



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

DOCTORADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE ECOLOGIA

Ecología Evolutiva de Procariontes en Cuatro Ciénegas

Coahuila, México

**T E S I S**

QUE PARA OBTENER EL GRADO ACADÉMICO DE

**DOCTORA EN CIENCIAS**

PRESENTA

**ANA ELENA ESCALANTE HERNÁNDEZ**



DIRECTORA DE TESIS: DRA. VALERIA SOUZA SALDIVAR

INSTITUTO  
DE ECOLOGIA  
UNAM

MÉXICO DF

2008

---

## AGRADECIMIENTOS

*A la UNAM, a la Facultad de Ciencias y al Instituto de Ecología, gracias a estas instituciones me he formado en la investigación biológica.*

*Al sistema de becas de la UNAM y de México que me han permitido seguir estudiando.*

*Al Dr. Daniel Piñero, mi primer maestro en el camino científico y por quién adquirí el interés en la evolución.*

*A la Dra. Valeria Souza por todas las ideas, por su energía incontenible y su paciencia inagotable.*

*Al Dr. Luis E. Eguiarte, por su ejemplo de disciplina, dedicación y rigor en el trabajo.*

*Al Dr. Miguel Angel Cevallos por estar siempre allí para ayudarme y enseñarme algo de biología molecular.*

*Al Dr. Pablo Vinuesa por la empatía y ayuda en todo este proceso.*

*Al Dr. Jesús Caballero por enseñarme la importancia de dedicarle tiempo a los pequeños y grandes asuntos de la investigación, por mostrarme otros aspectos de la ecología microbiana.*

*A la Dra. Gloria Soberón por sus comentarios enriquecedores y por darme una perspectiva aplicada de la microbiología.*

*Special thanks to my “back-up advisor”, Dr. Larry Forney, for his generous help, advice and friendship.*

*A Ana Noguez, por vivir de cerca lo bueno y sobre todo lo malo, gracias por mantenerme en una pieza.*

*A Laura Espinosa, por su ayuda como técnica y como amiga.*

*A René, por seguir este camino en paralelo.*

*A todos los miembros del laboratorio de Evolución Molecular y Experimental por compartir estos años: Andreina, Ale, Erika, Jaime, Eria, Morena, Germán, Esmeralda, Enrique, Sandra y Doña Silvia.*

*A Rodrigo...nadie lo vio de tan cerca, imposible sobrevivirlo sin tí.*

*A Paola, por los consejos, por escuchar, por la ayuda, por las discusiones, por estar aquí desde hace ya tantos años... esto es para siempre.*

*A mis padres Víctor y Pilar, por su ejemplo de trabajo y perseverancia, por su apoyo infinito en todos los aspectos de mi vida.*

*A mi hermana Mariana, por no rendirse y continuar trabajando...llegará el día de la cosecha.*

*A mi familia adoptiva: Gustavo, Grace, Ceci, Gerardo, Doña Elisa... por los ánimos.*

*A Cuatro Ciénegas, a sus paisajes impresionantes, a las preguntas interminables.*

---

# ÍNDICE

<i>Contenido</i>	<i>Pág.</i>
<b>Agradecimientos</b>	1
<b>Resumen</b>	6
<b>Abstract</b>	7
<b>Prólogo. Objetivo General y Presentación</b>	8
<b>Capítulo 1. Introducción General</b>	11
1.1. Importancia del estudio de la diversidad biológica	12
1.2. Estudio de la diversidad biológica a diferentes escalas	16
1.3. Ecología microbiana, breve reseña histórica	17
1.4. Diversidad de microorganismos	18
1.4.1. Interacciones tróficas	18
1.4.2. Perspectiva evolutiva	18
1.4.3. Heterogeneidad espacial	19
1.4.4. Heterogeneidad temporal	19
1.5. Estudios recientes y controversias actuales sobre la diversidad y la distribución de microorganismos.	19
<b>Capítulo 2. Área de estudio</b>	23
2.1. El área natural protegida	23
2.2. Condiciones generales del área de estudio	24
2.2.1. Ubicación geográfica	24
2.2.2. Clima	24
2.2.3. Hidrología	25
2.3. Sistema Churince	26
<b>Capítulo 3. Diversidad de comunidades acuáticas procariontes del valle de Cuatro ciénegas</b>	30
3.1. Manuscrito: “ <i>Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas basin</i> ”	31
<b>Capítulo 4. Diversidad de linajes ambientales de <i>Pseudomonas sp.</i></b>	48
4.1. Introducción general sobre genética de poblaciones	48
4.1.1. ¿Qué estudia la genética de poblaciones?	49
4.1.2. ¿Por qué incluir un estudio de genética de poblaciones en un contexto ecológico?	50
4.2. Estructura genética de poblaciones de <i>Pseudomonas sensu stricto</i> en un sistema acuático de Cuatro Ciénegas, Coahuila, México, implicaciones ecológicas y evolutivas	53
4.2.1. Antecedentes	53
4.2.2. Muestreo	56
4.2.2.1. Área de estudio y toma de muestras	56

## INDICE

<i>Contenido</i>	<i>Pág.</i>
<b>Capítulo 4.</b> Diversidad de linajes ambientales de <i>Pseudomonas sp</i>	
4.2.2.2. Microorganismos y condiciones de aislamiento	58
4.2.2.3. Muestras y análisis	59
4.2.3. Resultados	77
I. Distribución ecológica de genotipos	77
II. Análisis filogenético y genética de poblaciones	79
4.2.4. Discusión	105
I. Distribución ecológica de genotipos	105
II. Análisis filogenético y genética de poblaciones	107
4.2.5. Conclusiones	115
<b>Capítulo 5.</b> Conclusiones generales y perspectivas	117
5.1. Conclusiones	119
5.2. Perspectivas	120
<b>Referencias</b>	121
<b>Apéndice I.</b> Descripción de <i>Pseudomonas caenicarum sp. nov.</i> una nueva especie aislada de una laguna de evaporación en el valle de Cuatro Ciénegas en Coahuila, México.	135
<b>Apéndice II.</b> Ecología molecular en el estudio de comunidades bacterianas: métodos moleculares y de análisis.	154

## ÍNDICE DE TABLAS

	<i>Pág.</i>
<b>Tabla 1.1.</b> Estructura jerárquica de la biodiversidad genética y de especies y de los procesos involucrados a distintas escalas espaciales	15
<b>Tabla 3.1.</b> Relación de secuencias del gen 16S rRNA obtenidas de clonas ambientales. Se incluyen los mejores “hits” de RDPII y, en ciertos casos se incluye la información obtenida de NCBI (donde la filogenia no coincidía con la identidad reportada por RDPII). Se especifica el OTU al que corresponde cada clona/secuencia.	46
<b>Tabla 4.1.</b> Coordenadas geográficas y medidas fisicoquímicas del agua de los sitios de muestreo en el sistema Churince en Cuatro Ciénegas Coahuila México	57
<b>Tabla 4.2.</b> Oligonucleótidos utilizados para la amplificación por PCR de los genes estudiados	66
<b>Tabla 4.3.</b> Oligonucleótidos utilizados para la secuenciación de amplicones insertados en vectores de clonación	67
<b>Tabla 4.4.</b> Resumen de resultados de las pruebas de Mantel para distancia genética estimada por similitud de patrones rep-PCR y condiciones ambientales de los sitios de colecta	79
<b>Tabla 4.5.</b> Resultados del test de incongruencia filogenética (ILD)	89
<b>Tabla 4.6.</b> Resultados de la prueba de recombinación para un sitio y sitios múltiples	95
<b>Tabla 4.7.</b> Diversidad genética por locus de toda la muestra y linajes principales de <i>Pseudomonas</i>	98
<b>Tabla 4.8.</b> Resultados de la prueba de McDonald-Kreitman de selección en parejas de datos	99
<b>Tabla 4.9.</b> Valores de diferenciación genética entre las muestras de verano e invierno	100
<b>Tabla 4.10.</b> Valores de diferenciación genética entre sitios de colecta	101
<b>Tabla 4.11.</b> Valores de la prueba de $X^2$ para cada valor de $F_{ST}$	102
<b>Tabla 4.12.</b> Valores de diferenciación genética entre los tres linajes más abundantes de la muestra total.	104

## ÍNDICE DE FIGURAS

	<i>Pág.</i>
<b>Figura 2.1.</b> Mapa del área de Cuatro Ciénegas en el estado de Coahuila en México	28
<b>Figura 2.2.</b> Representación gráfica del sistema Churince	29
<b>Figura 4.1.</b> Representación gráfica de los sitios de muestreo en el sistema Churince en Cuatro Ciénegas, Coahuila, México	57
<b>Figura 4.2.</b> Cluster Pearson-UPGMA de los patrones BOX-PCR de las cepas de <i>Pseudomonas</i> estudiadas	78
<b>Figura 4.3a.</b> Genealogía del gen 16S rRNA para cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	84
<b>Figura 4.3b.</b> Genealogía del gen <i>acnB</i> para cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	85
<b>Figura 4.3c.</b> Genealogía del gen <i>gyrB</i> para cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	86
<b>Figura 4.3d.</b> Genealogía del gen <i>recA</i> para cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	87
<b>Figura 4.3e.</b> Genealogía del gen <i>rpoD</i> para cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	88
<b>Figura 4.4a.</b> Análisis de Split-Decomposition para todas las cepas de <i>Pseudomonas</i> colectadas en el sistema Churince	92
<b>Figura 4.4b.</b> Análisis de Split-Decomposition de los loci <i>acnB</i> , <i>gyrB</i> , <i>recA</i> y <i>rpoD</i> para cepas de <i>Pseudomonas</i> de los linajes <i>P. otitidis</i> , <i>P. psychrotolerans</i> y " <i>P. caenicarum</i> " colectadas en el sistema Churince	93
<b>Figura 4.4c.</b> Análisis de Split Decomposition de los loci <i>acnB</i> , <i>gyrB</i> , <i>recA</i> y <i>rpoD</i> para cada uno de los linajes numéricamente dominantes: <i>P. otitidis</i> , <i>P. psychrotolerans</i> y " <i>P. caenicarum</i> " colectadas en el sistema Churince	94
<b>Figura 4.5.</b> Representación gráfica de la distribución espacial de las cepas correspondientes a cada uno de los linajes más abundantes de la muestra de <i>Pseudomonas</i> de Churince en Cuatro Ciénegas, Coahuila	103
<b>Figura 4.6</b> Representación gráfica de las simulaciones de coalescencia (skyline plots) para cada linaje analizado.	105



## RESUMEN

En este trabajo se estudiaron los niveles de diversidad y distribución espacial de procariontes acuáticos en el valle de Cuatro Ciénegas a diferentes escalas espaciales y jerárquicas (dentro y entre especies). En primer lugar, se presenta un estudio de comunidades no cultivadas, con énfasis en los niveles de diversidad a distintas escalas espaciales y en la diferenciación de los sitios muestreados. Los resultados muestran que, la composición y abundancia de organismos de cada uno de los sitios es distinta, haciendo que la diversidad beta y gamma sean altas, no siendo así para la diversidad alfa. En segundo lugar, se presenta un estudio de diversidad genética a nivel de poblaciones cultivadas de *Pseudomonas*, en uno de los sistemas del valle que presenta un pronunciado gradiente ambiental y en donde el énfasis es en la influencia de la separación geográfica y de condiciones ambientales en la diversificación. Los resultados muestran que 1) Existe distribución espacial diferencial para ciertas variantes genéticas pero no se identifican los factores ambientales involucrados en ello; 2) Existe estructura clonal con poca o nula recombinación en las poblaciones sugiriendo una dinámica epidémica o de selección periódica. Las conclusiones del trabajo son: 1) La diversidad de microorganismos en Cuatro Ciénegas es fundamentalmente debida a la distribución discontinua de grupos genéticamente distintos tanto en espacio como en tiempo, siendo esto cierto tanto a escala de comunidades como de poblaciones 2) Aunque no fue posible determinar los factores ambientales que pudieran estar relacionados con la distribución observada de la diversidad, el estudio poblacional sugiere una dinámica definida por fuerzas evolutivas de carácter aleatorio.

## ABSTRACT

In the present work, we studied the levels and distribution of aquatic prokaryotic diversity within the Cuatro Ciénegas valley at different spatial and taxonomic scales (within and among species). In first place, a community level analysis is presented, with special attention to the diversity levels at different spatial scales and differentiation among sites. The results show that community composition of each site is considerable different, producing high levels of beta and gamma diversity, not being the case for alpha diversity. Secondly, a population genetics approach with cultured *Pseudomonas* was pursued in one of the hydrological systems of the valley, where a steep environmental gradient is evident, the emphasis here, was in the influence of geographic distance and environmental conditions in population diversification. Results show 1) Differential spatial distribution for genetic variants but no association of genetic distance with environmental conditions; 2) There is a clonal population structure with low or none recombination, suggesting an epidemic structure or periodic selection. The main conclusions of the study are: 1) Diversity of microorganisms in Cuatro Ciénegas is mainly a result of discontinuous distribution of genetically distinct groups both in space and time, being this true at community and population scales; 2) Despite the impossibility to determine the environmental factors involved with the observed diversity distribution, the population level study suggests a dynamic driven by stochastic evolutionary forces.

## PRÓLOGO

*“Since processes which produce biological diversities operate differently, and at different rates according to the position of biological phenomena along the scales of space, time and change, many theories and paradigms are probable more complementary than conflicting” Blondel, 1987.*

La diversidad de organismos que encontramos en la Tierra ha llamado la atención del hombre desde siempre, y esa diversidad ha sido el punto de partida para el estudio de la evolución y ecología de los organismos. La diversidad biológica tiene dos componentes inseparables; por un lado, existe la diversidad dentro de especies (que puede ser vista como variación genética) y por el otro lado, existe la diversidad entre especies, esto es la cuenta de linajes distintos. Además, las diferentes escalas espaciales y temporales a las que se puede estudiar la diversidad de organismos nos permiten aproximarnos a los factores que intervienen en el origen y mantenimiento de ésta. Por ejemplo, los estudios sobre diversidad de poblaciones de una misma especie nos acercan a entender los procesos que suceden en escalas temporales cortas al analizar los factores que influyen directamente en la sobrevivencia de los individuos y adaptación de las poblaciones, mientras que, estudios de la diversidad en comunidades a nivel regional nos permiten resolver cuestiones a escalas temporales mucho mayores en donde los factores involucrados resultan, por ejemplo, en patrones biogeográficos.

Este trabajo es una aproximación al estudio de la diversidad de procariontes e intenta presentar una perspectiva de la importancia del estudio de la diversidad biológica a diferentes escalas espaciales y jerárquicas (dentro y entre especies). Se da una visión personal de la contribución de este tipo de trabajos al conocimiento y de las perspectivas de los futuros estudios sobre diversidad de microorganismos en un contexto ecológico y

evolutivo. Finalmente se presentan algunas de las actuales dificultades prácticas y teóricas en el desarrollo de estudios sobre ecología microbiana, así como las herramientas técnicas y de análisis con las que se cuenta hoy en día, algunas de las cuales fueron utilizadas en el estudio desarrollado en la tesis.

### **Objetivo general y presentación**

Dentro del contexto de diversidad a diferentes escalas espaciales y temporales, la pregunta general del trabajo de investigación presentado en esta tesis es:

*¿Es posible entender los patrones de distribución y abundancia de especies (comunidades) si conocemos la dinámica evolutiva dentro de las especies (poblaciones)? o acaso en cada nivel jerárquico de biodiversidad ¿hay propiedades emergentes y no escalables?*

Para acercarnos a resolver esto, el objetivo general del trabajo fue estudiar los niveles de diversidad y distribución espacial de microorganismos acuáticos en el valle de Cuatro Ciénegas, Coahuila, una región especial por las condiciones únicas de sus sistemas acuáticos y poco estudiada en cuanto a su microbiota.

La primera parte del trabajo presenta el estudio de comunidades procariontes en el valle, con énfasis en los niveles de diversidad a distintas escalas espaciales y en la diferenciación de las comunidades de los sitios muestreados. La segunda parte del trabajo consta de un estudio sobre variación genética a nivel de poblaciones de bacterias del género *Pseudomonas* en uno de los sistemas del valle, el cual presenta un claro gradiente ambiental. En esta parte el énfasis es en la influencia de la separación geográfica y de las

condiciones ambientales en la diferenciación genética o diversificación de las poblaciones analizadas.

La tesis está compuesta por cinco capítulos y dos apéndices. El primer capítulo corresponde a la introducción general, en donde se abordan cuestiones sobre el estudio a diferentes escalas de la diversidad biológica, diversidad de microorganismos y controversias actuales sobre el tema. El segundo capítulo presenta información sobre el área de estudio. El tercer capítulo está integrado por el artículo “Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas basin” en donde se exploran los niveles y distribución de diversidad procarionte en el valle de Cuatro Ciénegas. El cuarto capítulo desarrolla una introducción a las ideas de la genética de poblaciones y a las herramientas usadas en el trabajo para evaluar la estructura genética de poblaciones bacterianas, así como los resultados del análisis de variación genética de poblaciones de *Pseudomonas* acuáticas de vida libre en el área de estudio. El quinto capítulo contiene las conclusiones generales y perspectivas del trabajo. Finalmente se encuentran: el Apéndice I, que se compone de un manuscrito en donde se describe una nueva especie del género *Pseudomonas* y que forma parte de los linajes estudiados en el capítulo cuarto; y el Apéndice II, que contiene un manuscrito que describe las diferentes estrategias moleculares y de análisis actualmente utilizadas en ecología microbiana.

Esta tesis fue desarrollada en el Laboratorio de Evolución Molecular y Experimental del Instituto de Ecología de la UNAM, bajo la dirección de la Dra. Valeria Souza Saldivar. El proyecto fue financiado con una beca doctoral de CONACyT y apoyos de NASA (NCCS2-1051) y SEMARNAT/CONACyT (2002-CO1-0246) a Valeria Souza.

## Capítulo 1. INTRODUCCIÓN GENERAL

La diversidad biológica es el conjunto de las variantes de organismos a todos los niveles de organización biológica, desde los genes hasta las comunidades. En biología uno de los objetivos fundamentales es entender cómo se genera y regula esta diversidad. Por ello es que la biodiversidad se ha estudiado a muchos niveles. Podemos estudiar desde la variación genética dentro de poblaciones de una especie y la diversidad entre poblaciones distantes de una misma especie, hasta la diversidad de especies que encontramos en comunidades o ecosistemas. Esto nos muestra que la diversidad biológica tiene muchas dimensiones y resumirla en un simple número de especies en muchas ocasiones puede limitar el entendimiento conceptual de ésta.

### *1.1. Importancia del estudio de la diversidad biológica*

Desde el punto de vista de la teoría evolutiva y ecológica, el estudio de la diversidad biológica constituye el centro del desarrollo de ambas. Por un lado, la teoría evolutiva, con las contribuciones conceptuales de Charles Darwin, logró establecer que la variación entre individuos de una misma especie era fundamental para que la selección natural, actuara sobre los individuos y de esta manera contribuyera al moldeamiento de las poblaciones en el proceso adaptativo. Es alrededor del proceso adaptativo y del estudio de la diversidad genética que se ha construido y desarrollado la teoría evolutiva tratando de resolver fundamentalmente cuestiones relacionadas con los mecanismos que mantienen, disminuyen u originan la variación genética dentro y entre especies. Por otro lado y muy ligada, la teoría ecológica se ha interesado en conocer los factores que regulan la diversidad en tiempo y espacio y las consecuencias de esta diversidad en los procesos ecosistémicos. Comprender la ecología de los organismos significa entender los cambios sutiles en el tiempo y espacio que constantemente llevan a la adaptación de estos o de sus poblaciones y

que resultan en la construcción de organizaciones biológicas tan complejas como ecosistemas tropicales. De esta forma es como la ecología y la evolución, se relacionan tan íntimamente alrededor del estudio de la diversidad biológica desde diferentes perspectivas. Así pues se puede afirmar que comprender la ecología es entender, en parte, la evolución.

### *1.2 Estudio de la diversidad biológica a diferentes escalas*

Dentro de la ecología los patrones de distribución de la diversidad de especies o patrones de las comunidades han sido muy estudiados. Algunos autores han sugerido que estos patrones son la consecuencia de procesos jerárquicos que interactúan de manera compleja para modelar la diversidad de la vida en el planeta (Morin y McGrady-Steed 2004). Debido a esto, ha surgido un gran interés en estudiar las diferentes escalas temporales y espaciales de la variación biológica. La perspectiva dada por una visión jerárquica nos permite relacionar las interacciones entre los procesos ecológicos locales y los procesos evolutivos que actúan a escalas regionales y continentales.

Una de las jerarquizaciones de la diversidad biológica más reconocidas es la que se refiere al área o escala espacial a la que se estudia. En este sentido la más utilizada es la propuesta por Cody (1975), en donde los términos diversidad alfa, beta y gamma se refieren a escalas local, de paisaje y regional respectivamente. Whittaker y colaboradores (2001) discuten la importancia de estudiar la diversidad de especies a distintas escalas para llegar a entender las variables que influyen en los niveles de diversidad de especies y en su distribución. Willis y Whittaker (2002) plantean el error que se ha cometido al tratar de encontrar “EL” mecanismo que explique patrones de riqueza de especies, con el supuesto implícito de que sin importar la escala a la que se estudie la diversidad, las explicaciones que resulten pueden ser escaladas hacia arriba o hacia abajo simplemente. Sin embargo,

como ha sido sugerido por algunos (Eldredge 1985, Gould 1998), los sistemas biológicos y ambientales son más complejos que esto.

Existe una variedad de factores a escala espacial y temporal que contribuyen al entendimiento de la biodiversidad, por esto es que una aproximación jerárquica podría ser más apropiada para modelar adecuadamente la distribución de la riqueza de especies. En un modelo jerárquico, los procesos podrían anidarse espacial y temporalmente y puede ser que alguna señal de los procesos de una escala dada se refleje en las escalas vecinas, pero ha sido demostrado de manera clara por distintos estudios que distintas variables o propiedades emergen a distintas escalas (Rahbeck y Graves 2001, Crawley y Herral 2001, Lennon *et al.* 2001). En este sentido hay que decir que, la diversidad genética es difícil de equiparar con la escala o jerarquías espaciales establecidas en ecología (Tabla 1.1., Mayr 1994), a pesar de que está intrínsecamente relacionada con la diversidad de especies a cualquier escala espacial y temporal. Esto es fundamentalmente debido a que, las jerarquías taxonómicas y de diversidad genética, aunque puedan encontrarse todas ellas en escalas locales, pueden reflejar grandes escalas temporales al implicar orígenes y relaciones evolutivas en contextos de miles a millones de años (Tabla 1.1.). Al mismo tiempo, las jerarquías espaciales involucran escalas temporales que van desde 1 año hasta 100 millones de años, desde procesos de microevolución (genética de poblaciones) hasta especiación y formación de grandes linajes (macroevolución). Sin embargo no podemos ignorar que la diversidad genética y los procesos que la generan y mantienen son participantes de importancia considerable en los patrones de distribución y de riqueza de especies tan estudiados en ecología.

En la Tabla 1.1, se presenta una propuesta de síntesis sobre las escalas de estudio de la diversidad, los factores posiblemente involucrados en los procesos a distintas escalas y la escala temporal correspondiente para estos procesos.

Tabla 1.1. Estructura Jerárquica de la Biodiversidad genética y de especies (modificado de Willis y Whittaker 2002) y de los procesos involucrados a distintas escalas espaciales

Diversidad genética	Escala taxonómica	Escala Espacial	Escala de estudio	Escala de riqueza de especies	Variabes ambientales predominantemente responsables	Escala temporal a la que los procesos ocurren
+	Poblaciones	Local	Alfa	Riqueza de especies dentro de comunidades, dentro de parches de habitat.	Interacciones bióticas y abióticas a escalas finas (estructura de habitat, disturbios por incendios, tormentas, competencia)	Procesos que ocurren en escalas temporales de ~1-100 años
++	Especies	Paisaje	Beta	Riqueza de especies entre comunidades; recambio de especies dentro del paisaje	Suelos, altura, efecto peninsular	Procesos que ocurren en escalas temporales de ~100-1000 años
+++	Género	Regional	Gamma	Riqueza de especies en grandes áreas geográficas dentro de continentes	Radiación y disponibilidad de agua, área, latitud	Procesos que han ocurrido en los últimos 10,000 años, desde la última glaciación
++++	Orden	Continental		Diferencias en linajes de especies y riqueza a través de continentes	Eventos de acidificación, ciclos glacial/interglacial, aparición de montañas	Procesos que han ocurrido en los últimos 1-10 millones de años
+++++	Familia	Global		Diferencias reflejadas en provincias biogeográficas	Movimientos de placas continentales, cambios en el nivel del mar	Procesos que han ocurrido en los últimos 10-100 millones
++++++	Reino					



información sobre diversidad, distribución e interacciones de microorganismos ha comenzado a acumularse y a contribuir al marco teórico de la ecología (Hughes *et al.* 2006). Algunas de las estrategias de estudio independientes del cultivo de microorganismos se explican y discuten en el Apéndice II de la tesis.

#### *1.4. Diversidad de microorganismos*

A pesar de la importancia que tiene en ecología el conocer los niveles de diversidad y distribución de los organismos, el mundo microbiano ha sido escasamente estudiado desde este punto de vista (Hughes *et al.* 2006). Como se mencionó, la razón principal es que la mayor parte de estos organismos no han podido ser recuperados del ambiente en cultivo. Sin embargo, en años recientes las herramientas independientes del cultivo han permitido empezar a conocer la diversidad del mundo microbiano. Entre los datos que se tienen actualmente sobre esto se encuentra el cálculo sobre la biomasa de microorganismos terrestres en la superficie de nuestro planeta, la cual ha sido estimada como igual a la de todas las plantas marinas y terrestres y tal vez sea el constituyente principal de la biomasa de la Tierra (Whitman *et al.* 1998, Pedersen 2000). Esto implica que, un porcentaje mayoritario de la vida en la Tierra ha sido poco estudiado en cuanto a su diversidad y ecología.

En los pocos años en los que se han hecho estudios microbiológicos con enfoque ecológico, se han logrado establecer algunas características sobre la dinámica microbiana en el ambiente y las características de diversidad biológica que ciertos ambientes presentan. Los ambientes acuáticos parecen presentar menos diversidad que los suelos y los sedimentos (Ritz *et al.* 1997, Torsvik *et al.* 1998, Øvreas *et al.* 2001). A pesar de que la diversidad genómica de agua y suelo puede ser distinta, el número máximo de individuos que integran el taxón dominante es el mismo  $\approx 10^4$ - $10^5$  individuos por gramo o mililitro.

Este hallazgo sugiere que, debe existir algún mecanismo o mecanismos que controlen la densidad de los taxa y que estos deben funcionar de manera más o menos similar en todos los ambientes, mientras que los mecanismos que controlan la abundancia total de la comunidad bacteriana funcionan de modos distintos en agua y suelo. Existen algunas ideas sobre los mecanismos que controlan la diversidad procarionte que se desarrollan a continuación.

#### 1.4.1. Interacciones tróficas

Hutchinson (1961) se preguntaba por qué hay tantas especies de fitoplancton en un medio aparentemente homogéneo, en donde todas las especies presentes parecen competir por el mismo tipo de nutrientes. Dentro de los procariontes, los competidores pueden coexistir si existe algún mecanismo de pérdida selectiva de especies que evite que los competidores más exitosos secuestren todos los recursos (Fuhrman 1999, Simek *et al.* 1999, Bohannan y Lenski 2000). Por ejemplo, puede existir depredación selectiva en tamaño llevada a cabo por protozoarios, lo que permitiría la coexistencia de bacterias y fitoplancton de diferentes tamaños, este tipo de interacciones determina cómo se distribuye la biomasa en grupos funcionales (Torsvik *et al.* 2002). Otra interacción frecuente en el mundo de las comunidades microscópicas es la que existe entre virus y bacterias, esta interacción es generalmente específica y ello permite la coexistencia de varios taxa de bacterias en una comunidad; e incluso ha sido asociada con la especiación dentro de grupos funcionales (Torsvik *et al.* 2002).

#### 1.4.2. Perspectiva evolutiva

Debido a las altas tasas de especiación observadas en trabajos sobre evolución experimental en bacterias se ha sugerido que la gran diversidad bacteriana que se detecta es simplemente una consecuencia de que las tasas de especiación son mayores que las tasas de

extinción (Dykhuizen 1998). Un ejemplo de la evidencia que soporta este argumento, es el caso del surgimiento de 3 “especies” ecológica y genéticamente distintas a partir de una especie original de *Escherichia coli* en el transcurso de 800 generaciones, lo que sugiere una rápida adaptación y partición del nicho llevando a la eventual diferenciación genética (Helling *et al.* 1987, Rosenzweig *et al.* 1994).

#### 1.4.3. Heterogeneidad espacial

La complejidad estructural del suelo y de los sedimentos es importante para la diversificación a nivel de poblaciones porque esta complejidad permite que los recursos sean fraccionados y de esta manera se creen nuevos nichos con los que se incrementa la especialización y división en especies ecológicas nuevas (Torsvik *et al.* 2002).

#### 1.4.4. Heterogeneidad temporal

La mayoría de las comunidades terrestres sufren perturbaciones intermitentemente como: escasez de alimento, sequía, congelamiento/descongelamiento o la actividad humana. Estas condiciones ambientales alteradas y los recursos que se liberan crean oportunidades para que se establezcan nuevas especies. Así, la perturbación asegurará que las comunidades incluyan una mezcla de los diferentes estadios de sucesión (Connell 1978).

Sin embargo, una perturbación fuerte y frecuente causará la desintegración de los microhabitats y la disrupción de los límites entre poblaciones, lo que permitirá que los recursos locales estén disponibles para una mayor proporción de la masa microbiana total (Torsvik *et al.* 2002), es decir más individuos pero menos especies.

### *1.5. Estudios recientes y controversias actuales sobre diversidad y distribución de microorganismos*

Como hemos mencionado, los ecólogos modernos buscan entender y explicar en términos generales cómo se genera y mantiene la biodiversidad, así como el origen y los mecanismos de las interacciones de los organismos entre sí y con el medio ambiente con el objetivo último de predecir el resultado de determinadas condiciones. Para cumplir con su objetivo se requiere de la construcción de teorías generales y de modelos de la realidad que generen predicciones a probar.

La construcción de modelos en ecología y en cualquier disciplina científica, requiere de la acumulación de información y de la búsqueda de patrones que resulten de dicha acumulación. Una de las ramas de la ecología es la biogeografía, la cual es el estudio de los patrones de distribución de la biodiversidad en espacio y tiempo con el objetivo de dilucidar los mecanismos que generan y mantienen la diversidad, tales como especiación, extinción, dispersión e interacciones entre especies. Esto implica que la información sobre diversidad y distribución nos habla de los procesos y mecanismos ecológicos y evolutivos de los organismos.

Los datos de distribución en microorganismos no son abundantes, por el rezago asociado a técnicas microbiológicas de cultivo. Sin embargo, las estrategias actuales independientes del cultivo han permitido resolver preguntas relacionadas con la diversidad de microorganismos, distribución y factores implicados en ella (McCaig *et al.* 1999, Dunbar *et al.* 1999 y 2000, Hoshino *et al.* 2001, Zhou *et al.* 2002). Todos estos esfuerzos están contribuyendo rápidamente a la generación de bases de datos sobre la abundancia y distribución de las especies microbianas, información valiosa para empezar a probar modelos ecológicos en este grupo de organismos hasta poco tiempo inexplorado.

Una de las discusiones actuales en ecología microbiana surge alrededor de la afirmación de que para los microorganismos “todo está en todos lados y el ambiente

*selecciona*” (Bass-Becking 1934). Esta afirmación es resultado de inferir que dado el tamaño y gran abundancia de los microbios existe una elevada tasa de sobrevivencia y dispersión geográfica, lo que genera distribuciones cosmopolitas (Finlay 2002). Esta frase implicaría en su primera parte, que en microorganismos no existen patrones de distribución y en la segunda parte que en el caso de que existiera distribución por parches, esta estaría definida únicamente por las condiciones ambientales.

En cuanto a la primera parte del argumento de Bass-Becking: “*todo está en todos lados...*” en la actualidad ya existen varias publicaciones enfocadas a analizar esto. Hay evidencia sobre algunos grupos de microorganismos que tienen distribuciones geográficamente restringidas debido a limitaciones en dispersión (Papke *et al.* 2003, Hughes *et al.* 2006) y a aislamiento geográfico de poblaciones de la misma especie (Papke y Ward 2004). Además, existen también datos sobre endemismos microbianos (Whitaker *et al.* 2003) y sobre patrones espaciales de diversidad microbiana no azarosos (Noguez *et al.* 2005, Green *et al.* 2004, Horner-Devine *et al.* 2004), lo cual es una demostración de que tienen biogeografía, es decir, no “*todo está en todos lados...*”.

Sin embargo, a pesar de la evidencia de biogeografía, la segunda parte del argumento mencionado “*...el ambiente selecciona*” podría ser al menos parcialmente cierta, pues en algunos casos hay datos de la fuerte influencia de factores ambientales en la distribución de microorganismos y en la diferenciación de comunidades y poblaciones (Casamayor *et al.* 2002, Yannarell y Triplett 2005, Hughes *et al.* 2006). En este sentido, es de suma importancia probar explícitamente hipótesis sobre los factores implicados en los patrones de distribución observados. Alrededor de este problema existen algunos trabajos que han estudiado patrones de distribución de diversidad tratando de separar el efecto de factores puramente ambientales, de factores históricos o asociados al aislamiento

geográfico de poblaciones de microorganismos (Whitaker 2003, Green *et al.* 2004, Reche *et al.* 2005, Fierer y Jackson 2006). Este tipo de aproximaciones son importantes porque, a pesar de que para los macroorganismos se observa que, la similitud genética en poblaciones y de composición de especies en comunidades decae con la distancia geográfica como reflejo de limitación en la dispersión de individuos o flujo génico (Tuomisto *et al.* 2003), poco se sabe de esto en microorganismos (Green y Bohannan 2006).

Aunque aún existe cierto grado de controversia en cuanto a la existencia de patrones biogeográficos en microorganismos y sobre todo en cuanto a los métodos para evaluarlos (Hughes *et al.* 2006), la discusión de este asunto, y los resultados de estudios que lo aborden a distintas escalas espaciales y temporales, es de gran importancia para el futuro de la ecología microbiana. Estudios que aborden el análisis de los patrones de distribución de diversidad microbiana, contribuirán a responder la pregunta subyacente de cuáles son los factores ecológicos y evolutivos que generan y mantienen la diversidad en estos organismos, enriqueciendo de esta manera el conocimiento biológico ya sea mostrando nuevos comportamientos o reforzando la teoría actual con datos del mundo microbiano.

## Capítulo 2. ÁREA DE ESTUDIO

Este proyecto aborda el estudio sobre diversidad de procariontes a dos escalas espaciales, ambas en los sistemas acuáticos de Cuatro Ciénegas en el estado de Coahuila. La primera parte consta de un estudio a escala regional, abordando la diversidad a nivel de comunidades y utilizando herramientas de análisis independientes del cultivo, en este caso se tomaron muestras de agua de cuatro de los siete sistemas descritos por Minckley (1969): Churince, Becerra/Garabatal, Río Mesquites y Tío Cándido/Escobedo. La segunda parte consta de un estudio a escala local, tomando como modelo uno de los sistemas de flujo de agua: Churince, en este análisis el enfoque fue a nivel de poblaciones de bacterias cultivables del género *Pseudomonas*.

### *2.1 El área natural protegida*

El valle de Cuatro Ciénegas es uno de los pocos humedales que quedan en el norte de América y ha sido comparado con las Islas Galápagos del Ecuador en cuanto a diversidad biológica y endemismos, debido a que un gran número de los organismos que se encuentran en el valle están en la lista de especies amenazadas, a que más de 70 especies son endémicas de la región (Desert Fishes Council Web, 2003) y a que todos ellos dependen de la permanencia de los sistemas acuáticos de la zona. Desde 1994 el valle de Cuatro Ciénegas fue decretado Área Natural Protegida (ANP) por el gobierno de México y es también considerado como un sitio prioritario para la conservación (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), World Wildlife Fund (WWF), Ramsar convention on wetlands), y más recientemente ha sido declarado como un sitio de interés por el proyecto del Hombre y la Biosfera (MAB/UNESCO). Además, muchos de los habitats acuáticos del valle permiten la existencia de estromatolitos vivos, comunidades microbianas completas que son consideradas como un ejemplo

---

de la vida en la Tierra de hace casi 3 mil millones de años (Madigan *et al.* 2000) esta situación es poco usual en la actualidad, y en todo el mundo existen tan sólo unos cuantos sitios en donde encontramos estromatolitos vivos (Madigan *et al.* 2000). Finalmente, cabe señalar que a pesar de que la información sobre diversidad de macroorganismos en la zona es abundante y es evidencia de que Cuatro Ciénegas es un reservorio importante de biodiversidad (Badino *et al.* 2004), la diversidad microbiana de este lugar ha sido escasamente estudiada (García-Pichel *et al.* 2002, Elser *et al.* 2005, Souza *et al.* 2006).

## *2.2 Condiciones generales del área de estudio*

### *2.2.1 Ubicación geográfica*

El valle de Cuatro Ciénegas está localizado en la zona central del estado de Coahuila, México (Fig. 2.1) y forma parte del Desierto Chihuahuense (26° 59' N, 102° 04' W). La extensión del valle es de aproximadamente 40 km de este a oeste y de 30 km de norte a sur a una altitud de 740 msnm (Minckley 1969).

Este valle está localizado entre los dos macizos montañosos más grandes de México, la Sierra Madre Oriental y la Sierra Madre Occidental. La humedad que viene de océano Pacífico y del Golfo de México no logra llegar al valle por la barrera montañosa, lo que en parte explica las condiciones áridas del lugar.

Rodeando al valle se encuentran montañas de más de 3000 msnm: la Sierra de la Madera, Sierra Purísima, Sierra Menchaca y Sierra la Fagua y es dividido en región este y oeste por la Sierra de San Marcos y Pinos (Fig. 2.1).

### *2.2.2 Clima*

El clima de Cuatro Ciénegas es árido, con menos de 200 mm de precipitación anual y más de 2000 mm de evaporación potencial (SEMARNAT 2003). Las lluvias ocurren entre mayo y

octubre, con la mayor parte de la precipitación durante junio. Las temperaturas en Cuatro Ciénegas pueden llegar al congelamiento (0°C) durante el invierno, cuando se acumula nieve en las partes altas de las montañas. Las temperaturas de verano suben a más de 44°C (Minckley 1969). Se estima que un clima similar al actual ha existido en el valle por más de 30,000 años con vegetación similar inferida por datos de polen y radioisótopos de carbono (Meyer 1973).

### 2.2.3 Hidrología

La poca precipitación anual reportada para la zona no puede ser generadora de cuerpos de agua en el valle, por lo que la gran cantidad de agua subterránea es la causa de la existencia de manantiales. Dentro de Cuatro Ciénegas, la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), ha reportado la existencia de cerca de 300 pozas, aunque la mayoría se han señalado como no permanentes (Minckley 1969). La densidad de estas pozas es de 12 a 15 por kilómetro cuadrado, siendo especialmente densas cerca de la Sierra San Marcos (Fig. 2.1, Winsborough 1990). El origen y procesos por los cuales han pasado los cuerpos de agua del valle han sido muy controvertidos históricamente. Resultados de análisis hidrológicos indican que la química del agua de Cuatro Ciénegas es consecuencia de la evaporación de agua principalmente, pero existen otros posibles procesos involucrados, que incluyen flujo subterráneo entre valles y mezcla con agua subterránea más profunda e incluso con valles vecinos (Johannesson *et al.* 2004, Evans 2005, Souza *et al.* 2006, Minckley y Jackson 2008). Debido a los escasos estudios sobre hidrogeología de la zona, la polémica sobre el origen del agua de las pozas del valle no ha podido resolverse del todo hasta la fecha.

Las pozas del valle presentan una variabilidad extrema en temperatura, salinidad, química del agua y descarga en escalas espaciales pequeñas o locales (Minckley 1969). Además, las características limnológicas son muy variadas en las más de sus 200 pozas. Las aguas son duras y

contienen gran cantidad de sales de calcio y magnesio; la elevada dureza de las aguas se incrementa a partir de los manantiales hacia las lagunas terminales. Los cationes dominantes, en orden de abundancia descendente son calcio, magnesio, sodio y potasio; mientras que los aniones son fuertemente dominados por los sulfatos, carbonatos y cloruros. Presentan una composición química similar; el pH varía desde la neutralidad (7.0-7.2) en los manantiales hasta la alcalinidad (8.0-9.7) en lagos terminales en fase de desecación (Alcocer y Kato 1995, Calegari 1997). Los cuerpos de agua tienen, además, un contenido muy bajo de fósforo y la pobreza de nutrientes no permite el desarrollo importante de algas, por lo que se ha sugerido que la base trófica del ecosistema está formada por bacterias (Souza *et al.* 2006). Sin embargo, aún no existen estudios en donde se pruebe esto explícitamente. Se ha sugerido también que el bajo contenido de fósforo es una firma de condiciones ambientales del PreCámbrico (Elser *et al.* 2005, 2006), en donde organismos metazoarios estaban limitados en su desarrollo. Esta característica de los sistemas acuáticos de Cuatro Ciénegas permite probar hipótesis sobre evolución y evolución temprana de la vida en un ambiente actual (Souza *et al.* 2008).

Los cuerpos de agua permanentes en el valle se han clasificado, de acuerdo con Minckley (1969), en siete sistemas de flujo, estos son: Churince, Becerra/Garabatal, Río Mesquites, Río Puente Chiquito, Tío Candido/Escobedo, Santa Tecla y Río Nadadores (Río Grande). Dentro de cada uno de los sistemas existen todos o algunos de los cuatro tipos principales de cuerpos de agua en el valle: manantiales (en donde surge el agua), pozas (alimentadas por flujos superficiales), ríos/canales y lagunas saladas.

### *2.3 Sistema Churince*

El sistema Churince está compuesto por cuatro pozas/lagunas (Fig. 2.2). Estas incluyen Poza Churince y Poza Bonita como sus manantiales, Laguna Intermedia como un lago intermedio

---

y Laguna Grande como el lago terminal o laguna salada de desecación. Se ha propuesto que hay conductos subterráneos que mantienen la conexión de este sistema con el Río Mesquites (Evans 2005), pero hasta ahora no hay evidencia para ello y se considera como un sistema aislado.

Un trabajo reciente sobre la hidrología de los sistemas acuáticos de Cuatro Ciénegas (Evans 2005) muestra datos fisicoquímicos (Fig. 2.2b) en donde se observan diferencias entre el manantial y la laguna intermedia, esta observación se correlaciona con ligeros incrementos en pH y ciertos iones (Fig. 2.2c).

Las diferencias químicas y fisicoquímicas entre Laguna Intermedia y Laguna Grande son más pronunciadas y es notable el incremento en la concentración de iones en la Laguna Grande en comparación con el resto del sistema (Fig. 2.2c).

Finalmente, es importante notar también las diferencias que existen entre los distintos sitios del sistema en cuanto a estabilidad temporal. Evan Carson (Arizona State University, comunicación personal) siguió los cambios de temperatura en distintos sitios del sistema a lo largo de 3 años y la Figura 2.2a muestra los resultados en donde es evidente que los cambios estacionales producen amplias fluctuaciones en temperatura hacia la Laguna Grande, mientras que en Poza Bonita y Churince estos cambios no son significativos.



Fig. 2.1. Mapa del área de Cuatro Ciénegas en el estado de Coahuila en México. Los nombres de las Sierras que rodean el valle se encuentran sobre la representación gráfica de las mismas (modificada de Johannesson *et al.* 2004)

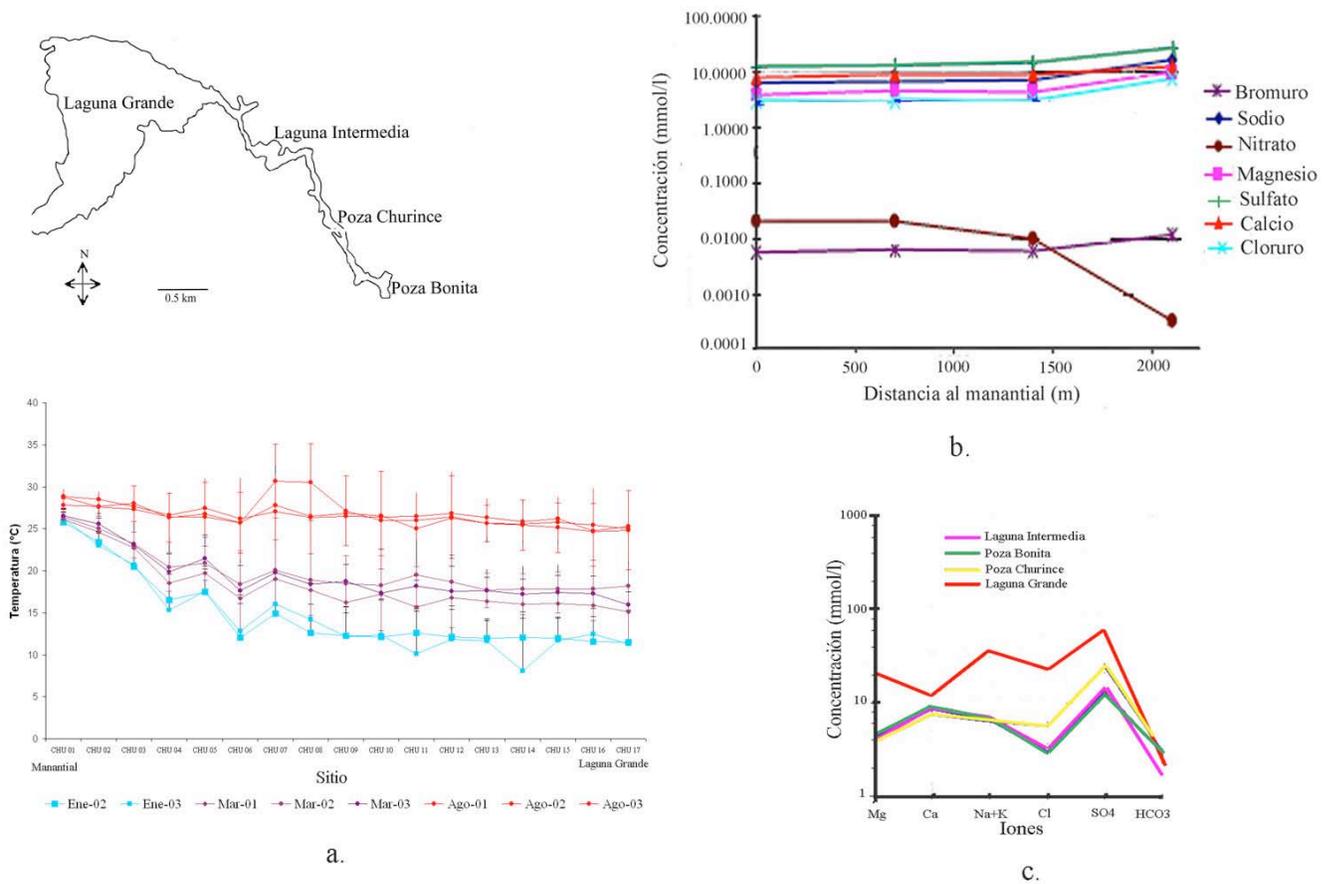


Fig.2.2.Representación del sistema Churince en donde se indican: Poza Bonita y Poza Churince como manantiales, Laguna Intermedia y Laguna Grande. a. Fluctuaciones estacionales en temperatura a lo largo del sistema Churince. b. Fisisicoquímica del sistema Churince. c. Concentración de iones en Poza Bonita, Poza Churince, Laguna Intermedia y Laguna Grande del sistema Churince (Evans 2005).

### **Capítulo 3. DIVERSIDAD DE COMUNIDADES ACUÁTICAS PROCARIONTES EN EL VALLE DE CUATRO CIÉNEGAS**

#### **Resumen**

El valle o bolsón de Cuatro Ciénegas (Coahuila, México) es un arreglo de diferentes sistemas acuáticos en el desierto Chihuahuense y tiene niveles inusualmente altos de endemismos de diferentes taxa. A pesar de que las poblaciones de organismos macroscópicos de Cuatro Ciénegas han sido bien descritas, se sabe poco sobre la diversidad taxonómica y distribución de microorganismos. En este trabajo, se tomaron datos de fragmentos terminales de restricción (T-RFLP) y librerías de clonas del gen 16S rRNA para describir la diversidad encontrada en comunidades acuáticas de procariontes y se probaron hipótesis históricas y ambientales en un esfuerzo para explicar la distribución de la diversidad. Se tomaron muestras de doce localidades en el valle de Cuatro Ciénegas. Los datos de librería de clonas muestran 69 filotipos diferentes dentro de las poblaciones numéricamente abundantes de los sitios muestreados. Dentro de los filotipos más abundantes se encuentran  $\gamma$  y  $\beta$  Proteobacteria, y halófilos extremos del dominio Archaea. Las diferencias entre sitios dadas por los datos de T-RFLP fueron significativas, 45 de los TRFs encontrados fueron localmente endémicos, y 38% del total de la diversidad está representada por diferencias entre sitios, sugiriendo alta diversidad  $\beta$  en el área de estudio. Los resultados indican que la similitud entre sitios dada por el análisis de comunidades no correlaciona con la distancia geográfica o algún parámetro ambiental medido, lo que sugiere un alto nivel de endemismos que podría ser debido a adaptaciones a condiciones locales.

## Diversity of aquatic prokaryotic communities in the Cuatro Ciénegas basin

Ana E. Escalante<sup>1</sup>, Luis E. Eguiarte<sup>1</sup>, Laura Espinosa-Asuar<sup>1</sup>, Larry J. Forney<sup>2</sup>, Ana M. Noguez<sup>1</sup> & Valeria Souza Saldivar<sup>1</sup>

<sup>1</sup>Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, México; and <sup>2</sup>Department of Biological Sciences, University of Idaho, Moscow, ID, USA

**Correspondence:** Valeria Souza Saldivar, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, México D.F. 04510, México. Tel.: +1 52 55 56229030; fax: +1 5255 5622 8995; e-mail: souza@servidor.unam.mx

Received 28 August 2007; revised 11 January 2008; accepted 9 March 2008.

DOI:10.1111/j.1574-6941.2008.00496.x

Editor: Patricia Sobczyk

### Keywords

diversity distribution; oligotrophic system; prokaryotic community; T-RFLP.

### Abstract

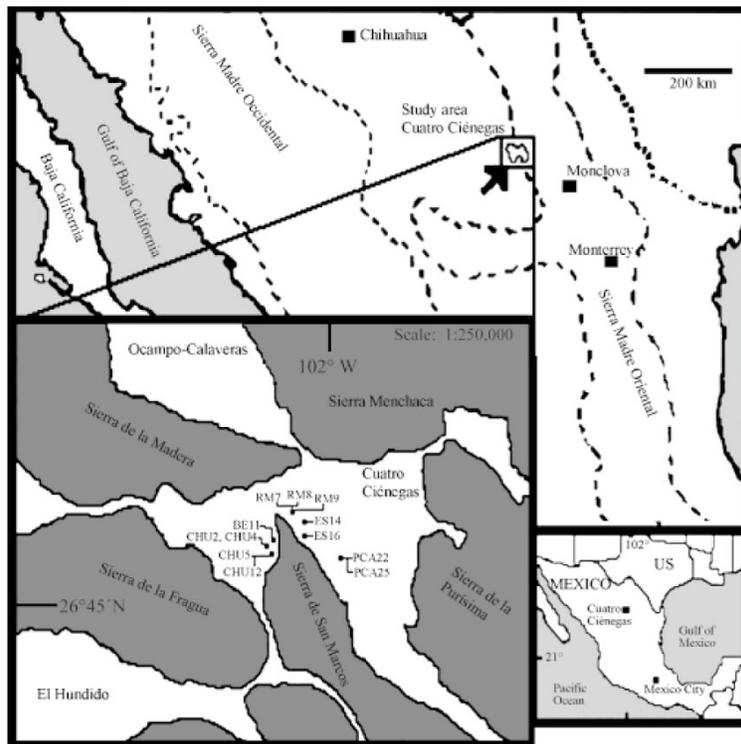
The Cuatro Ciénegas basin (Coahuila, México) is a composite of different water systems in the middle of the desert with unusually high levels of endemism and diversity in different taxa. Although the diversity of macrobiota has been well described, little is known about the diversity and distribution of microorganisms in the oligotrophic ponds. Here we describe the extent and distribution of diversity found in aquatic prokaryotic communities by analysis of terminal restriction fragment length polymorphisms (T-RFLP) of 16S rRNA genes and phylogenetic analysis of cloned genes. Twelve locations within the basin were sampled. Among all the samples, we found a total of 117 operational taxonomic units (OTUs) using T-RFLPs, which ranged in any single sample from four to 49. OTU richness and Shannon diversity indices for different sites varied, but none were particularly high. 16S rRNA gene sequence data showed 68 different phylotypes among 198 clones. The most abundant phylotypes were *Gamma*- and *Betaproteobacteria*, and extreme halophiles. The differences among sites were significant; 45 TRFs were found only once, and 37% of the total diversity was represented by differences between sites, suggesting high  $\beta$ -diversity. Further studies are needed to test whether this is a direct consequence of environmental heterogeneity in the basin.

### Introduction

The Cuatro Ciénegas basin in México has been described as an important biodiversity reservoir within the Chihuahuan desert. The area is a small (<840 km<sup>2</sup>) intermontane valley where different water systems are located. Most of the aquatic habitats are not permanent (Minckley, 1969), and the valley is divided into seven major and permanent drainage systems. Between locations, environmental conditions can vary in water chemistry, flow rate and the volume of spring discharge (Minckley, 1969, 1994; Dinger, 2001; Elser *et al.*, 2005). Moreover, most aquatic habitats in this area are extremely oligotrophic due to the almost negligible phosphorous levels (Elser *et al.*, 2005; F. García-Oliva, unpublished data). Despite this, the area is one of only two North American desert ecosystems characterized by high levels of species endemism hosting 70 endemic aquatic vertebrate and seven invertebrate species, including a

large number of snails and fish (Schmidt & Owens, 1944; Minckley, 1969; Cole, 1984; Hershler, 1985; Contreras-Balderas, 2004). These unusual levels of endemism have led individuals to compare Cuatro Ciénegas with the Galapagos Islands (Taylor & Minckley, 1966), and the unique biological characteristics of the area make it a priority for conservation efforts by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio), the World Wildlife Fund (WWF), the Ramsar Convention on Wetlands and Man and Biosphere (MAB)/UNESCO.

While previous studies have sought to describe and understand the biodiversity of plants, vertebrates and invertebrates within the Cuatro Ciénegas basin, comparatively little has been done to characterize the diversity of prokaryotes in the basin. Three studies on prokaryotic diversity in the aquatic habitats of Cuatro Ciénegas have been published recently (García-Pichel *et al.*, 2002; Elser *et al.*, 2005; Souza *et al.*, 2006). Two of these focused on the diversity found in



**Fig. 1.** Map of the study area. Cuatro Ciénegas basin in the state of Coahuila, México. Coordinates of sampling sites (CHU2, CHU4, CHU5, CHU12, BE11, RM7, RM8, RM9, ES14, ES16, PCA22, PCA25) are listed in Table 1. Modified from Johannesson *et al.* (2004).

the microbial mats and stromatolites of different sites within the study area: one described the population structure of *Cyanobacteria* (García-Pichel *et al.*, 2002) while the second examined the effect of nutrient enrichment on the biomass and diversity of the stromatolite community (Elser *et al.*, 2005). Souza *et al.* (2006) studied the community composition found in the water column at three sampling sites in the Cuatro Ciénegas basin and five in the surrounding areas. In this study, the authors referred to levels of microbial diversity but no detailed or statistical analysis of this was conducted. They mainly focused on the possible 'ancient' source of water at the site and the potential hydrological connectivity of Cuatro Ciénegas with the neighboring valleys placing special emphasis on the suggestion of a relic sea at Cuatro Ciénegas and the need to protect this area.

In this follow-up study, we broadened the Cuatro Ciénegas basin sampling and sought to explicitly assess not only the extent but the distribution of prokaryotic diversity among locations within the area, by obtaining estimates of  $\alpha$ - (local),  $\beta$ - (among sites) and  $\gamma$ - (regional) diversity at the site. The data presented here showed that diversity levels at a local scale were not particularly high, but differences in diversity among sites reflect a patchy distribution of diversity within the area.

## Materials and methods

### Study area and sampling

We sampled four of the seven major and permanent drainage systems of the Cuatro Ciénegas basin (Minckley, 1969): (1) Rio Churince (CHU), (2) La Becerra/Rio Garabatal/Juan Santos (BE), (3) Rio Mesquites (RM) (4) and Escobedo/Tio Candido/Los Hundidos/Los Gatos (ES, PCA) (Fig. 1). We analyzed a total of 12 samples from these sites (Table 1). All the sites sampled were rather small (not more than 100 m in diameter and no more than 3-m in depth), well-mixed pools because of water flow and wind. These features homogenize the water, eliminating spatial structuring within sampling sites (Reche *et al.*, 2005; Lindström, 2006). Moreover, preliminary surveys of replicate within-site samples using TRLP strategy (data not shown), revealed negligible within site variability in terms of community composition. However, environmental heterogeneity among sites has been reported previously (Dinger, 2001; Table 2); thus, we sought to assess patterns of diversity among pools. Each site was GPS referenced and the data were used to calculate the geographic distance between sampling locations. Over a 2-day period in the fall of 2002, samples of

**Table 1.** Chemical and biological features of sites sampled in Cuatro Ciénegas basin

Site	Latitude (N)	Longitude (W)	pH	Conductivity ( $\mu\text{S cm}^{-1}$ )	S	H ( $\pm$ SD)	J
Rio Churince							
CHU2	26°50.960'	102°08.750'	8.41	7500	36	2.70 (0.004)	0.75
CHU4	26°50.751'	102°08.336'	7.387	9000	38	3.08 (0.003)	0.85
<b>CHU5</b>	<b>26°50.421'</b>	<b>102°08.054'</b>	<b>7.46</b>	<b>2530</b>	<b>30</b>	<b>2.31 (0.004)</b>	<b>0.68</b>
CHU12	26°50.402'	102°08.028'	7.55	2030	14	2.21 (0.006)	0.84
Rio Mesquites							
<b>RM7*</b>	<b>26°55.215'</b>	<b>102°06.385'</b>	<b>8.15</b>	<b>3600</b>	<b>36</b>	<b>3.05 (0.003)</b>	<b>0.85</b>
<b>RM8</b>	<b>26°55.185'</b>	<b>102°06.341'</b>	<b>8.42</b>	<b>4500</b>	<b>16</b>	<b>2.54 (0.004)</b>	<b>0.91</b>
<b>RM9</b>	<b>26°55.175'</b>	<b>102°06.363'</b>	<b>7.01</b>	<b>6000</b>	<b>14</b>	<b>1.84 (0.006)</b>	<b>0.70</b>
La Becerra							
BE11	26°52.665'	102°08.262'	7.36	2112	4	1.31 (0.005)	0.94
Escobeda							
<b>ES14</b>	<b>26°52.296'</b>	<b>102°05.26'</b>	<b>7.9</b>	<b>2930</b>	<b>37</b>	<b>2.99 (0.006)</b>	<b>0.83</b>
<b>ES16</b>	<b>26°53.51'</b>	<b>102°05.178'</b>	<b>7.58</b>	<b>2040</b>	<b>31</b>	<b>2.49 (0.003)</b>	<b>0.72</b>
PCA22	26°52.081'	102°01.65'	7.62	3720	49	3.13 (0.003)	0.81
<b>PCA25</b>	<b>26°51.99'</b>	<b>102°01.94'</b>	<b>7.01</b>	<b>3720</b>	<b>15</b>	<b>2.27 (0.004)</b>	<b>0.84</b>

S, Richness; H, Shannon diversity index; J, equitability.

\*16S rRNA gene clone libraries were prepared from samples in bold font.

Geographic location of each site is presented as longitude and latitude in degrees. Physicochemical measurements correspond to pH and conductivity.

superficial water (< 12 cm depth) from three different locations within each site were collected in sterile 2 L Nalgene™ plastic bottles. Equal proportions of the samples were combined to make a single 2 L composite sample for each site. These were immediately filtered using 0.45  $\mu\text{m}$  filters and a field vacuum pump. Filters were placed into sterile BD Falcon™ vials (BD Biosciences, Bedford, MA) and immediately stored inside a liquid nitrogen container at  $-80^\circ\text{C}$ . After field collection, filters were maintained at  $-20^\circ\text{C}$  until they were used for the extraction of DNA.

### Physico-chemical measurements

The temperature, pH and conductivity (as a surrogate for salinity) of the water at each sampling site were measured using a YSI600QS Instrument equipped with a Hydrolab YSI600R sonde (YSI Inc., Yellow Springs, OH) to provide some insight into differences in the environmental conditions at the sites sampled.

### Analysis of terminal restriction fragment length polymorphisms (T-RFLP) of 16S rRNA genes

Genomic DNA was extracted from the filtered samples using a Ultra Clean Water DNA Kit (Mo Bio Lab Inc., Carlsbad, CA) within the first week after filters were stored at  $-20^\circ\text{C}$ . This DNA extraction procedure included a beadbeating step. DNA extraction efficiency was verified by total DNA electrophoresis in 1% agarose gels. The 16S rRNA genes in each sample were PCR amplified using the fluorescently labeled universal primers (Angert *et al.*, 1998) forward 515 VIC: 5'-GCG GAT CCT CTA GAC TGC AGT GCC AGC AGC CGC

GGT AA-3'; and reverse 1492 6FAM: 5'-GGC TCG AGC GGC CGC CCG GGT TAC CTT GTT ACG ACT T-3'. To obtain enough PCR product for further procedures, three independent 100  $\mu\text{L}$  PCR reactions were prepared from each sample as has been done previously (Dunbar *et al.*, 2001; Noguez *et al.*, 2005). Each reaction contained 1  $\times$  PCR buffer, 1.65 mM  $\text{MgCl}_2$ , 0.2 mM dNTP mixture, 0.6  $\mu\text{M}$  of each primer, 1 U *Taq* polymerase (Applied Biosystems, Foster City, CA) and 5% DMSO. All reactions were carried out in an MJ Research thermocycler (Watertown, MA) with the following program:  $94^\circ\text{C}$  for 4 min, followed by 35 cycles consisting of  $92^\circ\text{C}$  for 1.5 min,  $50^\circ\text{C}$  for 1.5 min,  $72^\circ\text{C}$  for 2 min and  $72^\circ\text{C}$  for 10 min. To reduce PCR bias due to preferential annealing of primers, we optimized the PCR conditions as suggested by Osborn *et al.* (2000). The same concentrations of DNA were used for all the reactions, and the number of cycles and annealing temperature were chosen to obtain high product yields without compromising PCR quality. The amplicons from the three replicate PCR reactions were combined, cleaned and concentrated (PCR Clean-up Kit, Millipore, Billerica, MA). The products from each PCR reaction were purified following electrophoresis in a 2% agarose gel (Gel extraction kit, Qiagen, Valencia, CA) to eliminate nonspecific products.

The T-RFLP of 16S rRNA genes in each sample were determined as follows: the amplicons were restricted using *AluI* for 3 h in a 20  $\mu\text{L}$  reactions that contained 10 U of *AluI* and c. 50 ng of the PCR product and were incubated in an MJ Research thermocycler for  $37^\circ\text{C}$  and  $65^\circ\text{C}$  for 30 min. The size and abundance of fluorescently labeled terminal restriction fragments (TRFs) were determined

**Table 2.** Water chemistry values for sites in the Cuatro Ciénegas basin area. Data are from Dinger (2001)

Site	Present study site equivalence	Temperature (°C)	pH	Dissolved oxygen (mg L <sup>-1</sup> )	Alkalinity (mg L <sup>-1</sup> )	Nutrient and ions mg L <sup>-1</sup> (±SE)							
						NO <sub>3</sub> <sup>-</sup>	PO <sub>4</sub> <sup>3-</sup>	SO <sub>4</sub> <sup>2-</sup>	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Na <sup>+</sup>	Cl <sup>-</sup>	
Becerra (BE)													
C.Becerra		34.7	8.4	11.0	167	2.24 (0.04)	NA	1267 (17.5)	343 (5.5)	118 (1.0)	153 (1.0)	105 (1.0)	
Garabatal		29.1	7.9	7.8	183	0.43 (0.03)	ND	2174 (243.4)	457 (3.5)	303 (7.5)	505 (5.0)	272 (23.7)	
P-Becerra	BE11	33.7	7.1	2.9	168	2.1 (0.19)	NA	1181 (0.5)	331 (9.5)	110 (4.0)	137 (1.0)	93 (0.5)	
Escobeda (ES/PCA)													
Hundidos	PCA22	25.5	7.9	5.5	189	ND	0.02 (0.0)	1871.2 (47.1)	364.5 (139.5)	162.5 (1.5)	241 (3.0)	139.6 (1.6)	
Los Gatos		30.3	8.5	8.0	162	ND	0.1 (0.0)	3695 (32.3)	425 (18.5)	710 (0.0)	1123 (37)	494 (5.6)	
P.Escobedo	ES14	34.2	7.2	2.3	184	1.83 (0.03)	ND	1290 (12.5)	320 (13.0)	108 (4.5)	129 (6.0)	99 (0.0)	
Río Mesquites (RM)													
Mojarra LW*		33.3	7.1	2.3	186	1.8 (0.5)	ND	971 (485.3)	326 (3.0)	99 (0.5)	150 (6.5)	8 (NA)	
Playitas*		25.3	8.5	6.9	113	ND	NA	3133 (212.5)	596 (4.0)	357 (5.0)	445 (2.0)	266 (19.0)	
Río Churince (CHU)													
P.Churince	CHU5	29.5	7.1	4.2	156	2.01 (0.01)	NA	1209 (6.0)	321 (16.5)	105 (5.0)	135 (3.0)	98 (0.5)	

\*These sampling sites are near to the present study RM-sampling sites.

NA, indicates not available.

ND, indicates values below detection levels.

Readings were collected in June and July 1999. Temperature, pH and dissolved oxygen were measured in the field with Hydrolab Minisonde water quality sonde. Alkalinity is expressed in terms of CaCO<sub>3</sub>. Alkalinity and ions were analyzed in the laboratory (*n* = 2).

using an ABI 3100-Avant Prism<sup>®</sup> Genetic Analyzer (Applied Biosystems) as described previously (Coolen *et al.*, 2005).

For each profile of TRFs, we established a baseline and only those TRFs with peak heights  $\geq 50$  fluorescent units (FU) were used in subsequent analyses (Dunbar *et al.*, 2001; Blackwood *et al.*, 2003). Each unique TRF was considered to be an operational taxonomic unit (OTU). Estimations of diversity were derived from matrices constructed based on the presence and abundance of TRFs using relative peak area as an estimate of abundance (Fierer & Jackson, 2006) calculated as:

$$Ap = (n_i/N) \times 100$$

in which  $n_i$  represents the peak area of one distinct TRF and  $N$  is the sum of all peak areas in a given T-RFLP pattern (Lukow *et al.*, 2000).

### Analysis of diversity

For all diversity analyses, each unique TRF was considered to be an OTU.

To describe the prokaryotic diversity at different sampling sites, we calculated Shannon diversity ( $H$ ) indices for each site using:

$$H = - \sum_{i=1}^S P_i \ln P_i$$

where  $P_i$  is the proportion of OTU  $i$ , such that  $\sum P_i = 1$ , and  $S$  is the total number of OTUs for each site (Magurran, 1988). Equitability indices ( $J$ ; Magurran, 1988) were also calculated per site using the formula:

$$J = \frac{H}{\ln S}$$

As a measure of how diversity was partitioned among sampling sites, we used the Shannon diversity indices with the following approach. Shannon diversity index ( $H$ ) or Shannon entropy of information has been described as a concave function; this implies that average local diversity ( $\bar{H}_{\text{within}}$ ) is always less than or equal to the regional or total diversity ( $H_T$ ) (Lande, 1996). For a concave measure like this, the total diversity or the entropy in a set of communities can be additively partitioned into nonnegative components within and among communities (Lande, 1996):

$$H_T = H_{\text{among}} + \bar{H}_{\text{within}}$$

where  $j$  = each local community that could be weighted based on its sample size or importance ( $q_j$ ), and  $\bar{H}_{\text{within}} = \sum_j q_j H_j$  or:  $H_T = H_\beta + \bar{H}_\alpha$ .

This additive model (Allan, 1975) has the advantage that the total diversity and its components can be compared directly (Lande, 1996). For the purposes of our work and as

a measure of how diversity is partitioned among sampling sites, the following value was calculated:

$$H_{\text{among}} = H_T - \bar{H}_{\text{within}},$$

where  $H_T$  is the total Shannon diversity of all sites pooled together, and  $\bar{H}_{\text{within}}$  is the average Shannon diversity per site (with the same weight for all sites). Hence, this value, if divided by  $H_T$ , could be taken as the proportion of the total diversity represented by differences among communities or sites and would be 0 if the composition and relative abundance of different taxa are identical among communities and approach 1 if there were no common OTU distribution among the sites.

Pairwise differences in the presence and abundance of OTUs at two sites were determined using an average-linkage clustering algorithm (uncentered and unweighted; Pielou, 1984). This analysis, which was carried out using CLUSTER 2.20 (Eisen *et al.*, 1998), compares the curves defined by the presence and abundance of OTUs at two sites, and calculates the degree of similarity of the two communities using the Pearson correlation coefficient.

Preliminary surveys of diversity within sites using the T-RFLP strategy revealed negligible within-site variability. Hence, we took the composite sample from each site as a good representation of the within-site diversity in order to further analyze diversity structure within the Cuatro Ciénegas basin as has been done in other similar studies (Reche *et al.*, 2005; Fierer & Jackson, 2006).

To identify sites that were similar in terms of community composition and structure, we performed principle component analysis (PCA) using STATISTICA (StatSoft Inc., Tulsa, OK) and the relative abundance of each TRF peak area. The relative position of each sample along the principle component axes can be used to describe the degree of community-level similarity between samples.

To test whether the community composition correlated with geographic location, we compared the Pearson similarity matrix with the geographic distance matrix between sampling sites using Mantel tests (Smouse *et al.*, 1986) and ARLEQUIN computer software (Schneider *et al.*, 2000). This is a test in which two dissimilarity matrices are compared to evaluate whether two variables are more closely associated than one would expect by chance (Smouse *et al.*, 1986). Finally, we performed linear regression analysis to evaluate the predictive power of the measured variable (latitude) on the levels of diversity observed at different sites (Fierer & Jackson, 2006).

### Clone library construction

Clone libraries were constructed from genomic DNA isolated from seven of the 12 sites within the Cuatro Ciénegas basin area. To avoid the formation of chimeric amplicons,

the two-step amplification protocol of Acinas *et al.* (2004) was used. The 16S rRNA genes in each sample were amplified in reactions that contained  $1 \times$  PCR buffer, 1.65 mM MgCl<sub>2</sub>, 0.2 mM dNTP mixture, 0.6  $\mu$ M of each primer, 1 U *Taq* polymerase (Applied Biosystems) and 5% DMSO. Unlabeled versions of the primers described above were used. All PCR reactions were carried out in an MJ Research thermocycler with the following program: 94 °C for 4 min, followed by 15 cycles consisting of 92 °C for 1.5 min, 50 °C for 1.5 min, 72 °C for 2 min, followed by 72 °C for 10 min. The PCR products from five replicate reaction mixtures from each sample were mixed and purified by electrophoresis in a 1% agarose gel (Gel extraction kit, Qiagen), followed by purification using a QIAquick Gel extraction kit (Qiagen Inc., the Netherlands). Each mixture of amplicons was reamplified by four cycles of PCR. The PCR products were cloned into the pCR2.1 vectors as per the manufacturer's instructions (Invitrogen Corp., San Diego, CA). The number of clones obtained from each sample was uneven but all clones obtained from each sample were picked and plasmids were extracted from each (30 clones average per library). Plasmid DNAs containing inserts were isolated for sequencing using the Plasmid Miniprep<sub>96</sub> kit (Millipore Corp.) and sequenced using the vector-based primers M13F and M13R using an ABI 3100-Avant Prism<sup>®</sup> Genetic Analyzer (Applied Biosystems).

### Phylogenetic analysis to describe community structure

All 16S rRNA gene sequences were exported to SEQUENCHER Version 4.1.4 (Gene Code, Ann Arbor, MI). The primer and vector sequences were eliminated and contigs were created. To identify possible chimeras, we conducted separate BLASTN searches (Altschul *et al.*, 1990) using the 5' and 3' regions of the genes, and compared the phylogenetic affiliations of the sequences, and those that were inconsistent were eliminated. All the remaining sequences were aligned using SEQUENCHER Version 4.1.4 (Gene Code) and required an overlap of 20 bp with a minimum sequence match of 75%. Sequences that were > 97% similar were defined as being the same OTU. A rarefaction curve (Gotelli & Colwell, 2001) of the abundance of these OTUs was constructed using ECOSIM (version 6.0) (Gotelli & Entsminger, 2001). To obtain an estimate of the actual number of OTUs that may be present in the sample, a nonparametric Chao1 richness estimator was calculated using ESTIMATES (Chao, 1984; Colwell, 2004).

To determine the phylogenetic relationships of partial 16S rRNA gene sequences, we performed Sequence Match at RDP (Cole *et al.*, 2007) and BLASTN searches and then constructed phylogenetic trees using maximum parsimony (MP) and neighbor-joining (NJ) (Kimura-2) using PAUP<sup>®</sup> 4.0 (Swofford, 2002).

### Nucleotide sequence accession numbers

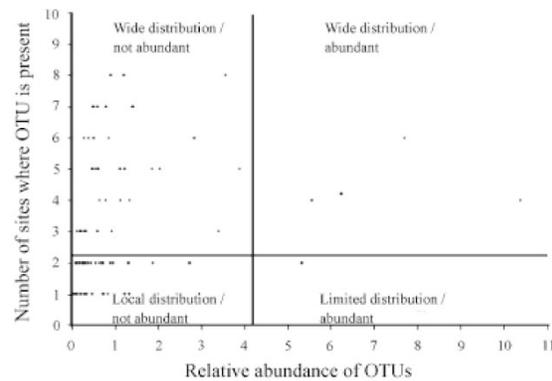
Unique sequences were submitted to GenBank and are available under accession numbers: EF011875, EF011877–EF011880, EF011882–EF011884, EF011887, EF011888, EF011891–EF011893, EF011896, EF011897, EF011899–EF0118903, EF011905–EF011909, EF011911–EF011917, EF011919, EF011922–EF011938, EF011940–EF011961, EF011963, EF011964 and EF011967–EF011971.

## Results

### Analysis of prokaryotic diversity

Because of the small size, mixing due to water flow and wind in the sampled pools (Reche *et al.*, 2005), as well as in light of previous surveys of within-site variability (data not shown), we chose to assess the prokaryotic diversity in 12 aquatic habitats within the Cuatro Ciénegas basin using one composite sample for each location. Profiles of T-RFLP of 16S rRNA genes were prepared from genomic DNA isolated from each sample, and each unique TRF was considered to be an OTU. Although TRFs of 16S rRNA genes do not have resolution for taxonomic identification of the populations in samples, the method can still be effectively used for comparisons between communities (Dunbar *et al.*, 1999). Moreover, the enzyme we chose (AluI) allows for the discrimination of OTUs without generating undesired products (Osborn *et al.*, 2000). We took special care to use the same amount of PCR product for digestion and to load the same amount of digested product from each sample for fragment analysis so that valid estimations of within- and among-site diversity could be made. As a measure of local diversity ( $\alpha$  diversity), the presence and abundance of each OTU in each T-RFLP profile was scored and this information was used to calculate OTU richness and Shannon diversity ( $H$ ) indices (Table 1). Among all the samples we found a total of 117 OTUs, and the number of OTUs in any single sample ranged from four to 49 OTUs. The OTU richness ( $S$ ) and Shannon diversity ( $H$ ) values for different sites varied, but none were particularly high. In general, equitability ( $J$ ) was high and very similar among sites, indicating that these communities were not dominated by a limited number of OTUs.

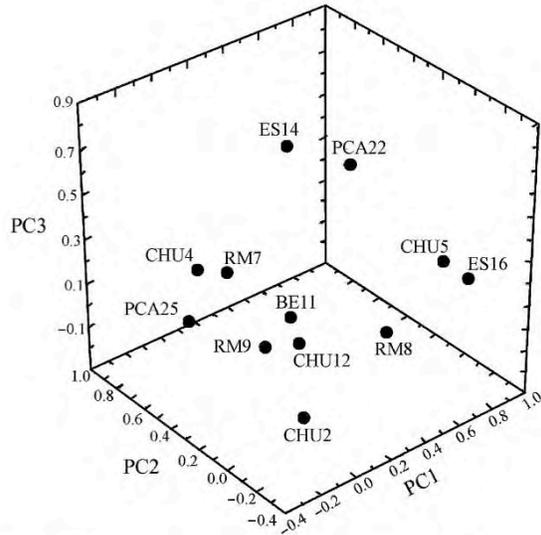
As a way to describe  $\beta$ -diversity within the area, the total diversity was partitioned (Lande, 1996) into its 'among-site' components. This showed that  $H_{\text{among}} = 1.49$ ; this means that almost 37% of the total diversity ( $H_T = 3.93$ ) of the area could be accounted for by differences in the diversity among sites and that the  $\beta$ -diversity of the area was high. In addition, we found that 38% of the OTUs (45 of 117) were unique to the site where they were found (Fig. 2) and appeared to be endemic. Most of these unique TRFs (90%) were not numerically dominant in the communities where



**Fig. 2.** Graphical representation of the spatial distribution and relative abundance of OTUs. Each dot represents a different OTU of the total sample (12 sites). The x-axis represents the relative abundance of each OTU in the total sample (fluorescence units of OTU/fluorescence of all OTUs from all samples) and the y-axis represents the number of sites where each OTU appeared. Hence, the dots or OTUs grouped in the (1) lower-left are considered as limited in their distribution and not abundant, (2) upper-left are considered as widely distributed and not abundant, (3) lower-right are considered as limited in their distribution but abundant and (4) upper-right are considered as widely distributed and not abundant.

they were found (Table 1, Fig. 2), and therefore may in large part account for the high  $\beta$ -diversity value. These results imply that diversity levels at each site may not be particularly high ( $\alpha$ -diversity) but that differentiation among sites is considerable ( $\beta$ -diversity).

We postulated that geographic distance might serve as a barrier to migration between sampling locations, and this might result in different communities becoming established in different locations. To test whether this were true, we first performed PCA of T-RFLP data to identify communities that were similar to one another. This analysis (Fig. 3) showed that none of the sites clustered with their nearest neighbor. Indeed, nearly the opposite was true. Sites CHU5 and ES16 are some distance from each other and separated by the Sierra San Marcos mountains, but nonetheless cluster together, as was the case for sites ES14, RM8, RM9 and PCA22, which are distant from CHU12 and BE11, and yet they clustered together. Thus, it appears that geographical proximity was not a good predictor of how related two communities might be in the Cuatro Ciénegas basin. This postulate was also tested by calculation of Pearson's correlation coefficients to assess the similarity among the communities. The data were used to construct matrices used for Mantel tests (Smouse *et al.*, 1986) that showed there was no statistically significant correlation between Pearson's correlation coefficients and geographic distance ( $r = 0.01$ ;  $P = 0.3$ ; with 1000 permutations). Finally, we performed linear regression analysis to evaluate the predictive power of

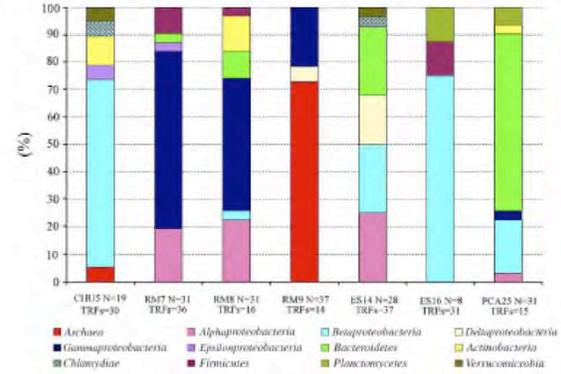


**Fig. 3.** PCA based on the relative abundance of T-RFLP data, showing clustering between superficial water samples of the Cuatro Ciénegas area. PC1, PC2 and PC3 explained 18%, 15% and 12% of the total variance, respectively.

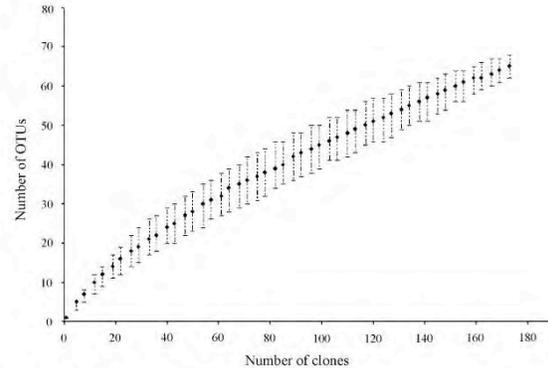
geographic location (latitude) on the levels of diversity observed at different sites. The linear regression test showed no significant correlation between diversity and geographic location (Shannon vs. latitude:  $r=0.08$ ). These results indicate that geographic location was not a significant determinant of community composition or the levels of diversity at the sites.

### Phylogeny of bacterial populations

To obtain more detailed information about the taxonomic diversity of the study area, we analyzed the sequences of 16S rRNA genes in clone libraries prepared from seven of the 12 sites within the Cuatro Ciénegas basin (CHU5, RM7, RM8, RM9, ES14, ES16, PCA25). The phylogenetic relationships of partial 16S rRNA gene sequences were determined using the Sequence Match algorithm of the RDP (Cole *et al.*, 2007) and BLASTN searches (Altschul *et al.*, 1990), and then constructed MP and NJ trees; both analyses gave approximately the same results (NJ trees as supplementary Fig. S1). A total of 198 clones from the clone libraries were analyzed; the number of clones obtained per site varied from eight to 37 (Fig. 4). Thirteen of these sequences were identified as possible chimeras and discarded for subsequent analysis (an average of 30 clones were obtained per site). The results from this analysis showed that sequences were dominated by *Proteobacteria* (107 clones/57.8%) including ones from the *Gamma* (44/23.8%), *Beta* (33/17.8%), *Alpha* (21/11.6%), *Delta* (7/3.8%) and *Epsilon* (2/1.1%) subdivisions. These



**Fig. 4.** Relative abundances of phylogenetic groups obtained from clone libraries data at different sites in the Cuatro Ciénegas basin. Each bar refers to a different site and the number of clones obtained from each site is indicated below each one as N.



**Fig. 5.** Rarefaction curve from analysis of the 16S rRNA gene clone libraries (185 clones).

were followed in abundance by clones of *Bacteroidetes* (31/16.8%) and Archaea of the Euryarcheota group (28/15.1%). Other groups with fewer sequences included *Actinobacteria* (7/23.8%), *Firmicutes* (5/2.7%), *Planctomycetes* (3/1.6%), *Chlamydiae* (2/1.1%) and *Verrucomicrobia* (2/1.1). In accordance with the T-RFLP data, the phylogenetic analyses showed that the phylogenetic identity of numerically abundant populations differed markedly from site to site. For example, site RM9 was dominated by Archaea, while sites ES16 and PCA 25 were dominated by *Betaproteobacteria* and *Bacteroidetes*, respectively (Fig. 4).

To evaluate our sampling efforts of clone libraries, and to have an idea of the total prokaryotic diversity within the Cuatro Ciénegas area ( $\gamma$ -diversity), a rarefaction curve was constructed based on the rank-abundance of OTUs at the site, where an OTU was defined by clones that shared > 97% sequence similarity (Fig. 5). In total, there were 68

## 8

OTUs retrieved from the Cuatro Ciénegas basin. The rarefaction curve generated with this information did not reach an asymptote, indicating that the phylotype richness within the study area had not been completely sampled. Consistent with this, the Chao1 nonparametric richness estimator also indicated that our sampling effort was far short of the number of OTUs predicted, which could be at least 180 [richness = 180.5, confidence interval (CI) = 15.9–332.17]. These results imply that a better estimation of total diversity within the area requires a more intensive sampling of clone libraries.

## Discussion

Scaling of diversity has been studied for macroorganisms (Gaston & Blackburn, 2000) and has contributed to ecological interpretations of spatial patterns of diversity distribution (Brown, 1995). Although such measures are also important to evaluate patterns of prokaryotic diversity, studies that do so are uncommon (Hughes *et al.*, 2006). Concepts such as  $\alpha$ ,  $\beta$ , and  $\gamma$  are not frequently invoked and few studies have been carried out to evaluate diversity at different spatial scales (Green *et al.*, 2004; Noguez *et al.*, 2005). Whittaker *et al.* (2001) discuss  $\alpha$ -,  $\beta$ - and  $\gamma$ -diversity in terms of the spatial scale of analysis, where  $\gamma$  refers to the regional spatial scale of analysis,  $\alpha$ -diversity refers to the local scale of analysis and  $\beta$ -diversity refers to the turnover between patches at the  $\alpha$ -scale. Given that different taxa may differ by many orders of magnitude in body size, the actual spatial scale of analysis may be chosen by convenience relating to the scale at which species have been mapped (Linder, 1991; Whittaker *et al.*, 2001).

For this work the levels of  $\alpha$ -diversity estimated from the T-RFLP data did not indicate the presence of high levels of OTU diversity within the sites sampled. There are at least three explanations for this observation. First, OTUs may underestimate diversity because each OTU may include multiple phylogenetic lineages. This is because multiple phylotypes can have TRFs with the same size (Marsh *et al.*, 2000). Second, only the numerically dominant populations are reflected in T-RFLP profiles and less abundant and rare organisms remain obscured (Liu *et al.*, 1997). Third, not all organisms are lysed with equal efficiency and thus the genomic DNA of some populations may not be recovered (Picard *et al.*, 1992). Moreover, there is known bias of PCR primers and some templates are not amplified at all (Liesack *et al.*, 1991) or not amplified as efficiently as others (Reysenbach *et al.*, 1992). It is unclear which of these factors, either alone or in combination, account for the apparently low numbers of OTUs observed, but it is likely that the true  $\alpha$ -diversity may be underestimated and not known with certainty. Nonetheless, as has been published in other studies using finger print techniques such as DGGE

or T-RFLP (Noguez *et al.*, 2005; Reche *et al.*, 2005; Fierer & Jackson, 2006), the data can be used both to estimate local diversity and to quantify differences in the kind and relative abundances of numerically abundant populations in the communities sampled and thereby gauge the patchiness of diversity and obtain an estimate of 'species turnover' or  $\beta$ -diversity (Noguez *et al.*, 2005). PCA analysis of T-RFLP data showed that none of the sites clustered with their nearest neighbor. Moreover, results from the partitioning of variance of total diversity into its among site components showed that almost 37% of the overall diversity of the area could be accounted for by differences among sampling sites. It should be noted that this might be a consequence of using unreplicated samples, which would cause the average diversity within communities to appear as diversity among communities. However, this seems unlikely to be the case because (1) other studies in freshwater habitats have shown that most of the diversity found is represented by the among-site component (Casamayor *et al.*, 2000; Yannarell & Triplett, 2004), (2) as we mentioned before, sampled pools are of a small size and well mixed because of wind and water flow; these features make it difficult to find evidence of significant spatial structuring within sampling sites (Reche *et al.*, 2005; Lindström, 2006). As a consequence of this, we think that it will also be difficult to find significant variation in the within-site diversity component contributing to the observed variation among sites, and (3) our previous surveys of replicate within-site samples have shown negligible within-site variation. Given this, we find that the diversity represented by the among-site variation is an interesting result when compared with other published studies in soils using T-RFLPs for diversity estimations; these studies show a similar pattern of few shared OTUs between sampling sites, making  $\beta$ -diversity high at different spatial scales (Noguez *et al.*, 2005; Wawrik *et al.*, 2007). Similarly, our data show a pattern of patchy OTU distribution, which is evidence that the  $\beta$ -diversity of prokaryotes in the Cuatro Ciénegas basin is high.

The apparent patchiness and high  $\beta$ -diversity observed in this study may reflect heterogeneity among the aquatic habitats sampled (Minckley, 1969; Dinger, 2001). The Cuatro Ciénegas area is reported to have aquatic habitats that belong to distinct drainage systems that differ in terms of their water chemistry and physicochemical parameters. These differences can be dramatic for different nutrients and ions as shown in a previous study by Dinger (2001), who reported point measurements of various environmental parameters. For example,  $\text{NO}_3^-$  was shown to range from nondetectable levels in the Escobeda system to  $2.24 \text{ mg L}^{-1}$  in the Becerra system, while  $\text{SO}_4^{2-}$  ranges from 971 in the Rio to  $3695 \text{ mg L}^{-1}$  in the Escobeda system (see Table 2 for more details).

Given this heterogeneity, it is not surprising that we did not find any simple correlation between community composition and geographic location. While differences in water chemistry and physicochemical parameters may be factors influencing the patchiness of diversity, we cannot exclude the possibility that it arises because the pool of existing species is diverse and community assembly is stochastic or that the lumping of different microbial lineages by T-RFLP may obscure the true spatial patterns of microbial diversity. More information about temporal changes in community composition and environmental conditions is needed to understand what factors account for the differences in community composition reported here.

Analysis of 16S rRNA gene sequences from clone library of different sites in the Cuatro Ciénegas basin show different compositions of dominant taxa in each site and, as a consequence, high levels of  $\gamma$ -diversity overall. This is interesting because it correlates very well with the previously described heterogeneity between drainage systems within the area.  $\gamma$ -Diversity values in the form of Chao1 diversity estimate are comparable to those found in an Alaskan soil in a study with a similar experimental design (Schloss & Handelsman, 2006). This suggests that in the Cuatro Ciénegas drainage systems sampled, resource heterogeneity, as well as seasonal (Carson & Dowling, 2005) and daily fluctuations, may result in overall high levels of diversity being maintained. However, given the low number of clones and the rarefaction curve results, we cannot make sweeping claims regarding  $\gamma$ -diversity at the site until more intensive sampling is performed.

The diversity levels at any scale of analysis could also be discussed in terms of the trophic status of the area studied, because one common feature of the sampling sites is the low availability of phosphorus in the aquatic habitats (Dinger, 2001; Elser *et al.*, 2005). It has been suggested previously that nutrient-enriched lakes have reduced species richness and increased evenness among key groups (Pearce *et al.*, 2005). For our case, it is unclear whether this relationship is generally true because a previous study with stromatolites within the area did not show a marked decrease in biomass or richness after phosphorous enrichment treatment (Elser *et al.*, 2005). Further studies addressing this question in the area are needed to evaluate the influence of trophic status on the diversity levels of prokaryotes in various kinds of aquatic habitats.

Phylogenetic analysis of the clone libraries showed an interesting result. There were dramatic differences in the community composition with regard to the numerically dominant populations present at various locations. This could be seen as evidence of local endemisms within the sampled area. Additional sampling could reveal that the phylotypes found at only one site may be present in other sites but in lower numbers. Nonetheless, because both clone

libraries and T-RFLP analysis show dramatic differences in the community composition between sites, it appears that bacterial populations at Cuatro Ciénegas may have a patchy distribution.

In conclusion, the data from this study show that different organisms may exist in different locations within the area studied, each community was rather unique and  $\beta$ -diversity was high. The analysis of cloned 16S rRNA gene sequences suggests that unique phylotypes were common and the diversity among sites was high. No clear geographical patterns in community composition were found, perhaps as a result of the complex hydrology in the area and a limited knowledge of the environmental conditions in the aquatic habitats studied. In future work, an effort should be made to sample more extensively the organisms present in the communities studied and characterize the physicochemical properties of the environments more fully.

## Acknowledgements

This research was supported by NASA to V.S. (NCCS2-1051), CONACyT-UNAM scholarship to A.E.E. and A.M.N. and SEMARNAT/CONACyT to V.S. (2002-CO1-0246). We thank S. Acinas, S. Bent, M. Wong, M. Schneider, I. Hernández and L. Falcón for technical support. We also thank La Venta Exploring Team, J. Gasca and A. Valera for the help during the sample collection.

## References

- Acinas SG, Klepac-Ceraj V, Hunt DE, Phaino C, Ceraj E, Distel DL & Polz MF (2004) Fine-scale phylogenetic analysis architecture of a complex bacterial community. *Nature* **430**: 551–554.
- Allan JD (1975) Components of diversity. *Oecologia* **18**: 359–367.
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.
- Angert ER, Northup DE, Reysenbach AL, Peek AS, Goebel BM & Pace NR (1998) Molecular phylogenetic analysis of a bacterial community in Sulphur River, Parker Cave, Kentucky. *Am Mineral* **83**: 1583–1592.
- Blackwood CB, Marsh T, Kim S-H & Paul EA (2003) Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. *Appl Environ Microbiol* **69**: 926–932.
- Brown JH (1995) *Macroecology*. Chicago University Press, Chicago, IL.
- Carson EW & Dowling TE (2005) Influence of hydrogeographic history and hybridization on the distribution of genetic variation in the pupfishes *Cyprinodon atrorus* and *C. bifasciatus*. *Mol Ecol* **15**: 667–679.
- Casamayor EO, Schafer H, Baneras I, Pedrós-Alió C & Muyzer G (2000) Identification of spatio temporal differences between microbial assemblages from two neighboring sulfurous lakes:

- comparison by microscopy and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **66**: 499–508.
- Chao A (1984) Non parametric estimation of the number of classes in a population. *Scan J St* **11**: 265–270.
- Cole GA (1984) Crustacea from the bolson of Cuatro Ciénegas, Coahuila, Mexico. *J Arizona-Nevada Aca Sci* **19**: 3–12.
- Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM & Tiedje JM (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acid Res* **35**: D169–D172.
- Colwell RK (2004) EstimateS: statistical estimation of species richness and shared species from samples version 7. Available at: <http://purl.oclc.org/estimates>.
- Contreras-Balderas S (2004) Ecosystems and water species in Cuatro Ciénegas. *Under the Desert: The Mysterious Waters of Cuatro Ciénegas* (Badino G, Bernabei T, Vito AD, Giulivo I & Savino G, eds), pp. 98–105. La Venta-Esplorazioni Geografiche, Instituto Coahuilense de Ecología, Edizioni Tintoretto (TV), Italy.
- Coolen MJL, Post E, Davis CC & Forney LJ (2005) Characterization of microbial communities found in the human vagina by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Appl Environ Microbiol* **71**: 8729–8737.
- Dinger E (2001) Aquatic Invertebrates of Cuatro Ciénegas, Coahuila, México and Effects of Fish on Stromatolite Invertebrate Assemblages. MS Thesis, Northern Arizona University, USA.
- Dunbar J, Takala S, Barns SM, Davis JA & Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol* **65**: 1662–1669.
- Dunbar J, Ticknor LO & Kuske CR (2001) Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. *Appl Environ Microbiol* **67**: 190–197.
- Eisen MB, Spellman PT, Brown PO & Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 1463–1468.
- Elser JJ, Schampel JH, García-Pichel F, Wade BD, Souza V, Eguiarte L, Escalante A & Farmer JD (2005) Effects of phosphorous enrichment and grazing snails on modern stromatolitic microbial communities. *Freshw Biol* **50**: 1808–1825.
- Fierer N & Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.
- García-Pichel F, Wade BD & Farmer JD (2002) Jet-suspended, calcite-ballasted cyanobacterial waterwarts in a desert spring. *J Phycol* **38**: 420–428.
- Gaston KJ & Blackburn TM (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.
- Gotelli NJ & Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* **4**: 379–391.
- Gotelli NJ & Entsminger GL (2001) *Ecosim: Null Models Software for Ecology Version 6.0*. Acquired Intelligence Inc., Kesey-Bear.
- Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M & Beattie AJ (2004) Spatial scaling and eukaryote diversity. *Nature* **432**: 747–750.
- Hershler R (1985) Systematic revision of the *Hydrobiidae* (*Gastropoda: Rissoaceae*) of the Cuatro Ciénegas basin, Coahuila, México. *Malacologia* **2**: 31–123.
- Hughes JBM, Bohannon BJM, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature Rev Microbiol* **4**: 102–112.
- Johannesson KH, Cortés A & Kilroy KC (2004) Reconnaissance isotopic and hydrochemical study of Cuatro Ciénegas groundwater, Coahuila, México. *South Am Earth Sci* **17**: 171–180.
- Lande R (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* **76**: 5–13.
- Liesack W, Weyland H & Stackebrandt E (1991) Potential risks of gene amplification by PCR determined by 16S rDNA analysis of mixed-culture of strict barophilic bacteria. *Microb Ecol* **21**: 191–198.
- Linder HP (1991) Environmental correlates of patterns of species richness in the south-western Cape Province of South Africa. *J Biogeogr* **18**: 509–518.
- Lindström E (2006) External control of bacterial community structure in lakes. *Limnol Oceanogr* **51**: 339–324.
- Liu W-T, Marsh TL, Cheng H & Forney L (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516–4522.
- Lukow T, Dunfield PF & Liesack W (2000) Use of T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. *FEMS Microbiol Ecol* **32**: 241–247.
- Magurran AE (1988) *Ecological Diversity and its Measurement*. Princeton University Press, Princeton, NJ.
- Marsh TL, Saxman P, Cole J & Tiedje J (2000) Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Appl Environ Microbiol* **66**: 3616–3620.
- Minckley WL (1969) Environments of the Bolson of Cuatrociénegas, Coahuila, Mexico, with special reference to the aquatic biota. *Univ Texas El Paso, Sci Series* **2**: 1–65.
- Minckley WL (1994) A bibliography for natural history of the Cuatro Ciénegas basin and environs, Coahuila, Mexico. *Proc Des Fishes Council* **25**: 47–64.
- Noguez AM, Arita HT, Escalante AE, Forney LJ, Garcia-Oliva F & Souza V (2005) Microbial macroecology: a highly structured prokaryotic soil assemblages is a tropical deciduous forest. *Global Ecol Biogeogr* **14**: 241–248.

- Osborn AM, Moore ERB & Timmis KN (2000) An evaluation of terminal restriction fragment length polymorphism (TRFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* **2**: 39–50.
- Pearce DA, van der Gast CJ, Woodward K & Newsham KK (2005) Significant changes in the bacterioplankton community structure of a maritime Antarctic freshwater lake following nutrient enrichment. *Microbiology* **151**: 3237–3248.
- Picard C, Ponsonnet C, Paget E, Nesme X & Simonet P (1992) Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Appl Environ Microbiol* **58**: 2717–2722.
- Pielou EC (1984) *The Interpretation of Ecological Data: A Primer of Classification and Ordination*. John Wiley & Sons, USA.
- Reche I, Pulido-Villena E, Morales-Baquero R & Casamayor EO (2005) Does ecosystem size determine aquatic bacterial richness? *Ecology* **86**: 1715–1722.
- Reysenbach A-L, Giver LJ, Wickham GS & Pace NR (1992) Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol* **58**: 3417–3418.
- Schloss PD & Handelsman J (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* **2**: 786–793.
- Schmidt KP & Owens DW (1944) Amphibians and reptiles of northern Coahuila Mexico. *Field Col Mus Nat, Zool Ser* **29**: 97–115.
- Schneider S, Roessli D & Excoffier L (2000) *Arlequin Ver. 2.00: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Smouse PE, Long JC & Sokal R (1986) Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Syst Zool* **35**: 627–632.
- Souza V, Espinosa-Asuar L, Escalante AE *et al.* (2006) An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc Natl Acad Sci USA* **103**: 6565–6570.
- Swofford DL (2002) *PAUP\*: Phylogenetic Analysis Using Parsimony (and other Methods) 4.0 Beta*. Sinauer Associates, Sunderland, MA.
- Taylor DW & Minckley WL (1966) New world for biologists. *Pac Discov* **19**: 18–22.
- Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ & Kerkhof L (2007) Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in new jersey and central Asia. *Appl Environ Microbiol* **73**: 2982–2989.
- Whittaker RJ, Willis KJ & Field R (2001) Scale and species richness: towards a general hierarchical theory of species diversity. *J Biogeogr* **28**: 453–470.
- Yannarell AC & Triplett EW (2004) Within- and between-lake variability in the composition of bacterioplankton communities: investigations using multiple spatial scales. *Appl Environ Microbiol* **70**: 214–233.

### Supplementary material

The following supplementary material is available for this article:

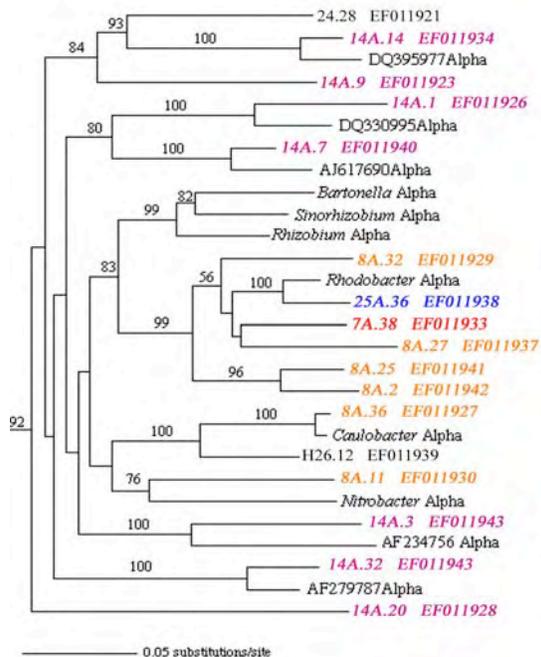
**Fig. S1.** Neighbor-Joining tree (Kimura-2) of 16SrRNA sequences obtained from clone libraries constructed from Cuatro Ciénegas water samples.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1574-6941.2008.00496.x> (This link will take you to the article abstract).

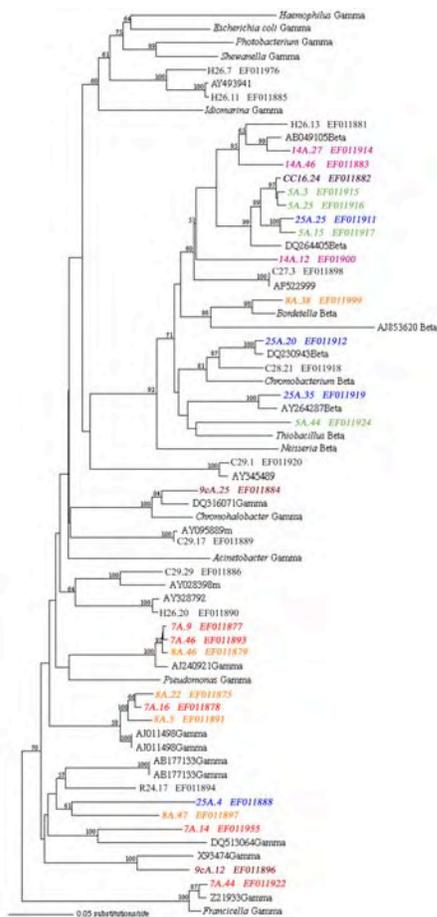
Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Figure S1. Neighbor-Joining tree (Kimura-2) of 16SrRNA sequences obtained from clone libraries constructed from Cuatro Ciénegas water samples. In bold are the sequences obtained from this study, the rest are the most closely related sequences from the GeneBank. Sequences sharing >97% of identity were considered as the same OTU. Numbers indicate bootstrap percentages from 1000 bootstrap replicates. Only bootstraps >50 are shown. The tree is shown in parts that correspond to different phylogenetic groups: a)  $\alpha$ -proteobacteria, b)  $\beta$ - and  $\gamma$ -proteobacteria, c)  $\delta$ -proteobacteria, d)  $\epsilon$ -proteobacteria, e) Bacteroidetes, f) Verrucomicrobia, Planctomycetes and Chlamydiae, g) Actinobacteria and Firmicutes, h) Archaea. The sites from which the clones were obtained are color coded: \*CHU5, \*RM7, \*RM8, \*RM9, \*ES14, \*ES16 and \*PCA25

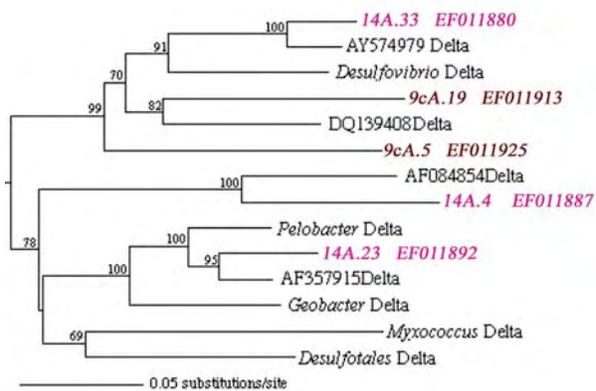
a)



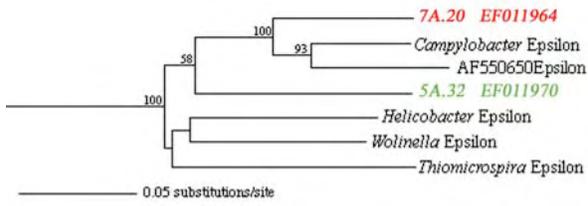
b)



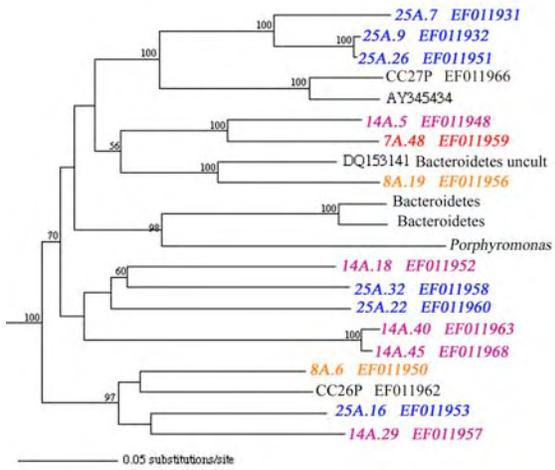
c)



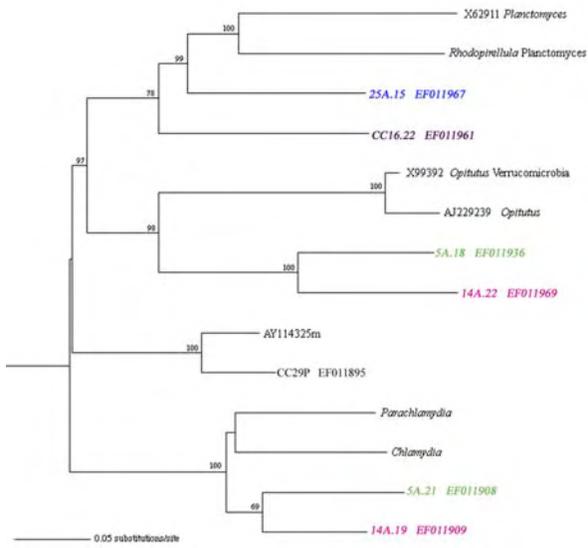
d)



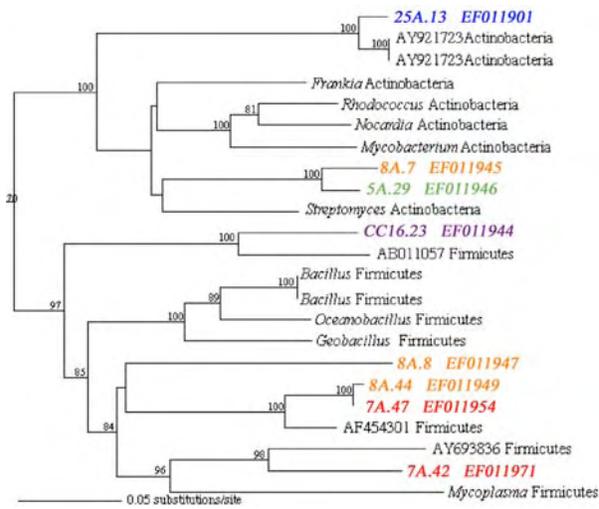
e)



f)



g)



h)

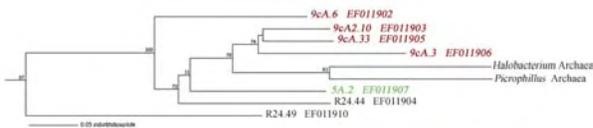


Tabla 3.1. Relación de secuencias del gen 16S rRNA obtenidas de clonas ambientales. Se incluyen los mejores “hits” de RDPII y, en ciertos casos se incluye la información obtenida de NCBI (donde la filogenia no coincidía con la identidad reportada por RDPII). Se especifica el OTU al que corresponde cada clona/secuencia. Continúa...

OTU	#clonas	Clonas representantes	GenBank ID	RDPII “best hit”			NCBI “best hit”
1	1	5A2	EF011907	Bacteria[100%]	Aquificae[37%]	Aquificae[37%]	Archaea
2	1	9cA33	EF011905	Bacteria[100%]	Aquificae[41%]	Aquificae[41%]	Archaea
3	15	9cA2	EF011903	Bacteria[100%]	Aquificae[41%]	Aquificae[41%]	Archaea
4	1	9cA6	EF011902	Bacteria[100%]	Aquificae[41%]	Aquificae[41%]	Archaea
5	1	7A20	EF011964	Bacteria[100%]	Proteobacteria[100%]	Epsilonproteobacteria[100%]	Archaea
6	2	14A29	EF011957	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[99%]	
7	1	25A16	EF011953	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[95%]	
8	1	8A6	EF011950	Bacteria[100%]	Bacteroidetes[99%]	Sphingobacteria[98%]	
9	2	14A40, 14A45	EF011887	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[100%]	
10	11	25A22	EF011960	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[100%]	
11	1	25A32	EF011958	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[86%]	
12	2	14A18	EF011952	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[54%]	
13	2	8A19	EF011956	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[69%]	
14	1	7A48	EF011959	Bacteria[100%]	Bacteroidetes[100%]	Bacteroidetes[92%]	
15	1	14A5	EF011948	Bacteria[100%]	Bacteroidetes[100%]	Sphingobacteria[79%]	
16	6	25A9, 25A26	EF011932	Bacteria[100%]	Bacteroidetes[100%]	Flavobacteria[100%]	
17	1	25A7	EF011931	Bacteria[100%]	Bacteroidetes[100%]	Flavobacteria[100%]	
18	1	14A19	EF011909	Bacteria[100%]	Chlamydiae[88%]	Chlamydiae[88%]	
19	1	5A21	EF011908	Bacteria[100%]	Chlamydiae[100%]	Chlamydiae[100%]	
20	3	14A23	EF011892	Bacteria[100%]	Proteobacteria[100%]	Deltaproteobacteria[100%]	
21	1	14A4	EF011887	Bacteria[100%]	Proteobacteria[69%]	Alphaproteobacteria[40%]	
22	1	9cA19	EF011913	Bacteria[100%]	Proteobacteria[94%]	Deltaproteobacteria[92%]	
23	1	14A33	EF011880	Bacteria[100%]	Proteobacteria[99%]	Deltaproteobacteria[99%]	
24	1	14A20	EF011928	Bacteria[100%]	Proteobacteria[88%]	Alphaproteobacteria[80%]	
25	1	14A7	EF011940	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
26	1	14A1	EF011926	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
27	1	14A32	EF011943	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[85%]	
28	1	14A3	EF011943	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[99%]	
29	1	8A11	EF011930	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
30	1	8A36	EF011927	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
31	1	8A2	EF011942	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
32	1	8A25	EF011930	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
33	2	8A27	EF011937	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
34	6	7A38	EF011933	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
35	1	25A36	EF011938	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
36	1	8A32	EF011929	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[100%]	
37	1	14A9	EF011923	Bacteria[100%]	Proteobacteria[79%]	Alphaproteobacteria[49%]	
38	1	14A14	EF011921	Bacteria[100%]	Proteobacteria[100%]	Alphaproteobacteria[98%]	
39	1	7A44	EF011922	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	

Tabla 3.1. Continuación.

OTU	#clonas	Clonas representantes	GenBank ID	RDPII "best hit"[Similitud]			NCBI "best hit"
40	6	9cA12	EF011896	Bacteria[100%]	Proteobacteria[98%]	Gammaproteobacteria[97%]	
41	1	7A14	EF011955	Bacteria[100%]	Proteobacteria[99%]	Gammaproteobacteria[98%]	
42	1	8A47	EF011897	Bacteria[100%]	Proteobacteria[99%]	Gammaproteobacteria[74%]	
43	1	25A4	EF011888	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	
44	2	9cA25	EF011884	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	
45	11	7A9, 7A46, 8A46	EF011877	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	
46	21	7A16, 8A22, 8A5	EF011875	Bacteria[100%]	Proteobacteria[100%]	Gammaproteobacteria[100%]	
47	1	25A35	EF011919	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
48	1	25A20	EF011912	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
49	1	5A44	EF011924	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
50	1	8A38	EF011999	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
51	4	14A12	EF011900	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
52	5	5A15, 25A25, 5A25, 5A3, CC16.24	EF011911	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
53	17		EF011882	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
54	1	14A46	EF011883	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
55	2	14A27	EF011914	Bacteria[100%]	Proteobacteria[100%]	Betaproteobacteria[100%]	
56	2	5A29	EF011946	Bacteria[100%]	Actinobacteria[100%]	Actinobacteria[100%]	
57	4	8A7	EF011945	Bacteria[100%]	Actinobacteria[100%]	Actinobacteria[100%]	
58	1	25A13	EF011901	Bacteria[100%]	Actinobacteria[98%]	Actinobacteria[98%]	
59	2	7A42	EF011971	Bacteria[100%]	Firmicutes[100%]	Mollicutes[99%]	
60	1	7A47	EF011954	Bacteria[100%]	Firmicutes[72%]	Clostridia[33%]	
61	1	CC16.23	EF011944	Bacteria[100%]	Firmicutes[72%]	Clostridia[33%]	
62	1	14A22	EF011969	Bacteria[100%]	Verrucomicrobia[95%]	Verrucomicrobiae[95%]	
63	1	5A18	EF011936	Bacteria[100%]	Verrucomicrobia[58%]	Verrucomicrobiae[58%]	
64	1	CC16.22	EF011961	Bacteria[100%]	Verrucomicrobia[58%]	Verrucomicrobiae[58%]	
65	2	25A15	EF011967	Bacteria[100%]	Planctomycetes[58%]		
66	1	9cA5	EF011925	Bacteria[100%]	Proteobacteria[100%]	Deltaproteobacteria[100%]	
67	1	8A8	EF011947	Bacteria[100%]	Firmicutes[85%]	Bacilli[42%]	
68	10	9cA3	EF011906	Bacteria[100%]	Aquificae[34%]	Aquificae[34%]	Haloarchaea

#### **Capítulo 4. DIVERSIDAD DE LINAJES AMBIENTALES DE *Pseudomonas sp.***

Después de la descripción y discusión de la diversidad de especies en comunidades procariontes del valle de Cuatro Ciénegas parecerá un poco brusco el salto que a continuación dará esta tesis pues se enfocará fundamentalmente en la diversidad de genes dentro de poblaciones o linajes específicos de un grupo de bacterias: *Pseudomonas*.

No cabe ninguna duda de que el concepto de diversidad puede resultar paradójico ya que, aunque es y será una piedra angular en ecología, existen muchos significados, usos y medidas para ella (Cousins 1991). En el contexto de la biología de la conservación, ha surgido el concepto de biodiversidad que es incluyente y sugiere la consideración de aproximaciones a la diversidad desde diferentes disciplinas de la biología (Eguiarte *et al.* 1999). En este sentido y para el caso de la aproximación al estudio de la biodiversidad en esta tesis, la definición de biodiversidad propuesta por di Castri y Younes (1996) engloba de manera muy adecuada la relación que guardan las escalas de estudio y la biodiversidad, siendo esta: "...el ensamble y las interacciones jerárquicas, a diferentes niveles de integración, de las escalas de organización genética, taxonómica y ecológica". Y de hecho, como sugieren Eguiarte *et al.* (1999), "...la diversidad filogenética es un concepto que recorre con solidez diferentes escalas jerárquicas en diferentes niveles de integración" desde poblaciones, pasando por especies, familias y géneros hasta llegar a grandes grupos como reinos; aunque parece que, conforme se avanza en la jerarquía, deben ser consideradas las propiedades emergentes de cada nivel.

El razonamiento planteado en las líneas anteriores justifica el hecho de que para la misma área de estudio sea relevante explorar la diversidad biológica a dos niveles

jerárquicos: diversidad entre especies (comunidades) y diversidad dentro de linajes o dentro de especies, como es el caso del estudio de la genética de poblaciones.

#### *4.1. Introducción general sobre genética de poblaciones*

Hoy en día estamos familiarizados con la muy citada frase de T. Dobzhansky: “*Nada tiene sentido en biología si no es a la luz de la evolución*”. Sin embargo, no somos del todo conscientes de que esto se vuelve posible, en parte, gracias al encuentro de la teoría evolutiva propuesta por Darwin-Wallace (1859) y las leyes de Mendel (1865) favorecido por R.A. Fisher, J.B.S. Haldane y S. Wright y resultando en el nacimiento de una disciplina denominada genética de poblaciones entre 1920 y 1930. A partir de entonces, los genetistas de poblaciones han buscado entender y predecir los efectos que tendrán en las poblaciones fenómenos de carácter genético tomando en cuenta, al mismo tiempo, factores de carácter ecológico y evolutivo como tamaños poblacionales, patrones de apareamiento, migración, selección natural y patrones de distribución espacial (Hartl y Clark 1989).

En este contexto ecológico y evolutivo es que se plantea en esta tesis el estudio de diversidad de comunidades previamente descrito y el estudio de diversidad a nivel de poblaciones a continuación.

##### *4.1.1. ¿Qué estudia la genética de poblaciones?*

La genética de poblaciones estudia las bases genéticas de la evolución. En este sentido la materia prima o el tipo de información que se requiere para este tipo de estudios es la relacionada con la diversidad genética de las poblaciones.

Con información sobre la diversidad genética de las poblaciones de estudio y la aplicación de modelos matemáticos que simplifican la realidad, la genética de poblaciones

intenta determinar cómo las poblaciones evolucionan dadas ciertas restricciones biológicas. En otras palabras, la genética de poblaciones intenta describir la dinámica microevolutiva y trata de resolver ¿cuáles son las principales fuerzas evolutivas que modelan las poblaciones?

*4.1.2. ¿Por qué incluir un estudio de genética de poblaciones en un contexto ecológico?*

Como se ha dicho aquí, la genética de poblaciones toma en cuenta en sus estudios, además de factores genéticos, factores de tipo ecológico; en este sentido podemos decir que hay mucho de evolución en la ecología y viceversa. Existen muchos estudios, en particular de bacterias, en donde las herramientas de la genética de poblaciones han ayudado a dilucidar los procesos evolutivos por los que pasan las poblaciones y las implicaciones ecológicas de los mismos. Sin embargo, aún existen controversias importantes ya que existe desconocimiento sobre la dinámica ecológica de bacterias y microorganismos en general.

Numerosos estudios ecológicos en plantas y animales han revelado que estos tienen rangos de distribución limitados determinados por causas ambientales o contingencias históricas (Futuyma 1998). El estudio de la distribución de la biodiversidad en espacio y tiempo es la biogeografía, y la información que surge de estas investigaciones nos ofrece una aproximación a los mecanismos que generan y mantienen la diversidad, tales como especiación, extinción, dispersión e interacciones entre especies (Brown y Lomolino 1998). Debido a esto, en macroorganismos se han estudiado ampliamente los mecanismos evolutivos como la especiación. En contraste, para microorganismos esta aproximación apenas se encuentra en desarrollo y ha generado una gran polémica que ha ocupado muchas publicaciones en los años recientes (Hughes *et al.* 2006).

Hacia 1934 Baas-Becking, un estudioso de microorganismos alemán, llegó a la misma conclusión a partir de lo que él había observado fundamentalmente en protozoarios: “...*todo está en todos lados y la naturaleza selecciona*”. En el siglo XXI, otro protozoólogo, Finlay (2002), publicó lo que despertaría la discusión actual sobre la distribución de microorganismos, que es fundamentalmente la misma idea planteada por Baas-Becking, pero extendida a todos los microorganismos. Los principales argumentos para postular que no hay barreras a la dispersión de microorganismos son la gran abundancia de individuos y su pequeñísimo tamaño corporal (Finlay 1998). A pesar de lo que algunos pudiéramos pensar, esta visión es aparentemente consistente con la falta de evidencia de patrones de distribución para morfoespecies de protozoarios ciliados (Finlay 2002). Debido a que los procariontes son aún más pequeños que los microorganismos eucariontes y aún más abundantes, estos han sido postulados, también por Finlay (2002), como ubicuos. Este postulado ha sido apoyado por algunos trabajos con procariontes marinos y terrestres al menos con marcadores moleculares como el gen del 16SrRNA (Fulthorpe *et al.* 1998, Cho y Tiedje 2000, Bent *et al.* 2003, Oda *et al.* 2003). Sin embargo, y en contraste con esto, al analizar la distribución de microorganismos a niveles de resolución taxonómica más finos, por ejemplo análisis multilocus de secuencias de genes codificadores de proteínas, o huellas genómicas basadas en la amplificación por PCR (polymerase chain reaction) de secuencias repetitivas extragénicas palindrómicas (rep-PCR) se ha mostrado la existencia de linajes endémicos en escalas espaciales que van desde 10 m en el suelo, diferentes continentes y habitats diferentes en agua y suelo (Fulthorpe *et al.* 1998, Cho y Tiedje 2000, Bent *et al.* 2003, Oda *et al.* 2003, Papke *et al.* 2003, Whitaker *et al.* 2003). Estos últimos estudios ponen en una situación de cuestionamiento si es que la idea de que “*todo está en todos lados y la naturaleza selecciona*” es aplicable *de facto* a los

microorganismos. Claramente, estudios poblacionales como los mencionados han ayudado a generar información sobre la distribución real de las poblaciones y sobre la dinámica de flujo entre poblaciones distantes. De manera paralela, análisis poblacionales también son importantes al tratar de dilucidar los factores que influyen en el establecimiento o que ofrecen una barrera a la dispersión de microorganismos y con ello conocer más sobre los mecanismos de especiación. De este modo es que el punto sobre la distribución de microorganismos es de gran relevancia pues lleva implícitos aspectos sobre la dinámica evolutiva de estos.

## 4.2. Estructura genética de poblaciones de *Pseudomonas sensu stricto* en un sistema acuático de Cuatro Ciénegas Coahuila, México, implicaciones ecológicas y evolutivas

En el espíritu de conocer la diversidad biológica de Cuatro Ciénegas a diferentes niveles taxonómicos y en consecuencia, evolutivos, se escogió como organismo modelo para la aproximación a nivel de poblaciones al grupo *Pseudomonas*, el cual nos permitiría entender con más detalle la ecología y evolución bacteriana en los sistemas acuáticos de Cuatro Ciénegas. Los motivos para esta elección fueron cuatro principalmente: primero, durante el análisis de comunidades obtuvimos evidencia de que el grupo  $\gamma$ -protoeubacteria, al cual pertenece el género mencionado, no sólo es el más abundante, si no que se encuentra distribuido ampliamente en los sistemas acuáticos del valle, lo cual era fundamental; segundo, la búsqueda en bases de datos de fragmentos terminales (MiCA: <http://mica.ibest.uidaho.edu/pat.php>) mostró probable presencia abundante de *Pseudomonas*, tercero, se requería de un grupo que pudiera cultivarse fácilmente para recuperar colonias aisladas y hacer el estudio poblacional; y cuarto, se trata de un grupo para el cual existe mucha información sobre diversidad y estructura genética lo cual permite comparación de los datos obtenidos para las cepas de Cuatro Ciénegas con los datos obtenidos de otras fuentes para llegar a conclusiones más generales.

### 4.2.1 Antecedentes

En los últimos años, las investigaciones sobre patrones de distribución de microorganismos han tenido un desarrollo notable (Hughes *et al.* 2006) sobre todo debido a la polémica afirmación “*todo está en todos lados y la naturaleza selecciona*” (Baas-Becking 1934). Muchos estudios se han enfocado en reunir evidencia de organismos no cultivables que ayude a resolver esta cuestión y de paso nos abran la puerta a hipótesis

sobre la dinámica evolutiva de los microorganismos. Sin embargo, la única manera de conocer la dinámica evolutiva de cualquier organismo, es mediante aproximaciones de genética de poblaciones. Hoy en día existen trabajos notables sobre la estructura genética de las poblaciones de diferentes microorganismos, desde organismos de vida libre como *Bacillus subtilis* (Istock 1992), patógenos y comensales humanos como *Escherichia coli* (Souza *et al.* 1999) y organismos simbiotes como *Rhizobium* (Piñero *et al.* 1988, Souza *et al.* 1992, Silva *et al.* 2002). A pesar del gran avance que para el entendimiento de la dinámica evolutiva de los microorganismos tienen estos trabajos, es notable el sesgo hacia organismos patógenos (Milkman 1973, Selander y Levin 1980, Ochman *et al.* 1983, Caugant *et al.* 1987, Musser *et al.* 1990a, 1990b, Smith *et al.* 1991) y el pobre conocimiento que se tiene sobre organismos no asociados a un hospedero o de vida libre, especialmente no extremófilos (Istock *et al.* 1992, Wise *et al.* 1995, Cho y Tiedje 2000, Oda *et al.* 2003).

Debido a la gran diversidad metabólica y al notable número de representantes patógenos del género *Pseudomonas*, éste ha sido ampliamente estudiado desde los inicios de la microbiología (Palleroni 1993). Existen varios estudios de genética de poblaciones de especies del género, sobre todo de patógenos de humanos como *P. aeruginosa* y de patógenos de plantas como *P. syringae* (Sarkar y Guttman 2004, Vernez *et al.* 2005). Estos estudios han sentado el precedente de altos niveles de clonalidad dentro de los linajes del género y una gran diversidad genética a nivel de grupo (Spiers *et al.* 2000). Sumado a esto, uno de los trabajos recientes más importantes sobre distribución espacial de microorganismos muestra una clara estructuración espacial a nivel continental de genotipos de distintas especies de vida libre del género (Cho y Tiedje 2000). Estos datos nos plantean un escenario de aislamiento sexual y de límites a la dispersión del grupo. Finalmente, es de

notar la falta de estudios tanto evolutivos como de distribución geográfica a nivel local. Los estudios sobre evolución toman en consideración cepas provenientes de regiones geográficas muy distantes y épocas distantes (Kiewitz y Tümmler 2000, Rius *et al.* 2001, Sarkar y Güttman 2004), mientras que lo que se conoce por Cho y Tiedje (2000) sobre la distribución espacial, involucra muestras de los cinco continentes y la escala espacial más pequeña es de alrededor de 100 km.

A pesar de que los estudios previos sobre diversidad y distribución de especies de *Pseudomonas* apuntan a la posibilidad de límites a la dispersión y a una estructura genética clonal, consideramos relevante realizar un estudio con cepas de vida libre a nivel local que involucre ambas aproximaciones: evolutiva con análisis de genética de poblaciones y ecológica: con análisis que exploren los factores involucrados en la limitación a la dispersión y/o establecimiento de ciertas cepas o linajes.

De este modo realizamos un estudio de diversidad genética y distribución geográfica con cepas del género *Pseudomonas* que se colectaron en verano e invierno en varios puntos de un sistema acuático de 2 km de longitud aproximadamente. En primer término se analizó la diversidad de genotipos y su distribución ecológica y en segundo lugar se realizó un estudio filogenético y de genética de poblaciones de los grupos genéticos más abundantes en la muestra. Por un lado, los resultados sobre el análisis de distribución de las variantes genéticas indican que la distribución no parece obedecer a ningún factor ambiental, aunque existe diferenciación genética entre los linajes presentes en invierno y verano. Por otro lado y atendiendo al análisis de genética de poblaciones, es notable la existencia de varios linajes evolutivos con muy bajos niveles de recombinación y estructura clonal. En suma los datos indican una estructura clonal de las poblaciones sugiriendo posibles eventos de selección periódica asociados a los cambios estacionales.

---

#### **4.2.2. Muestreo**

*4.2.2.1. Área de estudio y toma de muestras.*- El sistema acuático de donde se obtuvieron las cepas de este estudio es conocido como Churince, se encuentra en el valle de Cuatro Ciénegas en el estado de Coahuila en México y forma parte del Desierto Chihuahuense. La extensión del valle es de aproximadamente 40 km de este a oeste y de 30 km de norte a sur a una altitud de 740 msnm (Minckley 1969). A pesar de la poca precipitación en la zona (200 mm de precipitación anual y más de 2000 mm de evaporación potencial, SEMARNAT 2003) existen cientos de pozas y sistemas acuáticos de características contrastantes. Uno de estos sistemas es el sistema Churince, el cual consta de un manantial, una laguna intermedia y una gran laguna de desecación o laguna grande que se conectan en un solo sistema por pequeños riachuelos. Este sistema fue elegido para el estudio pues presenta condiciones en las características del agua que van cambiando desde al manantial hasta la laguna de desecación en menos de 2 km, por lo que resulta adecuado para probar hipótesis sobre la influencia de las condiciones ambientales en la dinámica poblacional y evolutiva del grupo biológico de interés.

El sistema Churince fue analizado fisicoquímicamente utilizando un equipo Hydrolab YSI600QS con una sonda YSI600R (YSI Inc., Yellow Springs, OH, USA). Los datos así obtenidos sirvieron para elegir ocho puntos de muestreo que resultaron diferentes en pH, oxígeno disuelto y temperatura (Tabla 4.1)

Tabla 4.1. Coordenadas geográficas y medidas fisicoquímicas del agua de los sitios de muestreo en el sistema Churince en Cuatro Ciénegas Coahuila, México.

Sitio	Latitud (N)	Longitud (W)	°T (°C)	$\Delta T^a$ (°C)	pH	Do <sup>b</sup> (mg/L)	Cond. <sup>c</sup> (mohms)
LG1	26°50.830′	102°09.335′	25.03	3.88	10	9.6	4.2
LG2	26°51.199′	102°09.009′	25.03	3.88	9.8	9.5	4.2
LG3	26°51.146′	102°08.964′	25.45	3.84	9.7	9	4.2
LI6	26°50.927′	102°08.539′	26.84	2.37	9.3	8	2.2
LI7	26°50.910′	102°08.438′	26.44	2.59	9.3	8	2
LI8	26°50.897′	102°08.558′	26.46	4.26	9.3	8	1.8
MAN9	26°50.717′	102°08.295′	27.6	1.16	8.9	7.5	1.6
MAN10	26°50.421′	102°08.047′	28.82	0.47	8.7	6	1.4

a. Variación estacional de la temperatura

b. Oxígeno Disuelto

c. Conductividad

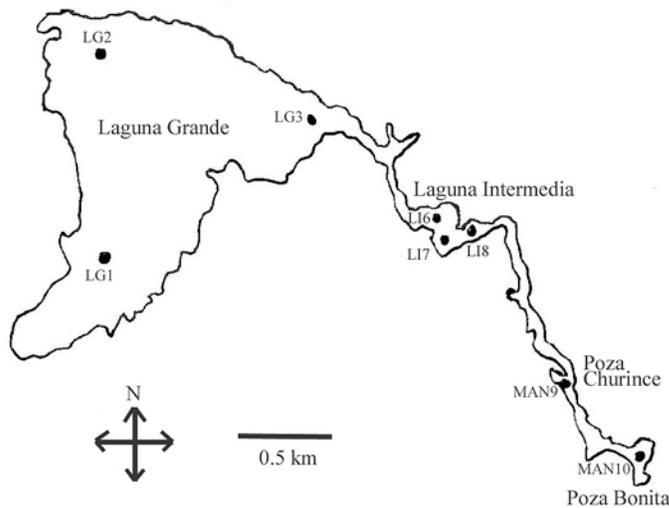


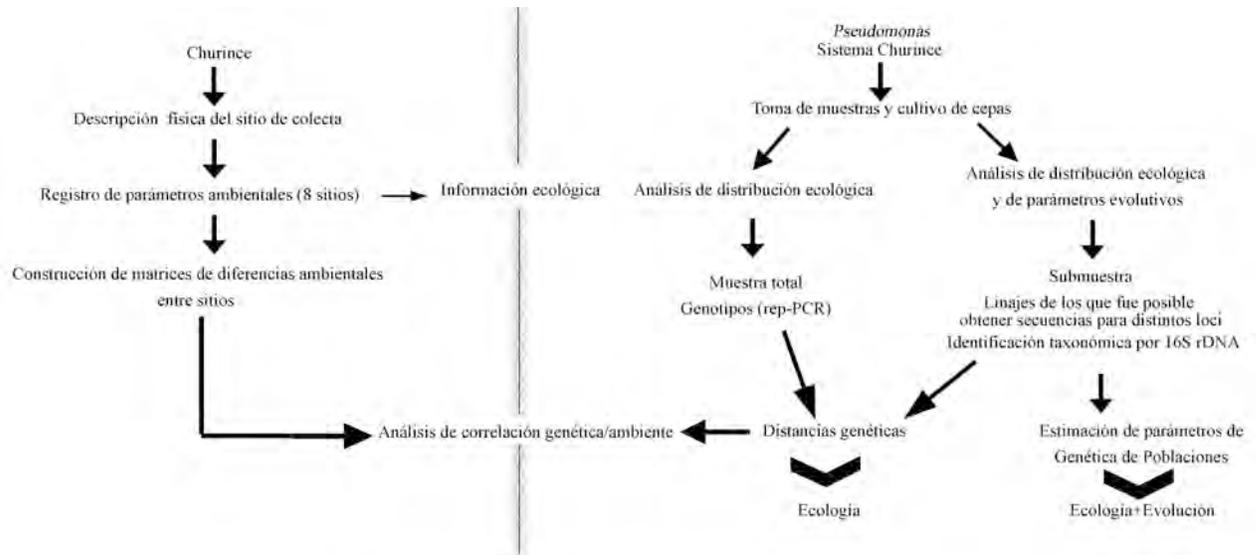
Fig. 4.1. Representación gráfica de los sitios de muestreo en el Sistema Churince en Cuatro Ciénegas, Coahuila México.

Se realizaron colectas en dos ocasiones (verano de 2003 e invierno de 2004). En cada ocasión se tomaron 3 muestras de 15 ml de agua superficial en los ocho puntos seleccionados (Fig. 4.1) utilizando tubos estériles cónicos BD Falcon™ (BD Biosciences Two Oak Park Bedford, MA 01730 USA). De cada una de las muestras se tomaron 3 submuestras de 200 µl para sembrarlos en cajas de Petri.

*4.2.2.2. Microorganismos y condiciones de aislamiento.*- Las cepas utilizadas en este estudio se aislaron utilizando el medio GSP (para un litro de medio GSP: glutamato de sodio L(+) $10.0$  g; almidón soluble  $20.0$  g; fosfato de potasio dihidrogeno  $2.0$ g; sulfato de magnesio  $0.5$  g; rojo de fenol  $0.36$  g; agar-agar  $12.0$  g; propuesto por Kielwein 1971). En este medio de cultivo, que contiene rojo de fenol así como glutamato y almidón como únicos nutrientes, crecen preferentemente los géneros *Pseudomonas* ( $\gamma$ -proteobacteria) y *Aeromonas* ( $\delta$ -protoebacteria), pues los géneros que normalmente los acompañan no pueden degradar estos compuestos (Stanier *et al.* 1966). El almidón es degradado por *Aeromonas* con producción de ácido, causando un cambio de color del rojo de fenol a amarillo, y en el caso de *Pseudomonas* no hay tal producción y el color se mantiene rojo-violeta. Gracias a esto es posible seleccionar fácilmente las cepas del género deseado. La temperatura de incubación fue  $28-29^{\circ}\text{C}$  y el crecimiento de las cepas de interés se dio entre uno y hasta tres días. En los subsecuentes procedimientos de análisis se siguió utilizando el mismo medio de cultivo.

*4.2.2.3. Muestras y análisis: aproximación general.*- A manera de resumen, a continuación se presenta un cuadro en donde se establecen las muestras que se

utilizaron durante los análisis con el objetivo de abordar dos enfoques, uno netamente ecológico y otro con un sesgo hacia genética de poblaciones y evolución.



## I. Distribución ecológica de genotipos

Como primera aproximación al estudio de la distribución de la diversidad de *Pseudomonas* en el sistema descrito, se obtuvieron patrones fingerprint de todas las cepas aisladas de tal modo que fuera posible identificar los diferentes genotipos y analizar su distribución espacial/ambiental.

*Extracción de DNA.*- La extracción de DNA genómico que se utilizó para todos los procedimientos de análisis genético se llevó a cabo utilizando el kit DNeasy™ (Qiagen™, Valencia, CA, USA).

*rep-PCR.*- El DNA genómico de los aislados fue utilizado para la genotipificación por medio de rep-PCRs. Esta estrategia molecular se basa en la existencia en el genoma procarionte de secuencias conservadas palindrómicas repetitivas extragénicas o REP por sus siglas en inglés (repetitive extragenic palindromic;

Versalovic *et al.* 1991, 1994). De este modo es posible amplificar los segmentos de DNA que se encuentran flanqueados por estas secuencias conservadas y obtener patrones de huella digital o *fingerprint* genómico con el uso de oligonucleóticos complementarios a las secuencias palindrómicas mencionadas. En este caso se utilizó el primer BOX (5'- CTA CGG CAA GGC GAC GCT GAC G-3'; Louws *et al.* 1994) y las condiciones de amplificación fueron para reacciones de 50 µl: Buffer 2X, MgCl<sub>2</sub> 4.6 mM, dNTP's 5 mM, BOX 100 µmol, DMSO 10%, Taq 2U, DNA 20 pmol. Para la amplificación se siguió el siguiente programa de temperaturas en un termociclador MJ Research (Watertown, MA, USA): 95°C por 7 min, 30 ciclos de 94°C por 1 min, 53°C por 1 min, 65°C por 8 min y extensión final de 65°C por 8 min. Los productos así obtenidos se sometieron a electroforesis en geles de agarosa 1.5% en buffer TAE 0.5X por 5 horas a 180 mV (5 V/cm). Los geles se tiñeron con bromuro de etidio y fotografiado en un transiluminador UV para obtener una imagen digital en formato tiff (tagged image file format).

*Análisis de diversidad y cluster de los patrones rep.-* Con el objetivo de visualizar posibles patrones de distribución de genotipos de *Pseudomonas* dentro de la colección, las imágenes digitales de los patrones de bandeo obtenidos se analizaron utilizando el software GelCompar (versión 4.0; Applied Maths, Kortrijk, Bélgica). Para cada patrón de bandeo (uno por cepa), se tomó en cuenta la presencia y abundancia de cada banda y se calculó un coeficiente de correlación de Pearson. Este índice representa la separación angular entre los vectores normalizados de datos medidos a partir de su media. En otras palabras, compara la distribución de

datos de dos variables (patrón de bandeo) y establece el grado de solapamiento de éstas. El cálculo del coeficiente de correlación lineal se realiza dividiendo la covarianza por el producto de las desviaciones estándar de ambas distribuciones (X y Y):

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

$\sigma_{XY}$  la covarianza de (X,Y)

$\sigma_X$  y  $\sigma_Y$  las desviaciones típicas de las distribuciones marginales.

Con este índice de correlación se construyó un dendrograma o cluster tipo UPGMA (unweighted pair group method with arithmetic distances; Sneath y Sokal 1973). La matriz de similitud generada indicaba 100 como identidad absoluta y 0 como completamente diferentes. Para poder evaluar posibles correlaciones entre diferencias ambientales y distancia genética dada por diferencias en los patrones rep, la matriz de similitud se transformó en una matriz de diferencias (100-valor de la matriz de similitud) y se dividió este valor entre 100 para hacer el rango de 0 a 1. Se hicieron dos tipos de análisis para evaluar posibles patrones de distribución de las variantes genéticas asociados a alguno de los parámetros ambientales medidos. En primer lugar se llevaron a cabo correlaciones simples (lineales) entre matrices. En segundo lugar se realizaron pruebas de autocorrelación de Mantel (Smouse *et al.* 1986) usando el software Arlequin (Schneider *et al.* 2000) con 1000 permutaciones. Para ambos análisis, se construyeron, además de las matrices de diferencias en patrones de bandeo mencionadas, matrices de distancia o diferencias ambientales.

Estas últimas se estimaron calculando las distancias euclidianas pareadas en una dimensión de los sitios en donde se colectó cada cepa, utilizando los valores normalizados de cada parámetro ambiental considerado (temperatura, pH, conductividad, oxígeno disuelto y variación estacional de la temperatura). Para el caso de la distancia geográfica, se calculó la distancia euclidiana o lineal en dos dimensiones entre sitios de muestreo.

La distancia euclidiana entre dos puntos  $P = (p_1, p_2, \dots, p_n)$  y  $Q = (q_1, q_2, \dots, q_n)$  en un espacio  $n$  euclidiano se define como

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} .$$

En el caso de una dimensión (un solo valor por parámetro, por ejemplo temperatura), para dos puntos  $P = (p_x)$  y  $Q = (q_x)$ , la distancia se calcula como:

$$\sqrt{(p_x - q_x)^2} = |p_x - q_x|$$

En el caso de dos dimensiones (dos valores por parámetro, por ejemplo coordenadas geográficas), para dos puntos  $P = (p_x, p_y)$  y  $Q = (q_x, q_y)$ , la distancia se calcula como:

$$\sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$$

## II. Diversidad filogenética y genética de poblaciones

Con la intención de conocer con más detalle la diversidad y la dinámica ecológica y evolutiva de las poblaciones de *Pseudomonas* del sistema descrito, se llevó a cabo un

estudio filogenético y de genética de poblaciones de todas las cepas aisladas pertenecientes al género, con énfasis en los linajes numéricamente más abundantes. Para cumplir con esto fue necesaria la selección de otros marcadores moleculares.

### **Estrategias moleculares**

*Selección de otros marcadores moleculares.*- Durante la historia de la genética de poblaciones en bacterias se han utilizado diversos marcadores moleculares, principalmente genes metabólicos. Los marcadores moleculares son útiles para cuantificar la diversidad genética y monitorear la dispersión de organismos y sus genes (Avice 1994). Sin embargo, las moléculas son más que marcadores de diversidad y distribución, ya que están involucradas en la adecuación de los organismos, por lo que entender las consecuencias de la variación genética es un problema central en la ecología (Avice 1994). En el presente estudio se utilizaron cinco genes marcadores para el estudio a nivel poblacional.

Recientemente se ha discutido (Yamamoto *et al.* 2000, Sarkar y Guttman 2004) que a pesar de la utilidad de la secuencia del gen 16SrRNA, se tendría una idea más clara de la estructura poblacional de las especies si se analizaran más genes de tipo “*housekeeping*” o esenciales. Estos genes son componentes del genoma llamado “*core*” y son menos propensos a sufrir transferencia horizontal. Los genes *housekeeping* son particularmente útiles para evaluar la importancia de recombinación en la evolución de linajes clonales, como en general es el caso del género *Pseudomonas* (Spiers *et al.* 2000) con sólo algunos eventos de recombinación reportados (Kiewitz y Tümmler 2000). Los genes *housekeeping* típicamente codifican proteínas esenciales para la sobrevivencia del organismo y evolucionan neutralmente o están selectivamente restringidos, este tipo de genes pueden ser usados para rastrear la historia evolutiva de linajes clonales en el tiempo.

De acuerdo con lo anteriormente dicho, se escogieron cuatro genes esenciales para el análisis poblacional del grupo de estudio, además del marcador filogenético 16S rRNA.

*a. DNA girasa subunidad B (gyrB).*- La proteína codificada por este gen es la enzima responsable de introducir superenrollamientos negativos en los cromosomas bacterianos y juega un papel crucial en la replicación (Watt y Hickson 1994)

*b. Factor Sigma70 (rpoD).*- La proteína codificada por este gen es uno de los factores que dan la señalización específica en el promotor para que la RNA polimerasa empiece la transcripción (Lonetto *et al.* 1992).

*c. Recombinasa A (recA).*- La proteína codificada por este gen es esencial en la recombinación homóloga y en la inducción del sistema SOS de reparación de DNA (Madigan *et al.* 2000).

*d. Aconitato Hidratasa B o Aconitasa (acnB).*- Enzima que actúa en el ciclo de Krebs, transforma el citrato en cis-aconitato y en este sentido es importante para la respiración celular de organismos aerobios (Madigan *et al.* 2000).

*Amplificación de genes, clonación y secuenciación.*- Para estudiar la variación y estructura genética de las poblaciones de *Pseudomonas* se amplificaron, clonaron y secuenciaron cuatro genes: *gyrB*, *rpoD*, *recA* y *acnB*, además de 16SrRNA para la asignación taxonómica de cada grupo. Para la amplificación por PCR se realizaron reacciones (50 µl) que contenían, para el gen 16SrRNA: 1X de Buffer para PCR (Invitrogen), 2 mM MgCL<sub>2</sub>, 0.8 mM dNTPs, 0.3 µM de cada oligonucleótido y una unidad de *Taq* polimerasa, y para los genes *gyrB*, *rpoD*, *recA* y *acnB*: 1X de Buffer para PCR (Invitrogen), 2 mM MgCL<sub>2</sub>, 0.8 mM dNTPs, 1 µM de cada oligonucleótido y una unidad de *Taq* polimerasa. Utilizando

un termociclador MJ Research (Watertown, MA, USA), se siguió el siguiente patrón de temperaturas para todas las amplificaciones: 10 min a 95°C, siguiendo con 30 ciclos de 1 min a 94°C, 1 min a la temperatura de alineación apropiada y 2 min de extensión a 72°C, por último un ciclo de extensión final a 72°C por 10 min. Las temperaturas de alineamiento y las secuencias de los oligonucleótidos utilizados para la amplificación de los distintos genes se muestran en la Tabla 4.2.

16S rDNA.- Es un marcador molecular ampliamente utilizado para establecer relaciones filogenéticas especialmente en bacterias, e incluso en la delimitación de especies (Woese 1987, Woese *et al.* 1990, Stackebrandt y Goebel 1994). Corresponde a la parte del DNA que codifica a la subunidad pequeña del rRNA. El uso de este marcador se ha difundido mucho pues aparentemente un valor de similitud  $\geq 97\%$  en la secuencia completa de este gen correlaciona bien con un 70% de identidad en todo el genoma revelado por experimentos de reasociación de DNA-DNA, ambas medidas han sido establecidas como valores estándar para delimitar especies en bacterias (Wayne *et al.* 1987, Stackebrandt y Goebel 1994). La secuencia de este gen en el presente trabajo se utilizó sólo en representantes de distintos grupos evaluados por rep-PCR y los genes anteriormente mencionados, con la intención de conocer la afiliación taxonómica mediante la comparación de las secuencias obtenidas con las bases de datos públicas y la reconstrucción filogenética usando estos mismos datos.

Todos los productos de PCR fueron sometidos a electroforesis en geles de agarosa para cortar la banda de interés y limpiarla utilizando el kit QIAquick gel extraction kit™ (Qiagen™, Valencia, CA, USA). Los productos de PCR limpios fueron ligados a un vector de clonación TOPO TA Cloning® (Invitrogen, 1600 Faraday Avenue Carlsbad, CA 92008

USA) y clonados en células DH5 $\alpha$ <sup>TM</sup>-T1R de *Escherichia coli* químicamente competentes. Finalmente, los plásmidos o vectores de clonación fueron extraídos y el fragmento de interés secuenciado. Las reacciones de secuenciación se realizaron utilizando reactivos de Applied Biosystems y corridos en un secuenciador ABI 3100- Avant Prism® (Applied Biosystems, Foster City, CA, USA). Para todos los genes se utilizaron los primers de secuenciación universales M13(-21) y M13R, y para el gen 16SrRNA se utilizaron también los primers internos: 530F, 790F, 357R, 530R y 981R (Tabla 4.3).

Tabla 4.2. Oligonucleótidos utilizados para la amplificación por PCR de los genes estudiados.

Amplicón	Primer	Tm(°C)	Secuencia (5'-3')	Pb	Tamaño amplicón	Referencia
16SrDNA	8F	52	AGTTGATCCTGGCTCAG	17	1500pb	Lane 1991
	1492R		ACCTTGTTACGACTT	15		
<i>rpoD</i>	70F	63	TGTAYATGMNGARATG GGNACNGT	25	850pb	Modificado de Yamamoto <i>et al.</i> 2000
	70R		GTTNGCYTCNACCATYTC YTTYTT	24		
<i>gyrB</i>	<i>gyrB</i> _F	60	ACCAYGSNNGGNGNAART TYRA	22	1000pb	Modificado de Yamamoto <i>et al.</i> 2000
	<i>gyrB</i> _R		AGTGCNNGRTCYTTYTCY TGRCA	23		
<i>acnB</i>	<i>acn</i> -Fp	60	ACATCCCCTGCACGCYCT GGCC	23	700pb	Sarkar y Guttman 2004
	<i>acn</i> -Rp		GTGGTGCTCTGGGAACCGA CGGTG	24		
<i>recA</i>	<i>recA</i> -F	60	CAGATCGAAMRNCARTTCGG	20	900pb	Este estudio <sup>a</sup>
	<i>recA</i> -R		YTTRCCCTGDCCGATCTT	18		

<sup>a</sup> Oligonucleótidos diseñados para este estudio a partir de alineación de secuencias disponibles en las bases de datos públicas (GenBank).

Tabla 4.3. Oligonucleótidos utilizados para la secuenciación de amplicones insertados en vectores de clonación.

Amplicón	Primer	Secuencia (5'-3')	pb	Referencia
16SrDNA	530F	CAGCAGCCGCGGTAATAC	18	Sacchi <i>et al.</i> 2002
	790F	ATTAGATACCCTGGTAG	17	Sacchi <i>et al.</i> 2002
	357R	CTGCTGCCTCCCGTA	15	Sacchi <i>et al.</i> 2002
	530R	GTATTACCGCGGCTGCTG	18	Sacchi <i>et al.</i> 2002
	981R	GGGTTGCGCTCGTTGCGGG	19	Sacchi <i>et al.</i> 2002
Vector	M13(-21)	TGAAAAACGACGGCCAGT	18	TOPO
	M13R	CAGGAAACAGCTATGACC	18	TOPO

### Análisis de datos

*Ensamble y edición de datos de secuencias.*- Las secuencias completas de cada gen se ensamblaron y revisaron con la ayuda del software PhredPhrap y Consed (Gordon *et al.* 1998, Gordon *et al.* 2001). Las secuencias del gen 16SrRNA se alinearon con las secuencias más parecidas encontradas en la base de datos del gen ribosomal (Ribosomal Database Project II o RDPII) y las secuencias de los otros genes con las secuencias más parecidas del GenBank (NCBI database). La alineación de las secuencias del gen 16SrRNA se hizo con MUSCLE (Edgar 2004) y revisada con NAST (DeSantis *et al.* 2006). La alineación de las secuencias del resto de los genes se hizo con DAMBE 4.5.50 (Xia y Xie 2001) utilizando como base el alineamiento de aminoácidos para alinear sobre este las secuencias de nucleótidos. Las secuencias alineadas automáticamente fueron revisadas manualmente.

*Relaciones filogenéticas.*- Con el objetivo de visualizar las relaciones filogenéticas entre las distintas cepas estudiadas y de establecer la existencia de los principales linajes evolutivos presentes en la colección de cepas, se evaluó el modelo de sustitución nucleotídica más adecuado para cada gen usando Modeltest implementado en el software PAUP\* (Swofford 2002). Para cada gen, y bajo el modelo de sustitución correspondiente, se calcularon matrices de distancia genética. Con esta información, se generaron árboles por Neighbor-Joining (Saitou y Nei 1987) y análisis tipo Bootstrap (1000 réplicas) para representar las relaciones evolutivas moleculares entre las cepas analizadas para cada uno de los cinco genes (16SrRNA, *gyrB*, *rpoD*, *acnB* y *recA*) usando el software MEGA 3.1 (Kumar *et al.* 2004). Los resultados de este análisis permitieron identificar tres linajes con mayor representación numérica en la colección total de cepas. De esta forma todos los análisis subsecuentes se realizaron tomando por separado el total de la muestra y cada uno de los linajes numéricamente más abundantes, los cuales se nombraron como: *Pseudomonas otitidis*, *P. psychrotolerans* y “*P. caenicarum*”, esta última resultó ser un nuevo linaje dentro del género y en el Apéndice I se encuentra su descripción.

### *Genética de poblaciones*

*Análisis para evaluar recombinación.*- Con la intención de evaluar los niveles de clonalidad en las poblaciones de *Pseudomonas* de este estudio, se realizaron varias pruebas que evalúan los niveles de recombinación. En la literatura existen cuatro tipos de evidencia que han sido utilizados para demostrar la participación de la recombinación en la evolución de organismos: a) presencia de desequilibrio de ligamiento entre alelos de diferentes loci de una población (Dykhuizen y Green 1986, Souza *et al.* 1992, Haubold *et al.* 1998); b) distribución de sitios polimórficos en la secuencia de un gen (Maynard-Smith 1992); c)

existencia de incongruencias filogenéticas entre genealogías de genes (Lecointre *et al.* 1998); d) la construcción de relaciones filogenéticas permitiendo que se construyan a manera de red y no necesariamente como árbol usando el test llamado Split-Decomposition (Huson 1998). Adicionalmente, de manera muy reciente se ha desarrollado una prueba para localizar sitios posibles de recombinación usando incongruencias filogenéticas y un algoritmo genético llamado GARD por sus siglas en inglés (Genetic Algorithm Recombination Detection, Pond *et al.* 2006).

De las pruebas mencionadas, en este trabajo aplicamos las primeras tres pruebas a continuación para evaluar recombinación intragénica y la cuarta para evaluar recombinación entre genes.

I) Prueba de incongruencia filogenética conocido como ILD (Incongruence Length Difference Test).- Este análisis pone a prueba la hipótesis nula de congruencia entre grupos de datos (Farris *et al.* 1995, Lecointre *et al.* 1998). La prueba se realizó tomando todas las posibles comparaciones pareadas entre loci. Para los 4 loci (*acnB*, *gyrB*, *recA* y *rpoD*) se realizaron pruebas pareadas para 1) todas las cepas y 2) cada uno de los tres linajes numéricamente más abundantes en la colección (*P.otitidis*, *P.psychrotolerans* y “*P.caenicarum*” (ver Apéndice I para descripción de esta especie). En todos los casos se usó parsimonia con búsqueda tipo Branch and Bound, con un tope máximo de árboles de 10,000 y 100 réplicas con el software PAUP\* (Swofford 2002). En este caso se hacen análisis independientes de parsimonia para dos conjuntos de datos individuales (las secuencias de dos loci: *x* y *y*), después las longitudes de cada árbol de máxima parsimonia se suman ( $L_x + L_y$ ), y esta longitud es comparada con la suma de longitudes ( $L_p + L_q$ ) de árboles de máxima parsimonia obtenidos de dos conjuntos de datos (*p* y *q*), del mismo tamaño que los datos originales y generados por particiones al azar del par original de

datos. Para  $W$  réplicas de esta prueba,  $S$  es el número de veces que la longitud de los árboles al azar es mayor que la de los árboles originales, y la hipótesis nula de congruencia se rechaza cuando la tasa de error (valor  $P = 1-S/W+1$ ) es menor a un valor límite (i.e. 5%) significa que hay más incongruencia entre las particiones que lo esperado al azar, por lo que se rechaza la hipótesis nula de congruencia (Lecointre *et al.* 1998).

II) Split-Decomposition.- Este análisis representa gráficamente conflictos filogenéticos entre sitios de un alineamiento múltiple y se realizó para los cuatro loci individualmente (*gyrB*, *acnB*, *rpoD* y *recA*) usando algoritmos de parsimonia, la distancia de Hamming y 1000 réplicas para el cálculo de bootstrap, con el software SplitsTree4 versión 4.8 (Huson y Bryant 2006). La prueba se llevó a cabo para la muestra completa y por separado también para los tres linajes filogenéticamente bien soportados y numéricamente más abundantes en la colección (*P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”). Esta prueba utiliza un método de parsimonia que no impone a los datos una estructura ramificada en forma de árbol (Huson 1998). El objetivo de la prueba es mostrar gráficamente y al mismo tiempo todas las mejores topologías o relaciones entre los taxa analizados dada la información de loci particulares. De esta manera el análisis resultará en una red si más de uno de los acomodos tiene buen soporte estadístico (Morrison 2005). Al visualizar la existencia de posiciones que producen señales filogenéticas incompatibles (reticulación) es posible hacer interpretaciones sobre recombinación homóloga como causa de dichos conflictos (Brown *et al.* 2002).

III) Prueba de recombinación Single Breakpoint y GARD.- Esta prueba se pensó originalmente para descartar falsos positivos en análisis de evolución no neutral (con selección) al localizar sitios probables de recombinación en grupos de datos de secuencias de DNA. La prueba se realizó usando el software disponible para pruebas de

recombinación en [www.datamonkey.org](http://www.datamonkey.org). Los análisis que se llevaron a cabo fueron *Single BreakPoint* y *GARD*, y se usó como modelo de sustitución el recomendado por los autores para grupos de datos pequeños (HKY85). Se realizaron pruebas por locus individuales para todas las cepas y por separado para tres linajes filogenéticamente bien sostenidos y más abundantes (*P. otitidis*, *P. psychrotolerans* y "*P. caenicarum*"). El método está diseñado para buscar el número máximo de sitios de ruptura en filogenias (incongruencias o *breakpoints*) pues infiere filogenias para cada fragmento no recombinante y elige el mejor resultado (sitios de recombinación) usando el criterio de Akaike (Sugiura 1978) derivado de un modelo de máxima verosimilitud (Pond *et al.* 2006).

IV) Comparación de topologías filogenéticas.- Esta fue la única prueba que se aplicó para evaluar posible recombinación entre genes. El método se basa en la comparación de dendogramas UPGMA (Nei 1975) construidos a partir del alineamiento completo de secuencias concatenadas de 1 a *n* genes. En este caso en particular se hizo el análisis por pares de genes (*acnB* y *gyrB*, *acnB* y *recA*, *acnB* y *rpoD*, *gyrB* y *recA*, *gyrB* y *rpoD*, *recA* y *rpoD*). El fundamento del análisis se basa en comparaciones de tres en tres secuencias, por ejemplo A, B y C, en donde A y B son más parecidas entre sí que con C. En primer lugar se descartan todos los sitios no informativos (aquellos que no permitan resolver las relaciones de similitud entre secuencias). Posteriormente se analizan todas las secuencias en pequeñas ventanas de bases avanzando de una en una y se estima el porcentaje de identidad para cada una de los tres posibles pares en cada posición. Las regiones con potencial de recombinación son aquellas en donde el porcentaje de identidad es mayor para el par A-C o B-C que para A-B. Este análisis se realizó en el programa RDP2 usando el análisis RDP (Martin *et al.* 2005).

---

**Análisis de variación genética.-** Se calcularon estimadores de diversidad genética estándar para 4 loci: *acnB*, *gyrB*, *rpoD* y *recA* usando el software DnaSP (Rozas y Rozas 1999). Todos los estimadores fueron calculados para la muestra total y por separado para los tres linajes ya mencionados de la colección (*P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”). Adicionalmente y con la intención de evaluar diferencias en diversidad influenciadas por las fluctuaciones estacionales de la temperatura del agua se calcularon los niveles de variación genética en el tiempo para toda la muestra. Los dos principales estimadores de diversidad genética que se calcularon son  $\pi$  (pi) y  $\theta$  (theta) para el modelo de sitios infinitos.

Diversidad nucleotídica.-  $\pi$  es una de las maneras más generales de cuantificar la variación nucleotídica en una población y consiste en determinar la proporción de diferencias de nucleótidos entre pares de secuencias y después dar peso a esta diferencia al incluir las frecuencias de las secuencias en el total de secuencias analizadas:

$$\pi = \sum_{ij} p_i p_j \pi_{ij} \quad (\text{Nei 1987})$$

donde  $p_i$  es la frecuencia de la secuencia  $i$  en la muestra,  $p_j$  es la proporción de la frecuencia  $j$  en la muestra y  $\pi_{ij}$  es la proporción de nucleótidos que difieren entre las secuencias  $i$  y  $j$ .  $\theta$  se estima a partir de la expresión:

$$\theta = 4N_e\mu \quad (\text{Kimura 1968})$$

donde  $N_e$  es el tamaño efectivo de la población y  $\mu$  es la tasa de mutación. Sin embargo, es difícil tener la determinación exacta de los parámetros  $N_e$  y  $\mu$ . De esta forma, una manera

indirecta de calcular  $\theta$  es utilizando el número total de sitios segregativos en un grupo de secuencias (sitios donde las secuencias difieren),

$$\theta = K/a \quad \text{ó} \quad \theta = K / \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n-1}\right) \quad (\text{Watterson 1975, Tajima 1983})$$

donde  $K$  es igual al número total de sitios segregativos en una muestra de secuencias y  $n$  es el número de secuencias en la muestra.

De las expresiones descritas se deduce que  $\pi$  se ve afectada de manera importante por los alelos de mayor frecuencia y es independiente del tamaño de muestra, mientras que  $\theta$  se ve afectada por el tamaño de muestra y alelos poco frecuentes (deriva génica).

#### Pruebas de selección basadas en la distribución del polimorfismo

*Prueba de D de Tajima de selección a nivel molecular.*- La relación entre ambos estimadores de diversidad ( $\pi$  y  $\theta$ ) permite determinar si las secuencias analizadas se encuentran bajo el modelo neutral o se desvían del mismo. Si ambos estimadores dan los mismos valores, el polimorfismo se considera neutro y distribuido aleatoriamente (Tajima 1983), si existen diferencias la selección podría estar participando en la diversidad observada (valores negativos = selección purificadora, valores positivos = selección diversificadora). La ecuación para la estimación del parámetro D de Tajima será:

$$D = \frac{\pi - K/a}{\sqrt{V(\pi - K/a)}}$$

donde  $a$  es igual a la definida para el cálculo de  $\theta$  y  $V$  es la varianza.

*Prueba Fu-Li de selección a nivel molecular.*- El método de detección de selección a nivel molecular de Fu y Li (1993) se desarrolló con la misma lógica que la prueba de Tajima, la diferencia fundamental reside en que para Fu y Li el parámetro a evaluar es la distribución de las mutaciones en una genealogía. Suponen que las mutaciones antiguas

tendrán que encontrarse con mayor probabilidad en las ramas más antiguas o internas de la genealogía que las mutaciones más recientes, las cuales se encontrarán en las ramas externas. Si existe selección negativa o purificadora existirá un exceso de ramas externas (jóvenes) porque los alelos deletéreos estarán en bajas frecuencias. Si existe selección positiva o diversificadora, algunos alelos serán antiguos y habrá deficiencia de mutaciones en las ramas externas. Es entonces que, al comparar el número de mutaciones en ramas internas y externas con los valores esperados de acuerdo al modelo neutro, se tendrá una prueba estadística para la detección de la participación de la selección. La prueba consiste en evaluar la siguiente expresión:

$$G = \frac{ne - ni/(a - 1)}{\sqrt{V[ne - ni/(a - 1)]}}$$

donde  $ne$  = número de mutaciones en ramas externas,  $ni$  = número de mutaciones en ramas internas,  $a$  es igual a la definida para el cálculo de  $\theta$  y donde  $V$  es la varianza. Al igual que la prueba de Tajima valores negativos indicarán selección purificadora y valores positivos, selección diversificadora.

*Prueba McDonald-Kreitman de selección a nivel molecular.* - Esta prueba desarrollada para evaluar el papel de la selección en la evolución del locus *Adh* en *Drosophila* (McDonald y Kreitman 1991), se basa en la relación entre las sustituciones sinónimas y no sinónimas de secuencias codificantes entre y dentro de poblaciones. La prueba evalúa si el cociente de las sustituciones no sinónimas a sinónimas entre las poblaciones es el mismo que dentro de las poblaciones. Para esto se establecen dos tipos de sitios variables: los sitios fijos y los sitios polimórficos. Los sitios fijos son iguales dentro de las poblaciones pero distintos entre poblaciones. Los sitios polimórficos son el resto de los sitios variables. La hipótesis nula de la prueba es neutralidad, en cuyo caso:

$$n_f/s_f = n_p/s_p$$

donde,

$n_f$  = sitios fijos con sustituciones no sinónimas

$s_f$  = sitios fijos con sustituciones sinónimas

$n_p$  = sitios polimórficos con sustituciones no sinónimas

$s_p$  = sitios polimórficos con sustituciones sinónimas

cualquier desvío de la igualdad sugiere selección positiva o diversificadora.

### Pruebas de selección basadas en sustituciones moleculares

*Coefficiente  $dN/dS$ .*- En secuencias de nucleótidos que codifican para proteínas, si existe una mutación que modifique la identidad del aminoácido de la proteína traducida, este cambio se conoce como no sinónimo, si la mutación no afecta esta identidad se conoce como cambio sinónimo. Dada esta situación el coeficiente que resulte de la tasa de sustitución no sinónima ( $dN$  o  $Ka$ ) entre la tasa de sustitución sinónima ( $dS$  o  $Ks$ ) nos dará una idea de si la variación genética estudiada se ajusta a un modelo neutro ( $dN/dS=1$ ), de selección purificadora ( $dN/dS<1$ ) o de selección diversificadora ( $dN/dS>1$ ). La estimación de estas tasas se realizó usando el método de Nei-Gojobori (1986).

**Análisis de diferenciación genética.**- El estadístico  $F_{ST}$  se calculó de acuerdo con lo establecido por Hudson *et al.* (1992), en donde  $F_{ST}=1-H_w/H_b$ , siendo  $H_w$ , la media del número de diferencias entre secuencias de una misma subpoblación o muestra y  $H_b$  es la media del número de diferencias entre secuencias de dos subpoblaciones o muestras. El fundamento del análisis se refiere al porcentaje de la diversidad que se encuentra representada por diferencias entre subpoblaciones, grupos o muestras (islas), por lo que a este tipo de estadísticos se les conoce también como de subdivisión poblacional. La interpretación del estadístico es simple, los valores van de 0 a 1, donde 1 es diferenciación

máxima (la diversidad genética de la muestra está representada por las diferencias entre subpoblaciones) y 0 es que no existe diferencia entre las muestras (la diversidad total puede encontrarse representada en cualquiera de las subpoblaciones). Para evaluar la significancia estadística de los resultados se realizó una prueba  $\chi^2$  para cada valor de  $F_{ST}$  obtenido (Hedrick 2000). Usando el software DnaSP (Rozas y Rozas 1999) y bajo el modelo de islas de Wright (1951) y con la intención de evaluar el nivel de diferenciación genética entre sitios dado por la composición de cepas del género *Pseudomonas* se estimó  $F_{ST}$  pareada por locus entre sitios de muestreo. De esta misma manera se estimaron los niveles de diferenciación genética por locus entre los linajes seleccionados ya mencionados (*P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”) y la diferenciación genética por locus en el tiempo excluyendo los sitios MAN9 y MAN10 (no presentan variación estacional en la temperatura, ver Tabla 4.1).

**Pruebas de correlación ambiental/genética.-** Con la intención de establecer posibles correlaciones en la distribución de las distintas variantes genéticas encontradas se realizaron pruebas de Mantel (Smouse *et al.* 1986) entre una matriz de diferenciación ( $F_{ST}$ ) y matrices de diferencias en los parámetros ambientales medidos (temperatura, pH, oxígeno disuelto, conductividad y distancia geográfica) entre los distintos sitios de muestreo. Estas pruebas se realizaron locus por locus. Las matrices de diferencias en los parámetros ambientales se construyeron calculando las distancias euclidianas pareadas utilizando los valores normalizados de cada parámetro ambiental considerado (ver arriba y Hughes *et al.* 2006).

**Análisis de demografía histórica.-** Se exploró la demografía histórica de las poblaciones de los 3 linajes numéricamente más abundantes usando “*classic skyline plots*” (Pybus *et al.* 2000). Este análisis es una representación gráfica de simulaciones de coalescencia que

estiman el tamaño efectivo de una población a través del tiempo. Usamos las secuencias disponibles para cada locus para construir un árbol de Máxima Parsimonia (ML) en PAUP\* (Swofford, 2002). Este árbol fue analizado como ultramétrico en GENIE version 3.0 (Pybus y Rambaut 2002) probando los modelos de población constante y crecimiento exponencial.

### **4.2.3. Resultados**

#### **I. Distribución ecológica de genotipos**

*a. Análisis de diversidad y cluster de los patrones rep.*- Los análisis de patrones rep muestran tres resultados relacionados con la distribución ecológica de las variantes genotípicas: 1) Como puede verse en la Figura 4.2, el número de variantes genotípicas idénticas es mayor en las muestras colectadas en invierno que en el verano y aunque hay cierto grado de agrupamiento de variantes genéticas por sitio y por temporada de colecta, esto no siempre es cierto; 2) Al evaluar posibles correlaciones simples entre matrices de diferencias en los parámetros ambientales y la matriz de diferencia en los patrones de bandeo, no se encontró ninguna correlación evidente (gráficas no presentadas); 3) Las pruebas de Mantel (Tabla 4.4) en general son consistentes con esta falta de patrones de distribución de genotipos asociados a parámetros ambientales, salvo en el caso del parámetro temperatura, ya que, el coeficiente de correlación (0.28), es estadísticamente significativo ( $P < 0.01$ ). Este último resultado es interesante, pues si regresamos al dendrograma de los patrones de bandeo (Figura 4.2) y vemos la posible asociación de diferentes genotipos a distintas épocas del año con temperaturas contrastantes (invierno y verano) no es imposible pensar en una dinámica ecológica diferente en ambas estaciones lo

cual estaría de acuerdo con la correlación de la temperatura y la distribución de distintos genotipos.

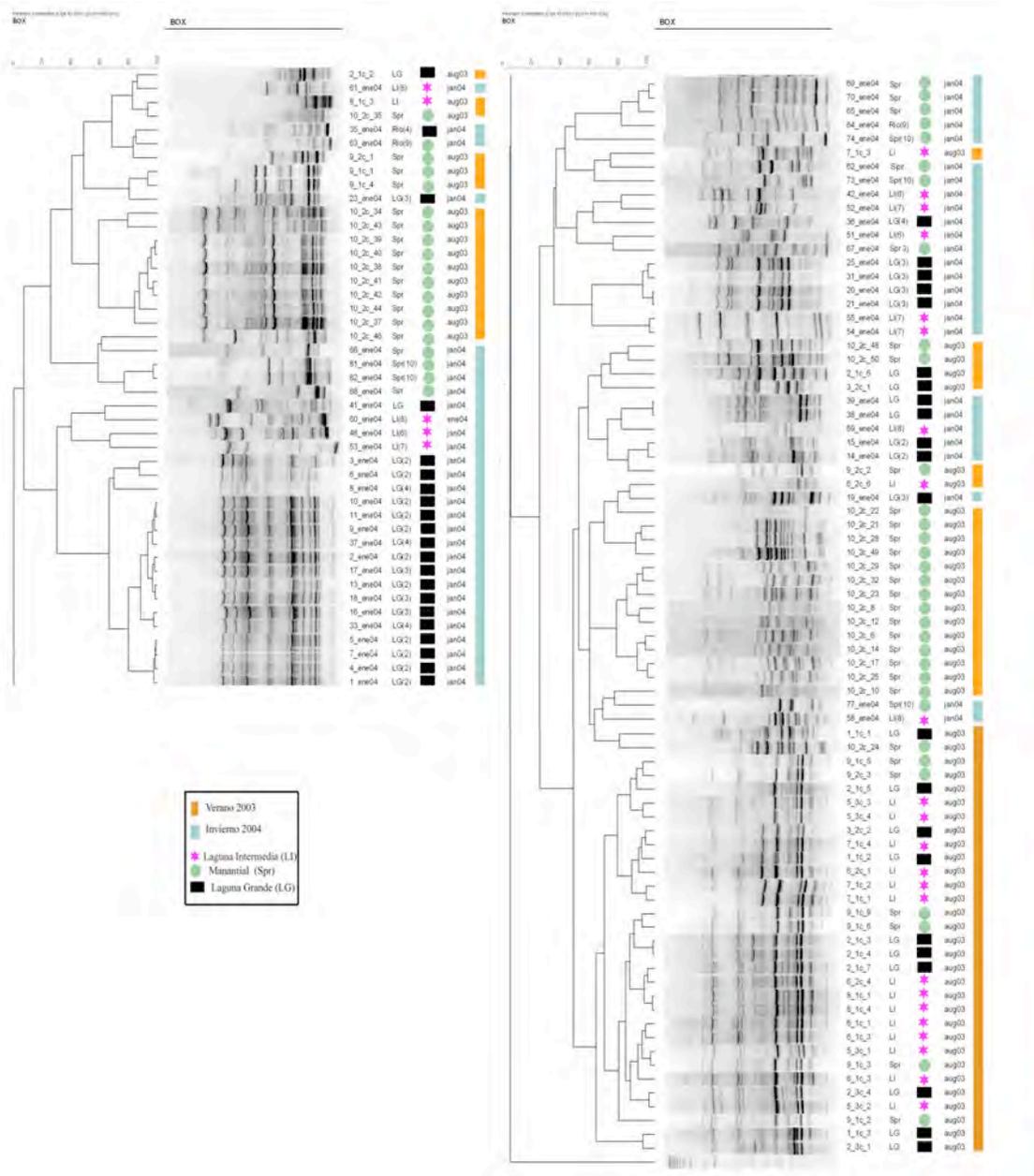


Figura 4.2. Cluster Pearson-UPGMA de los patrones BOX-PCR de las cepas de *Pseudomonas* estudiadas en el sistema Churince en Cuatro Ciénegas Coahuila, México.

Tabla 4.4. Resumen de resultados de las pruebas de Mantel para distancia genética estimada por similitud en los patrones rep-PCR y diferencias en las condiciones ambientales de los sitios de colecta. 1000 permutaciones.

Matrices	P	Coficiente de correlación
Similitud de Pearson vs. Distancia Geográfica <sup>a</sup>	0.285	0.04
Similitud de Pearson vs. Variación en Temperatura <sup>b</sup>	0.093	0.05
Similitud de Pearson vs. Diferencia de Temperatura <sup>b</sup>	0.001	0.28
Similitud de Pearson vs. Diferencia de pH <sup>b</sup>	0.254	0.04
Similitud de Pearson vs. Diferencia de Conductividad <sup>b</sup>	0.689	-0.04
Similitud de Pearson vs. Diferencia de Oxígeno disuelto <sup>b</sup>	0.116	0.1

<sup>a</sup>. La matriz de distancia geográfica consistió distancia euclideana entre pares de sitios en donde se colectó cada cepa.

<sup>b</sup>. Las matrices ambientales utilizadas se calcularon estimando las distancias euclideanas pareadas de los sitios en donde se colectó cada cepa, utilizando los valores normalizados de cada parámetro ambiental considerado (temperatura, pH, conductividad, oxígeno disuelto y variación estacional de la temperatura).

## II. Diversidad filogenética y genética de poblaciones

**a. Relaciones filogenéticas.-** El análisis filogenético de las secuencias de los genes *acnB*, *gyrB*, *recA*, *rpoD* y 16SrRNA tuvo como objetivos, establecer cuáles son los linajes evolutivos que se encuentran más representados en la colección de cepas del género *Pseudomonas*, así como evaluar niveles de clonalidad/recombinación en el interés de conocer la dinámica evolutiva de estos grupos en el área de estudio seleccionada. Los resultados de las reconstrucciones filogenéticas por locus están representados en las genealogías de la Figura 4.3a-e.

En el caso de la genealogía del gen 16SrRNA (Figura 4.3a), se usaron sólo algunas de las cepas de la muestra total, las cuales formaban parte de los linajes independientes en las genealogías de los otros loci, adicionalmente se usaron como grupos de referencia, secuencias obtenidas de bases de datos públicas (Ribosomal Database Project II) y que

corresponden a organismos cultivados y plenamente identificados como representantes o cepas tipo de la especie indicada, esto con la intención de asignar con mayor certeza un nombre específico a las cepas de este estudio. En esta genealogía se observa que las cepas incluidas en este estudio pertenecen efectivamente al grupo I de rRNA de *Pseudomonas* definido por Palleroni (1984), el cual es un grupo monofilético y corresponde a lo que ahora se conoce como *Pseudomonas sensu stricto*. Adicionalmente, y en congruencia con las filogenias del género publicadas (Yamamoto y Harayama 1998, Yamamoto *et al.* 2000), por un lado se observa una bifurcación en dos grupos principales, uno de ellos (Intragenetic Cluster I = IGCI, Yamamoto *et al.* 2000) incluye a *P. aeruginosa* y a sus parientes más cercanos y el otro grupo (IGCII) incluye a *P. putida* y a sus parientes. Ya en esta primera genealogía observamos también que, la colección de cepas del género de este estudio aparece como diversa en especies, existen representantes de al menos 7 linajes dentro del género: *P. mendocina* (IGCI), *P. alcaliphila* (IGCI), *P. psychrotolerans* (IGCI), *P. otitidis* (IGCI), *P. fluorescens* (IGCII), *P. putida* (IGCII), *P. mosselii* (IGCII). Por otro lado, encontramos un linaje de difícil afiliación al que denominaremos “*P. caenicarum*” cuyo más cercano linaje parece ser *P. peli* (ver Apéndice I), así como la existencia de cepas cercanas a linajes o grupos filogenéticamente definidos pero que no quedan anidadas en la reconstrucción del árbol (69ene04, 68ene04, 62ene04 y 10\_2c\_10). En cuanto a “*P. caenicarum*” vale la pena señalar que debido al considerable número de cepas incluidas en este grupo, se tomó la tarea de estudiar sus cualidades genéticas y fenotípicas a mayor detalle y se elaboró un manuscrito en donde se describe formalmente como una especie nueva, el cual puede encontrarse al final de esta tesis como Apéndice I.

En el caso de la genealogía del gen *rpoD* (Figura 4.3e), se incluyeron secuencias de referencia que fueron usadas en la reconstrucción de la filogenia del género (Yamamoto *et*

---

al. 2000) y que están disponibles en GenBank. En este caso encontramos una distribución de linajes o topología muy similar a la descrita anteriormente para el gen del 16SrRNA. Se encuentran representantes de los siguientes linajes: *P. otitidis*, *P. pseudoalcaligenes*, *P. alcaliphila*, "*P. caenicarum*", *P. psychrotolerans*, *P. mosselii* y *P. putida*. En esta genealogía, al igual que en las de, *acnB*, *gyrB* y *recA* (Figuras 4.3b-d), observamos la representación numérica de cada linaje en la colección. De esta manera es evidente que podemos hablar de dos linajes abundantes: *P. otitidis* y *P. psychrotolerans*. Estos dos linajes fueron considerados para análisis particulares en todo este estudio, además del linaje "*P. caenicarum*". "*P. caenicarum*", aunque numéricamente menos abundante que otros linajes como el asociado a *P. pseudoalcaligenes* o *P. alcaliphila* resulta de interés, al ser el menos relacionado con linajes previamente descritos del género, además de que el conjunto de datos para el resto de loci analizados es mayor para "*P. caenicarum*" que para *P. pseudoalcaligenes* o *P. alcaliphila*. La última razón para haber hecho esta selección es que, estos linajes parecen ser consistentemente encontrados como independientes y congruentes en las distintas genealogías. Finalmente, cabe notar que la diversidad de linajes en la colección que muestra el análisis de la genealogía de *rpoD* es consistente con la diversidad representada en la genealogía del gen 16SrRNA, pero además este análisis nos permite observar la poca diversidad que se encuentra dentro de cada linaje representado por el largo de las ramas que, en la mayoría de los casos, es cero o muy cercano a ese valor, lo que habla de que las cepas dentro de cada linaje y recuperadas de este sistema acuático son prácticamente idénticas en gen 16SrRNA.

Para la reconstrucción de la genealogía del gen *gyrB* (Figura 4.3c), también se incluyeron secuencias de referencia que fueron usadas en la reconstrucción de la filogenia del género (Yamamoto *et al.* 2000) y que están disponibles en GenBank. En este caso la

genealogía no presenta topología idéntica a 16SrRNA o *rpoD* a pesar de que, como se mencionó, este gen, junto con *rpoD* fue utilizado para la reconstrucción filogenética del género. La razón de esto, muy posiblemente, es que en el caso del presente estudio no se contó con la secuencia completa del gen, lo cual podría afectar la resolución de la reconstrucción. A pesar del acomodo topológico de los linajes, varios de ellos se distinguen con claridad: *P. otitidis*, *P. psychrotolerans*, "*P. caenicarum*", *P. alcaliphila*, *P. putida*, *P. fluorescens* y *P. mosselii*, sosteniéndose con valores de bootstrap > 50 y en la mayoría de los casos con valores >80. Nuevamente, es de notar la diversidad de linajes en la totalidad de la colección y la poca diversidad al interior de los grupos particulares. Finalmente, la posición de las cepas 10\_2c\_10 y 36ene04, hace pensar en una posible afiliación a otro género, lo cuál tendría que analizarse particularmente.

En los casos de las genealogías de los genes *acnB* (Figura 4.3b) y *recA* (Figura 4.3d) se utilizaron secuencias de referencia tomadas del GenBank, sin embargo la afiliación de estas no es totalmente segura, pues en la mayoría de los casos no corresponden a cepas tipo y no existen disponibles secuencias de estos genes para toda la diversidad de linajes dentro del género.

Para el caso de la genealogía del gen *acnB* (Figura 4.3b), la topología o posiciones relativas de los distintos linajes no parece ser congruente con las genealogías anteriores pero los linajes principales en general se conservan: *P. otitidis*, *P. pseudoalcaligenes/mendocina*, *P. psychrotolerans*, *P. alcaliphila*, "*P. caenicarum*", *P. mosselii*. Cabe notar una vez más, la poca diversidad dentro de los linajes, aunque el largo de las ramas al interior y fuera de ellos es más largo que en el resto de las genealogías, lo que sugiere una tasa de evolución más rápida que el resto de los loci analizados, de alguna forma habrá menos restricciones selectivas en la variabilidad de este gen.

Finalmente, la genealogía del gen *recA* (Figura 4.3d), tampoco es topológicamente congruente con las genealogías de 16SrRNA, *rpoD* y *gyrB*, aunque al igual que en el caso de *acnB*, los linajes principales en general se distinguen con claridad del resto: *P. otitidis*, *P. pseudoalcaligenes/mendocina*, *P. psychrotolerans*, *P. alcaliphila*, “*P. caenicarum*” y *P. mosselii*, los cuales forman grupos poco diversos al interior.

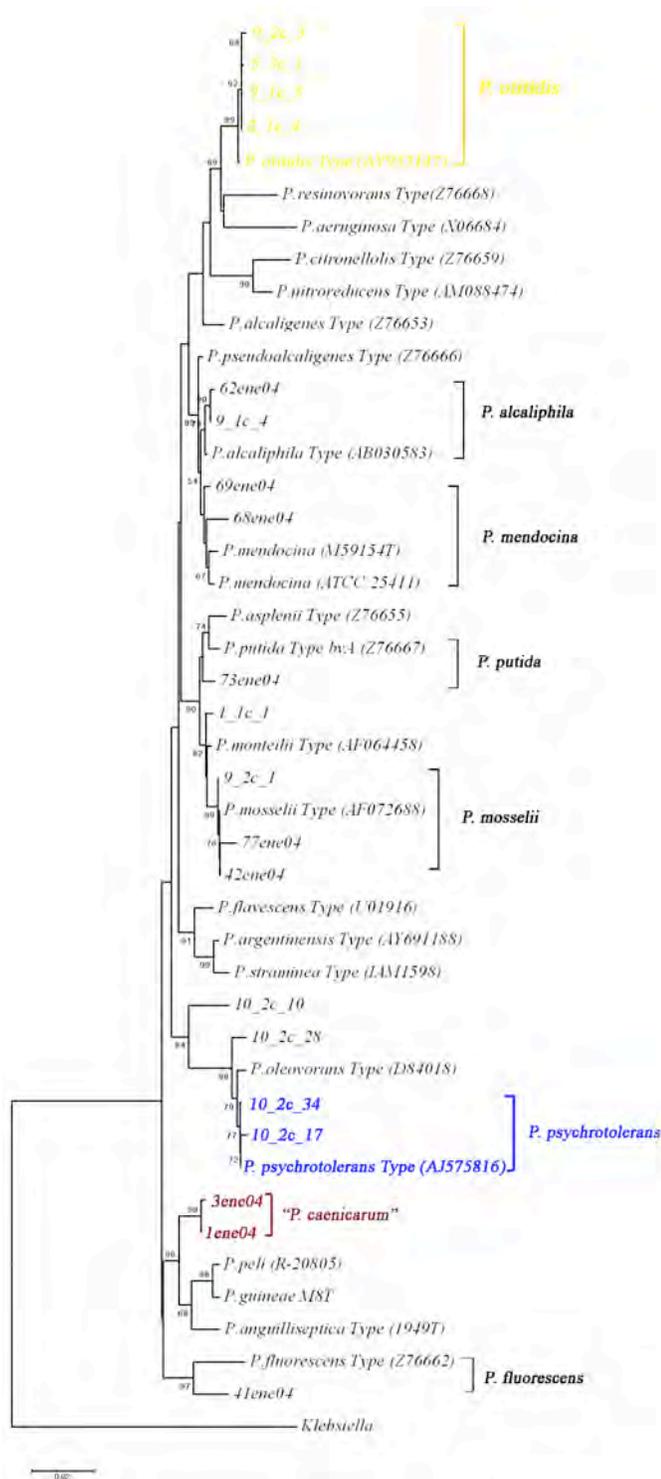


Fig. 4.3a. Genealogía del gen 16S rRNA para cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciénegas Coahuila, México. Se usaron secuencias de referencia de aislados tipo de distintas especies del género. El árbol presentado se construyó bajo un modelo de sustitución Tamura-Nei93 y algoritmo de distancia Neighbor-Joining con 1000 bootstrap con una longitud en el alineamiento de 1560 pb, incluyendo gaps y datos faltantes.

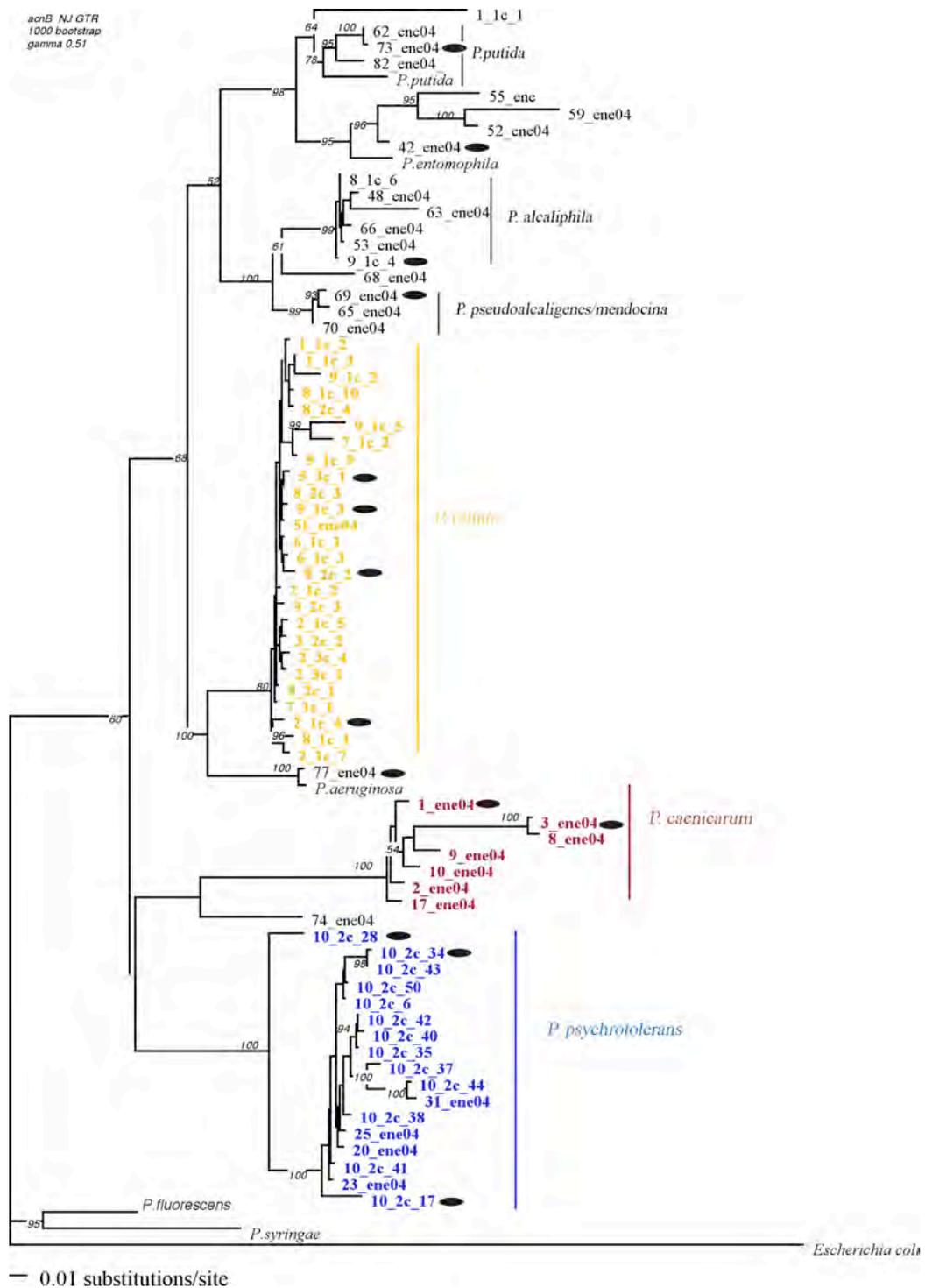


Fig. 4.3b. Genealogía del gen *acnB* para cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciénegas Coahuila, México. Se usaron secuencias de referencia del género tomadas del GenBank. El árbol presentado se construyó bajo un modelo de sustitución GTR y algoritmo de distancia Neighbor-Joining con 1000 bootstrap con una longitud en el alineamiento de 624 pb, incluyendo gaps y datos faltantes. Un círculo negro se refiere a que se cuenta con secuencia completa del gen 16S rRNA para la cepa señalada.

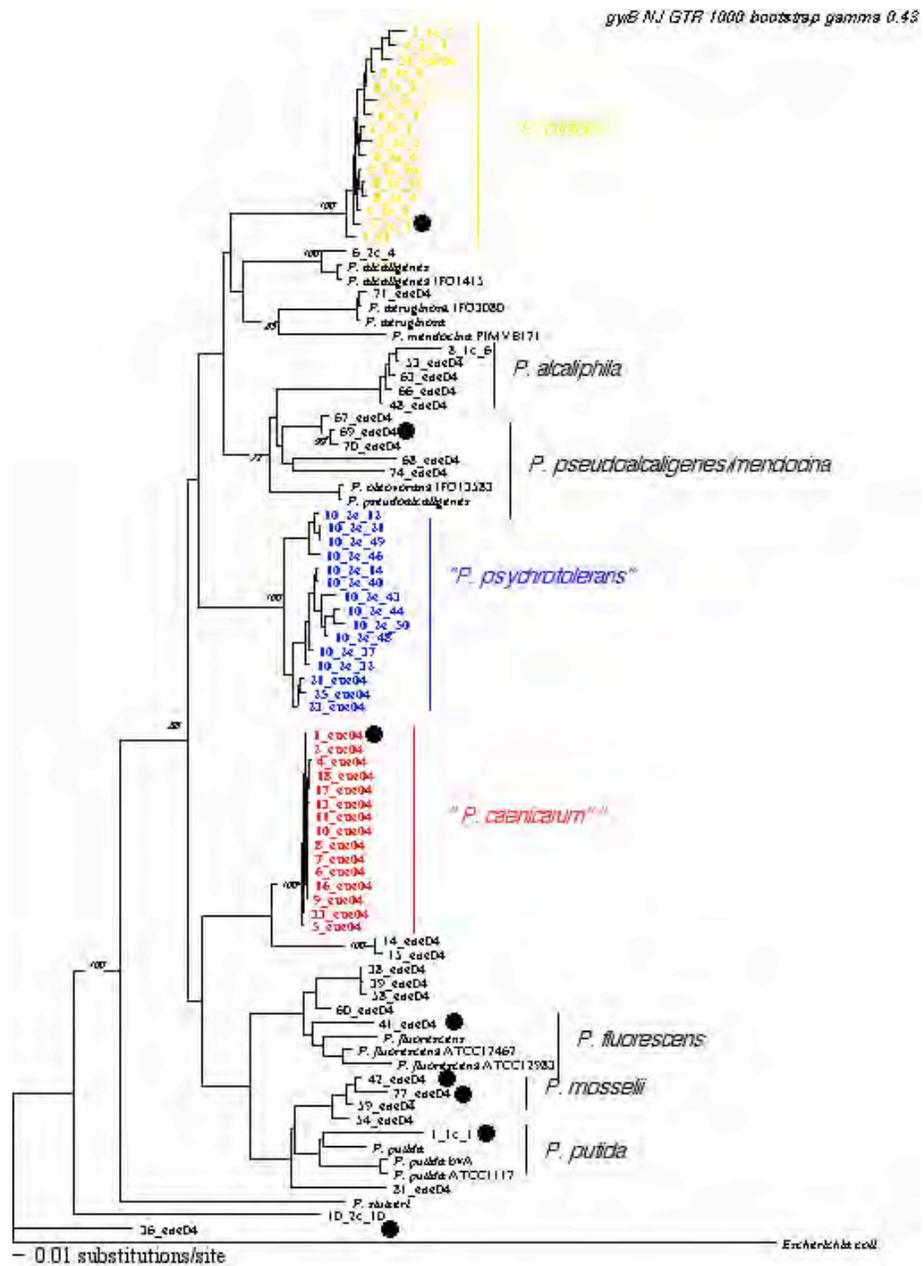


Fig. 4.3c. Genealogía del gen *gyrB* para cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciénegas Coahuila, México. Se usaron secuencias de referencia del género tomadas del GenBank y usadas por Yamamoto *et al.* (2000). El árbol presentado se construyó bajo un modelo de sustitución GTR y algoritmo de distancia Neighbor-Joining con 1000 bootstrap con una longitud en el alineamiento de 843 pb, incluyendo gaps y datos faltantes. Un círculo negro se refiere a que se cuenta con secuencia completa del gen 16S rRNA para la cepa señalada.

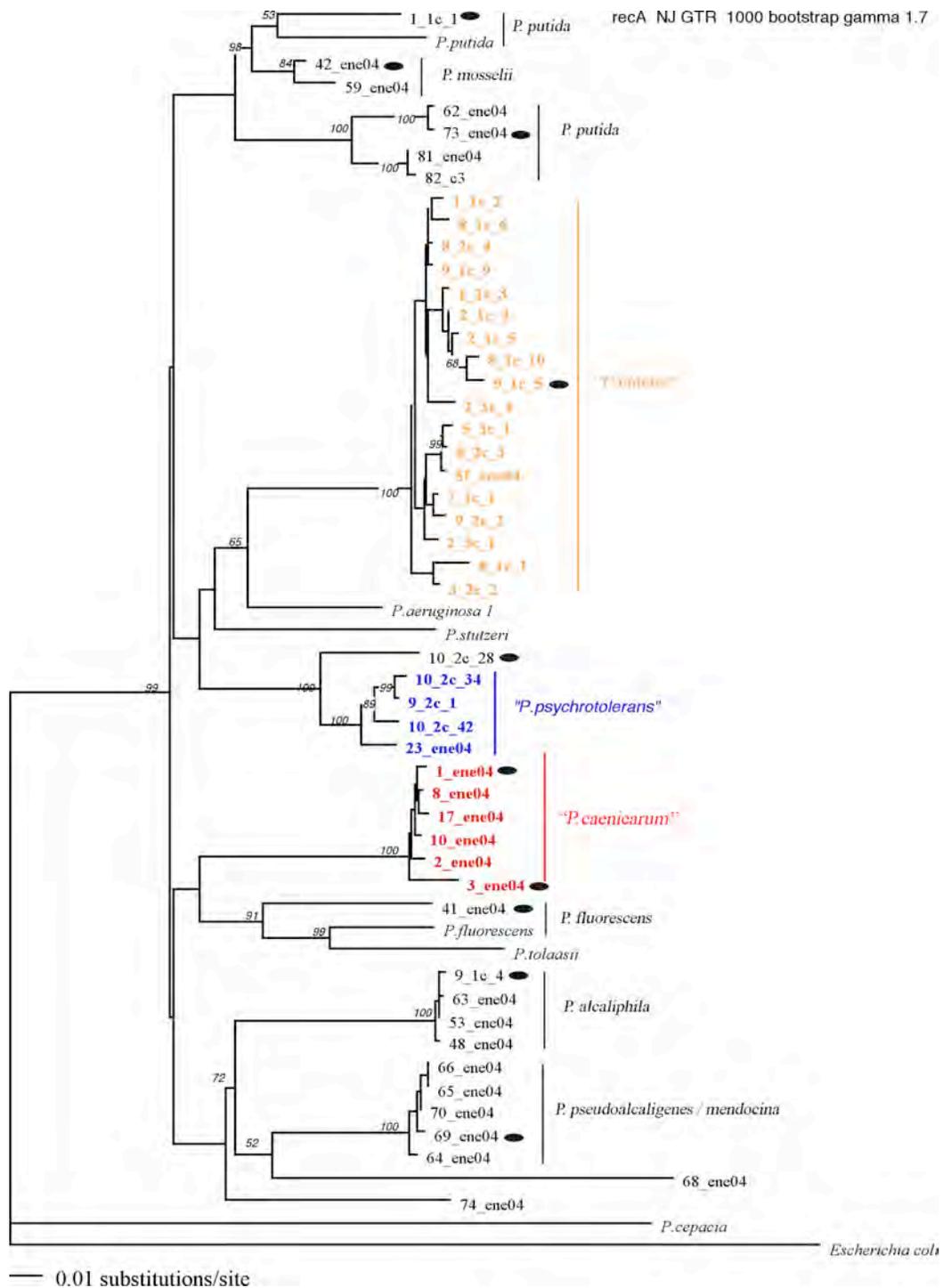


Fig. 4.3d. Genealogía del gen *recA* para cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciéngas Coahuila, México. Se usaron secuencias de referencia del género tomadas del GenBank. El árbol presentado se construyó bajo un modelo de sustitución GTR y algoritmo de distancia Neighbor-Joining con 1000 bootstrap con una longitud en el alineamiento de 843 pb, incluyendo gaps y datos faltantes. Un círculo negro se refiere a que se cuenta con secuencia completa del gen 16S rRNA para la cepa señalada.

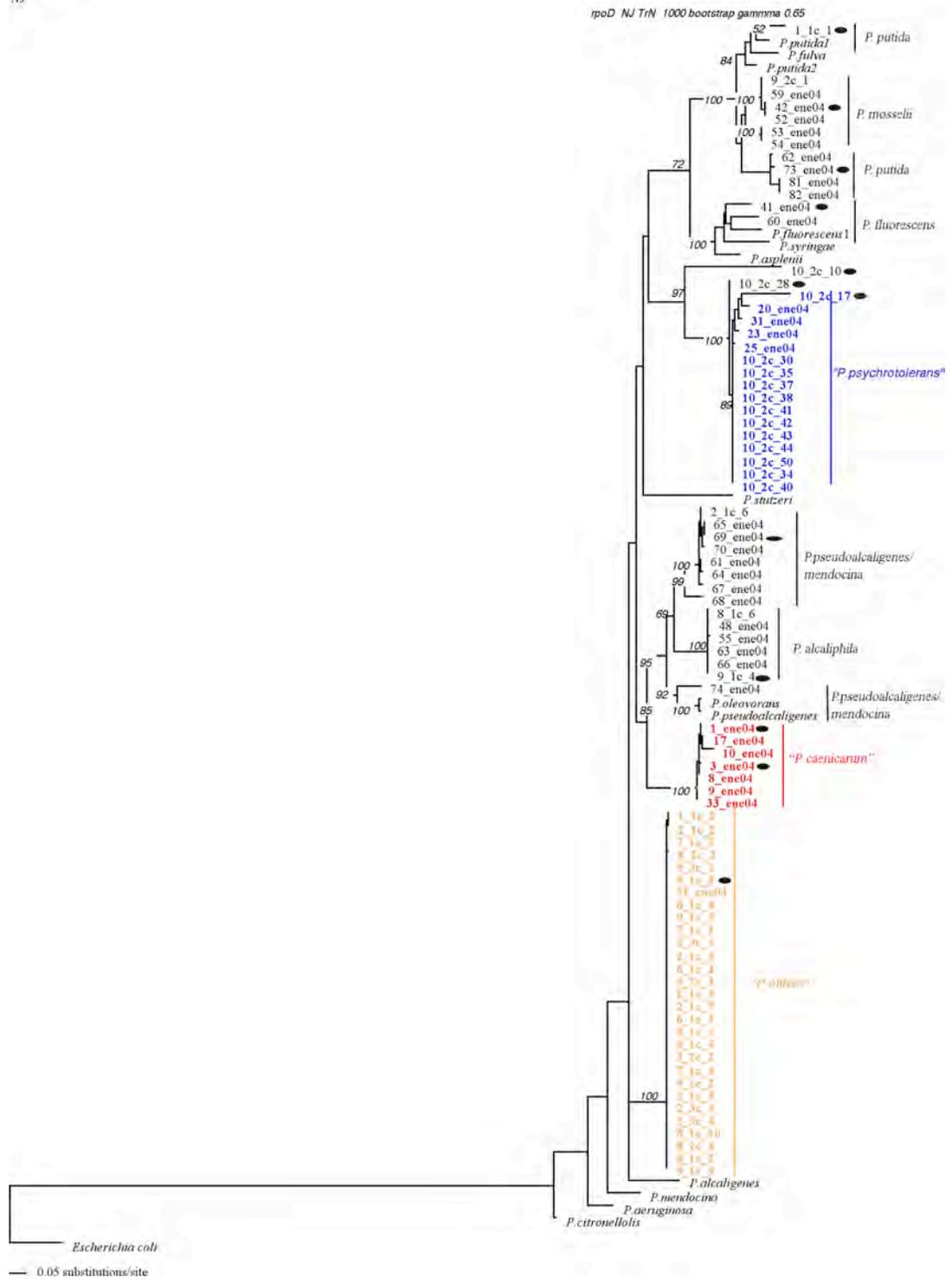


Fig. 4.3e. Genealogía del gen *rpoD* para cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciénegas Coahuila, México. Se usaron secuencias de referencia del género tomadas del GenBank y usadas por Yamamoto *et al.* (2000). El árbol presentado se construyó bajo un modelo de sustitución TrN y algoritmo de distancia Neighbor-Joining con 1000 bootstrap con una longitud en el alineamiento de 618 pb, incluyendo gaps y datos faltantes. Un círculo negro se refiere a que se cuenta con secuencia completa del gen 16S rRNA para la cepa señalada.

Análisis para evaluar recombinación

I. ILD.- Este análisis se realizó para evaluar de manera formal la congruencia de las diferentes genealogías y al mismo tiempo establecer la posibilidad de historias evolutivas diferentes dentro de lo que aparecen en las genealogías como linajes evolutivos claros. Los resultados de esta prueba se muestran en la Tabla 4.5. Como se muestra allí, la prueba se realizó tomando a toda la colección de cepas y también por linajes individuales, estos fueron *P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”. En el caso de las comparaciones que incluyen todas las cepas de la colección, todas resultan incongruentes. Esto puede ser debido a que se está tratando con distintas historias evolutivas al mezclar linajes diferentes. Las comparaciones por grupo en general resultan congruentes lo cual indicaría que todos los genes tienen historias evolutivas iguales o muy parecidas, esto nos hablaría de alto grado de clonalidad en cada uno de los linajes analizados y al mismo tiempo apoya la idea de linajes auténticos.

Tabla 4.5. Resultados del test de incongruencia filogenética para todas las cepas de la colección y tres linajes. 100 réplicas, 10,000 árboles, Branch and Bound Parsimonia

Todas las cepas	<i>rpoD</i>	<i>recA</i>	<i>gyrB</i>	<i>acnB</i>
<i>rpoD</i>	<b>1.00</b>			
<i>recA</i>	0.01 (39)	<b>1.00</b>		
<i>gyrB</i>	0.01 (36)	0.01 (24)	<b>1.00</b>	
<i>acnB</i>	0.01 (53)	0.01 (36)	0.01 (32)	<b>1.00</b>

<i>P. otitidis</i>	<i>rpoD</i>	<i>recA</i>	<i>gyrB</i>	<i>acnB</i>
<i>rpoD</i>	<b>1.00</b>			
<i>recA</i>	0.01 (13)	<b>1.00</b>		
<i>gyrB</i>	<b>0.32 (6)</b>	<b>1.00 (4)</b>	<b>1.00</b>	
<i>acnB</i>	0.01 (18)	<b>0.12 (13)</b>	<b>1.00 (6)</b>	<b>1.00</b>

Tabla 4.5. continuación

<i>P. psychrotolerans</i>	<i>rpoD</i>	<i>recA</i>	<i>gyrB</i>	<i>acnB</i>
<i>rpoD</i>	<b>1.00</b>			
<i>recA</i>	NP (2)	<b>1.00</b>		
<i>gyrB</i>	<b>1.00 (7)</b>	NP (1)	<b>1.00</b>	
<i>acnB</i>	0.01 (12)	NP (1)	<b>1.00 (7)</b>	<b>1.00</b>

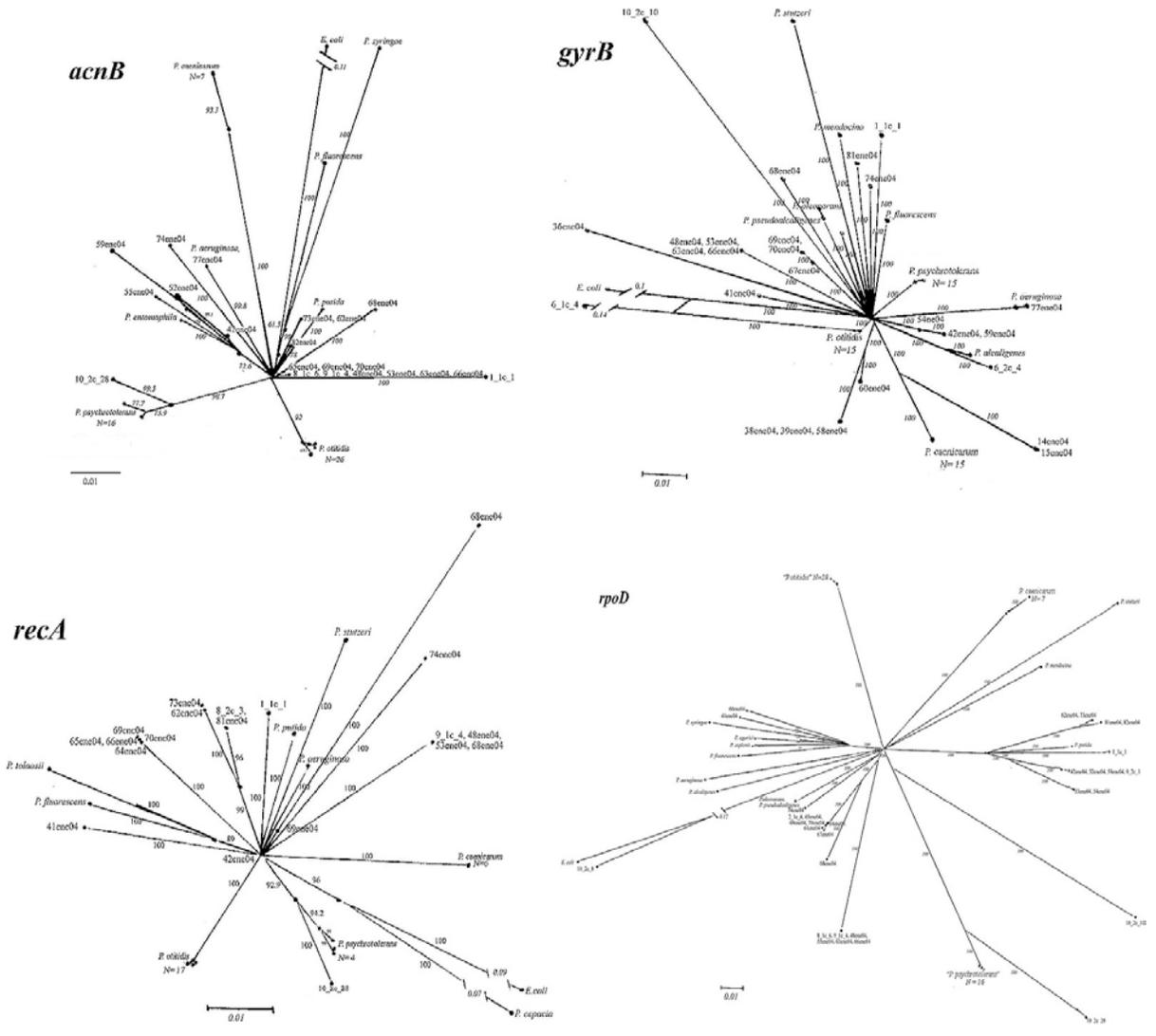
  

<i>“P. caenicarum”</i>	<i>rpoD</i>	<i>recA</i>	<i>gyrB</i>	<i>acnB</i>
<i>rpoD</i>	<b>1.00</b>			
<i>recA</i>	<b>1.00 (4)</b>	<b>1.00</b>		
<i>gyrB</i>	<b>1.00 (5)</b>	<b>1.00 (5)</b>	<b>1.00</b>	
<i>acnB</i>	<b>1.00 (5)</b>	<b>1.00 (6)</b>	<b>1.00 (6)</b>	<b>1.00</b>

NP = No fue posible la prueba porque el número de secuencias disponibles era < 3.

II. Split-Decomposition.- Otra manera de estudiar posibles incongruencias e inferir recombinación entre o dentro de linajes, es la reconstrucción en red de todas las posibles relaciones entre secuencias de DNA. Este análisis se realizó por locus individuales para toda la muestra y tomando en consideración solamente los tres linajes mencionados: *P. otitidis*, *P. psychrotolerans* y *“P. caenicarum”*. El resultado gráfico de los análisis se muestra en la Figura 4.4a-c. En el caso del análisis de toda la muestra para los cuatro loci (*acnB*, *gyrB*, *recA* y *rpoD*), en general no se aprecian eventos de recombinación entre cepas o linajes, salvo dos casos. El primer caso en el análisis de *acnB* se aprecia un posible evento de recombinación entre dos cepas de posible afiliación *P. mosselli* (52ene04 y 59ene04). El segundo caso en el análisis de *gyrB* nuevamente la cepa 6\_1c\_4 que en la genealogía de *gyrB* se comporta de manera diferente al resto de las cepas del linaje en donde se encuentra, aparece con un evento de recombinación con

*Escherichia coli*. Para el caso del análisis incluyendo sólo los tres linajes mencionados, *rpoD* y *recA* muestran una distribución perfecta de las cepas en tres linajes independientes sin eventos de recombinación dentro o entre linajes. El caso de *gyrB* muestra los tres linajes mencionados y la cepa 6\_1c\_4 como un linaje aparte. Finalmente *acnB* también muestra la separación de los tres linajes pero al igual que la genealogía de este gen se observa más diversidad dentro de linajes y un posible evento de recombinación en *P. otitidis*.



. Fig. 4.4a. Análisis de Split-Decomposition para todas las cepas de *Pseudomonas* colectadas en el sistema Churince en Cuatro Ciéngas Coahuila, México. Se presenta el análisis para cada locus: *acnB*, *gyrB*, *recA* y *rpoD*. En todos los casos se usaron secuencias de referencia del género tomadas del GenBank y para los locus *gyrB* y *rpoD* se incluyeron las secuencias de referencia usadas por Yamamoto *et al.* 2000.

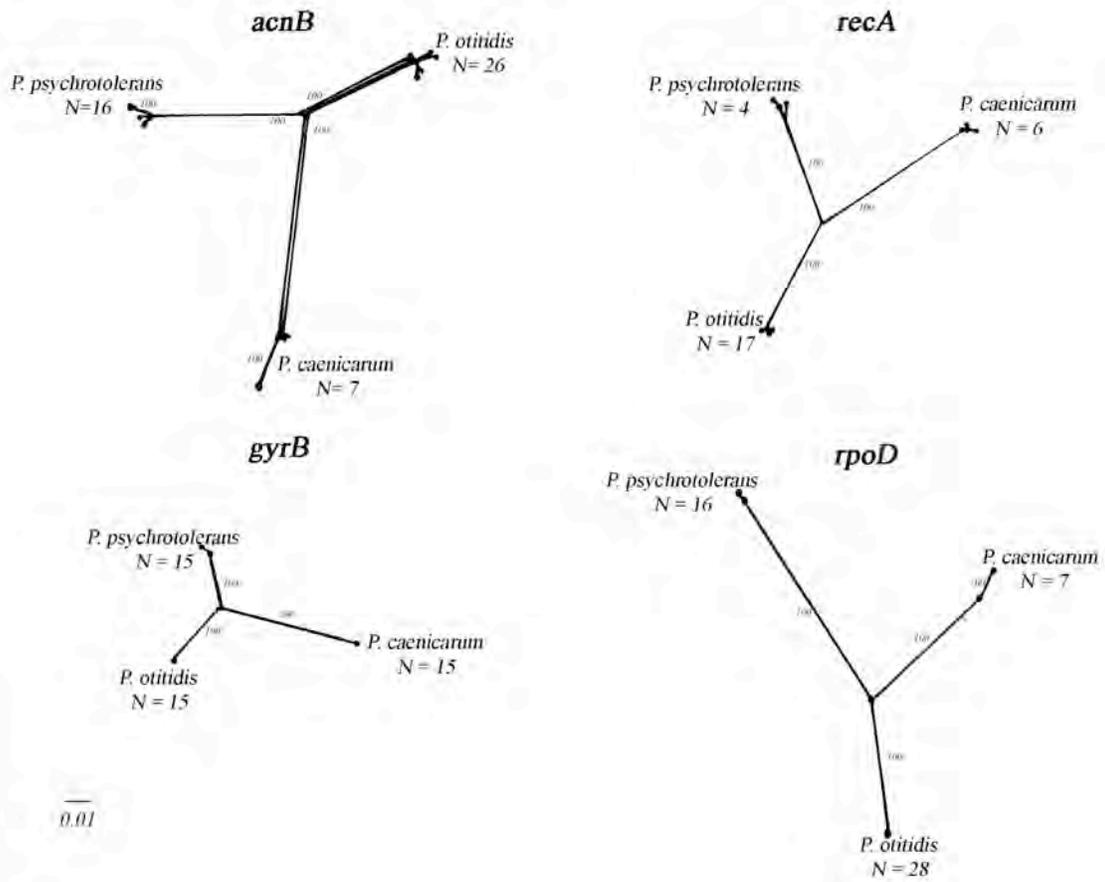


Fig. 4.4b. Análisis de Split-Decomposition de los loci *acnB*, *gyrB*, *recA* y *rpoD* para cepas de *Pseudomonas* de los linajes *P. otitidis*, *P. psychrotolerans* y "*P. caenicarum*" colectadas en Cuatro Ciénegas Coahuila, México.

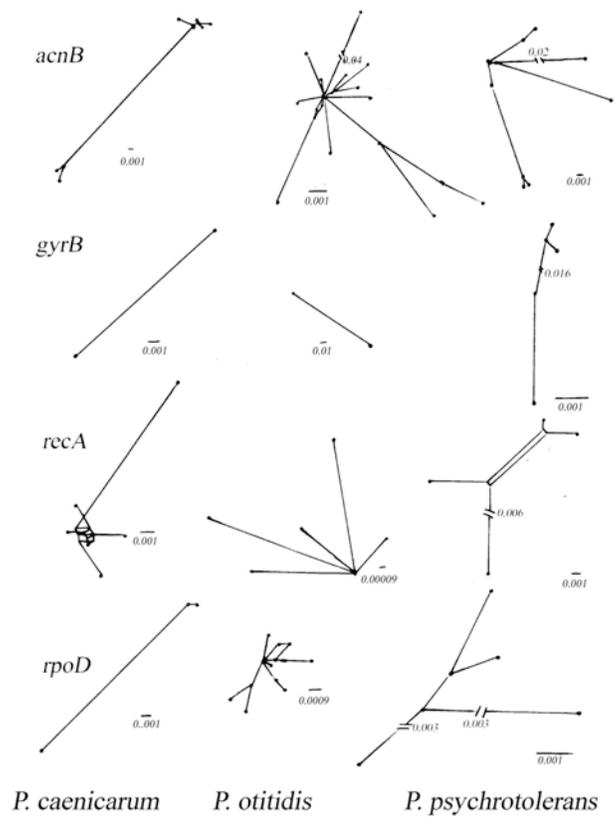


Fig. 4.4c. Análisis de Split Decomposition de los loci *acnB*, *gyrB*, *recA* y *rpoD* para cada uno de los linajes numéricamente dominantes: *P. otitidis*, *P. psychrotolerans* y "*P. caenicarum*" colectadas en el sistema Churince en Cuatro Ciénegas Coahuila, México.

III. Análisis de recombinación tipo GARD.- Para evaluar recombinación y niveles de clonalidad en nuestra muestra se realizó un análisis para detectar falsos positivos de selección molecular. Los resultados de la prueba se muestran en la Tabla 4.6. Para todos los casos se encontraron de cero a máximo tres posibles sitios de recombinación, siendo lo más común un sitio. Este resultado es interesante y resulta congruente con la señal de clonalidad dada por ILD. Es de notar el caso de “*P. caenicarum*” que no muestra señal de recombinación en 2 de los cuatro loci. Para el caso del gen *rpoD* no se encontró recombinación en ningún caso.

Tabla 4.6. Resultados de la prueba de recombinación para un sitio y sitios múltiples. Se presentan los resultados para 4 loci para toda la colecta de cepas de *Pseudomonas* y para tres linajes (*P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”)

Locus	Población	Breakpoints Single Point	AICc	Breakpoints GARD	$\Delta$ AICc
<i>acnB</i>	Todas	1 a 602pb	10328.3	1 a 432pb	75.24
	<i>P. otitidis</i>	1 a 664pb	3805.0	1 a 366pb	31.7
	<i>P. psychrotolerans</i>	1 a 533 pb	2418.35	1 a 630pb	84.2
	“ <i>P. caenicarum</i> ”	1 a 533 pb	2414.4	1 a 534pb	97.4
<i>gyrB</i>	Todas	1 a 545pb	23296.2	1 a 561pb	713.7
	<i>P. otitidis</i>	1 a 566pb	7651.57	3=1 196,567,886pb	?
	<i>P. psychrotolerans</i>	1 a 923pb	5092.5	2= 210, 846pb	?
	“ <i>P. caenicarum</i> ”	No	-	No	-
<i>recA</i>	Todas	1 a 42 pb	11953.8	1 a264pb	311.6
	<i>P. otitidis</i>	1 a 32pb	3810.5	1 a 97pb	368.1
	<i>P. psychrotolerans</i>	1 a 409pb	2667.56	2= 144, 410pb	4.5
	“ <i>P. caenicarum</i> ”	1 a 30pb	2761.0	1 a 31	106.3
<i>rpoD</i>	Todas	No	-	No	-
	<i>P. otitidis</i>	No	-	No	-
	<i>P. psychrotolerans</i>	No	-	No	-
	“ <i>P. caenicarum</i> ”	No	-	No	-

IV. Comparación de topologías filogenéticas.- Los resultados del análisis de recombinación entre genes fueron todos negativos, es decir, no se encontró evidencia de recombinación entre genes.

*Niveles de variación genética, pruebas de selección y estructura genética de las poblaciones.-* Con la intención de conocer con más detalle la dinámica evolutiva y ecológica de las poblaciones de *Pseudomonas* en el sistema Churince se estimaron los niveles de variación genética, se realizaron pruebas de selección a nivel molecular y de estructura genética, para todos los casos se consideró cada locus a nivel de toda la muestra, por linaje y en las dos muestras estacionales.

*Variación genética.-* Los resultados de este análisis se encuentran en la Tabla 4.7 y muestran que en la muestra total, los niveles de polimorfismo estimados como  $\pi$  y  $\theta$ , así como el porcentaje de sitios polimórficos no son constantes en magnitud para los distintos loci. Finalmente a nivel de linajes particulares, existe una notable caída en la magnitud de los niveles de diversidad al interior de los grupos, aunque los valores de  $\pi$  y  $\theta$  se muestran constantes en todos los loci para cada linaje y  $\pi$  y  $\theta$  en prácticamente todos los casos tienen los mismos valores, lo cual es indicio de evolución neutra en los loci estudiados.

*Selección a nivel molecular.-* Con la intención de evaluar posibles señales de selección en nuestra muestra y plantear hipótesis sobre la dinámica poblacional, se realizaron tres pruebas de selección a nivel molecular que toman en cuenta la distribución

de los polimorfismos: D de Tajima, prueba de Fu-Li, y la prueba de MacDonald-Kreitman. En el primer y segundo casos, prácticamente ninguna prueba resultó significativa pero es notable ver que la mayoría de los valores son negativos, lo cual sería indicio de selección purificadora, es decir, en purga de la diversidad a nivel de secuencias, además de que (salvo un caso) los resultados de ambas pruebas son congruentes (Tabla 4.7). Por otro lado, la prueba de MacDonald-Kreitman (Tabla 4.8) que evalúa diferencias en presiones de selección entre poblaciones muestra que no hay diferencias en presiones de selección para las muestras de invierno y verano pero sí las hay para el gen *gyrB* en todas las comparaciones entre linajes, y entre *P. psychrotolerans* y *P. otitidis* para *rpoD*. Esta situación es notable pues de alguna manera contradice la importante señal de clonalidad dada por los análisis anteriores, ya que si efectivamente se tratara de linajes completamente clonales, la señal de selección sería la misma para todos los marcadores. Finalmente, se evaluó la relación entre sustituciones sinónimas y no sinónimas, en la mayoría de los casos se encontró señal de selección negativa o purificadora, lo cual está de acuerdo con las pruebas de selección antes mencionadas en donde se analizó la distribución de los polimorfismos (Tabla 4.7).

Tabla 4.7. Diversidad genética por locus de los tres linajes evolutivos principales de *Pseudomonas* aislados del Sistema Churince en Cuatro Ciénegas, Coahuila.

Locus	Muestra	N	Longitud (pb) <sup>a</sup>	Haplotipos N	Sitios polimórficos (%)	$\pi$	$\theta$	D Tajima	D Fu Li	dN ( $\pm$ SD)	dS ( $\pm$ SD)	dN/dS
<i>acnB</i>	Toda la muestra	70	501	55	214 (78)	0.117	0.119	-0.045	-0.550	0.0743 (0.046)	0.3310 (0.172)	0.225
	Verano	41	510	34	166 (33)	0.085	0.091	-0.227	-0.865	0.0562 (0.050)	0.2277 (0.189)	0.247
	Invierno	29	558	23	214 (39)	0.141	0.127	0.429	0.814	0.0899 (0.052)	0.4160 (0.161)	0.216
	" <i>P. otitidis</i> "	26	576	23	48 (8)	0.011	0.022	-	-	0.0130 (0.008)	0.0060 (0.007)	2.210
	" <i>P. psychrotolerans</i> "	16	588	14	42 (7)	0.013	0.022	1.879*	2.577*	0.0050 (0.004)	0.0359 (0.039)	0.149
	" <i>P. caenicarum</i> "	7	624	7	33 (5)	0.022	0.022	0.225	0.708	0.0200 (0.019)	0.0323 (0.028)	0.614
<i>gyrB</i>	Toda la muestra	72	762	49	543 (71)	0.157	0.217	-0.969	-0.662	0.0656 (0.031)	0.6160 (0.3198)	0.107
	Verano	30	762	22	509 (67)	0.150	0.222	-1.254	-0.978	0.1250 (0.210)	0.5510 (0.481)	NP
	Invierno	42	825	27	347 (42)	0.133	0.143	-0.247	-0.118	0.0584 (0.031)	0.5960 (0.308)	0.098
	" <i>P. otitidis</i> "	15	777	11	247 (32)	0.048	0.099	-	-3.19*	0.0574 (0.144)	0.0660 (0.090)	NP
	" <i>P. psychrotolerans</i> "	15	825	12	66 (8)	0.030	0.025	2.309*	0.910	0.0066 (0.004)	0.1154 (0.079)	0.057
	" <i>P. caenicarum</i> "	15	843	2	2 (<1)	0.0003	0.0007	-1.491	-1.873	0.0002 (0.0005)	0.0007 (0.002)	0.320
<i>recA</i>	Toda la muestra	48	393	35	125 (32)	0.092	0.095	-0.119	0.0113	0.0196 (0.011)	0.4534 (0.239)	0.043
	Verano	23	501	19	120 (24)	0.060	0.077	-0.915	-1.068	0.0191 (0.020)	0.2520 (0.239)	0.076
	Invierno	25	687	20	222 (32)	0.111	0.119	-0.285	-	0.0198 (0.009)	0.5750 (0.266)	0.035
	" <i>P. otitidis</i> "	17	537	14	27 (5)	0.011	0.015	-0.969	-1.812	0.0020 (0.002)	0.0430 (0.027)	0.041
	" <i>P. psychrotolerans</i> "	4	846	4	26 (3)	0.0298	0.344	-1.011	-0.873	0.0050 (0.002)	0.1160 (0.085)	0.041
	" <i>P. caenicarum</i> "	6	843	6	21 (3)	0.00996	0.01299	-1.467	-1.284	0.0080 (0.008)	0.0166 (0.006)	0.474
<i>rpoD</i>	Toda la muestra	83	480	36	374 (78)	0.2311	0.2614	-0.401	-1.456	0.2290 (0.151)	0.5850 (0.286)	0.391
	Verano	49	480	19	363 (75)	0.1812	0.2572	-1.083	-1.994	0.1879 (0.184)	0.4423 (0.358)	0.425
	Invierno	34	627	21	350 (56)	0.2300	0.2140	0.287	0.764	0.2084 (0.120)	0.6174 (0.236)	0.338
	" <i>P. otitidis</i> "	28	513	9	9 (2)	0.0034	0.0045	-0.767	-0.343	0.0008 (0.001)	0.0121 (0.009)	NP
	" <i>P. psychrotolerans</i> "	16	651	5	12 (2)	0.0043	0.0056	-0.889	-1.568	0.0006 (0.001)	0.0160 (0.0176)	NP
	" <i>P. caenicarum</i> "	7	618	3	16 (3)	0.0074	0.0106	-	-	0.0062 (0.010)	0.0113 (0.0152)	NP

<sup>a</sup> pares de bases usados para el análisis sin tomar en cuenta gaps y "missing data".

NP= el análisis no fue posible, debido a la gran varianza en la estimación de los parámetros dN y dS

Tabla 4.8. Resultados de la prueba McDonald-Kreitman para probar selección en parejas de datos.

Locus	Comparación	Tipo de sustitución	Sitios fijos <sup>a</sup>	Sitios polimórficos <sup>b</sup>	Resultado*	P de Fisher <sup>c</sup>	P de G <sup>d</sup>
<i>acnB</i>	verano/invierno	Sinónima	0	124	-	-	-
		No sinónima	0	52			
	<i>P. otitidis/P. psychrotolerans</i>	Sinónima	34	36	f>p	NS	NS
		No sinónima	30	26			
	<i>P. otitidis/P. caenicarum</i>	Sinónima	43	23	f<p	NS	NS
		No sinónima	43	24			
<i>P. psychrotolerans/P. caenicarum</i>	Sinónima	51	32	f<p	NS	NS	
	No sinónima	41	29				
<i>gyrB</i>	verano/invierno	Sinónima	0	182	-	-	-
		No sinónima	0	99			
	<i>P. otitidis/P. psychrotolerans</i>	<b>Sinónima</b>	<b>47</b>	<b>107</b>	<b>f&lt;p</b>	<b>P&lt;0.01</b>	<b>P&lt;0.01</b>
		<b>No sinónima</b>	<b>32</b>	<b>169</b>			
	<i>P. otitidis/P. caenicarum</i>	<b>Sinónima</b>	<b>66</b>	<b>69</b>	<b>f&lt;p</b>	<b>P&lt;0.001</b>	<b>P&lt;0.001</b>
		<b>No sinónima</b>	<b>39</b>	<b>164</b>			
<i>P. psychrotolerans/P. caenicarum</i>	<b>Sinónima</b>	<b>57</b>	<b>54</b>	<b>f&gt;p</b>	<b>P&lt;0.001</b>	<b>P&lt;0.001</b>	
	<b>No sinónima</b>	<b>43</b>	<b>10</b>				
<i>recA</i>	verano/invierno	Sinónima	0	134	-	-	-
		No sinónima	0	15			
	<i>P. otitidis/P. psychrotolerans</i>	Sinónima	29	46	f>p	NS	NS
		No sinónima	13	8			
	<i>P. otitidis/P. caenicarum</i>	Sinónima	48	25	f>p	NS	NS
		No sinónima	12	6			
<i>P. psychrotolerans/P. caenicarum</i>	Sinónima	72	54	f<p	NS	NS	
	No sinónima	11	12				
<i>rpoD</i>	verano/invierno	Sinónima	0	61	-	-	-
		No sinónima	0	43			
	<i>P. otitidis/P. psychrotolerans</i>	<b>Sinónima</b>	<b>55</b>	<b>14</b>	<b>f&gt;p</b>	<b>P&lt;0.01</b>	<b>P&lt;0.001</b>
		<b>No sinónima</b>	<b>80</b>	<b>3</b>			
	<i>P. otitidis/P. caenicarum</i>	Sinónima	52	12	f>p	NS	NS
		No sinónima	52	9			
<i>P. psychrotolerans/P. caenicarum</i>	Sinónima	67	16	f>p	NS	NS	
	No sinónima	94	10				

\*f = cociente de sitios fijos = sitios no sinónimos fijos / sitios sinónimos fijos.

\* p = cociente de sitios polimórficos = sitios no sinónimos polimórficos / sitios sinónimos polimórficos.

<sup>a</sup> Sitios fijos = iguales dentro de las poblaciones, diferentes entre poblaciones.

<sup>b</sup> Sitios polimórficos = el resto de los sitios variables.

<sup>c</sup> Valores de P para la prueba de Fisher, la cual evalúa si existen diferencias estadísticamente significativas entre los cocientes de los sitios fijos y los sitios polimórficos.

<sup>d</sup> Valores de la prueba de G (con la corrección de Williams para continuidad (Sokal y Rohlf 1981) para probar la hipótesis nula de que la proporción de sustituciones no sinónimas es independiente de si éstas son fijas o polimórficas.

*Análisis de diferenciación genética.*- Se analizaron los niveles de diferenciación genética dentro de la muestra dados por la estacionalidad y el sitio de colecta. Además

como una prueba adicional a los análisis filogenéticos anteriores se evaluó la diferenciación entre linajes como una medida de aislamiento histórico. Los resultados del análisis de diferenciación en la muestra debida a la estacionalidad se encuentran en la Tabla 4.9. Los datos sugieren que existe aislamiento de grupos presentes en invierno y en verano, pues los valores de  $F_{ST}$  son considerables y significativos cuando las cepas de los sitios estacionalmente estables en los parámetros ambientales son excluidas del análisis (MAN9 y MAN10) (Tabla 4.11).

Tabla 4.9. Diferenciación genética entre las poblaciones de verano e invierno

Locus	Población 1	Población 2	$F_{ST}$
<i>acnB</i>	Verano	Invierno	0.19
	Verano (sin MAN9 y MAN10)	Invierno (sin MAN9 y MAN10)	0.4
<i>gyrB</i>	Verano	Invierno	0.19
	Verano(sin MAN9 y MAN10)	Invierno(sin MAN9 y MAN10)	0.33
<i>recA</i>	Verano	Invierno	0.32
	Verano (sin MAN9 y MAN10)	Invierno (sin MAN9 y MAN10)	0.48
<i>rpoD</i>	Verano	Invierno	0.18
	Verano (sin MAN9 y MAN10)	Invierno (sin MAN9 y MAN10)	0.37

Por otro lado, y en contraste con la diferenciación estacional de la muestra, la diferenciación asociada al sitio de colecta, parece no ser tan clara, aunque los valores son igualmente significativos como lo muestran las tablas 4.10 y 4.11.

Tabla 4.10. Valores de diferenciación genética por pares de sitios de colecta en el sistema Churince.

Locus: <i>acnB</i>	$F_{ST}$		
	LG1,2,3	LI6,7,8	MAN9,10
LG1,2,3	0		
LI6,7,8	0.17	0	
MAN9,10	0.11	0.18	0

Locus: <i>gyrB</i>	$F_{ST}$		
	LG1,2,3	LI6,7,8	MAN9,10
LG1,2,3	0		
LI6,7,8	0.24	0	
MAN9,10	0.29	0.12	0

Locus: <i>recA</i>	$F_{ST}$		
	LG1,2,3	LI6,7,8	MAN9,10
LG1,2,3	0		
LI6,7,8	0.10	0	
MAN9,10	0.15	0.17	0

Locus: <i>rpoD</i>	$F_{ST}$		
	LG1,2,3	LI6,7,8	MAN9,10
LG1,2,3	0		
LI6,7,8	0.09	0	
MAN9,10	0.10	0.15	0

Tabla 4.11. Valores de la prueba  $X^2$  para cada valor de  $F_{ST}$ .

Locus	Población 1	N	k alelos	Población 2	N	k alelos	$F_{ST}$	$X^2=2NF_{ST}(k-1)^a$	gl <sup>b</sup>	Valor de P <sup>c</sup>
<i>acnB</i>	Verano	41	34	Invierno	29	22	0.19	1463	55	0.0000
	Verano-9y10	21	20	Invierno-9y10	18	17	0.4	1123.2	36	0.0000
	<i>P. otitidis</i>	26	22	<i>P. psychrotolerans</i>	15	12	0.94	2543.64	33	0.0000
	<i>P. otitidis</i>	26	22	" <i>P. caenicarum</i> "	7	6	0.96	1710.72	27	0.0000
	" <i>P. caenicarum</i> "	7	6	<i>P. psychrotolerans</i>	15	12	0.91	680.68	17	0.0000
	LG	21	17	LI	18	15	0.17	411.06	31	0.0000
	LG	21	17	MAN	31	28	0.11	503.36	44	0.0000
	MAN	31	28	LI	18	15	0.18	740.88	42	0.0000
<i>gyrB</i>	Verano	30	22	Invierno	42	27	0.19	1313.28	48	0.0000
	Verano-9y10	15	14	Invierno-9y10	32	18	0.33	961.62	31	0.0000
	<i>P. otitidis</i>	15	11	<i>P. psychrotolerans</i>	15	11	0.78	982.8	21	0.0000
	<i>P. otitidis</i>	15	11	" <i>P. caenicarum</i> "	15	2	0.86	619.2	12	0.0000
	" <i>P. caenicarum</i> "	15	2	<i>P. psychrotolerans</i>	15	11	0.9	648	12	0.0000
	LG	25	11	LI	22	21	0.24	699.36	31	0.0000
	LG	25	11	MAN	24	18	0.29	795.76	28	0.0000
	MAN	24	18	LI	22	21	0.12	419.52	38	0.0000
<i>recA</i>	Verano	23	18	Invierno	25	17	0.32	1044.48	34	0.0000
	Verano-9y10	15	11	Invierno-9y10	14	10	0.48	556.8	20	0.0000
	<i>P. otitidis</i>	17	14	<i>P. psychrotolerans</i>	4	4	0.9	642.6	17	0.0000
	<i>P. otitidis</i>	17	14	" <i>P. caenicarum</i> "	6	5	0.95	786.6	18	0.0000
	" <i>P. caenicarum</i> "	6	5	<i>P. psychrotolerans</i>	4	4	0.92	147.2	8	0.0000
	LG	19	16	LI	12	9	0.1	148.8	24	0.0000
	LG	19	16	MAN	19	16	0.15	353.4	31	0.0000
	MAN	19	16	LI	12	9	0.17	252.96	24	0.0000
<i>rpoD</i>	Verano	49	18	Invierno	34	20	0.18	1105.56	37	0.0000
	Verano-9y10	25	12	Invierno-9y10	21	12	0.37	782.92	23	0.0000
	<i>P. otitidis</i>	28	9	<i>P. psychrotolerans</i>	16	5	0.99	1132.56	13	0.0000
	<i>P. otitidis</i>	28	9	" <i>P. caenicarum</i> "	7	3	0.99	762.3	11	0.0000
	" <i>P. caenicarum</i> "»	7	3	<i>P. psychrotolerans</i>	16	5	0.98	315.56	7	0.0000
	LG	23	12	LI	23	13	0.09	198.72	24	0.0000
	LG	23	12	MAN	36	21	0.1	377.6	32	0.0000
	MAN	36	21	LI	23	13	0.15	584.1	33	0.0000

<sup>a</sup> Hedrick 2000

<sup>b</sup> grados de libertad = (k-1)(s-1), donde k es el número de alelos y s el número de subpoblaciones (Hedrick 2000)

<sup>c</sup> Valores significativos de  $F_{ST}$  cuando  $P < 0.05$

Junto con los análisis de diferenciación por estación y por sitio de colecta, los cálculos de diferenciación entre los tres linajes escogidos para análisis individual muestran valores de  $F_{ST}$  cercanos a 1, siendo todos los valores significativos (Tablas 4.11 y 4.12), lo que sugiere aislamiento histórico.

Finalmente, se realizó un análisis de la distribución espacial de los tres linajes principales de la muestra: *P. otitidis*, *P. psychrotolerans* y “*P. caenicarum*”. Este análisis consistió en graficar la abundancia de cepas de cada linaje recuperadas de las tres zonas de colecta en Churince. Los resultados de esto se muestran en la Figura 4.5, en donde se aprecia claramente que existe una diferencia en cuanto a los sitios que preferencialmente ocupó cada linaje. *P. otitidis* se encontró distribuido ampliamente pero con mayor abundancia en Laguna Intermedia (LI), mientras que “*P. caenicarum*” se encontró únicamente en Laguna Grande (LG) y *P. psychrotolerans* fue recuperada de LG pero el mayor porcentaje de las cepas se encontró en el Manantial (MAN). Aunque la representación gráfica es clara, una prueba  $\chi^2$  de una tabla de contingencia 3x3 con la distribución de los linajes mencionados revela que efectivamente las diferencias en la distribución son significativas ( $\chi^2 = 63.3$ ,  $df = 4$ ).

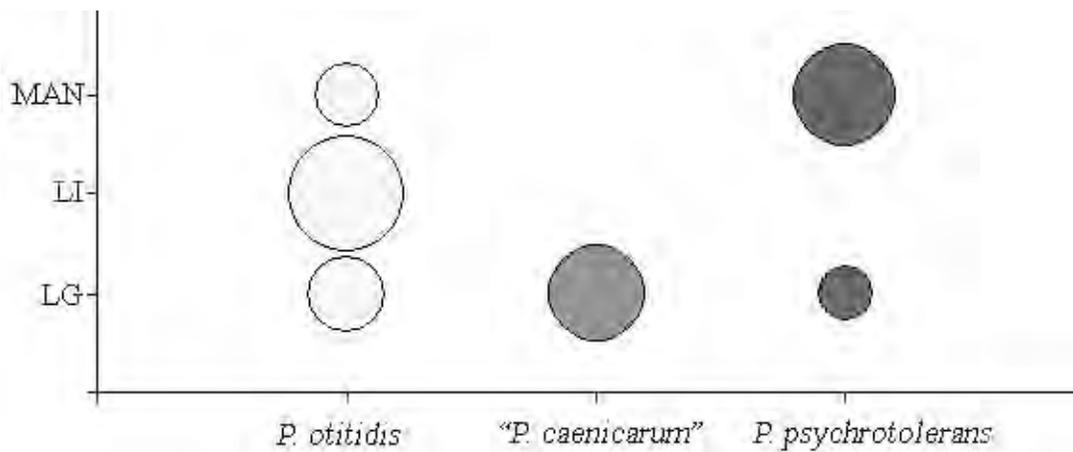


Figura 4.5. Representación gráfica de la distribución espacial de las cepas correspondientes a cada uno de los linajes más abundantes de la muestra de *Pseudomonas* de Churince en Cuatro Ciénegas, Coahuila. El tamaño del círculo representa el número de cepas recuperadas. Los sitios de colecta están agrupados por zonas: Laguna Grande (LG); Laguna Intermedia (LI) y Manantial (MAN).

Tabla 4.12. Valores de diferenciación genética entre los tres linajes principales de la muestra.

Locus	Linaje 1	Linaje 2	F <sub>ST</sub>
<i>acnB</i>	<i>P. otitidis</i>	<i>P. psychrotolerans</i>	0.94
	<i>P. otitidis</i>	" <i>P. caenicarum</i> "	0.96
	<i>P. psychrotolerans</i>	" <i>P. caenicarum</i> "	0.91
<i>gyrB</i>	<i>P. otitidis</i>	<i>P. psychrotolerans</i>	0.78
	<i>P. otitidis</i>	" <i>P. caenicarum</i> "	0.86
	<i>P. psychrotolerans</i>	" <i>P. caenicarum</i> "	0.90
<i>recA</i>	<i>P. otitidis</i>	<i>P. psychrotolerans</i>	0.90
	<i>P. otitidis</i>	" <i>P. caenicarum</i> "	0.95
	<i>P. psychrotolerans</i>	" <i>P. caenicarum</i> "	0.92
<i>rpoD</i>	<i>P. otitidis</i>	<i>P. psychrotolerans</i>	0.99
	<i>P. otitidis</i>	" <i>P. caenicarum</i> "	0.99
	<i>P. psychrotolerans</i>	" <i>P. caenicarum</i> "	0.98

*Análisis de correlación genética y ambiental.*- Se realizaron 6 pruebas de Mantel con matrices de diferenciación genética pareada por sitio ( $F_{ST}$ ) y matrices de diferencias ambientales (distancias euclidianas para posición geográfica, conductividad, pH, oxígeno disuelto, temperatura y variación en la temperatura). Ninguna de estas pruebas mostró correlaciones significativas entre las matrices, por lo que la hipótesis nula de que no existe asociación entre distribución y los parámetros ambientales medidos no se descarta.

**Demografía histórica.**- Con el objetivo de evaluar la dinámica poblacional en cuanto a tamaño efectivo de los tres linajes más abundantes, análisis de coalescencia tipo *Skyline plots* se realizaron y se muestran en la Figura 4.6. En estos análisis se nota para los tres linajes una curva que denota crecimiento exponencial, aunque las pendientes para cada

linaje parecen ser distintas. Las curvas para cada uno de los marcadores parecen variar también, sin embargo la tendencia en todos los casos es la misma.

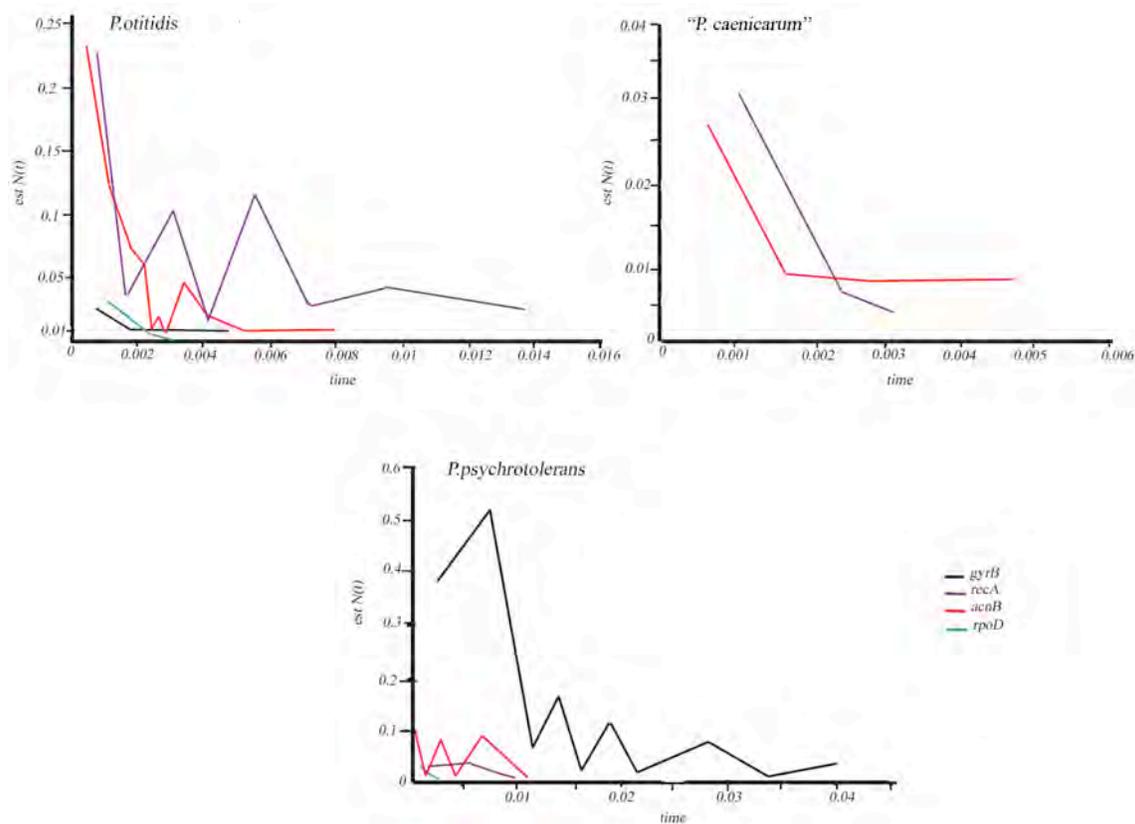


Figura 4.6. Representación gráfica de las simulaciones de coalescencia (*skyline plots*, Pybus y Rambaut 2002) para cada linaje analizado en donde cada línea representa la simulación para un locus.

#### 4.2.4. Discusión

##### I. Estudio de la distribución ecológica de genotipos

El análisis de huellas genómicas tipo rep-PCR, permite evaluar la posibilidad de patrones de distribución espacial o temporal de los genotipos estudiados (Cho y Tiedje 2000). Sin embargo, los resultados obtenidos del análisis de cluster y las pruebas de correlación de Mantel muestran, en general, una notable falta de patrones de distribución de

genotipos asociados a las condiciones ambientales medidas. Sólo en el caso de la diferencia estacional de la temperatura se encuentra una correlación significativa, aunque baja (coeficiente de correlación 0.28,  $p < 0.01$ ). Además, la visualización de los patrones en el dendrograma hace evidente que la muestra invernal tiene menos cepas idénticas que la muestra de verano, indicando una posible dinámica ecológica diferente en ambas estaciones, lo cual estaría de acuerdo con una posible correlación de la temperatura o facetas de ella y la distribución de distintos genotipos. Sin embargo, como se ha mencionado, para el resto de los parámetros ambientales, incluida la distancia geográfica, no parece haber evidencia de algún patrón de distribución de genotipos. Esta situación sorprende, pues estudios previos con *Pseudomonas* y otros grupos similares (Cho y Tiedje 2000, Bent *et al.* 2003, Oda *et al.* 2003) han mostrado estructuración espacial evidente. El resultado observado aquí puede deberse a una o cualquier combinación de las siguientes situaciones: 1) la falta de conocimiento de los factores ambientales o facetas de ellos, que en realidad determinan la distribución y ofrecen auténticas barreras a la dispersión de microorganismos y que por ello no fueron considerados; 2) la dispersión y establecimiento de microorganismos no obedece a factores ambientales o históricos sino a eventos aleatorios; 3) la escala espacial a la que aproximamos el estudio es demasiado pequeña y no ofrece límites a la dispersión del grupo de estudio por lo que no hay diferencias entre sitios de muestreo que reflejen historia de aislamiento poblacional; 4) *Pseudomonas* es un género muy adaptable a distintas fuentes de carbono y distintas situaciones ambientales (Spiers *et al.* 2000), por lo que los patrones de distribución sólo se ven a gran escala espacial (histórica) como en el caso del trabajo de Cho y Tiedje (2000) en donde la escala espacial menor es de 170 km; 5) como lo muestran los análisis con otros marcadores moleculares, hay pobre correspondencia entre la identidad filogenética y los patrones fingerprint, por lo

que es difícil hacer correlaciones de distancia genética con cualquier otro parámetro. Entonces, en general los patrones de distribución espacial de la variación genética en el caso de la muestra analizada en este estudio con rep-PCR no son evidentes y quedan por descartar varias posibilidades que afecten este resultado, sin embargo parece existir una distribución diferencial de genotipos en el tiempo tomado como verano e invierno.

## ***II. Análisis filogenéticos y de genética de poblaciones***

El análisis de secuencias de varios genes esenciales permite estudiar con detalle la dinámica evolutiva de los organismos en consideración. Con este tipo de datos, en este trabajo observamos en primer lugar, una notable diversidad filogenética revelada por el gen 16S rRNA y, en segundo lugar, con respecto a la estructura poblacional del grupo y usando otros loci, observamos clonalidad, alta diversidad de haplotipos, gran diferenciación genética entre linajes y entre temporadas de colecta y poca diferenciación genética entre sitios de muestreo.

### **a. Diversidad filogenética (gen 16S rRNA)**

El análisis del gen 16SrRNA de las cepas de este estudio y que forman parte de linajes independientes en las genealogías de los otros loci mostró que las cepas pertenecen a *Pseudomonas sensu stricto* y que en la muestra se encuentran representados los dos principales grupos filogenéticos descritos para el género (IGCI, IGCI) con al menos siete linajes: *P. mendocina* (IGCI), *P. alcaliphila* (IGCI), *P. psychrotolerans* (IGCI), *P. otitidis* (IGCI), *P. fluorescens* (IGCII), *P. putida* (IGCII), *P. mosselii* (IGCII) y un octavo linaje nuevo que denominamos "*P. caenicarum*". Este resultado es interesante sobre todo por el número de cepas totales analizadas (~ 80), lo que significa un linaje registrado por cada 10 cepas colectadas. Esta tasa de aparición de linajes por esfuerzo de muestreo podría, en

teoría, darnos una idea de la diversidad total de *Pseudomonas* en el lugar de muestreo, sin embargo es difícil determinar si esta es “alta” o “baja”, pues la mayoría de los estudios con microorganismos que evalúan diversidad filogenética, lo hacen con muestras totales (no cultivo) y en general el número de linajes para una muestra de 80 (clonas u observaciones) suele ser mayor a 10 (Hughes *et al.* 2001, Hughes y Bohannan 2004). Si consideramos que se estima que el total de la diversidad en cultivo es <5% de la diversidad total (Tiedje 1996, Torsvik *et al.* 1990), entonces sí podemos decir que la diversidad que recuperamos de un solo grupo (*Pseudomonas*) es muy alta. Por último, sobre esta parte del estudio podemos decir que la diversidad amplia del género *Pseudomonas* se encuentra representada en el sistema Churince al recuperar al menos siete linajes del grupo.

#### **b. Genética de Poblaciones (genes codificantes esenciales)**

Los datos presentados sobre distribución y estructura genética a nivel poblacional nos hablan de linajes fundamentalmente clonales, los cuales pueden entenderse como un grupo de organismos cuya divergencia es limitada y recurrentemente llevada a cero por eventos intermitentes de selección natural o cuellos de botella (Atwood *et al.* 1951). Sin embargo, los análisis explícitos de selección no dejan claro si esta es una fuerza determinante en la dinámica evolutiva de los linajes, por lo que es posible que otros mecanismos estocásticos (deriva génica y/o mutación) sean de mayor relevancia en cuanto a ello se refiera.

##### *i. Estructura genética (diversidad, recombinación, selección)*

La variación genética en cualquier especie nos habla de los procesos evolutivos por los que ésta ha pasado. En particular, los eventos de recombinación juegan un papel crucial en la evolución de las bacterias al homogeneizar la variación genética dentro de clonas y al introducir variación entre ellas. La importancia de esta fuerza evolutiva en bacterias ha sido

objeto de controversia y extensa investigación en casi todos los grupos conocidos y más recientemente de archaeas (Whitaker *et al.*, 2005; Papke *et al.*, 2007). En procariontes, hasta ahora parece no existir una regla general en cuanto a los niveles de recombinación, pues existe evidencia de grupos totalmente clonales como *Salmonella* (Boyd *et al.*, 1996), algunos grupos de *Pseudomonas* (Spiers *et al.*, 2000), y otros patógenos epidémicos como *Yersinia* (Achtman *et al.*, 1999) y por otro lado hay amplia evidencia de grupos donde la recombinación parece ser muy frecuente como *Helicobacter pylori* (Falush *et al.*, 2001), *Bacillus subtilis* (Istock *et al.*, 1992) e incluso estudios recientes muestran a archaeas extremófilas como altamente recombinantes (Whitaker *et al.*, 2005; Papke *et al.*, 2007). Esta diferencia en niveles de recombinación es responsable, en parte, de la dinámica evolutiva particular de los distintos grupos que van desde patógenos obligados, oportunistas, comensales hasta organismos de vida libre.

Nuestros análisis para la colección total de *Pseudomonas* y los linajes más abundantes muestran, en todos los casos, una estructura clonal con evidencia escasa o inexistente de recombinación en los datos de secuencias de DNA. Este resultado puede discutirse en extenso en cuanto a las posibles causas. De acuerdo con Maynard-Smith y colaboradores (1993), existen varios escenarios en donde puede haber señal de clonalidad en poblaciones de bacterias que son recombinantes:

- a) Las muestras analizadas pueden tratarse de una mezcla de varias poblaciones, dentro de las cuales la recombinación es común dentro pero no entre ellas por diferentes razones (aislamiento geográfico por ejemplo, Souza *et al.* 1992).
- b) Una posible estructura poblacional “epidémica” puede llevar a desequilibrio de ligamiento temporal (*P. aeruginosa* por ejemplo, Vernez *et al.* 2005).
- c) Por cuellos de botella severos o deriva génica.

Para distinguir las causas que generan el ligamiento o el resultado de clonalidad y, al mismo tiempo, tener certeza de los resultados, se puede evaluar:

- Análisis de subgrupos: si la señal de clonalidad desaparece cuando se analizan subgrupos de la muestra total, por ejemplo distintos linajes, es posible que nos encontremos ante el escenario de estructuración poblacional antes mencionado (punto a).
- Análisis de aislados individuales: si los aislados idénticos se tratan en grupo como un individuo, la señal de ligamiento suele desaparecer y es indicativo de una estructura epidémica. Aquí es importante señalar que el método de muestreo puede llevar a valores de clonalidad o ligamiento incluso en poblaciones con alta frecuencia de recombinación.

Dado lo anterior y con los datos presentados en este estudio podemos decir por un lado que, la señal de clonalidad no desaparece al analizar submuestras de la colección total, por lo que en este sentido hay certeza del resultado. Por otro lado, la posibilidad de una estructura epidémica no se analizó explícitamente y valdría la pena considerarlo, sobre todo en los casos en donde el número de haplotipos en relación al número de cepas es menor en al menos un 50%, por lo tanto esta es una posibilidad que no se descarta.

En este sentido, el resultado de clonalidad podría tener sentido en comparación con estudios previos con otras especies en donde se han establecido altos niveles de clonalidad *P. syringae* (Sarkar y Guttman 2004) dentro de los organismos patógenos y *P. stutzeri* (Rius *et al.* 2001) dentro de los de vida libre. Sin embargo no podemos descartar, como se mencionó, la posibilidad de una estructura epidémica como en el caso de *P. aeruginosa* (Vernez *et al.* 2005).

Por otro lado, el alto número de haplotipos por locus y la baja diversidad nucleotídica ( $\pi$  y  $\theta$ ) son consistentes con una estructura clonal en la que la recombinación como fuente de variación genética no existe o es muy poca, dejando como fuerza diversificadora a la mutación. Adicionalmente, al no existir intercambio genético (que además de ser fuerza diversificadora es al mismo tiempo un mecanismo cohesivo o de homogeneización genética), las variantes genéticas surgidas por mutación se mantienen y generan un alto porcentaje de haplotipos en cada población.

En relación a la diferenciación genética como consecuencia del flujo genético, los resultados son variables. La diferenciación genética entre sitios de muestreo es baja aunque significativa. Por otro lado, la diferenciación genética entre linajes es prácticamente de 100%, lo cual es totalmente consistente con los resultados de recombinación, splits decomposition e ILD. Finalmente la diferenciación genética como una medida indirecta de flujo o migración entre estaciones o temporadas de colecta muestra una notable diferencia entre ambas temporadas, lo que implica la existencia de diferentes genotipos en verano e invierno. En resumen, en este caso la estructura clonal y alta diferenciación genética entre linajes y estaciones del año sugiere otras fuerzas como selección, deriva génica y mutación como fuerzas evolutivas, incluso en un escenario de estructura epidémica.

En cuanto al papel de la selección en la dinámica evolutiva de los linajes de *Pseudomonas* en Churince, este no es del todo claro. En el caso de la prueba de neutralidad de Tajima, de Fu-Li y de la relación  $dN/dS$ , los valores sugieren selección negativa o purificadora, aunque en su mayoría no son significativos. Sin embargo, si tomamos en cuenta que aunque, los valores de selección por linaje y por locus no son significativos, al menos todos son en el mismo sentido (congruente con clonalidad). En este sentido podríamos hablar del mismo efecto selectivo en todo el genoma, lo cual de acuerdo con

Hein *et al.* (2004) nos podría estar dando una señal de crecimiento poblacional con cuellos de botella muy severos. Esta interpretación no es difícil de asociar con las observaciones ambientales del sitio de estudio y con los resultados arrojados por los “*skyline plots*”, por lo que lo considero una posibilidad que podría probarse en futuros estudios. Por otro lado, la prueba de MacDonald-Kreitman sí muestra diferencias significativas en presiones de selección entre linajes, pero sólo para un locus (*gyrB*), lo cual tendría sentido en cuanto a la distribución espacial diferente de los linajes pero contradice los datos sobre clonalidad, ya que si en efecto se trata de linajes clonales, la señal de selección tendría que existir en todos los loci, es posible que el uso de otros marcadores moleculares, o simplemente un mayor número de ellos resuelvan esta paradoja .

Dado lo anterior, y dado que los resultados sobre selección no son concluyentes, parece que las principales fuerzas en devenir evolutivo de los linajes estudiados quedarían reducidas a dos, las cuales son fundamentalmente resultado de procesos estocásticos y son: deriva génica y mutación. Estimar la participación relativa de estas dos fuerzas no es una tarea fácil, pues en el primer caso es necesario estimar el tamaño efectivo de las poblaciones, el cual se estimó mediante las simulaciones de coalescencia, sin embargo estas simulaciones parten de una “filogenia conocida” lo cual de inicio compromete la estimación. Otra forma que evitaría este pre-juicio sobre la evolución de los linajes sería hacer análisis más sofisticados que involucren reconstrucciones probabilísticas basadas en relaciones filogenéticas usando estadística bayesiana. En cualquier caso, los datos de coalescencia indicarían que es muy posible que el resultado actual en los linajes de *Pseudomonas* sea consecuencia de procesos estocásticos como deriva génica y/o mutación.

## *ii. Patrones de distribución de la variación genética*

Los resultados del análisis de estructura de la variación genética nos indican diferenciación genética entre la muestra de verano e invierno, así como la distribución diferencial de los tres linajes principales a lo largo de los sitios de estudio. Los valores encontrados para la diferenciación espacial son significativos aunque bajos, lo que puede ser indicativo de algún grado de migración entre sitios. Esta situación podría ser reflejo de una estructura de tipo *metapoblacional*. En una metapoblación, de acuerdo con Levins (1969), las subpoblaciones existen en parches de habitats discretos y estas subpoblaciones pueden cambiar con extinciones y recolonizaciones de otros parches. La dinámica de extinción y recolonización puede afectar notablemente el tamaño efectivo de las poblaciones, la variación genética y la distribución de ésta en las subpoblaciones (Slatkin 1977) y podría ser el mecanismo de estructuración de las poblaciones aquí estudiadas.

Los datos de distancias genéticas entre sitios de muestreo obtenidas a partir de la composición genotípica de las poblaciones colectadas en cada uno de ellos, permite probar hipótesis sobre factores ambientales relacionados con las diferencias observadas. En este caso no se observa ninguna correlación entre distancias genéticas y distancias geográficas o diferencias en los parámetros ambientales, a pesar de que es evidente la diferenciación genética entre la muestra de verano e invierno, así como la distribución diferencial de los tres linajes principales a lo largo de los sitios de estudio, lo cual se refleja en la diferenciación genética por sitio ( $F_{ST}$ ), que aunque baja, es significativa. El hecho de encontrar estructura genética de las poblaciones y no poder asociar esta a ningún factor ambiental puede deberse a tres factores: 1) Que en la muestra total, en la que es posible hacer correlaciones entre matrices genéticas y ambientales, existen muy pocas cepas que

permitan distinguir un patrón espacial/ambiental de distribución, como son los tres linajes numéricamente dominantes; 2) Aunque exista una distribución diferente de los grupos, esto no necesariamente implica que distancias genéticas mayores o linajes menos relacionados se encontrarán físicamente más distantes y 3) La escala filogenética de estudio no permite visualizar fenómenos como aislamiento por distancia o adaptaciones locales de genotipos pertenecientes a una misma especie. A pesar de esto, el resultado obtenido de los análisis de correlación ambiental desconcierta, pues existen reportes para *Pseudomonas* y otros grupos mesófilos y extremófilos en donde se encuentra clara correlación ambiental y/o geográfica con distancias genéticas (Cho y Tiedje, 2000; Bent *et al.*, 2003; Papke *et al.*, 2003; Whittaker *et al.*, 2003). Sin embargo, esta no parece ser una regla general pues, por ejemplo un estudio que involucra cepas colectadas de distintas fuentes y regiones geográficas de *P. syringae*, especie patógena de plantas, no encuentra asociación entre distancias genéticas y geográficas y tampoco encuentra asociación clara entre variantes genéticas determinadas y su hospedero (Sarkar y Guttman 2004). En resumen, en este caso la hipótesis nula de no asociación de la distribución de la diversidad genética con factores ambientales no puede descartarse.

### *iii. Demografía histórica*

Los datos obtenidos del análisis de demografía histórica muestran un posible crecimiento exponencial de las poblaciones estudiadas, lo cual podría explicar muchas de las observaciones anteriores, pues estamos hablando de tamaños poblacionales que han cambiando con el tiempo, probablemente se trate de una estructura clonal epidémica, en donde el tamaño efectivo original hubiera sido pequeño y resultado de un cuello de botella pasado.

Trabajos experimentales y teóricos han demostrado que si los tamaños efectivos de las poblaciones son pequeños debido, por ejemplo, a cuellos de botella, el papel de la selección natural es sobrepasado por fuerzas de carácter estocástico, en primer lugar deriva génica y por supuesto mutación; en este sentido se pueden fijar o propagar en la población, mutantes poco benéficas en términos de adecuación (Elena y Lenski, 2003). Los datos presentados aquí no hacen difícil imaginar un escenario en donde existan cuellos de botella periódicos que permitan establecer poblaciones clonales cada cierto tiempo, barriendo en cada evento la diversidad intra-linaje o intra-poblacional, en donde los factores periódicos de reducción poblacional podrían ser ambientales (variaciones estacionales) o biológicos (depredación). Esta posibilidad haría difícil encontrar evidencia del papel de la selección natural en el proceso evolutivo, lo cual, coincide con la pobre señal que observamos, así como con la elevada clonalidad de las poblaciones.

#### ***4.2.5. Conclusiones***

En conclusión, los datos presentados sobre distribución y estructura genética a nivel poblacional nos hablan de poblaciones clonales, las cuales pueden entenderse como un grupo de organismos cuya divergencia es limitada y recurrentemente llevada a cero por eventos intermitentes de selección natural o cuellos de botella (Atwood *et al.* 1951). Tales eventos pueden ser debidos a cambios estacionales de las condiciones ambientales que lleven a la reducción de la diversidad genética dentro de una población aunque sin prevenir la divergencia entre poblaciones (Cohan 1994). En este estudio los datos sugieren un posible papel protagónico de procesos estocásticos, lo cual dificulta la identificación de patrones de distribución espacial. A pesar de lo anterior, no es posible descartar del todo el papel de la selección natural. Es posible que estudios que tomen en consideración más

eventos estacionales aporten evidencia para determinar si los eventos de purga genética periódica se tratan de eventos selectivos de variantes genéticas (Ellis *et al.* 1999) o si se trata de eventos aleatorios de reducción de diversidad (cuellos de botella).

## Capítulo 5. CONCLUSIONES GENERALES Y PERSPECTIVAS

En el año de 1982, el ecólogo de poblaciones John L. Harper publica un ensayo: “*After description*” en el cual discute sobre la contribución de la descripción detallada de la variación individual dentro de las poblaciones y para entender los mecanismos genéticos de tal variación, esto en el ánimo de llegar a las explicaciones evolutivas últimas de lo que vemos hoy en día en su comportamiento y distribución. De alguna forma, esto es pasar del detalle de la genética de las poblaciones a dar generalizaciones aplicables incluso a predicciones a nivel de comunidades. La tesis aquí presentada sin duda comparte la visión de hace más de 20 años de Harper y la de otros ecólogos y ecólogos evolutivos que han trabajado en desarrollar métodos que permitan enriquecer la información obtenida bajo distintas escalas de análisis ecológico de los organismos. Entonces, este contexto filosófico obliga a que en este capítulo final se evalúe la contribución particular de los datos presentados en cada escala de análisis (poblaciones y comunidades) al entendimiento de los posibles mecanismos subyacentes a la distribución de la biodiversidad observada, o puesto en las mismas palabras con las que se presentó este manuscrito:

- ¿Es posible entender los patrones de distribución de especies (comunidades), al profundizar en el análisis de la dinámica evolutiva dentro de las especies (poblaciones)?
- ¿Es cada nivel jerárquico de biodiversidad independiente y posee características no escalables?

En cuanto a la primera pregunta, los datos presentados en este trabajo sobre diversidad a nivel de comunidades nos hablan de una distribución parchada, así como de la

falta de patrones de distribución relacionados con el ambiente. Por otro lado, los datos sobre diversidad de *Pseudomonas* nos permitieron ver algunos patrones de distribución estacional, y alguna señal de patrones de distribución espacial aunque no claramente asociada a algún parámetro ambiental. En términos generales, los datos no se contradicen y son consistentes en cuanto a que la diversidad entre sitios es alta. En cuanto a la contribución del estudio poblacional al entendimiento de los patrones de diversidad a nivel de comunidades, es claro que se refiere a los detalles de la dinámica evolutiva que puede revelarse con el análisis molecular de varios loci y varios linajes. En este sentido los datos encontrados en las poblaciones hacen evidente la posibilidad real de que la distribución esté asociada más a fenómenos estocásticos, como cuellos de botella estacionales, que a algún factor selectivo, lo cual no era posible evaluar en las comunidades con los datos disponibles.

A pesar de esta contribución sobre los detalles poblacionales, y volviendo a la visión de Harper, es difícil hacer de estas observaciones una generalización, pues es claro que la diversidad de microorganismos es tan amplia que se requieren muchas observaciones individuales de distintos grupos para empezar a conocer patrones a nivel de comunidades. Este trabajo es sólo una contribución al esfuerzo general.

Sobre la segunda interrogante, hablar de independencia entre niveles jerárquicos no es tarea fácil, pues sin duda la suma de especies individuales contribuye a la generación de los patrones de diversidad a escalas mayores, aunque este resultado no es igual a la suma de las partes, ya que características propias y emergentes de cada escala pueden ser independientes. Nuestros datos apuntan a que algunas características son escalables, en este

caso la distribución espacio/temporal restringida de algunos grupos o genotipos, es decir en las comunidades los valores de diversidad beta y en poblaciones los niveles de diferenciación genética. Sin embargo, consideramos que cada escala de estudio arroja información particular que puede ayudar a construir generalizaciones. Uno de los mejores ejemplos es la información evolutiva que nos da el análisis genético de las poblaciones en donde podemos inferir historias pasadas y contrastar esto con los patrones observados a una escala de paisaje. Finalmente, a pesar de que en las comunidades fue revelada una alta diversidad debido posiblemente a la diferenciación que encontramos en las poblaciones, en ningún caso resulta definitivo cuáles son las variables abióticas que influyen en esta diferenciación. Esto puede ser resultado de fallas en ambas aproximaciones en cuanto a la caracterización ambiental y a las hipótesis sobre adaptación de los microorganismos.

### *5.1. Conclusiones:*

1. El estudio detallado de la variación genética en poblaciones individuales ayuda a conocer las posibles causas de la distribución de la diversidad a escalas superiores y con ello a plantear hipótesis sobre patrones de distribución.
2. La diversidad de microorganismos en CCC es fundamentalmente debida a la distribución discontinua de grupos genéticamente distintos tanto en el espacio como en el tiempo. Esto es cierto a escala de comunidades y de poblaciones.
3. El grupo elegido para el estudio poblacional sugiere una dinámica metapoblacional o de cuellos de botella periódicos, posiblemente asociados a cambios estacionales, los cuales resultan en una estructura poblacional definida por fuerzas evolutivas de carácter fuertemente aleatorio como son deriva génica y mutación.

4. La clara diferenciación genética encontrada entre las muestras de verano e invierno, sugiere que las variaciones estacionales y los factores ambientales asociados a estas, sean relevantes en la estructuración de las poblaciones y las comunidades microbianas acuáticas del valle de Cuatro Ciénegas, Coahuila.

## *5.2. Perspectivas*

1. Se requieren más estudios de poblaciones con otros grupos en la zona que incluyan cepas de otros sitios e incluso de otras regiones que permitan hacer generalizaciones sobre la dinámica ecológica de microorganismos en Cuatro Ciénegas.
2. Estudios futuros sobre comunidades deberán enfocarse en grupos menos amplios que los procariontes estudiados aquí, para que sea posible empezar a comprender los factores de relevancia en el establecimiento y estructuración de comunidades.
3. Dados los altos niveles de clonalidad y la distribución en el tiempo de los linajes, son necesarios estudios a más largo plazo y con otros grupos cuyas características metabólicas les hagan más restrictiva la distribución para ampliar la visión de la dinámica del lugar y el conocimiento de la ecología microbiana en general.
4. Es necesaria la caracterización a largo plazo de las condiciones ambientales de la zona para formular hipótesis más claras sobre el posible papel del ambiente en la estructura y dinámica genéticas de las poblaciones y comunidades de microorganismos en Cuatro Ciénegas.

## REFERENCIAS

- Achtman, M., K. Zurth, G. Morelli, G. Torrea, A. Guiyoule y E. Carniel.** 1999. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA*. 96: 14043-14048.
- Alcocer, J y E. Kato.** 1995. Cuerpos acuáticos de Cuatro Ciénegas, Coahuila. En: G. de la Lanza y J.L. García C. (comps.), *Lagos y presas de México*. Centro de Ecología y Desarrollo, A.C. México, pp. 177-193.
- Atwood, K.C., L.K. Schneider y F.J. Ryan.** 1951. Periodic selection in *Escherichia coli*. *Proc Natl Acad Sci USA*. 37: 146-155.
- Avise, J.C.** 1994. *Molecular Markers, Natural History and Evolution*. Chapman y Hall. New York 511p.
- Badino, G., T. Bernabei, A. De Vivo, I. Giulivo y G. Savino.** 2004. Under the desert, the mysterious waters of Cuatro Ciénegas. La Venta Esplorazioni Geografiche y Gobierno Coahuila.
- Bass-Becking, L.G.M.** 1934. *Geobiologie of inleiding to de milieukunde*, W.P. van Stockum y Zoon N.V.
- Bent, S.J., C.L. Gucker, Y. Oda y L.J. Forney.** 2003. Spatial distribution of *Rhodopseudomonas palustris* ecotypes on a local scale. *Appl Environ Microbiol* 69: 5192-5197.
- Blondel, J.** 1987. From biogeography to life history theory: a multithematic approach illustrated by the biogeography of vertebrates. *J Biogeog* 14: 405-422.
- Bohannon B.J.M. y R.E. Lenski.** 2000. The relative importance of competition and predation varies with productivity in a model community. *Am Nat* 156: 329-340.
- Boyd, E.F., F.S. Wang, T.S. Whittan y R.K. Selander.** 1996. Molecular genetic relationships of the salmonellae. *Appl Environ Microbiol* 62: 804-808.
- Brown, E.W., M.L. Kotewicz y T.A. Cebula.** 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol Phylogenet Evol* 24: 102-120.
- Brown, J.H. y M.V. Lomolino.** 1998. Biogeography. Sinauer Sunderland, USA.
- Calegari, V.** 1997. Environmental perceptions and local conservation efforts in Cuatro Cienegas, Coahuila, Mexico. The University of Texas- Austin Dissertation.
-

- Casamayor E.O., R. Massana, S; Benlloch, L. Øvreås, B. Díez, V. Goddard, J.M. Gasol, I. Join, F. Rodríguez-Valera y C. Pedrós-Alió.** 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprint methods in a multipond solar saltern. *Environ Microbiol* 4: 338-348.
- Caugant, D.A., L.F. Mocca, C.E. Frash, L.O. Froholm, W.D. Zollinger y R.K. Selander.** 1987. Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and other membrane protein patterns. *J Bacteriol* 169: 2781-2792.
- Cho, J.C. y J.M. Tiedje.** 2000. Biogeography and degree of endemism of fluorescent *Pseudomonas* strains in soil. *Appl Environ Microbiol* 66:5448-5456
- Cody, M.L.** 1975. Towards a theory of continental species diversity: bird distribution over Mediterranean habitat gradients. *Ecology and Evolution of Communities* (eds M.L. Cody and J.M. Diamond) pp. 214-257. Harvard University Press, Cambridge, MA.
- Cohan, F.** 1994. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* 9: 175-180
- Connell J.H.** 1978. Diversity in tropical rain forests and coral reefs. *Science* 199: 1302-1310.
- Cousins, S.H.** 1991. Species diversity measurement: Choosing the right index. *Trends Ecol Evol* 6: 190-192.
- Crawley, M.J. y J.E. Hurrell.** 2001. Scale Dependence in Plant Biodiversity. *Science* 291: 864-868.
- Darwin, C.** 1859. On the Origin of Species.
- DeSantis, T.Z., P. Hugenholtz, K. Keller, E.L. Brodie, N. Larsen, Y.M. Piceno, R. Phan y G.L. Andersen.** 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 2006 July 1; 34(Web Server issue): W394–W399.
- Desert Fishes Council Web.** 2003. <http://www.desertfishes.org>
- Di Castri, F. y T. Younes.** 1996. Introduction: Biodiversity, the emergence of a new scientific field –its perspectives and constraints. En: di Castri F y T Younes (eds)

Biodiversity, Science and Development, towards a new partnership: 1-11 CAB International Wallingford, Oxon, UK.

- Dunbar J., L.O. Ticknor y C.R. Kuske.** 2000. Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S rRNA Gene Terminal Restriction Fragment Analysis. *Appl Environ Microbiol* 66(7): 2943-2950.
- Dunbar J., S. Takala, S.M. Barnes, J.A. Davis y C.R. Kuske.** 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol.* 65: 1662-1669.
- Dykhuisen D.E.** 1998. Santa Rosalia revisited: why are there so many species of bacteria? *Antoine van Leeuwenhoek* 73: 25-33.
- Dykhuisen, D.E. y L. Green.** 1986. DNA sequence variation, DNA phylogeny, and recombination in *Escherichia coli*. *Genetics* 113: S71.
- Edgar, R.C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5), 1792-97.
- Eguiarte, L.E., J. Larson-Guerra, J. Nuñez-Farfan, A. Martínez-Palacios, K. Santos del Prado y H.T. Arita.** 1999. Diversidad filogenética y conservación: ejemplos a diferentes escalas y una propuesta a nivel poblacional para *Agave victoriae-reginae* en el desierto de Chihuahua, México. *Rev Chil Hist Nat* 72: 475-492.
- Eldredge, N.** 1985. Unfinished Synthesis – Biological Hierarchies and Modern Evolutionary Thought. Oxford University Press. USA 237p.
- Elena, S.F. y R.R. Lenski.** 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* 4: 457-469.
- Ellis, R.J., I.P. Thompson y M.J. Bailey.** 1999. Temporal fluctuations in the pseudomonad population associated with sugar beet leaves. *FEMS Microbiol Ecol* 28: 345-356.
- Elser, J.J.** 2006. Biological Stoichiometry: A chemical bridge between ecosystem ecology and evolutionary biology. *Am Nat* 168: S25-S35.
- Elser, J.J., J.H. Schampel, F. García-Pichel, B.D. Wade, V.Souza, L. Eguiarte, A. Escalante y J.K. Farmer.** 2005. Effect of phosphorous enrichment and grazing snails on modern stromatolitic microbial communities. *Freshwater biology* 50: 1808-1825.
-

- Evans, S.B.** 2005. Using chemical data to define flow systems in Cuatro Ciéneas, Coahuila México: The University of Texas – Austin Dissertation 114p.
- Falush , D., C. Kraft, N.S. Taylor, P. Correa, J.G. Fox, M. Achtman y S. Suerbaum.** 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size and minimal age. *Proc Natl Acad Sci USA*. 98: 15056-15061.
- Farris, J.S., M. Källersjö, A.G. Kluge y C. Bult.** 1995. Constructing a Significance Test for Incongruence. *Syst Biol* 44: 570-572
- Fierer, N. y R.B. Jackson.** 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA*. 103: 626-631.
- Finlay, B.J.** 1998. The global diversity of protozoa and other small species. *Int J Parasitol* 28: 29-48.
- Finlay, B.J.** 2002. Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061-1063.
- Fu, Y.X. y W.H. Li.** 1993. Statistical tests of neutrality mutations. *Genetics* 133: 693-709.
- Fuhrman J.A.** 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541-548.
- Fulthorpe, R.R., A.N. Rhodes y J.M. Tiedje.** 1998. High levels of endemicity of 3-chlorobenzoate-degrading soil bacteria. *Appl Environ Microbiol* 64(5): 1620-1627.
- Futuyma, D.J.** 1998. Evolutionary Biology. Sinauer. Massachusetts USA
- García-Pichel, F., W.D. Wade y J.D. Farmer.** 2002. Jet-suspended, calcite-ballasted cyanobacterial waterwarts in a desert spring. *J Phycol* 38: 420-428.
- Gordon, D., C. Abajian y P. Green.** 1998. [Consed: a graphical tool for sequence finishing](#). *Genome Res* 8:195-202.
- Gordon, D., C. Desmarais y P. Green.** 2001. [Automated finishing with Autofinish](#). *Genome Res* 11:614-625.
- Gould, D.J.** 1998. [Gulliver's further travels: the necessity and difficulty of a hierarchical theory of selection](#). *Philos Trans R Soc London Ser B* 353: 307-314.
- Green, J. y B.J.M. Bohannan.** 2006. Spatial scaling of microbial biodiversity. *Trends Ecol Evol* 21: 501-507.
-

- Green, J.L., M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings y A.J. Beattie.** 2004. Spatial scaling of microbial eukaryote diversity. *Nature* **432**: 747-750.
- Harper, J.L.** 1982. After description. En: The plant community as a working mechanism, Newman, E.I. (ed) Special Publication series of the British Ecological Society No.1 Blacwell Scientific Publications. UK. p. 11-25.
- Hartl, D.L. y A.G. Clark.** 1989. Principles of Population Genetics. 2<sup>nd</sup> edition. Sinauer. Massachusetts USA. 682p.
- Haubold, B., M. Travisano, P.B. Rainey y R.R. Hudson.** 1998. Detecting linkage disequilibrium in bacterial populations. *Genetics* **150**: 1341-1348
- Hedrick, P.W.** 2000. Genetics of Populations. Johns and Bartlett Publishers, USA.
- Hein, J., M.H. Schierup y C. Wiuf.** 2004. Gene genealogies, variation and evolution: A primer in coalescent theory. Oxford University Press. 304 p.
- Helling R.B., C. Vargas y J. Adams.** 1987. Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* **116**: 349-358.
- Horner-Devine, M., M. Lage, J.B. Hughes y B.J.M. Bohannan.** 2004. A taxa-area relationship for bacteria. *Nature* **432**: 750-753.
- Hoshino T., N. Noda, S. Tsuneda, A. Hirata y Y. Inamori.** 2001. Direct detection by in situ PCR of the amoA gene in biofilm resulting from a nitrogen removal process. *Appl Environ Microbiol* **67**(11): 5261-5266.
- Hudson, R.R., M. Slatkin y W.P. Maddison.** 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583-589
- Hughes, J.B. y B.J.M. Bohannan.** 2004. Application of ecological diversity statistics in microbial ecology En: Molecular Microbial Ecology Manual 2<sup>nd</sup>. Edition 7.01: 1321-1244.
- Hughes, J.B., J.J. Hellman, T.H. Ricketts y B.J.M. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol.* **67**: 4399-4406.
- Hughes, J.B.M., B.J.M. Bohannan, J.H. Brown, R.K. Colwell, J.A. Fuhrman, J.L. Green, M.C. Horner-Devine, M. Kane, J.A. Krumins, C.R. Kuske, P.J. Morin, S. Naaem, L. Ovreas, A-L Reysenbach, V.H. Smith y J.T. Staley.** 2006.
-

- Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* 4: 102-112.
- Huson, D.H.** 1998. Splits Tree: a program for analyzing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.
- Huson, D.H. y D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2): 254-267.
- Hutchinson G.E.** 1961. The paradox of the plankton. *Am Nat* 95: 137-147.
- Istock, C.A., K.E. Duncan, N. Ferguson y X. Zhou X.** 1992. Sexuality in a natural population of bacteria, *Bacillus subtilis* challenges the clonal paradigm. *Mol Ecol* 1:95-103
- Johannesson, K.H., A. Cortés y K.C. Kilroy.** 2004. Reconnaissance isotopic and hydrochemical study of Cuatro Ciénegas groundwater, Coahuila, México. *South Am Earth Sci* 17:171-180.
- Kielwein, G.** 1971. die Isolierung und Differenzierung von Pseudomonaden aus Lebensmitteln. - *Arch. f. Lebensmittelhyg.* 22: 29-37.
- Kiewitz, C. y B. Tümmler.** 2000. Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J Bacteriol* 182: 3125-3135.
- Kimura, M.** 1968. Evolutionary rate at the molecular level. *Nature* 217: 624-626
- Kumar S., K. Tamura y M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150-163.
- Lane, D.J.** 1991. 16S/23S rDNA sequencing. Nucleic acid techniques, p. 115-175. In E. Strackebrandt and M. Goodfellow (ed), *Bacterial systematics*. John Wiley and Sons, New York, N. Y.
- Lecointre, G., L. Rachdi, P. Darlu y E. Denamur.** 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol Biol Evol* 15(12): 1685-1695
- Lennon, J.J., P. Koleff, J.J.D. Greenwood, K.J. Gaston.** 2001. The geographical structure of British bird distributions: diversity, spatial turnover and scale. *J Anim Ecol* 70: 966-979.
-

- Levins, R.** 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Amer* 15: 237-240.
- Lonetto, M., M. Gribskov, y C.A. Gross.** 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* 174: 3843-3849.
- Louws, F.J., D.W. Fulbright, C.T. Stephens y F.J. de Bruijn.** 1994. Specific Genomic Fingerprints of Phytopathogenic *Xantomonas* and *Pseudomonas* Pathovars and Strains Generated with Repetitive Sequences and PCR. *Appl Environ Microbiol* 60(7): 2286-2295.
- Madigan, M.T., J.M. Martinko y J. Parker.** 2000. Brock: Biology of Microorganisms. Prentice Hall. USA.
- Martin, D.P., C. Williamson y D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260-262.
- Maynard-Smith J., N.H. Smith, M. O'Rourke y B.G. Spratt.** 1993. How clonal are bacteria? *Proc Natl Acad Sci USA* 90: 4384-4388.
- Maynard-Smith, J.** 1992. Analyzing the mosaic structure of genes. *J Mol Evol* 34: 126-129.
- Mayr, R.M.** 1994. Conceptual aspects of the quantification of the extent of biological diversity. *Phil Trans R Soc Lond B* 345: 13-20.
- McCaig A.E., L. Glover y J.I. Prosser.** 1999. Molecular analysis of bacterial community structure and diversity of unimproved and improved upland grass pastures. *Appl Environ Microbiol* 65: 1721-1730.
- McDonald, J.H. y M. Kreitman.** 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652-654.
- Mendel, G.** 1865. "Meteorologische Beobachtungen aus Mähren und Schlesien für das Jahr 1864." *Verhandlungen des naturforschenden Vereines, Abhandlungen, Brünn* 3: 209-20
- Meyer, E. R.** 1973, Late-Quaternary paleoecology of the Cuatro Cienegas Basin, Coahuila, Mexico. *Ecology* 54: 982-995.
- Milkman, R.** 1973. Electrophoretic variation in *Escherichia coli* from natural sources. *Science* 182: 1024-1026
-

- Minckley, T.A. y S.T. Jackson.** 2008. Ecological stability in a changing world? Reassessment of the paleoenvironmental history of Cuatrociénegas, México. *J Biogeogr* 35: 188-190.
- Minckley, W.L.** 1969, Environments of the Bolson of Cuatro Ciengas, Coahuila, Mexico, With Special Reference to the Aquatic Biota: Texas Western Press, The University of Texas at El Paso, 65 pp.
- Morin, P. y J. McGrady-Steed.** 2004. Biodiversity and ecosystem functioning in aquatic microbial systems: a new análisis of temporal variation and species richness-predictability relations. *Oikos* 104: 458-466.
- Morrison, D.A.** 2005. Networks in phylogenetic analysis: new tools for population biology. *Intl J Parasitol* 25: 567-582
- Musser, J.M., J.S. Kroll, D.M. Granoff y otros 25 autores.** 1990. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. *Rev Infect Dis* 12: 75-111
- Musser, J.M., P.M. Shliever, A.W. Chow, P. Ewan, B.N. Kreiswirth, V.T. Rosdahl, A.S. Naidu, W. Witte y R.K. Selander.** 1990. A single clone of *Staphylococcus aureus* causes the majority of cases of toxic shock syndrome. *Proc Natl Acad Sci USA* 87: 225-229
- National Public Radio,** 2001, The Desert Springs of Mexico's Cuatro Cienegas: <http://www.npr.org/programs/re/archivesdate/2001/aug/20010827.cienegas.html>, Archive files, August 27, 2001, Morning Edition with John Burnett.
- Nei, M.** 1975. Molecular population genetics and evolution. North-Holland, Amsterdam.
- Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press. New York.
- Nei, M. y T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and non synonymous nucleotide substitutions. *Mol Biol Evol* 3: 418-426.
- Noguez, A.M. H.T. Arita, A.E: Escalante, L.J. Forney, F. García-Oliva y V. Souza.** 2005. Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Glob Ecol Biogeog* 14: 241-248.

- Ochman, H., T.S. Whittam, D.A. Caugant y R.K. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *J Gen Microbiol* 129: 2725-2726
- Oda, Y., B. Star, L.A. Huisman, J.C. Gottschal y L.J. Forney.** 2003. Biogeography of the purple non sulfur bacterium *Rhodospseudomonas palustris*. *Appl Environ Microbiol* 69: 5186-5191.
- Øvreås L., F.L. Daae, M. Hedal, F. Rodríguez-Valera y V. Torsvik.** 2001. Artículo presentado en el 9th Internacional Symposium on Microbial Ecology: Interaction in the Microbial World, Amsterdam 26 al 31 de Agosto de 2001
- Øvreås L., L. Forney, F.D. Daae y V. Torsvik.** 1997. Distribution of bacterioplankton in meromictic lake Sælevannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* 63: 3367-2273.
- Ovreas, A-L. Reysenbach, V.H. Smith y J.T. Staley.** 2006. Microbial biogeography: putting microorganisms on the map. *Nature Rev Microbiol* 4: 102–112.
- Palleroni, N.J.** 1984. Genus I. *Pseudomonas* Migula 1894. En: Krieg N.R. y Holt J.G. (eds) *Bergey's Manual of Systematic Bacteriology* (pp 141-199) The Williams y Wilkins Co. Baltimore
- Palleroni, N.J.** 1993. *Pseudomonas* classification. *Antonie van Leeuwenhoek* 64: 231-251
- Papke, R.T. y D. M. Ward.** 2004. The importance of physical isolation to microbial diversification. *FEMS Microbiol Ecol* 48: 293-303.
- Papke, R.T., N.B. Ramsing, M.M. Bateson y D.M. Ward.** 2003. Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* 5: 650-659.
- Papke, R.T., O. Zhaxybayeva, E. Feil, K. Sommerfeld, D. Muise y W.F. Doolittle.** 2007. Searching for species in haloarchaea. *Proc Natl Acad Sci USA* 104: 14092-14097.
- Pedersen, K.** 2000. Exploration of deep intraterrestrial life. *FEMS Microbiol Lett* 185: 9-16.
- Piñero, D., E. Martínez, y R.K. Selander.** 1988. Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. *Appl Environ Microbiol* 54: 2825-2832.
-

- Pond, S.L.K., D. Posada, M. B. Gravenor, C.H. Woelk y S.D.W. Frost.** 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 23: 1891-1901.
- Pybus, O.G. y A. Rambaut.** 2002. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18: 1401-1405
- Pybus, O.G., A. Rambaut y P.H. Harvey.** 2000. An Framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155: 1429-1437
- Rahbek, C. y R. Graves.** 2001. Multiscale assessment of patterns of avian species richness. *Proc Natl Acad Sci USA* 98: 4534-4539.
- Reche, I., E. Pulido-Villena, R. Morales-Baquero y E.O. Casamayor.** 2005. Does ecosystem size determine aquatic bacterial richness? *Ecology* 86: 1715-1722.
- Ritz, K., B.S. Griffiths, V.L. Torsvik. y N. B. Hendricksen.** 1997. Analysis of soil and bacterioplankton community DNA by melting profiles and reassociation kinetics. *FEMS Microbiol Lett* 149: 151-156.
- Rius, N., M.C. Fusté, C. Guasp, J. Laucat y J.G. Lorén.** 2001. Clonal population structure of *Pseudomonas stutzeri*, a species with exceptional genetic diversity. *J Bacteriol* 183: 736-744
- Rosenzweig R.F., R.R. Sharp, D.S. Treves y J. Adams.** 1994. Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics* 137: 903-917.
- Rozas, J. y R. Rozas.** 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174-175.
- Sacchi, C.T, A.M. Whitney, L.W. Mayer, R. Morey, A. Steigerwalt, A. Boras, R.S. Weyant y T. Popovic.** 2002. Sequencing of 16S rRNA gene: A rapid tool for identification of *Bacillus anthracis*. *Emer Infec Dis* 8: 1117- 1123.
- Saitou, N. y M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4): 406-425.
- Sarkar, S.F. y D.S. Guttman.** 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl Environ Microbiol* 70(4):1999-2012.

- Schneider, S., D. Roessli y L. Excoffier.** 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Selander, R.K. y B.R. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. *Science* 210: 545-547
- Silva C., P. Vinuesa, L.E. Eguiarte, E. Martínez-Romero y V. Souza.** 2002. *Rhizobium etli* and *Rhizobium gallicum* nodulate common bean (*Phaseolus vulgaris*) in a traditionally managed milpa plot in Mexico: population genetics and biogeographic implications. *Appl Environ Microbiol* 69(2): 884-893.
- Slatkin, M.** 1977. Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoret Pop Biol* 12: 253-262.
- Šimek, K., P. Kojecká, J. Nedoma, P. Hartman, J. Vrba y J.R. Dolan.** 1999. Shifts in bacterial community composition associated with different microzooplankton size fractions in a eutrophic reservoir. *Limnol Oceanogr* 44: 1634-1644.
- Smith, J.M., C.G. Dowson y B.J. Spratt.** 1991. Localized sex in bacteria. *Nature* 349: 29-31.
- Smouse, P.E., J.C. Long y R. Sokal.** 1986. Multiple regression and correlation extensions of the Mantel Test of matrix correspondence. *Syst Zool* 35: 627-632.
- Sneath, H.A. y R.R. Sokal.** 1973. Numerical Taxonomy. W.H. Freeman and Company, San Francisco, pp 230-234.
- Souza, V., L. Espinosa-Asuar, A.E. Escalante, L.E. Eguiarte, J. Farmer, L. Forney, L. Lloret, J.M. Rodríguez-Martínez, X. Soberón, R. Dirzo, y J.J. Elser.** 2006. An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc Natl Acad Sci* 103(17): 6565-6570.
- Souza, V., L.E. Eguiarte, J. Siefert y J.J. Elser.** 2008. Microbial endemism: does phosphorous limitation enhance speciation? *Nature Rev Microbiol* doi:10.1038.
- Souza, V., M. Rocha, A. Valera y L.E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. *Appl Environ Microbiol* 65(8): 3373-3385.

- Souza, V., T.T. Nguyen, R.R. Hudson, D. Piñero y R.E. Lenski.** 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proc Natl Acad Sci* 89: 8389-8393.
- Spiers, A.J., A. Buckling y P.B. Rainey.** 2000. The causes of *Pseudomonas* diversity. *Microbiology* 146: 2345-2350.
- Stackebrandt, E. y B.M. Goebel.** 1994. Taxonomic note: a place for DNA:DNA reassociation and 16SrRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44: 846-849.
- Stanier, R.Y. N.J. Palleroni y M. Doudoroff.** 1966. The aerobic *Pseudomonas* - a taxonomic study. - *J Gen Microbiol* 42; 159-271.
- Sugiura, N.** 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Commun Stat Theory Meth* A7: 13-26.
- Swofford, D.L.** 2002. PAUP\*: Phylogenetic analysis using parsimony and other methods (software). Sunderland, MA: Sinauer Associates.
- Tajima, F.** 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.
- Tajima, F.** 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.
- Tiedje, J.M.** 1996. Microbial Diversity and Ecosystem Function. En: D. Allsopp, R.R. Colwell y D.L. Hawksworth (eds). Cambridge University Press UK 73-87 pp.
- Torsvik V., F.L. Daae., R.-A. Sandaa y L. Øvreås.** 1998. Novel techniques for analysing diversity in natural and perturbed environments. *J Biotechnol* 64: 53-62.
- Torsvik V., L. Øvreås y T.F. Thingstad.** 2002. Prokaryotic Diversity – Magnitude, Dynamics and Controlling Factors. *Science* 296: 1064-1066.
- Torsvik, V., J. Goksoyr y F.L. Daae.** 1990. High diversity of DNA of soil bacteria. *Appl Environ Microbiol* 56: 782-787.
- Tuomisto, H., K. Ruokolainen y M. Yli-Halla.** 2003. Dispersal, environment and floristic variation of western Amazonian forests. *Science* 999: 241-244.
- Vernez, I., P. Hausser, M.V. Bernasconi y D.S. Blanc.** 2005. Population genetic analysis of *Pseudomonas aeruginosa* using multilocus sequence typing. *FEMS Immunol Med Microbiol* 43: 29-35
-

- Versalovic, J., M. Schneider, F.J. de Bruijn y J.R. Lupski.** 1994. Genomic fingerprinting of bacteria using repetitive sequence based PCR (rep-PCR). *Meth Cell Mol Biol* 5: 25-40
- Versalovic, J., T. Koeuth y J.R. Lupski.** 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucl Acids Res* 19: 6823-6831.
- Wagner R.** 1994. The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch Microbiol* 161: 100-109.
- Ward, D.M., M.M. Bateson, R. Weller y A.L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. En: K.C. Marshall (ed), *Advances in microbial ecology*. Plenum Press, New York, N.Y.
- Watt, P.M. y I.D. Hickson.** 1994. Structure and function of type II DNA topoisomerases. *Biochem J* 303: 681-695.
- Watterson, G.A.** 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2): 256-276.
- Wayne, L.G., D.J. Brenner, R.R. Colwell, P.A.D. Grimont, O.Kandler, M.I. Krichevsky, L.H. Moore, W.E.C. Moore, R.G.E. Murray, E.Stackebrandt, M.P. Starr y H.G. Trüper.** 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 35: 463–464.
- Whitaker, R.J., D.W. Grogan y J.W. Taylor.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301: 976-978.
- Whitaker, R.J., D.W. Grogan y J.W. Taylor.** 2005. Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus "islandicus"*. *Mol Biol Evol* 22: 2354-2361.
- Whitman W.B., D.C. Coleman y W.J. Wiebe.** 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* 95: 6578.
- Whittaker, R.J., K.J. Willis y R. Field.** 2001. Scale and species richness: towards a general hierarchical theory of species diversity. *J Biogeog* 28: 453-470.
- Willis, K.J. y R.J. Whittaker.** 2002. Species diversity – Scale matters. *Science* 295: 1245-1247.

- Winsborough, B.** 1990. Some ecological aspects of modern fresh-water stromatolites in lakes and streams of the Cuatro Ciénegas Basin, Coahuila, Mexico: The University of Texas- Austin Dissertation, 341 pp.
- Wise, M.G., L.J. Smickets y J.V. McArthur.** 1995. Genetic structure of a lotic population of *Burkholderia (Pseudomonas) cepacia*. *Appl Environ Microbiol* 61: 11791-1798.
- Woese, C.R.** 1987. Bacterial evolution. *Microbiol Rev* 51: 221-271.
- Woese, C.R., O. Kandler y M.L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87: 4576-4579.
- Wright, S.** 1951. The genetical structure of populations. *Ann Eugen* 15: 323-354
- Xia, X. y Z. Xie.** 2001. DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* 92: 371-373.
- Yamamoto, S. y S. Harayama.** 1998. Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *rpoD* and 16S rRNA genes. *Int J Syst Bacteriol* 48: 813-819.
- Yamamoto, S., H. Kasai, D.L. Arnold, R.W. Jackson, A. Vivian y S. Harayama.** 2000. Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* 146: 2385-2394.
- Yannarell, A.C. y E.W. Triplett.** 2005. Geographic and environmental sources of variation in lake bacterial community composition. *Appl Environ Microbiol* 71: 227-239.
- Zhou J., B. Xia, D.S. Treves, L.-Y. Wu, T.L. Marsh, R.V. O'Neill, A.V. Palumbo y J.M. Tiedje.** 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* 68(1): 326-334.

**Apéndice I.** Descripción de *Pseudomonas caenicarum* sp. nov. una nueva especie aislada de una laguna de evaporación en el valle de Cuatro Ciénegas en Coahuila, México.

*Pseudomonas caenicarum* sp. nov. a novel species isolated from an evaporating lagoon in the Cuatro Ciénegas valley in Coahuila, México.

Ana E. Escalante<sup>1</sup>, Jesús Caballero-Mellado<sup>2</sup>, Lourdes Martínez-Aguilar<sup>2</sup>, Alejandra Rodríguez-Verdugo<sup>1</sup>, Andrea González-González<sup>1</sup>, Jeiry Toribio-Jiménez<sup>3</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 DF.

<sup>2</sup>Programa de Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apartado Postal 565-A, Cuernavaca, Morelos, México

<sup>3</sup>Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Apartado Postal 70-275, México DF.

**Running title:** *Pseudomonas caenicarum* sp. nov.

\*Corresponding author

Valeria Souza Saldivar: (+52-55-56228995)

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

The GenBank EMBL/DDBJ accession numbers for the 16S rRNA sequences of strains

1N<sup>T</sup> and 3N are EU791281 and EU791282, respectively.

## ABSTRACT

Nine Gram-negative, rod-shaped, non-spore-forming isolates with identical or very similar rep-PCR patterns were recovered from an evaporative lagoon in Mexico. Two strains, 1N<sup>T</sup> and 3N, had virtually identical 16S rRNA gene sequences, and on this basis were identified as members of the genus *Pseudomonas* with *Pseudomonas peli* as closest relative. All of the 9 isolates displayed practically identical whole cell protein profiles, and major fatty acids (C<sub>16:0</sub> and C<sub>17:1w8c</sub>) supported the affiliation of the strains 1N<sup>T</sup> and 3N within the genus *Pseudomonas*. DNA-DNA reassociation values regarding *Pseudomonas peli* and other closely related *Pseudomonas* species were below 15%. Physiological and biochemical tests allowed phenotypic differentiation of the strains analyzed, including strain 1N<sup>T</sup>, from the five phylogenetically closest *Pseudomonas* species. On the basis of the polyphasic taxonomy approach undertaken, the 9 strains analyzed were classified as a novel species, for which the name *Pseudomonas caenicarum* sp. nov. is proposed, with strain 1N<sup>T</sup> (= LMGXXX = CIP XXX) as type strain.

The genus *Pseudomonas* was early described by Migula (1894). Since then there has been a steady increase in the description of a considerable number of *Pseudomonas* species, which has reached over 191 validly described species at present (Anzai *et al.*, 2000; Bacterial Nomenclature Up-to-Date available at URL: <http://www.dsmz.de/bactnom/bactname.htm>). *Pseudomonas* species are ubiquitous in nature and can be isolated from humans, clinical samples, and from the rhizosphere of plants, or found in soil, and water, including sea-water (Spiers *et al.*, 2000; Yamamoto *et al.*, 2000).

In the present study, a polyphasic approach was undertaken to determine the taxonomic status of 9 isolates recovered from an evaporative lagoon in the Churince system, a hydrological system of the Cuatro Ciénegas valley in Coahuila, México (26°50.830'N, 102°09.335'W).

A large collection of bacterial isolates was obtained from surface water samples collected in two different points, 10 m apart each one, in the “Laguna Grande” lagoon. Water samples of 15ml from each site were collected in sterile BD Falcon™ vials (BD Biosciences, MA, USA). 200 µl from each sample were inoculated on GSP agar plates Kielwein (1969, 1971), and then incubated at 29°C for 48 h. Grown colonies on different GSP plates were purified by subculturing on LB agar plates, and the cultures maintained at -80°C in LB broth with 20% (w/v) glycerol prior to analysis.

Isolates were grown in 5ml of LB liquid medium. DNA was extracted by using DNeasy Blood & Tissue Kit (Qiagen, CA, USA) according to the manufacturer's instructions. Repetitive sequence-based (rep) PCR profiles of the isolates and reference

strains were determined with BOX primers and conditions described previously (Versalovic *et al.*, 1991, 1994). The PCR conditions consisted of an initial denaturing of 95°C for 7 min, 30 cycles of 94°C for 1min, 53°C for 1 min, 65°C for 8 min and final extension of 65°C for 8 min. PCR products were electrophoresed in 1.5% agarose gels (TAE 0.5X) for 5 hours at 180mV (5 V/cm). Gels were stained with ethidium bromide and a digital image obtained for further analysis. Isolates that displayed the same number of bands and similar fragment size were considered identical and belonging to a same bacterial group or genotype. On this basis, 9 isolates with an identical or very similar rep-PCR pattern were identified, all of them recovered from different water samples. 16S rRNA gene sequences were obtained from two strains, 1N<sup>T</sup> and 3N, in order to define their taxonomic affiliation. The nearly complete 16S rRNA gene sequences were obtained by PCR amplification using the 27F and 1492R primers and conditions described previously (Lane 1991). 16S rRNA gene sequences (ca. 1450 bp) were obtained (High-Troughput Sequencing Enterprise of the University of Washington, USA). These sequences were compared to all publicly available complete 16S rRNA sequences of the type strains of validly described *Pseudomonas* species using the Ribosomal Database Project II (RDPII) database and software (Cole *et al.*, 2007). 16S rRNA sequence similarity analysis indicated that the closest relatives to strain 1N<sup>T</sup> (GenBank acc. no. EU791281) and 3N (acc. no. EU791282) were *P. peli* (98.5%), *P. argentinensis* (97.5 %), *P. flavescens* (97.4 %), and *P. anguilliseptica* (97.4 %). Very recently, a novel *Pseudomonas* species, *P. guineae*, has been described (Bozal *et al.*, 2007), and this one showed 97.9% identity with strain 1N<sup>T</sup>. The 16S rRNA phylogenetic tree shown in Fig. 1 illustrates the position of strains 1N<sup>T</sup> and 3N relative to the nearest *Pseudomonas* species and other ones related; both strains were clustered within a single group, clearly separated from the cluster formed by the closely related

species *P. peli*, *P. anguilliseptica* and *P. guineae*, as well as from the cluster formed by *P. argentinensis* and *P. flavescens*. These data suggested that the 9 isolates analyzed in the present study could constitute a novel *Pseudomonas* species.

Preparation of whole-cell proteins from all of the isolates and SDS-PAGE assays were performed as described previously (Estrada-de los Santos *et al.* 2001).

*Pseudomonas* isolates recovered from the lagoon showed almost identical protein profiles between them, however, their protein patterns were notably different from those of *P. anguilliseptica* and *P. peli*, as well as from other closely related *Pseudomonas* species (Fig. 2). It has previously been noted that bacteria with identical or similar protein patterns possess high genome similarity (Vandamme *et al.* 1996). On this basis, the SDS-PAGE results strongly suggested that all of these isolates represent a novel *Pseudomonas* species.

To identify qualitatively and quantitatively cellular fatty acids, strains 1N<sup>T</sup> and 3N were cultivated on LB for 2 days at 29°C. The whole-cell fatty acid composition was determined by using a gas chromatograph and the fatty acid methyl ester (FAME) analysis was done with Sherlock 4.5 Microbial Identification System (Microbial ID, Newark, DE, USA). Fatty acid composition of the strains 1N<sup>T</sup> and 3N, together with type strain data of the closest phylogenetic species, is shown in Table 1. The major fatty acid composition supports the affiliation of the strains 1N<sup>T</sup> and 3N within the genus *Pseudomonas*, and differences in relative abundance of fatty acids with *P. anguilliseptica*, *P. flavescens*, *P. guineae* and *P. peli* are shown as well. The strains 1N<sup>T</sup> and 3N contain clear quantitative differences of C<sub>12:0</sub>, C<sub>12:0</sub>3-OH and C<sub>16:0</sub> compared to the closest related species *P. anguilliseptica*, *P. guineae* and *P. peli*. In addition, these

strains present a combined feature (summed feature a = C<sub>16:1</sub>ω7 and C<sub>16:1</sub>ω6c) that is not present in the closely related species (Table 1). Although most of the fatty acids previously reported for the closest *Pseudomonas* species are present in the profile of strains 1N<sup>T</sup> and 3N, it is noticeable the absence of an unknown fatty acid, ECL 11.799, and fatty acid components in summed feature b (C<sub>16:1</sub>ω7 and iso-C<sub>15:0</sub> 2-OH).

To clarify taxonomic relationships at the species level, DNA-DNA hybridization experiments were performed as described previously (Estrada-de los Santos *et al.* 2001), which were based on relative levels of hybridization to <sup>32</sup>P-labeled DNA from strain 1N<sup>T</sup>. DNA-DNA relatedness assays were performed with four strains recovered from the lagoon, as well as with the type strain of the closest-related *Pseudomonas* species as indicated by 16S rRNA gene sequence data. DNA–DNA reassociation values between the strains recovered from the lagoon were in the range of 87-99 % [3N (87 %), 2N (96 %), and 5N (97 %)], indicating a close relationship at the species level (Vandamme *et al.*, 1996; Stackebrandt *et al.*, 2002). In contrast, low reassociation values, lower than 15 %, were obtained in hybridizations of strain 1N<sup>T</sup> with *P. peli* LMG 23201<sup>T</sup> (14 %), *P. anguilliseptica* LMG 21629<sup>T</sup> (14 %), *P. flavescens* LMG 18387<sup>T</sup> (13 %), and *P. stutzeri* LMG 11199<sup>T</sup> (12 %). These DNA-DNA reassociation data, together with the 16S rDNA sequences analysis, as well as SDS-PAGE protein patterns and fatty acids profiles, support the notion that the isolates group recovered from the evaporating lagoon belong to a new species within the genus *Pseudomonas*, for which the name *Pseudomonas caenicarum* is proposed.

In order to estimate the genome size of *P. caenicarum* sp. nov. strains 1N<sup>T</sup>, 2N, 3N, 6N, 8N, 9N, 10N and 11N were analyzed by Pulse Field Gel Electrophoresis (PFGE)

with I-*CeuI* endonuclease. Genomic DNA preparation and digestion of whole genome were carried out as described previously by Matushek *et al.* (1996) and some modifications suggested by Delgado *et al.* (in press). For *P. caenicarum* sp. nov. mean genome size was  $2995\pm 663$  kb, a similar genome size to those for *P. peli* (3450 kb) and *P. flavescens* (2665 kb), but notably smaller than *P. aeruginosa* PAO1 (6262 kb). Four RNA operons were detected in *P. caenicarum* sp. nov. 1N<sup>T</sup> (Supplementary Figure S1).

Bacteriological and biochemical characterization was carried out growing the isolates in LB medium for 48-72 h at 29°C. Phenotypic features were assayed with the API20NE and API50CH systems according to the manufacturer instructions (bioMérieux). The API 20NE system was used to determine nitrate reduction, gelatin liquefaction, aesculin hydrolysis and urease activity. All colonies were grown on LB agar plates to determine oxidase reaction, as a complementary test in API 20NE system. Data of carbon source usage were obtained with API 50CH system after incubation for 72 h at 29°C. Results of these tests showed that *P. caenicarum* strains differ from the closest related species in the utilization of specific carbon sources, for example, D-xylose, D-fucose and gluconate; complete data are described in Table 2.

#### Description of *Pseudomonas caenicarum* sp. nov.

*Pseudomonas caenicarum*. [qua-dri-cae-ni-caen'is. L. gen. n. *caenicarum* pertaining to the “Cuatro Ciénegas” valley, which is the region where the evaporative lagoon is located, and from where the described organism was isolated].

Cells are motile with one polar flagellum, rods ( $1.2\pm 0.13$  µm long and  $0.53\pm 0.02$  µm wide). Gram negative, oxidase and catalase positive. Good growth occurs on LB and GSP agar at 28-32°C, beige, mucoid and irregular colonies. Fatty acid profile shows

major presence of C<sub>18:1</sub>ω<sub>7</sub>c and also the presence of two unknown ECLs (11.819 and 12.502), C<sub>15:1</sub>ω<sub>6</sub>c, iso-C<sub>15:0</sub> 3-OH, iso-C<sub>16:0</sub>, 11 methyl 18:1<sub>ω</sub>7c and C<sub>16:1</sub>ω<sub>6</sub>c. Carbon source usage is indicated in Table 2. The use of D-xylose, rhamnose, D-fucose and potassium gluconate are the differentiation characters of the described species. The type strain is 1N<sup>T</sup> (= LMG XXX = CIP XXX). This strain was isolated from an evaporative lagoon in Cuatro Ciénegas, Coahuila state, México. Phenotypic and genomic characteristics of the strain 1N<sup>T</sup> are the same as described above for the species.

### **ACKNOWLEDGEMENTS**

We thank R. González-Chauvet, G.M. Rosas-Barrera, A. Casamitjana and Modern American School for sample collection and technical assistance. We gratefully acknowledge Dr. Bulmaro Reyes-Coria for help us with etymological construction of the new species name described in this paper, Gabriela Delgado Sapien for molecular techniques help, Araceli Patrón-Soberano for micrographs, and Luis E. Eguiarte for comments and ideas. This project was supported by grants from SEMARNAT/CONACyT to V.S. (2002-CO1-0246) and CONACyT-UNAM scholarship to A.E.E.

## References

- Anzai, Y., Kim, H., Park, J-Y., Wakabayashi, H. & Oyaizu, H. (2000). Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. *Int J Syst Evol Microbiol* **50**, 1563-1589.
- Bozal, N., Montes, M.J. & Mercadé, E. (2007). *Pseudomonas guineae* sp.nov., a novel psychrotolerant bacterium from an Antarctic environment. *Int J Syst Evol Microbiol* **57**, 2609-2612.
- Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam-Syed-Mohideen. A.S., McGarrell, D.M., Bandela, A.M., Cardenas, E., Garrity, G.M. & Tiedje, J.M. (2007), The Ribosomal Database Project (RDP-II): Introducing myRDP Space and quality controlled public data. *Nucl Acid Res* **35**, D169-D172.
- Delgado, S.G., Souza, V., Morales, R. & Cravioto A. (2008). Genomic Rearrangements at *rrn* Operons in *Citrobacter freundii*. *in press*.
- Estrada-de los Santos, P., Bustillos-Cristales, R. & Caballero-Mellado, J. (2001). *Burkholderia*, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Appl Environ Microbiol* **67**, 2790–2798.
- Kielwein, G., Gerlach, R. & Johne, H. (1969). Prevalence of *Aeromonas hydrophila* in raw milk. *Archiv furLebensmittelhygiene* **20**, 34-38.
- Kielwein, G. (1971). Die Isolierung und Differenzierung von Pseudomonaden aus Lebensmitteln. *Arch G. Lebensmillehyg* **22**, 29-37.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.

- Lane, D.J. (1991). 16S/23S rDNA sequencing. *Nucleic acid techniques*, p. 115-175. In E. Strackebrandt and M. Goodfellow (ed), *Bacterial systematics*. John Wiley and Sons, New York, N. Y.
- Matushek, M.G., Bonten, M.J. & Hayden, M.K. (1996). Rapid preparation of bacterial DNA for pulsed-field gel electrophoresis. *Journal of Clinical Microbiology* **34**, 2598-2600.
- Migula, W. (1894). Ube rein neues System der Bakterien. *Arb Bakteriol Inst Karlsruhe* **1**, 235–238 (in German).
- Spiers, A.J., Buckling, A. & Rainey, P. (2000). The causes of *Pseudomonas* diversity. *Microbiology* **146**, 2345-2350.
- Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kampf, P., Maiden, M.C.J., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C. & Whitman, W.B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043-1047.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* **60**, 407–438.
- Vanparys, B., Heylen, K., Lebbe, L. & De Vos, P. (2006). *Pseudomonas peli* sp. nov. and *Pseudomonas borbori* sp.nov., isolated from a nitrifying inoculum. *Int J Syst Evol Microbiol* **56**, 1875–1881.
- Versalovic, J., Koeuth, T. & Lupski, J.R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucl Acids Res* **19**, 6823-6831.

Versalovic, J., Schneider, M., de Bruijn, F.J. & Lupski, J.R. (1994). Genomic fingerprinting of bacteria using repetitive sequence based PCR (rep-PCR). *Meth Cell Mol Biol* **5**, 25-40.

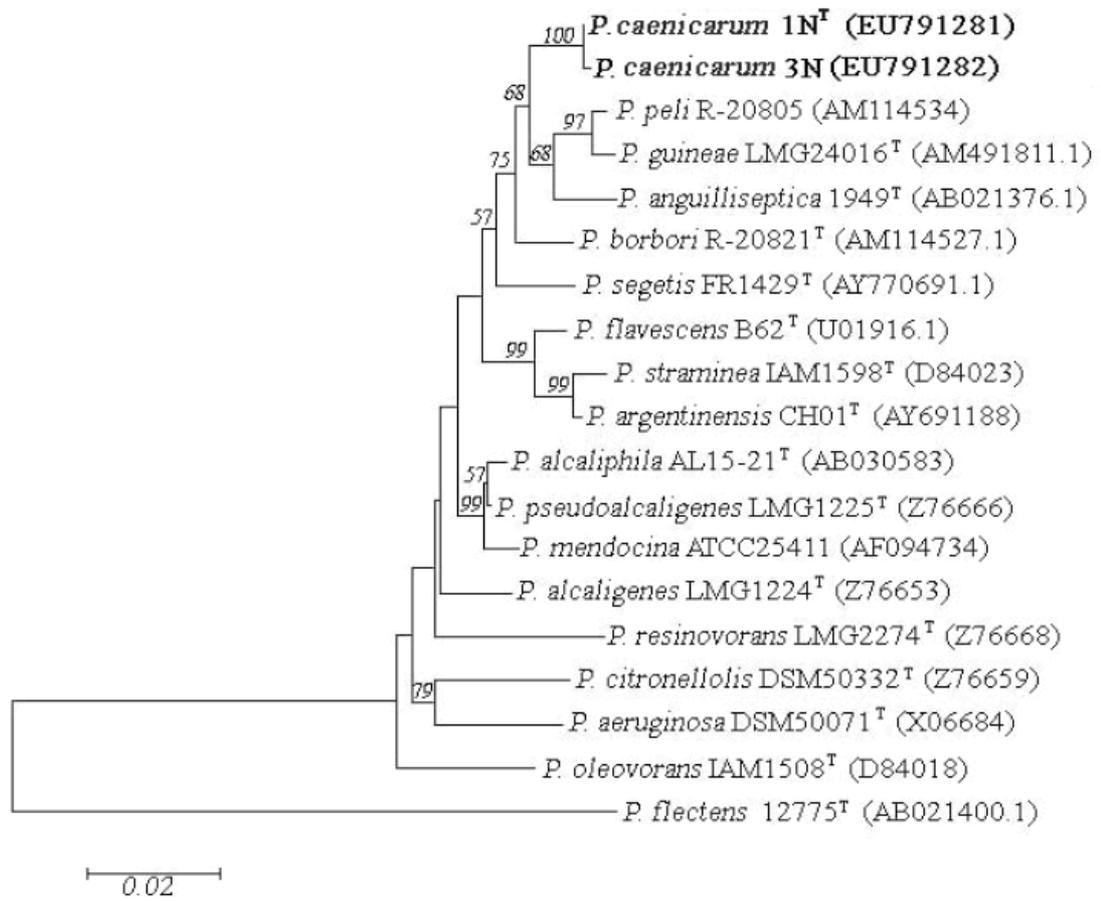
Yamamoto, S., Kasai, H., Arnold, D.L., Jackson, R.W., Vivian, A. & Harayama, S. (2000). Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. *Microbiology* **146**: 2385-2394.

Figure 1. Phylogenetic position of the analyzed isolates (1N<sup>T</sup>, 3N) based on the Kimura-2 parameter distance and Neighbor-Joining clustering after multiple alignment (1480bp) of the 16S rRNA gene sequences of the isolates and closest related members of the genus *Pseudomonas*. Bootstrap values higher than 50% (expressed as percentages after 1000 replications) are shown at branch points. Phylogenetic analysis was performed using the software package MEGA (Molecular Evolutionary Genetics Analysis) version 3.1 (Kumar *et al.*, 2004).

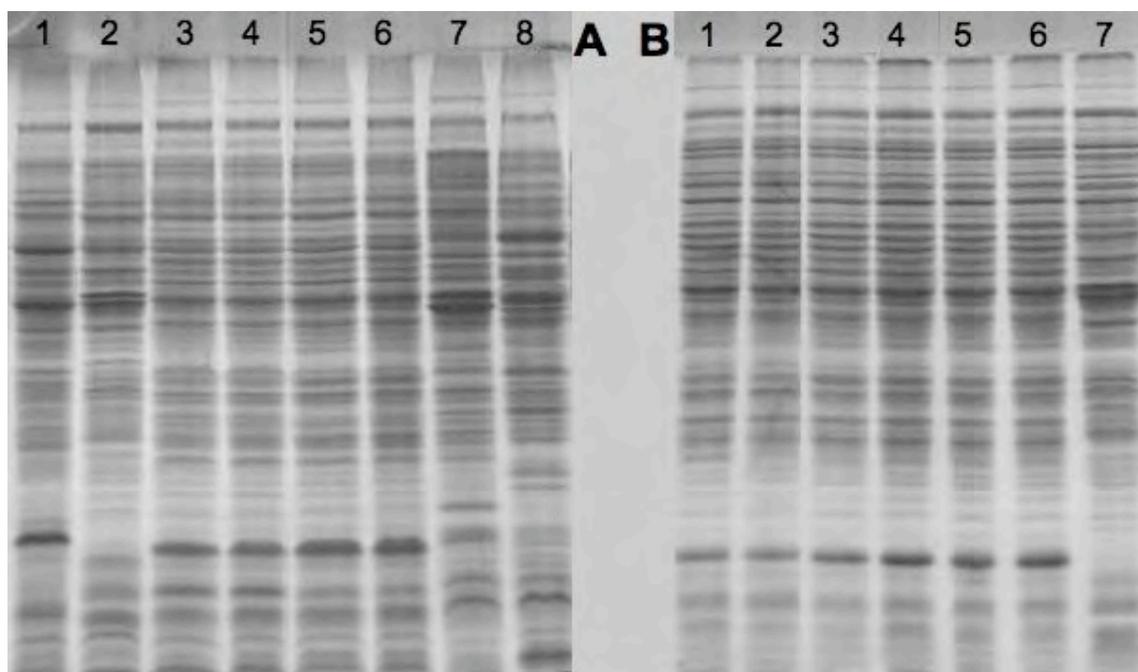
Figure 2. Whole-cell protein profiles of representative *P. caenicarum* strains recovered in the present study, type and reference strains of known *Pseudomonas* species. A) Lane 1, *P. anguilliseptica* LMG21629<sup>T</sup>; lane 2, *P. peli* LMG 23201<sup>T</sup>; lanes 3-6, *P. caenicarum*. 1N<sup>T</sup>, 11N, 10N, 9N; lane 7, *P. stutzeri* LMG 11199<sup>T</sup>; lane 8, *P. flavescens* LMG 18387<sup>T</sup>. B) Lanes 1-6, *P. caenicarum*, 1N<sup>T</sup>, 3N, 2N, 9N, 8N, 6N; lane 7, *P. peli* LMG 23201<sup>T</sup>.

Figure S1. Pulse Field Electrophoresis Gel of representative *P. caenicarum* strains recovered in the present study, type and reference strains of known *Pseudomonas* species: Lane 1, *Escherichia coli* strain O157:H7; Lane 2, *E. coli* strain K12; Lanes 3-10, *P. caenicarum* 1N<sup>T</sup>, 2N, 3N, 6N, 8N, 9N, 10N, 11N; Lane 11, *P. peli* LMG 23201<sup>T</sup>; Lane 12, *P. flavescens* LMG 18387<sup>T</sup>; Lane 13, *P. stutzeri* LMG 11199<sup>T</sup>; Lane 14, *P. aeruginosa* strain PAO1; Lane 15, *E. coli* strain K12

**Fig. 1.**



**Fig. 2.**



**Fig S1**

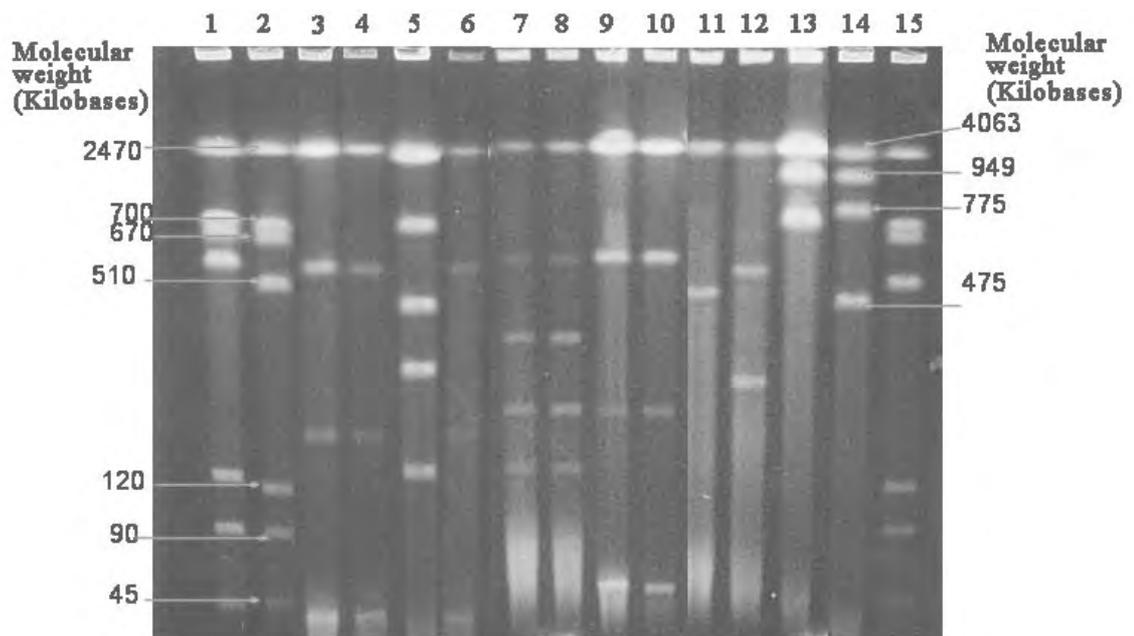


Table 1. Cellular fatty acid composition of the novel species characterized in this study and the closest *Pseudomonas* species. Results for groups of strains are expressed as means with standard deviations in parentheses. *P. caenicarum*: 1, 1N<sup>T</sup> and 2, 3N; 3, *P. peli* (LMG3201<sup>T</sup>, R20815); 4, *P. anguilliseptica* (LMG212629<sup>T</sup>); 5, *P. guineae* (three replicate lectures of M8<sup>T</sup>); 6, *P. flavescens* (LMG1837<sup>T</sup>). Data for *P. peli*, *P. anguilliseptica* and *P. flavescens* were taken from Vanparys *et al.* (2006), values for *P. guineae* were taken from Bozal *et al.* (2007). tr, trace (<0.1%); ---, not detected.

Fatty acid	1	2	3	4	5	6
C <sub>10:0</sub> 3-OH	3.11	3.17	4.10(0.23)	3.17	3.47(0.58)	5.23
Unknown (ECL 11.799)	---	---	4.01(0.06)	3.14	3.70(0.39)	tr
C <sub>12:0</sub>	8.08	8.08	4.92	3.87	3.99(0.49)	11.7
C <sub>12:0</sub> 3-OH	3.69	3.67	4.17(0.28)	2.7	3.61(0.28)	4.85
C <sub>16:0</sub>	14.85	15.16	10.01(0.04)	23.17	10.92(3.08)	19.15
C <sub>17:1</sub> ω8c	tr	tr	3.17(0.04)	tr	tr	tr
C <sub>18:0</sub>	1.03	1.00	tr	tr	tr	tr
C <sub>18:1</sub> ω7c	31.29	31.46	26.53(0.61)	33.57	31.02(1.08)	33.82
Summed feature a	34.79	34.70	---	---	---	22.07
Summed feature b	---	---	27.55	26.81	37.81(1.33)	

Summed feature a (C<sub>16:1</sub>ω7 and C<sub>16:1</sub>ω6c)

Summed feature b (C<sub>16:1</sub>ω7 and iso-C<sub>15:0</sub> 2-OH)

Table 2. Enzyme activity and assimilation of carbon sources. Characteristics are scored as: +, positive reaction; -, negative reaction. Strains: 1, *P. caenicarum* (1N<sup>T</sup>); 2, *P. caenicarum* (3N); 3, *P. peli* (LMG3201<sup>T</sup>); 4, *P. guineae* (M8<sup>T</sup>); 5, *P. anguilliseptica* (LMG212629<sup>T</sup>); 6, *P. flavescens* (LMG1837<sup>T</sup>). All strains were positive for oxidase and catalase activity and negative for: indole formation, glucose acidification, arginine dihydrolase, urease, aesculin hydrolysis, gelatin hydrolysis,  $\beta$ -galactosidase, N-Acetylglucosamine and assimilation of: D-Glucose, D-Ribose D-Maltose, Adipate, Phenilacetate, Erythritol, D-Arabinose, L-Xylose, D-Adonitol, Methyl-bD-Xylopyranoside, L-Sorbose, Dulcitol, Inositol, D-Sorbitol, Methyl-aD-Glucopyranoside, Methyl-aD-Mannopyranoside, N-Acetylglucosamine, Amygdalin, Arbutin, Esculin ferric citrate, Salicin, D-Cellobiose, D-Lactose (bovine origin), D-Melibiose, D-Saccharose (sucrose), Inulin, D-Melezitose, D-Rafinose, Starch, Glycogen, Xylitol, Gentiobiose, D-Turanose, D-Lyxose, D-Tagatose, L-Arabitol and 5 ketogluconate. Data for *P. guineae* were taken from Bozal *et al.* (2007).

Characteristic	1	2	3	4	5	6
Enzyme activity						
Nitrate reduction	-	-	-	+	-	-
Assimilation of						
Malate	+	+	+	+	+	+
D-Trehalose	-	-	-	-	-	+
L-Arabinose	-	-	-	-	-	+
D-Mannose	-	-	-	-	-	+
D-Mannitol	-	-	-	-	-	+
Caprate	+	+	-	+	+	+
Citrate (trisodic)	+	+	-	+	+	-
Glycerol	-	-	-	-	-	+
D-Xylose	+	+	-	-	-	-
D-Galactose	-	-	-	+	-	+
D-Fructose	-	-	-	-	-	weak +
D-Mannose	-	-	-	-	-	+
L-Rhamnose	+	+	-	-	-	-
D-Mannitol	-	-	-	-	-	+
D-Fucose	+	+	-	-	-	-
L-Fucose	-	-	-	-	-	-
D-Arabitol	-	-	-	-	-	-
Potassium Gluconate	+	+	-	-	-	+
Potassium 2-ketogluconate	-	+	-	-	-	-

## Apéndice II. Ecología molecular en el estudio de comunidades bacterianas: métodos moleculares y de análisis

A pesar de la posibilidad de aislar porcentajes mayores de microorganismos con estrategias novedosas de cultivo, el porcentaje de microorganismos no cultivables sigue siendo un problema para la caracterización de la comunidad total de microorganismos. En consecuencia, se han desarrollado diferentes estrategias para caracterizar a las comunidades microbianas sin necesidad de cultivo y a continuación explicaremos los fundamentos de las más utilizadas.

### 1. PFLA (Phospholipid Fatty Acid Analysis)

En todos los microorganismos encontramos fosfolípidos exclusivamente en las membranas celulares. Estas membranas pueden ser degradadas fácilmente dejando libre al componente fosfolipídico. Los fosfolípidos de las membranas son indicadores importantes de la biomasa microbiana activa y además se ha encontrado que existen lípidos únicos en grupos específicos de organismos, por lo que el análisis de fosfolípidos puede ser y ha sido usado para determinar la diversidad y abundancia microbianas (Hill *et al.* 2000, Widmer *et al.* 2001).

Esta estrategia requiere que la muestra ambiental de estudio sea analizada cuantitativamente por cromatografía de gases (GC) y espectrometría de masas (MS), lo que genera datos sobre la presencia y abundancia de diferentes fosfolípidos (Zelles y Bai 1993; Laczkó *et al.* 1997). Los datos obtenidos sobre los ácidos grasos de las muestras pueden ser comparados con bases de datos públicas (Laczkó *et al.* 1997), lo que permite la asociación de ciertos PFLA's con organismos o grupos específicos.

Diferencias en los patrones de PFLA pueden ser interpretadas en relación a cambios en la composición de las comunidades.

Este método ha sido utilizado con éxito en estudios que demuestran cambios grandes en comunidades microbianas asociadas a prácticas de manejo de suelos (Tunlid *et al.* 1989, Bossio *et al.* 1998).

A pesar de la utilidad de este método de análisis existen algunas limitaciones importantes (Haack *et al.* 1994). La primera es que no se conocen los patrones de ácidos grasos para todos los organismos, por lo que los datos obtenidos de muchas muestras no pueden relacionarse con algún grupo de microorganismos. La segunda limitación es que los patrones de ácidos grasos son muy susceptibles a cambios no necesariamente debidos a la composición de la comunidad, por lo que pueden darse patrones falsos que originen interpretaciones incorrectas. La tercera limitación es que bacterias y hongos producen diferentes cantidades de ácidos grasos dependiendo de las condiciones de crecimiento o estrés, por lo que a pesar de que los patrones de PFLA puedan correlacionarse con la presencia de algunos organismos, esto no significa necesariamente que esos patrones sean únicos para esos grupos en todas las condiciones.

## 2. Técnicas basadas en ácidos nucleicos

De entre todas las moléculas celulares probadas hasta ahora, los ácidos nucleicos han demostrado ser las moléculas más útiles, pues han contribuido con una nueva perspectiva y entendimiento sobre la estructura de las comunidades microbianas. Una de las principales ventajas asociadas a los métodos moleculares con ácidos nucleicos es que puede analizarse a la comunidad microbiana en su totalidad, incluyendo a aquellos organismos que no han podido ser cultivados en el laboratorio. Es por ello que estos

métodos se han vuelto cada vez más importantes en la ecología microbiana (Pickup 1991, Stackebrandt *et al.* 1993, Amann *et al.* 1995, Holben y Harris 1995).

Análisis de baja resolución y amplia escala, como la reasociación de DNA, permiten la determinación de la diversidad genética total de la comunidad bacteriana (Torsvik *et al.* 1996). Electroforesis de productos de PCR de genes de rRNA en geles desnaturizantes de gradiente (DGGE) permiten tener mayor resolución y proveen información sobre cambios en el grueso de la estructura de la comunidad (Muyzer *et al.* 1993). Cuando se combinan análisis de genes de rRNA en DGGE con hibridación usando sondas filogenéticas o secuenciando se logra obtener la afiliación filogenética de los miembros dominantes en número que se encuentran en la comunidad (Øvreås *et al.* 1997). La hibridación fluorescente in situ (FISH) de células microbianas con sondas filogenéticas provee información acerca de la composición taxonómica de la comunidad (Hahn *et al.* 1992, Amann *et al.* 1995). Finalmente, la clonación de productos de PCR de genes de rRNA de la comunidad arroja información específica de la comunidad microbiana no cultivable. Esta estrategia también permite la comparación de la estructura de la fracción cultivable con respecto a la no cultivable (Stackebrandt y Rainey 1995).

De esta manera, los ácidos nucleicos se han utilizado de diferentes maneras para caracterizar a las comunidades. A continuación ampliaremos la descripción sobre las estrategias mencionadas brevemente en párrafo anterior.

a. Cinética de reasociación de DNA.

El DNA de una comunidad microbiana es una mezcla de DNA de diferentes especies que están presentes en diferentes proporciones. Una curva de reasociación de DNA nos da una idea de que tan compleja es esa mezcla. La estrategia consiste en desnaturizar con calor la mezcla total de DNA y mediante la lectura de un espectrofotómetro (la monohebra de DNA y el DNA de doble cadena tienen picos de absorbancia a diferentes

longitudes de onda; ver Torsvik *et al.* 1995) se puede generar una cinética de reasociación de DNA. La cinética de reasociación de DNA es del tipo de segundo orden y cualquiera de este tipo de curvas tiene una constante de reasociación  $C_0t_{1/2}$ , donde  $C_0$  es la concentración molar de nucleótidos en un DNA monohebra al inicio de la reasociación y  $t_{1/2}$  es el tiempo en segundos en el que se da el 50% de la reasociación. Bajo condiciones definidas,  $C_0t_{1/2}$  es proporcional a la complejidad del DNA (heterogeneidad; Britten y Kohne 1968). Estos valores son siempre relativos a la cinética de reasociación de DNA de una sola especie (e.g. DNA de *Escherichia coli*), esta se considera la curva ideal y una curva de una comunidad siempre tendrá pendientes menores que la ideal (ver Fig. AII.1). Aunque en estos experimentos  $C_0t_{1/2}$  no tiene un significado preciso, este valor ha sido usado como similar a índices de diversidad de especies (Torsvik *et al.* 1995).

b. RAPD's (Random Amplified Polymorphic DNA).

El *fingerprint* o huella de DNA puede obtenerse de diferentes maneras, una de ellas es la conocida como RAPD. Dentro de los métodos de huellas de DNA basados en PCR (por sus siglas en inglés, Polymerase Chain Reaction, ver capítulo sobre PCR de este libro) quizás sea este el más ampliamente usado. El PCR de RAPD's utiliza oligonucleótidos de 10 a 12 pares de bases para amplificar segmentos al azar de DNA. RAPD funciona porque sólo un pequeño número de fragmentos (5 a 10) se amplificarán en un rango de longitud que puede ser fácilmente amplificado por PCR (usualmente menor a 3 o 4 kilobases) (Hoelzel y Green 1998).

Cuando el DNA que se encuentra entre las secuencias amplificadas consta de repeticiones en *tandem* y hay variaciones en su longitud, el fragmento amplificado será de longitud variable o polimórfico en la población o comunidad (ver Fig. AII.2).

Esta estrategia puede servir para dar un primer vistazo a la composición de distintas comunidades, sin embargo debemos mencionar que esta técnica es poco reproducible debido a que cualquier mínima variación en las condiciones de PCR o del DNA original puede generar cambios en los patrones, por lo que es recomendable utilizar técnicas adicionales para robustecer los datos sobre la estructura de las comunidades.

c. Técnicas relacionadas con la diversidad de genes de rRNA (SSU)

De las diferentes técnicas que usan a los ácidos nucleicos para estimar la diversidad de comunidades microbianas, la más útil y ampliamente utilizada es la determinación de secuencias o patrones del gen que codifica para la subunidad pequeña de los ribosomas (16S para procariontes y 5S o 18S para eucariontes) (Ward *et al.* 1992).

Esta subunidad pequeña (SSU) de rRNA es adecuada para este tipo de estudios por varias razones. La primera es que se encuentra en las tres formas de vida conocidas: Bacteria, Archaea y Eucarya (Woese *et al.* 1990). La segunda es que son moléculas que tienen regiones altamente conservadas y regiones con variación considerable en su secuencia (Woese 1987) y debido a estas tasas de evolución diferentes se pueden establecer relaciones a diferentes niveles jerárquicos en un análisis comparativo de secuencias (Hill *et al.* 2000). La tercera razón es que el rDNA puede amplificarse fácilmente por PCR y ser secuenciado. La cuarta es que los microorganismos no cultivables pueden ser estudiados y caracterizados sin cultivarse al obtener por PCR su SSU rDNA. Varios estudios han demostrado que más del 90% de los microorganismos que pueden observarse microscópicamente *in situ*, pueden ser extraídos y analizados (Steffan *et al.* 1988, Tsai y Olsen 1992, More *et al.* 1994, Zhou *et al.* 1996, Porteus *et al.* 1997, Steffan y Atlas 1998), en comparación con menos del 0.1% que pueden cultivarse (Hill *et al.* 2000).

Una vez amplificado el gen de rRNA por PCR es posible analizar la diversidad de este gen en la comunidad para tener una idea de la complejidad de la misma siguiendo diferentes estrategias que se explicarán a continuación.

- i. DGGE/TGGE (Denaturing Gradient/ Temperature Gradient Gel Electrophoresis).

Esta técnica permite separar mezclas de productos del PCR que son de la misma longitud pero que difieren en secuencia.

El poder de separación de este tipo de electroforesis se basa en que el comportamiento de desnaturalización del DNA está determinado de manera importante por la secuencia de nucleótidos. Al correr el DNA en el gel, la molécula se mantiene como doble cadena hasta que alcanza la concentración o temperatura desnaturalizante, reduciendo su movilidad en el gel (Hill *et al.* 2000).

En teoría, cualquier gen de rRNA que se encuentre en el DNA total de la comunidad puede ser amplificado por PCR y resuelto en un gel DGGE/TGGE.

En este tipo de análisis cada banda en el gel es considerada como una especie distinta en la comunidad y la intensidad de las bandas es tomada como un reflejo de la abundancia de esa secuencia en la comunidad (ver Fig. AII.3).

Existen sólo algunas consideraciones importantes sobre esta técnica y tienen que ver con la interpretación de los datos obtenidos. En primer lugar es posible que diferentes especies generen una misma banda y que los estimados de diversidad no sean del todo estrictos, por lo que lo ideal es hacer estimaciones de diversidad relativa entre comunidades y considerar filotipos en lugar de especies.

ii. T-RFLP's (Terminal Restriction Fragment Length Polymorphism).

El análisis de fragmentos de restricción terminales es actualmente uno de los métodos más poderosos que existen dentro del campo de la ecología microbiana para comparar rápidamente la diversidad de secuencias de DNA bacteriano amplificado por PCR de muestras ambientales (Marsh 1999, Tiedje *et al.* 1999).

El método se basa en la variación en la posición de sitios de restricción entre las secuencias y en la determinación de la longitud de fragmentos terminales de restricción (TRF's) marcados con fluorescencia por medio de electroforesis en geles de alta resolución en secuenciadores de DNA (Avanis-Aghahani *et al.* 1994, Liu *et al.* 1997). El resultado es una distribución de abundancia de fragmentos de diferentes tamaños (Fig. AII.4). Dentro de las características más importantes de este método está la gran velocidad de análisis de las muestras, lo que permite hacer réplicas de los experimentos y con ello lograr análisis estadísticos (Dunbar *et al.* 2001). El método de TRF puede ser usado para identificar diferenciación de comunidades, para comparación de riqueza relativa de filotipos y estructura de comunidades, y para identificar organismos específicos en una comunidad (Dunbar *et al.* 2001) a partir de la clonación y secuenciación de fragmentos específicos de 16S rRNA de la comunidad.

Sobre las limitaciones del método podemos decir que, a pesar de que es un método rápido de análisis comparativo de comunidades, tiene baja resolución (existe gran probabilidad de que muchos filotipos se encuentren representados por un solo tamaño de fragmento y representen un solo pico en la distribución) y esto puede llevar a estimados poco confiables de riqueza misma que puede ser calculada por otros métodos con mayor resolución filogenética. Esta situación fue puesta a prueba por Dunbar *et al.* (2001) al comparar los estimados de riqueza obtenidos de una librería de clonas de 16S

y de TRFLP's obtenidos de la misma muestra de donde se construyó la librería. La conclusión de Dunbar y colaboradores es que los patrones de TRFLP's son incapaces de proporcionar información confiable sobre riqueza relativa de los filotipos. Sin embargo, el método es muy efectivo en descubrir similitudes entre comunidades y tiene una buena sensibilidad (detecta genomas que se encuentran en concentraciones tan bajas como el 1% del total de la comunidad). Es por ello que este método se recomienda como útil para análisis rápidos de muestras de campo que quieran compararse en cuanto a su composición y parece ser una nueva herramienta útil para investigaciones de ecología microbiana.

iii. ARDRA (Amplified Ribosomal DNA-Restriction Análisis).

ARDRA es una estrategia de *fingerprint* o huella de DNA que se usa para caracterizar comunidades y poblaciones. La técnica consiste en amplificar por PCR el gen de 16S rRNA de la comunidad total, aislar una a una las diferentes copias que se encuentran en la mezcla y posteriormente someter a cada una de las copias a digestiones enzimáticas (restricción), también puede digerirse la muestra total (sin separar las copias) (Martin-Laurent *et al.* 2001).

La manera más común de separar las diferentes copias del el 16S rRNA (o cualquier otro) de la comunidad es clonando los diferentes fragmentos en *Escherichia coli*. Esto se hace al mezclar el PCR total con bacterias susceptibles a adquirir DNA extraño (competentes), de esta manera cada bacteria adquirirá una sola copia del gen y así quedarán separadas las diferentes copias de la mezcla total. Las bacterias que contienen los fragmentos se crecen en cajas de Petri y posteriormente se aísla su DNA junto con el gen de interés que se puede separar del DNA bacteriano. Cada gen obtenido de esta manera se somete a una digestión enzimática que resulta en un patrón de bandas en electroforesis en geles de agarosa (Fig. AII.5). Los patrones de restricción permiten

identificar diferentes grupos (diferentes patrones de restricción no significan necesariamente especies diferentes) o identificar comunidades distintas. Si el ARDRA se hace con copias separadas del gen 16S rRNA, el análisis de los diferentes patrones permite hacer estimaciones de diversidad y/o elegir representantes de los patrones distintos para obtener las secuencias e identificar los grupos taxonómicos específicos a los que pertenecen o con los que se relacionan.

#### iv. Secuencias

Cualquiera de las estrategias de ácidos nucleicos mencionadas pueden concluir con la secuenciación de algunas o todas las copias del gen 16S rRNA en la muestra. Sin embargo, se pueden evitar los análisis antes mencionados y hacer el PCR del DNA total e inmediatamente después secuenciar todas las copias con clonación previa.

El análisis de secuencias en comunidades microbianas requiere siempre separar las diferentes copias de la muestra, ya sea por clonación o cortando las diferentes bandas de geles TGGE.

Obtener la secuencia de 16S rRNA de los diferentes organismos de la muestra tiene varias ventajas. Una de ellas es que se construye una librería de clonas que mantiene las copias separadas y listas para otros análisis en el futuro. Otra ventaja y quizás la más importante para el análisis de comunidades microbianas es que la obtención de secuencias de librerías de clonas es la única manera de obtener estimados robustos sobre diversidad de comunidades complejas (como muchas comunidades microbianas) (Dunbar *et al.* 2000).

La principal limitante hasta ahora para la implementación general de análisis de comunidades con secuencias es el costo económico y de tiempo que esta técnica involucra.

d. Técnicas relacionadas con la expresión de genes (rRNA).

Recientemente varios análisis se han enfocado en la caracterización de comunidades microbianas del suelo basándose en el rRNA en contraste con los genes de rRNA codificados por el rDNA (Hahn *et al.* 1990, Moran *et al.* 1993, Felske *et al.* 1996, Purdy *et al.* 1996, Duarte *et al.* 1998, Felske y Akkermans 1998b). Al igual que el rDNA el rRNA tiene regiones tanto variables como altamente conservadas que permiten la discriminación de taxa a diferentes niveles taxonómicos. Además, el uso de rRNA ofrece tres ventajas principales sobre el uso de técnicas con rDNA.

La primera, debido a que los ribosomas son el sitio de síntesis de proteínas, el contenido de rRNA está directamente correlacionado con la actividad metabólica y la tasa de crecimiento (Wagner 1994), eso significa que una alta proporción de las secuencias detectadas en las muestras corresponderán a microorganismos metabólicamente activos y en crecimiento (Felske *et al.* 1996). La segunda ventaja es que debido a que las secuencias de rRNA normalmente se encuentran en las células en más copias que las secuencias de rDNA deben ser más fáciles de detectar. Y la tercera ventaja es que cuando los ribosomas son extraídos directamente de las muestras ambientales sólo se incluyen en el estudio a los microorganismos vivos y activos (Felske *et al.* 1997, Felske y Akkermans 1998a).

Cuando los amplicones de rRNA se separan en geles DGGE/TGGE, el patrón de bandas sirve como una huella digital de la comunidad microbiana. Si se asume que no hay sesgo en la amplificación, la intensidad de las bandas indica la abundancia en la comunidad de la secuencia de rRNA correspondiente (Felske *et al.* 1998b). Un factor que complica las interpretaciones es que el número de operones de rRNA puede variar entre los grupos taxonómicos (Rosado *et al.* 1997), por lo que es posible que exista heterogeneidad en las secuencias de rRNA dentro de un mismo grupo y en consecuencia

no puede asumirse que cada producto de PCR en el gel corresponda a un organismo diferente, y un mismo organismo puede estar representado por varias secuencias distintas (Hill *et al.* 2000). Por otro lado, la intensidad de la banda puede deberse a muchas copias de una misma secuencia en el mismo organismo o a que existen muchas células del organismo que posee la secuencia. Una manera de distinguir entre estas dos posibilidades es mediante hibridización *in situ* (ver ejemplo en Binder y Liu 1998). Otra manera es correr el mismo gel con productos de PCR de rRNA y rDNA y comparar la intensidad de las bandas para determinar si esta se encuentra relacionada con el número de copias del gen (Felske *et al.* 1997).

A pesar de que esta técnica puede ser muy útil para estudiar comunidades microbianas, existen algunas limitaciones en ella. Una de estas es que es necesario tener mucho cuidado con el almacenamiento y procesamiento de las muestras ambientales, ya que cualquier mínimo cambio en las condiciones ambientales originales puede alterar la actividad metabólica de los microbios (van Winzingerode *et al.* 1997), esto puede ser evitado con el procesamiento inmediato o congelamiento de las muestras. Otra limitación existe en relación a la eficiencia de extracción (Moran *et al.* 1993) y amplificación de rRNA (Suzuki y Giovannoni 1996, Zheng *et al.* 1996, van Winzingerode *et al.* 1997). Existe la posibilidad de que la amplificación sea sesgada por mayor abundancia de algunas secuencias en el DNA original o por mayor homología de ciertas secuencias por los oligonucleótidos del PCR. La mayoría de los estimados son en procariontes y aunque en teoría debería poder analizarse la comunidad eucarionte de rRNA de la misma manera que procarionte, los ribosomas eucariontes parecen ser más complejos en cuanto a los genes que los codifican y su regulación, además de que las bases de datos públicas con secuencias de rRNA casi no tienen representantes

eucariontes, por lo que identificar especies eucariontes con las secuencias de su rRNA es problemático (Hill *et al.* 2000).

e. FISH (Fluorescent in situ hybridization).

Este método ha sido usado principalmente con comunidades procariontes y permite la identificación y cuantificación directas de grupos taxonómicos generales o específicos dentro de su ambiente natural (Amann *et al.* 1995, Assmus *et al.* 1995, MacNaughton *et al.* 1996, Kenzaka *et al.* 1998). Con FISH todas las células son fijadas y su 16S o 23S rRNA se hibridiza con sondas de oligonucleótidos marcados con fluorescencia y taxón específicos. Posteriormente, las células que hibridaron y quedaron marcadas fluorescentemente, pueden verse en un microscopio confocal láser de fluorescencia. Con este método y gracias a que las células completas se hibridan, se evitan artefactos generados por sesgos en la extracción de DNA, amplificación por PCR y clonación (Ludwig *et al.* 1997, Wallner *et al.* 1997, Felske *et al.* 1998a). FISH puede usarse para visualizar organismos no cultivados, y es útil para estudiar la distribución ecológica de los organismos a través de habitats diversos (Ludwig *et al.* 1997, Zarda *et al.* 1997, Wullings *et al.* 1998). Tal vez el aspecto en el que se debe tener más cuidado cuando usamos FISH es en el diseño de la secuencia de las sondas de hibridación, ya que de ello depende que se haga una buena caracterización de la comunidad. Es necesario hacer una selección cuidadosa de las secuencias de los organismos representativos del taxón o taxa que se desea visualizar, pues la sonda será tan buena como lo sea esta selección (Amman *et al.* 1995). Errores en este diseño pueden ocasionar que no puedan visualizarse organismos no cultivables que pertenezcan al grupo de interés o que se den hibridaciones cruzadas con organismos de otros grupos (Hahn *et al.* 1992, MacNaughton *et al.* 1996, Felske *et al.* 1998a).

#### d. *Métodos estadísticos para el análisis de las comunidades bacterianas*

El análisis de cualquier comunidad involucra la determinación estadística de la diversidad a partir de la abundancia relativa de las especies presentes. Es por esto que la estimación de la diversidad en las comunidades bacterianas es actualmente uno de los retos y objetivos más ampliamente perseguidos en ecología microbiana.

Estimar la diversidad entraña dos problemas, determinar el tamaño de muestra necesario para tener representada a la comunidad y elegir el estimador de la diversidad más adecuado.

El problema del tamaño de muestra no es único para las comunidades bacterianas, los ecólogos del mundo macroscópico se han enfrentado al mismo y han diseñado herramientas para lidiar con el muestreo (Heck *et al.* 1975, Colwell y Coddington 1994, Magurran 1995), las cuales en la actualidad intentan aplicarse al mundo microscópico.

Recientemente, Hughes y colaboradores (2001) publicaron una revisión interesante sobre este problema, y evaluaron el tamaño de muestra que en general se requiere para tener un estimado útil sobre la diversidad de las comunidades. Además de esto evaluaron la utilidad de varios estimadores de diversidad para el caso de comunidades bacterianas. Sobre estos puntos centraremos este apartado.

#### ***El tamaño de muestra***

##### A. Curvas de acumulación de especies.

La relación entre el número de tipos (OTU's = especies, filotipos, etc.) observados y el esfuerzo del muestreo nos da información sobre el total de la diversidad de la comunidad que ha sido muestreada. Este patrón puede visualizarse graficando una curva de acumulación que consiste en relacionar el número acumulativo de OTU's observados contra el esfuerzo de muestreo. Este tipo de curvas revelan la calidad del muestreo, pues

ya que todas las comunidades contienen un número finito de especies si se continúan muestreando individuos, las curvas eventualmente alcanzarán una asíntota en el valor real del número de OTU's en la comunidad. Por lo que, entre más cóncava (hacia abajo) sea la curva, el muestreo será mejor (Fig. AII.6).

En la gran mayoría de las comunidades microbianas muestreadas no se alcanza la asíntota de las curvas debido a la enorme diversidad que existe. Sin embargo, aunque sería útil conocer la diversidad real de las comunidades microbianas, la mayor parte de las preguntas relacionadas con diversidad se refieren a cambios en la diversidad que suceden por cambios ambientales bióticos o abióticos, por lo que las respuestas a estas preguntas requieren únicamente estimados de diversidades relativas (Hughes *et al.* 2001).

La aproximación de diversidad relativa para evaluar el efecto de variables ambientales ha sido usada tanto en casos de insectos con diversidades altas a nivel de kilómetros cuadrados, como la de bacterias en la boca humana (Price 1975, Samways 1994) y esta estrategia parece estar difundiéndose y ser útil.

### B. Rarefacción

Esta aproximación ha sido adoptada recientemente por varios microbiólogos (Bills y Polishook 1994, Dunbar *et al.* 1999, Moyer *et al.* 1998). La rarefacción compara la diversidad observada entre sitios, tratamientos o habitats que no han sido muestreados equitativamente. Para hacer una comparación justa de entre sitios con tamaños de muestra distintos se extraen submuestras de la muestra más grande y se calcula la riqueza esperada de especies en las submuestras basándose en la distribución de abundancias de la muestra mayor. El proceso se repite para submuestras de diferentes

tamaños. Una curva de rarefacción muestra el cambio en el valor esperado de riqueza de especies de acuerdo al tamaño de la muestra (Ricklefs y Miller 2000).

### **Estimadores de diversidad:**

#### A. Extrapolación a partir de curvas de acumulación

La mayoría de los métodos de extrapolación a partir de curvas de acumulación usan la curva de acumulación observada para ajustarla a una forma funcional supuesta que modela el proceso de observar nuevas especies si aumenta el esfuerzo de muestreo (Hughes *et al.* 2001). Estos modelos incluyen la ecuación de Michaelis-Menten (Clench 1979, Raaijmakers 1987) y la función exponencial negativa (Soberón y Llorente 1993). El beneficio de estudiar la diversidad con estos métodos es que una vez que las especies se han contado no es necesario contarlas de nuevo y el esfuerzo sólo debe enfocarse en encontrar especies raras nuevas. La desventaja es que para comunidades muy diversas sólo es posible detectar una pequeña fracción de las especies y es posible que varias curvas de acumulación tengan un buen ajuste pero que cada una prediga diferente asíntota (Soberón y Llorente 1993). Esta aproximación de datos de comunidades relativamente bien muestreadas por lo que no parece útil por el momento para comunidades bacterianas (Hughes *et al.* 2001).

#### B. Estimadores paramétricos

Los métodos paramétricos estiman el número de especies no observadas en la comunidad al ajustar los datos del muestreo a modelos de abundancia relativa de especies. Estos modelos incluyen el modelo lognormal (Preston 1948) y el modelo Poisson normal (Bulmer 1974). Al ajustar los datos a estas distribuciones pueden obtenerse los parámetros de la curva, entre ellos la diversidad de la comunidad.

Sin embargo, existen algunos impedimentos para usar estimadores paramétricos para cualquier comunidad. Uno de los más importantes es que se asume que se conoce la distribución de abundancia para la comunidad. A pesar de que muchas comunidades de macroorganismos parecen tener una distribución lognormal de abundancia de especies (Sugihara 1980, Magurran 1995, DeVries *et al.* 1997), aún existe controversia sobre el modelo que se ajusta mejor (Hughes 1986, Harte *et al.* 1999). El problema es mucho mayor las comunidades microbianas, ya que en ausencia de grandes bases de datos sobre diversidad y abundancia no es claro si alguno de estos modelos se ajusta a la distribución real. Otra limitación, sobre todo para microorganismos, es que los estimadores paramétricos requieren grandes series de datos sobre abundancia para evaluar los parámetros de la distribución y en el caso de microorganismos las series de datos más grandes que existen incluyen sólo algunos cientos de individuos (Hughes *et al.* 2001).

### C. Estimadores no paramétricos

La diversidad de una comunidad depende de la riqueza (número de especies) y de la abundancia relativa de estas (equitatividad o *evenness*) (Begon *et al.* 1986).

Los estimadores no paramétricos que existen ponderan de diferente manera a las especies abundantes en relación a las especies raras.

La medida más simple de diversidad que toma en cuenta tanto la abundancia como la riqueza de especies es el índice de diversidad de Simpson. Este índice se calcula a partir de determinar la proporción de individuos de cada especie que contribuyen al total de la muestra (e.g. la proporción es  $iP_i$  para la especie  $i$ ).

$$D = \frac{1}{\sum_{i=1}^s p_i^2}$$

Donde  $S$  es el número total de especies de la comunidad (Begon *et al.* 1986). Este índice está basado en la probabilidad de que si se toman dos individuos, estos sean de la misma especie (Brown y Bowman 2001). Valores cercanos a 1 indican que sólo unas pocas especies predominan en la muestra.

Otro índice de diversidad muy utilizado es el índice de Shannon-Weaver el cual es frecuentemente llamado índice de Shannon-Wiener.

$$H = - \sum_{i=1}^S P_i (\ln(P_i))$$

Este índice calcula la probabilidad de predecir la especie de un individuo tomado al azar de la comunidad. El índice va de 0 para comunidades con una sola especie hasta valores altos para comunidades con muchas especies, cada una con pocos individuos (Brown y Bowman 2001).

Otros estimadores no paramétricos que parecen ser muy prometedores para la ecología microbiana han sido adaptados a partir de estadísticos de marcaje-liberación-recaptura (MRR) para estimar el tamaño de poblaciones animales (Seber 1973, Krebs 1989). Los estimadores no paramétricos basados en métodos MRR consideran la proporción de especies que han sido observadas antes (recapturadas) en relación a aquellas que han sido observadas sólo una vez. En una comunidad muy diversa, la probabilidad de que las especies sean observadas más de una vez será baja y la mayoría de las especies serán representadas por un solo individuo.

Los estimadores Chao1 y ACE (cobertura basada en la abundancia) usan la relación MRR para estimar la diversidad como riqueza total de especies al agregar un factor de corrección al número observado de especies (Chao 1984, Chao *et al.* 1993).

Chao1 estima la riqueza total de especies como:

$$S_{\text{Chao1}} = S_{\text{obs}} + (n_1^2 / 2n_2)$$

En donde  $S_{\text{Chao1}}$  es la riqueza total de especies,  $S_{\text{obs}}$  es el número de especies observadas,  $n_1$  es el número de especies observadas una vez y  $n_2$  es el número de especies observadas dos veces. Chao (1984) hace notar que este índice es particularmente útil para series de datos que tengan poca abundancia en sus clases, como es el caso de los microbios (Hughes *et al.* 2001).

ACE (Chao y Lee 1992) incorpora datos de todas las especies con menos de 10 individuos en lugar de sólo vistas 1 o 2 veces:

$$S_{\text{ACE}} = S_{\text{abund}} + (S_{\text{rare}}/C_{\text{ACE}}) + (F_1/C_{\text{ACE}}) \gamma^2_{\text{ACE}}$$

En donde  $S_{\text{ACE}}$  es la riqueza total de especies,  $S_{\text{abund}}$  es el número de muestras abundantes ( $>10$ ),  $S_{\text{rare}}$  es el número de muestras raras ( $\leq 10$ ),  $C_{\text{ACE}}$  es la cobertura de la muestra (proporción de especies raras).

Tanto Chao1 como ACE subestiman la riqueza verdadera cuando los tamaños de muestra son pequeños (Colwell y Coddington 1994).

En el análisis de los estimadores de diversidad para comunidades bacterianas realizado por Hughes y colaboradores (2001), se calcularon y compararon diversidades microbianas en variedad de ambientes y encontraron que aunque los estimadores dependen del tamaño de la muestra, la mayoría de los estimadores de riqueza se estabilizan con los tamaños de muestra que, en general, se tienen para microbios (200 a 1000 clonas).

Es importante señalar que todas las estimaciones estadísticas tienen sus limitaciones y una de las más importantes y frecuentes es que no existe rigor en la definición de los OTU's (Staley 1997), por lo que estudios distintos que calculen los mismos estimadores en general no son del todo comparables. De igual manera, la mayoría de estas aproximaciones requieren datos sobre frecuencias relativas de los diferentes OTU's y

muchos estudios han revelado que los análisis genéticos de diversidad microbiana van acompañados de sesgos de muestreo. Por ejemplo, los sesgos ya mencionados durante la extracción de DNA y PCR.

Sin embargo, el hecho de que la mayoría de las preguntas sobre estructura y función de las comunidades requieran comparaciones relativas sobrepasa muchos de los problemas sobre definición de especies y sesgos de muestreo (Hughes *et al.* 2001). Mientras que la unidad de medida sea definida y constante, la diversidad entre sitios o tratamientos puede ser comparada. De igual manera, para minimizar el efecto del sesgo de muestreo se pueden usar varias técnicas o genes que robustezcan las comparaciones (Nübel *et al.* 1999b).

*e. Revisión de estudios empíricos, crítica y perspectivas*

Como ya hemos revisado a lo largo de este capítulo, el estudio de las comunidades microbianas inició con la investigación de los organismos que era posible cultivar en el laboratorio, pero esta estrategia ha resultado ser poco eficaz en cuanto a la recuperación que se ha logrado en general ya que como se ha mencionado antes se ha sugerido que sólo el 0.1% de la comunidad total se logra aislar en cultivo (Hill *et al.* 2000). A pesar del poco porcentaje recuperado y estudiado, la diversidad encontrada es muy alta (alrededor de 4500 especies caracterizadas a la fecha, ver Torsvic *et al.* 2002).

Estudios con comunidades cultivables empezaron a darnos una idea de la cantidad de grupos o especies microbianas que pueden existir y de la diversidad y riqueza de especies que puede haber en diferentes habitats. Los típicos trabajos de caracterización de comunidades cultivables se basaron en los diferentes tipos de metabolismo que existían en la comunidad (Garland y Lehman 1999) y más recientemente este mismo enfoque ha sido industrializado en un producto conocido como BIOLOG™. Este producto está basado en el uso de un juego de 95 fuentes diferentes de carbono que han

sido descritas previamente (Garland y Mills 1991). Sin embargo, existen varias consideraciones importantes en el uso de este método para el análisis de comunidades. Una de las más importantes es que los sustratos utilizados por el sistema pueden no ser los relevantes ecológicamente y muy probablemente no reflejen la diversidad de sustratos que encontramos en el ambiente. Además, mientras que los perfiles de la fisiología de las comunidades pueden dar información sobre la diversidad microbiana, este y cualquier otro método basado en cultivo tiene aún el gran sesgo de muestreo generado por la limitante de las estrategias actuales de cultivo.

En prácticamente todos estos estudios se han descubierto nuevos linajes o probado nuevas técnicas de estudio y análisis de comunidades microbianas, además de contribuir con nuevo conocimiento sobre la distribución de varios grupos microbianos, tal es el caso de la presencia de miembros de Archaea (Ueda *et al.* 1995) a pesar de que se consideró por muchos años que los organismos de este grupo sólo se encontraban en ambientes extremos.

La gran mayoría de los trabajos actuales sobre comunidades microbianas buscan resolver preguntas sobre diversidad y el efecto de factores ambientales o humanos sobre ésta, así como la determinación de la diversidad en diferentes habitats (Dunbar *et al.* 1999, 2000, McCaig *et al.* 1999, Hoshino *et al.* 2001, Zhou *et al.* 2002). Además, es de resaltar que el porcentaje de estudios de comunidades bacterianas en relación con el de comunidades de Archaea o microorganismos eucariontes es muy superior (López-García *et al.* 2001).

Por ahora el objetivo primordial de la gran mayoría de los estudios moleculares sobre comunidades es estimar la diversidad por diferentes métodos y comparar datos para determinar la utilidad y alcances de las diferentes estrategias (Nübel *et al.* 1999a; Dunbar *et al.* 1999, 2000, 2001, Kroes *et al.* 1999, Zhou *et al.* 2002). Entre las

estrategias actualmente más utilizadas para determinar la diversidad microbiana se encuentran la construcción de librerías de clonas (Kroes 1999, McCaig 1999, Dunbar *et al.* 2000) y FISH para identificar la distribución de las especies (Hahn *et al.* 1992, Zarda *et al.* 1997, Wagner *et al.* 2002).

Todos estos esfuerzos están contribuyendo rápidamente a la generación de inmensas bases de datos sobre la abundancia y distribución de las especies microbianas. El carácter claramente descriptivo de los trabajos actuales sobre comunidades microbianas se debe a que es un campo que apenas empieza debido a que las herramientas para su estudio son relativamente recientes. Sin embargo, este es un primer paso para cambiar la tendencia que permaneció por mucho tiempo en el estudio de microbios, los microbiólogos pasaron décadas analizando organismos cultivables y caracterizándolos de acuerdo a su metabolismo. Ahora, gracias a los estudios sobre microorganismos no cultivables se sabe que la proporción de conocimiento que estos años de microbiología clásica han dejado es muy poca, aunque valiosa. A pesar de este cambio en la tendencia del estudio de microorganismos, nos encontramos apenas al pie de la montaña. Como hemos visto, por ahora, muchos grupos de investigadores se dedican a hacer listados sobre abundancia y diversidad de las comunidades, a probar nuevos métodos de obtención y análisis de datos y a muestrear la mayor cantidad de habitats diferentes. Es clara entonces la posición de recopilación de información que vivimos en la actualidad. Daría la impresión de que esta labor sería interminable pero poco a poco nos vamos dando cuenta de que no es imposible, las herramientas moleculares y tecnológicas dan la posibilidad de agilizar este proceso. Aún falta por dar el salto hacia interpretaciones que integren la información de estos listados, integración que esperamos nos de cada día una idea más clara de cómo funciona el mundo microscópico. Es de esperarse que todos los esfuerzos actuales para mejorar las técnicas de cultivo, la gran producción de datos

sobre diversidad y abundancia de especies obtenidos con métodos moleculares y el constante avance de la tecnología e informática, nos permitan acelerar el camino para responder las preguntas del mundo microbiano y contar en el futuro con información que ayude a analizar ampliamente la ecología y evolución de esta diversidad en ambientes naturales.

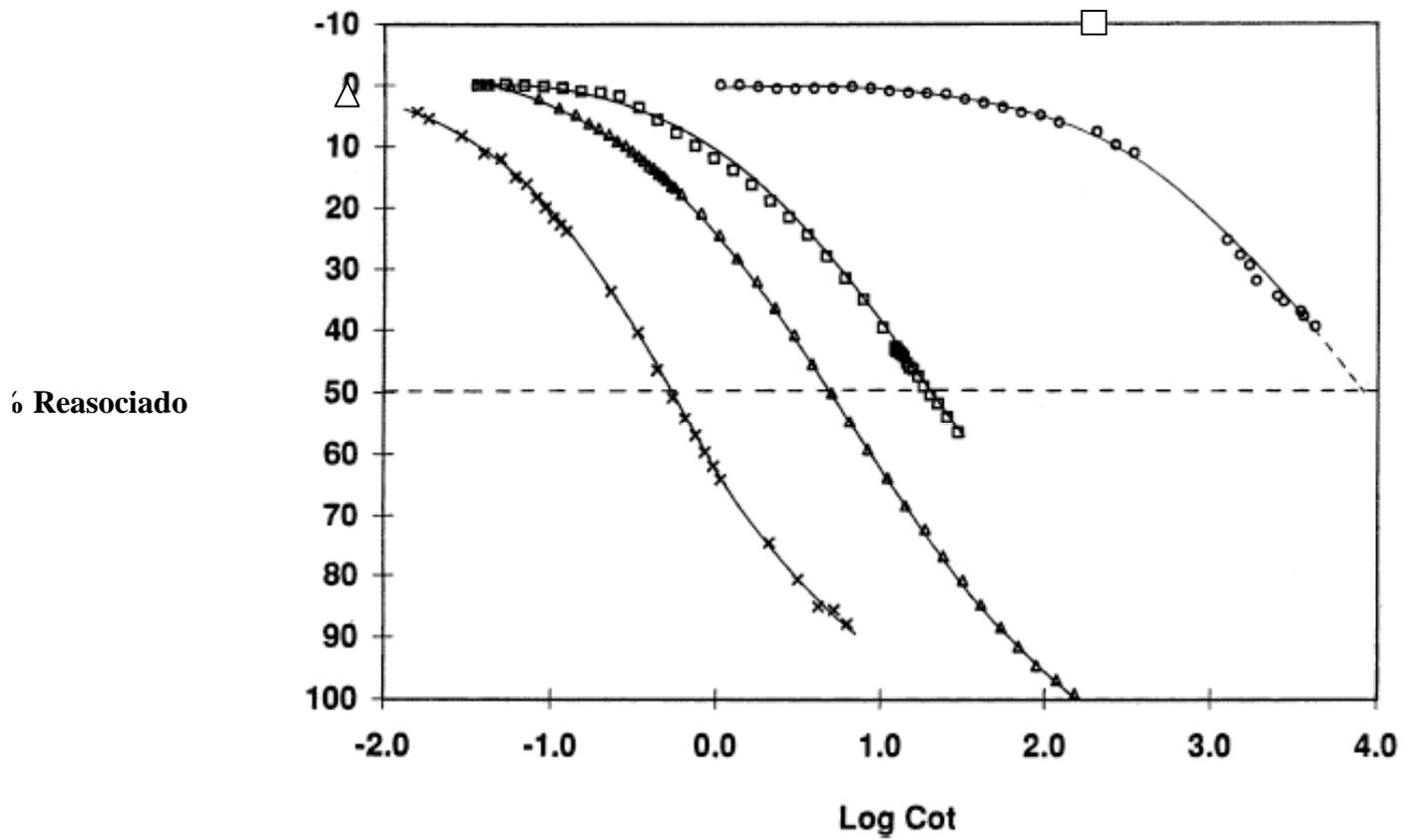


Fig. AII.1. Cinética de reasociación de DNA de la fracción bacteriana de sedimento marino (-o-), mesocosmos con sedimento de granja de peces (-Δ-), mesocosmos con sedimento de granja de peces dados 100ppm de ácido oxolinico (--Δ--), y *E. coli* (-x-) (tomado de Torsvik *et al.* 1998).

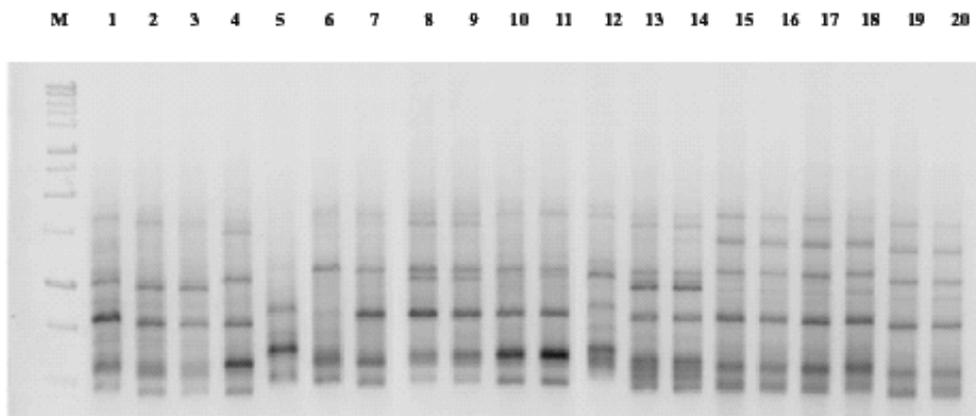


Fig. AII.2. RAPD, perfil de la planta medicinal *Podophyllum peltatum* (tomado de Lata *et al.* 2002).

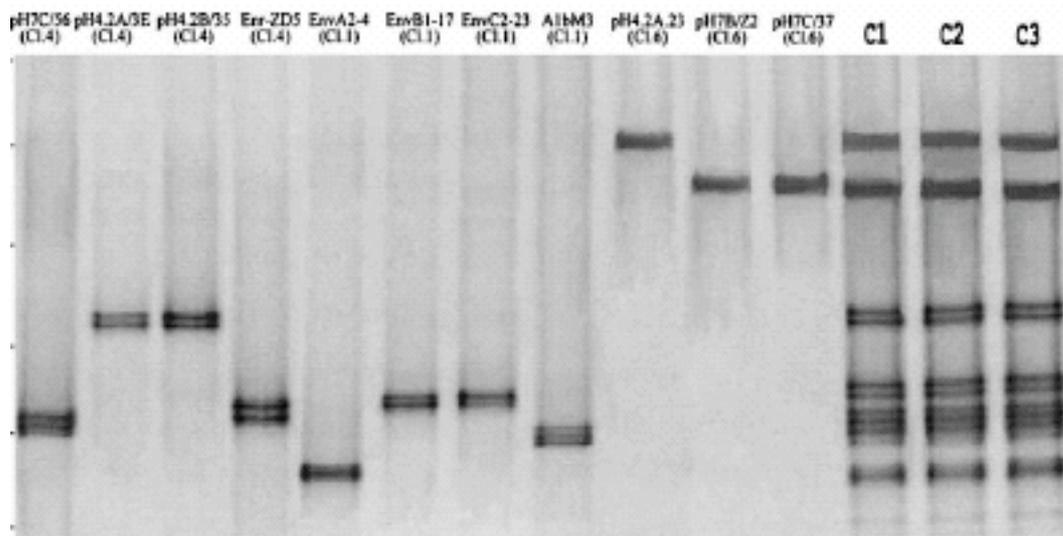


Fig. AII:3. Análisis de DGGE del gen 16S rRNA de bacterias de rizósfera (tomado de Kowalchuk *et al.* 2002).

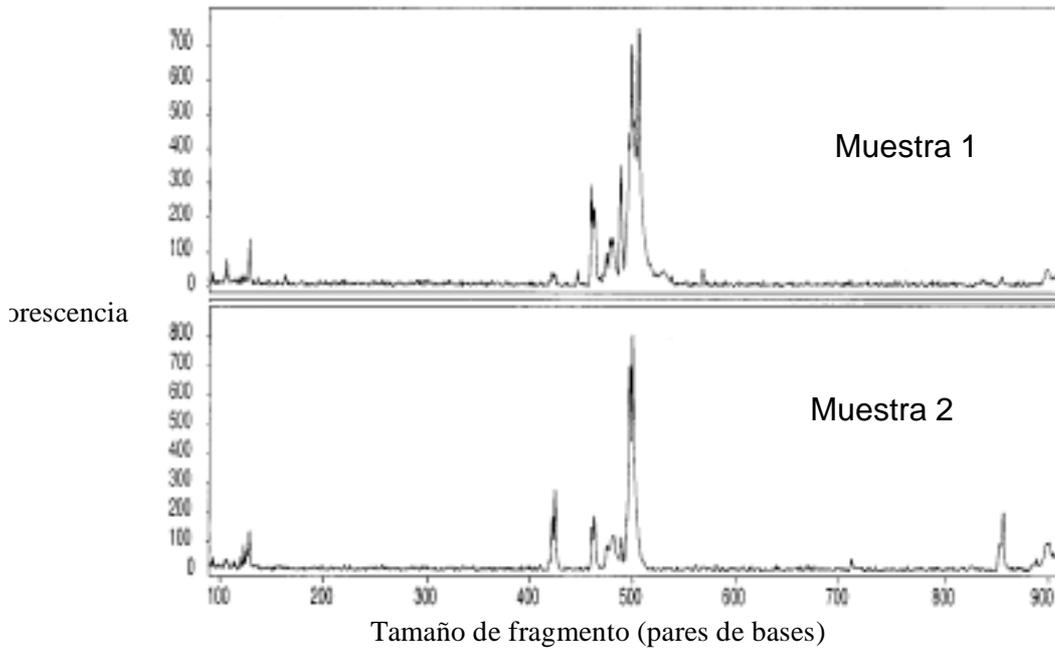


Fig. AII.4. Patrones de T-RFLP's con la enzima Rsa I en 16S rDNA amplificado directamente de muestras de suelo (tomado de Dunbar *et al.* 2000).

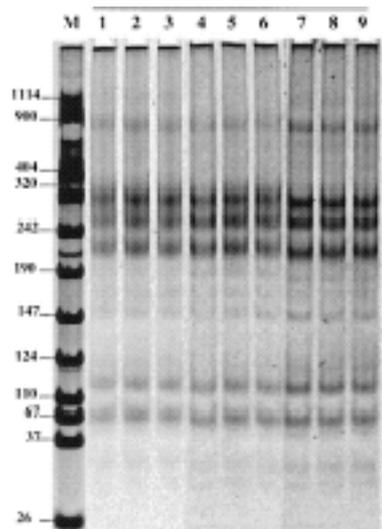


Fig. AII.5. Patrón de ARDRA de muestras ambientales (tomada de Martin-Laurent *et al.* 2001).

proporción  
de OTU's  
observadas

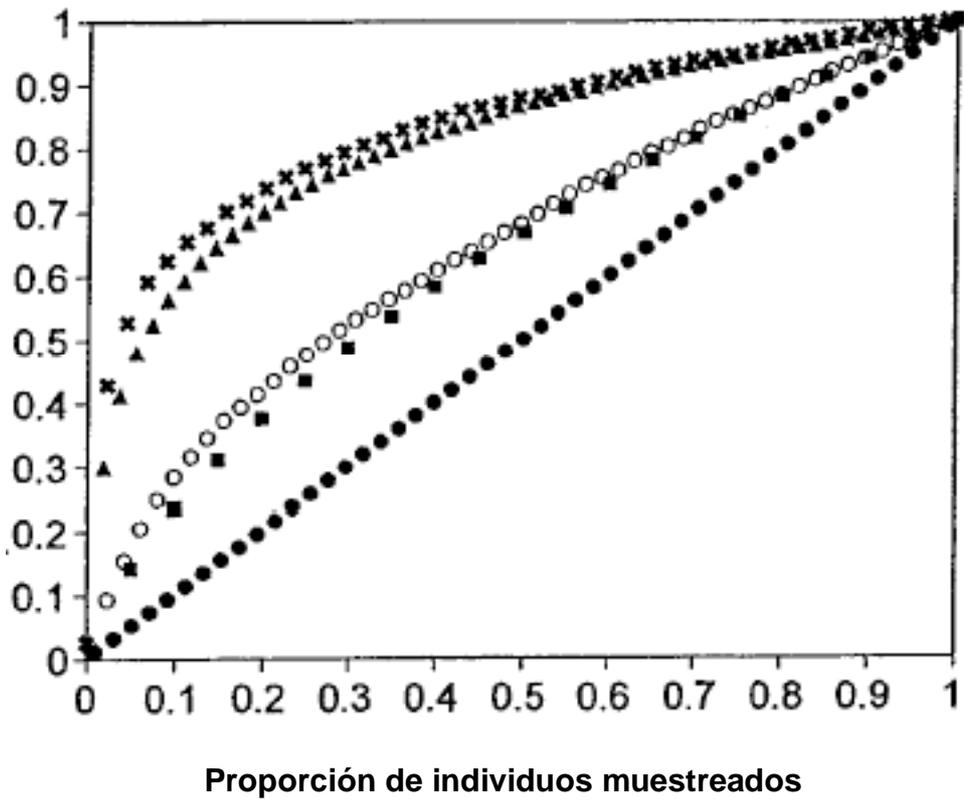


Fig. AII.6. Curvas de acumulación para plantas de Michigan (×), aves de Costa Rica (▲), bacterias de la cavidad oral humana (○), palomillas de Costa Rica (■), bacterias del suelo del este del Amazonas (●) (tomada de Hughes *et al.* 2001)

## Referencias

- Amann R., W. Ludwig y K.H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.
- Assmus B., P. Hutzler, G. Kirchhof, R. Amman, J.R. Lawrence y A. Hartmann.** 1995. *In situ* localization of *Azospirillum brasilense* in the rhizosphere of wheat with fluorescently labeled rRNA-targeted oligonucleotide probes and scanning confocal laser microscopy. *Appl Environ Microbiol* 61: 1013-1019.
- Avaniss-Aghajani, E., K. Jones, D. Chapman y C. Brunk.** 1994. A molecular technique for identification of bacteria using small subunit ribosomal RNA sequences. *BioTechniques* 17:144-149.
- Bills G.F. y J.D. Polishook.** 1994. Abundance and diversity of microfungi in leaf-litter of a lowland rain-forest in Costa Rica. *Mycologia* 86: 187-198.
- Binder, B.J. y Y.C. Liu.** 1998. Growth rate regulation of rRNA content of marine *Synechococcus* (cyanobacterium) strain. *Appl Environ Microbiol.* 64:3346-3351.
- Bossio D.A., K.M. Scow, N. Gunapala y K.J. Graham.** 1998. Determinants of soil microbial communities: effects of agricultural management season and soil type on phospholipid fatty acid profiles. *Microb Ecol* 36: 1-12
- Britten R.J. y D.E. Kohne.** 1968. Repeated sequences in DNA. *Science* 161: 529-540
- Brown, M.V. y J.P. Bowman.** 2001. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol* 35: 267-275.
- Bulmer, M.G.** 1974. On fitting the Poisson lognormal distribution to species abundance data. *Biometrics* 30: 101-110.
- Chao, A.** 1984. Non-parametric estimation of the number of classes in a population. *Scand J Stat* 11: 265-270.
- Chao, A. y S-M Lee.** 1992. Estimating the number of classes via sample coverage. *J Am Stat Assoc* 87: 210-217.
- Chao A., M-C. Ma y M.C.K. Yang.** 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrics* 43: 783-791.
- Clench, H.** 1979. How to make regional lists of butterflies: Some thoughts. *J Lepid Soc* 33: 216-231.
- Colwell R.K. y J.A. Coddington.** 1994. Estimating terrestrial biodiversity through extrapolation. *Phil Trans R Soc London B* 345: 101-118.
- DeVries P.J., D. Murray y R. Lande.** 1997. Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an Ecuadorian rainforest. *Biol J Linnean Soc* 62: 343-364.
- Duarte G.F., A.S. Rosado, L. Seldin, A.C. Keijzer-Wolters y J.D. van Elsas.** 1998. Extraction of ribosomal RNA and genomic DNA from soil for studying the diversity of the indigenous bacterial community. *J Microbiol Meth* 32: 21-29.
- Dunbar J., S. Takala, S.M. Barnes, J.A. Davis y C.R. Kuske.** 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol* 65: 1662-1669.
- Dunbar J., L.O. Ticknor y C.R. Kuske.** 2000. Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S rRNA Gene Terminal Restriction Fragment Analysis. *Appl Environ Microbiol* 66(7): 2943-2950.
- Dunbar J., L.O. Ticknor y C.R. Kuske.** 2001. Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction Fragment Profiles of 16S rRNA Genes from Bacterial Communities. *Appl Environ Microbiol* 66(7): 2943-2950.

- Felske A., B. Engelen, U. Nubel y H. Backhaus.** 1996. Direct ribosome isolation from soil to extract bacterial rRNA for community analysis. *Appl Environ Microbiol* 62: 4162-4167.
- Felske A., H. Rheims, A. Wolterink, E. Stakebrandt y A.D.L. Akkermans.** 1997. Ribosome analysis reveals prominent activity of an uncultured member of the class Actinobacteria in grassland soils. *Microbiology* 143: 2983-2989.
- Felske A. y A. D.L. Akkermans.** 1998a. Prominent occurrence of ribosomes from an uncultured bacterium of the Verrucomicrobiales cluster in grassland soils. *Lett Appl Microbiol* 26: 219-223.
- Felske A. y A. D.L. Akkermans.** 1998b. Spatial homogeneity of abundant bacterial 16S rRNA molecules in grassland soils. *Microb Ecol* 36: 31-36.
- Felske A., A.D.L. Akkermans y W. De Vos.** 1998a. In situ detection of an uncultured predominant *Bacillus* in Dutch grassland soils. *Appl Environ Microbiol* 64: 4588-4590.
- Felske A., A.D.L. Akkermans y W. De Vos.** 1998b. Quantification of 16S rRNAs in complex bacterial communities by multiple competitive reverse transcription-PCR in temperature gradient gel electrophoresis fingerprints. *Appl Environ Microbiol* 64: 4581-4587.
- Garland J.L. y R.M. Lehman.** 1999. Dilution/extinction of community phenotypic characters to estimate relative structural diversity in mixed communities. *FEMS Microbiol Ecol* 30: 333-343.
- Garland J.L. y A.L. Mills.** 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon source-utilization. *Appl Environ Microbiol* 57: 2351-2359.
- Haack S.K., H. Garchow, D.A. Odelson, L.J. Forney y M.J. Klug.** 1994. Accuracy, reproducibility, and interpretation of fatty acid methyl ester profiles of model bacterial communities. *Appl Environ Microbiol* 60: 2483-2493.
- Hahn D., R. Kester, M.J.C. Starrenburg y A.D.L. Akkermans.** 1990. Extraction of ribosomal RNA from soil for detection of *Frankia* with oligonucleotide probes. *Arch. Microbiol* 154: 329-335.
- Hahn D., R.I. Amann, W. Ludwig, A.D.L. Akkermans y K.H. Schelifer.** 1992. Detection of microorganisms in soil after in situ hybridization with rRNA-targeted, fluorescently labelled oligonucleotides. *J Gen Microbiol* 138: 1307-1312.
- Harte J., A. Kinzig y J. Green.** 1999. Self-similarity in the distribution and abundance of species. *Science* 284: 334-336.
- Heck K.L., G.V. Belle y D. Simberloff.** 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sampling size. *Ecology* 56: 1459-1461.
- Hill G.T., N.A. Mitkowski, L. Aldrich-Wolfe, L.R. Emele, D.D. Jurkonie, A. Ficke, S. Maldonado-Ramírez, S.T. Lynch y E.B. Nelson.** 2000. Methods for assessing the composition of soil microbial communities. *Applied Soil Ecology* 15: 25-36.
- Hoelzel A.R. y A. Green.** 1998. PCR protocols and analysis by direct DNA sequencing and PCR-based DNA fingerprinting, en A.L. Hoelzel (ed.) *Molecular Genetic Analysis of Populations – A practical approach*, Oxford University Press, New York. págs. 201-235.
- Holben W.E. y D. Harris.** 1995. DNA-based monitoring of total bacterial community structure in environmental samples. *Mol Ecol* 4: 627-631.
- Hoshino T., N. Noda, S. Tsuneda, A. Hirata y Y. Inamori.** 2001. Direct detection by *in situ* PCR of the *amoA* gene in biofilm resulting from a nitrogen removal process. *Appl Environ Microbiol* 67(11):5261-5266.

- Hughes J.B., J.J. Hellmann, T.H. Ricketts y J.M. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* 67: 4399-4406.
- Hughes, R.G.** 1986. Theories and models of species abundance. *Am Nat* 128: 897-899.
- Kenzaka T., N. Yamaguchi, K. Tani y M. Nasu.** 1998. rRNA –Targeted fluorescent in situ hybridization analysis of bacterial community structure in river water. *Microbiology* 144: 2085-2093.
- Krebs, C.J.** 1989. Ecological methodology. Harper and Row, New York, N.Y.
- Kroes, I., P.W. Lepp y D.A. Relman.** 1999. Bacterial diversity within the human subgingival crevice. *Proc Natl Acad Sci USA* 96: 14547-14552.
- Laczkó E., A. Rudaz y M. Aragno.** 1997. Diversity of antropogenically influenced or distributed soil microbial communities, en H. Insam y A. Rangger (eds.) *Microbial Communities – Functional versus structural approaches*. Springer Verlag, Heidelberg. págs. 57-67.
- Lata, H., R.M. Morales, A. Douglas y B.E. Scheffler.** 2002. Assessment of genetic diversity in *Podophyllum peltatum* by molecular markers. p. 537–544. En: J. Janick and A. Whipkey (eds.), *Trends in new crops and new uses*. ASHS Press, Alexandria, VA.
- Liu, W., T.L. Marsh, H. Cheng y L.J. Forney.** 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* 63: 4516-4522.
- López-García, P., F. Rodríguez-Valera, C. Pedrós-Alió y D. Moreira.** 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Science* 409: 603-607.
- Ludwing W., B.H. Bauer, M. Bauer, I. Held, G. Kirchhof, R. Chulze, I. Huber, S. Spring, A. Hartmann y K.H. Schleifer.** 1997. Detection and *in situ* identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol Lett* 153:181-190.
- MacNaughton S.J., T. Booth, T.M. Embley y A.G. O'Donnell.** 1996. Physical stabilization and confocal microscopy of bacteria on roots using 16S rRNA targeted fluorescent-labeled oligonucleotide probes. *J Microbiol Meth* 26: 279-285.
- Magurran A.E.** 1995. Ecological diversity and its measurement. Princeton University, Princeton, N.J.
- Marsh, T.L.** 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Curr Opin Microbiol* 2:323-327.
- Martin-Laurent F., L. Philippot, S. Hallet, R. Chaussod, J.C. Germon y G. Soulas.** 2001. DNA Extraction from soils: old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* 67(5): 2354-2359.
- McCaig A.E., L. Glover L. y J.I. Prosser.** 1999. Molecular analysis of bacterial community structure and diversity of unimproved and improved upland grass pastures. *Appl Environ Microbiol* 65: 1721-1730.
- Moran M.A., V.L. Torsvik, T. Torsvik y R.E. Hodson.** 1993. Direct extraction and purification of rRNA for ecological studies. *Appl Environ Microbiol* 59: 915-918.
- More M.I., J.B. Herrick, M.C. Silva, W.C. Ghiorse y E.L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. *Appl Environ Microbiol* 60: 1572-1580.
- Moyer C.L., J.M. Tiedje, F.C. Dobbs y D.M. Karl.** 1998. Diversity of deep-sea hydrothermal vent *Archaea* from Loihi Seamount, Hawaii. *Deep-Sea Res II* 45: 303-317.

- Muyzer G., E.C. de-Waal y A.G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* 59: 695-700.
- Nübel U., F. García-Pichel, M. Kühl y G. Muyzer.** 1999a. Spatial scale and diversity of benthic cyanobacteria and diatoms in a salina. *Hydrobiologia* 401: 199-206.
- Nübel U., F. García-Pichel, M. Kühl y G. Muyzer.** 1999b. Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Appl Environ Microbiol* 65: 422-430.
- Øvreås L., L. Forney, F.D. Daae y V. Torsvik.** 1997. Distribution of bacterioplankton in meromictic lake Sælevannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl Environ Microbiol* 63: 3367-2273.
- Pickup R.W.** 1991. Development of molecular methods for the detection of specific bacteria in the environment. *J Gen Microbiol* 137: 1009-1019.
- Porteus L.A., R.J. Seidler y L.S. Watrud.** 1997. An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. *Mol Ecol* 6: 787-791.
- Preston, F.W.** 1948. The commonness, and rarity, of species. *Ecology* 29: 254-283.
- Price, P.W.** 1975. *Insect Ecology*. Wiley, New York, N.Y.
- Purdy K.J., T.M. Embley, S. Takii y D.B. Nedwell.** 1996. Rapid extraction of DNA and rRNA from sediments by a novel hydroxiapatite spin-column method. *Appl Environ Microbiol* 62: 3905-3907.
- Raaijmakers, J.G.W.** 1987. Statistical analysis of the Michaelis-Menten equation. *Biometrics* 43: 793-803.
- Ricklefs R.E. y G.L. Miller.** 2000. *Ecology*. W.H. Freeman and Company, New York, N.Y.
- Rosado A.S., G.F. Duarte, L. Deldin y J.D. van Elsas.** 1997. Molecular microbial ecology: a mini review. *Revista de Microbiología* 28: 135-147.
- Samways, M.J.** 1994. *Insect conservation biology*. Chapman & Hall, London, England.
- Seber, G.A.F.** 1973. *The estimation of animal abundance and related parameters*. Griffin, London, England.
- Soberón J. y J. Llorente.** 1993. The use of species accumulation functions for the prediction of species richness. *Conserv Biol* 7: 480-488.
- Stackebrandt E., W. Liesak y D.M. Goebel.** 1993. Bacterial diversity in a soil sample of a subtropical Australian environment as determined by 16S rDNA analysis. *FASEB J* 7: 232-236.
- Stackebrandt E. y F.A. Rainey.** 1995. Partial and complete 16S rDNA sequences, their use in generation of 16S rDNA phylogenetic trees and their implications in molecular ecological studies, en A.D.L. Akkermans, J.D. van Elsas, F.J. de Bruijn (eds.) *Molecular Microbial Ecology Manual*. Kluwer, Dordrecht. págs. 3.1.1: 1-17.
- Staley, J.T.** 1997. Biodiversity: are microbial species threatened? *Curr Opin Biotechnol* 8: 340-345.
- Steffan R.J. y R.M. Atlas.** 1998. DNA amplification to enhance detection of genetically engineered bacteria in environmental samples. **Appl Environ Microbiol** 54: 2185-2191.
- Steffan R.J., J. Goksoyr, A.K. Bej y R.M. Atlas.** 1988. Recovery of DNA from soils and sediments. *Appl Environ Microbiol* 56: 776-781.
- Sugihara, G.** 1980. Minimal community structure: an explanation of species abundance patterns. *Am Nat* 116: 770-787.

- Suzuki M.T. y S.J. Giovannoni.** 1996. Bias caused by template annealing in the amplification mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* 62: 625-630.
- Tiedje, J.M., S. Asuming-Brempong, K. Nusslein, T.L. Marsh y S.J. Flynn.** 1999. Opening de black box of soil microbial diversity. *Appl Soil Ecol* 64:2894-2898.
- Torsvik V., F.L. Daae y J. Goksøyr.** 1995. Extraction, purification, and analysis of DNA from soil bacteria, en J.T. Trevors y J.D. van Elsas (eds.) *Nucleic Acids in the Environment: Methods and Applications*. Springer-Verlag, Berlin. págs 29-48.
- Torsvik V., R. Sørheim y J. Goksøyr.** 1996. Total bacterial diversity in soil and sediment communities – a review. *J Indust Microbiol* 17: 170-178.
- Torsvik V., L. Øvreås y T.F. Thingstad.** 2002. Prokaryotic Diversity – Magnitude, Dynamics and Controlling Factors. *Science* 296: 1064-1066.
- Tsai Y. y B.H. Olsen.** 1992. Detection of low numbers of bacterial cells in soils and sediments by polymerase chain rection. *Appl Environ Microbiol* 58: 754-757.
- Tunlid A., H.A.J. Hoitink, C. Low y D.C. White.** 1989. Characterization of bacteria that supress Rhizotocnia damping-off in bark compost media by analysis of fatty acid biomarkers. *Appl Environ Microbiol* 55: 1368-1374.
- Ueda T., Y. Suga y T. Matsuguchi.** 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. *Eur J Soil Sci* 46: 415-421.
- van Winzingerode F., U.B. Gobel y E. Stakebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* 21: 213-229.
- Wagner M., A. Loy, R. Nogueira, U. Purkhold, N. Lee y H. Daims.** 2002. Microbial community composition and function in waste water treatment plants. *Antoine van Leuwenhoek* 81: 65-680.
- Wagner R.** 1994. The regulation of ribosomal RNA synthesis and bacterial cell growth. *Arch. Microbiol.* 161: 100-109.
- Wallner G., B. Fuchs, S. Spring, W. Beisker y R. Amman.** 1997. Flow sorting of microorganisms for molecular analysis. *Appl Environ Microbiol* 63: 4223-4231.
- Ward D.M., M.M. Bateson, R. Weller y A.L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature, en K.C. Marshall (ed.) *Advances in Microbial Ecology*. Plenum Press, New York. págs. 219-286.
- Widmer F., A. Fließbach, E. Lackzó, J. Schulze-Aurich y J. Zeyer J.** 2001. Assesing soil biological characteristics: a comparison of bulk soil community DNA-, PFLA-, and BIOLOG<sup>TM</sup> – analyses. *Soil Biology & Biochemistry* 33: 1029-1036.
- Woese C.R.** 1987. Bacterial evolution. *Microbiol Rev* 51: 221-271.
- Woese C.R., O. Kandler y M.L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proc Natl Acad Sci USA* 87: 4576-4579.
- Wullings B.A., A.R. van Beuningen, J.D.Janse y A.D.L. Akkermans.**1998. Detection of *Ralstonia solanacearum* wich causes brown rot of potato, by fluorescent in situ hybridization with 23S rRNA-targeted probes. *Appl. Environ. Microbiol.* 64: 4546-4554.
- Zarda B., D. Hahn, A. Chatzinotas, W.Schonhuber, J.Neef, R.I. Amann y J. Zeyer.** 1997. Analysis of bacterial community structure in bulk soil by in situ hybridization. *Arch Microbiol* 168: 185-192.
- Zelles L. y Q.V. Bai.** 1993. Fractionation of fatty acids derived from soil lipids by solid phase extraction and their quantitative analysis by GC-MS. *Soil Biology & Biochemistry* 25: 495-507.

- Zheng D., E.W. Alm, D.A. Stahl y L. Raskin.** 1996. Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Appl Environ Microbiol* 62: 4504-4513.
- Zhou J., M.A. Bruns y J.M. Tiedje.** 1996. DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62: 312-322.
- Zhou J., B. Xia, D.S. Treves, L-Y. Wu, T.L. Marsh, R.V. O'Neill, A.V. Palumbo y J.M. Tiedje.** 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* 68(1): 326-334.