

Supplementary Material

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 and Kai-Ling Pan

Correspondence and phone calls about the paper should be directed to Jing-Feng Gao at the following address, phone and fax number, and e-mail address:

Jing-Feng Gao

College of Environmental and Energy Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

Tel: +86-10-6739-2627(office); Fax: +86-10-6739-1983

E-mail: gao.jingfeng@bjut.edu.cn or gao158@gmail.com

Text 1 Sequence analysis of Illumina MiSeq PE300 sequencing

Fig. S1. Concentrations of atmospheric pollutants (NO₂, SO₂, O₃ and CO) and chemical compositions of PM_{2.5} including OC, EC and water-soluble inorganic ions in BTH. (a) atmospheric pollutants; (b) OC and EC; (c) water-soluble inorganic ions.

Fig. S2. Rarefaction analysis of microbial community PM_{2.5} in BTH.

Table S1. Description of samples collected and corresponding meteorological conditions

Table S2. SRCCs between abundance, dominant phyla of bacterial community and meteorological data, atmospheric pollutants and chemical compositions of PM_{2.5} (** indicated $p < 0.01$; * indicate $p < 0.05$, two tailed. SRCC values with $p < 0.05$ are shown in blue.)

Table S3. Diversity and richness estimation of 16S rRNA gene libraries of PM_{2.5} using Illumina MiSeq PE300 sequencing

Table S4. Numbers of taxa classified by six different taxonomic levels

Table S5. The proportions of unclassified taxa by six different taxonomic levels

Table S6. Raw data of pathogenic bacteria screened in the total sequences

Text 1 Sequence analysis of Illumina MiSeq PE300 sequencing

The primary image analysis and basecalling of the reads were performed on the Illumina MiSeq PE300 sequencer. The reads obtained were de-multiplexed according to barcode sequences. Trimmomatic (Bolger *et al.*, 2014) and FLASH (Magoc and Salzberg, 2011) were used to screen the low quality reads and merge the PE reads based on the overlap, respectively. The merged sequences after removing the adapter, barcodes and primers were defined as “trimmed sequences”, which were filtered out the chimeras using the UPARSE pipeline (Edgar, 2013) and then used for the following analysis.

The Quantitative Insights into Microbial Ecology (QIIME) pipeline, an open-source software package was applied for post-sequencing analysis (Caporaso *et al.*, 2010). Operational taxonomic units (OTUs) were picked using the QIIME script `pick_de_novo_otus.py` at 97% similarity, which includes seven steps: 1) pick OTUs at 97% sequence similarity; 2) pick a representative sequence for each OTU; 3) align the sequences; 4) assign taxonomy according to the Greengenes database; 5) filter the alignment; 6) build a phylogenetic tree; and 7) make an OTU table. The most important output was OTU table, in which the representative sequences of OTUs were assigned to corresponding taxonomic ranks with the Greengenes database. The Venn analysis was also used to investigate the core microbiota, which could describe the shared and unique OTUs of some samples. The alpha diversity (Shannon, Chao 1, ACE, Simpson and Good’s coverage) and rarefaction curves were generated by the `alpha_rarefaction.py` workflow script.

Based on the composition of microbial communities, beta diversity could represent their explicit comparison, which could assess the differences between the communities. Beta diversity was performed with the QIIME script `beta_diversity_through_plots.py`.

To normalize all the samples at equal sequencing depth, the distance matrix was calculated between samples at a rarefaction level of 25462 (the minimum sequences of six PM_{2.5} samples), which could be visualized as the CA and PCoA. There are many beta diversity metrics the QIIME pipeline provided, we only calculated the default and extensively used beta diversity metrics (weighted and unweighted UniFrac) (Lozupone and Knight, 2005).

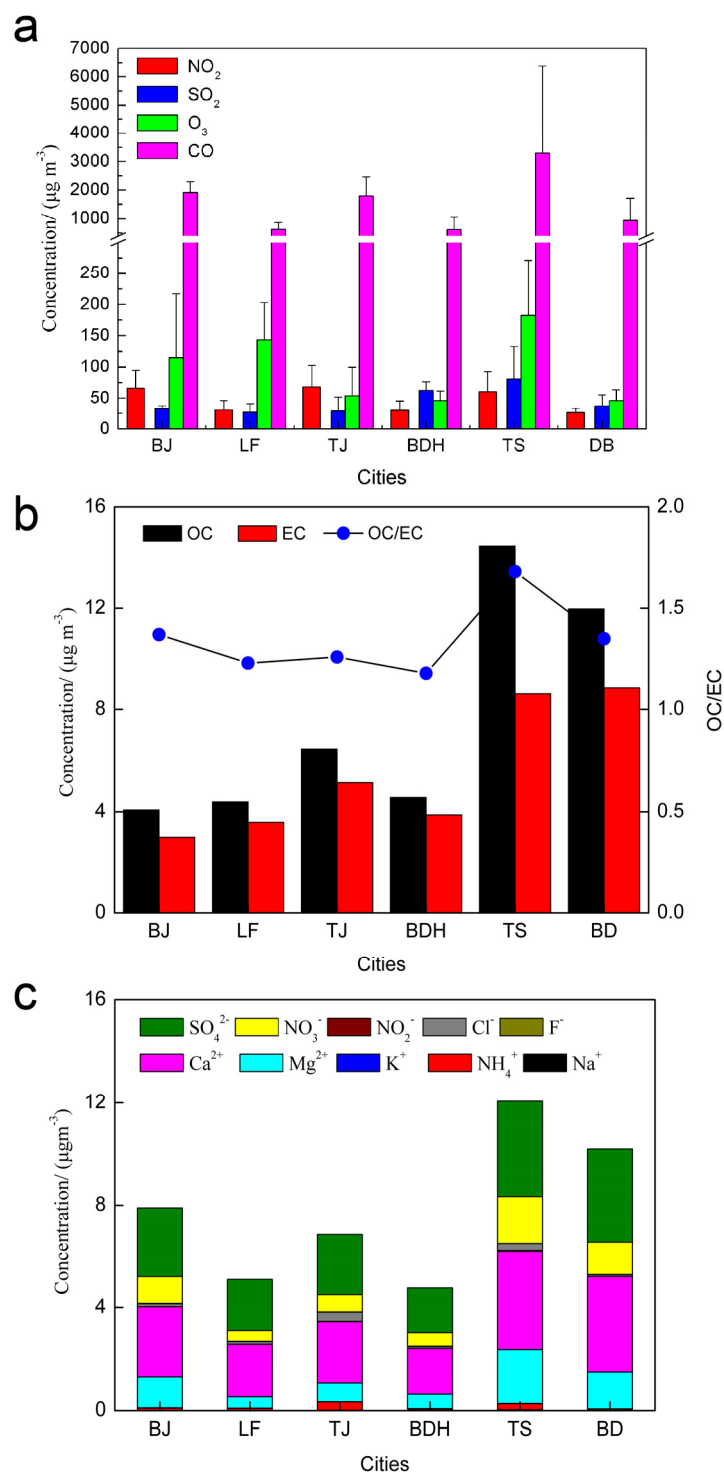


Fig. S1. Concentrations of atmospheric pollutants (NO_2 , SO_2 , O_3 and CO) and chemical compositions of $\text{PM}_{2.5}$ including OC, EC and water-soluble inorganic ions in BTH. (a) atmospheric pollutants; (b) OC and EC; (c) water-soluble inorganic ions.

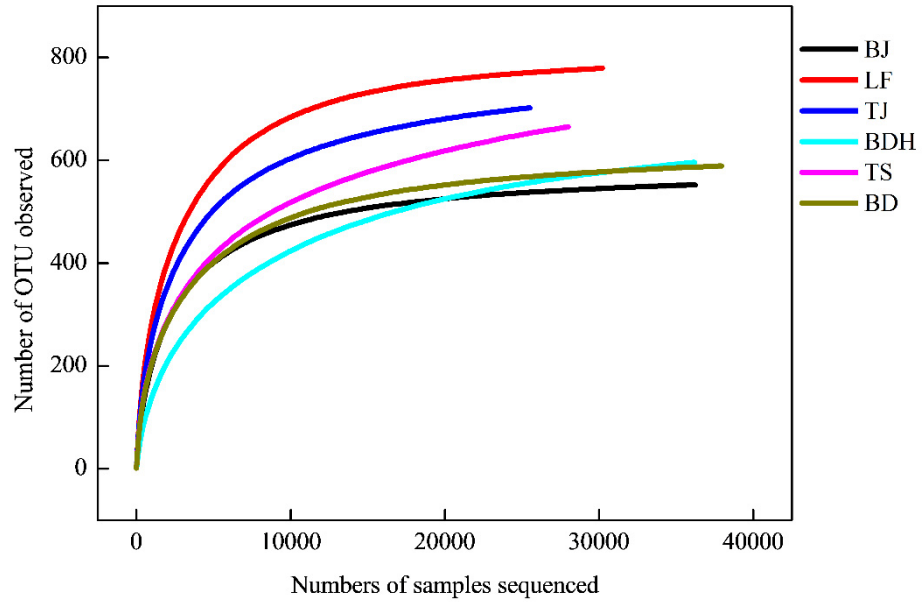


Fig. S2. Rarefaction analysis of microbial community PM_{2.5} in BTH.

Table S1. Description of samples collected and corresponding meteorological conditions

Cites	City type	Surroundings	Sampling date		T (°C)	RH (%)	AP (Hpa)	AQI
			Start time	End time				
Beijing (BJ)	Megacity	RCTMD	5-21		24	52	1005.25	183
				5-22	27	48	1006.11	188
Langfang (LF)	Suburban	RCTMD	5-23		31	34	1009.64	138
				5-24	23	75	1007.89	104
Tianjin (TJ)	Megacity	RCTMD	5-25		24	54	1006.31	134
				5-26	26	20	1000.35	82
Beidaihe (BDH)	Coastal urban	Bohai	5-27		17	62	1008.00	79
				5-28	19	60	1007.00	54
Tangshan (TS)	Industrial urban	RCTMD	5-29		23	51	1004.00	192
				5-30	32	54	1004.00	201
Baoding (BD)	Industrial urban	RCTMD	5-31		30	56	1004.00	158
				6-1	23	57	1005.00	73

Table S2. SRCCs between abundance, dominant phyla of bacterial community and meteorological data, atmospheric pollutants and chemical compositions of PM_{2.5} (** indicated $p < 0.01$; * indicate $p < 0.05$, two tailed. SRCC values with $p < 0.05$ are shown in blue.)

variables	Abundance		<i>Cyanobacteria</i>		<i>Proteobacteria</i>		<i>Actinobacteria</i>		<i>Firmicutes</i>		<i>Bacteroidetes</i>		<i>Deinococcus-Thermus</i>		<i>Chloroflexi</i>		Others	
	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values	SRCC	<i>p</i> values
T	-0.257	0.623	-0.371	0.468	0.543	0.266	0.714	0.111	0.029	0.957	0.829*	0.042	0.714	0.111	0.371	0.468	0.029	0.957
RH	0.029	0.957	0.600	0.208	-0.200	0.704	0.086	0.872	-0.714	0.111	-0.371	0.468	-0.029	0.957	-0.257	0.623	0.2	0.704
AP	0.600	0.208	0.429	0.397	-0.543	0.266	-0.086	0.872	-0.200	0.704	-0.314	0.544	-0.257	0.623	0.371	0.468	0.771	0.072
PM_{2.5}	-0.714	0.111	-0.714	0.111	0.600	0.208	0.543	0.266	0.257	0.623	0.829*	0.042	0.486	0.329	-0.086	0.872	-0.6	0.208
AQI	0.143	0.787	-0.143	0.787	0.543	0.266	0.314	0.544	0.029	0.957	0.6	0.208	0.657	0.156	0.371	0.468	0.029	0.957
NO₂	0.200	0.704	-0.486	0.329	-0.086	0.872	-0.371	0.468	0.600	0.208	0.257	0.623	-0.257	0.623	0.086	0.872	-0.257	0.623
SO₂	-0.029	0.957	0.543	0.266	0.314	0.544	-0.200	0.704	-0.886*	0.019	-0.086	0.872	0.257	0.623	-0.771	0.072	-0.657	0.156
O₃	0.086	0.872	-0.486	0.329	0.257	0.623	0.314	0.544	0.257	0.623	0.829*	0.042	0.371	0.468	0.371	0.468	-0.029	0.957
CO	-0.086	0.872	-0.257	0.623	0.657	0.156	0.086	0.872	0.143	0.787	0.486	0.329	0.543	0.266	0.029	0.957	-0.429	0.397
OC	-0.771	0.072	-0.257	0.623	0.543	0.266	0.429	0.397	-0.314	0.544	0.486	0.329	0.429	0.397	-0.543	0.266	-0.771	0.072
EC	-0.886*	0.019	-0.200	0.704	0.600	0.208	0.543	0.266	-0.257	0.623	0.314	0.544	0.486	0.329	-0.429	0.397	-0.6	0.208
Na⁺	-0.600	0.208	-0.771	0.072	0.200	0.704	0.429	0.397	0.200	0.704	0.886*	0.019	0.143	0.787	-0.2	0.704	-0.6	0.208
NH₄⁺	0.029	0.957	-0.600	0.208	-0.029	0.957	-0.257	0.623	0.486	0.329	0.486	0.329	-0.2	0.704	-0.086	0.872	-0.486	0.329
K⁺	0.543	0.266	0.829*	0.042	-0.371	0.468	-0.486	0.329	-0.371	0.468	-0.943**	0.005	-0.314	0.544	-0.029	0.957	0.486	0.329

Mg²⁺	-0.371	0.468	0.086	0.872	0.886*	0.019	0.257	0.623	-0.314	0.544	0.257	0.623	0.771	0.072	-0.257	0.623	-0.543	0.266
Ca²⁺	-0.429	0.397	-0.143	0.787	0.943**	0.005	0.486	0.329	-0.086	0.872	0.486	0.329	0.886*	0.019	0.029	0.957	-0.371	0.468
F⁻	-0.600	0.208	-0.771	0.072	0.200	0.704	0.429	0.397	0.200	0.704	0.886*	0.019	0.143	0.787	-0.2	0.704	-0.6	0.208
Cl⁻	0.029	0.957	-0.600	0.208	-0.029	0.957	-0.257	0.623	0.486	0.329	0.486	0.329	-0.2	0.704	-0.086	0.872	-0.486	0.329
NO₂⁻	0.543	0.266	0.829*	0.042	-0.371	0.468	-0.486	0.329	-0.371	0.468	-0.943**	0.005	-0.314	0.544	-0.029	0.957	0.486	0.329
NO₃⁻	-0.371	0.468	0.086	0.872	0.886*	0.019	0.257	0.623	-0.314	0.544	0.257	0.623	0.771	0.072	-0.257	0.623	-0.543	0.266
SO₄²⁻	-0.429	0.397	-0.143	0.787	0.943**	0.005	0.486	0.329	-0.086	0.872	0.486	0.329	0.886*	0.019	0.029	0.957	-0.371	0.468

Table S3. Diversity and richness estimation of 16S rRNA gene libraries of PM_{2.5} using Illumina MiSeq PE300 sequencing

Sample ID	3% cut off					
	OTUs	Good's coverage/ (%)	ACE	Chao1	Shannon	Simpson
BJ	552	99.90	565	571	3.8	0.1337
LF	779	99.85	793	800	4.97	0.0281
TJ	702	99.67	743	758	4.88	0.023
BDH	596	99.71	663	663	2.33	0.437
TS	665	99.51	765	794	4.48	0.0361
BD	589	99.88	605	613	4.37	0.0447

Table S4. Numbers of taxa classified by six different taxonomic levels

Taxonomic level	Total number	The number of taxa at each level						Range	Average
		BJ	LF	TJ	BDH	TS	BD		
Phylum	29	19	21	15	18	18	18	15-21	18.2
Class	57	37	45	35	33	35	37	33-45	37.0
Order	134	86	104	84	81	88	86	81-104	88.2
Family	259	166	215	171	159	177	163	159-215	175.2
Genus	526	318	423	365	308	340	319	308-423	345.5

Table S5. The proportions of unclassified taxa by six different taxonomic levels

Taxonomic level	Proportions of unclassified taxa*					
	BJ	LF	TJ	BDH	TS	BD
Phylum	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Class	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Order	0.21%	0.13%	0.07%	0.11%	0.02%	0.07%
Family	0.55%	0.67%	1.04%	0.50%	0.36%	0.44%
Genus	8.98%	9.36%	11.75%	2.22%	13.15%	5.35%
Species	76.04%	63.20%	66.11%	86.03%	68.90%	69.61%

*: Sequences that could not to be assigned to any known group were assigned as

‘Unclassified bacteria’.

Table S6. Raw data of pathogenic bacteria screened in the total sequences

OTU ID	Numbers of sequences						Genus	Species (Blast)
	BJ	LF	TJ	BDH	TS	BD		
OTU335	135	98	92	76	45	171	<i>Propionibacterium</i>	<i>Propionibacterium acnes</i>
OTU231	0	18	16	0	0	0	<i>Saccharomonospora</i>	<i>Saccharomonospora viridis</i>
OTU58	0	74	5	0	8	1	<i>Thermobifida</i>	<i>Thermobifida fusca</i>
OTU486	9	16	1094	10	102	616	<i>Enterococcus</i>	<i>Enterococcus faecium</i>
OTU76	23	75	34	16	23	121	<i>Streptococcus</i>	<i>Streptococcus mitis</i>
OTU733	39	55	62	14	99	54	<i>Streptococcus</i>	<i>Streptococcus gallolyticus</i>
OTU728	56	204	86	41	37	36	<i>Staphylococcus</i>	<i>Staphylococcus epidermidis</i>
OTU474	16	0	0	0	0	0	<i>Clostridium</i>	<i>Clostridium tetani</i>
OTU839	3	45	12	10	57	19	<i>Clostridium</i>	<i>Clostridium butyricum</i>
OTU936	8	33	21	2	5	12	<i>Clostridium</i>	<i>Clostridium perfringens</i>
OTU720	5	2	3	0	2	37	<i>Acinetobacter</i>	<i>Acinetobacter baumannii</i>
OTU507	0	0	0	0	1	18	<i>Aeromonas</i>	<i>Aeromonas hydrophila</i>
OTU894	41	20	22	9	10	47	<i>Aeromonas</i>	<i>Aeromonas hydrophila</i>
OTU121	804	0	36	0	129	403	<i>Escherichia</i>	<i>Escherichia coli</i>
OTU735	100	39	38	14	202	62	<i>Escherichia</i>	<i>Escherichia coli</i>
OTU717	58	67	249	23	79	51	<i>Enterobacter</i>	<i>Enterobacter cloacae</i>
OTU684	46	18	0	0	8	5	<i>Haemophilus</i>	<i>Haemophilus influenzae</i>
OTU332	0	0	0	1	11	0	<i>Pseudomonas</i>	<i>Pseudomonas aeruginosa</i>
OTU838	87	37	25	15	21	64	<i>Pseudomonas</i>	<i>Pseudomonas aeruginosa</i>
OTU315	24	28	15	3	3	29	<i>Stenotrophomonas</i>	<i>Stenotrophomonas maltophilia</i>
OTU447	12	9	4	3	4	9	<i>Vibrio</i>	<i>Vibrio parahaemolyticus</i>
OTU518	19	4	1	5	5	21	<i>Vibrio</i>	<i>Vibrio parahaemolyticus</i>
Total	1485	842	1815	242	851	1776		

References

- Bolger, A.M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30: 2114-2120.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Method.*7: 335-336.
- Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*. 10: 996-998.
- Lozupone, C. and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71: 8228-8235.
- Magoc, T. and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27:2957-2963.