

# METAGENÓMICA, GENÓMICA Y ECOLOGÍA MOLECULAR: LA NUEVA ECOLOGÍA EN EL BICENTENARIO DE DARWIN

Germán Bonilla-Rosso, Valeria Souza y Luis E. Eguarte<sup>a</sup>

*Depto. de Ecología Evolutiva, Instituto de Ecología, UNAM.  
Ciudad Universitaria, Apdo. Postal 70-275, México, D.F., 04510,  
México. <sup>a</sup>Autor de correspondencia. E-mail: [fruns@servidor.unam.mx](mailto:fruns@servidor.unam.mx)*

## RESUMEN

En años recientes hemos presenciado un avance enorme en el desarrollo de tecnologías moleculares, en particular en las relacionadas con la secuenciación del ADN. El impacto de la genómica y metagenómica es de tal magnitud que revolucionarán las ciencias ecológicas, dando origen, junto con la Ecología Molecular, a una Nueva Ecología, con preguntas y metodologías diferentes a las abordadas por la ecología clásica. Estas nuevas metodologías abren la puerta a la exploración de nuevas aproximaciones al estudio de las comunidades microbianas, pues permiten el análisis simultáneo de la caracterización taxonómica de las especies contenidas en la comunidad y las funciones que pueden desempeñar. Aquí revisamos algunos de los avances de esta Nueva Ecología, describiendo el potencial que puede alcanzar si se usa junto con otras disciplinas. En general, nos permite analizar comunidades microbianas naturales, incluyendo organismos que escapan a nuestras posibilidades de cultivo en el laboratorio. Asimismo, brinda información sobre la diversidad y estructura trófica de las comunidades al tiempo que permite ligarlas con las funciones ecológicas que llevan a cabo en el ciclo de nutrientes del ecosistema que las contiene. Éste representa un salto enorme para resolver cómo es que cambian los organismos a lo largo del tiempo en respuesta a su entorno ecológico, mientras que permite analizar a una escala más fina cómo es que el entorno ecológico afecta los patrones evolutivos dentro de los linajes filogenéticos contenidos en las comunidades. Finalmente, esbozamos nuestra perspectiva sobre el futuro de la Nueva Ecología.

**Palabras Clave:** ADN, Bacterias, Bioinformática, Comunidades Microbianas, Cuatrociénegas, Diversidad, Ecología Bacteriana, Ecología de Ecosistemas, Filogenia, Microbiología Ambiental.

## ABSTRACT

In recent years, we have witnessed an incredible advance in the development of molecular techniques, in particular new DNA sequencing technologies. Genomics and Metagenomics along with Molecular Ecology will revolutionize ecological sciences, giving birth to a New Ecology, with new questions and approaches to those studied by classical ecology. These new methodologies open the possibilities of new approaches to the study of microbial communities, allowing the simultaneous analysis of taxonomical and functional characterization of the community. In this review we discuss some of the advances of this New Ecology, describing its explanatory potential when used along with other disciplines. In general, it allows us to analyze natural microbial communities, including unculturable organisms. In addition, it produces information on the diversity, trophic structure of the community and functioning and nutrient cycling in the ecosystem. This represents a huge leap towards the understanding of organism evolution in its environment, and how this environment shapes evolutive patterns within phylogenetic lineages. Finally, we present our perspectives on the future development of this nascent discipline.

**Key Words:** DNA, Bacteria, Bioinformatics, Microbial Communities, Cuatrociénegas, Diversity, Bacterial Ecology, Ecosystem Ecology, Phylogeny, Environmental Microbiology.

## INTRODUCCIÓN

**H**ace 10 años, en el primer número de esta revista describimos lo que llamamos el “sueño Darwiniano”<sup>1</sup>, esto es, cómo con la ayuda de los marcadores moleculares podemos describir las relaciones genealógicas (filogenéticas) entre todos los seres vivos, o sea, reconstruir el gran árbol de la vida que tanto interesaba a Charles Darwin. En los últimos 10 años estas ideas han avanzado notablemente, pero aún nuestras predicciones más optimistas estaban lejos de concebir los espectaculares avances recientes en genómica<sup>2</sup>. Además de poder obtener con relativa facilidad los genomas completos de todo tipo de organismos, desde virus y bacterias (ver Alcaraz *et al.*<sup>3</sup> para el estudio más reciente publicado por mexicanos) hasta árboles y ornitorrincos<sup>4,5</sup>, se ha comenzado a analizar el genoma de TODOS los organismos de una comunidad dada. Estos estudios se conocen como de metagenómica, y representan un avance cuántico en nuestro entendimiento de cómo funciona la naturaleza. Estos datos nos permiten no sólo conocer a todos los organismos de las comunidades, especialmente los microscópicos, sino también todas sus funciones. Esta información promete abrir la “caja negra” que representan estos compartimientos en los ecosistemas y entender, por fin, la verdadera ecología de nuestro planeta. Estas herramientas, junto con la Ecología Molecular clásica que describimos en detalle en nuestro libro publicado recientemente<sup>6</sup> constituyen lo que llamamos la Nueva Ecología, que promete ser una de las ramas más importantes de la Ciencia en este siglo. En este artículo queremos revisar algunos de los avances más interesantes en esta Nueva Ecología, en particular los relacionados a metagenómica ecológica, para que sirvan como una de las primeras guías a esta rama del conocimiento en español.

La ecología es la ciencia que estudia los factores que determinan la distribución y abundancia de los seres vivos. Tradicionalmente su estudio se divide en cuatro niveles de complejidad: La Autoecología o Ecofisiología, que estudia cómo responden los individuos de una especie frente a cambios en su ambiente físico y cómo lo modifican, es decir, estudian la fisiología de los organismos en su entorno abiótico, la Ecología de Poblaciones estudia cómo se comportan los grupos de organismos de una misma especie en un mismo tiempo y espacio, la Ecología de Comunidades que trata de entender cómo se estructuran e interaccionan las poblaciones de diferentes especies dentro de una comunidad y la Ecología de Ecosistemas, que estudia los flujos de materia y energía entre una comunidad y su entorno abiótico.

El análisis ecológico se ve frecuentemente entorpecido por la complejidad intrínseca de los problemas que estudia. En particular, los tamaños muestrales necesarios para realizar comparaciones estadísticas entre poblaciones y comunidades pueden representar un reto formidable, como lo sabe cualquiera que haya ayudado en un estudio de este tipo y haya tenido que

contar y medir miles de muestras. Así podemos mencionar la complejidad en número y abundancia de especies que conforman una comunidad (por ejemplo en una selva tropical) y el problema que puede representar el sólo hecho de identificar muchas especies muy parecidas coexistiendo, la ausencia de bordes naturales que delimiten las comunidades y los ecosistemas, la extensa escala temporal a la que suceden cambios mesurables en poblaciones, comunidades y ecosistemas, y la magnitud espacial de los fenómenos ecológicos en los ecosistemas (por ejemplo, los ciclos biogeoquímicos y transferencia de energía entre niveles tróficos que a veces implican toda o amplias extensiones de la biosfera).

A pesar de que la importancia de los microorganismos procariontes (arqueas y bacterias) fue reconocida desde mediados del siglo pasado cuando Kluver y van Niel<sup>7</sup> estimaron que 95% del CO<sub>2</sub> disponible en la atmósfera era producido por la mineralización bacteriana, la aproximación ecológica a los microorganismos se ha mantenido tradicionalmente enfocada a la fisiología bacteriana con la microbiología clásica. Actualmente se reconoce que el dominio con mayor biomasa (aunque mucha gente insiste en que son los insectos, o los hongos, según con lo que trabajen) y mayor diversidad taxonómica y funcional es del dominio Bacteria.

Asimismo, han sido las bacterias las causantes del cambio atmosférico más importante (hasta ahora) en el planeta Tierra: la producción de oxígeno diatómico y con ello la conversión de una atmósfera reductora a una oxidante. Hoy en día, sólo los organismos procariontes son capaces de realizar la fijación de nitrógeno atmosférico, mineralizar nitrógeno hasta nitrato y respirar sulfatos o hierro, por mencionar algunos de los múltiples pasos de los ciclos biogeoquímicos globales.

Sin embargo, el estudio de los microorganismos, dado su pequeño tamaño y sus tasas generacionales muy variables (pero que pueden ser muy rápidas), implica una problemática natural. Actualmente, en general se acepta que se conoce tan sólo el 1% de la diversidad total de bacterias en el planeta<sup>8</sup>, y que en un gramo de suelo existen 10 mil millones de células bacterianas que pertenecen, al menos, a miles (según algunos autores millones) de especies<sup>9</sup>. Dado que las escalas espaciales y temporales de los organismos procariontes presentan una oportunidad única para el estudio de fenómenos ecológicos, el presente artículo pretende revisar los estudios recientes en ecología microbiana con un enfoque molecular alcanzado gracias al desarrollo de nuevas tecnologías.

La revolución genómica actual (entendida como el desarrollo de tecnologías de secuenciación de genomas completos) le ha dado un nuevo sentido al análisis de la ecología bacteriana, ya que al conocer el total de los genes contenidos dentro del genoma de una especie es posible resolver incógnitas fisiológicas y

ecológicas de ciertos organismos e inferir las necesidades de crecimiento de otros nuevos.

Por ejemplo, *Geobacter sulfurreducens* es una delta proteobacteria que se había identificado por la amplificación de su 16SrARN en suelos contaminados con material radiactivo y que podía cultivarse con dificultad en medios específicos. Tras la secuenciación de su genoma completo, resultó posible el diseño de medios de crecimiento específicos al descubrir su aerobiosis facultativa y su posibilidad de reducir sulfato como fuente de energía, así como su capacidad para producir electricidad<sup>10,11</sup>. Asimismo, el genoma del planctomicete *Kuenenia stuttgartensis* fue obtenido a partir de la secuenciación de ADN ambiental de lodos activados de una planta de tratamiento antes de poder cultivarlo, y su genoma reveló adaptaciones metabólicas relacionadas con la oxidación anaeróbica del amonio, proceso descrito bioquímicamente pero del cual se desconocía hasta entonces qué proteínas podrían realizarlo<sup>12</sup>.

En la Nueva Ecología se utilizan una serie de métodos moleculares y estadísticos para mejorar nuestro entendimiento en los cuatro niveles clásicos de la Ecología (Tabla I). La Nueva Ecología nos ayuda por ejemplo a describir la diversidad total, con énfasis en los organismos microscópicos (arqueas, bacterias, “protistas” y hongos microscópicos), junto con métodos genéticos y genómico/proteómicos para describir la funciones y las interacciones, en una perspectiva claramente filogenética y evolutiva.

### COMUNIDADES BACTERIANAS, GENÉTICA Y METAGENÓMICA

Se puede sugerir que la ecología de comunidades microbianas moderna surgió cuando Pace *et al.*<sup>13,14</sup> amplificaron y secuenciaron el gen de ARN ribosomal 16S (utilizado como marcador taxonómico universal en arqueas y bacterias) del ADN extraído

directamente de comunidades microbianas naturales (ADN ambiental: se toma una muestra de suelo o agua y se extrae TODO el ADN de la muestra, y se amplifica con métodos estándar de PCR<sup>15</sup>), encontrando una gran diversidad de organismos pertenecientes a grupos taxonómicos no descritos anteriormente a nivel de *phylum*. Antes de este estudio, la ecología microbiana se basaba en describir funciones (a veces enriqueciendo el ambiente) utilizando un enfoque de “caja negra” (o sea, se sabe lo que entra y lo que sale, pero no se sabe ni cómo se realiza el proceso ni quién lo lleva a cabo), o de tratar de cultivar las bacterias, un proceso muy ineficiente, ya que sólo una proporción muy baja de ellas son cultivables.

Desde entonces se ha descrito la composición taxonómica microbiana de una gran cantidad de ambientes distintos con una diversidad impresionante de grupos taxonómicos no descritos, en el extremo de los cuales se encuentra la definición del Reino Korarchaeota dentro del dominio Archaea, que no posee ningún representante cultivable<sup>16</sup>. Aunque las primeras filogenias universales de bacterias son sumamente recientes y reconocían alrededor de una decena de phyla<sup>17</sup>, actualmente se reconocen entre 40 y 82 phyla (dependiendo del autor) sin representantes cultivables y, por lo tanto, sin conocimiento sobre su biología y ecología<sup>18,19</sup>.

La metagenómica representa una aproximación totalmente nueva al estudio de las comunidades microbianas, definida como el análisis funcional y de secuencias de los genomas microbianos colectivos contenidos en una muestra ambiental, basándose ya sea en expresión o secuenciación<sup>20</sup>. El primer trabajo con datos tipo metagenoma, aunque parciales, fue el de Rondon *et al.*<sup>21</sup>, donde se describe una librería de clonas en BACs (Bacterial Artificial Chromosomes) que contenía fragmentos de ADN relativamente grandes obtenidos directamente de muestras de suelos agrícolas de la Estación de Investigación Agrícola de

| Nivel                          | Métodos tradicionales   | Nueva Ecología   |
|--------------------------------|---|--|
| Ecofisiología/<br>Autoecología | Curvas de tolerancia<br>Medidas en campo y laboratorio  | Genómica: Genes-procesos.<br>Expresión Genes: ARN<br>Proteómica y microchips.  |
| Poblaciones                    | Estimación tamaño de población: Captura/recaptura,<br>crecimiento.<br>Estructura/demografía.<br>Mecanismos de regulación. | Marcadores moleculares para análisis de paternidad.<br>Estimaciones de migración. Estimación de parámetros<br>históricos con coalescencia y filogeografía. |
| Comunidades                    | Listados de especies.<br>Número de individuos por especie.  | ADN ambiental.<br>Librerías de clonas de 16 y 18S.   |
| Ecosistemas                    | Análisis de nutrientes<br>Análisis de tasas y de entradas de energía, recursos.<br>Ciclos biogeoquímicos                  | Metagenomas<br>Metaproteomas<br>Genomas de especies clave  |

Tabla I. Los niveles clásicos de estudio y complejidad de la Ecología, los métodos como se han enfrentado tradicionalmente y las propuestas de la Nueva Ecología.

West Madison, Wisconsin. Cada una de las clonas fue sometida a diversos análisis diferenciales para la identificación de nuevas proteínas involucradas en actividades hemolíticas, lipolíticas y amilolíticas, demostrando así su potencial en el descubrimiento de nuevos genes.

Recientemente, el desarrollo y abaratamiento de tecnologías de secuenciación como el shotgun o la pirosecuenciación (Cuadro 1) ha brindado la posibilidad de análisis metagenómicos amplios. Las principales ventajas de estos métodos sobre las técnicas tradicionales de amplificación genética son: 1) un menor sesgo taxonómico en la secuenciación, y 2) la capacidad de analizar

Las tecnologías actuales se enfocan en la secuenciación no de genes u operones individuales, sino de grandes fragmentos de ADN como plásmidos o genomas completos. Todas involucran un paso previo en el cual el ADN total es fragmentado en pedazos más cortos y manejables, y por lo tanto todas involucran un proceso computacional posterior a la secuenciación que intenta reensamblar estos fragmentos en secuencias lineales individuales.

**Shotgun:** Rompe el ADN total en fragmentos y los separa por tamaños (3kb, 8kb y 40kb). Los fragmentos pequeños son insertados en cromosomas bacterianos artificiales (BAC) y los grandes en fósidos, para luego ser clonados en *Escherichia coli* para amplificarlos. Después los fragmentos son extraídos y purificados para ser secuenciados. Si las secuencias de los BACs no pueden ser ensambladas, se buscan en los fósidos para secuenciación completa. Este método es caro y lleva mucho tiempo construir y purificar las librerías, y es posible que regiones enteras no puedan ser clonables por contener genes letales. Sin embargo, produce fragmentos de 600-800 bp de longitud con una precisión de 99.97%. La cantidad de producto depende del número de clonas que se elija secuenciar. La secuenciación de un genoma menor a 5Mb, desde su preparación hasta la recuperación de datos, puede tomar un mes. La cobertura total es aproximadamente un 30x.

**Pirosecuenciación:** Nebuliza el ADN en fragmentos de 900bp. Toma una sola hebra de ADN de cada fragmento y añade "adaptadores" a cada extremo de la secuencia (un primer de 44 bp). Estos adaptadores fijan la secuencia a una superficie sólida donde duplica cada fragmento a partir del primer adaptador. Un lector óptico lee los fotones producidos por la hidrólisis de cada uno de los nucleótidos añadidos, generando la secuencia de ADN. El método prescinde de la construcción de librería de clonas, reduciendo drásticamente el precio y el tiempo de secuenciación. Un genoma menor a 5Mb, desde su preparación hasta la recuperación de datos, toma alrededor de 3 días con una cobertura aproximada de 5-8x (cada corrida de 4hrs produce 5 millones de bases). Sin embargo, la longitud promedio de las secuencias producidas es de 100-200 bp.

**Cuadro 1. Metodologías de Secuenciación a Gran Escala.**

simultáneamente el complemento taxonómico y funcional de una misma comunidad. Lo anterior permite que el desarrollo tecnológico sea recibido por el marco teórico robusto de la ecología de comunidades tradicional, por ejemplo, permitiendo implementar los análisis tradicionales utilizados por ecólogos vegetales y animales para evaluar la riqueza de especies, la estructura de comunidades y las diversidades alfa, beta y gamma<sup>15,22</sup>; a la par que es posible analizar el reconocimiento de gremios funcionales, inclusive con una mayor precisión y resolución que la permitida por los estudios macroecológicos tradicionales, pues la presencia de un gen permite medir el potencial taxonómico de la comunidad incluso en condiciones donde dicho gen no se encuentre en uso.

### UN EJEMPLO: EL ESTUDIO DEL MAR DE LOS SARGAZOS

Uno de los primeros trabajos publicados de metagenómica resulta de la secuenciación de la comunidad microbiana de aguas superficiales oceánicas del Mar de los Sargazos (Bermudas) realizado por Venter *et al.*<sup>23</sup> utilizando el ADN de 100 litros de agua en cada una de las 10 estaciones de muestreo, estas muestras se separan con tres filtros de diferentes tamaños (0.1-0.8 µm; 0.8-3 µm y 3-20 µm). Se analizaron la fracciones entre 0.8 y 3 micras de tamaño, que no incluyen ni las cosas relativamente grandes (como eucariontes planctónicos y cianobacterias filamentosas), ni las muy chicas (bacterias muy chicas a veces llamadas picoplancton y virus). Dado que el Mar de los Sargazos es un océano oligotrófico (pobre en nutrientes) se esperaba que la comunidad bacteriana que mantiene sería relativamente "simple", pero el metagenoma resultó en una cantidad aproximada de 1800 genomas diferentes y más de 1.2 millones de genes codificantes. La reconstrucción de genomas completos fue totalmente infructuosa por múltiples causas: i) La increíble diversidad genómica encontrada reduce la posibilidad de encontrar fragmentos correspondientes a la misma especie (o población u organismo), impidiendo ensamblar fragmentos razonablemente grandes de genoma bajo una cobertura estadísticamente confiable; ii) La abundancia natural relativa y diferencial de cada especie en la muestra la sesga hacia una mayor cobertura de las pocas especies más abundantes; iii) La variabilidad intrapoblacional (es decir, de los miembros de una misma especie) complica el ensamble, incrementando la probabilidad de ensamblar fragmentos que no se encuentran juntos en la naturaleza.

Los segmentos ensamblados con mayor cobertura (hasta 14x, lo que significa que el segmento se encuentra confirmado por el ensamble de al menos 14 fragmentos, ver Cuadro 2) pudieron ser exitosamente alineados con segmentos del genoma secuenciado de la cianobacteria unicelular *Prochlorococcus marinus* MED4, aislado originalmente de esta región oceánica. Sin embargo, los segmentos no pudieron ser ensamblados entre sí, y ninguno presenta una identidad del 100% con la cepa secuenciada, sugiriendo la presencia de una población altamente diversa y abundante de *P. marinus* con diversas variantes genómicas.

Cualquiera que sea la metodología de secuenciación, el proceso siguiente es el ensamble de las secuencias obtenidas. El ensamble se basa en el solapamiento de diferentes fragmentos de ADN que permitirán alinear los extremos similares y así reconstruir secuencias lineales completas conocidas como contigs y, con un poco de suerte, la secuencia lineal completa del genoma. El problema radica en que muchas veces los fragmentos secuenciados no cubren la totalidad del genoma (existen contigs que no pueden ensamblarse con ningún otro fragmento porque las regiones que los unirían no fueron secuenciadas) o existen regiones muy similares repetidas varias veces a lo largo del genoma, de manera que no es posible saber cuál contig se ensambla con cualquier otro. Obviamente, si los fragmentos producidos por la secuenciación son más grandes es más fácil ensamblar que si los fragmentos son más pequeños. Sin embargo, para saber con precisión si una secuencia es en realidad la secuencia del genoma y no un artefacto de la secuenciación, es necesario que cada base en cada sitio a lo largo de la secuencia lineal se encuentre repetida varias veces, es decir, que cada base en cada fragmento haya sido secuenciada varias veces y al alinear todos los diferentes fragmentos secuenciados de una misma región sean coherentes entre sí. A esto se le conoce como profundidad de la cobertura de secuenciación, y un genoma completo se considera confiable si tiene más de 6x. Esto quiere decir que la totalidad de bases secuenciadas es 6 veces el tamaño del genoma inferido, y por lo tanto se esperaría que cada base del genoma completo estuviera representado al menos en seis fragmentos independientes. Cabe señalar que el esfuerzo computacional que requiere este análisis es enorme y, aunque hemos observado enormes avances, el ensamblaje automático dista mucho de ser perfecto y frecuentemente requiere la asistencia humana.

Cuadro 2. Ensamble y Cobertura.

Uno de los descubrimientos más importantes producido por el análisis del metagenoma del Mar de los Sargazos es la revisión de la fotoautotofía (la capacidad de sintetizar nutrientes a partir de componentes inorgánicos utilizando la luz solar como fuente de energía) en comunidades marinas bacterianas, pues la enzima principal involucrada en el ciclo de Calvin (RubisCO) se encuentra en una muy baja abundancia y diversidad en las muestras, lo que podía explicarse por la sobredominancia del fotoautótrofo *Prochlorococcus* en la muestra. Sin embargo, una gran diversidad de proteorhodopsinas fue encontrada, con al menos 782 homólogos. Las proteorhodopsinas son una familia de rhodopsinas que son capaces de bombear protones en contra de un gradiente electroquímico transmembranal, inicialmente descritas en una alfaproteobacteria del clado SAR86<sup>24,25</sup>, cuya variante en tan sólo una sustitución nucleotídica que se traduce en la optimización en la absorbancia espectral de cada variante a diferente profundidad. La presencia de tal número de homólogos sugiere un gremio diverso y abundante de fotoautótrofos diferencialmente adaptado a nichos

marinos para optimizar la absorción luminosa.

## ECOLOGÍA DE ECOSISTEMAS, LA MINA DE HIERRO (AMD) Y LOS BIORREACTORES

El análisis funcional de comunidades permite realizar extrapolaciones respecto a la interacción entre el componente biótico y abiótico de un ecosistema. Por ejemplo, la presencia de una gran diversidad de proteorhodopsinas en el metagenoma del Mar de los Sargazos sugiere revisar las estimaciones de productividad primaria en los océanos y con ello los cálculos correspondientes al ciclo global de carbono. Sin embargo, dada la complejidad y versatilidad funcional de las comunidades marinas, las interpretaciones funcionales parecen tener más sentido cuando se analizan sistemas más simples en conjunto con estudios geoquímicos y bioquímicos.

El primer metagenoma “real” secuenciado fue publicado un poco antes que el estudio del Mar de los Sargazos por Tyson *et al.* también en el 2004. Los trabajos anteriores corresponden a metagenomas “dirigidos” o parciales como el que mencionamos del suelo de Wisconsin<sup>21</sup>. Tyson *et al.* analizaron una biopelícula simple que crece sobre el agua acidificada por escurrimiento dentro de una mina de hierro abandonada (AMD: Acid Mine Drainage). La simplicidad de la comunidad debido a la dominancia de dos organismos permitió la reconstrucción con una cobertura 10x de los genomas de una bacteria del grupo II del género *Leptospirillum* (Nitrospiraceae) y de una arquea menos abundante del género *Ferroplasma* tipo II (Euryarchaeota), así como el ensamble parcial de fragmentos del genoma de otras especies de *Leptospirillum* (grupo III,) y *Ferroplasma* (tipo III), mucho menos abundantes pero con alta transferencia horizontal, así como la presencia de otras tres especies procariontes con densidades tan bajas que no fue posible ensamblar sus fragmentos.

La cobertura de las dos especies dominantes (aproximadamente 10x cada una) permitió elucidar las rutas metabólicas involucradas en la oxidación de la pirita como fuente única de energía que permite la manutención de la comunidad. La continuidad de este trabajo mediante una aproximación proteómica<sup>26</sup> mediante espectrometría de masas por shotgun permitió analizar una fracción de las proteínas parcialmente degradadas en la mina, resultando en la identificación de la principal enzima oxidante de hierro en la comunidad: una secuencia nueva ácido-tolerante del *Leptospirillum* grupo II denominada Cyt579, demostrando la capacidad del enfoque combinado metagenómico-proteómico para develar nuevos procesos ecológicos.

Otra aplicación radicalmente diferente de las herramientas metagenómicas reside en el análisis de los biorreactores, como los usados en la eliminación de fósforo de aguas contaminadas (BPR), en los cuales se utilizan lodos con organismos que precipitan fosfatos en un tanque y después se les permite asentarse para su eliminación, dejando el cuerpo de agua

virtualmente libre de fosfatos. Sin embargo, la ignorancia respecto a la funcionalidad biológica de la comunidad BPR ocasiona reducción espontánea en la capacidad de eliminación del biorreactor. La secuenciación del metagenoma del lodo enriquecido<sup>27</sup> permitió la reconstrucción parcial del genoma de la especie dominante en el BPR: *Accumulibacter phosphatis*, un miembro del grupo Rhodocyclales (Betaproteobacteria) que en condiciones anaeróbicas acumula acetato y propionato, utilizando el ATP de sus reservas de polifosfato y glicógeno. Al cambiar a condiciones aeróbicas, el oxígeno disponible le permite respirar las reservas de acetato y propionato y reabastecer sus reservas de polifosfato tomando fósforo inorgánico (Pi) del medio (a través de su membrana plasmática por medio de un juego de transportadores de baja afinidad), eliminando el fósforo del agua. También cuenta con un juego de transportadores de alta afinidad que se espera se expresen sólo hacia el final de la fase aeróbica. De esta manera, fue posible hipotetizar que las condiciones de anaerobiosis o de microanaerobiosis ocasionarán una disminución en la precipitación de fosfatos por las comunidades BPR.

### BIOGEOGRAFÍA BACTERIANA Y FUNCIONALIDAD: EL GLOBAL OCEAN SAMPLING GOS

El proyecto biológico más ambicioso en lo que va del siglo XXI es sin lugar a dudas el Global Ocean Sampling (GOS), llevado a cabo por el J.C. Venter Institute<sup>28</sup>. En éste, se planteó muestrear 200 litros de agua cada 300 millas náuticas alrededor del mundo y secuenciar el metagenoma de la fracción de la comunidad cuyo tamaño estuviese comprendido entre los 0.8 y 1 µm. Este trabajo es comparable con los viajes de colecta de los naturalistas del siglo XIX, como el mítico viaje de Charles Darwin en el Beagle, o las exploraciones botánicas de Alexander Humboldt y Aimé Bonpland en el Continente Americano en cuanto a la cantidad de nueva información obtenida sobre nuevas especies.

La primera parte del proyecto publicada, que comprende 8,000 km entre el Atlántico Norte y el Pacífico Tropical del Este (desde Nova Scotia, Canadá, a través del Golfo de México, cruzando por el Canal de Panamá y hasta la Polinesia Francesa pasando por las Galápagos), representa el mayor aporte de secuencias nuevas y sin caracterizar a la base de datos del GenBank con 7.7 millones de fragmentos de ADN (reads) que corresponden a 6.3 millones de pares de bases (bp), 57% del cual representa secuencias únicas. En ellas se detectaron en total 811 ribotipos distintos entre 4,125 fragmentos diferentes del gen 16SrRNA con una similitud de 97%. De éstos, 52% representan ribotipos de los que no se tienen representantes actualmente en las bases de datos, es decir, especies nuevas, y alrededor del 16% son organismos nuevos a nivel de familia por lo menos. Funcionalmente, la predicción de proteínas de la base de datos aportó 6.12 millones de secuencias de proteínas, duplicando la cantidad de éstas conocidas en las bases de datos públicas, de las cuales un 23.4% representan familias de proteínas no descritas previamente y que se encuentran exclusivamente en la base de

datos del GOS (es decir, existen otras familias sin caracterizar pero de las que se conocen homólogos en otras bases de datos<sup>29</sup>).

Pero la información aportada por el GOS no se reduce a una cantidad indescriptible de especies y proteínas nuevas, sino que aporta información fundamental sobre la estructura y composición de especies microbianas en las comunidades marinas de aguas oceánicas superficiales:

- i) El 73% de los reads de 16S hallados corresponden a 60 de los 811 ribotipos, es decir, que las comunidades marinas se encuentran dominadas por 60 especies diferentes de las cuales sólo una no se había descrito anteriormente, pero la mayoría carece de miembros cultivables.
- ii) Ninguno de los organismos dominantes pertenece al dominio Archaea, aunque previamente se había sugerido que eran dominantes en los océanos.
- iii) De los 60 organismos dominantes, tan sólo cinco son ubicuos (es decir que están presentes en todas las muestras): dos son representantes del “grupo SAR11” (así llamado originalmente ya que no se habían descrito estas especies, actualmente se sabe que incluye a *Pelagibacter ubique*), otros dos pertenecen al “grupo SAR86” y el otro es cercano a SAR1.
- iv) Las muestras tropicales se encuentran dominadas por varios grupos del clado SAR86, un grupo de Rhodospirillaceae y las cianobacterias unicelulares *Synechococcus* y *Prochlorococcus*.
- v) Las muestras templadas se encuentran dominadas por grupos relacionados a Roseobacter RCA, SAR11 y otras gamaproteobacterias que se encuentran ausentes en las muestras tropicales.

Lo anterior no sólo describe la estructura de las comunidades marinas microbianas, sino que también sugiere la existencia de regiones biogeográficas bacterianas, fenómeno cuya existencia ha generado polémica desde su postulación<sup>30</sup>. Otro dato que lo fundamenta es el análisis de asociación por *clusters* de todo el contenido génico de la muestra, es decir, la comparación de todos los fragmentos de ADN y no sólo el 16SrRNA de un sitio contra otro sitio. A pesar de que el GOS incluye muestras a ambos lados del continente americano, las muestras no se agrupan por distancia geográfica, sino por temperatura, es decir, las muestras de los mares templados del Atlántico se agrupan en un mismo lado, mientras que las muestras del Mar de los Sargazos y del Golfo de México y el Caribe se agrupan junto con las muestras tropicales del Pacífico, sugiriendo la existencia de Regiones Oceánicas Biogeográficas delimitadas.

Asimismo, el estudio presenta la oportunidad única de analizar

simultáneamente la funcionalidad y su distribución espacial, delineando la existencia de límites funcionales biogeográficos: de las 2,674 proteorhodopsinas detectadas en las muestras, se conocen tres variantes fisiológicas correspondientes a la sustitución de un solo residuo de aminoácidos: las variantes con Metionina (M) o Leucina (L) presentan absorbancia máxima hacia el espectro verde, mientras que la variante con Glutamina (Q) presenta una absorbancia máxima hacia el espectro azul. En las muestras del GOS, se encontró que mientras la variante M se encuentra dispersa a lo largo de todas las muestras, la variante L domina las aguas templadas del Atlántico cercanas a las costas de US y Canadá, mientras que la variante Q es más abundante en aguas cálidas alejadas de las costas.

### EL METAGENOMA DE LOS SIMBIOTES DE UN GUSANO: INTERACCIONES Y LA NUEVA DEFINICIÓN DE ORGANISMO (*Olavius algarvensis*)

Una aplicación totalmente diferente del análisis metagenómico es demostrada por el análisis de la comunidad subcuticular asociada al gusano *O. algarvensis*<sup>31</sup>, un anélido del grupo de los oligoquetos (el mismo que las lombrices de tierra), que viven en los sedimentos costeros marinos del Mediterráneo, y que tiene la particularidad de carecer de boca, ano y tracto digestivo. Lo anterior sugiere la existencia de una relación simbiótica obligada con una comunidad de microorganismos que habitan por debajo de la cutícula del gusano. Tradicionalmente, el análisis de interacciones simbióticas se limitaba a estudiar la simbiosis a nivel fisiológico de una pareja de organismos (*Rhizobium* en las leguminosas, las hormigas en las acacias, los peces mágidos con las anémonas), o de un par de poblaciones respecto a sus tasas poblacionales en presencia y ausencia del otro simbiote. El estudio metagenómico de la comunidad subcuticular permite analizar el potencial genómico de sus bacterias simbiotes extracelulares simultáneamente. El trabajo reveló la presencia de cuatro especies bacterianas simbióticas diferentes: dos delta y dos gama proteobacterias, todas con capacidad de fijación de carbono que provee al hospedero múltiples fuentes de nutrientes.

Los simbiotes gama-1 y gama-2 poseen ambos los genes necesarios para un estilo de vida quimioautotrófico mediante la oxidación del azufre, lo que representa el primer reporte de una asociación simbiótica con dos organismos metabólicamente equivalentes. La diferencia entre los dos radica en que gama-1 es capaz de almacenar azufre en glóbulos, mientras que gama-2 posee dos genes más para la oxidación del azufre, posiblemente brindándole una mayor versatilidad ecológica, y también posee genes para acoplar la oxidación de azufre a la reducción disimilatoria de nitrato en condiciones de oxígeno limitado.

De esta manera, se sabe que también los dos grupos de simbiotes se encuentran en simbiosis entre sí, generando un sistema cerrado donde los simbiotes gama realizan sulfo-oxidación y fijan carbono vía el ciclo de Calvin, y en oposición los simbiotes delta realizan sulfo-reducción acoplada a la oxidación del

hidrógeno y fijan carbono vía la ruta del-acetil coA o el TCA inverso.

Respecto al gusano hospedero, las cuatro especies de bacterias simbiotes pueden aportar compuestos orgánicos sintetizados mediante la fijación de CO<sub>2</sub> y la incorporación de carbono orgánico disuelto del medio. También poseen la capacidad de sintetizar todos los aminoácidos y algunas vitaminas. La toma de los nutrimentos por parte del hospedero seguramente ocurre mediante la lisis y la digestión intracelular de las bacterias, pues no se encontraron exportadores de nutrientes en los genomas bacterianos y se ha corroborado la lisis de los simbiotes en la región epidérmica basal del hospedero. Asimismo, los productos de excreción del gusano son incorporados a las células bacterianas mediante importadores de urea y amonía, no sólo eliminando los compuestos tóxicos para el eucarionte sino reciclando compuestos ricos en nitrógeno. Por último, los genomas codifican una gran variedad de importadores de compuestos osmoreguladores orgánicos (glicinbetaína, taurina, TMAO) que se sabe son sintetizados por organismos osmoconformadores (es decir que cambia las concentraciones de sales dentro de su cuerpo para igualarlas al del medio) como *O. algarvensis*, y que son utilizados por los simbiotes como fuente de carbono y nitrógeno. Finalmente, el trabajo permite construir un modelo de comportamiento ecológico conjunto del sistema simbiótico, en el cual el gusano puede desplazarse de las regiones superficiales (ricas en oxígeno y sulfatos pero carente de ácido sulfídrico) en donde puede realizar sulfooxidación de las reservas de azufre del simbiote gama-1 utilizando oxígeno a la par que realiza respiración aeróbica; hacia la región media (rica en nitrato y sulfato y carente de oxígeno), en donde el simbiote gama-3 puede utilizar nitrato para oxidar compuestos reducidos de azufre que proveen los simbiotes delta mediante la sulforreducción, que en las dos regiones superficiales se comportan como heterótrofos. En las regiones más profundas (anóxicas carentes de nitrato pero ricas en compuestos reducidos de azufre), los simbiotes delta se comportan como autótrofos fijando CO<sub>2</sub> a partir de la oxidación del hidrógeno y los simbiotes gama realizan respiración anaeróbica de fumarato acoplada a la recarga de las reservas de azufre mediante la oxidación parcial de compuestos reducidos de azufre del medio.

### LIMITACIONES Y PROMESAS DE LA METAGENÓMICA

Como toda subdisciplina naciente, la metagenómica presenta nuevos retos y problemáticas a resolver dentro del análisis ecológico de comunidades procariontes. El más obvio se refiere a la ausencia de estandarización en las metodologías de obtención de datos, causando una enorme heterogeneidad en la información contenida en los análisis metagenómicos (pero ver Raes *et al.*<sup>32</sup>). Se usan diferentes protocolos de extracción de ADN que varían en sesgos y eficiencia, se usan diferentes volúmenes de muestra, y varía mucho el número de células por muestra y cantidad de ADN total utilizada en la secuenciación, por lo que la comparación confiable entre metagenomas actualmente es muy difícil. Este

problema se amplifica por falta de costumbre para emplear metodologías estadísticas para “normalizar” las bases de datos durante las comparaciones (para evitar problemas relacionados con la cobertura y redundancia de secuencias). Todo lo anterior conduce a la subestimación de la similitud entre muestras, composición de la comunidad y potencial funcional.

Nosotros consideramos que la mayoría de las inconsistencias ecológicas presentadas por la metagenómica se deben a la naturaleza de la abundancia relativa de las especies contenidas, dado que el mismo método de secuenciación por *shotgun* es sensible a diferencias en las abundancias relativas, y muchas veces la especie ecológicamente más importante no es la numéricamente más abundante (dominante). Así, no es imposible que secciones importantes de la comunidad sean pasadas por alto.

Un problema similar ocurre con el análisis de variabilidad intraespecífica, que comprende la presencia de polimorfismos y elementos móviles, complicando el ensamble confiable de segmentos; a pesar de que la variabilidad es la esencia del estudio biológico y los biólogos y ecólogos microbianos debemos desarrollar nuevas aproximaciones para el manejo y valoración de la variabilidad dentro de una comunidad.

La secuencia del metagenoma de una comunidad representa el potencial funcional a disposición de la comunidad y no la verdadera funcionalidad de la misma, dado que la presencia de genes en un genoma no implica su expresión ni su traducción. Para una comprensión total de la funcionalidad bioquímica y biogeoquímica de una comunidad, es necesario realizar análisis de expresión de ARN, medir las tasas de traducción con variaciones temporales y finalmente la cinética enzimática de las proteínas más abundantes en condiciones naturales para poder realizar conclusiones respecto a la funcionalidad real de la comunidad. Sin embargo, el metagenoma es el único análisis de todos los mencionados que permite intuir cómo responderá la comunidad en condiciones medioambientales diferentes a aquéllas en las que se encuentra.

Otra dificultad técnica para el desarrollo de la metagenómica es que la secuenciación *en masse* dista mucho de ser barata. Todo proyecto metagenómico debe contemplar la secuenciación de grandes cantidades de ADN porque representan varias especies diferentes. Afortunadamente, los costos de secuenciación se abaratan a una velocidad impresionante gracias al desarrollo de nuevas tecnologías de secuenciación masiva como la pirosecuenciación (454 Life Sciences) o la secuenciación reversible basada en términos (Solexa/Illumina) (ver Cuadro 1). Sin embargo, la longitud de las secuencias que estas tecnologías producen es muy pequeña en comparación con los métodos de secuenciación tradicionales (Sanger). Esto representa un enorme problema cuando se intenta ensamblar las secuencias, pues si son cortas (de menos de 100 pares de bases) las regiones cortas

no poseen suficiente información para asignarlas a una secuencia lineal única confiable.

El proceso de ensamblado de un genoma o peor, de un metagenoma, es análogo a reconstruir un rompecabezas: considerando que todas las secuencias de la comunidad son un rompecabezas completo, la extracción y secuenciación de ADN produce una imagen fragmentada y necesariamente incompleta de la imagen total. Las nuevas tecnologías de secuenciación permiten obtener más piezas individuales del rompecabezas, pero estas piezas son más pequeñas. Si tenemos una comunidad muy compleja, es probable que no podamos reconstruir ni siquiera algunas regiones del rompecabezas y nos quedemos solamente con piezas individuales que no son muy informativas por sí solas. Por ejemplo, la secuenciación de 71 millones de pares de bases del metagenoma de la comunidad simbiótica del tracto digestivo de las termitas (*Nasutitermes*)<sup>33</sup> se intuía sencillo, sin embargo resultó en una comunidad compleja con alrededor de 216 filotipos (grupos filogenéticos) dentro de 12 *phyla* diferentes. Esto ocasiona que los fragmentos individuales no puedan ser ensamblados pues de los 71Mbp, las secuencias lineales más largas tenían tan sólo 15 kbp (un genoma de una bacteria individual es de alrededor de 4 Mb). Por lo tanto, los esfuerzos tecnológicos deben dirigirse no sólo al mejoramiento de las tecnologías de secuenciación para obtener fragmentos más largos y de mejor calidad, sino también al desarrollo de algoritmos de cómputo alternativos para el ensamble de los fragmentos obtenidos, pues actualmente no tenemos la capacidad ni de cómputo ni de recursos humanos para realizar esta tarea. Esto nos lleva a mencionar otro posible problema: la capacidad de secuenciación aumenta mucho más rápido que la capacidad de analizar los datos producidos, de manera que hoy en día tenemos suficientes datos como para emplear los próximos diez años en su análisis.

## LA METAGENÓMICA Y LA NUEVA ECOLOGÍA EN CUATROCIÉNEGAS

Cuatrociénegas es un oasis en el Desierto Chihuahuense que presenta una serie de cuerpos de agua conocidos localmente como pozas, más de 300 cristalinas, muchas termales a 30-40°C, algunas con ricas en sales, en un valle que alberga una alta diversidad de todo tipo de organismos<sup>34-38</sup>, sobresaliendo las comunidades formadoras de tapetes microbianos y estromatolitos, que en este ecosistema en particular son la base de la cadena trófica como lo eran hace 500 millones de años<sup>37,39</sup>. Esto probablemente se debe al aislamiento del valle y a que la pobreza extrema de fósforo en el ecosistema lo mantuvo pobre en algas y otros organismos cosmopolitas modernos. Los estromatolitos fósiles son la evidencia más antigua que tenemos de vida sobre nuestro planeta y que en este tipo de comunidad microbiana primigenia se dieron las condiciones metabólicas para la cohesión de los grandes ciclos biogeoquímicos (S, N, C, O) que hicieron posible la evolución de los eucariontes. Un hecho aún más notable es que las redes tróficas (constituidas por

varios niveles que van desde los caracoles herbívoros hasta los peces carnívoros) se mantienen en este sitio al “filo de la navaja” nutricional debido a estas condiciones limitantes<sup>39</sup>. En nuestra búsqueda de la biodiversidad a nivel de comunidades y poblaciones encontramos que la microbiota constituye un ecosistema similar a ambientes del pasado de la tierra<sup>36</sup>. Sus organismos endémicos aparentemente han sobrevivido historias de cambios climáticos, manteniendo a las comunidades de estromatolitos como base de la cadena alimenticia. La mayor parte de los microorganismos secuenciados no sólo tienen una filiación marina, sino que son muy diversos y divergentes a todo lo conocido<sup>36,38</sup>. El análisis de sus comunidades acuáticas nos indica que no hay migración significativa entre las pozas, aun si éstas son similares químicamente y están a pocos kilómetros de distancia y hay un viento constante en el valle que podrían generar la migración. Los organismos de estas pozas tienen un alto endemismo y diferenciación ecológica (diversidad beta), es decir, cada poza tiene diversidad a nivel de comunidad diferente a la que sigue y la diversidad total debe de ser enorme. El que existan tantas especies en un sitio muy pobre en nutrientes libres y con muchas sales probablemente se debe a que no sólo el ambiente es fluctuante, sino a que los virus están eliminando predominantemente a los linajes más abundantes<sup>40</sup> explicando la distribución equitativa de las comunidades y el gran recambio observado entre comunidades, las cuales estarían sujetas a una selección dependiente de la frecuencia donde los linajes raros son exitosos, mientras no se vuelvan abundantes.

Consideramos que la obtención de datos metagenómicos en Cuatrociénegas es una excelente apuesta para poder comprender no sólo la diversidad, sino cómo la diversidad se liga a funciones ecosistémicas básicas como ciclaje de nutrientes (C, N, S, P, H y O) y la estructuración de la cadena alimenticia. La posibilidad de que el tapete microbiano de Cuatrociénegas sea un “ecosistema completo” (o sea que contiene muchos de los procesos o ciclos biogeoquímicos importantes), lo hace un sitio privilegiado para la metagenómica, ya que los metagenomas anteriores, incluido el GOS, nos han revelado la complejidad de la mayor parte de los ecosistemas, donde los ciclos biogeoquímicos requieren de grandes áreas y diversos compartimentos (por ejemplo: zonas con oxígeno y anóxicas). Por otra parte, creemos que en un sitio aislado y relativamente simple (dada su extrema oligotrofia) es más fácil entender la compleja red de la vida. Si a esto agregamos que en el valle de Cuatrociénegas es especialmente interesante en términos de diversidad y por sus características ambientales únicas que nos hablan de la tierra primitiva, un metagenoma, junto con genomas de sus especies dominantes/clave, nos puede dar elementos clave no sólo sobre su ecología, sino también sobre su evolución. En colaboración con el Laboratorio Nacional para la Genómica y Diversidad (LANGEBIO), el CINVESTAV-Irapuato, el Instituto de Biotecnología de la UNAM y la UAM-Cuajimalpa, estamos usando nuestra experiencia en el GOS<sup>28,30</sup>, en el genoma de *B. coahuilensis*<sup>3</sup> y en las comunidades bacterianas de la región<sup>36-40</sup> para realizar un detallado estudio

metagenómico de Cuatrociénegas. El proyecto ha resultado complicado por los problemas que representa extraer ADN ambiental de buena calidad en esta comunidad, pero creemos que va a constituir un estudio muy redondo dentro de la metagenómica y la Nueva Ecología, que esperamos poder visitar en unos años en esta revista.

## EL FUTURO DE LA NUEVA ECOLOGÍA

A pesar de sus costos y problemas técnicos, la Nueva Ecología, y en particular la Metagenómica ecológica nos promete por fin lograr entender cómo funciona realmente el planeta y su biósfera: cuáles son los ciclos y procesos dominantes, los indispensables para la vida y los globales o cuáles son sus procesos más locales, qué organismos son los realmente importantes, especies clave para que operen los ciclos biogeoquímicos y para poder mantener la vida en la tierra y cómo funciona. Éste es un sueño que hasta ahora sólo lo imaginábamos, el abrir y entender qué hay dentro de la “caja negra” y qué representaban los microbios en los ecosistemas. En el 2009 vamos a celebrar el bicentenario del nacimiento de Charles Darwin, junto con otras celebraciones, como los 200 años de la publicación de la “Filosofía zoológica” de Lamarck, los 150 años de la publicación del Origen de las Especies y los 150 años de la muerte de Alexander Humboldt. La aplicación de las herramientas genómicas y metagenómicas nos permiten actualmente abordar el estudio de la biosfera desde el enfoque que propuso Darwin: el entendimiento de la naturaleza a partir de la visualización de la historia evolutiva de los organismos como producto de su funcionamiento con su entorno natural tanto biótico como abiótico. Es decir, nos permite comprender simultáneamente que la diversidad biológica actual es producto de las interacciones funcionales de los organismos con su entorno (el teatro ecológico) que se mantiene y que cambia a lo largo del tiempo (es decir, el drama evolutivo, siguiendo las ideas de Hutchinson<sup>41</sup>). La metagenómica nos permite analizar la historia evolutiva de los componentes individuales de las comunidades naturales actuales mediante el análisis evolutivo de los genes contenidos en ella utilizando las herramientas filogenéticas desarrolladas durante el siglo pasado<sup>1,42</sup>, así como analizar las interacciones y funcionalidad de las comunidades mediante el análisis de los genes funcionales complementándolo con los enormes avances observados en bioquímica y fisiología.

Así como los naturalistas del siglo XIX necesitaban una sólida formación en la sistemática de plantas y animales, en el estudio de su morfología, en geología, climatología y cartografía, los nuevos naturalistas, esto es, los ecólogos del siglo XXI, necesitan sólidos conocimientos en la nueva sistemática y filogenia de los organismos, sin descuidar a los procariontes y virus, deben tener una buena formación en estadística, genética, biología molecular y bioinformática. Es claro que los nuevos ecólogos no pueden poseer todo el conocimiento de la biología actual como lo tenían los buenos naturalistas del siglo XIX, pero deben de tener una buena formación que les permita interactuar, entender y trabajar

armoniosamente en grandes equipos dentro de proyectos genómicos, de ecología molecular y metagenómicos. Así, tenemos que repensar en cómo formar una nueva generación de biólogos competentes tanto en los métodos de muestreo y diseños experimentales de la Ecología clásica, como en las técnicas de laboratorio de Biología Molecular y en los análisis básicos informáticos de las Secuencias de ADN y de proteínas. Para ello, necesitarán poseer conocimientos robustos en Ecología Teórica, Evolución, Bioquímica, Genética y Bioinformática, aunque ninguno de los planes de estudio actuales cubre un perfil como éste.

Efectivamente, los retos que enfrentamos los nuevos ecólogos son impresionantes, y requieren de una nueva visión y formación y los estudios van a ser complicados y costosos, pero proporcionalmente los riesgos, retos y costos económicos no son mayores que los que enfrentaron y resolvieron los naturalistas del siglo XVIII y XIX, como Linneo, Lamarck, Humboldt o el mismo Darwin. No somos tan listos, pero somos muchos y tenemos una sólida base gracias a sus trabajos pioneros.

#### AGRADECIMIENTOS

Muchos colegas, amigos y alumnos han contribuido a nuestra formación, en nuestros proyectos y al desarrollo de estas ideas. Con la Dra. Luisa Falcón y el Dr. René Cerritos iniciamos nuestras meditaciones sobre metagenómica. En el desarrollo de los métodos de ecología molecular debemos mencionar a la Dra. Ana Escalante, a la M. en C. Laura Espinosa y de nuevo a la Dra. Falcón, que han trabajado intensamente muchos años con nosotros, tratando de avanzar en estos complicados problemas técnicos y conceptuales. La Dra. Erika Aguirre revisó el manuscrito. Diferentes proyectos del Conacyt nos han apoyado en nuestros estudios, especialmente el de Metagenómica (Conacyt SEP 57507) y otros (Semarnat-Conacyt 44673Q y Conacyt SEP-2004-C01-46475-Q) y una beca del Conacyt para que Germán Bonilla-Rosso realizara los estudios de doctorado.

#### REFERENCIAS

- Eguiarte, L.E., Souza, V. & Núñez-Farfán, J. La revolución darwiniana y la evolución molecular. *TIP Revista Especializada en Ciencias Químico-Biológicas* **1(1)**, 8-16 (1998).
- National Research Council. The role of theory in advancing 21st century biology: Catalyzing transformative research (The National Academies Press, Washington, D.C., 2008).
- Alcaraz, L.D., *et al.* The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. *Proceedings of the National Academy of Sciences of the United States of America* **105(15)**, 5803-5808 (2008).
- Tuskan, G.A., *et al.* The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313(5793)**, 1596-1604 (2006).
- Warren, W.C., *et al.* Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453(7192)**, 175-183 (2008).
- Eguiarte, L.E., Souza, V. & Aguirre, X. Ecología Molecular (INE, SEMARNAT, CONABIO, UNAM, México, D.F., 2007).
- Kluyver, A.J. & van Niel, C.B. The Microbe's Contribution to Biology (Harvard University Press, Cambridge, MA, 1956).
- Amann, R.I., *et al.* Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and environmental microbiology* **56(6)**, 1919-1925 (1990).
- Torsvik, V., Øvreas, L. & Thingstad, T.F. Prokaryotic Diversity—Magnitude, Dynamics, and Controlling Factors. *Science* **296(5570)**, 1064-1066 (2002).
- Methé, B.A., *et al.* Genome of *Geobacter sulfurreducens*: Metal Reduction in Subsurface Environments. *Science* **302(5652)**, 1967-1969 (2003).
- Bond, D.R. & Lovley, D.R. Electricity Production by *Geobacter sulfurreducens* Attached to Electrodes. *Applied and environmental microbiology* **69(3)**, 1548-1555 (2003).
- Strous, M., *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440(7085)**, 790-794 (2006).
- Pace, N.R., *et al.* Analyzing natural microbial populations by rRNA sequences. *ASM News* **51**, 4-12 (1985).
- Escalante, A.E. in Ecología Molecular (eds. Eguiarte, L.E., Souza, V. & Aguirre, X.) (INE, SEMARNAT, CONABIO, UNAM, México, D.F., 2007), pp. 393.
- Auchtung, T.A., Takacs-Vesbach, C.D. & Cavanaugh, C.M. 16S rRNA Phylogenetic Investigation of the Candidate Division "Korarchaeota". *Applied and environmental microbiology* **72(7)**, 5077-5082 (2006).
- Woese, C.R. Bacterial evolution. *Microbiological reviews* **51(2)**, 221-271 (1987).
- Woese, C.R., Kandler, O. & Wheelis, M.L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87(12)**, 4576-4579 (1990).
- Hugenholtz, P., Goebel, B.M. & Pace, N.R. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *Journal of Bacteriology* **180(18)**, 4765-4774 (1998).
- Konstantinidis, K.T. & Tiedje, J.M. Microbial diversity and genomics (John Wiley & Sons, New Jersey, 2004).
- Handelsman, J., *et al.* Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology* **5(10)**, R245-R249 (1998).
- Rondon, M.R., *et al.* Cloning the Soil Metagenome: a Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Applied and environmental microbiology* **66(6)**, 2541-2547 (2000).
- Magurran, A.E. Ecological Diversity and Its Measurements (Princeton University Press, Princeton, NJ, 1988).
- Venter, J.C. *et al.* Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **304(5667)**, 66-74 (2004).
- Béjà, O., *et al.* Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. *Science* **289(5486)**, 1902-1906 (2000).
- Béjà, O., *et al.* Proteorhodopsin phototrophy in the ocean. *Nature* **411(6839)**, 786-789 (2001).
- Ram, R.J., *et al.* Community Proteomics of a Natural Microbial Biofilm. *Science* **308(5730)**, 1915-1920 (2005).
- García Martín, H., *et al.* Metagenomic analysis of two enhanced

- biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology* **24**(10), 1263-1269 (2006).
28. Rusch, D.B. et al. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* **5**(3), e77 (2007).
29. Yooseph, S., et al. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. *PLoS Biology* **5**(3), e16 (2007).
30. Falcón, L.I., et al. Evidence of biogeography in surface ocean bacterioplankton assemblages. *Marine Genomics* (aceptado).
31. Woyke, T., et al. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**(7114), 950-955 (2006).
32. Raes, J., Foerstner, K.U. & Bork, P.B. Get the most out of your metagenome: computational analysis of environmental sequence data. *Current opinion in microbiology* **10**(5), 490-498 (2007).
33. Warnecke, F., et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**(7169), 560-565 (2007).
34. Minckley, W.L. Cuatro Ciénegas fishes: research review of a local test of diversity versus habit size. *Arizona-Nevada Acad Sci* **19**, 13-21 (1984).
35. Minckley, W.L. A bibliography for natural history of the Cuatro Cienegas basin and environs, Coahuila, Mexico. *Proc Des Fishes Council* **25**, 47-64 (1994).
36. Souza, V., et al. An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proceedings of the National Academy of Sciences of the United States of America* **103**(17), 6565-6570 (2006).
37. Souza V., Eguiarte, L.E., Siefert J. & Elser J. J. Microbial endemism: does phosphorus limitation enhance speciation? *Nature Reviews Microbiology* **6**, 559-564 (2008).
38. Escalante, A.E., et al. Diversity of aquatic prokaryotic communities in the Cuatro Cienegas basin. *FEMS Microbiology Ecology* **65**(1), 50-60 (2008).
39. Elser, J. J., et al. Effects of PO<sub>4</sub> enrichment and grazing snails on microbial communities in an ecosystem with living stromatolites. *Freshwater Biology* **50**, 1808-1825 (2005).
40. Desnues, C., et al. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**(7185), 340 (2008).
41. Hutchinson, G.E. *The Ecological Theater and the Evolutionary Play* (Yale University Press, New Haven and London, 1965), pp. 139.
42. Eguiarte, L.E., et al. El análisis filogenético: Métodos, problemas y perspectivas. *Boletín de la Sociedad Botánica de México* **60**, 169-181 (1997).