

Please cite the Published Version

Brearley, FQ (2020) Sequence data describing the fungal community in a tropical quartzite soil. Data in Brief, 29.

DOI: <https://doi.org/10.1016/j.dib.2020.105112>

Publisher: Elsevier

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/625403/>

Usage rights:  [Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Additional Information: This is an Open Access article published in Data in Brief, published by Elsevier, copyright The Author(s).

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)



ELSEVIER

Contents lists available at ScienceDirect

Data in brief

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



Data Article

Sequence data describing the fungal community in a tropical quartzite soil



Francis Q. Brearley

Department of Natural Sciences, Manchester Metropolitan University, Chester Street, Manchester, M1 5GD, UK

ARTICLE INFO

Article history:

Received 23 May 2019

Received in revised form 16 December 2019

Accepted 3 January 2020

Available online 10 January 2020

Keywords:

Amplicon sequencing

Fungal diversity

Illumina

Metagenome

Quartz

Soil

Tropical

ABSTRACT

The soil fungal community of the Klang Gates quartz ridge in Malaysia was determined by ITS amplicon sequencing using the Illumina HiSeq platform. The community contained 2767 OTUs, 47% of which could not be assigned to a phylum, likely representing new lineages. Those that could be assigned were found within 5 phyla, 16 classes, 49 orders and 98 families with over 85% of these within the Ascomycota. Sequence data is available from the NCBI's Sequence Read Archive (PRJNA542066). This data illustrates the microbial diversity in a particularly nutrient poor tropical soil and can be used for broader-scale comparisons of microbial distributions.

© 2020 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/>).

1. Data

Due to their high diversity and key roles in biogeochemical cycles, it is important to better study microbial (and particularly fungal) communities in understudied tropical soils that may hold novel lineages with associated novel genes, biomolecular pathways and compounds. Klang Gates quartz ridge is found near Kuala Lumpur, Malaysia (3° 14' N, 101° 44' E) – it is one of the largest and longest quartz dykes in the world formed when residual magma crystallised and consolidated within the

E-mail address: f.q.brearley@mmu.ac.uk.

<https://doi.org/10.1016/j.dib.2020.105112>

2352-3409/© 2020 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 Table

Subject area	<i>Environmental microbiology</i>
More specific subject area	<i>Fungal community ecology</i>
Type of data	<i>Fastq files</i>
How data was acquired	<i>Illumina HiSeq amplicon sequencing of fungal community ITS regions from soil samples</i>
Data format	<i>Raw (trimmed) Fastq files</i>
Experimental factors	<i>Random sampling of environmental samples</i>
Experimental features	<i>Soil fungal DNA amplicon ITS sequencing by Illumina HiSeq with subsequent bioinformatic analysis</i>
Data source location	<i>Klang Gates quartz ridge, Kuala Lumpur, Malaysia (3° 14' N, 101° 44' E)</i>
Data accessibility	<i>Files available via NCBI's Sequence Read Archive https://www.ncbi.nlm.nih.gov/sra/PRJNA542066</i>

Value of the Data

- This data presents the first enumeration of fungal community structure in a very poorly studied tropical soil type, i.e. that developed on a nutrient-poor quartz dyke.
- The data can be used for broad-scale comparisons of fungal community diversity both in the tropics and with other biomes across the globe.
- The fungal community was dominated by Ascomycota, but nearly half of the OTUs could not be assigned to a phylum. This data therefore provides a data source for discovering new fungal taxa.

surrounding granitic rock [1]. On the top of the ridge, the quartz has weathered into coarse sand that is low in nutrient and water holding capacity. The vegetation here (Fig. 1a) is distinct from the forest surrounding the ridge as it is very stunted (<3 m tall) and has a number of endemic plant species [2,3].

Soil fungal DNA amplicon ITS sequencing from four soil samples by Illumina HiSeq and subsequent bioinformatic analysis generated 3,913,957 sequence reads belonging to 2767 fungal OTUs. The mean number of OTUs per sample was 1005 (\pm S.D. 396) with 3% of the OTUs found in all four samples and 68% found in only one sample. The fungal community showed a typical rank-abundance curve (Fig. 1b). Notably, 1303 OTUs (47%) could not be assigned to a phylum and may, therefore, represent new fungal lineages [4]. Those that could be further classified were found in 5 phyla, 16 classes, 49 orders, 98 families and 206 genera (Fig. 1c). However, these numbers should be considered lower estimates due to the lack of suitable reference sequences to assign all the taxa. Ascomycota dominated the community with over 95% of the sequence reads that could be assigned to a phylum (particularly in the class Sordariomycetes that comprised about 78% of the total reads) followed by Basidiomycota (about 5%) with other fungal phyla comprising less than 0.2% of the reads (Fig. 1c). The most abundant families were (in order of abundance): Ophiocordycipitaceae, Cordycipitaceae, Cephalothecaceae, Trichocormaceae, Herpotrichiellaceae, Nectriaceae and Mycosphaerellaceae each comprising over 1% of the total reads. Seventy-two percent of the OTUs could not be assigned to a genus but the most common genera (in order of abundance, each comprising more than 1% of the reads) were *Purpureocillium*, *Isaria*, *Cephalotheca*, *Penicillium*, *Exophiala* and *Fusarium*. In terms of numbers of OTUs, Ascomycota dominated the community again (85% of OTUs) with Basidiomycota contributing 13% of the total OTUs; other fungal phyla made up less than 1% of the OTUs each.

This data presents the first analysis of the microbial community in a low nutrient tropical quartz dyke soil that was dominated by Ascomycota but with a large proportion of OTUs, and indeed lineages, new to science and can be used in larger-scale comparisons of the influence of edaphic, and other, factors on soil fungal diversity and distributions.

2. Experimental design, materials and methods

Four soil samples were collected from the site (April 2016) and stored in Lifeguard soil preservation solution (Qiagen, Manchester, UK) until DNA extraction using a PowerSoil DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA) with one minute in a FastPrep 120 (Thermo Electron, Waltham, MA, USA) at the first stage. To amplify the ITS1 region, the primers ITS5 and 5.8S_fungi [5] were used in a first-round PCR for each of the four samples with 1 μ l of DNA in a reaction with Q5 Hot Start High-

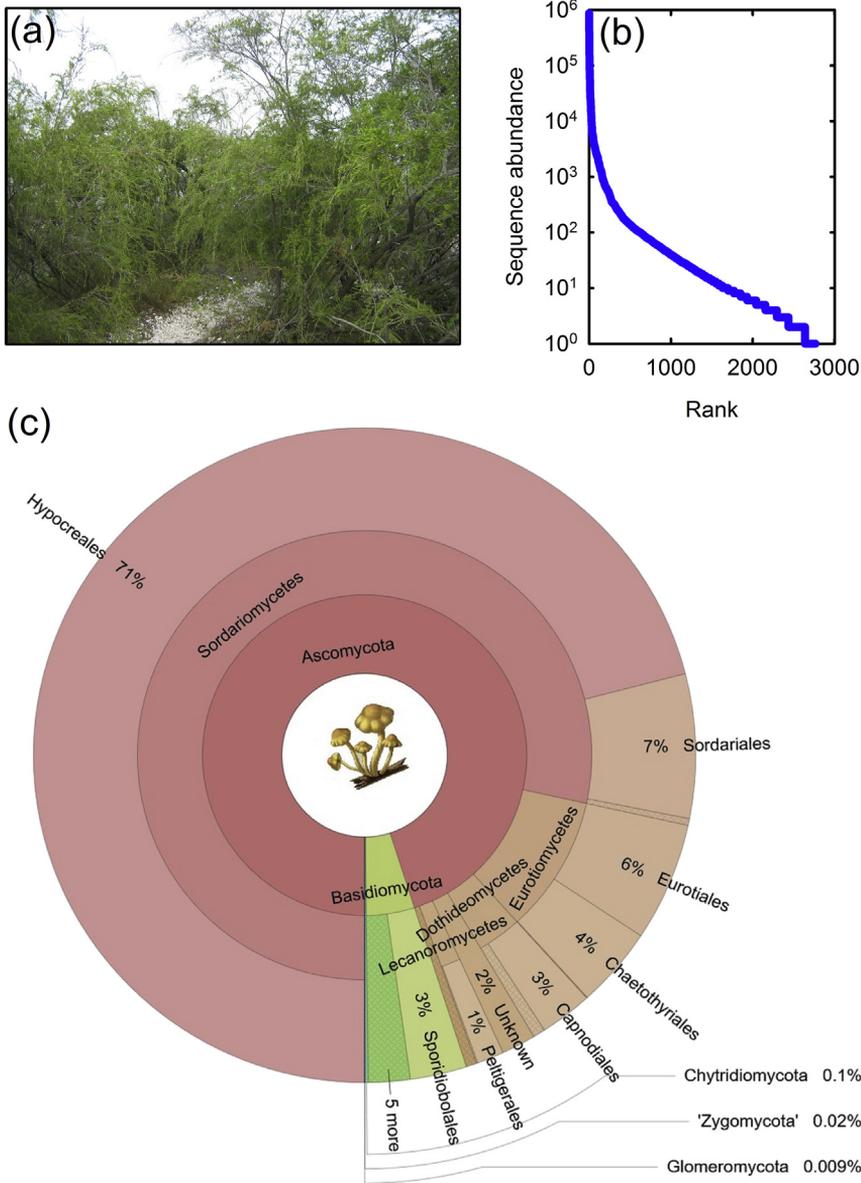


Fig. 1. **a:** View of the top of Klang Gates quartz ridge, Kuala Lumpur, Malaysia showing poor soil development **b:** Rank abundance curve of fungal community **c:** Krona [14] chart showing the taxonomic affiliation of fungal OTUs.

Fidelity Master Mix (New England BioLabs, Ipswich, MA, USA) totaling 25 μ l, with 10 sec at 98 $^{\circ}$ C, 30 sec at 58 $^{\circ}$ C, 30 sec at 72 $^{\circ}$ C for 15 cycles and then a 10 min extension at 72 $^{\circ}$ C. Samples were purified with AMPure SPRI beads (Beckman Coulter, Brea, CA, USA) before entering the second PCR to incorporate Illumina Nextera XT primers. Fifteen PCR cycles were performed with 30 sec at 95 $^{\circ}$ C, 30 sec at 55 $^{\circ}$ C and 30 sec at 72 $^{\circ}$ C, and then a 5 min final extension at 72 $^{\circ}$ C. Samples were again purified using AMPure SPRI beads before being quantified using a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA). Successfully generated amplicon libraries were pooled equimolar and size selected on a Pippin Prep

(Sage Science, Beverly, MA, USA) using a size range of 300–600 bp. The pool of libraries was denatured and loaded at 9 pM concentration. To help balance the complexity of the amplicon library, 15% PhiX was spiked in and the sequencing was carried out on one lane of an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) at 2×300 bp paired-end sequencing with v2 chemistry. Basecalling and demultiplexing was performed by CASAVA v. 1.8.2 (Illumina) with trimming of primers and adapter sequences using Cutadapt v. 1.2.1 [6] and removal of low quality bases using Sickle v. 1.200 [7]. Sequencing errors were corrected using BayesHammer implemented within SPAdes v. 3.1.0 [8]. PEAR v. 0.9.10 [9] was then used to align each pair of reads into a single sequence. Sequences shorter than 150 bp or longer than 600 bp were removed using a custom script, as were those matching PhiX. Sequences were clustered to generate a consensus sequence using Swarm v. 2.2.1 [10] using the smallest granularity, with clusters containing less than two sequences removed. Chimeras (18.4% of sequences) were removed using VSEARCH v. 1.3 [11] using a de-novo approach as well as a reference-based approach against the UNITE v. 12.11 database [12]. For each obtained cluster, the most abundant sequence was used as a representative, and the QIIME [13] script *assign_taxonomy.py* was used to match that sequence to the UNITE ITS database [12] using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Sequences not matching the Kingdom Fungi were removed from further analysis.

Acknowledgements

I thank Manchester Metropolitan University for funding this work, the Centre for Genomic Research at the University of Liverpool for sequencing, Luca Lenzi and Graeme Fox for bioinformatic support, and Central Selangor Regional Forest Office for access permission.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2020.105112>.

References

- [1] M.R. Umor, et al., Morphology, field observations and petrographic descriptions of Gombak Selangor quartz ridge, Hulu Klang, Selangor, Malaysia, *J. Fundam. Appl. Sci.* 10 (2018) 422–433.
- [2] R. Kiew, The Klang Gates ridge, Malaysia, *Malay. Nat.* 36 (1982) 22–28.
- [3] K.M. Wong, et al., Ecological aspects of endemic plant populations on Klang Gates quartz ridge, a habitat island in Peninsular Malaysia, *Biodiv. Conserv.* 19 (2010) 435–447.
- [4] L. Tedersoo, et al., Novel soil-inhabiting clades fill gaps in the fungal tree of life, *Microbiome* 5 (2017) 42.
- [5] L.S. Epp, et al., New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems, *Mol. Ecol.* 8 (2012) 1821–1833.
- [6] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet J.* 17 (2011) 10–12.
- [7] N.A. Joshi, J.N. Fass, Sickle: a Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33) [Software], 2011. Available at: <https://github.com/najoshi/sickle>.
- [8] A. Bankevich, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (2012) 455–477.
- [9] J. Zhang, et al., PEAR: a fast and accurate Illumina Paired-End reAd mergeR, *Bioinformatics* 30 (2014) 614–620.
- [10] F. Mahé, et al., Swarm v2: highly-scalable and high-resolution amplicon clustering, *PeerJ* 3 (2015), e1420.
- [11] T. Rognes, et al., VSEARCH: a versatile open source tool for metagenomics, *PeerJ* 4 (2016), e2584.
- [12] K. Abarenkov, et al., The UNITE database for molecular identification of fungi – recent updates and future perspectives, *New Phytol.* 186 (2010) 281–285.
- [13] J.G. Caporaso, et al., QIIME allows analysis of high-throughput community sequencing data, *Nat. Methods* 7 (2010) 335–336.
- [14] D. Ondov, et al., Interactive metagenomics visualisation in a web browser, *BMC Bioinf.* 12 (2011) 385.