



Data in Brief

Metagenomic analysis of bacterial and archaeal assemblages in the soil-mousse surrounding a geothermal spring



Sonu Bhatia ^a, Navneet Batra ^{a,*}, Ashish Pathak ^b, Amit Joshi ^c, Leila Souza ^d, Paulo Almeida ^d, Ashvini Chauhan ^b

^a Department of Biotechnology, GGSDS College, Sector 32-C, Chandigarh, India

^b School of the Environment, Environmental Biotechnology and Genomics Lab, 1515 S. MLK Blvd., Suite 305B, Building FSHSRC, Florida Agricultural and Mechanical University, Tallahassee, FL 32307, USA

^c Department of Biotechnology, SGGS College, Sector 26, Chandigarh, India

^d Laboratory of Biotechnology and Ecology of Microorganisms, Department of Biointeraction, Federal University of Bahia, Bahia, Brazil

ARTICLE INFO

Article history:

Received 6 May 2015

Received in revised form 13 May 2015

Accepted 21 May 2015

Available online 30 May 2015

Keywords:

Metagenome

Geothermal spring

Bacteria

Archaea

ABSTRACT

The soil-mousse surrounding a geothermal spring was analyzed for bacterial and archaeal diversity using 16S rRNA gene amplicon metagenomic sequencing which revealed the presence of 18 bacterial phyla distributed across 109 families and 219 genera. Firmicutes, Actinobacteria, and the Deinococcus-Thermus group were the predominant bacterial assemblages with Crenarchaeota and Thaumarchaeota as the main archaeal assemblages in this largely understudied geothermal habitat. Several metagenome sequences remained taxonomically unassigned suggesting the presence of a repertoire of hitherto undescribed microbes in this geothermal soil-mousse ecotone.

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Organism/cell line/tissue	Environmental metagenomes
Sex	Not applicable
Sequencer or array type	Ion Torrent PGM platform
Data format raw data:	FASTQ files
Experimental factors	Environmental samples
Experimental features	Bar-coded pyrosequencing and community structure diversity analysis using MG-RAST and STAMP portals
Consent	Not applicable
Sample source location	Soil-mousse from a geothermal spring, Manikaran, India

1. Direct link to deposited data

<http://www.ncbi.nlm.nih.gov/sra/SRX799662> [accn].

Geothermal springs provide a unique repository of thermophilic and hyperthermophilic microorganisms; some of which can possess valuable industrial enzymes [1]. Thus, both culture-dependent and culture-independent approaches have been utilized to obtain a better understanding of microbial diversity of thermophilic microbiota [2,3]. Several geothermal springs are located in North-India [4], such as Manikaran, from where we recently reported the microbial diversity

using 16S rRNA gene amplicon metagenomic sequencing. This revealed the dominance of Firmicutes, Aquificae, and the Deinococcus-Thermus group in this thermophilic environment [5]. The surface temperature of Manikaran springs are typically 96 °C but in the surrounding soils, which are continuously saturated by the spring water and have a mousse like appearance, temperature oscillates between 50 °C and 60 °C [4]. To gain a better understanding of the microbial diversity within the geothermal soil-mousse from Manikaran, duplicate samples were collected in sterile containers from two adjoining sites — SA and SB, near the springs. Both the sites are located within a range of 20 m and samples were collected, using sterile spatulas, from the top 2–5 cm depth. SA soil temperature typically oscillates between 65 and 70 °C, was slightly yellow in color likely due to the sulfur being discharged from the nearby outlet of the spring with hot geothermal water continuously draining over the site. Conversely, site B represents soils having temperature of 60 °C, due to the subsurface presence of geothermal spring conduits and were light brown in color. High temperature of site B is most likely due to the close proximity of the subsurface hot springs; thus, these soils were exposed to the steam originating from beneath. Thus thermophilic microorganisms are expected in the above soil samples. The samples were stored on ice in sterile containers and transported to GGSDS College, Chandigarh, where genomic DNA was extracted using the soil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA, USA). Illumina sequencing and processing of metagenomes followed our previously reported protocols [5]. Integration of obtained crude sequences was performed using CLC genomics version 7.5.1 (CLC bio,

* Corresponding author. Tel.: +91 172 491240x2045; fax: +91 172 2661077.
E-mail address: batranavneet@gmail.com (N. Batra).

Qiagen, Boston, MA, USA) and further quality trimmed (Q20) to obtain a total of 32.1 Mb of sequence data with total reads of 67,665. Sample SA1 contained 22.8 Mbp bases (G + C content 56.4%), SA2 contained 11.7 Mbp bases (G + C content 55.8%), SB1 contained 18.2 Mbp bases (G + C content 61.8%), and SB2 contained 5.3 Mbp bases (G + C content 61.7%), respectively. Quality trimmed reads were uploaded and analyzed using MG-RAST [6]. Additionally, metagenomic analysis data obtained by MG-RAST were also analyzed using Statistical Analysis of Metagenomic Profiles (STAMP) software version 2.0 [7].

Microbial diversity and richness analysis revealed that alpha diversity was highest in SA1 (80.54 spp.) followed by SA2 (60.84), SB1 (53.89) and SB2 (53.84), respectively. Rank abundance plots showed even

distribution of taxa across all 4 samples with Firmicutes as the predominating phyla with the exception of SA2, where unclassified bacteria were the most abundant group. Regardless of sites, the Gram-positive, endospore-forming Firmicutes (8–30%) dominated the geothermal soil-mousse (Fig. 1A–D), similar to the geothermal spring water shown by our recent report [5]. The other main bacterial assemblages included unclassified bacteria (21–31%), Actinobacteria (18–26%) and the Deinococcus-Thermus group (3–8%). Main bacterial genera included *Bacillus* spp., *Acidimicrobium* spp., *Arthrobacter* spp., *Actinomyces* spp., and *Thermomicrobium* spp., respectively. As shown in Fig. 2, differences observed between the two samples were mainly in Bacteroidetes, Gemmatimonadetes, Planctomycetes and

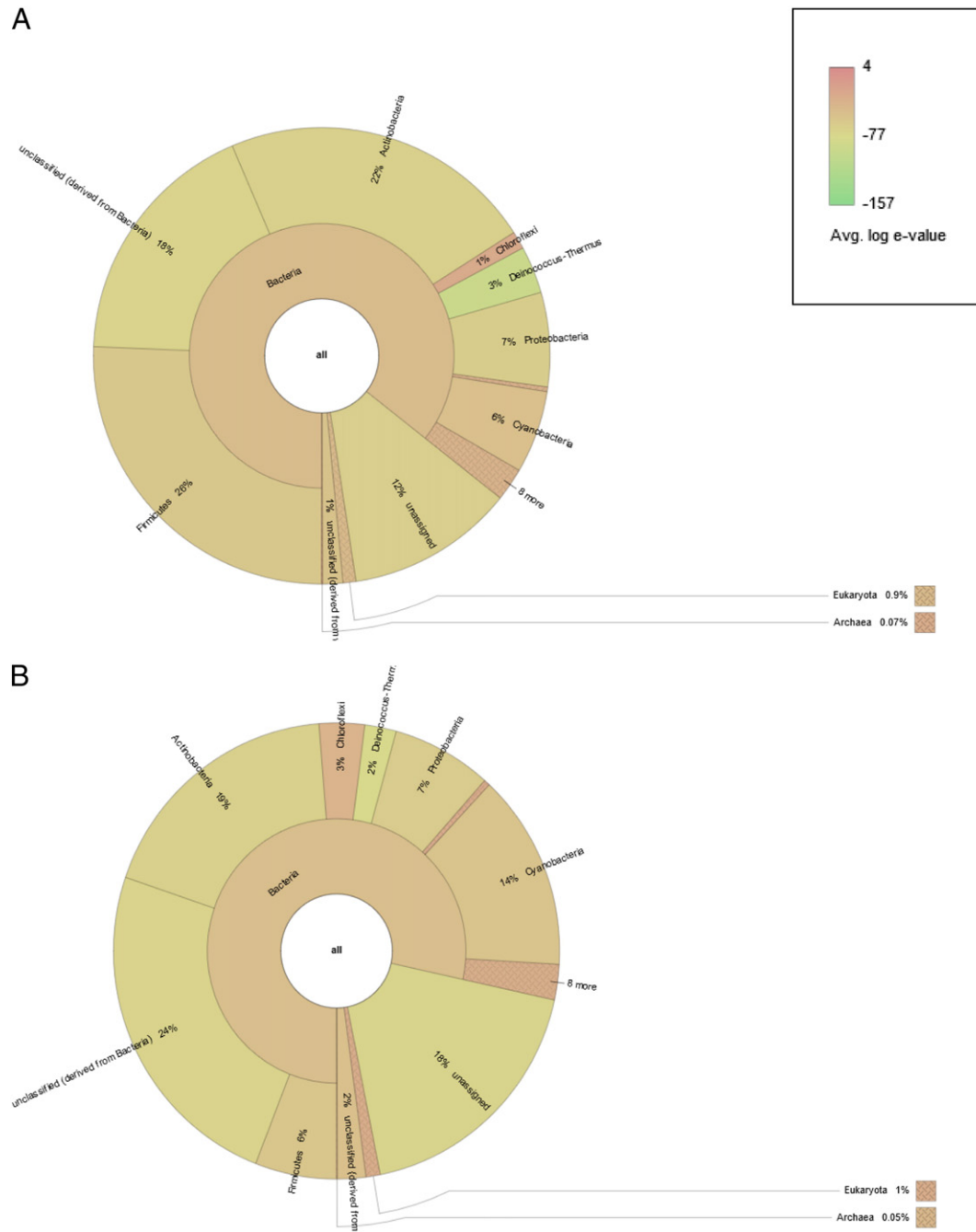


Fig. 1. Taxonomic distribution of the obtained metagenome sequences. Bacterial communities are dominant within the Manikaran Geothermal soil-mousse and Archaea are a minor fraction. The figure was prepared by the Krona interactive visualization program built within the MG-RAST portal. Shown are A (sample SA1), B (sample SA2), C (sample SB1) and D (sample SB2), respectively.

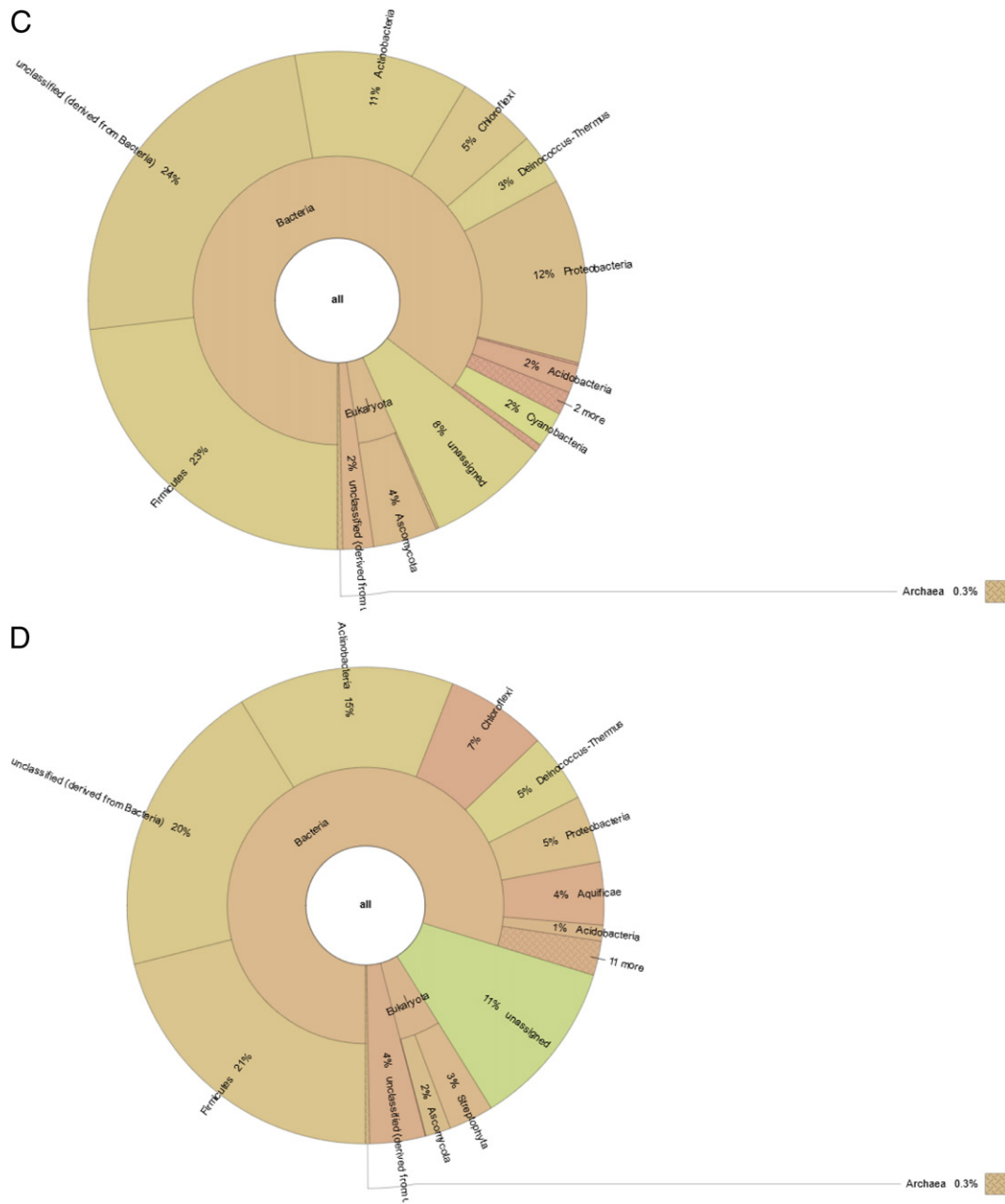


Fig. 1 (continued).

Verrucomicrobia – all these phyla predominated in SA soils; whereas, Acidobacteria predominated the SB soils, respectively. In particular, STAMP analysis further delineated differences observed between soils SA and SB, as shown in Fig. 3. This analysis revealed that, despite the predominance of the phylum Firmicutes across the analyzed soil-mousse environment, differences at 95% confidence intervals also existed among the genera that taxonomically clustered with Firmicutes from the two soil locations, as shown in red text. Moreover, the unassigned or unclassified groups also showed up as significant members of the soil-mousse econiche in SA1, SA2 (Fig. 3A) as well as SB1 and SB2 (Fig. 3B), respectively.

Crenarchaeota was the main archaeal phyla in the soil-mousse (50–67%), along with Thaumarchaeota (33–50%); in some of the samples, Thaumarchaeota was the only archaeal phyla found. Major archaeal genera included the ammonia-oxidizing *Nitrososphaera* spp.,

'Candidatus' *Nitrosocaldus* and *Nitrososphaera* spp. Moreover, 8–18% metagenomic sequences remained taxonomically unassigned suggesting the possibility of a plethora of hitherto undescribed microorganisms within this geothermal ecosystem.

2. Nucleotide sequence accession number

The DNA sequences from this metagenomic project were deposited in the Sequence Read Archive under the accession number SRX799662.

Acknowledgments

Partial funding for this study was provided by the U.S. Department of Defense (DOD) grants W911NF-10-1-0146 and W911NF-10-R-0006.

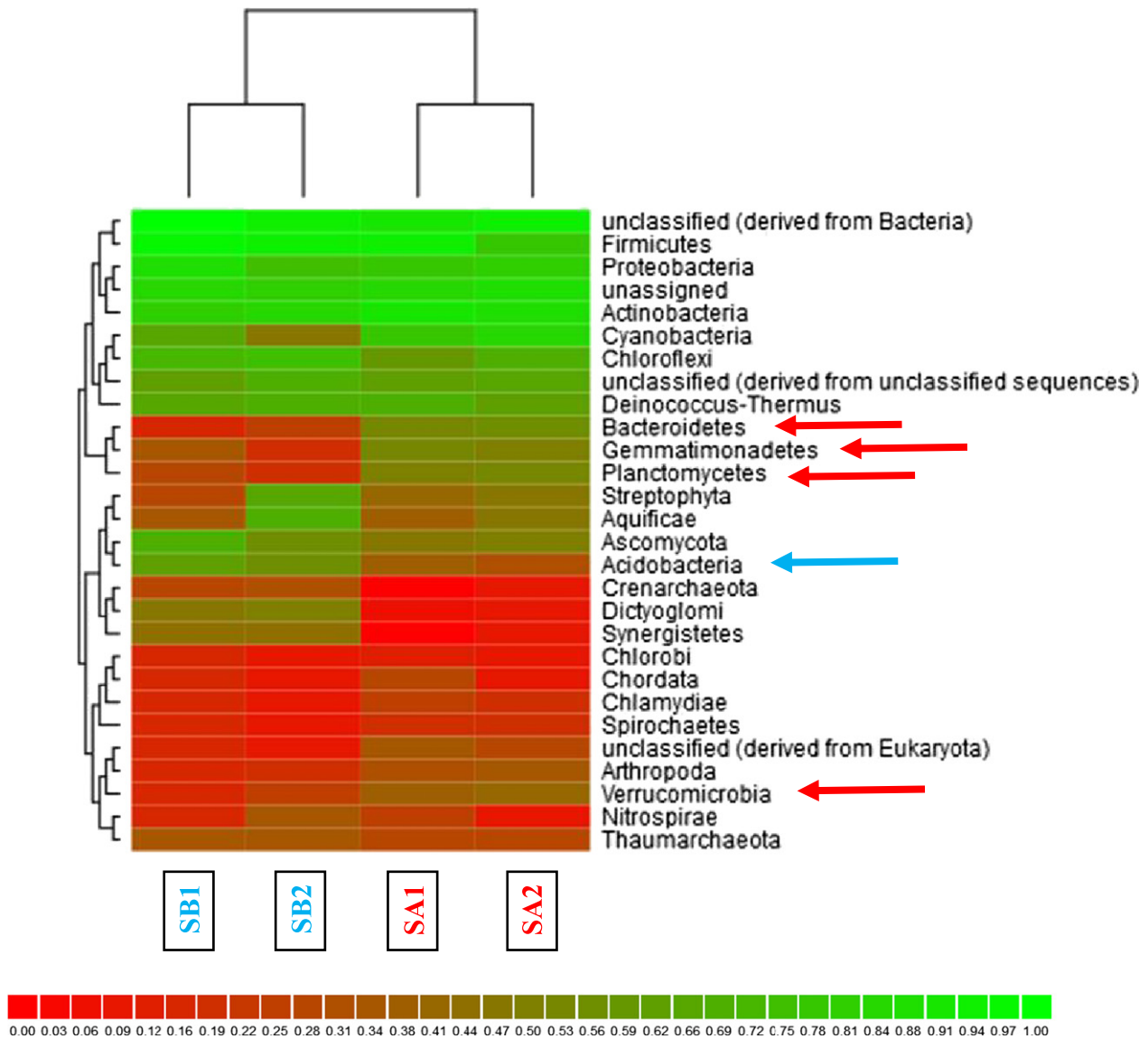


Fig. 2. Double hierarchical dendrogram showing bacterial distribution at the phylum-level from the Manikaran Geothermal soil-mousse metagenomes. The phylogenetic tree was calculated using the neighbor-joining method and the relationship among samples was determined by Bray–Curtis distance and the complete clustering method. The heatmap depicts the relative percentage of each bacterial phylum (variables clustering on the Y-axis) within each sample (X-axis clustering). The relative Euclidean distance values are depicted by the red and green colors indicating low and high abundance, respectively correlating with colors shown in the legend at the bottom of the figure. Clusters based on the distance of samples along the X-axis and the bacterial classes along the Y-axis are indicated in the upper and left of the figure, respectively. Arrows point to those bacterial taxa/phyla that differed between the SA1–SA2 and SB1–SB2 sample groups, based on Euclidean distances.

References

- [1] N. Batra, J. Singh, A. Joshi, S. Bhatia, Applications of β -gal-III isozyme from *Bacillus coagulans* RCS3, in lactose hydrolysis. *Int. J. Biol. Macromol.* 49 (2011) 879–884.
- [2] B.P. Hedlund, J.K. Cole, A.J. Williams, W. Hou, E. Zhou, W. Li, H. Dong, A review of the microbiology of the Rehai geothermal field in Tengchong, Yunnan Province, China. *Geosci. Front.* 3 (3) (2012) 273–288.
- [3] B. Kumar, P. Trivedi, A.K. Mishra, A. Pandey, L.M.S. Palni, Microbial diversity of soil from two hot springs in Uttarakhand Himalaya. *Microbiol. Res.* 159 (2) (2004) 141–146.
- [4] O.P. Pandey, J.G. Negi, Geothermal fields of India: a latest update. *Proc. World Geothermal Congress, Florence, Italy 1995*, pp. 163–171.
- [5] S. Bhatia, N. Batra, A. Pathak, S.J. Green, A. Joshi, A. Chauhan, Metagenomic evaluation of bacterial and archaeal diversity in the geothermal hot springs of Manikaran, India. *Genome Announc.* 3 (1) (2015) <http://dx.doi.org/10.1128/genomeA.01544-14> (pii: e01544-14).
- [6] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R.A. Edwards, The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9 (2008) 386.
- [7] D.H. Parks, G.W. Tyson, P. Hugenholtz, R.G. Beiko, STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30 (2014) 3123–3124.

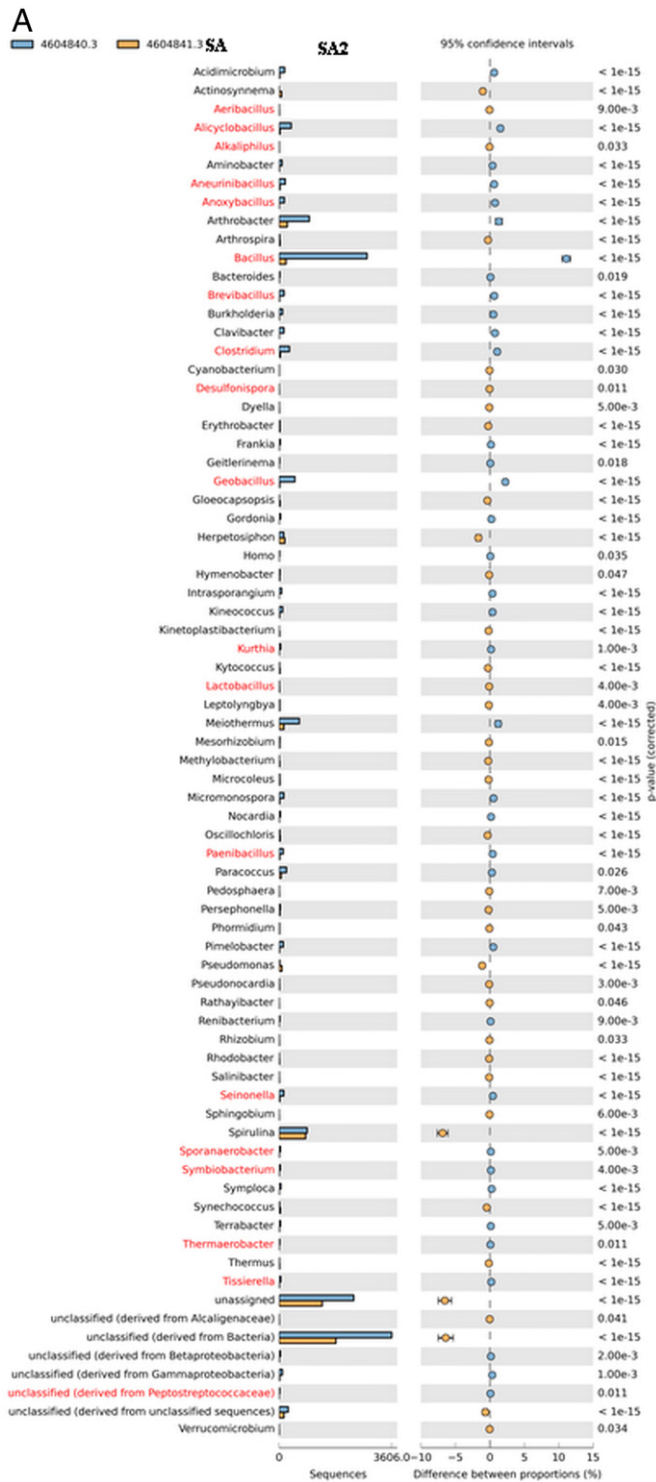


Fig. 3. Genera-level taxonomic statistical analysis of metagenomic sequences obtained from the Manikaran Geothermal soil-mouse using the STAMP software. Blue bars represent the SA1 and SB1 sample metagenomes and yellow bars represent the SB1 and SB2 metagenomes, respectively. Differences between samples are shown at 95% confidence intervals. Genera highlighted in red text belong to the predominant phyla (Firmicutes) that were identified across these soil-mouse habitats.

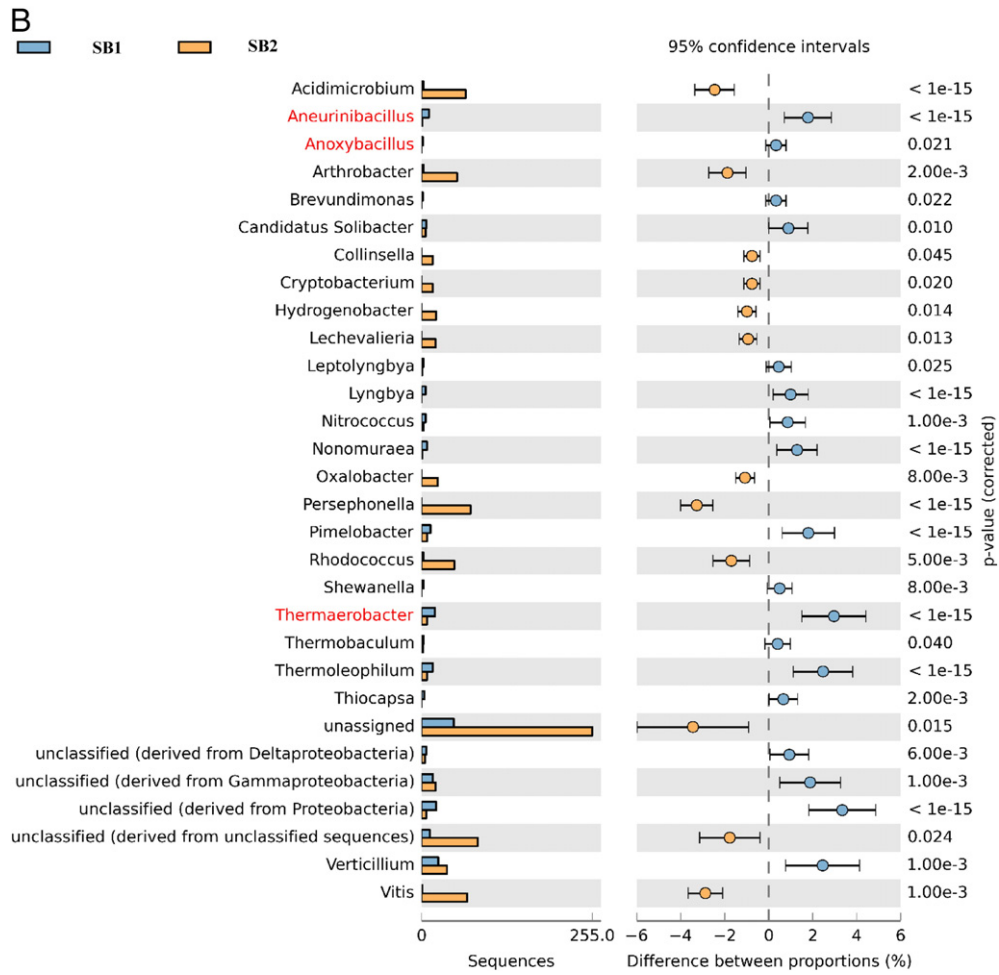


Fig. 3 (continued).