



HAL
open science

Characterization of bacterial diversity in three oligotrophic environments using high-throughput sequencing technology

Shu An

► **To cite this version:**

Shu An. Characterization of bacterial diversity in three oligotrophic environments using high-throughput sequencing technology. Agricultural sciences. Université Paris Sud - Paris XI, 2012. English. NNT: 2012PA112152 . tel-00859417

HAL Id: tel-00859417

<https://theses.hal.science/tel-00859417>

Submitted on 8 Sep 2013

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.

UNIVERSITE PARIS-SUD

THESE DE DOCTORAT

Présentée pour l'obtention du grade de

DOCTEUR EN SCIENCE

DE L'UNIVERSITE PARIS-SUD ORSAY

Discipline: Microbiologie

Par

SHU AN

**Characterization of bacterial diversity in three
oligotrophic environments using high-throughput
sequencing technology**

Directeur de thèse

Michael DUBOW

Jury

M. Denis FAURE	Rapporteur
M. David LANE	Rapporteur
M. Denis LE PASLIER	Examineur
Mme. Monique BOLOTIN-FUKUHARA	Examineur
M. Michael DUBOW	Examineur
M. Olivier TENAILLON	Examineur

Acknowledgements

Firstly I would like to thank my supervisor, Pr. Michael DuBow, for his constant encouragement and guidance. He has walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

I would like to express my heartfelt gratitude to Pr. Monique Bolotin-Fukuhara, who gave me her help and time in listening to me and helping me work out my problems during the difficult course of the thesis. Without her patient instruction, expert guidance and supports, the completion of this thesis would not have been possible.

I will also acknowledge sincerely Micheline Terrier, Jean-Baptist Poitelon and Asmaa El Hamidi-Jobic who helped me to improve the writing of this thesis.

I owe my deepest gratitude to my beloved family for their loving considerations, unconditional supports and great confidence in me all through these years.

Finally, I thank all the members of the team of Pr. Michael DuBow for the every warm moment we shared and their helps during my thesis.

List of contents

CHAPTER I: Introduction	7
I. A. Oligotrophic environments	8
I. A. a. The desert	9
I. A. b. Sandstorms	12
I. A. c. The drinking water distribution system.....	14
I. B. Environmental microbiology	16
I. B. a. General introduction.....	16
I. B. b. Microbial diversity	17
I. B. c. Technology tools	19
I. B. c. 1. Identification of bacteria	19
I. B. c. 3. Genetic fingerprint methods.....	24
I. B. c. 4. Conventional sequencing methods.....	29
I. B. c. 5. Next generation sequencing methods.....	30
I. B. c. 6. Quantitative detection and/or activity detection of bacteria	35
I. C. The bacterial diversity in oligotrophic environment.....	37
I. C. a. Deserts.....	38
I. C. b. Sandstorms	41
I. C. c. Drinking water systems.....	43
I. C. c. 1. Water	43
I. C. c. 2. Biofilm	46
I. D. Factors that can affect bacterial community diversity.....	49
I. D. a. In desert soil	49
I. D. b. In drinking water systems	54
I. E. Objective of my thesis.....	56
CHAPTER II: Bacterial diversity in surface sand of Asian deserts	57
II. A. Introduction d'étude	58
Article I: Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts	60

Figures and Tables	82
CHAPTER III: Particle-associated bacteria in Asian sandstorms	97
III. A. Introduction d'étude	98
Article II: Modification of atmospheric dust and sand-associated bacterial communities occurring during Asian sandstorms.	100
Figures and Tables	122
CHAPTER IV: Bacteria in Parisian drinking water distribution system.....	131
IV. A. Introduction d'étude	132
Article III: Characterization of bacterial populations in drinking water and biofilms present in part of Parisian distribution network.....	134
Figures and Tables	151
CHAPTER V: General discussion and Future perspectives	160
V. A. DNA extraction from oligotrophic environments	161
V. B. Partial 16S rDNA sequencing	161
V. C. Comparaison of high-throughput sequencing technology.....	163
V. D. Bacterial viability	164
V. E. Importance of detection of rare members in oligotrophic environments..	165
V. F. Pathogen detection.....	165
V. G. Deeper identification of unclassified bacteria	166
V. H. Bioinformatics for high-throughput sequencing analyses	167
V. I. Diversity indices accuracy	169
V. J. Microbial communities in oligotrophic environments.....	169
Reference	171
Abstract	188
Résumé.....	189

List of Figures

Figure 1: Map of the global extent of hyperarid, arid, and semi-arid regions and principal deserts. Generated by U. S. Geological Survey after Meigs, 1953 (151).	9
Figure 2: The desert landscape types: a. Sand desert, b. Stony desert, c. Rock desert, d. Plateau desert, e. Mountain desert. a,c,e were taken from Warhol, 2007 (225), b,d were from Johnson, 2003 (99).....	11
Figure 3: An evolutionary timeline showing the approximate appearance of life from 4.5 billion years ago to the present time (141).....	17
Figure 4: Universal phylogenetic tree, based on the comparison of sequence analysis of the 16S ribosomal RNA (140).	22
Figure 5: 16S rRNA secondary structure with 9 variable regions (77).....	23
Figure 6: Process for NGS 16S rDNA data analysis	33
Figure 7: Process of population analysis from environmental sample. Five different color arrow represent five possible analysis way.	36
Figure 8: the schematic presentation of five stages for the formation of biofilms in water supply system, adapted from Stoodley et al, 2002 (203).	47
Figure 9: The schematic presentation of multiregion pyrosequencing for bacterial identification.	167
Figure 10: Future research directions after the characterization of bacterial populations in oligotrophic environments.	170

List of Tables

Table 1: Avantages and disadvantages of different footprinting techniques adopted from (93)	28
Table 2: Culturability determined as a percentage of culturable bacteria in comparison with total cell counts.	37

List of abbreviations

ADE: Asian dust event

AOC: Assimilable organic carbon

ARISA: Ribosomal Intergenic Spacer Analysis

ATP: adenosine tri-phosphate

BLAST: Basic Local Alignment Search Tool

bTEFAP: Bacterial tag-encoded FLX amplicon pyrosequencing

CARD-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

HPC: heterotrophic plate counts

GAST: Global Alignment for Sequence Taxonomy

KOSA: name of Asian dust

NGS: next-generation sequencing

OTU: operational taxonomic unit

PAR: photosynthetically active radiation

PCA: principal component analysis

PCR: Polymerase Chain Reaction

PLFA: Phospholipid-derived fatty acids

PM_{2.5}: particulate matter <2.5µm

PM₁₀: particulate matter <10µm

qRT-PCR: real time quantitative PCR

RDP: ribosomal database project

rRNA: ribosomal ribonucleic acid

RNA: ribonucleic acid

SRA: Sequence Read Archive

SSCP: single-strand conformation polymorphism

T-RFLP: terminal restriction fragment length polymorphism

UNEP: United Nations Environment Programme

UPGMA: Unweighted Pair Group Method with Arithmetic Mean

UV: UltraViolet

WHO: World Health Organization

CHAPTER I: Introduction

I. A. Oligotrophic environments

Oligotrophic ecosystems can be loosely defined as environments that exhibit low ambient nutrient levels that are thus suboptimal for biological production. Additionally, oligotrophic systems are often characterized by a low nutrient flux (89, 193). Environments which meet these two criteria can be considered as oligotrophic environments, and can be found in water distribution systems (26), soil ecosystems (227), and air ecosystems (234). The microbiology of these environments is interesting because these are normally correlated with extreme conditions where relative special microbes can be detected. During my thesis, I had the opportunity to study bacterial diversity in three different environments, including one soil system (surface desert soil), one air system (the dust in air) and one water system (drinking water distribution system). These environments can all be classified as oligotrophic environments. Here, I will begin with an introduction to these three environments.

I. A. a. The desert

Approximately 33% of the Earth land surface is classified as desert regions (21). In 1952, Meigs estimated that 19% ($27.6 \times 10^6 \text{ km}^2$) of global land is composed by arid deserts and 14.6% ($21.2 \times 10^6 \text{ km}^2$) covers area defined as semi-arid deserts, excluding Antarctica (150). However deserts seems never to stop growing and, between 1975 and 1987, the desertification rate in China was $2,100 \text{ km}^2/\text{year}$ (244). An examination of the distribution of deserts (Fig. 1); deserts exist in all continents except Europe. Even in the polar regions, the low level of available water (most water there is frozen), are considered as regions belonging to deserts (62, 146).

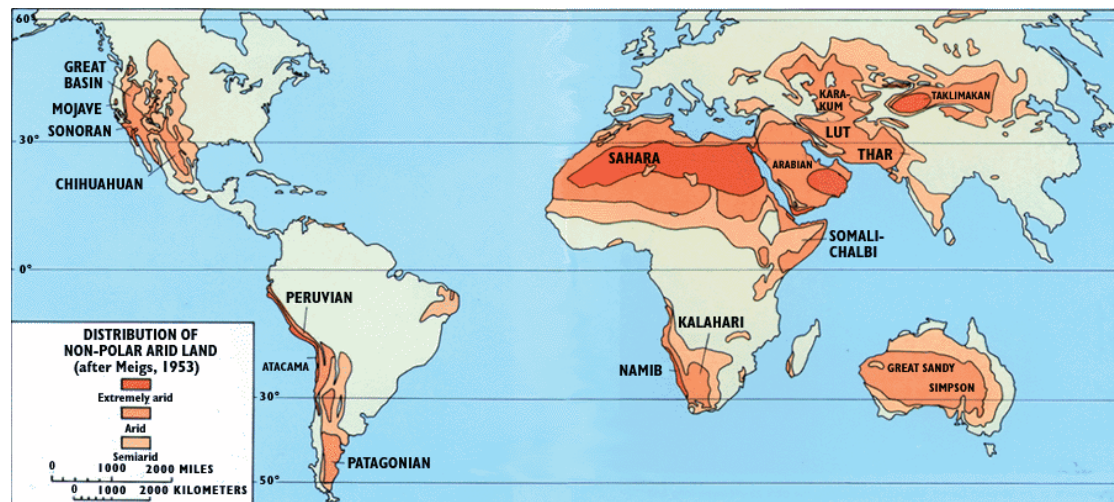


Figure 1: Map of the global extent of hyperarid, arid, and semi-arid regions and principal deserts. Generated by U. S. Geological Survey after Meigs, 1953 (151).

The desert regions on Earth vary in terms of their landscape, fauna, flora, water balance, and temperature. As a result of these variations, it is difficult to derive an exact definition of a desert environment. In general, deserts are defined in terms of aridity. There are many aridity indexes to measure the level of aridity for a region. One aridity index, adopted by UNEP (United Nations Environment Programme, 1992), P/ETP (where P = annual average precipitation, ETP = potential

evapotranspiration), is now frequently used to distinguish arid regions levels. This index is derived from monthly data on temperature and precipitation (P) over the period 1951-1980 from a worldwide network of meteorological stations. From the temperature data, together with monthly data on daylight hours, potential evapotranspiration (4) is calculated. In 1997, the UNEP World Atlas of Desertification (UNEP, 1997) classified deserts on the basis of the Aridity Index P/PET. Desert refers to regions with P/PET values less than 0.20, which includes both the arid and hyper arid land, and where the regional precipitation is less than 250 mm per year (21). Deserts are created by a lack of precipitation and not necessarily high temperature. The Antarctic is considered a desert, with annual precipitation of only 200 mm along the coast and far less inland (204). Soils that form in desert climates are predominantly mineral soils with low organic matter content. However, in certain desert soils, the repeated accumulation of water can cause salts to precipitate. When the water table rises to within about 2m of ground level, water may begin to rise to the surface by capillary action. When a rising water table intersects with salts, the salt will dissolve and be carried up to the surface, concentrating in the upper layers of the soil as water is evaporated. Most playa lakes (found in interior desert basins and adjacent to coasts within arid and semiarid regions) will consequently be highly saline.

Desert environments appear very monotonous, but they also have diversity levels. Using different parameters, deserts can be classified in to different groups. By temperature variation, deserts on Earth could be classified as cold deserts and hot deserts. The central parts of North American deserts have far higher temperatures and evaporation than African and Middle Eastern deserts, for example, leading to more extreme conditions (59). Besides temperature, there are five different landscapes in different deserts, called Sand desert (Fig. 2a), Stony desert (Fig. 2b), Rock desert (Fig. 2c), Plateau desert (Fig. 2d) and Mountain desert (Fig. 2e).

The effects of temperature in deserts are widely known. However, little emphasis has been placed on the large differences among deserts in ambient temperature (and seasonality) and how these differences affect the organisms that live there. Due to the lack of available moisture in deserts, plants are sparse or completely absent, creating

soil conditions with extremely low soil organic carbon and nitrogen levels, further limiting the potential diversity of microbial life (36, 148, 158). Examples of total organic carbon levels in a range of desert soil samples include 0.02-0.09% in samples taken along an elevational transect through the hyperarid core region of the Atacama Desert (48), 0.11%-0.7% in samples taken from surface sand of the Taklamaken Desert and 0.03%-0.05% in Gobi Desert in our study. By comparison, organic carbon in temperate soils ranges from 1 to 5% (191).



Figure 2: The desert landscape types: a. Sand desert, b. Stony desert, c. Rock desert, d. Plateau desert, e. Mountain desert. a,c,e were taken from Warhol, 2007 (225), b,d were from Johnson, 2003 (99).

I. A. b. Sandstorms

As humans, we have an intimate relationship with the air around us (234). The air that we breathe is like other visible environments such as water and soil. It has different physical and chemical parameters such as temperature, pressure and UV radiation exposure, etc. The air conditions can be influenced by weather, and Sandstorms can have large impact on air conditions.

Sandstorms are often generated in arid regions by high-energy winds. The two primary locations of sandstorm include the Sahara and Sahel regions of North Africa and the Gobi, TaklaMakan deserts of Asia (80). Every year, large quantity of the mobilized desert top soils move great distances through the atmosphere (72). It is estimated that 2×10^{12} to 5×10^{12} of desert sand dust are transported by wind each year in the world (177, 200). The migration of sand can be regional, or global. African sand dust from the Sahara desert and Sahel region can affect the air quality of the Middle East, Europe, Asia, the Caribbean, and the Americas (34, 70, 71). Asian sand dust from the Gobi, Taklamaken and Badain Jaran deserts can cross eastern Asia, the Pacific ocean, arrive in North America (24, 49), and even reach Europe (74). They can strongly impact the air quality of downwind regions by the small particles that abound including soot, salt, bacteria, fungal spores, and interstellar dust.

We have studied the sand-associated bacteria of sandstorms in Asia. Asian sandstorms seasonal phenomena, occurs every year between February and May. Asian sandstorms have a major impact on the air quality of the densely populated areas of China, Korea and Japan, and are important in the global dust cycle (197).

Concerns about health impacts due to changes in concentrations, size distributions and chemical composition of ambient pollutants and sand-associated microbes during sandstorms have been raised in recent years.

In Seoul, researchers studied the daily mortality between 1995 and 1998, with 28 Asian dust days. The percentage of increased deaths during the Asian dust days were 1.7% for all causes, 2.2% for deaths of persons aged 65 years and older, and 4.1% for

cardiovascular and respiratory causes. The results suggest that persons with advanced cardiovascular and respiratory disease may be more susceptible to Asian dust events (112). The increase of daily deaths due to Asian dust events were later confirmed using data from 2000 to 2004 in Seoul (121). In others cases, positive but not significant relations were reported to vulnerable population health during Asian dust days (236-238).

Most studies of the health impacts of dust storms in Taiwan have been based on hospital admission data. Children are considered to be particularly sensitive to ambient pollutant exposure (188). Based on the data registered in hospitals and clinics in Taipei City during 1997-2007, the rate of clinic visits for children increased 2.54% for preschool children (≤ 6 years of age) and 5.03% for school children (7-14 years of age) during Asian dust storms, compared with weeks before Asian dust storm events (32).

From January 1980 and May 2009, 19 epidemiologic studies of dust effects (3 for mortality, 13 for hospital visits or admissions and 3 for respiratory functions or symptoms) were performed, mainly from Taiwan and South Korea. The evidence from these studies was limited because exposure assessments were inadequately described and potential confounders were insufficiently controlled. Well-designed epidemiological studies are required to clarify any potential health effects of Asian dust events (83). Among these epidemiologic studies, the effects of sand-associated microbe are rarely mentioned. And the effects of sand-associated microbes may not be observed for days or weeks after a sandstorm. The long term studies are needed.

I. A. c. The drinking water distribution system

Among the different aquatic ecosystems, many can be considered as oligotrophic ecosystems, such as the deep ocean (16, 198, 235) and low level nutrient lakes (211), while drinking water systems are considered as an ultra oligotrophic ecosystem (230).

The sources of drinking water may come from natural surface water such as rivers and lakes, or groundwater. Potable water quality continues to be a major focus of environmental microbiology because new waterborne pathogens continue to emerge. Recent studies in 2009 suggest that 10 to 50% of diarrhea-associated illness is caused by waterborne microbial agents not yet identified (141).

Unlike the two previous environments (desert and sandstorm) mentioned, drinking (potable) water systems are artificial. Water from the source needs to undergo water treatment before arrival at the tap. Therefore, the environment of drinking water systems is well controlled and should meet the drinking water quality standards required by the World Health Organization (WHO). There are different treatment processes for drinking water in different regions. Different water sources may have different treatment solutions. They generally contain several programmed purification steps including filtration, disinfection, and chemical modification. The step of disinfection is to inactivate (kill) bacteria and viruses in water. In Europe, disinfectants used include chlorine, chloramines (chlorine plus ammonia), ozone, ultraviolet light, and chlorine dioxide. Ozone disinfection for drinking water was started in 1906, and has many advantages, including high-speed, high efficiency and low energy consumption, plus, no secondary pollution during the disinfection step. In drinking water standards, there are two cultivable indicators for pathogen control: *Escherichia coli* and *Enterococcus*. Using conventional culture methods, they should not be detected in sample of 100 ml. The detection of *E. coli* provides definite evidence of fecal pollution. But in practice, the detection of thermotolerant (fecal) coliform bacteria is an acceptable alternative (140). There are also many other standards such as up to 24 chemical indicators and physical aspect levels, according to the law in Europe. The

most discussed aspects include turbidity, pH, and chlorine levels. To ensure that a drinking water supply satisfies these guidelines, samples should be regularly examined. As drinking water safety is closely related to human health, the microbial diversity in drinking water systems has received more study than the previous two environments (deserts and sandstorms). The guidelines for drinking-water quality published by the World Health Organization (WHO) in 1997 mentioned that the principal risks to human health associated with the consumption of polluted water are microbiological in nature in most countries, although chemical contamination is also important (140). The level of contamination by pathogenic microorganisms in water directly increases the risk of acquiring a waterborne infection. An estimated 80% of all diseases and over one-third of deaths in developing countries are caused by the consumption of contaminated water (1). Outbreaks of waterborne infectious diseases via the use of contaminated drinking water still poses a serious health threat worldwide, despite the fact that drinking water is one of the most closely monitored and strictly regulated resources. Careful estimates indicate that each year about 350 million people are infected by waterborne pathogens, with 10 to 20 million succumbing to severe cases of infection (229). In the United States, it is reported that 126 drinking water-related disease outbreaks, 429,000 cases of illness, 653 hospitalizations, and 58 deaths occurred during the years 1991–1998 (43). It is clear that fast and accurate monitoring of chemical and microbiological parameters in drinking water is essential to safeguard the consumer and to improve the understanding of treatment and distribution systems (20).

I. B. Environmental microbiology

I. B. a. General introduction

Microbes play a critical role in human and environmental health. The more microbial populations are explored, the more complexity and diversity is uncovered. The biosphere contains between 10^{30} and 10^{31} microbial genomes, at least 2–3 orders of magnitude more than the number of plant and animal cells combined (228). Microbes control global utilization of nitrogen through nitrogen fixation, nitrification, and nitrate reduction, and drive the bulk of sulfur, iron and manganese biogeochemical cycles (162). They regulate the composition of the atmosphere, influence climates, recycle nutrients, and decompose pollutants. Without microbes, multi-cellular life on Earth would not have evolved and biology as we know it would not be sustainable (91). Environmental microbiology has expanded to the study of soil, water and air systems, including the interaction of indigenous microbes with organic and inorganic pollutants (182), the behavior of pathogens introduced into these systems (12, 173, 174, 181) and the discovery and application of new microbes and their products to benefit human health and welfare (101, 178, 216). The initial scientific focus of the field of environmental microbiology was on water quality and the fate of pathogens in the environment in the context of protection of public health (141). In modern environmental microbiology, pathogens and biomediation remain fundamental to the field (141). Over the past decade, there has been a virtual explosion of knowledge on microbial diversity and communities (2, 75, 106, 113, 234). The research has been greatly enhanced through the application of molecular and biotechnology tools (5, 141, 168, 205).

I. B. b. Microbial diversity

Microbial diversity is important to study, because it is fundamental to the maintenance and conservation of global genetic resources. Colwell (1997) indicated 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 (38). The exploration of microbial diversity has been reported in different systems, including soil (6), air (234) and aquatic environments (88, 113). Many studies have also been carried out on the human skin surface (28, 108), the human digestive system (47, 61) and the human oral environment (109).

Microorganisms can adapt to environmental variations much faster than other multi-cellular organisms. They can be detected in every extreme environment imposed by oligotrophy (low nutrients), temperature, pH, pressure, and radiation (141). In fact, microorganisms are pioneer colonizers and have had, over geologic time, a profound influence on the climate and environments found on Earth (Fig. 3) (141). As extreme environments are explored, the richness of microbial diversity is becoming increasingly evident (2, 6, 179).

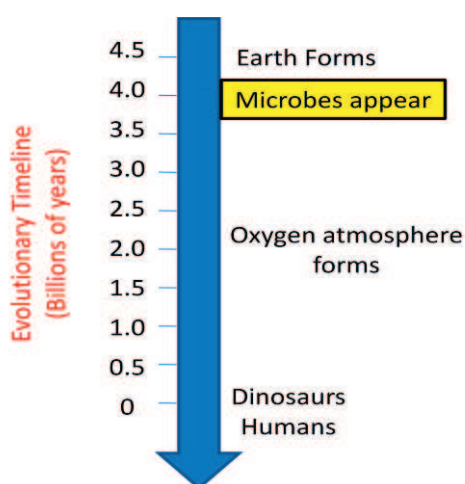


Figure 3: An evolutionary timeline showing the approximate appearance of life from 4.5 billion years ago to the present time (141).

Modern environmental microbiology has been greatly enhanced through the application of molecular genetics and biotechnology tools. Molecular-based tools are becoming available to allow the examination of microbial communities through analysis of microbial DNA and RNA, as well as proteins. The new generation techniques now allow us to search for new microbes in extreme environments such as hot springs, caves, deep-sea thermal vents without a culture step. Molecular sequence analysis of community DNA permits a new appreciation of microbial diversity (141). Among the microbiological community studies, the bacteria community has been mostly characterized in every environment type (136). It is estimated that there are more than 50 bacterial phyla based on the analysis of conserved 16S rRNA sequence (191).

During my thesis, I focused on bacterial diversity in three different oligotrophic environments, in order to begin to characterize their bacterial communities.

I. B. c. Technology tools

I. B. c. 1. Identification of bacteria

Until the 1970s, the classification of microorganisms was based primarily on physiological differences. The initial research on bacterial diversity was only for the cultivable stains. The bacteria needed to be isolated and cultured under laboratory conditions. The differences among bacteria were detected by their morphology, differential staining (typically begins with a gram stain) and biochemical tests; since so many bacteria look so much alike, it is frequently necessary to resort to biochemical tests for identification, along with phage typing using viruses which attack bacteria (bacteriophages). The database of phenotypic characteristics is limited to common species only (202, 209) and is not necessarily suitable for environmental isolates.

Culture-based methods are important in investigating the microbial ecology of natural and anthropogenically impacted environments, but they are extremely biased in their evaluation of microbial genetic diversity by selecting a particular population of microorganisms. They would miss most microbes and hence were not suitable for contemporary studies.

However, in the 1970s, techniques became available to allow the examination of nucleic acids; DNA-based identification and classification of the microorganism became possible (141). The application of molecular technique to define the bacteria classification can overcome the limitation of cultivable bacteria, as we know that only less than 1% of the total bacteria grow on conventional media (103).

For a long time, DNA-DNA hybridization was often used as the “gold standard” allowing the proposition of new species and definitive assignment of a strain with ambiguous properties to the correct taxonomic unit. Based upon DNA-DNA reassociation kinetics, the genetic definition of a species is quantifiable, i.e., (i) over 70% DNA-DNA relatedness and (ii) 5°C or less T_m for the stability of heteroduplex molecules. DNA hybridization assays are not without their shortcomings, however, being time-consuming, labor-intensive, and expensive to perform. Today, fewer and

fewer laboratories worldwide perform such assays, and many studies describing new species are solely based upon small subunit (SSU) 16S rDNA sequences or other polyphasic data (173). Particularly, ribosomal RNA or DNA sequencing is the basis for the classification of a microorganism and has now become the new gold standard for defining new genera and species (212, 215).

I. B. c. 2. 16S rDNA

The majority of research on the microbial populations now focuses on the analysis of the 16S rRNA gene (16S rDNA), which is believed to change less than other nucleic acids over time. A study of the similarity of rRNA of two different organisms is believed to show how far back in the past the two organisms branched off a common ancestor. The use of 16S rRNA gene sequences to study bacterial phylogeny and taxonomy has been the most common housekeeping genetic marker used for three reasons:

- I**, its presence all bacteria, often existing as a multigene family, or operons;
- II**, the function of the 16S rRNA gene over time has not changed, suggesting that random sequence changes are a more accurate measure of time (evolution);
- III**, the 16S rRNA gene (1,500 bp) is large enough for bioinformatic purposes (173).

Compared to other genes, the 16S rRNA gene has a number of clear advantages when it comes to diversity analyses:

- I**, it has highly conserved regions that permit effective PCR primer design,
- II**, regions of the gene are sufficiently variable to allow for accurate taxonomic and phylogenetic identification of community members,
- III**, since this gene has been widely sequenced in microbial diversity surveys, there is a large amount of accumulated 16S rRNA sequence data in databases that permit more accurate taxonomic identification and comparisons of community composition across studies (57).

The use of 16S rRNA gene sequencing is becoming commonplace for identifying biochemically unidentified bacteria or for providing reference identification for unusual strains (173).

In 1977, based on analysis of 16S rRNA, Carl Woese (233) identified an entirely new group of organisms – the Archaea – which eventually led to the modern classification of living things into a three-domain system consisting of Archaea, Eucarya, and Bacteria (Fig 4).

Bacterial 16S ribosomal RNA (rRNA) genes contain nine “hypervariable regions”

(V1–V9) (9), and it is commonly thought that these regions demonstrate considerable sequence diversity among different bacteria. In many cases, one or two variable regions were chosen to be sequenced instead of the total length of 16S rDNA (92, 155). But as no single region can differentiate among all bacteria, systematic studies that compare the relative advantage of each region for specific research goals are needed (30). The nine discriminating regions of 16S rRNA are shown in Fig. 5 of 16S rRNA (19).

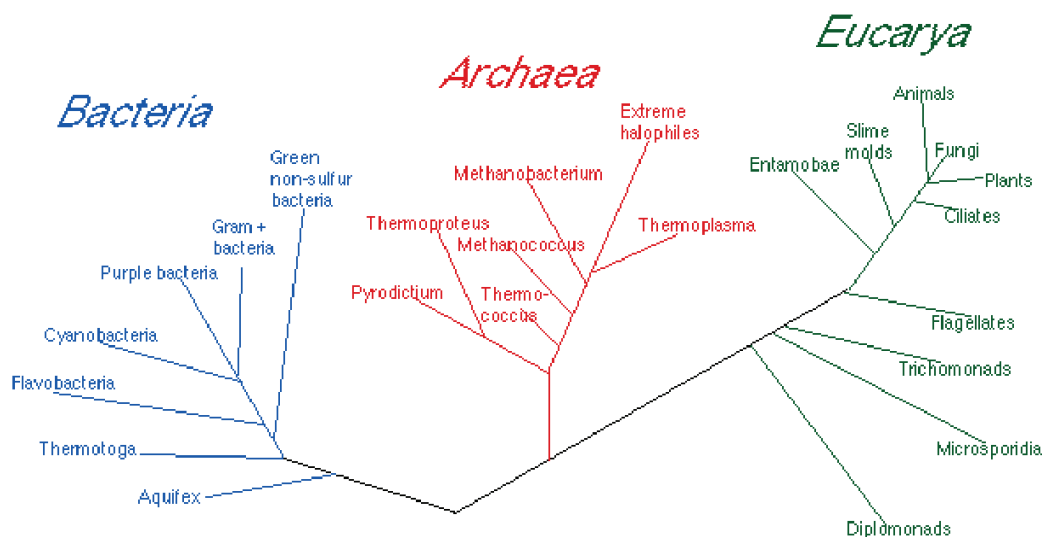


Figure 4: Universal phylogenetic tree, based on the comparison of sequence analysis of the 16S ribosomal RNA (140).

However, while 16S rDNA is the most popular biological molecular tool in modern microbial identification research, there are some difficulties in some situations:

I: In some cases, there is very little sequence variation observed between the 16S rRNA genes of closely related microorganisms, as Harmsen et al reported its failure to discriminate several similar species from same genus (82). 16S/23S ribosomal spacer region was proposed as a target for DNA probes to identify eubacteria (14).

II: The current GenBank database has a collection of 883 prokaryotic genomes representing 568 unique species, of which 425 species contained 2 to 15 copies of 16S rRNA genes per genome (2.22 ± 0.81) (175). In the case of *Aeromonas veronii*, the genome can contain up to six copies of the 16S rRNA gene that differ by up to 1.5% among themselves. This implies intragenomic heterogeneity of the 16S rRNA gene among aeromonads and could preclude the use of this technology alone for species

identification (173).

III: In addition there is always a debate as to whether it is enough to use only one molecule to identify one species. Sometimes, whole-genome molecular techniques are performed to offer a more comprehensive view of genetic diversity compared to PCR-based molecular approaches that target only a single or a few genes.

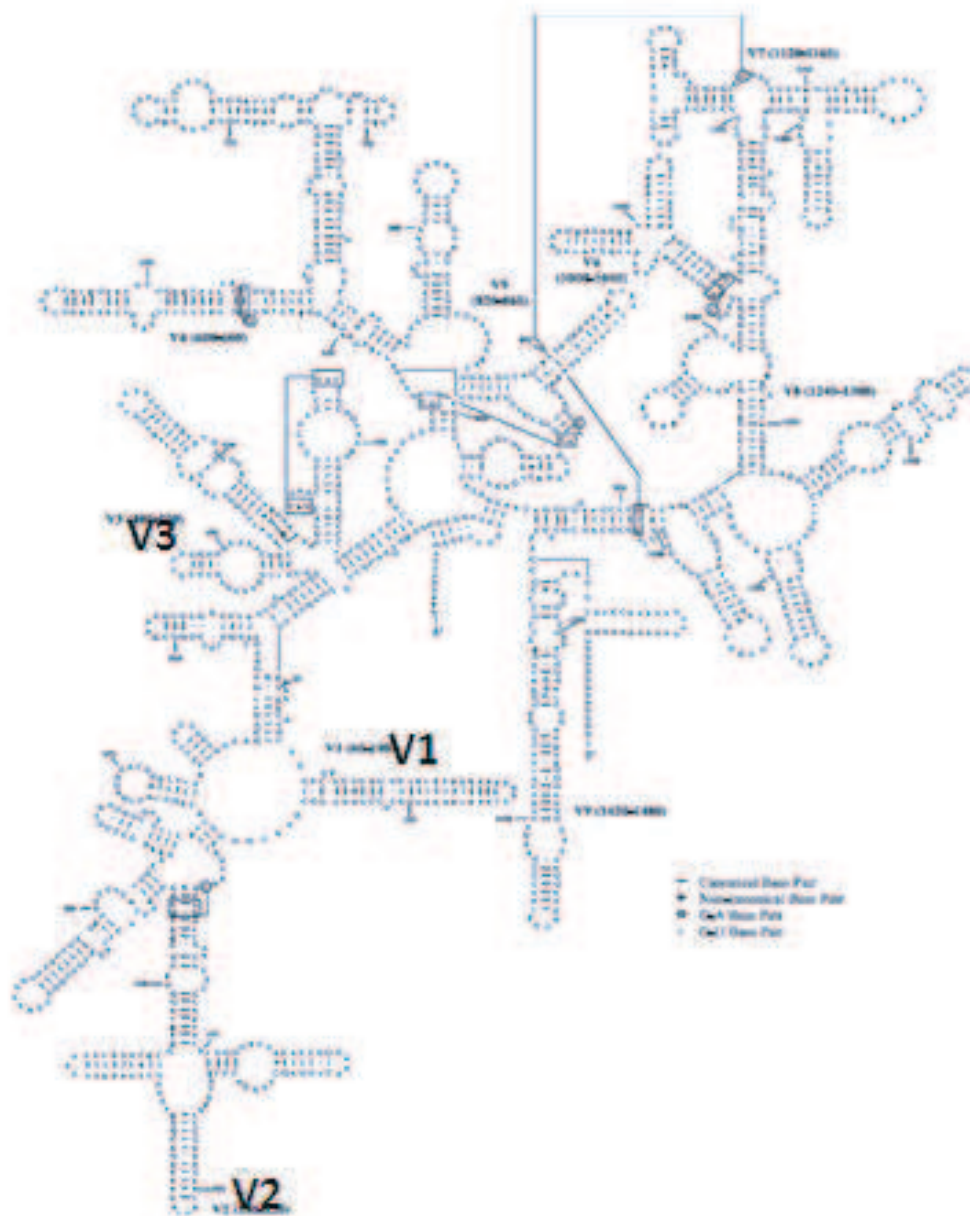


Figure 5: 16S rRNA secondary structure with 9 variable regions (77).

I. B. c. 3. Genetic fingerprint methods

Microbial fingerprinting methods are a category of techniques that differentiate microorganisms or groups of microorganisms based on unique characteristics of a universal component or section of a biomolecule (e.g., phospholipids, DNA, or RNA). Community fingerprinting is used by microbiologists studying a variety of microbial systems (e.g. marine, freshwater, soil, and human microbial communities) to measure biodiversity or track changes in community structure when environmental conditions change. Microbial fingerprinting methods provide an overall profile of the microbial community, and some can be used to identify subsets of the microorganisms present. Fingerprint techniques, as one of PCR-based methods, contain a variety of approaches corresponding to different analysis procedures of PCR products.

Polymerase Chain Reaction (PCR) -- when small amounts of the DNA of an organism are obtained, PCR can be used to increase the amount so that it can be tested. Since most species of microorganisms are not cultivable in the laboratory, the investigation of their communities is time-consuming using traditional culture procedure. In the past decades, many kinds of culture-independent methods have been developed. The advent of Polymerase Chain Reaction (PCR) by Kary B. Mullis (Nobel Prize in chemistry in 1993) in the mid-1980s revolutionized molecular biology as we know it (15). PCR is a fairly standard procedure now, and its use is extremely wide-ranging. PCR-based methods (community fingerprinting techniques) are currently the most widely used tools. This technique is normally combined with DNA fingerprinting or Ribosomal RNA or DNA sequencing.

Established since the 1990s, DGGE and T-RFLP are the most frequently used DNA fingerprint approaches on bacterial community analysis. DGGE and RFLP were widely used from the 1990s' for microbial community diversity without culture steps.

Denaturing Gradient Gel Electrophoresis (DGGE) is a microbial fingerprinting technique that separates amplicons of roughly the same size based on DNA sequence properties. These properties dictate the threshold at which DNA denatures. The DGGE

gel uses a gradient DNA denaturant (a mixture of urea and formamide), or a linear temperature gradient. When the fragment reaches its melting point (threshold of enough denaturant), it stops moving. This is due to the fact that a partially melted double-stranded DNA can no longer migrate through the gel. A GC clamp (about 40 bases with high GC content) is used as a special primer to anchor the PCR fragments together once they have denatured. DGGE is a nucleic acid (DNA or RNA)-based technique used to generate a genetic fingerprint of the microbial community and potentially identify dominant microorganisms. DGGE profiles are most often used to compare differences or changes in microbial community diversity and structure between samples, over time or space or in response to treatment. Each band in a different location on the gel represents a different phylotype (one unique sequence of a phylogenetic marker gene). For microbial communities, this method profiles many individual 16S rRNA sequences. The number of bands at differing horizontal positions can be used to estimate the level of biodiversity in that sample and infer phylogenetic affiliation.

After DGGE, the subset of the individual “bands” can be excised (physically cut) from the gel, the DNA sequence is determined for each excised band, and the resulting DNA sequence is compared to a database to identify the microbial population corresponding to each band (156). However, since individual “bands” are excised from the gel for sequencing, typically only 3–10 microorganisms are identified by DGGE analysis. DGGE is widely used to describe phylogenetic relationships between microbes since it can supply accurate and abundant information on biodiversity of a targeted ecosystem. DGGE was first used to detect DNA mutations in 1993 (156), and it is by far one of the most popular molecular techniques in microbiology (64). DGGE was applied to investigate the variation of microbial diversity among natural environments, including soil (167), desert (48) and the rhizosphere (50), etc.

This method does not require that the microbes be cultured in the lab and does not require any sequence data needed to design probes for hybridization methods. The main disadvantage is that this is a qualitative assessment of biodiversity and one must sequence the genes in order to make inferences about phylogenetic relatedness. Most

often, DGGE analysis is performed when identification of the predominant organisms is required but little is known about the microbial community of the sample prior to analysis.

Terminal restriction fragment length polymorphism (T-RFLP) is a method that uses fluorescently-labeled DNA fragments to produce a community fingerprint. To perform T-RFLP, one must select a target gene (e.g. 16S rRNA gene) to amplify by PCR. At least one primer used in the PCR reaction is fluorescently labeled at the 5' end. After PCR amplification, each copied DNA segment carries the fluorescent label. Next, restriction enzymes are used to cut the amplified DNA at specific recognition sites. The underlying assumption is that each microbe in the sample will have a different sequence on the target gene, so a restriction enzyme will cut each microbe's DNA in a different place; each different restriction site is considered to represent a single operational taxonomic unit (OTU). Thus, the enzyme will produce one fragment length for each type of microbe present in the sample. The result of digestion is a set of restriction fragments of different lengths, each of which is fluorescently labeled at one end. These are known as "terminal fragments" because they are labeled at the end where the PCR primer is attached. Next, the fragments are separated by size either through gel or capillary electrophoresis. Laser detection captures the size and fluorescence-intensity patterns of the terminal fragments. DNA standards of known size and fluorescence are included in the analysis as references.

The number of microorganisms that can be identified by T-RFLP can be 10 times greater than using DGGE, providing more comprehensive examination of the microbial community composition. The major advantages of T-RFLP are that it is fast and can easily accommodate many samples. Also, the visual output simplifies comparison of community structure patterns across different samples. The theoretical basis of T-RFLP assumes that peaks at different positions along the horizontal axis represent different types of organisms (or OTUs). However, a number of caveats must be taken into account. Different types of organisms may share a restriction site in the gene of interest. If that is the case, these organisms would not be distinguished as different peaks on the electropherogram. Furthermore, the area under a peak represents relative abundance

rather than absolute abundance, and there are biases in abundance measurement and PCR amplification. For example, organisms that are scarce in the original total DNA sample will not be amplified enough to be detected in the final analysis. This leads to an underestimation of community diversity. Liu *et al.* cite other possible factors that may distort results, including “differences in gene copy number between species and biases introduced during cell lysis, DNA extraction, and PCR amplification” (131). Also, direct identification of microbes in a sample is not possible through T-RFLP.

DGGE and T-RFLP provide a DNA-based profile of the microbial community and allow identification of the predominant organisms generally to the family or genus levels, but cannot quantify specific organisms or microbial functions.

Besides the DGGE and T-RFLP, other fingerprint methods may be sometimes used such as:

Ribosomal Intergenic Spacer Analysis (ARISA) Between the two genes 16S and 23S, there is an internal transcribed spacer (ITS) region. Due to the fact it is non-coding for proteins, it is a highly variable in nucleotide sequence and length. Once DNA is isolated from a community, PCR amplifies this spacer region. The fragments can be run on a gel (RISA), or the fluorescent primers can be translated into peaks in abundance of the different fragment lengths on an electropherogram (ARISA).

Single-strand conformation polymorphism (SSCP) fingerprinting of 16S rRNA was applied also in compare the bacterial community structure (51). SSCP is the electrophoretic separation of single-stranded nucleic acids based on subtle differences in sequence (often a single base pair) which results in a different secondary structure and a measurable difference in mobility through a gel.

These different techniques give us a number of choices for bacterial diversity analysis. Their advantages and disadvantages were listed in Table 1.

Table 1: *Avantages and disadvantages of different footprinting techniques adopted from (93)*

	DGGE	T-RFLP	ARISA	SSCP
Avantages	The bands of interest can be excised from the gel. Not expensive.	High sensitivity fast. Allows a comparison between different tests.	Can detect very slight differences in population Compatible with RFLP and sequencing for further analysis	No need for primer with GC clamp. The bands of interest can be excised from the gel. Compatible with automated analysis (CE-SSCP).
Disavantages	limited sensitivity The GC clamp reduces the profitability of PCR and promotes the formation of dimer Manipulation of gel requires experience	Incomplete restriction leads to an overestimation of a species Several restrictions are necessary for complete analysis The complex patterns are difficult to interpret phylogenetically	A single species can have multiple signals	High rate of re-pairing of DNA strands. A species may have several bands.

1. B. c. 4. Conventional sequencing methods

While the DNA fingerprint techniques revealed different bacterial communities from samples on gels, Sanger sequencing is often used to identify the dominant or interesting bands.

Sanger sequencing was named after its developer -- Frederick Sanger (Nobel Prize in Chemistry in 1980) (189). Sanger sequencing is a chain-terminator method. It is the process for the extension of complementary strand for the single-stranded DNA template using a chain terminating nucleotide (most commonly a di-deoxynucleotide) mixed with other normal nucleotides. The incorporation of the chain terminating nucleotide results the end of extension. Each reaction has added only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP) which are the chain-terminating nucleotides. A series of related DNA fragments are then obtained that are terminated only at positions where that particular nucleotide is used. The fragments are then size-separated by electrophoresis in a slab polyacrylamide gel, or more commonly now, in a narrow glass tube (capillary) filled with a viscous polymer. Sequence could then be read based in the position of bands. The sequences obtained by sequencing techniques are then aligned to a known reference sequence for identification (153). Today, 750-1000 base pairs of sequence can now be read from a single reaction by Sanger sequencing. The sequencing is now automated. The automated Sanger method is considered as a 'first-generation' technology, and newer methods are referred to as next-generation sequencing (NGS) (153).

I. B. c. 5. Next generation sequencing methods

With the development of microbial metagenomics, scientists have attempted to analyze all the genetic information present in total DNA extracted from an environmental sample or pure culture. Even for PCR-based methods, the quantities of DNA sequencing become a limit for the studies of microbial communities. Next generation sequencing was developed in the 1990s, but only after 2004 has their application greatly augmented (19).

The arrival of NGS technologies in the marketplace has changed the way we think about scientific approaches in basic, applied and clinical research. In some respects, the potential of NGS is akin to the early days of PCR, with one's imagination being the primary limitation to its use. The major advance offered by NGS is the ability to produce an enormous volume of data cheaply — in some cases in excess of one billion short reads per instrument run.

Next-generation sequencing (NGS) encompasses a number of different methodologies that have emerged since 2005. In numerous NGS methods, fragmented genomic DNA ligated to universal adaptors are amplified, on a solid phase, into PCR colonies of DNA or “polonies” (54). Each polony contains many copies of the same fragment, and all of the polonies can be sequenced in parallel using arrays allowing millions of reads per array. The general advantages of second-generation sequencing over Sanger sequencing are three-fold. First, since the preparations are done *in vitro*, bottlenecks like transformation of *E. coli* are avoided. Second, there is increased parallelism in second-generation methods because they are based on arrays rather than capillaries, which significantly reduces sequencing time. Third, since the polonies are all bound to the same array, they can be treated with single reagent volumes rather than multiple independent volumes, thereby dramatically cutting costs (76).

The emergence of next-generation sequencing (NGS) technologies offers an incredible opportunity to comprehensively study DNA sequence variation in different environments. Commercially available platforms from Roche (454), Illumina (Genome

Analyzer and Hiseq 2000), and Applied Biosystems (SOLiD) have the capability to completely sequence individual genomes to high levels of coverage (107).

The Roche 454 Genome Sequencer FLX utilizes massively parallel pyrosequencing of DNA fragments that are amplified en masse by emulsion PCR - amplifying complex DNA mixtures, based on the compartmentalization of genes in a water-in-oil (w/o) emulsion (143, 144, 232). The current FLX Titanium chemistry produces up to 1.25 million reads per run, with read lengths of ~400 nt. A single run yields almost half a billion nt (0.5 gb) of high-quality sequence. While the 454 platform is prone to indel errors near runs of multiple nucleotides (homopolymers), the substitution error rate is very low, and the relatively long reads are well-suited to alignment or *de novo* assembly.

Pyrosequencing uses an enzyme-cascade system, consisting of four enzymes and specific substrates, to produce light whenever a nucleotide is incorporated to form a base pair with the complementary base in a DNA template strand. The amount of light is proportional to the number of incorporated nucleotides. Each time, one type of nucleotide is added. When the light signal is detected, the base is registered and the next nucleotide is added. If the added nucleotide is not complementary to the next base in the template, no light will be generated.

The Illumina (Solexa) Genome Analyzer IIx utilizes sequencing-by-synthesis of surface-amplified DNA fragments (19) . While the reads produced on the Illumina platform were initially quite short (~32-40 bp), the current instrument yields ~500 million 100-bp reads (50 Gb) per run. The astonishing sequence throughput of Illumina machines present substantial bioinformatics challenges, particularly the alignment of short sequences of imperfect quality to large reference sequences. Indeed, an entire generation of novel algorithms (Maq, BWA, Novoalign, Bowtie, and others) has been developed to address the analysis challenges of Illumina sequencing.

The Applied Biosystems SOLiD sequencer uses a unique process catalyzed by DNA ligase, in which oligo adapter-linked DNA fragments are coupled to magnetic beads and amplified by emulsion PCR (149). The current instrument (SOLiD 3) produces up to 50 gb of high-quality sequence per slide in the form of 35 or 50 nt reads. A unique

advantage of the ABI SOLiD platform is its di-base encoding scheme, in which each base is effectively called twice in a sequencing read. The availability of two calls per base makes it possible to distinguish between sequencing errors and true variation, thereby improving the overall accuracy of reads from this platform.

Among the three NGS technology, 454 pyrosequencing is mostly applied because of its longer sequence reads. As paired reads are now available for 454 GS FLX, Solexa and SOLiD platforms, and read lengths are increasing for the higher-throughput technologies, it is likely that they will continue to play a part in future metagenomics projects (139).

Bioinformatic for NGS

The massive data produced by NGS also presents a significant challenge for data storage, analyses, and management solutions. Advanced bioinformatics' tools are essential for the successful application of NGS technology. As pyrosequencing of 16S rDNA was used in my research, I present the process of analysis of high-throughput 16S rDNA data and the analysis tools in Fig 6.

The analysis of NGS 16S rDNA data can be separated into several parts: First, the selection of good quality sequences. Due to the large quantity sequences obtained per run, the first job is to eliminate the poor quality sequences to ensure a good interpretation. During sequencing, the accuracy of each nucleotide is indicated by Phred quality scores which we can use to trim the sequences. Once the data is clean, the sequences can be classified by database BLAST. Several good databases are listed in Fig 6. Meanwhile, some statistical analyses can be carried out, such as rarefaction curves and the Chao1 estimator (widely used for total OTU number estimation) for OTU number. One extra step – normalizing the number of sequence to the same level for every sample - can lead to a normalized comparison among them. After the analyses for each sample, different similarity indices can be calculated to show the relations among samples.

Raw partial 16s DNA sequences data	Software to use
Quality Control Trim sequence	RDPII pyro pipeline: http://pyro.cme.msu.edu/ Greengenes Trim: http://greengenes.lbl.gov/cgi-bin/JD_Tutorial/nph-16S.cgi FASTX-Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/ Pangea : http://www.microgator.org/pangea/ PyroTagger: http://pyrotagger.jgi-psf.org/cgi-bin/index.pl ngs_backbone: http://bioinf.comav.upv.es/ngs_backbone/index.html
ChimeraCheck	B2C2: http://www.researchandtesting.com/B2C2 Chimera.slayer: http://www.mothur.org/wiki/Chimera.slayer Bellerophon: http://comp-bio.anu.edu.au/bellerophon/bellerophon.pl QIIME: http://qiime.org/tutorials/chimera_checking.html
Chloroplast Mitochondria check	RDPII Classifier: http://rdp.cme.msu.edu/classifier/classifier.jsp Metaxa: http://microbiology.se/software/metaxa/ BLAST: http://lslab.lscore.ucla.edu/Mitochondria/db/db.htm
Identification Classification	RDPII Classifier: http://rdp.cme.msu.edu/classifier/classifier.jsp TaxCollector: http://www.microgator.org/taxcollector/ MEGAN: http://ab.inf.uni-tuebingen.de/software/megan/ MG-RAST: http://metagenomics.anl.gov/metagenomics.cgi?page=Home Mothur: http://www.mothur.org/wiki/Classify.otu VITCOMIC: http://mg.bio.titech.ac.jp/vitcomic/
Statistical analysis	RDPII pyro: http://pyro.cme.msu.edu/abund/form.spr MG-RAST: http://metagenomics.anl.gov/metagenomics.cgi?page=Home QIIME: http://www.qiime.org/ Mothur: http://www.mothur.org/wiki/Rarefaction.single SPADE: http://chao.stat.nthu.edu.tw/softwareCE.html Pangea: http://www.microgator.org/pangea/
Relations among samples	RDPII pyro: http://pyro.cme.msu.edu/abund/form.spr FastUnifrac: http://bmf2.colorado.edu/fastunifrac/ MG-RAST: http://metagenomics.anl.gov/metagenomics.cgi?page=Home Ade4 in R: http://pbil.univ-lyon1.fr/Rweb/Rweb.general.html QIIME: http://www.qiime.org/ SPADE: http://chao.stat.nthu.edu.tw/softwareCE.html

Figure 6: Process for NGS 16S rDNA data analysis

Metagenomics

Sampling signature genes, such as 16S ribosomal RNA, does not provide much insight into the composition and activities of an environmental microbial community. This problem has recently been addressed by the emerging field of metagenomics. Metagenomics is a 'brute force' approach, whereby total DNA from a microbial and/or viral population is sequenced and compared with all previously sequenced genes. The high-throughput capability offered by next-generation sequencing methods makes them attractive for such an approach. Furthermore, traditional approaches to metagenomic sequencing have required an initial cloning step. This was problematic, because many cloned sequences are not stably maintained in the host (typically *E. coli*), leading to biases in the repertoire of sequences.

A cautionary note on the significance of phylogenetic surveys is always necessary. It is widely recognized by microbial molecular ecologists that a phylotypic signal (for example, a partial 16S rRNA gene sequence amplified by PCR) is no guarantee that the organism (phylotype) identified from that signal is a functional member of the microbial community. Extracted metagenomic DNA is typically derived from live, dormant, and dead cells and even from exogenous DNA from dead and lysed cells (40).

I. B. c. 6. Quantitative detection and/or activity detection of bacteria

While the sequencing-based methods are used to identify microorganisms present in a sample, other methods can provide other interesting information in a sample.

Real time quantitative PCR (qRT-PCR) provides very specific results — detection and quantification of a specific microorganism by dye-labeled primer during PCR..

DNA Chips contain different DNA sequences fixed on which DNA can be hybridized and can detect the target DNA.

When we focus on DNA in bacteria, it is hard to know the state of the bacteria, DNA can remain stable for a while after the death of a cell. There exist several techniques to detect the activity of the cell (100, 102, 185). Intracellular RNA is rapidly degraded in stressed cells and is more unstable outside the cell than DNA. As a result, RNA-based methods have been suggested to study the active microbial fraction in environmental matrices (185).

Fluorescence *in-situ* hybridization (FISH) can be used to detect the bacteria quantity in situ using proper fluorescent labeled probes (i.e. 16S rRNA). Bacterial FISH probes are often primers for the 16S rRNA region.

PLFA Analysis: Phospholipids are a primary structural component of the membranes of all living cells and break down rapidly upon cell death. Therefore, the mass of PLFAs in a sample is a direct measure of the viable biomass in the sample.

Flow cytometry is widely applied for counting the number of cells and detecting their activity in combination with fluorescence staining for membranes (100).

Total adenosine tri-phosphate (ATP) measurements can also be used to assess microbial viability in environments (20).

These techniques can often be combined. Techniques include microscopical and chemical analysis of samples collected from the field, culture-dependent methods of identification and function, and culture-independent genetic profiling and molecular-level isotopic measurements, which between them allow insights into what

organisms are present, how the community is structured and which organisms are playing what geochemical roles (161).

The selection of technique tools may depend on the research aims (Fig 7).

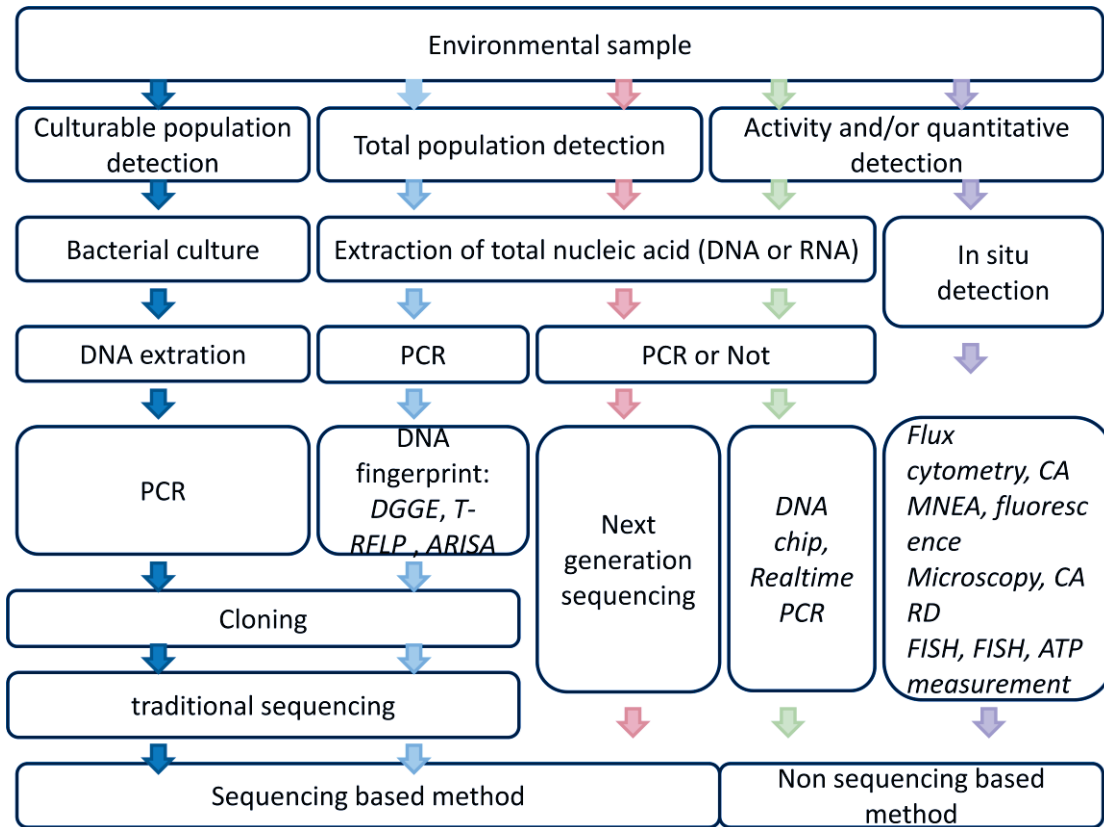


Figure 7: Process of population analysis from environmental sample. Five different color arrow represent five possible analysis way.

I. C. The bacterial diversity in oligotrophic environment

Previous environmental microbial studies revealed that only very few parts of environmental bacteria could be cultured under standard laboratory conditions, the percentage varied from 0.0001% to 15% in different environmental sample (7, 170), it is clear that the observed CFU counts can only show a fraction of the actual bacterial diversity. While recent studies confirmed low levels of cultivable microorganism in oligotrophic environment (Table 2), applications of new methods (culture independent techniques in microbial ecology) led to contrasting findings (29, 179). For example, Fierer & Jackson (56) found that the acidic soils of tropical forests harbour fewer bacterial taxa than the neutral pH soils of deserts in their global-scale study. In the McMurdo Dry Valley, a hyperarid polar desert, a broad range of previously unreported bacteria and fungal microbial soil communities were discovered (179). There are abundant microbial photoautotrophs, displaying unexpected phylogenetic diversity, in barren soils from just below the permanent ice line of the central Himalayas (159).

Table 2: *Culturability determined as a percentage of culturable bacteria in comparison with total cell counts.*

System	Culturability (%)	Reference
Alfresco air	0.5-3	(124)
Atacama desert surface	<1	(69, 123, 158)
Atacame desert subsurface	<0.1	(69, 123, 158)
Drinking water model system	4-9	(105)
Drinking water tap water	0.3-1.6	(20, 201)

I. C. a. Deserts

The living conditions of deserts are a challenge for microorganisms, as there is little available water and/or carbon, a very large range of temperatures and high exposure to UV irradiation from the Sun. The study of desert microorganisms may help in exploitation of desert and also offer the opportunity to discover novel biological activities such as new thermostable or alkaline-stable enzymes (125, 160). Meanwhile, many desert type bacteria were isolated, identified and classified using culture-dependent methods (208, 224, 226, 239, 241).

In earlier desert microbial research, culture dependent techniques were applied to estimate the number of cells per gram of sand in several deserts such as the Atacama Desert. Before these studies, deserts were considered as virtually lifeless. For the estimation of cell number in deserts, previous studies in Acatama Desert showed different results. It was reported that cultivable heterotrophic bacteria are present in the less arid regions of the Atacama Desert at levels of 10^7 colony-forming units (CFU) per gram of soil. However, in the desert's core regions, culturable heterotrophic bacteria are extremely low, corresponding to between 10^2 to 10^4 CFU/g of soil from 24° to 25° S (158). Based on the amount of adenine sublimed, Glavin *et al.* (2004) estimated a total bacterial density of 4.4×10^6 cells/g in the samples of Acatama desert soil (69). Using the DAPI staining method, bacterial counts for the surface and subsurface Atacama samples of 0.7×10^6 cells/g and 9.6×10^6 cells/g, respectively were measured. The DAPI counts for the Atacama soils are much higher than the total counts of viable, culturable heterotrophic bacteria (10 to 10^4 CFU/g) previously measured by serial dilution plating, which could indicate that soil samples from the Atacama Desert contain mostly nonculturable bacteria that are not detected by dilution plating (69) (**Table 1**). This wide range of values is due to the extent of the desert, the various levels of precipitation, as well as the chemical composition and elevations of various sample sites. Apparent differences in population size might also correspond to differences in efficiency of the recovery methods used. Compared to the

Acatama Desert, the data for other desert type environment is limited. By measuring the ATP value in surface soil samples of the Ross desert in Antarctica, Cowan *et al.* (42) obtained the values ranging from 2.6×10^6 to 4.0×10^8 cell/g, which are four orders of magnitude higher than previously reported microbial populations (90). Sand dunes from the Thar desert are also reported to have relatively smaller population (1.5×10^2 – 5×10^4 cell/g soil) (221).

By using PCR-based biological molecular techniques, microbial research can focus more precisely on the diversity of bacterial communities and their dominant microbes. In soils of the Atacama Desert, a high diversity of microorganisms known for life in hypersaline environments was found by analysis of DGGE profiles (186). Ríos *et al.* (2010) revealed that the evaporate rocks harbor communities predominantly made up of *Cyanobacteria*, along with heterotrophic bacteria (*Bacteroidetes*) and archaea in the Atacama Desert. They also confirmed this result with microscopy and traditional cultivation- dependent methods. In a very nutrient poor (oligotrophic) environment, photoautotrophic carbon input is critical to community development. The temporal and spatial variability in community composition involved non-dominant members of the community in a study of Sonoran Desert using DGGE (157). The most common phyla in Sonoran Desert are *Cyanobacteria*, *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, etc., Archaea were present as minor components by sampling 4 sites with 10 duplicates (157). Cowan (40) also noted that many of the phylotypes identified are putative nitrogen fixers and that these organisms are likely to be responsible for complementing the nitrogen deficiency of the Dry Valley soils in the Antarctic. Pointing *et al.* (179) revealed the highly specialized microbial diversity among the hypolith, chasmolith, endolith soil type of the McMurdo Dry Valley. Communities supported previously unreported polar bacteria and fungi, but Archaea were absent from all niches. In the Namib Desert, Prestel *et al.* (2008) found that the majority of the 16S rDNA sequences displayed considerable (94-98%) homology to members of the *Firmicutes*, and in particular to members of the genus *Bacillus*, though members of the *Bacteroidetes*, *Planctomycetes*, *Chloroflexi*, and *Betaproteobacteria* groups were also observed. Cannon *et al.* (2007) detected *Actinobacteria*, *Proteobacteria*, *Firmicutes* and TM7

division bacteria by analyzing DNA and PLFA from surface and subsurface soil samples of Atacama Desert, Chile. The dominant ribotypes were sequenced and shown to be phylogenetically affiliated with the phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Proteobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Cyanobacteria* and TM7 in soils of the Ross Sea region, Antarctica. Using high-throughput sequencing techniques (pyrosequencing) we explored the bacterial community in samples from the two largest deserts in Asia: The Taklamaken and Gobi deserts. We found the most dominant phyla are *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria*.

How many different kinds of microbes live in deserts? It is hard to answer this question since 16S rDNA sequencing focused mostly on the diversity, and the dominant bacteria, in desert samples. In 2011, one study (106), using high-throughput sequencing, revealed the richness of bacteria in the desert of Sekem (Egypt). They sequenced 20 000 sequences in the desert sample, and detected more than 4000 species (3% sequence similarity) and obtained the estimation of 13, 278 species by Chao 1 estimator in this desert sample. Lauber *et al.* (116) which compared 88 soil samples from different soil types by pyrosequencing, the samples from ten different desert soils were clustered together and found to be different from other soil types. Since there were only 1000-2000 reads sequenced per sample, it is hard to have the estimation of species richness. We sampled five surface soil samples in two Asian deserts with 3956 to 18382 sequences obtained by pyrosequencing. The species numbers estimated by Chao1 varied from 1172 to 2425 per sample. Comparing our results with Koberl's, it seems there is a large range of species richness in desert samples. It is clear that with the development of sequencing techniques, more information of species richness in desert samples will be revealed such as in other soil types (87, 210).

I. C. b. Sandstorms

Asian dust is considered as air pollution which affects air quality. It carries a potential risk for public health, and one important reason is that large numbers of microorganisms may be transported along with the dust (72, 73, 104). The transport of biological particles, including bacteria, fungi and virus, by Asian dust has been reported (122, 190, 199). More than 2×10^{13} to 4×10^{16} bacterial cells /km³ per month were estimated to be brought to Beijing by Asian dust (164). To identify bacteria in Asian dust, microbiological culture methods have been used to compare differences in atmospheric colony-forming units (CFU) between non dust days and dust days (33, 73, 95). Choi *et al.* (1997) showed a 4.3 fold increase in the number of CFU during an Asian dust in Taejon city (South Korea), while Jeon *et al.* (2010) indicated that the airborne culturable bacterial concentration was significantly increased during Asian dust events in Seoul (South Korea). During the days affected by ADE, culturable bacterial population levels showed significant positive correlations, with total suspended particles and particulate matter with aerodynamic diameter < 10 µm (PM10), whereas no significant correlation was found during non-dust days (97). A statistically significant correlation was observed between an increase in the prevalence of microorganisms recovered from atmospheric samples on dust days, versus that observed on non-dust days in Erdemli, Turkey (73). Culture-independent methods have been used recently to compare bacterial differences in population, such as DGGE (97, 122, 142, 164). Using terminal restriction fragment length polymorphism analysis (T-RFLP), Nishimura *et al.* (2010) revealed that the bacterial community structures in Asian dust samples differed greatly according to the scale of the dust event. The bacterial communities from major dust events were similar to those from an arid region of China (164). Jeon *et al.* (2010) revealed that the ambient air bacterial community structure was abruptly changed during Asian dust events in Seoul (South Korea) by comparing the denaturing gradient gel electrophoresis (DGGE) band patterns, and significant differences in the 16S rDNA clone library between Asian dust events and

non-dust days were observed (97).

Between 1.6×10^6 to 3.5×10^9 bacterial cells per gram were estimated in nine Asian dust samples at Beijing by quantitative realtime PCR in 2010 (164).

This bacterial composition of the dust is not known due to the techniques used (genetic fingerprint methods). *Serratia sp.* (*Gammaproteobacteria*), *Ralstonia sp.* (*Betaproteobacteria*) were identified in Beijing Asian dust (164). The Asian dust samples collected from 2007 to 2008 in Seoul contained members belonging to *Aquabacterium sp.*, *Bacillus sp.*, *Flavobacteriales bacterium sp.*, *Prevotellaceae bacterium sp.* and *Propionibacterium sp.* (122). During 2008 to 2010, the Asian dust samples collected in Seoul contained the *Firmicutes* (53%) dominated by *Bacillus* species members, *Actinobacteria* (18%), and *Alphaproteobacteria* (11.7%) (97). To our knowledge, NGS has not yet been applied to such studies.

I. C. c. Drinking water systems

I. C. c. 1. Water

Cultivation-based heterotrophic plate counts (HPC) are used worldwide as a general microbial quality parameter in drinking water treatment and distribution, even though it is well known that only 0.25% to 1.6 % of bacteria in drinking water systems can be cultured under laboratory conditions (79, 110). *E. coli* enumeration as the conventional indicator of fecal contamination in water system, largely underestimate the *E. coli* abundance in aquatic systems because they do not take into account cells that remain viable but have lost the ability to grow in or on culture media (194). The viable but not culturable state in bacteria has been reported many times (13). However, the majority of bacterial cells in natural communities are either nonculturable by current cultivation methods or may be present in a viable-but-nonculturable state (154, 207). More importantly, there exist some important waterborne pathogens (such as *Pseudomonas aeruginosa* or *Legionellae*) that may be present in water systems, but their presence shows no correlation with conventional indicator organism counts such as *E.coli*. Otherwise, sometimes water-related pathogens could find niches to be protected in water distribution systems by associating with biofilms or free-living amoebae, which makes their observation using conventional techniques more difficult (55). It is estimated that consumers' regularly consume more than 500 HPC/ml from drinking water taken from the household tap (176).

For all these reasons, NGS techniques are required in order to reveal the actual microbiological situation in drinking water system, and help to understand the potential risks for water consumers.

Due to the difficulty in sampling municipal water supply systems, most research has been done in pilot plants, rather than a actual water supply system (79, 231).

Total bacterial cell concentrations are normally not considered during drinking water treatment as an operative parameter in regular monitoring. This is mainly because easy,

rapid and inexpensive methods for quantification of total bacterial cell concentrations have, up to now, not been available (79). The flow cytometry method offers the opportunity to rapidly detect cell numbers in water systems. In 2008, Hammes *et al.* applied quantitative flow cytometry to determine the bacterial quantity using fluorescence staining of microbial cells with the nucleic acid stain SYBRs Green I in a drinking water pilot plant. They reported that flow cytometry successfully detected 1–2 log units bacterial quantity compared to the results obtained by culture-based methods (79). In the same year, Berney *et al.* also reported their study using flow cytometry and total adenosine tri-phosphate (ATP) measurements to assess microbial viability in different drinking water samples. In all their water samples, the total bacterial cell concentration varied in the range of $0.6\text{--}1.7\times 10^5$ cells/ml, including these from commercially available bottled water samples. The methods used in this study present an efficient way to assess the general microbial quality of drinking water (20). In 2004, Ultee *et al.* used microscopical methods (by staining the cells) to count the total cell numbers, for ground water samples from Germany. A range between 3.5×10^3 and 2.2×10^4 cells/ml for total counts, and between 8.1×10^2 and 3.3×10^3 cells/ml for viable count; were obtained (217). Eichler *et al.* (2006) indicated that the average total cell count for drinking water samples was about 3×10^5 cells per ml (51). Other studies focused on the quantity of specially functional bacterial groups such as Ammonia Oxidizing bacteria in drinking water using quantitative real-time PCR (219).

Early research in the 1980 by standard plate count methods found that *Actinomyces* and *Aeromonas* species were the two most common bacterial groups in chlorinated drinking water (118). Another culture-based method reported the *Alphaproteobacteria* as the most dominant group in samples of chlorinated drinking water and chloraminated distribution system simulator feed and discharge water (231). The ground water samples of a municipal water supply from Germany were reported to be dominated by *Betaproteobacteria* in the total bacterial population, while the isolated and cultured organisms mainly belonged to the *Proteobacteria* (Alpha, Beta and Gamma), *Flavobacteria* or *Actinobacteria* (217). One recent study applied high-throughput sequencing techniques to reveal the *Betaproteobacteria* as the major

bacterial group in a pilot scale water treatment plant (113). Members of the *Alphaproteobacteria*, *Betaproteobacteria*, and *Bacteroidetes* were found to be the major groups in different reservoirs from the Harz mountains of Northern Germany, but the species level composition was very different (51). Using partial 16S rDNA sequencing, *Alpha*, *Beta*, and *Gammaproteobacteria* were found to be the major groups representing bacteria in drinking water of Lebanon, and highlighted the common presence of *Sphingomonad* in drinking water (214).

Another interesting topic is to study the evolution of bacterial composition from the source water to the tap. Microbiologists followed the treatment of water from the source to the end point, revealing a significant decrease of cell abundance (51, 110).

By traditional 16S rDNA sequencing methods, one study of water pumping wells in Greece also detected the *Betaproteobacteria* as the most dominant group, and the tap water dominated by Mycobacterium-like phylotypes, related to biofilm bacterial communities (110).

One recent study applying high-throughput sequencing techniques revealed in a pilot scale plant system, that *Betaproteobacteria* predominated in the raw water, while *Alphaproteobacteria* were predominant in the membrane tank particulate matter and membrane biofilm. The bacterial community structure clearly differed for each sample at both the genus and species levels, suggesting that different environmental and growth conditions were generated during membrane filtration (113).

A major shift in the percentage of gram negative to gram-positive bacteria was observed in one study from wells to the distribution system to the tap (176).

I. C. c. 2. Biofilm

Biofilms are complex communities of microorganisms attached to surfaces or associated with interfaces. The different biofilm growth states can be separated into five stages (203) (Fig 8). Despite the focus of modern microbiology research on pure culture using planktonic (free-swimming) bacteria, it is now widely recognized that most bacteria found in natural, clinical, and industrial settings persist in association with surfaces. Furthermore, these microbial communities are often composed of multiple species that interact with each other and their environment (44, 120). Biofilms are considered as the most successful form of life on Earth and their members can tolerate high amounts of biocides (60). In general, four potential incentives behind the formation of biofilms by bacteria during infection are considered (96):

- (1) protection from harmful conditions in the environment (defense),
- (2) sequestration to a nutrient-rich area (colonization),
- (3) utilization of cooperative benefits (community),
- (4) bacteria normally grow as biofilms, and planktonic cultures are an *in vitro* artifact (biofilms as the default mode of growth).

Biofilms are ubiquitous. They can even be found inhabiting chlorinated drinking water distribution systems. In 2002, Liu *et al.* observed bacteria multiplied in the interior wall of water distribution pipes and in the crevice of pipes by scanning electron microscopy (130).

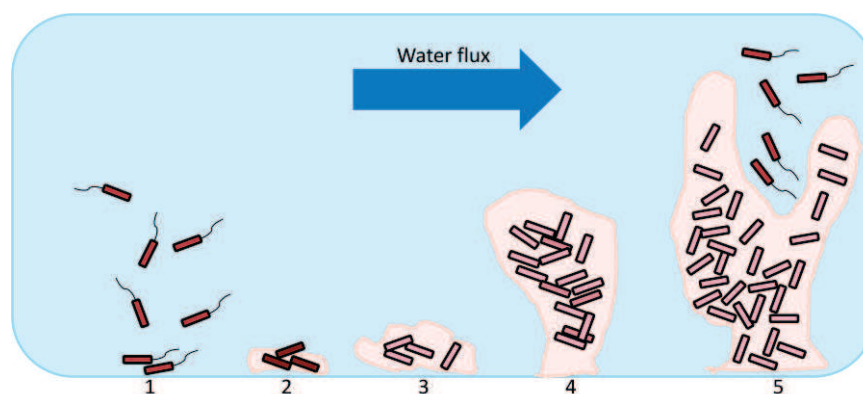


Figure 8: the schematic presentation of five stages for the formation of biofilms in water supply system, adapted from Stoodley et al, 2002 (203).

The biofilms on the surface of pipes in drinking water distribution systems are a niche for bacteria (214), as the bacteria in the biofilm can be protected from antibacterial agents (45). The bacteria of biofilms could modify the bacterial composition in water by their detachment. The study of bacterial diversity in drinking water biofilms can help to understand any potential threat to health. Compared to the sampling of water, that often needs a large volume to well understand their bacterial diversity; they are theoretically easier to collect. The water meter as a sampling site of biofilms is considered as a good choice to overcome the difficulty of sampling biofilm in water supply systems (88, 94). The interaction between planktonic bacteria and biofilm bacteria has already been suggested (105), and the microbial compositions of water supply system biofilms are now frequently studied.

Recent studies have applied high-throughput sequencing technology to better explore the bacterial diversity in such environments. In 2010, NGS techniques were used to reveal the bacterial diversity from two biofilms in an American drinking water supply system. The results suggested that the two biofilms were dominated by *Alpha/Betaproteobacteria*, or *Gammaproteobacteria* (88). The richness of bacteria in these two biofilms was at the level of 200 to 300 species by Chao1 estimator.

Hwang et al. (2011) reported the bacterial diversity of biofilms collected in different drinking water systems. The major groups contained the *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*. The richness of bacteria detected from different biofilms ranged from 200 to 1,600 species. It should be noted that different richness for the same sample was observed to vary from 600 to 1,600 using different DNA extraction methods (94).

In 2011, other research studied the microbial community composition in a pilot-scale microfiltration and chlorination plant for drinking water treatment. The bacterial diversity was analyzed for samples from raw water, membrane tank particulate matter, and membrane biofilm. Because of the large number of sequences obtained in the study (8,164–22,275), they observed large bacterial richness in each sample (1,133–1,731

OTUs). *Betaproteobacteria* predominated in the raw water, while *Alphaproteobacteria* were predominant in the membrane tank particulate matter and membrane biofilm. Moreover, signatures of potential pathogens, including *Legionella*, *Pseudomonas*, *Aeromonas*, and *Chromobacterium* were identified, and the proportions of *Legionella* and *Chromobacterium* were elevated in the membrane tank particulate matter, suggesting a potential threat to drinking water treated by membrane filtration (113).

I. D. Factors that can affect bacterial community diversity

I. D. a. In desert soil

It is important to study the influence of environmental factors on the bacterial community in soil samples. Several recent studies (8, 11, 56, 98, 116) have suggested environmental factors that may affect microbial diversity and richness in soil systems. Spatial heterogeneity in physical, chemical, and biological properties of soils allows for the proliferation of diverse microbial communities. Factors influencing the structuring of microbial communities, including availability of nutrients and water, pH, and soil texture, can vary considerably with soil depth and within soil aggregates.

pH: This factor was often reported as a main driver for bacterial communities in soil. Lauber *et al.* (116) examined soil samples from North and South America using pyrosequencing to analyze the relationship between their bacterial community structures versus environmental pH, with ten of these representing samples of desert soils such as these of the Mojave desert. These desert soils' bacterial community structures were distinguishable from the other soil samples, and could be explained by their higher pH. The range of pH for desert soils was found to be from pH 7-8.5. Similar results were observed by Koberl *et al.* (106) in their study of desert soil versus agricultural soil. However, Fierer *et al.* (2006) noted that because soils with pH levels >8.5 are rare, it is not clear whether the relationship between bacterial diversity and soil pH is truly unimodal.

Water availability: This factor is the key driver for all living things. Koberl *et al.* (2011) reported that soil bacterial communities were strongly driven by water supply (106). Bachar A *et al.* (2010) selected three sites with different levels of precipitation (400mm-Mediterranean site, 300mm-semi-arid site, and 100 mm-arid site). Their results indicated that soil bacterial abundance decreases with precipitation, but bacterial diversity was independent of precipitation gradients. Furthermore, community

composition was found to be unique to each ecosystem (11). Ben-David *et al.* (2011) studied the spatial distribution of soil microbial communities in the patchy arid and semi-arid landscapes of the Negev Desert. He further suggests that these drivers (shrub vs intershrub patches) not only act in concert but also in a way that is dependent on the aridity level (18). Other studies (21, 39) have also claimed that water availability is the primary controlling factor for microbial activity, diversity and community structure in desert soils.

Pointing *et al.* (2009) indicated in the McMurdo Dry Valleys of Antarctic, lithic community structure did not vary significantly with moisture input, as snowmelt resulted in increases in colonization frequency without significantly affecting diversity. The findings show that biodiversity near the cold-arid limit for life is more complex than previously appreciated, but that the communities lack variability probably due to the high selective pressures of this extreme environment (179).

Angel *et al.* (2010) surveyed the diversity of soil bacteria and archaea along a steep precipitation gradient ranging from the Negev Desert in the south of Israel (<100 mm annual rain) to the Mediterranean forests in the north (>900 mm annual rain). These differences of microbial communities evaluated by T-RFLP could largely be explained by the precipitation gradient combined with the vegetation cover. The archaeal and bacterial OTUs were unique to each climatic region, that is, arid, semiarid and Mediterranean, as well as patch type: plant interspaces or underneath the predominant perennial at each site (8).

Soil density and particle size: Jin *et al.* (2009) found that microbial density has an inverse relation with the bulk soil density and particle size, and a positive relationship with soil moisture content and porosity in shelter forests (98). Sessitsch *et al.* (2001) indicated that the soil particle size is one of abiotic drivers for microbial abundance (195). In addition, in surface soils, pore space seems to be an important factor. Pore spaces in microaggregates with neck diameters less than 6 μ m have more activity than pore spaces with larger diameters because the small pore necks protect resident bacteria from protozoal predation. Pore space also controls water content to some extent. Larger pores drain more quickly than smaller pores, and therefore the interior

of a small pore is generally wetter and more conducive to microbial activity. It has further been suggested that gram-negative bacteria prefer the interior of microaggregate pore space because of the increased moisture, whereas gram-positive bacteria, which are better adapted to withstand dry conditions, tend to occupy the microaggregate exteriors.(141)

Soil aggregate : Hansel *et al.* (2008) investigated changes in the microbial and functional communities within soil aggregates obtained along a soil profile spanning the surface, vadose zone, and saturated soil environments. The composition and diversity of microbial communities and specific functional groups involved in key pathways in the geochemical cycling of nitrogen, Fe, and sulfur were characterized using a coupled approach involving cultivation-independent analysis of both 16S rRNA (bacterial and archaeal) and as well as cultivation-based analysis of Fe (III)-reducing organisms. They found that the microbial communities and putative ammonia-oxidizing and Fe (III)-reducing communities varied greatly along the soil profile, likely reflecting differences in carbon availability, water content, and pH. In particular, the *Crenarchaeota* 16S rRNA sequences are largely unique to each type of soil, sharing a distribution and diversity similar to those of the putative (*amoA*-based) ammonia-oxidizing archaeal community. Anaerobic microenvironments within soil aggregates also appear to allow for both anaerobic- and aerobic-based metabolisms, further highlighting the complexity and spatial heterogeneity impacting microbial community structure and metabolic potential within soils (81).

Salinity: Lozupone *et al.* (133) 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. Aislabie *et al.* (2009) indicated that *Actinobacteria* and *Bacteroidetes* were more prevalent in dry alkaline soils and *Gammaproteobacteria* in dry saline soils. Firmicutes were extremely prevalent in saline ornithogenic Hallett soils. Elevated heavy metal levels may be an additional factor structuring bacterial diversity of the Hallett soils by comparing 5 location samples in soils of Ross Sea region, Antarctica surface (6).

Desert soil change: The changes in desert soil can influence the bacterial community

diversity. Orlando *et al.* (171) found that ammonia-oxidizers showed a significantly higher diversity of *amoA* gene clones after a “desert bloom” than during or at the beginning of this event in the Acatama desert, suggesting that the diversity of ammonia-oxidizing bacteria in this arid soil can be affected by the occurrences of “desert blooms”. Even shifts in community composition occur at the small vertical scale (stratification), particularly within the top 1 cm (65). It is also reported that soil chemical composition can affect the composition of cyanobacteria (66).

Koberl *et al.* (2011) attempted to compare the microbial community structure of desert soil with other kinds of soil types such as agricultural soils. After long-term farming, a drastic shift in the bacterial communities in desert soil was observed. Bacterial communities in agricultural soil showed a higher diversity and a better ecosystem function for plant health but a loss of extremophilic bacteria. Interestingly, they detected that indigenous desert microorganisms promoted plant health in desert agro-ecosystems (106).

Desert plants : Most of the desert microbial communities seem to be structured solely by abiotic processes (11, 98). But, if adapted desert plants occurred, (e.g. *Panicum* and *Stipagrostis* in the Sinai or *Reaumuria negevensis* in the Negev), they strongly shaped nearby soil microbial diversity (4, 85). In 1995, Herman *et al.* studied the distribution of heterotrophic bacteria in the Chihuahuan Desert soil using a culture-dependent method. They demonstrated that the resource abundance in resource islands predicts the distribution of heterotrophic bacterial numbers in desert soil, which can be called the “resource island” hypothesis (85). Later, in 1999, similar results were demonstrated by Aguilera *et al.* by studying the microorganisms in soil under shrubs in arid coastal Chile. They found that microorganisms, organic matter and nitrogen content were significantly higher underneath shrub canopies, especially in mid Fall. Microorganism abundances were positively correlated with nitrogen levels and soil moisture (4). Ben-David *et al.* (18) studied the spatial distribution of soil microbial communities in the patchy arid and semi-arid landscapes of the Negev Desert. Their results showed that in patchy desert landscapes, the microbial communities were primarily related to shrub (short plant) vs intershrub patches. Gram-positive bacteria were higher in soils under

the shrub canopies, while cyanobacteria and anaerobic bacteria were elevated in the intershrub soils. This suggests that the mere presence of a living shrub can be a dominant driving factor for the differential adaptation of desert microbial communities. However when biological soil crusts are formed in arid regions, their community structure seems less influenced by plant canopies. In 2005, Nagy *et al.* (2005) revealed that no differences in crust microbial diversity or composition were detected between crusts under plant canopies and those in plant interspaces, indicating a likely crust independence from higher plant resources (157). However, statistically significant variability with space and time could be detected, and samples within a site were more similar than samples between sites. The aridity level was also one of the main factors that shaped the microbial community structure in the patchy desert landscapes of Negev (18).

Others factors: There are many other factors that affect members of populations, including UVA, PAR (photosynthetically active radiation), altitude (41), organic carbon content (243), and oxygen concentration (137) can act as drivers of bacterial community diversity. Fierer *et al.* (2006) mentioned that the presence of a complex cryptic communities appears to strongly contradict the dogma that microbial diversity is inversely proportional to the severity of the climatic conditions (56). However, differences among the microbial communities of saline environments can also be partially attributed to physical and chemical factors and the nutrient-enriched water source (81).

I. D. b. In drinking water systems

The different water sources can have different bacterial composition and, after treatment, these differences can remain (51). In drinking water distribution systems, many factors can influence the final bacterial composition in tap water.

Coliform bacteria were detected to be associated with the following factors: filtration, temperature, disinfectant type and disinfectant level, assimilable organic carbon (AOC) level, corrosion control, and operational characteristics (119).

Assimilable organic carbon (AOC) level: The concentration of AOC is one of the key factors influencing regrowth of heterotrophic bacteria in drinking water supply systems (163). It is recommended that, in order to limit the growth of coliform bacteria in drinking water, assimilable organic carbon levels should be reduced to less than 50 micrograms/liter (117). The variation of assimilable organic carbon concentration followed different patterns in different distribution systems or different seasons due to their respective operational characteristics (130).

Disinfection: Chlorine concentration can affect the fate of bacteria. Different disinfection agent and different treatment process can influence the quantity and diversity of bacteria (51, 231). But in general, two primary disinfection regimens (chlorination and UV treatment) were not shown to have a measurable impact on the accumulation and fate of model microbial pathogens within a water distribution system (115).

Pipe material: Physicochemical characteristics of the material lining the distribution pipes are a key factor. The pipe surface of water distribution systems dramatically influenced the composition, activity, and disinfection resistance of biofilm bacteria (130, 166).

Water temperature: The presence of Coliform bacteria were detected to be associated with water temperature (119). Elevated temperatures in a hotter climate and compartmented household distribution systems resulted in very different microflora populations in comparison to those in the source water (176). It was shown that the number of HPC bacteria increases dramatically from the distribution system to the

consumers tap. Thus, the major source of bacteria ingested by the average consumer in Tucson originates from bacteria within the household distribution system or the household tap, rather than from source waters or the distribution system (176)

The temporal variation of bacterial composition in biofilm is much more important than the spatial variation in model drinking water distribution systems (145).

Biotic factors: The diversity of bacteria can affect cell abundance: Nitrifying microorganisms, for example, can contribute to the depletion of monochloramine and subsequently lead to increased overall microbial growth (51). Bacterial diversity can even affect disinfection efficacy and pathogen survival (174). The bacterial composition of biofilms can be influenced by biological factors. A significant decrease in bacterial biomass (up to 77%) was associated with the presence of amoebae (172).

Within a pilot-scale distribution system, desorption was one of the primary mechanisms affecting the loss of microspheres, *Legionellae*, and bacteriophages from biofilms, as well as disinfection and biological grazing (115)..

Manipulation biases: Up to five different DNA extraction methods were compared to characterize the samples of biofilms collected from the water meters. FastDNA Spin Kit for Soil (Qubigen) was considered as the best choice for such a study (94).

Others: There are other factors that were reported to influence the results such as residence time of the water in the distribution system, corrosion control, and operational characteristics (119).

I. E. Objective of my thesis

During my thesis, we focused on the bacterial diversity studies in oligotrophic environments by using pyrosequencing technology.

Our studies were separated into three topics:

I: Characterization of bacterial diversity in surface sand samples of two large Asian deserts- the Taklamaken Desert and Gobi Desert. We aimed to well character the bacterial composition in these samples using pyrosequencing, further discussed the factors that may impact the bacterial diversity and searched for the bacterial genera on common in all desert samples.

II: Comparison of particle-associated bacterial composition before and during sandstorms in five Asian cities. We wanted to reveal the modification of bacterial composition caused by the arrival of a sandstorm, tried to find the nature of particle-associated bacterial composition during a sandstorm.

III: Bacterial characterization of drinking water and biofilms present in the Parisian water distribution system. Water samples and biofilm samples were collected in different sites of Parisian drinking water distribution system. We used pyrosequencing to reveal their bacterial composition, observed the modification of bacterial composition during the water flux and discussed the potential threats from bacteria of the network biofilms.

CHAPTER II: Bacterial diversity in surface sand of Asian deserts

II. A. Introduction d'étude

Le désert est un environnement écologique important, il représente 30% de la surface de la terre. Les conditions de vie dans ces environnements sont un réel défi pour les micro-organismes à cause de nombreux facteurs limitants : peu d'eau et/ou de carbone disponible, une variation importante de température et une forte exposition aux irradiations UV. L'étude des micro-organismes du désert peut contribuer à l'exploitation de ce dernier et à découvrir de nouvelles activités biologiques telles que les nouvelles enzymes thermo- ou alcalino-stables (128, 160).

Notre étude a porté sur les deux grands déserts d'Asie : le désert Taklemaken (330.000 km²) et le désert de Gobi (60.000 km²). Ces deux déserts sont considérés comme le siège des tempêtes de sable d'Asie en particulier au printemps. En raison de l'augmentation de la désertification de ces régions, les tempêtes, de plus en plus fréquentes, peuvent avoir un impact à la fois sur l'environnement et sur la santé publique. Les micro-organismes à la surface des sables dans les deux déserts peuvent être transportés par les tempêtes et affecter l'environnement en aval.

Dans la littérature, de nombreuses bactéries de type « désert » ont été isolées, identifiées et classées en utilisant des méthodes de culture-dépendantes (46, 127-129, 147). En revanche, peu d'études existent sur l'analyse de la diversité microbienne dans ces deux déserts.

Le but de cette recherche est donc d'étudier la diversité bactérienne à la surface du sable du désert Taklemaken et du désert de Gobi en utilisant la technologie de séquençage à haut débit. Cinq échantillons de sable ont été collectés dans cinq sites différents : deux sites dans le désert Taklemaken, deux sites dans le désert de Gobi et un site situé entre les deux déserts. L'ADN total a été extrait et la région de V1V2 de l'ADNr 16S a été sélectionnée pour l'amplification et le séquençage (avec les amorces 8F et 357R). A notre connaissance, c'est la première fois que la technologie de séquençage à haut débit est utilisée pour étudier des échantillons de sable des déserts d'Asie. Les cinq échantillons possèdent des valeurs de pH comprises entre 8.5 et 9.8, rarement observés dans d'autres types de sols y compris le sol du désert.

Nos résultats ont révélé une grande diversité bactérienne dans le sol du désert comparable à d'autres types de sols. Une estimation de la totalité de la richesse bactérienne, utilisant l'estimateur non paramétrique Chao1, indique des valeurs entre 1172 et 2425 phylotypes totaux. Les phyla dominantes contiennent des *Firmicutes*, des *Proteobacteria*, des *Bacteroidetes* et des *Actinobacteria*. Les communautés bactériennes sont statistiquement différentes entre les cinq échantillons. Plus d'un millier d'espèces isolées dans nos échantillons sont inconnu de la base de données publique. En outre, nous avons observé une corrélation positive entre la richesse bactérienne et le rapport C/N du sol. Au total, 30 genres bactériens sont retrouvés en commun dans les cinq échantillons. Nous avons signalé quatre genres bactériens de type « désertique » détectés dans l'ensemble de nos échantillons et préalablement observés dans les études précédentes.

Article I: Bacterial diversity of surface sand samples from the Gobi and Taklamaken deserts

*Shu An, Cécile Couteau, Fan Luo, Julie Neveu and Michael S. DuBow**

Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay, France

* Correspondence: Michael DuBow, Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay ; tel: +33 1 69 15 46 12 ; Fax : + 33 1 69 15 78 08 ; e-mail: michael.dubow@igmors.u-psud.fr

Abstract

Arid regions represent nearly 30% of the Earth's terrestrial surface, but their microbial biodiversity is not yet well characterized. The surface sands of deserts, a subset of arid regions, are generally subjected to large temperature fluctuations plus high UV light exposure, and are low in organic matter. We examined surface sand samples from the Taklamaken (China: 3 samples) and Gobi (Mongolia: 2 samples) deserts, using pyrosequencing of PCR-amplified 16S V1/V2 rDNA sequences from total extracted DNA in order to gain an assessment of the bacterial population diversity. In total, 4,088 OTUs (using $\geq 97\%$ sequence similarity levels), with Chao1 estimates varying from 1,172 to 2,425 OTUs per sample, were discernable. These could be grouped into 102 families belonging to 15 phyla, with OTUs belonging to the *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* phyla being the most abundant. The bacterial population composition was statistically different among the samples, though members from 30 genera were found to be common among the five samples. An increase in phylotype numbers with increasing C/N ratio was noted, suggesting a possible role in the bacterial richness of these desert sand environments. Our results imply an unexpectedly large bacterial diversity residing in the harsh environment of these two Asian deserts, worthy of further investigation.

Keywords: Bacterial diversity, Pyrosequencing, Taklamaken and Gobi deserts.

Introduction

Arid regions represent nearly 30% of the Earth's terrestrial surface. Desert regions are defined as arid regions with less than 250 mm average annual precipitation, or where more water is lost by evaporation than precipitation. Deserts cover approximately 19% of the Earth's land surface and thus represent important ecological systems for study (1). The living conditions at the surface of deserts are a challenge for microorganisms, as there is little available water and/or carbon, a very large range of temperatures and high exposure to UV irradiation from the Sun. The study of these desert microorganisms may aid in efforts to prevent the spread of deserts, and/or to restore soil/vegetation cover and also offers the opportunity to discover novel biological activities such as new thermostable or alkaline-stable enzymes (2, 3).

Microbiological examinations of arid regions have begun to use culture independent methods to identify and characterize the endogenous bacterial community (4-6). Drees et al. (2006) studied the bacterial community from soil samples of the Atacama Desert using DGGE profiles. They demonstrated that microbial communities from the extreme hyperarid core of the Atacama Desert clustered separately from other bacterial communities by comparing their sequence data with that in Genbank. Pointing et al. (2009) analyzed, by terminal restriction fragment length polymorphism (T-RFLP), the bacterial community in the polar McMurdo Dry Valley desert in Antarctica. They suggested that the biodiversity near the cold-arid limit for life is complex, but that the communities lack a large degree of variability, likely due to the high selective pressures of this harsh environment. Neilson et al. (2012) combined the 454-pyrotag analysis with full length 16S rRNA gene library to reveal bacterial communities and to infer their functional metabolic potential inherent in unvegetated arid soils of the Atacama Desert. Their results revealed a wealth of novel bacteria and to support potentially viable communities with phylogenetic associations to non-phototrophic primary producers and bacteria capable of biogeochemical cycling in Atacama Desert.

The two largest deserts in Asia, the Taklamakan and Gobi deserts, are both located within and adjacent to China and have not been extensively examined for the

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

We found an unexpectedly large bacterial diversity residing in the surface sands of these two Asian deserts, in spite of large temperature fluctuations, high UV light exposure, low organic carbon and high alkalinity (pH 8.5–9.8). The bacterial populations were statistically different, but members from thirty genera were found to be present in all five samples.

Methods

Study sites and sampling

With a territory of 9.6 million km², China is one of the most severely desertified countries in the world (11). Deserts occupy approximately 27% (2.5 million km²) of China's total surface area (12). The Gobi Desert is located in the northern part of China and southern part of Mongolia. It is one of the ten largest deserts in the world. The Taklamakan Desert is located in the northwest of China, in the region of Xinjiang, and covers an area of 337,600 km² (13), two deserts are also the major source of sand for the seasonal dust storms in Asia that occur during the spring and summer (14). Two surface sand samples from the tops of dunes in the Taklamaken Desert were collected near Golmud (36°23' N, 94°46' E, October, 2007, 2818 m elevation, average temperature range during October: -1.2 °C to 12.9 °C), Korla (40°58' N, 85°57' E, March, 2008, 910 m elevation, average temperature range during March: 1 °C to 13.3 °C) in China. One surface sand samples from the tops of dunes from Dunhuang (40°06' N, 94°39' E, October, 2007, 1149 m elevation, average temperature range during October: 0.6 °C to 18.8 °C) located between the Taklemaken and Gobi deserts in China. Two surface sand samples from the tops of dunes in the Gobi Desert in Mongolia were collected 30 km south-east of Dzuunbayan (44°18' N, 110°06' E, December, 2007, 1001 m elevation, average temperature of December: -20.4 °C to -9.8 °C) (Fig. 1). The analyses for physico-chemical parameters were performed using standard methods by the Laboratoire d'Analyses de Sols (Arras, France).

DNA extraction

Total DNA was extracted from each sand sample using a protocol adapted from that of Zhou et al.(15). Five grams of sand were incubated at 30 °C with shaking for 1 hour after addition of 1 ml 1/4 TS (Tryptic Soy Broth) to rehydrate the bacteria and minimize nutrient shock (16). No significant change in bacterial composition was observed during this one hour incubation using ARISA-PCR (data not shown). Then, 13.5 ml extraction buffer (100 mM Tris-HCl pH 8, 100 mM Na EDTA pH 8, 100 mM Na₂HPO₄, 1.5 M NaCl, 1% [w/v] CTAB), containing 74 µg/ml pre-digested Pronase plus 6.7

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

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) PCR

An aliquot of extracted total DNA was adjusted to a final DNA concentration of 15 ng/µl in 1/10 TE buffer using a NanoVue spectrophotometer (GE Healthcare, Buckinghamshire, UK), and verified by ethidium bromide fluorescence after electrophoresis through a 1% agarose gel in TAE (2 mM Tris-acetate pH 8, 5 mM Na-EDTA) buffer. Then, multiple 50 µl PCR reactions were performed using the universal 16S rDNA bacterial primers 8F (*B*xxxxxxxAGAGTTTGATCMTGGCTCAG) and 357R (*A*xxxxxxxCTGCTGCCTYCCGTA), where *B* and *A* represent the adaptors B and A for pyrosequencing using the Gold pyrosequencing reaction (GS20, Roche/454 Life Sciences, Branford, CT, USA). The xxxxxx represents 6 nucleotide sequence tags designed for sample identification barcoding (17, 18). PCR amplification conditions were adapted for the use of three different thermostable DNA polymerases: I, Phusion High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland): 98 °C for 2 min followed by 25 cycles of 98 °C for 30 sec, 48 °C for 20 sec and 72 °C for 12 sec, and a final elongation step at 72 °C for 5 min; II, Pfu DNA Polymerase (Fermentas, ON, Canada): 95 °C for 3 min followed by 35 cycles of 95 °C for 30 sec, 48 °C for 30 sec

and 72 °C for 48 sec, and a final elongation step at 72 °C for 5 min; III, High Fidelity PCR Enzyme Mix (Fermentas): 94 °C for 3 min followed by 30 cycles of 94 °C for 30 sec, 48 °C for 30 sec and 72 °C for 24 sec, and a final elongation step at 72 °C for 5 min. Each 50 µl PCR reaction contained 15 ng DNA, 0.1 µM of each primer (Sigma-Aldrich, MO, USA), 0.2 mM dNTP mix (Fermentas), 1.25 units of thermostable DNA polymerase for I and II, or 2 units of thermostable DNA polymerase for III using the buffers supplied with each polymerase. Each DNA sample was subjected to 5-10 PCR reactions per thermostable DNA polymerase, and two different polymerases were used per sample to minimize bias. The resultant PCR products were pooled and loaded on a 1% agarose gel in TAE buffer. After electrophoreses and DNA visualization by ethidium bromide staining and long wave UV light illumination, the 16S amplified DNA fragment-containing regions were cut from the gel and purified using the NucleoSpin Extract II kit (Macherey-Nagel, North Rhine-Westphalia, Germany) according to the manufacturers' instructions. Fifty nanograms of PCR products from each sample were mixed for pyrosequencing use.

bTEFAP FLX pyrosequencing

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

bTEFAP Sequence processing pipeline and data analyses

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

The remaining sequences were then classified to the genus level using RDP-II Naive Bayesian Classifier (20, 21), with the confidence threshold set to 50%. The sequences were also classified using the RDPII all-bacteria database to find the closest species at $\geq 95\%$ similarity, with an e-value $\leq e^{-20}$ and bitscore >200 by Megablast. The

unclassified sequences were then compared with the database in VITCOMIC (22), and sequences with >80% similarity were classified to the phylum level according to the VITCOMIC results. We also normalized the number of sequences to 3,900 reads for each sample using the Pangea Pipeline (23) to compare the estimations for bacterial population richness and diversity. The Chao1 estimator and Shannon indices were calculated on the RDPII-pyro site, while the Bray-Curtis index was calculated using SPADE (24). The similarity level among the different OTUs was fixed at 97%, sequences with more than 3% difference were clustered into different OTUs by the complete-linkage clustering method on RDP's pyrosequencing pipeline (<http://pyro.cme.msu.edu/spring/cluster.spr>). We calculated the *p* value of the chi-square test for the bacterial populations from each pair of samples using R software (<http://www.r-project.org/>). Principal component analyses on the relative proportion of families or genera among the five samples using the ade-4 package adapted in R (25). All sequences have been deposited in the GenBank Sequence Read Archive (SRA).

Results

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

Physical and chemical properties of sand samples

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

Pyrosequencing data analysis

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

Unclassified sequences

Sequences with <95% similarity in the database of “all bacteria” of RDPII were considered as “new” sequences (1,991 new sequences detected). These were analyzed using unclassified sequence selection in PANGEA after Megablast against the RDPII bacterial database, yielding 1.36% to 20.56% of the total sequences per sample being classified as “new” sequences. These sequences could be classified as belonging to 740

OTUs at $\geq 97\%$ similarity level. Among these new sequences, less than 20% of the OTUs were shared between each pair of samples. Using naïve Bayesian rRNA classifier on RDPII, approximately two thirds (62%) could be classified at the phylum level. These were found to belong to the *Bacteroidetes*, *Proteobacteria*, *Actinobacteria* and *Firmicutes* phyla.

Bacterial richness and diversity in the sand samples

The OTUs, differing at $< 97\%$ sequence similarity and corresponding roughly to the level of species, ranged from 792 to 1445 per sample among the five samples (Table 2). We used the relative abundance rank curve, as a graphical representation of OTUs' composition. The curves here showed the relative abundance of the top 40 OTUs in each sample (Fig. S1). Among the five samples, there are no predominant OTUs observed. The most abundant OTU found was from sample Golmud with 7.7% coverage. The sample Korla showed the flatted curve than other samples with the most abundant OTU comprised only 4% of the sample, leading to a more even distribution of OTUs, thus can partly explain its higher diversity index and higher richness (Table 2). In total, the top 40 OTUs covered from 37.2% to 65.4% sequences in our samples. The numbers of OTUs with more than 1% sequences were ranged from 10 (Korla) to 22 (Gobi 2) OTUs in our samples. As over 1100 OTUs were detected in each sample, this figure showed that the high bacterial richness observed was a result of the large proportion of rare OTUs. The overall richness estimated by the Chao1 estimator ranged from 1,173 to 2,426 OTUs per sample using all the good quality sequences obtained after trimming, a result suggesting that our sampling revealed approximately 2/3 of the OTUs per sample, on average (Table 2 and Fig. S2). In total, 4,088 OTU phylotypes were discernible, with only 12 of them observed in all five samples, while just over half of the phylotypes (54.48%) were represented by a single sequence. The Shannon index revealed that the sample with the highest diversity level is that from Korla, while the sample with the lowest diversity level is that of Gobi2, though it is interesting to note that the Gobi2 sample contained the largest number of sequences. We also compared the richness estimated by the Chao1 estimator and the diversity measured using the Shannon index after normalizing all samples to the same numbers (3,900 sequences per

sample) of sequences using the selector.pl component of Pangea Pipeline. No significant modification to the levels of bacterial diversity was observed (Table 2).

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

Comparison of bacterial community relationships among the desert samples

The Jaccard index was calculated to reveal the unweighted dissimilarity, and the Bray-Curtis index was used to reveal the weighted dissimilarity among the five sequence samples for both the case of all good quality sequences and sequences which are normalized to 3,900 reads. For both of these indices, similar relationships among the samples were observed (Table 4). Among the five samples, the most closely-related populations, using SPADE analysis, represented the samples from Gobi1 and Gobi2, sharing approximately half of the total OTUs (49.2%). The most dissimilar samples were Golmud and Gobi2, which had only 8% of their OTUs in common (Table 4). A Jaccard's distance tree was generated on the RDP-pyro web server with the R Statistical Computing Package (UPGMA) in order to graphically reveal the relationships among these five samples (Fig. 2). The most similar bacterial populations were found in the Gobi1 and Gobi2 samples, while the bacterial population of the Golmud sample was found to be the most different from the others.

Bacterial composition of the desert samples

We classified the sequences with the RDP database, using RDP Classifier, to the genus level. Sequences representing bacteria belonging to 15 phyla, unequally distributed among the samples, were detected. The predominant phyla contained representatives from the *Firmicutes* (3.13%-82.37%), *Proteobacteria* (11.14%-50.79%), *Bacteroidetes* (1.66%-41.27%) and *Actinobacteria* (2.04%-18.50%) phyla (Fig. 3). The bacterial populations of the Gobi1 and Gobi2 samples, taken at a distance of five kilometers, were found to be significantly different ($p < 0.0001$) using the chi-square test. Bacteria representing the less predominant (<4% for each) phyla

contained members belonging to the *Gemmatimonadetes*, *Acidobacteria*, *Deinococcus-Thermus* and *Chloroflexi* phyla, while bacteria belonging to more rare phyla (*Tenericutes*, *Cyanobacteria*, *Verrucomicrobia*, *Nitrospira*, TM7, BRC1 and OD1) were not detected in all samples. Less than 5% of the sequences were unable to be classified using the RDPII database at the phylum level. Principal component analyses were performed based on the relative composition of family level members for the five samples (Fig. 4a). The samples could be grouped into two clusters: Gobi1, Gobi2 and Dunhuang, plus Korla and Golmud. Members of the *Rhodobacteraceae*, *Chitinophagaceae*, *Alteromonadaceae*, *Cytophagaceae* and *Flavobacteriaceae* families were found to play a key role underlying the variance among samples of the second group, while members of the *Bacillaceae* were relatively dominant in the first group. This relationship was maintained down to the level of genus, where members of the *Salinimicrobium*, *Pontibacter*, *Effviiibacter*, *Planococcus* and *Acinetobacter* genera were found to play a key role underlying the variance among samples of the second group, while members of the genus *Bacillus* were relatively dominant in the first group (Fig. 4b). Five-set Venn diagrams were established at the family (Fig. 5a) and genus (Fig. 5b) levels for the five samples. These diagrams showed the shared families or genera among two, three, four and all five samples. At the family level, members belonging to 30 of the families were shared among all samples. At the genus level, major differences of the five samples were revealed, as most genera were detected in only one sample.

A total of 413 genera were identified among the OTUs, with 189 genera belonging to the phylum *Proteobacteria*, 73 genera belonging to the phylum *Firmicutes*, and 68 genera belonging to the *Actinobacteria*. Among these, OTUs corresponding to 30 genera (Table 5) were detected in all five samples, and represent 56.13% of the total sequences. Twenty six of the genera contain members detected in various types of environmental samples, while members of four of the genera have been identified primarily in samples from desert-like environments (Table 5), as determined through Pubmed searches (not shown). Eight members of the 30 genera present in all five samples have been previously reported in samples from the Taklamaken or Gobi deserts.

The total proportions of these 30 genera in each sample varied from 19.4% (in sample Golmud) to 74.8% (in sample Gobi 2), while their distribution in each sample is displayed in Fig. 6, showing the varying proportions of each common genus among the 5 samples.

Discussion

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

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

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

The similar shape of rank abundance curves reflected the fact that the structures of

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

Since the sampling seasons for the five samples were different, the Gobi samples were taken under a much lower local temperature (-20 °C to -9.8 °C) comparing with other three samples (-1.2 °C to 18.8 °C). This large environmental difference might explain the less high richness and diversity observed for Gobi samples if we compared only the normalized data. The effect of temperature on the bacterial diversity in desert type environments were studied previously. In a study on Chihuahuan Desert grassland by Bell et al., they mentioned that bacteria were more able to respond to moisture pulses regardless of temperature (37). Although another desert cyanobacterial soil crusts research indicated that seasonality needs to be taken into consideration (38). The key environmental factors for total bacterial community may vary in different desert; a temporal variation of bacterial diversity research in Asian desert (same sampling site) would be necessary.

The five desert soil samples were found to be dominated by bacterial members belonging to four phyla: *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria*. Members of these phyla account for more than 90% of the sequences in each sample. Nonetheless, the percentage of the population they represent in each sample is varied. It

is not surprising that with the highly alkaline pH of these samples; fewer members of the *Acidobacteria* were observed compared to other soil samples.

Our results suggest that members belonging to four bacterial genera, *Pontibacter*, *Salinimicrobium*, *Planococcus* and *Marmoricola*, may be classified as indigenous desert types, as their members have been most often detected in desert-like environments, including the results of Lauber et al. (2009), which examined 35 soil types using 88 soil samples. Concerning the genus *Pontibacter* (phylum *Bacteroidetes*) three species have been detected in the Taklamaken desert region (39-41), while we discerned members of the genus *Pontibacter* once in a Mojave desert soil sample using the data of Lauber et al. (2009). Members of the genus *Salinimicrobium* (phylum *Bacteroidetes*) have been reported in a saline lake of the Taklamaken desert region (42), and all members are considered halophilic bacteria. For the genus *Planococcus* (phylum *Firmicutes*), members of this genus have been found in alkaline environments such as marine solar salterns, or harsh environments such as the cold deserts of the Himalayas or Antarctica. *Planococcus* members have been detected twice in Sahara desert dust events in Mali, West Africa (43) and in Erdemli, Turkey (44), and we detected members of this genus in three desert samples using the data of Lauber et al. (2009). Members of the genus *Marmoricola* (phylum *Actinobacteria*), have been isolated from volcanic ash (45), a marble statue (46), a beach (47) and a Korean soil sample (48). Analyses of the data of Lauber et al. (2009) showed that members of the genus *Marmoricola* were detected in all desert type soil samples, and in some samples from soil with minimal levels (300-800 mm per year) of precipitation.

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

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

characterized, from these unique and interesting ecosystems.

Acknowledgements

We thank Barry Holland and Denis Faure for interesting discussions. This work was supported by the Centre National de la Recherche Scientifique (CNRS), France and by the AQUAPHAGE program of the Agence Nationale de la Recherche (ANR), France.

References

1. Bhatnagar A, Bhatnagar M (2005) Microbial diversity in desert ecosystems. *Current Science* 89 (1):91-100
2. Lima DM, Oliveira RQ, Uetanabaro AP, Goes-Neto A, Rosa CA, Assis SA (2009) Thermostable inulinases secreted by yeast and yeast-like strains from the Brazilian semi-arid region. *Int J Food Sci Nutr* 60 (Suppl 7):63-71
3. Neveu J, Regeard C, DuBow MS (2011) Isolation and characterization of two serine proteases from metagenomic libraries of the Gobi and Death Valley deserts. *Appl Microbiol Biot* 91:635-644
4. Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA, Farrell RL (2009) Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci USA* 106 (47):19964-19969
5. 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. *Appl Environ Microbiol* 72 (12):7902-7908
6. Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, Tian F, LaComb M, Betancourt JL, Wing RA, Soderlund CA, Maier RM (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* 16:553-566
7. Liu M, Dai J, Liu Y, Cai F, Wang Y, Fang C (2011) *Desertibacter roseus* gen. nov., sp. nov., a gamma radiation-resistant bacterium in the family Rhodospirillaceae, isolated from desert sand. *Int J Syst Evol Microbiol* 61:1109-1113
8. Zhang L, Wu GL, Wang Y, Dai J, Fang CX (2010) *Bacillus deserti* sp. nov., a novel bacterium isolated from the desert of Xinjiang, China. *Antonie Van Leeuwenhoek* 99:221-229
9. Liu M, Liu Y, Wang Y, Luo X, Dai J, Fang C (2010) *Lysobacter xinjiangensis* sp. nov., a moderately thermotolerant and alkalitolerant bacterium isolated from gamma-irradiated sand soil sample. *Int J Syst Evol Microbiol* 61:433-437
10. An H, Xu M, Dai J, Wang Y, Cai F, Qi H, Peng F, Fang C (2010) *Sphingomonas xinjiangensis* sp. nov., isolated from the desert of Xinjiang, China. *Int J Syst Evol Microbiol* 61:1865-1869
11. Zha Y, Gao J (1997) Characteristics of desertification and its rehabilitation in China. *Journal of Arid Environments* 37 (1997):419-432
12. Ma J, Ding Z, Edmunds WM, Gates JB, Huang T (2009) Limits to recharge of groundwater from Tibetan plateau to the Gobi desert, implications for water management in the mountain front. *Journal of Hydrology* 364 (1-2):128-141
13. Li X, Yan P, Wu L (2003) Resources of Medicinal Plants Living in Sands in Xinjian Areas. *Primary Journal of Chinese Materia Medica* 17 (1):54-56
14. Xia X, Chen H, Wang P (2004) Aerosol properties in a Chinese semiarid region. *Atmospheric Environment* 38 (2004):4571-4581
15. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62 (2):316-322
16. Reasoner D, Geldreich E (1985) A new medium for the enumeration and subculture

- of bacteria from potable water. *Appl Environ Microbiol* 49 (1):1-7
17. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* 5 (3):235-237
 18. Binladen J, Gilbert MT, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2 (2):e197
 19. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72 (7):5069-5072
 20. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73 (16):5261-5267
 21. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* 35 (18):e120
 22. Mori H, Maruyama F, Kurokawa K (2010) VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences. *BMC Bioinformatics* 11:332
 23. Giongo A, Crabb DB, Davis-Richardson AG, Chauliac D, Mobberley JM, Gano KA, Mukherjee N, Casella G, Roesch LF, Walts B, Riva A, King G, Triplett EW (2010) PANGEA: pipeline for analysis of next generation amplicons. *ISME J* 4 (7):852-861
 24. Chao A, Shen T-J (2010) Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide published at <http://chao.stat.nthu.edu.tw>.
 25. Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I : One-table methods. *R-news* 4 (1):5-10
 26. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 8 (7):R143
 27. Teixeira LC, Peixoto RS, Cury JC, Sul WJ, Pellizari VH, Tiedje J, Rosado AS (2010) Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. *ISME J* 4 (8):989-1001
 28. Guss AM, Roeselers G, Newton IL, Young CR, Klepac-Ceraj V, Lory S, Cavanaugh CM (2011) Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 5 (1):20-29
 29. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* 75 (15):5111-5120
 30. Noha HY, Couger MB, Mostafa SE, Niyaz A (2010) Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere. *PLoS One* 5 (8):e12414
 31. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R (2010) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* 13 (1):135-144

32. Fulthorpe RR, Roesch LF, Riva A, Triplett EW (2008) Distantly sampled soils carry few species in common. *ISME J* 2 (9):901-910
33. Kirchman DL (2012) *Processes in microbial ecology*. Oxford University Press Inc., New York
34. Bachar A, Al-Ashhab A, Soares MI, Sklarz MY, Angel R, Ungar ED, Gillor O (2010) Soil microbial abundance and diversity along a low precipitation gradient. *Microb Ecol* 60 (2):453-461
35. Angel R, Soares MI, Ungar ED, Gillor O (2010) Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J* 4 (4):553-563
36. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103 (3):626-631
37. Bell C, McIntyre N, Cox S, Tissue D, Zak J (2008) Soil microbial responses to temporal variations of moisture and temperature in a chihuahuan desert grassland. *Microb Ecol* 56 (1):153-167
38. Bowker MA, Reed SC, Belnap J, Phillips SL (2002) Temporal variation in community composition, pigmentation, and F(v)/F(m) of desert cyanobacterial soil crusts. *Microb Ecol* 43 (1):13-25
39. Wang Y, Zhang K, Cai F, Zhang L, Tang Y, Dai J, Fang C (2010) *Pontibacter xinjiangensis* sp. nov., in the phylum 'Bacteroidetes', and reclassification of [*Effluviibacter*] *roseus* as *Pontibacter roseus* comb. nov. *Int J Syst Evol Microbiol* 60 (Pt 1):99-103
40. 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 (Pt 2):321-325
41. Zhang L, Zhang Q, Luo X, Tang Y, Dai J, Li Y, Wang Y, Chen G, Fang C (2008) *Pontibacter korlensis* sp. nov., isolated from the desert of Xinjiang, China. *Int J Syst Evol Microbiol* 58 (Pt 5):1210-1214
42. Lim JM, Jeon CO, Lee SS, Park DJ, Xu LH, Jiang CL, Kim CJ (2008) Reclassification of *Salegentibacter catena* Ying *et al.* 2007 as *Salinimicrobium catena* gen. nov., comb. nov. and description of *Salinimicrobium xinjiangense* sp. nov., a halophilic bacterium isolated from Xinjiang province in China. *Int J Syst Evol Microbiol* 58 (Pt 2):438-442
43. Kellogg CA, Griffin DW, Garrison VH, Peak KK, Royall N, Smith RR, Shinn EA (2004) Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. *Aerobiologia* 20 (2004):99-110
44. Griffina DW, Kubilayb N, Koc-akb M, Graya MA (2007) Airborne desert dust and aeromicrobiology over the Turkish Mediterranean coastline. *Atmospheric environment* 41 (2007):4050-4062
45. Lee DW, Lee SD (2010) *Marmoricola scoriae* sp. nov., isolated from volcanic ash. *Int J Syst Evol Microbiol* 60 (Pt 9):2135-2139
46. Urzi C, Salamone P, Schumann P, Stackebrandt E (2000) *Marmoricola aurantiacus* gen. nov., sp. nov., a coccoid member of the family Nocardioideae isolated from a marble statue. *Int J Syst Evol Microbiol* 50:529-536
47. Lee SD (2007) *Marmoricola aequoreus* sp. nov., a novel actinobacterium isolated

from marine sediment. Int J Syst Evol Microbiol 57 (Pt 7):1391-1395

48. Dastager SG, Lee JC, Ju YJ, Park DJ, Kim CJ (2008) *Marmoricola bigeumensis* sp. nov., a member of the family Nocardioidaceae. Int J Syst Evol Microbiol 58 (Pt 5):1060-1063

Figures and Tables

Table 1. Physical and chemical properties of the desert sand samples

	Korla	Golmud	Dunhuang	Gobi1	Gobi2
Clays (<2 μ m) [g/kg]	40	165	150	61	43
Silts (2-50 μ m) [g/kg]	3	308	472	5	4
Sands (50-2000 μ m) [g/kg]	957	527	378	934	953
Organic Carbon [g/kg]	1.12	4.22	7.07	0.515	0.296
Total nitrogen [g/kg]	0.101	2.47	0.58	0.108	0.102
C/N	11.1	1.71	12.2	4.77	2.91
Organic Material [g/kg]	1.94	7.31	12.2	0.891	0.512
pH	9.14	8.52	8.72	9.8	9.5

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

Sample	Number of sequences	OTUs	Chao1		Shannon index	
			After trimming	normalized (3900 reads)	After trimming	normalized (3900 reads)
Korla	8027	1445	2425	1858	6.12	5.99
Golmud	8435	988	1429	1075	5.58	5.39
Dunhuang	3958	874	1614	1643	5.5	5.54
Gobi1	5902	792	1172	1069	5.39	5.22
Gobi2	18403	999	1588	717	4.84	4.38

Table 3. Correlation coefficients for pH and C/N ratios.

	R	P-value (two tailed)	P-value (one tailed)
pH vs Richness	-0.486	0.407	0.203
pH vs Diversity	-0.456	0.441	0.221
C/N vs Richness	0.903*	0.035	0.018
C/N vs Diversity	0.665	0.221	0.110

* : Significant correlation

Table 4. Jaccard and Bray-Curtis dissimilarity indices of the desert samples. *

Jaccard index	Golmud	Dunhuang	Korla	Gobi1	Gobi2
Golmud	0				
Dunhuang	0.955 (0.949)	0			
Korla	0.807 (0.797)	0.740 (0.647)	0		
Gobi1	0.968 (0.963)	0.608 (0.562)	0.634 (0.622)	0	
Gobi2	0.670 (0.979)	0.521 (0.565)	0.650 (0.755)	0.203 (0.258)	0

Bray-Curtis index	Golmud	Dunhuang	Korla	Gobi1	Gobi2
Golmud	0				
Dunhuang	0.967 (0.955)	0			
Korla	0.923 (0.916)	0.912 (0.905)	0		
Gobi1	0.978 (0.978)	0.739 (0.742)	0.832 (0.831)	0	
Gobi2	0.980 (0.983)	0.856 (0.739)	0.921 (0.918)	0.681 (0.541)	0

* The numbers in parentheses are the values obtained using the normalized sequence (3900 sequences) data sets.

Table 5. The common Genera present in all the surface sand samples.

(a) The common Genera observed in Asian deserts and other environments.

Genus ID	Number of sequences	Mean % of population	Range of population
<i>Bacillus</i>	15608	28.86%	0.02%-62.10%
<i>Pseudomonas</i>	1428	1.65%	0.03%-7.40%
<i>Effluviibacter</i>	761	1.95%	0.03%-8.96%
<i>Acinetobacter</i>	671	1.69%	0.92%-2.67%
<i>Adhaeribacter</i>	499	1.33%	0.03%-2.85%
<i>Massilia</i>	491	1.00%	0.01%-1.88%
<i>Arthrobacter</i>	378	0.97%	0.22%-2.99%
<i>Flavisolibacter</i>	332	0.83%	0.08%-2.34%
<i>Lysobacter</i>	227	0.57%	0.13%-1.45%
<i>Herbaspirillum</i>	211	0.54%	0.01%-1.56%
<i>Devosia</i>	209	0.51%	0.04%-1.16%
<i>Gemmatimonas</i>	204	0.51%	0.05%-0.88%
<i>Porphyrobacter</i>	203	0.49%	0.03%-1.18%
<i>Gp6</i>	182	0.63%	0.01%-1.97%
<i>Paracoccus</i>	156	0.42%	0.03%-1.18%
<i>Planomicrobium</i>	135	0.36%	0.04%-1.01%
<i>Nocardioides</i>	124	0.35%	0.02%-1.23%
<i>Gp4</i>	113	0.33%	0.01%-0.73%
<i>Sphingomonas</i>	100	0.24%	0.09%-0.76%
<i>Rubellimicrobium</i>	86	0.21%	0.04%-0.60%
<i>Novosphingobium</i>	84	0.18%	0.03%-0.17%
<i>Truepera</i>	59	0.14%	0.01%-0.57%
<i>Comamonas</i>	44	0.10%	0.05%-0.19%
<i>Blastococcus</i>	43	0.10%	0.02%-0.31%
<i>Pedobacter</i>	42	0.10%	0.01%-0.24%
<i>Sphaerobacter</i>	27	0.09%	0.02%-0.23%

(b) The common Genera observed only in desert-like environments.

Genus ID	Number of sequences	Mean % of population	Range of population
<i>Pontibacter</i>	1101	3.06%	0.18%-7.09%
<i>Salinimicrobium</i>	944	2.27%	0.01%-8.60%
<i>Planococcus</i>	344	0.93%	0.08%-1.92%
<i>Marmoricola</i>	299	0.78%	0.03%-3.30%

Figure legends

Figure 1 Asian desert sampling sites. Asian desert sampling sites are indicated by arrowheads on the satellite photo of East Asia from a NASA World Wind screenshot http://an.wikipedia.org/wiki/Imachen:China_100.78713E_35.63718N.jpg). In the five sampling sites, Golmud and Korla are located in the Taklamaken Desert; Dunhuang is located between the Taklamaken and Gobi deserts. The dark arrowhead indicates the location of Beijing (34).

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

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

Figure 4 Principal component analyses of the desert samples at the family or genus level. The figures were generated by the ade-4 package in R software, simplified by grouping the less important, overlapped families or genera.

(a) Principal component analyses of the proportion of bacterial families in the desert sand samples. The samples in the boxes were placed by their relative abundance of families in the oval.

(b) Principal component analyses of the proportion of bacterial genera in the desert sand samples. The samples in the boxes were placed by their relative abundance of genera in the oval.

Figure 5 Venn diagrams. Venn diagrams of the bacterial groups of the desert sand samples representing the distribution of the number of families (a) or genera (b). The samples are represented by different circles. Golmud, Dunhuang, Korla, Gobi1 and Gobi2.

(a) Venn diagram representing the number of families detected in each sample and the overlap of families among the samples.

(b) Venn diagram representing the number of genera detected in each sample and the overlap of genera among the samples.

Figure 6 The relative abundances of each common genus in the sand samples.

Fig. 1



Fig. 2

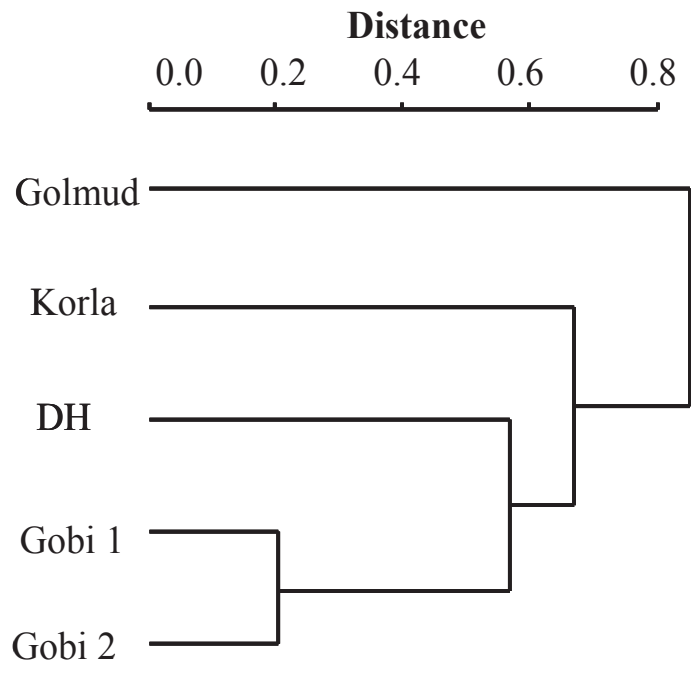


Fig. 3

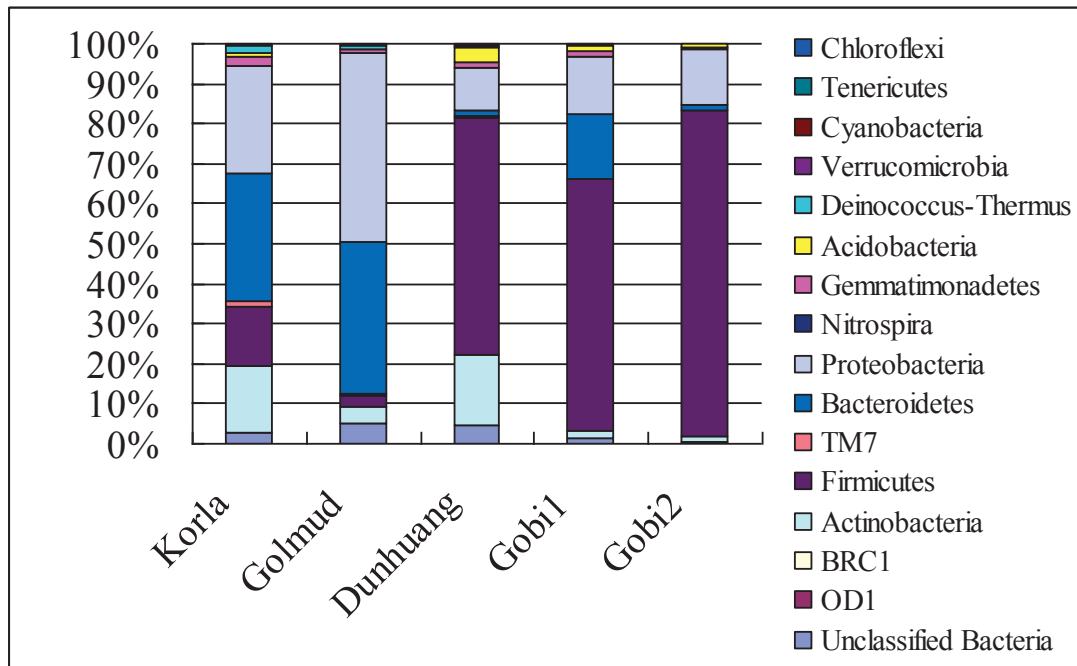


Fig. 4

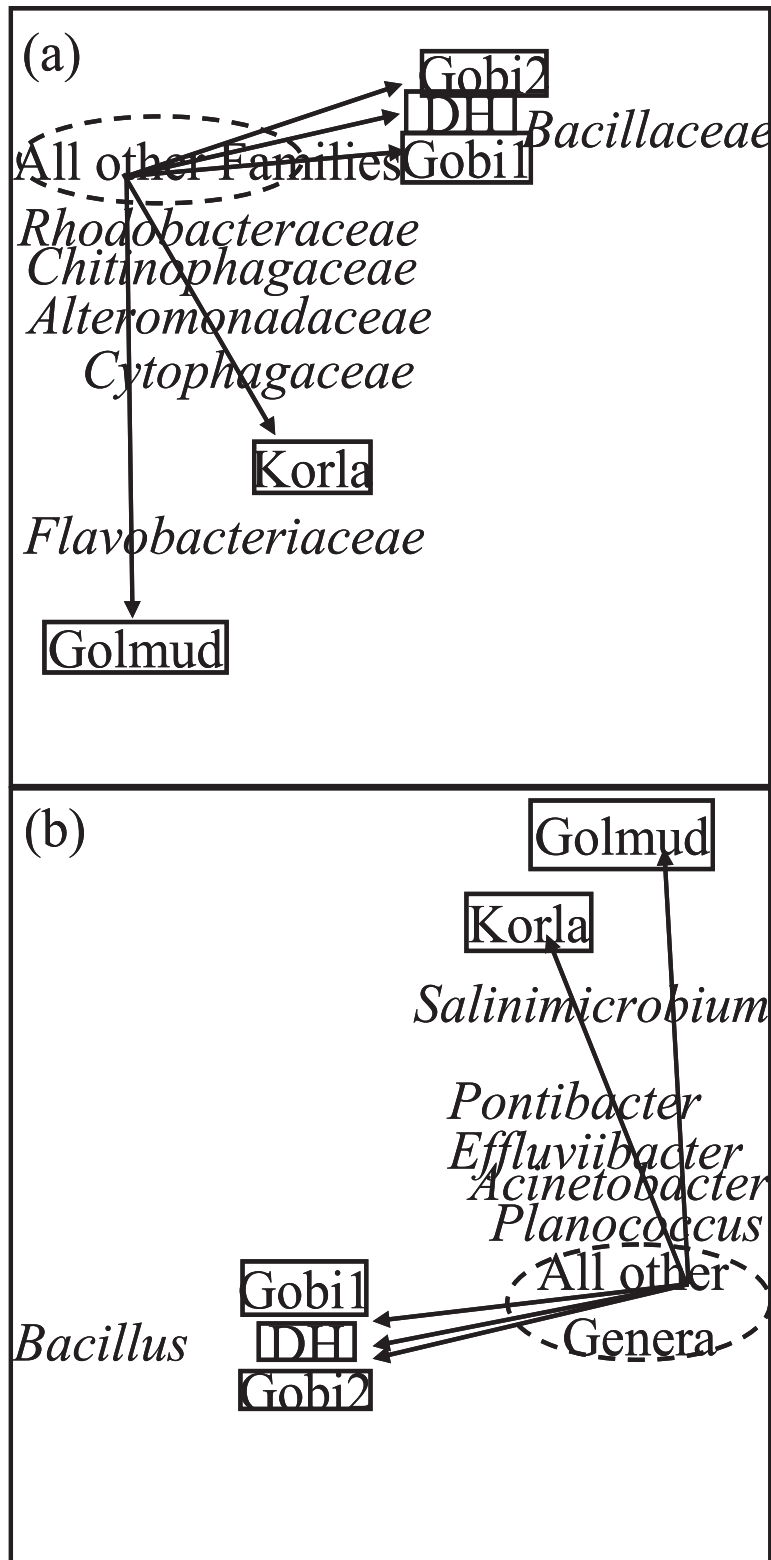


Fig. 5

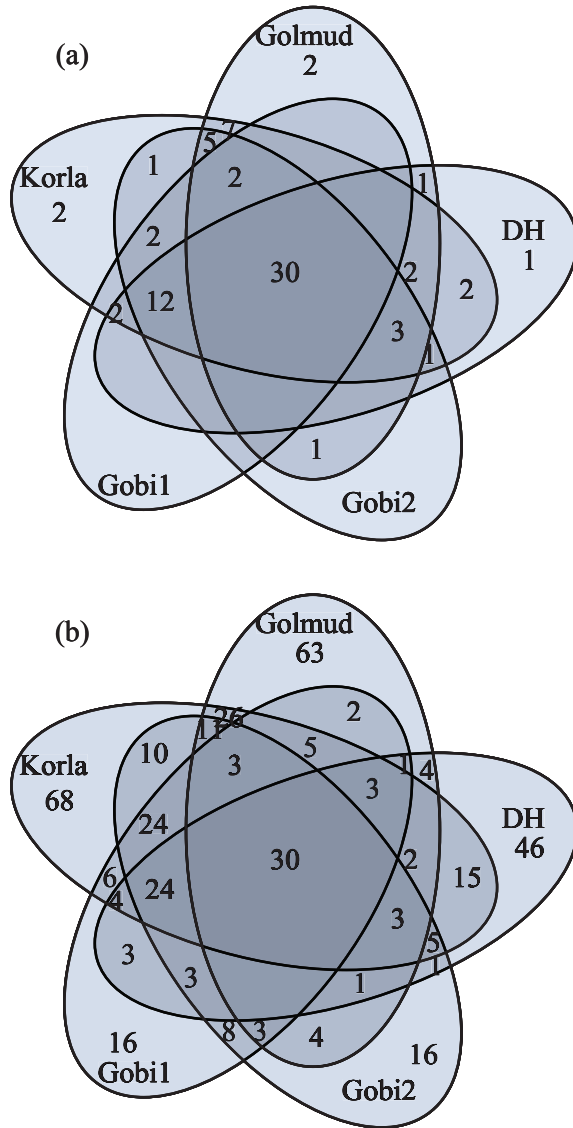
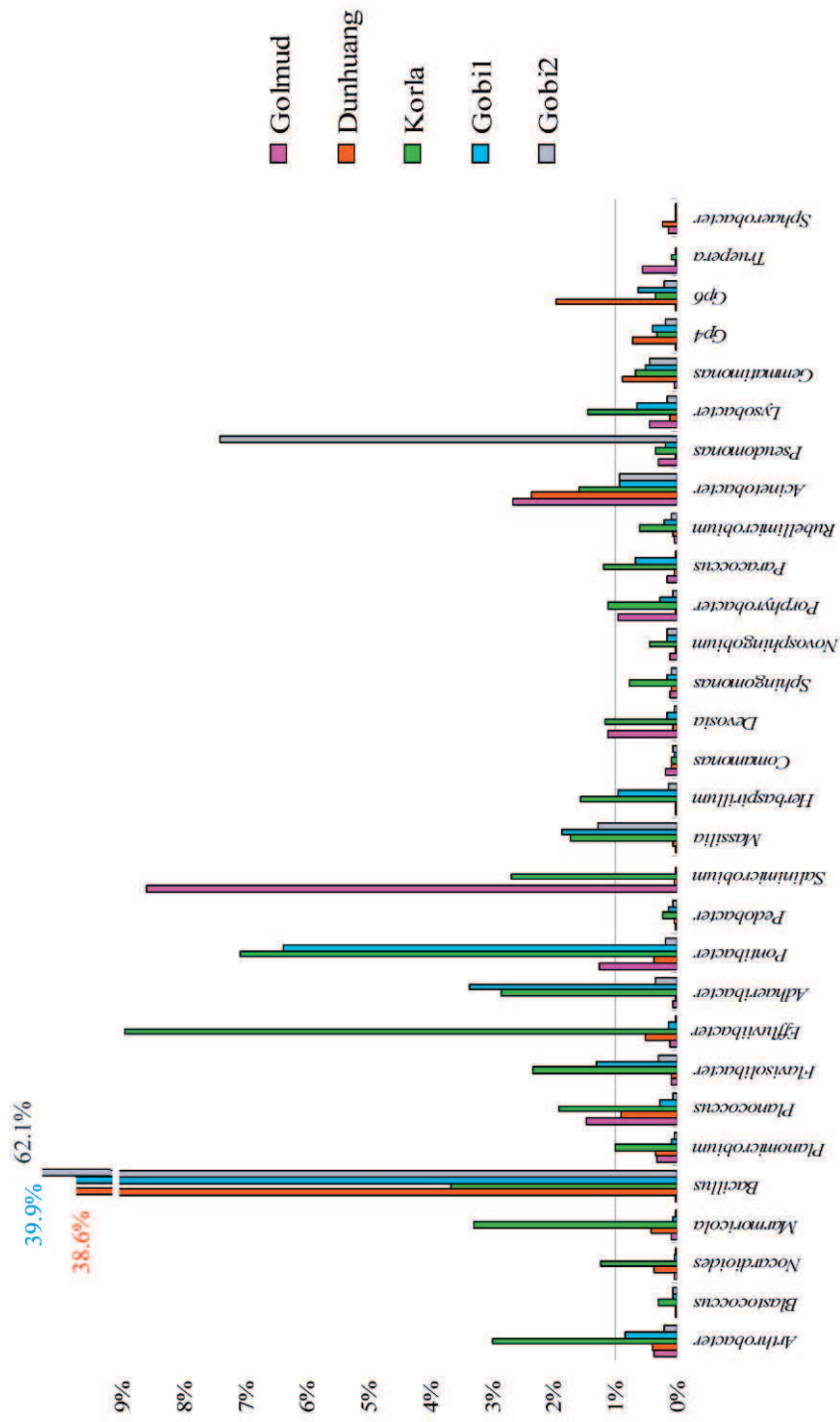
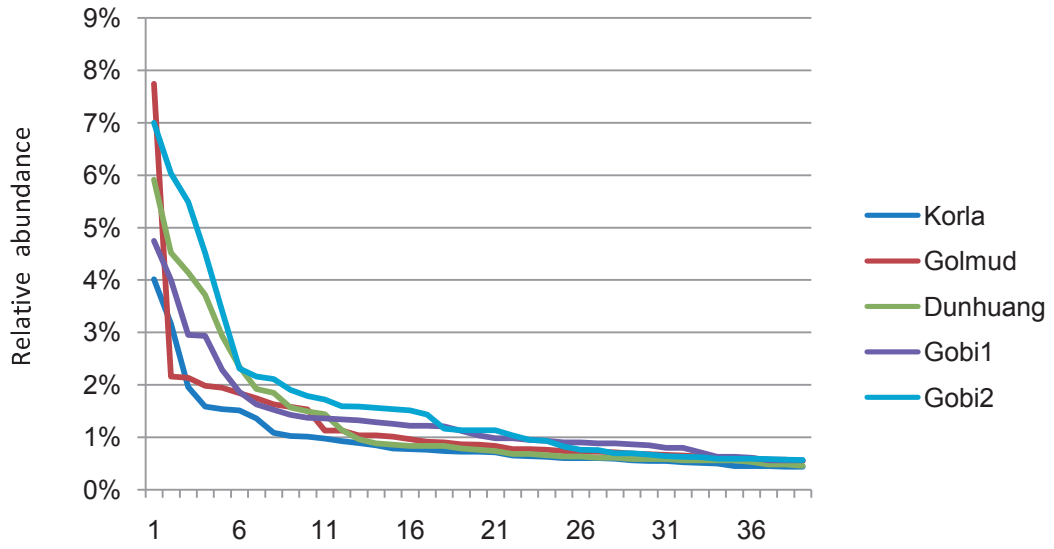


Fig. 6



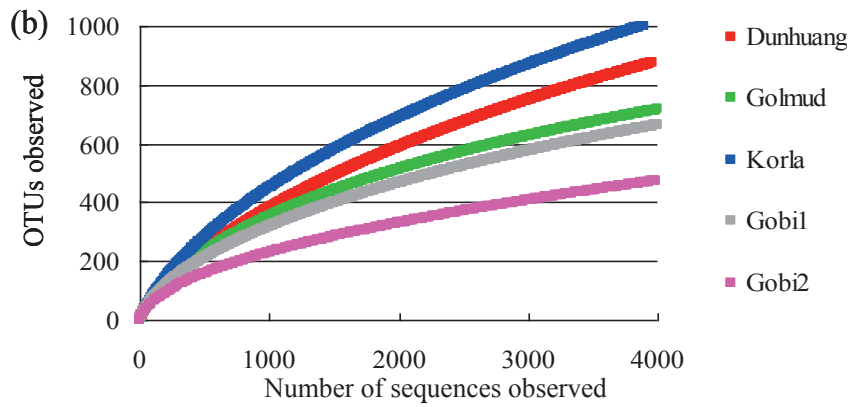
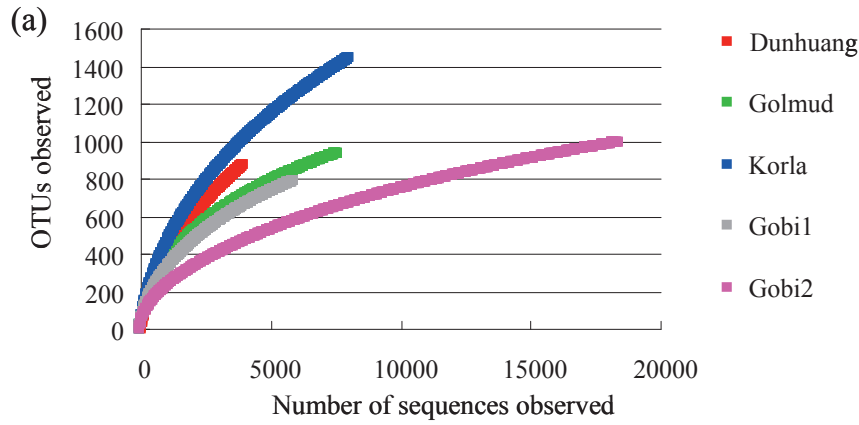
Supplementary figure 1. Rank abundance curves of OTUs based on a similarity threshold of 97% for each sample. Proportional abundances of top 40 OTUs were used for each sample.



Supplementary figure 2: Rarefaction curve for the five surface sand samples (using $\geq 97\%$ similarity) generated by the RDP-pyro web server.

(a) Rarefaction curve using all reads after trimming, including the quality control step.

(b) Rarefaction curve with normalized reads (3,900 sequences) for each sample.



**CHAPTER III: Particle-associated
bacteria in Asian sandstorms**

III. A. Introduction d'étude

Les tempêtes de sable d'Asie se produisent presque toujours au printemps, elles sont générées dans les régions arides d'Asie tels que le désert Taklamaken et le désert de Gobi. Les sables à la surface du désert peuvent être transportés vers d'autres régions du monde. L'arrivée des tempêtes de sable pourrait largement modifier l'environnement de l'air dans ces régions sous l'effet du vent, surtout dans les villes asiatiques qui sont le plus souvent touchées. Cette pollution affecte la qualité de l'air en Asie et comporte un risque potentiel pour la santé publique (72). Le transport des particules biologiques, y compris les bactéries, champignons et virus, au moment des tempêtes de sable en Asie a été étudié (122, 199). Une estimation de plus de 2×10^{13} à 4×10^{16} de cellules bactériennes par km^2 par mois, a été véhiculé à Pékin par les tempêtes d'Asie (164).

Les micro-organismes transportés par les sables peuvent augmenter et modifier la concentration et la composition de la population microbienne dans l'air de la région sous l'effet du vent. Ces modifications de structure microbienne de la composition locale pourraient avoir un effet important sur la qualité de l'air, et par conséquent sur la santé humaine, c'est pourquoi de plus en plus d'études sont établies sur les effets des tempêtes de sable en Asie. Des études antérieures ont indiqué que les microbes associés aux particules ne peuvent survivre pendant le transport à longue distance, c'est pourquoi notre travail a porté sur les bactéries associées aux particules plutôt que sur les cellules planctoniques.

Nos travaux visent à étudier la modification de la composition et la diversité des bactéries associées aux particules au moment de tempête de sable en Asie par la technologie de séquençage à haut débit. Nos échantillons ont été recueillis dans les sables de cinq villes différentes d'Asie entre 2009 et 2011 pendant les jours de tempête et par temps calme. À notre connaissance, c'est la première fois que la technologie de séquençage à haut débit a été appliquée dans une telle étude.

Nos résultats ont démontré que les compositions des bactéries associées aux particules sont modifiées pendant les tempêtes, en particulier, la proportion des *Proteobacteria* qui augmentent les jours de tempête. Nous avons signalé neuf genres bactériens

défectés en plus pendant les jours de tempêtes, cela nécessite des études plus approfondies.

Article II: Modification of atmospheric dust and sand-associated bacterial communities occurring during Asian sandstorms.

*Shu An, Fan Luo, Cécile Couteau and Michael S. DuBow**

Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay, France

* Correspondence: Michael DuBow, Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay ; tel: +33 1 69 15 46 12 ; Fax : + 33 1 69 15 78 08 ; e-mail: michael.dubow@igmors.u-psud.fr

Abstract

Desert sandstorms can affect the Earth's climate and environmental health. Asian desert sandstorms occur almost every year during the Spring, as the atmosphere in the Northern hemisphere warms. Previous studies have shown that these Asian desert sandstorms may transport microbes from the deserts over long distances. In this study, we examined the particle-associated (dust and sand-associated) bacterial populations of atmospheric sand in the absence (9 samples as controls) and presence (9 samples as sandstorm samples) of sandstorms in five Asian cities using DNA pyrosequencing of PCR amplified 16S rDNA. We obtained a total of 369,617 high quality sequences. Greater than 90% of the sequences from the sandstorm dust can be classified as representing bacteria belonging to four phyla: *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes*. The sand-associated bacterial populations in sandstorm samples were distinct from sand-associated bacteria in the absence of a sandstorm. Members of the phylum *Proteobacteria* were found to significantly increase in sandstorm samples ($P = 0.01$). Principal component analyses showed that the sandstorm-associated bacterial populations were clustered by sampling year, rather than location. Members belonging to nine bacterial genera (*Massilia*, *Planococcus*, *Carnobacterium*, *Planomicrobium*, *Pontibacter*, *Pedobacter*, *Lysobacter*, *Sanguibacter*, *Ohtaekwangia*) were observed to increase in sand-associated samples from sandstorms, versus the controls.

Importance: Asian sandstorms can spread approximately $4 - 8 \times 10^{14}$ g of sand to other parts of the world every year. One gram of sand can contain up to 3.5×10^9 bacteria. The arrival of desert bacteria on the sand can modify the local atmospheric bacterial population and may affect human and environmental health. Studies of Asian sandstorm particle-associated bacterial populations may determine potential risks that sandstorms may bring.

Key words: dust-associated bacteria, bacterial diversity, Asian sandstorm, pyrosequencing.

Introduction

Desert sandstorms are a meteorological phenomenon which have been postulated to affect the Earth's climate and environmental health. The largest sand sources on Earth for deserts sandstorm are the Sahara desert in North Africa and the Gobi plus Taklamaken deserts in Asia (24). With high-energy winds, the desert sand can be transported over long distances, which makes these sandstorms a global phenomenon. African sandstorms from the Sahara desert and Sahel region can affect the air quality of the Middle East, Europe, Asia, the Caribbean, and the Americas (14, 22, 23). Asian sandstorms from the Gobi, Taklamaken and Badain Jaran deserts can cross eastern Asia, the Pacific Ocean and arrive in North America (5, 17), and reach as far as Europe (27). It is estimated that up to 5 billion tons of desert sand is subjected to regional or global airborne migration every year (55). Desert dust is an important part of the atmospheric aerosol in some regions, and dust clouds intermittently transported over long distances can make a substantial contribution to the aerosol content of the air in distant regions. For example, approximately 50% of the particles (<10 μm) in Florida's atmosphere each summer can be African in origin (49, 50).

The potential for bacteria and other microorganisms to be transported over long distances through the air has long attracted microbiologists and been a focus of the field of aerobiology (9). The average residence time of microorganisms in the atmosphere can range from days to weeks, long enough for cells to travel between continents. During the air transport of microorganisms, they are exposed to many environmental stresses, including UV radiation, desiccation, and low pH within cloud water (9). Microorganisms, such as bacteria in aerosols, are often found to be attached to mineral dust or other larger aerosol particles (30). It has been hypothesized that bacteria attached to larger particles are more likely to retain culturability, perhaps because the particle protects them from environmental stresses (40). Wind-borne bacteria are typically transported <1 km from their source (6). However, dust-associated bacteria can be transported over 5,000 km (25, 30). The mix of microbiota and dust particles seems to protect microbiota from these stresses during long-range transport in the

atmosphere (30). The arrival of dust events increases the local concentration of PM₁₀ (particles <10 µm in size) and PM_{2.5} (particles <2.5 µm in size) (17). Of particular concern are that PM₁₀ particles can penetrate into the lungs and those of PM_{2.5} may penetrate into deep lung tissue and the subepithelial environment (24), thus leading to an increase of dust-associated bodily bacterial concentrations.

Asian dust, also sometimes called KOSA, is a seasonal meteorological phenomenon which occurs mostly in Spring. It has large effects on the environment, especially for eastern Asia, including China, Korea and Japan. Wind eroded desertification is one of the main reasons underlying sandstorm origins (15). In desertified areas, vegetation can be seriously degraded. Between 1975 and 1987, the desertification rate in China was 2,100 km²/year (65). Asian sandstorms are considered as air pollution which can affect environmental quality. They also pose a potential risk for public health, in that large numbers of microorganisms may be transported along with the dust (24, 37, 62). The transport of biological particles, including bacteria, fungi and viruses by Asian dust has been previously reported (39, 53, 54), while pathogens and allergens were detected that may potentially affect the health of downwind populations and ecosystems (32). More than 2×10^{13} to 4×10^{16} bacterial cells/km² per month were estimated to be transported to Beijing by Asian dust (46). To identify bacteria in Asian dust, classical microbiological culture methods have been used to compare differences in atmospheric colony-forming units (CFU) between normal days and days with Asian dust (13, 26, 30). Choi *et al.* (1997) showed a 4.3 fold increase in the number of CFU during an Asian dust storm in Daejeon city (South Korea), while Jeon *et al.* (2010) indicated that the culturable bacterial population levels showed significant positive correlations with total suspended particles (TSP) and particulate matter with PM₁₀ during Asian dust event days. As only very few environmental bacteria can be cultured under standard laboratory conditions (1, 47), the observed CFU counts show only a fraction of the actual bacterial diversity. Culture-independent methods such as DGGE and T-RFLP (31, 39, 43, 46) are being increasingly used to compare bacterial populations. Using terminal restriction fragment length polymorphism (T-RFLP) analysis, Nishimura *et al.* (2010) revealed that the bacterial community structures in Asian dust samples differed

according to the scale of the dust event. The bacterial communities from major dust events were similar to those from an arid region of China (46). Jeon *et al.* (2010) revealed that the ambient air bacterial community structure was abruptly changed during Asian dust events in Seoul (South Korea) by comparing denaturing gradient gel electrophoresis (DGGE)-band patterns, and significant differences in 16S rDNA clone libraries between Asian dust events and non-Asian dust days have been reported (31). As previous data showed that particle-associated microorganisms could better withstand environmental stress and thus survive long distance (6, 25, 30), a study of sand-associated microorganisms may illuminate the influence of long range bacterial transport.

In this work, we focused on sand-associated bacterial populations in Asian sandstorms. In order to obtain a more complete view of sand-associated bacterial communities in Asian sandstorm events, we used DNA pyrosequencing of 16S rDNA amplicons to explore the dust and sand-associated bacterial composition, in five Asian cities in China and South Korea, in the absence (as controls) and presence of Asian sandstorm events. Sequence level comparisons were used to compare the different sandstorm samples and control samples, and several bacterial genera were found to consistently increase during sandstorms in the five sampling cities, including one of genera which have been reported to contain potential human pathogens (*Massilia*).

Materials and methods

Study sites and sampling. Sampling was performed in five Asian cities over a span of three years. Two cities (Beijing and Taiyuan) are located in China, while the three others (Seoul, Gwangju and Incheon) are located in South Korea (Fig.1). Beijing and Taiyuan are located in the center of China, not far from the Gobi desert, a main Asian sandstorm source, and Beijing is located 80 km from the edge of the Gobi Desert. Twice-autoclaved sterile sand samples were placed on sterile plastic sheets on the roofs of buildings in the absence of Asian sandstorms and tossed into the air at random times over a 24-hour period as sand-associated bacteria controls (control samples). Sand and dust from sandstorms was collected during sandstorms using sterile plastic sheets to collect settled sand and dust at the same locations. The sampling dates and GPS site references are listed in Table 1. Samples in 2009 were taken at the end of August; while samples from 2010 and 2011 were collected during the Spring (from February to May) to compare the sand-associated bacterial composition between non-sandstorm days and sandstorm days. Samples were named by their first letter of the sampling city (B: Beijing; S: Seoul; G: Gwangju; T: Taiyuan and I: Incheon). The letters C or S after first letter refers to control (C) or sandstorm (S) and the name ends with the sampling year. The a or b after the sampling year was added when more than one sample were taken in the same year at the same location.

DNA extraction. Total DNA was extracted from each sand sample using a protocol adapted from that of Zhou *et al.* (63). From 0.5 to 2 grams of sand were incubated at 30 °C with shaking for 1 hour after addition of 1 ml 1/4 TS (Tryptic Soy Broth) to both rehydrate the bacteria and minimize nutrient shock (52). Then, 13.5 ml extraction buffer (100 mM Tris-HCl pH 8, 100 mM Na EDTA pH 8, 100 mM Na₂HPO₄, 1.5 M NaCl, 1% [w/v] CTAB), containing 74 µg/ml pre-digested Pronase plus 6.7 µg/ml RNase A, was added, followed by a 2 hour incubation at 37 °C with mild shaking. Following this, 1.5 ml of a 20% (w/v) SDS solution was added and incubation continued at 65 °C for a further 2 hours. The supernatant fluid was collected after a 10 min centrifugation at 6000 g at room temperature. The pellet was extracted one more

time with 4.5 ml extraction buffer plus 2% (w/v) SDS, mixed by vortexing for 10 sec, followed by incubation for 10 min at 65 °C and, after centrifugation, the supernatant fluids were pooled. The nucleic acids were extracted by the addition of an equal volume of chloroform/isoamyl alcohol (24:1) to the pooled supernatant fluids, and precipitated by the addition of 0.6 volumes of isopropanol for 1 hour at room temperature, followed by centrifugation at 16,000 g for 20 min at 20 °C. The DNA pellet was washed with 70% ethanol, followed by centrifugation at 16,000 g for 5 min at 20 °C. The DNA pellets were then air-dried and resuspended in 50 µl 1/10 TE buffer (1 mM Tris-HCl pH 8, 0.1 mM Na EDTA pH 8) at 4 °C overnight and stored at -20 °C until use.

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) PCR. An aliquot of extracted total DNA was adjusted to a final DNA concentration of 15 ng/µl in 1/10 TE buffer using a NanoVue spectrophotometer (GE Healthcare, Buckinghamshire, UK), and verified by ethidium bromide fluorescence after electrophoresis through a 1% agarose gel in TAE (2 mM Tris-acetate pH 8, 5 mM Na-EDTA) buffer. Then, multiple 50 µl PCR reactions were performed using the universal 16S rDNA bacterial primers: 27F (AxxxxxxAGAGTTTGATCMTGGCTCAG) and 357R (BxxxxxxCTGCTGCCTYCCGTA) where *A* and *B* represent the adaptors A and B for pyrosequencing using the Gold pyrosequencing reaction (GS-FLX, Roche/454 Life Sciences, CT, USA) for samples from 2009 and 2010. For the samples from 2011, primers 27F (AGAGTTTGATCMTGGCTCAG) and 517R (BxxxxxxxxxWTTACCGCGGCTGCTGG) were used, where *A* and *B* represent the adaptors A and B for pyrosequencing using the Titanium pyrosequencing reaction (GS-FLX Titanium, Roche/454 Life Sciences, CT, USA). The xxxxxx or xxxxxxxxxxx represent 6 or 10 nucleotide sequence tags designed for sample identification barcoding (23, 78). PCR amplification conditions were adapted for the use of five different thermostable DNA polymerases: **I**, Phusion High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland): 98 °C for 2 min followed by 25 cycles of 98 °C for 30 sec, 48 °C or 54 °C for 20 sec and 72 °C for 12 sec, and a final elongation step at 72 °C for 5 min; **II**, Pfu DNA Polymerase (Fermentas, ON, Canada): 95 °C for 3 min followed by 35 cycles of 95 °C for 30 sec, 48 °C or 54 °C for 30 sec and 72 °C for 48 sec, and a final

elongation step at 72 °C for 5 min; **III**, High Fidelity PCR Enzyme Mix (Fermentas): 94 °C for 3 min followed by 30 cycles of 94 °C for 30 sec, 48 °C or 54 °C for 30 sec and 72 °C for 24 sec, and a final elongation step at 72 °C for 5 min; **IV**, AccuPOL DNA Polymerase (Ampliqon, Herlev, Denmark): 96 °C for 2 min followed by 35-43 cycles of 96 °C for 45 sec, 49 °C for 45 sec and 72 °C for 1 min12 sec, and a final elongation step at 72 °C for 10 min; **V**, iProof High-Fidelity DNA Polymerase (Bio-Rad, Hercules, CA): 98 °C for 30 sec followed by 30 cycles of 98 °C for 10 sec, 57 °C for 30 sec and 72 °C for 18 sec, and a final elongation step at 72 °C for 5 min. Each 50 µl PCR reaction contained 15 ng DNA, 0.1 µM of each primer (Sigma-Aldrich, St. Louis, MO, USA), 0.2 mM dNTP mix (Fermentas), 1.25 units polymerase for **I** and **II**, 2 units polymerase for **III**, 2.5 units polymerase for **IV**, 1 unit polymerase for **V** using the buffers supplied with each polymerase. Each DNA sample was subjected to 5-10 PCR reactions per thermostable DNA polymerase, and two different polymerases were used per sample to minimize PCR bias. The resultant PCR products were pooled and loaded on a 1% agarose gel in TAE buffer. After electrophoreses and DNA visualization by ethidium bromide staining and long wave UV light illumination, the 16S amplified DNA fragment-containing regions were cut from the gel and purified using the NucleoSpin Extract II kit (Macherey-Nagel, North Rhine-Westphalia, Germany) according to the manufacturers' instructions. Fifty nanograms of PCR products from each sample were mixed for pyrosequencing.

bTEFAP FLX pyrosequencing. Pyrosequencing was performed using a Roche/454 FLX Pyrosequencer (Microsynth AG, Balgach, Switzerland). The sequences obtained for each sample were grouped according to the tag used, and the average sequence was found to be 227 nucleotides in length for those using the Gold reaction and 503 nucleotides in length for the samples using the Titanium reaction, after removal of the tags.

bTEFAP Sequence processing pipeline and data analyses. The sequences were selected by their length (>150 bases) and their quality score (90% of nucleotides with a quality score >25) using the Greengenes website (<http://greengenes.lbl.gov/cgi-bin/nph-index.cgi>) (16). Then, sequences with more than

two errors in the primer or more than one ambiguous base were removed using the RDP pyrosequencing pipeline (<http://pyro.cme.msu.edu/index.jsp>).

The remaining sequences were then classified to the genus level using RDP-II Naïve Bayesian Classifier (42, 59) with the confidence threshold set to 50% for Gold reaction samples and 80% for Titanium reaction samples. The DNA of chloroplasts and mitochondria may be amplified by PCR with the universal bacterial 16S rDNA primers used here (51, 56). The sequences classified as Chloroplast were then eliminated using FASTA Sequence Selection in RDP's pyro site. The non classified sequences were used for a further BLAST against the Genbank database to detect the sequences of mitochondria. In order to compare the estimations for bacterial population richness and diversity, we also normalized the sequence numbers for each sample to 3400 sequences by randomly selecting sequences from the fasta file using a perl script called *selector.pl* in Pangea (pipeline for analysis of next generation amplicons) (21). The Chao1 estimator and Shannon indices were calculated on the RDPII-pyro site, while the Bray-Curtis index was calculated using SPADE (2) to build the UPGMA clustering tree. The similarity level among the different OTUs was fixed at 97%. We calculated the p value of the chi-square test for the bacterial populations from each pair of samples using R software (<http://www.r-project.org/>). Principal component analyses on the relative proportion of phylum or OTU among the samples using the ade-4 package (11) adapted in R. All sequences data have been uploaded to the MG-RAST (the Metagenomics RAST) (<http://metagenomics.anl.gov/metagenomics.cgi?page=Home>) under number 4487160.3 - 4487161.3, 4487162.3 - 4487166.3, 4487168.3 - 4487172.3, 4487174.3 - 4487177.3, 4487179.3 - 4487181.3.

Results

Data quality. A total of 507,778 sequences were obtained and, after quality control checks and sequence elimination using Greengenes and RDP II pyro, 437,283 sequences remained. A B2C2 chimera check was used to eliminate the potential chimeric sequences, leaving 361,748 sequences. The average sequence length after the quality control steps were 229 nt for Gold reaction, and 373 nt for the Titanium reaction. Chloroplast sequences were detected in 13 of 18 samples, with their proportions varying between 0.04% and 4.2%. As the mitochondrial sequences would be classified as unclassified sequences in RDP classifier, we examined the unclassified sequences in Genbank using BLAST and found that the proportion of mitochondrial sequences was less than 0.01% per sample.

Richness and diversity in sand-associated bacterial populations. We collected samples in the absence and presence of Asian sandstorm events in five Asian cities from 2009 to 2011. We observed that the detectable OTU numbers varied from 321 to 4026 per sample. As shown in Table 2, a reduction in the number of DNA sequences using normalized samples influenced the estimated richness, using Chao1, while the Shannon diversity index remained stable. During the non-sandstorm event days, the samples from China displayed a significantly higher bacterial population richness (Chao1) than the samples from South Korea ($P = 0.03$) when comparing the normalized data.

Sequence level comparison. A UPGMA clustering tree (Fig. 2) was generated with the data of pairwise Bray-Curtis indices calculated using SPADE (sequence similarity level at 97%). The 18 samples were found to contain distinct bacterial communities. The control samples could be clustered by sampling year: 2009, 2010 and 2011. The sandstorm samples were also clustered by sampling year except the sandstorm sample of Gwangju 2011 which clustered with the control sample of Gwangju 2011. The exception of the sample from Gwangju 2011 was also detected in further analyses. The sandstorm samples of 2011 were found to be clustered by city and their country.

Bacterial composition. Fig. 3 shows the composition of our samples at the phylum

level. All the samples were dominated by members belonging to four phyla: *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes*. Over 90% of the sequences were classified as belonging to these four phyla. In general, the sandstorm samples contained more bacteria belonging to the *Proteobacteria*. This increase of members of the *Proteobacteria* in sandstorm samples is statistically significant ($P = 0.01$). Because of the increase of bacteria from *Proteobacteria* in sandstorm samples, we observed a higher average of gram negative cells in sandstorm samples ($74.5\% \pm 5.4\%$) than in control samples ($63.5\% \pm 15.5\%$), though these differences were not significant.

Statistical analyses. As seasonal variability of airborne bacteria has been reported (27), PCA was performed only for samples from the same season (Spring) and thus from 2010 and 2011. PCA analyses using the normalized data were generated using the *ade4* package of R (Fig.4). We observed that samples were able to be clustered by sampling year rather than sampling condition (control vs. sandstorm) or sampling city. A PCA at the phylum level explained 46.3% of the total variance, while at the OTU level, it explained only 28.6% of the observed variance. The large differences observed between GS2010 and the other samples appeared to be due to the higher abundance of members belonging to the phyla *Bacteroidetes*, *Cyanobacteria*, *Fibrobacteres*, *Deferribacteres*, OD1 and WS3 (see Supplementary Fig. 1). The differences between samples from 2010 versus samples from 2011 at the OTU level were more complicated. We observed that samples from 2010 contained more OTUs belonging to the genera *Delftia* and *Acinetobacter*, while samples from 2011 contained more OTUs belonging to the genera *Planococcus*, *Psychrobacter*, *Methylobacterium*, *Herbaspirillum*, *Rubellimicrobium* and *Microvirga*.

In total, 951 genera were detected in our samples, with 30 genera detected in all samples. In order to discern any modifications of bacterial populations at the genus level during sandstorms, we listed all genera that were always found to contain an increase in relative members in sandstorm samples, except the sample from Gwangju in 2011. The reason for the exception of the Gwangju sample from 2011 will be discussed later. Members belonging to nine genera were detected (Table 3). We also detected

members from five genera whose proportions were always reduced in sandstorm samples versus controls (Table 3).

Discussion

Richness and diversity of sand-associated bacterial populations. Previous studies have reported that the quantity of bacteria in the top soil of deserts can range from 0 to 10^9 bacteria per gram of sand (24). When desert soil is mobilized into the atmosphere, the concentrations of atmospheric sand-associated microorganisms increase in the environment and in the downwind regions (26, 32). In this work, we explored the richness and diversity of sand-associated bacterial populations in samples taken during sandstorms, as well as controls (sterile sand occasionally hurled into the air over a 24 hour period) to detect sand-associated bacteria in the absence of a sandstorm. As reported by Gihring *et al.* (2012), sample size can have an important effect on the Chao1 estimator (20). The richness estimated by Chao1 in each sample was dramatically reduced when we normalized the number of reads to 3400 sequences. It appears that 3400 reads in these samples is not enough for a complete richness estimation, thus suggesting more reads are necessary for an accurate estimation of total bacterial richness in our samples. Previous studies have reported that the amount of culturable bacteria is found to increase during sandstorm events in Seoul (31). We observed modifications of both richness and diversity for sand-associated bacterial communities during sandstorm event days, but there was no consistent trend for the modifications.

Similarity of bacterial composition. We sampled in Beijing and Gwangju each year from 2009 to 2011. As the PCA analyses showed that the samples did not cluster by sampling city, our results suggest a temporal variation in the structure of the sand-associated bacterial communities at the locations studied and are in agreement with other studies dealing with the annual variations in the diversity of atmospheric bacteria (18, 44). Since the samples were clustered by their sampling year, we can hypothesize that the temporal variability of local atmospheric bacterial population is larger than the spatial variability. Besides local meteorological conditions, the spatial variability of airborne bacterial richness was also reported as being influenced by different environmental factors, including land-use types (8) and distance from the sea

(48).

Dominant bacteria. Previous studies have concluded that the dominant bacteria in sandstorms were gram positive bacteria (26, 36, 39, 43), probably because they can form spores (such as members belonging to the genus *Bacillus*). Our results showed that gram negative bacteria, particularly members belonging to the phylum *Proteobacteria*, are the most predominant bacteria in particle-associated bacterial populations in Asian sandstorm samples. This difference may be caused by the different bacterial communities in sandstorm sources (deserts) and sampling methods applied. Previous research focused mainly on the culturable bacteria which are only a small part of bacterial populations (13). Culture-independent studies have used mostly cyclone air samplers with filters during airborne bacteria sampling which forcibly collected bacteria (31, 61). Nishimura *et al.* (2010) used the same sampling method as ours by collecting sands that fell on the top of a building (46). They reported two common dominant DGGE bands in Asian dust and soil samples from deserts which belonged to the *Proteobacteria* phylum, in agreement with our results.

The four dominant phyla detected in our samples were mentioned in previous research on bacteria in wind-eroded sediments (19). They indicated that members of *Proteobacteria* were the predominant group in coarse sediment, rather than fine dust (19). Members belonging to the *Bacteroidetes* and *Firmicutes* were more associated with fine dust (19), while members belonging to the *Actinobacteria* were more associated with source soil rather than wind-eroded sediments (19). These results agree with our source soil (deserts), as *Actinobacteria* were mostly detected in extreme environments such as hypersaline lakes, thermal springs, and arid soils (61).

Genus level modifications. Previous studies on the detection of bacteria in sandstorms often ignored the influence of locally present bacteria (26). We thus sampled both during non-sandstorm days and sandstorm days, and found members of nine genera (Table 3) that increased in sandstorm event days using pairwise comparisons.

Members of the genus *Massilia* appear to be very important in sandstorm samples, as they were always among the top 4 dominant groups at the genus level. In contrast,

members of the genus *Massilia* are not in the top five most abundant groups at the genus level in the control samples, except for the Gwangju 2011 control sample, which was collected four days after a sandstorm. Members belonging to the genus *Massilia* have been detected both in desert environments (10) and sandstorm samples (24, 26), and the sequences we detected display >96% similarity to the species *Massilia timonae*, isolated from patients (35) and reported to cause wound infections (58). The presence of members of the genus *Massilia* reveals a potential risk for humans during sandstorms. A more precise identification to the level of species or strain level would better demonstrate the potential risk brought by the arrival of *Massilia sp.* via sandstorms.

Among the nine genera that increased during sandstorm events, members of the genera *Pontibacter*, *Pedobacter* and *Lysobacter* were detected in the desert soil of China (41, 57, 64), while members of the genus *Planococcus* have been isolated from African desert dust events (26, 33) and in the cold desert of India (45). These results suggest that these bacteria can be transported from deserts to cities.

The exceptional behavior of Gwangju 2011 may be due to the sampling time of the control sample, which was four days after the passage of a sandstorm (see Table 1), suggesting that the atmospheric effects of particle-associated bacteria by sandstorms may last for several days.

Health threats. Dust-borne microbiological studies have detected bacterial dust-borne pathogens, and thus potential menaces to environmental health in downwind regions (32, 33). Besides the threats of pathogens, research has shown that exposure to microbial and microbial-component-laden airborne soils can cause respiratory stress (36). In addition, the lipopolysaccharide (LPS) of gram negative cells is an endotoxin which can elicit strong immune responses in animals and thus also augment the risk to human and animal health.

Previous dust-borne pathogen studies have examined the role of African dust events. African dust was reported to distribute potential coral pathogens (*Aspergillus sydowii*) to marine environments (60). Kellogg *et al.* (2004) identified bacterial populations in duststorm samples that suggest that approximately 10% were potential animal

pathogens, 5% were potential plant pathogens, and 25% were opportunistic human pathogens (33).

Species level identification of pathogens in Asian dust samples has rarely been reported. In Seoul, the increase of daily deaths due to Asian sandstorm events has been reported, especially for persons with advanced cardiovascular and respiratory disease (34, 38). In Taipei city (Taiwan), the rate of clinic visits for children increased 2.54% for preschool children and 5.03% for school children during Asian sandstorms (12).

Conclusions and perspectives. After potential long-range transport, the diversity of bacteria in sandstorms may play a significant role in ecosystem and human health (29, 62). As reported in the review of Kellogg *et al.* (32), a standard method is needed for the study of dust transported bacteria, such as sampling methods, extraction of DNA and sequencing methods for the identification of bacteria. A comparison of particle-associated microorganism and free living microorganism in the atmosphere is also important. Moreover, such studies should further extend to the domains of Archeae, virus and eukaryotic microbe.

Acknowledgements

This work was supported by the Centre National de la Recherche Scientifique (CNRS), France and by the AQUAPHAGE program of the Agence Nationale de la Recherche (ANR), France.

Reference

1. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**:143-69.
2. **Anne, C., and S. Tsung-Jen.** 2010. Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide published at <http://chao.stat.nthu.edu.tw>.
3. **Bell, M. L., J. K. Levy, and Z. Lin.** 2008. The effect of sandstorms and air pollution on cause-specific hospital admissions in Taipei, Taiwan. *Occup Environ Med* **65**:104-11.
4. **Binladen, J., M. T. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev.** 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* **2**:e197.
5. **Bishop, J. K., R. E. Davis, and J. T. Sherman.** 2002. Robotic observations of dust storm enhancement of carbon biomass in the North Pacific. *Science* **298**:817-21.
6. **Bovallius, A., R. Roffey, and E. Henningson.** 1980. Long-range transmission of bacteria. *Ann N Y Acad Sci* **353**:186-200.
7. **Bowers, R. M., I. B. McCubbin, A. G. Hallar, and N. Fierer.** 2012. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment* **50**:41-49.
8. **Bowers, R. M., S. McLetchie, R. Knight, and N. Fierer.** 2011. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *Isme J* **5**:601-12.
9. **Burrows, S. M., T. Butler, P. Jöckel, H. Tost, A. Kerckweg, U. Pöschl, and M. G. Lawrence.** 2009. Bacteria in the global atmosphere – Part 2: Modeling of emissions and transport between different ecosystems. *Atmos. Chem. Phys.*:9281-9297.
10. **Chanal, A., V. Chapon, K. Benzerara, M. Barakat, R. Christen, W. Achouak, F. Barras, and T. Heulin.** 2006. The desert of Tataouine: an extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environmental Microbiology* **8**:514-525.
11. **Chessel, D., A. B. Dufour, and J. Thioulouse.** 2004. The ade4 package - I : One-table methods. *R-news* **4**:5-10.
12. **Chien, L. C., C. H. Yang, and H. L. Yu.** 2012. Estimated Effects of Asian Dust Storms on Spatiotemporal Distributions of Clinic Visits for Respiratory Diseases in Taipei Children (Taiwan). *Environ Health Perspect*.
13. **Choi, D.-S., Y.-K. Park, S.-K. Oh, H.-J. Yoon, J. C. Kim, W.-J. Seo, and S.-H. Cha.** 1997. Distribution of Airborne Microorganisms in Yellow Sands of Korea. *The Journal of Microbiology* **35**:1-9.
14. **Chuvochina, M. S., D. Marie, S. Chevaillier, J. R. Petit, P. Normand, I. A. Alekhina, and S. A. Bulat.** 2011. Community Variability of Bacteria in Alpine Snow (Mont Blanc) Containing Saharan Dust Deposition and Their Snow Colonisation Potential. *Microbes Environ:Epub* 2011 Jun 8.

15. **Ci, L., and X. Yang.** 2010. desertification and its control in China, vol. High education press, Beijing.
16. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol* **72**:5069-72.
17. **Duce, R. A., C. K. Unni, B. J. Ray, J. M. Prospero, and J. T. Merrill.** 1980. Long-range atmospheric transport of soil dust from Asia to the tropical north pacific: temporal variability. *Science* **209**:1522-4.
18. **Fahlgren, C., A. Hagstrom, D. Nilsson, and U. L. Zweifel.** 2010. Annual variations in the diversity, viability, and origin of airborne bacteria. *Appl Environ Microbiol* **76**:3015-25.
19. **Gardner, T., V. Acosta-Martinez, F. J. Calderon, T. M. Zobeck, M. Baddock, R. S. Van Pelt, Z. Senwo, S. Dowd, and S. Cox.** 2011. Pyrosequencing reveals bacteria carried in different wind-eroded sediments. *J Environ Qual* **41**:744-53.
20. **Gihring, T. M., S. J. Green, and C. W. Schadt.** 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol* **14**:285-90.
21. **Giongo, A., D. B. Crabb, A. G. Davis-Richardson, D. Chauliac, J. M. Mobberley, K. A. Gano, N. Mukherjee, G. Casella, L. F. Roesch, B. Walts, A. Riva, G. King, and E. W. Triplett.** 2010. PANGEA: pipeline for analysis of next generation amplicons. *ISME J* **4**:852-61.
22. **Gorbushina, A. A., R. Kort, A. Schulte, D. Lazarus, B. Schnetger, H. J. Brumsack, W. J. Broughton, and J. Favet.** 2007. Life in Darwin's dust: intercontinental transport and survival of microbes in the nineteenth century. *Environ. Microbiol* **9**:2911-22.
23. **Goudie, A. S., and N. J. Middleton.** 2001. Saharan dust storms: nature and consequences. *Earth-Science Reviews* **56**:179-204.
24. **Griffin, D. W.** 2007. Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. *Clin. Microbiol Rev* **20**:459-77, table of contents.
25. **Griffin, D. W., C. A. Kellogg, V. H. Garrison, J. T. Lisle, T. C. Borden, and E. A. Shinn.** 2003. Atmospheric microbiology in the northern Caribbean during African dust events.
26. **Griffin, D. W., N. Kubilay, M. Kocak, M. A. Gray, T. C. Borden, and E. A. Shinn.** 2007. Airborne desert dust and aeromicrobiology over the Turkish Mediterranean coastline. *Atmospheric environment* **41**:4050-4062.
27. **Grousset, F. E., P. Ginoux, A. Bory, and P. E. Biscaye.** 2003. Case study of a Chinese dust plume reaching the French Alps. *Geophys. Res. Lett* **30**:1277.
28. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* **5**:235-7.
29. **Hwang, S. S., S. H. Cho, and H. J. Kwon.** 2005. [Effects of the severe asian dust events on daily mortality during the spring of 2002, in Seoul, Korea]. *J. Prev. Med.*

- Public Health **38**:197-202.
30. **Iwasaka, Y., G.-Y. Shi, M. Yamada, F. Kobayashi, M. Kakikawa, T. Maki, T. Naganuma, B. Chen, Y. Tobo, and C. S. Hong.** 2009. Mixture of Kosa (Asian dust) and bioaerosols detected in the atmosphere over the Kosa particles source regions with balloon-borne measurements: possibility of long-range transport. *Air Qual. Atmos. Health* **2**:29-38.
 31. **Jeon, E. M., H. J. Kim, K. Jeong, J. H. Kim, M. Y. Kim, Y. P. Kim, and J.-O. Ka.** 2010. Impact of Asian dust events on airborne bacterial community assessed by molecular analyses. *Atmospheric environment* **45**:4313-4321.
 32. **Kellogg, C. A., and D. W. Griffin.** 2006. Aerobiology and the global transport of desert dust. *Trends Ecol Evol* **21**:638-44.
 33. **Kellogg, C. A., D. W. Griffin, V. H. Garrison, K. K. Peak, N. Royall, R. R. Smith, and E. A. Shinn.** 2004. Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. *Aerobiologia* **20**:99-110.
 34. **Kwon, H. J., S. H. Cho, Y. Chun, F. Lagarde, and G. Pershagen.** 2002. Effects of the Asian dust events on daily mortality in Seoul, Korea. *Environ. Res* **90**:1-5.
 35. **La Scola, B., R. J. Birtles, M. N. Mallet, and D. Raoult.** 1998. *Massilia timonae* gen. nov., sp. nov., isolated from blood of an immunocompromised patient with cerebellar lesions. *J Clin Microbiol* **36**:2847-52.
 36. **Larsson, B. M., K. Larsson, P. Malmberg, and L. Palmberg.** 1999. Gram positive bacteria induce IL-6 and IL-8 production in human alveolar macrophages and epithelial cells. *Inflammation* **23**:217-30.
 37. **Lee, J. T., J. Y. Son, and Y. S. Cho.** 2007. A comparison of mortality related to urban air particles between periods with Asian dust days and without Asian dust days in Seoul, Korea, 2000-2004. *Environ. Res* **105**:409-13.
 38. **Lee, J. T., J. Y. Son, and Y. S. Cho.** 2007. A comparison of mortality related to urban air particles between periods with Asian dust days and without Asian dust days in Seoul, Korea, 2000-2004. *Environ Res* **105**:409-13.
 39. **Lee, S., B. Choi, S. M. Yi, and G. Ko.** 2009. Characterization of microbial community during Asian dust events in Korea. *Sci. Total Environ* **407**:5308-5314.
 40. **Lighthart, B.** 2000. Mini-review of the concentration variations found in the alfresco atmospheric bacterial populations. *Aerobiologia* **16**:7-16.
 41. **Liu, M., Y. Liu, Y. Wang, X. Luo, J. Dai, and C. Fang.** 2010. *Lysobacter xinjiangensis* sp. nov., a moderately thermotolerant and alkalitolerant bacterium isolated from gamma-irradiated sand soil sample. *Int J Syst Evol Microbiol* **61**:433-437.
 42. **Liu, Z., C. Lozupone, M. Hamady, F. D. Bushman, and R. Knight.** 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* **35**:e120.
 43. **Maki, T., S. Susuki, F. Kobayashi, M. Kakikawa, M. Yamada, T. Higashi, B. Chen, G. Shi, C. Hong, Y. Tobo, H. Hasegawa, K. Ueda, and Y. Iwasaka.** 2008. Phylogenetic diversity and vertical distribution of a halobacterial community in the atmosphere of an Asian dust (KOSA) source region, Dunhuang City. *Air Qual. Atmos. Health* **1**:81-89.

44. **Maron, P.-A., D. P. H. Lejona, E. Carvalho, K. Bizet, P. Lemanceau, L. Ranjard, and C. Mougel.** 2005. Assessing genetic structure and diversity of airborne bacterial communities by DNA fingerprinting and 16S rDNA clone library. *Atmospheric Environment* **39**:3687-3695.
45. **Mayilraj, S., G. S. Prasad, K. Suresh, H. S. Saini, S. Shivaji, and T. Chakrabarti.** 2005. *Planococcus stackebrandtii* sp. nov., isolated from a cold desert of the Himalayas, India. *Int J Syst Evol Microbiol* **55**:91-4.
46. **Nishimura, Y., T. Kenzaka, A. Sueyoshi, P. Li, H. Fujiyama, T. Baba, N. Yamaguchi, and M. Nasu.** 2010. Similarity of bacterial community structure between Asian dust and its sources determined by rRNA gene-targeted approaches. *Microbes Environ* **25**:22-27.
47. **Olsen, R. A., and L. R. Bakken.** 1987. Viability of Soil Bacteria: Optimization of Plate-Counting Technique and Comparison between Total Counts and Plate Counts within Different Size Groups. *Microb. Ecol* **13**:59-74.
48. **Pommier, T., N. P. R., G. J. M., C. Montserrat, A. S. G., and P.-A. Carlos.** 2010. Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat Microb Ecol* **61**:221-233.
49. **Prospero, J. M.** 1999. Long-range transport of mineral dust in the global atmosphere: impact of African dust on the environment of the southeastern United States. *Proc Natl Acad Sci U S A* **96**:3396-403.
50. **Prospero, J. M., and P. J. Lamb.** 2003. African droughts and dust transport to the Caribbean: climate change implications. *Science* **302**:1024-7.
51. **Raven, J. A., and J. F. Allen.** 2003. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol* **4**:209.
52. **Reasoner, D. J., and E. E. Geldreich.** 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl. Environ. Microbiol* **49**:1-7.
53. **Schlesinger, P., Y. Mamane, and I. Grishkan.** 2006. Transport of microorganisms to Israel during Saharan dust events *Aerobiologia* **22**:259-273.
54. **Shinn, E. A., D. W. Griffin, and D. B. Seba.** 2003. Atmospheric transport of mold spores in clouds of desert dust. *Arch. Environ. Health* **58**:498-504.
55. **Sid, P.** 2001. Dust, the thermostat. *Sci. News* **160**:200-201.
56. **Spencer, D. F., M. N. Schnare, and M. W. Gray.** 1984. Pronounced structural similarities between the small subunit ribosomal RNA genes of wheat mitochondria and *Escherichia coli*. *Proc Natl Acad Sci U S A* **81**:493-7.
57. **Tang, Y., Y. Wang, S. Ji, K. Zhang, J. Dai, L. Zhang, F. Peng, and C. Fang.** 2010. *Pedobacter xinjiangensis* sp. nov., from desert, Xinjiang. *J Microbiol Biotechnol* **20**:397-402.
58. **Van Craenenbroeck, A. H., K. Camps, P. Zachee, and K. L. Wu.** 2011. *Massilia timonae* infection presenting as generalized lymphadenopathy in a man returning to Belgium from Nigeria. *J Clin Microbiol* **49**:2763-5.
59. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol* **73**:5261-7.
60. **Weir-Rrush, J. R., V. H. Garrison, G. W. Smith, and E.A.Shinn.** 2004. The

relationship between gorgonian coral (Cnidaria: Gorgonacea) diseases and African dust storms. *Aerobiologia* **20**:119-126.

61. **Yamaguchi, N., T. Ichijo, A. Sakotani, T. Baba, and M. Nasu.** 2012. Global dispersion of bacterial cells on Asian dust. *Sci Rep* **2**:525.
62. **Yang, C. Y., S. S. Tsai, C. C. Chang, and S. C. Ho.** 2005. Effects of Asian dust storm events on daily admissions for asthma in Taipei, Taiwan. *Inhal. Toxicol* **17**:817-21.
63. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol* **62**:316-22.
64. **Zhou, Y., X. Wang, H. Liu, K. Y. Zhang, Y. Q. Zhang, R. Lai, and W. J. Li.** 2007. *Pontibacter akesuensis* sp. nov., isolated from a desert soil in China. *Int J Syst Evol Microbiol* **57**:321-5.
65. **Zhu, Z., and T. Wang.** 1993. The trends of desertification and its rehabilitation in China. *Desertification Control Bulletin* **22**:27-30.

Figures and Tables

Table 1: Coordinates of sampling sites.

Sample	date	Sandstorm/control	GPS coordinates
BC2009	28/08/2009	Control	39°58'45"N 116°18'57"E
SC2009	18/08/2009	Control	37°33'59"N 126°18'57"E
GC2009	21/08/2009	Control	35°09'40" N 126°56'05"E
BC2010	27/02/2010	Control	39°58'45"N 116°18'57"E
BS2010	20/05/2010	Sandstorm	39°58'45"N 116°18'57"E
GC2010	08/03/2010	Control	35°09'40" N 126°56'05"E
GS2010	19/03/2010	Sandstorm	35°09'40" N 126°56'05"E
BC2011	07/02/2011	Control	39°58'45"N 116°18'57"E
BS2011a	18/03/2011	Sandstorm	39°58'45"N 116°18'57"E
BS2011b	01/05/2011	Sandstorm	39°58'45"N 116°18'57"E
TC2011	07/02/2011	Control	37°50'02"N 112°32'34"E
TS2011a	18/03/2011	Sandstorm	37°50'02"N 112°32'34"E
TS2011b	01/05/2011	Sandstorm	37°50'02"N 112°32'34"E
GC2011	17/05/2011	Control	35°09'40" N 126°56'05"E
GS2011	13/05/2011	Sandstorm	35°09'40" N 126°56'05"E
IC2011	17/05/2011	Control	37°27'22"N 126°42'19"E
IS2011a	01/05/2011	Sandstorm	37°27'22"N 126°42'19"E
IS2011b	03/05/2011	Sandstorm	37°27'22"N 126°42'19"E

Table 2: Number of processed pyrosequencing reads, OTU richness and diversity for each sample.

Sample	Original reads					Normalized reads*		
	nb of seq	OTUs	Chao1	% coverage	Shannon index	OTUs	Chao1	Shannon index
BC2009	10895	2353	3517	66.9%	6.82	1179	2229	6.41
GC2009	4545	791	1203	65.8%	5.37	683	1108	5.34
SC2009	8383	1315	2086	63.1%	5.75	776	1381	5.59
BC2010	3386	603	1170	51.5%	4.35	603	1170	4.35
GC2010	3505	321	474	67.7%	2.90	315	473	2.90
BC2011	31959	3419	4247	80.5%	6.77	975	1496	6.05
TC2011	21603	3244	4452	72.9%	6.84	1074	1784	6.18
GC2011	27054	1915	2671	71.7%	5.48	550	942	5.01
IC2011	23466	506	554	91.4%	4.93	340	381	4.74
BS2010	6462	649	1133	57.3%	4.70	434	711	4.45
GS2010	3837	1227	2036	60.3%	6.39	1135	1938	6.35
BS2011a	22323	3025	4297	70.4%	6.37	872	1460	5.61
BS2011b	17960	2676	3644	73.4%	6.49	966	1733	5.91
TS2011a	49755	4026	5556	72.5%	6.19	748	1221	5.28
TS2011b	20139	2092	2841	73.6%	5.59	683	1045	5.00
IS2011a	22047	2666	3507	76.0%	6.55	897	1378	5.99
IS2011b	47026	2637	2990	88.2%	6.07	790	1332	5.50
GS2011	22893	1012	1365	74.1%	4.31	357	510	4.00

* : sequence numbers of each sample were normalized to 3400 reads using *Selector.pl* in Pangea.

Table 3: Genera observed to increase or decrease per site in each year, except for the Gwangju 2011 samples.

Increase in dust	Beijing 2010	Gwangju 2010	Beijing 2011	Taiyuan 2011	Incheon 2011	Gwangju 2011
<i>Massilia</i>	6276%	771%	1434%	89%	2843%	124%
<i>Planococcus</i>	71%	539%	99%	298%	+	-78%
<i>Carnobacterium</i>	251%	37%	105%	546%	+	-70%
<i>Planomicrobium</i>	105%	+	57%	343%	+	-16%
<i>Pontibacter</i>	2258%	+	141%	53%	403%	-45%
<i>Pedobacter</i>	1891%	296%	1514%	205%	1391%	-75%
<i>Lysobacter</i>	+	+	97%	85%	88%	-70%
<i>Sanguibacter</i>	179%	+	55%	634%	+	-68%
<i>Ohtaekwangia</i>	+	+	268%	579%	+	-100%
Decrease in dust	Beijing 2010	Gwangju 2010	Beijing 2011	Taiyuan 2011	Incheon 2011	Gwangju 2011
<i>Escherichia</i>						
<i>Shigella</i>	-69%	-95%	-56%	-89%	-97%	-70%
<i>Acinetobacter</i>	-82%	-32%	-40%	-37%	-72%	11%
<i>Propionibacterium</i>	-100%	-100%	-45%	-69%	-91%	71%
<i>Corynebacterium</i>	-95%	-85%	-24%	-40%	-76%	94%
<i>Streptococcus</i>	-96%	-96%	-68%	-67%	-91%	536%

+ indicates the genus was absent in non dust samples.

Figure Legends.

Figure 1: Sampling sites and the locations of the two Asian deserts. The edges of the deserts are indicated by dashed lines.

Figure 2: UPGMA clustering tree using the Bray-Curtis similarity index (97% cutoff) among the samples. Each sample was normalized to 3400 reads.

Figure 3: The relative composition of phyla in our samples. The samples were grouped by control samples and sandstorm samples and by sampling years.

Figure 4: Normalized principal component analysis (PCA) using phyla composition data (a) or OTU composition data (b) generated by R. Bold letters represent samples from 2011 and non-bold letters represent samples from 2010. PC1 and PC2 represent the two most important principal components.

Fig. 1

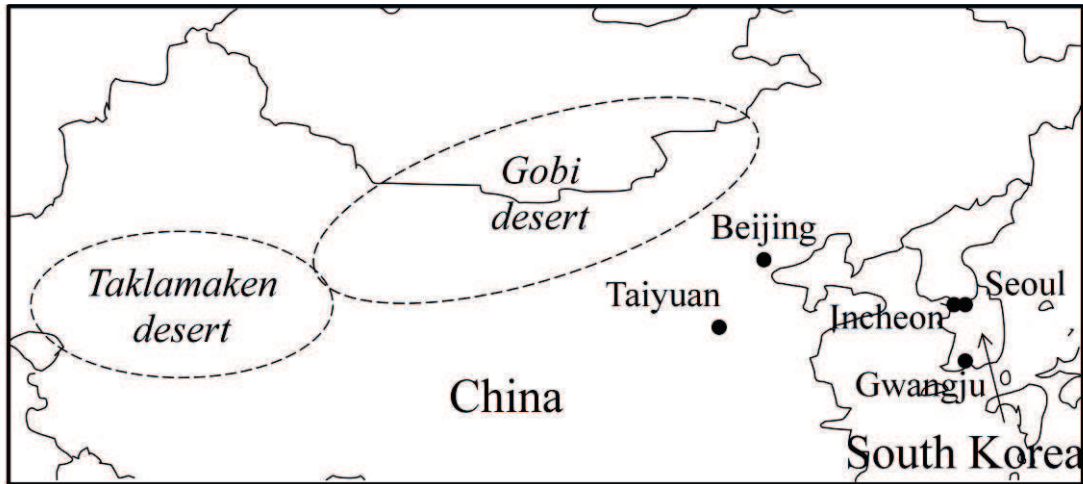


Fig. 2

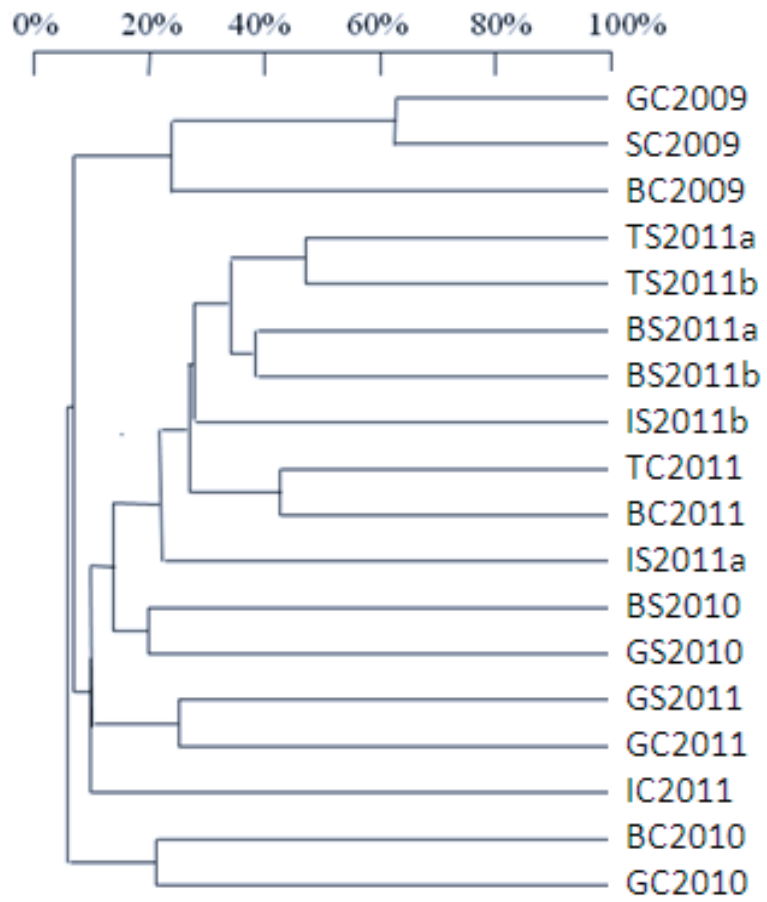


Fig. 3

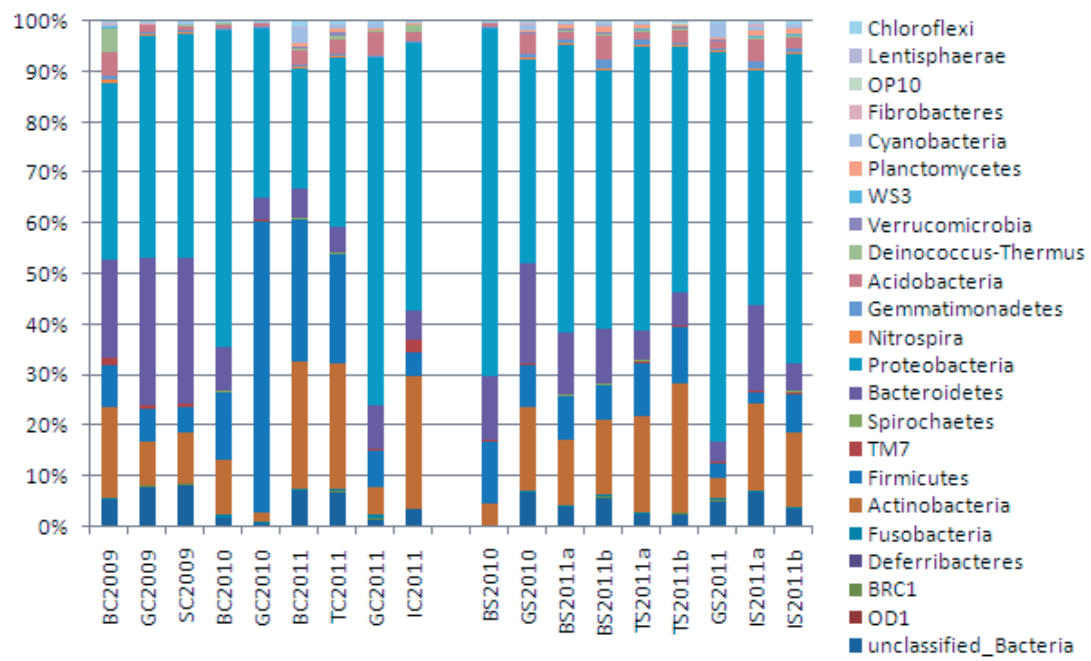
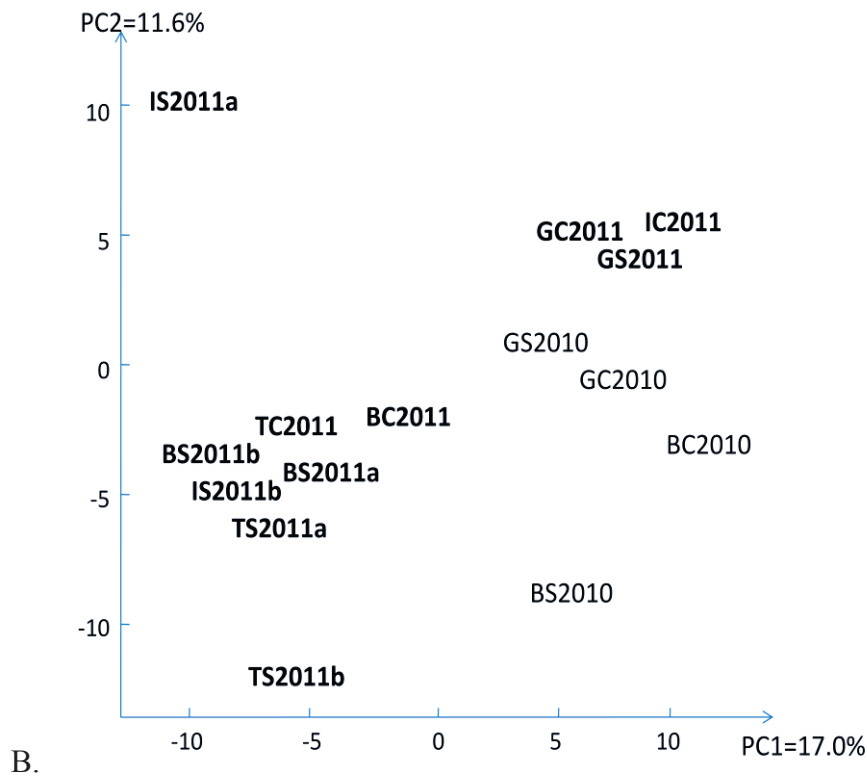
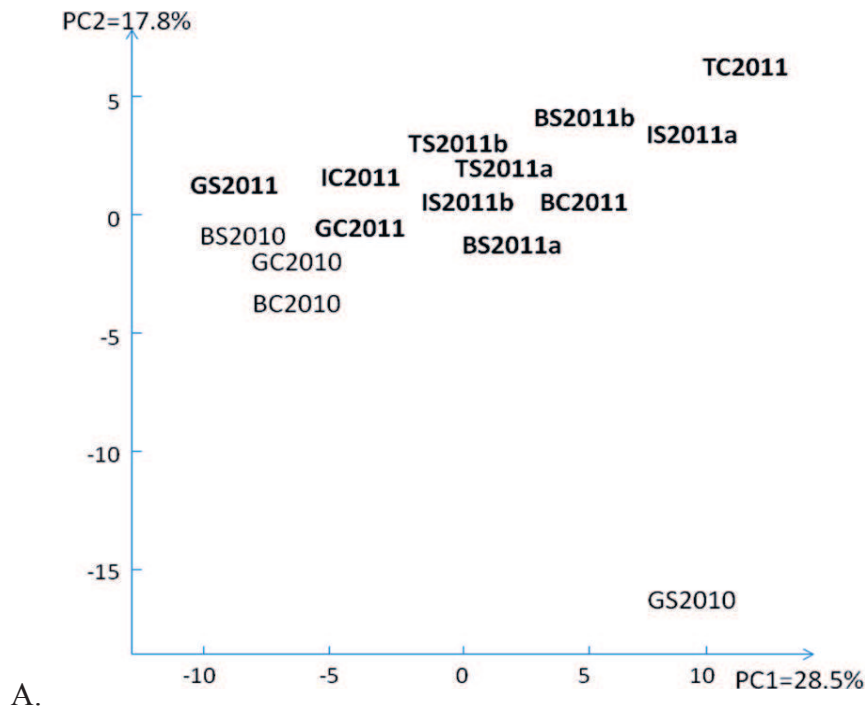
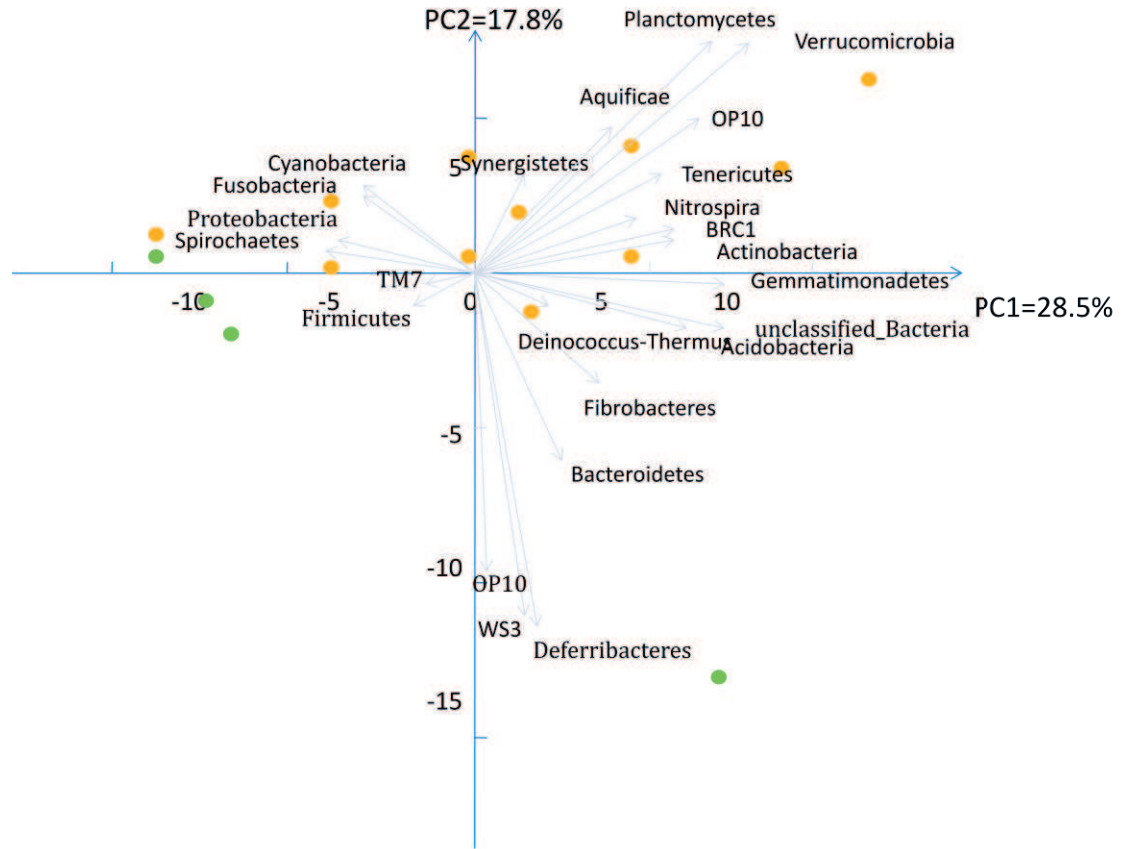


Fig. 4



Supplementary figure 1.

Normalized principal component analysis (PCA) using phyla composition data generated by R. Yellow points represent samples from 2011 and green points represent samples from 2010.



CHAPTER IV: Bacteria in Parisian drinking water distribution system

IV. A. Introduction d'étude

Après avoir analysé la population bactérienne dans les tempêtes de sable, et celles des déserts, nous poursuivons notre objectif de recherche à l'environnement aquatique. Ce travail est la continuation d'un travail antérieur sur l'étude de la communauté bactérienne dans les distributeurs d'Eau de Paris (180). En raison de la relation importante entre la présence de micro-organismes dans l'eau potable et la santé publique, des études de surveillance sont effectuées. La numérotation des bactéries hétérotrophes (HPC) à base de culture est utilisée dans le monde entier comme paramètre de qualité microbienne générale dans le traitement et la distribution de l'eau potable; même s'il est bien connu que seulement 0,25% à 1,6% des bactéries dans les réseaux d'eau potable peuvent être cultivé dans des conditions *in vitro* au laboratoire (79, 110).

La méthode conventionnelle de séquençage par clonage a été utilisée dans l'étude antérieure révélant un taux de couverture élevé de bactéries non classées dans les échantillons d'eau potable dans le distributeur d'Orly. Nous avons utilisé le séquençage à haut débit au sein de la population bactérienne sur trois échantillons d'eau et trois échantillons de biofilms provenant du distributeur de Paris. Le séquençage à haut débit peut aussi nous permettre de détecter les genres pathogènes en faible quantité.

Nos échantillons ont été prélevés dans une installation de distribution d'eau située à Orly, près de Paris. La source de l'eau est l'eau de surface de la Seine. À l'usine de traitement des eaux d'Orly, une ozonisation initiale est effectuée, suivie par les étapes de coagulation / floculation / décantation après l'injection de chlorure ferrique. L'eau est ensuite soumise à une filtration rapide sur sable (2-8 m/h), suivie par le chlore et la supplémentation d'acide phosphorique (180).

Nous avons suivi le flux d'eau provenant de l'usine d'Orly (DW-A) à l'entrée du réservoir (DW-B), et à la sortie du réservoir (DW-C). Nous avons constaté une forte variation de la communauté bactérienne, dans DW-A et DW-B, les bactéries prédominantes appartiennent aux populations des *Betaproteobacteria*, puis nous avons observé une conversion vers la population de *Alphaproteobacteria* dans DW-C.

L'analyse des trois échantillons de biofilm (BF-A, BF-B et BF-C) nous a montré que les bactéries prédominantes appartiennent aux populations des *Alphaproteobacteria* et des *Gammaproteobacteria*. Nous avons détecté un taux très élevé de bactéries non classées dans le DW-A et DW-B en accord avec l'étude antérieure. Parmi les séquences non classifiées, plus de 90% appartenaient aux deux OTUs (97% de similarité au niveau de séquence). Ces deux OTUs ont été signalés à plusieurs reprises dans le système d'eau potable indiquant leur prévalence dans un tel environnement (180, 184). Les relations de similitude entre les échantillons ont été montrées en utilisant des indices de similarité Morista. Le DW-C a montré une forte similitude avec un échantillon de biofilm (BF-C), ce qui suggère l'effet important du biofilm sur la modification des communautés bactériennes dans l'eau lors de la distribution.

Article III: Characterization of bacterial populations in drinking water and biofilms present in part of Parisian distribution network

*Shu An, Jean-Baptiste Poitelon and Michael S. DuBow**

Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay, France

* Correspondence: Michael DuBow, Univ. Paris-Sud, Laboratoire de Génomique et Biodiversité Microbienne des Biofilms (LGBMB), Institut de Génétique et Microbiologie, CNRS UMR 8621, Bâtiment 409, 91405 Orsay ; tel: +33 1 69 15 46 12 ; Fax : + 33 1 69 15 78 08 ; e-mail: michael.dubow@igmors.u-psud.fr

Abstract

Using pyrosequencing-based analyses of 16S rDNA amplicons, we characterized the bacterial communities in three water and three biofilm samples from one part of the Parisian drinking water distribution system. The total bacterial cell concentration in the water samples varied from 0.5 to 1.0×10^5 cells/ml. Our results showed that all samples were dominated by members belonging to the phylum *Proteobacteria*, ranging from 56.0% to 96.0% of the bacterial population. A dramatic change in bacterial population in the water during flow through the distribution of water was found. The dominant groups were changed from members belonging to phylum *Betaproteobacteria* to members belonging to phylum *Alphaproteobacteria* in the water leaving the reservoir. The richness of bacteria (number of OTUs) was reduced from the water treatment plant to the reservoir from 336 to 165 for water samples leaving the reservoir and from 947 to 275 for biofilm samples in the network. Two ubiquitous unclassified bacterial groups were detected in water samples before entering the reservoir (22.8% and 31.5%). Several OTUs belonging to pathogenic genera were detected in our samples, mostly in biofilm samples, thus suggesting that the biofilms may be an important source of bacteria during distribution to the consumers.

Key words: bacterial diversity, biofilm, drinking water system, pyrosequencing.

Introduction

Microbiological monitoring is crucial to maintain drinking water quality and public health. It is reported that 10 to 50% of diarrhea-associated illnesses are caused by waterborne microbial agents not yet identified (24). Previous studies have shown that only 0.25-1.6% of bacteria from drinking water can be cultured under laboratory conditions (2, 13, 20, 29). The microbiological monitoring of drinking water still relies heavily on conventional plating methods such as heterotrophic plate counts (HPC) and selective plating for pathogens such as *Escherichia coli* (2). Most importantly, there are some important waterborne pathogens (such as *Pseudomonas aeruginosa* or *Legionellae pneumophila*) that can be present in the water distribution system, but their presence shows no correlation with conventional indicator organism counts such as those for *E. coli* (9). Water-related pathogens can find niches to be protected in water systems by associating with biofilms or free-living amoebae, which makes their observation with conventional microbiological techniques more difficult (9).

To overcome the bias of culture-dependant methods, culture-independent methods are now being applied for microbial characterization, including in drinking water distribution systems (2, 21, 26, 28). Members belonging to phyla *Alphaproteobacteria*, *Betaproteobacteria*, and *Bacteroidetes* were the most reported dominant bacterial groups in drinking water distribution systems (8, 21, 31, 32).

The evolution of the bacterial population in water during flow within the distribution system was mostly examined during the water treatment (14, 21, 22, 25). After water treatment, the bacterial population of the water may also shift during flow in the water distribution network, but this has rarely been studied using non-cultivation techniques. The biofilms on the surface of pipes in drinking water distribution systems are a niche for bacteria (31), as the bacteria in the biofilm can be protected from antibacterial agents (6). The biofilm bacteria can affect the bacterial population in water by their detachment into the fluid phase. The interaction between planktonic bacteria and biofilm bacteria has already been suggested (19).

The microbial compositions of biofilms in drinking water distribution systems with the

major groups being detected using 454 DNA pyrosequencing included the *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* (15, 16). The richness levels varied from 200 to 1600 species per sample (15, 16). Moreover, the presence of potential pathogens, including members of the *Legionella*, *Pseudomonas*, *Aeromonas*, and *Chromobacterium* genera have been identified (21).

Here, we used pyrosequencing-based analysis of partial 16S rDNA amplicons to characterize the bacterial communities during the distribution of water in three locations in part of the Paris drinking water system. The water samples and biofilm samples of one Parisian drinking water distribution system were examined and the evolution of bacterial population during flow within the distribution was examined.

Materials and methods

Study site

The drinking water reservoir located near the town of l'Haÿ-les-Roses (GPS coordinates: 48°46'47"N, 2°20'40"E) was built in 1969. The reservoir collects and stores drinking water produced from the DWTP (drinking water treatment plant) located near the town of Orly (GPS coordinates: 48°44'55"N, 2°25'26"E). The Orly DWTP treats surface water retrieved from the Seine River. Following preozonation, the water is processed by combined coagulation-flocculation, rapid sand filtration, ozonation, granular activated carbon filtration, chlorination (free residual chlorine: >0.5 mg/l after 30 min contact time), and addition of phosphoric acid (residual concentration of at least 1mg/l). The distance between the Orly DWTP and the l'Haÿ-les-Roses reservoir is 7 km.

Sampling and sample processing

Drinking water (DW) was sampled from the Orly DWTP (DW-A), the inlet (DW-B) and outlet (DW-C) of the l'Haÿ-les-Roses reservoir (Fig. 1), on July 6, 2007. Water was collected using sterile bottles containing sodium thiosulfate, transported on ice and processed within 4 h after collection. Microorganisms were harvested by filtering 12 liters of water through a 0.2 µm pore size nylon filter (47 mm diameter, Millipore, France) and the filters were stored at -70°C until use.

The biofilm (BF) samples were collected on 17 July 2007 at the same location where drinking water was sampled during the yearly draining and maintenance, before cleaning. The biofilms were sampled from the inner surfaces of a cement-lined pipe (40 years in service) and the cement-lined walls (10 years of service) at the Orly location (BF-A) and the l'Haÿ-les-Roses reservoir (BF-B and BF-C) respectively (Fig 1). The biofilms were detached by scraping using sterile spatulas and rinsing with sterile deionized and filtered water. The samples were collected in 50 ml sterile conical-bottom centrifuge tubes. The surface areas sampled ranged from 105 to 45 cm². The samples were transported on ice and processed within 4 h. After centrifugation at 4,000 g for 20 min at 4°C, each supernatant fluid was collected and filtered through a

0.2 µm pore size nylon filter (47 mm diameter, Millipore, France). For each biofilm sample, the filter and pellet were stored at -70°C until use.

Nucleic acids extraction and purification

Nucleic acids from water samples were extracted from the frozen filters as previously described (25). For the biofilm samples, the frozen filters were thawed, cut into small strips with a sterile scalpel, transferred to centrifuge tube containing the biofilm pellets, and vortexed in 6.75 ml of extraction buffer (100 mM Tris-HCl, pH 8, 100 mM Na-EDTA, pH 8, 100 mM Na₂HPO₄, 1.5 M NaCl, 1% [wt/vol] CTAB). Following this, autodigested Pronase (75 µg.ml⁻¹ final concentration) was added and the tube incubated at 37°C for 2 h, with shaking. Then, 1.5 ml of 20% (wt/vol) SDS was added and incubation continued at 65°C for a further 2 h. After centrifugation at 6,000 g at room temperature for 20 min, the supernatant fluid was collected. The pellet was extracted two more times with 4.5 ml of extraction buffer supplemented with 0.5 ml of 20% SDS, followed by incubation at 65°C for 10 min. For each sample, the three supernatant fluids were pooled and the nucleic acids extracted by the addition of 1 vol of chloroform/isoamyl alcohol (24:1) and precipitated by the addition of 0.6 vol of isopropanol followed by an 1 h incubation at room temperature. After centrifugation at 30,000 g for 20 min, the nucleic acid pellet was air dried and resuspended in 50 µl of 0.1 x TE (1 mM Tris-HCl, pH 8, 0.125 mM Na-EDTA, pH 8). Since the biofilm sample from the Orly location was found to contain PCR inhibitors, nucleic acids from the filter and pellet were further extracted using the UltraClean™ Microbial DNA Isolation Kit (MoBio, France).

All purified nucleic acids were visualized by electrophoresis through a 0.8% agarose gel in TAE buffer (20 mM Tris-acetate pH 8, 5 mM Na-EDTA) and stored at -20°C until use.

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP)

The 16S rDNA sequences were amplified from the DNA samples using the forward primer 5'-GCCTTGCCAGCCCGCTCAGNNNNNNAGAGTTTGATCCTGGCTCAG-3' and the reverse primer 5'-GCCTCCCTCGCGCCATCAGNNNNNNCTGCTGCCTYCCGTA-3'. The underlined sequences represent the 454 primers and the bold

sequences correspond to the 8F and 357R 16S rRNA gene primers, respectively. The NNNNNN sequence designates the unique six base bar code used to tag each DNA sample (3, 12).

The PCR reactions were carried out in 50 μ l total final reaction volumes using 15 ng of DNA template, 0.1 μ M of each primer (Sigma-Aldrich, France), 0.2 mM dNTP mix (Fermentas, France). In order to reduce PCR amplification bias (3), the DNA templates were amplified in separate reactions with either 1.25 U Phusion DNA polymerase (Ozyme, France) or 1.25 U Pfu DNA Polymerase (Fermentas, France) using 1X Phusion Buffer (Ozyme, France) or 1X Pfu Buffer (Fermentas, France) respectively. Each DNA samples were subjected from five to ten PCR reactions per DNA polymerase to minimize single-tube amplification bias. The conditions of amplification applied with the Phusion DNA polymerase were: 98 °C for 2 min followed by 25 or 28 cycles of 98 °C for 30 sec, 48 °C for 20 sec and 72 °C for 12 sec, and a final elongation step at 72 °C for 5 min. The thermal conditions used with the Pfu DNA Polymerase (Fermentas, France) were: 95 °C for 3 min followed by 35 cycles of 95 °C for 30 sec, 48 °C for 30 sec and 72 °C for 48 sec, and a final elongation step at 72 °C for 5 min. The amplified fragments were purified using the Nucleospin Extract II Kit (Macherey-Nagel, France) after electrophoresis through a 1% agarose gel in TAE buffer. The PCR products from each sample were pooled and were sequenced using a Roche 454 FLX Pyrosequencer (Microsynth, Switzerland).

Sequence processing and data analyses

The total sequences obtained after pyrosequencing were first grouped into different samples according to the tag used. The average sequence was found to 227 nucleotides in length, without the tags.

The sequences with < 150 nt or more than 10% nt with quality scores < 25 (equal to 99.7% accuracy) were eliminated using the tools on the Greengenes website, trim (http://greengenes.lbl.gov/cgi-bin/nph-trim_fasta_by_qual.cgi) (7). Then, sequences with more than two errors in the primer or more than one ambiguous base were removed using the RDP pyrosequencing pipeline (<http://pyro.cme.msu.edu/index.jsp>). The chimera sequences were eliminated using B2C2 (Black Box Chemera Check)

software (11).

The remaining sequences were then classified to the genus level using RDP-II Naive Bayesian Classifier (23, 33) with the confidence threshold set to 50%. The richness estimated by Chao1 estimator and the diversity index by Shannon index were calculated on the RDP's pyro site (<http://pyro.cme.msu.edu/index.jsp>). The numbers of reads were normalized to 3030 reads using *Select.pl* of Pangea for better comparisons among samples (10). The similarity level among the different OTUs was fixed at 97%. The Morista similarity indices among samples were calculated using SPADE (5).

Results

Microbial and physico-chemical parameters were measured in the three water samples (Table 1). All six samples were also processed for 16S rDNA amplicon pyrosequencing. Over 104,480 sequences were obtained after pyrosequencing, and 61,203 high quality sequences were retained for further analyses.

The total bacterial cell concentration in all of the water samples varied from 0.5 to 1.0 x 10⁵ cells/ml. We observed an increase in cell concentration from 0.53 x 10⁵ to 1.0 x 10⁵ and a increase in pH of the water from 7.45 and 7.55 to 7.68 during flow from the treatment plant to the reservoir (see Fig. 1). The active percent cell ranged from 64.71% to 84.91% cells, measured by BacLight staining.

Bacterial composition analyses

The sequences in each sample were assigned using the RDP classifier with a 50% cutoff value, from the phylum to genus levels. Due to the length restriction of sequences from pyrosequencing, a species level classification was not performed.

The bacterial community composition at the phylum and class levels for *Proteobacteria* is summarized for each sample in Fig 2. In total, bacteria from 14 phyla were detected among samples. For the biofilm samples, BF-A contained bacteria from 13 different phyla, while the other sites contained bacteria from 6 different (BF-B) and 7 different (BF-C) phyla. In the water samples, DW-A and DW-B contained bacteria from 7 different phyla, while DW-C contained bacteria from 5 different phyla. The common phyla detected in all samples included *Proteobacteria*, *Firmicutes*, *Bacteroidetes* and *Actinobacteria*, with from 68.0% to 99.6% of sequences in each sample classified as belonging to one of these four phyla. Member from other phyla detected were grouped as “rare” phyla, including *Fusobacteria*, *Spirochaetes*, *Cyanobacteria*, *Nitrospira*, *Gemmatimonadetes*, *Acidobacteria*, *Deinococcus-Thermus*, *Verrucomicrobia*, *Planctomycetes* and *Chloroflexi*. The percentages of bacteria belonging to these phyla were relatively low (sum < 2% in each sample).

Members belonging to the phylum *Proteobacteria* were preponderant in all our samples with their proportional abundance ranging from 56% to 96%. At the class level

of Proteobacteria, the composition varied in the different water samples. We observed the evolution of *Proteobacteria* composition of the water from DW-A to DW-C. The predominant *Proteobacteria* class was the *Betaproteobacteria* in DW-A (53.0%) and DW-B (35.1%). The *Alphaproteobacteria* increased to 70.9% of the bacterial population in DW-C, though the *Betaproteobacteria* remained abundant in DW-C (23.1%). The *Proteobacteria* class in the three biofilm samples also varied (Fig 2). We note that, compared to the water samples, there was a significantly higher proportion of *Gammaproteobacteria* in the biofilm samples ($P = 0.04$).

Unclassified sequences

Between 22.8% and 31.5% sequences were assigned to unclassified phyla in DW-A and DW-B (Fig 2). Such a significant fraction of unclassified sequences were not detected in the other samples DW-C (2.9%), BF-A (2.1%), BF-B (0.4%) and BF-C (0.3%).

In total, 3054 sequences were assigned to unclassified bacteria using RDP classifier. These sequences were merged together, and clustered into OTUs at a 97% similarity level. One representative sequence was selected from each cluster to perform BLAST searches against Genbank. In total, 164 clusters were grouped, one dominant group (group A) comprising 79.0% of the sequences with another group (group B) comprising 8.2% of the sequences. The rest of the clusters represented rare OTUs with less than 1% relative abundance. The representative sequences from groups A and B had 99%-100% similarity with the unclassified bacterial sequences reported in other drinking water studies (26, 27).

Sequence diversity

Table 2 lists the statistical analyses data for each sample. The BF-A sample showed a much higher diversity and richness compared to the other samples. We observed that the number of OTUs was reduced from the Orly site to the reservoir exit (Table 2 and Fig 3). In total, 1216 OTUs were detected among all the samples. Only 8 OTUs were detected as present in all the samples. The rank relative abundance curves of OTUs were merged into a DW group and a BF group in Fig S1. We observed that the samples had few OTUs (4–15 OTUs) that contained more than 1% of the sequences. Most of the

OTUs belonged to rare groups. For example in the sample DW-A, the 6 most abundant OTUs represented a total of 92.3% of the sequences, while the 203 rare OTUs represented only 7.7% of the sequences in total.

Morista's similarity indices (34) were calculated using SPADE in order to compare the similarity relations among the samples at the sequence level. We observed a high similarity between DW-A and DW-B (0.841), and as well as between DW-C and BF-C (0.909), with the lowest similarity observed between the DW-C and BF-B samples (0.002) (Table 3).

The most abundant genera among the classifiable sequences

The sequences classifiable to the genus level were not equivalent among the samples. Using the 10 most abundant classifiable sequences for each sample (Table 4), the proportion of sequences classifiable to the genus level was 9.6% for DW-A, 21.4% for DW-B, 3.2% for DW-C, 66.1% for BF-A, 86.5% for BF-B and 33.1% for BF-C. The low level of classifiable sequences, particularly in the DW samples, suggests a high proportional of unknown bacteria. Most of the sequences in the DW samples could not be classified. In samples DW-A and DW-B, most sequences were clustered to unclassified bacteria. In sample BF-A and sample DW-C, most sequences were clustered as *unclassified Rhizobiales*.

We observed the genus *Nevskia* as one predominant genus in all three biofilm samples and in the DW-C sample, while it did not appear in the DW-A and DW-B samples. Members belonged to the genus *Acinetobacter* were present in the three water samples and two biofilm samples. Members of the genera *Burkholderia*, *Chryseobacterium*, *Orientia* were only found to be dominant in the water samples.

Potential waterborne pathogens

Members belonged to the potential pathogenic genera were detected in some of the samples including *Leptospira*, *Salmonella*, *Shigella*, *Legionella*, *Vibrio*, *Helicobacter*, *Mycobacterium*, *Giardia*, *Aeromonas*, *Pseudomonas* (Table 5). Most of them were found in the BF-A sample.

Discussion

The exploration of bacterial communities in drinking water distribution systems is essential for public health. Using high throughput sequencing technology, our study of the bacterial community from water and biofilm samples can provide a more accurate information to safeguard the consumer, and thus improve our understanding of the water distribution system. From the exit of the treatment plant to that of reservoir, we observed a dramatic shift of bacterial population in water samples. A high similarity of downstream water sample with the biofilm samples was observed, suggesting an important effect of biofilm bacteria on downstream water bacteria.

The average total cell count for the drinking water samples was 3×10^5 cells/ml (8). The BacLight kit was used to detect the live/dead cells in current work, and the cell concentrations observed (0.52×10^5 - 1.0×10^5 cells/ml) in three water samples showed a similar level with previous results (2).

The percentage of the live bacteria in samples can be between 50 and 84% of the total bacteria in the effluents, depending on the type of treatment processes used for drinking water (14). For drinking water samples obtained from a tap in Germany, the fraction of membrane intact cells counted microscopically accounted for $53\% \pm 6\%$ of the total bacteria (17). In non-chlorinated household tap water collected in Switzerland, 62% of intact cells were observed (2). In this study, the active bacterial numbers ranged from 64.71% to 84.91%.

The richness estimated by Chao1 in each sample was dramatically reduced when we normalized the number of reads to 3030 sequences. It appears that 3030 reads in these samples is not enough for a complete richness estimation, thus suggesting more reads are necessary for an accurate estimation of total bacterial richness in our samples (Table 2).

Bacterial communities throughout the drinking water distribution system can differ drastically from source water to tap water (8). We observed a large difference among the samples after water treatment. Similarly with many previous drinking water studies, members of the phylum *Proteobacteria* were reported as the dominant group in

drinking water (2, 17, 28). We observed a proportional increase of *Alphaproteobacteria* and a proportional decrease of the unclassified bacteria from DW-A to DW-C. Similarly, members of *Betaproteobacteria* have been reported as dominant bacterial group in raw surface water before water treatment. Members of *Alphaproteobacteria* was detected as a dominant group in the downstream water treatment area (21). The large proportion of *Gammaproteobacteria* in biofilm samples was not detected in the DW-C sample. During a bacterial live/dead status research in drinking water, the largest percentage of dead phylotypes (i.e. no cells detected in the live sorted fraction) was observed for the *Gammaproteobacteria* (38%) suggesting that the *Gammaproteobacteria* survive better in biofilms which can protect them from disinfection agents, as they are normally sensitive to disinfection stress (17).

The genus *Nevskia* was detected as a predominant genus in all three biofilm samples. The abundance of *Nevskia* in biofilms of drinking water distribution systems has been previously reported (18, 30), but rarely reported in drinking water samples. One hypothesis is that members of *Nevskia* may not withstand disinfection stress.

The high similarity between the DW-C and BF-C samples suggests that the biofilms play a important role in influencing the final tap water bacterial community, and the bacteria in water from the outlet of the reservoir could significantly differ from the bacteria in water from the inlet of reservoir. The detachment of biofilms further can contribute to compliance failures and impact water quality (4).

We observed a decrease of OTUs number from the DW-A to DW-C. The percentage of culturable bacteria in the bulk water also decreased as residence times increased (30). Our results confirm that longer times in disinfection agents for bacteria can decrease the richness of bacterial diversity. The BF-B and BF-C samples had comparable OTUs numbers with previous biofilm water meters studies (133 and 208 OTUs) (15). The bacterial richness for samples after water treatment were much less than one previous pilot-scale study for bacterial diversity in raw water, membrane tank particulate matter, and membrane biofilm (over 1100 OTUs) using pyrosequencing (21).

Similar groups of unclassified bacteria (at least two subgroups) have been reported as a significant fraction of the bacterial populations in drinking water systems (21, 26, 27).

As they were only detected in the drinking water system, we suggest that that they represent an important group of bacteria that can survive under disinfection stress.

The potential pathogens were mostly detected in BF-A which had the highest diversity and richness compared with the other samples. This sample was taken at the exit of the treatment plant. More efforts are required to avoid the biofilm formation in this location in order to better control the pathogens.

Reference

1. **Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz.** 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**:8966-9.
2. **Berney, M., M. Vital, I. Hulshoff, H. U. Weilenmann, T. Egli, and F. Hammes.** 2008. Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water Res* **42**:4010-8.
3. **Binladen, J., M. T. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev.** 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* **2**:e197.
4. **Chandy, J. P., and M. L. Angles.** 2001. Determination of nutrients limiting biofilm formation and the subsequent impact on disinfectant decay. *Water Res* **35**:2677-82.
5. **Chao, A., and T.-J. Shen.** 2010. Program SPADE (Species Prediction And Diversity Estimation). Program and User's Guide published at <http://chao.stat.nthu.edu.tw>.
6. **Davies, D.** 2003. Understanding biofilm resistance to antibacterial agents. *Nat Rev Drug Discov* **2**:114-22.
7. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol* **72**:5069-72.
8. **Eichler, S., R. Christen, C. Holtje, P. Westphal, J. Botel, I. Brettar, A. Mehling, and M. G. Hofle.** 2006. Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Appl Environ Microbiol* **72**:1858-72.
9. **Felfoldi, T., Z. Heeger, M. Vargha, and K. Marialigeti.** 2009. Detection of potentially pathogenic bacteria in the drinking water distribution system of a hospital in Hungary. *Clin Microbiol Infect* **16**:89-92.
10. **Gontcharova, V., E. Youn, R. D. Wolcott, E. B. Hollister, T. J. Gentry, and S. E. Dowd.** 2010. Black Box Chimera Check (B2C2): a Windows-Based Software for Batch Depletion of Chimeras from Bacterial 16S rRNA Gene Datasets. *Open Microbiol J* **4**:47-52.
11. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* **5**:235-7.
12. **Hammes, F., M. Berney, Y. Wang, M. Vital, O. Kosterb, and T. Egli.** 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Res* **42**.
13. **Ho, L., K. Braun, R. Fabris, D. Hoefel, J. Morran, P. Monis, and M. Drikas.** 2012. Comparison of drinking water treatment process streams for optimal bacteriological water quality. *Water Res* **46**:3934-42.
14. **Hong, P. Y., C. Hwang, F. Ling, G. L. Andersen, M. W. LeChevallier, and W. T. Liu.** 2010. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Appl Environ Microbiol* **76**:5631-5.
15. **Hwang, C., F. Ling, G. L. Andersen, M. W. Lechevallier, and W. T. Liu.** 2011. Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. *Microbes Environ* **27**:9-18.

16. **Kahlisch, L., K. Henne, L. Groebe, J. Draheim, M. G. Hofle, and I. Brettar.** 2010. Molecular analysis of the bacterial drinking water community with respect to live/dead status. *Water Sci Technol* **61**:9-14.
17. **Keinanen-Toivola, M. M., R. P. Revetta, and J. W. Santo Domingo.** 2006. Identification of active bacterial communities in a model drinking water biofilm system using 16S rRNA-based clone libraries. *FEMS Microbiol Lett* **257**:182-8.
18. **Kerr, C., K. Osborn, G. Robson, and P. Handley.** 1999. The relationship between pipe material and biofilm formation in a laboratory model system. *J Appl Microbiol* **85**:29S-38S.
19. **Kormas, K. A., C. Neofitou, M. Pachiadaki, and E. Koufostathi.** 2009. Changes of the bacterial assemblages throughout an urban drinking water distribution system. *Environ Monit Assess* **165**:27-38.
20. **Kwon, S., E. Moon, T. S. Kim, S. Hong, and H. D. Park.** 2011. Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. *Microbes Environ* **26**:149-55.
21. **Le Dantec, C., J. P. Duguet, A. Montiel, N. Dumoutier, S. Dubrou, and V. Vincent.** 2002. Occurrence of mycobacteria in water treatment lines and in water distribution systems. *Appl Environ Microbiol* **68**:5318-25.
22. **Liu, Z., C. Lozupone, M. Hamady, F. D. Bushman, and R. Knight.** 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* **35**:e120.
23. **Maier, R. M., I. L. Pepper, and C. P. Gerba.** 2009. *Environmental microbiology*, 2 ed, vol. Elsevier, Oxford.
24. **Norton, C. D., and M. W. LeChevallier.** 2000. A pilot study of bacteriological population changes through potable water treatment and distribution. *Appl Environ Microbiol* **66**:268-76.
25. **Poitelon, J. B., M. Joyeux, B. Welte, J. P. Duguet, E. Prestel, O. Lespinet, and M. S. DuBow.** 2009. Assessment of phylogenetic diversity of bacterial microflora in drinking water using serial analysis of ribosomal sequence tags. *Water Res* **43**:4197-206.
26. **Revetta, R. P., R. S. Matlib, and J. W. Santo Domingo.** 2011. 16S rRNA gene sequence analysis of drinking water using RNA and DNA extracts as targets for clone library development. *Curr Microbiol* **63**:50-9.
27. **Revetta, R. P., A. Pemberton, R. Lamendella, B. Iker, and J. W. Santo Domingo.** 2010. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water Res* **44**:1353-60.
28. **Siebel, E., Wang, Y., Egli, T., and Hammes, F.** 2008. Correlations between total cell concentration, total adenosine tri-phosphate concentration and heterotrophic plate counts during microbial monitoring of drinking water. *Drink. Water Eng. Sci.* **1**:1-6.
29. **Srinivasan, S., G. W. Harrington, I. Xagoraki, and R. Goel.** 2008. Factors affecting bulk to total bacteria ratio in drinking water distribution systems. *Water Res* **42**:3393-404.
30. **Tokajian, S. T., F. A. Hashwa, I. C. Hancock, and P. A. Zalloua.** 2005. Phylogenetic assessment of heterotrophic bacteria from a water distribution system using 16S rDNA sequencing. *Can J Microbiol* **51**:325-35.
31. **Ultee, A., N. Souvatzi, K. Maniadi, and H. Konig.** 2004. Identification of the culturable and nonculturable bacterial population in ground water of a municipal water supply in Germany. *J Appl Microbiol* **96**:560-8.

32. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**:5261-7.
33. **Wolda, H.** 1980. Similarity indices, sample size and diversity. *Oecologia* **50**:296-302.

Figures and Tables

Table 1: Sampling condition data and cell number in the three water samples. The cell numbers were counted by BacLight kitTM, results here were the average means from ten replicats for each sample.

	DW-A	DW-B	DW-C
pH	7.45	7.55	7.68
Turbidity (NFU) _a	0.02	0.03	0.04
Free chlorine (mg/l)	0.66	0.60	0.48
Total chlorine (mg/l)	0.73	0.74	0.57
Live cells/ml	4.50E+04	4.40E+04	6.70E+04
Dead cells/ml	8.60E+03	2.40E+04	3.30E+04
Total cells/ml	5.36E+04	6.80E+04	1.00E+05
% live cells	84.91%	64.71%	67.00%

_a FNU: *Formazin Nephelometric Units*

Table 2: Sequence data analyses. The bacterial population richness was estimated using Chao 1 estimator.

Sample	Sequence Number	Raw data			Normalised to 3030 sequences		
		OTUs	Chao1	Shannon index	OTUs	Chao1	Shannon index
DW-A	6686	195	336	1.88	137	276	1.98
DW-B	3030	127	201	2.41	127	201	2.41
DW-C	13493	100	165	1.11	44	58	1.01
BF-A	3398	671	947	5.56	641	915	5.54
BF-B	22653	284	463	2.92	84	151	2.36
BF-C	11943	162	275	1.46	84	147	1.32

Table 3: Morista's similarity indices among samples, they were calculated by SPADE. The values in parentheses were calculated by normalized data. The two highest similarity indices values are in bold.

	DW-A	DW-B	DW-C	BF-A	BF-B	BF-C
DW-A	1					
DW-B	0.841(0.842)	1				
DW-C	0.274(0.231)	0.202(0.168)	1			
BF-A	0.036(0.038)	0.073(0.063)	0.025(0.023)	1		
BF-B	0.002(0.002)	0.004(0.003)	0.203(0.268)	0.049(0.041)	1	
BF-C	0.004(0.002)	0.005(0.005)	0.909(0.933)	0.025(0.023)	0.236(0.301)	1

Table 4: The 10 genera that contained the most relative abundant bacteria in each sample. The values in parentheses showed the % of sequences which could be classified to the genus level. The genera in bold were genera appeared more than once.

	DW-A (9.5%)	DW-B (21.4%)	DW-C (3.2%)	BF-A (66.1%)	BF-B (86.4%)	BF-C (33.0%)					
<i>Bacillus</i>	5.325%	<i>Acinetobacter</i>	8.812%	<i>Hyphomicrobium</i>	1.141%	<i>Acinetobacter</i>	9.594%	<i>Nevskia</i>	38.397%	<i>Nevskia</i>	20.91%
<i>Chryseobacterium</i>	0.838%	<i>Undibacterium</i>	1.947%	<i>Acinetobacter</i>	0.371%	<i>Flavobacterium</i>	6.062%	<i>Undibacterium</i>	8.657%	<i>Sireptococcus</i>	2.45%
<i>Nitrospira</i>	0.568%	<i>Comamonas</i>	1.287%	<i>Bacillus</i>	0.371%	<i>Limnohabitans</i>	3.796%	<i>Methylibium</i>	7.518%	<i>Hyphomicrobium</i>	2.25%
<i>Orientia</i>	0.329%	<i>Acidovorax</i>	0.759%	<i>Undibacterium</i>	0.163%	<i>Pseudomonas</i>	3.620%	<i>Cellvibrio</i>	7.067%	<i>Sphingopyxis</i>	1.58%
<i>Undibacterium</i>	0.314%	<i>Chryseobacterium</i>	0.627%	<i>Orientia</i>	0.126%	<i>Herminiimonas</i>	3.031%	<i>Sphingopyxis</i>	6.644%	<i>Methylibium</i>	0.84%
<i>Flavobacterium</i>	0.269%	<i>Burkholderia</i>	0.627%	<i>Nevskia</i>	0.119%	<i>Bacillus</i>	2.972%	<i>Pseudomonas</i>	3.982%	<i>Acinetobacter</i>	0.65%
<i>Ideonella</i>	0.224%	<i>Cloacibacterium</i>	0.561%	<i>Burkholderia</i>	0.111%	<i>Sphingomonas</i>	1.560%	<i>Flavobacterium</i>	1.404%	<i>Nitrospira</i>	0.59%
<i>Acinetobacter</i>	0.150%	<i>Bacillus</i>	0.561%	<i>Brevundimonas</i>	0.104%	<i>Nevskia</i>	1.501%	<i>Brevundimonas</i>	1.232%	<i>Sphingobium</i>	0.56%
<i>Burkholderia</i>	0.120%	<i>Stenotrophomonas</i>	0.462%	<i>Delftia</i>	0.104%	<i>Comamonas</i>	1.413%	<i>Dongia</i>	1.192%	<i>Sphingomonas</i>	0.35%
<i>Bdellovibrio</i>	0.105%	<i>Sphingomonas</i>	0.429%	<i>Sphingobium</i>	0.082%	<i>Methylotenera</i>	1.324%	<i>Rugamonas</i>	0.998%	<i>Naxibacter</i>	0.29%

Table 5: The members belonged to potential pathogenic genera were detected in samples. The values showed the percentage of sequence in each sample.

Genus	DW-A	DW-B	DW-C	BF-A	BF-B	BF-C
<i>Leptospira</i>	0	0	0	0.09%	0	0
<i>Salmonella</i>	0	0	0.01%	0	0	0
<i>Escherichia/Shigella</i>	0	0.40%	0.06%	0.09%	0	0
<i>Legionella</i>	0	0	0.01%	0.35%	0.09%	0
<i>Vibrio</i>	0	0	0	0.06%	0	0.03%
<i>Helicobacter</i>	0	0	0	0.03%	0	0
<i>Mycobacterium</i>	0.07%	0	0.01%	0.06%	0.53%	0.10%
<i>Aeromonas</i>	0.03%	0	0.01%	0.26%	0	0
<i>Pseudomonas</i>	0.09%	0.20%	0.01%	3.62%	3.98%	0.15%
Total	0.19%	0.59%	0.12%	4.56%	4.60%	0.28%

Figure legends

Figure 1: Sampling positions in Orly drinking water distribution system (DWDS).

DW= Drinking water, BF=Biofilm. DW-A and biofilm sample BF-A were taken after the Orly treatment plant, DW-B was taken in the inlet of reservoir, DW-C was taken in outlet of reservoir. BF-B was taken in the reservoir near the inlet, BF-C was taken in the reservoir near the outlet.

Figure 2: Bacterial composition in our samples. a. The composition of phyla in our samples. b. The composition of class in phyla *Proteobacteria* in our samples.

Figure 3: Rarefaction curves in each sample. 3030 sequences were used to perform the rarefaction curve in each sample.

Fig. 1

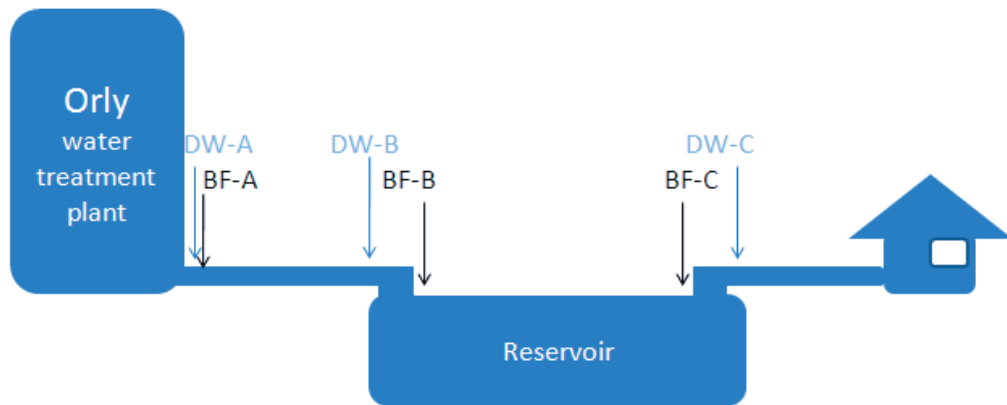
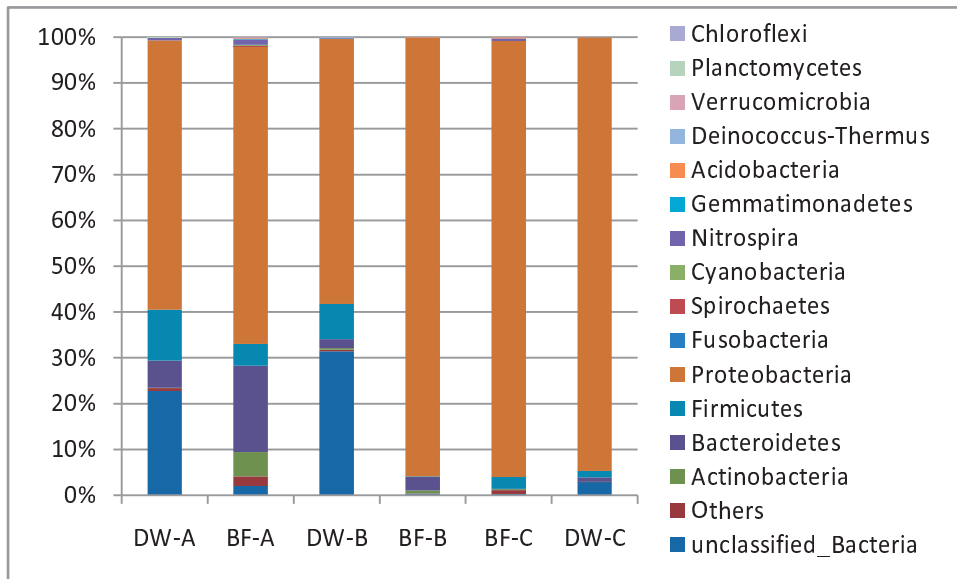


Fig. 2

a.



b.

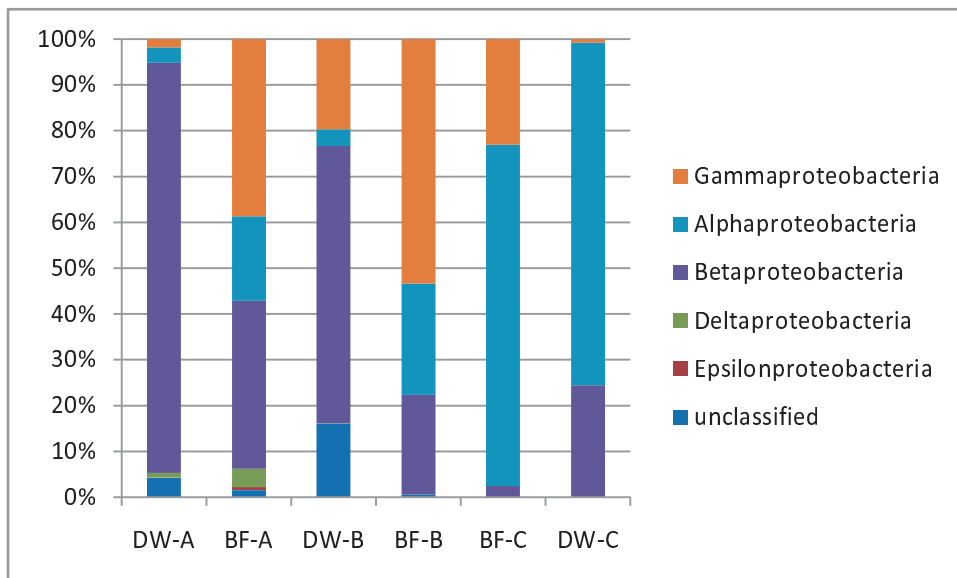
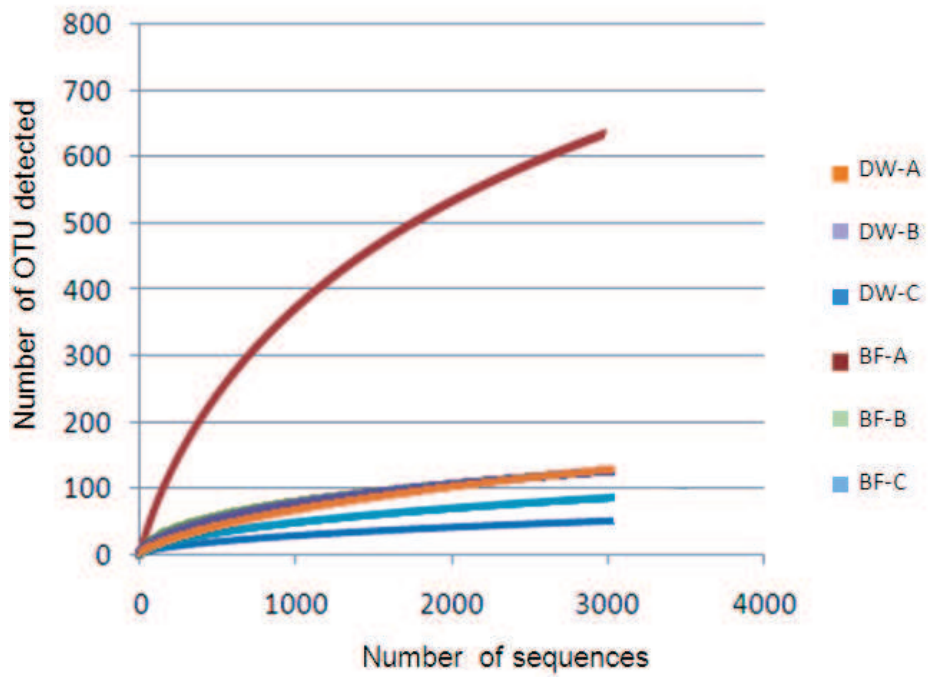


Fig. 3



CHAPTER V: General discussion and Future perspectives

High-throughput sequencing, in our case pyrosequencing, is a powerful technology that has been applied to several oligotrophic environments (88, 134, 218). During the course of my thesis, we used 16S rDNA amplicon pyrosequencing from samples from three different oligotrophic environments: the surface sand of Asian deserts, the sand and dust of sandstorms that fell in several Asian cities, and samples from Paris's water distribution system, in order to reveal the level of their bacterial diversity.

The first task of my thesis was to be able to prepare quality DNA for pyrosequencing from these oligotrophic environments.

V. A. DNA extraction from oligotrophic environments

Our aim was to be able to extract sufficient quantities DNA from environments with low organic matter content and other stresses for microbial (and DNA) survival including high iron content in sand causing free radical formation and DNA degradation. The bacteria on the surface sand of deserts and sandstorm samples were also under sources of stress including UV exposure, desiccation and high temperature variations (72), while bacteria from drinking water distribution systems were under stress, from chlorine exposure. In order to survive in oligotrophic environments, some bacteria may form spores with enhanced resistance to environmental stress, making their DNA isolation an extra challenge. We therefore applied the DNA extraction method of Zhou *et al.* (242) with a added pre-incubation step using 1/4 Tryptic Soy Broth to rehydrate bacteria while minimizing nutrient shock (183) to improve the efficacy of DNA extraction. Our protocol allowed us to detect from 649 to 4026 OTUs in dust samples, even though only 0.5 g to 2 g of sand was collected, as it is a challenge to obtain large samples from such samples.

V. B. Partial 16S rDNA sequencing

Pyrosequencing technology allows fast, qualitative and quantitative DNA analyses

without cloning steps, and is therefore less laborious and costly than classical Sanger DNA sequencing. The application of 16S rDNA amplicon pyrosequencing in oligotrophic environments makes it possible to detect their microbial community structures. The large number of obtained DNA sequences also allows the detection of more rare species than possible using conventional Sanger DNA sequencing techniques (52).

For pyrosequencing of 16S rDNA amplicons, we amplified the hypervariable V1V2 regions of 16S rDNA when using the 250nt ‘Gold’ DNA sequencing technology or the hypervariable V1V2V3 regions of 16S rDNA when using the 400nt ‘Titanium’ sequencing. Hypervariable 16S region rDNA sequencing is considered an effective means for assigning bacterial taxonomy and provides great advantages over traditional microbial cultivation sampling and Sanger DNA sequencing. Using Global Alignment for Sequence Taxonomy (GAST), the hypervariable region can map to the same taxonomic unit as a full length sequence with a $> 99\%$ correlation rate for commonly studied environments such as the human microbiome, and $\geq 96\%$ for less commonly studied environments such as deep-sea vents (91). The V2 region was reported to provide the maximum accuracy for phylogenetic determination using RDP classifier (223). The V1V2 and V3 regions were recommended by Liu *et al.* (2008) for assigning taxonomy (132). Using 16S rDNA amplicon pyrosequencing greatly enhances projects exploring the bacterial composition, diversity, and distribution of microbial populations (75, 87, 116). Youssef *et al.* (2009) reported that the V1V2 regions can overestimate the bacterial richness in samples, but the biases were slight when a 3% taxonomic cutoff was applied (240).

The advantage of hypervariable region sequencing is that it can use massively-parallel pyrosequencing, sampling to depths several orders of magnitude greater than previously achieved, and facilitating the exploration of the vast diversity of microbial populations and the rare biosphere (91). However, Sanger sequencing is still widely used for deep identification of bacteria with less massive quantities (normally for

hundreds of sequences), as longer sequences (such as full length 16S rDNA) can be obtained with a very inexpensive price per run.

Other genes, RNA, proteins, lipids or polysaccharides are sometimes used to detect microbial groups (10, 58). These molecules can not be used easily 16S rDNA to define new species due to cost plus the fact that the levels of resolution are not as high as for 16S rDNA. They are often used for known bacterial detection, and therefore rarely applied in oligotrophic environment studies where many unknown bacteria may be found.

V. C. Comparaison of high-throughput sequencing technology

The limit of sequence length is considered a weak point when comparing NGS with traditional (Sanger) sequencing. Common limiting factors include the degrading effects of lasers on DNA and enzymes used in sequencing, and the effects of cyclical washes, which slowly reduce the amount of DNA that can be sequenced. The phasing problem limits read length as well. This problem occurs because some of the molecules are not incorporated during one or more of the cycles. These errors build up cumulatively throughout the cycles. The technology providers are constantly improving protocols and reagent mixtures to increase read lengths (139). Among the most famous high-throughput sequencing technologies, including Roche (454) pyrosequencing, Illumina (Solexa) and ABI SOLiD, pyrosequencing is considered the most suitable for alignment, since it can provide the longest read lengths. Currently, 400 nt of sequence can be obtained by pyrosequencing, while only 100 nt of sequence can be obtained by Illumina (Solexa) and 50 nt by ABI SOLiD. The Illumina and ABI machines can yield 50 gb per run, while pyrosequencing can yield only 0.5 gb per run (107).

Besides the longer reads, the 454 pyrosequencing platform is prone to indel errors near runs of multiple nucleotides (homopolymers). On the positive side, the substitution

error rate is very low (68).

A unique advantage of the ABI SOLiD platform is its di-base encoding scheme, in which each base is effectively called twice in a sequencing read. The availability of two calls per base makes it possible to distinguish between sequencing errors and true variation, thereby improving the overall accuracy of reads from this platform (107). Recent developments of benchtop next-generation sequencers will allow the application of personal genome machine (PGM) on NGS in the near future becoming routine in laboratories (222). Three benchtop high-throughput sequencing instruments are now available including the 454 GS Junior (Roche, 2010), MiSeq (Illumina, 2011) and Ion Torrent PGM (Life Technologies, 2011). They are small and offer modest set-up and running costs (135). The 454 GS Junior continues to generate the longest reads (up to 600 nt) among the three instruments making it a best choice for population diversity studies, while the other two instruments generate about 100 nt but at higher throughput. It is worthy to note that the 454 GS Junior produces also homopolymer-associated indel errors (0.38 errors per 100 bases).

The SMRT sequencing (PacBio *RS*, 2011) can sequence up to 3000 nt with 75,000 sequences per SMRT cell. With such long reads, it aims to fully characterize genetic complexity (35).

V. D. Bacterial viability

When total DNA is extracted from a sample, it is possible that the DNA of dead cells is also collected. Due to the persistence of DNA in the environment after cells have lost their viability, DNA-based quantification methods can overestimate the number of (viable) cells in mixed populations, or even lead to false-positive results in the absence of viable cells (165). The distinction between viable and dead bacterial cells poses a major challenge in microbial community studies. One solution is the application of RNA-based sequencing technology such as Reverse transcription polymerase chain reaction (RT-PCR) to detect only live bacteria (152). Besides RNA-based sequencing,

another promising and easy-to-use method uses the ethidium monoazide bromide (EMA) treatment, which can penetrate only into dead cells and thus inhibit the amplification of DNA from dead cells (165). FISH/CARD-FISH can be used for microscopic differentiation to reveal the *in situ* bacterial community (230).

V. E. Importance of detection of rare members in oligotrophic environments

Due to the relatively small size of many Sanger-type culture-independent sequencing surveys, a detailed phylogenetic analysis of rare members of a bacterial community is lacking (52). Pyrosequencing can offer up to 4,000,000 sequences per run, other methods even more than that, making the detection of low abundance sequences possible. The greater sampling depth afforded by pyrosequencing uncovers not only the dominant microbial species, but many more members of the ‘rare biosphere’ (91).

Rare species have been reported to be important in oligotrophic environments (52, 63, 86, 206). Previous studies showed that community functioning following an environmental perturbation can depend on the presence of rare (as well as abundant) species in oligotrophic environments (206). The detection of rare species (OTUs) can provide clues regarding the origins and potential ecological roles of members of the rare biosphere (52). In the microbial biosphere study of the Arctic Ocean, Galan *et al.* (2009) concluded that the tremendous diversity of the rare biosphere is likely subject to ecological processes such as selection, speciation, and extinction (63). The members of the rare biosphere may act to form a seedbank that provide the genetic potential to respond to dramatic ecosystem changes (86), and thus any rare species may become a dominant species in the future.

V. F. Pathogen detection

The oligotrophic environments we examined have been implicated in human and

environmental health, and we noted potentially pathogenic members of genera in a water distribution system, in duststorm samples. The presence of pathogens in these environments may yield a potential health threat for humans, and therefore require more attention. The detection of pathogens by pyrosequencing at the genus level has been previously reported (22). DNA pyrosequencing-based bacterial identification can be a valuable tool that markedly improves bacterial pathogen identification (138), and can be applied to other (than 16S rDNA) gene sequences offering a better identification for pathogens (10, 220).

V. G. Deeper identification of unclassified bacteria

During our study, we detected a significant fraction of unclassified bacteria in desert and water samples which showed that there are still many unknown bacteria remaining to be characterized in oligotrophic environments. Almost all the unclassified sequences can be assigned to specific groups in Genbank with a high similarity (97%) by BLAST. These results indicated that many of these unclassified bacteria are ubiquitous, and thus worthy of further characterization. Others were often detected in oligotrophic environments and, because less data is found for them in public databases, leads to a lower level of taxonomic characterization of them (91).

Due to the sequence length limit of pyrosequencing, we may, in the future, amplify and sequence the 16S rDNA in two or three steps. This means for the same samples, we can perform two or three series of PCR reactions using different pairs of primers in order to amplify two or three amplicons for each 16S rDNA. The amplicons should contain overlapping regions to facilitate final 16S rDNA gene assembly. For example, a first primer pair for the V1 to V3 region, a second primer pair for the V3 to V6 region and the third pair primer for the V6 to V9 region. Assembly of short regions can provide the longer sequences more accurate by bacterial identification. However this would require more complex bioinformatic approaches (Fig 9). It will also show primer pair biases.

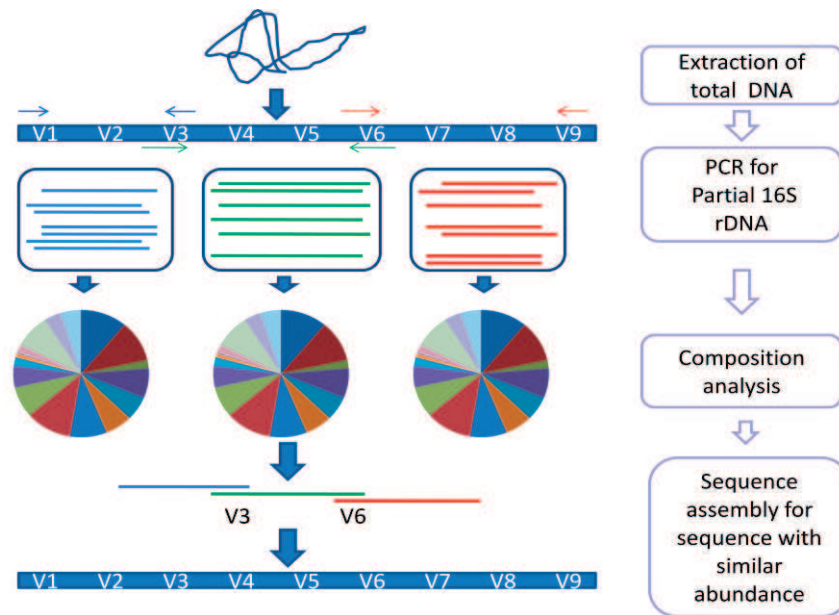


Figure 9: The schematic presentation of multiregion pyrosequencing for bacterial identification.

As the technology continues to improve, yielding greater read counts and longer sequences, NGS will provide even greater opportunities for sequencing, such as the use of longer hypervariable regions or combinations of variable regions, and ever-greater sampling depth (91).

V. H. Bioinformatics for high-throughput sequencing analyses

Next generation sequencing is revolutionizing biology. The development of bioinformatic applications for high-throughput sequence analyses is urgent and challenging. Over the past few years, the major challenge associated with environmental genome studies has shifted from generating sequences to analyzing them (187).

During my thesis, we mainly used web-based tools such as RDP-II, Greengenes, Ade4 of software R, since local analyses would require very powerful computer facilities. Different analysis tools are often used in different studies, making comparison among

different studies difficult. User friendly bioinformatic tools for each kind of high-throughput analysis are required. For any high throughput sequence data, a series of bioinformatic tools can be used for different analyses in order to extract as much information as possible. RDP's pyro and Mothur (192) are the two most used analysis tools for such studies.

RDP's pyro is an online web server which is provided by the Ribosome database project (RDP) team. The Ribosomal Database Project (RDP-II) provides the research community with aligned and annotated rRNA gene sequences, along with analysis services and a phylogenetically consistent taxonomic framework for these data. Updated monthly, these services are made available through the RDP-II website (<http://rdp.cme.msu.edu/>). RDP-II release 10.29 (June 1, 2012) contains 2,319,039 bacterial small subunit rRNA gene sequences in aligned and annotated format. High-throughput tools for initial taxonomic placement, identification of related sequences, probe and primer testing, data navigation and subalignment download are provided (37). The RDP's pyro site is widely used, as its article (37) has been cited 808 times as of 16 July 2012.

Mothur is program that requires downloading by the user. The mothur project was initiated by Dr. Patrick Schloss and his software development team in the Department of Microbiology & Immunology at The University of Michigan. This project seeks to develop a single piece of open-source, expandable software to fill the bioinformatics needs of the microbial ecology community. The first version was released in February 2009, which had accelerated versions of the popular DOTUR and SONS programs. Since then, they have added the functionality of a number of other popular tools including s-libshuff, TreeClimber (i.e. the parsimony test), UniFrac, distance calculation, visualization tools, a NAST-based aligner, and many other features. Their current goal is to release a new iteration of the project monthly (192). Their article has been cited 673 times as of 16 July 2012.

V. I. Diversity indices accuracy

In our work, we used the Chao1 estimator for bacterial richness estimation and Shannon diversity index for bacterial diversity analyses. With the improvement of sequences reads per sample, diversity analyses have been largely improved by high-throughput sequencing technology comparing with traditional (Sanger) sequencing technology.

Gihring *et al.* (2012) reiterated the need to equalize the number of reads being compared across libraries (67). The Chao 1 estimator and Shannon diversity index are dependent on sample size (31, 196). The application of Chao 1, which is based on the ratio of singleton to doubleton OTUs (31), suggests that in order to avoid bias due to artefactual singletons, we should use high strict sequence quality control parameters (92), and a 3% similarity level clustering (111).

V. J. Microbial communities in oligotrophic environments

Once the bacterial community structure is revealed, there are many interesting research directions to continue to analyse (summary in Fig. 10):

- selected bacterial groups for further study, such as the dominant groups: members of the phyla *Firmicutes* and *Proteobacteria* in deserts, members of the genus *Massilia* in Asian sandstorms, and members of the phylum *Proteobacteria* in DWDS. We can combine *in situ* techniques such as FISH to quantify them and test their live statut (66, 169, 171).
- functional analyses: under such extreme conditions, it will be interesting to study how bacterial members can survive, and what special characters they have to becoming thermostable (128), radiation-resistant (126), alkalitolerant (127) or disinfectant-resistant (154, 213). In order to study their metabolism, we need to isolate the bacteria and grow them in culture if possible (and RNA). Further

genomic DNA sequencing should be performed to find interesting genes alternatively, direct extraction of proteins, if we can obtain large quantities of isolated cells, are needed in order to characterize the proteins.

- the other microbial (Archeae, virus and eukaryotic microbe) members of the communities in oligotrophic environments (21, 29, 40). We can use universal Archaeal primer to amplify the DNA of Archeae, and characterize their taxonomic composition. It is important to note that the Archeae normally detected in extreme environments are difficult for DNA extraction. Therefore, the revelation of the Archaeal population composition is more difficult than for Bacteria (8, 234). Virus populations can be explored by random PCR, but the analysis of high-throughput sequencing is a challenge (53, 114). The use of 18S rDNA PCR sequencing for eukaryotic microbial community analyses is available (25, 84).

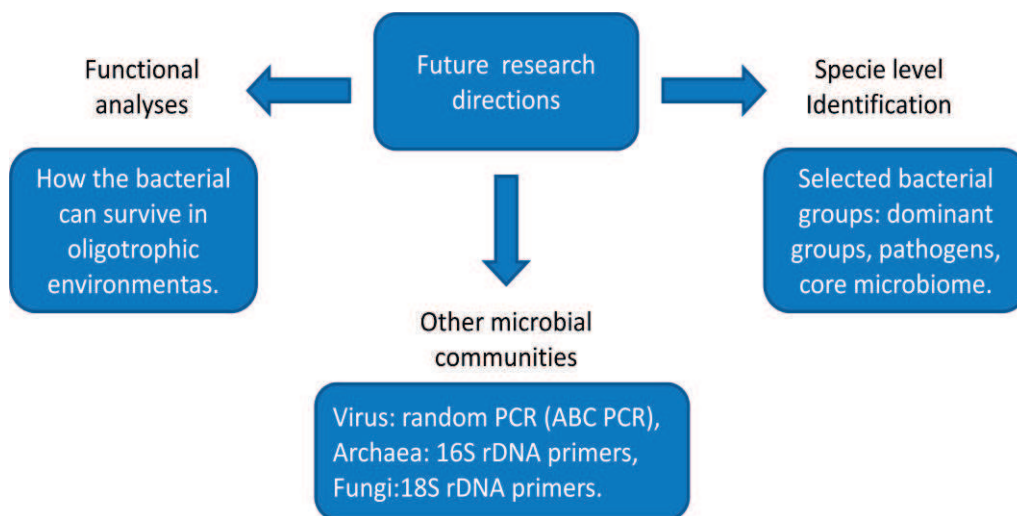


Figure 10: Future research directions after the characterization of bacterial populations in oligotrophic environments.

Reference

1. 1997. Guidelines for drinking-water quality, 2 ed, vol. 3. World health organization, Geneve.
2. **Abed, R. M., S. A. Kharusi, A. Schramm, and M. D. Robinson.** 2010. Bacterial diversity, pigments and nitrogen fixation of biological desert crusts from the Sultanate of Oman. *FEMS Microbiol Ecol* **72**:418-428.
3. **Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz.** 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**:8966-9.
4. **Aguilera, L. E., J. R. Gutiérrez, and P. L. Meserve.** 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.
5. **Ahmadian, A., M. Ehn, and S. Hober.** 2006. Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* **363**:83-94.
6. **Aislabie, J., S. Jordan, J. Ayton, J. L. Klassen, G. M. Barker, and S. Turner.** 2009. Bacterial diversity associated with ornithogenic soil of the Ross Sea region, Antarctica. *Can J Microbiol* **55**:21-36.
7. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**:143-69.
8. **Angel, R., M. I. Soares, E. D. Ungar, and O. Gillor.** 2010. Biogeography of soil archaea and bacteria along a steep precipitation gradient. *ISME J* **4**:553-563.
9. **Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman.** 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**:7724-36.
10. **Ausec, L., M. Zakrzewski, A. Goesmann, A. Schlüter, and I. Mandic-Mulec.** 2011. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PLoS ONE* **6**:e25724.
11. **Bachar, A., A. Al-Ashhab, M. I. Soares, M. Y. Sklarz, R. Angel, E. D. Ungar, and O. Gillor.** 2010. Soil microbial abundance and diversity along a low precipitation gradient. *Microb Ecol* **60**:453-461.
12. **Baldani, J. I., B. Pot, G. Kirchhof, E. Falsen, V. L. Baldani, F. L. Olivares, B. Hoste, K. Kersters, A. Hartmann, M. Gillis, and J. Dobereiner.** 1996. Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] *rubrisubalbicans*, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans comb. nov.*; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *Int J Syst Bacteriol* **46**:802-810.
13. **Barer, M. R.** 1997. Viable but non-culturable and dormant bacteria: time to resolve an oxymoron and a misnomer? *J Med Microbiol* **46**:629-31.
14. **Barry, T., G. Colleran, M. Glennon, L. K. Dunican, and F. Gannon.** 1991. The 16s/23s ribosomal spacer region as a target for DNA probes to identify eubacteria. *PCR Methods Appl* **1**:51-6.
15. **Bartlett, J. M., and D. Stirling.** 2003. A short history of the polymerase chain reaction. *Methods Mol Biol* **226**:3-6.
16. **Beckmann, A., and I. Hense.** 2009. A fresh look at the nutrient cycling in the oligotrophic

- ocean. *Biogeochemistry* **96**:1-11.
17. **Bell, M. L., J. K. Levy, and Z. Lin.** 2008. The effect of sandstorms and air pollution on cause-specific hospital admissions in Taipei, Taiwan. *Occup Environ Med* **65**:104-11.
 18. **Ben-David, E. A., E. Zaady, Y. Sher, and A. Nejidat.** 2011. Assessment of the spatial distribution of soil microbial communities in patchy arid and semi-arid landscapes of the Negev Desert using combined PLFA and DGGE analyses. *FEMS Microbiology Ecology* **76**:492-503.
 19. **Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, E. C. M. Chiara, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, et al.** 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**:53-9.
 20. **Berney, M., M. Vital, I. Hulshoff, H. U. Weilenmann, T. Egli, and F. Hammes.** 2008. Rapid, cultivation-independent assessment of microbial viability in drinking water. *Water Res* **42**:4010-8.
 21. **Bhatnagar, A., and M. Bhatnagar.** 2005. Microbial diversity in desert ecosystems. *Current Science* **89**:91-100.
 22. **Bibby, K., E. Viau, and J. Peccia.** 2010. Pyrosequencing of the 16S rRNA gene to reveal bacterial pathogen diversity in biosolids. *Water Res* **44**:4252-60.
 23. **Binladen, J., M. T. Gilbert, J. P. Bollback, F. Panitz, C. Bendixen, R. Nielsen, and E. Willerslev.** 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* **2**:e197.
 24. **Bishop, J. K., R. E. Davis, and J. T. Sherman.** 2002. Robotic observations of dust storm enhancement of carbon biomass in the North Pacific. *Science* **298**:817-21.
 25. **Borneman, J., and R. J. Hartin.** 2000. PCR primers that amplify fungal rRNA genes from environmental samples. *Appl Environ Microbiol* **66**:4356-4360.
 26. **Boutte, C. C., and S. Crosson.** 2011. The complex logic of stringent response regulation in *Caulobacter crescentus*: starvation signalling in an oligotrophic environment. *Mol Microbiol* **80**:695-714.
 27. **Bowers, R. M., I. B. McCubbin, A. G. Hallar, and N. Fierer.** 2012. Seasonal variability in airborne bacterial communities at a high-elevation site. *Atmospheric Environment* **50**:41-49.

28. **Capone, K. A., S. E. Dowd, G. N. Stamatas, and J. Nikolovski.** 2012. Diversity of the human skin microbiome early in life. *J Invest Dermatol* **131**:2026-32.
29. **Cary, S. C., I. R. McDonald, J. E. Barrett, and D. A. Cowan.** 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat Rev Microbiol* **8**:129-38.
30. **Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland.** 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* **69**:330-9.
31. **Chao, A.** 1984. Non-parametric estimation of the number of classes in a populaion. *Scan J Stat* **11**:265-270.
32. **Chien, L. C., C. H. Yang, and H. L. Yu.** 2012. Estimated Effects of Asian Dust Storms on Spatiotemporal Distributions of Clinic Visits for Respiratory Diseases in Taipei Children (Taiwan). *Environ Health Perspect.*
33. **Choi, D.-S., Y.-K. Park, S.-K. Oh, H.-J. Yoon, J. C. Kim, W.-J. Seo, and S.-H. Cha.** 1997. Distribution of Airborne Microorganisms in Yellow Sands of Korea. *The Journal of Microbiology* **35**:1-9.
34. **Chuvochina, M. S., D. Marie, S. Chevaillier, J. R. Petit, P. Normand, I. A. Alekhina, and S. A. Bulat.** 2011. Community Variability of Bacteria in Alpine Snow (Mont Blanc) Containing Saharan Dust Deposition and Their Snow Colonisation Potential. *Microbes Environ*:Epub 2011 Jun 8.
35. **Clark, T. A., I. A. Murray, R. D. Morgan, A. O. Kislyuk, K. E. Spittle, M. Boitano, A. Fomenkov, R. J. Roberts, and J. Korlach.** 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* **40**:e29.
36. **Cockell, C. S., C. P. McKay, K. Warren-Rhodes, and G. Horneck.** 2008. Ultraviolet radiation-induced limitation to epilithic microbial growth in arid deserts--dosimetric experiments in the hyperarid core of the Atacama Desert. *J Photochem Photobiol B* **90**:79-87.
37. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* **33**:D294-6.
38. **Colwell, R. R.** 1997. Microbial diversity: the importance of exploration and conservation. *J Ind Microbiol Biotechnol* **18**:302-7.
39. **Connon, S. A., E. D. Lester, H. S. Shafaat, D. C. Obenhuber, and A. Ponce.** 2007. Bacterial diversity in hyperarid Atacama Desert soils. *J Geophys Res* **112** G04S17, 9 pp.
40. **Cowan, D. A.** 2009. Cryptic microbial communities in Antarctic deserts. *Proc Natl Acad Sci USA* **106**:19749-19750.
41. **Cowan, D. A., S. B. Pointing, M. I. Stevens, S. C. Cary, F. Stomeo, and I. M. Tuffin.** 2010. Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. *Polar Biology* **34(2)**:307-311.
42. **Cowan, D. A., N. J. Russell, A. Mamais, and D. M. Sheppard.** 2002. Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass. *Extremophiles* **6**:431-6.
43. **Craun, G. F., N. Nwachuku, R. L. Calderon, and M. F. Craun.** 2002. Outbreaks in drinking-water systems, 1991-1998. *J Environ Health* **65**:16-23, 28; quiz 31-2.
44. **Davey, M. E., and A. O'Toole G.** 2000. Microbial biofilms: from ecology to molecular genetics.

- Microbiol Mol Biol Rev **64**:847-67.
45. **Davies, D.** 2003. Understanding biofilm resistance to antibacterial agents. *Nat Rev Drug Discov* **2**:114-22.
 46. **de Groot, A., V. Chapon, P. Servant, R. Christen, M. F. Saux, S. Sommer, and T. Heulin.** 2005. *Deinococcus deserti* sp. nov., a gamma-radiation-tolerant bacterium isolated from the Sahara Desert. *Int J Syst Evol Microbiol* **55**:2441-2446.
 47. **Dethlefsen, L., S. Huse, M. L. Sogin, and D. A. Relman.** 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**:e280.
 48. **Drees, K. P., J. W. Neilson, J. L. Betancourt, J. Quade, D. A. Henderson, B. M. Pryor, and R. M. Maier.** 2006. Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. *Appl Environ Microbiol* **72**:7902-7908.
 49. **Duce, R. A., C. K. Unni, B. J. Ray, J. M. Prospero, and J. T. Merrill.** 1980. Long-range atmospheric transport of soil dust from Asia to the tropical north pacific: temporal variability. *Science* **209**:1522-4.
 50. **Duineveld, B. M., G. A. Kowalchuk, A. Keijzer, J. D. van Elsas, and J. A. van Veen.** 2001. Analysis of bacterial communities in the rhizosphere of chrysanthemum via denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA as well as DNA fragments coding for 16S rRNA. *Appl Environ Microbiol* **67**:172-8.
 51. **Eichler, S., R. Christen, C. Holtje, P. Westphal, J. Botel, I. Brettar, A. Mehling, and M. G. Hofle.** 2006. Composition and dynamics of bacterial communities of a drinking water supply system as assessed by RNA- and DNA-based 16S rRNA gene fingerprinting. *Appl Environ Microbiol* **72**:1858-72.
 52. **Elshahed, M. S., N. H. Youssef, A. M. Spain, C. Sheik, F. Z. Najar, L. O. Sukharnikov, B. A. Roe, J. P. Davis, P. D. Schloss, V. L. Bailey, and L. R. Krumholz.** 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. *Appl Environ Microbiol* **74**:5422-5428.
 53. **Eriksson, N., L. Pachter, Y. Mitsuya, S. Y. Rhee, C. Wang, B. Gharizadeh, M. Ronaghi, R. W. Shafer, and N. Beerenwinkel.** 2008. Viral population estimation using pyrosequencing. *PLoS Comput Biol* **4**:e1000074.
 54. **Fan, J. B., M. S. Chee, and K. L. Gunderson.** 2006. Highly parallel genomic assays. *Nat Rev Genet* **7**:632-44.
 55. **Felfoldi, T., Z. Heeger, M. Vargha, and K. Marialigeti.** 2009. Detection of potentially pathogenic bacteria in the drinking water distribution system of a hospital in Hungary. *Clin Microbiol Infect* **16**:89-92.
 56. **Fierer, N., and R. B. Jackson.** 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**:626-631.
 57. **Fierer, N., and J. T. Lennon.** 2011. The generation and maintenance of diversity in microbial communities. *Am J Bot* **98**:439-48.
 58. **Fisher, M. M., and E. W. Triplett.** 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65**:4630-6.
 59. **Flegg, J.** 1993. *Deserts: a miracle of life*, vol. Facts on File, New York.

60. **Flemming, H. C.** 2002. Biofouling in water systems--cases, causes and countermeasures. *Appl Microbiol Biotechnol* **59**:629-40.
61. **Forno, E., A. B. Onderdonk, J. McCracken, A. A. Litonjua, D. Laskey, M. L. Delaney, A. M. Dubois, D. R. Gold, L. M. Ryan, S. T. Weiss, and J. C. Celedon.** 2008. Diversity of the gut microbiota and eczema in early life. *Clin Mol Allergy* **6**:11.
62. **Friedmann, E. I., L. Kappen, M. A. Meyer, and J. A. Nienow.** 1993. Long-term productivity in the cryptoendolithic microbial community of the Ross Desert, Antarctica. *Microb Ecol* **25**:51-69.
63. **Galand, P. E., E. O. Casamayor, D. L. Kirchman, and C. Lovejoy.** 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* **106**:22427-32.
64. **Gao, D., and Y. Tao.** 2011. Current molecular biologic techniques for characterizing environmental microbial community. *Frontiers of Environmental Science & Engineering in China*:DOI 10.1007/s11783-011-0306-6.
65. **Garcia-Pichel, F., S. L. Johnson, D. Youngkin, and J. Belnap.** 2003. Small-scale vertical distribution of bacterial biomass and diversity in biological soil crusts from arid lands in the Colorado plateau. *Microb Ecol* **46**:312-21.
66. **Garcia-Pichel, F., A. Lopez-Cortes, and U. Nubel.** 2001. Phylogenetic and morphological diversity of cyanobacteria in soil desert crusts from the Colorado plateau. *Appl Environ Microbiol* **67**:1902-10.
67. **Gihring, T. M., S. J. Green, and C. W. Schadt.** 2012. Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environ Microbiol* **14**:285-90.
68. **Gilles, A., E. Meglec, N. Pech, S. Ferreira, T. Malausa, and J. F. Martin.** 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* **12**:245.
69. **Glavin, D. P., H. J. Cleaves, M. Schubert, A. Aubrey, and J. L. Bada.** 2004. New method for estimating bacterial cell abundances in natural samples by use of sublimation. *Appl Environ Microbiol* **70**:5923-8.
70. **Gorbushina, A. A., R. Kort, A. Schulte, D. Lazarus, B. Schnetger, H. J. Brumsack, W. J. Broughton, and J. Favet.** 2007. Life in Darwin's dust: intercontinental transport and survival of microbes in the nineteenth century. *Environ. Microbiol* **9**:2911-22.
71. **Goudie, A. S., and N. J. Middleton.** 2001. Saharan dust storms: nature and consequences. *Earth-Science Reviews* **56**:179-204.
72. **Griffin, D. W.** 2007. Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. *Clin. Microbiol Rev* **20**:459-77,.
73. **Griffin, D. W., N. Kubilay, M. Kocak, M. A. Gray, T. C. Borden, and E. A. Shinn.** 2007. Airborne desert dust and aeromicrobiology over the Turkish Mediterranean coastline. *Atmospheric environment* **41**:4050-4062.
74. **Grousset, F. E., P. Ginoux, A. Bory, and P. E. Biscaye.** 2003. Case study of a Chinese dust plume reaching the French Alps. *Geophys. Res. Lett* **30**:1277.
75. **Guss, A. M., G. Roeselers, I. L. Newton, C. R. Young, V. Klepac-Ceraj, S. Lory, and C. M. Cavanaugh.** 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* **5**:20-29.

76. **Haimovich, A. D.** 2011. Methods, challenges, and promise of next-generation sequencing in cancer biology. *Yale J Biol Med* **84**:439-46.
77. **Hall, J. R., K. R. Mitchell, O. Jackson-Weaver, A. S. Kooser, B. R. Cron, L. J. Crossey, and C. D. Takacs-Vesbach.** 2008. Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. *Appl Environ Microbiol* **74**:4910-4922.
78. **Hamady, M., J. J. Walker, J. K. Harris, N. J. Gold, and R. Knight.** 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* **5**:235-7.
79. **Hammes, F., M. Berney, Y. Wang, M. Vital, O. Kosterb, and T. Eglia.** 2008. Flow-cytometric total bacterial cell counts as a descriptive microbiological parameter for drinking water treatment processes. *Water Res* **42**.
80. **Han, Y.-J., T. M. Holsen, P. K. Hopke, J.-P. Cheong, H. Kim, and S.-M. Yi.** 2004. Identification of source locations for atmospheric dry deposition of heavy metals during yellow-sand events in Seoul, Korea in 1998 using hybrid receptor models. *Atmospheric Environment* **38**:5353-5361.
81. **Hansel, C. M., S. Fendorf, P. M. Jardine, and C. A. Francis.** 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol* **74**:1620-33.
82. **Harmsen, D., S. Dostal, A. Roth, S. Niemann, J. Rothganger, M. Sammeth, J. Albert, M. Frosch, and E. Richter.** 2003. RIDOM: comprehensive and public sequence database for identification of Mycobacterium species. *BMC Infect Dis* **3**:26.
83. **Hashizume, M., K. Ueda, Y. Nishiwaki, T. Michikawa, and D. Onozuka.** 2010. Health effects of Asian dust events: a review of the literature. *Nihon Eiseigaku Zasshi* **65**:413-21.
84. **He, J., Z. Xu, and J. Hughes.** 2005. Analyses of soil fungal communities in adjacent natural forest and hoop pine plantation ecosystems of subtropical Australia using molecular approaches based on 18S rRNA genes. *FEMS Microbiol Lett* **247**:91-100.
85. **Herman, R. P., K. R. Provencio, J. Herrera-Matos, and R. J. Torrez.** 1995. Resource islands predict the distribution of heterotrophic bacteria in chihuahuan desert soils. *Appl Environ Microbiol* **61**:1816-21.
86. **Höfle, M. G., D. L. Kirchman, R. Christen, and I. Brettar.** 2008. Molecular diversity of bacterioplankton: link to a predictive biogeochemistry of pelagic ecosystems *Aquat Microb Ecol* **53**:39-58.
87. **Hollister, E. B., A. S. Engledow, A. J. Hammett, T. L. Provin, H. H. Wilkinson, and T. J. Gentry.** 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *Isme J* **4**:829-38.
88. **Hong, P. Y., C. Hwang, F. Ling, G. L. Andersen, M. W. LeChevallier, and W. T. Liu.** 2010. Pyrosequencing analysis of bacterial biofilm communities in water meters of a drinking water distribution system. *Appl Environ Microbiol* **76**:5631-5.
89. **Horikoshi, K., G. Antranikian, A. T. Bull, F. T. Robb, and K. O. Stetter.** 2011. *Extremophiles handbook*, 1 ed, vol. Springer, Tokyo.
90. **Horowitz, N. H., R. E. Cameron, and J. S. Hubbard.** 1972. Microbiology of the dry valleys of antarctica. *Science* **176**:242-5.

91. **Huse, S. M., L. Dethlefsen, J. A. Huber, D. Mark Welch, D. A. Relman, and M. L. Sogin.** 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* **4**:e1000255.
92. **Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**:R143.
93. **Huybens, N., J. Mainil, and D. Marlier.** 2009. Les techniques de biologie moléculaire d'analyse des populations bactériennes complexes. *Ann. Méd. Vét.* **153**:112-128.
94. **Hwang, C., F. Ling, G. L. Andersen, M. W. Lechevallier, and W. T. Liu.** 2011. Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. *Microbes Environ* **27**:9-18.
95. **Iwasaka, Y., G.-Y. Shi, M. Yamada, F. Kobayashi, M. Kakikawa, T. Maki, T. Naganuma, B. Chen, Y. Tobo, and C. S. Hong.** 2009. Mixture of Kosa (Asian dust) and bioaerosols detected in the atmosphere over the Kosa particles source regions with balloon-borne measurements: possibility of long-range transport. *Air Qual. Atmos. Health* **2**:29-38.
96. **Jefferson, K. K.** 2004. What drives bacteria to produce a biofilm? *FEMS Microbiol Lett* **236**:163-73.
97. **Jeon, E. M., H. J. Kim, K. Jeong, J. H. Kim, M. Y. Kim, Y. P. Kim, and J.-O. Ka.** 2010. Impact of Asian dust events on airborne bacterial community assessed by molecular analyses. *Atmospheric environment* **45**:4313-4321.
98. **Jin, Z. Z., J. Q. Lei, X. W. Xu, S. Y. Li, J. L. Fan, S. F. Zhao, H. W. Zhou, and F. Gu.** 2009. [Relationships of soil microbial biomass with soil environmental factors in Tarim Desert highway shelter-forest]. *Ying Yong Sheng Tai Xue Bao* **20**:51-57.
99. **Johnson, M.** 2003. *The ultimate desert handbook: a manual for desert hikers, campers, and travelers*, vol. Ragged Mountain Press, Camden.
100. **Kahlisch, L., K. Henne, L. Groebe, J. Draheim, M. G. Hofle, and I. Brettar.** 2010. Molecular analysis of the bacterial drinking water community with respect to live/dead status. *Water Sci Technol* **61**:9-14.
101. **Kakizaki, E., Y. Ogura, S. Kozawa, S. Nishida, T. Uchiyama, T. Hayashi, and N. Yukawa.** 2012. Detection of diverse aquatic microbes in blood and organs of drowning victims: First metagenomic approach using high-throughput 454-pyrosequencing. *Forensic Sci Int.*
102. **Keinanen-Toivola, M. M., R. P. Revetta, and J. W. Santo Domingo.** 2006. Identification of active bacterial communities in a model drinking water biofilm system using 16S rRNA-based clone libraries. *FEMS Microbiol Lett* **257**:182-8.
103. **Kell, D. B., A. S. Kaprelyants, D. H. Weichart, C. R. Harwood, and M. R. Barer.** 1998. Viability and activity in readily culturable bacteria: a review and discussion of the practical issues. *Antonie Van Leeuwenhoek* **73**:169-87.
104. **Kellogg, C. A., D. W. Griffin, V. H. Garrison, K. K. Peak, N. Royall, R. R. Smith, and E. A. Shinn.** 2004. Characterization of aerosolized bacteria and fungi from desert dust events in Mali, West Africa. *Aerobiologia* **20**:99-110.
105. **Kerr, C., K. Osborn, G. Robson, and P. Handley.** 1999. The relationship between pipe material and biofilm formation in a laboratory model system. *J Appl Microbiol* **85**:29S-38S.
106. **Koberl, M., H. Muller, E. M. Ramadan, and G. Berg.** 2011. Desert farming benefits from

- microbial potential in arid soils and promotes diversity and plant health. *PLoS One* **6**:e24452.
107. **Koboldt, D. C., D. E. Larson, K. Chen, L. Ding, and R. K. Wilson.** 2012. Massively parallel sequencing approaches for characterization of structural variation. *Methods Mol Biol* **838**:369-84.
108. **Kong, H. H.** 2012. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends Mol Med* **17**:320-8.
109. **Koren, O., A. Spor, J. Felin, F. Fak, J. Stombaugh, V. Tremaroli, C. J. Behre, R. Knight, B. Fagerberg, R. E. Ley, and F. Backhed.** 2010. Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc Natl Acad Sci U S A* **108 Suppl 1**:4592-8.
110. **Kormas, K. A., C. Neofitou, M. Pachiadaki, and E. Koufostathi.** 2009. Changes of the bacterial assemblages throughout an urban drinking water distribution system. *Environ Monit Assess* **165**:27-38.
111. **Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz.** 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**:118-23.
112. **Kwon, H. J., S. H. Cho, Y. Chun, F. Lagarde, and G. Pershagen.** 2002. Effects of the Asian dust events on daily mortality in Seoul, Korea. *Environ Res* **90**:1-5.
113. **Kwon, S., E. Moon, T. S. Kim, S. Hong, and H. D. Park.** 2011. Pyrosequencing demonstrated complex microbial communities in a membrane filtration system for a drinking water treatment plant. *Microbes Environ* **26**:149-55.
114. **Lahser, F. C., J. Wright-Minogue, A. Skelton, and B. A. Malcolm.** 2003. Quantitative estimation of viral fitness using Pyrosequencing. *Biotechniques* **34**:26-8.
115. **Langmark, J., M. V. Storey, N. J. Ashbolt, and T. A. Stenstrom.** 2005. Accumulation and fate of microorganisms and microspheres in biofilms formed in a pilot-scale water distribution system. *Appl Environ Microbiol* **71**:706-12.
116. **Lauber, C. L., M. Hamady, R. Knight, and N. Fierer.** 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.
117. **LeChevallier, M. W., W. Schulz, and R. G. Lee.** 1991. Bacterial nutrients in drinking water. *Appl Environ Microbiol* **57**:857-62.
118. **LeChevallier, M. W., R. J. Seidler, and T. M. Evans.** 1980. Enumeration and characterization of standard plate count bacteria in chlorinated and raw water supplies. *Appl Environ Microbiol* **40**:922-30.
119. **LeChevallier, M. W., N. J. Welch, and D. B. Smith.** 1996. Full-scale studies of factors related to coliform regrowth in drinking water. *Appl Environ Microbiol* **62**:2201-11.
120. **Lee, D. G., and S. J. Kim.** 2003. Bacterial species in biofilm cultivated from the end of the Seoul water distribution system. *J Appl Microbiol* **95**:317-24.
121. **Lee, J. T., J. Y. Son, and Y. S. Cho.** 2007. A comparison of mortality related to urban air particles between periods with Asian dust days and without Asian dust days in Seoul, Korea, 2000-2004. *Environ Res* **105**:409-13.
122. **Lee, S., B. Choi, S. M. Yi, and G. Ko.** 2009. Characterization of microbial community during Asian dust events in Korea. *Sci. Total Environ* **407**:5308-5314.

123. **Lester, E. D., M. Satomi, and A. Ponce.** 2007. Microflora of extreme arid Atacama Desert soils. *Soil Biology & Biochemistry* **39**:704-708.
124. **Lighthart, B., and Y. Tong.** 1998. Measurements of total and culturable bacteria in the alfresco atmosphere using a wet-cyclone sampler. *Aerobiologia* **14**:325-332.
125. **Lima, D. M., R. Q. Oliveira, A. P. Uetanabaro, A. Goes-Neto, C. A. Rosa, and S. A. Assis.** 2009. Thermostable inulinases secreted by yeast and yeast-like strains from the Brazilian semi-arid region. *Int J Food Sci Nutr* **60 (Suppl 7)**:63-71.
126. **Liu, M., J. Dai, Y. Liu, F. Cai, Y. Wang, and C. Fang.** 2011. *Desertibacter roseus gen. nov., sp. nov.*, a gamma radiation-resistant bacterium in the family Rhodospirillaceae, isolated from desert sand. *Int J Syst Evol Microbiol* **61**:1109-1113.
127. **Liu, M., Y. Liu, Y. Wang, X. Luo, J. Dai, and C. Fang.** 2010. *Lysobacter xinjiangensis sp. nov.*, a moderately thermotolerant and alkalitolerant bacterium isolated from gamma-irradiated sand soil sample. *Int J Syst Evol Microbiol* **61**:433-437.
128. **Liu, M., X. Luo, L. Zhang, J. Dai, Y. Wang, Y. Tang, J. Li, T. Sun, and C. Fang.** 2009. *Pseudomonas xinjiangensis sp. nov.*, a moderately thermotolerant bacterium isolated from desert sand. *Int J Syst Evol Microbiol* **59**:1286-9.
129. **Liu, M., F. Peng, Y. Wang, K. Zhang, G. Chen, and C. Fang.** 2009. *Kineococcus xinjiangensis sp. nov.*, isolated from desert sand. *Int J Syst Evol Microbiol* **59**:1090-1093.
130. **Liu, W., H. Wu, Z. Wang, S. L. Ong, J. Y. Hu, and W. J. Ng.** 2002. Investigation of assimilable organic carbon (AOC) and bacterial regrowth in drinking water distribution system. *Water Res* **36**:891-8.
131. **Liu, W. T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**:4516-22.
132. **Liu, Z., T. Z. DeSantis, G. L. Andersen, and R. Knight.** 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* **36**:e120.
133. **Liu, Z., C. Lozupone, M. Hamady, F. D. Bushman, and R. Knight.** 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res* **35**:e120.
134. **Logue, J. B., S. Langenheder, A. F. Andersson, S. Bertilsson, S. Drakare, A. Lanzen, and E. S. Lindstrom.** 2012. Freshwater bacterioplankton richness in oligotrophic lakes depends on nutrient availability rather than on species-area relationships. *Isme J* **6**:1127-36.
135. **Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. J. Pallen.** 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* **30**:434-9.
136. **Lozupone, C. A., and R. Knight.** 2007. Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**:11436-40.
137. **Ludemann, H., I. Arth, and W. Liesack.** 2000. Spatial changes in the bacterial community structure along a vertical oxygen gradient in flooded paddy soil cores. *Appl Environ Microbiol* **66**:754-62.
138. **Luna, R. A., L. R. Fasciano, S. C. Jones, B. L. Boyanton, Jr., T. T. Ton, and J. Versalovic.** 2007. DNA pyrosequencing-based bacterial pathogen identification in a pediatric hospital setting. *J*

- Clin Microbiol **45**:2985-92.
139. **MacLean, D., J. D. Jones, and D. J. Studholme.** 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**:287-96.
 140. **Madigan, M. T., J. M. Martinko, and J. Parker.** 1997. *Brock Biology of Microorganisms*, 8 ed, vol. Prentice Hall, London.
 141. **Maier, R. M., I. L. Pepper, and C. P. Gerba.** 2009. *Environmental microbiology*, 2 ed, vol. Elsevier, Oxford.
 142. **Maki, T., S. Susuki, F. Kobayashi, M. Kakikawa, M. Yamada, T. Higashi, B. Chen, G. Shi, C. Hong, Y. Tobo, H. Hasegawa, K. Ueda, and Y. Iwasaka.** 2008. Phylogenetic diversity and vertical distribution of a halobacterial community in the atmosphere of an Asian dust (KOSA) source region, Dunhuang City. *Air Qual. Atmos. Health* **1**:81-89.
 143. **Mardis, E. R.** 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**:133-41.
 144. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bembien, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-80.
 145. **Martiny, A. C., T. M. Jorgensen, H. J. Albrechtsen, E. Arvin, and S. Molin.** 2003. Long-term succession of structure and diversity of a biofilm formed in a model drinking water distribution system. *Appl Environ Microbiol* **69**:6899-907.
 146. **Matsumoto, G. I., J. A. Nienow, E. I. Friedmann, E. Sekiya, and R. Ocampo-Friedmann.** 2004. Biogeochemical features of lipids in endolithic microbial communities in the Ross Desert (McMurdo Dry Valleys), Antarctica. *Cell Mol Biol (Noisy-le-grand)* **50**:591-604.
 147. **Mayilraj, S., G. S. Prasad, K. Suresh, H. S. Saini, S. Shivaji, and T. Chakrabarti.** 2005. *Planococcus stackebrandtii* sp. nov., isolated from a cold desert of the Himalayas, India. *Int J Syst Evol Microbiol* **55**:91-4.
 148. **McKay, C. P., E. I. Friedmann, B. Gomez-Silva, L. Caceres-Villanueva, D. T. Andersen, and R. Landheim.** 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.
 149. **McKernan, K. J., H. E. Peckham, G. L. Costa, S. F. McLaughlin, Y. Fu, E. F. Tsung, C. R. Clouser, C. Duncan, J. K. Ichikawa, C. C. Lee, Z. Zhang, S. S. Ranade, E. T. Dimalanta, F. C. Hyland, T. D. Sokolsky, L. Zhang, A. Sheridan, H. Fu, C. L. Hendrickson, B. Li, L. Kotler, J. R. Stuart, J. A. Malek, J. M. Manning, A. A. Antipova, D. S. Perez, M. P. Moore, K. C. Hayashibara, M. R. Lyons, R. E. Beaudoin, B. E. Coleman, M. W. Laptewicz, A. E. Sannicandro, M. D. Rhodes, R. K. Gottimukkala, S. Yang, V. Bafna, A. Bashir, A. MacBride, C. Alkan, J. M. Kidd, E. E. Eichler, M. G. Reese, F. M. De La Vega, and A. P. Blanchard.** 2009. Sequence and structural variation

- in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* **19**:1527-41.
150. **Meigs, P.** 1952. Presented at the In Proceedings, Eighth General Assembly and Seventeenth International Congress, , Washington.
 151. **Meigs, P.** 1953. World distribution of arid and semi-arid homo climates. In: Reviews of research on arid zone hydrology. Paris: Arid zone program **1**:203-209.
 152. **Mereau, A., V. Anquetil, M. Cibois, M. Noiret, A. Primot, A. Vallee, and L. Paillard.** 2009. Analysis of splicing patterns by pyrosequencing. *Nucleic Acids Res* **37**:e126.
 153. **Metzker, M. L.** 2010. Sequencing technologies - the next generation. *Nat Rev Genet* **11**:31-46.
 154. **Mishra, A., N. Taneja, and M. Sharma.** 2011. Demonstration of viable but nonculturable *Vibrio cholerae* O1 in fresh water environment of India using ciprofloxacin DFA-DVC method. *Lett Appl Microbiol* **53**:124-6.
 155. **Mori, H., F. Maruyama, and K. Kurokawa.** 2010. VITCOMIC: visualization tool for taxonomic compositions of microbial communities based on 16S rRNA gene sequences. *BMC Bioinformatics* **11**:332.
 156. **Muyzer, G., E. C. de Waal, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**:695-700.
 157. **Nagy, M. L., A. Perez, and F. Garcia-Pichel.** 2005. The prokaryotic diversity of biological soil crusts in the Sonoran Desert (Organ Pipe Cactus National Monument, AZ). *FEMS Microbiol Ecol* **54**:233-45.
 158. **Navarro-Gonzalez, R., F. A. Rainey, P. Molina, D. R. Bagaley, B. J. Hollen, J. de la Rosa, A. M. Small, R. C. Quinn, F. J. Grunthaner, L. Caceres, B. Gomez-Silva, and C. P. McKay.** 2003. Mars-like soils in the Atacama Desert, Chile, and the dry limit of microbial life. *Science* **302**:1018-21.
 159. **Nemergut, D. R., E. K. Costello, M. Hamady, C. Lozupone, L. Jiang, S. K. Schmidt, N. Fierer, A. R. Townsend, C. C. Cleveland, L. Stanish, and R. Knight.** 2010. Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**:135-144.
 160. **Neveu, J., C. Regard, and M. S. DuBow.** 2011. Isolation and Characterization of Two Serine Proteases from Metagenomic Libraries of the Gobi and Death Valley Deserts. *Applied Microbiology and Biotechnology* **91**:635-644.
 161. **Newman, D. K., and J. F. Banfield.** 2002. Geomicrobiology: how molecular-scale interactions underpin biogeochemical systems. *Science* **296**:1071-7.
 162. **Nicolaus, B., M. Kambourova, and E. T. Oner.** 1993. Exopolysaccharides from extremophiles: from fundamentals to biotechnology. *Environ Technol* **31**:1145-58.
 163. **Niquette, P., P. Servais, and R. Savoir.** 2001. Bacterial dynamics in the drinking water distribution system of Brussels. *Water Res* **35**:675-82.
 164. **Nishimura, Y., T. Kenzaka, A. Sueyoshi, P. Li, H. Fujiyama, T. Baba, N. Yamaguchi, and M. Nasu.** 2010. Similarity of bacterial community structure between Asian dust and its sources determined by rRNA gene-targeted approaches. *Microbes Environ* **25**:22-27.
 165. **Nocker, A., and A. K. Camper.** 2006. Selective removal of DNA from dead cells of mixed bacterial communities by use of ethidium monoazide. *Appl Environ Microbiol* **72**:1997-2004.

166. **Norton, C. D., and M. W. LeChevallier.** 2000. A pilot study of bacteriological population changes through potable water treatment and distribution. *Appl Environ Microbiol* **66**:268-76.
167. **Nyberg, L., R. F. Turco, and L. Nies.** 2008. Assessing the impact of nanomaterials on anaerobic microbial communities. *Environ Sci Technol* **42**:1938-43.
168. **Nyren, P.** 2007. The history of pyrosequencing. *Methods Mol Biol* **373**:1-14.
169. **Ohad, I., H. Raanan, N. Keren, D. Tchernov, and A. Kaplan.** 2010. Light-induced changes within photosystem II protects *Microcoleus* sp. in biological desert sand crusts against excess light. *PLoS One* **5**:e11000.
170. **Olsen, R. A., and L. R. Bakken.** 1987. Viability of Soil Bacteria: Optimization of Plate-Counting Technique and Comparison between Total Counts and Plate Counts within Different Size Groups. *Microb. Ecol* **13**:59-74.
171. **Orlando, J., M. Alfaro, L. Bravo, R. Guevara, and M. Carú.** 2010. Bacterial diversity and occurrence of ammonia-oxidizing bacteria in the Atacama Desert soil during a "desert bloom" event. *Soil Biology and Biochemistry* **42**:1183-1188.
172. **Paris, T., S. Skali-Lami, and J. C. Block.** 2007. Effect of wall shear rate on biofilm deposition and grazing in drinking water flow chambers. *Biotechnol Bioeng* **97**:1550-61.
173. **Patel, J. B.** 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Mol Diagn* **6**:313-21.
174. **Pavlov, D., C. M. de Wet, W. O. Grabow, and M. M. Ehlers.** 2004. Potentially pathogenic features of heterotrophic plate count bacteria isolated from treated and untreated drinking water. *Int J Food Microbiol* **92**:275-87.
175. **Pei, A. Y., W. E. Oberdorf, C. W. Nossa, A. Agarwal, P. Chokshi, E. A. Gerz, Z. Jin, P. Lee, L. Yang, M. Poles, S. M. Brown, S. Sotero, T. Desantis, E. Brodie, K. Nelson, and Z. Pei.** 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl Environ Microbiol* **76**:3886-97.
176. **Pepper, I. L., P. Rusin, D. R. Quintanar, C. Haney, K. L. Josephson, and C. P. Gerba.** 2004. Tracking the concentration of heterotrophic plate count bacteria from the source to the consumer's tap. *Int J Food Microbiol* **92**:289-95.
177. **Perkins, S.** 2001. Dust, the Thermostat. *Science News* **160**:200.
178. **Piel, J.** 2012. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu Rev Microbiol* **65**:431-53.
179. **Pointing, S. B., Y. Chan, D. C. Lacap, M. C. Lau, J. A. Jurgens, and R. L. Farrell.** 2009. Highly specialized microbial diversity in hyper-arid polar desert. *Proc Natl Acad Sci USA* **106**:19964-19969.
180. **Poitelon, J. B., M. Joyeux, B. Welte, J. P. Duguet, E. Prestel, O. Lespinet, and M. S. DuBow.** 2009. Assessment of phylogenetic diversity of bacterial microflora in drinking water using serial analysis of ribosomal sequence tags. *Water Res* **43**:4197-206.
181. **Polymenakou, P. N., M. Mandalakis, E. G. Stephanou, and A. Tselepidis.** 2008. Particle size distribution of airborne microorganisms and pathogens during an Intense African dust event in the eastern Mediterranean. *Environ Health Perspect* **116**:292-6.
182. **Rajkumar, M., M. N. Vara Prasad, H. Freitas, and N. Ae.** 2009. Biotechnological applications

- of serpentine soil bacteria for phytoremediation of trace metals. *Crit Rev Biotechnol* **29**:120-30.
183. **Reasoner, D. J., and E. E. Geldreich.** 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Appl. Environ. Microbiol* **49**:1-7.
184. **Revetta, R. P., R. S. Matlib, and J. W. Santo Domingo.** 2011. 16S rRNA gene sequence analysis of drinking water using RNA and DNA extracts as targets for clone library development. *Curr Microbiol* **63**:50-9.
185. **Revetta, R. P., A. Pemberton, R. Lamendella, B. Iker, and J. W. Santo Domingo.** 2010. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. *Water Res* **44**:1353-60.
186. **Rios, A. d. L., S. Valea, C. Ascaso, A. Davila, J. Kastovsky, C. P. McKay, B. Gomez-Silva, and J. Wierchos.** 2010. Comparative analysis of the microbial communities inhabiting halite evaporites of the Atacama Desert. *Int Microbiol* **13**:79-89.
187. **Rodriguez-Ezpeleta, N., M. Hackenberg, and A. M. Aransay.** 2011. *Bioinformatics for High Throughput Sequencing*, vol. Springer.
188. **Salvi, S.** 2007. Health effects of ambient air pollution in children. *Paediatr Respir Rev* **8**:275-80.
189. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**:5463-7.
190. **Schlesinger, P., Y. Mamane, and I. Grishkan.** 2006. Transport of microorganisms to Israel during Saharan dust events *Aerobiologia* **22**:259-273.
191. **Schloss, P. D., and J. Handelsman.** 2004. Status of the microbial census. *Microbiol Mol Biol Rev* **68**:686-91.
192. **Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**:7537-41.
193. **Schut, F., R. A. Prins, and J. C. Gottschal.** 1997. Oligotrophy and pelagic marine bacteria: facts and fiction. *Aquat Microb Ecol* **12**:177-202.
194. **Servais, P., J. Prats, J. Passerat, and T. Garcia-Armisen.** 2009. Abundance of culturable versus viable *Escherichia coli* in freshwater. *Can J Microbiol* **55**:905-9.
195. **Sessitsch, A., A. Weilharter, M. H. Gerzabek, H. Kirchmann, and E. Kandeler.** 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl Environ Microbiol* **67**:4215-24.
196. **Shannon, C. E.** 1997. *The mathematical theory of communication.* 1963. *MD Comput* **14**:306-17.
197. **Shao, Y., and C. Dong.** 2006. A review on East Asian dust storm climate, modelling and monitoring. *Global and Planetary Change* **52**:1-22.
198. **Shi, X. L., D. Marie, L. Jardillier, D. J. Scanlan, and D. Vaultot.** 2009. Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS One* **4**:e7657.

199. **Shinn, E. A., D. W. Griffin, and D. B. Seba.** 2003. Atmospheric transport of mold spores in clouds of desert dust. *Arch. Environ. Health* **58**:498-504.
200. **Sid, P.** 2001. Dust, the thermostat. *Sci. News* **160**:200-201.
201. **Siebel, E., Y. Wang, T. Egli, and F. Hammes.** 2008. Correlations between total cell concentration, total adenosine tri-phosphate concentration and heterotrophic plate counts during microbial monitoring of drinking water. *Drink Water Eng Sci* **1**:1-6.
202. **Springer, B., L. Stockman, K. Teschner, G. D. Roberts, and E. C. Bottger.** 1996. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. *J Clin Microbiol* **34**:296-303.
203. **Stoodley, P., K. Sauer, D. G. Davies, and J. W. Costerton.** 2002. Biofilms as complex differentiated communities. *Annu Rev Microbiol* **56**:187-209.
204. **Strzalka, K., R. Szymanska, and M. Suwalsky.** 2011. Prenylipids and pigments content in selected antarctic lichens and mosses. *Journal of the Chilean Chemical Society* **56**:808-811.
205. **Sundquist, A., S. Bigdeli, R. Jalili, M. L. Druzin, S. Waller, K. M. Pullen, Y. Y. El-Sayed, M. M. Taslimi, S. Batzoglou, and M. Ronaghi.** 2007. Bacterial flora-typing with targeted, chip-based Pyrosequencing. *BMC Microbiol* **7**:108.
206. **Szabó, K. É., P. O. B. Itor, S. Bertilsson, L. Tranvik, and A. Eiler.** 2007. Importance of rare and abundant populations for the structure and functional potential of freshwater bacterial communities. *Aquat Microb Ecol* **47**:1-10.
207. **Szewzyk, U., R. Szewzyk, W. Manz, and K. H. Schleifer.** 2000. Microbiological safety of drinking water. *Annu Rev Microbiol* **54**:81-127.
208. **Tang, Y., J. Dai, L. Zhang, Z. Mo, Y. Wang, Y. Li, S. Ji, C. Fang, and C. Zheng.** 2009. *Dyadobacter alkalitolerans* sp. nov., isolated from desert sand. *Int J Syst Evol Microbiol* **59**:60-4.
209. **Tang, Y. W., N. M. Ellis, M. K. Hopkins, D. H. Smith, D. E. Dodge, and D. H. Persing.** 1998. Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. *J Clin Microbiol* **36**:3674-9.
210. **Teixeira, L. C., R. S. Peixoto, J. C. Cury, W. J. Sul, V. H. Pellizari, J. Tiedje, and A. S. Rosado.** 2010. Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. *ISME J* **4**:989-1001.
211. **Teubner, K.** 2003. Phytoplankton, pelagic community and nutrients in a deep oligotrophic alpine lake: ratios as sensitive indicators of the use of P-resources (DRP:DOP:PP and TN:TP:SRSi). *Water Res* **37**:1583-92.
212. **Therese, K. L., J. Bartell, P. Deepa, S. Mangaiyarkarasi, D. Ward, J. Dajcs, H. N. Madhavan, and D. Stroman.** 2009. DNA sequencing by Microseq kit targeting 16S rRNA gene for species level identification of mycobacteria. *Indian J Med Res* **129**:176-81.
213. **Tokajian, S., and F. Hashwa.** 2004. Incidence of antibiotic resistance in coliforms from drinking water and their identification using the Biolog and the API identification systems. *J Chemother* **16**:45-50.
214. **Tokajian, S. T., F. A. Hashwa, I. C. Hancock, and P. A. Zalloua.** 2005. Phylogenetic assessment of heterotrophic bacteria from a water distribution system using 16S rDNA sequencing. *Can J Microbiol* **51**:325-35.
215. **Tortoli, E.** 2003. Impact of genotypic studies on mycobacterial taxonomy: the new

- mycobacteria of the 1990s. *Clin Microbiol Rev* **16**:319-54.
216. **Tripathy, S., and R. H. Jiang.** 2012. Massively parallel sequencing technology in pathogenic microbes. *Methods Mol Biol* **835**:271-94.
 217. **Ultee, A., N. Souvatzi, K. Maniadi, and H. Konig.** 2004. Identification of the culturable and nonculturable bacterial population in ground water of a municipal water supply in Germany. *J Appl Microbiol* **96**:560-8.
 218. **Uroz, S., M. Buée, C. Murat, P. Frey-Klett, and F. Martin.** 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental Microbiology Reports* **2**:281-288.
 219. **van der Wielen, P. W., S. Voost, and D. van der Kooij.** 2009. Ammonia-oxidizing bacteria and archaea in groundwater treatment and drinking water distribution systems. *Appl Environ Microbiol* **75**:4687-95.
 220. **van Schaik, W., J. Top, D. R. Riley, J. Boekhorst, J. E. Vrijenhoek, C. M. Schapendonk, A. P. Hendrickx, I. J. Nijman, M. J. Bonten, H. Tettelin, and R. J. Willems.** 2010. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. *BMC Genomics* **11**:239.
 221. **Venkateswarlu, B., and A. V. Rao.** 1981. Distribution of microorganisms in stabilized and unstabilized sand dunes of the Indian desert. *Journal of Arid Environments* **4**:203-208.
 222. **Vogel, U., R. Szczepanowski, H. Claus, S. Junemann, K. Prior, and D. Harmsen.** 2012. Ion torrent personal genome machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information. *J Clin Microbiol* **50**:1889-94.
 223. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol* **73**:5261-7.
 224. **Wang, Y., J. Dai, L. Zhang, X. Luo, Y. Li, G. Chen, Y. Tang, Y. Meng, and C. Fang.** 2009. *Lysobacter ximonensis* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* **59**:786-789.
 225. **Warhol, T.** 2007. Desert, vol. Marshall Cavendish Benchmark, New York.
 226. **Weon, H. Y., B. Y. Kim, S. B. Hong, Y. A. Jeon, B. S. Koo, S. W. Kwon, and E. Stackebrandt.** 2009. *Massilia niabensis* sp. nov. and *Massilia niastensis* sp. nov., isolated from air samples. *Int J Syst Evol Microbiol* **59**:1656-60.
 227. **Whang, K., and T. Hattori.** 1988. Oligotrophic bacteria from rendzina forest soil. *Antonie Van Leeuwenhoek* **54**:19-36.
 228. **Whitman, W. B., D. C. Coleman, and W. J. Wiebe.** 1998. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**:6578-83.
 229. **WHO.** 1997. Division of Emerging and Communicable Diseases Surveillance and Control Annual Report—1996. .
 230. **Wilhartitz, I., R. L. Mach, E. Teira, T. Reinthaler, G. J. Herndl, and A. H. Farnleitner.** 2007. Prokaryotic community analysis with CARD-FISH in comparison with FISH in ultra-oligotrophic ground- and drinking water. *J Appl Microbiol* **103**:871-81.
 231. **Williams, M. M., J. W. Domingo, M. C. Meckes, C. A. Kelty, and H. S. Rochon.** 2004. Phylogenetic diversity of drinking water bacteria in a distribution system simulator. *J Appl*

- Microbiol **96**:954-64.
232. **Williams, R., S. G. Peisajovich, O. J. Miller, S. Magdassi, D. S. Tawfik, and A. D. Griffiths.** 2006. Amplification of complex gene libraries by emulsion PCR. *Nat Methods* **3**:545-50.
233. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**:5088-90.
234. **Womack, A. M., B. J. Bohannon, and J. L. Green.** 2010. Biodiversity and biogeography of the atmosphere. *Philos Trans R Soc Lond B Biol Sci* **365**:3645-53.
235. **Wu, J., E. Boyle, W. Sunda, and L. S. Wen.** 2001. Soluble and colloidal iron in the oligotrophic North Atlantic and North Pacific. *Science* **293**:847-9.
236. **Yang, C. Y.** 2006. Effects of Asian dust storm events on daily clinical visits for conjunctivitis in Taipei, Taiwan. *J Toxicol Environ Health A* **69**:1673-80.
237. **Yang, C. Y., Y. S. Chen, H. F. Chiu, and W. B. Goggins.** 2005. Effects of Asian dust storm events on daily stroke admissions in Taipei, Taiwan. *Environ Res* **99**:79-84.
238. **Yang, C. Y., M. H. Cheng, and C. C. Chen.** 2009. Effects of Asian dust storm events on hospital admissions for congestive heart failure in Taipei, Taiwan. *J Toxicol Environ Health A* **72**:324-8.
239. **Yang, S. Y., H. Liu, R. Liu, K. Y. Zhang, and R. Lai.** 2009. *Saccharibacillus kuerlensis* sp. nov., isolated from a desert soil. *Int J Syst Evol Microbiol* **59**:953-957.
240. **Youssef, N., C. S. Sheik, L. R. Krumholz, F. Z. Najjar, B. A. Roe, and M. S. Elshahed.** 2009. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Appl Environ Microbiol* **75**:5227-36.
241. **Zhang, L., G. L. Wu, Y. Wang, J. Dai, and C. X. Fang.** 2010. *Bacillus deserti* sp. nov., a novel bacterium isolated from the desert of Xinjiang, China. *Antonie Van Leeuwenhoek* **99**:221-229.
242. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. *Appl Environ Microbiol* **62**:316-322.
243. **Zhou, J., B. Xia, D. S. Treves, L. Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje.** 2002. Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* **68**:326-34.
244. **Zhu, Z., and T. Wang.** 1993. Trends of desertification and its rehabilitation in China. *Desertification Control Bulletin* **22**:27-29.

Abstract

Oligotrophic ecosystems can be loosely defined as environments that exhibit low ambient nutrient levels. During my thesis, I used 454 DNA pyrosequencing of partial 16S rDNA to explore the bacterial diversity in three different oligotrophic environments, including **A.** surface desert soil, **B.** Asian sandstorm dust and **C.** a section of the city of Paris's drinking water distribution system.

A. Arid regions represent nearly 30% of the Earth's terrestrial surface. The living conditions at the surface of deserts are a challenge for microorganisms, as there is little available water and/or carbon, a very large range of temperatures and high exposure to UV irradiation from the Sun. In surface sand samples from two large Asian deserts, unexpectedly large bacterial diversity residing was revealed. Sequences belonging to the *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* phyla were the most abundant. An increase in phylotype numbers with increasing C/N ratio was noted, suggesting a possible role in the bacterial richness of these desert sand environments.

B. Desert sandstorms are a meteorological phenomenon which have been postulated affect the Earth's climate and public health. We examined the particle-associated (dust and sand-associated) bacterial populations of atmospheric sand in the absence (as control) and presence of sandstorms in five Asian cities. Greater than 90% of the sequences can be classified as representing bacteria belonging to four phyla: *Proteobacteria*, *Bacteroidetes*, *Actinobacteria* and *Firmicutes*. Principal component analyses showed that the sandstorm-associated bacterial populations were clustered by sampling year, rather than location. Members belonging to nine bacterial genera (*Massilia*, *Planococcus*, *Carnobacterium*, *Planomicrobium*, *Pontibacter*, *Pedobacter*, *Lysobacter*, *Sanguibacter*, *Ohtaekwangia*) were observed to increase in sand-associated samples from sandstorms, versus the controls.

C. We characterized the bacterial communities in three water and three biofilm samples from one part of the Parisian drinking water distribution system. A dramatic change in bacterial population in the water during flow through the distribution system from the water treatment plant to the exit from the reservoir was found. The richness of the bacterial population was reduced from the water treatment plant to the reservoir (from 336 to 165 OTUs for water samples leaving the reservoir and from 947 to 275 for biofilm samples in the network). Several OTUs belonging to pathogenic genera were detected in our samples, mostly in the biofilm samples, thus suggesting that the biofilms may be an important source of bacteria during water distribution to the consumers.

Key words: Oligotrophic environment, Bacterial diversity, Asian desert, Asian sandstorm, Drinking water distribution system, Pyrosequencing.

Résumé

Les milieux oligotrophes sont pauvres en éléments nutritifs. En utilisant la technologie de séquençage à haut débit, on a étudié la diversité bactérienne dans trois environnements oligotrophes différents, y compris **A.** sables du désert, **B.** sables dans les tempêtes de l'Asie et **C.** l'eau et biofilms dans les réseaux de distribution d'eau potable.

A. Le désert représente 30% de la surface de la terre. Les conditions de vie dans ces environnements sont un réel défi pour les micro-organismes à cause de nombreux facteurs limitants : peu d'eau et/ou de carbone disponible, une variation importante de température et une forte exposition aux irradiations UV. Le but de cette recherche est donc d'étudier la diversité bactérienne à la surface du sable du désert Taklemaken et du désert de Gobi en utilisant la technologie de séquençage à haut débit. Nos résultats ont révélé une grande diversité bactérienne dans le sol du désert comparable à d'autres types de sols. En outre, nous avons observé une corrélation positive entre la richesse bactérienne et le rapport C/N du sol.

B. Les tempêtes de sable d'Asie se produisent presque toujours au printemps, elles sont générées dans les régions arides d'Asie telles que le désert Taklamaken et le désert de Gobi. L'arrivée des tempêtes de sable pourrait largement modifier l'environnement de l'air dans ces régions sous l'effet du vent, surtout dans les villes asiatiques qui sont le plus souvent touchées. Nos travaux visent à étudier la modification de la composition et la diversité des bactéries associées aux particules au moment de tempête de sable en Asie par la technologie de séquençage à haut débit. Nos résultats ont démontré que les compositions des bactéries associées aux particules sont modifiées pendant les tempêtes, en particulier, la proportion des *Proteobacteria* qui augmentent les jours de tempête. Nous avons signalé neuf genres bactériens détectés en plus pendant les jours de tempêtes, cela nécessite des études plus approfondies.

C. Après avoir analysé la population bactérienne dans les tempêtes de sable, et celles des déserts, nous poursuivons notre objectif de recherche à un environnement aquatique. Nous avons suivi le flux d'eau provenant de l'usine d'Orly (DW-A) à l'entrée du réservoir (DW-B), et à la sortie du réservoir (DW-C). Nous avons constaté une forte variation de la communauté bactérienne, dans DW-A et DW-B, les bactéries prédominantes appartiennent aux populations des *Betaproteobacteria*, puis nous avons observé une conversion vers la population de *Alphaproteobacteria* dans DW-C. Le DW-C a montré une forte similitude avec un échantillon de biofilm (BF-C), ce qui suggère l'effet important du biofilm sur la modification des communautés bactériennes dans l'eau lors de la distribution.

Mots clés: Milieux oligotrophes, la diversité bactérienne, déserts d'Asie, Tempêtes de sable, Réseaux de distribution d'eau potable, Pyroséquençage.

