Genomics Data 2 (2014) 132–134

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

GeoChip as a metagenomics tool to analyze the microbial gene diversity along an elevation gradient

Ying Gao [a], Shiping Wang [b], Depeng Xu [a], Hao Yu [c], Linwei Wu [a], Qiaoyan Lin [d], Yigang Hu [d,e], Xiangzhen Li [f], Zhili He [c], Ye Deng [c], Jizhong Zhou [a,c,g], Yunfeng Yang [a,*]

[a] State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China
[b] Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100085, China
[c] Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA
[d] Key Laboratory of Adaption and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China
[e] Shapotou Desert Experiment and Research Station, Cold and Arid Regions and Environmental & Engineering Research Institute, Chinese Academy of Sciences, Lanzhou 730000, China
[f] Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China
[g] Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form 3 June 2014
Accepted 3 June 2014
Available online 11 June 2014

Keywords:
Gene diversity
Soil microbial community
GeoChip 4.0
Genomic technology

ABSTRACT

To examine microbial responses to climate change, we used a microarray-based metagenomics tool named GeoChip 4.0 to profile soil microbial functional genes along four sites/elevations of a Tibetan mountainous grassland. We found that microbial communities differed among four elevations. Soil pH, temperature, NH_4^+–N and vegetation diversity were four major attributes affecting soil microbial communities. Here we describe in details the experiment design, the data normalization process, soil and vegetation analyses associated with the study published on ISME Journal in 2014 [1], whose raw data have been uploaded to Gene Expression Omnibus (accession number GSM1185243).

Specifications	
Organism	Uncultured bacterium
Sequencer or array type	GeoChip 4.0
Data format	Raw data: TXT, normalized data: TXT
Experimental factors	Soil samples were collected from four elevations: 3200 m, 3400 m, 3600 m and 3800 m.
Experimental features	Profiling microbial functional potentials with a microarray-based metagenomics tool named GeoChip 4.0 along an elevation gradient in a Tibetan grassland.
Consent	n/a
Sample source location	The Haibei Alpine Meadow Ecosystem Research Station (37°37′N, 101°12E′), Qinghai, China,

Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1185243

Experimental design, materials and methods

Description of the sites

This experiment was conducted at an alpine meadow in the Haibei Alpine Meadow Ecosystem Research Station of Chinese Academy of Science, which is located in a large valley surrounded by the Qilian Mountain of the northeast of Qinghai-Tibet Plateau (37°37′N, 101°12E′) in Qinghai province. It has a typical highland continental climate with cold and long winter but cool and short summer. The annual mean air temperature recorded at the station is −1.7 °C [2]. The day/night temperature variation is substantial due to strong sun radiation. The annual mean precipitation is 560 mm and 85% of rainfall is within the growing season from May to September [3].

The dominant soil type at the station is Mat Cryic Cambisols (a typical alpine grassland soil) and its pH values are 7.3 and 7.4 at depths of 10 and 20 cm, respectively. Aboveground plant biomass increases from May to July, reaches the maximum level in late July and early August, and withers in early October. Over 80% of vegetation species use C_3 photosynthetic pathway for carbon fixation [3].

This experiment, designed to study the effects of climate changes with the space-substitutes-time strategy, was set at four sites/elevations

* Corresponding author. Tel.: +86 10 62784692; fax: +86 10 62794006.
E-mail address: yangyf@tsinghua.edu.cn (Y. Yang).

http://dx.doi.org/10.1016/j.gdata.2014.06.003

of 3200, 3400, 3600 and 3800 m in May 2006. The spatial distances between adjacent sites are 6.2 km (3200–3400 m), 4.2 km (3400–3600 m) and 1.3 km (3600–3800 m), respectively.

The sites have typical vegetation and soil attributes for their respective elevations. The alpine meadow plant community at 3200 m is largely dominated by *Kobresia humilis*, *Elymus nutans*, *Stipa aliena*, *Potentilla anserine* and *Thalictrum alpinum*. The plant community at 3400 m is primarily dominated by *Potentilla fruticosa* shrub meadow and grass species of *K. humilis*, *E. nutans* and *Festuca ovina*. The plant community at 3600 m site is dominated by *K. humilis*, *Potentilla nivea*, *Thalictrum alpinum*, *Carex atrofusca*, *Poa crymophila* and *P. fruticosa*. At the 3800 site, the plant community is dominated by *K. humilis*, *P. crymophila*, *Androsace mariae*, *Polygonum macrophyllum* and *Kobresia pygmaea*. Due to the short growth period, aboveground plant biomass has low primary production and diversity.

Three $1.0 \times 1.0 \times 0.3$ m^3 plots were fenced at each elevation/site to prevent disturbance. The distance between two adjacent plots was roughly 0.6 m. In August 2009, soil at a depth of 0–20 cm was collected from all plots. Briefly, soil samples were collected randomly at five locations of every plot to ensure homogeneity. Then soil cores were mixed thoroughly on a clean tray. After materials such as roots, stones, pebbles and gravels were removed, soil was combined into a composite sample. Soil was sieved through a 2 mm sieve and stored at 4 °C for soil attribute measurements or −80 °C until DNA extraction. All tools were sterilized with 70% alcohol.

DNA extraction

Soil metagenomic DNA was extracted using a FastDNA spin kit for soil (MP Biomedical, Carlsbad, CA, USA) following the manufacturer's instructions and precipitated with 100% ethanol and 0.3 M NaOAc. DNA purity was assessed by UV absorbance ratios of A260/A280 (>1.8) and A260/A230 (>1.7), and DNA concentrations were measured with a PicoGreen method [4].

GeoChip 4.0 experiment

The labeling and hybridization of soil DNA were conducted as previously prescribed [5]. A total of 2 μg extracted DNA was mixed with 20 μl random primers, containing 2.5 μl deoxynucleoside triphosphate (dNTP) (5 mM dATP/dGTP/dCTP, 2.5 mM dTTP), 1 μl Cy5 dUTP (Amersham, Piscataway, NJ) and 80 U of the large Klenow fragment (Invitrogen, Carlsbad, CA). Then DNA mixture was treated at 99.9 °C for 5 min and chilled immediately to denature DNA, followed by addition of 2.5 μl of water and incubation at 37 °C for 3 h. Finally, the mixture was heated at 95 °C for 3 min to terminate DNA labeling.

Labeled DNA was purified using the QIA quick purification kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions and measured by NanoDrop ND-1000 spectrophotometer to assess labeling efficiency. DNA was then dried in the SpeedVac (ThermoSavant, Milford, MA, USA) at 45 °C for 45 min.

DNA hybridization

Labeled DNA was dissolved in 50 μl hybridization buffer (40% formamide, 25% SSC, 5 μg of unlabeled herring sperm DNA [Promega, Madison, WI], and 0.1% SDS) and 2 μl universal standard DNA (0.2 pmol μl^{-1}) labeled with fluorescent dye Cy5. The samples were then mixed by vortexing, incubated at 95 °C for 5 min, and maintained at 50 °C until hybridization. Microarrays were scanned by a NimbleGen MS 200 Microarray Scanner (Roche NimbleGen, Madison, WI) for approximately 16 h at 42 °C. Then scanned images were quantified by NimbleScan software as previously described [6].

Raw data processing

Data of signal intensities were uploaded to the laboratory's Microarray Data Manager System (http://ieg.ou.edu/microarray/) [1,5,6]. Then we processed them in the following steps: (i) spots of poor quality were removed, which were flagged as 1 or 3 by ImaGene (Arrayit, Sunnyvale, CA, USA) or with a signal to noise ratio of less than 2.0; (ii) the relative abundance of each sample was calculated by dividing the total intensity of the microarray, then multiplying by a constant and applying natural logarithm transformation; and (iii) probes detected in only one out of three replicates were removed to improve data quality.

Statistical analysis

Principal component analysis (PCA) was used to measure the overall functional gene structure. Bray–Curtis distance was used to obtain dissimilarity matrices in the adonis algorithm of the dissimilarity test for comparing GeoChip data of four elevations. The similarity test, Mantel test, Canonical correspondence analysis (CCA) and Variation partitioning analysis (VPA) were used to evaluate the linkages between microbial gene compositions and environmental attributes. In the similarity test, Euclidean distance was used to calculate the distance between samples, followed by calculation of Pearson correlation coefficient. To select attributes in CCA modeling, we used variation inflation factors (VIF) to examine whether the variance of canonical coefficients was inflated by the presence of correlations with other attributes. If an attribute had a variation inflation factor value larger than 20, we deemed it to depend on other attributes and consequently removed it from the CCA modeling. Correlation coefficients (r) were calculated using Pearson's correlation. The normalized total gene abundance for each functional gene was the average of the total gene abundance from all the replicates and all data are presented as mean ± s.e. The least significant difference (LSD) test was used to compare the significance of differences in relative abundance among four elevations. All of the analyses were performed with the Vegan package (v.1.15-1) using R, version 2.8.1 (R Foundation for Statistical Computing, Vienna, Austria).

Discussion

Here we describe a dataset of GeoChip 4.0 for profiling functional potentials of microbial community along four elevations in a grassland of the Tibetan plateau (Table 1). GeoChip is comprised of approximately 82,000 probes covering 410 functional gene families related to microbial carbon, nitrogen, sulfur, phosphorus cycling and others [6]. We showed that microbial gene abundances were correlated with greenhouse gas emissions. Therefore, it is possible to assess soil biogeochemical cycles based on measurements of microbial gene abundance.

Table 1
Number of detected genes at four elevations.

Gene categories	3200	3400	3600	3800	All elevations[a]
Antibiotic resistance	903	1668	1548	1495	1818
Bacteria phage	158	424	358	336	468
Bioleaching	192	389	341	329	431
Carbon cycling	2979	5953	5485	5384	6476
Energy process	258	516	479	471	559
Metal Resistance	2832	5352	4999	4887	5757
Nitrogen	2110	4181	3889	3820	4520
Organic Remediation	5926	10,692	10,333	10,122	11,490
Other category	561	1192	1101	1069	1324
Phosphorus	370	764	687	682	829
Stress	5389	11,010	9828	9616	11,958
Sulfur	773	1724	1591	1553	1894
Virulence	918	1820	1643	1574	1996
Total	23,369	45,685	42,282	41,338	49,520

[a] The number of genes detected at any of all four elevations.

Furthermore, it can be used to predict the impact of further climate changes in this region on functional potentials of microbial communities.

Conflict of interest

The authors declare that there is no conflict of interest on any work published in this paper.

Acknowledgments

We thank Haibei Research Station staff for sampling assistance. This research was supported by grants to Yunfeng Yang from the National Science Foundation of China (41171201) and the National >High Technology Research and Development Program of China (2012AA061401). To Shiping Wang from the National Basic Research Program (2010CB833502), to Jizhong Zhou from the United States Department of Energy, Biological Systems Research on the Role of Microbial Communities in C Cycling Program (DE-SC0004601), and Oklahoma Bioenergy Center (OBC). The GeoChips and associated computational pipelines used in this study were supported by ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assemblies through the Office of Science, Office of Biological and Environmental Research, of the US Department of Energy under Contract No. DE-AC02-05CH11231 and by the United States Department of Agriculture (Project 2007-35319-18305) through NSF-USDA Microbial Observatories Program.

References

[1] Y.F. Yang, Y. Gao, S.P. Wang, D.P. Xu, H. Yu, L.W. Wu, Q.Y. Lin, Y.G. Hu, X.Z. Li, Z.L. He, Y. Deng, J.Z. Zhou, The microbial gene diversity along an elevation gradient of the Tibetan grassland. ISME J. 8 (2014) 430–440.
[2] G.M. Cao, Y.H. Tang, W.H. Mo, Y.A. Wang, Y.N. Li, X.Q. Zhao, Grazing intensity alters soil respiration in an alpine meadow on the Tibetan plateau. Soil Biol. Biochem. 36 (2004) 237–243.
[3] L. Zhao, Y.N. Li, S.X. Xu, H.K. Zhou, S. Gu, G.R. Yu, X.Q. Zhao, Diurnal, seasonal and annual variation in net ecosystem CO_2 exchange of an alpine shrubland on Qinghai-Tibetan plateau. Glob. Chang. Biol. 12 (2006) 1940–1953.
[4] S.J. Ahn, J. Costa, J.R. Emanuel, PicoGreen quantitation of DNA: effective evaluation of samples pre- or post-PCR. Nucleic Acids Res. 24 (1996) (3282-3282).
[5] Y. Yang, L. Wu, Q. Lin, M. Yuan, D. Xu, H. Yu, Y. Hu, J. Duan, X. Li, Z. He, Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland. Glob. Chang. Biol. 19 (2013) 637–648.
[6] Q. Tu, H. Yu, Z. He, Y. Deng, L. Wu, J.D. Van Nostrand, A. Zhou, J. Voordeckers, Y.-J. Lee, Y. Qin, C.L. Hemme, Z. Shi, K. Xue, T. Yuan, A. Wang, J. Zhou, GeoChip 4: a functional gene array-based high throughput environmental technology for microbial community analysis. Mol. Ecol. Resour. (2014), http://dx.doi.org/10.1111/1755-0998.12239.