UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE ECOLOGÍA

Procesos de diversificación y adaptación de la familia Vibrionaceae.

El caso de Cuatro Ciénegas

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

Mirna Vázquez Rosas Landa

TUTORA PRINCIPAL DE TESIS: Dr. Valeria F. E. L. de M. de G. Souza Saldívar
Instituto de Ecología, UNAM
COMITÉ TUTOR: Dr. Ana E. Escalante
Instituto de Ecología, UNAM
COMITÉ TUTOR: Dr. Gloria Soberón Chávez
Instituto de Investigaciones Biomédicas, UNAM

MÉXICO, CIUDAD DE MÉXICO, NOVIEMBRE, 2017.

Agradecimientos

Posgrado en Ciencias Biomédicas.

Beca CONACyT (Becario No. 345250, CVU 473233).

Tutor Principal y miembros del Comité Tutor: Dra. Valeria Souza Saldívar del Instituto de Ecología, Dra. Ana E. Escalante del Instituto de Ecología y la Dra. Gloria Soberón Chavéz del Instituto de Investigaciones Biomédicas.

Agradecimientos a título personal

A la Dra. Valeria Souza Saldívar y al Dr. Luis Enrique Eguiarte Fruns por aceptarme en su equipo de trabajo y darme el apoyo necesario para que este trabajo llegará a término. Agradezco el apoyo técnico proporcionado por la Dra. Laura Espinosa Asuar y la Dra. Erika Aguirre Planter. A los miembros del jurado Dr. Diego Cortez Quezada, Dr. David René Romero Camarena, Dra. Gloria Soberón Chavéz, Dr. Luis David Alcaraz y Dr. Arturo Carlos II Becerra Bracho.

A Gabriel Yaxal Ponce Soto, Silvia Barrientos y Eneas Aguirre von Wobeser por su apoyo y enseñanzas durante el trabajo de campo y de laboratorio. A los miembros presentes y pasados del Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología de la UNAM, por su compañía, por las risas y por las pequeñas/grandes contribuciones que se vieron reflejadas en el trabajo de laboratorio, bioinformático y de escritura.

A Gabriel Yaxal Ponce Soto y Valerie de Anda Torres por su amistad, apoyo, compañía y por los talleres de bioinformática, que no solo nos dejaron muchas enseñanzas sino que constituyen una prueba de lo que podemos lograr juntos.

A los colaboradores y amigos. Al Dr. Enrique Ibarra Laclette en el Instituto de Ecología A.C. por su apoyo en el análisis bioinformático. Al Dr. David S. Guttman en la Universidad de Toronto por permitirme formar parte de su grupo de trabajo. A Shalabh Thakur por su valiosa amistad, por su ayuda durante el desarrollo de los análisis bioinformáticos y la discusión. Pero sobre todo por los momentos tan geniales que pasamos en Toronto.

Agradezco a la Dra. Gabriela Olmedo por su amistad, apoyo y por haberme abierto la puerta a este grupo de trabajo.

A mi familia, a mis padres Mirna Rosas Landa Castañeda y Aurelio Vázquez Bañuelos quienes siempre creyeron en mí, y fueron/son un apoyo y una motivación para seguir. A mis hermanas Monica Vázquez Rosas Landa y Mariana Vázquez Rosas Landa quienes por muy difícil que fuera el panorama no dejaron que abandonara mi camino, esto es gracias a ellas. A todos mis tíos en general, a mis abuelitos y principalmente a Rigoberto Alfaro Urbano y Leticia Rosas Landa Castañeda quienes con su apoyo me permitieron terminar esta etapa.

A mis profesores Dr. Luis Eguiarte, Dra. Valeria Souza, Dr. Luis David Alcaraz y Dr. Pablo Vinuesa, quienes con sus enseñanzas me permitieron culminar este trabajo. A mi comité turoral Dra. Valeria Souza, Dra. Ana E. Escalante y Dra. Gloria Soberon Chavez cuyas aportaciones fueron valiosas durante el desarrollo de este trabajo.

A la Dra. Valeria Souza, por haber confiado en mí y haberme dado la oportunidad de dar lo mejor.

A Gabriel Yaxal Ponce Soto, gracias por todo, por la amistad, por el apoyo técnico, por la paciencia. Fue un placer hacer equipo contigo.

Índice

Resumen

Una de las preguntas centrales de la biología es entender cómo se generan las especies. Esta pregunta resulta interesante en el contexto de la microbiología actual donde los avances tecnológicos han permitido develar una alta diversidad de especies bacterianas y de genes asociados a cada especie. Este hecho hace necesario una revisión de conceptos y metodologías referentes a la delimitación de especies en bacterias, así como una revisión de los factores evolutivos que moldean la diversidad de genes dentro de cada grupo. En este trabajo se utilizó una aproximación de genómica comparada para revisar y evaluar: *i)* cómo se definen las especies bacterianas, con el fin de entender los límites de la clasificación taxonómica y el impacto que tiene en las mediciones de riqueza y abundancia, *ii)* los factores evolutivos que moldean el pan-genoma y *iii)* la movilidad genética, a través de los mecanismos asociados a la patogénesis. Estos puntos se abordaron estudiando aislados bacterianos de la familia Vibrionaceae provenientes de Cuatro Ciénegas, Coahuila, un sitio que se caracteriza por su alta diversidad microbiana y baja perturbación humana. Nuestros resultados muestran que la clasificación tradicional basada en un valor de corte en un gen dado, no permite diferenciar especies que han divergido recientemente y mantienen poca variación dentro de su genoma, mientras que conceptos basados en la evaluación de procesos evolutivos si lo logran en algunos casos. Por otro lado, también observamos que Vibrionaceae dentro de Cuatro Ciénegas tiene un pan-genoma abierto, lo que indica que existe una gran variación genética y de especies; sin embargo, los linajes internos dentro de Vibrionaceae tienen en su mayoría pan-genomas cerrados. Así mismo, encontramos que existen señales de elementos móviles, no obstante, no se encontraron integrones completos, ni islas de patogenicidad, pero si una baja presencia de fagos con respecto a las secuencias CRISPR. Nuestros resultados sugieren que entender a los grupos bacterianos a través del proceso evolutivo, puede mejorar la identificación de especies que han divergido recientemente. Sin embargo, existen muchos casos donde la clasificación tradicional tiene la resolución suficiente para diferenciar grupos taxonómicos como es el caso de la familia Vibrionaceae dentro de Cuatro Ciénegas. Esta familia está conformada por muchos linajes cohesivos que mantienen pan-genomas cerrados, probablemente asociados con la especialización de cada linaje al ambiente. Esta especialización se ve reflejada en la probable erosión de elementos móviles como los integrones y en el mecanismo vía secuencias CRISPR asociado al contexto de la comunidad microbiana dentro de las pozas de Cuatro Ciénegas.

Abstract

One of the central questions in biology is to understand how species are generated, and this is especially interesting considering that as a result of technological advances it has been revealing a high microbial diversity at species and gene level has been described. Recent facts show the need for a revision of concepts and methods used to define bacteria species, as well as a revision of how evolutionary forces shape gene diversity. In this work we used a comparative genomic perspective to review and evaluate: i) how do we define bacteria species, which are the limits of taxonomy using classical methods and which could be the alternatives and implications, ii) we also evaluate the impact of the different evolutionary forces on pan-genome, which represents the gen level diversity of a given taxonomic group; and iii) the mobile genetic elements through the evolution of pathogenesis. We assessed those topics studying isolates of the Vibrionaceae family obtained from Cuatro Ciénegas (CCB), Coahuila, a place well known for the high microbial diversity that it harbors and the little human impact on their ponds. Our results show that classical methods to define a species do not allow to delimit genetic groups that recently diverge and posses low genetic variation within its genome. Furthermore, we also observed that Vibrionaceae within Cuatro Cienegas possesses open pan-genomes, which suggests that this group possesses high genetic variation and high species richness. Nevertheless, lineages within CCB Vibrionaceae posses closed pan-genomes. Regarding mobile genetic elements, we found neither complete integrons mechanisms nor pathogenic islands, but we did see a correlation between low phages presence and a high number of CRISPR sequences. Our results suggest that understanding microbial groups through the evolutionary perspective may separate groups that had recently diverge. Nevertheless, there are cases where classical methods have the resolution to define taxonomic groups, as is the fact of CCB Vibrionaceae. This family is conformed by several lineages highly specialized to the environment. This specialization is observed in the loss of functional mobile genetic elements as is the case of integrons and the mechanisms of CRISPR sequences related to the microbial community context within ponds in Cuatro Ciénegas.

Presentación

El presente trabajo se centra en el análisis genómico de diferentes especies de la familia Vibrionaceae aisladas de Cuatro Ciénegas, Coahuila (CCC). El objetivo general fue evaluar los factores evolutivos que moldean a diversidad genética dentro de los linajes de Vibrionaceae en CCC, tomando en cuenta la distribución espacial de los aislados así como su cercanía con especies patógenas de animales. Para este fin la presente tesis está estructurada en diferentes secciones que incluyen: una introducción general, un capítulo sobre el concepto de especie en bacterias, dos capítulos en formato de artículo, y un apartado de discusión y conclusiones generales. Finalmente se incluye un apéndice donde se detalla la metodología empleada en los análisis presentados en el primer capítulo.

En la introducción se proporciona un panorama general de la diversidad bacteriana a nivel de especie y a nivel genómico. Así mismo se describen las características generales de los miembros de la familia Vibrionaceae y del sitio de estudio. Se finaliza esta sección son los objetivos generales y particulares de la tesis.

Durante el primer capítulo se hace una revisión del concepto de especie en bacterias, la cual responde a la necesidad de definir grupos cohesivos que permitan la elaboración de análisis de tipo poblacional y se discute a detalle una propuesta para definir especies utilizando métodos evolutivos.

Los capítulos dos y tres constan cada uno de una introducción, métodos, resultados y la discusión de los mismos, que se relaciona con los objetivos particulares del presente trabajo. Los objetivos particulares fueron: *i)* evaluar la diversidad a nivel de especie y genética de los miembros de la familia Vibrionaceae en Pozas Rojas, CCC, y explicar dichos patrones a través de la teoría evolutiva y ecológica, *ii)* entender la relación de los linajes de Vibrionaceae en Pozas Rojas, CCC con los linajes patógenos a través de los genes asociados a la patogénesis y los mecanismos de transferencia horizontal.

A continuación se presenta una sección de discusión y conclusiones generales tomadas de los tres capítulos principales y relacionadas con los objetivos de este trabajo. Además, se incluye una sección de perspectivas en la que se realizan propuestas para estudios posteriores que permitan continuar avanzando en la comprensión de la distribución de Vibrionaceae dentro de Cuatro Ciénegas así como en los procesos evolutivos detrás de la diversidad genética.

Finalmente, se presenta un apéndice donde se detallan los métodos empleados en el capítulo uno.

Capítulo 1.
Introducción

Una de las grandes preguntas en biología es y sigue siendo ¿Cómo se generan las especies? En este sentido, las especies son unidades evolutivas, las cuales basados en el concepto biológico de especie (Mayr, 1942) se consideran grupos de individuos que intercambian material genético y permanecen aislados de otros grupos. La interacción de las diferentes fuerzas evolutivas mantendría a estos grupos unidos o bien los separaría promoviendo la especiación. La biología evolutiva ha interpretado la especiación como un resultado de la selección natural o de procesos aleatorios como la deriva génica y fluctuaciones en el tamaño poblacional. Se han descrito al menos tres mecanismos de especiación: alopátrica, parapátrica y simpátrica. En el primer caso, una misma poza génica diverge debido a que una barrera física aísla las pozas génicas evitando el flujo genético (Mayr, 1963). En el segundo caso, una fracción de la poza génica se aísla del resto y por medio del aislamiento las dos pozas génicas divergen. Por último, en el caso de la especiación simpátrica ninguna barrera física impide el aislamiento, sin embargo, la divergencia ocurre por aislamiento reproductivo adaptativo; es decir el flujo génico entre subpoblaciones de una misma poza génica es evitado por medio de la adaptación a otro nicho dentro de un mismo sito (Dieckmann & Doebeli, 1999).

Aparentemente la universalidad de estos conceptos se ve desafiada por los organismos asexuales como las bacterias. En esta sección discutiré cómo se interpretan estos conceptos cuando intentamos analizar la diversidad bacteriana, hablaré de la variabilidad genética que tienen las bacterias y su implicación; finalmente, introduciré al grupo bacteriano y al sitio de estudio con los que desarrollaré las preguntas e hipótesis a lo largo de los siguientes capítulos.

1.1 Diversidad microbiana

Recientemente y derivado de las técnicas de secuenciación masiva, se ha descrito una amplia diversidad bacteriana, incluso se han descrito linajes nuevos de los que únicamente se tiene conocimiento de ellos derivado de estudios metagenómicos (Hug *et al*., 2016). Por lo que, esto nos lleva a la siguiente pregunta: ¿Cómo ha surgido toda esta diversidad? Al respecto hay al menos dos modelos que nos permiten entender los procesos de especiación en bacterias, estos modelos retoman lo que se ha descrito en los macro-organismos, sin embargo consideran aspectos propios de la biología de los procariontes. Una diferencia fundamental entre los eucariontes y los procariontes es que la recombinación no está ligada a la reproducción. Durante finales del siglo pasado el paradigma de la

clonalidad de las bacterias fue ampliamente estudiado, la pregunta era: ¿Qué tan clonales son las bacterias? (Smith *et al*., 1993; Shapiro, 2016). La clonalidad hace referencia a una población que evoluciona a partir de mutaciones aleatorias en el genoma, es decir la ausencia de recombinación. La estructura génica de *Escherichia coli* a través del tiempo revelada por medio de Electroforesis de Enzimas Multilocus (MLEE) (Selander & Levin, 1980) hizo pensar que los eventos de recombinación eran aislados en la naturaleza, por lo tanto los procariontes eran altamente clonales. Sin embargo ha quedado establecido que la estructura genética de los procariontes es heterogénea siendo algunos organismos altamente recombinantes y otros más cercanos a la clonalidad (Shapiro, 2016).

Tomando en cuenta esta característica fundamental, surge el modelo de especiación de tipo clonal, el cual está plasmado en el modelo de ecotipo (Cohan, 2001; Cohan & Perry, 2007). En este modelo la recombinación es muy baja entre los individuos y la mutación es entonces la principal fuente de variación dentro del ecotipo. Si algún miembro de la población adquiere un carácter que le dé ventaja sobre el resto de los miembros, este individuo no intercambiará dicho carácter por medio de la recombinación, sino que desplazaría al resto de los individuos. Este modelo tiene a su vez un fuerte componente ecológico, donde cada ecotipo ocupa un nicho específico dentro del ambiente. Entonces, en el caso de que algún individuo dentro del ecotipo adquiriera por mutación un carácter que le permitiera explorar un nuevo nicho se generaría un nuevo ecotipo. Ante la ausencia de recombinación estos ecotipos quedan aislados de su linaje parental, de esta forma los grupos pueden divergir al grado de generar especies distintas (Bendall *et al*., 2016). En este caso la especiación es instantánea.

El modelo alterno de especiación sería el panmíctico, donde la variación que existe en un linaje se puede generar por medio de mutación o a través de la recombinación. Las variantes que existen dentro de este linaje se intercambian entre los miembros de la población. En el caso de que un individuo dentro de la población adquiera un carácter que le permita invadir un nuevo nicho, la recombinación al homogeneizar la población podría hacer que ese carácter se perdiera. En este modelo, al igual que en la especiación simpática o alopática, el aislamiento reproductivo o geográfico es necesario para que la divergencia ocurra (Cadillo-Quiroz et al., 2012; Shapiro et al., 2012; Cordero & Polz, 2014). En este caso la divergencia entre linajes es lenta ya que la recombinación homogeneiza las pozas génicas a menos que la adaptación las aislé.

1.2 Genómica microbiana.

Como se mencionó anteriormente, en general los procariontes y los eucariontes difieren en su mecanismo reproductivo. En la mayoría de los eucariontes, el intercambio genético está ligado directamente a la reproducción a través de la meiosis la cual se lleva a cabo en algún estadio dentro del ciclo de vida (Goondenough & Heitman., 2014; Speijer *et al.,* 2015) de tal forma que estos mecanismos previenen la inserción de material genético no homólogo (Mira *et al*., 2010). Por otro lado, para los procariontes la recombinación no está asociada a la reproducción y se pueden definir dos mecanismos de recombinación, la recombinación homóloga y la transferencia horizontal de genes (THG). La primera hace referencia al intercambio de secuencias homologas vía RecA entre dos bacterias, mientras que la transferencia horizontal se refiere a la incorporación de material genético homólogo o no homólogo a través de mecanismos como la transformación, transducción o conjugación (Lawrence & Retchless, 2009). Esto se traduce en una dinámica genómica menos astringente (Doolittle, 1999; Koonin *et al*., 2001; Ragan, 2001; Georgiades & Raoult, 2010). Por lo tanto, bacterias de la misma especie pueden variar en cuanto a su contenido génico (Mira *et al*., 2010).

A pesar de que esta variación genética en los genomas bacterianos se conoce desde el inicio de la biología molecular en los años 70´s y se hizo más evidente al mapear los genomas con técnicas tales como la electroforesis de campos pulsados (Bergthorsson & Ochman, 1998), la genómica comparada abrió un panorama insospechado (Welch, 2002; Mira *et al.,* 2010). Un estudio de genómica comparada de *E. coli* donde se analizaron diferentes serotipos demostró que al aumentar el número de genomas aumentaba la cantidad de genes no compartidos por todos (Welch, 2002) existiendo una parte del genoma constante y otra parte muy variable. Esto llevo a la postulación de la hipótesis del pan-genoma, que es el repertorio genético que tiene un grupo taxonómico (p.ej. especie, género). Por lo tanto, una sola bacteria no contiene todo el repertorio genético de una especie, sino más bien contiene una fracción del pan-genoma.

Derivado de esas observaciones el pan-genoma se puede dividir en: core, accesorio y único. El core genoma constituye el set de genes que se comparten por todos los individuos dentro de un grupo taxonómico, mientras que los genes accesorios se refiere al repertorio de genes que está presente en más de un aislado pero no en todos; finalmente los genes únicos son los que solo se encuentran en un

aislado. Se ha propuesto que los genes accesorios juegan un papel fundamental en la adaptación de los grupos y posiblemente se asocian a la divergencia de los mismos (McInerney *et al.*, 2017).

1.2.1 ¿Los eucariontes también tienen pan-genomas?

De manera práctica, una diferencia entre los eucariontes y procariontes es que existen pocos genomas aun en las bases de datos para poder definir propiamente un pan-genoma en eucariontes, respecto a como se ha hecho en bacterias, sin embargo se han descrito algunos (Golicz *et al.*, 2015), como es el pan-genoma de la especie humana (Li *et al.*, 2009). Por otro lado, se ha descrito una amplia variación a nivel genómico en las especies varias especies, un ejemplo son dos especies del género *Muntiacus*, donde la especie nativa de china (*Muntiacus reevesi*) posee cuarenta y seis cromosomas en ambos sexos y la especie hindú (*Muntiacus muntjak vaginalis*) posee seis y siete cromosomas en hembras y machos, respectivamente (Zhou *et al.*, 2006; Tsipouri *et al.*, 2008). Esto indica que los genomas en eucariontes pueden ser muy variables, lo que sugiere que existen pan-genomas, sin embargo, el origen de esta variación parece estar dado más que nada por duplicaciones cromosómicas y no por transferencia horizontal de genes (Lynch & Conery., 2003; Ku *et al.*, 2015).

En los siguientes capítulos se explorara el pan-genoma de Vibrionaceae en el Valle de Cuatro Ciénegas, en Coahuila con el objetivo de encontrar cómo es la dinámica evolutiva de este grupo microbiano.

1.3 Historia natural de la familia Vibrionaceae

1.3.1 Aspectos históricos

Los humanos hemos coevolucionado con las bacterias, sin embargo se empezaron a estudiar estos organismos hasta principios/mediados del siglo XIX, las principales razones giran en torno al contexto histórico que se vivía en ese momento y que estaba relacionado con la teoría de la generación espontánea, refutada por Louis Pasteur en este periodo. Sin duda los estudios de Louis Pasteur como los relacionados con la fermentación abrieron el camino de la microbiología actual "*fermentation is a result of life without air*" (Pasteur., 1879). La primera especie de *Vibrio* se caracterizó utilizando

microscopía en 1854 por Filippo Pacini, quien estudió los brotes de cólera que se dieron en la región de Florencia, Italia. Durante el mismo tiempo se estudió la epidemiología de cólera en varias ciudades de Inglaterra y se determinó que el agua contaminada era una posible causa de la enfermedad. Treinta años después Robert Koch y su equipo, siguieron un brote de cólera en India, a partir de estos estudios lograron obtener cultivos puros de *Vibrio cholerae* y definir que este organismo era la causa de la enfermedad. Robert Koch y su grupo hicieron otro gran descubrimiento, encontraron que otras formas de *Vibrio* se encuentran de manera ubicua en cuerpos acuáticos y muchos de ellos no son patógenos humanos. A finales de 1880 Martinus Beijerinck aisló por primera vez organismos no patógenos de este grupo: *V. fischeri*, *V. splendidus*, y *Photobacterium phosphoreum* son un ejemplo (Thompson *et al*., 2004; Al-saari et al., 2015).

1.3.2 Taxonomía y características generales

De acuerdo con el manual de Bergey´s los aislados de Vibrionaceae pertenecen a las Gammaproteobacteria, son Gram negativos, tienen un flagelo polar, son mesófilos, quimiorganótrofos y presentan fermentación facultativa. En la octava edición de este manual se propuso la formación de la familia Vibrionaceae (Veron, 1965), esta primer clasificación incluía los géneros *Aeromonas*, *Plesiomonas*, *Beneckea*, *Lucibacterium*, *Photobacterium* y *Vibrio* propuestos un año antes (Thompson *et al.,* 2004), sin embargo, el género *Aeromonas* fue removido dadas sus características en cuanto a contenido de GC (Baumann & Schubert, 1984; Colwell *et al.,* 1986). De al igual que el género *Pleisiomonas* basados en datos de 5S rRNA (MacDonell & Colwell, 1985). Mientras que el género *Beneckea* y *Lucibacterium* fueron desechados y reclasificados como *Vibrio. Luciferarium* contenia una sola especie *Luciferarium harveyi (Vibrio harveyi*; Buchanan & Gibbons, 1974)Baumman *et al.,* 1980). Actualmente la familia Vibrionaceae está constituida por alrededor de 150 especies distribuidas en siete géneros distintos: *Aliivibrio*, *Echinomonas*, *Enterovibrio*, *Grimontia*, *Photobacterium*, *Salinivibrio* y *Vibrio*. Sin embargo, aún existen controversias alrededor de este tema (Boyd *et al.,* 2015; Al-saari *et al.,* 2015).

Una característica particular de Vibrionaceae es la presencia de dos cromosomas, que contrasta con lo que se ha visto en familias cercanas como Aeromonadaceae y Plesiomonaceae, que solo tienen un cromosoma (Okada *et al.,* 2005). Aunque el contenido de guanina-citosina (GC) es muy similar

en ambos cromosomas, el tamaño varía considerablemente; en general el cromosoma 1 oscila entre los 3 a 4 Mpb, mientras que el cromosoma 2 mide entre 1-2 Mpb (Okada *et al.,* 2005). Una de las razones para no ser considerado un plásmido es que posee genes esenciales y que se replica junto con el cromosoma 1, mientras que los plásmidos se pueden replicar más veces (Hwan & Chattoraj, 2014). Utilizando métodos genómicos se ha encontrado que si bien en el cromosoma 2 hay genes ¨*housekeeping*¨ también hay muchas proteínas hipotéticas (Dikow & Smith, 2013), así mismo se ha visto que la tasa de sustitución es más alta en el cromosoma 2, lo que indica que este último evoluciona más rápido (Cooper et al., 2010). La hipótesis más fuerte respecto al origen del cromosoma 2 es que un mega-plásmido fue capturando genes esenciales convirtiéndose en un cromosoma en sí mismo, es por esto que se ha propuesto que el ancestro de todos los Vibrionaceae adquirió un plásmido que se mantiene hasta ahora, esta hipótesis se basa en que los orígenes de replicación en el cromosoma 2 son similares a los encontrados en los plasmidos (Heidelberg, 2000).

1.3.3 Hábitat y papel ecológico

Los miembros de la familia Vibrionaceae se encuentran generalmente en ambientes acuáticos tanto marinos como de agua dulce (Thompson *et al.,* 2004). Sin embargo se han encontrado en ambientes extremos, como en el océano ártico donde las temperaturas son cercanas a cero, así como en las profundidades del océano (Kahlke & Thorvaldsen, 2012). Muchos de los miembros de esta familia se asocian a animales acuáticos, como los corales y peces (Boyd *et al.,* 2015). Un ejemplo de esta asociación es *Aliivibrio fischeri* quien produce luminiscencia en *Euprymna scolopes* (Ruby & Lee, 1998). Sin embargo, hay muchos casos donde los miembros de la familia son patógenos, como es el caso de *Vibrio cholerae*. Su metabolismo es igualmente versátil, es heterótrofo y parece ser que tiene un papel importante en la liberación al ambiente de aminoácidos que obtiene a partir de la degradación de la pared de quitina de algunos crustáceos en los océanos, así mismo, se ha visto pueden participar en el ciclo del nitrógeno (Urdaci *et al.,* 1988; Thompson *et al.,* 2004; Takemura *et al.,* 2014).

1.4 Cuatro Ciénegas. Un modelo para estudiar la evolución microbiana.

El proyecto desarrollado en esta tesis forma parte del esfuerzo que se ha venido haciendo por más de diez años para entender la diversidad microbiana dentro del sistema acuático de Cuatro Ciénegas (Souza *et al.,* 2006). Este sitio estuvo conformado por cuatro grandes lagunas de las cuales en la actualidad solo quedan algunos ríos, pozas y manantiales. Durante la década de los 60's se clasificaron los sistemas acuáticos dentro de Cuatro Ciénegas en siete grupos: Churince, Becerra/Garabatal, Río Mesquites, Rio Puente Chiquito, Tio Candido/Escobeda, Santa Tecla y Río Salado/Río Grande (Minckle, 1969). El Valle de Cuatro Ciénegas se encuentra ubicado dentro del desierto Chihuahuense en el estado de Coahuila. Dada su importancia en el año de 1994 se estableció el Área de Protección de Flora y Fauna de Cuatro Ciénegas. Este sitio a demás es considerado prioritario para la conservación por la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) y Wold Wild life (WWF; Souza *et al.,* 2006).

El humedal del valle de Cuatro Ciénegas ha sido reportado como una de las regiones con más endemismos en Norteamérica (Stein *et al.,* 2000). Se han reportado al menos 70 especies endémicas, 9 reptiles, 10 peces, 13 moluscos, 7 crustáceos, 4 escorpiones, 3 insectos y 26 plantas de diferentes familias (Desert Fishes Council Web, 2015). Aunado a esto, se ha reportado una gran diversidad de bacterias dentro de los sistemas acuáticos del valle de Cuatro Ciénegas (Souza *et al.,* 2006; Escalante *et al.,* 2008; Bonilla-Rosso *et al.,* 2012; Peimbert *et al.,* 2012), esta diversidad parece estar relacionada con la escases de nutrientes, principalmente fósforo y el aislamiento del sitio.

1.4.1 Características generales del valle de Cuatro Ciénegas.

El valle de Cuatro Ciénegas se ubica en la zona centro del estado de Coahuila, México (26° 59' N, 102° 04' W), a una altitud promedio de 740 metros sobre el nivel del mar (Figura 1). El valle se encuentra rodeado al norte por la Sierra de la Madera, al sur con la Sierra de San Marcos y al oeste con la Sierra de la Fragua (Minckley, 1969). El clima de la región es árido con una presipitacion anual menor a los 200 mm (Muller, 1947) y una evaporación mayor a los 2,000 mm/año (Johannesson *et al.,* 2004). La temperatura oscila entre los 44 °C en verano a 0 °C en invierno (Minckley, 1969). El acuífero se encuentra en la región hidrológica Rio Bravo-Conchos, cuenca Presa Falcón-Rio Salado

y Subcuenca Río Salado-Nadadores. El valle tiene una densidad de 12 a 15 pozas por kilómetro cuadrado, estos cuerpos de agua oscilan de varios centímetros de diámetro hasta aproximadamente 1 kilómetro de longitud y de menos de 1 metro hasta más de 10 metros de profundidad (Minckley, 1969). En cuanto a las características fisicoquímicas del agua, estos cuerpos se caracterizan por mantener un pH cercano a la neutralidad (6.9-7.7), las aguas son duras y contienen gran cantidad de sales de calcio y magnesio. Sin embargo se destaca la escases de fosforo dentro de los cuerpos de agua (Souza *et al.,* 2006; Souza *et al.*, 2012).

Se sugiere que una fracción del agua subterránea proviene de la precipitación sobre las montañas circundantes (Johannesson *et al.,* 2004) las cuales al infiltrarse adquieren una gran cantidad de minerales. El agua regresa a la superficie al entrar en contacto con el magma que se bajó la Sierra de San Marcos y Pinos (Wolver and Diehl, 2011).

1.4.2 Sistema hidrológico Los Hundidos.

El presente trabajo se desarrolló con aislados obtenidos del sistema acuático Los Hundidos durante Marzo de 2013. Este sistema está conformado por al menos nueve pozas, estas pozas son consideradas pozas de desecación donde el nivel del agua se reduce dramáticamente durante el verano y se incrementa en el invierno (Rebollar *et al.,* 2012). El agua que alimenta estas pozas proviene de un rio subterráneo que emerge por diversos sumideros y conforman las pozas. Este sistema es conocido también como pozas rojas, debido al color de las mismas. La temperatura del agua de estas pozas fluctúa entre los 10 °C y 60 °C, con una fluctuación máxima diaria de 15 °C durante el invierno y 45 °C en verano (Peimbert *et al.,* 2012). El agua es acida (pH 5.5), sin embargo ha sido clasificado un sitio hipersalino con una concentración iónica única (Rebollar *et al.,* 2012).

1.4.3 Intercambio genético. Transferencia horizontal en bacterias de Cuatro Ciénegas.

Una de las preguntas que se ha buscado entender es cómo podemos explicar la diversidad de especies microbianas en los sistemas acuáticos de Cuatro Ciénegas. Los elementos hasta el año 2012 eran los siguientes: Alta diversidad beta (Escalante *et al.*, 2008), altos niveles de clonalidad en los aislados bacterianos de los géneros *Pseudomonas*, *Exiguobacterium y Bacillus* (Escalante, 2008; Rebollar *et al.,* 2012; Avitia *et al.,* 2014), genomas pequeños *Bacillus coahuilensis* y escasez de nutrientes, principalmente fósforo. Estos elementos llevaron a la formulación de una hipótesis donde la selección natural se encuentra detrás de la diversificación siendo la falta de nutrientes la principal presión selectiva, esto se ve reflejado en los genomas reducidos y en la alta clonalidad. En este sentido los eventos de transferencia horizontal de genes (THG) serían evitados debido al costo energético de replicar genomas más grandes. En este escenario, el DNA libre en lugar de ser una fuente de información genética se emplea como una fuente de nutrientes (Souza *et al.,* 2012). Recientemente el uso de DNA por las bacterias de Cuatro Ciénegas como fuente de energía ha sido demostrado (Tapia-Torres *et al.,* 2015).

Derivado de este modelo, es posible interpretar que la transferencia horizontal no juega un papel importante en la evolución de las bacterias en Cuatro Ciénegas, sin embargo esto no es del todo cierto. El genoma de *Bacillus coahuilensis* muestra varios eventos de THG como lo son la adquisición de los genes *sqd1* y *sqdX* asociados a la síntesis sulfolípidos en la membrana, así como genes asociados a la producción de rodopsinas, ambos probablemente adquiridos de una cianobacteria (Alcaraz *et al.,* 2008).

¿Clonalidad es sinónimo de ausencia de THG? la respuesta sería “No necesariamente”. Pensemos por ejemplo en una bacteria pandémica, dicha bacteria intercambia genes vía THG de manera frecuente. En alguno de esos eventos, adquiere un grupo de genes que le permite infectar con mayor eficacia a un hospedero, la bacteria se divide y se propaga rápidamente entre los diferentes hospederos y se mantiene de esta forma durante cierto tiempo. Bajo dicho escenario la estructura poblacional que se esperaría sería de tipo clonal.

Para realmente elucidar el impacto de la THG y los altos niveles de clonalidad, debemos empezar con hacernos ciertas preguntas. Primero, qué tan frecuentes son los eventos de THG y los que se han encontrado qué tan antiguos son. El éxito de un evento de THG va a estar asociado con el coeficiente de selección del gen sujeto a dicho evento. Por lo que, es posible que en las bacterias de Cuatro Ciénegas existan muchos eventos de transferencia horizontal, pero ninguno de ellos sea exitoso y por lo tanto no logramos observarlos. Sin embargo, siguiendo la hipótesis basada en la presión selectiva guiada por la falta de fósforo, ésta sugeriría que la THG se evita por el gasto energético que implica, sin embargo habrá casos donde el costo energético sea costeable en el sentido de que aumentaría la adecuación de cierto grupo bacteriano, como puede ser el caso de *Bacillus coahuilensis*.

Capítulo 2.

El concepto adaptativo de especie

En la siguiente sección se hace una revisión sobre el concepto de especie. Esta sección se justifica en el marco de esta tesis pues uno de sus objetivos es entender la diversidad a nivel intrapoblacional. Para que este objetivo se cumpla es necesario definir con claridad las unidades poblacionales y por ende las especies. Durante el siguiente capítulo se discutirán los principales problemas alrededor de la definición de una especie bacteriana y se abordan nuevas aproximaciones para hacer frente a este reto valiéndose de la información genómica disponible.

2.1 El concepto de especie microbiano.

¿Cómo se define en microbiología una especie? Esto es en realidad importante ya que las especies representan grupos genéticos cohesivos que comparten una misma historia evolutiva, por lo que si se pretende hacer un estudio de la evolución de algún grupo es necesario tener clara la unidad con la que se está trabajando.

Como se mencionó anteriormente, las bacterias tienen características inherentes a su biología que presuntamente no les permiten encajar en el concepto biológico de especie, por lo que los microbiólogos han tenido que desarrollar estrategias basadas principalmente en las características primero morfológicas, luego bioquímicas y ahora genómicas, para definir estos grupos.

La forma de definir especies más utilizada hasta ahora se basa en la similitud de la secuencia del gen 16S rRNA, donde se considera que dos organismos forman parte de la misma especie si presentan al menos el 97% de identidad en esa secuencia. Este criterio es análogo a un porcentaje de hibridación DNA-DNA igual o mayor al 70%, que hasta el momento ha resultado la técnica más certera pero metodológicamente inviable para definir una especie (Wayne *el al.*, 1987; Ogilvie & Hirsch, 2012).

El problema al que se han enfrentado los microbiólogos es la variación genética que existe dentro de las especies que se han definido utilizando este método. Por este motivo han surgido metodologías más exactas que consideran esta variación, un ejemplo de ello es el *multilocus sequence typing* (MLST) que se basa en la secuenciación de varios genes donde se recupera mayor diversidad genética, lo que permite identificar con mayor precisión grupos genéticos (Woodford & Johnson 2004).

Utilizando la información genómica disponible se han descrito algunos métodos que se basan en la teoría evolutiva para distinguir especies. Un método que se ha comprobado que es certero para diferenciar especies está basado en la cuantificación del flujo génico a través de homoplasias, este método se ha propuesto es universal pues se ha probado en eucariontes y procariontes (Bobay & Ochman, 2017). Por otro lado, se ha propuesto que la adaptación local es la responsable de la diversidad microbiana (Friedman *et al.*, 2013; Shapiro *et al.*, 2012). En este sentido se ha propuesto que es posible definir especies utilizando métodos basados en pruebas clásicas de selección, uno de ellos es el concepto adaptativo de especie (Vos, 2011). De manera general, este concepto plantea el uso de la prueba de McDonald-Kreitman (MK; 1991) en datos genómicos para definir especies. Esta prueba originalmente se planteó para identificar señales de selección utilizando un set de datos de dos especies distintas. En este caso particular, se plantea el uso de dicha prueba partiendo del hecho de no saber de forma explícita que dos grupos son especies diferentes, como en la prueba original. Entonces dos grupos aparentemente divergentes solo se pueden considerar como especies distintas si se identifican eventos de selección asociados, que sería observable a partir de la aplicación de esta prueba (Figura 2; Eyre-Walker & Charlesworth; 2006; Vos *et al.*, 2013).

Concretamente, la prueba se basa en el índice de neutralidad. De esta manera, cuando dos grupos bacterianos divergen a través de procesos neutrales las diferencias fijadas entre éstos grupos se espera que se vean reflejadas en la frecuencia de los polimorfismos sinónomos y no sinónimos dentro de ellos, tal como se expresa en el índice de neutralidad (NI):

$$\mathbf{NI=(Pn/Dn)/(Ps/Ds)}$$

Donde NI es el índice de neutralidad, Pn y Ps son los polimorfismos no sinónimos y sinónimos dentro de los grupos, y Dn y Ds son los polimorfismos no sinónimos y sinónimos entre los grupos, respectivamente. Cuando NI=1, no hay diferencias entre el patrón de las substituciones sinónimas y no sinónimas, entonces los grupos divergen de manera neutral. Cuando NI<1 se asume que las diferencias fijadas están bajo selección. Finalmente cuando NI >1 las diferencias entre especies están dadas por mutaciones sinónimas más frecuentemente de lo esperado. En este caso especifico, dos grupos se pueden clasificar como especies distintas cuando la prueba de MK da un valor mayor a uno para el índice de neutralidad y como una sola especie cuando el valor es menor a uno. Cuando el

índice de neutralidad es menor a uno, las diferencias fijadas entre los grupos estarán dadas por diferencias no sinónimas, asumiendo selección positiva (Vos, 2011).

Estas ideas resultan interesantes, sin embargo valdría la pena considerar que en la propuesta original solo se considera comparar regiones homólogas entre dos grupos bacterianos, sería pertinente tomar en cuenta el resto de la variación genética no homologa, donde probablemente existen genes adaptativos y ver el comportamiento de la prueba así como evaluar el impacto en la definición de especies (McInerney *et al.,* 2017).

2.2 La taxonomía de Vibrionaceae frente al concepto adaptativo de especie.

Una forma de determinar qué tan importante es la recombinación en un linaje bacteriano es utilizando secuencias de DNA y determinar la proporción de variación genética que entra por recombinación y mutación, a través de la relación de probabilidades de que un polimorfismo dado se explique por recombinación o mutación (r/m; Didelot & Falush, 2007). Utilizando esta aproximación se ha calculado r/m para diferentes bacterias con hábitats igualmente diversos (Vos & Didelot, 2009). *Vibrio parahaemolyticus* y *Vibrio vulnificus* tienen unos de los niveles más altos de r/m, siendo los valores 39.8 y 26.7 respectivamente. La recombinación en términos generales es una fuerza cohesiva pues mantiene los linajes unidos al intercambiar material genético evitando en todo caso la diversificación. Se ha observado en algunas poblaciones de *Vibrio cyclitrophicus* que la diversificación es posible debido a la fijación diferencial en dos grupos de bacterias de ciertos *loci* que aumentan la adecuación, creando un balance entre la intensidad de la selección y la recombinación, la cual se evita para no perder la combinación genómica adaptativa (Shapiro *et al.,* 2012).

Para probar la universalidad del concepto adaptativo de especie, se calcularon las proporciones de cambios sinónimos y no sinónimos entre grupos de bacterias que han sido descritas como especies distintas utilizando otros métodos dentro de la familia Vibrionaceae. Para el análisis, se eligieron especies cercanas como es el caso de *Vibrio cholerae* y *Vibrio mimicus; Vibrio ordalii* y *Vibrio anguillarum*, quienes son consideradas como especies diferentes, pero que durante la historia han sido clasificadas como la misma especie (Figura 2). En el análisis se incluyó también a *Vibrio cyclitrophicus* específicamente los genomas de los aislados que están en proceso de diversificación reciente (Shapiro *et al.*, 2012).

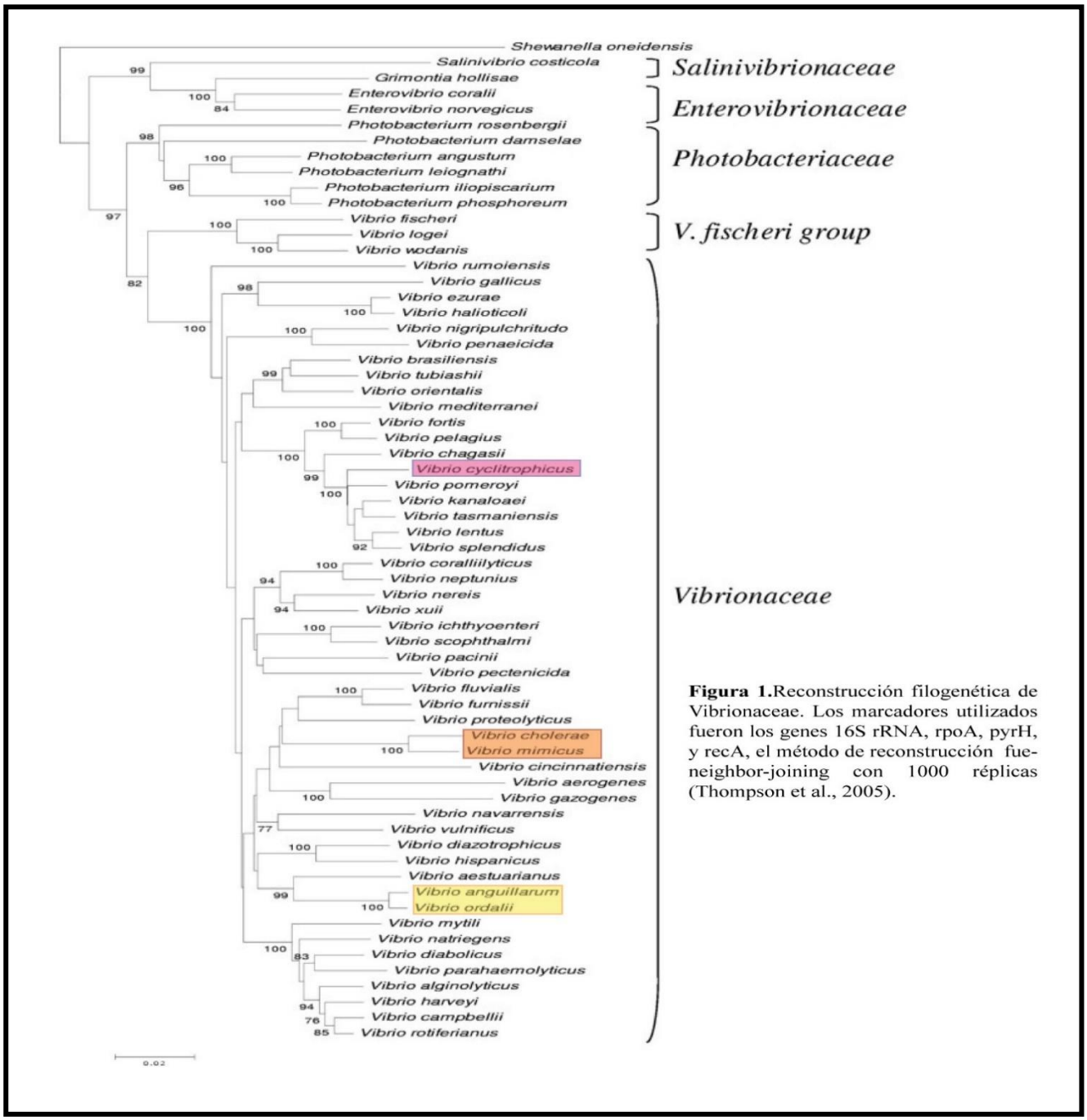


Figura 1. Reconstrucción filogenética de Vibrionaceae. Los marcadores utilizados fueron los genes 16S rRNA, rpoA, pyrH, y recA, el método de reconstrucción fue-neighbor-joining con 1000 réplicas (Thompson et al., 2005).

Se tomaron cinco genomas de cada una de las cuatro especies a analizar, para generar de esta forma set de datos poblacionales (Tabla 1). Los grupos de genomas fueron alineados mediante el programa Mugsy (Angiuoli & Salzberg, 2011), y a través de ventanas móviles se calculó el índice de neutralidad (Apéndice 1, Tabla 2).

Tabla 1. Genomas utilizados junto con el número de proyecto de NCBI.

Genero	Especie	Nombre en las bases de datos	BioProject - NCBI
Vibrio	*cholerae*	Vibrio_cholerae_O1_116059_uid195203	PRJNA76257
	cholerae	Vibrio_cholerae_O1_2010EL_1792_uid190214	PRJNA190214
	cholerae	Vibrio_cholerae_O1_116063_uid195205	PRJNA195205
	cholerae	Vibrio_cholerae_O1_2010EL_1798_uid190213	PRJNA190213
	cholerae	Vibrio_cholerae_O1_2010EL_1786_uid78933	PRJNA78933
Vibrio	*cholerae*	Vibrio_cholerae_HC_02A1_uid181562	PRJNA181562
	cholerae	Vibrio_cholerae_HC_06A1_uid180002	PRJNA180002
	cholerae	Vibrio_cholerae_HC_17A2_uid181734	PRJNA181734
	cholerae	Vibrio_cholerae_HC_02C1_uid181733	PRJNA181733
	cholerae	Vibrio_cholerae_HC_17A1_uid181723	PRJNA181723
Vibrio	*cholerae*	Vibrio_cholerae_HE_09_uid181560	PRJNA181560
	cholerae	Vibrio_cholerae_HE_25_uid180835	PRJNA180835
	cholerae	Vibrio_cholerae_HE_45_uid180836	PRJNA180836
	cholerae	Vibrio_cholerae_HE_16_uid181808	PRJNA181808
	cholerae	Vibrio_cholerae_HE_40_uid181731	PRJNA181731
Vibrio	*cholerae*	Vibrio_cholerae_V51_uid54329	PRJNA54329
	cholerae	Vibrio_cholerae_VC1761_uid190521	PRJNA190521
	cholerae	Vibrio_cholerae_VC4370_uid190522	PRJNA190522
	cholerae	Vibrio_cholerae_V52_uid54331	PRJNA54331
	cholerae	Vibrio_cholerae_VC35_uid190520	PRJNA190520
Vibrio	*cholerae*	Vibrio_cholerae_CP1030_3__uid181789	PRJNA181789
	cholerae	Vibrio_cholerae_CP1032_5__uid180826	PRJNA180826
	cholerae	Vibrio_cholerae_CP1033_6__uid180427	PRJNA180427
	cholerae	Vibrio_cholerae_CP1035_8__uid181722	PRJNA181722
	cholerae	Vibrio_cholerae_CP1037_10__uid181790	PRJNA181790
Vibrio	*anguillarum*	Listonella_anguillarum_M3	PRJNA211964
	anguillarum	Vibrio_anguillarum_775	PRJNA68057
	anguillarum	Vibrio_anguillarum_96F	PRJNA200507
	anguillarum	Vibrio_anguillarum_NB10_serovar_O1	PRJEB5701
	anguillarum	Vibrio_anguillarum_RV22	PRJNA200508
Vibrio	*mimicus*	Vibrio_mimicus_CAIM_602_uid189221	PRJNA189221
	mimicus	Vibrio_mimicus_MB_451_uid41415	PRJNA41415
	mimicus	Vibrio_mimicus_SX_4_uid182021	PRJNA182021
	mimicus	Vibrio_mimicus_VM223_uid41409	PRJNA41409
	mimicus	Vibrio_mimicus_VM573_uid56013	PRJNA56013
Vibrio	*cyclitrophicus*	Vibrio_cyclitrophicus_1F111_uid200006	PRJNA200006

	cyclitrophicus	Vibrio_cyclitrophicus_1F175_uid200021	PRJNA200021
	cyclitrophicus	Vibrio_cyclitrophicus_1F273_uid200022	PRJNA200022
	cyclitrophicus	Vibrio_cyclitrophicus_1F289_uid200023	PRJNA200023
	cyclitrophicus	Vibrio_cyclitrophicus_1F53_uid200019	PRJNA200019
Vibrio	*cyclitrophicus*	Vibrio_cyclitrophicus_ZF170_uid200031	PRJNA200031
	cyclitrophicus	Vibrio_cyclitrophicus_ZF205_uid200033	PRJNA200033
	cyclitrophicus	Vibrio_cyclitrophicus_ZF207_uid200032	PRJNA200032
	cyclitrophicus	Vibrio_cyclitrophicus_ZF255_uid200034	PRJNA200034
	cyclitrophicus	Vibrio_cyclitrophicus_ZF264_uid200035	PRJNA200035

Vibrio cholerae y *Vibrio mimicus*, son especies muy similares bioquímicamente y esta es una de las razones por las que *V. mimicus* posee ese nombre, a nivel de secuencia del gen 16S rRNA son también muy parecidas (*i.e.* difieren en 6 de 1456 nucleótidos) (Chun *et al*, 1999). A demás, la ecología y patología de estos grupos es análoga. Por un lado, las dos especies se distribuyen en ambientes de agua dulce y estuarinos, además de que ambas han sido aislada de pacientes con diarrea (Davis *et al*., 1981). Algunos genes de virulencia compartidos entre estas dos especies se encuentran asociados con bacteriófagos u otros elementos móviles, lo que refleja la importancia de la THG en este grupo (Thompson *et al*., 2006).

En el caso de *Vibrio cholerae* y *Vibrio mimicus*, solo en dos de los cinco casos probados se logró diferenciar a las dos especies. Las cepas *V. cholerae* O1, CP y V5_VC no se lograron diferenciar de *V. mimicus*. Estas tres cepas pertenecen a aislados patógenos, mientras que no todas las cepas de *V. mimicus* son patógenas, como es el caso de *V. mimicus* VM223 que fue aislada de un bivalvo (Hasan *et al.,* 2010). Sin embargo, estos dos grupos poseen el elemento CTX que codifica para la toxina (CT) de *V. cholerae* que tiene origen en un bacteriófago (Wang *et al.,* 2011), lo que puede estar influenciando este resultado.

Por otro lado, cuando contrastamos las cepas *V. cholerae* HC y HE encontramos que *V. cholerae* y *V. mimicus* sí se diferencian basados en el índice de neutralidad. Estas dos cepas de *V. cholerae* fueron obtenidas de un brote en Haití en 2010. El escenario previo al brote fue un terremoto en enero de 2010, seguido de un verano muy caliente; y después del primer diagnóstico de cólera, un huracán pasó por la zona (Figura 3; Hasan *et al.,* 2012). Sin embargo, se ha demostrado que el inicio del brote estuvo relacionado con la presencia de la Organización Mundial de la Salud en la zona, pues las cepas de *V. cholerae* encontradas son cercanas a las cepas Asiáticas (Enserink, 2011; Lantagne *et al.,* 2013; Robbins, 2014). En dado caso nuestros resultados indican que se trata de una cepa bien diferenciada respecto a *V. mimicus*.

En estudios recientes se ha reportado esta dificultad para delimitar claramente algunas cepas de *V. cholerae* y *V. mimicus* (Bobay & Ochman, 2017), estos casos sería necesario revisarlos dada la importancia de estas especies en la salud humana.

Finalmente, nuestros resultados indican que éste método falla al delimitar estos grupos en ciertos casos, sin embargo, probablemente si no se hubiera seccionado el grupo de *Vibrio cholerae* y se hubieran tomado todos los genomas como una sola población, sí se hubieran diferenciado, esto indica que estos grupos se comportan como ecotipos que pertenecen a una unidad mayor que es la especie.

Tabla 2. Resultados de las pruebas hechas con ventanas móviles y el índice de neutralidad.

Par de especies comparadas		**Número de eventos donde NI mayor a 1**	**Número de eventos donde NI menor a 1**	**Sitios sin variación**	**Total de puntos analizados**	**Especies distintas según NI**
Vibrio cholerae CP	*Vibrio mimicus*	324	308	0	632	
Vibrio cholerae O1	*Vibrio mimicus*	333	317	0	650	
Vibrio cholerae HE	*Vibrio mimicus*	331	345	0	681	*
Vibrio cholerae HC	*Vibrio mimicus*	308	330	0	638	*
Vibrio cholerae V5_VC	*Vibrio mimicus*	284	274	0	558	
V. ordalii	*V. anguillarum*	122	110	0	232	
Vibrio cyclitrophicus 1F	*Vibrio cyclitrophicus ZF*	52	49	64	165	

Al igual que las especies anteriores, *V. ordalii* y *V. anguillarum* son muy similares y fueron consideradas una sola especie hasta la descripción formal en los ochentas (Schiewe & Crosa 1981); sin embargo, recientemente se ha documentado la existencia de cepas de *V. anguillarum* que son indistinguibles de *V. ordalii* (Thompson *et al*., 2006). La proporción del índice de neutralidad en las muestras analizadas de *V. ordalii* y *V. anguillarum* son mayores a uno a lo largo del genoma, lo que indica que bajo el concepto adaptativo de especie son la misma unidad evolutiva (Figura 4).

Las cepas de *V. cyclitrophicus* analizadas provienen de un muestreo que supone un proceso de divergencia temprana. En el análisis se observó mayor proporción de valores iguales a cero para el

índice de neutralidad, lo que indica neutralidad. Este resultado es congruente con los análisis previos (Shapiro *et al*., 2012) donde observan que la divergencia se da en pocas regiones y no a través de barridos selectivos en todo el genoma, lo que explicaría los niveles de neutralidad. Sin embargo, si se descarta la proporción del genoma que es neutral, la proporción del índice de neutralidad que es mayor a uno es más alta que la que es menor a uno, por lo que deben de ser consideradas la misma especie.

En síntesis, los clados que son más cercanos entre sí, *V. cholerae* y *V. mimicus; V. ordalii* y *V. anguillarum;* y *V. cyclitrophicus* (1F y ZF) no fueron reconocidos como especie, utilizando el concepto adaptativo de especie. Una posibilidad para explicar este patrón es que lo que diferencia a estas especies se encuentra en el genoma flexible, información que queda excluida en este tipo de análisis. Dentro de esta posibilidad se asumiría que la mayor parte de las adaptaciones llegan a través de THG, por lo que las diferencias entre grupos no serían evidentes. Esto toma importancia para este caso específico debido las altas tasas de recombinación reportadas para el género (Mazel, 2006). Otra posibilidad radica en la cantidad de genomas analizados. Recientemente se ha descrito que *Vibrio parahaemolyticus* tiene una estructura poblacional global, este escenario indicaría que habría que muestrear la mayor parte de la variación para tener un resultado certero. La proporción de cambios sinónimos y no sinónimos tendría que ser evaluada junto con aspectos demográficos, para poder validar estas inferencias y no sea el reflejo que cambios en el tamaño efectivo de las poblaciones.

El concepto de especie es la consecuencia de entender el proceso evolutivo, que es la ventaja del concepto adaptativo de especie. Este concepto solo diferencia como especies distintas a grupos bacterianos que tienen alta variación a lo largo del genoma, por lo que grupos que estén en proceso de divergencia temprana (Shapiro *et al.*, 2012) se seguirían considerando como la misma especie. La teoría y algunas evidencias de organismos de vida libre indican que para los micro-organismos la divergencia a partir de la adaptación es más común que los procesos neutrales (Cadillo-Quiroz *et al.*, 2012; Shapiro *et al.*, 2012; Cordero & Polz, 2014). Sin embargo, sería interesante incorporar otros elementos como el tamaño efectivo, esto para poder probar si es posible identificar especies en grupos que estuvieran sujetos a cuellos de botella.

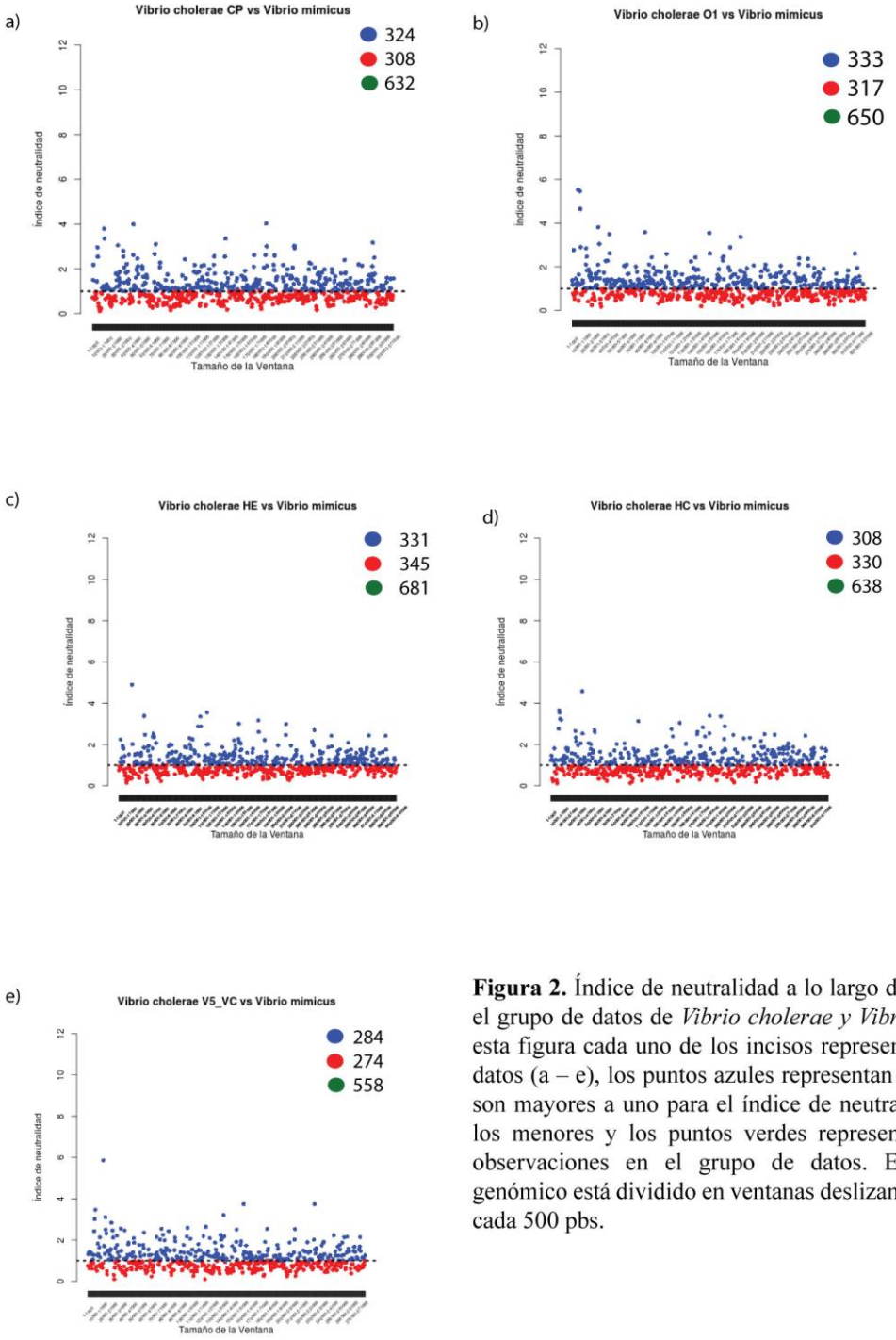


Figura 2. Índice de neutralidad a lo largo del genoma para el grupo de datos de *Vibrio cholerae y Vibrio mimicus.* En esta figura cada uno de los incisos representa un grupo de datos (a – e), los puntos azules representan los valores que son mayores a uno para el índice de neutralidad, los rojos los menores y los puntos verdes representan el total de observaciones en el grupo de datos. El alineamiento genómico está dividido en ventanas deslizantes de 1000 pbs cada 500 pbs.

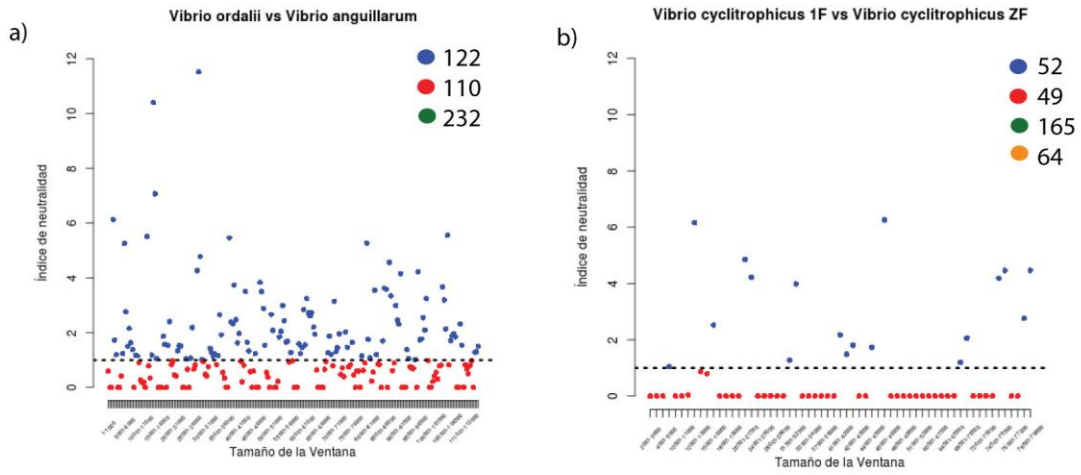


Figura 3.Ín dice de neutralidad a lo largo del genoma. *a) Vibrio ordalii y Vibrio anguillarum, b) Vibrio cyclitrophicus 1F y Vibrio cyclitrophicus ZF.* Los puntos azules representan los valores que son mayores a uno para el índice de neutralidad, los rojos los menores, los puntos verdes representan el total de observaciones en el grupo de datos y finalmente, la marca naranja indica el número de observaciones no incluidas. El alineamiento genómico está dividido en ventanas deslizantes de 1000 pbs cada 500 pbs.

Capítulo 3.

Vibrionaceae en Cuatro Ciénegas

En el capítulo anterior se discutió la forma en la que se define una especie y las nuevas propuestas basadas en conceptos evolutivos guiadas por la información genómica disponible. La definición de una especie es clave pues es la unidad operativa que se utiliza en los análisis ecológicos, y en el caso de los análisis evolutivos es la unidad de donde se distinguen poblaciones. En este sentido es necesario tener claro cómo se define la unidad que trabajamos con el fin de llegar a conclusiones certeras.

En este capítulo emplearemos los diferentes conceptos discutidos en los capítulos anteriores y los aplicaremos a un grupo particular de aislados de la familia Vibrionaceae obtenidos de Cuatro Ciénegas. Empezaremos desde una perspectiva ecológica hasta una perspectiva evolutiva, por lo que este capítulo está dividido en dos partes, en una primer parte hablaremos de la diversidad a nivel de especie en un contexto ecológico, seguido de un segundo apartado donde se analizan los procesos microevolutivos dentro de los linajes de la familia Vibrionaceae con énfasis particular en el pangenoma.

3.1 Diversidad de la familia Vibrionaceae dentro del Valle de Cuatro Ciénegas.

Durante más de diez años de estudio de la diversidad de Cuatro Ciénegas, no fue hasta 2012 que se empezaron a recuperar en cultivos miembros de la familia Vibrionaceae. En esta tesis presentamos el primer reporte de *Vibrio* como género dominante a nivel de cultivo en el sistema acuático Los Hundidos dentro del valle de Cuatro Ciénegas (Figura 5).

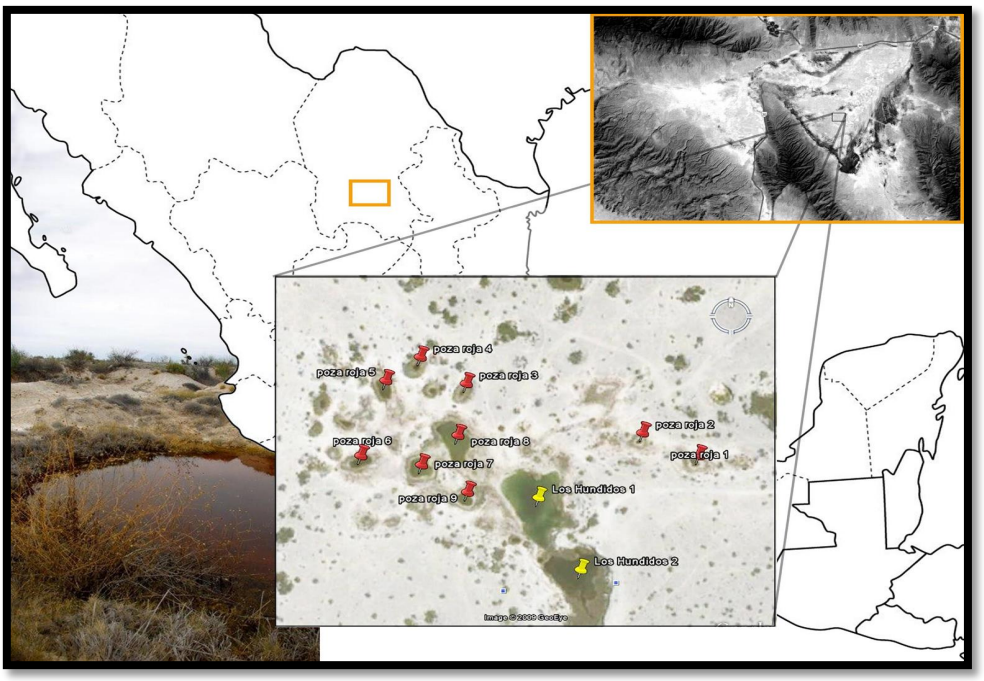


Figura 5. Sistema de flujo pozas rojas. Se observa un mapa de la República Mexicana, indicando donde se encuentra Cuatro Ciénegas y dentro el sistema de pozas rojas.

Los estudios de diversidad a nivel comunidades señalan que los tapetes microbianos del sistema acuático de Los Hundidos están dominados por bacterias del género *Pseudomonas*, a la vez que se resalta como el sitio con mayor desbalance estequiométrico 157N:1P dentro del valle (Bonilla-Rosso *et al.*, 2012; Peimbert *et al.*, 2012). Partiendo de dichos antecedentes y tomando en cuenta que lo que se refleja en los estudios metagenómicos no necesariamente se refleja a nivel de cultivo, el presente estudio inicio con el objetivo de generar una colección de aislados cultivables que permitiera hacer experimentos que ayudaran a corroborar hipótesis planteadas en los estudios previos. Sin embargo, durante los muestreos exploratorios durante los años 2011 y 2012, no se logró la recuperación de cepas del género *Pseudomonas*a pesar de que se utilizaron medios específicos, como lo fue el medio *Pseudomonas* isolation agar (PIA). En estos medios se recuperaron consistentemente los géneros *Aeromonas*, *Halomonas* y *Vibrio*. Dada esta eventualidad se exploró la diversidad del genero Vibrio a mayor profundidad.

Para el año 2013 se hizo un muestro específico para recuperar *Vibrio*, en este muestreo se utilizó un medios específicos para *Vibrio*, Thiosulfate-citrate-bile salts-sucrose agar (TCBS) y PIA donde previamente habían crecido los aislados. Así mismo se tomaron muestras ambientales para determinar parámetros fisicoquímicos. Adicionalmente, se realizaron mediciones de las concentraciones de

nutrientes en el sitio, observándose un aumento respecto a los previamente reportados (Tabla 3; Bonilla-Rosso *et al.,* 2012; Peimbert *et al.,* 2012). En 2010 se tiene registrada la presencia de un huracán en las costas del Golfo de México, este evento fue de tal magnitud que impacto la zona donde se encuentra Cuatro Ciénegas, inundando el sistema acuático de Los Hundidos. El impacto en la diversidad a nivel de la comunidad bacteriana es desconocido, sin embargo se observó un enriquecimiento de nutrientes en las pozas (Tabla 3). Esto sugiere que la comunidad pudo haber sufrido un cambio donde el incremento de nutrientes permitiera que miembros poco abundantes de la comunidad en condiciones normales se volvieran abundantes, algo así como el surgimiento de la biosfera rara (Sogin *et al.,* 2006). Sin embargo, existe la posibilidad de que sean miembros nuevos que llegaron con el arrastre de las lluvias. Ambas posibilidades podrían ser una explicación al hecho de no poder encontrar *Pseudomonas* y si *Vibrio* a nivel de bacterias cultivables.

Patrones similares se han observado en años recientes en otro sistema acuático llamado Churince. Donde a raíz de la perdida de agua se ha observado la disminución de *Pseudomonas* en cultivo y el aumento de *Vibrio*, en este caso aislados de *Vibrio cholerae*.

Tabla3. Relación de C:N:P en las pozas rojas en el tiempo.

Año de muestreo	**Tipo de muestra**	**Relación C:N:P**
2008	Tapete microbiano	15820:157:1
2013	Agua	350:9:1
2013	Sedimento	258:21:1

Observamos que la distribución de los nutrientes no era homogénea entre las dos interfaces muestreadas dentro de las pozas. Un análisis de componentes principales, mostro dos grupos principales asociados a la procedencia de la muestra, agua o sedimento (Figura 6). Los valores de nutrientes en la interfaz del agua presentan mayor dispersión y son más altos que en los sedimentos. Dadas estas características se definieron dos categorías para los análisis posteriores alto y bajo, que correlaciona con la interfaz agua y sedimento.

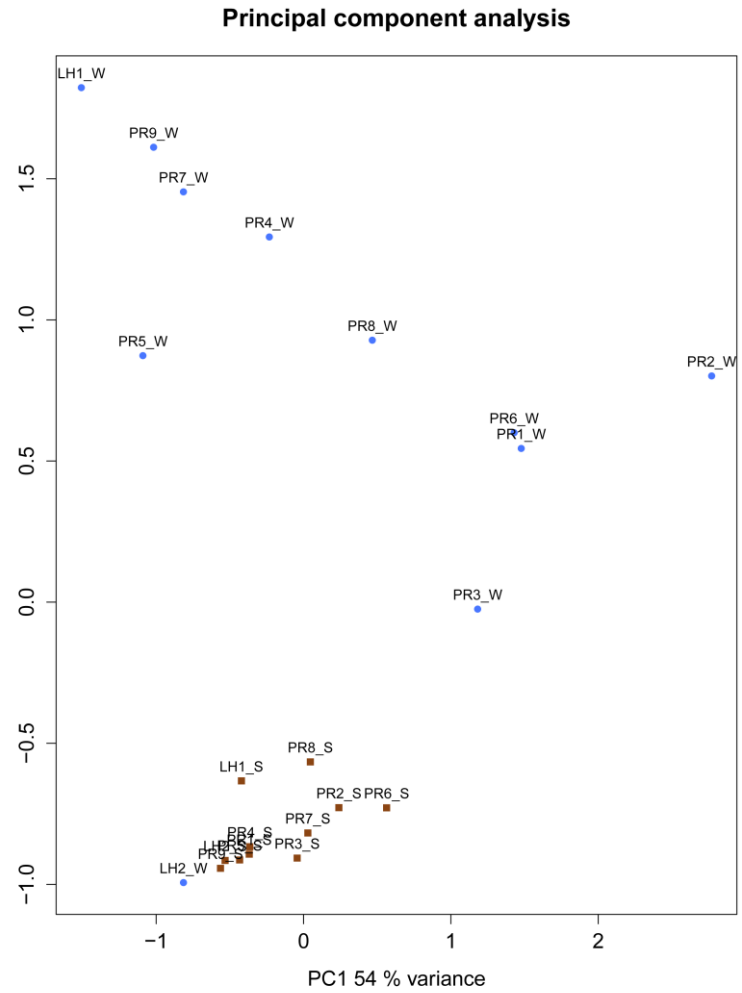


Figura 6. Análisis de componentes principales de las variables de nutrientes.

En total se obtuvieron 183 aislados, de los cuales solo 101 pertenecen a la familia Vibrionaceae. La identificación fue hecha atreves de la secuenciación parcial del gen 16S rRNA. Los linajes que se recuperaron eran cercanos a *V. cholerae*, *V. alginolitucus*, *V. parahaemoliticus*, *V. anguillarum* y *Photobacterium*. Con esta información se hicieron los primeros análisis referentes a la distribución de estos organismos dentro del sistema acuático Los Hundidos. Lo primero que quisimos saber con estos datos fue si la distribución de los linajes descritos de Vibrionaceae era aleatoria o si su distribución era dependiente de algún componente geográfico o ambiental. Para contestarnos esta pregunta, hicimos un análisis que infiere ecotipos, AdaptML (Hunt *et al.*, 2008) este programa permite predecir ecotipos a partir de la información genética y ambiental. Se utilizaron como variables la distancia de las pozas, las concentraciones de nutrientes y el ambiente de donde se obtuvo la muestra (agua/sedimento; Figura 7). Lo que observamos fue que los linajes se estructuran según la interfaz agua o sedimento, en la mayoría de los casos. Sin embargo, también observamos que el clado

más grande y abundante, que se conforma por miembros filogenéticamente cercanos a *V. anguillarum*, es generalista pues no está asociado específicamente a ninguno de los ambientes (agua/sedimento). Es decir dos estrategias ecológicas distintas dentro de los linajes de Vibrionaceae en los Hundidos, generalistas y especialistas.

Para el género *Exiguobacterium* se observó un patrón similar de asociación de filogrupos al ambiente agua o sedimento dentro del valle de Cuatro Ciénegas. Se ha propuesto que un ambiente como el agua representa hábitats menos estructurados donde la distribución de los recursos es homogénea debido a las corrientes, mientras que los sedimentos son ambientes estructurados donde la distribución de los recursos es heterogénea (Rebollar *et al.*, 2012). Por lo que, en el caso de Vibrionaceae probablemente la distribución que encontramos relacionada con los linajes tiene que ver con la estructura de los ambientes.

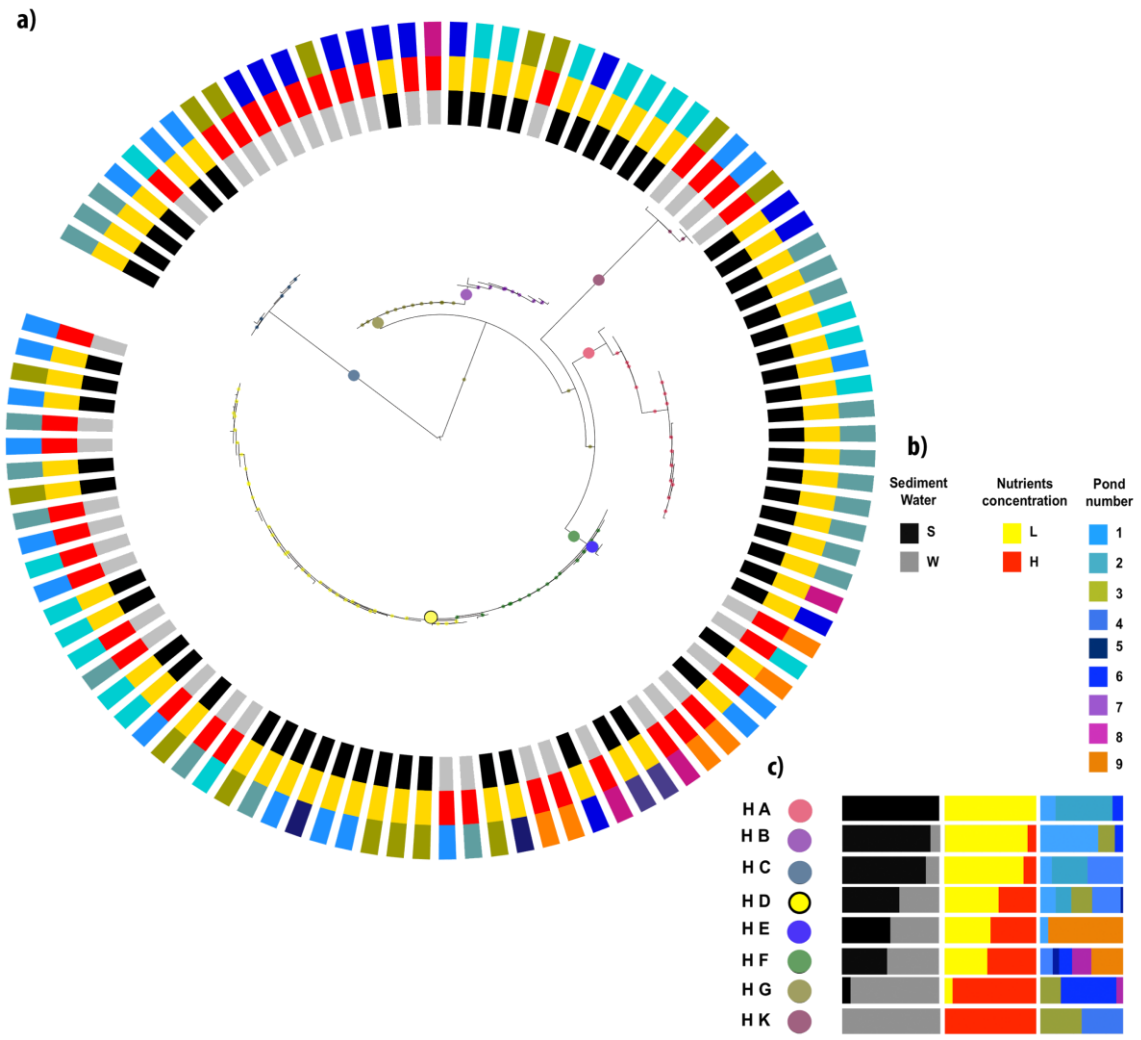


Figura 7. AdaptML. En el panel a) se muestra una filogenia construida con máxima verosimilitud, los puntos internos señalan los hábitats inferidos por el programa, el primer círculo interno indica agua (gris) y sedimento (negro), el segundo circulo indica concentraciones de nutrientes altas (rojo) y bajas (amarillo), el ultimo circulo indica el número de poza donde fueron aislados. El panel b) indica el código de color y el panel c) indica la proporción de cada variable ambiental en cada hábitat predicho.

3.2 Artículo 1. (En preparación). Insights on the role of recombination and effective population size in pan-genome. Lessons from Vibrionaceae within an endangered oasis.

En el siguiente apartado se retoman los análisis y las conclusiones de la sección anterior para contestar preguntas alrededor de los procesos evolutivos que moldean los linajes de Vibrionaceae en Cuatro Ciénegas. Cuando nos referimos a procesos evolutivos estamos hablando de la influencia de la selección natural, la deriva génica, la mutación y el flujo genético.

En esta segunda parte, se utiliza información obtenida de la secuenciación masiva de 42 aislados del género Vibrio y Photobacterium, derivados de los análisis anteriores. El criterio de selección de aislados se basó en su distribución en los dos ambientes, agua y sedimento, así como su filiación en los diferentes clados. Este análiss se llevó a cabo para entender si los estilos de vida generalistas o especialistas tienen una influencia en la diversidad genética y si éstos se ven reflejados en el comportamiento de los procesos evolutivos. Los análisis propios referentes a este apartado se detallan en el artículo que se presenta a continuación.

Artículo 1. (En preparación). Insights on the role of recombination and effective population size in pan-genome. Lessons from Vibrionaceae within an endangered oasis.

Mirna Vazquez-Rosas-Landa[1], Luis E. Eguiarte[1*], Jonás Aguirre-Ligori[1], Gabriel Yaxal Ponce-Soto[1], Shalabh Thakur[2], Enrique Scheinvar[1], David S. Guttman[2,3], and Valeria Souza[1*]

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

[2] Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada

[3] Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada.

Abstract 275 words

Prokaryotes have a wide range of genetic variation as has been shown in pan-genome studies. Nevertheless, the evolutionary forces that shape this genetic diversity have not been explored in detail. For instance, recombination may contribute to substantialy to pan-genome adding new variation through horizontal gene transfer; nevertheless, it can also act as a homogenizing force within the population. On the other hand, populations with large effective population sizes commonly have large accessory genomes. In this work, we explore the role of recombination and effective population size on the pan-genome of 42 Vibrionaceae bacterial strains isolated from an oligotrophic lagoon within Cuatro Cienegas, Mexico, a relatively isolated aquatic ecosystem surrounded by deserts and high mountains. We identified six clades within our strains, mostly defined by their isolation environment, except clade II identified as a generalist clade. Clade II showed the lower *r/m* value, 0.88, while clade IV showed the higher *r/m* value, 7.94. Despite its high *r/m* values, clade IV show a small N_e with a median value of 383,067, while clade III has the higher N_e with an amount of 15,018,880. Pan-genome metrics show that only clade II has open pan-genome, this particular clade was composed of at least three sub-clades with the lowest *r/m* and N_e values. Our results show that not always sizeable effective population sizes are correlated with open pan-genome and we suggest that the demographic history of populations, such as recent adaptive radiation could affect this correlation. We have shown that bacteria that habit extreme environments can show sizeable effective population sizes, which suggest that evolution by natural selection could be the driving force, also affecting the evolution of their pan-genome.

Introduction

Comparative genomics analyses have shown a wide range of genomic variation within bacteria at different phylogenetic levels. This variety could be explained by the ecological characteristics of each bacterial group (Lapierre & Gogarten, 2009; Collins & Higgs, 2012; Gordienko et al., 2013; Zhi et al., 2017). Nevertheless, the evolutionary mechanisms and dynamics that regulate the genome size, number of genes and diversity within each gene are still poorly understood. Tettelin *et al.* (2005) first described the pan-genome concept, showing that gene content among bacteria of the same species could be very variable. A closed pan-genome was defined as when the analyzes of new genomes of a bacterial species do not reveal new genes for the genetic pool of a species, while in an open pan-genome, every newly sequenced genome increases the total genetic pool (Tettelin *et al.,* 2005). It has been proposed that evolution by natural selection would be the most plausible explanation to the pan-genome (McInerney *et al.* 2017). If genetic drift were driving gene fixation, prokaryotic genomes would accumulate in times more and more genes tending to become as large as eukaryotic genomes, but instead, they remain in the range (Konstantinidis and Tiedje 2004). As well, empirical genome analyses have demonstrated that prokaryotic genomes are biased towards deletion of DNA if there are no selective pressures to maintain a given gene (Kuo and Ochman, 2009; Morris *et al*., 2012; Mas *et al.,* 2016).

Recently it has been shown that large effective population sizes in bacteria tend to be correlated with a large accessory genome (Andreani *et al.,* 2017). Nevertheless, the effect of other evolutionary forces, such as recombination on the shape of pan-genome has not been assessed. Recombination is one of the most critical evolutionary forces in bacteria evolution because it can introduce entirely new genetic material by horizontal gene transfer (HGT), or new alleles by homologous recombination within populations. Furthermore, recombination can homogenize populations, keeping cohesive groups (Souza *et al.*, 1992; Smith *et al.,* 1993; Souza & Eguiarte, 1997), and as a result slowing down diversification (Lawrence and Ochman, 1998).

Depending on the impact of recombination within bacteria population, we can describe two main microevolutionary models. In one hand, bacteria with low levels of recombination would follow the classical ecotype model (Cohan; 2001), where recombination between populations is uncommon, and mutations are the primary sources of diversity. This process is dominated by the intense natural selection that reduces all genetic variation in a process sometimes called clonal interferences or

periodic selection (Atwood *et al*., 1951; Guttman & Dykhuizen, 1994; Cohan; 2001; Cohan & Perry, 2007; Koeppel et al., 2008). On the other hand, in bacteria with high levels of recombination, the gene association brakes, and natural selection increases the frequencies of linked-loci to fixation, while the rest of the genome may remain the same. Diversification in this situation occurs by the decreasing of genetic exchange mediated by selection (Cadillo-Quiroz *et al.,* 2012; Shapiro *et al.,* 2012; Cordero & Polz, 2014).

From a genomic perspective, we could expect that bacteria with high homologous recombination and low HGT to show mainly closed pan-genome, whereas lineages with lower homologs recombination and higher HGT would have sizeable open pan-genomes. Hence, for a proper understanding of bacterial evolution and their pan-genome, it is necessary to describe if there is a correlation between natural selection, recombination rates and effective population size with pan-genome sizes.

To explore those ideas, we chose bacteria isolates from the Cuatro Cienegas Basin (CCB), a unique oasis in the Chihuahuan desert. It is composed of several aquatic systems with high microbial diversity and a particular stoichiometric condition, where phosphorous is extremely limited leading to a significant imbalance of nutrients (C:N:P) (Elser *et al.,* 2005; Tapia-Torres *et al.,* 2015; Valdivia-Anistro *et al.,* 2016). Recent population genetic studies of bacteria isolated from this extremely oligotrophic site show low levels of recombination in different lineages such as *Exiguobacterium*, *Bacillus,* and *Pseudomonas* (Escalante, 2008; Rebollar *et al.,* 2012; Avitia *et al.,* 2014). We proposed that nutrient constraints in this environment work as an ecological filter, reducing recombination due to the cost of replicating new DNA (Souza *et al.,* 2008; Souza *et al.,* 2012). As well, the metabolic co-dependence between neighbors builds strong community cohesion, apparently inhibiting efficient migration into the bacterial populations of CCB (Espinosa-Asuar *et al.,* 2015; Pajares *et al.,* 2016). As a result, we have suggested that mutation and natural selection, but no recombination, are the primary evolutionary forces shaping microbial diversity within CCB.

We used Vibrionaceae isolated from CCB to explore the role of the different evolutionary forces that shape the pan-genome. We decided to use these bacteria group given that most of the species within Vibrionaceae present high recombination rates along with large effective population sizes in the oceans (Vos & Didelot, 2009; Cui *et al.,* 2015). Based on those facts, we drive two hypothesis on the pan-genome behavior of CCB Vibrionaceae. In a null model, recombination is as high as has been

shown for other Vibrionaceae species in the oceans, which would be reflected as an open pan-genome. On an alternative model, based on previous population genetics studies, we could expect low recombination rates probably due to the high P limitation (Escalante, 2008; Rebollar *et al.*, 2012; Avitia *et al.*, 2014) which would lead to closed pan-genome. To explore these hypotheses, we performed comparative genomics analyses of 42 Vibrionaceae isolated from Cuatro Cienegas to examine the role of recombination and effective population size in its diversity. Our results show that recombination in CCB Vibrionaceae is lower than in the sea respect to what has been observed with MLST (Vos & Didelot, 2009) but in the range with the analysis of whole genome sequence we performed. Recombination among N_e values found to support the idea of evolution by natural selection, probably related to the site particular stoichiometric unbalance, leading, in general, to ecological specialization reflected in a close pan-genome pattern. However, the pan-genome of the nonspecialized clade is not closed, probably due to both subsampling and actual diversification process within of this particular population in our site.

Methods

Site description

We analyzed isolates from a particular site within CCB named Pozas Rojas, because of the color of the water. This aquatic system has been described as hypersaline in summer (Rebollar *et al*., 2012), with high fluctuations in temperature from winter to summer (around 1-60 °C; Peimbert *et al*., 2012). This site is composed of several small ponds that surround a bigger lagoon in a region named Los Hundidos (LH, Peimbert *et al*., 2012; Rebollar *et al*., 2012). The ponds are small but permanent and are separated from each other by ca. 90 meters along an arch around a larger lagoon. However, during summer 2010 this system was flooded by hurricane Alex, merging most of the ponds into a single bigger lagoon, until autumn 2011 when the water receded, leaving the moon shaped array of small red ponds in the same place (Figure 1).

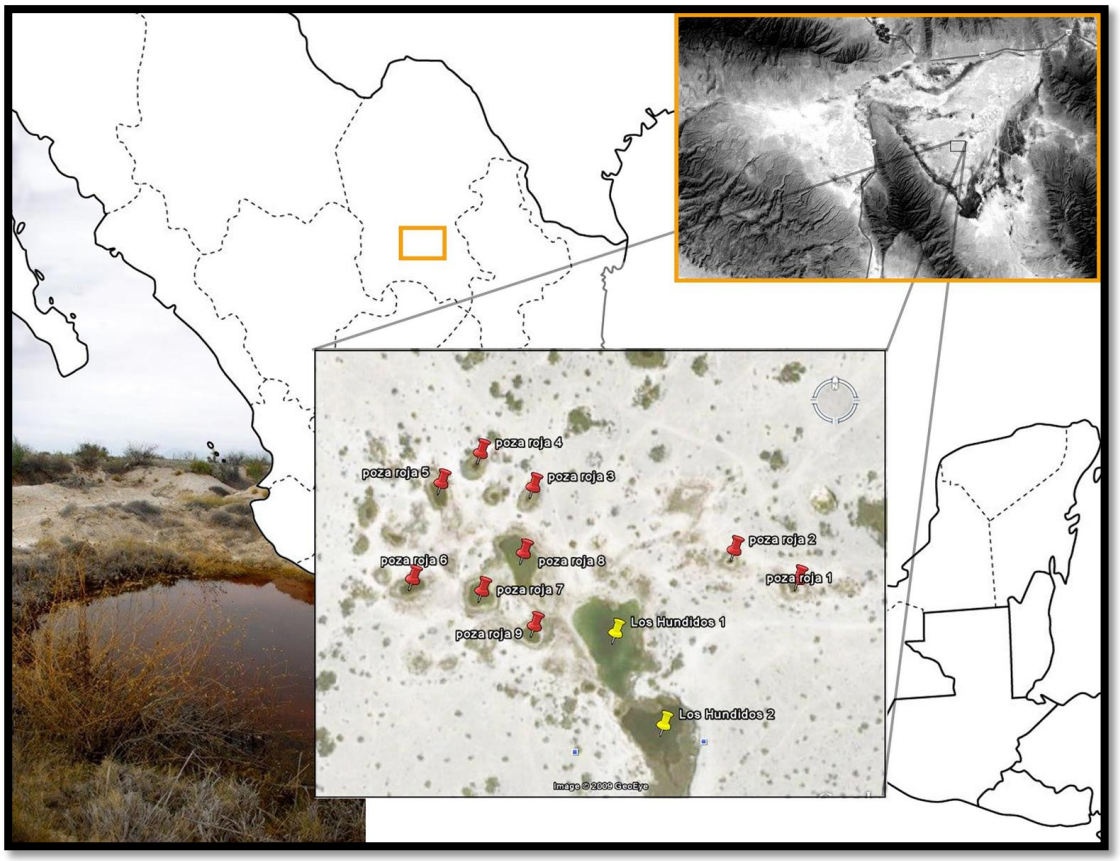


Figura 1. Sample site. Sampling sites are signal in red; a map is also show, which indicates location of Cuatro Ciénegas.

Sample collection and strains isolation.

Water and sediment samples were collected in duplicate from nine ponds (locally called pozas) located in Los Hundidos during March 2013 and stored at 4 °C until processing. Sediment was collected for nutrient analysis in 50 ml Falcon tubes and covered with aluminum foil before storage. Water was collected for nutrient quantification in 1 liter volumes and stored in the dark at 4 °C. Analyses were performed at the Instituto de Investigaciones en Ecosistemas y Sustentabilidad, UNAM, in Morelia, Mexico.

Cultivable strains were isolated in PIA (*Pseudomonas* isolation agar) and TCBS (Thiosulfate Citrate Bile Sucrose Agar) as previously described (Vázquez-Rosas-Landa *et al.*, 2017), obtaining a total of 183 isolates.

Environmental variables measurement.

For nutrient quantification, sediment samples were dried while water samples were filtered through a Millipore 0.42 µm filter. Total carbon (TC) and inorganic carbon (IC) were determined by combustion and colorimetric detection (Huffman, 1977), using a full carbon analyzer (UIC model CM5012, Chicago, USA). Total organic carbon (OC) was calculated as the difference between TC and IC. For total N (TN) and total P (TP) determination, samples were acid digested with H_2SO_4, H_2O_2, K_2SO_4 and $CuSO_4$ at 360°C. Soil N was determined by the macro-Kjeldahl method (Bremmer, 1996), while P was determined by the molybdate colorimetric method following ascorbic acid reduction (Murphy & Riley, 1962). The N and P forms analyzed were determined colorimetrically in a Bran-Luebbe Auto analyzer 3 (Norderstedt, Germany).

To explore the relationship between biogeochemical variables, ponds, and environment, we performed a principal component analysis, and based on this we conducted Student t-tests to analyze if there were differences between water and sediment in biogeochemical variables. All these analyses were performed using R (R core team) and the psych package (Gricelle, 2001; Jollife, 2002).

DNA Extraction and PCR Amplification of 16S rRNA.

For the 183 isolates obtained, DNA extraction was performed as described by Aljanabi & Martinez (1997). 16S rRNA genes were amplified using universal primers 27F (5′-AGA GTT TGA TCC TGG CTC AG-3′) and 1492R (5′-GGT TAC CTT GTT ACG ACT T-3′) (Lane, 1991). All reactions were carried out in an Applied Biosystems Veriti 96 Well Thermal cycler (California, USA) using an Amplificasa DNA polymerase (BioTecMol, Mexico) with the following program: 94°C for 5 min, followed by 30 cycles consisting of 94°C for 1 min, 50°C for 30 s, 72°C for 1 min and 72°C for 5 min. Polymerase chain reaction (PCR) amplification products were electrophoresed on 1% agarose gels. Sanger sequencing was performed at the University of Washington High-Throughput Genomics Center.

Phylogenetic analysis of 16S rRNA sequences.

Partial 16S rRNA gene sequences were alignment, and quality control were performed with Mothur (Schloss et al., 2009) and Clustalw (Larkin et al., 2007), and the alignments were manually revised. Genera level identification of the strains was made using the classifier tool (Wang et al., 2007) from the Ribosomal Database Project (RDP) Release 10, update 30 (Cole et al., 2009). Blastn searches were performed against Refseq_rna database from NCBI to select reference sequences. A total of 101 sequences were identified as a member of the Vibrionaceae family and were selected for subsequent analyses.

A maximum likelihood phylogenetic reconstruction was obtained with PhyML version 3.0 (Guidon et al., 2010) under the HKY+I+G substitution model estimated with jModelTest 2 (Darriba et al., 2012). The degree of support for the branches was determined with 1,000 bootstrap iterations. A *Halomonas* strain isolated from the same sampling was included as an outgroup.

Environmental association of phylogroups.

To test whether the community of cultivable strains was structured based on environmental variables, we performed an AdaptML analysis (Hunt *et al.*, 2008). A total of 101 isolates belonging to Vibrioneaceae and a *Halomonas* strain as outgroup were selected. Three categorical environmental variables were tested: pond of isolation, high and low nutrient concentrations, variables chosen from the PCA analysis described above, and the two sampled environments (water and sediment).

Genome sequencing, assembly, and annotation.

We selected 39 *Vibrio* isolates and 3 *Photobacterium* for whole-genome sequencing. DNA extractions were performed with the kit DNeasy Blood and Tissue (Qiagen). Quality and DNA concentration was estimated using a NanoDrop Lite (Thermo scientific). Sequencing was performed with Illumina MiSeq 2x250 technology, with insert libraries of 650 bps and an expected coverage of ca. 10x per genome. For the strain V15_P4S5T153, a second library was designed, this library was sequenced using the Jr 454 Roche technology. Sequencing was performed at the LANGEBIO, México.

Genomes assembly was performed with Newbler (Roche/ 454 Life Sciences). For scaffolding process we used SSPACE (Boetzer, 2001), gaps were closed using GapFiller (Nadalin, 2012) and final error correction was performed with iCORN (Otto et al., 2010). Coding sequences were inferred with Prodigal 2.0 (Hyatt et al., 2010) implemented in PROKKA software (Seemann, 2014). Each genome was annotated using InterProScan 5 (Jones et al., 2014) with the databases enabled by default.

Pangenome analysis

For the comparative genomics analysis, we used 42 strains of Vibrionaceae obtained from nine ponds in Pozas Rojas, following AdaptML results. Of these, 39 isolates belonging to the genera *Vibrio*, and three strains belong to *Photobacterium* (Vázquez-Rosas-Landa *et al.*, 2017). For comparison we also included genomes from fives reference *Vibrio* strains: V*ibrio alginolyticus NBRC 15630 ATCC 17749*, *V. anguillarum 775*, *V. furnissii NCTC 11218*, *V. parahaemolyticus BB22OP* and *V. metschnikovii CIP 69 14* (Supplementary table 1). Ortholog gene families were predicted from all 47 genomes using the DeNoGAP comparative genomics pipeline (Thakur & Guttman, 2016). As required by DeNoGAP to minimize false positive prediction of orthologs, we assigned *Photobacterium* genomes as outgroup. The completely sequenced genome of *V. anguillarum* strain 775 was used as seed reference (Supplementary table 1).

Following the prediction of ortholog families, we constructed a phylogenetic profile based on presence and absence of gene families across the genomes. This information was then used to classify ortholog families as the *core*, *accessory* and *unique*. Since most of the genomes in our dataset are not completely sequenced, we designate core ortholog families as those present in at least 95% of the genomes, to avoid the impact of missing genes due to sequencing or assembly artifacts. The accessory ortholog families were designated as those that were present in at least two genomes, but in less than

95% of the genomes. From the phylogenetic profile inferred with DeNoGAP, we extracted the *core*, *flexible* and *unique* genes per clade excluding references. With this information, the package Micropan (Snipen & Liland, 2015) within R was used to infer the open or closed nature of each pangenome dataset, following the heaps law proposed by Tettelin et al. (2008).

Core proteins were aligned using Kalign (Lassmann et al., 2009) to infer organism phylogeny. The resulting alignments of individual ortholog families were concatenated using a custom Perl script. A maximum likelihood phylogenetic tree was constructed based on concatenated alignment using FastTree program (Price et al., 2010). Six clades were defined from the estimated phylogeny (Figure 2).

Genetic diversity parameters.

We calculated for each coding domain sequence (CDS) on each clade, Pi, Theta and Tajima's *D* using BioPerl and the PopGen module (Stajich, 2005), implemented in an in-house script. For each clade, each CDS family was aligned using MUSCLE (Edgar, 2004) with the codon-alignment option performed in TranslatorX (Abascal et al., 2010). All the alignments were further improved by removing sequences with a large number of gaps using the MaxAlign program (Gouveia-Oliveira et al., 2007).

Recombination analysis of orthologues genes.

Of the total ortholog families in the *Vibrio* pangenome, we only used ortholog families found in at least three genomes for recombination analysis. Genetic recombination was examined on each CDS alignment by using inference of pairwise recombination sites, obtained with GENECONV (Sawyer, 1989) and by the identification of putative recombinant sequences through breakpoints using GARD (Kosakovsky et al., 2006). A whole-genome alignment for the 47 analyzed genomes was performed with MAUVE (Darling et al., 2010). The resulting alignment was used as input for Gubbins (Croucher et al., 2014) using RAxML (Stamatakis, 2014) and default parameters. Additionally, whole genome alignments were performed for each clade, excluding references, with progressive MAUVE (Darling et al., 2010). Each clade alignment was used as input for ClonalFrameML (Didelot & Wilson, 2015). Ancestral recombination events were inferred with the use of fastGEAR (Mostowy et al., 2017) with 15 iterations. The whole genome alignment previously obtained with Mauve was used as input.

Clustering analysis of clade II

Clade II is the most abundant lineage within Pozas Rojas. Recombination analysis showed that there are internal groups of recombination within this clade. To further explore the genetic structure of Clade II, the isolates were classified using a cluster analysis. For cluster analysis, we used Nei (1978) genetic distance and neighbor joining. Genomes with distance less of 0.001 were grouped and proved with a discriminant analysis of principal component analysis of the genetic variation, using the adegenet library in R (Jombart & Ahmed, 2011). For this study, we use 20 components and 3 discriminant functions.

Effective population size estimation

We followed an ABC framework approach to estimate the posterior distribution of effective population size (N_e) of each of the six clades. According to the previous clustering and recombination analysis, for clades I, III, IV, V and V1 we simulated a single population while for clade II we simulated three populations that diverged from an ancestral population. Simulations were performed using Fastsimcoal2 (Excoffier & Foll, 2011; Excoffier *et al.*, 2013). For each clade we simulated DNA sequences having a similar length equal to the number of nucleotides in the given clade, as well as a sample size equal to the number of sequences sampled for each clade. We assumed no recombination within the genome and used the *Escherichia coli* mutation rate of 2.2×10^{-10} mutations per nucleotide per generation (Lee *et al.*, 2012). We ran between two and four simulations for each clade. For the initial runs, we generated 100,000 replicates extracting N_e values from a prior distribution that ranged from 100,000 to 20,000,000 individuals, having a log-uniform distribution. For clade II, we also estimated the time of divergence of each sub-clade, by setting the prior distribution of time ranging from 1,000-4,000,000 generations. After a first run, we narrowed the prior ranges based on those simulations that had similar summary statistics compared to the observed data and performed another 100,000 simulations using the narrowed priors.

To compare simulated and observed data based on summary statistics we used the ape and pegas libraries in R to estimate the number of polymorphic sites and the entire genomes Tajima's *D.* Also we obtained 1000 slide windows to estimate the Tajima's *D* along the genomes, as well as the mean and standard deviation of Tajima's *D*. Tajima's *D* is commonly used to estimate demographic changes in populations. Based on the summary statistics, we used the abc function in the ABC package to

calculate the distribution of the N_e parameter based on a 0.05 % threshold distance between the simulated and observed data. For each clade, we report the median and the 95% interval confidence of N_e. For clade II we further report the average and 95% interval confidence of the number of generations since each sub-clade diverged from an ancestral clade.

Selection and enrichment analysis

We analyzed for signatures of positive selection among ortholog families found in at least three genomes using FUBAR program. We accounted for recombination breakpoints in the ortholog families while calculating positively selected sites based on GARD results (Kosakovsky et al., 2006). We considered any site to be positively selected if it showed P-value $\leq$ 0.05. To identify over-represented functions among genes with positive selection signals, we performed enrichment analyses with Fisher test implemented in R package and the InterProScan 5 annotation results.

Results

Biodiversity of Vibrionaceae in Pozas Rojas.

Before the hurricane Alex in summer 2010, the Pozas Rojas aquatic system had the stronger stoichiometric nutrient unbalance of the valley, with a ratio of C:N:P 15820:157:1 in sediments (Peimbert *et al*., 2012). After the drainage of the storm nutrient conditions became richer since at the time of the sampling the average nutrient ratio of C:N:P was 350:9:1 for water and 258:21:1 for sediment greatly reducing the stoichiometric unbalance for P, and limiting N.

The diversity analysis with the 16S rRNA gene found seven different Vibrionaceae lineages, most of which were closely related to *V. cholerae, V. alginolyticus, V. parahaemolyticus, V. ordalii, V. anguillarum, V. metschnikovii* and *Photobacterium* (Supplementary figure 1). Two of these lineages were not closely related to any genera within databases. Looking for a relationship between environmental variables and Vibrionaceae lineages, we performed an AdaptML analysis (Hunt *et al*., 2008). We found that lineages are structured according to their environment --water or sediment, -- showing two ecological strategies: most clades were specialist either to water or to sediment, while the most abundant lineage had a generalist ecology (Supplementary figure 2).

Pan-genome analysis of CCB Vibrioneaceae and lineages description.

From previous results, we select for sequencing 42 isolates that were found in water or sediment. The pan-genome analysis of 39 CCB *Vibrio*, 3 CCB *Photobacterium*, and 5 *Vibrio* references strains is composed of 20,121 orthologous gene families (Supplementary table 1). The genes that were present in at least 95% of the genomes conformed the core genome, which for this sample is composed of 1,254 genes, including reference genomes. The accessory genome is conformed of genes that were found in 2 to 44 genomes, 14,072 genes follow these criteria. The rest 4,795 genes were identified to be present only in one strain.

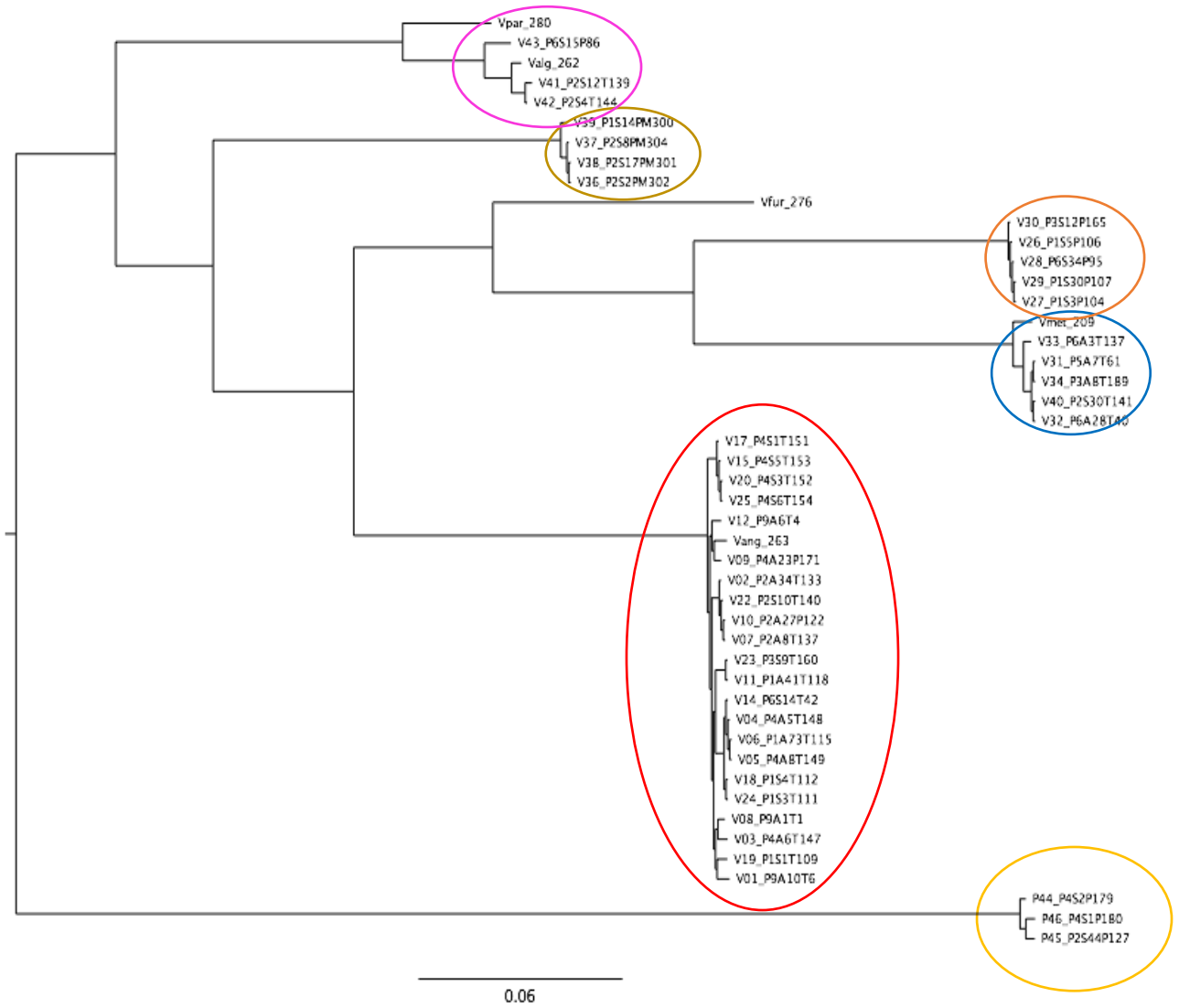
Core phylogeny was made with 1,254 genes (Figure 2). With this information, we identified the same six lineages observed with 16S rRNA gene (Supplementary figure 2) and a seventh clade, which is represented by a unique strain of marine *Vibrio furnissii sp.* Nov. 4 strain (NCTC 11218) (Lux *et al.,* 2011). In general, long branches were observed between coexisting lineages from CCB. In

particular, reference strain *V. anguillarum 775* isolated from a Coho salmon (Naka *et al.*, 2011) clusters within the large generalist clade II, while reference strain *V. metschnikovii CP 69-14*, which has been isolated in aquatic systems, is basal to clade III. Basal to clade VI, are reference strains *V. parahaemoliticus BB22OP* a pre-pandemic strain (Xu *et al.*, 2015) associated with seafood-borne gastroenteritis in humans and *V. alginolyticus NBRC 15630 = ATCC 17749*, an aquatic organism which in some cases can cause bacteremia. Clade 4 members are exclusive to CCB, given that there is no closely related strain sequenced on databases (Supplementary figure 3), same is the case for clades V and I, the last being closely related to *Photobacterium*. For each of the six defined clades with CCB strains, a classification of *core*, *accessory* and *unique* ortholog families was performed (Table 1).

Table 1. Pan-genome metrics of each clade. First column show the number of genomes used for the analysis, next there are the general metrics of pan-genome and last columns show the heaps values obtained.

Group clade	Number of CCB genomes included in each clade	Pan-genome metrics				Heaps law parameters	
		Core	Flexible	Unique	Total number of genes	Intercept value	Alpha
Clade 1	3	3617	346	603	4566	692.8508	1.1293
Clade 2	22	1746	5770	1745	9261	244.2096	0.7913
Clade 3	5	2672	718	324	3714	658.0634	1.6625
Clade 4	5	2055	1445	180	3680	2726.7580	2.0000
Clade 5	4	2853	1660	1332	5845	1196.2571	1.3109
Clade 6	3	2448	3476	1028	4992	3295.5770	2.0000
Vibrionaceae all clades	47	1254	14072	4795	20121	2263.7472	0.6621

Figure 2. Core gene phylogeny. Phylogenetic reconstruction of core genes. Circles represent different lineages and numbers represent the name of the lineage.



With this information, we used the Heaps law as proposed by Tettelin et al. (2008) to classify pan-genome of each clade as closed or open. The Heaps law model is fitted to the number of new gene clusters observed when genomes are ordered randomly. The model has two parameters: an intercept and a decay parameter called alpha: if alpha >1.0 the pan-genome is closed, if alpha <1.0 it is open. For the whole dataset, we observed an open pan-genome, with an alpha of 0.6621. Clade II has an alpha of 0.7913 and therefore it is considered an open pan-genome. The rest of the clades display a closed pan-genome pattern, suggesting that new genomes do not add more genes to the pan-genome in these clades (Table 1).

Genetic diversity parameters

We further decided to explore the genetic diversity within two sections of the pan-genome, *core* and *flexible* genes in each clade, to study whether any section had more genetic diversity and if it was constant over the clades (Table 2). In the case the *core* and *flexible* genes, clade IV had the lowest average diversity (average π= $8.08x10^{-4}$ and $1.48\ x10^{-3}$, average $\theta w = 5.21x10^{-4}$ and $9.15x10^{-4}$, respectively), while clade VI showed the highest average diversity (average π = 0.034 and 0.044, average θw = 0.027 and 0.035 respectively). As expected, flexible genes were always more diverse than core genes.

Table 2. Genetic diversity parameters. First panel shows nucleotide diversity and polymorphic sites for core genes within each clade, second panel show the same type of information for flexible genes within each clade.

Group clade	Core genes					
	π		θw		Polimorphic sites	
	Average	Sd (+/-)	Average	Sd (+/-)	Average	Sd (+/-)
Clade 1	0.0187	0.0489	0.0150	0.0393	32.2651	63.1413
Clade 2	0.0120	0.0177	0.0096	0.0101	51.6971	57.2944
Clade 3	0.0046	0.0157	0.0040	0.0089	13.6781	28.3986
Clade 4	0.0008	0.0102	0.0005	0.0057	1.9260	22.2293
Clade 5	0.0078	0.0334	0.0059	0.0239	15.2702	53.6281
Clade 6	0.0339	0.0301	0.0270	0.0218	77.5680	79.6851

Group clade	Flexible genes					
	π		θw		Polimorphic sites	
	Average	Sd (+/-)	Average	Sd (+/-)	Average	Sd (+/-)
Clade 1	0.0189	0.0460	0.0151	0.0365	28.6580	62.7688
Clade 2	0.0162	0.0343	0.0127	0.0243	45.8390	115.3718
Clade 3	0.0065	0.0239	0.0058	0.0186	14.9526	49.2038
Clade 4	0.0015	0.0180	0.0009	0.0108	2.7446	32.2741
Clade 5	0.0099	0.0389	0.0078	0.0298	16.1830	68.3097
Clade 6	0.0436	0.0424	0.0352	0.0334	72.5932	92.6384

Patterns of homologous recombination

Recombination is a process that can occur either within a bacterial species or among species. In our case, we were interested in exploring the levels of recombination within each clade and among them. To achieve these goals, we analyzed all coding domain sequences (CDS) that were found in at least three genomes to infer homologous recombination with GENECONV (Sawyer, 1989). From 15,380 genes that fulfill these criteria, only 2,928 genes show a significant signal of recombination, representing 14.5% of the pan-genome.

Our results show that there are more recombination events within clades than among clades, including reference genomes (Figure 3a). Also, this analysis shows a sub-structure for clade II, where we were able to delimitate three major genetic groups based using Nei (1978) genetic distance. The isolates that compose Sub-clade II-A are: V15_P4S5T153, V17_P4S1T151, V20_P4S3T152 and V25_P4S6T154, while, sub-clade II-D by the isolates: V04_P4A5T148, V05_P4A8T149, V06_P1A73T115, V14_P6S14T42, V18_P1S4T112 and V24_P1S3T111; and the sub-clade II-G composed by the isolates V02_P2A34T133, V07_P2A8T137, V10_P2A27P122 and V22_P2S10T140 (Figure 3b).One or two strains represented the other sub-clades.

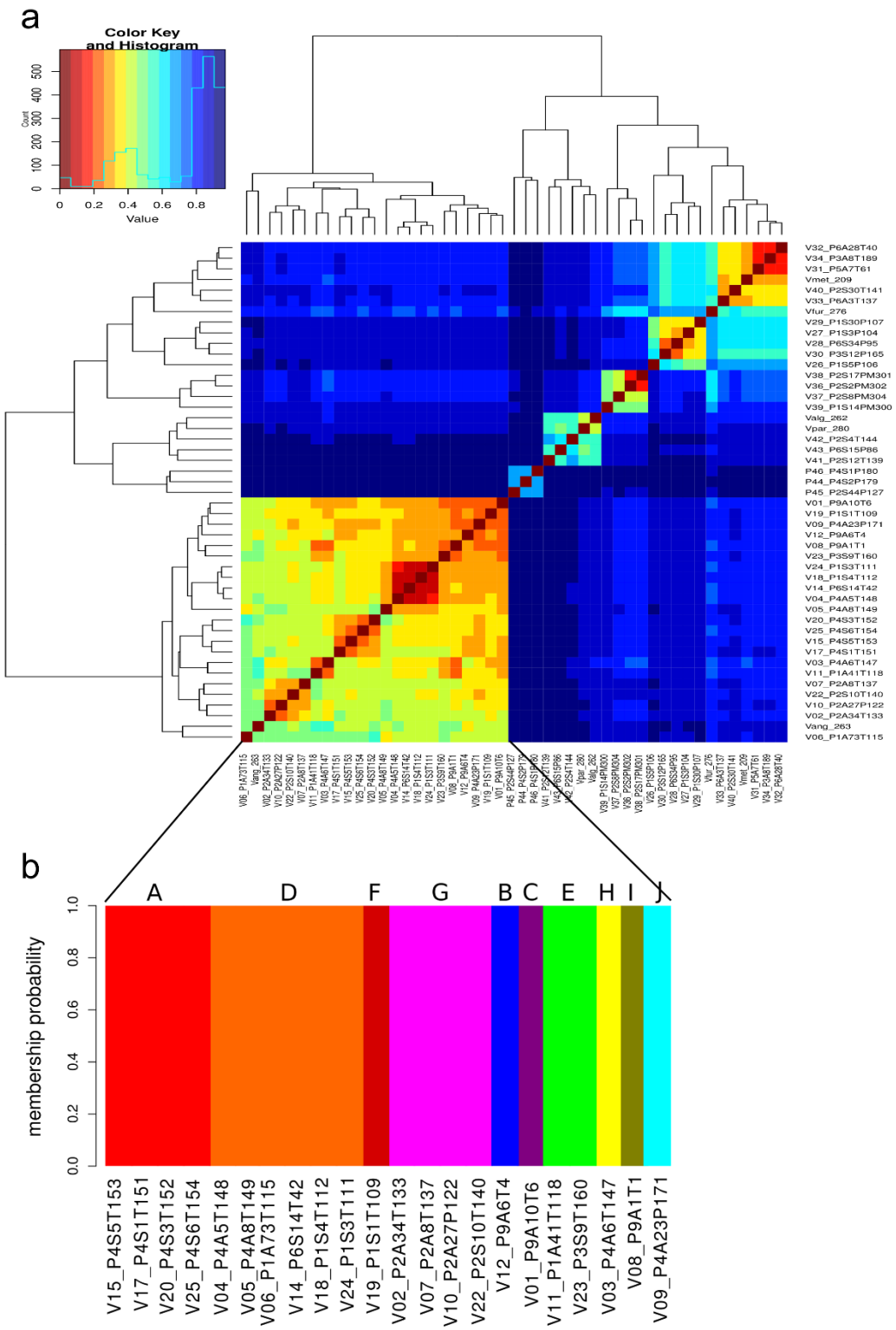


Figure 3. Patterns of recombination. Panel a) shows a *heatmap* of frequency of recombination events among different strains; red colors indicate more recombination events within strains while blue events indicate few recombination events. Panel b) shows clustering of Clade II strains using Nei (1978) genetic distance, each color represent a different genetic group.

Homologous recombination as a source of variation.

We evaluated the impact of homologous recombination and mutation within lineages estimating *r/m* (Didelot & Wilson, 2015); this measure reflects the ratio of probabilities that a given polymorphism is explained by recombination or by mutation (Rebollar *et al.*, 2012). For this analysis, we used whole genome alignment of each CCB clade, excluding references. Clades II and VI display the lowest *r/m* values, 0.88 and 1.13, respectively, while Clades I and IV have the highest values in our dataset, *r/m* 7.94 and 2.98, respectively (Table 3). We performed the same analysis on *V. parahaemolyticus*, *V.ordalii*, *V. anguillarum* and *P. leiognathi* reference genomes (Supplementary table 2), all isolated from marine environments, and their *r/m* estimate were higher than for clades II and 6 (2.84, 1.18, 5.46 and 1.64, respectively; Table 3). Values for these strains had been previously estimated (Vos & Didelot., 2008), however, these values are greater than those estimated in this work. This could be due to the different strategies employed, as previously they were estimated through a multilocus approach in contrast to our whole genome approach. In the same way, these values could be affected by the different sampling sizes for each estimate.

Table 3. Recombination vs mutation estimates. Values are presented for CCB clades but also for references datasets. Rho/theta and r/m estimates are presented for each dataset.

	Recombination vs mutation estimates	
Group clade	rho/theta	r/m
Clado 1	0.3417	7.9406
Clado 2	0.1598	0.8826
Clado 3	0.1760	2.9646
Clado 4	0.4024	2.9816
Clado 5	0.2240	2.4465
Clado 6	0.3913	1.1332
P. leiognathi	0.2611	1.6417
V. anguillarum	0.3060	5.4628
V. ordalii	0.1381	1.1846
V. parahaemolyticus	0.4763	2.8421

Gubbins analysis (*Croucher et al., 2014*) of whole genome alignment of 42 CCB Vibrionaceae genomes and 5 references indicates that some recombination events are shared with references strains, suggesting an ancient origin of the different lineages, before CCB diversification (Figure 4), while other recombination events happened more recently and only among CCB strains. This result is in line with what we found with fastGEAR (Mostowy et al., 2017) that indicates that 57.95 % of recombination events are ancient while the rest were recent.

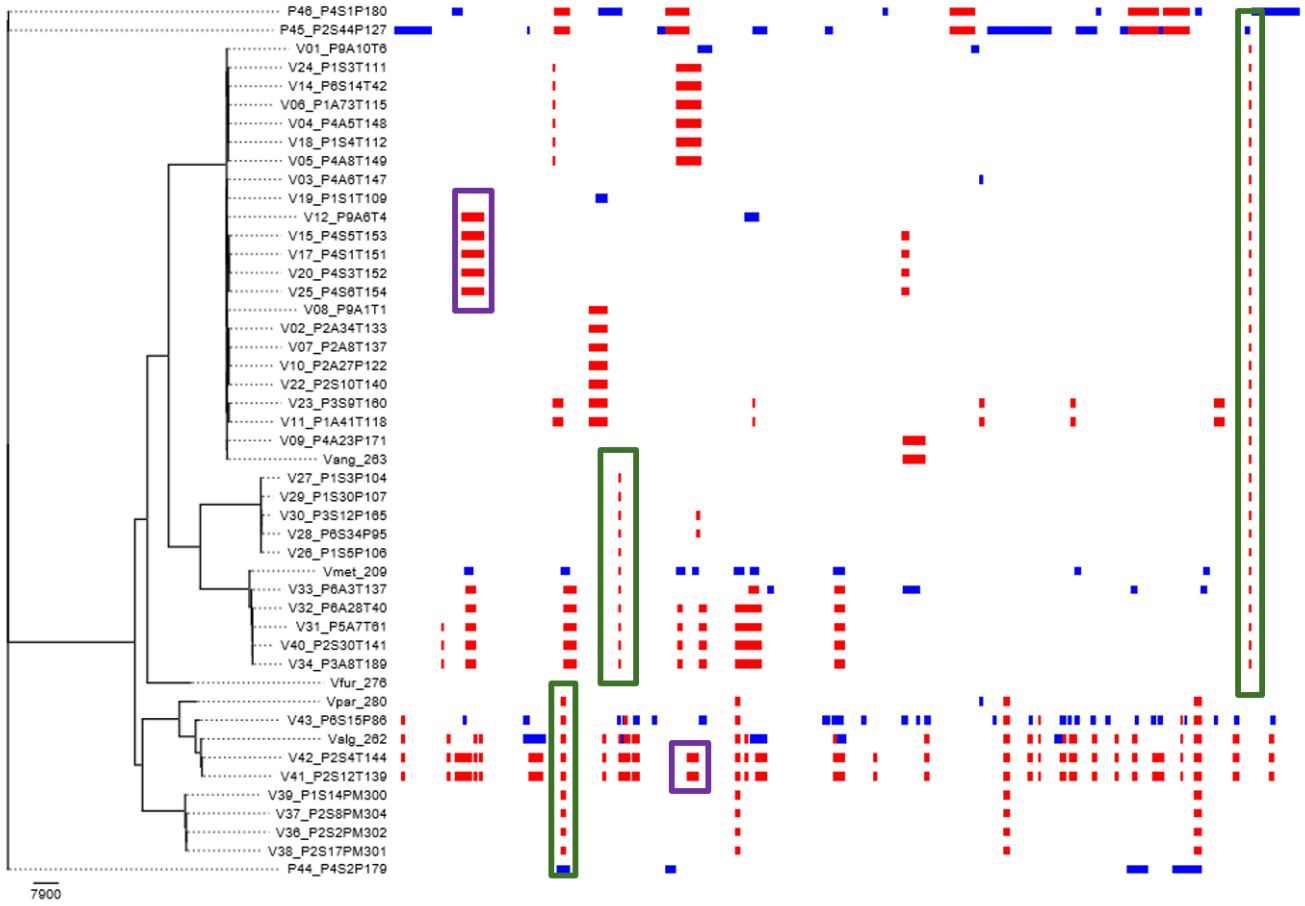


Figure 4. Gubbins inference of recombination. In this image, we are observing a whole genome alignment of reference strains and CCB strains. We are also observing a phylogenetic reconstruction. Red points indicates places within the alignment that had signals of recombination, blue points represent points of recombination wich origin was probably originated outside the alignment. Green boxed indicate some sites that are shared with references and purple boxes indicate some sites that had signals of recombination only within CCB strains.

Effective population size

According to ABC analyses, we found a wide range of effective population sizes (Table 4).

Clades I, III and V have larger effective population sizes, above 1,000,000, while Clade II and IV have smaller effective populations sizes, under 300,000 and 500,000, respectively. Also, we looked for the origin of sub-clades within clade II; we found that the in the most probably scenario sub-clade A was originated 850,113 (645,865-1,789,357) generations ago, while sub-clade D was originated 521,898 (210,418- 2,186,513) generations ago and sub-clade G was originated 550,158 (217,338 – 2,258,215) generations ago.

Table 4. Effective population sizes. Summary table of effective population sizes of CCB clades and prokaryotic and eukaryotic references.

Group Clade	Median Value	Range		Environment	Reference
		Lower value	Larger value		
Clade I	12,822,270	10,110,043	16,231,765	Sediment	This work
Clade II					
Sub-clade A	55,938	34,079	392,104	Sediment	This work
Sub-clade D	20,849	2,795	218,603	Water-Sediment	This work
Sub-clade G	29,791	6,174	226,658	Water-Sediment	This work
Clade III	15,018,880	8,970,283	22,432,331	Water-Sediment	This work
Clade IV	383,067	345,564	427,557	Sediment	This work
Clade V	9,594,874	5,894,074	12,914,770	Sediment	This work
H. pylori	39,665,437	-	-	-	Sung et al., 2012
S. enterica	348,991,354	-	-	-	Sung et al., 2012
E. coli	179,600,000	-	-	-	Foster et al., 2012
H. sapiens	20,348	-	-	-	Sung et al., 2012
A. thaliana	266,769	-	-	-	Gossmann et al., 2012
C. elegans	3,998,701	-	-	-	Denver et al., 2004
T. brucei	5,332,244	-	-	-	Sung et al., 2012

Positive selection analyses

We performed a test on each coding sequence present in at least three genomes looking at the *dn/ds* ratio for each site using FUBAR (Murrel *et al.*, 2013). We found that from 15,380 genes analyzed; only 367 had a significant signal of positive selection. Those genes are related to the carbohydrates metabolism, transport proteins and signaling enzymes (Table 5) and 75% of these genes were part of the flexible genome.

Table 5. Enrichment analysis of genes with signals of selection.

Highly represented functions among genes under seleccion	Fisher test	Significant
ABC transporter-like	p-value = 0.2319	
GNAT domain	p-value = 0.2364	
Glycosyl hydrolase, family 13, catalytic domain	p-value = 0.0007128	*
Major facilitator superfamily	p-value = 0.09123	*
ABC transporter type 1, transmembrane domain MetI-like	p-value = 0.5231	
AMP-dependent synthetase/ligase	p-value = 0.01116	*
Endonuclease/exonuclease/phosphatase	p-value = 0.01239	*
HAD-like domain	p-value = 0.04746	
LysR, substrate-binding	p-value = 0.4922	

Discussion

Pangenome of CCB Vibrionaceae.

In this study, we explored the role of different evolutionary forces in shaping the pan-genome of bacteria isolated from CCB through a comparative genomics analysis of 42 Vibrionaceae isolates. Pan-genome analysis showed 20,121 orthologues genes for the whole dataset including references from which core genome was composed of 1254 orthologue genes.

The core genes phylogeny of the 42 strains of CCB Vibrionaceae revealed six main clades. Two of these groups, clade IV and V, could be considered new species of *Vibrio,* given that they are not related to any other *Vibrio* within databases. Besides, clades IV and V present long branches and a notorious lack of recombination in between different *Vibrio* clades. Strains from Clade VI are closely related to pathogenic strains. Nevertheless, a recent study has shown that they do not harbor mobile genetic elements related to pathogenesis (Vázquez-Rosas-Landa *et al.,* 2017).

Heaps law estimation reveal an open pan-genome for the total sample of CCB Vibrionaceae. Nevertheless, most of the CCB lineages have a close pan-genome, except the most abundant lineage, clade II that has an open pan-genome. We also observed that the Vibrionaceae clades within Pozas Rojas were not distributed randomly: the clades were structured according to the isolation environment, water or sediment. Our results are in agreement with what was observed for *Exigoubacterium* isolated from this particular site (Rebollar *et al*., 2012). It has been proposed that water column is a less structured environment, where nutrients are homogeneously distributed, while sediments are structured environments, where resources are not homogenously distributed (Rebollar *et al.,* 2012). We observed that clade II has a widespread distribution in CCB, and it is found both in the water column and in the sediments. These features are an indication that these strains possess mechanisms to cope with different environments and the wider pan-genome would explain the adaptations to these.

A specialist organism would tend to lose genetic information, because of the selective pressure against DNA accumulation (Kuo and Ochman, 2009) and then show close pan-genomes. Nevertheless, this does not rule out that a specialist could also gain new genetic information that could be relevant for the adaptation to its specific environment. In other words, two factors should be

considered: the rate at which a recombination event occurred, and the effectiveness of each event that would be dependent on the selection coefficient of a given event. Therefore, a specialized genome could be full of HGT events but show closed pan-genomes due to the homogenizing effect of recombination. Closed pan-genome have been described for bacterial groups like *Bacillus anthracis* (Tettelin *et al.,* 2008), that live in isolated niches with limited access to the global microbial gene pool; something similar may happen in the specialized clades of CCB *Vibrio*, which are adapted to either water or sediment. As discussed above, Clade II is the most extensive and more generalist *Vibrio* group in CCB and displays an open pan-genome. This type of pan-genomes has been characterized for species such as *Escherichia coli* (Rasko et al., 2008) that live in multiple environments of mixed microbial communities and have different mechanisms of exchanging genetic material. In our case, this is paradoxical, since Clade II is the clade with less recombination, suggesting either a subsampling of this lineage in our study or, as our data shows, low recombination rates are a consequence of the internal structure of this clade. Core genes were first described as essential and flexible genes as dispensable (Tettelin *et al.,* 2005).

Accordingly, our genetic diversity analysis shows that core genes are always less diverse than flexible genes. Core genome is also a proxy of the variability within a phylogenetic group; small core genomes (in relation to the complete pan-genome) would represent highly variable groups in total genetic makeup.

A wide range of core genes has been calculated at a genus level, ranging from 522 to 2811 genes in *Streptococcus and Salmonella, respectively (*Donati *et al.,* 2010; Jacobsen *et al.,* 2011), and for the marine *Prochlorococcus* genus the core genome was calculated of 1273 genes using 12 strains representing its major lineages (Kettler *et al.,* 2007), close to what we found with the CCB Vibrionaceae strains. Nevertheless, to put these values in context we would have to consider genome size and the number of genomes analyzed. A small proportion of core vs. total genome size may reflect a wide habitat diversity, while a large core vs. total genome proportion may characterize a highly specialized specie. For instance, genome size of *Prochlorococcus* ranges between 1.6 Mbp and 2.7 Mbp (Kettler *et al.,* 2007), being the core a large proportion of the complete genome, while for CCB Vibrionaceae the genome size is around 5 Mbp. Therefore, if we consider for *Prochlorococcus* a genome of 2700 genes, core genome represents 47% of the whole genome, while for our Vibrionaceae it would represent around 25%. While both groups of organisms are commonly found in marine environments, they have different ecologies, as *Prochlorococcus* is the most

abundant photosynthetic cell in the oligotrophic oceans, contributing substantially to global photosynthesis, while marine Vibrionaceae species have a broad range of habitats, some free-living, others are pathogenic to aquatic animals while others even are bioluminescent symbionts of marine fishes and squids (Farmer *et al.,* 1986; Wachsmuth *et al.,* 1994), and some marine Vibrionaceae are extremophiles that can live in extreme cold, salt or pressure (Bartlett *et al.,* 2006). In the particular case of CCB Vibrionaceae, the small percentage of the genome represented in the core suggests that these strains occupy very diverse ecological niches within CCB.

Pan-genome size, effective population size and recombination

It has been suggested that effective population size has a major impact on pan-genome composition (Mclnerney *et al.,* 2017), and recently, it has been reported that large N_e are correlated with large accessory genomes (Andreani *et al.,* 2017) which would be reflected into open pan-genome. However, in our dataset, we found for Clade I, III and V, large N_e with closed pan-genome and high recombination rates *r/m* within their clades, ranging from 2.44 to 7.94. This could suggest that these clades are both highly successful in their environments explaining the higher N_e and are they tend to prefer either water or sediment. In this case, the upper recombination rates within lineage keep the cohesion of the lineage, since it is limited to its close relatives, but by increases effective population sizes it also, increases the genetic diversity and hence, the opportunity of adaptation to arise.

In contrast, clade IV has smaller effective population size, is a closed pan-genome and has high recombination rates. We propose that for this particular clade we could be observing a demographic effect, where this highly specialized lineage could be involved in an ongoing process of population expansion after the perturbation of 2011.

Finally, the most abundant clade is also the most puzzling of our study. On the one hand, the generalist clade II has lower recombination rates in their genes, but interestingly is the only lineage that exhibits an open pan-genome. Besides, we also observed that this lineage is composed of several sub-clades (A to J), which are differentiated from each other, but not enough as to be considered different clades. All of the sub-clades have small N_e, which suggest that these sub-clades are in the process of specialization. Therefore, our results indicate that large N_e are not always correlated with open pan-genomes. We could explain this sub-clades by two hypothesis, one these sub-clades are the result of a recent adaptive radiation process. Alternatively, this pattern is the result of subsampling, where clade II is probably a representation of a much larger diversity. To discriminate between both

hypothesis further sampling of *Vibrio* in the watershed is needed; however, this can prove to be difficult due to the general desiccation of CCB wetland due to groundwater overexploitation for agriculture (Souza et al., 2007, Wolaver 2013). The closing of the canals that are exporting the superficial water 80km away from the basin could help this situation, policies are underway to make this happens.

Genetic drift, natural selection and effective population size

Effective population size is a proxy to evaluate if natural selection or genetic drift are the primary drivers of evolution within populations. On the one hand, large effective population sizes indicate that natural selection could be driving the evolution of populations, while small effective population sizes suggest that populations are subjected to stochastic processes known as genetic drift. Our results suggest that Vibrionaceae within CCB have small effective population sizes if they are compared to the N_e of bacteria such as *S. enterica* (Sung *et al*., 2012) and *E. coli* (Foster *et al*., 2012). In the case of these bacteria, both have a broad distribution in all the planet and successful living strategies such as pathogenic strains of vertebrates (Souza *et al.,* 1999, Hidalgo-Vila *et al.,* 2007, Ahmed & Shimamoto, 2014). Also, these cosmopolitan bacteria have been reported to have high horizontal gene transfer rates (Gogarten *et al.,* 2002, Wiedenbeck & Cohan, 2011, Messerer *et al.,* 2017). In contrast, bacteria from CCB have been isolated from the oceans since the Oligocene (Moreno-Letelier *et al.,* 2012) and foreign DNA is used as food instead of as genetic information (Tapia-Torres *et al.,* 2015). These differences in lifestyle could be the explanation to this small N_e. Moreover, if we compare these population sizes with those reported to some eukaryotes, will be considered to be in the same range or higher for some clades. Therefore, what is happening within CCB Vibrionaceae, genetic drift or natural selection? Even though CCB Vibrionaceae have small N_e when compared to other bacteria, some clades have N_e values greater than those for eukaryotes, even all clade II sub-clades present values comparable to multicellular eukaryotes such as *H. sapiens* and *A. thaliana*. Given these results, it is possible that evolution of these populations is been driven by natural selection.

In CCB we have hypothesized that nutrient scarcity has played a significant role in the evolution of bacteria by promoting low recombination rates and local adaptation (Souza *et al.,* 2008; Souza *et al.,* 2012). Despite this reduced recombination rates, we found that 2% of the pan-genome have signals of positive selection, which suggests adaptation within Vibrionaceae. Among the functions

within the genes under positive selection in the studied genome of Vibrionaceae in CCB, we detected genes related to carbohydrates metabolism, transport polypeptides, intercellular signaling and AMP-dependent enzymes related to *quorum sensing* and central metabolism. We believe that divergence among Vibrionaceae clades can be related to these functions. In accordance with our data, it has been proposed that pan-genomes in prokaryotes are maintained by natural selection and not by genetic drift, given that we do not observe large genomes as in eukaryotes (McInerney *et al.* 2017). Our results show that 75.3% of the genes with signals of selection are within the flexible genome, which supports the idea that for Vibrionaceae within CCB, the pan-genome is also maintained by natural selection.

Our data shows that even if recombination is lower compared to what has been reported, Vibrionaceae from CCB posses wide variation at genetic diversity levels between core and flexible genes, which suggest that evolution on this lineages do not occur as genome-wide sweeps mode, but more like other *Vibrio* in the oceans (Shapiro *et al.*, 2012)

From ecology and species diversity to pan-genomes.

Members of Vibrionaceae family are always present in the aquatic systems in CCB, but usually at low abundance (Souza *et al.*, 2006; Escalante *et al.*, 2008; Bonilla-Rosso *et al.*, 2012; Peimbert *et al.*, 2012). We observed an increase C:N:P ratio in relation to the pre-hurricane sampling in the studied (Pozas Rojas) (Bonilla-Rosso *et al.*, 2012; Peimbert *et al.*, 2012), reducing the natural stoichiometry imbalance. We consider that the increase of nutrients provoked a shift in the community structure that increased the local abundance of *Vibrio* and *Photobacterium*, allowing us to cultivate these clades, while the competitive advantage of other previous abundant bacterial groups, such as *Pseudomonas*, that was previously very common (Peimbert et al., 2012) was drastically reduced. Recently the effect on nutrients in microbial communities within CCB has been assessed through an enrichment experiment (Lee *et al.*, 2017), observing that *Vibrio* at water column were abundant when nutrient stoichiometry remain un-enriched, while in sediments, *Vibrio* became abundant when nutrient stoichiometry were similar to what has been observed in marine environments N:P equal to 16:1. The present study supports the idea that in general communities are susceptible to stoichiometric variation.

Conclusion

Pan-genomes reflect, at least in part, the lifestyle of the organism. While open pan-genomes suggest that a bacterial clade can live in a broad range of habitats, close pan-genomes are usually associated with a highly specialized lifestyle. Recombination may or not favor open pan-genomes, depending on the rates at which a new allele is included and the time it takes to spread on the population, which is related to its selection coefficient. On the other hand, large effective population sizes can be found associated to closed pan-genomes, apparently as the result of demographic process. At CCB, despite the lower recombination rate detected compared with the values reported for related marine *Vibrio*, selection apparently acts as the main evolutionary force shaping pan-genomes. In our study we have shown that even in bacterial populations with smaller effective population sizes, we are able to find strong signals of natural selection, discarding genetic drift as the main evolutionary force of our lineages.

Acknowledgments

MVR-L is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 345250 from CONACYT. GYP-S is a doctoral student from Programa de Doctorado en Ciencias Biológicas, UNAM and acknowledges the scholarship by CONACYT (Grant no. 269818). This research was also supported by funding from PAPIIT project IG200215 and WWF-Alianza Carlos Slim, SEP-Ciencia Básica CONACYT grant 238245 to both VS and LEE. The paper was written during a sabbatical leave of LEE and VS at the University of Minnesota in Peter Tiffin and Michael Travisano laboratories, with support by scholarships from PASPA, DGAPA, UNAM. Biogeochemical analysis where performed by Rodrigo Velazquez-Duran at the Instituto de Investigaciones en Ecosistemas y Sustentabilidad, UNAM.

References

Ahmed, A.M., & Shimamoto, T. (2014). Isolation and molecular characterization of Salmonella enterica, Escherichia coli O157:H7 and Shigella spp. from meat and dairy products in Egypt. Int J Food Microbiol. 3:168-169:57-62. doi: 10.1016/j.ijfoodmicro.2013.10.014.

Andreani, N. A., Hesse, E., & Vos, M. (2017). Prokaryote genome fluidity is dependent on effective population size. The ISME Journal, 11(7), 1719–1721. http://doi.org/10.1038/ismej.2017.36

Avitia, M., Escalante, A. E., Rebollar, E. A., Moreno-Letelier, A., Eguiarte, L. E., & Souza, V. (2014). Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. PeerJ, 2, e696. http://doi.org/10.7717/peerj.696

Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., Krause, J.D., Whitaker, R. J. (2012). Patterns of Gene Flow Define Species of Thermophilic Archaea. PLoS Biology, 10(2), e1001265. http://doi.org/10.1371/journal.pbio.1001265

Cohan, F.M., Perry, E.B. (2007). A systematics for discovering the fundamental units of bacterial diversity. Curr Biol. 17(10):R373-86. Review.

Collins, R.E. & Higgs, P.G. (2012) Testing the infinitely many genes model for the evolution of the bacterial core genome and pangenome. Mol Biol Evol. 29(11):3413-3425. doi: 10.1093/molbev/mss163.

Cordero, O.X., & Polz, M.F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol. 12(4):263-73. doi:10.1038/nrmicro3218.

Cui, Y., Yang, X., Didelot, X., Guo, C., Li, D., Yan, Y., Zhang, Y., Yuan, Y., Yang, H., Wang, J., Wang, J., Song, Y., Zhou, D., Falush, D., Yang, R. (2015). Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen Vibrio parahaemolyticus. Mol Biol Evol. 32(6):1396-410. doi: 10.1093/molbev/msv009.

Didelot, X., & Falush, D. (2007). Inference of Bacterial Microevolution Using Multilocus Sequence Data. Genetics, 175(3), 1251–1266. http://doi.org/10.1534/genetics.106.063305

Didelot, X., & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. Trends in Microbiology, 18(7), 315–322. http://doi.org/10.1016/j.tim.2010.04.002

Elser, J. J., Schampel, J. H., Garcia-Pichel, F., Wade, B. D., Souza, V., Eguiarte, L., Escalante, A., Farmer, J. D. (2005), Effects of phosphorus enrichment and grazing snails on modern stromatolitic microbial communities. Freshwater Biology, 50: 1808–1825. doi:10.1111/j.1365-2427.2005.01451.x

Escalante, A.E., Eguiarte, L.E., Espinosa-Asuar, L., Forney, L.J., Noguez, A.M., Souza-Saldivar, V. (2008). Diversity of aquatic prokaryotic communities in the Cuatro Cienegas basin. FEMS Microbiol Ecol. 65(1):50-60. doi:10.1111/j.1574-6941.2008.00496.x.

Espinosa-Asuar, L., Escalante, A.E., Gasca-Pineda, J., Blaz, J., Peña, L., Eguiarte, L.E., Souza, V. (2015). Aquatic bacterial assemblage structure in Pozas Azules, Cuatro Cienegas Basin, Mexico: Deterministic vs. stochastic processes. Int Microbiol. 18(2):105-15. doi: 10.2436/20.1501.01.240. PubMed PMID: 26496618.

Excoffier, L. and M. Foll. (2011). Fastsimcoal: a continuous-time coalescent simulator of

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., and M. Foll. (2013). Robust demographic inference from genomic and SNP data. PLOS Genetics 9(10):e1003905.

Friedman, J., Alm, E. J., & Shapiro, B. J. (2013). Sympatric Speciation: When Is It Possible in Bacteria? PLoS ONE, 8(1), e53539. http://doi.org/10.1371/journal.pone.0053539

genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27: 1332-1334.

Gogarten, J.P., Doolittle, W.F., & Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. Mol Biol Evol. 19(12):2226-38

Gordienko, E. N., Kazanov, M. D., & Gelfand, M. S. (2013). Evolution of Pan-Genomes of Escherichia coli, Shigella spp., and Salmonella enterica. Journal of Bacteriology, 195(12), 2786–2792. http://doi.org/10.1128/JB.02285-12

Hidalgo-Vila, J., Díaz-Paniagua, C., de Frutos-Escobar, C., Jiménez-Martínez, C., Pérez-Santigosa, N. (2007). Salmonella in free living terrestrial and aquatic turtles. Vet Microbiol. 119(2-4):311-5.

Kimura, M. (1981). Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci U S A. 78(1), 454-8

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., … Cohan, F. M. (2008). Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. Proceedings of the National Academy of Sciences of the United States of America, 105(7), 2504–2509. http://doi.org/10.1073/pnas.0712205105

Konstantinidis, K. T., & Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences of the United States of America, 101(9), 3160–3165. http://doi.org/10.1073/pnas.0308653100

Kuo, C.-H., & Ochman, H. (2009). Deletional Bias across the Three Domains of Life. Genome Biology and Evolution, 1, 145–152. http://doi.org/10.1093/gbe/evp016

Lapierre, P. & Gogarten, J.P. (2009) Estimating the size of the bacterial pan-genome. Trends Genet. 25(3):107-10. doi: 10.1016/j.tig.2008.12.004

Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proceedings of the National Academy of Sciences of the United States of America, 109(41), E2774–E2783. http://doi.org/10.1073/pnas.1210309109

Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D. & Vandenkoornhuyse, P. (2016). Beyond the Black Queen Hypothesis. The ISME Journal. ePub. doi: 10.1038/ismej.2016.22

Messerer, M., Fischer, W., Schubert, S. (2017). Investigation of horizontal gene transfer of pathogenicity islands in Escherichia coli using next-generation sequencing. PLoS ONE, 12(7), e0179880. http://doi.org/10.1371/journal.pone.0179880

Moreno-Letelier, A., Olmedo-Alvarez, G., Eguiarte, L. E., & Souza, V. (2012). Divergence and Phylogeny of Firmicutes from the Cuatro Ciénegas Basin, Mexico: A Window to an Ancient Ocean. Astrobiology, 12(7), 674–684. http://doi.org/10.1089/ast.2011.0685

Morris, J.J., Lenski, R.E. & Zinser, E.R. (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. mBio, 3(2), e00036–12. http://doi.org/10.1128/mBio.00036-12

Naka, H., Dias, G. M., Thompson, C. C., Dubay, C., Thompson, F. L., Crosa, J. H. (2011). Complete Genome Sequence of the Marine Fish Pathogen Vibrio anguillarum Harboring the pJM1

Virulence Plasmid and Genomic Comparison with Other Virulent Strains of V. anguillarum and V. ordalii . Infection and Immunity, 79(7), 2889–2900. http://doi.org/10.1128/IAI.05138-11

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

Peimbert, M., Alcaraz, L.D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L., Eguiarte, L.E. and Souza, V. 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(7):648–658. doi.org/10.1089/ast.2011.0694

Ponce-Soto, G. Y., Aguirre-von-Wobeser, E., Eguiarte, L. E., Elser, J. J., Lee, Z. M.-P., & Souza, V. (2015). Enrichment experiment changes microbial interactions in an ultra-oligotrophic environment. Frontiers in Microbiology, 6, 246. http://doi.org/10.3389/fmicb.2015.00246

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., … Ravel, J. (2008). The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates . Journal of Bacteriology, 190(20), 6881–6893. http://doi.org/10.1128/JB.00619-08

Rebollar, E.A., Avitia, M., Eguiarte, L.E., González-González, A., Mora, L., Bonilla-Rosso, G., Souza, V. (2012). Water-sediment niche differentiation in ancient marine lineages of Exiguobacterium endemic to the Cuatro Cienegas Basin. Environ Microbiol. 14(9):2323-33. doi: 10.1111/j.1462-2920.2012.02784.x.

Rodríguez-Verdugo, A., Souza, V., Eguiarte, L. E., & Escalante, A. E. (2012). Diversity across Seasons of Culturable Pseudomonas from a Desiccation Lagoon in Cuatro Cienegas, Mexico. International Journal of Microbiology, 2012, 201389. http://doi.org/10.1155/2012/201389

Salemi, M., Vandamme, A.-M., & Lemey, P. (2009). The phylogenetic handbook: A practical approach to phylogenetic analysis and hypothesis testing. Cambridge, UK: Cambridge University Press.

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., Polz, M.F., Alm, E. J. (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science (New York, N.y.), 336(6077), 48–51. http://doi.org/10.1126/science.1218198

Smith, J. M., Smith, N. H., O'Rourke, M., & Spratt, B. G. (1993). How clonal are bacteria? Proceedings of the National Academy of Sciences of the United States of America, 90(10), 4384–4388.

Snipen, L., & Liland, K. H. (2015). micropan: an R-package for microbial pan-genomics. BMC Bioinformatics, 16, 79. http://doi.org/10.1186/s12859-015-0517-0

Souza, V., & Eguiarte, L. E. (1997). Bacteria gone native vs. bacteria gone awry?: Plasmidic transfer and bacterial evolution. Proceedings of the National Academy of Sciences of the United States of America, 94(11), 5501–5503.

Souza, V., Eguiarte, L. E., Travisano, M., Elser, J. J., Rooks, C., & Siefert, J. L. (2012). Travel, Sex, and Food: What's Speciation Got to Do with It? Astrobiology, 12(7), 634–640. http://doi.org/10.1089/ast.2011.0768

Souza, V., Eguiarte, L.E., Siefert, J., Elser, J.J. (2008). Microbial endemism: does phosphorus limitation enhance speciation? Nat Rev Microbiol. 6(7):559-64. doi:10.1038/nrmicro1917.

Souza, V., Nguyen, T. T., Hudson, R. R., Piñero, D., & Lenski, R. E. (1992). Hierarchical analysis of linkage disequilibrium in Rhizobium populations: evidence for sex? Proceedings of the National Academy of Sciences of the United States of America, 89(17), 8389–8393.

Souza, V., Rocha, M., Valera, A., & Eguiarte L.E. (1999). Genetic structure of natural populations of Escherichia coli in wild hosts on different continents. Applied and Environmental Microbiology 65: 3373-3385

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., & Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences of the United States of America, 109(45), 18488–18492. http://doi.org/10.1073/pnas.1216223109

Tapia-Torres, Y., Rodríguez-Torres, M. D., Elser, J. J., Islas, A., Souza, V., García-Oliva, F., & Olmedo-Álvarez, G. (2016). How To Live with Phosphorus Scarcity in Soil and Sediment: Lessons from Bacteria. Applied and Environmental Microbiology, 82(15), 4652–4662. http://doi.org/10.1128/AEM.00160-16

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Ros, I,M., Peterson, J.D., Hauser, C.R., Sundaram, J., Nelson, W., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pangenome." Proceedings of the National Academy of Sciences of the United States of America, 102(39), 13950–13955. http://doi.org/10.1073/pnas.0506758102

Vos, M., Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. 3(2):199-208. doi: 10.1038/ismej.2008.93.

Wiedenbeck J, Cohan FM. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev.35(5):957-76. doi: 10.1111/j.1574-6976.2011.00292.x

Xu, F., Ilyas, S., Hall, J. A., Jones, S. H., Cooper, V. S., & Whistler, C. A. (2015). Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. Frontiers in Microbiology, 6, 272. http://doi.org/10.3389/fmicb.2015.00272

Zhi, X.Y., Jiang, Z., Yang, L.L. & Huang, Y. (2017) The underlying mechanisms of genetic innovation and speciation in the family Corynebacteriaceae: A phylogenomics approach. Mol Phylogenet Evol. 107:246-255. doi: 10.1016/j.ympev.2016.11.009

Capítulo 4.

Patogénesis de Vibrionaceae en Cuatro Ciénegas

4.1 Artículo 2. (Publicado). Comparative genomics of free-living Gammaproteobacteria: pathogenesis-related genes or interaction-related genes?.

En este capítulo vamos se exploran dos cosas en relación a los linajes de Vibrionaceae del valle de Cuatro Ciénegas: *i)* su relación con otros linajes de Vibrionaceae fuera del valle, esto a través del estudio de los genes asociados a la patogénesis y *ii)* el impacto de la THG.

Como se mencionó anteriormente muchos miembros de la familia Vibrionaceae son patógenos, principalmente de animales, por otro lado, el valle de Cuatro Ciénegas es altamente diverso y sus comunidades diferentes a las encontradas en cuerpos acuáticos contientales (Souza *el al.,* 2006; Bonilla-Rosso *et al;* 2012). Por lo que una pregunta que se desprende de estos hechos es si los linajes de Vibrionaceae son patógenos también y si lo son cuales son las implicaciones de estos hechos.

En el siguiente apartado se presenta un análisis de genómica comparada donde se busca entender cuál es el papel de los genes asociados a la patogénesis en los genomas de los linajes del valle de Cuatro Ciénegas.

Pathogens and Disease, 75, 2017, ftx059

doi: 10.1093/femspd/ftx059
Advance Access Publication Date: 7 June 2017
Research Article

RESEARCH ARTICLE

Comparative genomics of free-living Gammaproteobacteria: pathogenesis-related genes or interaction-related genes?

Mirna Vázquez-Rosas-Landa†, Gabriel Yaxal Ponce-Soto†, Luis E. Eguiarte and V. Souza*

Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, Coyoacán 04510 Ciudad de México, México.

*Corresponding author: Laboratorio de Ecología Molecular y Experimental, Instituto de Ecología, Departamento de Ecología Evolutiva, Universidad Nacional Autónoma de México, AP 70-275, Coyoacán, DF, 04510, México. Tel: +52-55-5622-9006; E-mail: souza@unam.mx
†Equal contributions.

One sentence summary: Pathogenicity-related genes were found in a non-human-impacted ecosystem, which leads to the idea that these genes are not necessarily involved in pathogenesis, but in interactions within microbial communities.

Editor: Edmundo Calva

ABSTRACT

Bacteria have numerous strategies to interact with themselves and with their environment, but genes associated with these interactions are usually cataloged as pathogenic. To understand the role that these genes have not only in pathogenesis but also in bacterial interactions, we compared the genomes of eight bacteria from human-impacted environments with those of free-living bacteria from the Cuatro Ciénegas Basin (CCB), a relatively pristine oligotrophic site. Fifty-one genomes from CCB bacteria, including *Pseudomonas*, *Vibrio*, *Photobacterium* and *Aeromonas*, were analyzed. We found that the CCB strains had several virulence-related genes, 15 of which were common to all strains and were related to flagella and chemotaxis. We also identified the presence of Type III and VI secretion systems, which leads us to propose that these systems play an important role in interactions among bacterial communities beyond pathogenesis. None of the CCB strains had pathogenicity islands, despite having genes associated with antibiotics. Integrons were rare, while CRISPR elements were common. The idea that pathogenicity-related genes in many cases form part of a wider strategy used by bacteria to interact with other organisms could help us to understand the role of pathogenicity-related elements in an ecological and evolutionary framework leading toward a more inclusive One Health concept.

Keywords: comparative genomics; evolution; Gammaproteobacteria; one health; pathogenesis; secretion systems

INTRODUCTION

Interactions among microbes are key elements in the maintenance of healthy ecosystems, given that microbes have fundamental roles in nutrient cycling (Makino and Cotner 2005; Adams Krumins *et al.* 2009). There are several ways in which these interactions occur, some of which might be related to cross-feeding, while others involve secretion of molecules into the environment, as in the quorum-sensing response (Diggle *et al.* 2007). Nevertheless, other mechanisms of interaction require direct contact between cells. Many of these strategies are commonly considered as virulence elements, and their presence is used to predict the bacterial potential for infection (Métoth and Alizon 2014).

Received: 31 January 2017; **Accepted:** 2 June 2017

An interesting perspective on how to understand these complex interactions in an ecological context has been gained by using model systems of bacteria such as *Escherichia coli*. In this case, the pathogenic strains have different ecologies, ranging from generalist and opportunistic strains to very specific forms that cause disease, such as the uropathogenic strains that are adapted to the particular conditions of the urogenital system (Johnson 1991). We started to work on *E. coli* as a model to understand the evolution and ecology of wild (non-human) populations more than 20 years ago. We started by obtaining a collection of isolates different from *E. coli* of reference (ECOR; Ochman and Selander 1984) since the idea was to look at isolates from wild animals from different lineages and sites, in particular from Mexico. We observed a vast diversity among wild strains compared with human-related isolates, and we also found that plasmids and antibiotic resistance were more abundant in those environments impacted by humans than in the wild. However, even strains from whales in Antarctica were resistant to several antibiotics (Souza *et al.* 1999). We also explored the evolution of independent virulence factors by looking at the components of the locus of enterocyte effacement (LEE) in our wild *E. coli* collection from a range of animals (Sandners *et al.* 2001) and even in human-related isolates (Castillo, Eguiarte and Souza 2005). In particular, we analyzed the evolution of type 1 fimbria. This system consists of an adhesion protein that allows *E. coli* to attach to its host (Peek *et al.* 2001). Evidence for recombination was found, allowing natural selection to fine-tune the outer part of the proteins in response to the immune system, while the internal modules were much conserved. In summary, antibiotic resistance and attachment proteins are present in non-pathogenic *E. coli* strains; nevertheless, these genes are commonly associated with pathogenicity.

We are particularly interested in the virulence elements that are involved in a microbe's interactions. It is possible to describe two types of interactions, positive (such as mutualism) and negative (competence or predation), and while some interactions require a direct contact between cells, others do not (Epstein 2003). The secretion systems are mechanisms that do require contact among cells. In particular, Gram-negative bacteria possess six types of secretion systems (types I to VI), while other bacterial groups have similar systems (Horn *et al.* 2004). Among the functions described for the secretion system is a syringe-like structure that is able to 'inject' biomolecules as complex as DNA or as simple as a transduction signal through the cell envelope from one bacterial cell to another. Hence, these secretion systems have a panoply of functions that go from conjugation to killing the host cell (Koster, Bitter and Tommassen 2000; Kapitein and Mogk 2013; Costa *et al.* 2015). The Type III secretion system (T3SS) is considered of particular interest given that it mediates interactions between bacteria and eukaryotic cells and is essential for the establishment of infection for many pathogens, such as *Pseudomonas syringe* and *P. aeruginosa* (Hauser 2009; Gazi *et al.* 2012). However, T3SS is also essential for symbiotic mutualism, as is the case for some rhizobia (Viprey *et al.* 1998; Marie, Broughton and Deakin 2001; Marie *et al.* 2003; Yang *et al.* 2010; Okazaki *et al.* 2013; Tampakaki 2014). The Type VI secretion system (T6SS), as well as T3SS, seems to be widespread in nature, in particular among Gram-negative bacteria such as the Proteobacteria. However, T6SS has been proposed to have evolved for contact between prokaryotic cells (Bingle, Bailey and Pallen 2008).

Among the mechanisms employed by bacteria in a social context that do not require contact between cells we can mention antibiotics (Czárán *et al.* 2002; Parret and De Mot 2002; Riley and Wertz 2002). The traditional clinical view has linked antibiotic production and resistance with virulence. Nevertheless, antibiotics also play an important role within communities, and are found in bacteria from natural sources (Riley and Gordon, 1999; Fajardo *et al.* 2008). In a microbial community context, antibiotics can either inhibit growth or kill competitors. It has been shown that resistant microorganisms tend to be more abundant in human-impacted environments, which has been interpreted as the result of a human-mediated dispersal of these resistance genes and resistant bacteria (Cristobal-Azkarate *et al.* 2014). This dispersion is commonly associated with mobile genetic elements such as genomic islands, integrons and bacteriophages (Dobrindt *et al.* 2004; Hazen *et al.* 2010).

Integrons are gene-capturing elements that can host a wide range of 'cassettes' that promote the spread of antibiotic resistance genes or virulence factors (Cury *et al.* 2016). Such strategies have been described in several genera of Gammaproteobacteria such as *E. coli* (Salyers and Amabile-Cuevas 1997; Díaz-Mejia *et al.* 2008; Manyahi *et al.* 2017). In *E. coli* we observed that class 1 integrons were much more abundant in human-associated strains that in those from wild animals, suggesting a cost in expressing resistance to several antibiotics. This cost pays off when the selective pressure is high and regulation is tight. Probably this is the case for the origin and evolution of a cassette-like structure in human-associated strains. These genes are scattered along the chromosome in *E. coli* from wild animals (Díaz-Mejia *et al.* 2008).

However, the role of mobile elements, as well as of antibiotic synthesis and secretion systems, is complex; for example, not all genomic islands are pathogenicity islands, and most plasmids are not related to pathogenesis (Dobrindt *et al.* 2004). In some bacteria, such as *Vibrio*, integrons and the genes associated with them are usually relevant for their virulence (Mazel *et al.* 1998). However, it is also possible that they play a different role in non-pathogenic strains. On the other hand, phages are considered as one of the most efficient mechanisms of horizontal gene transfer (HGT; Jiang and Paul 1998; Canchaya *et al.* 2003; Weinbauer and Rassouldzadegan 2004; Moon *et al.* 2016). Nevertheless, the great diversity of phages that has been described in environments such as the ocean suggests that they might play also a relevant role in microbial communities as seed-banks for adaptive genes, which can potentially be acquired by any member of the community (Fancello *et al.* 2011; Modi *et al.* 2013; Subirats *et al.* 2016). Tightly associated to phages are the clustered regularly interspersed short palindromic repeats (CRISPR) loci. CRISPRs are non-contiguous direct repeats separated by small sequences called spacers and have been proposed to confer immunity against foreign genetic elements such as phages (Bland *et al.* 2007). For historical reasons, modern medical training has considered human health mostly as unlinked to ecosystems. Nevertheless, this idea has been changing since the One Health concept started to become more entrenched in public health policies when veterinary researchers, public health specialists and medical doctors began to talk to one other (http://www.onehealthinitiative.com). Recently we have witnessed a rise in microbiome studies that show that microbes are an essential element of our biology and health. Such microbiome studies have been an 'eye opener' in this aspect (Human Microbiome Project Consortium 2012; Bik 2016). We believe that to understand diseases and how to treat them, it is important to consider that bacteria and viruses form part of complex communities and that the elements related to virulence should be placed in this context.

Figure 1. Geographic location of the Cuatro Ciénegas Basin, Mexico. Small map shows regional localization of the basin. Two water systems were analyzed: (**a**) Churince and (**b**) Los Hundidos.

As a result of this new philosophy, we believe that it is important to evaluate all these pathogenic elements in an ecological context, so we can better understand their significance in a natural context. Therefore, we analyzed the frequency of occurrence of genes associated to the different mechanisms of interaction in strains from the Cuatro Ciénegas Basin (CCB), a hydrological system with little or no human activity, and in consequence with few or no human pathogens. CCB is a relatively pristine site where interactions among bacteria seem to be very important in structuring the communities (Souza *et al.* 2006, 2012b; Peimbert *et al.* 2012; Pérez-Gutiérrez *et al.* 2013; Ponce-Soto *et al.* 2015; Rodríguez-Torres *et al.* 2017).

Moreover, CCB is an oasis system in the desert in the northeast of Mexico, a veritable microbiological 'lost world', and a very diverse site with many relict microorganisms (Moreno-Letelier *et al.* 2011; Souza *et al.* 2012b). As a result of its extreme oligotrophy (P less than 0.5 μM; Elser *et al.* 2005), neither algae nor human-related bacterial strains survive (Souza *et al.* 2006, 2012b). *Escherichia coli*, *Salmonella* and their phages have been very uncommon during 17 years of microbial sampling in CCB for ecological, genomic and metagenomic studies (Desnues *et al.* 2008; Bonilla-Rosso *et al.* 2012). Paradoxically, CCB is one of the most biodiverse places in the world for microorganisms, due not only to the abundance of microbial mats and stromatolites but also to the high endemicity of its microbiota, as many lineages are very different from their relatives elsewhere (Alcaraz *et al.* 2008; Desnues *et al.* 2008; Escalante *et al.* 2008). This endemicity and diversity is enhanced by a high differentiation between sites within CCB, even on a scale of meters (López-Lozano *et al.* 2012, 2013; Espinosa-Asuar *et al.* 2015; Pajares *et al.* 2016; Lee *et al.* 2017). This differentiation is caused by a strong cohesion of each local community due to metabolic complementarity (Rodríguez-Torres *et al.* 2017), along with its large and diverse panoply of antibiotics and toxins. These metabolites inhibit effective migration of newcomers (Pérez-Gutierrez *et al.* 2013; Aguirre-von-Wobeser *et al.* 2014), as DNA of foreigners is a welcome source of P for these bacterial communities (Tapia-Torres *et al.* 2016). CCB is also unique in its macrobiota, in particular of endemic species of reptiles, snails, mollusks and fish (Hershler 1984; Minckley 1984; Carson *et al.* 2012; Velez *et al.* 2016).

In order to gain an understanding of the evolution and ecology of pathogenicity in well-known ubiquitous bacteria, we compared genomic elements commonly associated with pathogenicity in lineages of Gammaproteobacteria, namely *Pseudomonas*, *Vibrio*, *Photobacterium* and *Aeromonas*, from CCB with their close relatives isolated from human-impacted environments. These genera have been cultivated consistently in CCB in recent years (Rodríguez-Verdugo *et al.* 2012; Ponce-Soto *et al.* 2015). Likewise, some members of these genera have been characterized as pathogens of humans, animals and plants (Rainey and Bailey 1996; Austin 2010; Grosso-Becerra *et al.* 2014; Scully *et al.* 2016). Therefore, comparative genomic analysis using strains from CCB gives us an excellent opportunity to record and rethink the different roles of genes typically associated with pathogenicity in environmental strains.

METHODS

Data collection

Fifty-nine genomes in total were used for this analysis. Eight reference genomes were from four studied genera, *Pseudomonas*, *Aeromonas*, *Vibrio* and *Photobacterium* (Supplementary Table S1). Fifty-one genomes were obtained from isolates from two different aquatic systems within CCB, Churince and Pozas Rojas in Los Hundidos system (Fig. 1). Four *Pseudomonas* and three *Aeromonas* strains were isolated from the Churince aquatic system during 2011. We also analyzed *P. cuatrocienegensis* strain 21C1, an endemic species of CCB isolated in 2001 from the Churince system (Escalante *et al.* 2009). Thirty-nine *Vibrio*, three *Photobacterum* and one *Aeromonas* isolates were obtained from nine ponds in Pozas Rojas within the Los Hundidos system of CCB (Fig. 1) during 2013.

All the samples were collected from water and sediment using sterile BD Falcon vials (BD Biosciences, San Jose, CA, USA). The strains were obtained by plating 100 μl of each water sample or 100 μl of a 1:10 sediment dilution, prepared with 0.9% NaCl solution. Strains were isolated from water and sediment using *Pseudomonas* isolation agar (PIA) and thiosulfate citrate bile sucrose (TCBS) agar. Strains were incubated on agar plates at room temperature for 2 days, and then were kept at 4°C until isolation in the laboratory. Each isolate was subjected to two rounds of colony purification and then grown in liquid Luria–Bertani (LB) medium. Cryotubes were stored with an aliquot of each strain for future reference and cultivation.

DNA extraction and sequencing

DNA was extracted overnight in liquid LB medium from each strain using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). DNA quality and concentration were determined using a NanoDrop Lite spectrophotometer (Thermo Scientific). Purified DNA of *Pseudomonas* and *Aeromonas* strains was sent to the J. Craig Venter Institute (JCVI) for library preparation and sequencing. The estimated coverage of these strains ranged from 21- to 63-fold. *Vibrio* and *Photobacterium* strains were sequenced at Cinvestav-LANGEBIO, Mexico with an expected coverage of 10× per genome. All sequencing was performed with Illumina MiSeq 2 × 250 technology, with insert libraries of 650 bps.

The quality of raw reads was analyzed using FASTQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A minimum quality value of 25 was set, and low quality sequences were removed with fastq_quality_filter from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). After removing low quality sequences, adapter sequences were identified and removed using TagCleaner (Schmieder *et al.* 2010). Genome assembly was performed *de novo* with Newbler version 2.9 (Roche/454 Life Sciences). For the scaffolding process SSPACE (Boetzer *et al.* 2011) was used, gaps were closed using GapFiller (Nadalin, Vezzi and Policriti 2012) and final error correction was performed with iCORN (Sanger Institute). DeconSeq (Schmieder and Edwards 2011) was used to identify and remove contaminant sequences in the assembled genomes. Final assemblies were annotated with PROKKA software (Seeman 2014) and the Pfam database (Finn *et al.* 2014) using default parameters.

Virulence genes and core virulence phylogeny

To determine which of the so called virulence factors were present in the 59 bacterial genomes (including reference and CCB strains), we performed a BLASTp (Altschul *et al.* 1990) analysis against a curated database, the Virulence Factors Database (VFDB; Chen *et al.* 2005; Chet *et al.* 2016). We considered as valid hits the ones that exhibited 70% coverage and 70% identity against the query. With the selected genes, we conducted a core analysis with the GET_HOMOLOGUES package (Contreras-Moreira and Vinuesa 2013), in order to establish a core of virulence genes between all strains, and determine which were clade or strain specific.

In order to analyze whether the core virulence genes followed different evolutionary paths compared with the rest of the genes, a maximum likelihood phylogenetic reconstruction was performed with PhyML version 3.0 (Guindon *et al.* 2010) with the LG amino acids substitution model. Statistical support was determined with 1000 bootstrap replicates. Additionally, a 16S rRNA phylogenetic reconstruction of all the analyzed bacteria including reference and CCB strains was performed with PhyML version 3.0 with the HKY+I+G substitution model and 1000 bootstrap replicates.

Identification and phylogenetic reconstruction of Type III and VI secretion system genes

Genes related to secretion systems were extracted from a BLASTp (Altschul *et al.* 1990) search performed against the VFDB (Chen *et al.* 2005; Chet *et al.* 2016). Coding sequences for the nine T3SS core proteins previously described (Gazi *et al.* 2012; Tampakaki 2014) were retrieved from the analyzed genomes. Sequences for additional reference strains were manually retrieved from the UniProt Database (UniProt Consortium 2017). A protein alignment was performed with MUSCLE (Edgar 2004) for each of the nine sequences and phylogenetic analyses were performed with FastTree (Price, Dehal and Arkin 2010) using default parameters.

Identification of antibiotic resistance genes

In order to identify the presence of antibiotic resistance genes in all strains, a BLASTp search was performed against the Comprehensive Antibiotic Resistance Database (CARD; Jia *et al.* 2017). We considered valid hits the ones that presented 70% coverage and 70% similitude against the query.

Identification of mobile elements

A prediction of both genomic islands and pathogenicity islands for all the CCB and reference genomes was performed with the IslandViewer 3 server (Dhillon *et al.* 2015). As requested for IslandViewer predictions, the following available genomes were selected as reference: *Aeromonas veronii* B565 for the *Aeromonas* clade; *P. aeruginosa* PAO1 for the *P. aeruginosa* clade; *P. fluorescens* SBW25 for the *P. fluorescens* clade; *Vibrio anguillarum* 775 for *Vibrio* clade II to V and *Vibrio parahaemolyticus* BB22OP for *Vibrio* clade VI; and *Photobacterium profundum* SS9 for the *Photobacterium* clade. In the case of *Vibrio* and *Photobacterium* strains, predictions were done in duplicate, one for each chromosome.

We identified the presence of the three integron classes (complete integrons, CALIN and In0) in all genomes using the program IntegronFinder (Cury *et al.* 2016). The prediction of prophage sequences was done with PHASTER (Arndt *et al.* 2016). CRISPR regions were predicted through the CRISRP recognition tool (Bland *et al.* 2007). A database containing each spacer identified within the CRISPR regions was created. A BLASTn search for each spacer was performed against the created database in order to identify repeated spacers. Parameters for the programs and homemade scripts used have been deposited in GitHub (https://github.com/actevol/Comparative_genomics_of_virulence_factors).

RESULTS

General genomic features of CCB strains

16S rRNA analysis showed that the *Vibrio* formed different subclades (II–VI), related to reference strains *V. anguillarum* 775, *V. parahaemolyticus* BB22OP, *V. alginolyticus* NBRC 15630 and *V. metschnikovii* CIP 69.14 (Fig. 2a). For *Pseudomonas* we defined two clades, the *P. aeruginosa* clade and the *P. fluorescens* clade. In the case of *Aeromonas* and *Photobacterium* we only had one clade for each genus. This clustering was reflected in the genomic size, which for *Vibrio* ranged from 3.2 to 5 Mb and for *Pseudomonas* from 5.3 to 7.4 Mb, while a very narrow range was observed for *Photobacterium*, from 4.54 to 4.57 Mb and for *Aeromonas* 4.5 to 4.7 Mb (Supplementary Table S1). When compared with the GC content of references strains, most of the *Vibrio* and *Photobacterium* clades were observed to be similar to the reference strains. For *Pseudomonas* clades, we found that the endemic *P. cuatrocieneguensis* 21C1 from the *P. aeruginosa* clade had 61% GC content, which was lower than the rest of the clade, including the reference strain (66.3–66.6%). Within *Aeromonas* our strain A35_P exhibited a higher GC content (61.9%) than the rest of the group including the reference strain, which had an average of 58.3–58.7% (Supplementary Table S1).

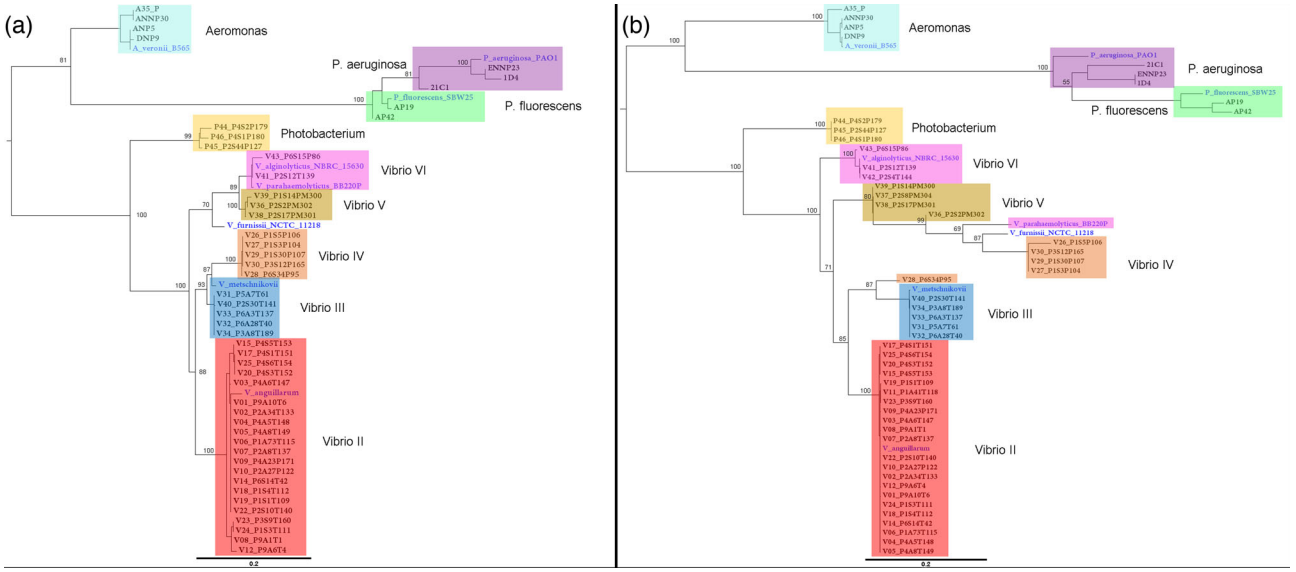


Figure 2. Comparison between phylogenetic relationships among 16S ribosomal RNA genes and 15 core virulence genes. Sequences from 51 isolates and eight reference strains (in blue) were analyzed. **(a)** 16S rRNA phylogenetic reconstruction; nine clades highlighted in colors were defined. **(b)** Virulence core phylogenetic reconstruction; clades are highlighted according to their 16S rRNA.

Virulence factor survey

We analyzed the presence of virulence factor genes in the 51 strains from CCB and the eight reference genomes, some of them pathogens, such as *P. aeruginosa* PAO1 and *V. alguinolyticus* NBRC 15630, and others from human-impacted environments, such as *A. veronii* B565 and *P. fluorescens* SBW25 (Supplementary Table S1).

We found from 62 to 164 genes encoding virulence factors in each strain of free-living CCB bacterium, while reference strains ranged from 52 to 305 (Supplementary Table S2). We further studied the presence of toxins, which may be related directly to infections. We found that the genes *ctx* from *V. cholerae* and *trh* and *tdh* from *V. parahaemolyticus* were not present in the CCB strains, and neither were the exoSTUY toxin genes from *P. aeruginosa*, while the toxin-coding gene *aexT* was present in *Aeromonas* strains ANP5, ANNP30 and DNP9, and the toxin RTX gene was present in *Vibrio* clade II reference strains *V. anguillarum* and *Aeromonas* A35_P. Nevertheless, it is possible that CCB strains possess other toxins that are not present in databases and for that reason were not identified.

Evolution of virulence factor genes

To understand the degree of conservation of the virulence factors among Gammaproteobacteria, we described a 'virulence core' set that included all virulence genes present in all analyzed strains. Fifteen genes were found to belong to this virulence core. These genes were mostly related to flagella and chemotaxis (Table 1).

When comparing the phylogenetic reconstruction of the 16S rRNA with the virulence core phylogeny (Fig. 2b), we obtained the same clades, with some exceptions within the *Vibrio* clades. In the virulence core phylogeny, the V36_P2S2PM302 strain belonging to *Vibrio* clade V seems to be ancestral to *Vibrio furnisii* and *V. parahaemolyticus*, which group within strains of clade IV, while V28_P6S34P95 splits completely from group IV to be ancestral to group III and *V. metschnikovii*.

Table 1. Virulence factor core genes.

Gene	Function	Category
flhA	Flagellar biosynthesis protein	Motility
flgB	Flagellar basal body rod protein FlgB	Motility
fliQ	Flagellar biosynthesis protein	Motility
fliR	Flagellar biosynthesis protein	Motility
htpB	Hsp60, 60K heat shock protein HtpB	Heat shock
cheW-2	Chemotaxis protein CheW	Chemotaxis
flgI	Flagellar P-ring protein precursor	Motility
fliG	Flagellar motor protein	Motility
cheB	Chemotaxis-specific methylesterase	Chemotaxis
flhG	Flagellar synthesis regulator FleN	Motility
fliN	Polar flagellar switch protein FliN	Motility
flgF	Flagellar basal-body rod protein FlgF	Motility
cheR	Chemotaxis protein methyltransferase CheR	Chemotaxis
fliE	Flagellar hook-basal body protein	Motility
fleR/flrC	Sigma-54-dependent response regulator	Motility

To identify which genes were influenced by HGT events, individual phylogenies for each core gene were performed. This analysis showed that gene *flhA* had a second copy. Some members of the *Aeromonas* clade, as well as all strains from *Photobacterium* and *Vibrio* clades IV, V and VI and the reference strain of *V. parahemolyticus* had this second copy of the gene related to flagella formation (Supplementary Figure S1). The second copy of this gene may represent a duplication event or an acquisition by HGT among different lineages.

Secretion systems, a cell–cell communication framework

To understand the importance of these complex bio-machines in the bacteria inhabiting CCB, we looked for the described core elements that composed T3SS (Gazi *et al.* 2012; Tampakaki, 2014) in CCB strains. We found between 23 and 37 genes related to this secretion system, which included regulators,

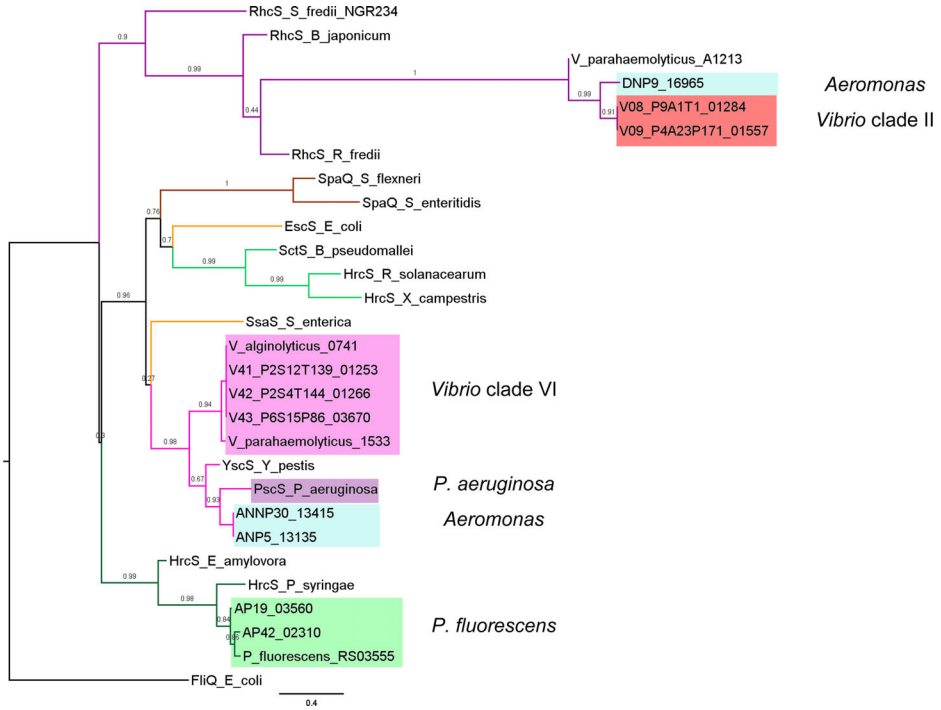


Figure 3. Phylogenetic reconstruction of *SctS* genes. Seven evolutionary groups of T3SS previously defined (Gazi *et al.* 2012; Tampakaki 2014) are represented with different branch colors within the phylogeny. Clades of CCB strains and the reference strains are highlighted using the previous color code. Gene *FliQ* from *E. coli* was used as the outgroup.

structural and effector genes. For example, many *Vibrio* from clade VI had genes with homology to T3SS1 from *V. parahaemolyticus*, as well as *Aeromonas* strains ANNP30 and ANP5. Both genera possess homologues to *ExsA* and *PscC*, a master regulator and a key protein in the formation of the injectisome, respectively (Hauser 2009). The presence of these two genes indicates that T3SS may be functional in CCB strains. Interestingly, *Aeromonas* strain DNP9 possesses only four proteins of the T3SS, but they are homologous to the T3SS2 of *V. parahaemolyticus* and *Sinorhizobium fredii* NGR234, as is the case for strains V08_P9A1T1 and V09_P4A23P171 of clade II (Suplementary Table S3). Although T3SS is widely distributed among environmental and pathogenic *Pseudomonas* strains, we found from one to four genes related to T3SS among genomes of *Pseudomonas* from CCB strains, which suggests that this system is incomplete and may not be functional in these environmental strains (Supplementary Table S3).

A phylogenetic analysis of the core elements of T3SS showed that these components have evolved in seven different groups (Gazi *et al.* 2012; Tampakaki 2014). To test whether CCB strains conform to a separate group, we performed a phylogenetic analysis using core genes of T3SS (Fig. 3). We found that *Vibrio* clade VI, *Aeromonas* strains ANNP30 and ANP5 cluster within the Ysc-T3SS family, which has been reported to confer resistance to phagocytosis and trigger macrophage apoptosis (Gazi *et al.* 2012). On the other hand, *Aeromonas* strain DNP9 and *Vibrio* clade II strains V08_P9A1T1 and V09_P4A23P171 cluster together with the T3SS2 copy of *V. parahaemolyticus* among the Rhizobiales T3SS family, which is dedicated to the intimate endosymbiosis of nitrogen fixation in the roots of leguminous plants. Genes of T3SS of CCB *Pseudomonas* form a cluster among the Hrc-Hrp-T3SS family, which is present in plant pathogenic bacteria of the genera *Pseudomonas, Erwinia, Ralstonia* and *Xanthomonas*.

We also searched for components related to the type VI secretion system (T6SS). As expected, we found that this system was present in many of the CCB strains (Fig. 4). Based on annotation, 12–22 genes were found to be associated to T6SS, among *Vibrio* clade II strains, the *P. fluorescens* clade, strain ENNP23 of the *P. aeruginosa* clade and most *Aeromonas* except A35_P, which only possesses four related proteins. This could suggest that in this particular environment, the interactions occurring among bacteria are more important than those with eukaryotes (Supplementary Table S4).

Antibiotic resistance

To estimate the potential of CCB strains for antibiotic resistance, we performed a search against an antibiotic resistance gene-curated database. This analysis showed different strategies among the analyzed genera (Fig. 4; Supplementary Table S5). *Pseudomonas*, in general, displayed a great abundance of multidrug efflux systems in both CCB and reference strains; nevertheless, genes for resistance to β-lactam were only identified in the reference strain *P. aeruginosa* PAO1. In contrast, CCB *Aeromonas* strains had a wide range of antibiotic resistances, including β-lactamases, multidrug efflux systems, and resistance to bacitracin, tetracycline and streptogramins. *Vibrio* and *Photobacterium* lacked resistance genes for β-lactams, except for two isolates closely related to *V. parahaemolyticus* from CCB. However, in all of them, we observed the *tet34* gene, which confers resistance to oxytetracycline, as well as tetracycline, chloramphenicol and the multidrug efflux system.

Mobile elements, genomic and pathogenicity islands

We analyzed the presence of genomic islands (GIs) and pathogenicity islands (PAIs) in our sample collection. The num-

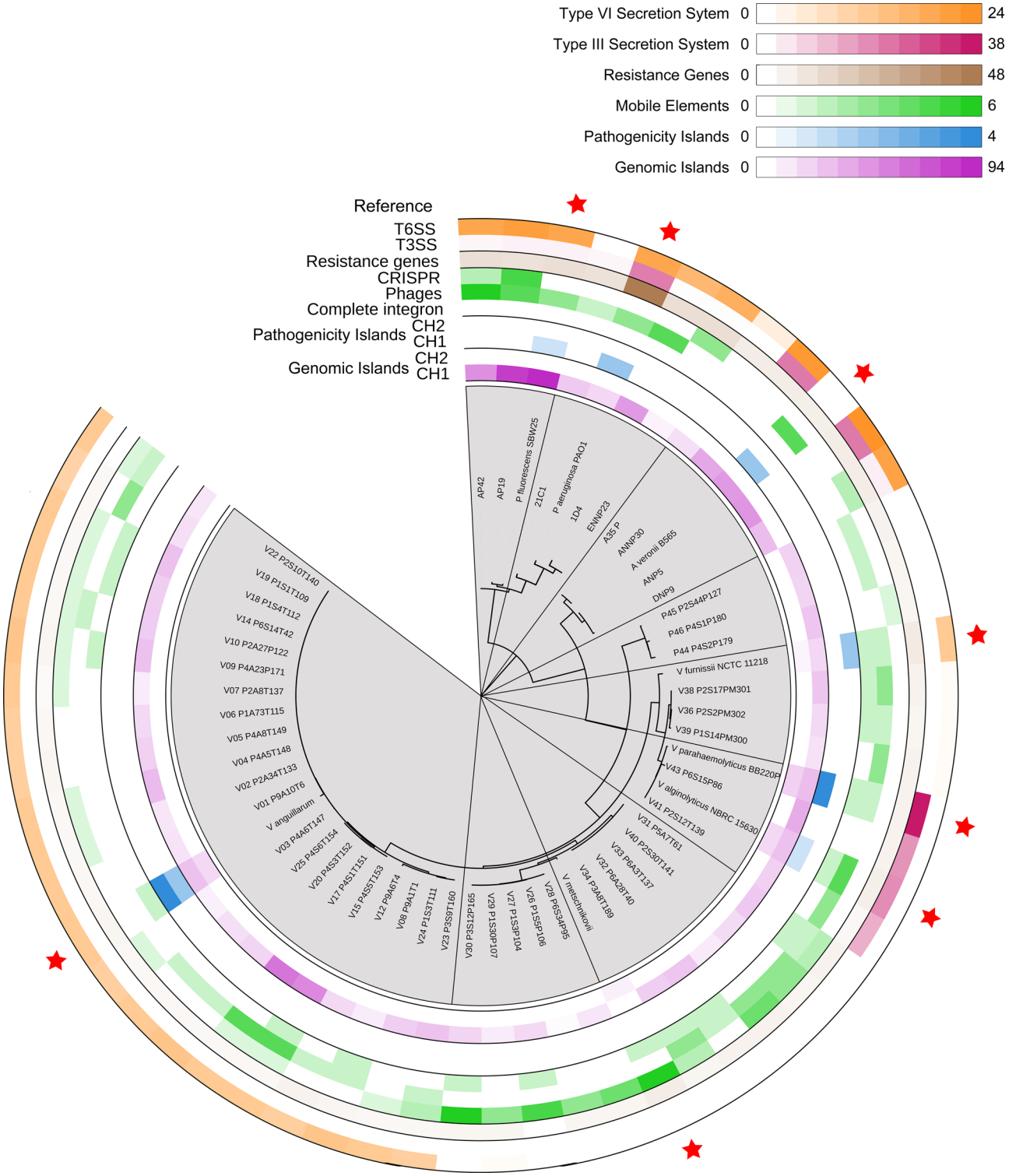


Figure 4. Schematic representation of virulence and pathogenicity elements. 16S rRNA phylogenetic reconstruction. Inner circles (from the inside outwards) represent the presence of genomic islands, pathogenicity islands, mobile elements, resistance genes, and abundance of secretion systems III and VI. Colored stars represent reference strains. CH1 and CH2 represent chromosome I and chromosome II in *Vibrio*. These values are only presented for reference strains of *Vibrio*, given that CCB strains are not are not assembled in chromosomes. For CCB strains, island values for both chromosomes are displayed as chromosome II. Previously defined clades are presented in gray clockwise: *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Aeromonas*, *Photobacterium* and *Vibrio* clades V, VI, III, IV and II. Image generated with iTOL (Letunic and Bork, 2016).

ber of predicted GIs ranged from 2 to 86 (Fig. 4). Interestingly, only in the reference strains were there found both PAIs and GIs (Fig. 4 and Supplementary Table S6). Genes that were represented in GIs of CCB strains were mostly hypothetical; nevertheless, we did find transposon-related genes, integrases and some genes of phage components. In the case of *Photobacterium* we found the genes *csgBACEFG*, involved in secretion of the extracellular matrix of biofilm. Also within this clade we found the genes *zraS* and *zraR*, which are related to the response to zinc and lead stress.

We searched for insertion sites as integrons in CCB and reference strains. CCB *Vibrio* strains contain the three types of previously defined integron elements: complete integrons, CALIN and In0 (Cury *et al.* 2016). Nevertheless, only 25.6% of these isolates possessed complete integrons, and attC sites were an order of magnitude lower in CCB *Vibrio* than in the reference strains. In the case of *Pseudomonas*, *Aeromonas* and *Photobacterium*, we observed strains containing only CALIN-type integrons, i.e. they do not contain integrases (Fig. 4; Supplementary Table S6).

Phages and CRISPRs

We analyzed lysogenic phages in our strains as these represent one of the most efficient mechanisms of HGT (Jiang and Paul 1998; Weinbauer and Rassouldzadegan 2004; Moon *et al.* 2016). The presence of prophages was variable among the analyzed strains. No complete prophages were found in either CCB *Aeromonas* or *Vibrio* clade IV strains. In contrast, the *P. fluorescens* clade, *Vibrio* clade III and *Vibrio* clade V strains harbored at least one complete prophage. It is of particular interest that these 'dormant' viruses were present in most of the reference strains, except *V. anguillarum* and *V. metschnikovii*, in which only incomplete prophages were found. The number of prophages identified in the reference strains varied from one to three in *Vibrio*, two for *Pseudomonas* and three in the case of *Aeromonas* (Fig. 4; Supplementary Table S6).

Prediction analysis using the CRISPR recognition tool (Bland *et al.* 2007) identified these elements in almost all analyzed clades from the CCB, varying from one to six CRISPR regions (Fig. 4; Supplementary Table S6). All the strains from the *Vibrio* clades III and IV and the *P. fluorescens* clade harbored CRISPR regions. In contrast, *Vibrio* clades V and VI did not exhibit any of these regions. *Aeromonas*, *Photobacterium* and *Vibrio* clade II showed a variable presence of these elements. Except for the reference strain *V. metschnikovii* CIP 69.14, none of the reference strains had CRISPR sequences. Each strain of the *Pseudomonas*, *Aeromonas* and *Photobacterium* clades harbored unique sequence spacers, except for two spacers shared within the *P. fluorescens* clade. In the case of *Vibrio*, we found that most spacers were shared between strains within each clade. This observation could suggest that each clade has a unique history regarding its phage protection system.

DISCUSSION

Probably since the early communities in the Achaean, bacteria have evolved a wide range of genetic strategies to interact with each other and with their environment (Bingle, Bailey and Pallen 2008; Schwarz *et al.* 2010; Riley and Wertz 2002; Morris *et al.* 2013). Since these ancestral genes are present in all lineages, including pathogens, there has been a historical confusion in the description of their role. This was due to our human-centric view of microbes and to most of the knowledge we have on gene function coming from the study of pathogenic strains associated with humans or plants of human interest. Therefore our work of comparative genomic analysis seeks to gain insight into the distribution of the so-called virulence/pathogenesis genes among genomes from pathogenic strains, human-impacted strains and strains from the extremely oligotrophic oasis of CCB.

First, we identified virulence factors in genomes from strains from CCB and from reference strains. We found a broad range of virulence factors in CCB strains from different genera analyzed in this study. The conservation of these genes in free-living bacteria of a relatively pristine environment suggests that they are not inherently related to pathogenicity, but they are part of the adaptations of free-living microbes. For instance, many virulence core genes are related to flagella formation and chemotaxis, features that allow bacteria to move towards attractants and away from repellents. In CCB these genes could affect the microbial community dynamics, as these communities are in part assembled based on a metabolic codependence (Rodríguez-Torres *et al.* 2017), as well as by active chemical warfare in the search for resources by destroying cheaters and foreigners (Pérez-Gutierrez *et al.* 2013; Aguirre-von-Wobeser *et al.* 2014; Ponce-Soto *et al.* 2015; Velez *et al.* 2016). The ability to sense the environment and move to optimal conditions may therefore increase bacterial fitness in the harsh CCB environments.

Interestingly, when we compared the 16S rRNA tree against the virulence core phylogeny, we found a shift in the phylogenetic position for three *Vibrio* strains. These phylogenetic incongruences may suggest early recombination events, which apparently occurred before the diversification of both *V. parahaemolyticus* and *V. furnissii*. Individual phylogenies of core genes showed a second copy of the flagella gene *flhA* in *V. parahaemolyticus* (Supplementary figure S1). This could represent either an HGT event or more likely a duplication that occurred before the divergence of these groups given the grouping of the sequences (Supplementary figure S1).

Moreover, this duplicated flagellar gene has a homolog in the T3SS (Gazi *et al.* 2012; Tampakaki 2014), making this story even more complex and in need of further exploration. As previously stated, secretion systems are involved in the transport of biomolecules through the cell envelope. Several studies have shown that these secretion systems have evolved by horizontal gene transfer events (Makino *et al.* 2003; Okada *et al.* 2010). However, when we analyzed the T3SS core elements (Fig. 3), we found that the seven evolutionary lineages of this system are conserved and that CCB strains do not constitute a separate group. Another probable horizontal gene transfer event is proposed among *Vibrio* clade II and T3SS2 of *V. parahaemolyticus* as well as with *Aeromonas* strain DNP9. This is particularly interesting since recombination events seem to be rare and ancient in CCB bacteria (Avitia *et al.* 2014). All the other recorded HGT events, such as the acquisition of a rhodopsin gene and the ability to produce sulpholipids in *Bacillus coahuilensis* (Alcaraz *et al.* 2008), and the movement of genes for phosphorus mobilizations in bacilli (Moreno-Letelier *et al.* 2011), are ancestral, adaptive and very well trimmed. In contrast with *Aeromonas* and *Vibrio*, we did not find most of the T3SS elements in *Pseudomonas* associated to CCB even though we detected some components. This may suggest that purifying selection has eroded the Type III secretion system in most of the local strains (Fig. 4), or that there are undescribed homologs, since eukaryotes are not only present but diverse and abundant as in the case of fungi (Velez *et al.* 2016).

Even though the Type VI secretion system is abundant in nature and has been proposed to mediate interactions among

bacteria, we did not find this system complete in most of our isolates from CCB (Fig. 4), suggesting that with the isolation of this ecosystem for a very long time, the bacteria have evolved the parallel T6SS whose homologies with those previously described are not clear. Further research should be done with more genomes as well as metagenomes for putative members of this secretion system.

For example, we found several resistance genes in CCB isolates and most of them involved the pump efflux system although these genes did not always match the phenotype. For the strains *Pseudomonas* 1D4, ENNP23, AP19 and AP42, and *Aeromonas* ANP5, ANNP30 and DNP9, antibiotic resistance assays were performed (Ponce-Soto *et al.* 2015). The genomic analysis of genes associated with antibiotic production and resistance, in particular β-lactams, aminoglycosides and tetracycline antibiotics, were not always congruent with the resistance phenotypes, as some strains that were sensitive to a given antibiotic carry the resistance genes. It is possible that these genes were not expressed under the growth conditions used for the assays, or more likely, that they are induced only under antagonistic conditions. We hypothesize that this variability in survival strategies (and possible associated trade-off costs) leads to an evolutionary dynamic based on competence and costs. This type of interaction may also happen in human-impacted environments, as is the case of *A. veronii* B565, which possess resistance genes to several antibiotics and was isolate from the sediment of an aquaculture pond (Li *et al.* 2011).

When we analyzed the presence of virulence factors among the CCB strains, we found that these strains had individual genes associated with virulence, such as toxin and antibiotic production and resistance, but in none of the strains were these different virulence genes contiguous (i.e. in tandem) in a cassette configuration such as an integron or a pathogenicity island. Our findings for CCB strains supports our previous ideas that pathogenic strains restructure their genome in order to be more efficient in expressing virulence genes together, in a cassette as is the case of the LEE locus in *E. coli* (McDaniel *et al.* 1995; Castillo, Eguiarte and Souza 2005).

Since several virulence factors and antibiotic resistance determinants are shared among bacteria, we looked for common markers of HGT/foreign DNA insertion. In contrast with what has been found in the reference strains, few isolates from the CCB contained mobile elements besides CRISPR sequences and a few *Vibrio* integrons. Recently, it has been proposed that a negative correlation exists between the presence of CRISPR regions, their spacers and the presence of horizontal transfer elements such as insertion sequences (García-Gutiérrez *et al.* 2015; Sheludchenko *et al.* 2015). It seems that CRISPR spacers act as barriers against mobile elements, by recognizing them as foreign elements. Our study corroborates the negative correlation between the presence of CRISPR sequences and mobile elements such as phages and integrons; in approximately 69% of the analyzed strains in which CRISPRs were present neither phages nor integrons were observed (Fig. 4; Supplementary Table S6). Our findings support the idea that the costs for extra DNA replication could be a counter-selection force against foreign DNA incorporation in this oligotrophic environment (Souza *et al.* 2008).

One Health and the ecology of pathogenicity

Bacteria have been evolving for almost 4 billion years and have survived because they have evolved many genetic systems that allow them to adjust to their changing environments including the changing microbial neighborhood. Genes for mobility, sensing and transport systems have been cataloged as pathogenesis and virulence related for historical reasons since they were first discovered in pathogenic strains. However, our study shows that they simply have a role in community structure either by signaling membership or by repelling non-members of the microbial market.

When referring to environmental bacteria, in this case strains isolated from CCB, we cannot specifically make a distinction between environmental and pathogenic behavior since bacterial coexistence entails both cooperation and competition. In these nutrient-limited communities, bacteria seem to have evolved to recognize 'members' from 'non-members', since membership would be essential for cost-effective ecological cooperation (Werner *et al.* 2014). In a place as diverse in bacteria as CCB, this membership appears to be particularly important, since each member of the community has a role in nutrient cycling and in the survival of the community (Peimbert *et al.* 2012). We have even formed the hypothesis that when food is scarce, HGT is rare and migration area is small (Souza *et al.* 2012a), keeping the membership local and strict. The lessons of CCB are clear and relevant: when anthropogenic influxes of nutrients or contaminants perturb an ecosystem, the resulting microbial communities are more likely to acquire genes by HGT to adapt faster to new environments. In CCB this has been experimentally shown as nutrient input changes the ecological dynamics of bacterial communities (Lee *et al.* 2017); the effect is similar to what happens when sewer-contaminated water goes through the food chain to our gut. This has also been demonstrated by the impressive and fast evolution of bacterial resistance to antibiotics in several anthropogenic environments, in particular in hospitals (Modi *et al.* 2013; Feugeas *et al.* 2016; Stalder and Top 2016). Additionally, recent studies have shown that the acquisition of integrons and resistance genes by microbiota of wild animals is associated with this contamination (Power, Emery and Gillings 2013; Cristobal-Azkarate *et al.* 2014), generating a concern regarding the propagation of antibiotic resistances in the environment, increasing the risk of cross infections among animals and humans in their transition zones.

The One Health concept aims to integrate concepts from disciplines such as medicine, veterinary science and ecology, among others, helping to eliminate the artificial barriers created in the modern treatment of diseases, barriers to which microbes are obviously indifferent. In this regard, our work contributes to the understanding that the genes considered as pathogenic are survival tools for bacteria in the environment. This work also shows how disturbances in the environment will be reflected in our health, directly by being exposed to contaminants, but also more subtly by being exposed to a microbiome that has a higher probability of acquiring virulence genes in cassettes such as pathogenicity islands.

SUPPLEMENTARY DATA

Supplementary data are available at *FEMSPD* online.

ACKNOWLEDGEMENTS

MV-R-L is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 345250 from CONACYT. GYP-S is a doctoral student from Programa de Doctorado en Ciencias Biológicas, UNAM and acknowledges a scholarship from CONACYT (Grant no. 269818). We acknowledge Dr Laura

Espinosa-Asuar and Dr Erika Aguirre Planter for their technical assistance. In addition, we strongly acknowledge the reviews of Dr Edmundo Calva, Dr Eneas Aguirre, Dr Eria Rebollar, Dr Gabriela Olmedo and Dr Jorge Valdivia. We are grateful to David Jaramillo for providing the photographs used in this article.

FUNDING

This study was conducted with financial support from Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) project IG200215.

REFERENCES

Adams Krumins J, Dighton J, Gray D *et al.* Soil microbial community response to nitrogen enrichment in two scrub oak forests. *Forest Ecol Manag* 2009;**258**:1383–90.

Aguirre-von-Wobeser E, Soberon-Chavez G, Eguiarte LE *et al.* Two-role model of an interaction network of free-living gamma-proteobacteria from an oligotrophic environment. *Environ Microbiol* 2014;**16**:1366–77.

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.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Arndt D, Grant JR, Marcu A *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;**44**:W16–21.

Austin B. Vibrios as causal agents of zoonoses. *Vet Microbiol* 2010;**140**:310–7.

Avitia M, Escalante AE, Rebollar EA *et al.* Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. *PeerJ* 2014;**2**:e696.

Bik EM. The hoops, hopes, and hypes of human microbiome research. *Yale J Biol Med* 2016;**89**:363–73.

Bingle LE, Bailey CM, Pallen MJ. Type VI secretion: a beginner's guide. *Curr Opin Microbiol* 2008;**11**:3–8.

Bland C, Ramsey TL, Sabree F *et al.* CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 2007;**8**:209.

Boetzer M, Henkel CV, Jansen HJ *et al.* Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 2011;**27**:578–9.

Bonilla-Rosso G, Peimbert M, Alcaraz LD *et al.* Comparative metagenomics of two microbial mats at Cuatro Cienegas Basin II: community structure and composition in oligotrophic environments. *Astrobiology* 2012;**12**:659–73.

Canchaya C, Fournous G, Chibani-Chennoufi S *et al.* Phage as agents of lateral gene transfer. *Curr Opin Microbiol* 2003;**6**:417–24.

Carson EW, Tobler M, Minckley WL *et al.* Relationships between spatio-temporal environmental and genetic variation reveal an important influence of exogenous selection in a pupfish hybrid zone. *Mol Ecol* 2012;**21**:1209–22.

Castillo A, Eguiarte LE, Souza V. A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: The search for the unit of selection. *Proc Natl Acad Sci U S A* 2005;**102**:1542–7.

Chen LH, Yang J, Yu J *et al.* VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 2005;**33**:D325–8.

Chen LH, Zheng DD, Liu B *et al.* VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 2016;**44**:D694–7.

Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 2013;**79**:7696–701.

Costa TR, Felisberto-Rodrigues C, Meir A *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* 2015;**13**:343–59.

Cristobal-Azkarate J, Dunn JC, Day JMW *et al.* Resistance to antibiotics of clinical relevance in the fecal microbiota of mexican wildlife. *PLoS One* 2014;**9**:e107719.

Cury J, Jove T, Touchon M *et al.* Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res* 2016;**44**:4539–50.

Czaran TL, Hoekstra RL, Pagie L. Chemical warfare between microbes promotes biodiversity. *Proc Natl Acad Sci U S A* 2002;**99**:786–90.

Desnues C, Rodriguez-Brito B, Rayhawk S *et al.* Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* 2008;**452**:340–3.

Dhillon BK, Laird MR, Shay JA *et al.* IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res* 2015;**43**:W104–8.

Díaz-Mejia JJ, Amabile-Cuevas CF, Rosas I *et al.* An analysis of the evolutionary relationships of integron integrases, with emphasis on the prevalence of class 1 integrons in *Escherichia coli* isolates from clinical and environmental origins. *Microbiology* 2008;**154**:94–102.

Diggle SP, Gardner A, West SA *et al.* Evolutionary theory of bacterial quorum sensing: when is a signal not a signal. *Philos Trans R Soc Lond B Biol Sci* 2007;**362**:1241–9.

Dobrindt U, Hochhut B, Hentschel U *et al.* Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2004;**2**:414–24.

Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.

Elser JJ, Schampel JH, Garcia-Pichel F *et al.* Effects of PO_4 enrichment and grazing snails on microbial communities in an ecosystem with living stromatolites. *Freshwater Biol* 2005;**50**:1808–25.

Epstein SS, Microbial interactions. *Els* 2003;0003285.

Escalante AE, Caballero-Mellado J, Martínez-Aguilar L *et al.* 2009. *Pseudomonas cuatrocienegasensis* sp. nov., isolated from an evaporating lagoon in the Cuatro Cienegas valley in Coahuila, Mexico. *Int J Syst Evol Microbiol* **59**:1416–20.

Escalante AE, Eguiarte LE, Espinosa-Asuar L *et al.* Diversity of aquatic prokaryotic communities in the Cuatro Cienegas basin. *FEMS Microbiol Ecol* 2008;**65**:50–60.

Espinosa-Asuar L, Escalante AE, Gasca-Pineda J *et al.* Aquatic bacterial assemblage structure in Pozas Azules, Cuatro Cienegas Basin, Mexico: Deterministic vs. stochastic processes. *Int Microbiol* 2015;**18**:105–15.

Fajardo A, Martinez-Martin N, Mercadillo M *et al.* The neglected intrinsic resistome of bacterial pathogens. *PLoS One* 2008;**3**:e1619.

Fancello L, Desnues C, Raoult D *et al.* Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *J Antimicrob Chemother* 2011;**66**:2448–54.

Feugeas JP, Tourret J, Launay A *et al.* Links between transcription, environmental adaptation and gene variability in *Escherichia coli*: correlations between gene expression and gene variability reflect growth efficiencies. *Mol Biol Evol* 2016;**33**:2515–29.

Finn RD, Bateman A, Clements J *et al.* Pfam: the protein families database. *Nucleic Acids Res* 2014;**42**:D222–30.

Garcia-Gutierrez E, Almendros C, Mojica FJ *et al.* CRISPR content correlates with the pathogenic potential of *Escherichia coli*. *PLoS One* 2015;**10**:e0131935.

Gazi AD, Sarris PF, Fadouloglou VE *et al.* Phylogenetic analysis of a gene cluster encoding an additional, rhizobial-like type III secretion system that is narrowly distributed among *Pseudomonas syringae* strains. *BMC Microbiol* 2012;**12**:188.

Grosso-Becerra M-V, Santos-Medellin C, Gonzalez-Valdez A *et al.* *Pseudomonas aeruginosa* clinical and environmental isolates constitute a single population with high phenotypic diversity. *BMC Genomics* 2014;**15**:318.

Guindon S, Dufayard JF, Lefort V *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**: 307–21.

Hauser AR. The Type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Microbiol* 2009;**7**:654–65.

Hazen TH, Pan L, Gu JD *et al.* The contribution of mobile genetic elements to the evolution and ecology of Vibrios. *FEMS Microbiol Ecol* 2010;**74**:485–99.

Hershler R. The hydrobiid snails (Gastropoda: Rissoacea) of the Cuatro Cienegas Basin: systematic relationships and ecology of a unique fauna. *J Arizona Nevada Acad Sci* 1984;**19**:61–76.

Horn M, Collingro A, Schmitz-Esser S *et al.* Illuminating the evolutionary history of Chlamydiae. *Science* 2004;**304**:728–7302:

Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207–14.

Jia B, Raphenya AR, Alcock B *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017;**45**:D566–73.

Jiang SC, Paul JH. Gene transfer by transduction in the marine environment. *Appl Environ Microbiol* 1998;**64**:2780–7.

Johnson JR. Virulence factors in *Escherichia coli* urinary tract infection. *Clin Microb Rev* 1991;**4**:80–128.

Kapitein N, Mogk A. Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. *Curr Opin Microbiol* 2013;**16**:52–8.

Koster M, Bitter W, Tommassen J. Protein secretion mechanisms in Gram-negative bacteria. *Int J Med Microbiol* 2000;**290**:325–31.

Lee ZM, Poret-Peterson AT, Siefert JL *et al.* Nutrient stoichiometry shapes microbial community structure in a shallow evaporitic pond. *Front Microbiol* 2017. doi: 10.3389/fmicb.2017.00949.

Lee ZM, Steger L, Corman JR *et al.* Response of a stoichiometrically imbalanced ecosystem to manipulation of nutrient supplies and ratios. *PLoS One* 2015. doi:10.1371/journal.pone.0123949.

Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;**44**:W242–5.

Li Y, Liu Y, Zhou Z *et al.* Complete genome sequence of *Aeromonas veronii* strain B565. *J Bacteriol* 2011;**193**:3389–90.

Lopez-Lozano NE, Eguiarte LE, Bonilla-Rosso G *et al.* Bacterial communities and the nitrogen cycle in the gypsum soils of Cuatro Cienegas Basin, Coahuila: A Mars analogue. *Astrobiology* 2012;**12**:699–709.

Lopez-Lozano NE, Heidelberg KB, Nelson WC *et al.* Microbial secondary succession in soil microcosms of a desert oasis in the Cuatro Cienegas Basin, Mexico. PeerJ 2013;**1**:e47.

McDaniel T, Jarvis K, Donnenberg M *et al.* A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci U S A* 1995;**92**:1664–8.

Makino W, Cotner JB. Elemental stoichiometry of a heterotrophic bacterial community in a freshwater lake: implications for growth- and resource-dependent variations. *Aquat Microbiol Ecol* 2005;**34**:33–41

Makino K, Oshima K, Kurokawa K *et al.* Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V cholerae. *Lancet* 2003;**361**: 743–9

Manyahi J, Tellevik MG, Ndugulile F *et al.* Molecular characterization of cotrimoxazole resistance genes and their associated integrons in clinical isolates of Gram-negative bacteria from Tanzania. *Microb Drug Resist* 2017;**23**:37–43.

Marie C, Broughton WJ, Deakin WJ. Rhizobium type III secretion systems: legume charmers or alarmers? *Curr Opin Plant Biol* 2001;**4**:336–42.

Marie C, Deakin WJ, Viprey V *et al.* Characterization of Nops, nodulation outer proteins, secreted via the type III secretion system of NGR234. *Mol Plant Microbe Interact* 2003;**16**:743–51.

Mazel D, Dychinco B, Webb VA *et al.* A distinctive class of integron in the *Vibrio cholerae* genome. *Science* 1998;**280**:605–8.

Méthot P-O, Alizon S. What is a pathogen? Toward a process view of host-parasite interactions. *Virulence* 2014;**5**:775–85.

Minckley W. Cuatro Cienegas fishes: research review and a local test of diversity versus habitat size. *J Arizona Nevada Acad Sci* 1984;**19**:13–21.

Modi SR, Lee HH, Spina CS *et al.* Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* 2013;**499**:219–22.

Moon BY, Park JY, Robinson DA *et al.* Mobilization of genomic islands of *Staphylococcus aureus* by temperate bacteriophage. *PLoS One* 2016;**11**:e0151409.

Moreno-Letelier A, Olmedo G, Eguiarte LE *et al.* Parallel evolution and horizontal gene transfer of the pst operon in Firmicutes from oligotrophic environments. *In J Evol Biol* 2011;**2011**:781642.

Morris B, Henneberger R, Huber H *et al.* Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* 2013;**37**:384–406.

Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 2012;**13**(Suppl 14):S8.

Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 1984;**157**:690–3.

Okada N, Matsuda S, Matsuyama J *et al.* Presence of genes for type III secretion system 2 in *Vibrio mimicus* strains. *BMC Microbiology* 2010;**10**:302.

Okazaki S, Kaneko T, Sato S *et al.* Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proc Natl Acad Sci U S A* 2013;**110**:17131–6.

Pajares S, Escalante AE, Noguez AM *et al.* Spatial heterogeneity of physicochemical properties explains differences in microbial composition in arid soils from Cuatro Cienegas, Mexico. *PeerJ* 2016;**4**:e2459.

Parret AH, De Mot R. Bacteria killing their own kind: novel bacteriocins of *Pseudomonas* and other γ-proteobacteria. *Trends Microbiol* 2002;**10**:107–12.

Peek AS, Souza V, Eguiarte LE *et al.* The interaction of protein structure, selection, and recombination on the evolution of

the type-1 fimbrial major subunit (fimA) from *Escherichia coli*. *J Mol Evol* 2001;**52**:193–204.

Peimbert M, Alcaraz LD, Bonilla-Rosso G *et al*. Comparative metagenomics of two microbial mats at Cuatro Cienegas Basin I: ancient lessons on how to cope with an environment under severe nutrient stress. *Astrobiology* 2012;**12**:648–58.

Perez-Gutierrez R-A, Lopez-Ramirez V, Islas A *et al*. Antagonism influences assembly of a *Bacillus* guild in a local community and is depicted as a food-chain network. *ISME J* 2013;**7**:487–97.

Ponce-Soto GY, Aguirre-von-Wobeser E, Eguiarte LE *et al*. Enrichment experiment changes microbial interactions in an ultra-oligotrophic environment. *Front Microbiol* 2015;**6**:246.

Power ML, Emery S, Gillings MR, Into the wild: dissemination of antibiotic resistance determinants via a species recovery program. *PLoS One* 2013;**8**:e63017.

Price MN, Dehal PS, Arkin AP, FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490.

Rainey PB, Bailey MJ. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol Microbiol* 1996;**19**:521–33.

Riley MA, Gordon DM, The ecological role of bacteriocins in bacterial competition. *Trends Microbiol* 1999;**7**:129–33.

Riley MA, Wertz JE. Bacteriocins: Evolution, ecology and application. *Annu Rev Microbiol* 2002;**56**:117–37.

Rodríguez-Torres M, Islas-Robles A, Hernandez-Gonzalez I *et al*. Phenotypic microdiversity and phylogenetic signal analysis of traits related to social interaction in *Bacillus* spp. from sediment communities. *Front Microbiol* 2017;**8**:29.

Rodríguez-Verdugo A, Souza V, Eguiarte LE *et al*. Diversity across seasons of culturable *Pseudomonas* from a desiccation lagoon in Cuatro Ciénegas, Mexico. *Int J Microbiol* 2012;**2012**, 201389.

Salyers AA, Amabile-Cuevas CF. Why are antibiotic resistance genes so resistant to elimination. *Antimicrob Agents Chemother* 1997;**41**:2321–5.

Sandner L, Eguiarte LE, Navarro A *et al*. The elements of the locus of enterocyte effacement in human and wild mammal isolates of *Escherichia coli*: evolution by assemblage or disruption. *Microbiology* 2001;**147**:3149–58.

Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 2011;**6**:e17288.

Schmieder R, Lim YW, Rohwer F *et al*. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* 2010;**11**:341.

Schwarz S, West E, Boyer F *et al*. Burkholderia Type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *Plos Pathog* 2010;**6**:e1001068.

Scully EP, Earp BE, Miller AL *et al*. Just a cut. *N Engl J Med* 2016;**375**:1780–6.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9.

Sheludchenko MS, Huygens F, Stratton H *et al*. CRISPR Diversity in *E. coli* isolates from australian animals, humans and environmental waters. *PLoS One* 2015;**10**:e0124090.

Souza V, Eguiarte LE, Siefert J *et al*. Microbial endemism: does phosphorus limitation enhance speciation? *Nat Rev Microbiol* 2008;**6**:559–64.

Souza V, Espinosa-Asuar L, Escalante AE *et al*. An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc Natl Acad Sci U S A* 2006;**103**:6565–70.

Souza V, Rocha M, Valera A *et al*. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl Environ Microbiol* 1999;**65**:3373–85.

Souza V, Eguiarte LE, Travisano M *et al*. Travel, sex, and food: what's speciation got to do with it? *Astrobiology* 2012a;**12**:634–40.

Souza V, Siefert JL, Escalante AE *et al*. The Cuatro Cienegas Basin in Coahuila, Mexico: An astrobiological Precambrian park. *Astrobiology* 2012b;**12**:641–7.

Stalder T, Top E. Plasmid transfer in biofilms: a perspective on limitations and opportunities. *NPJ Biofilms Microbiomes* 2016;**2**:16022.

Subirats J, Sanchez-Melsio A, Borrego CM *et al*. Metagenomic analysis reveals that bacteriophages are reservoirs of antibiotic resistance genes. *Int J Antimicrob Agents* 2016;**48**:163–7.

Tampakaki AP. Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria. *Front Plant Sci* 2014;**5**:114.

Tapia-Torres Y, Rodríguez-Torres MD, Elser JJ *et al*. How to live with phosphorus scarcity in soil and sediment: lessons from bacteria. *Appl Environ Microbiol* 2016;**82**:4652–62.

UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;**45**:D158–69.

Velez P, Gasca-Pineda J, Rosique-Gil E *et al*. Microfungal oasis in an oligotrophic desert: diversity patterns and community structure in three freshwater systems of Cuatro Cienegas, Mexico. *PeerJ* 2016;**4**:e2064.

Viprey V, Del Greco A, Golinowski W *et al*. Symbiotic implications of type III protein secretion machinery in *Rhizobium*. *Mol Microbiol* 1998;**28**:1381–9.

Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol* 2004;**6**:1–11.

Werner GD, Strassmann JE, Ivens AB *et al*. Evolution of microbial markets. *Proc Natl Acad Sci U S A* 2014;**111**:1237–44.

Yang S, Tang F, Gao MQ *et al*. R gene-controlled host specificity in the legume–rhizobia symbiosis. *Proc Natl Acad Sci U S A* 2010;**107**:18735–40.

Capítulo 5

Discusión y conclusión

El concepto de especie.

Como plantea Vos (2011), el concepto de especie debe reflejar procesos evolutivos. Actualmente, no existe una forma práctica que permita delimitar especies con sentido evolutivo, los esfuerzos que se han hecho para entender la diversidad microbiana están basados en métodos que pueden llegar a ser considerados arbitrarios pero prácticos (Achman & Wagner, 2008; Sogin *et al.,* 2006). La propuesta del concepto adaptativo de especie es atractiva pues busca explicar la diversidad microbiana basada en evidencias de divergencia explicadas por la selección. En nuestro análisis observamos dos cosas principales respecto al uso de este concepto. La primera es meramente conceptual, a pesar de que existen evidencias de adaptación y divergencia asociada a la estructuración ambiental (Cadillo-Quiroz *et al.,* 2012; Shapiro *et al.,* 2012), consideramos que antes de aplicarlo de manera recurrente se deberían de analizar un mayor número de grupos bacterianos, por ejemplo y aprovechando la secuenciación de tipo *single-cell* o metagenómica con alta cobertura (Bendall *et al*., 2016), intentar probar este concepto con los miembros de la biosfera rara (Sogin *et al.,* 2006), quienes probablemente tienen tamaños efectivos bajos y quizá entonces la deriva genética explique la divergencia de los grupos. En dado caso, la aplicación de este concepto, sin tomar en cuenta este tipo de organismos u otros como los simbiontes o patógenos estrictos, nos llevaría a una mala interpretación de la diversidad y a un concepto no incluyente. En segundo, en este trabajo observamos que el concepto adaptativo falla en dividir grupos que han divergido recientemente. Sin embargo, estos grupos que han divergido recientemente podrían ser considerados ecotipos.

Los análisis de recombinación que se hicieron con los aislados de Vibrionaceae de Cuatro Ciénegas y las cepas de referencia, mostraron que la mayor parte de los eventos de recombinación se dan entre ambos grupos y estos a su vez se consideran antiguos, ya que ocurren en las ramas internas de la filogenia. Sin embargo, los clados que observamos dentro de Cuatro Ciénegas están altamente diferenciados. Esto nos sugiere que a pesar del aislamiento, existe una señal evolutiva que marca estos grupos y a pesar de que se diferencian, no llega la diferenciación a ser tal que se puedan considerar especies distintas.

Hasta el momento sigue sin existir un concepto de especie que sea incluyente con la diversidad de organismos que existen, sin embargo, nuestros análisis muestran que hay especies bacterianas que son muy antiguas y que comparten una historia de recombinación, sin embargo estos linajes pueden formar ecotipos, como sería el caso de los linajes de Vibrionaceae dentro de Cuatro Ciénegas. Estos ecotipos pueden unirse en un grupo cohesivo y seguir formando parte del complejo que es la especie

(Cui *et al.,* 2015) o pueden seguir el proceso de divergencia hasta convertirse en unidades evolutivas independientes.

De la diversidad de especies a la diversidad de genes.

Los estudios de diversidad microbiana donde se observa estructuración asociada al ambiente o al espacio geográfico han inspirado trabajos de genómica de poblaciones, buscando las bases genéticas de la diversificación (Cadillo-Quiroz *et al.,* 2012; Shapiro *et al.,* 2012). En Cuatro Ciénegas se ha estudiado la diversidad microbiana tanto a nivel de comunidades (Escalante *et al.,* 2008; Espinosa-Asuar *et al.,* 2015; Pajares *et al.,* 2016) como poblacional (Rebollar *et al.,* 2012; Avitia *et al.,* 2014). En el caso de los estudios poblacionales, no se encontró estructura poblacional asociada al ambiente o al espacio geográfico en bacterias del genero *Bacillus*, sin embargo, si se encontró un patrón de estructura en el género *Exigobacterum*, el cual está asociado al ambiente de donde se obtuvieron los aislados, agua o sedimento. Siguiendo estos antecedentes, buscamos si existían patrones de estructuración en los aislados de la familia Vibrionaceae y encontramos que los linajes de este grupo se estructuran igual que los *Exigobacterum*, según el ambiente, agua o sedimento.

Utilizando la información genómica, hicimos un análisis de recombinación para determinar si existía un patrón de recombinación asociado al ambiente. Por un lado esperábamos que todos los aislados que fueran de agua recombinaran entre ellos con mayor frecuencia que con los aislados de los sedimentos, lo que de manera gráfica se vería representado en dos grupos diferenciados. Sin embargo lo que encontramos es que el patrón de recombinación nos muestra grupos asociados a los clados, es decir seis grupos que recombinan frecuentemente dentro de cada uno. Lo que sugiere que estos clados son antiguos y no existe un patrón de recombinación que este dado por el ambiente.

Descartando el ambiente, empezamos a analizar la variación genética, desde el punto de vista del contenido genético, y observamos que existe una alta variabilidad de genes a nivel de familia, pero baja dentro de cada clado, lo que nos sugiere especialización dentro de cada grupo. Esta variabilidad a nivel de genes es la descripción del pan-genoma donde siguiendo la propuesta de Tettelin y colaboradores (2008) llegamos a clasificar el pan-genoma de la familia Vibrionaceae como abierto, lo que indica que hay una gran diversidad genética dentro del grupo.

Ahora ¿Cómo podemos explicar este patrón de diversidad genética a nivel de familia y el hermetismo a nivel de clado? Para tratar de entender el comportamiento del pan-genoma se midió la tasa de recombinación a través del parámetro de r/m (Vos & Didelot, 2009) y el tamaño efectivo de las poblaciones utilizando métodos ABC (Excoffier & Foll, 2011; Excoffier *et al.,* 2013). La literatura sugiere que los tamaños efectivos grandes están correlacionados con una mayor cantidad de genes accesorios (Andreani *et al.,* 2017). Nosotros encontramos que 3 de nuestros clados tienen tamaños

efectivos grandes, pero pan-genomas cerrados. A su vez, contrastamos nuestros resultados con el patrón de recombinación. El intercambio genético puede a su vez afectar el tamaño del pan-genoma aumentándolo cuando se trata de eventos de transferencia horizontal o bien homogenizando la población, llevando a un patrón de pan-genoma cerrado. La mayoría de nuestros clados tienen valores de recombinación que se pueden considerar medios si se compara con, por ejemplo, con *Pelagibacter ubique* que tiene un valor de 63.1 o *Vibrio parahaemolyticus* que tiene un valor de 39.8 en los océanos (Vos & Didelot, 2009). En Cuatro Ciénegas se ha propuesto que la tasa de recombinación es baja, debido a la existencia de una presión de selección en contra de la adquisición de material genético, pues implica un gasto energético y en un lugar con pocos recursos como Cuatro Ciénegas, se evitaría. Bajo este escenario sería necesario considerar la tasa de adquisición de genes nuevos, que sería baja en el modelo de Cuatro Ciénegas y la tasa de fijación que sería dependiente de la tasa de recombinación para esparcir el alelo y el coeficiente de selección del mismo. Por lo que los pan-genomas cerrados en Cuatro Ciénegas pueden ser el resultado de escasos eventos de recombinación.

Los procariontes tienen una gran diversidad genética sin embargo, la forma en la que las fuerzas evolutivas explican este patrón aun no es clara. Hasta ahora parece que la selección promueve la diversidad de contenido genético por medio de la THG. Nosotros proponemos que las tasas de recombinación regulan el tamaño del pan-genoma.

Patógenos en Cuatro Ciénegas

Los miembros de la familia Vibrionaceae tienen un amplio rango de distribución y estilos de vida, muchos de ellos son patógenos de humanos u otros organismos acuáticos (Farmer *et al.,* 1986; Wachsmuth *et al.,* 1994). Dadas estas características y debido a que los clados 2 y 6 están relacionados con grupos de *Vibrio* patógenos decidimos buscar los elementos que se asocian a la patogénesis en los aislados de Cuatro Ciénegas. Nuestro sitio de estudio se caracteriza por poseer bajos niveles de nutrientes, y una alta diversidad de microorganismos con una alta afinidad con microorganismos marinos (Souza *et al.,* 2012), los modelos indican que las pozas de Cuatro Ciénegas son reminiscencias del mar antiguo (Moreno-Letelier *et al.,* 2012). Por lo que, partimos de la idea de que los elementos asociados a la patogénesis no debían de estar presentes en organismos que han estado aislados, sin embargo, encontramos que muchos de estos elementos si están presentes, lo que nos hizo replantearnos el papel de estos elementos en la biología del organismo.

Los elementos que encontramos más representados en los genomas de Cuatro Ciénegas fueron los sistemas de secesión, proteínas asociadas a la formación del flagelo y a la quimiotaxis. Los sistemas de secreción son estructuras que le permiten a los patógenos transferir toxinas de una célula a otra, en algunos casos transmiten ADN como es el caso de los sistemas de secreción tipo IV (Koster *et al.,*

2000; Kapitein and Mogk, 2013; Costa *et al.,* 2015). Nuestros datos muestran que los genes que conforman estos sistemas de secreción han pasado por diversos eventos de recombinación y probablemente de duplicaciones, que no solo forman parte de la familia Vibrionaceae si no que se comparten con otras Gammaproteobacteria. Por lo que se trata de familias génicas que son muy antiguas y que han venido evolucionando dentro de este grupo. La propuesta es que su función principal va más allá de la patogénesis, que son vías de contacto entre células que les permiten intercambiar no solo toxinas si no también señales u otros metabolitos que les permiten a los organismos coexistir dentro de una comunidad.

Por otro lado, se hizo la búsqueda de eleméntenos móviles como las islas de patogenicidad, integrones, resistencias a antibióticos, fagos y CRISPRs. En los genomas de Cuatro Ciénegas no encontramos islas de patogenicidad aunque si encontramos islas genómicas, por lo que, probablemente las cepas patógenas han pasado por re-arreglos genómicos que les han permitido mantener los genes que se emplean durante la infección juntos, quizá para regular con mayor facilidad su expresión. A su vez, encontramos una relación inversa en cuanto a la proporción de fagos y CRISPRs en los genomas de Cuatro Ciénegas, esto sugiere que las secuencias CRISPRs podrían estar funcionando como un sistema inmune que evita las infecciones por fagos.

Por lo que nuestros resultados sugieren que los mecanismos normalmente asociados a la patogénesis, pueden ser antiguos y tener un papel ecológico importante como puede ser comunicación o defensa. Este estudio abre la puerta a nuevos caminos de cómo entender estos elementos genéticos fuera del contexto clínico y que puede ser de gran utilidad para interpretar resultados provenientes del diagnóstico molecular.

Finalmente y de manera general, podemos concluir que la familia Vibrionaceae dentro de Cuatro Ciénegas lleva una historia paralela respecto a los miembros de la familia que habitan en los océanos o en ambientes impactados por el hombre. A pesar de esto, encontramos linajes que se encuentran en un proceso incipiente de diferenciación, por lo cual no pueden ser consideradas especies distintas, aunque en otros casos, esta divergencia es tal que podemos proponer que existen linajes endémicos. Este aislamiento en conjunto con las bajas tasas de eventos de THG ha impactado el pan-genoma de estos linajes. Como se reportó en este trabajo y en congruencia con trabajos previos, las comunidades en Cuatro Ciénegas son muy diversas; la gran diversidad y especialización que encontramos en la familia Vibrionacea puede estar ligada a esta especiación a través de la adaptación dentro de las pozas.

Literatura citada

Achman, M., and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. Nature Reviews Microbiology 6, 431-440.

Al-saari Nurhidayu, Gao, F., A.K.M. Rohul, A., Sato, K., Sato, K., Mino, S., Sawabe, T. (2015). Advanced Microbial Taxonomy Combined with Genome-Based-Approaches Reveals that Vibrio astriarenae sp. nov., an Agarolytic Marine Bacterium, Forms a New Clade in Vibrionaceae. PLoS ONE, 10(8), e0136279. http://doi.org/10.1371/journal.pone.0136279

Andreani, N.A., Hesse, E., Vos M. (2017). Prokaryote genome fluidity is dependent on effective population size. ISME J. 11(7):1719-1721

Angiuoli, S. V., & Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics, 27(3), 334–342. http://doi.org/10.1093/bioinformatics/btq665

Avitia, M., Escalante, A. E., Rebollar, E. A., Moreno-Letelier, A., Eguiarte, L. E., & Souza, V. (2014). Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. PeerJ, 2, e696. http://doi.org/10.7717/peerj.696

Bendall, M. L., Stevens, S. L., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., … Malmstrom, R. R. (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. The ISME Journal, 10(7), 1589–1601. http://doi.org/10.1038/ismej.2015.241

Bergthorsson, U., Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of Escherichia coli. Mol Biol Evol. 15(1):6-16.

Bonilla-Rosso, G., Peimbert, M., Alcaraz, L. D., Hernández, I., Eguiarte, L. E., Olmedo-Alvarez, G., & Souza, V. 2012. Comparative Metagenomics of Two Microbial Mats at Cuatro Ciénegas Basin II: Community Structure and Composition in Oligotrophic Environments. Astrobiology, 12(7), 659–673. http://doi.org/10.1089/ast.2011.0724

Boyd, E.F., Carpenter, M.R., Chowdhury, N., Cohen, A.L., Haines-Menges, B.L., Kalburge, S.S., Kingston, J.J., Lubin, J.B., Ongagna-Yhombi, S.Y., Whitaker, W.B. (2015). Post-Genomic Analysis of Members of the Family Vibrionaceae. Microbiol Spectr. 3(5). doi:10.1128/microbiolspec.VE-0009-2014.

Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., Whitaker, R. J. (2012). Patterns of Gene Flow Define Species of Thermophilic Archaea. PLoS Biology, 10(2), e1001265. http://doi.org/10.1371/journal.pbio.1001265

Chun, J., Huq, A., & Colwell, R. R. (1999). Analysis of 16S-23S rRNA Intergenic Spacer Regions of Vibrio cholerae and Vibrio mimicus. Applied and Environmental Microbiology, 65(5), 2202–2208.

Cohan, F.M. (2001). Bacterial species and speciation. Syst Biol. 50(4):513-524.

Cohan, F.M., Perry, E.B. (2007). A systematics for discovering the fundamental units of bacterial diversity. Curr Biol. 17(10):R373-86.

Cooper, V. S., Vohr, S. H., Wrocklage, S. C., & Hatcher, P. J. (2010). Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. PLoS Computational Biology, 6(4), e1000732. http://doi.org/10.1371/journal.pcbi.1000732

Cordero, O.X., Polz, M.F. (2014). Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol. 12(4):263-73. doi: 10.1038/nrmicro3218.

Costa, T.R., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M. and Waksman, G. 2015. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol. 13(3):343-359. doi:10.1038/nrmicro3456

Cui, Y., Yang, X., Didelot, X., Guo, C., Li, D., Yan, Y., Zhang, Y., Yuan, Y., Yang, H., Wang, J., Wang, J., Song, Y., Zhou, D., Falush, D., Yang, R. (2015). Epidemic Clones, Oceanic Gene Pools, and Eco-LD in the Free Living Marine Pathogen Vibrio parahaemolyticus. Mol Biol Evol. 32(6):1396-410. doi: 10.1093/molbev/msv009.

Davis, B. R., Fanning, G. R., Madden, J. M., Steigerwalt, A. G., Bradford, H. B., Smith, H. L., & Brenner, D. J. (1981). Characterization of biochemically atypical Vibrio cholerae strains and designation of a new pathogenic species, Vibrio mimicus. Journal of Clinical Microbiology, 14(6), 631–639.

Didelot, X., & Falush, D. (2007). Inference of Bacterial Microevolution Using Multilocus Sequence Data. Genetics, 175(3), 1251–1266. http://doi.org/10.1534/genetics.106.063305

Dieckmann, U & Doebeli, M. (1999). On the origin of species by sympatric speciation. Nature. 400(6742):354-357.

Dikow, R. B., & Smith, W. L. (2013). Genome-level homology and phylogeny of Vibrionaceae (Gammaproteobacteria: Vibrionales) with three new complete genome sequences. BMC Microbiology, 13, 80. http://doi.org/10.1186/1471-2180-13-80

Doolittle, W.F. (1999). Phylogenetic classification and the universal tree. Science. 284(5423):2124-9.

Elser, J.J., Schampel, J.H., Garcia-Pichel, F., Wade, B., Souza, V., Eguiarte, L.E., Escalante, A. and Farmer., J. 2005. Effects of PO4 enrichment and grazing snails on microbial communities in an ecosystem with living stromatolites. Freshwater Biol. 50: 1808-1825

Escalante, A.E., Eguiarte, L.E., Espinosa-Asuar, L., Forney, L J., Noguez, A.M. and Souza Saldivar, V. 2008. Diversity of aquatic prokaryotic communities in the Cuatro Cienegas basin. FEMS Microbiology Ecology, 65(1):50-60. doi:10.1111/j.1574-6941.2008.00496.x

Espinosa-Asuar, L. Escalante, A. E. Gasca-Pineda, J. Blaz, J. Pena, L. Eguiarte, L.E. Souza, V. 2015. Aquatic bacterial assemblage structure in Pozas Azules, Cuatro Cienegas Basin, Mexico: Deterministic vs. stochastic processes. International microbiology : the official journal of the Spanish Society for Microbiology 18 (2): 105-15. doi:10.2436/20.1501.01.240.

Excoffier, L. and M. Foll. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27: 1332-1334.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., and M. Foll (2013) Robust demographic inference from genomic and SNP data. PLOS Genetics 9(10):e1003905.

Friedman, J., Alm, E. J., & Shapiro, B. J. (2013). Sympatric Speciation: When Is It Possible in Bacteria? PLoS ONE, 8(1), e53539. http://doi.org/10.1371/journal.pone.0053539

Georgiades, K., & Raoult, D. (2010). Defining Pathogenic Bacterial Species in the Genomic Era. Frontiers in Microbiology, 1, 151. http://doi.org/10.3389/fmicb.2010.00151

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., … Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. Genome Research, 15(10), 1451–1455. http://doi.org/10.1101/gr.4086505

Hasan, N. A., Choi, S. Y., Eppinger, M., Clark, P. W., Chen, A., Alam, M., … Colwell, R. R. (2012). Genomic diversity of 2010 Haitian cholera outbreak strains. Proceedings of the National Academy of Sciences of the United States of America, 109(29), E2010–E2017. http://doi.org/10.1073/pnas.1207359109

Hasan, N. A., Grim, C. J., Haley, B. J., Chun, J., Alam, M., Taviani, E., … Colwell, R. R. (2010). Comparative genomics of clinical and environmental Vibrio mimicus. Proceedings of the National Academy of Sciences of the United States of America, 107(49), 21134–21139. http://doi.org/10.1073/pnas.1013825107

Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith. H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C., Fraser, C.M. (2002). DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. Nature.406(6795):477-483.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F. (2016). A new view of the tree of life. Nat Microbiol. 1:16048. doi: 10.1038/nmicrobiol.2016.48.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., Polz, M.F. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 320(5879):1081-5.

Kahlke, T., & Thorvaldsen, S. (2012). Molecular Characterization of Cold Adaptation of Membrane Proteins in the Vibrionaceae Core-Genome. PLoS ONE, 7(12), e51761. http://doi.org/10.1371/journal.pone.0051761

Kapitein, N.and Mogk, A. 2013. Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. Curr Opin Microbiol. 16(1):52-8. doi: 10.1016/j.mib.2012.11.009.

Koonin, E.V., Makarova, K.S., Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol. 55:709-42.

Koster, M., Bitter, W. and Tommassen, J. 2000. Protein secretion mechanisms in Gram-negative bacteria. Int J Med Microbiol. 290(4-5):325-331.

Lawrence, J.G., Retchless, A.C. (2009). The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol. 532:29-53. doi: 10.1007/978-1-60327-853-9_3.

Mayr, E. (1942). Systematics and the Origin of Species. New York, NY: Columbia University Press

Mayr, E. (1963). Animal species and evolution. Harvard University Press.

Mazel, D. (2006). Integrons: agents of bacterial evolution. Nat Rev Microbiol. 4(8):608-20.

McInerney, J.O., McNally, A., O'Connell, M.J. (2017). Why prokaryotes have pangenomes. Nat Microbiol. 2:17040. doi: 10.1038/nmicrobiol.2017.40.

Mira, A., Martín-Cuadrado, A.B., D'Auria, G., Rodríguez-Valera, F. (2010). The bacterial pan-genome:a new paradigm in microbiology. Int Microbiol. 13(2):45-57.

Moreno-Letelier, A., Olmedo-Alvarez, G., Eguiarte, L. E., & Souza, V. (2012). Divergence and Phylogeny of Firmicutes from the Cuatro Ciénegas Basin, Mexico: A Window to an Ancient Ocean. Astrobiology, 12(7), 674–684. http://doi.org/10.1089/ast.2011.0685

Ogilvie, L.A., & Hirsch, P.R. (2012). Microbial Ecological Theory. Caister Academic Press.

Okada, K., Iida, T., Kita-Tsukamoto, K., & Honda, T. (2005). Vibrios Commonly Possess Two Chromosomes. Journal of Bacteriology, 187(2), 752–757. http://doi.org/10.1128/JB.187.2.752-757.2005

Pajares, S., Escalante, A.E., Noguez, A. M., Garcia-Oliva, F., Martinez-Piedragil, C., Cram, S.S., Eguiarte, L.E. and Souza, V. 2016. Spatial heterogeneity of physicochemical properties explains differences in microbial composition in arid soils from Cuatro Cienegas, Mexico. PeerJ, 4, e2459. http://doi.org/10.7717/peerj.2459

Peimbert, M., Alcaraz, L.D., Bonilla-Rosso, G., Olmedo-Alvarez, G., García-Oliva, F., Segovia, L., Eguiarte, L.E. and Souza, V. 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(7):648–658. doi.org/10.1089/ast.2011.0694

Ragan, M.A. (2001). On surrogate methods for detecting lateral gene transfer. FEMS Microbiol Lett. 201(2):187-191.

Rebollar, E.A., Avitia, M., Eguiarte, L.E., González-González, A., Mora, L., Bonilla-Rosso, G., Souza, V. (2012). Water-sediment niche differentiation in ancient marine lineages of Exiguobacterium endemic to the Cuatro Cienegas Basin. Environ Microbiol. 14(9):2323-33. doi: 10.1111/j.1462-2920.2012.02784.x.

Ruby, E. G., & Lee, K.-H. (1998). The Vibrio fischeri-Euprymna scolopes Light Organ Association: Current Ecological Paradigms. Applied and Environmental Microbiology, 64(3), 805–812.

Schiewe, M.H, & Crosa, J.H. (1981). Molecular characterization of Vibrio anguillarum biotype 2. Canadian Journal of Microbiology. 27(10): 1011-1018, https://doi.org/10.1139/m81-158

Selander, R.K., Levin, B.R. (1980). Genetic diversity and structure in Escherichia coli populations. Science. 210(4469):545-547.

Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabó, G., Polz, M.F., Alm, E. J. (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science (New York, N.y.), 336(6077), 48–51. http://doi.org/10.1126/science.1218198

Shapiro, B.J. (2016). How clonal are bacteria over time?. Curr Opin Microbiol. 31:116-23. doi: 10.1016/j.mib.2016.03.013.

Smith, J. M., Smith, N. H., O'Rourke, M., & Spratt, B. G. (1993). How clonal are bacteria? Proceedings of the National Academy of Sciences of the United States of America, 90(10), 4384–4388.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J., Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." Proceedings of the National Academy of Sciences of the United States of America, 103(32), 12115–12120. http://doi.org/10.1073/pnas.0605127103

Souza, V., Eguiarte, L. E., Travisano, M., Elser, J. J., Rooks, C., & Siefert, J. L. (2012). Travel, Sex, and Food: What's Speciation Got to Do with It? Astrobiology, 12(7), 634–640. http://doi.org/10.1089/ast.2011.0768

Takemura, A. F., Chien, D. M., & Polz, M. F. (2014). Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. Frontiers in Microbiology, 5, 38. http://doi.org/10.3389/fmicb.2014.00038

Thompson, F. L., Iida, T., & Swings, J. (2004). Biodiversity of Vibrios. Microbiology and Molecular Biology Reviews, 68(3), 403–431. http://doi.org/10.1128/MMBR.68.3.403-431.2004

Thompson, F.L., Austin, B., Swings, J. (2006). The Biology of Vibrios. American Society for Microbiology

Urdaci, M.C., Stal,L.J., & Marchand, M.(1988).Occurrence of nitrogen fixation among Vibrio spp. Arch. Microbiol. 150, 224–229. doi: 10.1007/BF00407784

Vos, M. (2011). A species concept for bacteria based on adaptive divergence. Trends Microbiol. 19(1):1-7. doi: 10.1016/j.tim.2010.10.003.

Vos, M., Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. ISME J.3(2):199-208. doi: 10.1038/ismej.2008.93.

Wang, D., Wang, H., Zhou, Y., Zhang, Q., Zhang, F., Du, P., … Kan, B. (2011). Genome Sequencing Reveals Unique Mutations in Characteristic Metabolic Pathways and the Transfer of Virulence Genes between V. mimicus and V. cholerae. PLoS ONE, 6(6), e21299. http://doi.org/10.1371/journal.pone.0021299

Welch, R. A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., … Blattner, F. R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 99(26), 17020–17024. http://doi.org/10.1073/pnas.252529799

Woodford, N., Johnson, A.P. (2004). Genomics, Proteomics and Clinical Bacteriology. Human Press.

Apéndice 1

Métodos empleados para el análisis del concepto de especie adaptativo

Apéndice 1. Métodos utilizados en el análisis del concepto adaptativo de especie en Vibrionnaceae

Selección de cepas.

Un subgrupo de aislados de *Vibrio sp*. Fueron obtenidos de la base de datos del NCBI (Tabla 1). Para cada una de las especies analizadas se seleccionaron un total de cinco genomas, esto con el objetivo de tener una muestra representativa y del mismo tamaño para cada una de los grupos.

Alineamiento.

Las secuencias obtenidas se alinearon con el programa Mugsy (Angiuoli & Salzberg, 2011). El alineamiento obtenido para cada una de las combinaciones de los grupos fue convertido a formato fasta mediante el *script* MAFtoFASTA de la paquetería Galaxy (Giardine *et al*., 2005).

Inferencia de MKT.

El análisis de inferencia de MKT se llevó a cabo utilizando la librería PopGenome del paquete estadístico R. El tamaño de las ventanas se eligió con base en el tamaño de un gen promedio 1000 pbs con saltos de 500 pbs ya que permiten una región de empalme.