Illumina MiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas

Chen Hong · Yanxiao Si · Yi Xing · Yang Li

Received: 10 October 2014 / Accepted: 29 January 2015 / Published online: 12 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Mine activities leaked heavy metals into surrounding soil and may affected indigenous microbial communities. In the present study, the diversity and composition of the bacterial community in soil collected from three regions which have different pollution degree, heavy pollution, moderate pollution, and non-pollution, within the catchment of Chao River in Beijing City, were compared using the Illumina MiSeq sequencing technique. Rarefaction results showed that the polluted area had significant higher bacterial alpha diversity than those from unpolluted area. Principal component analysis (PCA) showed that microbial communities in the polluted areas had significant differences compared with the unpolluted area. Moreover, PCA at phylum level and Matastats results demonstrated that communities in locations

shared similar phyla diversity, indicating that the bacterial community changes under metal pollution were not reflected on phyla structure. At genus level, the relative abundance of dominant genera changed in sites with degrees of pollution. Genera *Bradyrhizobium*, *Rhodanobacter*, *Reyranella*, and *Rhizomicrobium* significantly decreased with increasing pollution degree, and their dominance decreased, whereas several genera (e.g., *Steroidobacter*, *Massilia*, *Arthrobacter*, *Flavisolibacter*, and *Roseiflexus*) increased and became new dominant genera in the heavily metal-polluted area. The potential resistant bacteria, found within the genera of *Thiobacillus*, *Pseudomonas*, *Arthrobacter*, *Microcoleus*, *Leptolyngbya*, and *Rhodobacter*, are less than 2.0 % in the indigenous bacterial communities, which play an important role in soil ecosystem. This effort to profile the background diversity may set the first stage for better understanding the mechanism underlying the community structure changes under in situ mild heavy metal pollution.

Responsible editor: Zhihong Xu

Electronic supplementary material The online version of this article (doi:10.1007/s11356-015-4186-3) contains supplementary material, which is available to authorized users.

C. Hong · Y. Si · Y. Xing (✉) · Y. Li
School of Civil and Environmental Engineering, and Key Laboratory of Metal and Mine Efficiently Exploiting and Safety, Ministry of Education, University of Science and Technology Beijing, 100083 Beijing, People's Republic of China
e-mail: xingyi@ustb.edu.cn

Y. Xing
e-mail: xingyi204@gmail.com

C. Hong
Research Center for Eco-Environmental Science, Chinese Academy of Sciences, Shuangqing Road No. 18, Haidian District, Beijing 100085, People's Republic of China

Y. Si
Tsinghua University, Shuangqing Road No. 30, Haidian District, Beijing 100084, People's Republic of China

Keywords Iron mine · MiSeq sequencing · Alpha diversity · Bacterial phyla · Potential resistance genera · Heavy metal

Introduction

Mining activities adversely affect the quality of surrounding soils and aquifers and threaten human health and ecosystems (Anjos et al. 2012; Rodríguez et al. 2009). Studies on the relationship between mining and microbial communities in surrounding soils mainly focused on nonferrous metals (e.g., Cu, Pb, and Sb), which are highly toxic to soil biota (Zhang et al. 2007). However, few studies have focused on iron mining despite its specific environmental impact. In iron mining area, great amounts of tailing dumps and iron deposit rocks

piled up directly on the ground and bring different kinds of harmful heavy metals (i.e., Cu, Zn, Cr, Cd, Pb) to surrounding soil through wind and hydraulic erosion. Besides, several surveys also indicated that iron mining areas, even heavily polluted mining area or tailing dam, always contain lower heavy metals contents than other nonferrous metal mining areas (Marmolejo-Rodríguez et al. 2007; Wang et al. 2012b; Huang et al. 2013). Pollution caused by iron mine activities was wider, gentler, and more recondite, and the microorganism which widely distributed in the heavy metal-polluted area may present more unique and subtle community changes. Study on soil microbe community structure could not only guide us in exploring the potential risk of affected soil but also provide reference for soil remediation with several indigenous resistance bacteria.

Recent developments in molecular fingerprinting techniques, such as denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism, and real-time PCR quantification, have allowed researchers to investigate the microbial diversity variations in soils contaminated with heavy metals (Berg et al. 2012; Cáliz et al. 2013; Giloteaux et al. 2011; Petrić et al. 2011; Zhao et al. 2014). Molecular fingerprint techniques have high resolution; however, the fingerprint complexity caused by the high richness of indigenous microbial communities in soil may mask the changes induced by heavy metals (Gomes et al. 2010). Rare ecologically significant species that may have been affected may avoid detection by DGGE because of the detection limit (Bruce et al. 2000). The recent application of the next-generation sequencing methods, such as pyrosequencing 454 and Illumina, may provide a more direct way to detect the microbial taxa, especially those with low-abundant species changes (Oberauner et al. 2013; Uroz et al. 2013). In addition, Illumina sequencing is cost-effective and could obtain tenfold or more sequences per sample than pyrosequencing 454, thereby allowing the analysis of a high number of detailed taxonomic profiles from samples (Kozich et al. 2013).

Recent studies have focused on either the intense in situ or exogenous heavy metal impact on taxonomic community diversity. Berg observed that the relative abundance of bacterial groups changed within a Cu gradient (20 to 3537 mg/kg), which originated from CuSO₄ contamination over 85 years ago (Berg et al. 2012). Margarita discussed the diversity of taxonomic groups after a 60-day incubation period with Cd concentrations reaching 12.5 mg/kg (Ros et al. 2009). However, these previous studies ignored the ubiquitous mild heavy metal pollution effect on soil microbes, which was apparent in the iron mine in our studies. The present study was conducted in the Anzigou iron mine, located north of Miyun County approximately 10 km upstream from the Miyun Reservoir in Beijing City. Frequent mining activities have caused heavy metal pollution in the Miyun Reservoir and have

directly affected water quality and safety. Several researchers have investigated the heavy metal distribution in the catchment area of Miyun Reservoir (Huang et al. 2013; Qin et al. 2014). However, few studies have characterized potential ecological risks of such distribution.

In this study, to obtain the comprehensive effect of bacterial communities structure, we compared all soil types located in three regions. The three regions were as follows: (1) ore treatment factory, ore mining region, and tailing dam from mining area; and (2) croplands, planted forest, and vegetable fields from two villages (X and Y). The soils from each location from the same region have different physical and chemical properties, but they have the same pollution level. The present study, therefore, aims to provide detailed variations of bacterial communities that came from different soil types under mild heavy metal pollution and explore the natural bioremedial consortia.

Materials and methods

Soil sampling and heavy metal determination

The study area was the Anzigou iron mine (40° 36' N, 117° 08' E), which is located in the north of Miyun County, Beijing City, China, upstream of the Miyun Reservoir. The total sampled surface soil (0 to 20 cm in depth) was collected from three regions, namely H, M, and C, respectively, at the Anzigou iron mine in October 2012. Three H samples were collected from the ore treatment factory, ore mining region, and tailing dam. Three M samples were collected from three different soil types (croplands, planted forest, and vegetable fields) in village X, which are 1 km downstream of the tailing dam near the Chao River. Three C samples were collected from three different soil types (croplands, planted forest, and vegetable fields) in village Y, which is 5 km downstream of the tailing dam (Fig. 1). Sample at each site was collected in triplicates. The soil samples were packed in sterile Ziploc bags, and each bag was sealed and transported to the laboratory for further study. A fraction of the mixed sample was stored at −20 °C for molecular analysis. Another fraction was dried at room temperature for a week and sieved through a 4-mm grid to remove stones and visible plant fragments. An aliquot was threaded through a 0.125-mm sieve for the metal content analyses. All samples were stored at 4 °C. Water content in samples was determined immediately after collection by drying at 105 °C until constant weight. The pH was determined with a pH meter (Starter-3C, OHAUS, USA) in a 1:2.5 suspension of ultrapure water. Organic matter was determined with a 2400II CHNS/O elemental analyzer (PerkinElmer, USA) after removing the inorganic C by digesting in 1 M HCl for 48 h, washing with ultrapure water until becoming neutral, and then drying at low temperature. Soil samples were extracted by

Fig. 1 The study area and sampling sites location

sodium bicarbonate to determine the available phosphorus using the molybdenum antimony colorimetric method. Ammonium nitrogen contents (NH_4^+) were determined using the indophenol blue colorimetric method (Keeney and Nelson 1982). Nitrate contents (NO_3^-) were measured by dual wavelength spectrometry (Norman et al. 1985). The total concentration of the metals was measured after digesting the soil sample in a mixture of H_2O_2 , HNO_3 , and HCl (1:4:2, v/v) in a sealed digestion tank at 190 °C for 30 h. The metal concentrations were determined using an inductively coupled plasma optical emission spectrometer (Varian, 720-ES, USA).

DNA extraction and PCR amplification

Microbial DNA was extracted from 0.5-g soil samples using the Power Soil DNA kit (MoBio Laboratories, Solana Beach, CA) according to the manufacturer's protocols. The V3–V4 regions of the bacterial 16S rRNA gene were amplified using PCR (95 °C for 2 min, followed by 25 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s, and a final extension at 72 °C for 10 min) using the primers (Dennis et al. 2013) 338 F 5'-barcode-ACTCCTACGGGAGGCAGCA-3' and 806R 5'-GGACTACHVGGGTWTCTAAT-3'. The barcode is an eight-base sequence unique to each sample. PCR reactions were performed in triplicate in a 20- μL mixture containing 4 μL of 5 $\times$ Fast Pfu buffer, 2 μL of 2.5 mM dNTPs, 0.4 μL of each primer (5 μM), 0.4 μL of Fast Pfu polymerase, and 10 ng of template DNA.

Illumina MiSeq sequencing

Amplicons were extracted from 2 % agarose gels and purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) according to the

manufacturer's instructions and quantified using QuantiFluor™-ST (Promega, USA). Purified amplicons were pooled in equimolar and paired-end sequenced (2×300) on an Illumina MiSeq platform (Majorbio, Shanghai) according to the standard protocols.

Processing and analyzing of sequencing data

Raw FASTQ files were de-multiplexed and quality-filtered using QIIME (version 1.17) with the following criteria: (i) The 300-bp reads were truncated at any site that obtained an average quality score of <20 over a 10-bp sliding window, and the truncated reads shorter than 50 bp were discarded; (ii) exact barcode matching, two nucleotide mismatch in primer matching, and reads containing ambiguous characters were removed; and (iii) only overlapping sequences longer than 10 bp were assembled according to their overlapped sequence. Reads that could not be assembled were discarded. Operational taxonomic units (OTUs) with 97 % similarity cutoff were clustered using UPARSE (version 7.1), and chimeric sequences were identified and removed using UCHIME.

The rarefaction analysis based on Mothur v.1.21.1 (Schloss et al. 2009) was conducted to reveal the diversity indices, including the Chao, ACE, and Shannon diversity indices. The beta diversity analysis was performed using UniFrac (Lozupone et al. 2011) to compare the results of the principal component analysis (PCA) using the community ecology package, R-forge (Vegan 2.0 package was used to generate a PCA figure). Venn diagrams were implemented by Venn Diagram, while Mantel test, Redundancy analysis (RDA), and Heatmap figures were performed in Vegan packages in R.

Results and discussion

Pollution site background

Table 1 presents the soil characteristics and heavy metal contents determined in the present study. Except for CV, pH value from other site changes slightly from 7.23 to 7.82. The soil acidification in site CV may be caused by excessive application of nitrogen fertilizer (Guo et al. 2010). Soil water contents ranged from 7.69 to 22.42 %, reaching the lowest value in region H. The organic matter contents had no obvious change in different sampling sites. In addition, the various values of available phosphorus, NH_4^+ and NO_3^- , were likely due to the fertilizing habit or land use pattern. The average Cr and Cd concentrations were 298.33 and 0.708 mg/kg, respectively, and maximum values were observed in group H. The Cd, Cu, and Cr concentrations in all samples exceeded the Chinese national background value (0.20, 35, and 90 mg/kg, the primary standard set by Chinese National Environmental Quality), with the average Cr and Cd

Table 1 pH value and heavy metal contents of soil samples

Sampling sites	pH	Water contents (%)	Organic matter (g/kg)	Available phosphorus (mg/kg)	NH ₄ ⁺ (mg/kg)	NO ₃ ⁻ (mg/kg)	Heavy metal contents				Nemerow index	
							Cd (mg/kg)	Cr (mg/kg)	Cu (mg/kg)	Zn (mg/kg)		Fe (g/kg)
HF	7.23 (±0.12)	7.69 (±0.35)	0.77 (±0.001)	4.12 (±0.11)	2.09 (±0.05)	0.14 (±0.01)	0.69 (±0.01)	280.60 (±10.09)	47.32 (±1.09)	101.23 (±3.48)	89.27 (±3.25)	3.01
HM	7.69 (±0.20)	9.10 (±0.19)	0.91 (±0.001)	5.54 (±0.14)	7.11 (±0.29)	0.57 (±0.01)	0.75 (±0.03)	324.44 (±9.20)	62.11 (±2.31)	79.12 (±3.29)	104.01 (±4.08)	3.17
HT	7.81 (±0.35)	8.34 (±0.31)	0.55 (±0.002)	3.79 (±0.32)	1.25 (±0.04)	6.03 (±0.04)	0.68 (±0.02)	294.25 (±7.28)	61.29 (±3.37)	102.30 (±3.24)	125.70 (±2.75)	3.02
Mean value	7.58	8.38	0.72	4.48	3.48	2.24	0.71	299.76	59.91	94.22	106.33	3.06
MC	7.26 (±0.11)	19.72 (±0.98)	1.01 (±0.004)	60.27 (±2.92)	1.09 (±0.03)	0.22 (±0.01)	0.49 (±0.03)	156.29 (±5.09)	35.76 (±0.97)	95.52 (±4.12)	29.36 (±1.02)	2.05
MF	7.19 (±0.42)	18.39 (±1.02)	1.14 (±0.004)	27.54 (±1.22)	0.78 (±0.01)	19.13 (±1.88)	0.50 (±0.02)	134.76 (±4.29)	50.67 (±4.05)	117.45 (±4.38)	31.83 (±0.96)	2.12
MV	7.56 (±0.18)	21.95 (±0.68)	1.39 (±0.002)	58.54 (±2.16)	0.21 (±0.01)	24.02 (±1.27)	0.49 (±0.01)	199.30 (±8.68)	47.44 (±3.55)	106.83 (±3.27)	30.02 (±1.88)	2.14
Mean value	7.34	20.02	1.18	48.78	0.69	14.45	0.49	163.45	44.62	106.60	30.40	2.10
CC	7.27 (±0.20)	19.31 (±0.87)	0.87 (±0.001)	5.24 (±0.41)	8.57 (±0.02)	5.47 (±0.02)	0.19 (±0.01)	110.49 (±3.94)	49.01 (±2.04)	87.10 (±4.03)	24.14 (±0.52)	0.98
CF	7.71 (±0.36)	17.91 (±1.05)	0.72 (±0.002)	26.12 (±1.52)	0.98 (±0.01)	12.42 (±0.54)	0.17 (±0.02)	98.75 (±2.05)	32.49 (±1.56)	76.38 (±2.40)	27.67 (±0.81)	0.88
CV	5.07 (±0.45)	22.42 (±1.51)	1.07 (±0.005)	13.88 (±0.79)	1.56 (±0.01)	51.33 (±1.88)	0.09 (±0.02)	132.37 (±5.91)	36.70 (±1.95)	92.39 (±4.57)	28.22 (±0.74)	0.76
Mean value	6.68	19.88	0.89	15.08	3.70	23.07	0.15	113.84	39.40	85.29	26.68	0.87
Background heavy metal contents in Beijing							0.12	29.8	18.7	57.5		
Background in China							0.20	90	35	100		

Values represent average±standard deviation

HF heavily polluted sites from ore treatment factory, HM heavily polluted sites from ore mining region, HT heavily polluted sites from tailing dam, MC moderately polluted sites from croplands, MF moderately polluted sites from planted forest, MV moderately polluted sites from vegetable fields from village X, CC clear sites from croplands, CF clear sites from planted forest, CV clear sites from vegetable fields from village Y

contents reaching values three to four times higher than the primary standard. The Zn concentration in all sites was higher than the background value in Beijing (57.5 mg/kg) but was not higher than the background value in China (100 mg/kg). The contents of the other heavy metals (Pb, Ni, and Co) were below the Beijing background (24.6, 26.8, and 7.5 mg/kg, data not shown). Although the Fe concentration was extremely high in all samples, it is not regarded as an indicator of metal pollution. The Nemerow composite index, a widely used pollution indicator and has been utilized by Wang (Wang et al. 2012b), showed that the soil samples from region H had the highest index values (Table 1), thereby indicating that the site was heavily polluted. Soil from region C was clear and used as a point of comparison. The surroundings of the mining area (MC, MF, and MV) were moderately polluted by a low diffusion of contaminants into the surrounding environment. However, considering the special characteristics of the Miyun Reservoir as a centralized urban drinking water source, stricter criteria (the primary standard) were selected to evaluate the heavy metal pollution level, and the results were much more serious than the other surveys which based on the secondary standard (Wang et al. 2012b).

Sequencing results and diversity indices

A total of 110,409 reads and 10,444 OTUs were obtained from nine samples through 468 MiSeq sequencing analysis. Each library contained 7959 to 22,327 reads, with different phylogenetic OTUs ranging from 724 to 1483. All rarefaction curves tended to approach the saturation plateau, indicating that the data volume of sequenced reads was reasonable, and the discovery of a high number of reads made a small contribution to the total number of OTUs. This rarefaction curve indicated the presence of a large variation in the total number of OTUs in different samples (Fig. 2). Compared with the contaminated soil (e.g., MC, MF, and MV), all samples from unpolluted soils have lower OTU density. The OTU densities of samples CF and CV were even lower than the average of most seriously contaminated soil samples (group H), and these samples' reads were not the least among the samples. In addition, the average OTUs of group C attained the lowest value (978), but its average reads (15,048) were significantly higher than those of group H (9851, Table 2).

Moreover, the calculation of the alpha diversity species richness (Chao), evenness (ACE), and Shannon index all confirmed the increasing diversity in the contaminated soil, especially in group H compared with group C (Table 2). The results of ANOVA analysis of the Shannon diversity indices also show significant differences with *P* value of 0.037, and the following post hoc pairwise test showed that groups H and M were significantly different from group C. The average Shannon diversity indices of the three groups were 6.15 (group H), 6.11 (group M), and 5.57 (group C). These results indicated that the Shannon diversity of polluted areas H and M was significantly higher than that of the unpolluted area C.

Fig. 2 Rarefaction curves of the OTU number at 97 % similarity boxplot for every samples

Taxonomy composition of the iron mining area

Sequences that could not be classified into any known group were assigned as unclassified. The bacterial OTUs were assigned into 35 different phyla, 251 families, or 381 genera. Six different phyla (*Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, and *Gemmatimonadetes*) out of the 35 total phylotypes were common to the nine libraries, which comprised more than 85 % of the total reads in every library. *Proteobacteria* was the most abundant group (Fig. 3a), comprising approximately 29.54 % (3086) of the OTUs and 39.95 %

Table 2 MiSeq sequencing results and diversity estimates for each sampling site

Sample	Sequencing results		Diversity estimates ^a		
	Total sequences	Total OTUs ^a	ACE	Chao	Shannon
HF	8877	894	1081±57	1091±78	5.97±0.03
HM	7957	1150	1407±65	1426±88	6.23±0.03
HT	12,721	1289	1492±48	1522±82	6.24±0.02
Mean value	9851	1111	1327	1346	6.15
MC	16,209	1340	1577±60	1603±89	6.16±0.02
MF	13,384	1323	1522±53	1534±73	6.08±0.03
MV	22,327	1483	1633±44	1670±71	6.08±0.02
Mean value	17,306	1382	1577	1602	6.11
CC	20,324	1327	1519±53	1535±75	5.98±0.02
CF	11,014	724	931±68	1029±138	5.43±0.03
CV	13,805	914	1157±69	1164±93	5.3±0.03
Mean value	15,048	978	1202	1242	5.57

ACE abundance-based coverage estimator, Chao Chao's species richness estimator, Shannon Shannon-Weiner Index

^a Species level, 97 % similarity threshold used to define operational taxonomic units (OTUs)

Fig. 3 Taxonomic distribution of soil samples. **a** Phylum distribution of all samples. **b** Class distribution of *Proteobacteria* phylum

(50,590) of the reads across all samples. *Acidobacteria*, the second most abundant phylum, (14.78 %, 1544 OTUs) consisted 18.70 % (23,682 reads) in all libraries. However, the proportion of *Proteobacteria* in the different samples showed high variability, in the range of 6.99 to 56.41 %. The read proportion reached lowest value in CV, which was different from that in group C (Fig. 3a). Meanwhile, *Acidobacteria* comprised the majority of the total reads in CV because of the strong acidic nature of the soil. The members from *Bacteroidetes* (13.61 %, 1422 OTUs), *Chloroflexi* (12.11 %, 1195 OTUs), *Actinobacteria* (10.45 %, 1091 OTUs), and *Gemmatimonadetes* (5.10 %, 533 OTUs) comprised 13.60 % (17,227 reads), 6.99 % (8848 reads), 8.11 % (10,266 reads), and 2.70 % (3424 reads) in all libraries, respectively. A much smaller fraction of the members of *Candidate_division_TM7* (3.11 %, 325 OTUs; 1.84 %, 2333 reads), *Verrucomicrobia* (2.85 %, 298 OTUs; 1.99 %, 2530 reads), *Cyanobacteria* (1.49 %, 156 OTUs; 1.65 %, 2085 reads), and *Nitrospirae* (1.37 %, 143 OTUs; 1.70 %, 2333 reads) were found. The average reads of the unclassified group account for 2.27 %, but this value fluctuated in different sites.

The members of *Alphaproteobacteria* dominated the *Proteobacteria* phylum and occupied 9.79 % (1023 OTUs) of all libraries but accounted for 22.44 % of the total reads (Fig. 3b). The subdivision of *Betaproteobacteria*, *Deltaproteobacteria*, and *Gammaproteobacteria* comprised 9.65, 4.34, and 9.34 % of the total reads, respectively. The *Epsilonproteobacteria* only contained two OTUs (0.019 %) and comprised a significantly small fraction (0.06 %) in the total reads.

Correlation between community structure and environmental factors

Mantel test was performed to analyze the environmental factors responsible in shaping soil bacterial community structure along a pollution gradient. Mantel test results showed that soil bacteria was significantly correlated with heavy metal

contents (mental $r=0.647$, $P<0.05$; Table S1), whereas soil properties had no obvious effect on bacteria (mental $r=0.364$; Table S1) indicating that heavy metal pollution played a more important role in shaping local bacteria community.

To further identify the major environmental variables controlling the soil microbial community structure, RDA was performed (Fig. 4). Parameters that could explain the variance of bacteria community structure were ranked by RDA results (Table 3). Results indicated that both heavy metal contents (Cd, Cr, Cu, Fe) ($P<0.05$) and soil pH ($P<0.01$) were identified to be the most influential environment factors to drive changes occurring in the community composition.

Comparison of bacterial community structure among groups

Principal component analysis (PCA) was used to identify the community structure differences under different pollution levels (Fig. 5). The resolutions of PCA between phylum, genus, and OTU levels were different. Except for CV, separating other locations at phylum level was different, indicating that communities in most locations shared similar phyla diversity. A much better discrimination was exhibited at the genus level, and group H grouped together, whereas the other groups had no clear boundary. The best resolution was obtained at the OTU level with all reads counted (Fig. 5c). At OTU level, PCA demonstrated that different soil locations from groups H and M were clustered, and the locations from group M (moderately polluted area) gathered together and were closer to group H (heavily polluted area communities) than to group C (unpolluted area). In addition, sites from group H gathered closer (similar) than the other groups, thereby indicating the higher number of similar community structures in heavily polluted areas. These results demonstrated that the microbial communities in the polluted areas had significant differences compared with the unpolluted area. This finding was consistent with the abovementioned alpha analysis results. However,

Fig. 4 RDA ordination plot showed the relationship between the bacteria community structure and soil properties (a) or heavy metal (b)

Table 3 Significance of environmental variables in explaining the bacteria community structure obtained from the RDA results

	r^2	P value
Soil		
Water contents	0.309	0.315
pH	0.952	0.001**
Organic matter	0.066	0.826
Available phosphorus	0.002	0.994
NH ₄ ⁺	0.038	0.899
NO ₃ ⁻	0.042	0.832
Heavy metal		
Cd	0.862	0.001**
Cr	0.674	0.011*
Cu	0.622	0.033*
Zn	0.375	0.232
Fe	0.576	0.046*

r^2 indicates the decision coefficient of environmental variables on the community structure

* $P < 0.05$; ** $P < 0.01$

the effect of soil properties on the community difference cannot be ignored. PCA at all levels demonstrated that the large gap between CV and the other regions in the unpolluted area was in accordance with the fact that the predominance of *Acidobacteria* in CV was closely related with the strong acidic nature of the site. Many research studies have confirmed that this phenomenon had close relationship with the strong acidic nature of soil (Jones et al. 2009).

Moreover, different from the PCA at genus and OTU level, PCA at phylum indicated that three groups have no clear boundary. To test the effect of heavy metal pollution on phyla structure, Matastats (White et al. 2009) were used to conduct differentially abundant features between three groups (Table 3). Matastats results demonstrated that three groups have significant difference on *Bacteroidetes*, groups H and C have significant difference on *Actinobacteria*, and groups have no obvious difference on the other phyla. Those findings raise the question of whether heavy metal pollution might have less affected on phyla structure.

Differences between the bacterial genera among different pollution levels

At genus level, the differences between the communities from various areas with different pollution levels were detected via a Venn diagram. A total of 341 genera were discovered, and 79.76 % of them belong to the shared genera (see Fig. S1). Group M alone contained more bacterial varieties (14 genera) than groups M (eight genera) and C (six genera), as shown in Fig. 6. An overlap between the detected genera of the three areas was also observed. The largest overlap was found between groups M/C (21 genera) followed by groups H/M (17) and groups H/C (3). Several new genera appeared in group H, indicating that these genera may thrive in heavily polluted areas. These genera included the following: *Betaproteobacteria* (36 reads, *Sulfuricella*, and 16 reads, *Candidatus Nitrotoga*), *Bacteroidetes* (4 reads, *Emticicia* and 3 reads, *Larkinella*), *Actinobacteria* (3 reads, *Modestobacter*), and *Cyanobacteria* (2 reads, *Arthronema*). However, their abundance was low, and their relative abundance was less than 0.05 %. Compared with group C, the five most abundant characteristic genera present in the polluted region (groups H and M), according to the read abundance, were as follows: *Alphaproteobacteria* (258 reads and 0.23 %, *Paracoccus*), *Cyanobacteria* (180 reads and 0.16 %, *Microcoleus*), *Cyanobacteria* (112 reads and 0.11 %, *Leptolyngbya*), *Betaproteobacteria* (98 reads and 0.09 %, *Thiobacillus*), and *Deltaproteobacteria* (68 reads and 0.06 %, *Peredibacter*).

Based on the relative abundance of the genera from Fig. 7, the genera with an average abundance of >1 % in at least one group were defined as dominant. Combined with the Venn diagram, most dominant genera belonged to the genera shared by the three

Fig. 5 Multivariate analysis based on information at phylum, genus, and OTU levels. **a** PCA based on phylum distribution, **b** PCA based on genus distribution, and **c** PCA based on total OTU level information

groups. Their relative abundance changed in sites with degrees of pollution, as shown in Fig. 8, except for the unknown genera. In group C, the dominant genera were *Sphingomonas*, *Bradyrhizobium*, *Blastocatella*, *Bryobacter*, *Rhodanobacter*, *Reyranelia*, *Rhizomicrobium*, and *Nitrospira*. *Sphingomonas* was the most abundant in the three groups, as seen in Fig. 6. The genera *Bradyrhizobium*, *Rhodanobacter*, *Reyranelia*, and

Rhizomicrobium significantly decreased with increasing pollution degree (from groups C to H), and their dominance decreased. Meanwhile, the relative abundances of *Blastocatella*, *Bryobacter*, and *Nitrospira* slightly decreased, but their dominance was maintained. The relative abundances of genera *Cellvibrio* and *Flexibacter* increased and reached more than 3 % in group M. However, their abundances significantly decreased in group H. The abundance of the following genera was still significantly higher in group H than in the other groups: *Steroidobacter*, *Massilia*, *Arthrobacter*, *Flavisolibacter*, and *Roseiflexus*, which belong to *Gamaproteobacteria*, *Betaproteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Chloroflexi*, respectively.

Fig. 6 Venn drawing showing the detected bacterial genera found in the three different groups of H, M, and C. Overlaps between the groups are indicated by the arrangement of the circles

Bacteria with potential resistance to heavy metals

To verify the potential resistance of abundant genera in heavy metal pollution area, Pearson correlation between the genera relative abundance and heavy metal contents was performed. *Arthrobacter*, which is atypical Cr^{6+} reducing bacteria (Asatiani et al. 2004), turned out to be closely related to Cr content in mining area ($R^2=0.726$, $P<0.05$) and became the dominant genera in group H. *Thiobacillus* was known as an important genus closely related to catalyzing ferrous ions and oxidizing the reducing inorganic sulfur (Dopson et al. 2003). In our samples, *Thiobacillus* had relative high abundance in the genera which only found in polluted area and its reads abundance in group H

Fig. 7 Bacterial distribution of the top 100 abundant genera among the nine samples. The bacterial phylogenetic tree was calculated using the neighbor-joining method. The heatmap plot depicts the relative abundance of each bacterial genera (variables clustering on the vertical axis) within each sample. The relative values for bacterial genera are indicated by *color intensity* with the legend indicated at the *top right corner*

is tripled in group M. Moreover, *Thiobacillus* relative abundance had close relationship with Fe content ($R^2=0.691$, $P<0.05$). In addition, most *Cyanobacteria* can produce extracellular polymeric substances, mainly polysaccharide, which could adsorb heavy metals dispersed in the environment (Pereira et al. 2009, 2011). In this study, *Cyanobacteria* was a rare group, but its total read abundances in groups H (795 reads) and M (1202 reads) were significantly higher than in group C (92 reads). *Microcoleus*, accounted for 8.75 % of the total *Cyanobacteria* species, displayed a strong correlation with Zn content ($R^2=0.789$, $P<0.05$). The genera *Rhodobacter* was also reported to be resistant against several metals and could show passive or active uptake of metals (Italiano et al. 2009; Mishra and Malik 2013). Based on the Venn diagram, *Rhodobacter* only appeared in polluted area. Besides, Pearson analysis showed that Zn content had a substantial correlation with *Rhodobacter* ($R^2=0.801$,

$P<0.01$). The population of resistant bacteria comprised less than 2 % of the indigenous microbial communities. However, the number of resistant bacteria can increase in contaminated soils, reaching 10 % or even 100 %, as the contaminants greatly change in terms of quantity and composition (Xu et al. 2012). Moreover, these bacteria serve a significant function in maintaining the ecological balance of the contaminated soil and could be used for the bioremediation of metal-polluted soil (Hayat et al. 2010; Kandeler et al. 2000). Numerous studies have reported a few naturally occurring microorganisms with high metal resistance in the rhizosphere that significantly increases the metal uptake in plants and reduces metal toxicity (Ma et al. 2011; Sessitsch et al. 2013).

Furthermore, several other genera were found similar to *Arthrobacter*, as the dominant genera in group H, and significantly related to heavy metals contents. *Steroidobacter* has

Fig. 8 Relative abundance of dominant genera in different groups. The genera with an average abundance of >1 % in at least one group were defined as dominant

been reported to utilize testosterone under both oxic and anoxic conditions. The genus *Massilia* was mainly isolated from the blood of an immunocompromised patient (Lindquist et al. 2003) and has been recently found in environmental samples, such as soil samples from China and in the drinking water distribution system in Seville (Spain) (Gallego et al. 2006; Wang et al. 2012a). In this study area, *Steroidobacter* was found to closely relate to Cd content ($R^2=0.729$, $P<0.05$), while *Massilia* had significant correlation with Cd and Cr contents ($R^2=0.718$, $P<0.05$; $R^2=0.671$, $P<0.05$). However, none of them were discovered to be resistant to metals in recent research studies. Their wide spread in heavily polluted area and close relation with heavy metals may be affected by

several other factors, such as the emission of beneficiation agents or wastewater, ore tailing accumulation, and availability of nutrients and oxygen, and water content in soil. These factors need further investigation.

Conclusions

In conclusion, the present study, using the high throughput Illumina sequencing method, provided a detailed picture of bacterial community variations on phylum and genus level under heavy metal pollution. Sequencing results and alpha diversity indices indicated that diversity of polluted areas, including ore treatment factory, ore mining region, tailing dam in H areas and three different fields (croplands, planted forest, and vegetable) in M areas which are only 1 km downstream of the tailing dam, was significantly higher than those of unpolluted areas. The total bacterial OTUs could be assigned into 35 different phyla, 251 families, or 381 genera. Among them, six different phyla (*Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, and *Gemmatimonadetes*) were common to the nine libraries, which comprised more than 85 % of the total reads in every library. PCA and Matastats results implied that heavy metal pollution has less affected on these phyla structure (Table 4). At genus level, some new genera appeared in polluted area, and the five most abundant genera were as follows: *Paracoccus*, *Microcoleus*, *Leptolyngbya*, *Thiobacillus*, and *Peredibacter*. Moreover, the relative abundance of dominant genera changed in sites with degrees of pollution, and several genera (e.g., *Steroidobacter*, *Massilia*, *Arthrobacter*, *Flavosolibacter*, and *Roseiflexus*) increased and became new dominant genera in the heavily metal-polluted area. The study found that some genera have close relation with heavy metals and may provide a new way to explore natural bioremedial genera. However, several genera which had not been reported for their resistance to metal

Table 4 Differentially abundant features at phyla level between three groups

Phylum	<i>P</i> value between groups HF, HM, and HT and groups MC, MF, and MV	<i>P</i> value between groups HF, HM, and HT and groups CC, CF, and CV	<i>P</i> value between groups MC, MF, and MV and groups CC, CF, and CV
Proteobacteria	0.317	0.159	0.909
Acidobacteria	0.347	0.609	0.329
Actinobacteria	0.060	0.028*	0.432
Bacteroidetes	0.011*	0.035*	0.006**
Chloroflexi	0.926	0.274	0.472
Gemmatimonadetes	0.655	0.946	0.853
Candidate_division_TM7	0.699	0.882	0.794
Verrucomicrobia	0.324	0.292	0.377
Cyanobacteria	0.803	0.349	0.180
Nitrospirae	0.419	0.324	0.782

P value means an individual measure of the false positive rate

* $P<0.05$; ** $P<0.01$

turned out to be dominant in polluted area and needed further study to explore their influencing factors.

Acknowledgments This work is supported by the National Natural Science Foundation of China (41273091), Beijing Nova Program (Z111106054511043).

Author contributions CH, YXS, and YX conceived and designed the experiments. YXS, CH, and YL performed the experiments. CH analyzed the data. CH and YXSYL contributed reagents/materials/analysis tools. CH and YXS wrote the paper.

References

- Anjos C, Magalhaes MCF, Abreu MM (2012) Metal (Al, Mn, Pb and Zn) soils extractable reagents for available fraction assessment: comparison using plants, and dry and moist soils from the Bracal abandoned lead mine area. *Portugal J Geochem Explor* 113:45–55
- Asatiani NV et al (2004) Effect of chromium (VI) action on *Arthrobacter oxydans*. *Curr Microbiol* 49:321–326
- Berg J, Brandt KK, Al-Soud WA, Holm PE, Hansen LH, Sørensen SJ, Nybroe O (2012) Selection for Cu-tolerant bacterial communities with altered composition, but unaltered richness, via long-term Cu exposure. *Appl Environ Microbiol* 78:7438–7446
- Bruce K, Jones T, Bezemer T, Thompson L, Ritchie D (2000) The effect of elevated atmospheric carbon dioxide levels on soil bacterial communities. *Glob Chang Biol* 6:427–434
- Cáliz J, Montserrat G, Martí E, Sierra J, Chung A, Morais P, Vila X (2013) Emerging resistant microbiota from an acidic soil exposed to toxicity of Cr, Cd and Pb is mainly influenced by the bioavailability of these metals. *J Soil Sediment* 13:413–428
- Dennis KL et al (2013) Adenomatous polyps are driven by microbe-instigated focal inflammation and are controlled by IL-10-producing T cells. *Cancer Res* 73:5905–5913
- Dopson M, Baker-Austin C, Koppineedi PR, Bond PL (2003) Growth in sulfidic mineral environments: metal resistance mechanisms in acidophilic micro-organisms. *Microbiology* 149:1959–1970
- Gallego V, Sánchez-Porro C, García MT, Ventosa A (2006) *Massilia aurea* sp. nov., isolated from drinking water. *Int J Syst Evol Microbiol* 56:2449–2453
- Giloteaux L, Sole A, Esteve I, Duran R (2011) Bacterial community composition characterization of a lead-contaminated *Microcoleus* sp consortium. *Environ Sci Pollut Res* 18:1147–1159
- Gomes NCM, Landi L, Smalla K, Nannipieri P, Brookes PC, Renella G (2010) Effects of Cd- and Zn-enriched sewage sludge on soil bacterial and fungal communities. *Ecotoxicol Environ Saf* 73:1255–1263
- Guo J et al (2010) Significant acidification in major Chinese croplands. *Science* 327:1008–1010
- Hayat R, Ali S, Amara U, Khalid R, Ahmed I (2010) Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann Microbiol* 60:579–598
- Huang X, Zhu Y, Ji H (2013) Distribution, speciation, and risk assessment of selected metals in the gold and iron mine soils of the catchment area of Miyun Reservoir, Beijing, China. *Environ Monit Assess* 185:8525–8545
- Italiano F, Buccolieri A, Giotta L, Agostiano A, Valli L, Milano F, Trotta M (2009) Response of the carotenoidless mutant *Rhodobacter sphaeroides* growing cells to cobalt and nickel exposure. *Int Biodeterior Biodegrad* 63:948–957
- Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* 3:442–453
- Kandeler E, Tschirko D, Bruce K, Stemmer M, Hobbs PJ, Bardgett RD, Amelung W (2000) Structure and function of the soil microbial community in microhabitats of a heavy metal polluted soil. *Biol Fertil Soils* 32:390–400
- Keeney D, Nelson D (1982) Nitrogen—inorganic forms. In: Page AL (eds) *Methods of soil analysis: part 2—chemical and microbiological properties*. Soil Sci Soc Am, Madison. pp: 643–698
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120
- Lindquist D, Murrill D, Burran WP, Winans G, Janda JM, Probert W (2003) Characteristics of *Massilia timonae* and *Massilia timonae*-like isolates from human patients, with an emended description of the species. *J Clin Microbiol* 41:192–196
- Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. *ISME J* 5:169
- Ma Y, Prasad M, Rajkumar M, Freitas H (2011) Plant growth promoting rhizobacteria and endophytes accelerate phytoremediation of metal-liferous soils. *Biotechnol Adv* 29:248–258
- Marmolejo-Rodríguez AJ, Prego R, Meyer-Willerer A, Shumilin E, Cobelo-García A (2007) Total and labile metals in surface sediments of the tropical river-estuary system of Marabasco (Pacific coast of Mexico): influence of an iron mine. *Mar Pollut Bull* 55:459–468
- Mishra A, Malik A (2013) Recent advances in microbial metal bioaccumulation. *Crit Rev Environ Sci Technol* 43:1162–1222
- Norman RJ, Edberg JC, Stucki JW (1985) Determination of nitrate in soil extracts by dual-wavelength ultraviolet spectrophotometry. *Soil Sci Soc Am J* 49:1182–1185
- Oberauer L, Zachow C, Lackner S, Högenauer C, Smolle K-H, Berg G (2013) The ignored diversity: complex bacterial communities in intensive care units revealed by 16S pyrosequencing. *Scientific reports* 3
- Pereira S, Zille A, Micheletti E, Moradas Ferreira P, De Philippis R, Tamagnini P (2009) Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiol Rev* 33:917–941
- Pereira S, Micheletti E, Zille A, Santos A, Moradas-Ferreira P, Tamagnini P, De Philippis R (2011) Using extracellular polymeric substances (EPS)-producing cyanobacteria for the bioremediation of heavy metals: do cations compete for the EPS functional groups and also accumulate inside the cell? *Microbiology* 157:451–458
- Petrić I, Bru D, Udiković-Kolić N, Hršak D, Philippot L, Martin-Laurent F (2011) Evidence for shifts in the structure and abundance of the microbial community in a long-term PCB-contaminated soil under bioremediation. *J Hazard Mater* 195:254–260
- Qin F, Ji H, Li Q, Guo X, Tang L, Feng J (2014) Evaluation of trace elements and identification of pollution sources in particle size fractions of soil from iron ore areas along the Chao River. *J Geochem Explor* 138:33–49
- Rodríguez L, Ruiz E, Alonso-Azcárate J, Rincón J (2009) Heavy metal distribution and chemical speciation in tailings and soils around a Pb-Zn mine in Spain. *J Environ Manag* 90:1106–1116
- Ros M, Pascual J, Moreno J, Hernandez MT, Garcia C (2009) Evaluation of microbial community activity, abundance and structure in a semi-arid soil under cadmium pollution at laboratory level. *Water Air Soil Pollut* 203:229–242
- Schloss PD et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for

- describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541
- Sessitsch A, Kuffner M, Kidd P, Vangronsveld J, Wenzel WW, Fallmann K, Puschenreiter M (2013) The role of plant-associated bacteria in the mobilization and phytoextraction of trace elements in contaminated soils. *Soil Biol Biochem* 60:182–194
- Uroz S, Ioannidis P, Lenglé J, Cébron A, Morin E, Buée M, Martin F (2013) Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation. *PLoS ONE* 8:e55929
- Wang J, Zhang J, Pang H, Zhang Y, Li Y, Fan J (2012a) *Massilia flava* sp. nov., isolated from soil. *Int J Syst Evol Microbiol* 62:580–585
- Wang Q, Wang R, Tian C, Yu Y, Zhang Y, Dai J (2012b) Using microbial community functioning as the complementary environmental condition indicator: a case study of an iron deposit tailing area. *Eur J Soil Biol* 51:22–29
- White JR, Nagarajan N, Pop M (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5:e1000352
- Xu ZY, Tang M, Chen H, Ban YH, Zhang HH (2012) Microbial community structure in the rhizosphere of *Sophora viciifolia* grown at a lead and zinc mine of northwest China. *Sci Total Environ* 435–436:453–464
- Zhang HB, Yang MX, Shi W, Zheng Y, Sha T, Zhao ZW (2007) Bacterial diversity in mine tailings compared by cultivation and cultivation-independent methods and their resistance to lead and cadmium. *Microb Ecol* 54:705–712
- Zhao J et al (2014) Responses of bacterial communities in arable soils in a rice-wheat cropping system to different fertilizer regimes and sampling times. *PLoS ONE* 9:e85301