

# Microbial alpha diversity data

When using these data, please cite:

Mainka, M.; Bakker, L.; Maier, A.; Ruethers, J.; Bakke Westergaard, K.; Fior, S.; Frossard, A.; Griepentrog, M.; Van de Broek, M.; Wilken, F, Dötterl, S.; Alexander, J.; Magnabosco, C. (2025): Arctic Greening Database. v 0.1. GFZ Data Services. <https://doi.org/10.5880/fidgeo.2025.074>

## Introduction

In this data sheet microbial alpha diversity per single bioplot is shown, as mean Shannon index and mean observed diversity calculated from aggregated extraction triplicates of 16S rRNA gene amplicon sequencing for prokaryotes. The data was quality checked and rarefied prior to analysis. The sheet contains the columns plot id, microbe\_shannon and observed.

Table 1: General information.

|                        |                                         |
|------------------------|-----------------------------------------|
| project_name           | ETH+ Arctic Greening                    |
| creation_date          | 2024-11-10                              |
| last_modification_date | 2024-11-10                              |
| creator                | Lena Bakker                             |
| contact_1              | lena.bakker@eaps.ethz.ch                |
| contact_2              | cara.magnabosco@eaps.ethz.ch            |
| number_samples         | 31                                      |
| sampling_period        | 2022-07-09 – 2023-08-09                 |
| processing_status      | incomplete = more data will be uploaded |
| upload_status          | structured                              |

Table 2: Variable overview.

| variable        | explanation                                                                                                                                                             | unit    |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| microbe_shannon | The Shannon diversity of prokaryotic soil microbial communities as a measure of alpha diversity of the rarefied sequencing data                                         | No unit |
| observed        | The observed diversity of prokaryotic soil microbial communities (numbers of different ASVs per sample) as a measure of alpha diversity of the rarefied sequencing data | No unit |

## Methods

### *Soil sampling*

Each of the samples were homogenized by mixing the bags, if the soil was loose enough, if it was not loose, by mixing in a bag using a mortar and pestle. The homogenized soil was then subsampled into cryovials in triplicates and shock frozen in liquid nitrogen. The remaining soil was kept for soil analysis. The molecular samples were kept in -80 °C and transported back in liquid nitrogen (one of the triplicates for potential RNA analysis) or frozen in a cooler (two of the triplicates for DNA analysis).

### *DNA Extraction*

DNA was extracted from 0.25 g of sample material using the commercial soil DNA extraction kit DNeasy PowerSoil Pro Kit (Qiagen®, Germany) with a bead beater, FastPrep 24 (MP Bio-medicals, Germany) for mechanical lysis. DNA concentrations were determined using Qubit fluorometric quantification using the high sensitivity kit (Invitrogen™, USA). All DNA extracts were normalized to 2 ng/μl for further processing.

### *16S rRNA Sequencing*

The 16S rRNA sequencing library was prepared following the protocol proposed in the Earth microbiome project (Parada et al., 2016). Using a one-step PCR barcoding of the V4 and V5 region with the primers 515F (GTGYCAGCMGCCGCGGTAA) (Caporaso et al., 2012) that was barcoded and 926R (CCGYCAATTYMTTTRAGTTT) (Quince et al., 2011). The normalized DNA extracts were amplified in triplicates with unique barcodes, the amplicon concentration was measured on the Qubit as described above and the correct amplicon length was checked on a gel electrophoresis gel (1.5% agarose gel). The amplified product was cleaned with 0.8X AMPure XP magnetic beads (Beckman Coulter, USA) and washed twice with 80% EtOH. Following the cleanup, the concentration of cleaned PCR product was measured on a Qubit as described above and used for the calculation for the normalization and pooling of the samples. The cleaned amplicons of samples were normalized to 2.1 ng/μl and pooled and then combined with the negative controls. The concentration of the final library was at 3.35 nM. The library was sequenced on the Illumina NextSeq2000 (Illumina, USA), on a P1 low output cell, set on PE 300 bp with a spike of 20 % PhiX. Custom primers were used with an index primer (AATGATACGGCGACCACCGAGATCTACACGCT), a forward sequencing primer (TATGGTAATTGTGTGYCAGCMGCCGCGGTAA) and a reverse sequencing primer (CGGCATACGAGATAGTCAGCCAGGGCCGYCAATTYMTTTRAGTTT).

### *Bioinformatic Processing*

Sequences were checked with FastQC (Andrews, 2010) and the quality of the reads was found to be good. All processing of sequencing data was performed on the ETH Zürich Euler cluster. The reads were trimmed with trimmomatic (Bolger et al., 2014) to a length of 230 bases and with a threshold of 28 on a 100 bases sliding window. Following this, the forward and reverse reads were joined using fastq-join (Aronesty, 2013) on a minimum of 20 bases of overlap. The processed full length reads of good quality were then processed using DADA2 (Callahan et al.,

2016), constructing an amplicon sequencing variant (ASV) table with removed chimeras. The taxonomy of the ASVs was done following the DADA2 pipeline and using the silva nr.99 v138.1 train set. Further analysis was done with the phyloseq package (McMurdie and Holmes, 2013) version 1.42.0 and the vegan package version 2.6-4 was used for diversity measurements.

## Sample processing

DNA extract from flash frozen soils stored at -80°C, sequencing on an Illumina NextSeq.

## References

Aronesty, E. Comparison of sequencing utility programs. *Open Bioinforma J* 7: 1-8. 2013.  
<https://doi.org/10.2174/1875036201307010001>

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30, no. 15 (2014): 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature methods* 13, no. 7 (2016): 581-583. <https://doi.org/10.1038/nmeth.3869>

McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." *PloS one* 8, no. 4 (2013): e61217.  
<https://doi.org/10.1371/journal.pone.0061217>

Parada, Alma E., David M. Needham, and Jed A. Fuhrman. "Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples." *Environmental microbiology* 18, no. 5 (2016): 1403-1414. <https://doi.org/10.1111/1462-2920.13023>

Quince, Christopher, Anders Lanzen, Russell J. Davenport, and Peter J. Turnbaugh. "Removing noise from pyrosequenced amplicons." *BMC bioinformatics* 12 (2011): 1-18.  
<https://doi.org/10.1186/1471-2105-12-38>