ORIGINAL PAPER

Hydrocarbon pollution does not influence bacterial diversity as much as geographic location: a Korean case study

D.-H. Cho · K.-H. Baek · R. Ramanan · C.-Y. Ahn · K.-H. Ahn · B.-D. Yoon · H.-M. Oh · H.-S. Kim

Received: 23 September 2013/Revised: 8 March 2014/Accepted: 18 March 2014/Published online: 29 April 2014 © Islamic Azad University (IAU) 2014

Abstract Bacterial diversity of hydrocarbon-contaminated sites in various regions of Korea was investigated to ascertain the influence of hydrocarbon pollution on bacterial diversity using terminal restriction fragment length polymorphism (T-RFLP) and differential gel gradient electrophoresis. Thirty-two hydrocarbon-contaminated soil samples were collected from seven different geographical locations in Korea. A dendrogram of T-RFLP profiles for the bacterial community structure in soil samples using Ward's method with Jaccard distance showed that samples from the same location clustered together. Principal components analysis (PCA) and selforganizing maps (SOM) of terminal restriction fragments were also used to characterize the associations among samples. PCA and SOM results also showed that soil bacterial communities were classified according to locations, but not by hydrocarbon pollution level. Moreover, correlation analyses prove a direct correlation between bacterial diversity and meteorological parameters, whereas no significant correlation was observed with

Electronic supplementary material The online version of this article (doi:10.1007/s13762-014-0578-z) contains supplementary material, which is available to authorized users.

D.-H. Cho · R. Ramanan · C.-Y. Ahn · K.-H. Ahn · B.-D. Yoon · H.-M. Oh · H.-S. Kim (⊠) Environmental Biotechnology Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea e-mail: hkim@kribb.re.kr

K.-H. Baek

Department of Life Science, Chung-Ang University, Seoul 156-756, Korea

hydrocarbon contamination levels. These results suggest that geographical origin, rather than soil contamination level, might be more important in determining the bacterial diversity of crude oil-contaminated soils. Environmental factors, which play a major role in determining natural bacterial diversity which in turn should be enriched for effective bioremediation, should be the central dogma while considering bioremediation of hydrocarboncontaminated sites.

Keywords Soil · Bacterial diversity · Crude oil · Contamination · Geographic location · Statistical analysis

Introduction

Widespread use of petroleum products without adequate infrastructure and measures to control their leakage into environment has resulted in many contaminated sites across the globe and Korea is no exception (Singh et al. 2009; McGenity 2014). These sites are on various stages of bioremediation, and one of the first questions to be answered is to determine the mode of remediation, in situ or ex situ (Tomei and Daugulis 2012; McGenity 2014; Gillespie and Philp 2013). While in situ bioremediation is usually preferred considering its applicability and low cost, there are still questions over use of indigenous microorganisms or external consortium. Many studies have been performed to examine to the role of total petroleum hydrocarbons (TPH) levels in determining the microbial diversity of contaminated soils (Maila et al. 2006; Benedek et al. 2013). However, biodegradation of TPH is also affected by chemical, geochemical, and biological factors, including soil physico-chemical characteristics, chemical structure and type of oil; quantity and bioavailability of the



contaminant; microbial diversity of the soil type and above all environmental and meteorological factors (Ji et al. 2013; Gadd 2010). More importantly, geographic location combined with climate of the region also plays a significant role in pollutant degradation and this is notably overlooked in most bioremediation studies (Juck et al. 2000). The influence of these factors on microbial community dynamics is also not well documented.

Although some studies have documented the role of geographical location on soil microbial diversity, the derived conclusions were different. Maila et al. (2006) and Sutton et al. (2013) proved that TPH levels significantly influence microbial diversity, whereas others concluded the vital role of geographic location (Liu et al. 2010; Juck et al. 2000). The main factor that led to the difference in their conclusions is considered as the lack of consideration for local environmental or meterological factors, which might play a major role in microbial diversity and soil conditions.

In the present study, we have employed molecular techniques such as denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (T-RFLP) in addition to powerful statistical analyses such as self-organizing map (SOM) and principal component analysis (PCA) (Dollhopf et al. 2001; Palumbo et al. 2004) to understand the effect of various parameters on microbial diversity in a polluted environment. We further analyzed the correlation between meteorological data, soil physico-chemical parameters, and microbial diversity with the level of hydrocarbon contaminants in different long-term variously contaminated sites across Korea. This study was performed in 2010–2011 at Environmental Biotechnology Research Center, KRIBB, Daejeon, South Korea.

Materials and methods

Soil sampling

Thirty-two soil samples were taken from seven hydrocarbon-contaminated sites in the Republic of Korea, *i.e.*, Incheon (IN), Daegu (DG), Gunsan (GS), Ulsan (UL), Pohang (PO), Daejeon (DJ), and Chungbuk (CH) (Fig. 1). Long-term contaminated soil samples were collected especially in the vicinity of leaks from storage tank and pipeline systems. Selected soils from the same site within 1-ha area were included in same group. A mountain soil from Sokcho (SC) with no history of hydrocarbon contamination was used as a control. All soil samples were taken from just below the surface to depth of approximately 10 cm, sealed in plastic bags, and transported on ice to the lab and maintained at 4 °C prior for analysis.





Fig. 1 A map showing the collection locations of soil samples

Soil chemical properties and meteorological data

Most samples contained petroleum hydrocarbons originating from diesel spills. The soil organic matter (OM) content was determined by loss on ignition for 2 h at 550 °C; soil pH was measured in a 1:1 soil-water solution; soil total nitrogen (TN), organic carbon (OC), and total phosphorous (TP) were measured using standard techniques (Clescerl et al. 1999) after extraction of total petroleum hydrocarbons (TPH). TPH were extracted with dichloromethane and analyzed using a gas chromatograph (Varian 3400cx, CA, USA) equipped with a flame ionization detector (FID). An ATM-5 capillary column with dimensions of $30 \text{ m} \times 0.32 \text{ mm}$ and $0.25 \mu \text{m}$ film thickness (Alltech Inc., Deerfield, IL, USA) was used, with a carrier gas of purified N₂ at a flow rate of 50 ml min⁻¹. The operating temperature program was started at 40 °C for 5 min, increased at a rate of 4 °C min⁻¹ to 170 °C for 3 min, then increased again at a rate of 5 °C min⁻¹ to 300 °C for 10 min. The injector and detector were maintained at 250 and 300 °C, respectively (Baek et al. 2007). Meteorological data for ten years including and prior to year of sam-(2001 - 2010)was obtained from Korea pling Meteorological Administration (KMA).

DNA extraction and PCR

Total genomic DNA was extracted from each soil sample (1 g, dry wt) using an Ultra soil extraction kit (MoBio Inc., Carlsbad, CA, USA) following the manufacturer's

instructions. The purity and quantity of DNA obtained from soil samples were examined by electrophoresis on 1 % agarose gel and measured at absorbance of 260 and 280 nm with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). One microliter of extracted DNA was amplified by PCR using an MJ mini Thermal cycler (Bio-Rad, Hercules, CA, USA). 16S rRNA genes from soil microbial communities were amplified by PCR using the primers, FAM27f (5'-FAM-AGA GTT TGA TCC TGG CTC GA-3') and 1542r (5'-AGA AAG GAG GTG ATC CAG CC-3'). Each 50 µl PCR mixture contained 100 µmol of each primer, 100 µM of each dNTP, 5 μ l 10× PCR buffer, 0.25 μ l 5 U Ex Taq polymerase (Takara Bio, Shiga, Japan), and the DNA. Samples were amplified as follows: 95 °C for 5 min, 30 cycles of denaturation (1 min at 94 °C), annealing (1 min at 60 °C), extension (1 min at 72 °C), and a final extension at 72 °C for 10 min. The amplified PCR product was examined by electrophoresis on 1 % agarose gel.

Terminal restriction fragment length polymorphism (T-RFLP) and diversity indices

Amplicons (50 µl) were purified using a PCR purification kit (Qiagen, Germany) as directed by the supplier and eluted with 20 µl sterile water. Purified PCR products (approximately 100 ng) were digested with 5 U of HaeIII (Fermentas, USA), in a 20 µl reaction volume. Restriction reactions were performed at 37 °C for 12 h. Aliquots (8 µl) of restriction digests were examined by 2.5 % agarose gel electrophoresis using SYBR Green I staining. For total restriction fragment analysis, 1 µl of digested samples were mixed with 1 µl of formamide (with loading buffer and DNA fragment length standard [Rox 2500, ABI]). The mixture was denatured at 94 °C for 5 min and snap-cooled on ice before electrophoresis on 7 % polyacrylamide gel for 10 h at 250 V using an ABI377 automated DNA sequencer (Applied Biosystems, CA, USA). T-RFLP profiles were analyzed using Genescan Software version 3.7 (Applied Biosystems, CA, USA).

Structural diversity was evaluated between samples using T-RFLP profiles by the Shannon (Shannon 1948) diversity index (H'), richness (S), evenness (E), and the reciprocal of Simpson's index (1/D). The Shannon diversity index was calculated as follows: $H' = -\Sigma$ pi(ln pi), where the summation is over all unique fragments i, and pi is the relative abundance of fragment i. The abundance of a particular fragment was determined using the peak height intensity in fluorescence units. Evenness was measured as follows: $E = H'/\ln S$. The reciprocal of Simpson's index (1/ D) for a fragment was calculated as follows: $1/D = 1/\Sigma$ pi2 (Simpson 1949). Richness (S) was defined as the number of unique T-RFs or operational taxonomic units in a profile. Statistically significant differences in H' and D among soil samples were tested using a randomization procedure, as described in the MS Excel instructions (Microsoft Office 2003, USA).

Denaturing gradient gel electrophoresis (DGGE)

Denaturing gradient gel electrophoresis was performed in a D-Code 16/16-cm gel system with a gel width of 1.0 mm (Bio-Rad, USA). Samples were electrophoresed on 8 % polyacrylamide gels with a 40–60 % denaturing gradient. Gels were run at 60 V for 18 h, then stained with Ethidium bromide and destained twice in $1 \times$ TAE buffer for 15 min. To evaluate DGGE band patterns, images were converted, normalized, and analyzed with the Kodak 1.0 software package (Eastman Kodak Co., USA) (Kim et al. 2010).

Cloning and sequencing of 16S rRNA genes

Bacterial 16S rRNA gene fragment amplified using the universal primer set, 27F and 1542R from each soil, were purified with a QIAquick PCR purification kit (Qiagen, Germany). Each purified product was ligated by using a pGEM-T easy vector (Promega Co., USA) and cloned into competent *E. coli* DH5 α cells as described by the manufacturer. The clones containing inserts of the correct size were sequenced with an ABI Prism 377 automatic sequencer (Applied Biosystems, CA, USA). The sequences were then compared with the GenBank database using the BLASTN facility of the National Center for Biotechnology Information (http://ncbi.nlm.nih.gov). 16S rRNA gene sequences of the clone library were aligned with Clone Manager 7 (Sci Ed Central, USA) to identify *Hae*III sites.

Statistical analyses

Various statistical methods have been applied to the analysis of associations among variables in multivariate measures of T-RFLP profiles. After standardization, T-RFLP profiles were normalized to make the cumulative peak



Fig. 2 Representative profile of Incheon soil bacterial diversity (IN-H1) by T-RFLP assays. **a** T-RFLP profiles, **b** numeric representation



height for each profile equal 10,000 fluorescence units (Fig. 2). This facilitated comparison of profiles based on relative peak heights (Dunbar et al. 2001). The standardized data were then subjected to statistical analysis. Jaccard distance metrics were used to analyze T-RFLP profiles, based on the presence or absence of T-RFs. The Jaccard coefficient indicated the presence/absence of T-RFs and was equal to the ratio of the number of T-RFs in common between the two profiles relative to the total number of T-RFs present in both profiles. Agglomerative hierarchical clustering (Ward 1963) was applied to obtain dendrograms for each distance metric. T test and ANOVA were performed using SPSS 18 (SPSS Inc., USA).

Correlation efficiency between the bacterial diversity, soil physico-chemical properties, and Meteorological data was determined. However, since the data are reported in unique units respective to that variable and to directly compare data between variables, all measurements were converted into a common scale through Z-transformation so that the mean and variance corresponded to 0 and 1. Changed Z-score was used to compute Pearson's linear correlation analysis (SPSS 18.0, USA) (Tan et al. 2003).

We also conducted PCA and SOM to detect associations among geographic location and/or hydrocarbon contamination level on the microbial diversity of soil. A correlation matrix was computed containing all variables for PCA using SPSS 18 (SPSS Inc., USA). These variables were transformed into normalized variables with a mean of zero. SOM allows neighborhood preservation and local resolution of the input space proportional to the data distribution



Results and discussion

In the present study, we have considered three sets of data to determine the real influence of TPH levels on soil microbial diversity. Experimental analyses were performed to understand the role of soil physico-chemical properties on microbial diversity, which was in turn determined by DGGE and T-RFLP. Historical meteorological data (2000–2009) and data of pertaining to the study period (2010) were collected and their role on soil microbial diversity was established through robust statistical analy-



ses. Finally, a multi-factor analysis was performed involving geographically different sampling sites with varied TPH levels, microbial diversity of those sites, and meterological and environmental data, covering all factors which might influence microbial diversity in soil samples contaminated by hydrocarbons.

Soil characterization

As shown in Supplementary Table 1, the content of organic matter in all the soil samples was generally low (0.3–3.0 %). Most soil samples had neutral pH, but two soil samples of GS group were highly acidic (pH 2.5 and 2.9). Both organic carbon and total nitrogen contents were higher for UL soils, compared with other soils. The C/N ratio was high in IN-H4 (15.48), UL-L3 (20.70), and DJ-H1 (15.57). TPH contamination levels in soil samples ranged from 100 ppm (<500 ppm; low contamination levels) to 100,000 ppm (>2,000 ppm; heavily contaminated). Most soil samples had similar organic carbon and total nitrogen values, despite differences in the petroleum hydrocarbon contamination range.

Microbial diversity analyses

T-RFLP analysis

Total fluorescence units in each T-RFLP profile were calculated after excluding peaks with peak height of <100 fluorescence units. T-RF profiles were compared and standardized to the profile with the smallest total fluorescence units. The range of total fluorescence units in the sample set varied between 3,097 and 10,471 fluorescence units. This procedure was repeated until the cumulative peak height in all samples was the same.

Bacterial diversity index values calculated using T-RF by the restriction enzyme HaeIII are presented in Supplementary Table 2. We observed distinct differences in the Shannon diversity index (H'), richness (S), and the reciprocal of Simpson's index (1/D) between groups, but not in eveness (E). Generally, high H' and low 1/D values indicate high diversity in the microbial community. H' values were highest at the IN site (3.02 ± 0.14) , and the S values were also highest at IN sites (34.25 \pm 6.09). The bacterial diversity of the soils in IN group was the highest, and the soils of UL and CH showed relatively low diversity. Overall, the diversity indices were very similar within soils sampled from the same location, regardless of the TPH contamination level. In particular, the soil samples in IN, DG, and DJ sites shared almost identical H' values at each site, although they had different TPH concentrations.

16S rRNA gene clone library analysis

The nucleotide sequences of clones from each 16S rRNA gene library from the 32 soil samples were randomly analyzed. Of the 272 sequenced clones, from a total of 320 clones, 87 clones were found to be uncultured bacteria and 26 species were assigned (17 T-RFs by HaeIII) (Table 1). The bacteria related to Mesorhizobium amorphae (DQ022832), Acinetobacter calcoaceticus (AY800383), Pseudomonas sp. (DQ821413), and Sphingomonas xenophaga (AY611716) were predominant in all sites. In particular, the bacteria related to A. calcoaceticus was observed in all soil samples from the PO and DG sites, Pseudomonas sp. in all soils from the PO and CH sites, M. amorphae in all soils from the GS and PO sites, and S. xenophaga in all soil samples from the DG site, regardless of TPH concentration. All 17 T-RFs were observed only at IN sites, whereas other locations were shown to contain from 9 to 13 T-RFs.

Cluster analysis: TPH levels versus geographic location

The effects of geographical location and hydrocarbon contamination level on the bacterial diversities of the crude oil-contaminated soils were investigated by performing a cluster analysis and producing a dendrogram of the patterns generated by T-RFLP of the soil samples from each site (Fig. 3). The cluster analysis showed that samples obtained at the same location clustered together. High similarity of band patterns was shown at the same locations using Jaccard coefficient. Sample sites could be divided into several clusters based on their geographic origin. GS and PO soil samples clustered together. The IN soils clustered together, separately from the soils in other group, but low TPH soils (IN-L1, IN-L2, and IN-L4) and high TPH soils (IN-H1, IN-H2, and IN-H3) did not cluster together within the IN group. The low contamination and high contamination soils from IN clustered together based on geographic location. The average similarity between soil samples based on the Jaccard distance was 61 % in IN, 85 % in DG, 84 % in GS, 89 % in UL, 88 % in PO, 87 % in DJ, and 84 % in CH. The PCA was performed to characterize the clustering of bacterial diversity with respect to TPH contamination level and geographic location. PCA also clustered the soil samples based on geographic location rather than the contamination level (Fig. 4). The results of the PCA corroborated with the dendrogram results. IN, DG, GS, and CH sites were clustered based on geographic location, respectively, but UL, PO, and DJ sites were not clustered on PCA. Thus, geographic location, rather than contamination level, might be more important in determining bacterial diversity.



Accession no.	The closest organism	HaeIII	TPH ^a		Location ^b						
		T-RF (bp)	High (8)	Low (23)	IN (8)	DG (5)	GS (5)	UL (3)	PO (3)	DJ (4)	CH (3)
AY642054	Propionibacterium acnes	66	4	10	5	4	1	2	1	1	0
CP000091	Ralstonia eutropha	78	4	8	3	3	4	0	1	0	1
AF300324	Ralstonia taiwanensis	78	4	8	3	3	4	0	1	0	1
EF173319	Rhizobium sp.	190	0	3	2	0	0	0	0	0	1
AE014292	Brucella suis	192	4	6	6	4	1	2	2	0	1
AF191737	Cupriavidus necator	199	4	5	4	3	0	0	0	0	2
AB246810	Alcaligenes sp.	201	3	7	3	2	3	0	0	1	1
AJ278452	Bordetella bronchiseptica	201	3	7	3	2	3	0	0	1	1
AY972267	Pseudomonas mosselii	201	3	7	3	2	3	0	0	1	1
EF031062	Burkholderia cepacia	201	3	7	3	2	3	0	0	1	1
Z93441	Acinetobacter lwoffii	201	3	7	3	2	3	0	0	1	1
AB245367	Dyella ginsengisoli	204	5	14	4	4	5	1	3	3	1
DQ022832	Mesorhizobium amorphae	204	5	14	4	4	5	1	3	3	1
AB245366	Rhodanobacter lindaniclasticus	206	2	9	4	3	1	0	1	0	2
AM184292	Aeromonas punctata	207	1	1	1	0	0	0	0	0	1
AJ516049	Steonotophomonas maltophilia	209	2	2	3	0	0	0	1	0	0
AY800383	Acinetobacter calcoaceticus	213	7	15	3	5	4	2	3	3	1
DQ821413	Pseudomonas sp.	216	6	12	3	1	4	2	3	2	3
AB245357	Panaciterramonas fulva	217	4	7	1	2	2	1	0	4	1
AF177667	Bordetella hinzii	217	4	7	1	2	2	1	0	4	1
EU275167	Achromobacter sp.	217	4	7	1	2	2	1	0	4	1
AB245357	Panaciterramonas fulva	219	3	5	1	3	1	1	1	1	0
DQ113453	Pseudomonas citronellolis	219	3	5	1	3	1	1	1	1	0
D13723	Sphingopyxis macrogoltabida	225	3	9	2	4	4	1	1	0	0
U37337	Sphingomonas paucimobilis	228	0	2	1	0	0	0	0	1	0
AY611716	Sphigomonas xenophaga	291	4	14	4	5	4	1	2	1	2

 Table 1
 Bacterial species identified by the 16S rRNA gene clone library and the number of soil samples containing the corresponding T-RF at each site

^a High means a TPH contamination level higher than 2,000 ppm, and low means less than 500 ppm. The total number of high contamination level soil samples was eight, and there were 23 low contamination soil samples

^b Soil samples were collected from seven different locations in Korea: IN, Incheon; DG, Daegu; GS, Gunsan; UL, Ulsan; PO, Pohang; DJ, Daejeon; CH, Chungbuk. The number of soil samples in each location is indicated in the parentheses

In SOM analyses, soil samples were clustered as groups according to their geographic location (Fig. 5). Using the U-matrix, sampling sites were classified into six groups. The IN sites are mostly located at the bottom of the map (Fig. 5a), whereas GS, DG, and DJ sites found on the right area of the map. The distances in the U-matrix were rescaled from 1.5 to 5.5. The IN sites are clustered in the bright areas with low U-matrix values, indicating close distances, whereas other sites cluster in dark areas with high values, indicating longer distances (Fig. 5b). The SOM results showed a similar trend to the results of the dendrogram and PCA.

Correlation analysis: microbial diversity and environmental parameters

Pearson's correlation analysis after Z-transformation of meteorological data and microbial diversity index (*H*') confirmed that significant correlation exist between microbial diversity and meteorological data such as fog and rainfall duration, total sunshine, temperature, and amount of rainfall (Supplementary Table 3). The rainfall duration showed high negative correlation (-0.624, P < 0.0001) with *H*', but showed high positive correlation with bacterial groupings in cluster analysis (0.731, P < 0.0001). Bacterial







Fig. 3 Dendrogram constructed using T-RFLP profiles of the soil DNA samples



Fig. 4 Principal component analysis using T-RFLP profiles of the soil DNA samples. a Jaccard distance. b Hellinger distance. Circle Incheon, square Daegu, triangle Gunsan, inverted triangle Ulsan, diamond Pohang, plus symbol Daejeon, five pointed star Chungbuk,

diversity is lower in high rainfall duration and low total sunshine duration. Total sunshine has negative correlation with diversity index (-0.752, P < 0.0001) and positive correlation with bacterial groupings in cluster analysis (0.53, P < 0.0001) Hence, these results indicate that bacterial diversity is strongly affected by temperature, rainfall, and light duration (Table 2). Bacterial groupings of DGGE bands in cluster analysis (Supplementary Fig. 1) showed a close relationship with rainfall duration (0.558, P < 0.001), total sunshine (-0.583, P < 0.0001), fog duration and cloudiness (-0.438, P < 0.0006), and similar results were obtained for T-RFs (0.424, P < 0.0008) in case of rainfall duration and other meteorological data (Supplementary Fig. 2). However, CN ratio, pH, TPH, and OC values showed a weak correlation in DGGE band cluster analysis (Supplementary Table 4).

The re-grouping of the bacterial community with respect to known TPH degrading community and soil bacteria and performance of Pearson's correlation analysis further confirms that TPH decomposers are strongly correlated with the average temperature and high day time temperatures, and high total nitrogen levels irrespective of the TPH levels (Supplementary Table 5 & Supplementary Fig. 2, 3). On the contrary, the common soil bacteria have a positive correlation with other meteorological parameters such as fog duration and total sunshine (Supplementary Table 6). This further confirms presence of TPH decomposers are directly dependent on environmental factors rather than the actual TPH levels.

Furthermore, basic statistical analyses such as T test and ANOVA also confirmed that microbial diversity indices



six pointed star Sokcho (Control). *Non-filled symbols* indicate low TPH contamination level (<500 ppm) and *filled symbols* indicate high TPH contamination level (>2,000 ppm)





Fig. 5 Hierarchical clustering of different soil samples using the trained SOM map. a U-matrix clustering SOM units, b Classification of the sampling sites on the SOM based on sampling site characteristics

 Table 2
 Pearson's correlation coefficients of microbial diversity and diversity cluster

Table 3	Statistical	analysis	of	correlation	between	microbial	diver-
sity and	TPH levels	/geograpl	nic	location			

	Pearson of	Pearson correlation		P value		
	H'	Cluster	H'	Cluster		
Cluster	-0.701	1	0.001			
рН	0.16	-0.416	0.191	0.009		
TPH	0.11	-0.001	0.275	0.497		
OC	-0.118	-0.032	0.26	0.431		
TN	-0.166	0.089	0.182	0.314		
C/N ratio	-0.012	-0.082	0.474	0.328		
Average temperature	-0.333	0.193	0.031	0.145		
High temperature	-0.39	0.377	0.014	0.017		
Low temperature	-0.043	-0.262	0.408	0.073		
Rainfall	-0.206	0.444	0.13	0.005		
Rainfall duration	-0.624	0.731	0.001	0.001		
Mean wind speed	0.33	-0.407	0.033	0.01		
Mean humidity	0.112	0.016	0.271	0.465		
Total sunshine	0.53	-0.752	0.001	0.001		
Fog duration	0.493	-0.477	0.002	0.003		
Cloudiness	-0.3	0.566	0.048	0.001		
Minimum gross temperature	-0.108	-0.048	0.278	0.397		
Surface temperature	-0.268	0.125	0.069	0.247		

-1.0 to -0.7 Strong negative correlation

-0.3 to -0.7 Clear negative correlation

-0.1 to -0.3 Weak negative correlation

-0.1 to +0.1 No correlation

+0.1 to +0.3 Weak positive correlation

+0.3 to +0.7 Clear positive correlation

+0.7 to +1.0 Strong positive correlation



(H') weakly correlated with TPH levels whereas geographic location strongly correlated (Table 3). Finally, to confirm the above findings, the T-RFLP results were taken together with cluster analysis, PCA, and SOM to identify any significant correlation between bacterial diversity and TPH concentration/geographical location. The statistical results for each soil sample indicated the clustering of soil samples based on geographical location; however, we found no clustering based on TPH contamination level. The bacterial groupings of DGGE bands and dendograms generated by cluster analysis also confirmed the strong



correlation of bacterial diversity with that of geographical location, which is influenced by environmental and meteorological factors, rather than levels of TPH. The results of the PCA and SOM corroborated with the cluster analysis. Both PCA and SOM analyses showed that geographic location determines microbial composition. It should be noted that both PCA and SOM makes no assumptions about the distribution of the input variables or their relationships to one another, making it an attractive alternative to traditional methods. Grouping by SOM was generally in accordance with PCA, where the patterned nodes of the SOM were classified according to geographic location. SOM was also useful for finely classifying the groups. Moreover, the known TPH degrading community reported so far, though present in some samples, were diverse in each sample and often dependent on the geographical origin rather than on the TPH levels. In fact, some known TPH decomposers, such as Pseudomonas sp., are present in all samples, irrespective of the TPH levels (Obayori et al. 2009).

Although many studies suggest that changes in soil hydrocarbon content result in characteristic shifts in microbial populations, particularly in the abundance of bacteria utilizing hydrocarbons (Benedek et al. 2013; Maila et al. 2006; Bundy et al. 2002; Baek et al. 2007), we present that environmental parameters should be taken into consideration while arriving at such conclusion. A multi-dimensional approach involving environmental factors, contamination levels, soil characteristics, and microbial community analysis is required to arrive at a definitive conclusion (Juck et al. 2000). This blaring neglect of environmental parameters in the above-mentioned studies should be one of the main reasons for differing opinions in establishing the effect of TPH levels or geographic origin on microbial communities in contaminated soils. This study definitively points to the role of geographic origin in determining bacterial diversity in contaminated soils and this would serve as a starting point for further studies to underscore the role of geographic location.

Conclusion

Bacterial diversity analysis in contaminated environments should be seen in totality with overall environmental conditions in the region and should be effectively investigated using appropriate statistical analyses. Here, we establish that geographical origin and environmental factors of the contaminated site play a more important role in determining bacterial diversity than soil TPH levels. In other words, enriching native bacterial diversity would also be an effective, long-term solution for sustainable environmental management of hydrocarbon-contaminated sites.

Acknowledgments This work was carried out with financial support from the project (KGM 3151312) under the KRIBB Research Initiative Program (www.kribb.re.kr).

References

- Alhoniemi E, Himberg J, Parhankangas J, Vesanto J (2000) Som Toolbox. http://www.cis.hut.fi/projects/somtoolbox/
- Baek KH, Yoon BD, Kim BH, Cho DH, Lee IS, Oh HM, Kim HS (2007) Monitoring of microbial diversity and activity during bioremediation of crude oil-contaminated soil with different treatments. J Microbiol Biotechnol 17(1):67–73
- Benedek T, Vajna B, Táncsics A, Márialigeti K, Lányi S, Máth I (2013) Remarkable impact of PAHs and TPHs on the richness and diversity of bacterial species in surface soils exposed to long-term hydrocarbon pollution. World J Microbiol Biotechnol 29:1989–2002
- Bundy JG, Paton GI, Campbell CD (2002) Microbial communities in different soil types do not converge after diesel contamination. J Appl Microbiol 92(2):276–288. doi:10.1046/j.1365-2672.2002. 01528.x
- Clescerl LS, Greenberg AE, Eaton AD (1999) Standard methods for the examination of water and wastewater, 20th edn. American Public Health Association, Washington
- Dollhopf SL, Hashsham SA, Tiedje JM (2001) Interpreting 16S rDNA T-RFLP data: application of self-organizing maps and principal component analysis to describe community dynamics and convergence. Microb Ecol 42(4):495–505. doi:10.1007/s00248-001-0027-7
- Dunbar J, Ticknor LO, Kuske CR (2001) Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. Appl Environ Microbiol 67(1):190–197. doi:10. 1128/aem.67.1.190-197.2001
- Gadd GM (2010) Metals, minerals and microbes: geomicrobiology and bioremediation. Microbiology 156:609–643
- Gillespie IM, Philp JC (2013) Bioremediation, an environmental remediation technology for the bioeconomy. Trends Biotechnol 31(6):329–332
- Ji X, Ripp S, Layton A, Sayler G, Debruyn J (2013) Assessing long term effects of bioremediation: soil bacterial communities 14 years after polycyclic aromatic hydrocarbon contamination and introduction of a genetically engineered microorganism. J Bioremed Biodeg 4:2
- Juck D, Charles T, Whyte LG, Greer CW (2000) Polyphasic microbial community analysis of petroleum hydrocarbon-contaminated soils from two northern Canadian communities. FEMS Microbiol Ecol 33:241–249
- Kim B-H, Baek K-H, Cho D-H, Sung Y, Koh S-C, Ahn C-Y, Oh H-M, Kim H-S (2010) Complete reductive dechlorination of tetrachloroethene to ethene by anaerobic microbial enrichment culture developed from sediment. Biotechnol Lett 32:1829–1835
- Kirk J, Zurada J (2004) Topography-enhanced BMU search in selforganizing maps. In: Yin F-L, Wang J, Guo C (eds) Advances in neural networks-ISNN 2004. Springer, Berlin, Heidelberg
- Liu W, Luo Y, Teng Y, Li Z, Ma L (2010) Bioremediation of oily sludge contaminated soil by stimulating indigenous microbes. Environ Geochem Health 32:23–29



- Maila MP, Randima P, Drønen K, Cloete TE (2006) Soil microbial communities: influence of geographic location and hydrocarbon pollutants. Soil Biol Biochem 38:303–310
- McGenity TJ (2014) Hydrocarbon biodegradation in intertidal wetland sediments. Curr Opin Biotechnol 27:46–54
- Obayori OS, Adebusoye SA, Adewale AO, Oyetibo GO, Oluyemi OO, Amokun RA, Ilori MO (2009) Differential degradation of crude oil (Bonny Light) by four Pseudomonas strains. J Environ Sci 21:243–248
- Palumbo A, Schryver J, Fields M, Bagwell C, Zhou J-Z, Yan T, Liu X, Brandt C (2004) Coupling of functional gene diversity and geochemical data from environmental samples. Appl Environ Microbiol 70:6525–6534

Simpson EH (1949) Measurement of diversity. Nature 163:688

Singh A, Kuhad RC, Ward OP (2009) Biological remediation of soil: an overview of global market and available technologies. In: Singh A, Kuhad RC, Ward, OP (eds) Advances in applied bioremediation. Springer, Berlin

- Shannon CE(1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
- Sutton NB, Maphosa F, Morillo JA, Abu Al-Soud W, Langenhoff AAM, Grotenhuis T, Rijnaarts HHM, Smidt H (2013) Impact of long-term diesel contamination on soil microbial community structure. Appl Environ Microbiol 79:619–630
- Tan TH, Wong KY, Cheng TK, Heng JT (2003) Coronary normograms and the coronary-aorta index: objective determinants of coronary artery dilatation. Pediatric Cardiol 24:328–335
- Tomei MC, Daugulis AJ (2012) Ex situ bioremediation of contaminated soils: an overview of conventional and innovative technologies. Crit Rev Environ Sci Technol 43:2107–2139
- Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244

