Microb Ecol (2013) 66:850–860
DOI 10.1007/s00248-013-0276-2

ENVIRONMENTAL MICROBIOLOGY

Bacterial Diversity of Surface Sand Samples from the Gobi and Taklamaken Deserts

Shu An · Cécile Couteau · Fan Luo · Julie Neveu · Michael S. DuBow

Received: 25 September 2012 / Accepted: 1 August 2013 / Published online: 22 August 2013

Abstract Arid regions represent nearly 30 % of the Earth's terrestrial surface, but their microbial biodiversity is not yet well characterized. The surface sands of deserts, a subset of arid regions, are generally subjected to large temperature fluctuations plus high UV light exposure and are low in organic matter. We examined surface sand samples from the Taklamaken (China, three samples) and Gobi (Mongolia, two samples) deserts, using pyrosequencing of PCR-amplified 16S V1/V2 rDNA sequences from total extracted DNA in order to gain an assessment of the bacterial population diversity. In total, 4,088 OTUs (using ≥97 % sequence similarity levels), with Chao1 estimates varying from 1,172 to 2,425 OTUs per sample, were discernable. These could be grouped into 102 families belonging to 15 phyla, with OTUs belonging to the *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria* phyla being the most abundant. The bacterial population composition was statistically different among the samples, though members from 30 genera were found to be common among the five samples. An increase in phylotype numbers with increasing C/N ratio was noted, suggesting a possible role in the bacterial richness of these desert sand environments. Our results imply an unexpectedly large bacterial diversity residing in the harsh environment of these two Asian deserts, worthy of further investigation.

Electronic supplementary material The online version of this article (doi:10.1007/s00248-013-0276-2) contains supplementary material, which is available to authorized users.

S. An · C. Couteau · F. Luo · J. Neveu · M. S. DuBow (✉)
Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay, France
e-mail: michael.dubow@u-psud.fr

Introduction

Arid regions, characterized by a dearth of available water, represent nearly 30 % of the Earth's terrestrial surface. Desert regions are defined as arid regions that receive <250 mm average annual precipitation or where evapotranspiration exceeds precipitation. Deserts cover approximately 19 % of the Earth's land surface and their microbial ecology has not been extensively studied [1]. The living conditions at the surface of deserts are a challenge for microorganisms, as there is little available water and/or carbon, a very large range of temperatures, and high exposure to UV irradiation from the sun. The study of these desert microorganisms may aid in efforts to prevent the spread of deserts and/or to restore soil/vegetation cover and also offer the opportunity to discover novel biological activities such as new thermostable or alkaline-stable enzymes [2, 3].

Microbiological examinations of arid regions have begun to use culture-independent methods to identify and characterize the endogenous bacterial community [4–6]. Drees et al. studied the bacterial community from soil samples of the Atacama Desert using DGGE profiles. They demonstrated that microbial communities from the extreme hyperarid core of the Atacama Desert clustered separately from other bacterial communities by comparing their sequence data with that in Genbank. Pointing et al. analyzed, by terminal restriction fragment length polymorphism, the bacterial community in the polar McMurdo Dry Valley desert in Antarctica. They suggested that the biodiversity near the cold-arid limit for life is complex, but that the communities lack a large degree of variability, likely due to the high selective pressures of this harsh environment. Neilson et al. combined 454 pyrotag analysis with a full-length 16S rRNA gene library to reveal bacterial communities and to infer their functional metabolic potential inherent in unvegetated arid soils of the Atacama Desert. Their results revealed a wealth of novel bacterial groups with phylogenetic associations to non-phototrophic

primary producers and bacteria capable of biogeochemical cycling.

The two largest deserts in Asia, the Taklamakan and Gobi deserts, are both located within and adjacent to China and have not been extensively examined for the biodiversity of their bacterial communities. Since desertification and regular dust storms in Asia have become large environmental problems, an understanding of the bacterial community diversity of these deserts may be desirable to aid in bolstering desert soil adhesion. Prior studies on these deserts often focused on the discovery of new cultivable bacterial groups displaying interesting properties such as alkali and/or drought/and or thermotolerance and radioresistance [7–10]. However, a more in-depth description of the bacterial community composition and structure of these deserts has not yet been reported. We chose to examine surface sand samples from the tops of dunes of the Taklamaken Desert ($n=2$) and Gobi Desert ($n=2$), plus one surface sand sample from the top of a dune at Dunhuang, located between these two deserts (usually classified as part of the Taklamaken Desert). The main objectives of this study were to determine the bacterial richness and diversity of these samples using pyrosequencing of PCR-amplified V1V2 regions of 16S rDNA by PCR of total DNA extracted from the sand. The amplified and sequenced 16S rDNA segments were then characterized using a variety of bioinformatic tools. As these surface sands are considered a harsh environment, we wanted to determine if there were only a few bacterial groups or many hundreds, and begin to ask other questions, such as the following: Are these bacterial populations similar? Do they share many bacterial groups? Among the shared bacterial groups, are certain indigenous to both Asian deserts or to deserts in general?

Given their large temperature fluctuations, low water availability, high UV light exposure, low organic carbon, and high alkalinity (pH 8.5–9.8), deserts are thought to be relatively harsh and perhaps lifeless environments. Nonetheless, we found a large bacterial diversity residing in the surface sands of these two Asian deserts. The bacterial populations were statistically different, but members from 30 genera were found to be present in all five samples.

Methods

Study Sites and Sampling

With a territory of 9.6 million km^2, China is one of the most severely desertified countries in the world [11]. Deserts occupy approximately 27 % (2.5 million km^2) of China's total surface area [12]. The Gobi Desert is located in the northern part of China and southern part of Mongolia. It is one of the ten largest deserts in the world. The Taklamakan Desert is located in the northwest of China, in the region of Xinjiang, and covers an area of 337,600 km^2 [13]. The two deserts are also the major source of sand for the seasonal dust storms in Asia that occur during the spring and summer [14]. Two surface sand samples from the tops of dunes in the Taklamaken Desert were collected near Golmud (36°23′ N, 94°46′ E, October 2007, 2,818 m elevation, average temperature range during October, −1.2 to 12.9 °C) and near Korla (40°58′ N, 85°57′ E, March 2008, 910 m elevation, average temperature range during March, 1 to 13.3 °C) in China. One surface sand sample from the tops of dunes from Dunhuang (40°06′ N, 94°39′ E, October 2007, 1,149 m elevation, average temperature range during October, 0.6 to 18.8 °C) located between the Taklamaken and Gobi deserts in China was also collected. Two surface sand samples from the tops of dunes in the Gobi Desert in Mongolia were collected 30 km southeast of Dzuunbayan (44°18′ N, 110°06′ E, December 2007, 1,001 m elevation, average temperature of December, −20.4 to −9.8 °C) (Fig. 1). The analyses for the physicochemical parameters of these five samples were performed using standard methods by the Laboratoire d'Analyses de Sols (Arras, France).

DNA Extraction

Total DNA was extracted from each sand sample using a protocol adapted from that of Zhou et al. [15]. Five grams of sand were incubated at 30 °C with shaking for 1 h after addition of 1 ml 1/4 TS (Tryptic Soy Broth) to rehydrate the bacteria and minimize nutrient shock [16]. No significant change in bacterial composition was observed during this 1 h incubation using ARISA-PCR (data not shown). Then, 13.5 ml extraction buffer (100 mM Tris–HCl pH 8, 100 mM

Fig. 1 Asian desert sampling sites. Asian desert sampling sites are indicated by *arrowheads* on the satellite photo of East Asia from a NASA World Wind screenshot {{PD-WorldWind}}(http://an.wikipedia.org/wiki/Imachen:China_100.78713E_35.63718N.jpg). In the five sampling sites, Golmud and Korla are located in the Taklamaken Desert; Dunhuang is located between the Taklamaken and Gobi deserts. The *dark arrowhead* indicates the location of Beijing (China)

Na EDTA pH 8, 100 mM Na_2HPO_4, 1.5 M NaCl, 1 % [*w*/*v*] CTAB), containing 74 μg/ml pre-digested Pronase plus 6.7 μg/ml RNAse A, was added, followed by a 2-h incubation at 37 °C with mild shaking. Following this, 1.5 ml of a 20 % (*w*/*v*) SDS solution was added and incubation continued at 65 °C for a further 2 h. The supernatant fluid was collected after a 10-min centrifugation at 6,000×*g* at room temperature. The pellet was extracted one more time with 4.5 ml extraction buffer plus 2 % (*w*/*v*) SDS, mixed by vortexing for 10 s, followed by incubation for 10 min at 65 °C and, after centrifugation, the supernatant fluids were pooled. The nucleic acids were extracted by the addition of an equal volume of chloroform/isoamyl alcohol (24:1) to the pooled supernatant fluids and precipitated by the addition of 0.6 volumes of isopropanol for 1 h at room temperature, followed by centrifugation at 16,000×*g* for 20 min at 20 °C. The DNA pellet was washed with 70 % ethanol, followed by centrifugation at 16,000×*g* for 5 min at 20 °C. The DNA pellets were then air-dried and resuspended in 50 μl 1/10 TE buffer (1 mM Tris–HCl pH 8, 0.1 mM Na EDTA pH 8) at 4 °C overnight and stored at −20 °C until use. On average, approximately 10 ng of DNA was recovered from 1 g of surface sand.

Bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) PCR

An aliquot of extracted total DNA was adjusted to a final DNA concentration of 15 ng/μl in 1/10 TE buffer using a NanoVue spectrophotometer (GE Healthcare, Buckinghamshire, UK) and verified by ethidium bromide fluorescence after electrophoresis through a 1 % agarose gel in TAE (2 mM Tris-acetate pH 8, 5 mM Na-EDTA) buffer. Then, multiple 50 μl PCR reactions were performed using the universal 16S rDNA bacterial primers 8F (*B*xxxxxxAGAGTTTGATC MTGGCTCAG) and 357R (*A*xxxxxxCTGCTGCCTYC CGTA), where *B* and *A* represent the adaptors B and A for pyrosequencing using the Gold pyrosequencing reaction (GS20, Roche/454 Life Sciences, Branford, CT, USA). The xxxxxx represents six nucleotide sequence tags designed for sample identification barcoding [17, 18]. PCR amplification conditions were adapted for the use of three different thermostable DNA polymerases: (I) Phusion High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland)—98 °C for 2 min followed by 25 cycles of 98 °C for 30 s, 48 °C for 20 s and 72 °C for 12 s, and a final elongation step at 72 °C for 5 min; (II) Pfu DNA Polymerase (Fermentas, ON, Canada)—95 °C for 3 min followed by 35 cycles of 95 °C for 30 s, 48 °C for 30 s and 72 °C for 48 s, and a final elongation step at 72 °C for 5 min; (III) High Fidelity PCR Enzyme Mix (Fermentas)—94 °C for 3 min followed by 30 cycles of 94 °C for 30 s, 48 °C for 30 s and 72 °C for 24 s, and a final elongation step at 72 °C for 5 min. Each 50 μl PCR reaction contained 1–5 ng DNA, 0.1 μM of each primer (Sigma-Aldrich, MO, USA), 0.2 mM dNTP mix (Fermentas), 1.25 units of thermostable DNA polymerase for I and II, or 2 units of thermostable DNA polymerase for III using the buffers supplied with each polymerase. Each DNA sample was subjected to 5–10 PCR reactions per thermostable DNA polymerase, and two different polymerases were used per sample to minimize bias. The resultant PCR products were pooled and loaded on a 1 % agarose gel in TAE buffer. After electrophoreses and DNA visualization by ethidium bromide staining and long wave UV light illumination, the 16S amplified DNA fragment-containing regions were cut from the gel and purified using the NucleoSpin Extract II kit (Macherey-Nagel, North Rhine-Westphalia, Germany) according to the manufacturers' instructions. Fifty nanograms of PCR products from each sample were mixed for pyrosequencing.

bTEFAP FLX Pyrosequencing

Pyrosequencing was performed using a Roche/454 FLX Pyrosequencer (GATC Biotech, Konstanz, Germany). The sequences obtained for each sample were grouped according to the tag used and, after removal of the tags, the average sequence was found to be 227 nucleotides in length.

bTEFAP Sequence Processing Pipeline and Data Analyses

The sequences were selected by their length (>150 bases) and their quality score (90 % of nucleotides with a quality score >25) using the Greengenes website (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) [19]. Then, sequences with more than two errors in the primer or more than one ambiguous base were removed using the RDP pyrosequencing pipeline (http://pyro.cme.msu.edu/index.jsp).

The remaining sequences were then classified to the genus level using RDP-II Naive Bayesian Classifier [20, 21] with the confidence threshold set to 50 %. The sequences were also classified using the RDPII all-bacteria database to find the closest species at ≥95 % similarity, with an *e*-value≤e^{-20} and bit score >200 by Megablast. The unclassified sequences were then compared with the database in VITCOMIC [22], and sequences with >80 % similarity were classified to the phylum level according to the VITCOMIC results. We also normalized the number of sequences to 3,900 reads for each sample using the Pangea Pipeline [23] to compare the estimations for bacterial population richness and diversity. The Chao1 estimator and Shannon indices were calculated on the RDPII-pyro site, while the Bray–Curtis index was calculated using Species Prediction And Diversity Estimation (SPADE) [24]. The similarity level among the different OTUs was fixed at 97 %, sequences with more than 3 % difference were clustered into different OTUs by the complete-linkage clustering method on RDP's pyrosequencing pipeline (http://pyro.cme.msu.edu/spring/cluster.spr). We calculated the *p* value of the chi-square

Table 1 Physical and chemical properties of the desert sand samples

	Korla	Golmud	Dunhuang	Gobi1	Gobi2
Clays (<2 μm) [g/kg]	40	165	150	61	43
Silts (2–50 μm) [g/kg]	3	308	472	5	4
Sands (50–2,000 μm) [g/kg]	957	527	378	934	953
Organic carbon [g/kg]	1.12	4.22	7.07	0.515	0.296
Total nitrogen [g/kg]	0.101	2.47	0.58	0.108	0.102
C/N	11.1	1.71	12.2	4.77	2.91
Organic Material [g/kg]	1.94	7.31	12.2	0.891	0.512
pH	9.14	8.52	8.72	9.8	9.5

test for the bacterial populations from each pair of samples using R software (http://www.r-project.org/). Principal component analyses on the relative proportion of families or genera among the five samples were performed using the ade4 package adapted in R [25]. All sequences have been deposited in the GenBank Sequence Read Archive.

Results

We examined five surface sand samples from the Taklamaken (China) and Gobi (Mongolia) deserts in order to discern several chemical parameters and describe the potential bacterial diversity of these two large Asian deserts. The two samples of Gobi Desert, taken at 5 km intervals, are called Gobi1 and Gobi2, while the other three samples are named by sampling locations: Golmud, Korla, and Dunhuang.

Physical and Chemical Properties of Sand Samples

Several physical and chemical parameters were measured, and the results, summarized in Table 1, show that the Golmud and Dunhuang samples contained sand particles of widely different sizes, while the Korla, Gobi1, and Gobi2 samples contained mostly particles larger than 50 μm. In addition, these two groupings were also observed for the pH values: pH 8.5–9 (Golmud and Dunhuang) versus pH >9 (Korla, Gobi1, and Gobi2). The C/N ratios and level of organic material varied over 10-fold among the samples, ranging from 1.71 to 12.2, and from 0.512 to 12.2 g organic material/kg sand, respectively.

Pyrosequencing Data Analysis

We obtained 60,610 sequences for the five sand DNA amplicon samples after pyrosequencing and tag grouping. These were first trimmed using the Greengenes trim tool and cleaned using the quality filter on the RDPII pyro pipeline. In total, 44,724 reads remained after application of the quality controls. The average sequence length, after trimming, was improved from 227 bases to 240 bases. The Dunhuang sample contained the smallest number of sequences (3,956 sequences), while the Gobi2 sample contained the largest number (18,382 sequences) (Table 2).

Unclassified Sequences

Sequences with <95 % similarity in the database of "all bacteria" of RDPII were considered as "new" sequences (1,991 new sequences detected). These were analyzed using unclassified sequence selection in PANGEA after Megablast against the RDPII bacterial database, yielding 1.36 to 20.56 % of the total sequences per sample being classified as "new" sequences. These sequences could be classified as belonging to 740 OTUs at ≥97 % similarity level. Among these new sequences, less than 20 % of the OTUs were shared between each pair of samples. Using naïve Bayesian rRNA classifier on RDPII, approximately two thirds (62 %) could be classified at the phylum level. These were found to belong to the *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, and *Firmicutes* phyla.

Table 2 Number of processed pyrosequencing reads, OTU richness and diversity for each desert sand sample

Sample	Number of sequences	OTUs	Chao1		Shannon index	
			After trimming	Normalized (3,900 reads)	After trimming	Normalized (3,900 reads)
Korla	8,027	1,445	2,425	1,858	6.12	5.99
Golmud	8,435	988	1,429	1,075	5.58	5.39
Dunhuang	3,958	874	1,614	1,643	5.5	5.54
Gobi1	5,902	792	1,172	1,069	5.39	5.22
Gobi2	18,403	999	1,588	717	4.84	4.38

Table 3 Correlation coefficients for pH and C/N ratios

	R	*p* value (two tailed)	*p* value (one tailed)
pH vs richness	−0.486	0.407	0.203
pH vs diversity	−0.456	0.441	0.221
C/N vs richness	0.903[a]	0.035	0.018
C/N vs diversity	0.665	0.221	0.110

[a] Significant correlation

Bacterial Richness and Diversity in the Sand Samples

The OTUs, differing at <97 % sequence similarity and corresponding roughly to the level of species, ranged from 792 to 1,445 per sample among the five samples (Table 2). We used the relative abundance rank curve as a graphical representation of OTU composition. The curves here showed the relative abundance of the most abundant 40 OTUs in each sample (Fig. S1). Among the five samples, there are no predominant OTUs observed. The most abundant OTU found was from the Golmud sample, comprising 7.7 % of the population. The Korla sample showed a more flattened curve than the other samples, with the most abundant OTUs comprising only 4 % of the sample, leading to a more even distribution of OTUs. These results can thus partly explain its higher diversity index and richness (Table 2). In total, the 40 most abundant OTUs comprised from 37.2 to 65.4 % of the sequences in our samples. The numbers of OTUs with more than 1 % of the sequences ranged from 10 (Korla) to 22 (Gobi 2) OTUs in our samples. As over 1,100 OTUs were detected in each sample, this figure shows that the relatively high observed bacterial richness was a result of a large proportion of rare OTUs. The overall richness estimated by the Chao1 estimator ranged from 1,173 to 2,426 OTUs per sample using all the good quality sequences obtained after trimming, a result suggesting that our sampling revealed approximately 2/3 of the OTUs per sample, on average (Table 2 and Fig. S2). In total, 4,088 OTU phylotypes were discernible, with only 12 of them observed in all five samples, while just over half of the phylotypes (54.48 %) were represented by a single sequence. The Shannon index revealed that the sample with the highest diversity level is that from Korla, while the sample with the lowest diversity level is that of Gobi2, though it is interesting to note that the Gobi2 sample contained the largest number of sequences, suggesting that the number of sequences obtained was not the only limit to discerning bacterial biodiversity levels. We also compared the richness estimated by the Chao1 estimator and the measured diversity using the Shannon index after normalizing all samples to the same numbers (3,900 sequences per sample) of sequences using the selector.pl component of the Pangea Pipeline. No significant modification to the levels of bacterial diversity was observed (Table 2).

Using the normalized sequence data set to avoid sequencing number bias, soil pH was not found to have a significant correlation with either bacterial richness (p (one tailed)=0.20) or diversity (p (one tailed)=0.22). A significant correlation was observed between the C/N ratio and population richness (p (one tailed)=0.02), though not between the C/N ratio and population diversity (p (one tailed)=0.11) (Table 3).

Comparison of Bacterial Community Relationships Among the Desert Samples

The Jaccard index was calculated to reveal the unweighted dissimilarity, and the Bray–Curtis index was used to reveal the weighted dissimilarity among the five sequence samples for both the case of all good quality sequences and sequences which are normalized to 3,900 reads. For both of these indices, similar relationships among the samples were observed (Table 4). Among the five samples, the most closely related populations, using SPADE analysis, represented the samples from Gobi1 and Gobi2, sharing approximately half of the total

Table 4 Jaccard and Bray–Curtis dissimilarity indices of the desert samples

	Golmud	Dunhuang	Korla	Gobi1	Gobi2
Jaccard index					
Golmud	0				
Dunhuang	0.955 (0.949)	0			
Korla	0.807 (0.797)	0.740 (0.647)	0		
Gobi1	0.968 (0.963)	0.608 (0.562)	0.634 (0.622)	0	
Gobi2	0.670 (0.979)	0.521 (0.565)	0.650 (0.755)	0.203 (0.258)	0
Bray–Curtis index					
Golmud	0				
Dunhuang	0.967 (0.955)	0			
Korla	0.923 (0.916)	0.912 (0.905)	0		
Gobi1	0.978 (0.978)	0.739 (0.742)	0.832 (0.831)	0	
Gobi2	0.980 (0.983)	0.856 (0.739)	0.921 (0.918)	0.681 (0.541)	0

The numbers in parentheses are the values obtained using the normalized sequence (3,900 sequences) data sets

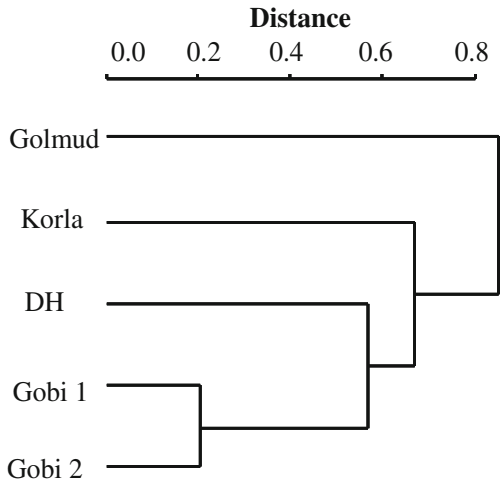

Fig. 2 A Jaccard's distance tree (at a ≥97 % sequence similarity level) of the bacterial population of the desert sand samples. The tree was generated on the RDP-pyro web server with the R Statistical Computing Package (UPGMA). The distance for each pair of samples is indicated by the position of the node between them, according to the Jaccard dissimilarity indices. *DH* = Dunhuang

OTUs (49.2 %). The most dissimilar samples were Golmud and Gobi2, which had only 8 % of their OTUs in common (Table 4). A Jaccard distance tree was generated on the RDP-pyro web server with the R Statistical Computing Package (UPGMA) in order to graphically reveal the relationships among these five samples (Fig. 2). The most similar bacterial populations were found in the Gobi1 and Gobi2 samples, while the bacterial population of the Golmud sample was found to be the most different from the others.

Bacterial Composition of the Desert Samples

We classified the sequences with the RDPII database, using RDP Classifier, to the genus level. Sequences representing bacteria belonging to 15 phyla, unequally distributed among the samples, were detected. The predominant phyla contained representatives from the *Firmicutes* (3.13–82.37 %), *Proteobacteria* (11.14–50.79 %), *Bacteroidetes* (1.66–41.27 %), and *Actinobacteria* (2.04–18.50 %) phyla (Fig. 3). The bacterial populations of the Gobi1 and Gobi2 samples, taken at a distance of 5 km, were found to be significantly different ($p < 0.0001$) using the chi-square test. Bacteria representing the less predominant (<4 % for each) phyla contained members belonging to the *Gemmatimonadetes*, *Acidobacteria*, *Deinococcus-Thermus*, and *Chloroflexi* phyla, while bacteria belonging to more rare phyla (*Tenericutes*, *Cyanobacteria*, *Verrucomicrobia*, *Nitrospira*, TM7, BRC1, and OD1) were not detected in all samples. Less than 5 % of the sequences were unable to be classified using the RDPII database at the phylum level. Principal component analyses were performed based on the relative composition of family level members for the five samples (Fig. 4a). The samples could be grouped into two clusters: Gobi1, Gobi2, and Dunhuang, plus Korla and Golmud. Members of the *Rhodobacteraceae*, *Chitinophagaceae*, *Alteromonadaceae*, *Cytophagaceae*, and *Flavobacteriaceae* families were found to play a key role underlying the variance among samples of the second group, while members of the *Bacillaceae* were relatively dominant in the first group. This relationship was maintained down to the level of genus, where members of the *Salinimicrobium*, *Pontibacter*, *Effiviibacter*, *Planococcus*, and *Acinetobacter* genera were found to play a key role underlying the variance among samples of the second group, while members of the genus *Bacillus* were relatively dominant in the first group (Fig. 4b). Five-set Venn diagrams were established at the family (Fig. 5a) and genus (Fig. 5b) levels for the five samples. These diagrams showed the shared families or genera among two, three, four, and all five samples. At the family level, members belonging to 30 of the families were shared among all samples. At the genus level, major differences of the five samples were revealed, as most genera were detected in only one sample.

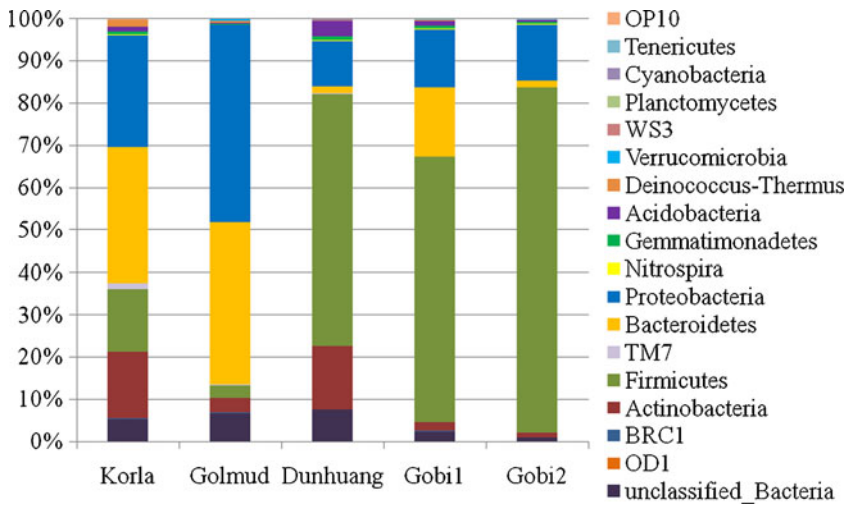

Fig. 3 Bacterial composition of the desert samples at the phylum level. The bacterial classification was performed using the RDPII Naive Bayesian Classifier with a 50 % confidence shoulder and VITCOMIC, as described in the "Methods"

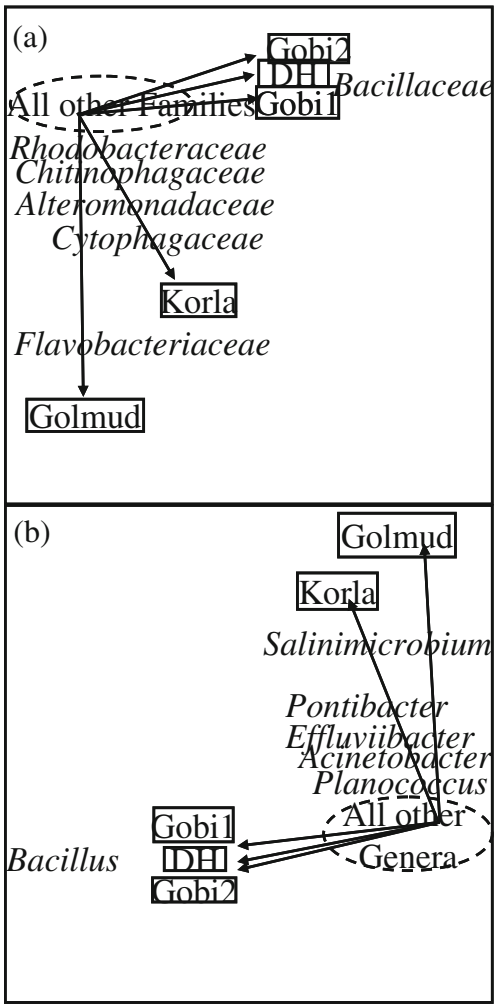

Fig. 4 Principal component analyses of the desert samples at the family or genus level. The figures were generated by the ade4 package in R software, simplified by grouping the less important, overlapped families or genera. **a** Principal component analyses of the proportion of bacterial families in the desert sand samples. The *samples in the boxes* were placed by their relative abundance of families in the *oval*. **b** Principal component analyses of the proportion of bacterial genera in the desert sand samples. The *samples in the boxes* were placed by their relative abundance of genera in the *oval*

A total of 413 genera were identified among the OTUs, with 189 genera belonging to the phylum *Proteobacteria*, 73 genera belonging to the phylum *Firmicutes*, and 68 genera belonging to the *Actinobacteria*. Among these, OTUs corresponding to 30 genera (Table 5) were detected in all five samples and represent 56.13 % of the total sequences. Twenty-six of the genera contain members detected in various types of environmental samples, while members of four of the genera have been identified primarily in samples from desert-like environments (Table 5), as determined through Pubmed searches (not shown). Eight members of the 30 genera present in all five samples have been previously reported in samples from the Taklamaken or Gobi deserts. The total proportions of these 30 genera in each sample varied from 19.4 % (in the Golmud sample) to 74.8 % (in the Gobi 2 sample), while their distribution in each sample is displayed in Fig. 6, showing the varying proportions of each common genus among the five samples.

Discussion

To reveal the potential bacterial diversity and populations in samples from the two largest Asian deserts, we examined five surface sand samples from the Taklamaken and Gobi deserts. Using pyrosequencing, we obtained 44,724 high quality sequences and examined the bacterial richness and diversity in each sample. We classified these sequences to the genus level,

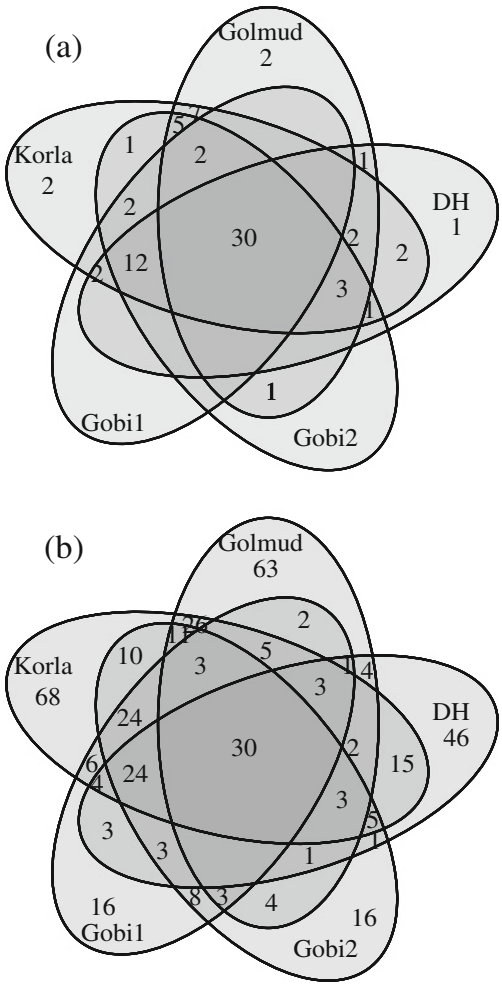

Fig. 5 Venn diagrams. Venn diagrams of the bacterial groups of the desert sand samples representing the distribution of the number of families (**a**) or genera (**b**). The samples are represented by *different circles*: Golmud, Dunhuang, Korla, Gobi1, and Gobi2. **a** Venn diagram representing the number of families detected in each sample and the overlap of families among the samples. **b** Venn diagram representing the number of genera detected in each sample and the overlap of genera among the samples

Table 5 The common genera present in all the surface sand samples

Genus ID	Number of sequences	Mean % of population	Range of population (%)
The common genera observed in Asian deserts and other environments			
Bacillus	15,608	28.86	0.02–62.10
Pseudomonas	1,428	1.65	0.03–7.40
Effluviibacter	761	1.95	0.03–8.96
Acinetobacter	671	1.69	0.92–2.67
Adhaeribacter	499	1.33	0.03–2.85
Massilia	491	1.00	0.01–1.88
Arthrobacter	378	0.97	0.22–2.99
Flavisolibacter	332	0.83	0.08–2.34
Lysobacter	227	0.57	0.13–1.45
Herbaspirillum	211	0.54	0.01–1.56
Devosia	209	0.51	0.04–1.16
Gemmatimonas	204	0.51	0.05–0.88
Porphyrobacter	203	0.49	0.03–1.18
Gp6	182	0.63	0.01–1.97
Paracoccus	156	0.42	0.03–1.18
Planomicrobium	135	0.36	0.04–1.01
Nocardioides	124	0.35	0.02–1.23
Gp4	113	0.33	0.01–0.73
Sphingomonas	100	0.24	0.09–0.76
Rubellimicrobium	86	0.21	0.04–0.60
Novosphingobium	84	0.18	0.03–0.17
Truepera	59	0.14	0.01–0.57
Comamonas	44	0.10	0.05–0.19
Blastococcus	43	0.10	0.02–0.31
Pedobacter	42	0.10	0.01–0.24
Sphaerobacter	27	0.09	0.02–0.23
The common genera observed only in desert-like environments			
Pontibacter	1,101	3.06	0.18–7.09
Salinimicrobium	944	2.27	0.01–8.60
Planococcus	344	0.93	0.08–1.92
Marmoricola	299	0.78	0.03–3.30

revealing a large degree of bacterial diversity and significant differences in bacterial populations among these samples.

Our study describes, for the first time, the potential bacterial composition and diversity in highly alkaline desert soils. We used the V1V2 variable region amplicons of 16S rDNA for this study, as the V1V2 region is considered an accurate region to examine uncultured bacterial diversity [21, 26]. We obtained an average of 8,945 reads per sample but were unable to reveal the entire bacterial population biodiversity for each sample, given the large absolute diversity we uncovered [27, 28]. Nonetheless, we were able to discern >60 % of the OTU phylotypes, calculated by the Chao1 estimator, of the projected microbial population, at the species (OTU) level for each sample. Even with over 18,000 sequences for the Gobi2 sample, only 62 % of the predicted OTUs were revealed. Our results showed a large degree of bacterial diversity in this dry and alkaline environment. The desert soil samples were each found to contain more than 1,000 phylotypes (at a ≥97 % sequence similarity level) by the Chao1 estimator, comparable to that observed in other soil types [29].

In total, though 4,088 OTUs, differing at <97 % sequence similarity and corresponding roughly to the level of species, were distinguishable, only 12 OTUs were found to be present in all five desert samples, suggesting that the variability of the bacterial populations among the five desert sand samples was relatively high. Since we found 30 genera in common among the five samples, the five samples likely contain different OTUs. The differences in bacterial composition among the samples were statistically significant at the OTU level, even for the two most similar samples (Gobi1 and Gobi2). Our results suggest that desert soil contains many different OTUs, as previously observed in other soil samples taken as close as 1 mm in distance [30–32].

The similar shape of rank abundance curves reflected the fact that the structures of bacterial diversity in the five samples were similar. We observed that there are no predominant phylotypes which comprised >10 % of total bacteria. Instead, about 40 phylotypes in each sample could be considered as higher abundance phylotypes (>0.5 %) than the others. As more than 792 phylotypes were detected in each of our samples, we found that most phylotypes represented relatively rare OTUs in our sand samples. Similar distributions of OTUs have been observed in other soil samples [29, 33]. It has been reported that pH may have the potential to predict bacterial richness and diversity in soil samples [29, 34, 35]. However, in soils with a very high pH (>8.5), and receiving very low precipitation levels, the results were more limited [29, 36]. Among our five sand samples, the pH was greater than 8.5 (pH 8.52 to pH 9.8), and no significant correlation with either bacterial population parameter versus pH was observed. However, we did observe a significant (r=0.90, r^2=81 %, p< 0.05) increase in phylotype numbers with increasing C/N ratio, as the C/N ratio of our samples ranged from 1.71 to 12.2.

Since the sampling seasons for the five samples were different, the Gobi samples were taken under a much lower local temperature (−20 to −9.8 °C) compared with the other three samples (−1.2 to 18.8 °C). This environmental difference might explain the decreased richness and diversity observed in the Gobi samples using only the normalized data. The effects of temperature on the bacterial diversity in desert-type environments have been previously studied. In a study on the Chihuahuan Desert grassland, it was found that bacteria were able to respond to moisture pulses regardless of temperature [37], though an examination of desert cyanobacterial soil crusts suggested that seasonality should also be taken into consideration [38]. The key environmental factors affecting bacterial community composition may vary among different

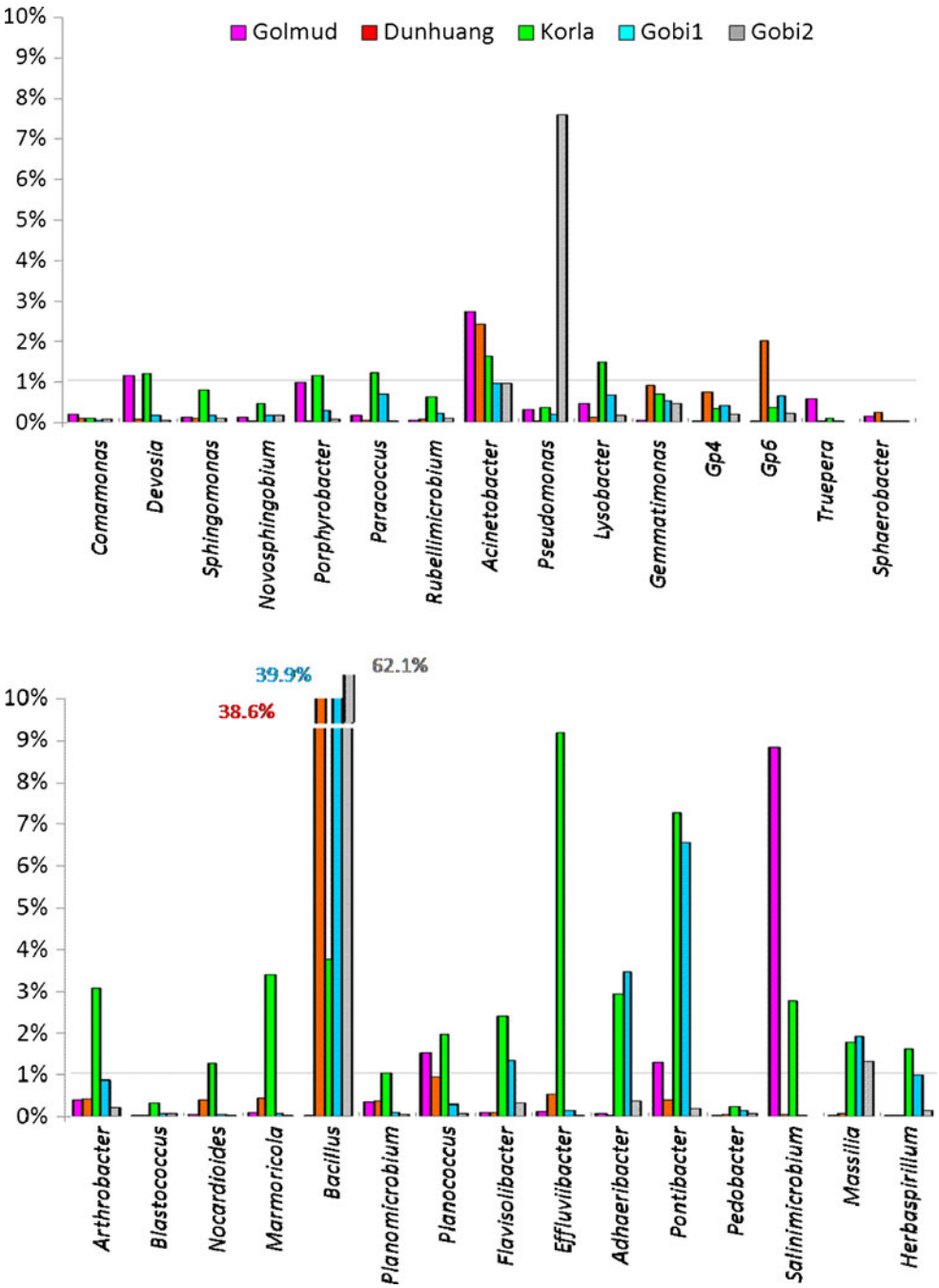

Fig. 6 The relative abundance of each common genus in the sand samples

desert environments. In advance of our sequencing results, the need for replicas was not apparent. Their absence will be rectified in the future, as the relatively large apparent bacterial diversity that we found means that the examination of the microbial diversity of surface desert sands will require an extensive sampling and sequencing effort. Clearly, further studies, including an examination of the temporal variation of bacterial diversity in (Asian) deserts at the same sampling sites, are necessary.

The five desert soil samples were found to be dominated by bacterial members belonging to four phyla: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria*. Members of these phyla account for more than 90 % of the sequences in each sample. Nonetheless, the percentage of the population they represent in each sample is varied. It is not surprising that with the highly alkaline pH of these samples, fewer members of the *Acidobacteria* were observed compared to other soil samples.

Our results suggest that members belonging to four bacterial genera, *Pontibacter*, *Salinimicrobium*, *Planococcus*, and *Marmoricola*, may be classified as indigenous desert types, as their members have been most often detected in desert-like environments, including the results of Lauber et al., which examined 35 soil types using 88 soil samples. Concerning the genus *Pontibacter* (phylum *Bacteroidetes*), three species have been detected in the Taklamaken Desert region [39–41], while we discerned members of the genus *Pontibacter* once in a Mojave Desert soil sample using the data of Lauber et al. Members of the genus *Salinimicrobium* (phylum *Bacteroidetes*) have been reported in a saline lake of the Taklamaken Desert region [42], and all members are considered halophilic bacteria. For the genus *Planococcus* (phylum *Firmicutes*), members of this genus have been found in alkaline environments such as marine solar salterns or harsh environments such as the cold deserts of the Himalayas or Antarctica. *Planococcus* members have been detected twice in Sahara Desert dust events in Mali, West Africa [43] and in Erdemli, Turkey [44], and we detected members of this genus in three desert samples using the data of Lauber et al. Members of the genus *Marmoricola* (phylum *Actinobacteria*) have been isolated from volcanic ash [45], a marble statue [46], a beach [47], and a Korean soil sample [48]. Analyses of the data of Lauber et al. showed that members of the genus *Marmoricola* were detected in all desert-type soil samples, and in some samples from soil with minimal levels (300–800 mm per year) of precipitation.

From the 1,991 new sequences that we detected, we could distinguish 740 OTUs, suggesting that there are likely unknown types of bacteria present in Asian desert sands. The Asian deserts are a potential source to discover new bacteria with particular properties such as alkali and/or drought tolerance, and as a source to isolate novel enzymes [3].

Thus, our results demonstrated a large microbial diversity, by culture-independent methods, present in surface sand samples from two Asian deserts. Many unknown bacterial species appeared to be present and remain to be identified and characterized, from these unique and interesting ecosystems.

Acknowledgments We thank Barry Holland (Institut de Génétique et Microbiologie, Université Paris-Sud, France) and Denis Faure (Institut des Sciences du Végétal, CNRS, Gif-sur-Yvette, France) for interesting discussions and the reviewers for their insightful comments and suggestions. This work was supported by the Centre National de la Recherche Scientifique, France and by the AQUAPHAGE program of the Agence Nationale de la Recherche, France.

References

1. Bhatnagar A, Bhatnagar M (2005) Microbial diversity in desert ecosystems. Curr Sci 89:91–100
2. Lima DM, Oliveira RQ, Uetanabaro AP, Goes-Neto A, Rosa CA, Assis SA (2009) Thermostable inulinases secreted by yeast and yeast-like strains from the Brazilian semi-arid region. Int J Food Sci Nutr 60(Suppl 7):63–71
3. Neveu J, Regeard C, DuBow MS (2011) Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. Appl Microbiol Biot 91:635–644
4. Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA, Farrell RL (2009) Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci USA 106:19964–19969
5. Drees KP, Neilson JW, Betancourt JL, Quade J, Henderson DA, Pryor BM et al (2006) Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. Appl Environ Microbiol 72:7902–7908
6. Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, Tian F et al (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. Extremophiles 16:553–566
7. Liu M, Dai J, Liu Y, Cai F, Wang Y, Fang C (2011) *Desertibacter roseus* gen. nov., sp. nov., a gamma radiation-resistant bacterium in the family *Rhodospirillaceae*, isolated from desert sand. Int J Syst Evol Microbiol 61:1109–1113
8. Zhang L, Wu GL, Wang Y, Dai J, Fang CX (2010) *Bacillus deserti* sp. nov., a novel bacterium isolated from the desert of Xinjiang, China. A Van Leeuw 99:221–229
9. Liu M, Liu Y, Wang Y, Luo X, Dai J, Fang C (2010) *Lysobacter xinjiangensis* sp. nov., a moderately thermotolerant and alkalitolerant bacterium isolated from gamma-irradiated sand soil sample. Int J Syst Evol Microbiol 61:433–437
10. An H, Xu M, Dai J, Wang Y, Cai F, Qi H et al (2010) *Sphingomonas xinjiangensis* sp. nov., isolated from the desert of Xinjiang, China. Int J Syst Evol Microbiol 61:1865–1869
11. Zha Y, Gao J (1997) Characteristics of desertification and its rehabilitation in China. J Arid Environ 37:419–432
12. Ma J, Ding Z, Edmunds WM, Gates JB, Huang T (2009) Limits to recharge of groundwater from Tibetan Plateau to the Gobi Desert, implications for water management in the mountain front. J Hydrol 364:128–141
13. Li X, Yan P, Wu L (2003) Resources of medicinal plants living in sands in Xinjian areas. Prim J Chinese Materia Medica 17:54–56
14. Xia X, Chen H, Wang P (2004) Aerosol properties in a Chinese semiarid region. Atmos Environ 38:4571–4581
15. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl Environ Microbiol 62:316–322
16. Reasoner D, Geldreich E (1985) A new medium for the enumeration and subculture of bacteria from potable water. Appl Environ Microbiol 49:1–7
17. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 5:235–237
18. Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, Nielsen R et al (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:e197
19. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072
20. 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
21. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. Nucl Acids Res 35:e120
22. Mori H, Maruyama F, Kurokawa K (2010) VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences. BMC Bioinfo 11:332
23. Giongo A, Crabb DB, Davis-Richardson AG, Chauliac D, Mobberley JM, Gano KA et al (2010) PANGEA: pipeline for analysis of next generation amplicons. ISME J 4:852–861

24. Chao A, Shen T-J (2010) Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide published at http://chao.stat.nthu.edu.tw. Accessed 12 Oct 2010
25. Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package-I: one-table methods. R news 4:5–10
26. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8:R143
27. Teixeira LC, Peixoto RS, Cury JC, Sul WJ, Pellizari VH, Tiedje J et al (2010) Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. ISME J 4:989–1001
28. Guss AM, Roeselers G, Newton IL, Young CR, Klepac-Ceraj V, Lory S et al (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. ISME J 5:20–29
29. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75: 5111–5120
30. Noha HY, Couger MB, Mostafa SE, Niyaz A (2010) Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere. PLoS One 5:e12414
31. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK et al (2010) Global patterns in the biogeography of bacterial taxa. Environ Microbiol 13:135–144
32. Fulthorpe RR, Roesch LF, Riva A, Triplett EW (2008) Distantly sampled soils carry few species in common. ISME J 2:901–910
33. Kirchman DL (2012) Processes in microbial ecology. Oxford University Press Inc., New York
34. Bachar A, Al-Ashhab A, Soares MI, Sklarz MY, Angel R, Ungar ED et al (2010) Soil microbial abundance and diversity along a low precipitation gradient. Microb Ecol 60:453–461
35. Angel R, Soares MI, Ungar ED, Gillor O (2010) Biogeography of soil archaea and bacteria along a steep precipitation gradient. ISME J 4:553–563
36. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103:626–631
37. Bell C, McIntyre N, Cox S, Tissue D, Zak J (2008) Soil microbial responses to temporal variations of moisture and temperature in a Chihuahuan Desert grassland. Microb Ecol 56:153–167
38. Bowker MA, Reed SC, Belnap J, Phillips SL (2002) Temporal variation in community composition, pigmentation, and $F_{(v)}/F_{(m)}$ of desert cyanobacterial soil crusts. Microb Ecol 43:13–25
39. Wang Y, Zhang K, Cai F, Zhang L, Tang Y, Dai J et al (2010) *Pontibacter xinjiangensis* sp. nov., in the phylum 'Bacteroidetes', and reclassification of [*Effluviibacter*] *roseus* as *Pontibacter roseus* comb. nov. Int J Syst Evol Microbiol 60:99–103
40. Zhou Y, Wang X, Liu H, Zhang KY, Zhang YQ, Lai R et al (2007) *Pontibacter akesuensis* sp. nov., isolated from a desert soil in China. Int J Syst Evol Microbiol 57:321–325
41. Zhang L, Zhang Q, Luo X, Tang Y, Dai J, Li Y et al (2008) *Pontibacter korlensis* sp. nov., isolated from the desert of Xinjiang, China. Int J Syst Evol Microbiol 58:1210–1214
42. Lim JM, Jeon CO, Lee SS, Park DJ, Xu LH, Jiang CL et al (2008) Reclassification of *Salegentibacter catena* Ying et al. 2007 as *Salinimicrobium catena* gen. nov., comb. nov. and description of *Salinimicrobium xinjiangense* sp. nov., a halophilic bacterium isolated from Xinjiang Province in China. Int J Syst Evol Microbiol 58: 438–442
43. Kellogg CA, Griffin DW, Garrison VH, Peak KK, Royall N, Smith RR et al (2004) Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. Aerobiologia 20:99–110
44. Griffina DW, Kubilayb N, Koc-akb M, Graya MA (2007) Airborne desert dust and aeromicrobiology over the Turkish Mediterranean coastline. Atmos Environ 41:4050–4062
45. Lee DW, Lee SD (2010) *Marmoricola scoriae* sp. nov., isolated from volcanic ash. Int J Syst Evol Microbiol 60: 2135–2139
46. Urzi C, Salamone P, Schumann P, Stackebrandt E (2000) *Marmoricola aurantiacus* gen. nov., sp. nov., a coccoid member of the family Nocardioidaceae isolated from a marble statue. Int J Syst Evol Microbiol 50:529–536
47. Lee SD (2007) *Marmoricola aequoreus* sp. nov., a novel actinobacterium isolated from marine sediment. Int J Syst Evol Microbiol 57:1391–1395
48. Dastager SG, Lee JC, Ju YJ, Park DJ, Kim CJ (2008) *Marmoricola bigeumensis* sp. nov., a member of the family Nocardioidaceae. Int J Syst Evol Microbiol 58:1060–1063