

Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



Programa de Posgrado en Ciencias en Ciencias de la vida con orientación en Microbiología

Caracterización de las comunidades microbianas en sedimentos del
Golfo de México, mediante análisis metagenómico

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de Maestro en Ciencias

Presenta:

Mirna Briceida Covarrubias Rodríguez

Ensenada, Baja California, México
2016

Tesis defendida por
Mirna Briceida Covarrubias Rodríguez

y aprobada por el siguiente Comité

Dra. María Asunción Lago Lestón
Director de tesis

Dra. Meritxell Riquelme Pérez

Dr. Helmut Maske Rubach



Dra. Clarisa Elizabeth Galindo Sánchez
Coordinador del Posgrado en Ciencias de la vida

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Mirna Briceida Covarrubias Rodríguez** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la vida con orientación en Microbiología.

Caracterización de las comunidades microbianas en sedimentos del Golfo de México, mediante análisis metagenómico.

Resumen aprobado por:

Dra. María Asunción Lago Lestón
Director de tesis

Los sedimentos marinos cubren aproximadamente el 70% de la superficie terrestre y presentan una alta diversidad de microorganismos, pero debido a la dificultad que existe para cultivarlos, la mayoría no han podido ser caracterizados. En el Golfo de México, la información disponible sobre la microbiota en sedimentos es escasa, principalmente en la Zona Económica Exclusiva Mexicana. Por tal motivo, el principal objetivo de este trabajo es caracterizar a las comunidades de bacterias y arqueas presentes en los sedimentos marinos mediante análisis metagenómico dirigido. Para ello se amplificó la región V4 del gen que codifica para el 16s RNA y los amplicones generados fueron secuenciados utilizando la plataforma Miseq de Illumina. Se obtuvieron muestras de 10 estaciones de la zona sur del Golfo, utilizando un multinucleador y un nucleador de caja Reinek. Por cada estación se obtuvieron dos muestras de la superficie del núcleo: de 0-5 cm y de 5-10 cm, y en dos estaciones también se tomó muestra de la parte más profunda del núcleo (más de 30 cm). Se obtuvieron 748,029 lecturas de buena calidad que correspondieron a 487,927 OTUs. Los resultados de los análisis de diversidad y riqueza de especies muestran que la profundidad de secuenciación es suficiente para representar la comunidad microbiana. El dominio *Arquea* representa alrededor del 40% de los OTUs totales en los sedimentos superficiales, sin embargo, en los sedimentos más profundos esta abundancia se reduce, mientras que su diversidad aumenta. Por su parte, el dominio *Bacteria* presentó mayor abundancia y diversidad en los sedimentos de fondo del núcleo. Los filo con abundancias más altas (89.7% de las lecturas) en los sedimentos superficiales son *Crenarchaeota*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Gemmatimonadetes* y *Chloroflexi*. En el caso de los sedimentos del fondo del núcleo los filo representativos con mayores abundancias son *Crenarqueota*, *Planctomycetes*, *Chloroflexi*, *Proteobacteria*, *Euryarqueota* y *OP8* (82.2% de las lecturas). El análisis de coordenadas principales (PCoA) reveló que no hay diferencias significativas en la composición de la comunidad microbiana entre las muestras de los sedimentos más superficiales así como entre las estaciones. Este hecho, posiblemente está relacionado a la bioturbación causada por la meiofauna. Sin embargo, en las estaciones donde se colectaron sedimentos a más de 30 cm, se observaron diferencias significativas respecto a los sedimentos superficiales, así como entre las dos muestras de fondo del núcleo. Estas diferencias podrían estar relacionadas con la zonación bioquímica de los sedimentos.

Palabras clave: sedimentos marinos, Golfo de México, análisis metagenómico, diversidad y abundancia de procariotas.

Abstract of the thesis presented by **Mirna Briceida Covarrubias Rodríguez** as a partial requirement to obtain the Master of Science degree in Science of life with orientation in Microbiology.

Metagenomic analysis of marine sediments from the Gulf of Mexico.

Abstract approved by:

Dra. María Asunción Lago Lestón
Thesis Director

Marine sediments cover nearly 70 % of the Earth's total surface. A wide variety of microbes inhabit these environments, but most of them have not been described yet due to difficulties to culture them. There are scarce reports on the microbiota of the Gulf of Mexico, particularly on the Exclusive Economic Zone of Mexico. The main goal of this study is to characterize bacterial and *archaea* communities present in marine sediments using a target metagenomic analysis. The V4 region of the 16s RNA gene was amplified and resulting amplicons were sequenced using the Illumina Miseq platform. Samples were collected from 10 different sampling sites from the Southern Gulf of Mexico using a multicore drill and a Reineck box corer. Two samples from the surface of each core (0-5 cm and 5-10 cm) were taken. In addition, a deeper sample from the core (30 cm) was collected in two stations. A total of 748,029 high-quality reads were obtained, representing 487,927 OTUs from the analysis. Species richness and diversity results support that our sequencing depth was appropriate and representative of the total microbe communities. The *Arquea* domain represents nearly 40 % of all OTUs from surface sediments but from deeper ones, this abundance decreases while diversity increases. The Bacteria domain showed higher abundance and diversity from deeper sediments. The most abundant phyla on surface sediments were *Crenarchaeota*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Gemmatimonadetes* y *Chloroflexi*, representing 89.7 % of total reads. Microbes of deeper sediments were represented by *Crenarqueota*, *Planctomycetes*, *Chloroflexi*, *Proteobacteria*, *Euryarqueota* and *OP8* (82.2% of total reads). Principal coordinates analysis (PCoA) revealed no significant differences on the composition of the microbial communities when comparing both, samples obtained from the same core and samples across all stations. This fact could probably be linked to the effect of bioturbation caused by the meiofauna. In stations where deeper sediments were collected, significant differences were observed between deep and top sediments, as well as between sampling sites. These differences could be related with the biogeochemical zonation of sediments.

Keywords: marine sediments, Gulf of Mexico, metegenomic analysis, prokaryotic abundance and diversity.

**A mi padre,
quien me enseñó a volar.**

*“Lo que perdemos, al final siempre vuelve a nosotros.
Aunque a veces no del modo que esperamos”
Luna Lovegood (J.K Rowling)*

Agradecimientos

Mi formación como maestra en ciencias estuvo llena de grandes acontecimientos que concluyeron en aprendizaje. El descubrimiento de nuevas áreas de la ciencia, y el enfoque de “Aquí todo se puede, si lo intentas y te esfuerzas” hicieron que viera el mundo de manera diferente.

Toda esta formación estuvo acompañada de buenos y malos ratos, de periodos donde se avanza rápido y periodos te estancas, y no estás seguro sobre como dar el siguiente paso. En mi caso, lo que ayudó en los momentos en que el trabajo resultaba difícil fueron: el compañerismo, los consejos, las observaciones, el apoyo y las buenas intenciones de las personas que encontré en CICESE.

Es importante mencionar, que entre esas personas destaca la Dra. Ma. Asunción Lago Iestón, mi directora de tesis, a quien le agradezco la confianza, la dedicación, los conocimientos, los ánimos y la seguridad que me dio.

También agradezco a mis sinodales, la Dra. Meritxell Pérez Riquelme y el Dr. Helmut Maske Rubach por las grandes ideas, críticas y observaciones excelentes que llevaron a mejorar el proyecto.

Ademas, quiero agradecer a mi equipo de trabajo, que aportaron sus conocimientos y habilidades a este proyecto. Especialmente a la Dra. Jennyfer Chong, el M. en C. Dante Magdaleno, la M. en C. Karla Sidón y la M. en C. Lluvia Vargas.

Al estar lejos de casa y por un periodo tan prolongado, me di cuenta que la vida se encarga de darte lo que necesitas, cuando lo necesitas. En Ensenda me encontré con una nueva familia, personas que en poco tiempo se ganaron mi confianza y mi cariño. Ellos se volvieron pilares en los momentos más decisivos, tengo que mencionar que en ellos recae algo del crédito de éste trabajo.

Aimée y Leonardo (señores De la Rosa), quiero que sepan que agradezco que me hayan adoptado como hermanita-hija, por hacerme sentir parte de su familia y darme amor fraternal cuando lo necesitaba. No sé que hubiera hecho sin ustedes.

Dulce, Héctor e Ivan, compañeros de estudio que se volvieron buenos amigos. Gracias por estar en las horas más negras (esto va de parte de mi familia también), su presencia siempre ayudó mucho. Pero sobre todo, gracias por las risas: nadie me hizo reír tanto como ustedes.

Anayatzin, agradezco que te acercaras a mí, aunque no tenías que hacerlo, que procurarás mi bienestar. Gracias por la compañía en las noches largas y difíciles.

Rocío, gracias por los ánimos y la confianza en el proyecto. Me diste otra forma de ver las cosas, me ayudaste en cosas que no soy muy buena. Pero sobre todo gracias por el cariño que me das.

Lupita, Nayeli, Danna, Ricardo y Martín, ¡Pequeños! Gracias por las risas, los abrazos y la diversión. Al final no sé si saben que los quiero muchísimo.

A todos gracias por estar en este viaje conmigo.

También agradezco a mi familia (aunque no forman parte de la comunidad CICESE) por el amor, el apoyo, los buenos deseos y las buenas vibras. Por que siempre me han dejado volar.

Por último agradezco al Centro Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE) y el Posgrado en Ciencias de la Vida por permitirme continuar con mi formación académica. Al Proyecto CiGoM al cual pertenece mi proyecto de tesis y a Conacyt por la beca de manutención otorgada (No. de becario 340065).

Índice

	Página
Resumen en español.....	ii
Resumen en inglés.....	iii
Dedicatorias.....	iv
Agradecimientos.....	v
Índice.....	vii
Lista de Figuras.....	x
Lista de tablas.....	xii
Lista de anexos.....	xii
 CAPÍTULO 1. INTRODUCCIÓN	1
1.1 Antecedentes	2
1.1.3 La Metagenómica	5
1.1.4 Herramientas bioinformáticas para el análisis metagenómico.	7
1.1.5 Caracterización microbiana en sedimentos marinos mediante metagenómica.....	8
1.1.6 Área de estudio: El Golfo de México.	10
1.2 Justificación	11
1.3 Hipótesis.....	11
1.4 Objetivos	12
1.4.1 Objetivo general.....	12
1.4.2. Objetivos específicos.....	12
 CAPÍTULO 2. METODOLOGÍA	13
2.1 Muestreo	13
2.1.1 Obtención y almacenamiento de las muestras.....	13
2.2 Preparación previa a la extracción de las muestras.....	14
2.3 Extracción de DNA genómico.....	15
2.4 Cuantificación de DNA.....	17
2.5 Re-Precipitación de DNA.....	17

2.6 Construcción de la biblioteca de secuenciación.	18
2.6.1 Diseño de los oligos.	18
2.6.2 Amplificación.	20
2.7 Electroforesis.	20
2.8 Purificación y normalización del producto de PCR.	20
2.9 Cuantificación de las bibliotecas de secuenciación.	21
2.10 Desnaturalización y dilución de bibliotecas y PhiX.	22
2.11 Secuenciación.	23
2.12 Análisis de calidad de la secuenciación.	25
2.13 Análisis de calidad de los datos.	25
2.14 Análisis en QIIME.	25
2.14.1 Mapping File.	25
2.14.2 “Picking” de OTUS y las secuencias representativas.	26
2.14.3 Análisis de alfa diversidad	27
2.14.4 Análisis de beta diversidad.	27
2.14 Visualización de la diversidad y abundancia de las muestras.	28
 CAPÍTULO 3. RESULTADOS	29
3.1 Muestreo	29
3.2 Extracción, cuantificación, concentración y electroforesis del DNA genómico total de muestras de sedimentos.	29
3.2.1 Extracción	29
3.2.2 Cuantificación y análisis de pureza del DNA extraído.	29
3.2.3 Electroforesis de los productos de extracción de las muestras de sedimentos de fondo del núcleo.	31
3.3 Construcción de las bibliotecas de secuenciación.	32
3.3.1 Amplificación de la región variable V4 del gen 16s rRNA.	32
3.3.2 Normalización y cuantificación de las bibliotecas.	34
3.4 Secuenciación.	34
3.4.1 Análisis del reporte de calidad de la corrida.	34
3.4.2 Análisis de calidad de las lecturas	36
3.5 Análisis de los datos	38

3.5.1 Obtención de los OTUs.....	38
3.5.2 Curva de rarefacción	39
3.5.3 Análisis de Alfa diversidad.....	39
3.5.4 Asignación taxonómica de los microbiomas.	41
3.5.5 Análisis de estructura de las comunidades microbianas de las muestras.	41
3.5.6. Comparación de la diversidad y abundancia de los grupos taxonómicos más abundantes entre muestras de fondo y superficie.....	44
3.5.7 Comparación de diversidad entre las estaciones G44 y H45 (superficie y fondo).....	49
3.5.8 Análisis de Beta diversidad.....	51
 CAPÍTULO 4. DISCUSIÓN	52
4.1 Consideraciones técnicas.	52
4.1.1 Extracción de DNA.....	52
4.1.2 Secuenciación.....	53
4.1.3 Análisis de la calidad de los datos obtenidos en la secuenciación.....	54
4.1.4 Curvas de rarefacción.....	54
4.2 Caracterización taxonómica de los microbiomas.....	55
4.2.1 Proteobacteria.....	56
4.2.2 Cloroflexi	57
4.2.3 Acidobacteria	57
4.2.4 <i>Actinobacteria</i>	58
4.2.5 Arquea	58
4.3 Comparación de la diversidad y abundancia de los microbiomas entre muestras de fondo y superficie.....	59
4.3.1 Alfa diversidad.....	59
4.3.2 Beta diversidad.....	60
 CAPÍTULO 5. CONCLUSIONES.....	62
Literatura citada	63
Anexos.....	66

Lista de figuras

Figura 1. Esquema representativo del gen 16s RNA Ribosomal.....	6
Figura 2. Pipeline de QIIME.....	8
Figura 3. Análisis metagenómico comparativo de 4 muestras de sedimentos del mar de china. Composición de la comunidad microbiana a nivel filo.....	9
Figura 4. Eco-regiones geomorfológicas del Golfo de México.....	10
Figura 5. Representación de las 10 estaciones de muestreo.....	13
Figura 6. Diagrama del flujo de la toma de submuestra de los núcleos.. ..	14
Figura 7. Flujo de trabajo para la construcción de los amplicones etiquetados	18
Figura 8. Flujo de trabajo para la cuantificación de las bibliotecas.	21
Figura 9. Extracciones de DNA de sedimentos de fondo del núcleo.	32
Figura 10. Electroforesis de extracciones de sedimentos con fenol cloroformo.. ..	32
Figura 11. Electroforesis de los productos de PCR	34
Figura 12. Resultado de los análisis de calidad en FastQC. S.....	37
Figura 13. Relación entre los OTU'S identificados y el número de lecturas.....	38
Figura 14. Curva de rarefacción.	39
Figura 15. Análisis de alfa diversidad con el índice de Shannon- Weaver.	40
Figura 16. Estimación y observación de riqueza de especies.	40
Figura 17. Composición de la diversidad y abundancia de los OTUS observados entre las muestras "B" (0-5 cm), "A" (5-10 cm) y F (más de 30 cm) a nivel Filo.....	42
Figura 18. . Diversidad y abundancia de los miembros de los OTU'S observados entre las muestras "B" (0-5 cm), "A" (5-10 cm) y F (más de 30 cm) a nivel clase a.	43

Figura 19. Diversidad y abundancia de <i>Proteobacterias</i> a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45.....	45
Figura 20. Diversidad y abundancia de <i>Cloroflexi</i> a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45.	46
Figura 21. Diversidad y abundancia de <i>Acidobacteria</i> a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45..	47
Figura 22. Diversidad y abundancia de Actinobacteria a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45.”.	48
Figura 23. Diversidad y abundancia de <i>Arquea</i> a nivel clase en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45.....	49
Figura 24. Diagrama de comparación de las muestras G44 y H45 de superficie y fondo	50
Figura 25. Análisis de Coordenadas Principales.....	51

Lista de tablas

Tabla 1. Oligos forward usados para la generación de amplicones.....	19
Tabla 2. Oligos Reverse usados para la generación de amplicones.....	19
Tabla 3. Concentraciones y pureza del DNA genómico extraído de las muestras de sedimentos.....	30
Tabla 4. Concentración y pureza de las muestras G44-1FF y H45-1FF extraídas utilizando el protocolo modificado de MoBio.	31
Tabla 5. Selección de las réplicas a amplificar e índices utilizados para identificar las muestras.	33
Tabla 6. Datos del reporte de la secuenciación..	35
Tabla 7. Lecturas obtenidas y OTU'S identificados por muestra.	36
Tabla 8. Número de OTUs identificados por nivel taxonómico..	41

Lista de Anexos

Anexo 1. La tabla muestra las muestras utilizadas para la secuenciación, la concentración de ADN y sus índices.....	66
Anexo 2. Reacción de PCR.....	66
Anexo 3. Programa para la generación de amplicones (Kozich <i>et al.</i> 2013).	67
Anexo 4. Esquema que representan una sección transversal de hábitats oceánicos (parte superior) y representaciones de zonificación biogeoquímica de los sedimentos (parte inferior)..	67
Anexo 5. Representación esquemática de la participación del domino archaea en los ciclos de: (a) Carbono, (b) de Nitrógeno, y (c) de Azufre.....	67

Capítulo 1. Introducción

Las bases de la Ecología Microbiana surgieron a finales del siglo XIX y principios del siglo XX, debido a la curiosidad científica que despertaban los microorganismos, su origen, su comportamiento y su relación con las otras formas de vida. Sin embargo, no es hasta 1947 cuando el microbiólogo Serguéi Vinogradsky en su libro *Microbiologie du Sol* (Madigan *et al.* 2010) estableció los conceptos e ideas relacionadas con la ecología microbiana. Anterior al trabajo de Vinogradsky los estudios de los microbiólogos estaban centrados en los cultivos puros y de los ecólogos en organismos superiores (Bertrand *et al.* 2015). Esto debido principalmente a la dificultad que presenta trabajar con microorganismos obtenidos directamente de muestras ambientales.

Actualmente la Ecología Microbiana es definida como el estudio de las relaciones de los microorganismos con el ambiente, así como sus interacciones con otros microorganismos y macroorganismos (Barton *et al.* 2011). La Ecología Microbiana evalúa la diversidad, abundancia y distribución de los microorganismos a nivel de ecosistema, de comunidad y de población (Rosas *et al.* 2004) con el fin de entender sus relaciones con el entorno.

Debido a los avances científicos y técnicos de las últimas décadas (los de biología molecular principalmente) ha habido grandes avances en el conocimiento de los microorganismos y los ecosistemas. Recientemente se realizan análisis exploratorios para caracterizar comunidades enteras, con el fin de identificar las funciones potenciales que presentan los microorganismos dentro de su ambiente, debido a que se ha demostrado que algunas de esas características metabólicas pueden presentar un potencial uso biotecnológico.

Aunque se ha avanzado significativamente en el estudio de los microorganismos y su relación con los ecosistemas en las últimas décadas debido a la revolución tecnológica, se piensa que aún hay mucho que dilucidar, debido principalmente a la gran variedad de especies de bacterias y arqueas, el hecho que son cosmopolitas y el rezago que hay en el área debido a la dificultad que presentó su estudio por años.

1.1 Antecedentes

1.1.1 Los sedimentos marinos como hábitat microbiano.

Los sedimentos marinos que cubren las plataformas, los márgenes y los taludes continentales, (también conocidos como la biosfera del subsuelo marino), son definidos como el hábitat que se encuentra bajo el lecho marino (Fry *et al.* 2008).

Al ser el segundo hábitat más extenso del planeta (sólo después de los océanos), presentan una gran variedad de ambientes (desde los sedimentos ricos de las plataformas continentales, lodos volcánicos, lodos carbonatados, de mar profundo y sedimentos pobres, entre otros) (Riding y Awramik, 2000). Además son el mayor reservorio de materia orgánica del planeta. Se ha demostrado que el ecosistema de los sedimentos marinos tiene un papel fundamental en las principales reacciones de mineralización que impulsan los ciclos de carbono, nitrógeno y azufre en los océanos (Madigan *et al.* 2010).

Debido a sus características físico-químicas y condiciones extremas, como temperaturas bajas y presiones elevadas, las comunidades que componen este ecosistema son mayormente meiofauna y microorganismos (bacterias y arqueas principalmente) (Snelgrove, 1997; Parkes *et al.* 2014).

Durante la década de 1990 se resaltó la importancia de los sedimentos marinos como hábitat. Realizando diferentes análisis microbiológicos en sedimentos de diferentes lugares y profundidades se concluyó que éste ofrece una amplia gama de condiciones de vida (Jørgensen y Boecio, 2007) y que es sede de ricas comunidades microbianas (Parkes *et al.* 1994). Además, se concluyó que la biosfera del subsuelo marino compone una parte importante de la biomasa procariota total (Whitman *et al.* 1998). A pesar de que es conocida la importancia de la comunidad procariota en los sedimentos marinos, su estudio ha estado limitado principalmente por las dificultades que presenta caracterizar a los miembros de la comunidad. Es por eso que en los últimos años se han venido implementando nuevos métodos para su estudio.

1.1.2 Herramientas utilizadas para la caracterización de microorganismos de suelos y/o sedimentos.

Es conocido que el principal limitante histórico para el estudio de la diversidad microbiana en suelos, es la dificultad que presenta aislar a los microorganismos. Se estima que solo alrededor del 1% de la población bacteriana presente en los suelos puede ser cultivada mediante el uso de prácticas de

laboratorio estándar, pero no se sabe si ese 1% es representativo (Kirk *et al.* 2004). Hasta hace algún tiempo esto era una limitante a la hora de investigar microbiomas en los ecosistemas en general, ya que al no poder cultivarlos no era posible detectarlos. Gracias al desarrollo de nuevas técnicas de Biología Molecular en las últimas décadas, la Ecología Microbiana se ha beneficiado de estos avances que han contribuido en gran medida al estudio y generación de conocimiento sobre de la composición de las comunidades microbianas sin la necesidad de cultivo (Escalante, 2008).

Entre las herramientas moleculares más importantes se destacan:

- DGGE/TGGE (*Denaturing Gradient/ Temperature Gradient Gel Electrophoresis*).
- FISH (*Fluorescence in situ hybridization* - hibridación fluorescente *in situ*).
- Secuenciación.

1.2.1.1 DGGE/TGGE (Denaturing Gradient/ Temperature Gradient Gel Electrophoresis).

Este tipo de análisis se basa en la elaboración de perfiles moleculares específicos de comunidades microbianas permitiendo la diferenciación de las mismas por la composición de sus miembros. Estos perfiles se caracterizan por el número, posición (presencia o ausencia de bandas particulares) e intensidad de las bandas, donde cada banda se toma como la representación de un grupo taxonómico diferente. De este modo se caracterizan las comunidades y además se cuantifican.

En estos análisis se emplean fragmentos de DNA específicos (16S) de los miembros de la comunidad usando cebadores específicos que serán amplificados mediante la Reacción en Cadena de la Polimerasa (PCR, por sus siglas en inglés). Posteriormente, los fragmentos son sometidos a electroforesis desnaturalizante, lo que permite su separación, que está determinada por la composición de la secuencia (contenido de G y C). Las moléculas de DNA con diferentes secuencias de nucleótidos tienen diferentes puntos de fusión por lo tanto terminan su migración en diferentes posiciones. El gradiente de desnaturalización es generado químicamente con una mezcla de urea y formamida generalmente (DGGE) o por un gradiente de temperatura (TGGE) (Stackebrandt *et al.* 2006).

1.1.2.2 FISH (Fluorescence in situ hybridization / Hibridación fluorescente in situ).

La técnica de FISH se basa en la hibridación de dos secuencias de ácidos nucleicos complementarias a sondas de oligonucleótidos específicos marcadas con fluoróforos para el taxón de interés y su observación mediante un microscopio de epifluorescencia directamente en el ambiente. La

muestra es clasificada según la presencia o ausencia de la señal emitida por la sonda específica para determinado grupo taxonómico .

En ecología microbiana resulta una técnica muy útil ya que permite determinar la identidad, la abundancia y la actividad relativa de microorganismos *in situ*. Ésta técnica permite la visualización de células individuales y su relación con otras células, lo que facilita el estudio de ecosistemas (Tortora *et al.* 2007). Además con ella es posible detectar microorganismos que no son cultivables

1.1.2.3 Secuenciación

Se define como el conjunto de métodos y técnicas bioquímicas cuya finalidad es la determinación de los nucleótidos presentes en un fragmento de DNA (Madigan *et al.* 2010).

En el año 1977 se desarrollaron dos de los métodos de secuenciación más importantes, basados en DNA. El método de Maxam y Gilbert, o método de rompimiento químico y el método enzimático o de Sanger. Estos dos métodos revolucionaron la secuenciación. El método de Sanger fue el más aceptado y usado por más de 30 años, ya que presentaba algunas ventajas como la reducción del manejo de productos tóxicos y radioisótopos. Pero con el proyecto del genoma humano, el requerimiento de un mayor rendimiento de secuenciación del DNA aumentó, lo que llevó al desarrollo de nuevas tecnologías tales como, la electroforesis capilar automatizada (Schuster, 2008).

Sin embargo, la comunidad científica necesitaba un mayor rendimiento de secuenciación, además una tecnología más accesible, lo cual fue posible con el desarrollo y evolución de tecnologías de secuenciación como las plataformas llamadas “*second-generation*” (la tecnología de secuenciación por síntesis, desarrollada por 454 Life Sciences y la plataforma Illumina-Solexa Genome Analyzer) (Schendure and Ji, 2008).

En los últimos años las tecnologías de secuenciación han avanzado rápidamente, lo que ha generado mejores condiciones para el análisis y comprensión de comunidades microbianas, gracias al desarrollo de la secuenciación masiva o secuenciación de última generación, NGS (por sus siglas en inglés “*Next-generation sequencing*”), que permite secuenciar miles de secuencias en paralelo permitiendo el análisis de muchas muestras en una sola secuenciación y obtener una alta cobertura de secuencia por muestra (Illumina, 2013).

Hasta el día de hoy existen diferentes plataformas de secuenciación entre ellas Illumina, PacBio, Ion Torrent y 454, todas con sus pros y contras, así que la plataforma de secuenciación es elegida tomando en cuenta muchos factores, tales como el objetivo del proyecto, los costos y el tiempo.

Sin embargo la plataforma de secuenciación líder es Illumina, principalmente porque presenta características que hacen que su uso sea más amigable para la comunidad científica, tales como: i) Rapidez, ii) Flujo de trabajo sencillo y optimizado, iii) Presenta diferentes opciones de secuenciación, iv) la cantidad de DNA de entrada es de nanogramos, v) Alto porcentaje de precisión, vi) Alto rendimiento, vii) Los datos provenientes de la secuenciación pueden ser transferidos, almacenados y analizados de manera fácil (Quail *et al.* 2012 e Illumina inc., 2015).

1.1.3 La Metagenómica

La Metagenómica, también conocida como genómica ambiental o genómica de comunidades, es una disciplina reciente. Se define como el estudio de los metagenomas (genomas colectivos de microorganismos provenientes de una muestra ambiental) para proporcionar información de la diversidad y ecología microbiana de un ambiente específico (<http://www.nature.com/subjects/metagenomics>). Supone la extracción, secuenciación y análisis de los genomas presentes dentro de una comunidad microbiana, lo que posibilita el estudio de una gran variedad de genes y sus productos (Uribe-Vélez, 2009).

El término “metagenómica” apareció por primera vez en el año 1998, acuñado por Jo Handelsman y su equipo de investigación, para referirse a la extracción y secuenciación de DNA de una muestra ambiental de suelo (Handelsman *et al.* 1998). Ellos partieron de la idea de que la filogenia molecular de RNA ribosomal (rRNA), en particular la pequeña subunidad (SSU o 16S) rRNA (Figura 1), es la clave para la identificación y diferenciación de los microorganismos basándose en los trabajos de Woese *et al.* (1977) y Woese (1987):

- I. Son moléculas que tienen regiones altamente conservadas que flanquean 9 regiones de secuencia variable (V1-V9), las regiones conservadas pueden ser utilizadas para amplificar las regiones variables (Woese *et al.* 1977; Woese 1987 y Dethlefsen *et al.* 2008).
- II. El DNA puede amplificarse fácilmente por PCR y ser secuenciado, con la ventaja de que los microorganismos no cultivables pueden ser estudiados y caracterizados (Hill *et al.* 2000).
- III. Se pueden establecer relaciones en diferentes niveles jerárquicos mediante un análisis comparativo de secuencias (Hill *et al.* 2000).

- IV. Una vez amplificado el gen 16s rRNA es posible analizar su diversidad en la comunidad para tener una idea de la complejidad de la misma (Hill *et al.* 2000).

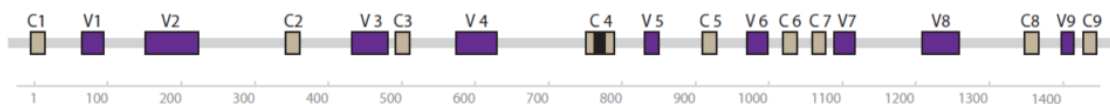


Figura 1. Esquema representativo del gen 16s RNA Ribosomal. Región variable (púrpura) y conservada (gris). (Tomado de Illumina inc., 2013).

Hoy en día la metagenómica usa la secuenciación de última generación lo que permite detectar miembros de la comunidad con baja abundancia, y que con otras técnicas no era posible. Éste método también permite analizar de forma exhaustiva todos los genes de todos los organismos presentes en una muestra compleja (Illumina, 2013).

Actualmente los estudios metagenómicos usan dos aproximaciones diferentes: secuenciación de amplicones y secuenciación *shotgun*. El primero es usado principalmente para resolver la composición o la dinámica de una comunidad específica y el segundo (conocido como metagenómica funcional) es mayormente utilizado para identificar el potencial metabólico presente en los genes de los miembros de una comunidad microbiana obtenidos directamente de una muestra ambiental (Illumina, 2013 y 2015).

- i) Secuenciación de amplicones: Este método es el más utilizado. En él se toma una muestra ambiental (suelo, agua, etc.) y se extrae el DNA de todas las células presentes en la muestra, posteriormente se utiliza un marcador taxonómicamente informativo que sea común en casi todos los organismos de interés (ej. el gen 16S del rRNA para bacterias y arqueas), el cual es amplificado mediante PCR. Los amplicones resultantes son secuenciados y caracterizados utilizando herramientas bioinformáticas para determinar qué microorganismos están presentes en la muestra y cuál es su abundancia (Illumina, 2013 y 2015).

La secuenciación de amplicones 16S ha sido utilizada en una gran variedad de ambientes. Pero aunque es una gran herramienta, presenta ciertas limitaciones, como la presencia de sesgos relacionados con la amplificación previa del marcador y la secuenciación (Poretsky *et al.* 2014). Además, esta aproximación solo proporciona información sobre la composición taxonómica de la comunidad, es imposible resolver

directamente las funciones biológicas asociadas a estos taxones utilizando este método (Langille *et al.* 2013). Por último, la secuenciación de amplicones está limitada al análisis de los taxones para los que se conocen marcadores informativos, microorganismos nuevos o muy divergentes, son difíciles de estudiar. Además, la resolución del gen 16s rRNA es limitada entre especies estrechamente relacionadas (Poretsky *et al.* 2014).

- ii) Secuenciación *shotgun*: En este enfoque, también se extrae el DNA genómico de todas las células contenidas en la muestra ambiental, pero en este caso no se utiliza solo un marcador informativo para la amplificación, sino todo el DNA presente en la muestra es secuenciado. Esto da lugar a estimaciones robustas de la composición y diversidad de la comunidad microbiana. Como todos los genes están presentes, tenemos información tanto de genes taxonómicamente informativos (como los procedentes de la subunidad 16S rRNA), así como de otros que nos permiten conocer las funciones biológicas codificadas en el genoma. Como resultado, los datos de metagenómica *shotgun* ofrecen la oportunidad de explorar simultáneamente dos aspectos de una comunidad microbiana: ¿quién está ahí? y ¿qué es lo que hace? (Sharpton, 2014). Aunque la metagenómica *shotgun* también presenta varios desafíos: la gran cantidad de datos que se generan, lo que complica el análisis, y además es mucho más costosa que la secuenciación de amplicones.

1.1.4 Herramientas bioinformáticas para el análisis metagenómico.

La implementación de las nuevas tecnologías de secuenciación trajo consigo el desarrollo de programas de cómputo capaces de analizar la gran cantidad de datos obtenidos. Uno de los más utilizados es el QIIME.

QIIME (Quantitative Insights Into Microbial Ecology) es un software de código abierto ampliamente utilizado para el análisis e interpretación de datos de secuenciación crudos procedentes de comunidades microbianas. Un análisis estándar en QIIME inicia con los datos de secuenciación de una o más tecnologías de secuenciación. Posteriormente para analizar los datos provenientes de una comunidad microbiana se escriben una serie de comandos en la ventana de la terminal y después se pueden ver de forma gráfica y textual (Kuczynski *et. al.*, 2011) (Figura 2). Este software permite la

realización de análisis de diversidad, asignación taxonómica, reconstrucción filogenética entre otras cosas. Pero solo se pueden analizar lecturas provenientes de secuenciación de amplicones.

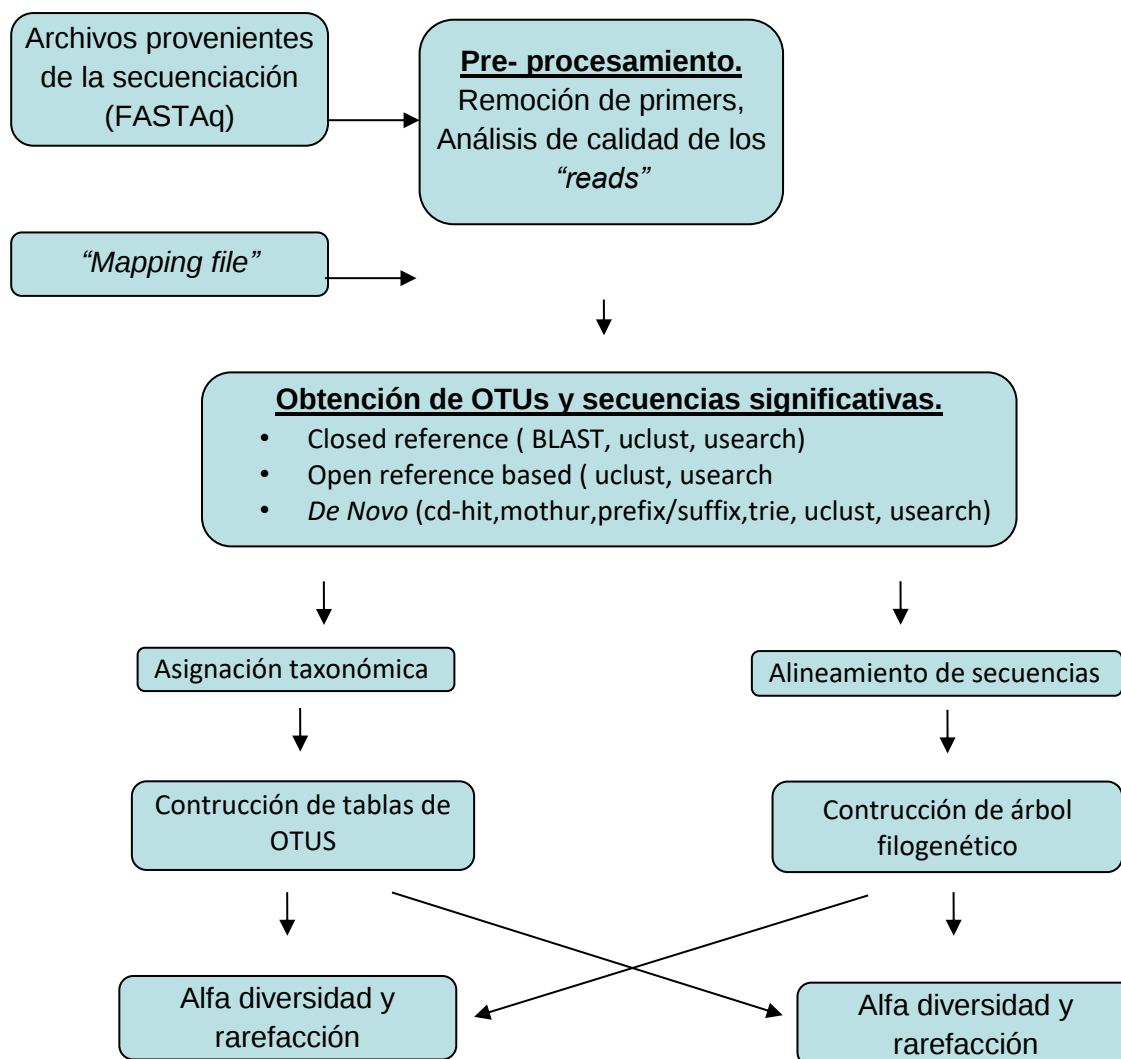


Figura 2. Pipeline de QIIME.

1.1.5 Caracterización microbiana en sedimentos marinos mediante metagenómica.

El desarrollo y abaratamiento de las herramientas utilizadas en la metagenómica en años recientes, han permitido la exploración del mundo microbiano con mayor precisión. Como consecuencia la comprensión de la diversidad microbiana del suelo y subsuelo marino ha incrementado de manera exponencial. Es por esto, que en la última década se han realizado grandes proyectos enfocados en el estudio de la diversidad microbiana en sedimentos de diversas partes del mundo.

Hace un par de años Parkes *et al.* (2014) realizaron un meta-análisis con datos de secuenciación obtenidos de diferentes plataformas, pertenecientes a comunidades procariotas de sedimentos provenientes de diferentes partes del mundo, con el fin de identificar a los miembros presentes en las comunidades procariotas de los sedimentos marinos a nivel global.

Como resultado obtuvieron que las comunidades procariotas en sedimentos superficiales en promedio están compuestas por los siguientes grupos bacterianos: *Cloroflexi*, *Planctomycetes*, JS1 (grupo candidato), *Alfa*, *Gamma*, *Delta* y *Epsilon proteobacteria*, *Bacteroidetes*, *Firmicutes*, *Actinobacteria*, *Cianobacterias*, *Spirochaetes*, OP8 y OP11. Respecto al filo Arquea, los grupos más representativos fueron: *Crenarqueota*, *Euriarqueota* y *Thaumarqueota*.

Además de los análisis en los que se pretende caracterizar a las comunidades del sedimento, hay también aquellos en los que se analiza la dinámica de las comunidades con respecto a variables físicas tales como la profundidad o la presencia de algún accidente geográfico. Como es el caso del estudio comparativo que realizaron Zang *et al.* (2015) en el mar de China, donde se analiza el impacto del ambiente sobre la dinámica de las comunidades de microorganismos (Figura 3), comprobando que las comunidades microbianas responden de manera diferente con respecto a la profundidad, ubicación y condiciones extremas.

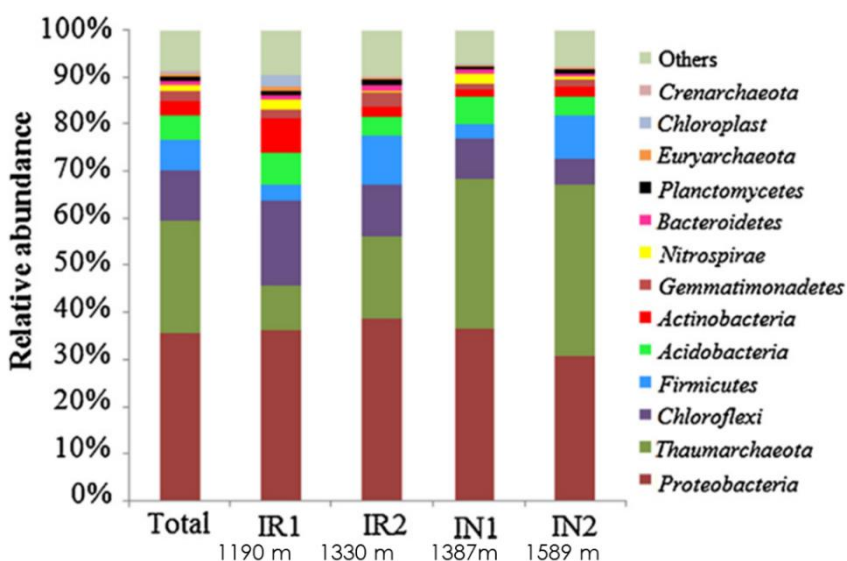


Figura 3. Análisis metagenómico por fragmentos de 4 muestras de sedimentos del mar de china. Composición de la comunidad microbiana a nivel filo. Se utilizaron cuatro muestras (IN1, IN2, IR1 e IR2), de las cuales IN1 e IN2 se encuentran en las regiones próximas y distantes, respectivamente, de chimeneas hidrotermales activas en Iheya Norte (IN), mientras IR1 e IR2 se encuentran en regiones cerca y distante, respectivamente, de las chimeneas hidrotermales activas en Iheya Ridge (IR). Cada color representa el porcentaje de las lecturas etiquetadas por filo respecto al total de lecturas etiquetadas del taxón, se observan claras diferencias entre las muestras (respecto a la diferencia de profundidades, así como con la cercanía a las chimeneas (Zang *et al.* 2015).

1.1.6 Área de estudio: El Golfo de México.

El Golfo de México se localiza en la parte central y meridional de Norteamérica. Es un sistema ambiental muy diverso, debido a sus dimensiones (superficie de 1.6×10^6 de km^2 y un volumen aproximado de 2.3×10^6 km^3), su ubicación geográfica y su geomorfología (Carson, 1980). La profundidad del Golfo es en promedio de 1615 m, siendo la región más profunda la fosa de Sigsbee, ubicada en el cuadrante suroeste, y donde las profundidades llegan hasta los 4384 m (Tunnell *et al.* 2013). El fondo del Golfo está compuesto principalmente por varios tipos de sedimentos: carbonatados biogénicos y terrígenos, de arena limosa y limo (Bouma, 1972).

Con base en las características geomorfológicas, la parte mexicana del Golfo ha sido dividido en 5 regiones: a) La plataforma del Golfo de México norte, b) Talud del Golfo de México norte, c) Plataforma del Golfo de México sur, d) Talud del Golfo de México sur, y e) Planicie del Golfo de México (Figura 4) (Lara-Lara *et al.* 2008).

El Golfo ha sido objeto de numerosas investigaciones, aunque en la mayoría de los casos los estudios se han basado en el análisis de los componentes bióticos presentes en la columna de agua, existiendo muy poca información referente a las comunidades que habitan en los sedimentos (Caso *et al.* 2004).

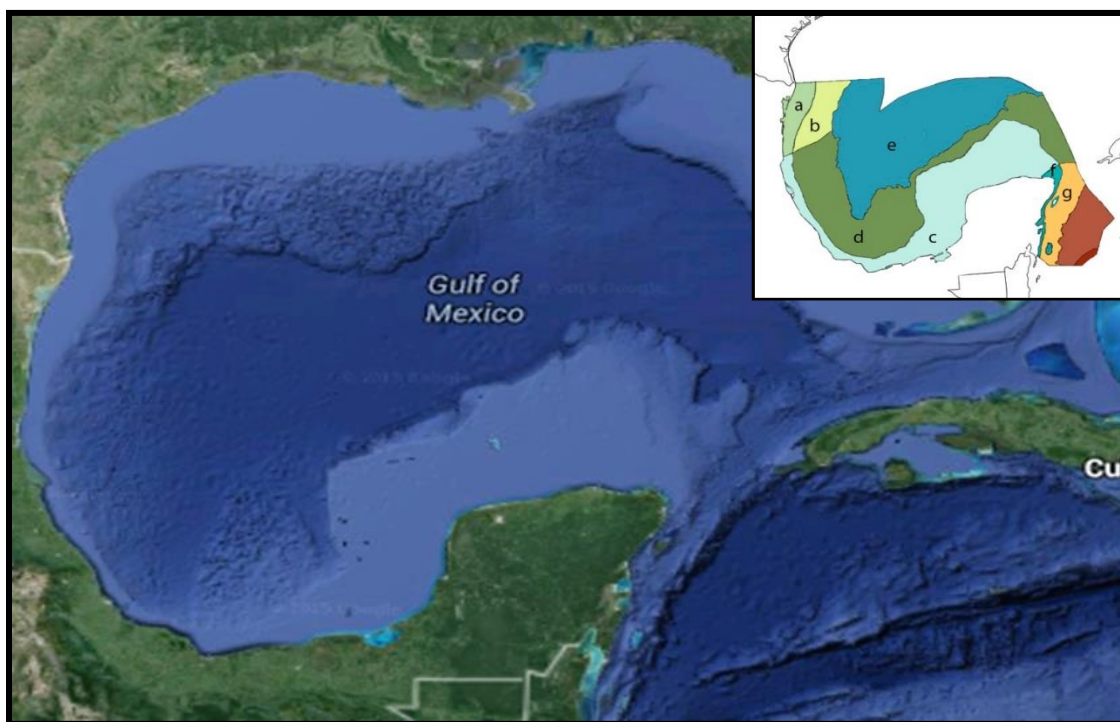


Figura 4. Eco-regiones geomorfológicas del Golfo de México: a) La plataforma del Golfo de México norte, b) Talud del Golfo de México norte c) Plataforma del Golfo de México sur, d) Talud del Golfo de México sur, e) Planicie del Golfo de México. Obtenido de Lara-Lara *et al.* (2008).

Sin embargo, a partir del derrame de *Deepwater horizon* en el 2010, los estudios enfocados en la microbiota presente en los sedimentos del Golfo de México se han incrementado. El principal enfoque que tienen estos estudios es conocer la respuesta de las comunidades microbianas respecto al derrame (Kimes *et al.* 2014). Además, se ha implementado la búsqueda de microorganismos endémicos degradadores de hidrocarburos mediante estudios metagenómicos (Koo *et al.* 2013), con el objetivo de identificarlos y estudiarlos, ya que en estudios previos al derrame del 2010 se detectó la presencia de poblaciones en sedimentos que eran capaces de degradar diferentes compuestos de hidrocarburos (principalmente de zonas del Norte del Golfo) y que fueron de gran ayuda para mitigar el impacto ambiental (Joye *et al.* 2004).

1.2 Justificación

El Golfo de México es una de las zonas con mayor producción de petróleo. En la parte mexicana destaca la bahía de Campeche, que además es considerada una región con alto riesgo para accidentes petroleros. Debido a que se conoce el riesgo de derrames petroleros en la zona económica exclusiva del golfo de México (ZEE), resulta alarmante que muchos de sus ecosistemas hayan sido poco estudiados, entre ellos los sedimentos marinos. Es por eso que en el presente estudio se pretende caracterizar a las comunidades de bacterias y arqueas presentes en dichos sedimentos, con el fin de evaluar su composición y así ser capaces de identificar posibles variaciones de la microbiota residente frente a una posible contingencia.

1.3 Hipótesis

H1: Debido a que las condiciones ambientales en los sedimentos de los 10 sitios de muestreo analizados del sur del Golfo de México son disímiles, la composición de las comunidades microbianas presentes es diferente.

H2: Como las características de los sedimentos cambian conforme aumenta la profundidad en la columna sedimentaria, las comunidades de bacterias y arqueas de los sedimentos más superficiales son diferentes respecto a los sedimentos del fondo del núcleo.

1.4 Objetivos

1.4.1 Objetivo general

Identificar las comunidades procariotas presentes en sedimentos del Golfo de México, mediante análisis metagenómico.

1.4.2. Objetivos específicos

- I. Generar una biblioteca metagenómica de las comunidades procariotas presentes en los diferentes sitios de muestreo.
- II. Identificar los principales grupos taxonómicos de bacterias y arqueas presentes en las zonas de muestreo.
- III. Analizar la distribución y diversidad de las comunidades entre los diferentes sitios de muestreo.
- IV. Analizar las diferencias en la composición de las comunidades microbianas respecto a la profundidad de la columna sedimentaria.

Capítulo 2. Metodología

2.1 Muestreo

El muestreo se realizó durante la campaña oceanográfica XIXIM-4 que tuvo lugar entre el 27 de agosto y el 21 de septiembre del 2015, en la Zona Económica Exclusiva (ZEE) del Golfo de México. Las muestras se colectaron en 10 sitios de muestreos distribuidos en la zona con profundidades que van de 1268 a 3768m (Figura 5).

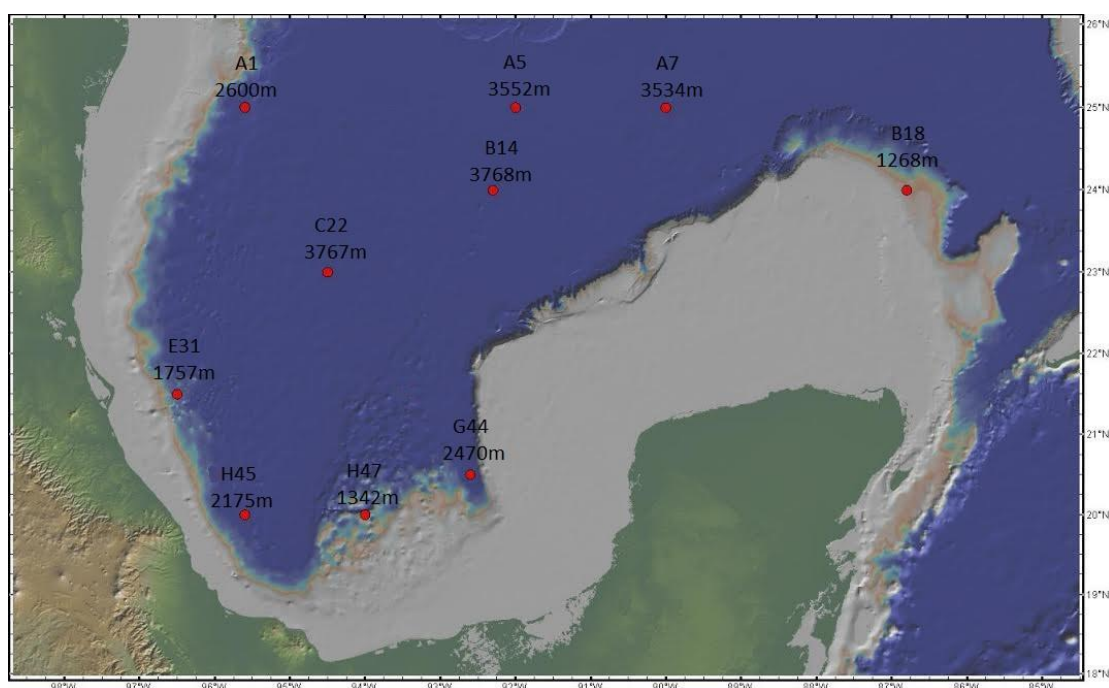


Figura 5. Representación de las 10 estaciones de muestreo donde se colectaron las muestras de sedimento durante el crucero oceanográfico XIXIMI-4.

2.1.1 Obtención y almacenamiento de las muestras.

La extracción de las muestras se realizó mediante un multinucleador (estaciones E31, C22, G44, H45 y H47) y un nucleador de caja Reinek (estaciones A1, A5, A7, B14 y B18). Una vez obtenidos los núcleos, se tomaron 5 muestras (réplicas) de uno de los núcleos por cada estación (aproximadamente 10 mL de sedimento por muestra) con una cilindro de aproximadamente 10 cm de largo. Posteriormente esta muestra fue dividida en dos para representar dos profundidades dentro de la columna sedimentaria. De los 0 a los 5 cm se le denominó “A” y de los 5 a los 10 cm se le dominó “B”. Además se

tomaron muestras (5 réplicas) de aproximadamente 5 mL de sedimento del fondo del núcleo (a una profundidad de 30cm aproximadamente) en dos estaciones.

El método utilizado para la obtención de las muestras fue el siguiente (Figura 6):

- I. Mediante el uso de jeringas (10 mL) sin punta se tomaron 5 submuestras.
- II. Se sacaron las jeringas y fueron vaciados los primeros 5 mL (correspondientes a una profundidad de 5 a 10 cm, anotación A) de sedimento en tubos *Falcon* de 15 mL que contenían amortiguador de sacarosa 25% w/v (previamente homogeneizados mediante agitación). Los 5 mL de sedimento restantes (correspondientes a los primeros 5 cm, anotación B) fueron colocados en otro set de tubos de 15 mL. Posteriormente los tubos con tapa fueron sellados con *Parafilm*.
- III. Las muestras se almacenaron en bolsas *Ziploc* a -20°C durante la campaña y su transporte al laboratorio. Posteriormente fueron almacenadas a -80°C hasta su procesamiento.

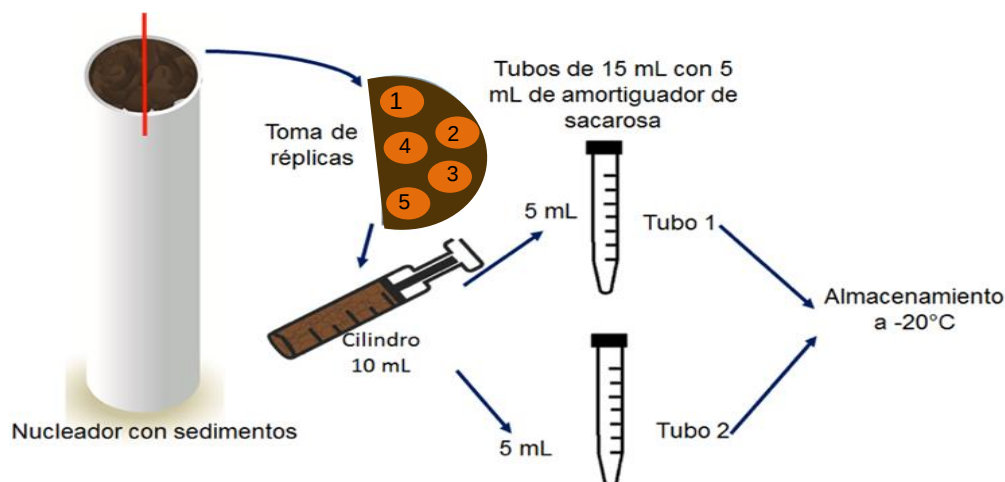


Figura 6. Diagrama del flujo de la toma de submuestra de los núcleos (Imagen tomada del protocolo, para la de obtención de muestras de sedimentos, equipo de metagenómica de sedimentos).

2.2 Preparación previa a la extracción de las muestras.

- I. Las muestras almacenadas a -80°C se sacaron del ultra congelador y se colocaron en gradillas a temperatura ambiente para que se descongelaran.
- II. Se verificó que cada uno de los tubos estuviera descongelado y se le dio vortex a máxima potencia hasta homogenizar la muestra.

Para separar el sedimento del buffer de sacarosa los tubos fueron centrifugados a 10,000 xg durante 1 minuto; posteriormente se descartó la fase acuosa.

2.3 Extracción de DNA genómico.

Se extrajo el DNA genómico de las muestras colectadas. Las muestras se dividieron en cuatro grupos para facilitar su extracción. El orden de extracción fue el siguiente:

Primer grupo: C22-A y B, B14-A y B, A7-A y B, y E3-A y B.

Segundo grupo: B18-A y B, H47-A y B, y A1-A y B.

Tercer grupo: A5-A y B, G44-A y B, y H45-A y B.

Cuarto grupo: G44-F y H45-F

Para la extracción del DNA se utilizó el kit *PowerSoil® DNA Isolation Kit*. Se llevó a cabo siguiendo el protocolo propuesto por el fabricante (detallado abajo) a excepción de dos muestras del cuarto grupo (G44-FF y H45-FF) en cuyas extracciones el protocolo se modificó ligeramente, añadiendo 200 µL de Fenol cloroformo isoamílico posterior al paso II del protocolo proporcionado por el kit, se mezcló por inversión por 1.5 min y posterior mente se le dio vortex por 30 s. También se modificó el paso del lavado del ADN, se usaron 600 µL de solución C4 y 600 µL de Etanol al 100%.

Todas las centrifugaciones se realizaron a temperatura ambiente.

- I. Se pesaron 0.25gr de sedimento por cada réplica, que fueron colocados en los tubos *Powerbead* proporcionados por el kit; se les dio vortex suavemente, con el objetivo de mezclar el buffer que los tubos contienen y homogenizar la muestra.
- II. Se agregaron 60 µL de la solución C1 y se invirtió varias veces el tubo.
La solución C1 contiene SDS y otros agentes disruptivos requeridos para una lisis celular completa.
- III. Los tubos se colocaron horizontalmente y se les dio vortex a máxima velocidad por 10 min, con el fin de ayudar en la lisis celular y a homogenizar la muestra.
- IV. Se retiraron los tubos del adaptador y fueron centrifugados a 10,000 xg durante 30 segundos.
- V. El sobrenadante fue transferido a un nuevo de 2mL al cual se le añadieron 250 µL de Solución C2 y se dio vortex por 5 segundos. Se incubó posteriormente a 4°C por 5 min.

La solución C2 contiene un agente que precipita el material inorgánico y orgánico (no DNA) incluyendo sustancias húmicas, restos celulares y proteínas. Esto para evitar la contaminación, que puede reducir la pureza del DNA.

- VI. Se centrifugaron los tubos por 1 minuto a 10,000 xg. Luego de centrifugar se transfirieron 600 μ L de sobrenadante (evitando el pellet formado, ya que contiene todo el material que precipitó la solución C2) a un tubo nuevo de 2mL y se le agregaron 200 μ L de la solución C3, y se agitó levemente. Posteriormente se incubó a 4°C por 5 min.

La solución C3 es un segundo agente que precipita el material inorgánico y orgánico (no DNA) incluyendo sustancias húmicas, restos celulares y proteínas.

- VII. Los tubos fueron centrifugados por 1 minuto a 10,000 xg, y se trasfirieron 750 μ L del sobrenadante a un tubo nuevo de 2 mL.

- VIII. Se agregaron 1.2 mL de la solución C4 (ésta se agitó antes de usarse) al tubo con los 750 μ L del sobrenadante y se le dio vortex por 5 segundos.

- IX. Se transfirieron 600 μ L de la solución al *Spin Filter* (proporcionado por el Kit), se centrifugó por 1 minuto a 10,000 xg y se descartó la solución filtrada. Se repitió éste paso hasta que la solución con el sobrenadante se terminó.

- X. Se le agregaron 500 μ L de la solución C5 al *Spin Filter* y se centrifugó por 30 segundos a 10,000 xg.

C5 es una solución basada en etanol, cuya función es lavar el DNA que está en la membrana del filtro de los Spin Filter.

- XI. Se descartó la solución filtrada y se centrifugó por 1 minuto a 10,000 xg.

Con el fin de eliminar los residuos de la solución C5, ya que el etanol puede interferir con muchas aplicaciones del DNA en extracción.

- XII. Cuidadosamente se cambió el filtro a un tubo nuevo de 2 mL, evitando salpicar el filtro con la solución C5. Se agregaron 15 μ L de la solución C6 en el centro de la membrana del filtro y

centrifugó por 30 segundos a 10,000 xg. Se repitió la acción hasta filtrar 50 µL. Se retiró el filtro del tubo y se colocó en uno nuevo.

La solución C6 es un buffer de elución.

- XIII. El DNA contenido en el primer tubo se almacenó a -20°C. Al tubo nuevo con el filtro se le agregaron 50 µL de la solución C6 en el centro de la membrana del filtro, después se centrifugó por 30 segundos a 10,000 xg. Se retiró el filtro con mucho cuidado y se descartó. El tubo con DNA se marcó y se almacenó a -20°C para tener DNA extra.

2.4 Cuantificación de DNA.

Después de realizar todas las extracciones se cuantificó el DNA, mediante espectrofotometría utilizando el *NanoDrop lite* (Thermo Scientific).

2.5 Re-Precipitación de DNA.

Después de cuantificar el DNA de las muestras superficiales (A y B), se obtuvieron valores negativos. Por esta razón se decidió realizar una re-precipitación mezclando las dos eluciones obtenidas en el proceso de extracción (volumen final de 100 µL).

Para precipitar el DNA se realizó el siguiente método:

- I. A los 100 µL de elución de DNA se le agregaron 4 µL de NaCl 5M. Se mezcló invirtiendo de 3 a 5 veces los tubos.
- II. Posteriormente se añadieron 200 µL de etanol absoluto frío, se invirtió de 3 a 5 veces para mezclar e incubó a -80°C por 15 min para precipitar el DNA.
- III. Se centrifugó a 10,000 xg durante 5 min y se decantó el etanol.
- IV. Se utilizó el “*speed vac*” *Eppendorf* a máxima velocidad para evaporar el líquido con etanol.
- V. El pellet se re-suspendió en 40 µL agua grado HPLC estéril.

Después de precipitar el DNA de las muestras se volvió a cuantificar usando un nanodrop.

2.6 Construcción de la biblioteca de secuenciación.

La construcción de la biblioteca de secuenciación consta de 2 etapas principales: la amplificación de la región de interés y la ligación de los adaptadores. Con el fin de optimizar la construcción de nuestras bibliotecas se utilizó el enfoque propuesto por Kozich *et al.* (2013), en la cual la amplificación y la ligación de los oligos se realizan en un solo paso. Para ello se utilizan oligos compuestos que contienen: el oligo para amplificar la región V4 del 16s rRNA (285pb), el adaptador Illumina (33pb), el índice (8pb), el PAD u oligo de secuenciación (10) y un *linker* (para unir estos fragmentos con la secuencia del oligo para la región específica) (Figura 7). Al unir la secuencia de los oligos específicos con las etiquetas necesarias para la secuenciación se disminuye el sesgo asociado al PCR, ya que solo hacemos un PCR, mejorando así la representatividad de nuestros resultados. Además esta aproximación minimiza costos.

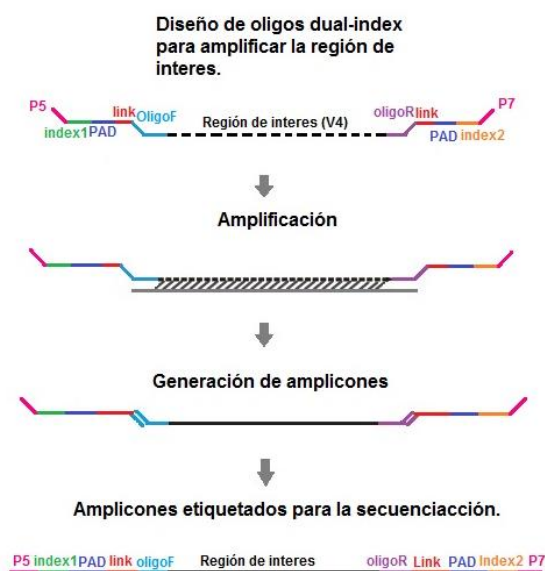


Figura 7. Flujo de trabajo para la construcción de los amplicones etiquetados.

2.6.1 Diseño de los oligos.

Los oligos para amplificar la región V4 del gen 16s RNA fueron los propuestos por Klindworth *et al.* (2013):

Oligo Forward = 5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG

Oligo Reverse = 5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

Después se realizó la construcción de los oligos compuestos utilizando la propuesta de Kozich *et al.* (2013) que consiste en añadir al oligo específico las secuencias del adaptador de Illumina

(AATGATACGGCGACCACCGAGATCTACAC), el índice (difiere por muestra), el oligo de secuenciación (ACTATCTGTA) y el linker (GT) (Tabla 1 y 2).

Tabla 1. Oligos *forward* usados para la generación de amplicones. Adaptador (rojo), índice (azul), PAD (verde), link (amarillo), oligo específico (negro).

ID Oligo	Secuencia (5'-3')
K501	AATGATACGGCGACCACCGAGATCTACACACTATCTGACTATCTGTATGGTGCCAGCMGCCGCGGTAA
K502	AATGATACGGCGACCACCGAGATCTACAC TAGCGAGTTATGGTAATTGGTGCCAGCMGCCGCGGTAA
K503	AATGATACGGCGACCACCGAGATCTACACCTAGTATGTATGGTAATTGGTGCCAGCMGCCGCGGTAA
K504	AATGATACGGCGACCACCGAGATCTACACGATAGCGTTATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT501	AATGATACGGCGACCACCGAGATCTACAC TAGTGTAGTATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT502	AATGATACGGCGACCACCGAGATCTACACGTCTAGTGTATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT503	AATGATACGGCGACCACCGAGATCTACACACGCGTGATATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT504	AATGATACGGCGACCACCGAGATCTACACAAGCAGCATATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT505	AATGATACGGCGACCACCGAGATCTACACCGATCTACTATGGTAATTGGTGCCAGCMGCCGCGGTAA
KT506	AATGATACGGCGACCACCGAGATCTACACTGCGTCACTATGGTAATTGGTGCCAGCMGCCGCGGTAA

Tabla 2. Oligos *Reverse* usados para la generación de amplicones. Adaptador (rojo), índice (azul), PAD (verde), link (amarillo), oligo específico (negro).

ID Oligo	Secuencia (5'-3')
K701	CAAGCAGAAGACGGCATACGAGATAACTCTCGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
K702	CAAGCAGAAGACGGCATACGAGATGCTGTAAACAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
K703	CAAGCAGAAGACGGCATACGAGATGGACGTTAAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
K704	CAAGCAGAAGACGGCATACGAGATGATCGTAGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
KT701	CAAGCAGAAGACGGCATACGAGATCACGATAGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
KT702	CAAGCAGAAGACGGCATACGAGATCGTATCGCAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
KT703	CAAGCAGAAGACGGCATACGAGATCTGCGACTAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
KT704	CAAGCAGAAGACGGCATACGAGATACGTACGTAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT
KT705	CAAGCAGAAGACGGCATACGAGATATATGCCGAGTCAGTCAGCCGGACTACHVGGGTWTCTAAT

2.6.2 Amplificación

Se seleccionó una de las réplicas por cada muestra para la secuenciación. La selección se basó en la concentración del DNA, las réplicas seleccionadas se muestran en el Anexo 1.

La reacción de PCR y el programa para la generación de los amplicones están en los Anexos 2 y 3.

2.7 Electroforesis

Los productos de PCR se corrieron en un gel para confirmar que la amplificación fue exitosa. Para ello se utilizaron geles de agarosa al 1.2 %, usando 3 μL de producto PCR y 2 μL de buffer de carga. Cada gel se corrió a 80 voltios durante 40 min. Posteriormente fueron observados y fotografiados mediante el uso de un fotodocumentador.

2.8 Purificación y normalización del producto de PCR.

La purificación y normalización de los productos PCR se realizaron mediante el *SequalPrep Normalization Plate kit (Invitrogen)*, siguiendo las instrucciones del fabricante, que son las siguientes:

- I. Se transfirieron 18 μL de cada producto de PCR a un pocillo de la placa de normalización.
- II. Se agregaron 18 μL de buffer de unión a cada pocillo, se mezcló e incubó por 1 hora a temperatura ambiente, esto para permitir la unión del DNA a la placa.
- III. Se eliminó la solución de cada uno de los pocillos con una pipeta sin rasparlos.
- IV. Se agregaron 50 μL de buffer de lavado y mezcló por pipeteo dos veces. Después se eliminó completamente el buffer de los pocillos.
- V. Se agregaron 20 μL de buffer de elución, mezcló por pipeteo cinco veces e incubó a temperatura ambiente durante 5 min
- VI. El *pool* fue hecho tomando 5 μL de DNA purificado de cada pocillo y se mezclaron en un tubo nuevo.

2.9 Cuantificación de las bibliotecas de secuenciación.

La cuantificación del *pool* se realizó mediante un fluorómetro Qubit[®] (*Thermo Fisher Scientific*) y el kit labeling detection (*Thermo Fisher Scientific*) siguiendo el protocolo proporcionado mediante el flujo de trabajo mostrado en la Figura 8.

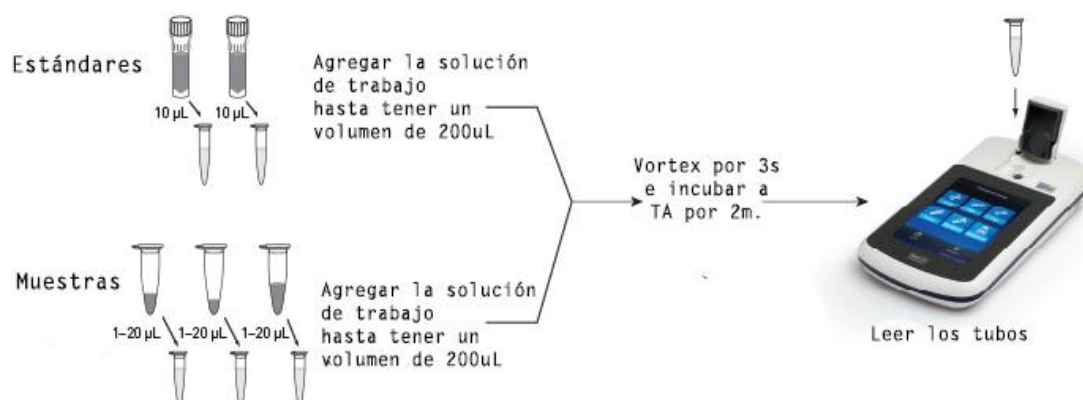


Figura 8. Flujo de trabajo para la cuantificación de las bibliotecas (imagen modificada del protocolo proporcionado por el kit).

El método para la cuantificación de las bibliotecas fue el siguiente:

- I. Se preparó la solución de trabajo a una relación 1:200 del buffer y el fluoróforo proporcionado por el kit.
- II. Se prepararon los estándares.
Con el fin de calibrar el fluorómetro se emplearon dos estándares (C y D) utilizando 10 µL de la solución estándar y 190 µL de la solución de trabajo.
- III. Se prepararon las muestras. Tomando 5 µL de muestra y 195 µL de la solución de trabajo.
- IV. Se les dio vórtex por 3 segundos y se incubaron a temperatura ambiente por 2 min.
- V. Se leyeron en el fluorómetro.

Después de tener las concentraciones de nuestras bibliotecas se calculó la concentración de DNA en nM, basándonos en el tamaño de nuestros amplicones y utilizando la siguiente fórmula:

$$\text{Concentración en nM} = \frac{\text{concentración en ng/}\mu\text{L}}{660\text{g/mol} \times \text{el tamaño de la biblioteca}} \times 10^6$$

Donde:

Concentración en ng/ μ L = concentración de DNA del *pool*: 0.372

Tamaño de la biblioteca = 254 pb

Concentración en nM= 2000 pM

2.10 Desnaturalización y dilución de bibliotecas y PhiX.

EL PhiX es una biblioteca control hecha con el genoma del virus PhiX, que debido a sus características (es pequeño y diverso) permite mitigar los cambios en las bibliotecas desbalanceadas y poco diversas (Illumina, 2016).

Para la desnaturalización y dilución se utilizó el kit v2 de illumina y se siguieron las instrucciones del fabricante.

2.10.1 Desnaturalización y dilución de las bibliotecas.

- I. Se preparó una dilución fresca de NaOH al 0.2 N.
- II. En un tubo se mezclaron 10 μ L del pool a 2000 pM y 10 μ L de NaOH al 0.2 N = a 20 μ L de pool desnaturalizado, se le dio vortex brevemente, se centrifugó a 280 xg por 1 minuto y se incubó a temperatura ambiente por 5 min.
- III. Se agregó 1980 μ L de HT1 (buffer de hibridación) a la solución desnaturalizada para tener nuestras bibliotecas desnaturalizadas a 20 pM.
- IV. Se mantuvo en hielo hasta la última dilución.
- V. Se realizó la última dilución de las bibliotecas para llevarlas a 8 pM.
- VI. 800 μ L de bibliotecas a 10 pM + 200 μ L de HT1= 1000 μ L a 8pM.

2.10.2 Desnaturalización y dilución del PhiX.

- I. En un tubo se mezclaron 2 μ L de PhiX a 10 nM con 3 μ L de H₂O de grado molecular para tener 5 μ L de PhiX a 4 nM.

- II. Se mezclaron 5 μ L de PhiX a 4 nM con 5 μ L de NaOH 0.2N para obtener una solución de PhiX desnaturalizada a 2 nM, se le dio vortex y se incubó 5 min a temperatura ambiente.
- III. Se agregaron 990 μ L del buffer HT1 para tener una biblioteca PhiX desnaturalizada a 20 pM.
- IV. Posteriormente se realizó la última dilución para llevar a 8 pM: 240 μ L de bibliotecas a 10 pM + 360 μ L de HT1= 600 μ L a 8 pM.

2.10.3 Combinación de la biblioteca de amplicones con la biblioteca control (PhiX).

En la secuenciación usó la biblioteca PhiX al 5 % con base en las recomendaciones del fabricante y también a lo reportado por Kozich *et al.* (2013).

Solución de trabajo final:

950 μ L de la biblioteca metagenómica (8 pM) + 50 μ L de biblioteca control (8 pM).

2.11 Secuenciación.

La secuenciación de las librerías se realizó usando la plataforma *Miseq de Illumina* recientemente adquirida por el Centro de Investigación Científica y Educación Superior de Ensenada (CICESE) como parte del proyecto CIGoM, siguiendo las instrucciones del fabricante.

2.11.1 Llenado de la “Sample sheet”

Se creó la “Sample sheet”, una hoja de cálculo donde se especifican los ID de las muestras que serán secuenciadas, así como sus índices para la identificación de las mismas.

2.11.2 Preparación del cartucho de reactivos.

El cartucho se descongeló colocándolo en un recipiente con agua, cuidando que ésta no pasara de la línea señalada.

Puesto que los amplicones que componen las bibliotecas de secuenciación poseen etiquetas diferentes a las que reconoce el kit de secuenciación Illumina fue necesario colocar los oligos de

secuenciación que corresponden a las etiquetas usadas en nuestros amplicones; es por eso que se realizó el procedimiento que se menciona a continuación.

- I. Después de descongelado el cartucho se secó perfectamente y se abrieron los pocillos 12, 13, 14 (que contienen los oligos) y 17 (lugar donde se coloca la muestra) con una punta de micropipeta de 1000 μ L.
- II. Las soluciones contenidas en los pocillos 12, 13 y 14 fueron retiradas para ser colocadas en tubos nuevos de 1.5 mL.
- III. Se le agregaron 5 μ L del oligo de secuenciación 1 (read1) al tubo que corresponde al pocillo 12 asegurándose de mezclar bien. Se repitió el proceso con el *index primer* y la solución del pocillo 13 y con el oligo de secuenciación 2 (Read2) y la solución del pocillo 14.
- IV. Se mezcló bien por pipeteo y un vortex de 30 segundos. Se regresaron las soluciones a su pocillo correspondientes.

2.11.3 Preparación de la “*flow cell*” y el reactivo PR2.

Se sacaron del almacenamiento a 4 °C la flow cell y el reactivo PR2. La flow cell fue enjuagada con agua milli –Q y secada con toallitas *KimWipes*, posteriormente se limpió con una toallita y alcohol al 100%.

2.11.4 Corrida del secuenciador.

- I. Se encendió el equipo y se fueron siguiendo las instrucciones para cargar la *flow cell*, el PR2 y el cartucho de reactivos.
- II. Se agregó la “*Sample sheet*”.
- III. Se inició la corrida.

2.12 Análisis de calidad de la secuenciación.

Después de la secuenciación se revisaron los índices de calidad que arroja el equipo para saber si la corrida fue exitosa. Los principales índices que se revisan son: número de bases secuenciadas, la densidad de los “clusters” (en miles por mm²) detectada por análisis de imagen, el porcentaje de “clusters” que pasaron el filtro de calidad, el porcentaje de la muestra que es alineada con el genoma PhiX, la tasa de error es calculada por el aumento del control PhiX, el rendimiento de la corrida, el porcentaje de alineamiento y el QScore.

2.13 Análisis de calidad de los datos.

Para el análisis de calidad de las secuencias se utilizó FastQC como herramienta. FastQC realiza un análisis completo evaluando la calidad en una serie de módulos: calidad por base, por contenido de GC, por contenido N de bases, calidad de las secuencias, calidad por longitud de las secuencias, por niveles de duplicación, por secuencias sobre representadas, y por el contenido de adaptadores.

Con base en los resultados se determinó la calidad de las lecturas

2.14 Análisis en QIIME.

Los datos obtenidos de la secuenciación fueron analizados usando el paquete de software QIIME (Quantitative Insights Into Microbial Ecology). Este software es un “pipeline” que utiliza gran número de herramientas para el análisis metagenómico mediante el uso de un “script” en lenguaje de programación Python.

Para iniciar los análisis fueron necesarios los archivos provenientes de la secuenciación en formato FASTA, el “mapping file” y la base de datos con la información taxonómica (Greengenes).

2.14.1 Mapping File.

Antes de iniciar con el análisis de las secuencias en QIIME se construyó el “mapping file”. Es un texto delimitado por tabuladores que describe cada una de las muestras que fueron incluidas en la secuenciación, incluye el ID de la muestra, el index de secuenciación y la secuencia de primers (Anexo 4).

2.14.2 “Picking” de OTUS y las secuencias representativas.

Las secuencias fueron agrupadas en Operational Taxonomic Units (OTUs) con base en su similitud. Esto se realiza con uclust, BLAST y otras herramientas.

Para realizar el “Picking” de OTUS existen tres protocolos basados en *scripts* maestros que realizan todas las etapas de procesamiento hasta construir la tabla de OTUS (producirán un archivo de asignación OTU (pick_otus.py), el “representative set of sequences” (archivo FASTA de pick_rep_set.py), el alineamiento de secuencias (archivo FASTA de align_seqs.py), La asignación taxonómica (assign_taxonomy.py) , un alineamiento de secuencias filtrado (filter_alignment.py), la construcción de un árbol filogenético (archivo de Newick make_phylogeny.py) y la tabla de OTUS (make_otu_table.py):

pick_de_novo_otus.py: Eficiente en estudios donde hay una mala caracterización de las comunidades o en su defecto no existe. En este protocolo los “reads” son agrupados sin una base de datos con secuencias de referencia.

pick_closed_reference_otus.py: Es implementado en estudios enfocados en taxones que están previamente bien caracterizados y no se quiere analizar nuevos. En este protocolo los “reads” son agrupados usando una base de datos con secuencias de referencia, los “reads” que no son agrupados son excluidos del análisis.

pick_open_reference_otus.py: Eficaz en estudios con cuyo objetivo es el análisis de taxones bien caracterizados y además se busca descubrir nuevos. En este protocolo las lecturas son agrupadas usando una base de datos con secuencias de referencia y aquellas que no presentan “hits” son agrupadas siguiendo el protocolo *de novo*.

El análisis de nuestros datos siguió el protocolo pick_closed_reference_otus.py, ya que nuestro objetivo es más bien exploratorio y descriptivo más que enfocado a la búsqueda de nuevos taxones.

El *Picking*” de OTUS se realiza mediante el siguiente *script*:

```
pick_open_reference_otus.py -r -i -o
```

Dónde:

-r Dirección del el archivo de las secuencias de referencia

-i Dirección de las secuencias a analizar.

-o Nombre del directorio de salida que tendrá el resultado después de que el *script* sea ejecutado.

2.14.3 Análisis de alfa diversidad

Para el análisis de alfa diversidad se utilizó el *script*

```
alpha_rarefaction.py -i -o -t -m -e 100
```

Dónde:

- i Dirección de la tabla de OTUs
- o Nombre del directorio de salida que tendrá el resultado después de que el *script* sea ejecutado.
- t dirección de árbol filogenético
- m Dirección del *mapping file*
- e Profundidad máxima de rarefacción.

Con los datos dados al *script*, éste realiza ordenadamente lo siguiente: La generación de las tablas de OTU rarificadas, el cálculo de los índices de alfa diversidad (Shannon y Chao1) de cada una de las tablas, la recopilación de los resultados y por último genera los gráficos de alfa rarefacción, las especies observadas y el índice de Chao1.

2.14.4 Análisis de beta diversidad.

El cálculo de la beta diversidad se realizará con el *script*.

```
jackknifed_beta_diversity.py -i -o -e 100 -m -t
```

Dónde:

- i Dirección de la tabla de OTUs
- o Nombre del directorio de salida que tendrá el resultado después de que el *script* sea ejecutado.
- t dirección de árbol filogenético
- m Dirección del *mapping file*
- e Profundidad máxima de rarefacción

El *script* jackknifing (re-muestreo repetido de un subconjunto de los datos disponibles de cada muestra) realiza el análisis de beta diversidad que arroja una matriz (*plots de* Análisis de coordenadas principales), mediante el uso del análisis de *clusters* UPGMA con el objeto de con cuantificar la distancia entre cada una las muestras.

2.14 Visualización de la diversidad y abundancia de las muestras.

Con el fin de visualizar, analizar, comparar y contrastar la información obtenida en las tablas de OTUs, se realizaron gráficos de porcentajes en Excel y en la herramienta de visualización Venny 2.1.0 (Oliveros, 2015).

Capítulo 3. Resultados

3.1 Muestreo

Se obtuvieron muestras de 10 estaciones diferentes (Tabla 3), 5 réplicas por estación de aproximadamente 10 mL cada una, que fueron divididas en 2 partes: 0-5 cm (B) y 5-10 cm (A). Además, se colectaron 2 muestras extras (con 5 réplicas por muestra) en las estaciones H47 y G44, de la parte del fondo del núcleo, aproximadamente a 30 cm de profundidad.

3.2 Extracción, cuantificación, concentración y electroforesis del DNA genómico total de muestras de sedimentos.

3.2.1 Extracción

Se realizaron 112 extracciones de DNA genómico, 50 correspondientes a las etiquetadas como “A”, 50 etiquetadas como “B”, 10 etiquetadas como “F” y 2 de las “FF”. El DNA se re suspendió en un volumen final de 50 μ L por muestra. Es importante mencionar que para el caso de las muestras de fondo se realizaron algunas modificaciones al protocolo proporcionado por el fabricante, con el fin de obtener un mejor rendimiento, como quedó descrito en la metodología.

3.2.2 Cuantificación y análisis de pureza del DNA extraído.

Las 100 extracciones provenientes de los sedimentos superficiales (0-10 cm) tuvieron que concentrarse ya que no fue posible cuantificarlas usando el *nanodrop* (Thermo Scientific). Después de concentrar el DNA se obtuvo un volumen final de 20 μ L. Posteriormente fueron cuantificadas (Tabla 3).

En promedio las muestras con mayores concentraciones de DNA genómico fueron las G44-B y las que presentaron menores concentraciones fueron las pertenecientes a G44-F. Individualmente la muestra A1-5B presentó la mayor concentración (26.5 ng/ μ L) y G44-1F la menor (2.3 ng/ μ L) (ver tabla 3). Las muestras etiquetadas como “FF” también fueron cuantificadas (Tabla 4). Se compararon los resultados de la extracciones de los sedimentos de fondo (F y FF) se observó que el rendimiento de la extracción incrementa considerablemente realizando la extracción complementando el *kit PowerSoil*[™] de Mobio con fenol-cloroformo-alcohol isoamílico y etanol al 100%.

Para evaluar la pureza del DNA se utilizó como indicador la relación de absorbancias a las longitudes de onda de 260nm y 280nm (relación A260/A280). Las muestras presentaron valores de A260/A280 en un rango de 1 a 1.77; 1.47 fue el valor promedio (Tabla 3).

Tabla 3. Concentraciones y pureza del DNA genómico extraído de las muestras de sedimentos.

ID Muestra	Concentración (ng/μl)	A260/A280	ID Muestra	Concentración (ng/μl)	A260/A280
C22-1A	6.8	1.35	A7-1B	7.1	1.13
C22-2A	8.1	1.52	A7-2B	13	1.3
C22-3A	10.2	1.3	A7-3B	7.5	1.31
C22-4A	6.9	1.13	A7-4B	8	1.4
C22-5A	6.3	1.25	A7-5B	8.1	1.32
C22-1B	10.5	1.55	E31-1A	10.4	1.53
C22-2B	20.3	1.51	E31-2A	10	1.47
C22-3B	10.6	1.34	E31-3A	7	1.45
C22-4B	9.4	1.47	E31-4A	31.3	1.23
C22-5B	7.9	1.34	E31-5A	12.9	1.53
B14-1A	10.3	1.29	E31-1B	13.6	1.56
B14-2A	7.3	1.32	E31-2B	18	1.53
B14-3A	6.5	1.08	E31-3B	16	1.59
B14-4A	8.2	1.38	E31-4B	13.1	1.54
B14-5A	10.8	1.02	E31-5B	13.7	1.53
B14-1B	16.5	1.37	B18-1A	13.8	1.56
B14-2B	7.4	1.26	B18-2A	17.3	1.52
B14-3B	7.2	1.12	B18-3A	16.4	1.44
B14-4B	8.5	1.25	B18-4A	12.2	1.55
B14-5B	9	1.32	B18-5A	10.1	1.24
A7-1A	7.6	1.28	B18-1B	10.9	1.36
A7-2A	6.4	1.05	B18-2B	13.9	1.43
A7-3A	6.3	1.02	B18-3B	14.1	1.39
A7-4A	6.5	1.04	B18-4B	13.9	1.42
A7-5A	6.7	1.24	B18-5B	12.3	1.42
H47-1A	10.3	1.2	A5-1B	13.7	1.47
H47-2A	10.2	1.25	A5-2B	12	1.43
H47-3A	10.6	1.31	A5-3B	11.9	1.38
H47-4A	9.7	1.35	A5-4B	11.1	1.41
H47-5A	23.7	1.25	A5-5B	13.1	1.31
A1-1A	16.3	1.4	G44-1B	22.4	1.76
A1-2A	16.7	1.49	G44-2B	22.5	1.67
A1-3A	35.7	1.29	G44-3B	23.1	1.7

A1-4A	15.2	1.46	G44-4B	19.6	1.74
A1-5A	14	1.44	G44-5B	59.9	1.54
A1-1B	14.5	1.45	H45-1A	14.7	1.69
A1-2B	16	1.42	H45-2A	16.7	1.52
A1-3B	18.7	1.5	H45-3A	19.5	1.53
A1-4B	14.6	1.54	H45-4A	16.3	1.64
A1-5B	26.2	1.47	H45-5A	13.2	1.73
A5-1A	12.5	1.45	H45-1B	13.7	1.73
A5-2A	11.6	1.46	H45-2B	14.3	1.72
A5-3A	11.7	1.46	H45-3B	17.1	1.74
A5-4A	12.6	1.46	H45-4B	16.9	1.62
A5-5A	12.6	1.46	H45-5B	18.8	1.66
G44-1F	2.3	1.00	H45-1F	3.3	1.63
G44-2F	3.3	1.26	H45-2F	2.8	1.47
G44-3F	4.4	1.75	H45-3F	3.1	1.12
G44-4F	3.6	1.16	H45-4F	2.6	1.21
G44-5F	4.3	1.23	H45-5F	3.5	1.23
H47-1B	11.2	1.4	G44-1A	21.3	1.57
H47-2B	13.1	1.33	G44-2A	16.6	1.52
H47-3B	11.7	1.23	G44-3A	25.3	1.61
H47-4B	17.8	1.3	G44-4A	21.9	1.53
H47-5B	12.6	1.39	G44-5A	18.9	1.67

Tabla 4. Concentración y pureza de las muestras G44-1FF y H45-1FF extraídas utilizando el protocolo modificado de MoBio.

Muestras	Concentración	A260/280
G44-1FF	39.3	1.74
H45-1FF	16.9	1.77

3.2.3 Electroforesis de los productos de extracción de las muestras de sedimentos de fondo del núcleo.

Puesto que las muestras de sedimentos del fondo del núcleo (F) fueron conservadas a -20 C° durante la campaña y almacenadas posteriormente a 80 C° sin buffer de sacarosa, se realizó una electroforesis para verificar la integridad del DNA obtenido (Figura 9). Lo mismo se hizo con las muestras que se extrajeron utilizando el protocolo modificado (FF) (Figura 10).

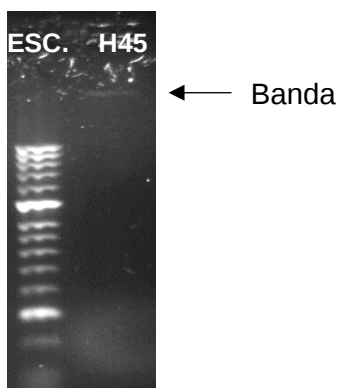


Figura 9. Extracciones de DNA de sedimentos de fondo del núcleo, muestra H45. Gel de agarrosa (1.5%) en TBE. Volumen cargado 7 μ L de DNA. Marcador de peso molecular de 2000 a 50 pb.

Al observar el gel se detectó una banda tenue pero definida, lo que resultó aceptable dada las bajas concentraciones de DNA obtenidas en la extracción (3.3 ng/ μ L) y la cantidad de DNA que se usó al momento de cargar el gel (7 μ L de la elución final). En el caso del DNA procedente de las muestras FF, se observó también bandas tenues pero además se detectó un ligero barrido en todas ellas, lo que podría indicar algo de degradación del DNA (Figura 10).

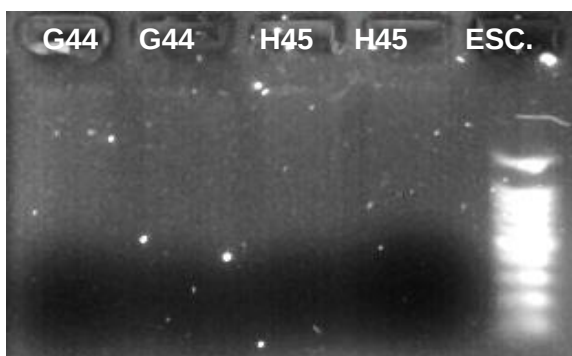


Figura 10. Electroforesis de extracciones de sedimentos con fenol cloroformo. Gel de agarosa al 1.2 %. Volumen de carga 10 μ L de DNA. Marcador de peso molecular de 2000 a 50 pb.

3.3 Construcción de las bibliotecas de secuenciación.

3.3.1 Amplificación de la región variable V4 del gen 16s rRNA.

Se seleccionó una de las réplicas de cada muestra para la construcción de las bibliotecas de secuenciación, basándonos en la concentración del DNA. A cada réplica seleccionada se le asignó un ID numérico, y un par de índices para su posterior identificación (Tabla 5).

Tabla 5. Selección de las réplicas a amplificar e índices utilizados para identificar las muestras.

MUESTRA	CONCENTRACIÓN (ng.µL)	INDICE F	INDICE R
C22-2A	29,2	ACTATCTG	AACTCTCG
C22-1B	34,7	ACTATCTG	GCTGTAAC
B14-4A	10,4	ACTATCTG	GGACGTTA
B14-1B	6,9	ACTATCTG	GGTCGTAG
A7-1A	7,6	ACTATCTG	TAAGTCTC
A7-5B	7,6	ACTATCTG	TACACAGT
E31-5A	11,9	TAGCGAGT	AACTCTCG
B18-1A	51,6	TAGCGAGT	GGACGTTA
A1-2A	6,6	CTAGTATG	AACTCTCG
A1-4B	8	CTAGTATG	GCTGTAAC
G44-5A	12,4	CTAGTATG	TAAGTCTC
H45-5A	13,6	GATAGCGT	AACTCTCG
H45-3B	9,4	GATAGCGT	GCTGTAAC
E31-3B	4,4	TAGCGAGT	GCTGTAAC
B18-2B	5,4	TAGCGAGT	GGTCGTAG
H47-4A	4,9	TAGCGAGT	TAAGTCTC
H47-5B	4,4	TAGCGAGT	TACACAGT
A5-5A	4,6	CTAGTATG	GGACGTTA
A5-5B	5	CTAGTATG	GGTCGTAG
G44-3B	7,2	CTAGTATG	TACACAGT
H45-5F	3,7	GATAGCGT	GGACGTTA
G44-3F	4,6	GATAGCGT	GGTCGTAG
H45-5FF	39.3	GATAGCGT	TAAGTCTC
G44-3FF	16.9	GATAGCGT	TACACAGT

El éxito de la amplificación se verificó mediante un gel de electroforesis al 1.6 %, confirmándose que la amplificación fue exitosa, ya que las bandas obtenidas presentaron el tamaño esperado (382pb) (Figura 11).

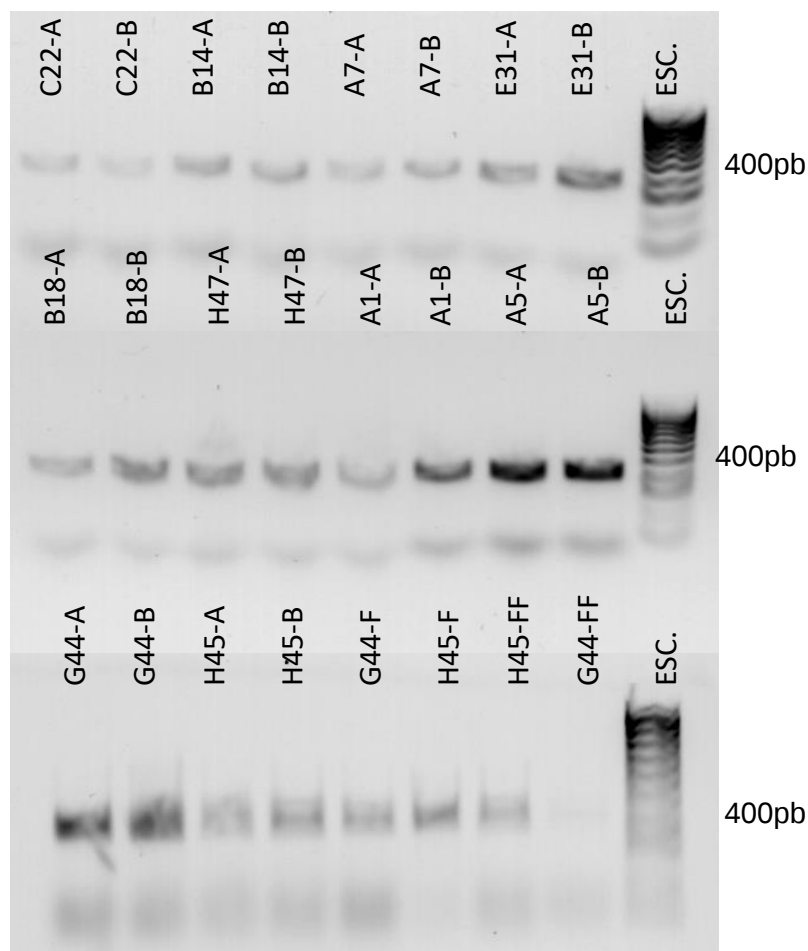


Figura 11. Electroforesis de los productos de PCR en gel de agarosa 1.6 %. Se obtuvieron los fragmentos del tamaño esperado (382pb) con base al marcador de peso molecular (2000 a 50 pb, *Sigma*).

3.3.2 Normalización y cuantificación de las bibliotecas.

La concentración del *pool* fue de 0.372 ng/μL. Todas las bibliotecas (incluyendo el PhiX) se desnaturalizaron y se diluyeron a la concentración de secuenciación a 8 pM.

3.4 Secuenciación.

3.4.1 Análisis del reporte de calidad de la corrida.

Al finalizar la corrida los reportes del secuenciador incluyen un informe que indica la calidad de la secuenciación (Tabla 6).

Tabla 6. Datos del reporte de la secuenciación. En la tabla se muestran los datos obtenidos de secuenciación paired * porcentaje de clusters que pasaron el filtro.

	Rendimiento	Densidad de los <i>clusters</i> (K/mm ²)	(%) Clusters PF *	Porcentaje Alineamiento	Porcentaje de Tasa de error	%≥Q30
Read 1	153.0 Mbp	762 ±1	89.71 ±0.04	6.39	0.35	97.3
Read 2	153.0 Mbp	762 ±1	89.71 ±0.04	6.29	0.35	92.47
Totales	320.3Mbp	-	-	6.34	0.35	94.59

De los resultados de calidad mostrados en la tabla, los más importantes son: el rendimiento de la corrida (*Yield*), el % de alineamiento (*Aligned %*) y el Qscore.

El rendimiento de la secuenciación fue de 320.33 Mbp, el porcentaje de alineamiento fue de 6.34 % (indica que la muestra se alineó bien con el PhiX), la tasa de error fue de 0.35 % (lo que indica que no hubo anomalías con el PhiX) y por último el Qscore que fue mayor a Q30 en el 94.59% de los casos, es decir la probabilidad de que nuestras secuencias presenten errores en sus bases es menor a 0.1%. También se analizó el porcentaje de *clusters* que pasaron los filtros de calidad, que fue cercano al 90%, lo que se traduce en un buen rendimiento de secuenciación.

Al observar los resultados de los parámetros de calidad concluimos que la secuenciación fue exitosa. De un total de 988,024 lecturas 870,980 pasaron los filtros de calidad, de las cuales el 75.7 % fueron asignadas a algún índice (748,029). En número de secuencias (de buena calidad y asignadas a algún índice) por muestras fue variable. La mayoría estuvo entre el rango de 30,000 y 40,000. El número máximo fue de 43,357 (B18.1A) y el mínimo de 22,886 (H45.5FF) (Tabla 7).

Tabla 7. Lecturas obtenidas y OTUs identificados por muestra.

Muestra	Lecturas	OTUs obtenidos
H45-5FF	22886	10054
H45-5F	36634	17812
H45-5A	30234	19763
H45-3B	30455	19502
G44-3FF	27993	14935
G44-3F	25776	12141
G44-3A	36690	25685
G44-3B	23038	15869
C22-2A	32594	20059
C22-1B	26629	18527
A7-1A	29922	20956
A7-5B	30412	21152
A1-2A	31926	20900
A1-4B	33464	23042
H47-4A	31766	22674
H47-5B	32869	23783
B18-1A	43357	31228
B18-2B	31112	21156
A5-5A	29007	19330
A5-1B	29043	18686
E31-5A	37081	24389
E31-3B	23778	15802
B14-4A	39664	28318
B14-1B	31699	22164
Total	748029	487927

3.4.2 Análisis de calidad de las lecturas

Se obtuvieron 48 archivos FASTAQ de la secuenciación, dos por muestra (Read1 y Read 2) lo que nos asegura que cada una de las muestras presentes en la biblioteca fue secuenciada. Se realizó un análisis de calidad con el programa FASTAQC (Figura 12).

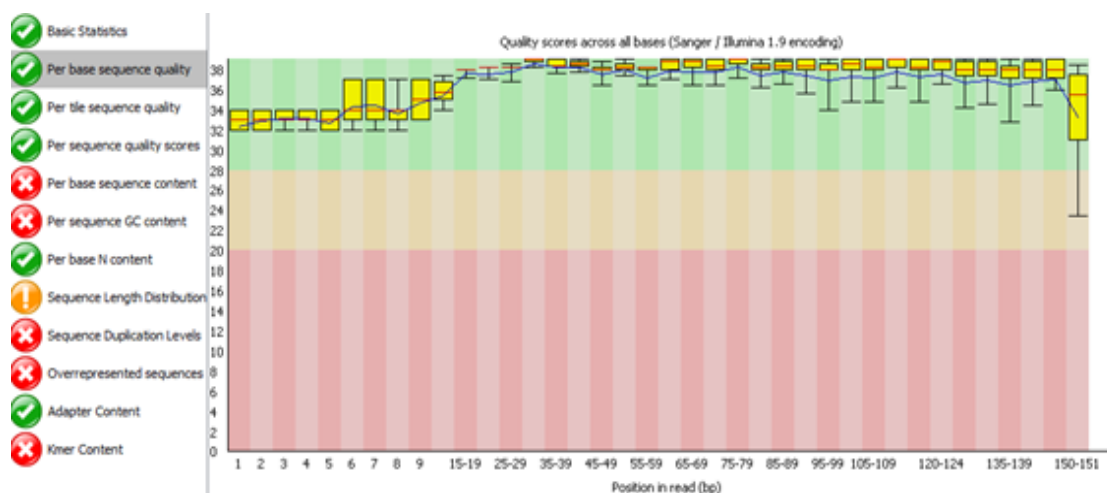


Figura 12. Resultado de los análisis de calidad en FastQC. En la Figura se muestra el módulo Per base sequence quality, se observa que todas nuestras secuencias son mayores a 30; lo que indica que la probabilidad de error en las bases es menor al 0.1 %. Se analizaron todas las secuencias y todas mostraron resultados similares.

El análisis de calidad de las secuencias por módulo, con base en la clasificación que arroja FastaQC fue:

1. *Basic statistics*: normal
2. *Per base sequence quality*: normal
3. *Per tile sequence quality*: normal
4. *Per sequence quality score*: normal
5. *Per base sequence content*: anormal
6. *Per sequence GC content*: inusual
7. *Per base N content*: normal
8. *Sequence length distribution*: inusual
9. *Sequence duplication levels*: anormal
10. *Overrepresented sequence*: anormal
11. *Adapter content*: normal
12. *Kimer content*: anormal

Después de analizar los índices de calidad se concluyó que nuestras secuencias son de buena calidad.

3.5 Análisis de los datos

3.5.1 Obtención de los OTUs

A partir de las 748,029 secuencias de buena calidad se obtuvieron en total 487,927 OTUs. La obtención de OTUs por muestra estuvo en un rango de entre 10,000 y 30,000. El valor máximo registrado fue de 31,228 OTUs (B18.1A), el mínimo de 10,054 (H45.5FF) y una media de 20,330 (Tabla 7).

La detección de OTUs está relacionada con la profundidad de muestro, es por eso que se realizó un análisis de correlación entre el número de lecturas y el número de OTUs obtenidos por muestra, el resultado fue un valor de 0.88, lo que nos indica que hay una correlación positiva entre ellos, es decir al aumentar la profundidad de secuenciación, la detección de OTUs también aumenta. Con el fin de hacer más visual la correlación se realizó un gráfico de dispersión (Figura 13).

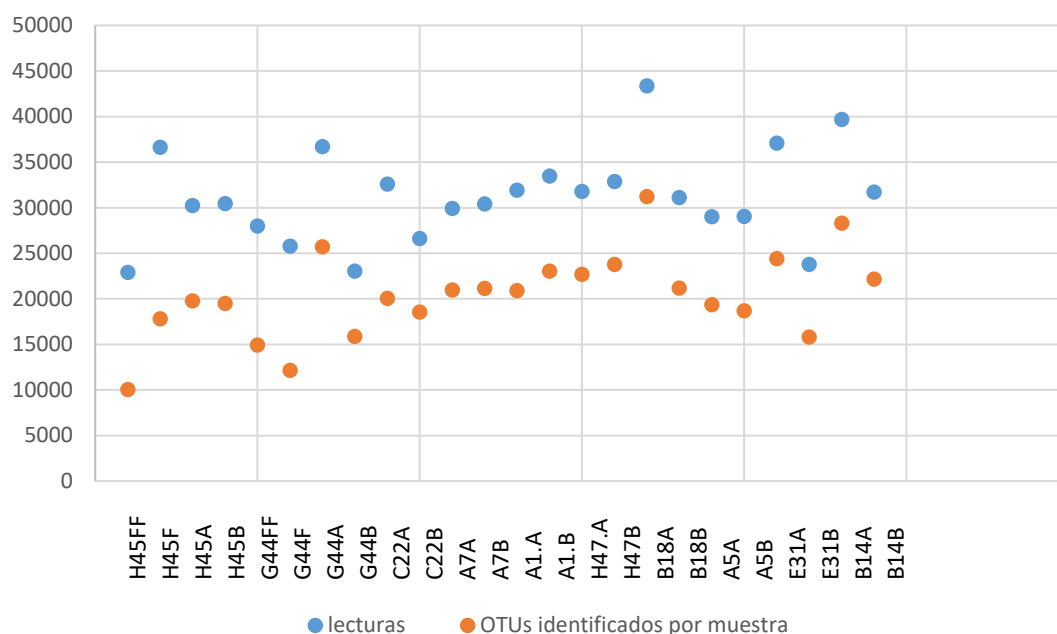


Figura 13. Relación entre los OTUs identificados y el número de lecturas.

En la Figura 13 se confirma la correlación positiva de la profundidad de secuenciación con la identificación de los OTUs (a excepción de H45-F), lo que indica que si aumentamos nuestra profundidad de secuenciación podríamos detectar mayor número de OTUs.

3.5.2 Curva de rarefacción

Ya que el número de secuencias en cada muestra afecta directamente la estimación de la diversidad de especies (menos secuencias = menos capacidad de detección) se realizó una curva de rarefacción, que nos ayuda a determinar si nuestra secuenciación tuvo la suficiente profundidad como para poder detectar a la mayoría de especies posibles dentro de la muestra (Figura 14).

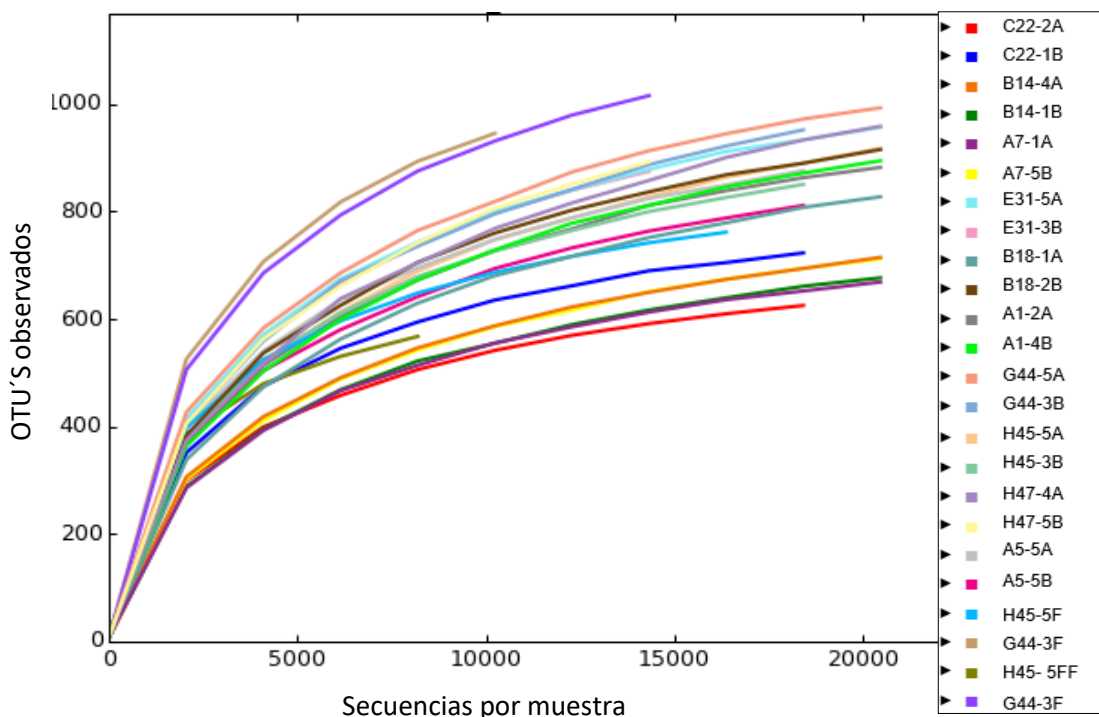


Figura 14. Curvas de rarefacción. Se observa que las curvas están cercanas a la asintota, lo que indica que la profundidad de secuenciación fue buena.

Las curvas de rarefacción que se obtuvieron no llegaron a la asintota, sin embargo se puede observar que la mayoría de las curvas presentan una tendencia de estar llegando a la asintota, es por eso que pensamos que nuestra profundidad de secuenciación es bastante aceptable.

3.5.3 Análisis de Alfa diversidad.

Los valores del índice de *Shannon–Weaver* obtenidos estuvieron entre 6.5 y 8. Los valores más altos fueron los de obtenidos en los sedimentos de fondo. Individualmente, la muestra G44-FF fue la que presentó el valor mayor (8.1) y las muestras con el valor menor fueron B14-By A7-A (6.5) (Figura 15). En general las muestras de fondo tuvieron valores más altos; esto indica que las comunidades de fondo son más diversas que las de la superficie.

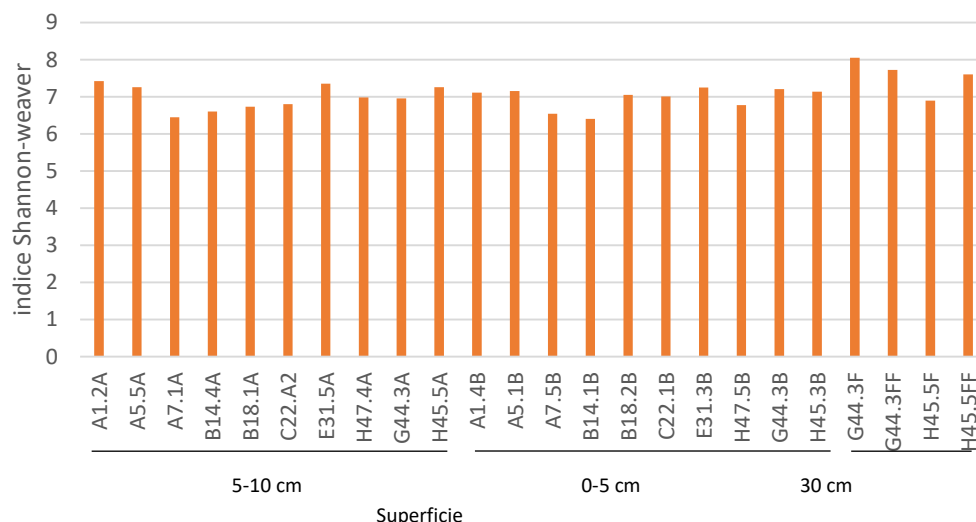


Figura 15. Análisis de alfa diversidad con el índice de Shannon- Weaver. Los índices de diversidad entre las muestras de superficie (A y B) del mismo sitio de muestreo son similares, a diferencia de los sitios de fondo cuyos valores de diversidad son mayores, a excepción de H45-FF.

En el análisis de estimación de la riqueza de especies mediante el índice Chao1 a una profundidad de secuenciación de 8,194 lecturas (para normalizar) por muestra (ver Figura 15) se obtuvo que el rango del número de OTUs estimados de la biblioteca es de 655 a 1114. Además se observó que la riqueza de especies es similar entre las muestras A y B, independientemente del sitio de muestreo. También se realizó una comparación de los OTUs estimados (Chao1) respecto a los OTUs observados (Figura 16), donde se visualiza que hay diferencias entre ellos. Esto podría significar falta de profundidad en la secuenciación.

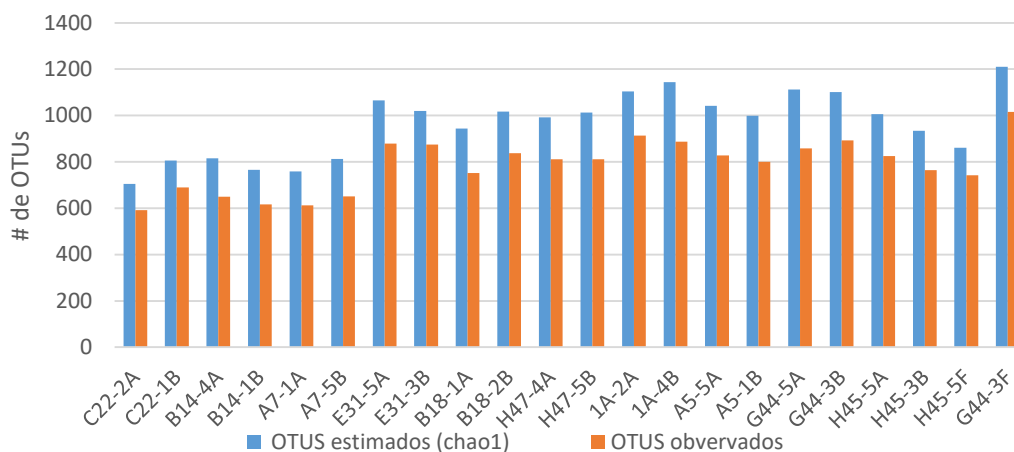


Figura 16. Estimación y observación de riqueza de especies. Se utilizó el índice Chao1 para la estimación de la riqueza de OTUs. Posteriormente estos datos se compararon con los datos de riqueza obtenidos.

Estimación y observación de riqueza de especies. Se utilizó el índice Chao1 para la estimación de la riqueza de OTUs. Posteriormente estos datos se compararon con los datos de riqueza obtenidos.

3.5.4 Asignación taxonómica de los microbiomas.

Se obtuvieron 5 tablas de identificación taxonómica de los OTUs a diferentes niveles taxonómicos (L2: Phylum, L3: Clase, L4: Orden, L5: Familia, L6: Género) (Tabla 13). En total se logró clasificar a 72 grupos del dominio arquea: 3 fila, 12 clases, 14 órdenes, 12 familias y 31 géneros; y 687 grupos del dominio bacteria: 56 fila, 134 clases, 171 órdenes, 173 familias y 153 géneros.

Tabla 8. Número de OTUs identificados por nivel taxonómico. La tabla además incluye el número de grupos que no se lograron identificar en el siguiente nivel de clasificación.

	Phyla	Clases	Orden	Familias	Géneros
Procariotas total	60	175	298	421	531
Arqueas	3	12	14	12	31
Bacterias	56	134	171	173	153
Grupos sin identificar (arquea)	0	0	6	15	23
Grupos sin identificar (bacteria)	1	29	107	221	347
Grupos sin identificar total	1	29	113	236	3703

De los grupos identificados, se detectó que el 93.62 % pertenecen a los fila *Crenarchaeota*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Chloroflexi*, NC10, *Gemmatimonadetes*, SBR1093 y *Bacteroidetes*.

3.5.5 Análisis de estructura de las comunidades microbianas de las muestras.

Se observó la composición y abundancia de los grupos de microorganismos presentes en las muestras de sedimento a nivel filo y clase (Figuras 16, 17). En las gráficas se puede apreciar como las muestras superficiales (A y B) son relativamente homogéneas en las 10 estaciones, sin embargo, al

comparar con las muestras de fondo (F) se aprecia como varían tanto en su composición como en su abundancia.

De las diferencias a nivel filo destacan que las abundancias de los grupos *Crenarqueota*, *Proteobacteria*, *SBR1093* y *Acidobacteria* representa al alrededor del 95% de los OTUs en superficie y en el fondo disminuyen al 40% (en promedio). Al contrario las abundancias de los grupos *Chloroflexi*, *Planctomycetes*, *Actinobacteria*, *OP3*, *Euriarqueota*, *GN04*, y *Espiroquetas* presentan mayor abundancia en el fondo, representan alrededor del 50% de los OTUs y en superficie son alrededor del 15%.

También fue posible detectar que existen grupos exclusivos en ambas zonas. En sedimentos superficiales se identificaron los grupos ZB3 y Anck, incluidos en la categoría <1%, y en sedimentos de fondo, OP8 y otros grupos que están dentro de la categoría "<1%". Otro aspecto importante a mencionar es que las abundancias de los grupos de bacterias y arqueas cambian considerablemente entre los sedimentos de fondo de núcleo. Fila como *Crenarqueota*, *Chloroflexi*, *NC10*, *Gemmatimonadetes*, *Planctomycetes* y grupos con abundancias menores al 1%, como *OP9*, *Fibrobacteres*, *LCP-89*, *LD1*, *Hyd24-12*, *Anck6* son más abundantes en un sitio que en el otro (Figura 17).

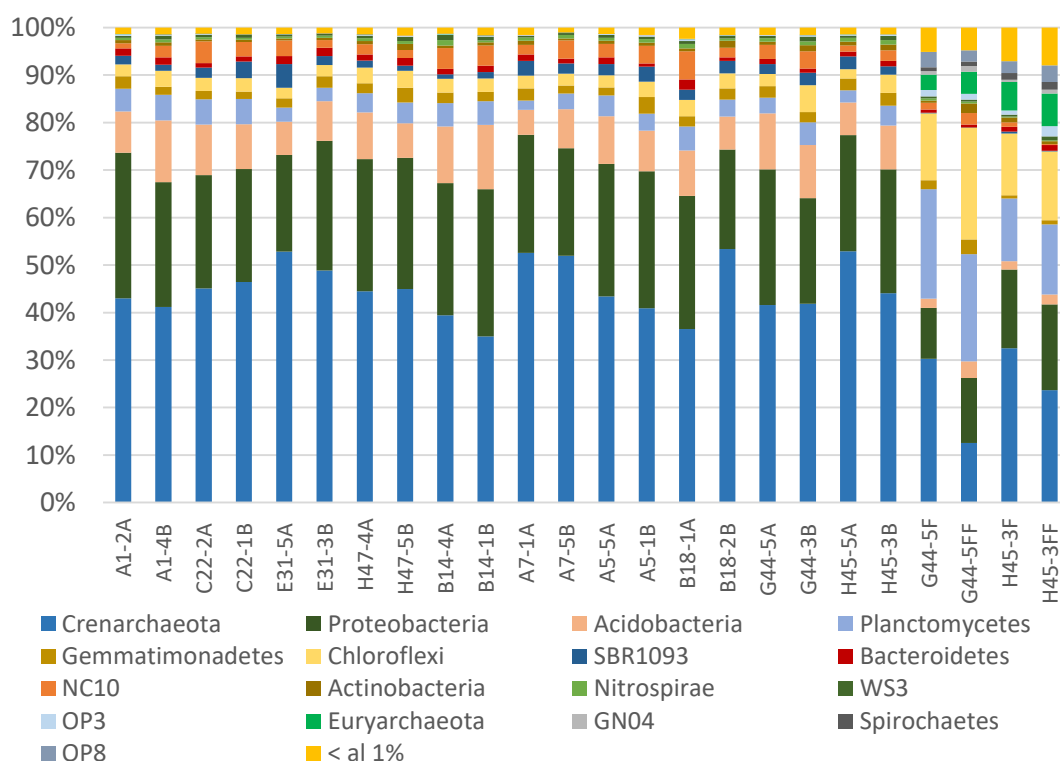


Figura 17. Composición de la diversidad y abundancia de los OTUs observados entre las muestras "B" (0-5 cm), "A" (5-10 cm) y F (más de 30 cm) a nivel Filo. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada "< al 1%".

En la figura 17 se observa que la estructura de la comunidad entre los sedimentos superficiales y de fondo es diferente a nivel filo, también hay diferencias entre las comunidades de fondo. En el recuadro rojo, se señalan los grupos cuyas abundancias disminuyen en los sedimentos del fondo y en azul se marcan los grupos que aumentan su abundancia en el fondo.

Al igual que en la figura 17 (composición de la comunidad a nivel filo), en la figura 18 se puede apreciar como las comunidades a nivel clase, son muy similares en los sedimentos superficiales, aunque a este nivel taxonómico se detectan pequeñas diferencias, pero con valores de abundancia bajos. Las diferencias se observan al comparar las muestras de fondo con las de superficie, y entre las muestras de fondo. También se observó que las clases: *MBGB*, *Dehalococcoidetes*, *OP8_1*, *BD4-9* y *MCG* son exclusivas de las muestras de fondo (F y FF) (Figura 18).

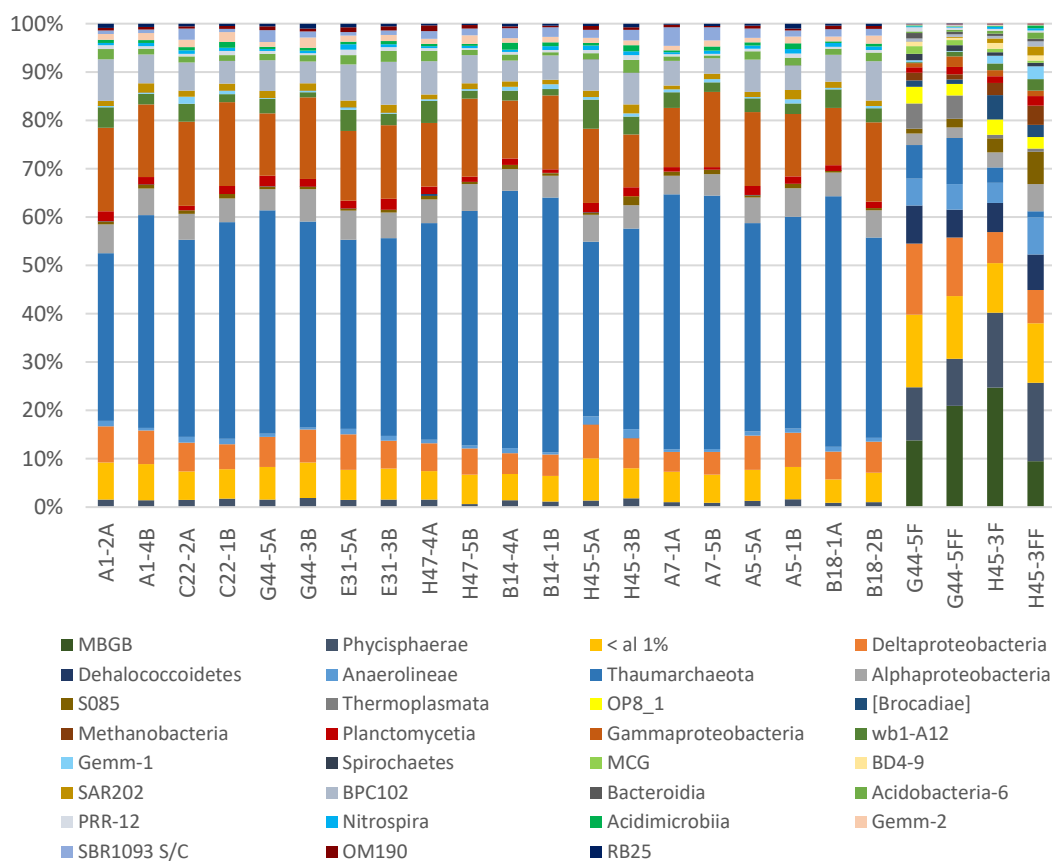


Figura 18. Diversidad y abundancia de los OTUs observados entre las muestras "B" (0-5 cm), "A" (5-10 cm) y F (más de 30 cm) a nivel clase. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada "< al 1%".

Respecto a las diferencias de abundancias de los grupos presentes en los sedimentos superficiales y de fondo, se observó que *Thaumarchaeota*, *Gammaproteobacteria*, *BPC102*, *Deltaproteobacteria*, el grupo etiquetado como "< al 1%", *Alphaproteobacteria*, *wb1-A12*,

Acidobacteria-6 y *Planctomycetia* representan alrededor del 90% de OTUs identificados en superficie, aunque estos grupos también se detectaron en los sedimentos de fondo, pero son menos abundantes. En cambio, en los sedimentos de fondo las clases *MBGB*, *Phycisphaerae*, el grupo etiquetado como "<al 1%", *Deltaproteobacteria*, *Dehalococcoidetes*, *Anaerolineae*, *Thaumarchaeota*, *Alphaproteobacteria*, *S085*, *Thermoplasmata*, *OP8_1*, *[Brocadiae]*, *Methanobacteria*, *Planctomycetia*, *Gammaproteobacteria*, *wb1-A12* y *Gemm-1* presentan alrededor del 90% de los OTUs identificados.

3.5.6. Comparación de la diversidad y abundancia de los grupos taxonómicos más abundantes entre muestras de fondo y superficie.

Con el fin de profundizar en el análisis de la composición de los microbiomas entre las estaciones, se compararon solo las estaciones G44 y H45, debido a que son las únicas donde se obtuvieron muestras de superficie y fondo y dado que los sedimentos superficiales son muy similares entre ellos.

Proteobacteria

La clase *Proteobacteria* fue analizada a nivel orden, en el análisis se observaron diferencias claras entre los sedimentos de superficie y fondo. En general se identificó un mayor número de órdenes en las muestras superficiales (ver Figura 19).

Se encontró que hay grupos exclusivos en las dos zonas del sedimento; *Thionalorhabdales* y los grupos con abundancias menores al 1%: *HOC36*, *PB19*, *Pasteurellales*, *Methylococcales*, *Aeromonadales*, son específicos de los sedimentos superficiales; en el caso de los sedimentos de fondo, se identificó el orden *Desulfarculales*, *DTB120* y algunos grupos en la categoría <1% (*IndB3-24*, *Neisseriales*).

Se identificó que las abundancias de algunos grupos varían dependiendo de la zona en el sedimento. Por ejemplo, *NBi-j*, *Cromatiales*, órdenes sin clasificar de *Gammaproteobacteria*, *Marinicenales*, y *Alteromonadales* son más abundantes en superficie que en fondo y los grupos *Desulfobacterales*, *Desulforcuales*, grupos sin identificar de la clase *Deltaproteobacteria* y *Sva-0853* son más abundantes en fondo que en superficie.

Además, se observaron diferencias de abundancia y composición entre las comunidades de las dos muestras de fondo. Los órdenes *Alteromonadales*, *Oceanospirillales* y algunos grupos incluidos en <1% (*Campylobacterales*, *IndB3-24*, *Mariprofundales*, *Neisseriales* y *Sva0485*) se encontraron en G44 y

no en H45. En cambio *DTB120*, *Nitrosomonadales* y *Sphingomonadales* (de los grupos con abundancia menor al 1%) fueron identificados en H45 y no en G44.

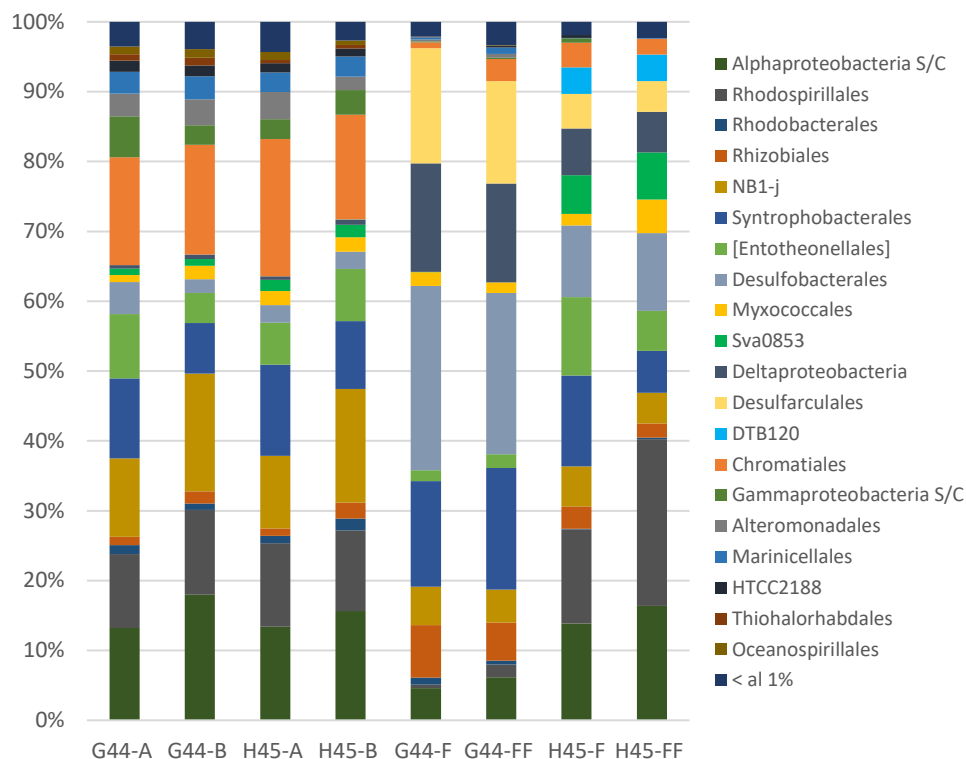


Figura 19. Diversidad y abundancia de *Proteobacterias* a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada “< al 1%”. En el gráfico se puede apreciar como varían las comunidades al comparar superficie y fondo. También se observan diferencias entre las muestras de Fondo (G44 y H45 F, FF).

Cloroflexi

Este grupo fue analizado a nivel orden. Igual que en los análisis anteriores no se encontraron diferencias de diversidad entre las muestras superficiales del sedimento, pero si en abundancia. En superficie se observó que los órdenes SB-34 y SAR202 S/C representan alrededor del 80% de los OTUs identificados, mientras que en fondo la abundancia de los mismos grupos es menor. En los sedimentos del fondo, GCA004 y *Dehalococcoidales* representan entre el 50 y 60 % del total de OTUs pertenecientes *Cloroflexi* en las muestras del fondo, mientras que superficie son grupos que representan alrededor del 10% (Figura 20).

También se detectaron grupos exclusivos en las muestras de superficie y fondo. En sedimentos superficiales estos órdenes tienen abundancias menores al 1%, así que fueron englobados en el grupo “<

al 1%” órdenes como: *MSB-1E9*, *OPB11* y *GIF9*; y órdenes sin identificar de las clases *Dehalococcoidetes* y *Anaerolineles* son propias de los sedimentos F y FF.

Después de analizar los datos de las muestras del fondo, se identificó que entre ellas también existen diferencias, de la abundancia principalmente, pero también referente a la diversidad de las comunidades. Se encontró que el orden GIF9 es exclusivo de los sedimentos de la estación G44, y un orden sin identificar perteneciente a la clase TK17 fue detectado en H45F.

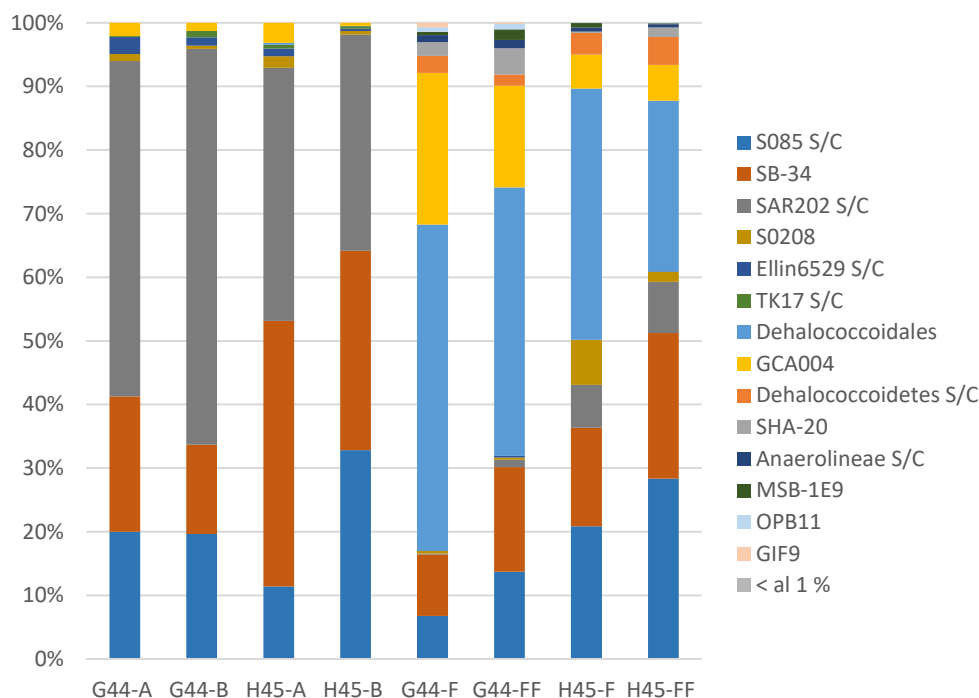


Figura 20. Diversidad y abundancia de *Cloroflexi* a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada “< al 1%”.

Acidobacteria

En el análisis de la diversidad y abundancia de *Acidobacteria* a nivel orden, no se detectó diferencia en la composición de las comunidades, pero si en las abundancias de los grupos presentes en los sedimentos superficiales. Sin embargo, se observó que las comunidades de los sedimentos de superficie y fondo son diferentes en diversidad y abundancia. Los grupos *Acidobacteria* S/C, DA052 y OS-k, presentaron mayores abundancias en los sedimentos F y FF; en cambio las abundancias de BPC102, RB25 y Sva0725 son mayores en los sedimentos superficiales.

Además, se encontró que hay diferencias en diversidad y abundancia entre las estaciones de fondo (como en los grupos bacterianos analizados anteriormente) (Figura 21). Se observó que EC1113 no está presente en G44F y FF, y el grupo RB25, es exclusivo de G44F y FF.

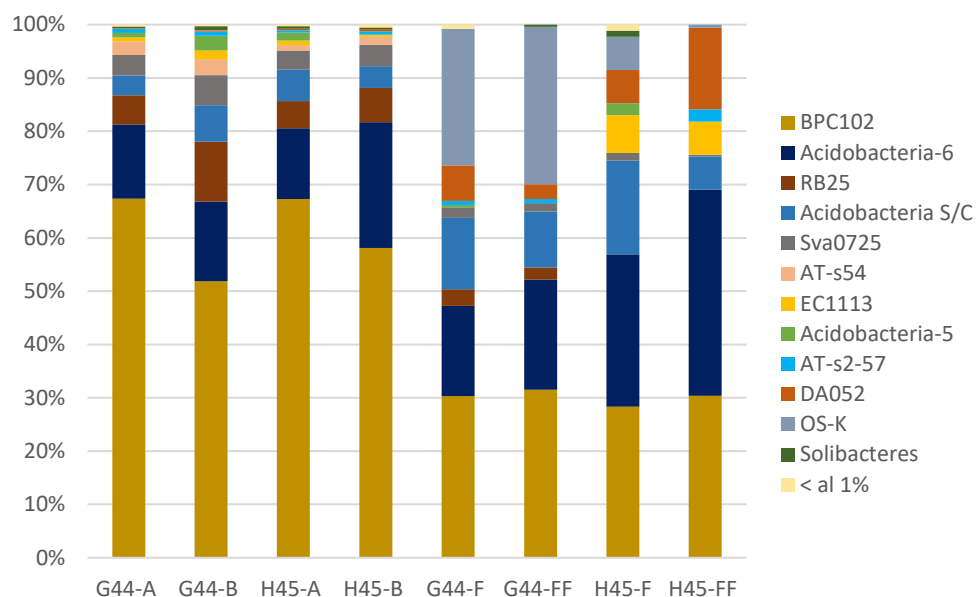


Figura 21. Diversidad y abundancia de *Acidobacteria* a nivel Orden en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada “< al 1%”.

Actinobacteria

El análisis de los OTUs de las muestras, a nivel familia en *actinobacteria* mostró que no hay diferencias en la composición de dichas comunidades, en los sedimentos superficiales entre las estaciones a excepción de H45, debido a la presencia del grupo OPB41 en esta estación. Además se observó que las comunidades de los sedimentos superficiales del núcleo son diferentes a las del fondo del núcleo, principalmente en abundancias, pero también en composición. El grupo Koll13 fue identificado solamente en G44 y H45 fondo (Figura 22).

En el caso de los sedimentos F y FF, existen diferencias entre estaciones, principalmente de abundancias. Pero también se observó que la familia At425_EubF1 es exclusiva de H45 F y FF.

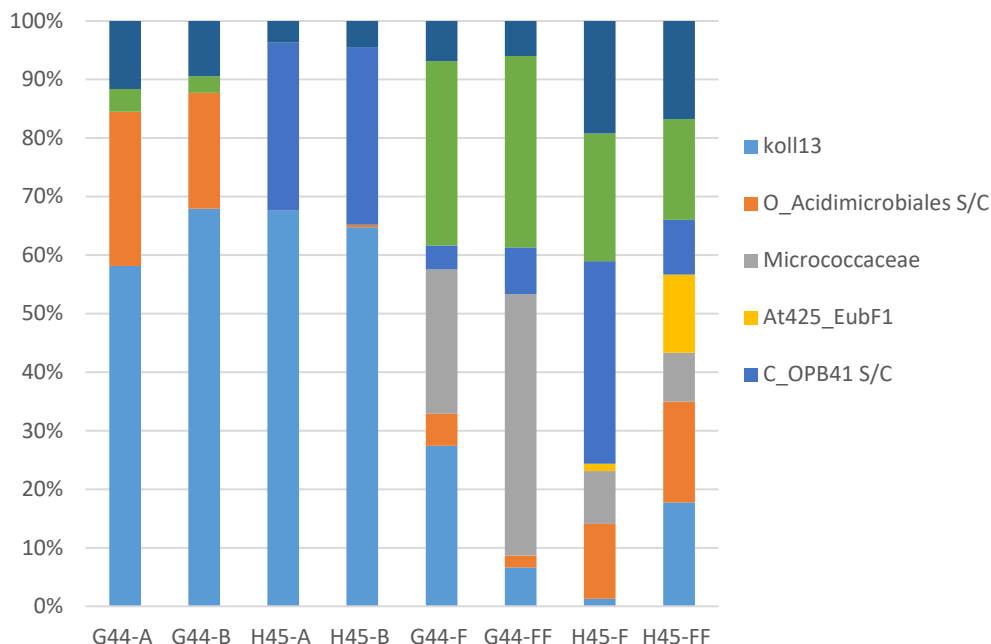


Figura 22. Diversidad y abundancia de Actinobacteria a nivel Orden en las muestras "B" (0-5 cm), "A" (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45. El gráfico muestra los grupos con una abundancia mayor al 1%, el resto se agrupó en una única categoría llamada "< al 1%".

Arquea

Se analizaron los OTUs correspondientes al dominio Arquea a nivel clase. Se encontraron diferencias muy marcadas entre los sedimentos superficiales y de fondo (Figura 23). En la superficie de los sedimentos se observa que el grupo dominante es *Thaumarchaeota*, se detectaron otros grupos, tales como miembros de *Crenarchaeota* sin clasificar, MBGA, MBGB, *Methanobacteria*, *Methanomicrobia* y *Thermoplasmata*. Pero todos ellos con abundancias menores al 1%. En cambio en los sedimentos F se detectaron mayor número de grupos (los grupos exclusivos fueron: clases pertenecientes a *Crenarchaeota* sin clasificar, Parvarchaea, AAG, MCG y MHVG). También se observó que la abundancia de los grupos aumentó. Los más abundantes fueron MBGB, MCG, *Methanobacteria*, *Thaumarchaeota* y *Thermoplasmata*.

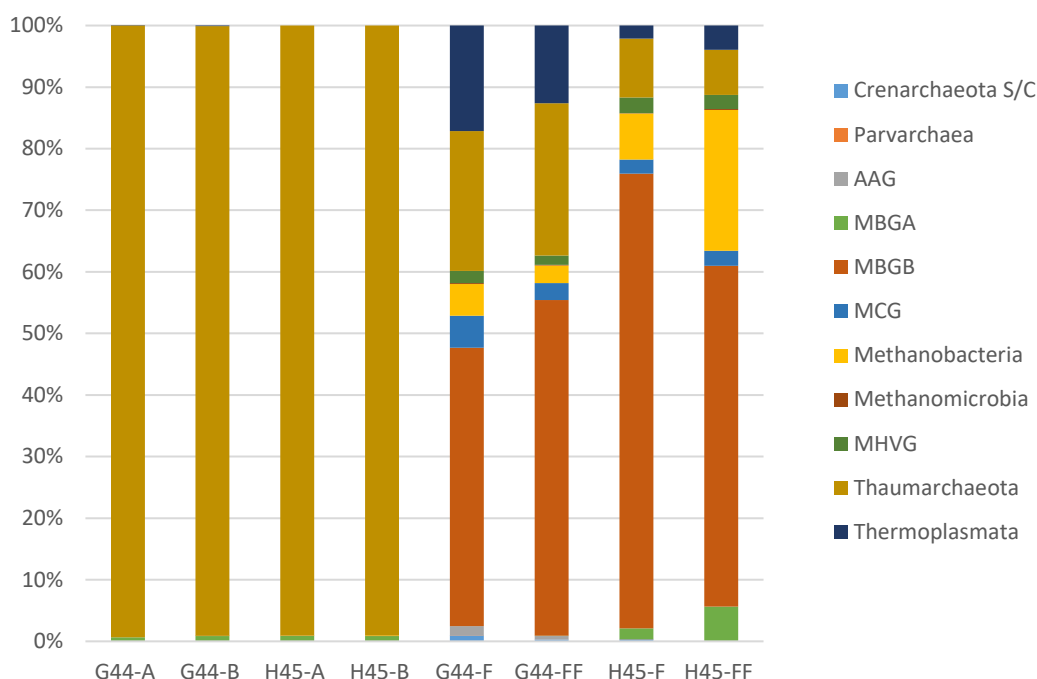


Figura 23. Diversidad y abundancia de *Arquea* a nivel clase en las muestras “B” (0-5 cm), “A” (5-10 cm), F y FF (más de 30 cm) de las estaciones G44 y H45.

3.5.7 Comparación de diversidad entre las estaciones G44 y H45 (superficie y fondo).

Se realizaron varios diagramas de Venn para identificar cuantos OTUs son compartidos y cuantos son detectados exclusivamente entre las muestras. Con este fin se sumaron los OTUs totales de G44 (A y B), G44 (F y FF), H45 (A y B) y H45 (F y FF) para aumentar el tamaño de muestra. Para minimizar las detecciones que puedan estar influenciadas por la profundidad de lectura solo se compararon aquellos de OTUS que presentaban al menos 50 lecturas por muestra (Figura 22).

Se encontraron 84 OTUs (64.1 %) comunes entre las muestras G44 y H45 de superficie (Figura 22A), también observamos que entre G44 y H45 de las muestras de fondo solo había un 24.6% de OTUs comunes (Figura 22B). En la figura 22C se muestra la comparación entre las cuatro muestras y en ella se observa que la similitud entre las muestras superficiales y las de fondo son muy pequeñas. Los OTUs comunes entre las 4 muestras fueron solo el 3.2% del total.

Cuando nos centramos en los OTUs detectados exclusivamente en las muestras, se identificaron 22 OTUs (16.8 %) solo presentes en G44, entre ellos destacan por su abundancia: miembros del orden *Chromatiales* y *NB1-j* (*Proteobacteria*), *Nitrosopumilus spp.* (filo *Crenarqueota*), miembros del orden *B110* (*Acidobacteria*) y miembros del orden *Phycisphaerales* (*Planctomycetes*). En H45 fueron detectados

exclusivamente 25 OTUs, entre los cuales se encuentran miembros del orden MSBL9 (*Planctomycetes*), las familias *Spirochaetaceae* (*Spirochaetes*), *SAGMEG-1 spp.* (*Euryarchaeota*), *Syntrophobacteraceae* (*Spirochaetes*) y *Piscirickettsiaceae* (*Proteobacteria*) (Figura 22a).

Respecto a los OTUs solo detectados en las muestras de fondo, en G44 se identificaron 65 OTUs exclusivos (37.1 %), siendo los más representativos miembros de la familia *Syntrophobacteraceae* (*Crenarqueota*), *Desulfococcus spp.* (*Proteobacteria*), miembros de la clase S085 (*Crenarqueota*), miembros de la familia *Cenarchaeaceae* (*Crenarqueota*) y miembros del orden p04_C01 (*Proteobacteria*). En el caso de H45 se detectaron 67 OTUs específicos (38.3%), por su por su abundancia: pMC1A4 spp. (*Euryarqueota*), miembros de la clase *Dehalococcoidetes* sin clasificar (*Cloroflexi*), miembros del orden MSBL9 (*Planctomycetes*), miembros de la clase *Alphaproteobacteria* sin clasificar (*Proteobacteria*) y *Desulfococcus spp.* (*Proteobacteria*) (Figura 22b).

Estos datos corroboran que las muestras de sedimentos superficiales son más similares entre ellas, que las muestras de fondo. Y que hay diferencias marcadas entre las muestras de superficie y fondo.

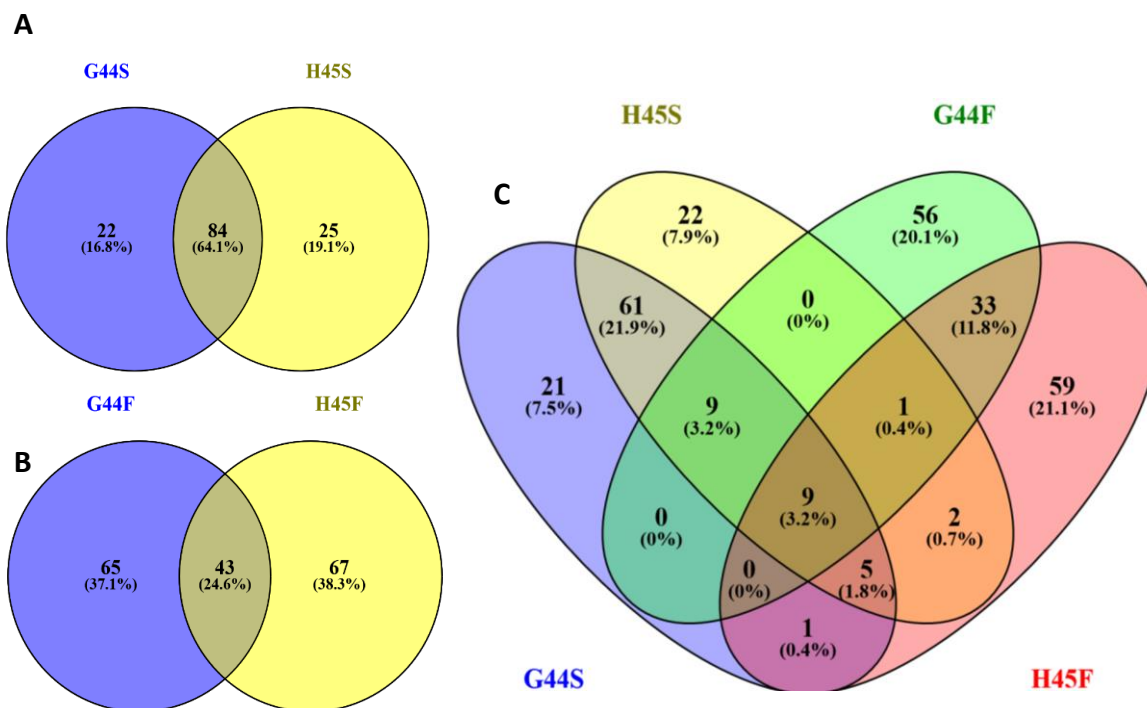


Figura 24. Diagramas de Venn, comparando las muestras G44 y H45 de superficie y fondo. A) Comparación entre las muestras superficiales de G44 y H45, B) comparación entre las muestras de fondo de G44 y H45. C) Comparación de las cuatro muestras juntas (G44 superficie, H45 superficie, G44 fondo y H45 fondo)

3.5.8 Análisis de Beta diversidad

En nuestros análisis se calculó la distancia entre 24 muestras (10 “A”, 10 “B”, 2F y 2FF), mediante un análisis de Coordenadas principales donde se visualiza como se forman 3 agrupaciones (Figura 25). La primera agrupación corresponde a todas las muestras A y B, en la segunda se agrupan las muestras G44 de fondo y la tercera agrupa a dos muestras H45 de fondo. Esto indica que las muestras de superficie (A y B) son muy similares entre ellas. Al contrario, las muestras de fondo (G44 y H45) difieren entre ellas. Además de la clara diferencia, entre las muestras de superficie y fondo.

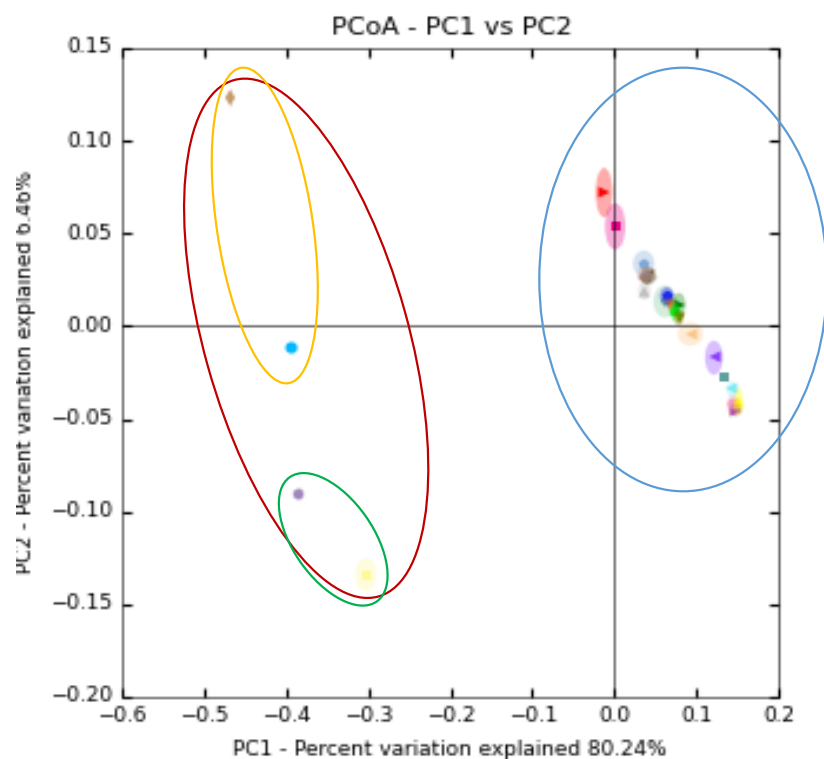


Figura 25. Análisis de coordenadas principales. Los tres gráficos muestran el plano desde Z, desde diferentes ángulos. Muestra 3 agrupaciones. En la primera (en azul) se agrupan las muestras de superficie (A y B). Las otras dos agrupaciones representan a las muestras de fondo del núcleo. En rojo observamos a H45 (F y FF) y G44 (F-FF) se muestra en verde. El 80.24% de la variación está explicada por PC1.

Capítulo 4. Discusión

Se analizaron los microbiomas de 24 muestras de sedimentos provenientes de 10 estaciones del Golfo de México. Las muestras se tomaron a 3 profundidades diferentes dentro del núcleo del sedimento: 0-5 cm, 5-10 cm y mayor a 30 cm. El análisis se realizó mediante la secuenciación de amplicones de la región V4 del gen 16s RNA usando la plataforma MiSeq® (Illumina, San Diego, CA).

Con base en los resultados obtenidos podemos afirmar que no se cumple la hipótesis 1: “Debido a que las condiciones ambientales en los sedimentos de los 10 sitios de muestreo analizados del sur del Golfo de México son disímiles, la composición de las comunidades microbianas presentes es diferente”. Puesto que observamos que los sedimentos superficiales (profundidad 0 a 10 cm dentro de la columna sedimentaria) son muy similares entre ellos independientemente del sitio de muestreo. Aceptamos la hipótesis 2: Con base a estudios previos (Orcut, 2013) planteamos que, “Como las características de los sedimentos cambian conforme aumenta la profundidad en la columna sedimentaria, las comunidades de bacterias y arqueas de los sedimentos más superficiales son diferentes respecto a las presentes en los sedimentos del fondo del núcleo”.

Además, se observó que los grupos de arqueas y bacterias difieren entre las estaciones a una profundidad mayor a 30 cm en la columna sedimentaria.

4.1 Consideraciones técnicas.

4.1.1 Extracción de DNA.

Se observó que al añadir fenol-cloroformo-alcohol isoamílico (25:24:1) a los reactivos de la lisis celular proporcionados por el kit PowerSoil® de MoBio con alcohol absoluto durante los lavados de DNA, el rendimiento de la extracción aumenta. Esto refuerza lo reportado en otros trabajos donde al realizar la lisis celular mediante el uso de disrupción física y el uso de fenol y/o cloroformo-alcohol isoamílico, se mejoraba el rendimiento de la extracción de DNA en sedimentos marinos (Miller *et al.* 1999). También Sagova-Mareckova *et al.* (2008) realizaron modificaciones al protocolo del kit PowerSoil® de MoBio, con el fin de mejorar el rendimiento de la extracción de DNA en suelo y sedimentos obteniendo resultados positivo. Esto indica que a pesar de que el kit está diseñado para su uso en suelos y sedimentos, aún se puede aumentar el rendimiento de la extracción con algunas modificaciones. De hecho, en la última versión publicada del protocolo se recomienda hacer alícuotas con la solución C5 y etanol al 100% en cantidades iguales, para utilizarla como solución de trabajo en el lavado del DNA. Es decir, el protocolo del kit se sigue modificando.

Sin embargo, hay que mencionar que con base en los resultados de secuenciación, podemos afirmar que aunque el complementar el kit con fenol-cloroformo-alcohol isoamílico (25:24:1) incrementa el rendimiento de la extracción, no así la detección de OTUs.

4.1.2 Secuenciación

Debido a que la secuenciación la realizamos nosotros, aquí en CICESE, en el secuenciador MiSeq (Illumina, San Diego, CA) recientemente adquirido, una parte importante de esta tesis ha sido la optimización de la secuenciación. Nuestras bibliotecas se prepararon a una concentración de 8 pM de DNA, se obtuvieron 988,024 lecturas con un porcentaje de alineamiento a PhiX de 6.34 %, la tasa de error fue de 0.35 % y el Qscore >Q30 en el 94.59% de los casos. La densidad de *clusters* fue de 762 ± 1 (K/mm²) y el porcentaje de *clusters* que pasaron los filtros de calidad fue cercano al 90%.

El análisis de estos parámetros nos indica que realizamos una “buena secuenciación”. El porcentaje de alineamiento y la tasa de error están relacionados con el porcentaje de PhiX utilizado, para el que no existe un valor estándar a utilizar pues depende del tipo de biblioteca que se desea secuenciar. Con el porcentaje de alineamiento se evalúa si la muestra se alinea con el genoma control, el valor obtenido debe de estar dentro de un rango cercano al porcentaje de PhiX utilizado, si el resultado del porcentaje de alineamiento es significativamente menor al PhiX utilizado, indica que la concentración de las bibliotecas fue subestimada, lo que llevaría a una sobrerrepresentación (Illumina, 2015). Nosotros usamos 5% de PhiX, y el porcentaje de alineamiento obtenido muy (6.3%) cercano a 5% lo que nos indica un alineamiento muy aceptable. Respecto a la densidad de *clusters*/mm², el valor óptimo reportado para el *Miseq* es entre 700-820, los valores obtenidos están dentro del rango. Por último, el Qscore es un valor de calidad que indica el porcentaje de que una de las bases en la secuencia sea errónea, el Qscore reportado para la plataforma Miseq es Q30 > 80% cada 150 pb (revisado en http://www.illumina.com/systems/hiseq_2500_1500/performance_specifications.html), como se mencionó anteriormente los resultados obtenidos fueron superiores a ese porcentaje.

Kozich *et al.* (2013), secuenciaron una biblioteca a 10 pM de DNA proveniente una comunidad microbiana, obtuvieron 130,411 lecturas con Qscore >Q30 en el 63.7 % de los casos, y dado que seguimos su propuesta de construcción de bibliotecas, podemos afirmar que nuestra secuenciación, tuvo menor número de errores que el reportado por ellos.

4.1.3 Análisis de la calidad de los datos obtenidos en la secuenciación.

El análisis de calidad de las secuencias con FASTAQC, clasificó nuestras lecturas como normales en los módulos: *Basic statistics*, *Per base sequence quality*, *Per tile sequence quality*, *Per sequence quality score*, *Per base N content*, *Adapter content*. Anormales en los módulos: *Per base sequence content*, *Sequence duplication levels*, *Overrepresented sequence*, *Kimer content*. E inusuales en: *Per sequence GC content* y *Sequence length distribution*.

Las clasificaciones normal, anormal e inusual en FastQC son lo indican la calidad de las secuencias. Pero se tiene que poner especial atención al contexto de la evaluación ya que algunas bibliotecas de secuenciación son diferentes a lo que FastQC considera una secuencia buena (normal), puesto que los lineamientos que FastQC emplea son específicos (Andrews S, 2010). Por ejemplo: para que las secuencias de una biblioteca sean clasificadas como “normales” deben ser diversas, es decir que no estén sesgada de ninguna manera.

Debido a que nuestra biblioteca, fue construida siguiendo la aproximación de secuenciación por amplicones, en consecuencia las lecturas obtenidas son del mismo fragmento, esto es detectado por el programa como fragmentos sobrerrepresentados. Es por eso que la anormalidad en los módulos “*Per base sequence content*” “*Per sequence GC content*” y “*Overrepresented sequence*” es esperada ya que son módulos que evalúan si hay secuencias sobre representadas en las bibliotecas. También en el módulo “*Sequence length distribution*” las lecturas se consideran anormales, ya que FASTAQC evalúa como normal a un grupo de secuencias con longitudes diferentes y como todas las lecturas tienen la misma longitud (150 pb), es un resultado esperado.

Por tanto, debido a las características particulares de nuestra biblioteca, el análisis de calidad nos indica que las secuencias que obtuvimos son de buena calidad.

4.1.4 Curvas de rarefacción

Se obtuvieron 487,927 OTUs totales a partir de las 748,029 lecturas de buena calidad, con un promedio de 20,330 OTUs por muestra, en contraste con los reportados por Kozich *et al.* (2013); estos autores reportaron 1,248.2 OTUs a partir de 130,411 lecturas. Aunque ellos partieron de un menor número de lecturas, nosotros, en proporción, obtuvimos mayor número de OTUs, lo que se traduce en mayor abundancia y diversidad de especies en nuestras muestras.

Se observó que el aumento de OTUs por muestra estaba correlacionado positivamente con el aumento en el número de lecturas. Es decir, si aumentamos nuestra profundidad de secuenciación aumentamos nuestras posibilidades de detección.

Con el fin de identificar si nuestro esfuerzo de secuenciación fue el adecuado para caracterizar con precisión las comunidades presentes, se realizaron las curvas de rarefacción para cada muestra. En promedio se obtuvieron 31,168 lecturas por muestra. Aunque, las curvas obtenidas no llegan a la asíntota, la mayoría se acercan bastante. Esto indica que nuestro esfuerzo de secuenciación no fue suficiente; sin embargo el estar cerca de la asíntota, y como la cantidad de OTUs identificados es alta, podemos decir que estamos cerca del valor máximo de detección. Es decir, que los datos de secuenciación obtenidos son suficientes para realizar la caracterización de la comunidad.

Nuestros datos son similares a los reportados por Zhang *et al.* (2015b); en su estudio de diversidad microbiana en sedimentos del mar de China, ellos mencionan que aunque la curva no presenta asíntota, se acerca, lo cual indica que las profundidades de muestreo fueron suficientes para representar la diversidad microbiana.

4.2 Caracterización taxonómica de los microbiomas.

Se identificaron taxonómicamente 3 fila, 12 clases, 14 órdenes, 12 familias y 31 géneros del dominio Arquea, además de 44 grupos que no pudieron ser asignados específicamente a un grupo taxonómico particular. Respecto al dominio Bacteria se identificaron 56 fila, 134 clases, 171 órdenes, 173 familias y 153 géneros, además de 705 grupos que no pudieron ser asignados específicamente a un grupo taxonómico.

En general el 93.62 % de las lecturas identificadas pertenecen a los fila *Crenarchaeota*, *Proteobacteria*, *Acidobacteria*, *Planctomycetes*, *Cloroflexi*, *NC10*, *Gemmatimonadetes*, *SBR1093* y *Bacteroidetes*. Esto coincide con lo reportado, ya que los fila Crenarqueota, Proteobacteria y Plactomicetes son de los grupos más abundantes a nivel mundial (Madigan, 2010).

A nivel clase, *Thaumarchaeota* (Crenarqueota), *Gammaproteobacteria*, *Deltaproteobacteria* y *Alphaproteobacteria* (Proteobacteria), BPC102 y *Acidobacteria-6* (Acidobacteria), wb1-A12 (NC10), SAR202 (Cloroflexi) y *Planctomycetia* (Plactomicetes) son los más abundantes en los sedimentos superficiales (representan alrededor del 90 % del total de lecturas). En los sedimentos de fondo las clases más abundantes son: *Thaumarchaeota*, MBGB y MCG (crenarqueota), *Phycisphaerae*, *Brocadiae* y *Planctomycetia* (Planctomycetes), *Deltaproteobacteria*, *Alphaproteobacteria* y *Gammaproteobacteria*

(*proteobacteria*), *Dehalococcoidetes*, S085 y *Anaerolineae* (*Chloroflexi*), OP8_1 (OP8), *Methanobacteria* y *Thermoplasmata* (*Euryarchaeota*), y *Gemm-1* (representan alrededor del 90% de las lecturas).

Algunos de los grupos detectados e identificados como abundantes en nuestras muestras han sido previamente reportados como característicos de algún tipo de sedimento. Parkes et al. (2014) y Orcutt et al. (2011) señalaron que *Gammaproteobacteria*, *Deltaproteobacteria*, *Alphaproteobacteria*, *Acidobacteria*, *Actinobacteria*, *Firmicutes*, *Bacteroidetes* y *Planctomycetes* son los grupos con mayores abundancias en sedimentos superficiales (primeros centímetros). Además, Parkes et al. (2014), mencionan a *Thaumarchaeota*, MBGB, miembros de *Chloroflexi* y *Deltaproteobacteria* como grupos abundantes en los primeros metros debajo de la superficie del sedimentos.

Al comparar los resultados obtenidos sobre los grupos más abundantes con estudios previos, observamos que algunos de los grupos identificados como los más abundantes en nuestras muestras, han sido reportados como característicos de sedimentos superficiales y profundos. Sin embargo, también detectamos grupos que no han sido reportados como característicos de sedimentos superficiales o de fondo. Basándonos en que la diversidad bacteriana está ligada al tipo de muestra (composición del sedimento) y profundidad dentro de la columna sedimentaria Orcutt et al. (2011) suponemos que hay grupos característicos de los sedimentos que analizamos, los cuales podrían ser OP8 y otros grupos como los candidatos BPC102, NC10.

4.2.1 Proteobacteria

La diversidad y abundancia relativa del filo *Proteobacteria* fue mayor en los sedimentos superficiales, coincidiendo con lo reportado anteriormente (Orcutt et al. 2011; Parkes et al. 2014). El filo *Proteobacteria* es muy diverso; presenta grupos con casi todos los tipos de metabolismos, autótrofos, heterótrofos, aerobios, anaerobios quimiolitótrofos, quimiorganótrofos, mixótrofos, etc. es por eso que el estudio de la actividad metabólica resulta complicada, incluso en niveles taxonómicos como orden o familia.

En este estudio se detectó que los órdenes *Pasteurellales*, HOC36, *Methylococcales* (del filo *Gammaproteobacteria*) y PB19 (*Deltaproteobacteria*) son exclusivos de los sedimentos superficiales, pero sus abundancias son muy bajas. En cambio en los sedimentos de fondo se identificó a *Desulfarculales* exclusivamente, con abundancias de entre el 10 y 5 %. También se observaron diferencias en las abundancias de los grupos de superficie y fondo. Grupos como *Chromatiales*, *Rhodospirillales* (de *Alfaproteobacteria*), *Alteromonadales*, *Marinicellales* y *Oceanospirillales* (de

Gammaproteobacteria) son mayores en superficie que en fondo. En fondo se observan mayormente *Desulfobacterales*, *Mixococales*, DTB120 y Sva0853 (*Deltaproteobacteria*).

Los resultados sobre los organismos indentificados en los sedimentos de fondo resultan convincentes ya que la mayoría de los grupos identificados son reductores de sulfato, tiosulfato o nitrato en condiciones anóxicas. Si nos basamos en la zonación biogeoquímica de los sedimentos propuesta por Orcutt (2011), quien menciona que a profundidades mayores a 10 cm en la columna sedimentaria (en sedimentos del talud aprox.). Respecto a los microorganismos de los sedimentos superficiales, solo podemos afirmar que son heterotróficos y que se ha reportado que miembros de *Alteromonadales* y *Oceanospirillales* son degradadores de hidrocarburos (Madigan, 2010).

4.2.2 Cloroflexi

El filo *Cloroflexi* fue más abundante y diverso en los sedimentos de más de 30 cm de profundidad. Se detectaron grupos exclusivos en los sedimentos superficiales, pero todos con abundancias menores al 1%. Este filo es comúnmente reportado en sedimentos marinos, especialmente el grupo *Sar202* (Stal y Cretoui, 2016), como en nuestras muestras. Sin embargo aún no hay mucha información sobre los miembros de éste filo en sedimentos marinos y se desconoce el papel que podrían tener (Stal y Cretoui, 2016).

4.2.3 Acidobacteria

El filo *Acidobacteria* fue más abundante en superficie que en fondo. Se cree que este filo, muy abundante en suelos y en sedimentos, cumple funciones bastante variadas e importantes en el medio ambiente puesto que tienen la capacidad de soportar ambientes contaminados y extremos (Wart et al. 2007). Además está el hecho de que es un filo metabólico y filogenéticamente diverso (aerobios, anaerobios facultativos, anaerobios, quimiorganótrofos, quimio heterótrofos, foto heterótrofos, reductores de nitrógeno, azufre, hierro y manganeso) (Madigan, 2010).

Aunque la mayoría de los grupos que se identificaron en nuestras muestras son candidatos y no hay suficiente información sobre ellos. Podemos suponer que tipo de metabolismo tienen, si sumamos la información disponible del filo con la información que tenemos del sedimento. Suponemos que en los primeros 10 cm de sedimento hay aerobios y/o aerobios facultativos debido a que hay disponibilidad de oxígeno en esta área principalmente por la presencia de bioturbadores (Cullen, 1973 y Marcelino-Barros,

2013). con base en lo reportado sobre la zonación bioquímica (Orcutt, 2011) pensamos que los microorganismos presentes son quimio heterótrofos, reductores de nitrógeno, acetato y /o azufre.

Respecto a los grupos presentes en fondo, suponemos que son microorganismos que están adaptados a bajas concentraciones de nutrientes, como los reportados por Barns *et al.* (2007), basándonos también en lo reportado sobre la zonación bioquímica (Orcutt, 2011) pensamos que los microorganismos presentes son reductores de hierro, azufre y manganeso.

4.2.4 Actinobacteria

El filo *Actinobacteria* fue más abundante en sedimentos de fondo. Se ha reportado anteriormente que *Actinobacteria* es más abundante en sub-suelo marino (Zhang *et al.* 2015b). Al igual que en el filo *Acidobacteria*, la mayoría de los grupos de *Actinobacteria* detectados son grupos candidatos, no están bien descritos hasta el momento y aún se desconoce mucho sobre ellos. Podemos suponer que son mixótrofos debido a que los miembros de este filo son quimiorganótrofos o mixótrofos (Madigan *et al.* 2010). Dadas las características del hábitat en el que se encuentran resulta más convincente que los miembros detectados sean mixótrofos, ya que no hay luz presente.

4.2.5 Arquea

El dominio arquea fue más abundante en los sedimentos superficiales que en los de fondo, sin embargo en el fondo se observó mayor diversidad. Observamos que el grupo dominante en los sedimentos superficiales es el filo *Thaumarqueota* (40% de los OTUs totales), lo cual podría ser consecuencia del sesgo relacionado a la amplificación, esto debido a que se cuenta con datos no publicados de bibliotecas *shotgun* donde su abundancia es menor (20- 30% aprox.). Como es sabido la amplificación genera sobre estimación de los grupos más abundantes (Poretsky *et al.* 2014), entre más abundante el grupo más alta la posibilidad de sobreestimación.

Respecto a la relación del grupo *Thaumarqueota* con los sedimentos, este grupo se caracteriza porque la mayoría de sus miembros son nitrificantes (reductores de amonio) (Pester *et al.* 2011) por lo que encontrarlo en sedimentos superficiales es muy común. Hasta hace poco se le clasificaba como una clase de *Crenarqueota* (Brochier-Armanet *et al.* 2008), y la base de datos que hemos usado, aún lo considera como tal, por eso en nuestras gráficas lo vemos como clase. Sin embargo, éste hecho no afecta

la caracterización de la comunidad, debido a que el error ya está tomando en cuenta y prácticamente el 100% de *Crenarqueota* es *Thaumarqueota*.

Respecto al incremento en la diversidad de arqueas en los sedimentos de fondo, es de suponerse, ya que el dominio arquea se caracteriza por tener miembros extremófilos (Madigan *et al.* 2010). Se ha reportado que *Euryarqueota* es comúnmente encontrado en los sedimentos más profundos debido a que es un grupo sumamente diverso. Entre sus miembros hay metanógenos, metano oxidadores, oxidadores de hierro, sulfato reductores, entre otros (Madigan *et al.* 2010). Con base en los resultados obtenidos se sugiere que los miembros de *Euryarqueota* detectados podrían ser metano oxidadores, debido al tipo de sedimento y a que numerosos estudios afirman que hay una simbiosis entre arqueas anaerobias metano oxidadoras y bacterias reductoras de sulfato (Hinrichs y Boetius, 2002), como las que detectaron con gran abundancia en nuestras muestras. Aunque no hay que negarla posibilidad de que miembros de *Euryarqueota* podrían ser oxidadores de hierro o sulfato reductores si se sustenta con la propuesta de zonación bioquímica propuesta por Orcutt (2011). Con base a esta zonación bioquímica y al análisis de la importancia de las arqueas en los ciclos biogeoquímicos (Offre *et al.* 2013) (ver Figura anexo 2), se sugiere que el resto de los grupos de arqueas encontrados en los sedimentos de fondo, son metanógenos, metano oxidadores, oxidadores de hierro, sulfato reductores, desnitrificadores y nitrificadores.

4.3 Comparación de la diversidad y abundancia de los microbiomas entre muestras de fondo y superficie.

Como se esperaba debido a los resultados de la clasificación taxonómica de las comunidades, la diversidad y abundancia de los grupos microbianos difiere entre los sedimentos superficiales y los de fondo del núcleo. El hecho de que la abundancia de algunos de los filos cambie respecto a la profundidad, es esperado pues las condiciones fisicoquímicas de los sedimentos cambian conforme aumenta la profundidad.

4.3.1 Alfa diversidad

La estimación de alfa diversidad nos ayuda a conocer la riqueza de especies dentro de una muestra. En este caso la estimación se realizó utilizando el índice de Shannon–Weaver y Chao1. Con el índice de diversidad de Shannon–Weaver se obtuvieron valores de entre 6.5 y 8.1 (Figura 15) y valores de estimación de riqueza Chao1, estuvieron entre 655 a 1114, similares a los observados (en rango de 506 y 894).

Los valores obtenidos son comparables con otros estudios de diversidad bacteriana en sedimentos, como los obtenidos por Verma et al. (2016) donde se reportó alta diversidad en sus muestras (valores 6.3 a 9), el estudio de Zhang et al. (2015) que reportaron valores del índice de Shannon de 7.4 a 8.2 y lo reportado por *Mamaeva et al. (2016)* quienes consideran una diversidad alta un índice de Shannon de 4.6.

Al analizar los datos referentes al índice Chao1 se observa que los valores estimados (en rango de 655 a 1114) son similares a los observados (en rango de 506 y 894), esto nos indica que nuestras muestras presentan índices de diversidad altos y que la riqueza de especies obtenida es muy similar a lo que se calcula debe ser la real. Además que esta diversidad es incluso mayor en los sedimentos del fondo del núcleo (Figura 16).

4.3.2 Beta diversidad

Se define beta diversidad como la diferencia en la composición de especies entre sitios. Y es utilizada para comparar comunidades biológicas diferentes. En resumen, el término beta diversidad hace referencia al grado de cambio en la composición de especies a lo largo de un gradiente ambiental o geográfico (Pirmank y Ros, 2002).

Al calcular la beta diversidad se observó la formación de tres grupos: el primero donde se agrupan las muestras A y B, en el segundo se agrupan las muestras G44 de fondo y el tercero que agrupa a las dos muestras H45 de fondo. Esto indica que no hay diferencias entre las muestras de sedimentos superficiales, independientemente de la profundidad en columna de agua (desde 1268 hasta 3768 m) y de la profundidad en núcleo del sedimento (10 primeros centímetros) del. Sin embargo, las muestras de superficie del núcleo son claramente diferentes de las más profundas (a más de 30 cm). Además también se observa que las muestras de fondo son diferentes entre ellas.

El hecho de no observar diferencias significativas entre las comunidades superficiales podría deberse a que la composición de los sedimentos analizados podía ser muy similar. La mayoría de nuestras muestras provienen de sedimentos de aguas profundas, ubicados en el talud o en la planicie del Golfo de México. Marcelino-Barros (2013) reportó que la composición granulométrica de las mismas estaciones, o cercanas a las nuestras, tomadas durante el crucero XIXIMI-1, es mayormente limosa, y que verticalmente (en los primeros 12 cm) su composición es homogénea.

Otro de los factores a considerar para explicar la similitud de las comunidades microbianas pertenecientes a las muestras de 0 a 10 cm de profundidad es la bioturbación de la macrofauna que vive en el sedimento. Se ha reportado que ésta contribuye a la circulación de nutrientes, de oxígeno, carbono, nitrógeno, azufre, etc. (Cullen, 1973; Snelgrove, 1997; Aller, 1982). En sedimentos del Golfo de México (campañas oceanográficas XIXIMI-1 y XIXIMI-2) se encontró actividad de microorganismos hasta a

10 cm de profundidad (Marcelino, 2013), esto contribuye a reforzar nuestros resultados de porque las comunidades de los primeros 10 cm son muy semejantes.

Las diferencias de las comunidades presentes en los primeros 10 cm y las presentes a 30 cm de profundidad resultan obvias ya que se están analizando dos ambientes totalmente distintos. Estas diferencias se explican principalmente por el tipo de sedimento. Los sedimentos analizados son limosos (Marcelino- Barros, 2013), lo que complica la circulación de agua ya que el sedimento es más compacto y si a esto le sumamos la disminución de la circulación de materia orgánica, nutrientes, oxígeno, carbono, nitrógeno, etc., proporcionada por la bioturbación, las diferencias se hacen más grandes.

Respecto a la diferencia de la composición de las comunidades microbiana entre las dos estaciones del fondo del núcleo (H45 y G44) colectadas a 2175 y 2470 m de profundidad respectivamente y cercanas a las zonas más productivas de Golfo debido a las desembocaduras de dos ríos: El Papaloapan (H45) y el Usumacinta-Grijalva (G44). Hubiéramos esperado que existieran diferencias entre las comunidades en los sedimentos superficiales. Incluso que los factores mencionados tuviesen más influencia en los contrastes que existen entre las muestras más superficiales y las más profundas. Sin embargo, al no ser así la explicación se puede deber a que la zonación biogeoquímica de los sedimentos profundos está más relacionada con los cambios de la estructura en las comunidades, que la profundidad en columna de agua y los aportes provenientes de los ríos, algo que ya reporto Orcutt *et al.* (2011). En este trabajo ellos explican que los aceptores de electrones dominantes en los sedimentos son diferentes conforme aumenta la profundidad, dependiendo de si son sedimentos de la plataforma continental o del talud, (ver anexo 1). Y debido a que su ubicación es diferente (una más cercana a la plataforma continental y la otra en el talud) los aceptores de electrones dominantes son diferentes, lo que se traduce en composición microbiana diferente.

Capítulo 5. Conclusiones

Una parte importante de este trabajo fue la estandarización de la secuenciación en la plataforma MiSeq de Illumina recientemente comprada en CICESE. Con base en los resultados de rendimiento y calidad de las secuencias, se puede afirmar que la secuenciación fue de buena calidad. Esto permitirá seguir secuenciando localmente y continuar con estudios similares al realizado.

En este estudio se caracterizaron, por primera vez, comunidades de bacterias y arqueas presentes en los sedimentos profundos del sur del Golfo de México, usando secuenciación masiva. Con ello se observó que dicha microbiota es abundante y muy diversa, y bastante similar (a nivel de filo) a lo encontrado en estudios de otros lugares del mundo. Aunque también, se observó que hay grupos que podrían ser específicos de los sedimentos del golfo de México.

La composición de la comunidad microbiana en cada estación no cambia significativamente dentro de los sedimentos más superficiales (A y B) y tampoco se observan diferencias significativas entre las estaciones. Posiblemente este hecho está ligado a la bioturbación causada por la meiofauna. Sin embargo, en los sedimentos colectados a más de 30 cm, sí hay diferencias significativas respecto a los superficiales así como entre las dos estaciones estudiadas (H45 y G44), estas diferencias podrían estar relacionadas con la zonación bioquímica de los sedimentos.

Respecto a grupos taxonómicos específicos, el dominio Arquea representa alrededor del 40 de los OTUs los sedimentos superficiales, sin embargo, en los sedimentos de fondo esta abundancia se reduce, pero en cambio su diversidad aumenta. Por otro lado el dominio Bacteria presenta tanto mayor abundancia como mayor diversidad en los sedimentos de fondo. En el caso particular de la gran abundancia del grupo *Thaumarqueota* en los sedimentos de superficie (40% en promedio) podría deberse a un ligero sesgo causado durante la construcción de las bibliotecas.

Finalmente, cabe mencionar que éste es un estudio explorativo y de diagnóstico. Representa el primer antecedente del estado actual de las comunidades de bacterias y arqueas presentes en los sedimentos profundos de la Zona Económica Exclusiva del Golfo de México. Y se espera sentar bases para estudios posteriores más ambiciosos.

Literatura citada

- Aller R. (1982). The effects of macrobenthos on chemical properties of marine sediment and overlying water. In *Animal-Sediment Relations*. Ed: P. McCall and M. Tevesz, pp. 53–102. New York: Plenum.
- Andrew S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Barns S., Cain C., Sommerville L., y Kuske C. (2007). Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl. Environ. Microbiol.* 73:3113-3116.
- Barton L. y Northrup D. (2011). *Microbial Ecology*. Oxford: Willey- Blackwell.
- Brochier-Armanet C., Boussau B., Gribaldo S., y Forterre P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*, 6(3), 245-252.
- Bouma A. (1972). Distribution of sediments and sedimentary structures in the Gulf of Mexico. Contribution on the geological and geophysical oceanography of the Gulf of Mexico: Texas A & M University Oceanography Studies 3, 35–65 p. 35-65.
- Carson R. (1980). *El Mar que Nos Rodea*. Ediciones Grijalbo, S.A., Barcelona España, 278 p.
- Caso M., Pisantry I. y Ezcurra E. (2004). Diagnóstico ambiental del golfo de México. México: Secretaría de Medio Ambiente y Recursos Naturales. Instituto Nacional de Ecología : Instituto de Ecología, A.C. : Harte Research Institute for Gulf of Mexico Studies. 1ra ed.
- Cullen, D. J. (1973). Bioturbation of superficial marine sediments by interstitial meiobenthos. *Nature* 242, 323 - 324 (30 March 1973).
- Dethlefsen L., Huse S., Sogin M., Relman D. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biology* 6: e280.
- Fry J., Parkes R., Cragg B., Weightman A., Webster G. (2008). Prokaryotic biodiversity and activity in the deep seafloor biosphere. *FEMS Microbiology Ecology* 66, 181–196.
- Handelsman J., Rondon M., Brady S., Clardy J., Goodman R. (1998). Molecular Biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol.* 5:R245–R249.
- Hill G.T., Mitkowski N.A., Aldrich-Wolfe L., Emele L.R., Jurkonie D.D., Ficke A., Maldonado-Ramírez S., Lynch S.T. y Nelson E.B. (2000). Methods for assessing the composition of soil microbial communities. *Applied Soil Ecology* 15:25-36
- Hinrichs K. y Boetius A. (2002). The anaerobic oxidation of methane: new insights in microbial ecology and biogeochemistry. In *Ocean Margin Systems* (pp. 457-477). Springer Berlin Heidelberg.
- Illumina, Inc. (2013). *Metagenomics 2nd Edition. A review of publications featuring Illumina Technology*
- Illumina, Inc. (2015). *Brochure sequencing systems portfolio*.
- Illumina, Inc. (2015) *MiSeq® System Guide*.
- Jørgensen B. y Boetius A. (2007). Feast and famine microbial life in the deep sea bed. *Nat Rev Microbiol* 5: 770–781.
- Joye S., Boetius A., Orcutt B., Montoya J., Schulz H., Erickson M., Lugo S. (2004). The anaerobic oxidation of methane and sulfate reduction in sediments from Gulf of Mexico cold seeps. *Chemical Geology* vol. 205 (3) p. 219-238
- Kimes N., Callaghan A., Aktas D, Smith W., Sunner j., Golding B., Morris P. J. (2013). Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Frontiers in Microbiology*: 4, 50.
- Kirk J., Beaudette L., Hart M., Moutoglou P., Klironomos J., Lee H. y Trevors J. T. (2004). Methods of studying soil microbial diversity. *Journal of Microbiological Methods*, 58 (2), 169-188.
- Koo H., Mojib N., Thacker R., Bej A. (2013). Comparative analysis of bacterial community-metagenomics in coastal Gulf of Mexico sediment microcosms following exposure to Macondo oil (MC252). *Antonie Van Leeuwenhoek*. 106(5):993-1009.

Kuczynski J., Stombaugh J., Walters W. A., González A., Caporaso J. G., y Knight R. (2011). Using QIIME to analyze 16S rRNA gene sequences from Microbial Communities. *Current Protocols in Bioinformatics / Editorial Board, Andreas D. Baxevanis*

Lara-Lara J.L., Arenas-Fuentes V., Bazán-Guzmán C., Díaz-Castañeda V., Escobar-Briones E., García-Abad M., Gaxiola-Castro G., Robles-Jarero G., Sosa-Ávalos R., Soto-González L.A., Tapia-García M., Valdez-Holguín J.E. (2008). Los ecosistemas Marinos. "Capital Natural de Mexico. Vol. I: Conocimiento actual de la biodiversidad". Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. 135-159 p.

Langille M. G., Zaneveld J., Caporaso J. G., McDonald D., Knights D., Reyes J. A., Clemente J.C., Burkepile D.E, Vega-Thurber R.L., Knight R., Beiko R.G., and Huttenhower C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821.

Madigan M.T., Martinko J.M. y Parker J. (2010). *Biología de los Microorganismos*. 10a edición, Pearson – Prentice Hall: Madrid, España.

Mamaeva E., Galach'yants P., Khabudaev V., Petrova P., Pogodaeva V., Khodzher B., y Zemskaya L. (2016). Metagenomic analysis of microbial communities of the sediments of the Kara Sea shelf and the Yenisei Bay. *Microbiology*, 85(2), 220-230.

Marcelino Barros (2013). Caracterización de la meiofauna del océano profundo del Golfo de México. . Tesis de Maestría en Ciencias. Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California. 108 pp.

Miller D. N., J. E. Bryant E. L. Madsen y Ghiorse W. C. (1999). Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* 65:4715–4724.

Offre P., Spang A., y Schleper C. (2013). Archaea in Biogeochemical Cycles. *Annu. Rev. Microbiol.* 67:437–57

Orcutt B., Sylvan J., Knab N. y Edwards (2011). Microbial Ecology of the Dark Ocean above at, and below the Seafloor, *Microbiol. Mol. Biol. Rev.* 2011, 75(2):361.

Parkes R.J., Cragg B.A., Bale S.J., Getliff J.M., Goodman K., Rochelle P.A., Fry J.C., Weightman A.J. and Harvey S.M. (1994). Deep bacterial biosphere in Pacific-Ocean sediments. *Nature* 371: 410–413.

Parkes R.J., Cragg B., Roussel E., Webster G., Weightman A., Sass H., (2014). A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions. *Marine Geology* 352: 409–425.

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.

Poretsky R., Rodriguez R. L.M, Luo C., Tsementzi D., Konstantinidis K.T. (2014) Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE* 9(4): e93827

Quail M. A., Smith M., Coupland P., Otto T., Harris S. R. Connor T.R., Berton A., Swerdlow H.P. and Gu Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341

Riding R. E. and Awramik S.M. (2000). *Microbial Sediments*. 1st ed. Springer-Verlag Berlin Heidelberg: Germany.

Rosas I., Cravioto A. y Ezcurra E. (2004). *Microbiología Ambiental*. 1ra ed., Instituto nacional de Ecología: Ciudad de México

Sagova-Mareckova M., Cermak L., Novotna J. , Plhachova K., Forstova J., y Kopecky J. (2008). Innovative Methods for Soil DNA Purification Tested in Soils with Widely Differing Characteristics. *Applied and environmental microbiology*. 65:4715–4724. , p. 2902–2907.

Stackebrandt E., Roselló-Mora R., Pukall R., Gilles V.C.P, Harayama S., Kasai H., Vauterin L., Vauterin P., Neufeld J.D., Mohn W.W., Glockner F.O., Meyerdierks A., Nubel U., Antwerpen M., Strommenger B. and witte W. (2006). *Molecular Identification, Systematics, and Population Structure of Prokaryotes*. 1st ed. Springer-Verlag Berlin Heidelberg: Germany.

Stal L. y Cretoui S. (2016). *The marine microbiome*. Springer International: Switzerland

Uribe-Vélez, D. (2009). Metagenómica. Una oportunidad para el estudio de la diversidad microbiana en Colombia. *Revista Colombiana de Biotecnología*, XI(2) Recuperado de <http://www.redalyc.org/articulo.oa?id=77613172001>

Tunnell J. W., Felder D.I., Earle S.A., Camp D.K., Buster N.A., Holmes C.W. (2013). *Gulf of Mexico origin, waters, and biota* College Station. 1st ed., Texas A&M University Press.

Verma P., Rangamaran V., Jeon C. Lee H., Priva P. y Dharani G., (2016). Complex bacterial communities in the deep-sea sediments of the Bay of Bengal and volcanic Barren Island in the Andaman Sea. *Mar Genomics*: S1874-7787(16)30089-7.

Woese C.R. and Fox G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74 (11): 5088–5090.

Woese C.R. (1987). Bacterial evolution. *Microbiol. Rev.* 51:221-271

Whitman WB, Coleman DC y Wiebe W. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* 95: 6578–6583.

Zhang J., Sun Q.L., Zeng Z.G., Chen S. and Sun L. (2015). Microbial diversity in the deep-sea sediments of Iheya North and Iheya Ridge, Okinawa Trough. *Microbiol Res.* 177:43-52

Zhang Y., Li X., Bartlett D., Xiao X. (2015b). Current developments in marine microbiology: high-pressure biotechnology and the genetic engineering of piezophiles. *Curr Opin Biotechnol* 33:157–164

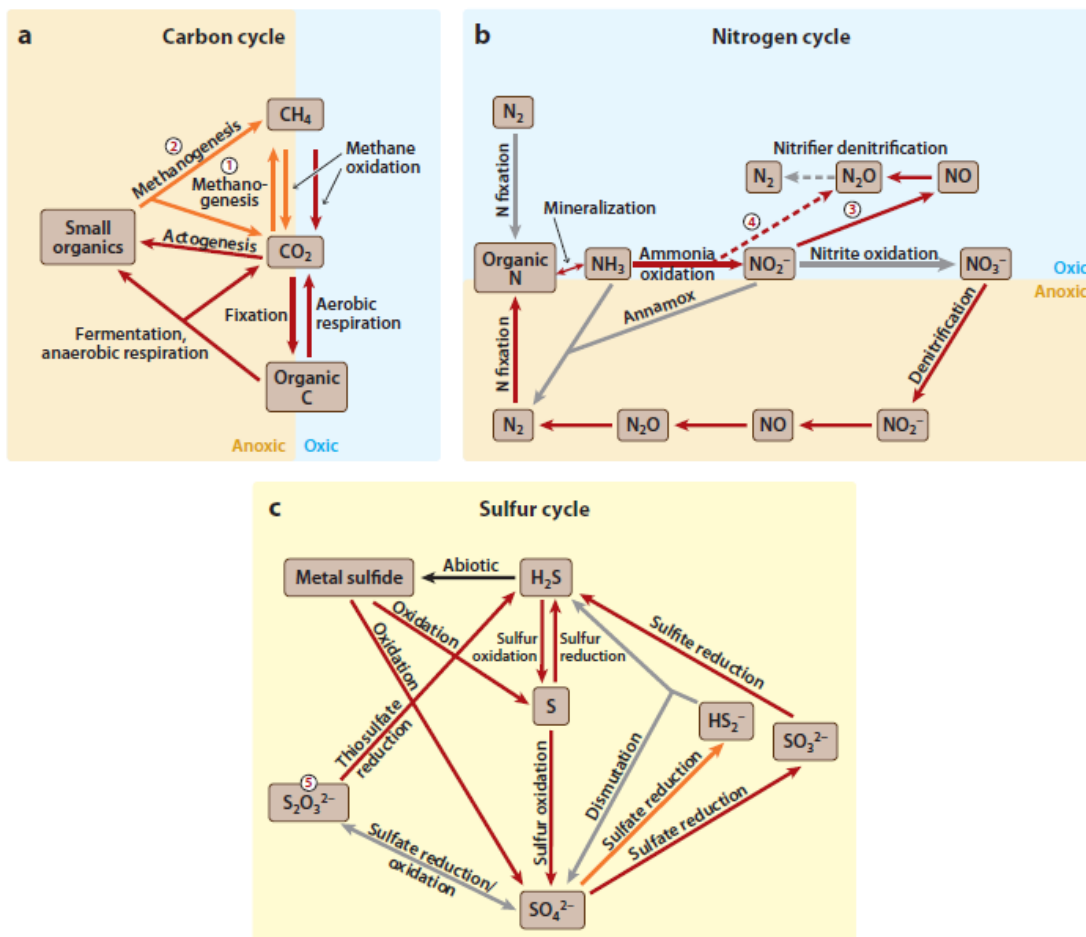
Anexos

Anexo 1. La tabla muestra las muestras utilizadas para la secuenciación, la concentración de ADN y sus índices.

MUESTRA	CONCENTRACIÓN (ng.µL)	INDICE F	INDICE R	SET DE OLIGOS
C22-2A	29,2	ACTATCTG	AACTCTCG	KR1-F1
C22-1B	34,7	ACTATCTG	GCTGTAAC	KR2-F1
B14-4A	10,4	ACTATCTG	GGACGTTA	KR3-F1
B14-1B	6,9	ACTATCTG	GGTCGTAG	KR4-F1
A7-1A	7,6	ACTATCTG	TAAGTCTC	KR5-F1
A7-5B	7,6	ACTATCTG	TACACAGT	KR6-F1
E31-5A	11,9	TAGCGAGT	AACTCTCG	KR1-F2
B18-1A	51,6	TAGCGAGT	GGACGTTA	KR3-F2
A1-2A	6,6	CTAGTATG	AACTCTCG	KR1-F3
A1-4B	8	CTAGTATG	GCTGTAAC	KR2-F3
G44-5A	12,4	CTAGTATG	TAAGTCTC	KR5-F3
H45-5A	13,6	GATAGCGT	AACTCTCG	KR1-F4
H45-3B	9,4	GATAGCGT	GCTGTAAC	KR2-F4
E31-3B	4,4	TAGCGAGT	GCTGTAAC	KR2-F2
B18-2B	5,4	TAGCGAGT	GGTCGTAG	KR4-F2
H47-4A	4,9	TAGCGAGT	TAAGTCTC	KR5-F2
H47-5B	4,4	TAGCGAGT	TACACAGT	KR6-F2
A5-5A	4,6	CTAGTATG	GGACGTTA	KR3-F3
A5-5B	5	CTAGTATG	GGTCGTAG	KR4-F3
G44-3B	7,2	CTAGTATG	TACACAGT	KR6-F3
H45-5F	3,7	GATAGCGT	GGACGTTA	KR3-F4
G44-3F	4,6	GATAGCGT	GGTCGTAG	KR4-F4

Anexo 2. Reacción de PCR

COMPONENTES	CONC.FINAL	µL
H ₂ O	-	-
Buffer	1x	5
Primer F	1.3µM	1
Primer R	1.3µM	1
My taq	1.25u	0.125
DNA	10ng	-
TOTAL		25µL



Anexo 5. Representación esquemática de la participación del dominio archaea en los ciclos de: (a) Carbono, (b) de Nitrógeno, y (c) de Azufre. Las flechas rojas indican los pasos metabólicos que se encuentran en arqueas y bacterias; flechas de color naranja indican las rutas metabólicas presentes exclusivamente en arqueas; y las flechas grises indican el metabolismo conocido sólo a partir de bacterias. Tomado de Offre *et al.* 2011.