



**UNIVERSIDAD AUTÓNOMA DEL ESTADO
DE MÉXICO**



FACULTAD DE CIENCIAS

“Caracterización metagenómica asociada a corrosión en un pozo petrolero mediante secuenciación del gen 16S rRNA”

TESIS

QUE PARA OBTENER EL TÍTULO DE:

Licenciado en Biotecnología

PRESENTA:

Andrea Casas Reyes

ASESOR:

Dr. Fernando Méndez Sánchez

CO-ASESOR:

Dr. Adrián Ochoa Leyva

Toluca, México 2018

ÍNDICE

DEDICATORIAS	4
AGRADECIMIENTOS	5
RESUMEN	6
1. Introducción	7
1.1 Corrosión de tuberías petroleras.....	7
1.2 Corrosión influenciada por microorganismos.....	7
1.3 Pérdidas económicas causadas por biocorrosión.....	8
1.4 Microorganismos implicados en las actividades corrosivas.....	9
1.4.1 Bacterias sulfato reductoras.....	11
1.4.2 Bacterias ácido productoras.....	11
2. Antecedentes	11
2.1 Caracterización de comunidades bacterianas.....	11
2.1.1 Cultivos bacterianos.....	12
2.1.2 Metagenómica.....	12
2.1.2.1 Metagenómica shotgun.....	12
2.1.2.2 Secuenciación de amplicones.....	13
2.2 Secuenciación de ADN de microbiota proveniente de pozos petrolero.....	13
3. Hipótesis	15
4. Objetivos	15
4.1 Objetivo general.....	15
4.2 Objetivos específicos.....	15
5. Materiales y métodos	15
5.1 Obtención de ADN bacteriano.....	16
5.2 Extracción de ADN bacteriano.....	16
5.3 Secuenciación.....	18
5.3.1 Obtención de amplicones.....	18
5.3.1.1 Ion Torrent.....	18
5.3.1.2 MiSeq.....	18

5.3.2 Purificación de amplicones.....	19
5.3.2.1 Ion Torrent.....	19
5.3.1.2 MiSeq.....	20
5.4 Filtrado por calidad de las secuencias.....	22
5.4.1 Secuencias Ion Torrent.....	22
5.4.2 Secuencias MiSeq.....	26
6. Resultados y discusión.....	29
6.1 Resultados experimentales.....	29
6.2 Filtrado por calidad de las secuencias.....	33
6.2.1 Ion Torrent.....	33
6.2.2 MiSeq.....	36
6.3 Selección de la base de datos ribosomal.....	41
6.3.1 Ion Torrent.....	41
6.3.2 MiSeq.....	44
6.4 Caracterización taxonómica Ion Torrent.....	46
6.5 Caracterización taxonómica MiSeq.....	53
6.6 Diversidad Ion Torrent (V2, V3, V4, V6, V7, V8 y V9) vs MiSeq (V3-V4).....	60
6.7 Comparación entre tecnologías de secuenciación (V3-V4).....	68
6.7.1 Ion Torrent (V2, V3, V4, V6, V7, V8 y V9) vs (V3-V4).....	68
6.7.2 MiSeq (Silva completo) vs (Silva V3-V4).....	73
6.7.3 Diversidad taxonómica V3-V4 Ion Torrent vs MiSeq.....	79
6.8 Diversidad de cultivos ACP y BSR de los pozos de Tabasco y Yucatán.....	89
7. Conclusiones	99
8. Bibliografía	101

RESUMEN

TÍTULO: Caracterización metagenómica asociada a corrosión en un pozo petrolero mediante secuenciación del gen 16S rRNA

Existen pocos estudios relacionados con la corrosión influenciada por microorganismos que logren dar un perfil más cercano de la comunidad bacteriana que se encuentra presente en los pozos petroleros, ya que en la mayoría de los trabajos realizados se utilizan técnicas moleculares y de cultivo que no logran cubrir en su mayoría el perfil taxonómico y de diversidad presente en las muestras analizadas. Es por eso que, en el presente trabajo de tesis se da un acercamiento más profundo de la comunidad bacteriana que existe en muestras de agua provenientes de un pozo petrolero ubicado en Tabasco, utilizando dos métodos de secuenciación masiva (Ion Torrent y MiSeq) los cuales secuencian distinto número de regiones hipervariables del gen ribosomal 16S. La comparación fue realizada para conocer qué plataforma de secuenciación resulta ser más conveniente al dar a conocer los resultados de diversidad y taxonomía del pozo petrolero.

Los resultados encontrados variaron según el método de secuenciación que fue utilizado para analizar las secuencias obtenidas del pozo petrolero de Tabasco, al comparar ambas plataformas, se obtuvo un perfil taxonómico más completo con la tecnología Ion Torrent. Por otro lado, sin importar la tecnología de secuenciación que fue utilizada, los tipos de microorganismos mayormente encontrados en las muestras se encuentran catalogados como sulfato reductores y productores de ácido, también fueron encontrados microorganismos que llevan a cabo actividades biorremediadoras, entre otros. Se cree que los microorganismos encontrados en el pozo petrolero se encuentran creciendo de manera sinérgica y de acuerdo con los principales tipos de bacterias encontradas, es probable que las tuberías del pozo analizadas en este trabajo estén corroídas por acción de los microorganismos.

1. Introducción

1.1 Corrosión de tuberías petroleras

La industria del petróleo consume una gran cantidad de material metálico. Cerca del 8% de la producción mundial de metales es utilizada en la producción, transporte y procesamiento de petróleo. El consumo de materiales metálicos en la industria del petróleo asciende a aproximadamente 32 kg de metal por cada tonelada de petróleo procesado. Las enormes cantidades de acero y otros metales utilizados en la industria petrolera están propensos a la corrosión en un grado más marcado que en otros ambientes industriales. La severidad del problema es tal que cerca de 1 kg de acero y sus aleaciones por tonelada de petróleo procesado son destruidos debido a la corrosión (Sastri, 2014).

Sin duda alguna, este tipo de instalaciones son vulnerables y se pueden deteriorar con el paso del tiempo. La ruptura de las paredes debido a la corrosión resulta ser un problema costoso para la industria petrolera que puede resultar en explosiones, interrupciones del producto, liberación de químicos peligrosos y daño ambiental (Duncan *et al.*, 2009).

Las tuberías están expuestas a muchos ambientes corrosivos distintos, tanto interna como externamente. El tipo de corrosión que existe en las tuberías metálicas es llamada electroquímica (Norsworthy, 2014). La corrosión electroquímica es una reacción que involucra la transferencia de electrones de un metal con estado de oxidación cero a un aceptor de electrones externo, causando la liberación de iones metálicos al medio y deteriorando el metal. Este proceso ocurre a través de una serie de reacciones de oxidación (anódicas) y reducción (catódicas) (Beech and Sunner, 2004).

La corrosión en las tuberías petroleras puede deberse también a la acción de microorganismos, ya que muchas veces las bacterias crecen y corroen los metales por mecanismos electroquímicos (Jones and Amy, 2002). Las actividades metabólicas de los microorganismos han estado implicadas en incidentes de fallas de las tuberías (Duncan *et al.*, 2009).

1.2 Corrosión influenciada por microorganismos

Tanto en ambientes naturales, como en ambientes hechos por el hombre, la corrosión ocurre cuando los materiales de metal puro o sus combinaciones (aleaciones) se someten a un cambio químico de su estado fundamental a una especie ionizada. Este tipo de reacciones puede estar influenciada por actividades microbianas, especialmente cuando los microorganismos se encuentran en contacto cercano con la superficie del metal (Beech and Gaylarde, 1999).

La deterioración del metal debido a actividades microbianas es llamada biocorrosión o corrosión influenciada por microorganismos. La corrosión influenciada por microorganismos es un proceso electroquímico donde los microorganismos inician, facilitan o aceleran una reacción de corrosión en una superficie metálica (Rajasekar *et al.*, 2010). La biocorrosión es definida también como la degradación de estructuras metálicas que son resultado de la actividad de una variedad de microorganismos, los cuales, pueden producir metabolitos agresivos para lograr hacer un ambiente corrosivo o son capaces de participar directamente en la reacción electroquímica que ocurre en la superficie del metal (Muthukumar *et al.*, 2003).

Muchas veces, existe la formación de biopelículas, siendo estas el resultado de un proceso de acumulación que inicia inmediatamente después de la inmersión del metal en un ambiente acuoso. En una primera etapa, se forma una película delgada debido a la deposición de iones inorgánicos y de compuestos orgánicos, esta delgada película es capaz de alterar las cargas electrostáticas y la humectabilidad de la superficie del metal facilitando así que pueda ser colonizado por bacterias (Videla and Herrera, 2005).

Las condiciones dentro de las tuberías e instalaciones de procesamiento suelen ser favorables para el crecimiento de microorganismos (Duncan *et al.*, 2009). En los microambientes ácidos anaerobios a menudo se forman biopelículas y pueden incrementar las tasas de corrosión. Aunque muchas veces las tasas de reacciones microbianas están limitadas por factores ambientales tales como disponibilidad de sustrato y nutrientes, aun así, son lo suficientemente rápidas para producir corrosión (Jones and Amy, 2002).

Los microorganismos pueden ser encontrados en una amplia gama de temperaturas, presión, salinidad y pH. Cualquiera que sean las condiciones ambientales, los microorganismos necesitan agua, una fuente de energía para impulsar su metabolismo y nutrientes para la renovación celular y crecimiento (Jack, 2002).

La temperatura resulta ser el principal factor limitante para el crecimiento microbiano, ya que esta incrementa a mayor profundidad a una tasa promedio de 3°C por 100 metros, por lo que los yacimientos de petróleo profundos que mantienen una temperatura *in situ* que excede los 130-150°C no pueden apoyar el crecimiento microbiano, lo que sugiere que la presencia de bacterias endógenas en los yacimientos petroleros puede estar limitada en un umbral de temperatura entre 80-90°C (Magot *et al.*, 2000).

Otro factor importante resulta ser la disponibilidad de donadores y aceptores de electrones, la cual gobierna el tipo de actividades metabólicas dentro del ambiente de los yacimientos petroleros, debido a que los campos petrolíferos son ambientes subterráneos profundos y generalmente están aislados de las aguas superficiales, sus potenciales redox son muy bajos y algunos aceptores de electrones están generalmente ausentes, como el oxígeno, el nitrato y el hierro; sin embargo, el agua contiene sulfato en distintas concentraciones y carbonato, cuyos factores llevan a suponer que los principales procesos metabólicos que ocurren en dichos ecosistemas son: la reducción de sulfato, metanogénesis, acetogénesis y procesos fermentativos (Magot *et al.*, 2000).

Para el caso de los sistemas cerrados y las instalaciones bajo tierra el metabolismo microbiano se basa en la energía derivada de las reacciones de oxidación y reducción. En algunos casos, los electrones liberados por la oxidación del metal son utilizados directamente en el metabolismo microbiano y junto con las condiciones químicas pueden promover la corrosión influenciada por microorganismos (Jack, 2002).

1.3 Pérdidas económicas causadas por biocorrosión

La corrosión del metal es uno de los principales daños que causa grandes pérdidas económicas en los sistemas de tuberías de la industria del petróleo (Neria-González *et al.*, 2006). El costo estimado total de corrosión metálica en los Estados Unidos representó aproximadamente 3.1% del Producto Interno Bruto en el año 2002, lo equivalente a 276 billones de dólares (US Federal

Highway Administration, 2002). Específicamente en los sectores de exploración y producción de las industrias de gas y petróleo se ha estimado un costo de 1.4 billones de dólares (Jan-Roblero *et al.*, 2004).

En los años 1950's, los costos de reparaciones y reemplazos relacionados con la corrosión influenciada por microorganismos en materiales de tuberías utilizados en distintos tipos de servicios en los Estados Unidos se estimaron en alrededor de \$ 0.5-2 billones anualmente (Beech and Gaylarde, 1999).

Booth (1964) en el Reino Unido sugirió que debido a la cantidad de consultas recibidas por corrosión que involucran fallas en las tuberías, una fracción significativa de dichas consultas (quizá el 50% o más) pudo ser atribuida directamente a causas microbiológicas.

Los efectos de la corrosión en la vida diaria se vuelven tan normales y cotidianos que no se les da la importancia debida, a pesar del costo que implica el mantenimiento y reposición de la infraestructura dañada; desafortunadamente, en México no se han hecho estudios completos que nos indiquen con claridad el impacto real que el fenómeno de la corrosión tiene en la economía del país (Ramírez Reyes *et al.*, 2011).

En la tabla 1 se muestran las principales industrias que se encuentran más afectadas por la corrosión influenciada por microorganismos.

Industria	Áreas afectadas
Industrias químicas	Tanques de acero inoxidable, tuberías y bridas
Generación de energía nuclear	Tuberías y tanques de carbono y acero inoxidable
Industrias de petróleo y gas en tierra y mar	Sistemas de inyección de agua, sistemas de manipulación de gas y petróleo
Industria de tuberías subterráneas	Suelos arcillosos saturados de agua con pH casi neutro y materia orgánica en descomposición
Industria de manejo y tratamiento de aguas residuales	Estructuras de concreto
Industria de mantenimiento de carreteras	Tuberías de alcantarillas
Industria de aviación	Tanques de aluminio de almacenamiento de combustible

Tabla 1. - Principales industrias y sus áreas afectadas debido a la corrosión.

Fuente: Tomado de Jones, 1996.

1.4 Microorganismos implicados en las actividades corrosivas

Los microorganismos que se han asociado con la corrosión involucran diversos géneros y especies y pueden dividirse en tres grandes grupos: 1) bacterias, 2) hongos y 3) algas (Iverson, 1987).

Cuando las bacterias crecen en las paredes de las tuberías, se desarrollan en colonias que contienen diversos tipos de bacterias, las cuales se ayudan mutuamente para sobrevivir. Muchas veces los microorganismos presentes en los ambientes acuosos tienden a formar biopelículas en

las superficies sólidas, dichas biopelículas consisten en poblaciones de microorganismos y sus secreciones poliméricas (Muthukumar *et al.*, 2003).

Las comunidades bacterianas de las biopelículas que se forman en las superficies de los materiales suelen ser heterogéneas por lo que existe una incertidumbre con respecto a cómo afectan estas comunidades en un ambiente dado. El conocimiento de la diversidad bacteriana de las biopelículas ayuda a entender las interacciones entre las bacterias corrosivas y la superficie del metal, así como las interacciones con otros microorganismos y provee las bases para el desarrollo de nuevos medios de detección y prevención de la corrosión (Neria-González *et al.*, 2006).

Los tipos de bacterias mejor conocidos que causan corrosión incluyen a las bacterias productoras de ácido, las bacterias reductoras de sulfato, las bacterias que oxidan hierro y manganeso, entre otras (Tabla 2). Normalmente las bacterias reductoras de sulfato y las bacterias productoras de ácido son los dos tipos comúnmente encontrados en tuberías de petróleo y gas. Las bacterias reductoras de sulfato producen sulfuro de hidrógeno, mientras que las bacterias productoras de ácido generan ácido acético/sulfúrico, los cuales son altamente corrosivos para la tubería (Muthukumar *et al.*, 2003).

Tipo de microorganismo	Característica	Efecto
Bacterias reductoras de sulfato: <i>Desulfovibrio sp.</i> <i>Desulfomonas sp.</i>	Utilizan H_2 para reducir SO_4^{2-} a S^{2-} : precipitación de H_2S y FeS	Despolarización catódica por el consumo de hidrógeno, despolarización anódica por sulfuros de hierro corrosivos
Bacterias que oxidan hierro y manganeso: <i>Gallionella sp.</i> <i>Leptothrix sp.</i> <i>Mariprofundus sp.</i>	Fe^{2+} a Fe^{3+} y Mn^{2+} a Mn^{3+} : formación de óxido de hierro y de dióxido de manganeso	Deposición de hierro y manganeso catódicamente activos
Bacterias reductoras de hierro: <i>Pseudomonas sp.</i> <i>Shewanella sp.</i> <i>Geothermobacter sp.</i>	Reducción de Fe^{2+} a Fe^{3+} : reducción de manganeso o de óxido de hierro	Reducción de óxidos de hierro y manganeso
Bacterias que oxidan sulfuro: <i>Thiobacillus sp.</i>	Oxida S^{2+} y SO_3^{2-} a H_2SO_4	Los ácidos corroen el metal
Bacterias y hongos que producen ácidos: <i>Clostridium sp.</i> <i>Fusarium sp.</i> <i>Penicillium sp.</i> <i>Hormoconis sp.</i>	Producción de ácidos como ácido nítrico, ácido sulfúrico y ácidos orgánicos	Disuelven el hierro, quelan el cobre, hierro y zinc
Bacterias que forman limo: <i>Clostridium sp.</i> <i>Bacillus sp.</i> <i>Desulfovibrio sp.</i> <i>Pseudomonas sp.</i>	Producción de sustancias poliméricas extracelulares	Los exopolímeros son capaces de unir iones metálicos

Tabla 2.- Principales bacterias y hongos relacionados con la corrosión del metal. Fuente: Tomado de Kip & A Van Veen, 2015.

1.4.1 Bacterias sulfato reductoras

El sulfato y el azufre son los aceptores de electrones para un gran grupo de deltaproteobacterias anaerobias que utilizan compuestos orgánicos o H_2 como donantes de electrones. El sulfuro de hidrógeno (H_2S) es el producto de la reducción tanto del sulfato como del azufre. Se conocen más de 40 géneros de estos organismos, denominados colectivamente como bacterias reductoras de sulfato (Madigan *et al.*, 2009).

Los organismos principalmente implicados y mejor estudiados en relación con la corrosión microbiana son las bacterias reductoras de sulfato que son un grupo de bacterias anaerobias estrictas, las cuales son taxonómicamente diversas, pero fisiológica y ecológicamente similares.

Aunque son consideradas anaerobias estrictas, pueden sobrevivir por largos periodos de tiempo en presencia de oxígeno. Las características de las bacterias reductoras de sulfato, además de sus diferencias morfológicas, es la amplia gama de fuentes de carbono que pueden utilizar, entre las cuales se encuentran principalmente el lactato, piruvato y malato, como también CO_2 y ácidos grasos (Iverson, 1987).

1.4.2 Bacterias ácido productoras

Las bacterias productoras de ácido comprenden una variedad de bacterias heterotróficas que comparten la habilidad común de producir ácidos orgánicos, cuando se desarrollan bajo condiciones reductoras. Estas bacterias son del tipo aerobio o anaerobias facultativas. Los ácidos producidos por estas bacterias se concentran en la superficie del metal, contribuyendo a la corrosión y generando pérdidas en la producción y calidad de los hidrocarburos (Gómez-Castellanos, 2009).

Las bacterias ácido productoras causan que el pH caiga significativamente de condiciones neutras a ácidas que van de niveles desde 3.5 a 5.5. Estas condiciones ácidas son lo suficientemente significativas para causar corrosión de cualquier estructura metálica. Debido a que estas actividades de producción de ácidos ocurren en ausencia de oxígeno, se ha encontrado que es muy probable que las bacterias productoras de ácido participen de forma significativa en las actividades de corrosión en asociación con las bacterias reductoras de sulfato, particularmente en las industrias de gas y petróleo (Droycon Bioconcepts, 2013).

2 Antecedentes

2.1 Caracterización de comunidades bacterianas

En la naturaleza, los microorganismos existen en poblaciones de células en interacción llamadas comunidades bacterianas, cuyas interacciones son estudiadas mediante la ecología microbiana. Los componentes principales de la ecología microbiana son la biodiversidad y la actividad microbiana, ya que cuando se conocen estas dos cuestiones, ayuda a menudo a aislar organismos de interés. Es por ello por lo que, la ecología microbiana utiliza numerosos métodos de cultivo y métodos independientes de cultivo con el fin de aislar y estudiar los microorganismos deseados (Madigan *et al.*, 2009).

2.1.1 Cultivos bacterianos

Se ha reconocido que las técnicas basadas en cultivos no son suficientes para la detección de microorganismos relacionados con la corrosión, ya que, en el pasado, la enumeración y la identificación de bacterias de muestras provenientes de campos petroleros fueron principalmente logrados por métodos basados en cultivos (Keasler *et al.*, 2013).

Sin embargo, existen varios riesgos al utilizar este tipo de métodos, incluyendo largos tiempos de incubación, como también la incapacidad de cultivar bacterias que se encuentran presentes únicamente en ambientes extremos, las cuales pueden ser comúnmente encontradas en los yacimientos productores de petróleo, por lo que usualmente, esto conduce a la subestimación de la población bacteriana presente en las muestras, existiendo así deficiencias al momento de presentar los resultados (Keasler *et al.*, 2013).

Debido a las deficiencias que pueden existir al recurrir a los métodos clásicos de cultivos bacterianos, se han adaptado diferentes estrategias para dar a conocer resultados que se encuentren más cercanos a la realidad de las poblaciones existentes en las muestras a analizar.

2.1.2 Metagenómica

El estudio de comunidades microbianas ha cambiado desde el primer reporte de microorganismos hecho por Leeuwenhoek y sus organismos orales en 1676, hasta la caracterización utilizando las técnicas moleculares de hoy en día (Escobar-Zepeda *et al.*, 2015).

A finales de 1970, Carl Woese y George Fox, propusieron el uso de genes de RNA ribosomal como marcadores moleculares para la clasificación de eucariontes y procariontes (Woese and Fox, 1977). Esta idea en conjunto con la secuenciación Sanger revolucionó el estudio y la clasificación de los microorganismos. El desarrollo de nuevas técnicas moleculares tales como la reacción en cadena de la polimerasa (PCR), clonación y secuenciación de genes de rRNA, hibridación fluorescente *in situ* (FISH), electroforesis en gel por gradiente de desnaturalización (DGGE); entre otras, dieron la pauta a una nueva área denominada hoy en día como metagenómica (Escobar-Zepeda *et al.*, 2015).

La metagenómica involucra el muestreo de las secuencias del genoma de una comunidad de organismos que habitan un ambiente común y provee una visión imparcial no sólo de la estructura de la comunidad (riqueza y distribución de las especies), sino también del potencial funcional (metabólico) de una comunidad (Hugenholtz and Tyson, 2008).

Las principales cuestiones que hay que resolver en ecología microbiana son ¿quién está ahí? y ¿qué está haciendo?, cuestiones que la metagenómica puede responder; por lo que la diversidad microbiana se puede determinar utilizando dos distintos acercamientos 1) Metagenómica shotgun y 2) Secuenciación de amplicones (Escobar-Zepeda *et al.*, 2015).

2.1.2.1 Metagenómica shotgun

Los métodos microbiológicos moleculares se basan en el análisis del material genético (DNA/RNA) y se puede conocer la diversidad a nivel de especie (analizando el gen 16S rRNA) o a nivel funcional (analizando genes que codifican para enzimas específicas) (Geurkink *et al.*, 2016).

Los recientes avances en la secuenciación de ADN y en bioinformática permiten la exploración de la diversidad genética de las comunidades microbianas que aún no se han cultivado. La secuenciación shotgun es un enfoque alternativo al estudio de dicha microbiota. En esta técnica, el ADN es extraído de una comunidad, pero en lugar de dirigirse a un locus genómico específico para su secuenciación, todo el ADN extraído es cortado en pequeños fragmentos que son secuenciados de manera independiente; esto resulta en secuencias de ADN (lecturas) que son alineadas a varias ubicaciones genómicas, algunas de estas lecturas alinearán a locus taxonómicamente informativos como el 16S y otras proporcionarán información sobre las funciones biológicas codificadas en el genoma, por lo que las muestras secuenciadas mediante metagenómica shotgun resultarán en informar dos aspectos de una comunidad microbiana: la diversidad presente en la muestra y lo que es capaz de hacer dicha comunidad (Sharpton, 2014).

2.1.2.2 Secuenciación de amplicones

La secuenciación de amplicones es el método más ampliamente utilizado para caracterizar la diversidad de microorganismos. En este método, se muestrea una comunidad y se extrae el ADN de todas las células de la muestra, el siguiente paso es amplificar mediante PCR un marcador genómico taxonómicamente informativo que es común a prácticamente todos los organismos de interés; los amplicones que se obtienen son secuenciados y posteriormente caracterizados mediante bioinformática para determinar qué microorganismos están presentes en la muestra y con qué abundancia relativa. En el caso de bacterias y arqueas, los estudios de amplicones usualmente van dirigidos al locus de la sub-unidad pequeña de ARN ribosomal (16S), que resulta ser tanto un marcador taxonómico como filogenético. Cuando se comparan los perfiles obtenidos de la secuenciación del gen 16S entre las muestras, se puede aclarar cómo la diversidad microbiana se asocia con las condiciones ambientales (Sharpton, 2014).

2.2 Secuenciación de ADN de microbiota proveniente de pozos petroleros

En la tabla 3 se presentan algunos antecedentes acerca de investigaciones realizadas en pozos petroleros utilizando distintos métodos de biología molecular, se presentan también los hallazgos encontrados en los mismos, fueron tomados en cuenta los artículos presentados en la tabla debido a las características que presentan, ya sea que el lugar de muestreo es el mismo de este trabajo, o que las técnicas tanto de secuenciación como del análisis de datos fueron muy similares a las utilizadas en este trabajo.

Objetivo	Tipo de muestra	Método	Conclusión	Autor
Analizar por métodos moleculares y de cultivo la diversidad de una biopelícula desarrollada en una tubería petrolera con signos de corrosión.	Dos muestras de biopelículas formadas dentro de dos distintas tuberías.	El ADN extraído fue amplificado con primers universales para el gen ribosomal 16S y secuenciado con ABI PRISM 310. Se realizó PCR anidada para buscar bacterias sulfato reductoras. Para estudiar la diversidad, se amplificó la región hipervariable V3 del gen 16S y se realizó un análisis de DGGE.	Los resultados del autor sugieren que la comunidad encontrada fue limitada, encontrando tres principales grupos de bacterias: anaerobias facultativas, fermentadoras anaerobias de carbohidratos y sulfato reductoras.	Isabel Neria-González, En Tao Wang, Florina Ramírez, Juan M. Romero y César Hernández Rodríguez, 2006.
Secuenciar el gen 16S rDNA para identificar las comunidades bacterianas presentes en las tuberías de transporte del petróleo.	Muestras de agua de las tuberías.	El ADN genómico aislado se amplificó con primers de eubacterias específicos para el gen 16S rDNA. La secuenciación del ADN se realizó en el equipo ABI PRISM 310. Las secuencias obtenidas se analizaron con BLAST.	Fueron encontradas 11 especies distintas de bacterias, entre ellas <i>Serratia</i> , <i>Bacillus</i> y <i>Pseudomonas</i> , no fueron encontradas bacterias SRB, los autores concluyen que las bacterias causantes de corrosión pueden ser de diferentes géneros, no necesariamente BSR.	Aruliah Rajasekar, Balakrishnan Anandkumar, Sundaram Maruthamuthu, Yen Peng Ting, Pattanathu K. S. M. Rahman, 2010.
Secuenciar el gen 16S de bacterias nitrato reductoras, sulfato reductoras y metanógenas para conocer su diversidad y distribución en un pozo petrolero para determinar si este tiene potencial para un tratamiento de recuperación del petróleo (MEOR).	Muestras de agua obtenidas de las tuberías de inyección y producción.	Fue secuenciada la región V4 del gen ribosomal 16S con la plataforma MiSeq y las secuencias obtenidas se analizaron con QIIME. Se construyeron clones de genes específicos para bacterias sulfato reductoras, nitrato reductoras y metanógenas.	Se encontró una gran diversidad de bacterias, la mayoría de ellas capaces de crecer con hidrocarburos como única fuente de carbono. Las bacterias nitrato reductoras, sulfato reductoras y metanógenas encontradas sugieren que en el pozo existe potencial para llevar a cabo un tratamiento MEOR.	Peike Gao, Huimei Tian, Guoqiang Li, Hongwen Sun, Ting Ma, 2015.
Secuenciar el gen 16S para determinar la composición microbiana durante un tratamiento de recuperación del petróleo (MEOR) en	Muestras de agua de tres distintas tuberías de producción en distintos días del tratamiento.	Fue extraído ADN de las muestras de agua y se secuenciaron las regiones V3, V4 y V5 del gen 16S con MiSeq. Las secuencias	Se encontraron phylum que resultaron benéficos para el tratamiento (MEOR), sin embargo, al finalizar dicho tratamiento,	Jing You, Gang Wu, Fuping Ren, Qi Chang, Bo Yu, Yanfen Xue, Bozhong Mu, 2016.

un pozo.		obtenidas fueron analizadas con QIIME usando como referencia la base de datos Greengenes.	los phylum benéficos fueron disminuyendo con el paso del tiempo.
----------	--	---	--

Tabla 3.- Antecedentes de secuenciación en pozos petroleros.

3 Hipótesis

La secuenciación de siete regiones hipervariables del gen 16S rRNA versus dos regiones permite una mejor estimación de la diversidad microbiana presente en un pozo petrolero.

4 Objetivos

4.1 Objetivo general

Analizar y caracterizar la diversidad de microorganismos presentes en un pozo petrolero ubicado en Tabasco mediante dos tecnologías de secuenciación masiva del gen ribosomal 16S y que puedan estar asociados a la actividad óxido reductora y ácido productora.

4.2 Objetivos específicos

- Estandarizar las técnicas de extracción, amplificación y secuenciación de las regiones V3 y V4 (MiSeq) y V2, V3, V4, V6, V7, V8 y V9 (Ion Torrent) del gen 16S rRNA.
- Diferenciar taxonómicamente los microorganismos encontrados en las muestras de agua de las tuberías y comparar con lo reportado en la literatura.
- Comparar la diversidad obtenida con los dos métodos de secuenciación: Ion Torrent y MiSeq.
- Encontrar cuál es la proporción del microbioma caracterizado que pertenece a bacterias sulfato reductoras y ácido productoras.

5 Materiales y métodos

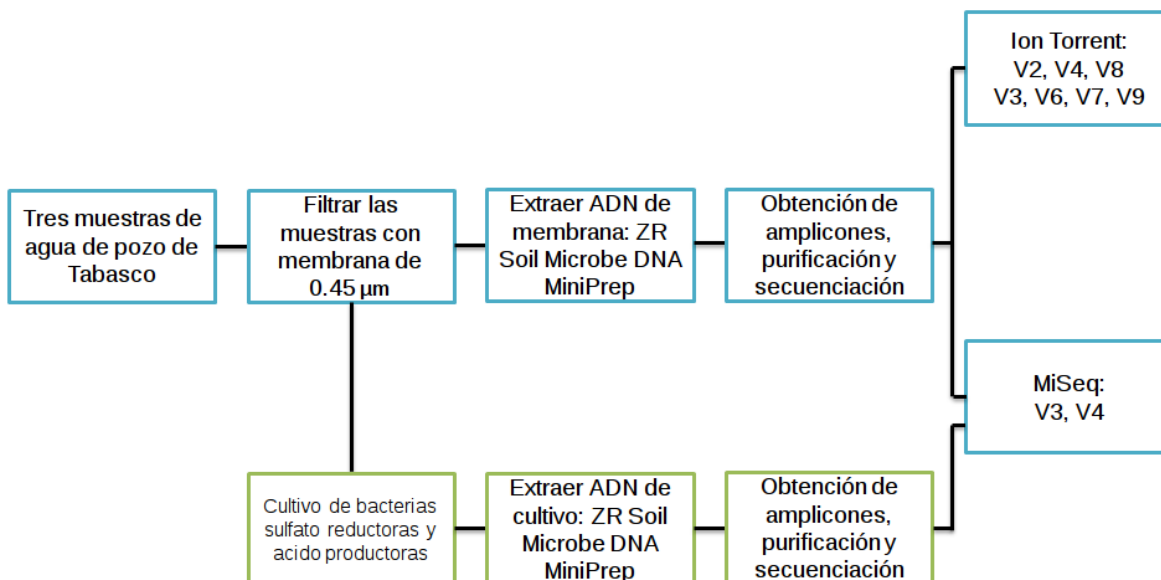


Figura 1.- Esquema general de la metodología aplicada a las muestras de agua obtenidas del pozo.

5.1 Obtención de ADN bacteriano

Se obtuvieron tres réplicas técnicas de muestras de agua de la tubería de un pozo petrolero ubicado en Tabasco, México, las cuales fueron etiquetadas como réplica 2, réplica 3 y réplica 4.

Fueron realizadas pruebas de diluciones seriadas a las muestras de agua del pozo para determinar la presencia de microorganismos del tipo sulfato reductor y ácido productor en cultivos específicos para las bacterias del tipo antes mencionado.

Las muestras de agua fueron filtradas con una membrana de 0.45 μm en unidades de filtrado Nalgene conectadas a una bomba de vacío, para poder retener los microorganismos presentes en la muestra en dicha membrana.

Se obtuvieron tres membranas correspondientes a cada una de las réplicas del pozo, las cuales fueron divididas en cuatro porciones, siendo distribuidas de la siguiente manera:

- $2/4$ de membrana fueron almacenados en buffer RLT + β Mercaptoetanol
- $1/4$ de membrana se almacenó en RNA Later
- $1/4$ de membrana se almacenó sin reactivo

Las membranas almacenadas con RLT + β Mercaptoetanol fueron utilizadas para la extracción de ADN bacteriano obtenidas de las tres réplicas.

Para conocer la diversidad del microbioma perteneciente a bacterias sulfato reductoras y ácido productoras de las réplicas, se realizaron cultivos específicos para dichas bacterias, inoculando una porción de membrana de la réplica 3, una vez inoculados los medios de cultivo, se incubaron a 35°C y 37°C para observar crecimiento de bacterias sulfato reductoras y ácido productoras, respectivamente.

Adicional a los cultivos inoculados con una porción de membrana de la réplica 3, fue inoculada agua de un pozo petrolero ubicado en Yucatán en medios de cultivo para bacterias sulfato reductoras y ácido productoras con la finalidad de comprobar si existen diferencias entre los microorganismos que pueden ser encontrados en dos distintos puntos de muestreo, los cultivos realizados con el agua de pozo de Yucatán fueron incubados con las mismas condiciones antes mencionadas.

5.2 Extracción de ADN bacteriano

Fue utilizado el protocolo Zymo Research Soil Microbe DNA MiniPrep para la extracción de ADN bacteriano de las tres réplicas del pozo y de los cultivos bacterianos que presentaron crecimiento; el protocolo fue seguido realizando modificaciones.

Se realizaron las extracciones de ADN de las tres réplicas del pozo las cuales fueron etiquetadas como 2/5, 3/5 y 4/5 para las muestras que serán secuenciadas con Ion Torrent, mientras que las muestras que serán secuenciadas con MiSeq fueron etiquetadas como ACR 2-5, ACR 3-5 y ACR 4-5. Los cultivos bacterianos fueron etiquetados como BPA y BSR.

El protocolo utilizado para la extracción de ADN bacteriano es el siguiente:

*De acuerdo con el protocolo de Zymo Research Soil Microbe DNA MiniPrep, originalmente, el siguiente paso era realizado únicamente en muestras de suelo. Antes de empezar con la extracción, se deben preparar las columnas IV-HRC del kit de la siguiente manera: 1) desechar la base, 2) insertar la columna en un tubo colector y 3) centrifugar a 8,000 x g por 3 minutos.

1. Agregar a un tubo de lisis (ZR Bashing Bead) la porción de membrana contenida en RLT + β mercaptoetanol de cada una de las 3 réplicas y añadir 750 μ l de buffer de lisis. Para el caso de los cultivos, se formó un pellet de bacterias, centrifugando a 10,000 rpm por 1 minuto en un tubo eppendorf que contenía 2 ml de cultivo, desechar el sobrenadante y re suspender el pellet en 750 μ l de buffer de lisis, el tubo eppendorf se llevó al vórtex para homogeneizar el pellet con el buffer y transferir todo el contenido a un tubo de lisis.
2. Llevar los tubos de lisis que contienen la porción de membrana y el pellet de los cultivos a un disruptor y procesar a máxima velocidad por 5 minutos.
3. Centrifugar el tubo de lisis a 10,000 x g por 1 minuto y transferir 400 μ l del sobrenadante a la columna que filtra el lisado del paso anterior (Spin Filter IV) y centrifugar a 7,000 x g por 1 minuto.
4. Añadir 1200 μ l de Soil DNA Binding Buffer al filtrado en el tubo colector del paso anterior y homogeneizar con la pipeta.
5. Transferir 800 μ l de la mezcla homogeneizada del paso anterior a la siguiente columna (Zymo Spin IIC) utilizando un nuevo tubo colector, centrifugar a 10,000 x g por 1 minuto. Desechar el sobrenadante y repetir el mismo paso.
6. Añadir 200 μ l de DNA Pre-Wash Buffer a la columna IIC utilizando un nuevo tubo colector y centrifugar a 10,000 x g por 1 minuto.
7. Añadir 500 μ l de Soil DNA Wash Buffer a la columna IIC y centrifugar a 10,000 x g por 1 minuto.
8. Transferir la columna IIC a un nuevo tubo colector y eluir dos veces el ADN como se indica. Para el caso de las tres réplicas, añadir 20 μ l de agua DEPC en la primera elución y centrifugar a 10,000 x g por 30 segundos y guardar el filtrado. Para la segunda elución, tomar otro tubo colector y añadir 100 μ l de agua DEPC, repitiendo el mismo proceso para centrifugar de la primera elución. Para realizar las eluciones de los cultivos bacterianos, se sigue el mismo procedimiento, esta vez eluyendo el ADN en 20 μ l y en 300 μ l.
9. Tomar el ADN del paso anterior y transferirlo a la columna que filtra el ADN de impurezas (Zymo-Spin IV-HRC), la cual fue previamente preparada antes de empezar con el procedimiento de extracción, centrifugar la columna con el ADN a 8,000 x g por 1 minuto, colectando el ADN en un tubo eppendorf de 1.5 ml, repitiendo el mismo procedimiento para las segundas eluciones, sin cambiar de columna.

Una vez que fue extraído el ADN de las réplicas del pozo y de los cultivos bacterianos de acuerdo con el protocolo antes mencionado, el siguiente paso fue obtener los amplicones que serán secuenciados mediante Ion Torrent y MiSeq.

5.3 Secuenciación

5.3.1 Obtención de amplicones

5.3.1.1 Ion Torrent

Para obtener los amplicones de las regiones V2, V3, V4, V6, V7, V8 y V9 fue utilizado el protocolo: Ion 16S Metagenomics Kit, el cual contiene un juego de dos oligonucleótidos que abarcan las siete regiones hipervariables, un juego de oligonucleótidos es específico para las regiones V2, V4 y V8 y el otro es específico para las regiones V3, V6, V7 y V9.

Se describen los reactivos y las cantidades necesarias para obtener los amplicones de las siete regiones hipervariables del gen ribosomal 16S.

Reactivo	Réplica 2	Réplica 3	Réplica 4
Master Mix + Oligos (V2, V3)	6 µl	6 µl	6 µl
ADN	3.2 µl	3.2 µl	2.9 µl
Agua	0.8 µl	0.8 µl	1.1 µl
Total	10 µl	10 µl	10 µl

Tabla 4.- Reactivos y cantidades necesarias para la obtención de amplicones de las tres réplicas del pozo.

El procedimiento y las condiciones de la PCR para obtener los amplicones se describen a continuación:

El ADN utilizado para la obtención de amplicones secuenciados con Ion Torrent fue de 300 ng. Una vez realizados los cálculos para la obtención de amplicones de cada una de las muestras a secuenciar, los tubos son llevados al termociclador empleando el siguiente programa:

- Mantener a 95°C por 10 minutos
- 30 ciclos de:
 - 95°C por 30 segundos
 - 58°C por 30 segundos
 - 72°C por 20 segundos
- Mantener a 72°C por 7 minutos

El tamaño de los amplicones obtenidos con Ion Torrent es de aproximadamente 200-300 pares de bases, por lo que fue verificada la presencia del producto en un gel de agarosa 2%.

5.3.1.2 MiSeq

Para la obtención de amplicones de las regiones V3 y v4 del gen 16S rRNA fue utilizado el protocolo: 16S Metagenomic Sequencing Library Preparation.

Los oligonucleótidos forward y reverse fueron sintetizados en la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología UNAM con las secuencias específicas para la obtención de amplicones de las regiones hipervariables V3 y V4 del gen ribosomal 16S, como lo indica el protocolo antes mencionado; los nombres de los oligos proporcionados por la Unidad de Secuenciación fueron 248 y 249. Se prepararon alícuotas de los oligos de acuerdo con la concentración que se requiere en el protocolo.

Las muestras de las que se obtuvieron los amplicones fueron las tres réplicas: 2/5, 3/5 y 4/5 y de los cultivos de bacterias sulfato reductoras y ácido productoras, nombradas como BPA-M para el cultivo obtenido del pozo de Tabasco y BSR, BPA para los cultivos obtenidos del pozo de Yucatán.

Se describen los reactivos y cantidades utilizadas para obtener los amplicones de las réplicas y los cultivos.

Reactivo	Réplica 2	Réplica 3	Réplica 4
ADN	0.8 µl	1.52 µl	0.52 µl
Oligo 248	5 µl	5 µl	5 µl
Oligo 249	5 µl	5 µl	5 µl
Master Mix	12.5 µl	12.5 µl	12.5 µl
Agua	1.7 µl	1.98 µl	1.98 µl
Total	25 µl	25 µl	25 µl

Tabla 5.- Reactivos y cantidades necesarias para la obtención de amplicones de las tres réplicas del pozo.

Reactivo	BPA-M	BPA	BSR
ADN	3.1 µl	1.2 µl	1.5 µl
Oligo 248	5 µl	5 µl	5 µl
Oligo 249	5 µl	5 µl	5 µl
Master Mix	12.5 µl	12.5 µl	12.5 µl
Agua	-	1.3 µl	1 µl
Total	25.6 µl	25 µl	25 µl

Tabla 6.- Reactivos y cantidades necesarias para la obtención de amplicones de los cultivos ácido-productores y sulfato reductores de los pozos de Tabasco y Yucatán.

Una vez que se realizaron los cálculos para la obtención de amplicones de cada una de las muestras a secuenciar, los tubos son llevados al termociclador empleando el siguiente programa:

- 95°C por 3 minutos
- 25 ciclos de:
 - 95°C por 30 segundos
 - 55°C por 30 segundos
 - 72°C por 30 segundos
- 72°C por 5 minutos
- Mantener a 4°C

Al finalizar la PCR, se deben obtener amplicones de un tamaño de aproximadamente 500-550 pares de bases, por lo que se verificó la obtención del producto en un gel de agarosa 2%.

5.3.2 Purificación de amplicones

5.3.2.1 Ion Torrent

Los amplicones obtenidos fueron purificados con el kit QIAquick Gel Extraction de QIAGEN cortando la banda que corresponde al tamaño del amplicón (200-300 pb) directamente del gel de agarosa, ya que al purificar con las perlas AMPure, se observaban barridos de bandas en el gel.

Los reactivos utilizados para la purificación del amplicón fueron los siguientes:

- Buffer QG (debe permanecer de color amarillo, lo cual indica un pH 7.5)
- Buffer PE (agregar etanol 96-100% antes de usarlo)
- Isopropanol 100%

Procedimiento:

1. Cortar el fragmento del gel con una navaja limpia tratando de no cortar agarosa extra. Antes de cortar, limpiar la superficie donde se van a colocar el gel y la navaja con cloro 10% y DNAzap.
2. Colocar el fragmento de agarosa en un tubo estéril de 1.5 ml y agregar 3 volúmenes de buffer QG por cada volumen de agarosa.
3. Incubar a 50°C durante 10 minutos o hasta que la agarosa se disuelva. Si el color de la mezcla es naranja o violeta, agregar 10 µl de acetato de sodio 3M pH 5.0 y mezclar, el color de la mezcla cambiará a amarillo, este es un paso importante ya que la unión de ADN a la membrana de las columnas sólo es eficiente a un pH < 7.5.
4. Después de que la agarosa se disuelva por completo, agregar un volumen de isopropanol y mezclar, posteriormente, agregar la mezcla a una columna QIAquick y centrifugar a 13,000 rpm por 1 minuto.
5. Descartar el sobrenadante, volver a colocar el mismo tubo colector y agregar 500 µl de buffer QG a la columna y centrifugar a 13,000 rpm por 1 minuto.
6. Agregar 750 µl de buffer PE a la columna y esperar de 5-10 minutos antes de centrifugar a 13,000 rpm por 1 minuto para lavar el ADN.
7. Descartar el sobrenadante, centrifugar a 13,000 rpm por 1 minuto para eliminar los residuos de etanol y colocar la columna en un tubo limpio de 1.5 ml para hacer la elución.
8. Agregar 32-42 µl de agua o buffer de elución en el centro de la membrana y esperar 1 minuto antes de centrifugar a 13,000 rpm por 1 minuto.

Una vez purificados los amplicones, se siguieron una serie de pasos para poder ser secuenciados mediante la plataforma de Ion Torrent.

5.3.2.2 MiSeq

Para purificar los amplicones previamente obtenidos, fue utilizado el protocolo 16S Metagenomic Sequencing Library Preparation; se describen a continuación los reactivos y cantidades necesarias para la purificación de los amplicones.

Reactivo	Cantidad
Tris 10 mM pH 8.5	52.5 µl por muestra
Perlas AMPure XP	20 µl por muestra
Etanol 80% recién preparado	400 µl por muestra

Tabla 7.- Reactivos y cantidades necesarias para purificar los amplicones obtenidos con MiSeq.

Nota: Las perlas AMPure XP deben estar a temperatura ambiente al menos dos horas antes de su uso.

Procedimiento:

1. Centrifugar los tubos de PCR donde se encuentran los amplicones, dando un pulso en la centrífuga para evitar que se queden gotas en las paredes.
2. Llevar al vórtex las perlas AMPure por 30 segundos para asegurar que las perlas se dispersen.
3. Añadir 20 μ l de perlas AMPure a cada muestra.
4. Pipetear arriba y abajo 10 veces para homogeneizar la muestra, posteriormente llevar al shaker a 1800 rpm por 2 minutos.
5. Incubar a temperatura ambiente sin mover el tubo, por 5 minutos.
6. Colocar los tubos en un stand magnético por 2 minutos, hasta que se haya aclarado el sobrenadante.
7. Sin quitar los tubos del stand magnético, remover y desechar el sobrenadante, con cuidado de no llevarse perlas.
8. Sin retirar los tubos del stand magnético, lavar las perlas con etanol 80% como se indica:
 - a. Agregar 200 μ l de etanol 80% a cada muestra.
 - b. Incubar los tubos en el stand por 30 segundos.
 - c. Con cuidado, remover y desechar todo el sobrenadante.
9. Realizar un segundo lavado con etanol, sin quitar los tubos del stand, siguiendo las instrucciones del paso anterior.
10. Sin mover los tubos del stand, permitir que las perlas se sequen al aire por 10-15 minutos.
11. Remover los tubos del stand magnético y añadir 52.5 μ l de Tris 10 mM a cada tubo, pipetear 10 veces arriba y abajo para homogeneizar la muestra y posteriormente llevar al shaker a 1800 rpm por 2 minutos (centrifugar con un pulso si se quedan gotas en las paredes).
12. Incubar los tubos a temperatura ambiente por 2 minutos, sin moverlos.
13. Colocar los tubos en el stand magnético por 2 minutos hasta que se haya aclarado el sobrenadante.
14. Transferir 50 μ l del sobrenadante a un tubo limpio.
15. Verificar en un gel de agarosa 2% los amplicones purificados.

Una vez purificados los amplicones, se siguieron una serie de pasos para poder ser secuenciados mediante la plataforma de MiSeq.

5.4 Filtrado por calidad de las secuencias

5.4.1 Secuencias Ion Torrent

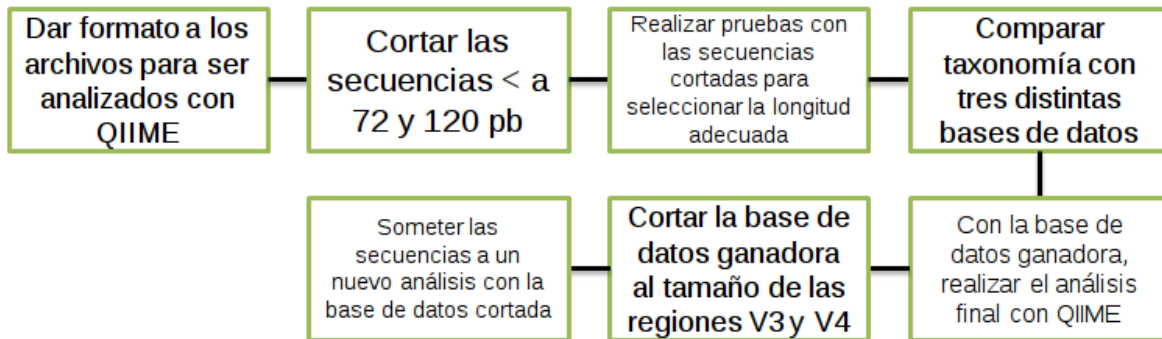


Figura 2.- Esquema general del método utilizado para filtrar por calidad las secuencias obtenidas con Ion Torrent.

Se obtuvieron los archivos que contienen las lecturas de las tres réplicas del pozo secuenciadas con la plataforma Ion Torrent. El primer paso para trabajar con las lecturas fue dar formato a las secuencias obtenidas para que pudieran ser analizadas con el programa QIIME (Caporaso *et al.*, 2010).

Las lecturas que arroja la plataforma de Ion Torrent vienen sin los oligos y sin códigos de barras utilizados para secuenciar las muestras; otra característica, es que cuentan ya con la calidad deseada, que es también personalizada desde la misma plataforma; para este caso, se estableció un puntaje de calidad de 25 (Phred Quality Score).

El puntaje de calidad es la métrica más común utilizada para evaluar la exactitud de un secuenciador, ya que indica la probabilidad de que una base dada sea nombrada de manera incorrecta, para lo cual se han establecido una serie de puntajes estándar, los cuales se muestran en la tabla 8.

Phred Quality Score	Probabilidad de base errónea	Exactitud de llamada de base
10	1 en 10	90%
20	1 en 100	99%
30	1 en 1,000	99.9%
40	1 en 10,000	99.99%
50	1 en 100,000	99.999%

Tabla 8.- Puntajes de calidad estándar.

Fuente: Tomado de Quality Scores for Next Generation Sequencing.

De acuerdo con los puntajes mostrados en la tabla 8, el puntaje de calidad que fue establecido para las secuencias de Ion Torrent, indica que más del 99% de las lecturas fueron nombradas correctamente por el secuenciador.

Como primer paso para trabajar con las secuencias de Ion Torrent, es necesario convertir las lecturas al formato fasta, que comienza con una descripción de la muestra en una sola línea, seguida por líneas de datos de secuencia (BLAST Topics); este es compatible con el programa QIIME, ya que es una extensión más sencilla que el formato original que arroja la plataforma Ion

Torrent; el cual se encuentra en extensión fastq, que contiene tanto la secuencia como un puntaje de calidad asociado por base (Cock et al., 2009).

El comando utilizado para convertir los archivos a extensión fasta es el siguiente:

```
awk 'BEGIN{P=1}{if(P==1||P==2){gsub(/^[@]/,">");print}; if(P==4)P=0; P++}' [path working directory replica 2 Ion Torrent] > replica2.fasta
```

Una vez que los archivos de las tres réplicas del pozo fueron convertidos al formato fasta, se necesita correr un programa escrito en lenguaje perl, para formatear las lecturas de acuerdo a como las utiliza QIIME. El comando utilizado para formatear las lecturas es el siguiente:

```
perl QIIMEfastaFormatter.pl -i fastas/
```

El resultado que arroja el comando anterior son las secuencias ya procesadas, con el número de lecturas contenidas en cada una, y en formato fasta, para así poder ser analizadas con QIIME; ya que las secuencias fueron formateadas por separado, antes de proceder con los posteriores análisis, se deben concatenar para trabajar con un solo archivo, el comando utilizado para juntar los archivos es el siguiente:

```
cat replica2.fna replica3.fna replica4.fna > seqs.fna
```

Una vez que se obtuvo el archivo que contiene las secuencias de las tres réplicas del pozo, el siguiente paso fue generar dos archivos con las secuencias; un archivo eliminando las lecturas menores a 72 pares de bases y el otro archivo, eliminando las lecturas menores a 120 pares de bases; con la finalidad de conservar las secuencias con las que se obtenga una mejor estimación de la diversidad en las muestras.

El comando utilizado para cortar las lecturas proviene de una barra de herramientas bioinformáticas llamada FASTX-Toolkit y es el siguiente:

```
fastx_clipper -l 72 -i [pwd seqs.fna] -o seqs_72.fna
```

```
fastx_clipper -l 120 -i [pwd seqs.fna] -o seqs_120.fna
```

Una vez que fueron cortadas las lecturas menores a 72 pb y 120 pb, las secuencias fueron sometidas a un análisis previo con QIIME y fueron realizadas distintas pruebas para conocer a que longitud resulta una mejor estimación de la diversidad presente en las tres muestras del pozo.

Las pruebas realizadas a las secuencias consistieron en el número de lecturas obtenidas con los dos cortes, ya que entre menor sea el número de lecturas que se pierden, es más probable encontrar una mejor estimación de la diversidad presente en las muestras; así como también, fue estimado el número de OTUs obtenidos, los cuales son grupos de secuencias similares agrupados a los niveles taxonómicos particulares de interés (Bikel *et al.*, 2015); OTUs se refiere a las unidades taxonómicas operacionales, por sus siglas en inglés (Operational Taxonomic Units); fue calculada la diversidad alfa, como los índices de Shannon, Simpson, de diversidad filogenética (PD) y el número de OTUs estimados; por último se realizaron diagramas de Venn de los géneros y especies compartidas.

Después de realizar las pruebas con las secuencias cuyas lecturas menores a 72 pb y 120 pb fueron removidas, y una vez seleccionadas las más óptimas para el análisis de las tres réplicas del pozo, el

siguiente paso, antes de realizar el análisis final con QIIME, fue comparar entre tres distintas bases de datos para conocer cuál de las tres provee una mejor información en cuanto a taxonomía y diversidad para el análisis del presente trabajo.

Las bases de datos comparadas fueron Greengenes (DeSantis *et al.*, 2006), la base de datos de Ribosomal Database Project (RDP) (Cole *et al.*, 2013) y Silva (Quast *et al.*, 2013). El programa QIIME utiliza por defecto la base de datos Greengenes, por lo que las bases de datos RDP y Silva tuvieron que ser descargadas.

El análisis con las tres bases de datos fue realizado al archivo *fna* previamente generado y para seleccionar la base de datos ganadora se tomaron en cuenta los siguientes aspectos: el número de lecturas, así como el número de OTUs obtenidos con cada una de las bases de datos, las métricas de diversidad alfa como los índices de Shannon cuya ecuación es la siguiente: $(H' = -\sum_{i=1}^S p_i \log_2 p_i)$, donde S representa el número de especies, p_i representa la abundancia relativa de la especie i representada por $\frac{n_i}{N}$, donde n_i es el número de individuos de la especie i y N el número de individuos de todas las especies (Shannon and Weaver, 1949), el índice de Simpson cuya ecuación es la siguiente: $(D = \frac{\sum_{i=1}^S n_i(n_i-1)}{N(N-1)})$, donde S representa al número de especies, N es el total de organismos presentes y n es el número de ejemplares por especie (Simpson E. H., 1949), y por último, el índice de diversidad filogenética (PD) y el número de OTUs estimados. Por último, se obtuvieron diagramas de Venn de los géneros y especies compartidas entre las tres bases de datos.

Una vez que fue seleccionada la base de datos ganadora, el siguiente paso fue realizar el análisis de diversidad y taxonomía con el programa QIIME (Quantitative Insights Into Microbial Ecology), que es un software bioinformático diseñado para el análisis de comunidades microbianas, el cual se basa en los datos de secuenciación de ADN.

El primer paso del análisis con QIIME consiste en la formación de OTUs, los cuales están agrupados de acuerdo con su similitud de secuencias; dichas secuencias se agrupan de acuerdo con un umbral definido por el usuario, el umbral es representado mediante un porcentaje (Navas-Molina *et al.*, 2013). El umbral definido para la asignación de OTUs utilizado en el presente trabajo fue de un 97% de similitud entre secuencias, para asignar taxonomía hasta nivel de especie.

Los distintos análisis que pueden ser realizados con QIIME para formar OTUs se resumen en la figura 3.

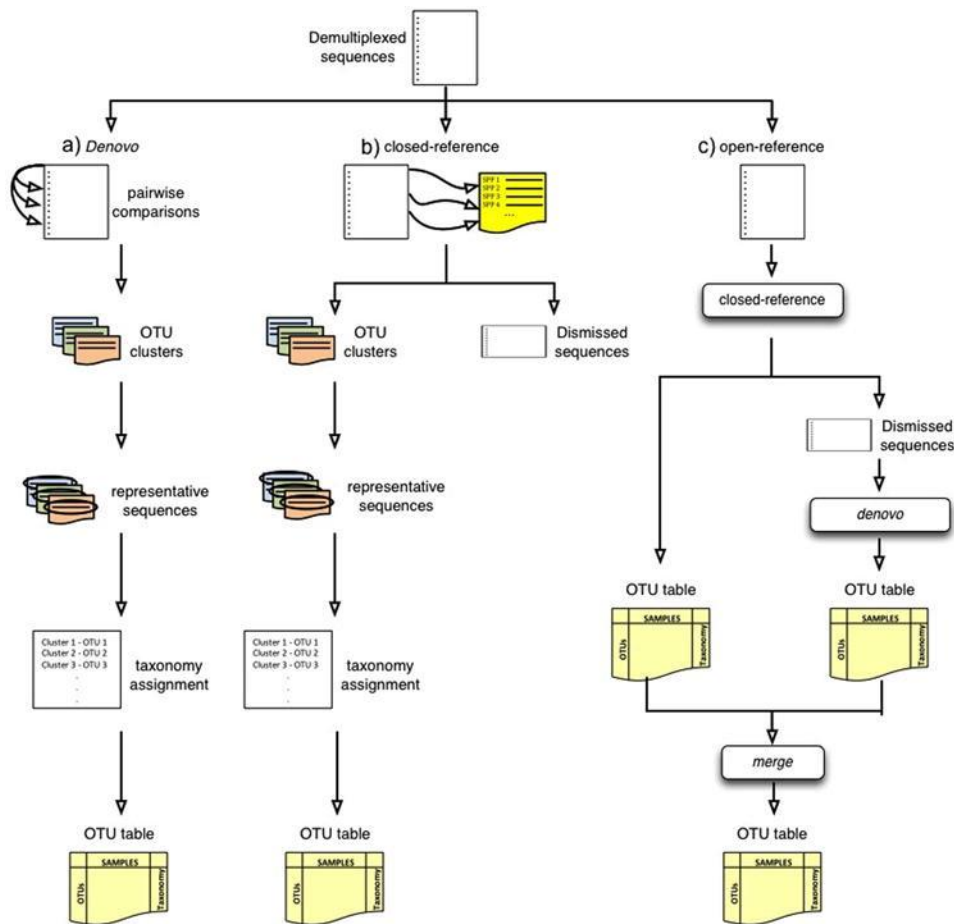


Figura 3.- Representación de los métodos utilizados por QIIME para la formación de OTUs.

Fuente: Tomado de Navas-Molina *et al.*, 2013.

El método utilizado para formar OTUs con las secuencias obtenidas de Ion Torrent fue el análisis de referencia ‘cerrada’, ya que, las secuencias obtenidas con Ion Torrent provienen de dos distintos grupos de amplicones. Una vez que se seleccionó el método de formación de OTUs, fue realizado el análisis de diversidad y taxonomía para conocer las bacterias presentes en las muestras de las tres réplicas del pozo.

En el presente trabajo, utilizando la tecnología de Ion Torrent, fueron secuenciadas siete de las nueve regiones hipervariables del gen rRNA 16S, lo cual abarca la mayoría de las regiones del gen, en comparación con las que son secuenciadas con la tecnología de MiSeq; con la finalidad de realizar una comparación entre tecnologías de secuenciación, además de haber realizado el análisis de las secuencias de Ion Torrent tomando en cuenta las siete regiones hipervariables, fue realizado un análisis tomando en cuenta únicamente las regiones V3 y V4, que son las mismas analizadas por MiSeq.

Para poder analizar únicamente las regiones V3 y V4 con las secuencias de Ion Torrent, fue preciso realizar un corte a la base de datos ganadora, de tal manera que abarcara solamente las regiones hipervariables V3 y V4, y posteriormente someter las secuencias de Ion Torrent a un nuevo análisis con la base de datos cortada con la finalidad de poder comparar la diversidad entre tecnologías de

secuenciación; por otro lado, también fue comparada la diversidad entre Ion Torrent al tomar en cuenta los resultados de secuenciación de las siete regiones hipervariables contra los resultados de las dos regiones (V3-V4).

Para corroborar que el corte realizado a la base de datos ganadora abarcara efectivamente las regiones V3-V4, las secuencias obtenidas con MiSeq fueron también analizadas tomando como referencia la base de datos cortada, comparando los resultados obtenidos, fueron realizados diagramas de Venn de los géneros y especies compartidas, así como también el número de OTUs obtenidos cuando se toman como referencia las bases de datos con y sin cortar.

5.4.2 Secuencias MiSeq

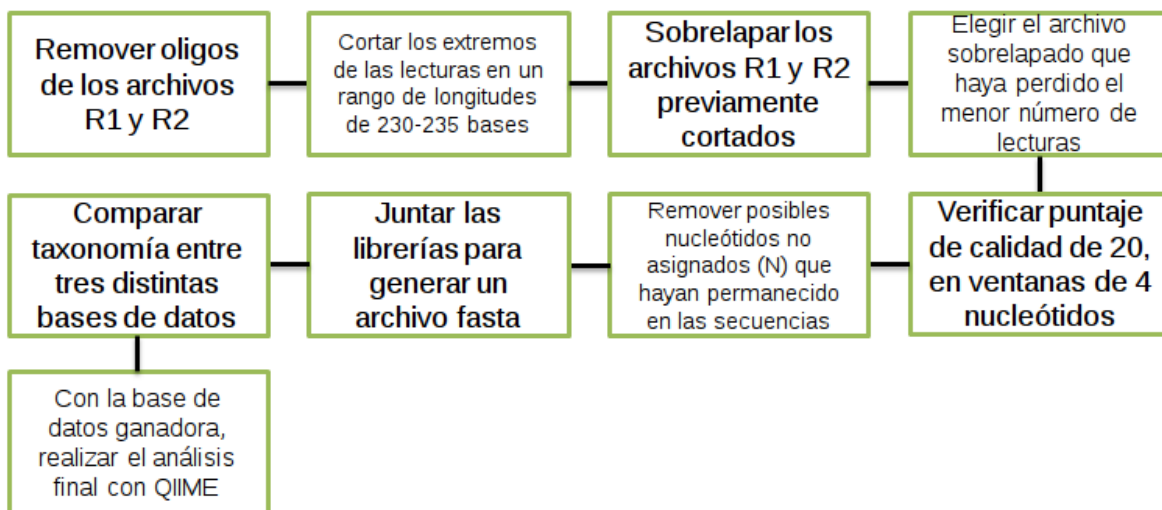


Figura 4. - Esquema general del método utilizado para filtrar por calidad las secuencias obtenidas con MiSeq.

Una vez obtenidas las secuencias de las tres réplicas del pozo y de los cultivos bacterianos que fueron secuenciados con MiSeq, estas fueron sometidas a distintos filtros de calidad antes de realizar el análisis final con el programa QIIME. Se obtuvieron dos archivos para cada una de las muestras que fueron secuenciadas nombrados como R1 y R2, los cuales indican la dirección de la secuenciación; el archivo R1 indica las secuencias obtenidas en dirección 'forward' y el archivo R2 indica las secuencias en dirección 'reverse'.

Al salir del secuenciador, el formato de los archivos R1 y R2 se encuentra en fastq, el cual es compatible con el programa FastQC (Andrews, 2010); en dicho programa se realizó un análisis de las secuencias, tomando en cuenta los siguientes aspectos: calidad por base de las secuencias y el contenido de bases que no pudieron ser nombradas por el secuenciador (N), las cuales pueden representar cualquier nucleótido.

Analizando las secuencias en el programa FastQC, se observó que el gráfico que indica el contenido de N mostró que las bases que no pudieron ser nombradas correctamente por el secuenciador se encontraban en las primeras 5 y 6 bases de las lecturas R1 y R2, respectivamente, por lo que, dichas bases fueron removidas mediante el programa Trimmomatic (Bolger *et al.*, 2014).

En el programa fue también observado el cromatograma de las secuencias, el cual mostró que al inicio de las lecturas se encontraban restos de los oligonucleótidos 'forward' y 'reverse' en los archivos R1 y R2, respectivamente, por lo que junto con las bases no nombradas (N) fueron también removidos utilizando el programa Trimmomatic.

El propósito de remover los restos de oligonucleótidos y los nucleótidos asignados como N que fueron encontrados al inicio de las secuencias se realizó debido a que al ser fragmentos de secuencias repetitivas, interferían con los resultados de diversidad y taxonomía según los análisis previos realizados con el programa QIIME, los cuales arrojaron una menor asignación de taxonomía e índices más bajos de diversidad en comparación con los resultados obtenidos con las secuencias a las que les fueron removidos los fragmentos de oligonucleótidos y de N.

El comando utilizado para remover los restos de oligos en las lecturas y las N es el siguiente:

```
TrimmomaticSE -phred33 [pwd archivo_R1.fastq] [archivo_R1_trimmed.fastq] HEADCROP:17
```

```
TrimmomaticSE -phred33 [pwd archive_R2.fastq] [archivo_R2_trimmed.fastq] HEADCROP:21
```

Una vez que fueron removidos los oligos y las N presentes en las lecturas, el siguiente paso es realizar cortes en los extremos de las lecturas en un rango definido; la finalidad de realizar dichos cortes es debido a que normalmente la calidad de las secuencias va decayendo conforme aumentan los ciclos de secuenciación, dejando las lecturas de menos calidad en los extremos de los archivos, otra de las razones para realizar los cortes a las lecturas fue para que al sobrelapar los archivos de las secuencias, el tamaño aproximado sea de 460 pares de bases, que es el tamaño cercano al de las regiones hipervariables V3-V4 del gen 16S.

Los extremos de las lecturas fueron cortados en un rango de longitud de 230-235 bases, haciendo cortes de uno en uno a las secuencias de ambos archivos, generando archivos cortados con las longitudes antes mencionadas, por lo que fue utilizado nuevamente el programa Trimmomatic. El comando utilizado para realizar los cortes en las secuencias fue el siguiente:

```
TrimmomaticSE -phred33 [pwd archivo_R1_trimmed.fastq] [archivo_R1_230-235.fastq]  
CROP:230-235
```

```
TrimmomaticSE -phred33 [pwd archivo_R2_trimmed.fastq] [archivo_R2_230-235.fastq]  
CROP:230-235
```

Cada uno de los archivos que fueron previamente cortados (R1 y R2) fueron sobrelapados para realizar los filtros de calidad posteriores sobre un solo archivo; al sobrelapar los archivos R1 y R2 con los respectivos cortes, fue elegido el que haya perdido el menor número de lecturas posible, ya que entre más lecturas se pierdan de las secuencias, puede perderse también cierto porcentaje de la diversidad presente en las muestras. Por lo que, además de aplicar filtros de calidad en las secuencias, lo que se busca es que con cada uno de los filtros se pierda la menor cantidad de lecturas posible en comparación con las que se obtuvieron directamente del secuenciador.

Para sobrelapar los archivos R1 y R2 cortados, fue utilizada una herramienta bioinformática llamada fastq-join; el comando utilizado es el siguiente:

```
fastq-join -p30 [archivo_R1_230.fastq] [archivo_R2_230.fastq] -o [archivo_pozo.fastq]
```

Con los archivos de las tres réplicas del pozo y de los cultivos bacterianos sobrelapados, el siguiente paso fue establecer un puntaje que asegurara la calidad de las secuencias, dicho puntaje establecido fue de 20, el cual de acuerdo con la tabla 8 asegura que el 99% de las lecturas fueron nombradas correctamente por el secuenciador. El puntaje de calidad establecido fue verificado utilizando nuevamente el programa Trimmomatic.

Verificar calidad de las secuencias mediante ventanas consiste en dar la orden al programa de ir recorriendo las lecturas por cada cuatro nucleótidos, dentro de esta ventana de nucleótidos establecida, se debe promediar el puntaje de 20; de no ser así, cuando la ventana no alcanza el promedio, la lectura es cortada y avanza a la siguiente línea de la secuencia.

El script utilizado para verificar calidad es el siguiente:

```
TrimmomaticSE -phred33 [pwd archivo_pozo.fastq] [archivo_pozo_ventana.fastq]  
SLIDINGWINDOW:4:20
```

Una vez que fue verificada la calidad de las secuencias, el último filtro aplicado consiste en remover los posibles nucleótidos no asignados (N) que pudieran encontrarse a lo largo de las lecturas; para remover las N's, fue utilizado un comando de la herramienta bioinformática FASTX-Toolkit, el comando utilizado es el siguiente:

```
fastx_clipper -i [archivo_pozo_ventana.fastq] -Q33 -C -M 45 -v -o  
[archivo_pozo_clipper.fastq]
```

Antes de realizar los análisis de taxonomía, se deben juntar los archivos de la secuenciación de las tres réplicas del pozo y de los cultivos bacterianos que previamente fueron filtrados y que cuentan con una buena calidad, además de un número considerable de lecturas; los archivos que han sido hasta el momento filtrados se encuentran en formato fastq, por lo que el siguiente paso es darles el formato de archivo fasta, el cual es compatible con QIIME.

Para lograr juntar las librerías en un solo archivo con extensión fasta, fue utilizado el programa QIIME, al cual se le debe proporcionar la dirección de cada uno de los archivos que contienen las secuencias y un nombre para que puedan ser identificadas, el cual será la etiqueta de las muestras en los análisis siguientes; el script utilizado para crear el archivo fasta es el siguiente:

```
split_libraries_fastq.py -i [pwd secuencias_limpias] -o seqs/ --sample_ids [nombres separados  
por comas] -barcode_type 'not-barcoded'
```

Una vez que fue creado el archivo fasta que contiene las secuencias de las réplicas del pozo y de los cultivos bacterianos; antes de realizar el análisis final con QIIME, fueron comparadas tres distintas bases de datos, con la finalidad de seleccionar la más óptima para los análisis de taxonomía y diversidad de las muestras.

Las bases de datos comparadas fueron Greengenes, la base de datos de Ribosomal Database Project (RDP) y Silva; el programa QIIME utiliza por defecto la base de datos Greengenes, por lo que las bases de datos RDP y Silva tuvieron que ser descargadas.

El análisis con las tres bases de datos fue realizado al archivo fna previamente generado y para seleccionar la base de datos ganadora se tomaron en cuenta los siguientes aspectos: el número de lecturas, así como el número de OTUs obtenidos con cada una de las bases de datos, las métricas

de diversidad alfa como los índices de Shannon, Simpson, de diversidad filogenética y el número de OTUs estimados y por último, diagramas de Venn de los géneros y especies compartidas entre las tres bases de datos.

Una vez que fue seleccionada la base de datos ganadora, el último paso es realizar el análisis de las muestras con QIIME; de acuerdo con los distintos métodos de selección de OTUs que se muestran en la figura 3, el método utilizado para analizar las secuencias obtenidas con MiSeq es el de referencia cerrada.

6 Resultados y discusión

6.1 Resultados experimentales

En las tablas 9 y 10 se muestran las cantidades de ADN que fueron extraídas del agua de las tres réplicas del pozo, la tabla 9 muestra el ADN extraído en la primera elución, el cual fue utilizado para construir los amplicones que fueron secuenciados con Ion Torrent y MiSeq, el ADN de la primera elución fue cuantificado utilizando Qubit; el ADN cuantificado por este equipo no permite conocer la pureza del material genético extraído mediante los valores de los radios 260/280 y 260/230, por lo que no se tiene el dato de los valores de pureza obtenidos con la primera elución.

Muestra	ng/ μ l	ng totales
Réplica 2	15.6	312
Réplica 3	23.8	476
Réplica 4	23.6	472

Tabla 9. - Cantidad de ADN extraído de las tres réplicas del pozo, primera elución.

El ADN de la segunda elución, mostrado en la tabla 10 fue cuantificado mediante un NanoDrop, con lo cual se pueden observar los radios 260/280 y 260/230 de cada una de las tres réplicas del pozo; los radios 260/280 de las tres réplicas se encuentran dentro del rango de ADN de pureza aceptable, mientras que los radios 260/230 pueden indicar presencia de ciertos contaminantes en la muestra; por lo que, como fue mencionado anteriormente, se tomaron las extracciones de ADN de la primera elución para construir los amplicones, ya que además, se obtuvo una mayor cantidad.

Muestra	ng/ μ l	Radio 260/280	Radio 260/230
Réplica 2	0.6	1.10	0.02
Réplica 3	5.3	1.71	0.03
Réplica 4	3.2	2.07	0.02

Tabla 10. - Cantidad de ADN extraído de las tres réplicas del pozo, segunda elución.

En la tabla 11 se muestran las cantidades de ADN extraído de los cultivos específicos para bacterias sulfato reductoras y ácido productoras. De los cultivos realizados con una porción de membrana del pozo de Tabasco, únicamente mostró crecimiento el cultivo para bacterias ácido productoras, debido al viraje de color en el medio; característica que no presentó el cultivo de bacterias sulfato reductoras. Por otro lado, las bacterias que fueron inoculadas con agua de un pozo distinto al muestreado en el presente trabajo presentaron crecimiento para los dos tipos de cultivos, por lo cual, fue extraído ADN de bacterias ácido productoras de membrana del pozo de

Tabasco y del agua de pozo de Yucatán, como también de bacterias sulfato reductoras provenientes de la misma agua.

Se muestra en la tabla 11 el ADN que fue cuantificado utilizando Qubit, tanto de la primera como segunda elución; la muestra de membrana del pozo de Tabasco que presentó crecimiento fue nombrada como BPA-M y las muestras provenientes del agua de pozo de Yucatán fueron etiquetadas como BPA y BSR para el ADN extraído de los cultivos de bacterias ácido productoras y sulfato reductoras, respectivamente.

Muestra	ng/ μ l	ng totales
BPA-M 1° elución	1.97	39.4
BPA 1° elución	15.4	308
BSR 1° elución	19.17	383.4
BPA-M 2° elución	0.53	159
BPA 2° elución	0.126	37.8
BSR 2° elución	0.67	201

Tabla 11.- Cantidad de ADN extraído de los cultivos para bacterias sulfato reductoras y ácido productoras.

Una vez que fue extraído el ADN de las tres réplicas del pozo y de los cultivos bacterianos, el siguiente paso fue construir los amplicones de las regiones específicas de las dos tecnologías de secuenciación a comparar (Ion Torrent y MiSeq), en las figuras 5, 6 y 7 se muestran las fotografías de los amplicones del gen rRNA 16S.

La figura 5 muestra la fotografía del gel de agarosa en donde se indica la presencia de los amplicones de las siete regiones hipervariables, construidos con los dos grupos de oligonucleótidos denominados como V2, el cual abarca las regiones V2, V4 y V8; y el oligonucleótido V3 que abarca las regiones V3, V6, V7 y V9. Las muestras se encuentran etiquetadas como 2/5, 3/5 y 4/5 las cuales corresponden a los amplicones obtenidos para la réplica 2, réplica 3 y réplica 4, respectivamente, fue corroborada la presencia de los amplicones de las tres réplicas del pozo por el tamaño aproximado que debe ser de 200-300 pares de bases de acuerdo con el protocolo.

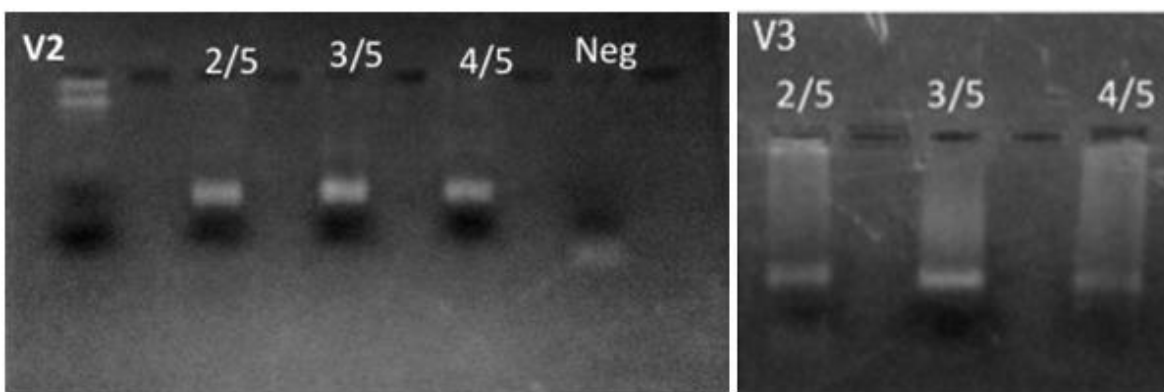


Figura 5.- Amplicones de las tres réplicas del pozo, secuenciados con Ion Torrent.

En la figura 6 se muestran los amplicones de las regiones hipervariables V3-V4, obtenidos de las tres réplicas del pozo que fueron secuenciados con MiSeq, junto con su control negativo, las muestras se encuentran etiquetadas como 2/5, 3/5 y 4/5 que corresponden a los amplicones de la

réplica 2, réplica 3 y réplica 4, respectivamente; fue corroborada la presencia de los amplicones, cuyo tamaño aproximado debe encontrarse entre los 500-550 pares de bases de acuerdo con el protocolo.

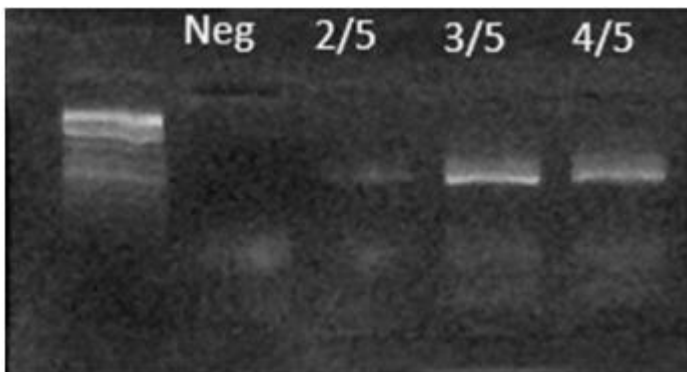


Figura 6. - Amplicones de las tres réplicas del pozo, secuenciados con MiSeq.

En la figura 7 se observa la fotografía del gel de agarosa donde se corroboró la presencia de los amplicones de las regiones hipervariables V3-V4, construidos a partir de los cultivos específicos para bacterias sulfato reductoras y ácido productoras de los pozos de Tabasco y Yucatán. Los cultivos pertenecientes a la porción de membrana del pozo de Tabasco fueron etiquetados como BPA-M para las bacterias ácido productoras, los números 300 y 20 corresponden a las eluciones en la extracción de ADN, donde 300 fue la segunda elución y 20, la primera. Los cultivos pertenecientes a las bacterias ácido productoras y sulfato reductoras del pozo de Yucatán fueron etiquetados como BPA y BSR, respectivamente, cuyos amplicones fueron construidos con la primera elución de ADN. De acuerdo con el protocolo de Illumina, los amplicones resultantes de las regiones V3-V4 deben tener un tamaño aproximado de 500-550 pares de bases.

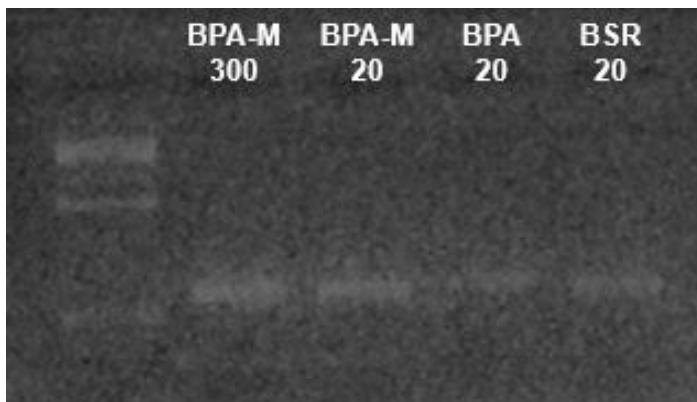


Figura 7. - Amplicones de los cultivos de bacterias sulfato reductoras y ácido productoras de los pozos de Tabasco y Yucatán, secuenciados con MiSeq.

Después de haber obtenido los amplicones de las muestras del pozo y de los cultivos para bacterias sulfato reductoras y ácido productoras, estos fueron purificados y tratados mediante una serie de pasos subsecuentes de acuerdo con lo establecido en los protocolos de las dos distintas tecnologías de secuenciación con la finalidad de obtener las secuencias de las muestras para lograr el análisis de diversidad presente en el pozo petrolero analizado en este trabajo, así como también

la comparación entre Ion Torrent y MiSeq. Antes de poder trabajar con las secuencias, fue preciso llevar a cabo una serie de pasos que aseguraran una buena calidad de las lecturas.

Por otro lado, se presentan en la figura 8 los resultados de las pruebas de diluciones que fueron realizadas a las muestras de agua para corroborar la presencia de microorganismos de tipo sulfato reductores y ácido productores dentro del pozo, de acuerdo con las diluciones seriadas realizadas a las muestras, es posible observar que existe un viraje de color en los dos primeros viales, lo cual indica la presencia de las bacterias antes mencionadas.

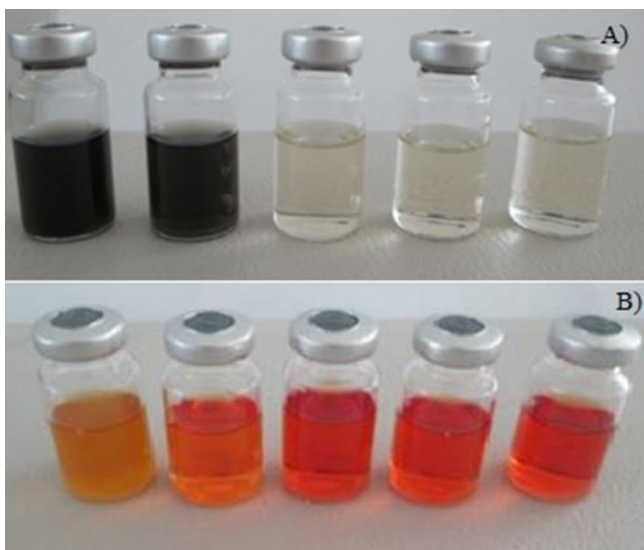


Figura 8.- Pruebas de diluciones de la actividad de A) bacterias sulfato reductoras y B) bacterias ácido productoras

En la tabla 12 se presentan los resultados de las diluciones realizadas para indicar la presencia de bacterias del tipo sulfato reductoras y ácido productoras, este tipo de diluciones se realizan como método de diagnóstico para conocer el estado en el que se encuentra el pozo de acuerdo con las UFC encontradas en las diluciones.

Como se indica en la tabla, la presencia de las bacterias de tipo sulfato reductor y ácido productor, tienen resultados negativos dentro del pozo, ya que puede ocurrir que debido a las actividades metabólicas que están llevando a cabo resulten en problemas que causen daños corrosivos en las tuberías; cabe resaltar que este método resulta ser un diagnóstico rápido para determinar la severidad del daño dentro de las tuberías basado en la presencia de los dos tipos de bacterias mejor estudiados que se sabe son los principales causantes de la corrosión, los resultados obtenidos de la secuenciación del gen ribosomal 16S revelarán con mayor detalle cuáles son las bacterias causantes del daño a las tuberías.

Número de botella	Dilución de la muestra	UFC/ml	Daño BSR	Daño BAP
1	1:10	10	Bajo	Bajo
2	1:100	100	Significante	Bajo
3	1:1,000	1,000	Muy significativa	Moderado
4	1:10,000	100,000	Severo	Significante
5	1:100,000	1,000,000	Muy severo	Severo

Tabla 12.- Tabla de diagnóstico de riesgo de corrosión mediante diluciones seriadas.

6.2 Filtrado por calidad de las secuencias

6.2.1 Ion Torrent

El primer paso para analizar las secuencias obtenidas con Ion Torrent fue el de dar formato a las lecturas para poder ser analizadas con QIIME, lo cual fue logrado mediante el uso de herramientas bioinformáticas, como el uso del programa perl y FASTX-Toolkit. Una vez que fue dado dicho formato requerido, se realizaron pruebas eliminando las lecturas de dos distintas longitudes, las lecturas menores a 72 pares de bases y 120 pares de bases, a continuación, se muestran los resultados de las pruebas realizadas a las secuencias para seleccionar la longitud adecuada.

En la figura 9 se muestra el histograma de la réplica 2, en donde se representan los distintos tamaños en los que se encuentran las secuencias obtenidas con Ion Torrent, de acuerdo con la gráfica, en el eje 'x' se muestra la longitud de la secuencia en pares de bases y en el eje 'y' se muestra el número de lecturas; las gráficas de las réplicas 3 y 4 presentan un patrón similar de comportamiento en cuanto a las longitudes de sus lecturas.

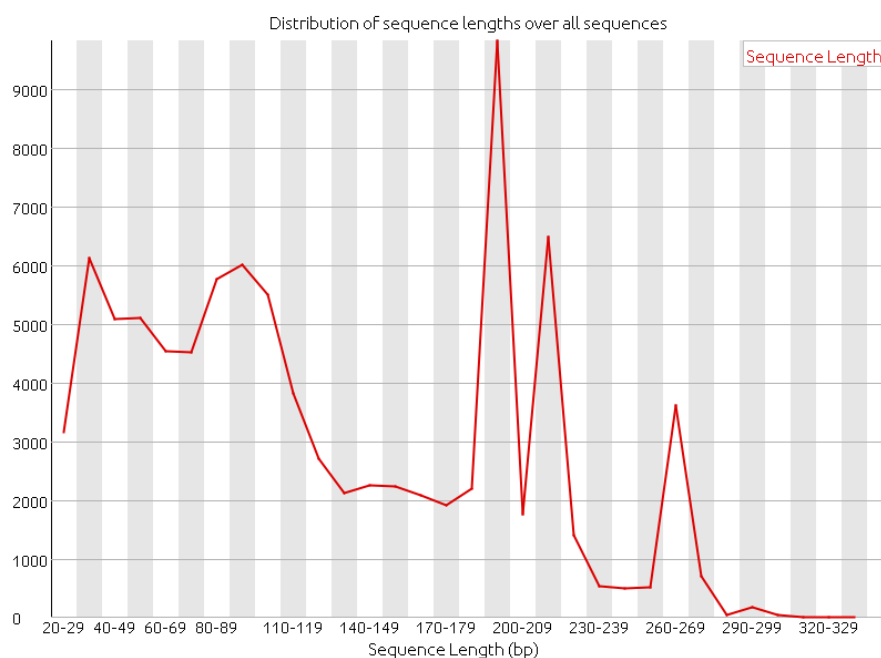


Figura 9.- Histograma de la distribución de longitudes de la réplica 2.

Uno de los objetivos principales de eliminar lecturas menores a 72 pb y 120, además de quitar lecturas de las secuencias que puedan no asignar a ningún OTU debido a su corto tamaño, fue el de determinar con cuál de las secuencias cortadas se pierde el menor número de lecturas posible, por lo que la primera prueba realizada fue la de conocer cuántas lecturas se conservan después de los cortes a las secuencias, dichos resultados se muestran en la tabla 13.

La tabla 13 muestra el número de lecturas crudas de cada una de las tres réplicas del pozo, es decir, el número de lecturas que se obtienen directamente al salir del secuenciador y las lecturas que se conservan después de realizar los dos cortes a las secuencias, a pesar de que con los dos cortes se pierde un gran número de lecturas, se puede observar que si son eliminadas las lecturas menores a 72 pares de bases, se pierden únicamente entre un 35 a 38% en comparación con las iniciales; mientras que al quitar las menores a 120 pares de bases, se pierden alrededor del 61 al 63% de las lecturas, eliminando más de la mitad, las cuales pudieran ser de gran utilidad al momento de asignar taxonomía, perdiendo entonces la mayor diversidad en las muestras.

De acuerdo con el número de lecturas que se obtienen con los dos distintos cortes, es evidente que la pérdida es mucho menor cuando son removidas las lecturas menores a 72 pares de bases, por lo que, hasta el momento, el corte de 72 pb resulta ser el mejor de los dos.

Muestra	Lecturas crudas	72 pb	120 pb
Réplica 2	90897	59152	34747
Réplica 3	81989	52444	30448
Réplica 4	112258	69775	40736

Tabla 13.- Número de lecturas obtenidas con cada uno de los cortes a las secuencias.

La siguiente prueba realizada a las secuencias consistió en determinar con cuál de los dos cortes se obtiene un mayor número de OTUs asignados a los seis niveles taxonómicos, desde phylum hasta especie; para fines prácticos del presente trabajo, las subsecuentes pruebas realizadas a las secuencias se realizaron en una sola de las librerías secuenciadas, siendo seleccionada la réplica 2 del pozo.

En la tabla 14 se muestra el número de OTUs asignados a los distintos niveles taxonómicos que se obtienen con los dos cortes realizados a las secuencias de la réplica 2 del pozo; como se puede observar, cuando son removidas las lecturas menores a 72 pares de bases se consigue un mayor número de OTUs en los distintos niveles taxonómicos, asignando casi el doble de los que se asignan cuando se eliminan las secuencias menores a 120 pares de bases, esto puede deberse al número de lecturas que se conservan con cada corte, ya que como lo indica la tabla 13, las secuencias cortadas a 72 pb conservan la mayor parte de las lecturas, lo que puede verse reflejado en la asignación de taxonomía.

Nuevamente las secuencias cuyas lecturas menores a 72 pares de bases fueron eliminadas, muestran una mejor asignación de taxonomía, siendo mejores dichas secuencias en la segunda prueba realizada.

Corte	Phylum	Clase	Orden	Familia	Género	Especie
72 pb	16	32	62	87	86	37
120 pb	7	11	24	30	30	9

Tabla 14.- Número de OTUs obtenidos a los distintos niveles taxonómicos con cada uno de los cortes a las secuencias.

En la tabla 15 se muestran los índices de diversidad alfa, los cuales determinan la riqueza de las especies que se encuentran presentes en la réplica 2 del pozo, ya que se desea conocer con qué corte se logra abarcar una mayor diversidad. Como primer paso y con la finalidad de que los índices obtenidos fueran comparables entre cortes, fue tomado el mismo número de secuencias por muestra. Como se muestra en la tabla, las secuencias que fueron cortadas eliminando las lecturas menores a 72 pares de bases obtienen mayores índices de diversidad, en comparación con los índices obtenidos para las secuencias con el corte a 120 pares de bases.

Además de los índices de diversidad alfa que se muestran en la tabla, fue estimado el número de OTUs que se encuentran en la muestra analizada; por lo que, como era de esperarse, el programa estimo la presencia de un mayor número de OTUs en las secuencias cuyo corte removió las lecturas menores a 72 pares de bases. Se puede observar también que el número de OTUs presentados en la tabla 15 resulta ser distinto a los OTUs que se presentan en la tabla 14, esto se debe a que los resultados de los OTUs estimados en la tabla 15 corresponden a la totalidad de OTUs que fueron encontrados por el programa, sin diferenciar a qué nivel taxonómico se encuentren asignados, mientras que en la tabla 14 se observa el número de OTUs que fueron asignados de phylum a especie.

Las secuencias a las cuales les fueron removidas las lecturas menores de 120 pares de bases muestran una menor diversidad, lo cual puede deberse a la pérdida de lecturas al realizar los cortes, pudiendo indicar que entre mayor sea el número de lecturas perdidas, es más probable perder una parte significativa de diversidad en la muestra, por lo que es importante mantener un número de lecturas adecuado.

Muestra	Secuencias/Muestra	PD (Diversidad Filogenética)	Shannon	Simpson	OTUs observados
Rep. 2 72 pb	6810	7.345	4.845	0.863	421
Rep. 2 120 pb	6810	3.295	3.011	0.685	132

Tabla 15. - Índices de diversidad alfa obtenidos con cada uno de los cortes a las secuencias.

Antes de decidir con que secuencias se realizaran los análisis de diversidad y taxonomía, se tomaron en cuenta el número de OTUs asignados a nivel de género y especie que se comparten entre los dos cortes, para conocer con que corte se obtiene una mayor variedad en cuanto a asignación de taxonomía.

En la figura 10 se muestra el diagrama de Venn comparando la diversidad de OTUs asignados a nivel de género y especie que son compartidos entre las dos secuencias; como puede observarse, del total de OTUs que pudieron ser asignados cuando se removieron las lecturas menores a 120 pares de bases, se encuentran todos compartidos con el número de OTUs asignados con las secuencias del corte de 72 pares de bases, cuyas secuencias, además del número de OTUs

compartidos, lograron asignar otros distintos, abarcando, por lo tanto, la mayor diversidad presente en la réplica 2 del pozo.

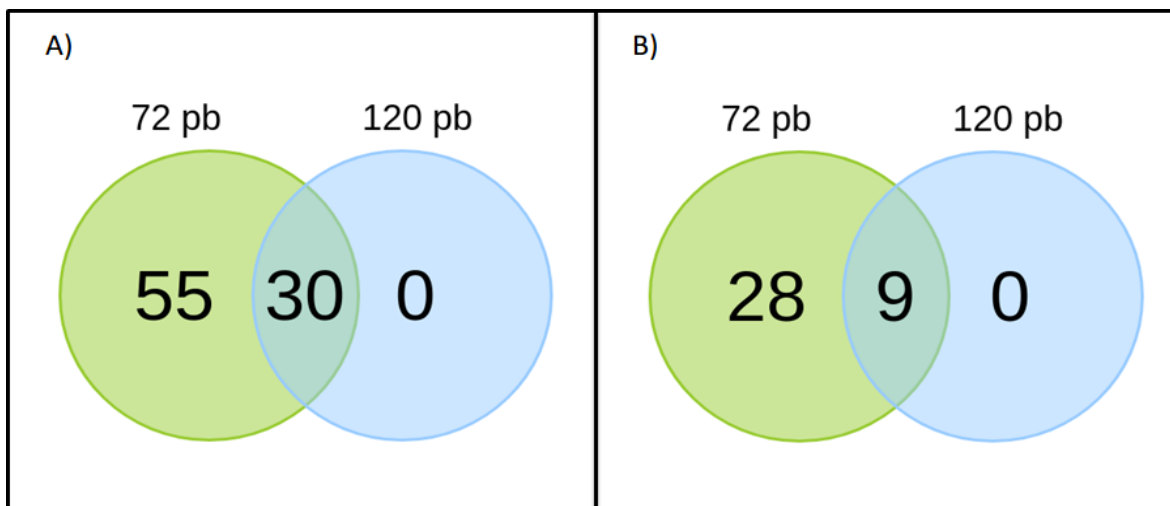


Figura 10.- Diagrama de Venn de A) géneros y B) especies compartidos entre los dos cortes realizados a las secuencias.

De acuerdo con las pruebas que fueron realizadas a las secuencias y con los resultados que se obtuvieron, resulta mejor para conocer la diversidad obtenida con la tecnología de secuenciación de Ion Torrent, eliminar de las secuencias las lecturas cuya longitud se encuentre debajo de los 72 pares de bases, ya que así se puede observar una mayor estimación de la diversidad presente en las muestras del pozo, en comparación con los resultados obtenidos cuando son removidas las lecturas menores a 120 pares de bases, por lo que los análisis definitivos serán con las secuencias a las que se les removieron las lecturas menores a 72 pares de bases.

6.2.2 MiSeq

Al trabajar con las secuencias obtenidas con MiSeq, se realizaron una serie de pasos para asegurar la calidad de las mismas, utilizando para ello diversos programas bioinformáticos como FastQC, Trimmomatic, herramientas de FASTX-Toolkit y pruebas previas con QIIME, esto con la finalidad de depurar las secuencias y, por lo tanto, generar resultados más confiables.

El primer paso para poder trabajar con las secuencias de MiSeq fue el de conocer la calidad de las lecturas mediante un análisis con el programa FastQC; la figura 11 muestra las gráficas de calidad por base de la librería ACR 2-5, tanto del archivo R1 (A) como R2 (B). En el eje 'x' de la gráfica se muestra la longitud de las secuencias en pares de bases, mientras que en el eje 'y' se observan los distintos puntajes de calidad los cuales van en un rango desde 0-30; el fondo del gráfico se divide en colores de acuerdo al rango de calidad con el que cuentan las secuencias, siendo el color verde para una muy buena calidad, el color amarillo para indicar una calidad aceptable y el color rojo para establecer una mala calidad. La línea roja que pasa por las cajas amarillas representa el valor de la mediana de calidad y la línea azul el valor de la media de calidad, mientras que las cajas amarillas explican los rangos de intercuartiles (25%-75%).

En la figura 11 A se muestra el gráfico de calidad del archivo R1, donde se observa como la calidad de las lecturas fue decayendo conforme avanzan los ciclos de secuenciación, tal y como lo

muestran los rangos de intercuartiles y las líneas de la media (azul) y la mediana (roja) de calidad; por otro lado, la gráfica B muestra la calidad por base del archivo R2, donde se observa que el archivo R2 cuenta con una mala calidad al inicio y al final de la secuenciación. Los gráficos de calidad de las otras dos réplicas del pozo y de los cultivos de bacterias sulfato reductoras y ácido productoras presentan los mismos patrones que los que se muestran en la figura 11.

Conociendo de manera gráfica la calidad con la que cuentan las secuencias obtenidas con MiSeq, es importante realizar una depuración de las mismas para obtener resultados confiables y cercanos a la diversidad presente en las muestras analizadas en el presente trabajo.

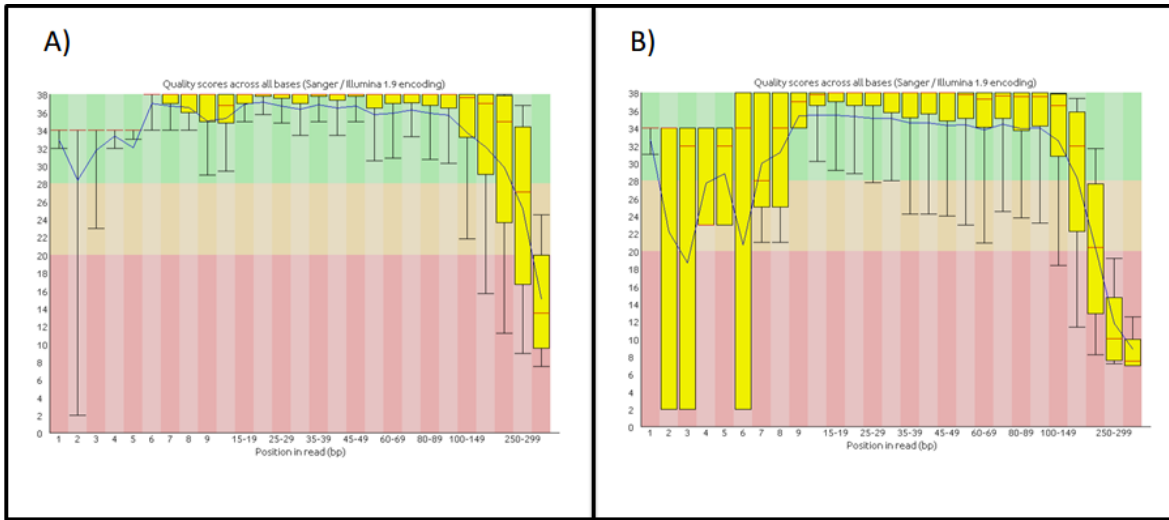


Figura 11.- Gráficos de calidad por base de la réplica 2 del pozo, antes de los filtros de calidad.

En la figura 12 se presentan los gráficos de los archivos R1 y R2 correspondientes a la muestra ACR 2-5 del pozo en donde se muestra el contenido de N por base de las secuencias; durante la secuenciación, existe la posibilidad de que el secuenciador coloque una N cuando este no es capaz de llamar una base con la suficiente confianza. En los gráficos de la figura 12, el eje 'x' muestra la posición de la lectura en la que se encuentra presente una N, mientras que el eje 'y' muestra el porcentaje de N's presentes en la secuencia.

De acuerdo con la figura 12, el contenido de N se encuentra al inicio de la secuenciación de los archivos R1 y R2; en el caso del archivo R1, se encontraba un porcentaje de entre 10-18% de N's en las primeras 5 bases, cuya evidencia son los dos picos mostrados en el gráfico, mientras que para el caso del archivo R2, fue encontrado un porcentaje de N de alrededor del 40-50% en las primeras seis bases de la secuencia, el comportamiento descrito fue el mismo no solo para la librería ACR 2-5, sino para las otras dos réplicas del pozo y de los cultivos de bacterias sulfato reductoras y ácido productoras.

Puede ser posible que la presencia de N's al inicio de las lecturas sea debido a que se encuentran restos de las secuencias de los oligos utilizados, y al no presentar variabilidad, siendo la misma secuencia repetidas veces, pudo ocurrir que la química del secuenciador empezó a ser estable después de cierto número de bases, sustituyendo entonces una base por la letra N. La gráfica del archivo R2 muestra un mayor porcentaje de N's presentes en la muestra, lo cual podría explicar las

secuencias de mala calidad que se encuentran al inicio de la secuenciación, tal y como lo muestra la figura 11.

Fue preciso remover las primeras bases de los archivos R1 y R2 en donde se encontraban las N's, así como también, los restos de secuencias de los oligos como el primer filtro de calidad. Después de los filtros posteriores aplicados para mantener la calidad de las secuencias, se presentan los gráficos de la librería ACR 2-5 después de la depuración.

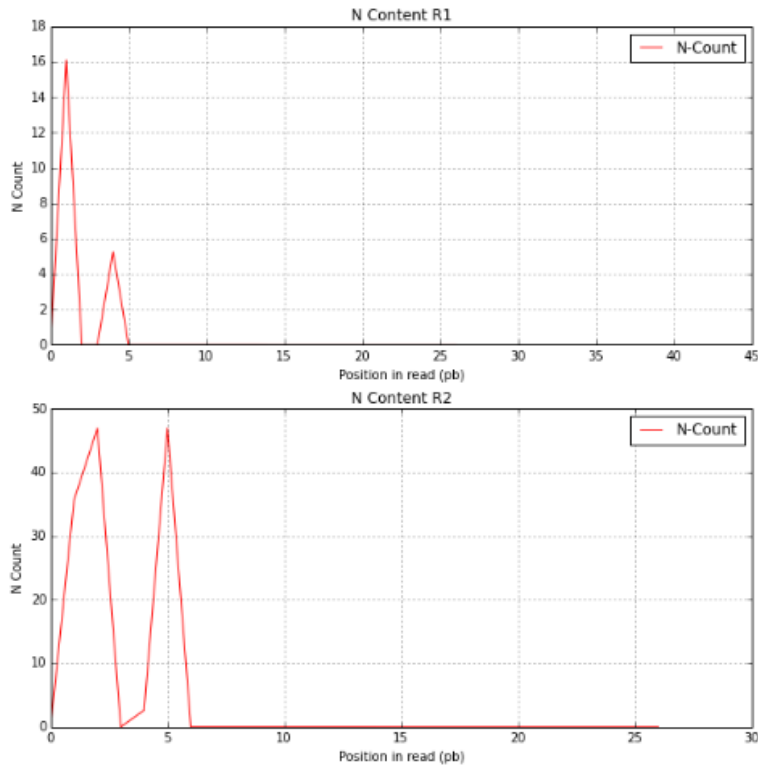


Figura 12.- Gráficos de nucleótidos no asignados (N) de la réplica 2, antes de los filtros de calidad.

La figura 13 muestra que después del primer filtro de calidad aplicado a las secuencias, el cual consistió en remover nucleótidos no nombrados, o N's, ya no se observan los picos que indican su presencia dentro de las primeras bases de las secuencias; las N's fueron removidas de las lecturas porque puede ser que, al realizar los análisis finales de taxonomía, se puedan detectar no como bases erróneas, sino como posibles nuevos OTUs, dando, por lo tanto, resultados falsos positivos.

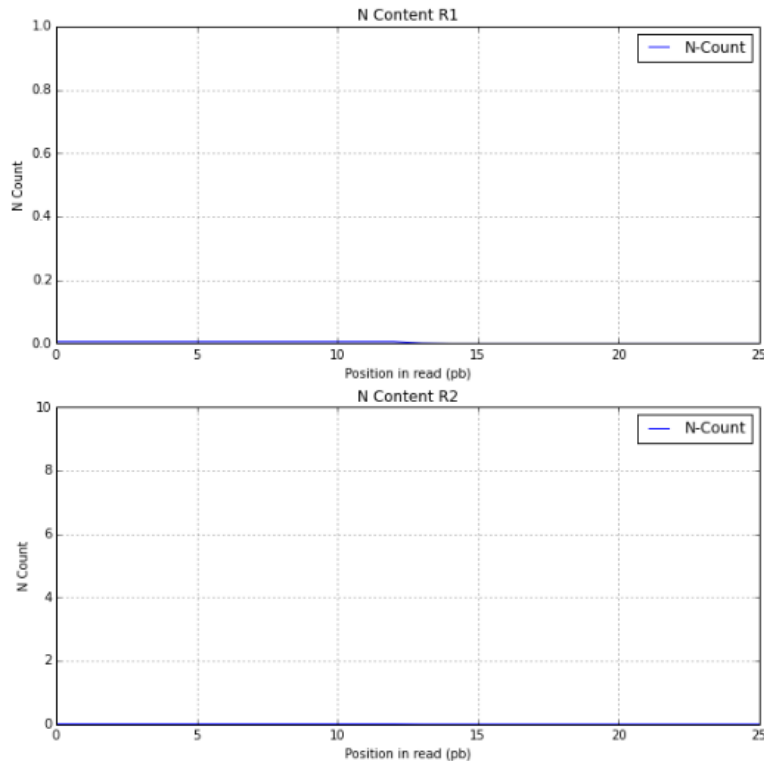


Figura 13.- Gráficos de nucleótidos no asignados (N) de la réplica 2, después de los filtros de calidad.

La secuenciación con la plataforma MiSeq regresa secuencias de longitud de 2 x 250 bases, a excepción que el número de ciclos durante el proceso sea aumentado, como fue el caso del presente trabajo, cuyo número de ciclos fue de 300, lo que significa que la mayoría de las lecturas obtenidas en los archivos R1 y R2 de cada una de las muestras secuenciadas cuentan con una longitud de 300 bases aproximadamente. Como se puede observar en la figura 11, cuando se cuenta con una longitud mayor a la esperada al momento de la secuenciación, es más probable que la calidad de las secuencias vaya disminuyendo, por lo que para depurar las secuencias, se cortaron los extremos de las lecturas dejándolas en un rango de longitud de 230 a 235 bases, probando después, sobrelapar cada uno de los cortes, esperando perder la menor cantidad posible de lecturas y además, obtener secuencias de un tamaño aproximado de 460 pares de bases, que es el tamaño real de las regiones V3 y V4 del gen 16S.

Las lecturas de la librería ACR 2-5 presentaron el mejor sobrelape, al perder una menor cantidad de lecturas, cuando estas fueron cortadas a una longitud de 232 bases; los cortes de las demás secuencias pertenecientes a las réplicas 3 y 4 del pozo y de los cultivos de bacterias sulfato reductoras y ácido productoras fueron dependientes de la muestra, cayendo todas dentro del rango de longitud de 230-232 bases.

Una vez que los archivos R1 y R2 de las muestras secuenciadas fueron sobrelapados, se estableció el puntaje de calidad con el que deben contar todas las secuencias de MiSeq, que fue de 20, el cual fue logrado mediante el uso de herramientas bioinformáticas.

Como último filtro, se removieron las posibles N's que se encontraran a lo largo de la secuencia para posteriormente formar un archivo con todas las muestras, el cual será compatible con QIIME.

Se muestran los resultados, tanto de las lecturas que se conservaron a lo largo de la depuración de las secuencias, como el grafico de calidad de la librería ACR 2-5.

La tabla 16 muestra el número de lecturas que se pierden después de aplicar a las secuencias cada uno de los filtros de calidad antes mencionados, como fue establecido, además de depurar las secuencias, también se busca perder la menor cantidad de lecturas posible con el fin de evitar perder diversidad; de acuerdo con la tabla 16, el número de lecturas perdidas para cada una de las muestras resulta ser mínima.

Muestra	Lecturas crudas	Lecturas unidas	Calidad por ventanas	Sin posibles N's
ACR 2-5	33485	27850	27841	27825
ACR 3-5	32561	27029	27016	27002
ACR 4-5	38682	32571	32564	32549

Tabla 16. - Lecturas de las tres réplicas del pozo después de cada filtro de calidad.

Como lo muestra la figura 14, la librería ACR 2-5 cuenta ya con una buena calidad, ya que, en comparación con la figura 11, ahora todas las lecturas se encuentran dentro de un rango de calidad de 32-37 debido a los filtros antes aplicados, por otro lado, cuenta también con la longitud aproximada de las regiones hipervariables V3-V4 que es de 460 pares de bases. Las librerías de las réplicas 3 y 4 y de los cultivos bacterianos presentan el mismo comportamiento, cuyos gráficos se observan similares al de la réplica 2.

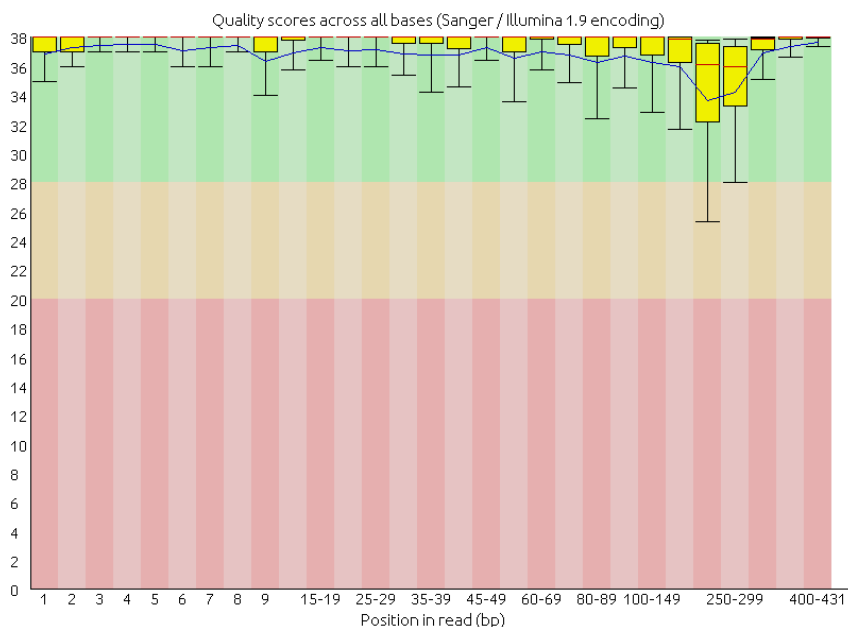


Figura 14. - Gráfico de calidad por base de la réplica 2 del pozo, después de los filtros de calidad.

El último paso fue juntar las librerías en un archivo con formato fasta, el cual es compatible con QIIME para así poder ser analizadas; una vez originado el archivo con las librerías, el siguiente paso fue seleccionar la base de datos ribosomal que mejor conviniera para asignación de diversidad y taxonomía.

6.3 Selección de la base de datos ribosomal

Antes de someter las secuencias ya depuradas por calidad, fue preciso llevar a cabo un análisis previo al final para conocer qué base de datos ribosomal explica la mejor diversidad presente en la muestra; se compararon tres bases de datos que fueron: Greengenes, Silva y RDP; el programa QIIME asigna diversidad por defecto con la base de datos de Greengenes, por lo que las otras dos bases de datos tuvieron que ser descargadas y utilizadas como referencia para las secuencias.

Nuevamente fue tomada la librería de la réplica 2 del pozo para realizar las pruebas de comparación entre las tres bases de datos, dichas pruebas consistieron en el número de lecturas que lograron asignar a nivel de género y especie cada una de las bases de datos, el número de OTUs clasificados a los distintos niveles taxonómicos, desde phylum hasta especie, métricas de diversidad alfa y diagramas de Venn de los géneros compartidos entre las tres bases de datos.

6.3.1 Ion Torrent

En la tabla 17 se muestra el número de lecturas de la réplica 2 del pozo antes y después de entrar al programa QIIME, como también las lecturas que fueron asignadas a los niveles taxonómicos de género y especie con cada una de las tres bases de datos que fueron comparadas.

A pesar de que al entrar al programa QIIME la réplica 2 pierde más de la mitad de lecturas con las que contaba, la base de datos con las que logra conservar una mayor cantidad de las mismas resulta ser Silva, representando una pérdida del 79%, mientras que la base de datos cuya pérdida representó un 89% de las lecturas fue RDP, por otro lado, Greengenes mostró un comportamiento un tanto similar al de Silva, representando una pérdida de lecturas del 81%.

En cuanto al número de lecturas asignadas a nivel de género y especie, nuevamente fue Silva la base de datos que logró conservar la mayoría de sus lecturas al asignar OTUs en los dos niveles taxonómicos antes mencionados, en comparación con Greengenes, cuya asignación de OTUs se ve reflejada en una gran pérdida de lecturas; por otro lado, la base de datos RDP no perdió ninguna lectura al asignar OTUs a nivel de género, aunque una de las grandes desventajas con las que cuenta, en comparación con las otras dos bases de datos, es que con RDP no es posible asignar OTUs a nivel de especie.

De acuerdo con la primera prueba realizada al comparar las tres distintas bases de datos, la cual fue revisar el número de lecturas que fue perdiendo la librería utilizada como ejemplo, es notorio que la base de datos que logró la menor pérdida fue Silva, seguida por Greengenes y por último RDP, teniendo esta la desventaja de no asignar OTUs a nivel de especie.

Muestra	Lecturas filtradas por calidad	Lecturas que entraron a QIIME	Lecturas asignadas hasta genero	Lecturas asignadas hasta especie
Rep. 2 GG	59152	11451	4585	265
Rep. 2 Silva	59152	12975	12387	11537
Rep. 2 RDP	59152	6807	6807	-

Tabla 17. - Lecturas de la librería de la réplica 2 con las tres distintas bases de datos.

La segunda prueba realizada para seleccionar la base de datos ribosomal consistió en comparar el número de OTUs asignados a los distintos niveles taxonómicos, que se obtienen con cada una de las bases de datos comparada.

En la tabla 18 se muestran los resultados de la comparación del número de OTUs que lograron ser asignados por las tres bases de datos, como se puede observar, la base de datos Silva resulta ser nuevamente la ganadora, consiguiendo asignar el mayor número de OTUs desde phylum hasta especie para la réplica 2 del pozo, logrando un total de 229 géneros y 327 especies, lo cual coincide con el número de lecturas que logró conservar y que se muestran en la tabla 17, que como se ha mencionado anteriormente, entre menor sea el número de lecturas perdidas, resulta más probable lograr asignar una mayor cantidad de OTUs. Seguida de Silva, resulta la base de datos Greengenes, la cual logra una buena asignación de taxonomía en comparación con RDP, hasta el nivel de familia, ya que RDP logró una mayor asignación de OTUs a nivel de género, por otro lado, el número de especies que logra asignar Greengenes es de 23, resultando en menos de la mitad de las que se logran asignar con Silva, las pocas especies obtenidas con Greengenes se pueden ver reflejadas con la pérdida de lecturas como se muestra en la tabla 18.

De acuerdo con la segunda prueba, nuevamente resulta ser Silva la base de datos que logra una mejor asignación de OTUs desde phylum hasta especie en comparación con Greengenes y RDP.

Muestra	Phylum	Clase	Orden	Familia	Género	Especie
Rep. 2 GG	16	32	62	87	86	23
Rep. 2 Silva	18	40	89	161	229	327
Rep. 2 RDP	9	21	42	77	119	-

Tabla 18. - Número de OTUs asignados a los distintos niveles taxonómicos por las tres bases de datos.

Antes de decidir con cuál base de datos ribosomal se realizará el análisis final de las secuencias obtenidas con Ion Torrent, se obtuvieron las métricas que representan la diversidad alfa de la réplica 2 del pozo, además de los diagramas de Venn a nivel de género que demuestran que tan similares resultan ser las tres bases de datos comparadas.

Para poder comparar los índices de diversidad alfa obtenidos con las tres bases de datos, se tomó en cuenta el mismo número de secuencias por muestra; como se muestra en la tabla 19, la base de datos que logró obtener un mayor índice de diversidad filogenética (PD) fue RDP, la cual, a pesar de ser la base de datos con el menor número de OTUs asignados a los distintos niveles taxonómicos, resultó ser la más diversa en cuanto a riqueza de las especies, seguida de Silva, cuyo índice de PD no muestra una gran diferencia con el obtenido con Greengenes. Por otro lado, en cuanto a los índices de Shannon, Simpson y el número de OTUs observados, Silva resulta ser la base de datos que muestra la mayor diversidad de la muestra de acuerdo con los valores obtenidos.

A pesar de la variabilidad de los índices de diversidad alfa obtenidos con las tres bases de datos, Silva sigue resultando la base de datos con la cual, además de obtener el mayor número de OTUs asignados en los niveles taxonómicos desde phylum hasta especie, se puede lograr explicar la diversidad de la muestra analizada de acuerdo con los altos índices de diversidad que se obtuvieron.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Rep. 2 GG	6800	7.501	4.864	0.864	425
Rep. 2 Silva	6800	7.725	5.196	0.898	462
Rep. 2 RDP	6800	8.947	4.571	0.904	208

Tabla 19. - Índices de diversidad alfa obtenidos con las tres bases de datos.

Como último paso, antes de conocer la base de datos ganadora para los análisis finales de las secuencias, se realizaron diagramas de Venn a nivel de género para conocer los OTUs compartidos entre las tres bases de datos y saber con cuál base de datos se obtiene la mayor cantidad de OTUs distintos a los compartidos.

En la figura 15 se observa el diagrama de Venn comparando las tres bases de datos; de acuerdo con la imagen, se puede ver que de los OTUs totales asignados por las tres distintas bases de datos, únicamente 28 géneros son compartidos entre las tres, mientras que la base de datos que logra obtener la mayor cantidad de OTUs distintos a los compartidos resulta ser Silva, seguida por la base de datos RDP y por último Greengenes.

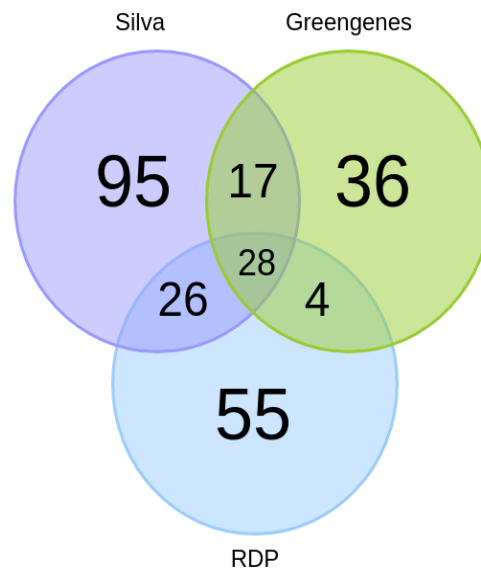


Figura 15. - Diagrama de Venn de los géneros compartidos entre las tres bases de datos.

De acuerdo con los resultados de las pruebas realizadas a las tres bases de datos que fueron comparadas en el presente trabajo, hubo dos de ellas que lograron los mejores resultados en cuanto a asignación de taxonomía e índices de diversidad alfa, las cuales fueron Silva y RDP, sin embargo, la base de datos RDP resulta no ser la más óptima para el análisis de las muestras del pozo, ya que, aunque a pesar de lograr una buena asignación de taxonomía, fue la única de las tres que descartó el mayor número de lecturas, además de ser la única base de datos que no es capaz de asignar OTUs a nivel de especie.

Como lo indican los resultados de las comparaciones entre las tres bases de datos, la mejor para realizar los análisis de diversidad de las tres réplicas del pozo, resulta ser Silva, ya que fue la base

de datos cuyos resultados de asignación de taxonomía e índices de diversidad alfa fueron más altos en comparación con los resultados obtenidos con las otras dos bases de datos evaluadas, por lo tanto, el análisis final de las réplicas del pozo será realizado utilizando como referencia la base de datos Silva.

6.3.2 MiSeq

Como se puede observar en la tabla 20, existen diferencias en cuanto al número de lecturas que se logran asignar a los niveles de género y especie con cada una de las bases de datos a comparar, siendo Silva la base de datos que logra conservar el mayor número de lecturas al someter las secuencias al programa, las mismas que conserva al ser asignadas a los niveles de género y especie, indicando entonces, que se logró asignar el 100% de OTUs a los niveles taxonómicos antes mencionados, por lo tanto, resulta ser la base de datos con mejor asignación en comparación con las otras dos.

Por otro lado, el análisis con la base de datos Greengenes resulta ser un tanto similar al análisis con Silva, ya que, al someter las secuencias al programa, las dos bases de datos conservan casi el mismo número de lecturas, la diferencia está en que, al asignar taxonomía a nivel de género y especie, se pierden más lecturas con Greengenes, además, no se logran asignar el 100% de OTUs a los niveles antes mencionados.

Por otro lado, cuando las secuencias son analizadas con RDP, es notorio que esta base de datos pierde un mayor número de lecturas en comparación con las otras dos, aunque logra asignar el 100% de OTUs a nivel de género, pero aun así se encuentra en desventaja, ya que, de las tres comparadas, resulta ser la única base de datos que no logra asignar OTUs a nivel de especie.

De acuerdo con la tabla 20, resulta ser Silva la base de datos que logra conservar el mayor número de lecturas y con la cual se obtiene el 100% de asignación de géneros y especies, por lo que, los resultados de la primera prueba hecha a las tres bases de datos indican que Silva podría ser óptima para el análisis de las réplicas del pozo, sin embargo, antes de decidir, se muestran los resultados de las demás pruebas, como la asignación de taxonomía, índices de diversidad alfa y diagrama de Venn de géneros compartidos entre las tres bases de datos.

Muestra	Lecturas filtradas por calidad	Lecturas que entraron a QIIME	Lecturas asignadas hasta genero	Lecturas asignadas hasta especie
ACR2 GG	7038	6852	5579	5161
ACR 2 Silva	7038	6867	6867	6867
ACR2 RDP	7038	5385	5385	-

Tabla 20. - Lecturas de la librería ACR2 con las tres distintas bases de datos.

En la tabla 21 se muestran los resultados de los OTUs que lograron ser asignados por las tres bases de datos; como se puede observar, los resultados obtenidos con Silva y Greengenes resultan ser un tanto similares en los cuatro primeros niveles taxonómicos, desde phylum hasta familia, sin embargo, al llegar a los niveles que resultan de mayor importancia como lo son género y especie es donde existen diferencias, siendo Silva la base de datos con la que se obtiene un mayor número de OTUs asignados a los niveles antes mencionados.

A pesar de que RDP es la base de datos que pierde un mayor número de lecturas en comparación con las otras dos, y que, además, no puede asignar OTUs a nivel de especie, los resultados mostrados en la tabla 21 indican que, en cuanto a asignación de taxonomía, las tres bases de datos presentan un comportamiento un tanto similar.

Muestra	Phylum	Clase	Orden	Familia	Género	Especie
ACR2 GG	7	8	11	15	12	7
ACR2 Silva	7	8	12	13	15	19
ACR2 RDP	6	7	9	11	13	-

Tabla 21. - Número de OTUs asignados a los distintos niveles taxonómicos por las tres bases de datos.

En la tabla 22 se muestran los resultados de las métricas de diversidad alfa que se obtuvieron con las tres bases de datos, fue tomado el mismo número de secuencias por muestra para que los resultados pudieran ser comparables. Como se observa en la tabla, de los índices que fueron comparados, se puede notar un comportamiento similar entre las tres bases de datos, ya que hay variaciones mínimas, incluyendo el número de OTUs que fueron estimados por el programa se encuentran muy similares, mismo comportamiento que presentan al asignar taxonomía, lo cual se puede observar en la tabla 21.

Con los resultados obtenidos de las métricas de diversidad alfa, podría decirse que, al no existir diferencias notorias entre las tres bases de datos, las tres resultarían óptimas al explicar la riqueza de las especies presentes en la muestra de la réplica 2 del pozo.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
ACR2 GG	3710	2.759	1.270	0.410	17
ACR2 Silva	3710	2.648	1.292	0.411	18
ACR2 RDP	3710	2.195	0.439	0.104	16

Tabla 22. - Índices de diversidad alfa obtenidas con las tres bases de datos.

En el diagrama de Venn se muestra la comparación entre las tres bases de datos en cuanto al número de OTUs asignados a nivel de género que se obtienen con cada una, por lo que se pretende conocer con cuántos OTUs distintos cuenta cada base de datos. Como se puede observar, la mayoría de OTUs asignados a nivel de género se encuentran compartidos entre las tres bases de datos, sin embargo, Silva resulta tener 3 distintos OTUs en comparación con las otras dos bases, siendo Greengenes la menos diversa, ya que únicamente cuenta con un OTU distinto.

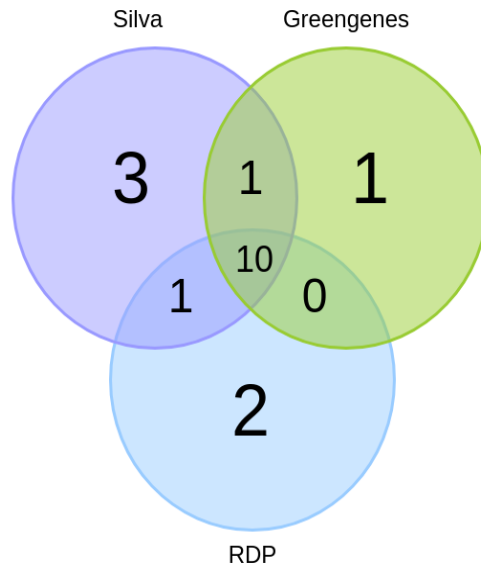


Figura 16. - Diagrama de Venn de los géneros compartidos entre las tres bases de datos.

De acuerdo con las pruebas realizadas a las tres bases de datos que fueron comparadas, y a pesar de que los resultados presentaron similitudes entre las tres, lo que se busca es una base de datos adecuada para el análisis de las muestras del pozo y de los cultivos bacterianos, siendo Silva la base de datos que cumple con dicha característica, ya que cuando las secuencias son sometidas al programa tomando Silva como referencia, resulta en una buena asignación de taxonomía y en índices de diversidad altos, por lo que los análisis definitivos que serán realizados a las réplicas del pozo y los cultivos bacterianos serán tomando como referencia la base de datos Silva.

6.4 Caracterización taxonómica Ion Torrent

Los análisis finales de taxonomía y diversidad de las muestras del pozo se realizaron tomando como referencia la base de datos Silva, se presentan los resultados obtenidos mediante gráficos que muestran la abundancia de OTUs a distinto nivel taxonómico, gráficos e índices de diversidad alfa, diagramas de Venn de géneros y especies compartidos entre las tres réplicas del pozo y correlaciones de Spearman para conocer la similitud entre las mismas y determinar la reproducibilidad técnica.

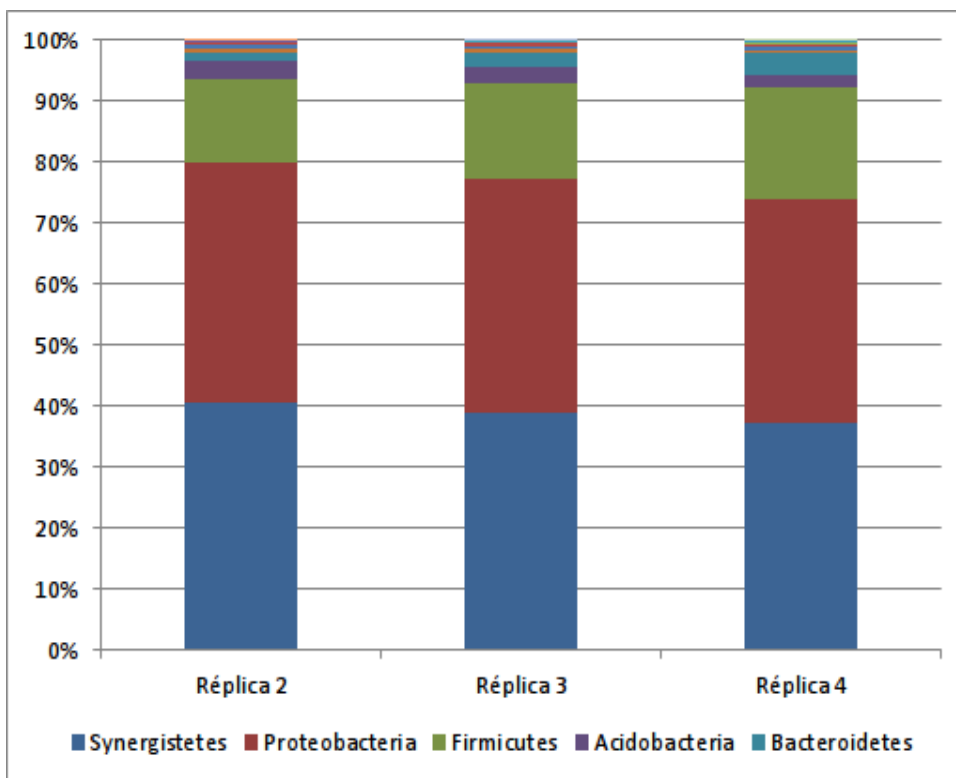


Figura 17. - Gráfica de OTUs a nivel de phylum de las tres réplicas del pozo.

La figura 17 muestra la asignación taxonómica obtenida con la base de datos Silva a nivel de phylum de las tres réplicas del pozo; como se puede observar en la gráfica, se presentan con mayor abundancia 5 distintos phylum, los cuales corresponden a la división *Synergistetes*, *Proteobacteria*, *Firmicutes*, *Acidobacteria* y *Bacteroidetes*, donde el phylum que resulta ser más abundante pertenece a la división *Synergistetes*, comprendiendo una frecuencia entre el 37%-40% para las tres réplicas del pozo, dicha división abarca la mayoría de la diversidad presente en las muestras. Las bacterias pertenecientes al phylum *Synergistetes* resultan ser bacilos Gram negativos estrictamente anaerobios y neutrófilos; se han encontrado en un amplio rango de hábitats anaerobios entre los que se incluyen tractos gastrointestinales de animales e insectos, sistemas de tratamiento de agua, suelos y pozos petroleros (Hugenholtz *et al.*, 2009). El segundo phylum más abundante está representado por las *Proteobacterias*, abarcando un total en las muestras entre el 36%-39% para las tres réplicas del pozo, seguidas en abundancia por el phylum *Firmicutes* cuya abundancia en las muestras se encuentra representada por el 13%-18%. Dentro de los 5 primeros phylum más abundantes encontrados en las réplicas del pozo que son mencionados en el presente trabajo, se encuentran también en menor proporción los phylum pertenecientes a las *Acidobacterias* y *Bacteroidetes*.

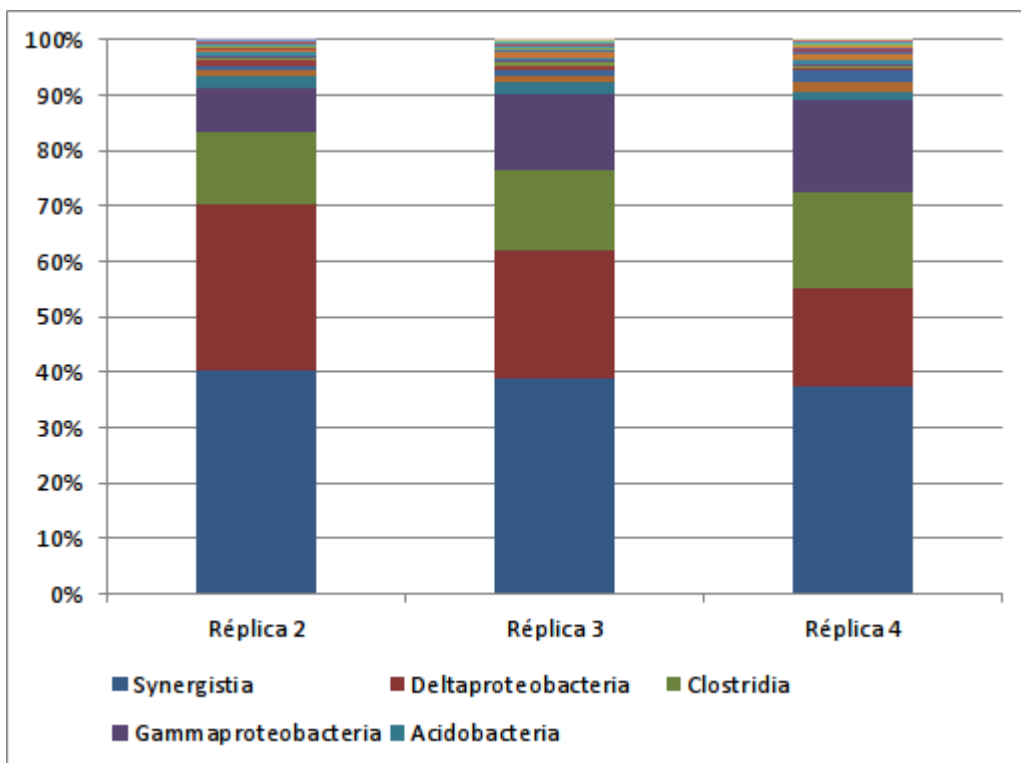


Figura 18. - Gráfica de OTUs a nivel de clase de las tres réplicas del pozo.

La figura 18 representa la gráfica de OTUs correspondiente al nivel de clases encontradas en las tres réplicas del pozo, al igual que en la figura 17, se muestran en el pie de gráfica las 5 clases más abundantes presentes en las muestras. La clase más abundante en las muestras secuenciadas con Ion Torrent resulta ser *Synergistia*, la cual comprende entre el 37%-40% del total de la muestra en las tres replicas; se encontró también la presencia de dos subdivisiones del phylum *Proteobacteria*, que son *Delta* y *Gammaproteobacteria*, cuya abundancia en las muestras se encuentra representada por el 17%-29% y 8%-16%, respectivamente, las bacterias pertenecientes a esta clase resultan ser metabólicamente las más diversas y constituyen la mayoría de bacterias con significancia médica, industrial y en la agricultura, como grupo, todas las bacterias son Gram negativas, fisiológicamente representan una gran diversidad, siendo anaerobias, microaerofílicas y aerobias facultativas (Madigan *et al.*, 2009). Otra de las clases que se encuentran presentes dentro de las 5 más abundantes de las tres réplicas del pozo resulta ser *Clostridia*, con una abundancia del 12%-17%; dicha clase representa un diverso grupo de bacterias dentro de las que pertenecen ambos tipos de organismos Gram positivos y negativos, una de las características de las bacterias pertenecientes a esta clase es que pueden fermentar aminoácidos para producir ATP en un proceso llamado reacción de Stickland, la cual genera amonio, sulfuro de hidrogeno y ácidos grasos durante la descomposición anaerobia de las proteínas (Willey *et al.*, 2014). Por último, representando una menor abundancia en comparación con las clases antes mencionadas, se encuentran las *Acidobacterias* con un 2% del total en las tres réplicas del pozo.

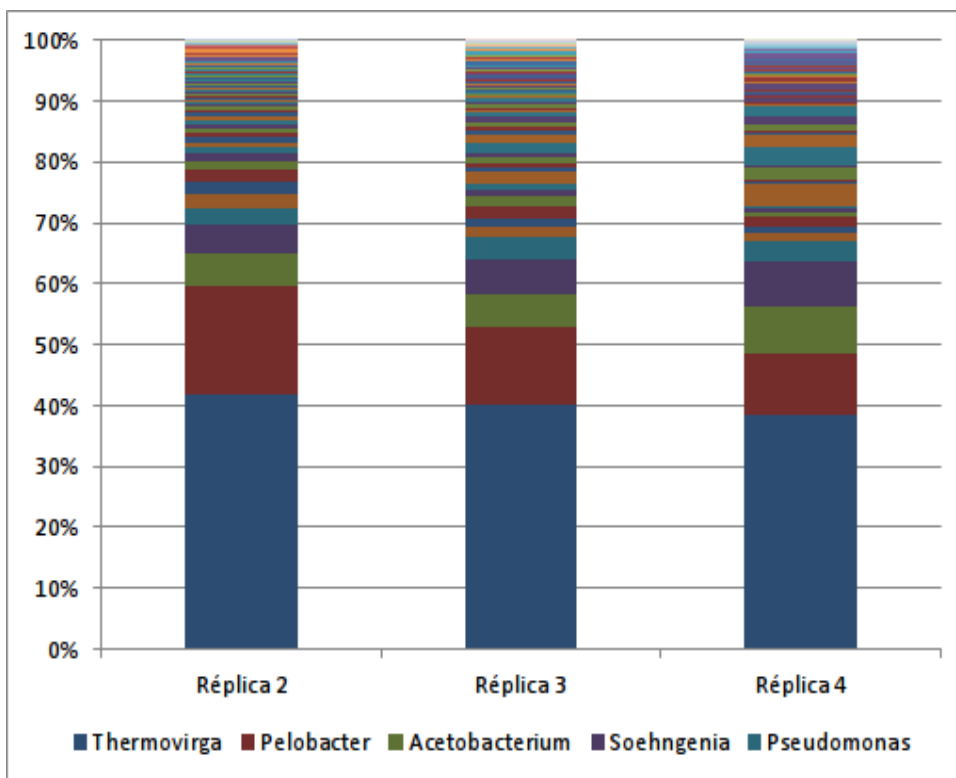


Figura 19. - Gráfica de OTUs a nivel de género de las tres réplicas del pozo.

En la figura 19 se muestra el gráfico de OTUs obtenidos a nivel de género de las tres réplicas del pozo, el género más abundante encontrado en las muestras resulta ser *Thermovirga* representando el 37%-40% en las tres réplicas del pozo, *Thermovirga* se clasifica dentro de los organismos denominados termófilos, los cuales pueden crecer a temperaturas entre 55 y 85°C, el crecimiento óptimo de estos organismos se encuentra en un rango de temperatura entre 55 y 65°C, dichos microorganismos poseen características como enzimas estables al calor y sistemas de síntesis de proteínas que funcionan adecuadamente a altas temperaturas (Willey *et al.*, 2014). Los microorganismos termófilos crecen rápidamente en muchos hábitats, incluyendo composta, pacas de heno, tuberías de agua caliente y aguas termales (Prescott *et al.*, 2004), Dahle en el 2006 logro aislar una nueva especie de *Thermovirga* proveniente de una muestra de agua producida de un yacimiento petrolero que fue nombrada *Thermovirga lienii*. Al momento de asignar OTUs a nivel de especie, la más abundante resultó ser una especie no cultivada de la familia *Synergistetes*, representando el 26%-29% de la abundancia en la muestra.

Otro de los géneros con mayor abundancia encontrado en las muestras de las tres réplicas resultó ser *Pelobacter*, con una abundancia del 9%-17% en las tres réplicas del pozo; dicho género es una entidad taxonómica de bacterias estrictamente anaerobias, Gram negativas, no formadoras de esporas, que utilizan un número muy limitado de sustratos; se ha encontrado que algunas especies de *Pelobacter* como *P. venetianus*, *P. carbinolicus* y *P. acidigallici* parecen estar altamente relacionadas con varias cepas de bacterias anaerobias pertenecientes a las sulfato reductoras (Schink, 2006). En cuanto a los OTUs asignados a nivel de especie encontrados en las tres réplicas del pozo, estas resultaron ser especies no cultivadas del género *Pelobacter*.

En una proporción entre el 4%-7% se encuentran las bacterias del género *Acetobacterium* que se clasifican dentro de las bacterias Gram positivas siendo estrictamente anaerobias, este tipo de bacterias presenta crecimiento cuando ocurre oxidación anaerobia de H_2 y reducción de CO_2 para formar ácido acético. Muchas de las especies encontradas de *Acetobacterium* se han encontrado en ambientes estrictamente anóxicos como ecosistemas de agua dulce, sedimentos anóxicos de estanques, zanjas, humedales y lodos de aguas residuales anóxicos (Bergey *et al.*, 2009). Los OTUs asignados a nivel de especie de *Acetobacterium* resultaron ser no cultivables del phylum *Firmicutes*.

Otro de los cinco principales géneros encontrados dentro de las tres réplicas del pozo es *Soehngenia* que se clasifica dentro de las bacterias Gram positivas, anaerobias pero aerotolerantes, dicha bacteria tiene la capacidad de fijar nitrógeno molecular y es capaz de crecer con un amplio rango de sustratos como carbohidratos y otras fuentes de carbono (Bergey *et al.*, 2009).

Por último, fue detectada también la presencia de bacterias pertenecientes al género *Pseudomonas* cuya abundancia en las tres réplicas del pozo es de entre un 2%-3%, esta bacteria es Gram negativa, aerobia, con un tipo de metabolismo que utiliza oxígeno como el aceptor terminal de electrones y puede crecer también de forma anaerobia (Bergey *et al.*, 2009). Una de las características principales de este tipo de bacterias es que se consideran dentro del grupo de bacterias biorremediadoras, encargadas de la eliminación de aceites, sustancias químicas tóxicas u otros contaminantes de determinado ambiente, debido a que poseen la capacidad para usar numerosos compuestos orgánicos como fuentes de carbono y energía, algunas especies utilizan más de 100 compuestos diferentes (Madigan *et al.*, 2009). Algunas cepas específicas de *Pseudomonas* poseen la característica de producir biosurfactantes que poseen la característica de remover contaminantes (Santa Anna *et al.*, 2002). Las especies de *Pseudomonas* que fueron encontradas en las réplicas del pozo no se encuentran descritas en la actualidad, pero son pertenecientes a *Pseudomonas sp. MA2*.

Cada una de las bacterias que fueron encontradas en las tres réplicas del pozo posee características que las diferencian unas de las otras, encontrando una gran diversidad entre las que destacan las bacterias ácido productoras, sulfato reductoras y biorremediadoras, la mayoría de las bacterias presentes en las muestras tienen la característica de ser anaerobias o anaerobias facultativas, mientras que otras como *Pseudomonas* presentan crecimiento en presencia de oxígeno, pero aun así logran sobrevivir en ambientes anóxicos como los pozos petroleros, esto puede deberse a que probablemente el conjunto de bacterias logra sobrevivir gracias a las relaciones simbióticas con los demás microorganismos presentes alrededor.

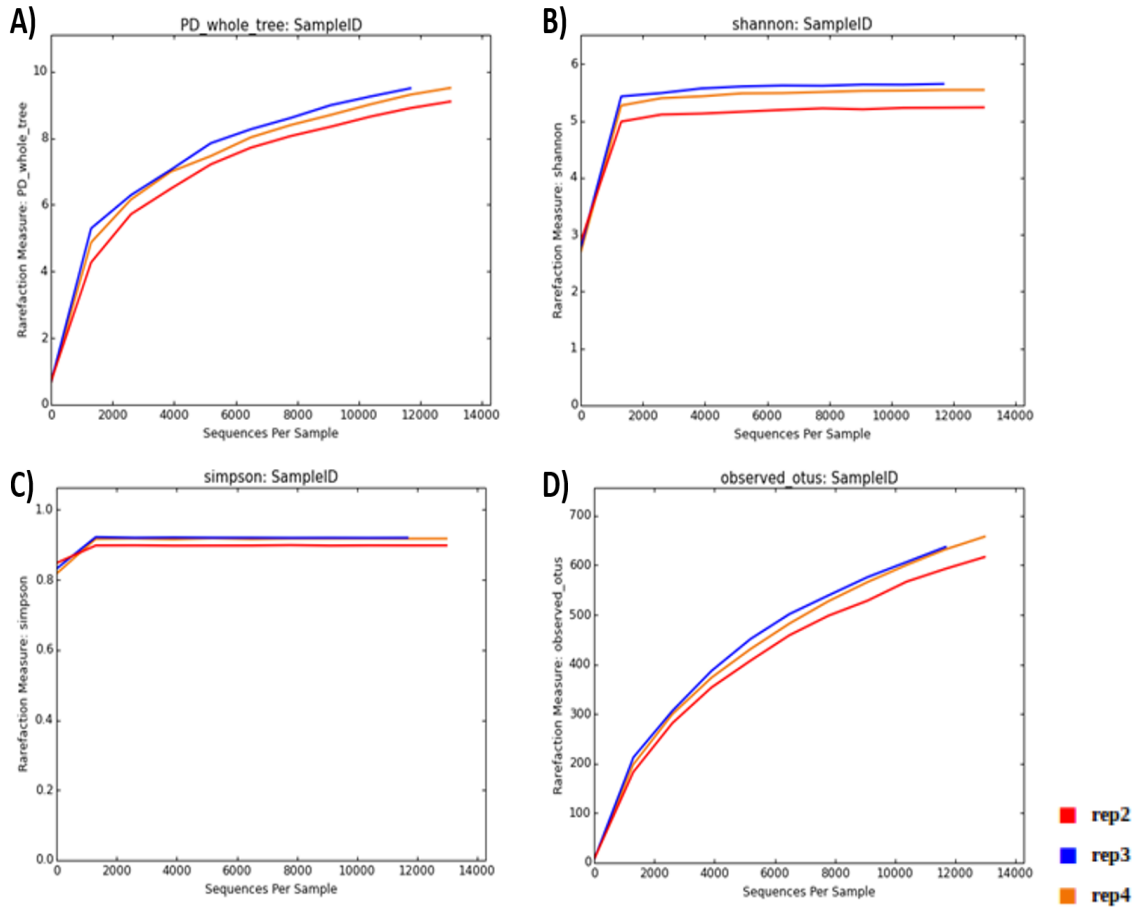


Figura 20. – Gráficas de los índices de A) Diversidad Filogenética (PD), B) Shannon, C) Simpson y D) OTUs observados de las tres réplicas del pozo.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Réplica 2	11674	8.900	5.231	0.898	593
Réplica 3	11674	9.494	5.649	0.920	637
Réplica 4	11674	9.304	5.543	0.917	632

Tabla 23. - Índices de diversidad alfa de las tres réplicas del pozo.

La figura 20 y la tabla 23 muestran las gráficas y los índices de diversidad alfa que se obtuvieron para las tres réplicas del pozo, respectivamente. De manera gráfica, se puede observar que existe una gran similitud entre las tres muestras de acuerdo con los índices obtenidos, mientras que en la tabla 23 se puede corroborar lo cercanos que se encuentran dichos índices, lo que indica que las tres réplicas resultan ser similares en cuanto a la diversidad y riqueza presente en cada una.

De acuerdo con el índice de diversidad filogenética (PD), resulta ser la réplica 2 la que representa el menor índice de PD, mientras que las réplicas 3 y 4 obtuvieron índices de diversidad muy similares entre ellas, el mismo comportamiento se puede observar cuando son comparados los índices de Shannon, Simpson y el número de OTUs estimados por el programa; el número de OTUs presentados en la tabla corresponden a los OTUs encontrados en los distintos niveles

taxonómicos. La similitud de la diversidad presente entre las tres muestras se ve reflejada por la reproducibilidad técnica que existe entre las réplicas; las réplicas 3 y 4 resultan ser las más similares entre ellas, como se puede observar en la tabla 23, aunque la réplica 2 presenta diferencias mínimas en comparación con las otras dos muestras.

Para conocer las similitudes que presentan las tres réplicas del pozo en cuanto a diversidad, se realizaron diagramas de Venn.

La figura 21 muestra el número de géneros y especies que se obtuvieron de las tres réplicas y que son compartidos; en las dos figuras se puede observar que, del total de la diversidad obtenida en las muestras, la mayoría es compartido por las tres réplicas; otra de las observaciones que se puede notar en los diagramas de Venn, es que las réplicas 2 y 3 en cuanto a asignación de géneros y especies, se encuentran muy similares entre ellas, mientras que, la réplica 4 logra ser la muestra con el mayor número de OTUs asignados a los niveles taxonómicos de género y especie y con la que las réplicas 2 y 3 logran compartir la mayor parte de dichos OTUs.

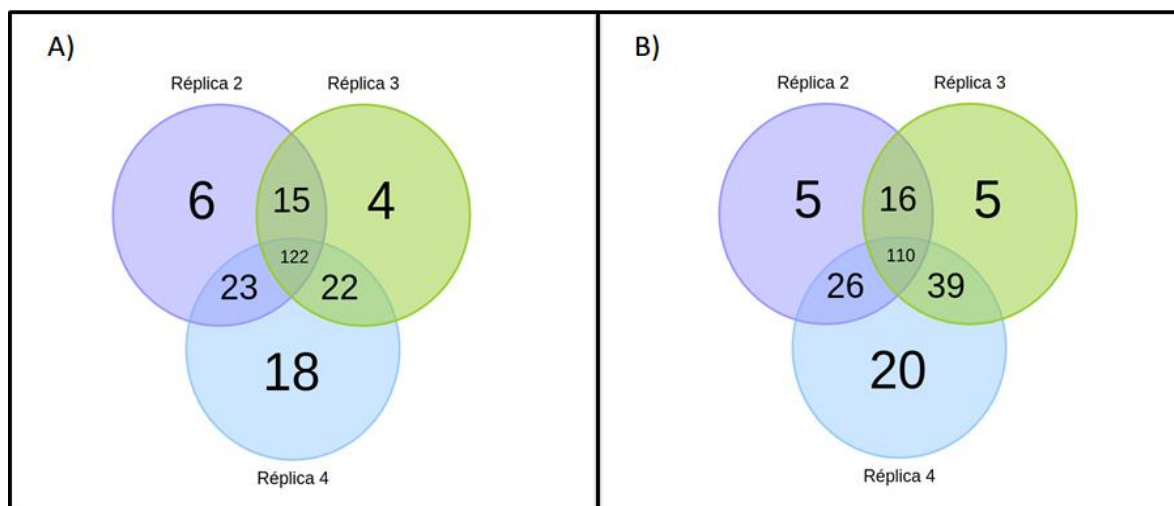


Figura 21.- Diagramas de Venn de A) géneros y B) especies compartidos entre las tres réplicas del pozo.

Al comparar la diversidad de géneros y especies encontradas en las tres réplicas de agua, con los índices de diversidad alfa obtenidos los cuales reflejan la riqueza entre las muestras, se puede observar que, de acuerdo con la tabla 23, las réplicas que resultan ser más parecidas en cuanto a riqueza, resultan ser las muestras 3 y 4, sin embargo, cuando se compara la diversidad de OTUs, resultan ser más parecidas las réplicas 2 y 3, ya que la réplica 4 resulta en un mayor número de OTUs asignados a los dos distintos niveles de género y especie. De acuerdo con los resultados de diversidad y riqueza obtenidos para las tres réplicas, se puede observar que entre las tres réplicas se comparte similitud, la cual es de esperarse, ya que provienen de réplicas tomadas de la misma tubería; antes de concluir que existe reproducibilidad técnica entre las tres muestras, se realizaron correlaciones de Spearman para determinar si estadísticamente existe similitud entre ellas.

En la tabla 24 se encuentran las correlaciones de Spearman obtenidas con el programa R Studio, las cuales van a ayudar a explicar las asociaciones estadísticamente existentes entre las tres réplicas del pozo; considerando los valores que se obtuvieron al comparar las muestras, fueron obtenidas correlaciones que se consideran positivas fuertes, ya que se encuentran en un rango \geq

0.5; en donde el valor de correlación más alto obtenido fue entre las réplicas 2 y 3, seguido del valor obtenido para las réplicas 3 y 4, lo cual indica que las réplicas antes mencionadas sostienen fuertes asociaciones o similitudes, dicha similitud pudo ser también observada al comparar la riqueza y la diversidad (figura 20 y tabla 23). Las réplicas que presentan una menor asociación entre ellas, resultan ser las réplicas 2 y 4, que, a pesar de tener una correlación positiva alta, no resultan ser del todo similares estadísticamente, lo mismo fue corroborado cuando fueron comparadas tanto la riqueza como la diversidad de géneros y especies encontradas en las dos muestras antes mencionadas.

Muestra	Réplica 2	Réplica 3	Réplica 4
Réplica 2	1	0.6461	0.5349
Réplica 3	0.6461	1	0.6351
Réplica 4	0.5349	0.6351	1

Tabla 24.- Correlaciones de Spearman entre las tres réplicas del pozo.

En general, se pueden observar las similitudes entre las tres réplicas del pozo, tomando en cuenta la diversidad obtenida a distintos niveles taxonómicos (figuras 17, 18 y 19), ya que al observar los gráficos obtenidos se pueden encontrar patrones semejantes entre ellos, tanto gráficamente como en abundancia de los OTUs presentes en las muestras, ocurriendo un comportamiento parecido al someter las tres réplicas a los análisis de diversidad alfa, cuyo gráfico (figura 20) muestra los distintos índices comparados, los cuales presentan tendencias parecidas en las tres réplicas, cuyo comportamiento puede ser corroborado con la tabla 23, mostrando las similitudes existentes en cuanto a riqueza de especies encontradas en las muestras; por último, al haber obtenido las correlaciones de Spearman, se puede corroborar que estadísticamente, las tres réplicas presentan asociaciones positivas entre ellas, con lo que es posible deducir que al realizar el muestreo, existe una alta reproducibilidad técnica.

Una vez que fue corroborada la reproducibilidad técnica que existe en el muestreo de las tres réplicas del pozo, el siguiente paso, antes de comparar la diversidad existente entre los dos métodos de secuenciación fue unir las tres secuencias correspondientes a las réplicas para realizar la caracterización general del pozo petrolero, ya que, para fines prácticos de análisis, resulta más conveniente analizar las secuencias como una única muestra, teniendo de respaldo los resultados antes presentados, los cuales confirman que las tres réplicas resultan ser semejantes entre ellas.

6.5 Caracterización taxonómica MiSeq

Los análisis finales de taxonomía y diversidad de las muestras del pozo se realizaron tomando como referencia la base de datos Silva, se presentan los resultados obtenidos mediante gráficos que muestran la abundancia de OTUs a distintos niveles taxonómicos, gráficos e índices de diversidad alfa, diagramas de Venn de géneros y especies compartidos entre las tres réplicas del pozo y correlaciones de Spearman para conocer la similitud entre las mismas y determinar la reproducibilidad técnica.

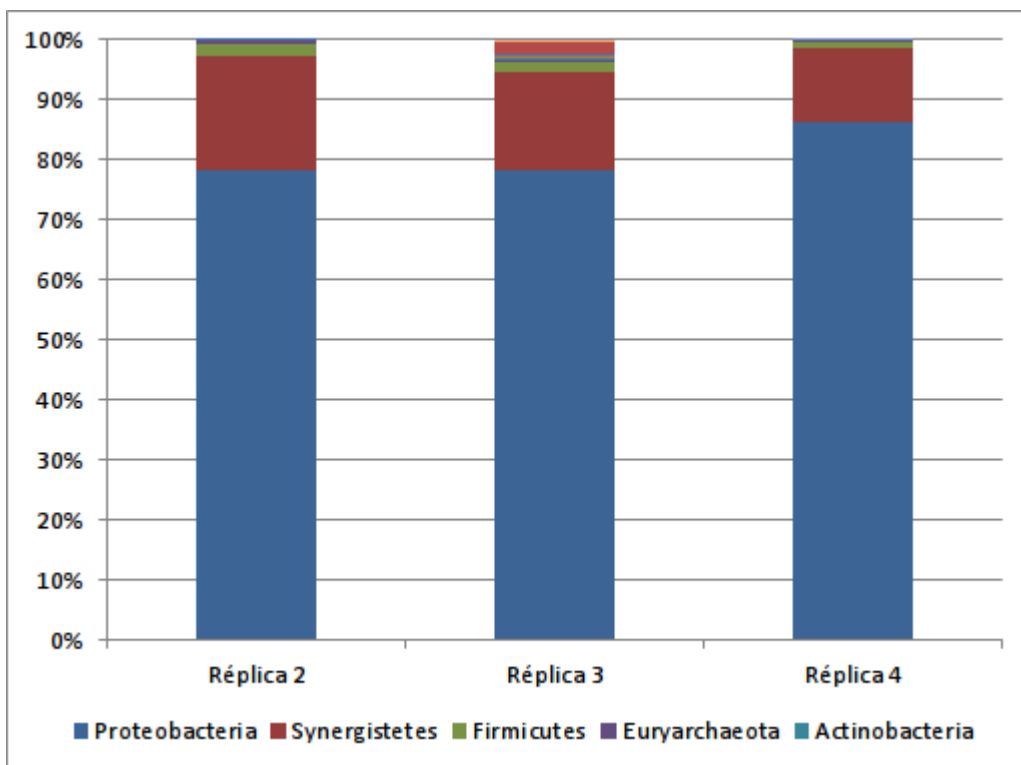


Figura 22. - Gráfica de OTUs a nivel de phylum de las tres réplicas del pozo.

En la figura 22 se muestra la asignación taxonómica obtenida con la base de datos Silva a nivel de phylum de las tres réplicas del pozo; en el pie de gráfico se muestran los 5 phylum más abundantes que se encuentran en la muestra que resultan ser: *Proteobacteria*, *Synergistetes*, *Firmicutes*, *Euryarchaeota* y *Actinobacteria*. Correspondiendo al phylum más abundante entre las tres replicas, se encuentran las *Proteobacterias*, comprendiendo una abundancia entre el 78%-86%. Las *Proteobacterias* resultan ser, como grupo, todas Gram negativas, mostrando una enorme diversidad de mecanismos de generación de energía, incluyendo especies quimiolitótrofas, quimioorganótrofas y fotótrofas, siendo también muy diversas fisiológicamente, en cuyo grupo se encuentran especies anaerobias, micro-aerobias y aerobias facultativas. Entre las *Proteobacterias* se incluyen las bacterias fototróficas rojas, nitrificantes, oxidantes del azufre y el hierro, oxidantes del hidrogeno, metanótrofas y metilótrofas, productoras de ácido acético, entre otras; comprendiendo el grupo metabólicamente más diverso (Madigan *et al.*, 2009). Por otro lado, el segundo phylum más abundante encontrado en las tres réplicas del pozo resulta ser *Synergistetes*, comprendiendo una abundancia del 12%-18%. Las bacterias pertenecientes al phylum *Synergistetes* se han encontrado en un amplio rango de hábitats anaerobios entre los que se incluyen tractos gastrointestinales de animales e insectos, sistemas de tratamiento de agua, suelos y pozos petroleros (Hugenholtz *et al.*, 2009). Los phylum antes mencionados comprenden un rango del 94%-98% de la abundancia presente en las tres réplicas del pozo, siendo, por lo tanto, los dos phylum más importantes obtenidos en las muestras. En menor abundancia se encuentran los phylum *Firmicutes*, *Euryarchaeota* y las *Actinobacterias*, comprendiendo una abundancia del 0.8%-1.8%, 0.3%-0.7% y 0.05%-0.4%, respectivamente.

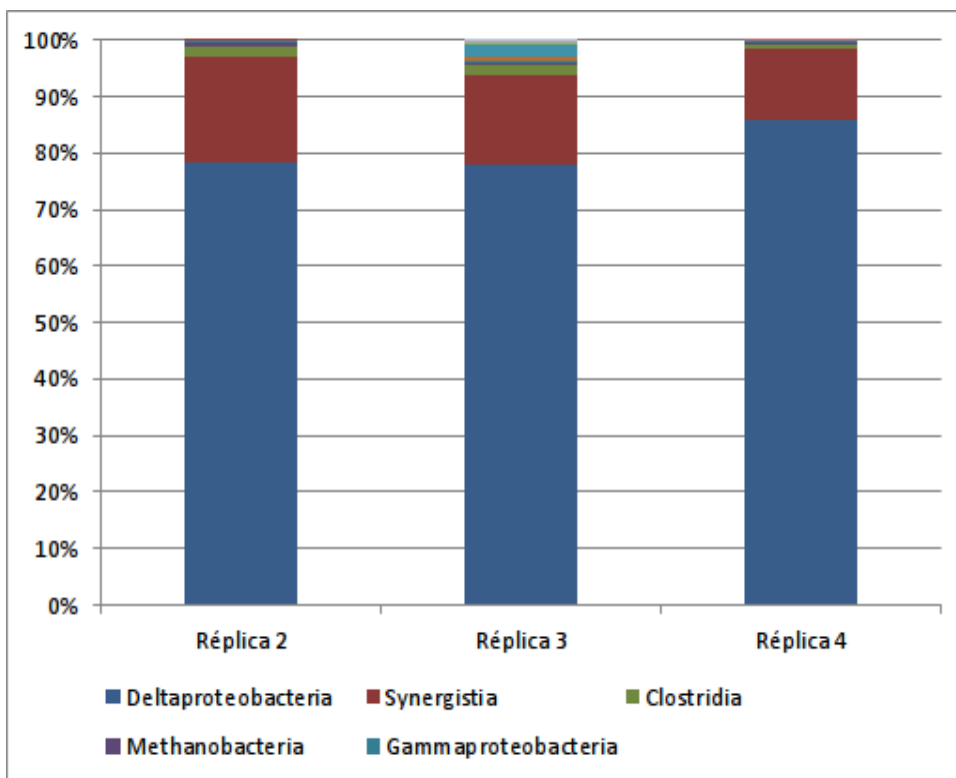


Figura 23. - Gráfica de OTUs a nivel de clase de las tres réplicas del pozo.

En la figura 23 se muestran los distintos OTUs que fueron obtenidos a nivel taxonómico de clase para las tres réplicas del pozo. La clase de bacterias que se encuentra con mayor abundancia comprendiendo un rango del 77%-85% en las muestras pertenece a las *Deltaproteobacterias*, una clase encontrada dentro del grupo del phylum más abundante, las *Proteobacterias*, esta clase comprende bacterias Gram negativas cuyo crecimiento puede ocurrir de manera tanto aerobia como anaerobia (Bergey *et al.*, 2009). Esta clase de bacterias contiene muchos de los géneros conocidos de bacterias reductoras de azufre y sulfato, como también bacterias reductoras de hierro. La segunda clase de bacterias más abundantes en las réplicas pertenece a *Synergistia*, en cuya clase se incluyen las bacterias del tipo anaerobio, en forma de bacilos, las cuales se han aislado principalmente de animales, bioreactores metanogénicos y hábitats terrestres (Bergey *et al.*, 2009). Otra de las clases presentes en las réplicas pertenece a *Clostridia*, cuyas bacterias presentan la característica de ser anaerobias con forma de bacilos, las bacterias de esta clase pueden encontrarse en el suelo, aguas residuales e incluso en los intestinos humanos y animales, incluso es bien sabido que algunas cepas pertenecientes a la clase *Clostridia* tienen papeles benéficos en la industria y el ambiente por la biorremediación de suelos contaminados y la producción de productos químicos de interés industrial (Greenbiologics.com, 2017). Por otro lado, en menor abundancia se encuentran las bacterias pertenecientes a la clase de las *Methanobacterias* y *Gammaproteobacterias*; las bacterias incluidas en la clase de las *Methanobacterias* tienen la característica de ser estrictamente anaerobias y presentan crecimiento oxidando hidrogeno, poseen además la característica de reducir azufre a sulfuro y los hábitats más comunes en los que se encuentran esta clase de bacterias suelen ser sedimentos acuáticos, suelos, digestores anaerobios de aguas residuales y el tracto gastrointestinal de

animales (Bergey *et al.*, 2009). Por último, las *Gammaproteobacterias* fueron encontradas en las muestras con una frecuencia mucho menor que las clases antes mencionadas, comprendiendo una abundancia del 0.18%-0.38%.

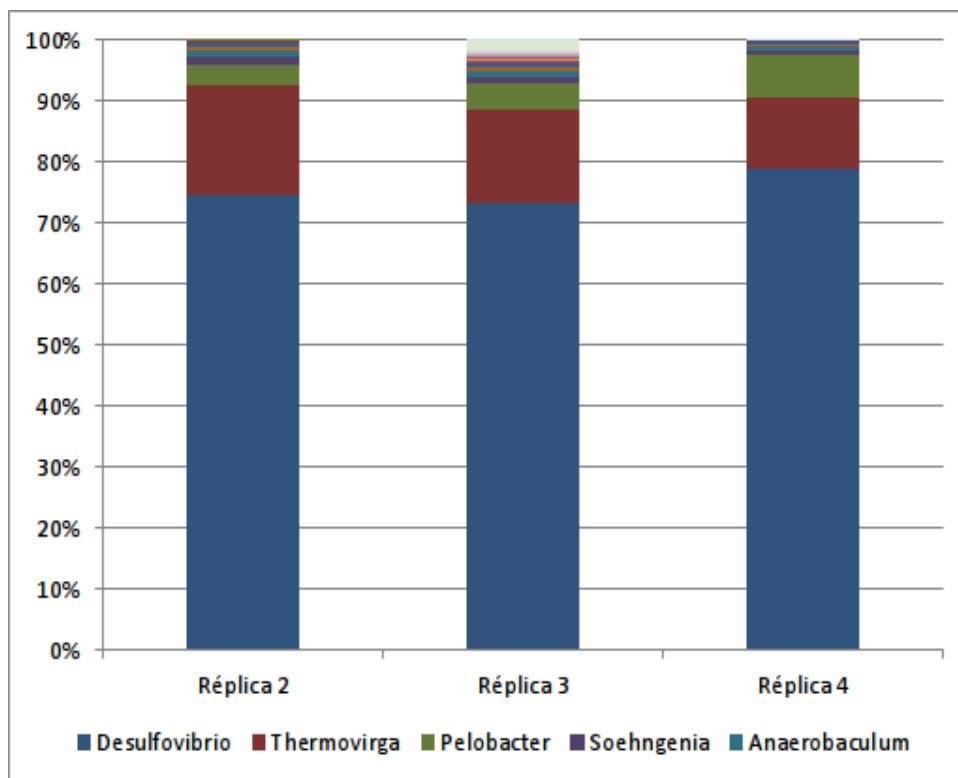


Figura 24. - Gráfica de OTUs a nivel de género de las tres réplicas del pozo.

En la figura 24 se observan los cinco géneros con mayor frecuencia presentes en las tres réplicas del pozo. Comprendiendo el género más abundante en las muestras, con una frecuencia del 71%-78%, se encuentra *Desulfovibrio*, las bacterias englobadas dentro de este género son catalogadas como bacterias reductoras de sulfato; la reducción del sulfato que llevan a cabo dichas bacterias requiere de la inversión de compuestos de gran energía y de distintos intermediarios para que pueda ocurrir la reducción hasta el producto final, sulfuro. El hidrógeno, el formiato, el lactato o el piruvato y muchos otros compuestos orgánicos, incluyendo los componentes del petróleo crudo y el hierro pueden servir como donantes de electrones para la reducción del sulfato a sulfuro de hidrogeno (Voordouw, 1995). Al someter las secuencias al análisis bioinformático para conocer las especies encontradas en las muestras, la especie más abundante con un rango de frecuencia del 70%-78% resulto ser *Desulfovibrio alaskensis*, cuya especie fue descrita por Feio *et al.*, (2004), la cual fue aislada de un pozo petrolero ubicado en Alaska, de acuerdo con las condiciones de crecimiento que fueron establecidas para *D. alaskensis*, se encontró que es una bacteria parcialmente oxidante de lactato, moderadamente halófila y perteneciente a las bacterias reductoras de sulfatos. Otro de los géneros más abundantes encontrados en las muestras resulto ser *Thermovirga*, la cual comprende una abundancia del 11%-17%; dicha especie, de acuerdo con el análisis bioinformático resulto ser una bacteria del phylum *Synergistetes* no cultivada aún. Una de las especies mejor estudiadas pertenecientes al género *Thermovirga*, resulta ser *Thermovirga lienii* que fue aislada de muestras de agua de producción provenientes de un pozo petrolero

ubicado en el mar del Norte y cuyas características la describen como una bacteria anaerobia, termófila y Gram negativa en forma de bacilo (Dahle and Birkeland, 2006). La bacteria descrita por Dahle y Birkeland proviene del mismo phylum y familia que la bacteria encontrada en el presente trabajo. Otro de los OTUs perteneciente al grupo de bacterias que tienen actividad sulfato reductora, aunque en una abundancia menor que *Desulfovibrio alaskensis*, fue una bacteria no cultivada del género *Pelobacter*, comprendiendo una abundancia del 3%-7%, esta puede ser encontrada principalmente en sedimentos marinos y yacimientos petrolíferos, Vigneron *et al.*, 2016 encontraron en sus análisis que más del 88% de las especies encontradas en sus secuencias pertenecían a los géneros *Desulfovibrio* y *Pelobacter*, sugiriendo que la presencia de estos dos géneros puede tener un rol importante en la corrosión encontrada en sus muestras. En menor abundancia, también encontradas en las tres réplicas del pozo, se encuentran las bacterias del género *Soehngenia* y *Anaerobaculum*, descritas como bacterias anaerobias mesófilas.

Las bacterias que fueron encontradas, resultado de la secuenciación con MiSeq, en su mayoría son pertenecientes al grupo de bacterias sulfato reductoras; este tipo de bacterias son conocidas debido a que el producto final de su metabolismo, el sulfuro de hidrogeno resulta ser un compuesto que se ha relacionado con daños a las tuberías petroleras. Normalmente las bacterias reductoras de sulfato y las bacterias productoras de ácido son los dos tipos comúnmente encontrados en tuberías de petróleo y gas (Muthukumar *et al.*, 2003). Por otro lado, también se encontró la presencia de distintas bacterias que metabólicamente realizan distintas funciones y que posiblemente estén creciendo de manera sinérgica con el resto de las bacterias encontradas en las tres muestras de agua del pozo.

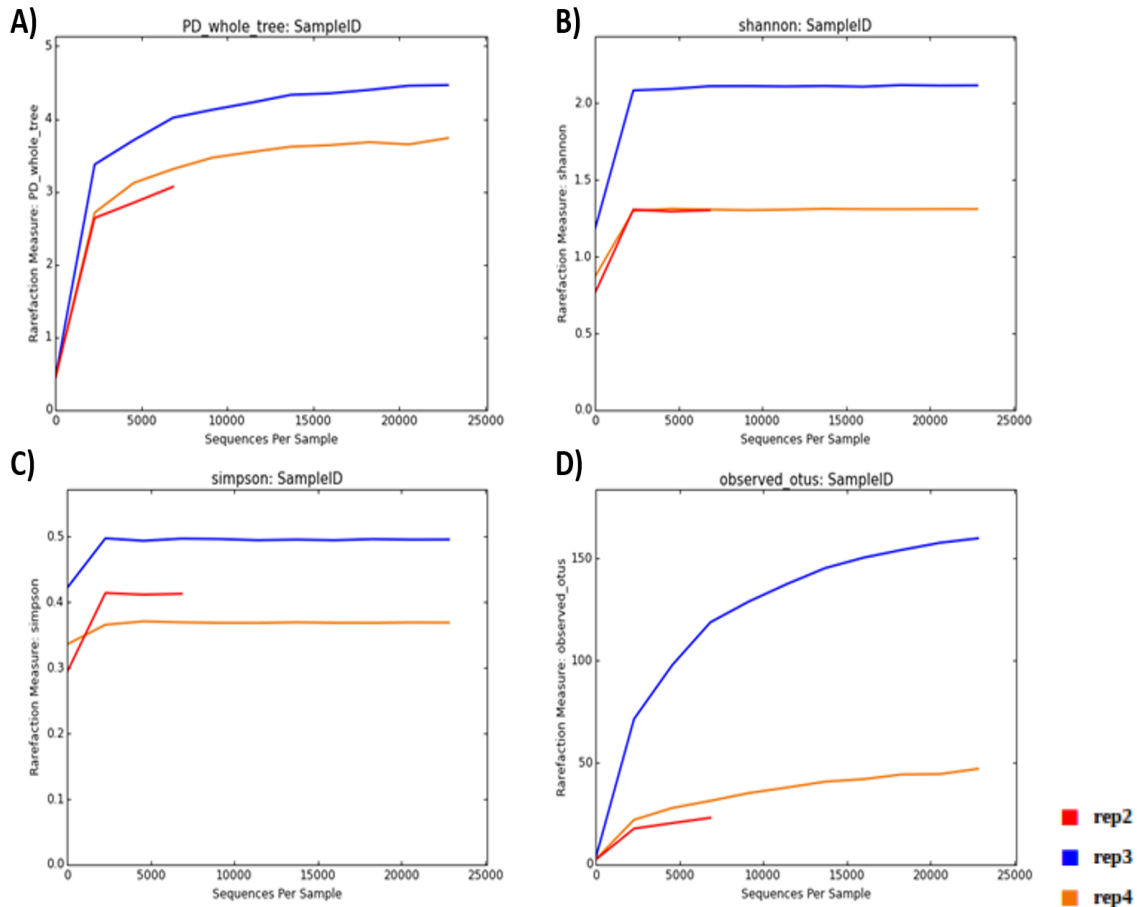


Figura 25. - Gráficas de los índices de A) Diversidad Filogenética (PD), B) Shannon, C) Simpson y D) OTUs observados de las tres réplicas del pozo.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Réplica 2	6856	3.069	1.303	0.413	23
Réplica 3	6856	4.017	2.110	0.497	118
Réplica 4	6856	3.312	1.307	0.370	31

Tabla 25. - Índices de diversidad alfa de las tres réplicas del pozo.

De acuerdo con los gráficos de diversidad presente en las muestras, se puede observar que en cuanto a la abundancia y proporción de OTUs encontrados en las tres réplicas se podría deducir que las réplicas son similares entre ellas; sin embargo, al analizar los índices que demuestran la riqueza presente en las muestras, ocurre un comportamiento totalmente distinto, ya que como lo muestra la figura 25, de manera gráfica se puede observar que las tres réplicas tienen índices de riqueza de especies distintos entre ellas, siendo la réplica 3 la más distinta, mientras que las réplicas 2 y 4 parecen ser las más similares. Por otro lado, de manera numérica, en la tabla 25 se muestran los índices de diversidad alfa de las tres réplicas del pozo, en donde se logra confirmar la similitud en cuanto a riqueza de las réplicas 2 y 4, mientras que la réplica 3 resulta tener la mayor riqueza de las tres de acuerdo con los índices obtenidos.

Para conocer las similitudes entre las réplicas en cuanto a diversidad de géneros y especies asignadas por el programa, se realizaron diagramas de Venn.

En la figura 26 se muestran los diagramas de Venn de los OTUs asignados a nivel de género y especie de las tres réplicas del pozo; al ser las tres muestras provenientes del mismo pozo, es evidente que comparten la mayoría de OTUs asignados a los niveles taxonómicos presentados en el diagrama de Venn; en ambos casos, es la réplica 3 la muestra que representa una mayor diversidad, teniendo 39 géneros y 26 especies distintas a las compartidas entre ellas, mientras que, las réplicas 2 y 4 se mantienen similares. De acuerdo con los diagramas de Venn presentados, es posible deducir que, a pesar de provenir de una misma zona de muestreo y de compartir ciertas similitudes entre las tres muestras, resulta ser la réplica 3 la que presenta una mayor diversidad de géneros y especies, así como también muestra los índices más altos de riqueza de especies de acuerdo con los resultados presentados en la tabla 26; puede suceder que la réplica 3 resulte ser más diversa en comparación con las otras dos, ya que al momento de muestrear, se hayan presentado ciertas dificultades técnicas. De acuerdo con los resultados presentados, se puede deducir que las diferencias entre las tres resultan ser mínimas, por lo que se realizaron correlaciones de Spearman para conocer estadísticamente si entre las secuencias de las réplicas puede existir reproducibilidad técnica.

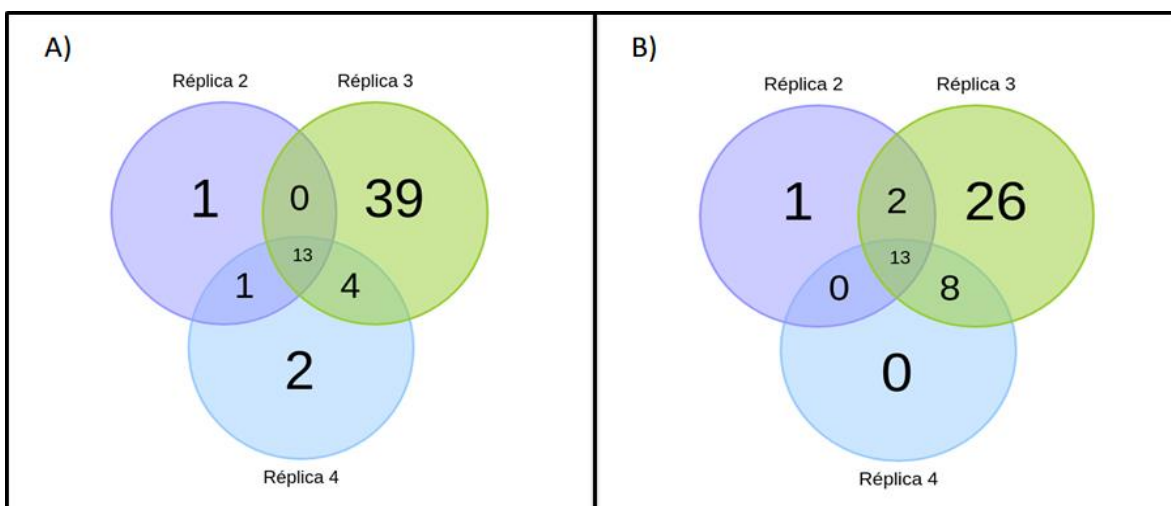


Figura 26.- Diagramas de Venn de A) géneros y B) especies compartidos entre las tres réplicas del pozo.

En la tabla 26 se presentan los resultados de las correlaciones de Spearman obtenidas con el programa R Studio existentes entre las tres réplicas, las cuales van a explicar si existe o no asociación entre las muestras; de acuerdo con las correlaciones presentadas en la tabla, existe una mayor interdependencia entre las réplicas 2 y 4, cuyo coeficiente de correlación resulta ser el más alto en comparación con los demás coeficientes obtenidos; la interdependencia que existe entre estas dos réplicas, ahora demostrada de manera estadística, pudo verse reflejada en los resultados de diversidad taxonómica y riqueza de las especies (diversidad alfa), debido al comportamiento similar presente entre las dos replicas mencionadas, por otro lado, los coeficientes obtenidos entre las réplicas 2, 3 y 3, 4 resultan ser regulares; al obtener coeficientes bajos, podría deducirse que las réplicas que obtuvieron dichos coeficientes tienen una menor interdependencia entre ellas, esto puede explicarse gracias a la diversidad y riqueza encontrada en la réplica 3, cuyos

resultados se comportan de manera distinta a las otras dos muestras. A pesar de haber obtenido correlaciones bajas, debido a los resultados obtenidos por la réplica 3, las interdependencias entre las muestras resultaron positivas, lo que puede indicar una buena reproducibilidad técnica.

Muestra	Réplica 2	Réplica 3	Réplica 4
Réplica 2	1	0.1849	0.5904
Réplica 3	0.1849	1	0.2598
Réplica 4	0.5904	0.2598	1

Tabla 26.- Correlaciones de Spearman entre las tres réplicas del pozo.

Al corroborar que existe reproducibilidad técnica en el muestreo de las tres réplicas del pozo, antes de comparar la diversidad existente entre los dos métodos de secuenciación fue unir las tres secuencias correspondientes a las réplicas para realizar la caracterización general del pozo petrolero, ya que, para fines prácticos de análisis, resulta más conveniente analizar las secuencias como una única muestra, teniendo de respaldo los resultados antes presentados, los cuales confirman que las tres réplicas resultan ser semejantes entre ellas.

6.6 Diversidad Ion Torrent (V2, V3, V4, V6, V7, V8 y V9) vs MiSeq (V3-V4)

Fueron comparados los resultados de diversidad taxonómica y riqueza de especies presentes en el pozo petrolero de Tabasco, que se obtuvieron con los dos métodos de secuenciación; Ion Torrent, cuyos amplicones secuenciados pertenecen a las regiones V2, V3, V4, V6, V7, V8 y V9 del gen 16S y MiSeq cuya secuenciación únicamente toma en cuenta las regiones hipervariables V3-V4.

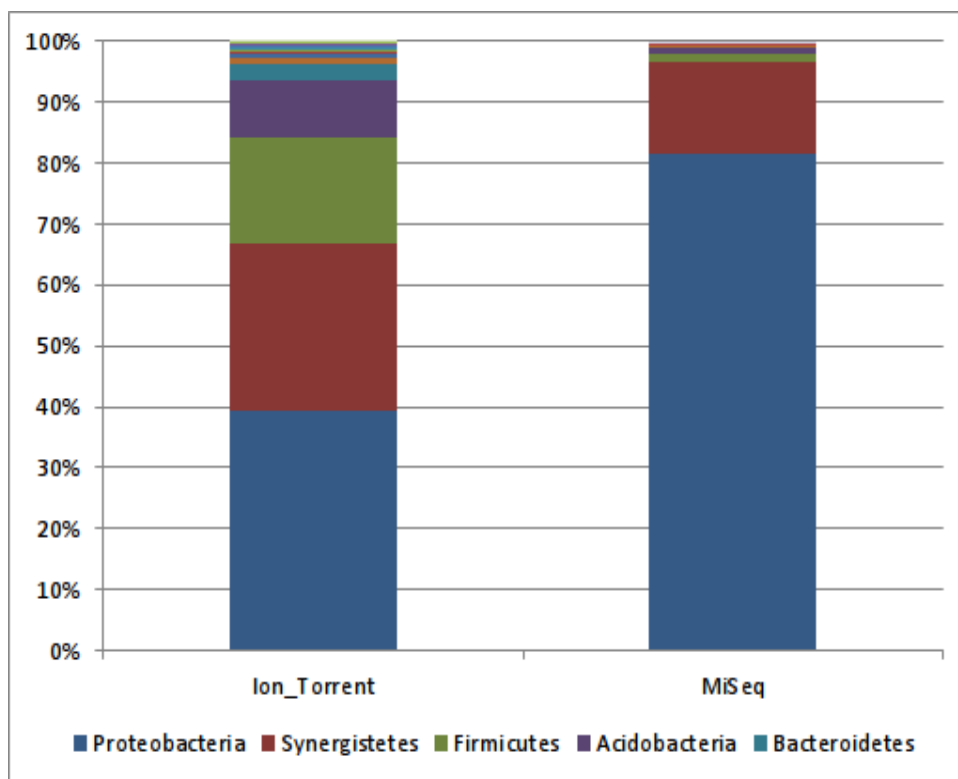


Figura 27. - Gráfica de OTUs a nivel de phylum; Ion Torrent vs MiSeq.

En la figura 27 se muestra la diversidad de OTUs encontrados en el pozo de Tabasco a nivel de phylum, en la cual se comparan los dos métodos de secuenciación con las distintas regiones hipervariables que fueron secuenciadas con cada uno, con Ion Torrent tomando en cuenta 7 regiones y MiSeq 2 regiones; de manera gráfica se pueden observar las notorias diferencias entre ambos métodos, siendo evidente que en la barra que representa la secuenciación con Ion Torrent, se representa un mayor número de OTUs, en comparación con los obtenidos por MiSeq; en total, a nivel de phylum, Ion Torrent logro recuperar de la muestra del pozo un total de 39 OTUs, mientras que con MiSeq, se lograron observar únicamente 20 OTUs.

En cuanto a los phylum encontrados por los dos métodos, los 5 con mayor porcentaje de abundancia resultaron ser los mismos en ambos resultados de secuenciación; el más abundante en ambos resultaron ser las *Proteobacterias*, con distinto porcentaje de abundancia en los dos casos, ya que este phylum se vio notoriamente aumentado en los resultados obtenidos con la secuenciación de MiSeq, abarcando un 81% de abundancia presente en la muestra, mientras que, en el caso de los resultados de la secuenciación con Ion Torrent, las *Proteobacterias* se encuentran en la muestra en una abundancia del 39%. El segundo phylum más abundante pertenece a *Synergistetes*, encontrado con una abundancia del 27% en los resultados de secuenciación con Ion Torrent, mientras que con MiSeq este mismo phylum se encuentra en una abundancia del 14%. Por otro lado, fue encontrada también la presencia del phylum *Firmicutes*, presentando el mismo patrón de comportamiento que los dos phylum antes mencionados, con una mayor abundancia en los resultados obtenidos con Ion Torrent, siendo del 17%, mientras que con MiSeq, fue encontrado con una abundancia del 1.3%.

Por último, se encontró también la presencia de los phylum pertenecientes a las *Acidobacterias* y *Bacteroidetes*, encontrados dentro del top 5 de los más abundantes en los dos métodos de secuenciación, presentando una mayor abundancia en los OTUs obtenidos con Ion Torrent, el mismo comportamiento que presentaron *Synergistetes* y *Firmicutes*.

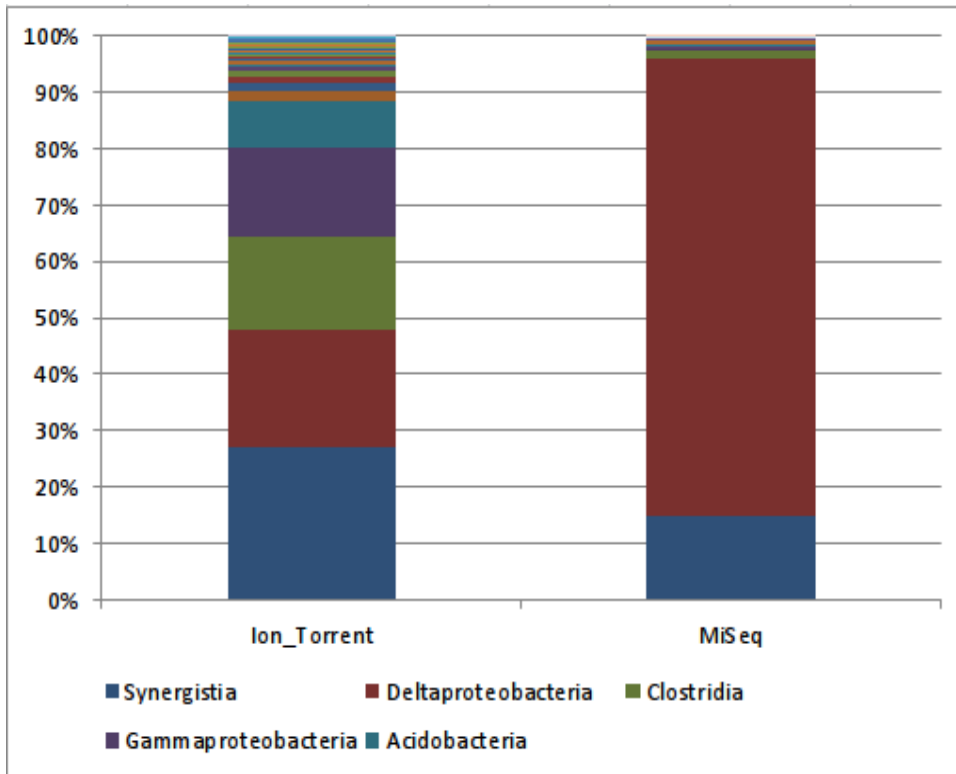


Figura 28. - Gráfica de OTUs a nivel de clase; Ion Torrent vs MiSeq.

En la figura 28 se muestran los OTUs obtenidos con los dos métodos de secuenciación. Dentro de las clases de bacterias más abundantes encontradas con Ion Torrent se encuentran: *Synergistia*, *Deltaproteobacteria*, *Clostridia*, *Gammaproteobacteria* y las *Acidobacterias*, con una abundancia dentro de la muestra del 27%, 20%, 16%, 15% y 8%, respectivamente, representando así el 86% de los OTUs encontrados a nivel de clase encontrados con la secuenciación de Ion Torrent.

Por otro lado, las cinco clases con mayor abundancia encontradas en los resultados de la secuenciación con MiSeq pertenecen a: *Deltaproteobacteria*, *Synergistia*, *Clostridia*, *Methanobacteria* y *Gammaproteobacteria*, con una abundancia dentro de la muestra del 81%, 14%, 1.3%, 0.4% y 0.3%, respectivamente, representando el 97% de los OTUs encontrados en la muestra del pozo secuenciada con MiSeq.

De acuerdo con ambos resultados, fueron encontradas las mismas clases de bacterias con distinta proporción según el método de secuenciación utilizado, presentando el mismo comportamiento que los phylum presentados en la figura 27; en este caso, las *Methanobacterias* se hicieron presentes dentro de las 5 clases más abundantes encontradas con la secuenciación de MiSeq, cuya presencia no se encuentra en los resultados obtenidos con Ion Torrent, aun así, el número de OTUs que lograron ser clasificados al nivel taxonómico de clase con Ion Torrent superan al número de OTUs encontrados con MiSeq, con un total de 94 y 37, respectivamente.

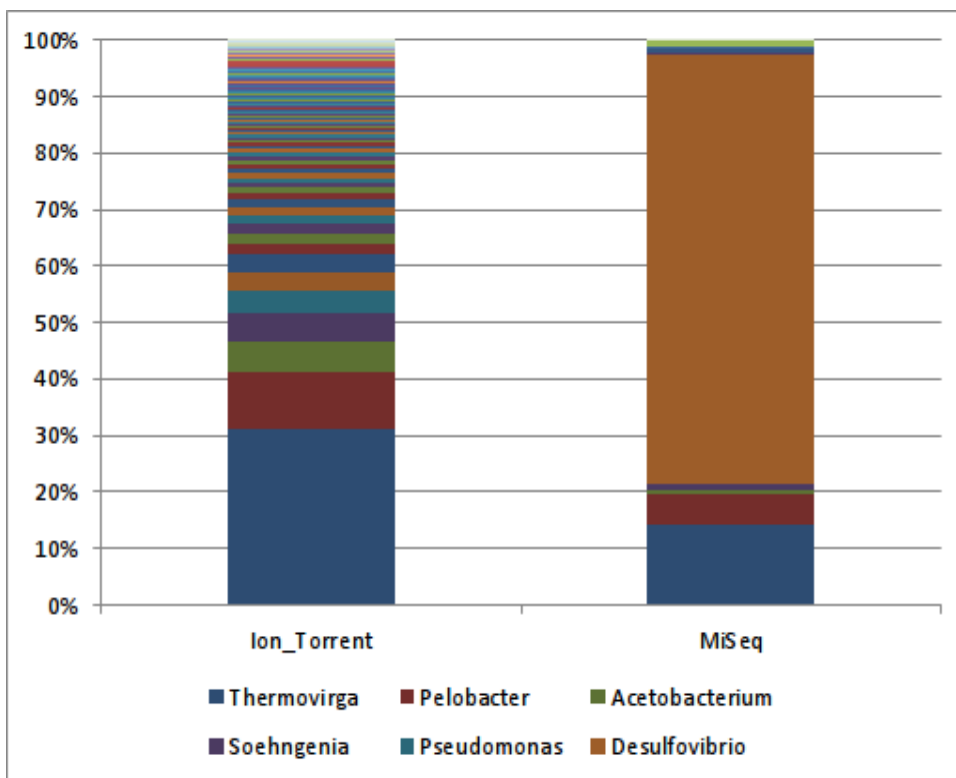


Figura 29. - Gráfica de OTUs a nivel de género; Ion Torrent vs MiSeq.

En la figura 29 se muestran los OTUs obtenidos con los dos métodos de secuenciación, los cuales lograron ser clasificados a nivel de género por el programa QIIME; los 5 géneros que fueron encontrados en mayor abundancia dentro de la muestra del pozo de acuerdo con los resultados de Ion Torrent son los siguientes: *Thermovirga*, *Pelobacter*, *Acetobacterium*, *Soehngenia*, *Pseudomonas* y un OTU identificado como una bacteria que reduce hierro, representando unas abundancias del: 26%, 8%, 4%, 4%, 3% y 2%, respectivamente, representando menos de la mitad de las abundancias de OTUs encontrados en el pozo.

Dentro de los OTUs más abundantes encontrados en los resultados obtenidos con MiSeq se encuentran principalmente: *Desulfovibrio*, *Thermovirga*, *Pelobacter*, *Anaerobaculum* y *Soehngenia* representando unas abundancias del 74%, 14%, 5%, 0.9% y 0.7%, respectivamente, abarcando así el 94% del total de abundancias presentes en la muestra.

De acuerdo con la figura 29, es evidente que de los dos métodos de secuenciación utilizados para analizar la muestra del pozo, resulta ser Ion Torrent con el cual se obtiene un mayor número de OTUs, logrando un total de 545, en comparación con los OTUs obtenidos por MiSeq, siendo un total de 105; del total del número de OTUs asignados a nivel de género por los dos métodos de secuenciación, entre los 5 más abundantes, se encuentran compartiendo 3 de ellos, aunque en distinta abundancia, como es el caso del género *Thermovirga*, que resultó ser el más abundante en los resultados obtenidos con Ion Torrent, mientras que con MiSeq ocupó el segundo lugar, siendo el más abundante para este método el género *Desulfovibrio*, el cual se encuentra también dentro de los OTUs encontrados por Ion Torrent, pero en una abundancia mucho menor que en la encontrada por MiSeq.

Según los OTUs que fueron asignados a los diferentes niveles taxonómicos, de los 5 más abundantes encontrados por los dos métodos de secuenciación, la mayoría se encuentran compartidos, aunque en distintas proporciones dentro de la muestra según el método utilizado; puede ocurrir que las proporciones de los OTUs encontrados cambien debido a la profundidad de secuenciación seleccionada para cada método, así como también al número de regiones hipervariables del gen 16S que fueron secuenciadas.

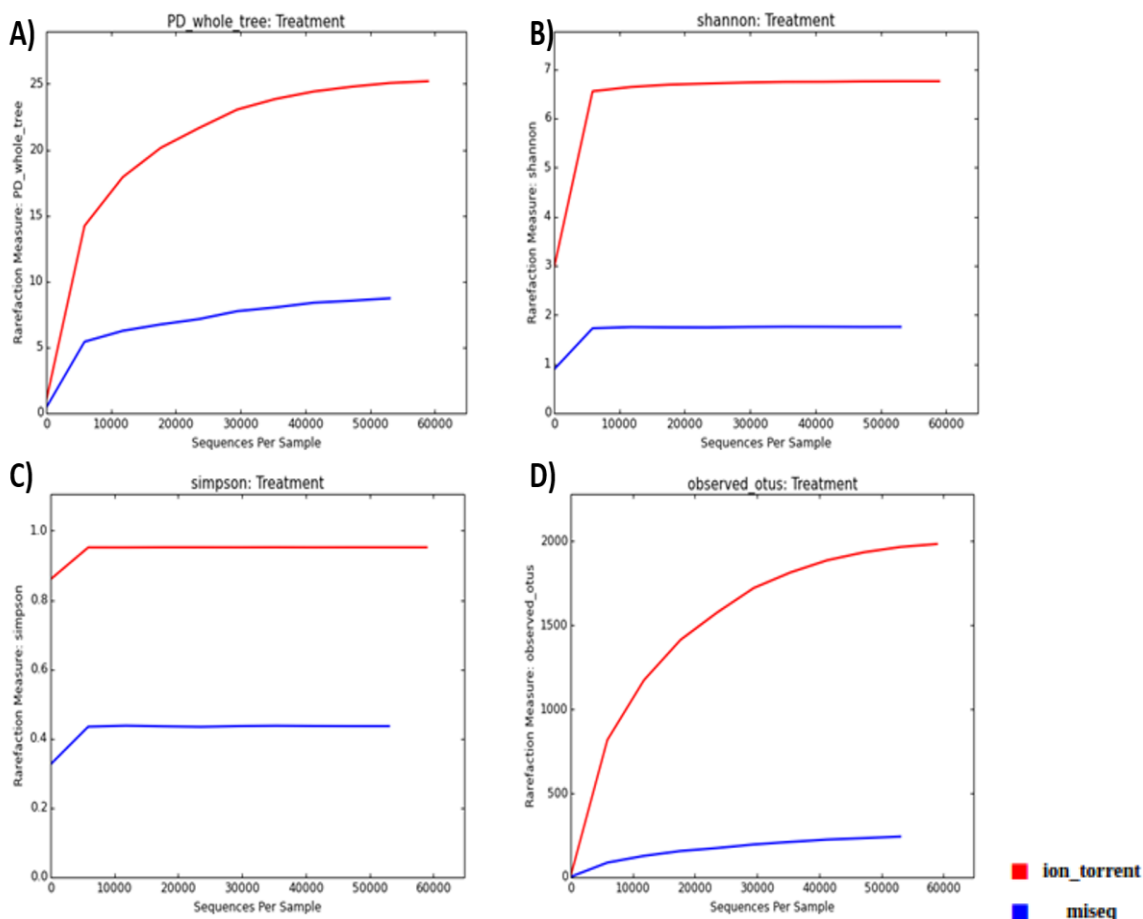


Figura 30. - Gráficas de los índices de A) Diversidad Filogenética (PD), B) Shannon, C) Simpson y D) OTUs observados; Ion Torrent vs MiSeq.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Ion Torrent	53029	25.067	6.754	0.952	1964
MiSeq	53029	8.726	1.757	0.436	242

Tabla 27. - Índices de diversidad alfa Ion Torrent vs MiSeq.

Fue comparada la diversidad con sus distintos índices para los dos métodos de secuenciación utilizados para conocer los microorganismos presentes en las réplicas del pozo; los resultados de la diversidad se presentan de manera gráfica en la figura 30 y los índices se muestran en la tabla 27.

En la figura 30 se muestran las gráficas de los distintos índices de diversidad alfa que fueron comparados, como el índice de PD, Shannon, Simpson y el número de OTUs observados; de acuerdo con las gráficas es posible observar que con Ion Torrent se logra captar una mayor diversidad de la que pudo obtenerse con MiSeq, lo cual puede verse reflejado en los resultados de la tabla 27; fue tomado el mismo número de secuencias por muestra con la finalidad de mantener una misma profundidad de secuenciación y que los resultados de los índices fueran comparables.

En primer lugar, fue tomado en cuenta el índice de diversidad filogenética (PD) el cual es una medida de la biodiversidad que se encuentra presente en la muestra, se puede esperar entonces que valores de PD más grandes pueden corresponder a una mayor diversidad (Faith & Baker, 2006), de acuerdo con ello, podría decirse que según los valores obtenidos, con Ion Torrent es posible ver un espectro más amplio de la diversidad que está presente en la muestra del pozo, ya que el índice calculado para los resultados con Ion Torrent fue de 25, siendo notoriamente mayor que el obtenido con MiSeq, cuyo índice es de 8.

Otros de los índices tomados en cuenta para comparar la diversidad obtenida con Ion Torrent y MiSeq es el índice de Shannon que es utilizado para conocer la composición de la comunidad que está siendo analizada; al igual que en el caso del índice de diversidad filogenética, entre mayor sea el índice, se indica una mayor riqueza de especies que se encuentran presentes en la muestra, por lo que nuevamente resulta ser Ion Torrent el método de secuenciación con un índice de Shannon mayor que el obtenido con MiSeq, indicando por lo tanto que existe una mayor riqueza de especies encontradas en los resultados obtenidos con Ion Torrent.

Fue considerado dentro de las medidas de diversidad alfa el índice de Simpson, el cual es utilizado para cuantificar la biodiversidad presente en cierto hábitat, tomando en cuenta el número de especies presentes, así como la abundancia de cada una, tomando en cuenta también las especies que se encuentran más abundantes dentro de la muestra; en el caso de las muestras del pozo analizadas en el presente trabajo, nuevamente resulta ser Ion Torrent el método de secuenciación con el mayor índice obtenido. Por último, se tomó en cuenta el número de OTUs estimados por el programa para cada uno de los métodos a comparar, siendo evidente que con Ion Torrent se pudo estimar la mayor cantidad de OTUs en comparación con los obtenidos con MiSeq.

De acuerdo con los índices de diversidad alfa que fueron comparados para poder conocer con que método de secuenciación es más probable obtener una mejor estimación de la riqueza y diversidad de especies presentes en la muestra del pozo petrolero, es evidente tanto de manera gráfica como numérica, que con Ion Torrent se puede cubrir u obtener una estimación más cercana de las bacterias que están presentes en las réplicas de agua que fueron muestreadas para el presente trabajo, lo cual puede estar ocurriendo debido a las regiones del gen 16S que se toman en cuenta al realizar la secuenciación con Ion Torrent, mientras que este método toma en cuenta 7 de las 9 regiones, con MiSeq únicamente se secuencian 2 regiones (V3-V4), pudiendo entonces explicar una mayor estimación de la diversidad, ya que al realizar los análisis de diversidad alfa fue posible observar que con los dos métodos se pudo lograr una profundidad de secuenciación similar, por lo que la diferencia entre la diversidad explicada por los dos puede ser bien atribuida como fue mencionado, a las regiones hipervariables secuenciadas.

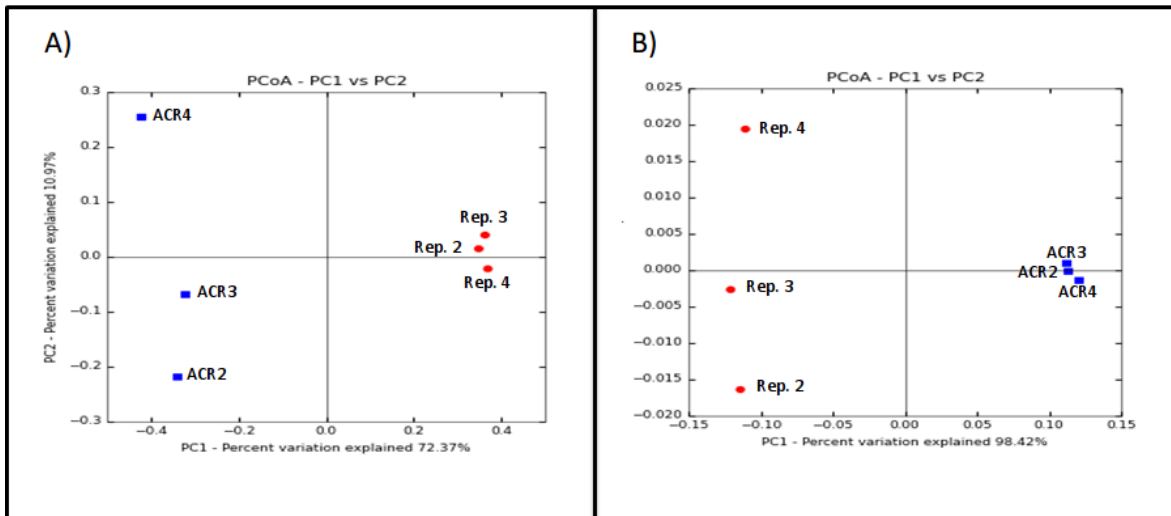


Figura 31.- Gráficas de análisis de coordenadas principales de las secuencias analizadas con Ion Torrent y MiSeq.

Otro de los análisis que fueron considerados para estudiar los microorganismos presentes en las muestras del pozo petrolero de Tabasco fue un análisis de diversidad beta, esto mediante gráficos de coordenadas principales de los OTUs encontrados en las tres réplicas del pozo, de acuerdo con el método de secuenciación utilizado; la diversidad beta es utilizada para medir las diferencias que componen a la muestra analizada (Koleff *et al.*, 2003; Anderson *et al.*, 2006).

Al realizar el análisis de diversidad beta, fueron obtenidas las gráficas de coordenadas principales, en cuyo análisis se utiliza una medida de similitud entre los individuos, partiendo de una matriz de similitud para obtener una representación geométrica de los individuos basada en una distancia que muestran si existe diferencia entre muestras (Cubero & Flores, 2003); en el panel A) de la figura 31 se muestra la gráfica de coordenadas principales no ponderada (medición cualitativa), la cual toma en cuenta la presencia de los OTUs para poder comparar las muestras, y en el panel B) se muestra la gráfica ponderada (medición cuantitativa) que toma en cuenta la presencia de los OTUs y sus abundancias.

Como se puede observar en la figura 31 A), existen dos agrupamientos entre las réplicas del pozo de acuerdo con el método que fue utilizado para ser secuenciadas; los puntos rojos etiquetados como Rep. 2, Rep. 3 y Rep. 4 pertenecen a los resultados de las réplicas que fueron secuenciadas con Ion Torrent; de acuerdo con el análisis realizado, es posible deducir que entre las tres réplicas existe gran similitud, ya que pudieron ser perfectamente agrupadas por el programa, a diferencia de las réplicas que fueron analizadas con MiSeq, etiquetadas como ACR2, ACR3 y ACR4 en cuadrados azules, entre las muestras se puede observar que existe cierta similitud entre ACR2 y ACR3 de acuerdo con las distancias entre ellas, mientras que la réplica ACR4 presenta una distancia mayor en comparación con las dos antes mencionadas. Dado que los resultados no ponderados se utilizan para comparar la composición de la comunidad bacteriana que está siendo analizada, es preciso decir que debido a las distancias entre las dos agrupaciones claramente vistas en la gráfica, y a pesar de pertenecer a una misma comunidad, es decir, las réplicas analizadas pertenecen al mismo punto de muestreo, existe una diferencia entre los dos métodos de secuenciación utilizados, lo cual indica que dependiendo de la plataforma de secuenciación

utilizada hay diferencias notorias entre la composición de las bacterias encontradas dentro del mismo pozo. Por otro lado, la réplica ACR4 secuenciada con MiSeq presenta también una distancia notoria de las réplicas ACR2 y ACR3, quien a pesar de pertenecer al mismo pozo que las dos anteriores y haber sido secuenciada en la misma plataforma, presenta diferencias con los OTUs que fueron encontrados.

En la figura 31 B) se observa la gráfica ponderada del análisis de coordenadas principales, la cual resulta ser una medida cuantitativa, ya que toma en cuenta la presencia del OTU, así como también la abundancia relativa de cada tipo de organismo que se encuentra presente en la muestra; en este caso se observa un patrón de comportamiento totalmente distinto al encontrado en la gráfica no ponderada, ya que resultan ser las muestras analizadas con MiSeq las que se logran agrupar, habiendo distancias cortas entre ellas, indicando entonces que de manera cuantitativa, existe una mayor similitud entre ellas; a diferencia de las muestras analizadas con Ion Torrent, cuyo comportamiento es el inverso al presentado en la gráfica no ponderada (A), ya que en este caso, las tres replicas no presentan un agrupamiento, siendo nuevamente la réplica 4 la que mantiene una distancia mayor en comparación con las otras dos replicas. A pesar de que las medidas cuantitativas y cualitativas de la diversidad se encuentran estrechamente relacionadas al evaluar la abundancia de las especies, empíricamente, estos tipos de medidas a menudo no están correlacionadas y pueden proporcionar puntos de vista diferentes (Lozupone *et al.*, 2007; Stirling & Wilsey, 2001), lo cual resulta ser el caso de la diversidad beta de las réplicas del pozo, ya que presentan comportamientos de manera inversa al realizar los análisis cualitativos y cuantitativos, debido a que puede influir la presencia o ausencia de OTUs, así como las abundancias relativas de los mismos dependiendo de la réplica a analizar.

La diversidad beta ayuda a identificar las similitudes o disimilitudes existentes entre las muestras analizadas; en este caso, sirvió para comprobar las disimilitudes que hay entre las dos plataformas de secuenciación utilizadas para conocer los microorganismos encontrados en el pozo petrolero de Tabasco, ya que al encontrar perfiles distintos entre las comunidades analizadas, fue posible agrupar por separado las réplicas analizadas con Ion Torrent y con MiSeq, por lo que también fue posible conocer lo similares que son entre ellas las réplicas, lo cual está claro que debería ser un resultado esperado, ya que provienen del mismo punto de muestreo siendo replicas técnicas, con la excepción de la réplica 4, la cual en los análisis tanto cuantitativos como cualitativos se mostró más alejada de las réplicas 2 y 3, pudiendo indicar que hay una diferencia entre la composición de la comunidad presente en dicha réplica, pudiendo ser atribuido a fallas técnicas.

De acuerdo con los OTUs obtenidos por los dos métodos de secuenciación, fueron realizados diagramas de Venn a los niveles de género y especie, con la finalidad de conocer cuántos OTUs son compartidos entre métodos de secuenciación y los distintos géneros y especies con los que cuentan cada uno; en la figura 32 A) se muestra el diagrama de Venn de los OTUs asignados a nivel de género dependiendo de la plataforma de secuenciación que fue utilizada, el número de OTUs que comparten entre los dos métodos es de 64, mientras que, indicados en el círculo azul, con MiSeq se obtienen 15 géneros distintos a los compartidos con Ion Torrent, dejando a esta plataforma de secuenciación con 350 géneros diferentes a los compartidos, comprobando una vez más que al secuenciar las muestras del pozo con la plataforma Ion Torrent, es posible describir la mayor diversidad posible de las bacterias que se encuentran presentes en el pozo petrolero.

Por otro lado, se muestra en el panel B) el diagrama de Venn de los OTUs que lograron ser asignados a nivel de especie, presentando el mismo comportamiento que el diagrama de Venn del panel A), ya que la mayoría de especies asignadas por MiSeq se encuentran compartidas entre los dos métodos de secuenciación, dejando el mayor número de especies asignadas diferentes a Ion Torrent, con lo cual es preciso deducir que al abarcar una mayor cantidad de regiones hipervariables del gen 16S, como en este caso, que fueron secuenciadas siete regiones contra dos regiones, es más probable una descripción más cercana de la comunidad bacteriana que se encuentra presente en el pozo petrolero.

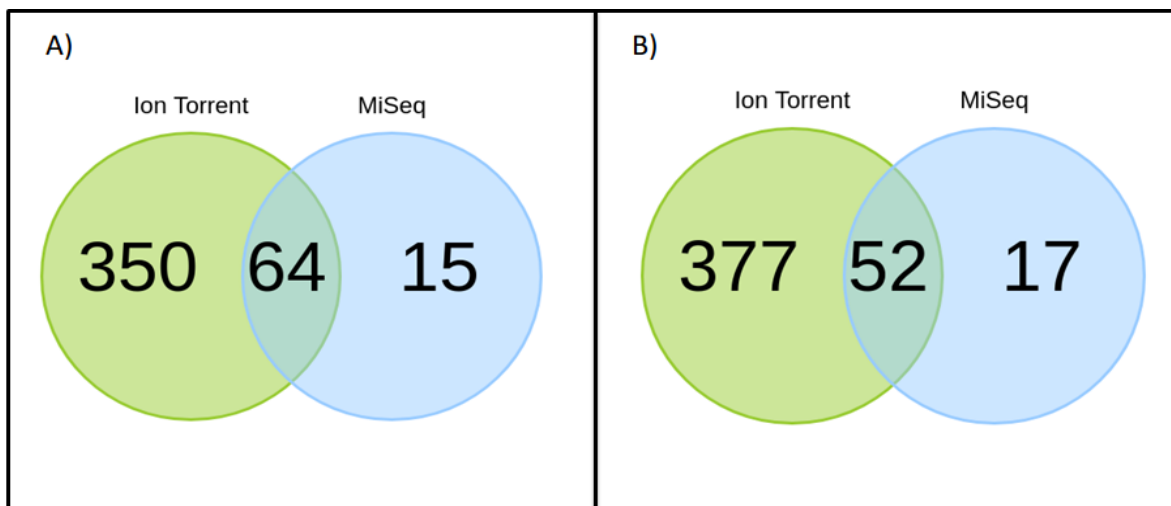


Figura 32.- Diagramas de Venn de A) géneros y B) especies compartidos entre las dos tecnologías de secuenciación.

De acuerdo con los resultados presentados tanto de los OTUs asignados a los distintos niveles taxonómicos, como los resultados de diversidad alfa y beta, se comprueba que con Ion Torrent se observan mejores resultados de la composición microbiana del pozo petrolero; sin embargo, podría decirse que no es posible comparar la secuenciación realizada por las dos plataformas debido al número distinto de regiones hipervariables del gen 16S que se utilizó para cada una, ya que la plataforma MiSeq se vería en una desventaja, por lo tanto, para poder realizar una comparación fue preciso modificar la base de datos que fue utilizada (Silva) con la finalidad de que tomara en cuenta únicamente los OTUs que pueden ser asignados por las regiones hipervariables V3-V4; los resultados con la base de datos modificada son presentados a continuación.

6.7 Comparación entre tecnologías de secuenciación (V3-V4)

6.7.1 Ion Torrent (V2, V3, V4, V6, V7, V8 y V9 vs V3-V4)

Como fue mencionado anteriormente, con tal de poder realizar comparaciones entre los dos métodos de secuenciación utilizados en el presente trabajo, se realizaron modificaciones a la base de datos Silva de tal manera que al someter las secuencias obtenidas con Ion Torrent al programa QIIME, este arrojará resultados únicamente de las regiones V3-V4, las mismas que fueron tomadas en cuenta al realizar la secuenciación con MiSeq. Los resultados presentados a continuación describen la comparación entre las secuencias obtenidas con Ion Torrent al tomar en cuenta las regiones V3-V4 y los resultados de las secuencias cuando se toman en cuenta siete regiones hipervariables del gen 16S.

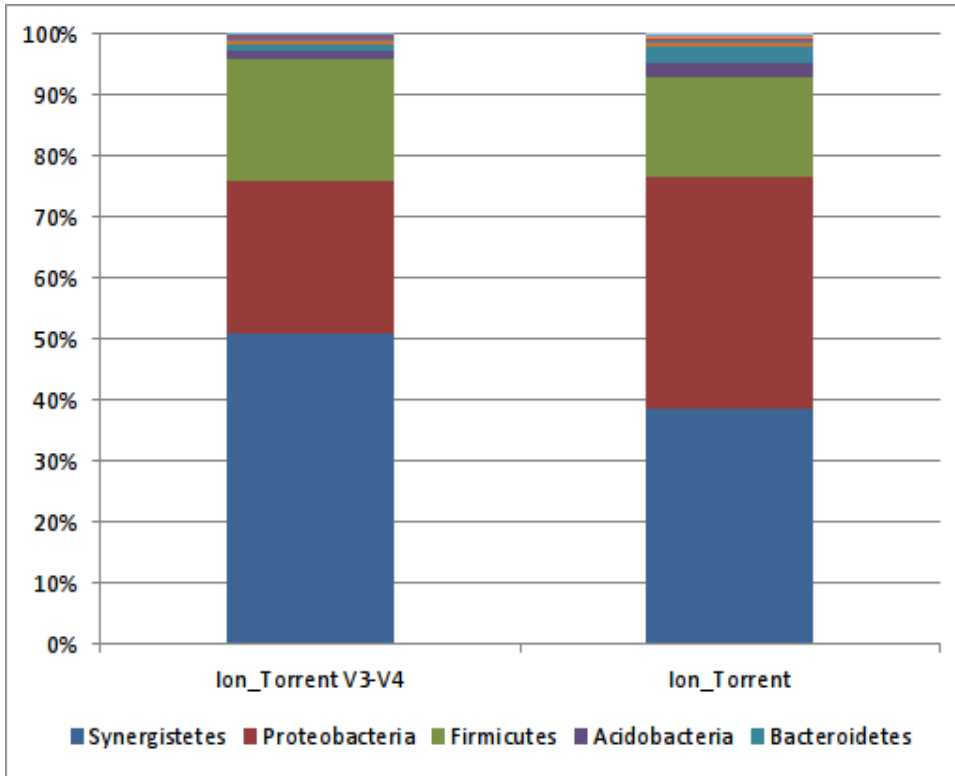


Figura 33.- Gráfica de OTUs a nivel de phylum comparando las secuencias de Ion Torrent V3-V4 y Ion Torrent V2, V3, V4, V6, V7, V8 y V9.

En la figura 33 se muestra la gráfica de OTUs que fueron asignados a nivel de phylum, comparando los resultados de la base de datos modificada con las regiones V3-V4 y la base de datos completa; el número de OTUs asignados a nivel de phylum para la base de datos modificada es de 17, mientras que los OTUs obtenidos con la base de datos completa resulta ser de 21, asignando 4 OTUs distintos a los que se obtuvieron con la base de datos cortada, por otro lado, los 5 OTUs más abundantes resultaron ser los mismos para los análisis con ambas bases de datos, de manera gráfica se puede observar que las proporciones de los phylum encontrados varían entre las dos bases de datos; como el phylum más abundante en las dos muestras, *Synergistetes* con una abundancia del 50% y del 38% en los resultados de la base de datos modificada y la base de datos completa, respectivamente, ocurriendo el mismo comportamiento para los 4 restantes phylum más abundantes.

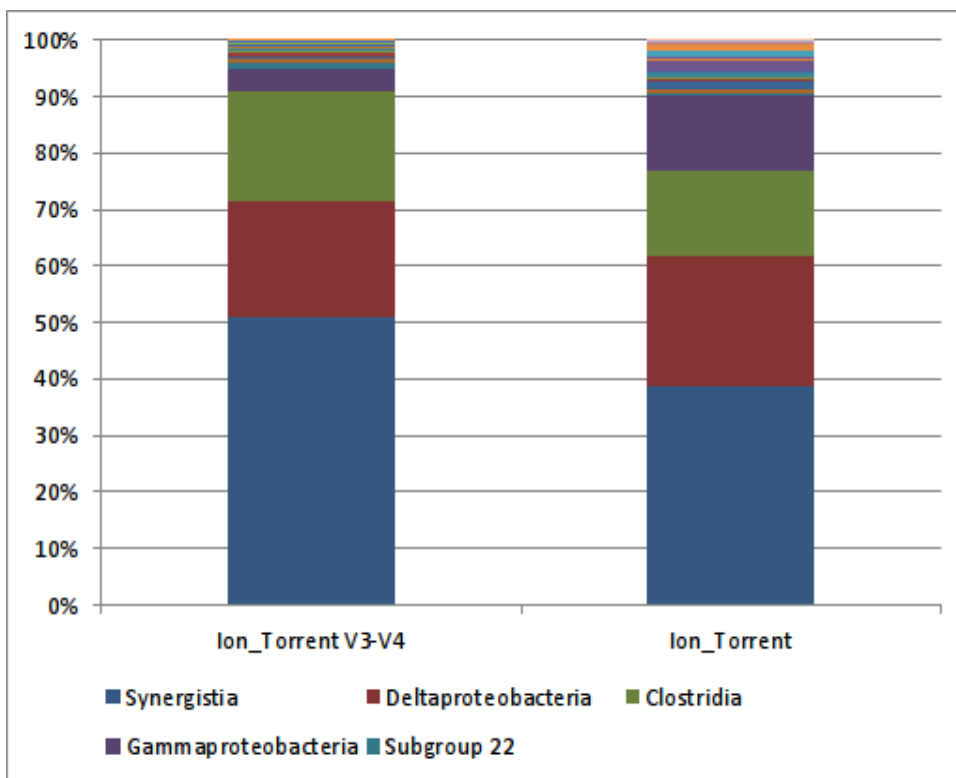


Figura 34.- Gráfica de OTUs a nivel de clase comparando las secuencias de Ion Torrent V3-V4 y Ion Torrent V2, V3, V4, V6, V7, V8 y V9.

En la figura 34 se presenta la gráfica de abundancias de los OTUs correspondientes a las clases encontradas en la muestra del pozo; las cinco clases más abundantes encontradas al comparar la base de datos modificada y completa resultan ser las mismas, aunque como puede observarse en la gráfica, las abundancias relativas cambian de acuerdo con la base de datos que fue utilizada para el análisis. La clase *Synergistia* que está presente en los dos resultados, y que resulta ser la más abundante en ambos casos se encuentra en una abundancia del 50% y 38% con la base de datos modificada y la base de datos completa, respectivamente, dicho comportamiento ocurre con las demás clases que se encuentran compartidas en ambas, esta conducta de las abundancias podría ser atribuida al número de OTUs encontrados por cada una de las bases de datos, ya que la base de datos que toma en cuenta únicamente las regiones V3-V4 logro asignar a nivel de clase 30 OTUs, mientras que se asignaron 45 OTUs con la base de datos que toma en cuenta siete regiones del gen 16S, por lo que puede ocurrir que al momento de asignar OTUs a los distintos niveles, se cambien las abundancias entre mayor o menor sea el número de OTUs asignados, lo que no cambia por el programa es la asignación, ya que la mayoría de las clases se encuentran compartidas entre ambos análisis.

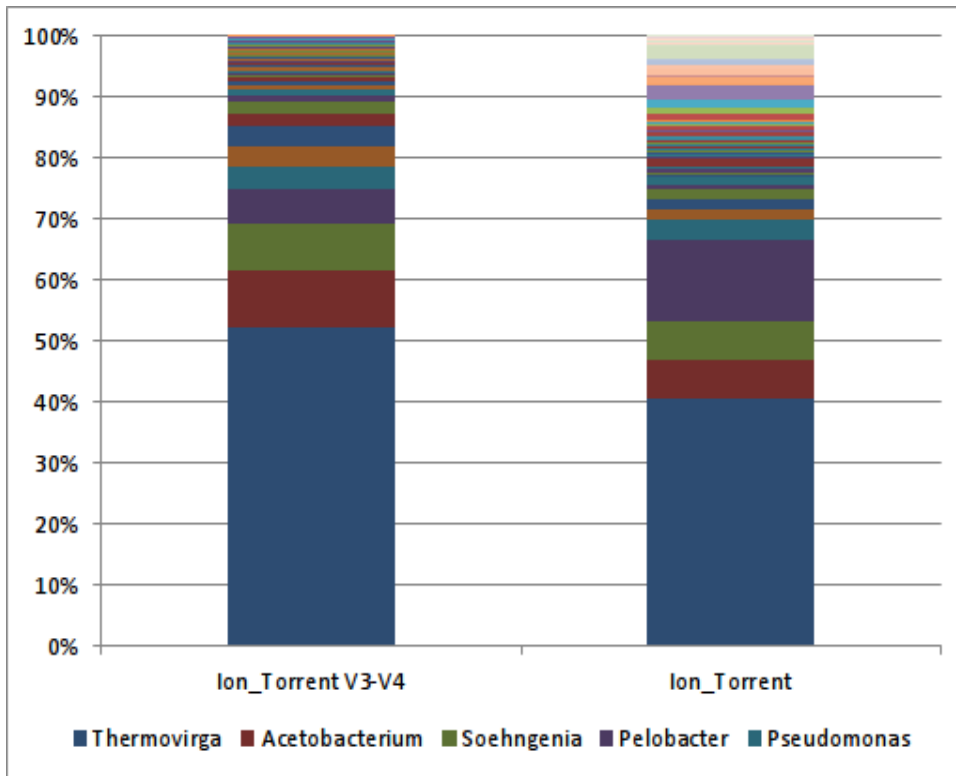


Figura 35.- Gráfica de OTUs a nivel de género comparando las secuencias de Ion Torrent V3-V4 y Ion Torrent V2, V3, V4, V6, V7, V8 y V9.

Por último se presenta en la figura 35 la gráfica de géneros encontrados al comparar las bases de datos modificada y completa; se puede distinguir que con la base de datos completa, es decir, la base de datos con la que se analizaron las siete regiones hipervariables del gen 16S, es posible observar una mayor cantidad de barras que indican la presencia de distintos OTUs, a diferencia de los encontrados con la base de datos modificada, lo cual se comprueba realizando el conteo de géneros asignados por cada una, siendo de 144 para el análisis con Silva que considera las regiones V3-V4 y 281 OTUs para la base de datos completa; se puede observar entonces un comportamiento similar en cuanto a asignación de OTUs que el observado cuando fueron comparados los dos métodos de secuenciación, ya que con Ion Torrent se obtuvo la mayor diversidad posible que pudiera explicar la comunidad total de la muestra del pozo; una explicación por la cual se obtiene un mayor número de OTUs cuando la base de datos Silva no es modificada es nuevamente debido al número de regiones hipervariables que fueron secuenciadas y que son tomadas en cuenta al momento de asignar taxonomía en la muestra.

Además de contemplar con cuál de las dos bases de datos utilizadas en el análisis se podría obtener una mejor asignación taxonómica a los distintos niveles, fue tomada en cuenta también la diversidad alfa de los resultados obtenidos con ambas bases de datos; en la tabla 28 se muestran los índices de los resultados de diversidad alfa, fue tomado el mismo número de secuencias por muestra para que pudieran ser comparables entre ellas. En primer lugar, se comparó el índice de diversidad filogenética, el cual como se puede observar en la tabla, resulta ser muy similar para ambas bases de datos, habiendo muy poca diferencia entre ellos, ocurriendo un patrón similar para los índices de Shannon y Simpson, lo cual podría indicar que a pesar de que la base de datos

modificada logre asignar un menor número de OTUs desde phylum hasta especie, la diversidad entre las muestras se mantiene de manera similar que cuando los análisis son realizados con la base de datos completa, a pesar de la gran diferencia que existe en cuanto a los OTUs asignados, los cuales se observan en la tabla 28, en donde la base de datos completa logra asignar más del doble de OTUs que la base de datos modificada.

Los resultados de diversidad alfa que arrojo el análisis con ambas bases de datos indican que a pesar de las diferencias que existen al asignar taxonomía desde phylum hasta especie, la diversidad que se obtiene con ambas bases de datos resulta ser muy similar, pudiendo indicar que no importa el número de OTUs asignados por cada una, con cualquiera de las dos bases de datos utilizada se conserva la diversidad que está presente en la muestra del pozo.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Ion Torrent V3-V4	31450	11.283	4.083	0.827	329
Ion Torrent	31450	11.169	5.563	0.913	856

Tabla 28.- Índices de diversidad alfa Ion Torrent V3-V4 vs Ion Torrent V2, V3, V4, V6, V7, V8 y V9.

Por último, fueron realizados diagramas de Venn de los géneros y especies que son compartidos entre las bases de datos comparadas, en el panel A) se observa el diagrama de los géneros compartidos entre ambas; como se puede observar, la mayoría de los OTUs que asigno la base de datos modificada se encuentran compartidos con los asignados por la base de datos completa, esta misma cuenta con una mayor cantidad de OTUs distintos a los compartidos, lo mismo ocurre cuando son comparadas las especies que lograron ser asignadas por cada una de las bases de datos, lo cual había sido observado tanto en las gráficas de abundancias de OTUs, como en la comparación de taxonomía asignada entre los dos métodos de secuenciación; Ion Torrent y MiSeq, lo cual podría indicar que al modificar la base de datos que fue utilizada (Silva), realmente se tomaron en cuenta únicamente las regiones hipervariables V3-V4 de acuerdo con los resultados presentados; por lo que, antes de poder realizar la comparación entre las plataformas de secuenciación, fue preciso realizar una comparación entre los datos secuenciados con MiSeq utilizando las mismas bases de datos, la modificada y la completa para conocer qué tan similares son y si en realidad la base de datos modificada toma en cuenta las regiones hipervariables V3-V4.

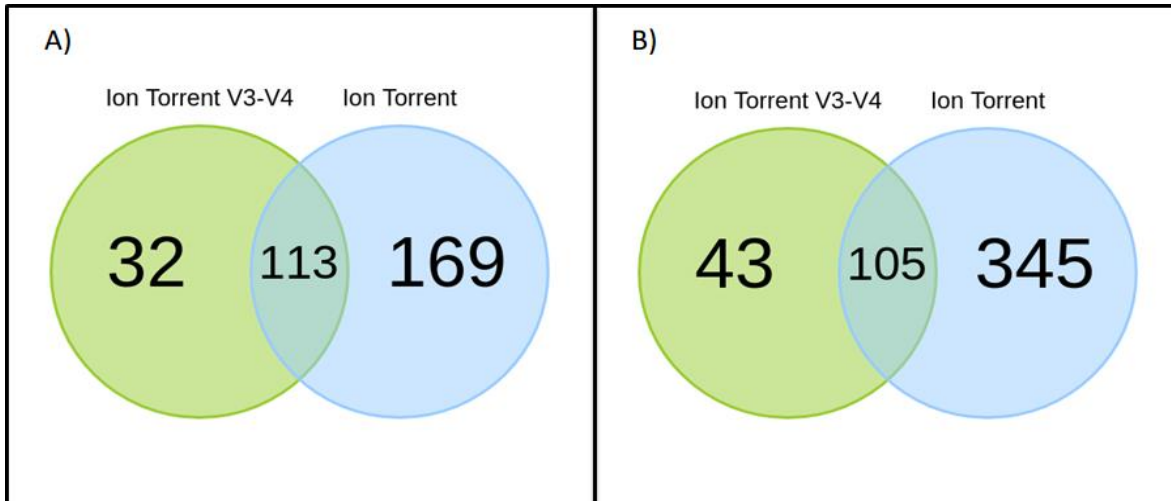


Figura 36- Diagrama de Venn de A) géneros y B) especies compartidas entre las secuencias analizadas con Ion Torrent (V3-V4) y Ion Torrent (V2, V3, V4, V6, V7, V8 y V9).

6.7.2 MiSeq (Silva completo vs Silva V3-V4)

Antes de realizar las comparaciones entre plataformas de secuenciación, fue preciso conocer si la base de datos que se modificó realmente tomaba en cuenta las regiones hipervariables V3-V4, por lo que esta misma fue utilizada en los datos de secuenciación con MiSeq, realizando el análisis bioinformático con ambas bases de datos, tal como fue realizado con los resultados de Ion Torrent, por lo que entre más similares sean los resultados de taxonomía y diversidad de MiSeq, significaría que la base de datos fue modificada de manera correcta y específica para las regiones antes mencionadas, los resultados son presentados a continuación.

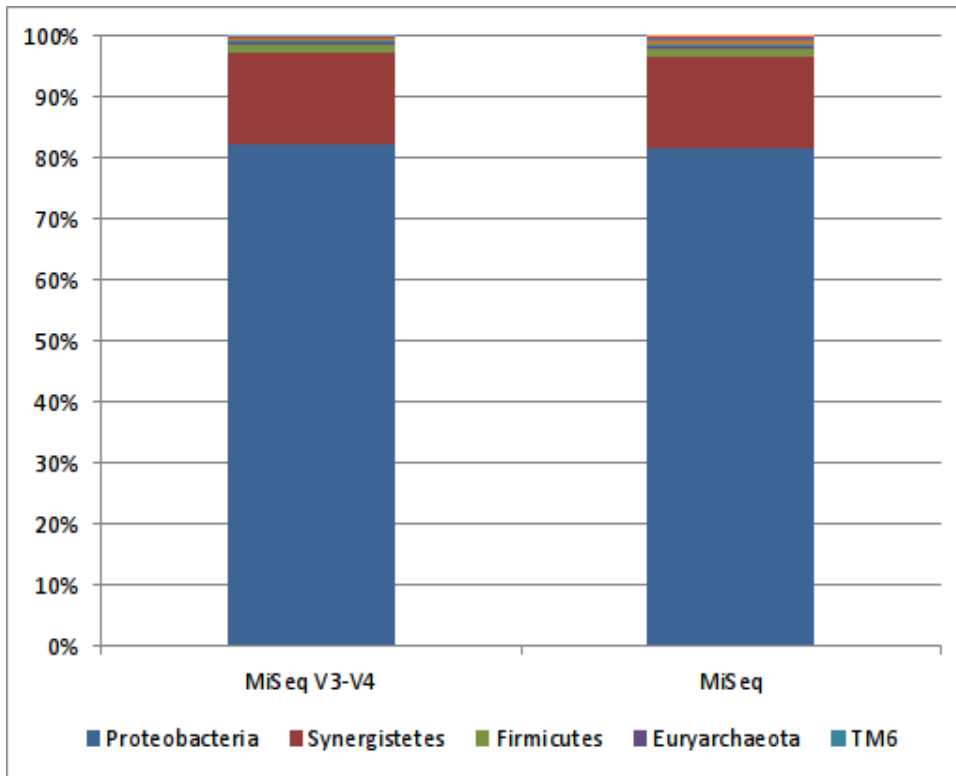


Figura 37.- Gráfica de OTUs a nivel de phylum comparando la base de datos Silva con y sin cortar las regiones V3-V4.

En la figura 37 se muestra la gráfica de abundancias relativas de los OTUs obtenidos a nivel de phylum al comparar la base de datos modificada y la base de datos completa; visualmente se puede observar que ambas graficas presentan cierto patrón de similitud en cuanto a los OTUs encontrados, así como los cinco phylum más abundantes resultan ser los mismos para ambos resultados, en este caso, las abundancias de los OTUs encontrados en ambos casos son también muy similares, ya que el phylum de las *Proteobacterias*, encontrado como el más abundante representa una abundancia del 82% y 81% en los resultados obtenidos con la base de datos modificada y la base de datos completa, respectivamente, ocurriendo lo mismo con los phylum más abundantes restantes; por otro lado, en cuanto al número de phylum asignados por cada una, con la base de datos modificada se obtuvieron 17, mientras que con la base de datos completa fueron asignados 18. De acuerdo con la figura 37, a nivel de phylum son muy similares los resultados obtenidos por ambas bases de datos.

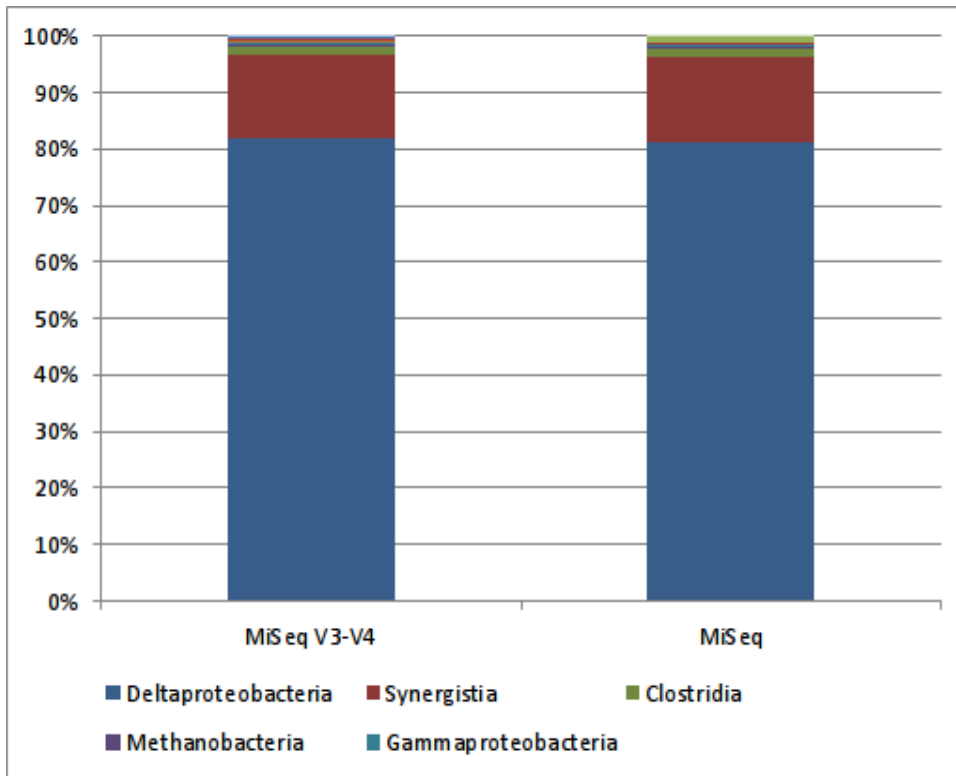


Figura 38.- Gráfica de OTUs a nivel de clase comparando la base de datos Silva con y sin cortar las regiones V3-V4.

En la figura 38 se observa la gráfica de las clases que fueron asignadas por ambas bases de datos, al igual que en la figura anterior, se observa también un patrón de similitud entre los resultados presentados tanto en las 5 clases más abundantes, como en el número de OTUs asignados por ambas, siendo de 31 OTUs asignados a nivel de clase con las dos bases de datos comparadas; de acuerdo con las cinco clases más abundantes, todas son compartidas entre los dos resultados, siendo *Deltaproteobacteria*, *Synergistia*, *Clostridia*, *Methanobacteria* y *Gammaproteobacteria* las principales clases cuya abundancia ocupa más del 90% de la muestra del pozo, además de compartir el top 5 de clases, entre las dos comparten las mismas abundancias dentro de la muestra, por lo que, al comparar a nivel de clase las dos bases de datos, resulta ser que con ambas se obtienen los mismos resultados, pudiendo indicar que la base de datos que fue modificada al ser sometida a los análisis de taxonomía, efectivamente arroja resultados correspondientes a las regiones hipervariables V3-V4.

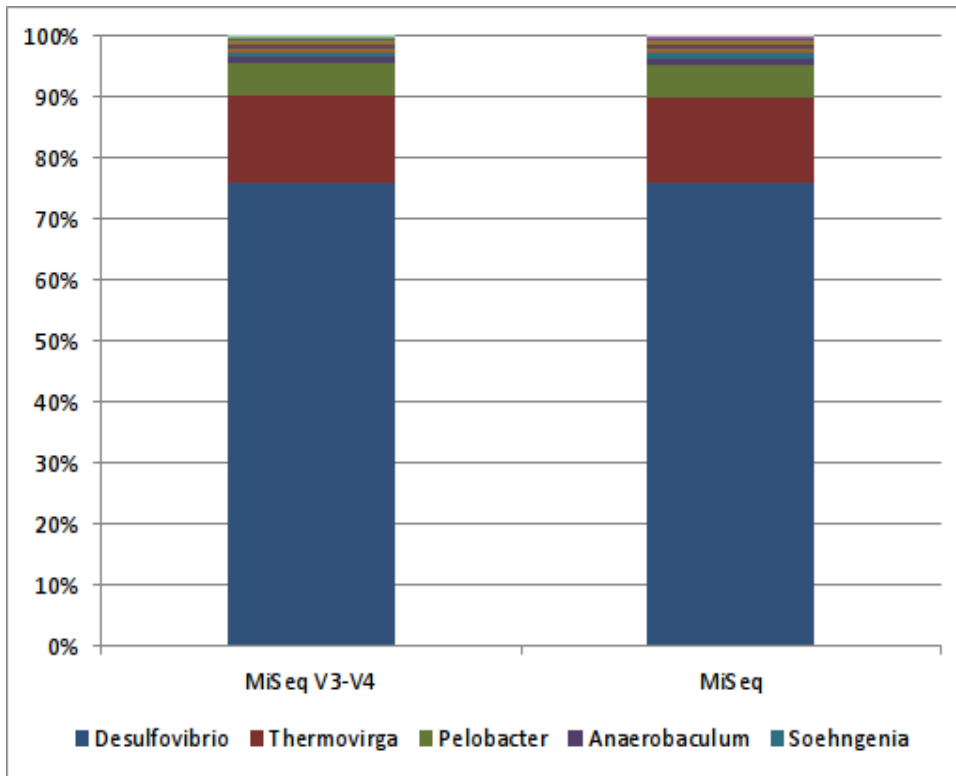


Figura 39.- Gráfica de OTUs a nivel de género comparando la base de datos Silva con y sin cortar las regiones V3-V4.

Por último, en la figura 39 se muestran los géneros asignados por ambas bases de datos; al igual que en los demás niveles taxonómicos, los dos análisis muestran similitud entre ellos, ya que nuevamente los cinco géneros más abundantes dentro de la muestra resultan ser los mismos, encontrados en abundancias similares, lo cual puede verse de manera gráfica en la figura 38, ya que ambas barras presentan un patrón similar; en cuanto al número de OTUs, a nivel de género hubo un comportamiento distinto al presentado en los niveles taxonómicos anteriores, ya que en este caso, con la base de datos que fue modificada se obtuvieron un total de 83 OTUs, mientras que con la base de datos completa fueron asignados 77 OTUs.

De acuerdo con los resultados de taxonomía al comparar las dos bases de datos, es evidente que con las dos se obtienen resultados muy similares, esto gracias a las modificaciones que fueron realizadas a Silva, haciendo cortes en la misma de tal manera que la taxonomía encontrada en cualquier muestra a analizar fuera únicamente la correspondiente a las regiones hipervariables V3-V4 del gen ribosomal 16S, como fue mencionado, en cuanto a taxonomía los resultados indican que la base de datos modificada corresponde efectivamente a las regiones V3-V4, por lo que el siguiente paso fue comparar si al verificar la diversidad alfa de las muestras se obtenían los mismos resultados.

En la tabla 29 se muestran los resultados de diversidad alfa que fueron obtenidos al comparar las bases de datos modificada y completa, fue tomado el mismo número de secuencias por muestra con la finalidad de que los datos fueran comparables entre ellos y se tomaron en cuenta los índices de diversidad filogenética, Shannon, Simpson y el número de OTUs que fueron asignados

para cada muestra. De acuerdo con lo observado en la tabla, los índices de PD, Shannon y Simpson presentan cierta similitud entre ellos, ya que las diferencias entre los índices obtenidos con la base de datos modificada y completa son mínimas, con lo cual podría decirse que la diversidad que se presenta en ambos casos es casi la misma, sin importar la base de datos que este siendo utilizada; lo mismo ocurre cuando son comparados el número de OTUs asignados por cada base de datos, resultan en una diferencia mínima de dos OTUs, ya que la base de datos modificada cuenta con 167 OTUs y la base de datos completa cuenta con 165 OTUs, por lo que con Silva modificada, se está obteniendo un mayor número de OTUs y también un índice de diversidad filogenética más alto que con la base de datos sin modificaciones; de acuerdo con estas observaciones, se podría decir incluso que con Silva modificada se obtienen mejores resultados en cuanto a diversidad alfa presente en las muestras del pozo.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
MiSeq V3-V4	53960	5.158	1.678	0.425	167
MiSeq	53960	4.586	1.735	0.434	165

Tabla 29.- Índices de diversidad alfa de las secuencias de MiSeq comparando la base de datos Silva con y sin cortar las regiones V3-V4.

Al realizar las comparaciones entre las dos bases de datos, se tomaron en cuenta también los diagramas de Venn de los géneros y especies que resultaron compartidas entre ambas, en la figura 40 A) se observa el diagrama de los géneros obtenidos por ambas bases de datos; como se puede ver, la mayoría de OTUs son compartidos por las dos, sin embargo, a pesar de compartir la mayor parte de los géneros, la base de datos que fue modificada cuenta con 15 géneros diferentes, y la base de datos completa únicamente cuenta con 9 distintos; lo mismo pasa cuando se toman en cuenta las especies asignadas, la mayoría se encuentran compartidas entre ambas, pero con Silva modificada se encuentran más especies distintas a las encontradas, por lo que en cuanto a asignación de taxonomía, resulta ser mejor la base de datos modificada, ahora bien, este comportamiento se presentó únicamente en los niveles de género y especie, ya que los niveles taxonómicos anteriores presentaron el mismo comportamiento, siendo muy similares en el número de OTUs asignados y en las abundancias en las que se encontraban dentro de la muestra.

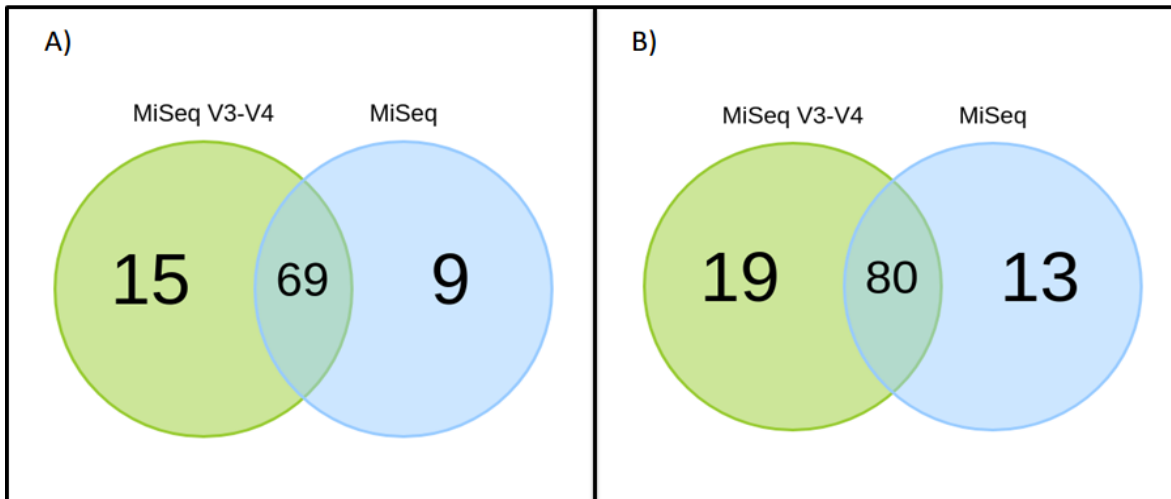


Figura 40.- Diagramas de Venn de A) géneros y B) especies compartidas comparando la base de datos Silva con y sin cortar las regiones V3-V4.

Se puede concluir que, de acuerdo con los resultados de diversidad alfa, se puede observar que los cambios entre ambas comparaciones son mínimas, y que incluso podría resultar mejor utilizar la base de datos que fue modificada. De acuerdo con las pruebas de comparación que fueron realizadas con los datos de secuenciación de la plataforma MiSeq y ambas bases de datos, puede concluirse que no existe diferencia significativa entre Silva con modificaciones y Silva completa, ya que con ambas se obtuvieron resultados muy similares lo cual indica que el corte realizado a la base de datos fue efectivamente correspondiente con las regiones hipervariables V3-V4 del gen 16S y que se pueden entonces realizar comparaciones entre las dos plataformas de secuenciación que son Ion Torrent y MiSeq tomando en cuenta las dos regiones antes mencionadas, ya que como se observó en resultados anteriores, al comparar 7 regiones contra 2 regiones pudo resultar desfavorable para la plataforma de MiSeq.

Los resultados que comparan la diversidad obtenida con Ion Torrent y MiSeq tomando en cuenta las regiones hipervariables V3-V4 del gen ribosomal 16S y que podrían indicar la presencia de corrosión en las tuberías del pozo petrolero de Tabasco se presentan a continuación.

6.7.3 Diversidad taxonómica V3-V4 Ion Torrent vs MiSeq

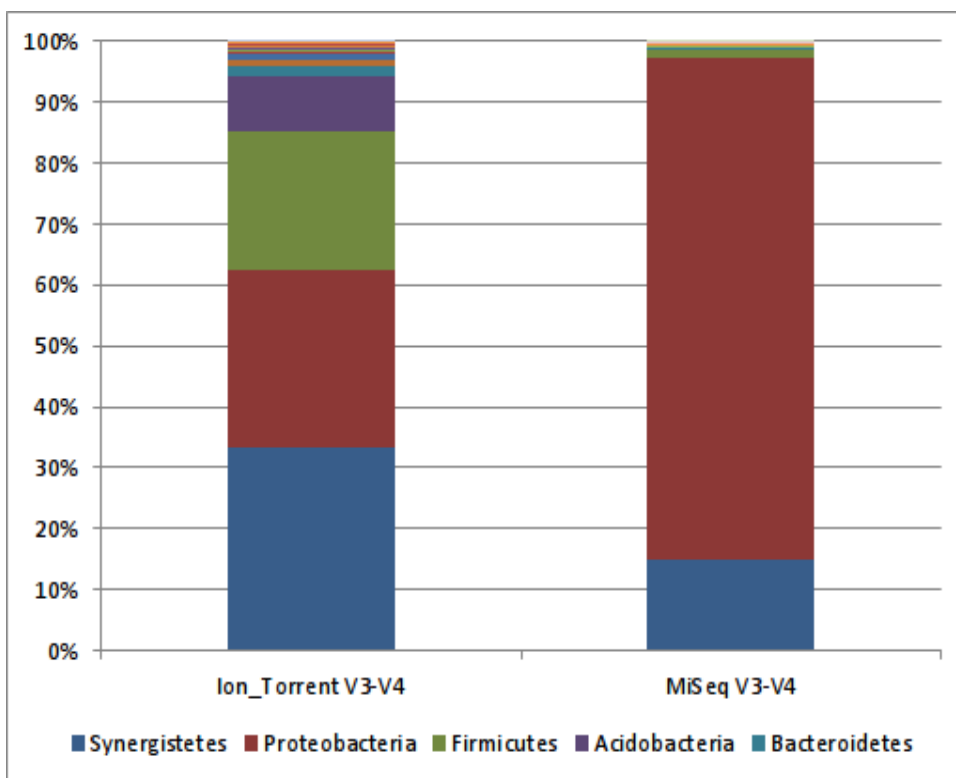


Figura 41.- Gráfica de OTUs a nivel de phylum comparando las regiones V3-V4 del gen ribosomal 16S con las dos tecnologías de secuenciación (Ion Torrent y MiSeq).

En la figura 41 se presenta la gráfica de OTUs asignados a nivel de phylum en la que se comparan las dos tecnologías de secuenciación, tomando en cuenta las regiones hipervariables V3-V4 del gen ribosomal 16S, el análisis fue realizado con la base de datos Silva que fue modificada para que diera como resultado la taxonomía y diversidad correspondiente a las regiones antes mencionadas. El número de OTUs que fueron asignados con Ion Torrent es de 31, mientras que los OTUs asignados con MiSeq son 19; a pesar de que fueron las mismas regiones hipervariables las que están siendo comparadas, sigue siendo Ion Torrent la plataforma de secuenciación con la que a nivel de phylum se obtiene un mayor número de OTUs. Los phylum más abundantes correspondientes a Ion Torrent son los siguientes: *Synergistetes*, *Proteobacterias*, *Firmicutes*, *Acidobacterias* y *Bacteroidetes*, con una abundancia de: 33%, 29%, 22%, 8% y 1%, respectivamente, representando el 93% de la muestra; mientras que los cinco phylum más abundantes encontrados en MiSeq son: *Proteobacterias*, *Synergistetes*, *Firmicutes*, *Euryarchaeota* y un OTU identificado como TM6, cuyas abundancias son las siguientes: 82%, 14%, 1%, 0.4% y 0.3%, representando un 97.7% del total de la comunidad presente en la muestra.

A pesar de haber realizado los análisis con la base de datos que fue modificada, aun así, es evidente que existen diferencias entre los dos métodos de secuenciación que fueron comparados, en donde nuevamente se puede observar que es con Ion Torrent con la plataforma con la que se obtiene un mayor número de OTUs que con MiSeq.

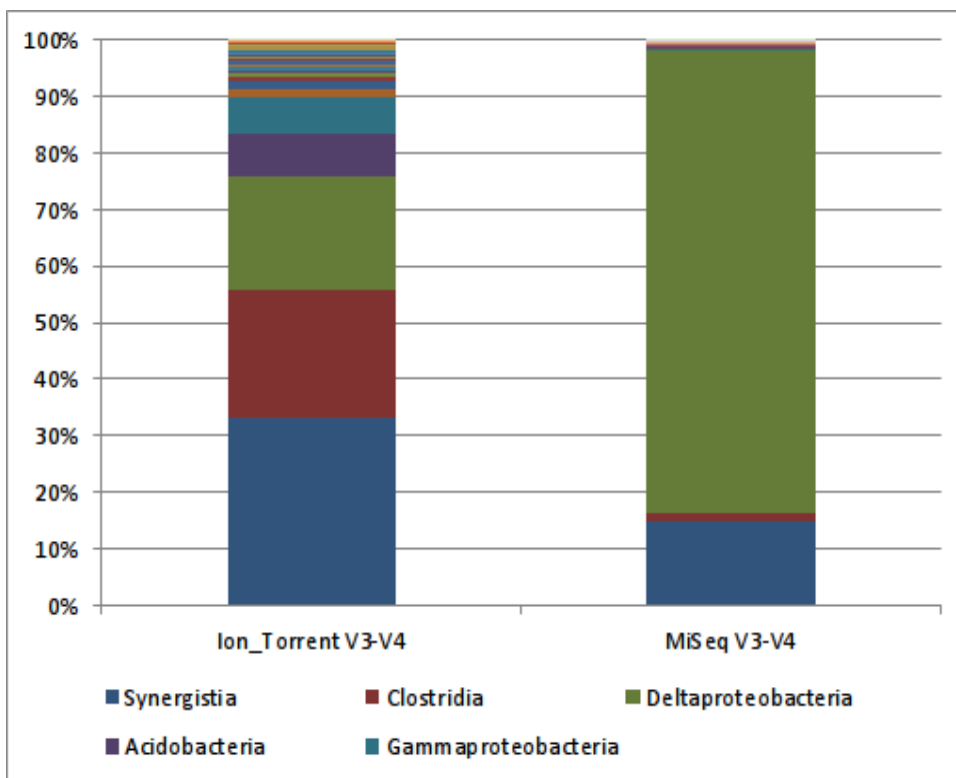


Figura 42.- Gráfica de OTUs a nivel de clase comparando las regiones V3-V4 del gen ribosomal 16S con las dos tecnologías de secuenciación (Ion Torrent y MiSeq).

La figura 42 muestra la gráfica de clases obtenidas con ambos métodos de secuenciación, los resultados de taxonomía presentados en la gráfica pertenecen a las regiones hipervariables V3-V4; en cuanto al número de OTUs asignados por cada plataforma, sigue siendo mucho mayor el obtenido con Ion Torrent, ya que con este método se obtuvieron un total de 69 OTUs asignados a nivel de clase, mientras que con MiSeq se lograron asignar 36 clases. Dentro de las 5 clases más abundantes encontradas con Ion Torrent están: *Synergistia*, *Clostridia*, *Deltaproteobacterias*, *Acidobacterias* y *Gammaproteobacterias*, representando las siguientes abundancias relativas dentro de la muestra: 33%, 22%, 20%, 7% y 6%, con lo cual abarcan el 88% de la abundancia total de la comunidad bacteriana, siendo la clase *Synergistia* la más abundante presente en los resultados obtenidos con Ion Torrent; por otro lado, las cinco clases más abundantes encontradas con MiSeq son las siguientes: *Deltaproteobacterias*, *Synergistia*, *Clostridia*, *Methanobacterias* y *Gammaproteobacterias*, las clases antes mencionadas se encuentran en una abundancia dentro de la muestra de: 81%, 14%, 1%, 0.4% y 0.3%, comprendiendo así el 96% de la comunidad encontrada con MiSeq.

A nivel de clase, es también evidente que con Ion Torrent se logra una mejor asignación taxonómica debido al número de OTUs que este asigno, por otro lado, al comparar las clases que fueron asignadas por ambos métodos, resulta ser que la mayoría de las que fueron encontradas con MiSeq se encuentran compartidas con Ion Torrent, lo que cambia entre ellas resulta ser la abundancia dentro de la muestra, ya que cuando son analizados con MiSeq, los OTUs se encuentran en mayor proporción que con Ion Torrent, lo cual podría explicarse debido al número

de OTUs asignados, ya que se ha observado ese mismo comportamiento, entre mayor número de OTUs haya en la muestra, la abundancia en la que se encuentran va disminuyendo.

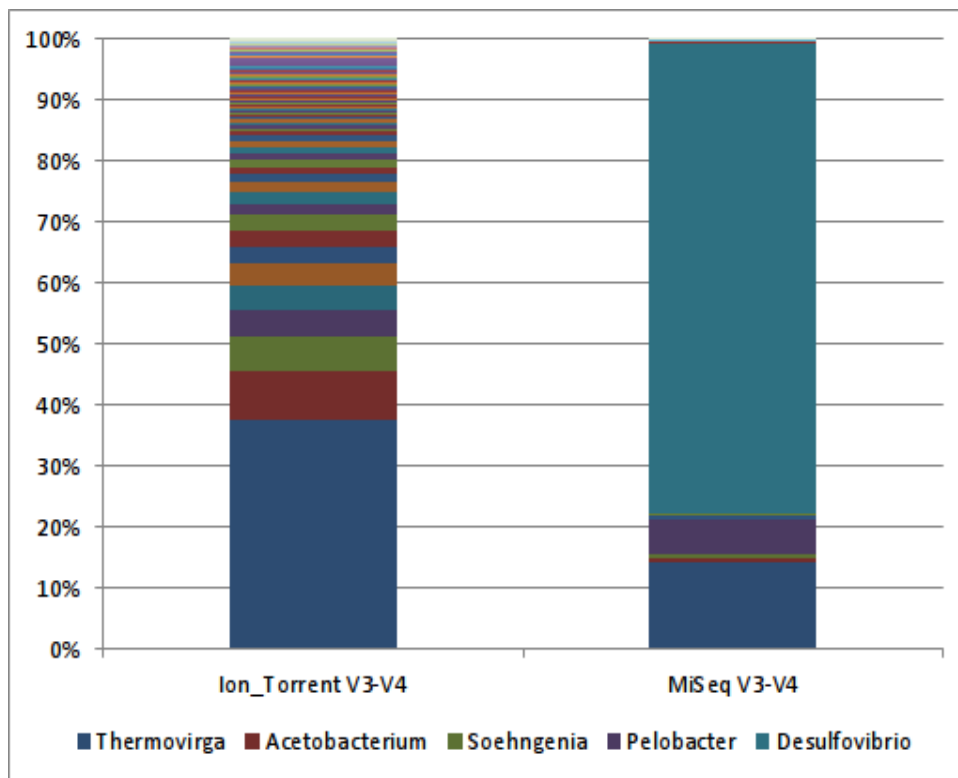


Figura 43.- Gráfica de OTUs a nivel de género comparando las regiones V3-V4 del gen ribosomal 16S con las dos tecnologías de secuenciación (Ion Torrent y MiSeq).

En la figura 43 se muestra la gráfica de géneros asignados por ambas plataformas de secuenciación, en donde se comparan los resultados obtenidos de las regiones hipervariables V3-V4 del gen 16S; en cuanto a taxonomía, resulta ser Ion Torrent la tecnología con la que se obtiene una mayor asignación de OTUs a nivel de género, ya que se consiguen asignar un total de 377, mientras que con MiSeq se obtienen únicamente 93 OTUs en este nivel. Con Ion Torrent, se obtienen los siguientes géneros en mayor abundancia dentro de la muestra: *Thermovirga*, *Acetobacterium*, *Soehngenia*, *Pelobacter*, un OTU identificado como iron-reducing bacterium y *Pseudomonas*, cuyas abundancias dentro de la muestra son las siguientes: 33%, 6%, 5%, 3%, 3% y 3%, respectivamente, con dichas abundancias se representa únicamente el 53% dentro de la muestra.

De acuerdo con los 5 géneros más abundantes encontrados con Ion Torrent dentro de la muestra de agua del pozo, *Thermovirga*, encontrado como el más abundante, indica la presencia de microorganismos del tipo termófilo, lo cual podría ser esperado, ya que la primera vez que este género fue descrito, el mismo fue aislado de una tubería de producción que alcanzaba temperaturas de hasta 68°C (Dahle & Birkeland, 2006); esta bacteria está caracterizada como degradadora de amino ácidos como la cisteína o reductora de azufre elemental, así como también el tiosulfato, siendo capaz de producir sulfuro de hidrogeno (Göker *et al.*, 2012; Lenhart *et al.*, 2014). El sulfuro de hidrogeno resulta ser toxico, altamente corrosivo y frecuentemente produce

precipitados de sulfuros metálicos en las tuberías y en los equipos de separación de fluidos, este compuesto, como ya fue mencionado, es producto principalmente de la actividad de bacterias reductoras de sulfato y de factores que tienen que ver con los nutrientes encontrados en las tuberías (Sturman & Goeres, 1999).

Por otro lado, encontrada en una abundancia menor, está presente el género *Acetobacterium*, la cual es considerada una bacteria acetogénica, capaz de producir ácidos orgánicos (Voordouw *et al.*, 1996); Se ha descrito que junto con las bacterias metanógenas, los microorganismos que conservan energía a través de la acetogénesis pueden participar en la corrosión determinada como EMIC, por sus siglas en inglés (Electrical Microbial Influenced Corrosion), este tipo de corrosión ha sido poco investigada y recientemente comprobada que puede tener efectos biocorrosivos en ambientes cuyas condiciones resultan ser estrictamente anóxicas, sin embargo, los efectos biocorrosivos que se han observado en donde tienen un papel las bacterias acetogénicas se ven aumentados en presencia de comunidades bacterianas más complejas (Kato *et al.*, 2015; Mand *et al.*, 2014).

Dentro de los 5 OTUs más abundantes encontrados con Ion Torrent, se encuentra también el género *Soehngenia*, este tipo de bacteria ha sido encontrada en la industria del petróleo (Kano *et al.*, 2008), además que ha sido relacionada con la biocorrosión gracias a que puede producir sulfuro de hidrógeno como parte de su metabolismo (Li *et al.*, 2016). Como ha sido mencionado, el sulfuro de hidrógeno es un compuesto que resulta ser altamente dañino para el material metálico, ya que puede causar corrosión; resulta entonces que *Soehngenia* es también una bacteria que puede llegar a producir este tipo de compuesto, pudiendo contribuir dañinamente al material metálico del pozo muestreado en este trabajo.

Otra de las bacterias que fueron encontradas con Ion Torrent es *Pseudomonas*, este tipo de bacteria fue catalogada como bacteria formadora de limo (Yuan & Pehkonen, 2007), es de tipo aerobio y tiene la característica de formar biopelículas en las superficies metálicas, las reacciones químicas que ocurren dentro de la biopelícula formada, se ha comprobado que pueden iniciar y sobre todo acelerar un proceso de corrosión (Abdolahi *et al.*, 2014). Otra de las bacterias encontradas dentro de la muestra, resulta ser *Pelobacter*, este tipo de bacteria normalmente se encuentra presente en sedimentos marinos y pozos petroleros, los cultivos de este tipo de bacterias son capaces de crecer fermentando varios tipos de compuestos hidrocarbonados generando así hidrógeno, cuyo producto final puede ser consumido por otras bacterias como las metanógenas y acetógenas como *Acetobacterium* (Vigneron *et al.*, 2016), podría decirse entonces que este género de bacteria se encuentra creciendo de manera sinérgica junto con *Acetobacterium* y algunas otras especies de bacterias acetogénicas beneficiándose una a la otra.

De acuerdo con las bacterias encontradas con Ion Torrent, se podría concluir que, al menos de las 5 más abundantes presentadas, todas se han encontrado teniendo un papel en actividades corrosivas en muestras tomadas de tuberías o pozos petroleros, la mayoría puede estar creciendo de manera independiente tomando nutrientes del medio que la rodea, siendo así enriquecida y desarrollando actividades corrosivas; o puede ocurrir el caso como *Pelobacter* y *Acetobacterium* que pueden estar creciendo de manera sinérgica, produciendo metabolitos que benefician a la otra, por lo que puede ocurrir también que los OTUs restantes asignados estén creciendo de las maneras antes mencionadas, pudiendo así provocar daños en las tuberías del pozo.

Por otro lado, los 5 OTUs más abundantes asignados a nivel de género que fueron obtenidos con MiSeq son los siguientes: *Desulfovibrio*, *Thermovirga*, *Pelobacter*, *Anaerobaculum* y *Soehngenia*, cuyas abundancias dentro de la muestra son: 75%, 13%, 5%, 0.9% y 0.7%, respectivamente, comprendiendo así el 94% de los OTUs encontrados con MiSeq en el pozo petrolero.

El género más abundante encontrado con MiSeq resultó ser *Desulfovibrio*, el cual pudo ser asignado hasta nivel de especie, dando como resultado *Desulfovibrio alaskensis*, esta especie en particular de bacteria fue aislado de un pozo petrolero ubicado en Alaska y de acuerdo con diversos estudios de caracterización, fue designado con el nombre de *Desulfovibrio alaskensis*; este tipo de bacterias pueden crecer bajo ciertas condiciones fisicoquímicas como un rango de pH entre 6.5 y 8.5 y temperaturas entre 10 y 45°C presentando un crecimiento óptimo a los 37°C (Feio *et al.*, 2004), dichos rangos de temperatura y pH en los que la bacteria puede presentar crecimiento se encuentran dentro de los que presenta el pozo petrolero estudiado, ya que el pH se encuentra en un rango de 6.35 y 7.10, mientras que la temperatura oscila entre 20.3 y 20.8°C, pudiendo facilitar el crecimiento de la especie ya mencionada. Esta bacteria pertenece a las denominadas bacterias sulfato reductoras, las cuales tienen la capacidad de producir sustancias poliméricas extracelulares; en algunos casos se ha visto que la producción de dichas sustancias puede actuar como una capa protectora del metal y causar una inhibición en el proceso de corrosión, tal es el caso de las sustancias poliméricas extracelulares que secreta *Desulfovibrio alaskensis*, las cuales ayudan a inhibir la corrosión (Kip & A van Veen, 2015); sin embargo, estas bacterias tienen la capacidad de formar biopelículas que podrían tener un papel en la corrosión.

Otro de los géneros que fueron encontrados con MiSeq, fue *Anaerobaculum*, este tipo de bacteria se caracteriza por ser anaerobia, moderadamente termófila; este tipo de bacteria puede crecer a un rango óptimo de pH y temperatura de 6.6-7.3 y 55-60°C, respectivamente, y tiene además la característica de poder fermentar ácidos orgánicos (Menes & Muxí, 2016); los ácidos orgánicos pueden ser producidos por algún otro género de bacteria, por lo cual *Anaerobaculum* puede estar creciendo de manera sinérgica con alguna otra bacteria presente en la muestra.

Por otro lado, los géneros de bacterias *Thermovirga*, *Pelobacter* y *Soehngenia* fueron descritas anteriormente, ya que de igual manera, se encuentran dentro de los 5 OTUs más abundantes encontrados con Ion Torrent; estos tres géneros de bacterias se han encontrado teniendo un papel en actividades corrosivas dentro de las tuberías petroleras, cuyo caso puede ser similar en el pozo petrolero muestreado en este trabajo, ya que fueron detectadas con ambos métodos de secuenciación, por lo que podrían estar siendo enriquecidas mediante acciones sinérgicas con otras bacterias encontradas dentro del pozo, pudiendo así llevar a cabo funciones ácido productoras o sulfato reductoras que puedan conllevar a acciones corrosivas y así mismo dañar las tuberías del pozo.

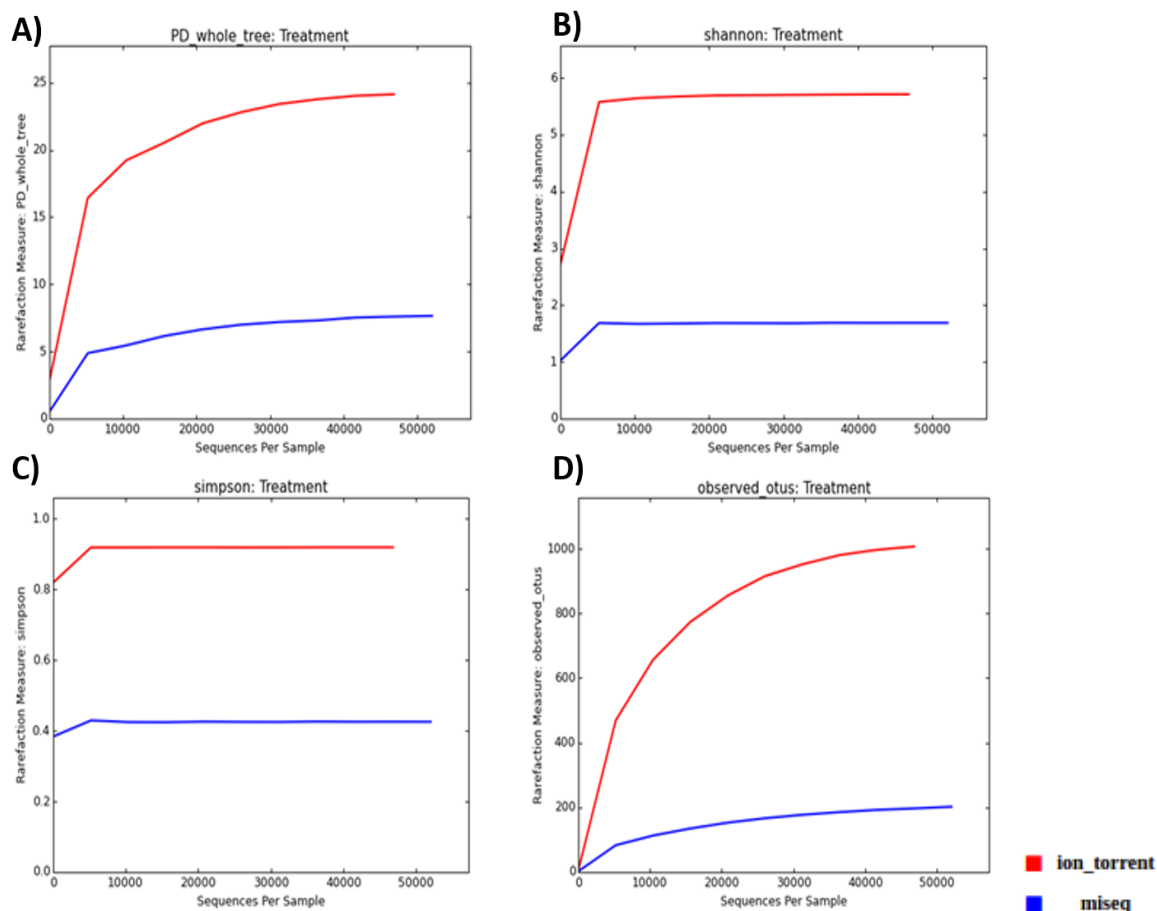


Figura 44.- Gráficas de los índices de A) Diversidad Filogenética (PD), B) Shannon, C) Simpson y D) OTUs observados; Ion Torrent V3-V4 vs MiSeq V3-v4.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
Ion Torrent	46837	24.138	5.712	0.920	1007
MiSeq	46837	7.596	1.689	0.426	197

Tabla 30. -Índices de diversidad alfa Ion Torrent V3-V4 y MiSeq V3-V4.

Fue comparada la diversidad alfa que se obtiene con los dos métodos de secuenciación (Ion Torrent y MiSeq) cuando únicamente se toman en cuenta las regiones hipervariables V3-V4 del gen ribosomal 16S, con la finalidad de conocer cuál de las dos plataformas proporciona la mayor diversidad y, por lo tanto, la más cercana a la comunidad presente en las muestras del pozo. Los índices de diversidad que fueron evaluados son el de diversidad filogenética, Shannon, Simpson y el número de OTUs que fue estimado por el programa QIIME para cada método de secuenciación, para que los índices antes mencionados tuvieran punto de comparación, fue tomado el mismo número de secuencias por muestra, es decir, la misma profundidad de secuenciación.

En la figura 44 se muestran de manera gráfica los índices antes mencionados, comparando los dos métodos de secuenciación utilizados, una vez más se puede observar el mismo comportamiento cuando son comparadas ambas plataformas de secuenciación, pero tomando en cuenta las siete

regiones hipervariables que son secuenciadas con Ion Torrent, la diferencia es que esta vez, fueron comparadas únicamente las regiones hipervariables V3-V4 tanto para Ion Torrent como para MiSeq; como se indica en la gráfica, la línea roja pertenece a los índices obtenidos por Ion Torrent, cuya diferencia en cuanto a la diversidad obtenida con MiSeq es claramente notoria, esto puede ser corroborado al observar la tabla 30, en donde se observa de manera numérica la gran diferencia en cuanto a diversidad obtenida.

Como puede notarse, las dos plataformas tienen una misma profundidad de secuenciación, que resulta ser alta en ambos casos, podría decirse que la máxima diversidad obtenida por las dos plataformas pudo ser alcanzada de acuerdo con lo observado en la figura 44, ya que de los índices que fueron comparados, las rectas logran 'acostarse' indicando que la secuenciación fue suficiente para poder observar la mayor diversidad posible que se encuentra presente en la muestra.

En cuanto al índice de diversidad filogenética (PD) obtenido en ambos casos, está claro que Ion Torrent representa el mayor, con un índice de 24, siendo mucho mayor en comparación con el obtenido por MiSeq, siendo de 7; de acuerdo con Faith (1992), es posible esperar que al lograr valores más altos de PD, estos correspondan a una mayor diversidad encontrada en la muestra que está siendo analizada, por lo que de acuerdo con los índices logrados con ambos métodos de secuenciación, resulta ser que con Ion Torrent se obtienen resultados que reflejan un muy alto índice de diversidad filogenética, lo que representa que con esta plataforma se puede observar una estimación de la diversidad bacteriana presente en la muestra del pozo mucho mayor que la que se puede obtener con MiSeq.

Otro de los índices evaluados para estimar la diversidad alfa estimada por ambas plataformas de secuenciación fue el índice de Shannon, este índice toma en cuenta tanto la diversidad como la igualdad de especies que se encuentran presentes en una muestra, con lo cual nos informa acerca de la composición de una comunidad tomando también en cuenta las abundancias relativas de las distintas especies encontradas (Stirling & Wilsey, 2001). Al igual que el índice de diversidad filogenética, un valor alto del índice de Shannon representa una mayor diversidad presente en la muestra analizada; en este caso, resulta ser nuevamente Ion Torrent la plataforma de secuenciación con la que se obtiene el índice más alto, siendo de 5, mientras que con MiSeq se obtiene un índice cercano a 2, lo que podría indicar que al secuenciar las muestras de agua del pozo petrolero analizado en este trabajo con Ion Torrent, es más probable conocer un poco más la diversidad de la comunidad presente en el pozo que si este es secuenciado con MiSeq.

El índice de Simpson fue otro de los índices evaluados para conocer cuál de los dos métodos de secuenciación resulta presentar una mayor estimación de la diversidad presente en las muestras de agua del pozo petrolero de Tabasco, cuando son tomadas en cuenta únicamente las regiones hipervariables V3-V4 del gen ribosomal 16S. El índice de Simpson indica, la biodiversidad que se encuentra en la comunidad analizada, así como también se centra especialmente en la diversidad de las especies dominantes dentro de la muestra; este índice al igual que los dos antes evaluados, entre mayor sea, indica una buena diversidad, por lo que entre ambos métodos de secuenciación resulta ser Ion Torrent quién logra un valor cercano a 1, mientras que con MiSeq se obtuvo un valor de 0.4, lo cual indica que con Ion Torrent se pudo obtener la mayor diversidad en cuanto a especies dominantes que se encuentran en el pozo petrolero de Tabasco.

Por último, también se evaluó el número de OTUs estimados por el programa que se obtuvieron con Ion Torrent y MiSeq, de acuerdo con la tabla 30, con Ion Torrent se logró observar un total de 1007 OTUs, mientras que con MiSeq, únicamente fueron estimados 197 OTUs; esto es tomando en cuenta que las secuencias obtenidas con ambas plataformas fueron analizadas con la base de datos modificada, cuyos resultados son específicos para las regiones hipervariables V3-V4 del gen ribosomal 16S; este mismo comportamiento fue observado cuando se evaluaron ambas plataformas, con la excepción que con Ion Torrent se tomaron en cuenta siete regiones hipervariables y con MiSeq únicamente dos, con lo cual, podría sugerirse que no importando el número de regiones hipervariables que son secuenciadas, con Ion Torrent puede obtenerse una mejor y mayor estimación de la diversidad que está presente en las muestras de agua del pozo petrolero de Tabasco, esto basándose en los resultados tanto de asignación de taxonomía como de diversidad alfa presentados anteriormente.

Fue también evaluada la diversidad beta de las réplicas del pozo, ya que esta diversidad puede medir la variabilidad de la composición de las especies entre unidades de muestreo (Anderson *et al.*, 2006), por lo que al realizar este análisis se desea conocer si existen diferencias entre ambos métodos de secuenciación utilizados.

En la figura 45 se observan las gráficas de coordenadas principales que fueron arrojadas por el programa QIIME; en el panel A) se muestra la gráfica de coordenadas no ponderada (gráfica cualitativa), la cual se centra específicamente en la presencia/ausencia de datos (OTUs) para poder comparar la composición de la comunidad y en el panel B) se presenta la gráfica de coordenadas ponderada (gráfica cuantitativa), para realizar esta gráfica se toma en cuenta además de la presencia/ausencia de los datos, la abundancia relativa de cada tipo de organismo (Lozupone *et al.*, 2007).

De acuerdo con la figura 45 A), es posible observar dos grupos completamente separados en polos distintos de la gráfica; los puntos azules etiquetados como ACR2, ACR3 y ACR4 corresponden a los resultados de las tres réplicas del pozo analizadas con MiSeq, mientras que los cuadros rojos etiquetados como Rep. 2, Rep. 3 y Rep. 4 pertenecen a los resultados de las tres réplicas del pozo analizadas con Ion Torrent; al realizar el análisis cualitativo de las réplicas del pozo, se puede observar que al tomar en cuenta si el OTU está presente o ausente dentro de las muestras, éstas se agrupan de acuerdo con sus similitudes, en el caso de las réplicas analizadas con Ion Torrent, existe la formación de un conjunto entre ellas dejando ver que los resultados obtenidos de las tres réplicas resultan ser muy similares, pudiendo indicar que los OTUs obtenidos con Ion Torrent resultan ser compartidos entre las tres, mientras que, de acuerdo con los resultados obtenidos de las réplicas analizadas con MiSeq, podría decirse que existe cierta agrupación entre ellas, ya que se encuentran en un mismo polo de la gráfica, aunque existen distancias entre las mismas que podrían sugerir que de manera cualitativa, los OTUs que fueron obtenidos por cada una no resultan ser tan similares o compartidos entre ellas. Otro de los comportamientos que pueden ser apreciados en la gráfica es la diferencia de los resultados obtenidos entre ambos métodos de secuenciación, ya que entre réplicas se pueden agrupar, pero en distintos polos de la gráfica, esto dependiendo de la plataforma de secuenciación que fue utilizada, por lo que podría decirse que los OTUs que resultaron de cada una no son compartidos en su totalidad, debido a la distancia a la que se encuentran ambos grupos de réplicas.

Por otro lado, en la figura 45, panel B), se encuentra la gráfica ponderada o cualitativa, que como fue mencionado anteriormente, toma en cuenta tanto la presencia o ausencia del OTU dentro de la muestra, así como sus abundancias relativas; de acuerdo con lo observado, es posible ver un comportamiento similar al del panel A), en donde se forman dos agrupaciones de las réplicas de acuerdo con el método de secuenciación que fue utilizado, los cuadros rojos etiquetados como Rep. 2, Rep. 3 y Rep. 4 pertenecen a los resultados obtenidos con Ion Torrent, mientras que los círculos azules etiquetados como ACR2, ACR3 y ACR4 corresponden a los resultados obtenidos con MiSeq. Al igual que la gráfica no ponderada, es posible ver que las réplicas forman dos grupos, en el caso de las réplicas analizadas con Ion Torrent, se puede distinguir que cuando se toma en cuenta también la abundancia relativa de los OTUs presentes en la muestra, estas no logran agruparse perfectamente como lo habían hecho en la gráfica A), ya que la réplica 2 se encuentra a una distancia significativa, incluso en un polo distinto en la gráfica, pudiendo indicar que, por un lado, las tres réplicas comparten los OTUs obtenidos, pero existen diferencias significativas entre la abundancia en la que se encuentran dichos OTUs dentro de la muestra, lo cual se ve reflejado en que la réplica 2 logre separarse de las réplicas 3 y 4; este mismo comportamiento fue también visto con las réplicas que fueron analizadas con MiSeq, ya que las réplicas 3 y 4 forman un conjunto entre ellas, con una distancia menor a la que se encuentra la réplica 2, que incluso también se encuentra en un polo distinto de la gráfica tal como la réplica que fue analizada con Ion Torrent, pudiendo ocurrir el mismo comportamiento, en donde a pesar de compartir los OTUs, las abundancias relativas de estos presentan diferencias, es por eso que en la gráfica se presentan separadas las réplicas 3 y 4 de la réplica 2.

Por otro lado, a pesar de las diferencias existentes entre la réplica 2 de las réplicas 3 y 4, estas se agrupan de manera similar que en la gráfica A), ya que dependiendo del método de secuenciación que fue utilizado para realizar el análisis, existen distancias entre los dos grupos de réplicas, lo cual podría indicar que los OTUs y las abundancias de cada uno que fueron encontrados por ambas resultan ser distintos entre réplicas, esto puede ser también observado con los resultados presentados en las gráficas 41, 42 y 43, ya que a pesar de que algunos OTUs son compartidos entre Ion Torrent y MiSeq, las abundancias en las que son encontrados dichos OTUs dentro de las muestras son totalmente diferentes.

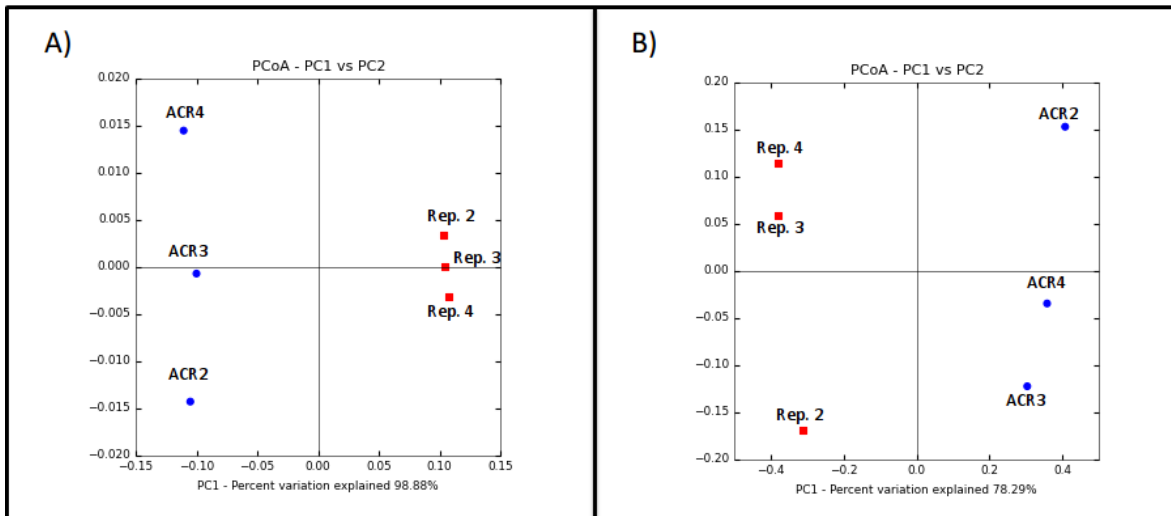


Figura 45.- Graficas de análisis de coordenadas principales de las secuencias analizadas con Ion Torrent V3-V4 y MiSeq.

Fueron también comparados mediante diagramas de Venn los OTUs asignados a nivel de género y especie por ambos métodos de secuenciación para conocer cuántos OTUs son compartidos y los distintos que fueron asignados.

En la figura 46 se muestran los diagramas de Venn de los géneros (A) y especies (B) que fueron obtenidos con ambos métodos de secuenciación, esto con la finalidad de conocer, del total de OTUs asignados a los niveles taxonómicos antes mencionados cuántos de ellos son compartidos por ambos métodos de secuenciación y cuántos de ellos resultan ser distintos, esto para conocer si se logra asignar una mayor cantidad de OTUs con Ion Torrent o con MiSeq.

En el panel A) se observa el diagrama de Venn de los géneros que se obtuvieron con cada método de secuenciación; el número de géneros que se encuentran compartiendo ambas plataformas es de 67, lo cual resulta ser casi la mayoría de géneros que lograron ser asignados con MiSeq, ya que con esta plataforma únicamente se obtienen 25 géneros que son distintos a los compartidos; por otro lado, además de los 67 géneros, con Ion Torrent se obtiene un total de 309 géneros diferentes, lo que hace evidente que con Ion Torrent se obtiene una mayor cantidad de OTUs que fueron asignados a nivel de género, cuyo comportamiento se pudo observar incluso desde el nivel de phylum y clase de acuerdo con las gráficas presentadas en las figuras 41 y 42.

Por otro lado, en el panel B) se muestra el diagrama de Venn de las especies que se obtuvieron con cada método de secuenciación; en este caso, resultan ser 63 especies las que se encuentran compartiendo ambas plataformas, mientras que con MiSeq se obtienen únicamente 50 de diferencia y con Ion Torrent son logradas 403 especies distintas a las compartidas, con lo que puede ser corroborado que efectivamente, con Ion Torrent es posible asignar una mayor cantidad de OTUs a los distintos niveles taxonómicos, esto puede resultar en una mejor estimación de la taxonomía que se encuentra presente en las muestras de agua del pozo petrolero de Tabasco; además, tomando en cuenta los resultados que fueron expuestos anteriormente, tanto de diversidad alfa y beta, es también posible deducir que con Ion Torrent podría conocerse de mejor manera la diversidad que se encuentra en el pozo, ya que como lo indican las figuras y tablas

presentadas anteriormente, Ion Torrent fue el mejor método de secuenciación para las muestras de los dos que fueron comparados, tomando en cuenta que la comparación se llevó a cabo basándose en los resultados de las regiones hipervariables V3-V4 del gen ribosomal 16S. En algunos casos (Salipante *et al.*, 2014; Quail *et al.*, 2012), se han reportado ciertas deficiencias o tasas de errores más altas provenientes de la secuenciación con Ion Torrent, cuando esta plataforma ha sido comparada con otras tecnologías distintas; sin embargo, en el caso particular de las muestras de agua del pozo petrolero de Tabasco, con Ion Torrent se lograron mejores resultados, ya que se pudo observar un perfil más amplio de las bacterias que se encuentran presentes en las muestras, lo cual era necesario para poder conocer si las bacterias que están en el pozo pueden tener actividades que resulten dañinas a los materiales metálicos y puedan ser las causantes de la corrosión; por otro lado, a pesar de haber obtenido muy poca diversidad, comparada con la obtenida con Ion Torrent, con MiSeq se obtuvo un perfil en donde casi la mayoría de bacterias presentes resultan ser bacterias con actividades sulfato reductoras principalmente, con lo cual podría decirse que, para este caso en particular, si se requiere conocer cuáles y cuántas bacterias que presentan dicha actividad están presentes en la muestra, se puede recurrir entonces a la plataforma de secuenciación de MiSeq, la cual resulta ser incluso económicamente más benéfica que Ion Torrent, por lo que podría deducirse que cualquiera de las dos tecnologías tiene sus beneficios, pero para el caso específicos de las muestras de agua que fueron secuenciadas en este trabajo, resulta más conveniente trabajar con los resultados provenientes de Ion Torrent, ya sea tomando en cuenta las siete regiones hipervariables secuenciadas en un principio o tomando en cuenta las regiones hipervariables V3-V4 del gen ribosomal 16S.

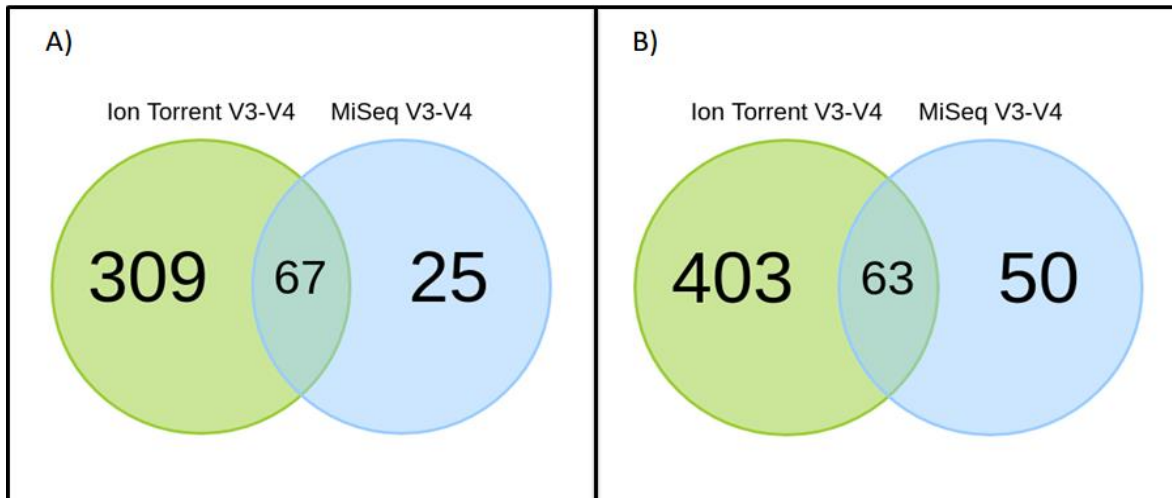


Figura 46.- Diagramas de Venn de A) géneros y B) especies compartidas entre Ion Torrent V3-V4 y MiSeq.

6.8 Diversidad de cultivos ACP y BSR de los pozos de Tabasco y Yucatán

Además de las muestras de agua que fueron secuenciadas con Ion Torrent y con MiSeq, también se realizaron cultivos de las membranas con las que se filtró el agua de las muestras antes mencionadas, estos cultivos fueron específicos para el crecimiento de bacterias con actividad sulfato reductora y ácido productora, los cultivos se realizaron para estos dos tipos de bacterias ya que la mayoría de actividad corrosiva que se presenta en la literatura es debido a estos dos tipos

de bacterias; por ello se presentan los resultados que fueron obtenidos de los cultivos que presentaron crecimiento, fueron además comparadas las bacterias sulfato reductoras y ácido productoras que lograron crecer en dos distintos pozos petroleros, uno ubicado en Tabasco, cuyos resultados de diversidad y taxonomía fueron presentados anteriormente, y el otro ubicado en Yucatán, esto con la finalidad de conocer las distintas bacterias con actividad sulfato reductora y ácido productora que pueden crecer en distintos puntos de muestreo.

6.8.1 Silva

Los cultivos que presentaron crecimiento en los medios específicos fueron para las bacterias sulfato reductoras y ácido productoras de las muestras de agua del pozo petrolero de Yucatán, mientras que las bacterias ácido productoras fueron las únicas que presentaron crecimiento de los cultivos realizados de membrana del pozo petrolero de Tabasco; de estos cultivos fue extraído el ADN y fueron secuenciadas las regiones hipervariables V3-V4 del gen ribosomal 16S, la secuenciación se realizó con MiSeq y el tratamiento de las lecturas, desde la limpieza, hasta el análisis fue realizado como es explicado en el método, la base de datos con la que se realizó el análisis final fue Silva, ya que resultó ser la más adecuada para explicar la diversidad presente en las muestras que fueron analizadas.

Se presentan los resultados obtenidos de la taxonomía que se encuentra presente en los cultivos realizados, para poder comparar la diversidad que puede crecer en dos distintos puntos de muestreo que son los pozos petroleros de Tabasco y Yucatán.

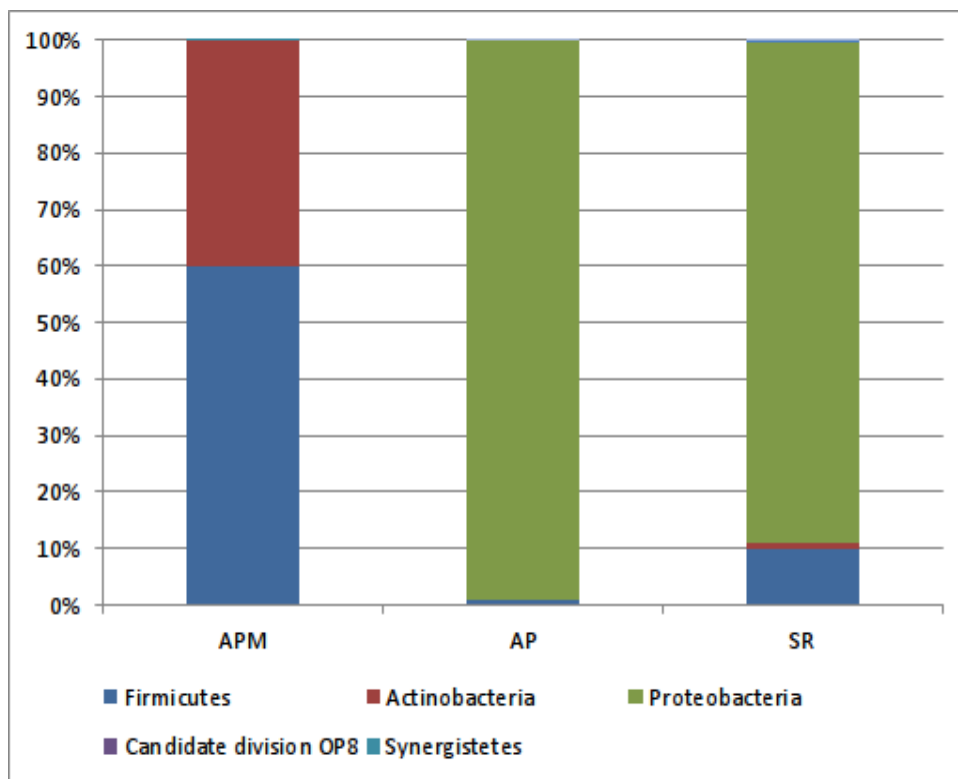


Figura 47.- Gráfica de OTUs a nivel de phylum de los cultivos para bacterias ácido productoras y sulfato reductoras de los pozos de Tabasco y Yucatán.

En la figura 47 se presenta la gráfica de abundancias relativas de los OTUs asignados a nivel de phylum, las barras se encuentran etiquetadas como APM, AP y SR que corresponden al cultivo de bacterias ácido productoras obtenidas de la membrana del pozo petrolero de Tabasco, y a las bacterias ácido-productoras y sulfato reductoras obtenidas de las muestras de agua del pozo petrolero de Yucatán, respectivamente. Como se puede observar en la gráfica, es posible distinguir un patrón de diferencia entre los resultados pertenecientes al pozo de Yucatán (AP y SR) y al pozo de Tabasco (APM). El número de OTUs que lograron ser asignados al nivel de phylum en cada una de las muestras resulta ser variable, en donde el cultivo de bacterias sulfato reductoras obtuvo el mayor número, siendo de 13 OTUs, seguido del cultivo de bacterias ácido productoras del mismo pozo, el cual fue de 7 OTUs, mientras que el cultivo de bacterias ácido productoras del pozo de Tabasco obtuvo un total de 5 OTUs.

Los OTUs asignados a nivel de phylum que se encontraron en el cultivo de bacterias ácido productoras del pozo petrolero de Tabasco fueron los siguientes: *Firmicutes*, *Actinobacterias*, *Proteobacterias*, un OTU denominado como Candidato a la división OP8 y por último, *Synergistetes*, cada uno con las siguientes abundancias relativas: 59%, 39%, 13%, 0.02% y 0.005%, respectivamente. Por otro lado, los cinco OTUs más abundantes encontrados en el cultivo de bacterias ácido productoras del pozo petrolero de Yucatán fueron los siguientes: *Proteobacterias*, *Firmicutes*, *Bacteroidetes*, *Actinobacterias* y *Synergistetes*, comprendiendo una abundancia relativa dentro de la muestra de: 98%, 0.9%, 0.15%, 0.02% y 0.01%, respectivamente, donde las *Proteobacterias* son predominantes en la muestra; al comparar ambos cultivos de bacterias ácido productoras en cuanto a los phylum resultantes, estas comparten 4 OTUs de los 5 más abundantes presentados.

Al analizar el cultivo de bacterias sulfato reductoras se obtuvieron 13 OTUs, los 5 más abundantes dentro de la muestra fueron los siguientes: *Proteobacterias*, *Firmicutes*, *Actinobacterias*, *Bacteroidetes* y las *Acidobacterias*, comprendiendo las siguientes abundancias relativas: 88%, 9%, 0.8%, 0.1% y 0.09%, respectivamente; de las bacterias sulfato reductoras más abundantes encontradas, 4 son compartidas con las bacterias ácido productoras provenientes del mismo pozo.

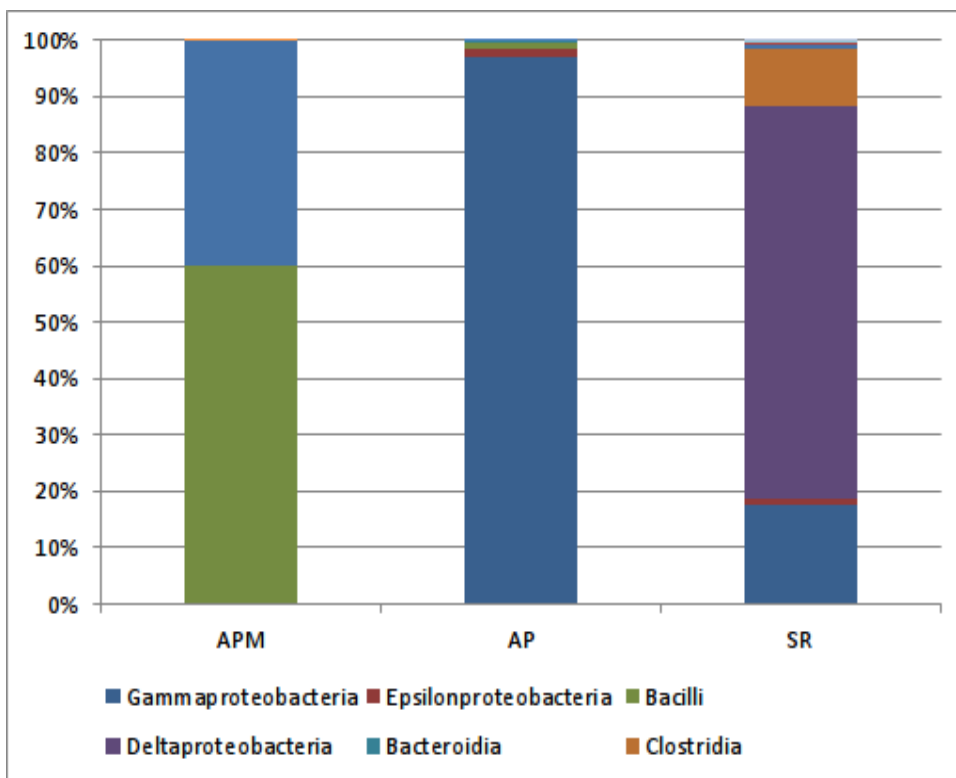


Figura 48.- Gráfica de OTUs a nivel de clase de los cultivos para bacterias ácido productoras y sulfato reductoras de los pozos de Tabasco y Yucatán.

En la figura 48 se muestra la gráfica de los OTUs obtenidos a nivel taxonómico de clase de los tres cultivos que están siendo comparados, los cultivos de bacterias ácido productoras y sulfato reductoras del pozo de Yucatán y el cultivo de bacterias ácido productoras de membrana del pozo de Tabasco; en este cultivo fueron encontrados un total de 9 OTUs, mientras que en el cultivo de bacterias ácido productoras de Yucatán se encontraron 13 OTUs y el cultivo de bacterias sulfato reductoras resultó ser nuevamente el más diverso de los tres que están siendo comparados, con un total de 22 OTUs encontrados en la muestra.

Los cinco OTUs más abundantes asignados a nivel de clase encontrados en el cultivo de membrana de bacterias ácido productoras fueron los siguientes: *Bacilli*, *Actinobacterias*, *Clostridia*, *Gammaproteobacterias* y un OTU clasificado como *Aminicenantes* sin cultivar, comprendiendo las siguientes abundancias relativas dentro de la muestra: 59%, 39%, 0.14%, 0.10% y 0.02%, respectivamente. Por otro lado, los cinco OTUs más abundantes encontrados en el cultivo de bacterias ácido productoras del pozo de Yucatán fueron los siguientes: *Gammaproteobacterias*, *Epsilonproteobacterias*, *Bacilli*, *Deltaproteobacterias* y *Bacteroidia*, cuyas abundancias relativas son las siguientes: 97%, 1.4%, 0.8%, 0.1% y 0.1%, respectivamente, en donde las *Gammaproteobacterias* resultan ser las más abundantes, mientras que esta misma clase que también fue encontrada en el cultivo del pozo de Tabasco, fue encontrada en menor abundancia en dicho pozo, en este caso, de los 5 OTUs más abundantes encontrados en ambos pozos, únicamente son compartidos 2 OTUs y fueron encontrados en diferentes abundancias, a nivel de phylum, pareció no haber tantas diferencias entre ambos cultivos, sin embargo, cuando son

analizados los OTUs a nivel de clase, empiezan a observarse las diferencias entre los microorganismos que pueden crecer en dos distintos puntos de muestreo.

En cuanto a los OTUs que fueron obtenidos en el cultivo de bacterias sulfato reductoras, los 5 más abundantes son los siguientes: *Deltaproteobacterias*, *Gammaproteobacterias*, *Clostridia*, *Epsilonproteobacterias* y *Actinobacterias*, cuyas abundancias relativas son las siguientes: 69%, 17%, 9.91%, 0.85% y 0.8%, respectivamente, el cultivo de bacterias sulfato reductoras resultó ser el más diverso taxonómicamente de los tres que están siendo comparados, debido a que no resultó el crecimiento del cultivo de bacterias sulfato reductoras del pozo petrolero de Tabasco, no es posible conocer y comparar las bacterias específicas antes mencionadas, esto para poder conocer cuál es la diversidad encontrada en distintos puntos de muestreo.

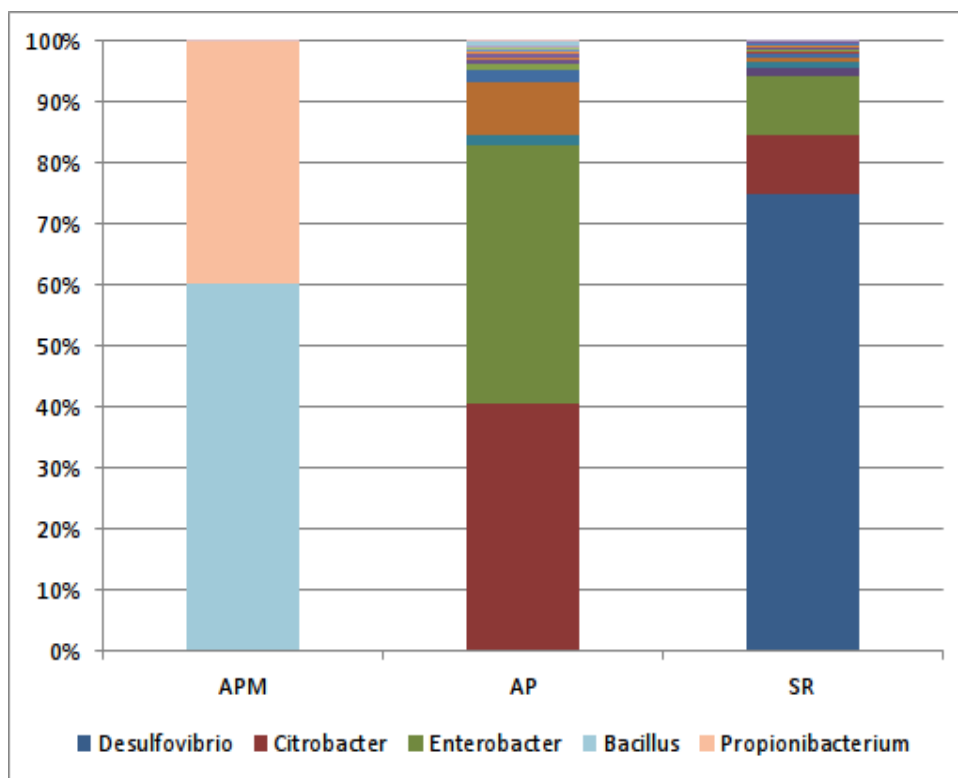


Figura 49.- Gráfica de OTUs a nivel de género de los cultivos para bacterias ácido productoras y sulfato reductoras de los pozos de Tabasco y Yucatán.

En la figura 49 se muestra la gráfica de géneros que fueron obtenidos de los cultivos de bacterias sulfato reductoras y ácido productoras de los pozos petroleros de Tabasco y Yucatán; dentro de los 5 géneros más abundantes encontrados en el cultivo de bacterias ácido productoras de Tabasco se encuentran los siguientes: *Bacillus*, *Propionibacterium*, *Sarcina*, *Photobacterium* y *Vibrio*, en las siguientes abundancias relativas dentro de la muestra: 59%, 39%, 0.13%, 0.05% y 0.03%, respectivamente, siendo las bacterias del género *Bacillus* las más abundantes dentro de la muestra; al analizar las secuencias al nivel de especie, los resultados fueron asignados a bacterias no cultivadas del mismo género antes mencionado, entre los OTUs que fueron asignados a nivel de especie se encuentran *Bacillus cereus*, *Bacillus mycoides*, *Bacillus weihenstephanensis*; diversos estudios han encontrado dentro de sus perfiles taxonómicos al género *Bacillus*, los cuales indican

que este género se ve favorecido en ambientes denominados micro-ácidos (Rajasekar *et al.*, 2010; Muthukumar *et al.*, 2006; López *et al.*, 2006; Albokari *et al.*, 2015). Otro de los géneros encontrados en el cultivo de membrana del pozo petrolero de Tabasco fue *Propionibacterium*, este género ha sido encontrado también en ambientes involucrados con la industria petrolera (Zhu *et al.*, 2003; Gao *et al.*, 2015; Albokari *et al.*, 2015), *Propionibacterium* se ha caracterizado dentro de los géneros que pueden producir ácidos orgánicos tales como el acético, butírico, fórmico, láctico, succínico y propiónico, los cuales pueden contribuir con la corrosión (Zhu *et al.*, 2003). Los géneros más abundantes encontrados en el cultivo del pozo pueden indicar que efectivamente se encuentran bacterias con actividades ácido productoras o algunas otras que puedan estar creciendo de manera sinérgica en el medio con dichas actividades, las cuales pueden estar contribuyendo a las actividades corrosivas dentro del pozo.

Dentro de los géneros encontrados en el cultivo de bacterias ácido productoras de Yucatán, fueron encontrados los siguientes: *Enterobacter*, *Citrobacter*, *Pseudomonas*, *Pantoea* y *Arcobacter* comprendiendo las abundancias relativas dentro de la muestra de: 42%, 40%, 8%, 1% y 1%, respectivamente, en donde el género *Enterobacter* resulta ser el más abundante dentro del cultivo; se ha descrito que las bacterias denominadas como *Enterobacterias* y *Pseudomonas* pueden generar ácidos débiles a partir de la degradación oxidativa de la glucosa, a pesar de no haber mucha información acerca del grupo de bacterias productoras de ácido, se han reportado miembros de la familia *Enterobacteriaceae* que han sido extraídos de ambientes corrosivos, tales como *Citrobacter*, *Enterobacter* y *Klebsiella* (Jan-Roblero *et al.*, 2008; Bermont *et al.*, 2007). Las bacterias encontradas en el cultivo del pozo de Yucatán pueden indicar que se están llevando a cabo actividades de producción de ácidos que puedan estar involucradas en ciertas actividades corrosivas.

Por otro lado, las cinco bacterias más abundantes encontradas en el cultivo específico para sulfato reductoras fueron las siguientes: *Desulfovibrio*, *Citrobacter*, *Enterobacter*, *Proteiniclasticum* y *Arcobacter*, con abundancias relativas de: 58%, 7%, 7%, 1% y 0.8%, respectivamente, cuando las secuencias fueron analizadas con el programa para obtener OTUs a nivel de especie, fue encontrada como la más abundante la especie *Desulfovibrio vulgaris*, esta bacteria se ha visto creciendo en sinergia junto con bacterias ácido productoras, ya que las bacterias ácido productoras pueden proveerle nutrientes necesarios bajo ciertas condiciones (Mand *et al.*, 2014).

De acuerdo con las bacterias que fueron obtenidas en cada uno de los tres cultivos que se compararon, cada OTU que se encuentra presente tiene cierta función dentro del cultivo específico, tanto de las bacterias ácido productoras como sulfato reductoras, por lo que puede ser posible que cada una de las bacterias encontradas tenga funciones específicas estando en conjunto con otras dentro del pozo del que fueron muestreadas, ya sea manteniendo relaciones de sinergia con el medio y las demás bacterias que las rodean, pudiendo provocar las actividades que provoquen corrosión en las tuberías.

Por otra parte, otro de los objetivos de secuenciar los cultivos provenientes de dos diferentes puntos de muestreo, fue para conocer si existen diferencias entre las bacterias encontradas de un punto a otro, desafortunadamente, el cultivo de bacterias sulfato reductoras proveniente del pozo petrolero analizado en este trabajo (Tabasco) no presentó crecimiento, por lo que únicamente pudo ser comparada la taxonomía encontrada de las bacterias ácido productoras, las cuales

presentaron evidentes diferencias en cuanto al número de OTUs encontrados en cada cultivo, como en su taxonomía, indicando entonces que cada punto de muestreo difiere del otro en cuanto a la comunidad encontrada, por lo que para conocer a detalle el perfil taxonómico y de diversidad de distintos pozos petroleros, es conveniente realizar un análisis de secuenciación como el que fue hecho en este trabajo.

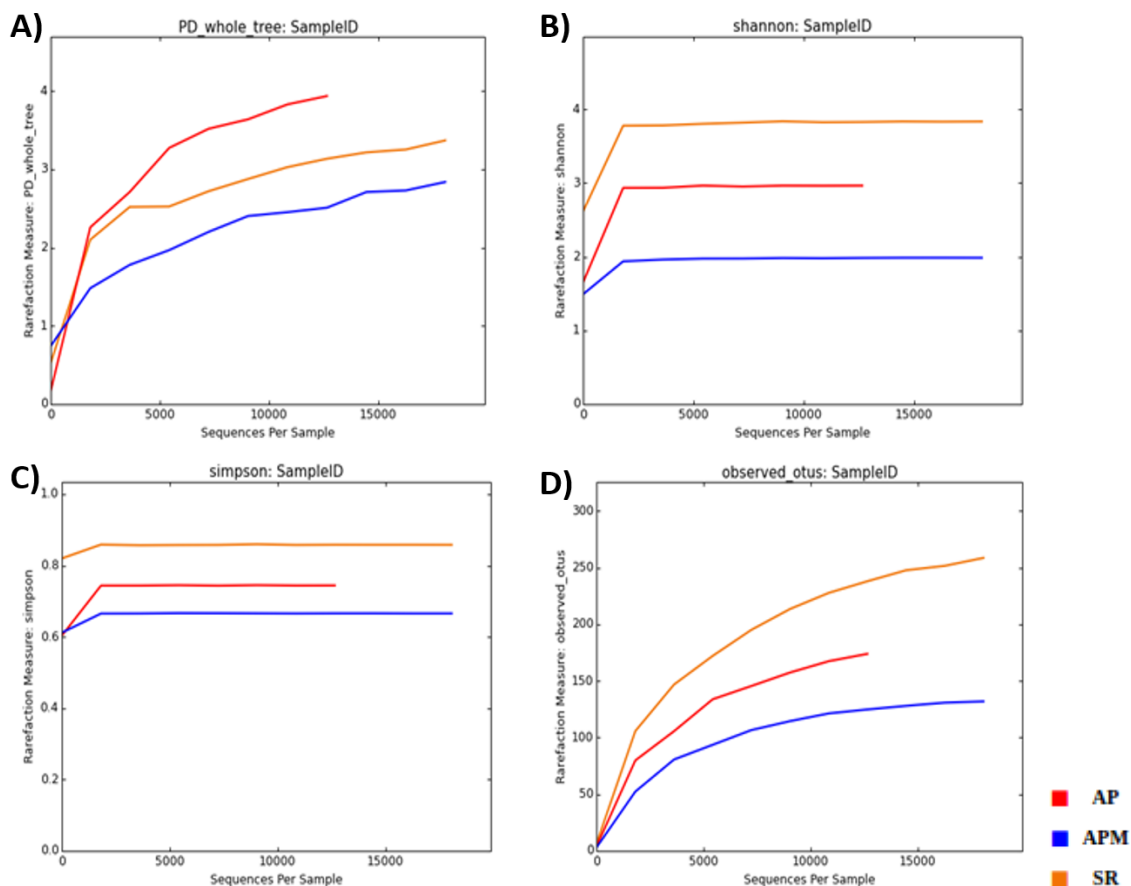


Figura 50.- Gráficas de los índices de A) Diversidad Filogenética (PD), B) Shannon, C) Simpson y D) OTUs observados de los cultivos para bacterias sulfato reductoras y ácido productoras.

Muestra	Secuencias/Muestra	PD	Shannon	Simpson	OTUs observados
APM	12645	2.508	1.984	0.666	124
AP	12645	3.933	2.966	0.744	174
SR	12645	3.133	3.829	0.858	238

Tabla 31.- Índices de diversidad alfa de los cultivos para bacterias sulfato reductoras y ácido productoras.

Además de analizar la taxonomía que se encuentra presente en los cultivos, también fue analizada la diversidad alfa de cada uno para poder conocer cuál de los tres resulta ser más diverso de acuerdo con los índices que fueron evaluados; para poder realizar las comparaciones, fue tomado el mismo número de secuencias por muestra de los tres cultivos, de manera gráfica se encuentran en la figura 50 los índices que fueron evaluados y en la tabla 31 se muestran los resultados de manera numérica, las muestras se encuentran etiquetadas como APM, AP y SR, las cuales

corresponden al cultivo de bacterias ácido productoras del pozo de Tabasco, y a los cultivos de bacterias ácido productoras y sulfato reductoras del pozo de Yucatán, respectivamente.

De manera gráfica (figura 50) se puede observar que la profundidad de secuenciación utilizada para los cultivos fue suficiente para poder alcanzar a cubrir la diversidad que se encuentra presente en las muestras, por otro lado, al analizar los índices de manera numérica, empezando por el índice de diversidad filogenética (PD), los cultivos inoculados con agua del pozo de Yucatán se ven más similares entre ellos, que el cultivo de bacterias ácido productoras del pozo de Tabasco, cuyo comportamiento permanece similar en los demás índices de diversidad que fueron analizados, de Shannon y Simpson, indicando que los cultivos de bacterias sulfato reductoras y ácido productoras del pozo petrolero de Yucatán son más diversos que el cultivo del pozo de Tabasco, lo cual puede ser también observado de acuerdo con el número de OTUs que fueron encontrados por cada uno, ya que el cultivo de bacterias ácido productoras de Tabasco resultó obtener el menor número de OTUs, mientras que el cultivo de bacterias sulfato reductoras de Yucatán obtuvo el mayor número de OTUs de los tres comparados, resultando así el más diverso de los tres, sin embargo, al no existir cultivo de bacterias sulfato reductoras de Tabasco, este no tiene punto de comparación.

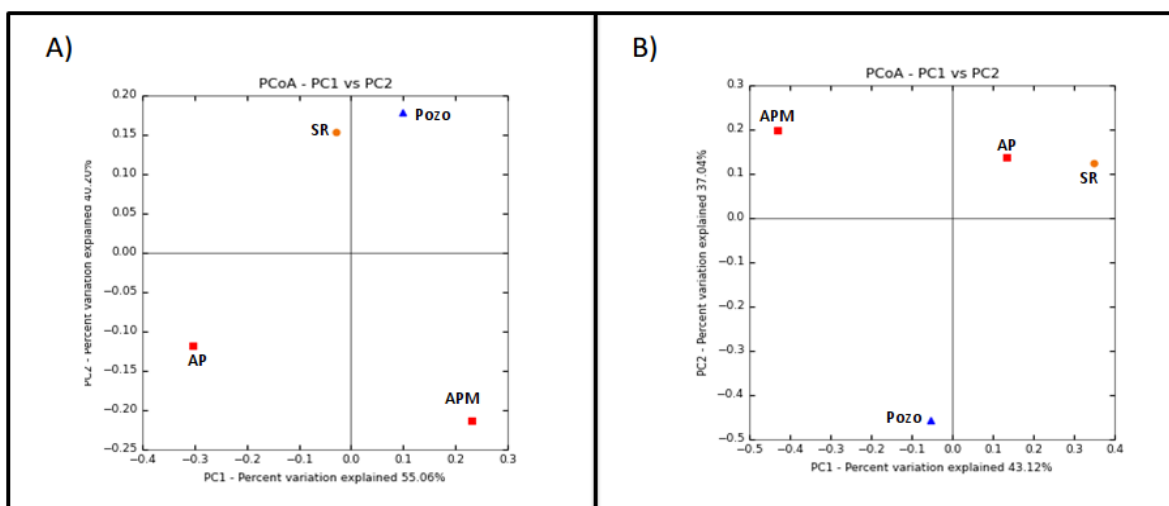


Figura 51.-Gráficas de análisis de coordenadas principales de los cultivos para bacterias sulfato reductoras y ácido productoras.

Además del análisis de diversidad alfa realizado a los cultivos, también fue hecho un análisis de diversidad beta para deducir si existen similitudes entre los cultivos, dentro del análisis fueron tomadas en cuenta las secuencias del pozo petrolero de Tabasco; en la figura 51 se presentan las gráficas de coordenadas principales no ponderada (panel A) y ponderada (panel B)) que se obtuvieron mediante el programa.

Fueron tomadas en cuenta las secuencias del pozo de Tabasco, ya que al realizar la comparación de la diversidad beta, uno de los resultados esperados era que el cultivo de bacterias ácido productoras se agrupara junto con los resultados del pozo petrolero, pero, como puede ser observado en ambas gráficas, no lograron agruparse, en el panel A) de la figura 51 fue realizado el análisis de coordenadas principales cualitativo, es decir, que únicamente se tomó en cuenta para graficar las diferencias, la presencia o ausencia de los OTUs asignados en las muestras, como se

puede observar, ninguna de las muestras analizadas logra agruparse, lo cual puede ser resultado de los distintos OTUs que fueron asignados por cada una, ya que los cultivos fueron específicos para el tipo de bacteria que fue analizada (sulfato reductoras o ácido productoras), sin embargo, podría decirse, de acuerdo con la gráfica, que existe una distancia menor entre las bacterias sulfato reductoras del pozo de Yucatán y las secuencias del pozo de Tabasco, pudiendo sugerir que las diferencias de los OTUs encontrados entre ellas son mínimas, a pesar de provenir de distintos puntos de muestreo, podría ser un tanto razonable, ya que dentro de la muestra del pozo, fueron encontrados una gran cantidad de OTUs pertenecientes al grupo de bacterias sulfato reductoras, sin embargo, una de las bacterias más abundantes encontradas en el cultivo fue *Desulfovibrio vulgaris*, la cual no se encuentra presente dentro de los OTUs encontrados en el pozo de Tabasco.

Por otro lado, en el panel B) se presenta la gráfica ponderada de coordenadas principales, es decir, la que toma en cuenta además de la presencia del OTU, su abundancia relativa; en este caso, se observa un patrón distinto al presentado en el panel A), ya que ahora los cultivos pertenecientes al pozo de Yucatán se encuentran a menor distancia uno del otro que como fue observado en el panel A), pudiendo formar una agrupación debido a que los cultivos pertenecen a un mismo punto de muestreo, lo cual, como ya fue mencionado, pudo haberse esperado del pozo y el cultivo de bacterias ácido productoras de Tabasco, sin embargo, en ambas gráficas, presentaron una distancia muy lejana una de la otra, lo que puede indicar que, a pesar de pertenecer a un mismo punto de muestreo, existen diferencias significativas entre lo que puede ser cultivado y lo que se encuentra dentro del pozo.

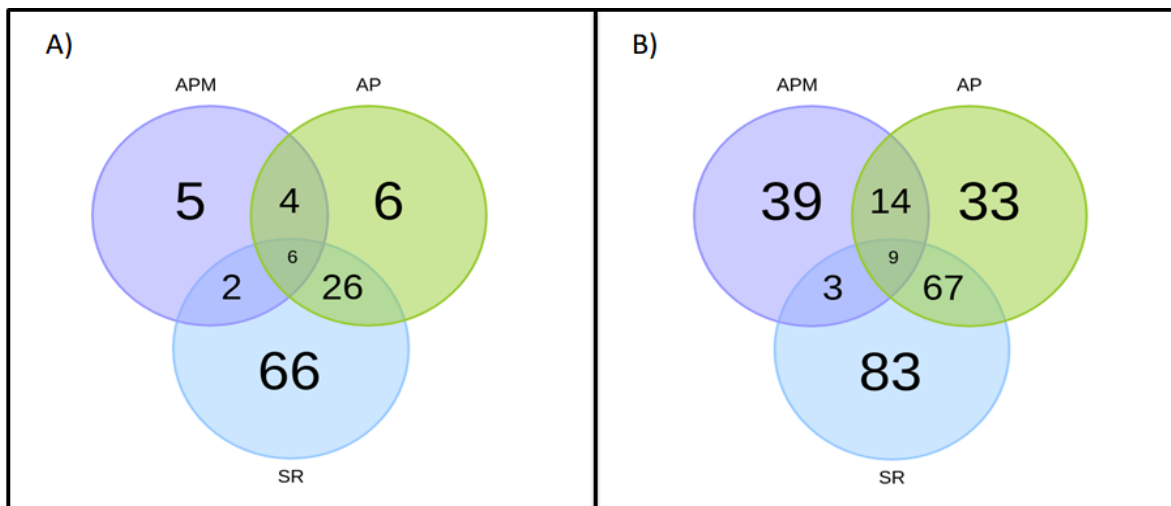


Figura 52.- Diagramas de Venn de los A) géneros y B) especies compartidas entre los cultivos de bacterias sulfato reductoras y ácido productoras.

Por último, fueron realizados diagramas de Venn de los OTUs asignados a género y especie que se encuentran compartiendo los cultivos que fueron comparados, en la figura 52 A) se observa el diagrama de Venn con el número de OTUs asignados hasta género de los tres cultivos; son muy pocos los géneros que se comparten entre los tres, cuyo resultado era de esperarse, ya que a pesar de provenir de dos diferentes puntos de muestreo, son también resultados de dos tipos distintos de cultivo, por otro lado, al comparar los OTUs obtenidos de los cultivos para bacterias

ácido productoras; este mismo comportamiento es visto cuando se comparan los OTUs obtenidos a nivel de especie de los tres cultivos.

De acuerdo con los resultados que se obtuvieron de taxonomía y diversidad alfa y beta de los cultivos para bacterias sulfato reductoras y ácido productoras, es posible deducir que a cada pozo petrolero lo constituye un perfil totalmente distinto de comunidades bacterianas que se encuentran en el mismo, ya sea de manera natural (como bacterias endógenas) o que hayan sido introducidas al pozo por las técnicas clásicas de recuperación del petróleo, de cualquier manera, las bacterias que se obtuvieron de los cultivos y de acuerdo con la literatura revisada, corresponden en efecto a llevar a cabo actividades del tipo sulfato reductor o ácido productor, cuyos efectos en conjunto con las demás bacterias y el medio que las rodea pueden ser efectivamente dañinas para los materiales metálicos de los pozos petroleros, llevando a cabo acciones que pueden resultar en actividades corrosivas.

7 Conclusiones

Según los resultados presentados para poder conocer el perfil taxonómico de las muestras de agua obtenidas de un pozo petrolero ubicado en Tabasco utilizando dos distintas plataformas de secuenciación masiva, fue posible conocer mediante dos perspectivas un análisis más completo de la comunidad que se encuentra en dicho pozo, ya que si hubieran sido utilizadas técnicas más básicas de biología molecular e incluso haber recurrido únicamente al uso de técnicas de cultivos bacterianos, no hubiera sido posible conocer más a fondo la comunidad analizada en este trabajo.

Por otro lado, los OTUs encontrados que fueron asignados a los distintos niveles taxonómicos gracias a las herramientas bioinformáticas utilizadas, independientemente del método de secuenciación que fue utilizado para su análisis, indican la presencia de una gran variedad de bacterias que desempeñan diferentes funciones, entre las principales, fueron encontradas bacterias del tipo sulfato reductoras, ácido productoras, metanógenas e incluso, aunque en proporciones menores dentro de la muestra, fueron también encontradas arqueas, estos tipos de bacterias, de acuerdo con la bibliografía, se han visto involucradas en procesos que resultan ser dañinos para las estructuras metálicas de los pozos petroleros, lo que resulta en casos de corrosión severa, esto gracias al metabolismo que desempeñan, ya que las bacterias encontradas en el pozo pueden estar creciendo de manera sinérgica unas con otras, llevando a cabo funciones específicas para la deterioración del metal.

Los resultados obtenidos de las diluciones seriadas en donde se obtuvo la presencia de las bacterias del tipo sulfato reductoras y ácido productoras dentro del pozo fue complementada con el perfil taxonómico que se dio a conocer según los resultados de la secuenciación del gen ribosomal 16S, independientemente del método de secuenciación que fue utilizado, ya que con los dos perfiles taxonómicos resultantes, fue posible identificar en proporciones mayores a las bacterias sulfato reductoras y ácido productoras, reforzando entonces los resultados de las diluciones, los cuales indican la severidad de la corrosión en las tuberías.

En este trabajo, otro de los objetivos, además de conocer el perfil taxonómico del pozo de Tabasco, fue el de llevar a cabo comparaciones entre dos plataformas de secuenciación masiva, las cuales fueron Ion Torrent y MiSeq; muy pocos estudios se han realizado comparando estas dos tecnologías de secuenciación, sin embargo, en los estudios que se han realizado se concluye que Ion Torrent presenta más errores que los que puede haber con MiSeq, pero en el caso de los resultados que se obtuvieron en este trabajo, fueron más convenientes los resultados arrojados con Ion Torrent que con MiSeq, ya que el objetivo era poder conocer un perfil más cercano de las bacterias encontradas en el pozo petrolero, lo cual fue logrado cuando el ADN obtenido fue secuenciado con Ion Torrent.

Las dos plataformas de secuenciación sugieren resultados interesantes que podrían ser utilizados de manera distinta, ya que, si se quisiera realizar un diagnóstico rápido y que resulte más económico, podría recurrirse entonces al método de secuenciación de MiSeq, por otro lado, si lo que se desea es tener un perfil taxonómico de las muestras a analizar lo más completo que sea posible, entonces la secuenciación podría realizarse con la plataforma de Ion Torrent.

Otro de los aspectos que deben ser tomados en cuenta es el análisis bioinformático que fue realizado en el presente trabajo, cuya metodología tuvo que ser estandarizada para lograr obtener

resultados que fueran certeros y de buena calidad, ya que se tuvieron que implementar distintas pruebas a lo largo del análisis bioinformático realizado, desde la limpieza de las secuencias, hasta la toma de decisión acerca de cuál base de datos pudiera ser óptima para los resultados provenientes de las secuencias del pozo petrolero de Tabasco, de igual manera se lograron también estandarizar las técnicas de extracción de ADN, purificación y secuenciación de las muestras, con la finalidad de poder mostrar un perfil taxonómico de la comunidad bacteriana lo más limpio posible, evitando siempre contaminación exterior que pudiera mostrar resultados inespecíficos.

Por último, fue también posible conocer mediante cultivos específicos, las bacterias que se encuentran involucradas en los procesos ácido productores dentro del pozo, lamentablemente no fue posible obtener datos de secuenciación de las bacterias que pudieran estar provocando los procesos de reducción del sulfato en el pozo, ya que el cultivo específico para este tipo de bacterias no presentó crecimiento, sin embargo, mediante estos cultivos fue posible conocer la diversidad de bacterias que se encuentra presente en un punto de muestreo distinto al presentado en este trabajo, con lo que se dedujo que, en cada pozo petrolero es posible encontrar distintos perfiles taxonómicos con los que se puede concluir si las bacterias presentes pueden estar causando actividades corrosivas, dañando así las estructuras metálicas.

Lo que puede concluirse es que, de acuerdo con el perfil taxonómico presentado en este trabajo, es que es muy probable que las bacterias encontradas en el pozo petrolero de Tabasco puedan estar llevando a cabo actividades que resulten de acción corrosiva y dañina a las estructuras metálicas que conforman el pozo.

8 Bibliografía

1. Abdolahi, A., Hamzah, E., Ibrahim, Z., & Hashim, S. (2014). Microbially influenced corrosion of steels by *Pseudomonas aeruginosa*. *Corrosion Reviews*, 32(3–4), 129–141.
2. Albokari, M., Mashhour, I., Alshehri, M., Boothman, C., & Al-Enezi, M. (2015). Characterization of microbial communities in heavy crude oil from Saudi Arabia. *Annals of Microbiology*, 65(1), 95–104.
3. Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters*, 9(6), 683–693.
4. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Disponible en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
5. Beech, I. B., & Gaylarde, C. C. (1999). Recent advances in the study of biocorrosion: an overview. *Revista de Microbiología*, 30(3), 177–190.
6. Beech, I. B., & Sunner, J. (2004). Biocorrosion: Towards understanding interactions between biofilms and metals. *Current Opinion in Biotechnology*, 15(3), 181–186.
7. Bermont-Bouis, D., Janvier, M., Grimont, P. A. D., Dupont, I., & Vallaeys, T. (2007). Both sulfate-reducing bacteria and Enterobacteriaceae take part in marine biocorrosion of carbon steel. *Journal of Applied Microbiology*, 102(1), 161–168.
8. Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., ... Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, 13, 390–401.
9. BLAST Topics. (2017). Query Input and database selection. Consultado el 24/08/2017.
10. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 30, 1–7.
11. Booth, G. H. (1964). Sulphur Bacteria in Relation to Corrosion. *Journal of Applied Bacteriology*, 27(1), 174–181.
12. Bradford, S. (1993). Corrosion Control. *Van Nostrand Reinhold*. 3-11.
13. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nature Publishing Group*, 7(5), 335–336.
14. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), 1767–1771.
15. Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., ... Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, 42(D1), 633–642.
16. Cubero, J. I., & Flores, F. (2003). Métodos Estadísticos Para El Estudio De La Estabilidad Varietal En Ensayos Agrícolas, 198.
17. Dahle, H., & Birkeland, N. K. (2006). *Thermovirga lienii* gen. nov., sp. nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *International Journal of Systematic and Evolutionary Microbiology*, 56(7), 1539–1545.
18. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 72(7), 5069–5072.
19. Droycon Bioconcepts. (2013). APB-BART Protocol, DBASOPO. 1-7. Consultado el 7/07/2016.
20. Duncan, K. E., Gieg, L. M., Parisi, V. A., Tanner, R. S., Green Tringe, S., Bristow, J., & Sufita, J. M. (2009). Biocorrosive Thermophilic Microbial Communities in Alaskan North Slope Oil Facilities. *Environmental Science Technology*, 43(2), 7977–7984.
21. Escobar-Zepeda, A., Vera-Ponce de León, A., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6(DEC), 1–15.

22. Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10.
23. Faith, D. P., & Baker, A. M. (2006). Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary Bioinformatics Online*, 2, 121–128.
24. Feio, M. J., Zinkevich, V., Beech, I. B., Llobet-Brossa, E., Eaton, P., Schmitt, J., & Guezennec, J. (2004). *Desulfovibrio alaskensis* sp. nov., a sulphate-reducing bacterium from a soured oil reservoir. *International Journal of Systematic and Evolutionary Microbiology*, 54(5), 1747–1752.
25. Gao, P., Tian, H., Li, G., Sun, H., & Ma, T. (2015). Microbial diversity and abundance in the Xinjiang Luliang long-term water-flooding petroleum reservoir. *MicrobiologyOpen*, 4(2), 332–342.
26. Geurkink, B., Doddema, S., & Euverink, G. J. (2016). Value of Next Generation Sequencing as monitoring tool for microbial corrosion A practical case from bioprophyling to tailor made MMM analysis. *Corrosion2016*, (7764), 1–14.
27. Göker, M., Saunders, E., Lapidus, A., Nolan, M., Lucas, S., Hammon, N., ... Klenk, H.-P. (2012). Genome sequence of the moderately thermophilic, amino-acid-degrading and sulfur-reducing bacterium *Thermovirga lienii* type strain (Cas60314(T)). *Standards in Genomic Sciences*, 6(2), 230–9.
28. Gómez-Castellanos. (2009). Aislamiento y caracterización morfológica y bioquímica de bacterias productoras de ácido (BPA) alojadas en aguas de producción de un campo petrolero. Tesis de Licenciatura. Universidad Industrial de Santander.
29. Green Biologics blog. (2017). <http://www.greenbiologics.com/blog/what-are-clostridia/>. Consultado el 05/01/2017.
30. Hugenholtz, P., & W. Tyson, G. (2008). Metagenomics. *Nature News and Views*, 455, 481–483.
31. Hugenholtz, P., Hooper, S. D., & Kyrpides, N. C. (2009). Focus: Synergistetes: Genomics update. *Environmental Microbiology*, 11(6), 1327–1329.
32. Illumina Technical Note. (2011). Quality Scores for Next-Generation Sequencing.
33. Iverson, W. P. (1987). Microbial Corrosion of Metals. *Advances in Applied Microbiology*, 32, 1–36.
34. Jack, T. R. (2002). Biological corrosion failures. *Failure Analysis and Prevention*, 11, 881–890.
35. Jan-Roblero, J., Posadas, A., Zavala Díaz De La Serna, J., García, R., & Hernández-Rodríguez, C. (2008). Phylogenetic characterization of bacterial consortia obtained of corroding gas pipelines in Mexico. *World Journal of Microbiology and Biotechnology*, 24(9), 1775–1784.
36. Jan-Roblero, J., Romero, J. M., Amaya, M., & Le Borgne, S. (2004). Phylogenetic characterization of a corrosive consortium isolated from a sour gas pipeline. *Applied Microbiology and Biotechnology*, 64(6), 862–867.
37. Jones, D. (1996). Principles and prevention of corrosion. Second edition. Prentice Hall. 372-381.
38. Jones, D. A., & Amy, P. S. (2002). A Thermodynamic Interpretation of Microbiologically Influenced Corrosion. *Corrosion*, 58(8), 638–645.
39. Kano, S., Mukaidani, T., Hattori, Y., Fujiwara, K., Miyagawa, Y., Takabayashi, K., ... Okatsu, K. (2009). Diversity of indigenous anaerobes and methane conversion system from reservoir oil by indigenous anaerobes in depleted oil fields. *Journal of the Japan Petroleum Institute*, 52(6), 297–306.
40. Kato, S., Yumoto, I., & Kamagata, Y. (2015). Isolation of Acetogenic Bacteria That Induce Biocorrosion by Utilizing Metallic Iron as the Sole Electron Donor. *Applied and Environmental Microbiology*, 81(1), 67–73.
41. Keasler, V., Bennett, B., Keller, C., Whalen, P., Cairns, J., & De Paula, R. M. (2013). Expanding the microbial monitoring toolkit: Evaluation of traditional and molecular monitoring methods. *International Biodeterioration and Biodegradation*, 81, 51–56.
42. Kip, N., & A van Veen, J. (2015). The dual role of microbes in corrosion. *The ISME Journal*, 9(3), 542–551.
43. Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, 72(3), 367–382.
44. Lenhart, T. R., Duncan, K. E., Beech, I. B., Sunner, J. a, Smith, W., Bonifay, V., ... Suflita, J. M. (2014). Identification and characterization of microbial biofilm communities associated with corroded oil pipeline surfaces. *Biofouling*, 30(7), 823–35.

45. Li, Y., Jia, R., Al-Mahamedh, H. H., Xu, D., & Gu, T. (2016). Enhanced biocide mitigation of field biofilm consortia by a mixture of D-amino acids. *Frontiers in Microbiology*, 7(JUN), 1–13.
46. Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and Qualitative B Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, 73(5), 1576–1585.
47. Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and Qualitative B Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology*, 73(5), 1576–1585.
48. Madigan, M., Martinko, J., Parker, J. (2009). Brock Biología de los Microorganismos. Treceava edición. Pearson.
49. Magot, M., Ollivier, B., & Patel, B. K. C. (2000). Microbiology of petroleum reservoirs. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 77(2), 103–116.
50. Mand, J., Park, H. S., Jack, T. R., & Voordouw, G. (2014). The role of acetogens in microbially influenced corrosion of steel. *Frontiers in Microbiology*, 5, 1–14.
51. Menes, R. J., & Muxí, L. (2002). *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. *International Journal of Systematic and Evolutionary Microbiology*, 52, 157–164.
52. Muthukumar, N., Rajasekar, A., Ponmari, S., Moh, S., Maruth, S., Muralidh, S., ... Raghavan, M. (2003). Microbiologically influenced corrosion in petroleum product pipelines - A review. *Indian Journal of Experimental Biology*, 41, 1012–1022.
53. Navas-Molina, J. A., Peralta-Sánchez, J. M., González, A., McMurdie, P. J., Vázquez-Baeza, Y., Xu, Z., ... Knight, R. (2013). *Advancing Our Understanding of the Human Microbiome Using QIIME. Methods in Enzymology* (1st ed., Vol. 531). Elsevier Inc.
54. Neria-González, I., Wang, E. T., Ramírez, F., Romero, J. M., & Hernández-Rodríguez, C. (2006). Characterization of bacterial community associated to biofilms of corroded oil pipelines from the southeast of Mexico. *Anaerobe*, 12(3), 122–133.
55. Norsworthy, R. (2014). Understanding corrosion in underground pipelines: basic principles. In *Underground pipeline corrosion: Detection, analysis and prevention* (pp. 3–34). Woodhead Publishing Limited.
56. Parshina, S. N., & Stams, A. J. M. (2015). *Soehngenia*. *Bergey's Manual of Systematics Archaea and Bacteria*, 0–2.
57. Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., ... Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1).
58. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), 590–596.
59. Rajasekar, A., Anandkumar, B., Maruthamuthu, S., Ting, Y.-P., & Rahman, P. K. S. M. (2010). Characterization of corrosive bacterial consortia isolated from petroleum-product-transporting pipelines. *Applied Microbiology and Biotechnology*, 85(4), 1175–1188.
60. Ramírez Reyes, J. L., et al., (2011). La corrosión atmosférica y su impacto en la economía veracruzana. *Revista de divulgación científica y tecnológica de la Universidad Veracruzana*. 1-2.
61. Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., ... Hoffman, N. G. (2014). Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. *Applied and Environmental Microbiology*, 80(24), 7583–7591.
62. Santa Anna, L. M., Sebastian, G. V., Menezes, E. P., Alves, T. L. M., Santos, A. S., Pereira, N., & Freire, D. M. G. (2002). Production of biosurfactants from *Pseudomonas aeruginosa* PA1 isolated in oil environments. *Brazilian Journal of Chemical Engineering*, 19(2), 159–166.
63. Sastri, V. S. (2014). Types of corrosion inhibitor for managing corrosion in underground pipelines. In *Underground Pipeline Corrosion* (pp. 166–211). Woodhead Publishing Limited.

64. Schink, B. (2006). The Genus *Pelobacter*. *The Prokaryotes*, 7, 5–11.
65. Shannon C.E., Weaver W. (1949). The mathematical theory of communication. University of Illinois Press. Urbana, IL, EEUU. 144pp.
66. Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, 5(June), 209.
67. Simankova, M. V., & Kotsyurbenko, O. R. (2009). *Acetobacterium*. *Bergey's Manual of Systematics Archaea and Bacteria*, 1–11.
68. Simpson E. H. (1949). Measurement of Diversity. *Nature*. 163, 688.
69. Stirling, G., & Wilsey, B. (2001). Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. *The American Naturalist*, 158(3), 286–299.
70. Sturman, P. J., Goeres, D. M., & Winters, M. A. (1999). Control of hydrogen sulfide in oil and gas wells with nitrite injection. *SPE 56772, SPE Annual Technical Conference and Exhibition*, 1–7.
71. US Federal Highway Administration. <https://www.fhwa.dot.gov/>. Consultado el 7/07/2016.
72. Videla, H. A., & Herrera, L. K. (2005). Microbiologically influenced corrosion: looking to the future. *International Microbiology*, 8, 169–180.
73. Vigneron, A., Alsop, E. B., Chambers, B., Lomans, B. P., Head, I. M., & Tsesmetzis, N. (2016). Complementary microorganisms in highly corrosive biofilms from an offshore oil production facility. *Applied and Environmental Microbiology*, 82(8), 2545–2554.
74. Voordouw, G. (1995). The genus *Desulfovibrio*: The Centennial. *Applied and Environmental Microbiology*, 61(8), 2813–2819.
75. Voordouw, G., Armstrong, S. M., Reimer, M. F., Fouts, B., Telang, a. J., Shen, Y., & Gevertz, D. (1996). Characterization of rRNA genes from oil field microbial communities indicates the presence of a variety of sulfate - reducing, fermentative, and sulfide - oxidizing bacteria. *Applied And Environmental Microbiology*, 62(5), 1623–1629.
76. Woese, C. R., & 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.
77. You, J., Wu, G., Ren, F., Chang, Q., Yu, B., Xue, Y., & Mu, B. (2016). Microbial community dynamics in Baolige oilfield during MEOR treatment, revealed by Illumina MiSeq sequencing. *Applied Microbiology and Biotechnology*, 100(3), 1469–1478.
78. Yuan, S. J., & Pehkonen, S. O. (2007). Microbiologically influenced corrosion of 304 stainless steel by aerobic *Pseudomonas* NCIMB 2021 bacteria: AFM and XPS study. *Colloids and Surfaces B: Biointerfaces*, 59(1), 87–99.
79. Zhu, X. Y., Lubeck, J., & Kilbane II, J. J. (2003). Characterization of Microbial Communities in Gas Industry Pipelines. *Applied and Environmental Microbiology*, 69(9), 5354–5363.