UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

DOCTORADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE ECOLOGÍA

# *"La metagenómica como herramienta de conservación de Cuatro Cienegas Coahuila: el caso del ciclo del azufre"*

TESIS

QUE PARA OPTAR POR EL GRADO DE DOCTOR EN CIENCIAS

P R E S E N T A:

VALERIE YSELLE DE ANDA TORRES

DIRECTOR DE TESIS:
DRA. VALERIA SOUZA SALDIVAR
INSTITUTO DE ECOLOGÍA

COMITÉ TUTORAL:

DRA. ICOQUIH ZAPATA PEÑASCO, INSTITUTO MEXICANO DEL PETRÓLEO IPN, MÉXICO
DR. LUIS DAVID ALCARAZ PERAZA, LABORATORIO NACIONAL DE CIENCIAS DE LA SOSTENIBILIDAD (LANCIS), MEXICO
TUTOR EXTERNO
DR. BRUNO CONTRERAS MOREIRA, AULA DEI CSIC ZARAGOZA ESPAÑA

CIUDAD DE MÉXICO JUNIO 2018

# Índice

*“There is still a window of time. Nature can win if we give her a chance.” - Jane Goodall*

*A Emilio, mi compañero de vida.*

*A mi madre*

*A mis cachorros*

Esta tesis se realizó bajo la dirección de la Dra. Valeria Souza Saldivar en el laboratorio de Evolución Molecular y Experimental en el lnstituto de Ecología de la Universidad Nacional Autónoma de México con el apoyo de la beca 356832 del Consejo Nacional de Ciencia y Tecnología (CONACYT).

Parte fundamental de esta tesis se llevo a cabo en el Laboratorio de biología computacional y estructural en la Estacion Experimental de Aula Dei CSIC Zaragoza España bajo la dirección del Dr. Bruno Contreras Moreira

El comité que participo en el desarrollo y realización de esta tesis doctoral estuvo conformado por:

Dra. Valeria Souza Saldivar

Dra Icoquih Zapata Peñasco

Dr. Luis David Alcaraz Peraza.

Dr. Bruno Contreras Moreira.

El jurado del exámen de grado estuvo conformado por:

| | |
|---|---|
| Presidente: | Dr. Pablo Vinuesa Fleischmann |
| Vocal: | Dra. Esperanza Martínez Romero |
| Vocal: | Dr. Víctor González Zúñiga |
| Vocal: | Dr. Luis José Delaye Arredondo |
| Secretario: | Dr. Luis David Alcaraz Peraza |

# Agradecimientos

A la Universidad Nacional Autónoma de México por ser la base de mi formación científica.

Al programa de Doctorado en Ciencias Biomedicas de la UNAM, y al Consejo Nacional de Ciencia y Tecnología (CONACyT) con numero de becario 356832 por permitirme realizar mis estudios de posgrado.

A la fundacion World Wildlife Fund (WWF)-Alianza Carlos Slim, Sep-Ciencia Básica Conacyt grant 238245 por el apoyo económico para realizar este trabajo de tesis doctoral. A la Fundación Morella-Soler por apoyarme en terminar en tiempo y forma esta tesis.

A la Dra. Valeria Souza y el Dr. Luis Eguiarte por abrirme las puertas de su laboratorio, por el apoyo, confianza y la guía brindada durante estos años.

A la Dra. Valeria Souza por el apoyo incondicional, la paciencia y la confianza depositada en mi. Te agradezco infinitamente los abrazos después de los tutorales, los cuidados cuando me desmayaba, la canasta que me alimentaba cuando no paraba de trabajar enfrente de la computadora, los consejos y el apoyo que me brindaste durante estos cinco años. Gracias por creer en mi, a enseñarme la humildad, y a nunca rendirme. Gracias por ser un ejemplo a seguir para luchar por un mundo mejor.

Al Dr. Luis Eguiarte. Gracias por el rigor científico, las arduas revisiones a los artículos y por enseñarme que el diablo esta en los detalles. Gracias por todo Luis!

A la Dra Icoquih Zapata por ayudarme a crecer como científica y como persona. Por los consejos y por el apoyo brindado durante estos años. Gracias por enseñarme el mundo anaerobio, el potencial *redox* y a darle un sentido integral a esta tesis. Gracias por tu amistad, por no dudar de mi y por apoyarme incodicionalmente.

Al Dr. Bruno Contreras Moreira por ser parte fundamental en mi formación científica. Gracias por recibirme en su laboratorio en Zaragoza y por enseñarme que el rigor científico no esta peleado con la calidez humana. Eres un extraordinario ser humano!. Gracias por tu calidad profesional y por la confianza que depositaste en mi desde un principio. Te agradezco infinitamente el apoyo que me has brindado en estos años, por sentarte conmigo a escribir el código de MEBS, por enseñarme Perl y apoyar mis ideas dispersas. Gracias por tu amistad y tu calidad como persona. Sin ti, esta tesis no hubiera sido posible.

Al Dr. Cesar Poot-Hernández. Hay tantas cosas que me gustaría agradecerte ya que, sin tu apoyo, asesoría, y sorporte emocional estaría totalmente perdida. Gracias por todos estos años de amistad, gracias por enseñarme el poder de Python, los notebooks de jupyter, y por apostar en que podíamos lograr cosas muy padres juntos Gracias!

Al comité sinodal, el comité tutoral y al Dr Luis Eguiarte por leer y corregir esta tesis. Gracias por sus valuosos comentarios que mejoraron sustancialmente este escrito.

Al Dr. Ernesto Igartua Arregui, Dra Ana Casas y Dr. Javier Ramos, por todo el apoyo brindado, los consejos, la paciencia y el soporte emocional. Gracias por aguantar mis risas desbordadas y por hacer que mi estancia en Zaragoza haya sido tan especial.

Al Dr. Carlos Cantalapiedra, por las discusiones que iban desde la interacción de los fotones hasta la diabetes flanil. No solo te agradezco las ayudas con awk, probabilidades de máximos teóricos, curvas ROC y combinatorias, si no también por darle sentido a las comidas y los almuerzos, a mis ideas dispersas y por dodo el apoyo. Sin ti definitivamente no hubiera sido lo mismo.

A la Dra Maribel Hernández y a Marcos Gonzales Laffitte por permitirme aprender el mundo de la inferencia de redes el cual fue crucial para el desarrollo de esta tesis. Gracias por el apoyo y la confianza.

A la Dra Niza Gamez por todos tus consejos, tu apoyo y sobre todo por tu amistad.

A la Dra. Mirna Vazquez, el Dr Gabriel Ponce, la M en C Jazmin Blaz. Gracias por compartir este camino conmigo, por escucharme, brindarme su apoyo y consejos para darle forma y sentido a preguntas fundamentales de la vida y del doctorado. Gracias por las risas compartidas en Cuatro Cienegas y en el laboratorio. Gracias por el apoyo y la amistad.

Al M.V y Z Manuel Rosas, por ser un ejemplo a seguir y un apoyo incondicional. Gracias por tu paz y tus consejos. Gracias por ayudarme en los muestreos en Cuatro Cienegas y por siempre estar ahí cuando mas lo he necesitado. Gracias por ser quien eres y por transmitir esa calidez humana a todos los que te rodean.

A los miembros del laboratorio de evolución molecular y experimental por compartir este camino conmigo, por las risas y por el apoyo. Gracias por soportarme todos estos años: Enrique, Niza, Mirna, Gabriel, Jazmin, Paola, Itzi, Manuel, Jonas, Josue, Yocelin Gracias por todo.

A la Dra Laura Espinosa, Dra Erika Aguirre, M.V y Z a Manuel Rosas y Sra Silvia Barrientos por su ayuda logística y técnica en el laboratorio.

Al Dr. Fernando García Guevara, por ayudarme con mis tareas de programación al inicio del doctorado, gracias por las asesorías y la ayuda.

A los miembros del la Estación Experimental de Aula Dei en Zaragoza España: Carlos, Ruben, Arantxa, Najla, Piere, Xesco, Alvaro, Uriel, Octavio. Gracias por hacer de mi estancia en Zaragoza una de las mejores experiencias de mi vida. Gracias por escucharme y apoyarme durante este tiempo. Os llevo en el corazón.

A Julia, Tomasa y Emilio. Gracias por abrirme las puertas de su casa y darme todo por el apoyo, comprensión y cariño que necesitaba. Gracias por alimentarme con la mejor comida aragonesa y valenciana que existe.

A Emilio. Por vivir conmigo todo este proceso, por tu apoyo, cariño, comprensión y consejos. No hay palabras para agradecerte todo lo que haces por mi. Eres pilar que me mantiene en pie, no sé que haría sin ti. el

A mi madre, por su amor y apoyo incondicional y a mi familia que siempre han estado pendientes de mi a pesar de la distancia. ¡GRACIAS!

A Dinguer, Sho, Tao, Hachi y pulguita. Gracias por transmitirme paz, y felicidad en momentos de crisis que me permitieron terminar esta tesis.

A Martin Alfonso Carrillo, por todo su apoyo en las colectas en Cuatro Cienégas, y por ayudarme a muestrear en el Churince con mi pie fracturado. Gracias Martin.

I would like to express my sincere gratitude to Dr Peter Stadler for the continuos support during the last year of my Ph.D, for his motivation and inmense knowledge who inspires everyone around him. Thanks for providing me the privilege not only to join their team at the Bioinformatics Group in Leipzig in several occations, but also for making me feel like family during my stays in Germany.

A Cuatro Cienégas por sus maravillosos paisajes y preguntas interminables. Gracias por permitirme observar por primera vez la vía láctea, los satélites alrededor de la tierra, las dunas de yeso, los tapetes microbianos y estromatolitos.

A los que fueron parte de este camino conmigo, por las experiencias las risas y el aprendizaje.

A quienes consulten esta tesis.

# 0.1 Introducción general

La alteración humana del medio ambiente ha causado una pérdida inmensurable de biodiversidad, alterando los procesos ecosistémicos, cambiando la resiliencia de los ecosistemas y modificando ciclos biogeoquímicos. Los cambios ambientales causados por las actividades humanas han aumentado la vulnerabilidad de los procesos ecosistémicos esenciales, mismos que son impulsados y sostenidos por las comunidades microbianas. Entender el impacto de las actividades humanas sobre dichas comunidades deben ser uno de los temas prioritarios en la ecología de la restauración y en la conservación a nivel de sistemas (Tilman, 2000; Oliver et al., 2015; Seymour, Lauro, & Brown, 2016)

Las actividades humanas impactan de manera directa e indirecta en todos los procesos ecosistémicos y en el flujo de los elementos en la Tierra, procesos que son mediados por la actividad microbiana. Por lo tanto, entender los mecanismos que subyacen a la respuesta de las comunidades microbianas frente a perturbaciones antropogénicas es de importancia imperativa para diseñar estrategias de manejo y conservación a nivel ecosistémico.

Generalmente, el efecto de las perturbaciones en las comunidades microbianas se traduce en una alteración en la disponibilidad de los recursos originales (nutrientes. o aceptores y donadores de electrones), o de las variables ambientales (pH, temperatura, incidencia de UV, etc.). Estas perturbaciones hacen que la comunidad cambie, debido a las nuevas condiciones, que implican un cambio en el nicho ecológico. Este proceso se le denomina sucesión, el cual implica un cambio en las proporciones de las especies que viven en la comunidad. Dependiendo de la magnitud y de la frecuencia temporal de la perturbación, una comunidad puede permanecer sin alteración alguna (resistencia), cambiar y después de un periodo de tiempo regresar a su estado original (resiliencia), cambiar en composición, pero recuperar sus funciones originales (redundancia funcional), o finalmente

cambiar en composición y funcionalidad modificando los procesos ecosistémicos en los que están involucrados.

Estas tres (resistencia, resiliencia y redundancia) son las propiedades emergentes que presentan las comunidades frente a una perturbación, y han sido objeto de revisiones recientes (Allison y Martiny, 2008; Konopka et al., 2015; Shade et al., 2012).

En ese sentido, se puede esperar un cambio en los servicios ecosistémicos proporcionados por las comunidades microbianas cuando la perturbación ambiental altera sus funciones originales, lo que generalmente implica un efecto negativo para la salud de los ecosistemas. A pesar de que existe un gran esfuerzo por entender cómo responden las comunidades microbianas frente a perturbación, las bases fundamentales que regulan dicha respuesta siguen sin entenderse. Por ejemplo, ¿Cuáles son los factores que contribuyen a la estabilidad de una comunidad microbiana frente a una perturbación? ¿Cuáles son las bases que dan lugar a la asociación que mantiene la dinámica de la comunidad en funcionamiento frente a una perturbación ambiental? ¿Cuáles son los mecanismos que contribuyen a la formación de comunidades resilientes y resistentes? y finalmente ¿Cuáles son los factores que estabilizan las comunidades microbianas y maximizan su diversidad? (Konopka et al., 2015).

En esta tesis, se buscó avanzar en la comprensión de los procesos ecológicos subyacentes a la estructura de comunidades microbianas particulares, identificando patrones comunes con otras comunidades bajo diferentes condiciones ambientales, con el objetivo final de mejorar el conocimiento de los mecanismos universales y construir un marco teórico sólido que nos permita tener bio-indicadores que puedan alerten sobre el estado del ecosistema bajo perturbación, como puede ser el cambio climático global.

Empleando la gran cantidad de información que se puede generar usando métodos de secuenciación masiva de alto rendimiento, en esta tesis estudie dos niveles de complejidad para entender la respuesta de las comunidades microbianas frente a perturbaciones ambientales, dividiendo así la tesis en cuatro capítulos.

En el Capítulo Uno, se resalta la importancia de los ciclos biogeoquímicos de las comunidades microbianas como una pieza de suma importancia para evaluar y modelar el posible impacto a nivel ecosistémico.

Posteriormente, en el Capítulo Dos, exploro la importancia de entender el impacto de las perturbaciones ambientales en comunidades microbianas estables y resistentes que son capaces de llevar a cabo el reciclaje y flujo de nutrientes, debido a que las comunidades microbianas son sistemas dinámicos en constante cambio, aún sin alteraciones externas. De esta forma, los tapetes microbianos resultan una valiosa herramienta como sistemas de estudio para distinguir la dinámica intrínseca de la comunidad, y de los factores ambientales que la afecta.

Posteriormente, estudio la problemática que existe en el sitio de estudio en el Capítulo Tres. El sitio la Cuenca de Cuatro Ciénegas (CCC), en el estado de Coahuila, al Norte de México, que cuenta con el mayor nivel de endemismo en América del Norte, y representa un oasis de alta biodiversidad en el desierto (Souza et al., 2006, 2012). A pesar de la gran importancia de este sitio único, la creciente demanda de agua para el desarrollo agrícola, forraje y alimentación animal ha afectado drásticamente los niveles de agua de una de las principales lagunas dentro de Cuatro Ciénegas, que es el sistema Churince, impactando en el funcionamiento de todos los ambientes que dependen del aporte de agua para su subsistencia. Enfatizando que es necesario la aplicación de aproximaciones robustas que no sólo analicen la dinámica de las comunidades microbianas en términos de estructura, composición y función, sino también del entendimiento de relaciones entre *taxa* que se llevan a cabo en dichas comunidades utilizando inferencia de redes.

Finalmente, en el último Capítulo se utilizaron todas las herramientas desarrolladas durante la tesis para determinar la pregunta central de este trabajo: ¿Cómo afecta la sobreexplotación de recursos naturales en los tapetes microbianos de Churince en Cuatro Ciénegas, Coahuila?

## 0.1.1 La importancia de evaluar los ciclos biogeoquímicos de manera integral.

A pesar de los grandes avances en ecología microbiana y de tecnologías como la secuenciación masiva, nuestra capacidad de entender e integrar los ciclos biogeoquímicos todavía es limitada. Esto se debe a que el flujo de los elementos esenciales para la vida en la Tierra (C, H, O, N, S, P) sucede través de una red de procesos biológicos, químicos y geofísicos que han co-evolucionado en la biósfera desde la aparición de los procesos metabólicos en la Tierra. Desde entonces, la historia evolutiva de la vida en nuestro planeta ha sido el resultado de complejas interacciones entre comunidades microbianas que difieren en sus capacidades metabólicas y ecológicas (Falkowski et al., 2008). Además, dichas comunidades microbianas acoplan efectivamente la transferencia de electrones (mediante reacciones *redox*), transformando los elementos y la energía derivados de procesos abióticos. Estos involucran el continuo suministro, así como la eliminación, de diversos depósitos de la superficie de la tierra, como así como diferentes procesos geotérmicos en el manto y la corteza del planeta, la meteorización de las rocas y diversos procesos fotoquímicos en la atmósfera, incluyendo el flujo constante de energía solar (Falkowski et al., 2008; Hedges, 1992; Newman and Banfield, 2002).

Debido a lo anterior, los ciclos biogeoquímicos han sido estudiados en diferentes disciplinas. Por ejemplo, la geoquímica y las ciencias atmosféricas se han enfocado en entender los grandes procesos abióticos en escalas planetarias, como son los procesos que influyen en el flujo de elementos desde y hacia los reservorios de la superficie de la Tierra y la atmósfera (Canfield et al., 2005; Falkowski et al., 2008). Por otra parte, en ecología microbiana se ha hecho hincapié en entender la relación entre actividades microbianas y procesos biogeoquímicos y ecosistémicos (Morales and Holben, 2011). Por ejemplo, el uso de sondas de isótopos estables (técnica mejor conocida en inglés como *stable-isotope probing*), combina el uso de métodos moleculares con la abundancia de isótopos estables para identificar microorganismos activamente involucrados en procesos metabólicos. De esta forma, se aísla el ADN de microorganismos que han asimilado sustratos enriquecidos

con isotopos estables para su crecimiento y posteriormente se utilizan técnicas de PCR (i.e., 16S rRNA) para su identificación taxonómica (Radajewski et al., 2000). Existen otros métodos que no requieren enriquecimiento en cultivos, y que se utilizan para establecer relaciones *in situ*; éstos métodos están basados en seleccionar genes marcadores específicos que participan en determinadas vías metabólicas relacionadas al reciclaje del carbono, nitrógeno, azufre y otros elementos (Ver Tabla 1). De esta forma, el ADN o el ARN obtenido directamente de ambientes naturales, es secuenciado y cuantificado, ya sea con PCR convencional, secuenciación Sanger, análisis de microarreglos, GeoChip (Tu et al., 2014a), metagenómica (Llorens-Marès et al., 2015; Quaiser et al., 2011; Swingley et al., 2012) o metatranscriptómica (Chen et al., 2014; Stewart et al., 2011).

Tabla 1. Genes marcadores moleculares más utilizados en estudios de ecología microbiana para analizar los ciclos biogeoquímicos en muestras ambientales

| *Marcador molecular* | *Proceso metabólico* | *Referencia* |
|---|---|---|
| *nifHD* | Fijación de nitrógeno | (Gaby and Buckley, 2012; López-Lozano et al., 2012; Severin, Lucas, 2007; Severin et al., 2010) |
| *amoAB* | Nitrificación | (Calvó and Garcia-Gil, 2004; Quaiser et al., 2011; Rotthauwe et al., 1997) |
| *nirS y nirK* | Denitrificación | (Braker et al., 2000) |
| *nrfA y nargG* | Reducción de nitrato a ammoniaco desasimilativa | (Jiménez et al., 2012; Philippot et al., 2002; Smith et al., 2007) |
| *dsrAB* | Reducción de sulfato desasimilativa | (Daly et al., 2000; Dar et al., 2007a; Kerkhof, 2005; Leloup et al., 2004; Loy et al., 2009a) |
| *aprAB* | Reducción y oxidación de azufre | (Hügler et al., 2010; Meyer and Kuever, 2007b; Watanabe et al., 2013a) |
| *soxB* | Oxidación de Azufre | (Hügler et al., 2010) |
| *sqr* | Oxidación de azufre | (Pham et al., 2008) |
| *mcrA* | Metanogénesis | (Luton et al., 2002; Ufnar et al., 2007) |
| *pmoA* | Oxidación de metano | (Freitag et al., 2010; Luesken et al., 2011) |
| *psbA* | Fotosíntesis oxigénica | (Zeidner et al., 2003) |
| *pufM* | Fotosíntesis anoxigénica | (Waidner and Kirchman, 2008) |
| *acLb* | Fijación autotrófica de Carbono | (Campbell et al., 2003; Hügler et al., 2010) |
| *ccbLM* | Fijación autotrófica de carbono | (Campbell et al., 2003; Hügler et al., 2010) |
| *coxLMS* | Fijación de Carbono | (Quaiser et al., 2011) |

Sin embargo, a pesar de los grandes avances en ecología microbiana, a la fecha no existen métodos sistemáticos que permitan evaluar, integrar, comparar y facilitar el estudio de los ciclos biogeoquímicos con datos genómicos y metagenómicos a gran escala. Por esta razón, en el Capítulo Uno de esta tesis, se valida un nuevo algoritmo computacional para capturar y evaluar la importancia de los principales ciclos biogeoquímicos (C, N, O, Fe, S y otras rutas metabólicas complejas) a escala multigenómica (usando hasta miles de genomas y/o metagenomas), utilizando un solo valor informativo denominado MEBS (por sus siglas en inglés: *Multigenomic Entropy Based Score*). El marco de referencia utilizado para implementar el algoritmo está basado en la divergencia de Kullback-Leibler, también conocida como entropía relativa *H'* (Kullback and Leibler, 1951), la cual ha sido ampliamente utilizada en física, en teoría de la comunicación y en inferencia estadística, siendo interpretada como medida de desorden, información e incertidumbre, respectivamente. En este trabajo se utilizó el concepto de teoría de la comunicación para resumir la información procedente de la maquinaria metabólica a partir de los datos genómicos y metagenómicos disponibles.

Para validar el algoritmo se realizó una anotación manual del ciclo del azufre, completando el primer inventario de representantes taxonómicos completamente secuenciados (Anexo 1), incluyendo genes, pasos enzimáticos, y vías metabólicas involucradas. Utilizando MEBS se evaluó un total de 2,107 genomas no redundantes, 900 metagenomas públicos, incluyendo metagenomas reportados en trabajos previos de (Bonilla-Rosso et al., 2012; Breitbart et al., 2009b; Nitti et al., 2012; Peimbert et al., 2012) y 35 metagenomas de Churince, Cuatro Ciénegas (analizados durante el trabajo de esta tesis doctoral), incluyendo 23 muestras de agua y sedimento antes de un experimento de enriquecimiento con tres diferentes proporciones de N:P (16:1, 75:1 y solo aumentar P) *vs.* el control (Lee et al., 2015, 2017), y 12 tapetes microbianos de la misma Laguna que se analiza en los Capítulo Tres y Cuatro de esta tesis.

Para evaluar la precisión y reproducibilidad del algoritmo, se utilizaron diferentes métodos, incluyendo muestreos aleatorios, modelos de regresión lineal y curvas ROC (por sus siglas en inglés: *receiver operator characteristic*). Los resultados obtenidos indican que

MEBS tiene una alta precisión para clasificar organismos previamente caracterizados a nivel bioquímico o fisiológico para metabolizar compuestos de azufre. Esta fase experimental bioinformática y de biología computacional permitió además identificar genomas sin representantes cultivados, como *Candidatus Desulforudis audaxviator* recolectados en ambientes relacionados con el azufre (e.g., ventilas hidrotermales). El análisis usando MEBS sirvió también para proponer el uso de 12 dominios de proteínas como marcadores del ciclo del azufre en datos metagenómicos.

El algoritmo actualmente cuenta con cuatro modalidades: i) entrenamiento y cálculo del MEBS *Score* en genomas y metagenomas; ii) detección de familias de proteínas como posibles marcadores moleculares en datos metagenómicos; iii) determinación de vías metabólicas completas y parciales analizadas por medio de patrones de presencia y ausencia de las familias de proteínas involucradas en dicho metabolismo, y iv) visualización de las familias de proteínas de un genoma o un metagenoma de interés, utilizando los mapas metabólicos de la base de datos de la *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (Kanehisa y Goto, 2000). Finalmente, se resalta la importancia de utilizar un valor cuantificable para evaluar cualquier metabolismo a escala local y lo cual puede potencialmente cambiar la visión actual para inferir capacidades metabólicas complejas en la era -omica (De Anda et al., 2017)

## 0.1.2 La complejidad del ciclo del azufre y su importancia global

El Capítulo Dos de esta tesis está dividido en dos secciones. En la primera sección se hace una recapitulación de ciclo del azufre y su importancia en la historia de la vida en la Tierra. Se encontró convergencia entre los datos geológicos, moleculares, genómicos y el algoritmo MEBS desarrollado, con el que se confirman los ambientes en donde la evidencia más reciente indica los posibles ambientes en donde se originó la vida en la Tierra. Además, en este Capítulo también se sintetiza la gran diversidad metabólica del ciclo del azufre dentro del árbol de la vida (Hug et al., 2016). Por otro lado, se presenta a los tapetes microbianos como modelos ecológicos de la evolución biogeoquímica temprana. La segunda sección del

Capítulo Dos, consta de un artículo de divulgación publicado en la revista *Oikos=,* del Instituto de Ecología de la UNAM, en donde se hace una recapitulación del proceso de formación de los átomos en el universo hasta el su reciclamiento en la Tierra y las dificultades actuales para entender los ciclos biogeoquímicos a escala global, enfocándose en el ciclo del azufre.

La gran cantidad de información acumulada en las últimas tres décadas (Anexo 2), y revisada en detalle en van Gemerden (1993), Seckbach y Aharon (2010), Foster y Green (2011), Bolhius et al. (2014) ha demostrado que los tapetes microbianos y los estromatolitos son comunidades estables y flexibles que se han adaptado a los continuos cambios ambientales en la Tierra desde el Arqueano. Debido a esto, en el Capítulo Tres se explica cómo es que los tapetes microbianos son un buen modelo de estudio para entender cómo es que cambian las comunidades frente a una perturbación antropogénica causada por sobre-explotación de agua.

### 0.1.3 La problemática en Cuatro Ciénegas

En el último Capítulo se realizó un estudio de serie temporal para evaluar la dinámica de los tapetes microbianos en la Laguna del Churince en Cuatro Ciénegas. Primero se realizó una descripción general de la diversidad y de las funciones potenciales de cada muestra en el tiempo y en el espacio. A partir de esa descripción general, se tomó como punto de partida la inferencia de las capacidades metabólicas para llevar a cabo los ciclos biogeoquímicos, los cuales, de verse afectados, se reflejaría en los procesos ecosistémicos. En estos análisis se empleó el algoritmo desarrollado durante esta tesis y que se describe en el Capítulo Uno.

Finalmente, en este Capitulo Cuarto se tomó en cuenta la compleja red de interrelación que llevan a cabo los miembros de las comunidades y cómo éstas cambiaban ante la perturbación, así como la recuperación del sistema para determinar la sensibilidad de la comunidad ante la perturbación. Los resultados indican que existe un incremento de las relaciones negativas bajo perturbación (e.g., competencia, antagonismo), mientras que existe un equilibro entre relaciones positivas y negativas cuando existe agua en la Laguna. Además, identificamos que ciertos organismos en muy baja abundancia, son cruciales en los

tapetes microbianos, actuando como especies clave que podrían llevar a cabo funciones cruciales para la estabilidad de estas comunidades bajo condiciones de perturbación.

# 0.2 Referencias


Allison, S. D., and Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. PNAS. Available at: http://www.pnas.org/content/105/suppl.1/11512.short.

Bolhuis, H., Cretoiu, M. S., and Stal, L. J. (2014). Molecular Ecology of Microbial Mats. FEMS Microbiol. Ecol. 90, 335–50. doi:10.1111/1574-6941.12408.

Bonilla-Rosso, G., Peimbert, M., Alcaraz, L. D., Hernández, I., Eguiarte, L. E., Olmedo-Alvarez, G., et al. (2012). Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments. Astrobiology 12, 659–673. doi:10.1089/ast.2011.0724.

Braker, G., Zhou, J., Wu, L., Devol, H., and Tiedje, J. M. (2000). Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities. Appl. Environ. Microbiol. 66, 2096–2104. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=101460%7B&%7Dtool=pmcentrez%7B&%7Drendertype=abstract.

Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., Dinsdale, E., et al. (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. Environ. Microbiol. 11, 16–34. doi:10.1111/j.1462-2920.2008.01725.x.

Calvó, L., and Garcia-Gil, L. J. (2004). Use of amoB as a new molecular marker for ammonia-oxidizing bacteria. J. Microbiol. Methods 57, 69–78. doi:10.1016/j.mimet.2003.11.019.

Campbell, B. J., Stein, J. L., and Cary, S. C. (2003). Evidence of Chemolithoautotrophy in the Bacterial Community Associated with Alvinella pompejana , a Hydrothermal Vent Polychaete Evidence of Chemolithoautotrophy in the Bacterial Community Associated with Alvinella pompejana , a Hydrothermal Vent Polychae. doi:10.1128/AEM.69.9.5070.

Canfield, D., Kristensen, E., and Bo, T. (2005). The Sulfur cycle. 1st ed. , eds. D. Canfield, E. Kristensen, and Thamdrup Bo Elsevier Academic Press.

Chen, L., Hu, M., Huang, L., Hua, Z., Kuang, J., Li, S., et al. (2014). Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. ISME J., 1–14. doi:10.1038/ismej.2014.245.

Daly, K., Sharp, R. J., and McCarthy, a J. (2000). Development of oligonucleotide probes and PCR primers for detecting phylogenetic subgroups of sulfate-reducing bacteria. Microbiology 146 ( Pt 7, 1693–705. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10878133.

Dar, S. a, Yao, L., van Dongen, U., Kuenen, J. G., and Muyzer, G. (2007). Analysis of diversity and activity of sulfate-reducing bacterial communities in sulfidogenic bioreactors using 16S rRNA and dsrB genes as molecular markers. Appl. Environ. Microbiol. 73, 594–604. doi:10.1128/AEM.01875-06.

De Anda, V., Zapata-Peñasco, I., Poot-Hernandez, A. C., Eguiarte, L. E., Contreras-Moreira, B., and Souza, V. (2017). MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle Authors. Gigascience 6, 1–17. doi:10.1093/gigascience/gix096.

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science 320, 1034–1039. doi:10.1126/science.1153213.

Foster, J. S., and Green, S. J. (2011). STROMATOLITES: Interaction of Microbes with Sediments. 18, 385–405. doi:10.1007/978-94-007-0397-1.

Freitag, T. E., Toet, S., Ineson, P., and Prosser, J. I. (2010). Links between methane flux and transcriptional activities of methanogens and methane oxidizers in a blanket peat bog. FEMS Microbiol. Ecol. 73, 157–65. doi:10.1111/j.1574-6941.2010.00871.x.

Gaby, J. C., and Buckley, D. H. (2012). A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. PLoS One 7, e42149. doi:10.1371/journal.pone.0042149.

Hedges, J. I. (1992). Global biogeochemical cycles: progress and problems. Mar. Chem. 39, 67–93. doi:10.1016/0304-4203(92)90096-S.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., et al. (2016). A new view of the tree of life. Nat. Microbiol. 1, 16048. doi:10.1038/nmicrobiol.2016.48.

Hügler, M., Gärtner, A., and Imhoff, J. F. (2010). Functional genes as markers for sulfur cycling and CO2 fixation in microbial communities of hydrothermal vents of the Logatchev field. FEMS Microbiol. Ecol. 73, 526–37. doi:10.1111/j.1574-6941.2010.00919.x.

Jiménez, D. J., Andreote, F. D., Chaves, D., Montaña, J. S., Osorio-Forero, C., Junca, H., et al. (2012). Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. PLoS One 7, e52069. doi:10.1371/journal.pone.0052069.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102409&tool=pmcentrez&rendertype=abstract.

Kerkhof, L. J. (2005). Phylogeography of Sulfate-Reducing Bacteria among Disturbed Sediments , Disclosed by Analysis of the Dissimilatory Sulfite Reductase Genes ( dsrAB ) Phylogeography of Sulfate-Reducing Bacteria among Disturbed Sediments , Disclosed by Analysis of the Dissi. doi:10.1128/AEM.71.2.1004.

Konopka, A. E., Lindemann, S., and Fredrickson, J. K. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. ISME J. 9, 1488–1495. doi:10.1038/ismej.2014.251.

Kullback, S., and Leibler, R. A. (1951). On Information and Sufficiency. Ann. Math. Stat. 22, 79–86. doi:10.1214/aoms/1177729694.

Labbate, M., Seymour, J. R., Lauro, F., and Brown, M. V (2016). Editorial Anthropogenic Impacts on the Microbial Ecology and Function of Aquatic Environments. 7, 6–8. doi:10.3389/fmicb.2016.01044.

Lee, Z., Poret-Peterson, A. T., Siefert, J. L., Kaul, D., Moustafa, A., Allen, A. E., et al. (2017). Nutrient stoichiometry Shapes Microbial Community Structure in an Evaporitic Shallow Pond. Front. Microbiol. 8, 1–15. doi:10.3389/fmicb.2017.00949.

Lee, Z., Steger, L., Corman, J. R., Neveu, M., Poret-Peterson, A. T., Souza, V., et al. (2015). Response of a stoichiometrically imbalanced ecosystem to manipulation of nutrient supplies and ratios. PLoS One 10, 1–17. doi:10.1371/journal.pone.0123949.

Leloup, J., Quillet, L., Oger, C., Boust, D., and Petit, F. (2004). Molecular quantification of sulfate-reducing microorganisms (carrying dsrAB genes) by competitive PCR in estuarine sediments. FEMS Microbiol. Ecol. 47, 207–14. doi:10.1016/S0168-6496(03)00262-9.

Llorens-Marès, T., Yooseph, S., Goll, J., Hoffman, J., Vila-Costa, M., Borrego, C. M., et al. (2015). Connecting biodiversity and potential functional role in modern euxinic environments by microbial metagenomics. ISME J., 1–14. doi:10.1038/ismej.2014.254.

López-Lozano, N. E., Eguiarte, L. E., Bonilla-Rosso, G., García-Oliva, F., Martínez-Piedragil, C., Rooks, C., et al. (2012). Bacterial communities and the nitrogen cycle in the gypsum soils of Cuatro Ciénegas Basin, coahuila: a Mars analogue. Astrobiology 12, 699–709. doi:10.1089/ast.2012.0840.

Loy, A., Duller, S., Baranyi, C., Mussmann, M., Ott, J., Sharon, I., et al. (2009). Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environ. Microbiol. 11, 289–99. doi:10.1111/j.1462-2920.2008.01760.x.

Luesken, F. a, Zhu, B., van Alen, T. a, Butler, M. K., Diaz, M. R., Song, B., et al. (2011). pmoA Primers for detection of anaerobic methanotrophs. Appl. Environ. Microbiol. 77, 3877–80. doi:10.1128/AEM.02960-10.

Luton, P. E., Wayne, J. M., Sharp, R. J., and Riley, P. W. (2002). The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill. Microbiology 148, 3521–3530. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12427943.

Meyer, B., and Kuever, J. (2007). Molecular Analysis of the Diversity of Prokaryotes in the Environment, Using aprA as Functional Marker Gene Molecular Analysis of the Diversity of Sulfate-Reducing and Sulfur-Oxidizing Prokaryotes in the Environment , Using aprA as Functional Marker Gene. doi:10.1128/AEM.01272-07.

Morales, S. E., and Holben, W. E. (2011). Linking bacterial identities and ecosystem processes: Can “omic” analyses be more than the sum of their parts? FEMS Microbiol. Ecol. 75, 2–16. doi:10.1111/j.1574-6941.2010.00938.x.

Newman, D. K., and Banfield, J. F. (2002). Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. Science 296, 1071–7. doi:10.1126/science.1010716.

Nitti, A., Daniels, C. a, Siefert, J., Souza, V., Hollander, D., and Breitbart, M. (2012). Spatially resolved genomic, stable isotopic, and lipid analyses of a modern freshwater microbialite from Cuatro Ciénegas, Mexico. Astrobiology 12, 685–98. doi:10.1089/ast.2011.0812.

Oliver, T. H., Isaac, N. J. B., August, T. A., Woodcock, B. A., Roy, D. B., and Bullock, J. M. (2015). Declining resilience of ecosystem functions under biodiversity loss. Nat. Commun. 6, 10122. doi:10.1038/ncomms10122.

Peimbert, M., Alcaraz, L. D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L., et al. (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, 648–658. doi:10.1089/ast.2011.0694.

Pham, V. H., Yong, J.-J., Park, S.-J., Yoon, D.-N., Chung, W.-H., and Rhee, S.-K. (2008). Molecular analysis of the diversity of the sulfide : quinone reductase (sqr) gene in sediment environments. Microbiology 154, 3112–21. doi:10.1099/mic.0.2008/018580-0.

Philippot, L., Piutti, S., Martin-Laurent, F., Hallet, S., and Germon, J. C. (2002). Molecular Analysis of the Nitrate-Reducing Community from Unplanted and Maize-Planted Soils. Appl. Environ. Microbiol. 68, 6121–6128. doi:10.1128/AEM.68.12.6121-6128.2002.

Prieto-Barajas, C. M., Valencia-Cantero, E., and Santoyo, G. (2017). Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. Electron. J. Biotechnol. 31, 48–56. doi:10.1016/j.ejbt.2017.11.001.

Quaiser, A., Zivanovic, Y., Moreira, D., and López-García, P. (2011). Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. ISME J. 5, 285–304. doi:10.1038/ismej.2010.113.

Radajewski, S., Ineson, P., Parekh, N. R., and Murrell, J. C. (2000). Stable-isotope probing as a tool in microbial ecology. Nature 403, 646–649. doi:10.1038/35001054.

Rotthauwe, J. H., Witzel, K. P., and Liesack, W. (1997). The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl. Environ. Microbiol. 63, 4704–4712. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=168793%7B&%7Dtool=pmcentrez%7B&%7Drendertype=abstract.

Seckbach, J., and Aharon, O. (2010). Microbial mats: modern and ancient microorganisms in stratified systems. , eds. J. Seckbach and O. Aharon Heidelberg: Springer doi:10.1007/978-90-481-3799-2.

Severin, Lucas, S. (2007). NifH expression by five groups of phototrophs compared with nitrogenase activity in coastal microbial mats. FEMS Microbiol. Ecol. 73, 55–67. doi:10.1111/j.1574-6941.2010.00875.x.

Severin, I., Acinas, S., and Stal, L. (2010). Diversity of nitrogen-fixing bacteria in cyanobacterial mats. FEMS Microbiol. Ecol. 73, 514–25. doi:10.1111/j.1574-6941.2010.00925.x.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., B??rgmann, H., et al. (2012). Fundamentals of microbial community resistance and resilience. Front. Microbiol. 3, 1–19. doi:10.3389/fmicb.2012.00417.

Smith, C. J., Nedwell, D. B., Dong, L. F., and Osborn, a M. (2007). Diversity and abundance of nitrate reductase genes (narG and napA), nitrite reductase genes (nirS and nrfA),

and their transcripts in estuarine sediments. Appl. Environ. Microbiol. 73, 3612–3622. doi:10.1128/AEM.02894-06.

Souza, V., Espinosa-Asuar, L., Escalante, A. E., Eguiarte, L. E., Farmer, J., Forney, L., et al. (2006). An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. Proc. Natl. Acad. Sci. U. S. A. 103, 6565–6570. doi:10.1073/pnas.0601434103.

Souza, V., Siefert, J. L., Escalante, A. E., Elser, J. J., and Eguiarte, L. E. (2012). The Cuatro Ciénegas Basin in Coahuila, Mexico: an astrobiological Precambrian Park. Astrobiology 12, 641–7. doi:10.1089/ast.2011.0675.

Stewart, F. J., Dmytrenko, O., Delong, E. F., and Cavanaugh, C. M. (2011). Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of solemya velum. Front. Microbiol. 2, 134. doi:10.3389/fmicb.2011.00134.

Swingley, W. D., Meyer-Dombard, D. R., Shock, E. L., Alsop, E. B., Falenski, H. D., Havig, J. R., et al. (2012). Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. PLoS One 7, e38108. doi:10.1371/journal.pone.0038108.

Tilman, D. (2000). Causes, consequences and ethics of biodiversity. Nature 405, 208–211.

Tu, Q., Yu, H., He, Z., Deng, Y., Wu, L., Van Nostrand, J. D., et al. (2014). GeoChip 4: A functional gene-array-based high-throughput environmental technology for microbial community analysis. Mol. Ecol. Resour. 14, 914–928. doi:10.1111/1755-0998.12239.

Ufnar, J. a, Ufnar, D. F., Wang, S. Y., and Ellender, R. D. (2007). Development of a swine-specific fecal pollution marker based on host differences in methanogen mcrA genes. Appl. Environ. Microbiol. 73, 5209–17. doi:10.1128/AEM.00319-07.

van Gemerden, H. (1993). Microbial mats: A joint venture. Mar. Geol. 113, 3–25. doi:10.1016/0025-3227(93)90146-M.

Waidner, L. a, and Kirchman, D. L. (2008). Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene pufM in the Delaware estuary. Appl. Environ. Microbiol. 74, 4012–21. doi:10.1128/AEM.02324-07.

Watanabe, T., Kojima, H., Takano, Y., and Fukui, M. (2013). Diversity of sulfur-cycle prokaryotes in freshwater lake sediments investigated using aprA as the functional marker gene. Syst. Appl. Microbiol. 36, 436–443. doi:10.1016/j.syapm.2013.04.009.

Zeidner, G., Preston, C. M., Delong, E. F., Massana, R., Post, A. F., Scanlan, D. J., et al. (2003). Molecular diversity among marine picophytoplankton as revealed by psbA analyses. Environ. Microbiol. 5, 212–216. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12588300.

# 0.3 Artículos

Los artículos que fueron producto de esta tesis doctoral se listan a continuación. El primero de ellos fue presentado como requisito para el trámite de titulación de Doctorado.

## Capitulo 1

1. **De Anda V*,** Zapata-Peñasco I, Poot-Hernandez AC, Eguiarte LE, Contreras-Moreira B*, Souza V*. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle. *GigaScience*. 2017;6(11):1-17. (*corresponding author doi:10.1093/gigascience/gix096.)

## Capitulo 2

2. **De Anda V**, Zapata-Peñasco I, Souza V. The sulfur cycle as the gear of the "clock of life": the point of convergence between geological and genomic data in the Cuatro Cienegas Basin. In Ecosystem ecology and geochemistry of Cuatro Cienegas: How to survive in an extremely oligotrophic site. Elser J, García-Oliva F, Souza V (ed) Springer. **Book Chapter In Press**
3. **De Anda V** and Souza V. Azufre: elemento incomprendido de la biogeoquímica planetaria En La Química de la Vida No. 16 Oikos Revista de divulgación del Instituto de Ecología UNAM

## Capitulo 3

4. **De Anda V**, Souza V, Zapata Peñasco. Towards a comprehensive understanding of environmental perturbations in microbial mats from Cuatro Cienegas by network modeling. In Ecosystem ecology and geochemistry of Cuatro Cienegas: How to survive in an extremely oligotrophic site. Elser J, García-Oliva F, Souza V (ed) Springer. **Book Chapter In Press**

## Capitulo 4

5. **De Anda V**, Zapata-Peñasco I, Blaz J, Poot-Hernández CA, Contreras Moreira B, Gonzales-Lattiffe M, Hernández-Rosales M, Gamez Tamariz N, Eguiarte L, Souza V. Highlighting the importance of negative interactions and rare biosphere within microbial mats under perturbation: a time-series metagenomics study. "Characterizing Modern Microbialites and the Geobiological Processes Underlying Their Formation". Frontiers in Microbiology. **In Press.**

# Capítulo 1


# MEBS: a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle

**Referencia:**

**De Anda V*,** Zapata-Peñasco I, Poot-Hernandez AC, Eguiarte LE, Contreras-Moreira B*, Souza V*. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle. *GigaScience*. 2017;6(11):1-17. (*corresponding author doi:10.1093/gigascience/gix096.)

**Reconocimientos derivados de este articulo:**

- Ganador de The Bioinformatics Peer Prize II en la categoria de estudiante: https://www.thinkable.org/vote_competitions/the-bioinformatics-peer-prize-ii
- Semifinalista de GigScience Prize Track en la Conferencia Internacional de Genomica ICG-12 https://academic.oup.com/gigascience/pages/prize_track
- Semifinalista del Premio Langebio 2017

**Algoritmo:**

El algoritmo incluyendo todos los scripts, asi como el manual y datasets de prueba se encuentra publicamente disponible en el siguiente repositorio de GitHub

https://eead-csic-compbio.github.io/metagenome_Pfam_score/

*GigaScience*, 6, 2017, 1–17

doi: 10.1093/gigascience/gix096
Advance Access Publication Date: 23 October 2017
Research

RESEARCH

# MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle

Valerie De Anda[1,*], Icoquih Zapata-Peñasco[2], Augusto Cesar Poot-Hernandez[3], Luis E. Eguiarte[1], Bruno Contreras-Moreira[4,5,*] and Valeria Souza[1,*]

[1]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, 70-275, Coyoacán 04510, D.F., México, [2]Dirección de Investigación en Transformación de Hidrocarburos, Instituto Mexicano del Petróleo, Eje Central Lázaro Cárdenas, Norte 152, Col. San Bartolo Atepehuacan, 07730, México, [3]Departamento de Ingeniería de Sistemas Computacionales y Automatización. Sección de Ingeniería de Sistemas Computacionales. Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas. Circuito Escolar 3000, Cd. Universitaria, 04510 Ciudad de México, [4]Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Avda. Montañana, 1005, Zaragoza 50059, Spain and [5]Fundación ARAID, calle María de Luna 11, 50018 Zaragoza, Spain

*Corresponding address. Valerie de Anda, Instituto de Ecología, Universidad Nacional Autónoma de México, 70-275, Coyoacán 04510, D.F., México; Tel: (+52)5556229006; Fax: +52-5556228995; E-mail valdeanda@ciencias.unam.mx; Bruno Contreras Moreira, Estación Experimental de Aula Dei, Avda. Montañana, 1005, Zaragoza 50059; Spain; Tel: (+34)976716089; Fax: +34 976716145; E-mail bcontreras@eead.csic.es; Valeria Souza, Instituto de Ecología, Universidad Nacional Autónoma de México, 70-275, Coyoacán 04510, D.F., México; Tel: (+52)5556229006; Fax: +52-5556228995; E-mail souza@unam.mx

## Abstract

The increasing number of metagenomic and genomic sequences has dramatically improved our understanding of microbial diversity, yet our ability to infer metabolic capabilities in such datasets remains challenging. We describe the Multigenomic Entropy Based Score pipeline (MEBS), a software platform designed to evaluate, compare, and infer complex metabolic pathways in large "omic" datasets, including entire biogeochemical cycles. MEBS is open source and available through https://github.com/eead-csic-compbio/metagenome_Pfam_score. To demonstrate its use, we modeled the sulfur cycle by exhaustively curating the molecular and ecological elements involved (compounds, genes, metabolic pathways, and microbial taxa). This information was reduced to a collection of 112 characteristic Pfam protein domains and a list of complete-sequenced sulfur genomes. Using the mathematical framework of relative entropy ($H'$), we quantitatively measured the enrichment of these domains among sulfur genomes. The entropy of each domain was used both to build up a final score that indicates whether a (meta)genomic sample contains the metabolic machinery of interest and to propose marker domains in metagenomic sequences such as DsrC (PF04358). MEBS was benchmarked with a dataset of 2107 non-redundant microbial genomes from RefSeq and 935 metagenomes from MG-RAST. Its performance, reproducibility,

**Received:** 13 June 2017; **Revised:** 9 September 2017; **Accepted:** 1 October 2017

and robustness were evaluated using several approaches, including random sampling, linear regression models, receiver operator characteristic plots, and the area under the curve metric (AUC). Our results support the broad applicability of this algorithm to accurately classify (AUC = 0.985) hard-to-culture genomes (e.g., *Candidatus Desulforudis audaxviator*), previously characterized ones, and metagenomic environments such as hydrothermal vents, or deep-sea sediment. Our benchmark indicates that an entropy-based score can capture the metabolic machinery of interest and can be used to efficiently classify large genomic and metagenomic datasets, including uncultivated/unexplored taxa.

*Keywords:* metabolic machinery; metagenomics; omic-datasets; Pfam domains; relative entropy; sulfur cycle; multigenomic entropy-based score

## Background

Over the last 15 years, the enormous advances in high-throughput sequencing technologies have revolutionized the field of microbial ecology, dramatically improving our understanding of life's microbial diversity to an unprecedented level of detail [1–4].

Nowadays, accessing the total repertoire of genomes within complex communities by means of metagenomics is becoming a standard and routine procedure in order to attain the full insight of the diversity, ecology, evolution, and functional makeup of the microbial world [5]. Furthermore, the accurate reconstruction of microbial genomes and draft-populations from environmental metagenomic studies has been shown to be a powerful approach [6–10], providing clues about the potential metabolic strategies of hard-to-culture microbial lineages by linking the functional mechanisms that support specific metabolisms with taxonomic, systematic, and ecological contexts of that lineage [8].

Despite the accelerated accumulation of large collections of metagenomic and genomic sequences, our ability to analyze, evaluate, and compare complex metabolic capabilities in large-scale "omic" datasets remains biologically and computationally challenging [11]. Predicting the metabolic potential is a key step in describing the relationship between a microbial community and its ecosystem function. This is largely performed by mapping the protein coding genes of "omic" data onto reference pathway databases such as MetaCyc [12] or KEGG [13] based on their homology to previously characterized genes [14]. The currently available methods for metabolic pathway prediction or reconstruction rely on the use of several metrics to infer the overall repertoire of metabolic pathways present in a given metagenomic dataset (e.g., MinPath [14], HUMAnN [15], PRMT [16], MetaPathways [17]).

However, due to the challenges involved in testing meaningful biological hypotheses with complex data, only a small proportion of the metabolic information derived from these datasets is eventually used to draw ecologically relevant conclusions. In this regard, most of the microbial ecology-derived "omic" studies have been mainly focused on either: (i) developing a broad description of the metabolic pathways within a certain environment [18, 19]; (ii) analyzing the relative abundance of marker genes involved in several metabolic processes and in certain ecosystems (e.g., primary productivity, decomposition, biogeochemical cycling) [20–24]; or (iii) discovering differentially abundant, shared or unique functional units (genes, proteins, or metabolic pathways) across several environmental metagenomic samples [25–27].

Therefore, in order to integrate all available omics data, we propose a novel approach to reduce the complexity of targeted metabolic pathways involved in several integral ecosystem processes—such as entire biogeochemical cycles—into a single informative score, called the Multigenomic Entropy-Based Score (MEBS). This approach is based on the mathematical rationalization of Kullback-Leibler divergence, also known as relative entropy $H'$ [28]. Relative entropy has been widely applied in physics, communication theory, and statistical inference, and it is interpreted as a measure of disorder, information, and uncertainty, respectively [29]. Here we use the communication theory concept of $H'$ to summarize the information derived from the metabolic machinery encoded by the protein coding genes of "omic" datasets. The application of this metric in biology was originally developed by Stormo and colleagues, who identified binding sites that regulate gene transcription [30].

In order to evaluate the performance of our approach, we selected the sulfur cycle (from now on, S-cycle) because this is one of the most metabolically and ecologically complex biogeochemical cycles, but there are few studies analyzing the complete repertoire (genes, proteins, or metabolic pathways) involved in the mobilization of inorganic-organic sulfur compounds through microbial-catalyzed reactions at a planetary scale [20, 31–35].

## MEBS Description

MEBS (MEBS, RRID:SCR_015708) runs in Linux systems [36]. For practical purposes, the MEBS algorithm was divided into 4 stages, summarized in Fig. 1 and explained below.

### Stage 1: manual curation of Sulfur cycle and "omic" datasets

#### *Sulfur taxonomic representatives*

A dataset comprehensively covering the currently known representatives of the S-cycle was obtained from primary literature and the MetaCyc database [12]. Each taxonomic representative (at the genus or species level) was selected under the criteria of having evidence suggesting their physiological and biochemical involvement in the degradation, reduction, oxidation, or disproportionation of sulfur compounds. Then, each taxonomic representative was scanned against our genomic dataset (see further details below) in order to obtain a list containing the completely sequenced and non-redundant genomes of the S-cycle. The resulting Sulfur list (or "Suli") currently contains 161 curated genomes and was used as the first input of the pipeline. Both the manually curated taxonomic representatives and Suli can be found in Table S1.

#### *Random taxonomic representatives (Rlist)*

As a negative control, we generated 1000 lists of genomes that are not particularly enriched on sulfur metabolic preferences. Each list contains 161 random genomes, the same number of microorganisms included in Suli. These lists were obtained by randomly subtracting from the genomic dataset (see below) 161 Refseq accession numbers and their corresponding names.

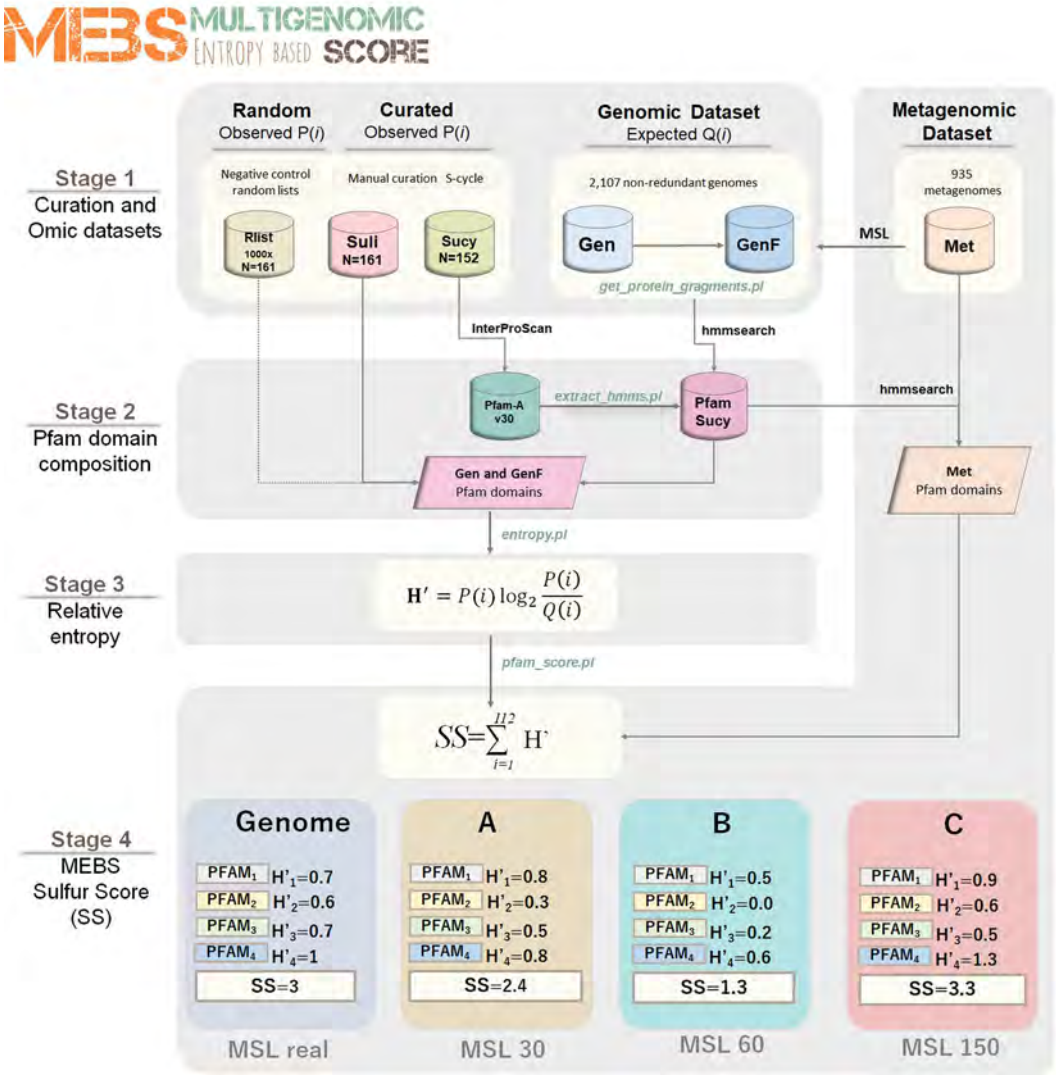



**Figure 1:** Schematic representation of the 4 stages of the MEBS algorithm focusing on the S-cycle. The first step consists of the systematic curation of a database containing the metabolic information of the S-cycle, which is reduced to a FASTA file of proteins involved (Sucy) and a list of 161 related microorganisms (Suli). A thousand lists of 161 random-sampled genomes were used as negative control (Rlist). The training dataset comprises 2107 genomes (Gen), which were fragmented in different sizes by considering the mean size length (MSL) of 935 metagenomes (Met). In the second stage, the domain composition of Sucy proteins is obtained by scanning Pfam-A, resulting in the Pfam-Sucy database. Then, the relative entropy ($H'$) of each Sucy-Pfam domain is obtained in the third stage. Finally, the precomputed entropies in Gen and GenF are used to evaluate full-length genomic sequences (real) and metagenomic sequences of variable MSL (in this example A, B, and C).

### *Metabolic pathways and genes*

We gathered and classified the metabolic pathways involved in the S-cycle from the primary literature and 2 experimentally validated curated databases: KEGG (KEGG, RRID:SCR_012773) [13] and MetaCyc (MetaCyc, RRID:SCR_007778) [12]. All the molecular information was then combined into a single database named Sucy (for sulfur cycle). Sucy currently contains 152 genes and 48 enzyme classification numbers annotated in the Enzyme classification [37] (Table S2). The 152 FASTA sequences of the proteins encoded by these genes were downloaded from UniProt [38] and used as the second input of the pipeline.

### *Genomic dataset (Gen)*

At the time of the analysis (December 21, 2016), a total of 4158 genomes were available from RefSeq database [39]. For comparative genomic purposes, we removed redundancy in this large dataset by using the Web interface [40, 41]. As a phylogenomic distance measure, we used a modified version of the Genomic Similarity Score (GSSb) [41]; we selected the most tolerant threshold of 0.95 (so as not to drop many sequenced genomes) and default parameters, resulting in 2107 clusters containing similar genomes, ordered by size (largest to smallest). Then, the largest genome representative for each group was searched in the NCBI genome assembly summary file [42] and downloaded from the NCBI FTP site [43].

### *Metagenomic dataset (Met)*

We used the Meta Genome Rapid Annotation using Subsystem Technology server (MG-RAST, RRID:SCR_004814) [44] to download metagenomes that: (i) were publicly available; (ii) contained associated metadata; and (iii) had been isolated from well-defined environments (i.e., rivers, soil, biofilms),

discarding host-associated microbiome sequences (i.e., human, cow, chicken). In addition, we included 35 unpublished metagenomes derived from sediment, water, and microbial mats from Cuatro Ciénegas, Coahuila (CCC), Mexico. The latter were also submitted and annotated in the MG-RAST server and will be described in depth elsewhere. The resulting collection of 935 FASTA files (≈500 GB), containing gene-called protein sequences (MG-RAST stage 350), were downloaded from the RESTful MG-RAST API [45]. While these metagenomes were evaluated and scored in STAGE 4, they were also analyzed to estimate their mean sequence length, considering that the fragmented nature of metagenomic sequences would have an impact on homology detection, depending on the length of the reads [46, 47]. Therefore, we measured the mean size length (MSL) of the peptide sequences of the 935 metagenomes in Met and the 152 curated proteins in Sucy, which are summarized in Fig. S1. It was observed that the MSL of Met varies broadly, with a majority of metagenomic peptides with MSL ≤ 30 aa, and that Sucy proteins range from 49 to 1020 aa, with MSL = 349 aa. According to this distribution, the metagenomes in Met were grouped into 7 well-defined categories: MSL ≤30, ≤60, ≤100, ≤150, ≤200, ≤250, ≤300 aa.

#### *Fragmented genomic dataset (GenF)*

In order to simulate the observed variability of MSL across metagenomes, protein sequences encoded in the genomic dataset (Gen, containing 2107 genomes) were *in silico* sheared with Perl script *get_protein_fragments.pl* into the 7 MSL categories defined above (30 to 300). This produced the GenF dataset, which currently requires up to 104 GB of disk space.

## Stage 2: domain composition of the input proteins

The annotation of protein domains in Sucy was conducted using Interproscan 5.21–60.0 [48] against databases Pfam-A v30 (Pfam, RRID:SCR_004726) [49], TIGRFAM v13 (JCVI TIGRFAMS, RRID:SCR_005493) [50], and Superfamily v1.75 (SUPERFAMILY, RRID:SCR_007952) [51]. Then, the Hidden Markov Models (HMMs) from matched Pfam domains (n = 112) were extracted from Pfam-A using script *extract_hmms.pl*. These selected HMMs were subsequently scanned against the Genomic, Genomic Fragmented, and Metagenomic datasets (from now on, "omic" datasets, see subsequent stages) using HMMER 3.0, the *hmmsearch* –cut_ga option [52].

## Stage 3: relative entropy and its use in detecting informative domains

In order to detect protein domains enriched among sulfur-based microorganisms (Suli), we used a derivative of the Kullback-Leibler divergence [28]—also known as relative entropy $H'(i)$—to measure the difference between probabilities P and Q (see Equation (1) below). In this context, $P(i)$ represents the frequency of protein domain $i$ in the 161 Suli genomes (observed frequency), while $Q(i)$ represents its frequency in the 2107 genomes in Gen (expected frequency). The script to compute the entropy (*entropy.pl*) requires the list of the genomes of interest (Suli) and the tabular output file obtained from the scanning of Gen and GenF against the Pfam-Sucy database. The obtained values of $H'$ (in bits) capture to what extent a given Pfam domain informs about the metabolism of interest. In this case, domains with $H'$ values close to or greater than 1 correspond to the most informative Pfam domains (enriched among S-based genomes), whereas low $H'$ values (close to 0) indicate non-informative ones. Negative values correspond to those observed less than expected.

$$H' = P(i) \log_2 \frac{P(i)}{Q(i)} \quad (1)$$

As a negative control, the $H'$ of the 112 Pfam domains were recalculated in both the Gen and GenF datasets, but replacing Suli with 1000 equally sized lists of random-sampled genomes (Rlist).

We evaluated the impact of the MSL in the computed entropy values using Gen and GenF. First, we focused on detecting informative Pfam domains that could be used as possible molecular marker genes in variable-length metagenomic sequences. Specifically, we looked for domains displaying stable $H'$ values across both Gen and GenF by using the script *plot_cluster_comparison.py*, which implements the following methods: K-Means, Affinity propagation, Mean-shift Spectral, Ward hierarchical, Agglomerative, DBSCAN, and Birch. All of these are part of the scikit-learn Machine Learning Python module [53].

## Stage 4: final score, interpretation, properties, and benchmark

Peptide sequences from a given genome or metagenome of interest are evaluated by first scanning their Pfam domains and then producing a final score, defined as the sum of the precomputed entropies of matched S-related Pfam domains (see Equation (2)). This score (Sulfur Score [SS] in our case) summarizes the information content of the metabolic machinery of interest. In this context, informative sulfur protein domains would contribute to higher SS, whereas non-informative ones would decrease it. This is an extension of procedures originally developed for the alignment of DNA and protein motifs, in which individual positions are independent and additive and can be simply summed up to obtain the total weight or information content [30]. Instead of aligning sequences, in our context we added up the entropy values of the Pfam domains matched in a given "omic" sample (resulting from scanning the sample of interest against Pfam-Sucy), from which a total weight (SS) is computed by using script *pfam_score.pl*.

$$SS = \sum_{i=1}^{112} H' \quad (2)$$

Datasets in which the majority of informative S-cycle protein domains are represented will yield a high SS; in contrast, low SS values should be expected if proteins involved in the S-cycle are not particularly enriched.

#### *MSL*

As the calculation of the SS depends on the MSL of the omic sample of interest, script *pfam_score.pl* supports option –size in amino acid residues (aa). In this way, appropriate precomputed $H'$ values for Pfam domains can be selected to produce the final score. Currently 30, 60, 100, 150, 200, 250, 300, and real sizes are supported.

#### *Metabolic pathway completeness and KEGG visualization*

The presence-absence patterns of Pfam domains belonging to particular pathways can be exploited to compute metabolic completeness. This optional task is invoked with

parameter –keggmap and a TAB-separated file mapping Pfam identifiers to KEGG Orthology entries (KO numbers) and the corresponding pathway in Sucy (see Table S3). To compute completeness, the total number of domains involved in a given pathway (i.e., sulfate reduction, sulfide oxidation) must be retrieved from the Sucy database (See Table S2). Then, the protein domains currently present in any given sample are divided by the total number of domains in the pre-defined pathway. The script produces: (i) a detailed report of the metabolic pathways of interest and (ii) a list of KO numbers with Hex color codes, corresponding to KO matches in the omic sample, which can be exported to the KEGG Mapper–Search & Color Pathway tool (see Fig. S2) [54].

*Properties and performance of SS*

Since the outcome of the final score (*SS*) largely depends on the list of microorganisms involved in the metabolism of interest (in our case, Suli) and the Pfam domains found in the input protein sequences (n = 112), we evaluated its robustness and reproducibility with several approaches. First, we compared our results with a benchmark performed 3 years ago in which we used Pfam-A v27 (instead of version 30), a genomic dataset containing 1528 non-redundant genomes (579 fewer genomes than our current Genomic dataset), and an input list of 156 genomes of interest (5 fewer that our current Suli). Second, *SS* estimates were compared with scores obtained by randomly selecting ≈50% of the 112 Pfam domains with both Gen and Met. This analysis was performed a thousand times with *pfam_score.pl* –random. Third, we benchmarked the predictive capacity of the *SS* in order to accurately classify genomes of S-related organisms (Suli, n = 161, positive instances), in contrast with a larger set of non-redundant genomes (Gen–Suli, n = 1.946, negative instances). Therefore, we computed the true positive rates (TPR), false positive rates (FPR), receiver operating characteristic (ROC) plots, and the resulting area under the curve (AUC) using the scikit-learn module [53].

## Results and Discussion

We present MEBS, a new open source software to evaluate, quantify, compare, and predict the metabolic machinery of interest in large "omic" datasets. The pipeline includes 4 stages. The first one consists of the systematic and targeted acquisition of the molecular and ecological information describing the metabolism of interest, represented by a list of curated microorganisms and a FASTA file of proteins involved in that metabolic network. In the second stage, the domain composition of the curated proteins is evaluated. Then, the domains enriched among the microorganisms of interest are identified by using the mathematical framework of the relative entropy (*H'*, third stage). Finally, the summation of the entropy of individual Pfam domains in a given genome or metagenomic dataset yields the final score (see Fig. 1).

To test the applicability of this approach, we evaluated the metabolic machinery of the S-cycle. Due to its multiple redox states and its consequences on microbiological and geochemical transformations, S-metabolism can be observed as a complex metabolic machinery, involving a myriad of genes, enzymes, organic substrates, and electron carriers, which largely depend on the surrounding geochemical and ecological conditions. For these reasons, the complete repertory involved in the metabolic machinery of the S-cycle has remained underexplored despite the massive data produced in "omic" experiments. Here, we performed an integral curation effort to describe all the elements involved in the S-cycle and then used, as explained in the following sections, to score genomic and metagenomic datasets in terms of their sulfur relevance.

### Manual curation: the complex metabolic machinery of the sulfur cycle

In order to integrate the complete biogeochemical S-cycle, we manually curated and modeled the major processes involved in the mobilization and use of S-compounds through Earth's biosphere. This effort resulted in 2 comprehensive databases. The first one includes most of the known microorganisms (with and without complete genomes) described in the literature to be closely involved in the S-cycle (Table S1). In this database, we included representative taxa from the following metabolic sulfur guilds: (i) chemolithotrophic, colorless sulfur bacteria (CLSB: 24 genera); (ii) anaerobic phototrophs, purple sulfur bacteria (PSB: 25 genera), and green sulfur bacteria (GSB: 9 genera); (iii) sulfate-reducing bacteria (SRB: 40 genera); and (iv) deep-branch sulfur hyperthermophilic microorganisms, such as elemental sulfur–reducing (SRM: 19 genera) and oxidizers (SO: 4 genera). From all the microorganisms described to be involved in the S-cycle, at the time of the analysis, a total of 161 were found to be completely sequenced and non-redundant genomes, and these were used as the first input of the pipeline (Suli).

The second database (Sucy) contains genes, proteins, and pathways with experimental evidence linking them to the S-cycle. To compile this database, we first gathered the most important S-compounds derived from biogeochemical processes and biological catalyzed reactions. Then we classified each S-compound according to its chemical and thermodynamic nature (Gibbs free energy of formation). Finally, we classified whether each compound can be used as a source of carbon, nitrogen, energy, or as an electron donor, fermentative substrate, or terminal electron acceptor in respiratory microbial processes. The schematic representation of the manually curated effort summarizing the complexity of the sulfur biogeochemical cycle in a global scale is shown in Fig. 2.

Once we selected the microorganisms, genes, and biogeochemical processes involved, we systematically divided the metabolic machinery of the S-cycle into 28 major metabolic pathways described in Table 1. In general terms, we included pathways involved in: (i) the oxidation/reduction of inorganic S-compounds, used as source of energy, electron donor or acceptor (P1-P7, P11 and P20 and P21); (ii) the degradation of organic S-compounds, such as aliphatic sulfonates, sulfur amino acids, and organosulfonates (P8-P10, P12-P19, P22, P23, P27); (iii) the methanogenesis from methylated thiols, such as dimethyl sulfide DMS (P24), metylthio-propanoate (P25), and methanethiol (P26), which are generated in nature by different biogeochemical processes [12]; and finally, (iv) the biosynthesis of sulfolipids (SQDG) (P28), because it has been observed that some bacteria living in S-rich and P-lacking environments are able to synthetize sulfolipids, instead of phospholipids, in the membrane as an adaptation to the selective pressures of these particular environments [55]. The synthetic pathway P29 is explained in further detail in the next sections (Table 1).

After the comprehensive metabolic inventory was compiled, we linked all the elements in a single network representation of the S-metabolic machinery (Fig. 3). To the best of our knowledge, this is the first molecular reconstruction of the cycle that considers all the sulfur compounds, genes, proteins, and the corresponding enzymatic steps resulting in higher-order molecular pathways. The latter representation also highlights the

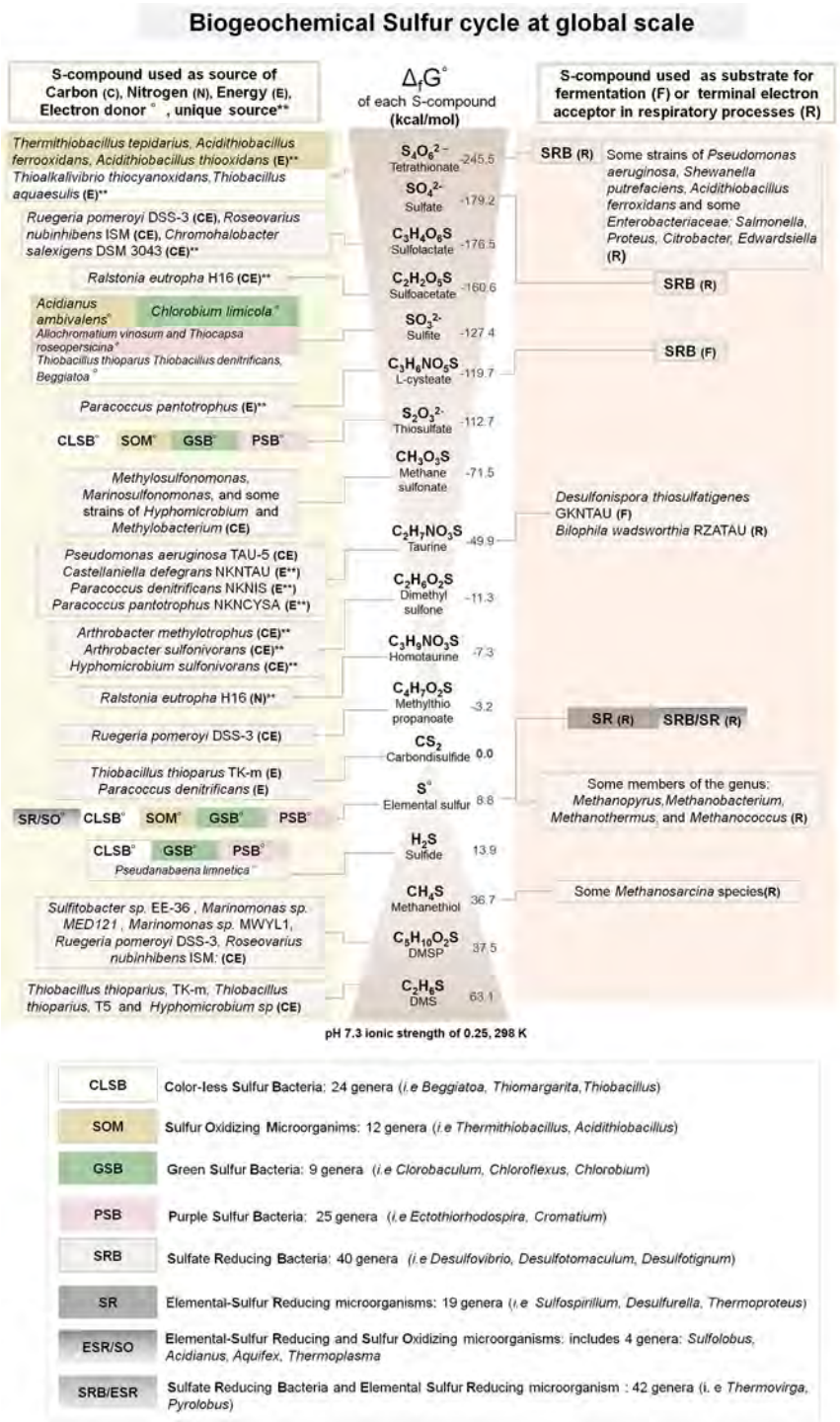



**Figure 2:** Sulfur cycle at a global scale. The most important organic and inorganic S-compounds derived from biogeochemical processes are arranged according to the Standard Gibbs free energy of formation described in Caspi et al. [12]. The left column indicates whether specific microorganisms are able to use those S-compounds as a source of carbon (C), nitrogen (N), energy (E), or electron donors (°). Double asterisks indicate whether the S-compound is used as sole source of C, N, or E. The corresponding electron acceptors in redox-coupled reactions using the S-compound as electron donor are not shown. The right column indicates whether the S-compound is used as fermentative substrate (F) or terminal electron acceptor in respiratory processes (R). Colored boxes summarize the metabolic guilds involved in the metabolism of S-compounds in oxidation (i.e., CLSB, SOM, PSB, and GSB) or reduction (SR, SRB) processes. The complete list of S-based microorganisms (Suli) is found in Table S1. Figure based on annotations from MetaCyc [12].

**Table 1:** Metabolic pathways of global biogeochemical S-cycle

| Pathway number | Metabolism[a] | Chemical process[b] | Sulfur compound | Type[c] | Chemical formula | Source[d] | Number of Pfam domains[e] |
|---|---|---|---|---|---|---|---|
| P1 | DS | O | Sulfite | I | $SO_3^{2-}$ | E | 9 |
| P2 | DS | O | Thiosulfate | I | $S_2O_3^{2-}$ | E | 10 |
| P3 | DS | O | Tetrathionate | I | $S_4O_6^{2-}$ | E | 2 |
| P4 | DS | R | Tetrathionate | I | $S_4O_6^{2-}$ | E | 17 |
| P5 | DS | R | Sulfate | I | $SO_4^{2-}$ | E | 20 |
| P6 | DS | R | Elemental sulfur | I | $S^o$ | E | 20 |
| P7 | DS | D | Thiosulfate | I | $S_2O_3^{2-}$ | E | 9 |
| P8 | DS | O | Carbon disulfide | O | $CS_2$ | E | 1 |
| P9 | A | DE | Alkanesulfonate | O | $CH_3O_3SR$ | S | 5 |
| P10 | A | R | Sulfate | I | $SO_4^{2-}$ | S | 20 |
| P11 | DS | O | Sulfide | I | $H_2S$ | E/S | 29 |
| P12 | A | DE | L-cysteate | O | $C_3H_6NO_5S$ | C/E | 1 |
| P13 | A | DE | Dimethyl sulfone | O | $C_2H_6O_2S$ | C/E | 3 |
| P14 | A | DE | Sulfoacetate | O | $C_2H_2O_5S$ | C/E | 2 |
| P15 | A | DE | Sulfolactate | O | $C_3H_4O_6S$ | C/S | 14 |
| P16 | A | DE | Dimethyl sulfide | O | $C_2H_6S$ | C/S | 16 |
| P17 | A | DE | Dimethylsulfoniopropionate | O | $C_5H_{10}O_2S$ | C/S/E | 12 |
| P18 | A | DE | Methylthiopropanoate | O | $C_4H_7O_2S$ | C/S | 7 |
| P19 | A | DE | Sulfoacetaldehyde | O | $C_2H_3O_4S$ | C/S | 7 |
| P20 | DS | O | Elemental sulfur | I | $S^\circ$ | C/S/E | 7 |
| P21 | DS | D | Elemental sulfur | I | $S^\circ$ | C/S/E | 1 |
| P22 | A | DE | Methanesulfonate | O | $CH_3O_3S$ | C/S/E | 7 |
| P23 | A | DE | Taurine | O | $C_2H_7NO_3S$ | C/S/E | 11 |
| P24 | DS | M | Dimethyl sulfide | O | $C_2H_6S$ | C | 1 |
| P25 | DS | M | Metylthio-propanoate | O | $C_4H_7O_2S$ | C | 1 |
| P26 | DS | M | Methanethiol | O | $CH_4S$ | C | 1 |
| P27 | A | DE | Homotaurine | O | $C_3H_9NO_3S$ | N | 1 |
| P28 | A | B | Sulfolipid | O | SQDG | | 4 |
| P29 | | | Markers | | Markers | | 12 |

[a]Metabolism: assimilative (A) inorganic compounds are reduced during biosynthesis; dissimilative (DS) inorganic compounds used as electron acceptors in energy metabolism. A large amount of electron acceptors is reduced, and the reduced product is secreted.
[b]Chemical process: oxidation (O): reduction (R), degradation (DE), biosynthesis (B), methanogenesis (M), disproportionation (D).
[c]Compound type: organic (O): sulfur atoms with covalent bonds to carbon atoms. Inorganic (I): sulfur compounds with non-carbon atoms.
[d]Source: sulfur compound used as source of energy (E), sulfur (S), carbon (C), nitrogen (N).
[e]Number of Pfam domains belonging to each metabolic pathway described in Sucy (Table S2).

interconnection of pathways in terms of energy flow and the interplay of the redox gradient (organic/inorganic) of the intermediate compounds that act as key axes of organic and inorganic reactions (e.g., sulfite).

## Annotation of Pfam domains within sulfur proteins

Our approach requires the detection of structural and evolutionary units, also known as domains, in the curated list of protein sequences involved in the metabolism of interest (S-cycle in this case). The annotation of protein domains against the Pfam-A database resulted in a total of 112 domains identified in 147 proteins (out of 152). These 112 domains constitute the Pfam-Sucy database and represent all the pathways listed in Table 1. Two other protein family databases were tested (TGRFAM and Superfamily), but the number of proteins with positive matches was lower than with Pfam (57 and 137, respectively), and thus the other protein family databases were not further considered.

## Preparation of omic datasets: Gen, GenF, and Met

The genomic dataset required for computing domain entropies (Gen) was obtained from public databases, as explained above in "MEBS Description." A fragmented version of Gen, called GenF, was generated by considering the MSL distribution of metagenomic sequences (Fig. S1).

In order to benchmark MEBS with real environmental metagenomic samples, a collection of 900 public metagenomes was obtained from MG-RAST, to which we added 35 metagenomes sampled from an ultra-oligotrophic shallow lake in México (CCC). Altogether, these 935 metagenomes set up the Met dataset.

## Using the relative entropy to recognize S-cycle domains and candidate markers

The next stage consists of the quantitative detection of informative domains (enriched among organisms in Suli) by computing their relative entropy ($H'$) using Equation (1). The occurrences of each of the 112 Pfam domains in Suli and the genomic datasets were taken as observed and expected frequencies, respectively. Fig. 4A summarizes the computed $H'$ values in real (Gen) and fragmented genomic sequences of increasing size (GenF). The results indicate that only a few Pfam domains are equally informative, regardless of the length of sequences. When $H'$ values inferred from real, full-length proteins are compared to those of fragmented sequences, it can be seen that shorter sequences (MSL 30 and 60 aa) yield larger entropy differences than sequences of length >100 aa (see in Fig. 4B). Therefore, in order

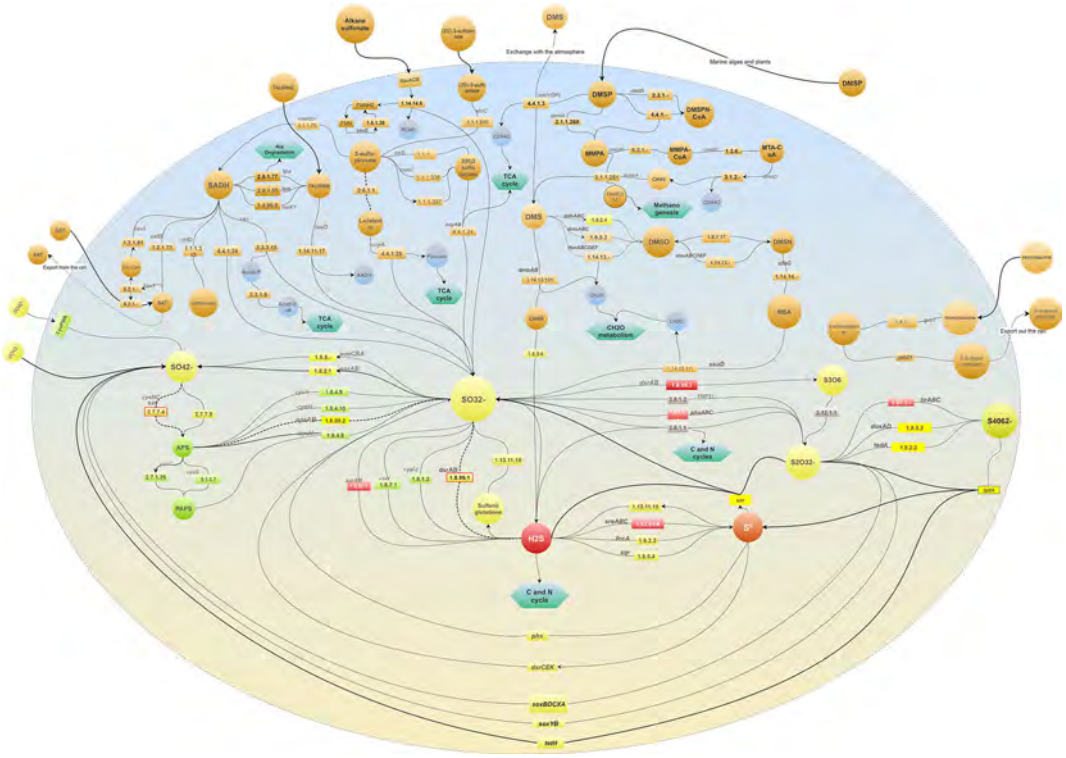



**Figure 3:** Comprehensive network representation of the machinery of the biogeochemical S-cycle in a single cell. The 28 molecular pathways involved in the metabolism of sulfur compounds described in Table 1 are included. The enzymatic steps are depicted as rectangles, followed by arrows indicating the direction of the reaction. Green hexagons represent metabolic links to other metabolisms. Bold dashed arrows indicate bidirectional reactions. Inorganic S-compounds have been arranged according to their reduction potential, from the most oxidized (yellow) to the most reduced (red) compounds. Gray rectangles indicate enzymes acting in disproportionation processes in which a reactant is both oxidized and reduced in the same chemical reaction, forming 2 separate compounds. Input biogeochemical S-compounds are shown outside and connected with bold arrows. Dashed arrows indicate S-compounds excreted out of the cell. The upper half of the modeled cell depicts the processes involved in the use of organic S-compounds (orange circles) found in natural environments and used as source of carbon, sulfur, and/or energy in several aerobic/anaerobic strains, described in Fig. 2.

to shortlist candidate marker genes, we selected those Pfam domains displaying constant, high mean $H'$ values in Gen and GenF, low $H'$ standard deviation (std), and a clear separation from the random distribution.

We tested several clustering methods, summarized in Fig. S3, with Ward and Birch performing best in grouping together informative protein domains with low std. However, the Ward classification was eventually selected as Birch failed to include a few Pfam domains relevant in the S-cycle (see Fig. S4). By using the Ward method, 3 well-defined clusters of Pfam domains were generated, as observed in Fig. 4C. Cluster 0 included 94 domains containing $H'$ values ranging from –0.4 to 0.4 and overlapping with the values obtained in the negative control explained in the next section. Cluster 1 consistently grouped together 12 Pfam domains, listed in Table 2, with high entropy and low std, and can therefore be proposed as molecular markers in metagenomic sequences of variable length. Among the proposed marker domains are APS-Reductase (PF12139: $H' = 1.2$), ATP-sulfurilase (PF01747: $H' = 1.03$), and DsrC (PF04358: $H' = 0.52$), key protein families in metabolic pathways involved in both sulfur oxidation/reduction processes. Finally, cluster 2 includes Pfam domains displaying high entropy values and high std, such as the PUA-like domain (PF14306: $H' = 1$). We presume that domains within this cluster are also key players in S-metabolism; however, their high std makes them unsuitable for markers, particularly with metagenomic sequences of variable MSL. We suggest that further analyses will be required to test the implication in S-energy conservation processes of proteins containing domains such as PF03916, PF02665, or PF14697 (see the complete list in Table S4).

## Is the entropy affected by the input list of microorganisms? Negative control test

In order to evaluate to what extent the $H'$ values depend on the curated list of microorganisms, we performed a negative control by replacing Suli in 1000 lists of randomly sampled genomes, and we used them to compute the observed frequencies (see Equation (1)). As expected, there was a clear difference between both $H'$ estimates (see Fig. S5). In particular, entropy values derived from the random test were found to be approximately symmetric and consistently low among the GenF size categories (compared with the real values), yielding values of –0.09 and 0.1 as 5th and 95th percentiles, respectively (Table S5).

## Sulfur Score and its predictive capacity to detect S-microbial players in a large genomic dataset

To test whether Pfam entropies can be combined to capture the S-metabolic machinery in "omic" samples, we calculated the final MEBS score, called, in this case, Sulfur Score. We computed the SS on each of the 2107 non-redundant genomes in Gen with the script *score_genomes.sh*. The individual genomes, along with their corresponding SS values and taxonomy according to NCBI, are found in Table S6.

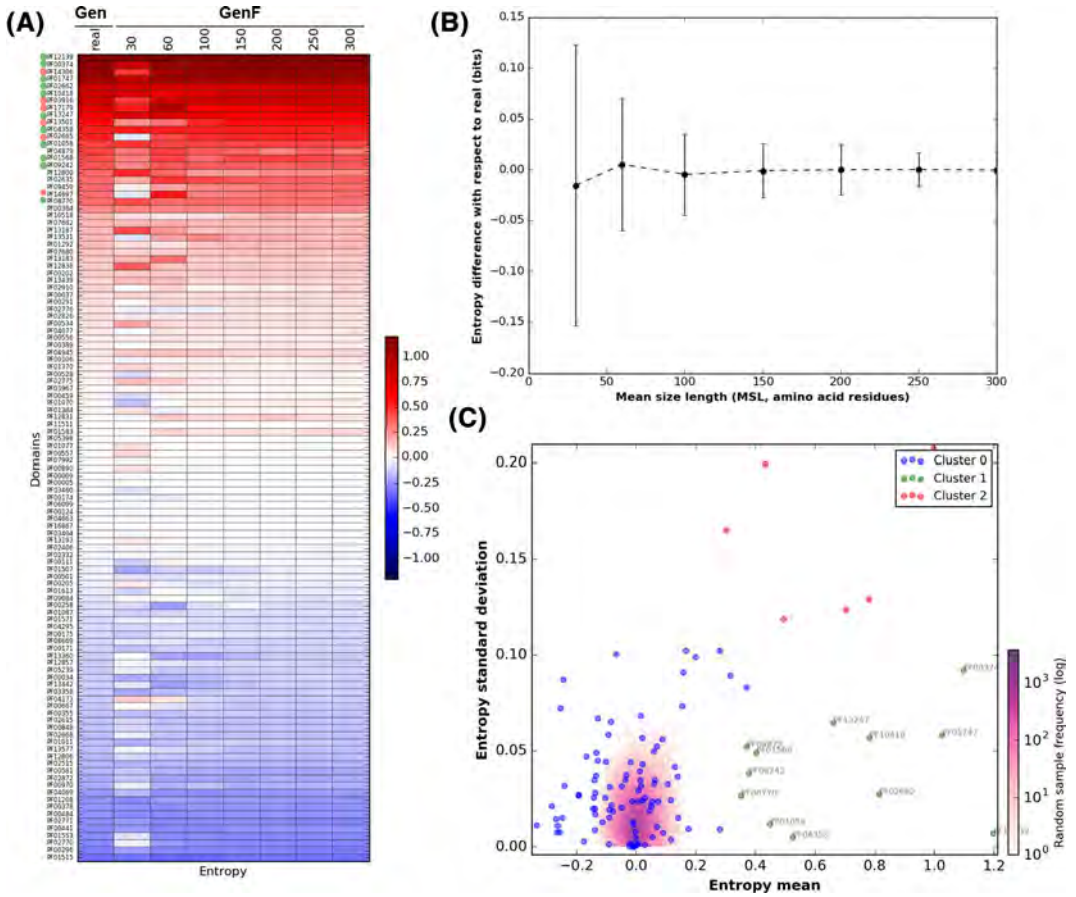



**Figure 4:** Entropy values of Sulfur-derived protein domains. **(A)** Heatmap showing the entropy values ($H'$) of the 112 Pfam domains identified in proteins curated in SuCy. **(B)** Difference between entropies estimated from sizes categories of growing peptide size (GenF) and the real values measured within complete genomes (Gen). Error bars show standard deviations. Both graphs were obtained with script *plot_entropy.py*. Clustering of the Pfam relative entropies obtained in Gen and GenF was produced with the Ward method. Log frequency of the entropy values computed in the random test is colored in purple (see scale bar). Cluster 0 (blue) groups protein domains with low relative entropy that overlap with the random distribution. Cluster 1 (green) includes the Pfam domains that fulfill the requirements to be used as molecular markers (high $H'$ and low standard deviation [std]). Red dots (cluster 2) correspond to Pfam domains with high $H'$ and std. The cluster was produced with script *F_meanVSstd.py*.

For evaluation purposes, we classified and manually annotated all the genomes in Gen according to their metabolic capabilities. First, we identified the 161 curated genomes belonging to Suli. Then, we focused on the remaining genomes. A set of 192 genomes with $SS > 4$ were labeled Sulfur unconsidered or related microorganisms (Sur). Finally, the rest of genomes in Gen were classified as NS (Non-Sulfur = Gen–(Suli + Sur)), including 1754 genomes. The boxplots in Fig. 5A summarize the scores obtained in these 3 subsets.

To double-check whether the Sur genomes—selected due to their *SS*—might be involved in the S-cycle, we manually annotated all of them, focusing on relevant genomic, biochemical, physiological, and environmental information that we might have missed since Suli was first curated (Table S7). Out of 192 genomes, 68 are reported to metabolize S-compounds under culture conditions in the literature. For instance, *Sideroxydans lithotrophicus ES-1*, a microaerophilic Fe-oxidizing bacterium, has been observed to also grow in thiosulfate as an energy source [56]. Another 59 Sur organisms have been isolated from Sulfur-rich environments, such as hot springs or solfataric muds. Remarkably, some of this species include hard-to culture genomes reconstructed from metagenomic sequences such as *Candidatus Desulforudis audaxviator MP104C* isolated from basalt-hosted fluids of the deep subsea floor [6], an unnamed endosymbiont of a scaly snail from a black smoker chimney [57], and archaeon *Geoglobus ahangari*, sampled from a 2000-meter-deep hydrothermal vent [58]. Furthermore, we also confirmed within Sur the implication of S-cycle of 20 species of the genus *Campylobacter*. These results are consistent with the ecological role of the involved taxa, that along with SRB and methanogens inhabiting host-gastrointestinal and low-oxygen environments, where several inorganic (e.g., sulfates, sulfites) or organic (e.g., dietary amino acids and host mucins) are highly metabolized by these metabolic guilds [59]. The implication of *Campylobacter* species in the S-cycle is also supported by the fact that some of them have been isolated from deep sea hydrothermal vents [60]. The remaining species in Sur were classified into different categories, including bioremediation (7), Fe-environment (2), marine (2), peat lands (2) and other environments (32) (see Fig. 5B).

When the *SS* values of genomes in Sur are compared to the S-metabolic guilds represented in Suli (e.g., PSB, SRB, GSB), it can be seen that they are indeed similar and clearly separated from the rest of the NS genomes (Fig. 5C). This strongly suggests that high-scoring genomes are indeed ecologically and metabolically implicated in the S-cycle.

Finally, in order to quantify the capacity of the *SS* to accurately classify S-related microorganisms, we computed a ROC

**Table 2:** Informative Pfam domains with high $H'$ and low std; novel proposed molecular marker domains in metagenomic data of variable MSL

| Pfam ID (Suli ocurrences) | $H'$ mean | $H'$ std | Description |
|---|---|---|---|
| PF12139 58/161 | 1.2 | 0.01 | Adenosine-5′-phosphosulfate reductase beta subunit: key protein domain for both sulfur oxidation/reduction metabolic pathways. Has been widely studied in the dissimilatory sulfate reduction metabolism. In all recognized sulfate-reducing prokaryotes, the dissimilatory process is mediated by 3 key enzymes: Sat, Apr, and Dsr. Homologous proteins are also present in the anoxygenic photolithotrophic and chemolithotrophic sulfur-oxidizing bacteria (CLSB, PSB, GSB), in different cluster organizations [35]. |
| PF00374 135/161 | 1.1 | 0.09 | Nickel-dependent hydrogenase: hydrogenases with S-cluster and selenium containing Cys-x-x-Cys motifs involved in the binding of nickel. Among the homologues of this hydrogenase domain is the alpha subunit of the sulfhydrogenase I complex of *Pyrococcus furiosus* that catalyzes the reduction of polysulfide to hydrogen sulfide with NADPH as the electron donor [69]. |
| PF01747 103/161 | 1.03 | 0.06 | ATP-sulfurylase: key protein domain for both sulfur oxidation and reduction processes. The enzyme catalyzes the transfer of the adenylyl group from ATP to inorganic sulfate, producing adenosine 5′-phosphosulfate (APS) and pyrophosphate, or the reverse reaction [70]. |
| PF02662 62/161 | 0.82 | 0.03 | Methyl-viologen-reducing hydrogenase, delta subunit: is 1 of the enzymes involved in methanogenesis and encoded in the mth-flp-mvh-mrt cluster of methane genes in *Methanothermobacter thermautotrophicus*. No specific functions have been assigned to the delta subunit [49]. |
| PF10418 122/161 | 0.78 | 0.06 | Iron-sulfur cluster binding domain of dihydroorotate dehydrogenase B: among the homologous genes in this family are *asrA* and *asrB* from *Salmonella enterica enterica serovar Typhimurium*, which encode (1) a dissimilatory sulfite reductase, (2) a gamma subunit of the sulfhydrogenase I complex of *Pyrococcus furiosus*, and (3) a gamma subunit of the sulfhydrogenase II complex of the same organism [12]. |
| PF13247 149/161 | 0.66 | 0.06 | 4Fe-4S dicluster domain: Homologues of this family include: (1) DsrO, a ferredoxin-like protein, related to the electron transfer subunits of respiratory enzymes, (2) dimethylsulfide dehydrogenase $\beta$ subunit (ddhB), involved in dimethyl sulfide degradation in *Rhodovulum sulfidophilum*, and (3) sulfur reductase FeS subunit (sreB) of *Acidianus ambivalens*, involved in sulfur reduction using $H_2$ or organic substrates as electron donors [12]. |
| PF04358 73/161 | 0.52 | 0 | DsrC like protein: DsrC is present in all organisms encoding a dsrAB sulfite reductase (sulfate/sulfite reducers or sulfur oxidizers). The physiological studies suggest that sulfate reduction rates are determined by cellular levels of this protein. The dissimilatory sulfate reduction couples the 4-electron reduction of the DsrC trisulfide to energy conservation [71]. DsrC was initially described as a subunit of DsrAB, forming a tight complex; however, it is not a subunit, but rather a protein with which DsrAB interacts. DsrC is involved in sulfur-transfer reactions; there is a disulfide bond between the 2 DsrC cysteines as a redox-active center in the sulfite reduction pathway. Moreover, DsrC is among the most highly expressed sulfur energy metabolism genes in isolated organisms and meta-transcriptomes [71]. |
| PF01058 158/161 | 0.45 | 0.01 | NADH ubiquinone oxidoreductase, 20 Kd subunit: homologous genes are found in the delta subunits of both sulfhydrogenase complexes of *Pyrococcus furiosus* [12]. |
| PF01568 156/161 | 0.4 | 0.05 | Molydopterin dinucleotide binding domain: this domain corresponds to the C-terminal domain IV in dimethyl sulfoxide (DMSO) reductase [49]. |
| PF09242 39/161 | 0.38 | 0.04 | Flavocytochrome c sulphide dehydrogenase, flavin-binding: enzymes found in S-oxidizing bacteria such as the purple phototrophic bacteria *Chromatium vinosum* [49]. |
| PF04879 151/161 | 0.37 | 0.05 | Molybdopterin oxidoreductase Fe4S4 domain: is found in a number of reductase/dehydrogenase families, which include the periplasmic nitrate reductase precursor and the formate dehydrogenase alpha chain, i.e., *Wolinella succinogenes* polysulfide reductase chain, *Salmonella typhimurium* thiosulfate reductase (gene phsA). |
| PF08770 45/161 | 0.35 | 0.03 | Sulphur oxidation protein SoxZ: SoxZ sulfur compound chelating protein, part of the complex known as the Sox enzyme system (for sulfur oxidation) that is able to oxidize thiosulfate to sulfate with no intermediates in *Paracoccus parantropus* [12]. |

curve (for a detailed description of ROC curves, see [61]). We thus defined genomes annotated in Suli as positive instances, and the rest as negative ones. The results are shown in Fig. 5D, with an estimated area under the curve (AUC) of 0.985, and the corresponding cut-off values of *SS* for several false positive rates (FPR). According to this test, an *SS* value of 8.705 is required to rule out all false positives in Gen, while $SS = 5.231$ is sufficient to achieve an FPR $< 0.05$.

Overall, these results indicate that MEBS is a powerful and broadly applicable approach to predict and classify microorganisms closely involved in the sulfur cycle, even in hard-to-culture microbial lineages.

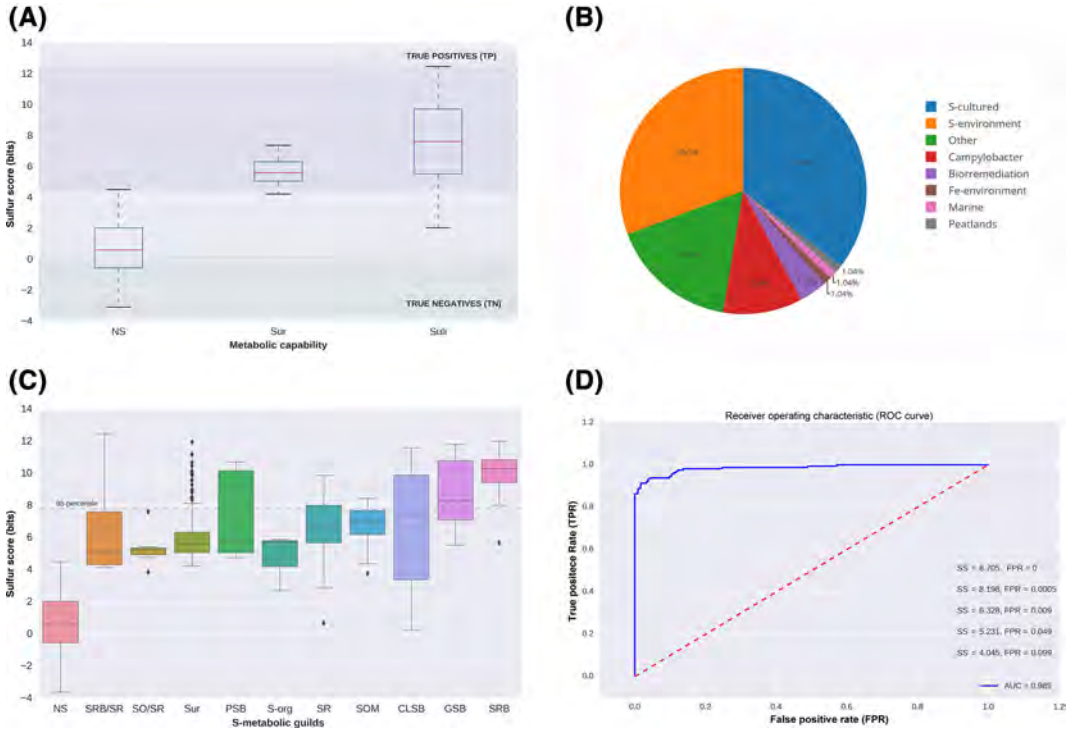



**Figure 5:** Distribution of Sulfur Score (*SS*) in 2107 non-redundant genomes (Gen). **(A)** Subsets of genomes annotated in Suli (n = 161). **(ii)** Sur, genomes not listed in Suli with *SS* > 4 and candidates to be S-related microorganisms (n = 192). **(iii)** Rest of the genomes in Gen (NS, n = 1754). According to the curated species, true positives can be defined as genomes with *SS* > max ($SS_{NS}$) distribution, whereas true negatives are those with *SS* < min ($SS_{Suli}$). **(B)** Assignment of the 192 genomes in Sur to ecological categories based on literature reports. **(C)** Distribution of *SS* for different S-metabolic guilds, and the genomes in Sur. **(D)** ROC curve with area under the curve (AUC) indicated together, with thresholds for some false positive rates (FPR).

## Sulfur Score and its predictive capacity to detect S-related environments in a large metagenomic collection

The *SS* was also computed for each metagenome in Met, using their corresponding MSL to choose the appropriate entropies previously calculated in dataset GenF (Table S8). In order to test whether *SS* values can be used to identify S-related environments, we performed the following analyses. First, we used the geographical metadata associated with each metagenome to map the global distribution of *SS*. In Fig. 6A, *SS* values are colored from yellow to red. The most informative S-environments (displaying *SS* values equal to or greater than the 95th percentile of each MSL category) are shown in blue.

Then, we sorted the metagenomes according to their environmental features, as proposed by the Genomic Standards Consortium (GSC) and implemented in MG-RAST. Each feature corresponds to 1 of 13 environmental packages (EPs) that standardize metadata describing particular habitats that are applicable across all GSC checklists and beyond [62]. Therefore, each EP represents a broad and general classification containing particular features. For example, the "water" EP includes 330 metagenomes from our dataset, belonging to several features such as freshwater, lakes, estuarine, marine, hydrothermal vents, etc. As each of these features has different ecological capabilities in terms of biogeochemical cycles, we can expect different behaviors among *SS* values, as shown in Fig. 6B. In general, all the metagenomes derived from hydrothermal vents (2), marine benthic (6), intertidal (8), and our unpublished CCC microbial mats had *SS* values above the 95th percentile, highlighting the importance of the S-cycle in these environments. In contrast, the metagenomes belonging to features such as sub-terrestrial habitat (7), saline evaporation pond (24), or organisms-associated habitat (7) displayed consistently low or even negative *SS* values, indicating a negligible presence of S-metabolic pathways in those environments. The remaining features have intermediate median *SS* values and contain occasionally individual metagenomes with *SS* values above the 95th percentile, such as freshwater, marine, ocean, or biofilm environments.

To validate the list of 50 high-scoring metagenomes (above the 95th percentile), we double-checked their annotations. According to the literature and associated metadata, all these environments are closely involved in the mineralization, uptake, and recycling processes of S-compounds, e.g., environmental sequences derived from the coastal Oligochaete worm *Olavius algarvensis*, hydrothermal vents, and marine deep-sea surface sediments around the Deep-Water Horizon spill in the Gulf of Mexico. The complete list of annotated metagenomes, along with their ecological capabilities, is found in Table S9.

## Evaluating the robustness of the sulfur score

To test the reproducibility and robustness of the MEBS final score (*SS*), we conducted 2 further analyses. In the first, we compared *SS* estimates derived from the Met dataset, computed with Pfam entropies obtained in the first MEBS benchmark performed 3 years ago (2014) with the current data described in this article (2017). Despite the changes of both databases (Pfam database version and the Suli list), we found a strong correlation ($r^2$ = 0.912) between the *SS* outcomes (Fig. S6 A). A kernel density analysis of the latter comparison suggests a different behavior of low

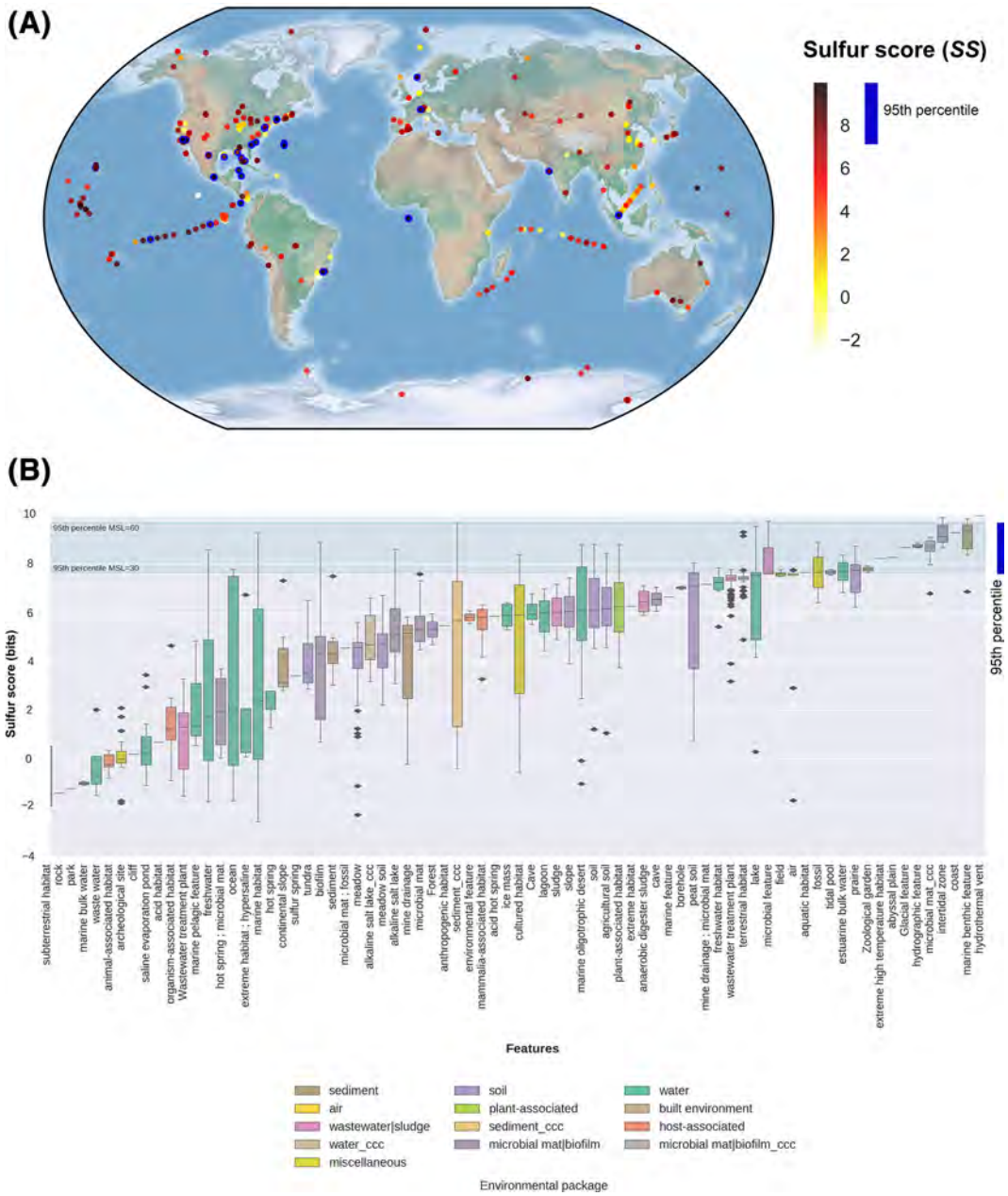



**Figure 6:** Distribution of Sulfur Score (*SS*) in the metagenomic dataset Met. **(A)** Geo-localized metagenomes sampled around the globe are colored according to their *SS* values. The following cut-off values correspond to the 95th percentiles of 7 mean size length classes (30, 60, 100, 150, 200, 250, and 300 aa): 7.66, 9.70, 8.81, 8.51, 8.18, 8.98, and 7.61, respectively. Circles with a thick blue border indicate metagenomes with $SS \geq$ the 95th percentile. **(B)** Distribution of *SS* values observed in 935 metagenomes classified in terms of features (x-axis) and colored according to their particular habitats. Features are sorted according to their median *SS* values. ccc: metagenomes from Cuatro Cienegas, Coahuila, Mexico. Green lines indicate the lowest and largest 95th percentiles observed across MSL classes.

and high *SS* scores, with the latter being more reproducible (see Fig. S6B).

In the second analysis, we quantitatively tested to what extent the entropy estimates of the 112 Pfam domains directly affect the outcome of the *SS* in Gen and Met. We randomly subsampled ≈50% of those domains to compute the *SS* a thousand times for each genome and metagenome in Gen and Met, respectively. The results, summarized in Table S10, confirm that *SS* values computed with random subsets of Pfam domains are generally lower than *SS* derived from the full list (n = 112) of Sucy-Pfam domains. To further inspect the distribution of *SS* values produced with random subsets of domains (random *SS*), we focused on the particular case of the metagenomes belonging to the category MSL = 60. As expected, the distribution of random *SS* oscillates between negative and positive values. Interestingly, metagenomes exhibiting only positive random *SS* are ranked above the 95th percentile according to their real *SS* values (see Fig. S7A). The latter indicates that even a random subset of Pfam domains used to compute the score is more likely to highly rank metagenomes containing the sulfur metabolic machinery (large number of high-entropy Pfam domains), than those lacking the sulfur metabolism or displaying a large number of non-informative Pfam domains. Furthermore, by comparing the median of random *SS* with the real scores, we observe

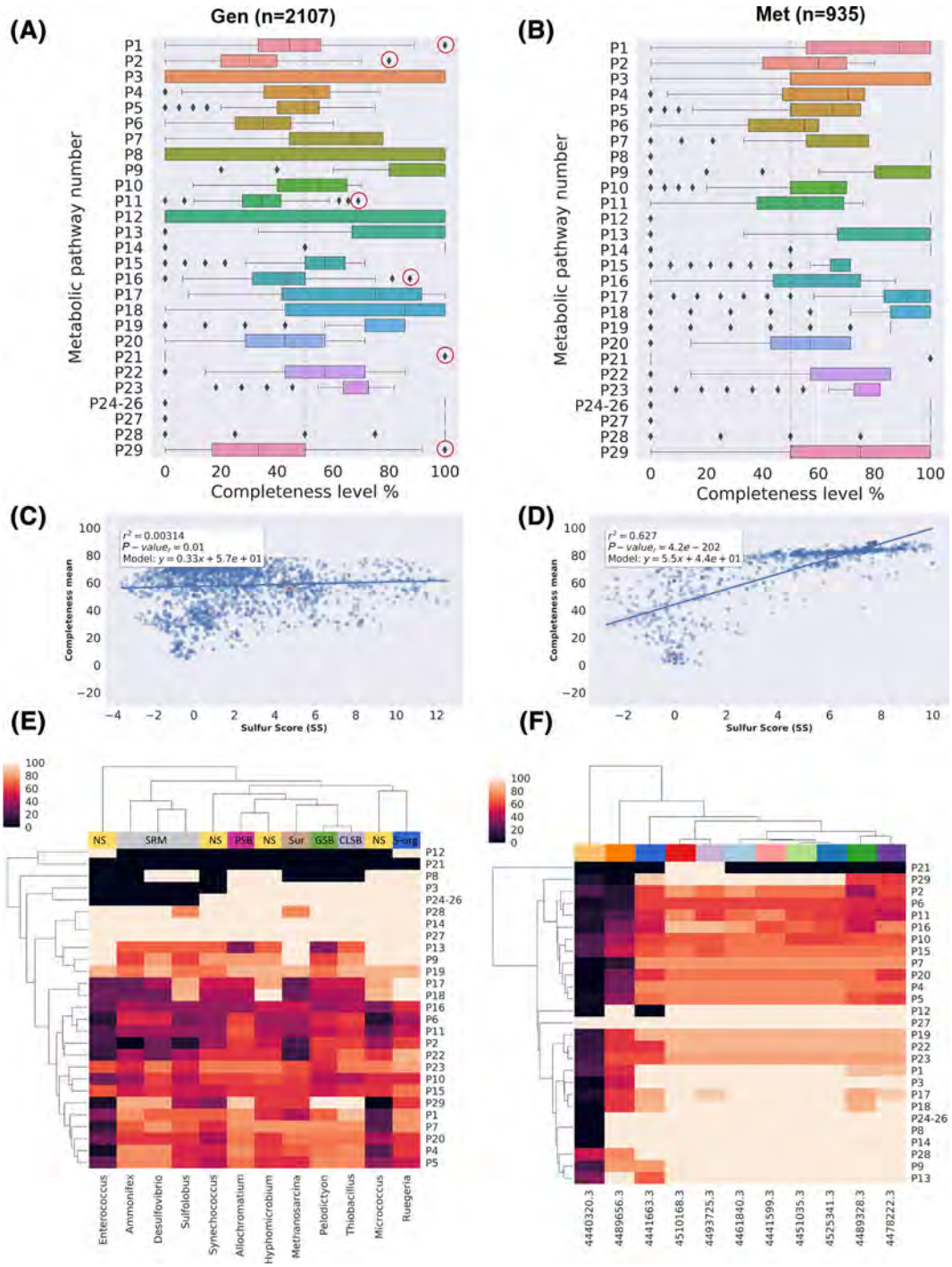



**Figure 7:** Metabolic completeness of the metabolic pathways described in Table 1. **(A)** Boxplot distribution of the pathway completeness in genomic and **(B)** metagenomic datasets. **(C)** Linear regression models of the Sulfur Score (SS) and the mean completeness in Gen and **(D)** Met dataset. **(E)** Heatmap showing the metabolic completeness of the following genomes: *Desulfovibrio vulgaris DP4* (SS = 11.442), *Ammonifex degensii KC4* (SS = 12.508); *Pelodictyon phaeoclathratiforme BU-1* (SS = 11.836); *Thiobacillus denitrificans* ATCC 25 259 (SS = 11.61); PSB: *Allochromatium vinosum DSM 180* (SS = 10.737); SUR: *Sur Methanosarcina barkeri* MS (SS = 5.93); *Sulfolobus acidocaldarius DSM 639* (SS = 5.457); *Synechococcus* sp. *JA-2–3Ba 2–13* (SS = 3.704); *Hyphomicrobium denitrificans 1NES1* (SS = 3.236); *Ruegeria pomeroyi DSS-3* (SS = 2.707); *Enterococcus durans* (SS = −0.194); *Micrococcus luteus NCTC_2665* (SS = −3.588). **(F)** Heatmap showing the metabolic completeness of the metagenomes with the following MG-RAST IDs and corresponding scores: 4489656.3 (SS = −2.649); 4440320.3 (SS = 0.1); 4441663.3 (SS = 9.986); 4510168.3 (SS = 7.781); 4493725.3 (SS = 9.547); 4461840.3 (SS = 8.813); 4441599.3 (SS = 9.274); 4451035.3 (SS = 9.918); 4525341.3 (SS = 9.287); 4489328.3 (SS = 4.958); 4478222.3 (SS = 4.88). The color codes at the top of the heatmap correspond to different environments. For a more detailed description of each metagenome, see Table S8.

a clear separation between those distributions (see Fig. S7B and Table S10).

## Completeness of S-metabolic pathways

As we described above, the MEBS pipeline models a metabolic network as an array of S-related protein domains (Sucy-Pfam) to ultimately use their entropies to produce the final score (*SS*). For a closer look, we also dissected the total contribution of independent domains at the network level in order to assess whether *SS* depends on the partial or complete detection of S-pathways. Consequently, we evaluated the pathway completeness in both genomic (Gen) and metagenomic (Met) datasets (see Tables S11 and S12, respectively). Since the number of Pfam domains per pathway goes from 1 to 29 (see Table 1 and Table S2), we suspect that pathways represented by a single domain might not reflect their complete metabolic function. For example, the pathways involved in the methanogenesis of compounds such as dimethylsulfide (DMS, P24), methyl-thiolpropanoate (MTPA, P25), and methanethiol (MeSH, P26) are represented by the same protein (MtsA, PF01208) in our Sucy database, as well as in Metacyc [12]. Therefore, we expect that pathways P24–26 will have identical presence-absence patterns in Gen and Met.

The boxplots in Fig. 7A and B summarize the distribution of completeness for each S-metabolic pathway, including the synthetic pathway (P29) composed by 12 candidate markers as described in Table 2. As expected, the observed completeness per pathway was higher in Met than in Gen, since microbial communities harbor a wider repertory of metabolic functions than single genomes. In the case of genomes, we noted that a few pathways were complete in most genomes, the majority being involved in the usage of organic sulfur compounds such as alkanesulfonates (P9), sulfoacetate (P14), and biosynthesis of sulfolipids (SQDG) and the single domain pathways P24–26. Remarkably, we also detected a few organisms displaying the highest levels of metabolic completeness in some S-energy based pathways. For example, we found that *Desulfosporosinus acidiphilus SJ4* (*SS* = 8.91) was the only genome harboring the complete repertory of Pfam domains described in Sucy for the sulfite oxidation (P1), strongly suggesting that it may oxidize sulfite. However, this activity remains to be tested in culture [63]. In the case of thiosulfate oxidation (P3), we detected 3 genomes displaying the highest levels of completeness, in agreement with their ecological features: *Hydrogenobaculum* sp. *Y04AAS1* (*SS* = 9.319) [64] and the CLSB: *Acidithiobacillus caldus ATCC 51756* (*SS* = 6.525) [65] and *Acidithiobacillus ferrivorans* (*SS* = 7.436) [66]. For the sulfate reduction dissimilative pathway (P5), out of 55 genomes displaying the higher completeness levels, 67% are actually SRB, 12% are Sur genomes, and the rest are sulfur oxidation microorganisms. Furthermore, the PSB *Thioflavicoccus mobilis 8321* (*SS* = 9.756), isolated from a microbial mat [67], was the genome displaying the most complete sulfide oxidation pathway (P11). Elemental sulfur disproportionation (P21) is represented by a single non-informative domain (PF07682, $H' = 0.172$) that remarkably is found in 14 sulfur respiring or related genomes such as *Sulfolobus tokodaii str. 7* (*SS* = 5.341) and *Acidianus hospitalis W1* (*SS* = 3.88). Finally, we identified 6 genomes encoding all 12 proposed markers. Among them, 3 were GSB (*Pelodictyon phaeoclathratiforme BU-1*, *SS* = 11.836, *Chlorobium chlorochromatii CaD3*, *SS* = 11.625, and *Chlorobium tepidum TLS*, *SS* = 11.354), 1 CLSB (Thiobacillus denitrificans ATCC 25259 *SS* = 11.61), another 1 PSB (*Thiocystis violascens DSM 198*, *SS* = 10.633), and finally 1 Sur (*Sedimenticola thiotaurini SS* = 10.109). For a complete description, see Table S13.

A global view of metabolic completeness was obtained by bulking the data from all pathways. Linear regression models between mean completeness and *SS* that were computed confirm the yielding $r^2$ values of 0.003 and 0.627 for Gen and Met, respectively (See Fig. 7C and D). Moreover, we also assessed the relationship between the mean completeness of the synthetic pathway of candidate markers (P29) and the *SS*. As expected, significant correlations were obtained in both datasets ($r^2 = 0.645$ and $r^2 = 0.881$ for Gen and Met, respectively) (see Fig. S8).

To get a more detailed insight into the completeness, we selected a few genomes and metagenomes displaying high and low *SS* values. Specifically, from the Gen dataset, we selected 1 representative from the main S-guilds, 1 Sur genome, and 2 genomes with low *SS* values (NS). As observed in Fig. 7, the low-scoring genomes *Enterococcus durans* (*SS* = −0.194), *Micrococcus luteus NCTC_2665* (*SS* = −3.588), *and Ruegeria pomeroyi DSS-3* (*SS* = 2.707) display unrelated patterns of sulfur metabolic completeness compared with the rest of the genomes and therefore are separated. In contrast, high-scoring S-respiring microorganisms *Desulfovibrio vulgaris DP4* (*SS* = 11.442), *Sulfolobus acidocaldarius DSM 639* (*SS* = 5.457), and *Ammonifex degensii KC4* (*SS* = 12.508) are clustered together. We also observed that mat-isolated cyanobacteria *Synechococcus* sp. *JA-2–3Ba 2–13*, classified as NS with *SS* = 3.704, was clustered together with other high-scoring genomes, in agreement with the lack of correlation reported above.

In the case of metagenomes (see Fig. 7E), we observed a clear correlation between *SS* and completeness. For example, metagenomes 4440320.3 and 4489656.3, with the lowest scores (*SS* = 0.1 and *SS* = −2.649, respectively), also exhibit the largest number of incomplete pathways. Similarly, high-scoring metagenomes derived from black smoker or marine sediment are grouped together in terms of completeness.

## Conclusions

Our study represents the first exploration of the Sulfur biogeochemical cycle in a large collection of genomes and metagenomes. The manually curated effort resulted in an inventory of the compounds, genes, proteins, molecular pathways, and microorganisms involved. This complex universe of articulated data was reduced to a list of microorganisms and Pfam domains encoded in the proteins that take part in that network. These domains were first ranked in terms of relative entropy, and then summed to produce a single S-score representing the relevance of a given genomic or metagenomic sample in terms of sulfur metabolic machinery. We took advantage of the mathematical framework of information theory, which has been widely used in computational biology.

The performance of the MEBS pipeline (designed for the above-mentioned tasks) was benchmarked on large genomic and metagenomic sets. Our results support the broad applicability of this algorithm in order to classify annotated genomes as well as newly sequenced environmental samples without prior culture. We also assessed to what extent the final score depended on the partial or complete detection of pathways and observed a higher completeness per pathway in metagenomic sequences than in individual genomes.

We demonstrated that a measurable score can be applied to evaluate any given metabolic machinery or biogeochemical cycle in large (meta)genomic scale, holding the potential to dramatically change the current view of inferring metabolic capabilities in the present “omic” era.

## Availability and requirements

Project name: MEBS

Project home page: https://github.com/eead-csic-compbio/metagenome_Pfam_score

Operating system(s): Linux

Programming language: Python 3, Perl5, Bash,

Other requirements: HMMER

License: GNU General Public License (GPL)

## Availability of supporting data

The datasets supporting the results of this article and snapshots of the MEBS code in GitHub are available in the *GigaDB* repository [68].

## Supplementary files

The supplementary pdf file contains the following information:

Supplementary figure S1: Histogram distribution of the mean size length of metagenomes in Met and the input sulfur proteins.

Supplementary figure S2: Visualization of the Pfam domains mapped onto KEGG metabolic pathways.

Supplementary figure S3: Comparison of clustering methods of the 112 Pfam entropies using script *plot_cluster_comparison.py*.

Supplementary figure S4: Clustering comparison between Birch and Ward clustering methods to stand out the Pfam entropies with high $H'$ and low std using the script.

Supplementary figure S5: Distribution of entropy values of 112 Pfam domains inferred from random-sampled and Suli genomes.

Supplementary figure S6: Comparison of Sulfur Scores (*SS*) with data obtained 3 years ago (2014), with the current data described in the article.

Supplementary table S4: Informative Pfam's with high $H'$ and high std (not used as molecular marker genes) in metagenomic fragmented data.

Supplementary table S5: Percentile distribution of the 112 Pfam entropies in the random test.

Supplementary table S10: Statistics of *SS* computed on genomic (Gen, real sequences) and metagenomic (Met, with increasing mean size length, from 30 to 300 aa) datasets.

In separated Excel files, the following supplementary tables are also provided:

Supplementary table S1: Table S1. Comprehensive list of the taxonomic representatives of sulfur cycle including Sulfur list, or "Suli," containing 161 curated genomes used as input for the pipeline.

Supplementary table S2: Sucy database containing the identifiers of the Sulfur proteins and their corresponding annotations derived from Interproscan and manual curation.

Supplementary table S3: Sulfur Pfam domains (Pfam-Sucy), and their corresponding mapping into KEGG (KO number), and the manual assignation into sulfur metabolic pathways.

Supplementary table S6: Gen dataset containing their corresponding *SS* and taxonomy assignment.

Supplementary table S7: Manual annotation of Sulfur uncosidered or related microorganisms (Sur) with Sulfur Score (SS) greater or equal to four.

Supplementary table S8: Met dataset with corresponding *SS* values and metadata.

Supplementary table S9: Manually annotated high-scoring metagenomes along with their ecological capabilities in terms of sulfur cycle.

Supplementary table S11: Metabolic completeness in Gen dataset for each of the 28 metabolic pathways of the S-cycle described in Table 1. (Pathway 29 contains the proposed marker genes.)

Supplementary table S12: Metabolic completeness in Men dataset for each of the 28 metabolic pathways of the S-cycle described in Table 1. (Pathway 29 contains the proposed marker genes.)

Supplementary table S13: Frequency and description of the most complete genomes in terms of S-cycle metabolic pathways.

## Abbreviations

AUC: area under the curve; CCC: Cuatro Cienegas, Coahuila; CLSB: color-less sulfur bacteria; EP: environmental packages; ESR: elemental sulfur-reducing microorganisms; FPR: false positive rate; Gen: genomic dataset; GenF: genomic fragmented dataset; GSB: green sulfur bacteria; GSC: Genomic Standards Consortium; HMM: Hidden Markov Models; MEBS: Multigenomic Entropy Based Score; Met: metagenomic dataset; MSL: mean size length, $H'$: relative entropy; NS: non-sulfur-related genomes; PSB: purple sulfur bacteria; Rlist: random list of taxonomic representatives; ROC: receiver operating characteristic; S: sulfur; S-cycle: sulfur cycle; SOM: sulfur-oxidizing microorganims; SRB: sulfate-reducing bacteria; *SS*: Sulfur Score; Sucy: Sulfur Cycle Database; Suli: Sulfur list; Sur: Sulfur unconsidered; TPR: true positive rate.

## Funding

Valerie De Anda is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), and received fellowship 356 832 from Consejo Nacional de Ciencia y Tecnología (CONACYT). This research was also supported by funding from World Wildlife Fund (WWF)-Alianza Carlos Slim, Sep-Ciencia Básica Conacyt grant 238 245 to both Valeria Souza and Luis Enrique Eguiarte and Spanish MINECO grant CSIC13–4E-2490. Bruno Contreras Moreira was funded by Fundación ARAID. The sabbatical leaves of Luis Enrique Eguiarte and Valeria Souza at the University of Minnesota were supported by scholarships from Programa de Apoyos para la Superación del Personal Académico de la UNAM (PASPA), Dirección General de Asuntos del Personal Académico (DGAPA), UNAM.

## Competing interest

The authors declare that they have no competing interests.

## Author contribution

V.D.A., B.C.M., and I.Z.P. wrote the paper. B.C.M., V.D.A., and A.C.P.H. developed and wrote the software and performed all the bioinformatics analyses. V.D.A. produced all the figures and wrote the documentation of the software. V.D.A. and I.Z.P. conceived the manual curation of the Sulfur Cycle Inventory and the microbiological, biogeochemical, and ecological interpretation. L.E. and V.S. provided the intellectual framework, expertise, and resources to develop and supervise the project. All the authors read and approved the final manuscript.

## Endnotes

We are currently finishing the analyses to demonstrate the applicability of this approach to other biogeochemical cycles (C, N, O, Fe, P). Thereby, we hope that the pipeline MEBS will facilitate analysis of biogeochemical cycles or complex metabolic networks carried out by specific prokaryotic guilds, such as bioremediation processes (i.e., degradation of hydrocarbons, toxic aromatic compounds, heavy metals, etc.). We look forward to collaborating with and helping other researchers by integrating comprehensive databases that might be helpful to the scientific community. Furthermore, we are currently working to improve the algorithm by using only a list of sequenced genomes involved in the metabolism of interest in order to reduce the manual curation effort. We are also considering taking *k-mers* instead of peptide Hidden Markov Models to increase the speed of the pipeline. We anticipate that our platform will stimulate interest and involvement among the scientific community to explore uncultured genomes derived from large metagenomic sequences.

## Acknowledgments

The authors gratefully acknowledge Emilio Morella for his valuable support, feedback, and comments throughout the development of this project. We really appreciate the reviewers Daan Speth and Thulani Makhalanyane, whose valuable comments and suggestions improved the manuscript and software. Our special thanks to Carlos P. Cantalapiedra, Seth Barribeau, Will Levitt, and anonymous reviewers for their comments on earlier versions of the manuscript; these comments immensely improved the algorithm and the final version of the article. The authors also thank the Laboratory of Computational and Structural Biology (EEAD-CSIC) and Laboratorio de Evolución Molecular y Experimental, Instituto de Ecología, UNAM, for providing the computational resources described in the article. The paper was written during a sabbatical leave of Luis Enrique Eguiarte and Valeria Souza (VS) at the University of Minnesota in Peter Tiffin's and Michael Travisano's laboratories.

## References

1. Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp 2012;**2**(1):3.
2. Oulas A, Pavloudi C, Polymenakou P et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinf Biol Insights 2015;**9**:75–88.
3. Morales SE, Holben WE. Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? FEMS Microbiol Ecol 2011;**75**(1):2–16.
4. Hug LA, Baker BJ, Anantharaman K et al. A new view of the tree of life. Nat Microbiol 2016;**1**(5):16048.
5. Marco D. Metagenomics: Current Innovations and Future Trends. Poole, UK: Caister Academic Press; 2011.
6. Jungbluth SP, Glavina Del Rio T, Tringe SG et al. Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. PeerJ 2017;**5**:e3134.
7. Iverson V, Morris RM, Frazar CD et al. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. Science 2012;**335**(6068):587–90.
8. Sangwan N, Xia F, Gilbert JA. Recovering complete and draft population genomes from metagenome datasets. Microbiome 2016;**4**(1):8.
9. Mehrshad M, Amoozegar MA, Ghai R et al. Genome reconstruction from metagenomic data sets reveals novel microbes in the brackish waters of the Caspian Sea. Appl Environ Microbiol 2016;**82**(5):1599–612.
10. Sharon I, Banfield JF. Genomes from metagenomics. Science 2013;**342**(6162):1057–8.
11. Hiraoka S, Yang C, Iwasaki W. Metagenomics and bioinformatics in microbial ecology: current status and beyond. Microbes Environ 2016;**31**(3):204–12.
12. Caspi R, Altman T, Dreher K et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 2012;**40**(D1):D742–53.
13. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;**28**(1):27–30.
14. Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for metagenomes. PLoS Comput Biol 2011;**5**:453–60.
15. Abubucker S, Segata N, Goll J et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol 2012;**8**(6):e1002358.
16. Larsen PE, Collart FR, Field D et al. Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. Microb Inform Exp 2011;**1**(1):4.
17. Hanson NW, Konwar KM, Hawley AK et al. Metabolic pathways for the whole community. BMC Genomics 2014;**15**:619.
18. Castañeda LE, Barbosa O. Metagenomic analysis exploring taxonomic and functional diversity of soil microbial communities in Chilean vineyards and surrounding native forests. PeerJ 2017;**5**:e3098.
19. Fierer N, Leff JW, Adams BJ et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 2012;**109**:21390–5.
20. Llorens-Marès T, Yooseph S, Goll J et al. Connecting biodiversity and potential functional role in modern euxinic environments by microbial metagenomics. ISME J 2015;1–14.
21. Quaiser A, Zivanovic Y, Moreira D et al. Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. ISME J 2011;**5**:285–304.
22. Xie W, Wang F, Guo L et al. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J 2011;**5**:414–26.
23. Delmont TO, Malandain C, Prestat E et al. Metagenomic mining for microbiologists. ISME J 2011;**5**:1837–43.
24. Ganesh S, Parris DJ, Delong EF et al. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J 2014;**8**:187–211.
25. Segata N, Izard J, Waldron L et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;**12**:R60.
26. Parks DH, Tyson GW, Hugenholtz P et al. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 2014;**30**:3123–4.
27. Sohn MB, Du R, An L. A robust approach for identifying differentially abundant features in metagenomic samples. Bioinformatics 2015;**31**:2269–75.
28. Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat 1951;**22**:79–86.
29. Commenges D. Information theory and statistics: an overview. 2015;1–22.
30. Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 1999;**15**:563–77.

31. Dar SA, Yao L, Van Dongen U et al. Analysis of diversity and activity of sulfate-reducing bacterial communities in sulfidogenic bioreactors using 16S rRNA and dsrB genes as molecular markers. Appl Environ Microbiol 2007;**73**:594–604.
32. Perez-Jimenez JR, Kerkhof LJ. Phylogeography of sulfate-reducing bacteria among disturbed sediments, disclosed by analysis of the dissimulatory sulfite reductase genes (dsrAB). Appl Environ Microbiol 2005;**71**:1004–11.
33. Loy A, Duller S, Baranyi C et al. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environ Microbiol 2009;**11**:289–99.
34. Hügler M, Gärtner A, Imhoff JF. Functional genes as markers for sulfur cycling and CO2 fixation in microbial communities of hydrothermal vents of the Logatchev field. FEMS Microbiol Ecol 2010;**73**:526–37.
35. Meyer B, Kuever J. Molecular analysis of the diversity of sulfate-reducing and sulfur-oxidizing prokaryotes in the environment, using aprA as functional marker gene. Appl Environ Microbiol 2007;**73**:7664–79.
36. Multigenomic Entropy Based Score pipeline (MEBS). https://github.com/eead-csic-compbio/metagenome_Pfam_score. Accessed 5 October 2017.
37. Enzyme nomenclature. http://enzyme.expasy.org/. Accessed 5 May 2016.
38. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database 2011;**2011**:bar009-.
39. Reference and representative genomes. https://www.ncbi.nlm.nih.gov/genome/browse/reference/. Accessed 21 December 2016.
40. Genome clusters. http://microbiome.wlu.ca/research/redundancy/redundancy.cgi. Accessed 21 December 2016.
41. Moreno-Hagelsieb G, Wang Z, Walsh S et al. Phylogenomic clustering for selecting non-redundant genomes for comparative genomics. Bioinformatics 2013;**29**:947–9.
42. NCBI genome assembly summary file. ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/assembly_summary_refseq.txt. Accessed 21 December 2016.
43. NCBI. NCBI FTP site. ftp://ftp.ncbi.nlm.nih.gov/genomes/all/. Accessed 21 December 2016.
44. MG-RAST. http://metagenomics.anl.gov/. Accessed 10 November 2016.
45. MG-RAST API. http://api.metagenomics.anl.gov/api.html. Accessed 15 November 2016.
46. Prakash T, Taylor TD. Functional assignment of metagenomic data: challenges and applications. Brief Bioinform 2012;**13**:711–27.
47. Zhong C, Yang Y, Yooseph S. GRASP: guided reference-based assembly of short peptides. Nucleic Acids Res 2015;**43**:e18.
48. Jones P, Binns D, Chang H-Y et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 2014;**30**:1236–40.
49. Finn RD, Tate J, Mistry J et al. The Pfam protein families database. Nucleic Acids Res 2008;**36**:D281–8.
50. Haft DH. The TIGRFAMs database of protein families. Nucleic Acids Res 2003;**31**:371–3.
51. Gough J, Chothia C. SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. Nucleic Acids Res 2002;**30**:268–72.
52. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 2011;**39**:W29–37.
53. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: machine learning in {P}ython. J Mach Learn Res 2011;**12**:2825–30.
54. KEGG Mapper. http://www.genome.jp/kegg/mapper.html. Accesed 4 September 2017.
55. Alcaraz LD, Olmedo G, Bonilla G et al. The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. Proc Natl Acad Sci U S A 2008;**105**:5803–8.
56. Emerson D, Field EK, Chertkov O et al. Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. Front Microbiol 2013;**4**:254.
57. Nakagawa S, Shimamura S, Takaki Y et al. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. ISME J 2014;**8**:40–51.
58. Manzella MP, Holmes DE, Rocheleau JM et al. The complete genome sequence and emendation of the hyperthermophilic, obligate iron-reducing archaeon "*Geoglobus ahangari*" strain 234T. Stand Genomic Sci 2015;**10**:77.
59. Carbonero F, Benefiel AC, Alizadeh-Ghamsari AH et al. Microbial pathways in colonic sulfur metabolism and links with health and disease. Front Physiol 2012;1–11.
60. Nakagawa S, Takaki Y, Shimamura S et al. Deep-sea vent—proteobacterial genomes provide insights into emergence of pathogens. Proc Natl Acad Sci U S A 2007;**104**:12146–50.
61. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006;**27**:861–74.
62. Field D, Sterk P, Kottmann R et al. Genomic standards consortium projects. Stand Genomic Sci 2014;**9**(3):599–601.
63. Alazard D, Joseph M, Battaglia-Brunet F et al. *Desulfosporosinus acidiphilus* sp. nov.: a moderately acidophilic sulfate-reducing bacterium isolated from acid mining drainage sediments. Extremophiles 2010;**14**:305–12.
64. Romano C, D'imperio S, Woyke T et al. Comparative genomic analysis of phylogenetically closely related Hydrogenobaculum sp. isolates from Yellowstone National Park. Appl Environ Microbiol 2013;**79**:2932–43.
65. Chen L, Ren Y, Lin J et al. *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. PLoS One 2012;**7**:e39470.
66. Liljeqvist M, Valdes J, Holmes DS et al. Draft genome of the psychrotolerant acidophile *Acidithiobacillus ferrivorans* SS3. J Bacteriol 2011;**193**:4304–5.
67. Imhoff JF, Pfennig N. *Thioflavicoccus mobilis* gen. nov., sp. nov., a novel purple sulfur bacterium with bacteriochlorophyll b. Int J Syst Evol Microbiol 2001;**51**:105–10.
68. De Anda V, Zapata-Penasco I, Poot-Hernandez AC et al. Multigenomic Entropy Based Score (MEBS): the molecular reconstruction of the sulfur cycle. GigaScience Database 2017. http://dx.doi.org/10.5524/100357.
69. Pedroni P, Volpe AD, Galli G et al. Characterization of the locus encoding the [Ni-Fe] sulfhydrogenase from the archaeon *Pyrococcus furiosus*: evidence for a relationship to bacterial sulfite reductases. Microbiology 1995;**141**:449–58.
70. Taguchi Y, Sugishima M, Fukuyama K. Crystal structure of a novel zinc-binding ATP sulfurylase from Thermus thermophilus HB8. Biochemistry 2004;**43**(14):4111–8.
71. Santos AA, Venceslau SS, Grein F et al. A protein trisulfide couples dissimilatory sulfate reduction to energy conservation. Science 2015;**350**:1541–5.

## 1.1. MEBS: Supplementary Information

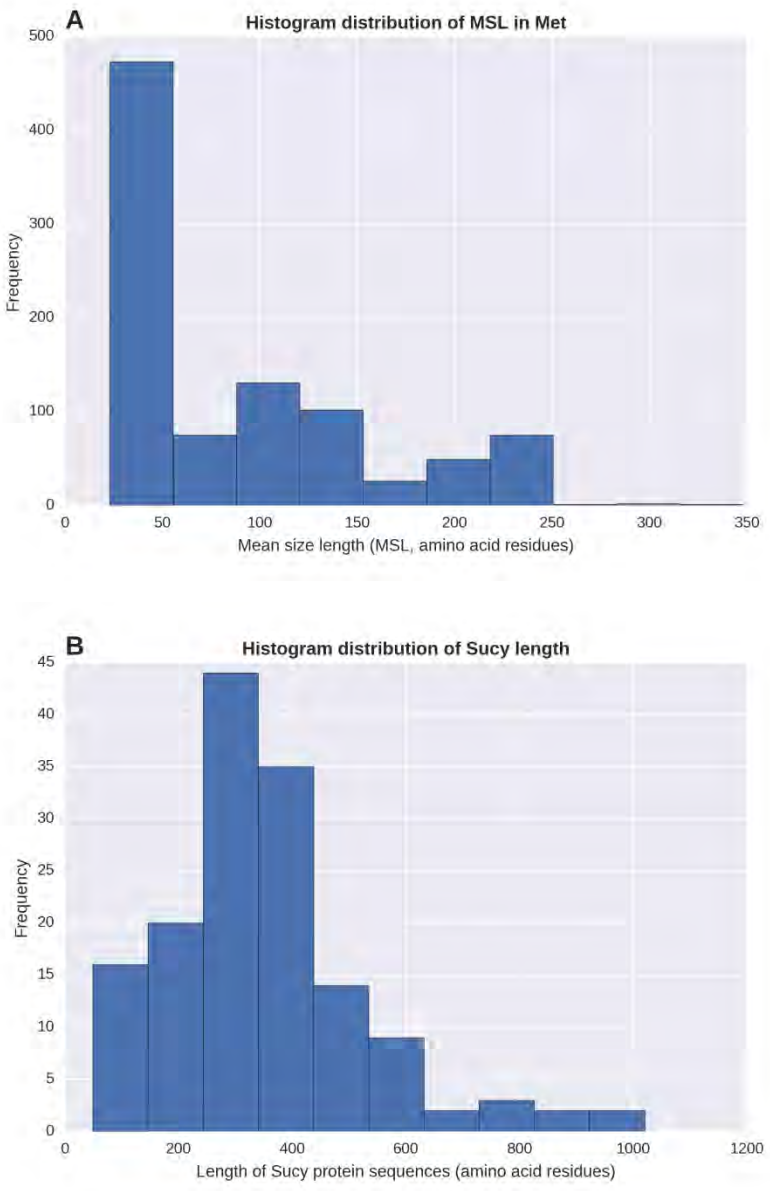



**Figure S1.** Histogram length distribution of the metagenomic dataset and input SuCy proteins. A) Mean size length of 935 metagenomes downloaded from MG-RAST. These gene-called protein sequences were allocated to these bins: 0-45, 46-80, 81-125, 126-175, 176-225, 226-275, >275. MSL ≤30=468, ≤60=66, ≤100=169, ≤150=99, ≤200=77, ≤250=58, ≤300=3. metagenomes), Sequences in the genomic dataset (Gen) were subsequently fragmented with matching sizes of 30, 60, 100, 150, 200, 250 and 300 amino acid residues (GenF categories). B) Length of sequences in the curated set S-related proteins (Sucy), comprising 152 proteins ranging from 49 to 1020 amino acid residues.

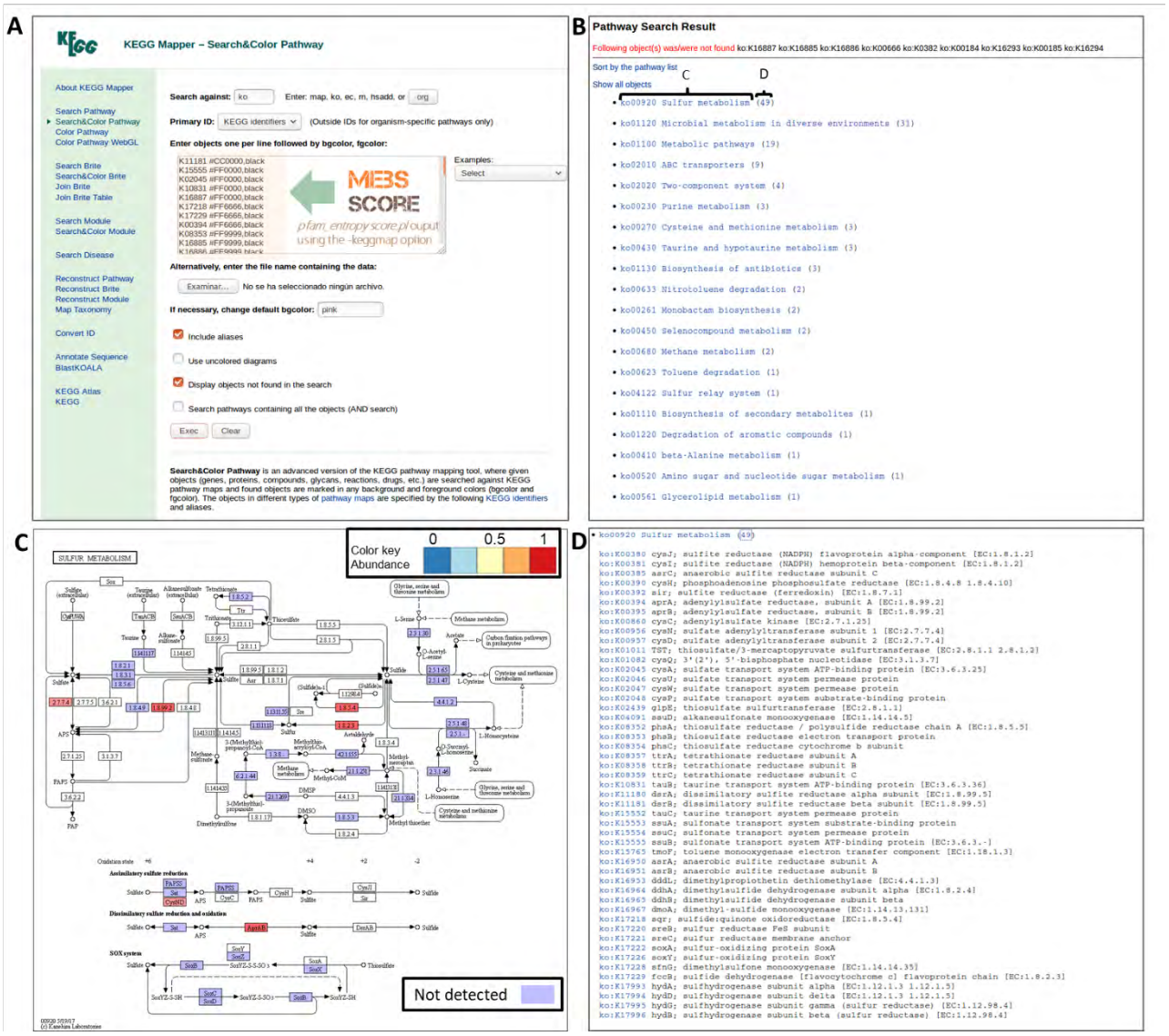



**Figure S2.** Visualization of the Pfam domains mapped onto KEGG metabolic pathways using script *pfam_score.pl* –keggmap option. A) By using a tab-separated file mapping Pfam identifiers to (KO numbers) and the corresponding pathway in Sucy (see Table S3), the output of the script can be exported to the KEGG Mapper – Search&Color Pathway tool(http://www.genome.jp/kegg/tool/map_pathway2.html and then select the metabolic map of interest as shown in B. In our case we selected the metabolic map00920 to map those KOs detected in the omic sample of interest (C). Due to the nature of the Pfam domains, is expected that several KO numbers might map to the same enzymatic number. These cases are colored in KEGG map with the color of the last given KO number in the output list, and therefore Finally a detailed description of the KO identifiers is also provided in KEGG (D).

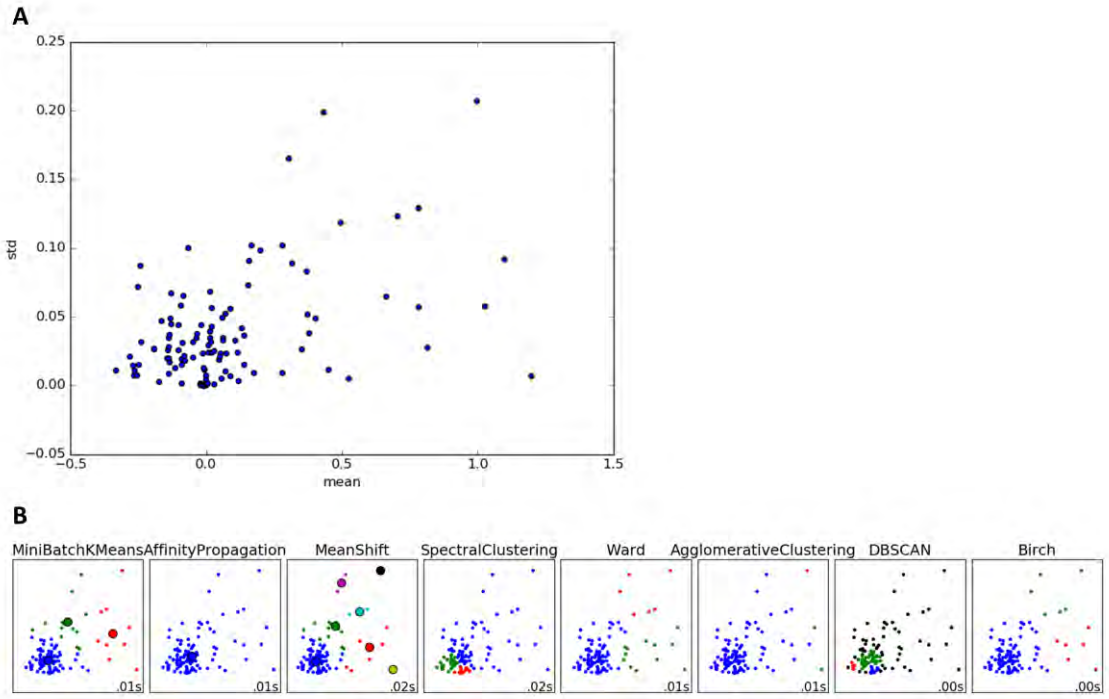



**Figure S3.** A) Scatter plot distribution of the 112 Pfam entropies according to their mean and std in GenF size categories (obtained with *plot_entropy.py* script). B) Comparison of clustering methods using scikit-learn Machine Learning module implemented in Python, using the script *plot_cluster_comparison.py* The clustering methods that clearly differentiate the Pfam's with high *H'* and low std, were: 1) Ward hierarchical clustering that uses as metric the distances between points and Birch branching factor that uses the Euclidean distance between points. The last line of squares represents the cluster separation of the Pfam entropies based in the scatter plot distribution showed above, where the entropy mean is in x-axis and std distribution is in y-axis

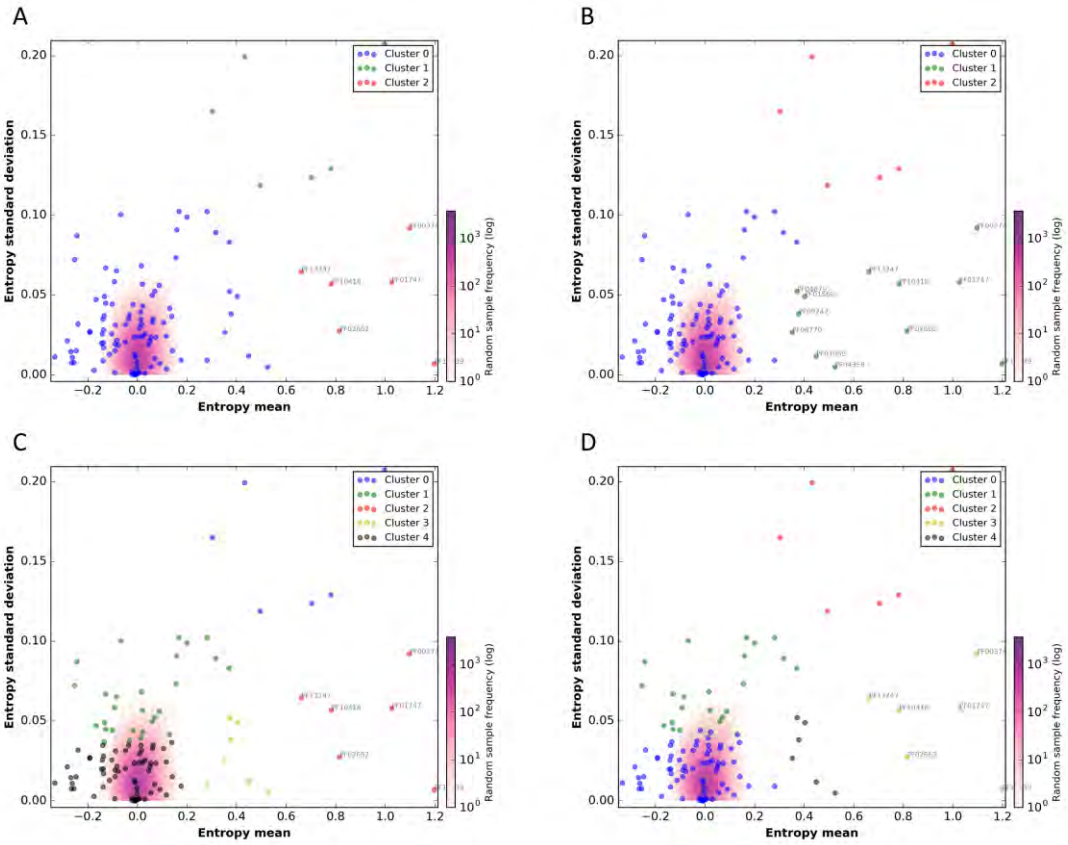



Figure S4. Clustering comparison between Birch and Ward clustering methods to stand out the Pfam entropies with high H' and low std using script *F_meanVSstd.py* A) Birch clustering algorithm k=3, B) Ward clustering algorithm K=3 C) Birch clustering algorithm K=5, D) Ward clustering algorithm K=5. Purple points indicate the Pfam values obtained in the random test (log frequency). The random values are located between -0.1 and 0.1, being the majority located around zero. Both methods clearly separate six Pfam entropies which are far from random values (H'>=1). However using Ward method, with three clusters (k=3) (B), six Pfam entropies are also considered. The Biochemical information indicates the importance of the latter Pfams in the Sulfur metabolism such as DsrC protein family (See Supplementary Table S6). Therefore we considered the Ward method k=3 to stand out the proposed molecular markers in metagenomic data.

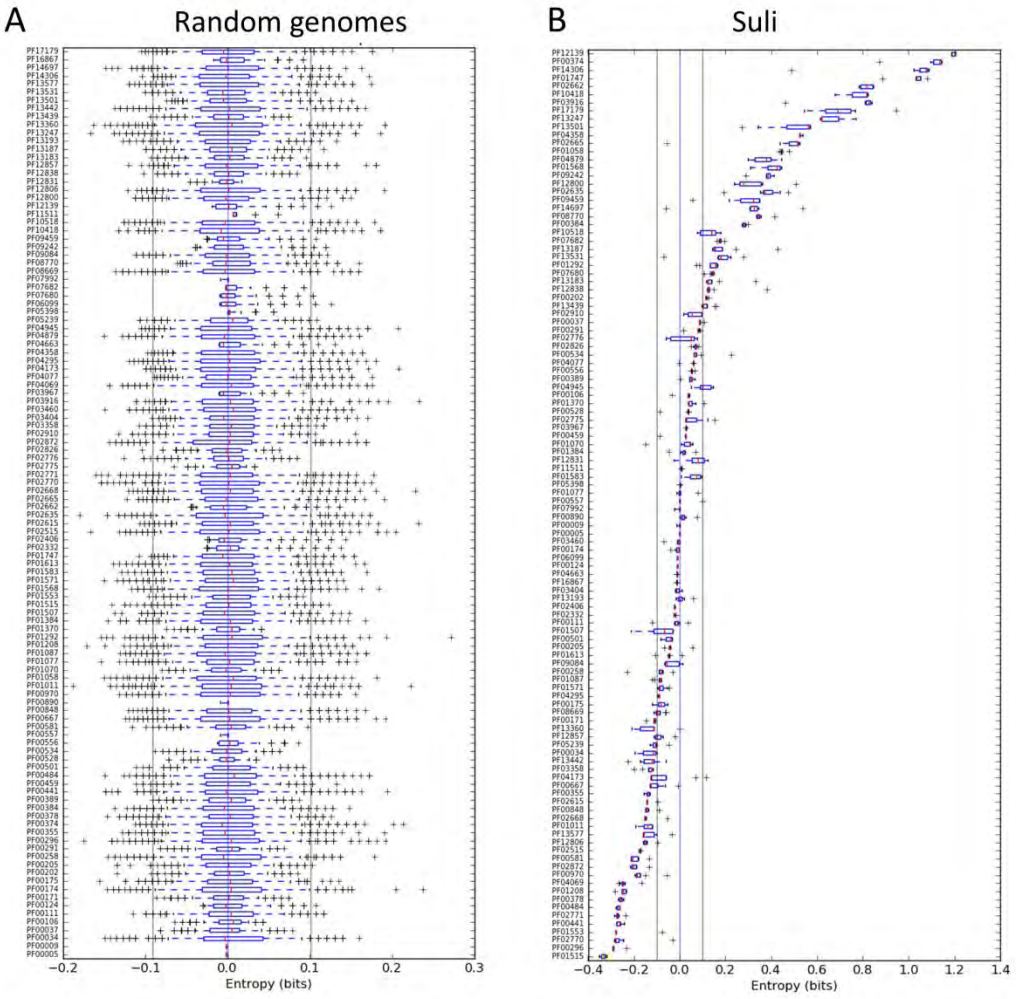



**Figure S5.** Distribution of entropy values of 112 Pfam domains inferred from random-sampled genomes (A) and S-energy based species in Suli (B). A) Boxplots of entropies computed from 1,000 lists of prokaryotic species sampled randomly and evaluated with all size categories in GenF. Mean percentiles 5% and 95% of the random distribution are shown as black lines. B) Boxplots of entropies computed from 161 prokaryotic species curated in Suli and evaluated with all size categories in GenF. Black lines indicate the [-0.1, 0,1] interval, where entropy values are likely to be obtained by chance. Plots were generated with scripts *extract_random_entropies.pl* and *plot_random_entropies.py*.

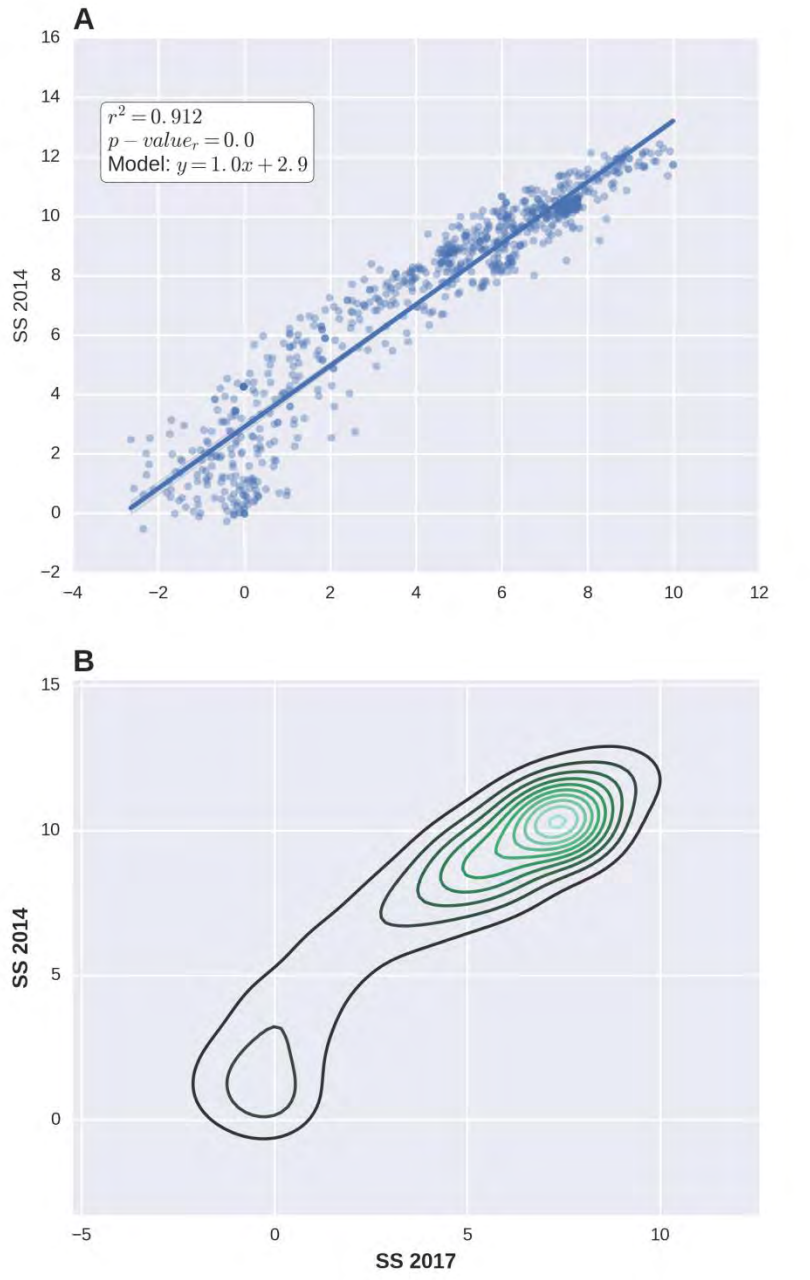
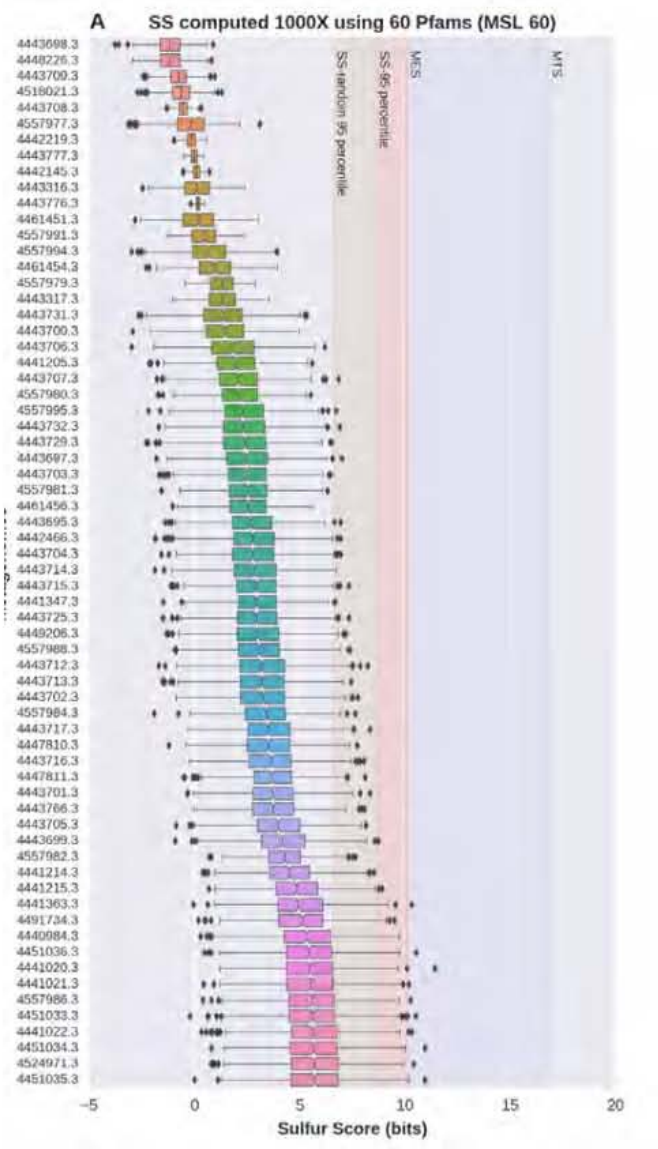



**Figure S6.** Comparison of Sulfur Scores (SS) computed in 2014 [Pfam v27, 1,528 non-redundant genomes and a sulfur list (Suli) of 156 microorganisms], and the SS computed in 2017 with Pfam v30, 2,107 non-redundant genomes and an updated Suli with 161 microorganisms. A) Linear regression of the SS computed in the same metagenomic dataset in both 2014 and 2017. B) Kernel density plot.


A
SS computed 1000X using 60 Pfams (MSL 60)
SS random 95 percentile
SS 95 percentile
MES
MTS
4443698.3
4448226.3
4443709.3
4518021.3
4443708.3
4557977.3
4442219.3
4443777.3
4442145.3
4443316.3
4443776.3
4461451.3
4557991.3
4557994.3
4461454.3
4557979.3
4443317.3
4443731.3
4443700.3
4443706.3
4441205.3
4443707.3
4557980.3
4557995.3
4443732.3
4443729.3
4443697.3
4443703.3
4557981.3
4461456.3
4443695.3
4442466.3
4443704.3
4443714.3
4443715.3
4441347.3
4443725.3
4449206.3
4557988.3
4443712.3
4443713.3
4443702.3
4557984.3
4443717.3
4447810.3
4443716.3
4447811.3
4443701.3
4443786.3
4443705.3
4443699.3
4557982.3
4441214.3
4441215.3
4441363.3
4491734.3
4440984.3
4451036.3
4441020.3
4441021.3
4557986.3
4451033.3
4441022.3
4451034.3
4524971.3
4451035.3
−5
0
5
10
15
20
Sulfur Score (bits)


B
SS variation in random test
SS
SS_random
Sulfur Score (bits)
10
8
6
4
2
0
−2
−4
0
10
20
30
40
50
60
Number of metagenomes

**Figure S7.** Sulfur-Scores (SS) obtained in a panel of 66 metagenomes by random sampling Pfam domains. The results shown correspond to metagenomes of Mean Size Length (MSL) of 60 amino acid residues. SS was computed 1000 times for each metagenome by randomly selecting 60 domains out of 112 S-related Pfam domains (described in Figure S2A). A) Boxplot of obtained SS in the context of the Maximum Theoretical Score (MST), the Maximum Expected Score (MES) and the 95% percentiles obtained with 112 and a subset of randomly sampled informative Pfam domains. MST is the sum of all Pfam domains with entropy > 0. MES is the sum of all 112 entropies in Figure S2A. B) Comparison of SS computed with all 112 domains and the median SS resulting from random sampling simulations on the same metagenomic datasets.

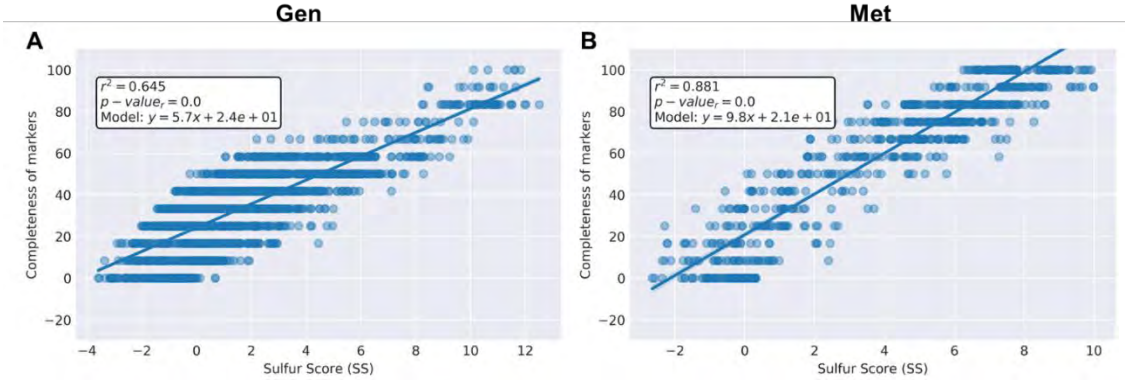



**Figure S8.** Linear regression models of the Sulfur Score (SS) and the mean level of completeness of a synthetic pathway of 12 candidate markers in Gen and Met datasets.

Table S4. Informative Pfam domains with high H' and high std (not used as molecular marker genes) in metagenomic fragmented data.

| Pfam | H' mean | H' Std | Description |
|---|---|---|---|
| PF14306 | 1 | 0.21 | PUA like domain is found at the N-terminus of ATP-sulfurylase enzymes. |
| PF03916 | 0.78 | 0.13 | NrfD is an integral transmembrane protein containing 8 conserved hydrophobic segments. It may be the naphthtoquinol oxidase although a role as a proton pump has not been excluded (Caspi et al., 2012a).Possibly involved in S-energy conservation purposes as is being observed a 30% of similarity with oxidoreductase membrane unit from SRB such as *Desulfitobacterium hafniense* Y51 (31% identity, 21% coverage)(Price and Arkin, 2017). |
| PF17179 | 0.7 | 0.12 | Part of a bifunctional enzyme complex that functions as an NADPH-dependent hydrogen-evolving hydrogenase with sulfur reducing activity (Price and Arkin, 2017) |
| PF13501 | 0.49 | 0.12 | This domain is found in the sulfur oxidation protein. SoxY plays a key role in sulfur oxidation SoxY is a sulfur covalently-binding protein, which binds sulfur anions, including sulfur, sulfite, thiosulfate and sulfate, to form a SoxY-thiocysteine-S-sulfur compound adducts , functioning as the sulfur substrate-binding protein that offers its sulfur substrate, which is covalently bound to a conserved C-terminal cysteine, to another oxidizing Sox enzyme (Caspi et al., 2012a; Finn et al., 2008) |
| PF02665 | 0.43 | 0.2 | Gamma subunit of the nitrate reductase enzyme, b-type cytochrome that receives electrons from the quinone pool It then transfers these via the iron-sulfur clusters of the beta subunit to the molybdenum cofactor found in the alpha subunit (Finn et al., 2008) |
| PF14697 | 0.3 | 0.17 | This family includes proteins containing domains which bind to iron-sulfur clusters. Members include bacterial ferredoxins, various dehydrogenases, and various reductases, the domain contains two 4Fe4S clusters (Finn et al., 2008). |

Table S5. Percentile distribution of the 112 Pfam entropies in the random test

| GenF size category | 5-percentile | 95-percentile |
|---|---|---|
| Real | -0.091 | 0.101 |
| 30 | -0.086 | 0.105 |
| 60 | -0.09 | 0.104 |
| 100 | -0.088 | 0.1 |
| 150 | -0.09 | 0.103 |
| 200 | -0.89 | 0.105 |
| 250 | -0.09 | 0.106 |
| 300 | -0.09 | 0.1 |

**Table S10 Statistics of SS computed on genomic (Gen, real sequences) and metagenomic (Met, with increasing Mean Size Length, from 30 to 300aa) datasets.**

| | | | | | | SS All Pfam domains (n=112) | | | | | Random SS sampled Pfam domains (1000x) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Category | N[a] | Pos[b] | MTS[c] | MES[d] | 95th[e] | max[f] | mean[g] | sem[h] | std[i] | 95th[e] | max[f] | mean[g] | sem[h] | std |
| Gen | real | 2,107 | 55 | 16.01 | 9.49 | 7.80 | 12.5 | 1.76 | 0.06 | 2.83 | 4.68 | 11.99 | 1.05 | 0.00 | 0.01 |
| Met | 30 | 468 | 51 | 13.67 | 7.76 | 7.66 | 7.78 | 4.63 | 0.15 | 3.14 | 5.41 | 9.50 | 2.25 | 0.07 | 1.54 |
| | 60 | 66 | 60 | 16.81 | 10.0 | 9.70 | 9.92 | 4.74 | 0.42 | 3.41 | 6.64 | 11.42 | 2.68 | 0.24 | 1.96 |
| | 100 | 164 | 57 | 15.56 | 8.98 | 8.81 | 9.09 | 5.78 | 0.20 | 2.52 | 6.38 | 11.20 | 3.17 | 0.11 | 1.38 |
| | 150 | 99 | 56 | 15.84 | 9.39 | 8.51 | 9.99 | 3.22 | 0.37 | 3.64 | 5.66 | 10.80 | 1.70 | 0.19 | 1.91 |
| | 200 | 77 | 56 | 15.88 | 9.48 | 8.18 | 9.55 | 4.07 | 0.30 | 2.64 | 5.36 | 10.19 | 2.11 | 0.16 | 1.40 |
| | 250 | 58 | 55 | 16.03 | 9.52 | 8.98 | 9.27 | 5.74 | 0.42 | 3.24 | 6.39 | 10.82 | 3.02 | 0.23 | 1.72 |
| | 300 | 3 | 57 | 15.92 | 9.46 | 7.61 | 8.38 | 2.51 | 3.00 | 5.20 | 6.26 | 8.74 | 1.41 | 1.64 | 2.84 |

[a]N: Total number of genomes or metagenomes considering on each category

[b]Pos: The number of randomly sampled domains corresponded to the number of Pfam families with positive entropy (H' > 0).

[c]MST: Maximum Theoretical Score is the sum of all Pfam domains with H' > 0. Represents a utopic scenario were only informative domains are present in any given 'omic sample'.

[d]MES: Maximum Expected Score is the sum of all 112 entropies

[e]95th: Percentile 95$^{th}$ of the SS distribution in the corresponding category

[f]max: Maximum *SS* value

[g]mean: Mean *SS* value

[h]sem: standard error of mean used as a measure of precision for the estimated distribution mean of *SS*

[i]std: standard deviation of the *SS* distribution in the corresponding category used as measure of data variability around mean

# Capítulo 2

# A. The sulfur cycle as the gear of the “clock of life”: the point of convergence between geological and genomic data in the Cuatro Cienegas Basin.

Valerie De Anda, Icoquih Zapata-Peñasco, and Valeria Souza

**Abstract**

Due to its chemical properties and several stable redox states, microbial transformations of sulfur compounds, have being affecting the geochemical features of the Earth’s biosphere since the Archaean. However, despite the great importance of sulfur cycling, reconciling the geologic record with genomic data has been challenging. Here we first review current state of the art evidence about the emergence of life on Earth in sulfur-rich environments, providing a conceptual framework that closely connects these two largely separated disciplines. Then, we summarize the current astonishing diversity of prokaryotes responsible for driving the sulfur cycle, suggesting that, due to their ancient origin, sulfur-associated taxa perhaps hold the greatest diversity of any group of microorganisms to metabolize a single element on Earth. Finally, because the guilds of sulfur metabolizing microbes co-occur at millimeter scales within microbial mats, we use the taxonomic and metabolic information derived from these primordial communities as ecological models to highlight sulfur as the guiding axis of these complex intersections, recapitulating a gear of the clock of life.

## 2.1. Oldest evidence of life on Earth: stromatolites, hydrothermal vents and hot springs - what do they have in common?

Since its discovery in the early 1980s, scientific consensus indicated that the Dresser Formation, Pilbara Craton, Western Australia contained evidence of the earliest forms of life on Earth. Paleobiological and chemical data suggest that the stromatolites within the Dresser formation (a region underlain by a 30-km thick sequence of relatively well-preserved sedimentary and volcanic rocks) are approximately 3.465 million year old (Ma) and hold the oldest and most convincing signs of life in the form of domical and coniform layered stromatolites (Allwood et al., 2006; Van Kranendonk et al., 2008). The biogenicity of these fossils was carefully confirmed by morphology, stable isotope signatures, seawater-like trace element signatures, and the presence of microfossils that were identified as the earliest known assemblage of cellular fossils (Allwood et al. 2006; Noffke et al. 2013; Schopf 1993; Tice and Lowe 2004; Van Kranendonk et al. 2008; Walter et al. 1980; Westall et al. 2006). More than two decades ago, this assemblage of microfossils was associated with the morphological taxa of prokaryotic, filamentous and coccoidal microorganisms (Schopf, 1993) and was recently validated using secondary ion mass spectroscopy. Other interpretations of the Dresser Formation, using sulfur isotopes, indicated that microbes disproportionated sulfur at that time, evidence of processing of elemental sulfur (Philippot et al., 2007). Evidence of microbe-driven sulfate reduction has also been found in sedimentary settings (Shen et al., 2001, 2009; Ueno et al., 2008a) and hydrothermal seafloor samples (Aoyama and Ueno, 2018; Shen et al., 2001, 2009; Ueno et al., 2008b). The presence of these prokaryotic assemblages, consisting of photosynthesizers, Archaeal methane producers, methane consumers, and sulfur users, (Schopf et al., 2018), as well as the diversity of the stromatolites, points to a biochemical sophistication of life by that time, 3.5 billion years ago, in which the biosphere had already given rise to an ancient but metabolically diverse ecosystem. Consequently, life must have originated significantly earlier on our planet. More recently, evidence points toward the limits between the Haedean and the Archeaean period,

4.1 billion years ago, where the life seems to have left an isotopic imprint in two independent locations (Dodd et al., 2017; Tashiro et al., 2017).

Recent evidence has confirmed that in fact life was present even hundreds of millions of years earlier than even the Dresser Formation. First, recently discovered stromatolites from the Isua Supracrustal Belt (ISB) in southwest Greenland (Nutman et al., 2016) are dated at 3.8-3.7000 billion years-old and indicate macroscopic layered structures produced by microbial communities that, according to trace elements signatures, grew in a shallow marine environment. (Dalton, 2004). Second, recent hypotheses point out that thermal sulfur-related environments, such as hot springs or hydrothermal vents, where high temperatures might have increased the rate of primordial biochemical enzymatic reactions and where high concentrations of reducing equivalents, made possible chemiosmosis gradients across a cell membrane, could indeed have been the right places for life to emerge (Dai, 2017; Djokic et al., 2017; Van Kranendonk et al., 2017). Clear evidence of this has been found in precipitates from the Nuvvuagittuq belt in Quebec, Canada, interpreted as putative fossilized microorganisms in ferruginous sedimentary rocks deduced as seafloor-hydrothermal vent-related that are at least from 3,770 and possibly 4,280 Ma (Dodd et al., 2017). Other fossil evidence, though not as ancient as the Quebec precipitates, points to an active volcanic landscape, in contrast to deep hydrothermal settings. Stratigraphic and geochemical evidence indicates that recently discovered Archean hot-spring deposits of 3.5 Ga, found also in the Dresser formation, were not so different from those present today and were affected by voluminous hydrothermal fluid circulation preserving a suite of microbial bio-signatures (Djokic et al., 2017; Van Kranendonk et al., 2017).

As can be seen, the recent evidence from geological record has shown that the right physical and chemical settings for the origin of life can be found in hot springs and hydrothermal vents, where the oldest evidence is possibly 4,280 Ma. Under this scenario, the theory of the Iron-Sulfur World suggests that life started chemoautotrophically in S-rich environments, such as aqueous flows of volcanic outlfows where the oxidative formation of pyrite satisfies

all the necessary conditions needed by an energy source for such an origin (Wächtershäuser, 1990, 2008)

The geologic record suggests that these oldest primordial environments (mats, hotsprings, and hydrothermal vents) are the sites where a bridge between sulfur chemistry and life arose through energetically favourable reactions that led to one of the most important biological processes on Earth, "the fixation of carbon dioxide into organic materials" (See Chapter 1 ). Among these energetically favourable reactions, the most accepted sources of energy available for chemosynthesis in early Earth were hydrogen gas (reacting with oxidants such as carbon dioxide) and sulfur dioxide (because, under anaerobic conditions, the reaction of $CO_2$ with $H_2$ releases energy). Current modern microorganisms using these ancient forms of chemical energy ( $H_2/CO_2$ and $H_2/H_2S$ couples) are found in sulfur-rich such deep environments such hydrothermal vents of the seafloor (Staley, 2002). In fact, several lines of evidence support sulfide ($H_2S$) over $H_2$ as the primordial energy source (reviewed in detail in (Olson et al., 2016)). For example due its chemical properties, sulfide can serve as an important organic product, reactant, catalyst (proto-enzyme), barrier (proto-membrane), and sustainable source of energy. (Olson et al., 2016) . Furthermore, recent studies suggest a critical role of FeS-dependent enzymes and of thioesters as carriers of chemical energy in the assembly of early metabolic networks. In fact, the exergonic reaction to form a thioester is the basis of the most ancient form of $CO_2$ fixation, the acetyl-CoA pathway (Goldford et al., 2017; Martin and Thauer, 2017; Semenov et al., 2016).

The paleontological and geological records depict the environment where the evolution of life occurred and based, on the assumption that molecular sequence determines biological function and that the ancient molecules would have necessarily been adapted to function in their surroundings, well-conserved molecules should provide evidence of past environmental conditions (Garcia et al., 2017). Thus, the molecular reconstruction of life using completely sequenced genomes (phylogenomics) can give evidence about the dominant conditions in the paleo-environment under study (Garcia et al., 2017).

Reconciliation with standard models of microbial evolution derived from phylogenies indicate that early branches are occupied by thermophilic and hyper-thermophilic prokaryotes capable of a wide range of S oxidation and reduction processes. This suggests that, in early ecosystems, microbial communities exploited the full range of the S-redox spectrum, with some organisms reducing and others oxidizing the reduced sulfur species. Hence, microbes from the "FeS world" were already building diverse stromatolites before the origin of cyanobacteria (Grassineau et al., 2001). In fact, the grouping of hyper-thermophilic species on short branches near the bases of both the bacterial and archaeal domains of the tree parsimoniously suggests that LUCA (Last Universal Common Ancestor) was a hyperthermophile (Gaucher et al., 2010; Lake et al., 2009). Moreover, respiration of sulfur compounds is thought to be one of the first metabolisms to evolve in a hot primitive Earth. Indeed, the placement of hyperthermophiles Archaea and deep-branching thermophilic bacteria is consistent with such an early origin. Isotopic data also suggest that sulfate reduction began around 3 billion years ago, and acquired global significance only after sulfate concentrations had significantly increased in the Precambrian oceans (Trudinger, 1992; Ueno et al., 2008b; Zhelezinskaia et al., 2014). The conservation of the sulfate reduction metabolism (using dissimilatory sulfite reductase) in both prokaryotic domains of life (Archaea and Bacteria), suggest two main hypotheses. The first one suggests that this trait could be present in the last common ancestor before the divergence of the two domains. The second hypothesis suggests that such a metabolic trait could have evolved in one of the domains soon after the divergence and was then transferred to the other domain by several early lateral gene transfer events. Given the expected parsimony of evolution, deep ancestry before divergence is the most accepted hypothesis (Müller et al., 2015; Muyzer and Stams, 2008; Wagner et al., 1998).

## 2.2 Using sulfur in early life: molecular and genomic evidence

In agreement with the geological record, the metabolic implications of so-called 'standard'rRNA models suggest that the full sulfur cycle pre-dated oxygenic photosynthesis and thus was in operation before cyanobacteria had evolved and were building diverse

stromatolites (Grassineau et al., 2001; Staley, 2002). In fact, the grouping of hyperthermophilic species on short branches near the bases of both the bacterial and archaeal domains of the tree parsimoniously suggest that LUCA (Last Universal Common Ancestor) was a hyperthermophile (Gaucher et al., 2010; Lake et al., 2009), and was confined to FeS compartments geologically produced at vents on the ocean floor that provided the initial cell 'wall' (Dai, 2017). Compartmentalization is a prerequisite for the evolution of any complex system and it seems most possible that these configuration consisted of iron sulfide (FeS) deposited at a warm (< 90 °C) submarine hydrothermal spring (Koonin and Martin, 2005).

## 2.3 Modern distribution of sulfur metabolism across the Tree of Life.

Currently, microbial usage of sulfur compounds is reflected in the astonishing diversity of sulfur-related metabolic pathways that strongly correlate with environmental and ecological influences. The great diversity of sulfur-related microorganisms on Earth can be observed as a result of continual innovations in molecular pathways, which involve large sets of enzymes, organic substrates, and electron carriers, which in turn ultimately depend on particular redox potentials, organic matter availability, electron donor or acceptor disposal, and temperature, among others (Canfield et al., 2005; Ghosh and Dam, 2009a; Hubas et al., 2011b)This complexity increases when considering the surrounding geochemical and ecological conditions of the sulfur players and the syntrophic relationship and shared chemical energy across space and time (Lau et al., 2016; Ozuolmez et al., 2015; Plugge et al., 2011). Figure 1 summarizes the distribution of sulfur metabolism guilds across the Tree of Life. Clearly, metabolic types are typically not confined to monophyletic clades. Conversely, species that do not use sulfur as an energy source are often closely related to sulfur users.

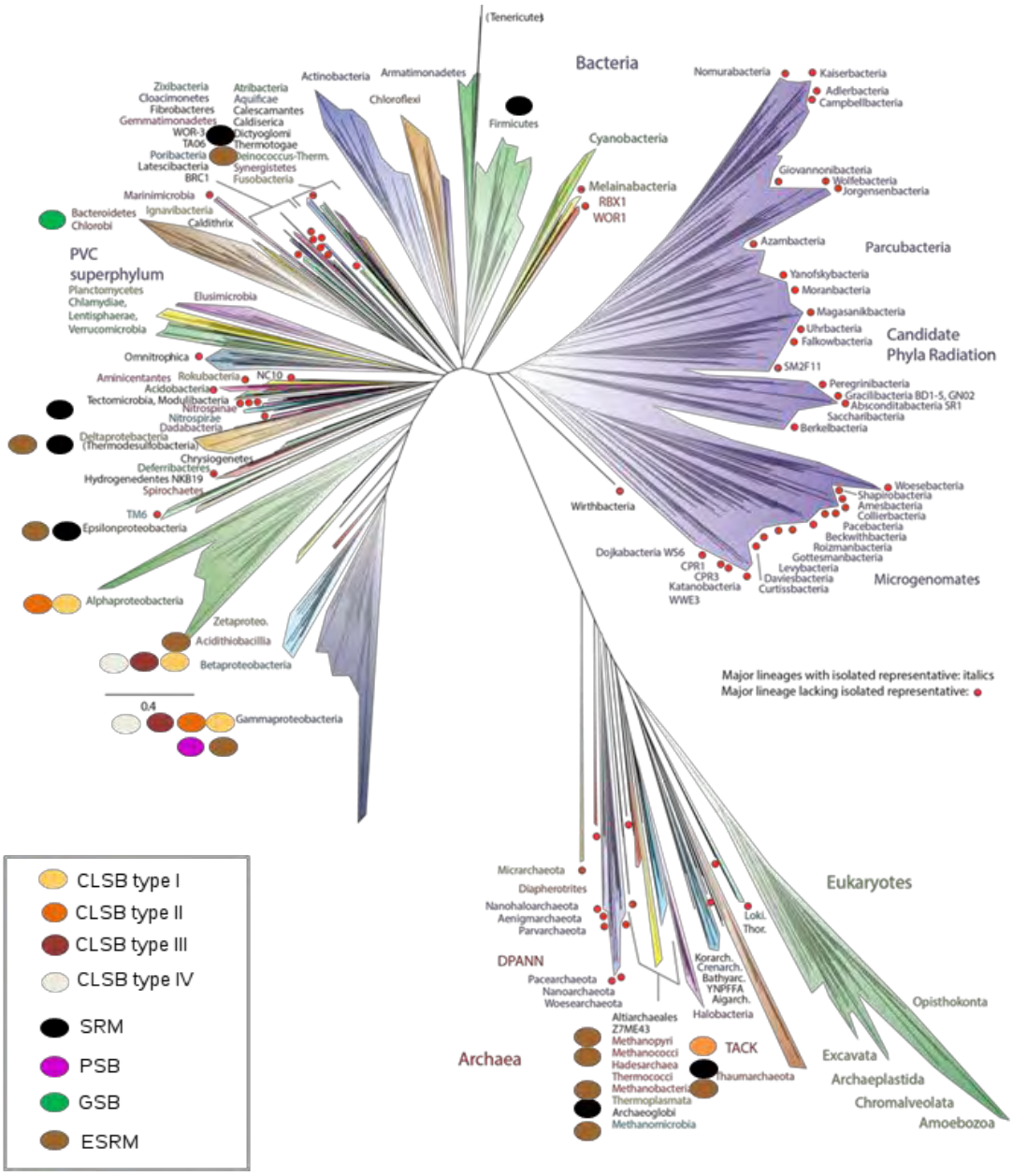



Figure 1. Current view of the tree of life, encompassing the total diversity represented by sequenced genomes. The distribution of the sulfur metabolic guilds is indicated by colored ellipses. Modified figure from Hug et al. (2016)

Historically, the S-taxa have been distinguished based on combined considerations derived from physiological, biochemical and evolutionary traits, e.g. membrane structure, pigment composition or phylogenetic position, and are grouped together into five well-established metabolic guilds that we briefly describe below:

### 2.3.1 Colorless Sulfur Bacteria (CLSB)

This guild is historically named for their lack of photopigments, although some are colorful (pink or brown) due to their high cytochrome content. They have been sub-divided with respect to their metabolic capabilities into the following four categories. (A) Obligate chemolithotrophs use inorganic reduced sulfur compounds as energy source and use the Calvin cycle for carbon fixation. In the most recent tree of life classification this subgroup of CLSBs is scattered among the Eubacteria and Archaea domains of life, including Betaproteobacteria (e.g. *Thiobacillus* and *Thiomonas*), Gammaproteobacteria *Thiomicrospira*, Aquificeae (e.g *Hydrogenobacter* and Crenarchaeota (e.g *Sulfolobus*). (B) The second class of CLSB is composed of the facultative chemolithotrophs, which can grow on mixtures of reduced sulfur compounds and organic substrates either chemolithoautotrophically or heterotrophically with complex organic compounds used as both carbon and energy sources. In contrast to the first group, the facultative chemolithotrophs are restricted within the proteobacteria group: Gammaproteobacteria (e.g. *Thiosphaera*, *Beggiatoa* and Alphaproteobacteria (e.g. *Paracoccus*). (C) The third type, the chemolithoheterotrophs also gains energy from the oxidation of reduced sulfur compounds, but does not fix carbon dioxide. This metabolism is also found in some species of Gamma- and Betaproteobacteria (e.g *Thiobacillus* and *Beggiatoa*, respectively). (D) The last group of CLSB encompasses chemoorganoheterotrophic microorganisms that possess the ability to oxidize reduced sulfur compounds, but do not appear to derive energy from them. These microorganisms are placed within the gamma and beta proteobacteria classes (e.g. *Thiobacteria*, *Thiotrix*, and *Macromonas,* respectively(Robertson and Kuenen, 2006)

### 2. 3.2 Purple sulfur bacteria (PSB)

These are named for their pigmentation due to the presence of bacteriochlorophyll *a* and *b*. These metabolically versatile bacteria can grow photolithoautotrophically using low light, relying on reduced sulfur compounds as electron donors, and fixing carbon through Calvin cycle. The PSB are divided into two families, the *Chromatiaceae* and the *Ectothiorhodospiraceae* within the Gammaproteobacteria division.

### 2.3.3. Green sulfur bacteria (GSB)

GSB are also named for their bacteriochlorophyll content, but unlike PSB, the GSB harbors c, d, e, and f type bacteriochlorophyll. They also use the reverse tricarboxylic acid cycle, also known as the Krebs cycle, for carbon dioxide fixation, instead of the Calvin Cycle. In addition, most GSB can assimilate a small number of simple, organic compounds such as acetate, but only in the presence of $CO_2$ and a photosynthetic electron donor. Most strains use electrons derived from oxidation of sulfide, but some strains can also oxidize elemental sulfur, thiosulfate, hydrogen and iron(II) (Frigaard and Bryant, 2008).

Overall, the oxidative sulfur metabolism in the photosynthetic anoxygenic bacteria shares some chemical and ecological similarities; however, their physiology and evolution are rather different. For example, the GSB are less physiologically versatile compared with the PSB (Friedrich et al., 2001; Frigaard and Bryant, 2008).

The metabolism involved in the reduction of sulfur compounds has also been classified based on the ecological, biochemical, and evolutionary traits of the microorganisms involved: sulfate reducing microorganisms (SRM) vs elemental sulfur reducing microorganisms (ESRM). There are also organisms that use other oxidized sulfur compounds, in particular thiosulfate and sulfite, as electron acceptors in their energy metabolism. However, very little is known about obligate thiosulfate or sulfite reducers. Usually, they are subsumed in the SRM class (Canfield and Thamdrup 2005).

### 2.3.4. Sulfate reducing microorganisms (SRM)

These anaerobic microorganisms mainly use sulfate as a terminal electron acceptor in an energy-gaining respiratory process linked to the oxidation of an electron donor. This electron-transfer pathway of this metabolism is not well understood (Pereira et al., 2011). The range of substrates used by SRM is very broad (e.g. dicarboxylic acids, alcohols, amino acids, sugars, a wide variety of aromatic compounds, benzene, straight chain alkanes, and inclusive man-made xenobiotic compounds and hydrogen) (Canfield and Thamdrup 2005). Such a large substrate amplitude suggests a series of horizontal gene transfer events as the origin of their broad metabolic and ecological versatility (Müller et al., 2015; Wagner et al., 1998). Supporting this hypothesis, within the tree of life the sulfate reduction is not only observed in deeply branched classes within Thermodesulfobacteria (e.g. *Thermodesulfobacterium*) but also is found in the nitrifying group Nitrospira (e.g *Thermodesulfovibrio*), and spore-forming Firmicutes within the Terrabacteria group (e.g. *Desulfotomaculum*) and the best-studied group of deltaproteobacterias (e.g. *Desulfovibrio*). In the archaea domain are representatives from both classes (Eurayarchaota (e.g. *Archaeoglobus*) and Crenarchaeota (*Thermocladium* and *Caldivirga*) (Canfield and Thamdrup 2005), suggesting again, either a very deep origin or several events of horizontal gene transfer.

### 2.3.5. Elemental Sulfur Reducing Microorganisms (ESRM)

Although elemental sulfur is not a good substrate for enzymatic reactions due to its low solubility in water (5 µg/L at 47ºC) (Caspi et al., 2012b), the ability to reduce elemental sulfur is widespread among several members of both prokaryotic domains of life. However, there are several biochemical and physiological differences in how sulfur is used and whether sulfur reduction serves as the principal energy-acquisition process or whether it is an alternative one. Chemolithoautotrophs use a heterotrophic reduction of elemental sulfur as the primary energy source; this particular metabolism is mostly found in very deep branched hyperthermophilic archaea from the Crenarchaeota division (TACK) (e.g. *Pyrodictium,*

*Pyrobaculum, Sulfolobus*) including a methanogenic member (Canfield and Thamdrup 2005), e.g. *Methanopyrus, Methanobacterium, Methanothermus*. It is also found in some Deltaproteobacteria (e.g. *Desulfurella acetivorans*, *Sulfospirillum arcachonense*). Although elemental sulfur is the main electron acceptor in this type of metabolism, other inorganic compounds such as nitrate, iron (III), or even oxygen can also be used. For the second group, it has been proposed that respiration of elemental sulfur constitutes an alternative metabolism for some versatile strains of Gammaproteobacteria such as Wolinella, *Pseudomonas, Shewanella* (Canfield and Thamdrup 2005) as well as the chemolithotroph *Acidithiobacillus ferrooxidans*, within the recent proposed class Zetaproteobacteria (Hoshino et al., 2016). The phylogenetic position of the deep-branched chemolithoautotrophs, as well as the unspecific nature of the reaction center, suggests that very early life used this metabolic pathway, although the second group seems to go back to ancient enzymes when there were no other sources of food.

## 2.4 Integrating the sulfur cycle within the microbial mat model

As explained in detail above, each of the metabolic reactions related to the SC are carried out by physiologically and phylogenetically diverse groups of microorganisms that live in widely varied environments and differ in their abilities to use various sulfur compounds. The great complexity of the biogeochemical sulfur cycle at the global scale is shown in Figure 2, where the most important organic and inorganic S-compounds derived from biogeochemical processes are arranged according to the Standard Gibbs free energy of formation and the boxes summarize the sulfur metabolic guilds. The compilation of the described genera for each guild along with their corresponding complete genome sequences is found in Supplementary information in De Anda et al. (2017): (http://gigadb.org/dataset/100357), and as Appendix 1 at the end of the thesis.

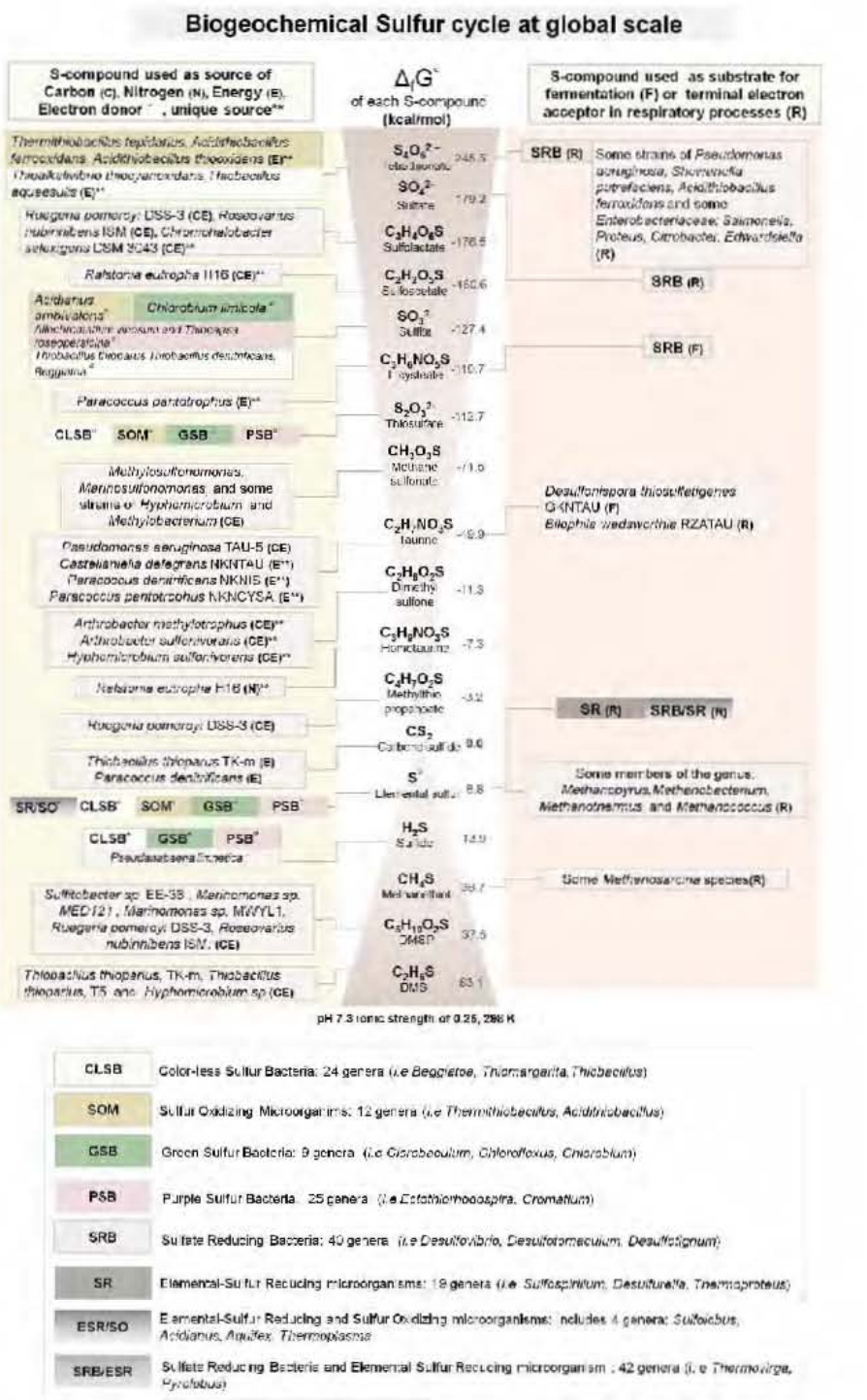



Figure 2. Sulfur cycle at the global scale, where the most important organic and inorganic S-compounds derived from biogeochemical processes are arranged according to the Standard Gibbs free energy of formation. Colored boxes indicate the metabolic guilds involved in the metabolism of S-compounds in oxidation (i.e., CLSB, SOM, PSB, and GSB) or reduction (SR, SRB) processes, which are summarized in the Figure 3. Figure from De Anda et al. (2017)

The great complexity of the sulfur cycle can be studied by considering the in microbial mats as ecological model. Microbial mats are compartmentalized organizations that have evolved over more than three billion years into the complex ecosystems that we know today (Herman and Kump 2005). Functionally, microbial mats are self-sufficient structures that support most of the major biogeochemical cycles within a vertical dimension of only a few millimeters in a multilayered space (Pinckney and Paerl, 1997). The metabolic guilds of microbial mats live in close proximity, exchanging nutrients and by-products such as organic matter, sulfide, and sulfate derived from photosynthesis, sulfate reduction and sulfide oxidation respectively. In this way, by creating an efficient turnover of major electron acceptor/donors, sulfur gradients contribute to a fine vertical distribution that sustain communities within which the sulfur cycle is the main crossing point of the major biogeochemical cycles (Dillon et al. 2007; Herman and Kump 2005; Prieto-Barajas et al. 2017; van Gemerden 1993; Visscher and Stolz 2005). Here, we classify the metabolic guilds within microbial mats according to their method of obtaining energy and carbon for biomass, including: i) photoautotrophs: photosynthetic cyanobacteria, PSB and GSB, ii) chemoautotrophs: sulfur oxidizing microorganisms such CLSB, and finally iii) heterotrophs: sulfate reducing bacteria, SRB (Bolhuis et al., 2014; Hubas et al., 2011b; van Gemerden, 1993)

The most representative reactions performed by these sulfur metabolic guilds within a typical microbial mat are shown schematically in Figure 3A. In this conceptual model, the upper layer of the mat is represented primarily by photosynthetic cyanobacteria that take advantage of abundant resources: sunlight and $H_2O$, which are used as source of energy and electrons (respectively) to fix $CO_2$. The oxygen and organic matter produced by photosynthesis rapidly diminish in deeper layers (Harris et al., 2013). The turnover of fixed carbon is caused by fermentative and heterotrophic components such as SRB that use inorganic sulfate as an external electron acceptor to oxidize organic matter compound substrates, resulting in the production of sulfide. The latter byproduct, as well as other reduced sulfur compounds, are used by PSB and CLSB as electron donors for anaerobic phototrophic and chemotrophic growth (respectively). The sulfate produced in turn is used by SRB.

In this context, the compartmentalization of microbial mats provides clear, natural boundaries that evoke the concept of a "minimum ecosystem". That is, specific parts of the cycle may be seen as parts of a whole. For instance, the redox level, reduced-oxidized compounds or even genes and enzymes implicated in certain routes, can be used as ecosystem boundaries. These assemblies set up a unit that represents the minimum ecosystem with minimum requirements for functionality, and therefore this can be applied as an ecological model to understand the intersection of the biogeochemical cycles. Considering the difficulties in delimiting the parameters that define a whole biogeochemical cycle, including its limits and scope or what elements should be considered as necessary to each cycle, we focus on metabolic coupling within microbial mats to highlight the cycles interplay at taxonomic and metabolic levels (Figure 3A).

In this schematic representation, the colored circles represent the metabolic guilds within the microbial mats ranging from the outer circle (layer 1), represented by cyanobacteria, followed by CLSB (layer2), PSB (layer 3), GSB (layer 4), and finally SRB (layer 5). The colors within each slice represent the metabolic intersection of the biogeochemical cycles (C, H, O, N, S, P, Fe). Below we explain how this interplay is represented, moving from the outer to the inner cycle. In layer 1, the Cyanobacteria carry out $CO_2$ fixation by Calvin cycle (C), releasing oxygen (O) using water $H_2O$ as electron donor (H). Some Cyanobacteria can perform nitrogen fixation (N), i.e. *Oscillatoria limnetica* that uses $H_2S$ as electron donor. Sulfur's intersection with the P cycle is seen in the biosynthesis of the sulfolipids sulfoquinovosyl diacylglycerol (SQDG) by some Cyanobacteria (*Synechococcus* and *Synechocystis*) and *Bacillus coahuilensis* at CCB (Alcaraz et al., 2008) The CLSB (layer 2 within the mat model) require an inorganic source of energy and obtain their cell carbon from the fixation of carbon dioxide by Calvin cycle (C). Using sulfide both as energy source and electron donor with oxygen (O), nitrate or nitrite as terminal acceptors the nitrogen oxides are reduced to nitrogen (N; denitrification) to obtain energy for growth. For example, *Thiobacillus denitrificans* is able to grow with O, nitrate, or nitrite as terminal electron donors. *Thiobacillus ferrooxidans* reduces ferric iron (Fe) under anaerobic conditions. The PSB (layer 3) use intermediate products of organic matter degradation by cyanobacteria and inorganic

sulfur compounds produced by sulfate-reducing bacteria as sulfide or molecular $H_2$ \ H as electron donors for the light-dependent reduction of $CO_2$ to cell material (C). Synthesis of SQDG is also reported in PSB such as in *Rhodobacter sphaeroides*. In contrast to the PSB, the GSB (layer 4) use the reverse TCA cycle to fix $CO_2$; however, all the biogeochemical intersections are shared as above. Finally, as mentioned above, dissimilatory sulfate reduction by SRB (layer 5) is an energy-gaining process in which the reduction of sulfate to sulfide is linked to either oxidation of an electron donor organic compound (C) or hydrogen (H). Pyrite ($FeS_2$) and iron monosulfide (FeS) are formed as the result of organic matter degradation by sulfate-reducing bacteria. In addition, there are also strains of SRB capable of fixing N using the nitrogenase (*nif*) complex. This entire complexity has been found in each of the sequenced microbial mats and stromatolites at CCB (De Anda 2017, submitted; Desnues et al. 2008; Nitti et al. 2012; Peimbert et al. 2012)

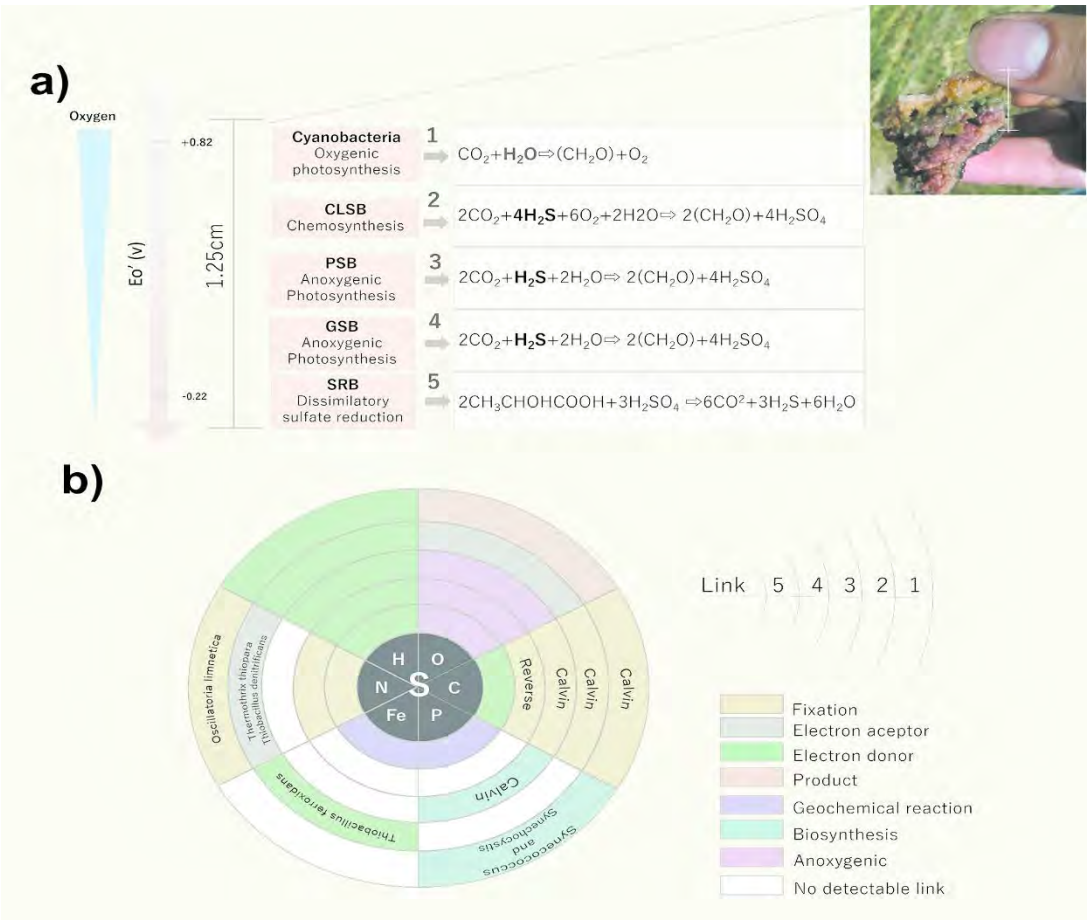



Figure 3. Microbial Mat Model: a) Representation of the vertical stratification of micro-organisms where physiological, metabolic and geochemical interactions according to redox potential are illustrated. b) Diagram of the intersection of the biogeochemical cycles in a microbial mat, where the sulfur cycle is the guiding axis of the intersection. The numbers indicate how microbial mat guilds from top to bottom, 1-5 indicating Cyanobacteria, Color-less sulfur bacteria, Purple Sulfur Bacteria, Green Sulfur Bacteria and finally Sulfate Reducing Bacteria, respectively

## 2.5 Capturing the importance of global sulfur cycle with a single value

The phylogenetic and metabolic complexity of sulfur-related microorganisms that extends over the Tree of Life has hampered their detection using a single gene-targeted marker, especially considering the great breadth of the metabolic pathways involved. Therefore, over the past decades, the use of several marker genes has been crucial to advancing the

understanding of evolutionary ecology of sulfur metabolism in environmental samples (Dar et al. 2007; Frigaard and Bryant 2004; Hügler et al. 2010; Loy et al. 2009; Meyer and Kuever 2007; Vladár et al. 2008; Watanabe et al. 2013). In the current omics-era emerging data have confirmed the importance and widespread distribution of organic sulfur assimilation in the ocean (Delmont et al., 2011; Todd et al., 2011a), sulfur redox metabolism in hot springs (Jiménez et al., 2012; Tang et al., 2013), stromatolites and microbial mats (Breitbart et al., 2009a; Khodadad and Foster, 2012a; Warden et al., 2016), biofilms (Wilbanks et al., 2014), and hydrothermal vents (Nakai et al., 2011). In addition, similar genomics-based studies have highlighted the genetic and functional bases for the recycling of sulfur in oxygen minimum zones (Canfield et al. 2010) and have provided insights into the genetic repertoire of poorly characterized microorganisms that reside deep within Earth's crust and sediments (Jungbluth et al., 2017; Mason et al., 2014). However, despite the great advances in microbial ecology and the explosion of high-throughput sequencing data, our ability to understand and integrate the global biogeochemical cycles is still limited. Recently, we have developed a computational algorithm aimed at classifying, evaluating, and comparing large-scale "omics" samples according to their metabolic machinery. To test our algorithm, we performed a reconstruction of the biogeochemical sulfur cycle, resulting in a comprehensive, manually curated inventory of the biotic players involved. This inventory includes genes, molecular pathways, enzymes, sulfur compounds, and the microorganisms involved (De Anda et al. 2017). To the best of our knowledge, this is the first attempt to use not only individual marker genes but also a comprehensive analysis of the entire sulfur-processing machinery to integrate the biotic and abiotic processes involved in the mobilization of inorganic-organic sulfur compounds through microbial-catalyzed reactions at a global scale. Our algorithm was tested at a multi-genomic scale (including thousands of public completely sequenced genomes and metagenomes). The results strongly highlight the broad ability of our proposed algorithm to accurately detect the enrichment of sulfur metabolism even in genomes derived from environmental samples without any cultured representatives (aka "microbial dark matter"). In this way, we were able to detect the most important organisms and environments worldwide in which the overall sulfur machinery was over-represented.

Consistent with oldest geological record for the emergence of life on Earth, we were able to identify microbial mats from Cuatro Cienegas (See Chapter 7), hydrothermal vents, and hot springs as the current environments that hold the complete repertoire of sulfur metabolic pathways (including the mobilization of organic and inorganic sulfur compounds) in worldwide environments (De Anda et al., 2017).

## 2.6 Conclusions and remarks

The ongoing debates over when life emerged on Earth or which signal is the oldest evidence of life has been subjected to continue renewal of evidence that had placed the origin of life earlier and earlier within geological time scales. However, despite the fragmentary nature of the fossil record, newly obtained geologic evidence indicates that hot springs, microbial mats, and hydrothermal vents were the primordial environments with the required conditions for the beginning of the life based on the sulfur cycle. Sulfur continues to play a critical role in modern analogues, such as the microbial mats from Cuatro Cienegas (see Chapter 8). In agreement with the geological record, molecular and genomic data, along with computational *in silico* modeling, also point out the critical role of sulfur in the beginning of life, suggesting that the use of sulfur compounds was the probably the most ancient metabolism that appeared on Earth, a process that continues to the modern day.

We also recapitulate the great metabolic diversity of the sulfur cycle within the recent tree of life and we use microbial mats as ecological models of early biogeochemical evolution. We can do so because they are nearly closed and self-sustaining ecosystems that encompass the major biogeochemical cycles, trophic levels, and food webs in a vertically laminated pattern. Our conceptualized microbial mat model allows us to highlight the important metabolic links of the major biogeochemical cycles performed by five sulfur metabolic guilds that are widespread in nature and encountered at the millimeter scale within microbial mats. Our emerging comparative framework and emphasis on historical patterns is helping to bridge barriers among organism-based research, community studies, phylogenomic analysis,

mathematical modelling, and paleogeochemical data, pointing out the critical role of sulfur in the early history of life.

## 2.7 References


Alcaraz LD, Olmedo G, Bonilla G, et al (2008) The genome of Bacillus coahuilensis reveals adaptations essential for survival in the relic of an ancient marine environment. Proc Natl Acad Sci U S A 105:5803–5808. doi: 10.1073/pnas.0800981105

Allwood AC, Walter MR, Kamber BS, et al (2006) Stromatolite reef from the Early Archaean era of Australia. Nature 441:714–8. doi: 10.1038/nature04764

Aoyama S, Ueno Y (2018) Multiple sulfur isotope constraints on microbial sulfate reduction below an Archean seafloor hydrothermal system. Geobiology 16:107–120. doi: 10.1111/gbi.12268

Bolhuis H, Cretoiu MS, Stal LJ (2014) Molecular Ecology of Microbial Mats. FEMS Microbiol Ecol 90:335–50. doi: 10.1111/1574-6941.12408

Breitbart M, Hoare A, Nitti A, et al (2009) Metagenomic and stable isotopic analyses of modern freshwater microbialites in {Cuatro Ciénegas, Mexico}. Environ Microbiol 11:16–34

Canfield D, Kristensen E, Bo T (2005) The Sulfur cycle, 1st edn. Elsevier Academic Press

Canfield D, Stewart F, Thamdrup B, et al (2010) A Cryptic Sulfur Cycle in off the Chilean Coast. Science (80- ) 330:1375. doi: 10.1126/science.1196889

Caspi R, Altman T, Dreher K, et al (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 40:D742--53. doi: 10.1093/nar/gkr1014

Dai J (2017) New insights into a hot environment for early life. Environ Microbiol Rep 9:1–26

Dalton R (2004) Fresh study questions oldest traces of life in Akilia rock. Nature 429:688

Dar S a, Yao L, van Dongen U, et al (2007) Analysis of diversity and activity of sulfate-reducing bacterial communities in sulfidogenic bioreactors using 16S rRNA and dsrB genes as molecular markers. Appl Environ Microbiol 73:594–604. doi: 10.1128/AEM.01875-06

De Anda V, Zapata-Peñasco I, Poot-Hernandez AC, et al (2017) MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle Authors. Gigascience 6:1–17. doi: 10.1093/gigascience/gix096

Delmont TO, Malandain C, Prestat E, et al (2011) Metagenomic mining for microbiologists. ISME J 5:1837–1843. doi: 10.1038/ismej.2011.61

Desnues C, Rodriguez-Brito B, Rayhawk S, et al (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452:340–343. doi: 10.1038/nature06735

Dillon JG, Fishbain S, Miller SR, et al (2007) High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. Appl Environ Microbiol 73:5218–5226. doi: 10.1128/AEM.00357-07

Djokic T, VanKranendonk MJ, Campbel KA, et al (2017) Earliest signs of life on land preserved in ca. 3.5 Ga hot spring deposits. Nat Commun 8:1–8. doi: 10.1038/ncomms15263

Dodd MS, Papineau D, Grenne T, et al (2017) Evidence for early life in Earth's oldest hydrothermal vent precipitates. Nature 543:60–64. doi: 10.1038/nature21377

Friedrich CG, Rother D, Bardischewsky F, et al (2001) Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? Appl Environ Microbiol 67:2873–82. doi: 10.1128/AEM.67.7.2873-2882.2001

Frigaard N-U, Bryant D a (2004) Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. Arch Microbiol 182:265–276. doi: 10.1007/s00203-004-0718-9

Frigaard N, Bryant DA (2008) Genomic and Evolutionary Perspectives on Sulfur Metabolism in Green Sulfur Bacteria. In: Dahl C and Cornelius F (ed) Microbial sulfur metabolism, Springer Berlin Heidelberg pp 60–76

Garcia AK, Schopf JW, Yokobori S, et al (2017) Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean. Proc Natl Acad Sci 114:4619–4624. doi: 10.1073/pnas.1702729114

Gaucher EA, Kratzer JT, Randall RN (2010) Deep phylogeny--how a tree can help characterize early life on Earth. Cold Spring Harb Perspect Biol 2:1–16. doi: 10.1101/cshperspect.a002238

Ghosh W, Dam B (2009) Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol Rev 33:999–1043. doi: 10.1111/j.1574-6976.2009.00187.x

Goldford JE, Hartman H, Smith TF, Segrè D (2017) Remnants of an Ancient Metabolism without Phosphate. Cell 168:1126–1134.e9. doi: 10.1016/j.cell.2017.02.001

Grassineau N V., Nisbet EG, Bickle MJ, et al (2001) Antiquity of the biological sulphur cycle: evidence from sulphur and carbon isotopes in 2700 million-year-old rocks of the Belingwe Belt, Zimbabwe. Proc R Soc B Biol Sci 268:113–119. doi: 10.1098/rspb.2000.1338

Harris JK, Caporaso JG, Walker JJ, et al (2013) Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. ISME J 7:50–60. doi: 10.1038/ismej.2012.79

Herman EK and Kump LR (2005) Biogeochemistry of microbial mats under Precambrian environmental conditions : a modelling study. Geobiology 3:77–92

Hoshino T, Kuratomi T, Morono Y, et al (2016) Ecophysiology of Zetaproteobacteria Associated with Shallow Hydrothermal Iron-Oxyhydroxide Deposits in Nagahama Bay of Satsuma Iwo-Jima, Japan. Front Microbiol 6:1554. doi: 10.3389/fmicb.2015.01554

Hubas C, Jesus B, Passarelli C, Jeanthon C (2011) Tools providing new insight into coastal

anoxygenic purple bacterial mats: review and perspectives. Res Microbiol 162:858–868

Hug LA, Baker BJ, Anantharaman K, et al (2016) A new view of the tree of life. Nat Microbiol 1:16048. doi: 10.1038/nmicrobiol.2016.48

Hügler M, Gärtner A, Imhoff JF (2010) Functional genes as markers for sulfur cycling and CO2 fixation in microbial communities of hydrothermal vents of the Logatchev field. FEMS Microbiol Ecol 73:526–37. doi: 10.1111/j.1574-6941.2010.00919.x

Jiménez DJ, Andreote FD, Chaves D, et al (2012) Structural and functional insights from the metagenome of an acidic hot spring microbial planktonic community in the Colombian Andes. PLoS One 7:e52069. doi: 10.1371/journal.pone.0052069

Jungbluth SP, Glavina del Rio T, Tringe SG, et al (2017) Genomic comparisons of a bacterial lineage that inhabits both marine and terrestrial deep subsurface systems. PeerJ 5:e3134. doi: 10.7717/peerj.3134

Khodadad CLM, Foster JS (2012) Metagenomic and metabolic profiling of nonlithifying and lithifying stromatolitic mats of Highborne Cay, The Bahamas. PLoS One 7:e38229. doi: 10.1371/journal.pone.0038229

Koonin E V., Martin W (2005) On the origin of genomes and cells within inorganic compartments. Trends Genet 21:647–654. doi: 10.1016/j.tig.2005.09.006

Lake JA, Skophammer RG, Herbold CW, Servin JA (2009) Genome beginnings: rooting the tree of life. Philos Trans R Soc B Biol Sci 364:2177–2185. doi: 10.1098/rstb.2009.0035

Lau MC, Kieft TL, Kuloyo O, et al (2016) An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. Proc Natl Acad Sci U S A 113:E7927–E7936

Loy A, Duller S, Baranyi C, et al (2009) Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environ Microbiol 11:289–299. doi: 10.1111/j.1462-2920.2008.01760.x

Martin WF, Thauer RK (2017) Energy in Ancient Metabolism. Cell 168:953–955. doi: 10.1016/j.cell.2017.02.032

Mason OU, Scott NM, Gonzalez A, et al (2014) Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. ISME J 8:1464–75. doi: 10.1038/ismej.2013.254

Meyer B, Kuever J (2007) Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5'-phosphosulfate reductase-encoding genes (aprBA) among sulfur-oxidizing prokaryotes. Microbiology 153:3478–3498. doi: 10.1099/mic.0.2007/008250-0

Müller AL, Kjeldsen KU, Rattei T, et al (2015) Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. ISME J 9:1152–1165. doi: 10.1038/ismej.2014.208

Muyzer G, Stams AJM (2008) The ecology and biotechnology of sulphate-reducing bacteria. Nat Microbiol 6:454

Nakai R, Abe T, Takeyama H, Naganuma T (2011) Metagenomic Analysis of 0.2-μm-Passable Microorganisms in Deep-Sea Hydrothermal Fluid. Mar Biotechnol 13:900–908. doi:

10.1007/s10126-010-9351-6
Nitti A, Daniels C a, Siefert J, et al (2012) Spatially resolved genomic, stable isotopic, and lipid analyses of a modern freshwater microbialite from Cuatro Ciénegas, Mexico. Astrobiology 12:685–98. doi: 10.1089/ast.2011.0812
Noffke N, Christian D, Wacey D, Hazen RM (2013) Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the ca. 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia. Astrobiology 13:1103–1124. doi: 10.1089/ast.2013.1030
Nutman AP, Bennett VC, Friend CRL, et al (2016) Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. Nature 537:535–538. doi: 10.1038/nature19355
Olson KR, Straub KD, Straub KD (2016) The Role of Hydrogen Sulfide in Evolution and the Evolution of Hydrogen Sulfide in Metabolism and Signaling. Physiology 31:60–72. doi: 10.1152/physiol.00024.2015
Ozuolmez D, Na H, Lever M, et al (2015) Methanogenic archaea and sulfate reducing bacteria co-cultured on acetate: teamwork or coexistence? Front Microbiol 6:492. doi: 10.3389/fmicb.2015.00492
Peimbert M, Alcaraz LD, Bonilla-Rosso G, et al (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 12:648–58. doi: 10.1089/ast.2011.0694
Pereira I a C, Ramos AR, Grein F, et al (2011) A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. Front Microbiol 2:69. doi: 10.3389/fmicb.2011.00069
Philippot P, Van Zuilen M, Lepot K, et al (2007) Early archaean microorganisms preferred elemental sulfur, not sulfate. Science (80- ) 317:1534–1537. doi: 10.1126/science.1145861
Pinckney JL, Paerl HW (1997) Anoxygenic photosynthesis and nitrogen fixation by a microbial mat community in a bahamian hypersaline lagoon. Appl Environ Microbiol 63:420–6
Plugge C, Zhang W, Scholten J, Stams A (2011) Metabolic Flexibility of Sulfate-Reducing Bacteria. Front Microbiol 2:81. doi: 10.3389/fmicb.2011.00081
Prieto-Barajas CM, Valencia-Cantero E, Santoyo G (2017) Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. Electron J Biotechnol 31:48–56. doi: 10.1016/j.ejbt.2017.11.001
Robertson LA, Kuenen JG (2006) The Colorless Sulfur Bacteria. In: Dworkin M, Falkow S, Rosenberg E, et al. (eds) The Prokaryotes: Volume 2: Ecophysiology and Biochemistry. Springer New York, New York, NY, pp 985–1011
Schopf JW (1993) Microfossils of the Early Archean Apex Chert: New Evidence of the Antiquity of Life. Science (80- ) 260:640–646. doi: 10.1126/science.260.5108.640
Schopf JW, Kitajima K, Spicuzza MJ, et al (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. Proc Natl Acad Sci 115:53–58. doi: 10.1073/pnas.1718063115

Semenov SN, Kraft LJ, Ainla A, et al (2016) Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. Nature 537:656–660. doi: 10.1038/nature19776

Shen Y, Buick R, Canfield D (2001) Isotopic evidence for microbial sulphate reduction in the early Archaea era. Nature 410:77–81

Shen Y, Farquhar J, Masterson A, et al (2009) Evaluating the role of microbial sulfate reduction in the early Archean using quadruple isotope systematics. Earth Planet Sci Lett 279:383–391. doi: 10.1016/j.epsl.2009.01.018

Staley J T. (2002) The Metabolism of Earth's First Organisms. In: American Astronomical Society Meeting Abstracts. p 1221

Tang K, Liu K, Jiao N, et al (2013) Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. PLoS One 8:e72958. doi: 10.1371/journal.pone.0072958

Tashiro T, Ishida A, Hori M, et al (2017) Early trace of life from 3.95 Ga sedimentary rocks in Labrador, Canada. Nature 549:516–518. doi: 10.1038/nature24019

Tice MM, Lowe DR (2004) Photosynthetic microbial mats in the 3 , 416-Myr-old ocean. Nature 431:549–552. doi: 10.1038/nature02920.1.

Todd JD, Curson ARJ, Kirkwood M, et al (2011) DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. Environ Microbiol 13:427–438. doi: 10.1111/j.1462-2920.2010.02348.x

Trudinger PA (1992) Bacterial sulfate reduction: current status and possible origin. Early Org Evol Implic Miner Energy Resour 367–377

Ueno Y, Ono S, Rumble D, Maruyama S (2008a) Quadruple sulfur isotope analysis of ca. 3.5 Ga Dresser Formation: New evidence for microbial sulfate reduction in the early Archean. Geochim Cosmochim Acta 72:5675–5691. doi: 10.1016/j.gca.2008.08.026

Ueno Y, Ono S, Rumble D, Maruyama S (2008b) Quadruple sulfur isotope analysis of ca. 3.5 Ga Dresser Formation: New evidence for microbial sulfate reduction in the early Archean. Geochim Cosmochim Acta 72:5675–5691. doi: 10.1016/j.gca.2008.08.026

van Gemerden H (1993) Microbial mats: A joint venture. Mar Geol 113:3–25. doi: 10.1016/0025-3227(93)90146-M

Van Kranendonk MJ, Deamer DW, Djokic T (2017) Life Springs. Sci Am 317:28–35. doi: 10.1038/scientificamerican0817-28

Van Kranendonk MJ, Philippot P, Lepot K, et al (2008) Geological setting of Earth's oldest fossils in the ca. 3.5 Ga Dresser Formation, Pilbara Craton, Western Australia. Precambrian Res 167:93–124. doi: 10.1016/j.precamres.2008.07.003

Visscher PT, Stolz JF (2005) Microbial mats as bioreactors: populations, processes, and products. Palaeogeogr Palaeoclimatol Palaeoecol 219:87–100. doi: 10.1016/j.palaeo.2004.10.016

Vladár P, Rusznyák A, Márialigeti K, Borsodi AK (2008) Diversity of sulfate-reducing bacteria inhabiting the rhizosphere of Phragmites australis in Lake Velencei (Hungary)

revealed by a combined cultivation-based and molecular approach. Microb Ecol 56:64–75. doi: 10.1007/s00248-007-9324-0

Wächtershäuser G (1990) The case for the chemoautotrophic origin of life in an iron-sulfur world. Orig life Evol Biosph 20:173–176. doi: 10.1007/BF01808279

Wächtershäuser G (2008) Iron-Sulfur World. In: Begley T. (ed) Wiley Encyclopedia of Chemical Biology. American Cancer Society, pp 1–8

Wagner M, Roger AJ, Flax JL, et al (1998) Phylogeny of Dissimilatory Sulfite Reductases Supports an Early Origin of Sulfate Respiration. J Bacteriol 180:2975–2982

Walter MR, Buick R, Dunlop JSR (1980) Stromatolites 3,400-3,500 Myr old from the North Pole area, Western Australia. Nature 284:443–445

Warden JG, Casaburi G, Omelon CR, et al (2016) Characterization of Microbial Mat Microbiomes in the Modern Thrombolite Ecosystem of Lake Clifton , Western Australia Using Shotgun Metagenomics. Front Microbiol 11:1–14

Watanabe T, Kojima H, Takano Y, Fukui M (2013) Diversity of sulfur-cycle prokaryotes in freshwater lake sediments investigated using aprA as the functional marker gene. Syst Appl Microbiol 36:436–443

Westall F, Vries ST, Nijman W, et al (2006) The 3.466 Ga “Kitty’s Gap Cheil,” an early Archean microbial ecosystem. GSA Spec. Pap. 405:105–131

Wilbanks EG, Jaekel U, Salman V, et al (2014) Microscale sulfur cycling in the phototrophic pink berry consortia of the Sippewissett Salt Marsh. Environ Microbiol 16:3398–3415. doi: 10.1111/1462-2920.12388

Zhelezinskaia I, Kaufman AJ, Farquhar J, Cliff J (2014) Large sulfur isotope fractionations associated with Neoarchean microbial sulfate reduction. Science (80- ) 346:742–744. doi: 10.1126/science.1256211

Capítulo 2

# B. Azufre: elemento incomprendido de la biogeoquímica planetaria

Valerie De Anda y Valeria Souza

A continuación, se muestra el trabajo de divulgación publicado en la revista de divulgación OIKOS del instituto de Ecología UNAM.

## Referencia

De Anda V y Souza V. Azufre: elemento incomprendido de la biogeoquímica planetaria En La Química de la Vida No. 16 Oikos Revista de divulgación del Instituto de Ecología UNAM

## Para acceder al texto en línea

http://web.ecologia.unam.mx/oikos3.0/index.php/todos-los-numeros/articulos-anteriores/187-azufre-y-biogeoquimica-planetaria

## Para obtener la versión en pdf del número completo

http://web.ecologia.unam.mx/oikos3.0/index.php/todos-los-numeros/2010-presente

*Artículo*

# Azufre: elemento incomprendido de la biogeoquímica planetaria

Valerie de Anda y Valeria Souza

**Recapitulando: la formación de los átomos en el universo, la Tierra primitiva y los ciclos biogeoquímicos**

Milisegundos después del *Big Bang*, el universo era una sopa caliente de plasma de partículas elementales que dieron lugar a toda la materia en el universo. Minutos después, ya se habían formado todos los neutrones y protones los cuales comenzaron a fusionarse entre sí, formando núcleos atómicos en un proceso conocido como nucleosíntesis del *Big Bang*. Trescientos mil (300,000) años después, el universo ya se había enfriado significativamente como para que los electrones se combinaran con los protones y neutrones formando los primeros átomos de helio e hidrógeno. Los gases de estos átomos, dieron lugar a nubes gigantes que al colapsar formaron las primeras estrellas y galaxias, generando en su interior los átomos esenciales para la vida en la Tierra: carbono, oxígeno, nitrógeno, azufre y fósforo (CHONP). Los primeros en formarse en el interior de las estrellas fueron el carbono y oxígeno, y los ciclos biogeoquímicos de estos dos elementos se conocen con detalle (ver en esta edición a Campo y a Souza y Viladomat-Jasso). Pero existen átomos más complejos, como el azufre, que se formaron posteriormente en las estrellas. La complejidad de los átomos se da al aumentar el número de electrones, ya que éstos les confieren la posibilidad de asociarse con otros elementos. Es decir, los electrones disponibles les confieren mayores posibilidades de asociarse con otros elementos, porque estos electrones son los responsables de la interacción entre los átomos: mientras más haya, aumentan los tipos de enlaces posibles que pueden formarse. Debido a esto, entender cómo se reciclan estos átomos "complejos" en nuestro planeta no ha sido algo trivial, ya que tienen la capacidad de interaccionar con una gran cantidad de átomos y elementos diferentes (ver glosario).

Los átomos que constituyen la vida, y que se formaron en las estrellas, se han reciclado en nuestro planeta y continúan haciéndolo por medio de diferentes procesos como son: 1. el movimiento del magma que proviene del centro de la Tierra 2. la tectónica de placas y 3. las reacciones que ocurren en la atmósfera debido al sol. Además, el reciclamiento de los átomos en la Tierra no sólo se ha dado por procesos abióticos, si no que fundamentalmente se ha llevado a cabo por los procesos metabólicos derivados de los organismos más pequeños: las bacterias y arqueas. Estos microorganismos procariontes, fueron las primeras formas de vida que aparecieron en la Tierra hace 3,800 millones de años, y con el paso del tiempo, evolucionaron hasta cambiar la química del planeta para siempre. Lo anterior dio lugar a los ciclos biogeoquímicos que son la razón fundamental de la existencia de la vida en el planeta Tierra, tal y como la conocemos hoy en día.

**Las primeras formas de vida en la Tierra se basaron en compuestos de azufre**

El planeta Tierra se formó hace aproximadamente 4,600 millones de años y por términos prácticos, la historia de la vida se divide en cuatro grandes eones: Hadeano, Archeano, Proterozoico y Fanerozoico (ver Figura 1). En el Hadeano (nombrado así por el Dios griego del inframundo Hades), la Tierra era excesivamente caliente, sin oxígeno, en donde una cantidad inimaginable de meteoritos hubiera destruido cualquier intento de vida, sin embargo, dichos impactos pudieron haber traído los compuestos esenciales para el inicio de la vida en la Tierra (moléculas orgánicas y posiblemente agua). La evidencia científica indica que la vida empezó al principio del eón Arqueano (3,800 millones de años), en un mar caliente sin oxígeno con abundante hierro. Las teorías más recientes indican que posiblemente la vida se originó en ambientes parecidos a las ventilas hidrotermales, las cuales son creadas a lo largo de las líneas de separación de las placas tectónicas en el fondo marino.

Existen principalmente dos tipos de ventilas hidrotermales, clasificadas por color del efluente: negro o blanco. Las ventilas o fumarolas negras, se encuentran cerca de los centros de expansión donde el magma calienta el agua de mar que se filtra a la corteza terrestre, emitiendo agua caliente (300-400°C), ácida (*pH* 2-3), rica en azufre reducido ($H_2S$ 3-110 mm/kg), dióxi-

| Glosario |
|---|
| ***Big Bang*:** Explosión cósmica de una muy pequeña cantidad de materia a grandes temperaturas que ocurrió hace 20 mil millones y que dio origen al universo. |
| **Partículas elementales:** Constituyentes más básicos de la materia que transmiten las interacciones fundamentales. |
| **Partículas subatómicas:** Partículas que constituyen los átomos, y que están formadas por algunas partículas elementales. |
| **Protón:** Partícula subatómica que se encuentra en el núcleo atómico y que tiene carga positiva. El número de protones en un átomo determina el elemento. |
| **Neutrón:** Partícula subatómica que se encuentra en el núcleo atómico; a diferencia del protón, no tiene carga eléctrica neta. El número de neutrones en un núcleo atómico determina el isótopo de ese elemento. |
| **Electrón:** partícula subatómica con carga negativa que se encuentra fuera del núcleo atómico. |
| **Nucleosíntesis:** Proceso por el cual protones y neutrones se fusionan para crear nuevos núcleos atómicos. Dependiendo el momento y el lugar donde se lleve a cabo, la nucleosíntesis se divide en 4: *i)* del *Big Bang* (primeros 3 minutos después de la formación del universo), *ii)* estelar (dentro de las estrellas donde se formaron los elementos esenciales para la vida, *iii)* explosiva (supernovas, donde se producen los elementos más pesados que el He) y *iv)* a espalación de rayos cósmicos (elementos ligeros). |
| **Átomo:** Partícula más pequeña en la que un elemento puede ser dividido. El átomo es un constituyente de la materia que está conformado por partículas subatómicas. |
| **Elemento:** Átomos que tienen el mismo número de protones en su núcleo (conocido como número atómico z) y que se encuentran en la tabla periódica. |
| **Arqueas:** Organismos unicelulares sin núcleo definido (procariontes) que difieren de las bacterias y eucariontes y que generalmente viven en ambientes extremos. |

do de carbono (4-215 mm/kg), hidrógeno disuelto $H_2$ (0.1-50 mm/kg) y metales de transición, en especial hierro ($Fe^{+}_{2}$). El azufre y el hierro producido en estas ventilas hidrotermales, reacciona para formar sulfuro de hierro (FeS), el cual precipita en contacto con el agua oxigenada del mar, generando el color negro característico de la ventila hidrotermal (ver en esta edición Zapata-Peñasco). Dichas características de las ventilas hidrotermales en donde se combina una alta presión y gran cantidad de calor, favorece reacciones químicas que no son posibles bajo otras condiciones. El calor de las ventilas favorece que los metales de sulfuros se mantengan en solución, y el *pH* favorece que se disocien, por tanto, aumentan las concentraciones de sulfuro de hidrogeno ($H_2S$). Este mismo escenario parece haber sido el de la Tierra primitiva, en donde el $H_2S$ y los sulfuros de metales reducidos, permanecían en solución en un mar sin oxígeno (anóxico) durante largo tiempo albergando las primeras formas de vida, cuyos metabolismos, en ausencia de oxígeno y en un ambiente hipercaliente, dependían del azufre para obtener energía.

En el árbol de la vida se encuentra evidencia de esta teoría, en donde los linajes más ancestrales son arqueas hipertermófilas que utilizan compuestos reducidos de azufre, así como hidrógeno molecular. Estos primeros organismos que respiraron azufre, fueron los predecesores de todos los metabolismos que surgieron posteriormente en la Tierra primitiva, como lo son la fotosíntesis anoxigénica llevada a cabo por bacterias púrpuras y por las verdes del azufre, así como la fotosíntesis oxigénica que por primera vez la llevaron a cabo las cianobacterias (Figura 1) que se describen con detalle en el artículo de Souza y Viladomat-Jasso de este número de *Oikos*=.

**La sucesión de eventos en la historia de la vida en la Tierra y los tapetes microbianos**

El dilucidar la historia de la vida en la Tierra primitiva no ha sido, ni es, tarea fácil. Para poder entender los eventos involucrados en la aparición de la vida en la Tierra y los consecuentes metabolismos anaerobios que respiraron azufre, hasta la fotosíntesis oxigénica, ha sido necesaria una gran cantidad de pruebas moleculares: fisicoquímicas, geoquímicas, fósiles, etcétera. Sin embargo, una evidencia muy clara de esta sucesión de eventos en la historia de la vida de la Tierra se encuentra en estructuras microbianas actuales que existen en muy pocos lugares en el mundo como Australia, Francia y México (Cuatro Ciénegas, Guerrero Negro y Puebla). Estas estructuras se conocen como microbialitos o tapetes microbianos (Figura 2).

El término microbialito se refiere a las estructuras órgano-sedimentarias que forman láminas arregladas de manera vertical que se desarrollan en interfases sólido/agua (Figura 2). Cuando

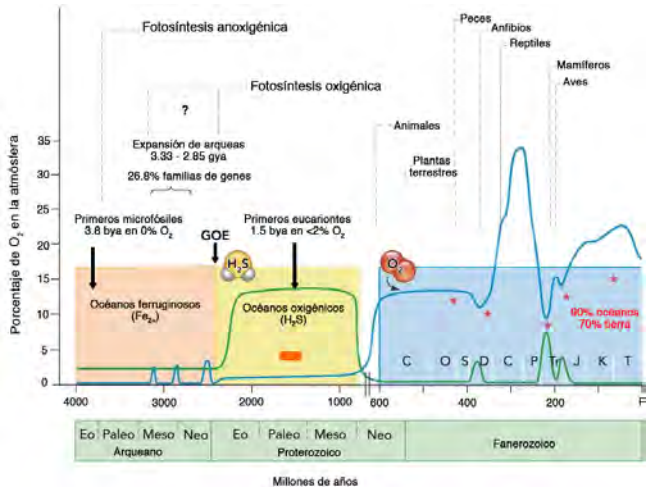



Figura 1. Línea del tiempo de la evolución relacionada a la atmósfera oxigénica ($O_2$, línea azul), y sulfuro de hidrógeno ($H_2S$, línea verde). Modificado de K.R. Olson y K.D. Straub (2016).

precipitan minerales como los silicatos o carbonatos de su entorno, se convierten en roca y son denominados estromatolitos. Cuando no precipitan minerales de su entorno son denominados tapetes microbianos. Los microorganismos que habitan los tapetes microbianos viven intercambiando entre sí nutrientes, materia orgánica y compuestos de azufre, lo que contribuye a que exista una fina formación de capas o estratificación vertical, donde se llevan a cabo la mayoría de los ciclos biogeoquímicos, de los cuales el ciclo del azufre es el eje rector de todos los demás. La organización de los microorganismos, de abajo hacia arriba, recapitulan la historia de la vida en la Tierra en tan sólo unos cuantos centímetros (Figura 2).

Durante más de dos décadas, la falta de métodos y herramientas adecuadas dificultó entender la diversidad microbiana en los tapetes. A principios de los años ochenta, el uso de microsensores de alta resolución (Figura 3), ayudó para caracterizar el medio ambiente físico y químico a pequeña escala espacial. El uso de esta tecnología resaltó la importancia de los gradientes de azufre y oxígeno, e identificó la reducción de sulfato y la fotosíntesis como principales procesos geoquímicos de los tapetes microbianos.

En los noventas, el uso de técnicas moleculares y microsensores de alta resolución (Figura 3) ayudó a estimar la estructura de la comunidad de muestras ambientales y reveló una inesperada gran diversidad microbiana, contrario a lo que se había observado con técnicas que dependen de cultivos y microscopía. A principios de siglo XXI, el uso de marcadores moleculares abrió la posibilidad de estudiar el potencial metabólico de los tapetes microbianos y resaltó la importancia de la reducción de sulfato y la fijación de nitrógeno. De igual manera, la combinación de medidas radiométricas, perfiles químicos y análisis filogenéticos sirvió para estudiar lo que hacen algunos tipos de bacterias en particular. Actualmente, técnicas como la metagenómica (descrita ampliamente en la revisión de S. Nikolaki y G. Tsiamis de 2013), que consiste en conocer la secuencia del ADN

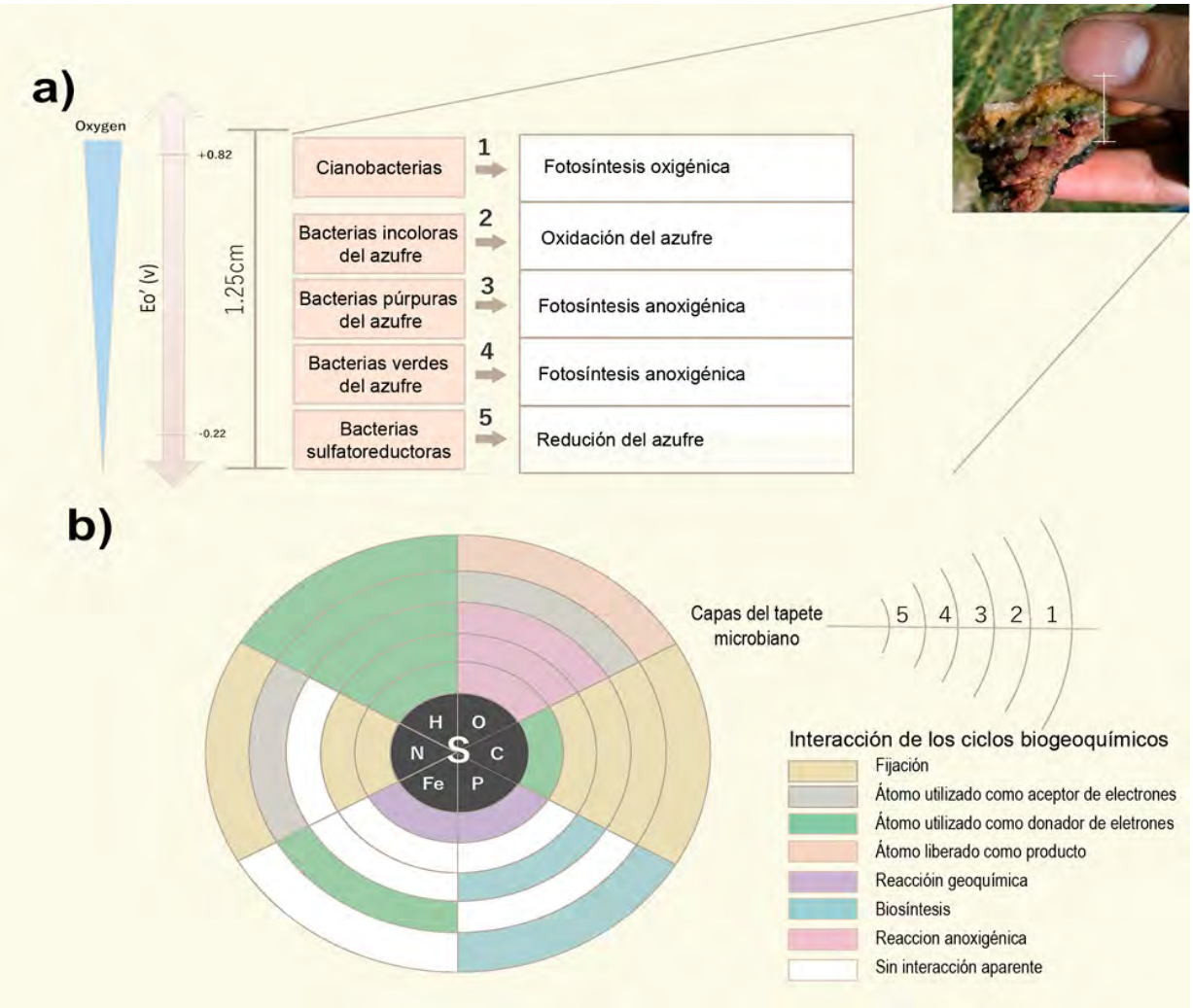



Figura 2. Modelo de un tapete microbiano. a) Representación de la estratificación vertical de los microorganismos del tapete donde se ilustra la interacción fisiológica, metabólica y geoquímica de acuerdo con el potencial redox. b) Esquema de la intersección de los ciclos biogeoquímicos en el tapete microbiano, donde el ciclo del azufre es el eje rector de la intersección.

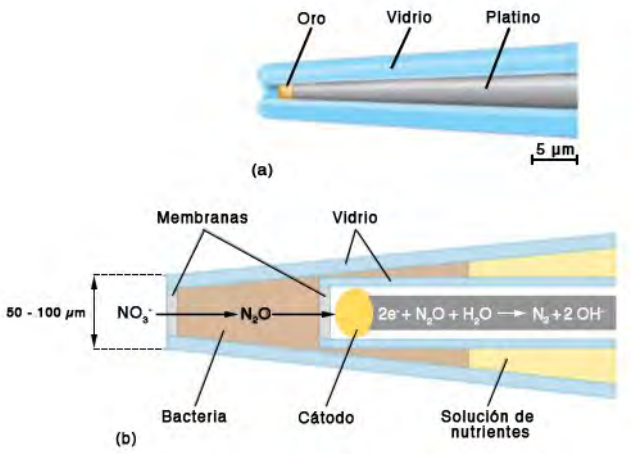



Figura 3. Esquema de un microelectrodo de oxígeno. La punta de platino funciona como un cátodo, por lo que cuando se aplica un voltaje, el $O_2$ es reducido a $H_2O$ generando corriente. La corriente resulta de la reducción de $O_2$ en la superficie de oro del cátodo y es proporcional al $O_2$ que contenga la muestra ambiental. b) Microsensor biológico para la detección de nitrato $NO_3$-. Las bacterias inmovilizadas en la punta del sensor denitrifican $NO_3$ a $N_2O$, el cual es detectado por la reducción de $N_2$ en el cátodo. Modificado de Brock *Biology of Microorganisms*, 13° ed.

de muestras obtenidas directamente del ambiente, han abierto la posibilidad de responder preguntas sobre las comunidades de microbios tales como ¿quién vive ahí? (potencial taxonómico) y ¿qué función llevan a cabo? (potencial metabólico). Así pues, los datos metagenómicos junto con información sobre genes marcadores moleculares, han ayudado a confirmar que los microrganismos que viven en los tapetes microbianos asemejan la historia de la vida en la Tierra, como se observa en la Figura 2a. El gradiente de oxígeno en los tapetes microbianos nos ayuda a entender lo que sucedió a lo largo de la historia de la vida en nuestro planeta. Primero había una atmósfera reductora, sin oxígeno, con condiciones extremas que permitieron el desarrollo de organismos que respiraban azufre, en forma de azufre elemental y sulfato. Estos organismos, son las mencionadas arqueas hipertermófilas y las bacterias sulfato reductoras (SRB por sus siglas en inglés). Estas últimas se encuentran en la parte más profunda de un tapete microbiano conviviendo con arqueas metanógenas. En la capa siguiente se encuentran las bacterias púrpuras y verdes del azufre (PSB y GSB por sus siglas en inglés respectivas) que necesitan el $H_2S$, producido por las sulfato reductoras. Estas bacterias púrpuras y verdes del azufre son capaces de hacer fotosíntesis sin necesidad de oxígeno, y a su vez producen sulfato que, como ya mencionamos, lo respiran las vecinas de abajo: las sulfato reductoras (ver Figura 2). Arriba de las PSB y GSB se encuentran las bacterias que oxidan compuestos de azufre, pero lo hacen en presencia de oxígeno y son conocidas como bacterias incoloras del azufre o CLSB (por sus siglas en inglés). Finalmente, la capa más superficial del tapete microbiano la dominan las cianobacterias, que son las responsables de haber cambiado la atmósfera reductora a una oxidante (Véase artículo de Souza y Viladomat-Jasso). Todos estos microrganismos que son esenciales para llevar a cabo el ciclo del azufre a escala global, se encuentran distribuidas en diferentes ambientes del planeta Tierra (ríos, mares, lagos, sedimentos, etcétera).

En la Figura 2b se representa lo que hemos descubierto sobre la interacción de los ciclos biogeoquímicos, cada rebanada en forma de triángulo esquematiza la relación del ciclo del azufre con otros elementos, y los círculos simbolizan las capas de los microorganismos en los tapetes microbianos, de la más profunda hasta la superficial. Las cianobacterias fijan al $CO_2$ por medio del ciclo de Calvin (y eso implica una interacción con el ciclo del carbono), liberando oxígeno (O) y utilizando agua como donador de electrones (H). Algunas cianobacterias como *Oscilatoria limnetica* pueden fijar nitrógeno y utilizar compuestos de azufre como fuente de energía. La relación del ciclo del azufre con el fósforo (P), la observamos en la membrana de algunas cianobacterias como *Synechococcus* y *Sayne-chocystis*, ya que sintetizan compuestos de azufre (sulfolípidos), en lugar de fosfolípidos, como respuesta a presiones selectivas de su ambiente. Las bacterias incoloras del azufre (por ejemplo, *Thiobacillus denitrificans* o *Thiobacillus ferooxidans*) que se encuentran en el segundo círculo (CLSB), también utilizan el ciclo del carbono ya que también fijan $CO_2$ a través del ciclo de Calvin, requiriendo un compuesto de azufre como fuente de energía, y además necesitan oxígeno (O), y nitrato o nitrito (N). En el tercer círculo están las bacterias púrpuras del azufre (PSB). Éstas necesitan compuestos reducidos de azufre (como $H_2S$) que producen las bacterias sulfato reductoras, y usan hidrógeno molecular para biosintetizar material celular a partir del $CO_2$. Al igual que las cianobacterias, la síntesis de sulfolípidos también está reportado en las PSB como *Rhodobacter sphaeroides*. En el cuarto círculo del modelo del tapete microbiano están representadas las bacterias verdes del azufre que, a diferencia de las PSB, utilizan el ciclo de los ácidos tricarboxílicos reverso para biosintetizar material celular usando $CO_2$, pero el patrón del ciclo del azufre con otros elementos es similar al de las PSB. Finalmente, también en la Figura 2b, se representa a las bacterias que reducen sulfato, proceso mediante el cual obtienen energía reduciendo el compuesto de azufre, un proceso que está ligado a la oxidación de compuestos orgánicos que contienen carbono (C) o hidrógeno (H).

**Tapetes microbianos para entender la importancia del azufre en la Tierra**

Bajo este esquema, nuestro trabajo de investigación propone utilizar al tapete microbiano como un modelo que nos permita entender la interacción biogeoquímica, molecular y ecológica de los microrganismos y el ciclo del azufre, usando así una novedosa perspectiva para analizar a los distintos ciclos biogeoquímicos.

Para generar un algoritmo computacional utilizamos el

orden y los microorganismos de cada una de las capas del tapete microbiano, que está determinado por el gradiente de azufre (Figura 2). Consideramos que éste es un modelo que representa a los ciclos biogeoquímicos, y con base en esta información generamos el algoritmo que nos permite entender el ciclo del azufre a escala global desde perspectivas diferentes: molecular, ecológica y biogeoquímica. Como parte de nuestra investigación realizamos un análisis bibliográfico para recopilar la información de todos los genomas secuenciados hasta el momento de arqueas y bacterias (incluyendo todos los organismos del azufre, PSB, GSB, CLSB, SRB) y todos los genomas de ambientes naturales secuenciados hasta ahora (incluyendo los tapetes microbianos y estromatolitos de Cuatro Ciénegas, Francia, Australia, Chile, etcétera) que se encuentran en las bases de datos públicas como el Centro Nacional para la Información Biotecnológica (NCBI por sus siglas en inglés). Desarrollamos un algoritmo que describe la importancia del ciclo del azufre, a escala global, por medio de un valor numérico. De esta manera podemos conocer qué tan importante es un ambiente determinado u organismo para el ciclaje del azufre, sin contar con ninguna prueba anterior, más que conociendo su genoma o en el caso de un ambiente, el conjunto de genomas o metagenoma. Nuestros resultados comprueban la importancia de los microorganismos que se encuentran en los tapetes microbianos y estromatolitos: todos los genomas del azufre secuenciados hasta ahora, tienen un valor muy alto en nuestro algoritmo. Además, los ambientes en donde el ciclo del azufre es más importante, se encuentran en los tapetes microbianos de Cuatro Ciénegas y en manglares de Brasil, confirmando que estos ambientes son claves para el reciclaje completo de las formas orgánicas e inorgánicas de azufre.

Con este algoritmo no sólo se puede estudiar el ciclo del azufre, sino que está diseñado para que se entienda cualquier ciclo biogeoquímico de nuestro planeta, desde un punto de vista ecológico, molecular y geoquímico. De esta forma se pueden generar valores numéricos de ciclos biogeoquímicos con los que es posible hacer análisis que integran la información de forma global, y detectar así ambientes clave para el ciclaje de los elementos esenciales para la vida en el planeta. Además, también se puede analizar y dar seguimiento en la misma escala global a las vías metabólicas de ciclos biogeoquímicos específicos. Por ejemplo, en el caso particular del ciclo del carbono, estamos trabajando con nuestro algoritmo para que sea capaz de detectar aquéllos ambientes importantes en los que está ocurriendo la degradación de hidrocarburos o compuestos aromáticos tóxicos (como el benceno nitrotolueno que contienen algunas gasolinas). De esta forma es posible encontrar en qué ambientes del mundo existen organismos potencialmente útiles para procesos como la biorremediación, entre otras cosas.

Pero, ¿por qué estudiar el ciclo del azufre en una escala global? y ¿por qué es necesario un algoritmo que detecte la importancia del ciclo del azufre?

**Las propiedades fisicoquímicas del azufre complican su entendimiento como ciclo biogeoquímico**

El azufre es un elemento no metal que se encuentra en la sexta familia de la tabla periódica, justo debajo del oxígeno. El azufre tiene un número atómico de 16 (es decir, tienen 16 protones) y una masa atómica de 32 (suma de protones y neutrones). En cuanto a su configuración electrónica 1s22s22p63s23p4, tiene seis electrones de valencia, los cuales son los responsables de la interacción entre átomos, es decir, el número potencial de enlaces con otros elementos químicos.

El azufre, a diferencia de los otros elementos que componen la vida en la tierra (CHONP), tiene ocho estados de oxidación, lo cual indica que el azufre puede tomar diferentes "formas químicas" que se pueden mover en la Tierra de manera abiótica y biótica por diferentes grupos bacterianos que son muy divergentes en términos evolutivos, ecológicos y metabólicos (las bacterias que encontramos en los tapetes microbianos).

Dependiendo del enfoque con el que se estudie el ciclo del azufre, se pueden estudiar sus transformaciones de manera parcial o general. Por ejemplo, la Oceanografía y la Geología sólo estudian los procesos meramente abióticos. Por otra parte, desde la perspectiva de la Ecología Microbiana, estudiamos los procesos y las relaciones entre los diferentes organismos que metabolizan compuestos de azufre, y el ambiente donde se llevan a cabo, involucrando tanto procesos bióticos como abióticos. De esta forma, reuniendo la información de los enfoques de Oceanografía, Geología y Ecología Microbiana, hoy sabemos que el ciclo del azufre influye en la respiración de la materia orgánica sedimentaria, el estado de oxidación de la atmósfera y los océanos, y la composición del agua de mar. Lo anterior está íntimamente relacionado con los otros ciclos biogeoquímicos (como por ejemplo el carbono). Sin embargo, debido a que los diferentes compuestos de azufre son movilizados por una gran cantidad de grupos bacterianos, el ciclo del azufre sólo se estudia por partes. Es decir, existen grupos de investigación que están enfocados en entender procesos, vías metabólicas y grupos bacterianos específicos de cada parte del ciclo del azufre. Por ejemplo, analizan la distribución de las bacterias sulfato reductoras por medio de un solo gen. Por lo tanto, los intentos por entender el ciclo del azufre, sólo han reconstruido pequeños fragmentos del rompecabezas. Debido a esto, existe poca información acerca de los patrones de diversidad y distribución de los microorganismos que llevan a cabo las reacciones *redox* del ciclo del azufre, y por lo tanto se desconocen los factores ambientales precisos que regulan su abundancia relativa. Aunado a esto, no hay evidencia directa del ciclaje activo a gran escala de este elemento. Lo anterior resalta la importancia de nuestro algoritmo que se apoya en la información de los tapetes microbianos, los cuales no sólo recapitulan la historia de la vida en la Tierra, sino que también llevan a cabo todos sus

ciclos biogeoquímicos. El utilizarlos como modelos de estudio es esencial para entender la vida en el pasado, y también para comprender la biogeoquímica del planeta actual, con ayuda de técnicas actuales.

Como ya lo hemos mencionado, el azufre ha sido y es un elemento esencial para la vida en el planeta. Lo encontramos en los aminoácidos esenciales como la cisteína y la metionina, y también en coenzimas y cofactores que se encuentran en todos los seres vivos (biotina, la tiamina, la coenzima A, y los *clusters* de hierro y azufre). Las proteínas que necesitan estos cofactores son esenciales en una gran variedad de procesos metabólicos y celulares fundamentales, como la fijación de carbono, la asimilación de lípidos y carbohidratos, la síntesis de pentosas, la duplicación celular, la replicación del ADN (ácido desoxirribonucleico), la fotosíntesis, la respiración y la regulación génica.

Sin el átomo de azufre no se podrían llevar a cabo procesos celulares indispensables para la vida en la Tierra, por lo que su importancia no sólo radica como un elemento fundamental en la biogeoquímica del planeta, sino también como un elemento clave para mantener funcionando la maquinaria celular de eucariontes y procariontes.

Lo anterior nos hace preguntarnos ¿la vida podido evolucionar en el universo si los valores de las constantes fundamentales o las condiciones iniciales del *Big Bang* hubieran sido diferentes?, ¿qué consecuencias hubiera tenido en la Tierra si el orden de aparición de los elementos en las estrellas hubiera sido diferente?, o ¿si las propiedades de los elementos que conforman la vida en la Tierra hubieran tenido un cambio menor en su termodinámica?, ¿qué hubiera pasado con el metabolismo y la evolución de la vida en la Tierra?. Quizás nunca lo sabremos, pero nada nos cuesta imaginar.

**Valerie de Anda.** Es Bióloga por la Facultad de Ciencias de la UNAM. Realizó su tesis de investigación en el Instituto de Biotecnología de la UNAM sobre la evolución dirigida de proteínas recombinantes. Actualmente imparte clases de microbiología en la Facultad de Ciencias UNAM, y estudia el doctorado en el Instituto de Ecología de la UNAM.

**Valeria Souza.** Es investigadora del Laboratorio de Evolución Molecular y Experimental del Departamento de Ecología Evolutiva. Estudia la ecología evolutiva de los microorganismos. Su trabajo ha sido reconocido con diversos premios nacionales.

**Para saber más**

• De Anda Torres V., C. Poot-Hernandez, I. Contreras-Moreira, L. Eguiarte y V. Souza. Enviado a *Environmental Microbiology*. A new multigenomic approach for the study of biogeochemical cycles at global scale: the complete molecular reconstruction of the sulfur cycle.

• Nikolaki S, Tsiamis G. 2013 Microbial diversity in the era of omic technologies. *Biomed Res Int*. 2013;2013:958719. doi: 10.1155/2013/958719. Epub 2013 Oct 24. Review. PubMed PMID: 24260747; PubMed Central PMCID: PMC3821902.

• Olson, K.R. y K.D. Straub. 2016. The Role of Hydrogen Sulfide in Evolution and the Evolution of Hydrogen Sulfide in Metabolism and Signaling. *Physiology* (Bethesda). 31: 60-72. doi: 10.1152/physiol.00024.2015. Review. PubMed PMID: 26674552.

Capítulo 3

# Towards a comprehensive understanding of environmental perturbations in microbial mats from Cuatro Cienegas by network modeling

Valerie De Anda, Icoquih Zapata-Peñasco and Valeria Souza

Abstract

The Cuatro Cienegas Basin (CCB) encompasses hundreds of aquatic systems that harbor diverse microbialites with different community structure com-position and with the highest level of endemism in North America. Thus, CCB represents a desert oasis of high biodiversity. Despite the great importance of this unique site, increasing demand on water for agricultural development (forage and feed livestock) was first manifested with extraction of groundwater in 2011, starting the drying process of aquifers and desertification of the Churince Lagoon. As consequence, water levels have been drastically fluctuating, affecting all ecosystem functions. This chapter reviews a recent network-based approach used to understand how the anthropogenic disturbances affect one of the most resistant microbial communities since the Archean, microbial mats.

## 3.1 General overview of CCB hydrogeology and its aquifer overexploitation.

The Cuatro Cienegas Basin (CCB) is an extremely low-P aquatic environment in a naturally isolated valley in the Chihuahuan Desert (Coahuila, México). CCB is listed as a Wetland of International Importance under the International RAMSAR Convention, an area of protection of flora and fauna (ANP) under Mexican government, and a priority site for Conservation of Nature by the World-Wide Fund for Nature and UNESCO (Souza et al. 2012). CCB has complex hydrogeological characteristics that have been shaped by basin's tectonic activity. Mesozoic faults associated with the opening of the Gulf of Mexico, created conditions that permitted the deposition of Creta-ceous carbonate rocks that now form a regional carbonate aquifer within CCB (Wolaver et al. 2012). According to isotopic data, this aquifer is affect-ed by the presence of older deeply penetrating faults that provide endogen-ic fluids from Earth's mantle and lower crust to surface water, producing the unique physicochemical characteristics of the current aquatic environments (Wolaver and Diehl 2010; Wolaver et al. 2012).

The aquatic systems within CCB exhibit two distinct classes of wa-ter that reflect the origin of that water (See the hydrogeological conceptual model described in detail in Wolaver et al. 2012). The first class includes Ca-SO4 rich waters derived from a regional carbonate aquifer discharge mixed with contributions from deeply sourced fluid from the mantle that ascend along basement-involved faults (e.g. La Becerra and Churince). The second class involves CaHCO3 waters with contributions made via the carbonate aquifer mixed with locally recharged mountain precipitation (e.g. Santa Tecla).

In recent decades, reduced deep-water inputs associated with intensively irrigated crops grown for forage and livestock feed prompted the desiccation of one of the principal aquatic systems within CCB, the Churince drainage. This anthropogenic disturbance has led to dramatic shifts in habi-tat conditions of the system, affecting turtles, fishes (Carson et al. 2015; Hernández et al. 2017), and microbial mats (De Anda et al., under review).

## 3.2 The Churince system a scenario for our study.

The Churince system is unique within the CCB because it is surrounded by pure gypsum (calcium sulfate) dunes that date to the Jurassic period. It contains no calcium carbonate deposits, indicating that this marine sediment was never buried by recent sediments (Wolaver et al., 2012) or organic matter from forest litter (Minckley and Jackson, 2007). This is particularly relevant in the Churince system since, as mentioned above, the water derived from the aquifer is shaped by deep water fluids influenced by mantle geochemistry (Wolaver et al., 2012; Wolaver and Diehl, 2010). Previous to 2006, the Churince system was represented by: a spring and a river system that flowed freely downstream, an intermediate lagoon with constant water supply, and finally a large desiccation lagoon (Laguna Grande), the depth of which varied with evaporation (Cerritos et al., 2011).

However, in summer 2007, Laguna Grande disappeared and the Laguna Intermedia started to have strong seasonal water fluctuations with extensive dessication during summer and recovery in winter due to reduced evaporation (Souza et al., 2007). Laguna Intermedia has experienced extensive water depletion during hot months since summer 2011. As observed in Figure 1, Lagunita Pond (lateral to Laguna Intermedia (see Chapter 4) started to display a greater desiccation during autumn from 2014 to 2017,and during 2016 and 2017 for the first time in our records, we observed a long period of complete desiccation in which water levels recover in spring and summer.

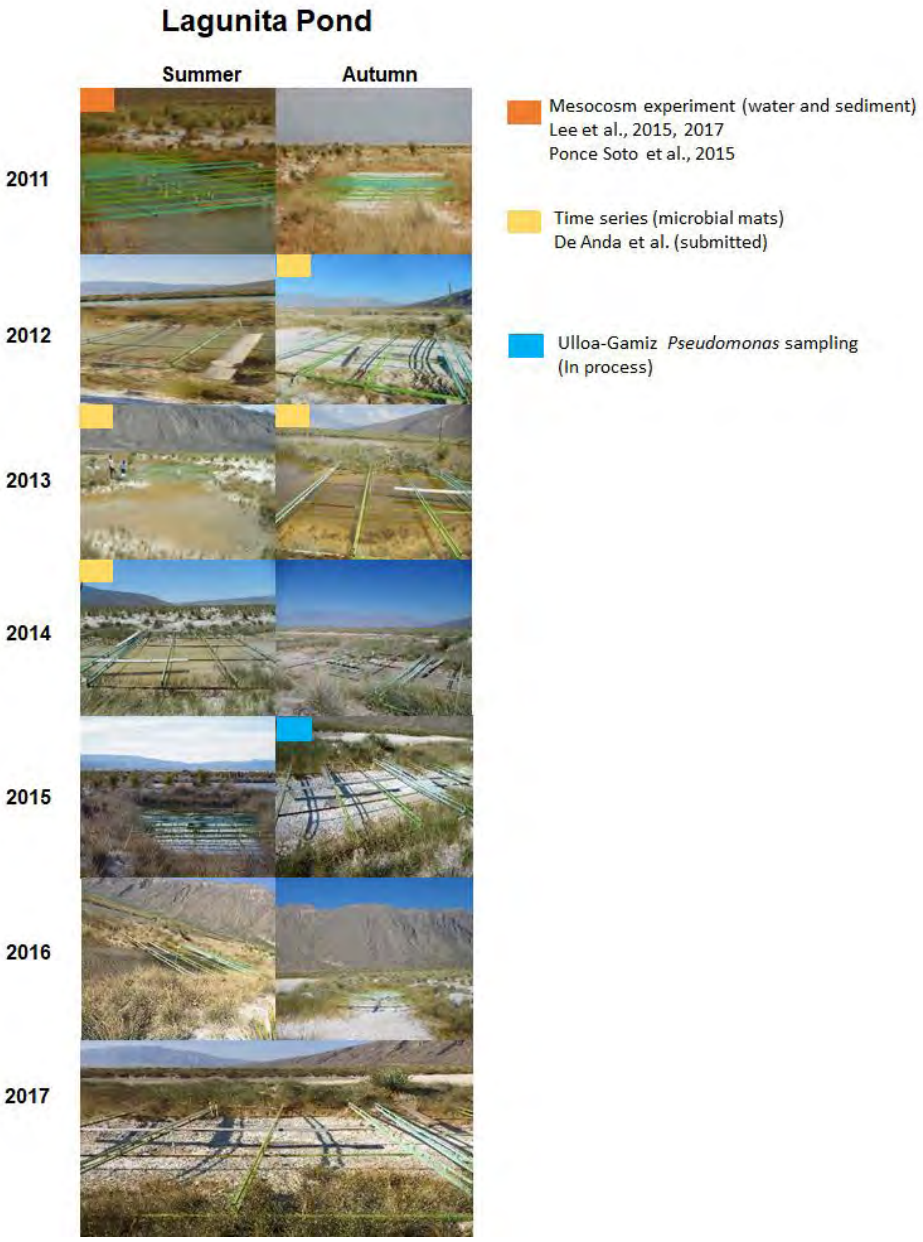



Figure 1. Water changes between autumn and spring from 2011-2017 with-in Lagunita Pond, (26.84810° N, -102.14160° W), lateral to the main Chu-rince system.

In spring 2013, the basin's main canal at La Becerra was closed due to a sustained conservation effort to recover the wetland. This closure made Churince's near-miraculous recovery possible; we record that recovery in this study. However, after summer 2014, that recovery stopped due to the reopening of the La Becerra canal for 2 weeks, changing the

recharge dynamics of the Churince wetland. During 2016-2017, the Churince only had water in the beginning of spring and even that supply is being limited, highlighting the urgency of closing all the canals that export water out of the basin as well as changing agricultural practices within the area (See blue rectangles in Figure 1). This water depletion in the Churince system is "a canary in the mine" that indicates larger threats due to hydrological disruption because this site is few meters higher (745 m asl) than the rest of the basin (average 730 masl) due to the tilt caused by the San Marcos Sierra (Wolaver et al., 2012).

An alternative explanation for the desiccation of the Churince system is extended drought associated with global warming trends. However, we believe that this is unlikely because the rainfall at CCB has actually been, on average, somewhat higher in the last 40 years (Montiel-González et al., 2017). Thus, the desiccation of Churince is most likely associated with continuous overexploitation of the aquifer for intensive agriculture in which spring waters are diverted toward a series of canals that irrigate farms even many kilometers away of the basin. This trend started in the 1960´s with the opening of deep canals in the main springs to export water to irrigated agricultural fields in the region as well as within the valley (Minckley, 1969) in order to water alfalfa for cattle feed. The problem started to become even more pressing during the last 15 years, with expansion of the agricultural frontier in neighboring valleys in 2002 and the opening of many deep wells within and outside the basin (Google Earth visual assessment and CNA (Comisión Nacional del Agua, data). This acceleration of water extraction precipitated the loss of Laguna Grande and a large proportion of the basin's wetlands. This collapse occurs because of subterranean connections between the 3 valleys undergoing agricultural expansion, including Ocampo to the north and el Hundido to the south (Figure 2) (Souza et al., 2006). To our surprise, the evidence gathered by our team prompted two Presidential decrees to declare two vedas, El hundido veda in 2007 and Ocampo-Cuatro Cienegas in 2013 (https://www.gob.mx/inecc/acciones-y-programas/analisis-de-la-variacion-del-nivel-de-los-principales-cuerpos-de-agua-de-cuatrocienegas).

The trend of the agricultural expansion in the southwestern region within CCB (the Hundido) from 1984 to 2016 can be seen in Figure 2.

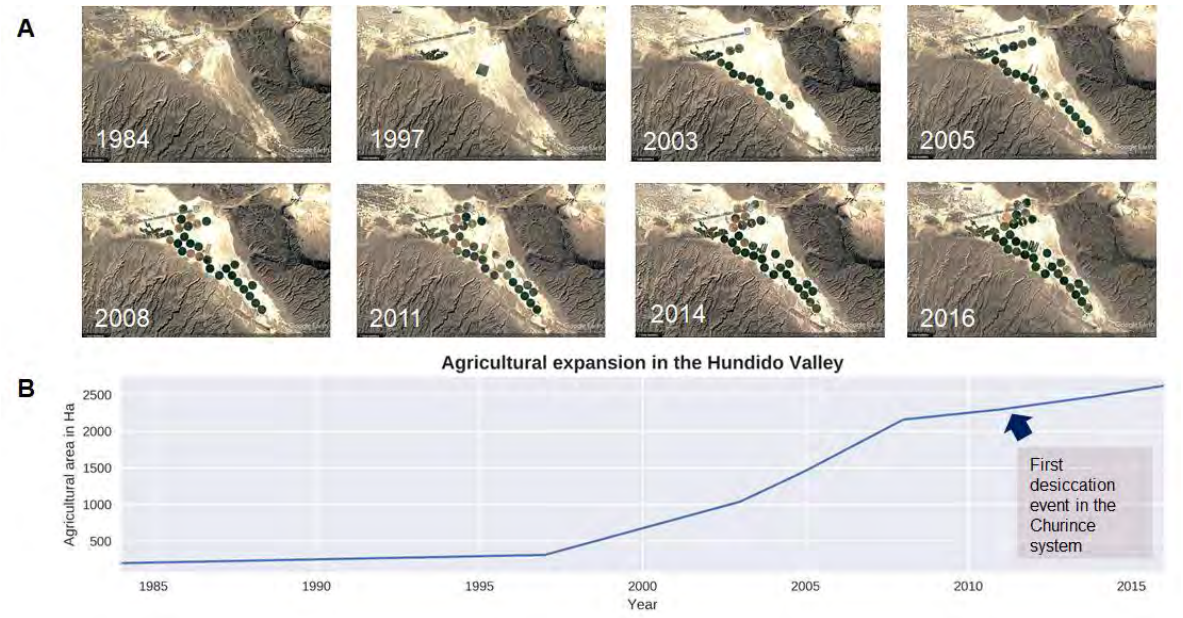



Figure 2. Agricultural expansion in the Hundido Valley 26,594838 °N, -102,214423°W. A) Images obtained with Google Earth Image Landsat-Copernicus from 1984 to 2016 B) Agricultural area in hectares was comput-ed by calculating the surface of crop land using Google Earth

During our first sampling (See Winter 2012 in Figure 1), we observed strong depletion of the aquifer, followed, as mentioned above, by an apparent recovery of the ecosystem in summer 2013 due to the closing of the "La Becerra" canal and the rebirth of the "el Garabatal" wetland that had been dry since the 1960´s (Minckley, 1969). The restoration of the water cycle in the wetland allowed Churince to recover its water flow for a small period that we were able to document (see yellow rectangles in Figure 1).

## 3.3 Microbial mats a bio-marker to understand environmental perturbation.

Because microbial communities are the heart of every ecosystem in our planet and given the intense overexploitation of water resources within CCB, one of our main aims has been to understand how microbial systems respond to ongoing environmental disturbance so that we can establish movitation for conservation and protection of the aquifer. Considering that

natural microbial communities are in constant flux even in the absence of environmental perturbations, we hypothesized that it is easier to identify the underlying mechanisms that arise from an environmental disturbance in very stable and resilient natural communities (for further information see Konopka et.al 2015)

Microbial mats are one of the oldest microbial ecosystems, having been established in the fossil record since the Archean. Thus, they are well known to be successful ecological communities that survived millions of years of environmental disruption due to their internal capacity to carry out all fundamental biogeochemical and energy transformations (van Gemerden 1993; Bolhuis et al. 2014; Prieto-Barajas et al. 2017). Therefore, they are excellent models for understanding the role of environmental disturbances on microbial communities. Our recent study (De Anda, under review) we evaluated microbial mats from Churince Lagoon in CCB over the course of two years during and after a desiccation event (see Figure 1 yellow rectangles). We hypothesized that if microbial mats are indeed stable over time (regardless of perturbations by water overexploitation), we should expect that samples taken at fixed time points in space would pre-sent similar community patterns. In contrast, if there is indeed a perturbation to the community, we expected systematic differences that would then reflect the degree of resistance, resilience, or functional redundancy of the microbial mats. To assess these potential changes, we used a network-based analytical framework. To our knowledge, this is the first network-based study that involves a time series approach in microbial mats.

In our study, we adopted the well-established macro-ecological concept that ecosystems with a greater number of species are more pro-ductive, more resilient to invasions of foreign species, and more stable in the face of perturbations such as drought. Our analysis is based on the pio-neering work of David Tilman, whose work has shown that, the greater number of species (especially plants), the larger number of limiting resource interactions are needed to fulfill different needs (Chapin et al. 1998; Tilman 1999, 2004; Tilman et al. 2006). In this macroecological context, almost ten years ago Bascompte and Stouffer (2009) suggested that, in face of the ongoing global change crisis, understanding the structure of species rela-

tionships, and how that structure relates to network disassembly, should be a high priority for systems-level conservation biology (Bascompte et al. 2003; Bascompte 2010; Guimarães et al. 2006, 2017; Rohr et al. 2014). However, in microbial ecology, little is known about the community-wide implications of human-induced perturbations our about the changes that such drivers induce in network interactions among microbial species.

## 3.4 Understanding microbial relationships by network inference.

Ecological interactions have been acknowledged to play a key role in shap-ing biodiversity (Guimarães et al. 2017). In contrast to macro-ecologists who have a long tradition of studying how species interact with each other (Bascompte 2010), microbial ecologists has largely focused on evaluating the relative abundances of community members (taxa and genes),and us-ing various ecological indexes (i.e. Shannon) to measure the degree of re-sistance, resilience, or functional redundancy following an environmental disturbance (Allison and Martiny 2008; Konopka 2009; Shade et al. 2012).

This disparity is likely due to the difficulties in inferring ecological interac-tions in the microbial world, where, in contrast to macro-ecological studies of plant or insect communities, ,it is very difficult to recognize interactions, such as who-eats-who (antagonism) or mutually beneficial relationships (i.e. a bee pollinating a type of flower) (Bascompte 2010).For microbial communities is close to impossible to determine every type of interaction just by observation, as microbial communities are composed of hundreds or thousands of species and exhibit much larger and more rapid changes in bi-omass, composition, and activity than do plant and animal population (Song et al. 2014). Therefore, it is necessary to use indirect, inferential methods that allow us to predict and detect different types of interactions within complex microbial communities.

Currently, advances in metagenomics and bioinformatics ap-proaches allow the identification and estimation of the relative abundances and functional capabilities of microbes in complex environmental samples. Prediction of the ecological relationships between microorganisms is achieved thorough network inference, using presence/absence

data or abundances profiles generated by several bioinformatics methods. In gen-eral, there are two main groups of network inference methods: those that can predict relationships between two species (namely pairwise relation-ships) and those that can predict more complex ones (for a review see Faust and Raes 2012; Song et al. 2014).

The predicted relationships are defined as competitive (-/-) (in which two species consume shared resources) if their abundances across samples are anti-correlated despite sharing environmental tolerance nich-es. In contrast, cooperative relationships (where the metabolites produced by one species are consumed by another and, potentially, vice-versa) dis-play similar abundance patterns. Furthermore, the relationships between species can be bidirectional, if the impact is mutually positive (mutualism if obligatory, or synergism if nonobligatory), mutually negative (competition), or positive on one side but negative on the other (antagonism); unidirec-tional, if the impact on one of the two is neutral, regardless of whether the impact on the other is positive (commensalism) or negative (amensalism); non-directional, if the impact on each other is negligible or insignificant (neutralism) (Freilich et al. 2011; Song et al. 2014). (For schematic represen-tation of the ecological interaction see Figure 1 in Faust and Raes 2012, or a more complete description in Table 1 in Song et al. 2014).

The network inference approach also offers potential insight into the stability of microbial communities by using temporal dynamic variation. As reviewed by Faust et al. (2015), several methods have been proposed to infer a single network of interaction from the entire time series. Existing methods require assuming a particular population dynamics model, which is not known a priori (Xiao et al. 2017). One of these methods exploits a direct comprehensive interacting network for microorganisms using Lotka-Volterra equations (Shaw et al. 2016), an approach that has been success-fully employed by macroecologists to describe a dynamic trophic web of more than two populations of macro-organisms. In microbial ecology stud-ies it has been observed that Lotka-Volterra equations can quantify microbial interactions and successfully predict microbiome temporal dynamics by modeling changes in abundance

as a function of taxon-specific growth rates and pair-wise interaction strengths (Shaw et al. 2016; Stein et al. 2013; Weng et al. 2017).

## 3.5 How the loss of water is affecting the microbial mats within Churince Lagoon?

We used a Lotka-Volterra-based network inference approach to understand how environmental perturbation affects the stability and dy-namics of microbial mats in the Churince system. To distinguish between in-trinsic shifts generated by the community per se and those derived from outside forces, we studied microbial mats from three sites within a small pond of the Churince system over the course of two years (Figure 3).

Water depletion altered the appearance of the microbial mats of the three sampling sites from two contrasting environments (dry and wet). We analyzed the shifts in community composition, structure, and function of these communities but also the extent to which the specific type of in-teraction and overall network structure was altered. Our results indicate that, even if the physicochemical environment was constant over time, the composition of microbial communities would be continually changing due to the intrinsic dynamics driven by ecological interactions. When environ-mental influences are not constant (as in the case of water depletion), those fluctuations further affect the community dynamics (e.g. community structure and composition). Consistent with the altered appearance of wet purple microbial mats, genera that were enriched under dry conditions were taxa implicated in the oxidation of inorganic reduced sulfur com-pounds. These were the purple sulfur bacterium -Ectothiorodospiraceae and the colorless sulfur bacterium Thioalkalivibrio, which is physiologically and metabolically adapted to hyper-saline (up to saturation) and alkaline (pH up to 10.5) conditions (Foti et al. 2006; Sorokin et al. 2011). These fluc-tuations in hydrology in the Lagunita pond would be expected to have a profound effect on redox potential within microbial mats, due to changes in nutrient and oxygen availability (Peralta et al. 2014). When moist conditions were reestablished by the restoration of the Garabatal wetland (also part of the CCB), the redox gradient was reinstated, allowing the growth of

micro-aerophilic and anaerobic taxa (i.e., increases of sulfate reducing bacteria) or Bacteroidetes. These dynamics indicate that, while few taxa are shared, many unique taxa exist among geographically separated mats in moist con-ditions. Our results confirm previous results showing that sediments of hyper-saline lakes and lagoons may support a rich community of anaerobic halophilic bacteria, given that the solubility of oxygen in hyper-saline brines is low and supplies of organic matter available are often high (Oren 1988, 2008). Our findings show that these two microbial groups could be used as indicators of water reestablishment due their continuous presence during moist conditions across broad spatial-temporal scales. The first taxon, Ornatilinea (Chloroflexi), is related to sequenced microbial mats from hot wa-ter emerging from a 2775-m deep well (Podosokorskaya et al. 2013), con-firming the deep-water origin hypothesis presented by Wolaver (2012) to explain the uniqueness of the CCB springs. The second taxon belongs to an unclassified genus from the Rikenellaceae family (Bacteroidetes), whose members are recognized as saccharolytic and able to ferment glucose to acid by-products by anaerobic metabolism (Graf 2014).

The similarities among the wet patches during dry conditions (de Anda et al., under revision) suggest a possible role of community members that are able to grow in acidic environments, gaining energy from the oxidation and inorganic sulfur compound and possible ferrous iron to obtain organic carbon from carbon dioxide. We suggest, that during water depletion events, small infiltrations of deep water rich in sulfur are patchy, resulting in some wet patches and another dry site. Wet patches possibly contain significant concentrations of sulfide derived from groundwater.

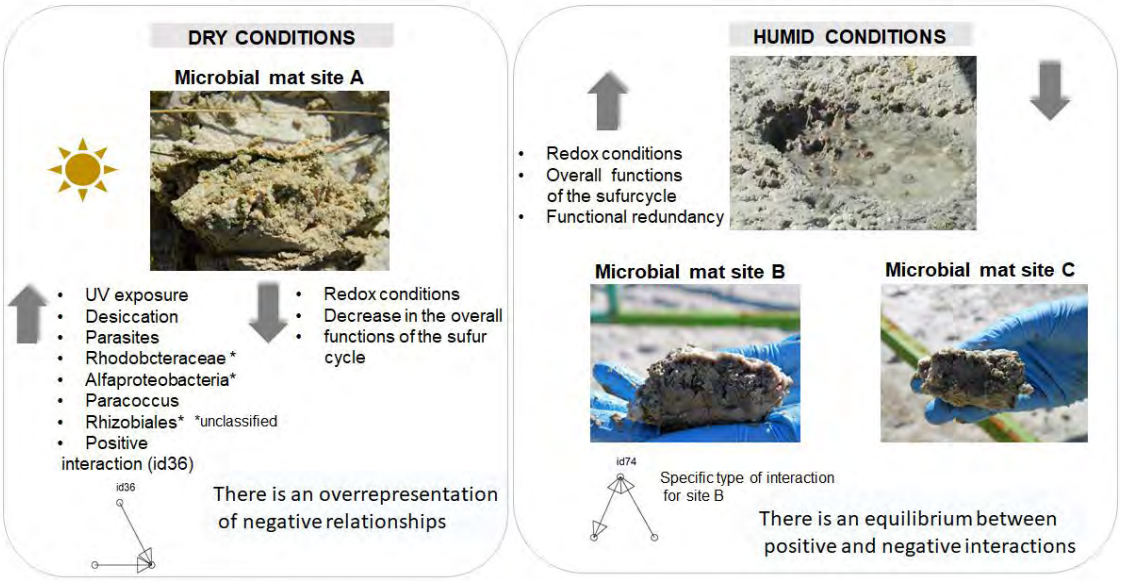



Figure 3. Main results obtained from our time series, Lotka-Volterra-based network inference approach for microbial mats collected during and after overexploitation of water resources within Churince system. Only the microbial mats obtained for our first sampling point (see figure 1) are shown, a period when La-gunita Pond was dry and there were only two remaining patches (sites B and C).

## 3.6 Network community structure indicates a completely overlapping niche within microbial mats.

Consistent with metabolic coupling within the microbial mats, we found that the overall community structure is densely inter-connected, pattern that is interpreted as indicating that the species in the community have overlapping niches (Faust and Raes 2012). This was assessed using a metric known as modularity (Q), which quantifies patterns of connectedness within and across groups (Newman 2006). Positive modularity values indicate that interactions occur predominantly within-groups while negative values show that interactions are more frequent between groups than within them.

Several studies suggest that modularity values for of biological networks fall in the range of 0.3 to 0.7, with higher values being rare (Newman and Girvan 2004). However, in our study

we obtained modularity values close to zero, indicating that, within the microbial mats of Churince, there is a notable lack of modules. Instead, the network of interactions represents "a higher-order module", in which all its members are related with each other (Grilli et al. 2016). This result suggests that the microbial mats of Churince have a unique ecological structure.

Microbial mats are constructed by a complex matrix of trophic and metabolic interactions that is constantly transforming their environment (Kerr et al. 2002). These successful ecological communities have survived millions of years of environmental disturbances due to their ability to con-struct their own "niche space". They do so because of metabolic complementation reactions with overlapping ecological functions at a millimeter scale in the absence of algae (that grow too fast and would outcompete other microbes for light) and particular conditions: water, a redox potential given by a sulfur source, and sunlight ( Bolhuis et al. 2014; Des Marais 2003; Pajares et al. 2012; Prieto-Barajas et al. 2017; van Gemerden 1993;). However, microbial mats grow slowly due to their cohesive development in the Precambrian, perhaps reflecting their evolutionary origins in a world where phosphorus (P) was very limited (Elser et al. 2006) and there was no opportunity to rapidly proliferate. Nowadays, fast-growing communities in a P-rich world have displaced microbial mats. Hence, microbialites like those in CCB constitute "time machines" that exist only in environments in which biogeochemical conditions sustain "multi-dimensional bubbles" that mimic past conditions, preventing competition by algal growth.

## 3.7 Conclusions

To date, understanding the complex links between environmental disturbance and microbial diversity that affect ecosystem stability has been based on the compositional and functional dynamics of taxa following an environmental disturbance. Understanding microbial interactions is essen-tial for revealing community assembly rules that determine how habitat traits affect the assembly of microbial communities. In this chapter, we highlight that the use of network inference approaches was crucial in understanding how anthropogenic

disturbance affects one of the most resistant microbial communities known, microbial mats. These analyses provided some of the first mechanistic insights into the response of microbial mats to environmental perturbation in nearly closed system. Our work indicates that improving the prediction and management of environmental ecosystems in face of anthropogenic perturbations requires examination of the interaction networks as a whole.

## 3.8 References


Allison SD, Martiny JBH (2008) Resistance, resilience, and redundancy in mi-crobial communities. PNAS 105.Suppl 1 (2008): 11512–11519 Bas-compte J (2010) Structure and dynamics of ecological networks. Sci-ence 329:765–766

Bascompte J, Jordano P, Melián CJ, et al (2003) The nested assembly of plant-animal mutualistic networks. PNAS 100:9383–7. doi: 10.1073/pnas.1633576100

Bascompte J, Stouffer DB (2009) The assembly and disassembly of ecologi-cal networks. Philos Trans R Soc B Biol Sci 364:1781–1787. doi: 10.1098/rstb.2008.0226

Bolhuis H, Cretoiu MS, Stal LJ (2014) Molecular Ecology of Microbial Mats. FEMS Microbiol Ecol 90:335–50. doi: 10.1111/1574-6941.12408

Cadotte MW, Arnillas CA, Livingstone SW, et al (2015) Predicting communi-ties from functional traits. Trends Ecol Evol 30:510–511. doi: 10.1016/j.tree.2015.07.001

Carson EW, Souza V, Espinosa-Pérez H, et al (2015) Mitochondrial DNA Di-versity and Phylogeography of Lucania interioris Inform Biodiversity Conservation in the Cuatro Ciénegas Basin, México. West North Am Nat 75:200–208. doi: 10.3398/064.075.0208

Cerritos R, Eguiarte LE, Avitia M, et al (2011) Diversity of culturable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, México. Antonie Van Leeuwenhoek 99:303–18. doi: 10.1007/s10482-010-9490-9

Chapin, FS, Sala OE, Burke IC, et al (1998) Ecosystem Consequences of Changing Biodiversity. Bioscience 48:45–52. doi: 10.2307/1313227

Des Marais DJ (2003) Biogeochemistry of hypersaline microbial mats illus-trates the dynamics of modern microbial ecosystems and the early evolution of the biosphere. Biol Bull 204:160–167

Elser JJ, Watts J, Schampel JH, et al (2006) Early Cambrian food webs on a trophic knife-edge? A hypothesis and preliminary data from a modern stromatolite-based ecosystem. Ecol Lett 9:292–300. doi: 10.1111/j.1461-0248.2005.00873.x

Faust K, Lahti L, Gonze D, et al (2015) Metagenomics meets time series analysis: unraveling microbial community dynamics. Curr Opin Micro-biol 25:56–66. doi: 10.1016/j.mib.2015.04.004

Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550. doi: 10.1038/nrmicro2832

Foti M, Ma S, Sorokin DY, et al (2006) Genetic diversity and biogeography of haloalkaliphilic sulphur-oxidizing bacteria belonging to the genus Thi-oalkalivibrio. FEMS Microbiol Ecol 56:95–101. doi: 10.1111/j.1574-6941.2006.00068.x

Freilich S, Zarecki R, Eilam O, et al (2011) Competitive and cooperative met-abolic interactions in bacterial communities. Nat Commun 2:587–589. doi: 10.1038/ncomms1597

Graf J (2014) The Family Rikenellaceae. In: Rosenberg E, DeLong EF, Lory S, et al. (eds) The Prokaryotes: Other Major Lineages of Bacteria and The Archaea. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 857–859

Grilli J, Rogers T, A|llesina S (2016) Modularity and stability in ecological communities. Nat Commun 7:12031. doi: 10.1038/ncomms12031

Guimarães PR, Pires MM, Jordano P, et al (2017) Indirect effects drive co-evolution in mutualistic networks. Nature 550:511–514. doi: 10.1038/nature24273

Guimarães PR, Rico-Gray V, dos Reis SF, Thompson JN (2006) Asymmetries in specialization in ant-plant mutualistic networks. Proc Biol Sci 273:2041–7. doi: 10.1098/rspb.2006.3548

Hernández A, Espinosa-Pérez HS, Souza V (2017) Trophic analysis of the fish community in the Ciénega Churince, Cuatro Ciénegas, Coahuila. PeerJ 5:e3637. doi: 10.7717/peerj.3637

Kerr B, Riley M a, Feldman MW, Bohannan BJM (2002) Local dispersal pro-motes biodiversity in a real-life game of rock-paper-scissors. Nature 418:171–4. doi: 10.1038/nature00823

Konopka A (2009) What is microbial community ecology? ISME J 3:1223–1230. doi: 10.1038/ismej.2009.88

Konopka, A. E., Lindemann, S., and Fredrickson, J. K. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. ISME J. 9, 1488–1495. doi:10.1038/ismej.2014.251.

Minckley TA, Jackson ST (2007) Ecological stability in a changing world? Re-assessment of the palaeoenvironmental history of Cuatrociénegas, Mexico. J Biogeogr 35:188–190. doi: 10.1111/j.1365-2699.2007.01829.x

Minckley W (1969) Environments of the Bolson of Cuatro Cienegas, Coahui-la, Mexico. Sci Ser 2:1–65

Montiel-González C, Tapia-Torres Y, Souza V, et al (2017) The response of soil microbial communities to variation in annual precipitation de-pends on soil nutritional status in an oligotrophic desert. PeerJ 5:e4007. doi: 10.7717/peerj.4007

Newman MEJ (2006) Modularity and community structure in networks. PNAS 103:8577–82. doi: 10.1073/pnas.0601602103

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E. American Physical Society; 2004;69: 26113. pmid:14995526 Oren A (1988) Anaerobic degradation of organic compounds at high salt concentrations. Antonie Van Leeuwenhoek 54:267–277. doi: 10.1007/BF00443585

Oren A (2008) Microbial life at high salt concentrations: Phylogenetic and metabolic diversity. Saline Systems 4:1–13. doi: 10.1186/1746-1448-4-2

Pajares S, Bonilla-Rosso G, Travisano M, et al (2012) Mesocosms of aquatic bacterial communities from the Cuatro Cienegas Basin (Mexico): a tool to test bacterial community response to environmental stress. Microb Ecol 64:346–58. doi: 10.1007/s00248-012-0045-7

Peralta AL, Ludmer S, Matthews JW, Kent AD (2014) Bacterial community response to changes in soil redox potential along a moisture gradient in restored wetlands. Ecol Eng 73:246–253. doi: 10.1016/j.ecoleng.2014.09.047

Podosokorskaya OA, Bonch-Osmolovskaya EA, Novikov AA, et al (2013) Or-natilinea apprima gen. nov., sp. nov., a cellulolytic representative of the class Anaerolineae. Int J Syst Evol Microbiol 63:86–92. doi: 10.1099/ijs.0.041012-0

Prieto-Barajas CM, Valencia-Cantero E, Santoyo G (2017) Microbial mat ecosystems: Structure types, functional diversity, and biotechnologi-cal application. Electron J Biotechnol 31:48–56. doi: 10.1016/j.ejbt.2017.11.001

Rohr RP, Saavedra S, Bascompte J (2014) On the structural stability of mu-tualistic systems. Science 345:1253497. doi: 10.1126/science.1253497

Shade A, Peter H, Allison SD, et al (2012) Fundamentals of microbial com-munity resistance and resilience. Front Microbiol 3:1–19. doi: 10.3389/fmicb.2012.00417

Shaw GTW, Pao YY, Wang D (2016) MetaMIS: a metagenomic microbial in-teraction simulator based on microbial community profiles. BMC Bio-informatics 17:488. doi: 10.1186/s12859-016-1359-0

Song HS, Cannon W, Beliaev A, Konopka A (2014) Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. Pro-cesses 2(4):711–752. doi: 10.3390/pr2040711

Sorokin DY, Kuenen JG, Muyzer G (2011) The Microbial Sulfur Cycle at Ex-tremely Haloalkaline Conditions of Soda Lakes. Front Microbiol 2:1-16 doi: 10.3389/fmicb.2011.00044

Souza V, Espinosa-Asuar L, Escalante AE, et al (2006) An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. PNAS 103:6565–6570. doi: 10.1073/pnas.0601434103

Souza V, Falcón LI, Elser JJ, et al (2007) Protecting a window into the an-cient Earth: Towards a Precambrian Park at Cuatro Cienegas, Mexi-co. The Citizen's Page, Evolutionary Ecology Research. Available online at http:// www.evolutionary-ecology.com/citizen/citizen.html Souza V, Siefert JL, Escalante AE, et al (2012) The Cuatro Ciénegas Basin in Coahuila, Mexico: an astrobiological Precambrian Park. As-trobiology 12:641–7. doi: 10.1089/ast.2011.0675

Stein RR, Bucci V, Toussaint NC, et al (2013) Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intesti-nal Microbiota. PLoS Comput Biol 9:31–36. doi: 10.1371/journal.pcbi.1003388

Tilman D (2004) Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. Proc Natl Acad Sci 101:10854–10861. doi: 10.1073/pnas.0403458101

Tilman D (1999) The ecological consequences of changes in biodiversity: A search for general principles. Ecol 80:1455–1474. doi: 10.2307/176540

Tilman D, Reich PB, Knops JMH (2006) Biodiversity and ecosystem stability in a decade-long grassland experiment. Nature 441:629–632. doi: 10.1038/nature04742

van Gemerden H (1993) Microbial mats: A joint venture. Mar Geol 113:3–25. doi: 10.1016/0025-3227(93)90146-M

Weng FCH, Shaw GTW, Weng CY, et al (2017) Inferring microbial interac-tions in the gut of the Hong Kong whipping frog (Polypedates mega-cephalus) and a validation using probiotics. Front Microbiol 8:1–11. doi: 10.3389/fmicb.2017.00525

Wolaver BD, Crossey LJ, Karlstrom KE, et al (2012) Identifying origins of and pathways for spring waters in a semiarid basin using He, Sr, and C iso-topes: Cuatrocienegas Basin, Mexico. Geosphere 9:113–125. doi: 10.1130/GES00849.1

Wolaver BD, Diehl TM (2010) Control of regional structural styles and fault-ing on Northeast Mexico spring distribution. Environ Earth Sci 62:1535–1549. doi: 10.1007/s12665-010-0639-7

Wong HL, Smith DL, Visscher PT, et al (2015) Niche differentiation of bacte-rial communities at a millimeter scale in Shark Bay microbial mats. Sci Rep 26;5:15607. doi: 10.1038/srep15607

Xiao Y, Angulo MT, Friedman J, et al (2017) Mapping the ecological net-works of microbial communities. Nat Commun 8:2042. doi: 10.1038/s41467-017-02090-2

Capítulo 4

# Highlighting the importance of negative interactions and rare biosphere within microbial mats under perturbation: a time-series metagenomic study.

Valerie De Anda, Icoquih Zapata-Peñasco, Jazmín Blaz, Augusto Cesar Poot-Hernández, Bruno Contreras Moreira, Marcos González Laffitte, Niza Gámez Tamariz, Maribel Hernández Rosales, Luis E. Eguiarte and Valeria Souza

## Algoritmo

El algoritmo que se desarrollo e implemento en este articulo se encuentra públicamente disponible en el siguiente repositorio de GitHub https://valdeanda.github.io/Networks/

## Información suplementaria

Los datos y scripts necesarios para recrear las figuras del artículo se encuentran publicamente disponibles en el siguiente repositorio de GitHub https://valdeanda.github.io/Time_series_mats/

# Highlighting the importance of negative interactions and rare biosphere within microbial mats under perturbation: a time-series metagenomic study.

Valerie De Anda[1], Icoquih Zapata-Peñasco[2], Jazmin Blaz[3], Augusto Cesar Poot-Hernandez[4], Bruno Contreras-Moreira[5], Marcos Gonzales Iaffitte[6], Maribel Hernandez Rosales[6], Luis E. Eguiarte[1], Valeria Souza[1*]

[1]Institute of Ecology, National Autonomous University of Mexico, Mexico, [2]Dirección de Investigación en Transformación de Hidrocarburos. Instituto Mexicano del Petróleo,, Mexican Institute of Petroleum, Mexico, [3]Laboratorio Nacional de Ciencias de la Sostenibilidad (LANCIS), Mexico, [4]Departamento de Ingeniería de Sistemas Computacionales y Automatización, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, Mexico, [5]Estación Experimental de Aula Dei (EEAD), Spain, [6]Instituto de Matemáticas, Universidad Nacional Autónoma de México, Mexico



*Submitted to Journal:*
Frontiers in Microbiology

*Specialty Section:*
Aquatic Microbiology

*Article type:*
Original Research Article

*Manuscript ID:*
380137

*Received on:*
29 Mar 2018

*Revised on:*
01 Jun 2018

*Frontiers website link:*
www.frontiersin.org

## Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

## Author contribution statement

VDA, IZP, LEE, VS: conceived the project and worked on the manuscript. VDA, JB: sampling, field work, processing, DNA extraction and ecological index analysis. VDA, CPH, BMC: conducted bioinformatics and data analysis (trimming, filtering, taxonomic and metabolic profile). BCM, MHR: provided computing resources and analysis interpretation. MGL, MHR: performed and coded the network analyses. MGL, MHR, NGT: provided ecological and mathematical interpretation of the networks. VDA, IZP: conducted the microbiological, biogeochemical, and ecological interpretation of the analysis. LEE, VS: sampling and fieldwork, provided expertise, logistics and resources to develop and supervised the project, as well as intellectual contributions to the work.

## Keywords

microbial mats, network motifs, time-series networks, rare biosphere, Interactions between organisms, MEBS, Environmental perturbation



## Abstract

Word count: 327

Understanding the responses of microbial communities in the face of environmental perturbations and the mechanisms, which constrain or promote microbial adaptation to ongoing climate change, is crucial to evaluate the ecosystem impact at global scales. Microbial mats are ideal as ecological models due to their persistence over geological time scales. We monitored with spatial and temporal resolution the response of water depletion in microbial mats in the endangered oasis of Cuatro Cienegas Basin (CCB), Coahuila Mexico. By using standard metagenomics in a time-series of 2 years in 3 close sampling sites, we observed that mats within CCB, are extremely diverse, each site has its own dynamic despite its proximity and each one of them had a large proportion of rare taxa. The comparison across sites suggests a clear involvement of anaerobic taxa in mats from wet zones, in contrast within dry conditions there is an increase in fungi, photosynthetic and halotolerant taxa. Our results indicate that microbial mats tend to evolve towards stable conditions by the reestablishment of redox conditions as the aquifer recovery momentarily, we observed this by observing that methane and sulfur cycles were over-represented at the end of the study. In order to understand the role of interactions in the recovery, we implement robust multi-layered time-series ecological networks for each sampling site. We observed that keystone taxa are members of the rare biosphere and of the “microbial mat core”, suggesting their crucial role as a buffer against environmental disturbance, by maintaining redundant separate modules and probably delivering public goods essential for the metabolic coupling within microbial mats. In humid conditions, there was an equilibrium between negative-positive interactions, while under desiccation there is an overrepresentation of antagonisms relationships in response to stress. Further studies are required to address whether the overrepresentation of certain motifs was environment-specific. Altogether, we provide a baseline data to develop a conceptual model for the complex microbial community dynamics under environmental disturbances, which give rise to patterns, forces, functions and interactions

## Funding statement

This work constitutes a partial fulfillment requirement for the Ph.D. degree of Valerie De Anda at the graduate program Doctorado en Ciencias Biomédicas of the Universidad Nacional Autónoma de México who received fellowship 356 832 of Consejo Nacional de Ciencia y Tecnología (CONACYT). The authors acknowledge the founding of WWF-Alianza Carlos Slim, as well as support by a Sep Conacyt Project to VS and LEE 1101OL34. The paper was written during VDA research stay in BCM laboratory with the support of Beca Mixta Conacyt. The manuscript was written during a sabbatical leave of LEE and VSS in the University of Minnesota in Peter Tiffin and Michael Travisano laboratories, with support of the program PASPA- DGAPA, UNAM.

## Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

*Does the study presented in the manuscript involve human or animal subjects:* No

# Highlighting the importance of negative interactions and rare biosphere within microbial mats under perturbation: a time-series metagenomic study.

Valerie De Anda[1], Icoquih Zapata-Peñasco[2], Jazmín Blaz[3], Augusto Cesar Poot-Hernández[4], Bruno Contreras Moreira[5,6], Marcos González Laffitte[7], Niza Gámez Tamariz[1], Maribel Hernández Rosales[7], Luis E. Eguiarte[1] and Valeria Souza[*1]

[1]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, 70-275, Ciudad de México 04510, México.

[2]Dirección de Investigación en Transformación de Hidrocarburos. Instituto Mexicano del Petróleo, Eje Central Lázaro Cárdenas, Norte 152, Col. San Bartolo Atepehuacan, Ciudad de México, 07730, México

[3]Laboratorio Nacional de Ciencias de la Sostenibilidad, Instituto de Ecología, Universidad Nacional Autónoma de México, 70-275, Ciudad de México 04510, México.

[4]Departamento de Ingeniería de Sistemas Computacionales y Automatización, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, UNAM, Ciudad Universitaria CP 04510 Ciudad de México, México.

[5]Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Avda. Montañana, 1005, Zaragoza 50059, Spain

[6]Fundación ARAID, Zaragoza, Spain.

[7]Instituto de Matemáticas, UNAM Juriquilla, Blvd. Juriquilla 3001, 76230 Juriquilla, Qro., México

*Correspondence:
souza@unam.mx

Keywords: time-series ecological networks, network motifs, microbial mats, MEBS, environmental perturbation, rare biosphere, interactions.

## Abstract

Understanding the responses of microbial communities in the face of environmental perturbations and the mechanisms, which constrain or promote microbial adaptation to ongoing climate change, is crucial to evaluate the ecosystem impact at global scales. Microbial mats are ideal as ecological models due to their persistence over geological time scales. We monitored with spatial and temporal resolution the response of water depletion in microbial mats in the endangered oasis of Cuatro Cienegas Basin (CCB), Coahuila Mexico. By using standard metagenomics in a time-series of 2 years in 3 close sampling sites, we observed that mats within CCB, are extremely diverse, each site has its own dynamic despite its proximity and each one of them had a large proportion of rare taxa. The comparison across sites suggests a clear involvement of anaerobic taxa in mats from wet zones, in contrast within dry conditions there is an increase in fungi, photosynthetic and halotolerant taxa. Our results indicate that microbial mats tend to evolve towards stable conditions by the reestablishment of redox conditions as the aquifer recovery momentarily, we observed this by observing that methane and sulfur cycles were over-represented at the end of the study. In order to understand the role of interactions in the recovery, we implement robust multi-layered time-series ecological networks for each sampling site. We observed that keystone taxa are members of the rare biosphere and of the "microbial mat core", suggesting their crucial role as a buffer against environmental disturbance, by maintaining redundant separate modules and probably delivering public goods essential for the metabolic coupling within microbial mats. In humid conditions, there was an equilibrium between negative-positive interactions, while under desiccation there is an overrepresentation of antagonisms relationships in response to stress. Further studies are required to address whether the overrepresentation of certain motifs was environment-specific. Altogether, we provide a baseline data to develop a conceptual model for the complex microbial community dynamics under environmental disturbances, which give rise to patterns, forces, functions and interactions.




## Introduction

Microbial mats and stromatolites are one of the oldest evidences of life on Earth, registered in the fossil record since the Archean (Nutman et al., 2016). These organo-sedimentary structures are well known to be successful ecological communities adapting to continuous environmental changes due to their capacity to perform most of the biogeochemical cycles in a nearly close way (Bolhuis et al., 2014; Guerrero et al., 2002; van Gemerden, 1993). However, that resilience seems to be faltering, their

survival is endangered by human perturbation (Souza et al., 2006). Understanding the responses of these complex microbial communities in the face of environmental perturbations can give us clues about their persistence in geological times. In this sense, networks inference has been shown to reveal patterns that both illuminating the ecological processes underlying microbial dynamics and promising a better understanding of the relationship between complexity and ecological stability (Montoya et al., 2006). Recently, microbial interactions and rare microbial taxa have been proposed as a buffer against environmental disturbances influencing both community assembly and stability (Jousset et al., 2017; Konopka et al., 2015), yet it is unclear how the microbial community dynamics (i.e. composition and membership relationships) are influenced by environmental constrains, which largely will shape the degree of resistance, resilience, functional redundancy of a microbial community (Bissett et al., 2013; Fuhrman, 2009; Konopka et al., 2015; Newton et al., 2012).

Due to their complexity, biological networks are commonly reduced into meaningful subnetworks to better characterize their structure, function and behavior (Alon, 2007; Prill et al., 2005). A type of subnetworks are the network motifs, which consist in the complete set of unique connected subgraphs containing *n* nodes (represented by OTUS, species, genes or proteins) (Alon, 2007; Baiser et al., 2016; Milo, 2002). These relationships are so frequent that they are unlikely to occur just by chance., and have been applied to the study of development, regulatory and neuronal networks (Prill et al., 2005; Shen-Orr et al., 2002; Tran et al., 2013). Besides, the study of network motifs has been explored theoretically and empirically in macro-ecological food webs since the 90's (Baiser et al., 2016; Borrelli et al., 2015; Stouffer et al., 2007). In community analysis, it has been shown that network motifs offer the opportunity to bridge the gap between the study of the dynamics of simple modules and the analysis of topological metrics describing the community as a whole (Delmas et al., 2017). Despite its success as an ecological tool, the estimation of network motifs in microbial ecology studies has been largely unexplored.

Additionally, the evaluation of microbial community networks, derived from OMIC studies has been largely focused on validating the interaction of a small group of microbes, or evaluating a small percentage of the strongest interaction pairs (i.e Tzun-Wen Shaw et al., 2017). Nevertheless, microbes, which are not very abundant and belong to rare biosphere, have been removed routinely from microbial ecological studies (Jousset et al., 2017) , where systematic approaches for evaluating the whole repertory of interactions in a given microbial community under environmental perturbation have also been absent. Considering that, natural microbial communities are highly complex systems that are in constant flux through spatial and temporal scales where even minor perturbation can significantly reorder the functions of each member of a community and the network of interactions between them

(Konopka et al., 2015). In this study, we focused on the hypothesis that by assessing general patterns of community relationships in stable microbial communities such as microbial mats during and after stressed conditions, we can determine the direct role of local interactions in the capability of resilience of the community. We believe that this type of analysis could be useful for accessing ecosystem stability and understanding the complex links between environmental change and microbial diversity. Modern microbial mats and stromatolites are rare, but occur in diverse aquatic habitats where extreme conditions limit the growth of algae that are strong competitors against the slow growing photosynthetic community within the microbial mats and stromatolites (Prieto-Barajas et al., 2017). Here we investigated microbial mats from the protected natural area of Cuatro Cienegas Basin (CCB), a naturally isolated valley in the Chihuahuan Desert, North of Mexico. CCB is a World Wild Fund (WWF) hotspot for biodiversity and a wetland of international importance under the RAMSAR convention. However, despite the great importance of this unique site, the increasing demand on water for agricultural development (mainly for forage and feed for livestock) has generated important conservation issues related to the drying of different aquatic systems in the basin. In particular, the desiccation process of one of the main aquatic systems within CCB, the Churince, a wetland with a particularly rich input of deep water with magmatic influence (Wolaver et al. 2012) forming a series of aquatic systems that had a high and endemic biodiversity, that now are in extreme danger of disappearing (De Anda et al., 2018; Moreno-Letelier et al., 2018; Souza et al., 2007, 2012). In our study, we analyzed microbial mats from 3-sites within the Churince hydrological system over the course of two years to address the following questions: What are the changes in the community dynamics over space and time? How are the networks assembled from their basic building blocks? Are these building blocks the same in each mat? What are the keystone taxa? It is possible to discriminate between intrinsic community changes and those given by the environmental degradation? Finally, can we mechanistically predict the overall behavior in the community in face of environmental stress using network modelling? To this end, we used high-throughput Illumina sequencing to assess the community composition, function as well as the change in the structure and membership networks between sites during the perturbation and after the fortuitous recovery of the ecosystem. Our results indicated that even though microbial mats are resilient and resistant taxonomically and metabolically, their network of interactions changed during the dry period, being that under stress, negative interactions predominate, while they are balanced under wet conditions.

## Methods

**Sample Collection and Processing.** Microbial mats were obtained for the course of two years from 2012 to 2014 from a small (ca. 12 m × 4 m) and shallow (<0.42 m, but variable) pond named "Lagunita" (26.84810° N, -102.14160° W) that is lateral to the main Churince lake ( Lee et al. 2015, 2017) (Figure 1). During initial fieldwork, the Lagunita pond dried (Figure 1D), with the exception for two wet patches. Microbial mats were sampled from the wet and dry areas of three different sites; A, B and C. The microbial mat A (26.848120°N, -102.141604°W) was exposed to the atmosphere and consequently quickly became dry, displaying green and yellow tonalities. The microbial mats from the wet purple patches were named site B (26.848093°N, -102.141608°W) and C (26.848084°N, -102.141577°W) and both were covered with a thick white surface, and a strong sulfide smell.
Microbial mats from three sites were sampled seasonally (Autumn and Spring) at four times: November 2012, May 2013, October 2014 and May 2014, resulting in a total of 12 samples (3 sites x 4 time points) (Figure 2). During the period of study, the water levels within Lagunita pond changed from being dry in the first sampling to a temporal recovery of the aquifer caused by the closure of a canal that drains the wetland. The first sampling point was considered the dry or perturbed/stressed condition and the rest of the sampling times were considered as humid, not stressed. Physicochemical water parameters were registered at each sampling time using Hydrolab Mini Sonde® 5 Multiprobe SE, with the exception of the initial (dry) sampling (as there was not water to measure).

Triplicate samples of each mat were obtained for each time point (5 x 5 x 5 cm), using sterile Falcon tubes (50 mL), and stored at 4 °C and subsequently frozen in liquid nitrogen until processing in the lab.

**DNA extraction and sequencing.** Total DNA for each replicate was extracted according to Purdy (2005). In total, approximately 118 DNA extractions were performed in order to accumulate sufficient amount of high quality DNA for sequencing. Replicates samples within each site and time points were pooled into a single sample, yielding a total of 12 DNA samples of ≈11 µg of high molecular weight DNA. After checking the quality of DNA in each sample (Supplementary Figure 1 and Table 1), DNA was sequenced by Illumina MiSeq Paired-End 2 x 300 bp (550 bp insert size) CINVESTAV-LANGEBIO, Irapuato, Mexico

**Sequence processing.** Quality of raw reads was analyzed using FastqC v0.11. TruSeq Indexed Adapter and barcodes were removed using cutadapt v1.12. Low quality sequences were discarded with

Trimomatic (Bolger et al., 2014), using a sliding window of 4 bp, an average quality per base of 20, and min read length of 36 bp. The assembly of the trimmed reads was conducted with Megahit v1.1.1-2 (Li et al., 2014) using the option – presets meta-large.

The coding regions were searched from the obtained contigs using Prodigal v2.6.3 (Hyatt et al., 2010) with the -a option, to obtain the translated amino acid sequences of the predicted coding regions and -p meta option. The peptide amino acid sequences were then scanned against Pfam-A v30. The abundance profile of each Pfam domain in the metagenomic samples was obtained from a Perl script (extract_pfam.pl) which is part of the MEBS software suite (De Anda et al., 2017). The FASTA files of the resulting sequence contigs have been deposited in the MG-RAST repository under project number mgp80319.

**Binning**. We used the *k-mer* based taxonomic classification algorithm of One Codex (Minot et al., 2015) in order to assign microbial taxonomy of the Megahit derived from contigs. Briefly, One Codex classifies unknown nucleotide sequences according to the set of signature sequences within it that are unique to a specific taxonomic group using oligonucleotides of 31 bp (k =31). The taxonomic profiles obtained from the reference-based approach were used for downstream analyses described below. A summary of the mapped reads annotated with One-Codex database at each taxonomic level is provided in Supplementary Table S2.

**Microbial mats diversity, structure and statistics:** Several descriptors of alpha diversity were obtained from Phyloseq-estimate_richness function (McMurdie and Holmes, 2013) implemented in R (R Development Core Team). In order to estimate the sampling effort as well as include the seasonal diversity in two years of sampling, cumulative rarefaction curves were obtained for each sampling site by adding up at each sampling time the genera abundance with respect to the previous sampling in that site. Several statistical analyses were performed in order to test for differences between samples and estimate components of variation due to year, site or water conditions. First, a hierarchical clustering with bootstrap support was accessed with R-functions pvclust and clusterboot. Then, we performed a Permutational multivariate analysis of variance (PERMANOVA) analysis using (R-vegan function Adonis) in order to establish the differences between sites and times. A pairwise comparison among samples for the same site and water conditions was performed using the STAMP v2.1.3 program (Parks et al., 2014) in order to identify the statistically significant difference among genera within each sample by using Welch's t-test type two-sided, with the confidence interval (CI) method of Welch's inverted adjustment of 0.95. Then, we compared the taxonomic presence/absence profile from the microbial

mats under dry condition versus those from the rest of the sampling, the particular and shared taxa across water conditions were plotted with Venn-Euler diagrams using the web-based toolhttp://bioinformatics.psb.ugent.be/webtools/Venn/). Finally, we identify the most conserved genera across space and time using the parse_pangenome_matrix.pl, script part of the software get_homologues (Contreras-Moreira and Vinuesa, 2013).

**Biogeochemical cycles through time.** We use MEBS: Multigenomic Entropy-Based Score (De Anda et al., 2017), to evaluate the metabolic machinery of C, O, N, S and Fe cycles in the microbial mats across time by using a single value measured in bits (informational units). Briefly, the fasta peptide amino acid sequences of each microbial mats obtained from Prodigal were as the input of the main script mebs.pl We use the –comp option to compute the metabolic completeness (currently, it is only supported the methane and sulfur cycle). The cycles enriched in the microbial mats were obtained by using an FDR (False discovery rate) of 0.0001 by using the –fdr option.

**Network inference**. We use the time-series Lotka-Volterra-based network inference approach: MetaMIS (Shaw et al., 2016), to infer the underlying interactions from microbial mats collected during and after perturbations. For each site, the non-normalized taxonomic classifications (ranging from Phylum to family) were used to compute the consensus networks. Due to the large amount of network interactions generated at lower hierarchical levels and limitation of computing power (Interl® Core™ i7-4500U CPU @3.20 GHz processor and 16 Gb RAM), we were not able to obtain the consensus networks at genera level. Considering that the three studied sites are very close together, we constructed a "high order-network" by concatenating the consensus network interaction within all sites at every hierarchical taxonomic level. We named it "global network" and it was constructed to find general patterns that could reflect the behavior of the overall microbial mats within Lagunita pond. MetaMIS can include six types of interaction patterns: interaction patterns, including mutualism (+/+), competition (-/-), parasitism/predation (+/-), commensalism (+/0), amensalism (-/0), and no effect (0/0), we decided to further separate them it into more simple networks displaying only either positive or negative relationships. In this way, for each site that we obtained from three different networks (consensus, positives and negatives) at every hierarchical level (phylum, class, order, family) resulting in a total of 36 Time-Series Ecological networks (TS-Ens). The global networks were also separated by type of interaction at every taxonomic level, resulting in 12 global (TS-Ens). In total, we obtained a total of 48 TS-Ens.

**Motifs discovery.**

In order to compare our data with that derived from the ecological theory developed in food webs, and the tractable number of networks motifs, we focused only on the three node subgraphs, which there are 13 possible motifs. For comparison, there are 199 and 9364 motifs for four and five node subgraphs, respectively (Stouffer et al., 2007). To calculate all significant network motifs of three nodes within the 48 TS-Ens, we used Mfinder v1.20 (Milo, 2002) with default options.

**Network statistics.**

To further evaluate the topological features of the 48 TS-Ens, we developed a software package (available at https://valdeanda.github.io/Networks) that was built on the broadly used python libraries that are freely available, such as Networkx (Hagberg et al., 2008). The script NetworkAnalysis.py receives an adjacency list either weighed or unweighted and computes several metrics. We focused on identifying key features that are interconnected, therefore we extracted topological features of this directed graphs networks, such as density, hubs with maximum in-degree and out-degree and clustering coefficient. Then, we assumed the networks as non-directed and calculated further topological features for simple graphs such modularity and communities from the data using the Louvain method (Blondel et al., 2008). Given that the real networks are very dense in terms of connections, we implemented in our software a method to generate random networks that resemble the real ones using the option gnm_random_graph from Networkx python module. Then, hundred random networks were generated for each real network with the same number of nodes and edges of real networks.



## Results

**Field study overview.** Water depletion generated different environmental conditions reflected in the visual structure of the microbial mats of the three sampling sites (a dry and two wet sites). Site A was initially found dry under bright sunlight (November 2012). In contrast, microbial mats from the wet patches within Lagunita pond (sites B and C) were covered in November 2012 with a thick white surface, which possibly protected the microbes of the uppermost layers from direct sunlight, leading to contrast conditions from those observed in site A (Figure 1D). Physicochemical parameters of water during wet conditions are shown in Supplementary Table 3. Lagunita pond shows a conductivity

ranging from 6.47 mS cm-1 to 11.59 mS cm-1. At the end of the study, May 2014 salinity was almost the double in concentration; nonetheless, pH remained constant (≈8) for two years.

**Microbial mats diversity and community structure.** The metagenomic dataset comprising 12 libraries consisting in more than 22 million read-pairs, of which ~9.3% were discarded during quality control. The filtered reads were subsequently assembled, yielding 4,685,929 of contigs (N50 ~417 bp ± 34.36 std) with a total of 3,731,180 genes. Around 5 million proteins (~426,300 ± 116,000 std. per metagenome) were detected with Prodigal (see details in Supplementary Table S4). Protein sequences were scanned against the PFAM-A 30 database for functional assignment. After taxonomy assignment, 100 bacterial phyla, 168 classes, 302 orders, and 539 families were detected.

Within the phyla identified in the microbial mats from Lagunita pond, the most abundant among samples were Proteobacteria (85.36% std. 0.05), Actinobacteria (7.26% std. 0.02), Firmicutes (3.71% std. 0.03) and Cyanobacteria (2.69% std. 0.017). Within the Proteobacteria phylum, the dominant classes were alpha and deltaproteobacteria (41.43% std 0.095, 18.86% std 0.075 respectively). For visual comparisons, the overall abundances shift across time are shown at family level (Supplementary Figure 2). Although, there is a shared pattern among sites in the overall community structure, each mat is different. To assess the changes of each community member over time with respect to the first sampling, the family abundance profile was evaluated using the Log10 ratio of fold-change (log10FC) (Supplementary Figure 3). In three sites, the largest variation in community abundances with respect to the first sampling point is observed in site C. For each site, the number of genera sampled within microbial mats were in the saturation phase at the end of the period of study, indicating that from each sample, we required at least all the time points to recover their entire diversity using illumina MiSeq sequencing (Supplementary Figure 4).

To characterize better the community composition of each site across time, we identified the proportion of sequences assigned to each genus that was enriched among dry or humid conditions. The scatter plots indicating the relative proportion of all the genera for each site (Figure 3), shown that the most of genera are present in low proportions (i.e., < 5%), highlighting the importance of the rare biosphere. We also observed that mats from site A display similar taxonomic proportions during dry and humid conditions ($r^2$=0.873), suggesting a more resilient microbial community at taxonomic level. During dry conditions, unclassified sequences from Rhodobacteraceae, Alphaproteobacteria Rhizobiales and the genus Paracoccus were over-represented, however when water conditions were restored, several unclassified taxa from Proteobacteria and Actinobacteria phyla increased in site A. In contrast,

microbial mats from wet patches differed in their genera proportions during dry and humid conditions ($r^2$=0.202, $r^2$=0.397, respectively), indicating drastic changes in the microbial mat community among environmental conditions, as it was previously observed in Figure 3. Consistently with the visual appearance of site B, the genera enriched among dry conditions comprehended taxa implicated in the oxidation of inorganic reduced sulfur compounds (i.e., Ectothiorhodospiraceae, Thialkalivibrio), suggesting that the observed purple layer corresponds to the sulfur associated bacterial community. Furthermore, the taxa, which were found to be common in dry conditions, are bio-indicators of halophytic and anaerobic environments. The fact that Desulfatitalea was enriched among dry conditions within site C, and enriched in humid conditions within site B, indicates the wide metabolic capabilities of this genera.

The alpha-diversity estimators across sites are shown in Supplementary figure 5. In general, we observe that Site B is the less diverse especially during dry conditions. Unexpectedly, site A displayed the most diverse microbial mats samples at taxonomic level. The trajectories of Shannon and Pielou indexes during the period of study indicate little variations, pointing out a resilient microbial mat community. In order to avoid arch effect caused by the unimodal species response curves, we use an ordination correspondence analysis (DCA) method (Supplementary Figure 6). The PERMANOVA analysis of the taxonomic profile suggests statistical differences $p<0.05$ among microbial mats from the different sites and seasons (sampling points), in contrast with the metabolic profile. We only found statistical differences across sites indicating that even if they are few meters apart geographically, each microbial mat is functionally different, but each one remains functionally constant in time. These data are also supported by the hierarchical cluster dendogram (Figure 4). In order to determine whether the structure observed in the clustering represents actual structure in the data, or it is an artifact of the clustering algorithm, we included a Hierarchical Clustering with P-Values via Multiscale Bootstrap. By cross validating the perseverance or stability of the groups by the Jaccard coefficient (a similarity measure), we observed at the taxonomic level that the microbial mats samples are grouped in two well-defined and stable groups. The wet patches at the first sampling point (B1, and C1) are clustered along the rest of mats from site B, in contrast site A and C are clustered together within a different group. The Jaccard similarities (0.952 and 0.945) suggest highly stable clusters grouping microbial mats according to sites. At the metabolic level, microbial mats are grouped in for clusters, being only two of them stable. The first one grouped the microbial mats from site A and one from site C (C3), with a stability of (0.744). The second, with a Jaccard similarity of 0.933, contains samples from sites B and C after the

disturbance event. The wet patches at the beginning of the study are separated into another cluster, all of them displaying less stability.

The comparison across sites suggests a clear involvement of anaerobic taxa in mats from wet patches whereas the dry site indicates mostly photosynthetic and heterotrophic genera. Among differentially abundant genera from microbial mats from site C compared with those from site A (Supplementary Figure 8), several representatives of sulfate reducing bacteria are observed: Desulfococcus, Desulfosarcina, Desulfonatronospira, Desulfatibacillum, and Geoalkalibacter. In contrast, site A was enriched with marine-related genera such as Plesiocystis, Hyphomonas, Enhygromyxa, and Arenimonas. The same pattern was observed in mean proportions between mats from site A and B, being enriched with the genera Spirochaeta, unclassified Candidatus Daviesbacteria, Desulfonatrospira, Desulfobacterium in wet patches, whereas unclassified cyanobacteria along with Synechococcus, Nostococales and Oscillatoriales are differentially abundant within site A (Supplementary Figure 9). The Filamentous, non-heterocysts cyanobacteria Phormidium (Marquardt and Palinska, 2007) and the denitrifying beta-proteobacterium Aromatoleum are associated with degradation of several aromatic compounds, were only present in site A, and never detected in mats either from site B or C, suggesting that site A was under more stressed conditions than its close neighbors. When comparing the differentially abundant genera from wet patches (site B and C), we observed a representation of an aerobic community enriched within site C (Rhizobacter, Rhodospirillum, Skermanella) (Supplementary Figure 10). Despite the geographic closeness of the three sites, each one presents many elements unique from each community and a unique response of such taxa for perturbation.

The contrasts among microbial mats from dry vs. humid indicate that regardless of geographic site within Lagunita pond, eight genera are differentially abundant ($p<0.005$) during dry conditions: Marinovum, Pannonibacter, Paracoccus, unclassified Rhodobacteraceae, Roseivivax, Oceanicola, Wenxinia and Citreicella. In contrast, we observed a total of 24 genera enriched during humid conditions, including the potential phosphorous removal by Candidatus Accumulibacter, unclassified sequences from Nanoarchaeota, and the microbial mat related to genera Ornatilinea. (Supplementary Figure 11).

The taxonomic presence/absence pattern confirms that each site displayed a specific response to contrasting water conditions. A diverse community of bacteria, but also Archaea, fungi, and protists,

and a lesser number of viruses (Supplementary Figure 12A) were observed during dry conditions. In general, the common denominator was the presence of several halophilic representatives of fungi, and genera associated with protozoa organisms, however no shared taxa were found among the three sites (Detailed list is found in Supplementary Table 5).

In the case of site A, we found 35 unique genera that are only shown in dry conditions (Autumn 2012). From the Archaea domain, we found unclassified sequences from the particularly interesting TACK group (Lake et al., 1984) as well as the recently proposed phylum Aigarchaeota (Guy and Ettema, 2011).While a total of 24 genera were unique for site B in dry conditions, most of bacteria, 85 were unique for site C, including Archaea representatives from hydrothermal systems and adapted to high salinity environments ( i.e Ignicoccus and Natrinema) (Huber et al. 2000; Mei et al. 2017). See Supplementary Table 5)

As we already mentioned, despite the geographical proximity among sites (less than 3 meters), no common taxa were found during dry conditions in all three sites. However, we detected some intersections of taxa among two sites (Supplementary Figure 12B). Unexpectedly, the dry mat and wet patch mat C shared six taxa that were never present during humid conditions. These taxa highlight the implication of a halotolerant and C-cycling community when water-depletion occurred within Lagunita pond (i.e., Gramella and Kordia genus). Three common taxa were found in sites A and B in dry conditions, including a halotolerant representative from the pathogenic order Spirochaetales (including Borreliella), and sequences similar to the endosymbionts of Acanthamoeba spp. Common to C and B were also the halophilic Planococcaceae bacteria and the anaerobic Clostridial Peptoclostridium (more details in Supplementary Table 5).

During humid conditions, representatives of Fungi and Protista were not found in sites B and C, contrasting with site A, where we detected a genus of the Leotiomyceta class (Ascomycota). As it was expected, while few taxa are shared, many unique taxa exist in each separated community (especially in site B). In contrast to dry conditions, we found two taxa across sites intersections, whose presence indicates the re-establishment of water conditions across spatio-temporal scales (Ornatilinea and Rikenellaceae) (Complete list in Supplementary Table 6).

**Microbial mat core**. To distinguish stochastic fluctuations from underlying statistical laws which govern microbial mats, we identified 373 genera that are shared across microbial mats during contrasting environmental conditions. Among them, 69.97% corresponds to Proteobacteria, 10.45% of Actinobacteria and Bacteroidetes 4.82%. Interestigly, two halophilic Archaea were part of the

microbial mat core: an unclassified genus from the Halobacteria class and Halorubrum, as well as unclassified sequences belonging to the Eukaryota domain. Although, it is difficult to assess the contributions and the role of unclassified Bacteria and Archaea within the mats, we can highlight their putative ecological niches and microbial functional roles within the microbial mat core (See Supplementary Table 7).

**Biogeochemical cycling.** We evaluated the temporal variability of the metabolic machinery of the main biogeochemical cycles by using a single value measured in information units, bits (Figure 5). Due to the fact that our benchmark work of the main cycles across a thousand metagenomes is not still published, we used in this study a restrictive FDR to avoid false positives and detect the most enriched cycles within the microbial mats. By using this approach, we observed than methane and nitrogen cycles where over-represented in the microbial mats (represented with asterisks in Supplementary Table 8 and bigger markers in Figure 5). In the first case, the methane cycle is always enriched within site B, and become significant in all microbial mats after water recovery at the end of the period of study (Autumn 13 and Spring 14). The dynamics of the nitrogen cycle follow an opposite behavior for site A and C, being significant at the first sampling point for both sites. Site C is the only one that becomes significant at the end of the period of study compared with the rest of the mats. In our previous benchmark for the Sulfur cycle, we detected that all samples of site B after the perturbance event and the site A in autumn 2013 where those among the most enriched when compared with nearly 1000 metagenomes, having a score above the $95^{th}$ distribution (De Anda et al., 2017). The latter indicates that the metabolic machinery within microbial mats from the wet patch B is reduced to a few specific pathways during dry conditions (i.e sulfide oxidation by purple sulfur bacteria). However, when water returns to the Lagoon, the whole repertory of the sulfur cycle is represented, therefore increasing the Sulfur Score. As observed in Figure 5A, the tendency indicates that during dry conditions in site A, the overall machinery of sulfur cycle was underrepresented, compared with the wet patches site B and C. These results agree with the anaerobic differentially abundant taxa within wet patches indicating that the deep aquifer fluctuations largely determine the redox gradient within the mats affecting the anaerobic organisms. After water recovery, the overall S cycle fluctuated seasonally, and the site A recovered the overall S-capabilities. The recovery of the metabolic usage of CH4 compounds by methanotrophs, methanogens, and methylotrophs (methane cycle), is evident at the end of the period of study within humid conditions.

For the case of the oxygen and iron cycle, our results indicate that there is neither enrichment nor impoverishment of the above-mentioned cycles in our samples. For example, for the case of photosynthesis (oxygen score), our results agree with the low proportion of cyanobacteria detected in the taxonomic profile indicating an under representation of the oxygenic photosynthetic pathways in our community during the period of study. In the case of Nitrogen cycle that includes pathways involved in the reduction and oxidation of both inorganic (nitrate (+5) to ammonia (-3)) and organic nitrogen compounds (i.e taurine, urea, and choline degradation), we found that for site A, the nitrogen cycle increases after the water recovery (spring 2013), but then it diminishes at the end of the period of study. The reduction and oxidation of F compounds including siderophores uptake, were under-represented during dry conditions within site B, but then they become stable over time.

For a closer look, we used the relative abundance of several protein domains as a proxy of the potential relevance of each cycle including the proposed markers of the S cycle described in De Anda et al. (2017). The heatmap representation of these protein domains (Figure 5B) was not enough to infer the overall cycling of the main biogeochemical cycles. In general, the carotenoid biosynthetic pathway (CYP97 family), represented by cytochrome P450 (PF00067) (Cui et al., 2013) and the bb3-type cytochrome c oxidases (cbb3-Cox) (PF13442), which enhance photosynthesis gene expression (Erici et al., 2012), were found to be more abundant in site A after water recovery and poorly represented in microbial mats from wet patches. Consistently, Cytochrome C and Quinol oxidase polypeptide I (PF00115), key enzymes for aerobic and respiratory processes in aerobic photosynthetic organisms (Lea-Smith et al. 2016) were more abundant in site A during the entire period of study, in contrast with site B, its abundance was increased after water recovering (May 2013, October 2013).

We also analyzed other protein domains involved in the biosynthesis of mycosporines and mycosporines-like aminoacides, such as shinorine, known to be crucial for the protection against high UV exposure and to other stress factors in cyanobacteria (Balskus and Walsh, 2011; Katoch et al., 2016). Unexpectedly, a luciferase related protein family PF00501 (acting via ATP-dependent covalent binding of AMP to their substrate including the small C terminal domain AMP-binding C (PF13193)), was among the most abundant protein sequences found during the study (See detailed picture in Supplementary Figure S13).

In the case of the S cycle, we observed several protein domains that were more abundant in site B after water recovery at the end of the period of study, highlighting the reestablishment of redox and anaerobic conditions with water recovering within Churince Lagoon. These domains include Pfam family PF02662, involved in anaerobic metabolism including the methanogenesis of organic sulfur compounds; the 4Fe-4S di-cluster domain (PF14697); and the NADH ubiquinone oxidoreductase (found in sulfhydrogenase complexes of Pyrococcus furiosus, PF01058). In the case of nitrogen and methane cycles, no particularly abundant protein domain was observed across samples (Figure 5B), contrasting with the iron cycle, represented by some essential components involved in the energy production and conversion, such as the Iron-Sulfur cluster PF00355 and the protein domain found in the cyc2 gene, essential for Fe(II) oxidation (PF00034).

We have also employed the metabolic completeness for a closer look into the dynamics of the cycles. As stablished previously, the metabolic completeness is defined as the full repertory of protein domains involved in a given metabolic pathway such as sulfate reduction or methanogenesis (De Anda et al., 2017). As observed in Figure 5C, several pathways remain constant through time. The complete repertory of protein domains required to perform the usage of organic sulfur compounds such as sulfonates as Dimethylsulfoniopropionate (DMSP), (3-(methylsulfanyl) propanoate (MTP) (Bullock et al., 2014; Curson et al., 2008; Jonkers et al., 1998), as well as the oxidation of inorganic sulfur compounds were found in all samples. In the case of the elemental sulfur disproportionation, we observe that it is absent in the stressed mat A and it was only present in autum 2013. In contrast, site C displayed the opposite behavior, being autumn 2013 the only time where the elemental sulfur disproportion was not complete. Remarkably, the dimethyl sulfide (DMS) oxidation pathway reported for aerobic microorganisms (such as *Thiobacillus*, *Hyphomicrobium* and anaerobic methanogens and sulfate reducing bacteria) was found to be 100% complete after the disturbance event in both sites (A and B), and in the site C during dry conditions. A similar pattern was observed in the case of the thiosulfate oxidation, which was found to be complete in site B, after and during disturbance, respectively (Figure 5C).

**Network inference method based on time-series data**

First, we evaluated whether the regenerated abundance profiles obtained from MetaMIS reproduce successfully the microbial abundance similar to the original by using the Bray Curtis Dissimilarity (BCD). As observed in Supplementary Table 9, a small BCD (mean 0.136 ±0.021 std.) was obtained in the three sites at all taxonomic levels, suggesting that interactions revealed successfully the underlying interactive relations of the microbial mat communities. We also confirmed that the majority of taxa within microbial mats, were found to be rare <0.1% (75% ±3.06 std.).

From 1600 intermediate networks generated by MetaMIS, we obtained 48 TSEns comprised by the 100% of the total interactions at every taxonomic level (See details in Supplementary Table 9). For the consensus TSEns, we focused on the amount of positive and negative interactions within sites. Approximately half of the interactions were found to be equally positive and negative in site B and C. In contrast, site A displayed a larger percentage of negative interactions (0.616 ±0.012 std.), compared with the positive ones (0.384±0.011, see Table 1).

In addition to this, to evaluate the 48 TSEns, we developed a software that computed several global measurements and generated for every given real network, a hundred of random networks matching the same number of nodes and edges than the real ones. The real measurements were, then, compared against those obtained from random networks. As observed in Table 2, due to the high number of relationships (highly dense networks) and due to method used to generate the random networks; it is not surprising that when we compared them, the structure of both real consensus and random networks are similar. These data suggest that a large number of taxa, probably from the diverse rare biosphere, co-occur randomly, and are not bacterial consortia that coevolved together. Although, several methods were tested to generate the random networks (data not shown), the values were also similar to the real ones, indicating that small world proprieties do not characterize our TSEns.

Furthermore, our results indicate that microbial mats are thought to be more robust to change due to functions that are likely shared among nodes or functional clades in the highly connected TSEns, with density values close to one ( ≈0.9) (Faust and Raes, 2012a; Montoya et al., 2006; Shaw et al., 2016; Steele et al., 2011). Moreover, the clustering coefficient gives information about the grouping of nodes within their immediate neighborhood, while modularity gives a similar information at larger scale. The clustering coefficient C1 is similar in the negative and the consensus networks (≈0.4); however, this association is present within the positive interactions to a lesser extent (≈0.2) (Peura et al., 2015). Networks with a high clustering coefficient are likely to contain more hubs or focal nodes than those with a lower coefficient (Proulx et al., 2005). Hence, the loss of these hub nodes, which

504 have been likened to 'keystone' species (Steele et al., 2011), reflects potential structural perturbations
505 to the community, and suggests some degree of community stress as bacterial associations have been
506 fractured. Our data indicate that negative relationships within mats may retain a greater number of hubs
507 than positive interactions, suggesting an important role for competitive interactions in stressed
508 conditions. As indicated by (Grilli et al., 2016), positive modularity values indicate that interactions
509 occur predominantly within groups, while negative values indicate that interactions are more frequent
510 among groups. In our study, while consensus networks indicate values of modularity close to cero,
511 indicating that taxa within microbial mats prefer those members from another subsystem due to the
512 highly densely connected network. However, when we separated the consensus networks by their type
513 of interaction (+ or -), we found that modularity increased in the negative associations (0.2), suggesting
514 clear preferences for competitive interactions within the same subsystem (Table 2). Our results indicate
515 that modularity and clustering coefficient of our networks were very low compared with other biological
516 networks (Baldassano and Bassett, 2016). The low modularity values indicated that our TSEns lack an
517 evident modular architecture. We can suggest that modularity values close to zero reassemble the idea
518 of microbial mats acting as a whole complete module per se, with all the parts interacting with each
519 other.

520 By evaluating the TSEns by their type of interaction and compared them again to random
521 networks (Table 3), our data indicate that negative relationships within mats may retain a greater
522 number of hubs than positive interactions, highlighting again, the important role for competitive
523 interactions within the mats. We found that modularity increased in the negative associations,
524 suggesting large groups of mutually excluding taxa and a clear dominance of competitive interaction
525 within the same subsystem, in agreement with the clustering coefficient. We also observe that both,
526 the number communities and members within each one, vary at every taxonomic level (Supplementary
527 Table 10).
528
529 The consensus TSEns at family level for each site are shown in Figure 6. To facilitate visual
530 representation, we displayed only the 0.05% of the total number of interactions, on the contrary if we
531 displayed the 100% of the interactions, the network will be highly complex and so densely connected
532 that patterns would not be appreciated clearly. At this level of resolution, we observed a crucial role of
533 the family Rhodobacteraceae as a hub within the three sites, having more negative associations within
534 site C. The family Desulfobacteraceae is also a key component within wet patches, having both
535 negative and positive relations with Fungi representatives (Aspergilacceae) in site B and C

respectively. Interestingly, low abundance taxa are represented as hubs within site A (Bacillaceae), and unclassified members of the cyanobacteria Oscillatoriales.

In order to identify whether the pattern observed in this low-resolution networks (Figure 6), where also presented in the TSEns composed by the total number of relationships, we identified highly connected nodes either affecting (max out degree), or affected (max out degree) by many links in the consensus, positive, negative and global TSEns. To get further detail or infer possible metabolic roles, we focused only on the nodes at family level summarized in Table 5. We identified the nodes that are part of the microbial mat core (C) (see Table S13), rare biosphere (RB) (with average relative abundance of less than 0.1% from the total number of microbial sequences) and capable of delivering public goods to the system (PG) (with reported evidence to performe such task). In the case stressed mat, we identified in the consensus TSEns, eight hubs affected by 337 nodes, indicating a high interconnectivity within the overall system, and one hub affecting 345 nodes represented by the Polyngiceae, a type of Myxobacteria that are known for excreting hydrolytic enzymes and decomposing various and complex biopolymers (Brinkhoff et al., 2012). The number of hubs in the positive and negative networks decreases to single keystone taxa that belong to both the rare biosphere and the microbial mat core (See details in Table 5).

Our results indicated that either the number of keystone taxa affected or affecting other members in the microbial mat community within site B is equivalent, being the majority classified as members of both rare biosphere and microbial mat core. An interesting example, within these hubs, is the family Thermoactinomycetaceae, which is known for its protein degradative capacities, strong lipolytic activity and alpha-amylase activity, as well as antimicrobial activity (Frikha Dammak et al., 2017). When the consensus network of site B is separated depending on the type of interaction, two hubs are observed within the positive TSEns, Staphylococcaceae and Fusobacteriaceae, both belonging to the rare biosphere. In contrast, within negative relationships within site B, we found Phyllobacteraceae as max-in degree (also a hub in the consensus TSEns), and two max-out-degree hubs (Cystobacterineae and Cloacimonetes), both unclassified sequences belonging to the rare biosphere and microbial mat core.

By analyzing the consensus network from the microbial mats of site C, we found two hubs, belonging to the microbial mat core, an in the case of Verrucromicrobia is also present in low abundance (rare biosphere). Interestingly, this phylum is known to perform sacharolytic lifestyle commensal and mutualistic relationships with ciliates (Vannini 2003). Besides, an unknown group of archaea seems to perform a crucial positive role within microbial mats, being affected by 252 nodes,

and the same for the family of Actinobacteria (Micrococcales), that was detected as max out degree hub within positive TSEns. When we focus only on the negative relationships within site C, six max-in-degree hubs appear. Four of them belonging to either the rare biosphere or microbial mat core (see more details in Table 5). Among the taxa that do not belong to the microbial mat core (but are part of the rare biosphere), we identified Alteromonadaceae and Beijerinckiaceae families. Within the same negative network of site C, we found only one max-out-degree hub, belonging to unclassified members of the phylum Bacteroidetes. Interestingly, these members are usually abundant during and following algal blooms, which are specialized in degrading high molecular weight compounds, showing a preference for consuming polymers rather than monomers (Fernández-Gómez et al., 2013).

Interestingly, we found that all the hubs identified in the global consensus TSEns belong to the rare biosphere. As an example, we identify the family Desulfarculaceae that is part of the microbiomes in particularly successful microbiota-host mutualistic configurations, and it has been considered to support longevity and the host homeostasis in complex environments (Debebe et al., 2017). Unexpectedly, we also found as max_ out-degree within the positive network many unclassified sequences from Eukarya domain whose presence would increase the energy transfer and therefore trophic complexity and potential resilience to environmental change (Duffy and Stachowicz, 2006).



**Network motifs**

Our approach focuses on the representation of three-node motifs over 48 TSEns that reflect the behavior of microbial mats from three sites during and after water depletion. The general pattern of representation of network motifs is shown in Figure 7. In our study, we considered only those motifs whose probability of appearing is lower than a cutoff value (here $P=0$ and $P\leq0.05$) according the distribution observed in randomized networks (Jin et al., 2007). These motifs appear numerous times on each particular network at every taxonomic level observed in Figure 7. For visual comparison, we normalize the abundance of each motif by the sum of the total across taxonomic levels. Therefore, the abundance of one indicates that a given motif is only found in that particular taxonomic level. As observed in Figure 7, particular motifs appear over-represented across the span of taxonomic levels when consensus networks are separated by type of interaction either positive or negative. For example, motifs 6, 36, 78, 102 and 110 are under-represented in the negative TSEns but are highly distributed across the positive interactions. This is particularly interesting since this motif 6 (also known as exploitative competition) occurs when some quantity of resources is consumed by an individual,

602 thereby depriving other individuals of it. This type of competition has also been called consumptive
603 competition (Kawata M, 1996), and was never found among the negative TSEns within our microbial
604 mats, suggesting that this motif was able to have different ecological connotation, for example one
605 bacteria delivering public goods to the system. The same pattern is observed for the apparent
606 competition motif (36), that occurs when two individuals, which do not compete directly for resources,
607 are affected each other indirectly by being prey for the same predator (Lang and Benbow, 2013).
608 However, in a microbial ecological context, the fact that this type of motif is mainly distributed across
609 positive relationships might suggest that is more likely to observe for example two bacterial species
610 eating detritus or protozoans feed on bacteria.

611 We found that motifs 12, 46, 108 and 238 are widely distributed across negative TSEns, but
612 under-represented within the positive ones. These motifs seem to display more complex relationships,
613 for example motif 238 represents the most mutually exclusive relationships in which all nodes exclude
614 each other, in a cycle form. Consistently, with the ecological theory (Stouffer et al., 2007), the tri-
615 trophic food chain motif, composed by prey, predator, and top-predator was over-represented within
616 negative networks compared with the positive ones. However, we found that this network motif could
617 also imply positive relationships due to it was only found within site B at phylum level. Furthermore,
618 the tri-trophic chain motif is also found within the consensus networks suggesting both beneficial and
619 negative interactions (Baiser et al., 2016).
620 In the case of the motif (ID102), we can argue that the relationships are entirely mutualistic or
621 beneficial since this motif was not found within the consensus or negative TSEns.

622 The Feed Forward Loop motif is one of the most significant in *E.coli* and yeast representing a
623 structural stable motif with no feedback interaction and it has been found to be over-represented in
624 transcriptional networks (Mangan and Alon, 2003; Shen-Orr et al., 2002). In ecological theory, the
625 same motif, involves omnivory representing a predator consuming species from two different but
626 lower trophic levels (Bascompte and Stouffer, 2009; Stouffer et al., 2007). However, in our microbial
627 mats TSEns, this motif does not involve entire positive or negative relationships, but it is more likely
628 to have at least one type of interaction since it is only found in the consensus TSEns.

629 While the Feed Back-Loop (FBL) has been associated as an unstable network motif under-
630 represented in gene regulatory neuronal networks (de-Leon and Davidson, 2007), in our study, we
631 observed an overrepresentation of these motifs within the positive networks consistently at every
632 taxonomic level (with exception of site A at phylum level). This motif could be an example of positive
633 metabolic coupling occurring within microbial mats.

Under the hypothesis that environment can influence the strength and type of interactions between taxa, we expected to find similar interaction patterns between the wet patches. Interestingly, we found that only site B displays a specific type of network motif ID74. The latter matter highlights a specific community dynamic and behavior even within the same environmental constrains, indicating local roles within site B that are not found within the microbial mats from the other sites.

## DISCUSION

Given the remarkable resilience of microbial mats due to their capacity of self-regulation, our hypothesis was that if microbial mats are stable over time -- regardless of the perturbation produced by human water overexploitation -- we should expect that samples taken through time in fixed time points in space would present similar community patterns. In contrast, if these communities have survived not by staying constant but by constantly changing, we would expect systematic differences that could be reflected in the emergent community properties.

In order to test such hypothesis, we used a time-series metagenomic network-based approach to understand how environmental perturbation affects the stability and dynamics of microbial communities. In order to distinguish among intrinsic shifts generated by the community *per se* from those derived from outside forces, we studied over the course of two years the microbial mats of three sites, close but geographically separated within a small Pond called Lagunita. The Lagunita pond is a well-studied site (Lee et al., 2015, 2017) since we had analyzed its microbial diversity in a nutrient enrichment treatment in 2011, observing that microbial communities of both water and sediment changed in respond to changes to the N:P ratio.

Using metagenomics and network analysis, we analyzed the shifts in bacterial community composition, structure and function in a time-series of 4 points and explored the extent to which all these changes were reflected in the specific building blocks of interactions within each community and in the overall networks structure. This contrast with similar studies where observations are commonly derived from single time points (Preisner et al., 2016).

Even though, the three sites are separated by a distance of a few meters, the particular environmental conditions found within wet patches in Lagunita pond during the desiccation event (November 2012) are fundamental characteristics of well-developed redox anaerobic environment, which could account for the predominance of anaerobic taxa and complete metabolic pathways (e.g., Desulfatitalea, Ectothiorodospira; sulfite oxidation) (Eren et al., 2014; Grimm et al., 2011; Higashioka et al., 2013; Sorokin et al., 2006). Due to the oxygen $O_2$ gradient, it is likely that microbial mats within

sites B and C were permanently anoxic and sulfidic. In contrast, the microbial mat from site A was exposed to bright sunlight, and in direct contact with the atmosphere, leading to a harsh photo-oxidative microenvironment at its surface. This is relevant, given the remarkable physicochemical-zonation of microbial mats on millimeter scales characterized by steep vertical gradients of oxygen, pH and sulfide (Chennu et al., 2015), and thus it is expected that lack of water had a deep effect not only in the community composition structure and function, but also in the community-level relationships and interactions, specially, when considering that metabolite exchanges in microbial communities give rise to the ecological interactions that govern ecosystem diversity and stability (Zomorrodi and Segrè, 2017).

**Microbial mats from ultraoligrotrophic pond are highly diverse**

According to the estimated coverage, our time-series metagenomic approach was more comprehensive that the previous single point (control) estimates of the same site (Lee et al., 2017; Okie et al., submitted). The follow up of the diversity for two years showed an accumulative curve that came to an asymptote in the last sampling. For instance, our data show that microbial mats from CCB exceed around three times the number of taxa within mats from Lake Clifton, Australia, indicating a highly diverse microbial community (Warden et al., 2016). This was not unexpected in wet conditions, since previous studies in water and sediment bacterial communities of the same pond showed a large diversity (Lee et al., 2017) , but to us was unexpected to find also such large diversity in the dry mat of November 2012. Given the extremely low P conditions, we know that this is not just dead DNA, since this is the favorite source of P on all the tested microbes (Tapia-Torres and Ólmedo-Álvarez, 2018) .

The observed differences in richness and diversity with Chao and Shannon indices may be because Shannon may fall short when examining a large number of low abundant organisms (the rare biosphere), that in our case, it represents the most of the taxa in our microbial community. The large amounts of low-abundant taxa indicate that this environment has a rich “seed pool” of genetic diversity, suggesting not only a biological complexity at the micron scale, but also a very dynamic structure with different ecotypes with apparently overlapping ecological features. As expected by the ecological theory on perturbations (Eng and Borenstain, 2018), the modifications of environmental conditions lead to changes in the proportion of the members of species in the community (measured at genera level). However, when we explored the relationship between taxonomic and metabolic diversity estimated with ecological metrics (i.e., Shannon, Pielou, hierarchical clustering, rarefaction curves),

and the functional potential for several biogeochemical cycles using MEBS algorithm, the site A (dry in November 2012) was much more diverse and resilient over time than expected. Although wet patches do not display the same taxonomical dynamics, they retain functional similarity that may be fulfill by different key members of the rare biosphere. A possible explanation for this result is the micro-dynamics of the deep aquifer, it is possible that site A had a more fluctuating environment in last years given that the deep source of water is a little bit farther, making it apparently dry as the aquifer gets smaller. Fluctuating environments have been shown to be more diverse in the CCB environments (Bonilla-Rosso et al., 2012; Pajares et al., 2015; Peimbert et al., 2012). Moreover, by using MEBS, we were able to capture the fluctuation dynamics of the whole metabolic machinery involved in the main cycles, not only by focusing in a few marker genes, but rather by following the importance of their dynamics over time. Our results indicate that the anaerobic cycles within microbial mats from the wet patches are maintained by the redox conditions that the deep sulfur rich aquifer gives to this ecosystem. This becomes more apparent when the aquifer recovers apparently, and the sulfur cycle becomes over-represented within the dry patch (site A, autumn 2013). This is also evident for the case of the methane cycle, where we observe that, when water returns to the Lagoon, microbial mats from the three sites have an overrepresentation at the two last sampling points (autumn 2013, spring 2014).

Following this logic, if we detected an overrepresentation of any given biogeochemical cycle we can argue that conditions allow the "seed bank" of the rare biosphere or dormant microbes (Shoemaker and Lennon, 2018) to promote the growth of microbes that show a particular metabolic repertory. For example, the sulfur cycle was only significant after the recovery from the desiccation event in the microbial mats from site A where they became over-represented compared with the rest of microbial mat samples. Even though, site A had the overall machinery of sulfur cycle underrepresented under initial dry conditions, this changed quickly when the water returned. Microbes were waiting for the proper ecological opportunity within the mat or maybe deep-water migrants that filled the empty niche. We do not know, both possibilities exist. However, on one hand, the high diversity within mats represented by the large number of rare taxa could indicate that the first possibility is likely since a large number of key stone taxa in the majority of hubs are either members of the rare biosphere or part of microbial mat core. On the other hand, the second possibility may imply that there is a larger "seed bank" under the mountain in crevices of the deep aquifer, this seed bank can fulfill several empty niches and explain their fast response of microbial mats for changing environmental conditions (Cadotte et al., 2015; Eng and Borenstain, 2018).

Along with this "seed bank", there are other important keystone taxa that are the dominant within mats, especially within dry conditions. For example, we can associate the abundance of the physiologically and metabolically versatile Rhodobacterales to their potential as primordial colonizers for the formation of biofilms in aquatic environments, which can explain their adaptation to dry environments by forming biofilms that are able to resist drying (Dang et al., 2008; Elifantz et al., 2013).

Moreover, our results suggest that taxonomical diversity per se is not a good indicator of community function and complexity in face of environmental disturbance, since we found no differences during and after water depletion within microbial mats from site A (dry patch), which appears to be "used" to perturbation. However, when comparing the diversity at metabolic level, the functional redundancy was observed within mats from the wet patches (site B and C), that share similar metabolic composition during humid conditions. The latter matter was also confirmed by the anaerobic community that was shared, such as purple sulfur bacteria and sulfur reducing bacteria, suggesting the reestablishment of redox conditions and the stability in wet patches, also proved with an overrepresentation of anaerobic cycles within wet patches. Indicating that although there were shifts in the community composition, the retention of functionality indicates a shared function in its response to water reestablishment towards a normal condition within the pond.

Although, the three sites are very close in space, the fact that they have different dynamics and diversity along the time-series indicates site specific community dynamics. This is not surprising, given the large microbial beta-diversity observed in CCB in general (Bonilla-Rosso et al., 2012; Espinosa-Asuar et al., 2015; Pajares et al., 2016) and in the studied pond in particular (Okie et al., submitted). We believe that such community differentiation in the space could be due not only to historical factors, as early colonizers could establish the ground rules of interactions, but also by stochastic events such as virus predation of the dominant taxa in a "king of the hill" model, which can keep a very dynamic system with a very large rare biosphere. The changes in diversity and function of site A, in particular, confirm our hypothesis that water conditions in the Churince Lagoon are important factors influencing metabolic function-composition within microbial mats. Even though, the lack of water constitutes an obvious environmental filtering for aquatic microbes (Monard et al., 2016; Pontarp et al., 2012), the fluctuation of such important recourse seems to be is playing an important role in the distribution and abundance of the taxa shaping microbial assemblies within the mats

**Particular interactions within microbial mats: network motifs**

Modeling of interactions through networks is considered a powerful strategy to understand the dynamic succession of communities through time and stability within a system (Coyte et al., 2015; Delmas et al., 2017; Thébault and Fontaine, 2010). Unlike, previous studies in which only the top 25% of interactions were considered for subsequent analysis (Weng et al., 2017), we used 100% of interactions in order to gain meaningful interpretation of the relationships among taxa. In this regard, given the complexity of the community, we required first a coarse-grained approach in order to detect general patterns consistent across spatio-temporal scales.

We used the consensus network relationships and then dissected them into positive and negative ones to obtain meaningful statistical properties. For instance, the network density or degree of network connectivity gives an idea of how quickly perturbations may spread, by providing a measure of how dense the network is; a small diameter indicates presence of a densely connected nodes, or hubs, hence fast propagation among nodes, which may make the network more sensitive to perturbation (Delmas et al., 2018). The large network density obtained, would indicate that in microbial mats, the networks of bacterial community are composed by highly connected groups, and unlike to low density networks (Sun et al., 2013), it is most likely that this largely connected community is more robust to change. In addition, it has been suggested that a modular organization of species interactions would benefit the dynamical stability of communities. The modular view of the networks dynamics allows to analyze different groups of nodes performing different functions with some degree of independence (Newman, 2006). In all our samples, the actual network values were similar to the random networks, indicating that small world and modular properties are not prevalent within our time-series. Values approaching $Q = 1$, which is the maximum, indicate strong community structure. In practice, values for such networks typically fall in the range from about 0.3 to 0.7, and higher values are rare. (Newman and Girvan, 2003). More recently, it has been suggested that the modularity maximum always does not correspond to the most pronounced community structure of a network (Fortunato and Hric, 2016), highlighting that probably complex densely connected networks, such those analyzed in this work, might point out that modularity values are indeed lower than as it has been previously suggested. (Blondel et al., 2008; Fortunato and Hric, 2016; Newman, 2006; Newman and Girvan, 2003; Poisot, 2013)

**Representation of network motifs in the microbial mats.**

Ecological networks literature names these module motifs, as they represent typical relationship among species (Freilich et al., 2010; Milo, 2002). There are thirteen possible three-node motifs in directed networks, each one represents a different relationship among three nodes (Milo, 2002). Among these motifs, some of them are absent or present with lower or higher frequency, depending on the type of interaction (either positive, negative or both). We focused on the frequencies of each one of these motifs to infer information about network structure within each site at every taxonomic level (Figure 7). We observed that the ecological motifs (Exploitative competition, tri-trophic food chain, apparent competition and Omnivoy or FFL), which have the greatest probability of being stable in most macro-ecological communities since they are the ones observed at higher frequency in food webs, were able to have different meanings when we analyze microbial communities. Furthermore, the distribution of the motif 102 in our microbial mats time-series is only present in positive type of relationships, this can explain why his type of motif has not been encountered within in silico metabolic networks of E. coli (Sonnenschein et al., 2012).

Under the hypothesis that environment can influence the strength and type of interactions among taxa, we expected to find similar interaction patterns among wet patches (Sites B and C). However, site B was the only site that displayed an specific type of interaction, type 74, a motif whose arrows suggest cross feeding among two members of the interaction and a type of "black queen" dynamic (Morris et al., 2012) with a third suggesting a very particular mutualistic community dynamic and behavior in this site, even within the same environmental constrains. Our hypothesis under this scenario is that motif abundance can be related either to the particular members of the community or to the particular history of environmental conditions. The overrepresentation of this particular motif 74 in other similar environments remains to be seen experimentally since all these relationships are part of microbial market and it may be context dependent (Mccully et al., 2017). Even though, the time-series ecological interactions that we analyzed in this work are highly dense and similar to those obtained from random networks, the network motifs are not random, therefore the fact of evaluating such motifs provides a link to understand the particular relationship dynamics within microbial mats.

Recently, these motifs have been used to define species trophic roles in the context of their community, and link these roles to the network's stability (Borrelli et al., 2015; Delmas et al., 2017). As far as we know, this is the first study to incorporate network motifs for the analysis of microbial mats under perturbation conditions, we anticipate that further studies are needed to corroborate whether the overrepresentation of network motifs is specific of microbial mats, compared for example with neuronal, transcription and food webs (Borrelli et al., 2015; Tran et al., 2013) . To better

825 understand why some motifs are found in high or low abundances within our microbial mats, we need
826 to explore not only the mathematical properties of such networks motif but also to design experiments
827 of direct interactions to understand the ecological meaning of generalist and specialist within each
828 node.

829 **Keystone taxa are part of the rare biosphere and "microbial mat core".**
830
831 In general, we found that site B has a higher number of hubs in max out degree (4 in total), meanwhile
832 site C has three and the site A has only one, suggesting that it is possibly more fragile, because if one
833 of them is removed, more relationships may be lost, compared to sites B and C (See Table 5). In
834 addition, we observed that most hubs belong to the rare biosphere and the core of mats and both seem
835 to provide public goods. This is particularly important, because generally in ecological studies there is
836 no way to discriminate their role within natural systems by simply highlighting their low abundance.
837 However, in our study we found that rare biosphere contains crucial factors for the resistance and
838 resilience of the whole system under environmental perturbation.
839 We also observed that some of the hubs, which were detected independently in the three sites,
840 were also detected in the global network. Interestingly, unclassified members of Eukarya domain were
841 detected as a hub in positive relationships. It has been observed that diverse communities of eukaryotes
842 live in microbial mats, not only a broad range of taxa but also a large functional diversity, including
843 phototrophs from several algal phyla and a variety of heterotrophic organisms, such as fungi and
844 protists. The microbial mat provides microhabitats in the desiccation/humid conditions, which gives
845 protection against reduce oxidative, osmotic, freeze-thaw, and dehydration stresses for all
846 microorganisms embedded within the microbial mat matrix (Jungblut et al., 2012). A global positive
847 interconnection among unknown eukaryotic taxa within the global network in microbial mats of
848 Churince is important to highlight regarding the presence of saprophytic, phagotrophic, parasitic and
849 predatory eukaryotes that would increase the number of links within a mat for nutrient and energy
850 transfer (Duffy and Stachowicz, 2006). It would be interesting, if the site persists to this very long
851 period of dryness to further investigate whether the removal of certain taxa can affect the entire system.
852 Sadly, we do not know if Churince will come back as a phoenix if the proper water policy is
853 implemented in order to restore this amazing biodiverse site.
854 **Drivers of stability in microbial mats.**

855 Our results are in agreement with other studies where it has been proposed that negative
856 interactions increase the resilience of microbial community (Coyte et al., 2015; Deng et al., 2016;

857 Foster and Bell, 2012; Zelezniak et al., 2015), in addition to the cooperative network of microbes (i.e.,
858 one characterized by mutualism) that are often unstable, while a higher proportion of competitive
859 interaction pairs (−/−) helps the host to maintain a stable microbial community (Coyte et al., 2015).
860 Despite variation in environmental parameter values, when the resources are limited some species
861 outcompete with others and the stability is reached when the final community with at most one species
862 per resource establishes different links (Borrelli et al., 2015). It has been suggested that modularity is
863 able to have a positive effect on stability only when (a) the system is composed of two subsystems of
864 about the same size, and (b) the overall mean interaction strength is negative (Grilli et al., 2016). Here,
865 we observed that in the stressed mat, the negative interactions are on average larger than the positive
866 ones, consistently with the above mentioned, was also the most resistant site measured by ecological
867 indexes. Our results indicate that under normal conditions there is an equilibrium between positive and
868 negative interactions, due to the metabolic interdependency based on cooperation or mutualistic
869 relationships and such exchanges can provide group advantage under nutrient-poor condition (Morris
870 et al., 2013). However, under environmental disturbance, the negative relationships exceed the positive
871 ones. This type of behavior is typically found for instance in food webs due to the low efficiency of
872 transformation of prey into predator (Allesina et al., 2015; Grilli et al., 2016).

873 On the other hand, it has been suggested that large systems, in which the positive effects
874 dominate the negative ones, are unstable and will likely lose stability through a hop bifurcation
875 (conversion efficiency of resources of consumers) (Allesina et al., 2015). Hop bifurcation should be
876 most common in the presence or an inverted biomass pyramid, typically occurring in planktonic or
877 other aquatic systems (Allesina et al., 2015; Grilli et al., 2016). In the microbial mats, the ratio of
878 cooperation versus competence is under equilibrium, and may indicate an ability of microbial mats to
879 increase, negative interactions under perturbation with specific taxa (see Table 5). Besides, it has been
880 suggested that a strong network of interactions among organisms can provide a buffer against
881 disturbance beyond the effect of functional redundancy, as alternative pathways, with different
882 combinations of microbes, can be recruited to fulfill specific functions, thus increasing the negative
883 interactions (Konopka et al., 2015).

884 Energetically and nutritionally, beneficial metabolites and nutrients are exchanged among
885 participants leading to optimal production and nutrient cycling of the community as a whole. It has
886 been tested the hypothesis that closely coordinated metabolic associations promote homeostasis and a
887 buffer against stressful, resource-limited conditions (Guerrero et al., 2002; Konopka et al., 2015; Wong
888 et al., 2017). Hence, there is a metabolic and ecological congruency, as long as biogeochemical and

environmental gradients allow individual niches to exist in close proximity, thus metabolic diverse microorganisms are oriented according to energetic, nutrient, and ecological requirements and tolerances (Guerrero et al., 2002; Wong et al., 2017).

Metabolic dependencies based on the Black Queen hypothesis (Morris et al., 2012), function as starting point for the evolution of cooperative behavior, where the cross-feeding (bidirectional dependency) is obligated in communities with essential functions that are costly for producers (Mas et al., 2016). To explore Black Queen ideas, we separated the effect of network complexity from specific traits of individual members in hubs. To find keystone candidate taxa important for the maintenance of structure and function of a community, we focused on microorganisms from the microbial mat core (present in all samples) and low abundant taxa (rare biosphere), to infer putative ecological niches and functional roles. We found that the average interaction strength on dry sampling mats had strong influence on the stability, pointed out as drivers of stability. Our study supports the notion that many species occur in trophic local clusters, commonly occurring with only a few other species (low degree), while some species serve as hubs, establishing mutualistic relationships, allowing co-occurrence of several small niches (Faust and Raes, 2012b; Mccully et al., 2017; Thébault and Fontaine, 2010; Zhou et al., 2010).



## Conclusions

In microbial mats, the biogeochemical cycles through networks of metabolite exchange are structured along energetic gradients (Guerrero et al., 2002; Wong et al., 2017). As energy yields become limiting, these networks promote co-metabolic interactions to maximize energy disequilibrium (Wong et al., 2017) and the recipient-biased competition for a cross-fed nutrient promotes mutualism stability. Thus, when there are more species than resources, some of them will invariable outcompete with others, in theory resulting in a final community with at most one specie per resource, and this community in theory should reach equilibrium and stability despite variable environmental parameter values, however, the links among species are different through time (Borrelli et al., 2015). The Churince Lagunita system gave us an extraordinary opportunity to explore the relationship among environmental conditions, diversity, function and interactions (all variables that constitute the multidimensional niche) within a unique Time-Series Ecological design.

We observed that even if each site has all metabolic functions, the diversity and the response of such diversity in time is unique. The analysis of the capabilities of each metagenome to perform the different biogeochemical cycles shows that when the aquifer reestablishes its deep flow, the anaerobic

dependent functions within the sulfur and methane cycle get also reestablished. This "rebound" could be due in part to a large potential "seed bank" that makes the microbial mat redundant and diverse along with resilient to change. This is confirmed in its interaction motifs, where there is a type of "fingerprint" of each unique site even though they are few meters apart. The microbial mat under stressful conditions, the dry mat, has more negative interactions than the other more wet communities were mutualistic interaction balance with antagonism; nevertheless, these motifs change as the same site recovers, suggesting that the rare biosphere has an extraordinary role in the permanence of these ancestral communities in geological time scale.

**Figures**

Figure 1. Location of the field study site. A) Cuatro Cienegas Basin is located at the North of Mexico. B) In the left-hand side of the basin, all aquatic systems are sulfate-rich receiving water input from an aquifer regional flow. The right-hand side of the basin is rich in carbonates, according to Wolaver et al. (2012). C) Churince system D) Lagunita pond within Churince system and the microbial mats sampled during desiccation event (November 2012).

**Figure 2.** Schematic representation of the spatial-temporal study ranging from Autumn 2012 to Spring 2014 within Lagunita pond. Microbial mats from each site are shown in the right-hand panel. Characteristics of the sampling site and microbial mats

**Figure 3**. Profile scatter plot of each site (A, B and C), x-axis (dry conditions) y (humid conditions) microbial mats with the difference in mean proportion among microbial mats within each site along with the associated confidence interval of this effect size (2th and 98th percentile). Points on each side of the grey dashed y = x line are enriched in one of the two samples. Statistical hypothesis test is required to determine if the observed difference is large enough, to discount it being a sampling artifact safely, however, in dry conditions there is only one mat for each site therefore no p values are not indicated.

**Figure 4.** Hierarchical cluster dendogram generated on the basis of the standardized abundances at genus level (panel A) and PFAM abundance profile (panel B). The hierarchical cluster analysis with multiscale bootstrap=1000, using ward method and maximum distance. The color of the branches in panel A indicates two clusters obtained from pvclust with an AU (Approximately Unbiased) p-value

larger than 95%. In the case of panel B, these clusters were formed by samples S1 and S10 and S6 and S12, which are not colored in branches (See supplementary Figure S7). The color codes in the figure legend indicated in pink, green and blue samples from site A, B and C, respectively. The group indicates the cross-validation performed with clusterboot with a bootstrapping of 1000. At taxonomic level, samples are grouped within two highly stable clusters according to the Jaccard similarities 0.952 and 0.945 light pink and light blue, respectively. At metabolic level (panel B), the samples were grouped within four clusters with the most stable cluster (jaccard similarity 0.93), which is the one formed with samples S8, S11, S6 and s12. Finally, the dendograms cut several groups k=2,K=3, K=4 using cut-tree to cross validate the samples grouped into the previously identify clusters.

**Figure 5**. Network representation of the consensus networks displaying only to top 0.05% of total interactions. These interactions represent less than 50 from around 450 consensus families for three sites. In site A (A) are displaying 61/122,841 interactions, from which 17/46,200 are positive and 44/76,641 negative. In site B (B) 53/107548, interactions are shown, being 28/37, 170 positive and 25/70,378 negative. From microbial mats of site C (C), the consensus network is composed from 60/121,056, from which 16/57,487 are positive and 40/63,569 negatives. The size of a circle (node) is proportional to the abundance of the family across the microbial mats from each site. The thickness of a connection (edge) is proportional to the strength of the interaction. Families are colored according to the Phylum.

**Figure 6**. Biogeochemical cycling within microbial mats across space and time. A) MEBS captured in bits the overall O, N, Fe, S, and CH4 cycles. B) Heat map displaying the profile abundances of specific protein domains specific to biogeochemical cycles. The corresponding numbers to the sampling point of 1-4: November 2012, May 2013, October 2013 and May 2014. C) Metabolic completeness of sulfur cycle divided into 28 pathways described previously in De Anda et al (2017).

**Figure 7**. Distribution of the 3-three-node subgraphs or network motifs across 48 microbial mat networks built across the taxonomic levels: Phylum (P), Class (C), Order (O), and Family (F). The top panel represents the motifs sorted by their ID identifier described in Mfinder. The motifs with specific ecological terminology are ID6, ID12, ID36, and ID38. The latter motif is also known as Feed Forward Loop (FFL) in regulatory networks. The ID98 represents Feed Back Loop (FBL). To facilitate visual comparison, the abundance of each motif was normalized by its appearance across taxonomic levels. The relative abundance of each one indicates that a given motif is only found in that particular

taxonomic lave (i.e id 36 Site A in negative networks at class level). It is observed that particular motifs appear over-represented across the span of taxonomic levels when consensus networks are separated by type of interaction either positive or negative. (i.e. Motif 108 in negative networks or motif 36 in positive ones)

## Tables

Tables are found at the end of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author contributions

VDA, IZP, LEE, VS: conceived the project and worked on the manuscript. VDA, JB: sampling, field work, processing, DNA extraction and ecological index analysis. VDA, CPH, BMC: conducted bioinformatics and data analysis (trimming, filtering, taxonomic and metabolic profile). BCM, MHR: provided computing resources and analysis interpretation. MGL, MHR: performed and coded the network analyses. NGT: provided ecological interpretation of the network analysis. VDA, IZP and VS: conducted the microbiological, biogeochemical, and ecological interpretation of the network-based time-series approach. LEE, VS: sampling and fieldwork, provided expertise, logistics and resources to develop and supervised the project, as well as intellectual contributions to the work.

## Funding

This work constitutes a partial fulfillment requirement for the Ph.D. degree of Valerie De Anda at the graduate program Doctorado en Ciencias Biomédicas of the Universidad Nacional Autónoma de México who received fellowship 356 832 of Consejo Nacional de Ciencia y Tecnología (CONACYT). The authors acknowledge the founding of WWF-Alianza Carlos Slim, as well as support by a Sep Conacyt Project to VS and LEE 1101OL34. The paper was written during VDA research stay in BCM laboratory with the support of Beca Mixta Conacyt. The manuscript was written during a sabbatical leave of LEE and VSS in the University of Minnesota in Peter Tiffin and Michael Travisano laboratories, with support of the program PASPA- DGAPA, UNAM.

**Acknowledgments**

We would like to thank Manuel Rosas, Paola Vazquez and Gabriel Ponce for their valuable help during fieldwork and sampling collections. Emilio Morella and Silvia Barrientos for the technical support during DNA extractions. Ernesto Igartua Arregui and Carlos Cantalapiedra, and Paul'Obrien whose valuable comments greatly improve the manuscript. The first author acknowledges the laboratory of computational & structural biology and Aula Dei CSIC, for computing resources and support during the research stay.

**Data Availability Statement**

The data and scripts to reproduce all the figures for this study can be found in the following repository. https://valdeanda.github.io/Time_series_mats/

The software developed for the network analysis is found in https://valdeanda.github.io/Networks/

1027 **Tables**

| | Site A | | Site B | | Site C | |
|---|---|---|---|---|---|---|
| | Positive | Negative | Positive | Negative | Positive | Negative |
| Phylum | 0,400 | 0,600 | 0,341 | 0,659 | 0,511 | 0,489 |
| Class | 0,375 | 0,625 | 0,849 | 0,151 | 0,516 | 0,484 |
| Order | 0,386 | 0,614 | 0,366 | 0,634 | 0,504 | 0,496 |
| Family | 0,376 | 0,624 | 0,346 | 0,654 | 0,475 | 0,525 |
| Mean and std. | 0,384 ±0.011 | 0,616 ±0.012 | 0,476 ±0.249 | 0,525 ±0.249 | 0,502 ±0.018 | 0,499 ±0.018 |

**Table 1** Percentage of positive and negative interactions in the microbial mat networks from each site

1028

1029

| | Site A | | Site B | | Site C | |
|---|---|---|---|---|---|---|
| | Real | Random | Real | Random | Real | Random |
| Clustering Coefficient | 0.494 (± 0.003) | 0,4964 (±0,0039 | 0.495 (± 0.001) | 0,4960 (±0,0009) | 0.493 (±0.004) | 0,4939 (±0,0029 |
| Density | 0.927 (± 0.043) | 0,9269 (±0,0498) | 0.911 (±0.008) | 0,9106 (±0,0090) | 0.892 (±0.024) | 0,8920 (±0,0278 |
| Diameter | 2.000 (± 0.707) | 1,9625 (±0,0750) | 2.000 (±0.000) | 2.000 (±0.000) | 2.500 (±0.500) | 2.000 (±0.000) |
| Modularity | 0.003 (± 0.002) | 0,0010 (±0,0014 | 0.002 (±0.002) | 0,0005 (±0,0008) | 0.004 (±0.002) | 0,0015 (±0,0014) |
| Radius | 1.250 (± 0.433) | 1,1250 (±0,2500 | 1.250 (±0.433) | 1,0000 (±0,0000) | 1.500 (±0.500) | 1,0625 (±0,1250) |
| Mean degree | 159.391 (± 97.214) | 159,3912 (±112,2527 | 164.246 (±88.870) | 164,2461 (±102,6182) | 168.933 (±92.229) | 168,9333 (±106,4964) |
| Communities | 2.500 (± 1.500) | 1,7475 0,9531 | 2.250 (±1.090) | 1,5600 (±0,5931) | 3.000 (±1.581) | 2,0850 (±0,8296) |

1030 Table 2 Global network measures obtained from real (consensus) and random networks Std. of the
1031 values obtained across the span of taxonomic levels from Phylum to Family

1032

1033

1034

| | Negative Interactions | | | Positive interactions | | |
|---|---|---|---|---|---|---|
| | A | B | C | A | B | C |
| Clustering Coefficient | 0.44 (±0.01) | 0.44 (±0.01) | 0.41 (±0.01) | 0.28 (±0.03) | 0.17 (±0.02) | 0.25 (±0.02) |
| Random | 0.41 (±0.01) | 0.42 (±0.00) | 0.35 (±0.01) | 0.29 (±0.02) | 0.27 (±0.01) | 0.35 (±0.02) |
| Modularity* | 0.09 (±0.02) | 0.13 (±0.02) | 0.24 (±0.01) | 0.05 (±0.01) | 0.06 (±0.02) | 0.02 (±0.01) |
| Random | 0.025 (±0.005) | 0.023 (±0.005) | 0.037 (±0.011) | 0.052 (±0.013) | 0.057 (±0.017) | 0.037 (±0.007) |
| Diameter* | 2.25 (±0.50) | 2.50 (±0.58) | 2.50 (±0.58) | 2.25 (±0.50) | 2.50 (±0.58) | 2.50 (±0.58) |
| Random* | 2 (±0.0) | 2 (±0.0) | 2 (±0.0) | 2 (±0.0) | 2 (±0.0) | 2 (±0.0) |
| Max In Degree | 140.75 (±103.91) | 158.00 (±107.09) | 125.50 (±94.31) | 140.00 (±103.06) | 126.75 (±74.91) | 126.50 (±93.43) |
| Random | 116.14 (±79.75) | 124.00 (±75.25) | 104.48 (±65.50) | 76.38 (±48.84) | 73.73 (±41.36) | 100.64 (±55.95) |
| Max Out Degree | 135.00 (±104.39) | 154.50 (±102.75) | 116.50 (±77.99) | 131.25 (±99.43) | 135.50 (±84.20) | 115.50 (±78.12) |
| Random | 116.21 (±79.61) | 123.98 (±75.05) | 104.23 (±65.17) | 76.29 (±48.72) | 73.83 (±41.15) | 100.86 (±56.18) |
| Number of Strongly connected components (scc) | 6.25 (±4.43) | 8.25 (±1.71) | 11.50 (±4.80) | 6.25 (±4.43) | 8.50 (±1.91) | 12.00 (±5.23) |
| Random | 1 (±0.0) | 1 (±0.0) | 1 (±0.0) | 1 (±0.0) | 1 (±0.0) | 1 (±0.0) |
| Number of Hubs with Max In Degree | 3.50 (±2.38) | 2.25 (±0.96) | 6.50 (±7.33) | 1.50 (±1.00) | 2.25 (±2.50) | 1.25 (±0.50) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Random | 1.32 (±0.13) | 1.27 (±0.14) | 1.29 (±0.13) | 1.26 (±0.11) | 1.26 (±0.04) | 1.26 (±0.11) |
| Number of Hubs with Max Out Degree | 1.00 (±0.00) | 1.75 (±0.50) | 1.00 (±0.00) | 1.75 (±1.50) | 1.50 (±1.00) | 6.50 (±10.34) |
| Random | 1.29 (±0.13) | 1.22 (±0.06) | 1.23 (±0.12) | 1.20 (±0.07) | 1.28 (±0.11) | 1.21 (±0.09) |

1035 Table 3. Comparison of topological parameters between random and real networks. Global network
1036 measures are divided by positive and negative interactions Std. (±) represent the values obtained
1037 across the span of taxonomic levels from Phylum to Family

1038

1039

**Table 4. Strongly interconnected taxa or hubs identify at family levels within TSEns of microbial mats**

| Site | Type of interaction | Max In-Degree (affected by many nodes) | Max out degree (affects many nodes) |
|---|---|---|---|
| **A** | *Consensus* | (337)<br>1. Nannocystaceae (PG)(RB)<br>2. Desulfarculaceae(C) (PG) (RB): *Desulfarculus*<br>3. Bacteroidaceae(C)(RB): *Bacteroides*<br>4. Caulobacteraceae(C)(RB): *unclassified Caulobacteraceae*<br>5. Bradyrhizobiaceae(C): Bosea, unclassified Bradyrhizobiaceae, Bradyrhizobium<br>6. unclassified Oscillatoriophycideae (UC)(PG)(RB)<br>7. Rhodobacterales (UC)(PG)(RB)<br>8. Actinobacteria phylum 201174(UC)(PG)(RB) | (345)<br>1. Polyangiaceae (PG)(RB) |
| | *Positives* | (275)<br>1. unclassified Chloroflexi (RB) | (263)<br>1. unclassified Armatimonadetes (UC)(PG)(RB) |
| | *Negatives* | (278)<br>1. unclassified Rhodobacterales (UC)(RB) | (275)<br>1. Pelobacteraceae(C)(RB)(PB) :*Pelobacter* |
| **B** | *Consensus* | (318)<br>1. unclassified Xanthomonadales (UC)(RB)(PG)<br>2.Beijenrincliaceae<br>3.Burkholderiaceae(C): *Burkholderia, unclassified Burkholderiaceae, Burkholderiales Genera incertae sedis unclassified, unclassified Burkholderiales*<br>4. Phyllobacteriaceae (C)(PG): *unclassified Phyllobacteriaceae*<br>5. Gemmantimonadaceae | (325)<br>1. Thermoactinomycetaceae(C)(RB)(PG): Thermoactinomyces<br>2. Actinomycetaceae(C)(RB)(PG): Actinomyces<br>3. Alcaligenaceae(C) (RB): *unclassified* Alcaligenaceae<br>4. Gemmantimonadaceae |
| | *Positives* | (226)<br>1. Staphylococcaceae(RB) | (246)<br>1. Fusobacteriaceae (RB) |
| | *Negatives* | (298)<br>1. Phyllobacteriaceae(C): *unclassified Phyllobacteriaceae* | (289)<br>1. Cystobacterineae (UC)(RB)(PG)<br>2. Cloacimonetes (UC)(RB) |
| **C** | *Consensus* | (335)<br>1. Unclassified Verrucomicrobia (UC)(RB)<br>2. Methylobacteriaceae(C) (PG):*( unclassified Methylobacteriaceae, Methylobacterium* | (343)<br>1.Chromobacteriaceae(C)(RB)(PG): *unclassified Chromobacteriaceae*<br>2.Mycobacteriaceae(C)(RB):*Mycobacterium*<br>3.Symbiobacteriaceae(RB) |
| | *Positives* | (252)<br>1. unclassified Archaea (RB) | (223)<br>1. Micrococcales (UC)(RB)(PG) |
| | *Negatives* | (251)<br>1. Rhizobiales (UC)<br>2. Alteromonadaceae<br>3. Methylococcaceae(C)(RB): *unclassified Methylococcaceae*<br>4. Bacillales (UC)(RB)<br>5. Corynebacteriales (UC)(RB)<br>6.Beijerinckiaceae(RB) | (222)<br>1. Bacteroidetes (UC)(RB) |
| **Global (A+B+C)** | *Consensus* | (391)<br>1.Desulfarculaceae(C) (RB): *Desulfarculus*<br>2 Pelobacteraceae(C)(RB): *Pelobacter*<br>3 Gemmatimonadaceae | (393)<br>1. Bacteroidales (UC)(RB)<br>2. Microgenomates group (UC) (RB)<br>3. Nocardiaceae(C) (RB): Nocardia<br>4. Nitrospiraceae (RB)<br>5. Prolixibacteraceae (RB) |
| | *Positive* | (349)<br>1. Moraxellaceae (RB) | (352)<br>1. unclassified Eukaryota (UC) |
| | *Negatives* | (367)<br>1. Phyllobacteriaceae(C): *unclassified Phyllobacteriaceae* | (362)<br>1. Desulfobacteraceae(C): Desulfatibacillum, Desulfatiglans, Desulfatirhabdium, Desulfatitalea, Desulfobacter, *unclassified* Desulfobacteraceae, *unclassified* Desulfobacterales |

C = Core microbial mat (presence in all samples regardless of environmental conditions and abundances)

UC=Unclassified sequences belonging to the core microbial mat

PG= Taxa with known members delivering public goods (i.e metabolites, enzymes, vitamins)

RB =Rare Biosphere >1% (0.01) abundances

## References


Allesina, S., Grilli, J., Barabás, G., Tang, S., Aljadeff, J., and Maritan, A. (2015). Predicting the stability of large structured food webs. *Nat. Commun.* 6, 7842. doi:10.1038/ncomms8842.

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461. doi:10.1038/nrg2102.

Baiser, B., Elhesha, R., and Kahveci, T. (2016). Motifs in the assembly of food web networks. *Oikos* 125, 480–491. doi:10.1111/oik.02532.

Baldassano, S. N., and Bassett, D. S. (2016). Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Sci. Rep.* 6, 1–14. doi:10.1038/srep26087.

Balskus, E. P., and Walsh, C. T. (2011). The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. 329, 1653–1656. doi:10.1126/science.1193637.The.

Bascompte, J., and Stouffer, D. B. (2009). The assembly and disassembly of ecological networks. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1781–1787. doi:10.1098/rstb.2008.0226.

Bissett, A., Brown, M. V., Siciliano, S. D., and Thrall, P. H. (2013). Microbial community responses to anthropogenically induced environmental change: Towards a systems approach. *Ecol. Lett.* 16, 128–139. doi:10.1111/ele.12109.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.*, 1–12. doi:10.1088/1742-5468/2008/10/P10008.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.

Bolhuis, H., Cretoiu, M. S., and Stal, L. J. (2014). Molecular Ecology of Microbial Mats. *FEMS Microbiol. Ecol.*, 1–16. doi:10.1111/1574-6941.12408.

Bonilla-Rosso, G., Peimbert, M., Alcaraz, L. D., Hernández, I., Eguiarte, L. E., Olmedo-Alvarez, G., et al. (2012). Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments. *Astrobiology* 12, 659–673.

doi:10.1089/ast.2011.0724.

Borrelli, J. J., Allesina, S., Amarasekare, P., Arditi, R., Chase, I., Damuth, J., et al. (2015). Selection on stability across ecological scales. *Trends Ecol. Evol.* 30, 417–425. doi:10.1016/j.tree.2015.05.001.

Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., Dinsdale, E., et al. (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Environ. Microbiol.* 11, 16–34. doi:10.1111/j.1462-2920.2008.01725.x.

Brinkhoff, T., Fischer, D., Vollmers, J., Voget, S., Beardsley, C., Thole, S., et al. (2012). Biogeography and phylogenetic diversity of a cluster of exclusively marine myxobacteria. *ISME J.* 6, 1260–1272. doi:10.1038/ismej.2011.190.

Bullock, H. a, Reisch, C. R., Burns, A. S., Moran, M. A., and Whitman, W. B. (2014). Regulatory and functional diversity of methylmercaptopropionate coenzyme A ligases from the dimethylsulfoniopropionate demethylation pathway in Ruegeria pomeroyi DSS-3 and other proteobacteria. *J. Bacteriol.* 196, 1275–85. doi:10.1128/JB.00026-14.

Cadotte, M. W., Arnillas, C. A., Livingstone, S. W., and Yasui, S. L. E. (2015). Predicting communities from functional traits. *Trends Ecol. Evol.* 30, 510–511. doi:10.1016/j.tree.2015.07.001.

Chennu, A., Grinham, A., Polerecky, L., de Beer, D., and Al-Najjar, M. A. A. (2015). Rapid reactivation of cyanobacterial photosynthesis and migration upon rehydration of desiccated marine microbial mats. *Front. Microbiol.* 6, 1–9. doi:10.3389/fmicb.2015.01472.

Conti, E., Franks, N. P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a super-family of adenylate-forming enzymes. *Structure* 4, 287–298. doi:10.1016/S0969-2126(96)00033-0.

Contreras-Moreira, B., and Vinuesa, P. (2013). GET{_}HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi:10.1128/AEM.02411-13.

Coyte, K. Z., Schluter, J., and Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Sci. (New York, NY)* 350, 663–666. doi:10.1126/science.aad2602.

Cui, H., Yu, X., Wang, Y., Cui, Y., Li, X., Liu, Z., et al. (2013). Evolutionary origins, molecular cloning and expression of carotenoid hydroxylases in eukaryotic photosynthetic algae. *BMC Genomics* 14, 457–476. doi:10.1186/1471-2164-14-457.

Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A., and Johnston, A. W. B. (2008). Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas

dimethylsulfide in several marine α-proteobacteria and Rhodobacter sphaeroides. *Environ. Microbiol.* 10, 757–767. doi:10.1111/j.1462-2920.2007.01499.x.

Dang, H., Li, T., Chen, M., and Huang, G. (2008). Cross-Ocean Distribution of Rhodobacterales Bacteria as Primary Surface Colonizers in Temperate Coastal Marine Waters. *Appl. Environ. Microbiol.* 74, 52–60. doi:10.1128/AEM.01400-07.

de-Leon, S. B.-T., and Davidson, E. H. (2007). Gene Regulation: Gene Control Network in Development. *Annu. Rev. Biophys. Biomol. Struct.* 36, 191–212. doi:10.1146/annurev.biophys.35.040405.102002.

De Anda, V., Zapata-Peñasco, I., Poot-Hernandez, A. C., Eguiarte, L. E., Contreras-Moreira, B., and Souza, V. (2017). MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle Authors. *Gigascience* 6, 1–17. doi:10.1093/gigascience/gix096.

De Anda, V., Zapata-Peñasco, I., and Souza, V. (2018). "Towards a comprehensive understanding of environmental perturbations in microbial mats from the Cuatro Ciene-gas Basin by network modeling," in *Ecosystem ecology and geochemistry of Cuatro Cienegas: How to survive in an extremely oligotrophic site*, ed. S. V Elser J García-Oliva F (Springer), In Press.

Debebe, T., Biagi, E., Soverini, M., Holtze, S., Hildebrandt, T. B., Birkemeyer, C., et al. (2017). Unraveling the gut microbiome of the long-lived naked mole-rat. *Sci. Rep.* 7, 1–9. doi:10.1038/s41598-017-10287-0.

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., et al. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nat. Commun.* 7, 1–8. doi:10.1038/ncomms10541.

Delmas, E., Besson, M., Brice, M.-H., and Burkle, L. A. (2017). Analyzing ecological networks of species in- teractions. 1–20. doi:10.1101/112540.

Delmas, E., Besson, M., Brice, M.-H., Burkle, L., Riva, G. V. D., Fortin, M.-J., et al. (2018). Analyzing ecological networks of species interactions. *bioRxiv*, 112540. doi:10.1101/112540.

Deng, Y., Zhang, P., Qin, Y., Tu, Q., Yang, Y., He, Z., et al. (2016). Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. *Environ. Microbiol.* 18, 205–218. doi:10.1111/1462-2920.12981.

Duffy, J. E., and Stachowicz, J. J. (2006). Why biodiversity is important to oceanography: Potential roles of genetic, species, and trophic diversity in pelagic ecosystem processes. *Mar. Ecol. Prog. Ser.* 311, 179–189. doi:10.3354/meps311179.

Elifantz, H., Horn, G., Ayon, M., Cohen, Y., and Minz, D. (2013). Rhodobacteraceae are the key

members of the microbial community of the initial biofilm formed in Eastern Mediterranean coastal seawater. *FEMS Microbiol. Ecol.* 85, 348–57. doi:10.1111/1574-6941.12122.

Eng, A., and Borenstain, E. (2018). Taxa-function robustness in microbial communities. *Microbiome* 6, 1–19. doi:10.1186/s40168-018-0425-4.

Eren, a M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., and Sogin, M. L. (2014). Minimum entropy decomposition : Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *Isme J* 9, 968–979. doi:10.1038/ismej.2014.195.

Erici, S., Pawlikb, G., Lohmeyerb, E., Koch, H.-G., and Daldal, F. (2012). Biogenesis of cbb3-type cytochrome c oxidase in Rhodobacter capsulatus. *Biochim. Biophys. Acta* 8, 2115–2126. doi:10.1021/ct300008d.Comparing.

Espinosa-Asuar, L., Escalante, A. E., Gasca-Pineda, J., Blaz, J., Peña, L., Eguiarte, L. E., et al. (2015). Aquatic bacterial assemblage structure in Pozas Azules, Cuatro Cienegas Basin, Mexico: Deterministic vs. stochastis processes. *Int. Microbiol.* 18, 105–115. doi:10.2436/20.1501.01.240.

Faust, K., and Raes, J. (2012a). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi:10.1038/nrmicro2832.

Faust, K., and Raes, J. (2012b). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi:10.1038/nrmicro2832.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S. G., González, J. M., et al. (2013). Ecology of marine bacteroidetes: A comparative genomics approach. *ISME J.* 7, 1026–1037. doi:10.1038/ismej.2012.169.

Fortunato, S., and Hric, D. (2016). Community detection in networks: A user guide. *Phys. Rep.* 659, 1–44. doi:https://doi.org/10.1016/j.physrep.2016.09.002.

Foster, K. R., and Bell, T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* 22, 1845–1850. doi:10.1016/j.cub.2012.08.005.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Ruppin, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.* 38, 3857–3868. doi:10.1093/nar/gkq118.

Frikha Dammak, D., Zarai, Z., Najah, S., Abdennabi, R., Belbahri, L., Rateb, M. E., et al. (2017). Antagonistic properties of some halophilic thermoactinomycetes isolated from superficial sediment of a solar saltern and production of cyclic antimicrobial peptides by the novel isolate paludifilum halophilum. *Biomed Res. Int.* 2017. doi:10.1155/2017/1205258.

Fuhrman, J. a (2009). Microbial community structure and its functional implications. *Nature* 459,

193–9. doi:10.1038/nature08058.

Grilli, J., Rogers, T., and Allesina, S. (2016). Modularity and stability in ecological communities. *Nat. Commun.* 7, 12031. doi:10.1038/ncomms12031.

Grimm, F., Franz, B., and Dahl, C. (2011). Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium allochromatium vinosum. *Front. Microbiol.* 2, 51. doi:10.3389/fmicb.2011.00051.

Guerrero, R., Piqueras, M., and Berlanga, M. (2002). Microbial mats and the search for minimal ecosystems. *Int. Microbiol.* 5, 177–188. doi:10.1007/s10123-002-0094-8.

Guy, L., and Ettema, T. J. G. (2011). The archaeal "TACK" superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587. doi:10.1016/j.tim.2011.09.002.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proc. 7th Python Sci. Conf. (SciPy 2008)*, 11–15.

Harris, J. (2012). Soil Microbial Communities and Restoration Ecology: Facilitators or Followers? *Science (80-. ).* 573, 573–574. doi:10.1126/science.1172975.

Harris, J. K., Caporaso, J. G., Walker, J. J., Spear, J. R., Gold, N. J., Robertson, C. E., et al. (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J.* 7, 50–60. doi:10.1038/ismej.2012.79.

Higashioka, Y., Kojima, H., Watanabe, M., and Fukui, M. (2013). Desulfatitalea tepidiphila gen. nov., sp. nov., a sulfate-reducing bacterium isolated from tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 63, 761–765. doi:10.1099/ijs.0.043356-0.

Huber, H., Burggraf, S., Mayer, T., Wyschkony, I., Rachel, R., and Stetter, K. O. (2000). Igniococcus gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species Igniococcus islandicus sp. nov. and Igniococcus pacificus sp. nov. *Int. J. Syst. Evol. Microbiol.* 50, 2093–2100.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi:10.1186/1471-2105-11-119.

Jeffries, T. C., Seymour, J. R., Newton, K., Smith, R. J., Seuront, L., and Mitchell, J. G. (2012). Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient. *Biogeosciences* 9, 815–825. doi:10.5194/bg-9-815-2012.

Jin, G., Zhang, S., Zhang, X. S., and Chen, L. (2007). Hubs with network motifs organize modularity dynamically in the protein-protein interaction network of yeast. *PLoS One* 2. doi:10.1371/journal.pone.0001207.

1206 Jonkers, H. M., Bruin, S., and Gemerden, H. (1998). Turnover of dimethylsulfoniopropionate
1207 (DMSP) by the purple sulfur bacterium Thiocapsa roseopersicina M11: ecological implications.
1208 *FEMS Microbiol. Ecol.* 27, 281–290. doi:10.1111/j.1574-6941.1998.tb00544.x.
1209 Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., et al. (2017). Where less
1210 may be more: How the rare biosphere pulls ecosystems strings. *ISME J.* 11, 853–862.
1211 doi:10.1038/ismej.2016.174.
1212 Jungblut, A. D., Vincent, W. F., and Lovejoy, C. (2012). Eukaryotes in Arctic and Antarctic
1213 cyanobacterial mats. *FEMS Microbiol. Ecol.* 82, 416–428. doi:10.1111/j.1574-
1214 6941.2012.01418.x.
1215 Katoch, M., Mazmouz, R., Chau, R., Pearson, L. A., Pickford, R., and Neilan, B. A. (2016).
1216 Heterologous production of cyanobacterial mycosporine-like amino acids mycosporine-
1217 ornithine and mycosporine-lysine in E. coli. *Appl. Environ. Microbiol.* 82, AEM.01632-16.
1218 doi:10.1128/AEM.01632-16.
1219 Konopka, A. E., Lindemann, S., and Fredrickson, J. K. (2015). Dynamics in microbial communities:
1220 unraveling mechanisms to identify principles. *ISME J.* 9, 1488–1495.
1221 doi:10.1038/ismej.2014.251.
1222 Kuever, J. (2014). "The Family Desulfarculaceae," in *The Prokaryotes: Deltaproteobacteria and*
1223 *Epsilonproteobacteria*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F.
1224 Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 41–44. doi:10.1007/978-3-642-
1225 39044-9_270.
1226 Lake, J. A., Henderson, E., Oakes, M., and Clark, M. W. (1984). Eocytes: a new ribosome structure
1227 indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 81,
1228 3786–3790. doi:10.1073/pnas.81.12.3786.
1229 Lea-Smith, D. J., Bombelli, P., Vasudevan, R., and Howe, C. J. (2016). Photosynthetic, respiratory
1230 and extracellular electron transport pathways in cyanobacteria. *Biochim. Biophys. Acta -*
1231 *Bioenerg.* 1857, 247–255. doi:10.1016/j.bbabio.2015.10.007.
1232 Lee, Z. M., Steger, L., Corman, J. R., Neveu, M., Poret-Peterson, A. T., Souza, V., et al. (2015).
1233 Response of a stoichiometrically imbalanced ecosystem to manipulation of nutrient supplies and
1234 ratios. *PLoS One* 10, 1–17. doi:10.1371/journal.pone.0123949.
1235 Lee, Z., Poret-Peterson, A. T., Siefert, J. L., Kaul, D., Moustafa, A., Allen, A. E., et al. (2017).
1236 Nutrient stoichiometry Shapes Microbial Community Structure in an Evaporitic Shallow Pond.
1237 *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.00949.
1238 Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2014). MEGAHIT: An ultra-fast single-

node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi:10.1093/bioinformatics/btv033.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11980–11985. doi:10.1073/pnas.2133841100.

Marquardt, J., and Palinska, K. A. (2007). Genotypic and phenotypic diversity of cyanobacteria assigned to the genus Phormidium (Oscillatoriales) from different habitats and geographical sites. *Arch. Microbiol.* 187, 397–413. doi:10.1007/s00203-006-0204-7.

Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D., and Vandenkoornhuyse, P. (2016). Beyond the Black Queen Hypothesis. *ISME J.* 10, 2085–2091. doi:10.1038/ismej.2016.22.

Mccully, A. L., Lasarre, B., and Mckinlay, J. B. (2017). Recipient-Biased Competition for an Intracellularly Generated Cross-Fed Nutrient Is Required for Coexistence of Microbial Mutualists. *MBio* 8, 1–13.

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8. doi:10.1371/journal.pone.0061217.

Mei, Y., Liu, H., Zhang, S., Yang, M., Hu, C., Zhang, J., et al. (2017). Effects of salinity on the cellular physiological responses of Natrinema sp . J7-2. 1–13.

Milo, R. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science (80-. ).* 298, 824–827. doi:10.1126/science.298.5594.824.

Minot, S. S., Krumm, N., and Greenfield, N. B. (2015). One Codex: A Sensitive and Accurate Data Platform for Genomic Microbial Identification. *bioRxiv*, 027607. doi:10.1101/027607.

Monard, C., Gantner, S., Bertilsson, S., Hallin, S., and Stenlid, J. (2016). Habitat generalists and specialists in microbial communities across a terrestrial-freshwater gradient. *Sci. Rep.* 6, 1–10. doi:10.1038/srep37719.

Montoya, J. M., Pimm, S. L., and Solé, R. V. (2006). Ecological networks and their fragility. *Nature* 442, 259–264. doi:10.1038/nature04927.

Moreno-Letelier, A., Souza, V., Travisano, M., Alcaraz, L. D., Olmedo, G., and Eguiarte, L. E. (2018). The lost world of Cuatro Cienegas Basin, a relictual bacterial niche in a desert oasis. *bioRxiv*. doi:10.1101/311381.

Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen Hypothesis : Evolution of Dependencies through Adaptative Gene Loss. *MBio* 3, 1–7. doi:10.1128/mBio.00036-12.Copyright.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci.*

*U. S. A.* 103, 8577–82. doi:10.1073/pnas.0601602103.

Newman, M. E. J., and Girvan, M. (2003). Finding and evaluating community structure in networks. doi:10.1103/physreve.69.026113.

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., Bertilsson, S., Stokstad, E., et al. (2012). Resistance, resilience, and redundancy in microbial communities. *PLoS One* 7, 1–7. doi:10.1146/annurev.micro.58.030603.123654.

Nutman, A. P., Bennett, V. C., Friend, C. R. L., Van Kranendonk, M. J., and Chivas, A. R. (2016). Rapid emergence of life shown by discovery of 3,700-million-year-old microbial structures. *Nature* 537, 535–538. doi:10.1038/nature19355.

Oliver, T. H., Isaac, N. J. B., August, T. A., Woodcock, B. A., Roy, D. B., and Bullock, J. M. (2015). Declining resilience of ecosystem functions under biodiversity loss. *Nat. Commun.* 6, 10122. doi:10.1038/ncomms10122.

Pajares, S., Escalante, A. E., Noguez, A. M., García-Oliva, F., Martínez-Piedragil, C., Cram, S. S., et al. (2016). Spatial heterogeneity of physicochemical properties explains differences in microbial composition in arid soils from Cuatro Cienegas, Mexico. *PeerJ* 4, e2459. doi:10.7717/peerj.2459.

Pajares, S., Souza, V., and Eguiarte, L. E. (2015). Multivariate and phylogenetic analyses assessing the response of bacterial mat communities from an ancient oligotrophic aquatic ecosystem to different scenarios of long-term environmental disturbance. *PLoS One* 10, 1–18. doi:10.1371/journal.pone.0119741.

Parks, D. H., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi:10.1093/bioinformatics/btu494.

Peimbert, M., Alcaraz, L. D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L., et al. (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, 648–658. doi:10.1089/ast.2011.0694.

Peura, S., Bertilsson, S., Jones, R. I., and Eiler, A. (2015). Resistant microbial cooccurrence patterns inferred by network topology. *Appl. Environ. Microbiol.* 81, 2090–2097. doi:10.1128/AEM.03660-14.

Poisot, T. (2013). An a posteriori measure of network modularity. *F1000Research*, 1–14. doi:10.12688/f1000research.2-130.v1.

Pontarp, M., Canbäck, B., Tunlid, A., and Lundberg, P. (2012). Phylogenetic Analysis Suggests That

1305 Habitat Filtering Is Structuring Marine Bacterial Communities Across the Globe. *Microb. Ecol.*
1306 64, 8–17. doi:10.1007/s00248-011-0005-7.
1307 Preisner, E. C., Fichot, E. B., and Norman, R. S. (2016). Microbial mat compositional and functional
1308 sensitivity to environmental disturbance. *Front. Microbiol.* 7, 1–16.
1309 doi:10.3389/fmicb.2016.01632.
1310 Prieto-Barajas, C. M., Valencia-Cantero, E., and Santoyo, G. (2017). Microbial mat ecosystems:
1311 Structure types, functional diversity, and biotechnological application. *Electron. J. Biotechnol.*
1312 31, 48–56. doi:10.1016/j.ejbt.2017.11.001.
1313 Prill, R. J., Iglesias, P. A., and Levchenko, A. (2005). Dynamic properties of network motifs
1314 contribute to biological network organization. *PLoS Biol.* 3, 1881–1892.
1315 doi:10.1371/journal.pbio.0030343.
1316 Proulx, S. R., Promislow, D. E. L., and Phillips, P. C. (2005). Network thinking in ecology and
1317 evolution. *Trends Ecol. Evol.* 20, 345–353. doi:10.1016/j.tree.2005.04.004.
1318 Purdy, K. J. (2005). Nucleic acid recovery from complex environmental samples. *Methods Enzymol.*
1319 397, 271–292. doi:10.1016/S0076-6879(05)97016-X.
1320 R Development Core Team R: A Language and Environment for Statistical Computing.
1321 Shaw, G. T.-W., Pao, Y.-Y., and Wang, D. (2016). MetaMIS: a metagenomic microbial interaction
1322 simulator based on microbial community profiles. *BMC Bioinformatics* 17, 488.
1323 doi:10.1186/s12859-016-1359-0.
1324 Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional
1325 regulation network of Escherichia coli. *Nat. Genet.* 31, 64–68. doi:10.1038/ng881.
1326 Shoemaker, W. R., and Lennon, J. T. (2018). Evolution with a seed bank: The population genetic
1327 consequences of microbial dormancy. *Evol. Appl.* 11, 60–75. doi:10.1111/eva.12557.
1328 Sonnenschein, N., Marr, C., and Hütt, M.-T. (2012). A Topological Characterization of Medium-
1329 Dependent Essential Metabolic Reactions. *Metabolites* 2, 632–647.
1330 doi:10.3390/metabo2030632.
1331 Sorokin, D. Y., Tourova, T. P., Lysenko, A. M., and Muyzer, G. (2006). Diversity of culturable
1332 halophilic sulfur-oxidizing bacteria in hypersaline habitats. *Microbiology* 152, 3013–3023.
1333 doi:10.1099/mic.0.29106-0.
1334 Souza, V., Eguiarte, L. E., Travisano, M., Elser, J. J., Rooks, C., and Siefert, J. L. (2012). Travel, sex,
1335 and food: what's speciation got to do with it? *Astrobiology* 12, 634–640.
1336 doi:10.1089/ast.2011.0768.
1337 Souza, V., Espinosa-Asuar, L., Escalante, A. E., Eguiarte, L. E., Farmer, J., Forney, L., et al. (2006).

An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6565–6570. doi:10.1073/pnas.0601434103.

Souza, V., Falcón, L. I., Elser, J. J., and Eguiarte, L. E. (2007). Protecting a window into the ancient Earth: Towards a Precambrian Park at Cuatro Cienegas, Mexico. *Citizen's Page, Evol. Ecol. Res.*, 1–22. Available at: http://www.evolutionary-ecology.com/citizen/citizen.html.

Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* 5, 1414–1425. doi:10.1038/ismej.2011.24.

Stouffer, D. B., Camacho, J., Jiang, W., and Nunes Amaral, L. A. (2007). Evidence for the existence of a robust pattern of prey selection in food webs. *Proc. R. Soc. B Biol. Sci.* 274, 1931–1940. doi:10.1098/rspb.2007.0571.

Sun, M. Y., Dafforn, K. A., Johnston, E. L., and Brown, M. V. (2013). Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. *Environ. Microbiol.* 15, 2517–2531. doi:10.1111/1462-2920.12133.

Tapia-Torres, Y., and Ólmedo-Álvarez, G. (2018). Life on Phosphite: A Metagenomics Tale. *Trends Microbiol.* 23, 170–172. doi:10.1016/j.tim.2018.01.002.

Thébault, E., and Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science (80-. ).* 329, 853–856. doi:10.1126/science.1188321.

Tran, N. H., Choi, K. P., and Zhang, L. (2013). Counting motifs in the human interactome. *Nat. Commun.* 4, 1–8. doi:10.1038/ncomms3241.

Tzun-Wen Shaw, G., Liu, A. C., Weng, C. Y., Chou, C. Y., and Wang, D. (2017). Inferring microbial interactions in thermophilic and mesophilic anaerobic digestion of HOG waste. *PLoS One* 12, 1–22. doi:10.1371/journal.pone.0181395.

van Gemerden, H. (1993). Microbial mats: A joint venture. *Mar. Geol.* 113, 3–25. doi:10.1016/0025-3227(93)90146-M.

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., and Corbeil, J. (2012). Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. *Appl. Environ. Microbiol.* 78, 549–559. doi:10.1128/AEM.06354-11.

Warden, J. G., Casaburi, G., Omelon, C. R., Bennett, P. C., Breecker, D. O., Foster, J. S., et al. (2016). Characterization of Microbial Mat Microbiomes in the Modern Thrombolite Ecosystem of Lake Clifton , Western Australia Using Shotgun Metagenomics. *Front. Microbiol.* 7, 1–14. doi:10.3389/fmicb.2016.01064.

Weng, F. C. H., Shaw, G. T. W., Weng, C. Y., Yang, Y. J., and Wang, D. (2017). Inferring microbial

1371 interactions in the gut of the Hong Kong whipping frog (Polypedates megacephalus) and a
1372 validation using probiotics. *Front. Microbiol.* 8, 1–11. doi:10.3389/fmicb.2017.00525.
1373 Wolaver, B. D., Crossey, L. J., Karlstrom, K. E., Banner, J. L., Cardenas, M. B., Ojeda, C. G., et al.
1374 (2012). Identifying origins of and pathways for spring waters in a semiarid basin using He, Sr,
1375 and C isotopes: Cuatrocienegas Basin, Mexico. *Geosphere* 9, 113–125.
1376 doi:10.1130/GES00849.1.
1377 Wong, H. L., Visscher, P. T., White, R. A., Smith, D. L., Patterson, M. M., and Burns, B. P. (2017).
1378 Dynamics of archaea at fine spatial scales in Shark Bay mat microbiomes. *Sci. Rep.* 7, 1–12.
1379 doi:10.1038/srep46160.
1380 Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015).
1381 Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc.*
1382 *Natl. Acad. Sci. U. S. A.* 112. doi:10.1073/pnas.1421834112.
1383 Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., and Zhi, X. (2010). Functional molecular ecological
1384 networks. *MBio* 1, e00169-10. doi:10.1128/mBio.00169-10.Editor.
1385 Zomorrodi, A. R., and Segrè, D. (2017). Genome-driven evolutionary game theory helps understand
1386 the rise of metabolic interdependencies in microbial communities. *Nat. Commun.* 8, 1–11.
1387 doi:10.1038/s41467-017-01407-5.
1388

Figure 1.TIF



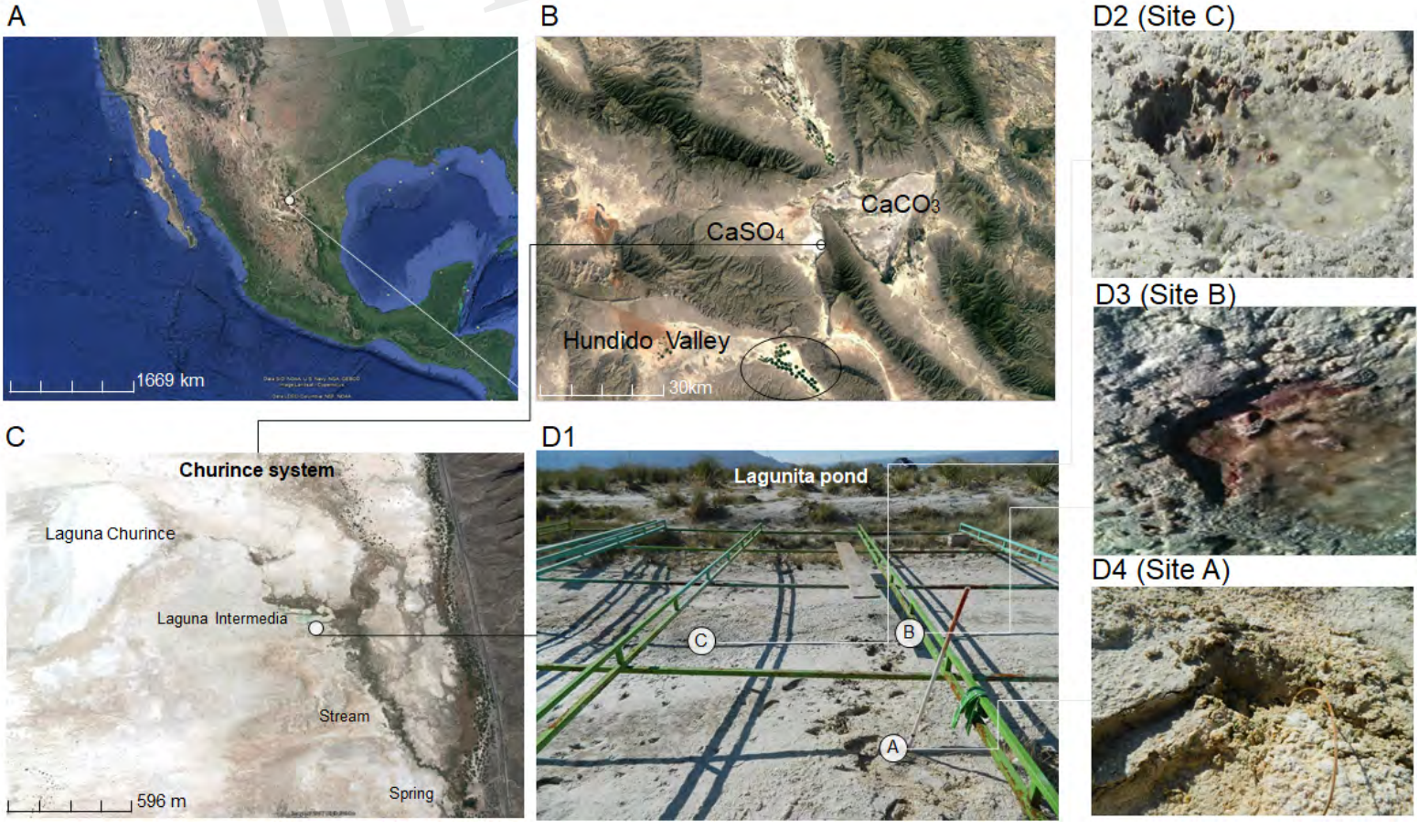


Figure 2.TIF

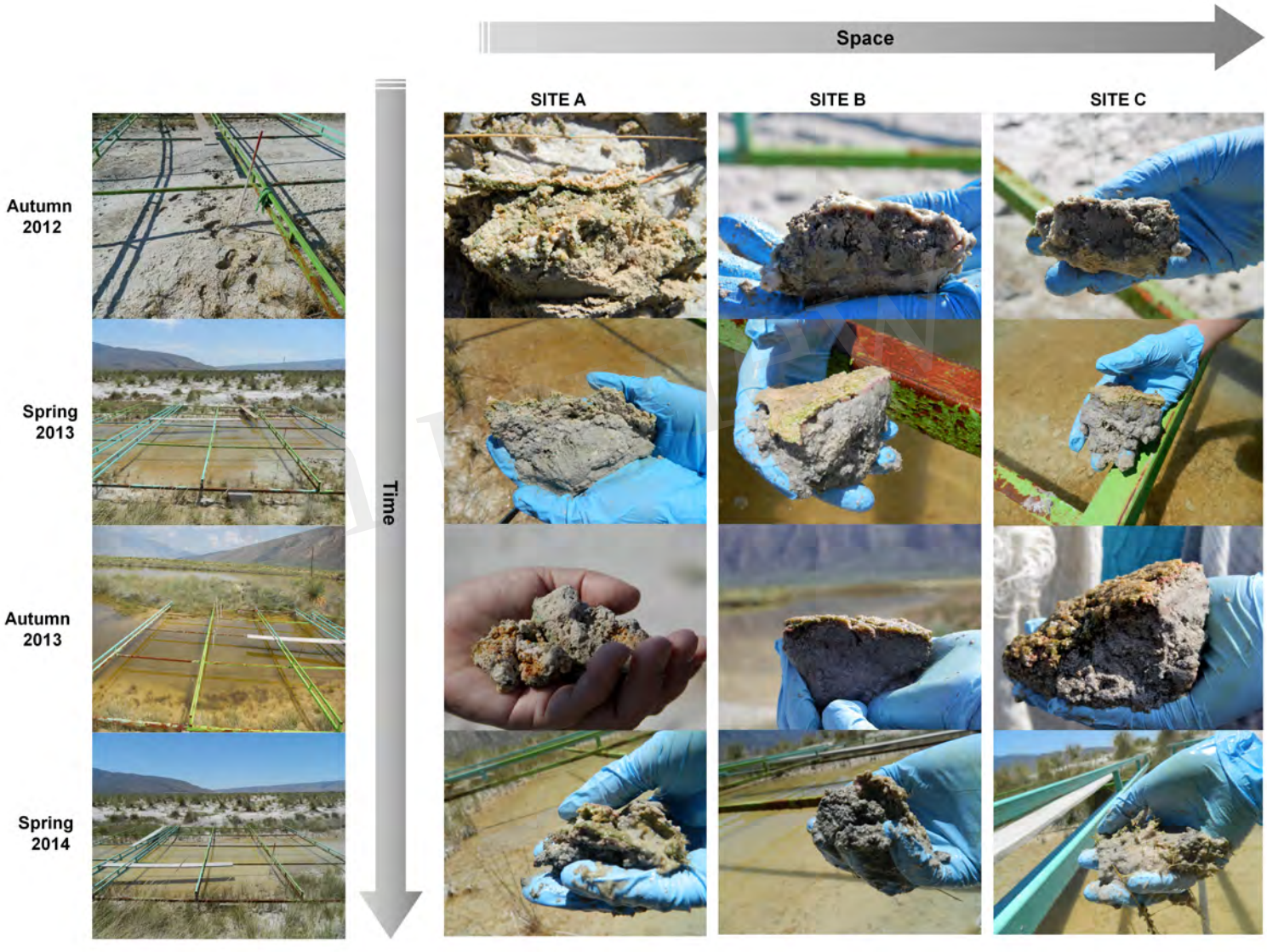
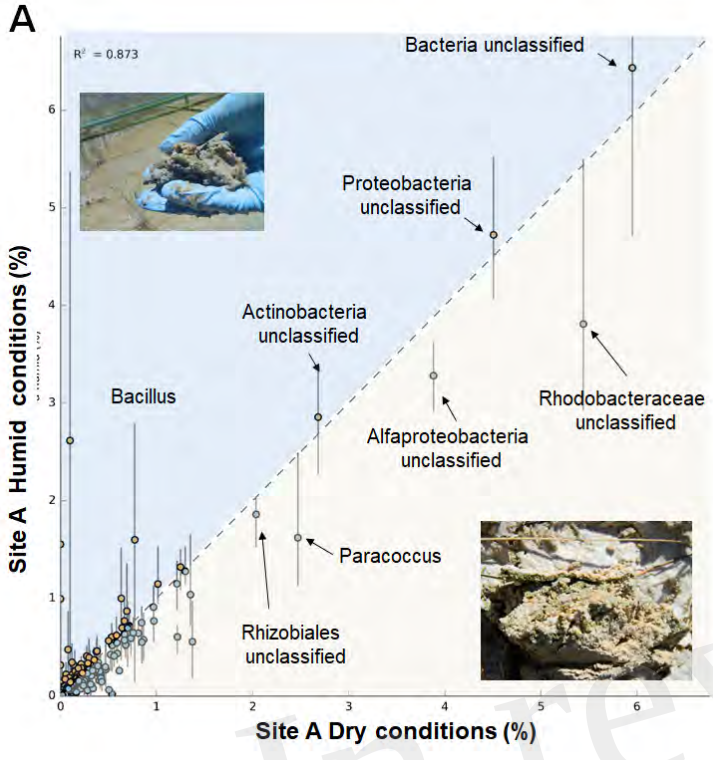
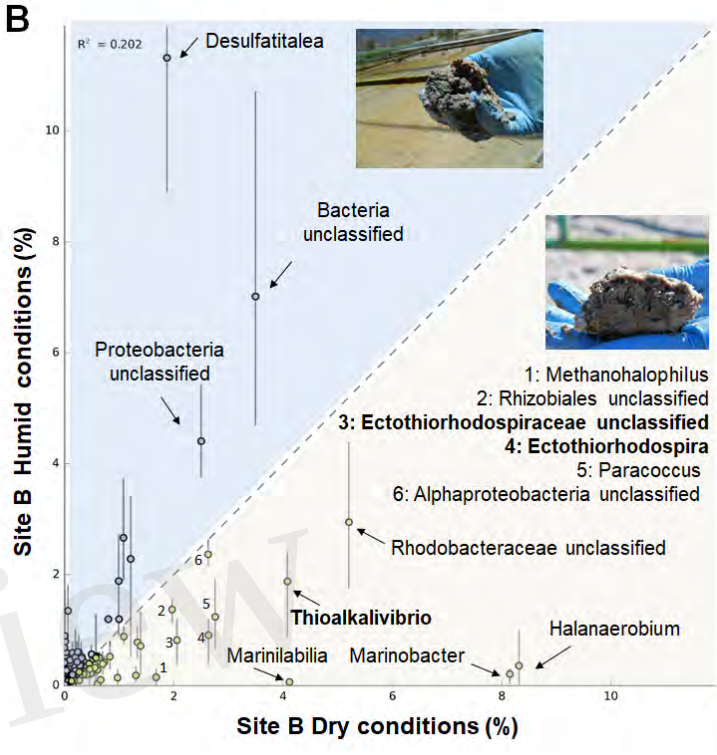

Space
SITE A
SITE B
SITE C
Time
Autumn 2012
Spring 2013
Autumn 2013
Spring 2014

Figure 3.TIF



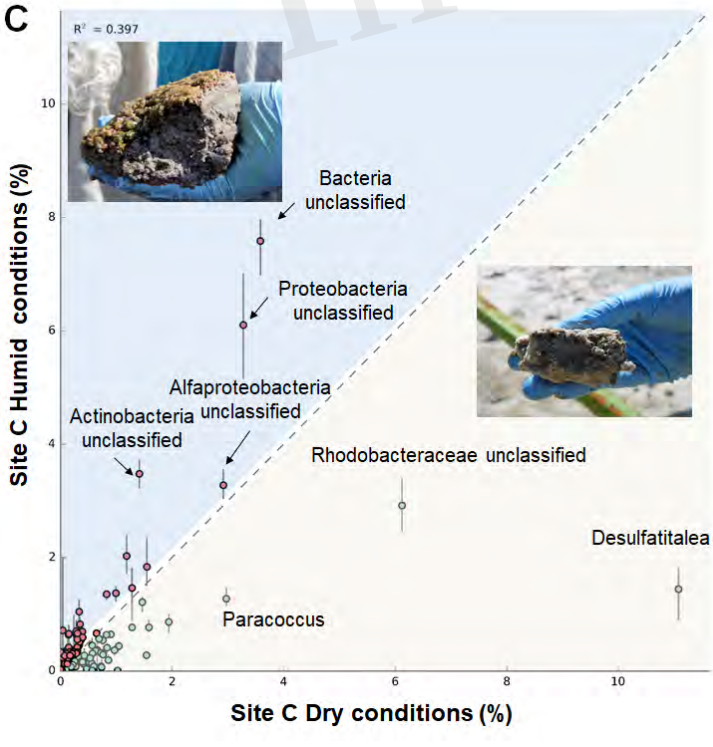



Figure 4.TIF

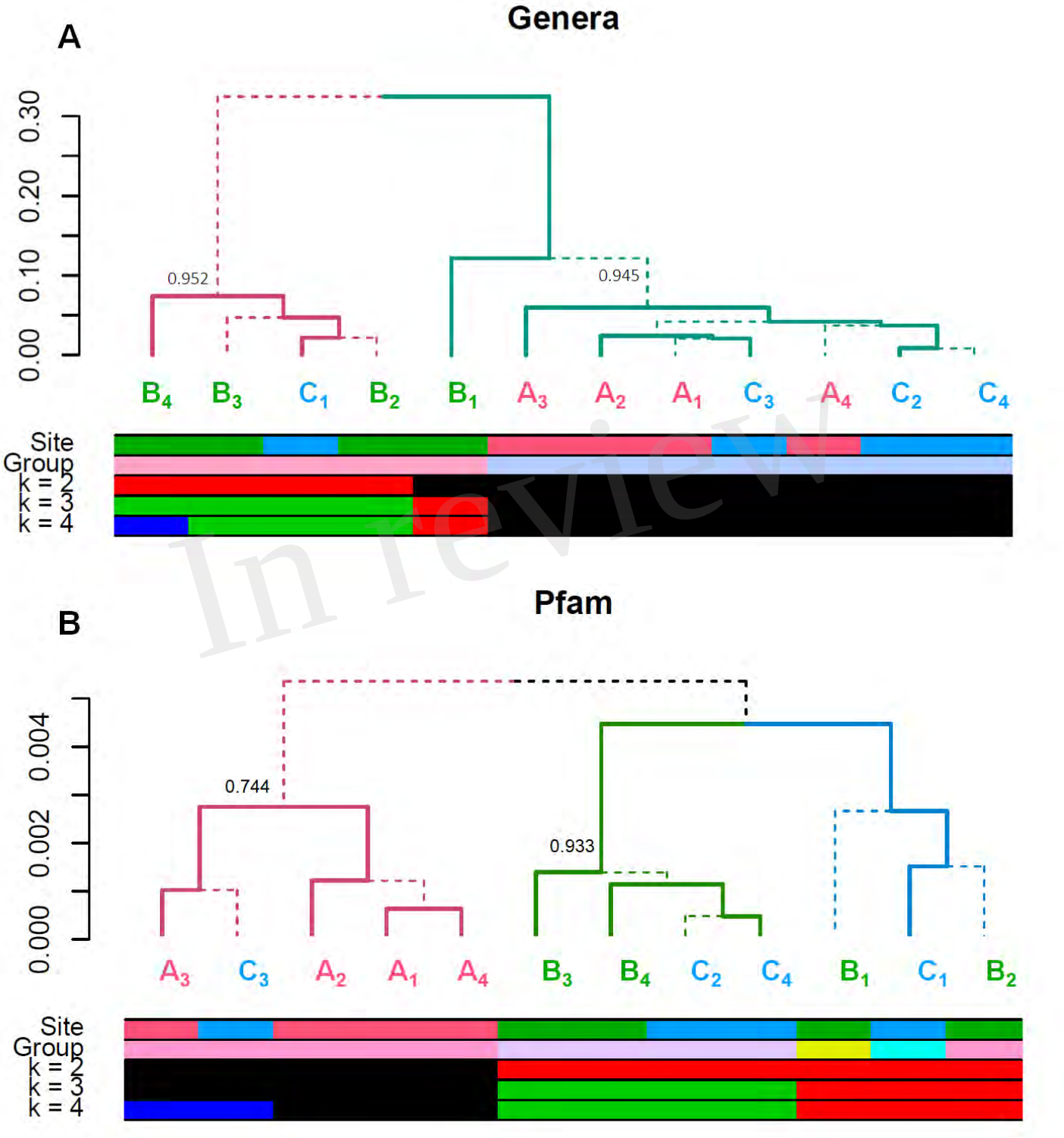


Figure 5.TIF

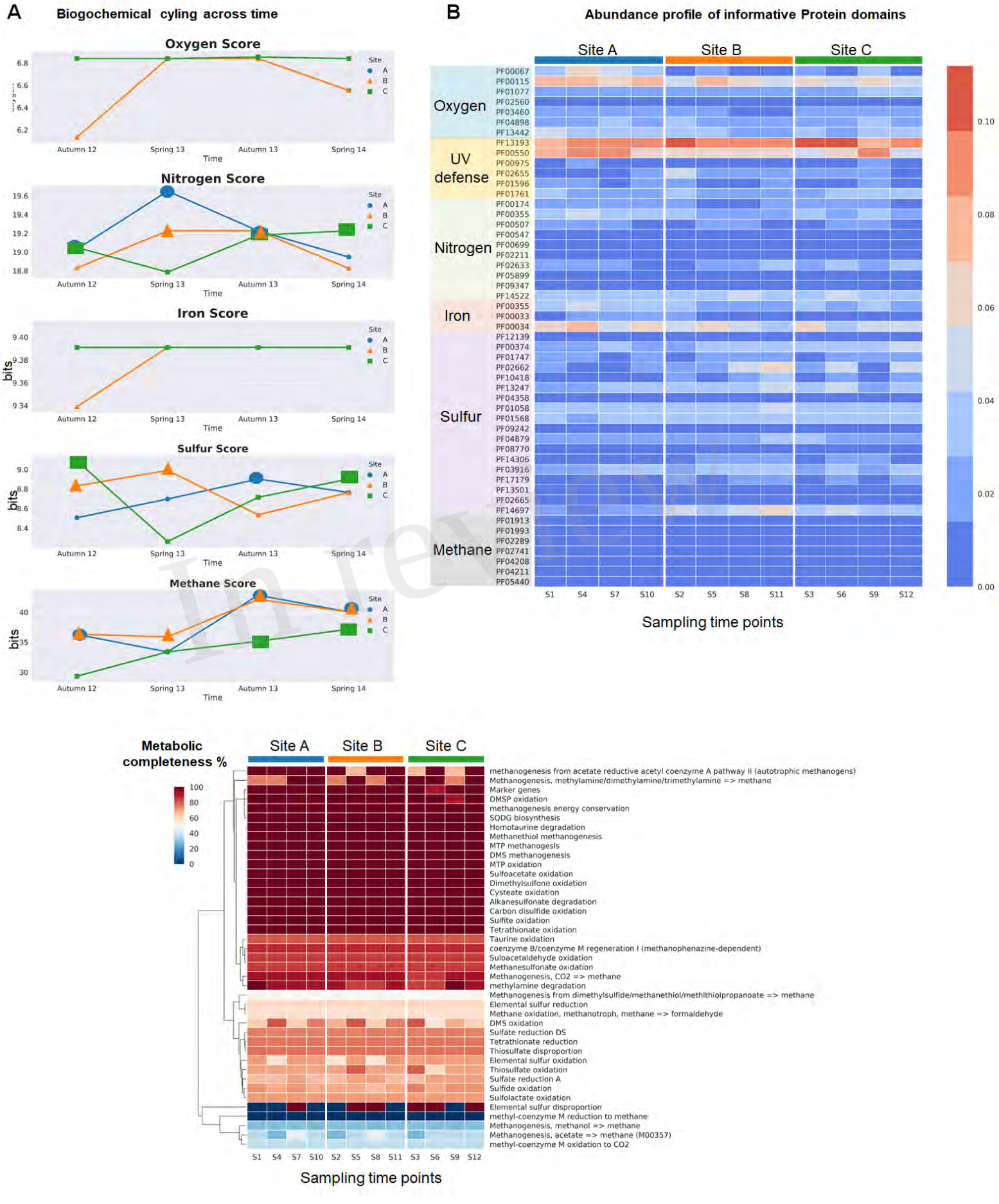

A
Biogochemical cyling across time
Oxygen Score
Nitrogen Score
Iron Score
Sulfur Score
Methane Score
Site
A
B
C
Autumn 12
Spring 13
Autumn 13
Spring 14
Time
bits
B
Abundance profile of informative Protein domains
Site A
Site B
Site C
Oxygen
UV defense
Nitrogen
Iron
Sulfur
Methane
S1 S4 S7 S10 S2 S5 S8 S11 S3 S6 S9 S12
Sampling time points
Metabolic completeness %
methanogenesis from acetate reductive acetyl coenzyme A pathway II (autotrophic methanogens)
Methanogenesis, methylamine/dimethylamine/trimethylamine => methane
Marker genes
DMSP oxidation
methanogenesis energy conservation
SQDG biosynthesis
Homotaurine degradation
Methanethiol methanogenesis
MTP methanogesis
DMS methanogenesis
MTP oxidation
Sulfoacetate oxidation
Dimethylsulfone oxidation
Cysteate oxidation
Alkanesulfonate degradation
Carbon disulfide oxidation
Sulfite oxidation
Tetrathionate oxidation
Taurine oxidation
coenzyme B/coenzyme M regeneration I (methanophenazine-dependent)
Suloacetaldehyde oxidation
Methanesulfonate oxidation
Methanogenesis, CO2 => methane
methylamine degradation
Methanogenesis from dimethylsulfide/methanethiol/methlthiolpropanoate => methane
Elemental sulfur reduction
Methane oxidation, methanotroph, methane => formaldehyde
DMS oxidation
Sulfate reduction DS
Tetrathionate reduction
Thiosulfate disproportion
Elemental sulfur oxidation
Thiosulfate oxidation
Sulfate reduction A
Sulfide oxidation
Sulfolactate oxidation
Elemental sulfur disproportion
methyl-coenzyme M reduction to methane
Methanogenesis, methanol => methane
Methanogenesis, acetate => methane (M00357)
methyl-coenzyme M oxidation to CO2
Sampling time points

Figure 6.TIF

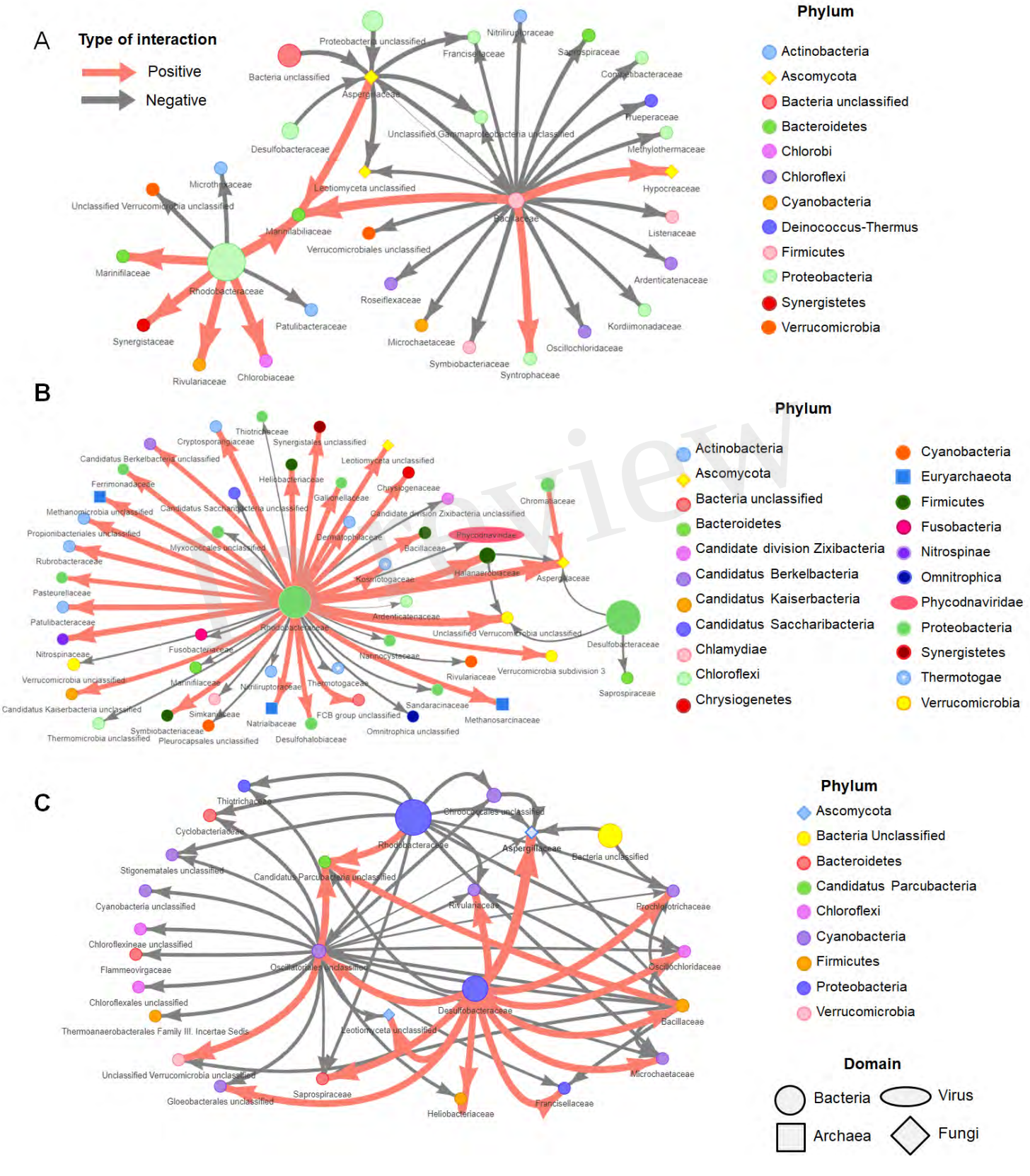
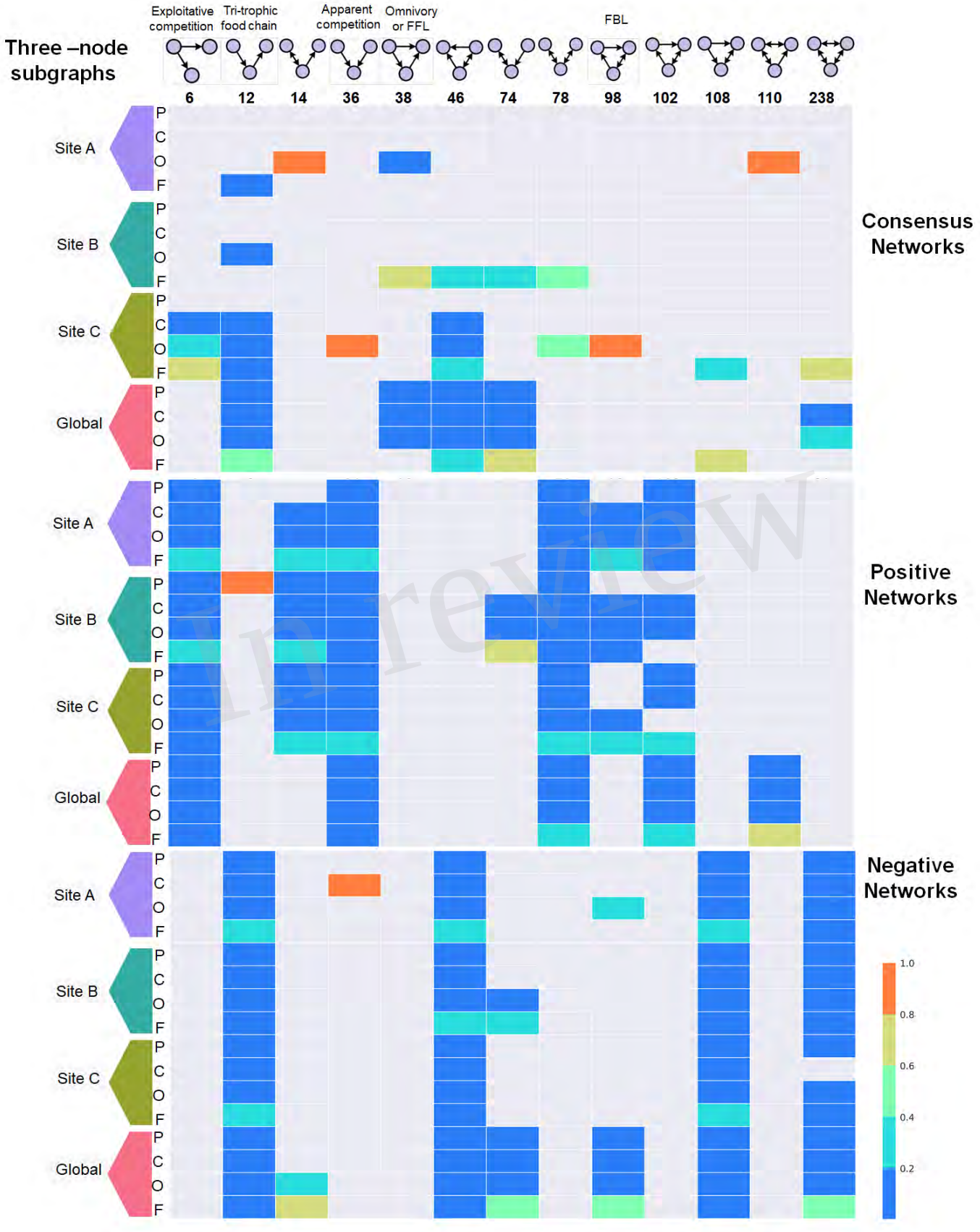

A
Type of interaction
Positive
Negative
Phylum
Actinobacteria
Ascomycota
Bacteria unclassified
Bacteroidetes
Chlorobi
Chloroflexi
Cyanobacteria
Deinococcus-Thermus
Firmicutes
Proteobacteria
Synergistetes
Verrucomicrobia
B
Phylum
Actinobacteria
Ascomycota
Bacteria unclassified
Bacteroidetes
Candidate division Zixibacteria
Candidatus Berkelbacteria
Candidatus Kaiserbacteria
Candidatus Saccharibacteria
Chlamydiae
Chloroflexi
Chrysiogenetes
Cyanobacteria
Euryarchaeota
Firmicutes
Fusobacteria
Nitrospinae
Omnitrophica
Phycodnaviridae
Proteobacteria
Synergistetes
Thermotogae
Verrucomicrobia
C
Phylum
Ascomycota
Bacteria Unclassified
Bacteroidetes
Candidatus Parcubacteria
Chloroflexi
Cyanobacteria
Firmicutes
Proteobacteria
Verrucomicrobia
Domain
Bacteria
Virus
Archaea
Fungi

Figure 7.TIF



# 0.4 Conclusiones generales

En mi tesis busqué entender cómo la sobreexplotación de recursos hídricos causado por actividades humanas en Cuatro Ciénegas Coahuila afecta a los tapetes microbianos de la Laguna del Churince. Para determinar los efectos de esta perturbación, analizamos los cambios en la estructura de la comunidad por dos años en los tapetes microbianos en condiciones contrastantes de disponibilidad de agua.

Por medio de métodos tradicionales de ecología microbiana se observó que los tapetes microbianos de CCC son sistemas con una alta diversidad, compuestos por un gran reservorio genético de miembros de la biosfera rara y que de manera general son comunidades bastante resistentes y resilientes frente a la desecación tanto a nivel taxonómico como funcional (metabólico).

Cuando aplicamos los métodos desarrollados en esta tesis doctoral (MEBS) y análisis de redes, encontramos que existe un efecto determinado por la falta de agua en los tapetes microbianos, afectando de manera general el ciclo del azufre y del metano. Estos, ciclos no solo son esenciales para el ensamblaje de estas comunidades, sino que también fueron los metabolismos que utilizaron de las primeras formas de vida en la Tierra. Consideramos que el uso de nuestro algoritmo MEBS fue más informativo que los métodos tradicionales derivados de la inferencia de perfiles de abundancias de genes marcadores moleculares.

También pudimos observar que cuando se recupera el sistema hidrológico y regresa el agua a la Laguna del Churince, se restablecen las funciones potenciales del ciclo biogeoquímico del azufre. Estos resultados son indicadores de la urgencia de medidas apropiadas le permitirán al gobierno federal (CONANP y CNA) tomar decisiones que redunden en la conservación del acuífero.

Por otra parte, esta tesis representa el primer estudio en llevar a cabo un enfoque de genómica y metagenómica comparada para el entendimiento global de ciclo biogeoquímico del azufre y su conexión con los demás ciclos. Así mismo, es un estudio

pionero en utilizar estas herramientas con fines de conservación en Cuatro Ciénegas Coahuila dada la intensa explotación de los recursos hídricos, con lo cual sumamos antecedentes directos para su conservación, y evidencia directa del impacto de la falta de agua en los tapetes microbianos.

Finalmente puedo enumerar mis conclusiones de esta manera:

1. La integración de datos metabólicos, ecológicos y genómicos del ciclo del azufre bajo un marco matemático sirvió para generar un nuevo algoritmo denominado MEBS, que permite evaluar con un solo valor informativo, la maquinaria metabólica de los ciclos biogeoquímicos a gran escala (miles de genomas y metagenomas).
2. Se implementó el uso de la entropía relativa para capturar la sobre-representación de la maquinaria metabólica de los ciclos biogeoquímicos.
3. Se realizó el primer inventario del ciclo del azufre de manera integral, y se recapitulo la gran diversidad de organismos responsables de metabolizar los compuestos del azufre en la biosfera.
4. Se propuso el uso del modelo del tapete microbiano como sistema ecológico de estudio para entender la intersección taxonómica y metabólica de los ciclos biogeoquímicos.
5. Se expuso la problemática de sobre-explotación de recursos hídricos en la Cuenca de Cuatro Ciénegas evaluando la dinámica de los tapetes microbianos bajo dichas condiciones de perturbación.
6. Al implementar el uso de *network motifs* en redes microbianas, este estudio permitió identificar interacciones específicas que podrían tener un significado diferente al propuesto en teoría ecológica propuesta para macro-organismos.
7. En este trabajo propongo que los cambios que dan lugar o que optimizan la diversidad y dan lugar a una mayor estabilidad en los tapetes microbianos podrían estar relacionados con cuatro factores: el aumento de relaciones de antagonismo y competencia bajo condiciones de perturbación, una baja modularidad en el que todos los miembros de la comunidad interaccionan entre sí, la gran diversidad

taxonómica siendo fundamentalmente los miembros de la biosfera rara y el *core* de microorganismos que pueden llevar a cabo funciones esenciales de liberación de bienes públicos.

8. Finalmente, una de las condiciones para que el funcionamiento ecosistémicos se mantenga frente a una perturbación ambiental, es la presencia de una amplia biosfera rara (poco abundante) ya que estos miembros de la comunidad que son raros en un momento son capaces de llevar acabo funciones metabólicas esenciales bajo diferentes condiciones ambientales.

# 0.5 Perspectivas

En ecología microbiana, el entender la compleja relación entre perturbación ambiental y estabilidad microbiana se ha basado en evaluar las propiedades emergentes de la comunidad después de una perturbación. Sin embargo, entender de manera integral dinámica de las comunidades aplicando inferencia de redes y representación de subredes (*network motifs*) así como el ciclaje de nutrientes, puede ser esencial para revelar las reglas de ensamblaje de la comunidad que permiten hacer frente a las perturbaciones ambientales.

Por lo tanto, se podría determinar si existe alguna correlación entre la dinámica de la comunidad y los procesos ecosistémicos, si se realiza una evaluación comprensiva de la composición, estructura, función y relaciones entre especies y genes o metabolitos. De esta forma, al evaluar de manera integral las comunidades microbianas, podremos generar principios que nos ayuden a construir modelos conceptuales e incluso predictivos frente una perturbación ambiental.

La dinámica de los ciclos biogeoquímicos en las comunidades microbianas puede ser abordada en varios niveles.

1. Tasas de respiración: Debido a que el $O_2$ es el primer aceptor de electrones en la mineralización de la materia orgánica, el medir las tasas respiratorias, permitirá cuantificar la importancia potencial de este proceso bajo distintos parámetros fisicoquímicos. Al cuantificar las tasas de desnitrificación (reducción del $NO_3^-$ a $N_2$) y $SO_4^{2-}$ se podría identificar la importancia relativa de cada uno de estos procesos bajo diferentes regímenes de nutrientes. La comprensión general de vías sitio-específicas funcionalmente importantes se puede obtener por medio de la correlación entre las tasas de respiración y las tasas de desnitrificación y reducción de $SO_4^{2-}$. (Cavalcanti et al., 2014; Fishbain et al., 2003; Kostka et al., 1999; Visscher et al., 1992)

2. Microelectrodos. Al utilizar sondas específicas para medir el oxígeno, y ácido sulfhídrico podrían servir para dar referencia de la actividad *in situ* (por ejemplo en tapetes

microbianos ser podría utilizar : http://www.coleparmer.com/buy/category/orion-silver-sulfide-ion-selective-electrode )

3. Isotopos: Medición de elementos en sus formas orgánicas e inorgánicas (CHNOSP). Por ejemplo, la composición isotópica de los compuestos de azufre en la naturaleza, puede indicar las vías microbianas por las cuales se forman los compuestos. Lo anterior aplica para ambientes modernos como en el tiempo geológicos (Aharon and Fu, 2000; Bontognali et al., 2012; Hoek and Canfield, 2008)

4. Geochip: Se puede evaluar la estructura, función y composición de las comunidades microbianas utilizando el microarreglo denominado GeoChip. La versión más reciente. GeoChip 5.0 contiene 167,044 sondas distintas, cubriendo un total de 395,894 secuencias codificantes (CDS) de aproximadamente 1500 familias funcionales involucradas en el metabolismo microbiano del carbono (degradación, fijación y metano), así como el ciclaje del nitrógeno, azufre y fosforo, metabolismo energético, homeostasis de metales, y otros genes como por ejemplo sistema CRISPR, metabolismo secuendario (por ejemplo antibióticos y pigmentos), respuesta a estrés, virus (bacteriófagos y virus de eucariontes), y virulencia http://www.glomics.com/gch-tech.html7 (Nikolaki and Tsiamis, 2013; Tu et al., 2014)

5. MEBS: Actualmente estamos trabajando en la segunda versión del algoritmo en la cual estamos implementado métodos de *machine learning* para que MEBS solo dependa de una lista de organismos que estén involucrados en algún metabolismo de interés (i.e degradación de xenobióticos, hidrocarburos etc). Esto permitirá generar bases de datos más robustas para evaluar de manera eficaz la dinámica de rutas metabólicas o ciclos biogeoquimos a escala global.

6. Tomando en cuenta los resultados de esta tesis, se plantea que combinando mayor número de réplicas y tiempos de muestreo (antes y después de la perturbación) se puede

relacionar la dinámica de los ciclos biogeoquímicos en función de la variabilidad de la diversidad taxonómica.

7. Al obtener más replicas, así como un mayor número de tiempos de muestreo, se podrían aplicar modelos matemáticos para conocer a estabilidad temporal de las comunidades microbianas. Por ejemplo, se podría utilizando la ley de Teylor con la que se puede conocer la dependencia entre la media de la abundancia relativa de los organismos y su dispersión temporal. Con esto se podría comparar comunidades bajo condiciones de perturbación vs comunidades no perturbadas y comprobar si realmente existe un patrón asociado a dicha perturbación (Dinleyici et al., 2018). Además, se podrían aplicar modelos matemáticos para predecir el comportamiento de las comunidades microbianas en función de perturbación (Barone et al., 2015; Steinway et al., 2015; Thébault and Fontaine, 2010)

Integrando las herramientas descritas anteriormente, se podrá determinar si existe alguna correlación entre la dinámica de los ciclos biogeoquímicos con la estructura y composición de la comunidad frente a alguna perturbación y los principios fundamentales que regulan la estabilidad microbiana y maximizan su diversidad funcional y taxonómica bajo condiciones de estrés.

# 0.6 Referencias

Aharon, P., and Fu, B. (2000). Microbial sulfate reduction rates and sulfur and oxygen isotope fractionations at oil and gas seeps in deepwater Gulf of Mexico. Geochim. Cosmochim. Acta 64, 233–246. doi:10.1016/S0016-7037(99)00292-6.

Barone, R., De Santi, C., Palma Esposito, F., Tedesco, P., Galati, F., Visone, M., et al. (2015). Metagenomics meets time series analysis: unraveling microbial community dynamics. Nat. Rev. Microbiol. 13, 217–229. doi:10.1038/nrmicro3400.

Bontognali, T. R. R., Sessions, A. L., Allwood, A. C., Fischer, W. W., Grotzinger, J. P., Summons, R. E., et al. (2012). Sulfur isotopes of organic matter preserved in 3.45-

billion-year-old stromatolites reveal microbial metabolism. Proc. Natl. Acad. Sci. U. S. A. 109, 15146–15151. doi:10.1073/pnas.1207491109.

Cavalcanti, G. S., Gregoracci, G. B., dos Santos, E. O., Silveira, C. B., Meirelles, P. M., Longo, L., et al. (2014). Physiologic and metagenomic attributes of the rhodoliths forming the largest CaCO3 bed in the South Atlantic Ocean. ISME J. 8, 52–62. doi:10.1038/ismej.2013.133.

Dinleyici, E. C., Martínez-martínez, D., Kara, A., Karbuz, A., and Dalgic, N. (2018). Time Series Analysis of the Microbiota of Children Suffering From Acute Infectious Diarrhea and Their Recovery After Treatment. 9, 1–11. doi:10.3389/fmicb.2018.01230.

Fishbain, S., Dillon, J. G., Gough, H. L., and Stahl, D. A. (2003). Linkage of High Rates of Sulfate Reduction in Yellowstone Hot Springs to Unique Sequence Types in the Dissimilatory Sulfate Respiration Pathway. 69, 3663–3667. doi:10.1128/AEM.69.6.3663.

Hoek, J., and Canfield, D. E. (2008). Controls on Isotope Fractionation During Dissimilatory Sulfate Reduction. 273–284.

Kostka, J. E., Thamdrup, B., Nohr, R., and Donald, G. (1999). Rates and pathways of carbon oxidation in permanently cold Arctic sediments. 180, 7–21.

Nikolaki, S., and Tsiamis, G. (2013). Microbial diversity in the era of omic technologies. Biomed Res. Int. 2013. doi:10.1155/2013/958719.

Steinway, S. N., Biggs, M. B., Loughran, T. P., Papin, J. A., and Albert, R. (2015). Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. PLOS Comput. Biol. 11, e1004338. doi:10.1371/journal.pcbi.1004338.

Thébault, E., and Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. Science (80-. ). 329, 853–856. doi:10.1126/science.1188321.

Tu, Q., Yu, H., He, Z., Deng, Y., Wu, L., Van Nostrand, J. D., et al. (2014). GeoChip 4: A functional gene-array-based high-throughput environmental technology for microbial community analysis. Mol. Ecol. Resour. 14, 914–928. doi:10.1111/1755-0998.12239.

Visscher, P. T., Prins, R. A., and Gemerden, H. Van (1992). Rates of sulfate reduction and thiosulfate consumption in a marine microbial mat. 86.

# Anexo 1

## Representantes taxonómicos del ciclo del azufre

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| **Colorless sulfur Bacteria (CLSB)** | | | | | | |
| 1 | CLSB | Achromatium | (Lindenmayer and Likens, 2010) | | | |
| 2 | CLSB | Beggiatoa | (Ce et al., 2007) | | | Beggiatoa leptomitiformis |
| 3 | CLSB | Macromonas | (Lindenmayer and Likens, 2010) | | | |
| 4 | CLSB | Paracoccus | (Canfield et al., 2005) | Paracoccus aminophilus JCM 7686 | Paracoccus aminophilus JCM 7686 | |
| | | | | Paracoccus denitrificans PD1222 | Paracoccus denitrificans PD1222 | |
| 5 | CLSB | Persephonella | (Ce et al., 2007) | Persephonella marina EX H1 | Persephonella marina EX-H1 | |
| 6 | CLSB | Sulfitobacter | (Curson et al., 2008) | | | Sulfitobacter sp. AM1-D1 |
| 7 | CLSB | Sulfurimonas | (Ce et al., 2007) | Sulfurimonas autotrophica DSM 16294 | Sulfurimonas autotrophica DSM 16294 | |
| | | | | Sulfurimonas denitrificans DSM 1251 | Sulfurimonas denitrificans DSM 1251 | |
| 8 | CLSB | Sulfurovum | (Ce et al., 2007) | Sulfurovum NBC37 1 | Sulfurovum sp. NBC37-1 | Sulfurovum lithotrophicum |
| 9 | CLSB | Thermothrix | (Lindenmayer and Likens, 2010) | | | |
| 10 | CLSB | Thioalkalimicrobium | (Sorokin et al., 2001) | Thioalkalimicrobium cyclicum ALM1 | Thioalkalimicrobium cyclicum ALM1 | Thioalkalimicrobium aerophilum AL3 |
| 11 | CLSB | Thiobacillus | (Canfield et al., 2005) | Thiobacillus denitrificans ATCC 25259 | Thiobacillus denitrificans ATCC 25259 | |
| 12 | CLSB | Thiobacterium | (Robertson and Kuenen, 2006) | | | |
| 13 | CLSB | Thiodendron | (Lindenmayer and Likens, 2010) | | | |
| 14 | CLSB | Thiomargarita | (Ce et al., 2007) | | | |
| 15 | CLSB | Thiomicrospira | (Canfield et al., 2005) | Thiomicrospira crunogena XCL 2 | Thiomicrospira crunogena XCL-2 | |
| 16 | CLSB | Thiomonas | (Ce et al., 2007) | Thiomonas 3As * no longer in refseq | Thiomonas arsenitoxydans * no lorger in refseq | *Only Thiomonas intermedia K12 |
| | | | | Thiomonas intermedia K12 | | |
| 17 | CLSB | Thioploca | (Ce et al., 2007) | | | Thioploca ingrica |
| 18 | CLSB | Thiosphaera | (Lindenmayer and Likens, 2010) | | | |
| 19 | CLSB | Thiospira | (Robertson and Kuenen, 2006) | | | |
| 20 | CLSB | Thiothrix | (Canfield et al., 2005) | | | |
| 21 | CLSB | Thiovulum | (Robertson and Kuenen, 2006) | | | |
| 22 | CLSB | Thioalkalivibrio | (Foti et al., 2006; Ghosh and Dam, 2009) | Thioalkalivibrio K90mix | Thioalkalivibrio sp. K90mix | Thioalkalivibrio paradoxus ARh 1 |
| | | | | Thioalkalivibrio nitratireducens DSM 14787 | Thioalkalivibrio nitratireducens DSM 14787 | Thioalkalivibrio versutus |
| | | | | Thioalkalivibrio sulfidophilus HL EbGr7 | Thioalkalivibrio sulfidophilus HL-EbGr7 | |
| 23 | CLSB | Thioalkalispira | (Foti et al., 2006; Ghosh and Dam, 2009) | | | |
| **Green Sulfur Bacteria (GSB)** | | | | | | |
| 1 | GSB | Ancalochloris | (Lindenmayer and Likens, 2010) | | | |
| 3 | GSB | Prosthecochloris | (Lindenmayer and Likens, 2010) | Prosthecochloris aestuarii DSM 271 | Prosthecochloris aestuarii DSM 271 | Prosthecochloris sp. CIB 2401 |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| 4 | GSB | Chlorobaculum | (Lindenmayer and Likens, 2010) | Chlorobaculum parvum NCIB 8327 | Chlorobaculum parvum NCIB 8327 | Chlorobaculum limnaeum |
| 5 | GSB | Chlorobium | (Lindenmayer and Likens, 2010) | Chlorobium chlorochromatii CaD3 | Chlorobium chlorochromatii CaD3 | |
| | | | | Chlorobium limicola DSM 245 | Chlorobium limicola DSM 245 | |
| | | | | Chlorobium luteolum DSM 273 | Chlorobium luteolum DSM 273 | |
| | | | | Chlorobium phaeobacteroides BS1 | Chlorobium phaeobacteroides BS1 | |
| | | | | Chlorobium phaeobacteroides DSM 266 | Chlorobium phaeobacteroides DSM 266 | |
| | | | | Chlorobium phaeovibrioides DSM 265 | Chlorobium phaeovibrioides DSM 265 | |
| | | | | Chlorobium tepidum TLS | Chlorobium tepidum TLS | |
| 6 | GSB | Chloroflexus | (Frigaard and Bryant, 2004) | Chloroflexus aggregans DSM 9485 | Chloroflexus aggregans DSM 9485 | |
| | | | | Chloroflexus aurantiacus J -0 -fl | Chloroflexus sp. Y-400-fl | |
| | | | | Chloroflexus sp. Y-400-fl | | |
| 7 | GSB | Chloroherpeton | (Lindenmayer and Likens, 2010) | Chloroherpeton thalassium ATCC 35110 | Chloroherpeton thalassium ATCC 35110 | |
| 8 | GSB | Pelodictyon | (Canfield et al., 2005) | Pelodictyon phaeoclathratiforme BU-1 | Pelodictyon phaeoclathratiforme BU-1 | |
| 9 | GSB | Heliobacterium | (Tang et al., 2010) | Heliobacterium modesticaldum Ice1 | Heliobacterium modesticaldum Ice1 | |
| **Purple Sulfur Bacteria (PSB)** | | | | | | |
| 1 | PSB | Allochromatium | (Lindenmayer and Likens, 2010) | Allochromatium vinosum DSM 180 | Allochromatium vinosum DSM 180 | |
| 2 | PSB | Amoeobobacter | (Canfield et al., 2005) | | | |
| 3 | PSB | Chromatium | (Canfield et al., 2005) | | | |
| 4 | PSB | Ectothiorhodospira | (Lindenmayer and Likens, 2010) | Ectothiorhodospiraceae bacterium M19 40 * no longer in refseq | | Ectothiorhodospira sp. BSL-9 |
| 5 | PSB | Halochromatium | (Lindenmayer and Likens, 2010) | | | |
| 6 | PSB | Halorhodospira | (Lindenmayer and Likens, 2010) | Halorhodospira halophila SL1 | Halorhodospira halophila SL1 | |
| 7 | PSB | Isochromatium | (Imhoff et al., 1998) | | | |
| 8 | PSB | Lamprobacter | (Lindenmayer and Likens, 2010) | | | |
| 9 | PSB | Marichromatium | (Imhoff et al., 1998) | | | Marichromatium purpuratum 984 |
| 10 | PSB | Rhabdochromatium | (Lindenmayer and Likens, 2010) | | | |
| 11 | PSB | Thermochromatium | (Lindenmayer and Likens, 2010) | | | |
| 12 | PSB | Thioalkalicoccus | (Lindenmayer and Likens, 2010) | | | |
| 13 | PSB | Thiobaca | (Frigaard and Dahl, 2009) | | | |
| 14 | PSB | Thiocapsa | (Lindenmayer and Likens, 2010) | | | |
| 15 | PSB | Thiococcus | (Lindenmayer and Likens, 2010) | | | |
| 16 | PSB | Thiocystis | (Canfield et al., 2005; Lindenmayer and Likens, 2010) | Thiocystis violascens DSM 198 | Thiocystis violascens DSM 198 | |
| 17 | PSB | Thiodictyon | (Lindenmayer and Likens, 2010) | | | |
| 18 | PSB | Thioflavicoccus | (Lindenmayer and Likens, 2010) | Thioflavicoccus mobilis 8321 uid184343 | Thioflavicoccus mobilis 8321 | |
| 19 | PSB | Thiohalocapsa | (Canfield et al., 2005; Lindenmayer and Likens, 2010) | | | |
| 20 | PSB | Thiolamprovum | (Lindenmayer and Likens, 2010) | | | |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| 21 | PSB | Thiopedia | (Canfield et al., 2005; Lindenmayer and Likens, 2010) | | | |
| 22 | PSB | Thiorhodococcus | (Lindenmayer and Likens, 2010) | | | |
| 23 | PSB | Thiorhodospira | (Lindenmayer and Likens, 2010) | | | |
| 24 | PSB | Thiorhodovibrio | (Lindenmayer and Likens, 2010) | | | |
| 25 | PSB | Thiospirillum | (Canfield et al., 2005) | | | |
| Sulfur oxidizer microorganisms (SOM) | | | | | | |
| 1 | SOM | Metallosphaera | (Auernik et al., 2008) | Metallosphaera cuprinaAr-4 | Metallosphaera cuprina Ar-4 | Methallosphaera sedula strains: ARS50-1,ARS50-2, ARS120-1 ,ARS120-2, SARC-M1 |
| | | | | Metallosphaera sedulaDSM 5348 | Metallosphaera sedula DSM 5348 | |
| 2 | SOM | Aeropyrum | (Nakagawa, 2004) | Aeropyrum camini SY1 =JCM 12091 (no longer in Refseq) | Aeropyrum camini SY1 = JCM 12091 | |
| | | | | Aeropyrum pernix K1 (no longer in Refseq) | Aeropyrum pernix K1 | |
| 3 | SOM | Sulfurihydrogenibium | (Ce et al., 2007) | Sulfurihydrogenibium azorenseAz-Fu1 | Sulfurihydrogenibium azorense Az-Fu1 | |
| | | | | Sulfurihydrogenibium sp.YO3AOP1 | Sulfurihydrogenibium sp. YO3AOP1 | |
| 4 | SOM | Alkalilimnicola | (Frigaard and Dahl, 2009) | Alkalilimnicola ehrlichii MLHE-1 | Alkalilimnicola ehrlichii MLHE-1 | |
| 5 | SOM | Acidithiobacillus | (Chen et al., 2012; Liljeqvist et al., 2011; Sorokin et al., 2006; Valdes et al., 2009) | Acidithiobacillus caldus SM-1 | Acidithiobacillus caldus SM-1 | Acidithiobacillus caldus ATCC 51756 |
| | | | | Acidithiobacillus ferrivorans SS3 | Acidithiobacillus ferrivorans SS3 | |
| | | | | Acidithiobacillus ferrooxidans ATCC 23270 | Acidithiobacillus ferrooxidans ATCC 23270 | |
| | | | | Acidithiobacillus ferrooxidans ATCC 53993 | | |
| 6 | SOM | Thermothiobacillus | (Sorokin et al., 2006) | | | |
| 7 | SOM | Halothiobacillus | (Sorokin et al., 2006) | Halothiobacillus neapolitanus c2 | Halothiobacillus neapolitanus c2 | Halothiobacillus sp. LS2 |
| 8 | SOM | Thermocrinis | (Canfield et al., 2010) | Thermocrinis albus DSM 14484 | Thermocrinis albus DSM 14484 | |
| 9 | SOM | Rhodobium | (Canfield et al., 2010) | | | |
| 10 | SOM | Aquaspirillum | (Canfield et al., 2010) | | | |
| 11 | SOM | Thiovirga | (Canfield et al., 2010) | | | |
| 12 | SOM | Leucothrix | (Canfield et al., 2010) | | | |
| 13 | SOM | Thiolapillus | (Nunoura et al., 2014) | Thiolapillus brandeum * not previosly considered (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104959) | | |
| Sulfur Oxidizer (SO) Sulfur Reducers (SR) and Sulfate Reducing Bacteria (SRB) | | | | | | |
| 1 | SO/SR | Sulfolobus | (Robertson and Kuenen, 2006) | Sulfolobus acidocaldarius DSM 639 | Sulfolobus acidocaldarius DSM 639 | Sulfolobus sp. A20 |
| | | | | Sulfolobus acidocaldarius N8 | Sulfolobus islandicus Y.N.15.51 | Sulfolobus sulfataricus strain SULB |
| | | | | Sulfolobus acidocaldarius Ron12/I | Sulfolobus solfataricus P2 | Sulfolobus sulfataricus strain 98/2 SULC |
| | | | | Sulfolobus acidocaldarius SUSAZ | Sulfolobus tokodaii str. 7 | Sulfolobus sulfataricus strain SULA |
| | | | | Sulfolobus islandicus HVE10/4 | | |
| | | | | Sulfolobus islandicus LAL14/1 | | |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| | | | | Sulfolobus islandicus L.D.8.5 * (no longer in Refseq) | | |
| | | | | Sulfolobus islandicus L.S.2.15 | | |
| | | | | Sulfolobus islandicus M.14.25 | | |
| | | | | Sulfolobus islandicus M.16.27 | | |
| | | | | Sulfolobus islandicus M.16.4 | | |
| | | | | Sulfolobus islandicus REY15A | | |
| | | | | Sulfolobus islandicus Y.G.57.14 | | |
| | | | | Sulfolobus islandicus Y.N.15.51 | | |
| | | | | Sulfolobus solfataricus 98/2 | | |
| | | | | Sulfolobus solfataricus P2 | | |
| | | | | Sulfolobus tokodaii str. 7 | | |
| 2 | SO/SR | Acidianus | (Lindenmayer and Likens, 2010) | Acidianus hospitalis W1 | Acidianus hospitalis W1 | |
| 3 | SO/SR | Aquifex | (Ce et al., 2007) | Aquifex aeolicus VF5 | Aquifex aeolicus VF5 | |
| 4 | SO/SR | Thermoplasma | (Schleper et al., 1995) | Thermoplasma acidophilum DSM 1728 | Thermoplasma acidophilum DSM 1728 | |
| | | | | Thermoplasmatales archaeon BRNA1 | Thermoplasma volcanium GSS1 | |
| | | | | Thermoplasma volcanium GSS1 | | |
| 1 | SR | Desulfurella | (Lindenmayer and Likens, 2010) | | | Desulfurella acetivorans A63 |
| 2 | SR | Desulfurobacterium | (Lindenmayer and Likens, 2010) | Desulfurobacterium thermolithotrophum DSM 11699 | Desulfurobacterium thermolithotrophum DSM 11699 | |
| 3 | SR | Desulfuromonas | (Lindenmayer and Likens, 2010) | | | |
| 4 | SR | Dethiosulfovibrio | (Lindenmayer and Likens, 2010) | | | |
| 5 | SR | Pelobacter | (Lindenmayer and Likens, 2010) | Pelobacter carbinolicus DSM 2380 | Pelobacter carbinolicus DSM 2380 | |
| | | | | Pelobacter propionicus DSM 2379 | Pelobacter propionicus DSM 2379 | |
| 6 | SR | Sulfurospirillum | (Lindenmayer and Likens, 2010) | Sulfurospirillum barnesii SES-3 | Sulfurospirillum barnesii SES-3 | Sulfurospirillum multivorans DSM 12446 |
| | | | | Sulfurospirillum deleyianum DSM 6946 | Sulfurospirillum deleyianum DSM 6946 | Sulfurospirillum sp. UCH001 |
| | | | | | | Sulfurospirillum cavolei |
| | | | | | | Sulfurospirillum halorespirans DSM 13726 |
| 7 | SR | Thermoproteus | (Lindenmayer and Likens, 2010) | Thermoproteus tenax Kra 1 | Thermoproteus tenax Kra 1 | |
| | | | | Thermoproteus uzoniensis 768-20 | Thermoproteus uzoniensis 768-20 | |
| 8 | SR | Desulfurococcus | (Ce et al., 2007) | Desulfurococcus fermentans DSM 16532 | Desulfurococcus fermentans DSM 16532 | |
| | | | | Desulfurococcus kamchatkensis 1221n | Desulfurococcus mucosus DSM 2162 | |
| | | | | Desulfurococcus mucosus DSM 2162 | | |
| 9 | SR | Desulfuromusa | (Ce et al., 2007) | | | |
| 10 | SR | Nautilia | (Ce et al., 2007) | Nautilia profundicola AmH | Nautilia profundicola AmH | |
| 11 | SR | Thermococcus | (Ce et al., 2007) | Thermococcus sp. 4557 | Thermococcus barophilus MP | Thermococcus cleftensis |
| | | | | Thermococcus sp. AM4 | Thermococcus gammatolerans EJ3 | Thermococcus paralvinellae |
| | | | | Thermococcus barophilus MP | Thermococcus kodakarensis KOD1 | Thermococcus eurythermalis |
| | | | | Thermococcus CL1 *no longer in refseq | Thermococcus litoralis DSM 5473 | Thermococcus guaymasensis DSM 11113 |
| | | | | Thermococcus gammatolerans | Thermococcus onnurineus NA1 | Thermococcus barophilus |
| | | | | Thermococcus kodakarensis KOD1 | Thermococcus sibiricus MM 739 | Thermococcus sp. 2319x1 |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| | | | | Thermococcus litoralis DSM 5473 | Thermococcus sp. 4557 | Thermococcus peptonophilus |
| | | | | Thermococcus onnurineus NA1 | Thermococcus sp. AM4 | Thermococcus piezophilus |
| | | | | Thermococcus sibiricus MM 739 | Thermococcus sp. CL1 | Thermococcus chitonophagu |
| 12 | SR | Caminibacter | (Ce et al., 2007) | | | |
| 13 | SR | Hyperthermus | (Rabus et al., 2006) | Hyperthermus butylicus DSM 5456 | Hyperthermus butylicus DSM 5456 | |
| 14 | SR | Ignicoccus | (Rabus et al., 2006) | Ignicoccus hospitalis KIN4/I | Ignicoccus hospitalis KIN4/I | |
| 15 | SR | Pyrobaculum | (Rabus et al., 2006) | Pyrobaculum 1860 *no longer in refseq | Pyrobaculum aerophilum str. IM2 | |
| | | | | Pyrobaculum aerophilum IM2 *no longer in refseq | Pyrobaculum arsenaticum DSM 13514 | |
| | | | | Pyrobaculum arsenaticum DSM 13514 *no longer in refseq | Pyrobaculum calidifontis JCM 11548 | |
| | | | | Pyrobaculum calidifontis JCM 11548 | Pyrobaculum islandicum DSM 4184 | |
| | | | | Pyrobaculum islandicum DSM 4184 | Pyrobaculum neutrophilum V24Sta | |
| | | | | Pyrobaculum neutrophilum V24Sta *no longer in refseq | Pyrobaculum oguniense TE7 | |
| | | | | Pyrobaculum oguniense TE7*no longer in refseq | Pyrobaculum sp. 1860 | |
| 16 | SR | Pyrococcus | (Rabus et al., 2006) | Pyrococcus abyssi GE5 | Pyrococcus abyssi GE5 | Pyrococcus kukulkanii |
| | | | | Pyrococcus furiosus COM1 | Pyrococcus furiosus COM1 | |
| | | | | Pyrococcus furiosus DSM 3638 | Pyrococcus horikoshii OT3 | |
| | | | | Pyrococcus horikoshii OT3 | Pyrococcus sp. NA2 | |
| | | | | Pyrococcus NA2 | Pyrococcus sp. ST04 | |
| | | | | Pyrococcus sp. ST04 | Pyrococcus yayanosii CH1 | |
| | | | | Pyrococcus yayanosii CH1 | | |
| 17 | SR | Pyrodictium | (Rabus et al., 2006) | | | Pyrodictium delaneyi |
| 18 | SR | Staphylothermus | (Rabus et al., 2006) | Staphylothermus hellenicus DSM 12710 | Staphylothermus hellenicus DSM 12710 | |
| | | | | Staphylothermus marinus F1 | Staphylothermus marinus F1 | |
| 19 | SR | Thermotoga | (Rabus et al., 2006) | Thermotoga lettingae TMO* no longer in refseq | Thermotoga lettingae TMO | Thermotoga sp. 2812B |
| | | | | Thermotoga maritima MSB8 | Thermotoga neapolitana DSM 4359 | Thermotoga sp. Cell2 |
| | | | | Thermotoga maritima MSB8 | Thermotoga sp. RQ2 | Thermotoga caldifontis AZM44c09 |
| | | | | Thermotoga maritima MSB8 | Thermotoga thermarum DSM 5069 | Thermotoga profunda AZM34c06 |
| | | | | Thermotoga naphthophila RKU-10 | | Thermotoga sp. RQ7 |
| | | | | Thermotoga neapolitana DSM 4359 | | Thermotoga maritima strain Tma100 |
| | | | | Thermotoga petrophila RKU-1 | | Thermotoga maritima strain Tma200 |
| | | | | Thermotoga sp. RQ2 | | |
| | | | | Thermotoga thermarum DSM 5069 | | |
| 1 | SRB | Desulfobacula | (Canfield et al., 2010) | Desulfobacula toluolica Tol2 | Desulfobacula toluolica Tol2 | |
| 2 | SRB | Desulfocaldus | (Canfield et al., 2010) | | | |
| 3 | SRB | Desulfocella | (Canfield et al., 2010) | | | |
| 4 | SRB | Desulfofrigus | (Canfield et al., 2010) | | | |
| 5 | SRB | Desulfonatronum | (Canfield et al., 2010) | | | |
| 6 | SRB | Desulfonauticus | (Canfield et al., 2010) | | | |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| 7 | SRB | Desulforegula | (Canfield et al., 2010) | | | |
| 8 | SRB | Desulforhabdus | (Canfield et al., 2010) | | | |
| 9 | SRB | Desulfospira | (Canfield et al., 2010) | | | |
| 10 | SRB | Desulfothermus | (Canfield et al., 2010) | | | |
| 11 | SRB | Desulfotignum | (Canfield et al., 2010) | | | |
| 12 | SRB | Thermodesulforhabdus | (Canfield et al., 2010) | | | |
| 13 | SRB | Archaeoglobus | (Pereira Cardoso et al., 2011) | Archaeoglobus fulgidus DSM 4304 | Archaeoglobus fulgidus DSM 4304 | Archaeoglobus fulgidus DSM 8774 |
| | | | | Archaeoglobus profundus DSM 5631 | Archaeoglobus profundus DSM 5631 | |
| | | | | Archaeoglobus sulfaticallidus PM70-1 | Archaeoglobus sulfaticallidus PM70-1 | |
| | | | | Archaeoglobus veneficus SNP6 | Archaeoglobus veneficus SNP6 | |
| 14 | SRB | Desulfoarculus | (Lindenmayer and Likens, 2010) | | | |
| 15 | SRB | Desulfobacca | (Lindenmayer and Likens, 2010) | Desulfobacca acetoxidans DSM 11109 | Desulfobacca acetoxidans DSM 11109 | |
| 16 | SRB | Desulfobacter | (Lindenmayer and Likens, 2010) | | | |
| 17 | SRB | Desulfobacterium | (Lindenmayer and Likens, 2010) | Desulfobacterium autotrophicum HRM2 | Desulfobacterium autotrophicum HRM2 | |
| 18 | SRB | Desulfobotulus | (Lindenmayer and Likens, 2010) | | | |
| 19 | SRB | Desulfobulbus | (Lindenmayer and Likens, 2010) | Desulfobulbus propionicus DSM 2032 | Desulfobulbus propionicus DSM 2032 | |
| 20 | SRB | Desulfocapsa | (Lindenmayer and Likens, 2010) | Desulfocapsa sulfexigens DSM 10523 | Desulfocapsa sulfexigens DSM 10523 | |
| 21 | SRB | Desulfococcus | (Pereira et al., 2011) | Desulfococcus oleovorans Hxd3 | Desulfococcus oleovorans Hxd3 | Desulfococcus multivorans |
| 22 | SRB | Desulfofustis | (Lindenmayer and Likens, 2010) | | | |
| 23 | SRB | Desulfohalobium | (Pereira et al., 2011) | Desulfohalobium retbaense DSM 5692 | Desulfohalobium retbaense DSM 5692 | |
| 24 | SRB | Desulfomonile | (Lindenmayer and Likens, 2010) | Desulfomonile tiedjei DSM 6799 | Desulfomonile tiedjei DSM 6799 | |
| 25 | SRB | Desulfonatronovibrio | (Lindenmayer and Likens, 2010) | | | |
| 26 | SRB | Desulfonema | (Ce et al., 2007) | | | |
| 27 | SRB | Desulforhopalus | (Lindenmayer and Likens, 2010) | | | |
| 28 | SRB | Desulfosarcina | (Lindenmayer and Likens, 2010) | | | |
| 29 | SRB | Desulfosporosinus | (Lindenmayer and Likens, 2010) | Desulfosporosinus acidiphilus SJ4 | Desulfosporosinus acidiphilus SJ4 | |
| | | | | Desulfosporosinus meridiei DSM 13257 | Desulfosporosinus meridiei DSM 13257 | |
| | | | | Desulfosporosinus orientis DSM 765 | Desulfosporosinus orientis DSM 765 | |
| 30 | SRB | Desulfotomaculum | (Ce et al., 2007) | Desulfotomaculum acetoxidans DSM 771 | Desulfotomaculum acetoxidans DSM 771 | |
| | | | | Desulfotomaculum carboxydivorans CO-1-SRB | Desulfotomaculum carboxydivorans CO-1-SRB | |
| | | | | Desulfotomaculum gibsoniae DSM 7213 | Desulfotomaculum gibsoniae DSM 7213 | |
| | | | | Desulfotomaculum kuznetsovii DSM 6115 | Desulfotomaculum kuznetsovii DSM 6115 | |
| | | | | Desulfotomaculum reducens MI-1 | Desulfotomaculum reducens MI-1 | |
| | | | | Desulfotomaculum ruminis DSM 2154 | Desulfotomaculum ruminis DSM 2154 | |
| 31 | SRB | Desulfovibrio | (Pereira et al., 2011) | Desulfovibrio aespoeensis Aspo 2 | Desulfovibrio aespoeensis Aspo-2 | Desulfovibrio fairfieldensis |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| | | | | Desulfovibrio africanus Walvis Bay | Desulfovibrio africanus str. Walvis Bay | Desulfovibrio sp. J2 |
| | | | | Desulfovibrio alaskensis G20 | Desulfovibrio alaskensis G20 | |
| | | | | Desulfovibrio desulfuricans ATCC 27774 | Desulfovibrio desulfuricans ND132 | |
| | | | | Desulfovibrio desulfuricans ND132 | Desulfovibrio desulfuricans subsp. desulfuricans str. ATCC 27774 | |
| | | | | Desulfovibrio gigas DSM 1382 *no longer in refseq | Desulfovibrio gigas DSM 1382 = ATCC 19364 | |
| | | | | Desulfovibrio hydrothermalis AM13 DSM 14728 *no longer in refseq | Desulfovibrio magneticus RS-1 | |
| | | | | Desulfovibrio magneticus RS-1 | Desulfovibrio piezophilus C1TLV30 | |
| | | | | Desulfovibrio piezophilus C1TLV30 *no longer in refseq | Desulfovibrio salexigens DSM 2638 | |
| | | | | Desulfovibrio salexigens DSM 2638 | Desulfovibrio vulgaris str. 'Miyazaki F' | |
| | | | | Desulfovibrio vulgaris DP4 | Desulfovibrio vulgaris str. Hildenborough | |
| | | | | Desulfovibrio vulgaris Hildenborough | | |
| | | | | Desulfovibrio vulgaris str. 'Miyazaki F' | | |
| | | | | Desulfovibrio vulgaris RCH1 | | |
| 32 | SRB | Thermodesulfatator | (Moussard et al., 2004) | Thermodesulfatator indicus DSM 15286 | Thermodesulfatator indicus DSM 15286 | |
| 33 | SRB | Desulfatibacillum | (Pereira et al., 2011) | Desulfatibacillum alkenivorans AK-01 | Desulfatibacillum alkenivorans AK-01 | |
| 34 | SRB | Desulfonatronospira | (Pereira et al., 2011) | | | |
| 35 | SRB | Desulfotalea | (Pereira et al., 2011; Ceet al., 2007) | Desulfotalea psychrophila LSv54 | Desulfotalea psychrophila LSv54 | |
| 36 | SRB | Desulfurivibrio | (Pereira et al., 2011) | Desulfurivibrio alkaliphilus AHT 2 | Desulfurivibrio alkaliphilus AHT2 | |
| 37 | SRB | Syntrophobacter | (Pereira et al., 2011) | Syntrophobacter fumaroxidans MPOB | Syntrophobacter fumaroxidans MPOB | |
| 38 | SRB | Thermodesulfovibrio | (Pereira et al., 2011) | Thermodesulfovibrio yellowstonii DSM 11347 | Thermodesulfovibrio yellowstonii DSM 11347 | |
| 39 | SRB | Thermodesulfobacterium | (Ce et al., 2007) | Thermodesulfobacterium geofontis OPF15 | Thermodesulfobacterium geofontis OPF15 | Thermodesulfobacterium commune DSM 2178 |
| 40 | SRB | Desulfovirga | (Rabus et al., 2006) | | | |
| 1 | SRB/SR | Desulfatiferula | (Rabus et al., 2006) | | | |
| 2 | SRB/SR | Ammonifex | (Pereira et al., 2011; Ce et al., 2007) | Ammonifex degensii KC4 | Ammonifex degensii KC4 | |
| 3 | SRB/SR | Caldivirga | (Pereira et al., 2011) | Caldivirga maquilingensis IC-167 | Caldivirga maquilingensis IC-167 | |
| 4 | SRB/SR | Desulfatirhabdium | (Rabus et al., 2006) | | | |
| 5 | SRB/SR | Desulfoluna | (Rabus et al., 2006) | | | |
| 6 | SRB/SR | Desulfomusa | (Rabus et al., 2006) | | | |
| 7 | SRB/SR | Desulfosalsimonas | (Rabus et al., 2006) | | | |
| 8 | SRB/SR | Desulfofaba | (Rabus et al., 2006) | | | |
| 9 | SRB/SR | Desulfopila | (Rabus et al., 2006) | | | |
| 10 | SRB/SR | Desulfovermiculus | (Rabus et al., 2006) | | | |
| 11 | SRB/SR | Desulfocurvus | (Rabus et al., 2006) | | | |
| 12 | SRB/SR | Malonomonas | (Rabus et al., 2006) | | | |
| 13 | SRB/SR | Desulfoglaeba | (Rabus et al., 2006) | | | |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| 14 | SRB/SR | Desulfosoma | (Rabus et al., 2006) | | | |
| 15 | SRB/SR | Desulfacinum | (Rabus et al., 2006) | | | |
| 17 | SRB/SR | Desulfitobacter | (Rabus et al., 2006) | | | |
| 18 | SRB/SR | Desulfitispora | (Rabus et al., 2006) | | | |
| 19 | SRB/SR | Desulfurispora | (Rabus et al., 2006) | | | |
| 20 | SRB/SR | Syntrophobotulus | (Rabus et al., 2006) | Syntrophobotulus glycolicus DSM 8271 | Syntrophobotulus glycolicus DSM 8271 | |
| 23 | SRB/SR | Caldimicrobium | (Rabus et al., 2006) | | | Caldimicrobium thiodismutans |
| 25 | SRB/SR | Thermovibrio | (Rabus et al., 2006) | Thermovibrio ammonificans HB 1 | Thermovibrio ammonificans HB-1 | Thermovibrio ammonificans HB-1 |
| 26 | SRB/SR | Hippea | (Rabus et al., 2006) | Hippea maritima DSM 10411 | Hippea maritima DSM 10411 | Hippea maritima DSM 10411 |
| 27 | SRB/SR | Desulfitobacterium | (Rabus et al., 2006) | Desulfitobacterium dehalogenans ATCC 51507 | Desulfitobacterium dehalogenans ATCC 51507 | Desulfitobacterium metallireducens DSM 15288 |
| | | | | Desulfitobacterium dichloroeliminans LMG P-21439 | Desulfitobacterium dichloroeliminans LMG P-21439 | |
| | | | | Desulfitobacterium hafniense DCB-2 | Desulfitobacterium hafniense Y51 | |
| | | | | Desulfitobacterium hafniense Y51 | | |
| 28 | SRB/SR | Thermovirga | (Rabus et al., 2006) | Thermovirga lienii DSM 17291 | Thermovirga lienii DSM 17291 | |
| 29 | SRB/SR | Dehalospirillum | (Rabus et al., 2006) | | | |
| 30 | SRB/SR | Sulfospirillum | (Rabus et al., 2006) | | | |
| 31 | SRB/SR | Desulfurispira | (Rabus et al., 2006) | | | |
| 32 | SRB/SR | Desulfurispirillum | (Rabus et al., 2006) | Desulfurispirillum indicum S5 | Desulfurispirillum indicum S5 | |
| 33 | SRB/SR | Sulfophobococcus | (Rabus et al., 2006) | | | |
| 34 | SRB/SR | Thermosphaera | (Rabus et al., 2006) | Thermosphaera aggregans DSM 11486 | Thermosphaera aggregans DSM 11486 | |
| 35 | SRB/SR | Stetteria | (Rabus et al., 2006) | | | |
| 36 | SRB/SR | Thermodiscus | (Rabus et al., 2006) | | | |
| 37 | SRB/SR | Pyrolobus | (Rabus et al., 2006) | Pyrolobus fumarii 1A | Pyrolobus fumarii 1A | |
| 38 | SRB/SR | Desulfurolobus | (Rabus et al., 2006) | | | |
| 39 | SRB/SR | Sulfurisphaera | (Rabus et al., 2006) | | | |
| 40 | SRB/SR | Stygiolobus | (Rabus et al., 2006) | | | |
| 41 | SRB/SR | Thermocladium | (Rabus et al., 2006) | | | |
| 42 | SRB/SR | Vulcanisaeta | (Rabus et al., 2006) | Vulcanisaeta distributa DSM 14429 | Vulcanisaeta distributa DSM 14429 | |
| | | | | Vulcanisaeta moutnovskia 768 28 | Vulcanisaeta moutnovskia 768-28 | |
| Other sulfur species | | | | | | |
| | | Geobacter sulfurreducens | (Caspi et al., 2012) | Geobacter sulfurreducens KN400 | Geobacter sulfurreducens PCA | |
| | | | | Geobacter sulfurreducens PCA | | |
| | | Ruegeria pomeroyi | (Bullock et al., 2014; Todd et al., 2011) | Ruegeria pomeroyi DSS 3 | Ruegeria pomeroyi DSS-3 | |
| | | Starkeya novella | (Benning, 1998; Kappler et al., 2004) | Starkeya novella DSM 506 | Starkeya novella DSM 506 | |

| | Guild | Genus | Reference | Redundant genomes (NCBI-2014) | Non-Redundant genomes (NCBI-2014) | Non-Redundant genomes Refseq 2016 |
|---|---|---|---|---|---|---|
| | | Wolinella succinogenes | (Canfield et al., 2005) | Wolinella succinogenes DSM 1740 | Wolinella succinogenes DSM 1740 | |
| | | Candidatus Desulforudis adaxviator MP104C | (Canfield et al., 2010) | Candidatus Desulforudis audaxviator MP104C | Candidatus Desulforudis audaxviator MP104C | |

## References


Auernik, K. S., Maezato, Y., Blum, P. H., and Kelly, R. M. (2008). The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon Metallosphaera sedula provides insights into bioleaching-associated metabolism. Appl. Environ. Microbiol. 74, 682–692. doi:10.1128/AEM.02019-07.

Benning, C. (1998). Biosynthesis and Function of the Sulfolipid Sulfoquinovosyl Diacylglycerol. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49, 53–75. doi:10.1146/annurev.arplant.49.1.53.

Bullock, H. a, Reisch, C. R., Burns, A. S., Moran, M. A., and Whitman, W. B. (2014). Regulatory and functional diversity of methylmercaptopropionate coenzyme A ligases from the dimethylsulfoniopropionate demethylation pathway in Ruegeria pomeroyi DSS-3 and other proteobacteria. J. Bacteriol. 196, 1275–85. doi:10.1128/JB.00026-14.

Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E. F., et al. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. Science 330, 1375–1378. doi:10.1126/science.1196889.

Canfield, D., Kristensen, E., and Bo, T. (2005). The Sulfur cycle. 1st ed. , eds. D. Canfield, E. Kristensen, and Thamdrup Bo Elsevier Academic Press.

Caspi, R., Altman, T., Dreher, K., Fulcher, C. a, Subhraveti, P., Keseler, I. M., et al. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40, D742--53. doi:10.1093/nar/gkr1014.

Ce, P., Cro, M., St, B., Sea, B., Sievert, S. M., and Kiene, P. (2007). The Sulfur Cycle. 117–123.

Chen, L., Ren, Y., Lin, J., Liu, X., Pang, X., and Lin, J. (2012). Acidithiobacillus caldus sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. PLoS One 7, e39470. doi:10.1371/journal.pone.0039470.

Curson, A. R. J., Rogers, R., Todd, J. D., Brearley, C. A., and Johnston, A. W. B. (2008). Molecular genetic analysis of a dimethylsulfoniopropionate lyase that liberates the climate-changing gas dimethylsulfide in several marine α-proteobacteria and Rhodobacter sphaeroides. Environ. Microbiol. 10, 757–767. doi:10.1111/j.1462-2920.2007.01499.x.

Foti, M., Ma, S., Sorokin, D. Y., Rademaker, J. L. W., Kuenen, J. G., and Muyzer, G. (2006). Genetic diversity and biogeography of haloalkaliphilic sulphur-oxidizing bacteria belonging to the

genus Thioalkalivibrio. FEMS Microbiol. Ecol. 56, 95–101. doi:10.1111/j.1574-6941.2006.00068.x.

Frigaard, N.-U., and Bryant, D. a (2004). Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. Arch. Microbiol. 182, 265–276. doi:10.1007/s00203-004-0718-9.

Frigaard, N.-U., and Dahl, C. (2009). Sulfur metabolism in phototrophic sulfur bacteria. doi:10.1016/S0065-2911(08)00002-7.

Ghosh, W., and Dam, B. (2009). Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol. Rev. 33, 999–1043.

Imhoff, J. F., Süling, J., and Petri, R. (1998). Phylogenetic relationships among the Chromatiaceae, their taxonomic reclassification and description of the new genera Allochromatium, Halochromatium, Isochromatium, Marichromatium, Thiococcus, Thiohalocapsa and Thermochromatium. Int. J. Syst. Bacteriol. 48 Pt 4, 1129–1143. doi:10.1099/00207713-48-4-1129.

Kappler, U., Aguey-Zinsou, K.-F., Hanson, G. R., Bernhardt, P. V, and McEwan, A. G. (2004). Cytochrome c551 from Starkeya novella: characterization, spectroscopic properties, and phylogeny of a diheme protein of the SoxAX family. J. Biol. Chem. 279, 6252–6260. doi:10.1074/jbc.M310644200.

Liljeqvist, M., Valdes, J., Holmes, D. S., and Dopson, M. (2011). Draft genome of the psychrotolerant acidophile Acidithiobacillus ferrivorans SS3. J. Bacteriol. 193, 4304–4305. doi:10.1128/JB.05373-11.

Lindenmayer, D. B., and Likens, G. E. (2010). The science and application of ecological monitoring. Biol. Conserv. 143, 1317–1328. doi:10.1016/j.biocon.2010.02.013.

Moussard, H., L'Haridon, S., Tindall, B. J., Banta, a., Schumann, P., Stackebrandt, E., et al. (2004). Thermodesulfatator indicus gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge. Int. J. Syst. Evol. Microbiol. 54, 227–233. doi:10.1099/ijs.0.02669-0.

Nakagawa, S. (2004). Aeropyrum camini sp. nov., a strictly aerobic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. Int. J. Syst. Evol. Microbiol. 54, 329–335. doi:10.1099/ijs.0.02826-0.

Nunoura, T., Takaki, Y., Kazama, H., Kakuta, J., Shimamura, S., Makita, H., et al. (2014). Physiological and genomic features of a novel sulfur-oxidizing gammaproteobacterium belonging to a previously uncultivated symbiotic lineage isolated from a hydrothermal vent. PLoS One 9. doi:10.1371/journal.pone.0104959.

Pereira Cardoso, I., Ramos, A. R., Grein, F., Marques, M. C., Marques, S., Bryant, D. A., et al. (2011). A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. 2, 1–22. doi:10.3389/fmicb.2011.00069.

Rabus, R., Hansen, T. A., and Widdel, F. (2006). D issi m i latory Sulfate- and Sulfur-Reduc i ng Prokaryote s.

Robertson, L. A., and Kuenen, J. G. (2006). “The Colorless Sulfur Bacteria,” in The Prokaryotes: Volume 2: Ecophysiology and Biochemistry, eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt (New York, NY: Springer New York), 985–1011. doi:10.1007/0-387-30742-7_31.

Schleper, C., Puehler, G., Holz, I., Gambacorta, a., Janekovic, D., Santarius, U., et al. (1995). Picrophilus gen. nov., fam. nov.: A novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. J. Bacteriol. 177, 7050–7059.

Sorokin, D. Y., Lysenko, A. M., Mityushina, L. L., Tourova, T. P., Jones, B. E., Rainey, F. A., et al. (2001). Thioalkalimicrobium aerophilum gen. nov., sp. nov. and Thioalkalimicrobium sibericum sp. nov., and Thioalkalivibrio versutus gen. nov., sp. nov., Thioalkalivibrio nitratis sp. nov. and Thioalkalivibrio denitrificans sp. nov., novel obligately alkaliphilic. Int. J. Syst. Evol. Microbiol. 51, 565–580. doi:10.1099/00207713-51-2-565.

Sorokin, D. Y. U., Banciu, H., Robertson, L. A., and Kuenen, J. G. (2006). Haloalkal i ph i l i c Sulfur-Ox i d i z i ng Bacter i a. 969–984.

Tang, K. H., Feng, X., Zhuang, W. Q., Alvarez-Cohen, L., Blankenship, R. E., and Tang, Y. J. (2010). Carbon flow of heliobacteria is related more to Clostridia than to the green sulfur bacteria. J. Biol. Chem. 285, 35104–35112. doi:10.1074/jbc.M110.163303.

Todd, J. D., Curson, A. R. J., Kirkwood, M., Sullivan, M. J., Green, R. T., and Johnston, A. W. B. (2011). DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. Environ. Microbiol. 13, 427–38. doi:10.1111/j.1462-2920.2010.02348.x.

Valdes, J., Quatrini, R., Hallberg, K., Dopson, M., Valenzuela, P. D. T., and Holmes, D. S. (2009). Draft genome sequence of the extremely acidophilic bacterium Acidithiobacillus caldus ATCC 51756 reveals metabolic versatility in the genus Acidithiobacillus. J. Bacteriol. 191, 5877–5878. doi:10.1128/JB.00843-09.

# Anexo 2

## Reseña histórica del estudio de tapetes microbianos y estromatolitos

| DESCRIPCIÓN | TIPO DE ESTUDIO | REFERENCIA |
|---|---|---|
| Mediciones de ácido sulfhídrico y oxígeno en tapetes microbianos. | Mediciones de parámetros geoquímicos | (Krumbein et al., 1991) |
| Medición de la actividad de nitrogenasa por métodos químicos, combinado con microscopia de contraste de fase en tapetes microbianos bajo diferentes condiciones. | Clasificación de las cianobacterias de acuerdo a sus características morfológicas. | (Stal et al., 1984) |
| Análisis fisicoquímicos combinados con microscopia para seguir la dinámica de crecimiento poblacional de cianobacterias en tres años sucesivos. | Clasificación de las cianobacterias de acuerdo a su morfología | (Stal et al., 1985) |
| Mediciones de sulfato reducción combinados con microelectrodos en tapetes microbianos de cianobacterias | Mediciones de parámetros geoquímicos. Ciclos diurnos de sulfato reducción en invierno y verano. | (Fründ and Cohen, 1992) |
| Mediciones de sulfato reducción combinado con mediciones de cuentas del número más probable para determinar la composición de bacterias sulfato reductoras. | Medición de parámetros geoquímicos combinados con diluciones seriales en un medio selectivo para bacterias sulfato reductoras | (Visscher et al., 1992) |
| Determinación de la abundancia y distribución de bacterias sulfato reductoras en un tapete microbiano hidrotermal utilizando sondas de hibridación complementarias a 16S rRNA. | Sondas de hibridación filogenéticas para la detección de grupos taxonómicos. | (Risatti et al., 1994). |
| Estimación de la diversidad y la estructura de la comunidad en un tapete microbiano hipersalino utilizando análisis. | Análisis taxonómico basado en librerías de 16s RNA y TRFLPs. | (Moyer et al., 1994) |
| Análisis de la diversidad bacteriana en un tapete microbiano hipersalino utilizando librerías de clonas de 16S. | Análisis taxonómico basado en librerías de 16s RNA y TRFLPs y secuenciación. | (Moyer et al., 1995) |
| Evaluación de la relación entre la zona oxigénica y anoxigénica, y fijación de nitrógeno en un tapete microbiano hipersalino en las Bahamas, utilizando análisis de fotopigmentos, microscópica cualitativa, tasas de consumo y producción y fijación de oxígeno, y actividad de nitrogenasa | Los fototrófos se determinaron por análisis de fotopigmentos y microscopia cuantitativa. (*M. chthonoplastes* y *Chromatium*) | (Pinckney and Paerl, 1997) |
| Evaluación de la presencia de bacterias sulfato reductoras con la disponibilidad de oxigeno en un tapete hipersalino de Sinai, Egipto. | Distribución de secuencias tipo DSR en intervalos óxicos y anóxicos del tapete microbiano | (Minz *et al.*, 1999) |
| Uno de los primeros trabajos en combinar microscopia con herramientas moleculares para analizar la diversidad de cianobacterias en un tapete microbiano de Lago Fryxell Antartica. | Análisis taxonómico y filogenético utilizando librerías de clonas del gen 16S rRNA, DGGA y secuenciación. | (Taton *et al.*, 2003) |
| Combinación de mediciones de microsensores *in situ* s (O, $H_2S$ and pH) junto con métodos tradicionales biogeoquímicos para determinar el ciclaje de carbono, oxigeno, azufre en un tapete hipersalino de Giraud, Francia | Mediciones de parámetros geoquímicos. | (Wieland et al., 2005) |
| Primer estudio integral en estudiar el papel del sulfato reducción en combinación con técnicas radiométricas, perfiles químicos y análisis moleculares filogenéticos. | Detección de sulfato reductor utilizando los primers DSR1f y DSR4r. Se analizaron 50 clonas. Solo se obtuvo un filotipo de dsrAB relacionado con el clado *Thermodesulfovibrio T. yellowstonii* y *T. islandicus* | (Dillon et al., 2007b) |
| Uno de los primeros estudios en utilizar tecnologías de secuenciación en tapetes microbianos para obtener la diversidad de cianobacterias mediante Genómica y Metagenómica comparativa. Secuenciación de genomas obtenidos de cultivos a partir de muestras de tapetes microbianas, además de la secuenciación total del DNA | Identificación de genes específicos (cluster de utilización de urea) de aislados de *Synecochoccus* obtenidos de la capa superior del tapete microbiano | (Bhaya *et al.*, 2007) |

| DESCRIPCIÓN | TIPO DE ESTUDIO | REFERENCIA |
|---|---|---|
| aislado de la capa fotosintética del tapete microbiano en La Poza Octopus en Yellowstone. | | |
| Combinación de microscopia de fluorescencia, hibridación *in situ*, mediciones geoquímicas y composición de comunidad mediante técnicas moleculares para investigar la biogeoquímica de dos tipos de tapetes microbianos. | Análisis taxonómico basado en librerías de clonas de 16 s RNA, la mayoría de las secuencias son cercanas a *Desulfobacteraceae, Desulfobulbaceae, Methylobacter marinus, Arcobacter sulfidicus and Sulfurimonas autotrophica, Sulfospirillum arcachonense.* | (Omoregie *et al.*, 2008) |
| Uno de los primeros estudios en utilizar técnicas isotópicas y análisis metagenómicos para determinar las comunidades microbianas y procesos químicos de dos microbialitos de Cuatro Ciénegas Coahuila México. | Análisis funcional de genes unos cuantos genes involucrados en el ciclaje de nutrientes (C,N,P,S) en microbialitos | (Breitbart et al., 2009b; Desnues et al., 2008b) |
| Estudio donde se combina microscopia, determinación isotópica de lípidos, alcoholes, e hidrocarbonos para caracterizar los posibles biomarcadores para identificar grupos microbianos activos en tapetes microbianos de Las Islas de la Republica de Kiribati. | Identificación e inferencia de cianobacterias, purpuraras del azufre y sulfato reductoras basado en microscopia y análisis órgano-geoquímicos. | (Bühring *et al.*, 2009) |
| Análisis integral de la comunidad microbiana dentro de las ventilas hidrotermales. Combinación de perfil químico y análisis filogenéticos moleculares | Aunque no es un estudio de microbialitos, es un ejemplo claro del uso de marcadores moleculares para dilucidar vías metabólicas clave en el ciclaje del carbón y azufre en ventilas hidrotermales, complementado con análisis taxonómico de gen 16s RNA p | (Hügler et al., 2010) |
| Primer estudio en medir la actividad de nitrogenasa y la expresión de *nifH* para elucidar los organismos responsables de la fijación del nitrógeno en tapetes microbianos. | La actividad de la nitrogenasa fue medida por el método de reducción de acetileno y se realizaron mediciones de los patrones de expresión de cinco organismos diazótrofos del tapete microbiano. | (Severin, Lucas, 2007) |
| Ejemplo de un estudio que utiliza marcadores moleculares para obtener información acerca de los organismos diazótrofos de los tapetes microbianos. | Análisis taxonómico y funcional para demostrar la relación entre la relación entre la estructura de la comunidad y el funcionamiento del ecosistema utilizando DGGE de 16s RNA y librerías de clonas de nifH. | (Severin *et al.*, 2010) |
| Revisión de las bacterias purpuras del azufre en los tapetes microbianos y microbialitos. | Combinación de análisis taxonómico utilizando librerías de clonas de 16s RNA y patrones de DGGE, con librerías de clonas de *primers* específicos del centro de reacción como es *pufM*. | (Hubas et al., 2011a) |
| Análisis metagenómico combinado con perfiles de utilización de sustrato utilizando microarreglos metabólicos en los tapetes microbianos de Highborne Cay, Bahamas. | Relación entre los patrones de utilización de sustratos (carbono, nitrógeno, fosforo y azufre) con análisis taxonómico y funcional de tapetes microbianos. | (Khodadad and Foster, 2012b) |
| Análisis metagenómico de tapetes microbianos de los hielos de Ártica y Antártica. | Búsqueda de genes funcionales de respuesta a estrés ambiental como (codificantes de exopolisacaridos, proteínas de respuesta al frio y modificaciones de membrana. | (Varin et al., 2012) |
| Análisis de Metagenomica comparativa junto con librerías de clonas de 16S para obtener los perfiles taxonómicos y funcionales de dos tapetes microbianos de Cuatro Ciénegas Coahuila, México. | Búsqueda de marcadores moleculares en metagenomas de dos tapetes microbianos para entender el papel de ciclos biogeoquímicos (utilización de fosfonato (*phnD, phnH, htxB*, y *ptxB*), metabolismo de polifosfatos ( *ppA, ppK*, y *ppX*), reciclaje de fosfato (*phoA, phoX*, y *pstS)*, utilización de fosfato y fosfito (*ptxB* y *phnH*), biosíntesis de sulfolípidos *sqdB* and *sqdX,* fijación de carbono( *rbcL, rbcS, codH*, y *aclY*,) fijación de nitrogeno (*nifH, nrfA, narG, napA, narB, nirS, norB, and nosZ)* | (Bonilla-Rosso et al., 2012; Peimbert et al., 2012) |
| Primer estudio en investigar el seguimiento de la interacción metabólica de azufre a escala microbiana en biofilms de color púrpura utilizando una combinación de secuenciación Metagenomica, análisis taxonómico de 16S rRNA, microscopia, e hibridación *in situ* así como una combinación de análisis geo microbiológicos isótopos estables | Análisis taxonómico y funcional de genes específicos del metabolismo del azufre (*dsrAB, aprAB*) | (Wilbanks et al., 2014) |

## References


Bhaya, D., Grossman, A. R., Steunou, A.-S., Khuri, N., Cohan, F. M., Hamamura, N., et al. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J. 1, 703–13. doi:10.1038/ismej.2007.46.

Bonilla-Rosso, G., Peimbert, M., Alcaraz, L. D., Hernández, I., Eguiarte, L. E., Olmedo-Alvarez, G., et al. (2012). Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: community structure and composition in oligotrophic environments. Astrobiology 12, 659–73. doi:10.1089/ast.2011.0724.

Breitbart, M., Hoare, A., Nitti, A., Siefert, J., Haynes, M., Dinsdale, E., et al. (2009). Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. Environ. Microbiol. 11, 16–34. doi:10.1111/j.1462-2920.2008.01725.x.

Bühring, S. I., Smittenberg, R. H., Sachse, D., Lipp, J. S., Golubic, S., Sachs, J. P., et al. (2009). A hypersaline microbial mat from the Pacific Atoll Kiritimati: insights into composition and carbon fixation using biomarker analyses and a 13C-labeling approach. Geobiology 7, 308–23. doi:10.1111/j.1472-4669.2009.00198.x.

Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008). Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452, 340–343. doi:10.1038/nature06735.

Dillon, J. G., Fishbain, S., Miller, S. R., Bebout, B. M., Habicht, K. S., Webb, S. M., et al. (2007). High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. Appl. Environ. Microbiol. 73, 5218–26. doi:10.1128/AEM.00357-07.

Fründ, C., and Cohen, Y. (1992). Diurnal Cycles of Sulfate Reduction under Oxic Conditions in Cyanobacterial Mats. Appl. Environ. Microbiol. 58, 70–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=195174&tool=pmcentrez&rendertype=abstract.

Hubas, C., Jesus, B., Passarelli, C., and Jeanthon, C. (2011). Tools providing new insight into coastal anoxygenic purple bacterial mats: review and perspectives. Res. Microbiol. 162, 858–68. doi:10.1016/j.resmic.2011.03.010.

Hügler, M., Gärtner, A., and Imhoff, J. F. (2010). Functional genes as markers for sulfur cycling and CO2 fixation in microbial communities of hydrothermal vents of the Logatchev field. FEMS Microbiol. Ecol. 73, 526–37. doi:10.1111/j.1574-6941.2010.00919.x.

Khodadad, C. L. M., and Foster, J. S. (2012). Metagenomic and metabolic profiling of nonlithifying and lithifying stromatolitic mats of Highborne Cay, The Bahamas. PLoS One 7, e38229. doi:10.1371/journal.pone.0038229.

Krumbein, W. E., Buchholz, H., Franke, P., Giani, D., Giele, C., and Wonneberger, K. (1991). O2 and H2S coexistence in stromatolites. Science (80-. ). 66, 381–389. doi:10.1007/BF00368068.

Minz, D., Fishbain, S., Green, S. J., Muyzer, G., Cohen, Y., Rittmann, B. E., et al. (1999). Unexpected population distribution in a microbial mat community: sulfate-reducing bacteria localized to the highly oxic chemocline in contrast to a eukaryotic preference for anoxia. Appl. Environ. Microbiol. 65, 4659–65. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=91621&tool=pmcentrez&rendertype=abstract.

Moyer, C. L., Dobbs, F. C., and Karl, D. M. (1994). Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. 60, 871–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=201404&tool=pmcentrez&rendertype=abstract.

Moyer, C. L., Dobbs, F. C., and Karl, D. M. (1995). Phylogenetic diversity of the bacterial community from a microbial mat at an active , hydrothermal vent system , Loihi Seamount , Phylogenetic Diversity of the Bacterial Community from a Microbial Mat at an Active , Hydrothermal Vent System , Loihi Seamou.

Omoregie, E. O., Mastalerz, V., de Lange, G., Straub, K. L., Kappler, A., Røy, H., et al. (2008). Biogeochemistry and community composition of iron- and sulfur-precipitating microbial mats at the Chefren mud volcano (Nile Deep Sea Fan, Eastern Mediterranean). Appl. Environ. Microbiol. 74, 3198–215. doi:10.1128/AEM.01751-07.

Peimbert, M., Alcaraz, L. D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L., et al. (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, 648–58. doi:10.1089/ast.2011.0694.

Pinckney, J. L., and Paerl, H. W. (1997). Anoxygenic photosynthesis and nitrogen fixation by a microbial mat community in a bahamian hypersaline lagoon. Appl. Environ. Microbiol. 63, 420–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1389512&tool=pmcentrez&rendertype=abstract.

Risatti, J. B., Capman, W. C., and Stahl, D. a (1994). Community structure of a microbial mat: the phylogenetic dimension. Proc. Natl. Acad. Sci. U. S. A. 91, 10173–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=44980&tool=pmcentrez&rendertype=abstract.

Severin, Lucas, S. (2007). NifH expression by five groups of phototrophs compared with nitrogenase activity in coastal microbial mats. FEMS Microbiol. Ecol. 73, 55–67. doi:10.1111/j.1574-6941.2010.00875.x.

Severin, I., Acinas, S., and Stal, L. (2010). Diversity of nitrogen-fixing bacteria in cyanobacterial mats. FEMS Microbiol. Ecol. 73, 514–25. doi:10.1111/j.1574-6941.2010.00925.x.

Stal, L. J., Gemerden, H. Van, and Krumbein, W. E. (1985). Structure and development of a benthic marine microbial mat. 31, 111–125.

Stal, L. J., Grossberger, S., Krumbein, W. E., and Division, G. (1984). Nitrogen fixation associated with the cyanobacterial mat of a marine laminated microbial ecosystem. 224, 217–224.

Taton, A., Grubisic, S., Brambilla, E., Wit, D., Wilmotte, A., and Wit, R. De (2003). Cyanobacterial Diversity in Natural and Artificial Microbial Mats of Lake Fryxell ( McMurdo Dry Valleys , Antarctica ): a Morphological and Molecular Approach Cyanobacterial Diversity in Natural and Artificial Microbial Mats of Lake Fryxell ( McMurdo Dry . doi:10.1128/AEM.69.9.5157.

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., and Corbeil, J. (2012). Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the High Arctic. Appl. Environ. Microbiol. 78, 549–559. doi:10.1128/AEM.06354-11.

Visscher, P. T., Prins, R. A., and Gemerden, H. Van (1992). Rates of sulfate reduction and thiosulfate consumption in a marine microbial mat. 86.

Wieland, A, Zopfi, J., Benthien, M., and Kühl, M. (2005). Biogeochemistry of an iron-rich hypersaline microbial mat (Camargue, France). Microb. Ecol. 49, 34–49. doi:10.1007/s00248-003-2033-4.

Wilbanks, E. G., Jaekel, U., Salman, V., Humphrey, P. T., Eisen, J. a, Facciotti, M. T., et al. (2014). Microscale sulfur cycling in the phototrophic pink berry consortia of the Sippewissett Salt Marsh. Environ. Microbiol. 16, 3398–3415. doi:10.1111/1462-2920.12388.

# Anexo 3

## Protocolo de extracción de DNA

El siguiente protocolo es una adaptación del método descrito en Purdy (2005) para obtener DNA de calidad para secuenciación en muestras de suelo, sedimento y tapetes microbianos de de Cuatro Cienegas. Para cualquier adaptación al método, ver la publicación original.

### 1. SOLUCIONES Y REACTIVOS

Antes de iniciar el protocolo, revisar que se tengan las siguientes SOLUCIONES, en caso contrario ir al anexo de preparación de soluciones.

1. S- Bio-Gel® HTP Hydroxyapatite, se almacena a 4°C*
2. S-TA Sephadex *
3. S-TA Fosfato de sodio 120mM pH 8.1, 1% (p/v) PVPP*
4. S-TA Fosfato de sodio 120mM pH 8*
5. S-TA Fosfato de sodio 120mM pH 7.2*
6. S- Tris-fenol equilibrado pH 8 (se ajusta con hidroxiquinolona y se almacena a 4°C)
7. R-TA SDS (20%) pH 8
8. S-TA Fosfato dipotásico 300mM pH 7.2*
9. S-TA Acetato de sodio 3M*
10. R- Alcohol 100 % se almacena a -20 °C
11. R- Alcohol 70% se almacena a -20°C
12. R-TA Agua libre de RNAsas*

TA almacenado a temperatura ambiente
S solución
R reactivo
* esterilizados en autoclave para cada ronda de extracción.

## 2. MATERIAL ESTÉRIL

Todos los siguientes materiales deben de estar esterilizados en autoclave con papel aluminio en cada ronda de extracción

1. Jeringas de 1 mL (deben ser nuevas: ya esterilizadas y empacadas)
2. Jeringas de 3 mL (deben ser nuevas: ya esterilizadas y empacadas)
3. Tubos falcón de 50 mL
4. Tapas perforadas para tubos falcón de 50 mL, con entrada para jeringa de 1 mL y de 3 mL
5. Pinzas metálicas
6. Puntas azules con puntas recortadas
7. Glass beads (perlitas)
8. Tubos eppendorf de 2 mL
9. Tubos eppendorf de 1.5 mL
10. Espátulas
11. Fibra de vidrio

## 3. MATERIAL Y EQUIPO ADICIONAL

1. Papel aluminio
2. Gradillas
3. Mascarilla o tapabocas
4. Guantes
5. Bata de laboratorio
6. Campana de flujo laminar
7. Campana de extracción
8. Tissue liser (QIAGEN. Model: TissueLyses LT. SN 23.1001/05431. Capacidad: 12)
9. Centrifuga con refrigeración para tubos falcón 50 mL(HERMLE Labortechnik Gmbh; Type: z 326 K; Max Drehzahal: 18000 1/min) Rotor (HERMLE 221.18; Max: 12000rpm; capacidad: 6)
10. Centrifuga para tubos de 1.5 mL
11. Cronometro
12. Hielo

## 4. PREPARACION DE SOLUCIONES

**HTP (Hidroxiapatita) para columna:** Utilizando una espátula estéril (NOTA: es mejor servir del recipiente y no usar la espátula para evitar contaminar el reactivo) poner 10 g de HTP en frasco estéril de 50 mL. Agregar ~50mL de agua libre de RNAsas y mezclar suavemente. Dejar reposar por 10 minutos (puede contener partículas suspendidas de HTP). Esterilizar y una vez frio almacenar a 4°C

**Sephadex para columna:** Utilizando una espátula estéril (NOTA: es mejor servir del recipiente y no usar la espátula para evitar contaminar el reactivo) poner 10 g en un frasco de 250 mL estéril y diluir en 200 mL de agua libre de RNAsas. Esterilizar la solución en autoclave.

**Soluciones de Fosfato de sodio: STOCKS.**

A) **Solución fosfato disodico 120 mM** ($Na_2HPO_4$): A 300mL de agua libre de RNAsas, agregar 6.8 g de $Na_2HPO_4$, aforar a 400 mL con agua libre de RNAsas y esterilizar en autoclave.

B) **Solución fosfato de sodio monosodico 120 mM** ($NaH_2PO_4$-$H_2O$): A 300mL de agua libre de RNAsas añadir 6.6 g de $NaH_2PO_4$, aforar a 400mL con agua libre de RNAsas y esterilizar en autoclave.

| *pH* | *mL de 120 mM $Na_2HPO_4$* | *mL de 120 mM $NaH_2PO_4$* |
|---|---|---|
| *5.8* | 8 | 92 |
| *6* | 12.3 | 87.7 |
| *6.2* | 18.5 | 81.5 |
| *6.4* | 26.5 | 73.5 |
| ***6.6*** | **37.5** | **62.5** |
| ***6.8*** | **49** | **51** |
| *7* | 61 | 39 |
| ***7.2*** | **72** | **28** |
| *7.4* | 81 | 19 |
| *7.6* | 87 | 13 |
| *7.8* | 91.5 | 8.5 |
| ***8*** | **94.7** | **5.3** |

## Soluciones de Fosfato de sodio:

**Fosfato de sodio 120 mM pH 7.2**.: A 72mL del stock (**A**) añadir 28mL del stock (**B**) hasta que el pH sea de 7.2.

**Fosfato de sodio 120 mM pH 8.0 + 1% PVPP**: A 95mL del stock (**A**) añadir 5mL del stock (**B**) hasta que el pH sea 8. Añadir 1gr (1% w/w) de PVPP

**Fosfato de sodio 120 mM pH 8**.: A 95mL del stock (**A**) agregar 5mL del stock (**B**) hasta que el pH sea **8.0**

**Fosfato dipotásico 300 mM pH 7.2:** A 80 mL de agua libre de RNAsas añadir 5.2 g de $K_2HPO_4$. Para ajustar el pH utilizar HCl 6 M. Aforar a 100 mL y manejar alícuotas pequeñas.

## Tris-fenol equilibrado pH 8: STOCKS

**Tris 1 M pH 8**: A 300 mL de agua libre de RNAsas añadir 48.4 g de Tris-HCl. Ajustar el pH con HCl 6 M y aforar a 400 mL. Esterilizar en autoclave.

**Tris 0.1 M pH 8**: A 360 mL de agua libre de RNAsas añadir 40 mL de Tris 1 M pH 8 dejar reposar todo un día a 4°C.

**Solución:** Añadir 0.1 g de 8-hidroxiquinolona (con mucho cuidado porque es muy tóxica, usar guantes, y cubre boca). Verificar el pH colocando una gota en papel indicador. Guardar en la oscuridad a 4ºC.

**SDS (20%) pH 8. NO ESTERILIZAR**. Tomar alícuotas pequeñas para manipular durante las extracciones.

**Acetato de sodio 3 M:** Pesar 18 g de acetato de sodio y diluir en 100 mL de agua libre de RNAsas

NOTA: Verificar que las muestras (suelo, sedimiento, etc.) estén descongeladas, antes de comenzar con los pasos del protocolo y mantener en hielo hasta el momento de su uso.

Para las muestras de tapete microbiano no se recomienda descongelar antes debido a que puede causar problemas de degradación de DNA.

## 5. PROCEDIMIENTO

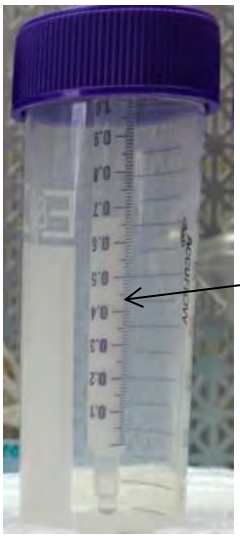
### Preparación de las columnas

1. Es necesario tener guantes, bata y mascarilla durante todo el procedimiento de extracción. Tambien es importante limpiar el área de trabajo con alcohol y sanitas.
2. Rellenar las jeringas de 1 mL y 3 mL con fibra de vidrio. Con ayuda de dos pinzas estériles se enrolla una sobre otra una sección larga de fibra de vidrio. se introduce el rollo en la jeringa con cuidado. Si es complicado hacer los rollos se puede introducir en pedazos muy pequeños. No introducir fibra de vidrio que se haya caído o tocado otra superficie.
3. Comprimir la fibra de vidrio con ayuda del embolo, cuidando de no hacer mucha presión. El tapón de fibra de vidrio debe llegar a un volumen de 0.1 mL (en las jeringas de 1 mL) y de 0.5 mL (para las de 3 mL).
4. Dejar el embolo dentro de la jeringa hasta realizar la extracción. Cubrir las jeringas con aluminio para evitar contaminación.

*La fibra de vidrio retiene las partículas en suspensión para preparar las columnas, y permite que el líquido fluya*

## Columnas de HTP

5. Se utilizan las **jeringas de 1 mL**.
6. La solución de HTP refrigerada a **4ºC** se agita suavemente para resuspender las partículas.
7. Se cargan las jeringas con ayuda de las **puntas recortadas** (azules) tomando la interfase de la solución, se debe tener cuidado de ir **gota a gota** ya que se puede tapar la jeringa e impedir que siga bajando el liquido.
8. Centrifugar a **1500 rpm/ 1 min** para que descienda la solución de HTP. El volumen final en la jeringa debe ser de **0.6 a 0.7 mL de HTP**, identificándolo fácilmente por su color blanco.
9. Rellenar y centrifugar las veces que sea necesario hasta alcanzar el volumen final.
10. Cuando estén listas se cubren con papel aluminio.




*¿Para qué sirve? La hidroxiapatita (HTP) permite separar diferencialmente proteínas y ácidos nucléicos, por lo que ésta retendrá los ácidos mientras las proteínas serán removidas al lavar la columna.*

## Columnas de Sephadex

11. Se utilizan las **jeringas 3 mL**.
12. La solución de sephadex tiene **dos fases** por lo cual es necesario agitarla suavemente y tomar en la interfase la mayor cantidad de liquido.
13. El llenado de las jeringas se realiza en la campana de flujo laminar.
14. Se llena la jeringa **hasta el tope** con el sephadex y después se **centrifuga a 6000 rpm/ 2 minutos**.
15. Este paso se repite hasta que el **volumen final** de la columna sea de **2 mL**.

16. Una vez alcanzado el volumen final, se procede a **acondicionar la columna**. Se agrega **0.4 mL de $H_2O$ libre de RNAsas** y después, se centrifuga a **6000 rpm/ 8 minutos**.
17. El paso anterior se repite **3 veces**.
18. Al terminar, cubrir con aluminio.

NOTA OPCIONAL

Al tener lista las columnas de Sephadex se recomienda montar el soporte y el tubo epperdorf de 1.5 mL tal como se muestra en el esquema de abajo; antes de iniciar con la extracción de DNA. Lo anterior debido a que cuando se llegue al paso en el cual se utilizan las columnas, se continúe el procedimiento sin detenerse.

*El sephadex es un gel que retiene moléculas de bajo peso molecular, por lo que remueve los ácidos húmicos y las sales de las soluciones de lisis y de lavado.*

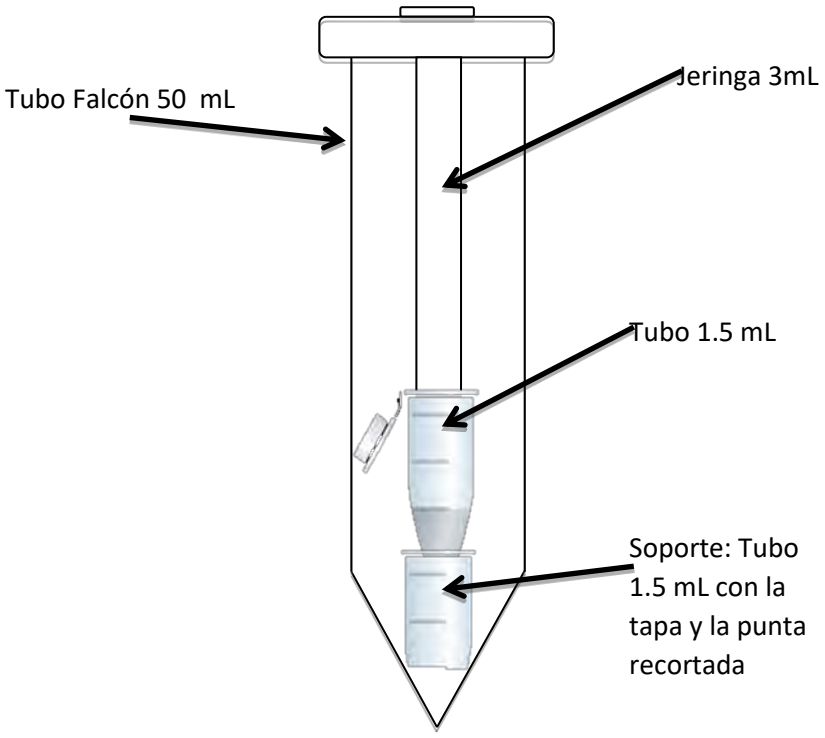



Imagen donada por Laura Méndez que optimizó el método para el desarrollo de su tesis "El papel de los microorganismos en solubilizadores de fósforo en Cuatro Ciénagas, Coahuila" https://www.uv.mx/dceb/generaciones/alumnos/gen12/s12015414/

Lisis de las muestras

1. En un tubo de 2 mL agregar aproximadamente 0.5 mL de glass beads (perlitas).
2. Agregar con ayuda de una espátula el mismo volumen de muestra (0.5 mL).

NOTA: Mantener los tubos con muestra en hielo en todo momento de la lisis y extracción.

*Las perlitas realizarán una lisis mecánica a todas las células en la muestra.*

3. Agregar en el tubo de 2 mL las siguientes soluciones de lisis en el orden en que aparecen:
   - 600 µl de fosfato de sodio 120 mM pH 8, 1% (p/v) PVPP. NOTA: agitar un poco la solución para tomar la mayor cantidad de partículas, tener cuidado porque la punta se tapa.
   - 400 µl de Tris-fenol equilibrado pH 8. NOTA: esta solución presenta dos fases y se utiliza la fase inferior, manipular en campana de extracción.
   - 40 µl de SDS (20%) pH 8

*El fosfato de sodio 120 mM pH 8, 1% (p/v) PVPP actúa como un buffer para mantener un pH neutro y al mismo tiempo es la fase acuosa en la que los ácidos nucléicos se mantendrán solubles; el PVPP es un polímero insoluble al cual se adsorben los ácidos húmicos. El fenol permitirá precipitar la mayor cantidad de proteínas y componentes de la membrana celular y el SDS actúa como detergente para romper las membranas celulares.*

4. Agitar los tubos de lisis de 2 mL en el Bead beat durante 30 segundos a máxima velocidad (50) y colocar rápidamente en hielo por 30 segundos. Repetir este paso 3 veces.

*Al agitar las muestras permitiremos que las perlas de vidrio rompan las membranas celulares, de igual forma ayudan a homogeneizar las soluciones en la muestra. Al colocar en hielo se inhibe la actividad enzimática de las DNAsas y RNAsas*

5. Centrifugar los tubos con muestra a 13 000 rpm/1 minuto.

*Al centrifugar se separa la fase acuosa donde en la que se encuentras los ácidos nucléicos, y se diferencian de las otras dos fases (fenol y pellet con partículas de suelo).*

6. Se observan **3 fases** dentro del tubo de 2 mL: **pellet, fenol y fase liquida**. Cuidadosamente colectar la fase liquida y cargar en la columna de HTP. **NOTA:** Si falto colectar parte de la fase liquida, reservar en hielo dentro de un eppedorf nuevo y cargar en la columna tras el paso 7, repitiendo las indicaciones del mismo.
7. Centrifugar la columna de HTP durante a **1500 rpm/ 2 minutos**.
8. En el tubo de 2 mL con la muestra se lava el pellet con **600 µl de fosfato de sodio 120 mM pH 8**, y se agita en vortex hasta que se desprenda la pastilla.
9. Agitar los tubos en el Bead beat durante **30 segundos a máxima velocidad** e inmediatamente colocar en **hielo por 30 segundos** (se repite una sola vez).
10. Centrifugar el tubo de 2 mL a **13 000 rpm/1 minuto**.
11. Colectar la fase liquida como en el paso 6 y centrifugar igual que en el paso 7.

*De esta forma se extraen más de 90% de los ácidos nucléicos extraíbles*

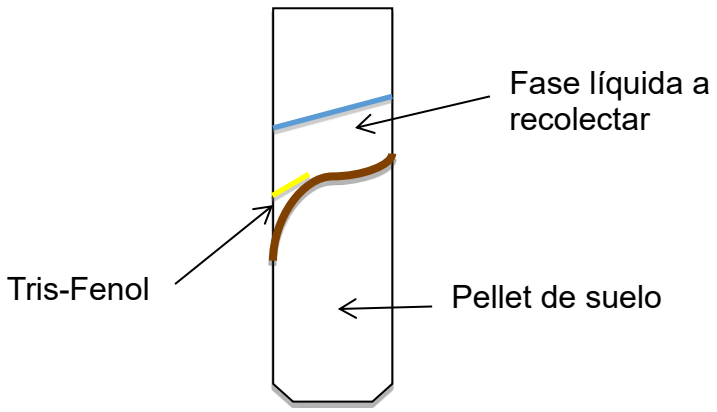



Imagen donada por Laura Méndez que optimizó el método para el desarrollo de su tesis "El papel de los microorganismos en solubilizadores de fósforo en Cuatro Ciénagas, Coahuila" https://www.uv.mx/dceb/generaciones/alumnos/gen12/s12015414/

*Purificación de los ácidos nucléicos totales*

1. Lavar la columna de HTP con **500 µl de fosfato de sodio 120 mM pH 7.2** y centrifugar a **1500 rpm por 2-6 minutos** o el tiempo que sea necesario para que pase todo el liquido por la columna (tapete microbiano: 5-10 minutos; suelo: 5 minutos). **Repetir este paso 3 veces.**

*¿Qué está pasando? Con esto se eluyen todas las proteínas que se encuentran unidas en la columna.*

NOTA. Es necesario cambiar de tubo falcón y montar el tubo eppedorf de 1.5 mL sobre el soporte, como se hizo antes con las columnas de Sephadex.

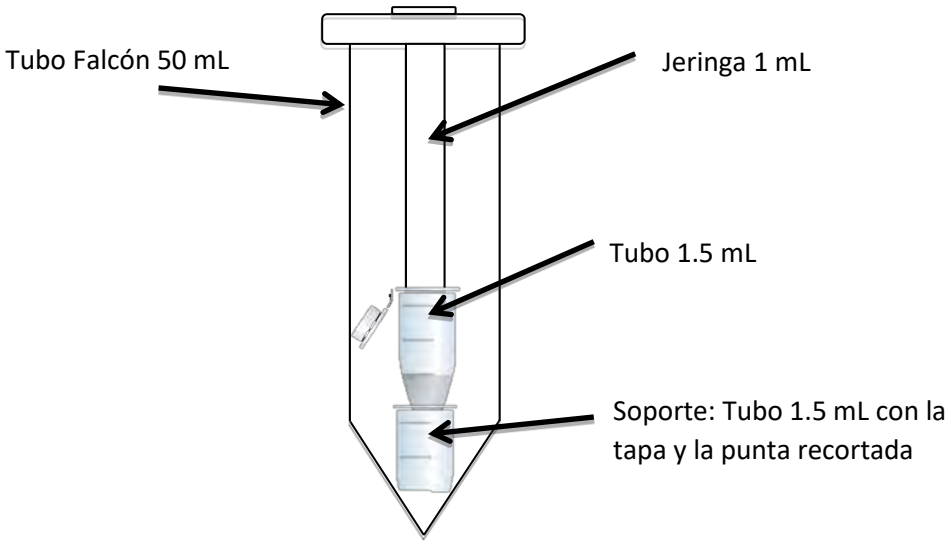



Imagen donada por Laura Méndez que optimizó el método para el desarrollo de su tesis "El papel de los microorganismos en solubilizadores de fósforo en Cuatro Ciénagas, Coahuila" https://www.uv.mx/dceb/generaciones/alumnos/gen12/s12015414/

2. Eluir los ácidos nucléicos de la columna en el tubo eppedorf 1.5 mL, agregando 400 µl de fosfato dipotásico 300 mM pH 7.2.
3. Centrifugar a 1500 rpm/ 2-6 minutos, el tiempo depende de la muestra.
4. Recuperar el tubo de 1.5 mL con el eluente y cargarlo en la columna de Sephadex (ya con el tubo de 1.5 mL y el soporte).

*¿Qué está pasando? En este paso se remueven las moléculas pequeñas que pueden contaminar el DNA, como ácidos húmicos y sales.*

5. Centrifugar el tubo falcón con la columna de Sephadex a **6000 rpm/ 8 minutos**, colectando los ácidos nucleicos en el tubo de 1.5 mL.

6. Agregar al tubo de 1.5 mL: **40 µl de acetato de sodio 3 M y 1 mL de etanol (frio) al 100%**. **NOTA:** si desea ver el DNA teñido, agregar primeramente, 2 µl de Glycoblue

*¿Qué está pasando? El DNA es insoluble en etanol por lo que se precipitará; el acetato de sodio permite que el DNA conserve la estructura de hélice.*

7. Dejar precipitando a -20ºC over night.
8. Centrifugar el tubo de 1.5 mL a **13 000 rpm/ 10 minutos** y decantar el sobrenadante por aspiración.

Agregar **1 mL de etanol (frío) al 70%** y agitar el pellet por vortex a baja velocidad por 10 segundos.

*¿Qué está pasando? Se elimina el etanol y la mayor parte de las sales que aún permanecen en el DNA.*

9. Centrifugar el tubo a **13 000 rpm/ 2 minutos** y remover cuidadosamente el sobrenadante por aspiración.
10. **Evaporar el etanol** en campana de flujo de laminar por varias horas (1-3 horas).

*¿Qué está pasando? En este paso se elimina casi todo el etanol de la muestra de DNA.*

11. Resuspender muy bien el DNA en **20 o 50 µl de $H_2O$ libre de RNAsas**, pipeteando lenta y cuidadosamente.

*¿Qué está pasando? En este paso se eliminan todas las sales que pueden contaminar al DNA, debido a que estas se disuelven en la fracción de agua de la solución de etanol al 70%.*