Genomics Data 9 (2016) 63–66

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Metagenomic analysis of microbial community of an Amazonian geothermal spring in Peru

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

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

ARTICLE INFO

Article history:
Received 5 June 2016
Received in revised form 21 June 2016
Accepted 22 June 2016
Available online 26 June 2016

Keywords:
Peruvian Amazon rainforest
Geothermal/hot spring
Illumina metagenome sequencing
16S rRNA based phylogeny
PICRUSt

ABSTRACT

Aguas Calientes (AC) is an isolated geothermal spring located deep into the Amazon rainforest (7°21′12″ S, 75° 00′54″ W) of Peru. This geothermal spring is slightly acidic (pH 5.0–7.0) in nature, with temperatures varying from 45 to 90 °C and continually fed by plant litter, resulting in a relatively high degree of total organic content (TOC). Pooled water sample was analyzed at 16S rRNA V3–V4 hypervariable region by amplicon metagenome sequencing on Illumina HiSeq platform. A total of 2,976,534 paired ends reads were generated which were assigned into 5434 numbers of OTUs. All the resulting 16S rRNA fragments were then classified into 58 bacterial phyla and 2 archaeal phyla. Proteobacteria (88.06%) was found to be the highest represented phyla followed by Thermi (6.43%), Firmicutes (3.41%) and Aquificae (1.10%), respectively. Crenarchaeota and Euryarchaeota were the only 2 archaeal phyla detected in this study with low abundance. Metagenomic sequences were deposited to SRA database which is available at NCBI with accession number SRX1809286. Functional categorization of the assigned OTUs was performed using PICRUSt tool. In COG analysis "Amino acid transport and metabolism" (8.5%) was found to be the highest represented category whereas among predicted KEGG pathways "Metabolism" (50.6%) was the most abundant. This is the first report of a high resolution microbial phylogenetic profile of an Amazonian hot spring.

Specifications	
Organism/cell line/tissue	Metagenome of Aguas Calientes (AC) hot spring
Sex	Not applicable
Sequencer or array type	Illumina Genome Analyzer IIx
Data format	Raw data: FASTAQ file
Experimental factors	Environmental sample
Experimental features	16S rRNA genes amplified from the metagenome using Illumina platform followed by bacterial community analysis using QIIME version 1.9.0
Consent	Not applicable
Sample source location	Water Sample, Aguas Calientes (AC) hot spring, Amazon rainforest, Peru

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/?term=SRX1809286.

* Corresponding author.
E-mail address: mgclmb@lamolina.edu.pe (M. Gutiérrez-Correa).

2. Experimental design, materials and methods

Microbial samples from extreme habitats (extremophiles) represent a huge reservoir of genetic diversity and a potent source for industrially important enzymes [1] but the culturability of microorganisms from extreme habitats is generally very poor. For 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 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 than other sequencing technologies [5].

Hot springs 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

http://dx.doi.org/10.1016/j.gdata.2016.06.013

microbial species and genes have been revealed in culture-independent microbial diversity assessment of hot springs [6]. The Peruvian 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 few hot springs but till date none of them was explored in detail to generate high resolution microbial profile. So, the main goal of this study is to generate a high resolution microbial phylogenetic profile of an Amazonian hot spring.

In this study water samples were collected from four random sampling points of AC and mixed in an equal ratio as previously stated by Chan et al. [5] during metagenomic study of a Malaysian hot spring. The temperature and pH were measured on-site. Metagenomic DNA was extracted using PowerWater® DNA Isolation Kits (Mo Bio Laboratories) following manufacturer's protocol. 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 are 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 in Illumina HiSeq platform at Genotypic Technology Private Limited, Bangalore, India. The Illumina paired end raw reads was quality checked using FastQC tool [7]. QIIME pipelines [8] was used for 16S RNA detection, clustering and OTU picking followed by Biom file generation and statistical analysis. SRA files were deposited to NCBI database under Accession Number SRX1809286. In the present study functional analysis of 16S amplicons was performed using the default settings of PICRUSt version 0.9.1 [9].

After quality filtration and adapter trimming of raw reads, clean sequences were clustered into 5434 operational taxonomic units (OTUs) using a 97% similarity cut off. Rarefaction curve indicated that a reasonable number of individuals were sampled (Fig. 1A). Good's coverage estimator revealed that >99% of the species were estimated, while high values of Chao1 richness estimator (9762) and Shannon diversity index (4.16) indicated that microbial communities in AC are highly rich and diverse. All the resulting fragments were then classified into 58 phyla, 165 classes, 300 orders, 520 families and 954 genera. The top 5 represented phyla were Proteobacteria (88.06%), Thermi (6.43%), Firmicutes (3.41%), Aquificae (1.10%) and Chloroflexi (0.41%) (Fig. 1B); however, an unidentified bacterial phylum (0.10%) was also found among top 10 bacterial phyla. An OTU based phylogenetic tree displayed the genetic diversity among AC microbial community (Fig. 1C). Gammaproteobacteria (86.1%) was found to be the highest represented class in AC microbial community followed by Deinococci (6.43%), Bacilli (3.22%), Betaproteobacteria (1.28%) and Aquificae (1.10%). A Krona chart was constructed to illustrate the distribution pattern of phyla Proteobacteria in AC hot spring (Fig. 2A). Further affiliation revealed that in AC microbial community the most abundant genus was *Acinetobacter* (71.09%) of Moraxellaceae family in which most of the species were unidentified (48.29%); remaining major genera included *Pseudomonas* (8.59%) of Pseudomonadaceae family, *Thermus* (5.99%) of Thermaceae family, *Enhydrobacter* (3.38%) of Moraxellaceae family and *Brevibacillus* (1.52%) of Paenibacillaceae family. The dominance of Proteobacteria, Firmicutes and Choloroflexi phyla in hot spring was reported earlier in few studies [5]. However, in contrast with our data, 16S rRNA based microbial diversity analysis from Little Hot Creek hot springs (temperature 78.7–82.5 °C and pH 6.75–6.97), California, similar to AC hot spring showed the dominance of the phyla Thermodesulfobacteria, *Deinococcus–Thermus*, Thermotogae and Dictyoglomi [10]. Aquificae and Thermotogae are two best known

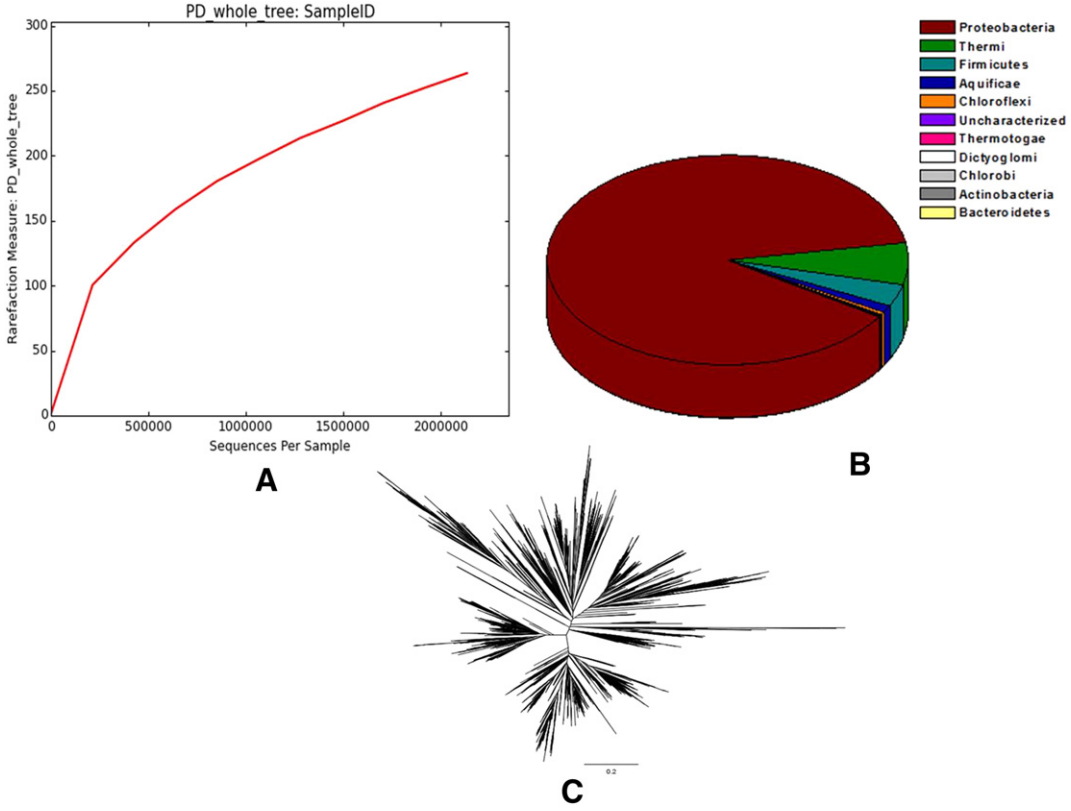

Fig. 1. Metagenomic diversity of Aguas Calientes hot spring (Peru). A: Rarefaction analysis curves of AC bacterial 16S rDNA sequences spanning the V3–V4 region; B: relative abundance of bacterial phyla in AC; and C: OTU based phylogenetic tree of AC microbial community.