



HAL
open science

The bacterial communities of sand-like surface soils of the San Rafael Swell (Utah, USA) and the Desert of Maine (USA)

Yang Wang

► **To cite this version:**

Yang Wang. The bacterial communities of sand-like surface soils of the San Rafael Swell (Utah, USA) and the Desert of Maine (USA). Agricultural sciences. Université Paris-Saclay, 2015. English. NNT : 2015SACLS120 . tel-01261518

HAL Id: tel-01261518

<https://theses.hal.science/tel-01261518>

Submitted on 25 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2015SACLS120

THESE DE DOCTORAT
DE L'UNIVERSITE PARIS-SACLAY,
préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N°577
Structure et Dynamique des Systèmes Vivants

Spécialité de doctorat : Sciences de la Vie et de la Santé

Par

Mme Yang WANG

The bacterial communities of sand-like surface soils of the San Rafael Swell (Utah, USA) and the Desert of Maine (USA)

Thèse présentée et soutenue à Orsay, le 23 Novembre 2015

Composition du Jury :

Mme. Marie-Claire Lett , Professeure, Université Strasbourg, Rapporteur
Mme. Corinne Cassier-Chauvat , Directeur de Recherche, CEA, Rapporteur
M. Armel Guyonvarch, Professeur, Université Paris-Sud, Président du Jury
M. Olivier Martin, Directeur de Recherche, INRA, Examineur
M. Michael DuBow, Professeur, Université Paris-Sud, Directeur de Thèse



Acknowledgements

There are so many persons I would like to thank for helping me in the study and in the life as a Ph.D. student in France. Firstly, I would like to express my sincere gratitude to my supervisor, Pr. Michael S. DuBow, for his continuous support and guidance. His guidance helped me in all the time of research and writing of this thesis, with his patience and immense knowledge.

Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Roland Matrippolito and Prof. Olivier Lespinet, for their insightful comments and encouragement.

My study in France was supported by the China Scholarship Council. I am very grateful for the great opportunity they offered me to become a Ph.D. student in the university of Paris-Sud and financial supports in life.

I am grateful to my fellow labmates of the team of Prof. Michael DuBow: Jorge Osman, Nga Nguyen, Gustavo Ribeiro, Christophe Regeard, and Micheline Terruer, for helping me solving the problems in research and their supports in my life in France.

Last but not the least, I would like to thank my family, my parents and my beloved husband for their encouragement all the time, and sharing with me all the happiness and sadness. Without their unconditional love and supports, I could not reach so far.

Table of contents.

Acknowledgements.....	3
List of Figures.....	7
List of Tables.....	8
List of abbreviations	9
CHAPTER 1. Introduction.....	10
1.1. Identification of bacterial diversity and Next Generation Sequencing technology.....	11
1.1.1. Introduction on the development of techniques for the identification of bacteria.....	11
1.1.2. Introduction to 16S ribosomal RNA.....	14
1.1.3. Metagenomics and the Next Generation Sequencing techniques	17
1.1.4. Bioinformatics for Meta-barcode Sequencing.....	24
1.2. Arid Environments	30
1.2.1. Introduction to Arid Environments	30
1.2.2. Characteristics of Drylands and Desert Soils.....	34
1.2.3. Drylands in North America.....	37
1.2.4. Analogue Sites for Mars	40
1.2.5. Desertification.....	41
1.3. Desert Biodiversity.....	43
1.3.1. Plants and Animals.....	43
1.3.2. Desert Microbiology	45
1.3.3. Bacterial Diversity	48
1.3.4. Bacteria in Desert-like Environments and Bacteria involved in Weathering	51
1.4. Relationship of Environmental Factors and Soil Bacterial Diversity	53
1.4.1. General introduction to the relationships of environmental factors and soil microbial diversity.....	53
1.4.2. Environmental factors in Deserts.....	55

1.4.3. Adaptation of Bacteria to Arid Environments	59
1.5. Objective of Thesis	61
CHAPTER 2: Bacterial Communities of the Desert of Maine	62
Abstract	64
Introduction	65
Materials and Methods	67
Results	70
Discussion	73
Acknowledgements	78
References	79
Tables	85
Figure Legends	89
CHAPTER 3: Bacterial communities of the deserts in Utah	93
Abstract	95
Introduction	96
Materials and Methods	99
Results	102
Discussion	107
Acknowledgements	112
References	113
Tables	121
Figure Legends	127
4. CHAPTER 4: Discussion and perspective	132
4.1. Discussion	133
4.1.1. Quality Control	134
4.1.2. Chimera Check	138
4.1.3. Taxonomic Classification of Sequences	142
4.1.4. OTU Clustering	145
4.2. Perspective	147

References..... 149

List of Figures

Figure 1.1. Nucleic acid-based methods in identifying bacteria.....	13
Figure 1.2. Distribution of variable regions in 16S rRNA from <i>E. coli</i> (Chuan et al 2014).	14
Figure 1.3. Three-domain system based on taxonomy of 16S rRNA, Carl Woese and George E. Fox, 1977	15
Figure 1.4. 16S rRNA secondary structure, adapted from Y. Pablo (2014).....	17
Figure 1.5. The 454 pyrosequencing chemistry (Petrosino et al 2009)	20
Figure 1.6. Process of Pyrosequencing (Margulies et al 2005).	21
Figure 1.7. Process of Illumina Sequencing	22
Figure 1.8. Analysis Process of Data from Meta-barcode Sequencing.....	24
Figure 1.9. Map of Global Distribution of Dry Land. Regions with declined color follows the aridity index (AI) of the UNEP. Arrows show the major intercontinental trajectories for desert dust (Laitty 2009).....	31
Figure 1.10. Classification of deserts. The pictures are from three different types of deserts: A. the Arabian Desert; B. the Gobi Desert; C. the Namibi Desert	32
Figure 1.11. The major types of desert landscapes: A. sand desert, B. stone desert, C. rock desert, D. plateau deserts, E. mountain deserts	35
Figure 1.12. Map of North American deserts (Laity 2009)	37
Figure 4.1. Quality score distribution of raw sequences.....	135
Figure 4.2. Comparison of quality score distribution of sequences trimmed using the three different methods.....	136
Figure 4.3. Number of chimeras detected by the Uchime and Decipher.	139
Figure 4.4. Number of sequences detected by the two programs in the Kumtagh sample	140

List of Tables

Table 1.1. Comparison of different methods to calculate distances between microbial communities	28
Table 1.2. Summary of studies of hot deserts (Makhalanyane et al 2015)	33
Table 4.1. The Phred quality score related to the base calling error probabilities.	134
Table 4.2. Comparison of different methods for quality control of sequences (original number of sequences is 22320).	136
Table 4.3. Diversity result before and after trimming.	137
Table 4.4. Percentage of potential non-chimeric sequences classified at different taxonomic levels.	140
Table 4.5. Summary of sequences in each database assigned to different taxonomic level	143
Table 4.6. Classification result of a sample from desert in Utah	143
Table 4.7. Summary of classification using different classifiers with 1000 sequences randomly selected from gold_database downloaded from Usearch.	144
Table 4.8. Number of OTUs clustered using different software at a distance 3%.	146

List of abbreviations

ATP: adenosine tri-phosphate

BLAST: Basic Local Alignment Search Tool

FISH: catalyzed reporter deposition fluorescent *in situ* hybridization

CFU: colony-forming units

DAPI: 4,6-diamidino-2-phenylindole

DGGE: Denaturing Gradient Gel Electrophoresis

DNA: Deoxyribonucleic acid

FISH: fluorescence *in-situ* hybridization

NGS: next-generation sequencing

OTU: operational taxonomic unit

PAR: photosynthetically active radiation

PCoA: Principal coordinates analysis

PCR: Polymerase Chain Reaction

qPCR: Real time quantitative PCR

*RD*P: Ribosomal Database Project

rRNA: Ribosomal Ribonucleic Acid

RNA: ribonucleic acid

SRA: Sequence Read Archive

T-RFLP: terminal restriction fragment length polymorphism

UNEP: United Nations Environment Programme

UPGMA: Unweighted Pair Group Method with Arithmetic Mean

UV: UltraViolet

CHAPTER 1. Introduction

1.1. Identification of bacterial diversity and Next Generation

Sequencing technology

1.1.1. Introduction on the development of techniques for the identification of bacteria

Since microorganisms were first recognized after the invention of microscopes in the 17th century, scientists have been looking for ways to isolate microorganisms in pure culture and to characterize isolates to differentiate them from each other [1]. Methods that can be used to characterize microbial isolates, as well as for classification of microorganisms, are essential for research in microbiology.

Traditional methods of bacterial identification based on phenotypic differences of organisms relied on cultivation under laboratory conditions, using gram staining, colony morphology differences and biochemical tests. However, these methods of bacterial identification had two major drawbacks. First, they can only be used for organisms that are able to be cultivated *in vitro*, which is extremely biased, as it selects only the small minority of bacteria that can grow in a laboratory situation [2]. As well, it frequently takes a long time to delineate characteristics of a clone culture, especially for those that grow slowly. Second, some strains exhibit unique phenotypic characteristics that do not fit into patterns that have been classified as marks of any known microbial categories [3]. These limits to databases will often not be suitable for identification and classification of environmental isolates.

The development of molecular biology offered a set of powerful new tools to accelerate the identification of microorganisms, which allows the examination of nucleic acids and detection of small variations within microbial species and even

within individual strains [4,5]. Accurate and definitive microorganism identification has been used in a wide variety of research and applications [6], including disease diagnosis associated with microbial infections, food production, agriculture and environmental studies [7].

Carl Woese, in 1977, first found that the 16S rRNA gene, one of the genes that makes ribosomal RNA, has evolutionary relationships in all prokaryotic organisms, and the distances of the sequences of 16S rRNA between different organisms likely indicates evolutionary distances [8]. With the development of the Polymerase Chain Reaction (PCR) techniques in the last decade of the 20th century and improvements in nucleic acid sequencing techniques, ribosomal RNA (or DNA) sequencing gradually became the basis for classification of microorganisms and the new gold standard for the taxonomy of microorganisms [9,10].

Now, methods of bacterial identification can be broadly sorted into two categories-genotypic techniques based on profiling an organism's genetic material, and phenotypic techniques based on profiling either metabolism or chemical composition of microorganisms [11]. The two methods focus on different characteristics of microorganisms and are frequently combined as complementary approaches in applications of microbiology research.

Phenotypic methods. Phenotypic techniques can generate direct functional information, such as activities of certain enzymes, and metabolic activities in specific groups of organisms [12]. They can also offer valuable information on physical and functional activities at the protein level. Phenotypic methods include traditional morphological characteristics, biochemical testing (presence of various enzymes), serological tests (test on specific antibodies, such as ELISA, and Western blotting), Phage typing (susceptibility to various phages), fatty acid profiles, etc.[13]

Genotypic methods. Genotypic techniques in general have the advantage over phenotypic methods of being independent of the physiological state of an organism (such as conditions or stresses in growth) [2]. They are based on the profile of a universal component (DNA or RNA), and certain of these fragments or sequences are unique to an individual organism or a group of closely related organism [14]. By comparing the information of different sequences generated by these fragments, organisms can be identified or classified [15]. Genotypic techniques can be classified into two major categories: fingerprint-based and sequence-based, and PCR is generally essential to both methods. The most commonly used nucleic acid-based methods in identifying bacteria are shown in **Fig 1**.

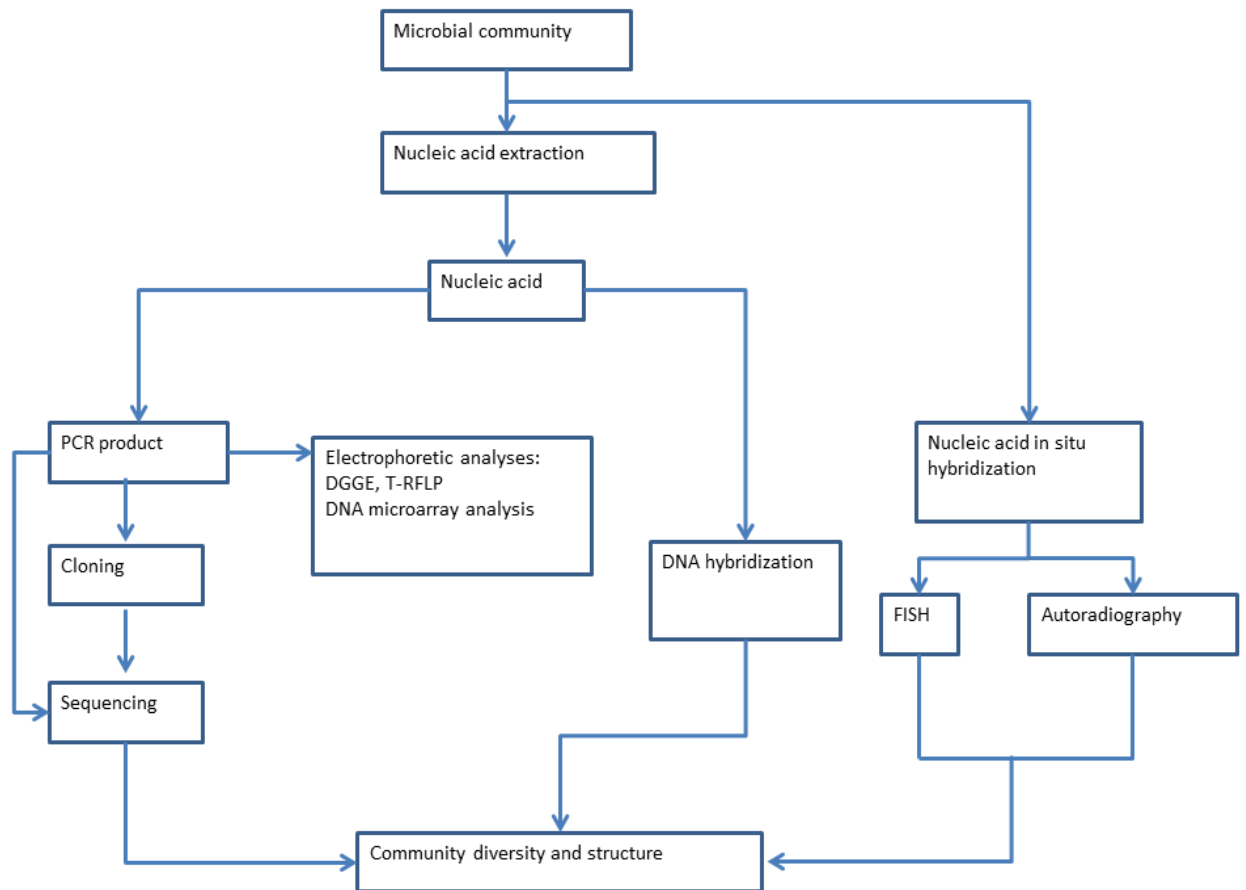


Figure 1.1. Nucleic acid-based methods in identifying bacteria.

1 1.1.2. Introduction to 16S ribosomal RNA

The 16S small ribosomal RNA gene (16S rRNA) is a component of the 30S small subunit of prokaryotic ribosomes. The genes for ribosomal RNA have evolved as organisms evolved, and the slight changes that have occurred can provide clues as to how closely or distantly various organisms are related. The 16S rRNA genes comprise nine hypervariable regions (V1-V9) that demonstrate considerable sequence diversity among different bacteria, and nine conserved regions (C1-C9) that remain consistent across different bacterial groups [16]. The distribution of these regions in 16S rRNA are shown in **Figure 2**.

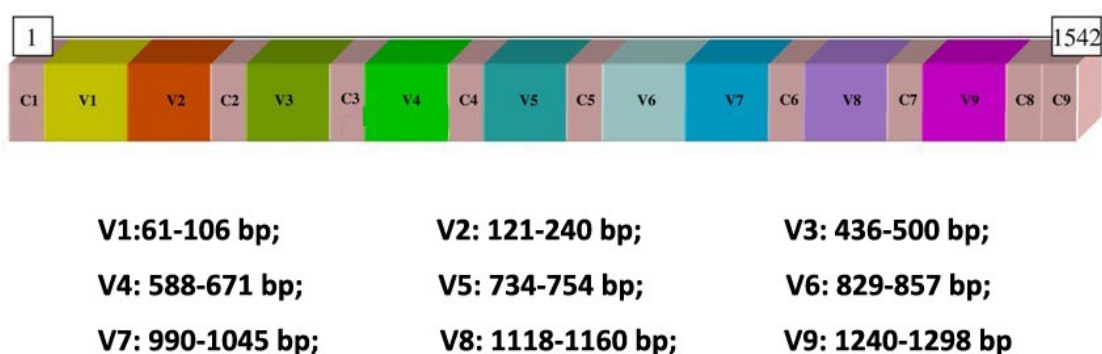


Figure 1.2. Distribution of variable regions in 16S rRNA from *E. coli* (Chuan et al 2014).

PCR amplification and analysis of the 16S rRNA genes have been widely used as a culture-independent method for documentation of the evolutionary history and taxonomic assignment of individual organisms, as well as in characterization of microbial communities [17]. The 16S rRNA gene has a number of clear advantages that make it optimal as a marker for these types of studies [18]:

- i. The length of this gene is convenient for amplification and sequencing, and certain length of its fragments (one or more variable regions) are sufficient for classification of sequences to deep level.
- ii. The highly conserved regions can be used for design of universal PCR primers;

- iii. The variable regions of this gene allow for accurate taxonomic classification and phylogenetic identification of microbial communities [19];
- iv. Lateral transfer of this gene between taxa are rare [19];
- v. Since this gene has been widely sequenced in microbial diversity research, there are many reference databases including a large amount of sequences with taxonomic information, which is convenient to assign query sequences to known taxonomic groups and compare community composition across studies [20].

In 1977, Carl Woese and George E. Fox introduced phylogenetic taxonomy of 16S ribosomal RNA and the three-domain system “tree of life” (**Figure 3**), through analysis of the 16S rRNA sequences [21]. They first discovered the third domain Archaea, and separated it from Bacteria. Over the past two decades, 16S rRNA gene sequencing analysis has made great studies, and has become the “gold standard” for the taxonomy of microorganisms.

Phylogenetic Tree of Life

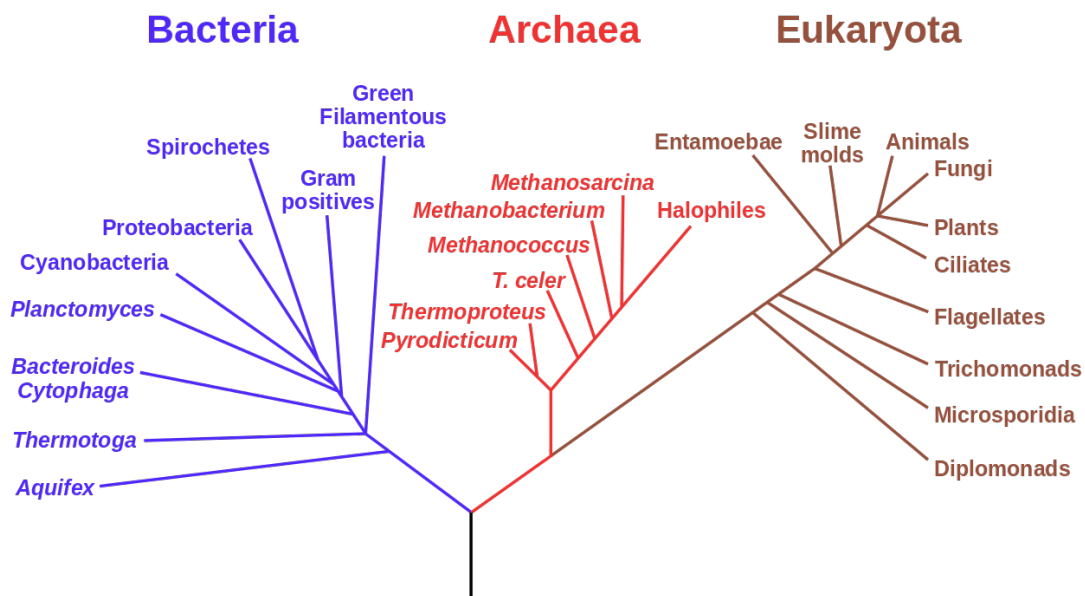


Figure 1.3. Three-domain system based on taxonomy of 16S rRNA, Carl Woese and George E. Fox, 1977

Despite the advantages of the 16s rRNA gene in the classification of bacteria, it is not perfect for some applications. For example, when sequence variation of the 16S rRNA between two microorganisms is very small, distance measurements of this gene may not be able to provide accurate information, such as the presence of similar species in the same genus [22]. The secondary structure of 16S rRNA (**Fig. 4**) may include valuable information [18], and has not yet been well explored in taxonomy studies. Also, in research about bacterial diversity and community structure using sequence profiles of the 16S rRNA genes, the copy numbers of this gene are regularly assumed to be consistent among different bacterial groups. These differences can affect the estimation of the abundance of different bacteria [23]. In addition, the classification of microorganisms based on 16S rRNA gene sequences relies on a well constructed database. However, current databases, including RDP, Silva, Greengenes and Genbank databases for 16S rRNA sequences, still have limits on the coverage of different taxonomic groups, especially at a deep level [24,25]. Thus multiple approaches, such as the whole-genome molecular techniques and the use of more target genes, combined with the 16S rRNA gene, could offer more accurate information in the classification and identification of microorganisms.

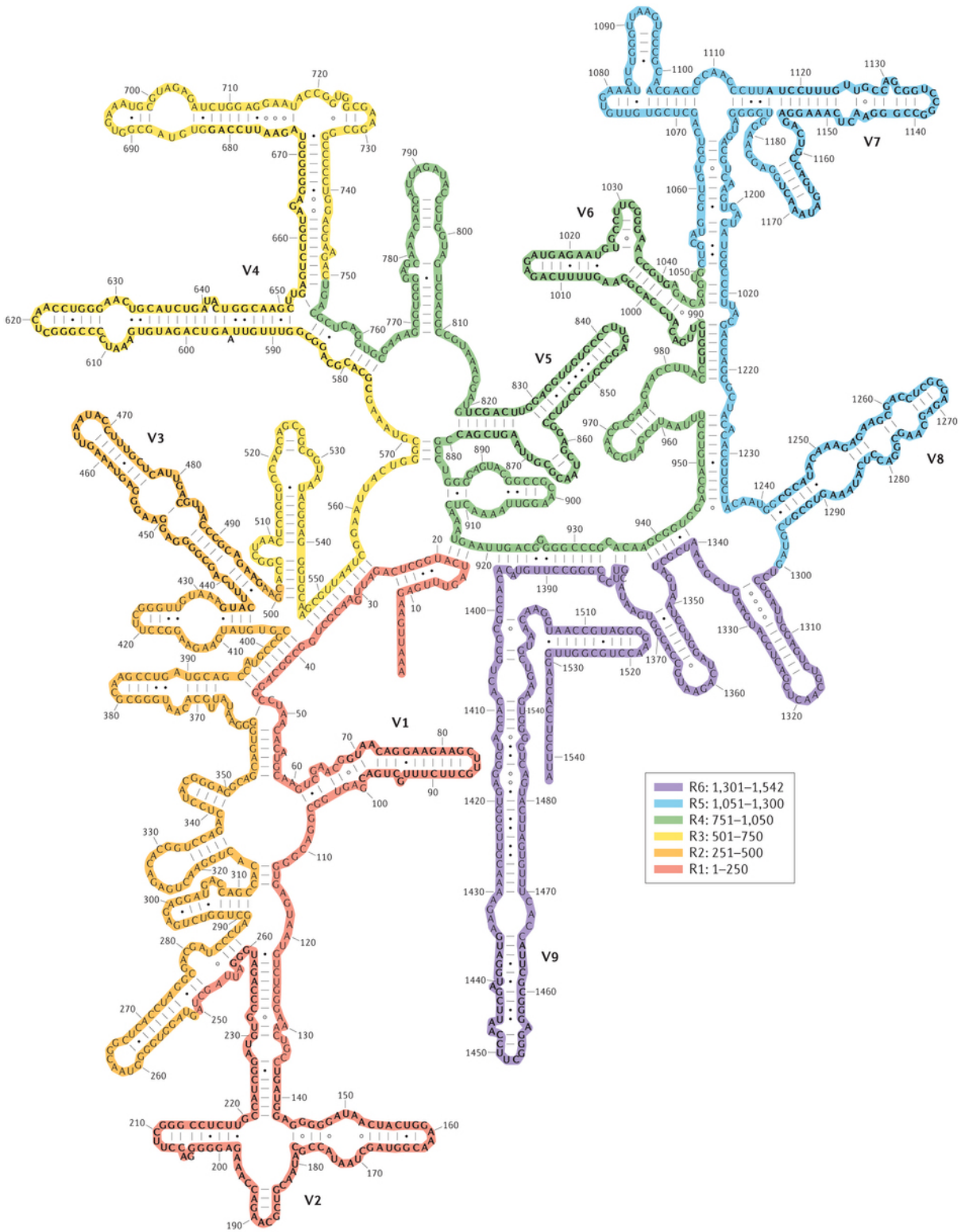


Figure 1.4. 16S rRNA secondary structure, adapted from Y. Pablo (2014)

1.1.3. Metagenomics and the Next Generation Sequencing techniques

Metagenomics is the study of the genetic material recovered directly from environmental samples [26]. With the development of next generation sequencing (NGS, also termed high-throughput sequencing; HTS) after 2004, metagenomics is helping to access the taxonomic and functional composition of microbial communities in any environmental biome, without the need to isolate or culture them in the laboratory [27]. While traditional microbiology relies on cultivated clonal cultures, the vast majority of microbial biodiversity had been missed [1]. Massively parallel sequencing techniques have revolutionized sequencing capabilities, far beyond the electrophoresis-based “first generation” sequencing, and launched the “next-generation” in genomic science. Metagenomics has already been successfully applied in many fields, including the analysis of the microbiome of natural water and soil environments, some extreme physical and chemical environments, food supply chains, animals, and human health. In recent years, to explore the taxonomic complexity, meta-barcode methods are broadly used, which is an amplicon-based approach, based on PCR-targeted sequencing of selected genetic species markers (such as some hypervariable regions of the 16S rRNA gene) [28].

Shotgun sequencing methods are also widely used in metagenomic samples, since shotgun metagenomics can provide information about both which organisms are present and what metabolic processes are possible in the community [29]. DNA sequences are randomly broken up into many small pieces and then reassembled by matching regions of overlap. Many organisms, which may be overlooked using traditional culturing techniques, may be retained as small sequence segments using shotgun sequencing. So to achieve the high coverage of different community members, especially under-represented ones, full sequencing of the genomes of large samples is often required.

Organisms and gene functions of the microbial communities in a specific environment can be obtained by shotgun sequencing of the whole metagenome. However, if we only need to analyse the taxonomic composition and biodiversity assessment of microbial community, an environmental DNA (e-DNA), meta-barcoding approaches can constitute an effective and less expensive solution [28]. It uses universal PCR primers, which are assumed to cover all the species belonging to the explored taxonomic range, and culture-independent sequencing of a selected genetic taxonomic marker (meta-barcode) from a mass collection of organisms or from e-DNA. The PCR products are then sequenced on a next generation sequencer, such as the Roche 454 and Illumina sequencers.

The 16S rRNA gene is commonly used to identify bacteria by meta-barcodes. In the case of recent versions of 454 technology, 16S rRNA gene sequences provide information of bacterial community biodiversity and relative taxa abundance down to the genus level [30].

The first metagenomic studies conducted high-throughput sequencing used massively parallel 454 pyrosequencing [31]. Three other technologies commonly applied to environmental sampling are the Ion Torrent Personal Genome Machine, the Illumina MiSeq or HiSeq and the Applied Biosystems SOLiD systems [32]. These techniques for sequencing DNA generate shorter fragments than traditional Sanger sequencing.

The Roche 454 Genome Sequencer

In 2008, 454 Sequencing launched the GS Titanium series reagents for use on the Genome Sequencer FLX instrument. This uses a large-scale parallel pyrosequencing system with the ability to sequence 400-600 million nt per run with 400-500 nt read lengths [33]. The system relies on fixing nebulized and adapter-ligated DNA fragments to small DNA-capture beads in a water-in-oil emulsion [34]. The DNA

fixed to these beads is then amplified by PCR. The technique is built on 4-enzyme real-time monitoring of DNA synthesis by bioluminescence, using a cascade that, upon nucleotide incorporation, ends in a detectable light signal (bioluminescence). The pyrosequencing chemistry is shown in **Figure 5**.

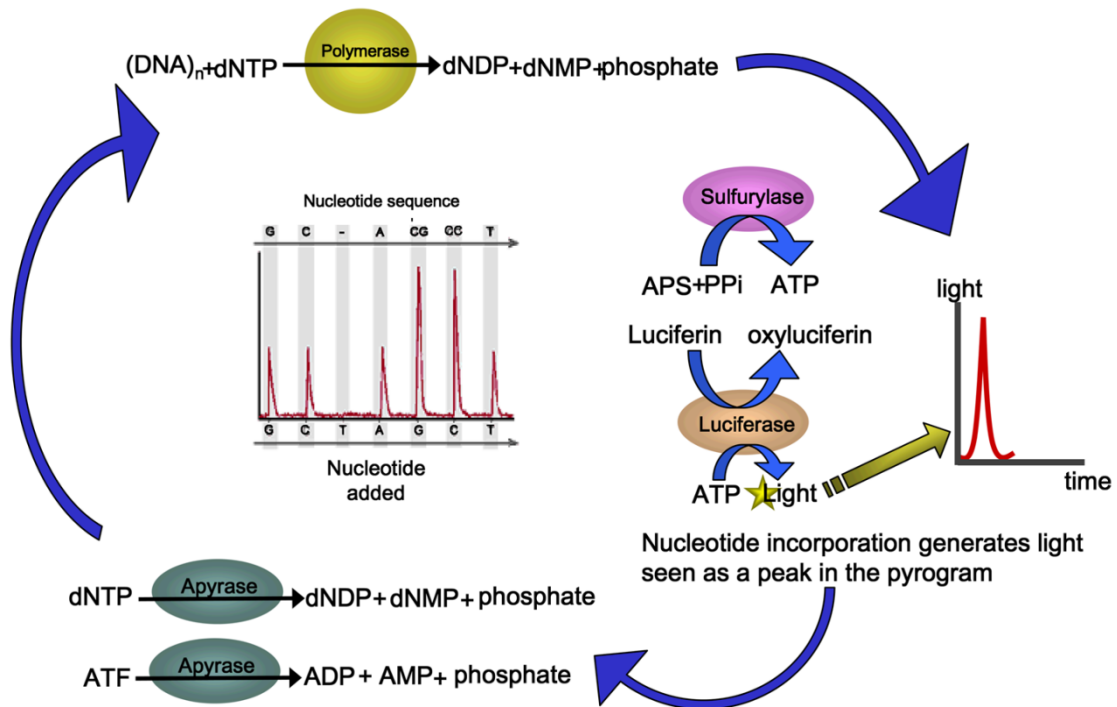


Figure 1.5. The 454 pyrosequencing chemistry (Petrosino et al 2009)

Genomic DNA fragments (such as 16S rRNA gene amplicons), are ligated to short adaptors, which provide priming sequences for both PCR amplification and sequencing of the sample-library fragments. Single-stranded template DNA (sstDNA) library is immobilized onto beads. The beads containing a library fragment carry a single sstDNA molecule. The bead-bound library is emulsified with the amplification reagents in a water-in-oil mixture. Each bead is captured within its own micro-reactor where PCR amplification occurs. Sequencing-by-synthesis then occurs by the DNA polymerase-driven generation of inorganic pyrophosphate, resulting in the formation of adenosine triphosphate (ATP) and ATP-dependent conversion of luciferin to oxyluciferin (**Fig. 6**). The generation of oxyluciferin results in the emission of photons of light, and the amplitude of each signal is directly related to the presence of

one or more nucleosides [35].

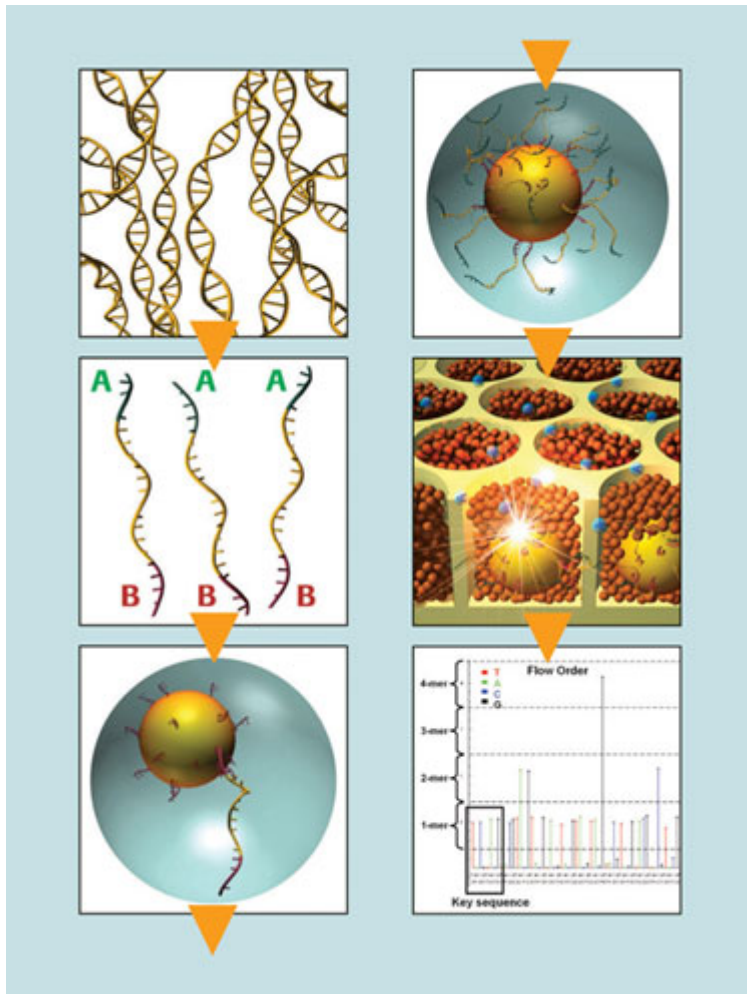


Figure 1.6. Process of Pyrosequencing (Margulies et al 2005).

The Illumina Genome Analyzer

Illumina sequencing also use sequencing-by-synthesis methods. It begins with the attachment of single strand template DNA (sstDNA) to primers on a slide, and the slide is flooded with nucleotides and DNA polymerase. Four different nucleotides (ATCG) are added, and these nucleotides are fluorescently labelled, with the colour corresponding to the base. They also have a terminator, so that only one base is added at a time. The four bases then compete for binding sites on the sstDNA to be sequenced and non-incorporated molecules are washed away [36]. After each

synthesis, a laser is used to excite the dyes and a photograph of the incorporated base is taken (**Figure 7**). Illumina sequencing only uses DNA polymerase, instead of multiple enzymes required by pyrosequencing. The current MiSeq platform can yield ~50 million 300 nt reads per run [37].

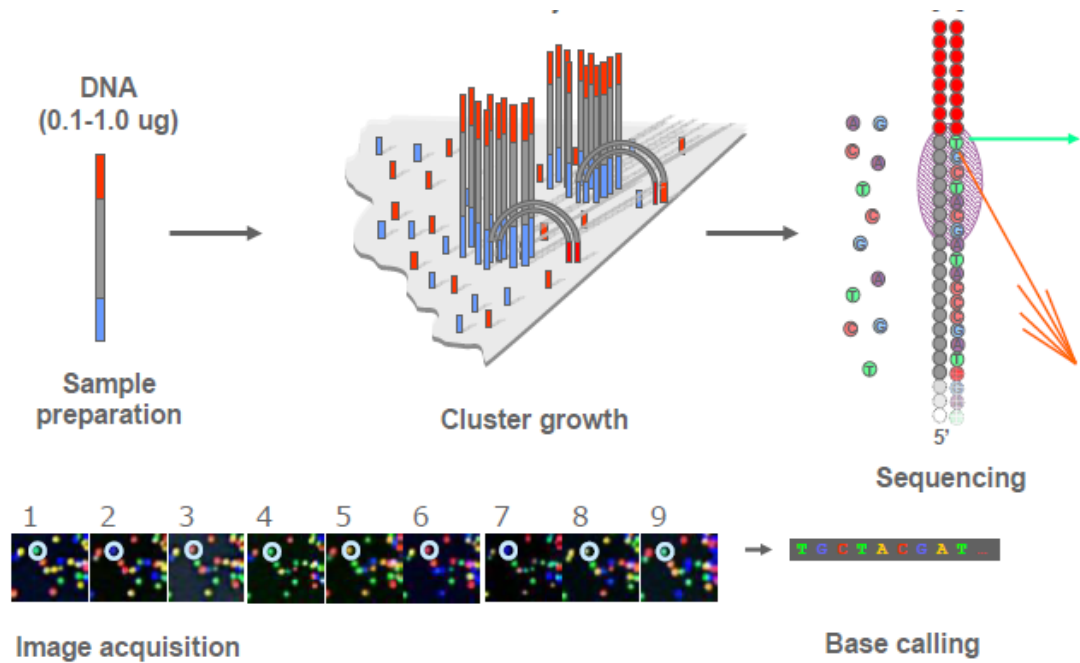


Figure 1.7. Process of Illumina Sequencing

Ion Torrent Personal Genome Machine

Unlike Illumina and 454, Ion torrent sequencing does not use optical signals. Instead, it exploits the fact that addition of a dNTP to a DNA polymer releases an H^+ ion [38]. DNA fragments of approximately 200 bp in length with adaptors are placed onto a bead. Then, DNA strands are amplified on the bead by emulsion PCR, with each bead in a single well of a slide. Like 454, the slide is flooded with a single type of dNTP, along with buffers and polymerase. The pH is detected in each of the wells, as each H^+ ion released will decrease the pH. So the sequences of the read are determined by detecting the changes in pH .

Iron Torrent has the same limitation as pyrosequencing, that it is difficult to enumerate

long repeats (homopolymers). The read length in achieved by Ion Torrent semiconductor sequencing is currently 400 nt, and the throughput is currently lower than that of other high-throughput sequencing technologies [39].

Among the three most widely used sequencing techniques, 454 pyrosequencing is broadly applied in meta-barcode analysis because of its longer sequence reads [40].

1.1.4. Bioinformatics for Meta-barcode Sequencing

The massively-parallel sequencing methods are capable of producing millions of reads, which presents a huge challenge for data storage, analyses and other manipulations [41]. Once sequencing is complete, raw sequence data must go through several analysis steps (Figure 8). A generalized data analysis pipeline for meta-barcode sequencing data includes: 1) preprocessing the data to remove adapter sequences and grouping to different libraries according to the specific barcodes, 2) checking the quality score of each base to remove low-quality reads, 3) mapping of the data to a reference database or stand-alone alignment of the sequence reads not relying on a reference database, and remove chimera sequences, 4) assign the sequences to their taxonomic classes, 5) cluster sequences to operational taxonomic unit (OTU) at a specific similarity level (normally 97%), and 6) statistical analysis of microbial community diversity within an environmental library (α -diversity) or between different libraries (β -diversity), based on OTU distribution or phylogenetic distances. Many free online tools and software packages exist to perform the bioinformatics necessary to analyze sequence data.

Data Analysis Process

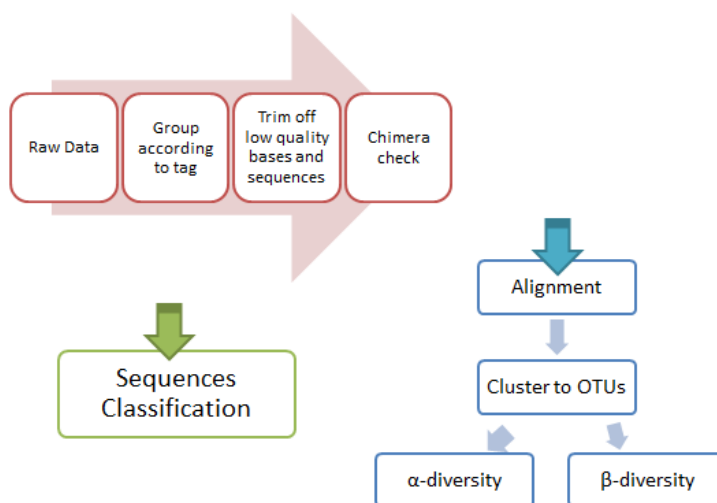


Figure 1.8. Analysis Process of Data from Meta-barcode Sequencing.

Cleaning of Raw Data

Raw data obtained from a sequencer needs to be grouped by barcodes (for multiple samples), and primers and adapters are removed from each read [42]. Then, the data are reformatted to fasta or fastQ format for further analysis. Several criteria are applied to remove sequence noise [43]: (i) noise length that extends the range of selected marker gene, (ii) more than one mismatch to the primer or barcode sequences, and (iii) the presence of homopolymers of > 8 bp in length. Then, the sequences are checked for quality score by trimming off the ends that normally contain low-quality bases, or filter off low-quality sequences (such as those that contain too many ambiguous bases, average quality scores are low, etc.). Regular pipelines available for these analysis procedures are Mothur [44], Qiime [45], and Galaxy [46].

Chimera sequences can be produced during the PCR process, originating from two or more parental sequences, that can be a large part of NGS sequencing errors. Several programs are used to detect chimeras from the data, using reference databases or de-novo methods, such as UChime [47], Decipher [48], and Chimera Slayer [49]. Since it is difficult to identify chimeras from data generated by massive sequence amplification, sometimes different methods can be combined.

Sequence Classification

Currently, the bioinformatics methods used to assign metagenomics sequences to their taxonomic classes adopt essentially three approaches, similarity-based methods, composition-based methods, and phylogenetic-based methods [28,50].

For the similarity-based methods, even short sequences (about 200 bp) can be classified, and the classification of sequences commonly relies on its comparison to curated collections of sequences in a reference database. So the absence of reference sequences for unknown species, especially for environmental samples, can affect the

assignment accuracy. Currently, the most widely used reference resources available for 16S rRNA gene-based identification of microorganisms are the Greengenes [51], Ribosomal Database Project II (RDP II) [52], and SILVA databases [53].

The composition-based methods use a feature space consisting of 6-8 base subsequences (words) to assign the metagenomic sequences to taxonomically annotated reference sequences [54]. The reference data are pre-treated using different methods, such as the naïve Bayesian classifier and the k-nearest-neighbor algorithms. This approach requires more computational abilities than the similarity-based methods.

The phylogenetic-based methods calculates the phylogenetic distances according to a reference evolutionary phylogenetic tree, and assigns the sequences using maximum likelihood, neighbor-joining algorithms, etc. This method requires huge computational demands, that are not available to all researchers. To achieve both accuracy and efficiency in the classification of meta-barcode sequences, some programs combine different algorithms [55].

Statistical Analysis

To study the biodiversity of microbial communities, Operational Taxonomical Units (OTU) are commonly used, since it is difficult to calculate on their real species numbers in different communities [56,57]. For meta-barcode sequencing data, not all sequences can be classified at deep taxonomic levels, which makes it more efficient to use OTU methods to reveal the structure of microbial communities [58]. OTUs are clustered based on sequence similarity, and the most accepted similarity is 97%, representing the potential distances among different species, although 95% or 99% are also used [59].

Among different methods adapted to group sequences into different OTU clusters,

two algorithms are broadly used, picking OTUs based on alignment to reference database (ie. Mothur and RDP II pipelines), and picking OTUs based on the calculation of sequence distances among each other, such as the Uparse) [60,61]. After OTU tables are generated, further statistical analyses can be performed.

Two biodiversity indices are broadly used to estimate the richness and evenness of a microbial community, Chao 1 and Shannon, based on the distribution of OTU populations in a community [62,63].

Chao 1 Estimator calculates the estimated true species diversity of a sample by the equation (The calculations for the bias-corrected Chao1 richness estimator in the program EstimateS).

$$S_{chao1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

where, \leftarrow

S_{chao1} = the estimated richness \leftarrow

S_{obs} = the observed number of species \leftarrow

n_1 = the number of OTUs with only one sequence \leftarrow

n_2 = the number of OTUs with only two sequences \leftarrow

The Shannon diversity index (H) is another index that is commonly used to characterize species diversity in a community. Shannon's index accounts for both abundance and evenness of the species present.

$$H_{shannon} = - \sum_{i=1}^{S_{obs}} \frac{n_i}{N} \ln \frac{n_i}{N}$$

where,

S_{obs} = the number of observed OTUs

n_i = the number of individuals in OTU i

N = the total number of individuals in the community

Other statistical methods, such as rarefaction curves and the Simpson Index, are also used.

Comparisons Between Communities

To investigate the differences between two microbial communities or among multiple libraries based on their composition, distance matrices are calculated using different algorithms, such as Bray-Curtis [64], Jaccard [65], Sorenson [66], weighted and un-weighted Unifrac [67]. OTU-based or phylogenetic-based distance matrices between every pair of community samples are presented in a square matrix. A comparison of different methods is shown in Table 1. Then, the matrix can be plotted using Principal Coordinate Analysis (PCoA) [68], UPMGA-tree (Unweighted Pair Group Method with Arithmetic Mean) or other bioinformatic methods [69]. Pearson correlation coefficients, to explore the relationship between microbial community structure and environmental factors, or between different taxa is also used [70].

Table 1.1. Comparison of different methods to calculate distances between microbial communities

Dissimilarity Measures	Species numbers	Phylogenetic
Incidence (presence or absence)	Sorensen	Un-weighted UniFrac
abundance	Bray-Curtis	Weighted UniFrac

List of Data Analysis Platforms and Pipelines for Meta-barcode Sequencing

1. Ribosomal Database Project (RDP II)
<http://rdp.cme.msu.edu/>
2. Mothur
http://www.mothur.org/wiki/Main_Page
3. Qiime
<http://qiime.org/>
4. Usearch
<http://www.drive5.com/usearch/>
5. SILVA rRNA database project
<http://www.arb-silva.de/>
6. Galaxy
<https://galaxyproject.org/>
7. MGRAST
<http://metagenomics.anl.gov/>
8. Prinseq
<http://prinseq.sourceforge.net/>
9. Greengenes
<http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>
10. ESPRIT
<http://www.ijbcb.org/ESPRITPIPE/php/onlinetool.php>
11. CAMERA
<http://metagenomics.anl.gov/>
12. R
<http://www.r-project.org/>

1.2. Arid Environments

1.2.1. Introduction to Arid Environments

Hot deserts are regions of land that have little rainfall, where few plants and animals generally exist. Aridity is the dominant climatic factor over about one-third of the land surface of the world, as approximately 7.5% of global land area is classified as extremely arid (hyperarid), 12.5% arid, and about 17.7% semiarid. If dry-subhumid areas (9.9%) are included in the classification, then drylands comprise about 47% of the Earth's land surface (United Nations Environment Programme, UNEP 1992). In total, about 49 million km² are affected by aridity. Of the total surface area of arid climate, Africa is 36.7%, Asia 31.7%, North America 12%, Australia 10.8%, and South America 8.8% [71,72]. To sum up, the dry areas of the world occupy more land than any other major climatic type [73].

Deserts are considered to be the hyperarid and arid regions, and semiarid and dry-subhumid regions the desert fringes. It is difficult to derive an exact definition of a desert, as aspects of climate (precipitation, evaporation, and temperature), geomorphic features, and flora and fauna, show considerable variation. Although they may share common features such as wide temperature variation, winds, geomorphology, shifting sands, and plant and animal life, they are not components of all arid environments. There are many aridity indexes to measure the level of aridity for a region, an assessment of the extent of drylands broadly applied is aridity index (AI) conducted by Hulme and Marsh (1990) on behalf of UNEP (1992), where

$$AI = P / PET,$$

P stands for the mean precipitation of a fixed time period, PET is potential evapotranspiration

An examination of the distribution of deserts based on the AI is shown in Fig 9. The delimitation of the different types of dryland environments by AI values are dry-subhumid (AI = 0.50 - < 0.65), semi-arid (AI = 0.20 - < 0.50), arid (AI = 0.05 - < 0.20), and hyper-arid (AI = < 0.05).

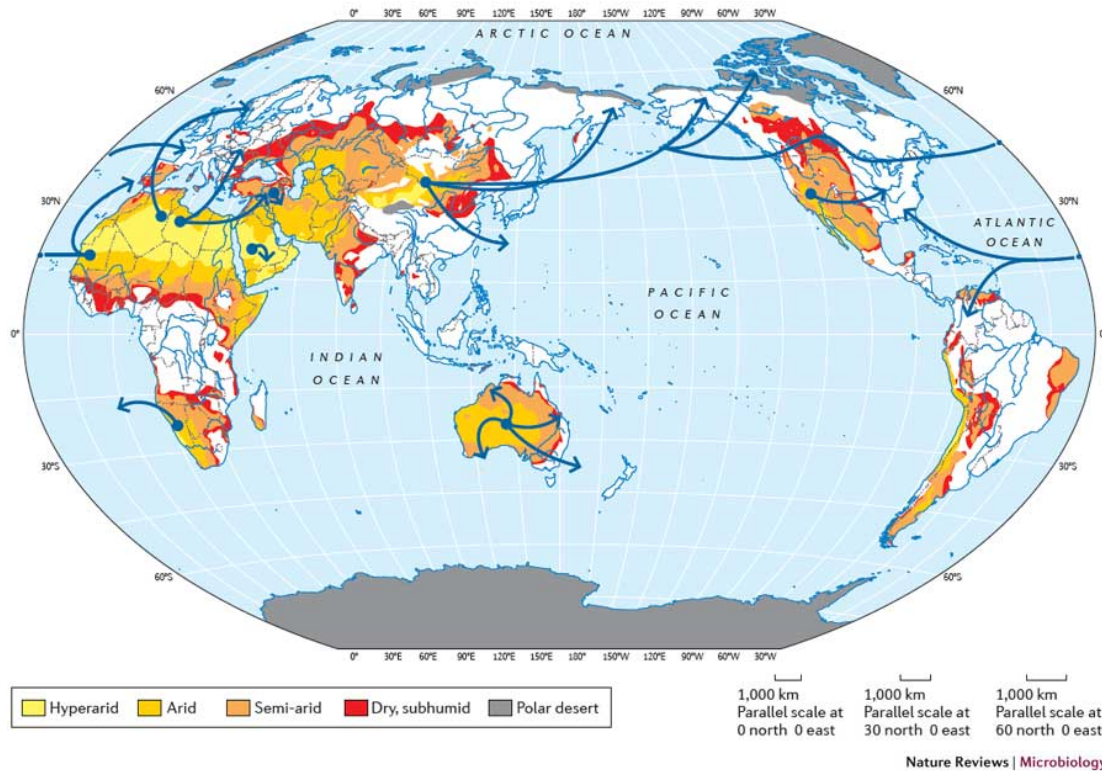


Figure 1.9. Map of Global Distribution of Dry Land. Regions with declined color follows the aridity index (AI) of the UNEP. Arrows show the major intercontinental trajectories for desert dust (Laitty 2009).

The classification of deserts relies on combinations of the total amount of annual rainfall, temperature, humidity, or other factors. In general, deserts are classified into 3 categories (**Fig 10**): Hot and dry deserts (such as South and Central America, Africa and Australia), Cold deserts (such as Antarctic, Greenland), Coastal deserts (mostly found on the western edges of a continent, and a famous one is the Atacama Desert in Chile). Table 2 shows a summary of studies of hot deserts [74].



A. Hot Desert B. Cold Desert C. Coastal Desert

Figure 1.10. Classification of deserts. The pictures are from three different types of deserts: A. the Arabian Desert; B. the Gobi Desert; C. the Namibi Desert

Table 1.2. Summary of studies of hot deserts (Makhalanyane et al 2015)

Classification	Country	Name	Approx. size (km ²)	Approx. temperature range (°C)	Approx. precipitation (mm/yr)	Average soil pH	Average soil organic carbon (%)	References
Hyperarid	Chile	Atacama-Sechura	105 000	-5~40	0~20	6.6~9.2	0.1~2.6	[75-77]
Hyperarid-arid	Southwestern Africa	Namib	81 000	5~45	5~100	7.9~8.5	0.1~0.3	[78,79]
	Northern Africa	Sahara	9 100 000	-5~45	5~150	7.6~7.9	0.1~1.2	[80-82]
Arid	Southern Mongolia	Gobi	53 000	-20~30	30-100	7.7~10.2	0.1-2.64	[83]
Arid-semiarid	Arabian Peninsula	Arabian	2 300 000	5~40	25~230	7~7.5	NA	[84]
	South Africa	Karoo	395 000	2~40	50~200	6.9~9	0.3~1.3	[85]
	Southwestern USA	Mojave	152 000	-10~50	30~300	7.1~9.4	0.04~0.1	[86]
	Central Australia	Simpson	180 000	5~40	50~400	6.5~7	0.1~0.3	[87,88]
	North Mexico							
	/Southwestern USA	Chihuahuan	455 000	10~40	70~400	5.9~6.2	0.2~1.9	[89]
	Southwestern USA	Sonoran	312 000	-10~50	70~400	5~8.6	0.4~2	[90]
	Southwestern Africa	Kalahari	520 000	-10~45	100~250	7.7~8.7	0.1~0.5	[91-93]
	Israel	Negev	13 000	5~40	100~300	7.2~8	0.5~0.7	[94]
	India/Pakistan	Thar	200 000	4~50	200~300	7.9~8.1	0.3~0.4	[95]
Semi-arid	Southern Australia	Gibson	156 000	6~40	200~400	NA	0.06	[96]
	Northwestern Australia	Great Sandy	285 000	10~40	250~370	5.8~6	0.1~1.1	[96]
	Northern Australia	Tanami	185 000	10~40	300~500	4.9~6.7	0.1~1.4	[97]

1.2.2. Characteristics of Drylands and Desert Soils

Climate. The world's deserts can be divided into three categories (Meigs 1953): hot deserts in tropical and subtropical latitudes, temperate deserts in higher latitudes in continental interiors, and coastal deserts found on the west coasts of continents in tropical latitudes. Hot inland deserts are identified by large temperature variations, and the persistence of high daily temperatures, with maximum temperatures commonly between 45 and 49°C [73]. Temperatures on the soil surface can be considerably higher, as much as 75-80°C [98]. Temperate deserts are characterized by considerable seasonal variations in temperature, and a dependable period of cold temperatures. These deserts are at higher latitudes than hot deserts, some precipitation occurs as snow, and soil moisture is often frozen. They have hot summers counterbalanced by relatively cold winters. In the arid regions of Antarctica, mean winter temperatures may be as low as -30°C [99]. Coastal deserts tend to have relatively low seasonal and diurnal ranges of temperature.

Precipitation. Deserts generally receive relatively low amounts of total annual precipitation. Spatial variation in rainfall is high, which leads to the high biodiversity in some desert areas [100]. Rainfall in deserts tends to fall in pulses [101], which can vary considerably in magnitude and timing. For some deserts at coastal areas, where rainfall is very low, fog becomes the major resource of precipitation [102]. The high level of fog in coastal desert fog zones provide habitats extremely favorable for lichen growth [103]. Mountain snow plays an important role in water storage and release in deserts that are in proximity to high-altitude mountain ranges. High runoff from desert slopes may offer sufficient water availability to agriculture and helps dampen the variability of the hydrograph [104].

Landscapes. There are five major types of desert landscapes that are commonly recognized: sand deserts, stone deserts, rock deserts, plateau deserts, and mountain

deserts (**Fig 11**). The sand desert landscape probably accounts for 15-20% of desert land [104], and is thus not as common as often perceived. Stone deserts usually have a gravel surface, which is covered by rocks too large to be carried away by wind or water, and also known as desert pavement. Rock desert landscapes normally have bare rock surfaces. Plateau landscapes are often found in a desert landform of mountain-and-basin deserts. Mountain deserts constitute a landscape form called shield desert, where wind is a more effective force than water, compared with mountain-and basin desert.



Figure 1.11. The major types of desert landscapes: A. sand desert, B. stone desert, C. rock desert, D. plateau deserts, E. mountain deserts

Saline Soils. Soils from desert environments are dominated by a mineral component with low organic matter, but the repeated accumulation of water in certain soils can cause salts to precipitate [105]. When the water table rises to within about 2 m of the ground level, water may begin to rise to the surface by capillary action. Then, dissolved salts will be carried up to the surface, and concentrate in the upper layers of the soil as water is evaporated [106]. Salinity changes the electrochemical balance of soil particles, which is harmful to plant cells, and can also increase soil erosion.

Radiation. The desert atmosphere is relatively clear. Clouds are rare, and the water

vapor content is low. The paucity of clouds has several important consequences. Incoming solar energy approaches a maximum in arid regions owing to the lack of cloud cover. Approximately 80% of solar radiation at the top of the atmosphere reaches the surface [73]. The distribution of radiation is characterized as seasonal, normally reaching a peak in June in the Northern Hemisphere and December in the Southern Hemisphere [107].

Desert Dust. Mineral dust is the most important export from the world's arid zones to the global Earth system, and affects atmospheric, oceanic, biological, terrestrial and human processes and systems [108]. It is responsible for suppression of rainfall [109], long-distance microorganisms transport, risks to human health, and agricultural soil erosion and productivity. Dust emission from arid environments in China represents as much as half of the global atmospheric loading of dust, while North America has only one very small zone located in the Great Basin with high values of dust loadings into the atmosphere [110]. The increasing greenhouse gas emissions have brought large changes to climate, and many parameters that control dust emission, such as vegetation, rainfall, soil moisture and surface wind speed are expected to change [111].

1.2.3. Drylands in North America

There are a diverse number of small deserts stretching from southeastern California to western Texas, and from Nevada and Utah to the Mexican states of Sonora, Chihuahua, and Coahuila and much of the peninsula of Baja California (**Figure 12**). Beyond these areas, semi-arid conditions extend north to eastern Washington, south to the central Mexican plateau, and east to link with the steppes of the High Plains. About 55% of the North American deserts are considered semiarid, 40% arid, and only 5% hyperarid [73].



Figure 1.12. Map of North American deserts (Laity 2009)

Many factors can work to create deserts. The rain shadow effect contributes to the form of North American deserts, which is caused by mountain ranges blocking moisture of Pacific origin in the winter and the Gulf of Mexico in summer. Rain shadow effects are major in the Great Basin and Mojave Deserts, while the effects of high pressure are important in the Chihuahuan and Sonoran Deserts. Increasing precipitation of some regions between July and mid-September over large areas of the southwestern USA and northwestern Mexico is called the North American monsoon [112], which is defined as sites that receive at least 50% of their annual precipitation during the summer period. North American monsoon systems develop in response to atmospheric moisture supplied by nearby warm oceans. The boundary between cold and warm deserts lies across southern Nevada and Utah [73].

The Great Basin, characterized by its high altitude at its northern position, is considered a “cold” desert. Approximately 60% of its precipitation in winter comes in the form of snow, and mean monthly temperatures is below 0°C from December to February. The mean annual temperature is 9°C [113]. Most regions have an elevation > 1200 m, with mountains up to 3000m in height. Mean precipitation varies from 2 to 300 mm among different sites and is evenly distributed throughout the year. Much of Nevada and Utah are in the Great Basin Desert [114].

The Colorado Plateau is a roughly circular area > 300,000 km² passing at higher elevation, that consists of plateaus and isolated mountains of Utah, Colorado, New Mexico, and Arizona [115]. Although the plateau lies mostly above 1500 m, it shows high internal variation [116]. Differential erosion characterizes a landscape dominated by canyons, cuesta scarps, and plains. Dune fields on the Colorado Plateau are not extensive, but have a variety of forms. Wind erosion features of the Colorado Plateau comprise deflation hollows, yardangs, wind-fluted cliffs, and blowouts [73].

The Mojave Desert is the smallest North American desert, and occupies

approximately 140,000 km² in southeastern California and southern Nevada, with elevations mostly above 1000 m [117]. The late Quaternary climate had a large impact on the surficial stratigraphy of the area [118]. It is roughly rectangular in shape, bounded by the Great Basin Desert to the north, and the Sonoran Desert to the south. It has an annual rainfall from ranging 76 to 102 mm across the desert floor, and reaching about 279 mm with increasing elevation. The Mojave Desert is characterized by numerous mountain ranges, valleys, endorheic basins, salt pans, and seasonal saline lakes. Most of the valleys are internally drained, such that all precipitation that falls within the valley does not eventually flow to the ocean [119].

The Sonoran Desert covers approximately 275,000 km² in large parts of the Southwestern United States in Arizona and California, and of Northwestern Mexico in Sonora, with elevations ranging from below sea level in California to about 1500 m in mountain foothills [120]. The Sonoran Desert is more subtropical than other North American deserts, with summer high temperatures reaching 49°C or more [121].

The Chihuahuan Desert occupies approximately 518,000 km² in southwestern North America, with major parts in northern Mexico, and one fourth ranging in western Texas and southern New Mexico. Most sites are located at elevations from 600 to 1,500 m. This desert has more rainfall than other warm desert ecoregions, with precipitation ranging from 150 to 400 mm [122]. The Chihuahuan Desert encompasses one of the most biologically diverse arid regions on Earth [123].

1.2.4. Analogue Sites for Mars

Some arid landscapes contain a range of Mars analogue features, relevant for geology and astrobiology studies. These features include wind erosion, moisture deficits, absence of vegetation, high UV radiation, etc., although surface temperature, atmospheric pressures, gravity and physiochemical composition are very different from that of Mars. The EuroGeoMars 2009 campaign was organized at the Mars Desert Research Station (MDRS) to perform multidisciplinary astrobiology research. MDRS in southeast Utah is situated in a cold arid region with mineralogy and erosion processes comparable to those on Mars [124].

Particular deserts reveal extreme Mars-like surface characteristics, such as the cold Antarctic desert McMurdo Dry Valley which is considered to have the coldest, driest and most oligotrophic soils [125]. Mars is the third largest planet in the solar system. It is often referred to as the "Red Planet" because of the iron oxide prevalent on its surface, which is similar to deserts in Utah, characterized by red-colored hills, soils and sandstones. MDRS is located in a cold arid desert with an temperature of 35°C at midday on the equator and -43 °C during the polar winters [124]. It consists of minerals containing silicon and oxygen, metals, and other elements that typically make up rock. Our ability to study the surface of other planets in the solar system is very limited, and studies on the terrestrial analogues located on Earth will help to build the foundation of terrestrial desert exploration.

1.2.5. Desertification

Desertification - the spread or intensification of land degradation towards greater aridity due to climatic changes or human activities - is occurring at an alarming rate around the world [126]. Dryland degradation results in huge economic losses and directly affects more than 1 billion people who depend on such areas for their livelihood, particularly small farmers (United Nations Environment Programme, UNEP 2012). UNEP estimates that 69% of agricultural drylands in the world are degraded or undergoing desertification.

Arid zones are the most vulnerable areas, and characterized by extreme drought. But drought alone cannot be responsible for desertification. Emanuel et al. (1985) predicted a dramatic increase in global desert lands due to climate changes with a doubling of the atmospheric CO₂ concentrations [127]. Soil salinization, agricultural development in marginal desert lands and housing developments can negatively affect arid environments. Soil salinization reduces soil quality, limits the growth of crops, constrains agricultural productivity, and in severe cases, leads to the abandonment of agricultural soils [128]. High soil salinity occurs naturally in deserts, but poor water management in irrigated areas raises the natural salinity of the soil to the soil surface [129].

Aquifer pumping in desert golf courses reduces the groundwater and increases soil salinity, as well as mineralization and chemical pollution of watercourses [130]. For example, the use of the water from the Colorado River for urban purposes in southern California has resulted in the river no longer reaching the sea. In North America, the replacement of grasslands by woody species with shrubs are particularly negative effects of desertification, making landscapes vulnerable to wind and water erosion [131], and soil erosion results in the loss of biodiversity. Drier conditions linked to increased demand for ground water pumped for agricultural irrigation, particularly in

the central and western US [132], results in the depletion of aquifers [133,134]. Thus, human use can create desert-like conditions in lands that were previously far more productive.

Desertification also contributes to other environmental and social crises, such as the mass migration of people and animals, species loss, climate change and the need for emergency assistance to human populations. It affects both developed and underdeveloped nations, up to 66% of the African continent is threatened by aridity, and almost 40% of land in the continental United States is vulnerable to desertification, estimated by the US Bureau of Land Management [135].

As humans make increasing use of dryland resources, hazards associated with aeolian and fluvial processes will be more intense [136]. It is clear that agricultural activities and water resource use in drylands may result in the potential acceleration of erosion.

1.3. Desert Biodiversity

1.3.1. Plants and Animals

Although it has been often suggested that deserts are relatively simple ecosystems characterized by low biodiversity, some research suggests that deserts are relatively complex and biologically rich [106]. Deserts supports various fauna and flora, including terrestrial plants and animals.

Plants have developed various morphological and physiological adaptations to live in the desert environment, classified as drought-escaping, drought-evading, or drought-resisting. There are some principal adaptations: 1) geophytes and other plants have special storage organs, 2) trees and shrubs with deep root systems are able to exploit deep aquifers in dry environments, 3) germinating immediately after the infrequent rains and completing their life cycles before summer heat [106], 4) plants have rapid gas exchange and small leaf surfaces to minimize heat input.

Local geographic factors, such as the mineral composition, nutrient reserve, organic content, and capacity to hold water, can affect the distribution and abundance of plants [73]. Plants can be classified as xerophytes, mesophytes, or phreatophytes according to their water requirements. Xerophytes plants, that can survive and reproduce when water is limited, dominate desert environments. Grasses are the most abundant species of plants in deserts [137]. There are several common desert plants found in desert environments, with different characteristics across the world owing to their different climatic conditions, including barrel cactus, brittle bush, palm trees, jumping cholla, saguaro cactus, etc. Vegetation in deserts can have large effects in different geological processes, including reducing sand transport and changing hydrology and dryland river forms.

Animals have evolved various strategies in order to cope with the extreme conditions in deserts, which can be classified into three categories: behavioral, morphological, and physiological. Morphological adaptations to heat include smaller body sizes and relatively larger surface areas, and light-colored surfaces to reflect radiation. Physiological adaptations are less common. They include dormancy during summer, uric acid as a major nitrogenous waste, the deposition of fat in tails or humps, salt glands that secrete salt without the loss of fluids, and an absence of sweat glands [106].

The Chihuahuan desert is one of the most biologically rich and diverse desert ecoregions in the world, others include the Great Sandy Tanmi Desert of Australia and the Namib-Karoo desert of southern Africa [114]. The high degree of local endemism is the result of the isolating effects of complex basin and range physiography, and dynamic changes in climate over the last 10,000 years [138]. Local species diversity is related to rainfall, more rich in semi-arid zones, while rockiness enhances species diversity because of the presence of many micro-niches.

1.3.2. Desert Microbiology

Arid lands account for the largest terrestrial biome [126], and stresses such as drought, temperature and radiation limit the scale of life extension. Research concerning microbial colonization and dispersion in deserts has been performed to estimate the function of microbial communities from desert sand which may play an important role in soil stability, nutrient cycles and environmental health.

Previous studies have predicted that there may be as many as 10^7 to 10^9 unique bacterial species on Earth [139], but with additional sequencing efforts, the species richness may increase. Species richness estimates are significantly higher in non-polar as compared with polar deserts [140]. Hot deserts supported significantly higher abundances of heterotrophic bacteria relative to photoautotrophic bacteria in cold deserts, and implied that productivity is higher in hot deserts and therefore capable of supporting greater biomass and trophic complexity than in cold deserts [141].

In many deserts, small poikilohydric life forms constitute a thin veneer on or within the top few centimeters of most soil and rock surface, which typically contain cyanobacteria, chlorophytes, fungi, heterotrophic bacteria, lichens and mosses [126]. Soil and rock surface communities are widespread and share general similarities in all the hot and cold deserts that have been examined worldwide. *Cyanobacteria* are major N_2 fixation hot spots in arid lands that support an array of heterotrophic microorganisms, and normally dominant the soil and rock surface community in arid environments. A study in the Atacama desert showed that the non-cyanobacteria phototrophic bacteria *Chloroflexi* are dominant in the hyper-arid core of the desert [76].

Biological soil crusts (BSCs) are specialized communities comprised of mosses, lichens, liverworts, cyanobacteria, and other organisms in the topsoil of terrestrial

environments [142,143]. They play important ecological roles in vegetation and ecological restoration in desert regions through aggregation of soil particles that reduce wind and water erosion, and different crust developmental and successional stages have different ecological functions [144]. BSCs increase infiltration and nitrogen fixation and contribute to local soil organic matter. They may constitute as much as 70% of the cover of biological organisms in a particular community [145]. These organisms are capable of withstanding desiccation, and often equilibrate their activities with soil moisture content [146]. BSCs are very sensitive to destruction by human activities such as grazing, agriculture, construction and outdoor activities [147].

Epiliths are observed in different types of arid environments. Lichens and mosses commonly occur on rock surfaces. Biogenic varnishes are well presented, that are associated with microorganisms, covering *Actinobacteria* and *Proteobacteria*.

Hypoliths are photosynthetic organisms that live underneath rocks in arid environments. Hypolithic colonization is protected by overlying mineral substrate from incident UV radiation and excessive photosynthetically active radiation (PAR) [148-150]. It has been reported that cyanobacterial hypoliths occurred on quartz in major deserts spanning every continent on Earth [151]. The *Actinobacteria*, *Alphaproteobacteria* and *Gammaproteobacteria* are ubiquitous in all hypoliths, and photoautotrophic cyanobacteria are more abundant in hypoliths of cold deserts [141,150,152]. *Deinococci* appear to be more abundant in warm and hot deserts but not common in polar hypoliths [76]. Few studies on multi-domain diversity of hypolithic microorganisms in deserts have been performed [153]. A study using quantitative PCR estimated the entire hypolithic communities in Tibetan and Antarctic Dry Valleys, and revealed that eukaryotic and archaeal taxa comprise less than 5% of recoverable phylotypes.

Endoliths are organisms (archaea, bacteria, fungi, lichen, algae) that live inside rocks, or in the pores between mineral grains of a rock. Rocks such as sandstone, limestone and weathered granite are normal habitats for endoliths in deserts. Endolithic colonization is widely observed among deserts of all aridity classes, indicating the advantage of adaptation to drought conditions.

Bioaerosols are suspensions of airborne particles that contain living organisms. They can disperse with the transportation of desert dust over intercontinental distances. Airborne microorganisms, such as spore-forming bacteria and mitosporic fungi in bioaerosols, pose large risks to human and animal health, with long distance traveling and potential pathogenic abilities.

Microbial communities and their functional structure in deserts are not well studied, and there are many unanswered questions regarding their biology, physiology and ecology. With the development of high-throughput technology, more work can be done using various omics methods such as genomics, proteomics or metabolomics, to fully understand the microorganisms in deserts.

1.3.3. Bacterial Diversity

Desert biomes have been shown to be remarkably different from other biomes in terms of soil microbial community composition and function [154]. Traditional culture-dependent methods for identification of bacterial diversity can only reveal a very small fraction of the actual bacteria in a community, and the percentage varies from 0.0001% to 15% in different environments [59]. Modern environmental microbiology has been greatly enhanced by the application of molecular genetic technology, which allows the examination of microbial communities through analyses of microbial DNA, RNA and proteins. Many studies have been performed on the bacterial community of soils in arid environments using next generation high-throughput sequencing methods.

The known bacterial diversity on Earth includes approximately 12000 different species (<http://www.bacterio.net/-number.html>). The unknown bacterial diversity is currently explored using molecular microbiology techniques, and many new bacterial taxa are being submitted to GenBank each year. Early research in the Atacama Desert based on cell numbers reported that cultivable heterotrophic bacteria are present in the less arid region of the Atacama Desert at levels of 10^7 colony-forming units (CFU) per gram of soil, while only present between 10^2 to 10^4 CFU/g of soil in the desert's core regions [155]. Populations of aerobic bacteria in deserts across the world are reported to vary from < 10 /g in the Atacama desert to 1.6×10^7 /g in soils of Nevada. Sand dunes from the Thar are reported to have a relatively smaller population ($1.5 \times 10^2 - 5 \times 10^4$ /g soil) [156]. Gram-positive spore formers are dominant and the populations do not decline significantly even during summer, and *Actinomycetes* may constitute ~50% of the total microbial bacterial population in desert soils [157].

Using PCR-based biological molecular techniques, microbial research can focus more precisely on the diversity of bacterial communities and their dominant microbes.

Pointing et al. (2008) reported that in the Namib Desert, the majority of 16S rDNA sequences displayed more than 94% homology to members of the *Firmicutes* (particularly to members of the genus *Bacillus*), and bacteria belonging to the *Bacteroidetes*, *Planctomycetes*, *Chloroflexi*, and *Betaproteobacteria* groups were also observed [158]. Connon et al. (2007) found that in the soil of Antarctica, dominant phylotypes were affiliated to the phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Cyanobacteria* and *TM7* [159]. An et al. (2013) studied the bacterial communities in samples from the two largest deserts in Asia, the Taklamaken and Gobi deserts, and found the most dominant phyla are *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* [160].

Photosynthetic cyanobacteria can be the primary inhabitants in arid environments, and typically live a few millimeters below the surface of translucent rocks, such as quartz, sandstone pebbles, halite and gypsum [161,162]. The capacity to benefit from a sufficient supply of CO₂, N₂ and light to allow photosynthesis and N₂ fixation while being protected from desert-like conditions (high radiation, desiccation, salt stress, etc.), allow these phototrophic communities to be prevalent in deserts. The distribution of cyanobacterial communities in desert pavements present more frequently in the form of patches, and their spatial distribution pattern in different sites are correlated with mean annual precipitation and temperatures [161]. Heterotrophic bacteria also occur widely in desert environments, including *Alphaproteobacteria*, *Actinobacteria*, *Flexibacteria*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes* and *Deinococcus-Thermus* [163]. Members of the CFB group (Cytophaga-Flavobacterium-Bacteroides) were found to be dominant in the hot desert of Tataouine of south Tunisia [164], and in the hyper-arid Taklamakan Desert in China.

Many studies show that members of the *Actinobacteria* can be dominant in arid environments, and the subclass *Actinobacteridae* are often found to prevail in desert

soil [165]. Members of the genera *Rubrobacter*, *Arthrobacter*, *Thermopolyspora* and *Streptomyces* have been found in both hot and cold deserts [166]. Members of the *Actinobacteria* have wide metabolic and sporulation capacities, as well as multiple UV repair abilities [167].

Bacterioidetes are also well represented in desert soils. Prestel et al. (2013) reported that samples from Death Valley soils showed a number of phylotypes with high homology to members of the *Flavobacteriales* and to the genus *Adhaeribacter* of the class *Sphingobacteria* [168]. In the Taklamakan Desert, an abundance of *Pontibacter* from the family *Cytophagaceae* were observed [169].

Proteobacteria are globally distributed and were thought to be prominent members of desert soil bacterial communities [170]. *Alpha-*, *Beta-* and *Gammaproteobacteria* are often linked to soils with higher rates of organic carbon inputs (Lopez et al. 2013). However, several studies have shown that *Proteobacteria* may be functionally important in nutrient-limited arid environments, since some members in this phylum are capable of photosynthesis [171].

Members of the phylum *Gemmatimonadetes* and *Firmicutes* are also widely observed in desert soils, and may be comparatively more abundant than in other biomes [154]. Some genera in the *Firmicutes* phylum, such as *Bacillus* and *Paenibacillus*, are able to form endospores that can facilitate survival under desiccation conditions. Some aerobic taxa in this phylum are characterized by rapid spore germination, non-fastidious growth requirements and short doubling times, which are a nice fit for desert environments.

1.3.4. Bacteria in Desert-like Environments and Bacteria involved in Weathering

Desert-like soil surface features such as desert pavements, surface accumulation of salt, calcium carbonate accumulation, and surface exposure of gypsum materials are manifestations of some kind of land deterioration [172]. This kind of land surface is commonly seen in semi-arid regions outside desert boundaries.

Mineral soil texture in desert-like condition is commonly sandy loam to loam sand. Sandy soils are formed by the weathering of the Earth's surface. Sand is the largest of all soil particle types and more spread apart than the particles of organic or clay soils, and can rarely retain surface water, resulting in less vegetation cover for surface protection. Sandy soils are formed from rock such as shale, granite, quartz and limestone. Sand allows air to freely circulate around it.

Desert-like sandy soils are presented as coastal sand soils, sandy loam in the forests and grassland, and local soil properties shape the bacterial diversity and communities. Coastal sandy soils in general lack three macronutrients: nitrogen, phosphorous and potassium [173], which is reported as the primary limit for growth of vegetation in this soil environment.

Russo *et al.* (2012) studied the bacterial communities in forests with sandy loam soil textures in the Lambir Hills National Park of Malaysia, which is sandstone-derived, nutrient-depleted and well-drained, and showed that *Proteobacteria* were dominant in sandy loam, while *Acidobacteria* were the most abundant group in clay [174]. They reported that *Actinobacteria*, *Betaproteobacteria*, *Clostridia*, *Bacilli* and *Gammaproteobacteria* were more abundant in sandy loam than in clay. Halliday *et al.* (2014) studied the bacterial community of dry beach sand in Avalon Bay Beach

(Catalina Island, USA), and reported that the phyla *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, *Planctomycetes* and *Acidobacteria* are dominant [175]. This study also claimed that bacterial communities of beach sand are broadly similar to soil communities at the phylum level and strongly influenced by soil pH and temperature. McHugh *et al.* (2014) reported that in the semi-arid grasslands of Arizona and New Mexico, bacterial communities were dominated by members the phyla *Actinobacteria* (53 %), *Proteobacteria* (16 %), and *Acidobacteria* (8.7 %) [176].

Microorganisms colonized on the surfaces of mineral soils contribute to precipitation of new minerals and to carbonate production, which plays an important role in the soil environment by contributing to the release of key nutrients. Several bacterial strains from different genera have been found to have mineral-weathering abilities, such as *Anabaena*, *Bradyrhizobium*, *Burkholderia* and *Collimonas* [177]. It has been reported that surface soil mineral particles appear to be inhabited by different communities: in limestone, the endolithic bacterial communities are comprised of Gram-positive bacteria and *Acidobacteria*, while the epilithic population are ~50% *Proteobacteria* [126]. Different primary minerals, such as granite, limestone, apatite, plagioclase, and quartz are colonized by different bacteria [178,179]. Studies of bacterial communities in soils with different mineral composition showed that concentrations of major elements, such as aluminium and calcium, seem to have a significant impact on the structure of the bacterial community [180].

1.4. Relationship of Environmental Factors and Soil Bacterial

Diversity

1.4.1. General introduction to the relationships of environmental factors and soil microbial diversity

Soil is a highly complex and important biome possesses immense bio-diversity and a large number of biological processes [181]. Bacteria constitute the largest portion of the biodiversity in soils, and play an important role in maintaining soil processes, which eventually affects the functioning of terrestrial ecosystems [182]. Richness and patterns of microbial diversity are affected by different environmental factors. Recent improvements in techniques have brought a revolution in our understanding of microbial diversity, and allow us to survey the diversity of microbial communities. Many studies have been performed to explore how the changes of environmental physiochemical parameters can affect the microbial diversity and community structures [183]. Several factors, such as pH, total carbon, organic materials, are reported to be able to shape the local microbial community.

Previous studies using isolates from soils based on pure cultures have revealed bacterial diversity within defined isolated taxa [184]. However, the taxa detected by culturing are known to not reflect all the taxa in an environment [185], and molecular methods based on diversity profiling of environmental metagenomics are more useful in the study of bacterial communities. Studies of microbial biogeography using NGS can often provide key insights into environmental tolerances and community structure of microbial taxa, particularly those difficult-to culture taxa that often dominate in natural environments [186].

Soil microbial communities are largely diverse and present geographic and

environmental specificity in the prevalence of different bacterial groups [187]. It is now widely accepted that soil microbial communities are significantly affected by environmental factors at different geographic scales [94]. Geographic location and physicochemical properties are believed to be major factors that affect soil bacterial diversity and community structures [90]. However, it was suggested that when soils are characterized by distinct environment factors, each will likely inhabit a unique microbial community, regardless of the geographic distance between them [74].

A study of microbial community structure from soil samples of the hyper-arid core in the Atacama desert indicated that salt content and water availability significantly correlated with the diversity of microbial communities. Nitrogen can be a limiting factor for biomass production and biological activities in soils of arid environments [75].

Andrew *et al.* (2012) reported that in Sonoran desert soils, microbial communities are shaped primarily by soil characteristics associated with geographic locations, while rhizosphere associations are secondary factors [187]. Similarly, a study across Israel and the United States found that the bacterial community structures in the phyllosphere of Tamarix trees are driven by climate instead of specific rhizosphere factors [188].

Lozupone *et al.* (2007) argued that the main environmental determinant of microbial diversity is salinity, rather than extremes of temperature, pH, or other physical and chemical factors represented in their samples [189]. El Hidri *et al.* (2013) Reported a huge phenotypic and phylogenetic diversity observed in arid and saline systems in southern Tunisia. This study indicated that extremely haloalkalitolerent bacteria were the most dominant group and were affiliated to *Bacillus*, *Nesterenkonia*, *Salinicoccus*, and *Marinococcus* genera [190].

1.4.2. Environmental factors in Deserts

An ubiquitous feature for deserts is the scant, low precipitation level, and they are also characterized by extreme fluctuations in temperature, generally low nutrient status, high levels of incident ultraviolet (UV) radiation and strong winds [170,191]. It has been demonstrated that deterministic factors drive bacterial community structure processes at both global and local scales [192,193]. Here, I summarize some environmental factors that have been reported as important drivers for bacterial communities in desert soil:

- 1) **Precipitation.** Lack of precipitation is the major feature of arid land, and water availability is considered to be a key limiting factor for all living things. Species richness declines with increasing aridity in deserts [152]. In the most extreme hyper-arid environment, microbial life retreats to isolated “oases” (sheltered niches or under rocks) that are formed as a result of biotic-abiotic relationships between the microorganisms and the available porous and deliquescent mineral substrates, permitting life in landscapes as a refuge from the great extinction pressures. The impact of precipitation on microbial diversity has been reported in many studies of deserts [194]. Some studies claimed that water availability is the primary controlling factor for microbial activity, diversity and community [195]. Gillor *et al.* (2010) performed bacterial diversity analysis along a large scale of a precipitation gradient using quantitative PCR and terminal restriction fragment length polymorphism (T-RFLP) and showed that, although soil bacterial abundance decreases with precipitation, bacterial diversity is independent of a precipitation gradient, and community composition is unique to each ecosystem [196].
- 2) **pH.** The effect of pH values are often considered as a main driver for bacterial communities in soil. Lauber *et al.* (2009) examined bacterial communities in 88

soils from across North and South America using high-throughput sequencing of PCR amplified 16S rDNA genes to investigate the relationships between the pattern of bacterial communities and their environmental factors. These contained ten samples from arid regions such as the Mojave desert, with pH values ranging from 7 to 8.5 [186]. Their results showed that these desert soils possess unique bacterial community structures that are distinguishable from other soil samples, and could be explained by their higher pH values. They also found that overall bacterial community composition was significantly correlated with differences in soil pH, largely driven by changes in the relative abundance of Acidobacteria, Actinobacteria, and Bacteroidetes across the range of soil pH examined. Other studies of microbial diversity in deserts soils illustrated that the impact of soil pH in arid environments may not be as important as moisture or radiation in deserts [197].

- 3) **UV Radiation.** UV Radiation in the deserts is markedly higher than in other biomes as a result of the low levels of atmospheric water vapor and tree cover, and it is an important factor limiting surface soil biodiversity. The flux of UVA and UVB is currently substantially higher on the Antarctic continent than elsewhere on Earth because of the polar ozone hole depletion and longer day lengths during the austral summer. The lithic environment can provide protection against UV exposure. A high diversity of bacteria associated with hypoliths has been reported in the Antarctic Dry Valleys, located in the polar region and exposed to high summer radiation levels [149].

- 4) **Salinity.** Saline soils are characterised by high concentrations of salts and by an uneven temporal and spatial water distribution [198], which is commonly found in deserts. The ions responsible for salination are Na^+ , K^+ , Ca^{2+} , Mg^{2+} and Cl^- . A high concentration of salt in soil changes the availability of water and nutrients for microorganisms, and can influence the size and the activity of soil microbial

biomass. Pointing *et al.* (2012) suggests that soil salinity is a major factor in community structure in desert ecosystems, and even modest levels of soil salinity may be an important determinant because biological water availability is reduced in the presence of soluble salts [126]. Aislabie *et al.* (2009) indicate that *Actinobacteria* and *Bacteroidetes* were more prevalent in dry alkaline soils and *Gammaproteobacteria* in dry saline soils [199].

- 5) **Temperature.** Extreme fluctuations in temperature and variations in diurnal and seasonal cycles also present major challenges in arid lands. Direct stresses are imposed by heat and cold shock and by freeze-thaw cycles, which are common in both hot and cold deserts [200]. In a comparative study of hot and cold deserts in China, it was reported that the presence of specific lineages of *Deinococci* may be related to mean annual temperature [152].
- 6) **Substrate mineralogy.** Soil is composed of a mixture of minerals that differ in element and organic compounds. Studies of environments indicate that different minerals in soil may select distinct bacterial communities in their microhabitats [201].
- 7) **Nutrient availability.** Nitrogen is often regarded as a limiting nutrient in oligotrophic environment like deserts, and it has been reported that microorganism abundances were positively correlated with nitrogen levels. However, Andrew *et al.* (2012) reported that available carbon, not nitrogen, is a limiting factor in driving local microbial diversity in the Sonoran Desert soils using pyrosequencing methods [90].
- 8) **Biotic Factors.** Plant communities affect the diversity of soil microorganisms locally through interactions within the rhizosphere, and microbial communities are directly affected by the microenvironment of plant root systems [202]. Early

studies have shown that resource abundance near vegetation can control the heterotrophic bacterial numbers in desert soils [203]. It has been reported that the microbial community structure of desert soils changes with agricultural activities, and long-term farming can induce a drastic shift in the bacterial communities in desert soil [194]. Bacterial communities in agricultural soil showed a higher diversity and a better ecosystem function for plant health, but with a reduction of extremophilic bacteria.

- 9) **Other Factors.** Many factors other than the above mentioned can affect the microbial diversity in deserts, such as soil geomorphology, geographic distance, altitude, concentration of heavy metals and etc. Sessitsch *et al.* (2001) claimed that the soil particle size is one of the abiotic drivers for microbial abundance [204]. Finkel *et al.* (2012) indicated that in Sonoran Desert soil, the geographic distance across different sample sites was the most important parameter which affected community composition, particularly that of *Betaproteobacteria* [188]. Seasonal climate factors such as monsoon precipitation in semiarid zones can significantly influence the bacterial population. For example, McHugh *et al.* (2015) reported that in the semi-arid zones of Arizona and New Mexico, the *Firmicutes* phylum experienced over a six-fold increase in relative abundance with increasing in response to monsoon precipitation. Conversely, *Actinobacteria*, the dominant taxa at the site, were reversely correlated to moisture availability [176].

1.4.3. Adaptation of Bacteria to Arid Environments

Microorganisms can adapt to environmental variations much faster than multi-cellular organisms. They are pioneer colonizers that have profound influence on the climate and environments on Earth [205]. Remarkable phylogenetic and metabolic diversity of bacteria, with the ability to develop biofilms, make them adapt and colonize extreme environments not tolerated by other organisms [206]. To survive in the deserts, an efficient metabolic stress response during growth and the ability to transition between active and dormant states are necessary to microorganisms.

It is well known that cells present stress tolerance strategies to avoid moisture, thermal and ultraviolet (UV) stress, but ecological studies are now revealing that bacteria exhibiting adaptation at the community level are also critical to the colonization of desert environments [207]. Desiccation tolerance is mediated both intracellularly by UV-absorbing compounds, and extracellularly by cell walls or polymeric substances, which is an indirect result of coping with stresses, such as osmotic, temperature, and oxygenic stresses [208].

Microorganisms in surface soils of deserts are directly exposed to high environmental stress levels, and possess a range of strategies against potential UV damage such as screening by pigments and damage-repair mechanisms [209]. Extreme ionizing-radiation resistance has been observed in several members of the domains Bacteria and Archaea, among which *Deinococcus* and *Rubrobacter* show the highest levels of resistance [210].

Huang *et al.* (2015) studied the diversity of the radiation-resistant microbes of the hyper-arid Taklamakan desert, and reported that radiation-resistant phylotypes belonged to the genera *Knoellia*, *Lysobacter*, *Nocardiodes*, *Paracoccus*, *Pontibacter*, *Rufibacter* and *Microvirga* [211]. *Cyanobacteria* are well presented in a range of hot

and cold deserts, which are important in biogeochemical cycling processes such as C or N utilization and stress response [74]. Studies found that members of the *Cyanobacteria* can maintain photosynthetic metabolism in desert-like conditions such as high radiation, desiccation, and salt stress [212].

1.5. Objective of Thesis

It is now widely accepted that measures must be taken to estimate, record, and conserve microbial diversity, not only to sustain human health but also to enrich the human condition globally through the wise use and conservation of genetic resources of the microbial world [213]. Desert environments, because of the range and severity of environmental factors, are an obvious target for fundamental research on the ecological and evolutionary processes which structure biological communities. They are also important because of the unique species that can occur there.

Arid and semi-arid ecosystems are important to study the potential impact of desertification on the microbial activity and communities structure of arid land [193]. By using NGS technology to study microbial communities, researchers have gained a new appreciation for the dynamics of microbial diversity in specific habitats, the variability of microbial diversity in spatial and temporal scale, and the environmental factors driving this variability [20].

In my thesis research, I used pyrosequencing of 16S rDNA amplicons to analyse the bacterial diversity and communities in the arid regions of San Rafeal Swell of Utah State, as well as a desert-like tourist site – the Desert of Maine, in the USA. We also examined some physicochemical parameters of sample sites to investigate the correlations between bacterial community structure and environmental drivers.

**CHAPTER 2: Bacterial Communities of the Desert of
Maine**

**Bacterial communities in the mineral sandy soil from the Desert of
Maine (USA) environment**

Yang Wang, Jorge R. Osman, Michael S. DuBow*

Laboratoire de Génomique et Biodiversité Microbienne des Biofilms, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Univ. Paris-Sud, Bâtiment 409, 91405 Orsay, France

*Address for correspondence: Laboratoire de Génomique et Biodiversité Microbienne des Biofilms, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Univ. Paris-Sud, Bâtiment 409, 91405 Orsay, France ;
Tel :+33(0)169154612 ; Fax : +33(0)169157808 ; E-mail : michael.dubow@u-psud.fr

Key Words: Mineral Soil, Bacteria, Biodiversity, 16s rDNA, Pyrosequencing

Running Head: Bacterial Communities of the Desert of Maine

Abstract

The Desert of Maine is a tract of glacial silt, surrounded by a pine forest, in the state of Maine located in the northeastern USA. The soil of the Desert of Maine has a sandy texture with poor water holding abilities, nutrient retention capabilities and a relatively low pH value (pH 5.09). Samples from this site thus present an interesting place to examine the bacterial diversity in mineral sandy loam soils with an acidic pH and low concentrations of organic materials. Two surface sand samples from the Desert of Maine were obtained, and pyrosequencing of PCR amplified 16S rDNA genes from total extracted DNA was used to assess bacterial diversity, community structure and the relative abundance of major bacterial taxa. We found that the soil samples from the Desert of Maine showed high levels of bacterial diversity, with a predominance of members belonging to the *Proteobacteria* and *Actinobacteria* phyla. Bacteria from the most abundant genus, *Acidiphilium*, represent 12.5% of the total 16S rDNA sequences. In total, 1394 OTUs were observed in the two samples, with the number of common OTUs observed in both samples being 668. By comparing our bacterial population results with studies on similar soil environments, we found that the samples contained less *Acidobacteria* than soils from acid soil forests, and less *Firmicutes* plus more *Proteobacteria* than soils from oligotrophic deserts.

Key Words: Mineral Soil, Bacteria, Biodiversity, 16s rDNA, Pyrosequencing

Introduction

Bacteria are integral and diverse components of soil, where their community structure and diversity have been found to be linked to many soil environmental characteristics, such as the physical and chemical properties of the soil (Lauber et al. 2009; Fierer et al. 2012a). Traditional microbial cultivation techniques frequently overlook the majority of microbes present in a sample (Amann et al. 1995), as most bacteria cannot be cultivated under laboratory conditions. Many recent studies have used high-throughput PCR amplified 16S rDNA sequencing to overcome this difficulty to identify the members of a prokaryotic community. The 16S rRNA gene is by far the most widely used genetic marker for phylogenetic and microbial community studies, as it has highly conserved regions that permit effective PCR primer design, and sufficient variable regions to allow for accurate taxonomic and phylogenetic identification of community members. Since this gene has been widely sequenced in microbial diversity surveys, there is a large amount of accumulated 16s rDNA sequence data in databases (Petrosino et al. 2009) such as the Greengenes, Silva and RDP databases. The 16S rDNA gene in bacteria includes a total of 9 hypervariable regions (V1–V9), and the V1-V3 regions have been shown to be effective for bacterial identification (Schloss 2010).

The Desert of Maine is a tract of glacial silt with a surface area of 160,000 m², surrounded by a pine forest, in southern Maine in the northeastern USA. It is not a true desert, as it receives an abundance of precipitation (76-120 cm/year), with a mean annual air temperature of 6-7 °C (<http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>, Natural Resources Conservation Service). Although it is a tourist attraction, there are no imported sand nor designer landscaped dunes. This surface area was formed approximately 11,000 years ago, during the end of the last Ice Age of the Pleistocene Period (Bahr and Friedman 2009). The parent material of the soil is sandy glaciofluvial deposits derived

from granite and gneiss. Soil and ground rocks were slowly scraped by glaciers into a sandy substance, forming a layer up to 25 meters deep. Then, over many centuries, surface soils formed a cap, concealing the “desert sand”, and allowing a forest to grow, followed by the subsequent development of agriculture. The glacial “desert” was once covered by a farm, and exposed because of severe soil erosion due to crop rotation mismanagement (<http://www.desertofmaine.com>).

The soil of the Desert of Maine has a sandy texture with poor water holding abilities, nutrient conservation capabilities and an acidic pH value. Mineral sandy loam soils contain less organic materials, with a basic pH (Griffiths et al. 2011; Crits-Christoph et al. 2013), while low pH soils generally contain more organic materials, such as the soils of forests and some grasslands (Nacke et al. 2011; Shah et al. 2011; Russo et al. 2012). The Desert of Maine presents an interesting example to observe the bacterial populations in a mineral sandy loam with a relatively low pH and concentration of organic materials. Here, we used pyrosequencing of PCR amplified 16S rDNA genes to assess bacterial diversity, community structure and the relative abundance of bacterial taxa within two sites of the surface soil of the Desert of Maine, and compared its bacterial community with those from several other sandy desert environments, as well as from other mineral soils.

Materials and Methods

Sampling. Two surface sand samples from the Desert of Maine (Fig. 1) were obtained on September 2, 2011. The two samples were collected by scooping surface sand into 50 ml sterile polyethylene conical centrifuge tubes, in an area cordoned off from tourists and without any measurable rainfall for at least four days. The average air temperature during the week of sampling was 22°C. After collection, the samples were treated as previously described (An et al. 2013). To perform the analyses of selected physicochemical parameters of the sample site soil, samples from the two sites were pooled and sent for analyses using standard methods to the Laboratoire d'Analyses de Sols (Institut National de la Recherche Agronomique, France).

Sample Preparation. Total DNA was extracted from each sample using the protocol of An et al. (2013). An aliquot of extracted DNA was adjusted to a final DNA concentration of 15 ng/μl in 1/10 TE buffer (1 mM Tris pH 8; 0.1 mM EDTA) using a NanoVue spectrophotometer (GE Healthcare, Little Chalfont, Buckinghamshire, UK), and the concentration verified by ethidium bromide fluorescence after electrophoresis through a 1% agarose gel in TAE buffer (2 mM Tris-acetate pH 8; 5 mM Na-EDTA). PCR reactions were performed in 25 μl reaction volumes. Each reaction contained one of two different thermostable DNA polymerases and their corresponding reaction buffers, 200 μM of each dNTP, 0.5 μM of each primer and 1 to 10 ng of extracted DNA. The 16S rRNA genes were amplified using the universal bacterial primers for pyrosequencing and covering hypervariable regions V1-V3: primer 27F (A adaptor + GAGTTTGATCMTGGCTCAG) and primer 518R (B adaptor + Mid + WTTACCGCGGCTGCTGG), where A and B represent the adaptors using the 454 Roche FLX Titanium pyrosequencing reaction platform. The Mid sequences are eight nucleotide tags designed for sample identification barcoding according to the 454 protocol. PCR amplification conditions were adapted for the use of two different thermostable DNA polymerases: A) Phusion High-Fidelity DNA Polymerase

(Finnzymes, Finland): 98°C for 2 min, followed by 28 cycles of 98°C for 30 s, 54°C for 20 s and 72°C for 15 s, and a final elongation step at 72°C for 5 min; B) Pfu DNA Polymerase (Fermentas, Canada): 95°C for 2 min, followed by 30 cycles of 95°C for 30 s, 48°C for 30 s and 72°C for 1 min, and a final elongation step at 72°C for 5 min. Each DNA sample was subjected to 3-5 different PCR reactions per DNA polymerase to minimize PCR bias. The PCR products were pooled and subjected to electrophoresis through a 1% agarose gel in TAE buffer. After electrophoresis and visualization of the PCR products by ethidium bromide staining and long wave UV light illumination, NucleoSpin Extract II kits (Macherey-Nagel, Germany) were used to purify the 16S rDNA PCR products. Then, 40 ng of PCR products from each sample were mixed for pyrosequencing, performed using a 454 Roche FLX Titanium Pyrosequencer (Microsynth AG, Switzerland).

DNA Sequence Data Processing. The raw DNA sequences were first assigned to each sample via their Mid tag using MOTHUR version 1.33 (Schloss et al. 2009), and reads were removed if at least one of the following criteria was met: (i) length less than 200 nt or longer than 600 nt, (ii) mismatch to the barcode sequences or more than one mismatch to the primer, and (iii) the presence of homopolymers of > 8 bp in length. Adaptor sequences were removed from the sequences using the “Cutadapt” tool (Martin 2011) implemented in the Galaxy server of the Institut de Génétique et Microbiologie (IGM) of the Université Paris-Sud (<http://galaxy.igmors.u-psud.fr>). Then, the sequences were checked for quality scores by ConDeTri version 2.2 (Smeds and Kunstner 2011), using the criteria that 80% of the nucleotides in a sequence have quality scores > 25. We used UCHIME (Edgar et al. 2011), with reference database Greengenes version 2013_May, and Decipher (through web tools available at <http://decipher.cce.wisc.edu/>) to detect chimera sequences (Wright et al. 2012). Sequences detected as chimeras by both programs were removed from the data sets. The raw sequences have been deposited in the GenBank short-read archive (SRA), with accession number SRP056525.

Taxonomy assignments of the remaining 16S rDNA reads were conducted using RDP II classifier with a confidence threshold cutoff of 80, using the Silva database release 119 reformatted in MOTHUR (database available on the download page at the website of MOTHUR) (Cole et al. 2014; Yilmaz et al. 2014). Sequences classified as Chloroplast, or that could not be classified as belonging to the Bacteria Kingdom, were removed. Diversity analyses were performed using the software package MOTHUR. The relations of the relative abundance of bacterial groups at different taxonomy levels between the two samples were calculated using the Pearson correlation coefficient measure with SPSS Statistics (Version 22). The clean reads were clustered into operational taxonomic units (OTUs) using UPARSE at a cutoff value of 97% sequence identity (Edgar 2010; Edgar 2013). The Chao1 and Shannon indices were calculated to estimate taxon richness and diversity (Schloss and Handelsman 2008). The significance of differences between two bacterial communities was calculated using Libshuff implemented in MOTHUR (Schloss et al. 2004), with 1500 randomly-selected sequences from each sample selected using PANGEA (Giongo et al. 2010).

Results

Chemical and Physical Properties of the Sand Samples

Two areas, separated by 3 m, of the Desert of Maine were sampled on September 2, 2011. The mean chemical and physical properties of the sand samples are shown in Table 1. The mean pH values of the soil at the sampling site was 5.09, indicating an acid soil environment. The levels of total organic carbon and organic material were less than 1 g/kg soil.

Diversity Analyses

The average length of the raw DNA sequences for the two samples were 479 nt (Maine 1) and 480 nt (Maine 2), respectively, while the total number of reads for each sample were 23,405 (Maine 1) and 28,983 (Maine 2), respectively. After bioinformatic cleaning, approximately 95% of the sequences remained (22,320 for Maine 1 and 27,680 for Maine 2). The sequences were further filtered for quality and examined for chimeric sequences, leaving 65% of the total reads (14,776 for Maine 1 and 19,085 for Maine 2). The average length of the sequences after processing were 396 nt for the two samples. The number of reads remaining after each step are presented in Table 2.

The clean sequences were clustered into OTUs at 97% similarity levels, excluding the unclassified sequences at the phylum level, and sequences classified as Chloroplast were removed. In total, 1394 OTUs were observed in the two samples, and the number of common OTUs observed in both samples was 668. The numbers of core taxons (most abundant OTUs with more than 1% of the sequences), were 18 in the Maine 1 sample and 14 in the Maine 2 sample. The core taxons comprised 30% of the bacterial population in the Maine 1 sample and 23% in the Maine 2 sample. We observed no differences between the Shannon diversity indices of the two samples (Table 2).

Classification of DNA Sequences

The sequences of the two samples were classified at 6 taxonomic levels with RDP Classifier, and comprise at least 22 phyla, 41 classes, 76 orders, 115 families and 172 genera, plus a number of unclassified sequences at various taxonomic levels. The distribution of sequences at the phylum level is shown in Fig. 2. Unclassified sequences at the phylum level represent 7.0% of the sequences in Maine 1 and 6.4% in Maine 2. The community structure at the phylum level shows a similar distribution for the two samples ($R=0.989$ with the Pearson coefficient measure). The significance of the differences between the two bacterial communities was calculated using the Libshuff package in MOTHUR, with a P value < 0.01 , indicating that the two bacterial communities do have not the same composition. The predominant phyla of the samples represent members of the *Proteobacteria*, *Actinobacteria*, *Chloroflexi*, *Bacteroidetes*, and *Acidobacteria* phyla, at 39.1%, 18.5%, 10.3%, 6.4%, and 5%, respectively as the average for the two samples, followed by members of the *Cyanobacteria* (3.4%), *Planctomycetes* (2.5%), *Gemmatimonadetes* (2.4%), WD272 (2.0%), *Armatimonadetes* (1.9%), TM7 (0.5%), and *Deinococcus-Thermus* (0.5%) phyla. There are 10 phyla that each comprise $< 0.5\%$ of bacterial sequences (*Verrucomicrobia*, *Chlorobi*, *Nitrospira*, *Elusimicrobia*, SM2F11, TM6, WCBH1-60, OP11, SHA-109 and *Firmicutes*), and these are grouped together as “rare phyla”. The percentage in the two samples assigned to each of these low abundance phyla are shown in Table 3.

Within the phylum *Proteobacteria*, members of the *Alphaproteobacteria* represent the most abundant group in both samples (67.8 % in Maine 1, 63.3% in Maine 2), followed by members of the *Betaproteobacteria* (15.2%, 16.8%), and *Deltaproteobacteria* (12.9%, 14.9%). The bacterial community structure at the Class level (Fig. 1) presents a similar pattern of distribution in the two samples ($R=0.959$ with the Pearson correlation coefficient measure). The members of the *Sphingobacteriales* family were the predominant members of the *Bacteroidetes*

phylum, with an average of 4.8 % of the total reads for the two samples. Members representing *Rhodospirillales* and *Burkholderiales* were the most abundant groups within the *Alphaproteobacteria* and *Betaproteobacteria* classes, respectively. The dominant Families in the two samples come from the *Acetobacteraceae* (16.8 %), *Chitinophagaceae* (4.6 %) and *Oxalobacteraceae* (4.3 %).

Among the 172 genera identified in the samples, 69 genera belong to the phylum *Proteobacteria*, 41 genera belong to the phylum *Actinobactia*, 24 genera belong to the phylum *Bacteroidetes*, and 9 genera belong to the phylum *Acidobacteria*. The most abundant 20 genera in each sample are shown in Table 4. The similarity between the two bacterial communities at the genus level is 0.975 using the Pearson correlation coefficient measure. Each of these abundant genera account for 0.6%-11.3% of bacteria identified in the Maine 1 sample and 0.8 %-13.6 % in the Maine 2 sample, with the most abundant genus being from *Acidiphilium* (in the phylum *Proteobacteria*) for both samples. Members of the genus *Crinalium* (1.9% in average for our two samples) represent a group of phototrophic bacteria belonging to the phylum *Cyanobacteria*, and species of this genus have been reported to be highly drought-resistant and to be isolated from coastal sand dunes (Ben de Winder et al. 1990; Ben de Winder and Mur 1994). Members of the genus *Arthrobacter* are abundant in both samples (1.4% in average for our two samples), and members of this genus are frequently involved in mineral weathering of soil and can secrete large amounts of oxalic acid (Uroz et al. 2009; Frey et al. 2010).

Discussion

In recent years, desert-related environmental effects have been increasing, as global warming and human activities contributing to desertification are increasingly threatening ecosystems around the world. Studies concerning microbial colonization and dispersion in deserts have been performed to estimate the function of microbial communities from desert sand which may play an important role in soil stability, nutrient cycles and environmental health. The Desert of Maine was previously productive agriculturally, and was covered by farm land. The mineral soils were exposed because of severe soil erosion due to crop rotation mismanagement. In this study, we used pyrosequencing of PCR amplified 16S rDNA gene to assess bacterial diversity and community structure of surface soil of the Desert of Maine. An examination of bacterial populations in samples of surface soil from this unique site provides an opportunity to investigate bacterial diversity and community structure in a desert-like environment, and its relation with those from other soils.

Previous studies have revealed that many environmental factors, including pH, the concentration of organic material and that of sodium, can have large effects on the presence and distribution of bacterial community members in soil (Rousk et al. 2010; Griffiths et al. 2011; Centeno et al. 2012). A large proportion of the variance in soil bacterial diversity and community composition appears to be strongly influenced by pH (Fierer et al. 2012b), at local (Osborne et al. 2011) and even continental scales (Lauber et al. 2009). In this study, we compared our data on the distribution of predominant phyla (Fig. 3) with those from studies using 16S rDNA gene pyrosequencing of samples taken from apparently similar soil environments or with similar physiochemical factors. The Desert of Maine is surrounded by a pine forest. Thus, its soil microbiome may be influenced by that of the pine forest environment around it. Uroz et al. (2010) studied the bacterial community of soils from an oak forest of Breuil-Chenue in Morvan (France), and found that members of

Acidobacteria account for about 36% of the total bacterial population in their sample. Shah et al. (2011) examined the sandy, acidic and nutrient-poor soil of a pine barrens region of Long Island (New York, USA), which is also composed of gravel deposited by the withdrawal of glaciers. This soil has a pH value of 4.9, total organic carbon of 20.9 g/kg, and Al and Fe concentration of approximately 0.1 g/100g. Other samples of surface soil from hot or cold deserts were compared with our results: 1) a sand sample (Gobi 1) from the Gobi Desert of Northwestern China (An et al. 2013), which also has a low concentration of organic carbon and organic materials (< 1 g/kg); 2) a sample (Altamira) from the Atacama Desert in Chile (Crits-Christoph et al. 2013); 3) a sample (Upper Wright Valley) from McMurdo Dry Valleys (Lee et al. 2012) in the area of the Antarctic continent (a cold desert). All these studies were performed using pyrosequencing of 16S rDNA amplicons. We used the Pearson correlation coefficient measure to estimate the similarity of the distribution of the predominant phyla among the different sites. The results showed that the distribution of the predominant phyla in both samples from Maine are more close to samples from the two forest soils (R values > 0.7), than to samples from both hot and cold deserts (R values < 0.5). For the Phylum *Proteobacteria*, there is no significant difference between our samples and these from the other two forest samples, but the relative abundance of this phylum in our samples is greater than that of the desert samples ($P < 0.05$). The Gobi Desert sample, however, contained a much higher proportion of members of the *Firmicutes* than the others. Samples from the Atacama Desert and the McMurdo Dry Valleys contain a larger population of *Actinobacteria* members when compared with other groups ($P < 0.05$). There is no significant difference in the percentage of *Bacteroidetes* among the different samples examined here. These results confirm those of earlier studies (Griffiths et al. 2011; Tiao et al. 2012; Crits-Christoph et al. 2013) that suggested that oligotrophic environments with a large mineral component and low levels of organic materials have a large population of Gram positive bacteria, such as those from the *Firmicutes* and *Actinobacteria*. Previous studies (Lauber et al. 2009) also indicate that high pH soils typically have a higher relative abundance of

members of the *Actinobacteria* and *Bacteroidetes* phyla, with a lower abundance of *Acidobacteria*, when compared with populations from more acidic soils. However, we did not find significant differences on the percentage of *Acidobacteria* among the different soils that we compared, suggesting that other factors may affect the bacterial community structure of these types of soils.

Lauber et al. (2009) examined bacterial communities in 88 soils from across North and South America using high-throughput sequencing of PCR amplified 16S rDNA genes, and their results showed that, in soils with pH values of 5-6, the dominant phyla were members of the *Acidobacteria*, *Alphaproteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Beta/Gammaproteobacteria* (Jones et al. 2009). In our samples, which had a similar pH range, we observed a lower proportion of *Acidobacteria* (5.1% vs 29.7%), and a higher proportion of *Actinobacteria* (18.5% vs 8.8%). Our data also reveal a high level (10.3%) of *Chloroflexi* phylum members, which is not commonly found in studies of deserts. However, a study of the Atacama Desert showed that the non-cyanobacteria phototrophic bacteria *Chloroflexi* was dominant in the hyper-arid core of the deserts (Lacap et al. 2011). Previous studies have shown that members of the *Chloroflexi* may play an important role as soil photoautotrophs and contribute to CO₂ uptake in the surface soil (Ley et al. 2006; Freeman et al. 2009). Members of the Family *Acetobacteraceae* were abundant in both our samples, comprising 16.8% of total sequences, on average. Members of this family have been described as nitrogen fixing bacteria able to act in plant growth promotion by a variety of mechanisms (Reis and Teixeira 2015). Koberl et al. (2011) reported a greater proportion of N-fixing bacterial groups in desert soils than farm soils, and suggest that this could be explained by the fact that plant growth promoting bacteria play an important role as a nitrogen donor in soils without compost treatment.

The abundant genera shown in Table 4 demonstrate that the dominant genus in our two samples is *Acidiphilium*, accounting for 11.3% of the bacteria in the Maine 1

sample and 13.6% for the Maine 2 sample. *Acidiphilium* is a genus in the phylum *Proteobacteria*, and many species from this genus are acidophilic bacteria isolated from acidic mineral environments (P. Harrison 1981; Wichlacz et al. 1986), which is consistent with the composition of our sample site. *Acidiphilium spp.* are also involved in the iron cycle, with the function of reducing ferric iron by oxidizing organic matter at low pH (Sanchez-Andrea et al. 2011). In mineral soils, Fe-oxidizing bacteria are well represented (Kan et al. 2011; Wu et al. 2015). Lithotrophs, such as members of *Microcoleus* in the phylum *Cyanobacteria*, are typically the dominant microorganisms in the microbial community in deserts, as well as sulfate-reducing bacteria like *Desulfobacterales*, but these two groups of bacteria do not appear to be abundant in the Desert of Maine soil (Pointing and Belnap 2012). Members of the *Alphaproteobacteria*, *Acidobacteria* and *Actinobacteria* are ubiquitous in mineral environments, which is consistent with the distribution of bacteria in our samples. We found that the samples contain relatively low levels of mineral weathering bacteria such as members of the *Burkholderia*, *Agrobacterium* and *Bacillus* genera, and only one abundant genus, *Arthrobacter*, was found to be correlated with mineral weathering (Uroz et al. 2009; Frey et al. 2010; Lepleux et al. 2012).

We also observed that 5% of the total OTUs in the dataset contained > 50% of the total sequences, while approximately 80% of the total OTUs were highly diverse and contained < 20% of the total sequences. The results of taxonomy assignment of the sequences showed that >30% of the sequences were not able to be classified at the genus level. This situation is frequent in studies of soil bacterial communities using high-throughput DNA sequencing techniques (Schutte et al. 2010; de Gannes et al. 2013; Hartmann et al. 2015).

In this study, we used Pyrosequencing of PCR amplified bacterial 16S rDNA genes to reveal a high degree of bacterial diversity and community structure in two soil samples from the Desert of Maine. This small sand-like environment presents unique

bacterial community patterns when compared with sand from hot deserts, and also presents differences on the abundance of predominant microorganisms within mineral soils.

Acknowledgements

We thank Ginger and Gary Currens from the Desert of Maine for their help in obtaining the sand samples, and all the members of the Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB) of the Institute for Integrative Biology of the Cell (I2BC) for interesting discussions, comments and suggestions. YW is a scholarship recipient from the China Scholarship Council (CSC). JO is a scholarship recipient from CONICYT Becas Chile. This work was supported by the Centre National de la Recherche Scientifique (CNRS), France.

References

- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143-169
- An S, Couteau C, Luo F, Neveu J, DuBow MS (2013) Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts. *Microb Ecol* 66:850-860. doi:10.1007/s00248-013-0276-2
- Bahr J, Friedman J (2009) *Amazing and Unusual USA*. Publications International, Ltd. Chicago, Illinois, USA
- Ben de Winder, Hans C.P. Matthijs, Mur LR (1990) The effect of dehydration and ion stress on carbon dioxide fixation in drought-tolerant phototrophic micro-organisms. *FEMS Microbiol Lett* 74:33-38
- Ben de Winder, Mur LR (1994) The sensitivity for salinity increase in the drought resistant cyanobacterium *Crinalium epipsammum* SAB 22.89. In: Stal L, Caumette P (eds) *Microbial Mats*. vol 35. NATO ASI Series. Springer Berlin Heidelberg. pp 117-123
- Centeno CM, Legendre P, Beltran Y, Alcantara-Hernandez RJ, Lidstrom UE, Ashby MN, Falcon LI (2012) Microbialite genetic diversity and composition relate to environmental variables. *FEMS Microbiol Ecol* 82:724-735. doi:10.1111/j.1574-6941.2012.01447.x
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A et al. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42:D633-642. doi:10.1093/nar/gkt1244
- Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, Jedynek B, McKay CP, Diruggiero J (2013) Colonization patterns of soil microbial communities in the Atacama Desert. *Microbiome* 1:28. doi:10.1186/2049-2618-1-28

- de Gannes V, Eudoxie G, Hickey WJ (2013) Prokaryotic successions and diversity in composts as revealed by 454-pyrosequencing. *Bioresour Technol* 133:573-580. doi:10.1016/j.biortech.2013.01.138
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-2461. doi:10.1093/bioinformatics/btq461
- Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996-998. doi:10.1038/nmeth.2604
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194-2200. doi:10.1093/bioinformatics/btr381
- Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R (2012a) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* 6:1007-1017. doi:10.1038/ismej.2011.159
- Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA et al. (2012b) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *PNAS* 109:21390-21395. doi:10.1073/pnas.1215210110
- Freeman KR, Pescador MY, Reed SC, Costello EK, Robeson MS, Schmidt SK (2009) Soil CO₂ flux and photoautotrophic community composition in high-elevation, 'barren' soil. *Environ Microbiol* 11:674-686. doi:10.1111/j.1462-2920.2008.01844.x
- Frey B, Rieder SR, Brunner I, Plotze M, Koetzsch S, Lapanje A, Brandl H, Furrer G (2010) Weathering-associated bacteria from the Damma glacier forefield: physiological capabilities and impact on granite dissolution. *Appl Environ Microbiol* 76:4788-4796. doi:10.1128/AEM.00657-10
- Giongo A, Crabb DB, Davis-Richardson AG, Chauillac D, Mobberley JM, Gano KA, Mukherjee N, Casella G et al. (2010) PANGEA: pipeline for analysis of next generation amplicons. *ISME J* 4:852-861. doi:10.1038/ismej.2010.16

- Griffiths RI, Thomson BC, James P, Bell T, Bailey M, Whiteley AS (2011) The bacterial biogeography of British soils. *Environ Microbiol* 13:1642-1654. doi:10.1111/j.1462-2920.2011.02480.x
- Hartmann M, Frey B, Mayer J, Mader P, Widmer F (2015) Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J* 9:1177-1194. doi:10.1038/ismej.2014.210
- 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. doi:10.1038/ismej.2008.127
- Kan J, Chellamuthu P, Obratzsova A, Moore JE, Nealson KH (2011) Diverse bacterial groups are associated with corrosive lesions at a Granite Mountain Record Vault (GMRV). *J Appl Microbiol* 111:329-337. doi:10.1111/j.1365-2672.2011.05055.x
- Koberl M, Muller H, Ramadan EM, Berg G (2011) Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. *PLoS One* 6:e24452. doi:10.1371/journal.pone.0024452
- Lacap DC, Warren-Rhodes KA, McKay CP, Pointing SB (2011) Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. *Extremophiles* 15:31-38. doi:10.1007/s00792-010-0334-3
- 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. doi:10.1128/AEM.00335-09
- Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC (2012) The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. *ISME J* 6:1046-1057. doi:10.1038/ismej.2011.170
- Lepleux C, Turpault MP, Oger P, Frey-Klett P, Uroz S (2012) Correlation of the abundance of betaproteobacteria on mineral surfaces with mineral weathering

- in forest soils. *Appl Environ Microbiol* 78:7114-7119.
doi:10.1128/AEM.00996-12
- Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, Maresca JA, Bryant DA et al. (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* 72:3685-3695.
doi:10.1128/AEM.72.5.3685-3695.2006
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10-12. doi:10.14806/ej.17.1.200
- Nacke H, Thurmer A, Wollherr A, Will C, Hodac L, Herold N, Schoning I, Schrumpp M et al. (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* 6:e17000. doi:10.1371/journal.pone.0017000
- Osborne CA, Zwart AB, Broadhurst LM, Young AG, Richardson AE (2011) The influence of sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiol Ecol* 78:70-79. doi:10.1111/j.1574-6941.2011.01105.x
- P. Harrison J (1981) *Acidiphilium cryptum* gen. nov., sp. nov., heterotrophic bacterium from acidic mineral environments. *Int. J. Syst. Bacteriol.* 31:327-332
- Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clin Chem* 55:856-866.
doi:10.1373/clinchem.2008.107565
- Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10:551-562. doi:10.1038/nrmicro2831
- Reis VM, Teixeira KRdS (2015) Nitrogen fixing bacteria in the family Acetobacteraceae and their role in agriculture. *J Basic Microbiol* 55:931-949.
doi:10.1002/jobm.201400898
- Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* 4:1340-1351. doi:10.1038/ismej.2010.58

- Russo SE, Legge R, Weber KA, Brodie EL, Goldfarb KC, Benson AK, Tan S (2012) Bacterial community structure of contrasting soils underlying Bornean rain forests: Inferences from microarray and next-generation sequencing methods. *Soil Biol Biochem* 55:48-59. doi:DOI 10.1016/j.soilbio.2012.05.021
- Sanchez-Andrea I, Rodriguez N, Amils R, Sanz JL (2011) Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. *Appl Environ Microbiol* 77:6085-6093. doi:10.1128/AEM.00654-11
- Schloss PD (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* 6:e1000844. doi:10.1371/journal.pcbi.1000844
- Schloss PD, Handelsman J (2008) A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics* 9:34. doi:10.1186/1471-2105-9-34
- Schloss PD, Larget BR, Handelsman J (2004) Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl Environ Microbiol* 70:5485-5492. doi:10.1128/AEM.70.9.5485-5492.2004
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537-7541. doi:10.1128/AEM.01541-09
- Schutte UM, Abdo Z, Foster J, Ravel J, Bunge J, Solheim B, Forney LJ (2010) Bacterial diversity in a glacier foreland of the high Arctic. *Mol Ecol* 19 Suppl 1:54-66. doi:10.1111/j.1365-294X.2009.04479.x
- Shah V, Shah S, Kambhampati MS, Ambrose J, Smith N, Dowd SE, McDonnell KT, Panigrahi B et al. (2011) Bacterial and archaea community present in the Pine Barrens Forest of Long Island, NY: unusually high percentage of ammonia oxidizing bacteria. *PLoS One* 6:e26263. doi:10.1371/journal.pone.0026263

- Smeds L, Kunstner A (2011) ConDeTri--a content dependent read trimmer for Illumina data. PLoS One 6:e26314. doi:10.1371/journal.pone.0026314
- Tiao G, Lee CK, McDonald IR, Cowan DA, Cary SC (2012) Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys. Nat Commun 3:660. doi:10.1038/ncomms1645
- Uroz S, Buee M, Murat C, Frey-Klett P, Martin F (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environ Microbiol Rep 2:281-288. doi:10.1111/j.1758-2229.2009.00117.x
- Uroz S, Calvaruso C, Turpault MP, Frey-Klett P (2009) Mineral weathering by bacteria: ecology, actors and mechanisms. Trends Microbiol 17:378-387. doi:10.1016/j.tim.2009.05.004
- Wichlacz PL, Unz RF, Langworthy TA (1986) *Acidiphilium angustum* sp. nov., *Acidiphilium facilis* sp. nov., and *Acidiphilium rubrum* sp. nov.: Acidophilic heterotrophic bacteria isolated from acidic coal mine drainage. Int J Syst Evol Microbiol 36:197-201. doi:doi:10.1099/00207713-36-2-197
- Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. Appl Environ Microbiol 78:717-725. doi:10.1128/AEM.06516-11
- Wu Y, Tan L, Liu W, Wang B, Wang J, Cai Y, Lin X (2015) Profiling bacterial diversity in a limestone cave of the western Loess Plateau of China. Front Microbiol 6:244. doi:10.3389/fmicb.2015.00244
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J et al. (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res 42:D643-648. doi:10.1093/nar/gkt1209

Tables

Table 1. Average chemical and physical properties of the Desert of Maine soil samples

	Mean Value
Organic carbon	0.43 g/kg
Total nitrogen	<0.02 g/kg
Organic material	0.75 g/kg
Aluminium (Al)	4.86 g/100g
Calcium (Ca)	0.93 g/100g
Iron (Fe)	1.41 g/100g
Magnesium (Mg)	0.39 g/100g
Phosphorous (P₂O₅)	0.08 g/100g
Potassium (K)	1.97 g/100g
Sodium (Na)	2.15 g/100g
Manganese (Mn)	420 mg/kg
pH	5.09

Table 2. Summary of the number of sequences and diversity indices for each sample

	Sequence numbers		Diversity Index		
	Raw	After cleaning	OTUs*	Chao1	Shannon
Maine1	23405	14776	924	1145	5.57
Maine2	28983	19085	1139	1693	5.71

*OTUs are clustered at 97% sequence identity.

Table 3. Relative sequence proportions belonging to rare phyla in the Desert of Maine samples

Rare phyla	Maine 1	Maine 2
OP11	0.1%	0.1%
SHA-109	0.3%	0.5%
WCHB1-60	0.3%	0.3%
<i>Verrucomicrobia</i>	0.2%	0.3%
TM6	<0.1%	<0.1%
<i>Firmicutes</i>	<0.1%	0.1%
<i>Elusimicrobia</i>	<0.1%	<0.1%
SM2F11	<0.1%	<0.1%
<i>Chlorobi</i>	<0.1%	<0.1%
<i>Nitrospirae</i>	<0.01%	<0.1%

Table 4. Relative abundance of the 20 most abundant bacterial genera in the Desert of Maine samples.

Genus	Maine1	Maine2
<i>Acidiphilium</i>	11.3%	13.6%
<i>Flavisolibacter</i>	2.4%	2.7%
<i>Crinalium</i>	2.4%	1.3%
<i>Methylobacterium</i>	2.2%	1.3%
<i>Noviherbaspirillum*</i>	0.9%	2.3%
<i>Chthonomonas</i>	1.1%	1.6%
<i>Amnibacterium</i>	1.6%	1.2%
<i>Arthrobacter</i>	1.5%	1.3%
<i>Actinomycetospora</i>	1.3%	1.3%
<i>Blastocatella</i>	1.1%	1.1%
<i>Anaeromyxobacter</i>	0.9%	1.3%
<i>Gemmatimonas</i>	1.0%	1.1%
<i>Segetibacter*</i>	0.7%	1.4%
<i>Acidobacterium</i>	1.0%	1.0%
<i>Singulisphaera</i>	0.9%	0.7%
<i>Lapillicoccus</i>	1.0%	0.6%
<i>Candidatus_Solibacter</i>	0.6%	0.9%
<i>Chamaesiphon*</i>	1.9%	0.5%
<i>Spirosoma</i>	0.8%	0.7%
<i>Hymenobacter</i>	0.6%	0.8%

*represent genera with > 2 fold differences between the two samples

Figure Legends

Fig. 1. Location of the Desert of Maine in the northeastern USA.

Fig. 2. Relative abundance of bacterial 16S rDNA gene sequences from the two samples at the Phylum (A) and Class (B) levels. See Materials and Methods for details.

Fig. 3. Comparison of the bacterial communities with those from selected other soil samples at the Phylum level.

Relative abundance of bacterial phyla in soil samples from: Two samples of the Desert of Maine; An oak forest in France (Uroz et al. 2010); A pine forest in the USA (soil pH 4.9, total organic carbon 20.9 g/kg, Al and Fe concentration of approximately 0.1 g/100g, Shah et al. 2011); A Gobi desert sand sample in Mongolia (pH 9.8, organic carbon 0.52 g/kg, organic material 0.9 g/kg, An et al. 2013); A sample from the Atacama Desert in Chile (pH 8.9, Ca 0.96 g/100g, Na 0.34 g/100g, Chris-Christoph et al. 2013); A sample from the McMurdo Dry Valley in the Antarctic (pH 7.0, total carbon 0.11 g/kg, Fe 1.39 g/100g, Na 0.35 g/100g, Lee et al. 2012)

Fig 1.

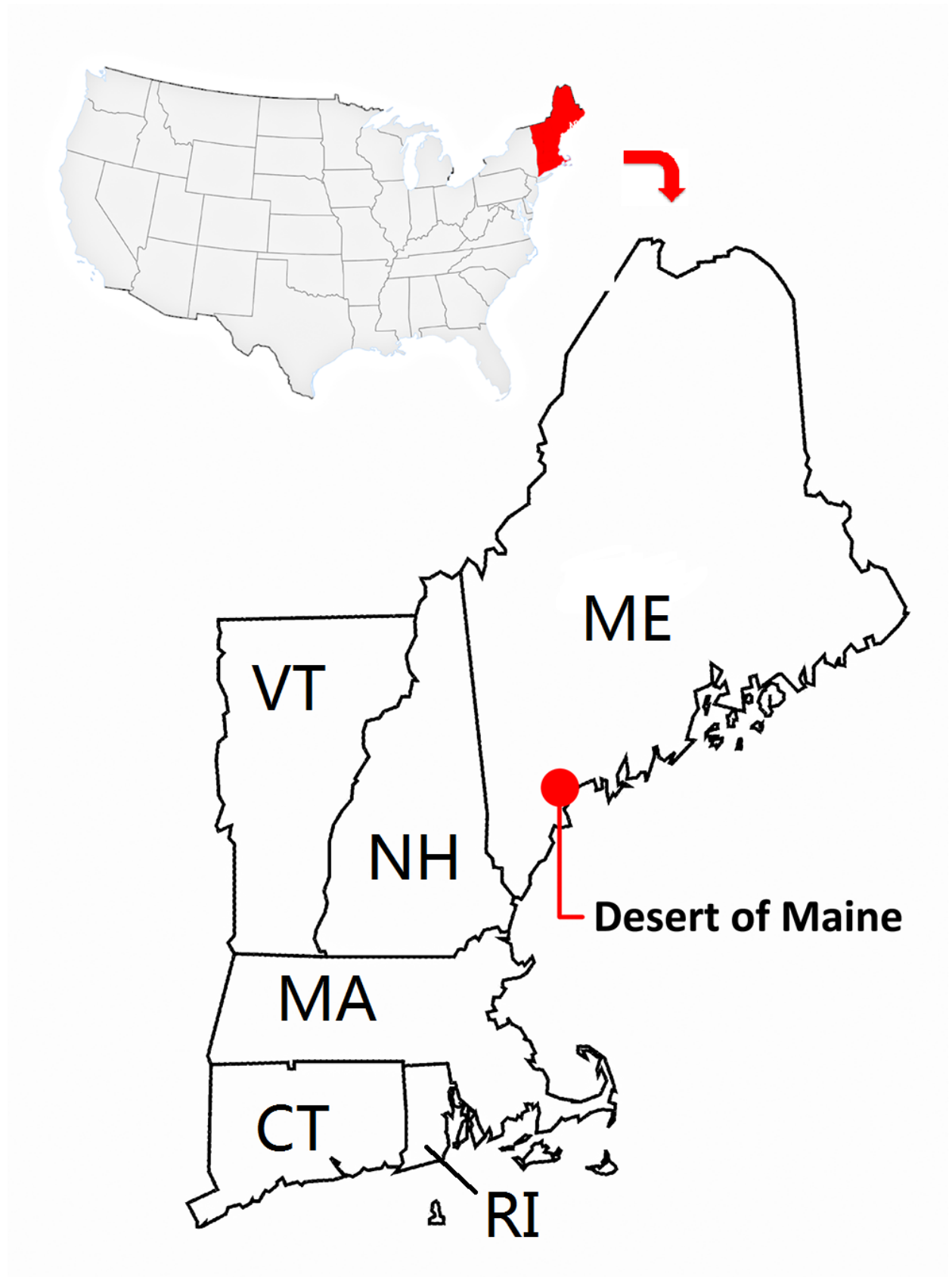


Fig 2.

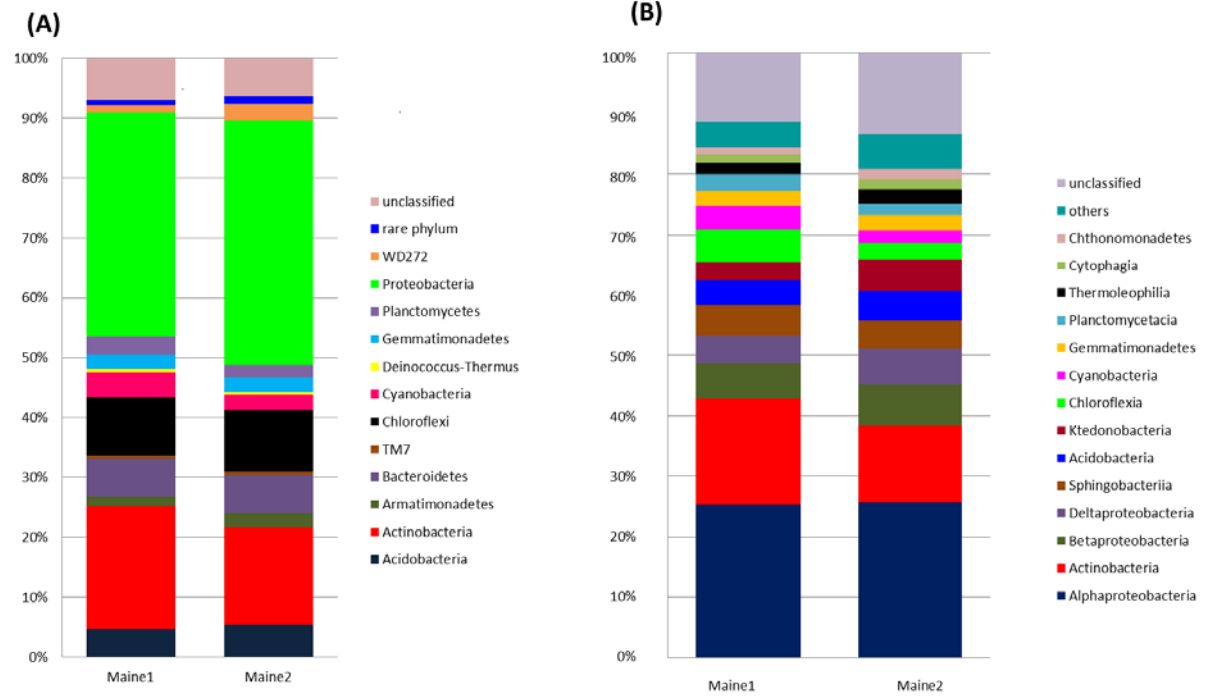
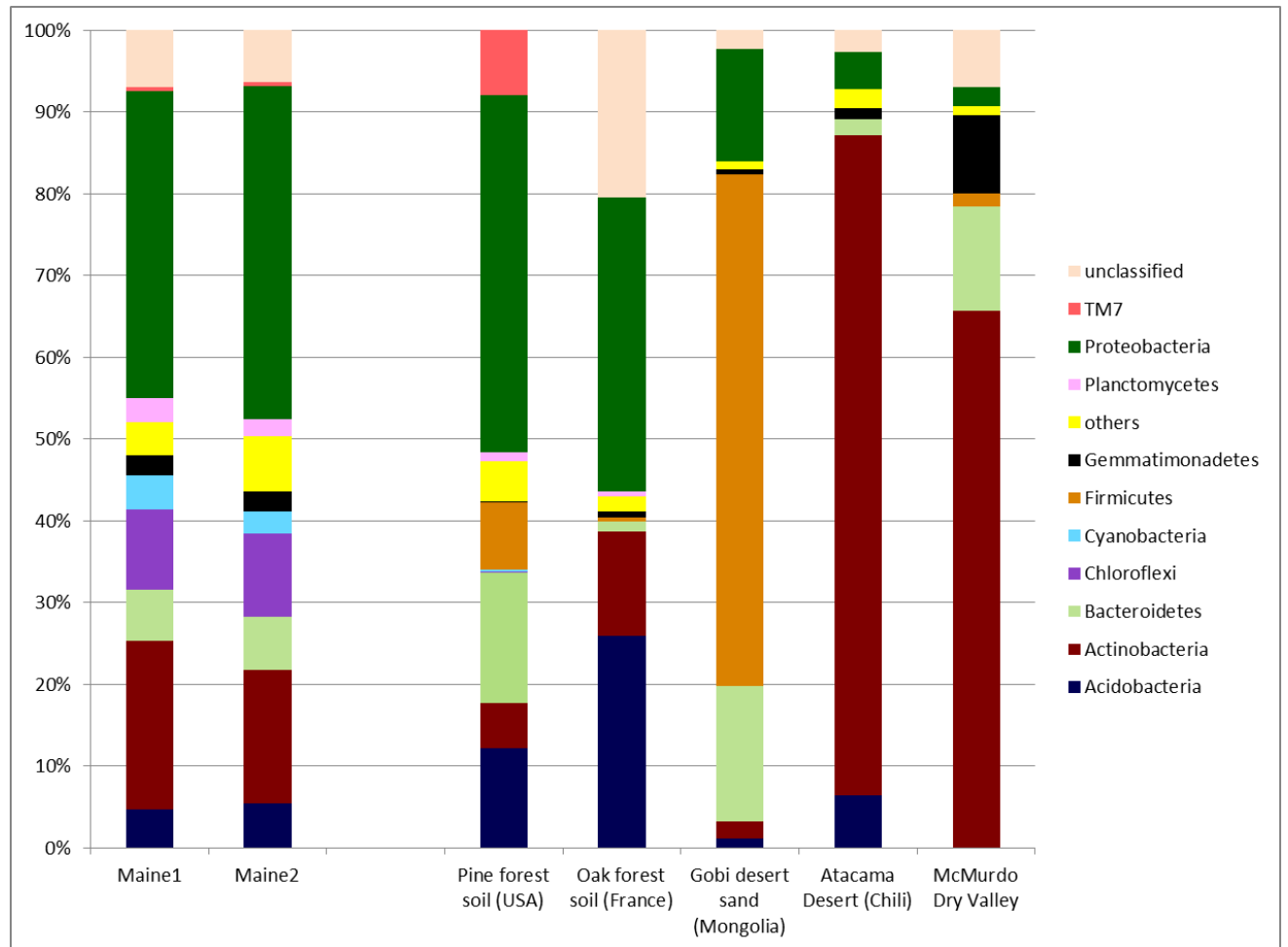


Fig 3.



CHAPTER 3: Bacterial communities of the desert in San

Rafael Swell (USA)

Bacterial community of surface soils from the San Rafael Swell

(Utah, USA)

Yang Wang¹, Chloé Jaubert², Jorge R. Osman¹, Tuyet-Nga Nguyen¹, Gustavo R. Fernandes¹, Michael S. DuBow^{1*}

¹Laboratoire de Génomique et Biodiversité Microbienne des Biofilms, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Univ. Paris-Sud, Bâtiment 409, 91405 Orsay, France

²Laboratoire de Microbiologie Fondamentale et Pathogénicité, UMR-CNRS 5234, Université de Bordeaux, 146 rue Léo Saignat, 33076 Bordeaux, France

*Address for correspondence: Laboratoire de Génomique et Biodiversité Microbienne des Biofilms, Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Univ. Paris-Sud, Bâtiment 409, 91405 Orsay, France ; Tel : +33(0)169154612 ; Fax : +33(0)169157808 ; E-mail : michael.dubow@u-psud.fr

Key Words: Deserts, Bacteria, Biodiversity, 16s rDNA, Pyrosequencing

Running Head: Bacterial communities of the deserts in Utah

Abstract

Deserts in Utah has geographic features that resemble the planet Mars, characterized by red-colored hills, soils and sandstones. In this study, we examined the bacterial diversity of surface soil samples from deserts in Utah using pyrosequencing of PCR amplified bacterial 16S rDNA genes. The sample sites cover the Goblin Valley State Park and nearby regions on the Colorado Plateau. We also examined physicochemical parameters of the soil samples to investigate any possible correlations between bacterial community structure and environmental drivers. The predominant phyla of the samples belong to members of the *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Gemmatimonadetes*. The most abundant genera in our samples are *Cesiribacter*, *Lysobacter*, *Adhaeribacter*, *Microvirga* and *Pontibacter*. We found that the relative abundance of Alphaproteobacteria and Gemmatimonadetes are significantly correlated to the environmental factors, such as pH and concentration of organic matters.

Introduction

Deserts are regions of land that have less than 25 centimeters rainfall, and where there is reduced diversity of plants and animals compared to that formed in more temperate zones. Aridity has been a major feature of the Earth's surface, and arid lands count for the largest terrestrial biome (Pointing and Belnap 2012b). In these environments, stresses such as drought, temperature and radiation are believed to limit the scale of biodiversity (Ward 2009). Research concerning microbial colonization and dispersion in deserts has been performed to estimate the function of microbial communities from desert sand, which may play an important role in soil stability, nutrient cycles and environmental health. Microbial community composition and function in desert biomes have been shown to be remarkably different from other biomes (Fierer et al. 2012). However, the diversity of microbial communities and their functional structure in deserts are still not well studied, and there are many unanswered questions regarding their biology, physiology and ecology (Pointing and Belnap 2012a). With the development of high-throughput sequence and analysis technologies, more work is being done to fully understanding the microorganisms their residing in deserts (van Belkum et al. 2001).

Amplification and analysis of 16S rRNA genes have been broadly used as a culture-independent method for documentation of the evolutionary history and taxonomic assignment of individual organisms, as well as in characterization of microbial communities (Head et al. 1998). More recently, meta-barcode methods are broadly used, which is an amplicon-based approach, based on PCR-targeted sequencing of selected hypervariable regions of the 16S rRNA gene (Bruno et al. 2015).

There are a diverse number of relatively small deserts stretching from southeastern California to western Texas, and from Nevada and Utah to the Mexican states of

Sonora, Chihuahua, and Coahuila and much of the peninsula of Baja California (Wickens 1998). The boundary between cold and warm deserts lies across southern Nevada and Utah. Deserts in Utah has geographic features that resemble Mars, relevant for geology and astrobiology studies (Chan et al. 1998). These features include extensive wind erosion, moisture deficits, absence of vegetation, and high UV radiation among others etc. (Makhalanyane et al. 2015). Also, the iron oxide prevalent on the surface of the Mars is similar to the deserts in Utah, characterized by red-colored hills, soils and sandstones. The Mars Society decided to set up the Mars Desert Research Station (MDRS) in the area as a Mars analog for such reasons (<http://mdrs.marssociety.org/>).

Bacteria constitute the largest portion of the biodiversity in soils, and play an important role in maintaining soil processes, which affect the functioning of terrestrial ecosystems (Epp et al. 2012). In this study, we examined the bacterial diversity research of surface soil samples from deserts on the Colorado Plateau (**Fig. 1**). The samples were collected in Goblin Valley (GV), the Little Wild Horse Canyon (LWH) located at south of Goblin Valley, south of the Utah State Route 24 (SR) and Temple Junction road (TJR) near the region of the Goblin Valley State Park, and 200 km southwest of the park near the Muddy Creek road (MCR).

Goblin Valley is a State Park in the USA, and lies on the southeast of the east edge of the San Rafael Swell, which is part of the Colorado Plateau physiographic region. The park is characterized by the presence of hoodoo rocks, referred to as "goblins". Those rock sculptures result from the weathering of Entrada Sandstone. The Entrada Sandstone was deposited in the Jurassic Period around 170 million years ago, and consists of debris eroded from former highlands and redeposited as alternating layers of sandstone, siltstone and shale (Milligan 1999). The distinct shape of these rocks comes from an erosion-resistant layer of hard rock atop softer sandstone which has eroded more quickly. The average daytime temperature in summer is between 32 °C

and 41 °C, and in winter daytime temperatures are above freezing most days but can drop to as low as -12 °C at night with occasional snow. The average annual precipitation is less than 20 cm (<http://www.usgs.gov/>, Precipitation of the Individual States and of the Conterminous States). During the summer, monsoons can arrive and increase precipitation in some regions between July and mid-September, and consist of at least 50% of the annual precipitation during this period.

The richness and patterns of microbial diversity in soils can be effected by many different environmental factors. Studies of microbial biogeography using metagenomics can often provide key insights into the physiologies, environmental tolerances, and ecological strategies of microbial taxa, particularly those difficult-to-culture taxa that often dominate in natural environments (Lauber et al. 2009). The bacterial community in desert environments can be regarded as a target for fundamental research on ecological and evolutionary processes, as the bacterial diversity in deserts have been found to be more rich than earlier expected (Cary et al. 2010). In this study, we demonstrated the bacterial diversity and community structures of surface soil in the Colorado Plateau in the Utah State using pyrosequencing of 16S rRNA amplicons. We built our pipeline for the analysis of 16S rRNA pyrosequencing data by combining several existing tools of metagenomics. We also examined correlations between certain environmental factors and bacterial diversity in the two deserts.

Materials and Methods

Sampling. To perform the bacterial diversity analysis of surface soils in the San Rafael Swell in Utah (USA), we recovered 18 samples of the soil (sand and soft rock particles) in 5 different regions: Dry river bed in Little Wild Horse Canyon, Goblin Valley State Park, near State Road 23 / I 70 Road, near Muddy Crack Road. All samples were taken at least 91 meters (100 yards) apart. The sampling GPS sites and description are shown in **Table 1**. The samples were collected by scooping surface sand into 50 ml sterile polyethylene conical centrifuge tubes, in different sites during September 2011. After collection the samples were treated as previously described (An *et al.* 2013). To perform the analyses of selected physicochemical parameters of the soil samples, the soil from each sites was pooled and sent for analyses, using standard methods, by the Laboratoire d'Analyses de Sols (INRA-ARAAS, France).

Sample Preparation. The total DNA was extracted from each sample using the protocol of An *et al.* (2013). An aliquot of extracted DNA was adjusted to a final DNA concentration of 15 ng/ μ l in 1/10 TE buffer (1 mM Tris pH 8; 0.1 mM EDTA) using a NanoVue spectrophotometer (GE Healthcare, Little Chalfont, Buckinghamshire, UK), and the concentration verified by ethidium bromide fluorescence after electrophoresis through a 1% agarose gel in TAE buffer (2 mM Tris-acetate pH 8; 5 mM Na-EDTA). PCR reactions were performed in 25 μ l reaction volumes. Each reaction contained one of two different thermostable DNA polymerases and their corresponding reaction buffers, 200 μ M of each dNTP, 0.5 μ M of each primer and 1 to 10 ng of extracted DNA. The 16S rRNA genes were amplified using the universal bacterial primers for pyrosequencing and covering hypervariable regions V1-V3: primer 27F (A adaptor + GAGTTTGATCMTGGCTCAG) and primer 518R (B adaptor + Mid + WTTACCGCGGCTGCTGG), where A and B represent the adaptors using the 454 Roche FLX Titanium pyrosequencing reaction platform. The Mid sequences are eight nucleotide tags designed for sample identification barcoding according to the 454

protocol. PCR amplification conditions were adapted for the use of two different thermostable DNA polymerases: A) Phusion High-Fidelity DNA Polymerase (Finnzymes, Finland): 98°C for 2 min, followed by 28 cycles of 98°C for 30 secs, 54°C for 20 secs and 72°C for 15 secs, and a final elongation step at 72°C for 5 min; B) Pfu DNA Polymerase (Fermentas, Canada): 95°C for 2 min, followed by 30 cycles of 95°C for 30 secs, 48°C for 30 secs and 72°C for 1 min, and a final elongation step at 72°C for 5 min. Each DNA sample was subjected to 3-5 different PCR reactions per DNA polymerase to minimize PCR bias. The PCR products were pooled and subjected to electrophoresis through a 1% agarose gel in TAE buffer. After electrophoresis and visualization of the PCR products by ethidium bromide staining and long wave UV light illumination, NucleoSpin Extract II kits (Macherey-Nagel, Germany) were used to purify the 16S rDNA PCR products. Then, 40 ng of PCR products from each sample were mixed for pyrosequencing, performed using a 454 Roche FLX Titanium Pyrosequencer (Microsynth AG, Switzerland).

DNA Sequence Data Processing. The raw DNA sequences were first assigned to each sample via their Mid tag using MOTHUR version 1.33 (Schloss et al. 2009), and reads were removed if at least one of the following criteria was met: (i) length less than 200 nt or longer than 600 nt, (ii) mismatch to the barcode sequences or more than one mismatch to the primer, and (iii) the presence of homopolymers of > 8 bp in length. Adaptor sequences were removed from the sequences using the “Cutadapt” tool (Martin 2011) implemented in the Galaxy server of the Institut de Génétique et Microbiologie (IGM) of the Université Paris-Sud (<http://galaxy.igmors.u-psud.fr>). Then, the sequences were checked for quality scores by ConDeTri version 2.2 (Smeds and Kunstner 2011), using the criteria that 80% of the nucleotides in a sequence have quality scores > 25. We used UCHIME (Edgar et al. 2011), with reference database Greengenes version 2013_May, and Decipher (through web tools available at <http://decipher.cee.wisc.edu/>) to detect chimera sequences (Wright et al. 2012). Sequences detected as chimeras by both programs were removed from the data sets.

The raw sequences have been deposited in the GenBank short-read archive (SRA), under the accession number SRP063276.

Taxonomy assignments of the remaining 16S rDNA reads were conducted using SILVAngs classifier (online server) with a similarity threshold of 90, using the Silva database release 123 (Cole et al. 2014; Yilmaz et al. 2014). Sequences classified as Chloroplast, or that could not be classified as belonging to the Bacteria Kingdom, were removed. Diversity analyses were performed using the software package MOTHUR. Normalization of 16S rDNA sequences based on the copy number of different species were performed using Tax4Fun package in R (Aßhauer et al. 2015).

Statistical Analysis. The clean reads were clustered into operational taxonomic units (OTUs) using UPARSE at a cutoff value of 97% sequence identity (Edgar 2010; Edgar 2013). The Chao1 and Shannon indices were calculated to estimate taxon richness and diversity (Schloss and Handelsman 2008). Bray-Curtis distance was used for calculating the distances of multiple samples. β -diversity of samples Principal coordinate analysis (PCoA) was conducted by R to group the microbial communities of different samples. The beta_diversity.py command was used to estimate the beta diversity of bacterial community using Qiime version 1.80 (Caporaso et al. 2010). Pearson correlations between relative abundances of sequences classified on specific phyla of different sample sites and environmental parameters were performed using SPSS version 22.

Results

Chemical properties of the sand samples

The chemical properties of the sand samples are shown in **Table 2**. For all the sample sites, the pH values of the sand are above 8, indicating an alkaline soil environment, which is normal for hot desert sand samples (Ward 2009). The pH values of the soil at the site of MCR and SRA are 10.1, higher than the other sites in the sample region (pH = 8.2 to 8.8). The chemical component results shows that the concentration of sodium (Na), potassium (K), iron (Fe) and aluminum (Al) were relatively higher in the two sites (MCR and SRA), which present a similar tendency for the pH value in the different sites. These results may indicate that the high concentrations of these metal ions, which come from decomposition of alkaline salt in the soil, contribute or lead to a high alkaline soil environment. On contrast, the concentration of organic materials and organic carbon were relatively lower in the two sites (MCR and SRA) and higher in TJR, which seems to inversely correspond to the concentration of mineral elements (Na, K, Fe, Al).

Sequence cleaning

We obtained 483364 sequences of raw reads in total from different groups of samples by 454 Roche Pyrosequencing (Microsynth AG, Switzerland), with an average length of 480 nt. After bioinformatic cleaning with Mothur, approximately 95% of the sequences remained (460475 reads). The sequences were further filtered for quality and examined for chimeric sequences, leaving 65% of the total reads (312358 reads). The average length of the sequences after processing were approximately 400 nt. The number of reads for each sample are presented in **Table 3**.

Classification of sequences

In total, 305,167 sequences from the 18 samples are classified into the bacteria domain (sequences classified into Chloroplast are removed from our data). The

sequences of were classified into 6 taxonomic levels by SILVAngs, including 37 phyla, 136 classes, 136 orders, 254 families and 565 genera. Sequences (97%) are classified into 37 phyla, with 13 common phyla for all the samples (**Fig. 1a**). The predominant phyla of the samples represent members of the *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Gemmatimonadetes*, at 28.3%, 25.7%, 19.7%, and 6.5%, as the respective averages for the 18 samples, followed by members of the *Acidobacteria* (3.9%), *Chloroflexi* (3.8%), *Firmicutes* (2.6%), *Armatimonadetes* (2.3%), *Denococcus-Thermus* (1.9%), *TM7* (1.3%), *Saccharibacteria* (1.3%), *Planctomycetes* (1.2%), *Cyanobacteria* (0.9%), and *Verrucomicrobia* (0.5%). Other phyla with less than 0.5% of bacterial sequences in total, are presented together as less common phyla.

The sequences assigned to the *Proteobacteria* phylum, representing 28.3% of the total, fell into the *Alpha*-, *Beta*-, and *Gamma*- three major class. We found that distributions of sequences belonging to the group of *Alphaproteobacteria* are significant different across sites ($P = 0.046$), with highest percentage in TJR (23.8% on average) and the lowest percentage in the region SR (5.9% on average). *Proteobacteria* are the most abundant in TJR_3 and TJR_4 (56.5% and 61.2% respectively), but the primary family in the two sites are different, with *Methylobacteriaceae* from *Alphaproteobacteria* dominant in TJR_3 (27.8%) and *Xanthomonadaceae* from *Gammaproteobacteria* dominant in TJR_4 (30.4%). Within the *Betaproteobacteria* class, members of the *Burkholderiales* family are the most abundant in all our samples.

The *Actinobacteria* was the second most abundant phylum found in our data, accounting for 25.7% of the total sequences. Three dominant classes within this phylum in our samples belong to the *Frankiales*, *Micrococcales* and *Acidimicrobiales* class. It has been reported that sequences assigned to *Actinobacteria* are the most abundant (>70%) in the soil sample of the Atacama Desert (Drees et al. 2006). Members of the

Actinobacteria are often found to prevail in desert soil (de la Torre et al. 2003; Neilson et al. 2012), such as *Rubrobacter*, *Arthrobacter*, and *Streptomyces*. However, many taxa belonging to *Actinobacteria* isolated from deserts soils appear to be new species (Mayilraj et al. 2006; Yung et al. 2007; Qin et al. 2009; Makhalanyane et al. 2015). In our study, we found a large proportion (> 50%) of sequences assigned to this phylum cannot be classified to genus level within the order *Frankiales* and *Acidimicrobiales*.

The *Bacteroidetes* was the third most abundant phylum with 19.7% of total sequences, which are commonly present in desert soils. Members of this phylum show optimum growth at high pH values, which is consistent with the alkaline pH seen in most deserts (Lauber et al. 2009). Members from four family were predominant groups in our samples: *Cytophagaceae*, *Flammeovirgaceae*, *Chitinophagaceae* and *Flavobacteriaceae*. *Bacteroidetes* are most abundant in MCR_3, presenting 64.3% of total sequences. *Firmicutes* are often observed to be more abundant in extreme environments, especially in hot deserts (Andrew et al. 2012; Marasco et al. 2012; Crits-Christoph et al. 2013). An *et al.* (2013) reported that sequences classified to the phylum Firmicutes account for > 60% of total sequences in a study of surface soils in the Gobi and Takalamaken deserts. We summarized the most abundant sequences classified at class level in **Fig. 2**.

Genus

Within the 565 genera identified among the sequences, 242 genera belong to the phylum *Proteobacteria*, 125 genera belong to the phylum *Actinobacteria*, 24 genera belong to the phylum *Bacteroidetes*, and 9 genera belong to the phylum *Acidobacteria*. The most abundant genera in sample sites are shown in **Table 4**. We found sequences classified to some genera can be predominant and counting for as much as more than 10% of total sequences in some samples. *Microvirga* are most abundant in TJR_3, representing 32.2% of the total sequences. Members of this genus have been reported to be thermophile and members have been isolated from a semi-arid site in Brazil

(Radl et al. 2014) and thermal aquifer (Radl et al. 2014). Sequences belonging to the *Cesiribacter* genus were predominant in MCR_3 and SRB_2, accounting for 30.8% and 19.9% of total sequences, respectively, in the two samples. Strains of this genus have been isolated from desert sand in China (Liu et al. 2012). Other predominant genera in a sample with relative abundance > 10% are *Lysobacter* (25.5% in TJR_4), *Salinimicrobium* (12.8% in TJR_2) and *Achromobacter* (11.7% in TJR_4).

Normalization of sequences

To study the real population of bacteria, sequences were normalized base on 16S rRNA gene copy number of different species by Tax4fun. We found that the structure of bacterial community shifts at all the taxonomic level. At the phylum level (**Fig. 1b**), the relative abundance of each phylum generally reduced since sequences that could not be aligned to the reference database with 16S rRNA copy number will be assigned to unclassified sequences (15.5% of total sequences at phylum level). However, relative abundance of the Actinobacteria increased from 25.7% before normalization to 35.3% after, which indicate the population of the Actinobacteria was underestimated by the method of simply summary of taxonomic classification based on number of sequences.

OTU-based analysis

The clean sequences were clustered into OTUs at 97% similarity, excluding the unclassified sequences at the phylum level and sequences classified as Chloroplast, Eukaryote or Mitochondria. In total, 9908 OTUs were observed in all the samples, ranging from 778 in the Drb_1 sample to 3595 in the Tjr_1 sample. The OTU numbers observed in each sample were shown in **Table 3**. Eleven core taxons (most abundant OTUs with more than 1% of total sequences) were found, comprising 37% of the population of all the samples. The number of singleton (OTUs with only one sequence) is 3926, comprising 37% of the total OTU numbers. Only 49 OTUs were observed in all the 18 samples. For different sample regions, the numbers of common OTUs within a

given sample site were: 599 for DRB (Drb_1 and Drb_2), 323 for GV (GV_1 to GV_5), 170 for MCR (Mcr_1 to Mcr_4), 323 for SR (SR_1 to SR_3), and 238 for TJR (Tjr_1 to Tjr_4).

The distance of the microbial communities between samples were calculated based on OTU abundance using a Bray-Curtis measure, and Principal Component Analysis (PCoA) was used to perceive the distances of the 18 samples (**Fig. 3**). The distance between different sites based on the OTU relative abundance are illustrated by a UPGMA tree (**Fig. 4**). The tree was generated by merging sequences in each site and normalized to the same number of reads (12501) in Qiime. The sample sites of GV and LWH are more close in distance compared with the other sites, and they are also close in geographic distance. The microbial population of the TJR site has the largest distance from the other sites, and these may correspond to the difference in environment parameters, as the concentration of organic carbon and mineral salts are significantly different at the TJR site ($P < 0.05$).

Discussion

The deserts in Utah present a range of arid conditions, with high salinity and low concentration of carbon, providing unique environment for microbial communities, and a combination of geochemistry and biodiversity data can be applied to study the environmental factors that may shape the community structure of soils (Crits-Christoph et al. 2013). In this study, we used pyrosequencing amplicons of 16S rRNA genes to analyse the bacterial diversity and communities in the deserts in the Utah State in the USA. We also examined some physicochemical parameters of sample sites to investigate the correlations between bacterial community structure and environmental drivers.

Fierer *et al.* (2012) reported that soils close to neutral had the highest bacterial diversity levels, compared with very basic (deserts) or acidic soils (rainforests and Arctic tundra) using metagenomic sequencing of total DNA extracted from soils from a wide range of ecosystems. The alkaline conditions may select for taxa most adapted to alkaline growth conditions, and the contribution of pH to bacterial diversity may thus be limited (Finkel et al. 2012). In this study, we used the Pearson's correlation analysis to illustrate the correlation between bacterial diversity and some environmental factors that may affect the bacterial community in soils (**Table 5**). We found that the most significant correlation with bacterial richness (Chao 1) is the concentration of organic carbon ($P < 0.01$), and the correlation between richness and the concentration of mineral elements (Na, K, Al, Fe) are also significant ($P < 0.05$). The two significant correlation was with the Shannon diversity index and the pH value of soils and the concentration of sodium ($P < 0.05$). The concentration of Ca, P, Mg and Mn are not found to be significantly correlated with bacterial diversity. However, concerning the limited number of sites in our study ($n=5$), the correlation between environmental factors and bacterial diversity should be comprehended cautiously.

Some bacterial community studies of surface soils in deserts found that geographical location plays an important role in the microbial communities (Finkel et al. 2011; Finkel et al. 2012; Nemergut et al. 2013). Beta-diversity of bacterial community structure among different samples were performed to test the hypothesis that variability of microbial communities found in desert soils in Utah across different sites is more than the variability within sites. We found the bacterial community in our samples shifts with different sites (**Fig.1**). However, in the SR, LWH and GV sites, the distribution of bacterial communities are more closely related, and different from MCR and TJR. Although, the MCR site is the most distant geographically from the other sample sites, we did not find a significant difference in the bacterial community structure compared with samples from other sites. The assessment of environmental factors in our study suggests that the concentration of mineral components appears to present a strong correlation with the bacterial community structure (**Table 6**). In the TJR site, the concentration of organic carbon and mineral salts are significantly different from the other sites ($P < 0.05$), and displays the largest dissimilarity on the OTUs distribution and relative abundance of some taxonomic groups (*Alphaproteobacteria*, *Acidimicrobiia*, *Gemmatimonadetes* and *Deinococci*, with P value < 0.05).

Desert environments is a selective environment that has been reported to present bacterial specificity even at the higher taxonomic levels (phylum to family) (Tamames et al. 2010). Salinity is a very important factor in shaping prokaryotic diversity (Li et al. 2013; Canfora et al. 2014; Geyer et al. 2014). In this study, we found that the relative abundance of sequences assigned to the phyla *Proteobacteria* and *Gemmatimonadetes* can be significantly related to certain environmental parameters. The relations between environmental factors and *Proteobacteria* relative abundances were performed using a Pearson's correlation analysis (**Table 6**). We found that the relative abundance of *Proteobacteria* was inversely correlated ($P < 0.05$) with soil pH value ($r^2 = -0.523$, with pH ranging from 8.2 to 10.1), as well as the

concentrations of Na, K, Fe, and Al. We also found that the relative abundance of sequences of this phylum was positively correlated with concentration of organic carbon and organic materials ($P < 0.05$). *Proteobacteria* have been previously reported to be abundant in desert soils (Rasuk et al. 2014; Li et al. 2015).

Sequences related to *Gemmatimonadetes* are commonly found in soils of various environments, including hot desert (Azua-Bustos et al. 2012), pasture (Chim Chan et al. 2008), crop agriculture (Montecchia et al. 2015), forests (Li et al. 2014) and freshwater sediments (Zhang et al. 2015), although only six cultured isolates have been reported (Fawaz 2013). DeBruyn *et al.* (2011) reported that *Gemmatimonadetes* relative abundances was inversely correlated to moisture in soils, and many *Gemmatimonadetes* phylotypes have higher relative abundances in semiarid and arid soils and deserts (DeBruyn et al. 2011). This also has been reinforced by a study on the distribution of the *Gemmatimonadetes* in agricultural soils, in that members of this phylum prefer dryer soils and tend to be more dependent on moisture availability. In our study, we found *Gemmatimonadetes* relative abundance was significantly different among sample sites ($P=0.017$), with the highest concentration in SR and MCR and lowest in TJR. Based on the analysis of environmental factors, SR and MCR sites present to be most “extreme” microenvironment compared with other sites, with the lowest concentration of organic materials and highest concentration of mineral elements, as well as the highest pH value (10.1). The relations between environmental factors and *Gemmatimonadetes* relative abundances were performed using the Pearsons correlation analysis (**Table 6**). Earlier studies have showed that relative abundances of *Gemmatimonadetes* are higher in soils with neutral pH versus these in acidic soils (28, 32, 48). We found that *Gemmatimonadetes* was significantly correlated ($P < 0.001$) with soil pH ($r^2=0.854$), with pH ranges from 8.2 to 10.1. We also found that the relative abundance of sequences of this phylum was significant correlated with the concentration of Na, Fe, K and Al, as well as the concentration of organic carbon and organic materials ($P < 0.05$). These results indicate that members

of the *Gemmatimonadetes* may present as potential members of biomarkers in soil communities of arid environment characterized by relative higher salinity and lower organic matter.

The percentage of genus level classifiable sequences varies among the phyla we found. For example, 91.4% sequences in the *Bacteroidetes* phylum could be classified to the genus level, while only 37.0% of the *Actinobacteria* phylum are able to be classified to this level. Also, sequences related to the *Acidobacteria* phylum can only rarely be classified to the genus level in our data. These indicated that sequences related to the phyla *Actinobacteria* and *Acidobacteria* in desert soil may present more as unknown species. Direito. *et al.* (2011) studied the microbial diversity near the MDRS in the Utah based on the DGGE profiles of ribosomal RNA genes, and found 239 clones of bacteria (Direito et al. 2011). All the phyla observed in their study have also been detected in our research and we recovered much more different bacterial taxa by using the pyrosequencing method. Sequences assigned to genera that are related to several types of extremophiles or isolated from desert-like environments were observed in all our samples, including radio-resistant, halophilic, thermophilic and endolithic bacteria. *Pontibacter* of the *Bacteroidetes* phylum are found to be abundant in deserts (Zhou et al. 2007; Subhash et al. 2014), and may possess unique abilities to adapt to desert environments (Makhalanyane et al. 2015). In the samples from the desert in Utah we observed that sequences belonging to this genus are abundant (2.2% of total sequences), and most abundant in the sample SRB_2 (7.7% of total sequences). Sequences classifiable to the genera *Rubrobacter*, *Truepera*, *Deinococcus* were reported to be radiotolerant (Cox and Battista 2005), and we found sequences classified to these genera to be abundant in the samples in deserts in Utah. Other genera that have been reported to be extremophiles were also observed, including: halophile bacteria such as *Salegentibacter*, *Halomonas* and *Salinimicrobium*, *Actinobacteria* (Rossello-Mora et al. 2003; Oren 2015); thermophilic bacteria such as *Thermoleophilum* and *Thermosporothrix* (Andrade et al. 1999); and *Rhodobacter*

known as endolithic bacteria (Stivaletta et al. 2010). Another group of bacteria that are found in these desert environments is *Pseudomonas*, and the members of the *Pseudomonas* family are believed to play a protective role for bacterial communities in many extreme environments because their ability for biofilm formation (Drenkard and Ausubel 2002; Selenska-Pobell et al. 2002).

In this study, we used Pyrosequencing of PCR amplified bacterial 16S rDNA genes to reveal the bacterial diversity and community structure in surface soil samples from the deserts in Utah, as well as their possible correlations with selected environmental factors. The research of bacterial diversity of soils in the desert in Utah presents a unique opportunity to understand bacterial communities in arid soils, characterized by high salinity and a low concentration of organic matter.

Acknowledgements

We thank all the members of the Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB) of the Institute for Integrative Biology of the Cell (I2BC) for interesting discussions, comments and suggestions. This work was supported by the Centre National de la Recherche Scientifique (CNRS), France.

References

- Abhauer KP, Wemheuer B, Daniel R, Meinicke P (2015) Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31:2882-2884. doi:10.1093/bioinformatics/btv287
- An S, Couteau C, Luo F, Neveu J, DuBow MS (2013) Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts. *Microb Ecol* 66:850-860. doi:10.1007/s00248-013-0276-2
- Andrade CMMC, Pereira Jr. N, Antranikian G (1999) Extremely thermophilic microorganisms and their polymer-hydrolytic enzymes. *Revista de Microbiologia* 30:287-298
- Andrew DR, Fitak RR, Munguia-Vega A, Racolta A, Martinson VG, Dontsova K (2012) Abiotic factors shape microbial diversity in Sonoran Desert soils. *Appl Environ Microbiol* 78:7527-7537. doi:10.1128/AEM.01459-12
- Azua-Bustos A, Urrejola C, Vicuña R (2012) Life at the dry edge: Microorganisms of the Atacama Desert. *FEBS Letters* 586:2939-2945. doi:http://dx.doi.org/10.1016/j.febslet.2012.07.025
- Bruno F, Marinella M, Santamaria M (2015) e-DNA Meta-Barcoding: From NGS Raw Data to Taxonomic Profiling. In: Picardi E (ed) *RNA Bioinformatics*. vol 1269. *Methods in Molecular Biology*. Springer New York. pp 257-278. doi:10.1007/978-1-4939-2291-8_16
- Canfora L, Bacci G, Pinzari F, Lo Papa G, Dazzi C, Benedetti A (2014) Salinity and bacterial diversity: to what extent does the concentration of salt affect the bacterial community in a saline soil? *PLoS One* 9:e106662. doi:10.1371/journal.pone.0106662
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG et al. (2010) QIIME allows analysis of high-throughput

- community sequencing data. *Nat Methods* 7:335-336.
doi:10.1038/nmeth.f.303
- Cary SC, McDonald IR, Barrett JE, Cowan DA (2010) On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat Rev Micro* 8:129-138
- Chan MA, Nicoll K, Ormo J, Okubo C, Komatsu G (1998) Great Basin–Mojave Desert Region vol 2. U.S. Department of the Interior, U.S. Geological Survey, Reston
- Chim Chan O, Casper P, Sha LQ, Feng ZL, Fu Y, Yang XD, Ulrich A, Zou XM (2008) Vegetation cover of forest, shrub and pasture strongly influences soil bacterial community structure as revealed by 16S rRNA gene T-RFLP analysis. vol 64. vol 3. doi:10.1111/j.1574-6941.2008.00488.x
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A et al. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42:D633-642. doi:10.1093/nar/gkt1244
- Cox MM, Battista JR (2005) *Deinococcus radiodurans* the consummate survivor. *Nat Rev Micro* 3:882-892
- Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, Jedynek B, McKay CP, DiRuggiero J (2013) Colonization patterns of soil microbial communities in the Atacama Desert. *Microbiome* 1:28-28. doi:10.1186/2049-2618-1-28
- de la Torre JR, Goebel BM, Friedmann EI, Pace NR (2003) Microbial Diversity of Cryptoendolithic Communities from the McMurdo Dry Valleys, Antarctica. *Applied and Environmental Microbiology* 69:3858-3867. doi:10.1128/AEM.69.7.3858-3867.2003
- DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M (2011) Global Biogeography and Quantitative Seasonal Dynamics of Gemmatimonadetes in Soil. *Applied and Environmental Microbiology* 77:6295-6300.

doi:10.1128/AEM.05005-11

Direito SOL, Ehrenfreund P, Marees A, Staats M, Foing B, Röling WFM (2011) A wide variety of putative extremophiles and large beta-diversity at the Mars Desert Research Station (Utah). *International Journal of Astrobiology* 10:191--207

Drees KP, Neilson JW, Betancourt JL, Quade J, Henderson DA, Pryor BM, Maier RM (2006) Bacterial Community Structure in the Hyperarid Core of the Atacama Desert, Chile. *Applied and Environmental Microbiology* 72:7902-7908.
doi:10.1128/AEM.01305-06

Drenkard E, Ausubel FM (2002) *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature* 416:740-743

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460-2461. doi:10.1093/bioinformatics/btq461

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996-998. doi:10.1038/nmeth.2604

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194-2200.
doi:10.1093/bioinformatics/btr381

Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, Riaz T, Erseus C, Gusarov VI et al. (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol Ecol* 21:1821-1833.
doi:10.1111/j.1365-294X.2012.05537.x

Fawaz MN (2013) *Revealing the Ecological Role of Gemmatimonadetes Through Cultivation and Molecular Analysis of Agricultural Soils*. University of Tennessee

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA et al. (2012) Cross-biome metagenomic analyses of soil microbial

- communities and their functional attributes. *Proc Natl Acad Sci U S A* 109:21390-21395. doi:10.1073/pnas.1215210110
- Finkel OM, Burch AY, Elad T, Huse SM, Lindow SE, Post AF, Belkin S (2012) Distance-Decay Relationships Partially Determine Diversity Patterns of Phyllosphere Bacteria on Tamrix Trees across the Sonoran Desert. *Applied and Environmental Microbiology* 78:6187-6193. doi:10.1128/AEM.00888-12
- Finkel OM, Burch AY, Lindow SE, Post AF, Belkin S (2011) Geographical Location Determines the Population Structure in Phyllosphere Microbial Communities of a Salt-Excreting Desert Tree. *Applied and Environmental Microbiology* 77:7647-7655. doi:10.1128/AEM.05565-11
- Geyer KM, Altrichter AE, Takacs-Vesbach CD, Van Horn DJ, Gooseff MN, Barrett JE (2014) Bacterial community composition of divergent soil habitats in a polar desert. *FEMS Microbiology Ecology* 89:490-494. doi:10.1111/1574-6941.12306
- Head IM, Saunders JR, Pickup RW (1998) Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microb Ecol* 35:1-21
- 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. *Applied and Environmental Microbiology* 75:5111-5120. doi:10.1128/aem.00335-09
- Li H, Ye D, Wang X, Settles M, Wang J, Hao Z, Zhou L, Dong P et al. (2014) Soil bacterial communities of different natural forest types in Northeast China. *Plant and Soil* 383:203-216. doi:10.1007/s11104-014-2165-y
- Li K, Bai Z, Zhang H (2015) Community succession of bacteria and eukaryotes in dune ecosystems of Gurbantünggüt Desert, Northwest China. *Extremophiles*

19:171-181. doi:10.1007/s00792-014-0696-z

Li K, Liu R, Zhang H, Yun J (2013) The Diversity and Abundance of Bacteria and Oxygenic Phototrophs in Saline Biological Desert Crusts in Xinjiang, Northwest China. *Microbial Ecology* 66:40-48. doi:10.1007/s00248-012-0164-1

Liu M, Qi H, Luo X, Dai J, Peng F, Fang C (2012) *Cesiribacter roseus* sp. nov., a pink-pigmented bacterium isolated from desert sand. *Int J Syst Evol Microbiol* 62:96-99. doi:10.1099/ijs.0.028423-0

Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond J-B, Cowan DA (2015) Microbial ecology of hot desert edaphic systems. *FEMS Microbiology Reviews*. doi: <http://dx.doi.org/10.1093/femsre/fuu011>

Marasco R, Rolli E, Ettoumi B, Vigani G, Mapelli F, Borin S, Abou-Hadid AF, El-Behairy UA et al. (2012) A drought resistance-promoting microbiome is selected by root system under desert farming. *PLoS ONE* 7:e48479. doi:10.1371/journal.pone.0048479

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10-12. doi:10.14806/ej.17.1.200

Mayilraj S, Kroppenstedt RM, Suresh K, Saini HS (2006) *Kocuria himachalensis* sp. nov., an actinobacterium isolated from the Indian Himalayas. *International Journal of Systematic and Evolutionary Microbiology* 56:1971-1975. doi:doi:10.1099/ijs.0.63915-0

Montecchia MS, Tosi M, Soria MA, Vogrig JA, Sydorenko O, Correa OS (2015) Pyrosequencing reveals changes in soil bacterial communities after conversion of Yungas forests to agriculture. *PLoS One* 10:e0119426. doi:10.1371/journal.pone.0119426

Neilson J, Quade J, Ortiz M, Nelson W, Legatzki A, Tian F, LaComb M, Betancourt J et al. (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils

- of the Atacama Desert, Chile. *Extremophiles* 16:553-566.
doi:10.1007/s00792-012-0454-z
- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL et al. (2013) Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* : MMBR 77:342-356. doi:10.1128/MMBR.00051-12
- Oren A (2015) Halophilic microbial communities and their environments. *Curr Opin Biotechnol* 33:119-124. doi:10.1016/j.copbio.2015.02.005
- Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10:551-562. doi:10.1038/nrmicro2831
- Qin S, Li J, Chen H-H, Zhao G-Z, Zhu W-Y, Jiang C-L, Xu L-H, Li W-J (2009) Isolation, Diversity, and Antimicrobial Activity of Rare Actinobacteria from Medicinal Plants of Tropical Rain Forests in Xishuangbanna, China. *Applied and Environmental Microbiology* 75:6176-6186. doi:10.1128/AEM.01034-09
- Radl V, Simoes-Araujo JL, Leite J, Passos SR, Martins LM, Xavier GR, Rumjanek NG, Baldani JI et al. (2014) *Microvirga vignae* sp. nov., a root nodule symbiotic bacterium isolated from cowpea grown in semi-arid Brazil. *Int J Syst Evol Microbiol* 64:725-730. doi:10.1099/ijms.0.053082-0
- Rasuk M, Kurth D, Flores M, Contreras M, Novoa F, Poire D, Farias M (2014) Microbial Characterization of Microbial Ecosystems Associated to Evaporites Domes of Gypsum in Salar de Llamara in Atacama Desert. *Microbial Ecology* 68:483-494. doi:10.1007/s00248-014-0431-4
- Rossello-Mora R, Lee N, Anton J, Wagner M (2003) Substrate uptake in extremely halophilic microbial communities revealed by microautoradiography and fluorescence in situ hybridization. *Extremophiles* 7:409-413.
doi:10.1007/s00792-003-0336-5

- Schloss PD, Handelsman J (2008) A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics* 9:34. doi:10.1186/1471-2105-9-34
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537-7541. doi:10.1128/AEM.01541-09
- Selenska-Pobell S, Flemming K, Tzvetkova T, Raff J, Schnorpfeil M, Geißler A (2002) Bacterial communities in uranium mining waste piles and their interaction with heavy metals. In: Merkel B, Planer-Friedrich B, Wolkersdorfer C (eds) *Uranium in the Aquatic Environment*. Springer Berlin Heidelberg. pp 455-464. doi:10.1007/978-3-642-55668-5_53
- Smeds L, Kunstner A (2011) ConDeTri--a content dependent read trimmer for Illumina data. *PLoS One* 6:e26314. doi:10.1371/journal.pone.0026314
- Stivaletta N, López-García P, Boihem L, Millie DF, Barbieri R (2010) Biomarkers of Endolithic Communities within Gypsum Crusts (Southern Tunisia). *Geomicrobiology Journal* 27:101-110. doi:10.1080/01490450903410431
- Subhash Y, Sasikala C, Ramana Ch V (2014) *Pontibacter ruber* sp. nov. and *Pontibacter deserti* sp. nov., isolated from the desert. *Int J Syst Evol Microbiol* 64:1006-1011. doi:10.1099/ijs.0.058842-0
- Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A (2010) Environmental distribution of prokaryotic taxa. *BMC Microbiol* 10:85. doi:10.1186/1471-2180-10-85
- van Belkum A, Struelens M, de Visser A, Verbrugh H, Tibayrenc M (2001) Role of Genomic Typing in Taxonomy, Evolutionary Genetics, and Microbial Epidemiology. *Clinical Microbiology Reviews* 14:547-560.

doi:10.1128/CMR.14.3.547-560.2001

- Ward D (2009) *The Biology of Deserts*. Oxford University Press. Oxford, UK
- Wickens G (1998) Arid and Semi-arid Environments of the World. In: *Ecophysiology of Economic Plants in Arid and Semi-Arid Lands. Adaptations of Desert Organisms*. Springer Berlin Heidelberg. pp 5-15. doi:10.1007/978-3-662-03700-3_2
- Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* 78:717-725. doi:10.1128/AEM.06516-11
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J et al. (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 42:D643-648. doi:10.1093/nar/gkt1209
- Yung PT, Lester ED, Bearman G, Ponce A (2007) An automated front-end monitor for anthrax surveillance systems based on the rapid detection of airborne endospores. *Biotechnology and Bioengineering* 98:864-871. doi:10.1002/bit.21466
- Zhang J, Yang Y, Zhao L, Li Y, Xie S, Liu Y (2015) Distribution of sediment bacterial and archaeal communities in plateau freshwater lakes. *Applied Microbiology and Biotechnology* 99:3291-3302. doi:10.1007/s00253-014-6262-x
- Zhou Y, Wang X, Liu H, Zhang KY, Zhang YQ, Lai R, Li WJ (2007) *Pontibacter akesuensis* sp. nov., isolated from a desert soil in China. *Int J Syst Evol Microbiol* 57:321-325. doi:10.1099/ijs.0.64716-0

Tables

Table 1. Information of location of Sample sites.

Region	Location Description	Lat/Lon	Sample Name
LWH	Dry river bed in Little Wild Horse Canyon	38°34'32.18" N	LHW_1
		110°46'55.06" W	LHW_2
GVA	Goblin Valley State Park	38°34'30.73" N	GVA_1
		110°43'47.24" W	GVA_2
			GVA_3
GVB		38°34'12.96" N	GVB_1
		110°44'31.89" W	GVB_2
MCR	Near Muddy Creeck Road	37°19'57.49" N	MCR_1
		112°41'42.87" W	MCR_2
			MCR_3
			MCR_4
SRA	Near State Road 24 / I 70 road	38°49'51.46" N	SRA_1
		110°22'52.35" W	
SRB		38°52'53.72" N	SRB_1
		110°20'8.87" W	SRB_2
TJR	Near Temple Junction Road / Goblin Valley Road	38°39'38.18" N	TJR_1
		110°38'46.79" W	TJR_2
			TJR_3
			TJR_4

Table 2. The Physicochemical data of the five soil sample sites

	LWH	GVA	MCR_1	SRA	TJR
PH	8.2	8.84	10.1	10.1	8.23
Ca (g/100g)	11.6	4.14	4.29	0.59	0.32
Fe (g/100g)	1.49	0.85	1.81	1.85	0.2
Mg (g/100g)	2.42	0.91	1.53	0.96	0.11
P (g/100g)	0.17	0.08	0.14	0.1	0.02
Na (g/100g)	0.37	0.55	0.72	0.81	0.08
K (g/100g)	1.46	1.49	2.26	2.14	0.92
Mn (mg/kg)	511	272	359	172	29.3
Al (g/100g)	3.19	3.11	4.8	4.8	1.11
Organic materials (g/kg)	2.61	1.17	1.15	0.21	3.49
Organic carbon (g/kg)	1.51	0.68	0.67	0.12	11.7

Table 3. Number of Sequences and Diversity Index of the 18 soil samples

	Sequence numbers		Diversity indices		
	Raw data	After cleaning	OTUs	Chao1	Shannon
LWH_1	11632	7572	1191	1764	6.11
LWH_2	20221	13515	1570	2372	5.88
GVA_1	28382	18721	1247	1924	5.07
GVA_2	30479	19357	1541	2431	5.56
GVA_3	28744	18533	1694	2605	5.83
GVB_1	25342	17198	1063	1763	4.25
GVB_2	27282	17258	1638	2756	5.79
MCR_1	21955	14976	1459	2349	5.75
MCR_2	25839	15756	1305	2216	5.59
MCR_3	45693	31143	778	1307	4.34
MCR_4	19433	13165	1628	2684	6.08
SRA_1	20331	14028	851	1532	4.85
SRB_1	33407	22134	972	1665	4.59
SRB_2	37157	25574	1170	2089	4.72
TJR_1	25380	16546	3595	6669	7.18
TJR_2	28836	15976	1320	2552	5.42
TJR_3	19900	12750	2196	4709	4.94
TJR_4	33351	16549	2015	5383	4.40

Table 4. The most abundant sequences at the genus level of soil sample sites

	GVA	GVB	LWH	MCR	SRA	SRB	TJR
<i>Cesiribacter</i>	0.1%	0.6%	1.0%	6.4%	6.0%	7.4%	0.2%
<i>Lysobacter</i>	0.1%	0.2%	0.5%	0.3%	0.2%	0.2%	10.1%
<i>Pontibacter</i>	0.9%	1.2%	1.1%	3.0%	3.5%	5.9%	1.2%
<i>Adhaeribacter</i>	2.8%	1.6%	1.9%	3.1%	1.6%	2.0%	0.8%
<i>Microvirga</i>	0.2%	0.2%	1.0%	0.4%	0.6%	0.2%	7.5%
<i>Sphingomonas</i>	1.2%	1.3%	2.4%	2.0%	0.1%	0.4%	3.3%
<i>Segetibacter</i>	4.0%	2.0%	1.2%	2.3%	0.1%	0.5%	0.8%
<i>Flavisolibacter</i>	3.9%	1.3%	0.8%	1.8%	0.2%	0.3%	1.8%
<i>Massilia</i>	0.9%	3.2%	1.9%	1.4%	5.7%	0.6%	0.9%
<i>Cytophaga</i>	0.3%	0.2%	0.8%	3.1%	3.2%	2.2%	0.1%
<i>Bacillus</i>	1.2%	0.5%	0.9%	2.4%	1.1%	0.2%	1.2%
<i>Truepera</i>	0.5%	0.5%	1.9%	0.6%	4.2%	4.0%	0.2%
<i>Rubellimicrobium</i>	0.9%	0.7%	2.4%	0.9%	1.3%	1.2%	1.1%
<i>Arthrobacter</i>	0.6%	1.5%	1.9%	0.7%	3.3%	1.3%	0.8%
<i>Hymenobacter</i>	1.9%	1.8%	0.8%	1.2%	0.1%	1.2%	0.4%
<i>Blastocatella</i>	0.8%	0.3%	0.6%	1.2%	0.0%	0.1%	2.4%
<i>Salinimicrobium</i>	0.1%	0.1%	0.5%	0.5%	0.7%	0.0%	3.2%
<i>Stenotrophomonas</i>	3.3%	0.1%	0.3%	0.0%	0.0%	1.4%	0.8%
<i>Achromobacter</i>	0.0%	0.1%	0.3%	0.0%	0.0%	0.0%	3.4%
<i>Acinetobacter</i>	3.0%	0.2%	1.0%	0.1%	0.0%	0.9%	0.2%
<i>Rubrobacter</i>	1.8%	0.1%	0.4%	0.5%	0.9%	1.8%	0.3%
<i>Euzebya</i>	0.4%	0.5%	0.3%	1.0%	0.2%	2.4%	0.3%
<i>Deinococcus</i>	0.3%	0.3%	0.8%	1.8%	0.6%	0.8%	0.1%
<i>Flavobacterium</i>	0.2%	4.0%	0.5%	0.0%	0.0%	0.0%	0.6%
<i>Nocardioides</i>	0.6%	0.6%	1.3%	0.3%	1.1%	0.3%	0.7%
<i>Streptomyces</i>	0.3%	0.3%	1.1%	0.2%	6.6%	0.1%	0.1%
<i>Altererythrobacter</i>	0.2%	0.9%	1.5%	0.2%	2.0%	0.1%	0.4%
<i>Pseudomonas</i>	0.6%	1.5%	0.5%	0.1%	0.9%	0.3%	0.6%
<i>Blastococcus</i>	0.3%	0.2%	1.1%	0.3%	0.7%	0.5%	1.0%
<i>Solirubrobacter</i>	1.8%	0.2%	0.3%	0.3%	0.1%	0.3%	0.4%
<i>Gillisia</i>	0.0%	0.1%	0.3%	0.0%	0.0%	0.0%	1.8%
<i>Nibribacter</i>	0.1%	0.4%	0.4%	0.8%	0.5%	1.0%	0.1%
<i>Patulibacter</i>	1.2%	0.5%	0.3%	0.5%	0.0%	0.1%	0.2%
<i>Gemmatimonas</i>	0.7%	0.6%	0.4%	0.6%	0.1%	0.2%	0.3%
<i>Noviherbaspirillum</i>	0.3%	1.0%	0.9%	0.5%	0.1%	0.2%	0.2%
<i>Planomicrobium</i>	0.0%	0.1%	0.3%	0.1%	0.0%	0.1%	1.5%

Table 5. Pearson's correlations between OTU richness and environmental factors

		PH	Organic carbon	Organic materials	Na	Ca	Fe	Al	P	K	Mg	Mn
OTUs	R ²	-.781	.912*	.956**	-.966**	-.028	-.852*	-.930*	-.529	-.875*	-.366	-.340
	P	.060	.016	.005	.004	.482	.033	.011	.180	.026	.272	.288
Chao1	R ²	-.657	.972**	.815*	-.895*	-.356	-.938**	-.938**	-.794	-.865*	-.667	-.644
	P	.114	.003	.047	.020	.278	.009	.009	.054	.029	.110	.120
Shannon	R ²	-.843*	.547	.879*	-.837*	.449	-.648	-.755	-.102	-.757	.087	.175
	P	.036	.170	.025	.039	.224	.119	.070	.435	.069	.445	.389

* with P value < 0.05

** with P value < 0.01

Table 6. Pearson's correlation between relative abundance of sequences and related environmental factors.

		pH	Organic carbon	Organic materials	Na	Ca	Fe	Al	P	K	Mg	Mn
<i>Proteobacteria</i>	R^2	-.523*	.595*	.585*	-.623*	-.249	-.569*	-.627*	-.464	-.617*	-.390	-.423
	P	.049	.027	.029	.020	.231	.034	.020	.075	.022	.118	.098
<i>Gemmatimonadetes</i>	R^2	.854**	-.751**	-.896**	.899**	.057	.695**	.844**	.426	.856**	.293	.341
	P	<.001	.004	<.001	<.001	.434	.009	.001	.095	<.001	.191	.152

* with P value < 0.05

** with P value < 0.01

Figure Legends

Fig. 1. Relative abundance of bacterial 16S rDNA gene sequences from the samples at the Phylum level.

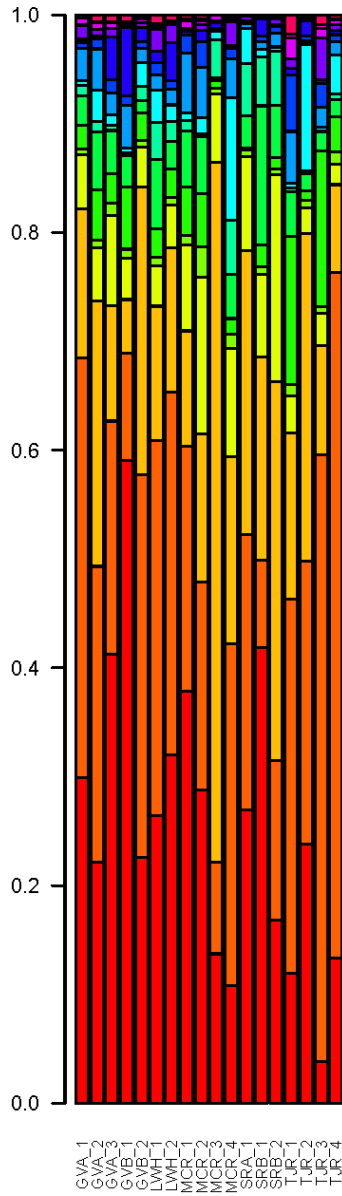
Fig. 2. Relative abundance of bacterial 16S rDNA gene sequences from the samples at the Class level.

Fig. 3. PCoA plot of distance of the microbial communities among the 18 samples based on OTU relative abundance using the Bray-Curtis measure.

Fig. 4. The distance between different sites based on the OTU relative abundance illustrated with a UPGMA tree.

Fig. 1

a Before normalization



b After normalization

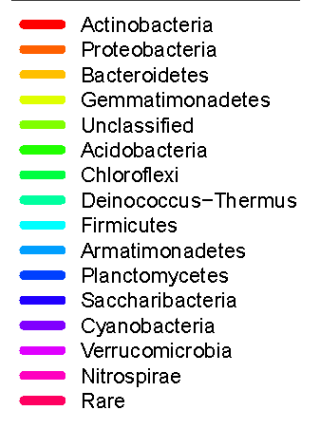
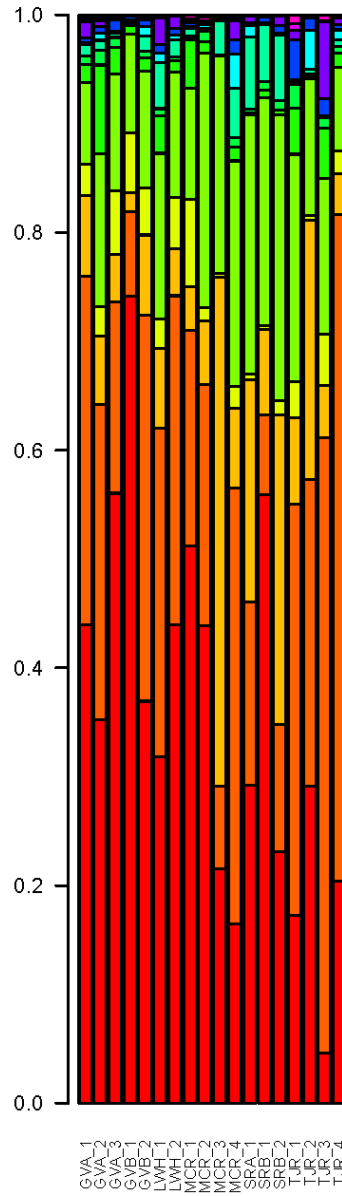
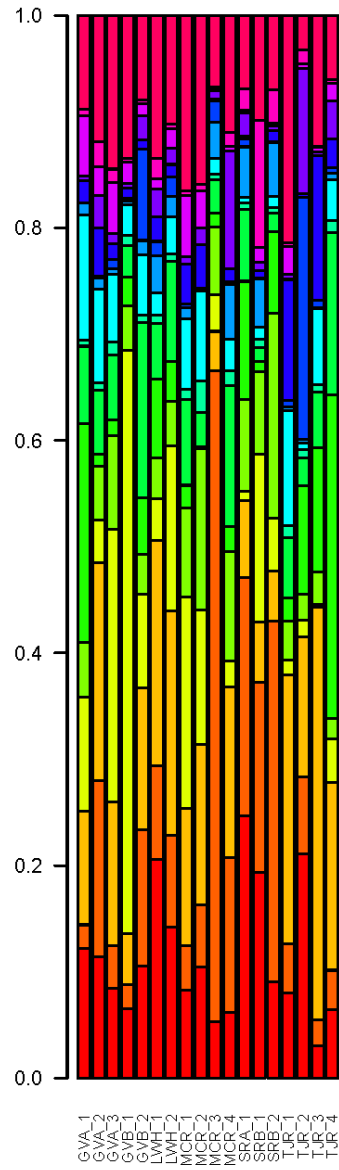


Fig 3.

a Before normalization



b After normalization

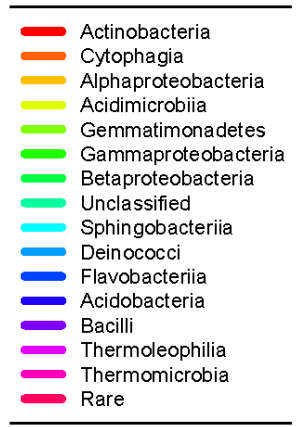
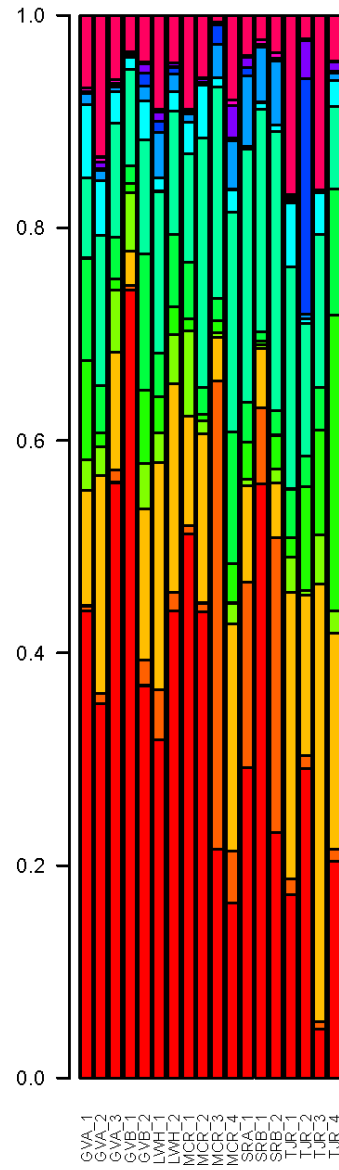


Fig 3.

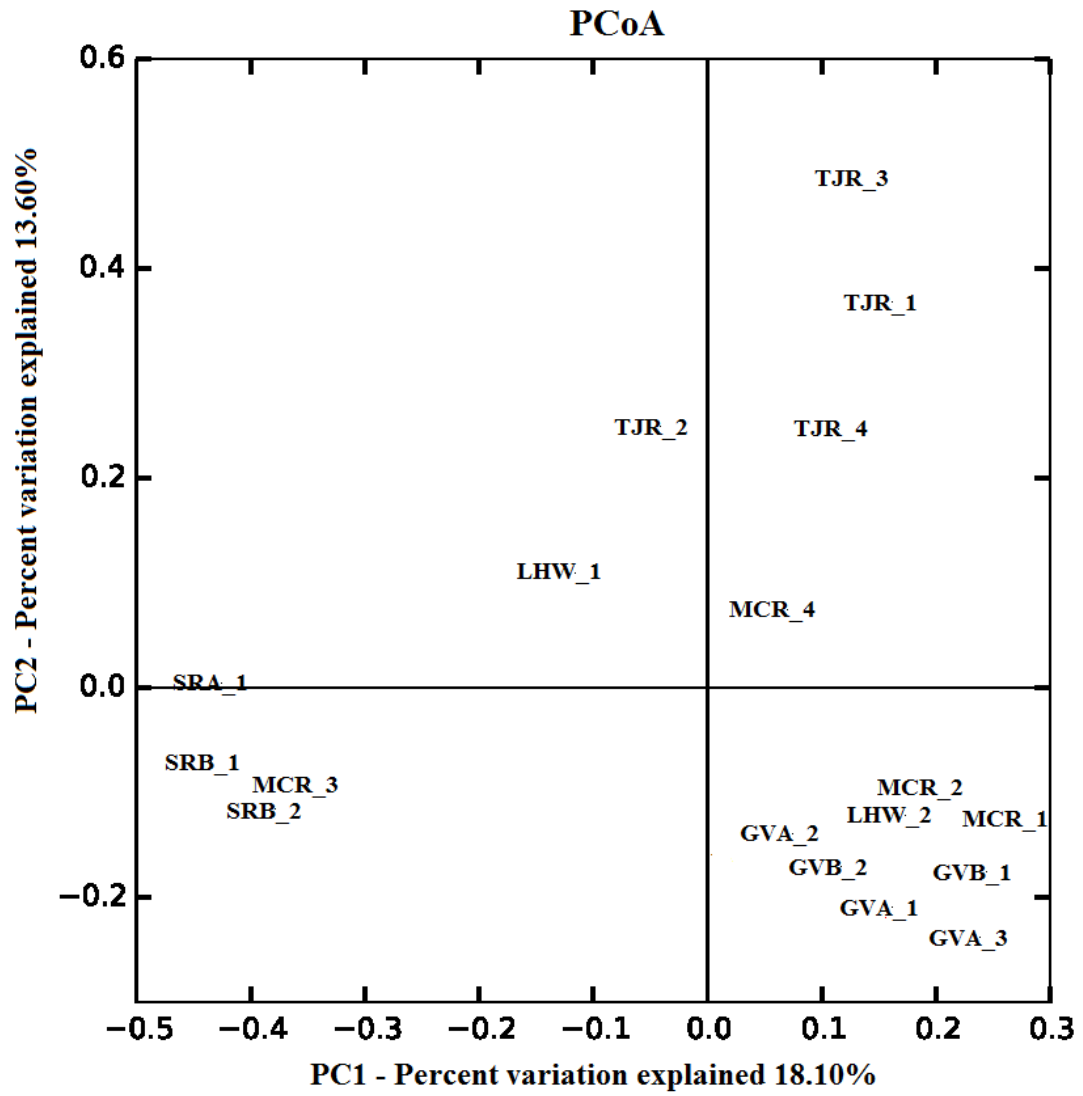
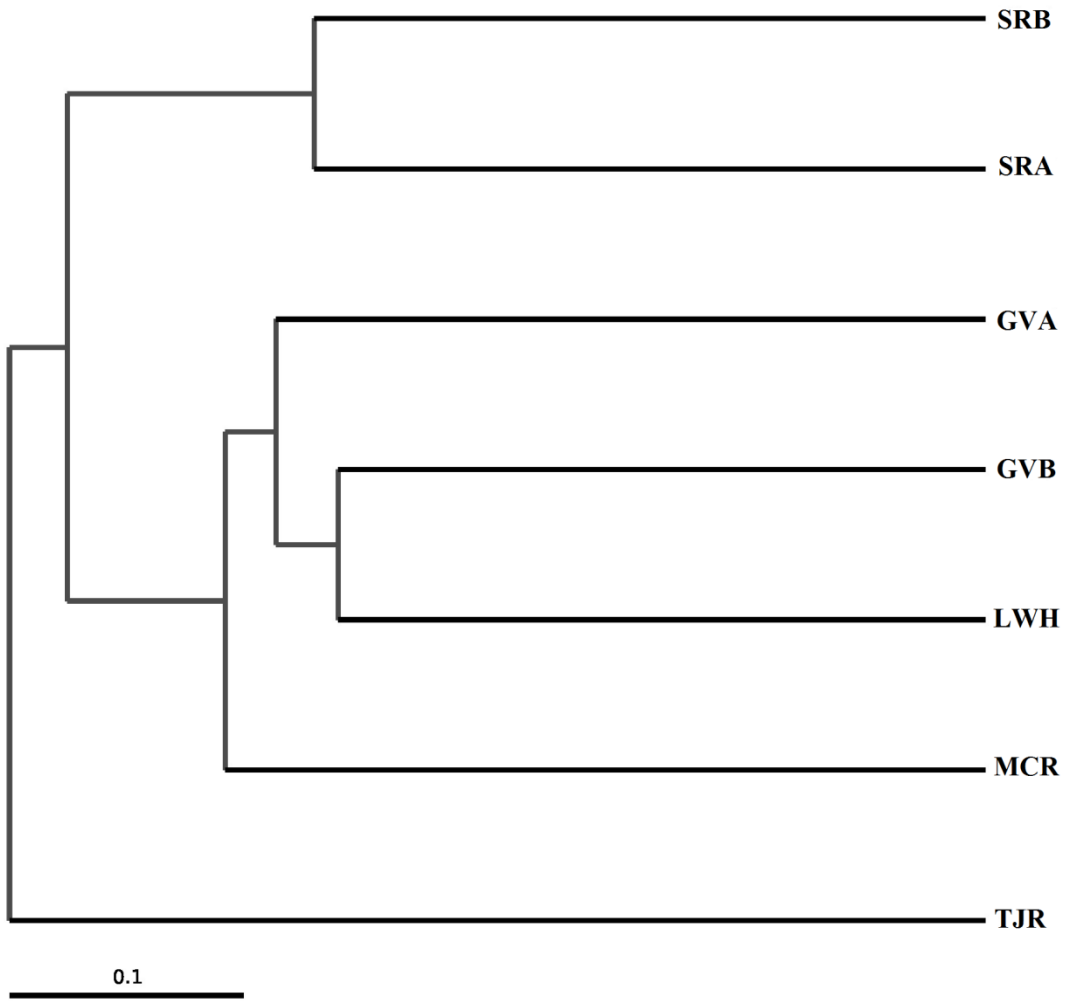


Fig. 4



CHAPTER 4: Discussion and perspective

4.1. Discussion

Pyrosequencing is one of the technologies supplanting Sanger sequencing for research on metagenomics and microbial diversity, with large coverage and sampling depth [214]. The application of pyrosequencing 16S rRNA amplicons to study the microbial community is now broadly used in environmental microbial ecology, and microbial diversity in arid environments is found to be much more abundant than previously surmised. The Titanium-based system is one of the most used pyrosequencing platform, and can produce 500 nt per read. However, data generated by pyrosequencing constantly faces the challenge on how to control the quality and subsequent statistical analyses with such large data sets, including raw sequence cleaning, sequence assignment at different taxonomic levels and clustering into Operational Taxonomic Unit (OTU). There are many different software for pyrosequencing data, as well as online pipelines such as the Mothur, Qiime, and Galaxy. It's crucial to choose the appropriate methods and tools in the analysis of pyrosequencing data, since different tools can general different results and thus lead to bias on the estimation of diversity and structure of a microbial community [215]. For the data produced from environmental samples, it is particular difficult to verify the estimation of microbial community diversity because a large proportion of sequences may be unknown or without a close relative in the databases.

In this chapter, I will discuss the tools and methods we have tested for the analyses of our data, including four important steps: i) quality control of raw reads to remove the noise in the sequences; ii) tools to check and remove chimeras in the data; iii) methods to classify sequences to different taxonomic levels and databases used in the taxonomic assignment of sequences; iiiii) software to cluster the sequences to OTUs.

4.1.1. Quality Control

The 454 pyrosequencing technology is based on sequencing-by-synthesis, and the chemical combination of nucleotide reagents with the template strand in a reaction generates light signal recorded by a camera. Thus the number and type of nucleotides (T, A, C, G) included in each flow of reaction are estimated [216]. The cyclically reaction and corresponding values estimated through light signals forms the basis for base type and per-base quality score calculations [217]. During the process of pyrosequencing, light intensities may not accurately reflect the real homopolymer length, and long homopolymers result in frequent miscalls, including insertions and deletions of nucleotides [218].

There are two major sources of errors that need to be considered in pyrosequencing data: those of the pyrosequencing reactions and those introduced the PCR amplification [43]. In our study, sequences reads generated by pyrosequencing were extracted and treated on FASTA (and QUAL) or FASTQ files. A quality score is assigned to each base as a phred equivalent related to the base calling error probabilities (P):

$$Q = - 10 \log_{10} P [219]$$

Table 4.1. The Phred quality score related to the base calling error probabilities.

Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%

Downstream analysis of sequences can be affected by low-quality sequences, artificial sequences (primers and barcodes) and sequence contamination [220]. Thus as the first step of cleaning sequences, we filtered sequences with the following criteria: (i) length less than 200 nt or longer than 600 nt, (ii) mismatch with barcode sequences or more than one mismatch to the primer, and (iii) the presence of homopolymers of > 8 bp in length. Then adaptor sequences were also removed from the sequences. These general cleaning criteria are commonly applied in research using pyrosequencing data, which can be performed using many metagenomics analysis tools such as Mothur and Qiime. In this step, 2-3% of total sequences were removed from our raw data.

Raw sequences generated from pyrosequencing normally have low Q score at the 3' end (Fig 4.1). Thus, truncate sequences to a specific length or trimming off bases under a specific Q score at the 3' end before filtering low-quality sequences will help conserve a number of sequences as well as remove low-quality bases, with sacrificing the length of sequences.

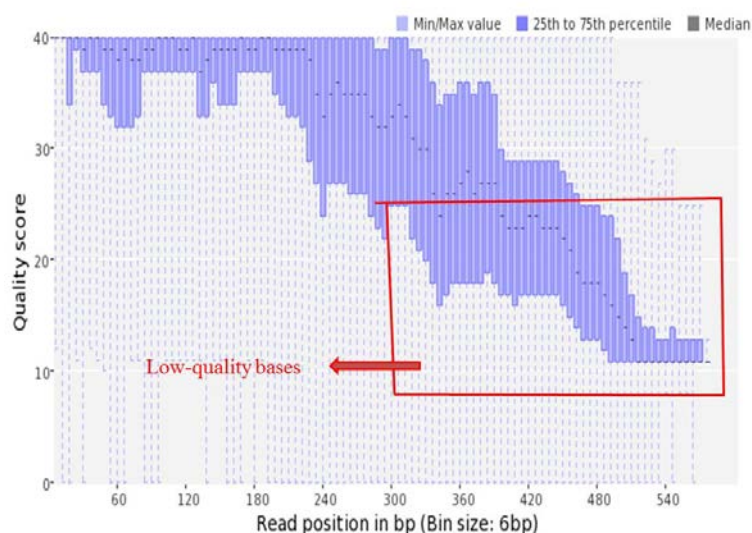


Figure 4.1. Quality score distribution of raw sequences

We compared various tools to trim the 3' end and filter low-quality sequences: a) truncate sequences to the same length (400 nt) using Prinseq [220]; b) trim by Q score

($Q = 25$) using ConDeTri [221]; c) trim by Q score (25) in sliding windows with Mothur. The result are demonstrated in Table 4.2, using sequences from one of our desert samples.

Table 4.2. Comparison of different methods for quality control of sequences (original number of sequences is 22320).

	Truncate at 400 nt	Sliding window	ConDeTri
Number of sequences	20378	17868	17049
Average length of reads	389 nt	279 nt	411 nt

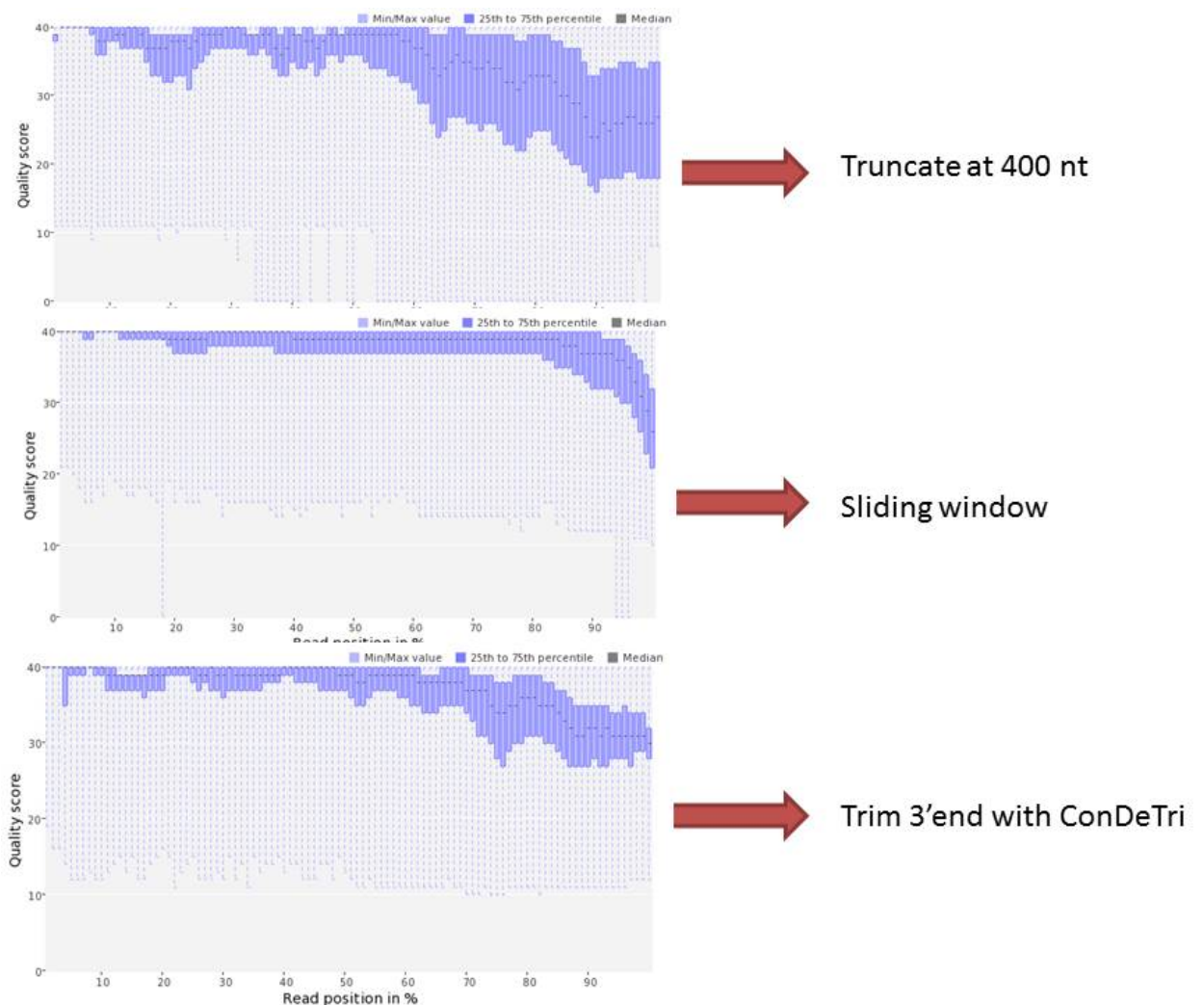


Figure 4.2. Comparison of quality score distribution of sequences trimmed using the three different methods.

The results indicated that the method by truncating sequencing to 400 nt with Prinseq can leave the most sequences, but the base quality at the 3' end is not as good as the other methods. Using sliding windows with the Mothur generated good quality sequences, while the average length of sequences dropped. Trimming the 3' end of sequences with ConDeTri appears to be the best method among the three in terms of number of sequences, average length and base quality.

The quality control step generally removes approximately 10-15% of total sequences from our samples. Sequences with low Q scores can lead to increased false-positive variant calls, resulting in an inaccurate estimation of sequence diversity. For example, we calculated the OTU numbers and Chao 1 index with sequences before and after trimming low quality bases (Table 4.3).

Table 4.3. Diversity result before and after trimming.

	Number of sequences	OTU	Chao 1
Before trimming	5000	1277	6393
After trimming	5000	673	1287

5000 Sequences are randomly selected (using a command in Pangea pipeline) from one of our desert samples (Maine_1).

4.1.2. Chimera Check

During the process of PCR, when incomplete extension occurs in a round, chimeras are generated which can act as a primer for a different sequence in the next round. Chimeras are thus composed of two or more parent sequences. These chimeras need to be detected and removed from data sets by aligning each sequence against a known reference database. We tested two programs for chimera checking that are most frequently used in research with pyrosequencing of 16S rRNA amplicons: UChime and Decipher.

In UChime, the query sequence is divided into four segments, each of which is aligned to a reference database. The best matched parents as candidates are then used to perform a multiple alignment, and chimera sequences will be reported when two candidate parents have an identity closer to (exceed a predetermined threshold) the query sequence than either candidate alone [47]. In Decipher, the query sequence is first assigned to one of the reference phylogenetic groups. Then a set of 30 nt long segments is formed from the sequence to be aligned with those of a reference database. The chimera sequence will be reported when some fragments have few matches within their own reference phylogenetic group but a large number of matches to another reference group [48]. Decipher has a higher rate of detection of short chimeric ranges (100-250 nt) and complex chimeras with multiple parents, while UChime can detect chimeras formed from closely related parents [48].

First, we used a control sample from mixing *E. coli* and *Deinococcus radiodurans* DNA to amplify in PCR. Before chimera checking we followed the cleaning and quality control steps as mentioned above, with 1424 reads remained. The chimeras found in this sample are shown in Fig 4.3. The Greengenes bacterial 16S rRNA database, version 2013_May (OTU_99) was used as reference sequences in UChime,

and the Decipher online version with the RDP database version 2012 was applied in our test. Chimera sequences found in the control samples are 26 for Uchime and 31 for Decipher, and the 26 chimeras observed by Uchime are detected by Decipher as well. Five chimeras are only detected by Decipher.

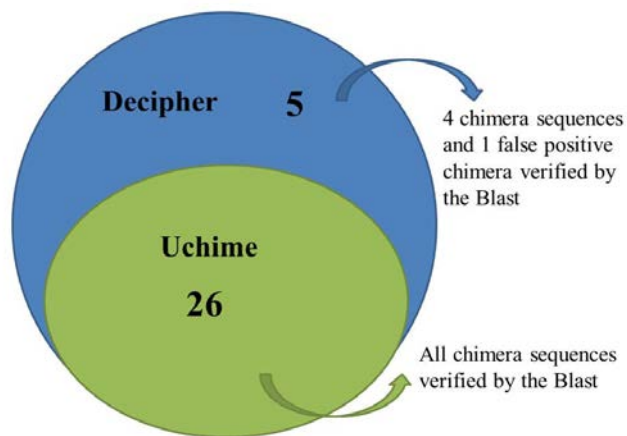


Figure 4.3. Number of chimeras detected by the Uchime and Decipher.

We found more chimera sequences detected by Decipher in the control sample, but with more false positive chimeras, then I further tested the two programs using an environmental sample from the a desert in China (Kumtagh). Sequences in the sample were also treated with cleaning step and quality control with ConDeTri, and 27770 sequences remained. The results are shown in Fig 4.4. We found that 1509 sequences were detected as chimeras by both programs, with 22201 sequences reported as good sequences. UChime and Decipher detected an additional 2550 and 1502 chimeras, respectively.

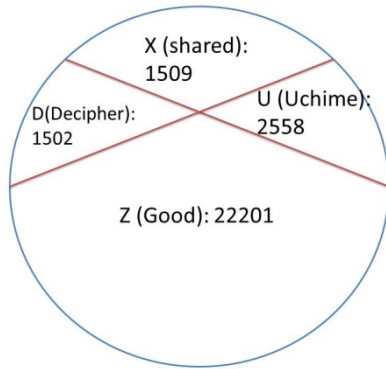


Figure 4.4. Number of sequences detected by the two programs in the Kumtagh sample

To further test the accuracy of the two programs, we classified sequences in the four parts of Fig 4.4 respectively with RDP II classifier and the Silva Database version 119. Table 4.4 demonstrates the proportion of total sequences in each part classified at different taxonomic levels.

Table 4.4. Percentage of potential non-chimeric sequences classified at different taxonomic levels.

	X	U	D	Z
Phylum	63%	86%	89%	99%
Class	56%	81%	83%	98%
Order	43%	75%	64%	92%
Family	35%	53%	46%	78%
Genus	11%	35%	30%	50%

We found the percentage of classified sequences is more in the non-chimera group (Z) than that of the other groups, and chimera sequences detected by both programs are more poorly classified. We also found at the genus level that Decipher removed most of the sequences classified to *Tunicatimonas* as chimeras in the Kumtagh sample,

which were in fact good sequences according to Blast results but absent from the reference database in Decipher. Thus, both the programs itself and the reference database used for chimera checking are essential in detecting chimeras. Based on these results, we removed chimeras detected by both programs (UChime and Decipher = X) in our samples for downstream analysis.

4.1.3. Taxonomic Classification of Sequences

Pyrosequencing generates many of 16S rRNA reads per sample. After sequences are cleaned for low quality bases and chimeras, a further classification step using specific 16S rRNA databases is required. There are two major methods to taxonomically classify 16S rRNA fragments: similarity-based tools such as BLAST and alignment-based taxonomic assignment such as RDP classifier [222]. The BLAST based method assigns the query 16S rRNA sequence to the best significant hit to a reference sequence in the database, which is adopted in several metagenomics analysis pipeline, such as MG-RAST [223] and CAMERA [224]. Since the BLAST approach tends to assign the query sequence to known taxonomic groups, it may lead to a large number of sequences remaining unclassified especially for environmental samples, that may belong to novel genera or even new phyla [225]. The RDP Classifier is based on the naïve Bayesian classification of 8-mer words belonging to the query sequence, which classifies the sequence to a similar taxonomic lineage according to the frequencies of 8-mer words identified to a taxonomic lineage in the database [226].

Reference databases are essential to the classification of sequences. There are three major 16S rRNA databases applied in the studies of environmental microbial communities: Greengenes, RDP and Silva. We summarized the taxonomic information of the three 16S rRNA database in Table 4.5 to show the percentage of sequences classified to each taxonomic level in the database. The RDP database has most of their sequences classified to a deep level (genus), but includes the least number of sequences.

Table 4.5. Summary of sequences in each database assigned to different taxonomic level

	Greengenes (198510) ^a		RDP (9665) ^b		Silva 119 (137878) ^c	
	Number of sequences	Percentage	Number of sequences	Percentage	Number of sequences	Percentage
phylum	198277	99.9%	9665	100%	137878	100%
class	196138	98.8%	9665	100%	135653	98.4%
order	186786	94.1%	9496	98.2%	130271	94.5%
family	153246	77.2%	9040	93.5%	120255	87.2%
genus	91122	45.9%	8833	91.4%	92443	67.1%

^aGreengenes database version 2013_May, 99_otus, with 198510 bacterial sequences

^bRDP database version 2012, with 9665 bacterial sequences

^cSilva database version 119, 99_non-redundant, with 137878 bacterial sequences.

We also tested the three databases using RDP Classifier with a sample from desert in Utah (Table 4.6). The Silva database_v119 presented better results with a higher proportion of sequences classified to the genus level, while the Greengenes database appeared to be better at higher taxonomic level (Phylum to Order).

Table 4.6. Classification result of a sample from desert in Utah

	Percentage of classified sequences		
	Silva	Greengenes	RDP
Phylum	89.4%	97.2%	86.0%
Class	87.3%	94.9%	84.7%
Order	82.2%	85.1%	77.2%
Family	73.1%	67.1%	59.5%
Genus	55.4%	45.9%	49.7%

The RDP Classifier is currently the most popular software used in the classification of 16S rRNA sequences in taxonomic studies of microbial communities [222]. To test the different classifier, we randomly selected 1000 sequences from the Gold_database (full length of 16S rDNA, non-chimera) downloaded in the Usearch website, whose genomic information is well characterized. Three different classifiers tested in our study are RDP II Classifier, Uclust implemented in Qiime and Crest (online version). The results were shown in Table 4.7, and RDP II presented better taxonomic classifications at each level.

Table 4.7. Summary of classification using different classifiers with 1000 sequences randomly selected from gold_database downloaded from Usearch.

	RDP II	Uclust	Crest
Phylum	100.00%	99.90%	99.90%
Class	100.00%	99.80%	93.90%
Order	99.70%	99.30%	85.90%
Family	98.40%	81.10%	64.60%
Genus	96.20%	74.30%	35.80%

Based on these results, the RDP II Classifier with the Silva 16S rRNA database v_119 database were generally applied for the classification of sequences in our samples.

4.1.4. OTU Clustering

Since the majority of microbial species and their 16S rRNA genes have not been taxonomically classified, diversity estimations using pyrosequencing data have to be frequently measured as the number of OTUs in a sample. OTUs are defined by applying a distance threshold corresponding approximately to a specific taxonomic level: 1-3% are typically used for species, 5% for genera, 15% for classes, etc. [227-230]. The number of OTUs are broadly used in microbial diversity studies to reflect the number of species in a sample. Further analyses are also based on the number of OTUs, such as the diversity Chao 1 and Shannon indices, as well as multiple sample comparisons on the microbial community structure.

To cluster 16S rDNA sequences to OTUs, the major methods can be categorized as taxonomic dependent or independent algorithms [231]. The taxonomic dependent methods such as RDP and Mothur, align sequences to a template database to calculate the distance among each other. In contrast, taxonomic independent methods compare sequences against each other to form a distance matrix, like Clustal and MUSCLE, or construct consensus sequences representing each cluster based on a greedy algorithm such as CD-HIT and Uclust.

In our study, we compared four different programs that have been used with an control sample (1393 sequences after quality control and removing chimeras) containing only two species (from *E. coli* and *Deinococcus. radiodurans*), including Uparse, ESPRIT-Tree, RDP and Mothur pipelines. Uparse is based on the CD-HIT greedy heuristic algorithm that can estimate the similarity between two sequences without performing multiple alignments of all pairs of sequences [60]. We chose representative software of each category to test: Uparse and ESPRIT-Tree for taxonomic independent algorithms; RDP and Mothur for taxonomic dependent

algorithms (Table 4.8).

Table 4.8. Number of OTUs clustered using different software at a distance 3%.

	Number of OTUs
Uparse	2
ESPRIT-Tree	3
RDP	4
Mothur	18

The algorithm implemented in the Mothur pipeline generated 18 OTUs, much more than the real number of species in the control sample (two species). We also extracted the representative sequences in each OTU and used BLAST against the NCBI 16S rRNA database. We found that the two representative sequences in Uparse are correctly assigned to the two species. Thus Uparse was chosen for the downstream analysis of our sample in deserts.

4.2. Perspective

In this study, we demonstrated the bacterial diversity and community structures of surface soil in deserts in the Utah State and the Desert of Maine using pyrosequencing of 16S rDNA amplicons. We built our pipeline for the analysis of 16S rDNA pyrosequencing data by combining several existing metagenomics tools. We also searched for correlations between environmental factors and bacterial diversity in the two deserts. Further studies stand be performed to reveal the microbial community and its ecological function in deserts:

- a) Some special bacterial groups, such as members of *Gemmatimonadetes* and *Alphaproteobacteria* , whose population we found significantly correlated to environmental factors in desert soils (pH and concentration of organic matters), and some extremophile genera observed in samples, can be selected for further study using qPCR to more precisely quantify their presence in many desert soils.
- b) Omics-based technologies, such as sequencing of total DNA and RNA in sample sites have the potential to yield major advances in our understanding of the functional capacity of microbial communities and their adaptive potential [232]. This research can help us to understand how microbial communities function in deserts their contribution to biogeochemical processes.
- c) Other microorganisms, such as viruses, fungi and Archaea play important roles in the deserts. NGS technologies can be applied to study their community structures. There are 18S rDNA sequence database available for eukaryote community analysis. It should be noted that the reference database for an Archaeal community study needs to be applied with caution, since the taxonomic classification of Archaeal 16S rDNA sequences are not as well organized as

those in bacterial 16S rDNA sequences.

References

1. Karine A, Querellou J (2009) Cultivating the uncultured: limits, advances and future challenges. *Extremophiles* 13: 583-594.
2. Rastogi G, Sani R (2011) Molecular Techniques to Assess Microbial Community Structure, Function, and Dynamics in the Environment. In: Ahmad I, Ahmad F, Pichtel J, editors. *Microbes and Microbial Technology*: Springer New York. pp. 29-57.
3. Friedrich T, Rahmann S, Weigel W, Rabsch W, Fruth A, et al. (2010) High-throughput microarray technology in diagnostics of enterobacteria based on genome-wide probe selection and regression analysis. *BMC Genomics* 11: 591.
4. van Passel MWJ, Marri PR, Ochman H (2008) The Emergence and Fate of Horizontally Acquired Genes in *Escherichia coli*. *PLoS Comput Biol* 4: e1000059.
5. Theron J, Cloete TE (2000) Molecular Techniques for Determining Microbial Diversity and Community Structure in Natural Environments. *Critical Reviews in Microbiology* 26: 37-57.
6. Reller LB, Weinstein MP, Petti CA (2007) Detection and Identification of Microorganisms by Gene Amplification and Sequencing. *Clinical Infectious Diseases* 44: 1108-1114.
7. López-Campos G, Martínez-Suárez J, Aguado-Urda M, López-Alonso V (2012) Detection, Identification, and Analysis of Foodborne Pathogens. *Microarray Detection and Characterization of Bacterial Foodborne Pathogens*: Springer US. pp. 13-32.
8. Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51: 221-271.
9. Raoult FAaD (2009) Exploring Microbial Diversity Using 16S rRNA High-Throughput Methods. *J Comput Sci Syst Biol* 2: 18.
10. Sarethy IP, Pan S, Danquah MK (2014) Modern Taxonomy for Microbial Diversity. In: Grillo O, editor. *Biodiversity - The Dynamic Balance of the*

Planet.

11. Prakash S, Bisen MD, G. B. Prasad (2012) *Microbes: Concepts and Applications*: Wiley-Blackwell.
12. Speers AE, Cravatt BF Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chemistry & Biology* 11: 535-546.
13. Emerson DA, Liane; Liu, Henry; Liu, Liping (2008) Identifying and characterizing bacteria in an era of genomics and proteomics. *BioScience* 58.
14. Allen EE, Banfield JF (2005) Community genomics in microbial ecology and evolution. *Nat Rev Microbiol* 3: 489-498.
15. Blow N (2008) Metagenomics: exploring unseen communities. *Nature* 453: 687-690.
16. Chuang P, Walsh E, Yen T, Cho I (2014) Chapter 4 - Applications of Next-Generation Sequencing Technologies to the Study of the Human Microbiome. In: Virginia García-Cañas AC, Carolina S, editors. *Comprehensive Analytical Chemistry*: Elsevier. pp. 75-106.
17. Head IM, Saunders JR, Pickup RW (1998) Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microb Ecol* 35: 1-21.
18. Siqueira JF, Fouad AF, Rôças IN (2012) Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology* 4: 10.3402/jom.v3404i3400.10743.
19. Clarridge JE (2004) Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews* 17: 840-862.
20. Fierer N, Lennon JT (2011) The generation and maintenance of diversity in microbial communities. *Am J Bot* 98: 439-448.
21. Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74: 5088-5090.
22. Janda JM, Abbott SL (2007) 16S rRNA Gene Sequencing for Bacterial

- Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology* 45: 2761-2764.
23. Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA Operon Copy Number Reflects Ecological Strategies of Bacteria. *Applied and Environmental Microbiology* 66: 1328-1333.
 24. Schloss PD, Westcott SL (2011) Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. *Applied and Environmental Microbiology* 77: 3219-3226.
 25. Rajendhran J, Gunasekaran P (2011) Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research* 166: 99-110.
 26. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* 5: R245-R249.
 27. Chen K, Pachter L (2005) Bioinformatics for whole-genome shotgun sequencing of microbial communities. *PLoS Comput Biol* 1: 106-112.
 28. Bruno F, Marinella M, Santamaria M (2015) e-DNA Meta-Barcoding: From NGS Raw Data to Taxonomic Profiling. In: Picardi E, editor. *RNA Bioinformatics: Springer New York*. pp. 257-278.
 29. Abubucker S, Segata N, Goll J, Schubert AM, Izard J, et al. (2012) Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. *PLoS Comput Biol* 8: e1002358.
 30. Handelsman J (2004) Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews* 68: 669-685.
 31. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, et al. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311: 392-394.
 32. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364.

33. Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 55: 641-658.
34. Masoudi-Nejad A, Narimani, Zahra, Hosseinkhan, Nazanin (2013) Next Generation Sequencing and Sequence Assembly.
35. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry* 55: 856-866.
36. Hao DC, Ge G, Xiao P, Zhang Y, Yang L (2011) The First Insight into the Tissue Specific *Taxus* Transcriptome via Illumina Second Generation Sequencing. *PLoS ONE* 6: e21220.
37. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, et al. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2: 6-6.
38. Rusk N (2011) Torrents of sequence. *Nat Meth* 8: 44-44.
39. Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Res* 15: 1767-1776.
40. Xu M, Fujita D, Hanagata N (2009) Perspectives and challenges of emerging single-molecule DNA sequencing technologies. *Small* 5: 2638-2649.
41. Dai L, Gao X, Guo Y, Xiao J, Zhang Z (2012) Bioinformatics clouds for big data manipulation. *Biology Direct* 7: 43-43.
42. Schmieder R, Lim YW, Rohwer F, Edwards R (2010) TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* 11: 341.
43. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12: 38.
44. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing *mothur*: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75: 7537-7541.

45. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7: 335-336.
46. Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11: R86.
47. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194-2200.
48. Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* 78: 717-725.
49. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 21: 494-504.
50. Chan ER, Hester J, Kalady M, Xiao H, Li X, et al. (2011) A novel method for determining microflora composition using dynamic phylogenetic analysis of 16S ribosomal RNA deep sequencing data. *Genomics* 98: 253-259.
51. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069-5072.
52. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, et al. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42: D633-642.
53. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590-596.
54. Garcia SP, Rodrigues JM, Santos S, Pratas D, Afreixo V, et al. (2013) A genomic distance for assembly comparison based on compressed maximal exact matches. *IEEE/ACM Trans Comput Biol Bioinform* 10: 793-798.

55. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, et al. (2005) Comparative metagenomics of microbial communities. *Science* 308: 554-557.
56. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, et al. (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360: 1935-1943.
57. Koeppel AF, Wu M (2013) Surprisingly extensive mixed phylogenetic and ecological signals among bacterial Operational Taxonomic Units. *Nucleic Acids Research*.
58. Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature* 459: 193-199.
59. Vetrovsky T, Baldrian P (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8: e57923.
60. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10: 996-998.
61. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658-1659.
62. Shannon CE (1997) The mathematical theory of communication. 1963. *MD Comput* 14: 306-317.
63. Hill MO (1973) Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* 54: 427-432.
64. Bray JR, Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* 27: 326-349.
65. Levandowsky M, Winter D (1971) Distance between Sets. *Nature* 234: 34-35.
66. Sørensen T (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol Skr* 5: 34.
67. Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71: 8228-8235.

68. (2007) Principal coordinate analysis and non-metric multidimensional scaling. *Analysing Ecological Data*: Springer New York. pp. 259-264.
69. Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2: 86-97.
70. Fisher RA (1915) Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* 10: 507-521.
71. Meigs P (1953) World distribution of arid and semi arid homoclimates. *Reviews of Research on Arid Zone Hydrology*. Paris: United Nations: UNESCO.
72. Warner TT (2009) *Desert Meteorology*. Cambridge: Cambridge University Press.
73. Laity JJ (2009) *Deserts and Desert Environments*; Oxford Chichester HNJ, editor: Wiley-Blackwell.
74. Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond JB, et al. (2015) Microbial ecology of hot desert edaphic systems. *FEMS Microbiol Rev* 39: 203-221.
75. Crits-Christoph A, Robinson CK, Barnum T, Fricke WF, Davila AF, et al. (2013) Colonization patterns of soil microbial communities in the Atacama Desert. *Microbiome* 1: 28.
76. Lacap DC, Warren-Rhodes KA, McKay CP, Pointing SB (2011) Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. *Extremophiles* 15: 31-38.
77. Lester RE, Close PG, Barton JL, Pope AJ, Brown SC (2014) Predicting the likely response of data-poor ecosystems to climate change using space-for-time substitution across domains. *Glob Chang Biol* 20: 3471-3481.
78. Stomeo F, Valverde A, Pointing SB, McKay CP, Warren-Rhodes KA, et al. (2013) Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles* 17: 329-337.
79. F.D. Eckardt, K. Soderberg, L.J. Coope, A.A. Mullera, K.J. Vickery, R.D.

- Grandine, C. Jacke, T.S. Kapalangaf, J. Henschelf (2013) The nature of moisture at Gobabeb, in the central Namib Desert. *Journal of Arid Environments* 93: 12.
80. Osborne PL (2012) *Tropical ecosystems and ecological concepts* (2nd Edition). Cambridge: Cambridge University Press.
81. Said M. Badr El-Din SMR, Moawad K. Zahra, and Wafaa M. H. Zidan (2006) Microbial Diversity in Two Egyptian Soils *Catrina Journal* 1.
82. Benslama O BA (2013) Ouided Benslama, Abderrahmane Boulahrouf. *Int J Curr Microbiol App Sci* 628: 35.
83. Kurapova AI, Zenova GM, Sudnitsyn, II, Kizilova AK, Manucharova NA, et al. (2012) [Thermotolerant and thermophilic soil Actinomycetes from desert steppes of Mongolia]. *Mikrobiologiya* 81: 105-116.
84. Mansour Almazrouia, , M. Nazrul Islama, P.D. Jonesa, b, H. Athara, M. Ashfaqr Rahmana (2012) Recent climate change in the Arabian Peninsula: Seasonal rainfall and temperature climatology of Saudi Arabia for 1979–2009. *Atmospheric Research* 111: 16.
85. Burke A (2002) Properties of soil pockets on arid Nama Karoo inselbergs—the effect of geology and derived landforms. *Journal of Arid Environments* 50: 15.
86. R. Hereford RHW, C.I. Longpre (2006) Precipitation history and ecosystem response to multidecadal precipitation variability in the Mojave Desert region, 1893–2001. *Journal of Arid Environments* 67: 21.
87. Free CL, Baxter GS, Dickman CR, Leung LK (2013) Resource pulses in desert river habitats: productivity-biodiversity hotspots, or mirages? *PLoS One* 8: e72690.
88. Ellis J, Nanopoulos DV, Olive KA (2013) No-scale supergravity realization of the Starobinsky model of inflation. *Phys Rev Lett* 111: 111301.
89. 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.
90. Andrew DR, Fitak RR, Munguia-Vega A, Racolta A, Martinson VG, et al. (2012) Abiotic factors shape microbial diversity in Sonoran Desert soils. *Appl Environ Microbiol* 78: 7527-7537.
91. W. R. J. Dean, S. J. Miltonw, Jeltsch F (1999) Large trees, fertile islands, and birds in arid savanna. *Journal of Arid Environments* 41: 61-78.
92. M.P. Lawson, Thomas DSG (2002) Late Quaternary lunette dune sedimentation in the southwestern Kalahari desert, South Africa: luminescence based chronologies of aeolian activity. *Quaternary Science Reviews* 21: 825-836.
93. Samuel Mosweu CM, Tibangayuka Kabanda (2013) Modification of soil properties by *Prosopis L.* in the Kalahari Desert, South-Western Botswana. *Open Journal of Ecology* 3: 5.
94. 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.
95. Tahir Rafique SN, Muhammad I. Bhangar, Tanzil H. Usmani (2008) Fluoride ion contamination in the groundwater of Mithi sub-district, the Thar Desert, Pakistan. *Environmental Geology* 56: 9.
96. Grace PR, Post WM, Hennessy K (2006) The potential impact of climate change on Australia's soil organic carbon resources. *Carbon Balance Manag* 1: 14.
97. Reith F, Brugger J, Zammit CM, Gregg AL, Goldfarb KC, et al. (2012) Influence of geogenic factors on microbial communities in metallogenic Australian soils. *ISME J* 6: 2107-2118.
98. Grove AT, Miles MR, Worthington EB, Doggett H, Dasgupta B, et al. (1977) The Geography of Semi-Arid Lands [and Discussion]. 457-475 p.
99. McKay CP, Friedmann EI, Gomez-Silva B, Caceres-Villanueva L, Andersen DT, et al. (2003) Temperature and moisture conditions for life in the extreme arid region of the Atacama desert: four years of observations including the El Nino of 1997-1998. *Astrobiology* 3: 393-406.
100. Noy-Meir I (1979) Structure and function of desert ecosystems. *Israel Journal of*

Botany 28: 1-19.

101. Webb RH, Fenstermaker, L.F., Heaton, J.S., Hughson, D.L., McDonald, E.V., and Miller, D.M. (2009) The Mojave Desert: Ecosystem Processes and Sustainability.
102. Chapin FS, Sala, Osvaldo E., Huber-Sannwald, Elisabeth (2001) Global Biodiversity in a Changing Environment.
103. Rundel PW (1978) Ecological Relationships of Desert Fog Zone Lichens. *The Bryologist* 81: 277-293.
104. Goudie AS (2013) *Arid and Semi-Arid Geomorphology*. Cambridge: Cambridge University Press.
105. Center PS (1997) *Desert Features*. U.S. Geological Survey.
106. Ward D (2009) *The Biology of Deserts*. Oxford, UK.: Oxford University Press.
107. R. A. Perry, Goodall DW (2009) *Arid Land Ecosystems*. London: Cambridge University Press.
108. Dunkerley DL (2011) Desert Soils. In: Thomas DSG, editor. *Arid Zone Geomorphology: Process, Form and Change in Drylands*, 3rd Edition: John Wiley & Sons.
109. Hui Wanching, Jacquie, Cook Benjamin I., Ravi Sujith, Fuentes José D., Paolo DO (2008) Dust-rainfall feedbacks in the West African Sahel. *Geophysical Research Letters* 42: 7563-7571.
110. Liu Xiaodong, Yin Zhiyong, Zhang Xiaoye, Xuchao Y (2004) Analyses of the spring dust storm frequency of northern China in relation to antecedent and concurrent wind, precipitation, vegetation, and soil moisture conditions. *Journal of Geophysical Research: Atmospheres* 109: 210.
111. Council NR (2010) *America's Climate Choices: Panel on Advancing the Science of Climate Change*.
112. Adams DK, Comrie AC (1997) The North American Monsoon. *Bulletin of the American Meteorological Society* 78: 2197-2213.
113. Laity JE (2002) Desert Environments. In *Physical Geography of North America*

- (A.R. Orme, Ed.). Oxford University Press: 21.
114. Taylor H. Ricketts ED (2000) Review of Terrestrial Ecoregions of North America: A Conservation Assessment. *Great Plains Research: A Journal of Natural and Social Sciences*: 532.
 115. Eugene P. Kiver DVH (1999) *Geology of U.S. Parklands*, 5th edition: John Wiley and Sons. 395 p.
 116. Hunt RM, Jr. (1974) The auditory bulla in Carnivora: an anatomical basis for reappraisal of carnivor evolution. *J Morphol* 143: 21-75.
 117. Basin H (1998) Great Basin–Mojave Desert Region. In: Chan MA, Nicoll K, Ormo J, Okubo C, Komatsu G, editors. Reston: U.S. Geological Survey. pp. 505-538.
 118. Wells SG, Dohrenwend, J.C., McFadden, L.D., Turrin, B.D., Mahrer, K.D. (1985) Late Cenozoic landscape evolution on lava flow surfaces of the Cima volcanic field, Mojave Desert, California. *Geological Society of America Bulletin* 96: 11.
 119. Survey USG (2014) *Geologic Provinces of the United States: Basin and Range Province*. *Geology in the Parks: USGS*
 120. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, et al. (2001) *Terrestrial Ecoregions of the World: A New Map of Life on Earth*. *BioScience* 51: 933-938.
 121. McGinnies WG (1976) *An Overview of The Sonoran Desert*. The Second Annual Conference of the Consortium of Arid Lands Institutions. Arizona: The University of Arizona.
 122. Schmidt RH (1979) A climatic delineation of the 'real' Chihuahuan Desert region. *US Dept Agric Bull* 16: 47.
 123. (2015) *Southern North America: Northern Mexico into southwestern United States*. Washington, DC: The World Wild Life Fond.
 124. Direito SOLE, Pascale; Marees, Andries; Staats, Martijn; Foing, Bernard; Röling, Wilfred F. M. (2011) *A wide variety of putative extremophiles and large*

- beta-diversity at the Mars Desert Research Station (Utah). Special issue: Astrobiology field research in Moon/Mars analogue environments 10: 16.
125. Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC (2012) The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. *ISME J* 6: 1046-1057.
 126. Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Micro* 10: 551-562.
 127. Emanuel W, Shugart H, Stevenson M (1985) Climatic change and the broad-scale distribution of terrestrial ecosystem complexes. *Climatic Change* 7: 29-43.
 128. Amezketa E (2006) An integrated methodology for assessing soil salinization, a pre-condition for land desertification. *Journal of Arid Environments* 67: 594-606.
 129. Cui Y, Shao J (2005) The role of ground water in arid/semiarid ecosystems, Northwest China. *Ground Water* 43: 471-477.
 130. Alyssa Kagel DB, and Karl Gawell (2007) *A Guide to Geothermal Energy and the Environment*. Washington, D.C: Geothermal Energy Association
 131. Ravi S, D'Odorico P, Huxman TE, Collins SL (2010) Interactions Between Soil Erosion Processes and Fires: Implications for the Dynamics of Fertility Islands. *Rangeland Ecology & Management* 63: 267-274.
 132. Cushman JC, Davis SC, Yang X, Borland AM (2015) Development and use of bioenergy feedstocks for semi-arid and arid lands. *Journal of Experimental Botany* 66: 4177-4193.
 133. Taylor RG, Scanlon B, Doll P, Rodell M, van Beek R, et al. (2013) Ground water and climate change. *Nature Clim Change* 3: 322-329.
 134. Scanlon BR, Faunt CC, Longuevergne L, Reedy RC, Alley WM, et al. (2012) Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the National Academy of Sciences* 109: 9320-9325.

135. H. Eswaran RL, P. F. Reich (2001) Land Degradation: An overview. In: E.M. IDH, L.R. Oldeman, F.W.T. Pening de Vries, S.J. Scherr, and S. Sompatpanit, editor. International Conference on Land Degradation and Desertification. India: United States Department of Agriculture
136. Belnap J, Munson SM, Field JP (2011) Aeolian and fluvial processes in dryland regions: the need for integrated studies. *Ecohydrology* 4: 615-622.
137. Wickens G. E. SEDAG, Sita, G., Nahal I. (1995) Role of Acacia species in the rural economy of dry Africa and the Near East. Rome: Food and Agriculture Organization of the United Nations. 137 p.
138. Mitchel P. McClaran WWB (1994) Arizona's Diverse Vegetation and Contributions to Plant Ecology. *Rangelands* 16: 208-217.
139. Reddy BV, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, et al. (2012) Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl Environ Microbiol* 78: 3744-3752.
140. Caruso T, Chan Y, Lacap DC, Lau MCY, McKay CP, et al. (2011) Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J* 5: 1406-1413.
141. Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA, et al. (2009) Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci U S A* 106: 19964-19969.
142. Rivera D, Mejías V, Jáuregui BM, Costa-Tenorio M, López-Archilla AI, et al. (2014) Spreading Topsoil Encourages Ecological Restoration on Embankments: Soil Fertility, Microbial Activity and Vegetation Cover. *PLoS ONE* 9: e101413.
143. Nunes da Rocha U, Cadillo-Quiroz H, Karaoz U, Rajeev L, Klitgord N, et al. (2015) Isolation of a significant fraction of non-phototroph diversity from a desert Biological Soil Crust. *Front Microbiol* 6: 277.
144. Mazor G, Kidron GJ, Vonshak A, Abeliovich A (1996) The role of cyanobacterial exopolysaccharides in structuring desert microbial crusts. *FEMS*

Microbiology Ecology 21: 121-130.

145. Kelley ST, Theisen U, Angenent LT, St Amand A, Pace NR (2004) Molecular analysis of shower curtain biofilm microbes. *Appl Environ Microbiol* 70: 4187-4192.
146. Belnap J, Büdel B, Lange OL (2003) Biological Soil Crusts: Characteristics and Distribution. In: Belnap J, Lange O, editors. *Biological Soil Crusts: Structure, Function, and Management*: Springer Berlin Heidelberg. pp. 3-30.
147. Montanarella L (2015) Soil at the interface between Agriculture and Environment.: The european mission of agriculture and environment.
148. Schlesinger WH, Phippen JS, Wallenstein MD, Hofmockel KS, Klepeis DM, et al. (2003) Community composition and photosynthesis by photoautotrophs under quartz pebbles, southern mojave desert. *Ecology* 84: 3222-3231.
149. Cowan Don A. KN, Pointing Stephen B., Craig Cary S. (2010) Diverse hypolithic refuge communities in the McMurdo Dry Valleys. Cambridge, ROYAUME-UNI: Cambridge University Press. 7 p.
150. Wong FK, Lacap DC, Lau MC, Aitchison JC, Cowan DA, et al. (2010) Hypolithic microbial community of quartz pavement in the high-altitude tundra of central Tibet. *Microb Ecol* 60: 730-739.
151. Bahl J, Lau MC, Smith GJ, Vijaykrishna D, Cary SC, et al. (2011) Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nat Commun* 2: 163.
152. Pointing SB, Warren-Rhodes KA, Lacap DC, Rhodes KL, McKay CP (2007) Hypolithic community shifts occur as a result of liquid water availability along environmental gradients in China's hot and cold hyperarid deserts. *Environmental Microbiology* 9: 414-424.
153. Chan Y, Lacap DC, Lau MCY, Ha KY, Warren-Rhodes KA, et al. (2012) Hypolithic microbial communities: between a rock and a hard place. *Environmental Microbiology* 14: 2272-2282.
154. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, et al. (2012) Cross-biome

- metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences of the United States of America* 109: 21390-21395.
155. Drees KP, Neilson JW, Betancourt JL, Quade J, Henderson DA, et al. (2006) Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. *Appl Environ Microbiol* 72: 7902-7908.
156. Ashish Bhatnagar MB (2005) Microbial diversity in desert ecosystems. *Current Science* 89: 10.
157. Kar A, Garg, B.K., Singh, M.P. and Kathju, S. (2009) Trends in Arid Zone Research In India.
158. Lau MC, Aitchison JC, Pointing SB (2009) Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles* 13: 139-149.
159. Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, et al. (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* 16: 553-566.
160. An S, Couteau C, Luo F, Neveu J, DuBow MS (2013) Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts. *Microb Ecol* 66: 850-860.
161. Warren-Rhodes KA, Kevin L. R, Boyle LN, Pointing SB, Chen Y, et al. (2007) Cyanobacterial ecology across environmental gradients and spatial scales in China's hot and cold deserts. 470-482 p.
162. Wierzchos J, Ascaso C, McKay CP (2006) Endolithic cyanobacteria in halite rocks from the hyperarid core of the Atacama Desert. *Astrobiology* 6: 415-422.
163. Chanal A, Chapon V, Benzerara K, Barakat M, Christen R, et al. (2006) The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environmental Microbiology* 8: 514-525.

164. Chanal A, Chapon V, Benzerara K, Barakat M, Christen R, et al. (2006) The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environmental Microbiology* 8: 514-525.
165. Makhalanyane TP, Valverde A, Lacap DC, Pointing SB, Tuffin MI, et al. (2013) Evidence of species recruitment and development of hot desert hypolithic communities. *Environ Microbiol Rep* 5: 219-224.
166. Santhanam R, Rong X, Huang Y, Andrews B, Asenjo J, et al. (2013) *Streptomyces bullii* sp. nov., isolated from a hyper-arid Atacama Desert soil. *Antonie van Leeuwenhoek* 103: 367-373.
167. Gao Q, Garcia-Pichel F (2011) Microbial ultraviolet sunscreens. *Nat Rev Micro* 9: 791-802.
168. Prestel E, Regard C, Salameh S, Neveu J, DuBow M (2013) The bacteria and bacteriophages from a Mesquite Flats site of the Death Valley desert. *Antonie van Leeuwenhoek* 103: 1329-1341.
169. Subhash Y, Sasikala C, Ramana Ch V (2014) *Pontibacter ruber* sp. nov. and *Pontibacter deserti* sp. nov., isolated from the desert. *Int J Syst Evol Microbiol* 64: 1006-1011.
170. Lester ED, Satomi M, Ponce A (2007) Microflora of extreme arid Atacama Desert soils. *Soil Biology and Biochemistry* 39: 704-708.
171. Zeng Y, Koblížek M, Feng F, Liu Y, Wu Z, et al. (2013) Whole-Genome Sequences of an Aerobic Anoxygenic Phototroph, *Blastomonas* sp. Strain AAP53, Isolated from a Freshwater Desert Lake in Inner Mongolia, China. *Genome Announcements* 1.
172. Margate D (2001) Desert-like soil surface features like desert pavements, surface accumulation of salt, calcium carbonate accumulation, and surface exposure.
173. Crawford RMM (2009) The Biology of Coastal Sand Dunes. *Annals of Botany* 104: 288.
174. Russo SE, Legge R, Weber KA, Brodie EL, Goldfarb KC, et al. (2012) Bacterial

- community structure of contrasting soils underlying Bornean rain forests: Inferences from microarray and next-generation sequencing methods. *Soil Biology and Biochemistry* 55: 48-59.
175. Halliday E, McLellan SL, Amaral-Zettler LA, Sogin ML, Gast RJ (2014) Comparison of Bacterial Communities in Sands and Water at Beaches with Bacterial Water Quality Violations. *PLoS ONE* 9: e90815.
176. McHugh T, Schwartz E (2015) Changes in plant community composition and reduced precipitation have limited effects on the structure of soil bacterial and fungal communities present in a semiarid grassland. *Plant and Soil* 388: 175-186.
177. Huang J, Sheng X-F, Xi J, He L-Y, Huang Z, et al. (2014) Depth-Related Changes in Community Structure of Culturable Mineral Weathering Bacteria and in Weathering Patterns Caused by Them along Two Contrasting Soil Profiles. *Applied and Environmental Microbiology* 80: 29-42.
178. Chen J, Blume H-P, Beyer L (2000) Weathering of rocks induced by lichen colonization - a review. *CATENA* 39: 121-146.
179. Uroz S, Calvaruso C, Turpault MP, Pierrat JC, Mustin C, et al. (2007) Effect of the Mycorrhizosphere on the Genotypic and Metabolic Diversity of the Bacterial Communities Involved in Mineral Weathering in a Forest Soil. *Applied and Environmental Microbiology* 73: 3019-3027.
180. Gadd GM (2010) Metals, minerals and microbes: geomicrobiology and bioremediation. *Microbiology* 156: 609-643.
181. Barrios E (2007) Soil biota, ecosystem services and land productivity. *Ecological Economics* 64: 269-285.
182. Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, et al. (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol Ecol* 21: 1821-1833.
183. Thomsen PF, Willerslev E (2015) Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological*

Conservation 183: 4-18.

184. Ramette A, Tiedje JM (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci U S A* 104: 2761-2766.
185. Amann R, Ludwig W (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiology Reviews* 24: 555-565.
186. 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. *Applied and Environmental Microbiology* 75: 5111-5120.
187. Andrew DR, Fitak RR, Munguia-Vega A, Racolta A, Martinson VG, et al. (2012) Abiotic Factors Shape Microbial Diversity in Sonoran Desert Soils. *Applied and Environmental Microbiology* 78: 7527-7537.
188. Finkel OM, Burch AY, Elad T, Huse SM, Lindow SE, et al. (2012) Distance-decay relationships partially determine diversity patterns of phyllosphere bacteria on Tamarix trees across the Sonoran Desert [corrected]. *Appl Environ Microbiol* 78: 6187-6193.
189. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* 104: 11436-11440.
190. El Hidri D, Guesmi A, Najjari A, Cherif H, Ettoumi B, et al. (2013) Cultivation-Dependant Assessment, Diversity, and Ecology of Haloalkaliphilic Bacteria in Arid Saline Systems of Southern Tunisia. *BioMed Research International* 2013: 15.
191. Chamizo S, Cantón Y, Miralles I, Domingo F (2012) Biological soil crust development affects physicochemical characteristics of soil surface in semiarid ecosystems. *Soil Biology and Biochemistry* 49: 96-105.
192. Wang J, Shen J, Wu Y, Tu C, Soininen J, et al. (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J* 7: 1310-1321.

193. Sher Y, Zaady E, Nejidat A (2013) Spatial and temporal diversity and abundance of ammonia oxidizers in semi-arid and arid soils: indications for a differential seasonal effect on archaeal and bacterial ammonia oxidizers. *FEMS Microbiol Ecol* 86: 544-556.
194. Koberl M, Muller H, Ramadan EM, Berg G (2011) Desert farming benefits from microbial potential in arid soils and promotes diversity and plant health. *PLoS One* 6: e24452.
195. Van Horn DJ, Okie JG, Buelow HN, Gooseff MN, Barrett JE, et al. (2014) Soil microbial responses to increased moisture and organic resources along a salinity gradient in a polar desert. *Appl Environ Microbiol* 80: 3034-3043.
196. Pasternak Z, Al-Ashhab A, Gatica J, Gafny R, Avraham S, et al. (2013) Spatial and temporal biogeography of soil microbial communities in arid and semiarid regions. *PLoS One* 8: e69705.
197. Van Horn DJ, Van Horn ML, Barrett JE, Gooseff MN, Altrichter AE, et al. (2013) Factors Controlling Soil Microbial Biomass and Bacterial Diversity and Community Composition in a Cold Desert Ecosystem: Role of Geographic Scale. *PLoS ONE* 8: e66103.
198. Canfora L, Bacci G, Pinzari F, Lo Papa G, Dazzi C, et al. (2014) Salinity and Bacterial Diversity: To What Extent Does the Concentration of Salt Affect the Bacterial Community in a Saline Soil? *PLoS ONE* 9: e106662.
199. Aislabie J, Jordan S, Ayton J, Klassen JL, Barker GM, et al. (2009) Bacterial diversity associated with ornithogenic soil of the Ross Sea region, Antarctica. *Can J Microbiol* 55: 21-36.
200. Derzelle S, Hallet B, Ferain T, Delcour J, Hols P (2003) Improved Adaptation to Cold-Shock, Stationary-Phase, and Freezing Stresses in *Lactobacillus plantarum* Overproducing Cold-Shock Proteins. *Applied and Environmental Microbiology* 69: 4285-4290.
201. Gleeson DB, Kennedy NM, Clipson N, Melville K, Gadd GM, et al. (2006) Characterization of bacterial community structure on a weathered pegmatitic

- granite. *Microbial Ecology* 51: 526-534.
202. Weinert N, Piceno Y, Ding GC, Meincke R, Heuer H, et al. (2011) PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol Ecol* 75: 497-506.
203. Aguilera LE, Gutiérrez JR, Meserve PL (1999) Variation in soil micro-organisms and nutrients underneath and outside the canopy of *Adesmia bedwellii* (Papilionaceae) shrubs in arid coastal Chile following drought and above average rainfall. *Journal of Arid Environments* 42: 61-70.
204. Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E (2001) Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl Environ Microbiol* 67: 4215-4224.
205. Ian Pepper CG, Terry J. Gentry (2014) *Environmental Microbiology*, 3rd edition. Amsterdam: Elsevier B.V.
206. Gorbushina AA (2007) Life on the rocks. *Environmental Microbiology* 9: 1613-1631.
207. Pointing SB, Belnap J (2012) Microbial colonization and controls in dryland systems. *Nat Rev Microbiol* 10: 551-562.
208. Vriezen JA, de Bruijn FJ, Nusslein K (2007) Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature. *Appl Environ Microbiol* 73: 3451-3459.
209. Wynn-Williams D, Edwards HM (2002) Environmental UV Radiation: Biological Strategies for Protection and Avoidance. In: Horneck G, Baumstark-Khan C, editors. *Astrobiology*: Springer Berlin Heidelberg. pp. 245-260.
210. Rainey FA, Ray K, Ferreira M, Gatz BZ, Nobre MF, et al. (2005) Extensive Diversity of Ionizing-Radiation-Resistant Bacteria Recovered from Sonoran Desert Soil and Description of Nine New Species of the Genus *Deinococcus* Obtained from a Single Soil Sample. *Applied and Environmental*

- Microbiology 71: 5225-5235.
211. Yu L, Luo X, Liu M, Huang Q (2015) Diversity of ionizing radiation-resistant bacteria obtained from the Taklimakan Desert. *Journal of Basic Microbiology* 55: 135-140.
 212. Kirti A, Rajaram H, Apte SK (2014) The Hypothetical Protein 'All4779', and Not the Annotated 'Alr0088' and 'Alr7579' Proteins, Is the Major Typical Single-Stranded DNA Binding Protein of the *Cyanobacterium, Anabaena sp.* PCC7120. *PLoS ONE* 9: e93592.
 213. Colwell RR (1997) Microbial diversity: the importance of exploration and conservation. *J Ind Microbiol Biotechnol* 18: 302-307.
 214. Kunin V, Engelbrekton 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.
 215. Barriuso J, Valverde J, Mellado R (2011) Estimation of bacterial diversity using next generation sequencing of 16S rDNA: a comparison of different workflows. *BMC Bioinformatics* 12: 1-11.
 216. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
 217. Balzer S, Malde K, Lanzen A, Sharma A, Jonassen I (2010) Characteristics of 454 pyrosequencing data--enabling realistic simulation with flowsim. *Bioinformatics* 26: i420-425.
 218. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* 6: 639-641.
 219. Richterich P (1998) Estimation of errors in "raw" DNA sequences: a validation study. *Genome Res* 8: 251-259.
 220. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863-864.

221. Smeds L, Kunstner A (2011) ConDeTri--a content dependent read trimmer for Illumina data. *PLoS One* 6: e26314.
222. Santamaria M, Fosso B, Consiglio A, De Caro G, Grillo G, et al. (2012) Reference databases for taxonomic assignment in metagenomics. *Briefings in Bioinformatics* 13: 682-695.
223. Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010) Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc* 2010: pdb prot5368.
224. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: a community resource for metagenomics. *PLoS Biol* 5: e75.
225. Ghosh TS, Gajjala P, Mohammed MH, Mande SS (2012) C16S - a Hidden Markov Model based algorithm for taxonomic classification of 16S rRNA gene sequences. *Genomics* 99: 195-201.
226. Lan Y, Wang Q, Cole JR, Rosen GL (2012) Using the RDP classifier to predict taxonomic novelty and reduce the search space for finding novel organisms. *PLoS One* 7: e32491.
227. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180: 4765-4774.
228. Sait M, Hugenholtz P, Janssen PH (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ Microbiol* 4: 654-666.
229. Schloss PD, Handelsman J (2006) Toward a census of bacteria in soil. *PLoS Comput Biol* 2: e92.
230. White JR, Navlakha S, Nagarajan N, Ghodsi MR, Kingsford C, et al. (2010) Alignment and clustering of phylogenetic markers--implications for microbial diversity studies. *BMC Bioinformatics* 11: 152.
231. Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, et al. (2012) A large-scale benchmark study of existing algorithms for taxonomy-independent microbial

community analysis. *Brief Bioinform* 13: 107-121.

232. Chan Y, Lacap DC, Lau MC, Ha KY, Warren-Rhodes KA, et al. (2012) Hypolithic microbial communities: between a rock and a hard place. *Environ Microbiol* 14: 2272-2282.

Titre : La diversité bactérienne dans les sols de surface de San Rafael Swell (Utah, USA) et le Desert de Maine (USA)

Synthèse en français :

Les zones arides couvrent environ un tiers de la surface terrestre de la planète. Dans cette étude, la diversité et la structure des communautés bactériennes de la surface du sol des déserts des États de l'Utah et du Maine ont été mises en évidence. Nous avons mis en œuvre une procédure permettant l'analyse des séquences de l'ADNr 16S en combinant des outils préexistants dédiés à la métagénomique. Ainsi, des corrélations entre certains facteurs environnementaux et la diversité bactérienne dans les deux déserts, ont pu être établis.

Le désert du Maine situé dans le nord-est États-Unis est une étendue de boue glaciaire, entourée par une forêt de pins. Nous avons observé que les échantillons de sol provenant du désert du Maine présentent une diversité bactérienne singulière, avec une prédominance de *Proteobacteria* et *Actinobacteria*. Les bactéries du genre le plus abondant, *Acidiphilium*, représentent 12,5% du total des séquences d'ADNr 16S.

Le Désert de l'Utah présente des caractéristiques géographiques qui ressemblent à Mars. En effet il est caractérisé par la présence de collines de couleur rouge et de sols constitués de grès. Les phylums prédominants sont les *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* et *Gemmatimonadetes*. Les genres les plus abondants dans nos échantillons sont *Cesiribacter*, *Lysobacter*, *Adhaeribacter*, *Microvirga* et *Pontibacter*. Mais de façon notable, il semble que l'abondance relative des *Alphaproteobacteria* et des *Gemmatimonadetes* est significativement corrélée aux certains facteurs environnementaux des sols, par exemple de pH et des concentration des matières organiques.

Titre : La diversité bactérienne dans les sols de surface de San Rafael Swell (Utah, USA) et le Desert de Maine (USA)

Mots clés : Diversité bactérienne, Désert, Pyroséquençage

Résumé : Les zones arides couvrent environ un tiers de la surface terrestre de la planète. Des études visant à comprendre la dispersion microbienne dans les déserts ont été réalisées. En effet, les communautés microbiennes du sable des déserts peuvent jouer un rôle important dans la stabilité des sols, les cycles de la matière et la santé environnementale. Le pyroséquençage pour les ARNr 16S à partir de l'ADN total extrait des sols des échantillons de sable peut donner des renseignements clés sur la structure des communautés bactériennes qui les composent. Dans cette étude, la diversité et la structure des communautés bactériennes de la surface du sol des déserts des États de l'Utah et du Maine ont été mises en évidence. Nous avons mise en œuvre une procédure permettant l'analyse des séquences de l'ADNr 16S en combinant des outils préexistants dédiés à la métagénomique. Ainsi, des corrélations entre certains facteurs environnementaux et la diversité bactérienne dans les deux déserts, ont pu être établis.

Le désert du Maine situé dans le nord-est Etats-Unis est une étendue de boue glaciaire, entourée par une forêt de pins. Le sol de ce désert possède les caractéristiques d'un sable avec de très faibles capacités de rétention d'eau, d'une rétention des éléments nutritifs, ainsi qu'une valeur de pH relativement faible (pH 5,09). Les échantillons provenant de ce site présentent donc des propriétés particulièrement intéressantes à étudier en lien avec la diversité bactérienne. Deux échantillons de sable de la surface du désert du Maine ont été obtenus, et le pyroséquençage des gènes d'ADNr 16S obtenus après amplification par PCR à partir de l'ADN total extrait a été utilisé pour évaluer la diversité bactérienne, la structure de la communauté bactérienne et l'abondance relative des principaux taxons. Nous avons observé que les échantillons de sol provenant du désert du Maine présentent une diversité bactérienne singulière, avec une prédominance de *Proteobacteria* et *Actinobacteria*. Les bactéries du genre le plus abondant, *Acidiphilium*, représentent 12,5% du total des séquences d'ADNr 16S. Au total, 1 394 OTU ont été comptabilisées. En comparant les résultats de notre population

bactérienne avec des études portant sur des sols avec caractéristiques similaires, nous avons constaté que les échantillons du Maine contiennent une faible diversité du phylum *Acidobacteria* que les sols acides des certains forêts, et moins de *Firmicutes* ainsi que plus de *Proteobacteria* que les sols des déserts oligotrophes.

Le Désert de l'Utah présente des caractéristiques géographiques qui ressemblent à Mars. En effet il est caractérisé par la présence de collines de couleur rouge et de sols constitués de grès. Les sites d'échantillonnage couvrent le Goblin Valley State Park et autour, notamment sur le plateau du Colorado. Avec des approches similaires à ceux utilisés pour le désert du Maine, des corrélations entre facteurs environnementaux (paramètres physico-chimiques) et diversité de structure des communautés bactériennes obtenus, ont été étudiés. Les phylums prédominants sont les *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* et *Gemmatimonadetes*. Les genres les plus abondants dans nos échantillons sont *Cesiribacter*, *Lysobacter*, *Adhaeribacter*, *Microvirga* et *Pontibacter*. Mais de façon notable, il semble que l'abondance relative des *Alphaproteobacteria* et des *Gemmatimonadetes* est significativement corrélée aux certains facteurs environnementaux des sols, par exemple de pH et des concentration des matières organiques.

Title : The bacterial communities of sand-like surface soils of the San Rafael Swell (Utah, USA) and the Desert of Maine (USA)

Keywords : Bacterial diversity, Deserts, Pyrosequencing

Abstract : Aridity is the dominant climatic factor over approximately 30% of the land surface of the world. Research concerning microbial populations in two U.S. deserts has been performed to determine the diversity of these bacteria. Pyrosequencing-based profiling of 16S rRNA amplicons from surface soils of sand samples can provide key insights into the structure of bacterial communities and their diversity. In this study, we demonstrated the bacterial diversity and community structures of surface soil in the Colorado Plateau in the Utah State and the Desert of Maine using pyrosequencing of 16S rRNA amplicons. We built our pipeline for the analysis of 16S rRNA pyrosequencing data by combining several existing tools of metagenomics. We also examined correlations between certain environmental factors and bacterial diversity in the two deserts.

The Desert of Maine is a tract of glacial silt, surrounded by a pine forest, in the state of Maine located in the northeastern USA. The soil of the Desert of Maine has a sandy texture with poor water holding abilities, nutrient retention capabilities and a relatively low pH value (pH 5.09). Samples from this site thus present an interesting place to examine the bacterial diversity in mineral sandy loam soils with an acidic pH and low concentrations of organic materials. Two surface sand samples from the Desert of Maine were obtained, and pyrosequencing of PCR amplified 16S rDNA genes from total extracted DNA was used to assess bacterial diversity, community structure and the relative abundance of major bacterial taxa. We found that the soil samples from the Desert of Maine showed high levels of bacterial diversity, with a predominance of members belonging to the Proteobacteria and Actinobacteria phyla. Bacteria from the most abundant genus, *Acidiphilium*, represent 12.5% of the total 16S rDNA sequences. In total, 1394 OTUs were observed in the two samples, with the

number of common OTUs observed in both samples being 668. By comparing our bacterial population results with studies on related soil environments, we found that the samples contained less *Acidobacteria* than soils from acid soil forests, and less *Firmicutes* plus more *Proteobacteria* than soils from oligotrophic deserts.

Deserts in Utah has geographic features that resemble Mars, characterized by red-colored hills, soils and sandstones. Our sample sites cover the Goblin Valley State Park and nearby regions on the Colorado Plateau. We also examined physicochemical parameters of soil from the sample sites to investigate correlations between bacterial community structure and environmental drivers. The predominant phyla of the samples represent members of the *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Gemmatimonadetes*. The most abundant genera in our samples are *Cesiribacter*, *Lysobacter*, *Adhaeribacter*, *Microvirga* and *Pontibacter*. We found that the relative abundance of Alphaproteobacteria and Gemmatimonadetes are significantly correlated to some environmental factors of soils, such as pH and concentration of organic matters.