Identification of microbial profiles in heavy metal contaminated soil from full length 16S rRNA reads sequenced by PacBio system

Moonsuk Hur[1] and Soo-Je Park[2]*

[1]Microorganism Resources Division, National Institute of Biological Resources, 42 Hwangyeong-ro, Incheon 22689, Republic of Korea

[2]Department of Biology, Jeju National University, 102 Jejudaehak-ro, Jeju 63243, Republic of Korea

***Correspondence:** Tel.: +82-64-754-3524; Fax: +82-64-756-3541; E-mail: sjpark@jejunu.ac.kr

Running title: microbial community analysis by PacBio

Abstract: Heavy metal pollution is a serious environmental problem as it adversely affects crop production and human activity. In addition, the microbial community structure and composition are altered in heavy metal contaminated soils. In this study, using full-length 16S rRNA gene sequences obtained by a PacBio RS II system, we determined the microbial diversity and community structure in heavy metal contaminated soil. Further, we investigated the microbial distribution, inferred their putative functional traits, and analyzed the environmental effects on the microbial compositions. The soil samples selected in this study were heavily and continuously contaminated with various heavy metals due to closed mines. We found that certain microorganisms (e.g. sulfur or iron oxidizers) might play an important role in the biogeochemical cycle. Using PICRUSt analysis, we predicted KEGG functional categories from abundances of microbial communities and revealed a high proportion belonging to transport, energy metabolism, and xenobiotics degradation in the studied sites. In addition, through full-length analysis, *Conexibacter*-like sequences, commonly identified by environmental metagenomics among the rare biosphere, were detected. In addition to microbial composition, we confirmed that environmental factors, including heavy metals, affect the microbial communities. Unexpectedly, among these environmental parameters, electrical conductivity (EC) might have more importance than other factors for a community description analysis.

Keywords: Heavy metal, soil, PacBio, 16S rRNA gene, mine

1. Introduction

As of 2017, 4,677 out of 5,544 mines (for coal, metal, and non-metal) in Korea have been abandoned owing to environmental concerns and/or economic factors. In particular, in the case of metal mines, most mines (about 95%, 2,084/2,184) located throughout Korea are in disuse due to soil pollution caused by waste rock and mine tailings with AMD (acid mine drainage) [1]. The mine tailings dams and/or AMD contaminated by heavy metals, including iron and cadmium, have led to serious environmental problems related to crops and public health through contamination of the water supply and food chain, respectively. Although these contaminated areas have been reclaimed by a long-term plan under the Ministry of Environments in Korea, there are rising concerns regarding soil contamination by heavy metals [2, 3].

Heavy metal pollution by anthropogenic activities affects microbial activities and community structures in terrestrial environments. There is a specific relationship between microorganisms and minerals in these extremely toxic environments. Soil microorganisms can affect plant growth and increase the accumulation of heavy metals in plants (i.e. phytoremediation) [4, 5]. Therefore, analysis of microbial communities might provide fundamental information for phytoextraction improvement [6]. Most previous studies have focused on identifying microbial community structures in polluted soils or isolating useful microorganisms for the removal of heavy metals [see review in 7]. Moreover, based on cultivation and metagenome approaches, we can easily understand and expand our knowledge of new microbe-mineral interactions [8-10]. Recently, microbial communities from various environments have been extensively sequenced and analyzed by next

generation sequencing (NGS) platforms using partial regions of the 16S rRNA gene (e.g. V1-V3, V3-V4, V4-V5, or V4-V6) [5, 11-17]. Additionally, numerous unrecognized Bacteria and Archaea have been identified by NGS technology, and the results can serve as fundamental taxonomic information in (meta)genome analysis with putatively deduced functional characterizations [18, 19].

Although previous studies using specific regions of the 16S rRNA gene have shown that taxonomic assignments are highly sensitive, advanced high throughput revealed unintended missing-classification with less accuracy, especially at the genus or species taxonomic level [20]. This might be because most of the naturally existing microorganisms have not yet been cultivated and identified.

Recently, a single molecular real-time (SMRT) DNA sequencing system has been developed by Pacific Biosciences (PacBio) and applied to microbial community and (meta)genome analyses [21-24]. This system is able to generate raw reads more than 10 kb long with a low error rate [25]. Although PacBio platform is less attractive than other short-read platform such as Illumina and Ion-Torrent due to higher cost, it has been applied to microbial community analysis without primer bias and with high quality [26-28]. Based on the advantages of PacBio platform, we also expect to obtain more accurate results from massive deep full-length reads generated by the PacBio system for the microbial community in heavy metal contaminated soil (from disused mines) and identify the microorganism(s) in the rare biosphere [29]. As mentioned above, most mines in Korea are in disuse and crops and human health might be affected by contaminated water and food chain. Additionally, very little is known about the composition and structure of the bacterial community in these contaminated areas in Korea. Therefore, in the present study, we characterized and compared the microbial communities in heavy-metal contaminated sites in the Republic of Korea. The main objectives of the present study are to demonstrate the practical application of the PacBio system in microbial community structure analysis and to investigate the microbial diversity and structure along with the effect of environmental parameters including heavy metals, on them.

2. Materials and methods

2.1. Sample collection and characterization

For the microbial community analysis, we selected three sites in regions with heavy metal contamination, namely, Hwaseong (H), Daegu (D), and Bonghwa (B), in the Republic of Korea (Supplementary Table S1). From each site, duplicate soil samples were collected from the surface (2–5 cm depth; F) and subsurface (60–75 cm depth; B) at five randomly selected spots (1-5) located at a distance of 1 m from each other for site-study replication [30]. Alphanumeric codes were assigned to each sample; for example, BF1 is the first surface sample collected at Bonghwa. The samples were transferred to sterile plastic tubes or bags and stored at -80 °C until microbial community structure analysis. The sampling depths were determined based on the recovery depth for heavy-metal contaminated soil from abandoned mines in the Republic of Korea [31]. Inductively coupled plasma atomic emission spectroscopy (ICP-OES; PerkinElmer Optima 7300 DV) was used to determine the concentrations of the heavy metals [copper (Cu), lead (Pb), arsenic (As), zinc (Zn) and cadmium (Cd)]. Before the analysis, the soil samples were sequentially filtered through 2.0 mm and 0.15 mm sieves, dried, and then digested with concentrated nitric acid. Calibration was performed using a distilled water blank and standard solution [31, 32]. The concentrations of the selected heavy metals in each sample were determined in duplicate. Total nitrogen (TN), total carbon (TC), and organic matter (OM) were measured using a CNS analyzer (US/Vario Max CN, Elementar Analysensysteme, GmbH, Hanau, Germany). pH/electrical conductivity (EC) was determined using a pH/EC meter (Sevenmulti S40, Mettler Toledo, Switzerland)[33]. *In situ* temperature was determined using a thermometer (Waterproof Digital Thermocouple Thermometer, A1.T9234). The particle size distribution of the soil samples was plotted on a particle size distribution curve (semi-log graph; Supplementary Table S1) to assess soil texture [33, 34] (Supplementary Table S1).

2.2. PCR amplification of bacterial 16S rRNA gene and sequencing by PacBio system

Total genomic DNA (gDNA) was extracted from each prepared sample using a Power Soil DNA kit (Mo Bio Laboratories, Solanan Beach, CA, USA). The quality and quantity of the extracted gDNA were determined using a DS-11 Plus Spectrophotometer (DeNovix, Inc., Wilmington, DE, USA) and by performing electrophoresis on a 1.0% (w/v) agarose gel, respectively.

For full-length bacterial 16S rRNA gene amplification, we performed PCR using the following mixture: 10 μL of 2× Dr. MAX Master Mix Solution (Doctor Protein Corp., Seoul, Republic of Korea), 1 μM of 27F (5′-AGRGTTYGATYMTGGCTCAG-3′) and 1492R (5′-RGYTACCTTGTTACGACTT-3′) primer set (final concentration), and ~10 ng of environmental DNA as a template. Where necessary, the template was diluted using 0.1× TRIS-EDTA buffer to decrease the concentration of PCR inhibitors. Cycling was performed with an initial denaturation at 95°C for 7 min, followed by 30 cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s, and a final extension at 72°C for 10 min. Amplification products were separated by agarose gel electrophoresis and purified using a PCR Clean-up Kit (LaboPass, Cosmo Genetech, Seoul, Republic of Korea). SMRTbell adapters were then ligated onto the purified PCR products and the libraries were sequenced by Pacific Biosciences using P6-C4 chemistry on a PacBio RS II SMRT DNA sequencing system (Pacific Biosciences, Menlo Park, CA, USA).

2.3. Sequence data and statistics analyses

Raw sequences were initially processed through the PacBio SMRT portal. All purified sequencing data were then processed using Mothur version 1.39.5 [24, 35, 36]. All sequences were aligned against a SILVA-based reference alignment and were classified against the greengenes (version gg_13_8_99) reference database using a negative Bayesian classifier implemented within Mothur [36]. Diversity indices [Shannon and Simpson indices, and the Chao1 nonparametric richness index] and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering, nonmetric multidimensional scaling (NMDS), and analysis of similarities (ANOSIM) were determined using the Mothur package [37]. A dissimilarity level of 3% was used in further investigations including diversity indices. Habitat specialization was calculated as described by Liu et al [38] using Levin's niche breadth index (B) [39], which assigns low niche breadth values to specialists (< 1.5) and higher values to generalists (> 3).

A Mann–Whitney U test was performed to compare the diversity indices between the microbial communities from each analyzed site. To estimate the relationship between microbial community and environmental factors, a canonical correspondence analysis (CCA) was performed using the VEGAN package [40-42]. The 32 highest-ranked genera (> 5% proportion in each sample) were assessed using CCA. Manual selection of environmental parameters through application of a Monte Carlo permutation test (999 random permutations) was conducted to determine their statistical significance ($p < 0.05$). The resulting ordination biplot approximated the weight of each OTU with respect to each environmental variable, represented as arrows. The length of the arrows indicated the relative importance of environmental factors that explained variations in the microbial communities. All figures were generated by R packages [43] and/or Origin Pro 2018 (OriginLab, Northampton, MA, USA)[44]. Further, for estimation of the interactions between microbial compositions and environmental parameters, Spearman's rank correlation coefficient (ϱ) was calculated simultaneously using the 'rcorr' function with the Hmics package [45] in R to analyze the significance of the correlation. A high correlation coefficient ($|\varrho| \geq 0.7$) with a p-value < 0.05 between microbial compositions and environmental parameters was visualized via Cytoscape (v.3.6.1)[46]. To estimate correlation values between environmental parameters, we used Spearman's rank correlation analysis [47]. Functional profiles of microbial communities were predicted by Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUST) [13]. Although functional predictions were assigned up to all KEGG Orthology (KO) numbers, only Xenobiotics biodegradation and metabolism, Energy metabolism, and Membrane transport were selected as contamination-related categories for analysis simplification and clarity. The categories related to "human disease" or "eukaryotes" were excluded due to poor relevance to environmental samples.

The full 16S rRNA gene sequences recovered in this study have been deposited in the DDBJ/ENA/GenBank Sequence Read Archive (SRA) under the accession number SRP137440.

3. Results

3.1. Environmental data

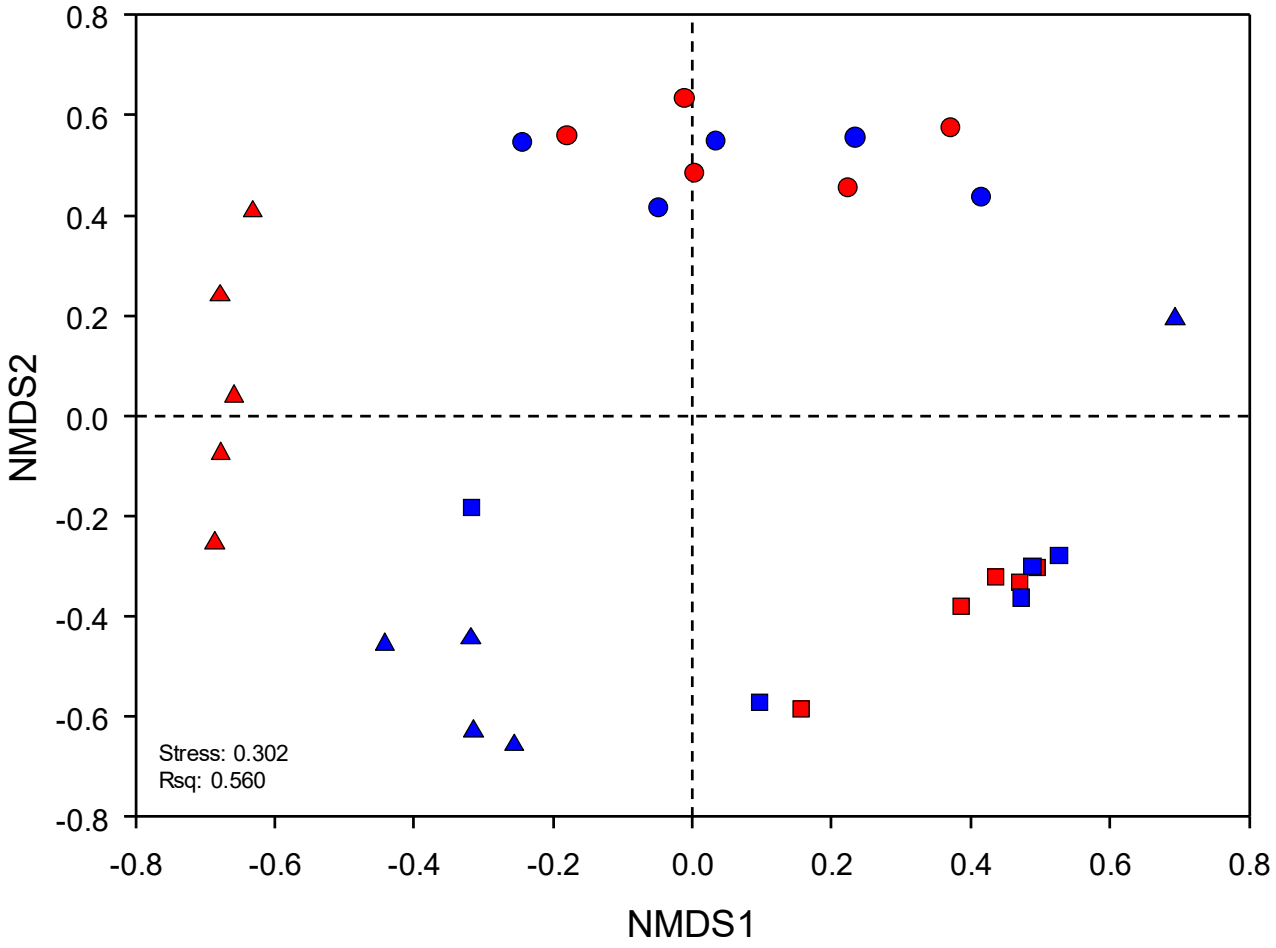
The geographical and physiochemical parameters of 30 soil samples from the three sites are summarized in Table S1. A pH range of 2.8-6.9 (average pH 4.9 ± 1.3) indicated acidic to near neutral pH. The subsurface (pH 4.3 ± 1.5) was slightly more acidic than the surface (pH 5.4 ± 1.2). The temperature reached an average of 23 ± 3.0 °C. The EC was in a range of 0.15-20.2 ± 5.5 ds/m. Except for the Bonghwa samples (11.1 ds/m), most samples had extremely low EC values (0.86 ds/m). In addition, we found significant differences in other environmental factors (e.g. TN, OM, and temperature) between sampling site and depth. The average concentrations (ppm) of major heavy metals (Cu, Pb, As, Zn, and Cd) were estimated as 322.6 ± 320.4, 1175.6 ± 1770.9, 5373.7 ± 9638.3, 3884.6 ± 7491.6, and 28.4 ± 51.7 ppm, respectively (Table S1). In particular, the concentration of As was higher than that of other heavy metals. However, in Hwaseong, the heavy metal concentrations were extremely low and As was undetectable. Although soil texture was similar, the difference in the soil composition (e.g. sand, silt, and clay) was significant between samples from different sites (Table S1).

3.2. General statistics for 16S rRNA gene amplification

A total of 122,702 sequences were obtained and analyzed from the 148,594 raw sequences after quality filtering supplied in the Mothur program (Table 1). The analyzed data for each sample ranged in size from 1,490 sequences for sample HF4 to 8,284 for sample DB5.

The whole analyzed sequences were classified into different well determined, candidate, and unclassified phyla (Fig. S1a). From all samples, the phylum Proteobacteria was identified as the highly accounted (i.e., dominant) group (comprised about 41% of total purified sequences), followed by Acidobacteria, Actinobacteria, Chloroflexi, AD3, Firmicutes, Planctomycetes, Nitrospirae, and Gemmatimonadetes, comprising 19.7%, 10.2%, 7.5%, 5.4%, 4.8%, 3.6%, 2.8%, and 1.3% of all sequences, respectively.

The results of NMDS showed that microbial communities were apparently clustered by sampling site (Rsq: 0.560 and Stress: 0.302), excluding two samples (DF1 and BF1) (Fig. 1). The stress value for NMDS indicated poor representation for microbial communities; however, the ANOSIM results indicated that the microbial community compositions (based on the 97% similarity level) were significantly different between most analyzed samples (Table S2). Further, with the exception of the DF1 sample, UPGMA showed that samples from the three distantly located sites formed a separated cluster (Fig. S2). In the case of Bonghwa, we only found a tendency to split into two parts depending on the sampling depth. In addition, samples from Hwaseong formed a separate and distant cluster outside the other samples, indicating a more distinguished relationship.



181

182 **Fig. 1.** Microbial community profiles of the samples from Hwaseong (circles), Daegu (squares), and Bonghwa
183 (triangles) based on nometric multidimensional scaling (NMDS) using the Mothur package. Distance matrix
184 was calculated using the Yue and Clayton theta supplied in the Mothur package. Operational taxonomic units
185 (OTUs) were determined based on 3% dissimilarity of nucleotide sequences. Blue and red denote surface and
186 subsurface, respectively.

Table 1. An overview of the soil samples and estimates of 16S rRNA gene sequence diversity and phylotype coverage of the PacBio data. Diversity was estimated using operational taxonomic units (OTUs). Diversity indices and richness estimators were calculated using the Mothur package (the mother project; http://www.mothur.org).

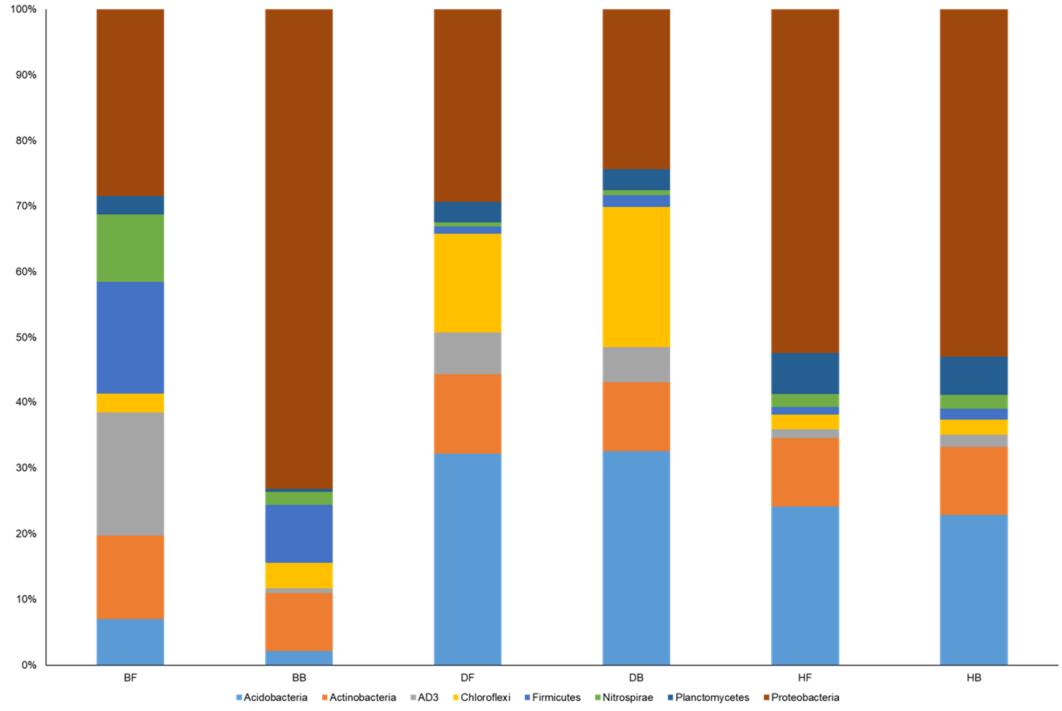
Sample *	Grouping name	No. of analyzed reads	OTU	Chao	Shannon	Simpson	Good's coverage	Sampling site (GPS, altitude)
BB1	BB	2738	388	1629.03	3.79	8.62	0.89	Bongwha (129°3'17.9"E, 36°51'45.6"N, 630m)
BB2		2931	544	2092.42	4.65	24.99	0.87	
BB3		2240	355	1814.62	3.98	13.51	0.88	
BB4		2404	600	4411.03	4.14	5.74	0.78	
BB5		7769	1492	7657.55	4.66	9.77	0.84	
BF1	BF	5677	486	2820.18	2.73	3.67	0.93	
BF2		4606	508	2393.16	4.01	12.06	0.92	
BF3		4420	547	2592.36	4.07	12.18	0.91	
BF4		3969	743	2787.00	5.25	53.47	0.87	
BF5		3784	601	3295.08	4.64	24.51	0.88	
DB1	DB	4107	792	2427.76	5.03	16.91	0.87	Daegu (128°40'18.6"E, 35°46'52.5"N, 243m)
DB2		2149	424	1513.36	4.30	8.74	0.86	
DB3		5250	1036	4241.84	5.09	13.70	0.86	
DB4		2762	551	1756.51	4.38	7.55	0.86	
DB5		8284	952	3784.81	4.99	40.10	0.92	
DF1	DF	5962	699	3354.82	4.17	13.56	0.91	
DF2		5489	712	2880.12	4.15	8.73	0.91	
DF3		4019	711	3157.89	4.39	8.37	0.87	
DF4		3880	702	2803.69	4.13	5.05	0.87	
DF5		3521	478	1580.64	4.69	36.54	0.91	
HB1	HB	4712	2425	9038.20	7.57	627.07	0.61	Hwaseong (126°55'47.6"E, 37°13'09.1"N, 105m)
HB2		4634	2167	7723.74	7.30	355.46	0.65	
HB3		4489	2308	8079.12	7.52	557.24	0.61	
HB4		3825	2057	6954.94	7.50	521.60	0.59	
HB5		3360	1954	7019.94	7.60	726.93	0.55	
hf1	HF	3142	1673	5914.55	7.30	539.23	0.60	
hf2		1662	1026	4310.03	7.09	348.38	0.51	
hf3		3720	2097	8645.22	7.57	607.90	0.56	
hf4		1490	1032	3987.65	7.42	861.26	0.45	
hf5		5707	2801	10771.69	7.62	661.09	0.63	

*Samples were named as follows: sampling site, depth [B: subsurface (below 60-75 cm), F: surface (below 2-15 cm)] and sampling replicates.

The analyzed sequences were affiliated with 42 phyla and the eight most abundant phyla (> 5% proportion of all reads in each sample) were designated as major phyla: Acidobacteria, Actinobacteria, AD3, Chloroflexi, Firmicutes, Nitrospirae, Planctomycetes, and Proteobacteria (Fig. 2a and Fig. S1). These phyla contributed from 87.4% to 99.7% (average 95.0%) of all sequences. In particular, among the phyla, we found that Proteobacteria accounted for the highest proportion, especially in the BB sample, which had the greatest proportion (about 70%). On the other hand, the proportion of Acidobacteria of the BB sample was less than that of the other sites.

At the genus taxonomic level, the aforementioned sequences were assigned to 720 genera including the unclassified group with a high taxonomic level (e.g. class, order, or family). For further analysis, we analyzed only 352 clearly assigned genera excluding the sequences assigned as "unclassified", although we might have lost many sequences (Fig. 2b) Finally, we selected 32 genera from each sample based on > 5% of total sequences for the following analyses such as CCA and interaction. Moreover, from a combined sample (as a grouping sample based on the depth of the sampling site), we found only 16 genera designated as dominant, with relatively high abundances (Fig. S1b). In this analysis, eight genera, *Leptospirillum*, *Rhodoplanes*, *Thiobacillus*, *Acidithiobacillus*, *Sulfobacillus*, *Conexibacter*, *Candidatus* Solibacter, and *Rhodovastum*, had the highest relative abundance, accounting for about 80% of total bacterial abundance from all samples. The genera *Acidothiobacillus* and *Sulfobacillus* were only identified in two samples (i.e. BF and BB) and in BF samples, respectively, as major taxa (Fig. 2b). Iron-oxidizing gram-positive acidophiles were identified in Daegu and Hwaseong samples as a minor taxon (less than 0.5% of total bacterial abundance); however, this iron-oxidizer was detected in BF with comparatively high abundance (5% of total bacterial abundance). Unexpectedly, the genus *Halothiobacillus*, isolated from marine environments including hydrothermal vents and considered as an obligately chemolithoautotrophic and sulfur oxidizer, was identified in BB.

(a)



(b)

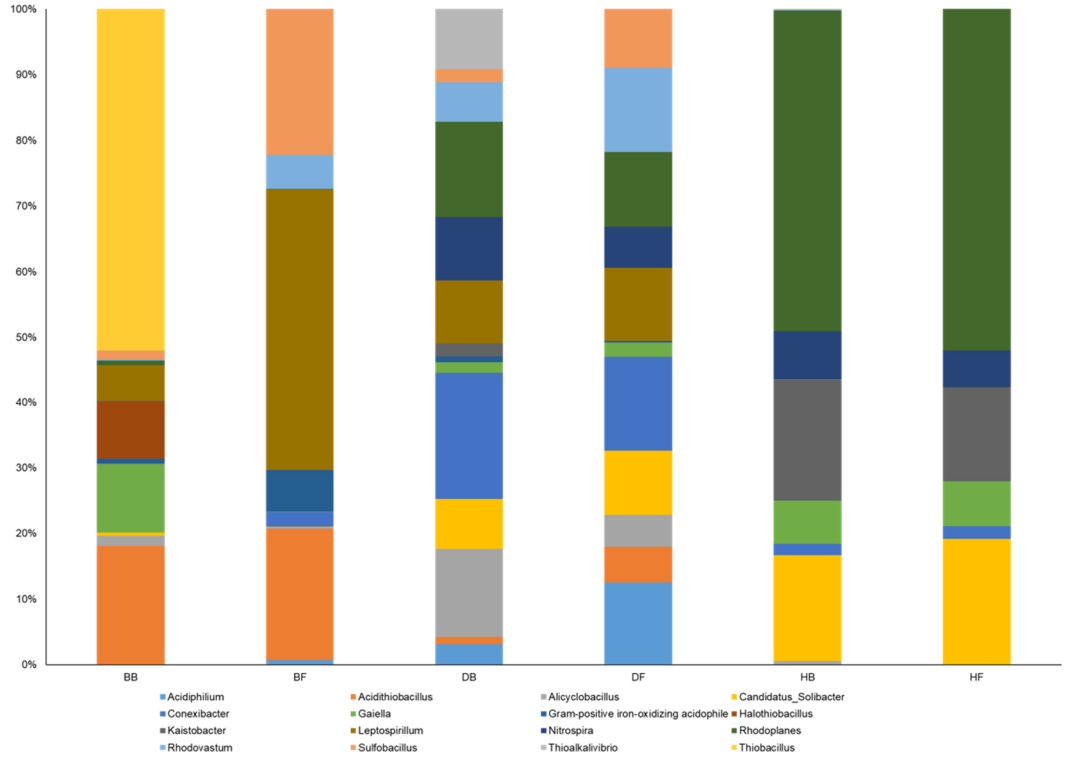


Fig. 2. Relative abundance of the most abundant 8 bacterial phyla (a) and 16 genera (b) in the contaminated soil samples. At the genus level, only those with proportions above 5% of the total reads in each sample are shown.

Subsequently, to determine generalist and specialist genera selected in this study (n=32), we calculated niche breadth (*B*) by Levin's index [38, 39] from all analyzed sites and depths. From this analysis, only 16 genera were determined as generalists (*B*>3, *Alicyclobacillus*, *Aquicella*, *Clostridium*, *Conexibacter*, *Nitrospira*, and *Rhodanobacter*) and specialists: (*B*<1.5, *Gallionella*, gram-positive iron-oxidizing_acidophile, *Halothiobacillus*, *Leptolyngbya*, *Leptospirillum*, *Rhodoferax*, *Sulfobacillus*, *Thioalkalivibrio*, *Thiobacillus*, and YNPFFP6 classified into the family *Sulfobacillaceae*).

In Table 1, the metrics for alpha diversity (number of OTUs, Chao1 richness, Shannon evenness, Simpson diversity, and Goods' coverage) of the microbial community are summarized. OTUs were defined as sequences with 97% sequence similarity. While variations in diversity indices were observed, there were no significant differences between surface and subsurface for intra-group and inter-group samples (estimated by Mann-Whitney *U* test). However, diversity indices of the H (Hwaseong) samples combined from surface and subsurface showed higher values than those of other samples (Fig. S4). In particular, the Simpson diversity index showed extremely remarkable differences (p = 0.0003 estimated by Mann-Whitney *U* test). On the other hand, the diversity indices of the samples from Bonghwa and Daegu showed similar ranges (p > 0.11 estimated by Mann-Whitney *U* test)(Fig. S4).

3.3. Environmental factors significantly affecting microbial community

To explain the variation in the microbial communities (i.e., the selected 32 genera) between sampling sites, CCA and association network analysis were performed. These analyses facilitated the investigation of the effect of environmental parameters including soil components. The CCA results indicated that with the exception of the pH and clay, most environmental variables exhibited significant effects on the microbial communities (p < 0.012 based on the 999 permutations)(Fig. S3). On the x axis (CA1), OM, TC, EC, and C/N showed highly positive positions (0.93-0.97), while temperature, pH, and silt showed highly positive positions (0.90-0.99) on the y axis (CA2). To estimate the interactions between the microorganism(s) and environmental

parameters, an associated network analysis was performed based on Spearman's rank correlation coefficient (ϱ)(Fig. 3). From the network analysis, 20 genera and 13 environmental factors were obtained based on the criteria described in the Materials and Methods. Further, a total of 76 correlations (41 and 35 for positive and negative correlations, respectively) were established between environmental factors and genera (Fig. 3). Based on this analysis, we found that As and EC, among the environmental factors, and *Rhodoplanes* genus, in the microbial composition, have higher interactions than others. Unexpectedly, four genera (*Leptospirillum*, *Sulfobacillus*, *Acidithiobacillus*, and gram-positive iron-oxidizer) have positive interactions with EC, but not pH and it was observed that pH exhibited a negative relationship only with Cu. These genera have been categorized as an acidophilic bacterial group [48]. In addition, four genera (*Thiobacillus*, *Halothiobacillus*, *Rhodovastum*, and *Acidiphilum*) and three environmental variables (TC, C/N, and OM) showed only an intra-group relationship between themselves (Fig. 3).

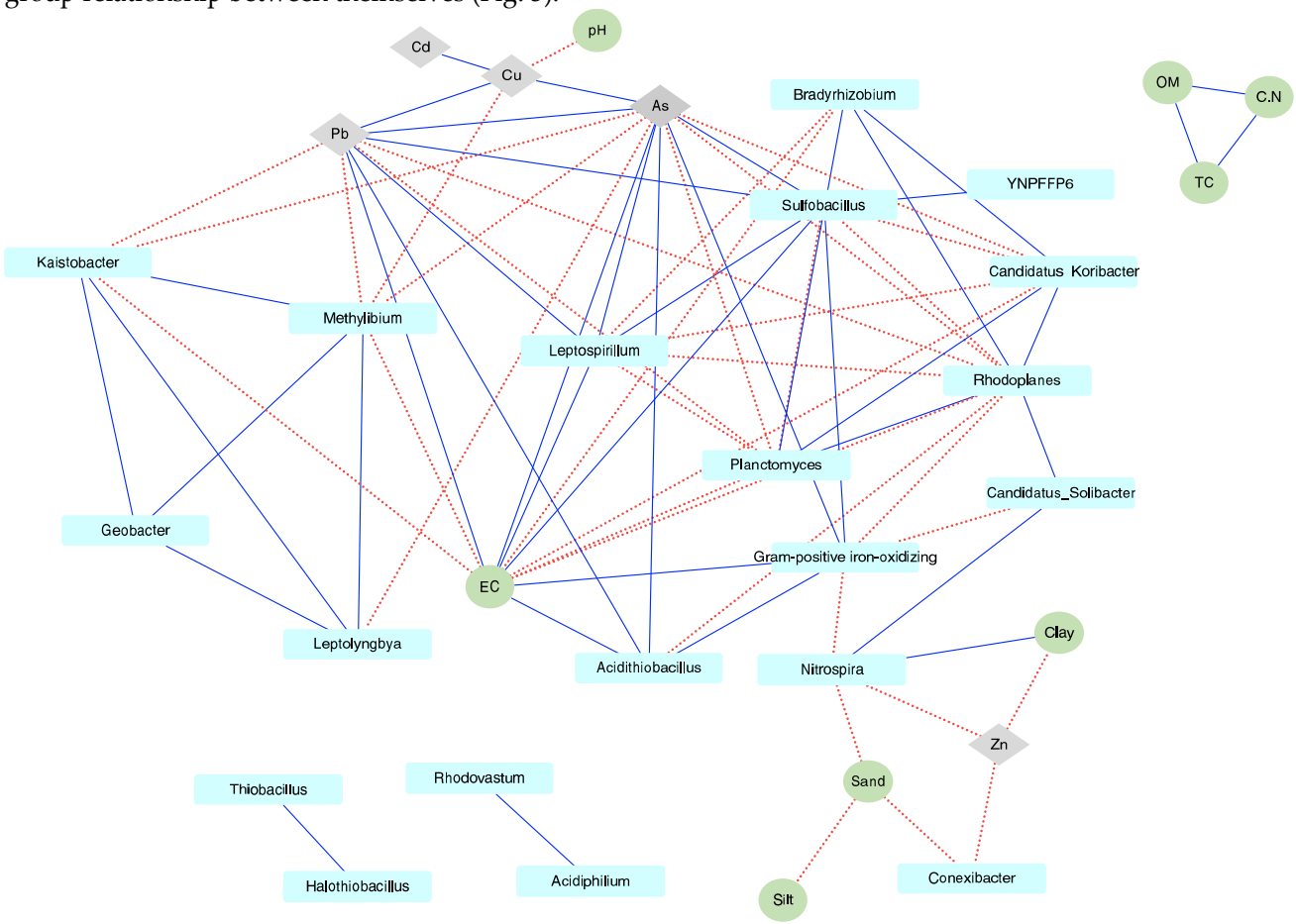


Fig. 3. Profile clustering Cytoscape network visualization of the 32 most abundant genera with environmental correlation (solid line, positive correlation; dotted line, negative correlation) estimated by Spearman's rank analysis. Analyzed genus, heavy metals, and other environmental factors are marked with cyan-squares, gray-diamonds, and green-circles, respectively.

3.4. Predictive functional profiles from microbial communities

Although we tried to understand microbial functional traits by analyzing abundances and distribution of microorganisms in each site, it is difficult to predict their ecological roles. Therefore, to observe and compare between study sites, putative functional profiles from 16S rRNA gene sequences were predicted via KEGG pathways of PICRUSt analysis. Among all KEGG pathways inferred by PICRUSt, the abundance of the KEGG category related to "Transport" was estimated as the highest (11.8-13.3% of total KEGG categories), after exclusion of the poorly relevant categories (see Materials and Methods). The proportions for Xenobiotics biodegradation and Energy metabolism were estimated to be 6.0% and 7.5%, respectively. Notably, between the studied sites, variation in some functional traits of third-tier KO was observed such as in degradation of

DDT, aminobenzoate, and nitrotoluene (Fig. S5). However, the proportions of other third-tier functional categories were similar between sampling sites.

4. Discussion

Over the past 10 years, NGS technology has been introduced and developed, and has played a central role in the field of microbial ecology for sequencing the small-subunit ribosomal RNA genes (e.g. 16S rRNA gene). Traditional strategies, i.e. clone-based sequencing and culture-dependent methods, for microbial community analysis have recently been extensively replaced by NGS platforms such as pyrosequencing (of Roche 454), paired-end sequencing (HiSeq or MiSeq of Illumina), and an ion semiconductor (IonTorrent). These platforms are now commonly used to generate hundreds of thousands of read sequences from various environmental samples from an amplicon of the variable region(s) of the 16S rRNA gene [49, 50]. However, it is well known that the amplicon approach might have an amplification bias that occurs by variable region selection associated with the primer choice [51]. The bias can affect the results for taxonomic classification and diversity indices [52]. Contrastingly, full-length 16S rRNA gene sequences enable the clear identification of taxonomy and phylogeny [53], despite amplification bias. In addition, some studies have reported molecular analysis of microbial community structures with no replication and randomization for experimental design [30, 54, 55]. Nevertheless, owing to the sequencing cost or analysis techniques that need computing ability, most previous studies have described the microbial community structures from various environments, including heavy-metal contaminated soils, using partial-length of the 16S rRNA gene sequenced by other NGS platforms. In response, this study sought to analyze microbial community structure and estimate the relationship between microbial compositions and environmental parameters in heavy-metal contaminated soils using a full-length bacterial 16S rRNA gene sequenced by the PacBio RS II platform with plentiful replicates for each studied site.

A number of studies have reported that environmental parameters can affect the microbial community structure and chemical processing (i.e. nutrient cycles) of soil. In the present study, we found that environmental factors influence microbial compositions (Fig. 3 and Fig. S3). In particular, heavy metals have significantly higher impact than other factors on microbial community structure. Contrary to the NMDS results, the CCA results showed that the microbial communities from Hwaseong are closely formed (Fig. 1 and Fig. S3). In addition, we identified that EC (ds m^{-1}) has higher positive correlation with heavy metals than with other environmental factors (Fig. 3 and Fig. S3). Although organic matter may influence the retention of heavy metals [56, 57], we found no significant correlation with heavy metals (Table S1). In fact, soil EC has been known as an indirect indicator of soil health. It affects yields and suitability of crop, and plant nutrient availability, as well as key soil processes such as the emission of greenhouse gases (e.g. nitrogen oxides and methane) [58, 59]. Some studies proposed that soil EC is a major factor that contributes to bacterial community and activity [60, 61]. Moreover, Jordán et al [62] contended that EC related to metal bioavailability shows higher correlation with heavy metal distribution. The associated network analysis conducted in the present study revealed that EC had a direct-positive relationship only with Pb and As; however, EC was also found to have an indirect-positive relationship with Cu and Cd. In addition, the aforementioned heavy metals and EC exhibited more relationships including negative association with microorganisms than other environmental factors. These findings suggest that EC is a major factor associated with the activities of microbial communities (see Fig. 3). Additionally, we observed that only 4 genera (*Acidothiobacillus*, *Letptosprillum*, *Sulfobacillus*, and gram-positive iron-oxidizing_acidophile) exhibited a positive relationship with EC. Although the members of these genera are known as acidophiles, our analysis did not reveal any relationship between pH and these microorganisms. Therefore, this finding possibly indicates that compared to other environmental factors, EC has a stronger relationship with the acidophiles. Collectively, these results indicate that metal concentrations solely do not reflect toxicity or environmental pollution. Measurement and analysis of EC and heavy metal concentration can be used to effectively assess contamination risk [62] and predict indigenous microbial activity [61].

With respect to the microbial composition, Proteobacteria, Acidobacteria, and Actinobacteria phyla were identified as predominant groups (Fig. 2a and Fig. S1a). Notably, these phyla are commonly found in terrestrial environments [63-65]. However, although Acidobacteria are considered to be the most dominant in soil, comprising an average of about 20% of total soil bacteria, the ecological roles of the phylum is still poorly

understood, with other phyla such as Verrucomicrobia, Chloroflexi, Planctomycetes, and Gemmatimonadetes [references in 66].

At the genus level, 16 genera were identified in this study as generalist or specialist, which were dominant. Among the genera, ten (*Gallionella*, gram-positive iron-oxidizing_acidophile, *Halothiobacillus*, *Leptolyngbya*, *Leptospirillum*, *Rhodoferax*, *Sulfobacillus*, *Thioalkalivibrio*, *Thiobacillus*, and YNPFFP6 classified into the family *Sulfobacillaceae*) were identified as specialist for their habitat. This suggests the possibility for development of a microbial indicator of contamination by heavy metals in soil. Interestingly, with the exception of samples from Bonghwa, microbial compositions of surface and subsurface samples were similar. This might be because the physiochemical characterizations of surface and subsurface soil at Daegu and Hwaseong were similar (Supplementary Table S1), indicating that environmental factors affect microbial community structures. In addition, some microorganisms in BF and BB were identified as unique compared to other samples. For example, *Leptospirillum* and *Thiobacillus* were only identified as predominant in BF and BB, respectively (Fig. 2b). *Leptosprillum* has been identified as an iron oxidizer and acidophile under oxic conditions [7]. They are classified into four groups [67]; in particular group IV is reported to be capable of hydrogen and iron oxidation [67]. Moreover, *Leptosprillum* spp. also significantly contribute to AMD processing and bioleaching [7, 68]. Since first being described in 1904, *Thiobacillus* spp. are known as an autotrophic sulfur-oxidizer using reduced sulfur compounds as an energy source [69]. Although some *Halothiobacillus* species are described as halotolerant and reclassified from the genus *Thiobacillus* [70], sequences related to the genus *Halothiobacillus* were not detected in other sites and samples. For this reason, *Halothiobacillus* identification might be because it is phylogenetically related to the genus *Thiobacillus*.

On the other hand, the genus *Rhodoplanes*, described as a primarily phototrophic purple nonsulfur bacterium [71], was only found in samples from Daegu and Hwaseong with high abundance (ranging from 8% to 15% of total bacterial abundance). In addition, we found sequences related to genera that are key players in geochemical cycling for nitrogen (i.e. *Nitrospira*) or sulfur (*Sulfobacillus, Desulfosporosinus*). Unexpectedly, in addition to microbial community analysis, PICRUSt inferred relatively little functional variations between sampling sites. Only a few functional categories were observed to exhibit variation between sampling sites. The "transport" category was identified as a dominant category in all studied sites with similar proportion. However, it is possible that our PICRUSt results support the relationship between microbial communities and environmental factors. Moreover, we hypothesize that the microorganism(s) are adapted to their local habitat and microbial community stability is affected during a long period of heavy-metal contamination, which acts as a selective pressure [72, 73]. Based on the present analysis, we observed that EC might have greater effect on microbial community structure compared to heavy metals or organic matter.

Although we successfully characterized the microbial distribution and their predicted functional traits from each sampling site, clear ecological roles and whole-metagenome analysis are lacking. Nevertheless, using full-length sequencing application, this study provides accurate information about microbial community structures and microbial interactions in heavy metal contaminated soil in Korea. In addition, the findings might also enable identification of the rare biosphere [28] using full-length sequencing. For example, in DB and DF samples (Fig. 2b), sequences classified into the genus *Conexibacter* were detected. Till date, in the genus *Conexibacter*, only two species have been isolated as novel representatives of the deep branch of the phylum Actinobacteria from soil [74, 75]. Moreover, *Conexibacter* spp. have been recently recognized in environmental metagenomics [76, 77] and massive sequencing [78], including Oxford Nanopore MinION [79]. Finally, the findings of the present study provide valuable insight into the decrease in microbial activity and diversity variations caused heavy metal pollution as determined by full-length sequencing.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Author Contributions: Conceived and designed the experiments: M.H. and S.J.P. Performed the experiments and analyzed the data: M.H. and S.J.P. Wrote the paper: S.J.P. All authors read and approved the final manuscript.

Ethics Statement: None required

Acknowledgements: This work was supported by grants from the National Institute of Biological Resources funded by the Ministry of Environment (no. NIBR201501119) and the National Research Foundation of Korea (no. 2018R1C1B6006762).

References

1. MIRECO. *2016 Yearbook of Mireco (Mine Reclamation Corp. In Korea) Statistics.* Wonju, Republic of Korea: Mine Reclamation Corp., 2017.
2. NIER. *Studies on the Assessment of Biological Exposure to Heavy Metals.* Incheon, Republic of Korea Press: National Institute of Environmental Research (NIER), 1996.
3. MOE. *Soil Environment Conservation Act in Korea*. Seoul, Republic of Korea.: Ministry of Environment, 2016.
4. Gałązka, Anna, and Rafał Gałązka. *Phytoremediation of Polycyclic Aromatic Hydrocarbons in Soils Artificially Polluted Using Plant-Associated-Endophytic Bacteria and Dactylis Glomerata as the Bioremediation Plant*. Vol. 64, 2015.
5. Galazka, A., J. Grzadziel, R. Galazka, A. Ukalska-Jaruga, J. Strzelecka, and B. Smreczak. "Genetic and Functional Diversity of Bacterial Microbiome in Soils with Long Term Impacts of Petroleum Hydrocarbons." *Front Microbiol* 9 (2018): 1923.
6. Wood, J. L., C. Zhang, E. R. Mathews, C. Tang, and A. E. Franks. "Microbial Community Dynamics in the Rhizosphere of a Cadmium Hyper-Accumulator." *Sci Rep* 6 (2016): 36067.
7. Baker, B. J., and J. F. Banfield. "Microbial Communities in Acid Mine Drainage." *FEMS Microbiol Ecol* 44, no. 2 (2003): 139-52.
8. Edwards, K. J., P. L. Bond, T. M. Gihring, and J. F. Banfield. "An Archaeal Iron-Oxidizing Extreme Acidophile Important in Acid Mine Drainage." *Science* 287, no. 5459 (2000): 1796-9.
9. Baker, B. J., G. W. Tyson, R. I. Webb, J. Flanagan, P. Hugenholtz, E. E. Allen, and J. F. Banfield. "Lineages of Acidophilic Archaea Revealed by Community Genomic Analysis." *Science* 314, no. 5807 (2006): 1933-5.
10. Baker, B. J., L. R. Comolli, G. J. Dick, L. J. Hauser, D. Hyatt, B. D. Dill, M. L. Land, N. C. Verberkmoes, R. L. Hettich, and J. F. Banfield. "Enigmatic, Ultrasmall, Uncultivated Archaea." *Proc Natl Acad Sci U S A* 107, no. 19 (2010): 8806-11.
11. Salipante, S. J., T. Kawashima, C. Rosenthal, D. R. Hoogestraat, L. A. Cummings, D. J. Sengupta, T. T. Harkins, B. T. Cookson, and N. G. Hoffman. "Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16s Rrna-Based Bacterial Community Profiling." *Appl Environ Microbiol* 80, no. 24 (2014): 7583-91.

12. Nelson, M. C., H. G. Morrison, J. Benjamino, S. L. Grim, and J. Graf. "Analysis, Optimization and Verification of Illumina-Generated 16s Rrna Gene Amplicon Surveys." *PLoS One* 9, no. 4 (2014): e94249.
13. Langille, M. G., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko, and C. Huttenhower. "Predictive Functional Profiling of Microbial Communities Using 16s Rrna Marker Gene Sequences." *Nat Biotechnol* 31, no. 9 (2013): 814-21.
14. Kozich, J. J., S. L. Westcott, N. T. Baxter, S. K. Highlander, and P. D. Schloss. "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, no. 17 (2013): 5112-20.
15. Choi, H., H. W. Koh, H. Kim, J. C. Chae, and S. J. Park. "Microbial Community Composition in the Marine Sediments of Jeju Island: Next-Generation Sequencing Surveys." *J Microbiol Biotechnol* 26, no. 5 (2016): 883-90.
16. Currie, A. R., K. Tait, H. Parry, B. de Francisco-Mora, N. Hicks, A. M. Osborn, S. Widdicombe, and H. Stahl. "Marine Microbial Gene Abundance and Community Composition in Response to Ocean Acidification and Elevated Temperature in Two Contrasting Coastal Marine Sediments." *Front Microbiol* 8 (2017): 1599.
17. Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, and R. Knight. "Ultra-High-Throughput Microbial Community Analysis on the Illumina Hiseq and Miseq Platforms." *ISME J* 6, no. 8 (2012): 1621-4.
18. Lasken, R. S. "Genomic Sequencing of Uncultured Microorganisms from Single Cells." *Nat Rev Microbiol* 10, no. 9 (2012): 631-40.
19. Lasken, R. S., and J. S. McLean. "Recent Advances in Genomic DNA Sequencing of Microbial Species from Single Cells." *Nat Rev Genet* 15, no. 9 (2014): 577-84.
20. Soergel, D. A., N. Dey, R. Knight, and S. E. Brenner. "Selection of Primers for Optimal Taxonomic Classification of Environmental 16s Rrna Gene Sequences." *ISME J* 6, no. 7 (2012): 1440-4.
21. Koren, S., and A. M. Phillippy. "One Chromosome, One Contig: Complete Microbial Genomes from Long-Read Sequencing and Assembly." *Curr Opin Microbiol* 23 (2015): 110-20.

22. Mosher, J. J., B. Bowman, E. L. Bernberg, O. Shevchenko, J. Kan, J. Korlach, and L. A. Kaplan. "Improved Performance of the Pacbio Smrt Technology for 16s Rdna Sequencing." *J Microbiol Methods* 104 (2014): 59-60.
23. Myer, P. R., M. Kim, H. C. Freetly, and T. P. Smith. "Metagenomic and near Full-Length 16s Rrna Sequence Data in Support of the Phylogenetic Analysis of the Rumen Bacterial Community in Steers." *Data Brief* 8 (2016): 1048-53.
24. Wagner, J., P. Coupland, H. P. Browne, T. D. Lawley, S. C. Francis, and J. Parkhill. "Evaluation of Pacbio Sequencing for Full-Length Bacterial 16s Rrna Gene Classification." *BMC Microbiol* 16, no. 1 (2016): 274.
25. Roberts, R. J., M. O. Carneiro, and M. C. Schatz. "The Advantages of Smrt Sequencing." *Genome Biol* 14, no. 7 (2013): 405.
26. Singer, E., B. Bushnell, D. Coleman-Derr, B. Bowman, R. M. Bowers, A. Levy, E. A. Gies, J. F. Cheng, A. Copeland, H. P. Klenk, S. J. Hallam, P. Hugenholtz, S. G. Tringe, and T. Woyke. "High-Resolution Phylogenetic Microbial Community Profiling." *ISME J* 10, no. 8 (2016): 2020-32.
27. Pootakham, W., W. Mhuantong, T. Yoocha, L. Putchim, C. Sonthirod, C. Naktang, N. Thongtham, and S. Tangphatsornruang. "High Resolution Profiling of Coral-Associated Bacterial Communities Using Full-Length 16s Rrna Sequence Data from Pacbio Smrt Sequencing System." *Sci Rep* 7, no. 1 (2017): 2774.
28. Karst, S. M., M. S. Dueholm, S. J. McIlroy, R. H. Kirkegaard, P. H. Nielsen, and M. Albertsen. "Retrieval of a Million High-Quality, Full-Length Microbial 16s and 18s Rrna Gene Sequences without Primer Bias." *Nat Biotechnol* 36, no. 2 (2018): 190-95.
29. Pedros-Alio, C. "The Rare Bacterial Biosphere." *Ann Rev Mar Sci* 4 (2012): 449-66.
30. Webster, R. "Replicate and Randomize, or Lie." *Environ Microbiol* 19, no. 1 (2017): 25-28.
31. MOE. "Official Test Methods of Soil Contamination.". Seoul, Republic of Korea.: Ministry of Environment, 2017.
32. Koh, H. W., M. Hur, M. S. Kang, Y. B. Ku, R. Ghai, and S. J. Park. "Physiological and Genomic Insights into the Lifestyle of Arsenite-Oxidizing *Herminiimonas Arsenitoxidans*." *Sci Rep* 7, no. 1 (2017): 15007.
33. Burt, Rebecca, States United, Service Natural Resources Conservation, Center National Soil Survey, and Laboratory National Soil Survey. *Soil Survey Laboratory Methods Manual*. [Washington, D.C.]: United States Dept. of Agriculture, Natural Resources Conservation Service, 2004.

34. Jahn, R., H. P. Blume, V. B. Asio, O. Spaargaren, and P. Schad. *Guidelines for Soil Description, 4th Edition*. Rome: FAO, 2006.
35. Myer, P. R., M. Kim, H. C. Freetly, and T. P. Smith. "Evaluation of 16s Rrna Amplicon Sequencing Using Two Next-Generation Sequencing Technologies for Phylogenetic Analysis of the Rumen Bacterial Community in Steers." *J Microbiol Methods* 127 (2016): 132-40.
36. Schloss, P. D., M. L. Jenior, C. C. Koumpouras, S. L. Westcott, and S. K. Highlander. "Sequencing 16s Rrna Gene Fragments Using the Pacbio Smrt DNA Sequencing System." *PeerJ* 4 (2016): e1869.
37. Schloss, P. D., D. Gevers, and S. L. Westcott. "Reducing the Effects of Pcr Amplification and Sequencing Artifacts on 16s Rrna-Based Studies." *PLoS One* 6, no. 12 (2011): e27310.
38. Liu, M., Z. Yu, X. Yu, Y. Xue, B. Huang, and J. Yang. "Invasion by Cordgrass Increases Microbial Diversity and Alters Community Composition in a Mangrove Nature Reserve." *Front Microbiol* 8 (2017): 2503.
39. Feinsinger, Peter, E. Eugene Spears, and Robert W. Poole. "A Simple Measure of Niche Breadth." *Ecology* 62, no. 1 (1981): 27-32.
40. Vegan: Community Ecology Package. R package version 2.4-5.
41. Buttigieg, P. L., and A. Ramette. "A Guide to Statistical Analysis in Microbial Ecology: A Community-Focused, Living Review of Multivariate Data Analyses." *FEMS Microbiol Ecol* 90, no. 3 (2014): 543-50.
42. Rani, S., H. W. Koh, S. K. Rhee, H. Fujitani, and S. J. Park. "Detection and Diversity of the Nitrite Oxidoreductase Alpha Subunit (*Nxra*) Gene of *Nitrospina* in Marine Sediments." *Microb Ecol* 73, no. 1 (2017): 111-22.
43. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
44. Moberly, J. G., M. T. Bernards, and K. V. Waynant. "Key Features and Updates for Origin 2018." *J Cheminform* 10, no. 1 (2018): 5.
45. Hmisc: Harrell Miscellaneous. R package version 4.1-1.
46. Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Res* 13, no. 11 (2003): 2498-504.
47. Bonett, Douglas G., and Thomas A. Wright. "Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations." 65, no. 1 (2000): 23-28.

48. González-Toril, Elena, Felipe Gómez, Moustafa Malki, and Ricardo Amils. "The Isolation and Study of Acidophilic Microorganisms." In *Methods in Microbiology*, 471-510: Academic Press, 2006.
49. Hamady, M., and R. Knight. "Microbial Community Profiling for Human Microbiome Projects: Tools, Techniques, and Challenges." *Genome Res* 19, no. 7 (2009): 1141-52.
50. Baker, G. C., J. J. Smith, and D. A. Cowan. "Review and Re-Analysis of Domain-Specific 16s Primers." *J Microbiol Methods* 55, no. 3 (2003): 541-55.
51. Forney, L. J., X. Zhou, and C. J. Brown. "Molecular Microbial Ecology: Land of the One-Eyed King." *Curr Opin Microbiol* 7, no. 3 (2004): 210-20.
52. Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland. "A Detailed Analysis of 16s Ribosomal Rna Gene Segments for the Diagnosis of Pathogenic Bacteria." *J Microbiol Methods* 69, no. 2 (2007): 330-9.
53. Vinje, H., K. H. Liland, T. Almoy, and L. Snipen. "Comparing K-Mer Based Methods for Improved Classification of 16s Sequences." *BMC Bioinformatics* 16 (2015): 205.
54. van Dorst, J., A. Bissett, A. S. Palmer, M. Brown, I. Snape, J. S. Stark, B. Raymond, J. McKinlay, M. Ji, T. Winsley, and B. C. Ferrari. "Community Fingerprinting in a Sequencing World." *FEMS Microbiol Ecol* 89, no. 2 (2014): 316-30.
55. Prosser, J. I. "Replicate or Lie." *Environ Microbiol* 12, no. 7 (2010): 1806-10.
56. Ferraz, M. C. M. Alvim, and J. C. N. Lourençlo. "The Influence of Organic Matter Content of Contaminated Soils on the Leaching Rate of Heavy Metals." *Environmental Progress* 19, no. 1 (2000): 53-58.
57. Uchimiya, Minori, Desmond I. Bannon, and Lynda H. Wartelle. "Retention of Heavy Metals by Carboxyl Functional Groups of Biochars in Small Arms Range Soil." *J. Agric. Food Chem.* 60, no. 7 (2012): 1798-809.
58. Adviento-Borbe, M. A., J. W. Doran, R. A. Drijber, and A. Dobermann. "Soil Electrical Conductivity and Water Content Affect Nitrous Oxide and Carbon Dioxide Emissions in Intensively Managed Soils." *J Environ Qual* 35, no. 6 (2006): 1999-2010.
59. Krauss, S., W. H. Schnitzler, J. Grassmann, and M. Woitke. "The Influence of Different Electrical Conductivity Values in a Simplified Recirculating Soilless System on Inner and Outer Fruit Quality Characteristics of Tomato." *J Agric Food Chem* 54, no. 2 (2006): 441-8.
60. Kim, J. M., A. S. Roh, S. C. Choi, E. J. Kim, M. T. Choi, B. K. Ahn, S. K. Kim, Y. H. Lee, J. H. Joa, S. S. Kang, S. A. Lee, J. H. Ahn, J. Song, and H. Y. Weon. "Soil Ph and Electrical Conductivity Are Key Edaphic Factors Shaping Bacterial Communities of Greenhouse Soils in Korea." *J Microbiol* 54, no. 12 (2016): 838-45.

61. Atekwana, Estella A., Eliot A. Atekwana, D. Dale Werkema, Jonathan P. Allen, Laura A. Smart, Joseph W. Duris, Daniel P. Cassidy, William A. Sauck, and Silvia Rossbach. "Evidence for Microbial Enhanced Electrical Conductivity in Hydrocarbon-Contaminated Sediments." 31, no. 23 (2004).
62. Jordán, Manuel M., Beatriz Rincón-Mora, and María Belén Almendro-Candel. "Heavy Metal Distribution and Electrical Conductivity Measurements in Biosolid Pellets." *J Soils Sediments.* 16, no. 4 (2016): 1176-82.
63. Janssen, P. H. "Identifying the Dominant Soil Bacterial Taxa in Libraries of 16s Rrna and 16s Rrna Genes." *Appl Environ Microbiol* 72, no. 3 (2006): 1719-28.
64. Zeng, Q., Y. Dong, and S. An. "Bacterial Community Responses to Soils Along a Latitudinal and Vegetation Gradient on the Loess Plateau, China." *PLoS One* 11, no. 4 (2016): e0152894.
65. Fierer, N., M. A. Bradford, and R. B. Jackson. "Toward an Ecological Classification of Soil Bacteria." *Ecology* 88, no. 6 (2007): 1354-64.
66. Greening, C., C. R. Carere, R. Rushton-Green, L. K. Harold, K. Hards, M. C. Taylor, S. E. Morales, M. B. Stott, and G. M. Cook. "Persistence of the Dominant Soil Phylum *Acidobacteria* by Trace Gas Scavenging." *Proc Natl Acad Sci U S A* 112, no. 33 (2015): 10497-502.
67. Goltsman, D. S., M. Dasari, B. C. Thomas, M. B. Shah, N. C. VerBerkmoes, R. L. Hettich, and J. F. Banfield. "New Group in the *Leptospirillum* Clade: Cultivation-Independent Community Genomics, Proteomics, and Transcriptomics of the New Species "*Leptospirillum* Group Iv Uba Bs"." *Appl Environ Microbiol* 79, no. 17 (2013): 5384-93.
68. Diaby, N., B. Dold, H. R. Pfeifer, C. Holliger, D. B. Johnson, and K. B. Hallberg. "Microbial Communities in a Porphyry Copper Tailings Impoundment and Their Impact on the Geochemical Dynamics of the Mine Waste." *Environ Microbiol* 9, no. 2 (2007): 298-307.
69. Robertson, Lesley A., and J. Gijs Kuenen. "The Genus *Thiobacillus*." In *The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses*, edited by Martin Dworkin, Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer and Erko Stackebrandt, 812-27. New York, NY: Springer New York, 2006.
70. Kelly, D P, and A P Wood. "Reclassification of Some Species of *Thiobacillus* to the Newly Designated Genera *Acidithiobacillus* Gen. Nov., *Halothiobacillus* Gen. Nov. And *Thermithiobacillus* Gen. Nov." *Int. J. Syst. Evol. Microbiol.* 50, no. 2 (2000): 511-16.
71. Hiraishi, A. , and Y Ueda. "*Rhodoplanes* Gen. Nov., a New Genus of Phototrophic Bacteria Including *Rhodopseudomonas Rosea* as *Rhodoplanes Roseus* Comb. Nov. And *Rhodoplanes Elegans* Sp. Nov." *Int.. Syst. Bacteriol.* 44, no. 4 (1994): 665-73.

570 72. Gómez, Pedro, Steve Paterson, Luc De Meester, Xuan Liu, Luca Lenzi, M. D. Sharma,
571 Kerensa McElroy, and Angus Buckling. "Local Adaptation of a Bacterium Is as Important as
572 Its Presence in Structuring a Natural Microbial Community." *Nature Communications* 7
573 (2016): 12453.
574 73. Brooks, A. N., S. Turkarslan, K. D. Beer, F. Y. Lo, and N. S. Baliga. "Adaptation of Cells to
575 New Environments." *Wiley Interdiscip Rev Syst Biol Med* 3, no. 5 (2011): 544-61.
576 74. Seki, Tamae, Atsuko Matsumoto, Risa Shimada, Yuki Inahashi, Satoshi Ōmura, and Yōko
577 Takahashi. "*Conexibacter Arvalis* Sp. Nov., Isolated from a Cultivated Field Soil Sample."
578 *Int J Syst Evol Microbiol.* 62, no. 10 (2012): 2400-04.
579 75. Monciardini, Paolo, Linda Cavaletti, Peter Schumann, Manfred Rohde, and Stefano Donadio.
580 "*Conexibacter Woesei* Gen. Nov., Sp. Nov., a Novel Representative of a Deep Evolutionary
581 Line of Descent within the Class Actinobacteria." *Int J Syst Evol Microbiol.* 53, no. 2 (2003):
582 569-76.
583 76. Deng, Jie, Yunfu Gu, Jin Zhang, Kai Xue, Yujia Qin, Mengting Yuan, Huaqun Yin, Zhili He,
584 Liyou Wu, Edward A. G. Schuur, James M. Tiedje, and Jizhong Zhou. "Shifts of Tundra
585 Bacterial and Archaeal Communities Along a Permafrost Thaw Gradient in Alaska."
586 *Molecular Ecology* 24, no. 1 (2015): 222-34.
587 77. Krivushin, Kirill, Fyodor Kondrashov, Lyubov Shmakova, Mariya Tutukina, Lada
588 Petrovskaya, and Elizaveta Rivkina. "Two Metagenomes from Late Pleistocene Northeast
589 Siberian Permafrost." *Genome Announc.* 3, no. 1 (2015): e01380-14.
590 78. Liu, Xiaojiao, Shuting Zhang, Qipeng Jiang, Yani Bai, Guihua Shen, Shili Li, and Wei Ding.
591 "Using Community Analysis to Explore Bacterial Indicators for Disease Suppression of
592 Tobacco Bacterial Wilt." *Sci Rep* 6 (2016): 36773.
593 79. Kerkhof, Lee J., Kevin P. Dillon, Max M. Häggblom, and Lora R. McGuinness. "Profiling
594 Bacterial Communities by Minion Sequencing of Ribosomal Operons." *Microbiome* 5 (2017):
595 116.

596