The ISME Journal (2010) 4, 829–838

www.nature.com/ismej

ORIGINAL ARTICLE

Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments

Emily B Hollister[1], Amanda S Engledow[1], Amy Jo M Hammett[2], Tony L Provin[1], Heather H Wilkinson[2] and Terry J Gentry[1]
[1]*Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA and* [2]*Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, USA*

Studies of hypersaline ecosystems often yield novel organisms and contribute to our understanding of extreme environments. Soils and sediments from La Sal del Rey, a previously uncharacterized, hypersaline lake located in southern Texas, USA, were surveyed to characterize the structure and diversity of their microbial communities. Samples were collected along a transect that spanned vegetated uplands, exposed lakebed sediments, and water-logged locations, capturing a wide range of environments and physical and chemical gradients. Community quantitative PCR (qPCR) was used in combination with tag-encoded pyrosequencing, 16S rRNA gene cloning, and Sanger sequencing to characterize the lake's soil and sediment microbial communities. Further, we used multivariate statistics to identify the relationships shared between sequence diversity and heterogeneity in the soil environment. The overall microbial communities were surprisingly diverse, harboring a wide variety of taxa, and sharing significant correlations with site water content, phosphorus and total organic carbon concentrations, and pH. Some individual populations, especially of *Archaea*, also correlated with sodium concentration and electrical conductivity salinity. Across the transect, *Bacteria* were numerically dominant relative to *Archaea*, and among them, three phyla—the *Proteobacteria*, *Bacteroidetes*, and *Firmicutes*—accounted for the majority of taxa detected. Although these taxa were detected with similar abundances to those described in other hypersaline ecosystems, the greater depth of sequencing achieved here resulted in the detection of taxa not described previously in hypersaline sediments. The results of this study provide new information regarding a previously uncharacterized ecosystem and show the value of high-throughput sequencing in the study of complex ecosystems.
The ISME Journal (2010) **4,** 829–838; doi:10.1038/ismej.2010.3; published online 4 February 2010
Subject Category: microbial ecology and functional diversity of natural habitats
Keywords: hypersaline; La Sal del Rey; pyrosequencing; sediment microbial communities

Introduction

Hypersaline ecosystems are distributed globally and represent a wide range of ecosystem types, including salt lakes, soda lakes, hypersaline springs, salt flats, playas, solar salterns, and ancient salt deposits (Terry *et al.*, 2000; Oren, 2002; Ventosa *et al.*, 2008). Hypersaline water, soils, and sediments harbor active and diverse microbial communities (Oren, 2002; Humayoun *et al.*, 2003; Caton *et al.*, 2004; Ley *et al.*, 2006; Mesbah *et al.*, 2007) and contain organisms representing all three domains of life (Oren, 2002). Studies of hypersaline ecosystems have improved our understanding of the biology of extreme environments, and many have resulted in the discovery of novel organisms and enzymes with enhanced potential for biotechnological applications (Caton *et al.*, 2004; Nicholson and Fathepure, 2005; Ventosa *et al.*, 2008). The majority of studies published on the microbiology of hypersaline ecosystems focus on aquatic communities (Oren, 2002; Humayoun *et al.*, 2003; Demergasso *et al.*, 2004; Mutlu *et al.*, 2008) and/or microbial mats (Mouné *et al.*, 2003; Ley *et al.*, 2006), whereas far fewer have attempted to characterize hypersaline soils and sediments (David *et al.*, 2005; Dong *et al.*, 2006; Jiang *et al.*, 2006; Mesbah *et al.*, 2007; Ventosa *et al.*, 2008).

The advent of pyrosequencing technology has provided the ability to examine microbial communities with unprecedented levels of coverage and detail (Sogin *et al.*, 2006; Roesch *et al.*, 2007; Sundquist *et al.*, 2007; Acosta-Martínez *et al.*, 2008). Pyrosequencing has enabled the production of sequence libraries that are orders of magnitude larger than conventional Sanger sequence libraries,

Correspondence: EB Hollister, Department of Soil and Crop Sciences, Texas A&M University, 370 Olsen Blvd., College Station, TAMU 2474, TX 77843-2474, USA.
E-mail: ehollister@tamu.edu
Received 25 September 2009; revised 6 January 2010; accepted 8 January 2010; published online 4 February 2010

and at a fraction of the cost (Cardenas and Tiedje, 2008). Such advances provide a major step forward in our ability to characterize microbial communities, particularly those that are highly complex, have been under-sampled historically, and/or may harbor members of the 'rare biosphere' (Sogin *et al.*, 2006). Furthermore, the high-throughput, low-cost nature of pyrosequencing affords new opportunities to study the dynamics of microbial communities in time and space or under fluctuating environmental conditions.

A recent global survey of bacterial communities from natural environments found that sediment communities may be more phylogenetically diverse than those from any other environment, including soils (Lozupone and Knight, 2007), and that salinity is a major factor in determining microbial community composition (Lozupone and Knight, 2007). Given these factors and the tremendous resolution afforded by new sequencing technologies, a tag-encoded pyrosequencing study was undertaken to characterize the structure and diversity of soil and sediment microbial communities at La Sal del Ray, a naturally occurring and previously uncharacterized salt lake ecosystem in southern Texas, USA. A transect approach was used in combination with community quantitative PCR (qPCR), 16S rRNA tag-pyrosequencing, 16S rRNA gene cloning and sequencing, and multivariate statistics to characterize the lake's soil and sediment microbial communities and identify the relationships they share with their environment.

Materials and methods

Site description and sampling

La Sal del Rey is a shallow (~1 m depth), 215-hectare hypersaline lake located within the Lower Rio Grande Valley National Wildlife Refuge, Hidalgo County, TX (26°31′55″ N, 98°03′50″ W). Mean annual temperature for the region is 23.8 °C, and annual rainfall averages 583 mm (based on National Climatic Data Center information for McAllen, TX) (NOAA, 2009). The geological history of La Sal del Rey is uncertain; however, the lake, which sits in a depression on the landscape and has no major outlets, is thought to have formed either through result wind erosion or the presence of a salt dome (Barton, 1925). Historical accounts describe its use, from pre-settlement times through the American Civil War, as a salt mine (Hawkins, 1947), and more recent descriptions of the lake report its total salinity to be as high as 177.9 g l^{-1}, with NaCl as its dominant salt (Cole and Brown, 1967).

Soil and sediment samples were collected in June 2008 from La Sal del Rey's shoreline and lakebed. Samples were collected at eight points (T3-0, T3-65, T3-130, T3-195 T3-260, T3-325, T3-390, and T3-455), spaced ~20 m from one another, along a 140 m transect. The transect originated in a vegetated, upland area (that is site T3-0), extended across exposed lakebed sediments, and continued toward the water's edge (that is sites T3-390 and T3-455). From point T3-0 to point T3-455, the sediments became increasingly water-logged, grew darker in color, and smelled increasingly of volatile sulfur compounds. Upland vegetation included mesquite (*Prosopis glandulosa*), cactus (*Opuntia lindheimeri*), and scattered grasses, including *Setaria leucopila*, *Hilaria berlangeri*, and *Digitaria californica*. Sparse patches of shoregrass (*Monanthochloe littoralis*) were also present. The lake appeared to have contracted in area over the course of the spring, exposing portions of the lakebed and producing an ice-like layer of salt over some points along the transect.

Surface soils and sediments were collected to a depth of ~10 cm using a cylindrical bulb planter with a diameter of 6.5 cm. Temperatures were collected at mid-depth to characterize the 'average' conditions at each sampling point, recognizing that the variation in temperature between the surface and depth has the potential to be large. In cases where solid salt crusts were present, they were removed before sampling. Samples were placed into sterile plastic bags, sealed, and held on ice until return to the laboratory. One portion of each sample was dried at 45 °C and used for physical and chemical analysis, and a second portion was frozen at −80 °C and used for DNA extraction.

Physical and chemical analysis

Samples were dried to a constant weight in a forced air oven at 45 °C. They were then sieved to 2 mm and submitted to the Soil, Water, and Forage Testing Laboratory at Texas A&M University for pH (Schofield and Taylor, 1955), detailed salinity (Rhoades and Clark, 1978), electrical conductivity (Rhoades, 1982), and phosphorus, magnesium, calcium, and sodium analyses (Mehlich, 1978), as well as organic carbon and total nitrogen analysis (McGeehan and Naylor, 1988).

DNA extraction and purification

Community DNA was extracted from transect sediments using a PowerMax DNA isolation kit (Mo Bio Laboratories Inc., Carlsbad, CA, USA) and a modified version of the manufacturer's protocol. Fifteen gram aliquots of sediment and 15 ml of bead solution were added to each bead-beating tube. After 5 min of bead beating, lysozyme was added (final concentration of 1 mg ml^{-1}), and samples were incubated at 37 °C for 1 h. Following lysozyme treatment, solution 'C1' was added and samples were incubated at 65 °C for 30 min. The manufacturer's protocol was followed from this point onward. Following elution, DNA samples concentrated by ethanol precipitation, resuspended in 10 mM Tris (pH 8.0), and purified using a 1.2%

(w/v) low-melting point agarose gel. DNA bands were excised from the gel, extracted using a Wizard SV Gel and PCR Clean Up kit (Promega, Madison, WI, USA), and quantified with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Community quantitative PCR

Community qPCR assays, based on Fierer *et al.* (2005) and Lima and Sleep (2007), were used to evaluate the relative abundances of *Bacteria* and *Archaea* along the survey transect. Assays were performed in triplicate, using an Eppendorf Mastercycler ep realplex thermal cycler (Eppendorf, Hamburg, Germany). Each 10 μl reaction contained: 4.5 μl 2.5 × RealMasterMix with 20 × SYBR solution (5Prime, Inc., Gaithersburg, MD, USA), 1.0 μl BSA (10 mg ml^{-1}), 0.5 μl of each primer (10 μM,), 2.5 μl molecular grade water, and 1.0 μl template DNA (2.5 ng μl^{-1}). Thermocycling consisted of an initial denaturation at 95 °C for 15 min, followed by 40 cycles of 95 °C for 1 min and annealing temperature (see below) for 30 s, and 72 °C for 1 min. Primer set Eub338/518 was used with an annealing temperature of 53 °C for *Bacteria* (Fierer *et al.*, 2005) and primer set ARCH85F/313R was used with an annealing temperature of 61 °C for *Archaea* (Lima and Sleep, 2007).

Plasmid standards for *Bacteria* and *Archaea* were generated from the genomic DNA of *Escherichia coli* DH10B(pUC19) (obtained from Carlos Gonzales, Texas A&M University) and *Methanosarcina acetivorans* C2A (obtained from William Metcalf, University of Illinois). Regions of interest were amplified from each organism using a FailSafe PCR kit (Epicentre Biotechnologies, Madison, WI, USA) and their corresponding qPCR primer sets. The PCR products were confirmed for size on a 1.2% (w/v) low-melting point agarose gel with UV light, excised, and extracted using a Wizard SV Gel and PCR Clean Up kit (Promega). Then, they were cloned into a pGEM-T Easy vector (Promega), and plasmids were isolated using a Wizard SV Miniprep kit (Promega). Plasmid DNA concentrations were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and serial dilutions of 5.0×10^{-3} to 5.0×10^{-7} ng μl^{-1} DNA were amplified to generate standard curves.

Amplification efficiency was calculated as described by Pfaffl (2001). rRNA gene copy numbers were estimated from the plasmid-based standard curves, and were calculated as a function of DNA concentration, the average molecular weight of a base pair of DNA (660 g mol^{-1}), the size of our cloning vector, and the size of our DNA fragment of interest. Standard curves generated from our bacterial and archaeal controls displayed linear relationships between the log-values of their gene copy numbers and calculated threshold cycles ($r^2 \geqslant 0.95$). Amplification efficiencies were 1.85 and 2.49 for *Archaea* and *Bacteria*, respectively, and were within the range of those reported by others (Fierer *et al.*, 2005 and references therein). Gene copy numbers were adjusted to accommodate the potential for *Bacteria* and *Archaea* to carry multiple copies of the 16S rRNA gene. Average values of 4.14 copies bacterium^{-1} and 1.70 copies archaeon^{-1} were obtained from the Ribosomal RNA Database (Lee *et al.*, 2009) (accessed 18 August 2009) and used in these calculations.

Bacterial 16S rRNA amplification, cloning, and sequencing

The 16S rRNA gene was PCR-amplified from community DNA using primers 27F and 1492R (Lane, 1991). Each 25 μl reaction contained 100 ng template DNA, FailSafe buffer E (Epicentre Biotechnologies), 1.25 U *Taq* polymerase, and forward and reverse primers at a final concentration of 0.1 μM each. Thermocycling was conducted in a 2720 Thermal Cycler (Applied Biosystems, Foster City, CA, USA) under the following conditions: initial denaturation at 95 °C for 1 min; 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min, 30 s; and a final extension at 72 °C for 7 min. Five replicate amplifications were performed for each sample, and their products were combined for downstream use.

PCR products were confirmed for as described above. Product bands were excised, purified using a Wizard SV Gel and PCR Clean Up kit (Promega), and cloned into a pGEM-T Easy vector (Promega). Approximately 300 positive colonies were selected from each sample, and glycerol stocks of each were submitted to Agencourt Bioscience Corporation (Beverly, MA, USA) for single-pass Sanger sequencing with primer 515F (Lane, 1991).

Bacterial tag-encoded amplicon pyrosequencing

Purified community DNA samples were submitted to the Research and Testing Laboratory (Lubbock, TX, USA) for tag-pyrosequencing. Samples were amplified with modified versions of primers 530F and primer 1100R, and the amplicons were sequenced using Roche 454 Titanium chemistry, producing reads from the forward direction from 530F (Acosta-Martínez *et al.*, 2008). Sequences were quality trimmed according to Acosta-Martínez *et al.* (2008).

Sequence analysis and community comparisons

Sanger sequence data were quality checked, aligned, trimmed to a similar length, and assigned putative identities using the Ribosomal Database Project (RDP) pipeline (Cole *et al.*, 2007) (accessed 26 January 2009). The RDP pyrosequencing pipeline and Classifier function were used to align and assign identities to the 454 sequence data (accessed 26 January 2009). Sequences from the Sanger clone

libraries were screened for chimeras using Bellerophon (Huber *et al.*, 2004), and potential chimeras were excluded from further analysis. Those 454 sequences that did not align with the 16S rRNA region were considered to be sequencing errors and were excluded from further analysis.

Distance matrices were constructed for both the Sanger and 454 sequence libraries using the dist.seqs function in MOTHUR, version 1.4.1 (Schloss *et al.*, 2009). MOTHUR was also used to assign sequences to operational taxonomic units (OTUs, 97% similarity) and calculate both Shannon's diversity index values (H') and ChaoI richness estimates. Phylogenetic trees were constructed using the relaxed neighbor-joining algorithm in Clearcut (version 1.0.9) (Sheneman *et al.*, 2006), and between-site comparisons of phylogenetic structure were conducted using the parsimony test in Treeclimber (Schloss and Handelsman, 2006). Parsimony test scores with *P*-values <0.05 were considered to represent significant differences, and pairwise comparisons of individual libraries were conducted only if the study-wide (that is global) test was found to be statistically significant (Schloss and Handelsman, 2006). Bonferonni correction for multiple comparisons was used among the pairwise comparisons, adjusting our significance level to $P \leqslant 0.0017$ (Neter, 1996).

Nonmetric multidimensional scaling (NMDS) of sites based on OTU composition, and Mantel tests of community correlation with environmental variables, were conducted using PC ORD, version 5.0 (MjM Software, Gleneden Beach, OR, USA) in default mode. NMDS and the Mantel tests were carried out on both the Sanger and 454 sequence libraries. OTUs occurring as singletons were removed from the data set before analysis to down-weight the effects of rare 'species.' Sørenson distance was used for both the NDMS ordination and the 'species' distance metric in our Mantel tests (McCune and Grace, 2002); Euclidian distance was used to describe the underlying structure of our environmental variables (Fierer and Jackson, 2006).

Sanger sequence reads were submitted to NCBI GenBank under accession numbers GQ893028 through GQ894720, and the 454 sequence reads were submitted to the NCBI Short Read Archive under accession number SRA009427.2.

Results

Soils and sediments chemical characteristics

The chemical composition of the transect sediments is summarized in Table 1. The sites exhibited a narrow range of pH values, and all contained sodium as their dominant cation. The sites varied widely with respect to water content and nutrient and cation concentrations. Strong positive correlations were observed with respect to: (1) water content, sodium or potassium concentration, and electrical conductivity ($r^2 = 0.71$–0.77); (2) sodium, magnesium, and potassium concentration ($r^2 = 0.79$–0.93); (3) sodium concentration and electrical conductivity ($r^2 = 0.98$); and (4) phosphorus concentration with both organic carbon content ($r^2 = 0.77$) and pH ($r^2 = 0.86$). Sediment organic carbon and total nitrogen concentrations, however, were not well correlated with one another ($r^2 = 0.41$).

Community qPCR

On the basis of the relative proportion of gene copies detected in the community qPCR assays, *Bacteria* were numerically dominant among the La Sal del Rey sediments, accounting for $\geqslant 97\%$ of the rRNA gene copies detected at each site. Adjusting these values to reflect the average number of rRNA copies that may be found in each organism type, *Bacteria* accounted for 92–99% of community membership. Although *Archaea* represented only a small proportion of the microbial community within each site, their relative abundance tended to increase along the transect. They were nearly absent at the terrestrial end of the transect (sites T3-0 and T3-65) but increased to $\sim 8\%$ of the community at the aquatic end of the transect. Archaeal abundance correlated well with sample water content ($r^2 = 0.80$) and to a lesser degree with electrical conductivity ($r^2 = 0.62$) and sodium content ($r^2 = 0.53$).

Community composition, diversity, and estimated richness

A total of 39 590 16S rRNA sequences were generated through 454 sequencing, with an average read length of 455 ± 85 bp (mean $\pm$ s.d.). The 454 sequence libraries ranged in size from 1403 sequences at site T3-0 to 6745 sequences at site T3-325 and contained between 744 OTUs at site T3-0 and 3 478 OTUs at site T3-65 (Table 2). An additional 1693 16S rRNA sequences, averaging 425 ± 5 bp in length (mean $\pm$ s.d.) after trimming, were generated through cloning and single-pass Sanger sequencing. The Sanger clone libraries ranged in size from 185 sequences at sites T3-0 and T3-130 to 230 sequences at site T3-390 and contained between 68 OTUs at site T3-390 and 143 OTUs at site T3-195 (Table 2). Overall, a total of 16 596 unique OTUs were identified among the 454 sequence libraries, and 658 unique OTUs were identified among the Sanger sequence libraries. No single OTU accounted for $>0.32\%$ of the entire data set, but to convey a better sense of the composition of each community, the 10 most abundant OTUs from each sampling point, along with their putative identities, have been provided in Supplementary Table 1.

The 454 libraries detected a greater variety of *Archaea* and low-abundance bacterial taxa than did the Sanger sequence libraries (Table 3); however, the most abundant taxa were detected with similar frequencies across both sequencing projects

Table 1 Soil and sediment physical and chemical properties

	Transect location							
	T3-0	*T3-65*	*T3-130*	*T3-195*	*T3-260*	*T3-325*	*T3-390*	*T3-455*
pH	8.2	9.1	8.8	9.1	9.1	9.1	8.7	8.8
Na (mg kg^{-1})	9966	14 717	21 017	12 522	8 363	12 060	24 687	25 365
P (mg kg^{-1})	113	15	25	29	18	17	40	45
K (mg kg^{-1})	470	240	564	403	345	446	716	1132
Ca (mg kg^{-1})	3968	1763	3905	4348	1967	6503	2723	3895
Mg (mg kg^{-1})	1276	1112	2030	1159	1121	1127	2981	1127
S (mg kg^{-1})	1441	2147	4188	1859	1292	1640	2211	2926
Organic C (mg kg^{-1})	14 371	1329	2662	5038	2584	7736	4581	6384
Total N (mg kg^{-1})	2216	1158	1639	944	1331	1523	1308	882
Water content (g kg^{-1})	8.29	9.18	16.35	16.07	16.57	16.31	20.31	23.45
EC (mS cm^{-1})	34.2	40.3	74.2	62.8	38.6	61.8	94.1	123
Temperature (°C)[a]	37.5	44	43.3	43	44	44.7	43	41.7

Abbreviation: EC, electrical conductivity.
[a]Soil temperature at the time of collection.

Table 2 Summary of sequence library sizes, operational taxonomic units (OTUs), and diversity and richness estimates

Transect location	*454 sequence libraries*				*Sanger sequence libraries*			
	Sequence library size	*Number of OTUs*[a]	*Shannon (H')*	*Chao I richness*	*Sequence library size*	*Number of OTUs*[a]	*Shannon (H')*	*Chao I richness*
All sites	39 590	16 596	—	—	1693	658	—	—
T3-0	1403	744	6.34	1403	185	128	4.64	314
T3-65	5947	3478	7.83	8159	211	136	4.72	338
T3-130	5448	2631	7.41	5183	185	99	4.32	198
T3-195	5002	2627	7.42	6123	229	143	4.68	452
T3-260	5195	2389	7.32	4869	219	97	4.23	168
T3-325	6745	2592	7.30	4556	218	73	3.74	141
T3-390	5008	2290	7.28	4408	230	68	3.50	116
T3-455	4842	2523	7.45	5285	216	85	3.89	177

[a]The number of OTUs identified at 'all sites' is not equivalent to the sum of OTUs across the study transect, as some OTUs were found at multiple locations.

Table 3 Detection and distribution of low abundance phyla

	Transect points							
	T3-0	*T3-65*	*T3-130*	*T3-195*	*T3-260*	*T3-325*	*T3-390*	*T3-455*
(A) Phyla detected exclusively with 454								
Chrysiogenetes								x
Deinococcus-Thermus		x	x	x		x	x	x
Gemmatimonadetes	x	x						
OP10		x		x				
OP11	x			x				
SR1						x	x	
Thermotogae				x		x	x	
TM7	x	x	x	x			x	
(B) Detection improved by 454[a]								
BRC1	x	x		x				x
Chlamydiae	x	x	x	x	x			x
Deferribacteres			x			x	x	x
OD1	x	x		x		x	x	

[a]Denotes taxa that were detected in multiple 454 libraries but only one Sanger sequence library.

(Figure 1). Members of at least 24 bacterial and 2 archaeal phyla were detected along the study transect. *Proteobacteria* and *Bacteroidetes* were encountered most frequently, representing 24–45% and 4–26% of each library's membership, respectively (Figure 1). Unclassified *Bacteria* (that is those

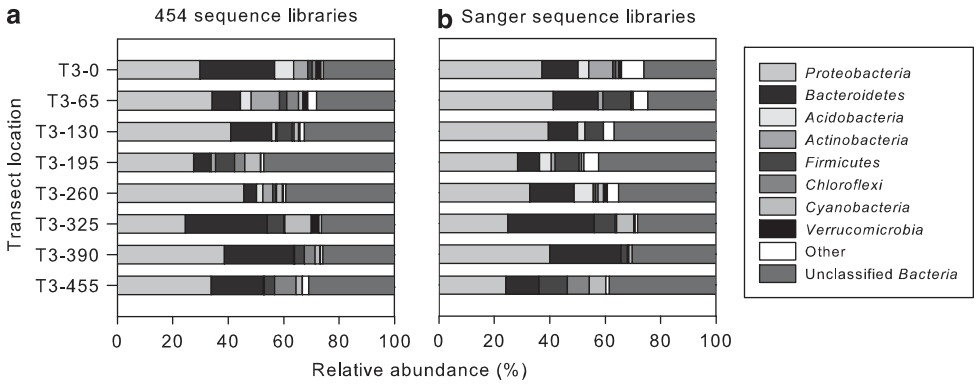

Figure 1 Bacterial community composition of soils and sediments located along a transect through a hypersaline lake ecosystem that included terrestrial sites (T3-0), exposed lakebed sediments (T3-65 to T3-260) and water-logged, 'aquatic' sites (T3-325 to T3-455). The 454 sequencing (panel **a**) and cloning and Sanger sequencing (panel **b**) were used to characterize the sediment communities. Both libraries were sequenced in the forward direction from positions 515–530 of the 16S rRNA gene.

that could not be identified at the phylum level with at least 80% confidence in RDP) accounted for an additional 25–47% of each sequence library. Among the *Archaea*, members of the phylum *Euryarchaeota* (particularly members of the order *Holobacteriales*) were encountered with the greatest frequency and accounted for the majority of archaeal sequences identified here.

Shannon diversity index values (H') suggest that diversity varied along the study transect. With the exception of the relatively small T3-0 454 sequence library, H' values tended to decrease as conditions became increasingly water-logged and salt-rich (Table 2), indicating that community richness and/or evenness declined along the transect. Chao I richness estimates suggest that approximately half of the estimated diversity contained within these communities was captured by our sequencing efforts (Table 2). They also suggest that further sequencing would be likely to yield additional unique OTUs; however, greater yields would be expected from the near-shore and intermediate-distance communities (for example sites T3-65, T3-130, and T3-195) than from those located farther away from shore (for example sites T3-260 to T3-455).

Analysis of communities via the Parsimony test

Parsimony tests conducted using the 1693 Sanger sequences found no significant differences among the transect communities with respect to their phylogenetic structure ($P = 0.132$). In contrast, when the parsimony test was conducted using the 39 950 sequences generated through 454 sequencing, all of the transect communities were significantly different from one another. The global test was statistically significant ($P < 0.001$), as were all of the pairwise comparisons of the individual transect locations ($P \leqslant 0.001$). Additional comparisons of sequence library similarity (that is overlap as measured by the Yue-Clayton index) can be found in Supplementary Table 2.

Microbial community correlations with physicochemical characteristics and NMDS ordination of communities based on OTUs

Mantel tests conducted with both the Sanger and 454 sequences found that La Sal del Rey's microbial communities correlated significantly with sample water, phosphorus, and total organic carbon content, as well as sample pH (Table 4). A significant correlation between sample calcium concentration and the Sanger communities was also detected.

The relationships shared between the transect microbial communities and the environmental variables that correlated well with them can be seen in the clustering of the transect sites in NMDS plots shown in Figure 2. Site water content, the best correlated variable measured here, can be seen in the segregation of the driest, most 'terrestrial' site T3-0, by itself, as well as the clustering of the most water-logged, 'aquatic' sites (T3-325, T3-390, and T3-455) with one another. The remaining four transect points formed a less cohesive 'intermediate' cluster. The influence of the other well-correlated variables can also be observed. For example, organic carbon concentration, phosphorus content, and pH appear to contribute to the segregation of site T3-0 from the other transect communities. Organic carbon concentration also appears to underlie the clustering of the transect into its three community types (Table 1), contributing to site T3-325's tendency to cluster with the 'aquatic' sites rather than the 'intermediate' sites despite its water content being more 'intermediate' in nature.

The distribution of OTUs across La Sal del Rey fell into one of three patterns: (1) occurring at most or all of the transect sites, but varying in abundance relative to site conditions; (2) occurring with cluster-specific distributions (that is terrestrial, intermediate, or aquatic); or (3) occurring with site-specific distributions (that is occurring exclusively in one of the eight transect sites). OTUs representing *Acidobacteria*, *Actinobacteria*, *Rhodobacteraceae*, and *Desulfohalobiaceae* were found at most or all

Table 4 Mantel correlations highlight the relationships shared between environmental traits and community composition

Environmental variable	*454 sequence libraries*		*Sanger sequence libraries*	
	Mantel score (r)	P-*value*	*Mantel score (r)*	P-*value*
pH	**0.441**	**0.023**	**0.489**	**0.035**
Na (mg kg^{-1})	0.231	0.142	0.101	0.265
P (mg kg^{-1})	**0.523**	**0.007**	**0.535**	**0.048**
K (mg kg^{-1})	0.108	0.203	0.044	0.330
Ca (mg kg^{-1})	−0.256	0.087	**−0.324**	**0.036**
Mg (mg kg^{-1})	0.220	0.139	0.101	0.288
S (mg kg^{-1})	−0.136	0.258	−0.120	0.379
Organic C (mg kg^{-1})	**0.475**	**0.014**	**0.541**	**0.034**
Total N (mg kg^{-1})	0.151	0.224	0.145	0.232
C:N ratio	−0.106	0.581	−0.088	0.362
Water (%)	**0.548**	**0.004**	**0.652**	**0.001**
Electrical conductivity (mS cm^{-1})	0.291	0.105	0.195	0.200

Bold-faced entries indicate significance values in which $P \leqslant 0.05$.

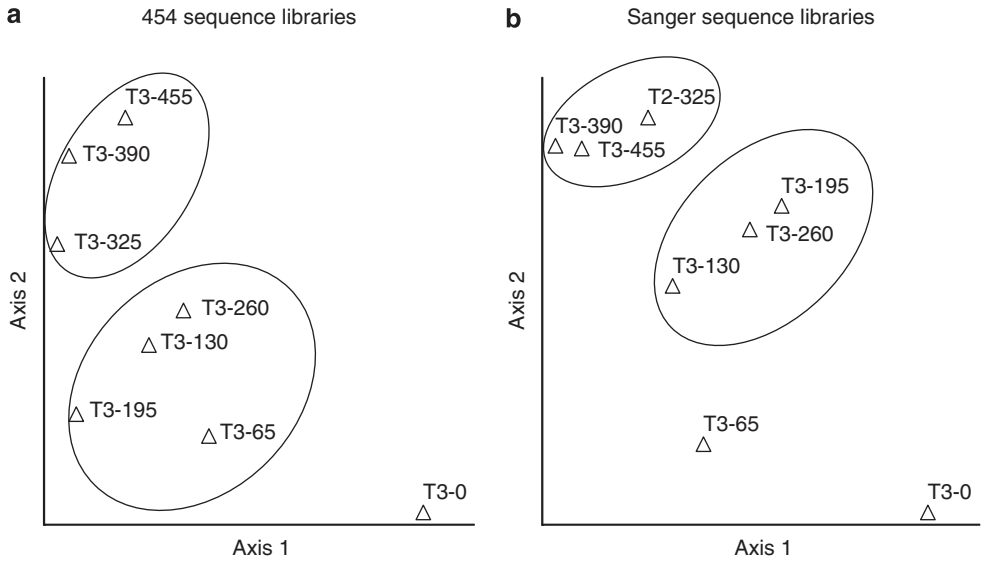

Figure 2 NMDS of transect points based on OTU composition determined by (**a**) 454 sequencing and (**b**) cloning and Sanger sequencing. Clustering of the transect communities into three groups, roughly corresponding to sample water content can be observed. Samples T3-390, T3-455, and T3-325 represent the most aquatic sites along the study transect, and site T3-0 represents the driest, most terrestrial site. Those clustered in the middle of the figure represent a set of conditions that were intermediate relative to the aquatic and terrestrial clusters. OTUs were defined as sequences sharing $\geqslant$97% similarity with one another.

of the sites along the study transect but varied in the abundance relative to site water content. *Acidobacteria* and *Actinobacteria* declined in abundance as conditions became more water-logged and salt-rich (Figure 1), whereas the abundances of the *Rhodobacteraceae* and *Desulfohalobiaceae* increased under the same conditions. The terrestrial-specific cluster of OTUs was composed of *Bacteroidetes* (particularly members of the genus *Balneola*), members of the GP10 division of the *Acidobacteria*, and *Myxococcales*-like organisms (Supplementary Table 1). Those OTUs occurring exclusively within the intermediate cluster (T3-65, T3-130, T3-195, and T3-260) included *Cyanobacteria*, other *Bacteroidetes*, *Chromatiales*, and a number of unclassified *Gammaproteobacteria* (Supplementary Table 1). The aquatic cluster (sites T3-325, T3-390, and T3-455) was characterized by OTUs representing *Rubrimonas*- and *Roseicyclus*-like bacteria, *Desulfohalobiaceae*, several unclassified *Bacteroidetes*-like organisms, and a variety of unclassified *Bacteria* (Supplementary Table 1). OTUs having site-specific distributions represented nearly 80% of the OTUs identified here, and the number (that is richness) of site-specific OTUs occurring within a given location was found to be negatively correlated with site water content ($r^2 = 0.69$, data not shown).

Discussion

In any discussion of PCR-based community characterizations, it should be acknowledged that these approaches are not without potential biases, including differential DNA extraction efficiencies,

the potential for species to carry multiple copies of the 16S rRNA gene, and biases in primer selectivity (von Wintzingerode *et al.*, 1997; Lee *et al.*, 2009). Furthermore, it is recognized that tag-pyrosequencing is also prone to error, including difficulty interpreting homopolymeric regions of sequence and producing amplicons sequences of anomalous length, both of which have the potential to overinflate diversity estimates (Kunin *et al.*, 2010). However, sequence library descriptions of community composition are often corroborated by other approaches to community characterization (Papineau *et al.*, 2005 and references therein), and recent evidence suggests that thorough controls on data quality and clustering can help to limit the effects of pyrosequencing errors on diversity estimates (Kunin *et al.*, 2010).

Although no technique is perfectly bias-free, it is clear that advances in sequencing technology are expanding our view of the microbial world in unprecedented ways. New methods are increasing dramatically the number of sequences that may feasibly be included in a single study, allowing deeper coverage and providing opportunities for new insights regarding microbial communities and their interactions with the environment (Sogin *et al.*, 2006; Turnbaugh *et al.*, 2006; Roesch *et al.*, 2007; Acosta-Martínez *et al.*, 2008). Although culture-independent studies of microbial communities have been carried out in hypersaline soils and sediments in the past, this study represents the first to have done so at La Sal del Rey, and it is the first among such studies to have used community qPCR and the combination of Sanger and tag-encoded pyrosequencing technologies.

The combination of techniques used in this study provided a detailed and powerful snapshot of the microbial communities occurring at La Sal del Rey. We found that the system's soil and sediment microbial communities harbored a wide variety of taxa and contained a large proportion of individuals that had no close match within the existing molecular databases. Despite our 454 sequence libraries being an order of magnitude larger than their corresponding Sanger libraries, and having been constructed using different primer sets, similar descriptions of community composition and similar correlations with environmental variables (Table 4) were obtained with each of the sequencing approaches. Major taxa were detected with similar relative abundances (Figure 1), corroborating the findings of other studies that have performed side-by-side comparisons of tag-encoded pyrosequencing versus Sanger clone libraries (Turnbaugh *et al.*, 2006; Jones *et al.*, 2009). Likewise, major taxa were detected with abundance levels comparable to those that have been described in other hypersaline ecosystems. The ratio of *Bacteria* to *Archaea* at La Sal del Rey resembles those reported by others who have attempted to quantify their abundance in other hypersaline sediments (Mouné *et al.*, 2003; Ley *et al.*, 2006). *Bacteria* appear to be numerically dominant in these systems, and among them three phyla—the *Proteobacteria*, *Bacteroidetes*, and *Firmicutes*—typically account for the majority of taxa found within each community (Mouné *et al.*, 2003; Dong *et al.*, 2006; Mesbah *et al.*, 2007).

Although the agreement among our 454 sequence libraries, Sanger sequence libraries, and those described by others in the literature was quite good relative to the abundances of the major taxa, we found that our 454 libraries included taxa that had not been captured by our Sanger sequencing efforts or those described in other studies (Humayoun *et al.*, 2003; Demergasso *et al.*, 2004; Dong *et al.*, 2006; Jiang *et al.*, 2006; Ley *et al.*, 2006; Mesbah *et al.*, 2007). These additional taxa generally occurred with low abundance and included members of the *Deinococcus-Thermus*, *Thermotogae*, and candidate division TM7, as well as a number of OTUs that could be identified, at best, as unclassified *Bacteria*. Akin to the detection of the 'rare biosphere' described by Sogin *et al.* (2006), the enhanced sampling depth achieved with tag-encoded pyrosequencing improved our detection of low abundance taxa (Table 3) and allowed us to profile our transect microbial communities with greater detail.

In addition to offering a broader view of the composition of La Sal del Rey's microbial communities, the increased sampling resolution achieved with the 454 libraries also improved our statistical power and enhanced our ability to detect differences among the transect communities. As communities increase in their complexity, it is thought that the minimum number of sequences (that is sampling effort) needed to distinguish differences among them should increase as well (Singleton *et al.*, 2001; Schloss and Handelsman, 2006). Parsimony tests conducted with our modestly sized Sanger libraries (average size = 211 sequences) did not find significant differences among the transect communities; however, when the parsimony test was repeated using the larger 454 libraries (average size = 4948 sequences), highly significant differences were detected among all transect sites.

Differences in microbial community diversity and phylogeny may result from a number of different factors, key among these being those characteristics, which define a community's physical and chemical environment. In sediment ecosystems, physical stratification and chemical gradients help to create and maintain high levels of diversity within and between bacterial communities (Torsvik *et al.*, 2002; Lozupone and Knight, 2007). Sodium concentration and electrical conductivity displayed some of the strongest gradients along our study transect, and given the importance of salinity in shaping microbial community structure at global scales, we expected that these factors would have an influential function at La Sal del Rey, as well. Although small portions of the communities at La Sal del Rey

(for example members of the archaeal class *Holobacteriales*) did display positive relationships with factors related to salinity, we found that overall community structure was better correlated with other environmental factors. These included sample water content, organic carbon concentrations, phosphorus content, and pH.

Perhaps the most striking among these factors, with respect to its influence on community structure, was sample water content. Despite the strong, positive correlations shared between sediment water content and factors related to salinity (for example sodium concentration and electrical conductivity), it appears that water's relationship with microbial community structure was primarily a function of its influence on sediment oxygen concentrations. Although we did not measure sediment oxygen concentrations directly, water content has been used as an indicator of sediment oxygen content in other systems (Dong *et al.*, 2006; Jiang *et al.*, 2006; Lozupone and Knight, 2007), and field observations of sediment color and odor suggest this to be the case at La Sal del Rey as well. Taken in whole, the Mantel correlations described above suggest that the microbial communities occurring along our study transect are shaped by their requirements for oxygen, the carbon substrates they metabolize, and their pH tolerances, rather than their tolerance or affinity for salt. Although salinity may have a major function in determining community composition at global scales (Lozupone and Knight, 2007), the results of this study suggest that its influence is less important at local scales, particularly in systems that are already salt-rich.

One of the central goals of microbial ecology is to understand the relationships that microbial communities share with their environment, and an extension of this goal is to understand how those relationships play out at local, regional, and global scales. Although this study provides a detailed snapshot and expands our understanding of microbial community variation at the local scale, it is clear that we have only just scratched the surface. We still have much to learn about these communities and the relationships that they share with one another, as well as the factors that shape these communities and others at larger scales. As sequencing technologies continue to advance, it is anticipated that progress toward these goals will improve steadily and provide greater insight into the diversity, complexity, and dynamics of microbial systems.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work was supported by the Texas AgriLife Research Bioenergy Program and the Texas A&M University Energy Resources Program. We acknowledge Mukul Sherekar, Justin Ng, and Heidi Mjelde for their assistance in the laboratory, the Lower Rio Grande Valley National Wildlife Refuge for access to and assistance in sampling La Sal del Rey, the Supercomputing Center at Texas A&M University for access to their computing facilities, and three anonymous reviewers, whose suggestions helped to improve this manuscript.

References

Acosta-Martínez V, Dowd S, Sun Y, Allen V. (2008). Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biol Biochem* **40**: 2762–2770.

Barton D. (1925). The salt domes of south Texas. *Am Assoc Pet Geol Bull* **9**: 536–589.

Cardenas E, Tiedje JM. (2008). New tools for discovering and characterizing microbial diversity. *Curr Opin Biotechnol* **19**: 544–549.

Caton TM, Witte LR, Ngyuen HD, Buchheim JA, Buchheim MA, Schneegurt MA. (2004). Halotolerant aerobic heterotrophic bacteria from the great salt plains of Oklahoma. *Microb Ecol* **48**: 449–462.

Cole GA, Brown RJ. (1967). Chemistry of *Artemia* habitats. *Ecology* **48**: 858–861.

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM *et al.* (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35**: D169–D172.

David AW, Papke RT, Doolittle WF. (2005). Archaeal diversity along a soil salinity gradient prone to disturbance. *Environ Microbiol* **7**: 1655–1666.

Demergasso C, Casamayor EO, Chong G, Galleguillos P, Escudero L, Pedrós-Alió C. (2004). Distribution of prokaryotic genetic diversity in athalassohaline lakes of the Atacama Desert, Northern Chile. *FEMS Microbiol Ecol* **48**: 57–69.

Dong HL, Zhang GX, Jiang HC, Yu BS, Chapman LR, Lucas CR *et al.* (2006). Microbial diversity in sediments of saline Qinghai lake, China: linking geochemical controls to microbial ecology. *Microb Ecol* **51**: 65–82.

Fierer N, Jackson JA, Vilgalys R, Jackson RB. (2005). Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. *Appl Environ Microbiol* **71**: 4117–4120.

Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.

Hawkins W. (1947). *El Sal del Rey*. Texas State Historical Association: Austin, TX.

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

Humayoun SB, Bano N, Hollibaugh JT. (2003). Depth distribution of microbial diversity in Mono lake, a meromictic soda lake in California. *Appl Environ Microbiol* **69**: 1030–1042.

Jiang H, Dong H, Zhang G, Yu B, Chapman LR, Fields MW. (2006). Microbial diversity in water and sediment of lake Chaka, an athalassohaline lake in Northwestern China. *Appl Environ Microbiol* **72**: 3832–3845.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. (2009). A comprehensive survey of soil

acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3**: 442–453.
Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118–123.
Lane DJ. (1991). 16S/23S rRNA sequencing. In: Stackenbrandt E, Goodfellow M (eds) *Nucelic Acid Techniques in Bacterial Systematics*. John Wiley & Sons: Chichester, UK. pp 115–157.
Lee ZM-P, Bussema III C, Schmidt TM. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and Archaea. *Nucleic Acids Res* **37**: D489–D493.
Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM *et al.* (2006). Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72**: 3685–3695.
Lima GdP, Sleep BE. (2007). The spatial distribution of eubacteria and archaea in sand-clay columns degrading carbon tetrachloride and methanol. *J Contam Hydrol* **94**: 34–48.
Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* **104**: 11436–11440.
McCune B, Grace JB. (2002). *Analysis of Ecological Communities*. MjM Software Design: Gleneden Beach, OR, 300 pp.
McGeehan SL, Naylor DV. (1988). Automated instrumental analysis of carbon and nitrogen in plant and soil samples. *Commun Soil Sci Plant Anal* **19**: 493–505.
Mehlich A. (1978). New extractant for soil test evaluation of phosphorus, potassium, magnesium, calcium, sodium, manganese and zinc. *Commun Soil Sci Plant Anal* **9**: 477–492.
Mesbah NM, Abou-El-Ela SH, Wiegel J. (2007). Novel and unexpected prokaryotic diversity in water and sediments of the alkaline, hypersaline lakes of the Wadi An Natrun, Egypt. *Microb Ecol* **54**: 598–617.
Mouné S, Caumette P, Matheron R, Willison JC. (2003). Molecular sequence analysis of prokaryotic diversity in the anoxic sediments underlying cyanobacterial mats of two hypersaline ponds in Mediterranean salterns. *FEMS Microbiol Ecol* **44**: 117–130.
Mutlu MB, Martínez-García M, Santos F, Peña A, Guven K, Antón J. (2008). Prokaryotic diversity in Tuz lake, a hypersaline environment in inland Turkey. *FEMS Microbiol Ecol* **65**: 474–483.
Neter J. (1996). *Applied Linear Regression Models*, 3rd edn. Irwin: Chicago, IL, xv, 720 pp.
Nicholson CA, Fathepure BZ. (2005). Aerobic biodegradation of benzene and toluene under hypersaline conditions at the Great Salt Plains, Oklahoma. *FEMS Microbiol Lett* **245**: 257–262.
NOAA (2009). *National Climatic Data Center* (http://cdo.ncdc.noaa.gov/ancsum/ACS), accessed 26 March 2009.
Oren A. (2002). *Halophilic Microorganisms and Their Environments*. Kluwer Academic: Boston, 575 pp.
Papineau D, Walker JJ, Mojzsis SJ, Pace NR. (2005). Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. *Appl Environ Microbiol* **71**: 4822–4832.
Pfaffl MW. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: 2002–2007.
Rhoades JD. (1982). Soluble salts. In: Page AL, Miller RH, Keeney DR (eds) *Methods of Soil Analysis: Part 2*, 2nd edn. ASA-SSSA: Madison, WI. pp 167–178.
Rhoades JD, Clark M. (1978). *Sampling Procedures and Chemical Methods in Use at the US Salinity Laboratory for Characterizing Salt-Affected Soils and Waters*. United States Department of Agriculture: Riverside, CA, pp 11–12.
Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1**: 283–290.
Schloss PD, Handelsman J. (2006). Introducing treeclimber, a test to compare microbial community structures. *Appl Environ Microbiol* **72**: 2379–2384.
Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
Schofield RK, Taylor AW. (1955). The measurement of soil pH. *Soil Sci Soc Am J* **19**: 164–167.
Sheneman L, Evans J, Foster JA. (2006). Clearcut: a fast implementation of relaxed neighbor joining. *Bioinformatics* **22**: 2823–2824.
Singleton DR, Furlong MA, Rathbun SL, Whitman WB. (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* **67**: 4374–4376.
Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci* **103**: 12115–12120.
Sundquist A, Bigdeli S, Jalili R, Druzin M, Waller S, Pullen K *et al.* (2007). Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* **7**: 108.
Terry JM, Renia TG, William DG, Helga S-L. (2000). *Environ Microbiol* pp 243–250.
Torsvik V, vreås L, Thingstad TF. (2002). Prokaryotic diversity—magnitude, dynamics, and controlling factors. *Science* **296**: 1064–1066.
Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1131.
Ventosa A, Mellado E, Sanchez-Porrro C, Marquez MC. (2008). Halophilic and halotolerant micro-organisms from soils. In: Dion P, Nautiyal CS (eds) *Microbiology of Extreme Soils*. Springer-Verlag: Berlin. pp 87–115.
von Wintzingerode F, Göbel UB, Stackebrandt E. (1997). Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)