Aquatic bacterial assemblage structure in Pozas Azules, Cuatro Cienegas Basin, Mexico: Deterministic vs. stochastic processes

Laura Espinosa-Asuar,[1] Ana Elena Escalante,[2] Jaime Gasca-Pineda ,[1] Jazmín Blaz ,[1] Lorena Peña,[1] Luis E. Eguiarte,[1] Valeria Souza[1]*

[1]Department of Evolutionary Ecology, Autonomous University of Mexico, Mexico City, Mexico. [2] National Laboratory of Sustainability Sciences, Institute of Ecology, Autonomous University of Mexico, Mexico City, Mexico

Summary. The aim of this study was to determine the contributions of stochastic vs. deterministic processes in the distribution of microbial diversity in four ponds (Pozas Azules) within a temporally stable aquatic system in the Cuatro Cienegas Basin, State of Coahuila, Mexico. A sampling strategy for sites that were geographically delimited and had low environmental variation was applied to avoid obscuring distance effects. Aquatic bacterial diversity was characterized following a culture-independent approach (16S sequencing of clone libraries). The results showed a correlation between bacterial beta diversity (1-Sorensen) and geographic distance (distance decay of similarity), which indicated the influence of stochastic processes related to dispersion in the assembly of the ponds' bacterial communities. Our findings are the first to show the influence of dispersal limitation in the prokaryotic diversity distribution of Cuatro Cienegas Basin. [**Int Microbiol** 2015; 18(2):105-115]

Keywords: bacterial assemblage structure · bacterial diversity · biogeography · inland waters· Cuatro Cienegas, Mexico

Introduction

The causes of the observed distribution of microbial diversity have been the subject of investigation in microbial ecology for almost two decades. To date, two main groups of causative factors are under consideration: (i) those in which dispersal limitation, neutral assembly, and mass effects, also known as stochastic processes, are preponderant and (ii) those in which local adaptation to environmental conditions, also known as deterministic processes, plays the larger role.

Biogeography is the study of organisms in space and time. It provides an understanding of the underlying causes of their distribution and ultimately of the drivers that give rise to biodiversity [28]. Traditionally, biogeographical studies have been conducted on macroorganisms, but in the past two decades the focus has expanded to include microorganisms, which have been investigated through culture-independent approaches. These studies have advanced the current understanding of the mechanisms that generate and maintain microbial diversity [15,29]. However, they have also generated debate regarding the processes that lead to the assembly of microbial communities, as both stochastic and deterministic models have been proposed.

*****Corresponding author:** V. Souza
Departamento de Ecología Evolutiva
Universidad Nacional Autónoma de México
Mexico City, Mexico
Tel. + 52-5556229006. Fax +52-5556228995.
E-mail: souza@servidor.unam.mx

To investigate the biogeographical patterns in microorganisms, Martiny and colleagues [28] introduced a framework with which to provide evidence of the potential causes of microbial diversity. Four scenarios explaining microbial distribution were proposed: (i) a random scenario, as the null hypothesis, in which microorganisms are arbitrarily distributed over space; (ii) a deterministic hypothesis, in which the observed distribution reflects differences in the contemporary environment; (iii) the existence of dispersal limits for bacteria, in which the dispersal history is reflected in the beta diversity (dissimilarity in composition) coupled with geographic distances; and (iv) an association between stochastic and deterministic processes. To properly test these four hypotheses, sampling design is critical. Thus, samples must be obtained from several locations that are geographically delimited and from systems with low environmental variation, to avoid obscuring distance effects [1].

The Cuatro Ciénegas Basin (CCB) is part of the Chihuahuan Desert (State of Coahuila, Mexico) (Fig. 1) and has been a protected wetland since 1994 (APFF, according to Mexican Federal Law). Despite its extreme oligotrophy [42], the CCB harbors a high microbial diversity in different environments, including soil [26,27], microbial mats [2,35], and water [33,34,42]. Studies conducted in the CCB investigated the association between the distribution of microbial diversity and both spatial and environmental variables. Geographical associations were not detected at either the community [12,26,27] or the population [4,39,40] level. However, sampling in those studies was performed in contrasting environmental systems, which probably interfered with the ability to detect statistically relevant geographic associations [1].

Thus, in the present study we designed a sampling scheme in non-contrasting and equivalent environments that allowed us to distinguish the influence of stochastic vs. deterministic processes in the observed distribution of microbial diversity. Specifically, we used a culture-independent approach to study the microbial diversity in an aquatic system, the Pozas Azules This system consists of many small, isolated, circular ponds of different diameters (6–50 m) and with maximum depths of 10 m. Locally, they are referred to as "pozas" (= wells or pits) (Fig. 1B). Both the temperature and the chemical properties of these ponds were previously shown to be stable over time [20,35]. Thus, the spatial distribution of the Pozas Azules is ideal for biogeographic exploration, as the ponds are essentially aquatic islands in a sea of land [9]. Because many of the reported spatial scales that have shown a historically contingent assembly (geographical association) in bacterial communities have been small (~1 km) [15], we selected four nearby ponds (maximum distance of ~1.8 km) with equivalent temperatures and low variations in pH and conductivity.

In accordance with the four scenarios proposed by Martiny et al. [28], we present evidence supporting an association between biogeographic patterns and the geographic distance of bacterial assemblages in Pozas Azules. This implies that bacterial dispersal at the CCB is limited, even at a scale of < 2 km. Thus, our study supports the third hypothesis: that community-structure patterns reflect stochastic processes related to dispersion.

Material and methods

Study site and sampling. The four ponds of the Pozas Azules aquatic system were sampled in February, 2007. Pozas Azules is located in the southwestern flank of the San Marcos Sierra, within the CCB (Fig. 1A), and has been part of a private preserve ranch since 2000 (Pronatura A.C. and the Nature Conservancy). A recent general description of the hydrology of the CCB referred to the Pozas Azules as a karstic system, rich in $CaHCO_3$. Its waters are a mixture of mountain recharge and carbonate aquifer groundwater [45]. A preliminary exploration of the ponds by our group showed that conductivity ranged from 1900 to 4500 μS cm^{-2}.

Our study was carried out in four GPS-referenced neighboring ponds (A–C, F). These ponds were visually similar (clear blue water; Fig. 1B), not visibly connected (maximum separation of 1.8 km; Fig. 1C), and had a diameter of 10–30 m. The pH and conductivity of the selected ponds ranged from 7.5 to 7.8 and from 1900 to 2800 μS cm^{-2}, respectively. In a preliminary analysis using terminal restriction fragment length polymorphism (T-RFLP), different samples taken from the same pond at the same time were very similar (data not shown). The relative homogeneity of the ponds in this area is due to the presence of a springtime stream that, together with the wind, mixes the waters and thus reduces spatial structuring—as suggested by for other aquatic environments [24]; the process is aided by the small size of the ponds. Based on these preliminary data, we concluded that a composite sample of each pond would adequately represent the bacterial community. Five sites, four near the edge (equidistant) and one at the center of the pond, were sampled in each of the four ponds. All samples consisted of 2.5 l of water collected 50 cm below the surface. The five samples from each pond were transferred into a single clean and sterile container (composite sample) and preserved at 4°C until processed a few hours later. The composite samples were then filtered and divided into subsamples for DNA (three 4-l subsamples) and nutrient (three 100-ml subsamples) analyses. Water samples for the DNA analysis were filtered first through a 10-μm Durapore filter and then through a 0.22-μm Durapore filter (Millipore, USA) using a field vacuum pump. The filters were placed into sterile 2-ml Eppendorf tubes, which were preserved in liquid nitrogen and transferred to the laboratory, where they were stored at −80°C until DNA extraction. Water samples for the nutrient analysis were filtered through a 0.22-μm nitrocellulose filter; the filtrate was preserved at 4°C until processing.

Physicochemical and nutrient analyses. Total dissolved solids (TDS), pH, salinity, conductivity (Cond), temperature, and dissolved oxygen were measured in the field at each of the five sampling points using a Hydrolab YSI600QS (YSI, Yellow Springs, OH, USA). Total carbon (TC) and inorganic carbon (IC), total nitrogen (TN) and dissolved inorganic nitrogen (DIN or NH_4^+), nitrite (NO_2^-), and total phosphorus (TP) and inorganic phospho-
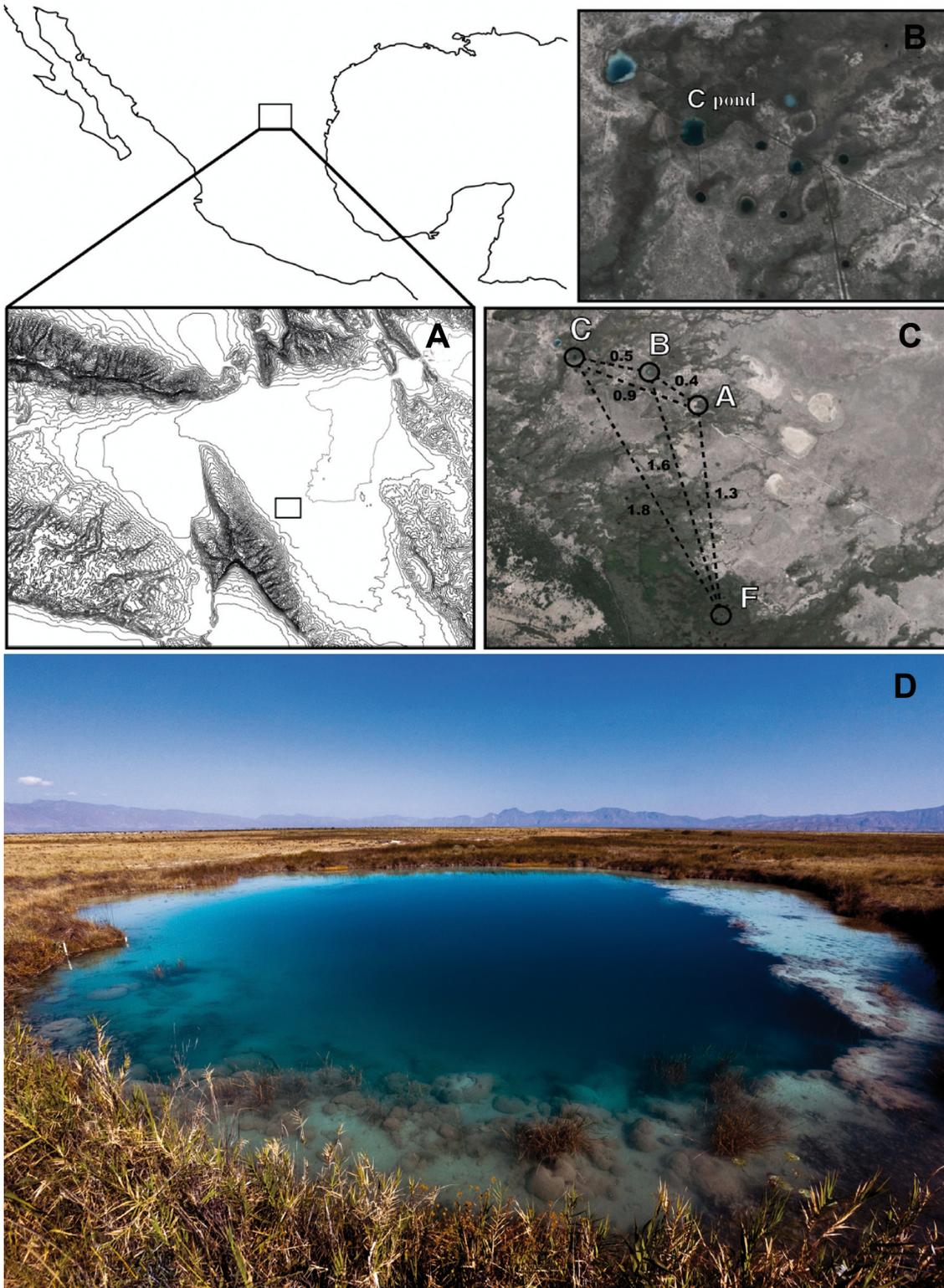
Fig. 1. The Pozas Azules system. (**A**) Map of México displaying the location of the Cuatro Ciénegas Basin (CCB) and the Pozas Azules system (squares) (modified from Rebollar et al. 2012). (**B**) Aerial view of pond C and the nearest ponds. The circular form of the pools can be visibly appreciated. Clear blue ponds associated with low conductivity values, such as pond C, were selected for this study (photograph used with permission of APFFCC). (**C**) Spatial distribution of the sampling sites at Pozas Azules (ponds A, B, C, and F); dashed lines show the distance between the sites in kilometers (Google Earth image). (**D**) Macroscopical aspect of one Poza azul.

rous (PO_3) were determined in the laboratory following standard procedures [35], using triplicate 100-ml subsamples from the composite samples. Dissolved organic carbon (DOC), phosphorus (DOP), and nitrogen (DON) concentrations were determined as the difference between the respective total values and the values of the inorganic components.

DNA isolation and sequencing of 16S rRNA clone libraries. Genomic DNA was extracted from the filtered samples (three replicates per pond) using an UltraClean water DNA kit (MoBio, Carlsbad, CA, USA) that included a final phenol-chloroform purification procedure. Each of the three DNA replicates per pond were independently subjected to PCR amplification targeting the 16S rRNA gene fragments of the bacterial domain using the universal primers 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′) [22]. The three PCR products of each pond were mixed prior to the cloning reaction. The cloning reaction, plasmid extraction, sequencing, and PCR conditions were performed as previously reported [33]. One clone library was generated per pond (n = 80±10). Partial sequences of the 16S rRNA gene were obtained using the primer 27F.

Sequence analysis. Sequences were trimmed and manually checked using BioEdit version 7.0.9.0 [14] . Sequences with an average length of 700 bp were aligned with the SILVA database and the NAST aligner [8] using Mothur v.1.31.2 [41] , they were then manually checked. Chimeric sequences were identified using Bellerophon software [17], UCHIME v4.2.40 [11], and DECIPHER [46] and were discarded from the dataset. Unique operational taxonomic units (OTUs) were defined with a threshold of 97% sequence similarity and were identified using Mothur v.1.31.2 [41].

Diversity and statistical analyses. Each OTU was phylogenetically identified using the classifier tool implemented on the Ribosomal Database Project (RDP) website [44]. Once unique sequences (the OTUs) were determined, their phylogenetic identities were confirmed using a neighbor-joining analysis (500 bootstrap replicates) in MEGA v.5 [43] and a maximum-likelihood analysis (1000 bootstrap replicates) using a TIM3+I+G model as determined by the JModel test [7] in PhyML v3.0 [13]. After the total number of unique OTUs was defined, a matrix that included all of the OTUs, their abundance, and the sampling sites was constructed and then used in the diversity, ordination, and correlation analyses described below.

Diversity analyses. To evaluate sampling efforts and to compare diversity among sampling sites, we determined coverage values, constructed rarefaction curves, and estimated alpha diversity per pond with the Chao 1, Shannon, and Simpson diversity indices for all samples. Coverage values were determined using the Good estimator, following the equation $C = (1 - n /N) \times 100$, where C is the percentage of coverage of the library, n is the number of singletons, and N is the total number of clones analyzed [37]. Beta diversity was calculated using the Sorensen binary index (using dissimilarity 1-Sorensen). Rarefaction curves and diversity indices were calculated using Mothur v.1.31.2 [41].

Ordination analyses. Multivariate and ordination approaches were used to visualize patterns in the distributions of environmental variables and to evaluate their correspondence with the OTU distributions [38]. A principal component analysis (PCA) was performed to visualize the environmental variance of the sampling sites. The rda function was carried out using the Vegan package [32] of R, version 3.0.1 (2013-05-16) [36]. In addition, a canonical correspondence analysis (CCA) was performed to model species' responses to environmental variables [13] using OTU abundance and standardized environmental data. The CCA function was implemented in the Ade4 package of R [10].

To standardize the environmental variables, raw data were transformed into z-scores using the decostand function (standardized method) implement-

ed in R (Vegan package) [32]. Non-redundant and explanatory environmental variables (except the geographic location) were identified using three statistical criteria for the standardized environmental variables: the PCA loadings (> 0.70), a paired correlation between variables (Pearson), and the colldiag index (>50; the table is available upon request). Ecological relevance was considered as well. The first component of the PCA accounted for 69.2% of the total variation and mainly involved DOC, Sal, IC, TDS, and Cond. Of these five variables, correlations between DOC, Sal, and IC were determined (data available upon request); therefore, these three variables were reduced and only IC was used as the representative explanatory variable. Likewise, TDS was discarded because it correlated with Cond, since both are derived measures. Consequently, from the first PCA component only IC and Cond were selected. The remaining four variables were retained either to explore nutrient variables (NO_2^-, TP) or because of their ecological relevance in determining aquatic bacterial communities (pH and TC) [25]. All of these calculations were performed using R packages as follows: The correlation between the variables was explored using the cor function of the Stats R package; graphics (order.single function) were created with Gclus [19] using the panelutils.r script. Condition indexes and variance decomposition proportions (colldiag function) were generated in Perturb v.2.05 [16].

Correlation analyses. To statistically determine the relative importance of geographic and contemporary environmental conditions, correlation analyses were conducted using OTU incidence data and environmental and geographical variables. These analyses were performed with functions implemented in the Vegan package [32] in R, unless otherwise specified. Three kinds of data matrices were constructed: a biological matrix composed of the 1-Sorensen index (estimated with the vegdist function), an environmental matrix using selected environmental variables (standardized z-scores) and constructed by calculating the Euclidian distance (dist function, Stats R package [36]), and a spatial matrix that contained distances between ponds as calculated from the coordinate data.

The effects of geographic distance vs. environmental dissimilarity on the composition of the assemblages (beta-diversity) were tested in pairwise Mantel tests among the three matrices using the Mantel function (Pearson's correlation and 9999 permutations). Environmental variables that best correlated with community dissimilarities (biological matrix) were identified using the BIOENV function [Pearson correlation and 1-Sorensen (Bray binary) arguments].

Nucleotide sequence accession numbers. The sequences were deposited in GenBank under the accession numbers KJ998817 to KJ999144.

Results

OTU composition: moderate turnover and phylum dominance.
The diversity estimated at the OTU and phylum levels was distributed differently among the sampling sites. At 97% sequence similarity, 48 OTUs, distributed across 10 phyla, were identified (Fig. 2A). Of these, 17 OTUs were shared (i.e., they occurred in more than one pond; Fig. 2B). Few OTUs were present in abundance (>15 sequences per OTU) in the four ponds but many OTUs were rare (one or two sequences each; Fig. 2B). OTUs that were abundant in one pond were not exclusive and were generally found in the other three ponds, but at lower numbers (Fig 2B). At the phylum
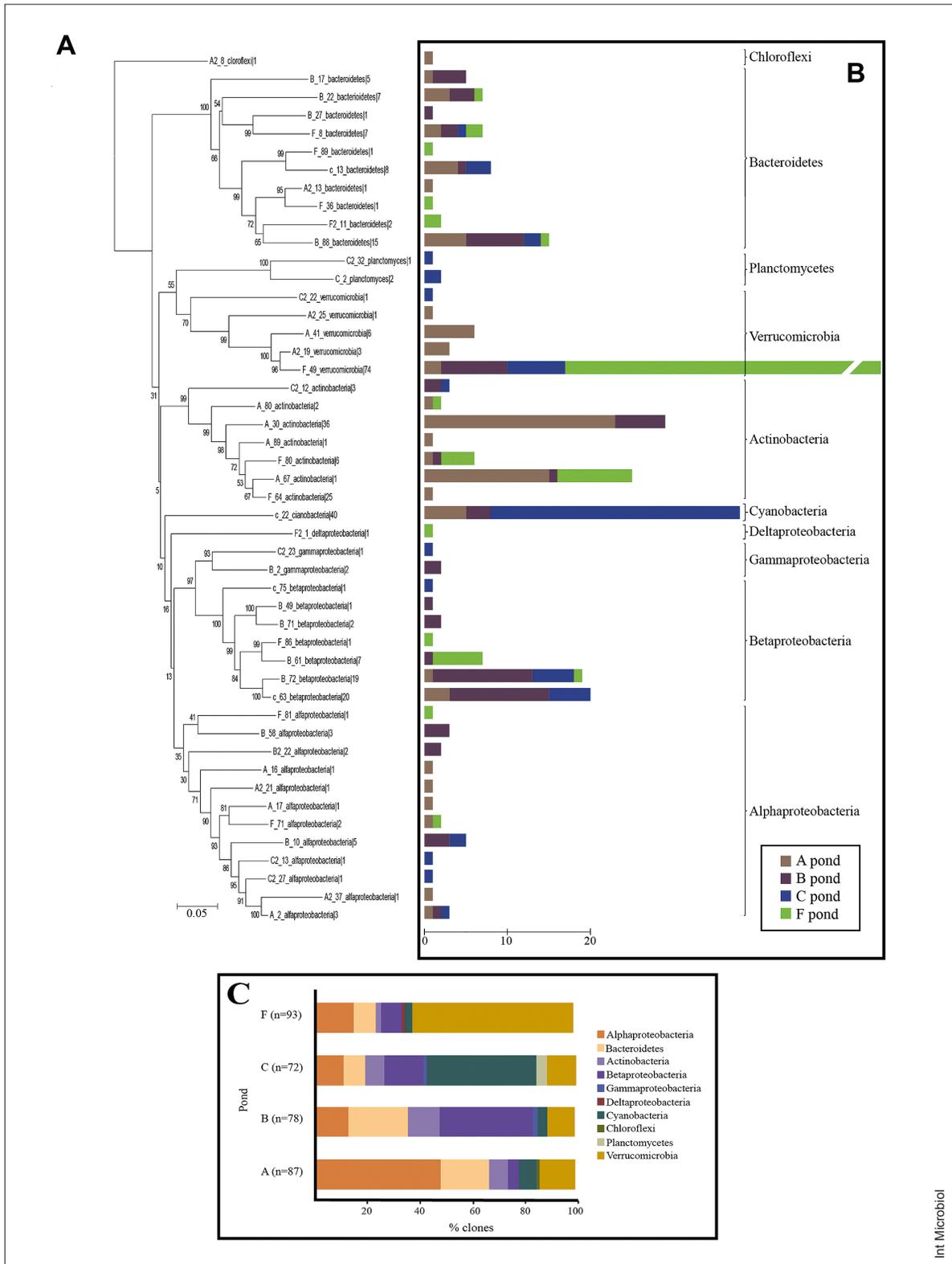
Fig. 2. Taxonomic description of the Pozas Azules bacterial communities. (**A**) Neighbor-joining tree of representative 16S rDNA gene sequences. The representative 48 OTUs of the sequences are defined at a 0.03 distance cutoff. The numbers after the OTU names correspond to sequence abundance (ML phylogeny gave very similar topologies; data not shown). (**B**) OTU abundance bars; each refers to a different pond. The F_49 OTU abundance bar is out of scale. (**C**) Relative abundance (in percent) of bacterial phyla from the clone library data of the four ponds. Each bar refers to a different pond. The number of clones obtained from each site (n) is indicated.
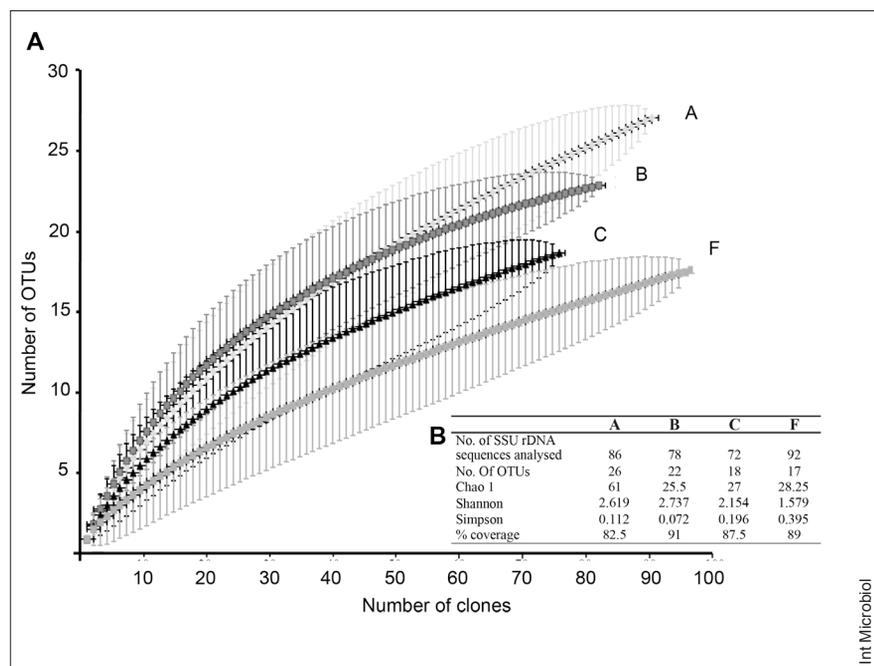
Fig. 3. Diversity-related data of bacterial communities in Pozas Azules CCB. (**A**) Rarefaction curves showing the 95% confidence interval of 16S rDNA gene sequences for each pond. The number of OTUs detected (at 97% cutoff) vs. the number of sequences analyzed per pond is shown. (**B**) The diversity data obtained with 16S rRNA sequences for each pond (OTUs at 97% cutoff).

level, differences between ponds were determined (Fig. 2C). Thus, the most abundant phyla were actinobacteria in pond A, betaproteobacteria in pond B, cyanobacteria in pond C, and verrucomicrobia in pond F. To describe the compositional differences between the ponds, a pairwise dissimilarity index (1-Sorensen, which ranges from 0, i.e., the identical composition of the samples, to 1, no similitude among the samples) based on OTU composition was calculated, followed by the average of all of the paired comparisons (0.548 ± 0.0973). The latter indicated moderate turnover between ponds.

Despite the fact that the rarefaction curves did not reach an asymptote (Fig. 3A), the coverage value for each gene library exceeded 80%. The diversity comparisons among sampling sites was compared based on different indices; rarefaction curves were then used to compare species richness in samples of different sizes. For the majority of indices (Chao 1, Shannon, and Simpson; Fig. 3B) and rarefaction curves (Fig. 3A), ponds A and B were the most diverse whereas pond F was the least diverse. Given the small number of sequences per sample, only the most abundant taxa per sample were recovered, which in the determination of biodiversity distribution patterns was previously shown to be robust [49,50] and thus valid despite the limited sampling depth.

Low environmental variation across aquatic ponds. Environmental variables among ponds did not show

major differences. For example, conductivity and pH ranged from 1940 to 2883.2 μS cm^{-2} and from 7.53 to 7.81, respectively (Table 1). Due to the extreme oligotrophy of these sites, the concentrations of some nutrients were below the limit of detection (Table 1). As shown in the PCA including all environmental variables, environmental variation was present across ponds (Fig. 4). The first component of the PCA explained 69.2% of the variation, which divided the ponds into two groups, pond F and ponds A–C, based on differences in Cond, Sal, TDS, IC, and DOC (Table 1).

Stochastic processes influencing the bacterial assemblages of Pozas Azules. The association between OTU composition and environmental variables was explored with a CCA using OTU abundance and environmental parameter data for each pond. The 15 original environmental variables (Table 1) were reduced to six non-redundant explanatory variables (except geographic location) based on the three statistical criteria described in the Methods section. The final environmental variables used in the analyses were TC, IC, TP, Cond, NO_2^-, and pH. Many OTUs were detected in only one of the four ponds. Although this may have been due to the low sequencing coverage, the abundances of these OTUs permitted their association with specific environmental variables (CCA data are available upon request).

Table 1. Environmental data obtained for each pond

	Pond			
	A	B	C	F
Latitude	N 26 49′24.4′′	N 26 49′31.8′′	N 26 49′34.7′′	N 26 48′37.08′′
Longitude	W 102 00′53.2′′	W 102 01′5′′	W 102 01′22.8′′	W 102 00′47.64′′
pH	7.76 (0)	7.53 (0)	7.712 (0.09)	7.81 (0.10)
Cond ($\mu S\ cm^{-2}$)	2501.6 (8.73)	2529 (3.67)	2883.2 (3.63)	1940 (0)
Sal (mg/l)	8.70 (0.53)	8.51 (1.24)	7.66 (0.82)	3.18 (0)
TDS (ppt)	1.3 (0)	1.3 (0)	1.5 (0)	1 (0)
TC (mg/l)	120.265(0.54)	109.22 (14.03)	110.41 (19.36)	102.28 (3.84)
IC (mg/l)	120.265 (0.54)	117.51 (7.13)	111.28 (18.12)	41.87 (1.38)
DOC (mg/l)	0	0	0	60.4 (3.54)
TP (mg/l)	0.83 (0.090)	0.89 (0.018)	0.87 (0.051)	0. 92 (0.025)
IP (PO_3) (mg/l)	0	0	0	0
DOP (mg/l)	0.83 (0.090)	0.89 (0.018)	0.87 (0.051)	0. 92 (0.025)
TN (mg/l)	0	0	0	0
DIN (NH_4^+) (mg/l)	0	0	0	0
DON (mg/l)	0	0	0	0
NO_2^- (mg/l)	0	0	0.0026 (0.0015)	0

Nutrient and other environmental data except for coordinates are the mean values. Standard deviations are in parentheses. (See Material and M methods for abbreviations.)

The relative importance of geographic (stochastic) and contemporary environmental (deterministic) conditions in the distribution of microbial diversity was determined in correlational analyses using environmental and geographical variables and by using the OTU incidence data. Environmental and geographic distance matrices were independently tested for their correlation with the biological distance matrix (differences among sites were represented by the dissimilarity index 1-Sorensen), using Mantel tests. There was no significant association between the biological distance matrix and the environmental matrix ($P = 0.1671$, $r = 0.5876$), whereas the biological distance matrix correlated significantly with geographic distance ($P = 0.04117$, $r = 0.7979$). To further evaluate the environmental variables that best correlated with community differences (i.e., the biological matrix), a BIO-ENV test was performed that identified TC, NO_2^-, pH, and Cond as the best-fitted variables; however, a Mantel correlation for these variables was not statistically significant ($P = 0.0862$, $r = 0.1635$).

Overall, our results provide evidence of a statistical correlation between bacterial composition and geographic distance.

Discussion

Diversity distribution studies have been used to investigate the influence of geographic and environmental factors in biogeographic patterns. For many years, the paucity of biogeographical studies of microbial diversity hindered an understanding of the ecological and evolutionary processes shaping diversity at the microbial scale. However, technological advances developed in the past two decades have allowed explorations of microbial diversity in natural environments and the testing of hypotheses regarding the factors determining its distribution.

To distinguish deterministic from stochastic dynamics, sampling design is critical. Specific environments should be chosen from locations that are geographically delimited and have low environmental variation, to avoid obscuring distance effects [1]. In this study, we adopted a culture-independent approach with a sampling strategy that minimized environmental variation, by looking at bacterial diversity in four ponds within the Pozas Azules aquatic system in the CCB, Mexico. Our findings show the influence of stochastic pro-

cesses in the distribution of prokaryotic diversity in the CCB.

OTU composition: moderate turnover and dominance by a particular phylum.

The phylogenetic analysis of microbial diversity showed one numerically dominant phylum in each pond that could nonetheless be found in all of the other ponds, albeit in lower numbers. These results provide further evidence of the complexity of bacterial community assembly mechanisms, in which "no two, naturally assembled bacterial communities appear to be the same" [6]. The average beta diversity between ponds indicated that nearly 50% of the OTUs were shared among them. The sharing of these taxa (many of them also abundant) is likely due to the fact that they are widely dispersed and likely to be generalist bacteria [23]. In fact, the same phyla have been reported in aquatic systems of the CCB valley [12,33,34] and in other freshwater systems [48]. Specifically, Actinobacteria are widely distributed in freshwater lakes, because of their dispersal capacity and ability to form spores [30]. Another potential generalist group shared among the studied ponds was Cyanobacteria, of which Chlorococcales [33] is the most abundant order in freshwater lakes [30]. Alphaproteobacteria (Rhodobacteriaceae) are common taxa of water bodies (e.g., OTU A_2) [5] and Betaproteobacteria predominated in the mesocosm experiments of Langenheder and Székely [23]. The most abundant verrucomicrobial OTU in our study (OTU F_49) was that of *Luteolibacter*, a widely distributed, non-motile bacterium [47].

In contrast to the dominant and shared phyla, other groups, such as Bacteroidetes, although shared, were not abundant, despite being registered as ubiquitous in aquatic environments [31]. We also found OTUs that were exclusive to one sampling site, even though most of them are highly abundant worldwide (e.g., the Hyphomicrobiaceae OTU A2_21) [18]. Although this discrepancy may be a consequence of the low sequencing coverage, whether shared or not these potentially generalist taxa occurred in different abundances across the four ponds. Langenheder and Székely [23] suggested that generalist groups usually represent neutrally assembled taxa.

Since it has been reported that patterns of abundance and diversity of bacterial communities are not substantially different whether or not rare taxa are included [49,50], the interpretation and discussion of the data and the observed patterns presented here are valid, even with the limited sampling.

Local factors in the Pozas Azules system: low environmental variation.

There were no major variations in the environmental conditions of the sampled ponds.

The similarity of the environmental variables at the study site was essential to statistically distinguish the influence of stochastic processes with respect to diversity distribution [1]. The 15 measured environmental variables were reduced to six non-redundant explanatory variables. Of these, the four variables selected by the BIOENV test (TC, Cond, NO_2^-, and pH), while not statistically correlated with community dissimilarities among ponds, were previously reported as important elements shaping the bacterial communities of lakes [25]. However, given our sampling strategy, in which the sites were selected based on the minimal environmental variation between them, a statistical correlation of these variables with assemblage dissimilarity was not expected. Despite the low environmental variability, one pond (F) deviated from the other three (Fig. 4). Thus, the use of finer measurements at smaller spatial scales may allow the detection of environmental variation, which can in turn explain otherwise confounding observations, such as differences in the abundance of some phyla between ponds.

Stochastic processes influence the Pozas Azules bacterial assemblages.

Our sampling strategy allowed us to independently evaluate the correlation between community dissimilarity among the ponds and geographic distance or environmental variables.

With respect to the environmental variables, the Mantel correlation test was not statistically significant, and in the CCA analysis most OTUs were not associated with any specific environmental condition. These non-conclusive correlations and associations were not surprising, given that the ponds were selected based on their minimal environmental variability. In terms of geographic effects on community composition, our results showed an association between aquatic bacterial composition and the geographic distances among the ponds.

Several studies have described the bacterial diversity in the CCB [2,4,12,26,27,33–35,39,40,42], but ours is the first to analyze environmentally similar sites to specifically determine the contribution of stochastic processes (dispersal limitation and neutral assembly) vs. deterministic processes (physicochemical conditions) to the observed distribution of microbial diversity. Two previous studies examined aspects of the microbial diversity from Pozas Azules [33,34], but their focus was the effect of environmental changes on the composition and diversity of bacterial communities, as determined in a long-term mesocosm experiment. The samples used in the studies of Pajares and colleagues were collected 2 years later than our samples [33,34]. One of the ponds ana-
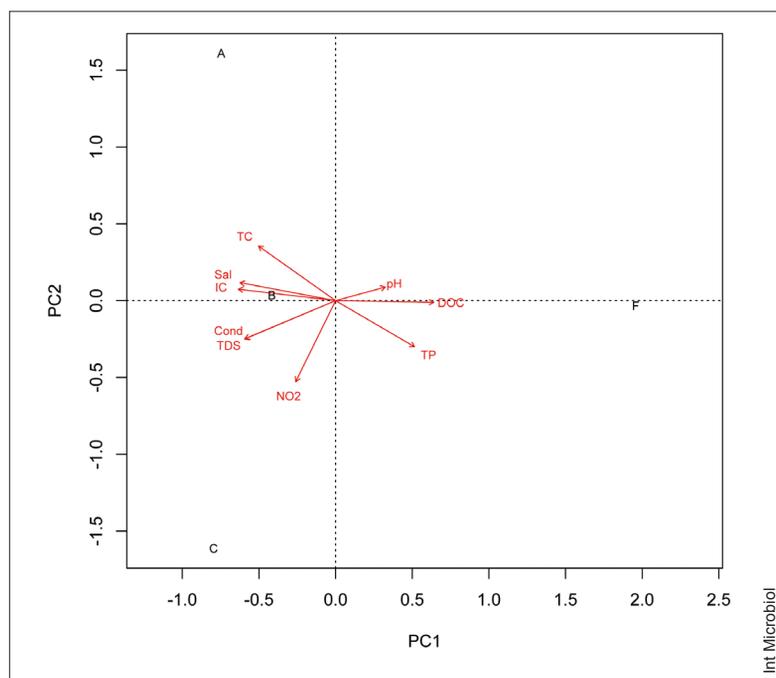
Fig. 4. Biplot generated from a principal component analysis (PCA) of the standardized environmental variables. The four ponds (A, B, C, and F) are shown. Vectors show scaled environmental variables. The first component of the PCA analysis accounted for 69.2% of the total variation, mainly involving DOC, Sal, IC, and Cond. (See Materials and methods for abbreviations.)

lyzed here was also included in their studies. It should be noted that ~50% of this study's OTUs were detected in the bacterial communities reported by those authors. This coincidence in OTU composition suggests the existence of a common pool of successful bacterial taxa whose presence is constant over time. If these taxa consist of generalists this would explain the neutrally assembled dynamics observed in this study [23]. Our results also suggest that the community dynamics in these ponds are associated with life-history traits—as reported by Nemergut et al. [29]—which are dependent on dispersion capacity, and with other bacterial ecological attributes (generalist or specialist). Thus, rare and slow-growing taxa would have no geographic association in aquatic communities [23], as was the case for the slow-growing phylum Planctomycetes in our samples (data not shown) [21].

The geographical pattern found in this study is consistent with the third hypothesis of Martiny et al., which states that there are limits to the dispersal of bacteria and that this dispersal history is reflected in the correlation between beta-diversity and geographic distances.

Conclusions and perspectives. Our results provide evidence of the primary importance of stochastic processes in determining the composition of aquatic bacterial communities in the Pozas Azules system. By using a sampling design that minimized the effect of environmental variability, we were able to show a correlation between beta diversity (1-Sorensen) and geographic distance. However, this evidence is suggestive, not conclusive, given the limited sample number. A larger number of samples and greater sequence coverage through a high-throughput sequencing approach should be part of future explorations of the CCB's aquatic bacterial communities. Many of the OTUs detected in Pozas Azules were similar to generalist taxa, which are usually neutrally assembled. More advanced investigations into the causes of the diversity distribution of prokaryotes should include controls for environmental variation (for example, using both contrasting and similar environmental sampling sites), to test the relative influence of stochastic and deterministic processes, and he study of specialist or rare taxa within the total communities (using specific primers or PhyloChip-designed approaches). Finally, the inclusion of life-history aspects, such as relative abundance, dispersion capacities, and ecological range (generalist *vs.* specialist), will lead to a better understanding of the processes underlying biogeographical patterns.

Competing interests. None declared.

References

1. Bissett A, Richardson AE, Baker G, Wakelin S, Thrall PH (2010) Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. Mol Ecol 19:4315-4327

2. Bonilla-Rosso G, Peimbert M, Alcaraz LD, Hernández I, Eguiarte LE, Olmedo-Alvarez G, Souza V (2012) Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin II: Community structure and composition in oligotrophic environments. Astrobiology 12:659-673

3. Ter Braak CF (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio 69:69-77

4. Cerritos R, Eguiarte LE, Avitia M, Siefert J, Travisano M, Rodríguez-Verdugo A, Souza V (2011) Diversity of culturable thermo-resistant aquatic bacteria along an environmental gradient in Cuatro Ciénegas, Coahuila, México. Ant van Leeuwen 99:303–318

5. De Corte D, Sintes E, Yokokawa T, Herndl GJ (2013) Comparison between MICRO–CARD–FISH and 16S rRNA gene clone libraries to assess the active versus total bacterial community in the coastal Arctic. Environ Microbiol Rep 5:272-281

6. Curtis TP, Sloan WT (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. Curr Opin Microbiol 7:221-226

7. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

8. DeSantis TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM, Phan R, Andersen GL (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. Nucleic Acids Res 34:W394-399

9. Dodson S (1992) Predicting crustacean zooplankton species richness. Limnol Oceanogr 37:848-856

10. Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22:1-20

11. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194-2200

12. Escalante AE, Eguiarte LE, Espinosa-Asuar L, Forney LJ, Noguez AM, Souza Saldivar V (2008) Diversity of aquatic prokaryotic communities in the Cuatro Cienegas basin. FEMS Microbiol Ecol 65:50-60

13. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321

14. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98

15. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 10:497-506

16. Hendrickx J (2012) Perturb: tools for evaluating collinearity. R package version 2.05. http://CRAN.R-project.org/package=perturb. (Accessed 2013 Nov.)

17. Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics 20:2317-2319

18. Huo YY, Cheng H, Han XF, Jiang XW, Sun C, Zhang XQ, Zhu XF, Liu YF, et al. (2012) Complete genome sequence of *Pelagibacterium halotolerans* B2 T. J Bacteriol 194:197-198

19. Hurley C (2012) Gclus: clustering graphics. R package version 1.3.1. http://CRAN.R-project.org/package=gclus. (Accessed 2013 Nov.)

20. Johannesson KH, Cortés A, Kilroy KC (2004) Reconnaissance isotopic and hydrochemical study of Cuatro Ciénegas groundwater, Coahuila, México. J South Am Earth Sci 17:171-180

21. Lage OM, Bondoso J (2012) Bringing Planctomycetes into pure culture. Front Microbiol 3:405

22. Lane DJ (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematics. Wiley, New York, NY, pp 115-175

23. Langenheder S, Székely AJ (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. ISME J 5:1086-1094

24. Lindström ES (2006) External control of bacterial community structure in lakes. Limnol Oceanogr 51:339-342

25. Logue JB, Lindström ES (2008) Biogeography of bacterioplankton in inland waters. Freshw Rev 1:99-114

26. López-Lozano NE, Eguiarte LE, Bonilla-Rosso G, García-Oliva F, Martínez-Piedragil C, Rooks C, Souza V (2012) Bacterial communities and the nitrogen cycle in the gypsum soils of Cuatro Ciénegas Basin, Coahuila: a Mars analogue. Astrobiology 12:699-709

27. López-Lozano NE, Heidelberg KB, Nelson WC, García-Oliva F, Eguiarte LE, Souza V (2013) Microbial secondary succession in soil microcosms of a desert oasis in the Cuatro Cienegas Basin, Mexico. Peer J 1:e47

28. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, et al. (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4:102-112

29. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL et al. (2013) Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev 77:342-356

30. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 75:14-49

31. O'Sullivan LA, Rinna J, Humphreys G, Weightman AJ, Fry JC (2006) Culturable phylogenetic diversity of the phylum "Bacteroidetes" from river epilithon and coastal water and description of novel members of the family Flavobacteriaceae: *Epilithonimonas tenax* gen. nov., sp. nov. and *Persicivirga xylanidelens* gen. nov., sp. Int J Syst Evol Microbiol 56:169-180

32. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. (2013) Vegan: Community Ecology Package. R package version 2.0-8. [http://CRAN.R-project.org/package=vegan]. (Accessed 2013 Nov.)

33. Pajares S, Bonilla-Rosso G, Travisano M, Eguiarte LE, Souza V (2012) Mesocosms of aquatic bacterial communities from the Cuatro Cienegas Basin (Mexico): a tool to test bacterial community response to environmental stress. Microb Ecol 64:346-358

34. Pajares S, Eguiarte LE, Bonilla-Rosso G, Souza V (2013) Drastic changes in aquatic bacterial populations from the Cuatro Cienegas Basin (Mexico) in response to long-term environmental stress. Ant. van Leeuwen 104:1159-1175

35. Peimbert M, Alcaraz LD, Bonilla-Rosso G, Olmedo-Alvarez G, García-Oliva F, Segovia L, Eguiarte LE, Souza V (2012) Comparative metagenomics of two microbial mats at Cuatro Ciénegas Basin I: ancient lessons on how to cope with an environment under severe nutrient stress. Astrobiology 12:648-658

36. R-Development Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

37. Ragon M, Fontaine MC, Moreira D, López-García P (2012) Different biogeographic patterns of prokaryotes and microbial eukaryotes in epilithic biofilms. Mol Ecol 21:3852-3868

38. Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62:142-160

39. Rebollar EA, Avitia M, Eguiarte LE, González-González A, Mora L, Bonilla-Rosso G, Souza V (2012) Water-sediment niche differentiation in ancient marine lineages of *Exiguobacterium* endemic to the Cuatro Cienegas Basin. Environ Microbiol 14:2323-2333

40. Rodríguez-Verdugo A, Souza V, Eguiarte LE, Escalante AE (2012) Diversity across seasons of culturable *Pseudomonas* from a desiccation lagoon in Cuatro Cienegas, Mexico. Int J Microbiol, Article ID 201389, 10 pp

41. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, et al. (2009) Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-7541

42. Souza V, Espinosa-Asuar L, Escalante AE, Eguiarte LE, Farmer J, Lloret L, Rodríguez-Martínez JM, Soberón X, et al. (2006) An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. Proc Natl Acad Sci USA 103:6565-6570

43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739

44. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-5267

45. Wolaver BD, Crossey LJ, Karlstrom KE, Banner JL, Cardenas MB, Ojeda CG, Sharp JM (2012) Identifying origins of and pathways for spring waters in a semiarid basin using He, Sr, and C isotopes: Cuatro cienegas Basin, Mexico. Geosphere 9:113-125

46. Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol 78:717-725

47. Yoon J, Matsuo Y, Kasai H (2012) *Portibacter lacus* gen. nov., sp . nov., a new member of the family Saprospiraceae isolated from a saline lake. J Gen Appl Microbiol 58:191-197

48. Zhao D, Huang R, Zeng J, Yan W, Wang J, Ma T, Wang M, Wu QL (2012) Diversity analysis of bacterial community compositions in sediments of urban lakes by terminal restriction fragment length polymorphism (T-RFLP). World J Microbiol Biotechnol 28:3159-3170

49. Zinger L, Boetius A, Ramette A (2014) Bacterial taxa-area and distance-decay relationships in marine environments. Mol Ecol 23:954-964

50. Zinger L, Gobet A, Pommier T (2012) Two decades of describing the unseen majority of aquatic microbial diversity. Mol Ecol 21:1878-1896