Statistical Analysis 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

Introduction to the biostatistical analysis

Objective

The project objective was to identify differences in the bacterial microbiome between plowed and unplowed soil samples. To meet this goal, 24 samples were collected: 12 from clay-type soil and 12 from sand-type soil. For each soil type, 6 plowed and 6 unplowed samples were collected.

Analysis

In this report, biostatistical analyses are performed and the results presented, building on the data generated and evaluated in the 2 prior reports (Report 1: Quality filtering and data processing, Report 2: Description and evaluation of the microbiome profiles).
Through biostatistical analysis, we compared the microbiome profiles between plowed and unplowed samples, within the respective soil types. We refer to the status of plowed and unplowed as the testing condition. The analyses therefore compare subgroups comprising 6 samples and the statistical analysis approaches were selected to respect the study design with small sample numbers and highly diverse communities.

A number of analyses were performed, as shortly introduced here:

1. The report initiates with a visual evaluation of the bacterial community composition and a statistical analysis to evaluate if differences are statistically significant (using Permutational Multivariate Analysis of Variance (ADONIS)).

2. This is followed by an analysis of variance, in which we evaluate if the homogeneity of the bacterial communities within the subgroups differ. If the testing condition disrupts the bacterial community in a non-specific manner and thereby makes samples more or less alike, this will be reflected by a difference in the distance between the samples within each subgroup.
3. We evaluated if there is a difference between the absolute bacterial abundance between the testing conditions. This provided information on the effect of the testing condition on the total bacterial load.
4. Alpha diversity is a measure of the within-sample diversity and we evaluated whether the alpha-diversity differed between testing conditions using three different measures of alpha-diversity.
5. Finally, we compared the abundance of single taxa between the testing conditions. This allows us to identify indicator organisms that may be of interest for further testing. We compared the abundance of organisms classified at 5 taxonomic levels; phyla, class, order, family and genera.

Spike-In of known bacteria

The use of next generation sequencing provides information of the relative abundance of bacteria within a sample and allow us to evaluate if the testing condition affects the composition of the community. However, it does not provide information on the bacterial load (the absolute abundance) of bacteria in a sample. To evaluate if the testing condition affected the absolute abundance of bacteria, spike-in of two bacterial taxa was included in the data generation. The addition of bacteria of known quantity allow for the calculation of the absolute bacterial load in each sample and thereby allow us to evaluate if the absolute abundance of the total bacterial load and of specific bacteria differ between the testing conditions. The two used bacteria are not found in the studied communities and will therefore not interfere with the analysis of the existing soil communities.

Differences in community composition (beta-diversity)

Visualization by ordination

As descripted in Report 2, beta-diversity is a measure of how similar or dissimilar the bacterial community is between each pair of samples. The measures are useful for statistical analysis and visualization of the overall microbiome community.

Here we evaluate the microbiome communities using the Bray-Curtis and Jaccard beta-diversity measures, where each sample is a point and the distance between the points increases with increasing dissimilarity in the microbiome community. In the two plots of Figure 1, all 24 samples are included and colored by soil type and by soil status. We see a clear separation of the clay and sand samples for both types, plowed and unplowed, highlighting that the bacterial composition between the two soil types and statuses are highly distinct.

In Report 2, we evaluated the summary statistics of the beta-diversity measures and saw that the microbiome profiles in the soil samples at the level of sequences are highly diverse with close to maximum possible distance between all samples (mean close to 1). Therefore, we initiate the evaluation of microbiome communities at the higher taxonomic level of genera, where the mean beta diversity decreases, indicating a possible structure in the dataset.

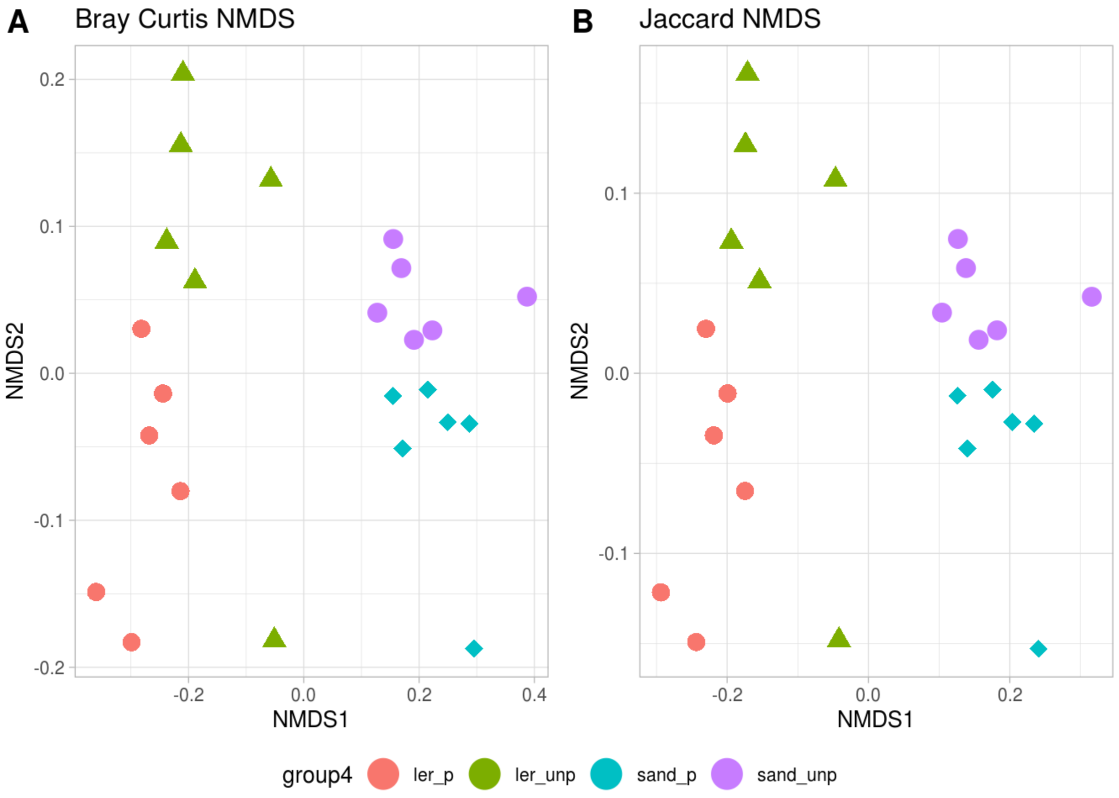

Figure 1: Visualization of the bacterial community in the 24 soil samples. Ordination plots using **A)** Bray distance and **B)** Jaccard disance and relative abundance. The plots were made using NMDS and beta-diversity at the genera level. Samples are coloured by soil type (clay or sand) and by status (plowed or unplowed).

Permutational Multivariate Analysis of Variance

To evaluate if the observed differences between the microbiome composition of plowed and unplowed samples is statistically significant, we perform an analysis named Permutational Multivariate Analysis of Variance (ADONIS). ADONIS partitions sums of squares of a multivariate data set, and is analogous to MANOVA (multivariate analysis of variance) using beta-diversity measures. It partitions distance matrices among sources of variation and fits linear models to the distance matrices using a permutation test with pseudo-F ratios and can therefore be considered as a “permutational manova”.

For the analysis we use Bray-Curtis and Jaccard beta-diversity measures and perform the analysis for both absolute and relative abundances, at the genera and class level.

Adonis for relative abundance data

Comparison of plowed and unplowed clay samples showed a significant difference using both Bray-Curtis (variance explained R2=18%, p-value=0.005) and Jaccard (variance explained R2=16%, p-value=0.002). Also for sand, the comparisons were significant with a notable amount of variation in the bacterial community explained by the testing condition (Bray-Curtis R2=14%, p-value=0.023; Jaccard R2=14%, p-value=0.018). The patterns were also evident at the higher taxonomic level of class (see details in summary statistics below). This is a noteworthy difference considering the small number of samples and highly diverse communities that are in all soil samples.

		Bray		Jaccard	
Taxa level	Soil type	R2	p	R2	p
Genera	Clay	0.1917	0.003	0.1666	0.005
	Sand	0.1584	0.023	0.1436	0.035

Class	Clay	0.1882	0.057	0.1749	0.06
	Sand	0.1909	0.02	0.1856	0.015

Adonis for absolute abundance data

In contrary to the relative abundance communities, comparison of plowed and unplowed clay samples with absolute abundance data showed no a significant difference using either Bray-Curtis (R2=0.09%, p-value=0.32) or Jaccard (R2=9.5%, p-value=0.41). However, the variance explained reached 9% which indicate that there is a difference that would be significant if the power was increased. Again, the inherent complexity of soil limits the statistical power to detect robust differences. A similar pattern is observed for sand and at the taxonomic level of class.

		Bray		Jaccard	
Taxa level	Soil type	R2	p	R2	p
Genera	Clay	0.0938	0.343	0.0937	0.373
	Sand	0.0901	0.437	0.0897	0.512
Class	Clay	0.0384	0.685	0.0466	0.752
	Sand	0.0394	0.801	0.0606	0.673

Multivariate homogeneity of group dispersions (variances)

The multivariate analysis of variance found in this section, is used to evaluate if there is a difference in the homogeneity of the bacterial communities between the plowed and unplowed samples. If plowing disrupts the bacterial community in a non-specific manner (e.g. a random disturbance more than a specific directional shift), this will be reflected by a difference in the mean distance of the samples within each subgroup.

For each subgroup of each sample, we calculated the average distance of group members to the group centroid in the multivariate space of the bacterial community. To test if the dispersions (variances) of the groups are different, the distances of group members to the group centroid are subject to an analysis of variance (ANOVA). The approach is a multivariate analogue of Levene’s test for homogeneity of variances if the distances between group members and group centroids is the Euclidean distance. However, as Euclidean distance is not an appropriate measure for microbiome data, we use Bray-Curtis dissimilarities.

The analysis for sand and clay using the relative abundance data showed no significant differences in variance between groups (p-value > 0.05). The analysis again supports a clear shift in the plowed and unplowed bacterial communities (see PCoA plots), however no support for a difference in variance is shown (see boxplots of the distance to centroid). As for relative abundance data the analysis for sand and clay using the absolute abundance data, showed no significant differences in variance between groups (p-value > 0.05).

The detailed results of the statistical analysis and illustrations of the results (PCoA and boxplots), can be seen below.

Clay, with relative abundance

```
## Analysis of Variance Table
##
## Response: Distances
##           Df    Sum Sq    Mean Sq F value Pr(>F)
## Groups     1 0.0002638 0.00026379  0.2014 0.6631
## Residuals 10 0.0130962 0.00130962
```

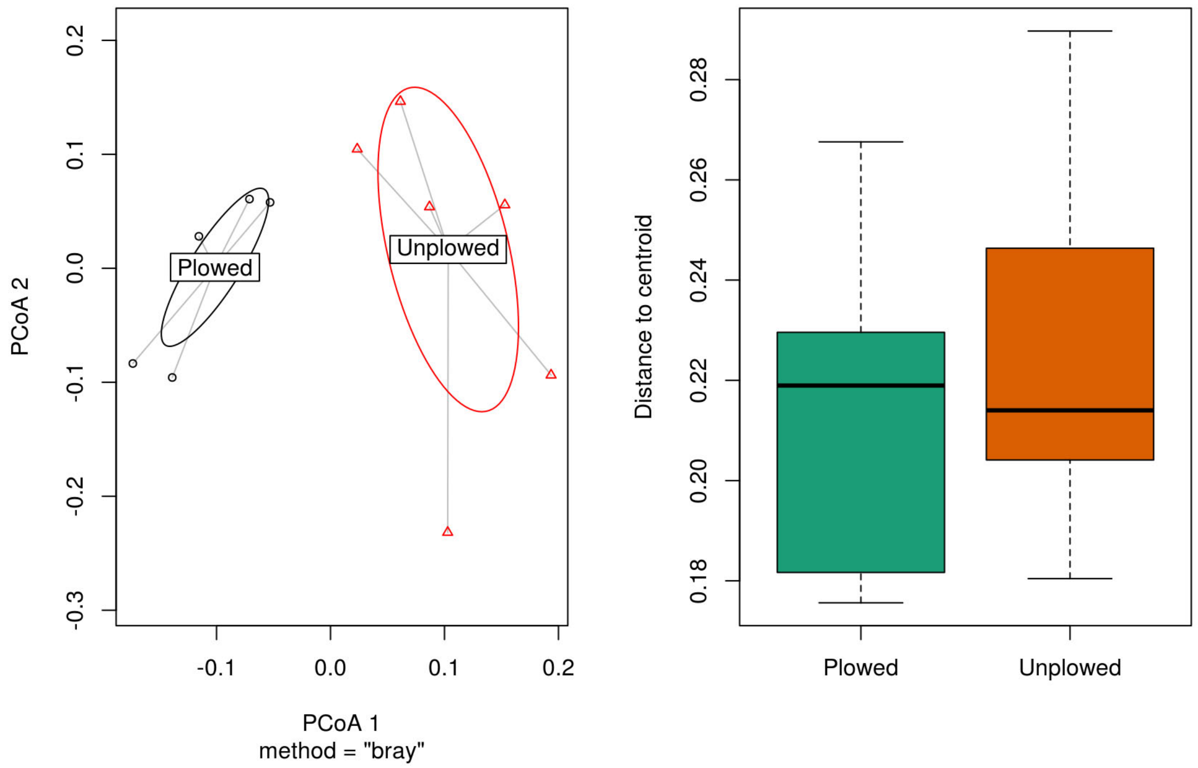

Figure 2: Visualization of the analysis of dispersion for clay samples. The left panel show a PCoA plot over plowed and unplowed clay samples based on relative abundance data, and the right panel show the distance to centroid for these categories.

Sand, with relative abundance

```
## Analysis of Variance Table
##
## Response: Distances
##           Df    Sum Sq   Mean Sq F value Pr(>F)
## Groups     1 0.0000849 0.0000849  0.0311 0.8636
## Residuals 10 0.0273155 0.0027316
```

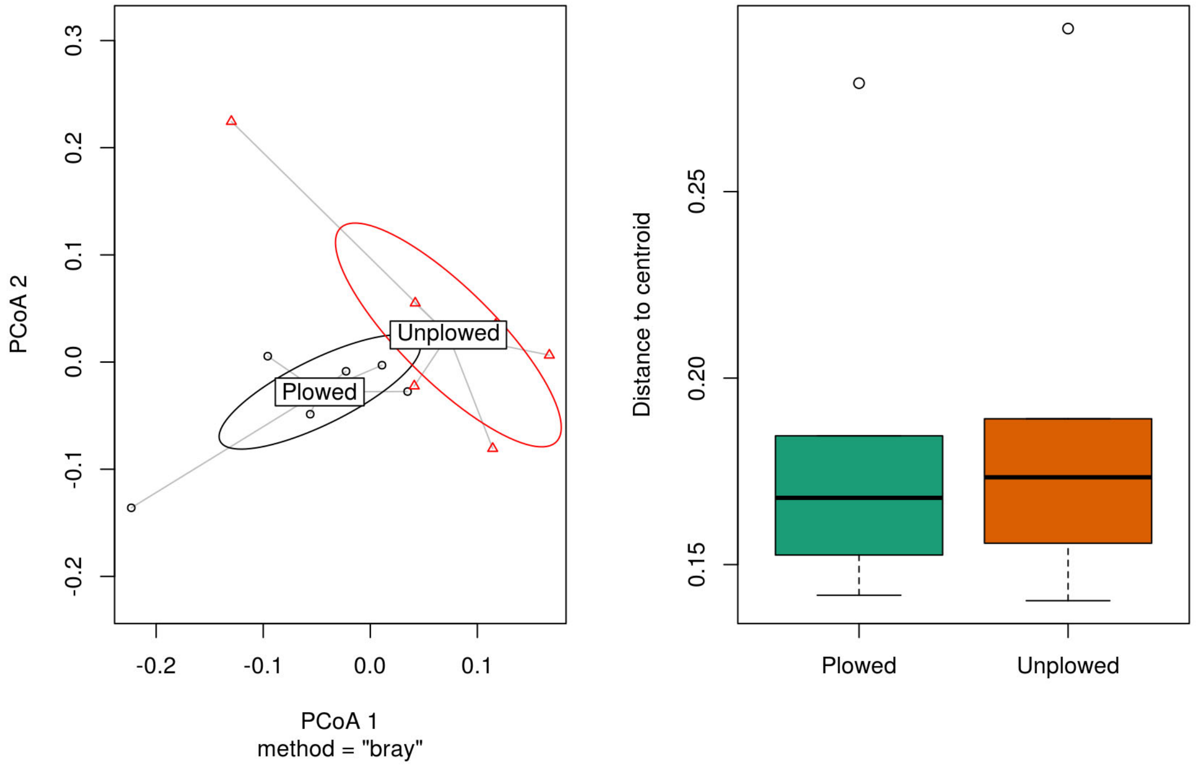

Figure 3: Visualization of the analysis of dispersion for sand samples. The left panel show a PCoA plot over plowed and unplowed sand samples based on relative abundance data and the right panel show the distance to centroid for these categories.

Clay, with absolute abundance

```
## Analysis of Variance Table
##
## Response: Distances
##           Df   Sum Sq   Mean Sq F value Pr(>F)
## Groups     1 0.001817 0.0018170  0.2098 0.6567
## Residuals 10 0.086588 0.0086588
```

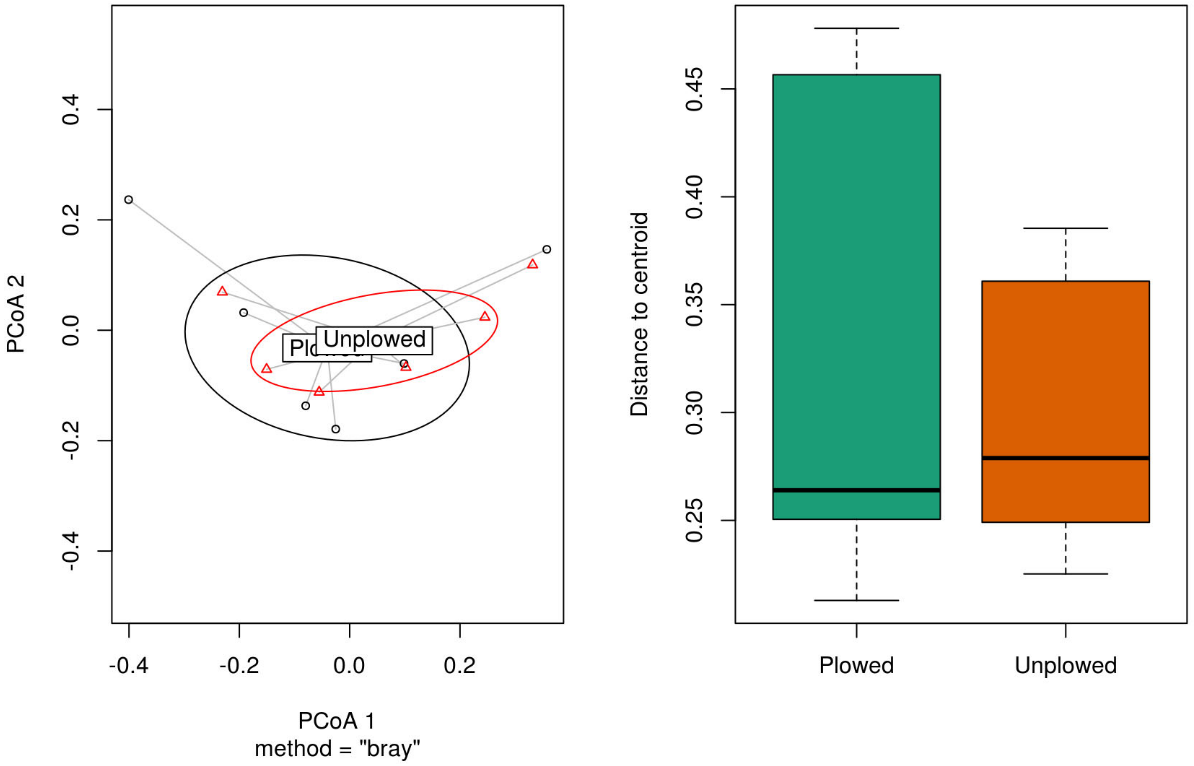

Figure 4: Visualization of the analysis of dispersion for clay samples. The left panel show a PCoA plot over plowed and unplowed clay samples based on absolute abundance data and the right panel show the distance to centroid for these categories.

Sand, with absolute abundance

```
## Analysis of Variance Table
##
## Response: Distances
##           Df   Sum Sq   Mean Sq F value Pr(>F)
## Groups     1 0.005538 0.0055378  0.8955 0.3663
## Residuals 10 0.061837 0.0061837
```

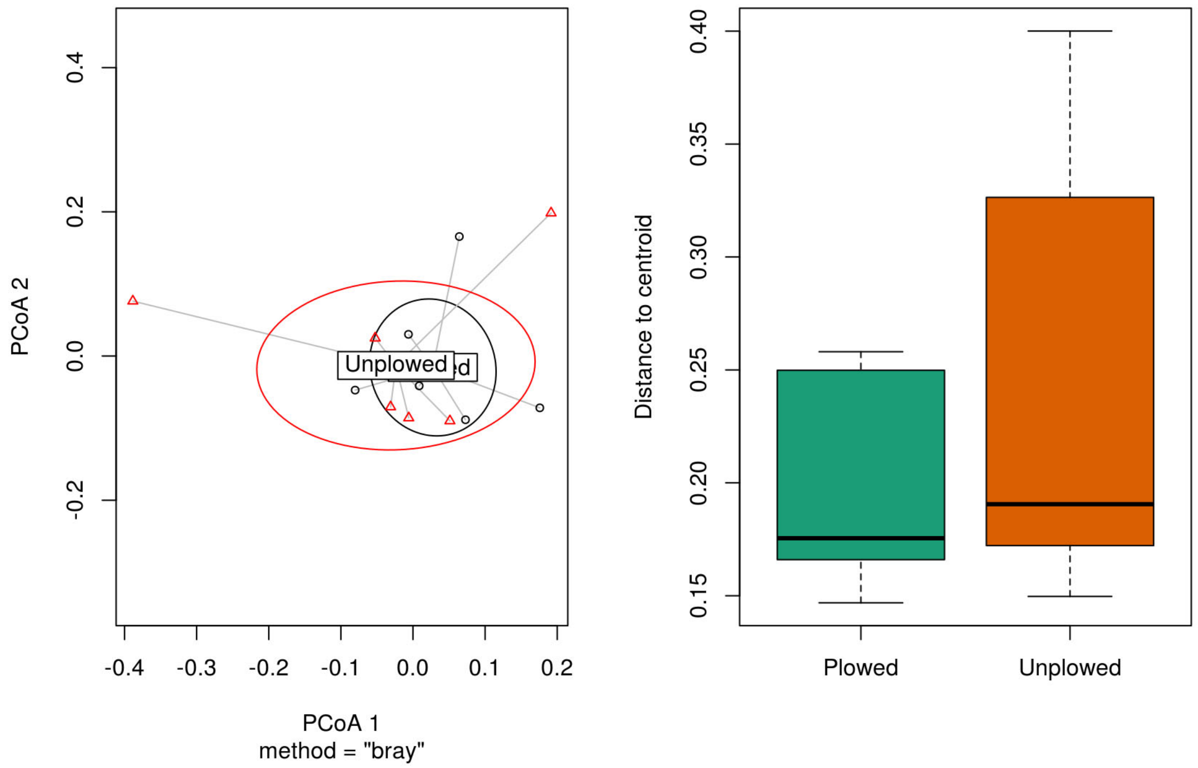

Figure 5: Visualization of the analysis of dispersion for sand samples. The left panel show a PCoA plot over plowed and unplowed sand samples based on absolute abundance data and the right panel show the distance to centroid for these categories.

Differences in absolute bacterial load

Information on the absolute bacterial abundance in a sample, calculated as the sum of detected bacteria in the sample, allow us to evaluate if plowing of soil affects the amount (or concentration) of bacteria in soil. We used a Wilcoxon rank sum test to evaluate the direction and potential significance of a change, and visualized the data using two boxplots (one for clay and one for sand). The statistical analysis did not support a difference in the absolute amount of bacteria between the plowed and unplowed samples, for either clay or sand (p-value > 0.05). This finding was supported by the visual inspection as can be seen in the below boxplots.

```
## CLAY
```

```
##
##  Wilcoxon rank sum test
##
## data:  asv.meta_clay$Sample_sum by asv.meta_clay$group4
## W = 19, p-value = 0.9372
## alternative hypothesis: true location shift is not equal to 0
```

```
## SAND
```

```
##
##  Wilcoxon rank sum test
##
## data:  asv.meta_sand$Sample_sum by asv.meta_sand$group4
## W = 14, p-value = 0.5887
## alternative hypothesis: true location shift is not equal to 0
```

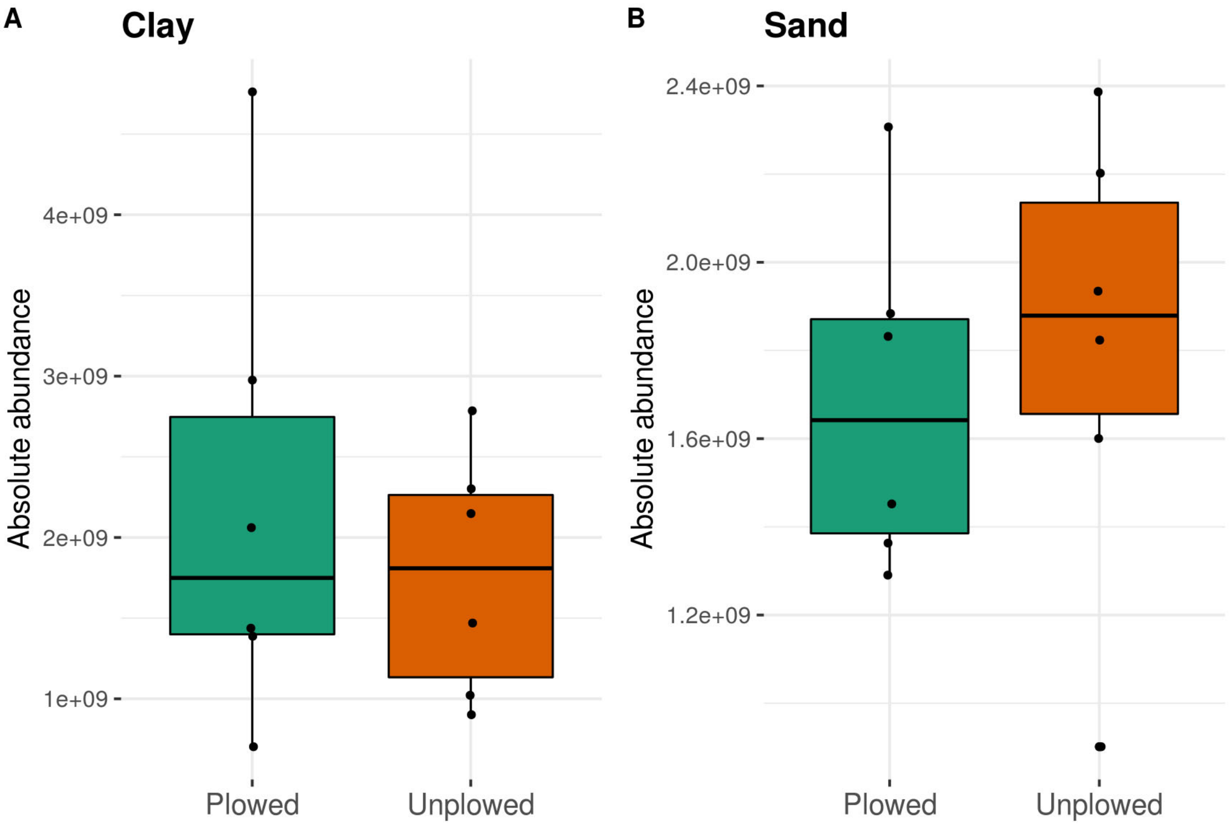

Figure 6: The total absolute abundance for sand and clay samples. The sum of the absolute abundances across plowed and unplowed samples of clay **(A)** and sand **(B)**.

Differences in alpha-diversity

As described in **Report 2**, alpha diversity is a measure of the diversity within (or complexity) within one microbiome community. We used a Wilcoxon rank sum test to evaluate the direction and potential significance of a change, and visualized the data using two boxplots (one for clay and one for sand). The analysis was performed for three measures of alpha diversity: observed species, Chao1 and Shannon. See **Report 2** for introduction to the different tupes of alpha diversity. The statistical analysis did not find a significant difference (p-value > 0.05), however the visual inspection indicated a consistent higher diversity in the unplowed samples, and the statistical analysis indicated that the difference was biggest for the sand samples. The lacking ability to detect a significant difference is probably due to the low power and high microbiome complexity of soil samples.

```
## Clay - Observed taxa
```

```
## 
##  Wilcoxon rank sum test
## 
## data:  ps.sub.relab_clay$Observed by ps.sub.relab_clay$group4
## W = 15, p-value = 0.6991
## alternative hypothesis: true location shift is not equal to 0
```

```
## Clay - Chao1
```

```
##
##  Wilcoxon rank sum test
##
## data:  ps.sub.relab_clay$Chao1 by ps.sub.relab_clay$group4
## W = 15, p-value = 0.6991
## alternative hypothesis: true location shift is not equal to 0
```

```
## Clay - Shannon
```

```
##
##  Wilcoxon rank sum test
##
## data:  ps.sub.relab_clay$Shannon by ps.sub.relab_clay$group4
## W = 15, p-value = 0.6991
## alternative hypothesis: true location shift is not equal to 0
```

```
## Sand - Observed taxa
```

```
##
##  Wilcoxon rank sum test
##
## data:  ps.sub.relab_sand$Observed by ps.sub.relab_sand$group4
## W = 9, p-value = 0.1797
## alternative hypothesis: true location shift is not equal to 0
```

```
## Sand - Chao1
```

```
##
##  Wilcoxon rank sum test
##
## data:  ps.sub.relab_sand$Chao1 by ps.sub.relab_sand$group4
## W = 9, p-value = 0.1797
## alternative hypothesis: true location shift is not equal to 0
```

```
## Sand - Shannon
```

```
##
##  Wilcoxon rank sum test
##
## data:  ps.sub.relab_sand$Shannon by ps.sub.relab_sand$group4
## W = 9, p-value = 0.1797
## alternative hypothesis: true location shift is not equal to 0
```

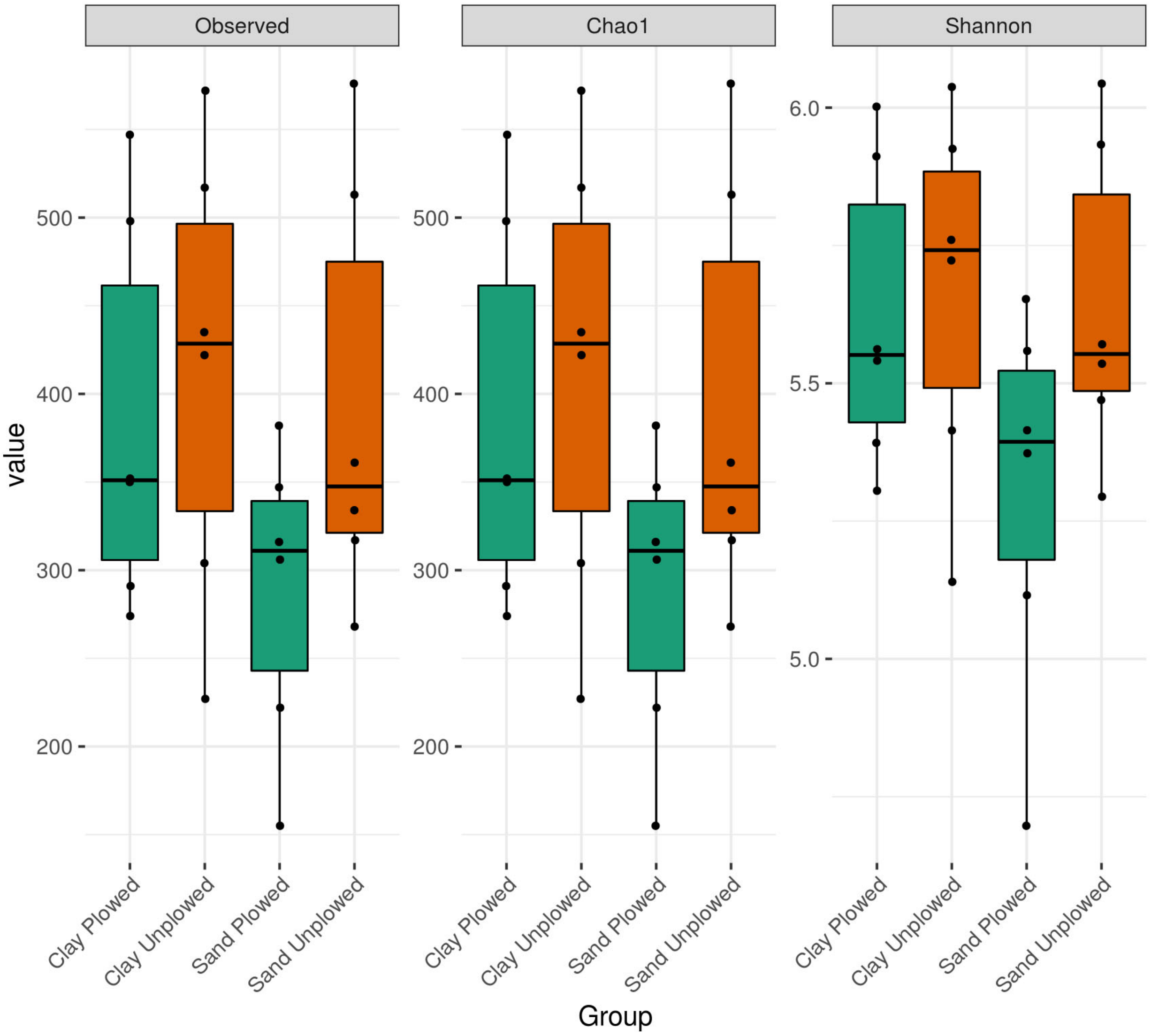

Figure 7: Alpha diversity metrics for normalised sampels. Values of alpha-diversity for all combinations of plowed, unplowed, clay and sand samples are illustrated for observed species, Chao1 and Shannon.

Differential abundance of single taxa

We now move from the evaluation of overall microbiome diversity and composition, to evaluate if there is a difference in the abundance of specific taxa between plowed and unplowed samples. This analysis provides the first insight into potential indicator organisms, and a first peak at the functional differences imposed by reducing plowing-activities. We use Wilcoxon rank sum test to evaluate the direction and potential significance of a change in each taxon. Below are the results for genera and phyla, while supplementary documents provide the results of the analysis of all taxonomic levels.

Boxplots are used to visualize the top 6 associated taxa (selected as taxa with lowest p-value in the Wilcoxon test). The boxplots show a clear difference in the abundance of a number single organism despite the low number of samples analyzed in this pilot project. The Wilcoxon rank sum test is used in the analysis as it is robust despite the small number of samples, non-normal distribution and inflation of zero values. However, the model is less sensitive than a parametric model and therefore provide a conservative estimate of significance. Despite these precautions, a number of organisms differ in abundance between the plowed and unplowed samples, including a significant decrease of Gemmatimonadetes in the unplowed samples in both sand and clay, as well as an increase of Latescibacteria that however is only significant in sand samples. Of interest Latescibacteria has previous been found increased in soil subject to reduce tillage (https://doi.org/10.3389/fmicb.2017.01127 (https://doi.org/10.3389/fmicb.2017.01127)). It is further worth noting that many top associated bacteria correspond between soil and sand, despite the clear differences in the overall microbiome composition observed in the **Figure 1**. This indicate that it can be possible to identify indicator species for tillage-associated soil conditions across soil types and geography.

Relative abundance

Phyla - Clay

Table 1

Table of top indicator phyla for clay

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV3664	Armatimonadetes	0.275	-0.267	5.00e-03	0.125	0.640
ASV723	Gemmatimonadetes	3.272	-1.052	2.60e-02	0.295	1.738
ASV5192	BRC1	0.018	0.017	7.40e-02	-0.060	0.000
ASV50	Nitrospirae	1.600	0.342	9.31e-02	-0.920	0.094
ASV1878	Planctomycetes	0.220	-0.130	1.32e-01	-0.053	0.315
ASV23	Firmicutes	2.133	0.569	1.32e-01	-1.274	0.440
ASV2404	Elusimicrobia	0.012	0.000	1.76e-01	0.000	0.029
ASV6395	Spirochaetes	0.003	0.000	1.76e-01	-0.015	0.000
ASV3069	Fibrobacteres	0.021	-0.008	1.82e-01	0.000	0.065
ASV38	Actinobacteria	11.406	2.894	2.40e-01	-6.061	0.540

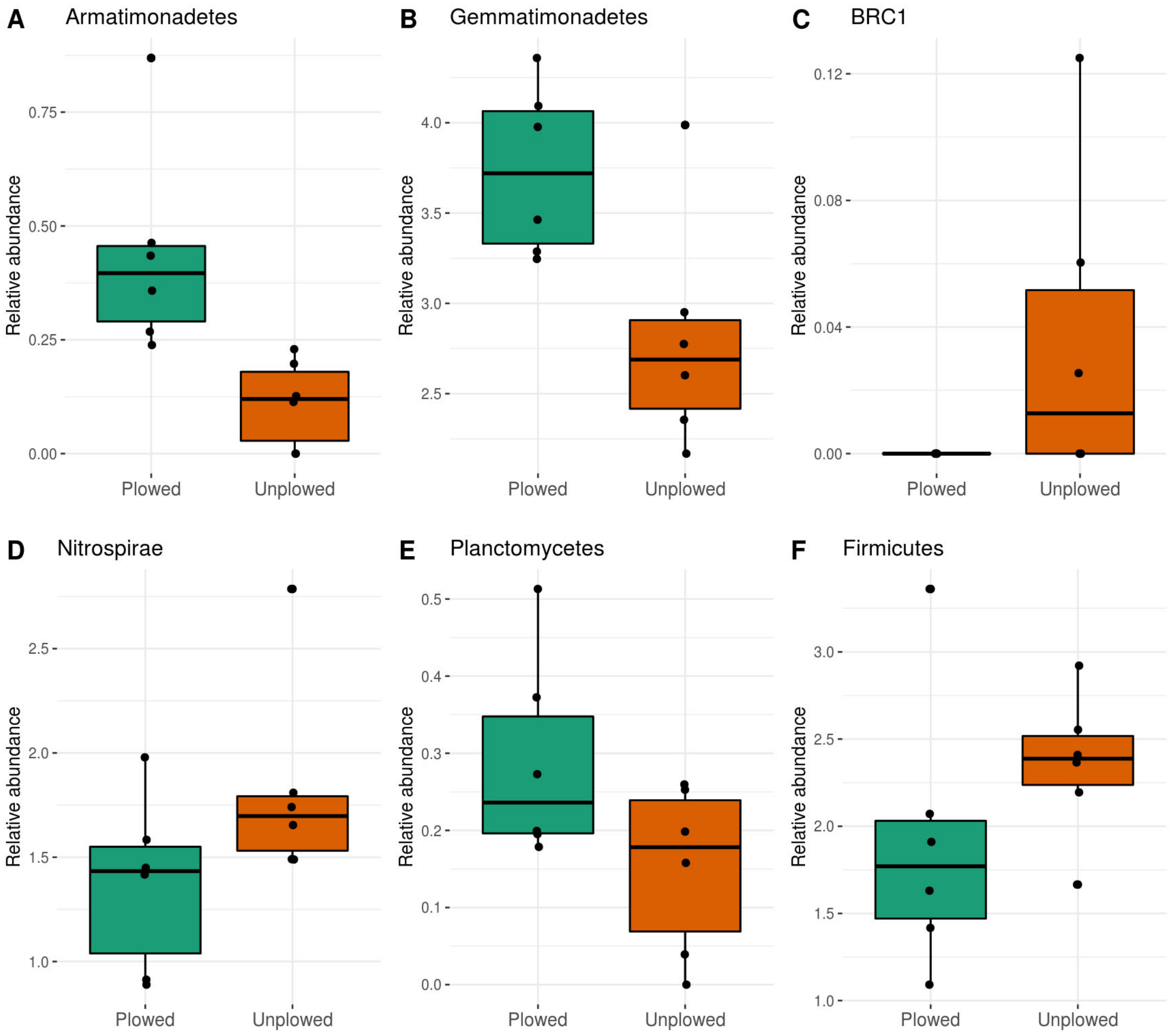

Figure 8:The top phyla with deviating relative abundance between plowed and Unplowed clay samples. The six phyla were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Phyla - Sand

Table 2

Table of top indicator phyla for sand

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV308	Latescibacteria	1.415	0.826	2.16e-03	-1.220	-0.358
ASV723	Gemmatimonadetes	3.555	-0.783	8.66e-03	0.223	1.497
ASV45	Proteobacteria	34.611	-3.491	4.11e-02	0.098	6.787
ASV25	Acidobacteria	28.830	4.386	6.49e-02	-5.736	0.030
ASV2404	Elusimicrobia	0.083	0.114	1.46e-01	-0.252	0.000
ASV5789	candidate_division_WPS-2	0.007	0.000	1.76e-01	-0.032	0.000
ASV6395	Spirochaetes	0.007	0.000	1.76e-01	-0.041	0.000

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV1878	Planctomycetes	0.050	0.030	2.95e-01	-0.101	0.031
ASV38	Actinobacteria	12.670	-2.292	3.10e-01	-4.185	3.786
ASV6814	Chlamydiae	0.006	0.000	4.05e-01	0.000	0.000

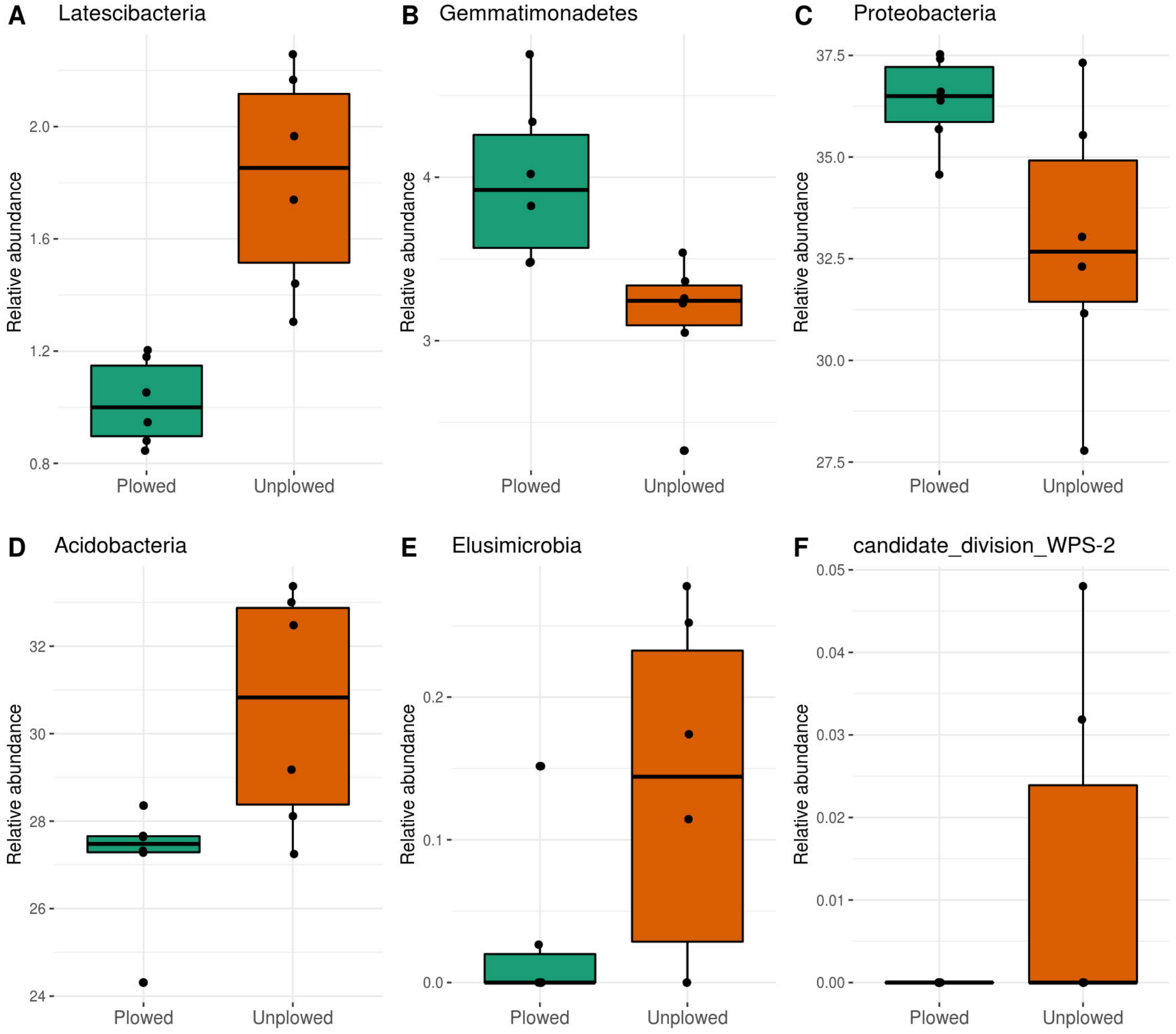

Figure 9: The top phyla with deviating relative abundance between plowed and Unplowed sand samples. The six phyla were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Genera - Clay

Table 3

Table of top indicator genera for clay

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV205	Lysobacter	1.810	-1.207	2.16e-03	0.665	1.964
ASV130	Pedomicrobium	0.360	0.436	5.00e-03	-0.644	-0.310
ASV5929	Rubrivirga	0.031	-0.042	1.15e-02	0.016	0.076

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV871	Parasegetibacter	0.377	-0.173	1.52e-02	0.075	0.308
ASV285	Flavihumibacter	0.179	-0.200	1.67e-02	0.000	0.581
ASV723	Gemmatimonas	3.272	-1.052	2.60e-02	0.295	1.738
ASV1477	Hyphomicrobium	0.201	-0.292	2.78e-02	0.023	0.336
ASV1532	Pontibacter	0.117	-0.160	3.41e-02	0.000	0.219
ASV2255	Phaselicystis	0.158	0.290	6.01e-02	-0.386	0.000
ASV106	Sphingomonas	1.433	-0.575	6.49e-02	-0.006	1.169
ASV1005	Devosia	0.062	-0.112	7.40e-02	0.000	0.272
ASV1171	Herminiimonas	0.106	-0.100	7.40e-02	0.000	0.326
ASV1230	Geobacter	0.026	-0.056	7.40e-02	0.000	0.119
ASV3037	Ensifer	0.075	0.137	7.40e-02	-0.302	0.000
ASV4166	Aquabacterium	0.037	-0.033	7.40e-02	0.000	0.136

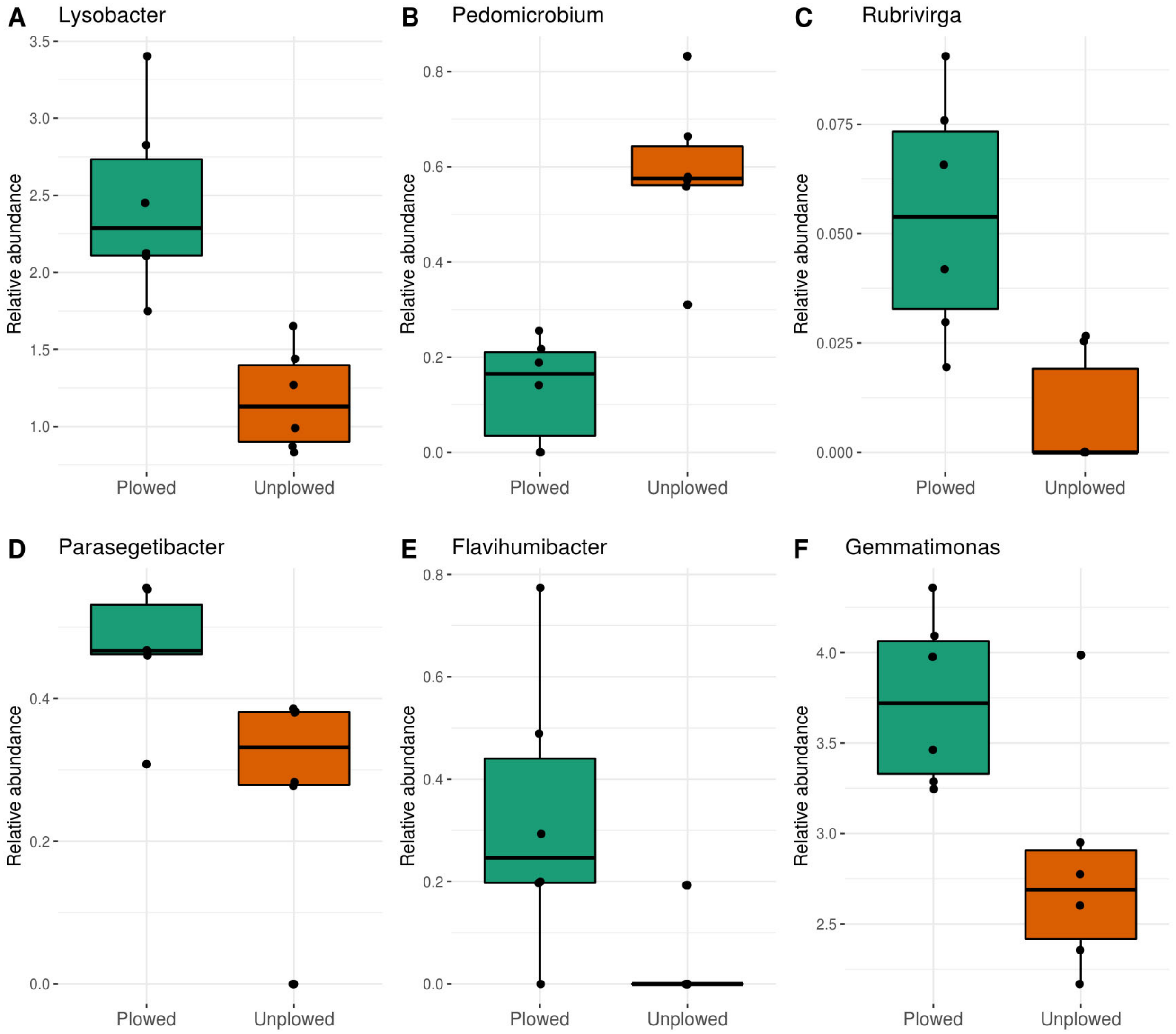

Figure 10: The top genera with deviating relative abundance between plowed and Unplowed clay samples. The six genera were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Genera - Sand Table 4 Table over top indicator genera for Sand

Table of top indicator genera for sand

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV45	Bradyrhizobium	0.984	-0.458	8.66e-03	0.086	0.838
ASV723	Gemmatimonas	3.555	-0.783	8.66e-03	0.223	1.497
ASV203	Aciditerrimonas	0.081	0.143	9.62e-03	-0.269	-0.125
ASV1965	Oryzihumus	0.099	0.206	2.84e-02	-0.429	0.000
ASV99	Schlegelella	1.705	-0.608	4.11e-02	0.029	1.469
ASV705	Pseudonocardia	0.130	0.137	6.01e-02	-0.354	0.000
ASV690	Microlunatus	0.016	0.035	7.40e-02	-0.075	0.000
ASV3827	Conexibacter	0.036	-0.033	7.40e-02	0.000	0.125

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV139	Nocardioides	0.975	-0.290	9.31e-02	-0.064	1.230
ASV2255	Phaselicystis	0.225	0.368	1.24e-01	-0.420	0.273
ASV106	Sphingomonas	0.654	-0.097	1.32e-01	-0.051	0.509
ASV469	Skermanella	0.232	0.068	1.42e-01	-0.378	0.097
ASV3273	Opitutus	0.077	-0.094	1.46e-01	0.000	0.211
ASV155	Chitinophaga	0.177	-0.099	1.58e-01	-0.191	0.409
ASV224	Portibacter	0.029	0.000	1.76e-01	-0.151	0.000

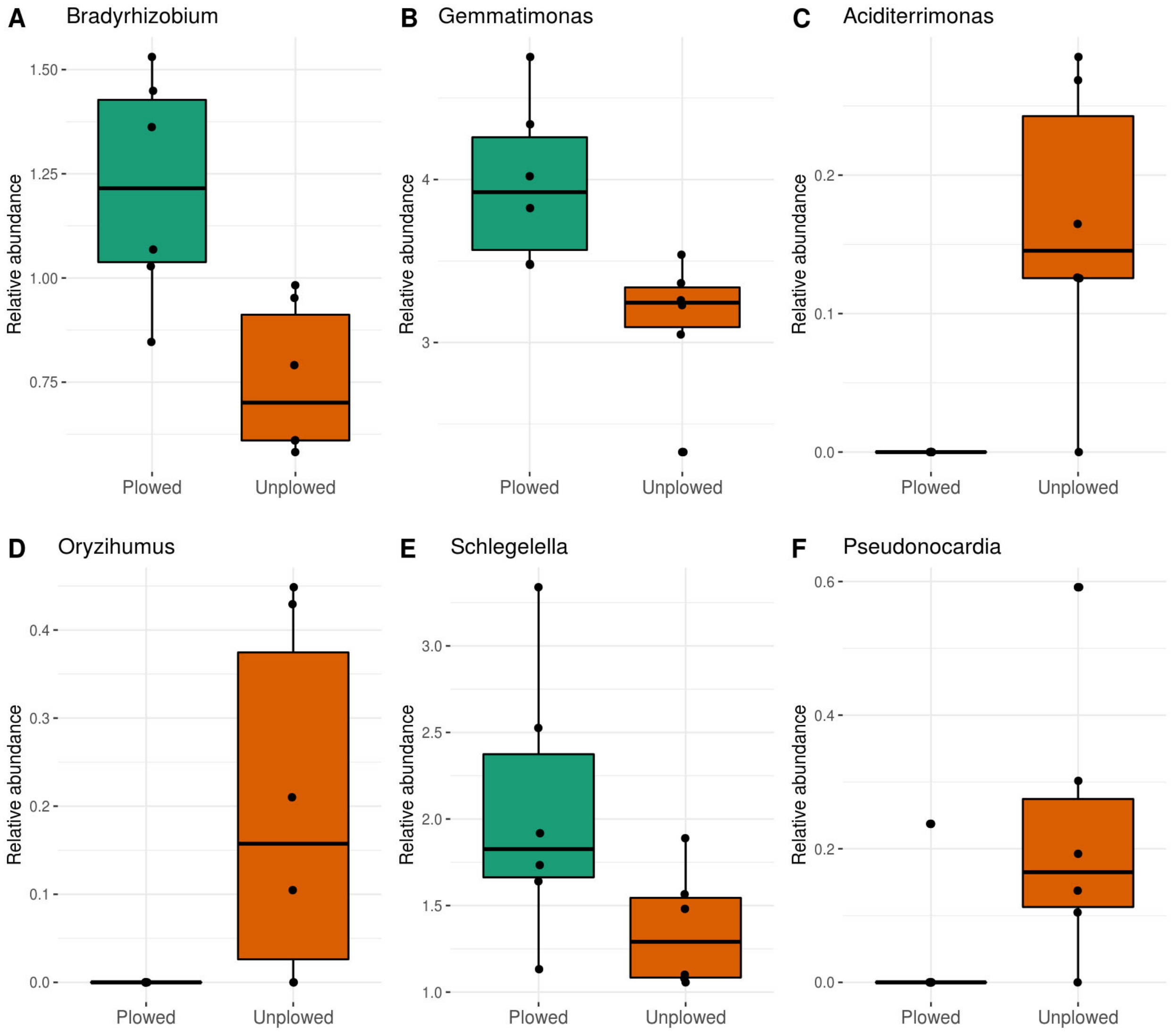

Figure 11: The top genera with deviating relative abundance between plowed and Unplowed sand samples. The six genera were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Absolute abundance

Phyla - Clay

Table 5

Table of top indicator phyla for clay

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV3479	Armatimonadetes	4.99e+08	-5.20e+08	5.00e-03	3.40e+08	1.43e+09
ASV5192	BRC1	2.01e+07	5.26e+07	7.40e-02	-5.85e+07	0.00e+00
ASV2730	Planctomycetes	4.38e+08	-2.12e+08	1.32e-01	-9.60e+07	1.23e+09
ASV6776	Spirochaetes	4.74e+06	5.24e-05	1.76e-01	-2.18e+07	0.00e+00
ASV3203	Elusimicrobia	1.77e+07	-5.29e-05	1.76e-01	0.00e+00	4.06e+07
ASV3069	Fibrobacteres	4.57e+07	-2.49e+07	1.82e-01	-9.66e-06	2.59e+08
ASV373	Gemmatimonadetes	6.69e+09	-2.77e+09	3.10e-01	-2.38e+09	9.43e+09
ASV217	Bacteroidetes	1.38e+10	-5.08e+09	3.10e-01	-3.42e+09	1.29e+10
ASV2640	Candidatus_Saccharibacteria	2.11e+08	-2.71e+07	3.10e-01	-1.58e+08	2.12e+08
ASV614	Chloroflexi	1.91e+09	5.19e+08	3.94e-01	-1.39e+09	3.85e+09

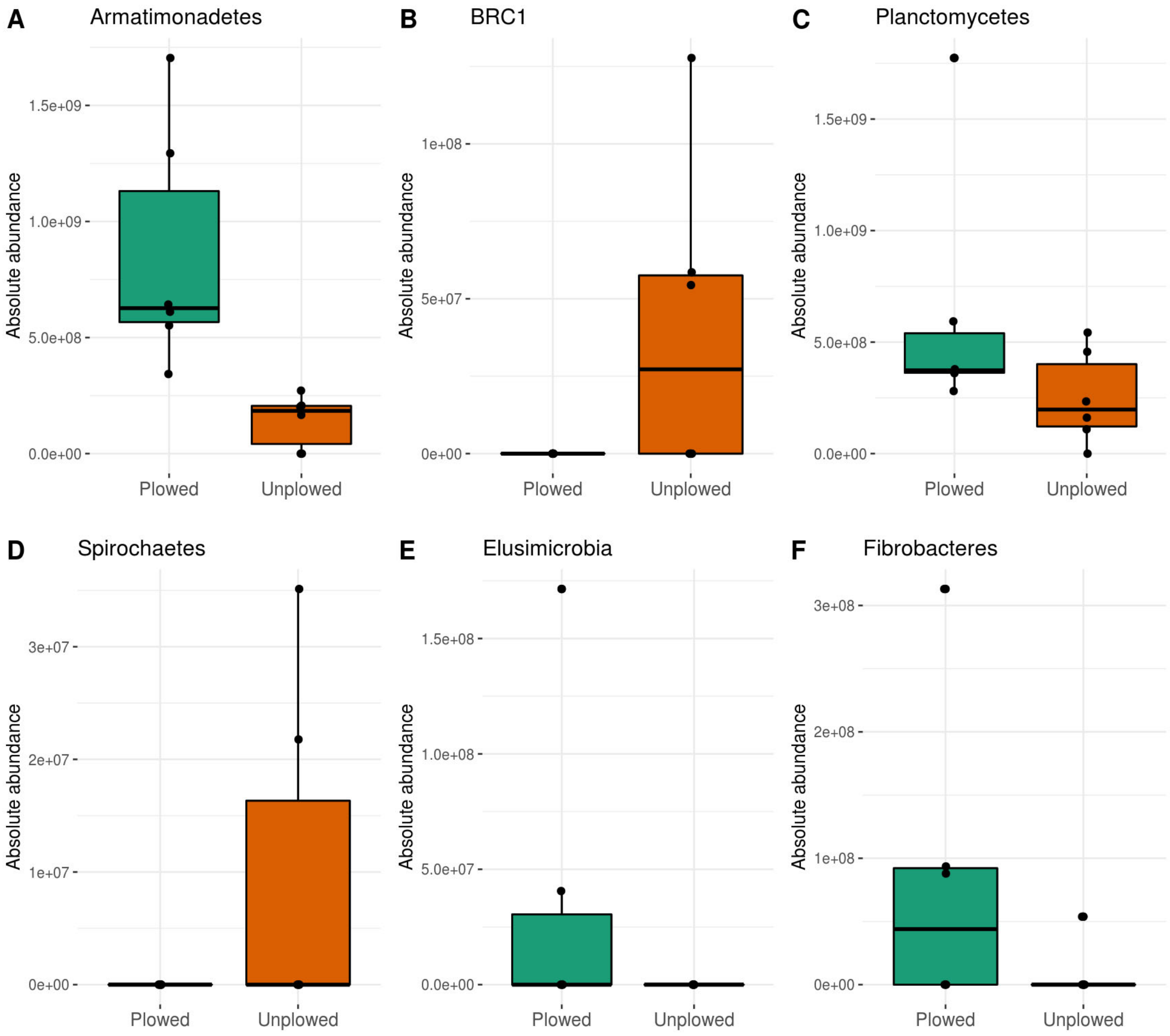

Figure 12: The top phyla with deviating absolute abundance between plowed and Unplowed clay samples. The six phyla were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Phyla - Sand

Table 6 Table over top indicator phyla for Sand

Table of top indicator phyla for sand

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV225	Latescibacteria	2.49e+09	1.34e+09	8.66e-03	-2.96e+09	-2.57e+08
ASV3203	Elusimicrobia	1.42e+08	2.24e+08	1.04e-01	-4.43e+08	4.91e-05
ASV6776	Spirochaetes	9.73e+06	5.88e-05	1.76e-01	-3.79e+07	0.00e+00
ASV5789	candidate_division_WPS-2	9.94e+06	5.67e-05	1.76e-01	-4.33e+07	0.00e+00

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV25	Acidobacteria	5.07e+10	1.16e+10	2.40e-01	-3.39e+10	1.13e+10
ASV614	Chloroflexi	1.06e+09	2.72e+08	3.10e-01	-8.56e+08	3.41e+08
ASV2730	Planctomycetes	9.81e+07	2.90e+07	3.75e-01	-1.90e+08	5.76e+07
ASV6814	Chlamydiae	7.15e+06	0.00e+00	4.05e-01	0.00e+00	5.93e-05
ASV7001	Euryarchaeota	1.81e+06	0.00e+00	4.05e-01	-6.01e-05	0.00e+00
ASV5192	BRC1	7.89e+06	0.00e+00	4.05e-01	-6.55e-05	0.00e+00

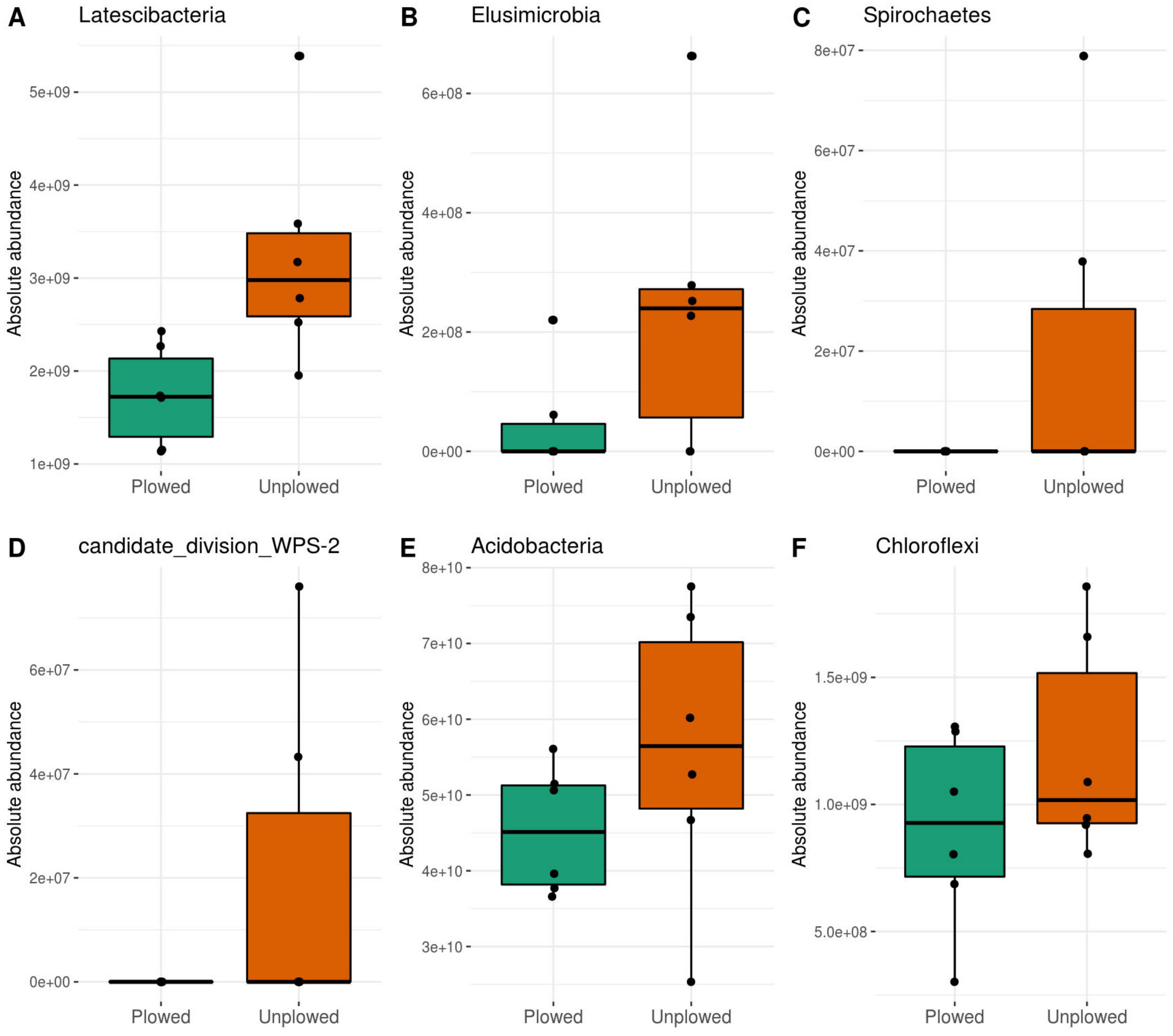

Figure 13: The top phyla with deviating absolute abundance between plowed and Unplowed sand samples. The six phyla were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Genera - Clay

Table 7

Table of top indicator genera for clay

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV285	Flavihumibacter	3.87e+08	-7.52e+08	1.67e-02	6.69e-05	1.46e+09
ASV5380	Rubrivirga	7.71e+07	-6.13e+07	2.78e-02	4.21e+06	2.70e+08
ASV1532	Pontibacter	2.21e+08	-2.86e+08	3.41e-02	2.23e-05	7.06e+08
ASV146	Lysobacter	3.82e+09	-2.39e+09	4.11e-02	1.92e+08	1.01e+10
ASV130	Pedomicrobium	6.75e+08	5.92e+08	4.50e-02	-1.23e+09	-5.96e+06
ASV1005	Devosia	1.07e+08	-1.59e+08	7.40e-02	0.00e+00	2.84e+08
ASV1230	Geobacter	7.73e+07	-1.66e+08	7.40e-02	0.00e+00	2.78e+08
ASV2841	Herminiimonas	2.03e+08	-4.00e+08	7.40e-02	0.00e+00	9.39e+08
ASV3037	Ensifer	1.23e+08	1.83e+08	7.40e-02	-2.76e+08	0.00e+00
ASV4166	Aquabacterium	4.46e+07	-7.05e+07	7.40e-02	0.00e+00	9.57e+07
ASV1192	Hyphomicrobium	4.60e+08	-4.83e+08	8.70e-02	-5.79e+07	1.39e+09
ASV3712	Pedobacter	2.68e+08	-2.03e+08	9.04e-02	-9.45e+07	4.09e+08
ASV1876	Phaselicystis	3.89e+08	3.48e+08	1.24e-01	-6.67e+08	3.82e+08
ASV6204	Tepidisphaera	1.01e+08	-7.79e+07	1.28e-01	-2.11e+07	2.04e+08
ASV1765	Thermomonas	9.00e+07	6.72e+07	1.29e-01	-3.04e+08	1.31e-05

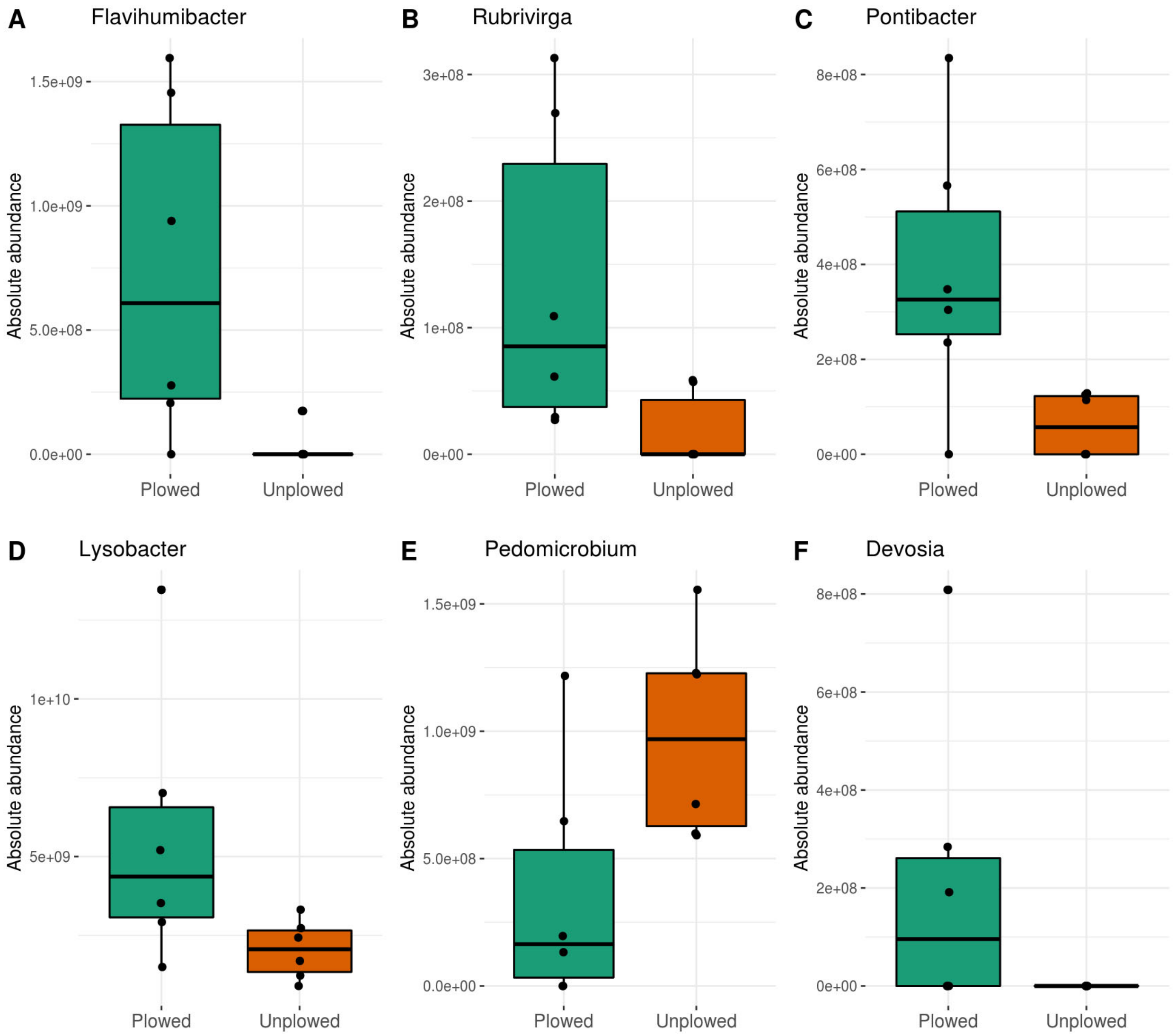

Figure 14: The top genera with deviating absolute abundance between plowed and Unplowed sand samples. The six genera were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Genera - Sand

Table 8

Table of top indicator genera for sand

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV203	Aciditerrimonas	1.50e+08	2.32e+08	9.62e-03	-5.52e+08	-1.14e+08
ASV2208	Oryzihumus	1.75e+08	2.25e+08	2.84e-02	-7.18e+08	0.00e+00
ASV705	Pseudonocardia	2.54e+08	2.50e+08	4.03e-02	-9.58e+08	-1.97e-05
ASV99	Schlegelella	2.87e+09	-6.73e+08	6.49e-02	-8.80e+07	1.83e+09
ASV1001	Microlunatus	3.44e+07	7.79e+07	7.40e-02	-1.37e+08	0.00e+00
ASV3827	Conexibacter	5.39e+07	-9.13e+07	7.40e-02	0.00e+00	1.82e+08
ASV1876	Phaselicystis	3.78e+08	6.71e+08	8.74e-02	-9.24e+08	7.36e-06

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV45	Bradyrhizobium	1.72e+09	-4.48e+08	9.31e-02	-3.71e+07	1.45e+09
ASV3711	Opitutus	1.33e+08	-1.37e+08	1.46e-01	-1.32e-05	4.59e+08
ASV4293	Phenylobacterium	2.71e+08	3.08e+08	1.58e-01	-5.34e+08	4.34e+07
ASV224	Portibacter	4.85e+07	5.74e-05	1.76e-01	-2.75e+08	0.00e+00
ASV366	Arenimonas	9.08e+07	6.23e-05	1.76e-01	-1.14e+08	0.00e+00
ASV521	Angustibacter	2.65e+08	9.40e-05	1.76e-01	-7.47e+08	0.00e+00
ASV1331	Luteolibacter	1.31e+07	7.29e-05	1.76e-01	-5.95e+07	0.00e+00
ASV2028	Caulobacter	9.94e+07	-6.47e-05	1.76e-01	0.00e+00	5.00e+08

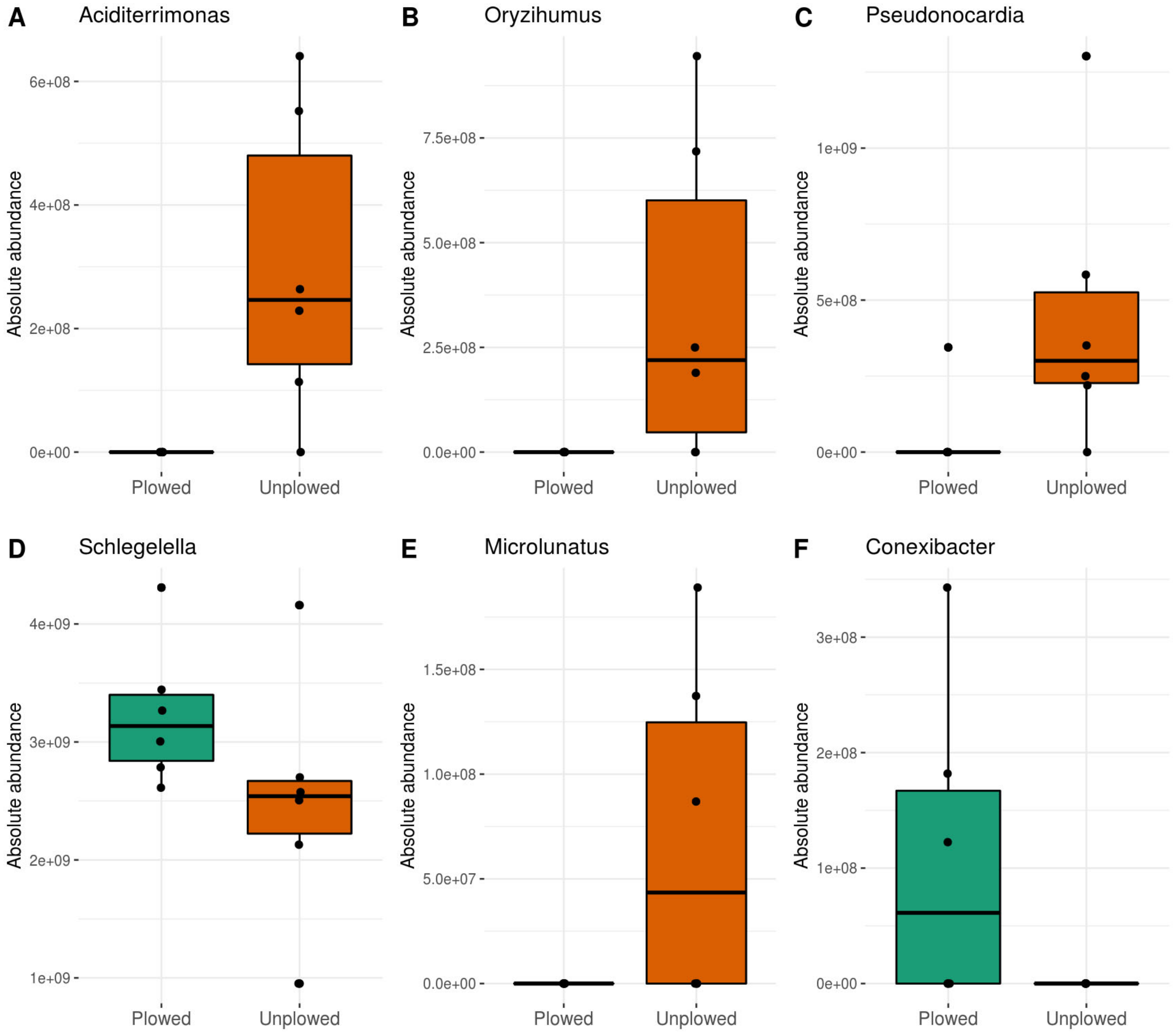

Figure 15: The top genera with deviating absolute abundance between plowed and Unplowed clay samples. The six genera were selected because they had the lowest p-value in the Wilcoxon rank sum test between plowed and Unplowed samples.

Version information

Table 9 Software and package versions

	version
OS	Ubuntu 16.04.6 LTS
R	3.6.1
readxl	1.3.1
backports	1.1.5
igraph	1.2.4.2
lazyeval	0.2.2
splines	3.6.1
usethis	1.5.1
digest	0.6.23
foreach	1.4.7
htmltools	0.4.0
memoise	1.1.0
cluster	2.1.0
remotes	2.1.0
recipes	0.1.7
Biostrings	2.54.0
modelr	0.1.5
gower	0.2.1
prettyunits	1.0.2
colorspace	1.4-1
rvest	0.3.5
rrcov	1.4-9
haven	2.2.0
xfun	0.11
callr	3.3.2
crayon	1.3.4
jsonlite	1.6
zeallot	0.1.0

	version
survival	2.44-1.1
iterators	1.0.12
ape	5.3
glue	1.3.1
gtable	0.3.0
ipred	0.9-9
zlibbioc	1.32.0
XVector	0.26.0
webshot	0.5.2
pkgbuild	1.0.6
Rhdf5lib	1.8.0
BiocGenerics	0.32.0
DEoptimR	1.0-8
mvtnorm	1.0-11
DBI	1.0.0
Rcpp	1.0.3
viridisLite	0.3.0
stats4	3.6.1
lava	1.6.6
prodlim	2019.11.13
httr	1.4.1
ellipsis	0.3.0
farver	2.0.1
pkgconfig	2.0.3
nnet	7.3-12
dbplyr	1.4.2
tidyselect	0.2.5
labeling	0.3
rlang	0.4.2
reshape2	1.4.3

	version
munsell	0.5.0
cellranger	1.1.0
tools	3.6.1
cli	1.1.0
generics	0.0.2
ade4	1.7-13
devtools	2.2.1
broom	0.5.2
evaluate	0.14
biomformat	1.14.0
yaml	2.2.0
ModelMetrics	1.2.2
processx	3.4.1
fs	1.3.1
robustbase	0.93-5
nlme	3.1-140
xml2	1.2.2
compiler	3.6.1
rstudioapi	0.10
curl	4.3
testthat	2.3.1
ggsignif	0.6.0
reprex	0.3.0
pcaPP	1.9-73
stringi	1.4.3
highr	0.8
ps	1.3.0
desc	1.2.0
Matrix	1.2-17
multtest	2.42.0

	version
vctrs	0.2.0
pillar	1.4.2
lifecycle	0.1.0
cowplot	1.0.0
R6	2.4.1
IRanges	2.20.1
sessioninfo	1.1.1
codetools	0.2-16
MASS	7.3-51.4
assertthat	0.2.1
pkgload	1.0.2
rhdf5	2.30.1
rprojroot	1.3-2
withr	2.1.2
S4Vectors	0.24.1
mgcv	1.8-28
parallel	3.6.1
hms	0.5.2
rpart	4.1-15
timeDate	3043.102
class	7.3-15
rmarkdown	1.18
Biobase	2.46.0
lubridate	1.7.4