptidasas son comunes en las bacterias, y las peptidasas de cisteína y serina son universales en los microorganismos (Caldwell, 2005). Las proteinasas tienen diversos roles bioquímicos, fisiológicos y reguladores en las bacterias y arqueas, sin embargo, su abundancia difiere entre los dos grupos de microorganismos y existen algunas familias únicas en cada uno de ellos (Nguyen et al., 2019). La distribución de las peptidasas también varía en función de los microhábitats ecológicos ocupados por diferentes taxones microbianos y las condiciones ambientales (Rawlings et al., 2019).

Lipasas

Las enzimas hidrolíticas de éster carboxílico constituyen un gran grupo de enzimas que pueden catalizar la hidrólisis, síntesis o transesterificación de un enlace éster. Se pueden encontrar en bacterias y arqueas hipertermófilas (Levisson et al., 2009). Hay dos grupos bien conocidos dentro de la familia de las hidrolasas de éster carboxílico: lipasas y estererasas. Las estererasas difieren de las lipasas al mostrar preferencia por los ésteres de acilo de cadena corta (más cortos que 10 átomos de carbono) y que no son activos en sustratos que forman micelas (Chahinian et al., 2002). Otros grupos incluyen, por ejemplo, arilestererasas y fosfolipasas. Según Kovacic y colaboradores (2018), la mayoría de las secuencias de lipasas disponibles han sido identificadas en bacterias. A pesar de la baja similitud entre las secuencias de aminoácidos entre las enzimas lipolíticas de bacterias, actualmente se clasifican en 19 familias según criterios filogenéticos, secuencias conservadas y funciones biológicas (Figura 4) (Kovacic et al., 2018). Las lipasas participan en diferentes mecanismos celulares tales como el crecimiento de las células, la adhesión a diferentes estructuras, tienen funciones específicas junto a otras enzimas, además tienen un rol muy importante en las infecciones microbianas (Stehr et al, 2003, Hausmann y Jaeger, 2010). El papel más destacado de las lipasas extracelulares para un microorganismo es la digestión de lípidos para la adquisición de nutrientes (Stehr et al, 2003).

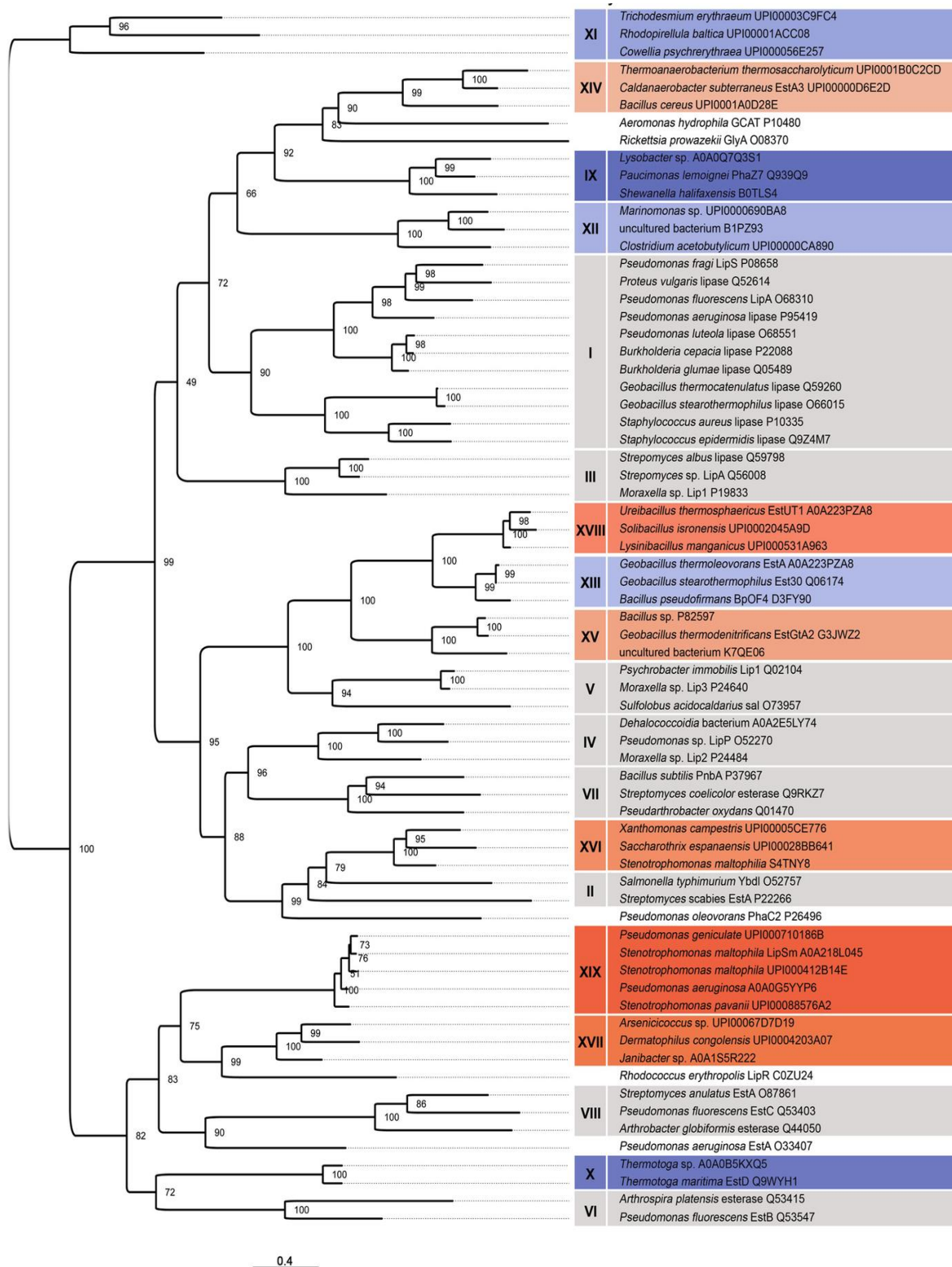


Figura 4. Árbol filogenético de las lipasas y esterases clasificadas en 19 subfamilias. Las familias están ordenadas por números romanos. Las familias I-VIII descritos por Hausmann y Jaeger 2010 están resaltados por recuadros grises. Las familias restantes IX al XIX están coloreadas de azul y rojo, las lipasas no clasificadas se encuentran en cajas sin colorear. Tomado de Kovacic et al., 2018.

Amilasas

Las amilasas se pueden clasificar en dos grandes grupos, las α -amilasas (EC 3.2.1.1), las cuales son enzimas extracelulares que hidrolizan los enlaces glicosídicos α -1,4 del almidón al azar, liberando dextrinas como subproductos, y las β -amilasas (EC3.2.1.2), exoenzimas que hidrolizan los enlaces α -1,4 del extremo no reductor, lo que provoca la inversión de la configuración anomérica de la maltosa liberada a su forma β (Mehta y Satyanarayana, 2016). Las α -amilasas muestran una amplia gama de degradación del sustrato, pueden degradar amilosa, amilopectina, ciclodextrinas, glucógeno y dextrinas, pero poseen la mayor especificidad hacia el almidón (Antranikian, 1992). Varios iones metálicos pueden influir en la actividad de estas enzimas, tal como lo es el Ca^{2+} , ya que puede mejorar la actividad de la enzima y le proporciona estabilidad cuando la temperatura aumenta (Khajeh et al., 2001). Actualmente, las α -amilasas se clasifican en las familias GH13, GH57 y GH119, siendo que las α -amilasas de las familias GH57 y GH119 se encuentran únicamente de procariotas (Janeček et al., 2012). La familia GH13 es la principal familia de α -amilasas, consta de más de 30 diferentes especificidades enzimáticas, y junto con GH70 y GH77 forma el clan GH-H (Cantarel et al., 2009). Los miembros del clan GH-H comparten el dominio α (β / α) y puede ser reconocido porque posee de 4 a 7 regiones conservadas de aminoácidos que contienen tres residuos catalíticos (Chen et al., 2012).

Existen bases de datos específicas como CAZy (*The Carbohydrate-Active EnZymes database*), la cual describe las familias que degradan, modifican o crean enlaces glicosídicos (Lombard et al., 2014). La mayoría de amilasas son extracelulares, sin embargo existen algunas que se encuentran dentro la célula, como lo son algunas encontradas en la bacteria hipertermófila *Thermotoga maritima* MSB8 (Ballschmiter et al., 2006) y la encontrada en *Streptococcus mutans* (Simpson y Russell, 1998).

Enzimas involucradas en la remoción de metales pesados y agroquímicos

Un xenobiótico es una sustancia contaminante con características químicas y estructurales complejas para las cuales se conocen pocas enzimas con la capacidad para degradarlo, y que han sido liberados al medio ambiente debido a actividades industriales y agrícolas (Gianfreda y Rao 2017). Algunas de las vías metabólicas relacionadas con el metabolismo, transporte intracelular y/o degradación de metales pesados, halocarbonos, herbicidas y pesticidas han sido descritas previamente en la base de datos como BacMet (<http://bacmet.biomedicine.gu.se/>) y KEGG (<https://www.genome.jp/kegg/>).

Los estudios basados en NGS permiten la identificación y caracterización de los genes que codifican por enzimas en los microorganismos independientemente de sistemas de cultivo a nivel de laboratorio. Esto amplía el panorama de estudio y podría revelar la presencia de enzimas con características únicas, entre las cuales están las involucradas en el metabolismo y degradación de xenobióticos (Saxena et al., 2017, Bao et al., 2017). La hidrólisis de plaguicidas por enzimas hidrolíticas de microorganismos puede servir como un mecanismo de desintoxicación o activación para la selectividad o resistencia de plaguicidas, ya que la especificidad de sustrato puede variar drásticamente entre microorganismos (Hoagland y Zablutowicz, 2001). Los estudios de metagenómica han permitido identificar y caracterizar enzimas hidrolíticas relacionadas con la degradación y metabolismo de contaminantes (Pushpanathan et al., 2014, Ufarté et al., 2015, Saxena et al., 2017, Jeffries et al., 2018), lo que ha generado nuevas herramientas e información relevante en aspectos como la biorremediación y otras aplicaciones industriales.

Es de gran importancia mencionar que existen muchos agroquímicos que utilizan en sus formulaciones metales tales como cobre y arsénico. Existen en el mercado herbicidas de cobre quelatado, herbicidas para plantas acuáticas a base de cobre (Mastin y Rodgers, 2000, Wagner et al., 2016), así como compuestos antimicrobianos a base de este metal (Lamichhane et al., 2018). En el caso del arsénico, existen insecticidas y fungicidas utilizados en la agricultura como arsenito de sodio (Na_2HAsO_3), arsenito de calcio ($\text{Ca}(\text{AsO}_2)_2$), cobre-acetoarsenito-ParisGreen ($\text{Cu}(\text{CH}_3\text{COO})_2 \cdot 3 \text{Cu}(\text{AsO}_2)_2$), piroarsenito de cobre ($\text{Cu}_2\text{As}_2\text{O}_5$),

arsenato de calcio ($\text{Ca}_3(\text{AsO}_4)_2$), y arsenato de sodio ($\text{Na}_2\text{HAsO}_4 \cdot 12 \text{H}_2\text{O}$) y arseniato de plomo (Bencko and Foong, 2017). El uso de arsénico en combinación con Pb, Ca, Mg, Mn, Fe y arsenicales inorgánicos está prohibido en Costa Rica desde 1999 (Servicio Fitosanitario del Estado, Ministerio de Agricultura y Ganadería, 2021). No obstante, existen herbicidas como glifosato, fungicidas como folpam y teldor e insecticidas como Pyrinex que pueden tener metales pesados como As, Br, Co, Pb y Ni en su formulación (Defarge et al., 2018). De estos compuestos se tiene reporte de importación a Costa Rica desde el año 2005 (Servicio Fitosanitario del Estado, MAG, 2021).

La mayoría de los estudios de remoción de metales se basan en procesos de adsorción a nivel de la membrana celular, en el que interfieren fuerzas electrostáticas. No obstante, existen otros mecanismos como la bioacumulación, el cual es un proceso metabólicamente activo en el que los microorganismos absorben metales en su espacio intracelular utilizando complejos importadores que crean una ruta de translocación a través de la bicapa lipídica (sistema de importación) (Diep et al., 2018). Una vez dentro del espacio intracelular, los metales pueden ser secuestrados por proteínas y ligandos peptídicos (Malik, 2004, Mishra y Malik, 2013).

Las bacterias se protegen de los compuestos tóxicos, para esto utilizan diversos mecanismos, entre ellos las bombas de eflujo. Algunos transportadores permiten la salida de un compuesto en específico o una clase de compuestos, y existen también transportadores de múltiples compuestos estructuralmente no relacionados (Ikeda y Yoshimura, 2002). Cinco superfamilias de bombas de flujo están asociadas a la resistencia/tolerancia de los microorganismos a diferentes drogas: 1) MDR: extrusión de múltiples fármacos y toxinas (MATE), 2) pequeña resistencia a múltiples fármacos (SMR), 3) superfamilia de facilitadores principales (MFS), 4) casete de unión a ATP (ABC) y 5) Resistencia-nodulación y división (RND). Según Alav y colaboradores (2018), las bombas de flujo de RND solo se han encontrado en bacterias Gram negativas y están organizadas como sistemas tripartitos que constan de una bomba de membrana citoplásmica, una proteína adaptadora periplásmica y un canal proteico de membrana externa. Todas las superfamilias de bombas de flujo utilizan energía de la fuerza motriz protón / sodio,

excepto la superfamilia ABC, que son transportadores primarios que utilizan energía de la hidrólisis de ATP. Las bombas son un componente clave para el flujo de fármacos, que es uno de los principales mecanismos de resistencia a los antibióticos en las bacterias. En las bacterias Gram positivas, la superfamilia de bombas de flujo de MFS es la más estudiada.

En las células bacterianas existen diferentes enzimas que regulan la concentración de los metales en su interior. Entre estas se encuentran las metalotioneínas, fitoquelatinas y metalochaperonas, las cuales pueden utilizar los metales como cofactores enzimáticos y transportadores de membrana, y además participan en procesos de homeostasis y desintoxicación de iones metálicos. El control de la expresión de genes que codifican por transportadores de metales y proteínas de almacenamiento generalmente se da mediante sensores de metales, que incluyen sistemas de dos componentes y más de siete familias de reguladores transcripcionales de unión a ADN y de unión a metales (Waldron y Robinson, 2009). En las bacterias, la limitación de metales activa las vías que participan en la importación y movilización, mientras que el exceso induce la salida y el almacenamiento (Chandrangsu, et al., 2017).

En estudios previos de cepas de cianobacterias aisladas de fuentes termales de nuestro país se ha demostrado la capacidad de tolerar y remover herbicidas como paraquat y bromacil (Brenes-Guillén et al., 2017), y metales pesados tales como aluminio, cobre, plomo y arsénico y hierro (Núñez, 2012). Estos resultados sugieren que algunos de los microorganismos aislados de ambientes con altas temperaturas son capaces de tolerar altas concentraciones de estos compuestos, y además remover sustancias tóxicas, lo que resulta de gran utilidad para el desarrollo de nuevas estrategias en biorremediación. Además del hecho que la mayoría de las aguas termales tienen alto contenido de algunos metales, por lo que las estrategias de resistencia a estos ya están activas en los microorganismos que ahí se encuentran.

Justificación

La metagenómica proporciona una serie de estrategias para evaluar el acervo genético completo de todos los microbios en un entorno particular. Con esta técnica se ha revelado una diversidad sin precedentes en la composición de la comunidad microbiana y la diversidad de funciones codificada en los genomas que revela genes y alelos nuevos. Las fuentes termales poseen nichos con condiciones extremas en los cuales se hipotetiza que se asemeja al ambiente en que se originó la forma de vida primitiva, lo que los convierte en recursos muy particulares para la identificación y caracterización de enzimas, así como sus variantes. Estos ambientes, con condiciones extremas de temperatura, pH y altas concentraciones de sales e iones, propician que los microorganismos que los habitan tengan genes que codifiquen para enzimas relacionadas con la sobrevivencia a dichas condiciones. Es por eso por lo que, son candidatos atractivos para la identificación y búsqueda de secuencias de ADN que codifican para moléculas como las enzimas con actividad hidrolítica, involucradas en la degradación de compuestos contaminantes como xenobióticos. La presencia de genes con capacidad de codificar enzimas que degradan compuestos complejos ha sido ya descrita en algunos estudios, sin embargo, no se ha realizado en nuestro país, cuyas fuentes termales tienen comunidades microbianas, diferentes tanto molecular como fenotípicamente a los reportados en otras fuentes termales.

La identificación y clasificación de las enzimas hidrolíticas y sus variantes alélicas provenientes de microorganismos termotolerantes ofrecerá una oportunidad para caracterizar estas secuencias y generar nuevas investigaciones. Las bases de datos actuales son un reservorio importante para comparar los datos generados en fuentes termales costarricenses con los de otros sitios, lo que permitiría realizar análisis de la micro-diversidad de ambientes extremos, así como comparaciones funcionales y metabólicas entre las diferentes muestras. Además, la información generada permitirá incrementar el conocimiento a nivel ecológico y metabólico de

estos ambientes, lo que justificaría futuros esfuerzos de investigación para la exploración y conservación de los recursos biológicos disponibles en las fuentes termales.

Objetivos

Objetivo General

Analizar y comparar datos de secuenciación de seis metagenomas de fuentes termales de Costa Rica y otros metagenomas públicos en búsqueda de secuencias de potenciales enzimas hidrolíticas.

Objetivos Específicos

- 1- Comparar la composición microbiana y anotación funcional de las fuentes termales costarricenses con 13 metagenomas de fuentes termales del Parque Nacional Yellowstone, Nueva Zelanda y Chile según las características fisicoquímicas como temperatura, pH y concentración de iones.
- 2- Identificar y clasificar secuencias de enzimas hidrolíticas tipo proteasas, lipasas, amilasas y esterases encontradas en seis fuentes termales costarricenses.
- 3- Clasificar y describir las enzimas posiblemente involucradas en la remoción de metales y agroquímicos en seis fuentes termales costarricenses

Materiales y Métodos

Características de las muestras

Las muestras utilizadas en este análisis pertenecen a la base de datos de metagenomas y microbiomas del Área de Microbiología Ambiental del CIBCM. Esta base está conformada por 14 metagenomas obtenidos a partir de secuenciación tipo *shotgun* de tapetes microbianos de diferentes partes del país y 30 muestras, de las cuales se tienen datos de secuenciación de amplicones 16S ARNr y algunas de estas, secuenciación del gen 18S ARNr. En la figura 5, se muestra una fotografía de cada una de las fuentes termales. En el cuadro 1, se muestra la ubicación, temperatura y pH de cada una de las fuentes termales de donde fueron colectados los tapetes microbianos utilizados en este estudio.

Cuadro 1. Localización geográfica, temperatura, pH de muestras utilizadas en este proyecto.

Muestra	Localización geográfica- asociación a volcán	Temperatura (°C)	pH
Rocas Calientes (RC1)	Zona Sur	63.0	6.0
Las Lilas	Volcán Rincón de la Vieja	74.0	6.0
Miravalles	Volcán Miravalles	51.0	7.03
La Luz	Cordillera de Tilarán	49.9	6.24
Salitral	Cordillera de Tilarán	34.9	6.1
Río Naranjo	Cordillera de Tilarán	59.8	6.53

Se extrajo el ADN de tapetes microbianos y agua de seis fuentes termales que han sido previamente caracterizados mediante secuenciación masiva de la región V3-V4 del gen 16S ARNr, y que poseen análisis fisicoquímicos asociados (Uribe-Lorío et al., 2019, Brenes-Guillén et al, datos no publicados). Para cada una de las muestras se tienen datos ambientales como pH y temperatura *in situ*, los cuales fueron obtenidos mediante una sonda multiparamétrica (Combo, Hanna Instruments). La extracción de ADN ambiental se realizó utilizando el kit de extracción Nucleospin Soil (Mackerey-Nagel) de acuerdo con el protocolo de la casa comercial. La calidad del ADN fue visualizada mediante electroforesis en un gel de agarosa al 1%. La concentración de ADN, así como la pureza relación 260/280 y 260/230 se cuantificó utilizando un espectrofotómetro Nanodrop 2000/2000c. El ADN extraído fue almacenado en un congelador a -20 °C. Su usó el TruSeq Nano DNA Kit para la preparación de librerías y se secuenciaron en la plataforma HiSeq (Illumina Inc., USA) por servicio de la empresa Macrogen Inc. (Corea). Las estadísticas generales de las secuencias se muestran en el Cuadro 2. Los archivos SRA (*Sequence Read Archive*) de las muestras de Nueva Zelanda, Parque Nacional Yellowstone (PNY) y Chile fueron obtenidas de GenBank, sus características generales se muestran en el cuadro 3. Para algunas de las muestras se obtuvieron fotografías de diferentes fuentes (Fig. 6).

Cuadro 2. Estadísticas de los datos crudos de secuenciación de las muestras de las fuentes termales de Costa Rica. Las columnas Q20 y Q30 corresponden al porcentaje de bases con un puntaje de calidad de 20 y 30 respectivamente.

ID	Total de bases	Total de secuencias (paired-end)	GC (%)	AT (%)	Q20 (%)	Q30 (%)
La Luz	16,368,673,878	162,066,078	55.25	44.75	97.49	92.21
Miravalles	9,027,602,402	89,382,202	55.03	44.97	96.55	89.60
Las Lilas	8,939,704,930	88,511,930	45.47	54.53	96.70	90.60
Salitral	14,987,773,194	148,393,794	55.32	44.68	97.18	91.49
Río Naranjo	13,632,200,280	134,972,280	51.14	48.86	97.26	91.60
Rocas Calientes	9,627,099,416	95,317,816	60.12	39.88	96.60	91.30

Cuadro 3. Características generales de cada muestra utilizada en el análisis.

ID	Bioproyecto	Secuenciador	Sitio	Temp. °C	pH
SRR5248290	PRJNA366268	Illumina HiSeq 2500	Obsidian Pool, Parque Nacional Yellowstone, Estados Unidos	42-90	~ 6.7
SRR5248299	PRJNA366343	Illumina HiSeq 2000	Obsidian Pool, Parque Nacional Yellowstone, Estados Unidos	42-90	~ 6.7
SRR5248365	PRJNA366346	Illumina HiSeq 2000	Obsidian Pool, Parque Nacional Yellowstone, Estados Unidos	42-90	~ 6.7
SRR5248302	PRJNA366345	Illumina HiSeq 2000	Obsidian Pool, Parque Nacional Yellowstone, Estados Unidos	42-90	~ 6.7
SRR5451033	PRJNA382437	Illumina HiSeq 2000	Tapete microbiano, Porcelana, Los Lagos, Chile	48	~ 6.5
SRR5451032	PRJNA382437	Illumina HiSeq 2000	Tapete microbiano, Porcelana, Los Lagos, Chile	58	~ 6.5
SRR5451031	PRJNA382437	Illumina HiSeq 2000	Tapete microbiano, Porcelana, Los Lagos, Chile	66	~ 6.5
SRR10063240	PRJNA412936	Illumina HiSeq 1500	Raoul Island (Green Lake), Nueva Zelanda	34.5	6.96
SRR10063241	PRJNA412936	Illumina HiSeq 1500	Raoul Island (Marker Bay Pool), Nueva Zelanda	29.9	5.9
SRR10063242	PRJNA412936	Illumina HiSeq 1500	Raoul Island (Eastern Pool), Nueva Zelanda	27.2	5.43
SRR5650826	PRJNA378813	Illumina NextSeq 500	Mound Spring, Parque Nacional Yellowstone, Estados Unidos	60-100	9
SRR5650808	PRJNA378813	Illumina NextSeq 500	Bijah Spring, Parque Nacional Yellowstone, Estados Unidos	60-100	7
SRR4030101	PRJNA336659	Illumina HiSeq 2000	Mammoth Hot Spring (Liberty Cap Streamers), Parque Nacional Yellowstone, Estados Unidos	72	6.5-7
Río Naranjo		Illumina HiSeq 4000	Tapete microbiano Salitral, Costa Rica	34.9	6.53

Salitral		Illumina HiSeq 4000	Tapete microbiano Río Naranjo, Costa Rica	59.8	6.1
La Luz		Illumina HiSeq 4000	Tapete microbiano La Luz, Costa Rica	49.9	6.24
Miravalles		Illumina HiSeq 4000	Tapete microbiano Miravalles, Costa Rica	51	7.03
Las Lilas		Illumina HiSeq 4000	Tapete microbiano Las Lilas, Costa Rica	74	6
Rocas Calientes		Illumina HiSeq 4000	Tapete microbiano Rocas Calientes, Costa Rica	63	6

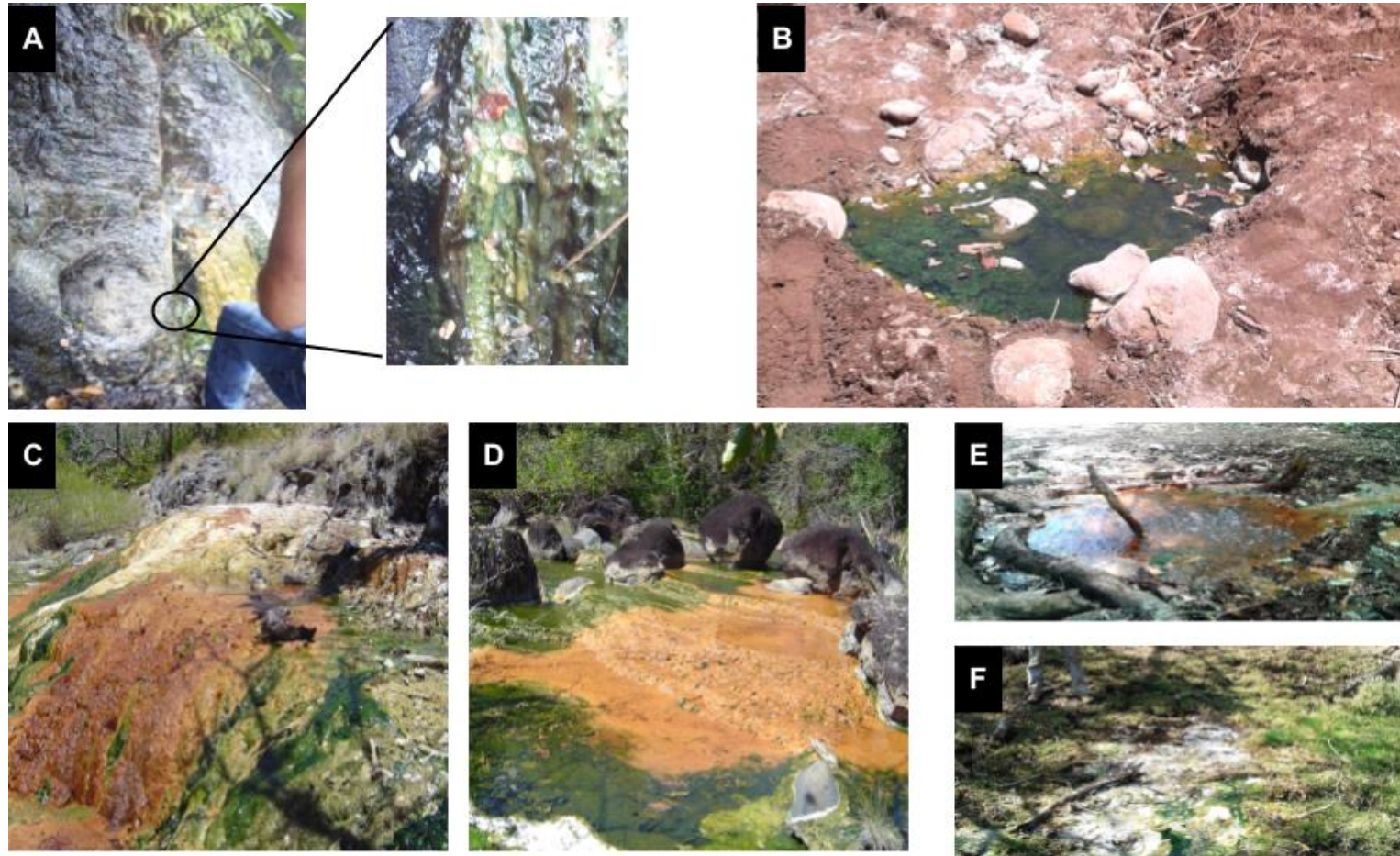


Figura 5. Fotografías de fuentes termales costarricenses. A. Rocas Calientes. B. Río Naranjo. C. Miravalles. D. Las Lilas. E. Salitral. F. La Luz. A: Fotografía facilitada por Lorena Uribe. B, E y F: Fotografías facilitadas por Guillermo Alvarado Induni. C y D: tomadas de Morales 2008 y Brenes Guillén et al., 2019.



Figura 6. Fotografías de fuentes termales de Yellowstone, Nueva Zelanda y Chile. A. Liberty Cap. B. OpSIDian Pool. C. Mound Spring, D. Bijah Spring, E. Fuentes termales en Nueva Zelanda. F. Fuente termal Porcelana, Chile. En los recuadros se señalan las diferentes fuentes bibliográficas de donde se tomaron las fotografías.

Análisis Bioinformático

Se realizó un análisis de los datos utilizando como referencia diferentes programas que han sido utilizados previamente para el análisis de datos de metagenómica *shotgun*. En la figura 7, se muestra un resumen de la metodología utilizada y a continuación se detallan las características de cada sección del análisis.

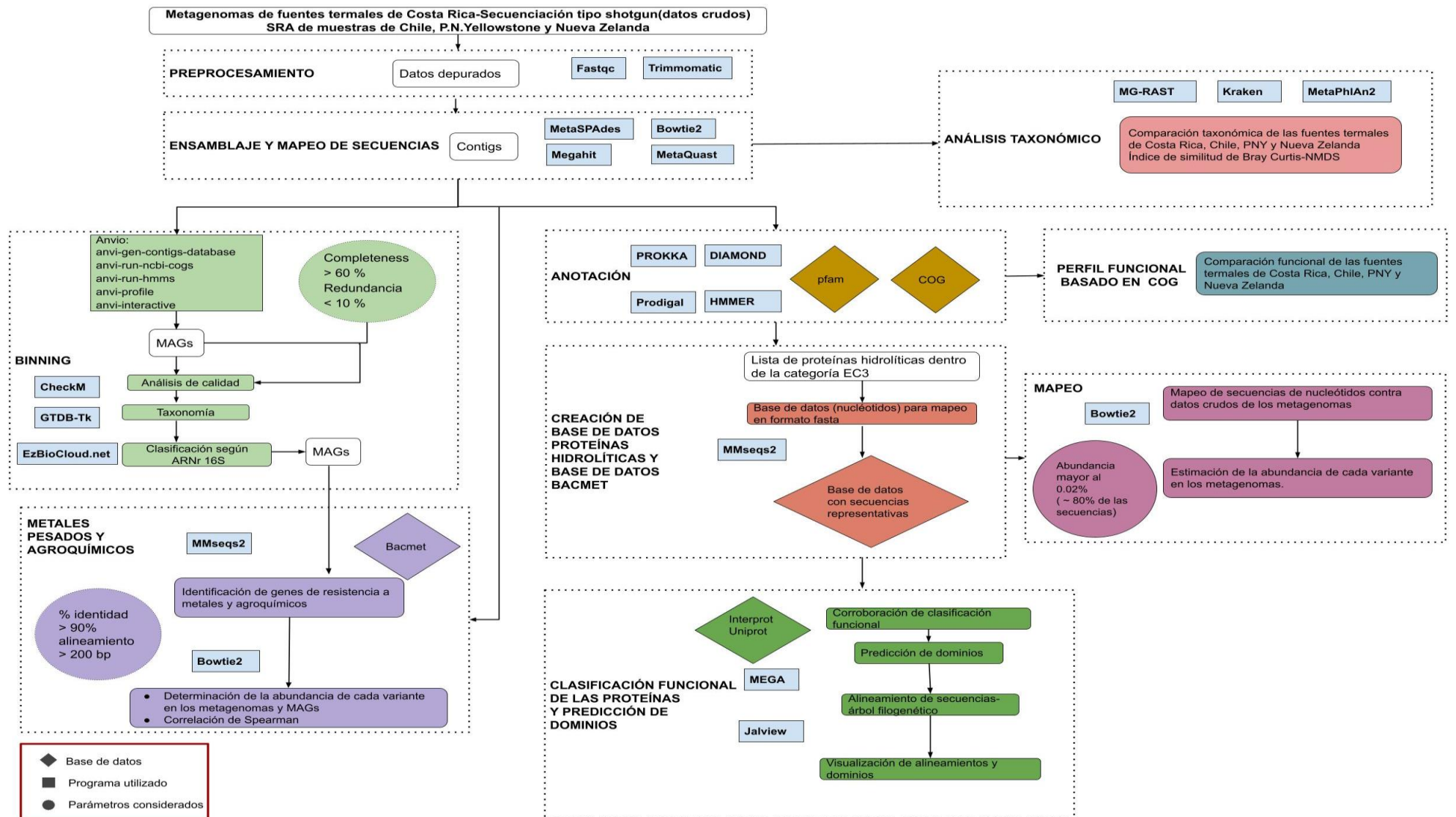


Figura 7. Diagrama general de la metodología de análisis utilizada en este trabajo.

Análisis de calidad de los datos de secuenciación

La mayoría de los análisis se realizaron utilizando el clúster Kabre (CENAT, CONARE). Para el análisis de calidad de las secuencias se utilizó el módulo FastQC (v 0.11.7). Este análisis permitió obtener datos relacionados con la calidad de la secuencia, cantidad de secuencias y permitió visualizar los valores de calidad de las secuencias. Posteriormente, se utilizó el programa Trimmomatic v0.32 (USADELLAB.org) con los parámetros SLIDINGWINDOW:4:20 y MINLEN: 100, para eliminar las secuencias con baja calidad (phred score menor a 20), eliminar adaptadores, *primers* y, secuencias cortas (menor a 100 pb).

Ensamblaje, alineamiento y anotación de metagenomas

La estandarización del análisis bioinformático de metagenomas de las fuentes termales se realizó utilizando como base el conjunto de programas ANVIO (Universidad de Chicago) desarrollado por Eren y colaboradores (2015). Para el ensamblaje de los metagenomas se utilizaron dos ensambladores: MEGAHIT (Li et al., 2015) y MetaSPAdes (Nurk et al., 2017). Debido a que se utilizó un ensamblaje de *novο*, se evaluaron ciertas métricas basadas en el número de *contigs* y *scaffolds* generados, la proporción de secuencias que fueron ensambladas, y el tamaño en pares de bases de esas secuencias (porcentaje de secuencias mapeadas al ensamblaje mayor al 60%), para escoger aquellas con un tamaño mayor a 1000 pb y un tamaño de k-mer menor o igual a 255 (parámetro por default). La evaluación de las estadísticas que se generen en cada ensamblaje se realizó con el programa MetaQUAST (Mikheenko et al., 2016). Para el alineamiento de las secuencias y obtención de estadísticas como porcentaje de *misassemblies* se utilizó la herramienta Bowtie2 (buscadores locales y globales) (Langmead y Salzberg, 2012). La predicción de genes se realizó utilizando el programa PRODIGAL (Hyatt et al., 2010) y la herramienta HMMER (Finn et al., 2015) para mapear los resultados de las anotaciones proteicas obtenidas entre las bases de datos GO (Gene Ontology) y Pfam 32.0 (<https://pfam.xfam.org>). Se utilizaron los parámetros por defecto de Prokka que incluyen un valor de *cut-off* de 1×10^{-6} (probabilidad que el alineamiento ocurra al azar y no por similitud, es significativo si es menor a 0.001) (Seemann, 2014).

La herramienta HMMER se basa en modelos probabilísticos ocultos de Markov (HMM) e incorpora la clasificación taxonómica a las muestras utilizando las bases de datos

propuestas por Rinke y colaboradores (2013) y Campbell y colaboradores (2013). Para la identificación de proteínas ya reportadas en la literatura se utilizó el programa DIAMOND (Buchfink et al., 2014) y las bases de datos COG (versión 2014) y Pfam (versión 32.0). Todo este proceso se realizó tanto para las muestras costarricenses como para los datos de 13 fuentes termales de Yellowstone, Nueva Zelanda y Chile.

Perfil taxonómico y funcional de las comunidades microbianas

Para determinar el perfil taxonómico de las comunidades se utilizaron dos clasificadores: Kraken2 dentro de OmicsBox (<https://www.biobam.com/omicsbox/>), MetaPhlAn2 (<https://huttenhower.sph.harvard.edu/metaphlan>) y MG-RAST (<https://www.mg-rast.org/>). El análisis en MG-RAST solamente se realizó para las muestras de Costa Rica. MG-RAST es una aplicación web en la cual se pueden elegir una serie de parámetros para el control de calidad. Luego, los datos crudos pasan automáticamente por varios pasos y, produce perfiles de abundancia a diferentes niveles taxonómicos, características funcionales y visualizaciones. Por otro lado, Kraken2, la última versión de Kraken, es un clasificador de secuencias que asigna etiquetas taxonómicas a lecturas cortas de ADN. Lo hace examinando los *k-mers* dentro de una lectura y consultando una base de datos con esos *k-mers*. Esta base de datos contiene un mapeo de cada *k-mer* en la biblioteca genómica de Kraken al ancestro común más bajo (LCA) (Wood et al., 2019). MetaPhlAn (Análisis filogenético metagenómico) es una herramienta computacional para perfilar la composición de comunidades microbiana, se basa en genes marcadores específicos de clado únicos identificados a partir de ~ 17.000 genomas de referencia (~ 13.500 bacterianos y arqueales, ~ 3500 virales y ~ 110 eucariotas) (Segata et al., 2012).

Una vez obtenida la abundancia de cada filo por muestra, se realizó un análisis de estadística multivariada para determinar la similitud entre las muestras, para ello se utilizó como insumo solamente los resultados obtenidos en Kraken. Para ello los datos fueron normalizados utilizando la raíz cuadrada y se realizó una prueba del perfil de similitud (SIMPROF), recomendada para muestras no divididas *a priori* en grupos (muestreo, los valores obtenidos en el cálculo del índice de similitud de Bray-Curtis y diseño experimental). SIMPROF utiliza como insumo permutaciones para estimar el patrón de agrupamiento entre las muestras. Posteriormente se realizó la prueba SIMPER (*Similarity percentage analysis*) y BEST para determinar la contribución de cada uno de los filios a las diferencias entre las muestras. BEST permite determinar las variables que "explican mejor" el patrón multivariado en las variables de respuesta. La prueba SIMPROF se utiliza para decidir qué subgrupos en

el análisis jerárquico son estadísticamente significativos. Posteriormente, se graficó el *shadeplot*, basado en los resultados de SIMPROF, en el cual se puede observar la abundancia de cada filo por muestra agrupadas, las líneas punteadas indican que el patrón de agrupamiento no tiene un soporte estadístico y por lo tanto no pertenecen al mismo grupo. El análisis de NMDS muestra el patrón de agrupamiento entre las muestras y el porcentaje de similitud entre las mismas según los análisis preliminares. Estos análisis se realizaron en el programa Primer 7 (*Plymouth Routines In Multivariate Ecological Research*, Clarke y Goyle 2015).

Para realizar las comparaciones del perfil funcional de las comunidades, se utilizaron los datos generados por Prokka (Pfam y COG). Para ello se determinó el número de secuencias de proteínas dentro de las 17 categorías funcionales COG (Cuadro 4). Posteriormente se construyó una base de datos para comparar las diferentes fuentes termales, para esto se utilizaron los 2448 COGs encontrados y se graficó la abundancia de cada COG con el programa estadístico Primer7. Es importante mencionar que los pares COG se distribuyen en una amplia gama de identidad de secuencia y valor esperado. Ni el porcentaje de identidad de secuencia ni el valor esperado del número de aciertos esperados de calidad similar (puntuación) que podrían encontrarse por casualidad (e-value), pueden dar una idea completa de la relación entre las dos proteínas. Si dos proteínas tienen una identidad de secuencia entre 60% - 70%, tienen cerca del 90% de probabilidad (quizás más) de compartir el mismo proceso biológico para los niveles del índice de GO.

Cuadro 4. Categorías funcionales según COG.

Clasificación	Categoría	Letra
Almacenamiento y procesamiento de información	Modificación y procesamiento de ARN	A
	Traducción, estructura ribosómica y biogénesis	J
	Transcripción	K
	Replicación, recombinación y reparación	L
Metabolismo	Conversión y producción de energía	C
	Metabolismo y transporte de aminoácidos	E
	Metabolismo y transporte de nucleótidos	F
	Metabolismo y transporte de carbohidratos	G
	Metabolismo y transporte de coenzimas	H
	Metabolismo y transporte de lípidos	I
	Metabolismo y transporte de sustancias inorgánicas	P
	Biosíntesis, transporte y catabolismo de metabolitos secundarios	Q
Señalización y procesamiento celular	Control del ciclo celular, división celular, partición cromosómica	D
	Biogénesis de la pared celular / membrana / envoltura	M
	Motilidad celular	N
	Modificación postraducciona, proteínas y chaperonas	O
	Mecanismos de traducción de señal	T
	Tráfico intracelular, secreción y transporte vesicular	U
	Mecanismos de defensa	V
	Estructuras extracelulares	W
	Citoesqueleto	Z
Poco caracterizado	Función general	R
	Función desconocida	S
Sin clasificación	Mobiloma: profagos y transposones	X

Genomas obtenidos a partir de (meta)genomas (MAGs)

El ensamblaje fue utilizado para la obtención de MAGs. Los MAGs de los tapetes microbianos fueron obtenidos mediante la interfaz interactiva de Anvio versión 5.0 (Eren et al., 2015). Posteriormente se filtraron aquellos que tenían un porcentaje de completitud mayor al 60% y de redundancia menor al 10%. Posteriormente se corroboró esta información utilizando CheckM (Parks et al., 2015) y su afiliación taxonómica con GTDB-Tk (Chaumeil et al., 2020) en la plataforma KBase (Arkin et al., 2018). Una vez obtenidos los MAGs se pasó a hacer el mapeo de los diferentes genes relacionados con la remoción de metales con la metodología antes mencionada.

Creación de base de datos

Para la creación de las bases de datos de secuencias proteicas, se utilizaron los resultados de la anotación de los 19 metagenomas de fuentes termales. De cada archivo producto de la anotación de las proteínas, se construyó un archivo que incluía todas las secuencias en formato fasta. Se utilizó la herramienta MMseqs2 (Steinegger y Soeding et al., 2017) para seleccionar secuencias de proteínas representativas de cada una de las bases de datos creadas, utilizando los parámetros `--min-seq-id 1 --cluster-mode 1 --kmer-per-seq 1 --alignment-mode 3`. Estos parámetros fueron seleccionados con el fin de agrupar secuencias con alta identidad (=100%), para ello se realiza un alineamiento que cubra al menos el 80% de las secuencias para crear el clúster. Se consideró una variante o alelo, aquella secuencia que tenga un porcentaje de identidad menor a 100% con otra secuencia en la base de datos y al menos un *mismatch* al realizar el mapeo. Posteriormente, se ordenó la clasificación de las enzimas generadas. Para ello se utilizó el esquema de clasificación numérica, basado en las reacciones químicas que catalizan (Comisión de Enzimas: número EC) (Cuadro 5). Se utilizaron diferentes términos de búsqueda para cuantificar y encontrar los diferentes grupos de proteínas dentro de la categoría de enzimas hidrolasas (EC3). Para la categoría de peptidasas se utilizaron los términos de búsqueda “peptidasa”, “proteínasa” y “proteas” en inglés. De cada metagenoma se extrajeron las secuencias clasificadas dentro de cada categoría. Esta información se utilizó como referencia para la construcción de la base de datos empleada en el mapeo de genes.

Cuadro 5. Esquema de clasificación numérica para enzimas, basado en las reacciones químicas que catalizan (Comisión de Enzimas: número CE)

Clasificación	Descripción
EC1 Oxidoreductasas	Reacciones de oxidación/reducción y de transferencia de átomos de H, O o electrones desde un sustrato a otro.
EC2 Transferasas	Transferencia de un grupo funcional desde un sustrato a otro. El grupo puede ser metil-, acil-, amino- o fosfato.
EC3 Hidrolasas	Formación dos productos a partir de un sustrato. Facilitan la ruptura de los enlaces de C-C, C-O, C-N, S-S, S-N, S-P, C-P.
EC4 Liasas	Adición o eliminación no hidrolítica de grupos de los sustratos. Pueden romper los enlaces C-C, C-N, C-O o C-S.
EC5 Isomerasas	Isomerización de una molécula.
EC6 Ligasas	Unión de dos moléculas por síntesis de nuevos enlaces C-O, C-S, C-N o C-C con la rotura simultánea de ATP.
EC7 Translocasas	Catalizar el movimiento de iones o moléculas a través de las membranas o su separación dentro de las membranas.

Mapeo y abundancia de variantes alélicas

Posterior al establecimiento de la base de datos, se realizó el mapeo. Para el mapeo de genes se utilizó el protocolo establecido por Castro y colaboradores (2019). Se utilizaron las secuencias de nucleótidos correspondientes a las secuencias proteicas, para ellos se utilizó Bowtie2 con los parámetros -N 1 y alineamiento local, donde N corresponde al número de *mismatches* permitidos. Si se utiliza un valor más alto, el alineamiento será más lento pero aumenta la sensibilidad, ya que el alineamiento local no requiere que las lecturas se alineen de un extremo a otro. Las alineaciones locales se pueden "recortar" en uno o ambos extremos de una manera que optimice la puntuación de alineación, además, genera alineamientos en regiones con alta similitud o regiones conservadas de secuencias con diferente longitud. En el caso de las bases de datos que contenían muchas secuencias, se

procedió a realizar un mapeo preliminar por muestra para seleccionar las variantes de los diferentes grupos de enzimas con un porcentaje de cobertura mayor a 0.02 % del total de secuencias para la graficación de la abundancia de cada una por muestra, donde este 0.02 % abarca el 80 % de las secuencias mapeadas. Para efectos de este estudio se va a considerar abundancia como el número de secuencias mapeadas. Para realizar comparaciones más detalladas de las secuencias de proteínas, se utilizó la herramienta UniProt-Blast (<https://www.uniprot.org/blast/>) (The UniProt Consortium), utilizando la base de datos UniRef90 (UniProt Reference Clusters 90% idéntidad).

Mapeo de genes de resistencia a metales pesados y agroquímicos en tapetes microbianos de fuentes termales.

Para el mapeo de genes de resistencia a metales (capacidad de tolerar altas concentraciones de metales) y agroquímicos se utilizó como referencia las bases de datos BacMet y KEGG. BacMet (v2.0, marzo 2018) es una base de datos curada manualmente de genes bacterianos, que se han confirmado experimentalmente que confieren a los microorganismos resistencia a metales y/o biocidas antibacterianos, con referencias completas. Además, incluye una base de datos de genes de resistencia putativos (Pal et al., 2014). Esta base contiene 753 genes, la mayoría localizados en los cromosomas, sin embargo, algunos se encuentran en plásmidos o transposones. En el caso del agroquímico atrazina, se construyó una base de datos con las secuencias de los genes relacionados con la degradación del compuesto basada en los datos de KEGG y la anotación con Prokka. El mapeo contra las secuencias crudas depuradas (secuencias R1 y R2) se llevó a cabo utilizando Bowtie2 (N-1, --local). Posteriormente se utilizó samtools para obtener la cantidad de secuencias mapeadas. Se utilizó la base de datos UniProt e Interprot para buscar información relacionada con los genes y las proteínas asociadas. La abundancia relativa de cada uno de los genes se normalizó usando el número total de secuencias para cada fuente termal y la longitud de cada secuencia. Se aplicó la correlación de Spearman para identificar la relación entre la abundancia de los genes asociados a la remoción y/o degradación de agroquímicos y la abundancia de los diferentes filos de microorganismos.

Resultados y discusión

Análisis de calidad de las secuencias

Al realizar el análisis de calidad y la depuración de los datos, se encontró que el número de secuencias se redujo entre un 25% a un 45% después del análisis de calidad para todas las muestras. El porcentaje de secuencias obtenidas en las muestras de Costa Rica después del proceso de depuración es similar a lo obtenido en las otras fuentes termales. Esas secuencias tienen un valor de calidad mayor a 20 (probabilidad de tener una base incorrecta es del 1%) y un tamaño mayor o igual a 100 pb. Algunas de las estadísticas generales de los datos antes y después del *trimming* se encuentran en anexos (anexos cuadro 1). Es importante mencionar, que solamente se consideraron las secuencias que se encontraban en el archivo *forward* y el archivo *reverse (paired)* para el ensamblaje. Los resultados del análisis de calidad antes y después del *trimming* de las secuencias de las muestras costarricenses y de los metagenomas de las fuentes termales de Yellowstone, Nueva Zelanda y Chile se muestran en el cuadro 6.

Cuadro 6. Número de secuencias en las muestras antes y después del *trimming*

Metagenoma	Antes de <i>trimming</i>	Después del <i>trimming</i> (%)
Rocas Calientes	47658908	35128948 (73.71)
La Luz	81033039	54559528 (67.33)
Miravalles	44691101	25368157 (56.76)
Las Lilas	44255965	24302038 (54.91)
Salitral	74196897	46154709 (62.21)
Río Naranjo	67486140	43069985 (63.82)
SRR5248290	80224270	47958068 (59.78)
SRR5248299	21303953	14037174 (65.89)
SRR5248365	42019147	29833694 (71.00)
SRR5248302	28094272	17157171 (61.07)
SRR5451033	140980559	96360212 (68.35)
SRR5451032	61953213	47016293 (75.89)
SRR5451031	10263362	6067699 (59.12)
SRR10063240	49791620	27659244 (55.55)
SRR10063241	48525851	29809430 (61.43)
SRR10063242	43957771	30317674 (68.97)
SRR5650826	54274165	35164231 (64.79)
SRR5650808	33491695	18674969 (55.76)
SRR4030101	21096521	15539697 (73.66)

Ensamblaje de los metagenomas

Después de realizar el análisis de calidad de las secuencias, se procedió con el ensamblaje de los 19 metagenomas utilizando MEGAHIT y MetaSPAdes. Posteriormente se determinaron las estadísticas utilizando MetaQUAST (Cuadro 7, 8). Los resultados obtenidos para las muestras de PNY, Nueva Zelanda y Chile se muestran en los anexos (Cuadro 2).

En la mayoría de las muestras se encontraron *contigs* más largos y N50 mayor utilizando el ensamblador MEGAHIT, por lo que se consideró solamente estos resultados para los análisis posteriores. Independientemente del ensamblador utilizado, la muestra de Las Lilas obtuvo el menor número de *contigs* (14129 utilizando MEGAHIT y 9010 con MetaSPAdes), mientras que Miravalles el mayor (124 400 para MEGAHIT y 54833 con MetaSPAdes). La longitud del *contig* más corto que constituye al menos la mitad del tamaño total fue mayor en Las Lilas (N50: 39671) y menor en Miravalles (N50: 8699), lo que indica que Miravalles es la muestra con mayor número de *contigs* pero estos son menores a 10000 pb. No se encontró una relación entre el número de secuencias y la cantidad de *contigs* generados y/o tamaño del metagenoma en pares de bases.

Cuadro 7. Resultados de MetaQUAST obtenido a partir de los ensamblajes con MEGAHIT

Estadística	Rocas Calientes	La Luz	Miravalles	Las Lilas	Salitral	Río Naranjo
# contigs (≥ 0 pb)	138743	223421	530724	73277	177452	113752
# contigs (≥ 1000 pb)	13942	26929	54546	7392	25917	12304
# contigs (≥ 5000 pb)	3064	4808	7694	1978	5725	2384
# contigs (≥ 10000 pb)	1620	2848	3757	1120	2984	1306
# contigs (≥ 25000 pb)	718	1421	1477	488	1208	669
# contigs (≥ 50000 pb)	369	700	709	252	531	308
Total longitud (≥ 0 pb)	133414300	250532910	427799133	87597710	226512672	113952765
Total longitud (≥ 1000 pb)	105204136	190879219	269131563	69474535	177044849	80882246
Total longitud (≥ 5000 pb)	83498292	148994840	180970879	58304943	136809633	61460996
Total longitud (≥ 10000 pb)	73428253	135340560	153841234	52324766	117683072	54103110
Total longitud (≥ 25000 pb)	59927273	112960886	119025889	42660651	90242270	44013454
Total longitud (≥ 50000 pb)	47615453	87249436	92718904	34473656	66644766	31228468
Número de contigs	26980	54706	124400	14129	48864	26940
Contig mayor longitud	1262733	1262974	1762222	1204961	951091	542363
Total longitud	114328417	210482669	317953893	74180527	193245949	91040718
N50	29996	31831	8699	39671	20202	22645
N75	4276	3260	1594	6931	3613	2757

Cuadro 8. Resultados de MetaQUAST obtenido a partir de los ensamblajes con MetaSPAdes

Estadística	Rocas Calientes	La Luz	Miravalles	Las Lilas	Salitral	Río Naranjo
# contigs (>= 0 pb)	20113	28202	54833	9010	27726	13954
# contigs (>= 1000 pb)	20113	28202	54833	9010	27726	13954
# contigs (>= 5000 pb)	3718	5384	7882	2303	5958	2715
# contigs (>= 10000 pb)	1899	3334	3943	1277	3053	1458
# contigs (>= 25000 pb)	668	1584	1555	559	1235	638
# contigs (>= 50000 pb)	295	680	692	255	526	247
Total longitud (>= 0 pb)	106592755	185831358	258737108	69198934	173688820	80441143
Total longitud (>= 1000 pb)	106592755	185831358	258737108	69198934	173688820	80441143
Total longitud (>= 5000 pb)	74881688	143240619	170901998	55573062	129248626	57395033
Total longitud (>= 10000 pb)	62277075	128807565	143535366	48336060	108918966	48839864
Total longitud (>= 25000 pb)	43403950	100908513	107017427	37043888	80721772	35836218
Total longitud (>= 50000 pb)	30586795	69355217	77139199	26376391	56272208	22372343
# contigs	20113	28202	54833	9010	27726	13954
Contig mayor longitud	449135	1448846	1219733	843767	867202	337684
Total longitud	1065922755	185831358	258737108	69198934	173688820	80441143
N50	15493	30051	14287	30003	20800	19236
N75	3817	6154	2977	7370	4808	4109

MEGAHIT usa un sucinto o breve dBg (*Brujin graph*) para reducir los requerimientos de almacenamiento, menor memoria RAM y tarda menos tiempo en comparación con otros ensambladores (Nurk et al., 2017, van der Walt et al., 2017). El ensamblaje metagenómico es un campo de investigación que aún se encuentra en desarrollo en comparación con el ensamblaje genómico, en los últimos años se ha observado un interés creciente en su potencial y la desarrollo de nuevos programas (Ayling et al., 2020). El ensamblaje de metagenomas se enfrenta a diversos retos (Ayling et al., 2020), entre ellos caben destacar un desconocimiento previo de la diversidad y abundancia de los microorganismos, por lo que los nodos con menor cobertura pueden tener su origen en genomas con menor abundancia, no de errores, por lo que no deben descartarse de inmediato. La baja cobertura de la mayoría de las especies significa que el ensamblaje de *novo* es poco probable a menos que el genoma en cuestión sea relativamente pequeño y, en cambio, dependemos de genomas de referencia o modelos / homología de predicción de genes para obtener evidencia de la presencia de especies.

Las especies o subespecies relacionadas introducen superposiciones extensas en un conjunto *kmer*. Las ramas o bucles entre estas regiones homólogas hacen que recorrer el grafo sea más complejo, y si alguna de las especies se presenta con una abundancia baja, entonces será difícil identificar la presencia de especies separadas y la descomposición del gráfico es extremadamente compleja, impidiendo que se obtenga la verdadera diversidad de la muestra. No existe una herramienta única que sea la mejor para todas las muestras y para todas las preguntas científicas. Las herramientas a menudo se pueden diferenciar por las compensaciones que hacen entre la contigüidad y la capacidad de capturar especies y diversidad de cepas.

Genomas obtenidos a partir de (meta)genomas

Se obtuvo un total de 84 MAGs con un valor de completitud mayor al 60% y de redundancia (contaminación) menor al 10%. Las estadísticas de cada uno, como tamaño, resultados de completitud y redundancia, así como clasificación taxonómica se muestran en el Cuadro 9 y en anexos información adicional (anexos cuadro 3). Este estudio describe por primera vez MAGs de tapetes microbianos de fuentes termales costarricenses. La distribución del número de MAGs por tapete microbiano y filo se muestran en la figura 8. Salitral es la fuente termal en donde se logró obtener el mayor número de MAGs (22) y Las Lilas el menor número (8). En total, se encontraron dentro del filo Chloroflexi (Chloroflexota) 21, 14

Bacteroidetes (Bacteroidota), y 10 de Cyanobacteria y Proteobacteria, y algunos filios con uno o dos MAGs.

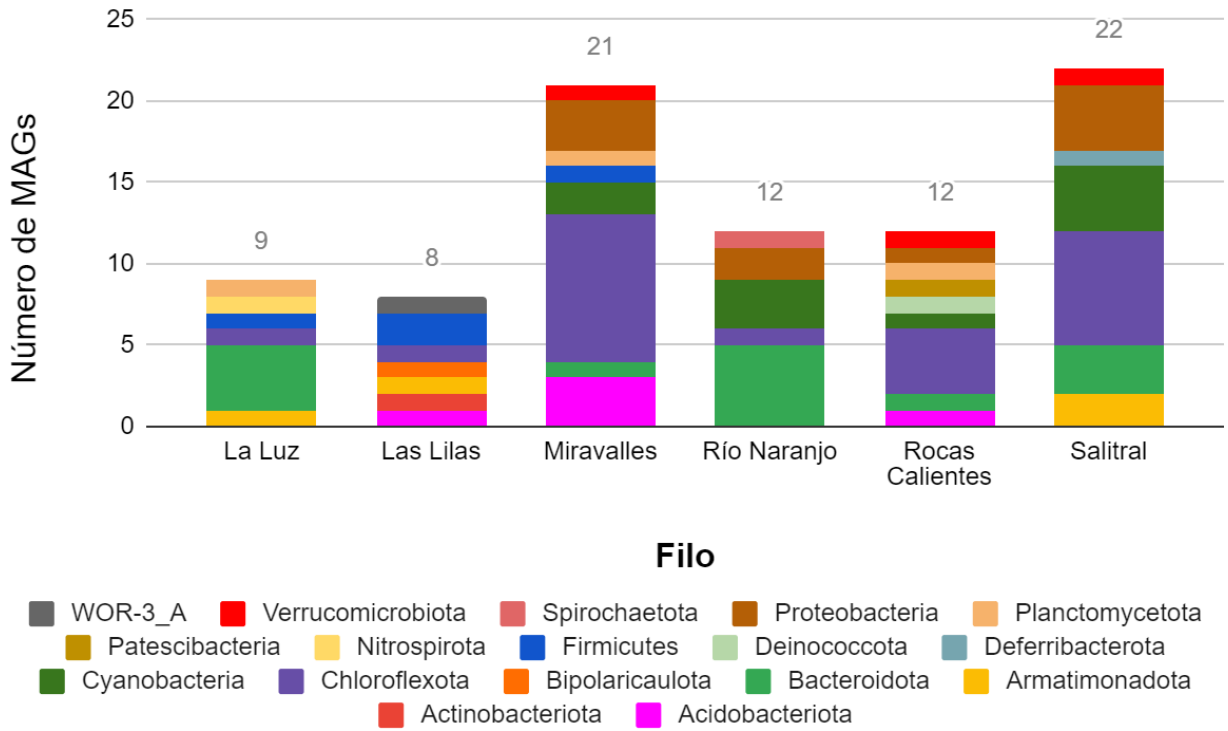


Figura 8. Cantidad de MAGs aislados de tapetes microbianos de seis fuentes termales según el filo al que pertenecen. El número encima de cada columna corresponde al total.

Cuadro 9. Resumen de los MAGs encontrados en los tapetes microbianos.

Metagenoma	MAG	Filo	GTDB-TK (última afiliación taxonómica)	Tamaño (pb)	Check M	
					% Completitud	% Redundancia
Miravalles	MAG_20	Acidobacteriota	Bryobacterales	3918455	62.51	8.72
Miravalles	MAG_2	Acidobacteriota	<i>Chloracidobacterium</i>	4098370	92.08	5.91
Miravalles	MAG_29	Acidobacteriota	Pyrinomonadales	5838842	94.87	5.56
Las Lilas	MAG_10	Acidobacteriota	<i>Blastocatellia</i> _HR10	3124095	96.58	3.63
Rocas Calientes	MAG_16	Acidobacteriota	Thermoanaerobaculia_UBA5704	2597779	92.95	3.7
Miravalles	MAG_79	Acidobacteriota	Bryobacteraceae	4051236	98.29	2.56
Las Lilas	MAG_5	Actinobacteriota	Actinobacteriota_UBA4738	1648098	88.03	0
Salitral	MAG_21	Armatimonadota	Fimbriimonadales	2355759	73.64	3.68
Salitral	MAG_14	Armatimonadota	Armatimonadota_UBA878	3022322	93.98	0.93
La Luz	MAG_8	Armatimonadota	Fimbriimonadales	2529723	93.21	2.78
Las Lilas	MAG_13	Armatimonadota	Armatimonadota_HRBIN17	3279848	93.53	0.65
Río Naranjo	MAG_17	Bacteroidota	<i>Algoriphagus</i>	3913899	94.8	2.9
Río Naranjo	MAG_16	Bacteroidota	Bacteroidia	2969208	91.09	2.39
La Luz	MAG_52	Bacteroidota	Rhodothermaceae	3960085	97.74	0.87
Miravalles	MAG_27	Bacteroidota	Bacteroidota_UBA10030	3612964	95.08	5.1
Salitral	MAG_46	Bacteroidota	Bacteroidia	2387108	89.25	1.47
Río Naranjo	MAG_26	Bacteroidota	<i>Raineya orbicola</i>	3175387	98.18	5.71
Salitral	MAG_47	Bacteroidota	Bacteroidia	3128688	99.45	6.38

La Luz	MAG_56	Bacteroidota	<i>Raineya</i>	3115527	98.18	1.46
La Luz	MAG_62	Bacteroidota	Bacteroidota_koll-22	2492654	99.63	0.54
Río Naranjo	MAG_27	Bacteroidota	<i>Thermoflexibacter</i>	4796393	96.34	0.56
La Luz	MAG_81	Bacteroidota	Chloroherpetonaceae	2528630	84.15	1.09
Salitral	MAG_48	Bacteroidota	<i>Raineya</i>	3155388	98.33	1.11
Rocas Calientes	MAG_19	Bacteroidota	Bacteroidota	3455443	98.51	0.99
Río Naranjo	MAG_20	Bacteroidota	Bacteroidota_B1	2629224	99.05	0.71
Las Lilas	MAG_11	Bipolaricaulota	Bipolaricaulota_RBG-16-55-9	1748052	89.83	0
Miravalles	MAG_34	Chloroflexota	Anaerolineales_UBA11579	3653485	65.83	5
Salitral	MAG_2	Chloroflexota	Anaerolineae_GCA-2794515	4105636	90.45	3.92
Miravalles	MAG_69	Chloroflexota	Anaerolineae_UBA6663	3208546	86.68	6.36
Rocas Calientes	MAG_25	Chloroflexota	Thermoflexales	4330212	94.9	0.91
Rocas Calientes	MAG_1	Chloroflexota	Caldilinea_GCA-2794515	5426742	98.18	6.92
Río Naranjo	MAG_12	Chloroflexota	Anaerolineae_SBR1031	4432444	97.27	1.92
Salitral	MAG_10	Chloroflexota	<i>Chloroflexus</i>	5280726	99.69	7.55
Salitral	MAG_23	Chloroflexota	Thermoflexales	4807199	96.74	2
Salitral	MAG_7	Chloroflexota	Anaerolineae_UBA5195	4379901	99.09	7.64
La Luz	MAG_53	Chloroflexota	Dehalococcoidia_G233	2753947	85.81	5.94
Rocas Calientes	MAG_12	Chloroflexota	Anaerolineae_CP2-2F	3614319	94.55	2.73
Salitral	MAG_25	Chloroflexota	Anaerolineae_Caldilinea	5909535	99.09	3.9
Miravalles	MAG_24	Chloroflexota	Anaerolineae_SBR1031	4419367	96.82	0
Salitral	MAG_26	Chloroflexota	<i>Roseiflexus</i>	5227688	98.74	0.47
Miravalles	MAG_80	Chloroflexota	Thermoflexales	4521693	96.82	0.91

Las Lilas	MAG_15	Chloroflexota	Anaerolineales_envOPS12	2490690	90.91	2.29
Miravalles	MAG_5	Chloroflexota	<i>Roseiflexus</i>	5385360	99.69	0.47
Salitral	MAG_19	Chloroflexota	Anaerolineae_OLB15	4131153	88.18	0.18
Rocas Calientes	MAG_26	Chloroflexota	Thermoflexales	2476037	92.27	0
Miravalles	MAG_3	Chloroflexota	Anaerolineae_SBR1031	5078292	92.73	0
Salitral	MAG_58	Cyanobacteria	Sericytochromatia_GCA-2770975	4035386	88.62	6.32
Río Naranjo	MAG_13	Cyanobacteria	<i>Synechococcus_A</i>	3429613	99.53	1.67
Salitral	MAG_57	Cyanobacteria	<i>Geminocystis</i>	4226853	98.03	2.26
Río Naranjo	MAG_19	Cyanobacteria	<i>Geminocystis</i>	3741656	97.6	1.42
Miravalles	MAG_33	Cyanobacteria	Leptolyngbyaceae	5511802	99.29	0
Salitral	MAG_9	Cyanobacteria	Elainellaceae_O-77	4847523	98.58	0.71
Salitral	MAG_3	Cyanobacteria	<i>Synechococcus_A</i>	2778176	97.64	0
Miravalles	MAG_32	Cyanobacteria	Phormidesmiales	4486119	98.58	0.71
Rocas Calientes	MAG_10	Cyanobacteria	<i>Synechococcus_JA-3-3Ab</i>	2921566	100	0
Salitral	MAG_49	Deferribacterota	<i>Calditerrivibrio</i>	2452381	99.07	10
Rocas Calientes	MAG_9	Deinococcota	<i>Thermus</i>	1912841	100	0.42
Las Lilas	MAG_6	Firmicutes	Firmicutes_CSP1-3	3472898	93.52	2.04
La Luz	MAG_74	Firmicutes	Bacillaceae_QFRU01	4140517	84.46	5.26
Miravalles	MAG_39	Firmicutes	<i>Bacillus</i>	4348778	98.67	1
Las Lilas	MAG_3	Firmicutes	Firmicutes_HRBIN32	2689334	93.52	1.85
La Luz	MAG_71	Nitrospirota	Thermodesulfovibrionales	2190319	98.18	1.82
Rocas Calientes	MAG_5	Patescibacteria	Microgenomatia_GWC2-37-13	906433	69.96	1.72
Rocas Calientes	MAG_14	Planctomycetota	Gemmataceae	4483130	86.69	5.58

Miravalles	MAG_16	Planctomycetota	Phycisphaerae_UBA1845	3264409	98.8	0
La Luz	MAG_51	Planctomycetota	Phycisphaerales_SM1A02	2722971	97.73	0
Río Naranjo	MAG_8	Proteobacteria	Rhodocyclaceae_Par-f-2	2102490	61.08	1.65
Miravalles	MAG_94	Proteobacteria	Xanthobacteraceae_Z2-YC6860	2442296	63.84	2.93
Rocas Calientes	MAG_7	Proteobacteria	Geminicoccaceae	4043285	98.51	2.49
Río Naranjo	MAG_4	Proteobacteria	<i>Gemmobacter</i>	4048689	98.94	0.76
Miravalles	MAG_51	Proteobacteria	Geminicoccaceae	4111005	96.07	1.99
Miravalles	MAG_48	Proteobacteria	Gammaproteobacteria_RBG-16-66-20	3590417	98.78	6.57
Salitral	MAG_27	Proteobacteria	<i>Rubrivivax</i>	3368313	97.66	0.45
Salitral	MAG_30	Proteobacteria	Burkholderiaceae	2774922	88.63	0.58
Salitral	MAG_33	Proteobacteria	Kordiimonadaceae	2691136	100	0
Río Naranjo	MAG_22	Spirochaetota	Leptospiraceae	3734004	97.45	0.7
Rocas Calientes	MAG_13	Verrucomicrobiota	Pedosphaerales	3833021	95.72	2.76
Miravalles	MAG_76	Verrucomicrobiota	Pedosphaerales_UBA11320	5208194	98.65	2.03
Salitral	MAG_13	Verrucomicrobiota	Methylacidiphilales	2265823	95.27	1.35
Las Lilas	MAG_36	WOR-3_A	Hydrothermaceae	1630143	91.53	0
Miravalles	MAG_31	Chloroflexota	Anaerolineales_OLB14	3242556	98.18	0.91

Comparación del perfil taxonómico de las diferentes fuentes termales

Una vez obtenido el ensamblaje, se comparó el perfil taxonómico de las diferentes muestras para comparar las comunidades microbianas de las muestras de Costa Rica y las de otros sitios alrededor del mundo. Los resultados obtenidos después del análisis en la plataforma Kraken a nivel de filo se muestran en la figura 9 los resultados de MetaPhlan2 y MG-RAST (Figura 1 en anexos). Según Kibegwa et al., (2020), Kraken y MG-RAST generan resultados comparables, sin embargo, Kraken genera una clasificación más acertada, y con MG-RAST se obtienen más secuencias “no clasificadas”. Wood y Salzberg (2014) mencionan que Kraken proporciona un 100% de precisión en la clasificación taxonómica, sin embargo, solamente logra clasificar el 89% de las secuencias. Estos mismos autores señalan que la categorización a nivel de filo es comparable entre ambas herramientas. A nivel de género Kraken clasifica más secuencias y en algunos casos puede no coincidir con los resultados obtenidos en MG-RAST.

En este estudio, la mayoría de los análisis de la abundancia de los filios no coinciden entre las diferentes plataformas, en algunos casos se presentan porcentajes de abundancia que difieren hasta en un 30% (anexos figura 1 y 2). En el caso de las muestras costarricenses, los resultados que se generaron con MG-RAST y Kraken son los más similares a los reportados previamente según análisis del amplicón ARNr 16S (Uribe-Lorío et al., 2019 Brenes-Guillén et al., datos no publicados). Por el contrario, en MetaPhlan2 se observa un aumento considerable de la abundancia del filo Chloroflexi, por lo que probablemente se encuentra sobreestimado, mientras que otros grupos como Cianobacteria desaparecen.

Basado en los resultados obtenidos, se decidió hacer el análisis y comparación de la estructura de las comunidades usando Kraken. Este análisis muestra que las fuentes termales costarricenses son 40% similares a las de Chile y Nueva Zelanda. Según el análisis de BEST, el filo Firmicutes contribuye al 85% de la variación entre las muestras, siendo más abundante en las muestras de Obsidian Pool, PNY. Las Proteobacterias y Aquificae contribuyen un 5% a la variación entre las muestras. El filo Aquificae es predominante en Liberty Cap, mientras que la abundancia de Proteobacterias es mayor en las muestras de Nueva Zelanda, Chile y Costa Rica (Fig. 9). En las fuentes termales con temperaturas mayores a los 50 °C aumenta la abundancia de filios como Chloroflexi y Deinococcus-Thermus (Fig. 9).

Todas las muestras que corresponden a tapetes microbianos (Chile y Costa Rica) se agrupan juntas con un 60% similitud, junto a las muestras de Nueva Zelanda que provienen de agua filtrada. Las fuentes termales de Rocas Calientes y La Luz que se caracterizan por tener una temperatura de 63 °C y 50 °C y son un 90% similar a las muestras de Porcelana, Chile, las cuales poseen una temperatura que varía entre los 48 °C y los 66 °C, siendo La Luz similar al tapete microbiano colectado en la temperatura más alta y Rocas Calientes a las tomadas a los 48 °C y 58 °C (Fig. 10). Las fuentes termales de Miravalles, Salitral y Río Naranjo tienen hasta un 80% de similitud con muestras de Nueva Zelanda. Las muestras Mound Spring, PNY y Las Lilas se agrupan juntas (similitud del 80%) teniendo en común la coloración anaranjada (indicador de la presencia de Chloroflexi y/o cianobacterias sometidas a largos períodos de radiación) y alta conductividad, lo que puede estar relacionado a altas concentraciones de sales (Alcamán et al., 2018). Crenarchaeota y Euryarchaeota comprenden menos de 2% en ambas muestras (Fig. 9). La comparación de la estructura de la comunidad a nivel de filo encontró que las muestras de Obsidian Pool y Liberty Cap son muy diferentes al resto de comunidades, esto probablemente relacionado a la temperatura, a las condiciones fisicoquímicas y a que estas muestras no provienen de un tapete microbiano, sino de sedimentos y agua. Los grupos de arqueas, Thermotoga y Aquificae se comportan igual en términos de abundancia en las muestras del PNY. Por otro lado, la diferencia entre el grupo de muestras de Chile y las de Nueva Zelanda está relacionada a la presencia de Deinococcus.

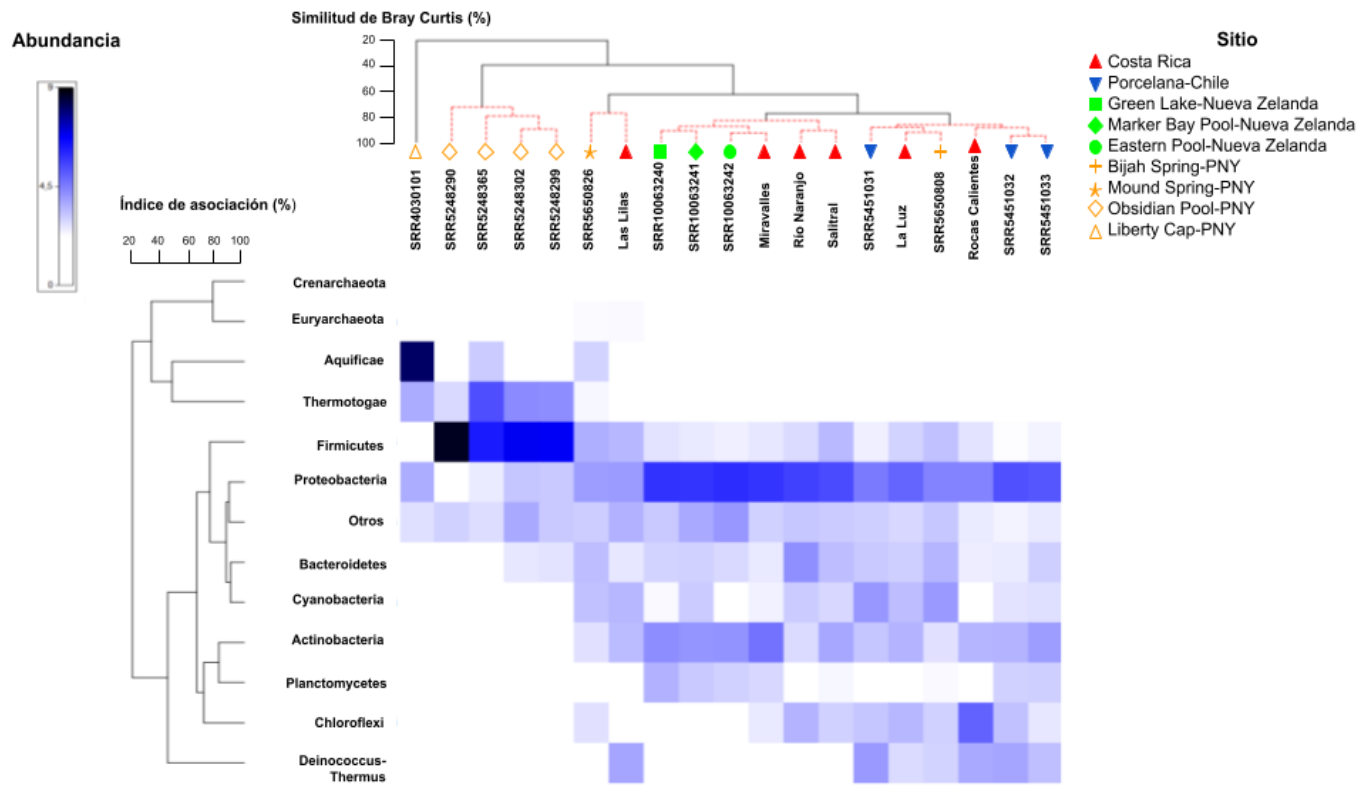


Figura 9. Diagrama de calor (Primer 7) que muestra las relaciones globales entre grupos de muestras. La intensidad del azul representa la abundancia de los filios presentes en las muestras. Las líneas del cladograma representan las agrupaciones estadísticamente significativas (vertical en color negro, nivel 1% / prueba de permutación Primer SIMPROF), mientras que las líneas rojas horizontales punteadas representan las agrupaciones que no son estadísticamente significativas.

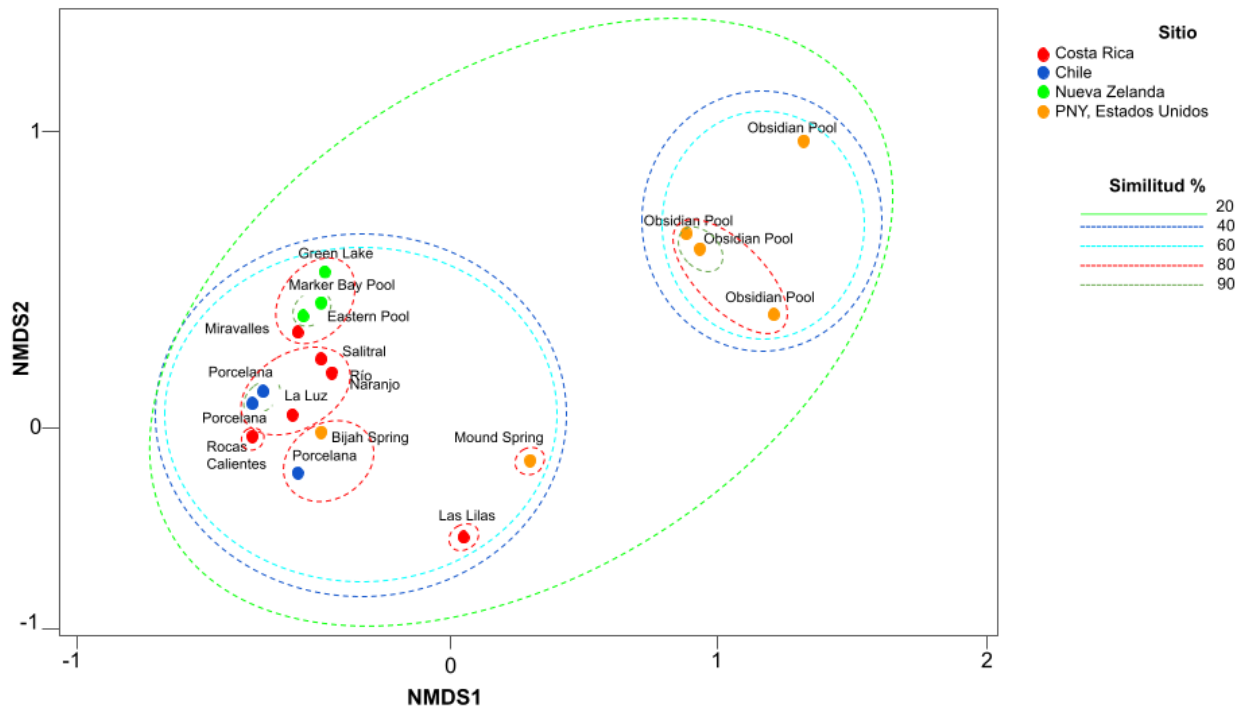


Figura 10. Ordenación de escala multidimensional no métrica (NMDS) de las comunidades bacterianas a nivel taxonómico de filo de las fuentes termales. Los círculos coloreados corresponden al porcentaje de similitud entre las muestras.

A continuación, se detallan y discuten los resultados obtenidos en cada una de las fuentes termales.

Fuentes termales de Costa Rica

En Rocas Calientes (63 °C y pH=6) se encontró que los filos Chloroflexi y Deinococcus-Thermus comprenden 53.4% de la comunidad. Estos resultados son menores a los mostrados por Uribe-Lorío y colaboradores (2019) en donde utilizaron el análisis del amplicón 16S ARNr, y encontraron que Chloroflexi y Deinococcus-Thermus comprendían el 84% de las secuencias. Rocas Calientes es una fuente termal muy particular, aunque Chloroflexi se encontró en otras fuentes termales, la abundancia en este sitio es significativamente mayor (Fig. 9). Chloroflexi es un grupo abundante en tapetes microbianos de fuentes termales (Urbietta et al., 2015, Wang et al., 2013), generalmente su abundancia aumenta conforme se incrementa la temperatura. Según los resultados, los géneros más abundantes son *Chloroflexus* y *Roseiflexus*. Las cianobacterias tienen una abundancia menor al 5%, siendo los géneros unicelulares como *Synechococcus*, *Microcystis* y *Cyanothece* los más abundantes. Aunque se encontraron secuencias asociadas a cepas filamentosas con *Leptolyngbya* y *Fischerella*, estas poseían una baja abundancia.

En La Luz, el filo Proteobacteria tiene un porcentaje de abundancia del 44.3%, Actinobacteria 11.4%, Chloroflexi 10.7%, Cianobacteria un 9.6% y Firmicutes un 16.0%. En Salitral por otro lado, las proteobacterias son las más abundantes con un 53% del total, seguido por las Actinobacterias con un 12.4%, luego Chloroflexi con 5.3% y Cyanobacteria comprende solamente el 4.5 % de la abundancia.

Las Lilas se caracteriza por tener un tapete microbiano de color rojo en algunos sitios, con una temperatura de 74 °C y pH de 6. utilizando Kraken, los filios Aquificae y Deinococcus-Thermus tienen una abundancia de 1.7 % y 17.8 % respectivamente, siendo Proteobacteria el filo más abundante (21.6%). El filo Aquificae está compuesto por bacterias termófilas e hipertermófilas, la mayoría quimiolitotróficas, que habitan generalmente en ambientes hidrotermales marinos y terrestres (Gupta 2014). La mayoría son bacilos Gram negativos, que no forman esporas, estrictamente termofílicos y su crecimiento óptimo ocurre por encima de 65 °C (Deckert et al., 1998, Reysenbach, 2001, Eder y Huber, 2002). Por otro lado, miembros del filo Deinococcus-Thermus generalmente se encuentran en ambientes acuáticos con temperaturas entre los 50 °C y los 85 °C, y se caracterizan por ser quimiorganotróficos aerobios. En Costa Rica, se han encontrado anteriormente microorganismos dentro del filo Aquificae como *Thermocrinis minervae* en ambientes termales de Rincón de la Vieja (Caldwell et al., 2010), a cuyo complejo pertenece Las Lilas. Los géneros *Hydrogenobacter* (Aquificae) y *Thermus* (Deinococcus-Thermus) comprenden cerca del 75% de la comunidad de esta muestra en específico. Es la única muestra de Costa Rica incluida en este análisis en la que se reporta la presencia de Crenarchaeota y Euryarchaeota (menor al 2%).

La muestra de Río Naranjo se caracteriza por tener una temperatura de 34 °C y pH cercano a 7. Según los resultados obtenidos en este análisis, las Proteobacterias comprenden el 55.2% del tapete microbiano, mientras que Chloroflexi y Cianobacteria tienen una baja abundancia (8.9% y 5.5% respectivamente).

Porcelana, Chile

Las tres muestras que corresponden a esta fuente termal se tomaron de tapetes microbianos que crecían a 66 °C, 58 °C y 48 °C. Análisis previos han demostrado que los filios Cyanobacteria y Chloroflexi son predominantes, siendo Cyanobacteria más abundante a la temperatura más baja (48 °C), resultados similares a los encontrados en las muestras costarricenses. Adicionalmente se han encontrado 51 filios, aunque en poca abundancia (Alcamán-Arias et al., 2018). Estas fuentes termales son similares (90%) a las muestras de

La Luz y Rocas Calientes (50 °C y 63 °C, respectivamente), las cuales también corresponden a un tapete microbiano.

Las aguas termales de Porcelana se encuentran en la Patagonia chilena al final del fiordo Comau y es parte de un área de eventos geotérmicos, rodeado por volcanes Cuaternarios activos como el Volcán Huequi (Duhart et al., 2000). Las paredes de esta fuente están colonizadas por esteras microbianas de color anaranjado y verde. Según estudios previos, las aguas termales están dominadas por cianobacterias, particularmente el género diazotrófico *Mastigocladus/Fischerella* (Stigonematales) y *Synechococcus* (Alcamán et al., 2015). Estos géneros junto a *Chloroflexus* fueron de los más abundantes según Kraken y MetaPhlan2. Mackenzie et al., (2013) mencionan que también se pueden encontrar microorganismos dentro del filo Proteobacteria, no obstante, en ninguno de los estudios reportados para esta fuente termal se han reportado valores de abundancia de hasta el 30% al 50% para este filo.

Mound y Bijah, PNY

Mound Spring corresponde a una muestra de sedimento rico en sílice cuya agua se encuentra a 43.3 °C. Generalmente los sedimentos cercanos a esta fuente termal poseen tapetes microbianos de color anaranjado rojizo y el agua puede tener temperaturas que van desde los 25 °C a los 90 °C, pH mayor a 8 y una conductividad mayor a 1300 $\mu\text{S}/\text{cm}$ (Reporte Universidad de Montana). Según estudios previos, el tapete microbiano de color anaranjado en esta fuente termal está dominado por cianobacterias como *Phormidium* y *Synechococcus* (Inagaki et al., 2001). En este estudio, utilizando Kraken, se encontró que las Proteobacterias son las más abundantes (21.45%), seguido de Firmicutes (15.4%), Bacteroidetes (11.44%) y Cyanobacteria (10.59%). No obstante, la abundancia relativa de estos grupos varía considerablemente cuando el análisis taxonómico se hace en Metaphlan2, siendo las cianobacterias el grupo más abundante (68% del total), seguido por Firmicutes (14.85%). Las Proteobacterias y arqueas fueron encontradas en menos del 1% (Fig 9).

Bijah Spring está ubicada a lo largo del corredor Mammoth Norris en el Parque Nacional Yellowstone, la muestra utilizada en este análisis corresponde a sedimento. Yu et al., (2017) mostraron que Chlorobi, Deferribacteres y Proteobacteria son los grupos más abundantes, a diferencia de los resultados encontrados en este análisis en donde Proteobacteria, Cyanobacteria y Bacteroidetes dominan. La comunidad bacteriana de esta fuente termal es similar a la encontrada en Porcelana, Rocas Calientes y La Luz utilizando Kraken.

Obsidian Pool, PNY

Obsidian Pool es una fuente termal persistente ubicada al sur de Lago Goose en el Área del Volcán Mud, el cual está localizado al suroeste del domo Sour Creek en la caldera del PNY, posee una profundidad que varía desde 10 cm hasta 100 cm, una temperatura desde los 42 °C a los 90 °C y un pH de 6.7 ± 0.4 . Está influenciada por gases magmáticos, como dióxido de carbono, hidrógeno y sulfuro de hidrógeno. Los estudios microbiológicos de muestras de sedimentos y agua de esta fuente termal empezaron en 1993 con los estudios de Barns y colaboradores (1994, 1996). Posteriormente se desarrollaron otras investigaciones de cultivo y caracterización de arqueas y bacterias (Hubert et al., 1995, Hugenholtz et al., 1998, Burggraf et al., 1997, 2001).

El filo Firmicutes es el más abundante en las cuatro muestras de este sitio (58% al 96%), sin embargo, aparecen otros grupos como Proteobacteria con abundancias entre el 1% y el 31% (Fig. 9). En Kraken se encontraron otros filos, sin embargo, su abundancia es menor al 5%. Es importante mencionar, que existen muy pocos metadatos asociados específicamente a estos metagenomas (Cuadro 3). Entre los géneros que tienen una mayor abundancia se encontraron *Caldicellulosiruptor obsidiansis* (aislado por primera vez en este sitio), *Thermoanaerobacter ethanolicus* y *Caldicellulosiruptor owensensis* (Firmicutes anaerobios, celulíticos), algunos de los cuales se han reportado previamente en esta fuente termal (Vishnivetskaya et al., 2015).

El resultado de este análisis mostró la dominancia del filo Firmicutes, lo cual difiere de los resultados mostrados por Graber y colaboradores (2001), en el cual determinaron que los Aquificales es el filo predominante. Estos resultados fueron corroborados en 2015 por Vishnivetskaya y colaboradores, en muestras del 2007 y 2008. Las muestras analizadas en este estudio fueron colectadas en 2011 y 2012, sin embargo, no se especifica el sitio exacto de colecta, por lo que es posible que sean de otros puntos diferentes a los utilizados en las investigaciones mencionadas. Sin embargo, estos autores mencionan que la fuente OP10 (68 °C, pH = 5), contiene una gran abundancia de Firmicutes los cuales podrían participar en diferentes etapas de degradación de la celulosa. Algunos estudios en microcosmos han demostrado que, al incrementar la temperatura de los sedimentos de esta fuente termal, aumenta la abundancia de Firmicutes, los cuales actúan como degradadores de biomasa a partir de plantas como *Juncus tweedyi*, la cual crece cerca de la fuente (Vishnivetskaya et al., 2015). Fishbain y colaboradores (2003) aislaron bacterias de los géneros *Thermodesulfobacterium* (Thermodesulfobacteria) y *Thermodesulfovibrio* (Nitrospirae). También se han realizado aislamientos de hipertermofilos ferredutores tales como

Geothermobacterium ferrireducens (Kashefi et al., 2003), reductores de sulfato (Hamilton-Brehm et al., 2013) y bacterias degradadoras de celulosa (Vishnivetskaya et al., 2015).

Liberty Cap, Mammoth springs, PNY.

En la muestra que corresponde a Liberty Cap (72 °C y pH 6-7) se encontró que el filo Aquificales es predominante, representando más del 90% en ambos análisis (Fig. 9). Proteobacteria comprende el 3.76% de la comunidad según Kraken. *Sulfurihydrogenibium* es el género más abundante en esta fuente termal (encontrado en menor cantidad en Obsidian). Son microorganismos quimiolitioautótrofos estrictos capaces de utilizar H₂, tiosulfato, azufre elemental, sulfito, Fe (II), arsenito y selenito como donantes de electrones y oxígeno molecular, nitrato, Fe (III), arsenato y selenato como aceptores de electrones (Nakagawa et al., 2005). Además, tiene un papel importante en el ciclo del azufre (Nakagawa y Fukui, 2003) y la mineralización de hierro en ambientes de aguas termales en Rusia y PNY principalmente (Reysenbach et al., 1999). Liberty Cap forma parte del complejo de fuentes termales de Mammoth en el PNY. Está formada por una cresta de depósitos de travertinos que son respiraderos de aguas termales producto de una fractura de la corteza. El agua emerge de un solo lugar a lo largo de una grieta en la terraza propiciando el crecimiento del cono (Bargar et al., 1978). Este cono se considera prehistórico, durante muchos años, la presión interna fue lo suficientemente alta como para empujar el agua a una gran altura, lo que permitió que se acumularan los depósitos minerales que se pueden ver hoy en día. La muestra de este análisis corresponde a aguas que fluyen en este sitio. Según lo consultado en la literatura no existen reportes científicos de microorganismos aislados de esta fuente o una caracterización de la comunidad microbiana.

Isla Raoult, Nueva Zelanda

Se utilizaron tres muestras, todas se encuentran en la isla Raoult en Nueva Zelanda: Green Lake, Marker Bay Pool y Eastern Pool. La temperatura y el pH de cada muestra se encuentran en entre 27 °C y 35 °C y pH cercanos a 6.5, según la metodología reportada, estas muestras corresponden a agua de la fuente filtrada. Raoult es una isla volcánica que se encuentra aproximadamente a 1000 km al noreste de Nueva Zelanda. Su caldera (Rim) contiene un lago volcánico y varias fuentes termales asociadas, todas alimentadas por un sistema hidrotermal procedente de agua meteórica y agua de mar (Stewart et al., 2018). Estas muestras son 60% similares en su estructura microbiana a los tapetes microbianos encontrados en Miravalles, Río Naranjo y Salitral (Fig. 10). Proteobacteria y Actinobacteria

comprenden cerca del 75% del total de la comunidad. Estudios anteriores han demostrado que Chloroflexi y Planctomycetes son muy abundantes en estas fuentes termales, mientras que las cianobacterias tienen una baja abundancia (Stewart et al., 2018).

Basado en los resultados descritos de los ambientes de fuentes termales, se concluye que la composición microbiana a nivel de filo de las fuentes termales de Costa Rica es similar a otras alrededor del mundo con características fisicoquímicas similares, sin embargo mantienen cierto grupo de microorganismos únicos, lo que apoya la teoría de que estos ambientes pueden ser considerados como islas. Es muy importante considerar que son comunidades dinámicas y cambiantes, que se ven influenciadas directamente por los cambios en las características fisicoquímicas del agua, y la estacionalidad, por lo que es necesario incluir estudios que involucren un monitoreo en el tiempo de las comunidades para tratar de caracterizarlas mejor. Asimismo, es importante determinar los microorganismos residentes o pasantes transitorios, para ellos se pueden realizar estudios espacio-temporales y utilizar diversos *primers* para amplificar diferentes regiones del ADN para clasificación taxonómica, entre otros.

Perfil funcional de las comunidades de las fuentes termales de Costa Rica, PNY, Chile y Nueva Zelanda.

Se encontraron 2448 secuencias de proteínas clasificadas dentro de las categorías COG según los resultados de Prokka, lo que representa entre el 25% y el 30% del total de secuencias encontradas en cada fuente termal, ya que la mayoría de las secuencias no tienen ninguna categoría de COG asociada. Se encontraron muy pocas secuencias asociadas a las categorías Modificación y procesamiento de ARN (A), Estructuras extracelulares (W) y Citoesqueleto (Z).

Rocas Calientes y Las Lilas son las muestras que más difieren en cuanto a su perfil funcional, como se mencionó anteriormente Rocas Calientes posee un tapete microbiano con una temperatura de 63 °C dominado por Chloroflexi y Deinococcus-Thermus, lo que podría ocasionar diferencias con las otras fuentes termales dominadas por Proteobacteria y Cianobacteria. Por otro lado, Las Lilas es una fuente termal muy particular con alta temperatura y conductividad (sugiere alta concentración de iones), puede ocasionar diferencias en cuanto a características metabólicas, reduciendo los COGs que se encontraron. Para esta fuente termal se encontraron dos COGs únicos: COG4237, que está

relacionado con la enzima hidrogenasa 4, y el COG4853, relacionado con Yycl, la cual es una enzima histidina quinasa.

Las fuentes termales de OpSIDIAN Pool no comparten la mayoría de COGs con las otras fuentes termales, no obstante, estas muestras tienen una comunidad bacteriana dominada por Firmicutes, lo que puede estar generando estas diferencias debido al perfil de los microorganismos presentes. Si se excluyen las muestras de OpSIDIAN Pool y Liberty Cap del PNY y se analiza solamente las muestras de Costa Rica, Chile y Nueva Zelanda (tienen un 60% de similitud según el perfil taxonómico), se encuentra que existen 1207 secuencias de proteínas en todas esas fuentes termales, de las cuales 203 se encuentran dentro del grupo de Señalización y procesamiento celular, 550 en Metabolismo, 246 en Almacenamiento y procesamiento de la información, 26 Poco caracterizadas (categorías R y S) y 2 de Mobiloma (profagos y transposones). Estas secuencias se podrían considerar como parte de la comunidad de proteínas típicas de estos tapetes microbianos y muestras de suelo de las fuentes termales con temperaturas de 27 °C a los 74 °C y pH cercano a 7

Es necesario ampliar el uso de otras bases de datos como Interprot, Swiss-Prot y TrEMBL para determinar el perfil funcional de las comunidades, sin embargo, los datos de COG podrían estar reflejando el comportamiento general de las comunidades. Cabe mencionar que algunas de las secuencias clasificadas como hidrolasas (según Pfam) utilizadas en la primera parte del análisis no tenían un COG asociado, por lo que es necesario utilizar estrategias complementarias para mejorar los resultados. De hecho, casi el 75% de las secuencias no tuvieron COG asignado, mientras que utilizando KEGG, la asignación a algunas de las categorías generalmente no era mayor al 50%. Además, según Pfam cerca del 50% de las secuencias se encontraban dentro de la categoría de putativas por lo que podrían ser proteínas poco estudiadas o con función desconocida.

Según Sommer y Salzberg (2012), todas las herramientas de software actuales predicen cientos o miles de genes "adicionales", es decir, genes que no coinciden con ningún gen con una función conocida y que generalmente reciben el nombre de "proteína hipotética". Muchos de estos genes hipotéticos probablemente representan secuencias codificantes de proteínas, pero muchos otros pueden ser predicciones falsas positivas, lo que puede crear problemas para los análisis posteriores de la función del genoma.

Para futuras investigaciones, se pueden utilizar estrategias para la identificación de la función de las secuencias de genes "putativas", entre estas i) métodos de clasificación basados en atributos de las secuencias de aminoácidos tales como composición, tamaño y propiedades fisicoquímicas, utilizando paradigmas de *machine learning* como modelos de

aprendizaje supervisados o no supervisados, y ii) contexto génico, basado en el supuesto de que los genes en posiciones similares van a tener la misma función (Bernardes y Pedreira, 2013, Yang et al., 2019).

Identificación y clasificación de las enzimas hidrolíticas

Una vez obtenido el ensamblaje de cada metagenoma, se procedió a realizar la anotación funcional de los mismos. Los resultados generales de la anotación de los metagenomas generado por Prokka se muestran en el cuadro 10. La cantidad de anotaciones encontradas se relaciona con el número de contigs de cada metagenoma, por lo que a mayor cantidad de *contigs* y mayor tamaño del metagenoma, mayor número de CDS.

Cuadro 10. Resultados generales de la anotación de los metagenomas utilizando Prokka.

Ubicación geográfica	Metagenoma	Contigs	Bases	CDS	ARNr	ARNt	ARNmt	Regiones repetidas
Costa Rica	Rocas Calientes	2888	12461302	11671	20	211	2	7
	La Luz	28202	185831358	172071	194	2752	52	141
	Miravalles	54833	258737108	238398	205	3003	45	88
	Las Lilas	9010	69198934	68248	105	1275	20	96
	Salitral	27726	173688820	159161	142	2300	46	99
	Río Naranjo	13954	80441143	71979	79	880	16	54
Chile	SRR5451033	24987	113471723	107355	78	1237	23	49
	SRR5451032	15183	81888823	77309	58	978	20	51
	SRR5451031	16388	70313291	67707	59	826	16	37
Nueva Zelanda	SRR10063240	77000	363523800	342579	302	3971	63	88
	SRR10063241	94457	437160376	393515	299	4013	63	93
	SRR10063242	75035	347893297	358284	209	2786	37	49
Mound Spring-PNY	SRR5650826	44554	125576690	121645	241	2095	34	99
Bijah Spring-PNY	SRR5650808	28439	103487621	102863	130	1679	35	92
Liberty Cap-PNY	SRR4030101	3650	20408352	19393	35	405	7	59
Obsidian Pool-PNY	SRR6208290	433	5382829	5043	12	104	3	25
	SRR5248299	1478	15106920	14482	29	313	7	46
	SRR5248365	888	6805706	6323	21	156	3	25
	SRR5248302	1731	10467517	9717	35	335	7	39

Se utilizaron los datos de las fuentes termales de Costa Rica, PNY, Nueva Zelanda y Chile para la construcción de la base de datos de las enzimas hidrolíticas tales como hidrolasas, proteasas, lipasas, esterases y amilasas. El número de secuencias en cada una de las categorías se muestra en el cuadro 11. Es importante recalcar que solamente se tomaron en cuenta las enzimas clasificadas en la categoría EC3. Las peptidasas son el grupo de enzimas hidrolíticas más abundante, en todas las muestras representan más del 50% del total, seguido por las hidrolasas y las esterases (~ 25 %). Se realizó una matriz de similitud pareada basado en la distancia p , entre las secuencias de proteínas y de nucleótidos de cada categoría, considerando que cuanto mayor es la distancia p , más divergentes son las secuencias. En todos los casos, la mayoría de los valores se encontraron entre 0.5 y 0.9, lo que indica que son muy diferentes entre sí.

Cuadro 11. Número total de variantes/alelos en cada metagenoma según la clasificación de la enzima (bajo la categoría EC3).

	Metagenoma	Peptidasas	Lipasas	Esterasas	Amilasas	Hidrolasas
Costa Rica	Rocas Calientes	166	1	47	2	77
	La Luz	2395	43	652	27	1261
	Miravalles	3217	41	866	23	1673
	Las Lilas	921	10	209	5	420
	Salitral	2159	35	610	24	1251
	R.Naranja	834	17	253	8	512
Obsidian Pool	SRR6208290	73	2	38	2	34
	SRR5248299	202	4	90	3	97
	SRR5248365	96	2	49	1	39
	SRR5248302	144	2	71	1	65
Chile	SRR5451033	1191	20	388	9	762
	SRR5451032	896	17	273	9	587
	SRR5451031	765	14	233	14	422

Nueva Zelanda	SRR10063240	3298	84	1140	33	2501
	SRR10063241	3449	94	1341	37	2712
	SRR10063242	2389	72	915	26	1909
Mound Spring-PNY	SRR5650826	1288	15	410	17	617
Bijah Spring-PNY	SRR5650808	1188	11	41	10	610
Liberty Cap-PNY	SRR4030101	271	8	75	8	119
	Total	24942	492	6657	259	15668

La base de datos de secuencias de proteínas agrupadas en cada tipo de enzima y depurada mediante MMseqs fue utilizada para mapear las secuencias en los diferentes metagenomas. El mapeo de las diferentes variantes se llevó a cabo solamente en las muestras costarricenses. La abundancia (número de secuencias mapeadas) de las diferentes variantes de cada gen cambia según la fuente termal, por lo que no se encontraron alelos con la misma abundancia entre las muestras. En la figura 11, se muestran todas las variantes de las enzimas dentro de las categorías que fueron identificadas como representativas después del análisis con MMseqs. Como se puede observar, algunos de los alelos son únicos de cada fuente termal, sin embargo, la mayoría de los alelos de cada gen se encuentran en todos los tapetes, aunque su abundancia difiere.

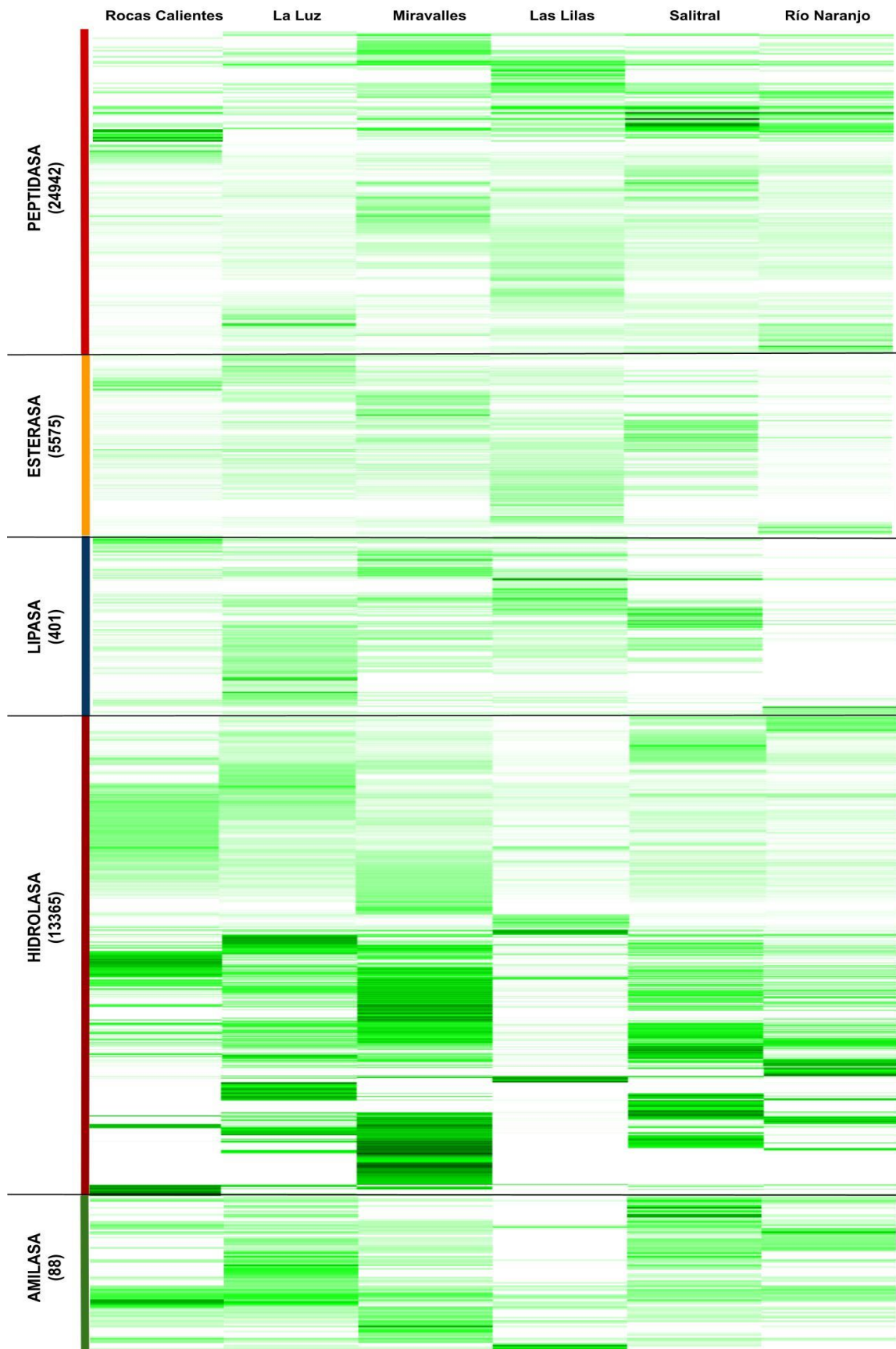


Figura 11. Mapa de calor de la abundancia estandarizada de cada una de las variantes encontradas en los tapetes microbianos. El número entre paréntesis corresponde al número de secuencias en cada categoría.

Las variantes alélicas de las enzimas hidrolíticas encontradas en este estudio son similares a las encontradas en bases de datos como UniProt. Es importante mencionar que, si dos secuencias de proteínas tienen una identidad superior al 70%, tienen aproximadamente un 90% de probabilidad o más de compartir el mismo proceso biológico para los niveles de índice de GO (Joshi y Xu, 2007), por lo que podrían tener similar estructura y función. No obstante, según Pearson (2013), dos proteínas que son más del 80% idénticas en los sitios activos conservados compartirán la misma función, pero es difícil inferir de manera confiable una función similar a distancias evolutivas mucho mayores. Este mismo autor señala que el porcentaje de identidad proporciona una aproximación útil a la distancia evolutiva, sin embargo, no es lineal con el porcentaje de identidad, por lo que un cambio de similitud del 80% al 70% podría reflejar una divergencia 200 millones de años, pero el cambio del 30% al 20% podría corresponder a un cambio de tiempo de divergencia de mil millones de años. Otro aspecto importante para considerar es que un porcentaje de similitud alto entre secuencias no garantiza homología (ancestro en común). La inferencia de similitud funcional a partir de homología es difícil. Una regla común es que dos secuencias son homólogas si son más del 30% idénticas en toda su longitud (Joshi y Xu, 2007), no obstante, en secuencias cortas se pueden obtener regiones idénticas fácilmente.

La comparación entre proteínas utilizando solamente las secuencias pueden generar resultados o conclusiones erróneas acerca de su estructura, función y homología, sin embargo, utilizar estas herramientas son un buen acercamiento y puede ser de utilidad para futuras investigaciones. Debido a las desventajas y sesgos que se generan a partir de las inferencias de funciones biológicas utilizando comparaciones y predicciones computacionales es importante realizar estudios funcionales que permitan evaluar en el laboratorio la función de la proteína. Esto se puede llevar a cabo mediante la clonación de la secuencia de ADN, análisis de expresión, diseño de mutantes, entre otros. En este trabajo para cada uno de los grupos de enzimas se realizó una breve descripción de las categorías, búsqueda de secuencias homólogas, clasificación y abundancia.

Por otro lado, es relevante considerar que la baja especificidad de las enzimas proporcionará a los organismos de vida libre plasticidad para sobrevivir en entornos extremos (Aharoni et al., 2005), ya que estas condiciones y las variaciones ambientales propiciarán la promiscuidad de estas, lo que permite al microorganismo alterar y reprogramar sus vías metabólicas (Singla y Bhardwaj, 2019). Algunos autores han planteado que, para sobrevivir en condiciones adversas, como altas concentraciones de iones o temperatura, los microorganismos tienen enzimas capaces de utilizar diversos tipos de sustratos, lo que

sugiere una baja especificidad. La especificidad de unión se encuentra en muchas secuencias conservada, a pesar de la baja especificidad que podría tener cada proteína. Por otro lado, se ha encontrado alta especificidad de sustrato a temperaturas altas de crecimiento del organismo (mayor a 60 °C), pero se vuelve promiscua a temperaturas más bajas (Lutz et al., 2007). Por lo tanto, a pesar de que las enzimas hidrolíticas de los microorganismos que se encuentran en los tapetes microbianos tienen sustratos específicos sobre los cuales actúan, podrían participar en la degradación de otros compuestos tales como agroquímicos. A continuación, se detalla por grupo de enzimas los resultados obtenidos.

Peptidasas

La base de datos de secuencias clasificadas como peptidasas contiene 24942 secuencias. Este estudio se enfocó en dos grupos: las serina proteasa y las metaloproteinasas, ambos grupos tienen su pH óptimo cercano a 7 y fueron las más abundantes. Cabe mencionar que de los demás grupos de metaloproteínas solamente se encontraron las cisteína peptidasas, específicamente la familia C56 (residuos conservados Cys100 y His101 en *Pyrococcus furiosus*), las cuales se encuentran tanto en bacterias como en arqueas.

Proteasas de serina termofílicas

La base de datos creada tiene 77 de secuencias clasificadas como serina proteasa termofílicas, de las cuales 71 eran representativas, donde todas las variantes tuvieron secuencias mapeadas. Basado en los resultados del mapeo de los genes y los análisis de abundancia (Fig. 12), se encontró que existen variantes con un mayor número de secuencias mapeadas (> 1500 secuencias), a diferencia de otras donde este valor fue muy bajo (~ 200 secuencias). Algunas de las variantes son predominantes en cada fuente termal, se clasificaron como serina extracelular y serina termofílicas, las cuales fueron mayoría y tuvieron porcentajes de similitud mayor al 80% con secuencias obtenidas de microorganismos termotolerantes.

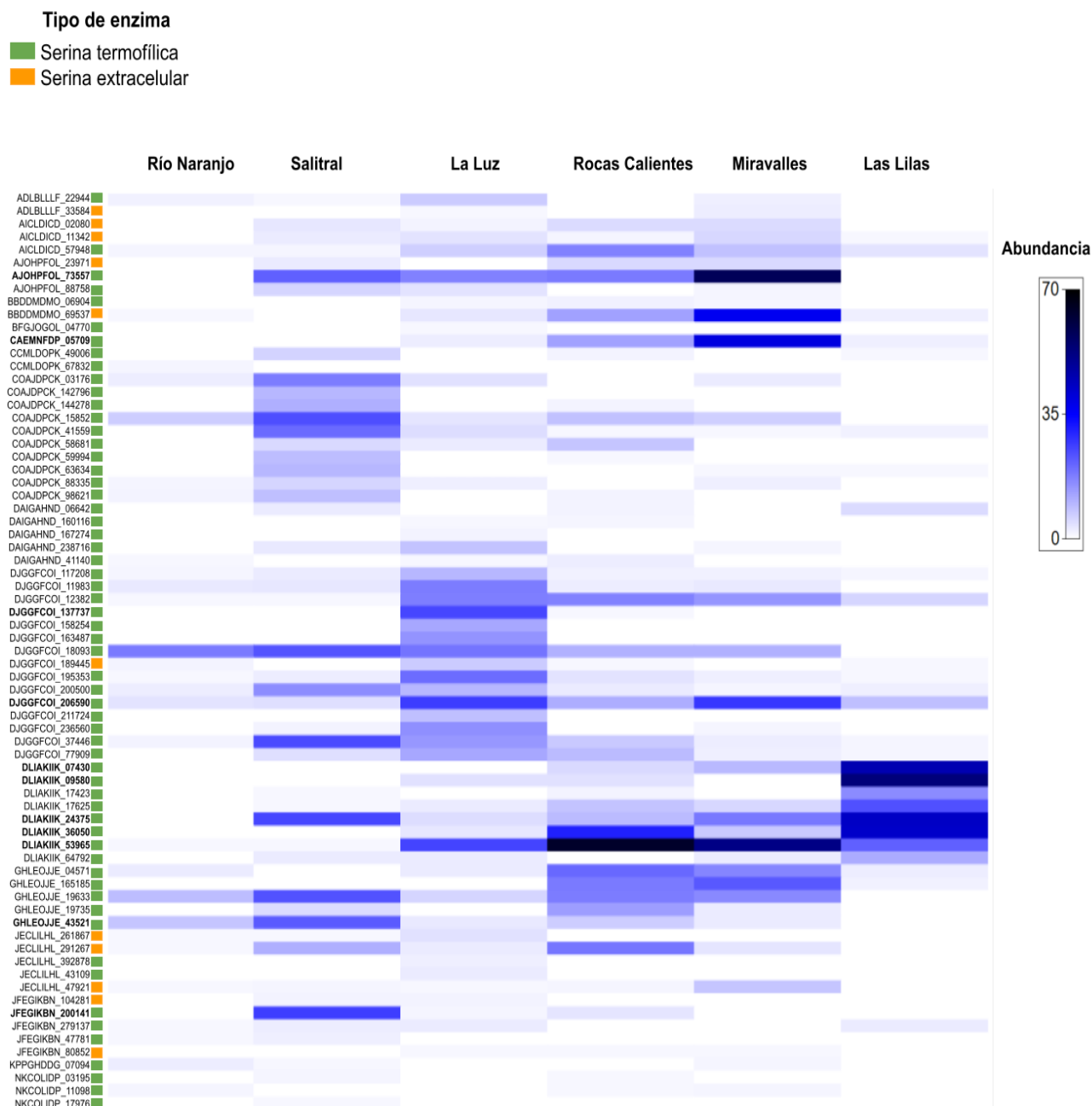


Figura 12. Abundancia de las enzimas clasificadas como serina proteasa en las fuentes termales. En negrita se señalan las secuencias mapeadas normalizadas con raíz cuadrada que se encontraron en mayor abundancia (más de 2000 secuencias mapeadas). Todas las secuencias pertenecen a la familia S8, la segunda familia más grande de peptidasas de serina. A pesar de que todas tienen el dominio característico de esta familia (tríada catalítica

Asp, His y Ser), algunas de ellas tienen un valor de identidad mayor o igual al 60% cuando se comparan con las secuencias proteicas en las bases de datos de UniProt.

Estos resultados son similares a los obtenidos por Takimura y colaboradores (2007), en el cual encuentran proteínas dentro de la misma familia con porcentajes de identidad cercanos al 40%, sin embargo, en las regiones que codifican para la tríada catalítica el porcentaje de similitud es mayor al 80%. El resultado del alineamiento se muestra en la Fig. 13, en donde se señalan además las regiones que corresponden al péptido señal y al dominio catalítico, donde se encuentra la tríada catalítica Asp-His-Ser. Algunas de las secuencias que fueron previamente clasificadas como peptidasas de serina poseen únicamente la diada catalítica Ser-Lys, la cual difiere de la tríada clásica Asp-Ser-His que se encuentra en la mayoría de las peptidasas de serina.

Los resultados muestran que las secuencias encontradas en los tapetes microbianos tienen los dominios funcionales característicos de las peptidasas de serina las cuales poseen algunos residuos de aminoácidos estrictamente conservados, lo que sugiere pueden tener una función similar. No obstante, se encontraron variantes en residuos específicos que podrían ser producto de la divergencia de las secuencias, que podrían ocasionar un cambio en la función de la proteína o una proteína no funcional. Con respecto a la región que corresponde al péptido señal se encontró que tienen algunos pocos residuos conservados, por lo que es muy variable. Como mencionan Ting y colaboradores (2016), los péptidos señal son extremadamente variables en secuencia y comprenden tres regiones distintas: una región N-terminal (N) cargada positivamente, una región hidrófoba (H) que atraviesa la membrana y una región polar C-terminal (C) que precede a la secuencia de la proteína madura. Un análisis en SignalP 5.0, muestra que la mayoría de péptidos señal en estas proteínas de tapetes microbianos tienen una vía de secreción independientes de Sec y Tat (traslocación de proteínas a la membrana interna). Las proteínas con péptidos señal diferentes a Sec y Tat podrían secretarse por mecanismos no convencionales o en dado caso que no tengan esta región pueden permanecer en el citoplasma (Bendtsen et al., 2005).

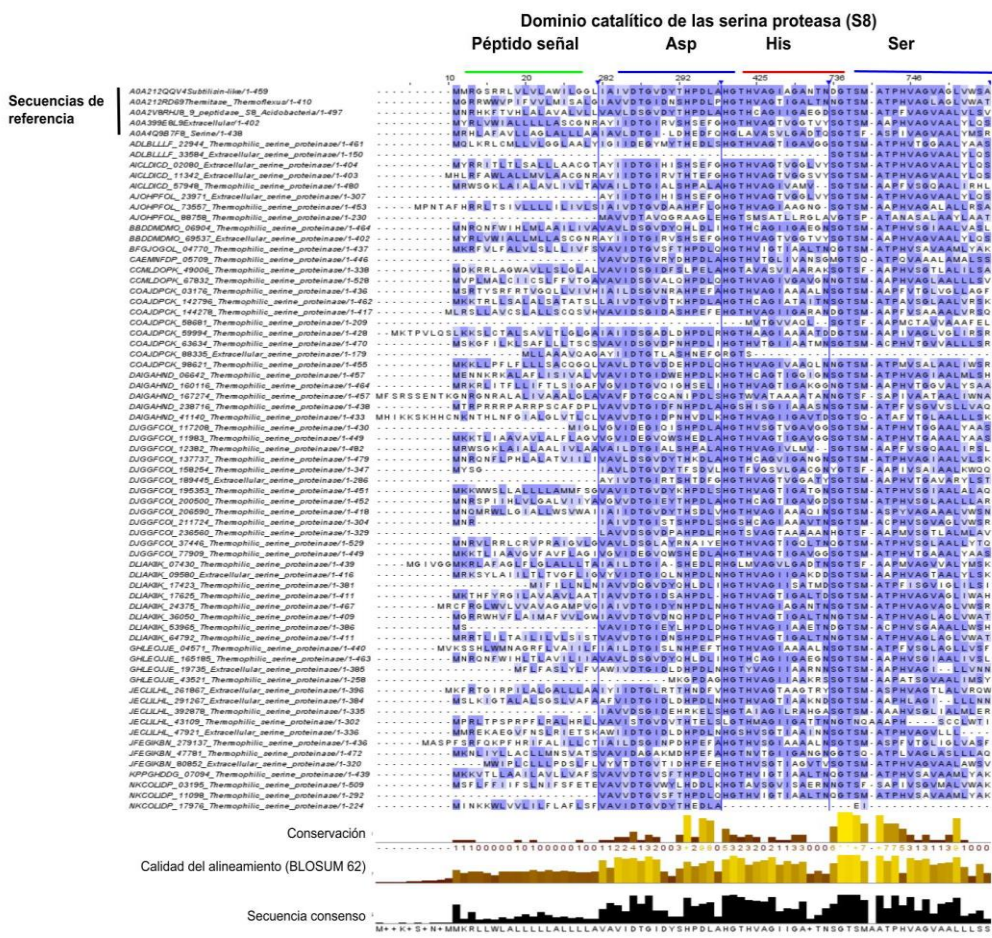


Figura 13. Alineamiento de secuencias clasificadas como serina proteasa encontradas en los tapetes microbianos de las fuentes termales costarricenses. Los colores señalan las secuencias que corresponden a la tríada catalítica. La conservación corresponde al índice que refleja la conservación de las propiedades fisicoquímicas en la alineación: las identidades obtienen la puntuación más alta y el siguiente grupo más conservado contiene sustituciones de aminoácidos pertenecientes a la misma clase fisicoquímica. La secuencia consenso que se muestra debajo de la alineación es el porcentaje del residuo por columna, si el valor es compartido por más de un residuo, se utiliza un símbolo "+"

En Rocas Calientes, se encontraron dos variantes predominantes. La variante CAEMNFDP_05709 es 90% similar a la serina proteasa de la familia S8 encontrada en el genoma de *Meiothermus cateniformans* (A0A399E307, datos no publicados). Este microorganismo se encontró en fuentes termales en el noreste de China con temperaturas entre 55 °C y 60 °C y pH superior a 8 (Zhang et al., 2009). Según la base de datos Pfam (<http://pfam.xfam.org/family/PF00082.22>), las subtilasas son la segunda familia más grande dentro de las serina proteasas. Al realizar el alineamiento de esta proteína en la base de datos UniProt, se confirmaron estos resultados y se mostró que la variante tiene regiones conservadas con otras proteínas de la familia S8 de *Meiothermus* sp. La variante AJOHPFOL_73557 es 83% idéntica a una peptidasa S8 (A5UX43) encontrada en *Roseiflexus* sp. cepa RS-1, por lo que son homólogos. La variante BBDDMDMO_69537, fue la que presentó el porcentaje de similitud más alto en este grupo de enzimas, siendo 97% similar a una proteinasa extracelular encontrada en el genoma de *Meiothermus cateniformans* (A0A399E8L9). Las variantes DLIAKIHK_53965 y DJGGFCOI_206590 son similares 88% y 91% a la peptidasa de serina termofílica de la cepa clasificada como Armatimonadetes GXS respectivamente. La variante DLIAKIHK_53965 se encontró en La Luz y en Las Lilas, mientras que la variante DJGGFCOI_206590 también se encontró en Miravalles. La variante DJGGFCOI_137737 solamente se encontró en la fuente termal Miravalles, es 57% idéntica a la peptidasa de la familia S8 (A0A2V8RHJ8), encontrada en una cepa bacteriana dentro del filo Acidobacteria. El estudio del microbioma de Rocas Calientes y Miravalles utilizando el gen 16S ARNr, demostró la presencia de OTUs asociados a los géneros *Chloroflexus*, *Roseiflexus* y *Meiothermus* (Uribe-Lorío et al., 2019), por lo que es de esperarse la presencia de secuencias de proteínas asociadas a estos microorganismos.

En Las Lilas se encontraron con mayor abundancia las variantes DLIAKIHK_53965, DLIAKIHK_36050, DLIAKIHK_24375, DLIAKIHK_09580 y DLIAKIHK_07430. La variante DLIAKIHK_24375 es 71.2 % similar a una serina proteasa encontrada en *Thermoflexus hugenholtzii* JAD2 (A0A212QQV4). DLIAKIHK_09580 es 88% similar a una proteasa encontrada en la bacteria Candidatus *Chrysopegis kryptomonas*. El genoma de esta bacteria fue encontrado utilizando la técnica Single Cell Genomics, en fuentes termales en China, caracterizadas por poseer temperaturas cercanas a los 77 °C y pH neutrales, y en fuentes termales en PNY mapeando secuencias similares (Eloe-Fadrosh et al., 2016). La variante DLIAKIHK_07430 es 64.5 % similar a una proteasa encontrada en el genoma de

Thermus thermamylovorans (A0A4Q9B7F8). En la fuente termal La Luz, se encontraron las variantes DLIAKIHK_53965 y DLIAKIHK_36050. La variante DLIAKIHK_36050 es 87.2 % similar a una proteasa de serina de *Thermoflexus hugenholtzii* JAD2 (A0A212RD69). En Salitral, la variante JFEGIKBN_200141 mostró un alto número de reads mapeados y es 60% similar a una secuencia asociada a serina proteasa encontrada en *Luteitalea* sp. Por otro lado, la variante GHLEOJJE_43521, que tiene una similitud de 69.8% con una secuencia encontrada en el genoma de *Raineya orbicola* (A0A2N3IK41), también se encontró en otras fuentes termales, pero con poca abundancia. En Río Naranjo se encontraron algunas variantes de serina proteasa, sin embargo, se observaron pocas secuencias mapeadas (< 200), por lo que la abundancia es muy baja.

Metalopeptidasa (Metaloproteínas)

Las metalopeptidasas (MPs) son enzimas proteolíticas que utilizan un metal para su mecanismo catalítico, la mayoría dependen del zinc, algunas también pueden utilizar cobalto, manganeso, níquel o cobre (UniProt, Rawlings y Barrett 2013). Estas enzimas se dividen en dos grupos en función del número de iones metálicos necesarios para la catálisis. Se reconocen 17 clanes y algunas familias sin asignar (Rawlings y Barrett 2013). La base de datos creada tenía un total de 2634 secuencias asociadas a metaloproteínas, agrupadas en 17 genes (*aprA*, *ftsH*, *ftsH1*, *ftsH2*, *ftsH3*, *ftsH4*, *irrE*, *loiP*, *mmpA*, *pmbA*, *prtV*, *rip1*, *rip2*, *rip3*, *stcE*, *tldD* y *ycaL*), tres familias o clanes, y del total 2353 fueron representativas. Algunos de los genes dentro de esta categoría son considerados *housekeeping* ya que son muy importantes para el mantenimiento de las funciones básicas de las células. Los resultados de este análisis de abundancia en cada gen se muestran en la figura 14, solamente se graficaron las variantes asociadas a los genes con una cobertura mayor al 0.02 % que fueron en total 72 variantes.

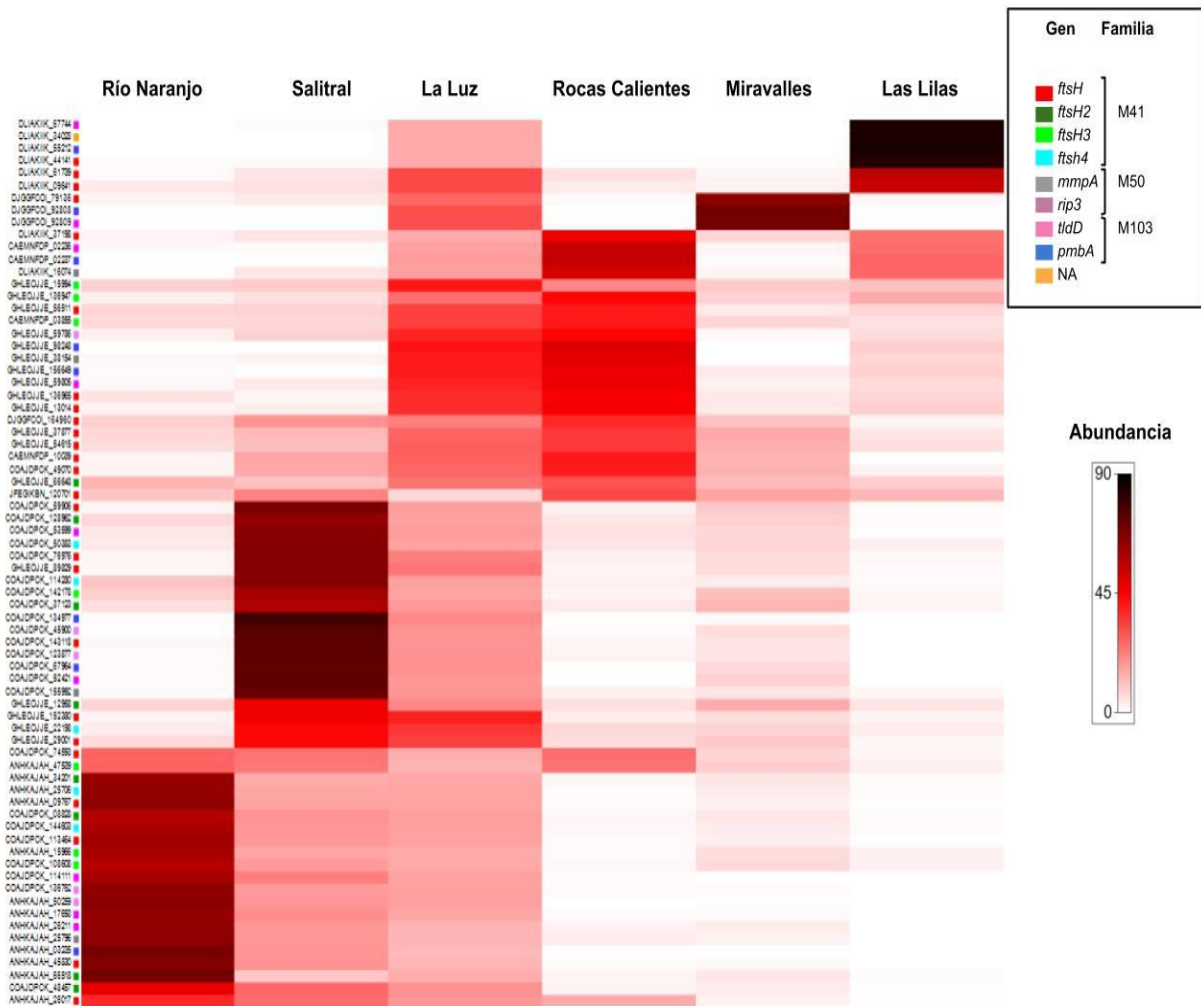


Figura 14. Abundancia de las diferentes variantes probablemente asociadas a metalopeptidasas en los tapetes microbianos de las fuentes termales costarricenses (secuencias mapeadas normalizadas con raíz cuadrada). En el cuadro de muestra la identificación por gen, NA significa que no hay un gen identificado previamente para esa secuencia.

La mayoría de las variantes de FtsH encontradas en estos tapetes microbianos están muy relacionadas a cianobacterias, las cuales son muy abundantes en tapetes microbianos, entre estas *Synechococcus* sp. y *Leptolyngbya* sp. Estos microorganismos se han reportado previamente en Miravalles, Rocas Calientes y Las Lilas (Uribe-Lorío et al., 2019, Brenes-Guillén et al., 2021). Los resultados de UniProt-Blast para las secuencias con una abundancia mayor a 2000 secuencias se muestran en el cuadro 12. La peptidasa FtsH puede encontrarse en bacterias, animales y plantas (Wagner et al., 2012). La mayoría de las bacterias no fotosintéticas contienen un solo gen *ftsH*, pero en cianobacterias como *Synechocystis* se han encontrado hasta cuatro copias del gen (Adam et al., 2004), lo que se puede relacionar con su importante rol en la homeostasis de los fotosistemas ya que es fundamental (*housekeeping gene*) para eliminar proteínas nocivas, participa en la regulación celular de la estabilidad proteica y tiene la capacidad de transportar sustratos de proteína de membrana fuera de la membrana (Ito y Akiyama, 2005).

Cuadro 12. Resultados de UniProt-Blast para las variantes de metaloproteínas más abundantes de las variantes *ftsH*

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
<i>ftsH</i>	GHLEOJJE_13014	A0A2G8PB46	<i>Synechococcus</i> sp. 60AY4M2	90.6	93.9
<i>ftsH</i>	GHLEOJJE_56511	UPI000C18B854	<i>Synechococcus</i> sp. 65AY6Li	88.2	92.3
<i>ftsH</i>	GHLEOJJE_29001	A0A081SIM7	Candidatus Thermochlorobacteriaceae bacterium GBChIB	73.1	85.0
<i>ftsH</i>	GHLEOJJE_152380	A0A081SJ28	Candidatus Thermochlorobacteriaceae bacterium GBChIB	77.2	88.4
<i>ftsH</i>	GHLEOJJE_136965	Q2JXF2	<i>Synechococcus</i> sp JA-3-3Ab	95.6	97.6
<i>ftsH2</i>	GHLEOJJE_66648	Q2JQW6	<i>Synechococcus</i> sp JA-3-3Ab	94.0	96.7
<i>ftsH3</i>	GHLEOJJE_15994	Q2JNP0	<i>Synechococcus</i> sp. JA-2-3 B'a (2-13)	96.4	98.7
<i>ftsH4</i>	GHLEOJJE_22198	A0A081SI35	Candidatus Thermochlorobacteriaceae bacterium GBChIB	76.7	87.8
<i>ftsH</i>	DJGGFCOI_79136	W7Z066	<i>Bacillus</i> sp. JCM 19045	99.7	99.8
<i>ftsH</i>	DLIAKIIK_44141	D3DKJ5	<i>Hydrogenobacter thermophilus</i> DSM 6534	95.3	98.9
<i>ftsH4</i>	COAJDPCK_50383	A0A0X8WY49	<i>Leptolyngbya</i> sp. O-77	92.0	95.3
<i>ftsH2</i>	COAJDPCK_128962	A0A0X8WUM0	<i>Leptolyngbya</i> sp. O-77	96.6	97.8
<i>ftsH4</i>	COAJDPCK_114280	A0A125STT5	<i>Leptolyngbya</i> sp. O-77	95.8	98.2
<i>ftsH</i>	COAJDPCK_143118	A0A0X8WW38	<i>Leptolyngbya</i> sp. O-77	91.7	95.6
<i>ftsH3</i>	COAJDPCK_142178	A0A0X8WVA6	<i>Leptolyngbya</i> sp. O-77	98.7	99.8
<i>ftsH2</i>	COAJDPCK_37123	A0A3M9ZEW7	<i>Leptolyngbya</i> sp. IPPAS B-1204	90.6	99.8

Al realizar el alineamiento y comparar con secuencias de proteínas curadas de bacterias como *Synechocystis*, *Thermomicrobium roseum*, *Alkaliphilus metalliredigens*, *Bacillus subtilis*, *E.coli* y *Thermus thermophilus*, se encontró en todas las secuencias los motivos característicos de la familia AAA (ATPasas asociadas con diversas actividades celulares), que combinan una dependencia de ATP y la degradación de proteínas. Dentro de esta región se encuentran los motivos Walker A y B (involucrados en la hidrólisis de ATP), así como los residuos de poros y la segunda región de homología (SRH) y el motivo característico de unión a zinc, HEXXH (Langklotz et al., 2012). La más frecuente fue HEAGH, sin embargo también se encontró la secuencia HEVGH, la cual se encontró únicamente en las secuencias de referencia de FtsH1, FtsH2 y FtsH4 de la cianobacteria *Synechocystis* PCC6803, primer organismo fototrófico en ser completamente secuenciado (Ikeuchi y Tabata, 2001) (Fig. 15), y en secuencias que tienen un porcentaje identidad mayor al 90% con secuencias de cianobacterias como *Leptolyngbya* O-77. Otra variante para esa región fue la HEIGH, la cual se encontró en secuencias con un porcentaje de identidad mayor al 90% de proteínas de la cianobacteria *Leptolyngbya* sp. IPPAS B-1204. Estos datos sugieren que las variantes encontradas en la región de unión a zinc pertenecen posiblemente a cianobacterias de los tapetes microbianos. Por otro lado, casi todas las secuencias tienen regiones altamente conservadas para los diferentes motivos de la proteína, sin embargo, algunas de estas no tienen el dominio de unión a zinc, por lo que podrían no ser dependientes de este metal y/o utilizar otro cofactor.

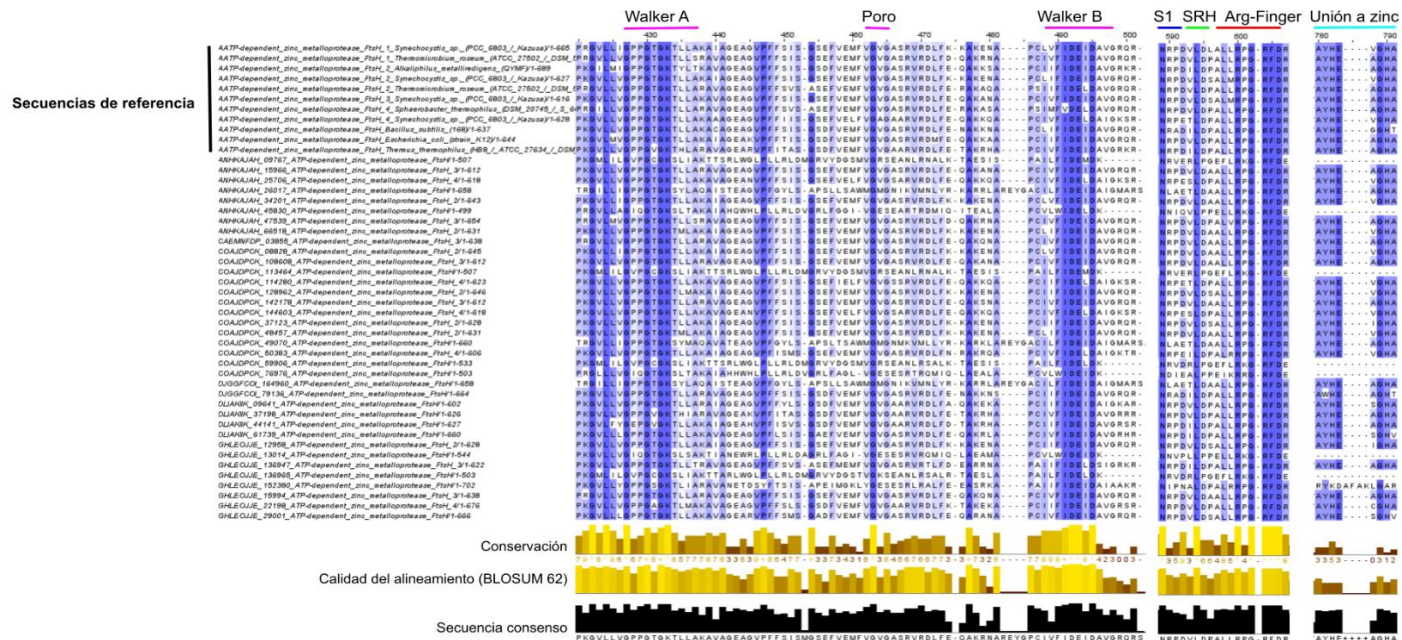


Figura 15. Alineamiento de secuencias clasificadas como FtsH encontradas en los tapetes microbianos de las fuentes termales costarricenses. S1: Sensor 1, Arg-finger: complejo catalíticamente activo con un monómero enzimático vecino a través de interacciones con el γ -fosfato del sustrato, SRH: segunda región de homología. La conservación: corresponde al índice que refleja la conservación de las propiedades fisicoquímicas en la alineación: las identidades obtienen la puntuación más alta y el siguiente grupo más conservado contiene sustituciones de aminoácidos pertenecientes a la misma clase fisicoquímica. El consenso que se muestra debajo de la alineación es el porcentaje del residuo modal por columna, si el valor modal es compartido por más de un residuo, se utiliza un símbolo "+".

Algunas variantes del gen *mmpA* fueron más abundantes en La Luz, Salitral y las Lilas (> 10000 secuencias) (Fig. 14). La mayoría de las variantes mostraron un porcentaje de identidad mayor al 90% con proteínas de cianobacterias (Cuadro 13). La proteína MmpA es una metaloproteasa que participa en la degradación de proteínas transmembrana como PodJS, la cual está involucrada en el establecimiento de la polaridad de la célula durante la división celular, por lo que debe ser degradada secuencialmente en momentos y lugares específicos durante el ciclo celular (Chen et al., 2005). El alto número de secuencias obtenido para algunas de estas variantes sugiere que tiene un rol muy importante en el tapete microbiano y que es un gen constitutivo (*housekeeping*).

Las variantes del gen *rip* son más abundantes en Salitral y la Luz (Fig. 14) (> 500 secuencias), y tienen una similitud mayor al 90% con proteínas encontradas en cianobacterias (Cuadro 13), lo que sugiere que probablemente las encontradas en este análisis son de este grupo de fotótrofos. Actualmente se considera como proteína putativa, por lo que no hay mucha información al respecto. Según la base de datos UniProt, es una proteína que participa en la regulación de la fase latente de la infección en *Mycobacterium* sp., además su actividad es inducida en respuesta al monóxido de carbono.

Los miembros de la familia de metalopeptidasas M50 incluyen proteasa del sitio 2 de la proteína de unión al elemento regulador de esteroides de mamíferos (SREBP), proteasa EcfE, proteínas relacionadas con la esporulación FB en estadio IV y varios homólogos hipotéticos en bacterias y eucariotas. Al realizar el alineamiento se encontraron los residuos de aminoácidos que se cree que son esenciales para la actividad catalítica ya que permiten la unión del metal (HEXXH), también los dominios HELGH (siete secuencias, incluida la secuencia de referencia Q9A710 de *Caulobacter*) y HELAH (2 secuencias, y la secuencia H8EUF2 de *Mycobacterium*), por lo que tenemos proteínas dentro de la familia que tienen variaciones de un aminoácido (ambos no polares y alifáticos), en el dominio de unión a zinc y que tiene un rol muy importante en la hidrólisis. En los motivos HEXXH se producen modificaciones con el doble de frecuencia que en las proteínas en general, aunque el papel biológico de las sustituciones en los sitios activos no está claro, ya que algunos de estos residuos cargados positivamente podrían no participar en la unión del ión zinc (Lenart et al., 2013). Se encontraron en todas las secuencias los residuos de glicina conservados, estos son fundamentales para la actividad proteolítica.

Cuadro 13. Resultados de Uniprot-Blast para las variantes de metaloproteínas más abundantes de las variantes de los genes *mmpA* y *rip*

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
<i>mmpA</i>	GHLEOJJE_38154	Q2JMR4	<i>Synechococcus</i> sp. JA-2-3B'a (2-13)	91.5	95.9
	COAJDPCK_155992	A0A0X8WRM0	<i>Leptolyngbya</i> sp. O-77	95.6	97.5
	DLIAKIIK_16074	K7R614	<i>Thermus oshimai</i> JL-2	91.7	94.6
	ANHKAJAH_25796	A0A2D2Q042	<i>Synechococcus lividus</i> PCC 6715	93.5	97.5
<i>rip</i>	GHLEOJJE_59786	Q2JNF0 (X aa)	<i>Synechococcus</i> sp. JA-2-3B'a (2-13))	83.0	89.8
	COAJDPCK_45900	A0A110AUP6	<i>Leptolyngbya</i> sp. O-77	96.5	99.2
	COAJDPCK_123877	A0A110AVA2	<i>Leptolyngbya</i> sp. O-77	90.1	95.3

Las variantes más abundantes de *pmbA* se encontraron principalmente en las Lilas, La Luz y Salitral (Fig. 14). Los resultados de UniProt-blast se muestran en el cuadro 14. La proteína PmbA es una enzima citoplasmática hidrofílica que posee 450 aminoácidos, en *Escherichia coli* participa en la biosíntesis y secreción del antibiótico Microcin B17 (Rodríguez-Sáinz et al., 1990, Murayama et al., 1996, Hu et al., 2012). Es una proteína moduladora de la girasa ADN, participa en el control de la división y muerte celular, así como en la regulación del almacenamiento de carbono (Rife et al., 2005).

Cuadro 14. Resultados de UniProt-Blast de las variantes de metaloproteínas más abundantes de los genes *pmbA*

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
<i>pmbA</i>	GHLEOJJE_98248	Q2JJ43	<i>Synechococcus</i> sp. JA-3-3Ab	89.2	94.8
	GHLEOJJE_156649	Q2JXJ7	<i>Synechococcus</i> sp. JA-3-3Ab	88.2	94.2
	DLIAKIIK_55212	D3DJQ3	<i>Hydrogenobacter thermophilus</i> DSM 6534	71.1	86.4
	DJGGFCOI_92808	W7ZKB4	<i>Bacillus</i> sp. JCM 19046	98.4	99.1
	COAJDPCK_134977	A0A0X8WUS2	<i>Leptolyngbya</i> sp. O-77	94.6	97.8

Las variantes con mayor abundancia de la proteína TIDd se encuentran en Las Lilas, Salitral, La Luz y Miravalles (Fig. 14). Todas estas variantes son similares a proteínas homólogas encontradas en los microorganismos más abundantes en estos ambientes (Cuadro 15). Existe evidencia en ensayos en laboratorio de que esta enzima participa en la proteólisis de CcdA (Proteína de biogénesis de tipo citocromo c), la cual es requerida para la síntesis de citocromo c y la etapa V de la esporulación (Allali et al., 2002).

La variante llamada NA (Metalopeptidasa), es muy abundante en Las Lilas, aunque se encontró en otras fuentes termales como Miravalles. Al realizar la anotación con Prokka se denominó proteína putativa. La comparación realizada en UniProt-Blast mostró que esta proteína es similar a una metaloproteasa encontrada en *Hydrogenobacter hydrogenophilus*, no obstante, ésta secuencia está en la categoría de “no revisada”, por lo que se le asignó esta clasificación por homología de secuencia y es necesario hacer una revisión más detallada de la misma. Con el fin de mejorar la clasificación dentro del grupo de las metalopeptidasas, se utilizó la herramienta en línea Conserved Domains del NCBI y la base de datos CDD (Conserved Domains Database) que contiene 55570 secuencias, los resultados obtenidos sugieren que es similar a una peptidasa de la familia M50 y tiene regiones similares a las proteínas de esta familia encontrada en otros microorganismos.

Cuadro 15. Resultados de UniProt-Blast para las variantes de metaloproteínas

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
NA	DLIAKIK_34028	A0A285P007	<i>Hydrogenobacter hydrogenophilus</i>	78.0	91.2
tIDd	GHLEOJJE_59805	Q2JT09	<i>Synechococcus</i> sp. JA-3-3Ab	91.3	97.6
	DLIAKIK_67744	A0A2H5WJL2	Uncultured bacterium HR13	85.3	93.8
	DJGGFCOI_92809	W7Z130	<i>Bacillus</i> sp. JCM 19045	98.9	98.9
	COAJDPCK_53599	A0A120MWJ1	<i>Leptolyngbya</i> sp. O-77	97.8	98.8
	COAJDPCK_52421	A0A3M1PMA1	Cyanobacteria bacterium J069	95.5	97.9

La mayoría de las secuencias de proteínas se encuentran relacionadas con microorganismos abundantes en estas fuentes termales tales como *Synechococcus* y *Leptolyngbya*. En Las Lilas todas estas variantes son similares a otras secuencias encontradas en *Hydrogenobacter hydrogenophilus*. Este microorganismo fue aislado por primera vez en fuentes termales de Japón, y tiene un rango de crecimiento óptimo que va desde los 70 °C a los 75 °C y pH neutro (Kawasumi et al., 1980, Ishii et al., 1989), sin embargo, se ha encontrado en fuentes termales con temperaturas entre los 37 °C y los 80 °C (Huber y Stetler, 2001). Esta cepa específicamente crece en condiciones anaerobias y usa nitrato como aceptor de electrones cuando el hidrógeno molecular es utilizado como fuente de energía, generando nitrito (Suzuki et al., 2001). Los Aquificales son un filo muy abundante en esta fuente termal, se caracterizan por utilizar tiosulfato o azufre como única fuente de energía, y forman ácido sulfúrico como producto final (Huber y Stetler, 2001).

En Miravalles, se encontraron tres variantes con una abundancia mayor en comparación con las otras fuentes termales. Todas ellas tienen una alta similitud con proteínas encontradas en el genoma de diferentes cepas de *Bacillus*. *Bacillus* JCM 19045 y JCM 19046 son alcalófilos (Cuadro 16), por lo que generalmente crecen en condiciones de pH mayor a 8 (Kudo et al., 2016). En el análisis taxonómico realizado previamente, se encontró que *Bacillus* spp es uno de los grupos más abundantes en esta fuente termal. En Rocas Calientes y Río Naranjo se encontraron variantes para cada uno de los genes, sin embargo, todas tenían poca abundancia. No se encontraron variantes únicas de este sitio cuando se realizó el mapeo.

Las metalopeptidasas representan casi la mitad de todas las enzimas (Valasatava et al., 2015). Algunas son responsables de catalizar funciones en procesos como la fotosíntesis, la respiración, la oxidación del agua, la reducción de oxígeno molecular, fijación de nitrógeno y participa en patógenesis (Lu et al., 2009). Las funciones específicas catalizadas por metaloenzimas incluyen (i) reacciones de oxidación y reducción, para las cuales los metales más importantes son Fe, Mn, Cu y Mo, (ii) reacciones de transposición basadas en radicales que requieren principalmente Fe o Co, (iii) transferencias de grupos metilo catalizadas por Co, (iv) hidrólisis que a menudo implica Zn, Fe, Mg, Mn o Ni, y (v) regulación del ADN, que generalmente requiere Zn (Pernil y Schleiff, 2019). Las metalopeptidasas se encuentran entre las primeras proteínas diseñadas *de novo* (*bottom up*) (DeGrado et al., 1999). En los últimos años, se ha utilizado la predicción computacional y la genómica comparativa de

genes que codifican metaloproteínas, proporcionando información significativa sobre su distribución, función y evolución en la naturaleza. Estos estudios se han realizado tanto en procariontes como en eucariontes, lo que demuestra patrones evolutivos divergentes y dinámicos de diferentes familias de metaloproteínas y metaloproteomas (Zhang y Zheng, 2020).

No se ha encontrado ninguna familia de peptidasa ubicua e incluso las consideradas “esenciales” para la vida están ausentes en algunos clados (Rawlings y Bateman, 2019). En las fuentes termales al igual que en otros ambientes extremos estas enzimas están involucradas en la degradación de diversas proteínas, respuesta al estrés, y pueden utilizar sustratos de fuentes primarias de carbono y energía mediante la hidrólisis extracelular (Ward et al., 2002). Las peptidasas son muy importantes para la nutrición bacteriana y en el metabolismo del nitrógeno. En sedimentos marinos se ha descrito que las peptidasas extracelulares son muy abundantes y diversas (Olivera et al., 2007, Zhou et al., 2009, 2013). Muchas de ellas participan en mecanismos celulares para la adaptación al frío (Chen et al., 2007, Yan et al., 2009, Yang et al., 2013), tolerancia a la salinidad (Yan et al., 2009), especificidad de sustrato y diversos mecanismos catalíticos (Ran et al., 2014). En microorganismos aislados de fuentes termales se han realizado esfuerzos para aislar, cuantificar y caracterizar diferentes proteasas (Prabhakar Rao y Rathod, 2017).

En los procesos de biorremediación, el primer paso para la degradación de algunos pesticidas es la hidrólisis, algunos microorganismos pueden utilizar los productos de la hidrólisis como fuente de carbono, nitrógeno y fósforo (Singh y Walker, 2006, Ufarté et al., 2017). Es importante considerar que las secuencias que posiblemente codifican para enzimas hidrolíticas dentro de la categoría peptidasas podrían tener un rol importante en la tolerancia de los microorganismos a las condiciones con alta concentración de metales y otros compuestos en las fuentes termales, por lo que podrían considerarse en estudios de biorremediación. Singh y colaboradores (2006), mencionan que la enzima OPAA (aislada de *Alteromonas undina*), participa en la hidrólisis de organofosforados y tiene un grado de homología a nivel de aminoácidos con *E. coli* X-Pro dipeptidasa (48%) y *E. coli* aminopeptidasa (31%). Estos autores señalan que, aunque aún no se conoce la función nativa de OPAA, se ha sugerido que desempeña un papel importante en el metabolismo de varios dipéptidos. En el mercado se encuentran disponibles diferentes herbicidas, antibióticos, bionutrientes y bioestimulantes que tienen dipéptidos o tripéptidos en sus formulaciones tales como el tripéptido (unión de dos aminoácidos alanina y glufosinato) de

fosfotricina (PTT, fosfotricilalanilalanina, bialafos), el cual es un antibiótico de producto natural y un herbicida potente producido por *Streptomyces viridochromogenes* y *Streptomyces hygroscopicus* (Schwartz et al., 2004, Blodgett et al., 2007). La actividad herbicida de los aminoácidos se ha estudiado previamente, como por ejemplo la metatirosina (aminoácido no proteico), liberado por la exudación de la raíz de plantas alelopáticas (Bertin et al., 2007) o por la sobreexcreción de patógenos (Nzioki et al., 2016). Algunas investigaciones y ensayos de laboratorio determinaron el posible efecto de los aminoácidos en el control de malezas (Vurro et al., 2006, Fernández-Aparicio et al., 2017).

Lipasas

La base de datos de lipasas contenía 492 secuencias, de las cuales 401 son representativas, y, de todas las 401 secuencias se obtuvieron secuencias mapeadas (Fig. 11), a excepción de Miravalles donde se encontraron tres secuencias únicas. El análisis de similitud con UniProt se realizó solamente para 16 secuencias que tenían un porcentaje de abundancia mayor al 0.02%. La mayoría de las secuencias de enzimas se encuentran relacionadas con secuencias encontradas en microorganismos como *Thermus oshimai*, *Synechococcus* y *Leptolyngbya*.

Se realizó un árbol filogenético similar al elaborado por Kovacic y colaboradores (2018) (Fig. 4), con el que se pretendía identificar clados de secuencias por familia y lograr clasificar en familias las secuencias de los tapetes microbianos. Este árbol se construyó únicamente con las secuencias que tuvieron un porcentaje de abundancia mayor al 0.02% y también se incluyeron las esterases, de las que se hablará más adelante (Figura 16). Como se puede observar en el árbol, muchas de las secuencias obtenidas en este estudio no forman clados con las familias ya establecidas, lo que es un indicativo de que probablemente son alelos o familias nuevas. Además, están más relacionadas a lipasas de bacterias termofílicas como *Thermoflexus*, *Thermus oshimai* y algunas cianobacterias. Por otro lado, algunas de las secuencias formaron clados dentro de las familias, lo que apoya los resultados de similitud obtenidos con UniProt, ya que son homólogas a otras encontradas en microorganismos que no necesariamente son termotolerantes.

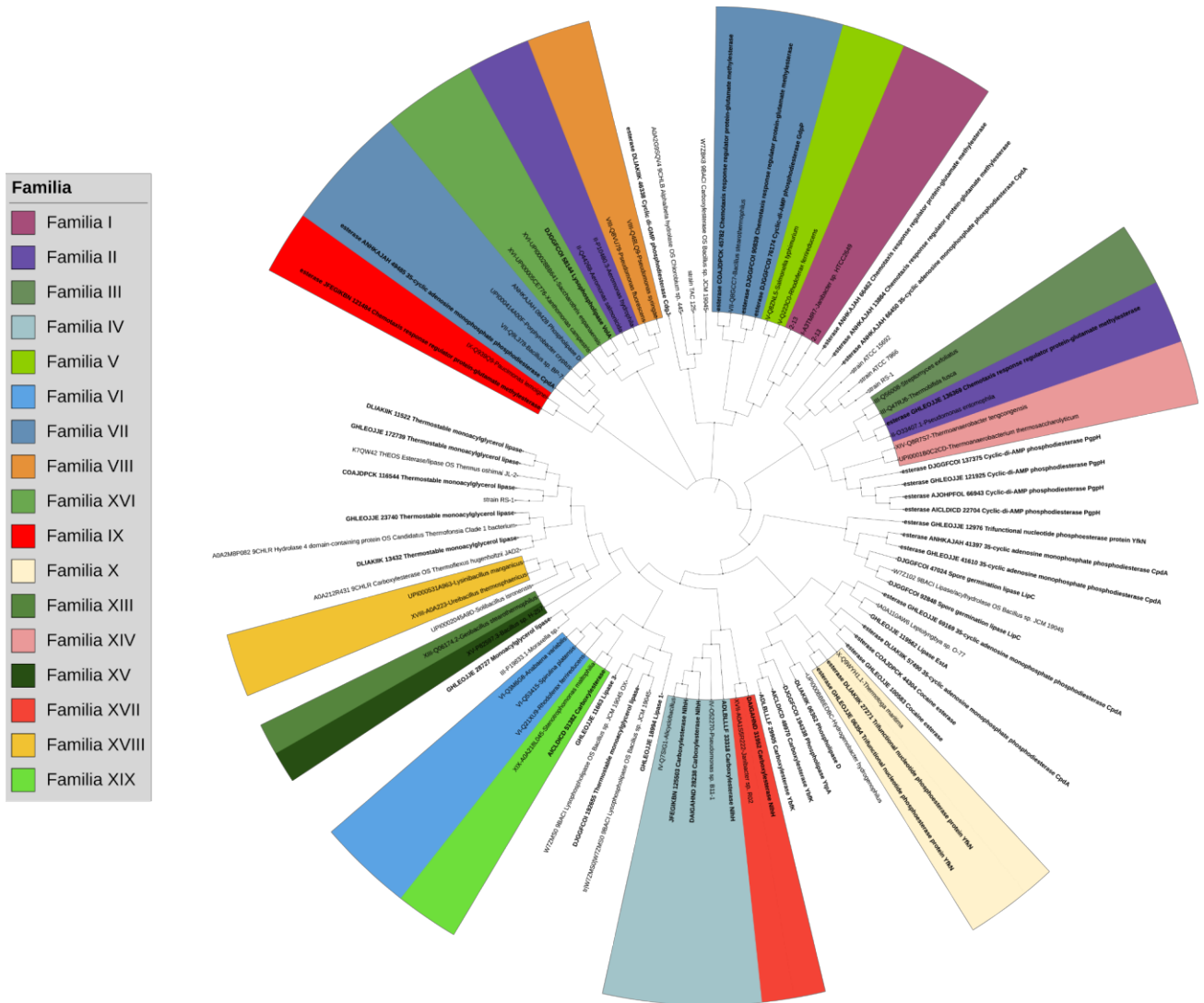


Figura 16. Árbol filogenético de las secuencias de lipasas y estereras encontradas en los tapetes microbianos con un porcentaje mayor al 0.02%. Para esto se utilizó el método de *clustering* de *Neighbor joining* en MEGA-X (version 10.1.8) y el diseño se realizó en Itol (Letunic y Bork 2006). Los identificadores en **negrita** corresponden a las muestras de Costa Rica y las blancas corresponden a secuencias que no se han asignado a ninguna familia.

Se encontraron secuencias con una similitud mayor al 60% y una cobertura mayor a 80% (Cuadro 16) en las variantes de los genes *estA*, *lip*, *ml*, *pld*, *tml*, *valA* y *ytpA*. Las lipasas fueron encontradas en varias fuentes termales, pero más abundantes en La Luz (> 40000 secuencias) (Fig. 11). La enzima EstA se encontró principalmente en Salitral (> 8000 secuencias), esta es producida y secretada por *Bacillus subtilis* como una lipasa de ésteres de p-nitrofenilo y triacilglicéridos (Eggert et al., 2001). En relación con las variantes de la proteína Lip, GHLEOJJE_18994 fue la más abundante en la fuente termal La Luz (18000 secuencias), sin embargo, se encontraron algunas otras variantes en Las Lilas Miravalles y Rocas Calientes (abundancia entre 200 y 5000 secuencias). Estas enzimas se encuentran relacionadas a proteínas de la familia I.8 de las triacilglicerol lipasas. Esta familia está conformada por secuencias como la Lip1 de la bacteria psicrófila *Pseudoalteromonas haloplanktis* (de Pascale et al., 2008). Lip1 tiene un peso molecular de 51 kDa, está asociada a la membrana externa y ha mostrado actividad contra sustratos de éster con cadenas de acilo largas (Kovacic et al., 2018). Las variantes encontradas en este estudio son similares a las encontradas en microorganismos termófilos tales como *Synechococcus* JA-2-3B'a (2-13), aislado de una fuente PNY, *Roseiflexus* spp, *Leptolyngbya* O-77 aislada Aso-Kuju en Japón (Nakamori et al., 2014) y la cepa JCM 1905 de *Bacillus*, encontrada en fuentes termales alcalinas de Japón (Kudo et al., 2014) (Cuadro 16). Una investigación realizada por Li y colaboradores (2018), demostró que los cambios de estabilidad inducidos por mutaciones en *lip1* conducen a un aumento de vida media del hongo *Candida rugosa* (patógeno, utilizado en fermentaciones de alimentos), a temperaturas cercanas a los 50 °C mientras que mutaciones específicamente en Asp457Ph provocan que el hongo pueda vivir a pH mayores a 7, por lo que es de gran importancia para tolerar altas temperaturas y pH alcalinos, lo que sugiere que podría tener el mismo rol en los microorganismos termotolerantes.

Las lipasas de monoacilglicerol termoestables (MLGP, EC 3.1.1.23) se encontraron en todas las fuentes termales, sin embargo, fueron más abundantes en La Luz y Rocas Calientes (50 °C - 63 °C) (19 000 y 1200 secuencias respectivamente). Las lipasas de monoacilglicerol (MGL) son enzimas que hidrolizan el monoacilglicerol en un ácido graso libre y glicerol. Los ácidos grasos se pueden usar para la síntesis de triacilglicerol, como fuente de energía, estos son componentes básicos para el almacenamiento de energía y precursores de los fosfolípidos de membrana (Aschauer, et al., 2018). Imamura y Kitaura (2000) aislaron de la cepa moderadamente termófila *Bacillus* sp. H-257 una MLGP y encontraron que estas enzimas tienen actividad óptima a los 75 °C y pH de 7.5, además son capaces de hidrolizar

monoacilglicerol. Resultados similares fueron descritos por Tang y colaboradores (2019), específicamente en una enzima aislada de la bacteria marina *Geobacillus* sp 12AMOR1 y expresada en *E. coli*, estos autores demostraron que la actividad máxima es cercana a los 60 °C y pH 8, además utiliza como sustrato el monoacilglicerol.

Las variantes del gen *pld* se encuentran principalmente en Las Lilas, la variante ANHKAJAH_08428 es abundante en Río Naranjo, Salitral y la Luz (> 400 secuencias), mientras que y DLIAKIHK_06362 se encontró únicamente en Las Lilas (1882 secuencias). La fosfolipasa D (*pld*) es una enzima que genera ácido fosfatídico para la señalización intracelular, y juega un papel importante en diversas funciones en procariontes, virus, levaduras, hongos, plantas y animales, y en mecanismos de virulencia (Bruntz et al., 2014). Además, cataliza la hidrólisis de fosfatidilcolina en ácido fosfatídico y colina (Houben AJS, Moolenaar, 2011, Bruntz et al., 2014). Estas enzimas participan en el mantenimiento de las funciones celulares basales que son esenciales para la célula.

La variante del gen *voA* fue encontrada únicamente en la fuente termal de Miravalles (3741), y la variante *ytpA* solamente en La Luz (19307 secuencias). En el caso de la variante *VoA*, se utilizó la herramienta en línea Conserved Domains del NCBI y la base de datos CDD (Conserved Domains Database) que contiene 55570 secuencias. Los resultados mostraron que esta secuencia comparte dominios conservados con las secuencias de la familia de lipasas Pla-1/cef, la cual es una lipasa extracelular encontrada principalmente en Gammaproteobacterias. Por otro lado, la proteína *YtpA* participa en la biosíntesis del antibiótico basiolocina producido por *Bacillus subtilis* 168, participa en la catálisis de fosfatidilglicerol (Tamehiro et al., 2002).

Cuadro 16. Resultados de UniProt-Blast para las variantes de lipasas más abundantes en todas las fuentes termales.

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
<i>estA</i>	GHLEOJJE_119562	A0A110AW66	<i>Leptolyngbya</i> sp. O-77	88.9	93.1
<i>lip</i>	DJGGFCOI_47024	A5V0A0	<i>Roseiflexus</i> sp. RS-1	86.4	91.9
	GHLEOJJE_18994	Q2JLM4	<i>Synechococcus</i> sp. JA-2-3B'a (2-13)	90.7	95.5
	GHLEOJJE_11663	A0A2G9SQV4	<i>Chlorobium</i> sp. 445	63.9	79.6
	DJGGFCOI_92848	W7Z102	<i>Bacillus</i> sp. JCM 19045	99.0	100.0
<i>ml</i>	GHLEOJJE_28727	Q2JJ03	<i>Synechococcus</i> sp. JA-2-3B'a (2-13)	78.9	87.8
<i>pld</i>	DLIAKIIK_06362	UPI000BBBEDBC	<i>Hydrogenobacter hydrogenophilus</i>	69.2	82.8
	ANHKAJAH_08428	UPI000414A00F	<i>Porphyrobacter cryptus</i>	91.1	94.3
<i>tml</i>	GHLEOJJE_23740	A0A2M8P082	Candidatus <i>Thermofonsia</i> Clado 1	90.0	96.2
	GHLEOJJE_172739	K7QW42	<i>Thermus oshimai</i> JL-2	87.4	93.0
	DLIAKIIK_13432	A0A212R431	<i>Thermoflexus hugenholtzii</i> JAD2	74.1	80.1
	DLIAKIIK_11522	K7QW42	<i>Thermus oshimai</i> JL-2	86.5	92.2
	DJGGFCOI_192655	W7ZBK8	<i>Bacillus</i> sp. JCM 19045	100.0	100.0
	COAJDPCK_116544	A5UVT0	<i>Roseiflexus</i> sp. RS-1	87.8	91.7
<i>volA</i>	DJGGFCOI_58144	A0A521WHD4	Gammaproteobacteria bacterium	45.7	60.4
<i>ytpA</i>	DJGGFCOI_194338	W7ZMS0	<i>Bacillus</i> sp. JCM 19045	99.6	100.0

Se han descrito algunas lipasas aisladas de microorganismos termotolerantes como *Bacillus* en fuentes termales de Islandia (Sigurgísladóttir et al., 1993), *Thermus thermophilus* HB27 (Fuciños et al., 2012), la cepa A30-1 de *Bacillus* (ATCC53841) de PNY (Wang et al., 1995), *Geobacillus* sp.TW1 en China (Li y Zhang, 2005, Yan et al., 2017), y *Bacillus* spp de fuentes termales de Setapak en Malasia (Sheikh Abdul Hamid et al., 2003, Zuridah et al., 2011). También se han encontrado en clonas obtenidas de fuentes termales de la India (Kumar et al., 2020). Específicamente en la cianobacteria *Synechocystis* PCC6803 se han descrito lipasas que sintetizan ester de fitol (precursor de diferentes vitaminas) a partir del fitol que se deriva de la descomposición de la clorofila (Aizouq et al., 2020). Por otro lado, se ha descrito que las células de *Synechocystis* tienen un mecanismo para equilibrar los lípidos o responder a cambios ambientales hidrolizando la membrana de fosfolípidos en ácidos grasos libres intracelulares a través de la actividad de lipasas (Gao et al., 2012, Eungrasamee et al., 2020). En cepas patógenas las lipasas junto a las proteasas tienen un papel muy importante en la formación de *biofilm* y la invasión de diferentes células huésped (Prabhawathi et al., 2014, Saggi et al., 2019). Se ha determinado la capacidad de algunas lipasas en la degradación de agroquímicos, Zhao y colaboradores (2021), demuestran que una nueva lipasa alcalina, tolerante a solventes orgánicos y activa entre temperaturas de 5 °C a 55 °C puede degradar el herbicida fenoxiaético 2,4-D, la cual fue aislada de *Bacillus licheniformis* NCU CS-5.

Esterasas

La base de datos consistió en 6660 secuencias de proteínas clasificadas como esterasas, que al realizar la depuración de la base mediante MMseqs, resultó en un total de 5575 secuencias, de estas 289 tenían una abundancia mayor al 0.02 % (> 5000 secuencias) del total de secuencias. Estas fueron incluídas en el mapa de calor (Figura 11), sin embargo, para la comparación con UniProt y el análisis filogenético se utilizaron solamente variantes que tenían más de 20% de abundancia en total y las carboxilesterasas con una abundancia mayor al 0.02%, por lo que este análisis se realizó solamente para 29 secuencias (Cuadro 18). Estas variantes están codificadas por los genes *cdgJ*, *yfkN*, *cpdA*, *sbcD*, *cocE*, *gdpP*, *cpdA* y *pgpH*, para algunos de los cuales se encontraron varios alelos. Para algunas de las variantes no existe aún un gen específico asociado, solamente se reconoce la función y el locus donde se puede localizar, esto a pesar de que utilizando prokka se le asignó una función específica. Como se puede observar en el árbol (Fig. 16), muchas de las secuencias

obtenidas en este estudio no forman clados con las familias ya establecidas, lo que es un indicativo de que probablemente son alelos o familias de esterasas no reportadas hasta el momento.

La mayoría de las esterasas encontradas en este estudio son codificadas por genes constitutivos, los cuales están involucrados en la hidrólisis del di guanilato cíclico (c-di-GMP), regulan positivamente la motilidad y negativamente la formación de biopelículas, liberación de fosfato inorgánico de 2', 3'-nucleótidos a través de actividades consecutivas 2', 3'-fosfodiesterasa. Estas pueden desempeñar un papel regulador importante en la modulación de la concentración intracelular de AMPc, lo que influye en los procesos dependientes de AMPc (UniProt).

Cabe destacar, que se encontraron diferentes alelos que codifican por el gen *cocE* (89), principalmente en el tapete microbiano de Río Naranjo en donde se encontraron 25249 secuencias, en contraste con las otras fuentes termales con menos de 50 secuencias mapeadas. Esta proteína hidroliza la cocaína en benzoato y éster metílico de ecgonina, les confieren a las bacterias la capacidad de utilizar la cocaína como única fuente de carbono y energía para el crecimiento, se han encontrado principalmente en bacterias de la rizosfera de las plantas de coca como *Rhodococcus* (Narasimhan et al., 2012). La falta de estabilidad de esta enzima a la temperatura fisiológica (37 °C) es un obstáculo para utilizarla en una terapia para la toxicidad de la cocaína (Narasimhan et al., 2012, Howell et al., 2014). Es posible que las secuencias encontradas en las cianobacterias de estos tapetes microbianos codifiquen para enzimas termotolerantes, lo que ofrece una oportunidad de futuros estudios relacionados con la hidrólisis de este compuesto. En los microorganismos de ambientes termofílicos podría tener una actividad hidrolasa, sin embargo, no ha sido caracterizada.

Es importante mencionar que se encontraron diferentes carboxilesterasas (CE), estas son una familia de enzimas que son importantes en la hidrólisis de una gran cantidad de compuestos endógenos y xenobióticos que contienen ésteres, como los plaguicidas organofosforados (OP) (Zhang et al., 2004, Singh 2014). Estas enzimas están codificadas por genes *ybfK* y *nlhH*, los alelos más abundantes se encuentran en el cuadro 17.

La carboxilesterasa YbfK también llamada CesB, pertenece a la familia de hidrolasas VI con un pliegue característico de hidrolasa α / β , algunas investigaciones demuestran que esta enzima participa en la degradación de insecticidas piretroides tales como β -cipermetrina, cipermetrina, β -ciflutrina, ciflutrina, λ -cialotrina y cialotrina (Bhatt et al., 2020, Xiao et al., 2020). Las secuencias encontradas en estos tapetes microbianos tienen un 89% de similitud con una secuencia de *Roseiflexus* sp, sin embargo, no existen reportes o artículos científicos

relacionados. Las diferentes cepas del filo Chloroflexi provenientes de tapetes microbianos de ambientes termales han sido poco caracterizados a nivel taxonómico y funcional, por lo que es de importancia para futuros estudios buscar estos genes y realizar ensayos en laboratorio para validar los resultados obtenidos mediante diferentes estrategias.

Los diferentes alelos del gen *nlhH* tuvieron un porcentaje de similitud del 74% al 100% con secuencias de bacterias como *Elioraea* sp. (PNY, bacteria termofílica), *Panacagrimonas perspica*, *Hyphomonas hirschiana* y *Sphingomonas formosensis* (aislada principalmente de suelo) (Cuadro 17). Esta enzima hidroliza varios ésteres de cadena corta, como triacilgliceroles y ésteres vinílicos (Canaan et al., 2004). Según las referencias consultadas, no existe evidencia de la participación de esta enzima en procesos de degradación de xenobióticos como los agroquímicos, sin embargo, podrían realizarse diferentes ensayos en el laboratorio y determinar su participación en el proceso de hidrólisis.

Cuadro 17. Resultados de Uniprot-Blast para las variantes de esterasas más abundantes (> 20%) y las carboxilesterasas con una abundancia mayor al 0.02% en todas las fuentes termales. Las secuencias marcadas con un asterisco fueron clasificadas como carboxilesterasas.

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
<i>cdgJ</i>	DLIAKIIK_46338	D3DG84	<i>Hydrogenobacter thermophilus</i> DSM 6534	90.2	96.1
<i>yfkN</i>	GHLEOJJE_12976	A0A0X8WTJ7	<i>Leptolyngbya</i> sp. O-77	93.5	96.4
<i>yfkN</i>	GHLEOJJE_06354	Q2JVQ9	<i>Synechococcus</i> sp. JA-3-3Ab	87.0	92.0
<i>yfkN</i>	DLIAKIIK_27271	K7QYA3	<i>Thermus oshimai</i> JL-2	92.8	96.6
<i>sbcD</i>	GHLEOJJE_69169	Q2JH63	<i>Synechococcus</i> sp. JA-2-3B'a (2-13)	84.1	90.6
<i>Coche</i>	COAJDPCK_44304	A0A0X8WYU2	<i>Leptolyngbya</i> sp. O-77	90.9	94.3
<i>gdpP</i>	DJGGFCOI_76174	W7ZQT2	<i>Alkalihalobacillus</i>	99.2	99.7
<i>cpdA</i>	GHLEOJJE_41610	Q2JRV1	<i>Synechococcus</i> sp. JA-3-3Ab	87.7	94.5
<i>cpdA</i>	ANHKAJAH_49485	A0A2D2Q0Y6	<i>Synechococcus lividus</i> PCC 6715	90.5	94.3
<i>cpdA</i>	ANHKAJAH_66450	A0A2D2Q1R0	<i>Synechococcus lividus</i> PCC 6715	84.9	89.6
<i>pgpH</i>	GHLEOJJE_121925	A0A2M8P3J3	Candidatus Thermofonsia Clade 1	90.8	96.3
<i>pgpH</i>	AJOHPFOL_66943	A9WAZ5	<i>Chloroflexus aurantiacus</i> ATCC 29366	97.5	98.3
<i>pgpH</i>	AICLDICD_22704	A9WAZ5	<i>Chloroflexus aurantiacus</i> ATCC 29366	95.5	98.3
Caur_0528	DJGGFCOI_90839	A9WEB5	<i>Chloroflexus aurantiacus</i> ATCC 29366	97.4	98.4
CYA_1044	GHLEOJJE_100583	Q2JQP2	<i>Synechococcus</i> sp. JA-3-3Ab	87.2	92.7
CFE21_21820	JFEGIKBN_121484	A0A257L0P5	Bacteroidetes bacterium B1(2017)	72.9	86.0

BRW62_08500	ANHKAJAH_41397	A0A2D2Q2Q2	<i>Synechococcus lividus</i> PCC 6715	84.0	92.1
CUN51_05065	GHLEOJJE_136369	A0A2M8P121	Candidatus Thermofonsia Clade 1	91.5	95.6
Proteína putativa	ANHKAJAH_13884	A0A2D2Q3T0	<i>Synechococcus lividus</i> PCC 6715	90.3	93.1
BRW62_11120	ANHKAJAH_66462	A0A2D2Q4K8	<i>Synechococcus lividus</i> PCC 6715	89.8	95.8
C3497_13595	COAJDPCK_45782	A0A2K9LDJ6	Zoogloeaceae bacterium Par-f-2	96.7	98.0
JCM19046_2533	DJGGFCOI_137375	W7ZE12	<i>Bacillus</i> sp. JCM 19046	99.7	99.7
RoseRS_2931	AICLDICD_51382	A5UXE6	<i>Roseiflexus</i> sp. RS-1	88.4	93.3
<i>ybfK</i>	AICLDICD_49970	A5UV33	<i>Roseiflexus</i> sp. RS-1	89.4	92.8
<i>ybfK</i>	ADLBLLLF_29905	A5UV33	<i>Roseiflexus</i> sp. RS-1	89.4	92.8
<i>nlhH</i>	ADLBLLLF_33318	A0A541BXA0	<i>Elioraea</i> sp. (Yellowstone)	100.0	100.0
<i>nlhH</i>	DAIGAHND_28238	A0A4R7PGC8	<i>Panacagrimonas perspica</i>	74.2	83.3
<i>nlhH</i>	DAIGAHND_31952	A0A059FSS3	<i>Hyphomonas hirschiana</i> VP5	75.0	85.8
<i>nlhH</i>	JFEGIKBN_125503	UPI0012F91CAC	<i>Sphingomonas formosensis</i>	74.9	84.2

Las esterasas participan en la degradación de ciertos agroquímicos como pesticidas piretroides, organofosforados y carbamatos (Hang et al., 2012, Bhatt et al., 2020, Bhatt et al., 2021), estas enzimas participan en el primer paso de la hidrólisis del enlace éster (-COOH), y esta reacción metabólica acelera la degradación primaria de los pesticidas. La hidrólisis de los enlaces éster puede aumentar la solubilidad en agua, la escisión hidrolítica genera alcoholes y ácidos carboxílicos. Existen algunos genes como *cehA*, *mcd* y *pytH* que participan en la hidrólisis de estos compuestos (categorías EC. 3.1.8 y EC 3.1.1), sin embargo, ninguno de estos genes fue encontrado en los tapetes microbianos, estos se han encontrado en análisis de metagenomas de rumen de vaca, aguas de tratamiento, suelo y bacterias como *Pseudomonas*, *Plesiomonas*, *Arthrobacter*, *Methylobacterium* y *Sphingobium*. La metagenómica funcional permite encontrar nuevos biocatalizadores al proporcionar acceso al acervo genético de la comunidad microbiana de un hábitat. Específicamente en ambientes termales, se han encontrado genes que codifican para esterasas novedosas, tal como *OB4Est*, que codifica una nueva esterasa de la familia VIII y tiene una secuencia similar a las β -lactamasas, pero con una especificidad de sustrato amplia (López-López et al., 2015), y *EstDZ2* que codifica para una esterasa y fue encontrada en datos de metagenómica shotgun y clonada en *Escherichia coli*, a partir de muestras ambientales de la fuente termal Kamchatka, Rusia (25 °C a 60 °C) (Zafareta et al., 2016).

Amilasas

La base de datos creada de amilasas contenía un total de 259 secuencias, de las cuales 88 eran representativas y se utilizaron para el mapeo. Estas variantes se agruparon en ocho genes, tales como *amy*, *amyA*, *amyB*, *amyM*, *amyS*, *cga*, *iam* y *treS*. Basado en los resultados del mapeo de los genes y los análisis de abundancia (Fig. 11), se encontró que las variantes de los genes *amy* se encuentran en todas las fuentes termales. En el cuadro 18, se muestran únicamente las variantes que tenían un número de secuencias mapeadas mayor a 2000 (> 0.02%). La mayoría de las variantes tienen un porcentaje de similitud mayor al 90% con proteínas encontradas en genomas de cianobacterias como *Leptolyngbya* o bacterias del filo Chloroflexi. Los genes *amy* codifican para α -amilasas (EC 3.2.1.1), estas enzimas son de la familia GH-13 de gliosil hidrolasas, las cuales hidrolizan al azar los enlaces α -1,4 glicosídicos del almidón, sin pasar por los puntos de ramificación y liberando el límite α -dextrinas como subproductos (Gupta et al., 2003, Sivaramakrishnan et al., 2006).

Diversos estudios demuestran que este grupo de enzimas pueden mantener sus funciones óptimas desde los 50 °C a los 100 °C y pH desde los 2 a 11 (Leal Dalmaso, et al., 2015, Sivaramakrishnan et al., 2006). Como se mencionó anteriormente, estas enzimas participan en la degradación de carbohidratos, principalmente almidón. Si bien la amilasa se considera una enzima específica para la hidrólisis del almidón, Karimi y Biria (2019) mencionan que la función de estas enzimas puede tener un amplio rango de sustratos (promiscuidad), ya que se ha encontrado que puede degradar otros compuestos como polietileno de baja densidad. Existe evidencia de que podrían estar involucradas en la degradación de insecticidas organofosforados como clorpirifos (Savitha y Raman, 2012). También se ha evidenciado que la aplicación de insecticidas en cultivos como maní incrementa la actividad de amilasas y celulasas en el suelo (Mohiddin et al., 2015), lo que sugiere que los microorganismos podrían estar utilizando estos compuestos como fuente de energía o tener mecanismos celulares para eliminarlos

Cuadro 18. Resultados de UniProt-Blast para las variantes de amilasas más abundantes en todas las fuentes termales.

Gen	ID	Secuencia de referencia	Organismo	% identidad	Fracción del alineamiento (%)
Alfa-amilasa	CAEMNFDP_05348	A5UPA4	<i>Roseiflexus</i> sp. RS-1	92.4	95.1
Alfa-amilasa	DJGGFCOI_75850	UPI00195A508E	<i>Alkalihalobacillus xiaoxiensis</i>	97.3	99.0
<i>amyA</i>	GHLEOJJE_99975	A0A2M8P087	Candidatus <i>Thermofonsia</i> Clado 1	86.8	93.1
<i>amyA</i>	DLIAKIIK_00805	A0A0N7MYV5	Candidatus <i>Chrysopegis kryptomonas</i>	86.6	95.1
<i>amyA</i>	GHLEOJJE_30810	A0A395M152	Candidatus <i>Thermochlorobacter aerophilum</i>	68.0	80.1
<i>amyB</i>	GHLEOJJE_99805	A0A2M8P2Y7	Candidatus <i>Thermofonsia</i> Clado 1	88.2	93.9
Amy	ANHKAJAH_33412	A9WA30	<i>Chloroflexus aurantiacus</i> ATCC 29366	98.3	99.0
Amy	COAJDPCK_114273	A0A6M8BDW7	<i>Thermoleptolyngbya</i> sp. PKUAC-SCTA183	90.7	94.3
<i>amyS</i>	COAJDPCK_129467	A0A0X8WRD9	<i>Leptolyngbya</i> sp. O-77	95.2	97.9
<i>amyS</i>	GHLEOJJE_49166	A0A0X8WRD9	<i>Leptolyngbya</i> sp. O-77	93.2	97.9

Genes involucrados en la remoción de metales y agroquímicos encontrados en los tapetes microbianos

Después de la anotación y la construcción de la base de datos, se procedió a realizar el mapeo y la búsqueda de genes posiblemente relacionados con la remoción de metales como arsénico, cobre, hierro y plomo, y agroquímicos como atrazina, paraquat, ametrina y bromacil.

Se han descrito algunas enzimas hidrolíticas relacionadas con la remoción de metales o degradación de agroquímicos (Dave et al., 1993, Russell et al., 2011, Jiang et al., 2020, Bhandari et al., 2021). En este estudio se describen los principales genes (16 en total) que codifican enzimas que participan en la remoción y degradación de los agroquímicos paraquat y atrazina, algunos de los cuales son transportadores que la célula utiliza para múltiples funciones. En el caso de bromacil y ametrina no se encontró información o secuencias de genes en los metagenomas que podrían participar en la remoción del mismo. Por otro lado, se describen los genes que le pueden conferir a los microorganismos de tapetes microbianos tolerancia a los diferentes metales, los cuales pueden participar en la salida (pasiva o activa) o acumulación. En la figura 15, se puede observar la abundancia de cada uno de los genes encontrados en los tapetes microbianos que confieren a los microorganismos tolerancia a arsénico, cobre y hierro. El porcentaje de similitud con las secuencias de referencia varía entre 60% al 90% en el 80% del alineamiento. En futuras investigaciones se podrían realizar mejoras a la continuidad en los *contigs*, como por ejemplo secuenciación de fragmentos largos, lo que permitiría encontrar operones completos.

A continuación, se detalla para cada metal los resultados obtenidos.

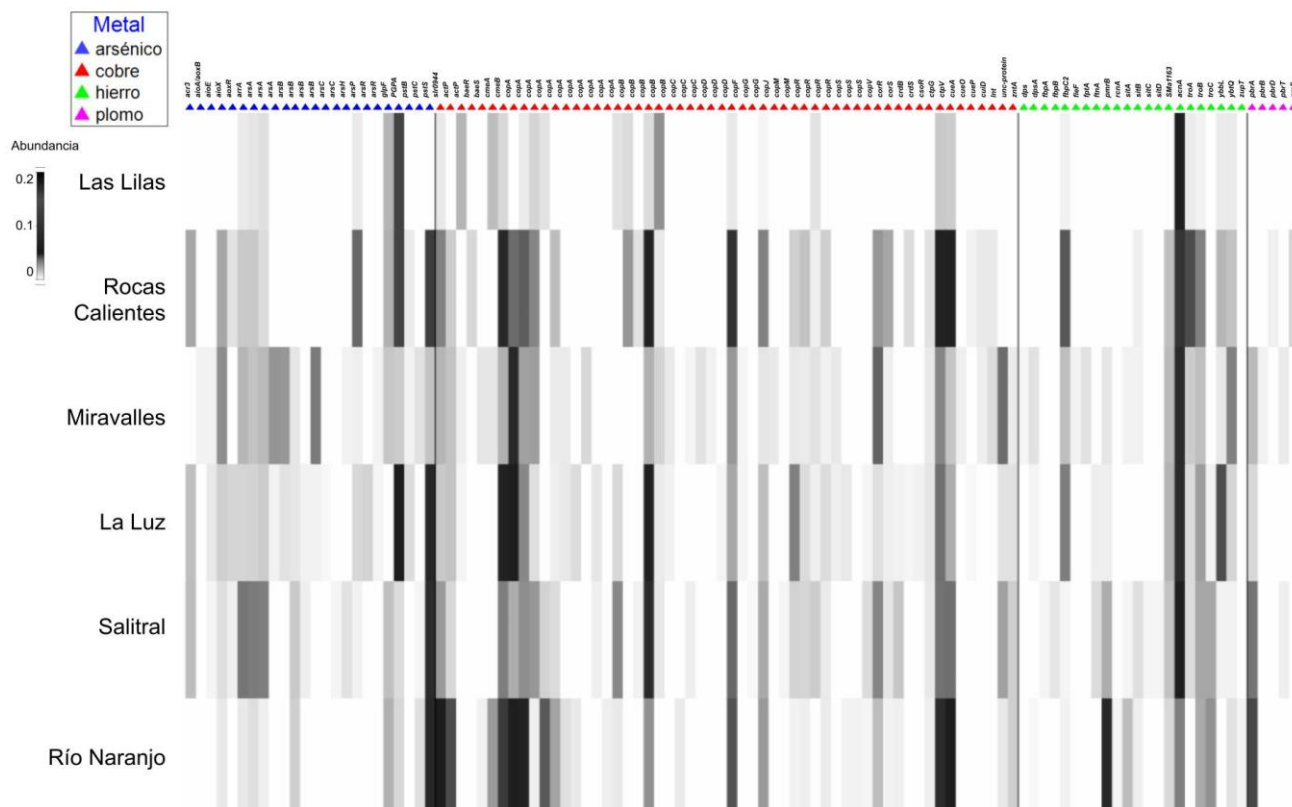


Figura 17. Gráfico de sombras de la abundancia relativa de cada uno de los genes involucrados en la tolerancia a metales encontrados en los tapetes microbianos. Datos previamente transformados utilizando raíz cuadrada.

Arsénico

Utilizando la base de datos BacMet, se encontraron 48 genes de resistencia y/o remoción a arsénico de los cuales solamente 25 fueron encontrados en los tapetes microbianos analizados. La abundancia de cada uno de los genes por muestra se observa en la figura 17, según esta base de datos y las referencias asociadas, algunos de estos se han encontrado en plásmidos principalmente. Basado en las clasificaciones propuestas por Escudero y colaboradores (2013), los genes encontrados en este estudio se agrupan en cuatro categorías: los operones *ars* y *arr* para la reducción de As (V), operón *aio* de oxidación As (III) y As (V) y el gen *acr3* para el transporte fuera de la célula. Los resultados obtenidos sugieren que los microorganismos de los ambientes termales tienen diferentes estrategias para tolerar este metal.

Al realizar el mapeo en los MAGs, los genes del operon *arsRDABC* y *aioXSRBA* fueron encontrados en 23 MAGs (de estos en 10 también se encontraron los operones de resistencia a cobre *copBAC* y *copMRS*). El porcentaje de identidad con los genes de estos MAGs y los de la base de datos BacMet, *Escherichia coli* (X80057), *Pseudomonas stutzeri* (AY702476), *Pseudomonas aeruginosa* (AF010234) y *Staphylococcus aureus* (M86824) varió entre un 70% a un 95% (Figura 18). Para el operón *aioXSRBA* en la mayoría de MAGs los genes tuvieron un valor de identidad menor al 70%, excepto en tres de los MAGs asociados a Rhodocyclaceae y Caldilinea, en los que se obtuvieron un valor mayor al 85%.

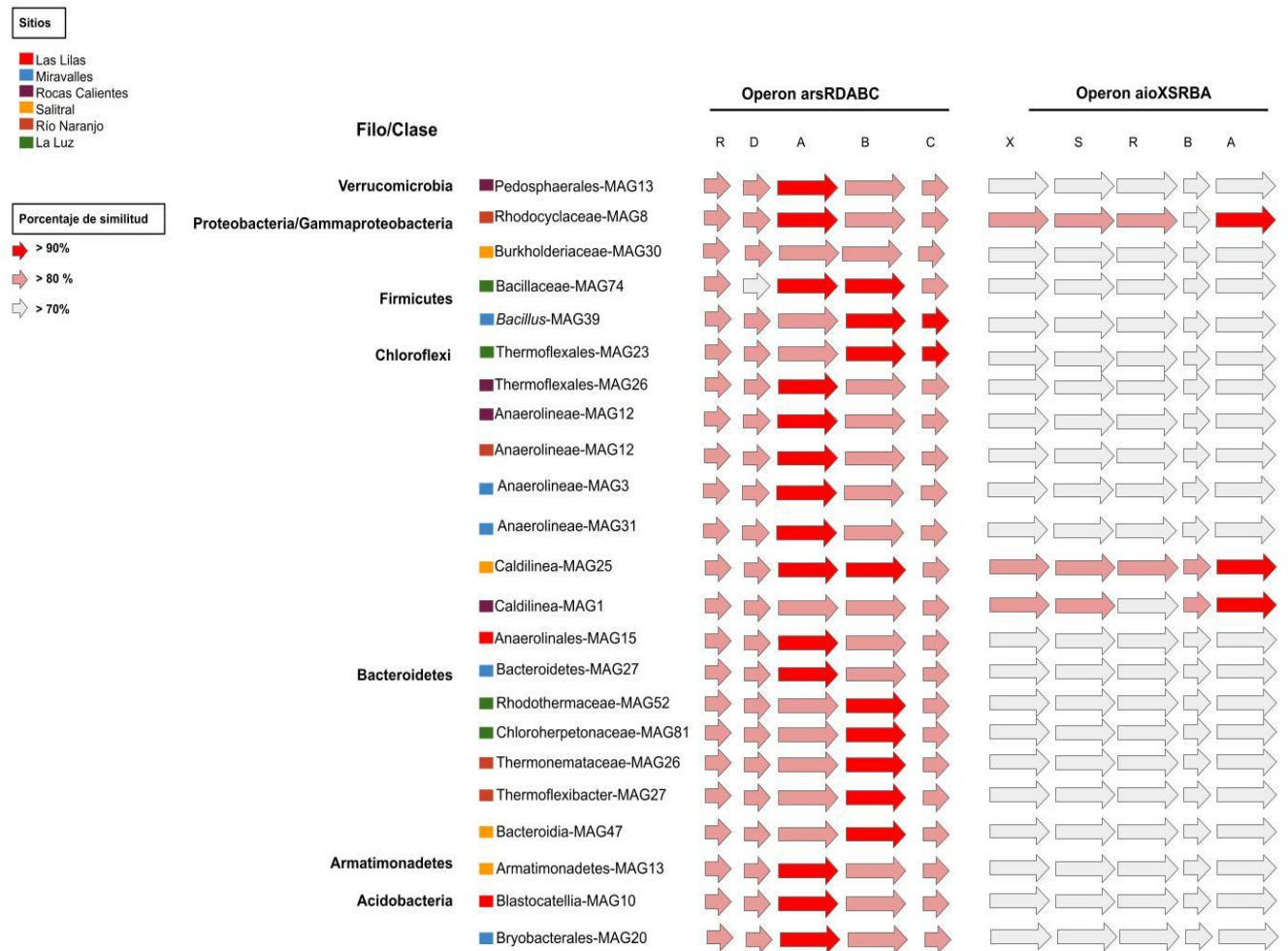


Figura 18. Representación de los operones arsRDABC y aioXSRBA en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio.

La mayoría de los genes encontrados en estas fuentes termales participan en procesos de oxidación-reducción del arsénico, tales como *aio/aox*, *arsC* y *arr*, otros como *pts* y *glpF* favorecen la entrada del metal a la célula (Badilla et al., 2018, Ospino et al., 2019), que podrían ser objeto de interés en investigaciones sobre la acumulación del metal dentro de la célula y la tolerancia de estos microorganismos (Fig. 19). Los genes *acr3*, *arsA* y *arsB*, que codifican para proteínas de membrana que actúan como canales para expulsar el arsénico de la célula, también fueron encontrados en los tapetes microbianos de estas fuentes termales.

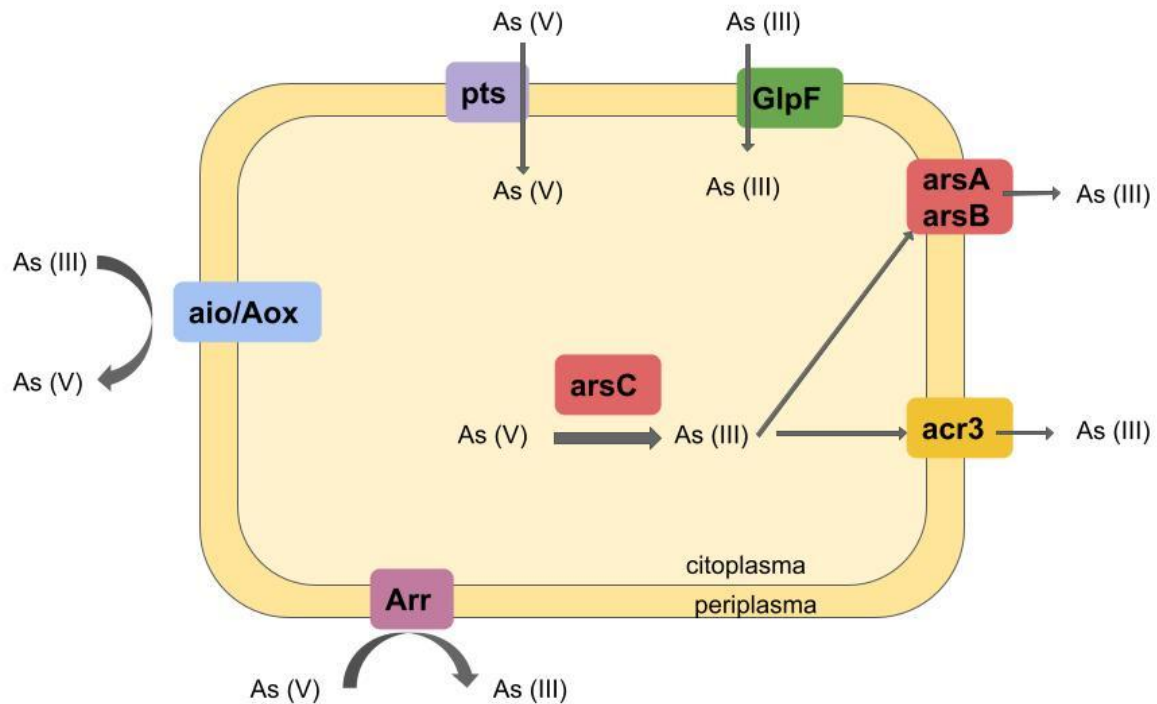


Figura 19. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al arsénico. Elaboración propia.

Todos estos sistemas han sido descritos principalmente en microorganismos termofílicos como *Hydrogenobacter*, *Hydrogenobaculum* y *Thermocrinis rubber* (Jiang et al., 2016), sin embargo, se han descrito en otras bacterias como *Staphylococcus*, *Bacillus* y *E. coli*. Chloroflexi es uno de los grupos más abundantes en la mayoría de las fuentes termales en este estudio (Fig. 9). Según Engel et al., (2012), el gen *aio* le confiere resistencia a arsénico a microorganismos como *Chloroflexus aurantiacus*. Además, la abundancia de esta bacteria es mayor en fuentes termales (< 70 °C) con una concentración de As (III) alta, por lo que se favorece la oxidación del arsénico (As III a As V) fuera de la célula. El As III puede unirse fácilmente a transportadores ubicados en la membrana y ser movilizado al interior de la célula, por lo que en proceso de oxidación reduce la entrada de arsénico a la célula. Por otro lado, en cianobacterias como *Synechocystis* PCC 6803 se han descrito detalladamente algunos mecanismos de resistencia y remoción de arsénico (López-Maury et al., 2003, Huértas et al., 2014) dentro de los que se encuentran algunos genes que fueron mapeados en este análisis.

Es muy importante mencionar que la mayoría de las fuentes termales poseen arsénico en concentraciones que se consideran tóxicas para el consumo humano (mayor a 10 µg/L).

Las muestras utilizadas en este análisis no cuentan con datos relacionados a la concentración de arsénico, sin embargo, otros investigadores han reportado que la concentración en fuentes termales costarricenses va desde los 0.01 mg/mL a los 25 mg/mL (López et al., 2012, Mollinedo, 2013). Según algunas investigaciones mencionadas por Connon y colaboradores (2008), la concentración de arsénico en aguas geotérmicas se encuentra entre 1 y 10 mg /L, sin embargo, en otras fuentes termales como algunas del Parque Nacional Yellowstone, EE. UU y el Tatio en Chile, se han reportado concentraciones de hasta 150 mg/L y 50 mg/L respectivamente (Langner et al., 2001, Romero et al., 2003). En las fuentes termales el arsénico puede estar presente como varias especies químicas, incluidos los oxianiones arsenito (AsO_3^{-3}) y arsenato (AsO_3^{-4}), así como los tioaniones de arsénico (Hug et al., 2014). La especiación de arsénico está controlada principalmente por el pH y el potencial redox (Vu et al., 2003), así como por procesos microbiológicos en donde ocurre la biosorción, precipitación, transformación redox y metilación (Crognale et al., 2017). Trabajos previos sobre la especiación de arsénico en ambientes termales han reportado la presencia principalmente de arsenito y arsenato (Macur et al., 2004, Hug et al., 2014). El arsenito tiene una alta afinidad por los grupos sulfhidrilo en los aminoácidos, lo que interrumpe la función de las proteínas (Oremland y Stolz, 2003). El arsenato es un análogo del fosfato, que desplaza los iones fosfato en las reacciones enzimáticas y, por lo tanto, interfiere con el metabolismo celular (Oremland y Stolz, 2003).

El arsenito, As III, puede ser la única especie de arsénico en aguas calientes anaeróbicas donde se convierte rápidamente en arsenato, As V, debido a oxidaciones microbianas (Jackson et al., 2001, Langner et al., 2001). En fuentes termales como Champagne Pool en Nueva Zelanda (Hug et al., 2014, Suhadolnik, et al., 2017), también se ha reportado la presencia de los genes de resistencia a arsénico encontrados en este estudio, lo que demuestra que los microorganismos de diferentes fuentes termales tienen estrategias similares de resistencia a diversos metales. En todas las fuentes estudiadas se encontraron genes que le proveen a los microorganismos diferentes estrategias para absorber, expulsar o transformar el arsénico. Algunos de ellos, como *arsA*, *arsB* y *arsC* se han encontrado en plásmidos de microorganismos como *E. coli*, *S. aureus*, *B. subtilis*, entre otros (Ben Fekih et al., 2018).

Existen diversos plásmidos reportados en la literatura que le confieren a los microorganismos resistencia arsénico, entre ellos el pR773, pl258, pSX267, arsRBC o arsRDABC encontrados en cepas de *E. coli*, *Yersinia* spp., *Acidiphilium multivorans* AIU 301,

Serratia marcescens, *Halobacterium* sp. NRC-1 y *Sinorhizobium* sp. cepa M14 (aislada del sedimento de suelo rico en arsénico de una mina de oro) (Ben et al., 2018). La presencia de genes de resistencia al arsénico en plásmidos o transposones constituye una oportunidad para que los microorganismos puedan transferirlos (mediante transferencia horizontal), lo que fomenta la diseminación y, por ende, se transfiera la tolerancia o resistencia a las condiciones fisicoquímicas de las fuentes termales.

Cobre

En BacMet existen 133 genes asociados a la resistencia a cobre, los cuales en algunos casos también confieren resistencia a otros metales como zinc y cobalto. En las fuentes termales solamente mapearon 56 de estos genes, los cuales fueron encontrados en diferentes abundancias en las muestras (Fig. 17). Estos se han descrito previamente en cromosomas, plásmidos y transposones de diferentes bacterias como *Mycobacterium tuberculosis*, *Xanthomonas* spp y *Pseudomonas* spp. En la figura 20 se muestra el esquema de una célula y todos los posibles mecanismos celulares que se encontraron en estos análisis relacionados con la resistencia al cobre. Algunos de ellos participan en la importación y acumulación del metal dentro de la célula o espacio periplásmico, lo que puede ser de utilidad para futuros estudios de remoción del metal.

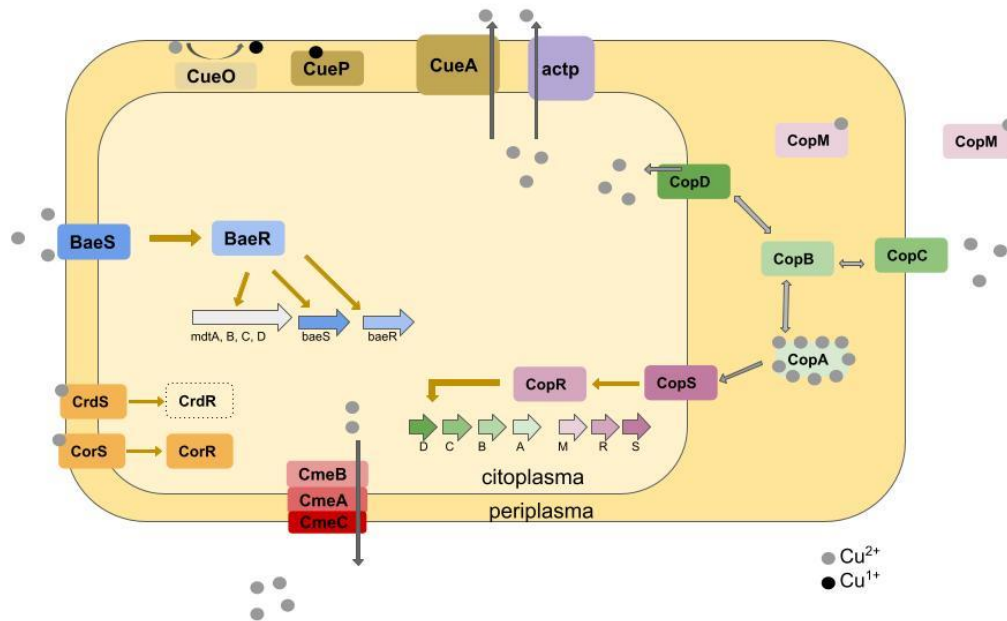


Figura 20. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al cobre. Elaboración propia.

El gen *actP* participa en la exportación de cobre fuera de la célula (González-Sánchez et al., 2018), según los resultados mostrados por Landeta y colaboradores (2011) en mutantes de *Rhizobium*, la ausencia de este gen causa hipersensibilidad al cobre. BaeRS es un sistema de dos componentes (TCS), el cual participa en la activación de genes de resistencia a metales. BaeS es un sensor de histidina quinasa que regula a BaeR, el cual es importante para la regulación de la salida de fármacos y metales, así como para la supervivencia celular (Appia-Ayme et al., 2011). Este sistema de dos componentes regula el operón *mdtABCD* y a los genes *baeR* y *baeS*, que codifica un sistema de transporte de múltiples fármacos y metales como cobre y zinc (Fig. 20) (Nishino, et al., 2007, Zoetendal et al., 2008). CmeABC, es una bomba de eflujo dentro de la familia MDR, que consta de una proteína de fusión periplásmica (CmeA), un transportador de flujo de la membrana interna que pertenece a RND (CmeB) y una proteína de la membrana externa (CmeC) (Lin et al., 2002) (Fig. 18). Los genes *cmeA* y *cmeB* presentes en plásmidos les confieren a bacterias patógenas como *Campylobacter jejuni* resistencia a antibióticos, metales pesados como cobre, sales biliares y agentes antimicrobianos (Lin et al., 2002, Pumbwe et al., 2004).

Los genes *copA*, *copB*, *copC*, *copD*, *copF*, *copG*, *copJ*, *copM*, *copR*, *copS* y *copV* participan en la resistencia al cobre en diversos microorganismos, algunos de ellos le confieren a la célula resistencia al metal y favorecen la acumulación de este, siendo de gran interés para

procesos de remoción de metales (Shafeeq et al., 2011), estos han sido unos de los genes más estudiados en cuanto a la resistencia al cobre. Según Puig y colaboradores (2002), el gen *copA* codifica una proteína periplásmica rica en metionina capaz de unir hasta 11 átomos de cobre y es homóloga a la ferroxidasa Fet3, la cual tiene alta afinidad por el hierro en levaduras, *copA* le confiere a las células bacterianas resistencia al cobre. Por otro lado, *copB*, codifica para una proteína rica en residuos de metionina, que es una proteína integral de la membrana externa que se une a CopA y CopC para expulsar el cobre fuera de la célula. La proteína CopC, es una pequeña proteína soluble que se encuentra en el espacio periplásmico y que se ha demostrado que se une a un átomo de cobre, mientras que CopD contiene varios dominios transmembrana putativos en la membrana interna. CopC y CopD participan en el transporte de cobre. Estos mismos autores señalan que el mecanismo de resistencia al cobre depende del secuestro y la acumulación en el periplasma y la membrana externa, lo que en algunas veces causa una coloración azul en las células.

Los operones *copMRS* y *copBAC* se encontraron en 67 MAGs, la similitud entre los diferentes genes de los MAGs y los de referencia varía desde un 60% a un 90% tomando en cuenta el 80% de la longitud de los fragmentos (e-value < 0.00001) (Figura 21). Este operón se encontró en diferentes filos de bacterias, pero principalmente en Chloroflexi (18) y Cyanobacteria (9). El gen *copA* fue en el que se encontró una mayor similitud (> 90%) en comparación con *Pseudomonas*, *Mycobacterium* (BacMet), *Xanthomonas citri* (HM362782) y *Synechocystis* sp. PCC 6803 plásmido pSYSX (AP006585), por lo que se podría considerar el más conservado, a diferencia de *copC*, *copB* y *copF* (ATPasa). También se encontraron los genes *copG*, *copV* y *copD* principalmente en Proteobacterias. Es importante mencionar que en Methylococcoides-MAG13 también se encontró el operon *cmeABC* (similitud mayor al 80%) y en el Roseiflexus-MAG26 se encontró el operón *baeRS* (similitud mayor al 90%), no obstante, no se encontraron en ninguno de los MAGs restantes.

En cianobacterias como *Synechocystis* la resistencia a cobre comprende el sistema de dos componentes, CopRS, CopM la cual es una proteína de unión de cobre periplásmica y extracelular y el sistema de exportación *copBAC*. CopS (histidina quinasa) es capaz de detectar cobre y probablemente activa CopR, que es capaz de unirse directamente a los promotores de los operones *copMRS* y *copBAC* (Fig. 20) (Huertas et al., 2014).

Los genes que forman el operon de regulación *copMRS* tuvieron una similitud entre 60% y 70% con respecto a los de referencia en la mayoría de MAGs (excepto en Cyanobacteria y Chloroflexi). Esto podría indicar variación genética o que en otros filos existen otros

mecanismos de regulación, y que el operon copBAC está regulado por otro sistema, o simplemente no forman parte del ensamblaje de los MAGs.

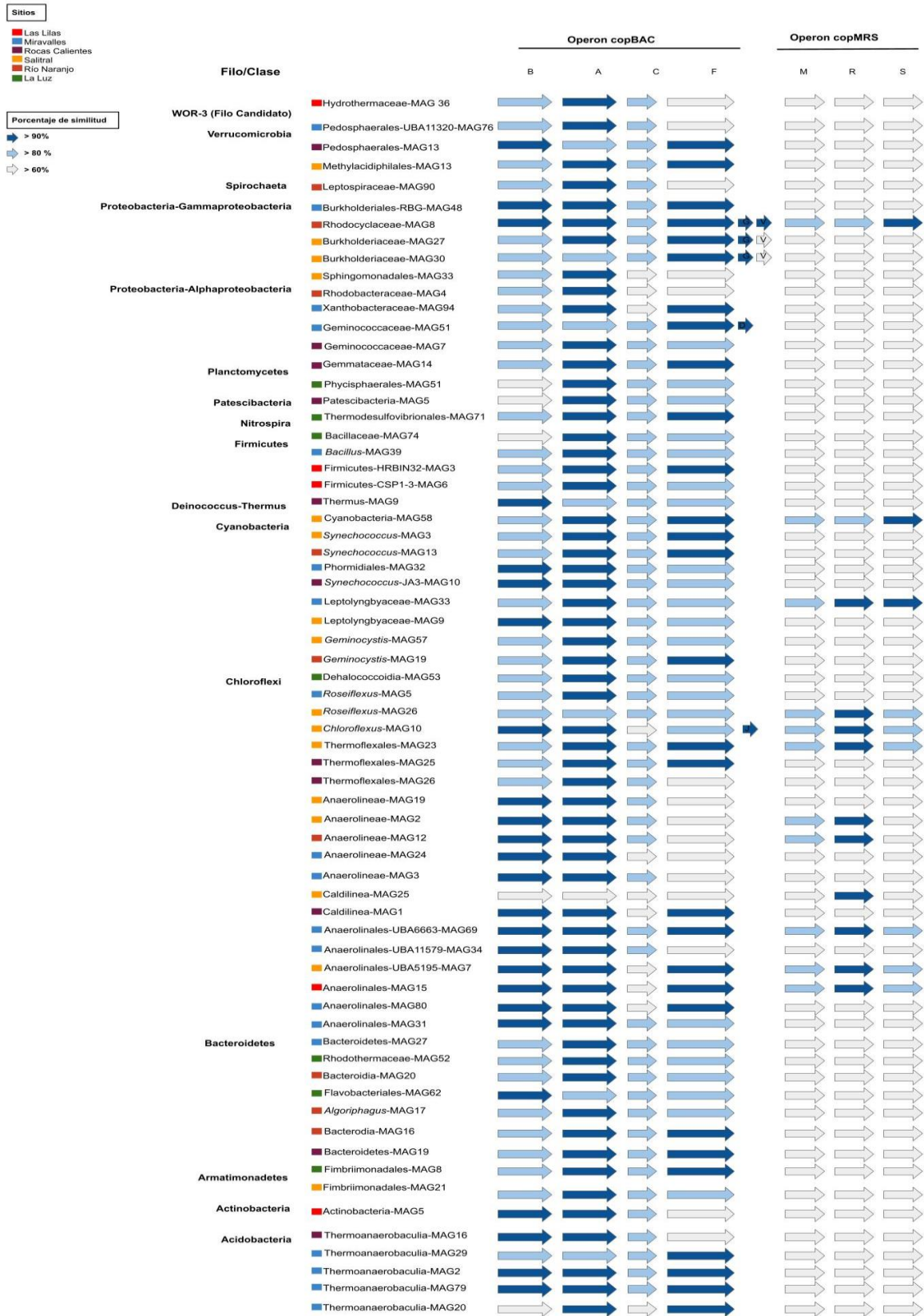


Figura 21. Representación de los operones copBAC y copRMS en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio.

CorS es una histidina quinasa que se encuentra en la membrana y es sensible al cobre y CorR es el regulador de respuesta (Sánchez-Sutil et al., 2013). Ambos participan en la activación de diversos mecanismos que regulan la concentración de cobre dentro de la célula. Por otro lado, en presencia de cobre, el sensor quinasa CrdS fosforila el regulador de respuesta CrdR desencadenando la activación de una proteína de resistencia y un complejo de expulsión del metal (Haley y Gaddy, 2015). En este análisis no se encontraron secuencias mapeadas para el gen *crdR*. El gen *crdS* es abundante (0.02%) en Rocas Calientes principalmente, sin embargo, se encontró también en La Luz.

CsoR es un factor de transcripción que fue descrito por primera vez en *Mycobacterium tuberculosis*, ayuda a mantener la homeostasis intracelular cuando la concentración de cobre aumenta, y además participa en la respuesta de protección al estrés oxidativo (Liu et al., 2007). CtpV se ha caracterizado como un exportador de cobre en *Mycobacterium tuberculosis* y participa en la virulencia de estos microorganismos (Ward et al., 2010).

El gen *cueA* codifica por una cobre-ATPasa, la cual es relevante en la tolerancia del cobre en células de vida libre (Rademacher y Masepohl, 2012). CueP, es esencial para la tolerancia al cobre, particularmente en condiciones anaeróbicas, además, contribuye a la tolerancia al cobre en condiciones aeróbicas. Esta enzima es más abundante en el periplasma. La inducción de CueP depende estrictamente de CueR, que se une a una secuencia promotora palindrómica altamente similar al sitio de unión (Espariz et al., 2007). Por otro lado, CueO participa en la reducción de Cu^{2+} a Cu^{+1} (Pontel et al., 2015).

No existen datos de la concentración de cobre para estas fuentes termales costarricenses relacionados directamente con las muestras de tapetes microbianos analizados. Según un análisis metagenómico en nueve fuentes termales en China con rango de temperatura de 46 °C a 96 °C, pH entre 2.3 y 9 y una concentración de cobre desde los 3 nM a los 38 nM, existe una amplia cantidad de genes asociados a la unión y oxidorreducción del metal, los cuales se encuentran en todas las fuentes termales estudiadas (Srivastava et al., 2020). La biodisponibilidad de cobre es una oportunidad y un desafío para todos los microorganismos, ya que participa como cofactor en enzimas que están involucradas en diversas vías metabólicas, y como ion libre puede ser tóxico (Grass et al., 2011). En algunas bacterias se han reportado concentraciones tóxicas mayores a 2 μM Cu^{2+} (Jaime-Pérez, 2019). En 2012, Núñez demuestra en su investigación la remoción en solución acuosa de este metal por la cianobacteria termofílica *Leptolyngbya* sp. 7M aislada de Miravalles, lo que sugiere que los microorganismos aislados de este sitio tienen una alta capacidad de remover este

compuesto, ya sea mediante interacción con la membrana o con el transporte de este dentro de la célula.

La regulación de cobre en la mayoría de las bacterias es producto de la actividad coordinada de diferentes familias de los exportadores de metales (ATPasas tipo P1B y bombas), enzimas desintoxicantes de cobre (multicopper oxidasas), metalochaperonas y reguladores sensibles al cobre (Rademacher y Masepohl, 2012). Según investigaciones en fuentes termales en China, el transporte de cobre fuera y dentro de la célula, así como la biodisponibilidad, se ven afectados por las condiciones fisicoquímicas como cambios en temperatura, pH y concentraciones de otros metales (Srivastava et al., 2020).

Plomo

Se encontraron nueve genes asociados a la resistencia a plomo en la base de datos BacMet, de los cuales solamente cinco mapearon en las fuentes termales de Costa Rica. Ninguno de estos genes se encontró en todas las fuentes termales, y en algunos casos su abundancia fue muy baja (< 0.0002) (Fig. 22). El operón de resistencia al plomo, *pbr*, les permite a los microorganismos capturar, acumular y expulsar plomo, se encuentra principalmente en plásmidos. El locus de resistencia al plomo *pbr* contiene los siguientes genes de resistencia: (i) *pbrT*, que codifica una proteína de captación de Pb (II), (ii) *pbrA*, codifica una ATPasa de exportación de Pb (II) de tipo P, (iii) *pbrB*, codifica una proteína de membrana integral predicha de función desconocida, y (iv) *pbrC*, la cual codifica una peptidasa señal de prolipoproteína y (v) *pbrD*, que codifica una proteína de unión a Pb (II), la cual es esencial para el secuestro de plomo (Borremans et al, 2001) (Fig. 22). En las fuentes termales costarricenses y otras partes del mundo la concentración de plomo es muy baja (cerca a 0), y en muchos casos no es detectable (Shakhatrehz et al, 2017), por lo tanto, es interesante encontrar estos genes en estos ambientes. Se puede presumir que la persistencia de un gen en plásmidos en ausencia de selección positiva puede estar determinada por varios factores como la estabilidad del plásmido, características del hospedador y las condiciones ambientales (Wein et al., 2019).

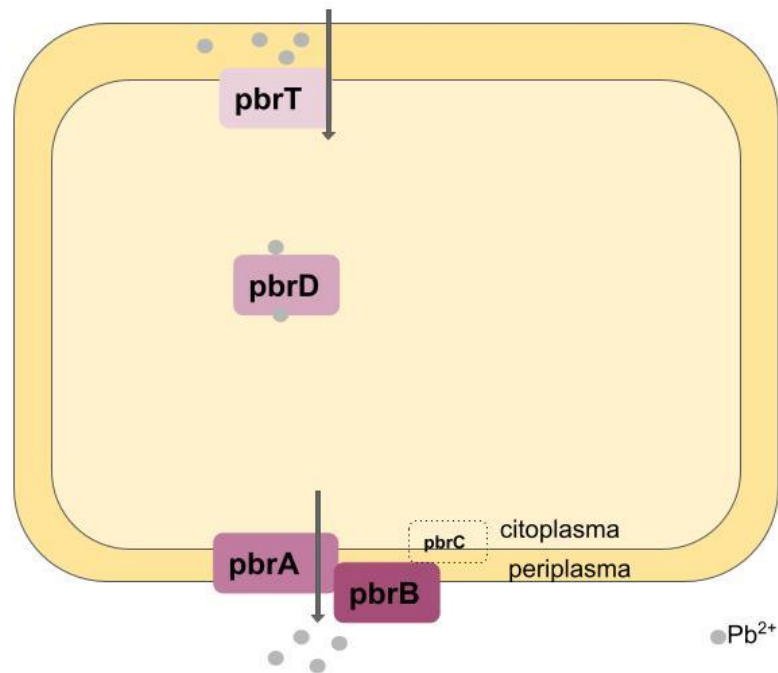


Figura 22. Representación de los diferentes genes encontrados en los tapetes microbianos de fuentes termales costarricenses que le confieren a las comunidades microbianas de tapetes microbianos de ambientes termales resistencia al plomo. Elaboración propia.

Hierro

Se encontraron 41 genes de resistencia a hierro, de los cuales solamente 22 mapearon contra los metagenomas de las fuentes termales (Fig. 14). Se encontraron específicamente genes relacionados a la importación y almacenamiento del metal dentro de la célula, los cuales pueden utilizarse como blanco de estudio en proyectos de biorremediación de metales en donde participen microorganismos aislados de los tapetes microbianos.

Dps es un gen que codifica para una proteína de unión a hierro, la cual participa en la homeostasis y biodisponibilidad del metal, también puede unirse al ADN y participan en la desintoxicación de H₂O₂ (Ishikawa et al., 2003, Moparthi et al., 2016). Estos genes se encontraron principalmente en el tapete microbiano de Miravalles.

El operón *fbpABC*, el cual codifica para un sistema transportador de hierro localizado en la membrana citoplasmática, se requiere para la utilización de múltiples sideróforos estructuralmente distintos, incluidos alcaligina, enterobactina, ferricromo y desferrioxamina B (Brickman et al., 2011, Amin et al., 2013). En respuesta a baja concentración de hierro, muchas bacterias producen sideróforos, quelantes de hierro de bajo peso molecular, que

median el transporte de hierro de alta afinidad a través de receptores de membrana (Ankenbauer y Quan, 1994).

En condiciones limitantes de hierro, el gen *ftpA* codifica para un sideróforo (quelante de Fe³⁺), por otro lado, FtnA es la principal proteína de almacenamiento de hierro en microorganismos como *E. coli*, lo que puede llegar a representar hasta el 50% del hierro total dentro de la célula (Nandal et al., 2010). PmrAB es un sistema bacteriano de dos componentes, el cual es necesario para una respuesta transcripcional adecuada a diversas concentraciones de hierro férrico. El hierro férrico induce a la histidina quinasa PmrB, lo que conduce a diversas respuestas transcripcionales (Guckes et al., 2017). En condiciones con altas concentraciones de hierro, los microorganismos tienen transportadores que reducen la concentración intracelular del metal favoreciendo su salida (Frawley y Fang, 2015)

El operón Tro (troABCDR), se encontró en los tapetes microbianos de todas las fuentes termales. Este operón pertenece a una familia altamente conservada de transportadores de casetes de unión a ATP, involucrados en la captación de cationes metálicos cuya expresión está controlada por TroR, una proteína reguladora sensible a cationes (Saraithong et al., 2020). El gen *ybbL* (*fetA*) codifica para un transportador de metal dependiente de ATP (Nicolaou et al., 2013).

El locus *sit*, que se encontró principalmente en Salitral, consta de cuatro genes que codifican un posible sistema de transporte ABC de metales como hierro. SitA es una proteína de unión periplásmica, SitB, es una proteína de unión a ATP y contiene las cajas de Walker canónicas, SitC y SitD son permeasas de la membrana interna basadas en dominios putativos que atraviesan la membrana (Runyen-Janecky et al., 2003).

YbtP y YbtQ son estructuralmente únicos entre la subfamilia de transportadores ABC asociados con el transporte de hierro, ambos contienen un dominio amino terminal que atraviesa la membrana y una ATPasa carboxilo terminal. Según Fetherston y colaboradores (1999), en *Yersinia pestis* este transportador es esencial para la importación de hierro dentro de la célula en bajas concentraciones del metal. El gen *yjiP* (también llamado gen *fieF*), se ha identificado principalmente en *E. coli* y codifica por un transportador de hierro que pertenece a la familia de facilitadores de difusión de cationes (CDF) (Grass et al., 2005).

En Río Naranjo se encontraron dos transportadores de metal, uno que participa en su exportación y otro que, por el contrario, favorece la entrada a la célula. El transportador RcnA se ha implicado previamente en la desintoxicación de Ni (II) y Co (II) y hierro en *E. coli*, probablemente a través de la exportación del metal (Koch et al., 2007). El gen *zupT* (*ygiE*) codifica una proteína de membrana citoplásmica (ZupT) relacionada con miembros

de la familia ZIP de transportadores de iones metálicos, que participa específicamente en la importación de hierro (Grass et al., 2005).

En las fuentes termales en estudio la concentración de hierro es muy baja o no detectable (no poseen color anaranjado característico), no obstante debido a la naturaleza de las fuentes termales y a las variaciones en pH y temperatura, la concentración del metal en el agua puede cambiar y por ende generar una selección natural positiva en los microorganismos, favoreciendo genes que puedan importar y acumular el metal en condiciones de baja concentración o por el contrario expulsarlo, bajo altas concentraciones (Andrews et al., 2003, Frawley y Fang, 2014).

En algunas fuentes termales, se forman abundantes depósitos de óxido de hierro los cuales albergan comunidades microbianas y podrían proporcionar información importante sobre los depósitos sedimentarios ricos en hierro del Precámbrico y los ecosistemas fototróficos (Walter, 1972, Parenteau y Cady, 2010, Des Marais y Walter, 2019). La acumulación de hierro en las células de cianobacterias como *Synechococcus* y *Cyanothece*, y también el filo Chloroflexi favorece la conservación de estas formando microfósiles. Los depósitos minerales asociados a manantiales ricos en hierro tienen un potencial considerable como objetivos de exploración para sus registros fósiles (Marais y Walter, 2019).

Aluminio

En BacMet se encuentran únicamente dos secuencias asociadas a la tolerancia a aluminio, siendo ambas mapeadas en La Luz y Salitral. Estas corresponden a los genes *queC*, que le confiere a la bacteria termofílica *Anoxybacillus gonensis* G2 resistencia a este metal (Beris et al., 2011), y *alu1*, que, según Jo et al., (1997), le confiere resistencia a aluminio a *Arthrobacter viscosus*. Según los estudios mencionados, estos genes participan en el proceso de biosíntesis de queuosina, un nucleósido modificado que está presente en algunos ARN de transferencia en bacterias, sin embargo, no se han realizado investigaciones en bacterias que amplíen la información y propongan relaciones más específicas entre estos genes y la resistencia a aluminio. En la cianobacteria *Leptolyngbya* sp. 7M aislada de la fuente termal Miravalles, se encontró una alta capacidad de remoción de este metal según análisis de adsorción (Núñez, 2012), sin embargo, no se ha estudiado el mecanismo de resistencia ni se ha realizado la búsqueda de estos genes en el genoma de la cianobacteria. Según datos no publicados del grupo de microbiología ambiental del CIBCM-UCR, la concentración de aluminio generalmente es menor a 3 µg/L o no detectable en algunas de las fuentes termales.

Paraquat

En la base de datos BacMet existen 29 genes de resistencia al paraquat, de las cuales solamente diez obtuvieron secuencias mapeadas en los metagenomas de las fuentes termales costarricenses. La mayoría de estos genes codifican para proteínas de membrana que están involucradas en la exportación de sustancias tóxicas. La abundancia de cada uno de esos genes fue diferente en cada una de las muestras. No se encontraron vías metabólicas relacionadas con la degradación de Paraquat en KEGG. Algunas de las funciones biológicas de las proteínas mapeadas están inferidas por homología, por lo que se encuentran dentro de la categoría de “no revisadas” en UniProt. Es importante mencionar que la mayoría de los genes reportados que confieren resistencia a paraquat codifican para bombas de eflujo de metales, antibióticos y diversas sustancias, descritas principalmente en *Acinetobacter baumannii*, por lo que tienen otras funciones dentro de la célula.

El gen *chr*, encontrado solamente en Salitral, forma parte del operón chrBACF del elemento transponible TnOtChr, el cual le confiere a la bacteria *Ochrobactrum tritici* resistencia al cromato. ChrC es una enzima superóxido dismutasa (SOD) involucrada en la detoxificación del estrés oxidativo (Sánchez-Calderón et al., 2019). Según Branco y colaboradores (2008), el promotor *chr* es fuertemente inducido por cromato o dicromato. La inducción del operón chrBACF suprime la acumulación de cromo en la célula mediante la actividad de una bomba de salida de cromato codificada por *chrA*. Estos mismos autores demostraron que la expresión de *chrB*, *chrC* o *chrF* en un doble mutante de *Escherichia coli* restauró su crecimiento aeróbico en un medio mínimo y confirió resistencia a los agentes generadores de superóxido como menadiona y paraquat. TnOtChr parece representar un sistema genético móvil para la distribución del operón de resistencia regulado por cromato. La presencia de genes que protegen contra la toxicidad del superóxido, debería proporcionar una ventaja adicional a las células que contienen TnOtChr en entornos con múltiples contaminantes. El gen *smfy* pertenece a la superfamilia MFS (Superfamilia de Facilitadores Principales). Existe muy poca información al respecto de las funciones de esta bomba en la célula, según Shahcheraghi y colaboradores (2007), en mutantes de *Escherichia coli* este gen participa en la salida de fármacos, incluidos DAPI, norfloxacin, cloruro de benzalconio, acriflavina y bromuro de etidio.

El operón AcrEF es homólogo a AcrAB (casi 80% de identidad) en *E. coli*, por lo que posiblemente, tiene una función similar (Rodionov et al., 2001). Estas bombas consisten en un conducto de salida de la membrana externa (Koronakis et al., 2000). Posee una proteína adaptadora periplásmica y un transportador de membrana interna, los adaptadores

periplásmicos generalmente tienen una estructura conservada, multidominio, que consta de dominios de membrana proximal (MP), barril β , horquilla y lipoil sintasa (Greene et al., 2013, Hinchliffe et al., 2014). El gen *acrF* codifica para una proteína que forma parte del sistema tripartito AcrEF-TolC. Por otro lado, el gen *acrE* codifica para una proteína que forma parte del sistema del flujo tripartito AcrEF-TolC (similar a AcrAB-TolC), es un sistema de secreción bacteriana que está involucrado en la salida de indol y resistencia a disolventes hidrofóbicos en *Escherichia coli* (Kobayashi et al., 2001). Estas bombas participan en la salida de compuestos como tintes, disolventes orgánicos, antibióticos, moléculas aniónicas, catiónicas, bipolares, aromáticas o alifáticas, que utilizan protones para expulsar estos compuestos de la célula (Pos et al., 2009).

La mayoría de los genes *adeA* se encontraron en Miravalles y Salitral. Los genes *ade*, codifican para proteínas dentro de la categoría de bombas de salida RND (Magnet et al., 2001). En *Acinetobacter* spp, esta bomba está compuesta por las proteínas AdeA, AdeB y AdeC (Wong et al., 2009). La bomba AdeABC consta de tres componentes: *adeA* es la proteína de fusión con la membrana, AdeB es el transportador y AdeC es la proteína externa a la membrana (Xu et al., 2019), *adeABC* está regulado por el operón *adeRS*, cuya expresión aumenta cuando la bomba de eflujo se expone a una concentración excesiva de agentes tóxicos o antibióticos (Mahmoudi et al., 2020).

Las proteínas dentro del sistema AdelJK forman parte del sistema de salida de RND de tres componentes que consta de Adel, que es similar a las proteínas de fusión de membrana, AdeJ, similar a las proteínas RND y AdeK, similar a los OMF (Damier-Piolle et al., 2008). En este mapeo solamente se encontraron secuenciadas para el gen que codifica para Adel en la fuente termal de Salitral. El gen *adeT1* confiere resistencia a antibióticos como cloranfenicol, eritromicina y novobiocina (Bharathi Srinivasan et al., 2011). Según la base de datos UniProt, puede ser parte de un sistema de transporte periplásmico tripartito independiente de ATP. Este gen se encontró únicamente en Miravalles.

El gen *QacE Δ 1* se encuentra principalmente en Rocas Calientes y Miravalles, sin embargo, su abundancia fue menor al 0.01%. Es un transportador SMR es una familia de transportadores de fármacos, está compuesta por cuatro hélices α transmembrana de aproximadamente 100-140 aminoácidos de longitud (Bay et al., 2008). Estos genes se encuentran principalmente en plásmidos, sin embargo, pueden estar también en el cromosoma bacteriano, se encontró que *QacE Δ 1* está completo lo que indica que es funcional. La familia de proteínas SMR tiene capacidad para conferir resistencia a una variedad de compuestos de amonio cuaternario y colorantes catiónicos (Bay et al., 2008).

El segmento 3' específico del integrón (*qacEDelta1-sul1-orf5Delta*) forma parte de un operón de resistencia a muchos antibióticos (Srinivasan et al., 2009). Cabe resaltar la asociación de este gen con integrones de clase 1, los cuales se encuentran en el 9% de los genomas bacterianos secuenciados, y es la clase más ubicua de integrones que existe, además ha sido la más reportada para bacterias (Deng et al., 2015)

Al realizar el mapeo en los MAGs, se encontró que la mayoría de los genes que podrían conferir resistencia a paraquat se encuentran en el filo Cyanobacteria (7 de 20 MAGs) principalmente en genomas identificados como *Synechococcus*, *Leptolyngbya* y *Geminocystis* (Fig.23). En el caso de los genes que forman parte del operón *adeABC* se encontró una similitud entre el 80% y el 90% para la mayoría de los casos, sin embargo, para el operon *chrBACF*, la similitud fue mayor al 70% y menor al 80%, excepto en *chrF*, en el cual se encontró una similitud mayor al 90% en el 85% del alineamiento. Por otro lado, el gen *QacEΔ1* se encontró solamente en *Roseiflexus*-MAG26, *Bacillus*-MAG39 y *Roseiflexus*-MAG5, con más de un 90% de similitud.



Figura 23. Representación de los operones chrBACF y adeABC en los MAGs encontrados en los diferentes tapetes microbianos de las fuentes termales de estudio.

Específicamente en Miravalles, se han aislado cepas de cianobacterias termotolerantes como *Leptolyngbya* sp. 7M con capacidad de remoción y tolerancia a este agroquímico (Brenes-Guillén et al., 2019), de la cual se tiene el genoma completo y actualmente se están buscando los genes relacionados con la resistencia a este herbicida (datos no publicados). Estudios anteriores muestran que las cianobacterias tienen una alta tolerancia a varios pesticidas y herbicidas y pueden metabolizar estas sustancias (Safi et al., 2014, Kumar y Shingh 2017). Las cepas bacterianas y fúngicas adaptadas para crecer y prosperar en ambientes extremos juegan un papel importante en el tratamiento biológico de ambientes contaminados (Margesin y Schinner, 2001, Deshmukh et al., 2016). Estas bombas que le confieren a los microorganismos resistencia al paraquat podrían participar en procesos para la expulsión de antimicrobianos o de algunos compuestos que se generan en procesos geoquímicos en estas fuentes termales.

Atrazina

En los metagenomas en estudio se encontraron los genes *atzB*, *atzD*, *atzE*, *atzF*, *biuH* y *triA* (Cuadro 19), los cuales tienen más de un 90% de similitud con los de la base de datos Bactmet en el 85% del alineamiento. Este es el primer estudio donde se reportan genes asociados a la degradación de este agroquímico en tapetes microbianos de fuentes termales. Se han reportado nueve enzimas que participan en la degradación de atrazina, las cuales fueron encontradas en este análisis, sin embargo, en la mayoría de los tapetes microbianos el número de secuencias mapeadas fue menor al 0.02% (menos de 100 secuencias), lo que indica que se encontraron muy pocas secuencias mapeadas (menos de 100). Bajo la metodología utilizada, al realizar el mapeo en los MAGs, ninguno de los genes encontrados en los metagenomas se encontró en estos genomas. Esto se puede deber a que el porcentaje de completitud que tienen los MAGs varía desde 61% al 100% por lo que la mayoría de los MAGs se encuentran incompletos. De hecho, la re-construcción de genomas completos de organismos a partir de metagenomas es un desafío ya que los genomas recuperados suelen estar muy fragmentados debido a la abundancia desigual de organismos, las repeticiones dentro y entre los genomas, los errores de secuenciación y la variación a nivel de cepa (Muralidharan, et al., 2021 y Yue et al., 2020). Algunas herramientas como MetaCarvel utilizan scaffolds para generar el ensamblaje, y con esto mejorar la contigüidad y la calidad de los MAGs resultantes y con esto obtener una mayor cantidad de genes de los organismos que se están reconstruyendo (Muralidharan, et al., 2021). Por otro lado, la estrategia para obtener estos genomas se basó en la presencia de *single-copy genes* (SCG, única copia en el 95% de genomas), la precisión de las predicciones puede ser limitada cuando este enfoque se aplica a genomas que pertenecen a clados poco estudiados, por lo que los SCG no pueden garantizar la ausencia de contaminación o el nivel de completitud real (Shaiber et al., 2020). Es importante considerar para futuros estudios, otras estrategias de binning, y revisar si los genes de degradación de atrazina se encuentran en algún genoma particular.

Cuadro 19. Genes que participan en la degradación de atrazina encontrados en las fuentes termales.

Número EC	ID	Gen	Organismo	Uniprot ID
3.5.4.43	Hidroxicloroatrazina etilaminohidrolasa	atzB	<i>Pseudomonas sp.</i> ADP	P95442
3.5.2.15	Ácido cianúrico amidohidrolasa	atzD	<i>Pseudomonas sp.</i> ADP	P58329
3.5.1.131	1-carboxibiuret hidrolasa	atzE	<i>Pseudomonas sp.</i> ADP	Q936X3
3.5.1.54	Alofanato hidrolasa	atzF	<i>Pseudomonas sp.</i> ADP	Q936X2
3.5.4.45	Desaminasa de melamina	triA	<i>Acidovorax citrulli</i> (<i>Acidovorax avenae</i> subsp. <i>citrulli</i>)	Q9EYU0

La degradación de atrazina puede involucrar dos vías diferentes: ya sea por desalquilación oxidativa que conduce a la formación intermedia de desisopropilatrazina o por hidrólisis para formar el producto deshalogenado intermedio hidroxatrazina (Kundu et al., 2019). Se ha demostrado que el metabolismo del herbicida atrazina está ligado a plásmidos catabólicos (Martinez-Vaz et al., 2001). Los tres primeros pasos enzimáticos, codificados por los genes *atzA*, *atzB* y *atzC*, transforman atrazina a ácido cianúrico. El ácido cianúrico es un compuesto intermediario en las vías de degradación de la melamina (2,4,6-triamino-s-triazina) y la atrazina en diferentes especies de *Pseudomonas* (Martinez-Vaz et al., 2001). Las poblaciones microbianas expuestas a compuestos clorados sintéticos, como la atrazina, generalmente producen enzimas que degradan estas moléculas, es interesante mencionar que se encontraron estos genes en muestras de ambientes prístinos lo que puede indicar algún grado de contaminación o alteración antropogénica.

Análisis de correlación entre los genes relacionados con la remoción de agroquímicos y los principales filos encontrados en los tapetes microbianos

La mayoría de los genes de remoción y/o degradación de paraquat y atrazina muestra una correlación positiva con microorganismos dentro de los filos Verrucomicrobia, Proteobacteria, Firmicutes, Cyanobacteria, Chlorobi y Acidobacteria principalmente (Fig. 24). Esto puede estar relacionado a que estos genes se encuentran en microorganismos dentro de estos grupos, específicamente en el caso de las Proteobacterias y Firmicutes, los

cuales son muy abundantes en todas las fuentes termales, excepto en Las Lilas y Río Naranjo. Para la elaboración de la figura se utilizaron los datos de abundancia relativa de los diferentes filos y la abundancia de cada uno de los genes por muestra, y con esto se realizó el análisis de correlación de Spearman.

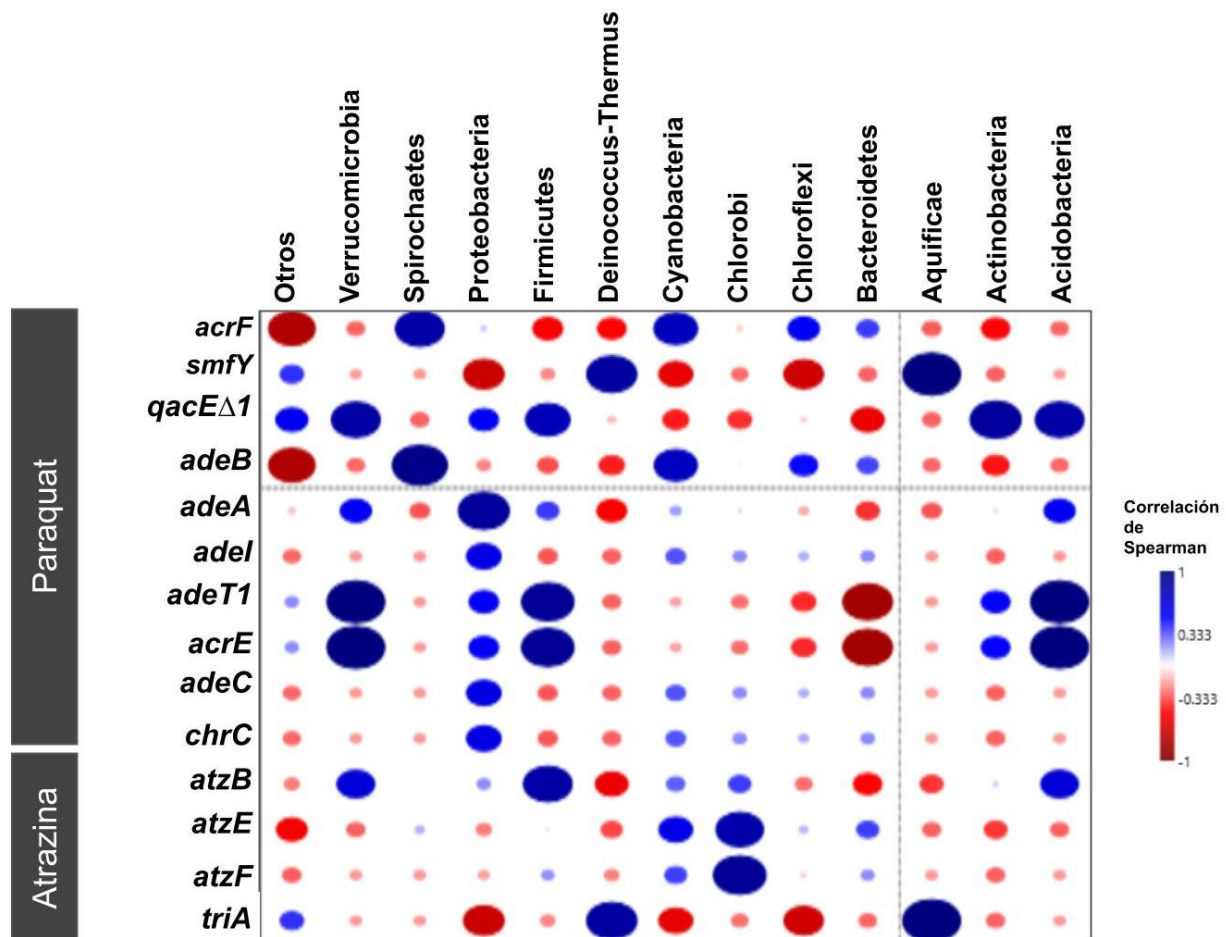


Figura 24. Resultados de correlación de Spearman entre los genes relacionados con la remoción/degradación de paraquat y atrazina, y los diferentes filos encontrados en los tapetes microbianos. La correlación positiva corresponde al color azul y la negativa al color rojo, el tamaño de los círculos es proporcional al valor absoluto de la correlación. En Otros se incluyen los grupos de Arqueas y secuencias clasificadas como Bacteria.

Como se mencionó anteriormente se encontró una correlación positiva de los genes que participan en la remoción/degradación de paraquat con el filo Proteobacteria, donde cabe mencionar que muchos reportes de cepas aisladas principalmente de suelo, y que han

demostrado su potencial de degradación del compuesto, tales como *Pseudomonas putida*, *Aerobacter aerogenes*, *Agrobacterium tumefaciens*, *Escherichia coli* y Firmicutes como *Bacillus spp* (Huang et al., 2019), pertenecen a este filo. Algunos de estos géneros se han encontrado en los tapetes microbianos de estas fuentes termales, específicamente *Bacillus* sp. En el caso específico de las cianobacterias, en donde se encontró una correlación positiva de este filo con la presencia de algunos genes de degradación de estos agroquímicos, estudios previos informan sobre la tolerancia y capacidad de remoción del paraquat de cianobacterias como *Nostoc punctiforme* ATCC 29133, *Synechocystis* sp. PCC6803 y *Leptolyngbya* sp. 7M (Forlani et al., 2008, Prosecka et al., 2009, Brenes-Guillén et al., 2019), los cuales se han encontrado en estos tapetes microbianos y en alta abundancias, específicamente en el caso de *Leptolyngbya* spp. Estos resultados sugieren la presencia de diversos mecanismos celulares que les confieren a estos microorganismos esta capacidad. Es importante mencionar que ninguno de los genes de degradación de atrazina encontrados en los metagenomas se localizó en los MAGs del filo Cyanobacteria, esto a pesar de que se encontró una correlación de estos con el filo. Esto se puede deber a que en los tapetes microbianos se encuentran cianobacterias de las cuáles no se obtuvo un MAG, sin embargo, estos microorganismos son muy abundantes en estas muestras según datos de secuenciación masiva del amplicón 16S ARNr, además se encontraron secuencias de este mismo gen en los metagenomas de *Gloeocapsa*, *Leptolyngbya*, *Fischerella*, *Synechococcus*, *Geminocystis*, entre otros, cianobacterias en los cuales se podrían encontrar estos genes, pero bajo la metodología utilizada no se pudieron asociar a algún género en específico. Lo mismo para el filo Chlorobi, en el que se encontró una correlación positiva sin embargo no obtuvieron MAGs dentro de este filo. Kyrgyzov y colaboradores (2020) señalan que el ensamblaje de metagenomas actualmente no es útil para recuperar genomas de baja abundancia, como lo es este filo en las muestras de tapetes microbianos costarricenses, por lo que plantean una estrategia de agrupamiento de secuencias antes del ensamblaje lo que permitiría la recuperación de genomas de microorganismos con baja abundancia (< 1%). Esta estrategia puede utilizarse en futuras investigaciones para lograr obtener los genomas de estos microorganismos en las muestras de tapetes microbianos. Deinococcus-Thermus es uno de los filos más abundantes en Rocas Calientes, caracterizada por ser una salida de agua con una temperatura cercana a los 63 °C y con ninguna alteración antropogénica visible. En el análisis de correlación de Spearman se encontró una relación negativa (desde -0.25 a -1) de los genes relacionados con la remoción y/o degradación de atrazina y paraquat, y este filo. Una relación similar se puede observar

en los filos Chloroflexi y Aquificae, los cuales son muy abundantes en Rocas Calientes y Las Lilas respectivamente. Es de gran relevancia mencionar que los tapetes microbianos de Salitral, La Luz y Río Naranjo se encuentran a orilla de ríos, por lo que tienen cierto grado de alteración antropogénica, y es posible que estén en contacto con xenobióticos que tienen las aguas del río aledaño, por otro lado los genes de remoción/degradación se encuentran en estos ambientes “prístinos”, ya que los principales transportadores de estos compuestos tienen funciones específicas en la célula expulsando otro tipo de sustancias. En algunas cepas mutantes de *Deinococcus radiodurans* se ha demostrado que tienen una alta sensibilidad al paraquat (Slade y Radman, 2011). Resulta de gran interés realizar una búsqueda más detallada de los genes antes mencionados en este grupo de microorganismos y realizar ensayos de tolerancia y/o remoción de estos compuestos.

Conclusiones

Al comparar a nivel taxonómico y funcional seis metagenomas de tapetes microbianos de fuentes termales de Costa Rica y los 13 metagenomas de Chile, Nueva Zelanda y P.N. Yellowstone, se encontró que las comunidades microbianas a nivel de filo de las fuentes termales costarricenses son similares a otras alrededor del mundo, sin embargo tienen algunos microorganismos únicos, probablemente producto de las condiciones ambientales específicas. La abundancia relativa de los filos presentes en metagenomas depende de la estrategia de análisis y bases de datos que utilicen las herramientas. En la mayoría de las muestras se encontraron *contigs* más largos y N50 mayor utilizando el ensamblador MEGAHIT. Es relevante utilizar datos crudos (archivos tipo SRA) si se requiere realizar comparaciones más detalladas y bajo los mismos parámetros de ensamblaje entre las diversas comunidades.

Se obtuvieron MAGs con un valor de completitud mayor al 60% y de redundancia menor al 10%, por lo que se consideran de media-alta calidad. Este estudio describe por primera vez MAGs de tapetes microbianos de fuentes termales costarricenses, la mayoría de los cuales se encontraron en el filo Chloroflexi, Cyanobacteria y Proteobacteria, y pertenecen a organismos que no se han cultivado, esto ofrece una oportunidad para realizar intentos de cultivos en medios específicos para la caracterización de estos microorganismos en estudios polifásicos. Diferentes géneros y especies de bacterias se encuentran agrupadas dentro del mismo filo, por lo que la estructura de la comunidad y las inferencias biológicas relacionadas van a cambiar si se disminuye el nivel taxonómico con el que se trabaja, esto es difícil de obtener si se trabaja con datos de metagenómica *shotgun*, por lo que es necesario implementar otras estrategias como el análisis del gen 16S ARNr y análisis filogenómicos. Es el primer análisis en donde se reporta y se encuentran MAGs pertenecientes a Crenarchaeota y Euryarchaeota en sitios de pH cercano a 7 y altas temperaturas.

El 75% de las secuencias encontradas en este estudio no tienen COG asignado, mientras que el 50% de las secuencias se encontraron dentro de la categoría de putativas, según Pfam, por lo que podrían ser proteínas novedosas, que han sido muy poco estudiadas, lo que ofrece una oportunidad muy amplia de investigación de estas secuencias y su posterior estudio en ensayos de laboratorio. Las peptidasas son el grupo de enzimas hidrolíticas más abundantes en las muestras de fuentes termales de todos los sitios estudiados, representan más del 50% del total, seguido por las hidrolasas y las esterases (~ 25 %). Algunas enzimas

dentro de estas categorías han sido utilizadas en la biorremediación de agroquímicos y otros xenobióticos, lo que ofrece una amplia base para estudios en este y otras áreas.

La mayoría de las variantes encontradas en tapetes microbianos de fuentes termales de Costa Rica son similares (> 90%) a las de microorganismos que toleran altas temperaturas, por lo que estas enzimas muy probablemente sean termotolerantes. Muchas de ellas podrían participar en procesos muy importantes como lo son la fotosíntesis, la degradación de compuestos específicos para el metabolismo y la producción de antibióticos. Algunas de las variantes se encontraron en todas las muestras, por lo que posiblemente pueden tolerar temperaturas que van desde los 34 °C a los 73 °C, así como diferente concentración de iones y pH, no obstante, algunas tienen una abundancia mayor en un rango de temperatura determinado.

En los MAGs obtenidos se encontraron genes que participan en el transporte, acumulación y oxidación-reducción metales y herbicidas, principalmente en los filos Cyanobacteria, Proteobacteria y Chloroflexi. Se encontró similitud entre el 70 % y el 90 % en las secuencias de estos genes con respecto a los de referencia que son en su mayoría de microorganismos no termales, lo que podría traducirse en diferencias fenotípicas entre las diferentes cepas y la respuesta a las diferentes condiciones ambientales. Esto es de gran relevancia, pues muestra que en fuentes termales son fuente de alelos nuevos que podrían ser objeto de estudio en futuras investigaciones.

Las comunidades bacterianas de ambientes termales participan en el ciclo biogeoquímico de diversos metales, lo que aporta significativamente a los estudios en geomicrobiología que se llevan a cabo en nuestro país. Futuras investigaciones y ensayos en laboratorio ayudarían comprender de una forma más integral el rol de estos microorganismos en estos ecosistemas.

Recomendaciones

- Realizar el mapeo de secuencias en metagenomas de otras fuentes termales para comparar los resultados obtenidos con las muestras costarricenses, esto con el fin de determinar variantes únicas en fuentes termales de zonas tropicales.
- Bajo la tendencia del análisis ómico integrativo, se deben combinar los resultados de diferentes ensambladores metagenómicos y herramientas para *binning*, esto con el fin de comprender de manera integral la composición y función de las comunidades microbianas. Además, se podrían utilizar la secuenciación de tercera generación como Pacific Biosciences (PacBio) y Oxford Nanopore Technologies (ONT) para la complementar los datos obtenidos con la plataforma Illumina.
- El aislamiento e identificación de los diferentes grupos de enzimas permitirá estudios más detallados de la función que tienen en los microorganismos y por ende en las comunidades microbianas. Los ensayos funcionales en el laboratorio son una herramienta muy importante que debe utilizarse de forma complementaria a los análisis metagenómicos, con el fin de esclarecer las funciones de las proteínas encontradas. Se debe verificar experimentalmente la presencia y función de los genes y reguladores sensibles a los diferentes metales.
- Los análisis taxonómicos y funcionales se deben realizar utilizando diferentes programas y bases de datos, esto con el fin de comparar y complementar los resultados obtenidos.
- Es de gran importancia continuar implementando en este tipo de análisis los principios FAIR (*Findable, Accesible, Interoperable y Reusable*) con el fin de tener una línea base en la administración de datos biológicos.

Referencias

Aanniz, T., Ouadghiri, M., Melloul, M., Swings, J., Elfahime, E., Ibijbijen, J., Ismaili, M., & Amar, M. (2015). Thermophilic bacteria in Moroccan hot springs, salt marshes and desert soils. In *Brazilian Journal of Microbiology* (Vol. 46, Issue 2, pp. 443–453). FapUNIFESP (SciELO). <https://doi.org/10.1590/s1517-838246220140219>

Adam, Z., Zaltsman, A., Sinvany-Villalobo, G., & Sakamoto, W. (2005). FtsH proteases in chloroplasts and cyanobacteria. In *Physiologia Plantarum* (Vol. 123, Issue 4, pp. 386–390). Wiley. <https://doi.org/10.1111/j.1399-3054.2004.00436.x>

Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., & Tawfik, D. S. (2005). High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. In *Chemistry & Biology* (Vol. 12, Issue 12, pp. 1281–1289). Elsevier BV. <https://doi.org/10.1016/j.chembiol.2005.09.012>

Aizouq, M., Peisker, H., Gutbrod, K., Melzer, M., Hölzl, G., & Dörmann, P. (2020). Triacylglycerol and phytol ester synthesis in *Synechocystis* sp. PCC6803. In *Proceedings of the National Academy of Sciences* (Vol. 117, Issue 11, pp. 6216–6222). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1915930117>

Allali, N., Afif, H., Couturier, M., & Van Melderren, L. (2002). The Highly Conserved TldD and TldE Proteins of *Escherichia coli* Are Involved in Microcin B17 Processing and in CcdA Degradation. In *Journal of Bacteriology* (Vol. 184, Issue 12, pp. 3224–3231). American Society for Microbiology. <https://doi.org/10.1128/jb.184.12.3224-3231.2002>

Alav, I., Sutton, J. M., & Rahman, K. M. (2018). Role of bacterial efflux pumps in biofilm formation. In *Journal of Antimicrobial Chemotherapy* (Vol. 73, Issue 8, pp. 2003–2020). Oxford University Press (OUP). <https://doi.org/10.1093/jac/dky042>

Alcamán-Arias, M. E., Pedrós-Alió, C., Tamames, J., Fernández, C., Pérez-Pantoja, D., Vásquez, M., & Díez, B. (2018). Diurnal Changes in Active Carbon and Nitrogen Pathways Along the Temperature Gradient in Porcelana Hot Spring Microbial Mat. In *Frontiers in Microbiology* (Vol. 9). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2018.02353>

Alvarado, G. E., & Vargas, A. (2017). Historia del descubrimiento y aprovechamiento de las fuentes termales en Costa Rica. In *Revista Geológica de América Central* (Issue 57). Universidad de Costa Rica. <https://doi.org/10.15517/rgac.v0i57.30148>

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. In *Journal of Molecular Biology* (Vol. 215, Issue 3, pp. 403–410). Elsevier BV. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)

Amin, A., Ahmed, I., Salam, N., Kim, B.-Y., Singh, D., Zhi, X.-Y., Xiao, M., & Li, W.-J. (2017). Diversity and Distribution of Thermophilic Bacteria in Hot Springs of Pakistan. In *Microbial Ecology* (Vol. 74, Issue 1, pp. 116–127). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00248-017-0930-1>

Andrews, S. C., Robinson, A. K., & Rodríguez-Quñones, F. (2003). Bacterial iron homeostasis. In *FEMS Microbiology Reviews* (Vol. 27, Issues 2–3, pp. 215–237). Oxford University Press (OUP). [https://doi.org/10.1016/s0168-6445\(03\)00055-x](https://doi.org/10.1016/s0168-6445(03)00055-x)

Ankenbauer, R. G., & Quan, H. N. (1994). FptA, the Fe (III)-pyochelin receptor of *Pseudomonas aeruginosa*: a phenolate siderophore receptor homologous to hydroxamate siderophore receptors. In *Journal of Bacteriology* (Vol. 176, Issue 2, pp. 307–319). American Society for Microbiology. <https://doi.org/10.1128/jb.176.2.307-319.1994>

Antranikian, G., Herzberg, C., & Gottschalk, G. (1987). Production of Thermostable α -Amylase, Pullulanase, and α -Glucosidase in Continuous Culture by a New *Clostridium* Isolate. In *Applied and Environmental Microbiology* (Vol. 53, Issue 7, pp. 1668–1673). American Society for Microbiology. <https://doi.org/10.1128/aem.53.7.1668-1673.1987>

Antranikian, G. (1992). "Microbial degradation of starch," in *Microbial Degradation of Natural Products*, ed A. Winkelmann (Germany: Weinheim VCH), 27–56.

Appia-Ayme, C., Patrick, E., J. Sullivan, M., Alston, M. J., Field, S. J., AbuOun, M., Anjum, M. F., & Rowley, G. (2011). Novel Inducers of the Envelope Stress Response BaeSR in *Salmonella Typhimurium*: BaeR Is Critically Required for Tungstate Waste Disposal. In M. A. Webber (Ed.), *PLoS ONE* (Vol. 6, Issue 8, p. e23713). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0023713>

Arce-Rodríguez, A., Puente-Sánchez, F., Avendaño, R., Martínez-Cruz, M., de Moor, J. M., Pieper, D. H., & Chavarría, M. (2019). Thermoplasmatales and sulfur-oxidizing bacteria dominate the microbial community at the surface water of a CO₂-rich hydrothermal spring located in Tenorio Volcano National Park, Costa Rica. In *Extremophiles* (Vol. 23, Issue 2, pp. 177–187). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-018-01072-6>

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., ... Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. In *Nature Biotechnology* (Vol. 36, Issue 7, pp. 566–569). Springer Science and Business Media LLC. <https://doi.org/10.1038/nbt.4163>

Aschauer, P., Zimmermann, R., Breinbauer, R., Pavkov-Keller, T., & Oberer, M. (2018). The crystal structure of monoacylglycerol lipase from *M. tuberculosis* reveals the basis for specific inhibition. In *Scientific Reports* (Vol. 8, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-018-27051-7>

Auguet, J.-C., Barberan, A., & Casamayor, E. O. (2009). Global ecological patterns in uncultured Archaea. In *The ISME Journal* (Vol. 4, Issue 2, pp. 182–190). Springer Science and Business Media LLC. <https://doi.org/10.1038/ismej.2009.109>

Ayling, M., Clark, M. D., & Leggett, R. M. (2019). New approaches for metagenome assembly with short reads. In *Briefings in Bioinformatics* (Vol. 21, Issue 2, pp. 584–594). Oxford University Press (OUP). <https://doi.org/10.1093/bib/bbz020>

Badilla, C., Osborne, T. H., Cole, A., Watson, C., Djordjevic, S., & Santini, J. M. (2018). A new family of periplasmic-binding proteins that sense arsenic oxyanions. In *Scientific Reports* (Vol. 8, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-018-24591-w>

Ballschmiter, M., Fütterer, O., & Liebl, W. (2006). Identification and Characterization of a Novel Intracellular Alkaline α -Amylase from the Hyperthermophilic Bacterium *Thermotoga maritima* MSB8. In *Applied and Environmental Microbiology* (Vol. 72, Issue 3, pp. 2206–2211). American Society for Microbiology. <https://doi.org/10.1128/aem.72.3.2206-2211.2006>

Baltaci, M. O., Genc, B., Arslan, S., Adiguzel, G., & Adiguzel, A. (2016). Isolation and Characterization of Thermophilic Bacteria from Geothermal Areas in Turkey and Preliminary Research on Biotechnologically Important Enzyme Production. In *Geomicrobiology Journal* (Vol. 34, Issue 1, pp. 53–62). Informa UK Limited. <https://doi.org/10.1080/01490451.2015.1137662>

Baker, M. (2012). De novo genome assembly: what every biologist should know. In *Nature Methods* (Vol. 9, Issue 4, pp. 333–337). Springer Science and Business Media LLC. <https://doi.org/10.1038/nmeth.1935>

Bao, Y.-J., Xu, Z., Li, Y., Yao, Z., Sun, J., & Song, H. (2017). High-throughput metagenomic analysis of petroleum-contaminated soil microbiome reveals the versatility in xenobiotic aromatics metabolism. In *Journal of Environmental Sciences* (Vol. 56, pp. 25–35). Elsevier BV. <https://doi.org/10.1016/j.jes.2016.08.022>

Bargar, K. (1978). *Geology and Thermal History of Mammoth Hot Springs, Yellowstone National Park, Wyoming*. US Geological Survey. <https://doi.org/10.3133/b1444>

Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. In *Proceedings of the National Academy of Sciences* (Vol. 93, Issue 17, pp. 9188–9193). Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.93.17.9188>

Batzoglou, S., Jaffe, D. B., Stanley, K., Butler, J., Gnerre, S., Mauceli, E., Berger, B., Mesirov, J. P., & Lander, E. S. (2002). ARACHNE: A Whole-Genome Shotgun Assembler. In *Genome Research* (Vol. 12, Issue 1, pp. 177–189). Cold Spring Harbor Laboratory. <https://doi.org/10.1101/gr.208902>

Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J., & Kelly, R. M. (1999). An endoglucanase, EglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1-->3),(1-->4)-beta-D-glucans and cellulose. *Journal of bacteriology*, 181(1), 284–290. <https://doi.org/10.1128/JB.181.1.284-290.1999>

Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: A multidrug transporter family that continues to grow. In *Biochimica et Biophysica Acta (BBA) - Biomembranes* (Vol. 1778, Issue 9, pp. 1814–1838). Elsevier BV. <https://doi.org/10.1016/j.bbamem.2007.08.015>

Ben Fekih, I., Zhang, C., Li, Y. P., Zhao, Y., Alwathnani, H. A., Saquib, Q., Rensing, C., & Cervantes, C. (2018). Distribution of Arsenic Resistance Genes in Prokaryotes. In *Frontiers in Microbiology* (Vol. 9). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2018.02473>

Bencko, V., & Yan Li Foong, F. (2017). The history of arsenical pesticides and health risks related to the use of Agent Blue. In *Annals of Agricultural and Environmental Medicine* (Vol. 24, Issue 2, pp. 312–316). Institute of Rural Health. <https://doi.org/10.26444/aaem/74715>

Beris, F. S., De Smet, L., Karaoglu, H., Canakci, S., Van Beeumen, J., & Belduz, A. O. (2011). The ATPase activity of the G2alt gene encoding an aluminium tolerance protein from *Anoxybacillus gonensis* G2. In *The Journal of Microbiology* (Vol. 49, Issue 4, pp. 641–650). Springer Science and Business Media LLC. <https://doi.org/10.1007/s12275-011-0522-6>

Bernardes, J., & Pedreira, C. (2013). A Review of Protein Function Prediction Under Machine Learning Perspective. In *Recent Patents on Biotechnology* (Vol. 7, Issue 2, pp. 122–141). Bentham Science Publishers Ltd. <https://doi.org/10.2174/18722083113079990006>

Bertin, C., Weston, L. A., Huang, T., Jander, G., Owens, T., Meinwald, J., & Schroeder, F. C. (2007). Grass roots chemistry: meta-Tyrosine, an herbicidal nonprotein amino acid. In *Proceedings of the National Academy of Sciences* (Vol. 104, Issue 43, pp. 16964–16969). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0707198104>

Bharathi Srinivasan, V., Rajamohan, G., Pancholi, P., Marcon, M., & Gebreyes, W. A. (2011). Molecular cloning and functional characterization of two novel membrane fusion proteins in conferring antimicrobial resistance in *Acinetobacter baumannii*. In *Journal of Antimicrobial Chemotherapy* (Vol. 66, Issue 3, pp. 499–504). Oxford University Press (OUP). <https://doi.org/10.1093/jac/dkq469>

Bhalla, A., Bansal, N., Kumar, S., Bischoff, K. M., & Sani, R. K. (2013). Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. In *Bioresource Technology* (Vol. 128, pp. 751–759). Elsevier BV. <https://doi.org/10.1016/j.biortech.2012.10.145>

Bhandari, S., Poudel, D. K., Marahatha, R., Dawadi, S., Khadayat, K., Phuyal, S., Shrestha, S., Gaire, S., Basnet, K., Khadka, U., & Parajuli, N. (2021). Microbial Enzymes Used in Bioremediation. In C. G. Silva (Ed.), *Journal of Chemistry* (Vol. 2021, pp. 1–17). Hindawi Limited. <https://doi.org/10.1155/2021/8849512>

Bhatt, P., Bhatt, K., Huang, Y., Lin, Z., & Chen, S. (2020). Esterase is a powerful tool for the biodegradation of pyrethroid insecticides. In *Chemosphere* (Vol. 244, p. 125507). Elsevier BV. <https://doi.org/10.1016/j.chemosphere.2019.125507>

Bhatt, P., Zhou, X., Huang, Y., Zhang, W., & Chen, S. (2021). Characterization of the role of esterases in the biodegradation of organophosphate, carbamate, and pyrethroid pesticides. In *Journal of Hazardous Materials* (Vol. 411, p. 125026). Elsevier BV. <https://doi.org/10.1016/j.jhazmat.2020.125026>

Bhushan, B., & Hoondal, G. S. (1998). In *Biotechnology Letters* (Vol. 20, Issue 2, pp. 157–159). Springer Science and Business Media LLC. <https://doi.org/10.1023/a:1005328508227>

Blodgett, J. A. V., Thomas, P. M., Li, G., Velasquez, J. E., van der Donk, W. A., Kelleher, N. L., & Metcalf, W. W. (2007). Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. In *Nature Chemical Biology* (Vol. 3, Issue 8, pp. 480–485). Springer Science and Business Media LLC. <https://doi.org/10.1038/nchembio.2007.9>

Borremans, B., Hobman, J. L., Provoost, A., Brown, N. L., & van der Lelie, D. (2001). Cloning and Functional Analysis of the pbr Lead Resistance Determinant of *Ralstonia metallidurans* CH34. In *Journal of Bacteriology* (Vol. 183, Issue 19, pp. 5651–5658). American Society for Microbiology. <https://doi.org/10.1128/jb.183.19.5651-5658.2001>

Boteva, N., & Kambourova, M. (2018). Thermophiles and Their Exploration for Thermostable Enzyme Production. In *Extremophiles in Eurasian Ecosystems: Ecology, Diversity, and Applications* (pp. 167–186). Springer Singapore. https://doi.org/10.1007/978-981-13-0329-6_6

Branco, R., Chung, A. P., Johnston, T., Gurel, V., Morais, P., & Zhitkovich, A. (2008). The Chromate-Inducible chrBACF Operon from the Transposable Element Tn OtChr Confers Resistance to Chromium(VI) and Superoxide. In *Journal of Bacteriology* (Vol. 190, Issue 21, pp. 6996–7003). American Society for Microbiology. <https://doi.org/10.1128/jb.00289-08>

Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. In *Briefings in Bioinformatics* (Vol. 20, Issue 4, pp. 1125–1136). Oxford University Press (OUP). <https://doi.org/10.1093/bib/bbx120>

Brenes-Guillén, L., Fuentes-Schweizer, P., García-Piñeres, A., & Uribe-Lorío, L. (2019). Tolerance and sorption of Bromacil and Paraquat by thermophilic cyanobacteria *Leptolyngbya* 7M from Costa Rican thermal springs. In *Journal of Contaminant Hydrology* (Vol. 226, p. 103539). Elsevier BV. <https://doi.org/10.1016/j.jconhyd.2019.103539>

Brenes-Guillén, L., Vidaurre-Barahona, D., Morales, S., Mora-López, M., Sittenfeld, A., & Uribe-Lorío, L. (2021). Novel Cyanobacterial Diversity Found in Costa Rican Thermal Springs Associated with Rincon de la Vieja and Miravalles Volcanoes: A Polyphasic Approach. In J. Collier (Ed.), *Journal of Phycology* (Vol. 57, Issue 1, pp. 183–198). Wiley. <https://doi.org/10.1111/jpy.13077>

Brickman, T. J., Cummings, C. A., Liew, S.-Y., Relman, D. A., & Armstrong, S. K. (2011). Transcriptional Profiling of the Iron Starvation Response in *Bordetella pertussis* Provides New Insights into Siderophore Utilization and Virulence Gene Expression. In *Journal of Bacteriology* (Vol. 193, Issue 18, pp. 4798–4812). American Society for Microbiology. <https://doi.org/10.1128/jb.05136-11>

Kay, D. (1980). Thermophilic microorganisms and life at high temperatures. T. D. Brock, Springer-Verlag, 1978. No. of pages: 465. In *Earth Surface Processes* (Vol. 5, Issue 1, pp. 101–102). Wiley. <https://doi.org/10.1002/esp.3760050112>

Brock, T. D. (2001). The Origins of Research on Thermophiles. In *Thermophiles Biodiversity, Ecology, and Evolution* (pp. 1–9). Springer US. https://doi.org/10.1007/978-1-4615-1197-7_1

Brock, T.D, & Mosser, J. L. (1975). Rate of Sulfuric-Acid Production in Yellowstone National Park. In *Geological Society of America Bulletin* (Vol. 86, Issue 2, p. 194). Geological Society of America. [https://doi.org/10.1130/0016-7606\(1975\)86<194:rospiy>2.0.co;2](https://doi.org/10.1130/0016-7606(1975)86<194:rospiy>2.0.co;2)

Bruntz, R. C., Lindsley, C. W., & Brown, H. A. (2014). Phospholipase D Signaling Pathways and Phosphatidic Acid as Therapeutic Targets in Cancer. In D. R. Sibley (Ed.), *Pharmacological Reviews* (Vol. 66, Issue 4, pp. 1033–1079). American Society for Pharmacology & Experimental Therapeutics (ASPET). <https://doi.org/10.1124/pr.114.009217>

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. In *Nature Methods* (Vol. 12, Issue 1, pp. 59–60). Springer Science and Business Media LLC. <https://doi.org/10.1038/nmeth.3176>

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. In *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* (Vol. 1842, Issue 10, pp. 1932–1941). Elsevier BV. <https://doi.org/10.1016/j.bbadis.2014.06.015>

Burggraf, S., Heyder, P., & Eis, N. (1997). A pivotal Archaea group. In *Nature* (Vol. 385, Issue 6619, pp. 780–780). Springer Science and Business Media LLC. <https://doi.org/10.1038/385780a0>

Burrows, W. (1973) *Text Book of Microbiology*. W.B. Saunders Company, Ontario

Caldwell, S. L., Liu, Y., Ferrera, I., Beveridge, T., & Reysenbach, A.-L. (2010). *Thermocrinis minervae* sp. nov., a hydrogen- and sulfur-oxidizing, thermophilic member of the Aquificales from a Costa Rican terrestrial hot spring. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 60, Issue 2, pp. 338–343). Microbiology Society. <https://doi.org/10.1099/ijs.0.010496-0>

Caldwell, B. A. (2005). Enzyme activities as a component of soil biodiversity: A review. In *Pedobiologia* (Vol. 49, Issue 6, pp. 637–644). Elsevier BV. <https://doi.org/10.1016/j.pedobi.2005.06.003>

Campbell, J. H., O'Donoghue, P., Campbell, A. G., Schwientek, P., Sczyrba, A., Woyke, T., Soll, D., & Podar, M. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. In *Proceedings of the National Academy of Sciences* (Vol. 110, Issue 14, pp. 5540–5545). Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.1303090110>

Campbell, K. M., Kouris, A., England, W., Anderson, R. E., McCleskey, R. B., Nordstrom, D. K., & Whitaker, R. J. (2017). *Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. In *Environmental Microbiology* (Vol. 19, Issue 6, pp. 2334–2347). Wiley. <https://doi.org/10.1111/1462-2920.13728>

Canaan, S., Maurin, D., Chahinian, H., Pouilly, B., Dourousseau, C., Frassinetti, F., Scappuccini-Calvo, L., Cambillau, C., & Bourne, Y. (2004). Expression and characterization of the protein Rv1399c from *Mycobacterium tuberculosis*. In *European Journal of Biochemistry* (Vol. 271, Issue 19, pp. 3953–3961). Wiley. <https://doi.org/10.1111/j.1432-1033.2004.04335.x>

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. In *Nucleic Acids Research* (Vol. 37, Issue Database, pp. D233–D238). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkn663>

Castro, E., et al. (2019). Búsqueda de genes de interés en datos de metagenómica shotgun. Disponible en: http://www.castrolab.org/isme/gene_search/gene_search.html

Cerdà-Costa, N., & Xavier Gomis-Rüth, F. (2013). Architecture and function of metallopeptidase catalytic domains. In *Protein Science* (Vol. 23, Issue 2, pp. 123–144). Wiley. <https://doi.org/10.1002/pro.2400>

Chahinian, H., Nini, L., Boitard, E., Dubès, J.-P., Comeau, L.-C., & Sarda, L. (2002). Distinction between esterases and lipases: A kinetic study with vinyl esters and TAG. In *Lipids* (Vol. 37, Issue 7, pp. 653–662). Wiley. <https://doi.org/10.1007/s11745-002-0946-7>

Chandrangsu, P., Rensing, C., & Helmann, J. D. (2017). Metal homeostasis and resistance in bacteria. In *Nature Reviews Microbiology* (Vol. 15, Issue 6, pp. 338–350). Springer Science and Business Media LLC. <https://doi.org/10.1038/nrmicro.2017.15>

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. In J. Hancock (Ed.), *Bioinformatics*. Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btz848>

Chen, J. C., Viollier, P. H., & Shapiro, L. (2004). A membrane metalloprotease participates in the sequential degradation of a *Caulobacter* polarity determinant. In *Molecular Microbiology* (Vol. 55, Issue 4, pp. 1085–1103). Wiley. <https://doi.org/10.1111/j.1365-2958.2004.04443.x>

Chen, X.-Lan., Xie, B.-Bin., Lu, J.-Tao., He, H.-Lun., & Zhang, Yuzhong. (2007). A novel type of subtilase from the psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: catalytic and structural properties of deseasin MCP-01. In *Microbiology* (Vol. 153, Issue 7, pp. 2116–2125). Microbiology Society. <https://doi.org/10.1099/mic.0.2007/006056-0>

Chen, W., Xie, T., Shao, Y., & Chen, F. (2012). Phylogenomic Relationships between Amylolytic Enzymes from 85 Strains of Fungi. In M. Freitag (Ed.), PLoS ONE (Vol. 7, Issue 11, p. e49679). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0049679>

Clarke, K.R. & Gorley, R.N. (2015) PRIMER v7: User Manual/Tutorial. PRIMER-E Plymouth.

Mahmoudi, H., Shokoohizadeh, L., Zare Fahim, N., Mohamadi Bardebari, A., Moradkhani, S., & Alikhani, M. Y. (2020). Detection of *adeABC* efflux pump encoding genes and antimicrobial effect of *Mentha longifolia* and Menthol on MICs of imipenem and ciprofloxacin in clinical isolates of *Acinetobacter baumannii*. In BMC Complementary Medicine and Therapies (Vol. 20, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12906-020-02887-7>

Connon, S. A., Koski, A. K., Neal, A. L., Wood, S. A., & Magnuson, T. S. (2008). Ecophysiology and geochemistry of microbial arsenic oxidation within a high arsenic, circumneutral hot spring system of the Alvord Desert. In FEMS Microbiology Ecology (Vol. 64, Issue 1, pp. 117–128). Oxford University Press (OUP). <https://doi.org/10.1111/j.1574-6941.2008.00456.x>

Cowan, D., Ramond, J.-B., Makhalanyane, T., & De Maayer, P. (2015). Metagenomics of extreme environments. In Current Opinion in Microbiology (Vol. 25, pp. 97–102). Elsevier BV. <https://doi.org/10.1016/j.mib.2015.05.005>

Crognale, S., Zecchin, S., Amalfitano, S., Fazi, S., Casentini, B., Corsini, A., Cavalca, L., & Rossetti, S. (2017). Phylogenetic Structure and Metabolic Properties of Microbial Communities in Arsenic-Rich Waters of Geothermal Origin. In Frontiers in Microbiology (Vol. 8). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2017.02468>

Damier-Piolle, L., Magnet, S., Brémont, S., Lambert, T., & Courvalin, P. (2008). AdelJK, a Resistance-Nodulation-Cell Division Pump Effluxing Multiple Antibiotics in *Acinetobacter baumannii*. In *Antimicrobial Agents and Chemotherapy* (Vol. 52, Issue 2, pp. 557–562). American Society for Microbiology. <https://doi.org/10.1128/aac.00732-07>

Dave, K. I., Miller, C. E., & Wild, J. R. (1993). Characterization of organophosphorus hydrolases and the genetic manipulation of the phosphotriesterase from *Pseudomonas diminuta*. In *Chemico-Biological Interactions* (Vol. 87, Issues 1–3, pp. 55–68). Elsevier BV. [https://doi.org/10.1016/0009-2797\(93\)90025-t](https://doi.org/10.1016/0009-2797(93)90025-t)

DeCastro, M.-E., Rodríguez-Belmonte, E., & González-Siso, M.-I. (2016). Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozyms. In *Frontiers in Microbiology* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2016.01521>

de Miguel Bouzas, T., Barros-Velazquez, J., & Gonzalez Villa, T. (2006). Industrial Applications of Hyperthermophilic Enzymes: A Review. In *Protein & Peptide Letters* (Vol. 13, Issue 7, pp. 645–651). Bentham Science Publishers Ltd. <https://doi.org/10.2174/092986606777790548>

Bowen De León, K., Gerlach, R., Peyton, B. M., & Fields, M. W. (2013). Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. In *Frontiers in Microbiology* (Vol. 4). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2013.00330>

de Pascale, D., Cusano, A. M., Autore, F., Parrilli, E., di Prisco, G., Marino, G., & Tutino, M. L. (2008). The cold-active Lip1 lipase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125 is a member of a new bacterial lipolytic enzyme family. In *Extremophiles* (Vol. 12, Issue 3, pp. 311–323). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-008-0163-9>

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., & Swanson, R. V. (1998). The complete genome of the hyperthermophilic

bacterium *Aquifex aeolicus*. In *Nature* (Vol. 392, Issue 6674, pp. 353–358). Springer Science and Business Media LLC. <https://doi.org/10.1038/32831>

Defarge, N., Spiroux de Vendômois, J., & Séralini, G. E. (2018). Toxicity of formulants and heavy metals in glyphosate-based herbicides and other pesticides. In *Toxicology Reports* (Vol. 5, pp. 156–163). Elsevier BV. <https://doi.org/10.1016/j.toxrep.2017.12.025>

DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F., & Lombardi, A. (1999). De Novo Design and Structural Characterization of Proteins and Metalloproteins. In *Annual Review of Biochemistry* (Vol. 68, Issue 1, pp. 779–819). Annual Reviews. <https://doi.org/10.1146/annurev.biochem.68.1.779>

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. In *Bioinformatics* (Vol. 23, Issue 6, pp. 673–679). Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btm009>

Des Marais, D. J., & Walter, M. R. (2019). Terrestrial Hot Spring Systems: Introduction. In *Astrobiology* (Vol. 19, Issue 12, pp. 1419–1432). Mary Ann Liebert Inc. <https://doi.org/10.1089/ast.2018.1976>

Demirjian, D. C., Morís-Varas, F., & Cassidy, C. S. (2001). Enzymes from extremophiles. In *Current Opinion in Chemical Biology* (Vol. 5, Issue 2, pp. 144–151). Elsevier BV. [https://doi.org/10.1016/s1367-5931\(00\)00183-6](https://doi.org/10.1016/s1367-5931(00)00183-6)

Deshmukh, R., Khardenavis, A. A., & Purohit, H. J. (2016). Diverse Metabolic Capacities of Fungi for Bioremediation. *Indian journal of microbiology*, 56(3), 247–264. <https://doi.org/10.1007/s12088-016-0584-6>

Diep, P., Mahadevan, R., & Yakunin, A. F. (2018). Heavy Metal Removal by Bioaccumulation Using Genetically Engineered Microorganisms. *Frontiers in bioengineering and biotechnology*, 6, 157. <https://doi.org/10.3389/fbioe.2018.00157>

Dick, J. M., & Shock, E. L. (2013). A Metastable Equilibrium Model for the Relative Abundances of Microbial Phyla in a Hot Spring. In J. H. Badger (Ed.), *PLoS ONE* (Vol. 8,

Issue 9, p. e72395). Public Library of Science (PLoS).
<https://doi.org/10.1371/journal.pone.0072395>

Djokic, T., Van Kranendonk, M. J., Campbell, K. A., Walter, M. R., & Ward, C. R. (2017). Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. In *Nature Communications* (Vol. 8, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/ncomms15263>

Do, J. H., & Choi, D.-K. (2006). Computational approaches to gene prediction. *The Journal of Microbiology*, 44(2), 137-144.

Driscoll, C. B., Otten, T. G., Brown, N. M., & Dreher, T. W. (2017). Towards long-read metagenomics: complete assembly of three novel genomes from bacteria dependent on a diazotrophic cyanobacterium in a freshwater lake co-culture. In *Standards in Genomic Sciences* (Vol. 12, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s40793-017-0224-8>

Dubinkina, V. B., Ischenko, D. S., Ulyantsev, V. I., Tyakht, A. V., & Alexeev, D. G. (2016). Assessment of k-mer spectrum applicability for metagenomic dissimilarity analysis. In *BMC Bioinformatics* (Vol. 17, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12859-015-0875-7>

Duhart, P., Crignola, P., Ordoñez, A., Muñoz, J. (2000). Franjas Metalogénicas en Chiloé Continental (41°-44°S). Congreso Geológico Chileno No. 9. Puerto Varas, Chile.

Eder, W., & Huber, R. (2002). New isolates and physiological properties of the Aquificales and description of *Thermocrinis albus* sp. nov. In *Extremophiles* (Vol. 6, Issue 4, pp. 309–318). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-001-0259-y>

Eggert, T., van Pouderooyen, G., Dijkstra, B. W., & Jaeger, K.-E. (2001). Lipolytic enzymes LipA and LipB from *Bacillus subtilis* differ in regulation of gene expression, biochemical properties, and three-dimensional structure. In *FEBS Letters* (Vol. 502, Issue 3, pp. 89–92). Wiley. [https://doi.org/10.1016/s0014-5793\(01\)02665-5](https://doi.org/10.1016/s0014-5793(01)02665-5)

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2018). The Pfam protein families database in 2019. In *Nucleic Acids Research* (Vol. 47, Issue D1, pp. D427–D432). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gky995>

Eloe-Fadrosh, E. A., Paez-Espino, D., Jarett, J., Dunfield, P. F., Hedlund, B. P., Dekas, A. E., Grasby, S. E., Brady, A. L., Dong, H., Briggs, B. R., Li, W.-J., Goudeau, D., Malmstrom, R., Pati, A., Pett-Ridge, J., Rubin, E. M., Woyke, T., Kyrpides, N. C., & Ivanova, N. N. (2016). Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. In *Nature Communications* (Vol. 7, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/ncomms10476>

Engel, A. S., Johnson, L. R., & Porter, M. L. (2012). Arsenite oxidase gene diversity among Chloroflexi and Proteobacteria from El Tatio Geysers Field, Chile. In *FEMS Microbiology Ecology* (Vol. 83, Issue 3, pp. 745–756). Oxford University Press (OUP). <https://doi.org/10.1111/1574-6941.12030>

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. In *PeerJ* (Vol. 3, p. e1319). PeerJ. <https://doi.org/10.7717/peerj.1319>

Escudero, L. V., Casamayor, E. O., Chong, G., Pedrós-Alió, C., & Demergasso, C. (2013). Distribution of Microbial Arsenic Reduction, Oxidation and Extrusion Genes along a Wide Range of Environmental Arsenic Concentrations. In M. Vos (Ed.), *PLoS ONE* (Vol. 8, Issue 10, p. e78890). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0078890>

Espariz, M., Checa, S. K., Audero, M. E. P., Pontel, L. B., & Soncini, F. C. (2007). Dissecting the Salmonella response to copper. In *Microbiology* (Vol. 153, Issue 9, pp. 2989–2997). Microbiology Society. <https://doi.org/10.1099/mic.0.2007/006536-0>

Eungrasamee, K., Incharoensakdi, A., Lindblad, P., & Jantaro, S. (2020). Synechocystis sp. PCC 6803 overexpressing genes involved in CBB cycle and free fatty acid cycling enhances

the significant levels of intracellular lipids and secreted free fatty acids. In *Scientific Reports* (Vol. 10, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-020-61100-4>

Fernández-Aparicio, M., Bernard, A., Falchetto, L., Marget, P., Chauvel, B., Steinberg, C., Morris, C. E., Gibot-Leclerc, S., Boari, A., Vurro, M., Bohan, D. A., Sands, D. C., & Reboud, X. (2017). Investigation of Amino Acids As Herbicides for Control of *Orobanche* minor Parasitism in Red Clover. In *Frontiers in Plant Science* (Vol. 8). Frontiers Media SA. <https://doi.org/10.3389/fpls.2017.00842>

Fetherston, J. D., Bertolino, V. J., & Perry, R. D. (1999). YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. In *Molecular Microbiology* (Vol. 32, Issue 2, pp. 289–299). Wiley. <https://doi.org/10.1046/j.1365-2958.1999.01348.x>

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. In *Nucleic Acids Research* (Vol. 39, Issue suppl, pp. W29–W37). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkr367>

Finsinger, K., Scholz, I., Serrano, A., Morales, S., Uribe-Lorio, L., Mora, M., Sittenfeld, A., Weckesser, J., & Hess, W. R. (2008). Characterization of true-branching cyanobacteria from geothermal sites and hot springs of Costa Rica. In *Environmental Microbiology* (Vol. 10, Issue 2, pp. 460–473). Wiley. <https://doi.org/10.1111/j.1462-2920.2007.01467.x>

Fischer, M. 1907. *The physiology of alimentation*. New York. J. Wiley & Sons.

Fishbain, S., Dillon, J. G., Gough, H. L., & Stahl, D. A. (2003). Linkage of High Rates of Sulfate Reduction in Yellowstone Hot Springs to Unique Sequence Types in the Dissimilatory Sulfate Respiration Pathway. In *Applied and Environmental Microbiology* (Vol. 69, Issue 6, pp. 3663–3667). American Society for Microbiology. <https://doi.org/10.1128/aem.69.6.3663-3667.2003>

Forlani, G., Pavan, M., Gramek, M., Kafarski, P., & Lipok, J. (2008). Biochemical Bases for a Widespread Tolerance of Cyanobacteria to the Phosphonate Herbicide Glyphosate. In

Plant and Cell Physiology (Vol. 49, Issue 3, pp. 443–456). Oxford University Press (OUP).
<https://doi.org/10.1093/pcp/pcn021>

Forouzan, E., Shariati, P., Mousavi Maleki, M. S., Karkhane, A. A., & Yakhchali, B. (2018). Practical evaluation of 11 de novo assemblers in metagenome assembly. In *Journal of Microbiological Methods* (Vol. 151, pp. 99–105). Elsevier BV.
<https://doi.org/10.1016/j.mimet.2018.06.007>

Frawley, E. R., & Fang, F. C. (2014). The ins and outs of bacterial iron metabolism. In *Molecular Microbiology* (Vol. 93, Issue 4, pp. 609–616). Wiley.
<https://doi.org/10.1111/mmi.12709>

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2014). Expanded microbial genome coverage and improved protein family annotation in the COG database. In *Nucleic Acids Research* (Vol. 43, Issue D1, pp. D261–D269). Oxford University Press (OUP).
<https://doi.org/10.1093/nar/gku1223>

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., & Koonin, E. V. (2020). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. In *Nucleic Acids Research* (Vol. 49, Issue D1, pp. D274–D281). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkaa1018>

Gao, Q., Tan, X., & Lü, X. (2012). Characterization of a key gene in membrane lipid cycle in *Synechocystis* sp. PCC6803. *Sheng wu Gong Cheng xue bao= Chinese Journal of Biotechnology*, 28(12), 1473-1481.

Ghai, R., Martin-Cuadrado, A.-B., Molto, A. G., Heredia, I. G., Cabrera, R., Martin, J., Verdú, M., Deschamps, P., Moreira, D., López-García, P., Mira, A., & Rodríguez-Valera, F. (2010). Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. In *The ISME Journal* (Vol. 4, Issue 9, pp. 1154–1166). Springer Science and Business Media LLC. <https://doi.org/10.1038/ismej.2010.44>

Gianfreda, L., & Rao, M. A. (2017). Soil Microbial and Enzymatic Diversity as Affected by the Presence of Xenobiotics. In *Soil Biology* (pp. 153–169). Springer International Publishing. https://doi.org/10.1007/978-3-319-47744-2_12

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. In *Computational and Structural Biotechnology Journal* (Vol. 18, pp. 9–19). Elsevier BV. <https://doi.org/10.1016/j.csbj.2019.11.002>

González-Sánchez, A., Cubillas, C. A., Miranda, F., Dávalos, A., & García-de los Santos, A. (2017). TheropAegene encodes a porin-like protein involved in copper transit in *Rhizobium etli* CFN42. In *MicrobiologyOpen* (Vol. 7, Issue 3, p. e00573). Wiley. <https://doi.org/10.1002/mbo3.573>

Graber, J. R., Kirshtein, J., Speck, M., & Reysenbach, A.-L. (2001). Community Structure along a Thermal Gradient in a Stream Near Obsidian Pool, Yellowstone National Park. In *Thermophiles Biodiversity, Ecology, and Evolution* (pp. 81–91). Springer US. https://doi.org/10.1007/978-1-4615-1197-7_7

Grass, G., Franke, S., Taudte, N., Nies, D. H., Kucharski, L. M., Maguire, M. E., & Rensing, C. (2005). The Metal Permease ZupT from *Escherichia coli* Is a Transporter with a Broad Substrate Spectrum. In *Journal of Bacteriology* (Vol. 187, Issue 5, pp. 1604–1611). American Society for Microbiology. <https://doi.org/10.1128/jb.187.5.1604-1611.2005>

Grass, G., Otto, M., Fricke, B., Haney, C. J., Rensing, C., Nies, D. H., & Munkelt, D. (2004). FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. In *Archives of Microbiology* (Vol. 183, Issue 1, pp. 9–18). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00203-004-0739-4>

Grass, G., Rensing, C., & Solioz, M. (2011). Metallic Copper as an Antimicrobial Surface. In *Applied and Environmental Microbiology* (Vol. 77, Issue 5, pp. 1541–1547). American Society for Microbiology. <https://doi.org/10.1128/aem.02766-10>

Greene, N. P., Hinchliffe, P., Crow, A., Ababou, A., Hughes, C., & Koronakis, V. (2013). Structure of an atypical periplasmic adaptor from a multidrug efflux pump of the spirochete *Borrelia burgdorferi*. In *FEBS Letters* (Vol. 587, Issue 18, pp. 2984–2988). Wiley. <https://doi.org/10.1016/j.febslet.2013.06.056>

Groboillot, A., Boadi, D. K., Poncelet, D., & Neufeld, R. J. (1994). Immobilization of Cells for Application in the Food Industry. In *Critical Reviews in Biotechnology* (Vol. 14, Issue 2, pp. 75–107). Informa UK Limited. <https://doi.org/10.3109/07388559409086963>

Guckes, K. R., Breland, E. J., Zhang, E. W., Hanks, S. C., Gill, N. K., Algood, H. M. S., Schmitz, J. E., Stratton, C. W., & Hadjifrangiskou, M. (2017). Signaling by two-component system noncognate partners promotes intrinsic tolerance to polymyxin B in uropathogenic *Escherichia coli*. In *Science Signaling* (Vol. 10, Issue 461). American Association for the Advancement of Science (AAAS). <https://doi.org/10.1126/scisignal.aag1775>

Gupta, R., Gigras, P., Mohapatra, H., Goswami, V. K., & Chauhan, B. (2003). Microbial α -amylases: a biotechnological perspective. In *Process Biochemistry* (Vol. 38, Issue 11, pp. 1599–1616). Elsevier BV. [https://doi.org/10.1016/s0032-9592\(03\)00053-0](https://doi.org/10.1016/s0032-9592(03)00053-0)

Gupta, R. S. (2014). The Phylum Aquificae. In *The Prokaryotes* (pp. 417–445). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38954-2_119

Hamilton-Brehm, S. D., Gibson, R. A., Green, S. J., Hopmans, E. C., Schouten, S., van der Meer, M. T. J., Shields, J. P., Damsté, J. S. S., & Elkins, J. G. (2013). *Thermodesulfobacterium geofontis* sp. nov., a hyperthermophilic, sulfate-reducing bacterium isolated from Obsidian Pool, Yellowstone National Park. In *Extremophiles* (Vol. 17, Issue 2, pp. 251–263). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-013-0512-1>

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. In *Chemistry & Biology* (Vol. 5, Issue 10, pp. R245–R249). Elsevier BV. [https://doi.org/10.1016/s1074-5521\(98\)90108-9](https://doi.org/10.1016/s1074-5521(98)90108-9)

Hang, B.-J., Hong, Q., Xie, X.-T., Huang, X., Wang, C.-H., He, J., & Li, S.-P. (2012). SulE, a Sulfonyleurea Herbicide De-Esterification Esterase from *Hansschlegelia zhihuaiae* S113. In *Applied and Environmental Microbiology* (Vol. 78, Issue 6, pp. 1962–1968). American Society for Microbiology. <https://doi.org/10.1128/aem.07440-11>

Hausmann, S., & Jaeger, K.-E. (2010). Lipolytic Enzymes from Bacteria. In *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 1099–1126). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77587-4_77

Hernández, W. (2012). Characterization of two microbial communities of extreme environment of Rincón de la Vieja National Park volcano, Guanacaste, Costa Rica [Magister Scientiae in Microbiology]. University of Costa Rica.

Hinchliffe, P., Greene, N. P., Paterson, N. G., Crow, A., Hughes, C., & Koronakis, V. (2014). Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. In *FEBS Letters* (Vol. 588, Issue 17, pp. 3147–3153). Wiley. <https://doi.org/10.1016/j.febslet.2014.06.055>

Hiraoka, S., Yang, C., & Iwasaki, W. (2016). Metagenomics and Bioinformatics in Microbial Ecology: Current Status and Beyond. In *Microbes and Environments* (Vol. 31, Issue 3, pp. 204–212). Japanese Society of Microbial Ecology. <https://doi.org/10.1264/jsme2.me16024>

Hoagland, R. E., & Zablutowicz, R. M. (2000). The Role of Plant and Microbial Hydrolytic Enzymes in Pesticide Metabolism. In *ACS Symposium Series* (pp. 58–88). American Chemical Society. <https://doi.org/10.1021/bk-2001-0777.ch004>

Houben, A. J. S., & Moolenaar, W. H. (2011). Autotaxin and LPA receptor signaling in cancer. In *Cancer and Metastasis Reviews* (Vol. 30, Issues 3–4, pp. 557–565). Springer Science and Business Media LLC. <https://doi.org/10.1007/s10555-011-9319-7>

Howell, L. L., Nye, J. A., Stehouwer, J. S., Voll, R. J., Mun, J., Narasimhan, D., Nichols, J., Sunahara, R., Goodman, M. M., Carroll, F. I., & Woods, J. H. (2014). A thermostable bacterial cocaine esterase rapidly eliminates cocaine from brain in nonhuman primates. In

Translational Psychiatry (Vol. 4, Issue 7, pp. e407–e407). Springer Science and Business Media LLC. <https://doi.org/10.1038/tp.2014.48>

Huang, X., & Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. In *Genome Research* (Vol. 9, Issue 9, pp. 868–877). Cold Spring Harbor Laboratory. <https://doi.org/10.1101/gr.9.9.868>

Huang, X., & Yang, S. (2005). Generating a Genome Assembly with PCAP. In *Current Protocols in Bioinformatics* (Vol. 11, Issue 1). Wiley. <https://doi.org/10.1002/0471250953.bi1103s11>

Huang, Q., Jiang, H., Briggs, B. R., Wang, S., Hou, W., Li, G., Wu, G., Solis, R., Arcilla, C. A., Abrajano, T., & Dong, H. (2013). Archaeal and bacterial diversity in acidic to circumneutral hot springs in the Philippines. In *FEMS Microbiology Ecology* (Vol. 85, Issue 3, pp. 452–464). Oxford University Press (OUP). <https://doi.org/10.1111/1574-6941.12134>

Huang, Y., Zhan, H., Bhatt, P., & Chen, S. (2019). Paraquat Degradation From Contaminated Environments: Current Achievements and Perspectives. In *Frontiers in Microbiology* (Vol. 10). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2019.01754>

Huber, R., Burggraf, S., Mayer, T., Barns, S. M., Rossnagel, P., & Stetter, K. O. (1995). Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. In *Nature* (Vol. 376, Issue 6535, pp. 57–58). Springer Science and Business Media LLC. <https://doi.org/10.1038/376057a0>

Huber, R., Burggraf, S., Mayer, T., Barns, S. M., Rossnagel, P., & Stetter, K. O. (1995). Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. In *Nature* (Vol. 376, Issue 6535, pp. 57–58). Springer Science and Business Media LLC. <https://doi.org/10.1038/376057a0>

Huertas, M., López-Maury, L., Giner-Lamia, J., Sánchez-Riego, A., & Florencio, F. (2014). Metals in Cyanobacteria: Analysis of the Copper, Nickel, Cobalt and Arsenic Homeostasis

Mechanisms. In *Life* (Vol. 4, Issue 4, pp. 865–886). MDPI AG. <https://doi.org/10.3390/life4040865>

Hug, K., Maher, W. A., Stott, M. B., Krikowa, F., Foster, S., & Moreau, J. W. (2014). Microbial contributions to coupled arsenic and sulfur cycling in the acid-sulfide hot spring Champagne Pool, New Zealand. In *Frontiers in Microbiology* (Vol. 5). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2014.00569>

Hugenholtz, P., Pitulle, C., Hershberger, K. L., & Pace, N. R. (1998). Novel Division Level Bacterial Diversity in a Yellowstone Hot Spring. In *Journal of Bacteriology* (Vol. 180, Issue 2, pp. 366–376). American Society for Microbiology. <https://doi.org/10.1128/jb.180.2.366-376.1998>

Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. In *BMC Bioinformatics* (Vol. 11, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/1471-2105-11-119>

Hynek, B. M., Rogers, K. L., Antunovich, M., Avard, G., & Alvarado, G. E. (2018). Lack of Microbial Diversity in an Extreme Mars Analog Setting: Poás Volcano, Costa Rica. In *Astrobiology* (Vol. 18, Issue 7, pp. 923–933). Mary Ann Liebert Inc. <https://doi.org/10.1089/ast.2017.1719>

Idury, R. M., & Waterman, M. S. (1995). A New Algorithm for DNA Sequence Assembly. In *Journal of Computational Biology* (Vol. 2, Issue 2, pp. 291–306). Mary Ann Liebert Inc. <https://doi.org/10.1089/cmb.1995.2.291>

Ikeda, T., & Yoshimura, F. (2002). A Resistance-Nodulation-Cell Division Family Xenobiotic Efflux Pump in an Obligate Anaerobe, *Porphyromonas gingivalis*. In *Antimicrobial Agents and Chemotherapy* (Vol. 46, Issue 10, pp. 3257–3260). American Society for Microbiology. <https://doi.org/10.1128/aac.46.10.3257-3260.2002>

Imamura, S., & Kitaura, S. (2000). Purification and Characterization of a Monoacylglycerol Lipase from the Moderately Thermophilic *Bacillus* sp. H-257. In *Journal of Biochemistry* (Vol.

- 127, Issue 3, pp. 419–425). Oxford University Press (OUP). <https://doi.org/10.1093/oxfordjournals.jbchem.a022623>
- Inagaki, F., Motomura, Y., Doi, K., Taguchi, S., Izawa, E., Lowe, D. R., & Ogata, S. (2001). Silicified Microbial Community at Steep Cone Hot Spring, Yellowstone National Park. In *Microbes and Environments* (Vol. 16, Issue 2, pp. 125–130). Japanese Society of Microbial Ecology. <https://doi.org/10.1264/jsme2.2001.125>
- Inskeep, W. P., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Rusch, D. B., Tringe, S. G., Macur, R. E., Jennings, R. deM., Boyd, E. S., Spear, J. R., & Roberto, F. F. (2013). Phylogenetic and Functional Analysis of Metagenome Sequence from High-Temperature Archaeal Habitats Demonstrate Linkages between Metabolic Potential and Geochemistry. In *Frontiers in Microbiology* (Vol. 4). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2013.00095>
- Ishii, M., Igarashi, Y., & Kodama, T. (1989). Purification and characterization of ATP:citrate lyase from *Hydrogenobacter thermophilus* TK-6. In *Journal of Bacteriology* (Vol. 171, Issue 4, pp. 1788–1792). American Society for Microbiology. <https://doi.org/10.1128/jb.171.4.1788-1792.1989>
- Ishikawa, T., Mizunoe, Y., Kawabata, S., Takade, A., Harada, M., Wai, S. N., & Yoshida, S. (2003). The Iron-Binding Protein Dps Confers Hydrogen Peroxide Stress Resistance to *Campylobacter jejuni*. In *Journal of Bacteriology* (Vol. 185, Issue 3, pp. 1010–1017). American Society for Microbiology. <https://doi.org/10.1128/jb.185.3.1010-1017.2003>
- Ito, K., & Akiyama, Y. (2005). Cellular functions, mechanism of action, and regulation of FtsH protease. In *Annual Review of Microbiology* (Vol. 59, Issue 1, pp. 211–231). Annual Reviews. <https://doi.org/10.1146/annurev.micro.59.030804.121316>
- Jackson, C. R., Langner, H. W., Donahoe-Christiansen, J., Inskeep, W. P., & McDermott, T. R. (2001). Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. In *Environmental Microbiology* (Vol. 3, Issue 8, pp. 532–542). Wiley. <https://doi.org/10.1046/j.1462-2920.2001.00221.x>
- Jaime-Pérez, N., Kaftan, D., Bína, D., Bokhari, S. N. H., Shreedhar, S., & Küpper, H. (2019). Mechanisms of sublethal copper toxicity damage to the photosynthetic apparatus of

Rhodospirillum rubrum. In *Biochimica et Biophysica Acta (BBA) - Bioenergetics* (Vol. 1860, Issue 8, pp. 640–650). Elsevier BV. <https://doi.org/10.1016/j.bbabbio.2019.06.004>

Janeček, Š., & Kuchtová, A. (2012). In silico identification of catalytic residues and domain fold of the family GH119 sharing the catalytic machinery with the α -amylase family GH57. In *FEBS Letters* (Vol. 586, Issue 19, pp. 3360–3366). Wiley. <https://doi.org/10.1016/j.febslet.2012.07.020>

Javaux, E. J. (2006). Extreme life on Earth—past, present and possibly beyond. In *Research in Microbiology* (Vol. 157, Issue 1, pp. 37–48). Elsevier BV. <https://doi.org/10.1016/j.resmic.2005.07.008>

Jeffries, T. C., Rayu, S., Nielsen, U. N., Lai, K., Ijaz, A., Nazaries, L., & Singh, B. K. (2018). Metagenomic Functional Potential Predicts Degradation Rates of a Model Organophosphorus Xenobiotic in Pesticide Contaminated Soils. In *Frontiers in Microbiology* (Vol. 9). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2018.00147>

Jiang, X., & Takacs-Vesbach, C. D. (2016). Microbial community analysis of pH 4 thermal springs in Yellowstone National Park. In *Extremophiles* (Vol. 21, Issue 1, pp. 135–152). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-016-0889-8>

Jiang, Z., Li, P., Van Nostrand, J. D., Zhang, P., Zhou, J., Wang, Y., Dai, X., Zhang, R., Jiang, D., & Wang, Y. (2016). Microbial communities and arsenic biogeochemistry at the outflow of an alkaline sulfide-rich hot spring. In *Scientific Reports* (Vol. 6, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/srep25262>

Jiang, W., Gao, Q., Zhang, L., Wang, H., Zhang, M., Liu, X., Zhou, Y., Ke, Z., Wu, C., Qiu, J., & Hong, Q. (2020). Identification of the key amino acid sites of the carbofuran hydrolase CehA from a newly isolated carbofuran-degrading strain *Sphingbium* sp. CFD-1. In *Ecotoxicology and Environmental Safety* (Vol. 189, p. 109938). Elsevier BV. <https://doi.org/10.1016/j.ecoenv.2019.109938>

Jo, J., Jang, Y.-S., Kim, K.-Y., Kim, M.-H., Kim, I.-J., & Chung, W.-I. (1997). Isolation of ALU1-PGene Encoding a Protein with Aluminum Tolerance Activity from *Arthrobacter viscosus*. In

Biochemical and Biophysical Research Communications (Vol. 239, Issue 3, pp. 835–839). Elsevier BV. <https://doi.org/10.1006/bbrc.1997.7567>

Joshi, T., & Xu, D. (2007). Quantitative assessment of relationship between sequence similarity and function similarity. In *BMC Genomics* (Vol. 8, Issue 1, p. 222). Springer Science and Business Media LLC. <https://doi.org/10.1186/1471-2164-8>

Jungbluth, S. P., Amend, J. P., & Rappé, M. S. (2017). Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. In *Scientific Data* (Vol. 4, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/sdata.2017.37>

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2018). New approach for understanding genome variations in KEGG. In *Nucleic Acids Research* (Vol. 47, Issue D1, pp. D590–D595). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gky962>

Karimi, M., & Biria, D. (2019). The promiscuous activity of alpha-amylase in biodegradation of low-density polyethylene in a polymer-starch blend. In *Scientific Reports* (Vol. 9, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-019-39366-0>

Kashefi, K., & Lovley, D. R. (2000). Reduction of Fe (III), Mn(IV), and Toxic Metals at 100°C by *Pyrobaculum islandicum*. In *Applied and Environmental Microbiology* (Vol. 66, Issue 3, pp. 1050–1056). American Society for Microbiology. <https://doi.org/10.1128/aem.66.3.1050-1056.2000>

Kashefi, K., Holmes, D. E., Reysenbach, A.-L., & Lovley, D. R. (2002). Use of Fe(III) as an Electron Acceptor To Recover Previously Uncultured Hyperthermophiles: Isolation and Characterization of *Geothermobacterium ferrireducens* gen. nov., sp. nov. In *Applied and Environmental Microbiology* (Vol. 68, Issue 4, pp. 1735–1742). American Society for Microbiology. <https://doi.org/10.1128/aem.68.4.1735-1742.2002>

Kawasumi, T., Igarashi, Y., Kodama, T., & Minoda, Y. (1980). Isolation of strictly thermophilic and obligately autotrophic hydrogen bacteria. In *Agricultural and Biological Chemistry* (Vol.

44, Issue 8, pp. 1985–1986). Oxford University Press (OUP).
<https://doi.org/10.1271/bbb1961.44.1985>

Khajeh, K., Naderi-Manesh, H., Ranjbar, B., Moosavi-Movahedi, A. akbar, & Nemat-Gorgani, M. (2001). Chemical modification of lysine residues in *Bacillus* α -amylases: effect on activity and stability. In *Enzyme and Microbial Technology* (Vol. 28, Issue 6, pp. 543–549). Elsevier BV. [https://doi.org/10.1016/s0141-0229\(01\)00296-4](https://doi.org/10.1016/s0141-0229(01)00296-4)

Kibegwa, F. M., Bett, R. C., Gachuri, C. K., Stomeo, F., & Mujibi, F. D. (2020). A Comparison of Two DNA Metagenomic Bioinformatic Pipelines While Evaluating the Microbial Diversity in Feces of Tanzanian Small Holder Dairy Cattle. In *BioMed Research International* (Vol. 2020, pp. 1–12). Hindawi Limited. <https://doi.org/10.1155/2020/2348560>

Kobayashi, K., Tsukagoshi, N., & Aono, R. (2001). Suppression of Hypersensitivity of *Escherichia coli* *acrB* Mutant to Organic Solvents by Integrational Activation of the *acrEF* Operon with the IS 1 or IS 2 Element. In *Journal of Bacteriology* (Vol. 183, Issue 8, pp. 2646–2653). American Society for Microbiology. <https://doi.org/10.1128/jb.183.8.2646-2653.2001>

Koch, D., Nies, D. H., & Grass, G. (2006). The RcnRA (YohLM) system of *Escherichia coli*: A connection between nickel, cobalt and iron homeostasis. In *BioMetals* (Vol. 20, Issue 5, pp. 759–771). Springer Science and Business Media LLC. <https://doi.org/10.1007/s10534-006-9039-6>

Kohli, U., Nigam, P., Singh, D., & Chaudhary, K. (2001). Thermostable, alkalophilic and cellulase free xylanase production by *Thermoactinomyces thalophilus* subgroup C. In *Enzyme and Microbial Technology* (Vol. 28, Issues 7–8, pp. 606–610). Elsevier BV. [https://doi.org/10.1016/s0141-0229\(01\)00320-9](https://doi.org/10.1016/s0141-0229(01)00320-9)

Konopka, A., Lindemann, S., & Fredrickson, J. (2014). Dynamics in microbial communities: unraveling mechanisms to identify principles. In *The ISME Journal* (Vol. 9, Issue 7, pp. 1488–1495). Springer Science and Business Media LLC. <https://doi.org/10.1038/ismej.2014.2>

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., & Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. In *Nature*

(Vol. 405, Issue 6789, pp. 914–919). Springer Science and Business Media LLC. <https://doi.org/10.1038/35016007>

Kovacic, F., Babic, N., Krauss, U., & Jaeger, K.-E. (2018). Classification of Lipolytic Enzymes from Bacteria. In *Aerobic Utilization of Hydrocarbons, Oils and Lipids* (pp. 1–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-39782-5_39-1

Kublanov, I. V., Perevalova, A. A., Slobodkina, G. B., Lebedinsky, A. V., Bidzhieva, S. K., Kolganova, T. V., Kaliberda, E. N., Rumsh, L. D., Haertlé, T., & Bonch-Osmolovskaya, E. A. (2009). Biodiversity of Thermophilic Prokaryotes with Hydrolytic Activities in Hot Springs of Uzon Caldera, Kamchatka (Russia). In *Applied and Environmental Microbiology* (Vol. 75, Issue 1, pp. 286–291). American Society for Microbiology. <https://doi.org/10.1128/aem.00607-08>

Kudo, T., Sakamoto, K., Akinaga, M., Kawauchi, A., Nakahara, T., Zhang, X., Yamada, A., Oshima, K., Suda, W., Kuwahara, H., Nakamura, N., Nogi, Y., Kitamura, K., Yuki, M., Iida, T., Moriya, S., Inoue, T., Hongoh, Y., Hattori, M., & Ohkuma, M. (2014). Draft Genome Sequences of Cyclodextrin-Producing Alkaliphilic *Bacillus* Strains JCM 19045, JCM 19046, and JCM 19047. In *Genome Announcements* (Vol. 2, Issue 2). American Society for Microbiology. <https://doi.org/10.1128/genomea.00211-14>

Kumar, S., Krishnani, K. K., Bhushan, B., & Brahmane, M. P. (2015). Metagenomics: Retrospect and Prospects in High Throughput Age. In *Biotechnology Research International* (Vol. 2015, pp. 1–13). Hindawi Limited. <https://doi.org/10.1155/2015/121735>

Kumar, A., & Singh, J. S. (2017). Cyanoremediation: A Green-Clean Tool for Decontamination of Synthetic Pesticides from Agro- and Aquatic Ecosystems. In *Agro-Environmental Sustainability* (pp. 59–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-49727-3_4

Kumar, A., Mukhia, S., Kumar, N., Acharya, V., Kumar, S., & Kumar, R. (2020). A Broad Temperature Active Lipase Purified From a Psychrotrophic Bacterium of Sikkim Himalaya With Potential Application in Detergent Formulation. In *Frontiers in Bioengineering and Biotechnology* (Vol. 8). Frontiers Media SA. <https://doi.org/10.3389/fbioe.2020.00642>

Kundu, K., Marozava, S., Ehrl, B., Merl-Pham, J., Griebler, C., & Elsner, M. (2019). Defining lower limits of biodegradation: atrazine degradation regulated by mass transfer and maintenance demand in *Arthrobacter aurescens* TC1. In *The ISME Journal* (Vol. 13, Issue 9, pp. 2236–2251). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41396-019-0430-z>

Kyrgyzov, O., Prost, V., Gazut, S., Farcy, B., & Bröls, T. (2020). Binning unassembled short reads based on k-mer abundance covariance using sparse coding. In *GigaScience* (Vol. 9, Issue 4). Oxford University Press (OUP). <https://doi.org/10.1093/gigascience/giaa028>

Laehnemann, D., Borkhardt, A., & McHardy, A. C. (2015). Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. In *Briefings in Bioinformatics* (Vol. 17, Issue 1, pp. 154–179). Oxford University Press (OUP). <https://doi.org/10.1093/bib/bbv029>

Lamichhane, J. R., Osdaghi, E., Behlau, F., Köhl, J., Jones, J. B., & Aubertot, J.-N. (2018). Thirteen decades of antimicrobial copper compounds applied in agriculture. A review. In *Agronomy for Sustainable Development* (Vol. 38, Issue 3). Springer Science and Business Media LLC. <https://doi.org/10.1007/s13593-018-0503-9>

Landeta, C., Dávalos, A., Cevallos, M. A., Geiger, O., Brom, S., & Romero, D. (2011). Plasmids with a Chromosome-Like Role in Rhizobia. In *Journal of Bacteriology* (Vol. 193, Issue 6, pp. 1317–1326). American Society for Microbiology. <https://doi.org/10.1128/jb.01184-10>

Langklotz, S., Baumann, U., & Narberhaus, F. (2012). Structure and function of the bacterial AAA protease FtsH. In *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* (Vol. 1823, Issue 1, pp. 40–48). Elsevier BV. <https://doi.org/10.1016/j.bbamcr.2011.08.015>

Langner, H. W., Jackson, C. R., McDermott, T. R., & Inskeep, W. P. (2001). Rapid Oxidation of Arsenite in a Hot Spring Ecosystem, Yellowstone National Park. In *Environmental Science & Technology* (Vol. 35, Issue 16, pp. 3302–3309). American Chemical Society (ACS). <https://doi.org/10.1021/es0105562>

Larsen, T., & Krogh, A. (2003). In BMC Bioinformatics (Vol. 4, Issue 1, p. 21). Springer Science and Business Media LLC. <https://doi.org/10.1186/1471-2105-4-21>

Dalmaso, G., Ferreira, D., & Vermelho, A. (2015). Marine Extremophiles: A Source of Hydrolases for Biotechnological Applications. In Marine Drugs (Vol. 13, Issue 4, pp. 1925–1965). MDPI AG. <https://doi.org/10.3390/md13041925>

Lenart, A., Dudkiewicz, M., Grynberg, M., & Pawłowski, K. (2013). CLCAs - A Family of Metalloproteases of Intriguing Phylogenetic Distribution and with Cases of Substituted Catalytic Sites. In S. Ranganathan (Ed.), PLoS ONE (Vol. 8, Issue 5, p. e62272). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0062272>

Letunic, I., & Bork, P. (2006). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. In Bioinformatics (Vol. 23, Issue 1, pp. 127–128). Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btl529>

Levisson, M., van der Oost, J., & Kengen, S. W. M. (2009). Carboxylic ester hydrolases from hyperthermophiles. In Extremophiles (Vol. 13, Issue 4, pp. 567–581). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-009-0260-4>

Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., Yang, B., & Fan, W. (2011). Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. In Briefings in Functional Genomics (Vol. 11, Issue 1, pp. 25–37). Oxford University Press (OUP). <https://doi.org/10.1093/bfpg/elr035>

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. In Bioinformatics (Vol. 31, Issue 10, pp. 1674–1676). Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btv033>

Li, G., Chen, Y., Fang, X., Su, F., Xu, L., & Yan, Y. (2018). Identification of a hot-spot to enhance *Candida rugosa* lipase thermostability by rational design methods. In RSC

Advances (Vol. 8, Issue 4, pp. 1948–1957). Royal Society of Chemistry (RSC). <https://doi.org/10.1039/c7ra11679a>

Li, H., & Zhang, X. (2005). Characterization of thermostable lipase from thermophilic *Geobacillus* sp. TW1. In *Protein Expression and Purification* (Vol. 42, Issue 1, pp. 153–159). Elsevier BV. <https://doi.org/10.1016/j.pep.2005.03.011>

Lin, K.-H., Liao, B.-Y., Chang, H.-W., Huang, S.-W., Chang, T.-Y., Yang, C.-Y., Wang, Y.-B., Lin, Y.-T. K., Wu, Y.-W., Tang, S.-L., & Yu, H.-T. (2015). Metabolic characteristics of dominant microbes and key rare species from an acidic hot spring in Taiwan revealed by metagenomics. In *BMC Genomics* (Vol. 16, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12864-015-2230-9>

Lin, K.-H., Liao, B.-Y., Chang, H.-W., Huang, S.-W., Chang, T.-Y., Yang, C.-Y., Wang, Y.-B., Lin, Y.-T. K., Wu, Y.-W., Tang, S.-L., & Yu, H.-T. (2015). Metabolic characteristics of dominant microbes and key rare species from an acidic hot spring in Taiwan revealed by metagenomics. In *BMC Genomics* (Vol. 16, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12864-015-2230-9>

Liu, T., Ramesh, A., Ma, Z., Ward, S. K., Zhang, L., George, G. N., Talaat, A. M., Sacchettini, J. C., & Giedroc, D. P. (2006). CsoR is a novel *Mycobacterium tuberculosis* copper-sensing transcriptional regulator. In *Nature Chemical Biology* (Vol. 3, Issue 1, pp. 60–68). Springer Science and Business Media LLC. <https://doi.org/10.1038/nchembio844>

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. In *Nucleic Acids Research* (Vol. 42, Issue D1, pp. D490–D495). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkt1178>

López-López, O., Cerdán, M., & González-Siso, M. (2013). Hot Spring Metagenomics. In *Life* (Vol. 3, Issue 2, pp. 308–320). MDPI AG. <https://doi.org/10.3390/life3020308>

López-López, O., Knapik, K., Cerdán, M.-E., & González-Siso, M.-I. (2015). Metagenomics of an Alkaline Hot Spring in Galicia (Spain): Microbial Diversity Analysis and Screening for

Novel Lipolytic Enzymes. In *Frontiers in Microbiology* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2015.01291>

López-Maury, L., Florencio, F. J., & Reyes, J. C. (2003). Arsenic Sensing and Resistance System in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. In *Journal of Bacteriology* (Vol. 185, Issue 18, pp. 5363–5371). American Society for Microbiology. <https://doi.org/10.1128/jb.185.18.5363-5371.2003>

Lozupone, C. A., & Knight, R. (2007). Global patterns in bacterial diversity. In *Proceedings of the National Academy of Sciences* (Vol. 104, Issue 27, pp. 11436–11440). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0611525104>

Lu, Y., Yeung, N., Sieracki, N., & Marshall, N. M. (2009). Design of functional metalloproteins. In *Nature* (Vol. 460, Issue 7257, pp. 855–862). Springer Science and Business Media LLC. <https://doi.org/10.1038/nature08304>

Lukashin, A. (1998). GeneMark.hmm: new solutions for gene finding. In *Nucleic Acids Research* (Vol. 26, Issue 4, pp. 1107–1115). Oxford University Press (OUP). <https://doi.org/10.1093/nar/26.4.1107>

Lutz, S., Lichter, J., & Liu, L. (2007). Exploiting Temperature-Dependent Substrate Promiscuity for Nucleoside Analogue Activation by Thymidine Kinase from *Thermotoga maritima*. In *Journal of the American Chemical Society* (Vol. 129, Issue 28, pp. 8714–8715). American Chemical Society (ACS). <https://doi.org/10.1021/ja0734391>

Mackenzie, R., Pedrós-Alió, C., & Díez, B. (2012). Bacterial composition of microbial mats in hot springs in Northern Patagonia: variations with seasons and temperature. In *Extremophiles* (Vol. 17, Issue 1, pp. 123–136). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-012-0499-z>

Macur, R. E., Langner, H. W., Kocar, B. D., & Inskeep, W. P. (2004). Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring. In *Geobiology* (Vol. 2, Issue 3, pp. 163–177). Wiley. <https://doi.org/10.1111/j.1472-4677.2004.00032.x>

Magnet, S., Courvalin, P., & Lambert, T. (2001). Resistance-Nodulation-Cell Division-Type Efflux Pump Involved in Aminoglycoside Resistance in *Acinetobacter baumannii* Strain BM4454. In *Antimicrobial Agents and Chemotherapy* (Vol. 45, Issue 12, pp. 3375–3380). American Society for Microbiology. <https://doi.org/10.1128/aac.45.12.3375-3380.2001>

Malik, A. (2004). Metal bioremediation through growing cells. In *Environment International* (Vol. 30, Issue 2, pp. 261–278). Elsevier BV. <https://doi.org/10.1016/j.envint.2003.08.001>

Mangrola, A. V., Dudhagara, P. R., Koringa, P. G., Joshi, C. G., & Patel, R. K. (2018). Metagenomic microbial community profiling of Unnai hot spring by Ion-Torrent based shotgun sequencing. In *Microbiology* (Vol. 87, Issue 1, pp. 143–146). Pleiades Publishing Ltd. <https://doi.org/10.1134/s0026261718010113>

Margesin, R., & Schinner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. In *Extremophiles* (Vol. 5, Issue 2, pp. 73–83). Springer Science and Business Media LLC. <https://doi.org/10.1007/s007920100184>

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. In *Nature* (Vol. 437, Issue 7057, pp. 376–380). Springer Science and Business Media LLC. <https://doi.org/10.1038/nature03959>

Martinez, B., Tomkins, J., Wackett, L. P., Wing, R., & Sadowsky, M. J. (2001). Complete Nucleotide Sequence and Organization of the Atrazine Catabolic Plasmid pADP-1 from *Pseudomonas* sp . Strain ADP. In *Journal of Bacteriology* (Vol. 183, Issue 19, pp. 5684–5697). American Society for Microbiology. <https://doi.org/10.1128/jb.183.19.5684-5697.2001>

Masaki, Y., Hirano, S. I., & Okibe, N. (2015). Microbial Community Structure Analysis of Blood Pond Hell Hot Spring in Japan and Search for Metal-Reducing Microbes. In *Advanced Materials Research* (Vol. 1130, pp. 45–49). Trans Tech Publications, Ltd. <https://doi.org/10.4028/www.scientific.net/amr.1130.45>

Mastin, B. J., & Rodgers, J. H. (2000). Toxicity and Bioavailability of Copper Herbicides (Clearigate, Cutrine-Plus, and Copper Sulfate) to Freshwater Animals. In *Archives of Environmental Contamination and Toxicology* (Vol. 39, Issue 4, pp. 445–451). Springer Science and Business Media LLC. <https://doi.org/10.1007/s002440010126>

Mehetre, G. T., Paranjpe, A., Dastager, S. G., & Dharne, M. S. (2016a). Investigation of Microbial Diversity in Geothermal Hot Springs in Unkeshwar, India, Based on 16S rRNA Amplicon Metagenome Sequencing. In *Genome Announcements* (Vol. 4, Issue 1). American Society for Microbiology. <https://doi.org/10.1128/genomea.01766-15>

Mehetre, G. T., Paranjpe, A. S., Dastager, S. G., & Dharne, M. S. (2016b). Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. In *Genomics Data* (Vol. 7, pp. 140–143). Elsevier BV. <https://doi.org/10.1016/j.gdata.2015.12.031>

Mehetre, G., Shah, M., Dastager, S. G., & Dharne, M. S. (2018). Untapped bacterial diversity and metabolic potential within Unkeshwar hot springs, India. In *Archives of Microbiology* (Vol. 200, Issue 5, pp. 753–770). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00203-018-1484-4>

Mehta, D., & Satyanarayana, T. (2016). Bacterial and Archaeal α -Amylases: Diversity and Amelioration of the Desirable Characteristics for Industrial Applications. In *Frontiers in Microbiology* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2016.01129>

Mikheenko, A., Saveliev, V., & Gurevich, A. (2015). MetaQUAST: evaluation of metagenome assemblies. In *Bioinformatics* (Vol. 32, Issue 7, pp. 1088–1090). Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btv697>

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turrone, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., Lugli, G. A., Rodriguez, J. M., Bode, L., de Vos, W., Gueimonde, M., Margolles, A., van Sinderen, D., & Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of

the Infant Gut Microbiota. In *Microbiology and Molecular Biology Reviews* (Vol. 81, Issue 4). American Society for Microbiology. <https://doi.org/10.1128/mnbr.00036-17>

Miller, S. R., & Castenholz, R. W. (2000). Evolution of Thermotolerance in Hot Spring Cyanobacteria of the Genus *Synechococcus*. In *Applied and Environmental Microbiology* (Vol. 66, Issue 10, pp. 4222–4229). American Society for Microbiology. <https://doi.org/10.1128/aem.66.10.4222-4229.2000>

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. In *Genomics* (Vol. 95, Issue 6, pp. 315–327). Elsevier BV. <https://doi.org/10.1016/j.ygeno.2010.03.001>

Mishra, A., & Malik, A. (2013). Recent Advances in Microbial Metal Bioaccumulation. In *Critical Reviews in Environmental Science and Technology* (Vol. 43, Issue 11, pp. 1162–1222). Informa UK Limited. <https://doi.org/10.1080/10934529.2011.627044>

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2020). Pfam: The protein families database in 2021. In *Nucleic Acids Research* (Vol. 49, Issue D1, pp. D412–D419). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkaa913>

Mohammad, B. T., Al Daghistani, H. I., Jaouani, A., Abdel-Latif, S., & Kennes, C. (2017). Isolation and Characterization of Thermophilic Bacteria from Jordanian Hot Springs: *Bacillus licheniformis* and *Thermomonas hydrothermalis* Isolates as Potential Producers of Thermostable Enzymes. In *International Journal of Microbiology* (Vol. 2017, pp. 1–12). Hindawi Limited. <https://doi.org/10.1155/2017/6943952>

Mohiddin, G. J., Srinivasulu, M., Subramanyam, K., Madakka, M., Meghana, D., & Rangaswamy, V. (2013). Influence of insecticides flubendiamide and spinosad on biological activities in tropical black and red clay soils. In *3 Biotech* (Vol. 5, Issue 1, pp. 13–21). Springer Science and Business Media LLC. <https://doi.org/10.1007/s13205-013-0188-3>

Mollinedo, N. (2013). Investigación hidrogeológica para determinar el origen del arsénico en aguas para consumo humano en la Región Noroeste de Aguas Zarcas, San Carlos, Alajuela,

Costa Rica. [Maestría Académica en Hidrogeología y Manejo de Recursos Hídricos]. University of Costa Rica.

Moparthy, V. K., Li, X., Vavitsas, K., Dzhygyr, I., Sandh, G., Magnuson, A., & Stensjö, K. (2016). The two Dps proteins, NpDps2 and NpDps5, are involved in light-induced oxidative stress tolerance in the N₂-fixing cyanobacterium *Nostoc punctiforme*. In *Biochimica et Biophysica Acta (BBA) - Bioenergetics* (Vol. 1857, Issue 11, pp. 1766–1776). Elsevier BV. <https://doi.org/10.1016/j.bbabi.2016.08.003>

Morales, S. (2008). Diversidad morfológica y posición filogenética de cianobacterias encontradas en fuentes termales y volcanes de Costa Rica. [Microbiology Master's degree]. University of Costa Rica.

Muralidharan, H. S., Shah, N., Meisel, J. S., & Pop, M. (2021). Binnacle: Using Scaffolds to Improve the Contiguity and Quality of Metagenomic Bins. In *Frontiers in Microbiology* (Vol. 12). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2021.638561>

Murayama, N., Shimizu, H., Takiguchi, S., Baba, Y., Amino, H., Horiuchi, T., Sekimizu, K., & Miki, T. (1996). Evidence for Involvement of *Escherichia coli* Genes *pmbA*, *csrA* and a Previously Unrecognized GenetIdD, in the Control of DNA Gyrase by *letD(ccdB)* of Sex Factor F. In *Journal of Molecular Biology* (Vol. 256, Issue 3, pp. 483–502). Elsevier BV. <https://doi.org/10.1006/jmbi.1996.0103>

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., Horreard, F., Amann, R., Jørgensen, B. B., & Kuypers, M. M. M. (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. In *Proceedings of the National Academy of Sciences* (Vol. 105, Issue 46, pp. 17861–17866). Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.0809329105>

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H. J., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H.-H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., ... Venter, J. C. (2000). A Whole-Genome Assembly of *Drosophila*. In *Science* (Vol. 287,

Issue 5461, pp. 2196–2204). American Association for the Advancement of Science (AAAS). <https://doi.org/10.1126/science.287.5461.2196>

Nagarajan, N., & Pop, M. (2013). Sequence assembly demystified. In *Nature Reviews Genetics* (Vol. 14, Issue 3, pp. 157–167). Springer Science and Business Media LLC. <https://doi.org/10.1038/nrg3367>

Nakagawa, T., & Fukui, M. (2003). Molecular Characterization of Community Structures and Sulfur Metabolism within Microbial Streamers in Japanese Hot Springs. In *Applied and Environmental Microbiology* (Vol. 69, Issue 12, pp. 7044–7057). American Society for Microbiology. <https://doi.org/10.1128/aem.69.12.7044-7057.2003>

Nakagawa, S., Shtaih, Z., Banta, A., Beveridge, T. J., Sako, Y., & Reysenbach, A.-L. (2005). *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium azorense*. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 55, Issue 6, pp. 2263–2268). Microbiology Society. <https://doi.org/10.1099/ijs.0.63708-0>

Nakamori, H., Yatabe, T., Yoon, K.-S., & Ogo, S. (2014). Purification and characterization of an oxygen-evolving photosystem II from *Leptolyngbya* sp. strain O-77. In *Journal of Bioscience and Bioengineering* (Vol. 118, Issue 2, pp. 119–124). Elsevier BV. <https://doi.org/10.1016/j.jbiosc.2014.01.009>

Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. In *Nucleic Acids Research* (Vol. 40, Issue 20, pp. e155–e155). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gks678>

Nandal, A., Huggins, C. C. O., Woodhall, M. R., McHugh, J., Rodríguez-Quiñones, F., Quail, M. A., Guest, J. R., & Andrews, S. C. (2009). Induction of the ferritin gene (*ftnA*) of *Escherichia coli* by Fe²⁺-Fur is mediated by reversal of H-NS silencing and is RyhB independent. In

Molecular Microbiology (Vol. 75, Issue 3, pp. 637–657). Wiley. <https://doi.org/10.1111/j.1365-2958.2009.06977.x>

Narasimhan, D., Woods, J. H., & Sunahara, R. K. (2012). Bacterial cocaine esterase: a protein-based therapy for cocaine overdose and addiction. In *Future Medicinal Chemistry* (Vol. 4, Issue 2, pp. 137–150). Future Science Ltd. <https://doi.org/10.4155/fmc.11.194>

Nguyen, T. T. H., Myrold, D. D., & Mueller, R. S. (2019). Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat. In *Frontiers in Microbiology* (Vol. 10). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2019.00413>

Nicolaou, S. A., Fast, A. G., Nakamaru-Ogiso, E., & Papoutsakis, E. T. (2013). Overexpression of *offetA*(ybbL) and *offetB*(ybbM), Encoding an Iron Exporter, Enhances Resistance to Oxidative Stress in *Escherichia coli*. In *Applied and Environmental Microbiology* (Vol. 79, Issue 23, pp. 7210–7219). American Society for Microbiology. <https://doi.org/10.1128/aem.02322-13>

Nishino, K., Nikaido, E., & Yamaguchi, A. (2007). Regulation of Multidrug Efflux Systems Involved in Multidrug and Metal Resistance of *Salmonella enterica* Serovar Typhimurium. In *Journal of Bacteriology* (Vol. 189, Issue 24, pp. 9066–9075). American Society for Microbiology. <https://doi.org/10.1128/jb.01045-07>

Núñez, J. 2012. Evaluación de la capacidad de retención de Cu^{2+} , Zn^{2+} y Al^{3+} de cianobacterias para su recuperación de aguas residuales. [Licenciatura en Química. Escuela de Química]. University of Costa Rica. Costa Rica.

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. In *Genome Research* (Vol. 27, Issue 5, pp. 824–834). Cold Spring Harbor Laboratory. <https://doi.org/10.1101/gr.213959.116>

Nzioki, H. S., Oyosi, F., Morris, C. E., Kaya, E., Pilgeram, A. L., Baker, C. S., & Sands, D. C. (2016). *Striga* Biocontrol on a Toothpick: A Readily Deployable and Inexpensive Method for Smallholder Farmers. In *Frontiers in Plant Science* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fpls.2016.01121>

Olivera, N. L., Sequeiros, C., & Nievas, M. L. (2007). Diversity and enzyme properties of protease-producing bacteria isolated from sub-Antarctic sediments of Isla de Los Estados, Argentina. In *Extremophiles* (Vol. 11, Issue 3, pp. 517–526). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-007-0064-3>

Oliverio, A. M., Power, J. F., Washburne, A., Cary, S. C., Stott, M. B., & Fierer, N. (2018). The ecology and diversity of microbial eukaryotes in geothermal springs. In *The ISME Journal* (Vol. 12, Issue 8, pp. 1918–1928). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41396-018-0104-2>

Olson, N. D., Treangen, T. J., Hill, C. M., Cepeda-Espinoza, V., Ghurye, J., Koren, S., & Pop, M. (2017). Metagenomic assembly through the lens of validation: recent advances in assessing and improving the quality of genomes assembled from metagenomes. In *Briefings in Bioinformatics* (Vol. 20, Issue 4, pp. 1140–1150). Oxford University Press (OUP). <https://doi.org/10.1093/bib/bbx098>

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

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

Pagaling, E., Grant, W. D., Cowan, D. A., Jones, B. E., Ma, Y., Ventosa, A., & Heaphy, S. (2012). Bacterial and archaeal diversity in two hot spring microbial mats from the geothermal region of Tengchong, China. In *Extremophiles* (Vol. 16, Issue 4, pp. 607–618). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-012-0460-1>

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., & Larsson, D. G. J. (2013). BacMet: antibacterial biocide and metal resistance genes database. In *Nucleic Acids Research* (Vol. 42, Issue D1, pp. D737–D743). Oxford University Press (OUP). <https://doi.org/10.1093/nar/gkt1252>

Parenteau, M. N., & Cady, C. L. (2010). Microbial biosignatures in iron-mineralized phototrophic mats at chocolate pots hot springs, Yellowstone National Park, United States. In *Palaios* (Vol. 25, Issue 2, pp. 97–111). Society for Sedimentary Geology. <https://doi.org/10.2110/palo.2008.p08-133r>

Pearson, W. R. (2013). An Introduction to Sequence Similarity (“Homology”) Searching. In *Current Protocols in Bioinformatics* (Vol. 42, Issue 1). Wiley. <https://doi.org/10.1002/0471250953.bi0301s42>

Pell, J., Hintze, A., Canino-Koning, R., Howe, A., Tiedje, J. M., & Brown, C. T. (2012). Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. In *Proceedings of the National Academy of Sciences* (Vol. 109, Issue 33, pp. 13272–13277). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1121464109>

Pernil, R., & Schleiff, E. (2019). Metalloproteins in the Biology of Heterocysts. In *Life* (Vol. 9, Issue 2, p. 32). MDPI AG. <https://doi.org/10.3390/life9020032>

Petersen, W., & Short, C. A. (1918). On the Relation of the Serum Ereptase (Peptidase) Titer to the Clinical Course in Pneumonia. In *Journal of Infectious Diseases* (Vol. 22, Issue 2, pp. 147–153). Oxford University Press (OUP). <https://doi.org/10.1093/infdis/22.2.147>

Pevzner, P. A., Tang, H., & Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. In *Proceedings of the National Academy of Sciences* (Vol. 98, Issue 17, pp. 9748–9753). *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.171285098>

Pontel, L. B., Checa, S. K., & Soncini, F. C. (2015). Bacterial Copper Resistance and Virulence. In *Bacteria-Metal Interactions* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-18570-5_1

Pos, K. M. (2009). Drug transport mechanism of the AcrB efflux pump. In *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* (Vol. 1794, Issue 5, pp. 782–793). Elsevier BV. <https://doi.org/10.1016/j.bbapap.2008.12.015>

Pathak, A. P., & Rathod, M. G. (2017). EXPLORATION OF A HOT SPRING FOR THERMOSTABLE PROTEASE PRODUCERS. In *Journal of Microbiology, Biotechnology and Food Sciences* (Vol. 7, Issue 2, pp. 101–109). Slovak University of Agriculture in Nitra. <https://doi.org/10.15414/jmbfs.2017.7.2.101-109>

Prosecka, J., Orlov, A. V., Fantin, Y. S., Zinchenko, V. V., Babykin, M. M., & Tichy, M. (2009). A novel ATP-binding cassette transporter is responsible for resistance to viologen herbicides in the cyanobacterium *Synechocystis* sp. PCC 6803. In *FEBS Journal* (Vol. 276, Issue 15, pp. 4001–4011). Wiley. <https://doi.org/10.1111/j.1742-4658.2009.07109.x>

Puig, S., Rees, E. M., & Thiele, D. J. (2002). The ABCDs of Periplasmic Copper Trafficking. In *Structure* (Vol. 10, Issue 10, pp. 1292–1295). Elsevier BV. [https://doi.org/10.1016/s0969-2126\(02\)00863-8](https://doi.org/10.1016/s0969-2126(02)00863-8)

Pumbwe, L. (2004). Expression of the efflux pump genes *cmeB*, *cmeF* and the porin gene *porA* in multiple-antibiotic-resistant *Campylobacter jejuni*. In *Journal of Antimicrobial Chemotherapy* (Vol. 54, Issue 2, pp. 341–347). Oxford University Press (OUP). <https://doi.org/10.1093/jac/dkh331>

Pushpanathan, M., Jayashree, S., Gunasekaran, P., & Rajendhran, J. (2014). Microbial Bioremediation. In *Microbial Biodegradation and Bioremediation* (pp. 407–419). Elsevier. <https://doi.org/10.1016/b978-0-12-800021-2.00017-0>

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. In *BMC Genomics* (Vol. 13, Issue 1, p. 341). Springer Science and Business Media LLC. <https://doi.org/10.1186/1471-2164-13-341>

Rademacher, C., & Masepohl, B. (2012). Copper-responsive gene regulation in bacteria. In *Microbiology* (Vol. 158, Issue 10, pp. 2451–2464). Microbiology Society. <https://doi.org/10.1099/mic.0.058487-0>

Ran, L.-Y., Su, H.-N., Zhou, M.-Y., Wang, L., Chen, X.-L., Xie, B.-B., Song, X.-Y., Shi, M., Qin, Q.-L., Pang, X., Zhou, B.-C., Zhang, Y.-Z., & Zhang, X.-Y. (2014). Characterization of a Novel Subtilisin-like Protease Myroicolsin from Deep Sea Bacterium *Myroides profundus* D25 and Molecular Insight into Its Collagenolytic Mechanism. In *Journal of Biological Chemistry* (Vol. 289, Issue 9, pp. 6041–6053). Elsevier BV. <https://doi.org/10.1074/jbc.m113.513861>

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. In *Biochemical and Biophysical Research Communications* (Vol. 469, Issue 4, pp. 967–977). Elsevier BV. <https://doi.org/10.1016/j.bbrc.2015.12.083>

Rawlings, N. D. (2013). Protease Families, Evolution and Mechanism of Action. In *Proteases: Structure and Function* (pp. 1–36). Springer Vienna. https://doi.org/10.1007/978-3-7091-0885-7_1

Rawlings, N. D., & Barrett, A. J. (2013). Introduction. In *Handbook of Proteolytic Enzymes* (pp. 325–370). Elsevier. <https://doi.org/10.1016/b978-0-12-382219-2.00077-6>

Rawlings, N. D., & Bateman, A. (2019). Origins of peptidases. In *Biochimie* (Vol. 166, pp. 4–18). Elsevier BV. <https://doi.org/10.1016/j.biochi.2019.07.026>

Reysenbach, A.-L., Huber, R., Stetter, K. O., Ishii, M., Kawasumi, T., Igarashi, Y., Eder, W., L'Haridon, S., & Jeanthon, C. (2001). Phylum BI. Aquificae phy. nov. In *Bergey's Manual® of Systematic Bacteriology* (pp. 359–367). Springer New York. https://doi.org/10.1007/978-0-387-21609-6_18

Reysenbach, A. L., Seitzinger, S., Kirshtein, J., & McLaughlin, E. (1999). Molecular Constraints on a High-Temperature Evolution of Early Life. In *The Biological Bulletin* (Vol. 196, Issue 3, pp. 367–372). University of Chicago Press. <https://doi.org/10.2307/1542972>

Richardson, E. J., & Watson, M. (2012). The automatic annotation of bacterial genomes. In *Briefings in Bioinformatics* (Vol. 14, Issue 1, pp. 1–12). Oxford University Press (OUP). <https://doi.org/10.1093/bib/bbs007>

Rife, C., Schwarzenbacher, R., McMullan, D., Abdubek, P., Ambing, E., Axelrod, H., Biorac, T., Canaves, J. M., Chiu, H.-J., Deacon, A. M., DiDonato, M., Elsliger, M.-A., Godzik, A., Grittini, C., Grzechnik, S. K., Hale, J., Hampton, E., Han, G. W., Haugen, J., ... Wilson, I. A. (2005). Crystal structure of a putative modulator of DNA gyrase (pmbA) from *Thermotoga maritima* at 1.95 Å resolution reveals a new fold. In *Proteins: Structure, Function, and Bioinformatics* (Vol. 61, Issue 2, pp. 444–448). Wiley. <https://doi.org/10.1002/prot.20468>

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N. N., Anderson, I. J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B. K., Gies, E. A., Dodsworth, J. A., Hedlund, B. P., Tsiamis, G., Sievert, S. M., Liu, W.-T., Eisen, J. A., Hallam, S. J., Kyrpides, N. C., Stepanauskas, R., ... Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. In *Nature* (Vol. 499, Issue 7459, pp. 431–437). Springer Science and Business Media LLC. <https://doi.org/10.1038/nature12352>

Roberts, W., 1880. On the Digestive Ferments and on the Preparation and Use of Artificially Digested Food. (1880). In *Nature* (Vol. 23, Issue 582, pp. 169–169). Springer Science and Business Media LLC. <https://doi.org/10.1038/023169a0>

Rodionov, D. A., Gelfand, M. S., Mironov, A. A., & Rakhmaninova, A. B. (2001). Comparative approach to analysis of regulation in complete genomes: Multidrug resistance systems in gamma-proteobacteria. *Journal of Molecular Microbiology and Biotechnology*, 3(2), 319-324.

Rodríguez-Sáinz, M. C., Hernández-Chico, C., & Moreno, F. (1990). Molecular characterization of pmbA, an *Escherichia coli* chromosomal gene required for the production of the antibiotic peptide MccB17. In *Molecular Microbiology* (Vol. 4, Issue 11, pp. 1921–1932). Wiley. <https://doi.org/10.1111/j.1365-2958.1990.tb02041.x>

Ross, K. A., Feazel, L. M., Robertson, C. E., Fathepure, B. Z., Wright, K. E., Turk-MacLeod, R. M., Chan, M. M., Held, N. L., Spear, J. R., & Pace, N. R. (2012). Phototrophic Phylotypes Dominate Mesothermal Microbial Mats Associated with Hot Springs in Yellowstone National Park. In *Microbial Ecology* (Vol. 64, Issue 1, pp. 162–170). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00248-012-001>

Rothschild, L. J., & Mancinelli, R. L. (2001). Life in extreme environments. In *Nature* (Vol. 409, Issue 6823, pp. 1092–1101). Springer Science and Business Media LLC. <https://doi.org/10.1038/35059215>

Roumpeka, D. D., Wallace, R. J., Escalettes, F., Fotheringham, I., & Watson, M. (2017). A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. In *Frontiers in Genetics* (Vol. 8). Frontiers Media SA. <https://doi.org/10.3389/fgene.2017.00023>

Runyen-Janecky, L. J., Reeves, S. A., Gonzales, E. G., & Payne, S. M. (2003). Contribution of the *Shigella flexneri* Sit, Iuc, and Feo Iron Acquisition Systems to Iron Acquisition In Vitro and in Cultured Cells. In *Infection and Immunity* (Vol. 71, Issue 4, pp. 1919–1928). American Society for Microbiology. <https://doi.org/10.1128/iai.71.4.1919-1928.2003>

Russell, R. J., Scott, C., Jackson, C. J., Pandey, R., Pandey, G., Taylor, M. C., Coppin, C. W., Liu, J.-W., & Oakeshott, J. G. (2011). The evolution of new enzyme function: lessons from xenobiotic metabolizing bacteria versus insecticide-resistant insects. In *Evolutionary Applications* (Vol. 4, Issue 2, pp. 225–248). Wiley. <https://doi.org/10.1111/j.1752-4571.2010.00175.x>

Safi, J., Awad, Y., & El-Nahal, Y. (2014). Bioremediation of Diuron in Soil Environment: Influence of Cyanobacterial Mat. In *American Journal of Plant Sciences* (Vol. 05, Issue 08, pp. 1081–1089). Scientific Research Publishing, Inc. <https://doi.org/10.4236/ajps.2014.58120>

Saggu, S. K., Jha, G., & Mishra, P. C. (2019). Enzymatic Degradation of Biofilm by Metalloprotease From *Microbacterium* sp. SKS10. In *Frontiers in Bioengineering and Biotechnology* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fbioe.2019.00192>

Sahay, H., Yadav, A. N., Singh, A. K., Singh, S., Kaushik, R., & Saxena, A. K. (2017). Hot springs of Indian Himalayas: potential sources of microbial diversity and thermostable hydrolytic enzymes. In *3 Biotech* (Vol. 7, Issue 2). Springer Science and Business Media LLC. <https://doi.org/10.1007/s13205-017-0762-1>

Saini, J., Kaur, R., & Sharma, P. K. (2017). Metagenomic Insights into Microbial Diversity and Metabolic Potential of Hot Spring Ecosystems. In *Mining of Microbial Wealth and MetaGenomics* (pp. 143–159). Springer Singapore. https://doi.org/10.1007/978-981-10-5708-3_9

Sánchez-Calderón, L., Chávez-Avilés, M., Gómez, B., & Díaz, C. (2019). Análisis estructural de la superóxido dismutasa ChrC de *Ochrobactrum tritici*. *Visum Mundi*. 3, 2.

Sánchez-Sutil, M. C., Pérez, J., Gómez-Santos, N., Shimkets, L. J., Moraleda-Muñoz, A., & Muñoz-Dorado, J. (2013). The *Myxococcus xanthus* Two-Component System CorSR Regulates Expression of a Gene Cluster Involved in Maintaining Copper Tolerance during Growth and Development. In E. Cascales (Ed.), *PLoS ONE* (Vol. 8, Issue 7, p. e68240). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0068240>

Saraithong, P., Goetting-Minesky, M. P., Durbin, P. M., Olson, S. W., Gherardini, F. C., & Fenno, J. C. (2020). Roles of TroA and TroR in Metalloregulated Growth and Gene Expression in *Treponema denticola*. In T. J. Silhavy (Ed.), *Journal of Bacteriology* (Vol. 202, Issue 7). American Society for Microbiology. <https://doi.org/10.1128/jb.00770-19>

Sato, M., Suda, M., Okuma, J., Kato, T., Hirose, Y., Nishimura, A., Kondo, Y., & Shibata, D. (2017). Isolation of highly thermostable β -xylosidases from a hot spring soil microbial community using a metagenomic approach. In *DNA Research* (Vol. 24, Issue 6, pp. 649–656). Oxford University Press (OUP). <https://doi.org/10.1093/dnares/dsx032>

Savitha, K., & S. Raman. (2012). Isolation, Characterisation and Enhanced Amylase Activity of a Chlorpyrifos Degrading Bacterial Strain, *Bacillus stearothermophilus*. *Nature Environment and Pollution Technology*. 11:37-40

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., & Sharma, V. K. (2017). Metagenomic Analysis of Hot Springs in Central India Reveals Hydrocarbon Degrading Thermophiles and Pathways Essential for Survival in Extreme Environments. In *Frontiers in Microbiology* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2016.02123>

Schwartz, D., Berger, S., Heinzemann, E., Muschko, K., Welzel, K., & Wohlleben, W. (2004). Biosynthetic Gene Cluster of the Herbicide Phosphinothricin Tripeptide from *Streptomyces viridochromogenes* Tü494. In *Applied and Environmental Microbiology* (Vol. 70, Issue 12, pp. 7093–7102). American Society for Microbiology. <https://doi.org/10.1128/aem.70.12.7093-7102.2004>

Sczyrba, A., Hofmann, P., Belmann, P., Koslicki, D., Janssen, S., Dröge, J., Gregor, I., Majda, S., Fiedler, J., Dahms, E., Bremges, A., Fritz, A., Garrido-Oter, R., Jørgensen, T. S., Shapiro, N., Blood, P. D., Gurevich, A., Bai, Y., Turaev, D., ... McHardy, A. C. (2017). Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. In *Nature Methods* (Vol. 14, Issue 11, pp. 1063–1071). Springer Science and Business Media LLC. <https://doi.org/10.1038/nmeth.4458>

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. In *Nature Methods* (Vol. 9, Issue 8, pp. 811–814). Springer Science and Business Media LLC. <https://doi.org/10.1038/nmeth.2066>

Segretin, A. B., Lima, M. A., Willis Poratti, G., Urbietta, M. S., Bernardelli, C., Giaveno, A., & Donati, E. (2016). Assessment of Microbial Patagonian Communities for Using in Heavy Metal Bioremediation. In *Biology and Biotechnology of Patagonian Microorganisms* (pp. 71–90). Springer International Publishing. https://doi.org/10.1007/978-3-319-42801-7_5

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. In *Bioinformatics* (Vol. 30, Issue 14, pp. 2068–2069). Oxford University Press (OUP). <https://doi.org/10.1093/bioinformatics/btu153>

Servicio Fitosanitario del Estado MAG. (2021). Lista de productos prohibidos. Tomado de: https://www.sfe.go.cr/DocsStatusRegistro/Listado_de_prohibidos.pdf

Shaiber, A., Willis, A. D., Delmont, T. O., Roux, S., Chen, L.-X., Schmid, A. C., Yousef, M., Watson, A. R., Lolans, K., Esen, Ö. C., Lee, S. T. M., Downey, N., Morrison, H. G., Dewhirst, F. E., Mark Welch, J. L., & Eren, A. M. (2020). Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. In *Genome Biology* (Vol. 21, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s13059-020-02195-w>

Shafeeq, S., Yesilkaya, H., Kloosterman, T. G., Narayanan, G., Wandel, M., Andrew, P. W., Kuipers, O. P., & Morrissey, J. A. (2011). The cop operon is required for copper homeostasis and contributes to virulence in *Streptococcus pneumoniae*. In *Molecular Microbiology* (Vol. 81, Issue 5, pp. 1255–1270). Wiley. <https://doi.org/10.1111/j.1365-2958.2011.07758.x>

Shahcheraghi, F., Minato, Y., Chen, J., Mizushima, T., Ogawa, W., Kuroda, T., & Tsuchiya, T. (2007). Molecular Cloning and Characterization of a Multidrug Efflux Pump, SmfY, from *Serratia marcescens*. In *Biological and Pharmaceutical Bulletin* (Vol. 30, Issue 4, pp. 798–800). Pharmaceutical Society of Japan. <https://doi.org/10.1248/bpb.30.79>

Shakhatreh, M. A. K., Jacob, J. H., Hussein, E. I., Masadeh, M. M., Obeidat, S. M., Juhmani, A. F., & Abd Al-razaq, M. A. (2017). Microbiological analysis, antimicrobial activity, and heavy-metals content of Jordanian Ma'in hot-springs water. In *Journal of Infection and Public Health* (Vol. 10, Issue 6, pp. 789–793). Elsevier BV. <https://doi.org/10.1016/j.jiph.2017.01.01>

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. In *Frontiers in Plant Science* (Vol. 5). Frontiers Media SA. <https://doi.org/10.3389/fpls.2014.00209>

Sheikh Abdul Hamid, N., Zen, H. B., Tein, O. B., Halifah, Y. M., Saari, N., & Bakar, F. A. (2003). Screening and identification of extracellular lipase-producing thermophilic bacteria from a Malaysian hot spring. In *World Journal of Microbiology and Biotechnology* (Vol. 19, Issue 9, pp. 961–968). Springer Science and Business Media LLC. <https://doi.org/10.1023/b:wibi.0000007330.84569.39>

Simpson, C. L., & Russell, R. R. B. (1998). Intracellular α -Amylase of *Streptococcus mutans*. In *Journal of Bacteriology* (Vol. 180, Issue 17, pp. 4711–4717). American Society for Microbiology. <https://doi.org/10.1128/jb.180.17.4711-4717.1998>

Singh, B., Kaur, J., & Singh, K. (2013). Microbial degradation of an organophosphate pesticide, malathion. In *Critical Reviews in Microbiology* (Vol. 40, Issue 2, pp. 146–154). Informa UK Limited. <https://doi.org/10.3109/1040841x.2013.763222>

Singh, B. K., & Naidu, R. (2012). Cleaning contaminated environment: a growing challenge. In *Biodegradation* (Vol. 23, Issue 6, pp. 785–786). Springer Science and Business Media LLC. <https://doi.org/10.1007/s10532-012-9590-5>

Singh, B. K., & Walker, A. (2006). Microbial degradation of organophosphorus compounds. In *FEMS Microbiology Reviews* (Vol. 30, Issue 3, pp. 428–471). Oxford University Press (OUP). <https://doi.org/10.1111/j.1574-6976.2006.00018.x>

Singla, P., & Bhardwaj, R. D. (2019). Enzyme promiscuity – A light on the “darker” side of enzyme specificity. In *Biocatalysis and Biotransformation* (Vol. 38, Issue 2, pp. 81–92). Informa UK Limited. <https://doi.org/10.1080/10242422.2019.1696779>

Sittenfeld, A., Mora, M., Ortega, J. M., Albertazzi, F., Cordero, A., Roncel, M., SÁjnchez, E., Vargas, M., FernÁjndez, M., Weckesser, J., & Serrano, A. (2002). Characterization of a photosynthetic *Euglena* strain isolated from an acidic hot mud pool of a volcanic area of Costa Rica¹. In *FEMS Microbiology Ecology* (Vol. 42, Issue 1, pp. 151–161). Oxford University Press (OUP). <https://doi.org/10.1111/j.1574-6941.2002.tb01004.x>

Sittenfeld, Ana, Vargas, Maribelle, Sánchez, Ethel, Mora, Marielos, & Serrano, Aurelio. (2004). Una nueva especie de *Euglena* (Euglenozoa: Euglenales) aislada de ambientes extremófilos en las Pailas de Barro del Volcán Rincón de la Vieja, Costa Rica. *Revista de Biología Tropical*, 52(1), 27-30.

Sivaramakrishnan, S., Gangadharan, D., Nampoothiri, K.M., Soccol, C.R. (2006) α -Amylases from microbial sources—An overview on recent developments. *Food Technology and Biotechnology*, 44, 173-184.

Slade, D., & Radman, M. (2011). Oxidative Stress Resistance in *Deinococcus radiodurans*. In *Microbiology and Molecular Biology Reviews* (Vol. 75, Issue 1, pp. 133–191). American Society for Microbiology. <https://doi.org/10.1128/mmbr.00015-10>

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” In *Proceedings of the National Academy of Sciences* (Vol. 103, Issue 32, pp.

12115–12120). Proceedings of the National Academy of Sciences. <https://doi.org/10.1073/pnas.0605127103>

Sommer, M. J., & Salzberg, S. L. (2021). Balrog: A universal protein model for prokaryotic gene prediction. In C. A. Ouzounis (Ed.), *PLOS Computational Biology* (Vol. 17, Issue 2, p. e1008727). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pcbi.1008727>

Srinivasan, V. B., Rajamohan, G., & Gebreyes, W. A. (2009). Role of AbeS, a Novel Efflux Pump of the SMR Family of Transporters, in Resistance to Antimicrobial Agents in *Acinetobacter baumannii*. In *Antimicrobial Agents and Chemotherapy* (Vol. 53, Issue 12, pp. 5312–5316). American Society for Microbiology. <https://doi.org/10.1128/aac.00748-09>

Srivastava, S., Dong, H., & Briggs, B. R. (2020). The Effect of Spring Water Geochemistry on Copper Proteins in Tengchong Hot Springs, China. In H. Atomi (Ed.), *Applied and Environmental Microbiology* (Vol. 86, Issue 13). American Society for Microbiology. <https://doi.org/10.1128/aem.00581-20>

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. In *Nucleic Acids Research* (Vol. 6, Issue 7, pp. 2601–2610). Oxford University Press (OUP). <https://doi.org/10.1093/nar/6.7.2601>

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. In *Nucleic Acids Research* (Vol. 8, Issue 16, pp. 3673–3694). Oxford University Press (OUP). <https://doi.org/10.1093/nar/8.16.3673>

Stehr, F., Kretschmar, M., Kröger, C., Hube, B., & Schäfer, W. (2003). Microbial lipases as virulence factors. In *Journal of Molecular Catalysis B: Enzymatic* (Vol. 22, Issues 5–6, pp. 347–355). Elsevier BV. [https://doi.org/10.1016/s1381-1177\(03\)00049-3](https://doi.org/10.1016/s1381-1177(03)00049-3)

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. In *Nature Biotechnology* (Vol. 35, Issue 11, pp. 1026–1028). Springer Science and Business Media LLC. <https://doi.org/10.1038/nbt.3988>

Stewart, L. C., Stucker, V. K., Stott, M. B., & de Ronde, C. E. J. (2018). Marine-influenced microbial communities inhabit terrestrial hot springs on a remote island volcano. In *Extremophiles* (Vol. 22, Issue 4, pp. 687–698). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-018-1029-4>

Stewart, R. D., Auffret, M. D., Warr, A., Wisner, A. H., Press, M. O., Langford, K. W., Liachko, I., Snelling, T. J., Dewhurst, R. J., Walker, A. W., Roehe, R., & Watson, M. (2018). Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. In *Nature Communications* (Vol. 9, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41467-018-03317-6>

Suhadolnik, M. L. S., Salgado, A. P. C., Scholte, L. L. S., Bleicher, L., Costa, P. S., Reis, M. P., Dias, M. F., Ávila, M. P., Barbosa, F. A. R., Chartone-Souza, E., & Nascimento, A. M. A. (2017). Novel arsenic-transforming bacteria and the diversity of their arsenic-related genes and enzymes arising from arsenic-polluted freshwater sediment. In *Scientific Reports* (Vol. 7, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-017-11548-8>

Suzuki, M., Cui, Z. J., Ishii, M., & Igarashi, Y. (2001). Nitrate respiratory metabolism in an obligately autotrophic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus* TK-6. In *Archives of Microbiology* (Vol. 175, Issue 1, pp. 75–78). Springer Science and Business Media LLC. <https://doi.org/10.1007/s0020300>

Takimura, Y., Saito, K., Okuda, M., Kageyama, Y., Saeki, K., Ozaki, K., Ito, S., & Kobayashi, T. (2007). Alkaliphilic *Bacillus* sp. strain KSM-LD1 contains a record number of subtilisin-like serine proteases genes. In *Applied Microbiology and Biotechnology* (Vol. 76, Issue 2, pp. 395–405). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00253-007-1022-9>

Tamames, J., & Puente-Sánchez, F. (2018). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. Cold Spring Harbor Laboratory. <https://doi.org/10.1101/347559>

Tamehiro, N., Okamoto-Hosoya, Y., Okamoto, S., Ubukata, M., Hamada, M., Naganawa, H., & Ochi, K. (2002). Bacilysocin, a Novel Phospholipid Antibiotic Produced by *Bacillus subtilis* 168. In *Antimicrobial Agents and Chemotherapy* (Vol. 46, Issue 2, pp. 315–320). American Society for Microbiology. <https://doi.org/10.1128/aac.46.2.315-320.2002>

Tang, W., Lan, D., Zhao, Z., Li, S., Li, X., & Wang, Y. (2019). A Thermostable Monoacylglycerol Lipase from Marine *Geobacillus* sp. 12AMOR1: Biochemical Characterization and Mutagenesis Study. In *International Journal of Molecular Sciences* (Vol. 20, Issue 3, p. 780). MDPI AG. <https://doi.org/10.3390/ijms20030780>

Tank, M., Thiel, V., Ward, D. M., & Bryant, D. A. (2017). A Panoply of Phototrophs: An Overview of the Thermophilic Chlorophototrophs of the Microbial Mats of Alkaline Siliceous Hot Springs in Yellowstone National Park, WY, USA. In *Modern Topics in the Phototrophic Prokaryotes* (pp. 87–137). Springer International Publishing. https://doi.org/10.1007/978-3-319-46261-5_3

Tatusov, R. L. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. In *Nucleic Acids Research* (Vol. 28, Issue 1, pp. 33–36). Oxford University Press (OUP). <https://doi.org/10.1093/nar/28.1>.

Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A Genomic Perspective on Protein Families. In *Science* (Vol. 278, Issue 5338, pp. 631–637). American Association for the Advancement of Science (AAAS). <https://doi.org/10.1126/science.278.5338.631>

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., Segovia, B. T., Lansac-Toha, F. A., Lemke, M., DeSalle, R., Mason, C. E., & Brugler, M. R. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. In *Scientific Reports* (Vol. 7, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-017-06665-3>

Thiel, V., Wood, J. M., Olsen, M. T., Tank, M., Klatt, C. G., Ward, D. M., & Bryant, D. A. (2016). The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and

Metagenomic Sequencing. In *Frontiers in Microbiology* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2016.00919>

Ting, Y. T., Harris, P. W. R., Batot, G., Brimble, M. A., Baker, E. N., & Young, P. G. (2016). Peptide binding to a bacterial signal peptidase visualized by peptide tethering and carrier-driven crystallization. In *IUCrJ* (Vol. 3, Issue 1, pp. 10–19). International Union of Crystallography (IUCr). <https://doi.org/10.1107/s2052252515019971>

Tully, B. J., Graham, E. D., & Heidelberg, J. F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. In *Scientific Data* (Vol. 5, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/sdata.2017.203>

Turner, P., Mamo, G., & Karlsson, E. N. (2007). Potential and utilization of thermophiles and thermostable enzymes in biorefining. In *Microbial Cell Factories* (Vol. 6, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/1475-2859-6-9>

Ueno, K., Ibarra, M., & Gojobori, T. (2016). Structural adaption of extremophile proteins to the environments with special reference to hydrophobic networks. In *Ecological Genetics and Genomics* (Vol. 1, pp. 1–5). Elsevier BV. <https://doi.org/10.1016/j.egg.2015.10.001>

Ufarté, L., Potocki-Veronese, G., & Laville, Á. (2015). Discovery of new protein families and functions: new challenges in functional metagenomics for biotechnologies and microbial ecology. In *Frontiers in Microbiology* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2015.00563>

Urbietta, M. S., González-Toril, E., Bazán, Á. A., Giaveno, M. A., & Donati, E. (2015). Comparison of the microbial communities of hot springs waters and the microbial biofilms in the acidic geothermal area of Copahue (Neuquén, Argentina). In *Extremophiles* (Vol. 19, Issue 2, pp. 437–450). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-015-0729-2>

Uribe-Lorío, L., Brenes-Guillén, L., Hernández-Ascencio, W., Mora-Amador, R., González, G., Ramírez-Umaña, C. J., Díez, B., & Pedrós-Alió, C. (2019). The influence of temperature

and pH on bacterial community composition of microbial mats in hot springs from Costa Rica. In *MicrobiologyOpen* (Vol. 8, Issue 10). Wiley. <https://doi.org/10.1002/mbo3.893>

Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. In *Microbiome* (Vol. 6, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s40168-018-0541-1>

Valasatava, Y., Andreini, C., & Rosato, A. (2015). Hidden relationships between metalloproteins unveiled by structural comparison of their metal sites. In *Scientific Reports* (Vol. 5, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/srep09486>

van der Walt, A. J., van Goethem, M. W., Ramond, J.-B., Makhalanyane, T. P., Reva, O., & Cowan, D. A. (2017). Assembling metagenomes, one community at a time. In *BMC Genomics* (Vol. 18, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12864-017-3918-9>

Vargas, A. G., & Alvarado, G. E. (2007). Los primeros catálogos de fuentes termales en Costa Rica realizados por el naturalista Alemán Alexander Von Frantzius en 1862 y 1873. In *Revista Geológica de América Central* (Issue 37). Universidad de Costa Rica. <https://doi.org/10.15517/rgac.v0i37.4202>

Vargas, M., Kashfi, K., Blunt-Harris, E. L., & Lovley, D. R. (1998). Microbiological evidence for Fe(III) reduction on early Earth. In *Nature* (Vol. 395, Issue 6697, pp. 65–67). Springer Science and Business Media LLC. <https://doi.org/10.1038/25720>

Vines, S.H. (1904). The Proteases of Plants. In *Annals of Botany: Vol. os-18* (Issue 2, pp. 289–317). Oxford University Press (OUP). <https://doi.org/10.1093/oxfordjournals.aob.a088963>

Vishnivetskaya, T. A., Hamilton-Brehm, S. D., Podar, M., Mosher, J. J., Palumbo, A. V., Phelps, T. J., Keller, M., & Elkins, J. G. (2014). Community Analysis of Plant Biomass-Degrading Microorganisms from Obsidian Pool, Yellowstone National Park. In *Microbial*

Ecology (Vol. 69, Issue 2, pp. 333–345). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00248-014-0500-8>

Vollmers, J., Wiegand, S., & Kaster, A.K. (2017). Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! In F. Rodriguez-Valera (Ed.), PLOS ONE (Vol. 12, Issue 1, p. e0169662). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0169662>

Vu, K. B., Kaminski, M. D. & Nuñez, L. (2003). Review of Arsenic Removal Technologies for Contaminated Groundwaters. Tomado de <https://publications.anl.gov/anlpubs/2003/05/46522.pdf>

Vurro, M., Boari, A., Pilgeram, A. L., & Sands, D. C. (2006). Exogenous amino acids inhibit seed germination and tubercle formation by *Orobanche ramosa* (Broomrape): Potential application for management of parasitic weeds. In *Biological Control* (Vol. 36, Issue 2, pp. 258–265). Elsevier BV. <https://doi.org/10.1016/j.biocontrol.2005.09.017>

Wagner, R., Aigner, H., & Funk, C. (2012). FtsH proteases located in the plant chloroplast. In *Physiologia Plantarum* (Vol. 145, Issue 1, pp. 203–214). Wiley. <https://doi.org/10.1111/j.1399-3054.2011.01548.x>

Wagner, V., Antunes, P. M., Irvine, M., & Nelson, C. R. (2016). Herbicide usage for invasive non-native plant management in wildland areas of North America. In J. Finn (Ed.), *Journal of Applied Ecology* (Vol. 54, Issue 1, pp. 198–204). Wiley. <https://doi.org/10.1111/1365-2664.12711>

Waldron, K. J., & Robinson, N. J. (2009). How do bacterial cells ensure that metalloproteins get the correct metal? In *Nature Reviews Microbiology* (Vol. 7, Issue 1, pp. 25–35). Springer Science and Business Media LLC. <https://doi.org/10.1038/nrmicro2057>

Walter, M. R. (1972). A hot spring analog for the depositional environment of Precambrian iron formations of the Lake Superior Region. In *Economic Geology* (Vol. 67, Issue 7, pp. 965–972). Society of Economic Geologists. <https://doi.org/10.2113/gsecongeo.67.7.965>

Wang, Y., Srivastava, K. C., Shen, G.-J., & Wang, H. Y. (1995). Thermostable alkaline lipase from a newly isolated thermophilic *Bacillus*, strain A30-1 (ATCC 53841). In *Journal of Fermentation and Bioengineering* (Vol. 79, Issue 5, pp. 433–438). Elsevier BV. [https://doi.org/10.1016/0922-338x\(95\)912](https://doi.org/10.1016/0922-338x(95)912)

Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G., Zhang, C., Song, Z., Zhang, Y., Ren, H., Zhang, J., & Zhang, L. (2013). Control of Temperature on Microbial Community Structure in Hot Springs of the Tibetan Plateau. In J. Neufeld (Ed.), *PLoS ONE* (Vol. 8, Issue 5, p. e62901). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0062901>

Ward, S. K., Abomoelak, B., Hoyer, E. A., Steinberg, H., & Talaat, A. M. (2010). CtpV: a putative copper exporter required for full virulence of *Mycobacterium tuberculosis*. In *Molecular Microbiology* (Vol. 77, Issue 5, pp. 1096–1110). Wiley. <https://doi.org/10.1111/j.1365-2958.2010.07273.x>

Ward, D. E., Shockley, K. R., Chang, L. S., Levy, R. D., Michel, J. K., Connors, S. B., & Kelly, R. M. (2002). Proteolysis in hyperthermophilic microorganisms. In *Archaea* (Vol. 1, Issue 1, pp. 63–74). Hindawi Limited. <https://doi.org/10.1155/2002/503191>

Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E., & Fischer, W. W. (2018). Evolution of Phototrophy in the Chloroflexi Phylum Driven by Horizontal Gene Transfer. In *Frontiers in Microbiology* (Vol. 9). Frontiers Media SA. <https://doi.org/10.3389/fmicb.2018.00260>

Wein, T., Hülter, N. F., Mizrahi, I., & Dagan, T. (2019). Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. In *Nature Communications* (Vol. 10, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41467-019-10600-7>

Wong, E., Yusof, M., Mansor, M., Anbazhagan, D., Ong, S., & Sekaran, S. (2009). Disruption of *adeB* gene has a greater effect on resistance to meropenems than *adeA* gene in *Acinetobacter* spp. Isolated from University Malaya Medical Centre. *Singapore medical journal*, 50(8), 822-826.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. In *Genome Biology* (Vol. 15, Issue 3, p. R46). Springer Science and Business Media LLC. <https://doi.org/10.1186/gb-2014-15-3-r46>

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. In *Genome Biology* (Vol. 20, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s13059-019-1891-0>

Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., Hettich, R. L., Lipton, M. S., Williams, K. H., Long, P. E., & Banfield, J. F. (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. In *Science* (Vol. 337, Issue 6102, pp. 1661–1665). American Association for the Advancement of Science (AAAS). <https://doi.org/10.1126/science.1224041>

Xiao, Y., Lu, Q., Yi, X., Zhong, G., & Liu, J. (2020). Synergistic Degradation of Pyrethroids by the Quorum Sensing-Regulated Carboxylesterase of *Bacillus subtilis* BSF01. In *Frontiers in Bioengineering and Biotechnology* (Vol. 8). Frontiers Media SA. <https://doi.org/10.3389/fbioe.2020.00889>

Xu, C. F., Bilya, S. R., & Xu, W. (2019). *adeABC* efflux gene in *Acinetobacter baumannii*. In *New Microbes and New Infections* (Vol. 30, p. 100549). Elsevier BV. <https://doi.org/10.1016/j.nmni.2019.100549>

Yan, B.-Q., Chen, X.-L., Hou, X.-Y., He, H., Zhou, B.-C., & Zhang, Y.-Z. (2009). Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains. In *Extremophiles* (Vol. 13, Issue 4, pp. 725–733). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00792-009-0263-1>

Yan, W., Li, F., Wang, L., Zhu, Y., Dong, Z., & Bai, L. (2017). Discovery and characterization of a novel lipase with transesterification activity from hot spring metagenomic library. In

Biotechnology Reports (Vol. 14, pp. 27–33). Elsevier BV. <https://doi.org/10.1016/j.btre.2016.12.007>

Yang, J., Li, J., Mai, Z., Tian, X., & Zhang, S. (2013). Purification, characterization, and gene cloning of a cold-adapted thermolysin-like protease from *Halobacillus* sp. SCSIO 20089. In *Journal of Bioscience and Bioengineering* (Vol. 115, Issue 6, pp. 628–632). Elsevier BV. <https://doi.org/10.1016/j.jbiosc.2012.12.013>

Yang, K. K., Wu, Z., & Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. In *Nature Methods* (Vol. 16, Issue 8, pp. 687–694). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41592-019-0496-6>

Yu, F. B., Blainey, P. C., Schulz, F., Woyke, T., Horowitz, M. A., & Quake, S. R. (2017). Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. In *eLife* (Vol. 6). eLife Sciences Publications, Ltd. <https://doi.org/10.7554/elife.26580>

Yue, Y., Huang, H., Qi, Z., Dou, H.-M., Liu, X.-Y., Han, T.-F., Chen, Y., Song, X.-J., Zhang, Y.-H., & Tu, J. (2020). Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets. In *BMC Bioinformatics* (Vol. 21, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12859-020-03667-3>

Zarafeta, D., Moschidi, D., Ladoukakis, E., Gavrillov, S., Chrysina, E. D., Chatziioannou, A., Kublanov, I., Skretas, G., & Kollis, F. N. (2016). Metagenomic mining for thermostable esterolytic enzymes uncovers a new family of bacterial esterases. In *Scientific Reports* (Vol. 6, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/srep38886>

Zhang, Y., & Zheng, J. (2020). Bioinformatics of Metalloproteins and Metalloproteomes. In *Molecules* (Vol. 25, Issue 15, p. 3366). MDPI AG. <https://doi.org/10.3390/molecules25153366>

Zhang, C. L., Ye, Q., Huang, Z., Li, W., Chen, J., Song, Z., Zhao, W., Bagwell, C., Inskip, W. P., Ross, C., Gao, L., Wiegel, J., Romanek, C. S., Shock, E. L., & Hedlund, B. P. (2008). Global Occurrence of Archaeal *amoA* Genes in Terrestrial Hot Springs. In *Applied and*

Environmental Microbiology (Vol. 74, Issue 20, pp. 6417–6426). American Society for Microbiology. <https://doi.org/10.1128/aem.00843-08>

Zhang, J., Lan, W., Qiao, C., Jiang, H., Mulchandani, A., & Chen, W. (2004). Bioremediation of Organophosphorus Pesticides by Surface-Expressed Carboxylesterase from Mosquito on *Escherichia Coli*. In *Biotechnology Progress* (Vol. 20, Issue 5, pp. 1567–1571). Wiley. <https://doi.org/10.1021/bp049903c>

Zhang XQ, Zhang WJ, Wei BP, Xu XW, Zhu XF, Wu M. 2009. *Meiothermus cateniformans* sp. Zhang, X.-Q., Zhang, W.-J., Wei, B.-P., Xu, X.-W., Zhu, X.-F., & Wu, M. (2010). *Meiothermus cateniformans* sp. nov., a slightly thermophilic species from north-eastern China. In *International Journal of Systematic and Evolutionary Microbiology* (Vol. 60, Issue 4, pp. 840–844). Microbiology Society. <https://doi.org/10.1099/ijs.0.0>

Zhao, J., Liu, S., Gao, Y., Ma, M., Yan, X., Cheng, D., Wan, D., Zeng, Z., Yu, P., & Gong, D. (2021). Characterization of a novel lipase from *Bacillus licheniformis* NCU CS-5 for applications in detergent industry and biodegradation of 2,4-D butyl ester. In *International Journal of Biological Macromolecules* (Vol. 176, pp. 126–136). Elsevier BV. <https://doi.org/10.1016/j.ijbiomac.2021.01.214>

Zhou, M.-Y., Chen, X.-L., Zhao, H.-L., Dang, H.-Y., Luan, X.-W., Zhang, X.-Y., He, H.-L., Zhou, B.-C., & Zhang, Y.-Z. (2009). Diversity of Both the Cultivable Protease-Producing Bacteria and Their Extracellular Proteases in the Sediments of the South China Sea. In *Microbial Ecology* (Vol. 58, Issue 3, pp. 582–590). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00248-009-9506-z>

Zoetendal, E. G., Smith, A. H., Sundset, M. A., & Mackie, R. I. (2008). The BaeSR Two-Component Regulatory System Mediates Resistance to Condensed Tannins in *Escherichia coli*. In *Applied and Environmental Microbiology* (Vol. 74, Issue 2, pp. 535–539). American Society for Microbiology. <https://doi.org/10.1128/aem.02271-07>

Zuridah, H. (2011). Identification of lipase producing thermophilic bacteria from Malaysian hot springs. In *African Journal of Microbiology Research* (Vol. 5, Issue 21). Academic Journals. <https://doi.org/10.5897/ajmr11.777>

Anexos

Cuadro 1. Estadísticas generales de los datos crudos de los metagenomas utilizados en este estudio.

Metagenoma	Número de secuencias (single end)	Total de bases secuenciadas (Gpb)	Cobertura (#secuencias/Total de bases)	Tamaño (Gigabytes)	Tamaño de secuencias (bp)	Porcentaje de secuencias restantes si se deduplica ¹	Tamaño de secuencias	Porcentaje de secuencias restantes si se deduplica ¹
SRR5248290	80224270	24.1	33.29	14.3	149-151	12.07	100-151	15.09
SRR5248299	21303953	6.4	33.29	3.5	149-151	39.07	100-151	42.8
SRR5248365	42019147	12.6	33.35	7.3	149-151	16.29	100-151	23.56
SRR5248302	28094272	8.4	33.45	4.7	149-151	27.09	100-151	31.67
SRR5451033	140980559	18.2	77.46	6.2	30-116	60.72	100-116	71.02
SRR5451032	61953213	10.7	57.90	3.8	30-116	65.94	100-116	67.89
SRR5451031	10263362	2.2	46.65	690.3Mb	36-126	89.79	100-126	93.12
SRR10063240	49791620	14.9	33.42	5.9	149-151	85.42	100-151	93.15
SRR10063241	48525851	14.6	33.24	5.8	149-151	86.73	100-151	89.02
SRR10063242	43957771	13.2	33.30	5.3	149-151	89.9	100-151	91.12
SRR5650826	54274165	17.4	31.19	6.6	149-151	74.32	100-151	82.12
SRR5650808	33491695	10	33.49	3.4	149-151	33.12	100-151	38.5
SRR4030101	21096521	9.7	21.75	5.3	149-150	65.43	100-151	74.2
Salitral	74196897	14.98	49.53	14	100-102	51.34	100-101	58.1
Río Naranjo	67486140	14.98	45.05	13	100-102	36.99	100-101	43.1
La Luz	81033039	16.36	49.53	16	100-102	57.72	100-101	63.98
Miravalles	44691101	9	49.66	9	100-102	80.88	100-101	85.93
Las Lilas	44255965	8.9	49.73	8	100-102	40.22	100-101	45.75
Rocas Calientes	47658908	3.3	144.42	4.7 (Mb)	100-102	74.43	100-101	78.8

¹ Un alto nivel de duplicación indica algún tipo de sesgo de enriquecimiento (valores más bajos en estadística).

Cuadro 2. Resultados del ensamblaje de muestras de Chile, Nueva Zelanda y PNY obtenido a partir de los ensamblajes con Megahit

Estadística	SRR5248290	SRR5248299	SRR5248365	SRR5248302	SRR5451033	SRR5451032	SRR5451031	SRR10063240	SRR10063241	SRR10063242	SRR5650826	SRR5650808	SRR4030101
# contigs (>= 0 bp)	433	1478	888	1731	24987	15183	16388	77000	94457	75035	44554	28439	3650
# contigs (>= 1000 bp)	433	1478	888	1731	24987	15183	16388	77000	94457	75035	44554	28439	3650
# contigs (>= 5000 bp)	205	511	320	318	2946	2150	2531	10987	9514	13328	3366	3997	825
# contigs (>= 10000 bp)	138	374	156	162	1592	1224	1241	4982	5155	5981	1350	1487	470
# contigs (>= 25000 bp)	64	176	44	90	685	593	402	1920	2223	1499	373	396	144
# contigs (>= 50000 bp)	21	61	16	38	320	292	143	916	1210	592	142	138	38
Total length (>= 0 bp)	5382829	15106920	6805706	10467517	113471723	81888823	70313291	363523800	437160376	347893297	125576690	103487621	20408352
Total length (>= 1000 bp)	5382829	15106920	6805706	10467517	113471723	81888823	70313291	363523800	437160376	347893297	125576690	103487621	20408352
Total length (>= 5000 bp)	4875428	13228052	5444607	7557085	74682341	58134659	44447590	239830753	290259482	229760442	53200981	55457750	15096851
Total length (>= 10000 bp)	4401386	12254256	4299394	6524261	65336588	51642058	35483209	198804753	260180539	178378847	39542512	38381793	12595750
Total length (>= 25000 bp)	3228444	8977837	2605605	5342991	51334899	41822040	22887755	152641511	214847756	111535774	24844718	22199953	7456666
Total length (>= 50000 bp)	1751905	5003048	1572781	3400966	38442146	31163394	13985892	117895436	179153273	80948757	17064931	13506247	3845316
# contigs	433	1478	888	1731	24987	15183	16388	77000	94457	75035	44554	28439	3650
Largest contig	171926	172824	227851	184882	2643357	438562	627002	1124019	2048419	1719666	893503	310439	268933
Total length	5382829	15106920	6805706	10467517	113471723	81888823	70313291	363523800	437160376	347893297	125576690	103487621	20408352
N50	32123	32556	15772	26681	17930	26509	10248	13762	23451	10611	3508	5695	16445
N75	14266	14611	6163	4257	2652	3615	2674	3045	2616	3163	1649	2297	4613
L50	45	126	84	86	943	559	1209	3520	2378	5551	5697	3301	284
L75	108	298	259	383	5928	2929	4852	19523	20375	21214	19517	10741	869

Cuadro 3. Estadísticas generales de los MAGs obtenidos en este estudio

Metagenoma	MAG	Tamaño (bp)	Número de contigs	N50	%GC	Anvio		Check M	
						% Completitud	% Redundancia	% Completitud	% Redundancia
Miravalles	MAG_20	3918455	838	6732	62.1	78.9	9.9	62.5	8.7
Miravalles	MAG_2	4098370	337	342585	60.4	83.1	5.6	92.1	5.9
Miravalles	MAG_29	5838842	246	36536	56.1	97.2	4.2	94.9	5.6
Las Lilas	MAG_10	3124095	79	89606	62.0	98.6	4.2	96.6	3.6
Rocas Calientes	MAG_16	2597779	78	144431	59.2	83.5	3.6	93.0	3.7
Miravalles	MAG_79	4051236	71	86007	64.9	98.6	1.4	98.3	2.6
Las Lilas	MAG_5	1648098	36	127294	66.9	94.4	0.0	88.0	0.0
Salitral	MAG_21	2355759	887	3164	61.0	78.9	8.5	73.6	3.7
Salitral	MAG_14	3022322	224	18380	67.3	93.0	2.8	94.0	0.9
La Luz	MAG_8	2529723	185	34922	60.1	94.4	2.8	93.2	2.8
Las Lilas	MAG_13	3279848	86	88381	52.7	95.8	0.0	93.5	0.7
Río Naranjo	MAG_17	3913899	886	6277	44.2	100.0	1.4	94.8	2.9
Río Naranjo	MAG_16	2969208	780	4811	37.7	93.0	4.2	91.1	2.4
La Luz	MAG_52	3960085	306	25515	67.3	100.0	4.2	97.7	0.9
Miravalles	MAG_27	3612964	297	21405	56.6	95.8	0.0	95.1	5.1
Salitral	MAG_46	2387108	205	18720	38.8	85.9	1.4	89.3	1.5
Río Naranjo	MAG_26	3175387	139	35261	37.4	98.6	2.8	98.2	5.7
Salitral	MAG_47	3128688	133	99515	39.4	98.6	4.2	99.5	6.4
La Luz	MAG_56	3115527	96	50708	38.3	98.6	2.8	98.2	1.5

La Luz	MAG_62	2492654	94	39878	36.9	98.6	1.4	99.6	0.5
Río Naranjo	MAG_27	4796393	85	99177	36.6	93.0	0.0	96.3	0.6
La Luz	MAG_81	2528630	76	55626	47.6	84.5	1.4	84.2	1.1
Salitral	MAG_48	3155388	72	83804	38.4	98.6	2.8	98.3	1.1
Rocas Calientes	MAG_19	3455443	39	155203	55.5	97.1	0.0	98.5	1.0
Río Naranjo	MAG_20	2629224	33	138335	36.9	97.2	1.4	99.1	0.7
Las Lilas	MAG_11	1748052	98	27888	58.5	91.5	2.8	89.8	0.0
Miravalles	MAG_34	3653485	1753	2270	54.3	45.1	1.4	65.8	5.0
Salitral	MAG_2	4105636	1072	5085	48.9	97.2	8.5	90.5	3.9
Miravalles	MAG_69	3208546	1067	3678	61.3	83.1	8.5	86.7	6.4
Rocas Calientes	MAG_25	4330212	289	29913	65.3	97.8	2.2	94.9	0.9
Rocas Calientes	MAG_1	5426742	273	62937	59.9	99.3	6.5	98.2	6.9
Río Naranjo	MAG_12	4432444	264	27858	49.4	98.6	1.4	97.3	1.9
Salitral	MAG_10	5280726	237	59684	55.9	98.6	7.0	99.7	7.6
Salitral	MAG_23	4807199	231	33368	62.7	95.8	4.2	96.7	2.0
Salitral	MAG_7	4379901	228	89442	57.3	91.5	8.5	99.1	7.6
La Luz	MAG_53	2753947	209	26844	71.2	60.6	4.2	85.8	5.9
Rocas Calientes	MAG_12	3614319	172	36685	58.1	97.8	1.4	94.6	2.7
Salitral	MAG_25	5909535	125	100787	62.1	97.2	7.0	99.1	3.9
Miravalles	MAG_24	4419367	100	95161	54.3	98.6	1.4	96.8	0.0
Salitral	MAG_26	5227688	94	85635	59.8	95.8	4.2	98.7	0.5
Miravalles	MAG_80	4521693	77	87718	64.5	98.6	0.0	96.8	0.9

Las Lilas	MAG_15	2490690	76	119997	57.3	94.4	7.0	90.9	2.3
Miravalles	MAG_5	5385360	73	151573	59.9	95.8	4.2	99.7	0.5
Salitral	MAG_19	4131153	68	106691	62.2	93.0	2.8	88.2	0.2
Rocas Calientes	MAG_26	2476037	40	100696	63.6	97.8	0.0	92.3	0.0
Miravalles	MAG_3	5078292	36	249717	56.3	97.2	0.0	92.7	0.0
Salitral	MAG_58	4035386	738	9103	43.6	93.0	4.2	88.6	6.3
Río Naranjo	MAG_13	3429613	403	48207	52.5	97.2	7.0	99.5	1.7
Salitral	MAG_57	4226853	393	20364	34.3	97.2	4.2	98.0	2.3
Río Naranjo	MAG_19	3741656	379	20376	36.0	97.2	4.2	97.6	1.4
Miravalles	MAG_33	5511802	287	36307	49.4	97.2	2.8	99.3	0.0
Salitral	MAG_9	4847523	181	38343	56.1	97.2	2.8	98.6	0.7
Salitral	MAG_3	2778176	110	66152	52.7	93.0	0.0	97.6	0.0
Miravalles	MAG_32	4486119	108	65991	53.8	98.6	4.2	98.6	0.7
Rocas Calientes	MAG_10	2921566	103	44470	59.8	100.0	1.4	100.0	0.0
Salitral	MAG_49	2452381	243	46679	35.9	95.8	8.5	99.1	10.0
Rocas Calientes	MAG_9	1912841	236	13148	68.2	98.6	0.0	100.0	0.4
Las Lilas	MAG_6	3472898	410	25663	63.0	93.0	4.2	93.5	2.0
La Luz	MAG_74	4140517	253	42394	38.2	56.3	2.8	84.5	5.3
Miravalles	MAG_39	4348778	50	745061	40.6	98.6	4.2	98.7	1.0
Las Lilas	MAG_3	2689334	151	56077	71.6	94.4	1.4	93.5	1.9
La Luz	MAG_71	2190319	99	41660	43.4	95.8	0.0	98.2	1.8
Rocas Calientes	MAG_5	906433	185	8670	32.6	76.3	2.9	70.0	1.7

Rocas Calientes	MAG_14	4483130	416	80026	62.4	89.2	6.5	86.7	5.6
Miravalles	MAG_16	3264409	87	52819	60.0	98.6	0.0	98.8	0.0
La Luz	MAG_51	2722971	33	142765	65.9	98.6	0.0	97.7	0.0
Río Naranjo	MAG_8	2102490	680	4010	62.3	80.3	2.8	61.1	1.7
Miravalles	MAG_94	2442296	520	6570	64.9	77.5	8.5	63.8	2.9
Rocas Calientes	MAG_7	4043285	480	25180	71.3	100.0	2.9	98.5	2.5
Río Naranjo	MAG_4	4048689	316	22904	71.1	100.0	4.2	98.9	0.8
Miravalles	MAG_51	4111005	305	28620	73.6	88.7	4.2	96.1	2.0
Miravalles	MAG_48	3590417	178	94307	69.3	98.6	14.1	98.8	6.6
Salitral	MAG_27	3368313	176	32262	72.5	90.1	0.0	97.7	0.5
Salitral	MAG_30	2774922	151	36423	68.6	67.6	2.8	88.6	0.6
Salitral	MAG_33	2691136	25	175021	66.4	100.0	0.0	100.0	0.0
Río Naranjo	MAG_22	3734004	280	22385	38.6	93.0	4.2	97.5	0.7
Rocas Calientes	MAG_13	3833021	327	17000	65.0	97.8	2.9	95.7	2.8
Miravalles	MAG_76	5208194	63	117024	65.2	97.2	5.6	98.7	2.0
Salitral	MAG_13	2265823	7	628101	62.7	95.8	2.8	95.3	1.4
Las Lilas	MAG_36	1630143	40	111166	29.2	90.1	1.4	91.5	0.0
Miravalles	MAG_31	3242556	72	93455	46.0	91.5	1.4	98.2	0.9

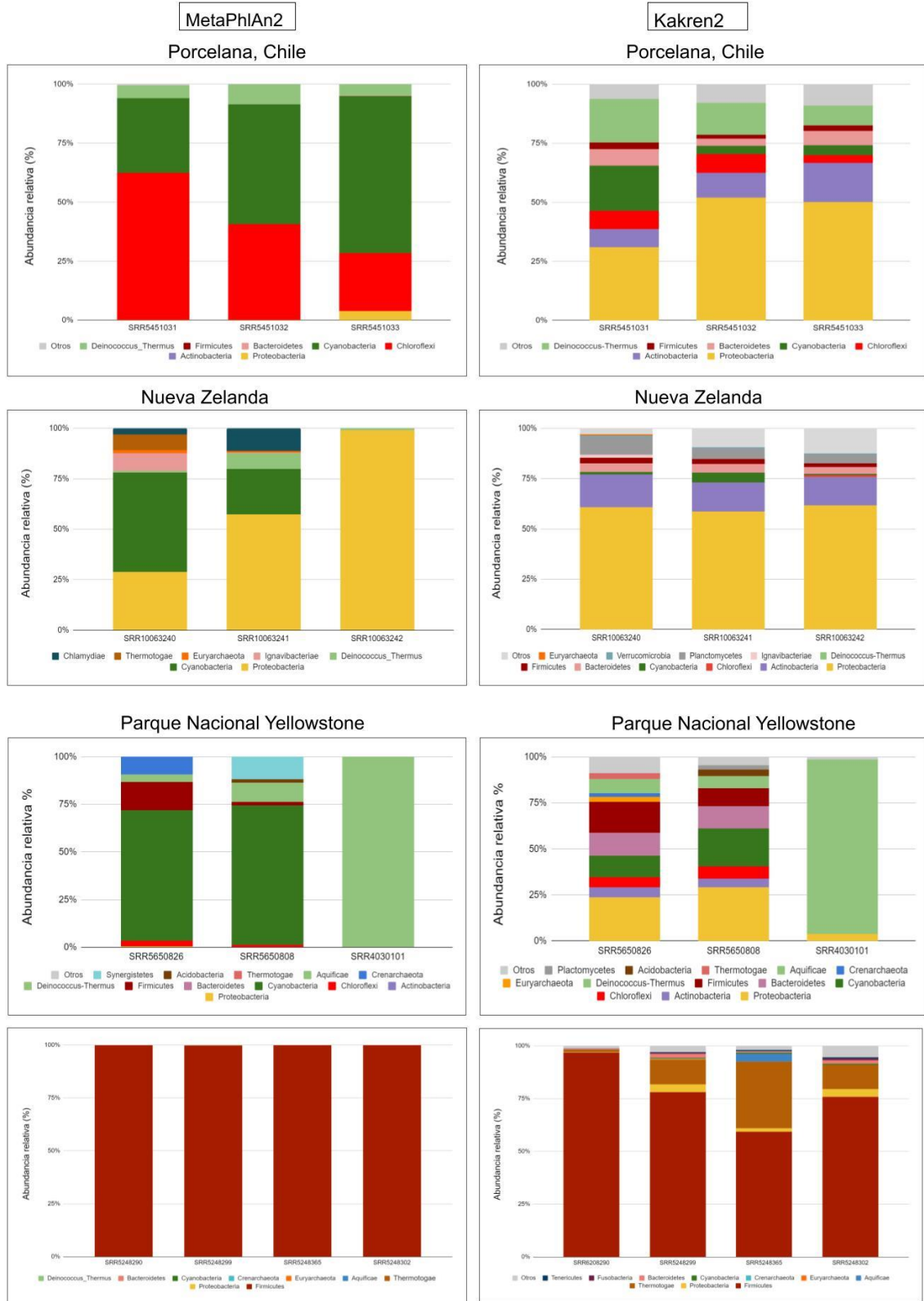
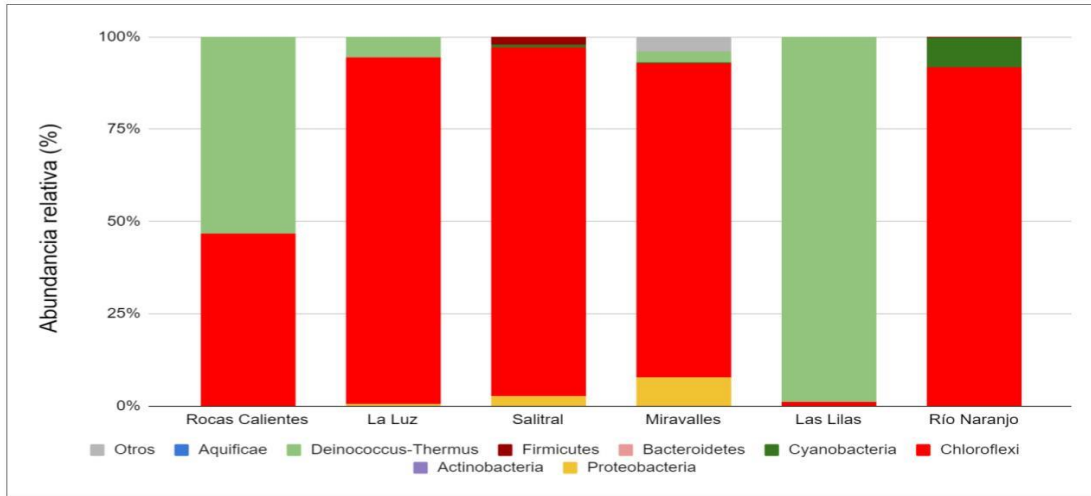
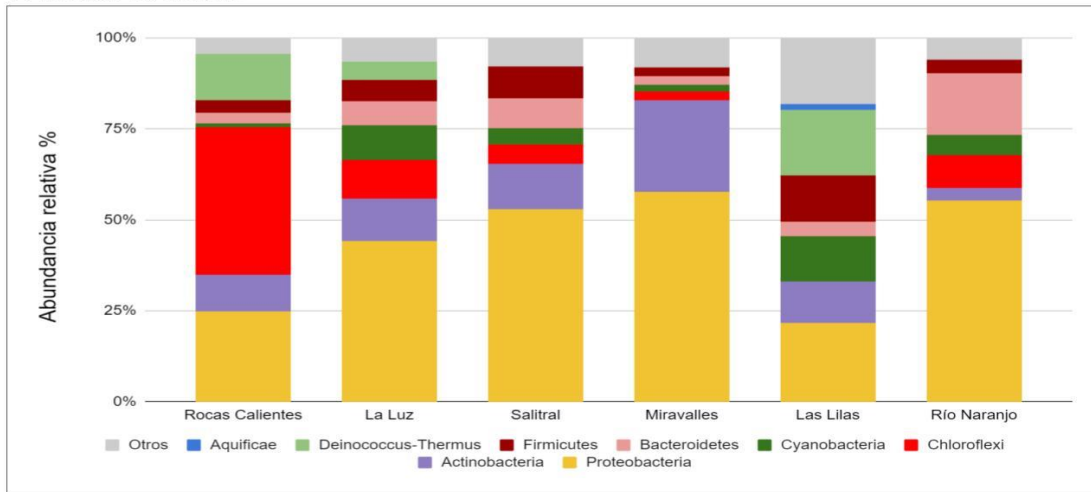


Figura 1. Gráfico de barras de la abundancia relativa de cada filo según los resultados del análisis taxonómico utilizando MetaPIAn2 y Kraken2 para las muestras de Chile, Nueva Zelanda y Yellowstone.

A- Análisis en MetaPhlan2



B- Análisis en Kraken2



C- Análisis en MG-RAST

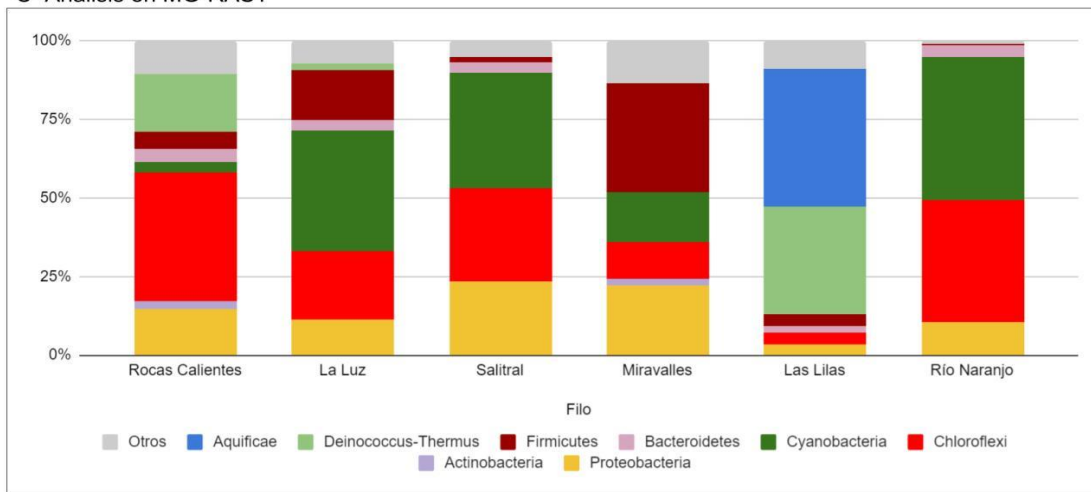


Figura 2. Gráfico de barras de la abundancia relativa de cada filo según los resultados del análisis taxonómico utilizando MetaPhlan2, MG-RAST y Kraken2 para las muestras de Costa Rica.