Data in Brief

Metagenomic exploration of microbial community in mine tailings of Malanjkhand copper project, India

Abhishek Gupta a, Avishek Dutta b, Jayeeta Sarkar a, Dhiraj Paul a, Mruganka Kumar Panigrahi c, Pinaki Sar a,⁎

a Department of Biotechnology, Indian Institute of Technology Kharagpur, West Bengal 721302, India
b School of Bioscience, Indian Institute of Technology Kharagpur, West Bengal 721302, India
c Department of Geology and Geophysics, Indian Institute of Technology Kharagpur, West Bengal 721302, India

A R T I C L E   I N F O

Article history:
Received 26 January 2017
Accepted 3 February 2017
Available online 04 February 2017

Keywords:
Mine tailings
Metagenomics
Next generation sequencing
Microbial community
Acid mine drainage

A B S T R A C T

Mine tailings from copper mines are considered as one of the sources of highly hazardous acid mine drainage (AMD) due to bio-oxidation of its sulfidic constituents. This study was designed to understand microbial community composition and potential for acid generation using samples from mine tailings of Malanjkhand copper project (MCP), India through 16S RNA gene based amplicon sequencing approach (targeting V4 region). Three tailings samples (T1, T2 and T3) with varied physiochemical properties selected for the study revealed distinct microbial assemblages. Sample (T3) with most extreme nature (pH < 2.0) harbored Proteobacteria, Actinobacteria, Chloroflexi while the samples (T1 and T3) with slightly moderate nature (pH < 4.0 and > 3.0) exhibited abundance of Proteobacteria, Fimicutes, Actinobacteria and/or Nitrospirae. Metagenomic sequences are available under the BioProject ID PRJNA361456.

© 2017 The Authors. 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: Mine tailings sediment
Sex: Not applicable
Sequencer or array type: Ion Torrent S5 Platform
Data format: Raw data: FASTQ file
Experimental factors: Environmental samples
Experimental features: 16S rRNA gene amplicon (V4 region) sequencing using Ion Torrent platform and diversity analysis using QIME 1.9.0
Consent: Not applicable
Sample source location: Tailing dam of Malanjkhand copper project, Balaghat district, M.P., India. T1 (N21°59.096″ E080°42.427″), T2 (N21°59.096″ E080°42.427″) and T3 (N21°59.915″ E080°42.081″)

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/bioproject/361456

2. Experimental design, materials and methods

Mine tailings are characterized with low –pH, – organic matters and other nutrients, as well as high concentrations of toxic metals and sulfate which pose a severe environmental threat [2 and 3]. Several studies have been conducted worldwide to understand microbial community composition within mine tailings and their metabolic role in biogeochemical cycles of acid generation using either clone library/cultivation or pyrosequencing based approach [3,4,5,6,7,8, and 9]. In recent years, advances in semiconductor based next generation sequencing technology has shown high promises to gain better insights into the microbial community structure and dynamics in natural ecosystems. Numerous studies have been conducted to understand the microbial community structure and function of AMD ecosystem [10,11,12,13,14,15 and 16]. However, dynamics of microbial community composition from mine tailings (source) to the generation of AMD (product) remains less explored. This study was designed to target unexplored mine tailings of Malanjkhand copper project (MCP), Asia’s biggest copper mine, situated in Balaghat district of Madhya Pradesh, India to analyze in-depth microbial community composition and their potential role in acid generation.

2.1. Sampling, DNA extraction and sequencing

Tailing sediments (T1 and T2) were collected from tailing dam and its seepage point (T3) of MCP during April 2014 (T1 and T3) and September 2016 (T2). Samples were collected in sterile container and stored at 4 °C, till further analysis. Metagenome was extracted from 0.25 g of each sample using Power soil DNA isolation kit (MoBio laboratories) according to the manufacturer protocol, quantified in Qubit 3.0...
fluorometer (Invitrogen, Thermo Fisher Scientific) and V4 region of 16S rRNA gene was amplified with V4 specific primers [17]. Each forward primer was tagged with 10-12 bp barcode for multiplexing during sequencing run. PCR was performed in 25 μl of reaction using AmpliTaq Gold 360 Master Mix (Invitrogen, Thermo Fisher Scientific), 40 pico mole each fusion primers and 10–50 ng of template DNA. PCR amplified products (V4 amplicons libraries) were extracted using E gel base system (Invitrogen, Thermo Fisher Scientific). Purified amplicons were processed and sequenced in Ion S5 sequencer.

2.2. Bioinformatic processing of 16S rRNA gene sequence and diversity analysis

Total 0.97 million raw reads were obtained after sequencing. These reads were analyzed in QIIME pipeline [18]. Quality filtering was performed using split_libraries.py to remove primers, sequences with homopolymers run of > 6 bp and read length beyond the range of 230-300 bp. Minimum 3 primer mismatches were allowed in this step due to degeneracy of primer set. The output files from three samples were merged together and pick_closed_reference_otus.py was used to pick OTUs, assign taxonomy and create OTU table against a SILVA 119 reference database (www.arb-silva.de/documentation/release-119). Number of OTUs was obtained using alpha_diversity.py. Agglomerative hierarchical clustering on the abundance of major genera (cumulative abundance > 0.5%) were performed with weighted pair average (WPGA) clustering method using Bray-Curtis dissimilarity distance matrix in XLSTAT 2014 software.

Out of 0.7 million operational reads (obtained after quality filtering), 0.58 million reads were assigned to taxonomy using closed reference method. Alpha diversity indicator revealed that T1 was highly diverse (OTUs 4885) followed by T3 (OTUs 3776) and T2 (OTUs 211). Sample (T3) with most extreme nature (pH < 2.0) harbored Proteobacteria, Actinobacteria and Chloroflexi while the samples (T1 and T2) with slightly moderate nature (pH > 4.0 and < 3.0) exhibited abundance of Proteobacteria, Firmicutes, Actinobacteria and/or Nitrospirae. Archaeabacterial phyla, Euryarchaeota and Thaumarchaeota were abundant in the extreme sample (T3) and accounted for 7.2% of total assigned reads. (Fig. 1).

**Fig. 1.** Distribution of taxa at phylum level across the samples (cumulative abundance > 0.5%).

**Fig. 2.** Weighted pair average group based hierarchical clustering with Bray Curtis dissimilarity distance matrix of major abundant genera (cumulative abundance > 0.5%) across the samples.
T1 and T2 samples showed commonality with presence of genera *Sulfobacillus* and *Alicyclobacillus*. T1 showed specific abundance of *Halomonas*, *Corynebacterium*, *Staphylococcus*, *Rhizobium*, *Acinetobacter* and *Stenotrophomonas* while T2 harbored *Leptospirillum*, *Acidithiobacillus*, *Acidiferrobacter* and *Acidiphilium*. In contrast, T3 sample showed distinct abundance of *Ralstonia*, *Metallobacterium*, *Acidovorax*, *Rhodanobacter*, uncultured-*Chloroflexi*, -*Chitinophagaceae*, -*Thermoplasmatales*, -*Thaumarcheota* and *Thiobacillus*. Hierarchical clustering on abundance of these genera across the samples was also evident in Fig. 2, indicating distinct patterns of assemblage among the organisms with common metabolic/biogeochemical functions. Detection of major genera such as *Sulfobacillus*, *Leptospirillum*, *Acidithiobacillus*, *Acidiferrobacter*, *Metallobacterium*, *Acidiphilium* and *Alicyclobacillus* in the tailings samples indicates the intrinsic acid generation potential of the microbial community. Observed variation in community composition among the samples could be attributed to the prevalent physicochemical conditions and their role in shaping the community structure. This study provides an insight into the microbial community structure and their relation to biogeochemistry of copper mine tailings.

**Acknowledgements**

This work was financially supported by the Department of Biotechnology (DBT), Govt. of India (BT/PR 7533/BCE/8/959/2013). Next generation facility with Ion S5 is funded by the Indian Institute of Technology (IIT) Kharagpur (IIT/SRIC/BT/ODM/2015–16/141) through SCBSI challenge grant. Abhishek Gupta (DBT-JRF) is a recipient of DBT/2013-14/131BT91P01) institutional fellowship (with Ion S5) is funded by the Indian Institute of Technology (DBT), Govt. of India (BT/PR 7533/BCE/8/959/2013). Next generation technology at Iron Mountain, California. Geochem. Trans. 5 (2) (2003) 37–84.

**References**

[1] A. Schippers, A. Breuker, A. Blazejak, K. Bosecker, D. Kock, T.L. Wright, The biogeochemistry and microbiology of sulfidic mine waste and bioleaching dumps and heaps, and novel Fe (II)-oxidizing bacteria. Hydrometallurgy 104 (3) (2010) 342–350.

[2] M.O. Mendez, J.W. Neilson, R.M. Maier, Characterization of a bacterial community in an abandoned semi-arid lead-zinc mine tailing site. Appl. Environ. Microbiol. 74 (12) (2008) 3899–3907.

[3] G.L. Tan, W.S. Shu, K.B. Hallberg, F. Li, C.Y. Lan, W.H. Zhou, L.N. Huang, Culturable and molecular phylogenetic diversity of microorganisms in an open-dumped, extremely acidic Pb/Zn mine tailings. Extremophiles 12 (5) (2008) 657–664.

[4] D. Kock, A. Schippers, Geomicrobiological investigation of two different mine waste tailings generating acid mine drainage. Hydrometallurgy 83 (1) (2006) 167–175.

[5] N. Diaby, B. Dold, H.R. Pfeifer, C. Holliger, D.B. Johnson, K.B. Hallberg, Microbial communities in a porphyry copper tailings impoundment and their impact on the geochemical dynamics of the mine waste. Environ. Microbiol. 9 (2) (2007) 298–307.

[6] D. Kock, A. Schippers, Quantitative microbial community analysis of three different sulfidic mine tailing dumps generating acid mine drainage. Appl. Environ. Microbiol. 74 (16) (2008) 5211–5219.

[7] L.N. Huang, W.H. Zhou, K.B. Hallberg, C.Y. Wan, J. Li, W.S. Shu, Spatial and temporal analysis of the microbial community in the tailings of a Pb-Zn mine generating acidic drainage. Appl. Environ. Microbiol. 77 (15) (2011) 5540–5544.

[8] L.X. Chen, J.T. Li, Y.T. Chen, L.N. Huang, Z.S. Hua, M. Hu, W.S. Shu, Shifts in microbial community composition and function in the acidification of lead/zinc mine tailings. Environ. Microbiol. 15 (9) (2013) 2431–2444.

[9] P.K. Dha, P. Sar, Microbial communities in uranium mine tailings and mine water sediment from Jaduguda U mine, India: a culture-independent analysis. J. Environ. Sci. Health A 6 (2014 May 49).

[10] K.B. Hallberg, D.B. Johnson, Biodiversity of acidophilic prokaryotes. Adv. Appl. Microbiol. 49 (2001) 37–84.

[11] B.J. Baker, J.F. Banfield, Microbial communities in acid mine drainage. FEMS Microbiol. Ecol. 44 (2) (2002) 139–152.

[12] G.L. Tan, W.S. Shu, K.B. Hallberg, F. Li, C.Y. Lan, L.N. Huang, Cultivation-dependent and cultivation-independent characterization of the microbial community in acid mine drainage associated with acidic Pb/Zn mine tailings at Lechang, Guangdong, China. FEMS Microbiol. Ecol. 59 (1) (2007) 118–126.

[13] E. González-Torilli, E. Llobet-Brossa, E.O. Casamayor, R. Amann, R. Amils, Microbial ecology of an extreme acidic environment, the Tinto River. Appl. Environ. Microbiol. 69 (8) (2003) 4853–4865.

[14] G.K. Druschel, B.J. Baker, T.M. Ghring, J.F. Banfield, Acid mine drainage biogeochemistry at Iron Mountain, California. Geochem. Trans. 5 (2) (2003) 4865–4865.

[15] P.N. Bertin, A. Heinrich-Salmeron, E. Pelletier, F. Gouilhen-Chollet, F. Arsène-Ploetze, S. Gallien, V. Bonmerfoz. Metalic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta–proteogenomics. ISME J. 5 (11) (2011) 1733–1747.

[16] J.L. Kuang, L.N. Huang, L.X. Chen, Z.S. Hua, S.J. Li, M. Hu, W.S. Shu, Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. ISME J. 7 (5) (2013) 1038–1050.

[17] S.T. Bates, D. Berg-Lyons, J.G. Caporaso, W.A. Walters, R. Knight, N. Fierer, Examining the global distribution of dominant archeal populations in soil. ISME J. 5 (5) (2011) 908–917.

[18] J.G. Caporaso, R. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, G.A. Huttenle, QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7 (5) (2010) 335–336.