



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE DOCTORADO EN CIENCIAS BIOMÉDICAS

GENÉTICA DE POBLACIONES Y EVOLUCIÓN  
MOLECULAR DE *Bacillus sp* EN CUATRO CIÉNEGAS  
COAHUILA, MÉXICO

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

MORENA AVITIA CAO ROMERO

TUTORA: DRA. VALERIA SOUZA SALDÍVAR

INSTITUTO DE ECOLOGÍA

MÉXICO. D.F ABRIL 2015



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Esta tesis se realizó bajo la asesoría de la Dra. Valeria Souza Saldívar, en el Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología de la Universidad Nacional Autónoma de México, con el apoyo financiero de los proyectos SEP-CONACyT 0237A1 y World Wildlife Found (WWF) - Alianza Carlos Slim ECOIE410.

El comité tutorial que asesoró el desarrollo de esta tesis estuvo formado por:

Dra. Valeria Souza Saldívar, Instituto de Ecología, UNAM.

Dr. Luis E. Eguiarte Fruns, Instituto de Ecología, UNAM.

Dr. Mario Soberón Chávez, Instituto de Biotecnología, UNAM.

El jurado de examen doctoral estuvo constituido por:

|                              |            |
|------------------------------|------------|
| Dra. Gloria Soberón Chávez   | Presidente |
| Dr. David Romero Camarena    | Vocal      |
| Dr. Georges Dreyfus Cortés   | Vocal      |
| Dra. Gabriela Olmedo Álvarez | Vocal      |
| Dra Valeria Souza Saldívar   | Secretaria |

## ÍNDICE

|                               |            |
|-------------------------------|------------|
| <b>RESUMEN/ABSTRACT .....</b> | <b>3-4</b> |
|-------------------------------|------------|

|  |          |
|--|----------|
| <b>Capítulo 1. INTRODUCCIÓN GENERAL.....</b> | <b>5</b> |
|--|----------|

|   |   |
|---|---|
| 1.1 Entendiendo las causas de la diversidad en procariontes: genética de poblaciones y evolución molecular en bacterias ..... | 6 |
|---|---|

|  |   |
|--|---|
| 1.1.1 Aspectos metodológicos del estudio de la diversidad genética en poblaciones de bacterias ..... | 8 |
|--|---|

|  |    |
|--|----|
| 1.1.2 Consideraciones teóricas del estudio de poblaciones en bacterias ..... | 14 |
|--|----|

|   |    |
|---|----|
| 1.1.2.1 El problema del concepto de especie en bacterias..... | 14 |
|---|----|

|  |    |
|--|----|
| 1.1.2.2 El papel de la recombinación en las poblaciones de bacterias ..... | 23 |
|--|----|

|  |    |
|--|----|
| 1.1.2.3 La importancia de la geografía y el ambiente ..... | 29 |
|--|----|

|   |           |
|---|-----------|
| <b>Capítulo 2. ANTECEDENTES Y JUSTIFICACIÓN DEL PROYECTO.....</b> | <b>38</b> |
|---|-----------|

|  |    |
|--|----|
| 2.1 Sitio de estudio: Cuatro Ciénegas Coahuila ..... | 38 |
|--|----|

|   |    |
|---|----|
| 2.2 Grupo de estudio: el género <i>Bacillus</i> ..... | 42 |
|---|----|

|  |    |
|--|----|
| 2.2.1 El género <i>Bacillus</i> en Cuatro Ciénegas ..... | 48 |
|--|----|

|   |    |
|---|----|
| 2.3 Justificación y presentación del proyecto ..... | 50 |
|---|----|



|  |            |
|--|------------|
| <b>Capítulo 3. ARTÍCULO 1: “<i>Bacillus cuatrociénegasensis</i> sp. nov., una nueva especie moderadamente halotolerante aislada del valle Cuatro Ciénegas, México”</b> ..... | <b>53</b>  |
| 3.1 Resumen.....   | 53         |
| 3.2 Artículo en inglés.....  | 55         |
| <br>   |            |
| <b>Capítulo 4. Diversidad genética, distribución y recombinación en las poblaciones de <i>Bacillus</i> en CCC</b> .....  | <b>80</b>  |
| 4.1 Diversidad genética y recombinación en los filogrupos B1 y B2.....   | 87         |
| <br>   |            |
| <b>Capítulo 5. ARTÍCULO 2: “Expansiones poblacionales compartidas entre linajes bacterianos que coexisten son reveladas por la evidencia genética”</b> .....                 | <b>100</b> |
| 5.1 Resumen.....   | 100        |
| 5.2 Artículo en inglés.....  | 101        |
| <br>   |            |
| <b>Capítulo 6 CONCLUSIONES Y PERSPECTIVAS</b> .....  | <b>150</b> |

## RESUMEN

En este trabajo se estudió la diversidad genética y los patrones de evolución molecular de poblaciones de bacterias del género *Bacillus* aisladas del valle de Cuatro Ciénegas, Coahuila, México. Durante tres años consecutivos se aislaron bacterias de muestras de agua y sedimento provenientes de los principales sistemas acuáticos dentro del valle. Para analizar la variación genética de los aislados se utilizaron seis marcadores moleculares incluyendo al gen 16S rRNA y cinco genes constitutivos. Los análisis de la secuencia del gen 16S rRNA permitieron identificar a los aislados como una nueva especie dentro del género *Bacillus*, cuya identidad taxonómica fue confirmada con una aproximación polifásica. A partir de la filogenia construida con la secuencia concatenada de los cinco genes constitutivos se pudieron identificar dos filogrupos principales en los que se agrupó la mayor parte de los aislados. Dichos filogrupos no se encuentran asociados a un sistema acuático particular dentro del valle ni al ambiente específico del cual fueron aislados (agua o sedimento). Los análisis genéticos a nivel de filogrupo sugieren que existe cierto grado de recombinación entre los miembros de cada filogrupo pero no entre ellos. Finalmente, el análisis más fino de la estructura poblacional al interior de cada filogrupo y su comparación con otras poblaciones de bacterias de Cuatro Ciénegas permitió descubrir que los eventos históricos han tenido un papel importante en moldear la diversidad genética tanto de las poblaciones de *Bacillus* como en las de otros géneros de bacterias que ahí habitan.

## **ABSTRACT**

This work studied the genetic diversity and molecular evolution patterns of bacterial populations of the genus *Bacillus* isolated from the Cuatro Ciénegas valley in Coahuila, México. Bacteria were isolated from water and sediment samples collected during three consecutive years on the main aquatic systems within the valley. Six molecular markers, including 16S rRNA gene and five housekeeping genes, were included in order to analyze the genetic diversity of the studied populations. The genetic analyses based on the 16S rRNA sequence identified the bacterial isolates of this study as a new species within the genus *Bacillus*, and this result was confirmed by a polyphasic approach. The phylogeny constructed using the concatenated sequence of the five housekeeping genes placed most of the isolates in two main phylogenetic groups. These two phylogenetic groups were not associated with a particular aquatic system within the valley or to the specific isolation environment (water or sediment). Genetic analyses within each of the phylogenetic groups showed that there is a degree of recombination within each phylogroup but not between them. Finally, a more detailed analysis of population structure within each phylogenetic group and its comparison with other bacterial populations from Cuatro Ciénegas showed that historical events have had an important role in shaping the genetic diversity of the studied *Bacillus* populations.

## Capítulo 1. INTRODUCCIÓN GENERAL

Los procariontes son los organismos más abundantes y diversos del planeta. Se ha calculado que en toda la Tierra existen entre  $4$  y  $6 \cdot 10^{30}$  células de este tipo de organismos (Whitman, Coleman & Wiebe, 1998). Debido a su abundancia, los procariontes representan la mayor reserva de nutrientes del planeta, llegando a estar incluso por encima de la concentración de nutrientes presentes en las plantas (Whitman, Coleman & Wiebe, 1998). La mayoría se distribuyen en los océanos, en el suelo y en la superficie terrestre. Sin embargo, además de estar presentes en todos los hábitats colonizados por los eucariontes, también habitan muchos otros sitios en los cuales sólo los procariontes extremófilos han podido desarrollarse (Schlegel & Jannasch, 2013). La diversidad de los procariontes se ve reflejada en la gran cantidad de especies y linajes que los representan. Se ha estimado, por ejemplo, que en el océano existen alrededor de 160 especies por mililitro mientras que en el suelo este número asciende hasta 3,800-6,400 especies por gramo (Curtis, Sloan & Scannell, 2002). Por otro lado, los procariontes se caracterizan también por tener una gran diversidad a nivel metabólico (Zinder & Dworkin, 2006). Lo anterior, sumado a su abundancia a nivel global, hace de los procariontes actores centrales en la regulación de todos los ciclos biogeoquímicos en el planeta (Kinzig, Pacala & Tilman, 2002). Finalmente, los procariontes también son organismos muy diversos a nivel genético. Sus tiempos generacionales cortos y sus tamaños poblacionales grandes son algunas

de las características que, se cree, promueven la diversidad a este nivel (Dykhuizen, 1998). Sin embargo, a pesar de la importancia y abundancia de los procariontes a nivel global, existen todavía preguntas importantes en torno a cuáles son los mecanismos que generan dicha diversidad.

En la siguiente sección de la introducción se presentan algunas de las ideas y trabajos más importantes que se han desarrollado en las últimas décadas para tratar de explicar la gran diversidad de los procariontes que existe en nuestro planeta, enfocándose en particular en el análisis de los mecanismos evolutivos que permiten generar diversidad a nivel poblacional y cómo éstos se relacionan con la divergencia entre linajes y los procesos de especiación.

### **1.1 Entendiendo las causas de la diversidad en procariontes: genética de poblaciones y evolución molecular en bacterias**

Para poder comprender la enorme diversidad de grupos procariontes presentes en nuestro planeta es necesario considerar los mecanismos que han generado dicha diversidad. Entendiendo que la diversidad es resultado de un proceso de divergencia evolutiva, es importante entender los procesos de generación de diversidad a nivel poblacional, para poder también entender los mecanismos de divergencia genética de los linajes y la especiación en procariontes.

La genética de poblaciones es una herramienta muy valiosa que justamente nos permite entender, por un lado, los procesos por medio de los cuales se genera la diversidad en las poblaciones y, por el otro, cómo las diferentes fuerzas evolutivas actúan moldeando dicha diversidad. Gracias a la genética de poblaciones hemos podido entender la dinámica evolutiva y los mecanismos de generación de diversidad de un número importante de plantas y animales. Sin embargo, el estudio de los microorganismos, en particular las bacterias, presenta una serie de complicaciones metodológicas (dificultad de aislamiento y cultivo) (Amann, Ludwig & Schleifer, 1995; Whitman, Coleman & Wiebe, 1998) y algunos problemas teóricos (dificultad para definir especies y poblaciones y conceptos como aislamiento geográfico y reproductivo) (Ward, 1998; Silva & Vinuesa, 2007; Fraser et al., 2009) que han limitado, en cierta medida, el avance de los trabajos de genética de poblaciones y evolución molecular en este grupo de organismos. A continuación se presenta una breve revisión de algunos de estos aspectos metodológicos y teóricos centrales para el estudio de las poblaciones de bacterias.

### **1.1.1 Aspectos metodológicos del estudio de la diversidad genética en poblaciones de bacterias**

Un problema metodológico importante que enfrenta el estudio de las poblaciones bacterianas es la dificultad que significa su aislamiento y cultivo. Se ha calculado que las especies cultivables representan menos del 1% del total de la diversidad

bacteriana (Amann, Ludwig & Schleifer, 1995; Oren, 2004). En cierta medida, este problema ha dejado de ser un impedimento grave gracias al desarrollo de técnicas moleculares independientes de cultivo. En años recientes, el desarrollo de dichas técnicas ha permitido estudiar e inventariar la gran diversidad de bacterias y arqueas del planeta (DeLong & Pace, 2001) así como detectar la presencia de estos grupos de organismos en sitios en los que hasta hace poco se creía imposible encontrar vida (Pikuta, Hoover & Tang, 2007; Mesbah & Wiegel, 2012; Seckbach, Oren & Stan-Lotter, 2013; Stevenson et al., 2014) En particular, la secuencia del gen 16S ribosomal (*16S rRNA*) ha sido utilizada ampliamente como el marcador universal para describir la diversidad procarionte de nuestro planeta (Woese, Kandler & Wheelis, 1990; Acinas et al., 2004; Konstantinidis, Ramette & Tiedje, 2006). Junto con la secuenciación del *16S rRNA*, el desarrollo de la metagenómica ha permitido el análisis detallado de comunidades microbianas completas a partir de la secuenciación del repertorio completo de genes contenido en una muestra ambiental particular (Allen & Banfield, 2005; Medini et al., 2008). Las técnicas moleculares independientes de cultivo no solamente han permitido hacer un inventario detallado de la diversidad microbiana a nivel de comunidades sino que también han permitido empezar a responder algunas de las preguntas centrales de la ecología microbiana. De este modo, nuestra comprensión el modo en que se ensamblan las comunidades bacterianas, cuáles son las funciones de los diferentes tipos de bacterias y arqueas en una comunidad y, finalmente, cuál es el papel de este tipo de organismos en los ecosistemas ha crecido

notablemente en las últimas décadas gracias al desarrollo de estas nuevas tecnologías (Forney, Zhou & Brown, 2004; Whitaker & Banfield, 2006; Dinsdale & Rohwer, 2008).

Sin embargo, el nivel de resolución necesario para describir y analizar la diversidad procarionte a nivel de poblaciones difícilmente se alcanza con técnicas metagenómicas independientes de cultivo (Medini et al., 2008). Para la genética de poblaciones es necesario contar con información genética lo más detallada posible de una población cuyos individuos se puedan identificar claramente. De esta manera, el aislamiento y obtención de cultivos axénicos, mediante técnicas clásicas de microbiología, parece seguir siendo un primer paso indispensable cuando se busca conocer la distribución de la diversidad dentro de una población.

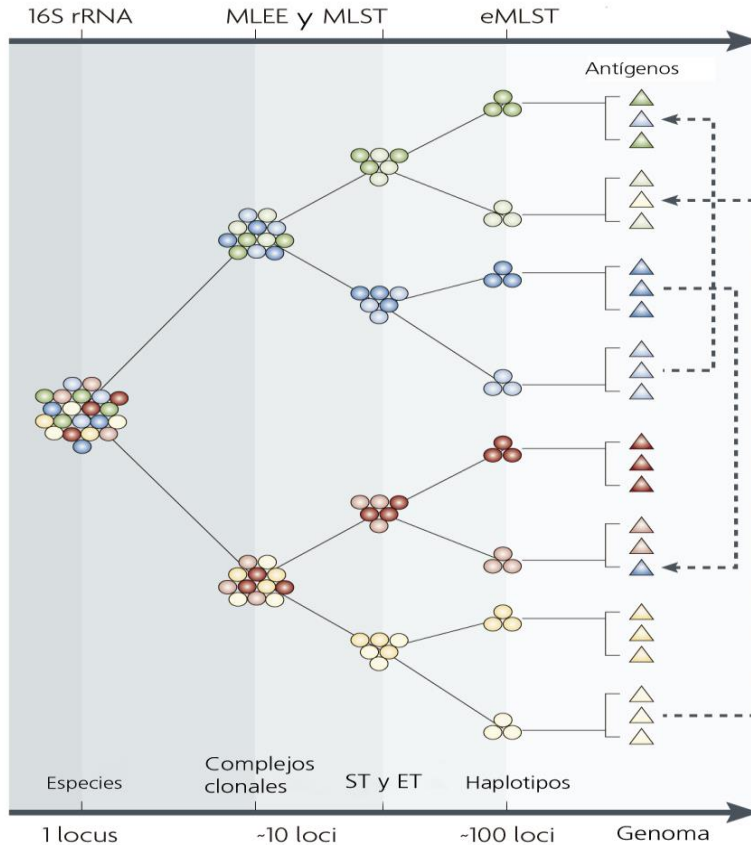
A partir de organismos cultivables, los primeros trabajos de genética de poblaciones en bacterias se hicieron por medio de electroforesis multilocus de enzimas (MLEE, por sus siglas en inglés, *multilocus enzyme electroforesis*). La técnica de MLEE permitió la exploración de la diversidad genética de poblaciones bacterianas basándose en los patrones de separación, por electroforesis, de enzimas metabólicas. Dicho método permitió trabajos pioneros como los de *Escherichia coli* (Milkman, 1973) y su posterior aplicación en varias bacterias cultivables patógenas (Selander et al., 1986) como *Neisseria meningitidis* (Caugant et al., 1987) y *Salmonella* (Selander, Li & Nelson, 1996) y no patógenas



como *Rhizobium leguminosarum* (Piñero, Martínez & Selander, 1988), *Bacillus subtilis* (Istock et al., 1992) y *Rhizobium etli* (Souza et al., 1992). Posteriormente, gracias al desarrollo de técnicas moleculares que permiten la secuenciación directa de genes de manera relativamente rápida y fácil, los trabajos basados en MLEE empezaron a ser sustituidos por análisis de secuencias, que aunque al principio utilizaron sólo pocos genes (Nelson, Whittam & Selander, 1991) rápidamente dieron lugar a los análisis multilocus. La tipificación multilocus (MLST, por sus siglas en inglés, *multilocus sequence typing*) fue originalmente propuesta como un método que permitiría la caracterización de diferentes tipos de bacterias y la comparación directa de resultados generados en laboratorios distintos (Maiden et al., 1998). La idea es utilizar las secuencias parciales de 7 genes *housekeeping* (genes esenciales o de mantenimiento) y las mismas herramientas de análisis para generar bases de datos comunes que pueden ser utilizadas por laboratorios en diversas regiones del mundo. La explosión de trabajos MLST ha permitido generar una base de datos grande (MLST Public Repository; <http://www.mlst.net>) que cuenta actualmente con unas 30 especies representadas y otras 15 en proceso.

Una de las limitaciones de los análisis de tipo MLST es que existen algunas especies de bacterias en las cuales los genes *housekeeping* presentan una muy baja diversidad nucleotídica e incluso existen otras (conocidas como especies monomórficas, como son *Salmonella typhi* o *Bacillus anthracis*), en las que los

genes *housekeeping* son tan uniformes que todos los aislados tienen la misma secuencia y parecen ser el mismo en términos genéticos. Además de esto, los análisis de tipo MLST presentan el problema de la baja cobertura del genoma completo lo cual permite conocer sólo una fracción muy pequeña del total de la diversidad genética de una especie. En este sentido, la secuenciación de cientos de loci en lugar de unos cuantos (conocida como *extended MLST* o eMLST) así como la secuenciación de genomas completos se utilizan cada vez con mayor frecuencia como herramientas fundamentales que permiten describir y analizar con mucha mayor resolución la diversidad genética que existe al interior de las poblaciones (ver Figura 1).



**Figura 1.** Representación esquemática de los diferentes niveles de resolución que se pueden alcanzar dependiendo del o los marcadores moleculares utilizados y su incidencia sobre la interpretación de la estructura poblacional en bacterias. El gen 16S ribosomal (16S rRNA) se utiliza como el marcador universal para diferenciar especies, sin embargo, su poder de resolución es muy limitado para detectar la diversidad intraespecífica. Los análisis multilocus de menos de 10 loci, ya sea a través de electroforesis de enzimas (MLEE) o de secuenciación de genes housekeeping (MLST), proporcionan resolución dentro de las especies y permite definir complejos clonales así como los tipos electroforéticos (ETs) y de secuencia (STs) al interior de dichos complejos. Al ampliar los esquemas multilocus e incluir información de cientos de genes (eMLST) es posible incrementar la resolución y llegar a detectar haplotipos únicos en las poblaciones. Sin embargo, pueden existir genes que se dejen fuera en los esquemas eMLST y que estén relacionados, por ejemplo, con aspectos muy particulares del estilo de vida del organismo analizado (por ejemplo proteínas antigénicas en bacterias patógenas). De este modo, sólo en análisis de genomas completos podrá detectar toda la variación genética no clonal que dan forma a la estructura poblacional de bacterias a escala más fina (modificada de Medini et al., 2008)

Los análisis a nivel genómico han permitido entender la estructura poblacional de varias bacterias patógenas (Retchless & Lawrence, 2007; Falush, 2009; Wilson, 2012) y algunas no patógenas (Makarova et al., 2006; Shapiro & Polz, 2014) a una escala mucho más fina. De este modo se ha podido resolver, por ejemplo, la estructura genética de poblaciones de especies monomórficas como *B. anthracis* (Read et al., 2002) o entender la asociación de ciertos genotipos con características ambientales muy específicas en poblaciones ambientales del género *Vibrio* (Shapiro et al., 2012). Es claro entonces que cuando se quiere resolver la estructura poblacional a un nivel muy fino lo deseable es contar con información de tipo eMLST o de genomas completos. Sin embargo también hay que considerar que el uso de esquemas de tipo MSLT puede seguir siendo importante y valioso si los tamaños de muestra de la población a analizar son muy grandes. Es válido obtener conclusiones sobre la estructura genética de poblaciones bacterianas a partir de la secuencia de algunos cuantos genes pero siempre teniendo en cuenta el nivel de resolución de los marcadores genéticos que se están utilizando y las preguntas particulares que se busca responder sobre dicha población.

Finalmente cabe mencionar aquí que una alternativa para estudio de las poblaciones bacterianas sin pasar necesariamente por el aislamiento y cultivo es el conjunto de técnicas conocidas como secuenciación de célula única (SCG, por sus siglas en inglés, *single cell genomics*). Dichas técnicas se han desarrollado de

manera importante en la última década (Lasken, 2012; Stepanauskas, 2012; Blainey, 2013) y han permitido, por ejemplo, generar genomas de referencia de bacterias que nunca se han cultivado, facilitando así la interpretación de datos metagenómicos de comunidades bacterianas complejas (Woyke et al., 2009; Swan et al., 2011). La obtención de genomas de organismos no cultivados ha permitido también resolver filogenias de algunos grupos muy divergentes de bacterias y arqueas (Ghai et al., 2011; Blainey, 2013). Finalmente, la gran promesa de las técnicas de SCG es que su aplicación a mayor escala permitirá obtener información detallada sobre la diversidad genética y genómica en poblaciones naturales de bacterias, lo cual será de gran ayuda para tener una idea mucho más precisa de los procesos de microevolución en bacterias.

En conjunto, las nuevas tecnologías moleculares han permitido importantes avances en las áreas de la evolución y ecología microbiana y, en los años por venir, seguramente serán muy útiles para seguir profundizando nuestro conocimiento en dichos temas. Sin embargo, siguen existiendo muchas preguntas sin resolver en el estudio de las poblaciones de bacterias y su evolución. Algunos aspectos teóricos relevantes en el estudio de las poblaciones bacterianas se discuten a continuación.

### **1.1.2 Consideraciones teóricas del estudio de poblaciones en bacterias**

El estudio de la genética de poblaciones y evolución molecular en bacterias es esencial para poder entender los mecanismos de generación de diversidad en las mismas. Gracias a la genética de poblaciones hemos podido entender la dinámica evolutiva y los mecanismos de generación de diversidad de un número importante de macroorganismos. Sin embargo las bacterias y arqueas presentan una serie de características que plantean la pregunta de si es posible o adecuado utilizar para su estudio las mismas herramientas de análisis y principios teóricos generados para macroorganismos. A continuación se presentan algunos de los aspectos teóricos que deben ser tomados en cuenta para entender mejor el modo en que los procariontes evolucionan.

#### **1.1.2.1 El problema del concepto de especie en bacterias**

El debate sobre el concepto de especie en bacterias sigue vigente y es central en la microbiología actual debido a que está directamente relacionado con una serie de aspectos de la biología de los procariontes que siguen sin poder explicarse del todo. Por un lado sigue sin quedar claro el modo en que las diferentes fuerzas evolutivas permiten la generación y extinción de los linajes genéticos en bacterias. Existen también lagunas importantes en el conocimiento de los mecanismos de diferenciación entre subpoblaciones con ancestría común así como sobre los

procesos de adaptación a nichos nuevos y ambientes cambiantes. A pesar de que se han hecho notables avances encaminados a esclarecer algunos de estos puntos (Gevers et al., 2005; Koepfel et al., 2008; Achtman & Wagner, 2008; Fraser et al., 2009; Kopac & Cohan, 2011; Shapiro & Polz, 2014), la sistemática bacteriana no ha logrado incorporar del todo los conocimientos generados en los campos de la ecología y evolución procarionte a su propio campo de estudio.

En el caso de los macroorganismos, a lo largo de la historia de la biología, se han desarrollado y aplicado distintos conceptos de especie para delimitar las unidades básicas de clasificación de la diversidad biológica. Dentro de todos ellos pueden reconocerse tres tipos principales: los que se basan en el aislamiento reproductivo, los que se basan en la similitud y los que se basan en la monofilia (Cerritos, 2007; Silva & Vinuesa, 2007; Hart, 2011). Dentro de los conceptos de especie que se basan en el aislamiento reproductivo, el concepto biológico de especie, que plantea que las especies son grupos de poblaciones entre las cuales existe intercambio genético actual o potencialmente además de que están aisladas reproductivamente de otras especies (Dobzhansky, 1937; Mayr, 1942). La mayoría de los conceptos de especie están basados en similitud, bajo la idea intuitiva de que un grupo de organismos pertenecen a una misma especie si comparten un número importante de características que pueden ser morfológicas, fisiológicas, ecológicas, conductuales, etc. Este tipo de reconocimiento de las especies se formalizó antes de la teoría evolutiva y tiene el grave problema de que no

reconoce a las especies como entidades que son resultado de un proceso histórico evolutivo (Mayden, 1997). Finalmente, el concepto basado en monofilia más reconocido y utilizado es el concepto filogenético de especie. Éste busca definir especies a partir del reconocimiento de los grupos más pequeños de organismos que se pueden diferenciar de otros organismos similares y dentro de los cuales se puede reconocer una relación ancestro-descendiente (Cerritos, 2007). Los caracteres tomados en cuenta para reconocer este tipo de especies pueden ser fenotípicos o genotípicos siempre y cuando se vea reflejada la relación de ancestría común.

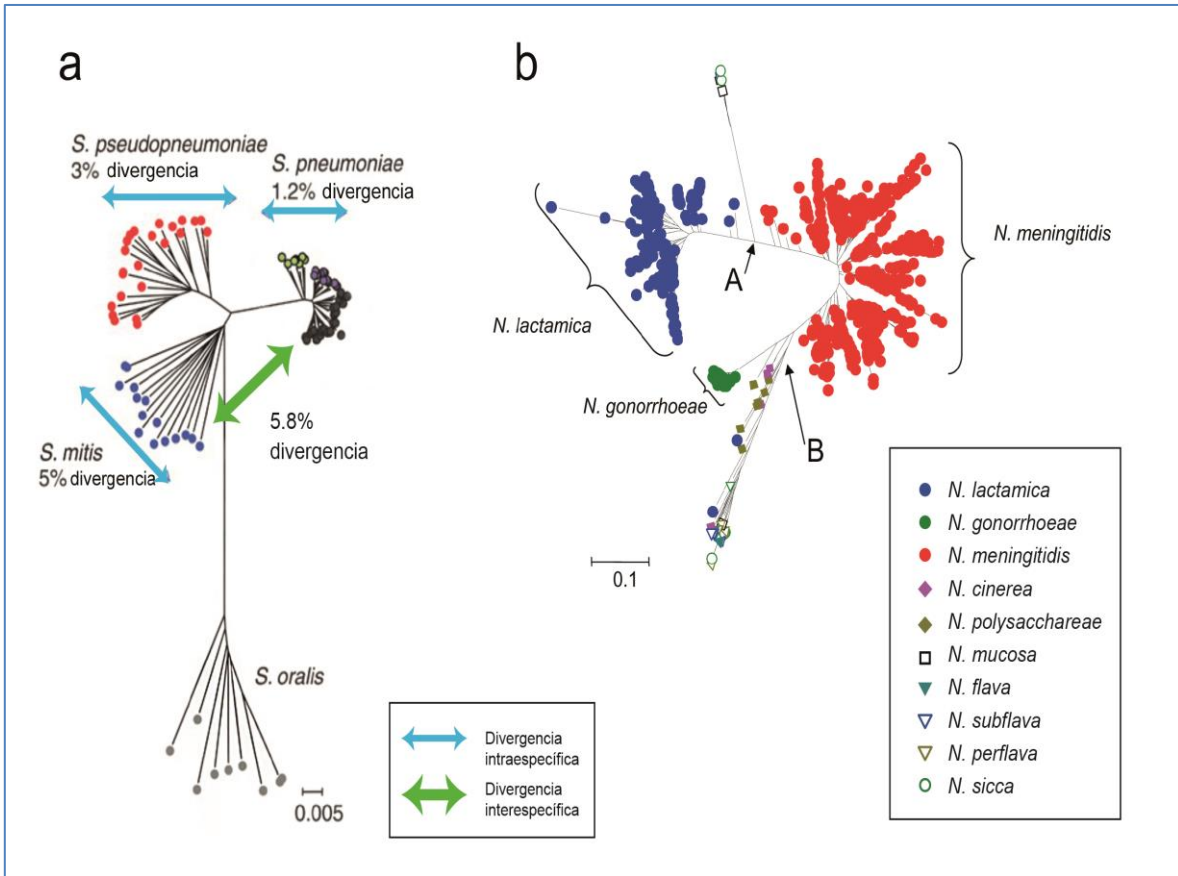
Cada uno de los diferentes tipos de conceptos de especie tiene sus problemas particulares aún en su aplicación a los macroorganismos, sin embargo, el problema de definir especies en microorganismos es aún mayor, debido a que la mayoría de ellos, en particular las bacterias y arqueas no cuentan con varias de las características con base en las cuales dichos conceptos fueron propuestos. El ejemplo más claro es la imposibilidad de aplicar el concepto biológico de especie a los organismos procariontes que tienen reproducción asexual. También resulta complicado reconocer unidades fenotípicamente distintas de otras basándose solamente en el limitado número de características morfológicas y fisiológicas de los procariontes que podrían ser informativas.



A pesar de las dificultades teóricas que implica la definición de especies en procariontes, los taxónomos de bacterias y arqueas han tenido que encontrar un conjunto de criterios prácticos que les permitan nombrar y describir especies. Actualmente la llamada aproximación polifásica para definir especies es la más aceptada en la sistemática bacteriana. Dicha aproximación busca llegar a un consenso sobre la identidad taxonómica de las especies a partir de incorporar diferentes tipos de información en la clasificación de las entidades biológicas (Vandamme et al., 1996). En primer lugar, se utiliza un conjunto de rasgos fenotípicos que incluyen características morfológicas, bioquímicas y fisiológicas y que pueden ser fácilmente reconocidas y descritas en un laboratorio. Se incluyen también características genotípicas y filogenéticas. En este caso, se da especial atención al porcentaje de hibridación DNA/DNA del genoma completo y a la secuencia de genes ribosomales de subunidad pequeña (16S rRNA). Para que dos aislados bacterianos se consideren pertenecientes a una misma especie, deben presentar un porcentaje de hibridación DNA-DNA de cuando menos un 70% (Stackebrandt & Goebel, 1994). De modo similar, se considera que dos aislados bacterianos con una identidad mayor o igual al 97% en la secuencia de sus genes 16S rRNA pertenecen a la misma especie (Stackebrandt et al., 2002). Sin embargo, el uso de este gen para distinguir especies bacterianas es controvertido debido a que, en muchos casos, dicho marcador molecular no tiene el poder de resolución necesario para distinguir entre dos especies cercanas (Fraser et al., 2009). En muchos casos los análisis de tipo MLST y los estudios

genómicos han revelado que existe gran diversidad genética dentro de los grupos definidos sólo por el 16s rRNA, por lo que es importante desarrollar explicaciones teóricas que den cuenta de dicha diversidad.

Es claro que se pueden reconocer grupos discretos de organismos a partir de analizar la similitud de las secuencias de diferentes marcadores moleculares. Sin embargo, un problema que está asociado al uso de secuencias es la dificultad de definir un punto de corte estricto y aplicable a todos los grupos de procariontes, a partir del cual deba reconocerse una especie. Por ejemplo, se ha observado que existen diferencias grandes en el grado de diversidad en las secuencias entre las especies de *Streptococcus* reconocidas como *Streptococcus pneumoniae*, *Streptococcus pseudopneumoniae* y *Streptococcus mitis* (ver FIGURA 2a). La diferencia promedio entre dos aislados de *S. mitis* (5 %) es similar a la distancia promedio entre dos aislados de *S. mitis* y *S. pneumoniae* (5.8 %) (Hanage, Fraser & Spratt, 2005, 2006; Fraser et al., 2009). Es decir, si se usara un nivel de corte fijo para distinguir estas especies tendríamos que agrupar a *S. pneumoniae* y *S. mitis* en una sola especie o separar a cada aislado de *S. mitis* en especies diferentes. Claramente, ninguna de las dos alternativas es aceptable si se busca reflejar en la delimitación de especies el proceso evolutivo que llevó a cada uno de estos grupos a mantenerse como una unidad genética discreta.



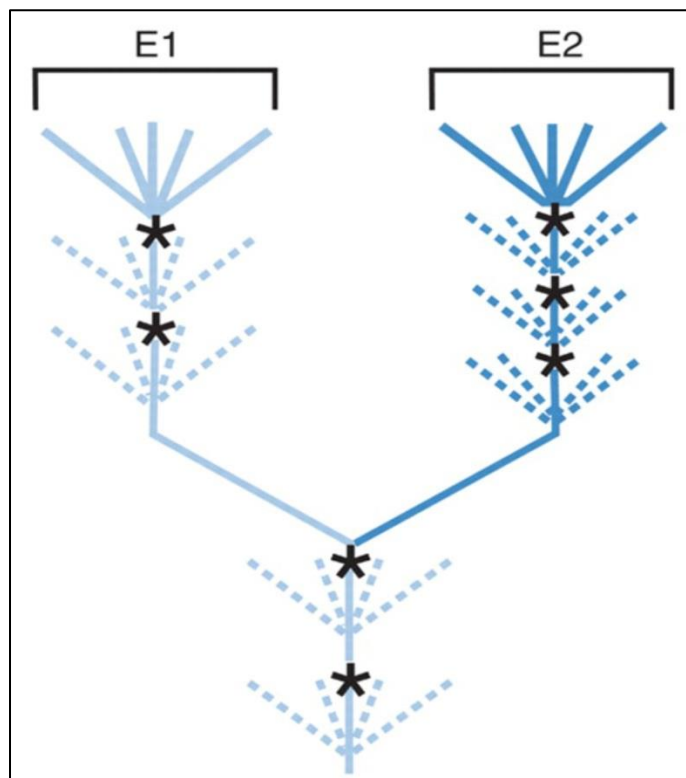
**Figura 2.** Filogenias de especies de *Streptococcus* y *Neisseria* construidas a partir de secuencias parciales de seis genes housekeeping. En el panel **a** se pueden identificar tres especies reconocidas como *S. mitis*, *S. pneumoniae* y *S. pseudopneumoniae* y los niveles de divergencia genética al interior de cada una de ellas. La divergencia fue calculada usando el porcentaje de sitios variables entre dos secuencias y se presenta el promedio de los valores calculados para cada par de secuencias. Puede observarse que la diferencia promedio entre dos aislados de *S. mitis* es similar a la distancia promedio entre dos aislados de *S. mitis* y *S. pneumoniae*. En el panel **b** se muestra la filogenia de las especies de *Neisseria* en la que pueden identificarse tres agrupamientos bien delimitados correspondientes a *N. meningitidis*, *N. lactamica* y *N. gonorrhoeae*, sin embargo puede verse también que existen otras especies como *N. sicca* o *N. mucosa* que no forman agrupamientos bien delimitados y podrían ser consideradas como “especies borrosas” (modificada de Hanage et al., 2005 y 2006 y Fraser et al., 2009).

Aunado al problema de no poder establecer un valor fijo de similitud de secuencias para definir especies que sea aplicable a todos los grupos procariontes, está el hecho de que no en todos los grupos de bacterias y arqueas podemos reconocer estos agrupamientos genéticos claramente circunscritos. Por ejemplo, en algunos géneros de bacterias como *Neisseria* los “límites” de las especies se rompen por eventos de recombinación ocasionales entre grupos genéticos distintos a los que se les ha dado la categoría de especies (ver figura 2b). Este tipo de ejemplos llevó a los autores de este trabajo a emplear el término de “especies borrosas” (“*fuzzy species*”, es el término original usado en inglés) para reconocer la existencia de estos grupos de bacterias que no necesariamente forman grupos genéticamente bien definidos y claramente distintos de otros (Hanage, Fraser & Spratt, 2005). La existencia de este tipo de especies con límites genéticos “borrosos” se ha comprobado en trabajos posteriores con otros géneros de bacterias recombinantes como *Burkholderia* y *Streptococcus* (Hanage, Fraser & Spratt, 2006) así como con arqueas del género *Halorubrum* (Papke et al., 2007). A partir de estos trabajos ha quedado claro que los límites genéticos de las especies no siempre son claros y que la recombinación frecuente puede llegar a impedir el establecimiento de dichos límites (Hanage, 2013). El papel de la recombinación en generar diversidad y promover divergencia en bacterias será tratada más ampliamente en otro apartado de esta introducción. Sin embargo cabe mencionar desde ahora que la frecuencia con la que la recombinación sucede en los diferentes grupos de bacterias es muy variable. Es la frecuencia de la

recombinación la que le permitirá actuar como una fuerza cohesiva que mantiene la unidad genética de los linajes bacterianos o como un proceso capaz también de “borrar” los límites de dichos linajes (Fraser et al., 2009).

Otro aspecto central que debe considerarse en la discusión relativa al concepto de especies en procariontes es la necesidad de incorporar aspectos ecológicos en la delimitación de las especies. En este sentido, el modelo de los ecotipos propuesto por Cohan (Cohan, 2002) presenta la idea de que las especies bacterianas deberían ser delimitadas a partir de reconocer la existencia de ecotipos en la naturaleza. Los ecotipos son grupos de organismos que comparten el mismo nicho ecológico y que son resultado de una historia evolutiva común de selección ecológica. De este modo, los organismos que integran un ecotipo deberían ser “ecológicamente homogéneos” entre ellos (Cohan & Perry, 2007; Kopac & Cohan, 2011). En este modelo, una población clonal de bacterias puede diversificar como resultado de la acumulación de mutaciones neutrales y la deriva génica actuando sobre dichas mutaciones. Ocasionalmente, una mutación benéfica puede surgir y provocar que todos los individuos que no presenten dicha mutación desaparezcan como resultado de un barrido selectivo. Los barridos selectivos recurrentes restringirán así la divergencia genética al interior de estos grupos o ecotipos. Los ecotipos que habitan nichos diferentes serán afectados por barridos selectivos diferentes, de modo que en cada uno de ellos se fijarán mutaciones adaptativas distintas (ver figura 3). Una vez que se han separado los

ecotipos se propagan clonalmente y permanecen como unidades independientes adaptadas a nichos específicos distintos. La teoría de los ecotipos es importante porque propone una explicación sobre algunos de los posibles mecanismos de especiación y reconoce la importancia del componente ecológico en formar grupos cohesivos en bacterias. Sin embargo, tiene el gran problema de asumir que la mayoría de las poblaciones procariontes son clonales. El papel de la recombinación y cómo podría ésta impedir, hasta cierto punto, la formación de los ecotipos no está claramente explicada en esta teoría.



**Figura 3.** Modelo de ecotipos para explicar la diferenciación de poblaciones de bacterias. El árbol muestra un linaje inicial que se después diferencia de dos sublinajes E1 y E2) que difieren en algún

aspecto de su ecología. La selección periódica (barridos selectivos) ocurre en los puntos marcados por asteriscos y elimina prácticamente toda la diversidad que se ha acumulado en las poblaciones desde el último barrido selectivo. Las líneas punteadas en el árbol se refieren a la diversidad purgada por la selección y mientras que las líneas continuas se refieren a la diversidad existente. Como las dos poblaciones son ecológicamente distintas (i.e. ecotipos), la selección periódica en un sublinaje no afecta la diversidad en el otro sublinaje y viceversa. Cada ecotipo entonces puede divergir y llegar a formar una especie separada (tomada de Fraser et al., 2009).

A partir de la discusión sobre el concepto de especies en bacterias y de los avances en los campos de la evolución y ecología microbianas queda claro que para entender los procesos de diversificación en procariontes hay que considerar, por un lado, el papel de la recombinación y, por el otro, la influencia de los factores ecológicos. A continuación se abordan cada uno de estos aspectos un poco más ampliamente.

### **1.1.2.2 El papel de la recombinación en las poblaciones de bacterias**

Las bacterias son organismos asexuales que se reproducen por bipartición, una célula madre da lugar a dos células hijas, cada una de las cuales contiene información genética que es idéntica a la de su progenitora. Bajo este escenario esperaríamos tener estructura poblacional totalmente clonal. La diversidad genética en las poblaciones de bacterias sería resultado exclusivo de la acumulación de mutaciones sucesivas. Sin embargo, sabemos que las bacterias tienen mecanismos (conjugación, transformación y transducción) que les permiten

llevar a cabo el intercambio horizontal (o lateral) de información genética. Estos mecanismos parasexuales de intercambio de información genética son muy diferentes de los mecanismos sexuales característicos de los eucariontes. En primer lugar están desacoplados de la reproducción de modo que no son eventos que ocurran en cada generación, en segundo lugar son unidireccionales (de un donador a un receptor) y, finalmente, sólo se transfiere en cada evento una fracción del genoma que puede involucrar desde algunos pares de bases hasta cientos de kilobases. Los mecanismos de transferencia horizontal de los procariontes pueden concluir con dos tipos de eventos de recombinación, homóloga y no homóloga, cada una de los cuales tienen consecuencias importantes sobre la estructura genética y los mecanismos evolutivos en bacterias.

La recombinación no homóloga (también conocida como recombinación ilegítima o sitio específica) es el mecanismo mediante el cual un organismo puede adquirir información genética de otro genoma diferente e incorporarlo al propio. Una característica importante de este tipo de recombinación es que los linajes entre los que se comparte la información pueden ser filogenéticamente muy lejanos. De este modo, los genes adquiridos por recombinación no homóloga pueden llevar incluso a la adquisición de nuevas funciones y provocar cambios significativos en el nicho y fenotipo de una especie (Ochman, Lawrence & Groisman, 2000; Lawrence, 2002; Gogarten & Townsend, 2005). Se ha estimado que la recombinación ilegítima puede resultar en un incremento de entre  $10^4$  y  $10^6$



en la tasa de innovación genómica entonces la tasa de mutación se vuelve despreciable (Jain et al., 2003). La comparación del creciente número de genomas bacterianos que han sido secuenciados muestra que este tipo de transferencia de información genética ha tenido un papel muy importante en la evolución de los genomas bacterianos (Ochman, Lawrence & Groisman, 2000; Gogarten, Doolittle & Lawrence, 2002; Narra & Ochman, 2006).

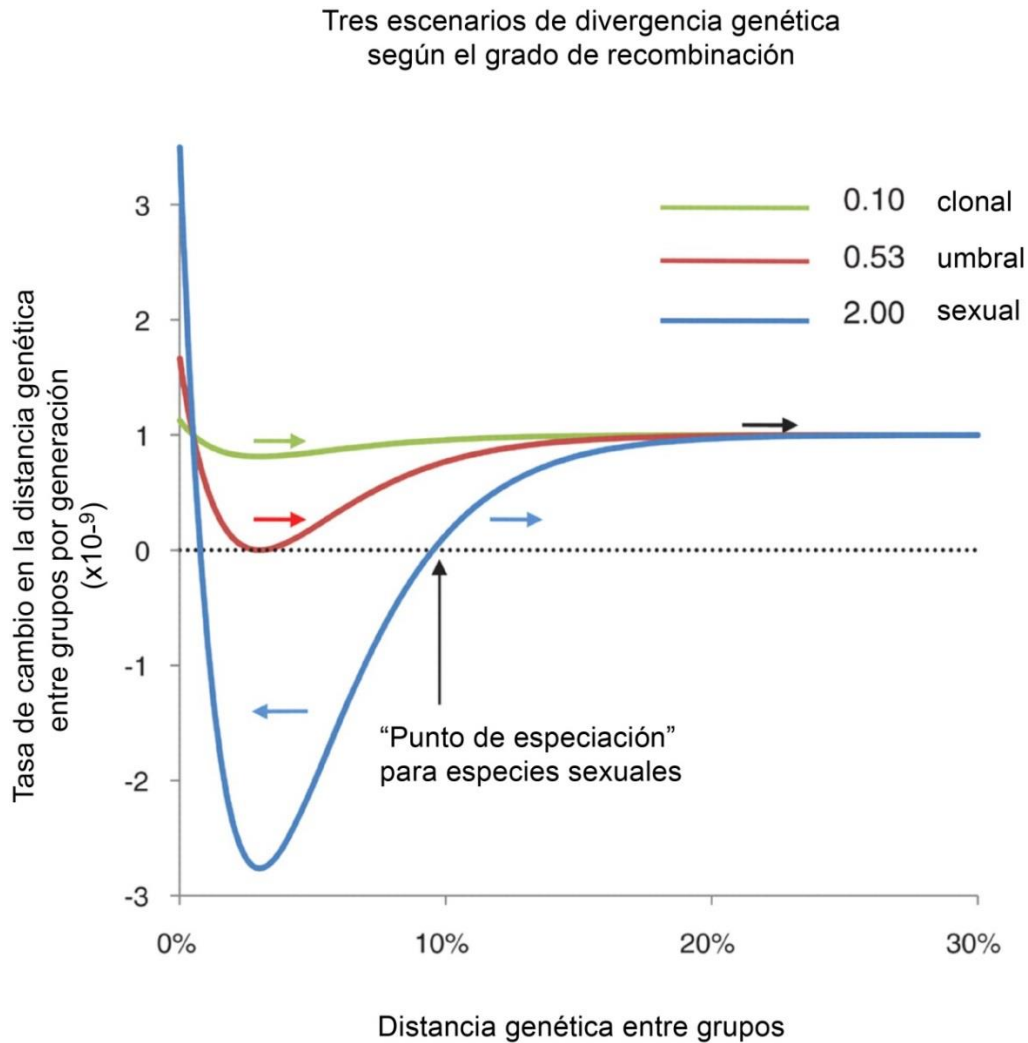
Por su parte, la recombinación homóloga es aquella que sucede entre organismos genéticamente cercanos. Para que se lleve a es necesario que el fragmento de DNA externo que se incorpora tenga su homólogo correspondiente en el genoma receptor. La recombinación de este tipo se vuelve menos probable conforme incrementa la divergencia de las secuencias entre donador y receptor. Esta característica provoca que la recombinación homóloga sea más frecuente entre organismos de la misma especie pero no significa que no pueda darse también entre organismos de especies cercanas. Debido a esta recombinación entre especies, es muy probable que casi cualquier aislado de una especie particular contenga cuando menos una proporción de material genético de alguna especie filogenéticamente cercana. Esta alta frecuencia de los eventos de recombinación ha llevado incluso a que algunos autores sugieran que las bacterias en realidad no forman especies debido a que no hay un límite al flujo de genes entre ellas (Doolittle & Papke, 2006). Sin embargo, a pesar de la alta prevalencia de la recombinación aún entre linajes diferentes, existen también

evidencias que sugieren que sí es posible reconocer la presencia de grupos genéticos discretos y que es posible reconstruir las filogenias de los linajes procariontes (Papke et al., 2007; Fraser et al., 2009; Vos, 2010; Hanage, 2013; Shapiro & Polz, 2014).

Un gran número de trabajos desarrollados en las últimas décadas han demostrado que la recombinación es mucho más frecuente en las poblaciones procariontes de lo que previamente se pensaba (Souza et al., 1992; Istock et al., 1992; Feil & Spratt, 2001; Silva et al., 2003; Gogarten & Townsend, 2005; Hanage, Fraser & Spratt, 2005; Narra & Ochman, 2006; Vos & Didelot, 2009). Se reconoce a la recombinación como un fenómeno que contribuye de manera importante a la generación de la diversidad, la redistribución de la misma en las poblaciones e incluso a la adaptación a diferentes ambientes. A partir de estos trabajos hoy sabemos que existe una amplia gama de tasas de recombinación en los diferentes linajes de bacterias y arqueas que han sido analizados. Existen especies que pueden ser consideradas clonales como *Staphylococcus aureus*, *Salmonella sp* o *Lactobacillus casei* (Maynard Smith et al., 1993; Enright et al., 2000; Feil et al., 2003) mientras que existen otras que son altamente recombinantes y pueden considerarse como poblaciones panmícticas como son *Helicobacter pylori*, *Streptococcus pneumoniae* y varias especies del género *Neisseria* (Suerbaum et al., 1998; Hanage, Fraser & Spratt, 2005; Vos & Didelot, 2009). Es precisamente la frecuencia con la que sucede la recombinación la que determina su efecto sobre

las poblaciones. Si es muy intensa puede diluir la cohesividad histórica de un linaje y producir filogenias en forma de red. Sin embargo, también puede actuar como una fuerza de cohesión dentro de los linajes cuando es más frecuente al interior de las especies que entre ellas (actuando así como una barrera al flujo génico que puede equipararse con el aislamiento sexual de los eucariontes). La intensidad de la recombinación determinará así la facilidad con la que se promueve la divergencia genética entre grupos de organismos (Fraser, Hanage & Spratt, 2007; Fraser et al., 2009). En la figura 4 se explica cómo estos procesos de divergencia suceden en función de las tasa de recombinación. En el eje Y se muestra la tasa de divergencia entre dos linajes a lo largo del tiempo. Considerando la divergencia como la tasa de cambio de la distancia genética entre dos linajes, los valores negativos indican convergencia y homogeneización entre linajes mientras que los valores positivos se refieren a la divergencia entre linajes. Puede observarse como cuando la tasa de recombinación es muy baja los linajes tienden a divergir más rápidamente mientras que, conforme la recombinación aumenta, ésta aumenta la cohesividad al interior de los linajes, se homogeneiza la poza génica y la divergencia es mucho más lenta. Para que los linajes con altas tasas de recombinación puedan divergir es necesario que exista algún proceso que disminuya, cuando menos por un periodo de tiempo, el intercambio genético entre linajes, por ejemplo, un período de alopatría o de diferenciación ecológica. En estas especies, el “punto de especiación” es la cantidad de divergencia entre

linajes que necesita acumularse para impedir que vuelvan a convergir en un solo linaje cuando las barreras a la recombinación dejan de existir.



**Figura 4.** Divergencia genética entre dos grupos a partir de modelos de poblaciones contrastantes en los niveles de recombinación (clonal vs sexual o recombinante). Cuando se tienen tasas de recombinación bajas, la población se define como clonal y los grupos se diferenciarán genéticamente de manera rápida (**línea y flechas verdes**). Conforme aumenta la tasa de recombinación, dicho proceso actuará como fuerza cohesiva en las poblaciones y retrasará el proceso de divergencia. Si la tasa de recombinación es mayor al umbral (**línea roja**) las poblaciones tenderán a la convergencia (**línea y flechas azules**). De este modo, en grupos con altas tasa de recombinación los procesos de divergencia serán promovidos por otros factores como

barreras geográficas y/o ecológicas que eviten la convergencia genética por homogeneización de la poza génica. El punto de especiación para los grupos recombinantes se refiere a la cantidad de divergencia genética (entre linajes) que necesita acumularse para impedir que los linajes vuelvan a convergir en unos solo una vez que las barreras externas que limitan la recombinación dejen de existir.

Además de su contribución como fuerza cohesiva o diversificadora en los linajes procariontes, la recombinación también tiene efectos al interior de las poblaciones que son muy importantes y que afectan de manera directa la dinámica evolutiva de las mismas. La recombinación genera nuevas combinaciones de alelos y rompe el ligamiento entre genes. De este modo, cuando la recombinación es frecuente, se espera un escenario en el que las poblaciones tengan una diversidad genética alta y exista escaso ligamiento entre genes. Este tipo de poblaciones pueden entenderse bajo los supuestos clásicos de la genética de poblaciones debido a que, en términos genéticos, se comportan como poblaciones de organismos eucariontes con reproducción sexual. Sin embargo, en poblaciones completamente clonales donde todos los genes están ligados, la unidad sobre la cual la selección natural actuaría sería el genoma completo. De este modo, cualquier mutación adaptativa que surgiera se fijaría en la población gracias a la selección periódica. Linajes nuevos se estarían formando todo el tiempo y desplazando a los anteriores en barridos selectivos. La diversidad genética se mantendría en niveles muy bajos debido a los constantes barridos selectivos, Así, las poblaciones de bacterias evolucionarían de acuerdo con el modelo de los ecotipos (Cohan, 2002) y, en el extremo, cada clona debería considerarse como

una población. Los supuestos clásicos de la genética de poblaciones no podrían utilizarse para el análisis de este tipo de poblaciones. En este sentido, el gran reto para la genética de poblaciones de bacterias está en determinar, por un lado, las contribuciones relativas de los procesos evolutivos que generan variación genética (mutaciones puntuales y recombinación) y, por el otro, entender los factores que subsecuentemente determinan sus caminos evolutivos (la selección y la deriva génica).

### **1.1.2.3 La importancia de la geografía y el ambiente**

Parte importante de la información que tenemos actualmente sobre los procesos de diversificación y evolución en procariontes han surgido a partir del estudio de microorganismos patógenos. Sin embargo, considerando que la mayor parte de la diversidad procarionte se encuentra en ambientes naturales no asociados al humano (Whitman, Coleman & Wiebe, 1998; DeLong & Pace, 2001), es en esos ambientes donde es necesario ampliar nuestro conocimiento sobre los procesos que determinan la evolución de bacterias y arqueas. Además de entender el papel de la mutación y la recombinación como mecanismos generadores de diversidad en procariontes, es esencial considerar también el papel de las barreras geográficas y ecológicas en el proceso de formación y diversificación de linajes bacterianos (Whitaker, 2006).

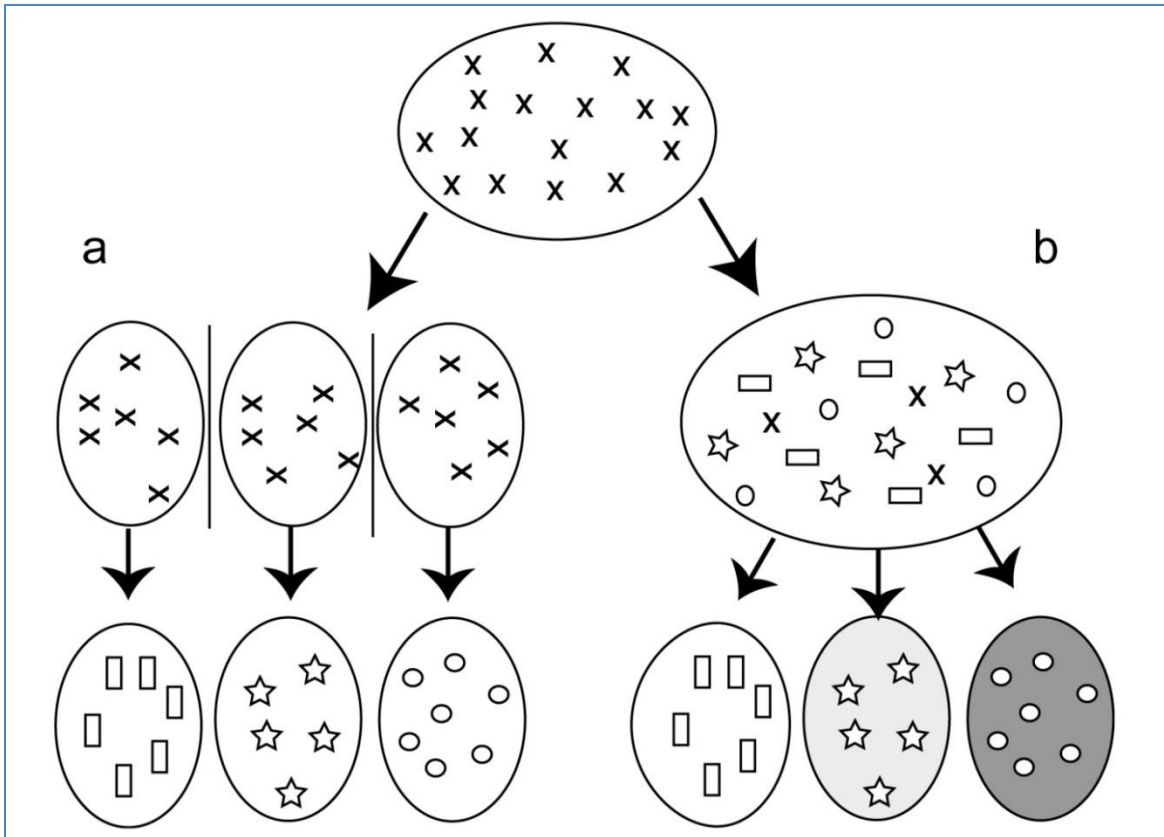
El estudio de los mecanismos de diversificación en microorganismos implica identificar los patrones de variación genética que se pueden reconocer tanto al interior como entre las poblaciones así como las barreras que estructuran dichos patrones y dan lugar a linajes genéticos que se mantienen a lo largo del tiempo. Gracias a las herramientas moleculares desarrolladas en las últimas décadas se ha vuelto posible el análisis de muestras ambientales de mayor tamaño. Por otra parte, algunas de las preguntas de los ecólogos evolutivos, que históricamente han trabajado con macroorganismos, empiezan a tener atención en el campo de la microbiología. Por ejemplo, la importancia de las barreras geográficas en microorganismos ha empezado a cuestionarse en años recientes (Whitfield, 2005; Martiny et al., 2006; Whitaker, 2006), mientras que para macroorganismos, la especiación alopátrica es uno de los mecanismos más aceptados (Coyne & Orr, 2004).

La especiación alopátrica resulta de la evolución divergente de poblaciones geográficamente aisladas. Ha sido particularmente complicado reconocer la diferenciación fenotípica entre poblaciones aisladas así como identificar las barreras geográficas potenciales que ocasionen una interrupción en el flujo génico entre ellas debido al tamaño pequeño de bacterias y arqueas. La visión clásica acerca de los factores que determinan la distribución de los microorganismos, plantea un modelo en el cual las poblaciones microbianas no tienen ningún límite a su dispersión gracias a su tamaño pequeño, su gran plasticidad metabólica y su

abundancia (Finlay, 2002). Sin embargo, conforme los estudios de biogeografía microbiana aumentan, empieza a quedar claro que algunas especies sí presentan patrones de distribución particulares relacionados ya sea con barreras geográficas o con procesos de adaptación local (Papke et al., 2003; Whitaker, Grogan & Taylor, 2003; Vinuesa et al., 2005; Whitaker, 2006; Vos & Velicer, 2008; Dvořák, Hašler & Poulíčková, 2012; Sul et al., 2013)

Los patrones biogeográficos observados en los organismos se refieren a la división y distribución de la variación en grupos únicos y pueden ser resultado tanto de barreras físicas que limitan la dispersión (divergencia alopátrica) como de factores ambientales (divergencia ecológica) (Lomolino, Riddle & Brown, 2005). En la figura 5 se representan de manera esquemática estos dos modelos que pueden explicar los patrones de divergencia en los organismos asociados a la geografía o el ambiente.





**Figura 5.** Dos modelos de diferenciación genética a partir de una población inicial homogénea. En el panel **a** se muestra el modelo de divergencia alopátrica en la que una población inicial es separada en tres subpoblaciones aisladas unas de otras por barreras geográficas; en este caso el aislamiento geográfico permitirá la diferenciación de cada población en un linaje nuevo distinto genéticamente al que le dio origen. En el panel **b** se muestra el modelo de divergencia ecológica en la que la población homogénea evoluciona en una población heterogénea que diverge a partir de que cada genotipo explora un nicho ecológico independiente (tomada de Whitaker, 2006).

Los patrones de distribución cosmopolitas son resultado de una tasa alta de dispersión y de la homogenización de la poza génica (a través de la recombinación homóloga. La capacidad de dispersión tiene que ver con las capacidades físicas de transporte y sobrevivencia de los organismos. Por lo tanto, suponemos que aquellos microorganismos que tienen rangos de distribución

amplios serán los aquellos que poseen estructuras que les permiten un movimiento activo (flagelos, por ejemplo) así como aquellos cuyo movimiento pasivo es facilitado por la formación de estructuras de resistencia como endosporas o biopelículas. La dispersión puede ser facilitada también por la asociación con macroorganismos migratorios incluidos los humanos. De este modo, no es sorprendente encontrar muchos estudios de bacterias patógenas que muestran una aparente distribución cosmopolita (Smith, 1991; Smith, Feil & Smith, 2000; Feil & Spratt, 2001). La tasa de migración o dispersión también está relacionada con la densidad de las poblaciones microbianas debido a que la probabilidad de colonizar un nuevo hábitat depende directamente de que haya un mayor número de organismos que se pueden propagar. Recientemente se han desarrollado algunos modelos teóricos para poner a prueba esta idea y se ha mostrado que la tasa de migración aumenta conforme aumenta el tamaño de las poblaciones (Sloan et al., 2006). Otro componente importante que determina la probabilidad de la dispersión es la estructura espacial del hábitat microbiano. Por ejemplo, los microorganismos adaptados a hábitats separados por distancias grandes (que se asemejan a un modelo de islas) tienen tasas de migración menores que aquellos que se encuentran en ambientes con una distribución más continua (Staley & Gosink, 1999) Finalmente, es importante considerar que la identificación de barreras geográficas en microorganismos depende en gran medida de la escala espacial usada. Mientras que existen algunas especies de bacterias en las que se pueden reconocer las barreras geográficas a una escala

de centímetros (Oda et al., 2003; Vogel et al., 2003) puede haber otras en las que los patrones sólo se vuelvan aparentes cuando se analizan con una escala global o de varios cientos de kilómetros (Cho & Tiedje, 2000; Vos & Velicer, 2008).

Gracias al avance de la biogeografía de microorganismos ha quedado claro que las barreras geográficas pueden afectar los patrones de distribución de arqueas y bacterias y que existen algunas especies en las cuales el efecto del aislamiento por distancia es muy claro, cuando menos a escalas continentales (Cho & Tiedje, 2000; Whitaker, Grogan & Taylor, 2003; Vos & Velicer, 2008). Sin embargo, para poder explicar mejor algunos de los patrones de divergencia encontrados en bacterias y arqueas debemos tener en cuenta también que la selección natural es una fuerza fundamental que puede promover la adaptación a escalas locales y así generar divergencia entre linajes (Vos, 2010; Shapiro et al., 2012). A pesar de que sabemos que las condiciones ambientales son fundamentales para el proceso de adaptación local, existen poco trabajos que hayan encontrado evidencias claras de factores ambientales y/o ecológicos directamente relacionados con los patrones de divergencia genética a nivel poblacional. En gran medida esta falta de avances se debe a que existen pocos estudios de genética de poblaciones y evolución molecular que se centren en poblaciones silvestres de procariontes. Además, determinar el nicho ecológico y los factores ambientales a escalas de micro hábitats supone una serie de dificultades técnicas importantes y puede resultar muy complicado.

Algunos de los trabajos que han explorado la importancia que los factores ambientales y/o ecológicos tienen en determinar la estructura de las poblaciones de bacterias de vida libre sí han podido identificar algunas condiciones ambientales particulares asociadas a la distribución de la diversidad genética de ciertas especies bacterianas. Por ejemplo, estudios con poblaciones de *Bacillus simplex* en los suelos de dos regiones de Israel con condiciones climáticas contrastantes permitieron determinar que la radiación solar es un factor que afecta la estructura de las poblaciones de estas bacterias (Sikorski & Nevo, 2005, 2007). En otro caso, las poblaciones de otra bacteria ambiental recuperada de suelos del género *Desulfolobus* mostraron una estructura poblacional que responde a las condiciones de salinidad presentes en su ambiente (Oakley et al., 2010). En un trabajo más reciente se analizaron más de 1000 aislados del género *Vibrio* recuperados en el océano y se pudo determinar que existen subgrupos genéticos dentro de las poblaciones asociados a partículas planctónicas de diferentes tamaños (Hunt et al., 2008). Además, el análisis genómico de 20 de estos aislados permitió comprobar que la diferenciación genética es resultado de un proceso de diferenciación ecológica reciente al interior de las poblaciones. La comparación de los genomas permitió detectar genes y polimorfismos únicos asociados a cepas con nichos ecológicos diferentes, lo cual permite proponer que estas subpoblaciones están en un proceso de especiación incipiente (Shapiro et al., 2012). Evidencias similares se encontraron en el trabajo de Rebollar y

colaboradores (Rebollar et al., 2012) en el cual, trabajando con poblaciones ambientales del género *Exiguobacterium*, se pudo detectar la existencia de subgrupos compuestos por organismos asociados exclusivamente al sedimento o al agua. Todos los trabajos anteriores sugieren que la divergencia genética en poblaciones simpátricas de bacterias puede, en efecto, ser promovida por procesos de adaptación a ambientes y condiciones ecológicas particulares.

Finalmente, los estudios de genómica comparada también han permitido reunir evidencias de que los procesos de divergencia ecológica y adaptación local resultado de la selección positiva puede moldear la diversidad genética de poblaciones ambientales de bacterias (Bielawski et al., 2004; Shapiro et al., 2009). Por su parte, la secuenciación de genomas completos de algunos organismos recuperados de sitios con características ambientales particulares ha permitido detectar la importancia de eventos de transferencia horizontal que están asociados a la adaptación a nichos ecológicos específicos. Tal es el caso de especies como *Bacillus coahuilensis* (Alcaraz et al., 2008) y *Sulfolobus islandicus* (Cadillo-Quiroz et al., 2012).

Con base en la revisión aquí presentada queda claro que si bien en las últimas décadas los avances en los campos de la ecología y evolución de procariontes han sido notables, siguen existiendo muchas preguntas importantes sin resolver. En este sentido cabe mencionar que, en parte debido a las

dificultades que representa su aislamiento y también en parte debido al interés que históricamente ha existido por entender la causa de las enfermedades, el estudio de poblaciones de bacterias de vida libre (o ambientales) ha sido mucho menor que el de las bacterias patógenas. Sin embargo, la mayor parte de la diversidad procarionte se encuentra en ambientes naturales (DeLong & Pace, 2001). Seguir ampliando el conocimiento sobre poblaciones de bacterias de vida libre es lo que nos permitirá entender los procesos generales que operan en su adaptación y evolución y explicar así la gran diversidad de la vida procarionte sobre el planeta.

## **Capítulo 2. ANTECEDENTES Y JUSTIFICACIÓN DEL PROYECTO**

En este apartado de la tesis se describirán brevemente algunas de las características del sitio de estudio que lo hacen particularmente adecuado para estudiar la evolución y ecología de poblaciones naturales de bacterias. Se presentarán también algunos trabajos previos de genética de poblaciones y evolución molecular llevados a cabo con el género *Bacillus*, que dieron lugar al desarrollo del presente trabajo.

### **2.1 Sitio de estudio: Cuatro Ciénegas Coahuila**

El valle de Cuatro Ciénegas, Coahuila (CCC) es el humedal más importante del desierto Chihuahuense y el más vasto de América del norte. Se encuentra rodeado por grandes cadenas montañosas y su formación es resultado de una historia geológica compleja (Wilson & Pitts, 2010; Souza et al., 2012). Este valle es el lugar con más especies endémicas por kilómetro cuadrado del continente, albergando más de 70 especies endémicas de plantas y animales (Souza et al., 2006). Se estima que en CCC existen alrededor de unas 200 pozas además de contar con otros cuerpos de agua como son manantiales, lagunas, ciénegas, canales artificiales y ríos (Johannesson, Cortés & Kilroy, 2004). Dichos sistemas acuáticos se caracterizan por tener condiciones ambientales y de disponibilidad de nutrientes contrastantes entre sí (Elser et al., 2005; Rebollar et al., 2012). En

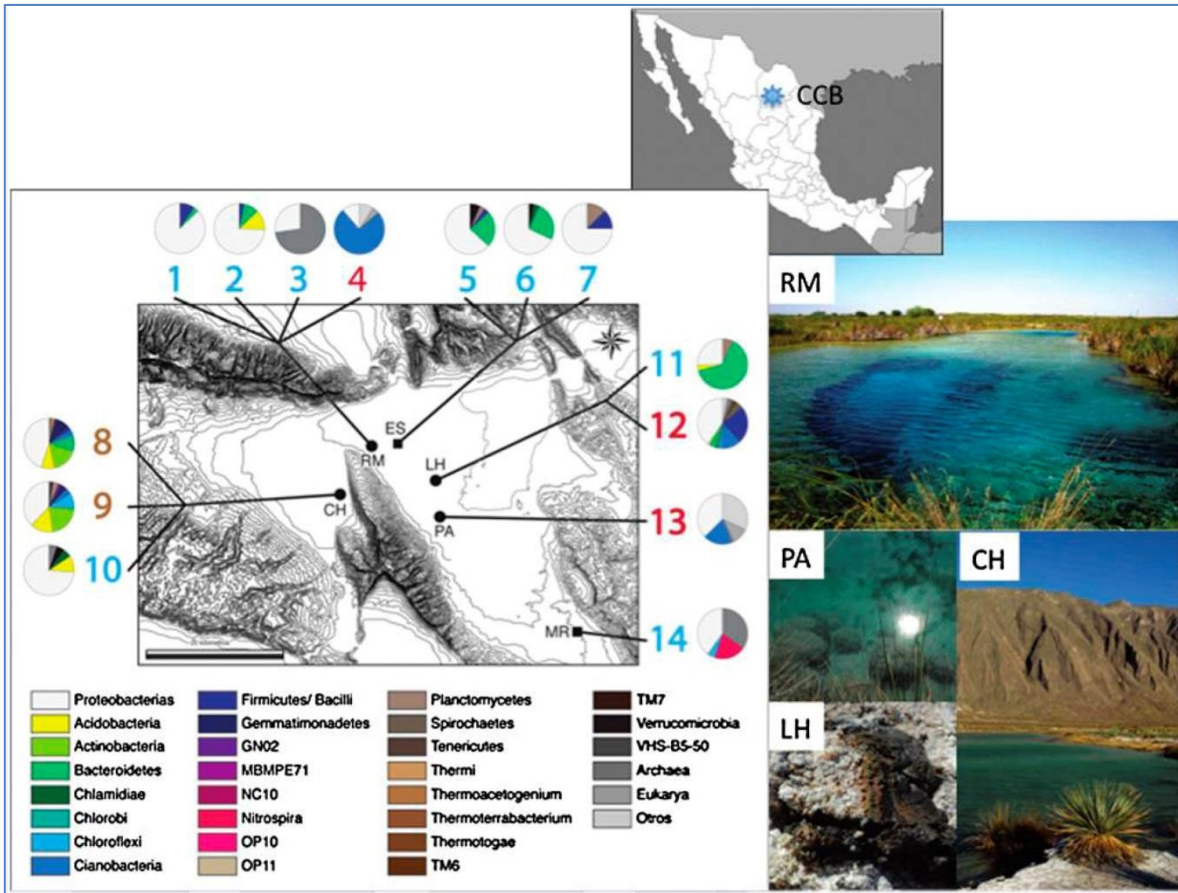
particular, los sistemas acuáticos del valle se caracterizan por ser oligotróficos y tener muy bajos niveles de fósforo disponible, de hecho, dichos sistemas presentan la concentración más baja de fósforo de los sistemas acuáticos que se han estudiado hasta la fecha (Elser et al., 2005; Souza et al., 2008). La limitación de nutrientes ha favorecido que los sistemas acuáticos del valle estén dominados por comunidades microbianas abundantes y diversas (Souza et al., 2006; Escalante et al., 2008; Cerritos et al., 2011). Es así como el sistema aislado que forma el valle de Cuatro Ciénegas, su geología y las características ambientales anteriormente mencionadas lo convierten en un sitio adecuado para entender los mecanismos ecológicos y evolutivos importantes en la generación de diversidad en poblaciones ambientales de bacterias.

La primera descripción de la composición de la diversidad de las comunidades bacterianas de CCC se obtuvo gracias a un muestreo amplio que incluyó los principales sistemas acuáticos dentro del valle y sus alrededores. En este primer trabajo se detectó una gran diversidad de grupos bacterianos, la mayoría de los cuales además mostraron estar filogenéticamente relacionados con bacterias de origen marino (Souza et al., 2006). La afinidad con linajes marinos también se ha observado en comunidades de virus que pueden encontrarse en estromatolitos de CCC (Desnues et al., 2008). Para explicar la afinidad de las comunidades de microorganismos de CCC con linajes marinos se ha propuesto



que dichas comunidades son un relicto de la época en la que el valle estaba todavía bajo el mar (hace unos 150 millones de años; (Wolaver et al., 2008).

Estudios posteriores con comunidades microbianas de organismos cultivables recuperadas del valle mostraron que dichas comunidades tienen una alta diversidad  $\beta$  (Escalante et al., 2008), es decir, cada sitio de muestreo tiene una composición de grupos bacterianos particular, probablemente debido a la heterogeneidad ambiental en el valle. Otro trabajo mostró que en particular la comunidad de firmicutes y actinobacterias provenientes de uno de los sistemas acuáticos de CCC conocido como Churince, presenta un patrón de distribución de dichos grupos bacterianos que está relacionado con el gradiente de salinidad ambiental (Cerritos et al., 2011). En años más recientes también se han estudiado las comunidades bacterianas con técnicas independientes de cultivo. Trabajos metagenómicos de algunos tapetes microbianos y estromatolitos de CCC (Desnues et al., 2008; Breitbart et al., 2009; Bonilla-Rosso et al., 2012; Peimbert et al., 2012) han mostrado también que las comunidades ahí encontradas son muy diversas, que son complejas en términos metabólicos y que dicha complejidad metabólica parece estar relacionada con las condiciones ambientales únicas de cada sistema acuático del valle. En la figura 6 se muestra un resumen de la diversidad microbiana a nivel de phylum en los sistemas acuáticos de CCC hasta ahora más estudiados.



**Figura 6.** Diversidad microbiana a nivel de phylum en diferentes sitios de Cuatro Ciénegas, Coahuila. La identidad taxonómica fue determinada usando la secuencia del gen 16S rRNA amplificada a partir de DNA ambiental. Los números se refieren al total de muestras tomadas en cada región. CH se refiere a Churince, RM se refiere a Río Mezquites (que incluye pozas de desecación y un río con oncolitos), ES se refiere a Escobedo, LH se refiere a Los Hundidos, PA se refiere a Pozas Azules y MR se refiere a una zona minera donde la muestra fue obtenida 600 m por debajo de la superficie. El mapa de México muestra el valle de Cuatro Ciénegas en el centro del estado de Coahuila y las fotografías corresponden a los sitios de muestreo de Cuatro Ciénegas (Tomada de Souza et al., 2012).

Por otro lado, a partir de algunos aislados cultivables obtenidos en CCC, se han podido describir y caracterizar especies que parecen ser únicas de este sitio

de estudio, como el caso de *Pseudomonas cuatrocienegasensis* (Escalante et al., 2009) y *Bacillus coahuilensis* cuatro (Cerritos et al., 2008). En particular, el análisis detallado del genoma de *B. coahuilensis* ha mostrado que esta especie presenta adaptaciones a las condiciones locales. Por ejemplo, esta bacteria posee en su genoma genes que le permiten sintetizar sulfolípidos de membrana en lugar de fosfolípidos (Alcaraz et al., 2008). Dicho rasgo representa una gran ventaja adaptativa en un ambiente en donde la disponibilidad de fósforo es un factor limitante. Finalmente, los trabajos de ecología y evolución molecular llevados a cabo hasta ahora en CCC han permitido proponer un modelo de diversificación microbiana en el cual la falta de fósforo podría ser responsable de inhibir algunos de los procesos de homogeneización (principalmente el intercambio genético por recombinación) de las poblaciones naturales de bacterias que ahí habitan (Souza et al., 2008).

## **2.2 Grupo de estudio: el género *Bacillus***

El género *Bacillus* es un género de bacterias Gram positivas con bajo contenido de G+C (Reino Bacteria; Phylum Firmicutes; Clase Bacilli; Orden Bacillales; Familia Bacillaceae) que actualmente está integrado por más de 300 especies reconocidas (<http://www.bacterio.net/b/bacillus.html>). Se caracterizan por tener una gran variedad de estilos de vida y ser ubicuos en la naturaleza (Claus & Berkeley, 1986). Una característica distintiva del género es su capacidad de

formar endosporas como estructuras de resistencia que contribuyen a su capacidad para colonizar ambientes muy diversos (Slepecky & Hemphill, 2006). Representantes del género *Bacillus* han sido aislados de ambientes tan diversos como cuerpos de agua continentales y agua de mar, suelo, plantas, animales e incluso el aire (Pignatelli, Moya & Tamames, 2009). La diversidad fenotípica del género es muy amplia, son capaces de sobrevivir a altas temperaturas, salinidades extremas, condiciones ácidas y también existen algunos grupos especializados en burlar al sistema inmune de varios animales (Holt, 1986). Su diversidad metabólica es notable también. *B. arsenicoselenatis* y *B. selenitireducens*, por ejemplo, pueden vivir a partir de usar selenio y arsénico como aceptores de electrones (Switzer Blum et al., 1998). Otros como *B. carboniphilus* pueden crecer sobre superficies de carbón activado (Fujita et al., 1996) o en sitios tan peculiares como la superficie de pinturas, como *B. barbaricus* (Täubel et al., 2003). La diversidad metabólica del género *Bacillus* ha permitido la utilización de algunos de sus miembros en procesos industriales de obtención de riboflavina, streptavidina,  $\beta$ -lactamasa y diversas toxinas de insectos y nematodos usadas como insecticidas (Maughan & Van der Auwera, 2011). Finalmente, otros grupos de *Bacillus* como *B. anthracis* y *B. cereus* se caracterizan por provocar enfermedades en humanos.

Algunas especies de este género, como *B. subtilis* y *B. anthracis*, están dentro de las primeras especies de bacterias descritas en la historia de la

microbiología. Ferdinand Cohn fue el primero en describir el proceso de esporulación en *B. subtilis* (Cohn, 1962) y Robert Koch probó su famosa teoría de la causa de que las enfermedades son causadas por gérmenes usando a *B. anthracis* (Koch, 1962). Aunque la diversidad de *Bacillus* es muy grande, *B. subtilis* y el linaje de *B. anthracis* siguen siendo las especies más ampliamente estudiadas hasta la fecha.

Algunos de los primeros trabajos de genética de poblaciones con bacterias de vida libre fue llevado a cabo por Istock y colaboradores (Istock et al., 1992) con una población de *B. subtilis* proveniente del Desierto de Sonora. Usando 60 aislados de dicha especie y técnicas de MLEE los autores buscaron describir la estructura genética de dicha población y encontraron evidencia de que, en la población estudiada, la recombinación era frecuente (Istock et al., 1992). De hecho, este trabajo es uno de los primeros que muestran que la recombinación juega un papel muy importante en moldear la estructura genética de las poblaciones de bacterias, oponiéndose así a la teoría más aceptada en esa época, la cual asumía que la estructura poblacional de las bacterias debía ser clonal. En un trabajo posterior, Duncan y colaboradores (Duncan et al., 1994) analizaron la estructura genética de dos poblaciones simpátricas de *B. subtilis* y *B. licheniformis* del Desierto de Sonora y encontraron que la recombinación es frecuente dentro de cada especie pero no entre especies, lo cual reafirma la idea de que la recombinación puede actuar como fuerza cohesiva al interior de los linajes

bacterianos. Finalmente, en un trabajo más reciente, Istock y colaboradores (Istock et al., 2001) aportaron nuevos datos sobre la diversidad geográfica de los linajes de *B. subtilis*. Al analizar un total de 106 aislados provenientes de Arizona, California, Utah, México, Chile, Chad, Túnez y China, dichos autores encontraron seis linajes genómicos principales dentro de la muestra. Cuatro de los seis linajes estuvieron presentes en tres de los cuatro continentes revisados, mientras que los otros dos fueron encontrados sólo en uno o dos continentes. Este resultado sugiere que dichos linajes tienen una distribución cosmopolita y debieron haberse dispersado desde su sitio de origen. Dentro de cada gran linaje se encontraron sublinajes, algunos de los cuales mostraron una diferenciación geográfica local en sus genomas mientras que otros presentaron una mezcla de genomas similares de regiones geográficas distantes.

El grupo de *Bacillus cereus* sensu lato en el cual se incluye a *B. anthracis* es otro de los que ha sido más ampliamente estudiado dentro de este género. Este grupo incluye a seis especies muy cercanamente relacionadas: *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides* y *B. weinhestephanensis*. La identidad taxonómica de cada una de estas especies ha sido objeto de debate en los últimos años, en particular a partir que el análisis de datos moleculares ha mostrado que los límites entre los miembros de este grupo pueden resultar muy difíciles de definir y que posiblemente es necesario replantear la definición de estas y otras especies del género *Bacillus* (Maughan & Van der

Auwera, 2011). *B. anthracis* es un patógeno obligado de mamíferos y es el agente causante del ántrax o carbunco (Passalacqua & Bergman, 2006; Pilo & Frey, 2011). *B. cereus* es un patógeno oportunista de humanos que genera envenenamiento por comida y es responsable de contaminaciones en hospitales y sitios de fabricación de alimentos (Tourasse et al., 2011). *B. thuringiensis* es un patógeno de insectos bien conocido por su uso como biopesticida en todo el mundo (Soberón et al., 2007) aunque ocasionalmente puede también actuar como patógeno oportunista en humanos causando necrosis de tejidos, infección pulmonar e intoxicación por alimentos (Ghelardi et al., 2007; McIntyre et al., 2008). *B. weinhephanensis* es una bacteria tolerante al frío y, junto con *B. mycoides* y *B. pseudomycoides*, son generalmente consideradas las especies no patógenas del grupo de *B. cereus* (Maughan & Van der Auwera, 2011). Sin embargo, los organismos del grupo de *B. cereus* en general están muy cercanamente relacionados en términos genómicos (Helgason et al., 2000; Tourasse et al., 2006; Kolstø, Tourasse & Økstad, 2009) y todas las especies que lo conforman tienen el potencial de ser patógenas debido a que tienen en su cromosoma un conjunto común de enterotoxinas y enzimas que pueden causar diarrea y otras enfermedades gastrointestinales (Tourasse et al., 2011).

Debido a su importancia médica y económica, la estructura genética y las relaciones filogenéticas dentro del grupo de *Bacillus cereus* han sido ampliamente estudiadas usando diversos métodos de tipificación buscando

reconstruir la historia evolutiva del surgimiento de las especies y linajes que lo integran así como identificar los grupos clonales responsables de las enfermedades que generan. Análisis a gran escala de este grupo han incluido la caracterización usando MLEE (Helgason et al., 1998, 2000; Vilas-Boas et al., 2002), análisis de polimorfismos en la longitud de fragmentos amplificados (AFLP por sus siglas en inglés, *Amplified Fragment length Polymorphism*) y, más recientemente, análisis de tipos MLST (Helgason et al., 2004; Ko et al., 2004; Priest et al., 2004; Sorokin et al., 2006; Tourasse et al., 2006; Cardazzo et al., 2008; Hoffmaster et al., 2008; Didelot et al., 2009). Los diferentes análisis de tipificación llevados a cabo con el grupo de *B. cereus* han mostrado que el grupo *B. cereus sensu lato* forma un grupo filogenético en el cual los miembros de cada especie frecuentemente se encuentran entremezclados (Tourasse et al., 2011). Sin embargo, la excepción está representada por *B. anthracis* que forma un linaje altamente clonal al interior de *B. cereus sensu lato* y *B. pseudomycooides* que forma un grupo muy divergente que se comporta como grupo externo del grupo de *B. cereus* (Fagerlund et al., 2007; Auger et al., 2008; Lapidus et al., 2008). Por su parte, los análisis genéticos del grupo de *B. cereus* han mostrado de manera independiente que dicha especie está dividida en múltiples grupos filogenéticos y que las cepas relacionadas con el envenenamiento por alimentos están distribuidos a lo largo de toda la filogenia (Tourasse et al., 2011). Además, estos subgrupos filogenéticos muestran diferencias ecológicas, en particular relacionadas con la tolerancia a la temperatura (Guinebretière et al., 2008). La



población del grupo *B. cereus* parece ser sólo débilmente clonal, aunque es posible identificar algunos complejos clonales de cepas relacionadas con aislados clínicos o contaminantes de alimentos, también existen evidencias de recombinación y transferencia horizontal de genes (Tourasse et al., 2011). Análisis de tipo MLST al interior de este grupo producen filogenias que no son congruentes, lo cual es indicativo de que la recombinación ha tenido un papel importante en su evolución (Didelot & Falush, 2007; Didelot et al., 2009).

Sin embargo, un problema con la mayoría de los trabajos que se refieren al grupo de *B. cereus* sensu lato es que casi todos ellos se han llevado a cabo con colecciones de cepas de orígenes muy heterogéneos en lugar de tratar de entender primero la estructura poblacional a una escala local. En este sentido, el trabajo de Vilas-Bôas y colaboradores (Vilas-Boas et al., 2002) es particularmente interesante porque es uno de los pocos que analizan cepas obtenidas de poblaciones naturales de dos especies del grupo de *B. cereus*. En este estudio, los autores trabajaron con cepas de *B. thuringiensis* y *B. cereus* aisladas al mismo tiempo del suelo de dos sitios geográficamente distantes. Con las muestras, por medio de MLEE, los autores analizaron la estructura genética de dichas poblaciones para explorar la importancia relativa del intercambio genético intra e interespecífico, con la intención de probar así si la distancia genética entre estas dos especies está relacionada con la distancia geográfica. Los resultados

mostraron que sí existe una diferenciación genética entre poblaciones de *B. cereus* y *B. thuringiensis*; las poblaciones geográficamente distantes de cada una de las especies de *Bacillus* mostraron ser genéticamente más similares entre sí que con respecto a las poblaciones de la otra especie, aun cuando compartieran un mismo espacio geográfico. Lo anterior sugiere que el intercambio genético entre cepas de la misma especie es mucho mayor que entre especies. En el caso de este trabajo también se propone que la recombinación actúa como una fuerza que unifica las pozas génicas al interior de las poblaciones.

### **2.2.1 El género *Bacillus* en Cuatro Ciénegas**

Dentro de la diversidad bacteriana de Cuatro Ciénegas existen algunos géneros que han sido recuperados frecuentemente y cultivados con éxito a partir de muestras de diferentes sistemas acuáticos dentro del valle. Algunos de estos son *Pseudomonas*, *Exiguobacterium* y *Bacillus* (Escalante et al., 2008; Cerritos et al., 2011; Rebollar et al., 2012).

El trabajo previo llevado a cabo con el género *Bacillus* a partir de aislados recuperados en CCC permitió describir la diversidad genética y estructura de algunos grupos de dicho género. Usando la secuencia del gen 16S rRNA se pudo comprobar que los *Bacillus* aislados de CCC están filogenéticamente relacionados

con *Bacillus* marinos. Uno de estos grupos fue descrito como una especie endémica de esta región, distinta a especies de *Bacillus* previamente descritas en la literatura (Cerritos et al., 2008) y nombrado como *B. coahuilensis*. Junto con *B. coahuilensis* se identificaron otros 5 grupos del mismo género de los cuales, además de la secuencia 16s rRNA se obtuvieron las secuencias de dos genes housekeeping (*recA* y *gltX*). Dichas secuencias fueron utilizadas para determinar la estructura poblacional dentro de cada grupo identificado. Los resultados mostraron que al interior de cada una de las poblaciones la diversidad genética es baja y se pudo identificar una estructura poblacional altamente clonal para cada uno de estos grupos. Sin embargo, el número de muestra de cada población fue siempre menor a 15 aislados y se usaron solamente dos genes para el análisis multilocus (Cerritos, 2008). Es posible entonces que los resultados observados pudieran deberse a un efecto tanto del tamaño de muestra como al número limitado de marcadores utilizados. Finalmente, cabe mencionar aquí que a pesar de que este trabajo con *Bacillus* de CCC se observó una estructura poblacional clonal, un trabajo posterior identificó evidencias de dos eventos importantes de transferencia horizontal en el genoma de *B. coahuilensis* (adquisición de genes de rodopsina y sulfoquinovosa), que pudieron haber contribuido a la adaptación de dicha especie a las condiciones de oligotrofia de CCC (Alcaraz et al., 2008)(Alcaraz et al., 2008).

### **2.3 Justificación y presentación del proyecto**

A pesar de los avances hechos en los campos de la ecología y evolución molecular de procariontes en las últimas décadas, las principales preguntas que siguen siendo centrales para ambos campos están relacionadas con las causas detrás de la enorme diversidad de grupos procariontes que existe en nuestro planeta. Para poder responder estas preguntas es necesario considerar los mecanismos que han generado dicha diversidad. Entendiendo que la diversidad es resultado de un proceso de divergencia evolutiva, es importante entender los procesos de generación de diversidad a nivel poblacional, para poder también entender los mecanismos de divergencia genética de los linajes y la especiación en procariontes.

El estudio de poblaciones de bacterias se ha centrado en gran medida en el análisis de organismos patógenos, justamente para tratar de entender procesos como las epidemias y el surgimiento de cepas patógenas. Sin embargo, el estudio de poblaciones de bacterias de vida libre (o ambientales) ha sido mucho menor que el de las bacterias patógenas y apenas en los últimos años ha empezado a ampliarse. Como sabemos, la mayor parte de la diversidad procarionte se encuentra en ambientes naturales (DeLong & Pace, 2001), por lo que resulta indispensable seguir ampliando el conocimiento sobre poblaciones de bacterias de vida libre para poder entender los procesos generales que operan en su

adaptación y evolución y explicar así la gran diversidad de la vida procarionte sobre el planeta.

En el presente trabajo se exploró la genética de poblaciones y evolución molecular del género *Bacillus* a partir de una colección amplia de aislados recuperada de los cuatro principales sistemas acuáticos dentro del valle de CCC. El objetivo principal fue describir la diversidad genética de las poblaciones de *Bacillus* del valle y determinar los factores que han moldeado dicha diversidad.

Los primeros análisis genéticos llevados a cabo con la colección de *Bacillus* de CCC mostraron que se trata de una especie que no ha sido previamente descrita en la literatura. Así, en el capítulo tres de esta tesis se presenta el primer artículo derivado de este proyecto en el cual se llevó a cabo la determinación de las relaciones filogenéticas de este grupo de *Bacillus* aislado de CCC y se llevó a cabo la descripción polifásica de esta nueva especie, para la cual se propuso el nombre de *Bacillus cuatrocieneegasensis*.

En el capítulo cuatro se presentan los análisis de genética de poblaciones y evolución molecular llevados a cabo con el objetivo de determinar la importancia relativa de la mutación y la recombinación en la generación de la diversidad genética de las poblaciones de *Bacillus* estudiadas y la posible relación de la distribución de dicha diversidad con los factores ambientales de valle de CCC.

Finalmente, en el capítulo cinco se analiza con mayor profundidad la estructura poblacional de la muestra de *Bacillus* de CCC y se compara con otras poblaciones de los géneros *Exiguobacterium* y *Pseudomonas* también aisladas de este sitio. Lo anterior con el objetivo de descubrir la posible influencia de los factores históricos sobre los patrones de diversidad de poblaciones bacterianas que coexisten en un área geográfica dada.

**Capítulo 3. ARTÍCULO 1:** “*Bacillus cuatrociénegasensis* sp. nov., una nueva especie moderadamente halotolerante aislada del valle Cuatro Ciénegas, México”

### **3.1 Resumen**

En este trabajo se describió una nueva cepa bacteriana, P19<sup>T</sup>, con forma de bacilo, Gram positiva, formadora de endosporas y halotolerante, que fue aislada de una muestra de agua colectada en el valle de Cuatro Ciénegas, México. Para determinar su posición taxonómica, la cepa fue objeto de una caracterización polifásica. La cepa fue capaz de crecer a 10-45 °C (óptimo a 30-37 °C), en un rango de pH de 6-10 (óptimo a pH 8) y en 2.5-10 % (w/v) NaCl (óptimo a 7.5 %). Los análisis filogenéticos basados en la secuencia del gen 16S rRNA colocaron al aislado dentro del género *Bacillus* y además mostraron que la cepa P19<sup>T</sup> tiene la mayor similitud de secuencia con *Bacillus vietnamensis* (98%), *Bacillus aquimaris* (97.8%) y *Bacillus marisflavi* (97.5%). Aunque los porcentajes de similitud de las secuencias del 16S rRNA resultaron mayores a 97%, la similitud DNA-DNA entre la cepa P19<sup>T</sup> y las cepas tipo de las especies cercanamente relacionadas fueron menores que el 30%. Los ácidos grasos más abundantes fueron identificados como iso-C15:0 (36.8 %), anteiso-C15:0 (19.95 %) y anteiso-C17:0 (14.31 %). Los lípidos polares predominantes fueron identificados como fosfatidilglicerol, difosfatidilglicerol y fosfatidiletanolamina. MK-7 fue identificada como la

menaquinona predominante y el tipo de peptidoglicano presente en la pared celular se identificó como A1 $\alpha$  (L-Lys directo). El contenido de G+C genómico fue de 42.1 %. Con base en los datos fenotípicos, quimiotaxonómicos y filogenéticos, la cepa P19<sup>T</sup> puede ser distinguida de otros taxa cercanamente relacionados y, por lo tanto, consideramos que representa una nueva especie dentro del género *Bacillus*. El nombre propuesto para la cepa P19<sup>T</sup> (=CECT 7983T =LMG 26558T) es *Bacillus cuatrocienegasensis* sp. nov.



***Bacillus cuatrocienegasensis* sp. nov., a moderately halotolerant bacterium isolated from the Cuatro Ciénegas Basin, Mexico**

Morena Avitia<sup>1</sup>, René Cerritos<sup>2</sup>, Eria A. Rebollar<sup>1,3</sup>, Jesus Caballero-Mellado<sup>4†</sup>, Lourdes Martínez-Aguilar<sup>4</sup>, Alejandra Padua<sup>1</sup>, Jorge Molina-Torres<sup>5</sup>, Enrique Ramírez-Chávez<sup>5</sup>, and Valeria Souza<sup>1\*</sup>

<sup>1</sup>Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, México D.F. 04510, México.

<sup>2</sup>Departamento de Medicina Experimental, Facultad de Medicina, Universidad Nacional Autónoma de México, Apartado Postal 70-275, México D.F. 04510, México.

<sup>3</sup>Biology Department, James Madison University, Bioscience Building 2001, MSC 7801, Harrisonburg VA, USA.

<sup>4</sup>Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apartado Postal 565-A, Cuernavaca, Morelos, México.

<sup>5</sup>Departamento de Biotecnología y Bioquímica, Cinvestav Unidad Irapuato, Apartado Postal 629, Irapuato, Gto. 36821, México.

\*Corresponding author:

Valeria Souza Saldívar

Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Circuito exterior s/n anexo al Jardín Botánico Exterior. Apartado Postal 70-275, México D.F. 04510, México.

Tel. (+52-55 56229006) Fax: (+52-55 56228995)

E-mail: [souza@servidor.unam.mx](mailto:souza@servidor.unam.mx)

## **ABSTRACT**

A rod shaped, endospore forming, Gram-stain positive and halotolerant novel bacterial strain, designated as P19<sup>T</sup>, was isolated from a water sample collected from the Cuatro Ciénegas Basin (CCB), Mexico. To determine its taxonomic position, the strain was subjected to polyphasic characterization. The strain P19<sup>T</sup> is able to grow at 10-45 °C (optimum at 30-37 °C), in a pH range of 6-10 (optimum at pH 8) and in 2.5-10 % (w/v) NaCl (optimum at 7.5 %). Phylogenetic analysis based on 16S rRNA gene sequences placed the strain P19<sup>T</sup> within the *Bacillus* genus and revealed that it has the highest sequence similarity with *Bacillus vietnamensis* (98 %), *Bacillus aquimaris* (97.8 %) and *Bacillus marisflavi* (97.5 %). Although the 16S rRNA sequence similarities were above 97%, DNA-DNA relatedness between strain P19<sup>T</sup> and the type strains of closely related species was lower than 30 %. The major fatty acids were identified as iso-C<sub>15:0</sub> (36.8 %), anteiso-C<sub>15:0</sub> (19.95 %) and anteiso-C<sub>17:0</sub> (14.31 %). The predominant polar lipids were identified as phosphatidylglycerol, diphosphatidylglycerol and phosphatidylethanolamine. MK-7 was identified as the predominant menaquinone and the cell wall peptidoglycan type was identified as A1α (L-Lys direct). The genomic DNA G+C content was determined to

be 42.1%. On the basis of phenotypic, chemotaxonomic and phylogenetic data, strain P19<sup>T</sup> can be distinguished from its closely related taxa and thus represents a novel species within the genus *Bacillus*. The name *Bacillus cuatrocieneegasensis* sp. nov., is proposed, with the type strain P19<sup>T</sup> (=CECT 7983<sup>T</sup> =LMG 26558<sup>T</sup>)

## INTRODUCTION

The genus *Bacillus* (Gordon, 1981) currently consists of almost 300 recognized species (<http://www.bacterio.net/b/bacillus.html>) which have a wide range of physiological life styles and are ubiquitous in nature (Claus & Berkeley, 1986; Slepecky & Hemphill, 2006). The genus *Bacillus* is well known for being capable of forming resistant endospores that contribute to their ubiquity in natural environments. In recent years, several halotolerant or moderately halophilic, endospore forming, rod-shaped bacteria of the genus *Bacillus* have been described (Ventosa, Nieto & Oren, 1998; Yoon, 2003; Noguchi et al., 2004; Cerritos et al., 2008; Singh et al., 2014). Most of these former species have been isolated from marine environments, however, little is known about species inhabiting non-marine, high salinity aquatic environments (Lim et al., 2006; Cerritos et al., 2008) In order to describe the halotolerant bacterial diversity of high salinity non-marine aquatic environments, samples were collected from the Cuatro Ciénegas Basin (CCB), Mexico (Rebollar et al., 2012; Avitia et al., 2014). During these studies, a strain designated as P19<sup>T</sup> was isolated. The aim of the present study is to determine the taxonomic status of strain P19<sup>T</sup> using a polyphasic approach (including phenotypic properties, chemotaxonomic, phylogenetic analyses, and genomic relatedness), to meet the minimal standards for the description of novel taxa (Logan et al., 2009) On the basis of the results, strain P19<sup>T</sup> is considered to represent a novel species of the genus *Bacillus*, for which the name *Bacillus cuatrociénegasensis* sp. nov. is proposed, since this strain was isolated from the CCB.

## MATERIALS AND METHODS

### Isolation and cultivation

A collection of 151 isolates were obtained from surface water and sediment from four aquatic systems within the CCB: Churince, Los Hundidos, Rio Mesquites, and Pozas Azules (Avitia et al., 2014). In particular, isolate P19<sup>T</sup> was recovered from a desiccation pond near the Rio Mesquites (26°55.133' N, 102° 6.19' W). All isolates were obtained using previously described methods (Rebollar et al., 2012). Briefly, water and sediment samples were diluted at 1:10; 1:100; 1:1,000; and 1:10,000 using saline solution (1% NaCl). Subsequently, 200 µl of each dilution was plated in marine agar 2216 (MA, Difco) and incubated for 24 h at 37°C. Colonies were purified by single-colony isolation. All isolates were stored at -80 °C in 20% (w/v) glycerol. From this large collection, five strains (P19<sup>T</sup>, PM43e, PE8e, A81 and 105M) were chosen for being representative of all sources of isolation. The five strains were subjected to morphological, phenotypic and genetic characterization. All strains were routinely grown on MA at 37 °C for all the characterization experiments, except when stated otherwise. The type strains of closely related taxa *Bacillus aquimaris* TF-12<sup>T</sup>, *Bacillus marisflavi* TF-11<sup>T</sup> and *Bacillus vietnamensis* NRIC 0531<sup>T</sup> were used as reference strains in all of the experiments (except where mentioned otherwise) under the same laboratory conditions.

Repetitive sequence-based PCR profiles with BOX primers as well as SDS-PAGE assays of whole-cell proteins of the five isolates and reference strains were performed (see

Supplementary material for details). Among these five strains, the strain P19<sup>T</sup> was chosen as the type strain for the novel species described in this study and it was subjected to a more detailed characterization.

### **Morphology and phenotypic characterization**

The colony morphology of the strains was observed after growth on MA for 72 hours. The cell morphology and sporulation process of the selected strains were examined using phase contrast microscopy. Samples were fixed with formaldehyde and stained with 1% (w/v) crystal violet. Preparations were visualized with a Nikon microscope (model microphotFXA) using bright field and differential interference contrast (DIC). Images were captured and processed using a Nikon digital camera (model DFX100) and acquisition software (ACT-1, Nikon). A more detailed characterization of cell morphology of the P19<sup>T</sup> strain was achieved by electron microscopy. A sample of an overnight culture was taken and washed with 10 mM Tris-HCl, pH 7.0. Cell samples were stained with 1% PTA, pH 7.0, and observed using a JEM-1200EXII electron microscope (JEOL, Tokyo, Japan). Micrographs were taken at an accelerating voltage of 100 kV.

Motility assays were performed on P19<sup>T</sup> strain. For motility assays, a medium consisting of tryptone 1.0%, MgSO<sub>4</sub> 35mM, CaCl<sub>2</sub> 7mM, KCl 7mM and NaCl 120mM, also known as tryptone-based seawater (TBSW) (O'Shea et al., 2005), was used with different agar concentrations. 5- $\mu$ l samples of a stationary-phase culture were inoculated on Petri dishes containing two soft agar concentrations (0.3% and 0.6%) and incubated at 37

°C for 10 or 48h. Swimming and swarming ability was recorded as the ability of bacteria to move away from the inoculation point. Images of the soft agar plates were taken using a Canon Power shot A700 zoom digital camera.

The biochemical profile of each strain was characterized by single carbon source assimilation tests. The tests were performed on MA, replacing yeast extract and peptone with specific carbon sources. Acid production from carbohydrates was determined as previously described by Leifson (Leifson, 1963).

Temperature and NaCl (%w/v) tolerance were evaluated by culturing each strain on MB at 10, 15, 20, 28, 37, 40 and 45°C and 0, 2.5, 5, 7.5, 10, 12.5, 15, 17.5 and 20 % NaCl (w/v) respectively. Microtiter plates were inoculated with 5ml of each strain and the growth was evaluated by spectrophotometry (at 600 nm) during 24 h. pH tolerance was assessed by measuring the diameter of colonies after 24 h of incubation at 37 °C on MA with pH values between 4 and 12. pH was modified by adding hydrochloric acid (HCl) or sodium hydroxide (NaOH) to the MA.

### **16S rRNA gene sequencing and phylogenetic analyses**

For each strain, DNA was extracted using a DNeasy tissue Kit (Qiagen, Valencia CA) following the manufacturer's instructions. The 16S rRNA gene was amplified using the eubacterial universal primers 27F and 1492R, according to the conditions described by Lane (Lane, 1991). PCR amplification was performed according to Cerritos *et al.* (2008).

All PCR products were sequenced at the University of Washington's High Throughput Sequencing Solutions Laboratory. A nearly full length (1350 bp) 16S rRNA gene was obtained and used for further phylogenetic analyses.

The identity of each isolate was determined by comparing 16S rRNA sequences with the Ribosomal Database Project including only type strains. Sequences were aligned using ClustalW program (Thompson, Higgins & Gibson, 1994). Phylogenetic reconstructions were performed using two algorithms: neighbor-joining (NJ) and maximum likelihood (ML) as implemented in MEGA6 (Tamura et al., 2013). The stability of the phylogenetic relationship was assessed with bootstrap analysis, by performing 10,000 re-sampling for the tree topology of the neighbor-joining data.

#### **DNA base composition, DNA-DNA hybridization**

To confirm taxonomic relationships at the species level, DNA base composition and DNA-DNA hybridization were performed by the LMG consortium. DNA base compositions were determined using the HPLC technique (Mesbah, Premachandran & Whitman, 1989). Hybridizations were performed in the presence of 50% formamide at 39 °C according to a modification of the method described by Ezaki et al. (Ezaki, Hashimoto & Yabuuchi, 1989). The DNA-DNA relatedness percentages reported are the means of at least five hybridizations. With this technique, the average standard deviation is 14 units (Goris et al., 1998). Deviations of up to 20-25 units are still acceptable.



## Chemotaxonomic analyses

The quantitative analysis of whole-cell fatty acids of the P19<sup>T</sup> strain was carried out as previously described by Teska *et al.* (Teska, Coyne & Ezzell, 1998). Fatty Acid Methyl Esters (FAMES) were analyzed using a gas chromatograph, coupled to an electron impact quadrupole mass spectrometer (GC/EIMS\_Agilent Technologies 7890 A, coupled to a Hewlett Packard 5973 MSD) with a 7683 auto sampler (Agilent Technologies). A fused silica capillary column HP-INNOWax (30 m x 250 µm x 0.25 µm) was used for the separation of FAMES. Aliquots of 1 µl of each sample were injected with a split ratio of 1:10 at 250 °C. The initial oven temperature (50 °C) was initially maintained for 3 min, then increased at a rate of 10 °C/min to 250 °C and held for 15 min. Helium was used as carrier gas at a flow rate of 2 mL/min. FAME standards and GC/EIMS were used to identify peaks with the aid of the AMDIS deconvolution program and NIST 2008 Mass Spectra Database.

For the polar lipid determination, strains were grown in MB for 20 h at 37 °C, and the polar lipids were extracted as described by Bligh and Dyer (Bligh & Dyer, 1959). The chloroform phase was used for develop on TLC plates (high performance TLC aluminum sheets, silica gel 60; Merck) by two-dimensional separation (Geiger et al., 1999). Total polar lipids were observed either by iodine vapor (Sims & Larose, 1962) or by spraying with ANS reagent (8-anilino-1-naphthalenesulfonic acid) (Zbierzak, Dormann & Holzl, 2011). Polar lipids were identified on the basis of their migration and identified by Rf comparisons with standards for phosphatidic acid, phosphatidylserine,

phosphatidylethanolamine, phosphatidylglycerol and diphosphatidylglycerol (Sigma) in combination with specific staining. Lipids containing amino groups, phosphorus, quaternary nitrogen and glycolipids were identified with ninhydrin (Kates, 1986), Phospray-SUPELCO 33047, dragendorff reagent (Kates, 1986) and a-naphtol-sulfuric acid (Tindall et al., 2007) stain respectively.

Analyses of respiratory quinones and peptidoglycan of the cell wall content of strain P19<sup>T</sup> were carried out by the DSMZ Identification Service according to previously published protocols (Schumann, 2011). The cell biomass for these analyses was obtained by cultivation on MB at 37 °C for 48 h.

## RESULTS AND DISCUSSION

From a large collection of 151 bacterial isolates from the CCB that had previously been identified as *Bacillus* (Avitia et al., 2014), five strains (representative of all the sites of isolation) were initially selected to describe the novel species of *Bacillus* presented in this study. The almost complete 16S rRNA gene sequences (1,350 bp) of the five strains were obtained (Gen-Bank accession numbers JN639059, JN639107, JN639159 and JN634520 for strains P19<sup>T</sup>, PMs43e, PE8e, A81 and 105M respectively) and shown to have more than 99.8% of sequence similarity among them. In order to meet the standards for the description of novel taxa suggested by Logan et al (Logan et al., 2009), the strains were subjected to further phenotypic and genetic characterization. The results of such analyses showed that the five strains have very different genetic and phenotypic profiles with respect to the reference strains but also among them (Figures S1 and S2). The strain P19<sup>T</sup> was chosen as the type strains for being representative of the sampling site from which most of the isolates from the large collection were recovered and therefore it was subjected to a more detailed characterization.

### **Morphology and phenotypic characterization**

Colonies of strain P19<sup>T</sup> were observed to be pale orange on MA. Cells were observed to be Gram-positive rods, 1.0 x 1.3-2.0 µm wide, occurring as single cells or pairs, with central endospores present (Figure S3). Cells showed to have flagellum (Figure S4) and proved to be motile by swimming (Figure S5). Strain P19<sup>T</sup> was found to grow at 10-45 °C

temperature, pH 6-10 and with 2.5-10 % (w/v) NaCl, and optimum growth was observed at 30-37 °C, pH 8 and 7.5 % (w/v) NaCl concentration. The detailed physiological and biochemical characteristics of strain P19<sup>T</sup> are presented in Table 1 and summarized in the species description. In spite of sharing many common phenotypic characteristics, a number of observations differentiated strain P19<sup>T</sup> from the closely related strains of the genus *Bacillus*. For example, strain P19<sup>T</sup> was positive for acid production from L-Arabinose and L-Malic Acid as opposed to *B.aquimaris*, *B.marisflavi* and *B.vietnamensis*, which proved to be negative for this feature (Table 1).

#### **Phylogenetic analysis, DNA-DNA relatedness and DNA base composition**

The comparison among the 16S rRNA sequences of strain P19<sup>T</sup> and the closely related type strains of validly named species on the Ribosomal Data Base Project showed the highest similarity with *B. vietnamensis* NRIC 0531<sup>T</sup> (AB099708) with 98%, 97.8 % with *B. aquimaris* TF-12<sup>T</sup> (AF483625) and 97.5 % with *B. marisflavi* TF-11<sup>T</sup> (AF483624). A neighbor-joining phylogenetic tree (Fig. 1) showed that strain P19<sup>T</sup> is within a cluster comprised of *B. aquimaris*, *B. marisflavi*, *B. vietnamensis*, *Bacillus coahuilensis* and *Bacillus seohaenensis* with a relatively low bootstrap value (53 %). A similar tree topology of the strain P19<sup>T</sup> with the closely related species was observed in analyses performed by the ML algorithm (Avitia et al, 2014 and data not shown). The species *B. coahuilensis* and *B. seohaenensis* 16S rRNA sequences showed less than 97% similarity values with strain P19<sup>T</sup>, thus they were not included in the rest of the analyses.

The DNA-DNA relatedness was 13 (2) % with *B. aquimaris* TF-12T, 9 (11) % with *B. marisflavi* TF-11T and 9 (9) % with *B. vietnamensis* NRIC 0531T. All these values are less than the 70 % threshold recommended to assign the strain to a novel species (Stackebrandt & Goebel, 1994).

The DNA G+C content of strain P19<sup>T</sup> is 42.1 (Table 1) which is consistent with the taxonomic position of the strain within the genus *Bacillus* (Shida, 1997).

### **Chemotaxonomic analyses**

The cellular fatty acid profile of strain P19<sup>T</sup> included iso- C<sub>15:0</sub> (36.8 %) and anteiso- C<sub>15:0</sub> (19.95 %) as the major fatty acids (Table 2). This is in agreement with the profiles of the closely related type strains (Yoon, 2003; Lee et al., 2006) Importantly, strain P19<sup>T</sup> contained some fatty acids such as iso-C<sub>17:1</sub> ω10c that are not present or are present in small amounts in the profiles of the closely related species' type strains. Nevertheless, the cellular fatty acid profile obtained for strain P19<sup>T</sup> is similar to those of members of the genus *Bacillus* (Claus & Berkeley, 1986; Shida, 1997).

Strain P19<sup>T</sup> shared a similar polar lipids profile with *B. aquimaris* TF-12<sup>T</sup>, *B. marisflavi* TF-11<sup>T</sup> and *B. vietnamensis* NRIC 0531<sup>T</sup> which consisted predominantly of phosphatidylglycerol (PG), diphosphatidylglycerol (DPG) and phosphatidylethanol-amine (PE) (Figure S6). However, the presence and patterns of one unknown phosphorous-containing lipid (UPL) in the case of *B. aquimaris* and *B. vietnamensis* or

phosphatidylserine (PS) in *B. vietnamensis* or two UPL in the case of *B. marisflavi* could differentiate strain P19<sup>T</sup> from the rest of the type strains.

Strain P19<sup>T</sup> was found to contain MK-7 as the major isoprenoid quinone which is in agreement with the core characteristic described for the genus *Bacillus* (Kämpfer et al., 2006). The peptidoglycan type of P19<sup>T</sup> was found to be A1 $\alpha$  (L-Lys direct) (Schleifer & Kandler, 1972) with  $\alpha$ -carboxyl group of D-Glu partially substituted by Gly (type A11.gly according to the DSMZ catalogue of strains [www.peptidoglycan-types.info](http://www.peptidoglycan-types.info)). The rest of the closely related *Bacillus* species contain meso-diaminopimelic acid as the diagnostic amino acid of the cell wall peptidoglycan (Table 1). Therefore, strain P19<sup>T</sup> could be distinguished from closely related species based on cell wall peptidoglycan type.

According to the recommendations of Kämpfer et al. (2006) for the inclusion of any species in the genus *Bacillus*, strain P19<sup>T</sup> shares almost all of the core characteristics (presence of major menaquinone as MK-7; dominance of iso and anteiso fatty acids and profile of polar lipids mainly comprised of phosphatidylglycerol, diphosphatidylglycerol and phosphatidylethanolamine) that are present in the genus *Bacillus*. However, strain P19<sup>T</sup> differs from the type species of closely related *Bacillus* in some phenotypic and chemotaxonomic characteristics (Table 1).

On the basis of physiological, phylogenetic, chemotaxonomic and genomic characteristics, we propose that strain P19<sup>T</sup> should be assigned to the genus *Bacillus* within a novel species for which the name *Bacillus cuatrocienezasensis* sp. nov. is proposed.

### **Description of *Bacillus cuatrocienegasensis* sp. nov (P19<sup>T</sup>)**

*Bacillus cuatrocienegasensis* (cu.a.tro.cie.ne.gas.en'sis. N.L. fem. adj. *cuatrocienegasensis* pertaining to the Cuatro Cienegas Basin, the location of the aquatic environment from which the strain was isolated).

Cells are Gram-stain positive, rod-shaped (1.3-2  $\mu\text{m}$  long and 1  $\mu\text{m}$  wide), occurring as single cells or pairs; with central endospores present (Supplementary Fig. ). Motile by means of a single flagellum (Supplementary Fig. 2). Colonies are smooth, circular to slightly irregular, slightly raised and pale orange in color and 1-2 mm in diameter after 72 h at 37 °C on marine agar. Cells can grow at a temperature range of 10-45 °C (optimum growth occurs at 30-37 °C). The pH range for growth is 6-10 (optimum at pH 8). Growth occurs in the presence of 2.5-10 % (w/v) NaCl (optimum at 7.5%). Positive for catalase but negative for oxidase. Can grow on glycerol, L-arabinose, D-mannose, glucose, L-malic acid, D-maltose, sucrose, D-trehalose and dulcitol when supplied as the sole source of carbon and energy but cannot growth only on D-xylose, D-mannitol, D-sorbitol, salicin, D-raffinose, starch fructose or lactose. Positive for acid production from glycerol, L-arabinose, D-mannose, D-maltose, sucrose, L-malic acid, D-trehalose and dulcitol, but negative for D-xylose, D-mannitol, D-sorbitol, salicin, D-raffinose, starch, glucose, fructose and lactose. The major cellular fatty acids are iso-C<sub>15:0</sub> (36.8 %), anteiso-C<sub>15:0</sub> (19.95 %) and anteiso-C<sub>17:0</sub> (14.31 %), followed by iso-C<sub>17:0</sub> (6.69 %), iso-C<sub>16:0</sub> (6.35 %), iso-C<sub>17:1</sub>  $\omega$ 10c (5.15 %), C<sub>16:1</sub>  $\omega$ 7c (3.72 %), C<sub>16:0</sub> (2.27 %), C<sub>16:1</sub>  $\omega$ 11c (1.66 %), iso-C<sub>14:0</sub> (1.39 %), C<sub>14:0</sub>, C<sub>18:0</sub> and iso-C<sub>13:0</sub> (<1 % each one). The major polar lipids consist of phosphatidylglycerol, diphosphatidylglycerol, phosphatidylethanolamine and a positive  $\alpha$ -naphthol lipid. The

predominant respiratory quinone is MK-7. The cell-wall peptidoglycan type is A1 $\alpha$  (L-Lys direct). The DNA G+C content of the type strain is 42.1 mol % (as analyzed by using HPLC).

The type strain P19<sup>T</sup> (=CECT 7983<sup>T</sup> =LMG 26558<sup>T</sup>), was isolated from a desiccation pond near the Rio Mesquites (26°55.133' N, 102° 6.19' W), in CCB, México.

The GenBank accession number for the 16S rRNA gene sequence of strain P19<sup>T</sup> ((=CECT 7983<sup>T</sup> =LMG 26558<sup>T</sup>) is JN639068.

## ACKNOWLEDGMENTS

We thank Dr. Santiago Ramírez-Barahona and Dr. Christine Rooks for their constructive reviews of the manuscript. We especially thank Diana Sahonero for polar lipid determination. We also thank Javier de la Mora and Yael González for assistance with electron microscopy and motility assays as well as Marco Gudiño for technical support with phase contrast microscopy and photomicrographs. We also thank Dr. Laura Espinosa and Dr. Erika Aguirre for technical support during the development of the project. We thank Rodrigo González Chauvet for support during field work. Special thanks to Dr. Janet L. Siefert for providing useful *Bacillus* type strains. We thank Dr. Julia Pozas for her advice in the species name description. This paper constitutes part of the doctoral research of the first author, who thanks the Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and acknowledges the scholarship (no. 210335) provided by the Consejo Nacional de Ciencia y Tecnología (CONACyT). The project was supported by grants 2006-C01-23459-CONACyT-FONSEC SEMARNAT, 57507-SEP-CONACyT and OL039-WWF-Alianza Carlos Slim to V. S.



## REFERENCES

- Avitia M, Escalante AE, Rebollar E a, Moreno-letelier A, Eguiarte LE, Souza V. 2014. Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. *Peer J*:1–17.
- Bligh E, Dyer W. 1959. A rapid method for total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*:911–917.
- Cerritos R, Vinuesa P, Eguiarte LE, Herrera-Estrella L, Alcaraz-Peraza LD, Arvizu-Gómez JL, Olmedo G, Ramirez E, Siefert JL, Souza V. 2008. *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico. *International journal of systematic and evolutionary microbiology* 58:919–23.
- Claus DR, Berkeley CW. 1986. The genus *Bacillus*. In: *Bergey's manual of systematic bacteriology*. 1105–1139.
- Ezaki T, Hashimoto Y, Yabuuchi E. 1989. Fluorometric Deoxyribonucleic Acid-Deoxyribonucleic Acid Hybridization in Microdilution Wells as an Alternative to Membrane Filter Hybridization in which Radioisotopes Are Used To Determine Genetic Relatedness among Bacterial Strains. :224–229.
- Geiger O, Rohrs V, Weissenmayer B, Finan T, Thomas-Oates J. 1999. The regulator gene *phoB* mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl- N , N , N -trimethylhomoserine in *Rhizobium* ( *Sinorhizobium* ) *meliloti*. *Molecular microbiology*:63–67.
- Gordon R. 1981. *One hundred and seven years of the genus Bacillus*. London: Academic Press.
- Goris J, Suzuki K, De Vos P, Nakase T, Kersters K. 1998. Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. *Can J Microbiol*:1148–1153.
- Kämpfer P, Rosselló-Mora R, Falsen E, Busse H-J, Tindall BJ. 2006. *Cohnella thermotolerans* gen. nov., sp. nov., and classification of “*Paenibacillus hongkongensis*” as *Cohnella hongkongensis* sp. nov. *International journal of systematic and evolutionary microbiology* 56:781–6.
- Kates M. 1986. *Techniques of lipidology. Isolation, analysis and identification of lipids*. Amsterdam, New York, Oxford: Elsevier.

- Lane DJ. 1991. 16S/23S rDNA sequencing. Nucleic acid techniques. In: Stackebrandt E, Goodfellow M eds. *Bacterial Systematics*. New York, USA: John Wiley & Sons, 115–175.
- Lee J, Lim J, Park D, Jeon CO, Li W, Kim C. 2006. *Bacillus seohaeanensis* sp. nov., a halotolerant bacterium that contains L -lysine in its cell wall. :1893–1898.
- Leifson E. 1963. Determination of Carbohydrate Metabolism of. *Growth (Lakeland)* 85:5–7.
- Lim J-M, Jeon CO, Lee S-M, Lee J-C, Xu L-H, Jiang C-L, Kim C-J. 2006. *Bacillus salarius* sp. nov., a halophilic, spore-forming bacterium isolated from a salt lake in China. *International journal of systematic and evolutionary microbiology* 56:373–7.
- Logan N a, Berge O, Bishop a H, Busse H-J, De Vos P, Fritze D, Heyndrickx M, Kämpfer P, Rabinovitch L, Salkinoja-Salonen MS, Seldin L, Ventosa a. 2009. Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *International journal of systematic and evolutionary microbiology* 59:2114–21.
- Mesbah M, Premachandran U, Whitman WB. 1989. Precise Measurement of the G+C Content of Deoxyribonucleic Acid by High-Performance Liquid Chromatography. *International Journal of Systematic Bacteriology* 39:159–167.
- Noguchi H, Uchino M, Shida O, Takano K, Nakamura LK, Komagata K. 2004. *Bacillus vietnamensis* sp. nov., a moderately halotolerant, aerobic, endospore-forming bacterium isolated from Vietnamese fish sauce. :2117–2120.
- O’Shea TM, DeLoney-Marino CR, Shibata S, Aizawa SI, Wolfe AJ, Visick KL. 2005. Magnesium promotes flagellation of *Vibrio fischeri*. *Journal of Bacteriology* 187:2058–2065.
- Rebollar EA, Avitia M, Eguiarte LE, 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 Ciénegas Basin. *Environmental Microbiology* 14:2323–2333.
- Schleifer KH, Kandler O. 1972. Peptidoglycan. Types of Bacterial Cell Walls and their Taxonomic Implications. 36:407–477.
- Schumann P. 2011. Peptidoglycan structure. *Methods in microbiology* 38:101–129.
- Shida O. 1997. *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*,. :289–298.

- Sims R, Larose J. 1962. The use of iodine vapor as a general detecting agent in the thin layer chromatography of lipids. *Journal of the American Oil Chemists Society* 39:232.
- Singh NK, Kaur C, Kumar N, Velmurugan S, Citarasu T, Mayilraj S. 2014. *Bacillus aequororis* sp. nov., Isolated From Marine Sediment. *Current microbiology*.
- Slepecky RA, Hemphill HE. 2006. *The Genus Bacillus — Nonmedical*.
- Stackebrandt E, Goebel BM. 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology* 44:846–849.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30:2725–9.
- Teska J, Coyne S, Ezzell J. 1998. Identification of *Bacillus anthracis* Using Gas Chromatographic Analysis of Cellular Fatty Acids and a Commercially Available Database Application. *Agilent Technologies*:1–6.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673–4680.
- Tindall B, Sikorski J, Mibert R, Kreig N. 2007. Phenotypic characterization and the principles of comparative systematics. In: Reddy C, Beveridge T, Breznak J, Marzluf G, Schmidt T, Snyder L eds. *Methods for general and molecular microbiology*. Whashington, DC: ASM press, 330–393.
- Ventosa A, Nieto JJ, Oren A. 1998. Biology of Moderately Halophilic Aerobic Bacteria *Biology of Moderately Halophilic Aerobic Bacteria*. 62.
- Yoon J-H. 2003. *Bacillus marisflavi* sp. nov. and *Bacillus aquimaris* sp. nov., isolated from sea water of a tidal flat of the Yellow Sea in Korea. *International Journal of Systematic and Evolutionary Microbiology* 53:1297–1303.
- Zbierzak A, Dormann P, Holz G. 2011. Analysis of lipid content and quality in arabidopsis plastids. In: *Methods in molecular biology*. 411–426.

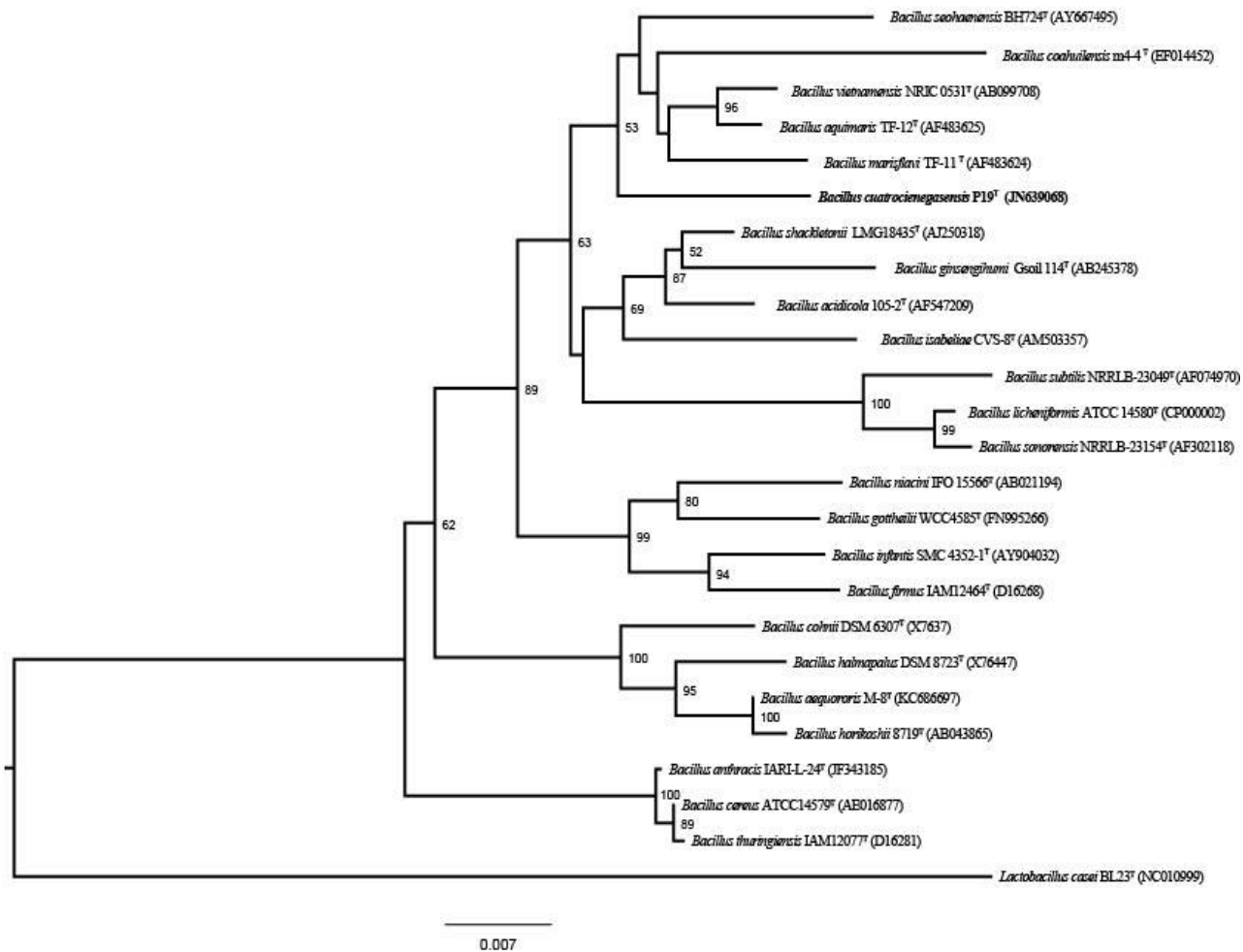
## Tables and figures

**Table 1.** Differentiating phenotypic and chemotaxonomic characteristics of strain P19<sup>T</sup> and the type strain of closely related species

|                                  | <i>Bacillus<br/>cuatrocieneegasensis</i><br>P19 <sup>T</sup>                                  | <i>Bacillus aquimari</i> s<br>TF-12 <sup>T</sup>   | <i>Bacillus marisflavi</i> TF-<br>11 <sup>T</sup>  | <i>Bacillus<br/>vietnamensis</i><br>NRIC 0531 <sup>T</sup>                                     |
|----------------------------------|---|--|--|--|
| <b>Gram staining</b>             | Positive  | Variable*  | Positive (or variable)*  | Positive*  |
| <b>Flagellum type</b>            | Single polar  | Peritrichous*  | Single polar*  | Peritrichous*  |
| <b>Spore position</b>            | Central   | Central*   | Subterminal or central*  | Central*   |
| <b>Colony colour</b>             | Pale orange   | Pale orange-yellow*  | Pale yellow*   | Orange*  |
| <b>Growth at:</b>                |   |  |  |  |
| Temperature (°C) range (optimum) | 10-45 (30-37)   | 10-45 (28-37)  | 15-47 (30-37)  | 15-45 (28-37)  |
| pH range (optimum)               | 6-10 (8)  | 6-11 (10)  | 5-11 (9)   | 5-8 (6)  |
| NaCl range (% w/v) (optimum)     | 2.5-10 (7.5)  | 0-10 (7.5)   | 0-10 (5)   | 0-15 (5)   |
| <b>Oxidase reaction</b>          | -   | -  | -  | +  |
| <b>Growth on<sup>a</sup>:</b>    |   |  |  |  |
| D-Xylose                         | -   | +  | -  | +  |
| D-Mannitol                       | -   | +  | +  | +  |
| D-Sorbitol                       | -   | +  | +  | +  |
| Salicin                          | -   | +  | +  | +  |
| D-Maltose                        | +   | +  | -  | +  |
| Sucrose                          | +   | +  | -  | +  |
| D-Rafinose                       | -   | +  | +  | +  |
| Starch                           | -   | +  | +  | +  |
| Fructose                         | -   | +  | +  | +  |
| D-Trehalose                      | +   | +  | -  | +  |
| Lactose                          | -   | +  | +  | +  |
| <b>Acid production from:</b>     |   |  |  |  |
| Glycerol                         | +   | +  | -  | +  |
| L-Arabinose                      | +   | -  | -  | -  |
| D-Mannose                        | +   | -  | +  | +  |
| D-Mannitol                       | -   | +  | +  | +  |
| D-Maltose                        | +   | +  | -  | +  |
| Sucrose                          | +   | +  | -  | +  |
| D-Rafinose                       | -   | -  | -  | +  |
| Glucose                          | -   | +  | -  | +  |
| Fructose                         | -   | -  | -  | +  |
| L-Malic Acid                     | +   | -  | -  | -  |
| D-Trehalose                      | +   | +  | -  | +  |
| Dulcitol                         | +   | -  | -  | -  |
| <b>G+C content</b>               | 42.1  | 38*  | 49*  | 43*  |
| <b>Major fatty acids</b>         | i-C <sub>15:0</sub> (36.8%)<br>ai-C <sub>15:0</sub> (19.95%)<br>ai-C <sub>17:0</sub> (14.31%) | i-C <sub>15:0</sub> (46.9%)*<br>ai-C <sub>15:0</sub> (22.2%)*<br>C <sub>16:1</sub> ω7c a (8.4%)* | ai-C <sub>15:0</sub> (27.7%)*<br>i-C <sub>15:0</sub> (22.9%)*<br>i-C <sub>14:0</sub> (9.1%)* | ai-C <sub>15:0</sub> (48.3%)*<br>i-C <sub>15:0</sub> (16.2%)*<br>ai-C <sub>17:0</sub> (13.6%)* |
| <b>Cell wall type</b>            | A1α (L-Lys direct)  | A1γ (m-DAP)*   | A1γ (m-DAP)*   | A1γ (m-DAP)*   |

All strains were positive for catalase reaction. All strains could grow on glycerol, L-arabinose, D-mannose, glucose, L-malic acid and dulcitol when supplied as the sole source of carbon and energy. All strains were negative for acid production from D-xylose, D-sorbitol, salicin, starch and lactose. <sup>a</sup>: When supplied as the sole source of carbon and energy. + Positive; - negative. i: iso; ai: anteiso. All data are from this study except if mentioned otherwise.

\* Data from Yoon et al. (2003) and Noguchi et al. (2004)



**Figure 1.** Neighbour-joining phylogenetic tree showing the relationships of strain P19<sup>T</sup> and closely related type strains of genus *Bacillus* inferred from sequences of the 16S rRNA gene. The tree is rooted by using *Lactobacillus casei* BL23<sup>T</sup> (NC010999) as an out-group. The tree was generated using the MEGA 6 software package (Tamura et al. 2013) based on a comparison of 1,350 nucleotides. Bootstrap values (only >50%), expressed as percentage of 10, 000 replications, are given at the branching points. The bar represents 0.007 substitutions per nucleotide position. The maximum likelihood method was also applied to the same sequences and gave approximately the same results, with the same clustering for the novel species. The accession number of each type strain is shown in parentheses

**Table 2.** Cellular fatty acid profiles (%) of strain P19<sup>T</sup> in comparison with closely related taxa.

| Fatty acids                   | <i>Bacillus cuatrocienegasensis</i><br>P19 <sup>T</sup> | <i>Bacillus aquimaris</i><br>TF-12 <sup>T</sup> | <i>Bacillus marisflavi</i><br>TF-11 <sup>T</sup> | <i>Bacillus vietnamensis</i><br>NRIC 0531 <sup>T</sup> |
|-------------------------------|---|---|--|--|
| iso-C <sub>13:0</sub>         | 0.09  | 0.4   | -  | -  |
| iso-C <sub>14:0</sub>         | 1.39  | 6.5   | 9.1  | 3.7  |
| iso-C <sub>14:0</sub> 3-OH    | -   | -   | 2  | -  |
| C <sub>14:0</sub>             | 0.77  | 0.9   | 1.7  | 1.3  |
| iso-C <sub>15:0</sub>         | 36.8  | 46.9  | 22.9   | 16.2   |
| iso-C <sub>15:0</sub> 3-OH    | -   | -   | 6.1  | -  |
| C <sub>15:1</sub>             | -   | 0.7   | -  | -  |
| anteiso-C <sub>15:0</sub>     | 19.95   | 22.2  | 27.7   | 48.3   |
| C <sub>15:0</sub>             | 0.63  | 0.5   | 1.1  | 1  |
| C <sub>15:0</sub> 2-OH        | -   | -   | 5.5  | -  |
| iso-C <sub>16:0</sub>         | 6.35  | 4.2   | 7.7  | 11.2   |
| C <sub>16:0</sub>             | 2.27  | 0.5   | 1  | 2.8  |
| iso-C <sub>16:0</sub> 3-OH    | -   | -   | 1.5  | -  |
| C <sub>16:1</sub> ω7c alcohol | -   | 8.4   | 2.5  | -  |
| C <sub>16:1</sub> ω7c         | 3.72  | -   | -  | -  |
| C <sub>16:1</sub> ω11c        | 1.66  | 1.3   | -  | -  |
| iso-C <sub>17:0</sub>         | 6.69  | 0.8   | 1.2  | 1.3  |
| iso-C <sub>17:1</sub> ω10c    | 5.15  | 1.8   | -  | -  |
| anteiso-C <sub>17:0</sub>     | 14.31   | 2.3   | 7.1  | 13.6   |
| C <sub>17:0</sub> 2-OH        | -   | -   | 1.2  | -  |
| C <sub>18:0</sub>             | 0.22  | -   | -  | -  |
| Summed feature 4*             | -   | 2.9   | 1.8  | -  |

Values are percentages of total fatty acids. – not detected.

Data for strain P19<sup>T</sup> were obtained in this study. Data for reference strains were obtained from Yoon et al. (2003) and Noguchi et al. (2004).

\* Summed features represent groups of two or three fatty acids that cannot be separated by GC. Summed feature 4 contains one or more of iso-C<sub>17:1</sub> I and/or anteiso-C<sub>17:1</sub> B.

## **SUPPLEMENTARY MATERIAL**

### **Supplementary methods**

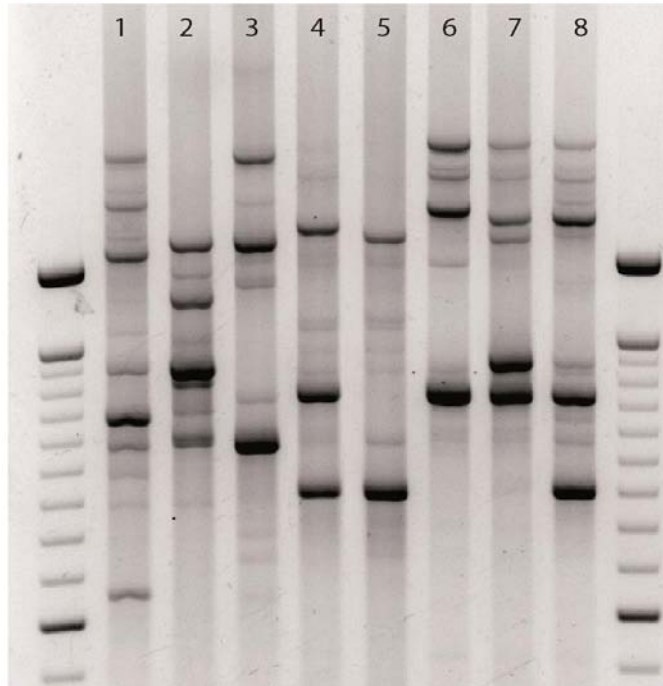
#### DNA finger printing by BOX-PCR

Repetitive extragenic palindromic PCR (rep-PCR) genomic fingerprinting of the isolates was carried out with a Box-A1R primer (5'-CTACGGCAAGGCGACGCTGACG-3') according to the protocol of (Häne et al., 1993). The following PCR conditions were used: 2 min at 95°C, followed by 30 cycles of 93°C for 45 s, 50°C for 1 min, 65°C for 8 min, and a final extension at 65°C for 8 min. PCR products were analyzed on 1.5% (w/v) agarose gels containing 0.5X TAE-buffer (200 mM trisacetate, 0.5 mM EDTA, pH8). The electrophoresis was performed for 5 hours at 180 mV (5V cm<sup>-1</sup>). A 1-kb Plus DNA size ladder (INVITROGEN) was run at both sides of the gel. The gels were stained with ethidium bromide.

#### SDS-PAGE

Overnight cultures (grown in MB medium) were inoculated at 50-fold dilution into the same growth medium and subsequently cultured for 10 h at 37 °C (OD<sub>600nm</sub>=0.6). 500-µl samples were harvested by centrifugation at 14,000 rpm for 2min. The pellets were resuspended in 50 µl 1X DTT. The protein concentration was determined using the method described previously (Bradford, 1976) and samples were analyzed in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels (Laemmli, 1970).

## Supplementary figures

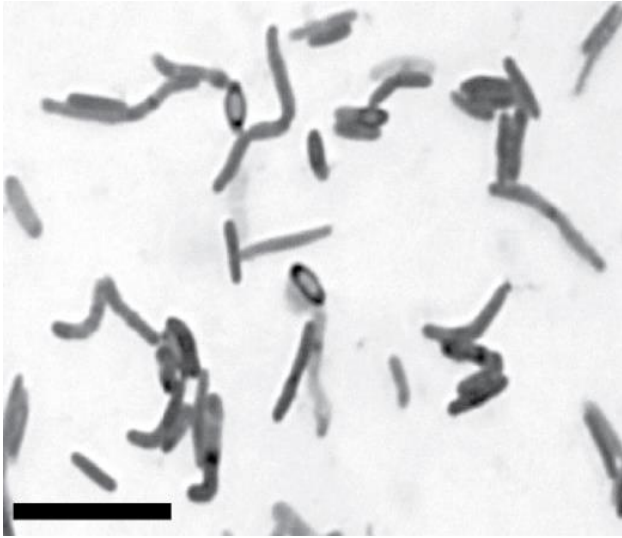


**Figure S1.** BOX-PCR genomic fingerprints of strains P19<sup>T</sup>, PMs43e, PE8e, A81 and 105M and of type strains of closely related species of *Bacillus*. 1) *Bacillus aquimaris* TF-12<sup>T</sup>; 2) *Bacillus marisflavi* TF-11<sup>T</sup>; 3) *Bacillus vietnamensis* NRIC 0531<sup>T</sup> and 4-8) strains P19<sup>T</sup>, PMs43e, PE8e, A81 and 105M respectively.

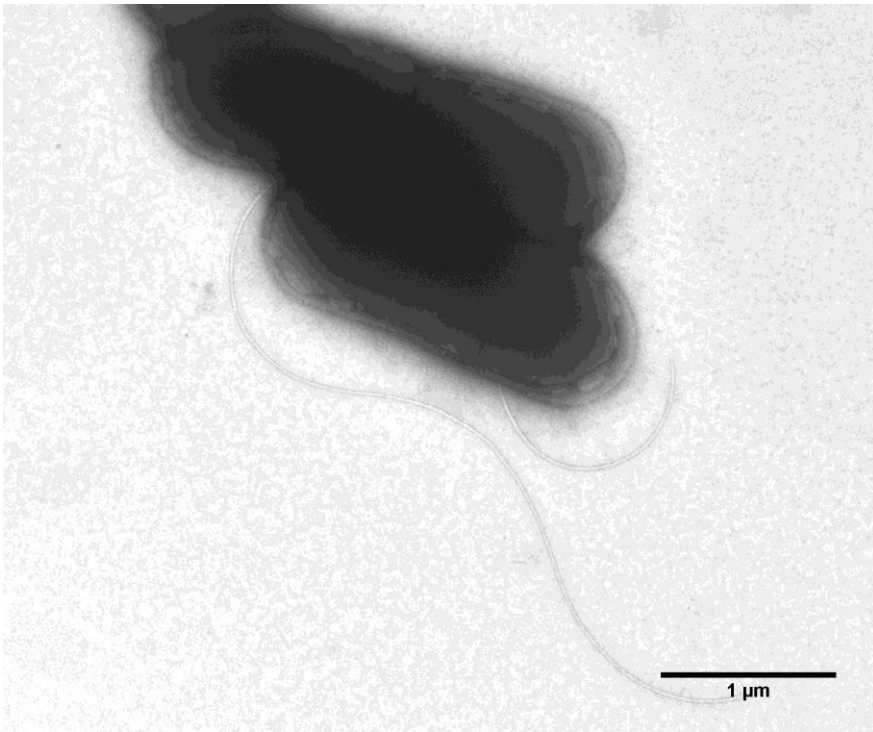




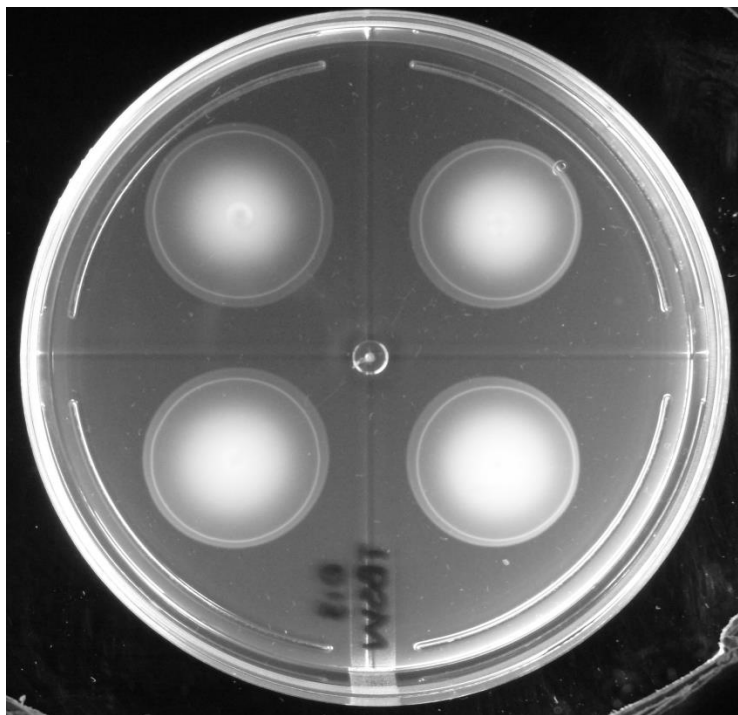
**Figure S2.** Whole-cell protein electrophoregrams (SDS-PAGE) of strains P19<sup>T</sup>, PMs43e, PE8e, A81 and 105M and of type strains of closely related species of *Bacillus*. 1) *Bacillus aquimaris* TF-12<sup>T</sup>; 2) *Bacillus marisflavi* TF-11<sup>T</sup>; 3) *Bacillus vietnamensis* NRIC 0531<sup>T</sup> and 4-8) strains P19<sup>T</sup>, PMs43e, PE8e, A81 and 105M respectively.



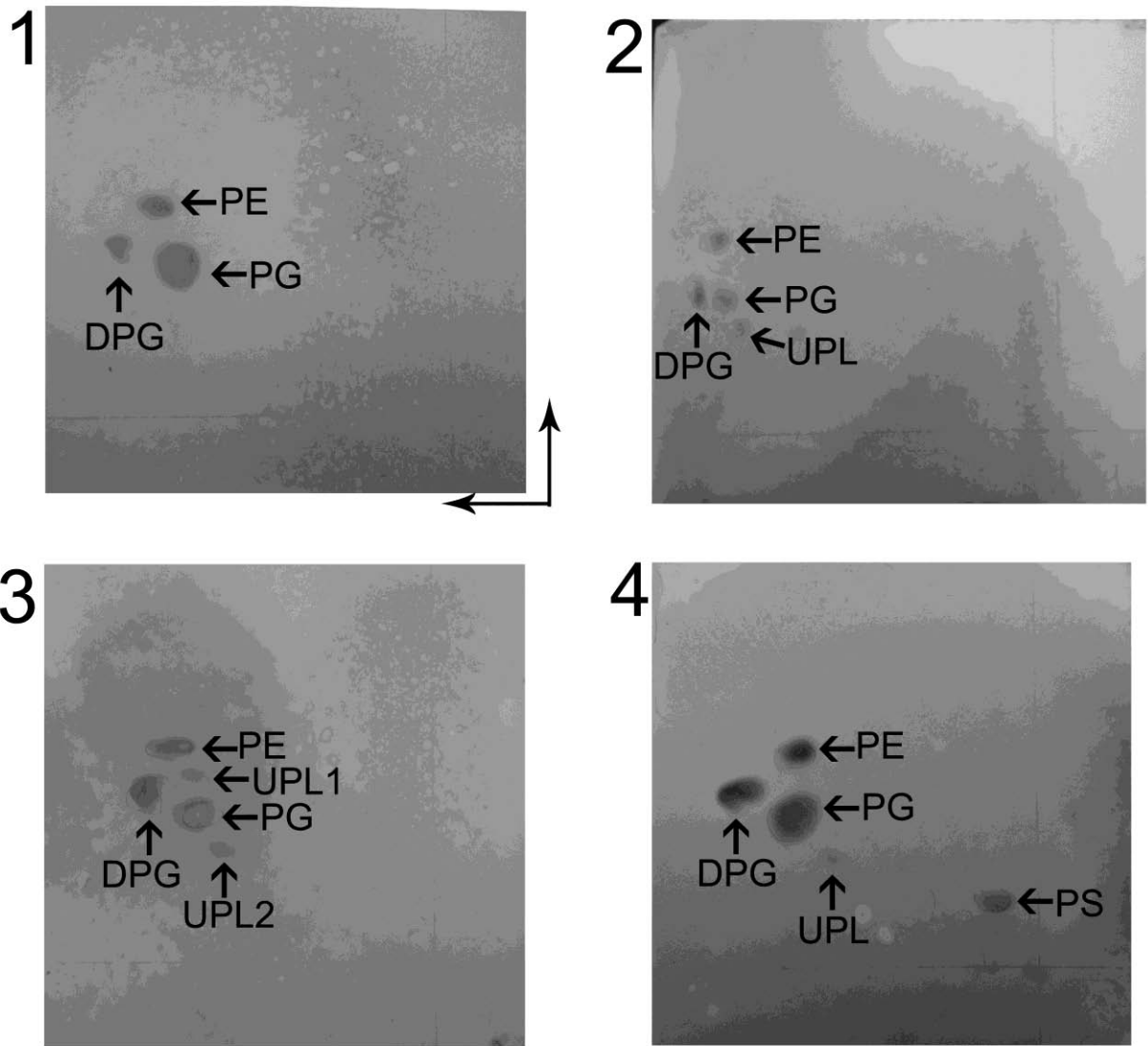
**Figure S3.** Photomicrographs of vegetative cells and endospores of strain P19<sup>T</sup>. Bar, 5  $\mu\text{m}$ .



**Figure S4.** Electron micrograph of strain P19<sup>T</sup> isolated from an overnight culture stained with 1% PTA as indicated under Materials and methods. Bar, 1  $\mu\text{m}$ .



**Figure S5.** Swimming assays of strain P19<sup>T</sup> in 0.3 % soft agar plates. Soft agar plates were made with 0.3% agar in TBSW medium; see Materials and methods. Cells were incubated at 37 °C for 10 h.



**Figure S6.** Polar lipid profile of *Bacillus* species, 1) P19<sup>T</sup>, 2) *B. aquimaris*, 3) *B. marisflavi* 4) *B. vietnamensis*. The lipids phosphatidylserine (PS), phosphatidylethanolamine (PE), phosphatidylglycerol (PG) diphosphatidylglycerol (DPG), unknown phospholipid (UPL or UPL1 and UPL2) are indicated. Arrow heading top, first dimension. Arrow heading left, second dimension.

**References:**

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.

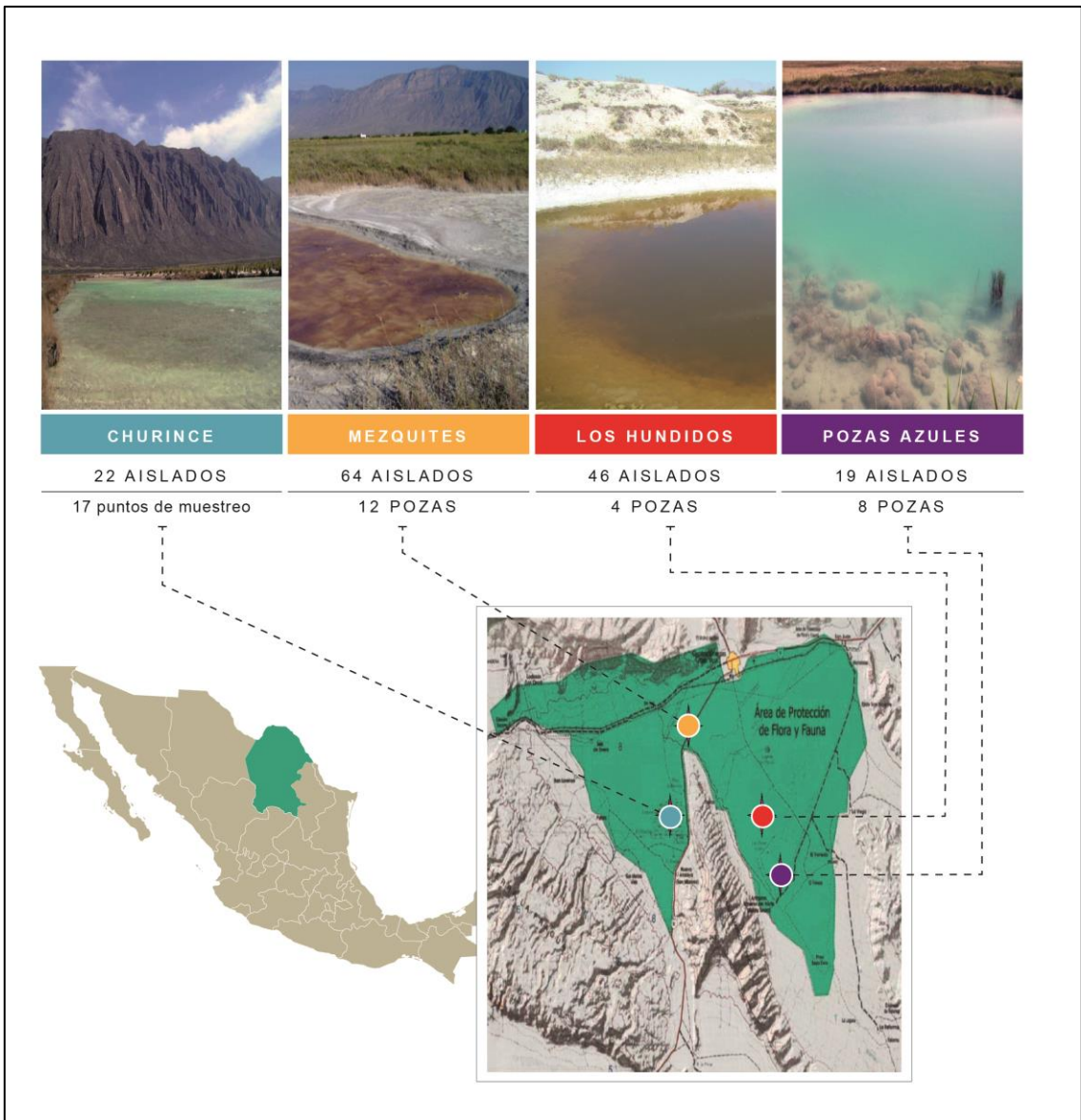
Häne BG, Jäger K, Drexler HG. 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. *Electrophoresis* **14**: 967-972.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

#### **Capítulo 4. Diversidad genética, distribución y recombinación en las poblaciones de *Bacillus* en CCC**

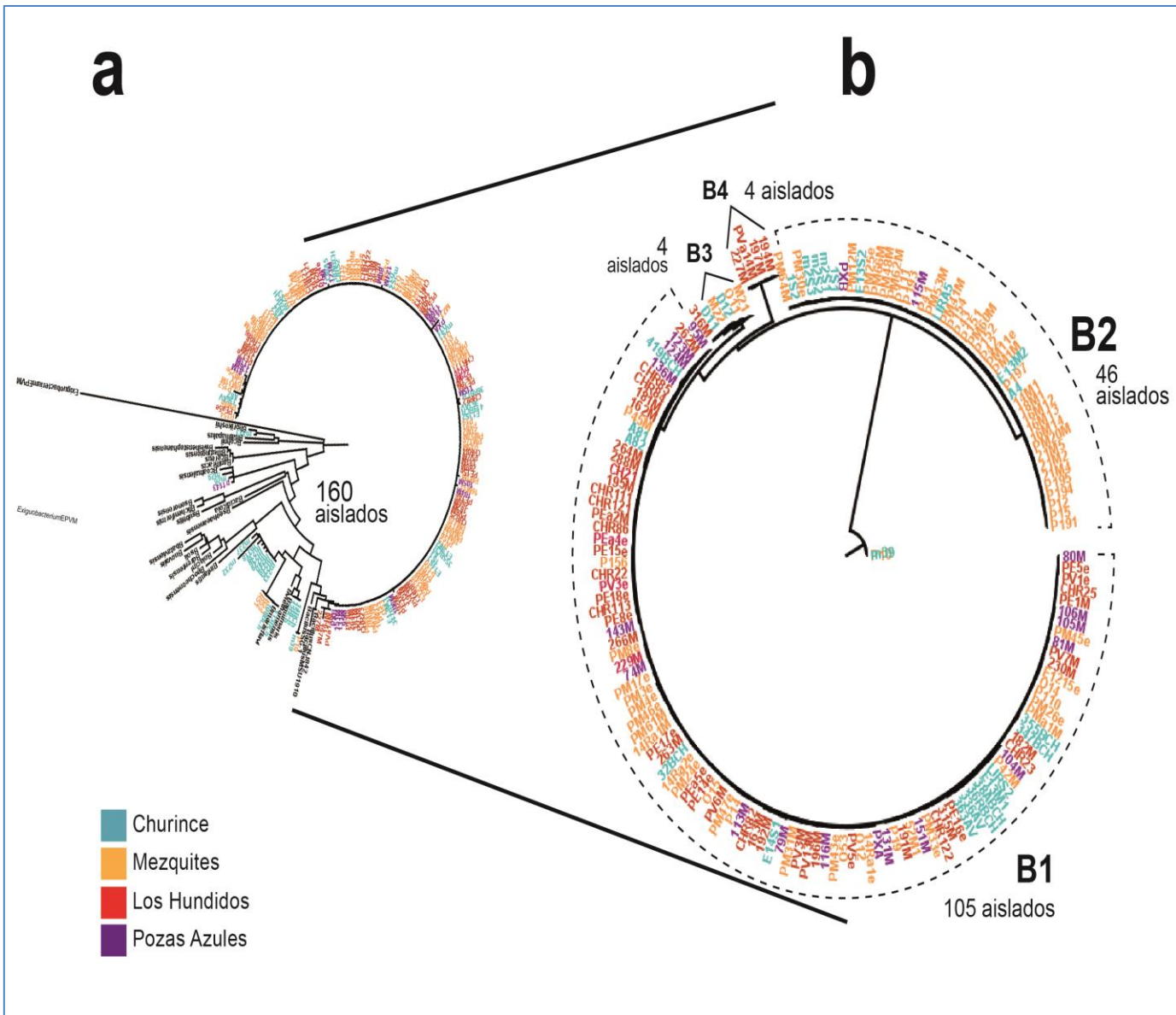
A partir de la colección de 160 aislados obtenida en este proyecto, se procedió a la amplificación y obtención de secuencias parciales de 5 genes *housekeeping* (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*) para poder llevar a cabo un análisis multilocus y poder analizar con mayor detalle la diversidad genética y la dinámica poblacional de los grupos de *Bacillus* recuperados en el valle. Los detalles de los sitios de muestreo y el número de cepas obtenidas en cada sitio se muestran en la figura 7.

La reconstrucción filogenética obtenida con la secuencia completa del gen 16S rRNA reunió a los 160 aislados de la colección en un solo filogrupo, como puede verse en la figura 8a. Sin embargo, la filogenia reconstruida a partir de la secuencia concatenada de los cinco genes *housekeeping* dividió al total de los aislados en cuatro grupos, nombrados como B1, B2, B3 y B4, que pueden verse en la figura 8b. En dicha figura se presentan los aislados coloreados de acuerdo con el sitio de origen de la muestra de la cual se obtuvieron. Dentro de los cuatro filogrupos identificados, los que reunieron a la mayor parte de los aislados son los identificados como B1 y B2, por lo que el resto de los análisis se hicieron usando sólo los 151 aislados incluidos en estos dos filogrupos.



**FIGURA 7.** Mapa de los sitios de muestreo incluidos en este trabajo. Se muestra la ubicación de cada sistema acuático en el valle de CCC, el número de puntos de muestreo de cada sitio y así el número de aislados obtenidos de cada ubicación. El código de colores presentado en esta figura (**azul** para Churince, **naranja** para Mezquites, **rojo** para Los Hundidos y **morado** para Pozas Azules) se utiliza en el resto del capítulo para hacer referencia al sitio de obtención de los aislados





**Figura 8.** Filogenias de máxima verosimilitud de los grupos de *Bacillus* de CCC. En **a** se muestra la filogenia reconstruida usando la secuencia completa del gen 16S rRNA de los 160 aislados recuperados en este trabajo así como las secuencias de algunos otros aislados recuperados en trabajos previos en CCC y de la cepa *Exiguobacterium* EPVM como grupo externo. En **b** se muestra la filogenia reconstruida usando la secuencia parcial concatenada de cinco genes (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*; 2330 nucleótidos en



total) así como el número de aislados que se agrupan en cada filogrupo (**B1**, **B2**, **B3** y **B4**). El código de colores se refiere al sitio de origen de los aislados.

Como puede apreciarse en la filogenia obtenida a partir de las secuencias concatenadas, los filogrupos B1 y B2 están claramente separados entre ellos y forman grupos bien definidos. Dicha separación se confirmó con los valores de  $F_{ST}$  obtenidos para dichos filogrupos, mostrados en la tabla 1. Los valores de  $F_{ST}$  obtenidos entre los dos filogrupos fueron muy altos para los cinco genes housekeeping pero muy bajos cuando se usó la secuencia del gen 16s rRNA. La anterior confirma, por un lado, que estos filogrupos se encuentran genéticamente diferenciados y, por el otro, que dicha diferenciación no puede observarse con el bajo nivel de resolución alcanzado con la secuencia del gen 16S rRNA.

**Tabla 1. Índice de diferenciación genética ( $F_{ST}$ ) entre filogrupos de *Bacillus* y entre sistemas acuáticos de CCC.**

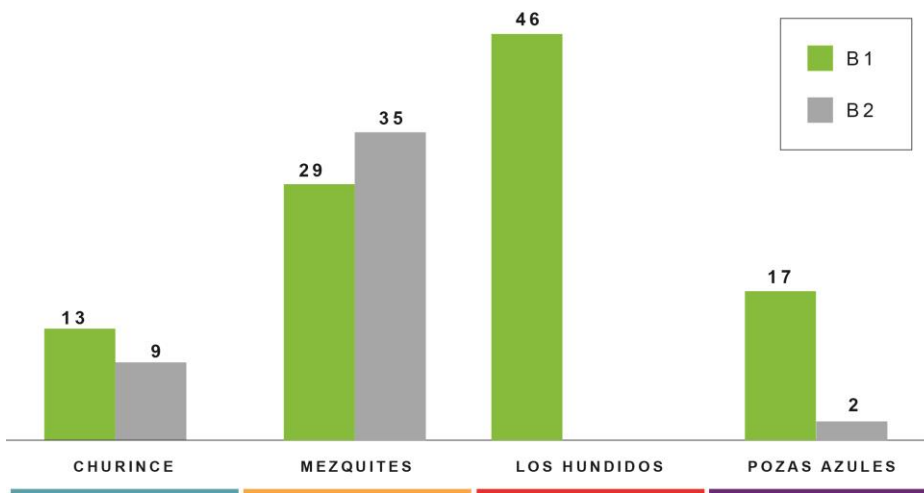
|                 | Filogrupos B1-B2 | CH-M    | CH-H    | CH-PA   | M-H     | M-PA   | H-PA    |
|-----------------|------------------|---------|---------|---------|---------|--------|---------|
| <b>16s rRNA</b> | 0.00758          | -       | -       | -       | -       | -      | -       |
| <b>citC</b>     | 0.95             | 0.00743 | 0.242   | 0.1384  | 0.3768  | 0.2819 | 0.0326  |
| <b>gltX</b>     | 0.935            | 0.031   | 0.235   | 0.155   | 0.344   | 0.257  | 0.00055 |
| <b>hsp70</b>    | 0.962            | 0.00203 | 0.224   | 0.1889  | 0.359   | 0.329  | 0.00808 |
| <b>recA</b>     | 0.931            | 0.00132 | 0.16021 | 0.12515 | 0.277   | 0.2545 | 0.0229  |
| <b>spo0A</b>    | 0.924488         | 0.01726 | 0.31792 | 0.1338  | 0.45855 | 0.2806 | 0.0208  |

Valores de  $F_{ST}$  pareada ( $p < 0.0001$ ) (Hudson et al., 1992) entre filogrupos (**B1** y **B2**) y entre sistemas acuáticos (**CH**, Churince; **H**, Hundidos; **M**, Mezquites; **PA**; Pozas Azules).

Por otra parte, es interesante notar que el grupo B2 está integrado en su mayoría por aislados de Mezquites, mientras que el grupo B1 está integrado por aislados recuperados de los cuatro sitios de muestreo. Para explorar la posible influencia del sitio de aislamiento de la muestra sobre la estructura genética de los filogrupos obtenidos, se calcularon también los valores de  $F_{ST}$  entre los cuatro sistemas acuáticos. Como puede verse en la tabla 1, los valores de  $F_{ST}$  entre los sitios de muestreo son bajos, lo cual indica que no existe una estructura poblacional asociada al sitio de origen de los aislados.

Considerando que estudios previos sobre la distribución de especies bacterianas en CCC han mostrado que existen diferencias en la composición bacteriana entre sitios, se exploró la idea de que al interior de los filogrupos de *Bacillus* pudiera existir una estructuración de la diversidad genética asociada al sitio de muestreo. Como se muestra en la figura 9, B1 presenta una distribución más amplia ya que se localiza en los cuatro sistemas acuáticos muestreados, mientras que B2 no está presente en Los Hundidos. Basándose en la distribución de los filogrupos y en los valores de  $F_{ST}$  es claro que el sitio de aislamiento no es

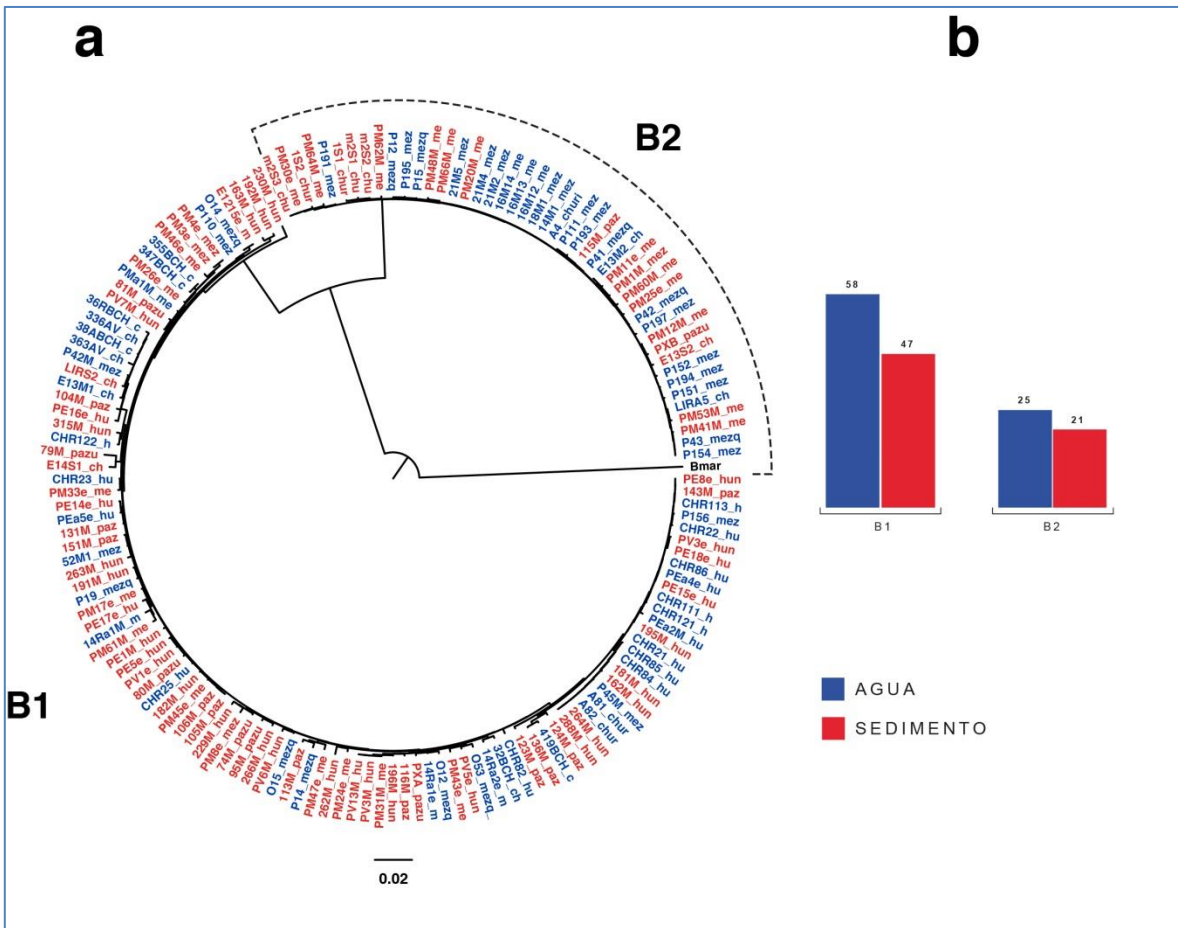
un factor que se relacione con la distribución de la diversidad genética en las poblaciones de *Bacillus* estudiadas. El patrón de dispersión amplia que se observa en el grupo B1 podría explicarse en cierta medida a la capacidad del género *Bacillus* de formar endosporas, lo cual podría contribuir a darle a los linajes de *Bacillus* de CCC una mayor capacidad de dispersión que a otros linajes que no pueden formar dichas estructuras de resistencia.



**Figura 9.** Distribución de los aislados en los cuatro sistemas acuáticos de CCC. Cada color representa un filogrupo (B1 y B2).

Por otra parte, existe evidencia de que si bien el sitio de origen del muestreo puede no ser importante en la estructura genética, existen otros valores ambientales en CCC que pueden afectar la distribución de las poblaciones

bacterianas en el valle. En un trabajo reciente con poblaciones del género *Exiguobacterium* aisladas en CCC, Rebollar y colaboradores (2012) pudieron detectar la existencia de subgrupos compuestos por organismos asociados exclusivamente al sedimento o al agua de los diferentes sitios de muestreo dentro del valle. Tomando en cuenta este antecedente, se exploró en las poblaciones de *Bacillus* de este estudio la posibilidad de que el ambiente de procedencia de la muestra (agua o sedimento) pudiera tener una influencia sobre la distribución de la diversidad genética entre o al interior de los filogrupos. En la figura 10 se muestra por la filogenia de los 2 grupos (B1 y B2) sobre la cual se colorea el ambiente de procedencia de los aislados y también se grafica la distribución del número de aislados provenientes de cada ambiente en cada filogrupo.



**Figura 10.** Filogenia de máxima verosimilitud y distribución de los filogrupos B1 y B2 en los ambientes de aislamiento. En **a** se muestra la filogenia reconstruida usando la secuencia parcial concatenada de cinco genes (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*; 2330 nucleótidos en total) de los filogrupos B1 y B2. En **b** se muestra la proporción de aislados provenientes de agua o sedimento al interior de cada filigrupo. El código de colores se refiere al ambiente del cual fueron recuperados los aislados (agua o sedimento).

En este caso, tampoco se encontró un patrón evidente de asociación entre el ambiente de origen de la muestra y los filogrupos de *Bacillus* de CCC. Este resultado fue confirmado por los resultados de la prueba UniFrac (Lozupone,

Hamady & Knight, 2006) en la que no se tuvieron resultados significativos que permitieran explicar la distribución de la diversidad genética ni entre filogrupos ni al interior de ellos asociada al ambiente de aislamiento de las cepas.

Con todos los resultados anteriores queda claro que existen dos filogrupos bien diferenciados (B1 y B2) al interior de la colección de aislados de *Bacillus* de CCC recuperados en este estudio. Por otro lado, la divergencia entre estos filogrupos no parece estar asociada con el sitio de origen de la muestra (Churince, Mezquites, Los Hundidos, Pozas Azules) ni con el ambiente particular del cual fueron recuperados (agua y sedimento). Para tratar de profundizar un poco más en los factores que podrían explicar la divergencia de estos dos filogrupos mayores se procedió a explorar la diversidad genética al interior de cada filogrupo así como la posible influencia de la mutación y la recombinación en generar los patrones de diversidad observados.

#### **4.1 Diversidad genética y recombinación en los filogrupos B1 y B2**

Para poder explorar con más detalle la estructura genética de las poblaciones del género *Bacillus* se calcularon los índices de diversidad genética al interior de cada uno de los filogrupos B1 y B2. Los resultados obtenidos se presentan en la tabla 2, en la que puede observarse que los valores de diversidad nucleotídica medidos como  $\pi$  y  $\theta$  son bajos. Los valores obtenidos en ambos filogrupos son

comparables con los valores obtenidos en el trabajo de Cerritos (Cerritos, 2008), lo cual es notable debido a que los tamaños de muestra de ese trabajo eran mucho menores y podrían haber resultado en los bajos niveles de diversidad nucleotídica observada, sin embargo, el hecho de que sean similares a los valores obtenidos en el presente trabajo da evidencia de que los niveles de diversidad de los *Bacillus* de CCC son, en general bajos, sobre todo si se comparan con los valores obtenidos para poblaciones de otros géneros dentro de CCC como *Pseudomonas* y *Exiguobacterium* (Escalante, 2008; Rebollar et al., 2012). En particular, los valores obtenidos para poblaciones de *Exiguobacterium* son de un orden de magnitud mayor que los obtenidos para las poblaciones de *Bacillus* (Rebollar et al., 2012).

**Tabla 2. Diversidad genética de los filogrupos B1 y B2 con seis marcadores moleculares.**

| <b>GRUPO B1 (N=105)</b> |          |         |          |          |           |
|-------------------------|----------|---------|----------|----------|-----------|
| <b>Gen</b>              | <b>S</b> | $\pi$   | $\theta$ | <b>H</b> | <b>Hd</b> |
| <i>16s rRNA</i>         | 7/1330   | 0.00018 | 0.00101  | 12       | 0.199     |
| <i>citC</i>             | 23/489   | 0.0064  | 0.009    | 25       | 0.903     |
| <i>gltX</i>             | 30/387   | 0.01503 | 0.01483  | 22       | 0.906     |
| <i>hsp70</i>            | 26/527   | 0.00445 | 0.00944  | 25       | 0.829     |
| <i>recA</i>             | 24/525   | 0.00427 | 0.00875  | 17       | 0.763     |
| <i>spo0A</i>            | 38/390   | 0.00878 | 0.01864  | 33       | 0.878     |
| <b>GRUPO B2 (N=46)</b>  |          |         |          |          |           |

| Gen             | S       | $\pi$   | $\theta$ | H  | Hd    |
|-----------------|---------|---------|----------|----|-------|
| <i>16s rRNA</i> | 10/1329 | 0.00042 | 0.00171  | 9  | 0.388 |
| <i>citC</i>     | 9/489   | 0.00308 | 0.00419  | 12 | 0.809 |
| <i>gltX</i>     | 2/387   | 0.00139 | 0.00118  | 3  | 0.513 |
| <i>hsp70</i>    | 6/527   | 0.00141 | 0.00259  | 7  | 0.506 |
| <i>recA</i>     | 6/531   | 0.00144 | 0.00257  | 7  | 0.579 |
| <i>spo0A</i>    | 5/390   | 0.00192 | 0.00292  | 8  | 0.581 |

**S** es el número de sitios segregantes/el número total de nucleótidos,  $\pi$  es la diversidad nucleotídica pareada de Nei,  $\theta$  es la teta por sitio de Watterson =  $2Ne\mu$ , **H** es el número de haplotipos y **Hd** es la diversidad nucleotídica, calculados según Rozas y Rozas, 1999.

Además de la baja diversidad genética, en el trabajo de trabajo de Cerritos (Cerritos, 2008) con poblaciones de *Bacillus* de CCC se observó una estructura poblacional altamente clonal que podría estar relacionada con los bajos niveles de diversidad. Con el fin de probar si esta estructura poblacional clonal es también la regla en las poblaciones de *Bacillus* de este estudio, se analizó con detalle el papel que la mutación y la recombinación podrían tener en generar la diversidad genética observada en dichas poblaciones.

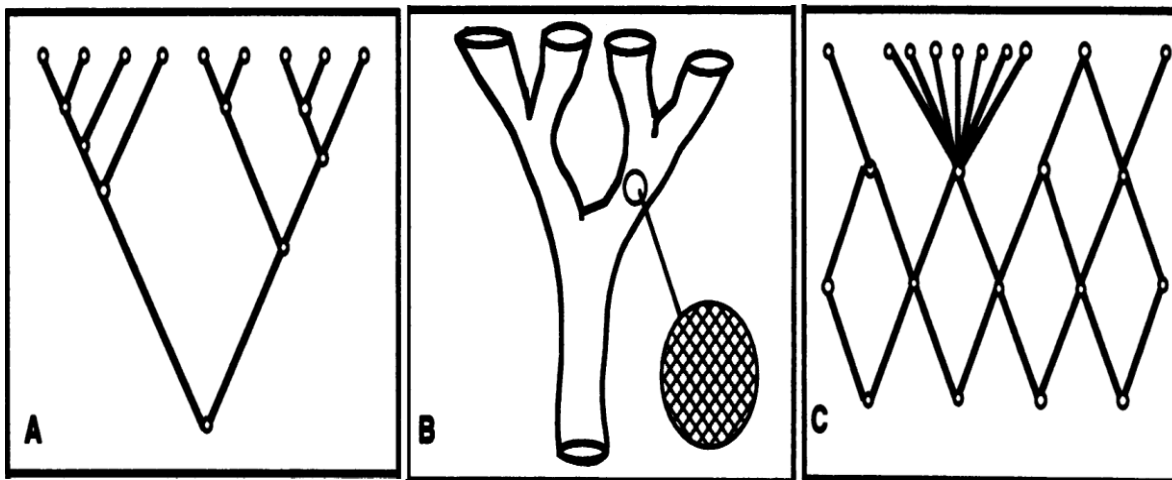
El estudio de la recombinación en bacterias se ha basado históricamente en tres tipos de evidencias: 1) la estructura de mosaico dentro de los genes, que proporciona la evidencia molecular de eventos de recombinación intragénica, 2) el equilibrio de ligamiento entre alelos de diferentes loci, que proporciona evidencia estadística de la recombinación intragénica y 3) la comparación entre árboles genealógicos derivados de genes diferentes, que evidencia historias evolutivas



diferentes para regiones diferentes del genoma (Maynard Smith & Smith, 1993; Spratt & Maiden, 1999; Feil & Spratt, 2001). En particular en este trabajo, considerando que las secuencias utilizadas en este estudio son secuencias parciales que no cubren la totalidad de los genes analizados, los análisis se centraron en detectar recombinación entre genes y no al interior de cada uno de ellos.

La primera medida de recombinación obtenida en las poblaciones de este estudio fue el índice de asociación. El índice de asociación ( $I_A$ ) (Maynard Smith et al., 1993) es una medida que permite detectar el desequilibrio de ligamiento entre alelos de varios loci. Si la población es completamente clonal, siempre se presentarán los mismos alelos de varios loci juntos mientras que si es completamente recombinante (panmíctica) se observarán todas las posibles combinaciones de alelos. El índice de asociación se obtiene entonces al calcular la diferencia entre la varianza que se esperaría si todos los alelos estuvieran distribuidos al azar y la varianza observada de esta distribución. Si no hay diferencia significativa entre las varianzas ( $I_A = 0$ ), se puede suponer que la recombinación es libre en la muestra (Salas, 2007). Los valores de  $I_A$  diferentes de cero deben tratarse con cuidado debido a que efectivamente pueden significar que la recombinación está ausente en las poblaciones sin embargo existen otra serie de razones que pueden provocar que se observe desequilibrio de ligamiento al interior de las poblaciones. Procesos como la epistasis (donde la selección natural

favorece ciertas combinaciones de alelos) y la deriva génica pueden provocar que en las poblaciones donde la recombinación no está ausente se observe desequilibrio de ligamiento (Maynard Smith et al., 1993). Por otro lado se puede observar desequilibrio de ligamiento si la muestra analizada consiste de una mezcla de varias poblaciones dentro de cada una de las cuales la recombinación es común pero es rara entre poblaciones (Souza et al., 1992). Finalmente, el desequilibrio de ligamiento puede observarse también en poblaciones con una estructura epidémica. En este caso existirá una asociación significativa entre loci pero debido al incremento reciente y explosivo de una clona particular que puede confundirse con una estructura poblacional efectivamente clonal (Maynard Smith et al., 1993, figura 11).



**FIGURA 11.** Representación de diferentes tipos de estructuras poblacionales en bacterias. **A** y **B** son poblaciones que consisten de aislados separados en dos ramas principales. La estructura de la población en **A** es clonal en todos los niveles, de modo tal que el dendograma representado es un árbol evolutivo: no hay recombinación ni entre

aislados de la misma rama ni entre las dos ramas del árbol. En **B**, la recombinación no ocurre entre aislados de las dos ramas principales, pero sí hay recombinación frecuente entre los aislados de cada una de las ramas. En **C** se representa una estructura poblacional epidémica en la cual existe recombinación frecuente entre todos los miembros de la población, de modo que la estructura poblacional debe representarse como una red en lugar de un árbol. Sin embargo, ocasionalmente surge un individuo muy exitoso cuya frecuencia aumenta rápidamente en la población produciendo así una clona epidémica (modificada de Maynard-Smith et al., 1993).

Al índice de asociación se le han hecho variaciones para calcular las varianzas de las diferencias pareadas sin asumir una distribución normal, obteniendo así una distribución exacta por medio de simulaciones Monte Carlo (Haubold et al., 1998; Haubold & Hudson, 2000). LIAN (Haubold & Hudson, 2000) es un programa que permite poner a prueba la hipótesis nula del equilibrio de ligamiento ( $V_D=V_e$ ). Se calcula el número de loci en los cuales cada par de holotipos difiere y a partir de la distribución de los mismatches se calcula la varianza ( $V_D$ ). Esta varianza calculada se compara con la varianza esperada bajo el equilibrio de ligamiento ( $V_e$ ) y la hipótesis nula de una población en equilibrio de ligamiento se pone a prueba con simulaciones de Monte Carlo. Usando este programa se calculó el índice de asociación estandarizado ( $I_A^S$ ) para las poblaciones de *Bacillus* de este estudio y se puso a prueba la hipótesis nula de que las poblaciones están en equilibrio de ligamiento. Los análisis se llevaron a cabo en el total de la muestra así como para los grupos B1 y B2 de manera independiente. En ambos casos los análisis se llevaron a cabo con el total de las

secuencias y también incluyendo sólo los secuencia tipos (STs) diferentes. Los STs se obtuvieron con base en la secuencia concatenada de los cinco genes *housekeeping* (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*) y se definieron como secuencias únicas por medio de un análisis de haplotipos utilizando DNAsp (Rozas et al., 2003). A partir de este análisis se definieron 142 STs de los 151 aislados del género *Bacillus*.

**Tabla 3. Análisis de equilibrio de ligamiento a partir de los datos multilocus usando LIAN (Haubold y Hudson, 2000)**

|                                  | $V_D$  | $V_e$  | $I_A^s$ | $P_{MC}$                            |
|----------------------------------|--------|--------|---------|-------------------------------------|
| <b>Total secuencias (N=151)</b>  | 0.5232 | 0.3269 | 0.2001  | <b>&lt; 1.00 x 10<sup>-04</sup></b> |
| <b>Total STs (N=142)</b>         | 0.4089 | 0.3077 | 0.1096  | <b>&lt; 1.00 x 10<sup>-04</sup></b> |
| <b>Gpo B1 secuencias (N=105)</b> | 0.5744 | 0.4647 | 0.0787  | <b>&lt; 1.00 x 10<sup>-04</sup></b> |
| <b>Grupo B1 STs (N=94)</b>       | 0.5017 | 0.4626 | 0.0282  | <b>9.10 x 10<sup>-03</sup></b>      |
| <b>Grupo B2 secuencias (46)</b>  | 1.0754 | 0.897  | 0.0663  | <b>1.22 x 10<sup>-02</sup></b>      |
| <b>Grupo B2 STs (N=35)</b>       | 0.7757 | 0.7997 | -0.01   | 0.623                               |

Los valores en negritas muestran los valores significativos  $P < 0.05$ .

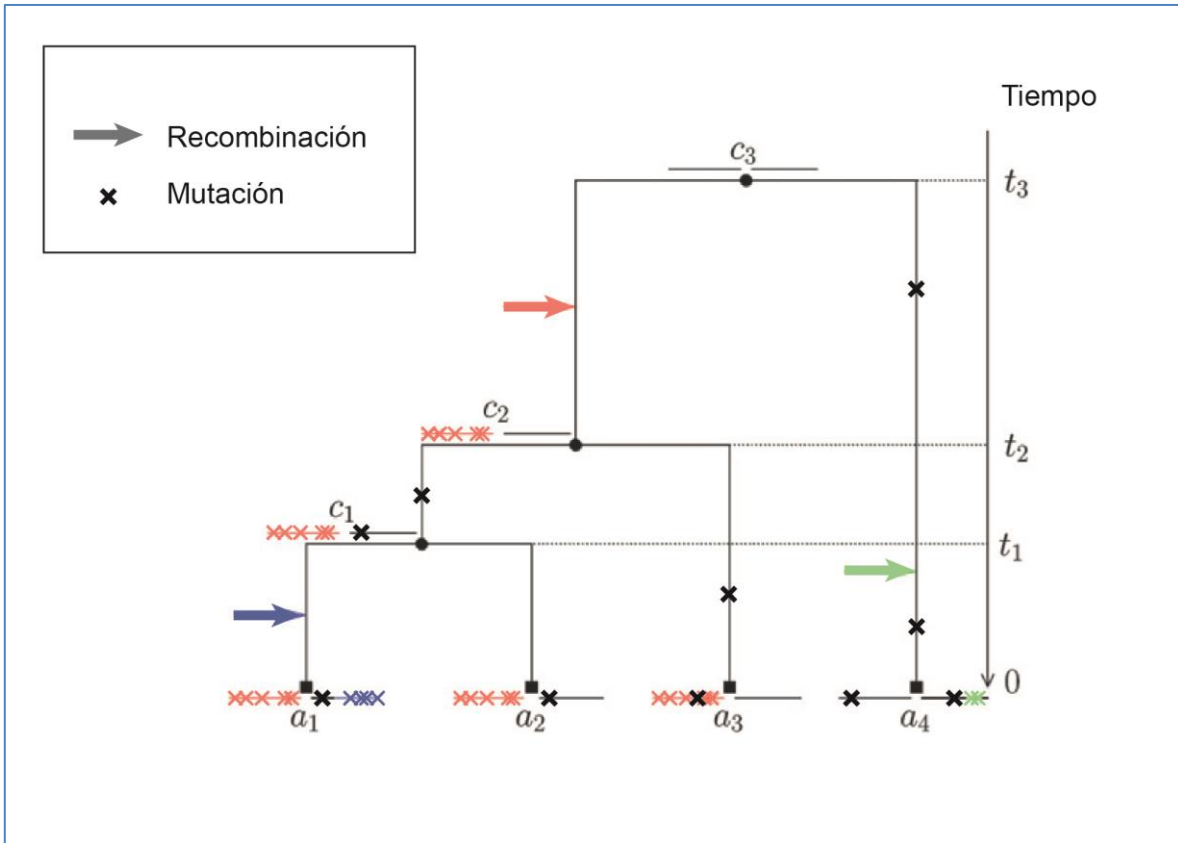
$V_D$  es la varianza observada de la distribución de mismatches,  $V_e$  es la varianza esperada,  $I_A^s$  es el índice de asociación estandarizado y  $P_{MC}$  es el valor de la prueba de significancia que pone a prueba la hipótesis nula de equilibrio de ligamiento ( $H_0: V_D=V_e$ ) con simulaciones Monte Carlo.

Como puede verse en la tabla 3 la mayoría de los valores  $P_{MC}$  obtenidos para los  $I_A^s$  resultaron significativos, por lo que en prácticamente todos los grupos analizados se pudo rechazar la hipótesis nula de que existe equilibrio de ligamiento. Este resultado indica que la recombinación homóloga entre alelos, si es que sucede en estas poblaciones, no ha sido suficiente para alcanzar el equilibrio de ligamiento entre alelos. Por otro lado, es interesante notar que en el grupo B2, cuando se analizaron todas las secuencias el resultado fue lo esperado para una población que no está en equilibrio de ligamiento, sin embargo, cuando se analizaron solamente las secuencias diferentes pudo observarse que el valor de  $P_{MC}$  resultó no significativo, lo cual puede interpretarse como que en este grupo podría existir una clona dominante que, al ser eliminada en los STs, permite reconocer que los niveles de recombinación al interior de este grupo no son tan bajos como se pensaría al analizar la muestra completa.

Aunque el índice de asociación es una medida tradicionalmente utilizada para detectar desequilibrio de ligamiento en las poblaciones, es importante tener en cuenta el hecho de que los valores diferentes de cero pueden no deberse necesariamente a la ausencia de recombinación en las poblaciones sino a otros factores como epistasis, estructuras poblacionales epidémicas, etc. Para analizar con mayor profundidad la contribución relativa de la mutación y la recombinación en la diversidad genética observada en los grupos de *Bacillus* aquí analizados, se

llevó a cabo un análisis de recombinación usando ConalFrame (Didelot & Falush, 2007).

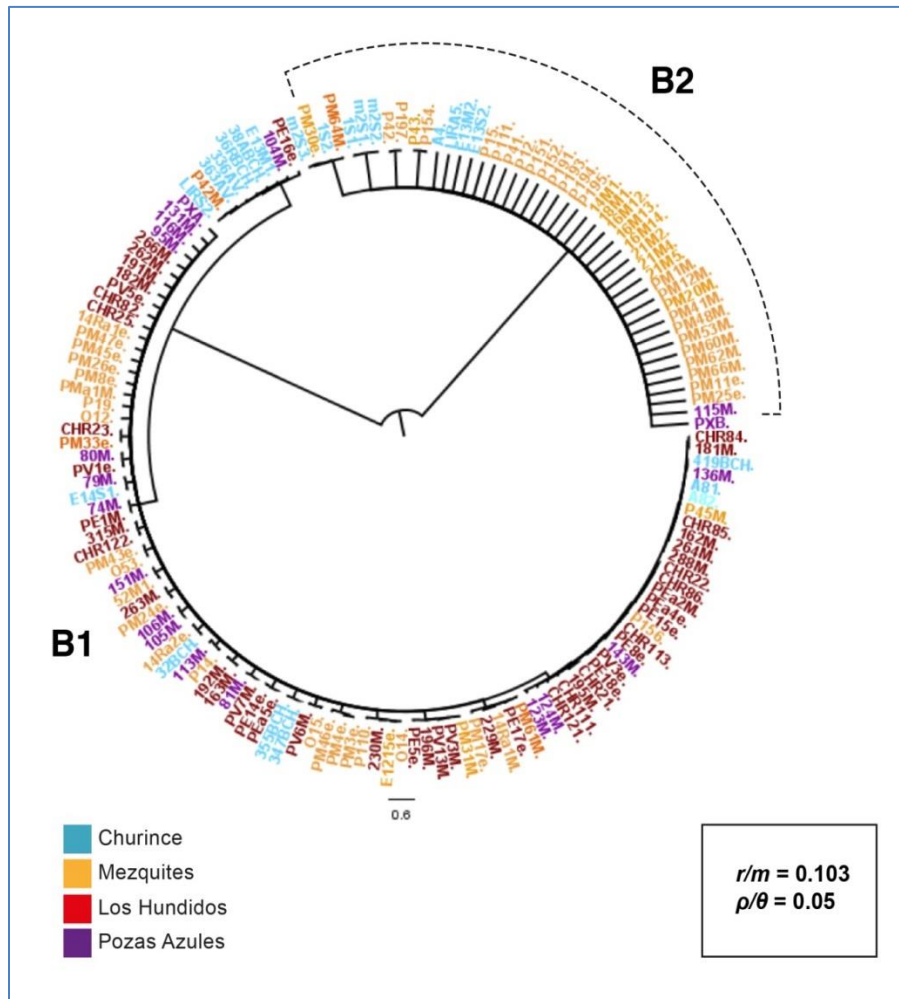
ConalFrame es un programa que desarrolla un método para inferir relaciones clonales entre bacterias así como los eventos de recombinación que podrían interrumpir el patrón de herencia clonal. Busca reconstruir la genealogía de una muestra de cepas, así como los eventos de mutación y recombinación de las ramas de dicha genealogía, basándose en un modelo de coalescencia (figura 12). El programa permite, usando estadística Bayesiana, inferir la genealogía de un grupo de bacterias y los parámetros que pueden afectarla (mutación, recombinación, localización y longitud de los fragmentos recombinantes). Así mismo permite calcular las probabilidades relativas de que las sustituciones nucleotídicas observadas en un conjunto de secuencias sean resultado de recombinación o mutación ( $r/m$ ). De este modo, da una medida de impacto relativo de la recombinación en la generación de la diversidad genética observada (Didelot & Falush, 2007). Finalmente, ConalFrame tiene la ventaja de que el cálculo de  $r/m$  que hace permite la comparación directa de los resultados obtenidos para una muestra o población particular con los valores de  $r/m$  calculados para cualquier otra especie de bacterias o arqueas (Vos & Didelot, 2009).



**Figura 12.** Ilustración de modelo del ClonalFrame. Dos bloques (líneas horizontales) evolucionan por mutaciones puntuales (cruces negras) y por recombinación (flechas de colores que inducen las sustituciones marcadas por las flechas de colores).  $a_1$ ,  $a_2$ ,  $a_3$  y  $a_4$  corresponden a las secuencias observadas en la muestra analizada mientras que  $c_1$ ,  $c_2$  y  $c_3$  corresponden a las secuencias de los nodos internos (tomada de Didelot y Falush, 2007).

Usando ClonalFrame y las secuencias concatenadas de los cinco genes *housekeeping* (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*) se construyeron las genealogías de los filogrupos B1 y B2 y se calculó la frecuencia a la cual suceden los eventos de recombinación y mutación ( $\rho/\theta$ ) así como las probabilidades relativas de que un

polimorfismo dado pueda ser explicado como producto de un evento de recombinación o de mutación ( $r/m$ ). Este último cálculo es una medida de que tan importante es la recombinación en relación con la mutación en generar los patrones de diversidad genética que se observan en la población analizada.



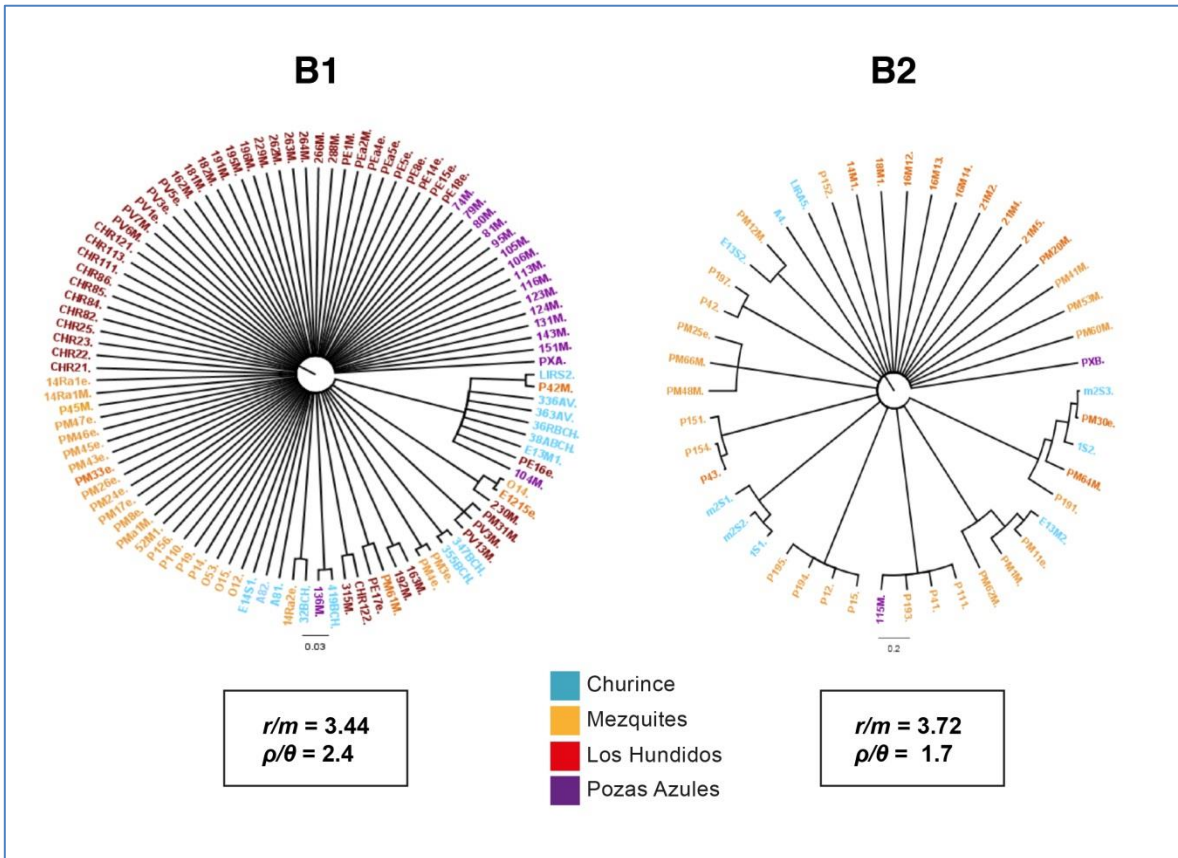
**Figura 13.** Árbol consenso que muestra la relación entre los 151 aislados de la muestra de *Bacillus* de CCC usando cinco genes concatenados (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*; 2330 nucleótidos en total). El dendograma se construyó con la combinación de tres



corridas independientes de ClonalFrame. En el recuadro se muestran los valores obtenidos para los cálculos de la frecuencia de la recombinación con respecto a la mutación ( $\rho/\theta$ ) así como de la contribución relativa de la recombinación con respecto a la mutación en generar la diversidad genética observada en la muestra ( $r/m$ ).

Como puede verse en la figura 13 la genealogía construida con ClonalFrame separó a los 151 aislados en los dos filogrupos previamente identificados (B1 y B2). Por otro lado, los cálculos de  $r/m$  y  $\rho/\theta$  (0.103 y 0.05 respectivamente) resultaron en valores muy bajos. Dichos valores pueden ser interpretados como que la recombinación es mucho menos frecuente que la mutación (aproximadamente 10 veces menor, considerando el valor de  $\rho/\theta = 0.05$ ) y que además la contribución relativa de la recombinación en generar la diversidad genética observada es muy baja, cuando se considera la muestra completa. Sin embargo, cuando los análisis se hicieron al interior de cada uno de los filogrupos (ver figura 14) se pudo observar que tanto los valores de  $\rho/\theta$  como de  $r/m$  aumentaron considerablemente. En ambos casos la contribución relativa de la recombinación con respecto a la mutación pasó de ser prácticamente 10 veces menor en toda la muestra ( $r/m=0.103$ ) a ser aproximadamente dos veces mayor al interior de cada filogrupo ( $r/m = 2.4$  y  $1.7$  en B1 y B2 respectivamente). Esta observación puede ser interpretada como que al interior de cada uno de los filogrupos la recombinación es un poco más frecuente pero los eventos de recombinación prácticamente no suceden entre filogrupos. Esta observación es consistente con las observaciones hechas con otras especies de *Bacillus* en los

cuales la recombinación puede suceder de manera más frecuente al interior de las poblaciones pero no entre ellas (Duncan et al., 1994; Vilas-Boas et al., 2002).



**Figura 14.** Árboles consenso para cada uno de los filogrupos B1 y B2 usando cinco genes concatenados (*citC*, *gltX*, *hsp70*, *recA* y *spo0A*; 2330 nucleótidos en total). Los dendogramas se construyeron con la combinación de tres corridas independientes de ClonalFrame. En los recuadros se muestran los valores obtenidos para los cálculos de la frecuencia de la recombinación con respecto a la mutación ( $\rho/\theta$ ) así como de la contribución relativa de la recombinación con respecto a la mutación en generar la diversidad genética observada en la muestra ( $r/m$ ).

Los resultados de los análisis de recombinación en los grupos de *Bacillus* de CCC de este estudio muestran que la recombinación tiene un papel importante en mantener la cohesión al interior de los filogrupos. Por otro lado, los resultados obtenidos con los análisis de desequilibrio de ligamiento permiten afirmar que, aunque la recombinación sucede al interior de los filogrupos, esta no es tan frecuente como para romper el ligamiento entre genes, cuando menos en el filogrupo B1.

## **Capítulo 5. ARTÍCULO 2:** “Expansiones poblacionales compartidas entre linajes bacterianos que coexisten son reveladas por la evidencia genética”

### **5.1 Resumen**

Los estudios de poblaciones comparativos pueden ayudar a descubrir la influencia de los eventos históricos sobre los patrones actuales de biodiversidad en taxa que coexisten en cierta área geográfica. En particular, los estudios comparativos derivados de la genética de poblaciones y la teoría de coalescencia han sido usados para investigar la dinámica poblacional de bacterias patógenas con el fin de entender los fenómenos epidémicos. En contraste, y a pesar de su relevancia ecológica, existe poco conocimiento sobre los procesos que determinan la diversidad de poblaciones naturales de bacterias. En este trabajo se analizan los patrones de diversidad genética de poblaciones de tres géneros de bacterias (*Bacillus*, *Exiguobacterium* y *Pseudomonas*) que son abundantes y coexisten en los sistemas acuáticos del valle de Cuatro Ciénegas, México. Se pone a prueba la hipótesis de que un hábitat común deja una huella sobre la variación genética de las poblaciones bacterianas, independientemente de las relaciones filogenéticas. Usando marcadores multilocus 1) se llevaron a cabo análisis filogenéticos, 2) se describió la estructura genética de las poblaciones bacterianas, 3) se calcularon parámetros descriptivos de la diversidad genética, 4) se llevaron a cabo pruebas de neutralidad y 5) se llevaron a cabo reconstrucciones históricas basadas en

coalescencia. Los resultados obtenidos muestran una marca de expansiones poblacionales sincrónicas en tiempo que son independientes tanto de la afinidad filogenética como del sitio de muestreo de los aislados. Así, este estudio proporciona evidencia que apoya el análisis de linajes bacterianos que coexisten en ambientes naturales como una herramienta para avanzar en nuestro entendimiento de la evolución bacteriana más allá de los microbios con importancia médica.

# Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence

Morena Avitia<sup>1</sup>, Ana E. Escalante<sup>2</sup>, Eria A. Rebollar<sup>1,3</sup>,  
Alejandra Moreno-Letelier<sup>4</sup>, Luis E. Eguiarte<sup>1</sup> and Valeria Souza<sup>1</sup>

<sup>1</sup> Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, México DF, México

<sup>2</sup> Departamento de Ecología de la Biodiversidad, Laboratorio Nacional de Ciencias de la Sostenibilidad, Instituto de Ecología, Universidad Nacional Autónoma de México, México DF, México

<sup>3</sup> Biology Department, James Madison University, Harrisonburg VA, USA

<sup>4</sup> Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, UK

## ABSTRACT

Comparative population studies can help elucidate the influence of historical events upon current patterns of biodiversity among taxa that coexist in a given geographic area. In particular, comparative assessments derived from population genetics and coalescent theory have been used to investigate population dynamics of bacterial pathogens in order to understand disease epidemics. In contrast, and despite the ecological relevance of non-host associated and naturally occurring bacteria, there is little understanding of the processes determining their diversity. Here we analyzed the patterns of genetic diversity in coexisting populations of three genera of bacteria (*Bacillus*, *Exiguobacterium*, and *Pseudomonas*) that are abundant in the aquatic systems of the Cuatro Ciénegas Basin, Mexico. We tested the hypothesis that a common habitat leaves a signature upon the genetic variation present in bacterial populations, independent of phylogenetic relationships. We used multilocus markers to assess genetic diversity and (1) performed comparative phylogenetic analyses, (2) described the genetic structure of bacterial populations, (3) calculated descriptive parameters of genetic diversity, (4) performed neutrality tests, and (5) conducted coalescent-based historical reconstructions. Our results show a trend of synchronic expansions across most populations independent of both lineage and sampling site. Thus, we provide empirical evidence supporting the analysis of coexisting bacterial lineages in natural environments to advance our understanding of bacterial evolution beyond medical or health-related microbes.

**Subjects** Biodiversity, Evolutionary Studies, Microbiology

**Keywords** Historical demography, Coalescence, Cuatro cienegas, Bacterial population genetics, MLST

## INTRODUCTION

The present-day distribution of biodiversity is a consequence of the evolutionary dynamics of populations and the history of the regions in which they occur (Lomolino, Riddle & Brown, 2006). Lineage diversification results from interactions between the intrinsic

Submitted 19 May 2014

Accepted 22 November 2014

Published 16 December 2014

Corresponding author

Valeria Souza, souza@unam.mx

Academic editor

Frederick Cohan

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.696

© Copyright  
2014 Avitia et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**



biological constraints of organisms and extrinsic environmental factors (Dawson, 2012). The demographic patterns of coexisting populations have often been analyzed in the context of the history of a region to better understand how the evolution of resident lineages has been affected by their environment (Lessa, Cook & Patton, 2003; Carnaval & Moritz, 2008; Ramírez-Barahona & Eguiarte, 2013; Chan, Schanzenbach & Hickerson, 2014; Hope et al., 2014). The analysis of genetic traits through the application of coalescent theory has provided evidence of the impacts of past geological and environmental events on the demographic history of resident species, such as the effect of Quaternary glaciations upon the distribution of animal species in North America (Lessa, Cook & Patton, 2003; Hope et al., 2014).

It is clear that patterns of biodiversity across phylogenetically distinct taxa may be influenced by shared historical factors (Chan, Schanzenbach & Hickerson, 2014; Hope et al., 2014). Despite the importance of microorganisms in ecosystems (Allison & Martiny, 2008; Strickland et al., 2009), the scale of the impact of historical factors remains poorly understood and little is known about the population dynamics of natural microbial populations, which has greatly hindered our understanding of the processes determining diversity. Most studies on microbial population dynamics have analyzed demographic patterns of bacterial pathogens, undertaken primarily to understand disease epidemics and population expansion events of human pathogens (Pybus et al., 2001; Wirth et al., 2007; Tazi et al., 2010). Studies of demographic trends in natural bacterial populations have been scant (Guttman, Morgan & Wang, 2008), due in part to limited sampling of populations from different lineages at similar temporal and geographic scales.

In this context, the Cuatro Ciénegas Basin (CCB) harbors high levels of bacterial diversity, arguably due to environmental variation (e.g., salinity) across the aquatic systems within the basin (Cerritos et al., 2011). However, there is no conclusive evidence that identifies environmental conditions or geographic distance as predictors for the presence of certain bacterial groups (Escalante et al., 2008; Rebollar et al., 2012). Thus the influence of shared historical factors in population dynamics could be a plausible explanation for the observed diversity in CCB. A collection of coexisting bacterial isolates from the aquatic systems in this area has been built for over a decade (Cerritos et al., 2011; Rebollar et al., 2012; Rodríguez-Verdugo et al., 2012) and represents a unique opportunity to investigate the historical population patterns of coexisting bacterial lineages in a natural setting.

In the present work, we tested the hypothesis that a shared history in the CCB region has left common genetic signatures across phylogenetically diverse microbial populations. Using population genetics and coalescent-based approaches, we assessed the population history of lineages from two closely related genera of Firmicutes, *Bacillus* and *Exiguobacterium*, as well as lineages of Gammaproteobacteria from the genus *Pseudomonas*. The selected lineages are all common in the natural setting of the CCB aquatic system, and we anticipated that the genetic variation present among coexisting lineages would reveal a signature of common historical dynamics independent of phylogenetic relationships. Multilocus sequence typing (MLST) markers were used to (1) perform comparative phylogenetic analyses, (2) describe the genetic structure of bacterial populations,



(3) calculate descriptive parameters of genetic diversity, (4) perform neutrality tests, and (5) conduct historical demography reconstructions.

Consistent with our hypothesis, we identified a common trend of expansion in the studied populations over a similar time frame that occurred independently of phylogenetic relationships. These results provide empirical evidence that analyzing coexisting bacterial lineages in natural environments can advance our understanding of bacterial evolution, beyond medical or health-related species.

## MATERIALS & METHODS

### Study sites and sampling

Surface water and sediment samples were collected between 2003 and 2009 (Rodríguez-Verdugo *et al.*, 2012; Rebollar *et al.*, 2012) from four aquatic systems within the CCB: Churince (C), Los Hundidos (H), Mesquites (M), and Pozas Azules (Pa). *Pseudomonas* isolates were collected only from C. All isolates were obtained using previously described methods (Rebollar *et al.*, 2012). All samples were collected under the “Vida Silvestre” permit 0531 FAUT-0230 granted by the Mexican government agency: “Comisión Nacional de Áreas Protegidas” (CONANP).

Water and sediment samples were diluted at 1:10; 1:100; 1:1,000; and 1:10,000 using saline solution (1% NaCl). Subsequently, 200 µl of each dilution was plated using either marine agar (Difco 2216) for Firmicutes or GSP agar (Kielwein, 1971) for *Pseudomonas*. Firmicutes isolates were incubated for 24 h at 37 °C; *Bacillus* and *Exiguobacterium* colonies were pale or bright orange, respectively. All orange colonies were purified by single-colony isolation. *Pseudomonas* isolates were incubated for 48 h at 30 °C. *Pseudomonas* colonies were selected based on a change in the color of the GSP media to purple. All isolates were stored at −80 °C in 20% (w/v) glycerol.

### Molecular markers and sequencing

DNA was extracted from all isolates using a DNAeasy Tissue Kit (Qiagen, Valencia, CA, USA), following the manufacturer’s instructions. The 16S rRNA gene was amplified and sequenced using 27F and 1492R primers, according to the conditions described by Lane (1991). PCR products were sequenced and used to identify isolates. Once a strain was identified as *Bacillus*, *Exiguobacterium* or *Pseudomonas*, we amplified and sequenced a set of housekeeping genes commonly used for MLST (Cerritos *et al.*, 2011; Yamamoto *et al.*, 2000; Sarkar & Guttman, 2004; Rodrigues *et al.*, 2006). Genus-specific PCR primers (Table S1) were used. All PCR products were sequenced at the University of Washington’s High Throughput Genomics Center. Sequences were deposited in the GenBank database with the following accession numbers: *Bacillus citC* (KC900996–KC901178); *Bacillus gltx* (JQ241465–JQ241624); *Bacillus hsp70* (KC901179–KC901361); *Bacillus recA* (JQ247793–JQ247952); *Bacillus spo0A* (KC901362–KC901540); *Exiguobacterium citC* (JF916988–JF917080, JF952020–JF952109); *Exiguobacterium hsp70* (JF952111–JF952292); *Exiguobacterium recA* (JF952293–JF952475); *Exiguobacterium rpoB* (JF952476–JF952658);



*Pseudomonas acnB* (KC953704–KC953753); *Pseudomonas gyrB* (KC920532–KC920576); *Pseudomonas recA* (KC961435–KC961462); *Pseudomonas rpoD* (KC920481–KC920531).

### Phylogenetic reconstructions

The complete sequences of the *16S rRNA* gene and partial sequences of the MLST genes were edited and aligned using BioEdit (Hall, 1999). Representative *16S rRNA* sequences for the three genera were obtained from GenBank and were included in the alignments as references (see Figs. S1A–S1C for accession numbers).

We calculated the Pairwise Homoplasy Index ( $\Phi_W$ ) using the “PHI test recombination” function implemented in SplitsTree4 (Huson & Bryant, 2006) to verify clonality for all lineages as has been reported for *Bacillus*, *Exiguobacterium* and *Pseudomonas* (Roberts & Cohan, 1995; Spiers, Buckling & Rainey, 2000; Rebollar et al., 2012). Since all lineages are clonal (Table S2), we were able to use concatenated data sets for phylogenetic reconstruction and population genetics analyses.

For all genera, maximum likelihood (ML) phylogenies were constructed using (i) *16S rRNA* sequences and (ii) concatenated alignments of MLST sequences. The program jModelTest v.2.1.3 was used to determine the best nucleotide substitution model for each alignment (Posada, 2008; Darriba et al., 2012). For the *16S rRNA* gene, TPM2uf+I+G, TIM1+I+G and GTR+I+G were the substitution models selected for *Bacillus*, *Exiguobacterium* and *Pseudomonas*, respectively. GTR+I+G, TIM2+I+G and GTR+I+G were the models selected for the *Bacillus*, *Exiguobacterium*, and *Pseudomonas* concatenated MLST loci alignments, respectively. Phylogenetic relationships were reconstructed using PhyML v.3.0 (Guindon & Gascuel, 2003), with the corresponding DNA substitution models, tree improvement was carried out by Subtree Pruning and Regrafting, and branch support was evaluated by 1,000 bootstrap pseudo-replicates.

### Population structure

We defined populations as the minimal study unit in our collection. Populations have to be defined considering both geographic and genetic criteria. Given that it is possible to find populations (genetic pools) that are not site-restricted we investigated population structure by genetic differentiation, which is defined as changes in the frequency distribution of haplotype variants among subpopulations (Hartl & Clark, 1997). We estimated pairwise  $F_{ST}$  values among sampling sites within lineages. With this approach we defined a single population as all isolates from a lineage that exhibited no significant genetic differentiation as measured by pairwise  $F_{ST}$  estimates. Pairwise  $F_{ST}$  estimates were obtained for each monophyletic lineage across all sampling sites using Arlequin 3.5 (Excoffier & Lischer, 2010), with 1,000 iterations. For the *Pseudomonas* isolates,  $F_{ST}$  analysis was not performed since all isolates were from the same sampling site.

The  $F_{ST}$  approach implicitly incorporates geographic location and genetic criteria in the investigation of population structure. However, it is possible that genetic structure exists for each lineage when looking at individual genetic variants within each population. Thus, we investigated potential substructure within the groups defined by  $F_{ST}$  by taking a Bayesian approach implemented in BAPS 6 (Tang et al., 2009). The Bayesian analysis of

population structure was performed using the option for linked loci, specifically developed for MLST data (Tang *et al.*, 2009). A maximum number of clusters (K) was set to ten, or equal to the number of individuals if these were fewer than ten. Each analysis was replicated ten times.

Once populations were defined, standard measures of nucleotide diversity ( $\pi$ ) and the mutation parameter Watterson's  $\theta$  were estimated, together with neutrality tests (Tajima's  $D$ , Fu and Li's  $F^*$  and  $D^*$  and Fu's  $F_S$ ) using DNAsp v4.1 (Rozas *et al.*, 2003). Details of these calculations can be found in the [Supplemental Information](#).

### Historical population dynamics

We performed an Extended Bayesian Skyline Plot (EBSP) analysis as implemented in Beast v.1.7.5 (Drummond & Rambaut, 2007; Heled & Drummond, 2008). The EBSP infers past population dynamics from a sample of contemporary sequences, taking into account the genealogical stochasticity of the coalescent (Ho & Shapiro, 2011; Drummond *et al.*, 2005). Additionally, this method does not depend on a specific *a priori* demographic model, but infers the number of population size changes directly from the data. As a result, it provides a measure of statistical credibility of the inferred number of population size changes compared to the alternative of constant population size (Heled & Drummond, 2008).

For the EBSP analysis, we used all MLST genes, a strict molecular clock, and the same nucleotide substitution models used for phylogenetic reconstruction to estimate changes in population size for each genetically homogeneous population. The substitution rate used for the two Firmicutes genera (*Bacillus* and *Exiguobacterium*) was  $7 \times 10^{-9}$  substitutions/per site/generation, obtained from a dated phylogenetic tree (Maughan, 2007). The substitution rate used for *Pseudomonas* was  $2.8 \times 10^{-8}$  substitutions/per site/generation, based on an experimental evolution study of *P. fluorescens* (Barrett, MacLean & Bell, 2006). All time estimates obtained were expressed in number of generations. Changes in population size were expressed as a function of the product of  $N_e$  and the generation time ( $N_e * t$ ). All analyses were run for 50–100 million generations, until adequate mixing was achieved. Ten percent of burn-in was removed and the sampling was done every 1,000 chains. The rest of the parameters were set according to the guidelines of Heled & Drummond (2008). Results were analyzed with Tracer v.1.5 and LogCombiner v.1.7.5 (Drummond & Rambaut, 2007).

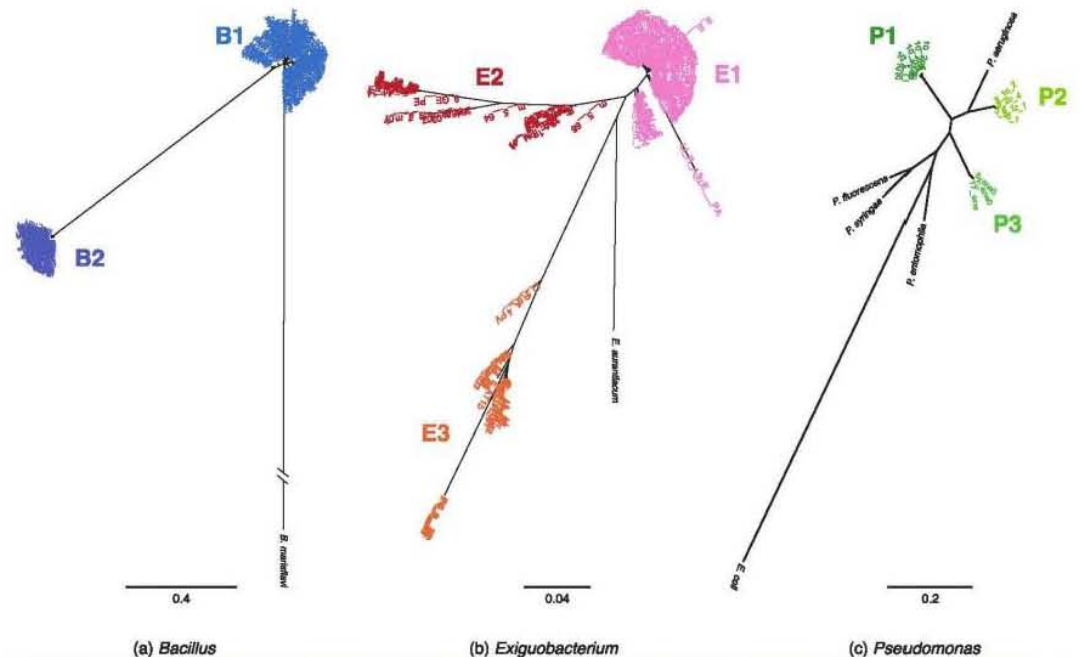
## RESULTS

To investigate the influence of a common habitat in the genetic variation of bacterial populations, we analyzed MLST data from a collection of isolates across three bacterial genera (*Bacillus*, *Exiguobacterium* and *Pseudomonas*) sampled from water bodies in the geographic region of CCB, Mexico.

### The phylogenetic history of CCB lineages

Phylogenetic reconstructions<sup>117</sup> based on MLST and 16S *rRNA* sequences were used to assign isolates to monophyletic lineages (Figs. 1 and S1 respectively). MLST phylogenies





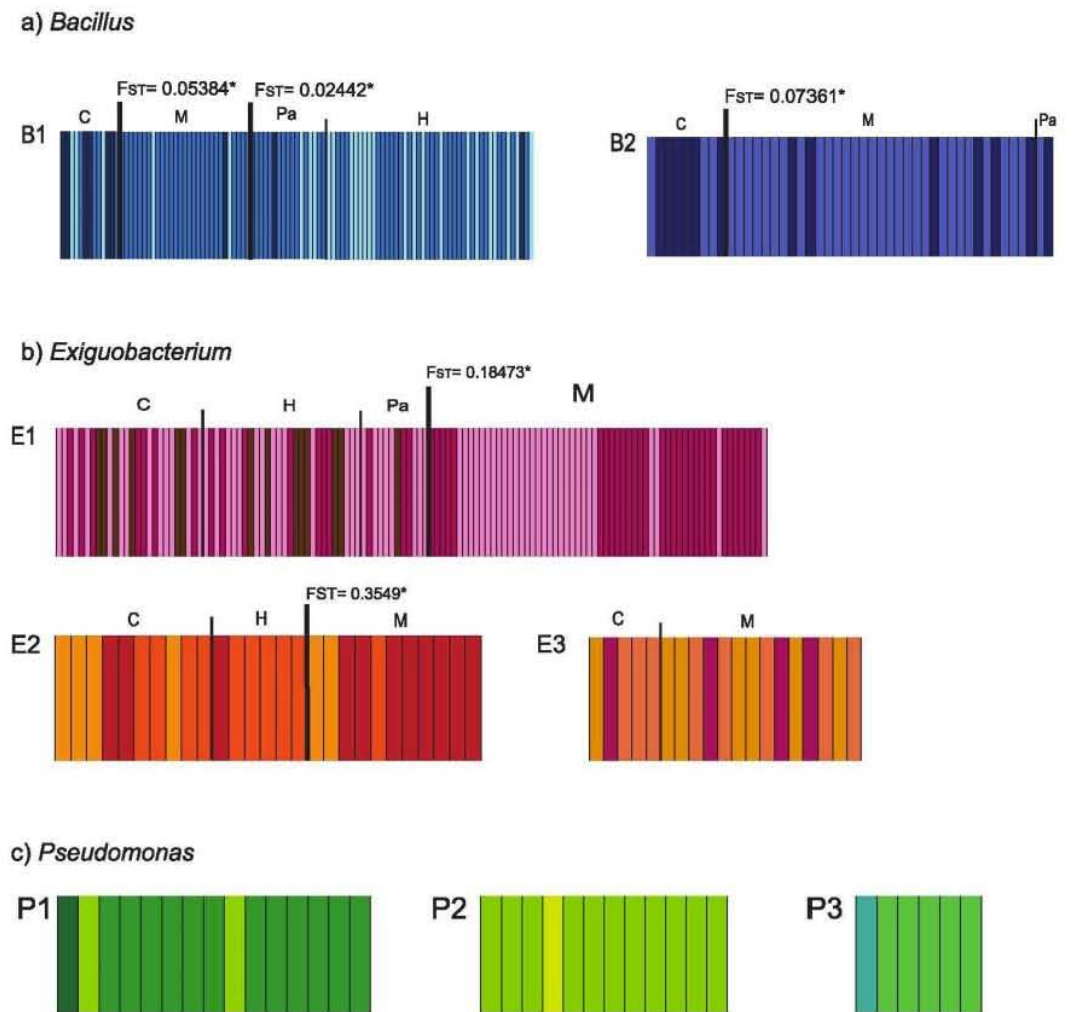
**Figure 1** MLST phylogenetic reconstructions of isolates collected in the Cuatro Ciénegas Basin (CCB), Mexico. Phylogenies were constructed using the maximum likelihood approach and are based on concatenated sequences of at least four MLST loci. (A) *Bacillus* lineages with *B. marisflavi* as outgroup, and *citC*, *gltX*, *hsp70*, *recA* and *spo0A* loci, (B) *Exiguobacterium* lineages with *E. aurantiacum* as outgroup, and *citC*, *hsp70*, *recA* and *rpoB* loci, and (C) *Pseudomonas* lineages with *Escherichia coli* as outgroup, and *acnB*, *gyrB*, *recA* and *rpoD* loci. The scale of the bar represents the number of substitutions per site.

identified at least two lineages within each genus. We named the lineages according to genus: “B” for *Bacillus*, “E” for *Exiguobacterium*, and “P” for *Pseudomonas*.

Among *Bacillus* isolates, we identified two clusters, B1 and B2 (Fig. 1A); their closest relatives were *B. aquimaris*, *B. vietnamensis*, *B. marisflavi*, and *B. coahuilensis* (Fig. S1A). *Exiguobacterium* isolates clustered into three well-defined groups: E1, E2, and E3 (Fig. 1B). For clusters E1 and E2, the most closely related species was *E. aurantiacum*, and *E. profundum* was the closest relative of cluster E3. *Pseudomonas* isolates were divided into three groups: P1, P2, and P3 (Fig. 1C). P1 included the majority of isolates and, according to *16S rRNA* sequences, was closely related to *P. oryzihabitans*. P2 isolates were closely related to *P. otitidis* and *P. aeruginosa*, while P3 had recently been described as a new species, *P. cuatrocienegasensis* (Escalante et al., 2009) (Fig. S1C).

### Population structure

Populations defined by pairwise  $F_{ST}$  values corresponded with sampling-site pairs with non-significant  $F_{ST}$  values (Table S3). We named the resulting populations indicating lineage and sampling site as: B1\_C, B1\_HPa, B1\_M, B2\_C, B2\_MPa, E1\_CHPa, E1\_M, E2\_CH, E2\_M, E3\_CM, P1\_C, P2\_C, and P3\_C. Further exploration of structure within lineages was performed with BAPS and showed clusters that are shared among sites within lineages. In all cases, differences in frequency distribution of haplotypes (clusters) correspond with the population structure that we identified through pairwise  $F_{ST}$  (Fig. 2).



**Figure 2** Bayesian Analysis of Population Structure (BAPS) of isolates collected in the Cuatro Ciéngas Basin (CCB), Mexico. The analysis was conducted for each studied lineage within each genus (*Bacillus* = B1, B2; *Exiguobacterium* = E1, E2, E3; *Pseudomonas* = P1, P2, P3) using the option for linked loci. The maximum number of clusters (K) was set to 10, or equal to the number of individuals if these were fewer than 10. Each analysis was replicated 10 times. The columns represent a single multilocus genotype and are identified with distinctive colors corresponding to different genetic clusters. Different populations are graphically denoted with a dark and wide line between populations with significant  $F_{ST}$  values, which are also shown.

Therefore, we delimited populations based on  $F_{ST}$  estimates since it incorporates both the geographic and genetic components in the definition of populations.

### Population history: population genetics and coalescent approaches

Results of the genetic diversity analyses and neutrality tests (including Fu's  $F_S$ ; Fu, 1997) are presented in Table S4. Neutrality tests suggested expansion events but were inconclusive. In our data set, most values for Tajima's D and Fu & Li's tests were negative, although non-significant, which is suggestive of expansion events of populations. However, Fu's  $F_S$ ,



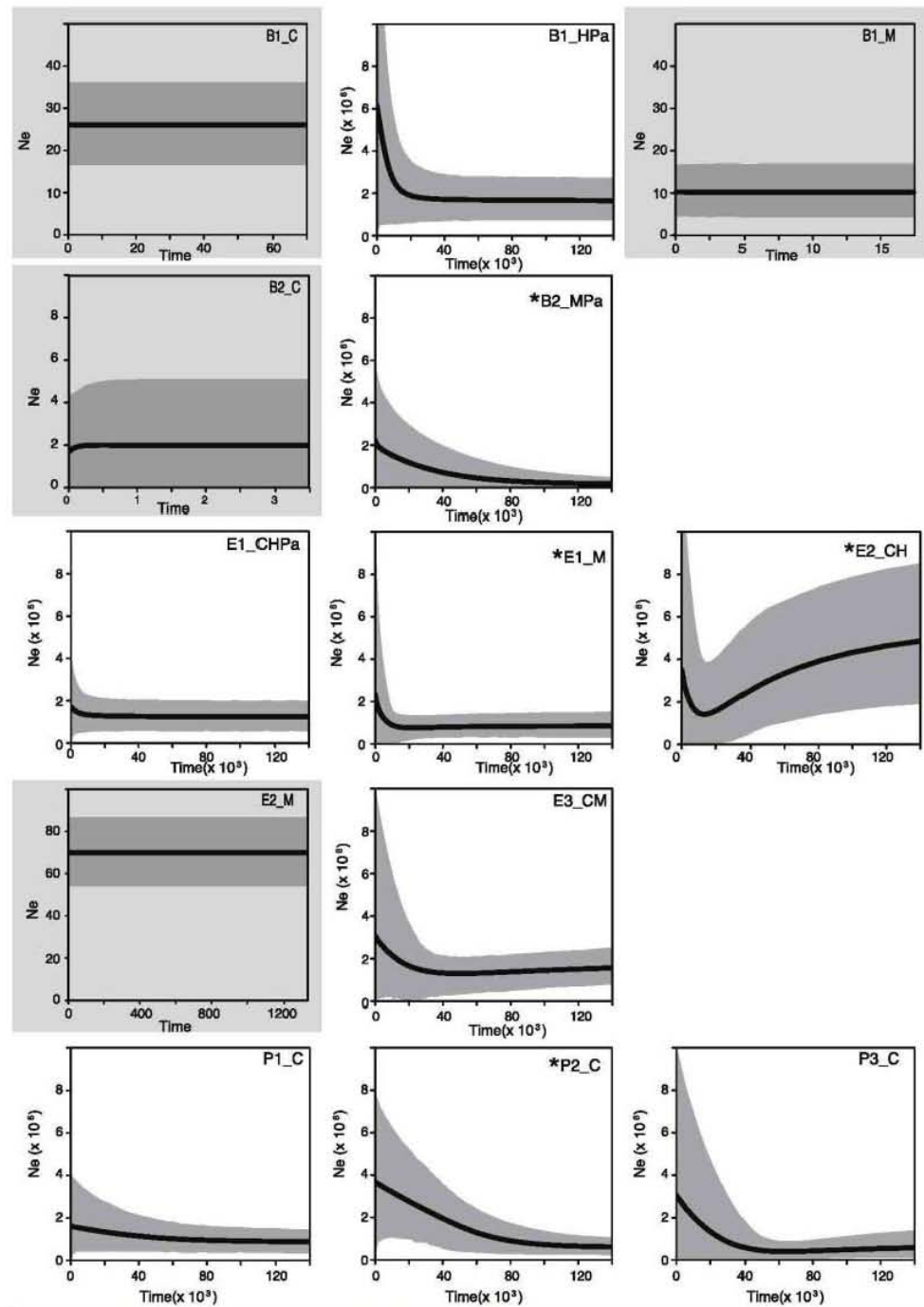
a test that was explicitly designed to evaluate population expansions, was not significant in most cases. Overall, these classic population genetics estimates did not allow for strong conclusions about historical population events.

To further explore the influence of historical processes in shaping the diversity of these bacterial populations, we used a coalescent approach. Extended Bayesian Skyline Plot (EBSP) analyses of populations (genetic pools) showed signals of expansion in 9 of the 13 populations, of which 4 have statistical support for at least one change in population size (Fig. 3; for statistical support values of the population size changes see Table S4). Within *Bacillus*, an expansion (although not statistically supported) was detected in one population (B1\_HPα) that included most B1 isolates. Plots for the other two B1 populations (B1\_M and B1\_C) showed constant population size across generations (Fig. 3). The B2 plots showed statistical support for change in population size in at least one population (B2\_MPα) and constant population size in another (B2\_C). Within *Exiguobacterium*, there was evidence of expansion in one of two E1 populations (E1\_M), in one of the two E2 populations (E2\_CH) and a trend of expansion (not statistically supported) in the only E3 population (E3\_CM). Finally, the plots for all *Pseudomonas* populations displayed signals of expansion, although only P2\_C was statistically supported. Overall, a trend of expansion was observed for the studied populations independent of phylogenetic relationships or sampling sites. Moreover, the skyline plots revealed a common time frame for the start of expansions (20,000 to 30,000 generations ago, Fig. 3), regardless of lineage- or locus-specific parameters.

## DISCUSSION

Diversity patterns in nature are a consequence of the evolutionary dynamics of populations and the history of the regions in which populations occur (Lomolino, Riddle & Brown, 2006). A powerful approach to unveil the influence of historical factors is to analyze current genetic variation to reconstruct the demographic history of populations (Lessa, Cook & Patton, 2003; Hope et al., 2014). In microorganisms, genetic studies looking at historical demography have mainly focused on host-associated bacteria to gain insight in disease epidemics (Mhedbi-Hajri et al., 2013; Wirth et al., 2006; Comas et al., 2013; Holt et al., 2013). However, little is known about population dynamics of non-host associated and naturally occurring bacteria.

When analyzing bacterial populations caution should be taken in the delimitation of populations, which has to consider key aspects such as the degree of genetic and ecological diversity. It is also important to consider the molecular resolution at which populations are defined (Kopac & Cohan, 2011). The more variable loci studied, the more populations may be identified. In our case, few and highly conserved loci may result in a coarse population delimitation. Nonetheless, to include both genetic and ecological aspects in the delimitation of populations, we defined individual populations for each lineage by analyzing genetic structure across sampling sites. The analysis showed that populations defined by the  $F_{ST}$  statistic include different sampling sites with contrasting environmental conditions (Rebollar et al., 2012), which is in accordance with previous work where no



**Figure 3** Bayesian skyline plots for lineages collected in the Cuatro Ciénegas Basin (CCB), Mexico. Showing the effective population size through time (black line). The shaded area represents 95% credibility intervals. Labels denote the lineage and sampling sites of each population. B, E, and P stand for *Bacillus*, *Exiguobacterium*, and *Pseudomonas*, respectively. C, H, M, and Pa stand for the sampling sites Churince, Los Hundidos, Mesquites, and Pozas Azules, respectively. Shaded plots represent populations with constant population sizes. Asterisks indicate populations with statistically supported demographic changes (Table S5).



evidence of geographic or environmental barriers to dispersal exists (Escalante *et al.*, 2008; Rebollar *et al.*, 2012). We explored the existence of subpopulations within the  $F_{ST}$ -defined populations following a Bayesian approach implemented in BAPS (Tang *et al.*, 2009), and we found more clusters that subdivided even further the populations defined with  $F_{ST}$  estimates. However, genetic diversity within the identified populations ( $\pi = 0.00032\text{--}0.04698$  and Watterson's  $\theta = 0.00073\text{--}0.047$ ; Table S4) is in the range of values reported for natural bacterial populations (Roberts & Cohan, 1995; Vos & Velicer, 2006), meaning that the observed diversity is within the range of normal polymorphism in bacterial populations. Overall,  $F_{ST}$ , BAPS, and diversity estimates support our population delimitation and indicate that we are not including unusually divergent individuals into the studied populations which might confound the population genetics and expansion estimations. Given that classic population genetics estimates did not allow for strong conclusions about historical population events, we used a coalescent approach to further explore the influence of historical processes. From this approach, we found evidence showing that common historical events were population expansions that occurred in 9 out of 13 populations (Fig. 3). Even more, it is noteworthy that at least one population of each genus presented evidence of expansion that was statistically supported. In clonal and free-living populations of bacteria, evolutionary models predict that populations experience strong fluctuations, either by selective sweeps or by metapopulation dynamics (Fraser *et al.*, 2009; Kopac & Cohan, 2011; Shapiro & Polz, 2014). The results obtained in this study can be explained by these models since the analyzed lineages come from a natural setting and present no evidence of recombination (Table S2; Roberts & Cohan, 1995; Spiers, Buckling & Rainey, 2000; Rebollar *et al.*, 2012). Without recombination, clonal clusters emerge by random mutation, but can quickly drift to extinction unless they have a major selective advantage, in which case their presence could suggest the emergence of new ecologically distinct populations (Shapiro & Polz, 2014) or ecotypes (Kopac & Cohan, 2011). Nonetheless, given that we do not have conclusive evidence that identifies environmental conditions or geographic distance as predictors for the presence of certain bacterial groups or populations (Escalante *et al.*, 2008; Rebollar *et al.*, 2012), the influence of shared historical factors in population dynamics could be a plausible explanation for the observed population expansion events in CCB. Given that we used just a few loci to delimit the studied populations, it should be noted here that we cannot rule out the possibility that the observed expansions may be a reflection of the emergence of closely related populations (undetectable to our molecular markers), and not population expansions in size. This ambiguity could be resolved if more variable loci, even complete genomes, are analyzed. Despite the alternative interpretations of the results, we consider notable the fact that we identified the same type of events (either diversification or population expansions) in most populations in the same time frame.

Population expansions in our data were identified to occur during the same time frame ( $\cong 20,000$  generations ago). Our estimation of time to the most recent population change is calculated in number of generations. Although generation times may vary considerably among bacterial lineages (Mason, 1934; Errington, Powell & Thompson,



1965), laboratory observations of growth rates of the lineages included in this study do not suggest major differences in generation times between them, which supports the synchrony argument. In this study, we used previously reported estimates for Firmicutes (*Bacillus* and *Exiguobacterium*) and *Pseudomonas* to parameterize the coalescent simulations. We used the same substitution rate for *Bacillus* and *Exiguobacterium* since little information is available for *Exiguobacterium* and also because [Maughan \(2007\)](#) did not detect a considerable rate of heterogeneity between sporulating and non-sporulating Firmicutes. In this respect, it should be acknowledged that values of substitution rates are critical for the estimation of the temporal scale for the demographic events that result from ESPB analyses ([Ho et al., 2011](#)) and ideally should be directly estimated for the studied sample. We are aware that this could impact the temporal scales of our skyline plots, and be of particular relevance when comparing lineages with different substitution rates, but we do not have any reason to doubt the substitution rates reported in the literature. Further work should try to directly estimate substitution rates for the studied lineages to minimize the possibility of different time scales for the expansion events.

Despite the fact that expansions are expected in bacterial populations, the time of such expansions may differ among lineages because they depend on specific traits such as mutation rate, life history, and population structure ([Avice, 2000](#); [Avice, 2009](#); [Kuo & Ochman, 2009](#); [Maughan, 2007](#); [Woolfit & Bromham, 2003](#)). The synchrony of expansions in coexisting populations of different lineages are not expected to be observed unless shared environmental and historical factors have a major influence on population dynamics obscuring evidence of lineage specific adaptation, as has been observed in different organisms including “human associated” pathogens ([Ornelas et al., 2013](#); [Comas et al., 2013](#); [Falush et al., 2003](#)). The synchrony of population expansions supports the argument that current genetic variation may be the result of a major regional event over all populations.

Currently, we do not have the necessary information or samples from the appropriate temporal and spatial scales to determine the environmental changes associated with the historical events that influenced population dynamics of the studied lineages. However, we can propose an approximate timespan in which these demographic events occurred and hypothesize on the environmental causes behind our observations. Most estimates of generation times for bacteria have been determined in laboratory conditions, which may be considerably shorter than generation times in the wild (for *Escherichia coli* see [Pierucci, 1972](#); [Ochman, Elwyn & Moran, 1999](#)). Thus, we estimated the approximate absolute time of population expansions based on generation times for aquatic bacteria estimated in conditions that are similar to their natural habitat conditions (approximately between 75 and 300 generations per year; [Jannasch, 1969](#); [Hendricks, 1972](#)). We selected a generation time at the lower limit (75 gen/yr), because it is possible that the bacteria in our study may grow slowly, given the extreme limitation of nutrients, in particular phosphorous, found in the CCB aquatic system ([Souza et al., 2006](#); [Souza et al., 2012](#); [Souza et al., 2008](#)). Thus, if the synchronic event occurred 200–300 years ago for most populations, it is tempting to suggest that the population expansions we observed reflect changes resulting from the



recent human activity in the CCB area. In particular, an increase in anthropogenic activity during the late eighteenth and nineteenth centuries has been reported, primarily centered around agriculture and ranching (*Alessio-Robles, 1946*) which coincides with the time frame of expansions found here.

## CONCLUSIONS

This study investigates historical population events of coexisting but phylogenetically distant bacterial lineages. To our knowledge, this is the first report of a synchronous signature of population expansion that occurred independently of phylogenetic relationships, which is in accordance with our initial hypothesis. Our findings provide strong evidence for the potential of comparative population genetics to address questions about the influence of shared historical factors in the population evolutionary processes of naturally occurring non-host associated bacteria. This information may be important for natural resource management in the context of the ecosystem services that microorganisms provide.

## ACKNOWLEDGEMENTS

We thank Dr. Evan Carson, Dr. Michael Travisano, Dr. Daniel Piñero, Dr. Juan Pablo Jaramillo and Dr. Santiago Ramírez-Barahona for their constructive reviews of the manuscript. We also thank Laura Espinosa, Dr. Erika Aguirre and Jaime Gasca-Pineda for technical support during the development of the project.

This paper constitutes part of the doctoral research of the first author, who thanks the Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and the Consejo Nacional de Ciencia y Tecnología (CONACyT).

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The first author received a scholarship and financial support from the Consejo Nacional de Ciencia y Tecnología (CONACyT, grant no. 210335). This work constitutes part of the doctoral research of the first author who received scholarship and financial support from the Consejo Nacional de Ciencia y Tecnología (CONACyT, grant no. 210335). The project was supported by a grant from WWF-Alianza Carlos Slim. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:  
Consejo Nacional de Ciencia y Tecnología (CONACyT): 210335.  
WWF-Alianza Carlos Slim.

### Competing Interests<sup>124</sup>

Valeria Souza and Luis E. Eguiarte are Academic Editors for PeerJ.

## Author Contributions

- Morena Avitia, Ana E. Escalante and Eria A. Rebolgar conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Alejandra Moreno-Letelier conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Luis E. Eguiarte and Valeria Souza conceived and designed the experiments, contributed reagents/materials/analysis tools, reviewed drafts of the paper.

## Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

All samples were collected under the “Vida Silvestre” permit 0531 FAUT-0230 granted by the Mexican government agency: “Comisión Nacional de Áreas Protegidas” (CONANP).

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Sequences were deposited in the GenBank database with the following accession numbers: *Bacillus citC* (KC900996–KC901178), *Bacillus gltx* (JQ241465–JQ241624), *Bacillus hsp70* (KC901179–KC901361), *Bacillus recA* (JQ247793–JQ247952), *Bacillus spo0A* (KC901362–KC901540), *Exiguobacterium citC* (JF916988–JF917080, JF952020–JF952109), *Exiguobacterium hsp70* (JF952111–JF952292), *Exiguobacterium recA* (JF952293–JF952475), *Exiguobacterium rpoB* (JF952476–JF952658), *Pseudomonas acnB* (KC953704–KC953753), *Pseudomonas gyrB* (KC920532–KC920576), *Pseudomonas recA* (KC961435–KC961462), *Pseudomonas rpoD* (KC920481–KC920531).

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.696#supplemental-information>.

## REFERENCES

- Alessio-Robles M. 1946. *Historia Política de La Revolución*. Mexico: Botas.
- Allison SD, Martiny JBH. 2008. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* **105**:11512–11519 DOI [10.1073/pnas.0801925105](https://doi.org/10.1073/pnas.0801925105).
- Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge, Massachusetts: Harvard University Press.
- Avise JC. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**:3–15 DOI [10.1111/j.1365-2699.2008.02032.x](https://doi.org/10.1111/j.1365-2699.2008.02032.x).



- Barrett RDH, MacLean RC, Bell G. 2006. Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. *Biology Letters* 2(2):236–238 DOI 10.1098/rsbl.2006.0439.
- Carnaval AC, Moritz C. 2008. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* 35(7):1187–1201 DOI 10.1111/j.1365-2699.2007.01870.x.
- Cerritos R, Eguiarte LE, Avitia M, Siefert J, Travisano M, Rodríguez-Verdugo A, Souza V. 2011. Diversity of culturable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, México. *Antonie van Leeuwenhoek* 99(2):303–318 DOI 10.1007/s10482-010-9490-9.
- Chan YL, Schanzenbach D, Hickerson MJ. 2014. Detecting concerted demographic response across community assemblages using hierarchical approximate Bayesian computation. *Molecular Biology and Evolution* 31(9):2501–2515 DOI 10.1093/molbev/msu187.
- Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. 2013. Out-of-Africa migration and neolithic coexpansion of mycobacterium tuberculosis with modern humans. *Nature Genetics* 45(10):1176–1182 DOI 10.1038/ng.2744.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9:772 DOI 10.1038/nmeth.2109.
- Dawson MN. 2012. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. *Molecular Ecology* 21(4):987–1004 DOI 10.1111/j.1365-294X.2011.05417.x.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7:214 DOI 10.1186/1471-2148-7-214.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22(5):1185–1192 DOI 10.1093/molbev/msi103.
- Errington FP, Powell E, Thompson N. 1965. Growth characteristics of some gram-negative bacteria. *Journal of General Microbiology* 39:109–123 DOI 10.1099/00221287-39-1-109.
- Escalante AE, Caballero-Mellado J, Martínez-Aguilar L, Rodríguez-Verdugo A, González-González A, Toribio-Jiménez J, Souza V. 2009. *Pseudomonas* *cuatrociénegasensis* Sp. Nov., isolated from an evaporating lagoon in the Cuatro Cienegas Valley in Coahuila, Mexico. *International Journal of Systematic and Evolutionary Microbiology* 59(Pt 6):1416–1420 DOI 10.1099/ijs.0.006189-0.
- Escalante AE, Eguiarte LE, Espinosa-Asuar L, Forney LJ, Noguez AM, Saldivar VS. 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.
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Molecular Ecology Resources* 10(3):564–567 DOI 10.1111/j.1755-0998.2010.02847.x.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S. 2003. Traces of human migrations in helicobacter pylori populations. *Science* 299(5562):1582–1585 DOI 10.1126/science.1080857.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The Bacterial species challenge: ecological diversity. *Science* 323:741–746 DOI 10.1126/science.1159388.

- Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**(5):696–704 DOI [10.1080/10635150390235520](https://doi.org/10.1080/10635150390235520).
- Guttman DS, Morgan RL, Wang PW. 2008. The evolution of the pseudomonads. In: Fami MB, Collmer A, Iacobellis NS, Mansfield JW, Murillo J, Schaad NW, Ullrich M, eds. *Pseudomonas syringae pathovars and related pathogens—identification, epidemiology and genomics*. Springer Science & Business Media BV, 307–319.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.
- Hartl DL, Clark AG. 1997. *Principles of population genetics*. 3rd ed. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**:289 DOI [10.1186/1471-2148-8-289](https://doi.org/10.1186/1471-2148-8-289).
- Hendricks CW. 1972. Enteric bacterial growth rates in river water. *Applied Microbiology* **24**(2):168–174.
- Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011. Time-dependent rates of molecular evolution. *Molecular Ecology* **20**(15):3087–3101 DOI [10.1111/j.1365-294X.2011.05178.x](https://doi.org/10.1111/j.1365-294X.2011.05178.x).
- Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources* **11**(3):423–434 DOI [10.1111/j.1755-0998.2011.02988.x](https://doi.org/10.1111/j.1755-0998.2011.02988.x).
- Holt KE, Vu T, Nga T, Pham D, Vinh H, Wook D, Phan M, Tra V. 2013. Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proceedings of the National Academy of Sciences of the United States of America* **110**(43):17522–17527 DOI [10.1073/pnas.1308632110](https://doi.org/10.1073/pnas.1308632110).
- Hope AG, Ho SYW, Malaney JL, Cook JA, Talbot SL. 2014. Accounting for rate variation among lineages in comparative demographic analyses. *Evolution; International Journal of Organic Evolution* **68**(9):2689–2700 DOI [10.1111/evo.12469](https://doi.org/10.1111/evo.12469).
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**(2):254–267 DOI [10.1093/molbev/msj030](https://doi.org/10.1093/molbev/msj030).
- Jannasch HW. 1969. Estimations of bacterial growth rates in natural waters. *Journal of Bacteriology* **99**(1):156–160.
- Kielwein G. 1971. Die Isolierung Und Differenzierung von Pseudomonaden Aus Lebensmitteln. *Archiv fur Lebensmittelhygiene* **22**:29–37.
- Kopac S, Cohan FM. 2011. A theory-based pragmatism for discovering and classifying newly divergent bacterial species. In: Tibayrenc M, ed. *Genetics and evolution of infectious diseases*. Burlington, MA: Elsevier, 21–41.
- Kuo C-H, Ochman H. 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biology Direct* **4**:35–44 DOI [10.1186/1745-6150-4-35](https://doi.org/10.1186/1745-6150-4-35).
- Lane DJ. 1991. 16S/23S rDNA sequencing. Nucleic acid techniques. In: Stackebrandt E, Goodfellow M, eds. *Bacterial systematics*. New York: John Wiley & Sons, 115–175.
- Lessa EP, Cook JA, Patton J. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the late quaternary. *Proceedings of the National Academy of Sciences of the United States of America* **100**(18):10331–10334 DOI [10.1073/pnas.1730921100](https://doi.org/10.1073/pnas.1730921100).



- Lomolino MV, Riddle BR, Brown JH. 2006.** *Biogeography*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Mason MM. 1934.** A comparison of the maximal growth rates of various bacteria under optimal conditions. *Journal of Bacteriology* **29**:103–110.
- Maughan H. 2007.** Rates of molecular evolution in bacteria are relatively constant despite spore dormancy. *Evolution* **61**(2):280–288 DOI [10.1111/j.1558-5646.2007.00026.x](https://doi.org/10.1111/j.1558-5646.2007.00026.x).
- Mhedbi-Hajri N, Hajri A, Boureau T, Darrasse A, Durand K, Brin C, Saux MF-L, Manceau C, Poussier S, Pruvost O, Lemaire C, Jacques MA. 2013.** Evolutionary history of the plant pathogenic bacterium *xanthomonas axonopodis*. *PLoS ONE* **8**(3):e58474 DOI [10.1371/journal.pone.0058474](https://doi.org/10.1371/journal.pone.0058474).
- Ochman H, Elwyn S, Moran NA. 1999.** Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences of the United States of America* **96**(22):12638–12643 DOI [10.1073/pnas.96.22.12638](https://doi.org/10.1073/pnas.96.22.12638).
- Ornelas JF, Sosa V, Soltis DE, Daza JM, González C, Soltis PS, Gutiérrez-Rodríguez C, de los Monteros AE, Castoe TA, Bell C, Ruiz-Sanchez E. 2013.** Comparative phylogeographic analyses illustrate the complex evolutionary history of threatened cloud forests of Northern Mesoamerica. *PLoS ONE* **8**(2):e56283 DOI [10.1371/journal.pone.0056283](https://doi.org/10.1371/journal.pone.0056283).
- Pierucci O. 1972.** Chromosome replication and cell division in *Escherichia coli* at various temperatures of growth. *Journal of Bacteriology* **109**(2):848–854.
- Posada D. 2008.** jModeltest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**(7):1253–1256 DOI [10.1093/molbev/msn083](https://doi.org/10.1093/molbev/msn083).
- Pybus OG, Charleston MA, Gupta S, Rambaut A, Holmes EC, Harvey PH. 2001.** The epidemic behavior of the hepatitis c virus. *Science* **292**(5525):2323–2325 DOI [10.1126/science.1058321](https://doi.org/10.1126/science.1058321).
- Ramírez-Barahona S, Eguiarte LE. 2013.** The role of glacial cycles in promoting genetic diversity in the neotropics: the case of cloud forests during the last glacial maximum. *Ecology and Evolution* **3**(3):725–738 DOI [10.1002/ece3.483](https://doi.org/10.1002/ece3.483).
- Rebollar EA, Avitia M, Eguiarte LE, 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 Ciénegas basin. *Environmental Microbiology* **14**(9):2323–2333 DOI [10.1111/j.1462-2920.2012.02784.x](https://doi.org/10.1111/j.1462-2920.2012.02784.x).
- Roberts MS, Cohan FM. 1995.** Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49**:1081–1094 DOI [10.2307/2410433](https://doi.org/10.2307/2410433).
- Rodrigues DF, Goris J, Vishnivetskaya T, Gilichinsky D, Thomashow MF, Tiedje JM. 2006.** Characterization of *Exiguobacterium* isolates from the Siberian Permafrost. Description of *Exiguobacterium sibiricum* sp. nov. *Extremophiles* **10**(4):285–294 DOI [10.1007/s00792-005-0497-5](https://doi.org/10.1007/s00792-005-0497-5).
- Rodríguez-Verdugo A, Souza V, Eguiarte LE, Escalante AE, Roda A. 2012.** Diversity across seasons of culturable *Pseudomonas* from a desiccation lagoon in Cuatro Ciénegas, Mexico. *International Journal of Microbiology* **2012**:Article 201389 DOI [10.1155/2012/201389](https://doi.org/10.1155/2012/201389).
- Rozas J, Sanchez-DelBarrio JC, Messeguier X, Rozas R. 2003.** DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**(18):2496–2497 DOI [10.1093/bioinformatics/btg359](https://doi.org/10.1093/bioinformatics/btg359).
- Sarkar SF, Guttman DS. 2004.** Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Applied and Environmental Microbiology* **70**(4):1999–2012 DOI [10.1128/AEM.70.4.1999](https://doi.org/10.1128/AEM.70.4.1999).

- Shapiro BJ, Polz MF. 2014.** Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in Microbiology* 22(5):235–247 DOI 10.1016/j.tim.2014.02.006.
- Souza V, Eguiarte LE, Siefert J, Elser JJ. 2008.** Microbial endemism: does phosphorus limitation enhance speciation? *Nature Reviews Microbiology* 6(7):559–564 DOI 10.1038/nrmicro1917.
- Souza V, Espinosa-Asuar L, Escalante AE, Eguiarte LE, Farmer J, Forney L, Lloret L, Rodriguez-Martinez JM, Soberon X, Dirzo R, Elser JJ. 2006.** An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proceedings of the National Academy of Sciences of the United States of America* 103(17):6565–6570 DOI 10.1073/pnas.0601434103.
- Souza V, Siefert JL, Escalante AE, Elser JJ, Eguiarte LE. 2012.** The cuatro ciénegas basin in Coahuila, Mexico: an astrobiological precambrian park. *Astrobiology* 12(7):641–647 DOI 10.1089/ast.2011.0675.
- Spiers AJ, Angus B, Rainey PB. 2000.** The causes of pseudomonas diversity. *Microbiology* 146:2345–2350.
- Strickland MS, Lauber C, Fierer N, Bradford MA. 2009.** Testing the functional significance of microbial community composition. *Ecology* 90(2):441–451 DOI 10.1890/08-0296.1.
- Tang J, Hanage WP, Fraser C, Corander J. 2009.** Identifying currents in the gene pool for bacterial populations using an integrative approach. *PLoS Computational Biology* 5(8):e1000455 DOI 10.1371/journal.pcbi.1000455.
- Tazi L, Pérez-Losada M, Gu W, Yang Y, Xue L, Crandall KA, Viscidi RP. 2010.** Population dynamics of neisseria gonorrhoeae in Shanghai, China: a comparative study. *BMC Infectious Diseases* 10:13 DOI 10.1186/1471-2334-10-13.
- Vos M, Velicer GJ. 2006.** Genetic population structure of the soil bacterium Myxococcus Xanthus at the centimeter scale. *Applied and Environmental Microbiology* 72(5):3615–3625 DOI 10.1128/AEM.72.5.3615-3625.2006.
- Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MCJ, Ochman H, Achtman M. 2006.** Sex and virulence in escherichia coli: an evolutionary perspective. *Molecular Microbiology* 60(5):1136–1151 DOI 10.1111/j.1365-2958.2006.05172.x.
- Wirth T, Morelli G, Kusecek B, Van Belkum A, Van der Schee C, Meyer A, Achtman M. 2007.** The rise and spread of a new pathogen: seroresistant moraxella catarrhalis. *Genome Research* 17:1647–1656 DOI 10.1101/gr.6122607.7.
- Woolfit M, Bromham L. 2003.** Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution* 20(9):1545–1555 DOI 10.1093/molbev/msg167.
- Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S. 2000.** Phylogeny of the genus pseudomonas: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. *Microbiology* 146:2385–2394.

## SUPPLEMENTARY INFORMATION

### Materials and Methods

#### Genetic diversity and neutrality tests

Once populations were defined, genetic diversity analyses for populations within each lineage and gene were performed using DNAsp v4.1 (Rozas et al., 2003). Standard measures of nucleotide diversity ( $\pi$ ) and the mutation parameter Waterson's  $\theta$  were estimated, together with neutrality tests (Tajima's  $D$ , Fu and Li's  $F^*$  and  $D^*$ , as well as  $F_S$ ). Briefly, both  $\pi$  and Waterson's  $\theta$  are measures of intraspecific genetic variation. The former is based on pairwise comparisons of nucleotide diversity and can be viewed as the average level of polymorphism, offering a description of the current condition of the population. By contrast, Waterson's  $\theta$  is based on the number of segregating sites present and incorporates the effective population size ( $N_e$ ) and thus it can be used to gain insight into historical demography (Hedrick, 2011). Tajima's  $D$  (Tajima, 1989) evaluates discrepancies between  $\pi$  and Waterson's  $\theta$ . Under the Wright-Fisher neutral model (or mutation-drift model),  $\pi$  and Waterson's  $\theta$  are expected to be equal; deviations from neutrality can give information about population history, serving as a first approach to demographic events. Tests developed by Fu & Li (1993) analyze whether the frequency of singleton variants (restricted to the external or terminal branches (Hartl & Clark, 1997) are consistent with the mutation-drift expectation (Fu & Li 1993) and are considered a more direct method of analyzing background selection. In addition, Fu's  $F_S$  (Fu, 1997) is sensitive to population

growth and has been used as supporting evidence for demographic processes (Lessa, Cook & Patton, 2003). Significance for all neutrality tests was assessed using 10,000 coalescent simulations based on the observed number of segregating sites (Rozas et al., 2003).

## Figures and tables

**Figure S1. 16S rRNA phylogenetic reconstructions of lineages studied.** Phylogenies were constructed using the maximum likelihood approach and the complete sequence of the 16S rRNA gene. **(a)** *Bacillus* lineages with *Geobacillus jurassicus* as outgroup, **(b)** *Exiguobacterium* lineages with *E. auranticum* as outgroup, and **(c)** *Pseudomonas* lineages with *Escherichia coli* as outgroup. Bold font denotes representative sequences of the lineages studied. The scale of the bar represents number of substitutions per site.

**Table S1.** PCR Primers used to amplify MLST loci in the three lineages studied.

**Table S2. Pairwise Homoplasy Index ( $\Phi_w$ ) P values.** Values were obtained for all the housekeeping loci analyzed for each studied lineage using “PHI test recombination” function implemented in Splits Tree4 (Huson & Bryant, 2006). Bold fonts denote significant P values, which account for recombination. *NC* denotes non-computable



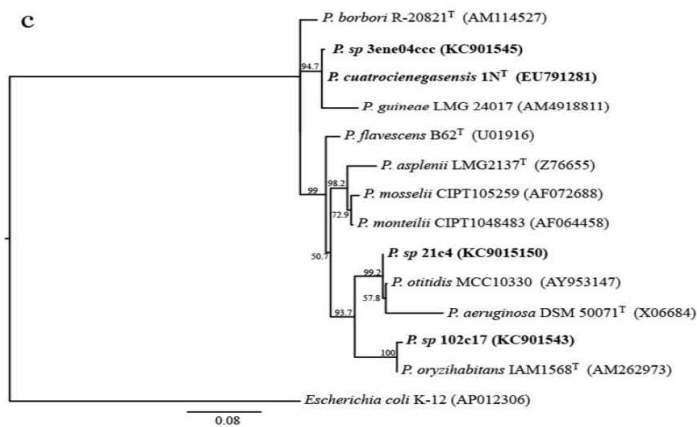
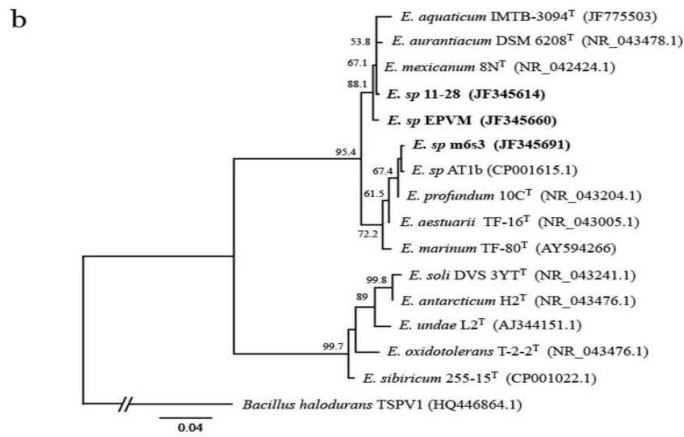
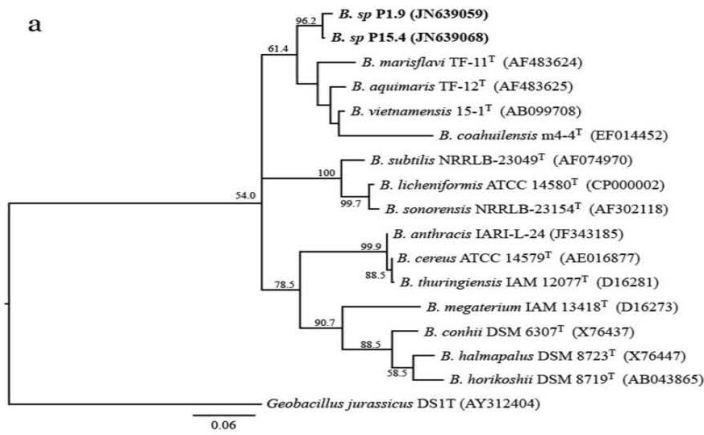
estimations due to small number of informative sites. E1, E2 and E3 values were previously published in Rebollar et al. (2012).

**Table S3.** Pairwise  $F_{ST}$  estimates. Values are shown for *Exiguobacterium* and *Bacillus* lineages. Bold font denotes significant  $F_{ST}$  values ( $P < 0.05$ ).

**Table S4. Summary statistics for genetic diversity and neutrality parameters for each population at each locus.**  $\pi$  is Nei's pairwise nucleotide diversity value and  $\theta$  is Watterson's theta per site =  $2Ne\pi$ ; neutrality was tested using Tajima's D, Fu & Li's D\* and F\*, and Fu's  $F_S$ . Bold fonts denote significant values ( $P < 0.05$ ) for neutrality tests (Rozas et al., 2003). N = number of sequences used in each analysis.

**Table S5.** Credibility intervals (95%) for estimates of number of population size changes. Values different from 0 (statistically supported expansions) are shown in bold.

Figure S1



**Table S1.** PCR Primers used to amplify MLST loci in the three lineages studied.

| Genus                  | Locus        | Primers | T <sub>m</sub> (°C) | Sequence (5'-3')                | PCR product size (bp) | Reference                                 |
|------------------------|--------------|---------|---------------------|---------------------------------|-----------------------|---|
| <i>Exiguobacterium</i> | <i>citC</i>  | citC-F  | 49                  | GGDGAYGGMACWGGWCCWG<br>AYATTTGG | 1,165                 | Rodrigues, 2006                           |
|                        |              | citC-R  |                     | AATTCWGAACATTTMACTTCT<br>GT     |                       | Rodrigues, 2006                           |
|                        | <i>hsp70</i> | hsp70-F | 47                  | GGTATTGAYTTAGGAACAACA<br>AACT   | 1455                  | Rodrigues, 2006                           |
|                        |              | hsp70-R |                     | CTTCTGCWTCCTTKACCAT             |                       | Rodrigues, 2006                           |
|                        | <i>recA</i>  | recA-F  | 47                  | GARAARCAATTYGGBAAAGGT<br>TC     | 869                   | Rodrigues, 2006                           |
|                        |              | recA-R  |                     | TGTTYGMATTTTCACGKCCTT<br>G      |                       | Rodrigues, 2006                           |
|                        | <i>rpoB</i>  | rpoB-F  | 53                  | CGAACATGCAACGTCAGGC             | 1078                  | Rodrigues, 2006                           |
|                        |              | rpoB-R  |                     | ACATCYTCYTACGNGCACC             |                       | Rodrigues, 2006                           |
| <i>Pseudomonas</i>     | <i>acnB</i>  | acn-Fp  | 60                  | ACATCCCGCTGCACGCYCTGG<br>CC     | 700                   | Sarkar and Guttman 2004                   |
|                        |              | acn-Rp  |                     | GTGGTGTCTGGGAACCGACG<br>GTG     |                       | Sarkar and Guttman 2004                   |
|                        | <i>gyrB</i>  | gyrB_F  | 60                  | ACCAYGSNGGNGNAARTTYR<br>A       | 1000                  | Modified from Yamamoto <i>et al.</i> 2000 |
|                        |              | gyrB_R  |                     | AGTGCNGGRTCYYTTYCYTGR<br>CA     |                       | Modified from Yamamoto <i>et al.</i> 2000 |
|                        | <i>recA</i>  | recA-F  | 60                  | CAGATCGAAMRNCARTTCGG            | 900                   | This study <sup>a</sup>                   |

|                 |              |          |           |                                 |       |   |
|-----------------|--------------|----------|-----------|---------------------------------|-------|---|
|                 |              | recA-R   |           | YTTRCCCTGDCCGATCTT              |       | This study <sup>a</sup>                   |
|                 | <i>rpoD</i>  | 70F      | 63        | TGTAYATGMGNGARATGGGN<br>ACNGT   | 850   | Modified from Yamamoto <i>et al.</i> 2000 |
|                 |              | 70R      |           | GTTNGCYTCNACCATYTCYTT<br>YTT    |       | Modified from Yamamoto <i>et al.</i> 2000 |
| <i>Bacillus</i> | <i>citC</i>  | citC-F   | 52-54.5   | GGDGAYGGMACWGGWCCWG<br>AYATTTGG | 1,165 | Rodrigues, 2006                           |
|                 |              | citC-R   |           | AATTCWGAACATTTMACTTCT<br>GT     |       | Rodrigues, 2006                           |
|                 | <i>gltX</i>  | gltx-F   | 52        | CGYGGBGADGAYCAYATYT             | 488   | This study <sup>a</sup>                   |
|                 |              | gltX-R   |           | CRATTCMGCDCRWARCT               |       | This study <sup>a</sup>                   |
|                 | <i>hsp70</i> | hsp70-F  | 52.5-54.5 | GGTATTGAYTTAGGAACAACA<br>AACT   | 1455  | Rodrigues, 2006                           |
|                 |              | hsp70-R  |           | CTTCTGCWTCITTKACCAT             |       | Rodrigues, 2006                           |
|                 | <i>recA</i>  | recA-F   | 51        | GATCGTCARGCAGSCYTWGAT           | 574   | This study <sup>a</sup>                   |
|                 |              | recA-R   |           | TTWCCRACCATAACSCCRAC            |       | This study <sup>a</sup>                   |
|                 | <i>spo0A</i> | spo0A-F1 | 49-52     | GATGAYAATCGCGAGYTKGTH           | 505   | This study <sup>a</sup>                   |
|                 |              | spo0A-F2 |           | TWDTDGATGAYAATCGNGAR<br>Y       | 511   | This study <sup>a</sup>                   |
|                 |              | spo0A-R  |           | KARATAACCYTTATTRTGGGC           |       | This study <sup>a</sup>                   |

<sup>a</sup> Primers specifically designed for this study from sequences obtained from GenBank.

**Table S2. Pairwise Homoplasly Index ( $\Phi_w$ ) P values** obtained for all the housekeeping loci analyzed for each studied lineage using “PHI test recombination” function implemented in Splits Tree4 (Huson & Bryant, 2006). Bold fonts denote significant P values, which account for recombination. NC denotes non-computable estimations due to small number of informative sites. E1, E2 and E3 values were previously published in Rebollar et al. (2012).

|                        | Genetic marker | $\Phi_w$ P value |              |
|------------------------|----------------|------------------|--------------|
| <i>Bacillus</i>        | <b>B1</b>      | <i>citC</i>      | 0.58         |
|                        |                | <i>gltX</i>      | 0.375        |
|                        |                | <i>hsp70</i>     | 0.199        |
|                        |                | <i>recA</i>      | 0.254        |
|                        |                | <i>spo0A</i>     | 0.261        |
|                        | <b>B2</b>      | <i>citC</i>      | 0.057        |
|                        |                | <i>gltX</i>      | NC           |
|                        |                | <i>hsp70</i>     | 1            |
|                        |                | <i>recA</i>      | 1            |
|                        |                | <i>spo0A</i>     | 1            |
| <i>Exiguobacterium</i> | <b>E1</b>      | <i>citC</i>      | 0.317        |
|                        |                | <i>hsp70</i>     | 0.927        |
|                        |                | <i>recA</i>      | 1            |
|                        |                | <i>rpoB</i>      | 1            |
|                        | <b>E2</b>      | <i>citC</i>      | <b>0.001</b> |
|                        |                | <i>hsp70</i>     | 0.184        |
|                        |                | <i>recA</i>      | <b>0.028</b> |

|                    |             |              |              |
|--------------------|-------------|--------------|--------------|
| <i>Pseudomonas</i> | <b>E3</b>   | <i>rpoB</i>  | 0.828        |
|                    |             | <i>citC</i>  | 0.346        |
|                    |             | <i>hsp70</i> | 0.29         |
|                    |             | <i>recA</i>  | 0.404        |
|                    |             | <i>rpoB</i>  | 0.108        |
|                    | <b>P1</b>   | <i>acnB</i>  | 0.9157       |
|                    |             | <i>gyrB</i>  | <b>0.039</b> |
|                    |             | <i>recA</i>  | 0.769        |
|                    |             | <i>rpoD</i>  | 1            |
|                    | <b>P2</b>   | <i>acnB</i>  | 0.123        |
|                    |             | <i>gyrB</i>  | 0.891        |
|                    |             | <i>recA</i>  | 0.064        |
|                    |             | <i>rpoD</i>  | 0.189        |
| <b>P3</b>          | <i>acnB</i> | 1            |              |
|                    | <i>gyrB</i> | NC           |              |
|                    | <i>recA</i> | NC           |              |
|                    | <i>rpoD</i> | NC           |              |

**Table S3.** Pairwise  $F_{ST}$  estimates for *Exiguobacterium* and *Bacillus* lineages. Bold font denotes significant  $F_{ST}$  values ( $P < 0.05$ ).

***Bacillus* B1**

|              | Churince       | Los Hundidos   | Mesquites      | Pozas Azules |
|--------------|----------------|----------------|----------------|--------------|
| Churince     | 0              |                |                |              |
| Los Hundidos | <b>0.13122</b> | 0              |                |              |
| Mesquites    | <b>0.05384</b> | <b>0.05815</b> | 0              |              |
| Pozas Azules | <b>0.06417</b> | 0.01184        | <b>0.02442</b> | 0            |

***Bacillus* B2**

|              | Churince       | Mesquites | Pozas Azules |
|--------------|----------------|-----------|--------------|
| Churince     | 0              |           |              |
| Mesquites    | <b>0.07361</b> | 0         |              |
| Pozas Azules | -0.08534       | -0.06249  | 0            |

***Exiguobacterium* E1**

|          | Churince | Los Hundidos | Mesquites | Pozas Azules |
|----------|----------|--------------|-----------|--------------|
| Churince | 0        |              |           |              |

|              |                |                |   |                |
|--------------|----------------|----------------|---|----------------|
| Los Hundidos | -0.00447       | 0              |   |                |
| Mesquites    | <b>0.15847</b> | <b>0.19081</b> | 0 | <b>0.18473</b> |
| Pozas Azules | <b>0.07145</b> | 0.04933        |   | 0              |

---

***Exiguobacterium E2***

|              | Churince       | Los Hundidos  | Mesquites |
|--------------|----------------|---------------|-----------|
| Churince     | 0              |               |           |
| Los Hundidos | 0.05388        | 0             |           |
| Mesquites    | <b>0.19335</b> | <b>0.3549</b> | 0         |

---

***Exiguobacterium E3***

|           | Churince | Mesquites |
|-----------|----------|-----------|
| Churince  | 0        |           |
| Mesquites | -0.01191 | 0         |

---



**Table S4. Summary statistics for genetic diversity and neutrality parameters for each population**

**at each locus.**  $\pi$  is Nei's pairwise nucleotide diversity value and  $\theta$  is Watterson's theta per site =  $2N\mu$ ; neutrality was tested using Tajima's D, Fu & Li's D\* and F\*, and Fu's  $F_S$ . Bold fonts denote significant values ( $P < 0.05$ ) for neutrality tests (Rozas et al., 2003). N = number of sequences used in each analysis.

|                 |               | Genetic marker (N) | $\pi$   | $\theta$ | D               | D*              | F*              | $F_S$         |
|-----------------|---------------|--------------------|---------|----------|-----------------|-----------------|-----------------|---------------|
| <b>Bacillus</b> | <b>B1_C</b>   | <i>citC</i> (13)   | 0.00634 | 0.00659  | -0.14865        | 0.20065         | 0.12398         | -0.001        |
|                 |               | <i>gltx</i> (13)   | 0.02418 | 0.01665  | 1.93412         | 1.55213         | 1.88975         | 6.931         |
|                 |               | <i>hsp70</i> (13)  | 0.00676 | 0.00917  | -1.09708        | -1.08285        | -1.24064        | -1.684        |
|                 |               | <i>recA</i> (13)   | 0.00579 | 0.00546  | 0.24002         | 0.95617         | 0.87547         | -2.14         |
|                 |               | <i>spo0A</i> (13)  | 0.00677 | 0.00992  | -1.29439        | -1.33265        | -1.50892        | -0.415        |
|                 | <b>B1_HPα</b> | <i>citC</i> (63)   | 0.00636 | 0.00911  | -0.93631        | <b>-2.8527</b>  | <b>-2.57788</b> | <b>-9.203</b> |
|                 |               | <i>gltx</i> (63)   | 0.01236 | 0.01481  | -0.52713        | -0.37074        | -0.50997        | -1.966        |
|                 |               | <i>hsp70</i> (63)  | 0.00388 | 0.00765  | <b>-1.5085</b>  | -0.78297        | -1.23732        | <b>-5.974</b> |
|                 |               | <i>recA</i> (63)   | 0.00482 | 0.00889  | -1.42751        | <b>-3.78042</b> | <b>-3.49337</b> | -3.499        |
|                 |               | <i>spo0A</i> (63)  | 0.00792 | 0.01796  | <b>-1.81687</b> | <b>-4.26053</b> | <b>-4.00576</b> | -4.839        |
|                 | <b>B1_M</b>   | <i>citC</i> (29)   | 0.00584 | 0.00677  | -0.45529        | -0.68539        | -0.71918        | -3.595        |
|                 |               | <i>gltx</i> (29)   | 0.0124  | 0.01842  | -1.18276        | -1.76197        | -1.85484        | -1.998        |
|                 |               | <i>hsp70</i> (29)  | 0.00459 | 0.00918  | <b>-1.7444</b>  | -1.6035         | -1.93523        | <b>-7.155</b> |
|                 |               | <i>recA</i> (29)   | 0.00212 | 0.00432  | <b>-1.59693</b> | -0.36113        | -0.86063        | <b>-4.852</b> |

|                        |                   |                   |         |                 |                 |                 |                 |                 |
|------------------------|-------------------|-------------------|---------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <i>Exiguobacterium</i> |                   | <i>spo0A</i> (29) | 0.00976 | 0.0111          | -0.41454        | -0.88609        | -0.8659         | <b>-13.094</b>  |
|                        | <b>B2_C</b>       | <i>citC</i> (9)   | 0.00329 | 0.00376         | -0.52629        | 0.04471         | -0.10194        | -1.338          |
|                        |                   | <i>gltx</i> (9)   | 0.00129 | 0.00095         | 0.98627         | 0.8404          | 0.96219         | 0.849           |
|                        |                   | <i>hsp70</i> (9)  | 0.00285 | 0.00279         | 0.07803         | 0.50315         | 0.44642         | -0.208          |
|                        |                   | <i>recA</i> (9)   | 0.00149 | 0.00226         | -0.88249        | -0.7022         | -0.88249        | -2.106          |
|                        |                   | <i>spo0A</i> (9)  | 0.001   | 0.00094         | 0.15647         | 0.8404          | 0.74837         | 0.477           |
|                        | <b>B2_MPa</b>     | <i>citC</i> (37)  | 0.0029  | 0.00392         | -0.75486        | -0.02029        | -0.29221        | <b>-5.51971</b> |
|                        |                   | <i>gltx</i> (37)  | 0.00135 | 0.00124         | 0.17481         | -0.80429        | -0.60451        | 0.285           |
|                        |                   | <i>hsp70</i> (37) | 0.001   | 0.00182         | -1.09093        | -1.06705        | -1.25012        | -2.126          |
|                        |                   | <i>recA</i> (37)  | 0.00126 | 0.00208         | -1.51297        | -1.68268        | -1.82046        | -0.38           |
|                        |                   | <i>spo0A</i> (37) | 0.00216 | 0.00307         | -0.77205        | -0.7022         | -0.84352        | <b>-4.305</b>   |
|                        | <b>E1_CHPa</b>    | <i>citC</i> (69)  | 0.00499 | 0.01563         | <b>-2.25866</b> | <b>-6.22055</b> | <b>-5.61834</b> | 1.465           |
|                        |                   | <i>hsp70</i> (69) | 0.00544 | 0.02513         | <b>-2.66051</b> | <b>-5.27413</b> | <b>-5.08687</b> | -1.285          |
|                        |                   | <i>recA</i> (68)  | 0.01998 | 0.02007         | -0.01436        | -1.7318         | -1.2783         | 1.678           |
|                        |                   | <i>rpoB</i> (68)  | 0.00128 | 0.00533         | -2.22123        | <b>-4.3899</b>  | <b>-4.3083</b>  | -5.028          |
|                        | <b>E1_M</b>       | <i>citC</i> (63)  | 0.00599 | 0.00375         | 1.69688         | -0.80621        | 0.0581          | 3.985           |
| <i>hsp70</i> (64)      |                   | 0.00127           | 0.00279 | -1.45722        | <b>-2.82187</b> | <b>-2.79375</b> | -2.175          |                 |
| <i>recA</i> (62)       |                   | 0.01022           | 0.0382  | <b>-2.54278</b> | <b>-6.32914</b> | <b>-5.81342</b> | 2.901           |                 |
| <i>rpoB</i> (64)       |                   | 0.00113           | 0.00326 | <b>-1.74904</b> | <b>-4.20979</b> | <b>-4.00043</b> | -2.069          |                 |
| <b>E2_CH</b>           | <i>citC</i> (16)  | 0.03332           | 0.02845 | 0.73142         | 0.5417          | 0.68842         | 7.808           |                 |
|                        | <i>hsp70</i> (16) | 0.04698           | 0.03728 | 1.1183          | 1.20473         | 1.36438         | 7.98            |                 |
|                        | <i>recA</i> (16)  | 0.0249            | 0.02303 | 0.34381         | 0.42753         | 0.4665          | 3.896           |                 |
|                        | <i>rpoB</i> (16)  | 0.02074           | 0.02193 | -0.22758        | -0.50261        | -0.49062        | 0.955           |                 |

|                    |                  |                   |         |                 |                 |                 |                 |                |
|--------------------|------------------|-------------------|---------|-----------------|-----------------|-----------------|-----------------|----------------|
| <i>Pseudomonas</i> | E2_M             | <i>citC</i> (12)  | 0.02912 | 0.03077         | -0.24782        | -0.23508        | -0.2719         | 8.152          |
|                    |                  | <i>hsp70</i> (12) | 0.04065 | 0.04328         | -0.28249        | -0.1994         | -0.25244        | 10.151         |
|                    |                  | <i>recA</i> (12)  | 0.0229  | 0.02467         | -0.3278         | -0.27522        | -0.32949        | 1.598          |
|                    |                  | <i>rpoB</i> (12)  | 0.01538 | 0.01778         | -0.60888        | -0.19975        | -0.34979        | 4.372          |
|                    | E3_CM            | <i>citC</i> (19)  | 0.02814 | 0.0169          | 2.68291         | 2.683           | 2.08208         | 10.933         |
|                    |                  | <i>hsp70</i> (19) | 0.0169  | 0.00932         | 0.53515         | -0.24331        | -0.01875        | 1.08           |
|                    |                  | <i>recA</i> (19)  | 0.02018 | 0.01753         | 0.60746         | 1.0072          | 1.03414         | 6.199          |
|                    |                  | <i>rpoB</i> (19)  | 0.00471 | 0.047           | 0.0054          | -0.45778        | -0.37705        | -1.132         |
|                    | P1_C             | <i>acnB</i> (16)  | 0.01041 | 0.01874         | <b>-1.85733</b> | <b>-2.05057</b> | <b>-2.3048</b>  | <b>-4.793</b>  |
|                    |                  | <i>gyrB</i> (15)  | 0.02787 | 0.02237         | 1.065           | 0.74481         | 0.96143         | 2.495          |
|                    |                  | <i>recA</i> (5)   | 0.03062 | 0.03349         | -0.647          | -0.96758        | -0.97904        | 0.717          |
|                    |                  | <i>rpoD</i> (16)  | 0.00426 | 0.00556         | -0.88876        | -1.56786        | -1.58834        | 1.2            |
|                    | P2_C             | <i>acnB</i> (26)  | 0.00909 | 0.01869         | <b>-1.94317</b> | <b>-2.64783</b> | <b>-2.84909</b> | <b>-13.503</b> |
|                    |                  | <i>gyrB</i> (15)  | 0.00816 | 0.00867         | -0.2464         | -0.16294        | -0.21393        | -2.4           |
|                    |                  | <i>recA</i> (17)  | 0.01131 | 0.01487         | -0.96926        | -1.81228        | -1.8181         | <b>-5.679</b>  |
|                    |                  | <i>rpoD</i> (28)  | 0.00341 | 0.00451         | -0.76648        | -0.34344        | -0.55018        | -2.886         |
| P3_C               | <i>acnB</i> (7)  | 0.02005           | 0.01953 | 0.15214         | 0.61134         | 0.56146         | -1.169          |                |
|                    | <i>gyrB</i> (15) | 0.00032           | 0.00073 | -1.49051        | -1.87275        | -2.0195         | 0.235           |                |
|                    | <i>recA</i> (6)  | 0.00248           | 0.00326 | -1.36732        | -1.39992        | -1.48974        | <b>-2.112</b>   |                |
|                    | <i>rpoD</i> (7)  | 0.00093           | 0.00132 | <b>-1.23716</b> | -1.29591        | -1.37408        | -0.922          |                |

**Table S5.** Credibility intervals (95%) for estimates of number of population size changes. Values different from 0 (statistically supported expansions) are shown in bold.

| Lineage_population | Mean         | 95% HPD lower | 95% HPD higher |
|--------------------|--------------|---------------|----------------|
| E1_CHPa            | 0.428        | 0             | 2              |
| <b>E1_M</b>        | <b>2.3</b>   | <b>1</b>      | <b>4</b>       |
| E2_M               | 0            | na            | na             |
| <b>E2_CH</b>       | <b>1.94</b>  | <b>1</b>      | <b>3</b>       |
| E3_CM              | 1.16         | 0             | 3              |
| B1_HPpa            | 0.679        | 0             | 2              |
| B1_M               | 2.60E-2      | 0             | 0              |
| B1_C               | 0            | na            | na             |
| <b>B2_MPpa</b>     | <b>1.47</b>  | <b>1</b>      | <b>3</b>       |
| B2_C               | 0.786        | 0             | 3              |
| P1_C               | 0.817        | 0             | 2              |
| <b>P2_C</b>        | <b>1.355</b> | <b>1</b>      | <b>3</b>       |
| P3_C               | 1.58         | 0             | 3              |

## Capítulo 6. CONCLUSIONES Y PERSPECTIVAS

Para poder comprender la gran diversidad tanto genética como ecológica que existe en los procariontes es necesario entender los mecanismos que han generado dicha diversidad. Entendiendo que la diversidad es resultado de un proceso de divergencia evolutiva, es importante entender los procesos de generación de diversidad a nivel poblacional, para poder también entender los mecanismos de divergencia genética de los linajes y la especiación en procariontes. Un paso importante para lograrlo es ampliar nuestro conocimiento sobre poblaciones de bacterias de vida libre, porque es en los ambientes naturales donde se encuentra la mayor parte de la diversidad procarionte.

Tomando en cuenta lo anterior, el trabajo presentado en esta tesis tuvo como motivación principal ampliar el estudio de poblaciones ambientales de bacterias. En particular, a partir de una colección amplia de aislados recuperada de los cuatro principales sistemas acuáticos dentro del valle de CCC, se exploró la genética de poblaciones y evolución molecular de un grupo de bacterias del género *Bacillus*. Se describió la diversidad genética de las poblaciones de *Bacillus* del valle a diferentes niveles buscando determinar los factores que han moldeado dicha diversidad. En cada uno de los tres capítulos en los que se condensan los resultados de este trabajo se discuten los hallazgos particulares. Sin embargo, a

continuación se presentan y discuten brevemente las conclusiones generales derivadas del conjunto de los análisis.

La primera conclusión derivada de este trabajo es que el grupo de *Bacillus* estudiado representa una especie dentro de dicho género que no ha sido previamente descrita en la literatura. Además de ser una especie nueva dentro del género *Bacillus*, los análisis filogenéticos permitieron determinar que el linaje descrito es filogenéticamente cercano a especies de *Bacillus* aisladas de ambientes marinos. Este resultado es consistente con trabajos previos que muestran que la mayor parte de los linajes bacterianos de CCC tienen una afiliación con linajes marinos (Souza et al., 2006; Cerritos et al., 2008; Rebollar et al., 2012). En el trabajo de Rebollar et al., (2012) se pudo determinar, a través del análisis detallado del contenido de sales contenidas en el agua de CCC, que las proporciones iónicas de las muestras del valle son muy similares a las proporciones del agua de mar. Es posible que estas condiciones ambientales hayan permitido el desarrollo de linajes bacterianos en CCC con rasgos similares a los linajes marinos. Sin embargo, para poder entender los mecanismos evolutivos que han permitido la existencia de estos linajes bacterianos en CCC sería necesario contar con información genómica de varios de estos aislados para poder llevar a cabo estudios comparativos y poder determinar las bases genéticas de la adaptación a estos ambientes similares a los marinos.

El siguiente resultado importante de este trabajo es que la divergencia entre los dos filogrupos identificados en la muestra estudiada no parece tener relación con el sitio de aislamiento de las cepas ni con el ambiente particular del cual fueron obtenidas (agua/sedimento) pero sí con la recombinación actuando como fuerza cohesiva al interior de cada filogrupo. Este resultado es contrastante con lo obtenido en el trabajo de Rebollar *et al* (2012) en el que se encontró que los linajes del género *Exiguobacterium* también aislados de CCC muestran un patrón de diferenciación asociada a los ambientes de agua y sedimento. Si bien es notable el hecho de que estos patrones contrastantes se den en dos géneros de bacterias filogenéticamente relacionados, el hecho de que los miembros del género *Bacillus* sean capaces de formar esporas podría explicar en cierta medida el hecho de que las distribuciones de los filogrupos dentro del valle no parecen estar determinadas por los factores ambientales relacionados con nichos particulares (agua/sedimento).

Los análisis de las contribuciones relativas que la mutación y la recombinación han tenido en generar los patrones de diversidad genética observados en los grupos de *Bacillus* estudiados mostraron que la recombinación es frecuente al interior de los filogrupos y prácticamente inexistente entre

filogrupos. Los valores de  $r/m$  obtenidos con estos análisis son similares a los obtenidos para otros grupos de *Bacillus* (Vos y Didelot, 2009) y se considera que están en el rango de los valores “intermedios” de  $r/m$ , considerando que valores altos de  $r/m$  son los que están por encima de 10. Este resultado es consistente con los resultados de los análisis de desequilibrio de ligamiento, que apoyan la idea de que, a pesar de que existe cierto nivel de recombinación al interior de los filogrupos de *Bacillus* estudiados en este trabajo, la frecuencia de los eventos de recombinación no es tan alta como para que se rompa el ligamiento entre alelos y las poblaciones de dichos filogrupos se comporten como poblaciones totalmente panmícticas.

El resultado más importante presentado en esta tesis se desprende del trabajo comparativo de las poblaciones de *Bacillus* con otros géneros de bacterias de CCC, presentado en el capítulo cinco. En este trabajo, además de llevar a cabo un análisis fino de la estructura de las poblaciones estudiadas se pudo concluir que los eventos históricos compartidos han jugado un papel importante en moldear la diversidad genética de las poblaciones de bacterias estudiadas. La distribución de la diversidad es una consecuencia tanto de la dinámica evolutiva de las poblaciones como de la historia de las regiones geográficas en las que habitan (Lomolino et al., 2016). Así, la diversificación de los linajes resulta de las interacciones entre las características biológicas de cada organismo y los factores



ambientales externos (Dawson, 2012). Los patrones demográficos de poblaciones de macroorganismos que coexisten en una región dada, estudiados en el contexto de la historia de dicha región, han ayudado a entender cómo la evolución de los linajes que ahí habitan han sido afectados por dicha historia (Carnaval y Moritz, 2008; Ramírez-Barahona y Eguiarte, 2013). Sin embargo, a pesar de la importancia que los microorganismos tienen en las funciones de los ecosistemas, (Allison y Martiny, 2008; Strickland et al., 2009) la magnitud de impacto que los factores históricos tienen sobre la dinámica poblacional de bacterias y arqueas no se ha explorado. La mayoría de los estudios sobre la dinámica poblacional de bacterias se han centrado en el análisis de patrones demográficos de bacterias patógenas (Pybus et al., 2001; Wirth et al., 2007; Tazi et al., 2010) y los estudios demográficos de bacterias de vida libre han sido muy escasos (Guttman et al., 2008). En este sentido, el trabajo presentado en esta tesis es uno de los primeros en explorar la importancia que los factores históricos pueden tener sobre la dinámica poblacional de grupos de bacterias de vida libre. Los resultados obtenidos aportan evidencia sobre el potencial que los estudios comparativos de genética de poblaciones pueden tener para ayudarnos a entender algunos de los asuntos que siguen siendo centrales para los campos de la microbiología evolutiva y ecología molecular de bacterias.

Finalmente, cabe mencionar aquí que si bien se obtuvieron resultados importantes sobre la dinámica poblacional de los grupos de *Bacillus* estudiados, una limitante importante de este trabajo fue el contar con tan pocos marcadores moleculares. Para poder profundizar la caracterización de los linajes de *Bacillus* aquí presentados sería muy importante contar con una mayor cantidad de información tanto genotípica como fenotípica. Con este objetivo se ha iniciado la caracterización por medio de BOX-PCR de una parte de la colección de aislados. Este tipo de genotipificación permitirá describir la plasticidad genómica de los aislados y hacer una comparación con la filogenia previamente obtenida usando los genes constitutivos (Avitia et al., 2012). La intención de usar este tipo de marcadores es que en trabajos previos con grupos ambientales de bacterias los patrones de diversidad a nivel genómico han permitido discernir patrones que relacionan la diversidad genética a este nivel con condiciones ambientales particulares y/o barreras geográficas (Cho y Tiedje, 2000; Sikorski y Nevo; 2005). En el trabajo presentado en esta tesis no se pudo reconocer la influencia de las condiciones ambientales particulares y la diversidad genética de los grupos de *Bacillus* analizados, sin embargo una de las razones pudo haber sido el bajo nivel de resolución obtenido con los marcadores moleculares elegidos. Además de los análisis genotípicos se inició también la caracterización fenotípica de algunos de los aislados. Para lograrlo se ha empezado a trabajar con curvas de crecimiento de los aislados en condiciones diferentes de temperatura, pH y salinidad. Aunque los resultados preliminares no han permitido reconocer un patrón que asocie la

pertenencia a uno de los filogrupos (B1 y B2) con la respuesta a las diferentes condiciones probadas permitirá hacer generalizaciones al respecto. Por último, la secuenciación y análisis de genomas completos de algunos de estos aislados sería muy adecuada para seguir explorando las causas detrás del patrón de divergencia observado entre filogrupos.

## REFERENCIAS

- Achtman M, Wagner M. 2008. Microbial diversity and the genetic nature of microbial species. *Nature reviews. Microbiology* 6:431–40.
- Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430:551–554.
- Alcaraz LD, Olmedo G, Bonilla-Rosso G, Cerritos R, Hernández G, Cruz A, Ramíerz E, Putonti C, Jiménez B, Martínez E, López V, Arvizu J, Ayala F, Razo F, Caballero J, Siefert J, Eguiarte L, Vielle J, Martínez O, Souza V, Herrera-Estrella A, Herrera-Estrella L. 2008. The genome of *Bacillus coahuilensis* reveals adaptations essential for survival in the relic of an ancient marine environment. *Proceedings of the National Academy of Sciences* 105:5803–5808.
- Allen EE, Banfield JF. 2005. Community genomics in microbial ecology and evolution. *Nature reviews. Microbiology* 3:489–98.
- Amann RI, Ludwig W, Schleifer K. 1995. Phylogenetic Identification and In Situ Detection of Individual Microbial Cells without Cultivation. 59:143–169.
- Auger S, Galleron N, Bidnenko E, Ehrlich SD, Lapidus A, Sorokin A. 2008. The genetically remote pathogenic strain NVH391-98 of the *Bacillus cereus* group is representative of a cluster of thermophilic strains. *Applied and Environmental Microbiology* 74:1276–1280.
- Avitia M, Escalante AE, Rebollar E a, Moreno-Ietelier A, Eguiarte LE, Souza V. 2014. Population expansions shared among coexisting bacterial lineages are revealed by genetic evidence. *Peer J*:1–17.
- Bielawski JP, Dunn K a, Sabehi G, Béjà O. 2004. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proceedings of the National Academy of Sciences of the United States of America* 101:14824–14829.
- Blainey PC. 2013. The future is now: single-cell genomics of bacteria and archaea. *FEMS microbiology reviews* 37:407–27.
- Bligh E, Dyer W. 1959. A rapid method for total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*:911–917.

- Bonilla-Rosso G, Peimbert M, Alcaraz LD, Hernández I, Eguiarte LE, 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* 12:659–673.
- Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, Dinsdale E, Edwards R, Souza V, Rohwer F, Hollander D. 2009. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro Ciénegas, Mexico. *Environmental Microbiology* 11:16–34.
- Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ. 2012. Patterns of gene flow define species of thermophilic Archaea. *PLoS Biology* 10.
- Cardazzo B, Negrisoló E, Carraro L, Alberghini L, Patarnello T, Giaccone V. 2008. Multiple-locus sequence typing and analysis of toxin genes in *Bacillus cereus* food-borne isolates. *Applied and Environmental Microbiology* 74:850–860.
- Caugant DA, Mocca LF, Frasch CE, Froholm LO, Zollinger WD, Selander RK. 1987. Genetic Structure of *Neisseria meningitidis* Populations in Relation to Serogroup, Serotype, and Outer Membrane Protein Pattern. *Journal of Bacteriology* 169:2781–2792.
- Cerritos R. 2007. La especie como unidad evolutiva: uso de marcadores moleculares para su reconocimiento y delimitación, con especial énfasis en microorganismos. In: Eguiarte LE, Souza V, Aguirre X eds. *Ecología Molecular*. Secretaría de Medio Ambiente y Recursos Naturales, Instituto Nacional de Ecología, Universidad Nacional Autónoma de México Y Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, 325–350.
- Cerritos R. 2008. Análisis del concepto biológico, cohesivo, ecológico y filogenético de especie en bacterias halófilas del valle de Cuatro Ciénegas, Coahuila. UNAM.
- Cerritos R, Vinuesa P, Eguiarte LE, Herrera-Estrella L, Alcaraz-Peraza LD, Arvizu-Gómez JL, Olmedo G, Ramirez E, Siefert JL, Souza V. 2008. *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Ciénegas Valley in Coahuila, Mexico. *International journal of systematic and evolutionary microbiology* 58:919–23.
- Cerritos R, Eguiarte LE, Avitia M, Siefert J, Travisano M, Rodríguez-Verdugo A, Souza V. 2011. Diversity of culturable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, México. *Antonie van*

- Leeuwenhoek, International Journal of General and Molecular Microbiology* 99:303–318.
- Cho JC, Tiedje JM. 2000. Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Applied and Environmental Microbiology* 66:5448–5456.
- Claus DR, Berkeley CW. 1986. The genus *Bacillus*. In: *Bergey's manual of systematic bacteriology*. 1105–1139.
- Cohan FM. 2002. What are bacterial species? *Annual reviews of microbiology* 56:457–487.
- Cohan FM, Perry EB. 2007. A systematics for discovering the fundamental units of bacterial diversity. *Current biology: CB* 17:R373–86.
- Cohn F. 1962. Studies on the biology of the bacilli. In: Brock T ed. *Milestones in Microbiology*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 49–56.
- Coyne JA, Orr HA. 2004. *Speciation*. Sinauer Associates, Inc.
- Curtis TP, Sloan WT, Scannell JW. 2002. Estimating prokaryotic diversity and its limits.
- DeLong EF, Pace NR. 2001. Environmental diversity of bacteria and archaea. *Systematic biology* 50:470–478.
- Desnues C, Rodriguez-brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H, Furlan M, Wegley L, Chau B, Ruan Y, Hall D, Angly FE, Edwards RA, Li L, Thurber RV, Reid RP, Siefert J, Souza V, Valentine DL, Swan BK, Breitbart M, Rohwer F. 2008. Phages apart. *Group* 6:6735–6735.
- Didelot X, Barker M, Falush D, Priest FG. 2009. Evolution of pathogenicity in the *Bacillus cereus* group. *Syst Appl Microbiol* 32:81–90.
- Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175:1251–66.
- Dinsdale EA, Rohwer F. 2008. Dissecting microbial employment. *Nature biotechnology* 26:997–998.

- Dobzhansky T. 1937. *Genetics and the origin of species*. Columbia University Press.
- Doolittle WF, Papke RT. 2006. Genomics and the bacterial species problem. *Genome biology* 7:116.
- Duncan KE, Ferguson N, Kimura K, Zhou X, Istock CA. 1994. Fine-Scale Genetic and Phenotypic Structure in Natural Populations of *Bacillus subtilis* and *Bacillus licheniformis* : Implications for Bacterial Evolution and Speciation. *Evolution* 48:2002–2025.
- Dvořák P, Hašler P, Poulíčková A. 2012. Phylogeography of the *Microcoleus vaginatus* (Cyanobacteria) from three continents--a spatial and temporal characterization. *PloS one* 7:e40153.
- Dykhuizen DE. 1998. Santa Rosalia revisited : Why are there so many species of bacteria ? :25–33.
- Elser JJ, Schampel JH, Garcia-Pichel F, Wade BD, Souza V, Eguiarte L, Escalante A, Farmer JD. 2005. Effects of phosphorus enrichment and grazing snails on modern stromatolitic microbial communities. *Freshwater Biology* 50:1808–1825.
- Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. 2000. Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus* Multilocus Sequence Typing for Characterization of Methicillin- Resistant and Methicillin-Susceptible Clones of *Staphyloco*. *Journal of clinical microbiology* 38:1008–1015.
- Escalante AE, Eguiarte LE, Espinosa-asuar L, Forney LJ, Noguez AM, Saldivar VS. 2008. Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas basin. 65:50–60.
- Escalante AE. 2008. Ecología evolutiva de procariontes en Cuatro Ciénegas, Coahuila. UNAM.
- Escalante AE, Caballero-Mellado J, Martínez-Aguilar L, Rodríguez-Verdugo A, González-González A, Toribio-Jiménez J, Souza V. 2009. *Pseudomonas cuatrociénegasensis* sp. nov., isolated from an evaporating lagoon in the Cuatro Ciénegas valley in Coahuila, Mexico. *International journal of systematic and evolutionary microbiology* 59:1416–20.

- Ezaki T, Hashimoto Y, Yabuuchi E. 1989. Fluorometric Deoxyribonucleic Acid-Deoxyribonucleic Acid Hybridization in Microdilution Wells as an Alternative to Membrane Filter Hybridization in which Radioisotopes Are Used To Determine Genetic Relatedness among Bacterial Strains. :224–229.
- Fagerlund A, Brillard J, Fürst R, Guinebretière M-H, Granum PE. 2007. Toxin production in a rare and genetically remote cluster of strains of the *Bacillus cereus* group. *BMC microbiology* 7:43.
- Falush D. 2009. Toward the use of genomics to study microevolutionary change in bacteria. *PLoS genetics* 5:e1000627.
- Feil EJ, Cooper JE, Grundmann H, Robinson DA, Enright MC, Berendt T, Peacock SJ, Smith JM, Murphy M, Spratt BG, Moore CE, Day NPJ. 2003. How Clonal Is *Staphylococcus aureus* ? 185:3307–3316.
- Feil EJ, Spratt BG. 2001. Recombination and the population structures of bacterial pathogens. *Annual Reviews of Microbiology* 55:561–590.
- Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species. *Science (New York, N.Y.)* 296:1061–1063.
- Forney LJ, Zhou X, Brown CJ. 2004. Molecular microbial ecology: land of the one-eyed king. *Current opinion in microbiology* 7:210–20.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. 2009. The bacterial species challenge: making sense of genetic and ecological diversity. *Science (New York, N.Y.)* 323:741–6.
- Fraser C, Hanage WP, Spratt BG. 2007. Recombination and the nature of bacterial speciation. *Science (New York, N.Y.)* 315:476–80.
- Fu Y-X. 1997. Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* 147:915–925.
- Fu Y-X, Li W-H. 1993. Statistical Tests of Neutrality of Mutations. *Genetics* 133:693–709.
- Fujita T, Shida O, Takagi H, Kunugita K, Pankrushina a. N, Matsushashi M. 1996. Description of *Bacillus carboniphilus* sp. nov. *International Journal of Systematic Bacteriology* 46:116–118.



- Geiger O, Rohrs V, Weissenmayer B, Finan T, Thomas-Oates J. 1999. The regulator gene *phoB* mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl- N , N , N -trimethylhomoserine in *Rhizobium* ( *Sinorhizobium* ) *meliloti*. *Molecular microbiology*:63–67.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Peer Y Van De, Vandamme P, Thompson FL, Swings J. 2005. Defining prokaryotic species Re-evaluating prokaryotic species. *3*:733–739.
- Ghai R, Pašić L, Fernández AB, Martin-Cuadrado A-B, Mizuno CM, McMahon KD, Papke RT, Stepanauskas R, Rodriguez-Brito B, Rohwer F, Sánchez-Porro C, Ventosa A, Rodríguez-Valera F. 2011. New Abundant Microbial Groups in Aquatic Hypersaline Environments. *Scientific Reports* 1:sreport00135.
- Ghelardi E, Celandroni F, Salvetti S, Fiscarelli E, Senesi S. 2007. *Bacillus thuringiensis* pulmonary infection: critical role for bacterial membrane-damaging toxins and host neutrophils. *Microbes and Infection* 9:591–598.
- Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic Evolution in Light of Gene Transfer. :2226–2238.
- Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. *Nature reviews. Microbiology* 3:679–687.
- Goris J, Suzuki K, De Vos P, Nakase T, Kersters K. 1998. Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. *Can J Microbiol*:1148–1153.
- Guinebretière MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M, De Vos P. 2008. Ecological diversification in the *Bacillus cereus* Group. *Environmental Microbiology* 10:851–865.
- Hanage WP. 2013. Fuzzy species revisited. *BMC biology* 11:41.
- Hanage WP, Fraser C, Spratt BG. 2005. Fuzzy species among recombinogenic bacteria. *BMC biology* 3:6.
- Hanage WP, Fraser C, Spratt BG. 2006. The impact of homologous recombination on the generation of diversity in bacteria. *239*:210–219.
- Hart MW. 2011. The species concept as an emergent property of population biology. *Evolution; international journal of organic evolution* 65:613–6.

- Hartl DL, Clark AG. 1997. *Principles of Population Geneticd*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Haubold B, Travisano M, Rainey PB, Hudson RR. 1998. Detecting linkage disequilibrium in bacterial populations [In Process Citation]. *Genetics* 150:1341–1348.
- Haubold B, Hudson RR. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Linkage Analysis. Bioinformatics (Oxford, England)* 16:847–848.
- Hedrick PW. 2011. *Genetics of Populations*. Sudbury, Massachusetts: Jones & Bartlett Publishers.
- Helgason E, Caugant D a., Lecadet MM, Chen Y, Mahillon J, Lövgren A, Hegna I, Kvaløy K, Kolstø AB. 1998. Genetic diversity of *Bacillus cereus*/*B. thuringiensis* isolates from natural sources. *Current Microbiology* 37:80–87.
- Helgason E, Økstad OLEA, Caugant DA, Mock LE, Hegna IDA, Johansen HA, Fouet A. 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* - One Species on the Basis of Genetic Evidence. *Applied and environmental microbiology* 66:2627–2630.
- Helgason E, Tourasse NJ, Meisal R, Caugant DA, Kolstø A. 2004. Multilocus Sequence Typing Scheme for Bacteria of the *Bacillus cereus* Group. 70:191–201.
- Hoffmaster AR, Novak RT, Marston CK, Gee JE, Hessel L, Pruckler JM, Wilkins PP. 2008. Genetic diversity of clinical isolates of *Bacillus cereus* using multilocus sequence typing. *BMC microbiology* 8:191.
- Holt J. 1986. Gram-Positive Bacteria other than Actinomycetes. In: Sneath P, Mair N, Sharpe M, Holt J eds. *Bergey's manual of systematic bacteriology*. Baltimore, MD: Williams & Williams,.
- Hunt DE, David L a, Gevers D, Preheim SP, Alm EJ, Polz MF. 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science (New York, N.Y.)* 320:1081–5.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution* 23:254–67.

- Istock CA, Duncan KE, Ferguson N, Zhou X. 1992. Sexuality in a natural population of bacteria--*Bacillus subtilis* challenges the clonal paradigm. *Molecular ecology* 1:95–103.
- Istock CA, Ferguson N, Istock NL, Duncan KE. 2001. Geographical diversity of genomic lineages in *Bacillus subtilis* ( Ehrenberg ) Cohn sensu lato. 1:179–191.
- Jain R, Rivera MC, Moore JE, Lake J a. 2003. Horizontal gene transfer accelerates genome innovation and evolution. *Molecular Biology and Evolution* 20:1598–1602.
- Johannesson KH, Cortés A, Kilroy KC. 2004. Reconnaissance isotopic and hydrochemical study of Cuatro Ciénegas groundwater, Coahuila, México. *Journal of South American Earth Sciences* 17:171–180.
- Kämpfer P, Rosselló-Mora R, Falsen E, Busse H-J, Tindall BJ. 2006. *Cohnella thermotolerans* gen. nov., sp. nov., and classification of “*Paenibacillus hongkongensis*” as *Cohnella hongkongensis* sp. nov. *International journal of systematic and evolutionary microbiology* 56:781–6.
- Kates M. 1986. *Techniques of lipidology. Isolation, analysis and identification of lipids*. Amsterdam, New York, Oxford: Elsevier.
- Kinzig AP, Pacala S, Tilman GD. 2002. The Functional Consequences of Biodiversity. In: *The Functional Consequences of Biodiversity*. 265–293.
- Ko KS, Kim J-W, Kim J-M, Kim W, Chung S, Kim IJ, Kook Y-H. 2004. Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* Gene. *Infection and immunity* 72:5253–61.
- Koch R. 1962. The etiology of anthrax, based on the life history of *Bacillus anthracis*. In: Brock T ed. *Milestones in Microbiology*. Englewood Cliffs, NJ: Prentic-Hall, Inc., 89–95.
- Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, Rooney AP, Brambilla E, Connor N, Ratcliff RM, Nevo E, Cohan FM. 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* 105:2504–9.

- Kolstø A-B, Tourasse NJ, Økstad OA. 2009. What sets *Bacillus anthracis* apart from other *Bacillus* species? *Annual review of microbiology* 63:451–476.
- Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361:1929–40.
- Kopac S, Cohan FM. 2011. A theory-based pragmatism for discovering and classifying newly divergent bacterial species. In: Tibayrenc M ed. *Genetics and Evolution of Infectious Diseases*. Elsevier,.
- Lane DJ. 1991. 16S/23S rDNA sequencing. Nucleic acid techniques. In: Stackebrandt E, Goodfellow M eds. *Bacterial Systematics*. New York, USA: John Wiley & Sons, 115–175.
- Lapidus A, Goltsman E, Auger S, Galleron N, Ségurens B, Dossat C, Land ML, Broussolle V, Brillard J, Guinebretiere MH, Sanchis V, Nguen-the C, Lereclus D, Richardson P, Wincker P, Weissenbach J, Ehrlich SD, Sorokin A. 2008. Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. *Chemico-Biological Interactions* 171:236–249.
- Lasken RS. 2012. Genomic sequencing of uncultured microorganisms from single cells. 10.
- Lawrence JG. 2002. Gene transfer in bacteria: speciation without species? *Theoretical population biology* 61:449–460.
- Lee J, Lim J, Park D, Jeon CO, Li W, Kim C. 2006. *Bacillus seohaeanensis* sp. nov., a halotolerant bacterium that contains L-lysine in its cell wall. :1893–1898.
- Leifson E. 1963. Determination of Carbohydrate Metabolism of. *Growth (Lakeland)* 85:5–7.
- Lessa EP, Cook JA, Patton JL. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences of the United States of America* 100:10331–4.
- Lim J-M, Jeon CO, Lee S-M, Lee J-C, Xu L-H, Jiang C-L, Kim C-J. 2006. *Bacillus salarius* sp. nov., a halophilic, spore-forming bacterium isolated from a salt lake in China. *International journal of systematic and evolutionary microbiology* 56:373–7.

- Logan N a, Berge O, Bishop a H, Busse H-J, De Vos P, Fritze D, Heyndrickx M, Kämpfer P, Rabinovitch L, Salkinoja-Salonen MS, Seldin L, Ventosa a. 2009. Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *International journal of systematic and evolutionary microbiology* 59:2114–21.
- Lomolino M V, Riddle BR, Brown JH. 2005. *Biogeography*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Lozupone C, Hamady M, Knight R. 2006. UniFrac – An online tool for comparing microbial community diversity in a phylogenetic context. 14:1–14.
- Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences* 95:3140–3145.
- Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee J-H, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O’Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D. 2006. Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 103:15611–6.
- Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman J a, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach A-L, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. *Nature reviews. Microbiology* 4:102–12.
- Maughan H, Van der Auwera G. 2011. Bacillus taxonomy in the genomic era finds phenotypes to be essential though often misleading. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 11:789–97.
- Mayden R. 1997. A hierarchy of species concept: the denouement in the saga of the species problem. In: *Species:the units of biodiversity*. London: Chapman and Hall, 381–424.

- Maynard Smith J, Smith NH, Rourke MO, Spratt BG. 1993. How clonal are bacteria ? *EhJ-X*. 90:4384–4388.
- Maynard Smith J, Smith NH. 1993. Detecting Recombination from Gene Trees. :590–599.
- Mayr E. 1942. *Systematics and the origin of species*. Columbia University Press.
- McIntyre L, Bernard K, Beniac D, Isaac-Renton JL, Naseby DC. 2008. Identification of *Bacillus cereus* group species associated with food poisoning outbreaks in British Columbia, Canada. *Applied and Environmental Microbiology* 74:7451–7453.
- Medini D, Serruto D, Parkhill J, Relman D a, Donati C, Moxon R, Falkow S, Rappuoli R. 2008. Microbiology in the post-genomic era. *Nature reviews. Microbiology* 6:419–30.
- Mesbah M, Premachandran U, Whitman WB. 1989. Precise Measurement of the G+C Content of Deoxyribonucleic Acid by High-Performance Liquid Chromatography. *International Journal of Systematic Bacteriology* 39:159–167.
- Mesbah NM, Wiegel J. 2012. Life under multiple extreme conditions: diversity and physiology of the halophilic alkalithermophiles. *Applied and environmental microbiology* 78:4074–82.
- Milkman R. 1973. Electrophoretic Variation in *Escherichia coli* from Natural Sources. *Science* 182:1024–1026.
- Narra HP, Ochman H. 2006. Of What Use Is Sex to Bacteria ? :705–710.
- Nelson K, Whittam TS, Selander RK. 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3- phosphate dehydrogenase gene ( *gapA* ) in natural populations of *Salmonella* and *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 88:6667–6671.
- Noguchi H, Uchino M, Shida O, Takano K, Nakamura LK, Komagata K. 2004. *Bacillus vietnamensis* sp . nov ., a moderately halotolerant , aerobic , endospore-forming bacterium isolated from Vietnamese fish sauce. :2117–2120.

- O'Shea TM, DeLoney-Marino CR, Shibata S, Aizawa SI, Wolfe AJ, Visick KL. 2005. Magnesium promotes flagellation of *Vibrio fischeri*. *Journal of Bacteriology* 187:2058–2065.
- Oakley BB, Carbonero F, van der Gast CJ, Hawkins RJ, Purdy KJ. 2010. Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. *The ISME journal* 4:488–497.
- Ochman H, Lawrence JG, Groisman E a. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.
- Oda Y, Star B, Huisman L a., Gottschal JC, Forney LJ. 2003. Biogeography of the purple nonsulfur bacterium *Rhodospseudomonas palustris*. *Applied and Environmental Microbiology* 69:5186–5191.
- Oren A. 2004. Prokaryote diversity and taxonomy: current status and future challenges. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 359:623–38.
- Papke RT, Ramsing NB, Bateson MM, Ward DM. 2003. Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology* 5:650–659.
- Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muike D, Doolittle WF. 2007. Searching for species in haloarchaea. *Proceedings of the National Academy of Sciences of the United States of America* 104:14092–7.
- Passalacqua K, Bergman N. 2006. *Bacillus anthracis*: interactions with the host and establishment of inhalational anthrax. *Future Microbiology* 1:397–415.
- Peimbert M, Alcaraz LD, Bonilla-Rosso G, Olmedo-Alvarez G, García-Oliva F, Segovia L, Eguiarte LE, 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* 12:648–658.
- Pignatelli M, Moya A, Tamames J. 2009. EnvDB, a database for describing the environmental distribution of prokaryotic taxa. *Environmental Microbiology Reports* 1:191–197.
- Pikuta E V, Hoover RB, Tang J. 2007. Microbial extremophiles at the limits of life. *Critical reviews in microbiology* 33:183–209.
- Pilo P, Frey J. 2011. *Bacillus anthracis*: molecular taxonomy, population genetics, phylogeny and patho-evolution. *Infection, genetics and evolution : journal of*

*molecular epidemiology and evolutionary genetics in infectious diseases*  
11:1218–24.

- Piñero D, Martinez E, Selander RK. 1988. Genetic Diversity and Relationships among Isolates of *Rhizobium leguminosarum* biovar phaseoli. *Applied and Environmental Microbiology* 54:2825–2832.
- Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. 2004. Population structure and evolution of the *Bacillus cereus* group. *Journal of bacteriology* 186:7959–70.
- Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM. 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296:2028–33.
- Rebollar EA, Avitia M, Eguiarte LE, 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 Ciénegas Basin. *Environmental Microbiology* 14:2323–2333.
- Retchless AC, Lawrence JG. 2007. Temporal fragmentation of speciation in bacteria. *Science* 317:1093–6.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Schlegel HG, Jannasch HW. 2013. Prokaryotes and Their Habitats. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F eds. *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 39–80.
- Schleifer KH, Kandler O. 1972. Peptidoglycan . Types of Bacterial Cell Walls and their Taxonomic Implications. 36:407–477.
- Schumann P. 2011. Peptidoglycan structure. *Methods in microbiology* 38:101–129.
- Seckbach J, Oren A, Stan-Lotter H. (eds.) 2013. *Polyextremophiles*. Dordrecht: Springer Netherlands.
- Selander RK, Caugant DA, Ochman Ho, Musser JM, Gilmour MN, Whittam TS. 1986. Methods of Multilocus Enzyme Electrophoresis for Bacterial Population



- Genetics and Systematics. *Applied and Environmental Microbiology* 51:873–884.
- Selander RK, Li K, Nelson K. 1996. Evolutionary genetics of *Salmonella enterica*. In: Neidhardt F. ed. *Escherichia coli and Salmonella*. Washington DC: ASM press,.
- Shapiro BJ, David L a., Friedman J, Alm EJ. 2009. Looking for Darwin's footprints in the microbial world. *Trends in Microbiology* 17:196–204.
- Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* 336:1093–6.
- Shapiro BJ, Polz MF. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in microbiology* 22:235–247.
- Shida O. 1997. *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, . :289–298.
- Sikorski J, Nevo E. 2005. Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at “Evolution Canyons” I and II, Israel. *Proceedings of the National Academy of Sciences of the United States of America* 102:15924–9.
- Sikorski J, Nevo E. 2007. Patterns of thermal adaptation of *Bacillus simplex* to the microclimatically contrasting slopes of “Evolution Canyons” I and II, Israel. *Environmental microbiology* 9:716–26.
- Silva C, Vinuesa P, Eguiarte LE, Martí E, Souza V. 2003. *Rhizobium etli* and *Rhizobium gallicum* Nodulate Common Bean ( *Phaseolus vulgaris* ) in a Traditionally Managed Milpa Plot in Mexico : Population Genetics and Biogeographic Implications. 69:884–893.
- Silva C, Vinuesa P. 2007. Ecología evolutiva de bacterias y el concepto de especie: el caso de los Rizobios. In: Eguiarte L, Souza V, Aguirre X eds. *Ecología Molecular*. secretaria de Medio Ambiente y Recursos Naturales, Instituto Nacional de Ecología, Universidad Nacional Autónoma de México Y Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, 351–392.
- Sims R, Larose J. 1962. The use of iodine vapor as a general detecting agent in the thin layer chromatography of lipids. *Journal of the American Oil Chemists Society* 39:232.

- Singh NK, Kaur C, Kumar N, Velmurugan S, Citarasu T, Mayilraj S. 2014. *Bacillus aequororis* sp. nov., Isolated From Marine Sediment. *Current microbiology*.
- Slepecky RA, Hemphill HE. 2006. *The Genus Bacillus — Nonmedical*.
- Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology* 8:732–740.
- Smith JM. 1991. The Population Genetics of Bacteria. *Proceedings of the Royal Society B: Biological Sciences* 245:37–41.
- Smith JM, Feil EJ, Smith NH. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays: news and reviews in molecular, cellular and developmental biology* 22:1115–1122.
- Soberón M, Pardo-López L, López I, Gómez I, Tabashnik BE, Bravo A. 2007. Engineering modified Bt toxins to counter insect resistance. *Science (New York, N.Y.)* 318:1640–1642.
- Sorokin A, Candelon B, Guilloux K, Galleron N, Wackerow-Kouzova N, Ehrlich SD, Bourguet D, Sanchis V. 2006. Multiple-locus sequence typing analysis of *Bacillus cereus* and *Bacillus thuringiensis* reveals separate clustering and a distinct population structure of psychrotrophic strains. *Applied and environmental microbiology* 72:1569–78.
- Souza V, Nguyen TT, Hudson RR, Piñero D, Lenski RE. 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* 89:8389–8393.
- Souza V, Espinosa-asuar L, Escalante AE, Eguiarte LE, Farmer J, Forney L, Sobero X, Dirzo R, Elser JJ, Lloret L, Rodri JM. 2006. An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. 103:6565–6570.
- Souza V, Eguiarte LE, Siefert J, Elser JJ. 2008. Microbial endemism: does phosphorus limitation enhance speciation? *Nature reviews. Microbiology* 6:559–64.
- Souza V, Siefert JL, Escalante AE, Elser JJ, Eguiarte LE. 2012. The Cuatro Ciénegas Basin in Coahuila, Mexico: an astrobiological Precambrian Park. *Astrobiology* 12:641–7.

- Spratt BG, Maiden MC. 1999. Bacterial population genetics, evolution and epidemiology. *Philosophical Transactions of the Royal Society B: Biological Sciences* 354:701–710.
- Stackebrandt E, Stackebrandt E, Frederiksen W, Frederiksen W, Garrity GM, Garrity GM, Grimont P a D, Grimont P a D, Ka P, Ka P, Maiden MCJ, Maiden MCJ, Nesme X, Nesme X, Swings J, Swings J, Tru HG, Tru HG, Rossello R, Rossello R, Vauterin L, Vauterin L, Ward AC, Ward AC, Whitman WB, Whitman WB. 2002. Taxonomic note 1. *International Journal of Systematic and Evolutionary Microbiology*:1043–1047.
- Stackebrandt E, Goebel BM. 1994. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology* 44:846–849.
- Staley JT, Gosink J. 1999. Poles apart: biodiversity and biogeography of polar sea ice bacteria. *Annual Reviews of Microbiology* 53:189–215.
- Stepanauskas R. 2012. Single cell genomics : an individual look at microbes. *Current Opinion in Microbiology* 15:613–620.
- Stevenson A, Burkhardt J, Cockell CS, Cray J a, Dijksterhuis J, Fox-Powell M, Kee TP, Kminek G, McGenity TJ, Timmis KN, Timson DJ, Voytek M a, Westall F, Yakimov MM, Hallsworth JE. 2014. Multiplication of microbes below 0.690 water activity: implications for terrestrial and extraterrestrial life. *Environmental microbiology*.
- Suerbaum S, Smith JM, Bapumia K, Morelli G, Smith NH, Kunstmann E, Dyrek I, Achtman M. 1998. Free recombination within *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America* 95:12619–12624.
- Sul WJ, Oliver T a, Ducklow HW, Amaral-Zettler L a, Sogin ML. 2013. Marine bacteria exhibit a bipolar distribution. *Proceedings of the National Academy of Sciences of the United States of America* 110:2342–7.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland EDP, Gomez ML, Sieracki ME, DeLong EF, Herndl GJ, Stepanauskas R. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science (New York, N.Y.)* 333:1296–300.

- Switzer Blum J, Burns Bindi A, Buzzelli J, Stolz JF, Oremland RS. 1998. *Bacillus arsenicoselenatis*, sp. nov and *Bacillus selenitireducens*, sp. nov.: two haloalkaliphiles from Mono Lake , California that respire oxyanions of selenium and arsenic. *Archives of microbiology* 171:19–30.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* 30:2725–9.
- Täubel M, Kämpfer P, Buczolits S, Lubitz W, Busse HJ. 2003. *Bacillus barbaricus* sp. nov., isolated from an experimental wall painting. *International Journal of Systematic and Evolutionary Microbiology* 53:725–730.
- Teska J, Coyne S, Ezzell J. 1998. Identification of *Bacillus anthracis* Using Gas Chromatographic Analysis of Cellular Fatty Acids and a Commercially Available Database Application. *Agilent Technologies*:1–6.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22:4673–4680.
- Tindall B, Sikorski J, SMibert R, Kreig N. 2007. Phenotypic characterization and the principles of comparative systematics. In: Reddy C, Beveridge T, Breznak J, Marzluf G, Schmidt T, Snyder L eds. *Methods for general and molecular microbiology*. Whashington, DC: ASM press, 330–393.
- Tourasse NJ, Helgason E, Økstad O a, Hegna IK, Kolstø a-B. 2006. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *Journal of applied microbiology* 101:579–93.
- Tourasse NJ, Helgason E, Klevan A, Sylvestre P, Moya M, Haustant M, Økstad OA, Fouet A, Mock M, Kolstø AB. 2011. Extended and global phylogenetic view of the *Bacillus cereus* group population by combination of MLST, AFLP, and MLEE genotyping data. *Food Microbiology* 28:236–244.
- Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological reviews* 60:407–438.

- Ventosa A, Nieto JJ, Oren A. 1998. Biology of Moderately Halophilic Aerobic Bacteria *Biology of Moderately Halophilic Aerobic Bacteria*. 62.
- Vilas-Boas G, Sanchis V, Lereclus D, Lemos MVF, Bourguet D. 2002. Genetic Differentiation between Sympatric Populations of *Bacillus cereus* and *Bacillus thuringiensis*. *Applied and Environmental Microbiology* 68:1414–1424.
- Vinuesa P, Silva C, Werner D, Martı E. 2005. Population genetics and phylogenetic inference in bacterial molecular systematics : the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. 34:29–54.
- Vogel J, Normand P, Thioulouse J, Nesme X, Grundmann GL. 2003. Relationship between Spatial and Genetic Distance in. *Applied and environmental microbiology* 69:1482–1487.
- Vos M. 2010. A species concept for bacteria based on adaptive divergence. *Trends in Microbiology*:1–7.
- Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *The ISME journal* 3:199–208.
- Vos M, Velicer GJ. 2008. Report Isolation by Distance in the Spore-Forming Soil Bacterium *Myxococcus xanthus*. :386–391.
- Ward DM. 1998. A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1:271–277.
- Whitaker RJ. 2006. Allopatric origins of microbial species. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361:1975–84.
- Whitaker RJ, Banfield JF. 2006. Population genomics in natural microbial communities. *Trends in ecology & evolution* 21:508–16.
- Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science (New York, N.Y.)* 301:976–978.
- Whitfield J. 2005. Biogeography. Is everything everywhere? *Science* 310:960–961.
- Whitman WB, Coleman DC, Wiebe WJ. 1998. Perspective Prokaryotes : The unseen majority. 95:6578–6583.

- Wilson DJ. 2012. Insights from genomics into bacterial pathogen populations. *PLoS pathogens* 8:e1002874.
- Wilson JS, Pitts JP. 2010. Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: A resource for biologists. *Progress in Physical Geography* 34:419–441.
- Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms : Proposal for the domains. 87:4576–4579.
- Wolaver BD, Sharp JM, Rodriguez JM, Ibarra Flores JC. 2008. Delineation of regional arid karstic aquifers: An integrative data approach. *Ground Water* 46:396–413.
- Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, Saw JH, Senin P, Yang C, Chatterji S, Cheng J-F, Eisen JA, Sieracki ME, Stepanauskas R. 2009. Assembling the marine metagenome, one cell at a time. *PloS one* 4:e5299.
- Yoon J-H. 2003. *Bacillus marisflavi* sp. nov. and *Bacillus aquimaris* sp. nov., isolated from sea water of a tidal flat of the Yellow Sea in Korea. *International Journal of Systematic and Evolutionary Microbiology* 53:1297–1303.
- Zbierzak A, Dormann P, Holz G. 2011. Analysis of lipid content and quality in arabidopsis plastids. In: *Methods in molecular biology*. 411–426.
- Zinder SH, Dworkin M. 2006. Morphological and Physiological Diversity. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E eds. *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 185–220.