Centro de Investigación Científica y de Educación

Superior de Ensenada, Baja California



# Maestría en Ciencias en Ciencias de la Vida con orientación en Biología Ambiental

Caracterización de las comunidades microbianas presentes en los sedimentos de Perdido y Coatzacoalcos del Golfo de México mediante análisis metagenómicos

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de

Maestro en Ciencias

Presenta

**Clara Barcelos Santiago** 

Ensenada, Baja California, México

2018

Tesis defendida por

# **Clara Barcelos Santiago**

y aprobada por el siguiente comité

Dra. María Asunción Lago Lestón Directora de tesis

Dra. Meritxell Riquelme Pérez

Dr. Juan Carlos Herguera García

**Dra. Jennyfers Chong Robles** 



**Dra. Clara Elizabeth Galindo Sánchez** Coordinadora del Posgrado en Ciencias de la Vida

> **Dra. Rufina Hernández Martínez** Directora de Estudios de Posgrado

Clara Barcelos Santiago © 2018 Queda prohibida la reproducción parcial o total de esta obra sin el permiso formal y explícito del autor y director de la tesis. Resumen de la tesis que presenta **Clara Barcelos Santiago** como requisito parcial para la obtención del grado de Maestro en Ciencias de la Vida con orientación en Biología Ambiental.

### Caracterización de las comunidades microbianas presentes en los sedimentos de Perdido y Coatzacoalcos del Golfo de México mediante análisis metagenómicos

Resumen aprobado por:

Dra. María Asunción Lago Lestón Directora de tesis

La metagenómica hace uso de las tecnologías de secuenciación masiva para estudiar el contenido genómico presente en una muestra tomada del ambiente. Esta disciplina ha permitido expandir el conocimiento sobre los microorganismos que habitan en los diversos ecosistemas del planeta sin la necesidad de aislarlos o cultivarlos. Un gran número de muestras tomadas de la columna de agua y de sitios impactados por derrames de hidrocarburos en el golfo de México ha sido estudiado con aproximaciones metagenómicas mientras que la caracterización de la microbiota de sus sedimentos ha sido relativamente escasa. Por esta razón en este trabajo se analizaron muestras de sedimentos tomados de profundidades que van desde 550 hasta 3,200 m de profundidad en 18 puntos de dos regiones del Golfo de México (Perdido y Coatzacoalcos), con el fin de conocer la comunidad procariota que los habita. Para ello se utilizaron dos aproximaciones metagenómicas: secuenciación de amplicones de la región V4 del gen 16S del ARNr, y secuenciación shotqun. Las secuencias de amplicones obtenidas se procesaron informáticamente con herramientas de USEARCH y comandos de QIIME. La asignación taxonómica de las OTUs se realizó de novo usando la base de datos SILVA 128. El limpiado, ensamble y anotación de las lecturas obtenidas de la secuenciación shotqun se realizó con las herramientas Trimmommatic, MEGAHIT y Prokka, respectivamente. Los resultados obtenidos muestran que la profundidad de la toma de muestra tiene un mayor efecto en la rigueza y abundancia de especies que la región de procedencia de los sedimentos. Por ejemplo los miembros de Planctomycetacia (Bacteria) disminuyen al aumentar la profundidad mientras que los del Marine Group I (Archaea) aumentan. También se encontró que los filos predominantes son: Proteobacteria, Thaumarchaeota, Planctomycetes, Acidobacteria y Chloroflexi, y que los sedimentos de profundidades mayores a 2,000 m son similares entre sí en términos de composición microbiana. En los metagenomas de los sedimentos de las tres estaciones analizadas se identificaron genes involucrados en reacciones de óxido-reducción de compuestos de azufre y nitrógeno y en la fijación del carbono, además de genes relacionados con la degradación de hidrocarburos aromáticos. El estudio de la microbiota que habita en estos ambientes es necesario para identificar patrones que permitan inferir los procesos biogeoquímicos que son impulsados por los microorganismos, y proporcionar una línea base para el planteamiento de estudios dirigidos a la obtención de productos de interés biotecnológico.

**Palabras clave:** Metagenómica, sedimentos, amplicones, *shotgun*, Perdido, Coatzacoalcos, golfo de México.

Abstract of the thesis presented by **Clara Barcelos Santiago** as a partial requirement to obtain the Master of Science degree in Life Sciences with orientation in Environmental Biology.

### Metagenomic analysis of seafloor microbial communities from Perdido and Coatzacoalcos in the Gulf of Mexico

Abstract approved by:

Dr. María Asunción Lago Lestón Thesis Director

Metagenomics refers to the study of the genomic content directly extracted from an environmental sample and has been used to increase the knowledge about ecosystems and their microbial inhabitants. Marine environments have been extensively studied using metagenomic approaches with the aim to characterize the microbial diversity and their interactions with the geosphere. In the Gulf of Mexico, metagenomics studies have been conducted primarily in samples taken from the water column and from sites impacted by oil spills, whereas sediments have traditionally received less attention. Sediments were taken from 18 sites of the regions of Perdido and Coatzacoalcos in the Gulf of Mexico, from depths ranging from 500 to 3,200 m under the sea level, in order to know the prokaryotic community inhabit them. Samples were analyzed using two metagenomic approaches: amplicon based and shotgun sequencing. Reads from the sequencing of amplicons of the V4 region of the 16S rRNA gene were processed using bioinformatics tools: USEARCH algorithm was used to merge the forward and reverse sequences and to clean up the sequences; QIIME's commands were used to do the taxonomic assignment, OTUs were assigned de novo using the SILVA 128 database. The cleaning, assembly, and annotation of the reads from the shotgun sequencing were performed using Trimmommatic, MEGAHIT, and Prokka, respectively. Results show that the sample depth rather than the location influences the richness and abundance of the microbial inhabitants of the sediments. For example, members of Planctomycetacia (Bacteria) decrease with increasing depth while, the abundance of the Marine Group I (Archaea) members increases. It was also found that the predominant species from the regions studied are members of the Proteobacteria, Thaumarchaeota, Planctomycetes, Acidobacteria, and Chloroflexi fila and, that sediments of depths greater than 2,000 m are similar to each other in terms of microbial composition. The microbiome of the sediments contains genes involved in oxidation-reduction reactions of sulfur and nitrogen compounds and in carbon fixation; also genes related to the degradation of aromatic hydrocarbons were identified. Studies about microbial communities from sediments are necessary to identify patterns that allow us to infer the biogeochemical processes driven by microorganisms, and provide a baseline for studies aimed to obtain products of biotechnological interest.

Keywords: Metagenomics, marine sediments, 16S, shotgun, Perdido, Coatzacoalcos, gulf of Mexico.

A mis padres,

A Donovan y

A todos aquellos que creen en mí.

# Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada con el número de registro 613345 en el periodo comprendido entre el 01-09-2016 y el 31-08-2018.

Al CIGoM, por el hermoso proyecto, por el apoyo económico para mi estancia en la UC Davis y por la beca terminal con folio 25647 del 1-09-2018 al 30-11-2018. Esta investigación ha sido financiada por el Fondo Sectorial CONACYT-SENER-Hidrocarburos, proyecto 201441. Esta es una contribución del Consorcio de Investigación del Golfo de México (CIGoM).

Al CICESE, por aceptarme en el posgrado y el apoyo otorgado para mi estancia en Davis.

A la Dra. Asunción Lago por su confianza, sus enseñanzas, por permitirme formar parte de su equipo de trabajo y por fomentarme el amor por la ciencia.

Al Dr. Juan Carlos Herguera, por su apoyo para mi estancia en Davis, por sus sugerencias, sus historias, sus consejos y sobre todo por su ejemplo.

A la Dra. Meritxell Riquelme, por hacerme dudar siempre, por su interés en el proyecto y la disponibilidad de ayudarme.

A la Dra. Jennyfers Chong, por todo el invaluable apoyo técnico, por las asesorías y revisiones, por el interés constante y ayudarme siempre que lo necesitaba. Este trabajo no hubiera sido posible sin toda tu ayuda.

A los miembros de nuestro laboratorio: Dante, Yamne, Karla, Haydee, Arlette, Mateo y Balam. Gracias por toda su ayuda y paciencia y por hacer que estos años fueran más llevaderos.

A Fernando, por todas las historias, por escucharme, por las sugerencias de estilo, los consejos y por ayudarme a ser una mejor persona.

A Luis y Aldo, por los buenos momentos, por los regaños necesarios y por haberme guiado cuando más perdida estaba.

Al personal administrativo del departamento: Adriana, Yoli, Anabel y Gaby. Por hacer mi vida más sencilla y por brindarme su amistad.

A mis papás. Gracias por mi vida, por seguir creyendo en mí, por no limitarme nunca, por su apoyo incondicional, por su ejemplo, por su disciplina, por inculcarme a hacer las cosas con amor y por todo lo que las palabras no pueden expresar. Los amo infinitamente.

Y a mis sobrinos, por recordarme día a día las razones para seguir haciendo ciencia.

# Tabla de contenido

Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	ix
Lista de tablas	xi
Capítulo I. Introducción	1
1. El estudio de muestras ambientales	3
1.1. La secuenciación	3
1.2. La metagenómica	5
1.2.1. La secuenciación de amplicones	5
1.2.2. La secuenciación shotgun	6
1.2.3. Estudios ambientales realizados mediante el análisis de metagenomas	7
1.3. El Golfo de México	8
1.4. Justificación	10

Capítulo 2. Metodología	12
2.1. Toma de muestra	12
2.2. Metagenómica dirigida	13
2.2.1. Extracción de ADN	13
2.2.2. Preparación de la biblioteca de secuenciación	14

1.5. Preguntas de investigación \_\_\_\_\_\_ 11

1.6. Objetivo General \_\_\_\_\_\_ 11

1.7. Objetivos específicos 11

2.2.3. Secuenciación	16
2.2.4. Procesamiento informático de las secuencias	16
2.2.4.1. Pre-procesamiento de las secuencias	16
2.2.4.2. Obtención de las Unidades Taxonómicas Operacionales	17
2.2.4.3. Análisis de la diversidad microbiana	17
2.2.4.4. Tests estadísticos	18
2.2.5. Inferencia funcional usando PICRUSt	19
2.3. Metagenómica shotgun	19
2.3.1. Concentración del ADN	20
2.3.2. Preparación de la biblioteca	20
2.3.3. Secuenciación shotgun	20
2.3.4. Procesamiento informático de las secuencias	21

Capítulo 3. Resultados	_ 22
3.1. Metagenómica dirigida	_ 22
3.1.1. Extracción de ADN	_ 22
3.1.2. Amplificación de la región de interés	_ 24
3.1.3 Secuenciación y procesamiento informático	_ 24
3.1.4. Caracterización taxonómica de la microbiota de las zonas de estudio	_ 28
I. Composición de la microbiota de los sedimentos entre 550 y 800 m de profundidad	_ 34
II. Composición de la microbiota de los sedimentos entre 1,000 y 1,500 m de profundidad	_ 35
III. Composición de la microbiota de los sedimentos entre 2,000 y 3,200 m de profundidad	_ 37
3.1.5. Inferencia de la función de la microbiota de los sedimentos de Perdido y Coatzacoalcos	_ 38
3.2. Secuenciación shotgun	_ 39
3.2.1. Secuenciación y el procesamiento informático	_ 40

Capítulo 4. Discusión	43
4.1. Microorganismos en el fondo marino de Perdido y Coatzacoalcos	43
4.2. Potencial metabólico de la microbiota de los sedimentos marinos	47
4.3. Limitaciones y perspectivas	48
4.4. Sugerencias técnicas	48
Capítulo 5. Conclusiones	50
Literatura citada	51
Anexos	58
Anexo A. Protocolo para la extracción de ADN	58
Anexo B. Información sobre los primers de amplificación	60
Anexo C. Información adicional sobre la Reacción en Cadena de la Polimerasa	61
Anexo D. Otros datos sobre las muestras	62
Anexo E. Número de OTU en las estaciones de muestreo entre 550 y 800 m	65
Anexo F. Número de OTU en las estaciones de muestreo entre 1,000 y 1,500 m	66
Anexo G. Número de OTU en las estaciones muestreadas entre 2,000 y 3,200 m	67
Anexo H. Porcentaje de abundancia relativa de los dominios procariontes en los sedimentos de de muestreo	las zonas 68
Anexo I. Composición porcentual de las 7 categorías del nivel 2 obtenidas en PICRUSt	69
Anexo J. Genes identificados en los contigs generados de las lecturas obtenidas por la secue shotgun	enciación 70
Anexo K. Fila de las arqueas identificadas en las estaciones de muestreo	77

# Lista de figuras

Figura 1. Fechas de lanzamiento comercial de las plataformas de secuenciación	_ 4
Figura 2. Aproximaciones metagenómicas para el estudio de muestras ambientales	5
Figura 3. Ubicación de las 10 estaciones de muestreo donde se colectaron las muestras de sediment analizadas por Covarrubias-Rodríguez (2016)	os _ 9
Figura 4. Mapa de las estaciones muestreadas en la campaña oceanográfica MMF-01	13
Figura 5. Gel del agarosa del ADN extraído de algunas de las muestras de sedimentos2	23
Figura 6. Electroforesis del producto de PCR de algunas de las réplicas realizadas 2	24
Figura 7. Número de secuencias procesadas y OTU asignadas por estación2	25
Figura 8. Número de OTU asignadas con respecto al número de secuencias obtenidas2	26
Figura 9. Curvas de rarefacción2	26
Figura 10. Número de especies observadas y esperadas (Chao1) por estación de muestreo	27
Figura 11. Número de OTUs únicas y compartidas entre las estaciones de muestreo 2	28
Figura 12. Abundancia relativa de los fila procariontes en las dos zonas de muestreo 2	29
Figura 13. Porcentajes de abundancia relativa promedio de los microorganismos procariont identificados en las réplicas por estación de muestreo a nivel filo	es 30
Figura 14. Análisis de Coordenadas Principales (PCoA) de las réplicas considerando a la zona de muestre como la principal fuente de variación	eo 31
Figura 15. Análisis de Coordenadas Principales (PCoA) de las réplicas considerando a la profundidad muestreo como la principal fuente de variación	de 32
Figura 16. Abundancia relativa promedio a nivel clase de los microorganismos procariontes identificad en las estaciones de muestreo ordenadas de menor a mayor profundidad	os 33
Figura 17. Porcentaje de abundancia relativa promedio a nivel orden de los microorganismos procariont identificados en las estaciones entre 550 y 800 m de profundidad	es 34
Figura 18. Abundancia relativa promedio a nivel orden de los microorganismos procariontes identificad en las estaciones entre 1,000 y 1,500 m de profundidad	os 36

gura 19. Abundancia relativa promedio a nivel orden de los procariotas identificados en las estacion entre 2,000 y 3,200 m de profundidad	nes 37
gura 20. Categorías funcionales inferidas en los metagenomas de amplicones de las muestras analizad 	das 39
gura 21. Clases de microorganismos procariontes con abundancia relativa mayor al 1% en los sedimen de las estaciones C10, D18 y C13	tos 40
<b>gura 22.</b> Representación porcentual de las categorías funcionales identificadas con KEGG Ghost Koala los genes identificados y anotados de las estaciones C10, D18 y C13	de 42
gura 23. Flujo de trabajo para la construcción de los amplicones etiquetado	61
gura 24. Número de OTUs únicas y compartidas en las estaciones de muestreo entre 550 y 800 m profundidad	de 65
gura 25. Número de OTUs únicas y compartidas en las estaciones de muestreo entre 1,000 y 2,000 m profundidad	de 66
gura 26. Número de OTUs únicas y compartidas en las estaciones de muestreo entre 2,000 y 3,000 m profundidad	de 67
gura 27. Grafica de la distribución de los Dominios de microorganismos procariontes identificados	68
gura 28. Composición porcentual promedio de las 7 categorías del nivel 2 obtenidas en PICRUSt	69
gura 29. Abundancia de los fila de arqueas identificados	77

х

# Lista de tablas

Tabla 1. Estaciones de muestreo y profundidad de la columna de agua a la que fueron colectadas	_ 12
Tabla 2. Valor promedio de la concentración y pureza del ADN extraído de cada muestra de Perdido	_ 22
Tabla 3. Valores de las concentraciones y purezas del ADN extraído de las muestras de Coatzacoalcos	23
Tabla 4. Tabla de resultados generales del procesamiento de las secuencias shotgun	_ 41
Tabla 6. Condiciones de los ciclos de PCR	_ 61
Tabla 7. Variables consideradas para la obtención de los amplicones en las réplicas de Perdido	_ 62
Tabla 8. Variables consideradas para la obtención de los amplicones en las réplicas de Coatzacoalcos _	_ 63
Tabla 9. Índices de riqueza de especies por estación de muestreo	_ 64
Tabla 10. Genes identificados en el microbioma de los sedimentos de la estación C10	_ 70
Tabla 11. Genes identificados en el microbioma de los sedimentos de la estación C13	_ 71
Tabla 12. Genes identificados en el microbioma de los sedimentos de la estación D18	_ 76

# Capítulo 1. Introducción

Los microorganismos son los responsables de catalizar la mayoría de las reacciones químicas que dan forma a los ambientes de la Tierra y sus océanos. Diferentes especies de bacterias y arqueas intervienen en procesos que transforman y reciclan las sustancias orgánicas e inorgánicas necesarias para el mantenimiento de la biosfera (Falkowski *et al.,* 2008). Estos procesos, denominados en conjunto como ciclos biogeoquímicos, ocurren naturalmente en suelos, cuerpos de agua y sedimentos. Las reacciones químicas que componen estos ciclos son extremadamente complejas y están mediadas por una diversidad microbiana desconocida o pobremente caracterizada (Madsen, 2011).

Los microorganismos que habitan la biosfera son altamente diversos y sólo un pequeño porcentaje de ellos (1%) ha podido ser cultivado (Rappe y Giovannoni, 2003). El estudio del porcentaje de microorganismos restante es uno de los retos más importantes para la microbiología actual. El desarrollo de métodos de cultivo ha sido extremadamente complicado puesto que en el crecimiento de los microorganismos intervienen una gran variedad de factores que aún permanecen desconocidos (Epstein, 2013).

Ante la dificultad que representa cultivar a los microrganismos se han desarrollado nuevas estrategias y metodologías que permiten estudiarlos sin la necesidad de aislarlos y/o cultivarlos. Para la realización de estos estudios se requiere la extracción del material genómico de una muestra y su posterior secuenciación. En la actualidad esta secuenciación se realiza mayoritariamente usando las tecnologías de secuenciación de alto rendimiento o *Next Generation Sequencing* (NGS); puesto que permiten generar gigabases (Gb) de datos genómicos a un costo sustancialmente más bajo que el que era posible hace sólo 15 años (http://www.genome.gov/sequencingcosts/). Una de las disciplinas que más se ha beneficiado de estas mejoras, es la Metagenómica, que permite el estudio genómico de los microorganismos no cultivables de manera más rápida, económica y reproducible (Wooley *et al.,* 2010).

La metagenómica ha usado tradicionalmente dos aproximaciones para el estudio de muestras ambientales: (i) la secuenciación de amplicones o *metabarcoding* y (ii) la secuenciación *shotgun*. La primera implica la elección de un gen marcador taxonómicamente informativo con el que se identifica la diversidad microbiana presente en la muestra (los organismos procariontes se han identificado comúnmente con el uso de una región del gen 16S que codifica para la subunidad menor del ARN ribosomal). En la segunda, todo el ADN extraído es fragmentado y secuenciado. Con esta metodología además de identificar a los microorganismos presentes en la muestra (su microbiota), se puede conocer su potencial metabólico, es decir la función de la comunidad microbiana. Usando estas aproximaciones se han logrado estudiar diferentes microbiomas, que son el conjunto de genes que posee una microbiota particular (https://learn.genetics.utah.edu/content/microbiome/); por ejemplo, el humano (Quin, *et al.*, 2010) y el de gran variedad de muestras de suelo, agua y sedimentos (Craig, 2004; Mason *et al.*, 2014; Lamendella *et al.*, 2014).

La microbiota que habita los ecosistemas marinos ha sido estudiada con la finalidad de identificar su composición y la relación que tiene con los ciclos biogeoquímicos. Para ello se han analizado muestras de agua tomadas en diferentes profundidades y en diversos puntos del planeta (Craig, 2004; Petersen *et al.*, 2011; Trevors *et al.*, 2011; Kennedy, 2011), así como muestras de sedimentos (Biddle *et al.* 2006; Mason *et al.*, 2014; Marshall *et al.*, 2018; Vigneron *et al.*, 2018).

El incremento actual del interés en los sedimentos marinos proviene del descubrimiento de que la microbiota que los habita no sólo está viva sino que está activa. Esto ha llevado a pensar que en estos ambientes hay microorganismos que podrían ser, o poseer genes, de interés biotecnológico. Al ser parte de una provincia petrolera, los sedimentos del golfo de México (GdeM) han estado en constante contacto con hidrocarburos del petróleo provenientes de las actividades diarias de extracción, de derrames y de emanaciones naturales (chapopoteras) por lo que se piensa que los microorganismos adaptados a estos ambientes podrían estar involucrados en la degradación de hidrocarburos (HC) (MacDonald *et al.,* 2015).

Los derrames de petróleo en el GdeM son comunes y son una de las causas principales de la entrada al océano de una gran cantidad y variedad de HC. Estos pueden volatilizarse, ser degradados en la superficie, transportarse o aglomerarse y sedimentarse dependiendo de sus características fisicoquímicas y del volumen de petróleo derramado (Harvey, 1987). El derrame más grande de la industria petrolera se produjo en el GdeM a causa de la explosión del pozo de aguas profundas (Macondo) que provocó el vertimiento de 5 millones de barriles de petróleo crudo al mar que impactaron gravemente la flora y fauna de la zona.

Para evaluar los cambios en la microbiota de la zona afectada por el derrame se realizaron estudios metagenómicos que evidenciaron una sucesión de las comunidades microbianas tanto en la columna de agua como en las costas que fueron alcanzadas por el petróleo (Hazen *et al.*, 2010; Valentine *et al.*, 2010; Kessler *et al.*, 2012; Bik, 2012; Lamendella *et al.*, 2014). Esta catástrofe puso de manifiesto la falta de información microbiológica que se tiene del GdeM y la escasa atención que la microbiota de sus sedimentos ha recibido. Esta microbiota podría desempeñar un papel clave en procesos ecológicos y

geológicos (Falkowski *et al.,* 2008) y ser una fuente de compuestos bioactivos desconocidos no explotados (Kodzius y Gojobori, 2015) por lo que identificar su composición proporcionaría una línea base sobre la cual plantear nuevas investigaciones.

### 1. El estudio de muestras ambientales

La posibilidad de estudiar a los microorganismos presentes en una muestra ambiental comenzó con la publicación de los análisis comparativos de secuencias del gen del ARNr (Ácido Ribonucleico ribosomal) realizados por Woese (1987). Estos estudios definieron por primera vez a los tres dominios de la vida (Archaea, Bacteria y Eukarya) desde una perspectiva genética y proporcionaron las bases necesarias para su categorización mediante el uso de las moléculas de ARNr que se encuentran en todas las formas de vida celulares. La identificación de la microbiota con el uso del ARNr permite hacer inferencias ecológicas representativas sobre su función en un ambiente particular ya que proporciona información sobre toda la comunidad y no sólo de la porción que puede ser cultivada (Head et al., 1997).

Los análisis comparativos de secuencias de ARNr se han realizado con el uso de herramientas moleculares como la electroforesis en gel con gradiente de desnaturalización (Duineveld *et al.*, 2001; Simpson y White, 1997), la hibridación celular completa (*whole cell hybridization*) (Amann, 1995; Musat y Widdel, 2007) y la secuenciación, siendo esta última la más usada en la actualidad para análisis de muestras ambientales (Quin, *et al.*, 2010; Craig, 2004; Mason *et al.*, 2014; Lamendella *et al.*, 2014).

### 1.1. La secuenciación

Los métodos actuales para el estudio de los microbios de una muestra ambiental se basan en la extracción del ADN genómico de la muestra y su posterior secuenciación. La secuenciación comenzó en los años 70s con el uso de la cromatografía en dos dimensiones que permitía obtener secuencias de fragmentos cortos de ADN. Con la publicación del método de Sanger de terminación de la cadena (Sanger *et al.*, 1977) se comenzaron a obtener secuencias más largas, en menor tiempo y de manera más confiable y reproducible. Desde entonces, la secuenciación ha pasado por cambios importantes que la han convertido en una de las técnicas más usadas para el estudio del material genómico.

La secuenciación de alto rendimiento (o *Next Generation Sequencing, NGS*) fue desarrollada a partir de la aplicación de un nuevo enfoque para secuenciar fragmentos cortos (llamados lecturas o *reads*) de manera paralela y masiva y provocando un incremento en la capacidad de secuenciación (Ronaghi, 2001). A raíz de esto se desarrollaron una gran variedad de plataformas con diferentes características que permiten obtener un gran número de datos (Figura 1).



**Figura 1.** Fechas de lanzamiento comercial de las plataformas de secuenciación contra los millones de pares de bases obtenidas por ejecución. Los números dentro de los puntos de datos denotan las longitudes de lectura actuales. Las plataformas de secuenciación están codificadas por colores (Reuter et al., 2015).

Las tecnologías de secuenciación han revolucionado las ciencias biológicas; su alto rendimiento, reproducibilidad y velocidad ahora permiten una amplia variedad de aplicaciones y el estudio de sistemas biológicos que antes no era posible.

## 1.2. La metagenómica

La Metagenómica es una de las disciplinas que más se ha visto beneficiada por las mejoras en las tecnologías de secuenciación masiva. El término fue acuñado por Handelsman *et al.* (1998) y hace referencia al análisis de las secuencias de los genomas microbianos que se encuentran en una muestra ambiental como si fueran un único genoma, por lo que se le denomina comúnmente metagenoma. Actualmente estas secuencias se obtienen tanto con la secuenciación de amplicones como con la secuenciación *shotgun* (Figura 2).



Figura 2. Aproximaciones metagenómicas para el estudio de muestras ambientales. Modificado de Morgan y Huttenhower, 2012.

#### 1.2.1. La secuenciación de amplicones

La secuenciación de amplicones se ha usado para conocer la composición de las comunidades de microorganismos en diversos ambientes (Quin, *et al.,* 2010; Craig, 2004; Mason *et al.,* 2014; Lamendella *et al.,* 2014). Los procariontes (Bacteria y Archaea) se han identificado con el análisis de secuencias de regiones del gen 16S (que codifica para la subunidad menor del ARNr) por tres razones principales: (i) Se

encuentra presente en todas las especies de los dominios Bacteria y Archaea. (ii) Contiene regiones altamente conservadas que pueden emplearse para diseñar cebadores de PCR de amplio espectro (teóricamente universales) y sus regiones variables son útiles para la discriminación entre microorganismos y su clasificación taxonómica (Kuczynski *et al.*, 2011). (iii) Es el marcador taxonómicamente informativo del que se tienen bases de datos con un gran número de secuencias depositadas (Yarza *et al.*, 2014).

El gen 16S consta de nueve regiones hipervariables (V1-V9); algunas presentan mayor variabilidad que otras y permiten realizar una asignación taxonómica precisa (Yu y Morrison, 2004; Youssef *et al.*, 2009). Las regiones V1-V4 presentan dos ventajas cuando se trata de asignación taxonómica: (i) proporcionan una mayor resolución para distinguir entre especies debido a su elevada variabilidad; (ii) existen un mayor número de secuencias parciales correspondientes a estas regiones en las bases de datos, facilitando la asignación taxonómica (Kim *et al.*, 2011).

La secuenciación del gen 16S ha permitido la identificación de microorganismos que permanecían desconocidos en una gran variedad de ambientes y se ha usado para crear perfiles de diversidad microbiana que permiten asociar a los microorganismos a un entorno particular (Rappé y Giovannoni, 2003; Lozupone y Knight, 2007; McCliment *et al.*, 2006; Soo *et al.*, 2009). Sin embargo, esta metodología sólo provee información sobre la composición de las comunidades microbianas y aunque se puede realizar una inferencia funcional, la precisión de ésta depende del número de genomas disponibles en las bases de datos (Langille *et al.*, 2013).

### 1.2.2. La secuenciación shotgun

La secuenciación *shotgun* permite el estudio de todos los genes de todos los organismos presentes en una muestra compleja, incluyendo a eucariontes, procariontes y virus. Gracias a este método se puede evaluar la diversidad microbiana y detectar su abundancia en varios ambientes además de proveer información sobre la posible función que estos microorganismos tienen en esto ambientes (Illumina, 2015).

En la secuenciación *shotgun*, el ADN genómico extraído de la muestra es cortado en fragmentos que son posteriormente amplificados y secuenciados. A partir de esto se obtienen millones de secuencias

de ADN genómico que mediante programas bioinformáticos se alinean unos con otros para obtener secuencias más largas (*contigs*). Algunos de estos *contigs* pueden contener genes taxonómicamente informativos y/o codificantes que proveen información sobre las funciones biológicas de la comunidad microbiana estudiada.

La secuenciación *shotgun* ha sido usada para identificar nuevos virus, caracterizar la diversidad genómica y funcional de bacterias no cultivadas y revelar proteínas ecológicamente importantes (Yozwiak *et al.,* 2012; Wrighton *et al.,* 2012; Godzik, 2011).

### 1.2.3. Estudios ambientales realizados mediante el análisis de metagenomas

El estudio de ecosistemas marinos a escala global ha sido impulsado por proyectos como el *Ocean Drilling Program* (ODP) (Graham y Julson, 1989), el *Global Ocean Sampling* del Instituto J. Craig Venter (Parthasarathy *et al.*, 2007), el *Earth Microbiome Project* (Gilbert *et al.*, 2010) y la iniciativa Europea *Tara Oceans Expedition* (Hingamp *et al.*, 2013).

En 2004, Venter *et al.,* reportaron la secuenciación *shotgun* del ADN extraído de muestras de agua de mar tomada en el Mar de los Sargazos. Como resultado, generaron 1,045 millones de pares de bases de secuencias. Los autores estimaron que estos datos provenían de al menos 1,800 especies genómicas, incluyendo 148 filotipos bacterianos antes desconocidos. Además, identificaron cerca de 1.2 millones de genes en las muestras que no se habían reportado, incluyendo más de 782 nuevos fotoreceptores *rhodopsin-like*.

Como parte del ODP, Biddle *et al* (2006) estudiaron muestras de sedimentos provenientes de regiones generadoras de metano y reductoras de sulfatos, encontrando que los miembros de Crenarchaeota, Chloroflexi y Euryarchaeota eran los más abundantes. Por las características de la zona, los autores esperaban encontrar un gran número de genes implicados en la metanogénesis y reducción de sulfatos, sin embargo lograron identificar relativamente pocos por lo que sugirieron que las condiciones a las que están sujetos los microorganismos del fondo marino han propiciado que estos tengan características que aún no han sido descritas (Biddle *et al.* 2006).

La composición de una comunidad de microorganismos depende las características fisicoquímicas de su hábitat, por lo que las perturbaciones que este pueda sufrir tienen un efecto directo sobre la abundancia de las especies que la componen. La metagenómica se usó como herramienta para detectar estos cambios después de la perturbación que sufrió la zona noroeste del GdeM a causa del *Deepwater Horizon Oil Spill* (DHOS) en el que se vertieron al océano cerca de 5 millones de barriles de petróleo.

Un estudio realizado sobre muestras tomadas al inicio del derrame en un punto cercano a los 1,100 metros bajo el nivel del mar mostró un cambio en la composición de las comunidades microbianas, habiendo una proliferación de miembros de las *Gammaproteobacteria* (Hazen *et al.*, 2010). Poco después, se encontró que miembros no cultivables de los *Oceanospirillales* dominaban la pluma de aguas profundas pero con el paso del tiempo eran reemplazados por especies de *Colwellia* y *Cycloclasticus* (Valentine *et al.*, 2010). En muestras de agua tomadas cuando finalizó el derrame los microorganismos que dominaban la región eran los oxidadores de metano (Kessler *et al.*, 2012).

A partir de esto, se realizaron estudios para identificar qué microorganismos se relacionaban con fracciones específicas del petróleo y se encontró que miembros de los *Oceanospirillales* son capaces de degradar cicloalcanos (Mason *et al.,* 2012) y que especies de *Colwellia* probablemente están involucradas en la oxidación de etano, propano y benceno (Redmond y Valentine, 2012).

Todos los estudios realizados en el GoM después del derrame han proporcionado información valiosa sobre el impacto que el petróleo crudo puede causar en la columna de agua y zonas costeras. El conocimiento de las comunidades microbianas antes y después del derrame permitió dilucidar estas sucesiones ecológicas. Sin embargo, este evento evidenció la falta de información sobre las comunidades microbianas en las zonas del GoM perteneciente a la Zona Económica Exclusiva de México (ZEE).

# 1.3. El Golfo de México

La ZEE mexicana se extiende desde Tamaulipas hasta Yucatán y contiene más del 80% de las reservas totales de petróleo del país (PEMEX). Ante la creciente demanda de hidrocarburos en el mundo, en la actualidad se busca explotar yacimientos en aguas profundas de las regiones de Perdido (Tamaulipas) y Coatzacoalcos (Veracruz-Tabasco-Campeche). Sin embargo, la perforación en aguas profundas conlleva un mayor riesgo de explosión, y con ello de derrame de hidrocarburos al océano, debido a las presiones

extremadamente elevadas a las que se trabaja, la baja visibilidad en esos entornos y la imposibilidad del acceso humano a la boca de los pozos (Cummings *et al.*, 2015)

Los sedimentos del GdeM han estado en constante contacto con fracciones del petróleo provenientes tanto de derrames como de emanaciones naturales de hidrocarburos (chapopoteras) (MacDonald, 2015; Asl *et al.*, 2016). Sin embargo, los estudios para conocer la diversidad microbiana que los habita han sido relativamente escasos. Esto se ha atribuido a lo complicado que es extraer y conservar muestras de sedimentos con fines biológicos (Parthasarathy *et al.*, 2007). Las mejoras en los protocolos de muestreo y en las técnicas de almacenamiento han permitido estudiar muestras de sedimentos de distintos puntos del GdeM con aproximaciones metagenómicas y han dado lugar a la identificación de la microbiota que los habita.



**Figura 3.** Ubicación de las 10 estaciones de muestreo donde se colectaron las muestras de sedimentos analizadas por Covarrubias-Rodríguez (2016).

En un estudio previo se analizaron muestras de sedimentos someros (entre 0-5 cm y 5-10 cm de profundidad de la columna sedimentaria) y sedimentos profundos (más de 30 cm en la columna sedimentaria) tomadas en 10 puntos del GdeM (entre 1,268 y 3,768 m bajo el nivel del mar) (Figura 3) (Covarrubias-Rodríguez, 2016).

Los resultados mostraron que los sedimentos de los primeros 10 cm de la columna sedimentar son similares entre sí en términos de composición microbiana, en comparación con los sedimentos profundos, aquellos que están a más de 30 cm. En los sedimentos someros, los fila Crenarchaeota, Proteobacteria, SBR1093 (filo bacteriano identificado por primera vez en lodos activados de un reactor *batch*) y Acidobacteria representan alrededor de 95% de los OTU identificados y en los sedimentos profundos los fila más abundantes son Chloroflexi, Planctomycetes, Actinobacteria, OP3 (actualmente conocido como Omnitrophica), Euryarchaeota, GNO4 (filo de bacterias halotolerantes) y *Spirochaetas*, representando a cerca de 50% de los OTU identificados. Este estudio fue una primera caracterización del microbioma de los sedimentos de diversos puntos del GdeM, sin embargo, se enfocó en muestras de sedimentos profundos, por lo que aún hace falta estudiar muestras de sedimentos del talud continental (de profundidad menor a 1,000 m y más cercanas a la costa), observar si existen diferencias entre puntos de muestreo más cercanos entre sí (relativamente) y (de existir estas) tratar de identificar a que se deben.

### 1.4. Justificación

Después del DHOS en el Golfo de México se hizo evidente la falta de información biológica que se tiene de la ZEE mexicana. Ante esto, y debido a la intención de explorar y explotar zonas más profundas para la extracción de petróleo se hace necesaria la identificación de las comunidades procariontes presentes en los sedimentos que podrían resultar afectadas por potenciales derrames así como su caracterización metagenómica para identificar microorganismos (o genes) adaptados a estas regiones y que podrían ser de interés biotecnológico.

Debido a que los microorganismos procariontes de estos sitios han sido poco estudiados y no pueden ser cultivados, las aproximaciones que usan la NGS son la herramienta obligada para estudiarlos. Su identificación y caracterización es fundamental para generar una línea base de conocimiento para el planteamiento de estudios que se dirijan a la obtención de enzimas o productos de interés biotecnológico, además de brindar información sobre las posibles funciones que estos microorganismos tienen en este ambiente.

# 1.5. Preguntas de investigación

Debido a que este es un estudio exploratorio se consideró pertinente el planteamiento de varias preguntas que guiaron la investigación:

- ¿Qué microorganismos procariotas se encuentran en los sedimentos de las regiones de Perdido y Coatzacoalcos?
- ¿Existen diferencias entre la diversidad de microorganismos procariontes de las dos zonas de muestreo?
- ¿Cuál es el potencial metabólico de los procariontes que habitan los sedimentos muestreados?
- ¿Existen microorganismos o genes implicados en la degradación específica de hidrocarburos?

# 1.6. Objetivo General

Identificar a los microorganismos procariontes que habitan los sedimentos de distintas profundidades de las regiones de Perdido y Coatzacoalcos del Golfo de México y su potencial metabólico.

# 1.7. Objetivos específicos

- Identificar a las comunidades de procariontes presentes en los sedimentos marinos muestreados en las regiones de Perdido y Coatzacoalcos mediante la secuenciación de fragmentos de la región V4 del gen 16S del ARNr.
- Identificar si existen diferencias entre las regiones de Perdido y Coatzacoalcos en términos de composición microbiana usando los resultados de la metagenómica dirigida.
- III. Observar posibles diferencias en la composición de la microbiota procarionte en las estaciones de muestreo con respecto a la profundidad de toma de muestra usando herramientas estadísticas y los resultados de la metagenómica dirigida.
- Identificar el potencial metabólico de las comunidades procariontes de los sedimentos marinos de Perdido y Coatzacoalcos mediante PICRUSt y la secuenciación shotgun.

## 2.1. Toma de muestra

Este trabajo fue realizado con muestras de sedimentos superficiales (entre 0-10 cm de profundidad en la columna sedimentaria) colectados durante la campaña oceanográfica Metagenómica – Malla Fina 01 (MMF-01), realizada del 25 de febrero al 18 de marzo del 2016. Se muestreó en 18 estaciones: 9 en la zona de Perdido y 9 en Coatzacoalcos. La profundidad de las estaciones osciló entre 550 y 3,200 m bajo el nivel del mar (Tabla 1, Figura 4). Las muestras se colectaron con un nucleador de caja. Por cada estación se tomaron 9 jeringas de 20 mL. Las muestras se congelaron inmediatamente en nitrógeno líquido durante el crucero. En el laboratorio se almacenaron 80°C hasta su procesamiento.

Región	Estación	Profundidad (m)	Región	Estación	Profundidad (m)
	N1	788		C10	550
	A1	1,000		C11	1,000
	A3	2,500		C12	1,500
Perdido	A4	3,200	Coatzacoalcos	C13	2,500
	B5	550		C14	3,200
	B6	1,000		D15	600
	Β7	1,200		D16	680
	B8	2,400		D17	1,000
	B9	3,200		D18	1,500

Tabla 1. Estaciones de muestreo y profundidad de la columna de agua a la que fueron colectadas. Crucero MMF-01.

**Nota:** A menos que se indique lo contrario, cuando se habla de profundidad se hace referencia a los metros *bajo el nivel del mar* en que fue tomada la muestra.



Figura 4. Mapa de las estaciones muestreadas en la campaña oceanográfica Metagenómica-Malla Fina.

# 2.2. Metagenómica dirigida

### 2.2.1. Extracción de ADN

La extracción de ADN fue realizada a temperatura ambiente a partir de 250 mg de sedimento húmedo. Se realizaron tres extracciones por estación, cada una de una jeringa diferente. La manipulación en el laboratorio de cada una de las jeringas se realizó con guantes nuevos y material estéril. La limpieza de áreas de trabajo se complementó con RNaseZap. Se procuró no mantener las muestras a temperatura ambiente por un tiempo prolongado y una vez tomada la alícuota las jeringas fueron almacenadas nuevamente a -80°C.

La extracción se realizó usando el kit *PowerSoil® DNA Isolation* de Qiagen® siguiendo las instrucciones del fabricante y con la adición de Fenol:Cloroformo:Alcohol isoamílico (25:24:1) con el propósito de mejorar la lisis celular (Anexo A). La concentración del ADN obtenido se cuantificó en un *NanoDrop Lite Spectrophotometer* de Thermo Scientific®, usando como blanco la solución C6 del kit de extracción. Se tomaron como mínimo dos mediciones por cada muestra. La integridad del ADN se observó mediante electroforesis en gel de agarosa al 0.8%.

#### 2.2.2. Preparación de la biblioteca de secuenciación

La preparación de la biblioteca de secuenciación involucró tres etapas: la amplificación de la región V4 del gen 16S del ARNr mediante la Reacción en Cadena de la Polimerasa (PCR, por sus siglas en inglés), la normalización y purificación del producto de PCR, y la formación, la desnaturalización y dilución de la biblioteca.

La amplificación se realizó con la estrategia y primers propuestos por Kozich *et al.* (2013). Se amplifica la región de interés y se añaden los índices y adaptadores en un solo paso de la PCR. Esto disminuye los sesgos asociados a la amplificación que se generan cuando se realizan dos pasos de PCR, así como el número de artefactos de secuenciación. Los *primers* fueron optimizados y usados anteriormente (Covarrubias-Rodríguez, 2016) y están compuestos por el oligo específico para amplificar la región V4 del 16S (propuestos por Klindworth *et al.*, 2013), un *linker*, un oligo de secuenciación (o PAD), un índice (que es distinto para cada muestra) y el adaptador de *Illumina* (Anexo B).

Para facilitar el proceso de amplificación se diluyó el ADN en agua libre de DNasas hasta una concentración final de 3 ng/µL. En cada reacción se usaron 4 µL de buffer de reacción MyTaq<sup>M</sup> (5x), 0.4 µM de cada primer (0.8 µL de cada uno con 10 µM de concentración), 0.1 U de ADN polimerasa MyTaq<sup>M</sup> de Bioline, el volumen equivalente a la concentración de ADN deseado y agua libre de DNasas hasta completar un volumen de 20 µL.

Las reacciones de PCR se prepararon con 20 ng de ADN por reacción y 25 ciclos de PCR (para detalles sobre la reacción de PCR ver el anexo C), sin embargo, en algunos casos no se logró la amplificación

con estas condiciones, por lo que fue necesario disminuir hasta 6 ng el ADN por reacción, e incluso en algunos casos se incrementaron los ciclos de PCR. A este respecto, en este trabajo se decidió usar como máximo 30 ciclos para evitar la sobreamplificación. En los casos en que aún con la modificación de los parámetros anteriores no se obtuvo la región de interés se recurrió al uso de Albúmina de Suero Bovino (BSA, por sus siglas en inglés). Para comprobar que la amplificación se realizó correctamente, las amplificaciones se visualizaron mediate electroforesis en gel de agarosa al 1.5% durante 40 minutos a 80 volts.

Una vez confirmada la obtención de los fragmentos, se procedió a la normalización y purificación de los productos de PCR con el kit *SequalPrep™ Normalization Plate* (*Invitrogen*) siguiendo las indicaciones del fabricante. El protocolo involucró tres pasos principales: el de unión de los amplicones a la placa, el lavado de la misma y la elución de los fragmentos.

Tomando en cuenta la intensidad y tamaño de los fragmentos observados en el gel de agarosa, se procedió a mezclar 10  $\mu$ L de los productos PCR normalizados que fueran más similares. A cada uno de los tubos mezclados se le denomina comúnmente *pool*. La concentración de los diferentes *pool* se cuantificó en un equipo *Qubit*<sup>®</sup> usando el *dsDNA Broad Range Assay kit* de Invitrogen<sup>M</sup> y 2  $\mu$ L de cada *pool*.

Con los valores de concentración de los *pool* se calculó el valor de la concentración de ADN en unidades de molaridad (pM) de cada *pool* usando la fórmula 1. El *pool* final se construyó con una concentración de 2,000 pM, considerando un tamaño de fragmento de 390 bp.

$$[pM] = \frac{\left[\frac{ng}{\mu L}\right]}{\frac{660 \frac{g}{mol} \cdot número \ de \ pb \ del \ fragmento} \cdot 1x10^9} \cdot 1x10^9$$
 Fórmula 1

Finalmente, el *pool* fue desnaturalizado con NaOH, diluido con buffer de hibridación y desnaturalizado con calor siguiendo las indicaciones para el *kit MiSeq® Reagent v2* de Illumina®. Como método de control interno para las bibliotecas de baja diversidad la biblioteca de secuenciación final se preparó a una concentración al 5% de *PhiX*. Se usa *PhiX* como control debido a que las características del genoma del virus (pequeño y diverso) permiten mitigar variaciones en las bibliotecas (Illumina, 2015).

### 2.2.3. Secuenciación

La secuenciación pareada (*paired-end*) de los fragmentos se realizó en la plataforma MiSeq<sup>™</sup> de *Illumina*<sup>®</sup> en el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) y con el uso de *micro*-celdas v2, que tienen un rendimiento de hasta 1.2 Gb de datos. Debido a que se siguió una estrategia diferente a la de *Illumina* para la amplificación del fragmento de interés se añadieron al kit de reactivos los primers necesarios para llevar acabo la secuenciación. Finalmente, se proporcionó a la plataforma un archivo de texto (separado por tabulaciones) donde se indicó el nombre de cada muestra y el par de índices que la identificaban, esta información es usada para la eliminación de los artefactos de secuenciación y para la separación de las secuencias de cada muestra en un archivo distinto.

### 2.2.4. Procesamiento informático de las secuencias

Las secuencias sin artefactos pueden ser descargadas de la plataforma de almacenamiento (*BaseSpace Sequence Hub*) de *Illumina*. Las secuencias sentido (*forward*) y las antisentido (*reverse*) por muestra son devueltas en formato *.fastq* en archivos separados. Para identificar la calidad (*Q-score, QS*) de los amplicones, corroborar su longitud y saber el número de las secuencias por muestra se usó el software *FastQC*.

#### 2.2.4.1. Pre-procesamiento de las secuencias

Para poder obtener información biológica de las secuencias es necesario procesarlas con programas computacionales. Debido a que los archivos a procesar sobrepasaban en conjunto los 2 Gb, todo el procesamiento informático se realizó en el *cluster* de cómputo del CICESE. Todos los comandos usados fueron activados mediante el uso de archivos *slurm*.

El pre-procesamiento involucró la unión de la secuencia sentido y la antisentido, el filtrado de las secuencias y la eliminación de quimeras. La unión de las secuencias (sentido y antisentido) se realizó con el comando *fastq\_mergepairs* de la herramienta de análisis *USEARCH;* se permitieron como máximo 10 *mismatches* en el alineamiento y una longitud mínima y máxima de 200 y 300 nucleótidos (respectivamente) para la unión (Edgar *et al.,* 2011). Todas las secuencias ya unidas se concentraron en un solo archivo con el uso de la opción *relabel @* que añade el nombre de las muestra en la etiqueta asignada a la lectura.

El archivo en formato .*fq* fue filtrado y transformado a formato .*fasta* con el comando *fastq\_filter*. Los puntos (.) encontrados en este archivo fueron cambiados por guiones bajo (\_) (usando *sed 's/\./\_/g'*) para poder continuar con el procesamiento.

Finalmente, la identificación y eliminación de quimeras se realizó con el comando *uchime*, tomando como referencia las secuencias del gen 16S del ARNr depositadas en la base de datos *gold* (https://sourceforge.net/projects/microbiomeutil/files/).

#### 2.2.4.2. Obtención de las Unidades Taxonómicas Operacionales

El agrupamiento de OTUs y la generación de la tabla de los mismos se realizó con comandos de *QIIME (Quantitative Insights Into Microbial Ecology)* (Caporaso *et al.,* 2010). Esto consistió en agrupar las secuencias que tuvieran un 97% de identidad; esto significa que las secuencias se comparan unas con otras y se agrupan aquellas que difieran como máximo en un 3% de sus bases.

La agrupación de los OTU se realizó *de novo* con una serie de comandos (*pick\_otus.py*, *pick\_rep\_set.py*, *align\_seqs.py*, *make\_phylogeny.py*) y la asignación taxonómica posterior se realizó con el comando *assign\_taxonomy.py* y con referencia a la base de datos SILVA\_128 (Quast *et al.*, 2013). El árbol filogenético obtenido se proyectó en forma de dendograma usando el software *FigTree* versión *1.4.3*. La tabla de OTU se generó con *make\_otu\_table.py* y fue filtrada para eliminar las secuencias de mitocondrias y cloroplastos. El número de secuencias que se procesaron y de OTU encontrados por muestra fue obtenido mediante el comando *biom summarize-table* (Caporaso *et al.*, 2010).

#### 2.2.4.3. Análisis de la diversidad microbiana

Los datos de diversidad microbiana y sus abundancias relativas promedio en las muestras se obtuvieron con el comando *summarize\_taxa\_through\_plots.py*, para ello se proporcionó un archivo con los metadatos disponibles de las muestras y la librería de Python para la generación de los gráficos (*matplotlib*).

Mediante el comando *alpha\_rarefaction.py* se obtuvieron las curvas de rarefacción de la diversidad microbiana con respecto al número de secuencias por muestra (profundidad de secuenciación).

Estas curvas permiten visualizar el número mínimo de secuencias que se necesitan para identificar el mayor número posible de especies presentes en la muestra. Los índices de diversidad alfa (Shannon (diversidad) y Chao 1 (especies esperadas)), diversidad filogenética y el número de especies observadas fueron obtenidos con *alpha\_diversity.py* tomando como referencia el árbol filogenético de la base de datos SILVA\_128 (Yilmaz *et al.,* 2013).

Para los análisis de diversidad beta se obtuvieron matrices de distancia (UniFrac) (Lozupone y Knight, 2005) con el comando *beta\_diversity\_through\_plots.py*, tomando como valor de corte el número de secuencias obtenidas por la muestra con menor profundidad de secuenciación (19,072) y el árbol filogenético generado durante la identificación de las OTU. Los Análisis de Coordenadas Principales (PcoA, por sus siglas en inglés) en dos dimensiones se generaron con el comando *make\_2d\_plots.py* de la matriz ponderada (que toman en cuenta las abundancias relativas de los taxones) y no ponderada (basadas únicamente en la presenta/ausencia de los taxones).

Para la visualización de las OTU compartidas entre estaciones de muestreo se realizó un diagrama de Venn en versión *UpSet R*, el cual permite la visualización de las intersecciones entre más de tres sets de datos (Lex *et al.,* 2014). Los gráficos fueron obtenidos en la plataforma online del software (<u>http://gehlenborglab.shinyapps.io/upsetr/</u>), para poder realizar esto, es necesario transformar la tabla de OTU de formato *.biom* a *.txt* usando el comando *biom convert* con las opciones *--to-tsv* y *--header-key taxonomy.* De esta manera la tabla puede ser modificada con un editor de texto (Excel o similar) para convertirla en una tabla que indique únicamente la presencia/ausencia (0 y 1) de los OTU. También es necesario eliminar la columna correspondiente a la asignación taxonómica así como la primera fila de la tabla que indica que los datos corresponden a una tabla de OTU.

### 2.2.4.4. Tests estadísticos

Para analizar la significancia estadística entre las estaciones de muestreo se realizaron dos test. Por un lado se realizó la evaluación con ANOSIM y por otro a través de PERMANOVA de factores como la zona de muestreo (Perdido o Coatzacoalcos), la profundidad bajo el nivel del mar de la toma de muestra y las réplicas por estación de muestreo siendo la entrada principal las matrices de distancia generadas durante el análisis de beta diversidad, un valor de p<0.05 se consideró estadísticamente significativo.

### 2.2.5. Inferencia funcional usando PICRUSt

La inferencia del potencial metabólico de los microorganismos se realizó mediante el software PICRUSt (*Phylogenetic Investigation of Communities by Reconstruction of Unobserved States*) (Langille *et al.,* 2013). Este software está diseñado para predecir la función de los microorganismos contenidos en un metagenoma de amplicones a partir de los genomas completos que se encuentren disponibles en la base de datos *GreenGenes* (McDonald *et al.,* 2012). Para el uso de esta herramienta fue necesario generar la tabla de OTU con referencia cerrada a la base de datos de *GreenGenes*. La tabla fue filtrada para eliminar cloroplastos y mitocondrias y se transformó en una tabla de presencia-ausencia de OTU.

El análisis de la tabla con la versión en línea de PICRUSt (*Galaxy*, opción *Langille Lab*) se divide en tres pasos: en el paso uno se realizó la normalización del número de copias del gen 16S y el cálculo del NSTI (*Nearest Sequenced Taxon Index*), en el paso dos se predijeron los metagenomas de las estaciones con respecto a los metagenomas completos secuenciados de las especies presentes y en el paso tres se categorizaron las posibles funciones de los microorganismos tomando en cuenta la base de datos *KEGG* (*Kyoto Encyclopedia of Genes and Genomes*) (<u>http://galaxy.morganlangille.com/</u>).

### 2.3. Metagenómica shotgun

Tomando en cuenta los resultados de la metagenómica dirigida se seleccionaron tres estaciones para realizar la secuenciación del metagenoma completo. Las estaciones seleccionadas fueron C10, C13 y D18, ubicadas en la región de Coatzacoalcos (a 550, 2500 y 1500 m de profundidad, respectivamente). La elección de estas estaciones se realizó tomando en cuenta dos factores: el perfil de diversidad obtenido con la metagenómica dirigida y la información proporcionada (comunicación directa) con los encargados del muestreo. Fue comentado que la estación C10 se ubicó en sitios cercanos a las tuberías de extracción de hidrocarburos de PEMEX y que la estación D18 se planteó inicialmente con la intención de muestrear una chapopotera. La estación C13 se eligió debido a que el perfil de diversidad de las muestras a más de 2,000 m bajo el nivel del mar es similar y por tanto la información obtenida de esta estación podría ser representativa.

### 2.3.1. Concentración del ADN

La calidad del ADN que se extrajo durante este trabajo fue mayoritariamente de alto peso molecular (10,000 bp), por lo que se decidió usar el mismo para la preparación de la biblioteca *shotgun*. Para garantizar la pureza de la muestra se trató con Ribonucleasa A (*stock* con concentración de 100 µg/mL) al 10% en volumen de ADN por tratar que posteriormente se limpió y concentró con el kit *Zymo: Genomic DNA Clean & Concentrator* ™. La presencia e integridad del ADN obtenido se confirmó con electroforesis en gel de agarosa al 0.8% a 80 volts durante 30 minutos.

### 2.3.2. Preparación de la biblioteca

El ADN purificado fue cuantificado por fluorescencia en un *Qubit*<sup>®</sup> usando el *dsDNA Broad Range Assay kit* de Invitrogen<sup>M</sup> y 2 µL de muestra. Considerando los valores obtenidos se hicieron diluciones del ADN hasta una concentración de 2 ng/µL que se cuantificaron y diluyeron nuevamente hasta 0.2 ng/µL.

Para la preparación de la biblioteca el ADN es fragmentado aleatoriamente y acto seguido se ligan a sus extremos 3' y 5' los adaptadores necesarios para la secuenciación; a este proceso se le denomina *tagmentación*. La biblioteca de secuenciación fue construida siguiendo las instrucciones del kit *Nextera XT Library Preparation* de *Illumina* (Illumina, 2015) con una modificación en el proceso de limpieza (protocolo no publicado; Dra. Chong-Robles). Para visualizar los tamaños de fragmento obtenidos se realizó electroforesis en gel de agarosa al 2% durante 40 minutos a 80 volts.

Tomando en cuenta la intensidad y tamaño de los fragmentos observados en el gel se realizaron los cálculos para construir la biblioteca con una concentración final de 13 pM (Formula 1). La biblioteca de secuenciación se preparó siguiendo las indicaciones del kit *Nextera XT Library Preparation* usando un 2% de *PhiX*.

#### 2.3.3. Secuenciación shotgun

La secuenciación de la biblioteca se realizó en un solo sentido (*single-end*) con el kit *v3* (150 ciclos) para la plataforma *MiSeq* de *Illumina*, con el objetivo de obtener hasta 25 millones de secuencias. La secuenciación fue realizada en el equipo de secuenciación perteneciente a CICESE.

### 2.3.4. Procesamiento informático de las secuencias

Los datos de la secuenciación *shotgun* fueron descargados de la plataforma de almacenamiento de *Illumina* y colocados en el *cluster* –*OMICA* de CICESE. Para visualizar la calidad de los fragmentos obtenidos, su tamaño y el posible contenido de adaptadores se usó la herramienta *fastqc* instalada en el cluster (Andrews, 2010). A continuación, se realizó el cortado (*trimming*) de los adaptadores de secuenciación de cada uno de los archivos obtenidos, usando la opción *single-end* del programa *Trimmomatic* y el archivo *TruSeq2-SE.fa* para el reconocimiento de los adaptadores correspondientes al kit usado para la secuenciación (Bolger *et al.,* 2014). Al terminar se verificó nuevamente la calidad de los fragmentos y que se hubieran eliminado los adaptadores correctamente.

El ensamblaje de los fragmentos en secuencias más largas (*contigs*) se hizo con *Megahit* (Li *et al.,* 2015). El ensamblaje (en formato *.fa*) realizado se evaluó con *Quast* (Gurevich *et al.,* 2013) y posteriormente fue anotado usando la herramienta *Prokka*, que permite la anotación rápida de genomas de procariontes (Seemann, 2014). Finalmente, el archivo resultante de este procesamiento en formato *.faa* fue descargado y cargado a la plataforma *KEGG Ghost Koala* en su versión e línea (Kanehisa *et al.,* 2016).

## 3.1. Metagenómica dirigida

### 3.1.1. Extracción de ADN

El ADN se extrajo de tres jeringas de sedimento (SD) por estación de muestreo. Por cada extracción se utilizaron 250 mg de sedimento. La concentración del ADN extraído fue distinta entre réplicas y estaciones, teniendo un valor promedio global de 4.33 ng de ADN/mg de sedimento, con un mínimo de 1.22 ng/mg en la muestra A3c y un máximo de 8.08 ng/mg en la muestra C10a (Tabla 2 y 3). La presencia e integridad del ADN extraído se confirmó mediante electroforesis en geles de agarosa (Figura 5). En todas las réplicas se observaron fragmentos de ADN de alto peso molecular y con integridad suficiente para el procesamiento planeado.

Muestra	A260/280	ng de ADN/mg SD	Muestra	A260/280	ng de ADN/mg SD
A1a	1.97	2.92	B6c	1.63	6.36
A1b	2.02	4.84	B7a	2.32	2.36
A1c	1.54	2.48	B7b	1.66	3.66
A3a	1.33	3.72	B7c	1.77	3.94
A3b	1.81	3.86	B8a	2.05	4.98
A3c	1.74	1.22	B8b	1.65	3.66
A4a	2.5	1.86	B8c	1.69	3.44
A4b	2.3	1.92	B9a	2.02	3.54
A4c	1.91	4.06	B9b	2.17	2.18
B5a	2.16	3.34	B9c	1.92	3.82
B5b	2	3.38	N1a	2.19	4.02
B5c	1.68	3.48	N1b	1.74	1.7
B6a	1.92	4.54	N1c	1.25	4.6
B6b	1.87	6.34			

Tabla 2. Valor promedio de la concentración y pureza del ADN extraído de cada muestra de Perdido.

Muestra	A260/280	ng de ADN/mg SD	Muestra	A260/280	ng de ADN/mg SD
C10a	1.9	8.08	C14c	1.66	4.14
C10b	1.86	7.8	D15a	1.84	5.92
C10c	1.81	6.04	D15b	1.83	5.8
C11a	1.87	5.772	D15c	1.69	4.6
C11b	1.93	5.5	D16a	1.88	6.68
C11c	1.92	5.78	D16b	1.78	7.12
C12a	2.06	4.26	D16c	1.68	7.56
C12b	2.00	4.62	D17a	2.15	3.68
C12c	1.87	4.7	D17b	1.79	3.18
C13a	2.06	3.12	D17c	2.29	3.18
C13b	1.82	4.08	D18a	2.08	4.32
C13c	1.59	5.04	D18b	1.54	2.06
C14a	1.8	3.82	D18c	1.21	7.32
C14b	1.76	3.32			

Tabla 3. Valores de las concentraciones y purezas del ADN extraído de las muestras de Coatzacoalcos.



**Figura 5.** Gel del agarosa del ADN extraído de algunas de las muestras de sedimentos de las regiones de Perdido y Coatzacoalcos. Se cargaron 5 µL del ADN en un gel de agarosa al 0.8% (80 V por 35 min).

### 3.1.2. Amplificación de la región de interés

La región V4 del gen 16S del ARN ribosomal se amplificó correctamente en todas las muestras, obteniéndose el tamaño del fragmento esperado (~390 pb) (Figura 3.2). Se logró amplificar la región de interés con 18 ng, 15 ng, 12 ng y 6 ng de ADN, con 25, 27 y 30 ciclos y usando BSA en los casos en que no se logró la amplificación con la modificación de concentración o ciclos de PCR (Anexo D).



**Figura 6.** Electroforesis del producto de PCR de algunas de las réplicas realizadas, que confirmó la amplificación del fragmento de interés. 3 µL del producto de la PCR se corrieron en geles de agarosa al 2.0% (80 V por 35 min).

### 3.1.3 Secuenciación y procesamiento informático

La secuenciación de los amplicones se realizó en tres corridas distintas. Las secuencias de todas las muestras tuvieron un *Quality Score* superior a 25. Como resultado del pre-procesamiento se obtuvieron 2,917,498 secuencias pareadas y libres de quimeras, en las cuales se pudieron identificar 55,264 OTU distintas.

El mayor número de OTUs fue obtenido para las réplicas C14c (3,117 OTUs a partir de 49,256 secuencias) y B5b (3,193 OTUs de 19,072 secuencias). La réplica con mayor número de secuencias fue C11c con 118,774 de las que se obtuvieron 7,293 OTUs (Figura 7).


**Figura 7.** Número de secuencias procesadas y OTU asignadas por estación. En el eje y se presenta el número de secuencias u OTU en escala logarítmica (base 2).

Aunque intuitivamente se podría pensar que de las réplicas con mayor número de secuencias se obtendría un mayor número de OTU, en la figura 3.4 se puede observar que esto no es así necesariamente. Por ejemplo, en las réplicas A4c y D16b se obtuvieron 63,899 y 64,075 secuencias respectivamente, sin embargo en D16b se identificaron 3,571 OTU más que en A4c (Figura 8).

Las curvas de rarefacción obtenidas (Figura 9) muestran que se la profundidad de secuenciación fue suficiente para identificar un gran número de especies presentes en las muestras y que un valor de ~20,000 secuencias fue suficiente para la realización de los análisis comparativos, al observarse una tendencia a la asíntota en este valor.



Figura 8. Número de OTU asignadas con respecto al número de secuencias obtenidas.



Figura 9. Curvas de rarefacción que indican el número de especies esperadas (Chao1) con respecto a la profundidad de secuenciación.

Los índices de diversidad de especies esperadas y observadas permitieron identificar que las estaciones de profundidades superiores a 2,000 m tienen un índice de riqueza de especies similar. Además, se puede observar que los sedimentos de las estaciones D16, C11 y D18 son más diversos que el resto de las estaciones; siendo esta última la que presenta una mayor diferencia entre el número de OTU observados y esperados (Figura 10).



**Figura 10.** Número de especies observadas y esperadas (Chao1) por estación de muestreo. Las réplicas están ordenadas con respecto a los metros bajo el nivel del mar en que fueron tomados los sedimentos analizados.

A este respecto, de las OTU identificadas por estación de muestreo, D18 es la que tiene un mayor número de OTU únicas (3,645), seguida por D16 con 3,170 y C11 con 2,696 (Figura 11). Así mismo, las estaciones de profundidades superiores a 2,000 m son las que tienen un menor número de OTU únicos. Todas las estaciones comparten 467 OTU, siendo D18 y C10 las que más OTUs comparten entre sí.



**Figura 11.** Número de OTUs únicas y compartidas entre las estaciones de muestreo. El número total de OTUs identificadas por estación se indica en las barras horizontales del lado izquierdo. Las OTUs únicas y compartidas se indican en las barras verticales. Los puntos aislados señalan la barra de OTUs únicas por estación y las que están unidas por líneas indican el número de OTUs compartidas entre los puntos unidos. El gráfico fue realizado en la plataforma en línea del software UpSetR.

#### 3.1.4. Caracterización taxonómica de la microbiota de las zonas de estudio

Aunque se trató de identificar hasta el nivel de especie las OTUs formadas, la asignación taxonómica se visualizó correctamente hasta el nivel de orden. Esto es debido a que los microorganismos en los sedimentos marinos han sido poco estudiados y las variantes específicas de su gen 16S que permiten discernir entre una especie y otra no se encuentran disponibles, dificultando por consiguiente la identificación en las categorías taxonómicas más bajas.

Se identificaron 81 fila diferentes en las zonas de muestreo. En Perdido, se identificaron 81 y en Coatzacoalcos 77; de éstos, 13 y 14 fila (respectivamente) están presentes en porcentajes mayores al 1%. De estos fila, 21 son Archaea y el resto Bacteria. En promedio en las estaciones de muestreo, el filo más abundante es el de las Proteobacteria (30%), seguido de las Thaumarchaeota (14%), Planctomycetes (13%), Acidobacteria (8%) y Chloroflexi (8%) (Figura 12).

Considerando los promedios de este nivel taxonómico, la diferencia más notoria entre estaciones es que en Coatzacoalcos el porcentaje de Thaumarchaeota es menor que en Perdido (12% y 17%, respectivamente) y que el de Planctomycetes es mayor (15% y 12%, respectivamente). La composición taxinómica del área es similar, el porcentaje de abundancia relativa de los fila Proteobacteria, Thaumarchaeota, Planctomycetes, Acidobacteria, Chloroflexi y Nitrospirae es muy cercano entre ambas regiones por lo que podría concluirse que los sedimentos vistos como hábitat microbiano son homogéneos.



**Figura 12.** Abundancia relativa de los fila procariontes en las dos zonas de muestreo. Se presenta un promedio de la abundancia relativa realizado con todas las réplicas de las estaciones de muestreo. Se muestran los fila con abundancia relativa mayor al 1%, el resto se agrupa en las categoría otros. En la parte superior se identifica a los fila de bacteria y en la inferior a los de Archaea (en tonalidades verdes).



**Figura 13.** Porcentajes de abundancia relativa promedio de los microorganismos procariontes identificados en las réplicas por estación de muestreo a nivel filo. Del lado derecho se presentan las estaciones muestreadas en Perdido y del izquierdo las de Coatzacoalcos. Los fila en tonalidades verdes de la parte inferior corresponden al dominio Archaea y el resto al dominio Bacteria. Los fila en porcentaje de abundancia menor a 1% se agruparon en las categorías otros.

Sin embargo, cuando se observan los perfiles de diversidad de cada una de las estaciones de muestreo, se hacen notorio que existen diferencias en las abundancias relativas de los filos. Aunque existen diferencias en los porcentajes de abundancia relativa, los fila más abundantes parecen ser los mismos en las estaciones excepto en D18, que tiene el menor porcentaje de Thaumarchaeota y Proteobacteria y una mayor abundancia de Chloroflexi (Figura 13).

El Análisis de Coordenadas Principales (PCoA) que se realizó tomando a la zona de muestreo como la principal fuente de variación no mostró que las OTUs de Perdido se agruparan separados de las de Coatzacoalcos (Figura 14). Los valores p obtenidos de los test estadísticos (PERMANOVA y ANOSIM) fueron de 0.03 (pseudo-F=2.76) y 0.06 (R=0.08). Con base en estos resultados se puede mencionar que el efecto de la zona de muestreo en la diversidad de especies no es estadísticamente significativo.



**Figura 14**. Análisis de Coordenadas Principales (PCoA) de las réplicas considerando a la zona de muestreo como la principal fuente de variación. Los puntos azules corresponden a las réplicas de Perdido y los rojos a Coatzacoalcos. a) PCoA de la matriz no ponderada (matriz de presencia-ausencia) y b) PCoA de la matriz ponderada (que considera la abundancia relativa).

Posteriormente, los mismos análisis fueron realizados considerando la profundidad a la que fue tomada la muestra como la principal fuente de variación. Tanto en el análisis de ANOSIM como el de PERMANOVA se obtuvo un valor de p igual a 0.01 (valor pseudo-F igual a 2.39 y R de 0.5), que indica diferencias estadísticamente significativas con respecto a esta variable. Estas diferencias pueden observarse claramente en el PCoA de la figura 15. Pese a que la estación D18 se ordena de manera dispar en el PCoA, se puede observar cómo las réplicas parecen distribuirse en relación a la profundidad a que fueron muestreadas. Las agrupación más evidente se observa entre las muestras a profundidades mayores a los 1,200 m. Las estaciones ubicadas entre 550 y 788 m de profundidad se ordenan juntas aunque no tan cercanas como las más profundas; y las ubicadas a 1,000 m de profundidad se ordenan de manera tal que parecen unir a las zonas más profundas con las someras.



**Figura 15**. Análisis de Coordenadas Principales (PCoA) de las réplicas considerando a la profundidad de muestreo como la principal fuente de variación. Del lado derecho se presenta el PCoA de la matriz no ponderada y del lado izquierdo la de los datos ponderados.

Esta diferencia de hace notoria en la figura 16, en donde se presenta un promedio de las abundancias relativas de las clases de las tres réplicas. Se puede observar que las clases con porcentaje de abundancia relativa mayor al 1% identificadas por estación de muestreo de menor (550 m) a mayor (3,200 m) profundidad. Para facilitar la visualización se realizó el promedio de las tres réplicas por estación y las clases con abundancia relativa menor a 1% se agruparon en la categoría *otros*. De las 352 clases identificadas 17 contribuyen con más de 1% a la abundancia relativa.

En la figura 16 se puede observar que la abundancia de los miembros del *Marine Group I* incrementa a mayor profundidad de muestreo, de un 6% al 22% en 550m y 3,200m, respectivamente. Las *Deltaproteobacteria* parecen mantenerse constantes (entre 9-13%) y las *Alphaproteobacteria*  incrementan sutilmente sus porcentajes (de 4 a 12%). Por el contrario, miembros de las *Planctomycetacia* disminuyen en abundancia al incrementar la profundidad, de un 10% a un 2%.

Sin embargo, la estación D18 no sigue el mismo patrón que el resto. El porcentaje del *Marine Group I* y *Alphaproteobacteria* es de menos del 1% y los porcentajes de *Anaerolineae* y *Phycisphaera* son del 10% cada uno. Además, en esta estación, el porcentaje de OTU no asignadas es del 6%, el mayor de todas las estaciones.

Ante estas diferencias y para poder obtener una mejor visualización de los órdenes identificados, las estaciones se separaron en tres grupos: I) estaciones entre 550 y 800 m de profundidad, II) estaciones entre 1,000 y 2,000 m de profundidad y III) estaciones entre 2,000 y 3,200 m de profundidad.



**Figura 16**. Abundancia relativa promedio a nivel clase de los microorganismos procariontes identificados en las estaciones de muestreo ordenadas de menor a mayor profundidad. Las tonalidades verdes corresponden a clases de arqueas y el resto a bacteria. Se presentan las clases con abundancia relativa promedio mayor a 1%, el resto se agrupan en las categorías Otros Bacteria y Otros Archaea.

#### I. Composición de la microbiota de los sedimentos entre 550 y 800 m de profundidad

Se identificaron 578 órdenes, de los cuales 24 aportan más del 1% a la abundancia relativa. En promedio, los *Brocadiales* (8%), *Xhantomonadales* (5%), *NB1-j* (4%), *MSBL9* (4%) y *Rhodospirillales* (4%) son los órdenes más abundantes en estas profundidades. Órdenes del *Marine Group I* (desconocidos, no cultivados y otros) suman el 5% (Figura 17) y los miembros del *Marine Benthic Group E* son más abundantes que en el resto de las estaciones (Anexo K).



**Figura 17.** Porcentaje de abundancia relativa promedio a nivel orden de los microorganismos procariontes identificados en las estaciones entre 550 y 800 m de profundidad. Se presentan los órdenes con porcentajes de abundancia mayores al 1%, el resto se agrupan en la categoría Otros. En la parte superior se indica la profundidad de la toma de muestra en metros.

En D16, el porcentaje de *Xhantomonadales* es del 8%, los *Brocadiales* se encuentran en un 7% de abundancia relativa promedio y NB1-j en un 6%. En la estación C10, el porcentaje de los miembros de los *Brocadiales, Xhantomonadales, MSBL9,* No asignados y *Anaerolineales* es del 5% en cada uno. En esta estación, miembros de los *Desulfobacterales* y *Syntrophobacterales* se encuentran presentes en un 2.3%, cada uno. En N1, los *Brocadiales* son los más abundantes (10%), seguidos de los *Nitrospirales* (5%) y MSBL1 (4.5%).

En estas estaciones de muestreo se identificó que D16 es la que tiene el mayor número de OTU únicas (5,780), seguida por C10 con 3,622. Las cinco estaciones comparten 2,233 y C10 y D16 son las que comparten más OTU entre sí (1,060) (Anexo E).

En la estación C10, las familias más abundantes son las *MSBL9: no cultivada, Brocadiacea* y *Plancomicetacea* (pertenecientes al filo de los Planctomycetes) que suman el 13%. Las familias de Proteobacteria (*JTB255* (5%), *Rhodospirillaceae* (3%), *Syntrophobacteraceae* (2%), *Desulfobacteracea* (2%), *Deltaproteobacteria* (*NB1-j*, 4%) y algunas *Gammaproteobacteria: no cultivadas* (2%) suman un 18%. Los miembros de Chloroflexi (*Anaerolineaceae*) tienen el mismo porcentaje que las OTU no asignadas (5%). El resto de las familias se encuentran en porcentajes menores al 2%.

#### II. Composición de la microbiota de los sedimentos entre 1,000 y 1,500 m de profundidad

Las estaciones A1, B6, C11 y D17 se muestrearon a 1,000 m de profundidad; B7 a 1,200 m y A3, C12 y D18 a 1,500 m. De ellas, A1 y D18 fue donde se observaron mayores diferencias en los porcentajes de abundancia en las OTU identificadas.

En general, en estas profundidades el orden más abundante es el de los *Rhodospirillales* (7%), los órdenes del *Marine Group I* (no cultivados, desconocidos, otros y ambiguos) suman un 15%, los *Xhantomonadales* y los miembros NB1-j suman un 5% cada uno (Figura 18). A excepción de A1 y D18, las estaciones a estas profundidades tienen a estos órdenes en porcentajes de abundancia relativa muy cercanos.



**Figura 18.** Abundancia relativa promedio a nivel orden de los microorganismos procariontes identificados en las estaciones entre 1,000 y 1,500 m de profundidad. Se presentan los órdenes con porcentajes de abundancia mayores al 1%, el resto se agrupan en Otros.

En la estación A1, los *Brocadiales* suman un 5% al igual que los *Rhodospirillales y los NB1-*, porcentajes que son similares a los observados en las estaciones de menores profundidades. Por otro lado se hace aún más evidente que D18 es diferente al resto de las estaciones. En ella, el porcentaje del *Marine Group I* es muy cercano al 0, los miembros del MSBL9 suman el 6% y los *Anaerolineales* el 10%. En este sentido, la estación D18 tiene 6,599 OTU únicas y comparte con el resto de las estaciones únicamente 733 (Anexo F).

La estación B7 (1,200 m) parece seguir la tendencia del incremento de miembros de las Thaumarchaeota (*Marine Group I:* no cultivadas, deconocidas, otras; 19%) y un alto porcentaje de Proteobacteria correspondiente a familias de *Alphaproteobacteria (Rhodospirillaceae,* 9%), Gammaproteoacterias (JTB255, 3%; no cultivadas, 2%, y Xhantomonadales, 2%) y Deltaproteobacterias (NB1-j, 4% y Desulfurellaceae, 1%).

Por otro lado, la estación D18 tiene un porcentaje menor al 1% de miembros de la *Marine Group I* y *Alphaprotecobacteria*, las *Deltaproteobacteria* (*Desulfobacteraceae, 4*%; *Syntrophobacteraceae, 2*%; *Desulfobulbaceae, 2*%) suman un 8%, las *Anaerolineaceae* (*Chloroflexi*) son la familia más abundante (10%) y las familias de *Planctomycetes* (*MSBL9, 6%,* y Plantomycetaceae, 2%) aportan el 6%. Además, se pueden identificar en porcentajes mayores al 1% otras familias que en el resto de las estaciones se encuentran en porcentajes casi nulos (*Aminicenantes (2%), Woesearchaeota* (2%) y *Lokiarchaeota* (2%), por ejemplo).



#### III. Composición de la microbiota de los sedimentos entre 2,000 y 3,200 m de profundidad

**Figura 19.** Abundancia relativa promedio a nivel orden de los procariotas identificados en las estaciones entre 2,000 y 3,200 m de profundidad. Se presentan los órdenes con porcentajes de abundancia mayores al 1%, el resto se agrupan en Otros.

Las estaciones muestreadas entre 2,000 y 3,200 m de profundidad parecen ser más homogéneas con respecto a los porcentajes de abundancia relativa promedio de los órdenes identificados (Figura 19). En general, los *Rhodospirillales* son el orden dominante en estas profundidades estando en promedio de abundancia relativa un 11%. Las *Marine Group I* en sus distintas familias (no cultivadas, desconocidas, ambiguas y otras) se encuentran en un 22%, las *Deltaproteobacteria* (NB1-j, *Desulfurellales* y *Myxococcales*) están en un 7% de abundancia promedio. El porcentaje de *Nitrospirales* es cercano al 3%, los miembros de los *Planctomycetes* suman el 2% total y los de *Chloroflexi* un 6%. El número de OTUs compartidas entre estas estaciones es de 2,055. La estación A4, es la que tiene un el mayor número de OTU únicas (2,106) y la de menor número es C14 con 1,348 (Anexo G).

En la estación C13, se identificó el mayor porcentaje de *Euryarchaeota* (*Thermoplasmata*, 5%). Las *Proteobacteria* más abundantes son las *Alpha*- (10%), *Gamma*- (5%) y *Delta*- (4%) y las familias de *Thaumarchaeota* se encuentran en un 20%.

#### 3.1.5. Inferencia de la función de la microbiota de los sedimentos de Perdido y Coatzacoalcos

Con el uso de PICRUSt y la tabla de OTU generada con referencia a la base de datos *GreenGenes* se realizó la inferencia de la función que los microorganismos podrían estar desempeñando en los sedimentos de las regiones de Perdido y Coatzacoalcos. Sin embargo, es importante tomar en cuenta que la inferencia de la función que tienen los microorganismos en ambientes poco caracterizados no es recomendable.

Para la utilización de PICRUSt se normalizó la tabla de OTU (tomando en cuenta el número de copias del gen 16S que puede haber en el genoma de los microorganismos) y se obtuvo el valor del *Nearest Sequenced Taxon Index* (NSTI) que indica la disponibilidad de los genomas para cada estación. Estos valores oscilaron entre 0.16 y 0.29 e indican que para las OTU identificadas en las estaciones se pueden encontrar hasta un 84% de los genomas completos. En general, la exactitud de PICRUST disminuye con valores más altos de NSTI (mayores a 0.10) por lo cual la evaluación funcional debe hacerse con cautela (Langille *et al*, 2013).

Como resultado del análisis, se obtuvo que es posible que los microorganismos identificados tengan genes principalmente relacionados con el metabolismo (51%) y el procesamiento de información genética (18%). Los genes relacionados al procesamiento de información ambiental podrían estar

representados en un 11% y los ligados a enfermedades humanas en un porcentaje cercano al 1%. El 14% de los posibles genes no pudieron ser clasificados y por tanto no se pueden hacer aseveraciones concluyentes (Figura 20).



**Figura 20.** Categorías funcionales inferidas en los metagenomas de amplicones de las muestras analizadas. Se muestra la composición porcentual de las 7 categorías del nivel 1 obtenidas en PICRUSt.

De todos los posibles genes, el 11% están implicados en el metabolismo de los aminoácidos, el 10% en el metabolismo de carbohidratos, el 9% en el transporte de membrana y el 7% en la replicación y reparación del material genético. El 5% de los genes están pobremente caracterizados, el 2% están relacionados con la biodegradación y metabolismo de xenobióticos y el 2% con el metabolismo de terpenoides y policétidos (Anexo I). Sin embargo, debido a que este tipo de muestras ha sido poco estudiado, el porcentaje de microorganismos con un genoma completo secuenciado y disponible es bajo, por lo que esta inferencia funcional no es confiable.

#### 3.2. Secuenciación shotgun

Con base en la asignación taxonómica realizada con la metagenómica dirigida y las diferencias observadas entre estaciones de muestreo a distintas profundidades se eligieron muestras de ADN de las

estaciones C10 (550 m), D18 (1,500 m) y C13 (2,500 m) para realizar la secuenciación *shotgun* y con ello identificar el potencial metabólico de los microorganismos.



**Figura 21.** Clases de microorganismos procariontes con abundancia relativa mayor al 1% en los sedimentos de las estaciones a) C10, b) D18 y c) C13. En la categoría otros se agrupan en resto de las familias identificadas.

#### 3.2.1. Secuenciación y el procesamiento informático

La biblioteca de secuenciación se construyó con ADN que tuvo valores de concentración iguales a 11.3 ng/ $\mu$ L, 5.3 ng/ $\mu$ L y 5.9 ng/ $\mu$ L, para C10, C13 y D18, respectivamente después de la limpieza y concentración.

Como resultado de la secuenciación se obtuvieron 4.36 Gb de datos en total, que corresponden a 14,510,746 lecturas. De estas, para C10 se obtuvieron un 48.4%, para C13 un 30.1% y para D18un 9.09%. La longitud de las lecturas osciló entre 35 y 151 bp con valores de calidad superiores a 28. En las tres muestras se obtuvieron fragmentos de tamaños entre 31 y 151 bp con valores de calidad mayores a 28 antes del procesamiento con *Trimmomatic;* después del cortado, el valor de calidad de las secuencias fue mayor 30 (Tabla 4).

El archivo de la anotación de los contigs obtenido con *Prokka*, para cada una de las estaciones, se cargó en *KEGG Ghost Koala* para la caracterización funcional de las secuencias. De las 2,627 entradas

41

Herramienta de evaluación	Archivo procesado	Característica	C10	C13	D18
FastQC -	Secuencias crudas	Número de secuencias	7,027,036	4,367,831	1,320,069
		Longitud (bp)	35-151	35-151	35-151
		QC	>33	>28	>32
	Secuencias curadas	Número de secuencias	6,966,160	4,355,617	1,317,287
		Longitud (bp)	35-151	35-151	35-151
		QC	>34	>30	>32
QUAST		# contigs	2,004	7,626	329
		# contigs (>= 0 bp)	12,401	28,355	3,173
	Ensamblaje	# contigs (>= 1000 bp)	103	1,089	12
	(Megahit)	# contigs (>= 5000 bp)	0	0	1
		contig más largo	1,924	4,211	5,021
		Longitud total	1,324,819	5,810,450	209,807

 Tabla 4. Tabla de resultados generales del procesamiento de las secuencias shotgun.

La caracterización funcional indicó que el mayor porcentaje de genes identificados están relacionados con el procesamiento de la información genética (Figura 22). Los genes relacionados con el metabolismo de los carbohidratos son más abundantes en la estación C13, seguida de C10. Mientras que en D18, hay un mayor porcentaje de genes involucrados en el metabolismo de la energía.



**Figura 22.** Representación porcentual de las categorías funcionales identificadas con KEGG Ghost Koala de los genes identificados y anotados de las estaciones a) C10, b) D18 y c) C13.

Mediante este procesamiento de las secuencias fue posible identificar genes relacionados con la reducción de nitrato a nitrito, la oxidación de nitrito a nitrato y la reducción de sulfato a APS (adenosine 5'-phosphosulfate) en la estación C10. En D18, se identificaron genes implicados en la oxidación de tiosulfato a sulfato por el complejo SOX y codificantes para proteínas citocromo c oxidasas así como deshidrogenasas de metanoato (Anexo I).

En C13, fue posible identificar genes implicados en las rutas de la reducción disimilatoria de nitrato a amoniaco, la nitrificación de amoniaco a nitrito, la nitrificación completa de amoniaco->nitrito->nitrato (*commamox*) y de la reducción asimilatoria y disimilatoria de sulfato a H<sub>2</sub>S. También se identificaron genes relacionados con la oxidación del metano a formaldehído y la asimilación de este último, así como genes relacionados con la degradación de ftalato a protocatecuato. (Anexo J).

En C10 fue posible identificar enzimas implicadas en la degradación de nitrotolueno (*Benzoyl acetyl-CoA thiolase*) y aminobenzoato (*terephtalate decarboxylase*) y un gen posiblemente relacionado con la degradación de estireno.

# Capítulo 4. Discusión

La diversidad microbiana de la biosfera es muy grande y hasta hace un par de décadas se conocía un porcentaje muy pequeño. Mediante el análisis de metagenomas de muestras ambientales se ha podido identificar un gran número de los microorganismos que hasta la fecha no han podido ser cultivados. En este trabajo se analizaron los metagenomas de las comunidades procariontes presentes en 18 muestras de sedimentos tomadas en el golfo de México durante la campaña oceanográfica MMF-01 (Metagenómica-Malla Fina 01), llevada a cabo en 2016. Para la caracterización de la microbiota de estas muestras se secuenció la región V4 del gen 16S del ARNr y para el estudio de los metagenomas completos se realizó secuenciación *shotgun* del ADN extraído de tres de las muestras.

Las secuencias de los amplicones y su procesamiento informático permitieron identificar un gran número de OTUs en los sedimentos del GdeM y, aunque se podría pensar que un mayor número de secuencias permite identificar más OTUs, esto no es necesariamente cierto. Como ejemplo de lo anterior se observó que el número de OTU y secuencias obtenidas en las réplicas A4c y D16b y nos indica que una mayor profundidad de secuenciación no garantiza que se identifique un mayor número de especies.

Los resultados obtenidos permitieron identificar a los principales géneros de bacterias y arqueas que habitan los sedimentos de las regiones de Perdido y Coatzacoalcos y que las diferencias en la composición de las comunidades microbianas están influenciadas en mayor medida por la profundidad a la que se colectaron los sedimentos que por la zona de muestreo. Además, la información obtenida de la secuenciación *shotgun* sugiere que los microorganismos procariontes de los sedimentos del GdeM están involucrados en las reacciones redox de compuestos de azufre y nitrógeno.

### 4.1. Microorganismos en el fondo marino de Perdido y Coatzacoalcos

Los resultados obtenidos a través de la secuenciación de amplicones permitieron conocer que los microorganismos más abundantes en los sedimentos de Perdido y Coatzacoalcos pertenecen a los fila Proteobacteria, Thaumarchaeota y Planctomycetes, filos que habían sido previamente reportados como característicos de los sedimentos marinos del Golfo de México (Covarrubias-Rodríguez, 2016).

Se identificaron 350 clases de microorganismos procariotas, tres veces más que las encontradas por Godoy-Lozano *et al* (2018) en sedimentos someros del sur-oeste del GdeM. Además, en este trabajo

se identificaron 20 fila de arqueas, que suman el 20% de la abundancia relativa promedio; mientras que Godoy-Lozano *et al.*, reportan únicamente 4 en abundancias menores al 1%.

El filo de las Proteobacteria fue el más abundante en términos de abundancia relativa promedio de todas las estaciones; se ha reportado que sus miembros están relacionados con la fijación del nitrógeno usando compuestos de azufre como agente oxidante (Battistuzzi y Hedges, 2008). El segundo filo más abundante es el de las Thaumarchaeota, que son arqueas quimiolitótrofas nitrificantes y por ello es común encontrarlas en ambientes marinos (Pester *et al.*, 2011). La información sobre los Planctomycetes, Chloroflexi y Acidobacteria hallados en sedimentos marinos es limitada, pero se piensa que su presencia está relacionada con la presencia de compuestos de nitrógeno (Vigneron, *et al.*, 2017).

Las especies más abundantes en las dos zonas de muestreo son las mismas. Las variaciones se observan en el porcentaje de abundancia relativa de ellas con respecto a la profundidad en que habitan. El porcentaje de abundancia de especies de *Planctomycetacia* en menores profundidades podría estar relacionado con la mayor presencia de polisacáridos, posiblemente procedentes de tejidos de invertebrados marinos y algas de la superficie (Woebken *et al.*, 2007). Además, estas bacterias podrían estar degradando compuestos sulfonados provenientes de algas y compuestos clorados (clorometano y cloroalcanos) derivados de fitoplancton y otras algas que habitan la superficie oceánica (Kertesz, 2000; Vigneron *el at.*, 2017). En este sentido, podría haber una relación inversa entre la presencia de miembros del *Marine Group I* y los *Planctomycetacia* atribuida a que con el incremento de la profundidad disminuye la disponibilidad de polisacáridos y se incrementa la concentración de compuestos de azufre (Vigneron *et al.*, 2017, Sánchez-Soto, 2018).

El mayor porcentaje de Thaumarchaeota en sedimentos profundos podría deberse a que las especies de este filo están mejor adaptadas para sobrevivir en los entornos con bajos flujos de energía y mayores presiones del fondo marino (Valentine, 2010), lo que explicaría por qué su abundancia relativa incrementa con la profundidad. Por otro lado, los miembros de las Deltaproteobacteria están estrechamente relacionados con los ciclos biogeoquímicos (Battistuzzi y Hedges, 2008). En particular, los miembros de los *Desulfobacteraceae* están implicados en las rutas de reducción de sulfatos, fijación del nitrógeno y la oxidación de metano; y al igual que las *Syntrophyobacteraceae* podrían estar relacionados con la degradación de tolueno (Vigneron *et al.* 2018).

En los sedimentos de entre 550 y 788 m de profundidad el orden más abundante es el de los *Brocadiales*, que son bacterias que realizan la oxidación anaeróbica del amonio. Los miembros de la familia

*Anaerolineae* fueron identificados hasta en un 5% de abundancia relativa, y aunque no se tiene mucha información sobre ellos Vigneron *et al.* (2017) sugieren que podrían tener genes (desconocidos) implicados en la degradación de hidrocarburos aromáticos.

En profundidades superiores a los 1,000 hay un incremento en el porcentaje de *Rhodospirillales*, que pueden crecer quimioheterotróficamente en la obscuridad o heterotróficamente bajo condiciones aeróbicas o microaeróbicas (Baldani *et al.*, 2014). La presencia de *Rhodospirillales* en ocasiones está relacionada con la fermentación anaeróbica de lactato a acetato que después es usado por especies de Archaea para producir metano (Madigan y Martinko, 2005).

Estas diferencias en la presencia y abundancia de especies se han explicado comúnmente con respecto a la profundidad de la columna de agua (con sus respectivas variaciones de temperatura y presión) (Orcutt *et al.*, 2010). Los valores de riqueza de especies en profundidades menores a 1,000 m obtenidos en este estudio no permiten validar la sugerencia de Sánchez-Soto *et al* (2018) sobre que las comunidades microbianas presentes en sedimentos de estas profundidades son más ricas en especies. En este sentido, los sedimentos menos profundos podrían estar influenciados por el influjo de materia orgánica y otros compuestos provenientes de la superficie continental y oceánica por lo que la composición de las comunidades microbianas que los habitan obedecen a factores más variados que las que habitan en sedimentos más profundos.

Por otro lado, los valores de riqueza de especies en las estaciones profundas (más de 1,500 m de profundidad) coinciden con lo sugerido por Jochens y DiMarco (2008) sobre que los sedimentos de zonas profundas son menos diversos que los someros. Estos resultados refuerzan lo mencionado por Covarrubias-Rodríguez (2016) sobre que la diversidad microbiana en sedimentos profundos del GdeM es homogénea en términos de composición microbiana.

En sedimentos colectados en profundidades superiores a los 2,000 m abundan los organismos pertenecientes a las familias de los *Rhodospirillaceae* y del *Marine Group I*. Además, se hizo notorio que los sedimentos de estas profundidades son habitados por miembros del filo de las *Euryarchaeota*. Se ha reportado que los *Thermoplasmatales* habitan en sedimentos profundos o ventas hidrotermales y que tienen potencial para la reducción de nitratos, la metanogénesis y la oxidación de metano (Madigan *et al.,* 2010), siendo la estación C13 en donde se identificó el mayor porcentaje ellos. Covarrubias-Rodríguez (2016) sugirió que estos microorganismos podrían ser oxidadores de metano y que se encuentran en una

relación mutualista (*syntrophy*) con otras arqueas reductoras de sulfatos, en este caso, las pertenecientes al *Marine Group I*.

Aunque parece ser que la profundidad de los sedimentos es la variable que mejor explica las diferencias en los perfiles taxonómicos de las estaciones (p<0.05), el perfil taxonómico en la estación D18 no parece depender de la profundidad de la toma de muestra. Las estaciones A3, C12 y D18 fueron muestreadas a 1,500 m de profundidad, siendo la primera de la región de Perdido y las dos últimas de Coatzacoalcos. Las estaciones A3 y C12 tienen un perfil de diversidad similar entre ellas y con las estaciones muestreadas a 1,000 y 1,200 m, sin embargo, el perfil taxonómico de la estación D18 es diferente al resto de las estaciones. Por lo anterior, no se pueden atribuir estas diferencias ni a la profundidad ni a la zona de muestreo sino a otros factores más específicos del sitio de muestreo.

En la estación D18 se identificaron principalmente miembros de las familias *Anaerolineaceae*, *Desulfobacteraceae* y *Syntrophobacteraceae* que previamente fueron reportadas como microorganismos característicos de sitios de emanación de hidrocarburos (metano y petróleo) en el norte del golfo de México (Vigneron *et al.*, 2017). Adicionalmente, Vigneron *et al*, mencionan que los miembros de *Anaerolineae* podrían tener genes relacionados con la degradación de hidrocarburos (como el naftaleno) y con la reducción de sulfatos. En este sentido, cuando se planeó la campaña MMF-01, se proyectó que las estaciones B7, C13 y D18 estarían cerca de sitios de emanaciones naturales de hidrocarburos pero dado que no se puede controlar el movimiento del nucleador una vez que es lanzado no se puede garantizar que los sedimentos sean tomados exactamente del sitio planeado. Sin embargo, la identificación de las especies arriba mencionadas llevó a pensar que los sedimentos fueron, en efecto, tomados de una chapopotera o un sitio muy cercano a una.

La presencia de especies de *Aminicenantes* en D18 podría ser otro indicador de la cercanía de una chapopotera, puesto que estas bacterias abundan en sitios contaminados por hidrocarburos o respiraderos hidrotermales, además, habitan en sitios con bajos niveles de oxígeno y participan en la degradación del carbono a metano en conjunto con especies de *Planctomycetes* (Farag *et al.,* 2014; Robbins *et al.,* 2016). En los sedimentos de esta estación se identificó un porcentaje relativamente alto de bacterias reductoras de sulfatos que no es común de identificar en sedimentos superficiales (Teske, 2012). Además, los miembros del *Marine Group I* se encuentran en un porcentaje mínimo contrario a la tendencia que se observa en la figura 16 y se identificaron arqueas (*Lokiarchaeota* y *Woesearchaeota*) que son características de sedimentos más profundos (Orcutt *et al.,* 2013).

### 4.2. Potencial metabólico de la microbiota de los sedimentos marinos

La identificación de la función de las comunidades de este estudio se realizó mediante dos estrategias: infiriéndola con la herramienta PICRUSt (Langille *et al.*, 2013) y a través de la secuenciación *shotgun* del ADN de las estaciones C10, C13 y D18.

La inferencia del potencial de las comunidades identificadas por la metagenómica dirigida no permitió observar diferencias en la función de los diferentes perfiles taxonómicos de las estaciones de muestreo. Este era un resultado esperado ya que los valores de NSTI obtenidos estuvieron entre 0.16 y 0.29 y nos indicó que la disponibilidad de los genomas completos en la base de datos de *GreenGenes* era limitada para la diversidad identificada. Sin embargo, los resultados de esta inferencia arrojaron que posiblemente la función principal de los microorganismos en el fondo marino estaba más relacionada con su metabolismo primario y el procesamiento de la información genética.

Dada la posible presencia de hidrocarburos en los sedimentos y con el objetivo de identificar genes relacionados con su degradación se realizó la secuenciación *shotgun* del ADN de las estaciones C10, C13 y D18. Estas estaciones se eligieron considerando su perfil taxonómico y las características del sitio en que fueron tomadas. Los sedimentos de la estación C10 fueron tomados a 550 m de profundidad en un sitio cercano a tuberías de extracción de hidrocarburos, en base a informaciones previas de la zona; la estación D18 estaba cerca de una chapopotera y los sedimentos de la C13 se colectaron a una profundidad mayor de 2,500 m.

Como cabía esperar, se identificaron principalmente genes implicados en el procesamiento de la información genética y en el metabolismo de la energía (Eipstein, 2013; Head *et al.*, 1997). Aunque Sánchez-Soto *et al.* (2018) sugirieron que la microbiota de los sedimentos del Golfo de México tenía el potencial para la degradación de compuestos aromáticos como nitrotolueno, etilbenceno y clorociclohexano basados en la inferencia funcional con PICRUSt, los datos obtenidos de la anotación con *Prokka* en este trabajo no permitieron confirmar esta predicción. Sin embargo, esto puede deberse a que no se ensamblaron secuencias suficientemente largas y por lo tanto los *contigs* ensamblados no permitieron identificar todos los genes de la microbiota procarionte de los sitios muestreados.

El potencial metabólico de las comunidades microbianas de los sedimentos marinos es el resultado de las condiciones ambientales a las que han estado sometidas y aún podrían identificarse microorganismos y/o genes desconocidos e implicados en procesos biogeoquímicos y, en el caso particular de las zonas de estudio, implicados en la degradación de hidrocarburos. Estos microrganismos o genes podrían ser de gran ayuda en el tratamiento *in situ* de sitios afectados por hidrocarburos del petróleo.

### 4.3. Limitaciones y perspectivas

A pesar de la gran cantidad de información obtenida en este trabajo, estudiar muestras ambientales y realizar inferencias ecológicas basadas en resultados aislados representa un gran reto por la variabilidad intrínseca de las muestras. En este sentido, es necesario que los estudios microbiológicos sean respaldados y/o complementados por información fisicoquímica, pues de esta manera podrían realizarse inferencias más acertadas sobre las interacciones entre la biosfera y la geosfera que contribuyan al entendimiento de los procesos en que los microrganismos están involucrados.

En este trabajo, la profundidad de la secuenciación *shotgun* sólo permitió ensamblar *contigs* cortos y por ello no fue posible identificar rutas metabólicas completas sino sólo puntos aislados. Por esto sería ideal la obtención de secuencias *shotgun* que permitan el ensamble de *contigs* más largos y la posibilidad de obtener los genomas completos de estos microorganismos para así conocer el potencial de los microorganismos del fondo marino.

#### 4.4. Sugerencias técnicas

Durante el proceso de amplificación de las muestras se hizo evidente que en los sedimentos marinos existen sustancias capaces de inhibir la Reacción en Cadena de la Polimerasa por lo que se sugiere usar diluciones del ADN extraído y concentraciones menores a 15 ng en la reacción. En este sentido, podría ser de ayuda realizar la identificación de estas sustancias para desarrollar estrategias que permitan la limpieza de estas muestras y posiblemente mejoras en los métodos de extracción y amplificación.

La asignación taxonómica de las OTU formadas fue realizada *close reference, open reference* y *de novo*, reportándose en este trabajo los resultados *de novo* porque implica menos sesgo asociado con la asignación de los OTU, además de que permite observar el porcentaje de OTU que no son asignadas. La secuenciación *shotgun* debe ser realizada preferentemente pareada y en plataformas que devuelvan como mínimo 4 Gb de datos por muestra, con fragmentos que permitan ensamblar contigs de mayor longitud (de mínimo 5,000 bp). De esta manera es más probable la identificación correcta de un mayor número de genes e incluso podría ser posible identificar genomas completos de los microorganismos presentes en la muestra.

# **Capítulo 5. Conclusiones**

- Las comunidades de procariotas de los sedimentos de Perdido y Coatzacoalcos está compuesta principalmente por los fila Proteobacteria, Thaumarchaeota, Planctomycetes, Acidobacteria, Chloroflexi y Nitrospirae.
- La región de muestreo no presentó una influencia estadísticamente significativa en la composición las comunidades microbianas estudiadas. Por otro lado, la profundidad bajo el nivel del mar a la que fueron tomadas fue el factor que presentó una influencia estadísticamente significativa (p<0.01) tanto en la riqueza como en la abundancia de las comunidades de estudio.</li>
- La abundancia relativa de los *Planctomycetacia* en los sedimentos disminuye con el incremento de la profundidad, al contrario de lo que ocurre con el *Marine Group I*, cuya abundancia aumenta con la profundidad.
- Los sedimentos del golfo de México de profundidades superiores a los 2,000 m son similares entre sí, en términos de composición microbiana y están dominados por miembros del Marine Group I y Rhodospirillales.
- La microbiota procarionte en el fondo marino de las regiones de Perdido y Coatzacoalcos es altamente diversa. Los resultados de la secuenciación *shotgun* sugieren que las especies de Thaumarchaeota, Planctomycetes, Chloroflexi, Delta-, Gamma-, y Alphaproteobacteria están implicadas en los ciclos biogeoquímicos del planeta.
- Se identificaron genes implicados en la oxidación-reducción de compuestos de azufre y nitrógeno y en la fijación del carbono, como resultado del análisis funcional.

## Literatura citada

- Amann, R. I. (1995). In situ identification of micro-organisms by whole cell hybridization with rRNAtargeted nucleic acid probes. *Molecular Microbial Ecology Manual*, 331-345. doi:10.1007/978-94-011-0351-0\_23
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
- Asl, S. D., Amos, J., Woods, P., Garcia-Pineda, O., y Macdonald, I. R. (2016). Chronic, Anthropogenic Hydrocarbon Discharges in the Gulf of Mexico. *Deep Sea Research Part II: Topical Studies in Oceanography, 129,* 187-195. doi:10.1016/j.dsr2.2014.12.006
- Baldani, J. I., Videira, S. S., Teixeira, K. R., Reis, V. M., Oliveira, A. L., Schwab, S., y Hartmann, A. (2014). The Family Rhodospirillaceae. *The Prokaryotes*, 533-618. doi:10.1007/978-3-642-30197-1\_300
- Battistuzzi, F. U., y Hedges, S. B. (2008). A Major Clade of Prokaryotes with Ancient Adaptations to Life on Land. *Molecular Biology and Evolution*, *26*(2), 335-343. doi:10.1093/molbev/msn247
- Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., y House, C. H. (2006). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. Proceedings of the National Academy of Sciences, 105(30), 10583-10588. doi:10.1073/pnas.0709942105
- Bik, H. M., Halanych, K. M., Sharma, J., y Thomas, W. K. (2012). Dramatic Shifts in Benthic Microbial Eukaryote Communities following the Deepwater Horizon Oil Spill. *PLoS ONE*, 7(6). doi:10.1371/journal.pone.0038550
- Bolger, A. M., Lohse, M., & Usadel, B. (2014, 04). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi:10.1093/bioinformatics/btu170
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Knight, R. (2010, 04). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335-336. doi:10.1038/nmeth.f.303
- Covarrubias-Rodríguez, M. B. (2016). Caracterización de las comunidades microbianas en sedimentos del Golfo de México, mediante análisis metagenómico. Tesis de Maestría. CICESE. 2016
- Cummings, R., García, C., Hawthorn, A., y Holicek, R. (2015). Más allá de las profundidades: Los desafíos de la región de aguas ultra-profundas. Oilfield Review: 04.
- Dinghua, L., Chi-Ma,n L., Ruibang, L., Kunihiko, S., y Tak-Wah, L.; MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, Bioinformatics, Volume 31, Issue 10, 15 May 2015, Pages 1674–1676, <a href="https://doi.org/10.1093/bioinformatics/btv033">https://doi.org/10.1093/bioinformatics/btv033</a>
- Duineveld, B. M., Kowalchuk, G. A., Keijzer, A., Elsas, J. D., y Veen, J. A. (2001). Analysis of Bacterial Communities in the Rhizosphere of Chrysanthemum via Denaturing Gradient Gel Electrophoresis of PCR-Amplified 16S rRNA as Well as DNA Fragments Coding for 16S rRNA. Applied and Environmental Microbiology, 67(1), 172-178. doi:10.1128/aem.67.1.172-178.2001

- Edgar, Robert C., Haas, Brian J., Clemente, Jose C., Quince, C., y Knight, R. 2011. UCHIME Improves Sensitivity and Speed of Chimera Detection 27 (16): 2194–220010. doi:10.1093/bioinformatics/btr381
- Epstein, S. (2013, 10). The phenomenon of microbial uncultivability. *Current Opinion in Microbiology*, *16*(5), 636-642. doi:10.1016/j.mib.2013.08.003
- Falkowski, P. G., Fenchel, T., y Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*(5879), 1034-1039. doi:10.1126/science.1153213
- Farag, I. F., Davis, J. P., Youssef, N. H., y Elshahed, M. S. (2014). Global Patterns of Abundance, Diversity and Community Structure of the Aminicenantes (Candidate Phylum OP8). *PLoS ONE*, 9(3). doi:10.1371/journal.pone.0092139
- Gilbert, J. A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C. T., Brown, C. T., y Stevens, R. (2010). Meeting Report: The Terabase Metagenomics Workshop and the Vision of an Earth Microbiome Project. *Standards in Genomic Sciences*, *3*(3), 243-248. doi:10.4056/sigs.1433550
- Godzik, A. (2011). Metagenomics and the protein universe. Current Opinion in Structural Biology, 21(3), 398-403. doi:10.1016/j.sbi.2011.03.010
- Graham, A., y Julson, A. (1989). Introduction to the Ocean Drilling Program. Ocean Drilling Program Technical Notes. doi:10.2973/odp.tn.11.1989
- Gurevich, A., Saveliev, V., Vyahhi, N., y Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. doi:10.1093/bioinformatics/btt086
- Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., y Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology, 5(10), R245-R249.
- Harvey, R. G., 1987. A personal overview of oil in the marine environment. Caribb Jour. Sci., 23: 5-9.
- Hazen, T. C., Dubinsky, E. A., Desantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., y Mason, O. U. (2010).
   Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. Science, 330(6001), 204-208.
   doi:10.1126/science.1195979
- Head, I., Saunders, J., y Pickup, R. (1997). Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microbial Ecology*, 35(1), 1-21. doi:10.1007/s002489900056
- Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., y Ogata, H. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal*, 7(9), 1678-1695. doi:10.1038/ismej.2013.59
- Hinrichs, K., y Boetius, A. (2002). The Anaerobic Oxidation of Methane: New Insights in Microbial Ecology and Biogeochemistry. *Ocean Margin Systems*, 457-477. doi:10.1007/978-3-662-05127-6\_28

Illumina, Inc. (2015). Brochure sequencing systems portfolio. MiSeq ®System Guide.

- Jochens, A. E., y Dimarco, S. F. (2008). Physical oceanographic conditions in the deepwater Gulf of Mexico in summer 2000–2002. *Deep Sea Research Part II: Topical Studies in Oceanography, 55*(24-26), 2541-2554. doi:10.1016/j.dsr2.2008.07.003
- Kessler, J. D., Valentine, D. L., Redmond, M. C., y Du, M. (2011). Response to Comment on "A Persistent Oxygen Anomaly Reveals the Fate of Spilled Methane in the Deep Gulf of Mexico". Science, 332(6033), 1033-1033. doi:10.1126/science.1203428
- Kertesz, M. A. (2000). Riding the sulfur cycle metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiology Reviews*, 24(2), 135-175. doi:10.1111/j.1574-6976.2000.tb00537.x
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., y Glöckner, F. O., (2013). Evaluation of General 16S Ribosomal RNA Gene PCR Primers for Classical and Next-Generation Sequencing-Based Diversity Studies. Nucleic Acids Research 41 (1): 1–11. doi:10.1093/nar/gks808
- Kodzius, R., y Gojobori, T. (2015). Marine metagenomics as a source for bioprospecting. *Marine Genomics,* 24, 21-30. doi:10.1016/j.margen.2015.07.001
- Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., y Schloss, P. D. (2013). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, 79(17), 5112-5120. doi:10.1128/aem.01043-13
- Kanehisa, M., Sato, Y., y Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4), 726-731. doi:10.1016/j.jmb.2015.11.006
- Kennedy, J., O'Leary, N., Kiran, G., Morrissey, J., O'Gara, F., Selvin, J., y Dobson, A. (2011). Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *Journal of Applied Microbiology*, 111(4), 787-799. doi:10.1111/j.1365-2672.2011.05106.x
- Kim, M., Morrison, M., y Yu, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods*, 84(1), 81-87. doi:10.1016/j.mimet.2010.10.020
- Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D., y Knight, R. (2012).
   Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*, 13(1), 47-58. doi:10.1038/nrg3129
- Lamendella, R., Strutt, S., Borglin, S., Chakraborty, R., Tas, N., Mason, O. U., y Jansson, J. K. (2014). Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. *Frontiers in Microbiology, 5.* doi:10.3389/fmicb.2014.00130
- Langille, M. G., Zaneveld, J., Caporaso, J. G., Mcdonald, D., Knights, D., Reyes, J. A., y Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology, 31(9), 814-821. doi:10.1038/nbt.2676
- Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., y Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1983-1992. doi:10.1109/tvcg.2014.2346248

- Lozupone, C., y Knight, R. (2005). UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, 71(12), 8228-8235. doi:10.1128/aem.71.12.8228-8235.2005
- Lozupone, C. A., y Knight, R. (2007). Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences, 104(27), 11436-11440. doi:10.1073/pnas.0611525104
- MacDonald, I. R., Garcia-Pineda, O., Beet, A., Asl, S. D., Feng, L., Graettinger, G., y Swayze, G. (2015). Natural and unnatural oil slicks in the Gulf of Mexico. *Journal of Geophysical Research: Oceans*, 120(12), 8364-8380. doi:10.1002/2015jc011062
- Madigan M.T., Martinko J.M. y Parker J. (2010). Biología de los Microorganismos. 10a edición, Pearson Prentice Hall: Madrid, España.
- Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Current Opinion in Biotechnology*, 22(3), 456-464. doi:10.1016/j.copbio.2011.01.008
- Marshall, I. P., Karst, S. M., Nielsen, P. H., y Jørgensen, B. B. (2018). Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition. *Marine Genomics*, *37*, 58-68. doi:10.1016/j.margen.2017.08.004
- Mason, O. U., Hazen, T. C., Borglin, S., Chain, P. S., Dubinsky, E. A., Fortney, J. L., y Jansson, J. K. (2012). Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. The ISME Journal, 6(9), 1715-1727. doi:10.1038/ismej.2012.59
- Mason, O. U., Scott, N. M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., y Jansson, J. K. (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *The ISME Journal*, 8(7), 1464-1475. doi:10.1038/ismej.2013.254
- Mccliment, E. A., Voglesonger, K. M., O'day, P. A., Dunn, E. E., Holloway, J. R., y Cary, S. C. (2006). Colonization of nascent, deep-sea hydrothermal vents by a novel Archaeal and Nanoarchaeal assemblage. Environmental Microbiology, 8(1), 114-125. doi:10.1111/j.1462-2920.2005.00874.x
- Mcdonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., y Hugenholtz, P. (2011). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal, 6*(3), 610-618. doi:10.1038/ismej.2011.139
- Morgan, X. C., y Huttenhower, C. (2012). Human microbiome analysis. PLoS computational biology, 8(12), e1002808
- Musat, F., y Widdel, F. (2007). Anaerobic degradation of benzene by a marine sulfate-reducing enrichment culture, and cell hybridization of the dominant phylotype. *Environmental Microbiology, 0*(0). doi:10.1111/j.1462-2920.2007.01425.x
- Muyzer, G., de Waal, E. C., y Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology, 59(3), 695–700
- National Research Council, Transportation Research Board, Ocean Studies Board, Marine Board, y Committee on Oil. (2002, May 23). Oil in the Sea III: Inputs, Fates, and Effects. Retrieved from https://doi.org/10.17226/10388

- Olsen, G. (1986). Microbial Ecology and Evolution: A Ribosomal RNA Approach. Annual Review of Microbiology, 40(1), 337-365. doi:10.1146/annurev.micro.40.1.337
- Orcutt, B. N., Joye, S. B., Kleindienst, S., Knittel, K., Ramette, A., Reitz, A., y Boetius, A. (2010). Impact of natural oil and higher hydrocarbons on microbial diversity, distribution, and activity in Gulf of Mexico cold-seep sediments. *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(21-23), 2008-2021. doi:10.1016/j.dsr2.2010.05.014
- Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., y Iliopoulos, L. (2015). Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. Bioinformatics and Biology Insights, 9. doi:10.4137/bbi.s12462
- Parthasarathy, H., Hill, E., y MacCallum, C. (n.d.). Global Ocean Sampling Collection. Retrieved from https://doi.org/10.1371/journal.pbio.0050083
- Pester M., Schleper C., y Wagner M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Current opinion in microbiology*, *14*(3), 300-306.
- Petersen, J. M., Zielinski, F. U., Pape, T., Seifert, R., Moraru, C., Amann, R., y Dubilier, N. (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature*, 476(7359), 176-180. doi:10.1038/nature10325
- Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Jörg P., J. y Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013; 41(D1):D590–596. [PubMed: 23193283]
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., y Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59-65. doi:10.1038/nature08821
- Rappé, M. S., y Giovannoni, S. J. (2003). The Uncultured Microbial Majority. *Annual Review of Microbiology,* 57(1), 369-394. doi:10.1146/annurev.micro.57.030502.090759
- Reuter, J., Spacek, D. V., y Snyder, M. (2015). High-Throughput Sequencing Technologies. *Molecular Cell,* 58(4), 586-597. doi:10.1016/j.molcel.2015.05.004
- Riesenfeld, C. S., Schloss, P. D., y Handelsman, J. (2004). Metagenomics: Genomic Analysis of Microbial Communities. Annual Review of Genetics, 38(1), 525-552. doi:10.1146/annurev.genet.38.072902.091216
- Robbins, S. J., Evans, P. N., Parks, D. H., Golding, S. D., y Tyson, G. W. (2016). Genome-Centric Analysis of Microbial Populations Enriched by Hydraulic Fracture Fluid Additives in a Coal Bed Methane Production Well. *Frontiers in Microbiology*, 7. doi:10.3389/fmicb.2016.00731
- Ronaghi, M. (2001). Pyrosequencing Sheds Light on DNA Sequencing. *Genome Research*, 11(1), 3-11. doi:10.1101/gr.11.1.3
- Sánchez-Soto Jiménez, M. F., Cerqueda-García, D., Montero-Muñoz, J. L., Aguirre-Macedo, M. L., y García-Maldonado, J. Q. (2018). Assessment of the bacterial community structure in shallow and deep sediments of the Perdido Fold Belt region in the Gulf of Mexico. *PeerJ*, 6. doi:10.7717/peerj.5583

- Sanger, F., Nicklen, S., y Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74(12), 5463-5467. doi:10.1073/pnas.74.12.5463
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069. doi:10.1093/bioinformatics/btu153
- Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. Frontiers in plant science, 5, 209
- Simpson, J., y White, B. (1997). Microbial community structure of the rumen as assessed by denaturing gradient gel electrophoresis of polymerase chain-reaction amplified 16S rDNA genes. *Reproduction Nutrition Development, 37*(Suppl. 1), 29-30. doi:10.1051/rnd:19970709
- Soo, R. M., Wood, S. A., Grzymski, J. J., Mcdonald, I. R., y Cary, S. C. (2009). Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. Environmental Microbiology, 11(3), 715-728. doi:10.1111/j.1462-2920.2009.01859.x
- Teske, A. P. (2012). Tracking microbial habitats in subseafloor sediments. *Proceedings of the National Academy of Sciences, 109*(42), 16756-16757. doi:10.1073/pnas.1215867109
- Trevors, J. (2011). Viable but non-culturable (VBNC) bacteria: Gene expression in planktonic and biofilm cells. *Journal of Microbiological Methods*, *86*(2), 266-273. doi:10.1016/j.mimet.2011.04.018
- Valentine, D. L., Kessler, J. D., Redmond, M. C., Mendes, S. D., Heintz, M. B., Farwell, C., y Villanueva, C. J. (2010). Propane Respiration Jump-Starts Microbial Response to a Deep Oil Spill. Science, 330(6001), 208-211. doi:10.1126/science.1196830
- Venter, J. C. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, 304(5667), 66-74. doi:10.1126/science.1093857
- Vigneron, A., Alsop, E. B., Cruaud, P., Philibert, G., King, B., Baksmaty, L., y Tsesmetzis, N. (2017). Comparative metagenomics of hydrocarbon and methane seeps of the Gulf of Mexico. *Scientific Reports*, 7(1). doi:10.1038/s41598-017-16375-5
- Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K., y Rosselló-Móra, R. (2008). The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. Systematic and Applied Microbiology, 31(4), 241-250. doi:10.1016/j.syapm.2008.07.001
- Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., y Glöckner, F. O. (2013). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, 42(D1). doi:10.1093/nar/gkt1209
- Youssef, N., Sheik, C. S., Krumholz, L. R., Najar, F. Z., Roe, B. A., y Elshahed, M. S. (2009). Comparison of Species Richness Estimates Obtained Using Nearly Complete Fragments and Simulated Pyrosequencing-Generated Fragments in 16S rRNA Gene-Based Environmental Surveys. *Applied* and Environmental Microbiology, 75(16), 5227-5236. doi:10.1128/aem.00592-09
- Yozwiak, N. L., Skewes-Cox, P., Stenglein, M. D., Balmaseda, A., Harris, E., y Derisi, J. L. (2012). Virus Identification in Unknown Tropical Febrile Illness Cases Using Deep Sequencing. PLoS Neglected Tropical Diseases, 6(2). doi:10.1371/journal.pntd.000148

- Yu, Z., y Morrison, M. (2004). Comparisons of Different Hypervariable Regions of rrs Genes for Use in Fingerprinting of Microbial Communities by PCR-Denaturing Gradient Gel Electrophoresis. *Applied* and Environmental Microbiology, 70(8), 4800-4806. doi:10.1128/aem.70.8.4800-4806.2004
- Wasmund, K., Mußmann, M., y Loy, A. (2017). The life sulfuric: Microbial ecology of sulfur cycling in marine sediments. *Environmental Microbiology Reports*, *9*(4), 323-344. doi:10.1111/1758-2229.12538
- Woebken, D., Teeling, H., Wecker, P., Dumitriu, A., Kostadinov, I., Delong, E. F., y Glöckner, F. O. (2007).
   Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with planctomycete genomes. *The ISME Journal*, 1(5), 419-435. doi:10.1038/ismej.2007.63
- Wooley, J.C., Godzik, A., y Friedberg, I. (2010) A Primer on Metagenomics. PLoS Comput Biol 6(2): e1000667. doi:10.1371/journal.pcbi.1000667
- Woese, C. R. (1987). Bacterial evolution. Microbiological reviews, 51(2), 221.
- Woese, C. R., Kandler, O., y Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences, 87(12), 4576-4579. doi:10.1073/pnas.87.12.4576
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., Verberkmoes, N. C., y Banfield, J. F. (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. Science, 337(6102), 1661-1665. doi:10.1126/science.1224041

### Anexo A. Protocolo para la extracción de ADN

Para la extracción del ADN de las muestras se usó el *PowerSoil® DNA Isolation kit* de Qiagen®. El protocolo descrito a continuación son las instrucciones proporcionadas por el fabricante con modificaciones de la Dra. Jennyfers Chong Robles.

Antes de comenzar con la extracción es necesario realizar los siguientes pasos.

- I. Colocar el Fenol:Cloroformo:Alcohol Isoamílico (PCI) previamente preparado (en proporción 25:24:1) y mantenido a 4°C en la campana de extracción.
- II. La solución C1 y C6 se colocan a temperaturas de 55°C y 37°C, respectivamente.
- III. Por muestra a procesar se usan 1 tubo con perlas, 5 tubos libres de DN/RNasas y una columna de filtrado.

La extracción se realiza de la siguiente manera.

(Todos los pasos de centrifugación se realizaron a temperatura ambiente y 14,000 x g).

- 1. Añadir 60 µL de solución C1 al tubo con perlas. Mezclar por inversión.
- 2. Añadir no más de 250 mg de sedimento (peso húmedo). Mezclar por inversión.
- 3. Usando puntas con filtro, añadir 273 µL de PCI. Mezclar por inversión e incubar por 5 minutos.
- 4. Mezclar con vortex a máxima velocidad durante 5 minutos.
- 5. Centrifugar durante un minuto.
- En campana y usando puntas con filtro transferir sin perturbar la interface formada ni arrastrar partículas de sedimento el sobrenadante a un tubo limpio con 250 μL de C2. Mezclar e incubar a 4°C durante 5 minutos.
- 7. Centrifugar durante 2 minutos.
- Transferir 600 μL del sobrenadante (sin tocar el pellet) a un tubo nuevo con 200 μL de C3. Mezclar e incubar a 4°C durante 5 minutos.
- 9. Centrifugar durante 3 minutos.
- En un tubo limpio añadir 650 μL de C4 y 450 μL de etanol al 100%, mezclar perfectamente.
   Transferir como máximo 850 μL del sobrenadante del punto 9, mezclar.
- Usando puntas con filtro transferir 650 μL a la columna. Centrifugar durante 15 segundos.
   Descartar el fluido. Repetir dos veces más.

- 12. Centrifugar durante 2 minutos. Transferir cuidadosamente la columna a un tubo nuevo.
- 13. Añadir 650 µL de alcohol al 100%. Centrifugar durante 30 segundos. Desechar fluido.
- 14. Centrifugar durante 3 minutos. Desechar fluido.
- 15. Añadir 650 μL de solución C5. Centrifugar durante 30 segundos. Desechar fluido.
- 16. Centrifugar durante 4 minutos. Transferir columna a tubo nuevo.
- 17. Añadir 50 μL de solución C6 directamente a la matriz de la columna. Incubar a temperatura ambiente por 10 minutos. Centrifugar por 1 minuto.
- 18. Repetir punto 17. Centrifugar por 2 minutos. Desechar columna. Si se desea, se puede cuantificar el ADN o almacenar a -20°C.

## Anexo B. Información sobre los primers de amplificación

La construcción de las bibliotecas de secuenciación usando la estrategia de Kozich *et al.* (2013), implica que la amplificación de la región de interés y la ligación de los adaptadores necesarios para su secuenciación se realice en un solo paso. En este enfoque se utilizan *primers* compuestos que contienen:

 Los oligos para amplificar la región V4 del gen 16s ARNr que fueron los propuestos por Klindworth et al. (2013) y son los siguientes:
 Oligo Forward = 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG

Oligo Reverse = 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

- El adaptador de *Illumina™* de 33 pb
   AATGATACGGCGACCACCGAGATCTACAC
- El oligo de secuenciación o PAD de 10 pb ACTATCTGTA
- Un linker de 2pb que une al oligo forward con el de secuenciación (GT)
- Un índice de 8 pb, que es distinto para cada muestra y por lo tanto permite su identificación (Tabla A).

ID índice Reverse	Índice <i>Reverse</i> i7	ID índice Forward
16sK701R	CGAGAGTT	16sK501F
16sK702R	GTTACAGC	16sK502F
16sK703R	TAACGTCC	16sK503F
16sK704R	CTACGACC	16sK504F
16sK705R	GAGACTTA	16sK505F
16sK706R	ACTGTGTA	16sK506F
16sK707R	TGCGTCAA	16sK507F
16sK708R	ACGTGCGC	16sK508F
16sK709R	ATAGTACC	16sK509F
16sK710R	GCGTATAC	
16sK711R	TGCTCGTA	
16sK712R	AACGCTGA	
16sK713R	CGCGATAT	

#### Tabla A. Índices forward y reverse disponibles para la secuenciación
## Anexo C. Información adicional sobre la Reacción en Cadena de la Polimerasa

Precalenta	Despaturalización		Extonsión	Final		
miento		Desnaturalización	Alineamiento	Extensión	Extension	i mai
95°C, 30 s	95°C, 2 min	95°C, 20 s	55.5°C, 15 s	72°C, 5 min	72°C, 7 min	4°C, ∞





**Figura 23.** Flujo de trabajo para la construcción de los amplicones etiquetados (Tomado de Covarrubias-Rodríguez, 2016).

#### Anexo D. Otros datos sobre las muestras

En las siguientes tablas se muestran las variables que se modificaron para lograr la amplificación de la región V4 del gen 16S del ARNr, estas fueron el uso de BSA, en número de ciclos de PCR y la concentración de ADN usada. Así mismo, se indican las lecturas que se procesaron con QIIME, el número de OTU identificadas en estas secuencias y los índices de diversidad alfa.

**Tabla 6.** Variables consideradas para la obtención de los amplicones en las réplicas de Perdido, el número de secuencias de cada una de las réplicas y el número de OTU identificadas.

Muestra	Uso de BSA	ADN en PCR	Ciclos de PCR	Secuenciación	Secuencias codificantes para OTU	Número de OTU
N1a	no	12	25	21	44229	4898
N1b	no	12	25	21	46133	4123
N1c	no	12	25	21	59983	5377
A1a	no	15	25	21	48090	4544
A1b	no	12	25	21	64521	5229
A1c	no	15	30	23	59383	5895
A3a	no	15	25	21	55949	3560
A3b	no	15	25	21	50101	3929
A3c	no	18	25	23	53526	4151
A4a	no	15	25	21	60694	4072
A4b	no	15	30	23	60335	4390
A4c	no	18	25	23	63899	3920
B5a	no	15	25	21	28544	3717
B5b	no	18	25	23	19072	3193
B5c	no	18	25	23	23800	3608
B6a	no	18	25	23	51431	3986
B6b	no	15	25	21	39339	4597
B6c	no	18	25	23	39327	3603
B7a	no	15	25	21	44092	4821
B7b	no	15	25	21	38715	4026
B7c	no	18	25	23	58914	4567
B8a	no	15	30	23	46508	3233
B8b	no	15	25	21	61354	3678
B8c	no	15	25	21	58498	3912
B9a	no	15	30	23	55897	4502
B9b	no	15	30	23	54342	3623
B9c	no	15	25	21	50393	3416

 Tabla 7. Variables consideradas para la obtención de los amplicones en las réplicas de Coatzacoalcos, el número de secuencias de cada una de las réplicas y el número de OTU identificadas.

Muestra	Uso de BSA	ADN en PCR	Ciclos de PCR	Secuenciación	Secuencias codificantes para OTU	Número de OTU
C10a	no	6	27	21	71172	5000
C10b	no	6	27	21	62428	5645
C10c	no	6	27	23	39112	5963
C11a	no	15	25	21	85689	7686
C11b	no	18	25	23	96153	6361
C11c	no	18	25	23	118774	7293
C12a	no	15	25	21	45795	3549
C12b	no	15	30	23	59396	4654
C12c	no	15	30	23	53190	4907
C13a	no	15	25	21	64585	3872
C13b	no	18	25	23	89070	4265
C13c	no	18	25	23	77215	3918
C14a	no	15	25	21	53047	3382
C14b	no	18	25	23	62194	3919
C14c	no	18	25	23	49256	3117
D15a	no	12	27	21	28215	4231
D15b	no	15	25	21	27649	4274
D15c	no	6	27	23	28231	4050
D16a	no	15	25	21	59127	7013
D16b	no	18	25	23	64075	7491
D16c	no	18	25	23	63152	6406
D17a	no	15	30	23	45572	5286
D17b	no	15	25	21	54238	5547
D17c	no	18	25	23	43939	3766
D18a	no	18	25	23	50255	5491
D18b	Sí	15	25	24	44126	5509
D18c	Sí	15	25	24	44774	5976

Estación	Observadas	Esperadas (Chao1)	PD_whole_tree	shannon	Estación	Observadas	Esperadas (Chao1)	PD_whole_tree	shannon
N1a	4898	5930.531	200.3777	9.941829	C10a	5000	5700.785	195.8046	10.22779
N1b	4123	5001.408	170.7387	9.166986	C10b	5645	6625.059	223.5509	10.32521
N1c	5377	6385.135	211.0807	9.790719	C10c	5963	7426.713	236.161	10.68898
A1a	4544	5271.742	189.3196	9.989515	C11a	7686	9395.294	275.4183	10.32631
A1b	5229	6334.66	187.9072	9.62611	C11b	6361	7905.507	215.4476	9.698075
A1c	5895	7534.448	225.8244	10.10857	C11c	7293	8770.179	252.3444	9.783601
A3a	3560	4436.421	140.4066	8.867675	C12a	3549	4255.18	145.3268	8.888054
A3b	3929	4820.945	156.167	9.09008	C12b	4654	5725.003	173.5171	9.39578
A3c	4151	4964.161	159.9563	9.447366	C12c	4907	6289.031	181.094	9.574868
A4a	4072	5081.133	159.2005	8.909919	C13a	3872	4977.204	141.8049	8.632475
A4b	4390	5322.727	175.9672	9.314086	C13b	4265	5054.905	157.809	8.874462
A4c	3920	4725.514	153.0809	8.926492	C13c	3918	4761.073	146.6797	8.728074
B5a	3717	4476.104	161.876	9.928575	C14a	3382	4189.779	133.5349	8.945942
B5b	3193	3898.32	148.5971	10.03937	C14b	3919	4772.663	152.0237	8.796073
B5c	3608	4633.942	153.833	9.406996	C14c	3117	3862.141	123.8473	8.57639
B6a	3986	4924.191	155.2271	9.240451	D15a	4231	5042.357	179.6534	10.32366
B6b	4597	5839.14	174.3454	9.715353	D15b	4274	5171.693	182.254	10.26751
B6c	3603	4273.444	148.9209	9.267452	D15c	4050	4964.834	173.5801	9.777937
B7a	4821	5998.534	181.9124	9.771551	D16a	7013	8642.38	261.0345	10.52113
B7b	4026	4845.282	154.0757	9.618097	D16b	7491	9242.637	275.5956	10.55484
B7c	4567	5509.66	166.1157	9.227872	D16c	6406	7816.705	241.8965	10.32531
B8a	3233	4113.043	126.1473	8.668049	D17a	5286	6414.475	192.775	9.995452
B8b	3678	4426.31	143.4469	8.672754	D17b	5547	6888.528	201.3579	9.840398
B8c	3912	4868.216	149.2186	8.921358	D17c	3766	4592.214	148.7357	9.336133
B9a	4502	5535.219	183.8252	9.203207	D18a	5491	6902.389	217.8497	10.2296
B9b	3623	4551.923	139.2711	8.900272	D18b	5509	7836.296	219.0991	10.06015
B9c	3416	4188.35	138.3465	8.866524	D18c	5976	8369.149	234.4849	10.46593

**Tabla 8.** Número de especies observadas, esperadas (Chao1), PD Whole Tree e índice de Shannon de cada una de las réplicas.

## Anexo E. Número de OTU en las estaciones de muestreo entre 550 y 800 m



**Figura 24**. Número de OTUs únicas y compartidas en las estaciones de muestreo entre 550 y 800 m de profundidad. El número total de OTUs identificadas por estación se indica en las barras horizontales del lado izquierdo. Las OTUs únicas y compartidas se indican en las barras verticales. Los puntos aislados señalan la barra de OTUs únicas por estación y las que están unidas por líneas indican el número de OTUs compartidas entre los puntos unidos. El gráfico fue realizado en la plataforma en línea del software UpSetR.





**Figura 25.** Número de OTUs únicas y compartidas en las estaciones de muestreo entre 1,000 y 2,000 m de profundidad. El número total de OTUs identificadas por estación se indica en las barras horizontales del lado izquierdo. Las OTUs únicas y compartidas se indican en las barras verticales. Los puntos aislados señalan la barra de OTUs únicas por estación y las que están unidas por líneas indican el número de OTUs compartidas entre los puntos unidos. El gráfico fue realizado en la plataforma en línea del software UpSetR

#### Anexo G. Número de OTU en las estaciones muestreadas entre 2,000 y 3,200 m



**Figura 26**. Número de OTUs únicas y compartidas en las estaciones de muestreo entre 2,000 y 3,000 m de profundidad. El número total de OTUs identificadas por estación se indica en las barras horizontales del lado izquierdo. Las OTUs únicas y compartidas se indican en las barras verticales. Los puntos aislados señalan la barra de OTUs únicas por estación y las que están unidas por líneas indican el número de OTUs compartidas entre los puntos unidos. El gráfico fue realizado en la plataforma en línea del software UpSetR.

Anexo H. Porcentaje de abundancia relativa de los dominios procariontes en los sedimentos de las zonas de muestreo.



**Figura 27.** Grafica de la distribución de los Dominios de microorganismos procariontes identificados en a) los sedimentos muestreados, b) los sedimentos de Perdido y c) de los sedimentos de Coatzacoalcos.



- Metabolismo de aminoácidos
- Transporte de membrana
- Replicación y reparación
- Poco caracterizados
- Movilidad ccelular
- Metabolismo de lípidos
- Procesamiento de información genéticca
- Metabolismo
- Transcripción
- Metabolismo y biosíntesis de glicano
- Familias de enzimas
- Biosíntesis de otros metabolitos secundarios
- Enfermedades infecciosas
- Sistema endócrino
- Signaling Molecules and Interaction
- Cancer
- Sistema nervioso
- Sistema inmune
- Sistema digestivo
- Comunicación celular
- Sistema secretor

- Metabolismo de carbohidratos
- Energy Metabolism
- Traduccción
- Metabolismo de cofactores y vitaminas
- Metabolismo de nucleótidos
- Procesos celulares y señalización
- Folding, Sorting and Degradation
- Metabolismo y biodegradación de xenobióticos
- Transducción de señales
- Metabolismo de Terpenoides and Polyketidos
- Metabolismo de otros aminoácidos
- Crecimiento celular y muerte
- Enfermedades neuridegenerativas
- Transporte y catabolismo
- Adaptación ambiental
- Enfermedades metabólicas
- Sistema circulatorio
- Sistema excretorio
- Enfermedades del sistema inmune
- Enfermedades cardiovasculares

**Figura 28.** Composición porcentual promedio de las 7 categorías del nivel 2 obtenidas en PICRUSt de los microorganismos identificados en todas las estaciones de muestreo.

#### /

## Anexo J. Genes identificados en los contigs generados de las lecturas obtenidas

# por la secuenciación shotgun

 Tabla 9. Genes identificados en el microbioma de los sedimentos de la estación C10.

C10
Cyclic di-GMP phosphodiesterase PA4108 => Cyclic di-GMP phosphodiesterase
UPF0234 protein XC_3703 => hypothetical protein
Probable multidrug resistance ABC transporter ATP-binding/permease protein YheH => putative
multidrug resistance ABC transporter ATP-binding/permease protein YheH
Bifunctional protein FoID => Bifunctional protein FoID protein
Uncharacterized oxidoreductase YtbE => putative oxidoreductase YtbE
Uncharacterized protein YgaU => putative protein YgaU
Probable phospholipid-binding protein MIaC => putative phospholipid-binding protein MIaC
Desiccation/radiation resistance protein DR_1769 => Desiccation/radiation resistance protein
Segregation and condensation protein B homolog => Segregation and condensation protein B
UPF0382 membrane protein YwdK => hypothetical protein
Probable phospholipid ABC transporter permease protein MIaE => putative phospholipid ABC
transporter permease protein MlaE
Probable formate transporter 1 => putative formate transporter 1
Uncharacterized protein YxaL => putative protein YxaL
Putative multidrug export ATP-binding/permease protein SA1683 => Putative multidrug export ATP-
binding/permease protein
Uncharacterized protein YxaL => putative protein YxaL
Uncharacterized zinc protease Rv2782c => putative zinc protease
Uncharacterized protein Rv2307c => putative protein
Uncharacterized FAD-linked oxidoreductase Rv2280 => putative FAD-linked oxidoreductase
Fluoroquinolones export ATP-binding protein Rv2688c => Fluoroquinolones export ATP-binding protein
Putative multidrug export ATP-binding/permease protein SA1683 => Putative multidrug export ATP-
binding/permease protein
UPF0053 inner membrane protein YgdQ => hypothetical protein
UPF0213 protein BH0048 => hypothetical protein
Probable peptidoglycan glycosyltransferase FtsW => putative peptidoglycan glycosyltransferase FtsW
Putative transport protein Rv1101c => Putative transport protein
Probable enoyl-CoA hydratase echA8 => putative enoyl-CoA hydratase echA8
Bifunctional protein Rv2228c => Bifunctional protein
UPF0060 membrane protein YnfA => hypothetical protein
Cyclic di-GMP phosphodiesterase VC_1295 => Cyclic di-GMP phosphodiesterase
Probable FMN/FAD exporter YeeO => putative FMN/FAD exporter YeeO
Uncharacterized signaling protein PA1727 => putative signaling protein
Uncharacterized protein HI_0933 => putative protein
Uncharacterized protein Rv1829 => putative protein
Cyclic di-GMP phosphodiesterase PA4781 => Cyclic di-GMP phosphodiesterase
Probable 3-hydroxybutyryl-CoA dehydrogenase => putative 3-hydroxybutyryl-CoA dehydrogenase
Probable cysteine desulfurase => putative cysteine desulfurase

Uncharacterized ABC transporter ATP-binding protein Rv0986 => putative ABC transporter ATP-binding protein

Cyclic di-GMP phosphodiesterase PA4108 => Cyclic di-GMP phosphodiesterase

Uncharacterized protein YpbG => putative protein YpbG

Uncharacterized glycosyltransferase Rv0539 => putative glycosyltransferase

Probable transcriptional regulatory protein DVU\_2259 => putative transcriptional regulatory protein Cyclic di-GMP phosphodiesterase PA4108 => Cyclic di-GMP phosphodiesterase

Probable endonuclease 4 => putative endonuclease 4

Uncharacterized protein Rv1339 => putative protein

Uncharacterized protein YxaL => putative protein YxaL

ABC transporter ATP-binding/permease protein Rv1747 => ABC transporter ATP-binding/permease protein

Probable polyglutamine synthesis accessory protein MT0602 => putative polyglutamine synthesis accessory protein

Probable type II secretion system protein HxcR => putative type II secretion system protein HxcR Uncharacterized protein HI\_0755 => putative protein

Universal stress protein Rv1636 => Universal stress protein

 Tabla 10. Genes identificados en el microbioma de los sedimentos de la estación C13.

#### C13

Uncharacterized protein YccU => putative protein YccU Bifunctional protein FoID => Bifunctional protein FoID protein Bifunctional protein FoID => Bifunctional protein FoID protein Uncharacterized FAD-linked oxidoreductase Rv2280 => putative FAD-linked oxidoreductase Uncharacterized protein MSMEG 2782/MSMEI 2713 => putative protein/MSMEI 2713 UPF0051 protein SA0778 => hypothetical protein Uncharacterized protein YnbD => putative protein YnbD UPF0060 membrane protein YnfA => hypothetical protein UPF0060 membrane protein YnfA => hypothetical protein Universal stress protein MT2061 => Universal stress protein Probable glycine dehydrogenase (decarboxylating) subunit  $1 \Rightarrow$  putative glycine dehydrogenase (decarbox UPF0111 protein HI 1603 => hypothetical protein Probable adenylyl-sulfate kinase => putative adenylyl-sulfate kinase Putative acetyltransferase SA2342 => Putative acetyltransferase Probable endonuclease 4 => putative endonuclease 4 Cytochrome c oxidase subunit 1 homolog, bacteroid => Cytochrome c oxidase subunit 1, bacteroid UPF0111 protein HI 1603 => hypothetical protein Probable glycine dehydrogenase (decarboxylating) subunit 2 => putative glycine dehydrogenase (decarbox Probable glutamine ABC transporter permease protein GlnM => putative glutamine ABC transporter permeas Segregation and condensation protein B homolog => Segregation and condensation protein B Uncharacterized transporter Rv2685 => putative transporter Probable glycine dehydrogenase (decarboxylating) subunit 1 => putative glycine dehydrogenase (decarbox

Probable peptidyl-prolyl cis-trans isomerase => putative peptidyl-prolyl cis-trans isomerase Uncharacterized HTH-type transcriptional regulator YybR => putative HTH-type transcriptional regulator

Putative phosphoribosyl transferase MT0597 => Putative phosphoribosyl transferase

UPF0073 inner membrane protein YqfA => hypothetical protein

Uncharacterized protein YisK => putative protein YisK

Uncharacterized protein YqeY => putative protein YqeY

Uncharacterized protein MT2060 => putative protein

Probable polyketide biosynthesis zinc-dependent hydrolase PksB => putative polyketide biosynthesis zin

Acetyltransferase SACOL1063 => Acetyltransferase

Uncharacterized protein HI\_1162 => putative protein

Uncharacterized ABC transporter ATP-binding protein Rv1281c => putative ABC transporter ATPbinding pr

F420H(2)-dependent reductase Rv1155 => F420H(2)-dependent reductase

Probable polyketide biosynthesis zinc-dependent hydrolase PksB => putative polyketide biosynthesis zin

Probable cyclic di-GMP phosphodiesterase VC\_1348 => putative cyclic di-GMP phosphodiesterase UPF0758 protein YsxA => hypothetical protein

Ribosome-associated protein L7Ae-like => Ribosome-associated protein L7Ae-like protein

Uncharacterized oxidoreductase SA2266 => putative oxidoreductase

Uncharacterized oxidoreductase SA2266 => putative oxidoreductase

Probable peptidoglycan glycosyltransferase FtsW => putative peptidoglycan glycosyltransferase FtsW UPF0051 protein Rv1461 => hypothetical protein

UPF0301 protein VC\_0467 => hypothetical protein

Uncharacterized ABC transporter ATP-binding protein Rv0986 => putative ABC transporter ATP-binding pro

Probable glutamine ABC transporter permease protein GlnM => putative glutamine ABC transporter permeas

Probable lipid II flippase MurJ => putative lipid II flippase MurJ

Uncharacterized protein Rv1507c => putative protein

Uncharacterized N-acetyltransferase YjaB => putative N-acetyltransferase YjaB

Uncharacterized HTH-type transcriptional regulator YybR => putative HTH-type transcriptional regulator

Uncharacterized HTH-type transcriptional regulator YtcD => putative HTH-type transcriptional regulator Probable bifunctional SAT/APS kinase => putative bifunctional SAT/APS kinase

Probable S-adenosylmethionine-dependent methyltransferase MSMEG\_2350/MSMEI\_2290 => putative S-adenosyl

UPF0098 protein CT\_736 => hypothetical protein

Probable ATP-dependent helicase DinG homolog => putative ATP-dependent helicase DinG UPF0051 protein SA0778 => hypothetical protein

Uncharacterized ABC transporter ATP-binding protein YxIF => putative ABC transporter ATP-binding prote

Probable transcriptional regulatory protein pdtaR => putative transcriptional regulatory protein pdtaR Uncharacterized protein Rv1488 => putative protein

Probable intracellular septation protein A => putative intracellular septation protein A

Uncharacterized protein MT2089 => putative protein

NAD(P)H-dependent FMN reductase PA1204 => NAD(P)H-dependent FMN reductase

Uncharacterized protein YciH => putative protein YciH

Uncharacterized protein YccU => putative protein YccU

Uncharacterized protein MSMEG\_2782/MSMEI\_2713 => putative protein/MSMEI\_2713

Probable ABC transporter ATP-binding protein M6\_Spy0273 => putative ABC transporter ATP-binding protei

Probable peptidoglycan D,D-transpeptidase PenA => putative peptidoglycan D,D-transpeptidase PenA Uncharacterized protein Rv1829 => putative protein

Uncharacterized protein YisK => putative protein YisK

UPF0758 protein => hypothetical protein

Uncharacterized protein HI\_1163 => putative protein

Ribonuclease TTHA0252 => Ribonuclease

Phosphatase Rv3376 => Phosphatase

Probable HTH-type transcriptional regulator YgaV => putative HTH-type transcriptional regulator YgaV Uncharacterized HTH-type transcriptional regulator YtcD => putative HTH-type transcriptional regulator Probable branched-chain-amino-acid aminotransferase => putative branched-chain-amino-acid aminotransfe

Uncharacterized HTH-type transcriptional regulator YbaQ => putative HTH-type transcriptional regulator

Uncharacterized ABC transporter ATP-binding protein YadG => putative ABC transporter ATP-binding prote

UPF0051 protein SA0778 => hypothetical protein

Probable transcriptional regulatory protein YedW => putative transcriptional regulatory protein YedW Uncharacterized oxidoreductase SA2266 => putative oxidoreductase

Probable sulfoacetate--CoA ligase => putative sulfoacetate--CoA ligase

Putative aminotransferase MSMEG\_6286/MSMEI\_6121 => Putative aminotransferase/MSMEI\_6121 Probable adenylyltransferase/sulfurtransferase MoeZ => putative adenylyltransferase/sulfurtransferase Probable Fe(2+)-trafficking protein => putative Fe(2+)-trafficking protein

Uncharacterized amino acid permease YhdG => putative amino acid permease YhdG

Probable type II secretion system protein HxcR => putative type II secretion system protein HxcR Probable pyridine nucleotide-disulfide oxidoreductase RcIA => putative pyridine nucleotide-disulfide o Virulence protein STM3117 => Virulence protein

Ribonuclease TTHA0252 => Ribonuclease

Probable nicotinate-nucleotide pyrophosphorylase [carboxylating] => putative nicotinate-nucleotide pyr

Probable Fe(2+)-trafficking protein => putative Fe(2+)-trafficking protein

Probable ATP-dependent transporter SufC => putative ATP-dependent transporter SufC

Putative esterase Rv1847 => Putative esterase

Probable multidrug ABC transporter ATP-binding protein YbhF => putative multidrug ABC transporter ATP-

Uncharacterized metal-dependent hydrolase YcfH => putative metal-dependent hydrolase YcfH Uncharacterized protein YnbD => putative protein YnbD

Dihydroorotate dehydrogenase B (NAD(+)), electron transfer subunit homolog => Dihydroorotate dehydroge

Uncharacterized ABC transporter ATP-binding protein Rv0073 => putative ABC transporter ATP-binding pro

Probable ATP-dependent transporter SufC => putative ATP-dependent transporter SufC Uncharacterized HTH-type transcriptional regulator YybR => putative HTH-type transcriptional regulator Uncharacterized protein YqgN => putative protein YqgN

Nucleotide-binding protein Rv1421 => Nucleotide-binding protein

UPF0434 protein PFL\_1779 => hypothetical protein

Uncharacterized protein YibN => putative protein YibN

Uncharacterized HTH-type transcriptional regulator YtcD => putative HTH-type transcriptional regulator Monocarboxylate 2-oxoacid-binding periplasmic protein all3028 => Monocarboxylate 2-oxoacidbinding per

Uncharacterized protein HP\_1423 => putative protein

UPF0098 protein CT\_736 => hypothetical protein

Nucleoid-associated protein Cthe\_2143 => Nucleoid-associated protein

Uncharacterized protein HI\_0828 => putative protein

Uncharacterized sugar kinase YdjH => putative sugar kinase YdjH

Ribonuclease TTHA0252 => Ribonuclease

Ribonuclease TTHA0252 => Ribonuclease

Probable cytochrome c oxidase subunit 1 => putative cytochrome c oxidase subunit 1

Putative glutaredoxin Rv3198.1 => Putative glutaredoxin.1

Putative universal stress protein SA1532 => Putative universal stress protein

Probable glycine dehydrogenase (decarboxylating) subunit 1 => putative glycine dehydrogenase (decarbox

Probable bacterial non-heme ferritin => putative bacterial non-heme ferritin

Uncharacterized ABC transporter ATP-binding protein YejF => putative ABC transporter ATP-binding prote

Probable UbiX-like flavin prenyltransferase => putative UbiX-like flavin prenyltransferase

Uncharacterized protein YccU => putative protein YccU

Uncharacterized protein MSMEG\_2782/MSMEI\_2713 => putative protein/MSMEI\_2713

Uncharacterized methyltransferase YcgJ => putative methyltransferase YcgJ

UPF0051 protein Rv1461 => hypothetical protein

Uncharacterized protein YciH => putative protein YciH

Probable cysteine desulfurase => putative cysteine desulfurase

Probable bacterial non-heme ferritin => putative bacterial non-heme ferritin

UPF0173 metal-dependent hydrolase SA1529 => hypothetical protein

Sulfurtransferase Alvin\_2599 => Sulfurtransferase

Segregation and condensation protein B homolog => Segregation and condensation protein B

Probable ATP-dependent transporter SufC => putative ATP-dependent transporter SufC

Multifunctional alkaline phosphatase superfamily protein pRL90232 => Multifunctional alkaline phosphat

Uncharacterized protein Rv2307c => putative protein

Putative anti-sigma factor antagonist TM\_1442 => Putative anti-sigma factor antagonist

Uncharacterized protein BHWA1\_00569 => putative protein

Putative universal stress protein SA1532 => Putative universal stress protein

Aminopeptidase HP\_1037 => Aminopeptidase

Probable transcriptional regulatory protein aq\_1575 => putative transcriptional regulatory protein Purine-binding protein BAB2\_0673 => Purine-binding protein

Uncharacterized oxidoreductase SA2266 => putative oxidoreductase

Probable glycine dehydrogenase (decarboxylating) subunit 2 => putative glycine dehydrogenase (decarbox

Uncharacterized protein Rv2307c => putative protein

Ribonuclease TTHA0252 => Ribonuclease

Uncharacterized protein Rv0628c => putative protein

Uncharacterized protein Rv2307c => putative protein

UPF0047 protein Rv2556c => hypothetical protein

Probable UbiX-like flavin prenyltransferase => putative UbiX-like flavin prenyltransferase

Probable polyketide biosynthesis zinc-dependent hydrolase PksB => putative polyketide biosynthesis zin

Probable serine/threonine-protein kinase YbdM => putative serine/threonine-protein kinase YbdM UPF0051 protein SA0778 => hypothetical protein

Uncharacterized protein Rv1829 => putative protein

Desiccation/radiation resistance protein DR\_1769 => Desiccation/radiation resistance protein

Probable peptidoglycan glycosyltransferase FtsW => putative peptidoglycan glycosyltransferase FtsW UPF0098 protein CT\_736 => hypothetical protein

Nucleotide-binding protein Rv1421 => Nucleotide-binding protein

UPF0235 protein YggU => hypothetical protein

Putative universal stress protein SA1532 => Putative universal stress protein

Cyclic di-GMP phosphodiesterase PA4108 => Cyclic di-GMP phosphodiesterase

Probable nicotinate-nucleotide pyrophosphorylase [carboxylating] => putative nicotinate-nucleotide pyr

Bifunctional protein FoID => Bifunctional protein FoID protein

Uncharacterized protein YyaP => putative protein YyaP

NAD(P)H-dependent FAD/FMN reductase GTNG\_3158 => NAD(P)H-dependent FAD/FMN reductase UPF0716 protein FxsA => hypothetical protein

Probable cytochrome c oxidase subunit 1 => putative cytochrome c oxidase subunit 1

Probable UbiX-like flavin prenyltransferase => putative UbiX-like flavin prenyltransferase

Probable ABC transporter phosphonate/phosphite binding protein PhnD2 => putative ABC transporter phosp

Probable UbiX-like flavin prenyltransferase => putative UbiX-like flavin prenyltransferase

Probable endonuclease 4 => putative endonuclease 4

Probable diguanylate cyclase DgcC => putative diguanylate cyclase DgcC

Uncharacterized ABC transporter ATP-binding protein Rv2564 => putative ABC transporter ATP-binding pro

Uncharacterized protein YciH => putative protein YciH

Sulfurtransferase Alvin\_2599 => Sulfurtransferase

Probable phospholipid-binding protein MlaC => putative phospholipid-binding protein MlaC

Probable phospholipid ABC transporter-binding protein MlaD => putative phospholipid ABC transporter-bi

Uncharacterized protein MT2089 => putative protein

Probable cytochrome c oxidase subunit 1 => putative cytochrome c oxidase subunit 1

Probable tRNA-dihydrouridine synthase => putative tRNA-dihydrouridine synthase

Probable endonuclease 4 => putative endonuclease 4

Probable ATP-dependent transporter SufC => putative ATP-dependent transporter SufC

Putative monooxygenase Rv0793 => Putative monooxygenase

Uncharacterized protein YedJ => putative protein YedJ

Uncharacterized protein Rv0887c => putative protein

3'-5' exoribonuclease MT2234.1 => 3'-5' exoribonuclease.1

UPF0126 inner membrane protein YicG => hypothetical protein

Uncharacterized signaling protein PA1727 => putative signaling protein Probable 4-hydroxyproline 2-epimerase => putative 4-hydroxyproline 2-epimerase Ribonuclease TTHA0252 => Ribonuclease Uncharacterized ABC transporter ATP-binding protein YheS => putative ABC transporter ATP-binding prote Probable cysteine desulfurase => putative cysteine desulfurase UPF0701 protein YicC => hypothetical protein Probable adenylyl-sulfate kinase => putative adenylyl-sulfate kinase Uncharacterized Nudix hydrolase NudL => putative Nudix hydrolase NudL Putative ribosomal protein L7Ae-like => Putative ribosomal protein L7Ae-like protein Uncharacterized Nudix hydrolase NudL => putative Nudix hydrolase NudL Putative ribosomal protein L7Ae-like => Putative ribosomal protein L7Ae-like protein Uncharacterized Nudix hydrolase NudL => putative Nudix hydrolase NudL Uncharacterized Nudix hydrolase NudL => putative Nudix hydrolase NudL Putative ribosomal protein L7Ae-like => Putative ribosomal protein L7Ae-like protein Translation initiation factor IF-3' at k141 12997 position 366 Bifunctional protein FolD protein' at k141 14302 position 607 Beta sliding clamp' at k141\_15017 position 528 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase' at k141 16683 position 1192 Methionine synthase' at k141 22347 position 909 Superoxide dismutase [Mn]' at k141\_25870 position 539

 Tabla 11. Genes identificados en el microbioma de los sedimentos de la estación D18.

D18
UPF0045 protein Rv1898 => hypothetical protein
Uncharacterized amino acid permease YhdG => putative amino acid permease YhdG
Uncharacterized signaling protein PA1727 => putative signaling protein
Bifunctional protein Rv2228c => Bifunctional protein
Probable cytochrome c oxidase subunit 1 => putative cytochrome c oxidase subunit 1
Nucleoid-associated protein Cthe_2143 => Nucleoid-associated protein
Putative formate dehydrogenase SA2102 => Putative formate dehydrogenase
Uncharacterized HTH-type transcriptional regulator YtcD => putative HTH-type tra nscriptional
regulator YtcD
Universal stress protein A homolog 2 => Universal stress protein A



#### Anexo K. Fila de las arqueas identificadas en las estaciones de muestreo.

**Figura 29.** Abundancia de los fila de arqueas identificados mediante la secuenciación de amplicones en las estaciones de muestreo. Las estaciones están ordenadas de menor a mayor profundidad (550-3,200 m bajo el nivel del mar)