Airborne Bacterial Communities of PM$_{2.5}$ in Beijing-Tianjin-Hebei Megalopolis, China as Revealed by Illumina MiSeq Sequencing: A Case Study

Jing-Feng Gao*, Xiao-Yan Fan, Hong-Yu Li, Kai-Ling Pan

College of Environmental and Energy Engineering, Beijing University of Technology, Beijing 100124, China

ABSTRACT

Bacteria are ubiquitous and abundant in the atmosphere and some of them are potential pathogens known to cause diseases or allergies in humans. However, the quantities and compositions of total airborne bacterial community and their relationships with environmental factors remain poorly investigated. Here, a case study of the total airborne bacteria of PM$_{2.5}$ collected at six cities in Beijing-Tianjin-Hebei (BTH) megalopolis, China were profiled using quantitative polymerase chain reaction (qPCR) and Illumina MiSeq (PE300) sequencing. qPCR results showed the high abundance of total airborne bacteria of PM$_{2.5}$ in BTH, ranging from $4.82 \times 10^4 \pm 1.58 \times 10^3$ to $2.64 \times 10^5 \pm 9.63 \times 10^4$ cell m$^{-3}$ air, and averaged $1.19 \times 10^5$ cell m$^{-3}$ air. The six PM$_{2.5}$ samples were classified into three groups. Proteobacteria, Cyanobacteria, Actinobacteria and Firmicutes were the four dominant phyla of PM$_{2.5}$. 18 common potential pathogens with extremely low percentage (3.61%) were observed, which were dominated by Enterococcus faecium and Escherichia coli. Plants and soil are probably the main sources of bacteria in PM$_{2.5}$, as suggested by the high percentages of Chloroplast, plant-associated bacteria (e.g., Rhizobiales and Sphingomonadales) and soil-inhabiting bacteria (e.g., Burkholderiales and Pseudomonadales). Variation partitioning analysis (VPA) indicated that the atmospheric pollutants explained the most of the variation (31.90%) in community structure of PM$_{2.5}$, followed by meteorological conditions (15.73%) and the chemical compositions of PM$_{2.5}$ (11.32%). The case study furthers our understanding of the diversity and composition of airborne bacterial communities of PM$_{2.5}$ in BTH, and also identified the main factors shaping the bacterial communities.

Keywords: PM$_{2.5}$; Illumina MiSeq sequencing; Airborne bacterial communities; Potential pathogens; Beijing-Tianjin-Hebei megalopolis.

INTRODUCTION

Atmospheric fine particulate matter, PM$_{2.5}$, has been considered as the major air pollutant in urban cities (Cao et al., 2014), especially in Beijing-Tianjin-Hebei (BTH) megalopolis, one of the rapidly developing regions of China. PM$_{2.5}$ is likely to penetrate and deposit in tracheobronchial and alveolar regions due to the extremely small aerodynamic diameters ($\leq 2.5$ µm). The morbidity and mortality of people will increase when expose to high concentrations of PM$_{2.5}$ (Pope et al., 2009). The majority of studies published extensively focus on the readily measured physical and chemical properties of PM$_{2.5}$, e.g., organic carbon (OC), element carbon (EC) (Cao et al., 2007) and water-soluble inorganic ions (Li et al., 2013). However, comparatively little is known regarding the biological fraction of PM$_{2.5}$. The biological fraction, especially bacteria, is ubiquitous and abundant in atmosphere (Bowers et al., 2010), which has important potential impacts on human health. For example, bacterial pathogens can cause respiratory diseases and tissue inflammation (Cao et al., 2014). It has been suggested that the relative abundances of several respiratory microbial allergens and pathogens increased with the increasing concentrations of particulate matter in urban area (Cao et al., 2014). Meanwhile, plant, livestock and human pathogens are also dispersed through the atmosphere (Pillai and Ricke, 2002). Moreover, the large number of humans in public places in urban areas, such as Residents-Commercial-Transportation Mixed Districts (RCTMD), maybe the key locales for natural pathogens transmission (Robertson et al., 2013). Therefore, it is important and necessary to investigate the diversity and composition of total airborne bacteria and the potential pathogens of PM$_{2.5}$ in urban areas.

Although culture dependent methods were considered to be the standard techniques to investigate the airborne bacterial communities for a long time, the culture dependent methods might significantly underestimate the bacterial diversity for only a low proportion of the viable microorganisms could be cultured (Pace, 1997). In recent years, the culture
independent methods, especially the high-throughput sequencing technology has been widely applied to investigate the diversity of total airborne bacterial communities, including cultivable microorganisms, viable but non-cultivable microorganisms and dead microorganisms. Pyrosequencing has been applied to investigate the bacterial communities of air samples collected by using an air filtration collection system from a high-elevation site (3200 m) and three dominant land-use types (agricultural fields, suburban areas and forests) in Colorado, USA (Bowers et al., 2009; Bowers et al., 2010). Pyrosequencing has also been used to analyze the diversity of bioaerosols collected by fluid impingement (> 90% of the particles captured were > 3 µm) in a heavily trafficked metropolitan subway system in New York City, USA (Robertson et al., 2013). The Illumina Genome Analyzer Ix sequencing (2 × 150 bp) has been applied to investigate the bacterial populations of total suspended particulate matter (TSP) collected in an urban area of Milan (Italy) (Bertolini et al., 2013). Overall, the previous studies mentioned above mainly focus on the bacterial communities in the coarse particulate matters (TSP or > 3 µm), little is known about the total bacterial communities in PM2.5. Although the airborne bacterial communities of PM2.5 have been investigated at four cities in the Midwestern of USA based on the pyrosequencing (Bowers et al., 2011) and at one city in China (Beijing province) during a severe smog event based on the metagenomic sequencing (Cao et al., 2014), there is still a deep gap in the knowledge of total airborne bacterial communities in PM2.5. Compared with pyrosequencing and Illumina Genome Analyzer Ix sequencing (2 × 150 bp), Illumina MiSeq (PE300) sequencing accomplishes a more sensitive and higher accuracy assessment of the microbial community dynamics. Therefore, in this study, Illumina MiSeq (PE300) sequencing was applied to investigate the total airborne bacterial communities in PM2.5. In addition, quantification of airborne bacteria is essential for the risk assessment. The abundance of total airborne bacteria was widely investigated by quantitative polymerase chain reaction (qPCR), a culture-independent molecular method, based on the 16S rRNA gene (Lee et al., 2010; Bertolini et al., 2013). Although the obtained copies of 16S rRNA gene are not directly related to cell number due to the presence of multiple ribosomal operons in the bacterial genomes, the qPCR based on the 16S rRNA gene also could be used to investigate the shifts of total bacterial abundances across samples (Bertolini et al., 2013).

This study is a case study of the characteristics of total airborne bacteria of PM2.5 in urban areas. In this study, PM2.5 samples were collected at six cities in BTH. First, the abundance and composition of total bacterial communities of PM2.5 were investigated by qPCR and Illumina MiSeq (PE300) sequencing. Clustering Analysis (CA) and Principal Coordinates Analysis (PCoA) were applied to assess the similarity of the six samples. Second, the potential sources of the bacteria in PM2.5 were speculated from the general knowledge of the ecology of the dominant bacterial taxa at order, family and genus levels. Third, the diversity and distribution of potential pathogens in PM2.5 in BTH were investigated at the species level due to microbial species within the same family or genus may by significantly different in pathogenic potential (Sjoling et al., 2007). Environmental variables have important effects on the structure of the total airborne bacterial community (Bertolini et al., 2013). Finally, the relationship between total airborne bacterial communities and the environmental factors was further discussed based on the Spearman’s rank correlation coefficients (SRCC), Principal Components Analysis (PCA), Redundancy Analysis (RDA) and Variation Partitioning Analysis (VPA).

METHODS

Sample Collection and Meteorological Conditions

PM2.5 sampling was carried out from 21 May 2014 to 1 June 2014 at six cities in BTH (Fig. 1(a)). There are two megacity sites (Beijing and Tianjin, BJ and TJ), two industrial urban cites (Tangshan and Baoding, TS and BD), a suburban site (Langfang, LF) and a coastal site (Beidaile, BDH). Detailed information of sampling sites, sampling time and corresponding meteorological conditions was presented in Table S1 (in supplementary material).

A model KC-6120 comprehensive atmospheric sampler (Laoshan Electronic Instrument Factory, Qingdao, China) with glass fiber filters being used as the sampling media was applied for the collection of PM2.5 at an average flow rate of 100 L min−1 for 24 h. The glass fiber filters were pre-heated at 450°C for 4 h to remove organic material in a muffle furnace and their weights were measured by a microbalance before PM2.5 collection. After PM2.5 collection, the filters were wrapped with sterile aluminum foil and placed in a freezer (−20°C). For the case study, we sampled one site for 24 h and rotated the collection order first from northwest to southeast axis of BTH (namely BJ-LF-TJ) due to the fact that the wind direction of BTH is predominantly from southeast during summer and then from northeast to southwest axis of BTH (namely BDH-TS-BD). The longer sampling time and application of Illumina MiSeq PE300 sequencing could minimize the effect of single sampling at each site on our investigation of bacterial abundance and community composition because the atmosphere is moving.

The meteorological data including temperature (T), relative humidity (RH) and atmospheric pressure (AP) were recorded concurrently (Table S1 in supplementary material). The atmospheric pollutants (i.e., NO2, SO2, O3 and CO) and the air pollution index (AQI) were also recorded according to the China National Environmental Monitoring Center. (http://113.108.142.147:20035/emcpublish/).

Chemical Analyses of PM2.5

The OC, EC and water-soluble ions of PM2.5 were measured. A thermal/optical carbon aerosol analyzer (DRI Model 2001A, Desert Research Institute, USA) was used to analyze OC and EC according to IMPROVE_A with slight modification (Chow et al., 2007). The water-soluble ions were analyzed by an ion chromatography (IC) system (ICS-90, Dionex, USA) according to the procedures of Zhao and Gao (2008).
**DNA Extraction**

1/4 of the whole glass fiber was cut into pieces using the sterilized scissors and tweezers, which was taken to extract genomic DNA using a Fast-DNA® SPIN Kit following the manufacturer’s protocol (Qiagen, CA, USA). The concentration and quality of DNA extracted were measured by Nanodrop Spectrophotometer ND-1000 (Thermo Fisher Scientific, USA). Then, the DNA was stored at –20°C for subsequent analysis.

**qPCR of 16S rRNA Gene**

qPCR was performed to determine the abundance of total bacterial 16S rRNA gene in the investigated PM$_{2.5}$ samples using primer set Uni1055F and 1392R (Ferris et al., 1996) on a MX3005P Thermocycler (Agilent Technologies, USA). The 20 µL of qPCR mixture contained aseptic water (7.2 µL), each primer (0.4 µL), GoTaq® qPCR Master Mix (Promega, USA) (10 µL) and template DNA (2 µL). The qPCR conditions were as follows: 95°C for 10 min, 40 cycles of 30 s at 95°C, 60 s at 53°C and 60 s at 72°C. The standard curve was generated by using serial dilutions of a plasmid DNA from one representative positive clone containing bacterial 16S rRNA gene. Linear plot was obtained with correlation coefficient of $R^2$ of 0.996. All the qPCR assays (standards, samples and no template control) were carried out in triplicates. The amplification efficiency of the standard curve was 90.6%.

**MiSeq PE300 Sequencing and Sequence Analysis**

The hypervariable V3-V4 regions of 16S rRNA gene were amplified with designed barcoded universal primers 338F (Ben and Ampe, 2000) and 806R (McBain et al., 2003) in triplicates in an Applied Biosystems 9700 PCR System (ABI GeneAmp® 9700). The three PCR products were pooled and visualized on a 2% agarose gel. The correct PCR products were purified and quantified with AxyPrep DNA Gel Extraction Kit (Axygen Biosciences) and QuantiFluor™-ST (Promega), respectively. Then the equimolar amounts of amplicons from different samples were pooled to achieve equal mass concentrations in the final mixture. Subsequently, the mixture was used for the MiSeq PE library construction. Finally, the libraries were sequenced on Illumina MiSeq PE300 sequencer (Illumina, USA) in Majorbio Co., Ltd (Shanghai, China). All the raw data obtained were archived at NCBI Sequence Read Archive (SRA) under accession number of SRP057007.

Quantitative Insights into Microbial Ecology (QIIME) pipeline, an open-source software package, was applied for post-sequencing analysis (Caporaso et al., 2010). Details of sequence analysis were provided in Text 1 in supplementary material.

**Analysis of Potential Pathogens Sequences**

The potential pathogens sequences were also screened (at species level) to investigate their distribution and diversity in PM$_{2.5}$ in BTH. In previous studies, a reference human pathogenic bacteria list was provided, including genus names, species names and the diseases caused (Ye and Zhang, 2011; Cao et al., 2014). The list might not be complete, but it covered a broad range of human pathogenic bacteria. In this study, the list was used to find the potential pathogens sequences from total sequences. First, the operational taxonomic units (OTUs) belonged to the genera mentioned in the reference list were selected. Second, the representative sequences of the selected OTUs were compared against database sequences using the NCBI BLAST to obtain the nearest phylogenetic relatives and determine the species of the OTUs. Third, if the species names obtained were included in the reference list, the sequence numbers of the OTUs for each sample were used for the subsequent analysis of potential pathogens in PM$_{2.5}$.

**Statistical Analyses**

The similarity analysis of bacterial communities in BTH was carried out by CA and PCoA based on the weighted UniFrac distance matrix. The SRCC, PCA and RDA were applied to investigate the correlations between environmental variables (meteorological data, atmospheric pollutants and chemical compositions of PM$_{2.5}$) and airborne bacterial abundances and communities. The significance of each variable on the microbial community was evaluated using the Monte Carlo permutation test. VPA was performed to further determine the relative importance of the environmental variables on the bacterial communities. All the statistical analyses were performed using the functions in the vegan package in R software 2.15.

**RESULTS AND DISCUSSION**

**Chemical Composition and the Abundance of Total Airborne Bacteria of PM$_{2.5}$**

The concentrations of PM$_{2.5}$ ranged from 35.42 µg m$^{-3}$ (BDH) to 194.44 µg m$^{-3}$ (TS), which were higher in TS, BD and TJ than LF, BJ and BDH (Fig. 1(a)). The concentrations of atmospheric pollutants and the AQI were also higher in TS and BD than the other four sampling sites (Fig. S1(a) and Table S1, in supplementary material). The results suggested that air pollution in TS and BD was greater than the other four cities. The concentrations of chemical compositions of PM$_{2.5}$ were depicted in Figs. S1(b) and S1(c) (in supplementary material). The results suggested that air pollution in TS and BD was greater than the other four cities. The concentrations of chemical compositions of PM$_{2.5}$ were depicted in Figs. S1(b) and S1(c) (in supplementary material). The concentrations of OC ranged from 4.07 µg m$^{-3}$ (BJ) to 14.47 µg m$^{-3}$ (TS) with an average of 7.65 µg m$^{-3}$, and which of EC were in the range of 2.97 µg m$^{-3}$ (BJ) to 8.64 µg m$^{-3}$ (BD) with an average of 5.50 µg m$^{-3}$. Similar as OC and EC, the concentrations of water-soluble inorganic ions of PM$_{2.5}$ were also higher in TS and BD. The secondary ions (SO$_4^{2-}$ and NO$_3^-$) and two cations (Ca$^{2+}$ and Mg$^{2+}$) were among the most abundant, and the relative abundance of which in the water-soluble inorganic ions of PM$_{2.5}$ was in the range of 89.63% (TJ) to 98.59% (BD) with an average of 95.51%. The average T, RH and AP during the sampling period were (24.92 ± 4.56)°C, (51.90 ± 13.83)% and (1005.63 ± 2.45) Hpa, respectively (Table S1 in supplementary material).

qPCR was applied to determine the abundance of total airborne bacteria based on the 16S rRNA gene, the result of which was depicted in Fig. 1(b). The 16S rRNA gene abundance was high and relatively constant, which occurred
in the range of $1.93 \times 10^5 \pm 6.31 \times 10^3$ (BD) to $1.06 \times 10^6 \pm 3.85 \times 10^5$ (BJ) copies m$^{-3}$ air and averaged $4.76 \times 10^5$ copies m$^{-3}$ air. The correlations between the environmental factors and the total airborne bacteria 16S rRNA gene abundance were explored via calculation of SRCC. As shown in Table S2 (in supplementary material), the abundance of the airborne bacteria was only negatively correlated to the concentration of EC (SRCC = –0.886, $p = 0.019$).

For comparisons with previous studies, we estimated the number of bacteria in air samples by dividing the number of gene copies by four (Lee et al., 2010). Therefore, the numbers of bacteria cells ranged from $4.82 \times 10^5 \pm 1.58 \times 10^3$ (BD) to $2.64 \times 10^5 \pm 9.63 \times 10^2$ (BJ) cell m$^{-3}$ air, and averaged $1.19 \times 10^5$ cell m$^{-3}$ air. The abundance of total bacteria in PM$_{2.5}$ collected at four urban cities in USA was quantified via flow cytometry, which was $10^5$–$10^6$ cell m$^{-3}$ air (Bowers et al., 2011) and was similar to the bacterial abundance in this study. The bacterial abundance of PM$_{2.5}$ in this study was higher than that of TSP in Seoul measured by qPCR ($5.19 \times 10^1$ to $4.31 \times 10^3$ cell m$^{-3}$ air) (Lee et al., 2010). The bacterial abundance of PM$_{2.5}$ was confirmed to be high in summer (Bowers et al., 2011), which was in accordance with this study. The high bacterial abundance in summer might be related to the increasing release of bacteria from plants (Bertolini et al., 2013). In addition, the abundance of bioaerosols (> 3 µm) collected in a metropolitan subway system in New York city for a 1.5-year period was constant, ranging from $1 \times 10^4$ to $4 \times 10^5$ cell m$^{-3}$ air (Bertolini et al., 2013). The total bacterial abundances were also relatively constant ($9.5 \times 10^5$ to $6.6 \times 10^6$ cell m$^{-3}$ air) at a high-elevation site for 2-week collection period (Bowers et al., 2009). The abundance of PM$_{2.5}$ collected at four
urban cities in USA was in the range of $10^5$–$10^6$ cell m$^{-3}$ air (Bowers et al., 2011). The results suggested that the bacterial abundance was constrained in a range across the samples despite the changing sampling sites, sampling time and meteorological conditions, which was consistent with this study ($10^4$–$10^5$ cell m$^{-3}$ air).

**Richness and Diversity Analysis**

In total, 1054 OTUs (647 ± 85 OTUs sample$^{-1}$) were observed for 193969 sequences in BTH. The airborne bacterial diversity levels observed in BTH were higher than those reported in previous studies of coarse particulate matters (e.g., TSP or > 3 µm) collected from a high-elevation site and a metropolitan subway system (Bowers et al., 2009; Robertson et al., 2013). However, the diversity levels in BTH were lower than those reported in the urban aerosols (TSP) collected in both San Antonio and Austin (Brodie et al., 2007).

The bacterial diversity of LF was highest based on the OTU numbers and diversity indexes including ACE, Chao1 and Shannon (Table S3 in supplementary material), which was further indicated by the rarefaction curves (Fig. S2 in supplementary material). The TJ, TS, BD and BJ samples had the moderate richness based on the Shannon index. Moreover, the Shannon indexes in LF, TJ, TS and BD were close, ranging from 4.37 to 4.97 and the abundances of PM$_{2.5}$ in these four cities were also in the same order of magnitude. The Shannon indexes in BJ and BDH were much lower than the other four cities, however, the abundances of which were one order of magnitude higher than them. The results suggested that higher diversity does not mean higher bacterial abundance, and the relationship between them remains unclear. The rarefaction curves reached the saturation plateau and the Good’s coverage values of the six libraries were above 99%, indicating that the sequences obtained by Miseq PE300 were enough to cover the majority of the bacterial community.

The sequences obtained were classified from phylum to genus. As seen in Table S4 (in supplementary material), in total, 29 phyla, 57 classes, 134 orders, 259 families, and 526 genera were retrieved. For each sample, the numbers of taxa from phylum to genus were in the range of 15–21, 33–45, 81–104, 159–215 and 308–423, respectively. There were still a portion of sequences could not be assigned to any known group, indicating the extent of novel sequences captured by this study. The proportions of the unclassified taxa were 0.07%–0.21%, 0.36%–1.04% and 2.22%–13.15% from order to genus, respectively (Table S5, in supplementary material).

**Bacterial Community Composition of PM$_{2.5}$ at High Taxonomic Levels**

There were dissimilar bacterial communities even at the phylum level for the six PM$_{2.5}$ samples (Fig. 2). *Cyanobacteria* was the dominant phylum in BJ and BDH, accounting for 42.53% and 67.89% of the sequences in each library, however, which only accounted for a small proportion of the air sample collected from three different land-use types in Colorado, USA (Bowers et al., 2010). Differences in the relative abundance of bacterial community might be related to the different sampling locations (Brown, 2002). *Proteobacteria*, *Actinobacteria* and *Firmicutes* were dominant in LF, TJ, TS and BD, and the relative abundance of the three in the sequences of each library was 81.70%, 83.72%, 76.69% and 76.10%, respectively. *Proteobacteria* and *Actinobacteria* were also the dominant phyla in PM$_{2.5}$ collected during a severe smog event (Cao et al., 2014). However, *Proteobacteria* and *Cyanobacteria* were Gram-negative bacteria, which

![Fig. 2. The bacterial communities at phylum and class levels. The phylum or class with relative abundance greater than 1% of the total sequences were displayed, and the relative abundance less than 1% of the total sequences were defined as “Others”.](image-url)
were sharply inconsistent with the airborne bacterial communities detected by the culture-based methods, which suggested that 90% of the cultured airborne bacterial communities were Gram-positive bacteria (Lighthart, 1997). It is well known that only a low proportion of viable microorganisms could be cultured (Pace, 1997), therefore, the culture-based methods could overlook dominant airborne bacterial communities and underestimate their diversity. The other major phyla with average abundance higher than 1% of the total sequences were Bacteroidetes (2.61%), Deinococcus-Thermus (1.67%) and Chloroflexi (1.25%). The abundance of “Others” was less than 1% of total sequences in all the samples. These seven phyla were also the common inhabitants in the outdoor air, which had been frequently detected across cities in different countries and environments (Cao et al., 2014).

Among 53 classes detected in the six samples, 11 of them were dominant (> 1% abundance at least one sample), as depicted in Fig. 2, Cyanobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Actinobacteria and Bacilli were shared by the six samples with higher relative abundances than Clostridia, Deinococci, Cytophagia, Sphingobacteria and Thermomicrobia. Cyanobacteria accounted for 28.64% of the total sequences. Alphaproteobacteria (23.23% of the total sequences) were dominated in the phylum Proteobacteria, followed by Betaproteobacteria (13.06% of the total sequences) and Gammaproteobacteria (10.56% of the total sequences). The results were different with the previous study where aerosol samples collected in a rural environment in the south region of Paris (France) were dominated by Betaproteobacteria and Gammaproteobacteria and other subdivisions accounted for a small proportion (Maron et al., 2005). The relative abundance of Actinobacteria of the total sequences was 7.83%. Bacilli was the dominant class within phylum Firmicutes, followed by Clostridia, which accounted for 6.89% and 2.76% of the total sequences, respectively. The other major classes with relative abundances lower than 2% of the total sequences were Deinococci (1.67%), Cytophagia (1.35%), Sphingobacteria (0.79%) and Thermomicrobia (0.73%). Actinobacteria, Bacilli and Cytophagia were the cellulose-decomposing bacteria, which have been extensively studied in various environments (Wu et al., 2012).

For comparing the bacterial communities across samples, PCoA and CA were applied based on the weighted UniFrac distance matrix. As depicted in Fig. 3(a), the P1 and P2 explained 83.61% and 8.04% of the variance in overall community structure. It demonstrated that samples from BJ and BDH, LF and TJ, TS and BD tended to cluster together, respectively. However, the samples BJ and BDH were far from the samples LF, TJ, TS and BD (Fig. 3(a)), indicating that the bacterial communities in BJ and BDH were different from the other four samples. Also, as shown in Fig. 3(b), the results of CA revealed that total bacterial communities in the six samples could be clustered into three groups: (1) Group I contained two samples from BJ and BDH; (2) Group II contained two samples from LF and TJ; (3) Group III contained two samples from TS and BD. Additionally, the Group II and Group III were grouped together, indicating the similar bacterial community structure.
The shared OTUs of each group were further analyzed (Fig. 3(c)). Within Group I (BJ and BDH), 378 OTUs were shared, accounting for 49.09% of the total richness in the group (770 OTUs). For Group II (LF and TJ), the percentage of species shared was 62.92% (572 out of 909 OTUs). For Group III (TS and BD), the shared OTUs accounted for 64.56% of the total 762 OTUs.

**Sources Analysis Based on the Core Orders and Genera**

Most of bacterial aerosols originate from natural and anthropogenic sources, e.g., soil, plants, vegetables, water bodies, wastewater treatment plant and agriculture activities (Bowers et al., 2011; Bertolini et al., 2013). Plants are one of the main sources of airborne bacteria, so that the Chloroplast sequences among the total sequences were included for the bacterial community analysis, which may provide information on the potential sources of airborne bacteria (Brodie et al., 2007).

At the order level, 22 of 134 orders were dominant (> 1% abundance at least one sample) and shared by all the six samples. As shown in Fig. 4(a), each panel represented one of the 22 orders. The Chloroplast (panel 1) was more abundant in samples from BDH (67.56%) and BJ (41.06%) than the other four samples, suggesting that plants were the dominant sources for PM$_{2.5}$ in the two cities. The sampling surroundings of BJ and BDH were different, however, the bacterial communities of the two cities were clustered into group I (Figs. 3(a) and 3(b)), which might due to the similar dominant source (plants) of PM$_{2.5}$. The previous study confirmed that the dominant sources of bacteria in a given period of time have an important effect on the structure of the total airborne bacterial community (Bertolini et al., 2013). Another 12 orders (panel 2 to panel 13) with average abundance higher than 2% of the total sequences were Subsection IV (14.76%), Rhodobacterales (8.44%), Rhodospirillales (6.60%), Rhizobiales (5.94%), Rickettsiales (4.11%), Sphingomonadales (3.71%), Burkholderiales (2.90%), Enterobacteriales (2.86%), Pseudomonadales (2.76%), Micrococcales (2.21%), Frankiales (2.15%) and Corynebacteriales (2.13%). The relative importance of different potential sources of airborne bacteria could be speculated from the general knowledge of the ecology of the dominant bacterial taxa (Franzetti et al., 2011). Rhizobiales and Sphingomonadales were the plant-associated bacteria (Bertolini et al., 2013), especially Sphingomonadales was the common inhabitant of leaves (Murakami et al., 2010). Burkholderiales and Pseudomonadales were the soil-inhabiting bacteria (Bertolini et al., 2013). The results suggested that plants and soil maybe the main sources of the airborne bacteria in PM$_{2.5}$ in BTH.

The 35 dominant genera (> 1% abundance at least one sample) were selected from a total of 526 genera, as shown in Fig. 4(b). Streptophyta was abundant in BJ and BDH. Eight genera were abundant in three or four samples from TJ, BD, LF and TS, i.e., Paracoccus, Planomicrobiurn, Massilia, Kocuria, Arthrobacter, Deinococcus, Rubellimicrobiurn and Sphingomonas. In addition, Acinetobacter, Pseudomonas, Exiguobacterium, Psychrobacter and Shigella were also the

![Fig. 4. Distributions of core orders (a) and genera (b) in six samples. Heatmap was applied for the analysis of 35 dominant genera selected from 526 genera, which depicts the relative abundance of each bacterial genus (variables clustering on the Y-axis) within each sample (X-axis clustering).](image-url)
other megalopolis in the world in previous studies. Forty BTH, which were consistent with the results obtained at et al. The results suggested that soil and plants are the sources of these samples were investigated by Illumina sequencing. Bacteria (Bowers et al., 2010). In addition to the main sources (soils and plants), fecal material, most likely dog feces, is another main source of bacteria in 96 PM2.5 samples collected (soils and plants), fecal material). 18 known pathogens species were found, the highest in LF (1776), and highest in TJ (1815) (Table S6, in supplementary material). There were eight potential pathogen species, including Escherichia coli, Staphylococcus epidermidis, Streptococcus viridans, and Propionibacterium acnes. In this study, 7011 potential pathogen sequences were obtained, accounting for 3.61% of the total sequences (193969). The numbers of potential pathogen sequences were different across six samples, which were lower in BDH (242), higher in LF (842), TS (851), and BDH (1776), and highest in TJ (1815) (Table S6, in supplementary material). 18 known pathogens species were found, the relative abundance of which was 60.28% and 34.68%, respectively. There were eight potential pathogen species, the relative abundances of which were in the range of 2% to 9%, including Propionibacterium acnes (8.80%), Enterobacter cloacae (7.52%), Staphylococcus epidermidis (6.56%), Streptococcus galloylacticus (4.61%), Streptococcus mitis (4.16%), Pseudomonas aeruginosa (3.72%), Aeromonas hydrophila (2.40%), and Clostridium butyricum (2.08%). Moreover, there were another eight potential pathogen species with the relative abundance lower than 1.5%, i.e., Stenotrophomonas maltophilia, Vibrio parahaemolyticus, Thermobifida fusca, Clostridium perfringens, Haemophilus influenza, Acinetobacter baumannii, Saccharomonospora viridis and Clostridium tetani.

The relative abundances of 18 microbial pathogens species within the six PM2.5 libraries were depicted in Fig. 5(b). BJ, TJ, TS and BD contained 16 pathogens species. There were 17 and 13 pathogens species in LF and BDH, respectively. Escherichia coli was the dominant pathogens species in BJ, TS and BD, accounting for 60.88%, 38.9% and 26.18% of the pathogen sequences, respectively, which was also shared by the six libraries. This species is known to cause diseases including urinary tract infections, diarrhea, hemorrhagic colitis and hemolytic-uremic syndrome in humans (Ye and Zhang, 2011). Staphylococcus epidermidis, a causative agent of infections of implanted prostheses (e.g., heart valves and catheters), was dominant in LF, taking 24.23% of the pathogens sequences of LF, which was also the secondary species in BDH (16.94%). Propionibacterium acnes known to cause acne and other skin conditions (Bojar and Holland, 2004) was dominant in BDH, accounting for 31.40% of the pathogens sequences in BDH. The most dominant species in TJ and BD was Enterococcus faecium, and the relative abundance of which was 60.28% and 34.68%, respectively, and which is known to cause nosocomial infections (Ye and Zhang, 2011). Among the 18 pathogens species occurred in BTH, Clostridium tetani only occurred in BJ; Saccharomonospora viridis was shared by LF and TJ; Thermobifida fusca and Haemophilus influenza were both shared by four samples in BTH; Acinetobacter baumannii was shared by five samples except BDH. The other 13 pathogens species were shared by all the six samples. In summary, extremely low percentage of potential pathogens sequences was found in the total sequences of PM2.5 in BTH, and Enterococcus faecium and Escherichia coli showed high relative abundance. The dominant potential pathogens of PM2.5 samples in BTH were different and the potential pathogens were also shared by different PM2.5 samples. However, the risks of the potential pathogens to human health are still need to be further investigated.

Distribution of Microbial Potential Pathogens of PM2.5

In this study, 7011 potential pathogen sequences were obtained, accounting for 3.61% of the total sequences (193969). The numbers of potential pathogen sequences were different across six samples, which were lower in BDH (242), higher in LF (842), TS (851), and BD (1776), and highest in TJ (1815) (Table S6, in supplementary material). 18 known pathogens species were found, the relative abundance of which was 60.28% and 34.68%, respectively. There were eight potential pathogen species, the relative abundances of which were in the range of 2% to 9%, including Propionibacterium acnes (8.80%), Enterobacter cloacae (7.52%), Staphylococcus epidermidis (6.56%), Streptococcus galloylacticus (4.61%), Streptococcus mitis (4.16%), Pseudomonas aeruginosa (3.72%), Aeromonas hydrophila (2.40%), and Clostridium butyricum (2.08%). Moreover, there were another eight potential pathogen species with the relative abundance lower than 1.5%, i.e., Stenotrophomonas maltophilia, Vibrio parahaemolyticus, Thermobifida fusca, Clostridium perfringens, Haemophilus influenza, Acinetobacter baumannii, Saccharomonospora viridis and Clostridium tetani.

The relative abundances of 18 microbial pathogens species within the six PM2.5 libraries were depicted in Fig. 5(b). BJ, TJ, TS and BD contained 16 pathogens species. There were 17 and 13 pathogens species in LF and BDH, respectively. Escherichia coli was the dominant pathogens species in BJ, TS and BD, accounting for 60.88%, 38.9% and 26.18% of the pathogen sequences of each sample, respectively, which was also shared by the six libraries. This species is known to cause diseases including urinary tract infections, diarrhea, hemorrhagic colitis and hemolytic-uremic syndrome in humans (Ye and Zhang, 2011). Staphylococcus epidermidis, a causative agent of infections of implanted prostheses (e.g., heart valves and catheters), was dominant in LF, taking 24.23% of the pathogens sequences of LF, which was also the secondary species in BDH (16.94%). Propionibacterium acnes known to cause acne and other skin conditions (Bojar and Holland, 2004) was dominant in BDH, accounting for 31.40% of the pathogens sequences in BDH. The most dominant species in TJ and BD was Enterococcus faecium, and the relative abundance of which was 60.28% and 34.68%, respectively, and which is known to cause nosocomial infections (Ye and Zhang, 2011). Among the 18 pathogens species occurred in BTH, Clostridium tetani only occurred in BJ; Saccharomonospora viridis was shared by LF and TJ; Thermobifida fusca and Haemophilus influenza were both shared by four samples in BTH; Acinetobacter baumannii was shared by five samples except BDH. The other 13 pathogens species were shared by all the six samples. In summary, extremely low percentage of potential pathogens sequences was found in the total sequences of PM2.5 in BTH, and Enterococcus faecium and Escherichia coli showed high relative abundance. The dominant potential pathogens of PM2.5 samples in BTH were different and the potential pathogens were also shared by different PM2.5 samples. However, the risks of the potential pathogens to human health are still need to be further investigated.

Correlations between Environmental Factors and Bacterial Community Structure

The correlations between environmental factors and bacterial community structure (at phylum level) were analyzed via calculation of SRCC (Table S2 in supplementary material). Cyanobacteria were positively correlated to K+ and NO3–. Proteobacteria were positively correlated to Mg2+, Ca2+, NO3– and SO42–. Firmicutes showed negative correlation with SO2. Bacteroidetes were positively correlated to T, PM2.5, O3, Na+ and F and negatively correlated to K+ and NO2–. Deinococcus-Thermus were positively correlated to Ca2+ and SO42–.

Fig. 6(a) depicted the results of PCA and RDA analysis based on the environmental factors and bacterial community structure. Principal Component 1 (PC1) and PC2 explained 53.71% and 20.38% of the variance in overall community structure. Cyanobacteria showed significant negative correlation with PC1. Firmicutes, Bacteroidetes, Chloroflexi and others were positively correlated to PC2. RDA analysis suggested that the phyla showed different correlations with environment factors. Here, only some significantly positive correlation operational parameters are discussed. The results showed that Cyanobacteria had significantly positive correlation with NO3–, K+, SO2 and RH. Firmicutes and others were positively correlated to O3. T showed positive correlation with Bacteroidetes, Chloroflexi and Actinobacteria. PM2.5 was another factor positively correlated to Actinobacteria, SO2, Ca2+, EC, OC, NO3– and Mg2+ showed positive correlations with Proteobacteria and Deinococcus-Thermus. However, the results of Monte Carlo permutation tests for 21 constrained eigenvalues indicated that the null hypothesis could not be rejected. Although the relationships between the environmental factors and bacterial community structure could be investigated via SRCC, PCA and RDA, the results of these
analyses could not be used to evaluate the relative importance of environmental factors in shaping the bacterial community structure. The VPA could be applied to quantitatively assess the relative importance (or effects) of multiple environmental factors in affecting bacterial community structure. In this study, the 21 environmental variables were classified into three groups: meteorological conditions (named as M, including T, RH, AP and AQI), atmospheric pollutants (named as P, including PM$_{2.5}$, NO$_2$, SO$_2$, O$_3$ and CO) and the chemical compositions of PM$_{2.5}$ samples (named as C, including OC, EC and ten water-soluble ions) to investigate their variations in the airborne bacterial community. The result of VPA analysis was shown in Fig. 6(b). A total of 81.83% of the observed variations in the bacterial community structures (at phylum level) could be explained by these three environmental variables. Atmospheric pollutants explained the most of the variation (31.90%), followed by meteorological conditions (15.73%) and the chemical compositions (11.32%). The $p$ values of the relative effects of each factor or a combination of factors were further calculated. As shown in Fig. 6(b), all the $p$ values were much higher than 0.05, ranging from 0.13 to 0.40. The results indicated that there were no significant effects of environmental factors on the bacterial community structures. Based on the RDA and VPA analysis, it was difficult to pinpoint one environmental factor or one group of factors that shaped the bacterial community structure, and further investigations are needed.
Fig. 6. The PCA, RDA and VPA analysis based on the phylum level. (a) The PCA and RDA analysis. The purple point denoted phyla of bacterial community. The numbers: 1 = *Cyanobacteria*, 2 = others, 3 = *Firmicutes*, 4 = *Bacteroidetes*, 5 = *Chloroflexi*, 6 = *Actinobacteria*, 7 = *Proteobacteria*, 8 = *Deinococcus-Thermus*. The colored arrows denoted the environmental factors. (b) The VPA analysis. “M”, “P” and “C” represented meteorological conditions, atmospheric pollutants and chemical compositions of PM$_{2.5}$ samples, respectively. Each diagram represents the biological variation partitioned into the relative effects of each factor or a combination of factors. The values in the parentheses were the $p$ values of the relative effects.

CONCLUSIONS

The high-throughput MiSeq sequencing was used to investigate the bacterial communities of PM$_{2.5}$ in BTH. The abundance of total airborne bacteria of PM$_{2.5}$ in BTH was high, with the average of 1.19 $\times$ 10$^5$ cell m$^{-3}$ air. The samples collected form six cities were clustered into three groups: Group I (BJ and BDH), Group II (LF and TJ) and Group III (TS and BD). *Proteobacteria*, *Cyanobacteria* and *Actinobacteria* were the most dominant phyla of PM$_{2.5}$. Extremely low percentage (3.61%) of 18 common potential pathogens of the total sequences were observed. The pathogens were dominated by the *Enterococcus faecium* and *Escherichia coli*. Chloroplast showed high relative abundance of PM$_{2.5}$ and some soil-associated bacteria were found, suggesting plants and soil maybe the main sources of the airborne bacteria in PM$_{2.5}$ in BTH. Finally, VPA results indicated that the atmospheric pollutants explained the most of the variation (31.90%) in community structure of PM$_{2.5}$, followed by meteorological conditions (15.73%) and the chemical compositions of PM$_{2.5}$ (11.32%).

ACKNOWLEDGMENTS

We would like to thank the National Natural Science Foundation of China (51078007, 51378027, 51578015), the Natural Science Foundation of Beijing (8162010) and Beijing Talent Foundation of BJUT (2013-JH-L06) for the financial supports of this study, and we would like to thank State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences for the supports of the OC/EC and water-soluble ions analysis in PM$_{2.5}$.

DISCLAIMER

Reference to any companies or specific commercial products does not constitute its endorsement or recommendation by the NSFC.

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at http://www.aaqr.org.

REFERENCES

Ben, O.N. and Ampe, F. (2000). Microbial community dynamics during production of the Mexican fermented maize dough pozol. Appl. Environ. Microbiol. 66: 3664–3673.

Bertolini, V., Gandolfi, I., Ambrosini, R., Bestetti, G., Innocente, E., Rampazzo, G. and Franzetti, A. (2013). Temporal variability and effect of environmental variables on airborne bacterial communities in an urban area of Northern Italy. Appl. Microbiol. Biotechnol. 97: 6561–6570.

Bojar, R.A. and Holland, K.T. (2004). Acne and *propionibacterium acnes*. Clin. Dermatol. 22: 375–379.

Bowers, R.M., Lauber, C.L., Wiedinmyer, C., Hamady, M., Hallar, A.G., Fall, R., Knight, R. and Fierer, N. (2009). Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei. Appl. Environ. Microbiol. 75: 5121–5130.

Bowers, R.M., McLetchie, S., Knight, R. and Fierer, N. (2010). Spatial variability in airborne bacterial communities across land-use types and their relationship to the...
Cao, C., Jiang, W.J., Wang, B.Y., Fang, J.H., Lang, J.D., Tian, G., Jiang, J.K. and Zhu, T.F. (2014). Inhalable microorganisms in Beijing’s PM2.5 and PM10 pollutants during a severe smog event. Environ. Sci. Technol. 48: 1499–1507.

Cao, J.J., Lee, S.C., Chow, J.C., Watson, J.G., Ho, K.F., Zhang, R.J., Jin, Z.D., Shen, Z.X., Chen, G.C., Kang, Y.M., Zou, S.C., Zhang, L.Z., Qi, S.H., Dai, M.H., Cheng, Y. and Hu, K. (2007). Spatial and seasonal distributions of carbonaceous aerosols over China. J. Geophys. Res. 112: D22S11.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Cao, J.J., Chow, J.C., Watson, J.G., Ho, K.F., Cao, C., Jiang, W., Fang, J.H., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7: 335–336.

Chow, J.C., Watson, J.G., Chen, L.W., Chang, M.C., Robinson, N.F., Trimble, D. and Kohl, S. (2007). The IMPROVE_A temperature protocol for thermal/optical carbon analysis: Maintaining consistency with a long-term database. J. Air Waste Manage. Assoc. 57: 1014–1023.

Ferris, M.J., Muyzer, G. and Ward, D.M. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. 62: 340–346.

Franzetti, A., Gandolfi, I., Gaspari, E., Ambrosini, R. and Bestetti, G. (2011). Seasonal variability of bacteria in fine and coarse urban air particulate matter. Appl. Microbiol. Biotechnol. 90: 745–753.

Lee, S.H., Lee, H.J., Kim, S.J., Lee, H.M., Kang, H. and Kim, Y.P. (2010). Identification of airborne bacterial and fungal community structures in an urban area by T-RFLP analysis and quantitative real-time PCR. Sci. Total Environ. 408: 1349–1357.

Li, X.R., Wang, L.L., Ji, D.S., Wen, T.X., Pan, Y.P., Sun, Y. and Wang, Y.S. (2013). Characterization of the size-segregated water-soluble inorganic ions in the Jing-Jin-Ji urban agglomeration: Spatial/temporal variability, size distribution and sources. Atmos. Environ. 77: 250–259.

Lighthart, B. (1997). The ecology of bacteria in the alfresco atmosphere. FEMS Microbiol. Ecol. 23: 263–274.

Maron, P.A., Lejon, D.P.H., Carvalho, E., Bizet, K., Lemanceau, P., Ranjard, L. and Mougel, C. (2005). Assessing genetic structure and diversity of airborne bacterial communities by DNA fingerprinting and 16S rDNA clone library. Atmos. Environ. 39: 3687–3695.

McBain, A.J., Bartolo, R.G., Catrenich, C.E., Charbonneau, D., Ledder, R.G., Rickard, A.H., Symmons, S.A. and Gilbert, P. (2003). Microbial characterization of biofilms in domestic drains and the establishment of stable biofilm microcosms. Appl. Environ. Microbiol. 69: 177–185.

Murakami, Y., Otsuka, S. and Senoo, K. (2010). Abundance and community structure of sphingomonads in leaf residues and nearby bulk soil. Microbes Environ. 25: 183–189.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276:734–740.

Pillai, S.D. and Ricke, S.C. (2002). Bioaerosols from municipal and animal wastes: Background and contemporary issues. Can. J. Microbiol. 48: 681–696.

Pope, C.R., Ezzati, M. and Dockery, D.W. (2009). Fine-particle air pollution and life expectancy in the United States. N. Engl. J. Med. 360: 376–386.

Robertson, C.E., Baumgartner, L.K., Harris, J.K., Peterson, K.L., Stevens, M.J., Frank, D.N. and Pace, N.R. (2013). Culture-independent analysis of aerosol microbiology in a metropolitan subway system. Appl. Environ. Microbiol. 79: 3485–3493.

Sjoling, A., Wiklund, G., Savarino, S.J., Cohen, D.I. and Svensson, A.M. (2007). Comparative analyses of phenotypic and genotypic methods for detection of enterotoxigenic Escherichia coli toxins and colonization factors. J. Clin. Microbiol. 45: 3295–3301.

Wu, S.G., Wang, G.T., Angert, E.R., Wang, W.W., Li, W.X. and Zou, H. (2012). Composition, diversity, and origin of the bacterial community in grass carp intestine. PLoS One 7: e30440.

Ye, L. and Zhang, T. (2011). Pathogenic bacteria in sewage treatment plants as revealed by 454 pyrosequencing. Environ. Sci. Technol. 45: 7173–7179.

Zhang, R.J., Jin, Z.D., Shen, Z.X., Chen, G.C., Kang, Y.M., Zou, S.C., Zhang, L.Z., Qi, S.H., Dai, M.H., Cheng, Y. and Hu, K. (2007). Spatial and seasonal distributions of carbonaceous aerosols over China. J. Geophys. Res. 112: D22S11.