Genomics Data 12 (2017) 54–57

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Cataloguing the bacterial community of the Great Salt Plains, Oklahoma using 16S rRNA based metagenomics pyrosequencing

Ahmed H. Gad

Department of Biological Science, University of Tulsa, 800 South Tucker Drive, Oliphant Hall 312, Tulsa, OK 74104, United States

ARTICLE INFO

Article history:
Received 15 February 2017
Received in revised form 3 March 2017
Accepted 4 March 2017
Available online 8 March 2017

Keywords:
Metagenome
Pyrosequencing
Great Salt Plain
Bacteria
Salt flat
Vegetation
GSP
GAS
GAB

ABSTRACT

The Great Salt Plains of Oklahoma (GSP) is an extreme region, a hypersaline environment from marine origin and a unique area of the Salt National Wild Refuge in the north-central region of Oklahoma. In this study we analyzed the diversity and distribution of bacteria in two habitats; vegetated areas (GAB) and salt flat areas (GAS) in the sediments of GSP using the high-throughput techniques of 16S rRNA gene amplicon (V1-V2 regions) metagenomics-454 pyrosequencing. The filtered sequences resulted to a total of 303,723 paired end reads were generated, assigned into 1646 numbers of OTUs and 56.4% G + C content for GAB, and a total of 144,496 paired end reads were generated, assigned into 785 numbers of OTUs and 56.7% G + C content for GAS. All the resulting 16S rRNA was of an average length ~ 187 bp, assigned to 37 bacterial phyla and candidate divisions. The abundant OTUs were affiliated with Proteobacteria (36.2% in GAB and 31.5% in GAS), Alphaproteobacteria (13.3% in GAB and 8.7% in GAS), Gammaproteobacteria (13% in GAB and 14.2% in GAS), Deltaproteobacteria (6.5% in GAB and 6.1% in GAS), Betaproteobacteria (2.6% in GAB and 1.14% in GAS), Bacteroidetes (16.8% in GAB and 24.3% in GAS), Chloroflexi (8.7% in GAB and 6% in GAS), Actinobacteria (8.5% in GAB and 5.8% in GAS) and Firmicutes (6.5% in GAB and 6.6% in GAS). This is the first study of a high resolution microbial phylogenetic profile of the GSP and the findings stipulate evidence of the bacterial heterogeneity that might be originated by surface and subsurface environments and better understanding of the ecosystem dynamics of GSP. Metagenome sequence data are available at NCBI with accession numbers; LT699840-LT700186.

Specifications	
Organism/cell line/tissue	Great salt plain sediment metagenome
Sex	Not applicable
Sequencer	Sequencer or array Type 454 GS FLX Titanium pyrosequencer
Data format	Raw data FAST file
Experimental factors	Environmental sample
Experimental features	16S rRNA genes amplified from the metagenome using pyrosequencing followed by bacterial community analysis using Mothur version 1.35.1
Consent	Not applicable
Sample source location	Salty deposits, Vegetated soil, Great Salt Plains, Oklahoma, USA.

1. Direct link to deposited data

http://www.ebi.ac.uk/ena/data/view/LT699840-LT700186

E-mail address: ahmed-gad@utulsa.edu.

North America harbors broadly terrestrial saline environments, including the Great Salt Plains district. The GPS is one of the most remarkable reservoirs for microbial community with variability of tolerance to saline stress. It is a distinguish area of the Salt National Wild Refuge in the north-central region of Oklahoma. The area is featured by a very large salt flat (~65 km^2) of mud flats and sand bars, which covered by a small layer of thasassohaline salt evaporates. The few studies that have investigated the microbial diversity in GPS sediments in USA have improved our knowledge on the microbial community [1–7], but the comprehensive knowledge of microbial diversity and function in the GSP is still rarely uncovered. There are no studies available on the diversity of bacterial community in GSP using high-throughput DNA pyrosequencing technologies and bioinformatics. Therefore, in order to gain new perception into bacterial diversity, and reveal the impact of soil types in shaping the microbial communities in GSP, we investigated the microbial diversity of thirty samples collected from vegetated and salt flat soils during the months of June to December 2009. We used 454-pyrosequencing to establish foundational information for future research. This study focuses on the GSP soil samples that were collected from vegetated soils, and salt flat soils that comprise the most habitats

http://dx.doi.org/10.1016/j.gdata.2017.03.002

in the ecosystem. Samples were collected in triplicate in a sealed sterile containers from the surface (8 cm) and subsurface (15 cm) soils of the GSP. Soil metagenomic DNA was recovered using (PowerSoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA). We analyzed the bacterial community by amplifying the V1 and part of V2 regions of bacterial 16S rRNA gene using PCR. Pyrosequencing was implemented for 200 cycles on a Roche 454 GS-Junior sequencing instrument according to the manufacturer's protocol (454 Life Science, USA). The quality filtration, trimming and chimeric processes for the raw data were performed by Mothur software [8]. The clean sequences were analyzed with SILVE analysis pipeline [9], then clustered using a 97% similarity cut off into a total of 303,723 paired end reads and 56.4% G + C content for GAB (1646 OTUs), and a total of 144,496 paired end reads and 56.7% G + C content for GAS (785 OTUs). The final analysis showed that the metagenome sequences were represented by 37 bacterial phyla and candidate divisions (23 common phyla and candidate divisions between GAB and GAS) (Fig. 1A). We found less frequency of phyla in the salt flat soils, possibly because of physico-chemical impact of the salinity, and lack of root exudates. The top five represented phyla were Proteobacteria being the most dominant phylum in both samples (36.2% in GAB and 31.5% in GAS respectively), Bacteroidetes formed the second abundant one (16.8% in GAB and 24.3% in GAS respectively), Chloroflexi (8.7% in GAB and 6% in GAS respectively), Actinobacteria (8.5% in GAB and 5.8% in GAS respectively) and Firmicutes accounted the less abundant phylum in both samples (6.5% in GAB and 6.6% in GAS respectively) (Fig. 1B). However, an unidentified bacterial phyla were detected among the top 13 bacterial phyla in GAB (0.54%) and among the top 11 bacterial phyla in GAS (2%) (Fig. 1B). The following groups were found in less abundance: Lentisphaerae, Chlamydiae, Armatimondetes, Fibrobacteres, Deferribacteres, Tenericutes, Spirochaetae, Elusimicrobia, Fusobacteria, WCHB1-32, TM6, RF3, Candidate division OD1, Candidate division OP3, Candidate division OP11, Candidate division SR1, and Candidate division WS3. It is worth mentioning that the Operational Taxonomic Units (OTUs) richness was higher in the GAB samples than in the GAS samples, and with only 167 OTUs shared in both samples. The rare phyla such as NPL-UPA2, WCHB1-32, Nitrospira, Chlamydiae, TM6, Tenericutes, Elusimicrobia and Candidate division OP3 exhibited a preference for growth in GAB soil samples, whereas phyla; Fusobacteria, Deinococcus, TA06, Candidate divisions OP8 and BRC1 showed preference for growth in GAS (Fig. 2). The next generation pyrosequencing evidently revealed the metagenomics of microbiota in the GSP sediments, also can effectively

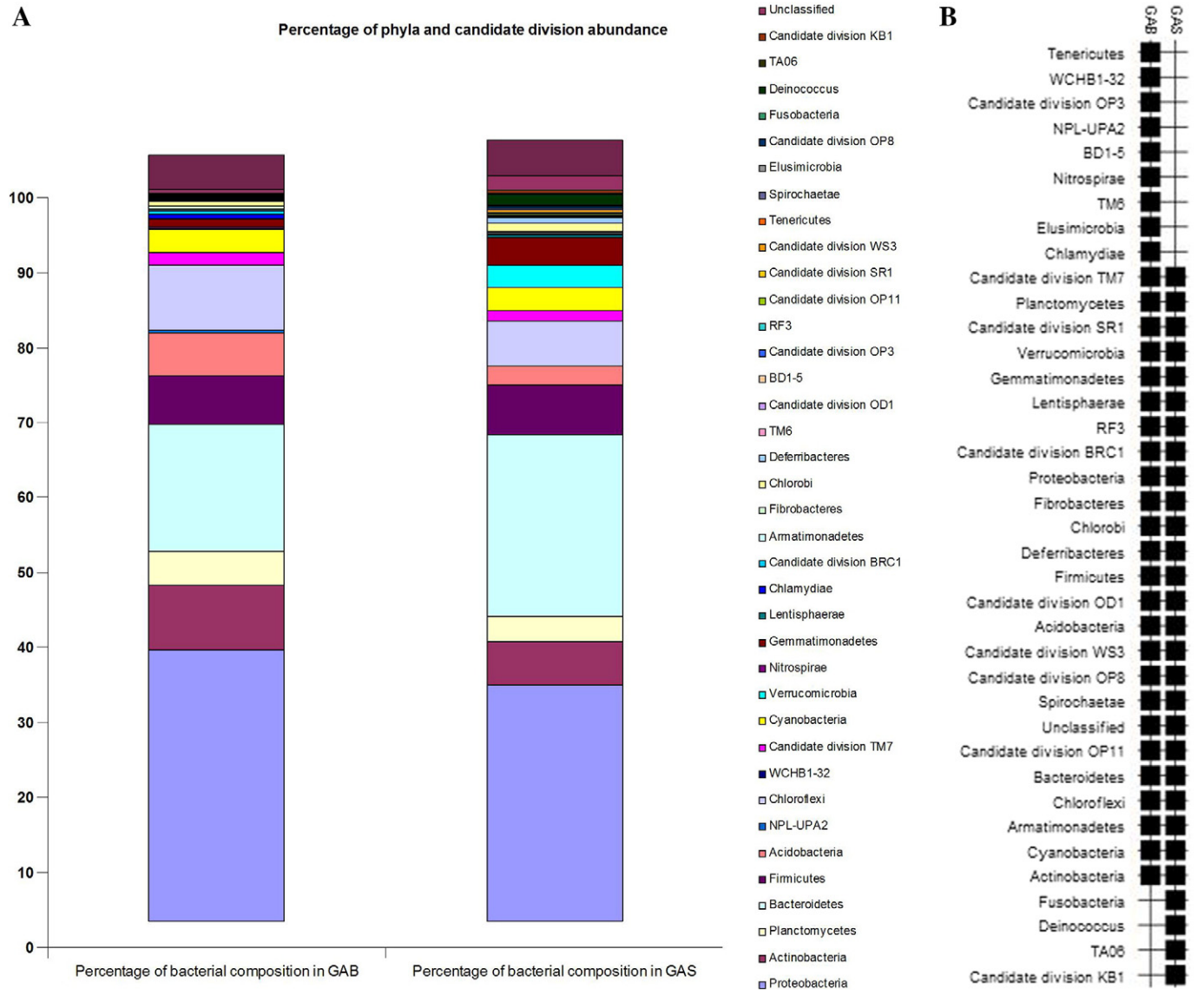

Fig. 1. A. Stratified column charts representing the percentage of frequencies of recovered phyla and candidate divisions from GAB and GAS samples. **B.** Serration visualization of distribution of phyla and candidate divisions between GAB and GAS samples.

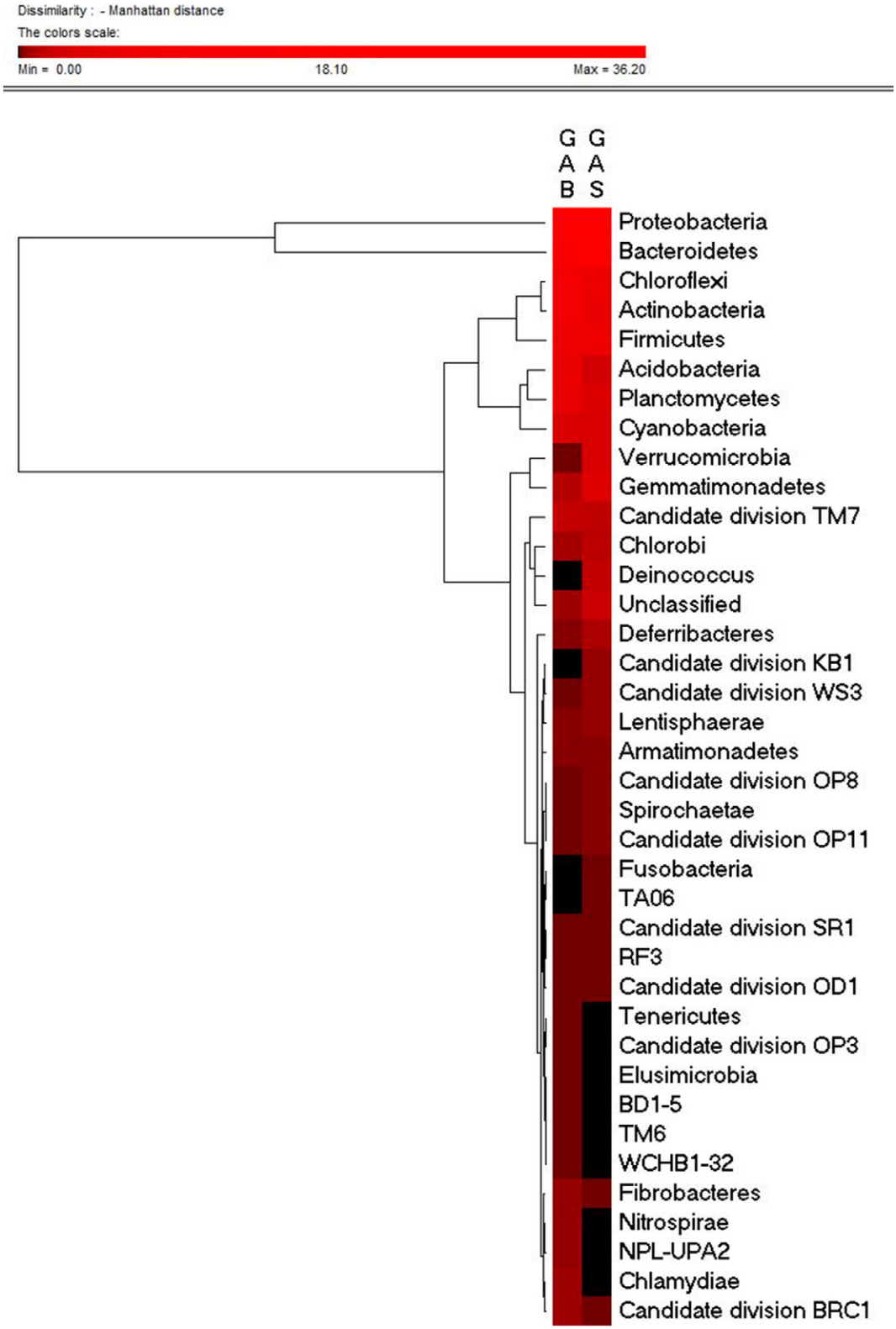

Fig. 2. Hierarchical clustering heat map generated from taxonomic abundance profile reflecting distribution of bacterial phyla and candidate divisions in the sediments of GAB and GAS samples of the great salt plains, Oklahoma.

improve our understanding of the microbial composition of the extreme environments [10], their abundance, diversity and distribution [11–12]. However, several metagenome sequences have still remained taxonomically unidentified indicating the existence of a repertoire of unassigned microbes in different habitats of the sediment of GSP. These findings provide molecular evidence of the bacterial community heterogeneity in the GSP, which might be controlled by both subsurface and surface environments. Therefore, the analysis of bacterial community datasets obtained from GSP sediments revealed significant difference between vegetated and salt flat samples. Finally, the bacterial communities in soils of the GSP including GAB and GAS samples are highly diverse and taxonomically obvious, when reflecting the environmental conditions,

soil structures, the influence of rhizosphere on bacterial members' growth and emerging of new or even different lineages.

2. Nucleotide sequence accession number

Metagenome sequence data are available at ENA Accession No. http://www.ebi.ac.uk/ena/data/view/LT699840-LT700186

Competing interests

The author declares that there are no competing interests.

Acknowledgments

This study was financially supported by the Office of Research and Sponsored Programs (The Student Research Grant Program) of University of Tulsa (20-2-1211825-59520), USA. We acknowledge Dr. N. Youssef, and Dr. M. Elshahed for technical support regarding our study and Dr. M. Fakhr for enthusiasm during the study.

References

[1] T.M. Caton, L.R. Witte, H.D. Ngyuen, J.A. Buchheim, M.A. Schneegurt, Halotolerant aerobic heterotrophic bacteria from the Great Salt Plains of *Oklahoma*. Microb. Ecol. 48 (4) (2004) 449–462.

[2] T.M. Caton, I.R. Caton, L.R. Witte, M.A. Schneegurt, Archeal diversity at the Great Salt Plains of *Oklahoma* described by cultivation and molecular analyses. Microb. Ecol. 58 (3) (2009) 519–528.

[3] I.R. Caton, M.A. Schneegurt, Culture-independent analysis of the soil bacterial assemblage at the Great Salt Plains of *Oklahoma*. J. Basic Microbiol. 52 (1) (2012) 16–26.

[4] C.A. Nicholson, B.Z. Fathepure, Aerobic biodegradation of benzene and toluene under hypersaline conditions at the great salt plains, *Oklahoma*. FEMS Microbiol. Lett. 245 (2) (2005) 257–267.

[5] R. Litzner, T.M. Caton, M.A. Schneegurt, Carbon substrate utilization, antibiotic sensitivity, and numerical taxonomy of bacterial isolates from the great salt plains of *Oklahoma*. Arch. Microbiol. 185 (4) (2006) 286–296.

[6] M.A. Schneegurt, Colorimetric microbial diversity analysis and halotolerance along a soil salinity gradient at the great salt plains of *Oklahoma*. Res. Microbiol. 164 (1) (2013) 83–90.

[7] C. Wilson, T.M. Caton, J.A. Buchheim, M.A. Buchheim, M.A. Schneegurt, R.V. Miller, DNA-repair potential of *Halomonas* spp. From the salt plains microbial observatory of Oklahoma. Microb. Ecol. 48 (4) (2004) 541–549.

[8] P.D. Schloss, S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75 (23) (2009) 7537–7541.

[9] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glockner, The SILVA ribosomal RNA gene database project: improved data processing and web- based tools. Nucleic Acids Res. 41 (D1) (2013) (D590-D569).

[10] N. Youssef, C.S. Sheik, L.R. Krumholz, F.Z. Najar, B.A. Roe, M.S. Elshahed, Comparison of species richness estimation obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl. Environ. Microbiol. 75 (16) (2009) 5227–5236.

[11] N.H. Youssef, K.N. Ashlock-Savage, M.S. Elshahed, Phylogenetic diversities and community structure of members of the extremely halophilic Archaea (order Halobacteriales) in multiple saline sediment habitats. Appl. Environ. Microbiol. 78 (5) (2012) 1332–1344.

[12] M.S. Elshahed, F.Z. Najar, M. Aycock, C. Qu, B.A. Roe, L.R. Krumholz, Metagenome analysis of the microbial community at Zodletone spring *Oklahoma*: insights into the genome of a member of the novel candidate division OD1. Appl. Environ. Microbiol. 71 (2005) 7598–7602.