

Comparación de tres métodos moleculares para el análisis de procariontes ambientales en el mar del canal de Yucatán, México

Analysis of three molecular methods to describe sea Yucatan, Mexico environmental prokaryotic taxa

Laura Espinosa-Asuar,¹ Ana E. Escalante,²
Luisa I. Falcón,¹ Germán Bonilla-Rosso,¹
Santiago Ramírez-Barahona,¹ Luis E. Eguiarte¹
y Valeria Souza¹

¹ Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, México 04510

² Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, México 04510

e-mail: souza@unam.mx

Espinosa-Asuar L., A. E. Escalante, L. I. Falcón, G. Bonilla-Rosso, S. Ramírez-Barahona, L. E. Eguiarte y V. Souza. 2014. Comparación de tres métodos moleculares para el análisis de procariontes ambientales en el mar del canal de Yucatán. *Hidrobiológica* 24 (3): 257-270.

RESUMEN

En este trabajo describimos la comunidad procarionte de una muestra de agua marina del canal de Yucatán. Para la determinación de la diversidad microbiana se usaron tres métodos moleculares: 1) T-RFLPs. 2) Secuenciación de amplicones (bibliotecas de clones). 3) Secuenciación *shotgun* de un metagenoma. Como un segundo objetivo, se presenta una comparación de los alcances y los límites de cada uno de estos tres métodos. Para esta comparación, se tomaron en cuenta tres criterios: el número de unidades taxonómicas detectadas, la precisión en la asignación taxonómica y el costo del estudio. Los *taxa* más abundantes fueron *Candidatus* Portiera OTU 3744 (equivalente al clado SAR86) y *Candidatus* Pelagibacter. Los resultados mostraron que la estrategia de secuenciación *shotgun* de ADN es la más poderosa en términos de unidades taxonómicas detectadas, mientras que los datos que se obtuvieron por T-RFLPs y con la biblioteca de clones, representan sólo una submuestra de la biblioteca de fragmentos generados mediante *shotgun*. En cuanto a resolución filogenética (determinación taxonómica), la aproximación más precisa fue la secuenciación de bibliotecas de clones. Los costos de las tres estrategias varían considerablemente, pero sus alcances también lo hacen. Por lo tanto, es importante tomar en consideración que algunas preguntas ecológicas y evolutivas sólo pueden ser contestadas específicamente por una u otra metodología.

Palabras clave: Clonación, diversidad procarionte, NGS, *shotgun*, T-RFLPs.

ABSTRACT

In this study we describe a prokaryotic community from a seawater sample obtained in the Yucatan Channel, using for this purpose three molecular methods: 1) T-RFLPs. 2) Sequencing of amplicons (clone libraries). 3) Metagenome *shotgun* sequencing; the three are useful for the determination of microbial diversity. We also present a comparison of the scope and limits of each method. The comparison took into account three criteria: the number of taxonomic units detected, the taxonomic assignment accuracy and the cost of the study. The most abundant taxa were *Candidatus* Portiera OTU 3744 (equivalent to SAR86 clade) and *Candidatus* Pelagibacter. The results showed that the *shotgun*

sequencing strategy is the most powerful in terms of detected taxonomic units, while the data obtained by T-RFLPs and clone library methods represent only a subsample of the shotgun fragment library. Regarding phylogenetic resolution (taxonomical determination), the more accurate approach is the sequencing of clone libraries. The costs of the three strategies vary considerably, but so does its scope. Therefore, it is important to consider that one, or another methodology, can only specifically answer some ecological and evolutionary questions.

Key words: Cloning, prokaryotic diversity, NGS, *shotgun*, T-RFLPs.

INTRODUCCIÓN

La diversidad microbiana es una pieza de conocimiento indispensable para comprender el funcionamiento de los ecosistemas, ya que los microbios son un componente fundamental en los procesos biogeoquímicos (Fierer & Lennon, 2011; Nazaries *et al.*, 2013). Sin embargo, este componente fue pasado por alto durante años en la literatura ecológica, debido en parte a la falta de herramientas técnicas que permitieran analizar y modelar el papel de los microorganismos en los ecosistemas (Hughes *et al.*, 2006). En las dos últimas décadas ha surgido una gran variedad de herramientas moleculares para el estudio de la composición de comunidades microbianas enteras (Green *et al.*, 2008; Segata *et al.*, 2013; Zarraonandia *et al.* 2013). Estos avances tecnológicos han abierto la puerta al crecimiento sin precedentes de estudios ecológicos de los microorganismos en sus ambientes naturales (Allison *et al.*, 2013; Bonilla-Rosso *et al.*, 2008, 2012; Peimbert *et al.*, 2012).

Los estudios modernos sobre diversidad microbiana inician con la extracción de ADN total de una muestra ambiental. Posteriormente puede trabajarse con técnicas que analizan los genomas completos de la muestra o con otras que se enfocan en la amplificación de un solo gen. Dentro de esta segunda opción existen marcadores moleculares con los que se analiza la diversidad de fragmentos de genes o de regiones específicas (*fingerprints*) o marcadores que se utilizan para analizar secuencias individuales de genes, que generalmente involucran la clonación de los amplicones. Por otro lado, están las técnicas que secuencian directamente el ADN total de la muestra, como es el caso de la secuenciación *shotgun* (secuenciación de fragmentos al azar, la única de este tipo que no precisa de amplificación de regiones particulares) o más recientemente, las técnicas conocidas como secuenciación de nueva generación (*Next Generation Sequencing* o NGS por sus siglas en inglés) (Eguarte *et al.*, 2013; Williamson & Yooseph, 2012). En estas últimas (por ejemplo, pirosecuenciación 454, SOLiD o Illumina/Solexa), sí existe un paso de amplificación al enriquecer las bibliotecas; la amplificación puede hacerse en todos los fragmentos que se generan o también es posible amplificar regiones específicas.

América Latina es una de las regiones a nivel mundial con mayor biodiversidad en macroorganismos (CONABIO, 1988; Rzedowski, 2006). Sin embargo, de acuerdo con una revisión realizada en ISI Web of Knowledge (rango: 1980-2009, palabra clave:

microbial diversity), esta región del mundo ha publicado el menor número de reportes en comparación con EE UU, Europa y Asia. La revisión realizada también reflejó un incremento de publicaciones en todo el mundo en la última década analizada (2000-2009). Una probable causa de estas observaciones es que los análisis moleculares se han vuelto más accesibles en los últimos años, por lo que se ha hecho posible describir y estudiar la diversidad de microorganismos en cualquier lugar del mundo (para una revisión véase Lindstrom & Langenheder, 2012).

El objetivo de este trabajo es la evaluación de tres métodos moleculares: *fingerprints*, en particular Fragmentos Terminales de Restricción Polimórficos por Tamaño (T-RFLPs por sus siglas en inglés), biblioteca de clones derivadas de PCR y biblioteca de secuencias generadas por *shotgun*, utilizados para analizar la diversidad procarionte de una muestra ambiental de agua marina del canal de Yucatán (muestra GS017, reportada por Rusch *et al.*, 2007). La composición taxonómica de la comunidad microbiana fue analizada aplicando el marcador del gen 16S ADNr. Los tres criterios de análisis para evaluar las metodologías empleadas fueron: 1) la comparación del número de unidades taxonómicas identificadas por cada método, 2) la precisión en la determinación de los grupos taxonómicos y 3) los costos para cada una de las metodologías.

MATERIALES Y MÉTODOS

Se analizó la composición microbiana en una muestra de agua marina (GS017) que fue tomada a 2 m de profundidad en el mar abierto del canal de Yucatán (20°31'21"N; 85°24'49"O), a unos 150 km al este de la isla de Cozumel, México (Rusch *et al.*, 2007).

Extracción de ADN. El ADN se extrajo de una porción del filtro obtenido en la expedición del grupo de Rusch (Rusch *et al.*, 2007), utilizando el método reportado en ese mismo trabajo.

Métodos asociados a PCR (T-RFLPs y biblioteca de clones). El gen que codifica para el ARN ribosomal 16S (gen 16S ADNr) fue amplificado por PCR con los oligonucleótidos universales para bacterias y arqueas F515 y R1492 (Turner *et al.*, 1999), cuyas secuencias son: F515 (5'-GCGGATCCTCTAGACTGCAGTGCCAGCAGCCGCGGTAA-3') R1492 (5'-GGCTCGAGCGGCCGCCCGGTTACTTGTACGACTT-3').

Para la amplificación de T-RFLPs, el oligonucleótido F515 estaba marcado con fluorescencia VIC en el extremo 5' (Applied

Biosystems). Las condiciones de la reacción de PCR fueron las siguientes: *buffer* para PCR 1X, 1.65 mM de MgCl₂, 0.2 mM de dNTPs, 0.06 mM de cada oligonucleótido, 1 U de Taq polimerasa (Applied Biosystems, Foster City, CA) y 5% de DMSO. Todas las reacciones se llevaron a cabo en un termociclador MJ Research (Watertown, MA) con los siguientes ciclos de reacción: 94 °C 4 min, 35 ciclos de 92 °C 1.5 min, 50 °C 1.5 min, 72 °C 2 min, completando con 72 °C 10 min. Para tener una mejor representación de la muestra, se hicieron tres reacciones de PCR por separado (tanto para clones como para T-RFLPs). El amplificado (aprox. 977 pb) de las tres reacciones se mezcló y purificó a partir de un gel de agarosa al 2%, con el kit QIAquick (Qiagen Inc., Netherlands).

La técnica de T-RFLPs se fundamenta en una reacción de restricción, con la que se generan distintos fragmentos de ADN que componen un patrón característico (denominado "la huella" o *fingerprint*) de la comunidad analizada. El tamaño de cada fragmento es considerado como equivalente a una unidad taxonómica operativa (OTU, por sus siglas en inglés), y corresponde al sitio de restricción más cercano al oligonucleótido fluorescente que se utilizó en la reacción de PCR (Osborn *et al.*, 2000). El producto obtenido por PCR (fragmento del gen 16S ADNr) fue digerido enzimáticamente con *Alu I* (Promega, Madison, WI) en reacciones de 20 µl. Para cada reacción se utilizaron 1 unidad de *Alu I* y 50 ng de producto de PCR. Las reacciones fueron incubadas en un termociclador MJ Research durante 3 h a 37 °C, seguidas de la inactivación de la enzima durante 30 min a 65 °C. Se realizaron tres corridas independientes de la muestra, usando un secuenciador automático ABI Avant 3100 PRISM® (Applied Biosystems, Foster City, CA). Con esto se obtuvieron los tamaños y alturas de los Fragmentos de Restricción Terminales (TRFs, por sus siglas en inglés), en forma de picos en un cromatograma. Para evitar fluorescencia inespecífica, sólo los picos que estuvieron presentes en las tres réplicas, y que presentaban alturas > 10 unidades de fluorescencia, fueron considerados en el análisis. Cada TRF obtenido en este análisis se consideró una OTU.

Biblioteca de clones. La mezcla de amplificadas del gen 16S ADNr ya purificada, se clonó en el vector plasmídico pCR2.1, de acuerdo con las instrucciones del fabricante (Invitrogen Corp., San Diego, CA). Los plásmidos se aislaron con el *kit* de extracción SNAP® Miniprep Kit (Invitrogen Corp.) y aquellos que contenían el inserto fueron secuenciados usando los oligonucleótidos M13F y M13R, en un secuenciador ABI Avant 3100 PRISM® (Applied Biosystems, Foster City, CA). Todas las secuencias fueron revisadas manualmente en BioEdit (Atschul *et al.*, 1990). Para descartar quimeras (artefacto producido por la técnica de PCR, que origina fragmentos formados con secuencias de distinta procedencia), las secuencias obtenidas fueron analizadas con Bellerophon (Huber *et al.* 2004) y Chimera-Check en RDP (Cole *et al.*, 2003). Quimera-Check es una herramienta que actualmente no se encuentra en línea, pero en su lugar puede utilizarse UCHIME (Edgar *et al.*, 2011) o DECIPHER (Wright *et al.*, 2012), disponibles también

en RDP. También se consideraron quimeras aquellas secuencias que produjeron discrepancias en el orden de las ramas de dos árboles filogenéticos generados con regiones distintas del gen 16S ADNr (533-873 y 874-1215, de acuerdo con la numeración de este gen en *Escherichia coli*). El tamaño final de la biblioteca, después de eliminar las quimeras, fue de 93 secuencias. Se hizo una selección de una sola secuencia para cada filotipo (se define filotipo como secuencias distintas). El corte utilizado para diferenciar estos filotipos o secuencias tipo (con una diferencia mayor al 98% entre cada filotipo) fue sugerido por Stackebrandt y Ebers en 2006. De esta manera, las 93 secuencias que se obtuvieron en la biblioteca quedaron representadas por un total de 42 secuencias, las cuales fueron registradas en GenBank con los números de acceso de KF303144 a KF303185.

Método independiente de PCR (secuenciación genómica por *shotgun*). Las secuencias del gen 16S ADNr en el metagenoma de la muestra GS017 correspondiente al Canal de Yucatán (Rusch *et al.*, 2007) fueron detectadas mediante rRNA-HMM (Huang *et al.*, 2009) y sólo se retuvieron aquellas mayores que 200 pb. Estas secuencias fueron inspeccionadas para la detección de quimeras mediante Mallard (Ashelford *et al.*, 2006) y aquellas que pasaron la prueba fueron alineadas y clasificadas en Greengenes (DeSantis *et al.*, 2006b). La alineación fue inspeccionada manualmente. Dada la naturaleza de la base de datos metagenómica obtenida por *shotgun*, los fragmentos recuperados no son superpuestos a lo largo de regiones extensas, lo que previene la construcción de OTUs mediante agrupación por similitud a nivel de secuencia.

Determinación taxonómica. Para determinar las categorías taxonómicas a nivel *phylum* de ambas bibliotecas (clones y *shotgun*), se utilizó la herramienta Classify de Greengenes en octubre del 2011 (DeSantis *et al.*, 2006a, 2006b). La asignación taxonómica a nivel de orden, sólo para la biblioteca de clones, se determinó con Simrank (también herramienta de Greengenes), utilizando las filogenias de referencia que recientemente se actualizaron en esta base de datos (McDonald *et al.*, 2012) y se clasificaron aquellas secuencias con > 98% de similitud para una asignación taxonómica más confiable (Stackebrandt & Ebers, 2006). En caso de haber obtenido < 98%, el índice de similitud correspondiente se especifica en el texto.

Análisis filogenético. Se realizó un análisis filogenético con las secuencias del gen 16S ADNr obtenidas en la biblioteca de clones con el fin de corroborar la identidad taxonómica generada en Greengenes, utilizando las 42 secuencias tipo o filotipos. Para cada una se descargaron las dos secuencias con mayor similitud de las bases de datos públicas (una obtenida mediante BLAST (Altschul *et al.*, 1990), en junio de 2011, y otra en Greengenes en esa misma fecha) y se usaron como referencia para la filogenia. Todas estas secuencias fueron alineadas en Greengenes (DeSantis *et al.*, 2006a) y el alineamiento se revisó manualmente en BioEdit (Hall, 1999). Para determinar la posición filogenética de

estas secuencias, se realizaron análisis de Neighbor-Joining en MEGA5 (Tamura *et al.*, 2011), de máxima parsimonia en PAUP 4.0 (Swofford, 2002) y de máxima verosimilitud en phyML 3.0 (Guindon *et al.*, 2010), usando las secuencias de arquea como grupo externo. El mejor modelo de sustitución se obtuvo con Model Generator 0.85 (Keane *et al.*, 2006).

Análisis de TRFs con la biblioteca de clones. Para la clasificación taxonómica de los TRFs obtenidos con T-RFLPs y para la comparación de estos datos con la biblioteca de clones, se utilizaron 59 secuencias de esta biblioteca, seleccionadas en función de la calidad que tenían en el extremo 5'. Se generó un mapa de restricción (digestión *in silico*) con *Alu I* en BioEdit (Altschul *et al.*, 1990) para cada una de estas secuencias, cuya determinación taxonómica ya se había hecho previamente, con el fin de identificar el tamaño de TRF correspondiente a cada taxón.

Comparación de costos de las estrategias moleculares. Los costos para cada estrategia, se calcularon tomando en cuenta los precios en promedio de acuerdo con las cotizaciones de distintas compañías. En el caso de las secuencias obtenidas por *shotgun*, se tomó como referencia a Thomas *et al.* (2012). Para clones y T-RFLPs, el costo fue calculado enlistando el material necesario requerido para ambas técnicas (Taq polimerasa, dNTPs, materiales), añadiéndose un costo aproximado para la extracción de ADN. Para los T-RFLPs se consideró el costo por corrida en el secuenciador automático de capilares (rendimiento de capilares, materiales utilizados, POP6, formamida, *buffer* y marcador de peso molecular fluorescente) y para clonación se consideró el costo de los *kits* de clonación y extracción de plásmidos. Se tomó en cuenta el menor precio de secuenciación por muestra, comparando dos unidades de secuenciación: 1) Washington University High Throughput Sequencing (cotización) y 2) la Unidad de Síntesis y Secuenciación del Instituto de Biotecnología (Universidad Nacional Autónoma de México, disponible en línea, consultado 2 noviembre 2013). Con fines comparativos, también se calcularon costos de técnicas de secuenciación de nueva generación (NGS), consultando para ello las siguientes páginas: Langebio (cotización), UCDavis, University of Tennessee y University of Virginia (disponibles en línea, consultados 2 noviembre 2013).

RESULTADOS

Diversidad, medida como número de unidades taxonómicas detectada. La Tabla 1 resume el número de grupos taxonómicos que fue posible determinar para cada técnica, y en cada apartado se discutirán estos datos con detalle.

T-RFLPs versus biblioteca de clones. Con la técnica de T-RFLPs se detectaron 39 OTUs (fragmentos); la mayoría de ellos resultó con poca abundancia y sólo algunos fueron muy abundantes (Fig. 1). Para comparar estos datos con la biblioteca de clones, se utilizaron las secuencias de esa biblioteca ($n = 59$) cuyos cromogramas mostraron buena calidad en el extremo 5' (con picos bien

Tabla 1. Número de unidades taxonómicas detectadas por las tres metodologías utilizadas: biblioteca de clones, *shotgun* y T-RFLPs.

	Biblioteca de clones	<i>Shotgun</i>	T-RFLPs
OTUs (fragmentos)	17	ND	39
OTUs (secuencias tipo)	42	ND	ND
<i>Phyla</i>	7	34	ND

definidos, separados entre sí y con poco ruido de fondo). Con estas secuencias, se realizó una restricción *in silico* que generó 17 fragmentos distintos (OTUs), cuya clasificación taxonómica (correspondiente a cada una de las 59 secuencias seleccionadas) se especifica en la Tabla 2. Sólo nueve OTUs coincidieron por tamaño en ambas metodologías (Fig. 1) y el OTU más abundante obtenido para ambas metodologías fue el fragmento de 123 pares de bases. Este fragmento puede provenir de una gran variedad de secuencias de la biblioteca de clones correspondientes a distintos *taxa* de bacterias: *Candidatus Pelagibacter* y Rhodospirillaceae, que son Alfaproteobacterias. *Vibrionales* y el OTU 3370, en el caso de Gammaproteobacteria, y por último del orden Sphingobacteriales que pertenece al *phylum Bacteroidetes*.

Debido a que en la mayoría de los casos los fragmentos de *shotgun* no tienen la misma región terminal que los productos de PCR (con los cuales se generan los fragmentos de restricción terminales), no fue posible realizar una digestión *in silico* para compararlos con las otras dos metodologías.

Biblioteca de clones versus *shotgun*. El tamaño promedio de las 860 secuencias del gen 16S ADN_r de la secuenciación por *shotgun* fue de 400 pares de bases y se detectaron 34 *phyla*, mientras que en la biblioteca de clones para este mismo gen sólo se pudieron detectar siete *phyla* (Fig. 2), con una extensión de cerca de 700 pares de bases por secuencia. Todos los *taxa* que componen la biblioteca de clones fueron detectados en la biblioteca de *shotgun*. Alfaproteobacteria y Gammaproteobacteria fueron los *phyla* más abundantes recuperados por ambas técnicas (20.4% Alfaproteobacteria y 18.8% Gammaproteobacteria en la biblioteca por *shotgun* y 53.7% y 34.4% en la biblioteca de clones, respectivamente). Bacteroidetes fue uno de los *phyla* más abundantes (15.11%) en las secuencias obtenidas por *shotgun*, aunque fue detectado en menor proporción en la biblioteca de clones (4.3%). Representantes poco abundantes, que resultaron únicos o con dos secuencias en la biblioteca de *shotgun* (con 0.11% o con 22% del total) fueron los *phyla* Acidobacteria, Chlorobi, Fusobacteria, Nitrospirae, OP11, SR1, Thermi, TM6 y WS3. En el caso de la biblioteca de clones, los *phyla* con un solo representante (Euryarchaeota, Archaea), Cyanobacteria y *Marine group A* (o Grupo A marino), con ~1% del total, tuvieron una abundancia intermedia en la biblioteca obtenida por *shotgun*.

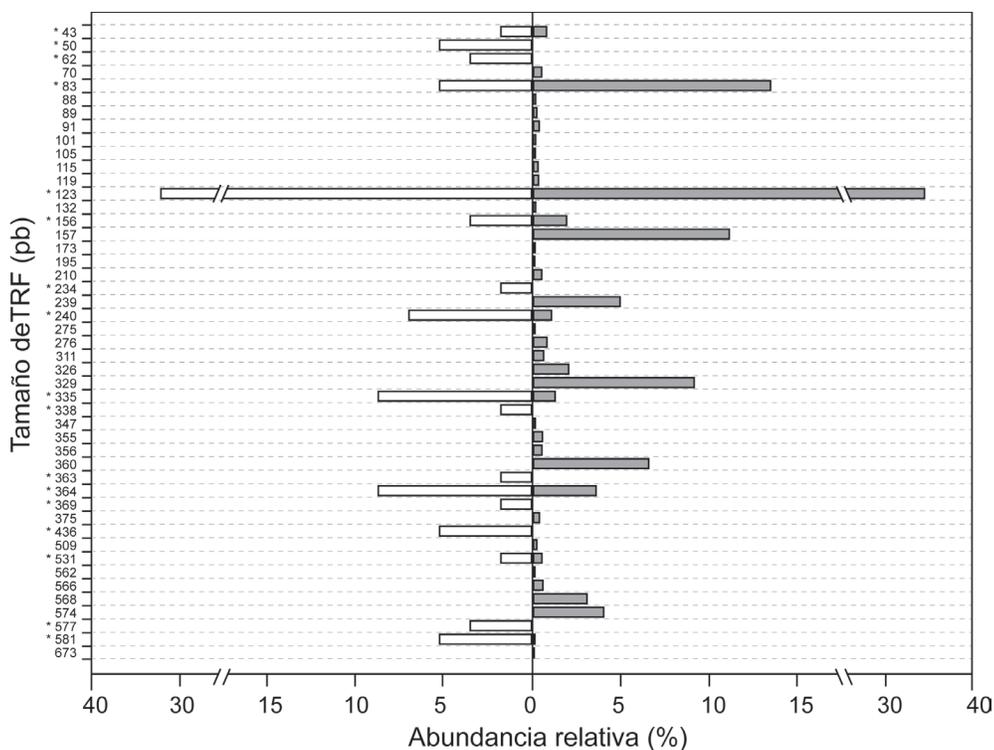


Figura 1. Fragmentos de restricción del gen 16S ADN_r generados con T-RFLPs y con la biblioteca de clones. El largo de las barras representa la abundancia relativa de cada fragmento en la muestra total. Las barras blancas corresponden a fragmentos de restricción obtenidos *in silico* sobre las secuencias provenientes de la biblioteca de clones y las barras grises corresponden a fragmentos de restricción generados con T-RFLPs. Con asterisco se resaltan los fragmentos a los que fue posible asignarles clasificación taxonómica, la cual se especifica en la Tabla 2.

Tabla 2. Clasificación taxonómica de los OTUs que fue posible designar, utilizando los fragmentos *in silico* provenientes de la biblioteca de clones. Todas las secuencias se clasificaron utilizando un índice de similitud (Simrank) > 98%, exceptuando los OTUs especificados en el paréntesis.

TRF (pb)	Clasificación Simrank de acuerdo con los fragmentos <i>in silico</i>
43	<i>Candidatus</i> Portiera (OTU 3744)
50	<i>Candidatus</i> Pelagibacter (OTU 2827)
62	<i>Candidatus</i> Portiera (OTU 3744)
83	<i>Candidatus</i> Pelagibacter (OTU 2828)
123	<i>Candidatus</i> Pelagibacter (OTU 2827), Rhodospirillaceae (OTU 2793 sc), Sphingobacteriales (OTU1170), <i>Photobacterium angustum</i> (OTU 3974, 93%), Gammaproteobacteria (OTU 3370)
156	<i>Candidatus</i> Portiera (OTU 3744), Rhodospirillaceae (OTU 2793, 87.74%)
234	<i>Candidatus</i> Portiera (OTU 3744)
240	<i>Candidatus</i> Portiera (OTU 3744)
335	<i>Candidatus</i> Portiera (OTU 3744)
338	<i>Pseudoalteromonas</i> (OTU 3440)
363	<i>Candidatus</i> Portiera (OTU 3744, 94.69%)
364	<i>Candidatus</i> Pelagibacter (OTU 2827, 89.29%), Rhodospirillaceae (OTU 2793, 94.14%)
369	Vibrionales (swat-3)
436	<i>Candidatus</i> Portiera (OTU 3744), OM60 (OTU 3802)
531	<i>Marine group</i> II (OTU 198)
577	Deltaproteobacteria (OTU 3167)
581	<i>Candidatus</i> Pelagibacter (OTU 2827, 94.34%)

Por otra parte, la biblioteca de clones se analizó a un nivel jerárquico más restringido, para lo cual, las 93 secuencias fueron agrupadas con un 98 % de similitud (Stackebrandt & Ebers, 2006) en 42 tipos de secuencias (filotipos o Sec Tipo en Fig. 3). Estas 42 secuencias son equivalentes a 42 unidades taxonómicas (OTUs). Veinticuatro de ellas resultaron únicas y cuatro resultaron ser las más abundantes, con más de cinco clones por tipo, asociadas a los géneros marinos no cultivables *Candidatus Pelagibacter* y *Candidatus Portiera* OTU 3744. El resto tuvo menor abundancia (de dos a cuatro clones por tipo). Sólo 17 de los 42 filotipos pudieron ser clasificados en estos niveles taxonómicos más bajos (Fig. 3).

Grupos taxonómicos. El árbol filogenético generado con los 42 filotipos de la biblioteca de clones muestra que la clasificación de la base de datos de Greengenes es congruente con la distribución filogenética de las ramas (Fig. 3) a los dos niveles analizados (*phylum* y orden). Se muestra la clasificación obtenida a nivel orden o menor sólo para aquellas secuencias que dieron > 98% de similitud en Simrank. En contraste, con las secuencias de *shotgun* no fue posible hacer una filogenia, ya que corresponden a fragmentos que cubren distintas regiones del gen 16S ADNr, lo cual no permite un alineamiento adecuado.

Costos: El costo aproximado actual de una biblioteca (una sola muestra), obtenida utilizando secuenciación genómica por *shotgun*, es de 400,000 dólares (EE UU) (Thomas *et al.*, 2012).

Para hacer una biblioteca de 96 clones se necesita una inversión de 465 dólares y, por último, una muestra de ADN analiza-

da con T-RFLPs tiene un costo de 25 dólares. La Tabla 3 contiene los detalles de estos estimados y los costos actuales para metodologías de nueva generación (NGS).

DISCUSIÓN

En este trabajo presentamos la comparación de tres metodologías para el estudio de diversidad de comunidades microbianas para lo que se usó como referencia una muestra ambiental del canal de Yucatán. Debido a la naturaleza de este estudio, la interpretación ecológica de los datos es necesariamente limitada en cuanto a los aspectos ecológicos de la muestra analizada. Sin embargo, es de gran utilidad al comparar detalladamente los alcances de las tres metodologías utilizadas: T-RFLPs, bibliotecas de clones y *shotgun*.

Diversidad de la muestra. Independientemente del método empleado, la distribución de la estructura de la comunidad observada fue similar a la que se ha descrito para la mayoría de las comunidades bacterianas en diversos ambientes (Woodcock *et al.*, 2007; Pedrós-Alió, 2012): pocos *taxa* muy abundantes y muchos poco abundantes (Figs. 1 y 2). Adicionalmente, se observó gran diversidad de fragmentos (T-RFLPs) y *phyla* (*shotgun*), lo cual podría explicarse por la alta productividad primaria que ha sido reportada en la zona de donde proviene la muestra (CONABIO, 2013; Barton *et al.*, 2010).

Más allá de los patrones generales de diversidad, la información taxonómica fina sólo pudo lograrse con las secuencias de la biblioteca de clones al detectar *taxa* a niveles jerárquicos

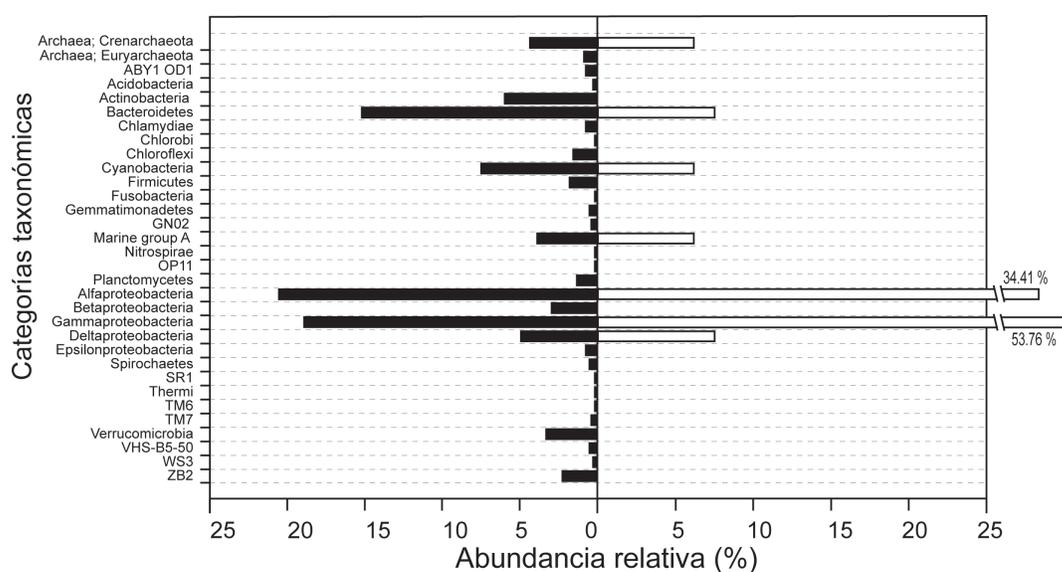


Figura 2. Diversidad de *phyla* detectados con biblioteca de clones y secuenciación genómica por *shotgun* del gen 16S ADNr. El largo de las barras representa la abundancia relativa de cada *phylum* en la muestra total. Las barras negras corresponden a *phyla* identificados en la base de datos de la secuenciación de ADN por *shotgun* y las barras blancas corresponden a *phyla* identificados en la biblioteca de clones.

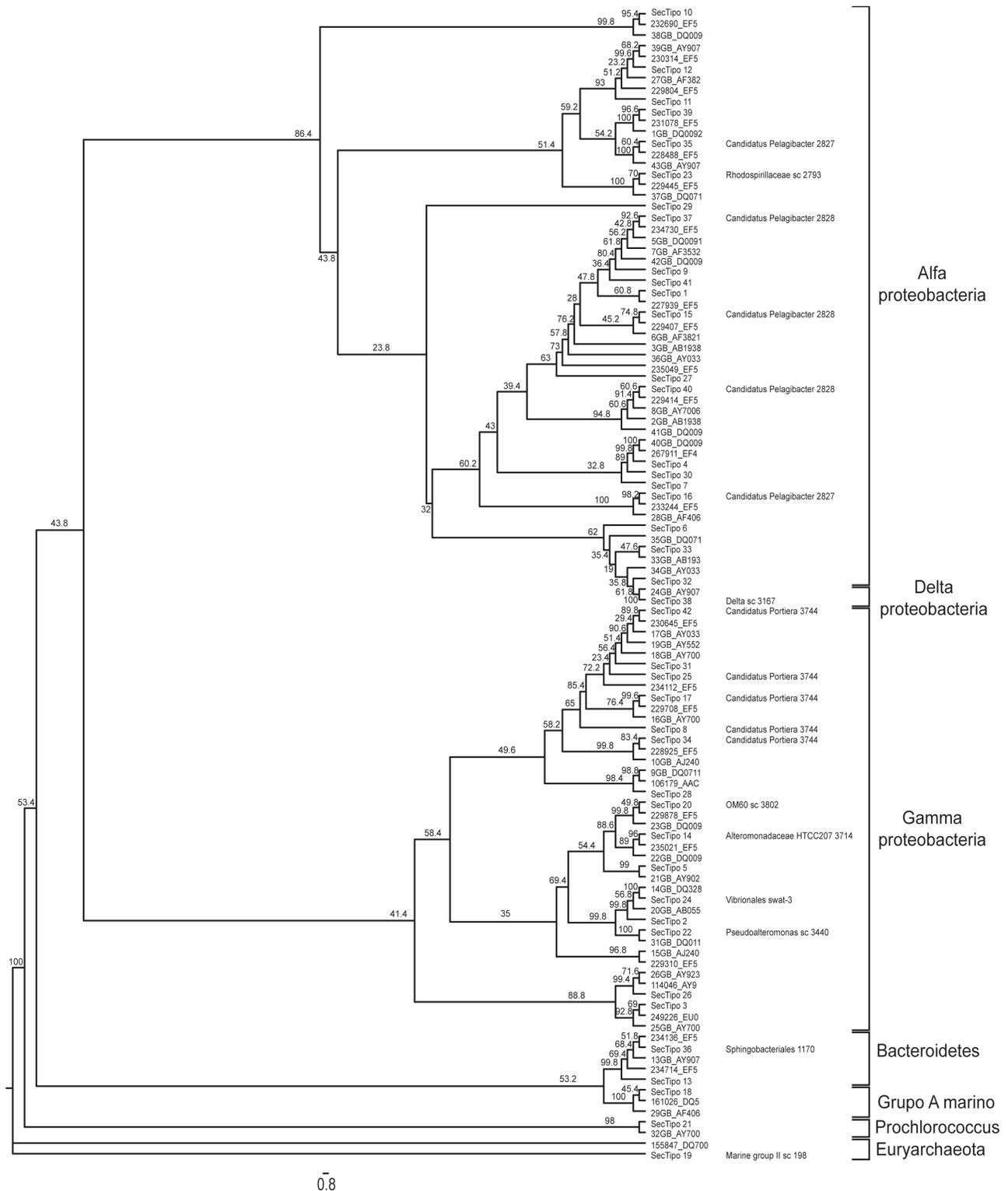


Figura 3. Representación filogenética de la diversidad del gen 16S ribosomal encontrada en la biblioteca de clones de una muestra de agua del canal de Yucatán. El árbol fue construido con un algoritmo de máxima verosimilitud (no se muestran los árboles construidos con los otros dos algoritmos, ya que generaron una topología similar). Las secuencias SecTipo corresponden a las 42 secuencias tipo generadas en la biblioteca de clones, y el resto son secuencias obtenidas de bases de datos públicas. La clasificación correspondiente (utilizando Simrank), a nivel orden o menor, se asignó sólo en las secuencias tipo con similitud > 98 % a la base de datos. El número en cada rama corresponde al valor de *bootstrap* (%) sobre 500 réplicas.

Tabla 3. Detalles de los costos en USD estimados para los tres métodos moleculares utilizados, así como el costo de algunas técnicas de secuenciación de nueva generación.

	Métodos en este trabajo		Plataformas de secuenciación de Nueva Generación (NGS)						
	T-RFLPs	Shotgun	Clonas	454 GS Junior Langebio	454 GSxL Langebio	454 GS Titanium Langebio	454 GS Titanium U. de Tennessee	Illumina UCDavis	Ion Proton Langebio
Extracción ADN	22	22	22	22	22	22	22	22	22
PCR	1		1						
Clonación			15						
Extracción plásmido			152						
Secuenciación	4		275						
Precio reportado		400 000 (Thomas <i>et al.</i> , 2012)		2 772	5 343	5 300 y 50 900*	16 000	3 553	2 825
Total por muestra	25	400 022	465	2 794	5 365	5 322 y 50 922*	16 022	3 575	2 847
Tipo de datos obtenidos	Un perfil de fragmentos por muestra	1 millón y medio de secuencias mayores que 700 pb	96 secuencias	92 000 secuencias de 31 pb	1 millón de secuencias de 500 pb	1 millón de secuencias de 380 pb	1 millón de secuencias de 380 pb	90 millones de secuencias de 100 pb	20 millones de secuencias de 200 pb

*Costo total de una placa que incluye la preparación de 192 bibliotecas (para más información, consultar el texto).

más restringidos, incluso a nivel de género (Fig. 3). Las secuencias corresponden a bacterias que han sido reportadas como parte de comunidades oceánicas en todo el mundo. Los dos *taxa* más abundantes fueron *Candidatus Pelagibacter* (Alfaproteobacteria) y *Candidatus Portiera* OTU 3744 (Gammaproteobacteria, equivalente al clado SAR86). Son bacterias heterótrofas aeróbicas que se han especializado en utilizar distintas fuentes de carbono, aparentemente para reducir la competencia (Dupont *et al.*, 2011). En contraste también se encontraron potenciales competidores como *Pseudoalteromonas* 3440 y Vibrionales (*swat* 3), capaces de inhibir el crecimiento de muchas especies de bacterias (Skovhus *et al.*, 2007, J. Craig Venter Institute, disponible en línea, consultado 2 noviembre 2013). Otro grupo detectado fue el de las bacterias fotótrofas de la familia Rhodospirillaceae (e.g. OTU 2793 y OM60; Madigan *et al.*, 2003; Zheng *et al.*, 2011) y las arqueas del grupo marino II (OTU 198), de las que recientemente se ha reportado una proteorodopsina que es ancestral a la de las bacterias (Iverson *et al.*, 2012). Adicionalmente fueron identificados algunos *taxa* bacterianos con utilidad industrial o ecológica, como Alteromonadaceae (HTCC207 OTU 3714), el cual es importante para el reciclaje y biodegradación de materia orgánica, en particular de hidrocarburos (Jin *et al.*, 2012). También encontramos a las Deltaproteobacteria (OTU 3167), cuyos linajes de mares fríos (equivalentes al clado SAR324 en la taxonomía de Silva (Pruesse *et al.*, 2007)) desempeñan un papel importante en el reciclaje de fosfato orgánico disuelto (Brown & Donachie, 2007). Por último, se identificaron secuencias del orden Sphingobacteriales (OTU 1170), cuya capacidad de adherencia con algas ha sido recientemente reportada (Gómez-Pereira *et al.*, 2012).

El inventario de la diversidad de microorganismos en distintos hábitats abre la posibilidad de investigar la relación entre ciertos grupos microbianos (algas, protozoarios, bacterias, hongos, virus) y aspectos funcionales de los ecosistemas. La caracterización de tal diversidad de organismos a través de estrategias genómicas independientes de cultivo, permite el monitoreo de ambientes naturales y el efecto de perturbaciones en la diversidad, como uno de los controladores de las funciones. Por ejemplo, el clado SAR11, al que pertenece *Candidatus Pelagibacter*, ha sido utilizado para monitorear los patrones de calentamiento en los mares, lo que potencialmente permitirá un mejor entendimiento de los patrones de respuesta de éstas y otras especies de bacterioplancton al cambio climático (Giovannoni & Vergin, 2012). Por otra parte, existe una estrecha relación entre la estructura de las comunidades bacterianas en el océano con la productividad primaria, medida como clorofila *a* (Falcón *et al.*, 2008, Yilmaz *et al.*, 2012) que podría determinarse en el canal de Yucatán con estudios estacionales en esta zona. Sin embargo, dependiendo del objetivo con el que se realice el estudio, el diseño experimental, número de muestras y número de secuencias o datos genómicos obtenidos, son cruciales en la interpretación ecológica de

la información. De esta manera, los datos aquí presentados son ilustrativos del tipo de información que puede obtenerse, aunque no nos permiten por el momento llegar a conclusiones ecológicas definitivas.

Comparación de metodologías para el análisis de diversidad procarionte. La comparación de los patrones obtenidos con los tres métodos, arrojó claras diferencias en la caracterización de la diversidad de la muestra analizada (i.e. número de *taxa* observados, Tabla 1). En cuanto a las diferencias entre T-RFLPs y la biblioteca de clones, la observación más importante es la evidencia de un muestreo incompleto en ambas técnicas, reflejado en la falta de coincidencias en la presencia y/o ausencia de OTUs. Esta discrepancia es producida principalmente por aquellos fragmentos con menor abundancia que pueden o no ser detectados por la técnica utilizada, efecto característico de una intensidad de muestreo moderada. A pesar de que en el análisis por T-RFLPs un OTU realmente no representa un taxón particular (ya que el fragmento puede ser generado por distintos *taxa* que tengan un mismo sitio de restricción), la caracterización de una muestra a partir del análisis de la diversidad de fragmentos sí permite hacer inferencias indirectas sobre la diversidad de la muestra, como ha sido reportado en análisis de comunidades de suelos por Noguez *et al.* (2005), por Fierer y Jackson (2006) o recientemente por Reed y Martiny (2013). Así, el patrón de fragmentos obtenidos por T-RFLPs para esta muestra indica que la diversidad de *taxa* en el canal de Yucatán es alta.

Por otra parte, la diferencia entre la biblioteca de clones y la de *shotgun* hizo evidente que esta última estrategia de secuenciación permite un muestreo más exhaustivo. La secuenciación tipo Sanger asociada a *shotgun* tiene bajas tasas de error y lecturas largas (hasta 700 pb) (Thomas *et al.*, 2012) lo que permite que la clasificación de las secuencias que se obtienen sea precisa (Hao & Chen, 2012). Además, la secuenciación por *shotgun*, al no tener el sesgo de amplificación (PCR) que otras técnicas sí tienen, ha resultado un excelente modelo para comparar con las dos metodologías que se llevaron a cabo en el presente trabajo. De esta forma, es posible afirmar que la abundancia relativa que se obtuvo con la base de datos de la biblioteca por *shotgun* representa con mayor precisión la estructura del ensamble procarionte del canal de Yucatán. La abundancia relativa de las Alfaproteobacteria y Gammaproteobacteria en la biblioteca de clones (53.7% y 34.4%, respectivamente), por ejemplo, hace pensar que durante la PCR se incorpora un sesgo hacia estos grupos, ya que la abundancia obtenida para estos mismos grupos por *shotgun* fue mucho menor (20.4% y 18.8%). Este sesgo por PCR se da por amplificación preferencial de ciertos grupos filogenéticos y ha sido reportado previamente (ver por ejemplo Frank *et al.*, 2008; Bogert *et al.*, 2011), un efecto que se agudiza si consideramos que el tamaño de muestra (número de clonas analizadas) no es muy grande. Aún con las ventajas y precisión que tiene la secuenciación de ADN por *shotgun* sobre otras, actualmente está utilizándose cada vez

menos en estudios de diversidad metagenómica, ya que la secuenciación por el método Sanger que emplea es muy costosa y, además, requiere mayor tiempo que las técnicas de nueva generación (NGS) para el procesamiento de las muestras. Un ejemplo de estas nuevas técnicas de secuenciación es la pirosecuenciación por *barcoding* (amplicones provenientes de un solo gen, Huse *et al.*, 2008), que además de tener un menor costo, produce un mayor número de secuencias (Suenaga, 2012; Williamson & Josephs, 2012). En promedio (dependiendo de la plataforma utilizada y el tipo de corrida que se elija) se pueden obtener bibliotecas desde unas 1500 hasta un millón de secuencias (Tabla 3) con una longitud más corta que las de clones (desde 100 hasta 500 pb hoy en día). Posibilidades como ésta actualmente han mejorado en gran medida el problema del número de secuencias obtenidas en el análisis de las comunidades microbianas.

La biblioteca de clones aporta información más precisa en cuanto a resolución filogenética, ya que produce secuencias más largas, aunque en menor número. El largo de las secuencias obtenidas permite una detallada asignación taxonómica de los integrantes de la comunidad incluso a nivel de especie. La secuenciación genómica por *shotgun* en cambio tiene una resolución taxonómica menor, debido a la reducida longitud de los fragmentos. Además, estas fracciones de ADN provienen de distintas regiones del gen, por lo que realizar una alineación o un análisis filogenético es casi imposible.

De los tres métodos utilizados, no es difícil ver que la aproximación por T-RFLPs es la más pobre en su resolución taxonómica, ya que se obtienen fragmentos del mismo tamaño, que pueden resultar de grupos filogenéticos muy distintos (Tabla 2). Esta limitación metodológica es evidenciada por la reducida asignación taxonómica que puede hacerse de los fragmentos (Fig. 1 y Tabla 2), algunos de ellos posiblemente producidos por distintos *taxa*, o por grupos filogenéticos que tienen más de un fragmento característico. Sin embargo, es necesario considerar que para laboratorios con bajo presupuesto, la técnica de T-RFLPs representa una buena opción, ya que es metodológicamente sencilla, pero poderosa en cuanto al conocimiento que puede generarse. Además de su bajo costo y sencillez, publicaciones recientes han demostrado la utilidad de los T-RFLPs en investigaciones ecológicas que requieran seguimiento de comunidades en el tiempo o en tratamientos experimentales, en las cuales uno de los requisitos es la identificación de cambios generales en la composición de las comunidades (ver por ejemplo Erwin *et al.*, 2011; Escalante *et al.*, 2008; Lavin *et al.*, 2008). También se han publicado patrones de diversidad o composición microbiana de T-RFLPs que están asociados a factores ambientales (Fierer & Jackson, 2006; Noguez *et al.*, 2008; Reed & Martiny, 2013; Zhao *et al.*, 2012), lo que permite hacer importantes inferencias ecológicas sobre la estructura de las comunidades de bacterias. Esta técnica permite igualmente, de una manera rápida y precisa, comparar *grosso modo* varias localidades al mismo tiempo, así como estimar la diversidad de

cada una, utilizando herramientas clásicas de la Ecología (Noguez *et al.*, 2005; Reed & Martiny, 2013).

La elección de uno u otro método para la caracterización de diversidad microbiana requiere, necesariamente, de la consideración simultánea de los costos y de las capacidades de cada uno. En este sentido, ya se han mencionado detalladamente alcances y limitaciones de los tres métodos en cuanto a diversidad (número de *taxa* observados) y asignación taxonómica, que junto con los datos de costo, proporcionan los argumentos necesarios para sugerir qué tipo de estrategia es más adecuada para distintas investigaciones de orden ecológico o evolutivo. Ya se ha indicado también que con el desarrollo de tecnologías NGS ha sido posible incrementar el número de secuencias por muestra a costos cada vez más bajos, por lo cual también es necesario considerarlas. En términos generales, obtener datos con una biblioteca *shotgun* constituye la aproximación más costosa (400,022 dólares), seguida por la construcción de bibliotecas de clones (465 dólares) y por último la de los T-RFLPs (25 dólares) (precios considerando la extracción de ADN, Tabla 3). En cuanto a NGS, el mejor precio que se encontró lo ofrece Langebio, en Irapuato, México. Por muestra se cobran ≈265 dólares, cifra menor que el precio estimado por biblioteca de clones (460 dólares); para este ejemplo en particular cada muestra generaría, en promedio, unas 2000 secuencias de 380 pb de largo con la plataforma GS Titanium. Este cálculo se obtuvo del presupuesto para la corrida de una placa, dividida en 16 regiones, y utilizando 12 marcas (*barcoding*) distintas en cada región, para un total de 192 muestras. Ésta y otras compañías generalmente no ofrecen el servicio por muestra, por lo que es necesario pagar la placa entera (≈3000 dólares). Si a un laboratorio le interesan 192 bibliotecas con estas características, es posible juntar todas las muestras en una sola placa o corrida (192 es sólo un ejemplo para obtener el mejor precio por muestra, aunque hay opciones para menor número de muestras), lo cual significa una inversión total de 50,943 dólares. La diferencia de precio entre secuenciar una sola muestra o varias (en este caso, 192) radica en el trabajo de preparación de las muestras individuales (*barcoding* y bibliotecas). Ya sea una sola muestra o múltiples (multiplex), con este tipo de tecnología (pirosecuenciación) se obtienen desde 92 000 hasta un millón de secuencias (divididas entre el número de muestras), sin embargo, la variación de precios es grande (Tabla 3). Por ejemplo el costo de la plataforma GS Titanium es de 5322 dólares en Langebio, y el precio por este mismo tipo de secuenciación en la Universidad de Tennessee es de 16,000 dólares. Con otras tecnologías NGS (Illumina, por ejemplo) el precio es menor por muestra: 3553 dólares para una biblioteca de 90 millones de fragmentos de 100 pb en UCDavis, y es aún menor utilizando Ion Proton, por ejemplo en Langebio, el precio por muestra de esta tecnología es de 2825 dólares por 20 millones de fragmentos de 200 pb.

Estas comparaciones son considerando únicamente escenarios en donde el objetivo sea analizar la diversidad de amplico-

nes específicos (e.g. 16S rDNA, ITS1), sin embargo, el potencial de estrategias como *shotgun* e incluso estas nuevas tecnologías de secuenciación incluye secuenciar genomas completos de organismos únicos o metagenomas (genomas de comunidades completas). Con ello, en ambos esquemas (*shotgun* o NGS), se genera información que permite hacer inferencias sobre el funcionamiento de la comunidad microbiana, además de conocer la composición taxonómica (Bonilla-Rosso *et al.*, 2008). Con este enfoque se han reportado un número cada vez mayor de estudios evolutivos que analizan la relación entre la diversidad taxonómica de los microorganismos con la función ecosistémica (Bonilla-Rosso *et al.*, 2012; Peimbert *et al.*, 2012).

Aunque puede ser obvio para este momento, cabe mencionar que la mayor demanda de trabajo al utilizar una técnica como la secuenciación genómica por *shotgun* (o métodos de NGS) se da en la parte de análisis informático de las extensas bases de datos generadas en el proceso de secuenciación, por lo que tendría que considerarse una inversión en equipos de cómputo en los costos finales. Muchos programas y plataformas de programas fueron desarrollados en los últimos años, disponibles de manera gratuita para los investigadores (por ejemplo, la plataforma QIIME, Caporaso *et al.*, 2010), pero aquellos análisis en los que se requiera hacer modificaciones de los *pipelines* (relacionados con el proceso de programación) requerirán de ciertas habilidades para programar que también es necesario considerar. De esta forma, la inversión del tiempo necesario para los análisis, así como los costos que ello genere y el *software* son factores importantes para tomar en cuenta (Yoccoz, 2012).

En resumen, los estudios de tipo experimental, en donde sean requeridas muchas réplicas o seguimiento temporal, así como el estudio de patrones de diversidad asociados a factores ambientales, podrían verse beneficiados por los bajos costos y la alta reproducibilidad de los T-RFLPs, considerando por supuesto la limitación de la asignación taxonómica (Erwin *et al.*, 2011; Fierer & Jackson, 2006; Reed & Martiny, 2013; Zhao *et al.*, 2012). En contraste, estudios que requieran una caracterización minuciosa de la diversidad y abundancia de grupos taxonómicos con precisión filogenética deberán apostar por una estrategia de bibliotecas de clones o pirosecuenciación con *barcoding* (Huse *et al.*, 2008), dependiendo del presupuesto con que se cuente. En cuanto a la estrategia por *shotgun*, ésta es ideal en estudios exploratorios, de descripción gruesa de diversidad y de bioprospección (de genes y funciones), en donde no se requiere el análisis de múltiples muestras o réplicas experimentales (Simon & Daniel, 2011).

AGRADECIMIENTOS

Esta investigación fue financiada por Semarnat/Conacyt 2002-C01-0246 y además constituye un requisito parcial del programa del Posgrado en Ciencias Biomédicas de la Universidad Nacional Autónoma de México (UNAM). Laura Espinosa-Asuar agradece

el apoyo de la beca proporcionada por el Consejo Nacional de Ciencia y Tecnología (Conacyt) no. 113 997, y de la UNAM. Agradecemos a la doctora Erika Aguirre Planter por su apoyo técnico, a la doctora Karla Heidelberg por facilitarnos la porción del filtro y su colaboración en la extracción del ADN, al biólogo Jaime Gasca Pineda por su asesoría en las filogenias y a la M. en Edición Concepción Asuar por su ayuda en la edición y corrección de estilo de este artículo.

REFERENCIAS

- ALLISON, S. D., Y. LU, C. WEIHE, M. L. GOULDEN, A. C. MARTINY, K. K. TRESEDER & J. B. MARTINY. 2013. Microbial abundance and composition influence litter decomposition response to environmental change. *Ecology* 94: 714-725.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. MYERS & D. J. LIPMAN. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 522-529.
- ASHELFORD, K. E., N. A. CHUZHANOVA, J. C. FRY, A. J. JONES & A. J. WEIGHTMAN. 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Applied Environmental Microbiology* 72: 5734-5741.
- BARTON, A. D., S. DUTKIEWICZ, G. FLIERL, J. BRAGG & J. F. FOLLOWS. 2010. Patterns of diversity in marine phytoplankton. *Science* 327: 1509-1511.
- BOGERT, B. VAN DEN, M. DE VOS WILLEM, E. G. ZOETENDAL & M. KLEEREBEZEM. 2011. Microarray analysis and barcoded pyrosequencing provide consistent microbial profiles depending on the source of human intestinal samples. *Applied and Environmental Microbiology* 77: 2071-2080.
- BONILLA-ROSSO, G., V. SOUZA & L. E. EGUIARTE. 2008. Metagenómica, genómica y ecología molecular: la nueva ecología en el bicentenario de Darwin. *TIP Revista Especializada en Ciencias Químico-Biológicas* 11: 41-51.
- BONILLA-ROSSO, G., M. PEIMBERT, L. D. ALCARAZ, I. HERNÁNDEZ, L. E. EGUIARTE, G. OLMEDO-ÁLVAREZ & V. SOUZA. 2012. Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments. *Astrobiology* 12: 659-673.
- BROWN, M. V. & S. P. DONACHIE. 2007. Evidence for tropical endemism in the Deltaproteobacteria Marine Group B/SAR324 bacterioplankton clade. *Aquatic Microbial Ecology* 46: 107-115.
- CAPORASO, J. G., J. KUCZYNSKI, J. STOMBAUGH, K. BITTINGER & F. D. BUSHMAN. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335-336.
- COLE, J. R., B. CHAI, T. L. MARSH, R. J. FARRIS, Q. WANG. 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research* 31: 442-443.

- CONABIO. 1988. *La diversidad biológica de México: Estudio de país*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México.
- CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad). 2013. Disponible en línea en: http://www.conabio.gob.mx/gap/index.php/Procesos_oceanograficos (consultado el 2 noviembre 2013).
- DESANTIS, T. Z., P. HUGENHOLTZ, N. LARSEN, M. ROJAS & E. L. BRODIE. 2006a. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* 72: 5069-5072.
- DESANTIS, T. Z., P. HUGENHOLTS, K. KELLER, E. L. BRODIE, N. LARSEN, Y. M. PICEÑO, R. PHAN & G. L. ANDERSEN. 2006b. Nast: a multiple sequence alignment server for comparative analysis of 16s rRNA genes. *Nucleic Acids Research* 34: W394-W399.
- DUPONT, C. L., D. B. RUSCH, S. YOOSEPH, M. J. LOMBARDO & R. A. RICHTER. 2011. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal* 6: 1186-1199.
- EDGAR, R. C., B. J. HAAS, J. C. CLEMENTE, C. QUINCE & R. KNIGHT. 2011. Uchime improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194-2200.
- EGUIARTE, L. E., J. A. AGUIRRE-LIGUORI, L. JARDÓN-BARBOLLA, E. AGUIRRE-PLANTER & V. SOUZA. 2013. Genómica de poblaciones: nada en evolución va a tener sentido si no es a la luz de la genómica, y nada en genómica tendrá sentido si no es a la luz de la evolución. *TIP Revista Especializada en Ciencias Químico-Biológicas* 16: 5-19.
- ERWIN, P. M., J. B. OLSON & R. W. THACKER. 2011. Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the Northern Gulf of Mexico. *Plos One* 6: 1-16.
- ESCALANTE, A. E., L. E. EGUIARTE, L. ESPINOSA-ASUAR, L. J. FORNEY, A. M. NOGUEZ & V. SOUZA SALDÍVAR. 2008. Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas Basin. *FEMS Microbiology Ecology* 65: 50-60.
- FALCÓN, L. I., A. M. NOGUEZ, L. ESPINOSA-ASUAR, L. E. EGUIARTE & V. SOUZA. 2008. Evidence of biogeography in surface ocean bacterioplankton assemblages. *Marine Genomics* 1: 55-61.
- FIERER, N. & R. B. JACKSON. 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* 103: 626-631.
- FIERER, N. & J. T. LENNON. 2011. The generation and maintenance of diversity in microbial communities. *American Journal of Botany* 98: 439-448.
- FRANK, J. A., C. I. REICH, S. SHARMA, J. S. WEISBAUM, B. A. WILSON & G. J. OLSEN. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and Environmental Microbiology* 74: 2461-2470.
- GIOVANNONI, S. J. & K. L. VERGIN. 2012. Seasonality in ocean microbial communities. *Science* 335: 671-676.
- GÓMEZ-PEREIRA, P. R., M. SCHÜLER, B. M. FUCHS, C. BENNKE & H. TEELING. 2012. Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environmental Microbiology* 14: 52-66.
- GREEN, J. L., B. J. M. BOHANNAN & R. J. WHITAKER. 2008. Microbial Biogeography: from taxonomy to traits. *Science* 320: 1039-1043.
- GUINDON, S., J. F. DUFAYARD, V. LEFORT, M. ANISIMOVA, W. HORDIJK & O. GASCUEL. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307-321.
- HALL, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series* 41: 95-98.
- HAO, X. & T. CHEN. 2012. OTU analysis using metagenomic shotgun sequencing data. *PLoS ONE* 7: e49785.
- HUANG, Y., P. GILNA & W. LI. 2009. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 25: 1338-1340.
- HUBER, T., G. FAULKNER & P. HUGENHOLTZ. 2004. Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20: 2317-2319.
- HUGHES, J. B., B. J. BOHANNAN, J. H. BROWN, R. K. COLWELL, J. A. FUHRMAN & J. L. GREEN. 2006. Microbial biogeography: putting microorganisms on the map. *Natural Review of Microbiology* 4: 102-112.
- HUSE, S. M., D. LES, J. A. HUBER, D. M. WELCH, D. A. RELMAN & M. L. SOGIN. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLOS Genetics* 4: 1-10.
- IVERSON, V., R. M. MORRIS, C. D. FRAZAR, C. T. BERTHIAUME, R. L. MORALES & E. V. ARMBRUST. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335: 587-590.
- J. CRAIG VENTER INSTITUTE (disponible en línea en): <https://moore.jcvi.org/moore/SingleOrganism.do?speciesTag=VSWAT3> (consultado 2 noviembre 2013).
- JIN, H. M., J. M. KIM, H. J. LEE, E. L. MADSEN & C. O. JEON. 2012. Alteromonas as a key agent of polycyclic aromatic hydrocarbon biodegradation in crude oil-contaminated coastal sediment. *Environmental Scientific Technology* 46: 7731-7740.
- KEANE, T. M., C. J. CREEVEY, M. M. PENTONY, T. J. NAUGHTON & J. O. MCINERNEY. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolution Biology* 6: 29.
- LANGE BIO (disponible en línea en): <http://www.Langebio.cinvestav.mx/lab-sergen/> (consultado 2 noviembre 2013).
- LAVIN, P., P. GÓMEZ, B. GONZÁLEZ & O. ULLOA. 2008. Diversity of the marine picocyanobacteria Prochlorococcus and Synechococcus assessed

- by terminal restriction fragment length polymorphisms of 16S-23S rRNA internal transcribed spacer sequences. *Revista Chilena de Historia Natural* 81: 515-531.
- LINDSTROM E. S. & S. LANGENHEDER. 2012. Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports* 4: 1-9.
- MADIGAN, M. T., J. M. MARTINKO & J. PARKER. 2003. Brock Biology of Microorganisms. Prentice-Hall: New Jersey, decimal edición.
- MCDONALD, D., M. N. PRICE, J. GOODRICH, E. P. NAWROCKI & T. Z. DESANTIS. 2012. An improved GreenGenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal* 6: 610-618.
- NAZARIES, L., Y. PAN, L. BODROSSY, E. M. BAGGS, P. MILLARD, J. C. MURRELL & B. K. SINGH. 2013. Microbial regulation of biogeochemical cycles: evidence from a study on methane flux and land-use change. *Applied and Environmental Microbiology* 79: 4031-4040.
- NOGUEZ, A. M., H. T. ARITA, A. E. ESCALANTE, L. J. FORNEY, F. GARCÍA-OLIVA & V. SOUZA. 2005. Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Global Ecology Biogeography* 14: 241-248.
- NOGUEZ, A. M., A. E. ESCALANTE, L. J. FORNEY, M. NAVA-MENDOZA, I. ROSAS, V. SOUZA & F. GARCÍA-OLIVA. 2008. Soil aggregates in a tropical deciduous forest: effects on C and N dynamics, and microbial communities as determined by t-RFLPs. *Biogeochemistry* 89: 209-220.
- OSBORN, A. M., E. R. MOORE & K. N. TIMMIS. 2000. An evaluation of terminal-restriction fragment length polymorphism. T-RFLP analysis for the study of microbial community structure and dynamics. *Environmental Microbiology* 2: 39-50.
- PEDRÓS-ALIÓ, C. 2012. The rare bacterial biosphere. *Annual Review of Marine Science* 4: 449-466.
- PEIMBERT, M., L. D. ALCARAZ, G. BONILLA-ROSSO, G. OLMEDO-ÁLVAREZ, F. GARCÍA-OLIVA, L. SEGOVIA, L. E. EGUIARTE & V. SOUZA. 2012. Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin I: Ancient lessons on how to cope with an environment under severe nutrient stress. *Astrobiology* 12: 648-657.
- PRUESSE, E., C. QUAST, K. KNITTEL, B. FUCHS, W. LUDWIG, J. PEPLIES & F. O. GLÖCKNER. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35: 7188-7196.
- REED, H. E. & J. B. MARTINY. 2013. Microbial composition affects the functioning of estuarine sediments. *The ISME Journal* 7: 868-879.
- RUSCH, D. B., A. L. HALPERN, G. SUTTON, K. B. HEIDELBERG & S. WILLIAMSON. 2007. The sorcerer II global ocean sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* 5: e77.
- RZEDOWSKI, J. 2006. *Vegetación de México*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad: México. Primera edición digital, 504 págs.
- SEGATA, N., D. BOERNIGEN, T. L. TICKLE, X. C. MORGAN, W. S. GARRETT & C. HUTTENHOWER. 2013. Computational meta'omics for microbial community studies. *Molecular Systematic Biology* 9: 666.
- SIMON, C. & R. DANIEL. 2011. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology* 77: 1153-1161.
- SKOVHUS, T. L., C. HOLMSTRÖM, S. KJELLEBERG & I. DAHLÖF. 2007. Molecular investigation of the distribution, abundance and diversity of the genus *Pseudoalteromonas* in marine samples. *FEMS Microbiology Ecology* 61: 348-61.
- STACKBRANDT, E. & J. EBERS. 2006. Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today* 33: 152-155.
- SUENAGA, H. 2012. Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology* 14: 13-22.
- SWOFFORD, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA (disponible en línea en): [http://www.tsu.edu/PDFFiles/CBER/Miranda/PAUP %20Manual.pdf](http://www.tsu.edu/PDFFiles/CBER/Miranda/PAUP%20Manual.pdf) (consultado 2 noviembre 2013).
- TAMURA, K., D. PETERSON, N. PETERSON, G. STECHER, M. NEI & S. KUMA. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.
- THOMAS, T., J. GILBERT & F. MEYER. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics Experimentation* 2: 3.
- TURNER, S., K. M. PRYER, V. P. W. MIAO, & J. D. PALMER. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryotic Microbiology* 46: 327-338.
- UCDAVIS GENOME CENTER (disponible en línea en): <http://dnatech.genome-center.ucdavis.edu/prices.html> (consultado 2 noviembre 2013).
- UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO (Unidad de Síntesis y Secuenciación, Instituto de Biotecnología) (disponible en línea en): <http://www.ibt.unam.mx/sintesis/secuenciacion.html> (consultado 2 noviembre 2013).
- UNIVERSITY OF TENNESSEE (Center for Environmental Biotechnology) (disponible en línea en): <http://www.ceb.utk.edu/dnasequence.html> (consultado 2 noviembre 2013).
- UNIVERSITY OF VIRGINIA MEDICAL SCHOOL (Center for Public Health Genomics) (disponible en línea en): <http://cphg.virginia.edu/mackey/projects/sequencing-pipelines/sweetseq/next-gen-sequencing-costs-at-ua/> (consultado 2 noviembre 2013).
- WILLIAMSON, S. J. & S. YOUSEPH. 2012. From bacterial to microbial ecosystems (metagenomics). In Helden J van, A. Toussaint & D. Thieffry (eds.) *Bacterial Molecular Networks: Methods and Protocols. Series Methods in Molecular Biology* 804: 35-55.

- WOODCOCK, S., C. J. VAN DER GAST, T. BELL, M. LUNN, T.P. CURTIS, I. M. HEAD & W. T. SLOAN. 2007. Neutral assembly of bacterial communities. *FEMS Microbiology Ecology* 62: 171-180.
- WRIGHT, E. S., L. S. YILMAZ & D. R. NOGUERA. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Applied and Environmental Microbiology* 78: 717-725.
- YILMAZ, P., M. H. IVERSEN, W. HANKELN, R. KOTTMANN, C. QUAST & F. O. GLÖCKNER. 2012. Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiology Ecology* 81: 373-385.
- YOCOZ, N. G. 2012. The future of environmental DNA in ecology. *Molecular Ecology* 21(8) (Special Issue environmental DNA): 2031-2038.
- ZARRAONAINDIA, I., D. P. SMITH & J. A. GILBERT. 2013. Beyond the genome: community-level analysis of the microbial world. *Biology and Philosophy* 28: 261-282.
- ZHAO, D., R. HUANG, J. ZENG, W. YAN, J. WANG, T. MA, M. WANG & Q. L. WU. 2012. Diversity analysis of bacterial community compositions in sediments of urban lakes by terminal restriction fragment length polymorphism (T-RFLP). *World Journal of Microbiology and Biotechnology* 28: 3159-3170.
- ZHENG, Q., R. ZHANG, M. KOBLÍŽEK, E. N. BOLDAREVA, V. YURKOV, S. YAN & N. JIAO. 2011. Diverse arrangement of photosynthetic gene clusters in aerobic anoxygenic phototrophic bacteria. *PLoS One* 6: e25050.

Recibido: 12 de noviembre del 2013.

Aceptado: 31 de marzo de 2014.