Electronic Journal of Biotechnology 45 (2020) 30–37

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research article

Next generation sequencing and microbiome's taxonomical characterization of frozen soil of north western Himalayas of Jammu and Kashmir, India

Verruchi Gupta [a], Inderpal Singh [a,b], Shafaq Rasool [a], Vijeshwar Verma [a,b,*]

[a] School of Biotechnology, Shri Mata Vaishno Devi University, Kakryal, Katra, Jammu and Kashmir, India
[b] Bioinformatics Infrastructure Facility (BIF), School of Biotechnology, Shri Mata Vaishno Devi University, Kakryal, Katra, Jammu and Kashmir, India

ARTICLE INFO

Article history:
Received 11 September 2019
Accepted 18 March 2020
Available online 24 March 2020

Keywords:
Annotation
Bioinformatics
Frozen soil
Glacier
Himalayas
Metagenomics
Microbial biodiversity
Microbial taxonomy
Microbiome
Next generation sequencing
Taxonomical characterization

ABSTRACT

Background: Traditionally, microbial genome sequencing has been restrained to the species grown in pure culture. The development of culture-independent techniques over the last decade allows scientists to sequence microbial communities directly from environmental samples. Metagenomics is the study of complex genome by the isolation of DNA of the whole community. Next generation sequencing (NGS) of metagenomic DNA gives information about the microbial and taxonomical characterization of a particular niche. The objective of the present research is to study the microbial and taxonomical characterization of the metagenomic DNA, isolated from the frozen soil sample of a glacier in the north western Himalayas through NGS.
Results: The glacier community comprised of 16 phyla with the representation of members belonging to Proteobacteria *and* Acidobacteria. The number of genes annotated through the Kyoto Encyclopedia of Genes and Genomes (KEGG), GO, Pfam, Clusters of Orthologous Groups of proteins (COGs), and FIG databases were generated by COGNIZER. The annotation of genes assigned in each group from the metagenomics data through COG database and the number of genes annotated in different pathways through KEGG database were reported.
Conclusion: Results indicate that the glacier soil taken in the present study, harbors taxonomically and metabolically diverse communities. The major bacterial group present in the niche is Proteobacteria followed by Acidobacteria, and Actinobacteria, etc. Different genes were annotated through COG and KEGG databases that integrate genomic, chemical, and systemic functional information.
How to cite: Gupta V, Singh I, Rasool S, et al. Next Generation sequencing and microbiome's taxonomical characterization of frozen soil of North Western Himalayas of Jammu and Kashmir, India. Electron J Biotechnol 2020;45. https://doi.org/10.1016/j.ejbt.2020.03.003.

1. Introduction

More than 10% of the Earth's land surface is covered with glacial ice [1]. Microbes harbor these huge reservoirs, such as glaciers, icebergs, and glacial masses, with significant biodiversity and activities in the frozen environments of the Earth. Extreme conditions at subzero temperatures had earlier been considered to exist without life, or only as repositories for wind-transported microorganisms trapped in the ice [2]. However, now microbes have been found to be the major component of ecosystems of low to high temperatures providing conditions for their growth and survival [3]. A study conducted in gigantic reservoirs of microbial biodiversity in the Antarctic and Greenland glaciers report 9.61×10^{25} microbes [4]. Microbial biodiversity present at low temperature plays a major role in soil development and other biochemical processes therein [5]. Novel findings of industrially important bioactive molecules like enzymes, antibiotics, etc. can be produced from the biodiversity of microbes present in extreme niches [6]. Recently, next generation sequencing (NGS) has developed as a powerful technique for analyzing the complex samples for taxonomical biodiversity and studying metabolic pathways [7]. Statistical or computational tools and databases have been developed to study and manage huge metagenomic data for faster and detailed genomic or genetic profiling of environmental samples at a very affordable cost [8,9,10]. Currently, a metagenomic approach is widely used to find microbial biodiversity in extreme and unique niches of the environment along with the relative abundance of genes, biochemical processes, and metabolic pathways [11,12,13]. Studies showed that bacterial groups are being prominently found in

* Corresponding author.
E-mail address: verma211@gmail.com (V. Verma).
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

https://doi.org/10.1016/j.ejbt.2020.03.003

the remote area of the Canadian high arctic [14], and 10 bacterial genera have been identified based on 16S ribotyping from permafrost samples of an Arctic site [15,16]. These studies clearly revealed the presence of vast reservoirs of microbial life and communities in extreme environments like glaciers, snow ice, and ice bergs, ice caps which impact the dynamics of glacial world, and play a crucial role in soil formation and other biogeochemical processes [17,18,19]. In the present study, the whole taxonomical microbial community structure of a frozen soil sample from a glacier of the North western Himalayas (NWH) of Jammu and Kashmir (J&K) (India), up to species level along with the functional annotations of the predicted genes has been analyzed based on NGS by using the Kyoto Encyclopedia of Genes and Genomes (KEGG), GO, Clusters of Orthologous Groups of proteins (COGs), FIG, and PFAM databases.

2. Materials and Methods

2.1. Sample collection

A soil sample of 10–100 g each was collected 1–1.5 ft deep under the surface in triplicate in the month of October, 2015 (autumn) from the glacier of NWH of J&K, India. The temperature of the glacier was -20°C. The height of the glacier is 4700 m above the sea level. The longitude and latitude of the glacier is 34° 9′ 49″ N, 75°19′ 49″ E. The soil was moist and it was collected in sterile bags. The samples were carried in airtight sterile bags to the SMVDU and stored in 4°C till further use.

2.2. Preparation of soil metagenome

A soil sample of 1–2 g, in triplicate, was used for the extraction of DNA. Total soil DNA was extracted using ultrapure ultraclean Mega preparation soil DNA kit (Mobio Laboratory, USA) as per the manufacturer's instructions and then precipitated overnight with 5 M NaCl and ethanol at -20°C. The sample was centrifuged at 13,500 × g for 30 min. The pellet was washed with 70% ethanol twice by centrifuging at 13,500 × g for 10 min. Then, the pellet was air dried, dissolved in 1× TE, and analyzed on 0.8% agarose gel.

2.3. Quality control of DNA, its amplification, sequencing, and NGS analysis

Agarose gel (1%) was run for accessing the quality of the genomic DNA (3 μl) at 110 V for 30 min using 1 × TAE buffer. A DNA sample of 1 μl was used for determining the concentration using Qubit® 2.0 fluorometer as per manufacturer's protocol. Truseq Nano DNA library preparation kit was used for the preparation of the paired-end sequencing library from 200 ng of g-DNA. Covaris S2 system is the instrument used for shearing DNA. Ultrasonication was used for the mechanical shearing of the g-DNA into smaller fragments. The fragment size is ~300 bp. Following this, continuous end repairing was done by ligating 'A' at the 3′ ends of DNA fragments to allow platform-specific adapters to be ligated to both ends of DNA fragments. These adapters contain sequences that are essential for binding the dual-barcoded libraries to a flow cell for sequencing, and thus allow the PCR amplification of adapter-ligated fragments and binding the standard Illumina sequencing primers. Through HiFi PCR master mix, a high-fidelity amplification of fragments was performed for attaining maximum yield from the starting DNA with limited quantity. The DNA thus amplified was analyzed through Bioanalyzer 2100 (Agilent Technologies) using high sensitivity DNA chip as per manufacturer's guide. DNA concentration of the library was again analyzed through Qubit and mean peak size was analyzed through a Bioanalyzer. The library was subsequently loaded on the Illumina platform for clustering and paired-end sequencing of the templates in both forward and reverse directions.

As this was a metagenomic DNA library, thus *de novo* assembly of high quality paired end reads was carried out through the metaSPADES (v3.11.1) package using default parameters and scaffolds were generated post adapter trimming and the removal of low quality reads based on the Phred score, which was over 30 [20]. Genes were predicted through the Prodigal tool (v2.6.3) using the metagenomics mode at default parameters [21]. Kaiju, metagenomic classifier was used to calculate the taxonomic abundance from predicted genes [22]. The functional annotation assignment of predicted genes from the metagenomic data was done using the COGNIZER tool [23]. In house Perl scripts were also used for the analysis of NGS data.

3. Results

3.1. Qualitative and quantitative analysis of g-DNA

The quality of the DNA was evaluated first through running it at 1% agarose gel at 100 V for 30 min (Fig. 1). Quantification using Qubit Fluorometer revealed DNA concentration 292 ng/μl and a yield of 5.84

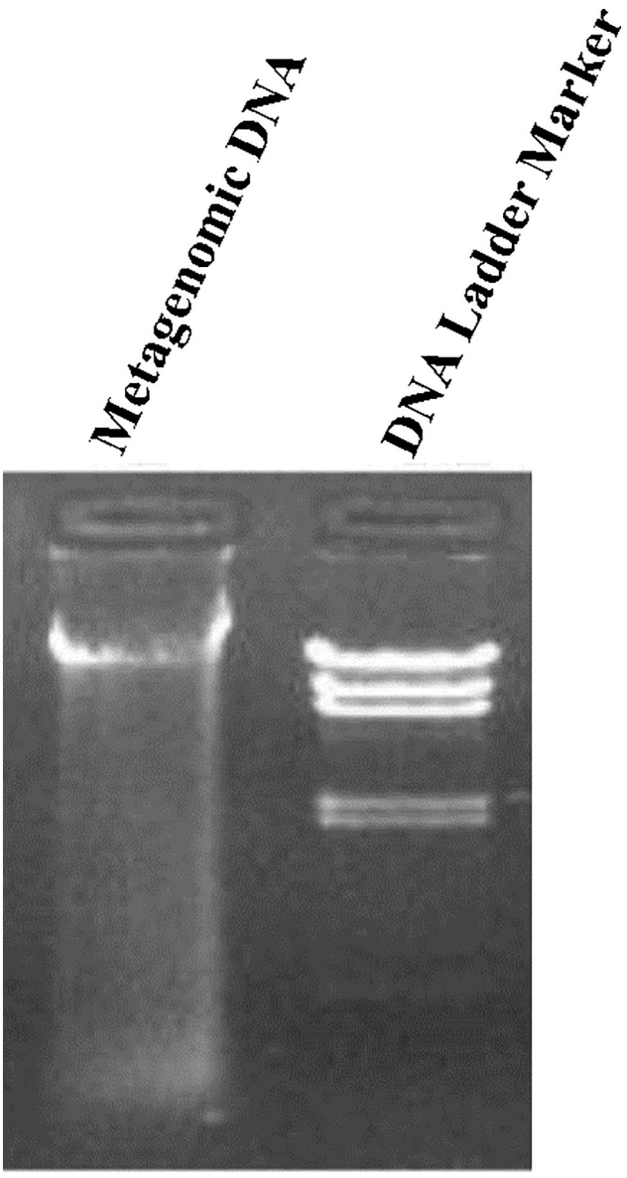

Fig. 1. Agarose gel (1%) showing the extraction of metagenomic DNA. Lane1: Metagenomic DNA isolated from soil sample of Kolahoi Glacier and Lane 2: λ *Hind* III marker.

μg. The analysis of the library thus amplified through Bioanalyzer 2100 (Agilent Technologies) (Fig. S1).

3.2. Metagenome assembly and gene prediction

The metagenomic DNA library was prepared from the sample to prepare the average size of library 453 bp. The library was sequenced through Illumina platform (2 × 150 bp chemistry) to generate ~9 GB data. Following this, the *de novo* assembly of the high quality paired end reads was done using metaSPADES (v3.11.1) [20]. Fasta sequences were then generated from the assembly (**Fig. S2**). Statistics of the scaffolds' length distribution was calculated. Thus, the scaffolds generated were subjected to gene prediction using the Prodigal Tool (v2.6.3) with default parameters [21]. A total of 256,699 scaffolds were generated for soil sample, which were subjected to genes prediction by Prodigal (v2.6.3) at the metagenome mode (-meta option). Prodigal is a widely used gene prediction program that can identify genes in short, anonymous coding sequences occur with a high degree of accuracy. The novel value of the method consists of enhanced translation initiation site identification, the ability to identify sequences that use alternate genetic codes, and confidence values for each gene. These predicted genes were then taken further for taxonomic and functional analysis. Fasta sequences of genes (and its corresponding protein as well as the GFF file) predicted were also generated. Total 395,119 genes were predicted, total gene size was 183,172,093 NT and an average length of genes was 446 NT (Fig. S3).

3.3. Accession number

The nucleotide sequence of the whole NGS data has been submitted in SRA and NCBI, and the accession no is **PRJNA543600**.

3.4. Taxonomic classification and abundance

Taxonomic classification of the metagenomic DNA at phylum, class, order, family, genus, and species level was done through the Kaiju tool [22]. It is a fast and freely accessible tool for classifying the metagenome's taxonomy. Its algorithm searches for exact protein matches using a set of reference databases containing the annotated protein sequences. Kaiju utilizes complete genome databases from the NCBI Refseq NCBI Blast etc. Based on the alignment and identification of identical/homologous protein sequences, the classification is completed. Accordingly, the taxonomic identifier is linked to the following scaffold at various levels of taxonomy. Genes predicted from the Prodigal tool were taken as an input for the Kaiju tool. A total of 202,669 genes from the total of 395,119 genes are having annotation by at least one database. Through the classification, it was observed that most abundant taxons from the metagenomic sample are the Proteobacteria followed by the Acidobacteria (Fig. 2).

Through the analysis of the taxonomic abundance at the Phylum level, it was observed that the Proteobacteria (with 99,213 genes) were the most abundant phylum followed by Acidobacteria (with 45,027 genes), Actinobacteria (with 17,640 genes), Verrucomicrobia (with 16,016 genes), Chloroflexi (with 8727 genes) and so on (top five most abundant phyla are reported above (Fig. 3), for a more comprehensive bar plot of phylum abundance at phylum level reporting top 30 most abundant Phyla).

Furthermore, through the analysis of the taxonomic abundance at the class level, it was inferred that the Alphaproteobacteria (with 44,865 genes) were the most abundant class followed by Betaproteobacteria (with 27,540 genes), and the Gammaproteobacteria (with 14,602 genes) in the 30 most abundant classes observed (**Fig. S4**).

Digging deep into the analysis, taxonomic abundance at the Order (**Fig. S5**) and Family levels (Fig. 4) revealed that the Rhizobiales (with 30,305 genes), Acidobacteriales (with 13,623 genes), Burkholderiales (with 11,372 genes), Rhodospirillales (with 5137 genes), and the Myxococcalles (with 4489 genes) were the top five most abundant orders observed in top 30 abundant orders while, *Bradyrhizobiaceae* (with 18,312 genes), *Acidobacteriaceae* (with 7319 genes), *Nitrospiraceae*, (with 3540 genes), *Burkholderiaceae* (with 3329 genes), and *Comamonadaceae* (with 3264 genes) were the most abundant families observed. At the genus level (Fig. 5), most abundant genera observed were *Bradyrhizobium* (with 14,096 genes), *Candidatus entotheonella* (with 3881 genes), *Nitrospira* (with 3385 genes), *Candidatus solibacter* (with 2479 genes), *Gemmatimonas* (with 1751 genes), *Ktedonobacter* (with 1674 genes), *Candidatus koribacter* (with 1654 genes), and *Streptomyces* (with 1618 genes) in the top 30 most abundant genera. The analysis of taxonomic abundance at species level led to the identification of the *Bradyrhizobium erythrophlei* (with 3327 genes), Actinobacteria *bacterium* (with 3247 genes), Acidobacterium *bacterium* (with 2869 genes), *Acidobacteriales bacterium* (with 2469 genes), *Candidatus solibacter usitatus* (with 2479

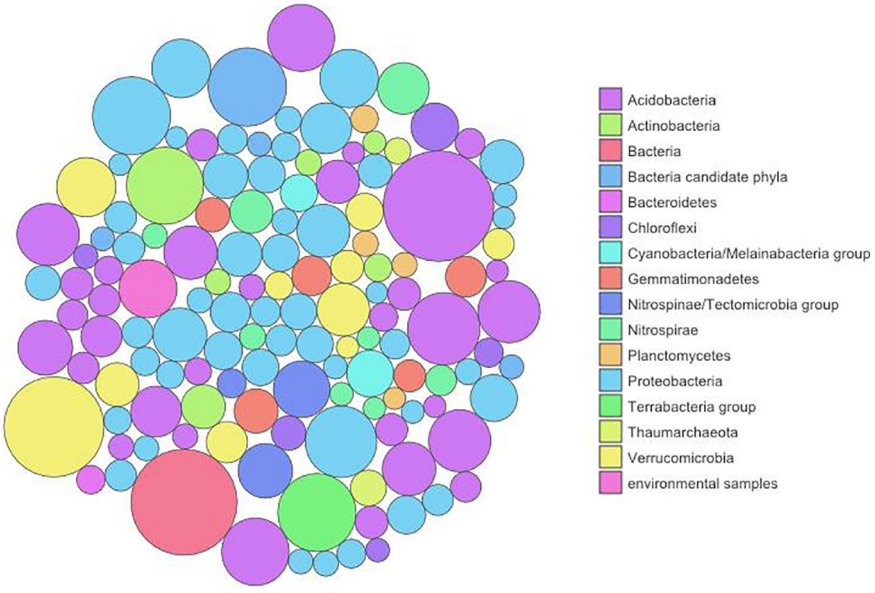

Fig. 2. A bubble plot showing the relative taxonomic abundance in the metagenomic sample. Bubble size indicates taxon abundance relative to its maximum abundance (largest bubble size). The size of the circle is scaled logarithmically to represent the number of sequences assigned directly to the taxon.

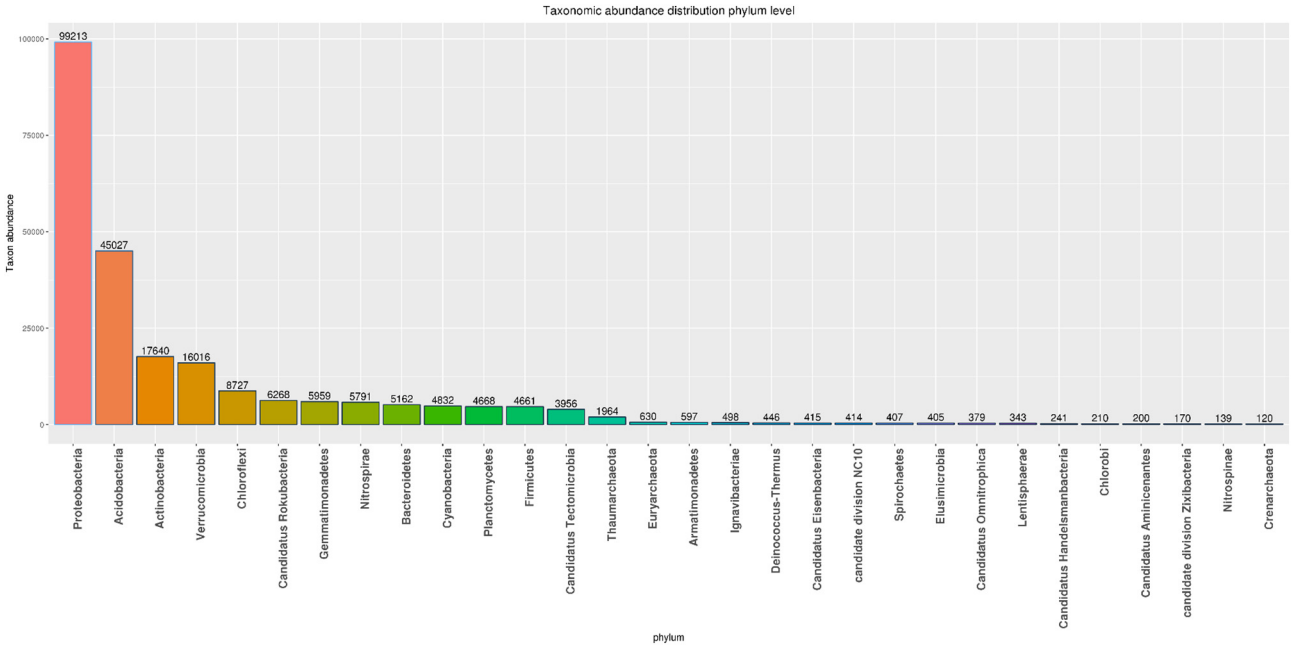

Fig. 3. A bar chart showing the taxonomic abundance of metagenomics sample at phylum level.

genes) as top five most abundant species in the top 30 most abundant species observed in the presented metagenomic sample (**Fig. S6**). *Bradyrhizobium erythrophlei* is a bacterium from the genus of *Bradyrhizobium,* which has been isolated from the nodules of the tree *Erythrophleum fordii* [24]. Actinobacteria *bacterium* are of great economic importance to humans as agriculture and forests depend on their contributions to soil systems [25]. Acidobacterium *bacteria* can be found in a variety of environments including soil, hot springs, oceans, caves, and metal-contaminated soils [26]. *Candidatus solibacter usitatus* produces enzymes to break down the organic carbon available in its environment for metabolism and participates in nitrate and nitrite reduction resulting in the heterogeneous nutrient deposition throughout the soil where this species inhabits [27].

3.5. Functional annotation

After detailed taxonomical analysis, the predicted genes derived from the Prodigal Tool were input to the COGNIZER tool, a fast standalone package, which utilizes multiple annotation databases such as COGs, KEGG, Gene Ontology, protein families database (Pfam), and FIGfam (proteins to be iso functional homologs) for a comprehensive

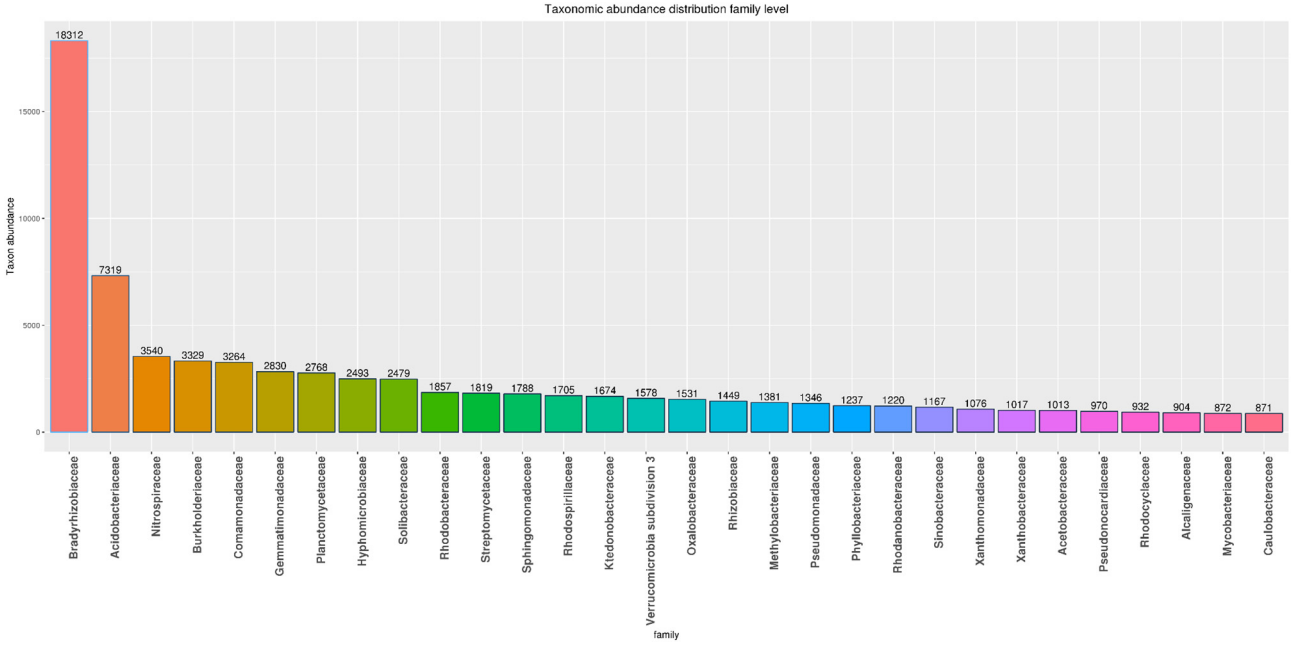

Fig. 4. A bar chart shows the taxonomic abundance at family level. From the figure it can be inferred that *Bradyrhizobiaceae* (with 18,312 genes) is the most abundant family followed by *Acidobacteriaceae* (with 7319 genes). This distribution plot is made from top thirty families.

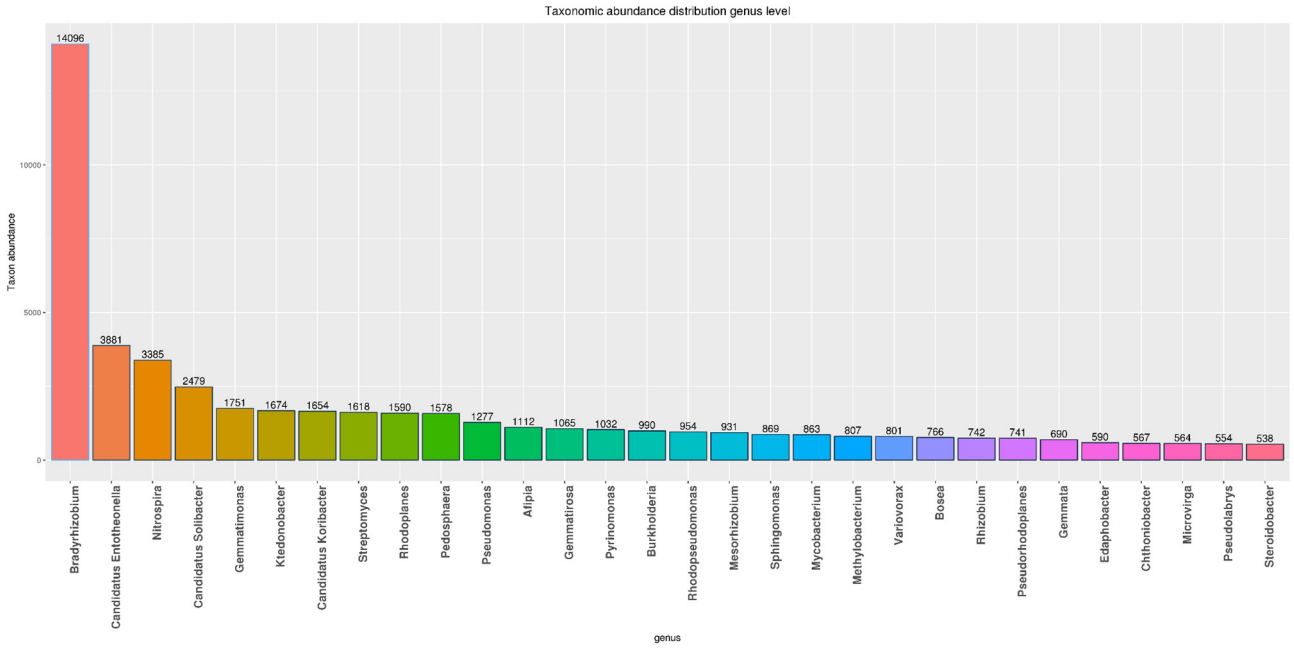

Fig. 5. A bar chart shows taxonomic abundance at genus level. From the figure it can be inferred that *Bradyrhizobium* (with 14,096 genes) is the most abundant genus followed by *Candidatus Entotheonella* (with 3881 genes).

functional annotation of individual sequences from metagenomic data (**Fig. S7**) [23,28,29,30,31,32].

3.6. Clusters of Orthologous Groups of proteins (COGs) annotation

COGs attempts the phylogenetic classification of proteins based on proteins encoded in a set of complete genomes of archaea, bacteria, and eukaryotes. COG database helps in the rapid functional characterization of the role of one microorganism in its community; in addition, it is much smaller than the NCBI nonredundant. Through exhaustive comparison of the predicted proteins from the metagenomic data, COGs were constructed, which represented the annotated genes clustered under different categories of function (Fig. 6).

3.7. KEGG pathway annotation

KEGG is a database of the manually curated network pathways representing various cellular processes like Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, and Organismal System. KEGG Functional analysis showed that 250,808 genes have been assigned with 4617 KEGG classes. Maximum KEGG Orthology's (KO) assigned were belonging to

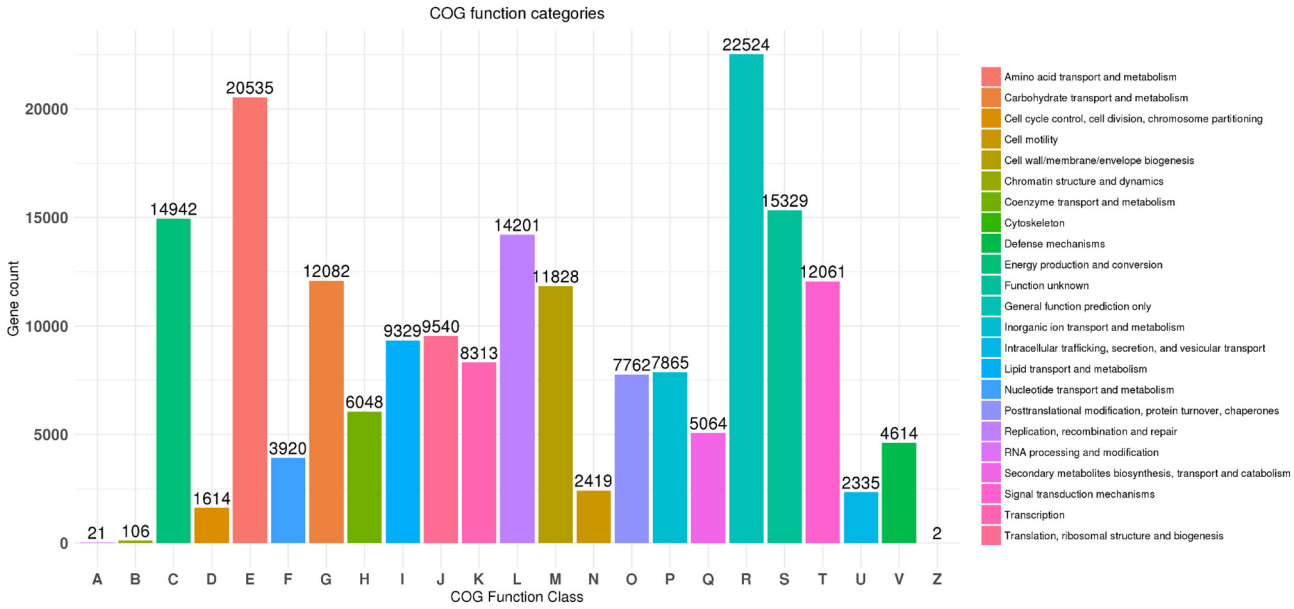

Fig. 6. COG-based annotation of genes. The Clusters of Orthologous Groups of proteins (COGs) is an attempt on a phylogenetic classification of the proteins encoded in 21 complete genomes of bacteria, archaea, and eukaryotes (http://www.ncbi.nlm.nih.gov/COG). The COGs were constructed by applying the criterion of consistency of genome-specific best hits to the results of an exhaustive comparison of all protein sequences from these genomes. The figure represents genes annotated under different functional categories found under different orthologous groups of COG database.

the Metabolism category followed by Environmental Information Processing (**Fig. S8**). The dominant pathways are serine or threonine protein kinase, bacterial [EC: 2.7.11.1], and adenylate cyclase [EC: 4.6.1.1] having 2362 and 1732 gene hits, respectively.

4. Discussion

Advancement in DNA sequencing technologies in the interpretation of sequencing data at an economical price by using the bioinformatic tools is now widely used [33]. Collection of soil samples, extraction of their metagenomic DNA, and their data analysis are the three most important factors for the metagenomic analysis [34,35]. Bioinformatic analysis of the NGS data using Kaiju software identified 16 different phyla in which the Proteobacteria was found abundantly and followed by Acidobacteria, Verrucomicrobia, Thaumarchaeota, Actinobacteria, etc. Similar findings have been reported by authors [36,37] from the cold environments in which Proteobacteria (Alpha-, Beta-, and Gamma - proteobacteria), Bacteroides, and Actinobacteria are the most common bacterial phyla found as compared to Archaea. Literature surveys of metagenomic studies show that the *Proteobacteria*, *Acidobacteria,* and *Actinobacteria* were dominant bacterial phyla at high altitude soil samples, while the *Bacteroidetes* and *Fermicutes* were dominant at low altitude soils [38]. Besides, high altitude samples have reported more of *Alpha – Proteobacteria,* while the increased presence of *Beta-proteobacteria* is reported at low altitudes [38]. *Cytophaga, Flavobacterium, and Bacteroides* (CFB) have also been reported in high concentrations at high altitudes [38]. Actinobacteria, Acidobacteria, and Proteobacteria have also been reported as the dominant phyla found in the active layer and the permafrost communities [39,40,41,42,43,44] playing crucial roles in the decomposition of organic matter and nutrient cycling in polar ecosystems with limited trophic complexity [45]. Bacterial phyla: *Bacteriodetes (CFB cluster), Actinobacteria, Firmicutes*, and *Proteobacteria,* including representatives from the *Alpha-, Beta-*, and *Gamma – Proteobacteria* classes have been reported earlier in glaciers, ice, and super cooled cloud droplets [15,46]. Microbial abundance in snow has been reported as a result of the postdepositional processes [47]. In the Canadian High Arctic samples, the Alpha - Proteobacteria and *Methylocystaceae* family may play a crucial role in methyl and methane oxidation in the active layer [5,48,49]. Besides, more complexity and diversity of soil bacterial communities have been reported from high altitudes [50]. The Alpha-Proteobacteria are widespread in natural environments including cold regions and are adapted to the oligotrophic lifestyle assimilating a wide range of organic compounds [51]. Many ecosystems have been reported to contain similar dominant bacterial communities [52,53,54]. For example, *Acidobacteria, Alpha - Proteobacteria, Actinobacteria, Beta – Proteobacteria, and Gamma- Proteobacteria* have been reported that more than 75% of the bacteria are found in the Changbai Mountain, in the region of northeastern China [54]. *Acidobacteria* and *Proteobacteria* are to be major representatives in a mountainous forest in eastern Peru [52]; *Proteobacteria* represented more than 50% of the soil bacterial community in the Yuanyang Lake ecosystem in Taiwan [55]; *Proteobacteria*, *Acidobacteria*, *Actinobacteria,* and *Bacteroidetes* are the major phyla reported from Mount Fuji in Japan [56]. Studies carried out in the western Himalayas, India report Psychrophilic and Psychrotolerant bacteria, which are abundantly available in the Himalayan habitats (60%) [57]. Lower taxonomical characterization shows that *Bradyrhizobium* is the most abundant genus followed by *Candidatus* and *Nitrospira,* etc. Contrary to this, the genus *Polaromonas* has been reported to be plentiful in the supraglacial area of the Arctic [19] and Antarctic glaciers [58]. Besides, Arctic and Antarctic glaciers possess the pigmented algae in large scale such algae are *Chlamydomonas*, *Chloromonas*, *Raphidonema*, and *Chrysophyceae* [59,17] and Fungi, particularly Basidiomycetous yeasts and Chytridiomycota are also found in abundance [17,60].

The present study conducted at the geography of Pangong lake, Ladakh [61] reported bacteria (83.86%), archaea (0.24%), eukaryotes (0.42%), viruses (0.41%), and unclassified (15.02%). The major phyla reported and represented were *Proteobacteria, Bacteroidetes*, *Firmicutes*, *Actinobacteria, Balneolaeota*, *Cyanobacteria, Verrucomicrobia, Euryarchaeota*, *Planctomycetes* and *Ascomycota*. However, *Methylophaga* was the most abundant genus found in the Pangong lake [61]. Thus, results for the present study are in conformity with the earlier reports up to the phyla level; however, bacterial communities reported from the cold samples of arctic, permafrost, and other such niches differ at taxa level.

5. Conclusion

This is the maiden NGS analysis reported from the glacier soil of the North Western Himalayas, J&K, taken in the present study, which harbors taxonomically and metabolically diverse communities. This study will open vistas for the detailed NGS studies of the other glaciers of the NWH of J&K. The data obtained were used for understanding the biome diversity and presence of various genes in this important ecological niche. In the NGS, the community was found to be comprised of 16 phyla with the representation of members belonging to Proteobacteria, Acidobacteria, Verrumicrobia, Thaumarchaeota, Actinobacteria, Bacteriodetes, and Chloroflexi, etc. At the genus level, *Bradyrhizobium* was the most abundant as compared to other genera like *Candidatus entotheonella*, *Nitrospora*, *Candidatus solibacter*, *Gemmatimonas,* and *Ktedonobacter.* The number of genes annotated through KEGG, GO, Pfam, COG, and FIG databases were generated by COGNIZER. The annotation of genes assigned in each group from the metagenomics data through COG database and the number of genes annotated in different pathways through KEGG database was also reported successfully.

Financial support

The work was supported by the Department of Biotechnology, Govt. of India (File no: BT/PR12632/BPA/118/51/2014), which is highly acknowledged for the present study.

Conflict of interest

The authors of this original manuscript declare that they have no conflict of interest, and express their consent for the publication process.

Acknowledgments

We gratefully acknowledge Dr. Anurag Kumar, assistant professor, School of Languages and Literature, SMVDU, Katra for proofreading the manuscript. BIF centre, SMVDU, Katra is also acknowledged for the bioinformatics analysis of the NGS data.

Supplementary material

https://doi.org/10.1016/j.ejbt.2020.03.003

References

[1] Choudhari S, Smith S, Owens S, et al. Metagenome sequencing of prokaryotic microbiota collected from Byron Glacier, Alaska. Genome Announc 2013;1(2):e0009913. https://doi.org/10.1128/genomeA.00099-13 PMid: 23516217.
[2] Cowan DA, Tow LA. Endangered Antarctic environments. Annu Rev Microbiol 2004; 58:649–90. https://doi.org/10.1146/annurev.micro.57.030502.090811 PMid: 15487951.
[3] Rawat N, Joshi GK. Bacterial community structure analysis of a hot spring soil by next generation sequencing of ribosomal RNA. Genomics 2018;111(5):1053–8. https://doi.org/10.1016/j.ygeno.2018.06.008 PMid: 31533897.

[4] Priscu JC, Christner BC. Earth's icy biosphere. In: Bull AT, editor. Microbial diversity and bioprospecting. Washington, DC: ASM Press; 2004. p. 130–45. https://doi.org/10.1128/9781555817770.ch13.
[5] Schutte UM, Abdo Z, Foster J, et al. Bacterial diversity in a glacier foreland of the high Arctic. Mol Ecol 2010;19(Suppl. 1):54–66. https://doi.org/10.1111/j.1365-294X.2009.04479.x PMid: 20331770.
[6] Kumar M, Yadav AN, Tiwari R, et al. Deciphering the diversity of culturable thermotolerant bacteria from Manikaran hot springs. Ann Microbiol 2014;64(2): 741–51. https://doi.org/10.1007/s13213-013-0709-7.
[7] Chan CS, Chan KG, Tay YL, et al. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. Front Microbiol 2015;6:177. https://doi.org/10.3389/fmicb.2015.00177 PMid: 25798135.
[8] Rondon MR, August PR, Bettermann AD, et al. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 2000;66(6):2541–7. https://doi.org/10.1128/AEM.66.6.2541-2547.2000 PMid: 10831436.
[9] von Mering C, Jensen LJ, Snel B, et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res 2005;33 (Suppl1):D433–7. https://doi.org/10.1093/nar/gki005 PMid: 15608232.
[10] Oulas A, Pavloudi C, Polymenakou P, et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinform Biol Insights 2015;9:75–88. https://doi.org/10.4137/BBI.S12462 PMid: 25983555.
[11] Simon C, Wiezer A, Strittmatter AW, et al. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. Appl Environ Microbiol 2009;75(23): 7519–26. https://doi.org/10.1128/AEM.00946-09 PMid: 19801459.
[12] Frank-Fahle BA, Yergeau E, Greer CW, et al. Microbial functional potential and community composition in permafrost-affected soils of the NW Canadian Arctic. PLoS One 2014;9(1):e84761. https://doi.org/10.1371/journal.pone.0084761 PMid: 24416279.
[13] von Mering C, Hugenholtz P, Raes J, et al. Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 2007;315(5815): 1126–30. https://doi.org/10.1126/science.1133420 PMid: 17272687.
[14] Skidmore M, Anderson SP, Sharp M, et al. Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. Appl Environ Microbiol 2005;71(11):6986–7. https://doi.org/10.1128/AEM.71.11.6986-6997.2005 PMid: 16269734.
[15] Amato A, Hennebelle R, Magand O, et al. Bacterial characterization of the snow cover at Spitzberg, Svalbard. FEMS Microbiol Ecol 2007;59(2):255–64. https://doi.org/10.1111/j.1574-6941.2006.00198.x PMid: 17328766.
[16] Amato A, Kooistra WH, Ghiron JH, et al. Reproductive isolation among sympatric cryptic species in marine diatoms. Protist 2007;158(2):193–207. https://doi.org/10.1016/j.protis.2006.10.001 PMid:17145201.
[17] Anesio AM, Lutz S, Chrismas NAM, et al. The microbiome of glaciers and ice sheets. NPJ Biofilms Microbiomes 2017;3:10. https://doi.org/10.1038/s41522-017-0019-0 PMid: 28649411.
[18] Edwards A. Coming in from the cold: potential microbial threats from the terrestrial cryosphere. Front Earth Sci 2015;3:12. https://doi.org/10.3389/feart.2015.00012.
[19] Garcia-Lopez E, Rodriguez-Lorente I, Alcazar P, et al. Microbial communities in coastal glaciers and tidewater tongues of Svalbard Archipelago, Norway. Front Mar Sci 2019;5:512. https://doi.org/10.3389/fmars.2018.00512.
[20] Nurk S, Meleshko D, Korobeynikov A. metaSPAdes: A new versatile metagenomic assembler. Genome Res 2017;27(5):824–34. https://doi.org/10.1101/gr.213959.116 PMid: 28298430.
[21] Hyatt D, Chen GL, Locascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119. https://doi.org/10.1186/1471-2105-11-119 PMid: 20211023.
[22] Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju, 7; 2016; 11257 https://doi.org/10.1038/ncomms11257 PMid: 27071849.
[23] Bose T, Haque MM, Reddy C, et al. COGNIZER: A framework for functional annotation of metagenomic datasets. PLoS One 2015;10(11):e0142102. https://doi.org/10.1371/journal.pone.0142102 PMid: 26561344.
[24] Yao Y, Sui XH, Zhang XX, et al. *Bradyrhizobium erythrophlei* sp. nov. and *Bradyrhizobium ferriligni* sp. nov., isolated from effective nodules of *Erythrophleum fordii*. Int J Syst Evol Microbiol 2015;65(6):1831–7. https://doi.org/10.1099/ijs.0.000183 PMid: 25754551.
[25] Barka EA, Vatsa P, Sanchez L, et al. Taxonomy, physiology, and natural products of *Actinobacteria*. Microbiol Mol Biol Rev 2016;80(1):1–43. https://doi.org/10.1128/MMBR.00019-15 PMid: 26609051.
[26] Quaiser A, Ochsenreiter T, Lanz C, et al. Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. Mol Microbiol 2003;50(2):563–75. https://doi.org/10.1046/j.1365-2958.2003.03707.x PMid: 14617179.
[27] Ward NL, Challacombe JF, Janssen PH, et al. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol 2009;75(7):2046–56. https://doi.org/10.1128/AEM.02294-08 PMid: 19201974.
[28] Finn RD, Bateman A, Clements J, et al. Pfam: The protein families database. Nucleic Acids Res 2014;42(D1):D222–30. https://doi.org/10.1093/nar/gkt1223 PMid: 24288371.
[29] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27 PMid: 10592173.
[30] Meyer F, Overbeek R, Rodriguez A. FIGfams: Yet another set of protein families. Nucleic Acids Res 2009;37(20):6643–54. https://doi.org/10.1093/nar/gkp698 PMid: 19762480.
[31] Tatusov RL, Galperin MY, Natale DA, et al. The COG database: A tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 2000;28(1): 33–6. https://doi.org/10.1093/nar/28.1.33 PMid: 10592175.
[32] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: Tool for the unification of biology. Nat Genet 2000;25(1):25–9. https://doi.org/10.1038/75556 PMid: 10802651.
[33] Kircher M, Kelso J. High-throughput DNA sequencing—Concepts and limitations. Bioessays 2010;32(6):524–36. https://doi.org/10.1002/bies.200900181 PMid: 20486139.
[34] Bakken LR. Separation and purification of bacteria from soil. Appl Environ Microbiol 1985;49(6):1482–7. https://doi.org/10.1128/AEM.49.6.1482-1487.1985 PMid: 16346816.
[35] Lombard N, Prestat E, van Elsas JD, et al. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. FEMS Microbiol Ecol 2011;78 (1):31–49. https://doi.org/10.1111/j.1574-6941.2011.01140.x PMid: 21631545.
[36] Yergeau E, Hogues H, Whyte LG, et al. The functional potential of high Arctic permafrost revealed by metagenomic sequencing, qPCR and microarray analyses. ISME J 2010;4(9):1206–14. https://doi.org/10.1038/ismej.2010.41 PMid: 20393573.
[37] Lewin A, Wentzel A, Valla S. Metagenomics of microbial life in extreme temperature environments. Curr Opin Biotechnol 2013;24(3):516–25. https://doi.org/10.1016/j.copbio.2012.10.012 PMid: 23146837.
[38] Kumar S, Suyal DC, Yadav A, et al. Microbial diversity and soil physiochemical characteristic of higher altitude. PLoS ONE 2019;14(3):e0213844. https://doi.org/10.1371/journal.pone.0213844 PMid: 30875404.
[39] Zhou J, Davey ME, Figueras JB, et al. Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. Microbiology 1997;143(12):3913–9. https://doi.org/10.1099/00221287-143-12-3913 PMid: 9421915.
[40] Vishnivetskaya TA, Petrova MA, Urbance J, et al. Bacterial community in ancient Siberian permafrost as characterized by culture and culture-independent methods. Astrobiology 2006;6(3):400–14. https://doi.org/10.1089/ast.2006.6.400 PMid: 16805696.
[41] Gilichinsky DA, Wilson GS, Friedmann EI, et al. Microbial populations in Antarctic permafrost: biodiversity, state, age, and implication for astrobiology. Astrobiology 2007;7(2):275–311. https://doi.org/10.1089/ast.2006.0012 PMid: 17480161.
[42] Gilichinsky D, Vishnivetskaya T, Petrova M, et al. Bacteria in Permafrost. In: Margesin R, Schinner F, Marx JC, et al, editors. Psychrophiles: from Biodiversity to Biotechnology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 83–102. https://doi.org/10.1007/978-3-540-74335-4_6.
[43] Steven B, Briggs G, McKay CP, et al. Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. FEMS Microbiol Ecol 2007;59(2):513–23. https://doi.org/10.1111/j.1574-6941.2006.00247.x PMid: 17313587.
[44] Wilhelm RC, Niederberger TD, Greer C, et al. Microbial diversity of active layer and permafrost in an acidic wetland from the Canadian High Arctic. Can J Microbiol 2011;57(4):303–15. https://doi.org/10.1139/w11-004 PMid: 21491982.
[45] Yergeau E, Schoondermark-Stolk SA, Brodie EL, et al. Environmental microarray analyses of Antarctic soil microbial communities. ISME J 2009;3(3):340–51. https://doi.org/10.1038/ismej.2008.111 PMid: 19020556.
[46] Sattler B, Puxbaum H, Psenner R. Bacterial growth in supercooled cloud droplets. Geophys Res Lett 2001;28(2):239–42. https://doi.org/10.1029/2000GL011684.
[47] Xiang SR, Shang TC, Chen Y, et al. Deposition and postdeposition mechanisms as possible drivers of microbial population variability in glacier ice. FEMS Microbiol Ecol 2009;70(2):9–20. https://doi.org/10.1111/j.1574-6941.2009.00759.x PMid: 19796140.
[48] Martineau C, Whyte LG, Greer CW. Stable isotope probing analysis of the diversity and activity of methanotrophic bacteria in soils from the Canadian high Arctic. Appl Environ Microbiol 2010;76(17):5773–84. https://doi.org/10.1128/AEM.03094-09 PMid: 20622133.
[49] Vishnivetskaya TA, Layton AC, Lau MCY, et al. Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. FEMS Microbiol Ecol 2014;87(1):217–30. https://doi.org/10.1111/1574-6941.12219 PMid: 24102625.
[50] Shen C, Ni Y, Liang W, et al. Distinct soil bacterial communities along a small-scale elevational gradient in alpine tundra. Front Microbiol 2015;6:582. https://doi.org/10.3389/fmicb.2015.00582 PMid: 26217308.
[51] Petrovskaya LE, Novototskaya-Vlasova KA, Spirina EV, et al. Expression and characterization of a new esterase with GCSAG motif from a permafrost metagenomic library. FEMS Microbiol Ecol 2016;92(5):fiw046. https://doi.org/10.1093/femsec/fiw046 PMid: 26929439.
[52] Fierer N, McCain CM, Meir P, et al. Microbes do not follow the elevational diversity patterns of plants and animals. Ecology 2011;92(4):797–804. https://doi.org/10.1890/10-1170.1 PMid: 21661542.
[53] Wang J, Soininen J, Zhang Y, et al. Contrasting patterns in elevational diversity between microorganisms and macroorganisms. J Biogeogr 2011;38(3):595–603. https://doi.org/10.1111/j.1365-2699.2010.02423.x.
[54] Shen C, Xiong J, Zhang H, et al. Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. Soil Biol Biochem 2013;57: 204–11. https://doi.org/10.1016/j.soilbio.2012.07.013.
[55] Myers RT, Zak DR, White DC, et al. Landscape-level patterns of microbial community composition and substrate use in upland forest ecosystems. Soil Sci Soc Am J 2001; 65(2):359–67. https://doi.org/10.2136/sssaj2001.652359x.
[56] Singh D, Takahashi K, Kim M, et al. A hump-backed trend in bacterial diversity with elevation on Mount Fuji, Japan. Microb Ecol 2012;63(2):429–37. https://doi.org/10.1007/s00248-011-9900-1 PMid: 21735154.
[57] Gangwar P, Alam SI, Bansod S, et al. Bacterial diversity of soil samples from the western Himalayas, India. Can J Microbiol 2009;55(5):564–77. https://doi.org/10.1139/W09-011 PMid: 19483785.

[58] Irgens RL, Gosink JJ, Staley JT. *Polaromonas vacuolata* gen. nov., sp. nov., a psychrophilic, marine, gas vacuolate bacterium from Antarctica. Int J Syst Bacteriol 1996;46(3):822–6. https://doi.org/10.1099/00207713-46-3-822 PMid: 8782696.

[59] Lutz S, Anesio AM, Edwards A, et al. Microbial diversity on Icelandic glaciers and ice caps. Front Microbiol 2015;6:307. https://doi.org/10.3389/fmicb.2015.00307 PMid: 25941518.

[60] Butinar L, Spencer-Martins I, Gunde-Cimerman N. Yeasts in high Arctic glaciers: The discovery of a new habitat for eukaryotic microorganisms. Anton Leeuw Int J G 2007;91(3):277–89. https://doi.org/10.1007/s10482-006-9117-3 PMid: 17072534.

[61] Rathour R, Gupta J, Kumar M, et al. Metagenomic sequencing of microbial communities from Brackish water of Pangong Lake of the Northwest Indian Himalayas. Genome Announc 2017;5(40):e01029-17. https://doi.org/10.1128/genomeA.01029-17 PMid: 28982995.