

Microbiome Profile Report

by BiomCare

10/12/2019

Customer	SEGES
Customer ID	DA00201-19
Project	Differences between bacetria and fungi in plowed and unplowed clay and sand samples
Sample type	Soil
Number of samples	24 samples
Type of data	Sequencing of bacteria (16S rRNA) with spike-in and fungi (ITS2)

Table of contents

- Table of contents
- Overview of microbiome profiles
- Alpha diversity
- Beta diversity
 - Summary statistics of the beta-diversity matrices

Overview of microbiome profiles

Processing of the 16S rRNA gene sequencing data across the 24 samples, resulted in the detection of 7131 unique bacterial sequences. When assigning the sequences at higher taxonomic levels, a total of 209 genera, 136 family, 99 order, 60 class and 24 phyla were detected. At the level of kingdom, 3 Archea and 7128 bacteria were detected, while 0 sequences were not annotated to a known organism at the kingdom level. At the lower taxonomic levels, an increasing number of sequences did not match a know microorganism in the reference database, potentially due to the high complexity of soil samples with many organism only know getting identified and characterized (e.g. 5372 unassigned sequences at the genus level).

The mean read depth across samples was 14365.92 with a minimum read depth of 6953 In further processing and analysis, we have included all 24 samples.

Summary statistics of the microbiome profiles across all samples, and the changes imposed at each step of filtering is listed here:

```
## Overview of the microbiome profile before filtering. At this step, only t
he quality-control samples has been removed.
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 7131 taxa and 24 samples ]
## sample_data() Sample Data: [ 24 samples by 3 sample variables ]
## tax_table() Taxonomy Table: [ 7131 taxa by 6 taxonomic ranks ]
## refseq() DNASTringSet: [ 7131 reference sequences ]
```

```
## Overview of the microbiome profile after
## - filtering of samples to remove any sample with a total read count less than 6000.
## - removal of the two bacteria used in spike-in
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 7092 taxa and 24 samples ]
## sample_data() Sample Data: [ 24 samples by 3 sample variables ]
## tax_table() Taxonomy Table: [ 7092 taxa by 6 taxonomic ranks ]
## refseq() DNASTringSet: [ 7092 reference sequences ]
```

```
## Overview of the microbiome profile after
## - normalization by spike-in bacteria to obtain absolute bacterial masses
## - filtering of samples to remove any sample with a total read count less than 1e8
## - removal of the two bacteria used in spike-in
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 7092 taxa and 24 samples ]
## sample_data() Sample Data: [ 24 samples by 3 sample variables ]
## tax_table() Taxonomy Table: [ 7092 taxa by 6 taxonomic ranks ]
## refseq() DNASTringSet: [ 7092 reference sequences ]
```

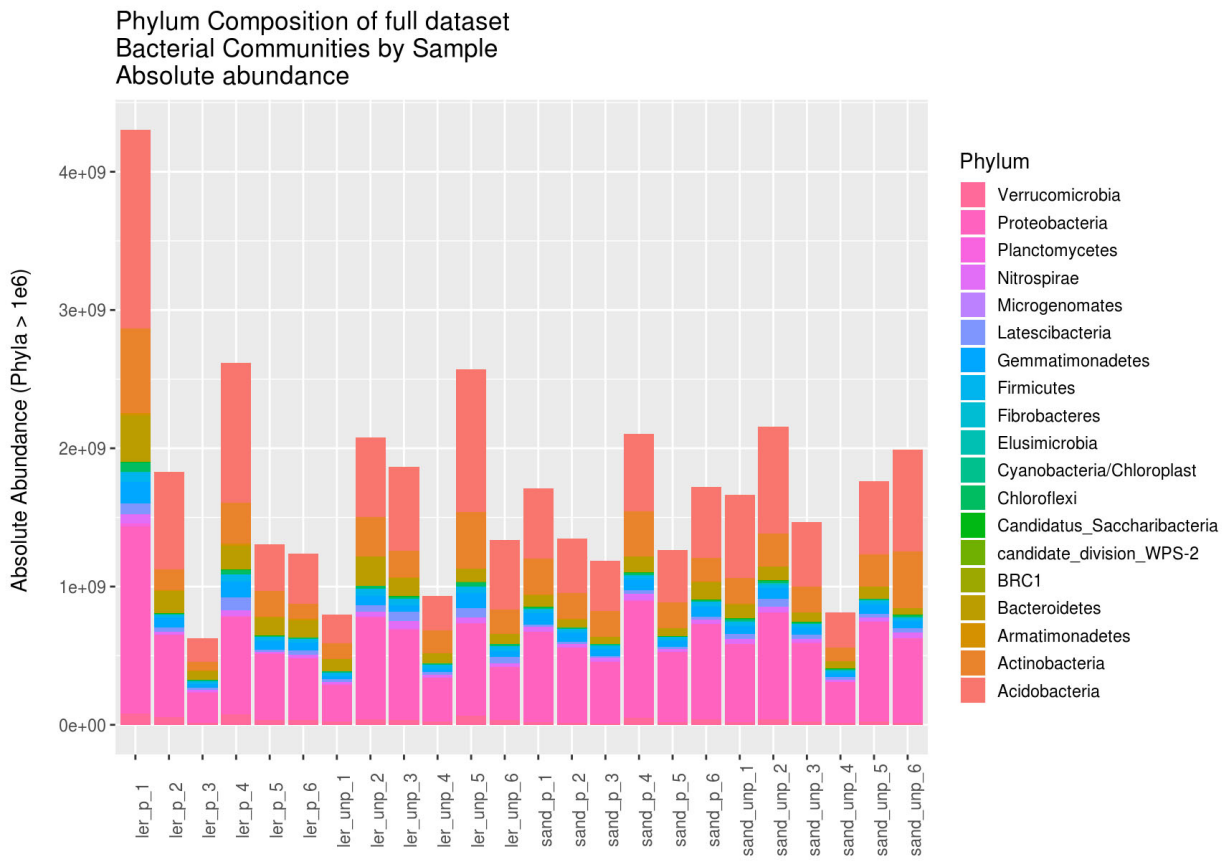
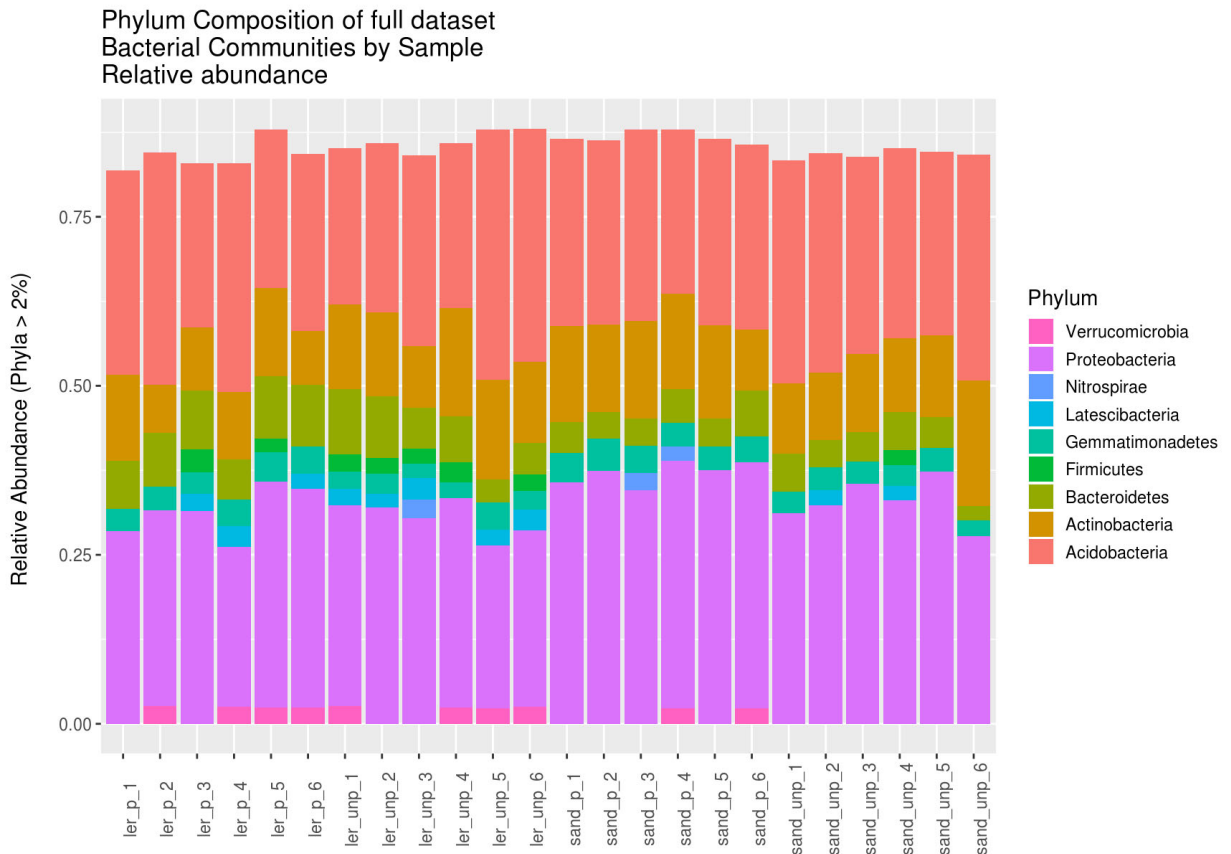


Figure 1: Stacked barplots over relative and absolute abundances.

Alpha diversity

Alpha-diversity is a measure of the diversity within (or complexity) within one microbiome community. Different samples' alpha-diversity score is thus independent of any other sample's score, unlike beta-diversity, described below. The alpha-diversity measures are stored in a matrix which can then be used for later statistical analysis and visualization.

Alpha-diversity is commonly plotted using different measures, most commonly Shannon, Observed and Chao1, explained below. These plots are useful for distinguishing any outliers or trends in sample alpha-diversity.

The measures we plot in Plot X and Y are Shannon, Observed and Chao1. Shannon diversity is the uncertainty in naming the species of a individual at random, explaining diversity in the fact that if one species dominates the sample, it is easy to predict the species of an individual taken at random. Observed is simply the number of observed species within the sample. Chao1 explains diversity by looking at the abundance data for each species, looking specifically at the number of singletons and doubletons in relation to the total number of individuals. For all these measures it is true that a higher value indicates a more diverse sample, and lower value a less diverse sample.

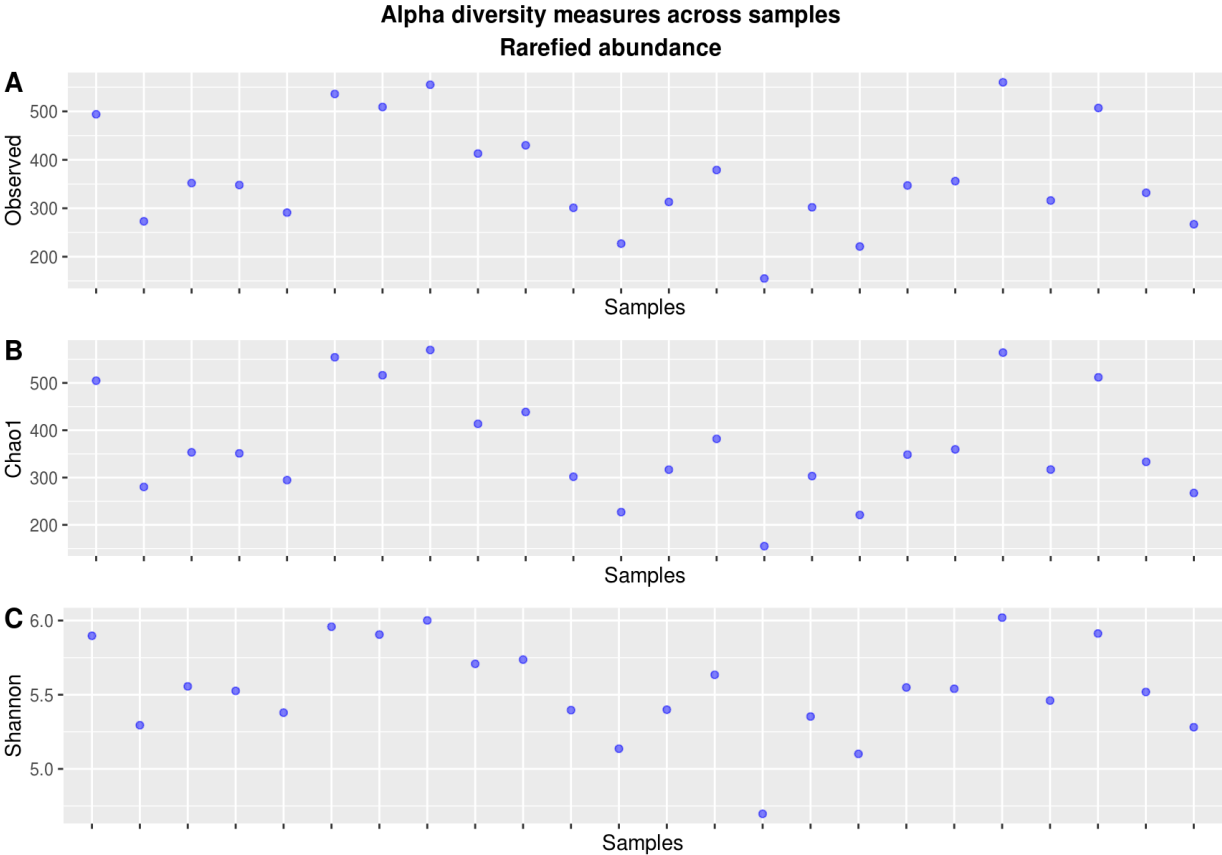


Figure 2: Alpha diversity across samples using rarefied abundance

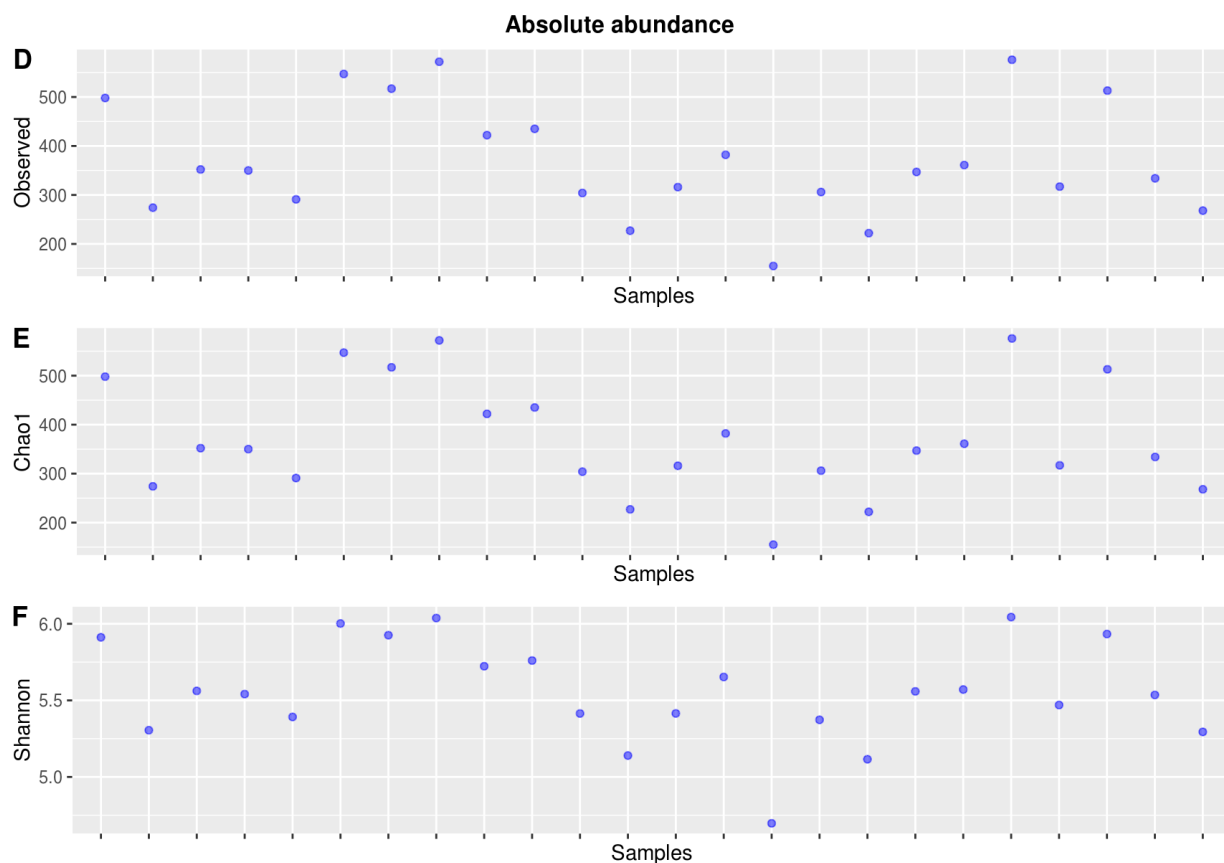


Figure 3: Alpha diversity across samples using absolute abundance

Beta diversity

Beta-diversity is a measure of the diversity (or complexity) of the microbiome community between samples, and therefore differ from alpha-diversity which is a measure of the diversity within samples. Therefore, beta-diversity is a measure of how similar or dissimilar each pair of samples are. The beta-diversity measures are stored in a distance matrix which can then be used for later statistical analysis and visualization.

Beta-diversity measures are often inspected in ordination plots, where each sample is a point and the distance between the points increases with increasing dissimilarity in the microbiome community. Such plots are usefull to explore the data and see which variables explain the similarity or dissimilarity of groups of samples.

We have calculated two beta-diversity measures; Bray-Curtis dissimilarity and Jaccard distance. Bray-Curtis include the taxa abundance in its algorithm, while Jaccard transforms the data to presense/absence (1 or 0). Using both measures allow us to evaluate if abundance is a more important driver than presence/absence of taxa in explaining the patterns of the projects. We calculate both beta-diversity measures, both for the normalized to relative abundance data (normalized by total read count per sample) and for the absolute abundance data. The summary statistics for the calculated matrices are listed below, and can be used to evaluate the overall dissimilarity pattern across the dataset. Both Bray-Curtis and Jaccard range from 0 to 1, with 1 indicating maximum possible difference between samples (no shared sequences). A mean value for a beta-diversity matrix that is clsoe to 1 therefore indicate high dissimilarity between all samples.

Summary statistics of the beta-diversity matrices

Relative abundance

Sequences

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.6385 1.0000  1.0000  0.9851 1.0000  1.0000
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.4689 1.0000  1.0000  0.9757 1.0000  1.0000
```

Genera

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.3638 0.5054  0.5811  0.5644 0.6249  0.7741
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.2224 0.3382  0.4095  0.3982 0.4545  0.6315
```

Family

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.2519 0.4113  0.4944  0.4783 0.5451  0.6710
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.1441 0.2589  0.3284  0.3187 0.3747  0.5049
```

Order

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.1350 0.2694  0.4251  0.3882 0.4816  0.6661
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
## 0.0724 0.1556  0.2700  0.2476 0.3172  0.4993
```

Class

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.08101 0.22443 0.36161 0.32754 0.40952 0.60220
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04222 0.12640 0.22071 0.20116 0.25748 0.43082
```

```
## Phyla
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.04024 0.14932 0.18887 0.19074 0.23848 0.33481
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02053 0.08068 0.10428 0.10679 0.13538 0.20107
```

Absolute abundance

```
## Sequences
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.6327 1.0000 1.0000 0.9869 1.0000 1.0000
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.4627 1.0000 1.0000 0.9783 1.0000 1.0000
```

```
## Genera
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3729 0.5664 0.6356 0.6294 0.7039 0.8884
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.2292 0.3951 0.4659 0.4684 0.5431 0.7992
```

```
## Family
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.2854  0.4862  0.5638  0.5581  0.6268  0.8594
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.1665  0.3212  0.3926  0.3954  0.4564  0.7534
```

```
## Order
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.1862  0.4088  0.5056  0.4956  0.5869  0.8589
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.1026  0.2569  0.3383  0.3409  0.4153  0.7526
```

```
## Class
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.1216  0.3641  0.4531  0.4563  0.5562  0.8541
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.06471 0.22260 0.29290 0.30786 0.38528 0.74538
```

```
## Phyla
```

```
## $jaccard
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.07126 0.24814 0.35132 0.38514 0.51847 0.85420
##
## $bray
##   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
##  0.03695 0.14165 0.21309 0.25361 0.34996 0.74550
```

Nonmetric Multidimensional Scaling (NMDS)

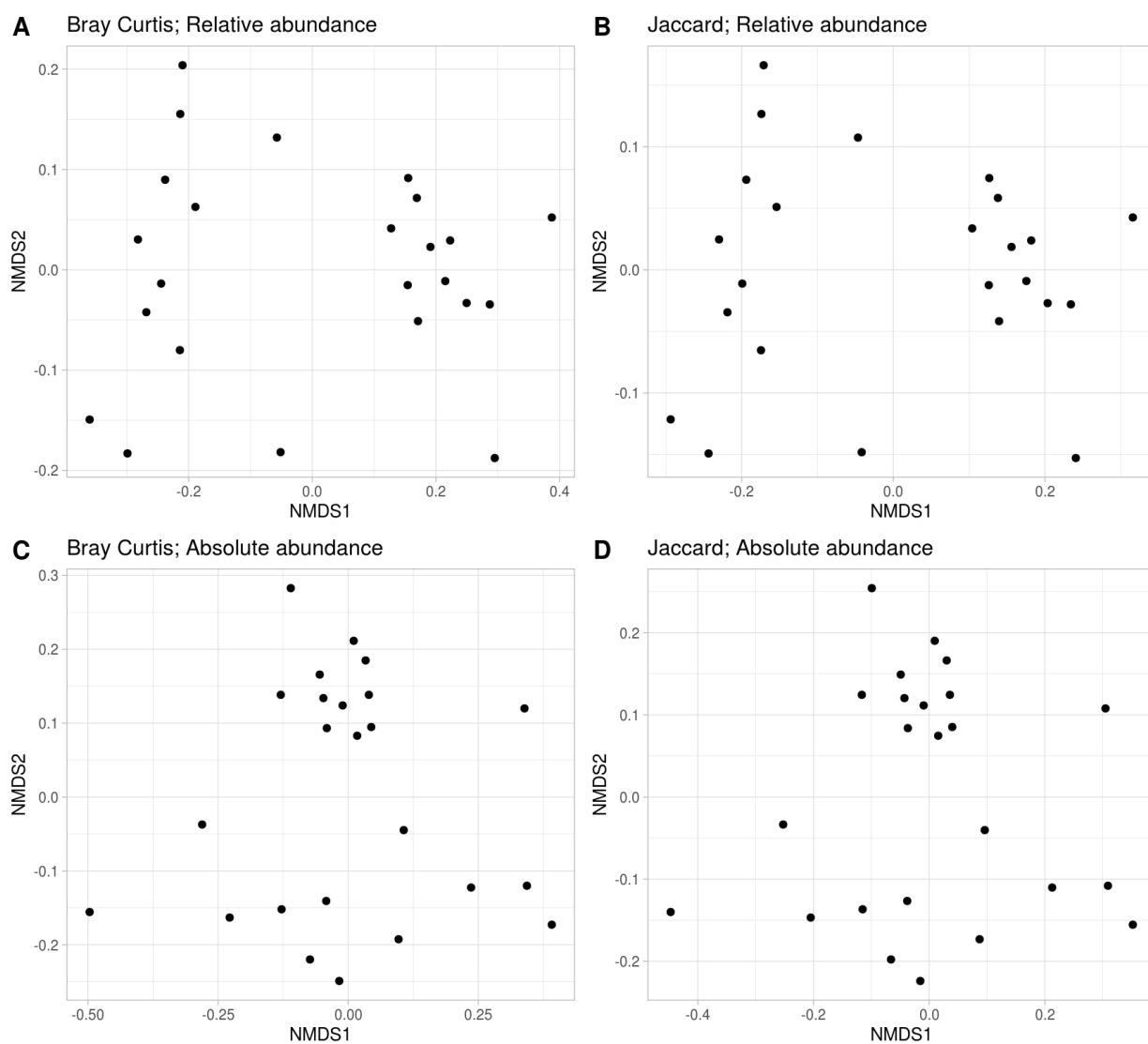


Figure 4: Nonmetric multidimensional scaling across the 24 soil samples. Plots are made using relative abundance data with Bray Curtis dissimilarity (**A**) and Jaccard (**B**), and using absolute abundance data with Bray Curtis dissimilarity (**C**) and Jaccard (**D**).

Version information

Table 1: Software and package versions

	version
OS	Ubuntu 16.04.6 LTS
R	3.6.1
colorspace	1.4-1
ggsignif	0.6.0
ellipsis	0.3.0
rprojroot	1.3-2

	version
XVector	0.26.0
fs	1.3.1
rstudioapi	0.10
farver	2.0.1
remotes	2.1.0
lubridate	1.7.4
xml2	1.2.2
codetools	0.2-16
splines	3.6.1
knitr	1.26
pkgload	1.0.2
zeallot	0.1.0
ade4	1.7-13
jsonlite	1.6
broom	0.5.2
cluster	2.1.0
dbplyr	1.4.2
compiler	3.6.1
httr	1.4.1
backports	1.1.5
assertthat	0.2.1
Matrix	1.2-17
lazyeval	0.2.2
cli	1.1.0
htmltools	0.4.0

	version
prettyunits	1.0.2
tools	3.6.1
igraph	1.2.4.2
gtable	0.3.0
glue	1.3.1
reshape2	1.4.3
Rcpp	1.0.3
Biobase	2.46.0
cellranger	1.1.0
vctrs	0.2.0
Biostrings	2.54.0
multtest	2.42.0
ape	5.3
nlme	3.1-140
iterators	1.0.12
xfun	0.11
ps	1.3.0
testthat	2.3.1
rvest	0.3.5
lifecycle	0.1.0
devtools	2.2.1
zlibbioc	1.32.0
MASS	7.3-51.4
hms	0.5.2
parallel	3.6.1

	version
biomformat	1.14.0
rhdf5	2.30.1
yaml	2.2.0
curl	4.3
memoise	1.1.0
stringi	1.4.3
highr	0.8
S4Vectors	0.24.1
desc	1.2.0
foreach	1.4.7
BiocGenerics	0.32.0
pkgbuild	1.0.6
rlang	0.4.2
pkgconfig	2.0.3
evaluate	0.14
Rhdf5lib	1.8.0
labeling	0.3
cowplot	1.0.0
tidyselect	0.2.5
processx	3.4.1
R6	2.4.1
IRanges	2.20.1
generics	0.0.2
DBI	1.0.0
pillar	1.4.2

	version
haven	2.2.0
withr	2.1.2
mgcv	1.8-28
survival	2.44-1.1
modelr	0.1.5
crayon	1.3.4
rmarkdown	1.18
usethis	1.5.1
readxl	1.3.1
callr	3.3.2
reprex	0.3.0
digest	0.6.23
webshot	0.5.2
stats4	3.6.1
munsell	0.5.0
viridisLite	0.3.0
sessioninfo	1.1.1