



The latitudinal gradient in rock-inhabiting bacterial community compositions in Victoria Land, Antarctica



Jaejin Lee ^{a,b}, Junho Cho ^{a,c}, Yong-Joon Cho ^a, Ahnna Cho ^a, Jusun Woo ^d, Jongik Lee ^d, Soon Gyu Hong ^a, Woo Jun Sul ^c, Ok-Sun Kim ^{a,*}

^a Division of Polar Life Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea

^b Department of Agricultural & Biosystems Engineering, Iowa State University, Ames, IA 50011, United States of America

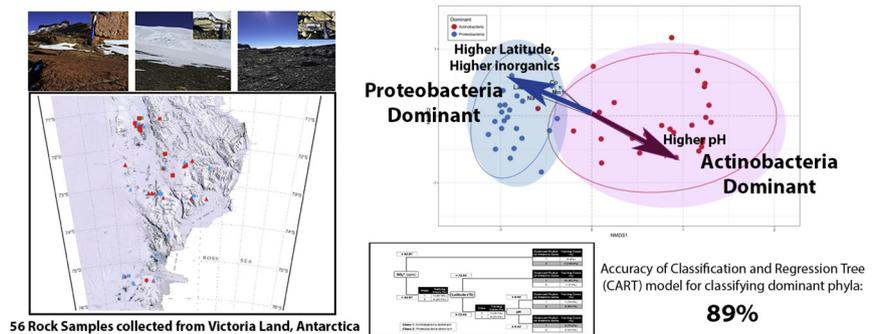
^c Department of Systems Biotechnology, Chung-Ang University, Anseong 17546, Republic of Korea

^d Division of Polar Earth-System Sciences, Korea Polar Research Institute, Incheon 21990, Republic of Korea

HIGHLIGHTS

- Antarctic rock is a less studied ecosystem due to poor logistics and accessibility.
- Relationships between lithic bacteria and the environment are not well understood.
- MiSeq and statistical analyses were conducted for 56 rock samples from Antarctica.
- Latitude was significantly correlated with the dominant phyla, pH, and NO₃.
- The CART model classified the dominant phylum with 89% accuracy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 11 October 2018

Received in revised form 5 December 2018

Accepted 5 December 2018

Available online 6 December 2018

Editor: Frederic Coulon

Keywords:

Antarctica

Rock-inhabiting bacterial community

Latitudinal gradient

Correlation

Machine learning

ABSTRACT

The harsh conditions in Victoria Land, Antarctica have formed a simple ecosystem dominated by microbes that use rocks as shelters to avoid environmental stressors. The area is composed of basement rocks that illustrate the history of complex deformation, thus it is highly valuable not only in perspectives of geology but also in biological aspects. Because this region is inhospitable to higher-level organisms and receives least external influences, it can be an ideal environment to investigate the relationship between rock-inhabiting bacterial communities and environmental factors. In such conditions, inorganics dissolved from minerals can be considered as key factors influencing rock-inhabiting bacterial communities. Thus, the present study attempted to explore rock-inhabiting bacterial communities throughout Victoria Land, to identify environmental parameters that are more influential on bacterial community compositions, and to investigate latitudinal gradients in environmental parameters and rock-inhabiting bacterial communities. The results suggested that (i) rock-inhabiting bacterial communities in Victoria Land predominately consisted of either *Actinobacteria* or *Proteobacteria*; (ii) latitudinal gradients in rock-inhabiting bacterial community compositions and some environmental parameters were observed; (iii) latitude, pH, nitrate, and sulfate significantly correlated with the dominant phyla; and (iv) the Classification and Regression Tree (CART) analysis demonstrated that latitude, pH, and sulfate concentrations could explain the dominant phylum with 89% accuracy. These findings can provide important insight into the link between environmental factors and rock-inhabiting bacterial community compositions in conditions with extremely cold temperatures and high levels of radiation, which could also be of interest in astrobiology.

© 2018 Elsevier B.V. All rights reserved.

* Corresponding author at: Division of Polar Life Sciences, Korea Polar Research Institute, 26, Songdomirae-ro, Yeosu-gu, Incheon 21990, Republic of Korea.

E-mail address: oskim@kopri.re.kr (O.-S. Kim).

1. Introduction

Antarctica is one of the driest, coldest, and most remote places on Earth (Selbmann et al., 2008; Cary et al., 2010). While most of the continent is covered with ice and snow, a tiny portion (i.e., 0.32%) is ice-free and is mostly composed of exposed rocks and oligotrophic mineral soils (Ugolini and Bockheim, 2008). Victoria Land, located in eastern Antarctica, is one of the largest ice-free regions on the Antarctic continent. The region is exposed to severe environmental stressors such as extremely low temperature, high ultraviolet radiation, and low moisture levels, which are inhospitable to most higher-level organisms (Cowan and Ah Tow, 2004; Makhalyane et al., 2014). Northern and southern Victoria Land have a wide area of rocks exposed to various environmental conditions (i.e., mountainous terrains ranging from 70° S to 76° S and cold deserts from 76° S to 78° S). Since most of the area is composed of basement rocks that illustrate the complex deformational history (Williams et al., 1971; Faure and Mensing, 2010), it is not only geologically significant but also biologically interesting. Due to the unfavorable living conditions and least external influences, this region has often been considered as an analogue for the rough characteristics of the Martian surface (Wynn-Williams and Edwards, 2000; Wierzbos and Ascaso, 2002; Dartnell et al., 2010). Therefore, investigations of rock-inhabiting microbes can provide insight into microbial survival in such harsh conditions including extraterrestrial life.

These geographical and climatic properties have formed a simple ecosystem dominated by microbial communities that use rocks as shelters from these stressors (Cary et al., 2010; de los Ríos et al., 2014). In addition to the climatic barriers, there is little chance of organic matter deposited from the surrounding environment, because higher-level organisms, including plants, cannot survive in this region. Thus, low levels of organic material have been reported as one of critical environmental factors that inhibit microbial life in this region (Vishniac, 1993). Under the harsh conditions, inorganics dissolved from minerals which can be potentially used by microbes are considered as key factors influencing bacterial communities (Uroz et al., 2009). Although this simple ecosystem dominated by microbes can be an ideal environment to investigate the relationship between environmental factors and bacterial community composition, there have been limited attempts to understand the link (Barrett et al., 2006; Van Goethem et al., 2016).

While the link between rock-inhabiting microbial communities and environmental factors have remained unknown, several studies have been conducted to understand which environmental factors are locally or regionally related to the composition of soil bacterial communities in Antarctica (well summarized in Bottos et al., 2014). For example, Sokol et al. demonstrated that dominant groups in soil microbial communities in the McMurdo Dry Valleys were influenced by gradients in pH and moisture at the Transantarctic scale (Sokol et al., 2013). New Zealand's Latitudinal Gradient Project (LGP) in Antarctica was also one of extensive efforts made to understand which environmental factors affect the composition of bacterial communities and to find patterns in ecosystem structure and function changed by latitude (Howard-Williams et al., 2010). After a decade of research, the LGP concluded that the effects of the latitudinal gradient on ecosystem structure and function did not overwhelm local microscale variability. (Howard-Williams et al., 2010; Smith et al., 2010). However, it is still unclear which environmental factors are influential to rock-inhabiting bacterial communities throughout northern and southern Victoria Land or whether latitudinal patterns can be found.

In a recent study, Yung et al. showed that different types of *Cyanobacteria* predominated in chasmoendolithic microbial communities depending on how much sunlight the slope they were inhabiting received (Yung et al., 2014). The study provided a clue to link rock-inhabiting bacterial communities with environmental factors, but other geographical (e.g. latitude) and chemical parameters (e.g. pH) were not considered. The majority of previous studies on rock-inhabiting bacteria in Antarctica have focused on the microbial communities that inhabit a specific location of rocks; for example, studies have explored endolithic

microbial communities using microscopy (Omelson, 2008), examined the predominance of *Cyanobacteria* in hypolithic microbial communities (Chan et al., 2012), explored the predominance of different phyla (i.e., *Proteobacteria*, *Cyanobacteria*, and *Actinobacteria*) in hypolithic bacterial communities in the Dry Valleys (Makhalyane et al., 2013), and examined the endolithic microbial diversity in different color layers within rocks (Archer et al., 2017). Although research focused on niche-specific bacterial communities has expanded the understanding of microbes inhabiting Antarctic rocks, the approach is not suitable for conducting an overall survey of rock-inhabiting bacterial communities in a large area.

The aims of this study were to explore rock-inhabiting bacterial communities throughout Victoria Land, to identify environmental parameters that are more influential on bacterial community compositions, and to investigate latitudinal gradients in environmental parameters and rock-inhabiting bacterial communities. To achieve these goals, MiSeq 16S rRNA gene sequencing was conducted without differentiating surface and internal fractions for 56 rock samples collected from northern and southern Victoria Land at latitudes ranging from 71° S to 77° S. Among various environmental factors, we focused on inorganics (i.e., cations, anions, and trace elements) together with geographical and chemical parameters. Classification and Regression Tree machine learning algorithms were applied to the dataset to rank the relative importance of influential factors and to see if the algorithm could discover so far unrecognized patterns between rock-inhabiting bacterial communities and environmental parameters.

2. Materials and methods

2.1. Sample collection

Rock samples were collected at 56 locations in northern and southern Victoria Land, Antarctica, during the summer season (December and January) from 2013 to 2015. While traveling on a helicopter for sampling, geologically representative rocks were randomly collected. Sampling sites were located at latitudes ranging from 71° 24' S to 77° 33' S and longitudes ranging from 159° 4' E to 167° 27' E (Fig. 1). Collected samples were contained in sterile Whirl-Pak bags (Nasco, USA) and stored at −20 °C during transportation. The rock samples were crushed using a sterile hammer and pestle without differentiating surface and internal fractions, to include all rock-inhabiting microorganisms (Choe et al., 2018). After homogenization, the samples were stored at −80 °C until use. Rock samples were classified into four types based on the process of formation: 8 metamorphic, 8 plutonic, 23 volcanic, and 17 sedimentary rocks (Fig. 2, Table S1).

2.2. Geochemical analysis

For the geochemical analysis, 40 g of each powdered sample was mixed with 40 mL of distilled water and shaken for 6 h using a Thermo Scientific Vari-Mix™ Test Tube Rocker. After the mixture was centrifuged, the supernatant was filtered using a 0.45 µm polypropylene syringe filter. The filtered supernatant was used to measure anion concentrations and to determine pH, as previously described (Banerjee et al., 2000). For cation analysis, the filtered samples were dried and redissolved in 5% HNO₃. Trace elements, cation concentrations, and anion concentrations eluted from the rock powder samples were determined using an inductively coupled plasma atomic emission spectrometer (ICP-AES; Perkin Elmer Optima 8300), an inductively coupled plasma mass spectrometer (ICP-MS; Thermo Fisher iCAP Q), and an ion chromatography system (ICS; Thermo Fisher Dionex ICS-1100), respectively, that were operated by the Korea Basic Science Institute (Table S2).

2.3. DNA extraction and MiSeq 16S rRNA gene sequence analysis

Genomic DNA was extracted from 0.3 g of freeze-dried rock powder using the FastDNA SPIN Kit for Soil (MP Biomedicals, USA), following

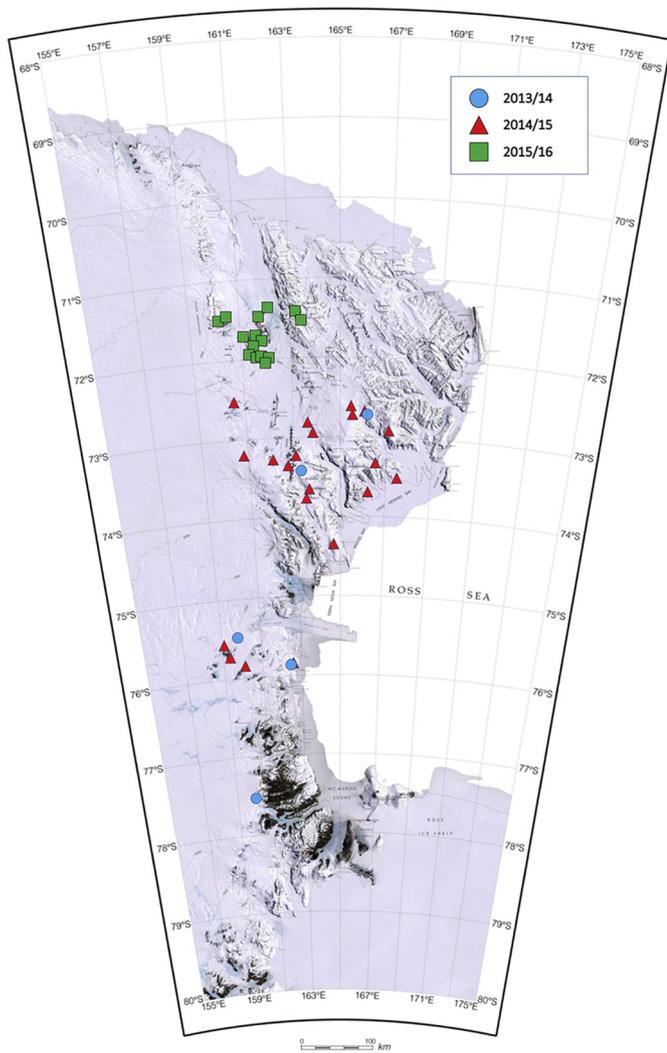


Fig. 1. Sampling locations in Victoria Land, Antarctica at latitudes ranging from 71° 24' 17" S to 77° 33' 57" S. Of 56 samples, 9 samples were collected in 2013 (blue circles), 32 samples in 2014 (red triangles), and 15 samples in 2015 (green squares).

the manufacturer's instructions. From the extracted DNA, the V3-V4 region of the bacterial 16S rRNA gene was amplified using the 341F and 805R primer set with the dual-index strategy (Illumina, 2013). Purified amplicons were pooled in equal molar amounts to prepare the library. The high-throughput sequencing of the library was performed using the MiSeq platform (Illumina, USA). Generated raw files were processed and analyzed using Mothur v1.39.5 following the MiSeq SOP (Schloss et al., 2009). The sequences were filtered by read length over 400 bp and aligned to Silva v128 reference files from Mothur. Chimeric sequences were removed using the UCHIME algorithm v4.2 (Edgar et al., 2011). After screening, the remaining reads were clustered into the Operational Taxonomical Units (OTUs) with a minimum 97% similarity threshold, and the resulting OTUs were assigned to taxons using the EzTaxon-e 16S rDNA database v2017.05 (Kim et al., 2012).

2.4. Statistical analysis and data mining

Spearman's rank correlation coefficient (ρ) was calculated to assess correlations between variables using SPSS Statistics 22 (SPSS Inc., USA). Further data analyses and visualizations were conducted in R v.3.4.1. Nonmetric multidimensional scaling (NMDS) ordinations were

plotted based on Bray-Curtis distances. The significance between groups was determined through analysis of similarity (ANOSIM) using vegan package (v2.4–4) (Oksanen et al., 2017). F-tests and two-sample *t*-tests were used to examine differences in variables between *Actinobacteria*- and *Proteobacteria*-dominant groups.

Data mining approaches using machine learning algorithms are promising tools for not only developing predictive models but also discovering unrecognized trends or patterns in large datasets; thus, those have been widely used for environmental research (Cortés et al., 2000; De'ath and Fabricius, 2000; Yoo et al., 2016; Yoo et al., 2018). In the present study, machine learning-based data mining was performed using Classification and Regression Tree (CART) algorithm, which uses the Gini impurity as a splitting criterion (Breiman et al., 1984). Six environmental variables (i.e., pH, Latitude, NO_3^- , SO_4^{2-} , Cd, and Se) which had significant correlations (i.e., $p < 0.05$) with relative abundances of *Actinobacteria* and *Proteobacteria* were used as input variables, and the dominant phylum class was used as the target variable in the data mining analysis. Of 53 samples predominated by *Actinobacteria* or *Proteobacteria*, 70% (i.e., 35 samples) were randomly chosen as training cases, and the rest were used as the test set. For coherence test, 10-fold cross-validation was carried out using 100 randomly shuffled data sets. The performance of each model was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), a scalar value calculated from the ROC graph (Fawcett, 2006). Based on the performances, a classification model with the highest AUC value for the training and the test sets was selected as the representative model (Lee et al., 2016). The data mining analysis was performed using the SAS Enterprise Miner Workstation 14.2 (SAS Inc., USA).

3. Results

3.1. Rock-inhabiting bacterial community compositions

A total of 7,408,605 16S rRNA sequences were clustered at 97% sequence similarity, which resulted in 9561 OTUs. The compositions of rock-inhabiting bacterial communities were distinctly divided into two groups: *Actinobacteria*- or *Proteobacteria*-dominant (Fig. 3). *Actinobacteria* were predominant in 28 rock samples, with abundances ranging from 35.0% to 91.2%. In the 25 rock samples predominated by *Proteobacteria*, the abundances of *Proteobacteria* ranged from 28.1% to 85.2%. Most of the abundant OTUs (i.e., 64 out of the top 100) were classified as either *Actinobacteria* or *Proteobacteria* at the phylum level. However, at the genus level, most of the OTUs belonging to the two phyla remained unclassified, although some were classified as *Rubrobacter* (*Actinobacteria*), *Commamonas*, *Pseudomonas*, or *Sphingomonas* (*Proteobacteria*). Other phyla, such as *Bacteroidetes*, *Firmicutes*, *Chloroflexi*, and *Acidobacteria*, consistently existed across all samples. *Cyanobacteria* was predominant (i.e., 34% relative abundance) in one sample collected from Little Rocks (LTR; 71° 24' S and 162° 0' E); however, *Cyanobacteria* abundance was below 1% for most of the samples.

3.2. Correlations between bacterial communities and environmental factors

Significant differences in bacterial community compositions between *Actinobacteria*- and *Proteobacteria*-dominant samples were also observed through NMDS ordination plots based on the OTU table (Fig. 4). Among the 43 geographical and chemical factors, five parameters (i.e., pH, Latitude, Na^+ , Mn^{2+} , and Co) were significantly correlated ($p < 0.01$) with rock-inhabiting bacterial communities in Victoria Land (Table S3).

Correlations between the dominant phyla and environmental variables were also calculated using the Spearman's rank correlation coefficient. Variables that were correlated (i.e., $p < 0.05$) with the dominant phyla/classes are shown in Table 1. Latitude, NO_3^- , and Se had negative correlations with the abundance of *Actinobacteria*, while they had positive

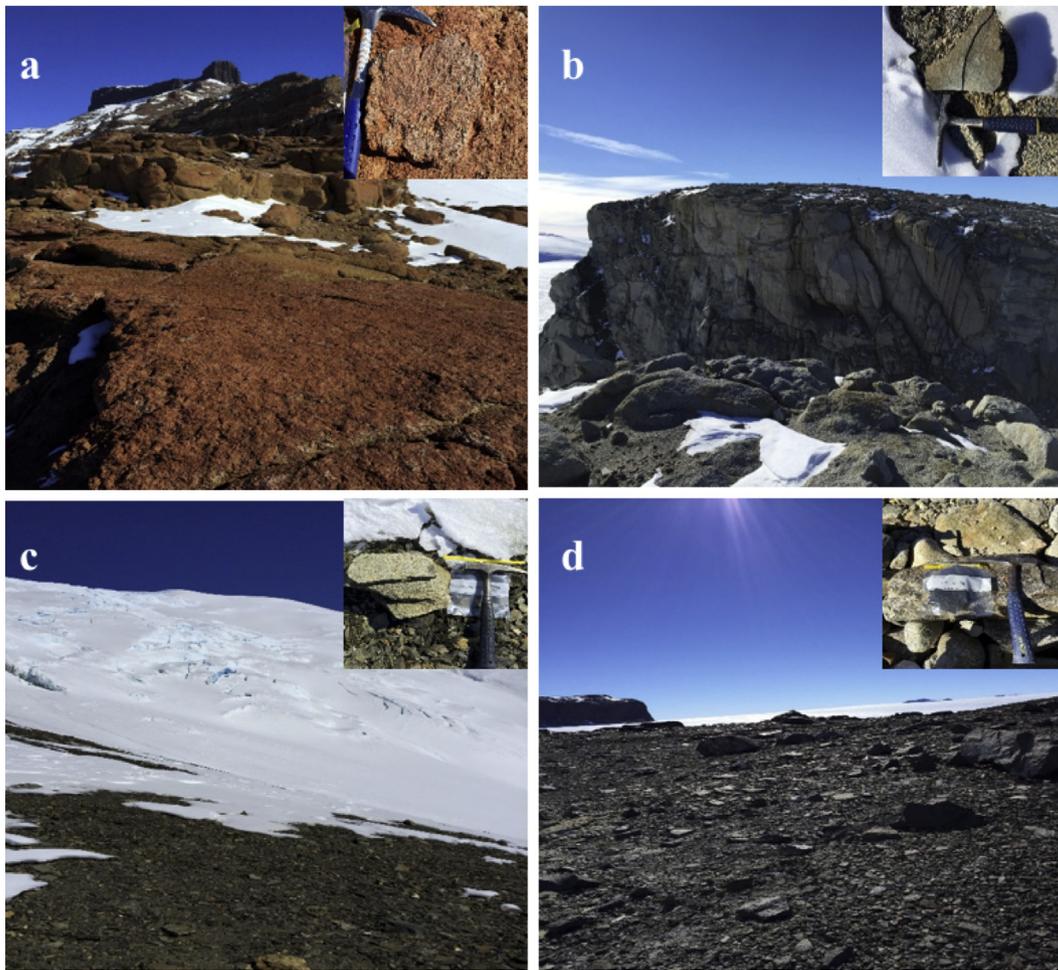


Fig. 2. Sampling sites in Victoria Land. (a) volcanic rock sample collected from Haban Spur; (b) plutonic rock from Star Nunatak; (c) metamorphic rock from Kyanite; (d) sedimentary rock from Beta Peak.

correlations with the abundance of *Proteobacteria*. In contrast, pH showed a negative correlation with *Proteobacteria*, whereas it was positively correlated with *Actinobacteria*. For example, relatively high pH was a strong

indicator of the dominance of *Actinobacteria* in rock-inhabiting bacterial communities. Similarly, relatively high latitude and high nitrate concentration were indicators of the dominance of *Proteobacteria* (Fig. 5).

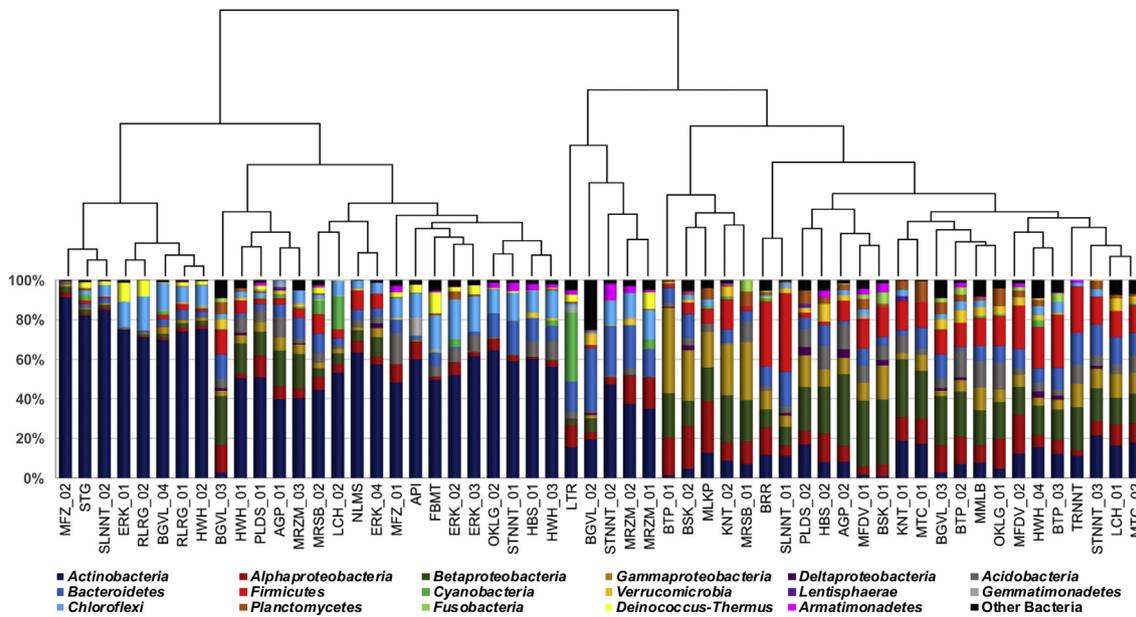


Fig. 3. Rock-inhabiting bacterial community structures at the phylum/class level combined with hierarchical clustering based on Euclidean distance.

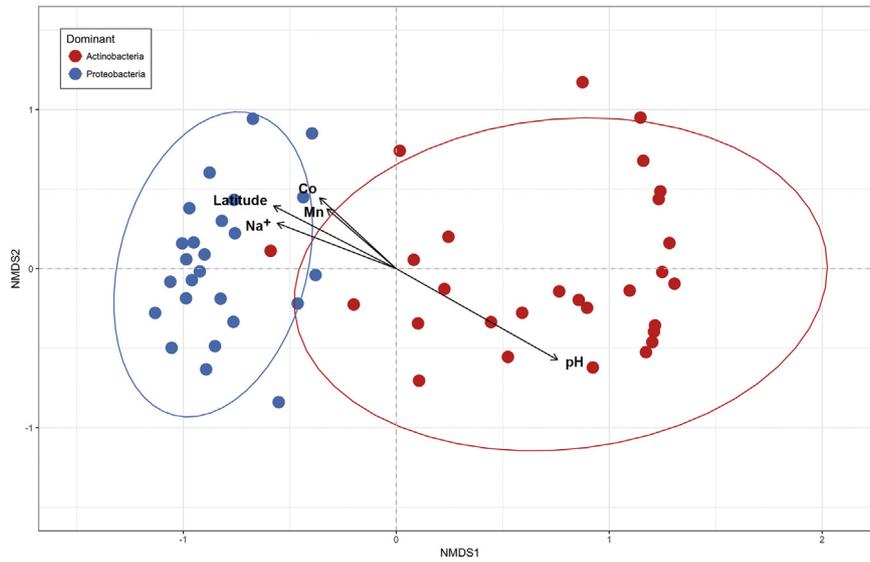


Fig. 4. NMDS of rock-inhabiting bacterial community compositions (Stress = 0.18; ANOSIM R = 0.72, $p < 0.001$) based on OTU abundances of rock samples dominated by *Actinobacteria* or *Proteobacteria*. The arrows indicate environmental parameters which were significantly correlated ($p < 0.01$) with rock-inhabiting bacterial communities.

Table 1

Spearman's rank correlation coefficients of comparisons between the dominant phyla/classes and the major environmental parameters (N = 53). Environmental factors that had significant correlations with the dominant phyla are selectively shown.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) <i>Actinobacteria</i> (%)	1.000										
(2) <i>Alphaproteobacteria</i> (%)	-.696 ^a	1.000									
(3) <i>Betaproteobacteria</i> (%)	-.730 ^a	.417 ^a	1.000								
(4) <i>Gammaproteobacteria</i> (%)	-.697 ^a	.441 ^a	.811 ^a	1.000							
(5) <i>Firmicutes</i> (%)	-.591 ^a	.296 ^b	.726 ^a	.724 ^a	1.000						
(6) Latitude (S°)	-.397 ^a	.274 ^b	.252	.380 ^a	.318 ^b	1.000					
(7) pH	.542 ^a	-.471 ^a	-.453 ^a	-.505 ^a	-.439 ^a	-.458 ^a	1.000				
(8) NO ₃ ⁻ (ppm)	-.580 ^a	.371 ^a	.594 ^a	.661 ^a	.566 ^a	.391 ^a	-.577 ^a	1.000			
(9) SO ₄ ²⁻ (ppm)	-.206	.269 ^b	.210	.246	.232	.043	-.339 ^b	.493 ^a	1.000		
(10) Se (ppb)	-.470 ^a	.355 ^a	.370 ^a	.389 ^a	.380 ^a	.383 ^a	-.536 ^a	.580 ^a	.549 ^a	1.000	
(11) Cd (ppb)	-.215	.306 ^b	.250	.216	.314 ^b	.325 ^b	-.465 ^a	.440 ^a	.387 ^a	.623 ^a	1.000

^a $p < 0.01$.

^b $p < 0.05$.

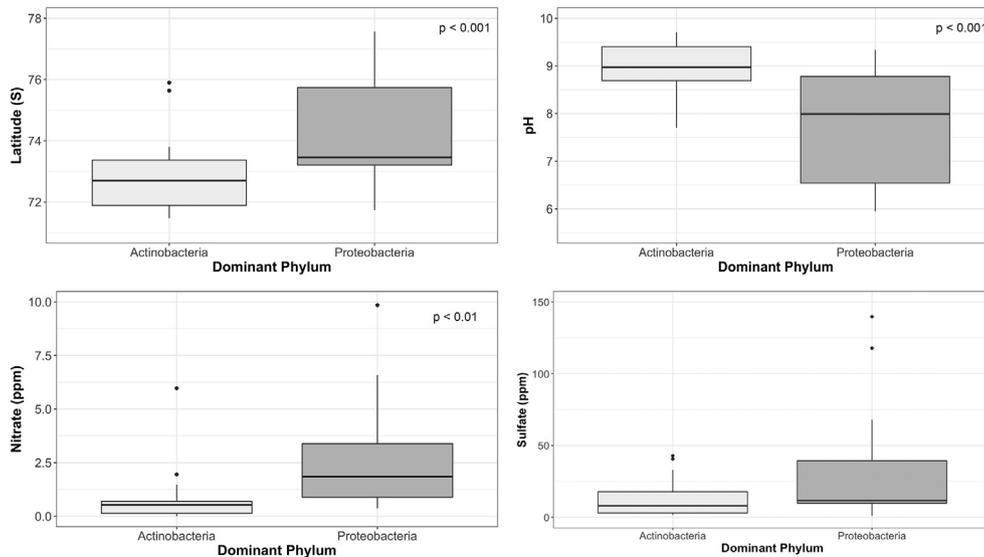


Fig. 5. Boxplots indicating the ranges and the median values for latitude, pH, nitrate, and sulfate concentrations for *Actinobacteria*- and *Proteobacteria*-dominant samples.

Table 2
Percentage of acceptable models^a and the performance indices of the best CART model.

Data partitioning ratio (Training set:Test set)		70:30	
Percentage of acceptable trees (AUC > 0.7)	Training sets	100%	
	Test sets	72%	
AUC of the best CART model	Training set	0.88	
	Test set	0.89	
True positive rates of the best CART model	Training set	Class 1 ^b	86%
		Class 2 ^b	93%
		Overall	89%
	Test set	Class 1 ^b	90%
		Class 2 ^b	88%
	Overall	89%	

^a Classification models with an Average Area Under the ROC Curve (AUC) above 0.7.^b Class 1: *Actinobacteria* dominant, Class 2: *Proteobacteria* dominant.**Table 3**

Relative importance values of input variables used to construct the best classification model.

Input variable	Relative importance
SO ₄ ²⁻	1.0000
Latitude	0.7501
pH	0.7305

Except for latitude, other measures of geographic parameters (i.e., longitude, elevation, and distance from the coast) had no significant correlations ($p > 0.05$) with the dominant phylum/classes. Latitude was significantly correlated with other geochemical parameters, such as pH, NO₃⁻, Se, and Cd. While pH had a negative correlation with latitude, NO₃⁻, Se, and Cd had a positive correlation with latitude. Although SO₄²⁻ was not correlated with latitude, it had a negative correlation with pH and a positive correlation with *Alphaproteobacteria* and NO₃⁻.

3.3. The best CART model and classification rules

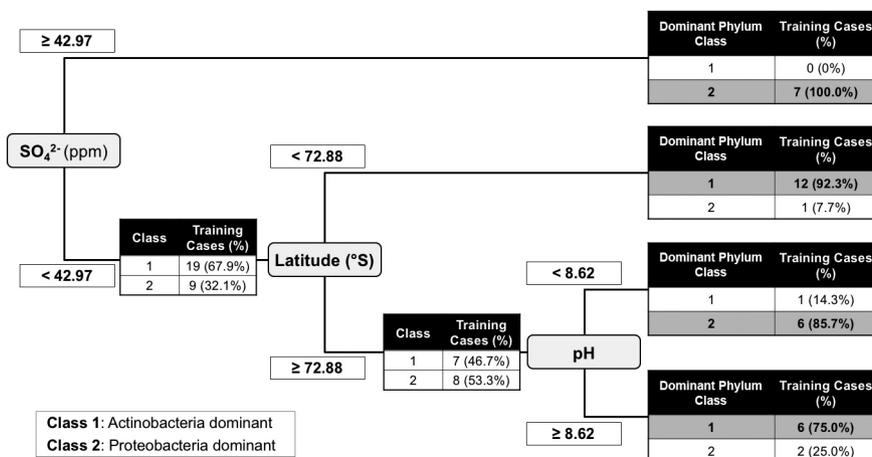
Six geochemical variables (pH, Latitude, NO₃⁻, SO₄²⁻, Cd, and Se) that were significantly correlated with the dominant phyla were chosen as input variables for data mining analysis. Of 100 individual classification trees resulting from the 10-fold cross-validation for 100 randomly shuffled data sets, the best CART model was selected based on the area under the ROC curve (AUC) value and performance indices (Table 2). The overall true positive rates of the best CART model were 89% for both the training set and the test set, which means the model correctly classified which phylum would be dominant using given environmental factors with 89% accuracy. SO₄²⁻, latitude, and pH were essential

parameters to construct the best CART model. Relative importance values calculated by the CART algorithm demonstrated that SO₄²⁻ was the most useful parameter for the classification, followed by latitude and pH (Table 3). The best CART model is shown in Fig. 6. The numbers in the right column of each node are the training cases that meet the splitting criteria, and a shaded row in a termination node indicates the dominant phylum determined by the CART algorithm (i.e., Class 1, *Actinobacteria*-dominant; Class 2, *Proteobacteria*-dominant). The first split was due to SO₄²⁻ concentrations. Seven training cases that had SO₄²⁻ concentrations exceeding 42.97 ppm belonged to *Proteobacteria*-dominant samples. For the 28 cases with SO₄²⁻ concentrations below 42.97 ppm, the next split was latitude. When latitude was lower than 72.88 °S, most of the training cases (92.3%) were *Actinobacteria*-dominant. For the 15 cases with a latitude higher than 72.88 °S, pH was the third split. With a pH below 8.62, the majority of training cases (85.7%) were *Proteobacteria*-dominant, while 75% of the training cases with a pH exceeding 8.62 belonged to *Actinobacteria*-dominant samples.

4. Discussion

The present study showed a latitudinal gradient in rock-inhabiting bacterial community compositions and environmental parameters in Victoria Land, Antarctica. Latitude was significantly correlated with the abundances of the dominant phyla and some geochemical parameters, including pH and NO₃⁻. Geochemical parameters, such as pH, NO₃⁻, SO₄²⁻, Se, and Cd, also showed significant correlations with the abundances of the dominant phyla. The results showed that the phylum *Actinobacteria* tended to be predominant at relatively low latitudes and high pH. In contrast, *Proteobacteria* was more likely to be dominant at relatively high latitudes, low pH, and sites with more inorganic nutrients. The CART algorithm was applied to the dataset to investigate most influential parameters and the synergetic effects of multiple parameters that affect rock-inhabiting bacterial communities in Victoria Land, Antarctica. The CART algorithm selected SO₄²⁻, latitude, and pH as essential variables to build a classification tree, which demonstrated that those variables could explain the dominant phylum with 89% accuracy.

Unlike the majority of previous studies on rock-inhabiting microbial communities in Antarctica, among 56 rock samples, only one sample was predominated by *Cyanobacteria*. Although one previous study showed a predominance of *Proteobacteria*, *Actinobacteria*, and *Cyanobacteria* in different hypolithic bacterial communities in the McMurdo Dry Valleys (Makhalyane et al., 2013), most studies demonstrated the predominance of *Cyanobacteria* in hypolithic microbial communities (Chan et al., 2012), in chasmoendolithic communities (Yung et al., 2014), and in endolithic microbial communities in sandstone and granite (Archer et al., 2017). The difference was mainly due

**Fig. 6.** The representative classification model classifying the dominant phylum of rock-inhabiting bacterial communities in Victoria Land, Antarctica. See text for details.

to the sampling and analytical methods. Previous studies focused on visual evidence (e.g. green or black colored-bands) of life on the rocks when collecting samples and analyzing bacterial communities. In the present study, we randomly collected the geologically representative rocks regardless of visual evidence of life on the rocks. Additionally, the analytical method without differentiating surface and internal fractions could result in the difference.

New Zealand's Latitudinal Gradient Project (LGP) suggested that the changes in the environment and biota shown within five degrees of latitude are due to other geographical and climatic factors including altitude, distance from the coast, topography, and microclimate (Howard-Williams et al., 2010). However, we did not observe any significant correlations between rock-inhabiting microbial community structures and other geographical factors such as longitude, elevation, and distance from the coast (i.e., $p > 0.05$). It has been known that pH is one of major factors controlling lithotrophic microbial populations (Banfield et al., 1999). Our observations on rock-inhabiting bacterial communities and environmental parameters in Victoria Land also confirm that dominant phyla and bacterial community compositions are affected by changes in pH along latitudinal gradients.

Because a rock itself is an oligotrophic environment and the extremely cold environments receive least external influences, availability of inorganics can be a key factor for bacterial survival and growth (Uroz et al., 2009). For example, microbes are required to obtain enzyme co-factors (eg. Mo, Cu, Zn, Mg, Ni, etc.) from dissolution of sulfide minerals. They also need phosphorus (to construct DNA, RNA, ATP, etc.), which can be obtained from dissolution of apatite and other secondary phosphates (Banfield et al., 1999). Our results showed that inorganic parameters (i.e., cations, anions, and trace elements), which had correlation with bacterial community compositions, were positively correlated with *Proteobacteria*-dominant samples. Since many bacterial species in the phylum *Proteobacteria* are capable of utilizing NO_3^- , Fe_3^+ , or SO_4^{2-} as electron acceptors, the correlations between the abundance of *Proteobacteria* and NO_3^- or SO_4^{2-} can be readily understood. However, explaining the correlations between rock-inhabiting bacterial communities and trace elements or cations is more challenging. A previous study showed that K^+ was the only parameter correlated with bacterial communities in weathered rocks from a karst ecosystem in China and *Actinobacteria* was the only phylum that showed a significant correlation with K^+ content (Yun et al., 2016). Further investigations will be required to get better understanding on the relationships between rock-inhabiting bacterial communities and cations or trace elements.

Interestingly, the SO_4^{2-} concentration was the most useful variable to construct the classification model, although the correlation analysis showed that latitude and pH were the strongest indicators for rock-inhabiting microbial communities and the dominant phyla. One possible reason can be the principle that the Gini impurity index works. The Gini impurity adopted as a splitting criterion for the CART algorithm continues splitting the given data into two subgroups based on various splitting rules to make the subgroups as homogenous as possible (De'ath and Fabricius, 2000). Since the threshold SO_4^{2-} concentrations exceeding 42.97 ppm resulted in a homogeneous group (i.e., all 7 cases were *Proteobacteria*-dominant), SO_4^{2-} was ranked as the first split by the algorithm for the current dataset. However, it should be noted that SO_4^{2-} may not be the most useful variable for other datasets. Thus, continuous efforts to build a comprehensive database are required to deeply understand the relationships between rock-inhabiting bacterial communities and environmental factors.

Bacteria have been found even in the most inhospitable environments on Earth, such as deep subsurfaces, Arctic glaciers, and extremely cold deserts in Antarctica. The overall survey on rock-inhabiting bacterial communities throughout Victoria Land suggests that even under such unfavorable conditions towards living organisms, there could be certain patterns of microbial survival and bacterial community structure that can be explained based on given environmental factors. Due to the harsh conditions, such as extremely cold temperatures and high levels

of radiation, the Dry Valleys are often referred to as a Martian surface analogue. Because rocks are one of the few inhabitable places where microbes can avoid the environmental stresses of such conditions, our findings can provide insight into which environmental parameters affect microbial survival in Martian-like extreme conditions.

Acknowledgements

This research is a part of the project "Crustal evolution of Victoria Land, Antarctica and formative process of planets (20140409, PM18030)" funded by the Ministry of Ocean and Fisheries, Korea. Junho Cho was supported by Chung-Ang University Graduate Research Scholarship in 2017. We also acknowledge and thank Tae Hwan Kim and Yeong Min Kim at Seoul National University and Dr. Changhwan Oh at Korea Polar Research Institute for collecting rock samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2018.12.073>.

References

- Archer, S.D.J., de los Ríos, A., Lee, K.C., Niederberger, T.S., Cary, S.C., Coyne, K.J., Douglas, S., Lacap-Bugler, D.C., Pointing, S.B., 2017. Endolithic microbial diversity in sandstone and granite from the McMurdo Dry Valleys, Antarctica. *Polar Biol.* 40, 997–1006.
- Banerjee, M., Whitton, B.A., Wynn-Williams, D.D., 2000. Phosphatase activities of endolithic communities in rocks of the Antarctic Dry Valleys. *Microb. Ecol.* 39, 80–91.
- Banfield, J.F., Barker, W.W., Welch, S.A., Taunton, A., 1999. Biological impact on mineral dissolution: application of the lichen model to understanding mineral weathering in the rhizosphere. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3404–3411.
- Barrett, J.E., Virginia, R.A., Hopkins, D.W., Aislabie, J., Bargagli, R., Bockheim, J.G., Campbell, I.B., Lyons, W.B., Moorhead, D.L., Nkem, J.N., Sletten, R.S., Steltzer, H., Wall, D.H., Wallenstein, M.D., 2006. Terrestrial ecosystem processes of Victoria Land, Antarctica. *Soil Biol. Biochem.* 123, 93–105.
- Bottos, E.M., Scarrow, J.W., Archer, S.D.J., McDonald, I.R., Cary, S.C., 2014. Bacterial community structures of Antarctic soils. In: Cowan, D.A. (Ed.), *Antarctic Terrestrial Microbiology*. Springer-Verlag, Berlin Heidelberg, pp. 9–33.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Chapman & Hall/CRC.
- Cary, S.C., McDonald, I.R., Barrett, J.E., Cowan, D.A., 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. *Nat. Rev. Microbiol.* 8, 129–138.
- Chan, Y., Lacap, D.C., Lau, M.C.Y., Ha, K.Y., Warren-Rhodes, K.A., Cockell, C.S., Cowan, D.A., McKay, C.P., Pointing, S.B., 2012. Hypolithic microbial communities: between a rock and a hard place. *Environ. Microbiol.* 14, 2272–2282.
- Choe, Y.-H., Kim, M., Woo, J., Lee, M.J., Lee, J.I., Lee, E.J., Lee, Y.K., 2018. Comparing rock-inhabiting microbial communities in different rock types from a high arctic polar desert. *FEMS Microbiol. Ecol.* 94, fyy070.
- Cortés, U., Sánchez-Marré, M., Ceccaroni, L., R-Roda, I., Poch, M., 2000. Artificial intelligence and environmental decision support systems. *Appl. Intell.* 13, 77–91.
- Cowan, D.A., Ah Tow, L., 2004. Endangered Antarctic environments. *Annu. Rev. Microbiol.* 58, 649–690.
- Dartnell, L.R., Hunter, S.J., Lovell, K.V., Coates, A.J., Ward, J.M., 2010. Low-temperature ionizing radiation resistance of *Deinococcus radiodurans* and Antarctic Dry Valley bacteria. *Astrobiology* 10, 717–732.
- de los Ríos, A., Wierzbos, J., Ascaso, C., 2014. The lithic microbial ecosystems of Antarctica's McMurdo Dry Valleys. *Antarct. Sci.* 26, 459–477.
- De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81, 3178–3192.
- Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improved sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200.
- Faure, G., Mensing, T.M., 2010. Southern Victoria Land; basement rocks. The Transantarctic Mountains: Rocks, Ice, Meteorites and Water. Springer Science & Business Media, pp. 67–98.
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874.
- Howard-Williams, C., Hawes, I., Gordon, S., 2010. The environmental basis of ecosystem variability in Antarctica: research in the Latitudinal Gradient Project. *Antarct. Sci.* 22, 591–602.
- llumina, 2013. 16S Metagenomic sequencing library preparation guide. [WWW document]. http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.ilmn.
- Kim, O.S., Cho, Y.J., Lee, K., Yoon, S.H., Kim, M., Na, H., Park, S.C., Jeon, Y.S., Lee, J.H., Yi, H., Won, S., Chun, J., 2012. Introducing EXTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62, 716–721.
- Lee, J., Im, J., Kim, U., Löffler, F.E., 2016. A data mining approach to predict in situ detoxification potential of chlorinated ethenes. *Environ. Sci. Technol.* 50, 5181–5188.

- Makhalanyane, T.P., Valverde, A., Birkeland, N., Cary, S.C., Tuffin, I.M., Cowan, D.A., 2013. Evidence for successional development in Antarctic hypolithic bacterial communities. *ISME J.* 7, 2080–2090.
- Makhalanyane, T.P., Pointing, S.B., Cowan, D.A., 2014. Lithobionts: cryptic and refuge niches. In: Cowan, D.A. (Ed.), *Antarctic Terrestrial Microbiology*. Springer-Verlag, Berlin Heidelberg, pp. 163–179.
- Oksanen, J., Blanchet, G.B., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2017. *Vegan: Community Ecology Package*. R Package Version 2.1.4. <http://CRAN.R-project.org/package=vegan>.
- Omelson, C.R., 2008. Endolithic microbial communities in Polar Desert habitats. *Geomicrobiol J.* 25, 404–414.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Horn, D.J.V., Webber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.
- Selbmann, L., de Hoog, G.S., Zucconi, L., Isola, D., Ruisi, S., van den Ende, A.H.G., Ruibal, C., De Leo, F., Urzì, C., Onofri, S., 2008. Drought meets acid: three new genera in dothidealean clade of extremotolerant fungi. *Stud. Mycol.* 61, 1–20.
- Smith, J.L., Barrett, J.E., Tusnády, G., Rejtő, L., Cary, S.C., 2010. Resolving environmental drivers of microbial community structure in Antarctic soils. *Antarct. Sci.* 22, 673–680.
- Sokol, E.R., Herbold, C.W., Lee, C.K., Cary, S.C., Barrett, J.E., 2013. Local and regional influences over soil microbial metacommunities in the Transantarctic Mountains. *Ecosphere* 4, art136.
- Ugolini, F.C., Bockheim, J.G., 2008. Antarctic soils and soil formation in a changing environment: a review. *Geoderma* 144, 1–8.
- Uroz, S., Calvaruso, C., Turpault, M.-P., Frey-Klett, P., 2009. Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends Microbiol.* 17, 378–387.
- van Goethem, M.W., Makhalanyane, T.P., Valverde, A., Cary, S.C., Cowan, D.A., 2016. Characterization of bacterial communities in lithobionts and soil niches from Victoria Valley, Antarctica. *FEMS Microbiol. Ecol.* 92, fiw051.
- Vishniac, H.S., 1993. The microbiology of Antarctic soils. In: Friedmann, E.I. (Ed.), *Antarctic Microbiology*. Wiley-Liss, New York, pp. 297–341.
- Wierzchos, J., Ascaso, C., 2002. Microbial fossil record of rocks from the Ross Desert, Antarctica: implication in the research for past life on Mars. *Int. J. Astrobiol.* 1, 51–59.
- Williams, P.F., Hobbs, B.E., Vernon, R.H., Anderson, D.E., 1971. The structural and metamorphic geology of basement rocks in the McMurdo sound area, Antarctica. *J. Geol. Soc. Aust.* 18, 127–142.
- Wynn-Williams, D.D., Edwards, H.G.M., 2000. Antarctic ecosystems as model for extraterrestrial surface habitats. *Planet. Space Sci.* 48, 1065–1075.
- Yoo, K., Shukla, S.K., Ahn, J.J., Oh, K., Park, J., 2016. Decision tree-based data mining and rule induction for identifying hydrogeological parameters that influence groundwater pollution sensitivity. *J. Clean. Prod.* 122, 277–286.
- Yoo, K., Yoo, H., Lee, J.M., Shukla, S.K., Park, J., 2018. Classification and regression tree approach for prediction of potential hazards of urban airborne bacteria during Asian Dust events. *Sci. Rep.* 8, 11823.
- Yun, Y., Wang, H., Man, B., Xiang, X., Zhou, J., Qiu, X., Duan, Y., Engel, A.S., 2016. The relationship between pH and bacterial communities in a single karst ecosystem and its implication for soil acidification. *Front. Microbiol.* 7, 1955.
- Yung, C.C.M., Chan, Y., Lacap, D.C., Pérez-Ortega, S., de los Rios-Murillo, A., Lee, C.K., Cary, S.C., Ponting, S.B., 2014. Characterization of chasmoendolithic community in Miers Valley, McMurdo Dry Valleys, Antarctica. *Microb. Ecol.* 68, 351–359.