British Biotechnology Journal
15(1): 1-11, 2016, Article no.BBJ.27519
ISSN: 2231–2927, NLM ID: 101616695

SCIENCEDOMAIN *international*
www.sciencedomain.org

Metagenomic Analysis of Microbial Communities in the Soil-mousse Surrounding of an Amazonian Geothermal Spring in Peru

Sujay Paul[1], Yolanda Cortez[1], Nadia Vera[1], Gretty K. Villena[1] and Marcel Gutiérrez-Correa[1*]

[1]Laboratorio de Micología y Biotecnología, Universidad Nacional Agraria La Molina, Av. La Molina s/n, Lima 12, Peru.

Authors' contributions

This work was carried out in collaboration between all authors. Authors SP, YC and NV performed the experiments. Author SP did some analysis and wrote the first draft of the manuscript. Authors GKV and MGC designed the study, helped to do the analysis and to write the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BBJ/2016/27519
Editor(s):
(1) Chung-Jen Chiang, Department of Medical Laboratory Science and Biotechnology, China Medical University, Taiwan.
(2) Chan Yean Yean, Department of Medical Microbiology and Parasitology, School of Medical Sciences, Universiti Sains Malaysia, Malaysia.
Reviewers:
(1) Anita Shete, Aavanira Biotech Pvt. Ltd., Pune, India.
(2) Kingsley Anukam, Uzobiogene Genomics, London, Ontario, Canada.
Complete Peer review History: http://www.sciencedomain.org/review-history/15591

Original Research Article

Received 6th June 2016
Accepted 22nd July 2016
Published 30th July 2016

ABSTRACT

Aim: Aguas Calientes (AC) is one of the very few geothermal springs located deep into the Amazon rainforest in Peru. The main aim of this study was to generate an illumina based high resolution microbial phylogenetic profile of the soil-mousse surrounding of an Amazonian geothermal spring like AC.
Study Design: Soil-mousse surrounding of AC geothermal spring was subjected to metagenome sequencing using Illumina HiSeq platform.
Place and Duration of Study: Soil samples were collected from the surrounding of Aguas Calientes (7°21'12" S, 75°00'54" W). The duration of the study was from 2013-2016.
Methodology: Metagenomic DNA was extracted from pooled soil samples using PowerSoil® DNA isolation kit and analyzed at 16S rRNA V3-V4 hypervariable region by amplicon metagenome sequencing on Illumina Hiseq platform. QIIME pipeline was used for 16S RNA detection, clustering

**Corresponding author: E-mail: mgclmb@lamolina.edu.pe;*

and Operational Taxonomic Units (OTUs) picking followed by Biom file generation and statistical analysis. Functional analysis of 16S amplicons was performed using the default settings of PICRUSt.

Results: A total of 72 bacterial phyla and 3 archaeal phyla were detected in AC soil. Proteobacteria (50.09%) was found to be the highest represented phylum among bacterial communities while among archaeal communities Crenarchaeota (0.26%) dominated the sample. More than 50% of the sequences in AC soil were found unidentified/novel at the genus level. A plausible facultative mutualistic relationship was predicted among some members of the communities. In Clusters of Orthologs (COGs) analysis, most of the sequences were found to be associated with “Amino acid transport and metabolism” (8.3%) category, while among predicted Kyoto Encyclopedia of Genes and Genomes (KEGGs) pathways "Membrane transport" (12.1%) was the most abundant one.

Conclusion: This is the first report of a high resolution microbial phylogenetic profile of an Amazonian geothermal spring soil-mousse surrounding. Although a very diverse group of the bacterial and archaeal population was observed in the sample, a large portion of unidentified thermophilic microbial members was also noticed which need to be studied.

Keywords: Peruvian Amazon rainforest; geothermal spring; Illumina metagenome sequencing; 16S rRNA based phylogeny; soil-mousse surrounding; QIIME; PICRUSt.

ABBREVIATIONS

AC (Aguas Calientes); TOC (Total Organic Content); QC (Quality Control); QIIME (Quantitative Insights Into Microbial Ecology); NCBI (National Center for Biotechnology Information); COGs (Clusters of Orthologs); OTUs (Operational Taxonomic Units); KEGGs (Kyoto Encyclopedia of Genes and Genomes); PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States).

1. INTRODUCTION

Microbial samples from extreme habitats (extremophiles) represent a huge reservoir of genetic diversity and a potent source of industrially important enzymes [1], but the culturability of microorganisms from extreme habitats is generally very poor. For a long time, full-length sequences generated by Sanger sequencing of 16S rRNA clone libraries were considered to be the gold standard for phylogenetic analysis but soon it was realized that this procedure is not only expensive but also have many limitations. To overcome the limitations and to gain high taxonomic resolution of the bacterial population in extreme habitats, culture-independent 16S rRNAs amplicon based metagenome sequencing became a common practice [2]. Both 454 (Roche) and Illumina platforms are now largely used to study metagenome and microbial diversity [3,4] but the most recent and advanced Illumina MiSeq/HiSeq sequencing platform provides a very distinguish and high quality view of microbial composition which include the detection of extremely low abundant strains in a sample than other sequencing technologies [5].

Geothermal springs and their surrounding are unique sites for extremophilic microorganisms and are of great interest for many years because enzymes obtained from them have been proved to be extremely valuable as biocatalysts for industrial and biotechnological purposes. Moreover, many unknown microbial species and genes have been revealed in culture-independent microbial diversity assessment of geothermal springs [6]. Studies revealed that microbial diversity in geothermal springs is greatly affected by the pH and temperature of the water [7-11,4]. Water samples from an acidic geothermal spring (pH 3.5 - 4) and a circumneutral geothermal spring (pH 7.2 - 7.4) from Russia were analyzed previously [8] and the results revealed that Thermotogae and Gammaproteobacteria were the most abundant class in the acidic geothermal spring, while Thermodesulfobacteria, Gammaproteobacteria, and Betaproteobacteria dominated the circumneutral geothermal spring. Other studies with alkaline geothermal spring’s metagenomics revealed that Thermus [9], Hydrogenobacter [7], Caldicellulosiruptor, Dictyoglomus, Fervidobacterium [10] and Synechococcus [11] were the most represented groups. Swingley et al. [12] reported that in Bison Pool geothermal spring outflow channel, the most prevalent bacterial community was found to be chemotrophic at high temperature (92°C), while phototropic community dominated a biofilm at

low temperature (56°C). Later, a similar kind of findings was also reported [9]. Among Archaeal communities, generally, Euryarchaeota predominates the acidic geothermal springs [8] while Crenarchaeota monopolizes the alkaline ones [7,10]. Analyses of microbial assemblages in the soil of few geothermal spring's surrounding revealed the friendly coexistence of extremophiles and non-extremophiles as well as aerobic and anaerobic microorganisms [13,14].

Amazon rainforest is one of the most biologically diverse areas on earth and a rich source of several novel microbial species. Peruvian Amazon is endowed with very few geothermal springs but due to the remoteness of those geothermal springs till date, none of them were explored. Aguas Calientes (AC) is a nearly undisturbed geothermal spring located deep in the Peruvian Amazon rainforest (7°21'12" S, 75°00'54" W) (Fig. 1). This geothermal spring is a shallow, sulfur-rich (typical rotten egg smell as an indicator), and slightly acidic (pH 5.0-7.0) in nature, with temperatures varying from 45 to 90°C. Additionally, it is continually fed by plant litter, resulting in a relatively high degree of 'Total Organic Content' (TOC). The surrounding soil of AC, which is continuously saturated by the spring water appears like a mousse with an oscillating temperature between 30°C to 60°C; thus thermophilic microorganisms are quite expected in the sample. The metagenomic analysis of water sample is already communicated elsewhere (Paul et al., 2016 communicated). In this study, the main objective is to generate a high resolution microbial taxonomic profile of AC geothermal spring surrounding soil to trap novel microorganisms.

2. MATERIALS AND METHODS

2.1 Sample Collection

Soil samples were collected from four random sampling points of AC surrounding and mixed in an equal ratio as previously stated by Chan et al. [5] during the metagenomic study of a Malaysian geothermal spring. The temperature and pH were measured on-site.

2.2 DNA Extraction, Library Preparation, and Sequencing

Metagenomic DNA was extracted using Powersoil® DNA isolation kit (Mo bio laboratories) following manufacturer's protocol. Briefly, the sample was added to a bead

Fig. 1. Aerial and side view of Aguas Calientes (AC)

beating tube provided with Powersoil® DNA isolation kit for rapid and thorough homogenization followed by vortex with lysis buffer in proprietary bead tubes. After that, proteins and inhibitors were removed and pure DNA was eluted using Mobio laboratories silica spin column. DNA quantity and quality were then checked in Nanodrop 2000 spectrophotometer (Thermo fisher scientific Inc). Twenty-five nanograms of nanodrop quantified DNA was used for amplifying the V3-V4 region of 16s rRNA with specific primers which also have a 'tag' sequence that is complementary to Illumina sequence adapter and index primers from the Nextera XT Index kit V2. This round of PCR generates single amplicons of ~530 bp. In the next round of PCR (indexing PCR) Illumina sequencing adapters and dual indexing barcodes are added using limited cycle PCR to give a final product of ~610 bp. The libraries were cleaned using Highprep PCR (Magbio, Cat # AC-60050) magnetic beads, Qubit quantified and validated for quality by running an aliquot on High Sensitivity Bioanalyzer chip (Agilent). Finally, the cleaned libraries were sequenced on Illumina HiSeq platform at Genotypic Technology Private Limited, Bangalore, India.

2.3 Analysis of Sequencing Data and Functional Annotation

The Illumina paired-end raw reads were quality checked using FastQC tool [15]. Processing of reads was based on the sequence and its quality value (ASCII characters) under three categories: Adapter 'Quality Control' (QC), B-block QC and Low quality end QC. Microbial diversity in the sample was assessed by alpha diversity analysis (Chao1, Shannon, and Simpson indices) and rarefaction curve. Chao1 estimator calculates the estimated true species diversity of a sample by the equation: $S_{estimate} = S_{observed} + F_1^2 / 2F_2$ where F_1 is the number of singletons sampled (i.e., the number of species with only a single occurrence in the sample), and F_2 is the number of doublets (the number of species with exactly two occurrences in the sample). The Shannon and Simpson diversity indices account for both abundance and evenness of the species present in a community. 'Quantitative Insights into Microbial Ecology' (QIIME) pipelines [16] was used for 16S RNA detection, clustering and OTU picking followed by Biom file generation and statistical analysis. Species richness and diversity was further assessed using Good's coverage estimator and ACE. Good's coverage estimator, in which Coverage = 1 - (number of individuals in species / total number of individuals), allow researchers to have a good idea of how their limited sampling relates to the entire sampled population, while Abundance based Coverage Estimator (ACE) uses information on the frequency of rare species in a sample to estimate the number of undetected species in an assemblage. Krona, a powerful visualization tool that allows intuitive exploration of relative abundances and confidences within the complex hierarchies of metagenomic classifications, was also used in this study. The OTU picked by the sample was visualized as a heatmap where each row corresponds to an OTU. The higher the relative abundance of an OTU in the sample, the more intense the color at the corresponding position in the heatmap. The SRA file was deposited to 'National Center for Biotechnology Information' (NCBI) database under Accession Number SRX1809287. In the present study functional analysis of 16S amplicons was performed using the default settings of 'Phylogenetic Investigation of Communities by Reconstruction of Unobserved States' (PICRUSt) tool version 0.9.1 (Langille et al. 2013). In the default setting, PICRUSt uses relatively new Biological Observation Matrix (BIOM) format which is motivated by several goals. First, to facilitate efficient handling and storage of large, sparse biological contingency tables; second, to support encapsulation of core study data (contingency table data and sample/observation metadata) in a single file; and third, to facilitate the use of these tables between tools that support this format. PICRUSt does metagenome functional predictions by multiplying each normalized OTU abundance by each predicted functional trait abundance to produce a table of functions. It performs functional analysis for KEGG orthologs and COGs.

3. RESULTS AND DISCUSSION

3.1 Phylogenetic Composition and Microbial Communities in AC Soil

Illumina HiSeq 2500 generated a total of 3,183,656 paired-end raw reads from AC soil sample. After quality filtration and adapter trimming clean sequences were clustered into 12,749 OTUs using a 97% similarity cut-off value. A steep slope in the rarefaction curve indicated that a large fraction of the species diversity remains to be discovered (Fig. 2). If the curve becomes flattened to the right, it will indicate a reasonable number of individuals is

sampled and more intensive sampling is likely to yield only a few additional species. Sampling curves generally rise very quickly at first and then level off towards an asymptote as fewer new species are found per unit of individuals collected. Higher Good's coverage (0.996), Chao1 (19,466), ACE (19,167), Shannon (8.11) and Simpson (0.979) indices indicated that the microbial communities in AC soil are rich and very diverse in nature, and this could be due to the fact that the soil is rich in TOC. In this study, a total of 99.57% of the 16S rRNA gene fragments were assigned to a bacterial domain and 0.37% to Archaeal domain. All the resulting fragments were classified into 72 phyla, 224 classes, 444 orders, 720 families and 1388 genera. AC soil sample was dominated mainly by 5 phyla Proteobacteria (50.09%), Firmicutes (18.73%), Choloroflexi (16.31%), Acidobacteria (3.31%), and Bacteroidetes (2.31%) (Fig. 3). Community structure of top two phyla Proteobacteria and Firmicutes were represented by Krona charts (Figs. 4 and 5). The dominance of the phylum Proteobacteria, Firmicutes and Choloroflexi in geothermal springs and industrial wastewater was reported earlier in few studies from different physiochemical environments [5,4,17]. However, in contrast with our data 16S rRNA-based microbial diversity analysis from Little Hot Creek LHC) geothermal springs (temperature 78.7 - 82.5°C and pH 6.75 - 6.97), California, similar to AC geothermal spring showed the dominance of the phyla Thermodesulfobacteria, Deinococcus–Thermus, Thermotogae and Dictyoglomi [18]; similarly microbial community analysis of surrounding soil of Manikaran geothermal spring, India revealed the dominance of Actinobacteria and Deinococcus-Thermus group [14]. Both the results indicated that microbial diversity in geothermal springs or surrounding soils is greatly influenced by physicochemical environments. Betaproteobacteria (28.07%) was found to be the most dominant class in AC soil while other prevailing classes were Anaerolineae (15.78%), Bacilli (14.32%), Gammaproteobacteria (9.86%), Alphaproteobacteria (8.99%) and Clostridia (4.40%). Although the class Betaproteobacteria was the most abundant class in the sample, however at the order level the most abundant order was Bacillales (14.19%) of the phylum Firmicutes.

Other orders which dominated the samples were Burkholderiales (12.26%) and Thiobacterales (11.89%) of the phylum Proteobacteria and Anaerolineales (10.71%) of the phylum Choloroflexi. Around 13% sequences were found novel or unidentified at the family level and most of them belonged to phylum Chloroflexi (approx. 4.5%), Acidobacteria (approx. 2.0%), Proteobacteria (approx 1.5%) and Bacteroidetes (approx. 1%). Further classification revealed that more than 50% of the sequences in AC soil were unidentified at genus level which is quite interesting because it seems that it contains several novel genera which need to be explored. There are several factors which in combination made AC surrounding soil a unique site for microbial diversity compare to other related geothermal springs or surrounding soil, and they are: i) shallowness of the spring, ii) fluctuating temperature of surrounding soil from 30°C to 60°C, iii) non-uniform range of pH throughout the surrounding soil, and the most important, iv) the plant litter, which is one of the natural sources that enhance its carbon contents (TOC). The most abundant unidentified genera in AC surrounding soil belonged to the Family Thiobacteraceae (Phylum: Proteobacteria; >30%), Alicyclobacillaceae (Phylum: Firmicutes >11%) and Comamonadaceae (Phylum: Proteobacteria; >8%). Among known genera *Anaerolinea* (7.24%), *Thermomonas* (3.78%) and *Alicyclobacillus* (1.94%) were found to be the dominating ones. *Anaerolinea thermophila* (4.51%), *Anaerolinea thermolimosa* (1.41%), *Sulfuricurvum kujiense* (1.20%) and Alicyclobacillus *acidocaldarius* (0.78%) were the most abundant strains identified at species level. Members of thermophilic anaerobic genus *Anaerolinea* (Phylum: Choroflexi) grow chemo-organotrophically on a variety of carbohydrates [19]. Notably, ATP synthesis in *A. thermophila* is predicted to proceed primarily via oxidative phosphorylation, using an aerobic electron transport chain. Interestingly, the genes present among the members of this genus are extensively coded for sugar metabolism, including those encoding enzymes involved in the utilization of fructose, galactose, maltose, sorbitol, starch, sucrose, trehalose, xylose, and xylulose [20]. The presence of high organic contents in AC was probably one of the reasons behind the abundance of this genus. *Sulfuricurvum kujiense* (Phylum: Proteobacteria) is a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium belongs to the phylum Proteobacteria and frequently found in crude oil or oil sands. It solely utilizes various reduced sulfur compounds such as elemental sulfur, sulfide and thiosulfate as electron donors and biologically remove sulfide from a liquid or gaseous phase. Several known

and unknown members of *Alicyclobacilli* (Phylum: Firmicutes) also dominated AC soil at genera and species level. They are strictly aerobic, acidophilic, thermophilic, soil-dwelling organisms, and have been shown to grow at temperatures between 20 and 70°C (with the optimum temperature range being 42–60°C) and pH values of 2.0 to 6.0 [21]. Regarding archaeal community structure Crenarchaeota (0.26%) and Euryarchaeota (0.11%) were the two major represented Archaeal phyla detected in AC soil. Another recently proposed archaeal phylum Parvarchaeota [22] was also detected. Further affiliation revealed that different unidentified families of the order pGrfC26 under Crenarchaeota dominated the AC soil samples (0.24%). Few methanogenic members of the family Methanomassiliicoccaceae (0.06%), Methanobacteriaceae (0.03%), and Methanosaetaceae (0.01%) under Euryarchaeota were also detected, among which the complete genome sequence of a strain of *Methanobacterium thermoautotrophicum*, one of the best methane producing thermophilic archaeon are available in the database [23]. From the analysis of microbial communities in AC soil it is clear that due to moderate temperature both aerobic and anaerobic members distributed equally and they can use either organic or inorganic substances for their growth and development. It could also be possible that some members of AC microbial communities were in a facultative mutualistic relationship which may be a part of their survival strategy, although it is not very clear. For example, several species of photosynthetic bacterial genera like *Rhodobacter*, *Rhodopila, Rhodocyclus, Chromatium* were detected in AC soil. Due to the higher temperature and mousse like appearance, the amount of oxygen level in AC soil was expected to be low which could be a threat for other aerobic microorganisms, but the presence of photosynthetic cyanobacteria (*Cylindrospermopsis*, *Thermosynechococcus*, *Scytonema*) likely supply additional oxygen in the soil-mousse environment. On the other hand, higher abundance of sulfur-oxidizing bacterial strains like *S. kujiense* expectedly remove toxic sulfur compounds from the environment for the betterment of other nonsulfur bacteria. Another example is the dominance of denitrifying bacterial group of the family Comamonadaceae. Atmospheric nitrogen often reacts with rainwater and forms toxic nitrate compounds in the soil; a denitrifying group of bacteria metabolizes nitrate compounds and detoxify soils. Finally, the higher abundance of the members of acetogenic clostridia (Phylum: Firmicutes) in AC soil also indicated the possibility of syntrophic relationships among acetogenic bacteria and methanogenic Archaea with respect to the methanogenesis process as described previously by Ragsdale and Pierce, 2008 [24].

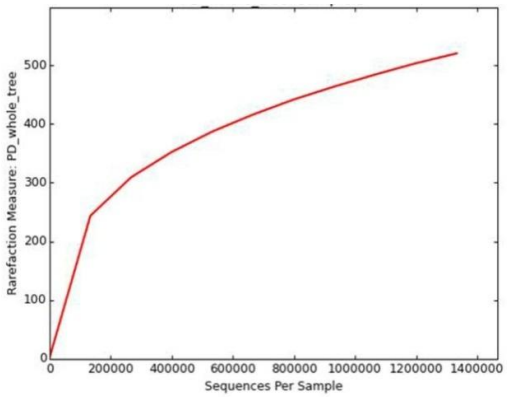

Fig. 2. Rarefaction analysis curves of bacterial 16S rDNA sequences spanning the V3-V4 region (PD=Phylogenetic Diversity)

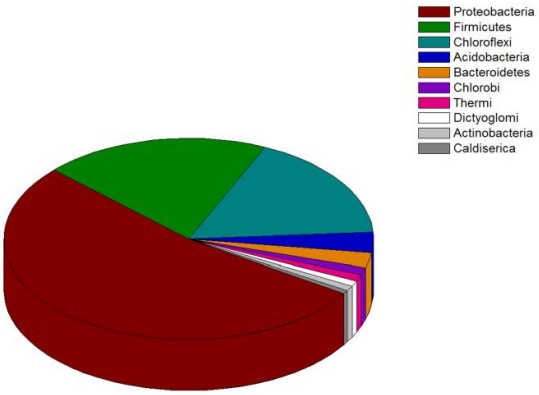

Fig. 3. Relative abundance of bacterial phyla in AC soil sample

3.2 Predicted Functional Metagenomes of AC Soil

PICRUSt, a recent tool designed to infer metagenomic information from 16S amplicon sequencing data, was used in this study for functional analysis [25]. PICRUSt did metagenome functional predictions by multiplying each normalized OTU abundances by each predicted functional trait abundance to produce a table of functions. It performed functional analysis for COGs and KEGG orthologs. Both annotation schemes contain categories of

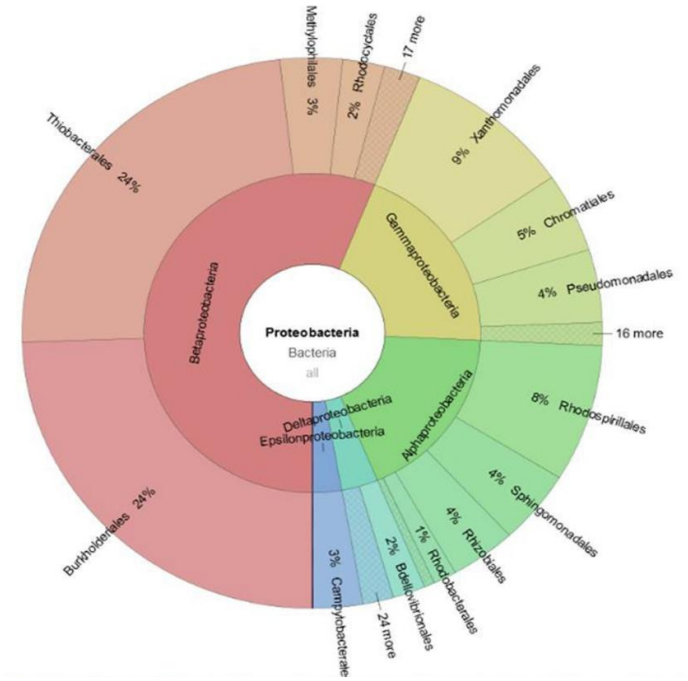

Fig. 4. Summary of the phylum proteobacteria in a multilevel chart (Krona)

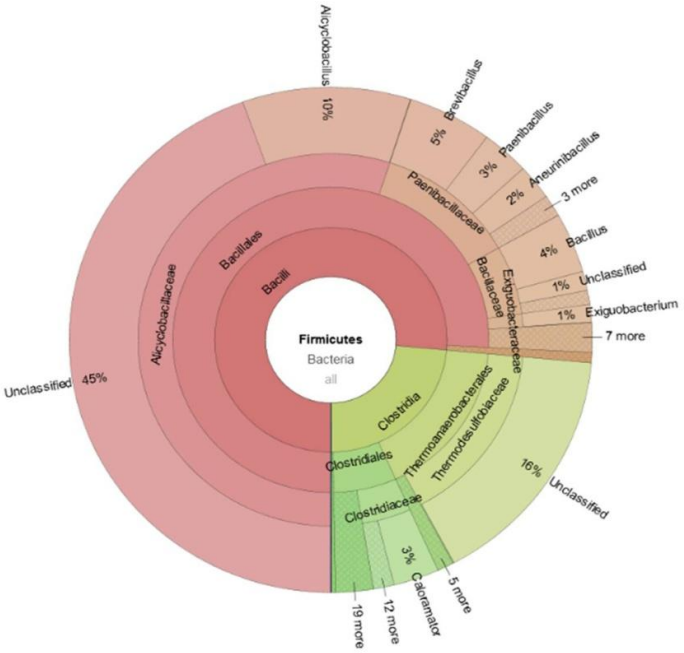

Fig. 5. Summary of the phylum firmicutes in a multilevel chart (Krona)

metabolic functions organized in multiple hierarchical levels: COG analysis uses evolutionary relations (orthologs) to group functionally related genes while KEGG analysis maps enzymes onto known metabolic pathways. COG analysis of AC soil disclosed that despite "General" and "Unknown" functional categories, the highest represented category at second tier was "Amino acid transport and metabolism" (8.3%) followed by "Energy production and conversion" (6.7%) and "Transcription" (6.6%) (Fig. 6). Among predicted KEGG pathways most

of the sequences were assigned to "Membrane transport" (12.1%), “Amino acid metabolism” (10.3%) and “Carbohydrate metabolism” (10.2%) categories (Fig. 7). Further analysis revealed that Maximum numbers of categories (12 categories) were found under the “Metabolism” super category. In agreement with our KEGG functional analysis data, highest numbers of sequence assignment under "Metabolism" super category was also reported previously in a similar Malaysian geothermal spring [5] but in contrast with our data “Carbohydrate metabolism” was the highest represented category in that study. A heat map (Fig. 8) and a Phylogenetic tree (Fig. 9) indicated significant variation among OTUs.

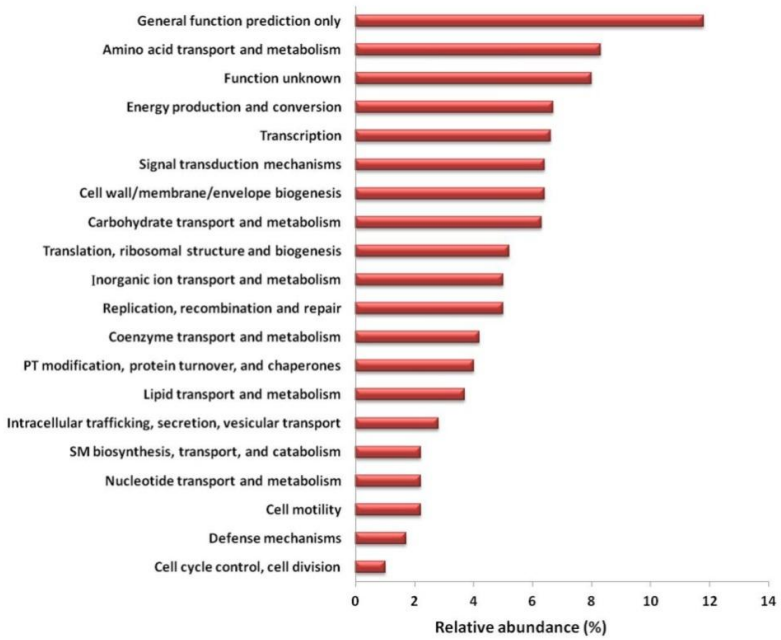

Fig. 6. COGs metagenome functional prediction of identified OTUs using PICRUSt tool

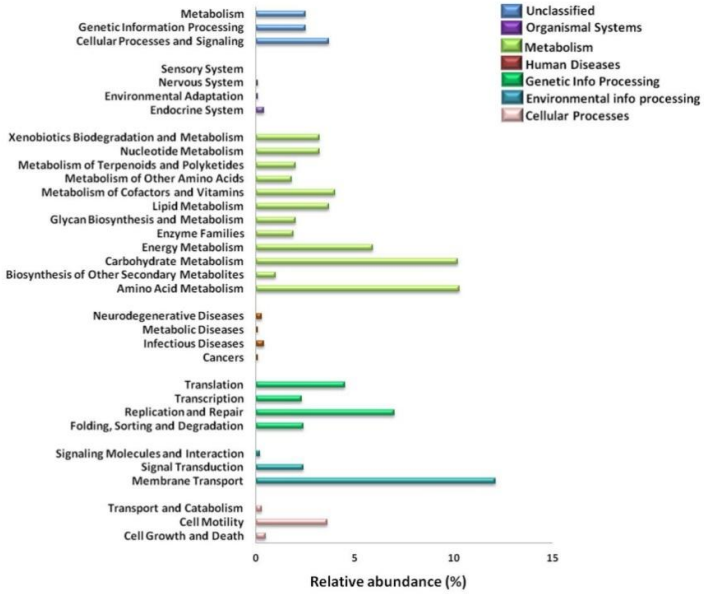

Fig. 7. KEGGs metagenome functional prediction of identified OTUs using PICRUSt tool

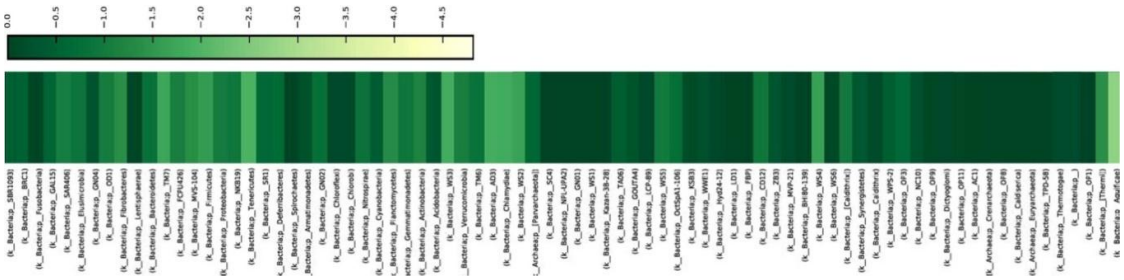

Fig. 8. OTUs picked by the sample were visualized at phylum level as a heat map

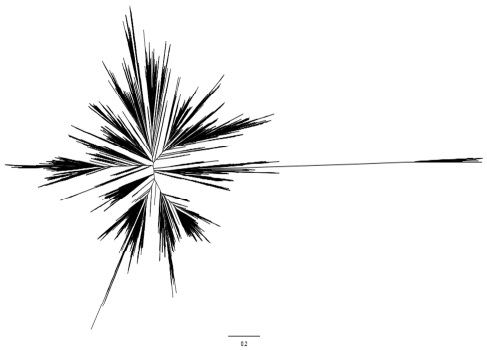

Fig. 9. OTU-based phylogenetic tree

4. CONCLUSION

The diversity and functional analysis of a microbial assemblage derived from a geothermal spring surrounded soil-mousse environment of Amazon rainforest has been illuminated for the first time by high end Illumina next generation sequencing data set. Analyses of the taxonomic composition and the metabolic potential revealed that the AC surrounding soil is dominated by aerobic and facultative aerobic bacteria (mainly Proteobacteria). A wide metabolic diversity was established, including the ability to assimilate inorganic and organic nitrogen and sulfur sources. A plausible mutualistic survival strategy was also predicted among some members of the microbial communities. This report can contribute to better understanding the microbial diversity and activity at extreme habitats like high temperature and low pH in a deep rainforest. However, more detail study using high-end approaches will be required to fully assess the association of metabolic routes with particular taxa and their relevance to community dynamics in AC soil.

ACKNOWLEDGEMENTS

This work was supported by grants from Concejo Nacional de Ciencia y Tecnología of Peru [Nº 002-2013-CONCYTEC-FONDECYT] and Programa Nacional de Innovación para la Competitividad y Productividad, Ministry of Production of Peru [Nº 109-FINCyT-FIDECOM-PIPEA-2012 and Nº 143-FINCyT-IB-2013].

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Satyanarayana T, Raghukumar C, Shivaji S. Extremophilic microbes: Diversity and perspectives. Current Science. 2005;89: 78–90.
2. Dudhagara P, Ghelani A, Patel R, Chaudhuri R, Bhatt S. Bacterial tag encoded FLX titanium amplicon pyrosequencing (bTEFAP) based assessment of prokaryotic diversity in metagenome of Lonar soda lake, India. Genomics Data. 2015;4:8–11.
3. Gloor GB, Hummelen R, Macklaim JM, Russel J, Fernandes AD, Macphee R, et al. Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged CR products. ArXiv. 2010;5:1–28.
4. Sekar S, Zintchem AAEA, Keshri J, Kamalika I, Momba M. Bacterial profiling in brine samples of the Emalahleni water reclamation plant, South Africa, using 454-pyrosequencing method. FEMS Microbiol Lett. 2014;359:55–63.
5. Chan CS, Chan K, Tay Y, Chua Y, Goh K. Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. Front Microbiol. 2015;6:177.
DOI: 10.3389/fmicb.2015.00177

6. Ghelani A, Patel R, Mangrola A, Dudhagara P. Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. Genomics Data. 2015;4:54–56.
7. Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. a comprehensive census of microbial diversity in hot springs of tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. PLoS One. 2013;8(1):e53350.
   DOI: 10.1371/journal.pone.0053350
8. Wemheuer B, Taube R, Akyol P, Wemheuer F, Daniel R. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka peninsula. Archaea. 2013;13. Article ID 136714.
   DOI: 10.1155/2013/136714
9. De León KB, Gerlach R, Peyton BM, Fields MW. Archaeal and bacterial communities in three alkaline hot springs in heart lake geyser basin, Yellowstone national park. Front Microbiol. 2013;4:330.
   DOI: 10.3389/fmicb.2013.00330
10. Sahm K, John P, Nacke H, Wemheuer B, Grote R, Daniel R, et al. High abundance of heterotrophic prokaryotes in hydrothermal springs of the Azores as revealed by a network of 16S rRNA gene-based methods. Extremophiles. 2013;17: 649–662.
11. Miller SR, Weltzer M. Investigating bacterial diversity along alkaline hot spring thermal gradients by barcoded pyrosequencing. In Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats. Ed. F. J. de Bruijn (Hoboken, NJ: John Wiley and Sons, Inc.); 2011.
12. Swingley WD, Meyer-Dombard DR, Shock EL, Alsop EB, Falenski HD, Havig JR, et al. Coordinating environmental genomics and geochemistry reveals metabolic transitions in a hot spring ecosystem. PLoS One. 2012;7(6):e38108.
   Doi: 10.1371/journal.pone.0038108
13. Kumar B, Trivedi P, Mishra AK, Pandey A, Palni LM. Microbial diversity of soil from two hot springs in Uttaranchal Himalaya. Microbiol Res. 2004;159:141–146.
14. Bhatia S, Batra N, Pathak A, Joshi A, Souza L, Almeida P, et al. Metagenomic analysis of bacterial and archaeal assemblages in the soil-mousse surrounding a geothermal spring. Genomics Data. 2015;5:195–200.
15. Andrews S. FastQC: A quality control tool for high throughput sequence data; 2010.
   Available:http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
16. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–336.
17. Mangrola A, Dudhagara P, Koringa P, Joshi CG, Parmar M, Patel R. Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. Genomics Data. 2015;4:153–155.
18. Vick TJ, Dodsworth JA, Costa KC, Shock EL, Hedlund BP. Microbiology and geochemistry of little hot creek, a hot spring environment in the long valley caldera. Geobiology. 2010;8:140–154.
19. Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, et al. *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. Int J Syst Evol Microbiol. 2006;56:1331-1340.
20. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et al. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome. 2013;1:22.
   DOI: 10.1186/2049-2618-1-22
21. Chang S-S, Kang D-H. *Alicyclobacillus* spp. in the fruit juice industry: History, characteristics, and current isolation/ detection procedures. Crit Rev Microbiol. 2004;30:55–74.
22. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013;499:431–437.
23. Smith DR, Doucette-Stamm LA, Deloughery C, et al. Complete genome sequence of methanobacterium thermoautotrophicum deltaH: Functional

analysis and comparative genomics. J Bacteriol. 1997;179:7135–7155.

24. Ragsdale SW, Pierce E. Acetogenesis and the Wood-Ljungdahl pathway of CO2 fixation. Biochim Biophys Acta. 2008;1784: 1873–1898.

25. Langille MGI, Zaneveld J, Caporaso JG, Mcdonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31:814–21.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/15591