

Statistical Analysis Report

BiomCare ApS

04/02/2020

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 fungi (ITS2)

Table of contents

- Table of contents
- Introduction to the biostatistical analysis
- Differences in community composition (beta-diversity)
 - Visualization by ordination
 - Permutational Multivariate Analysis of Variance
 - Multivariate homogeneity of group dispersions (variances)
- Differences in alpha-diversity
- Differential abundance of single taxa
 - Relative abundance
- Version information

Introduction to the biostatistical analysis

Objective

The project objective was to identify differences in the fungal 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: Microbiome profiling**). Through the biostatistical analysis we compare the microbiome profiles between plowed and unplowed samples for each soil type (sand and clay). We refer to the status of plowed/unplowed, as the testing condition. The different statistical analysis approaches were selected to respect the study design with small sample numbers and highly diverse communities with interventions expected to have notable effect on the microbiome.

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

1. The report initiates with a visual evaluation of the fungal 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 fungal communities within the subgroups differ. If the testing condition disrupts the fungal 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. Alpha-diversity is a measure of the within-sample diversity and we evaluate whether the alpha-diversity differ between testing conditions using four different measures of alpha-diversity.
4. Finally, we compare 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 compare the abundance of organisms classified at 5 taxonomic levels; phylum, class, order, family and genus.

Differences in community composition (beta-diversity)

Visualization by ordination

As described in **Report 2**, beta-diversity is a measure of how similar or dissimilar the fungal community is between each pair of samples. The measures are useful for statistical analysis and visualization of the overall fungal 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, highlighting that the fungal composition between the two soil types are highly distinct. Comparing the plowed and unplowed samples shows a strong separation along the second axis (NMDS2) for clay samples, however no clear separation for the sand samples.

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.

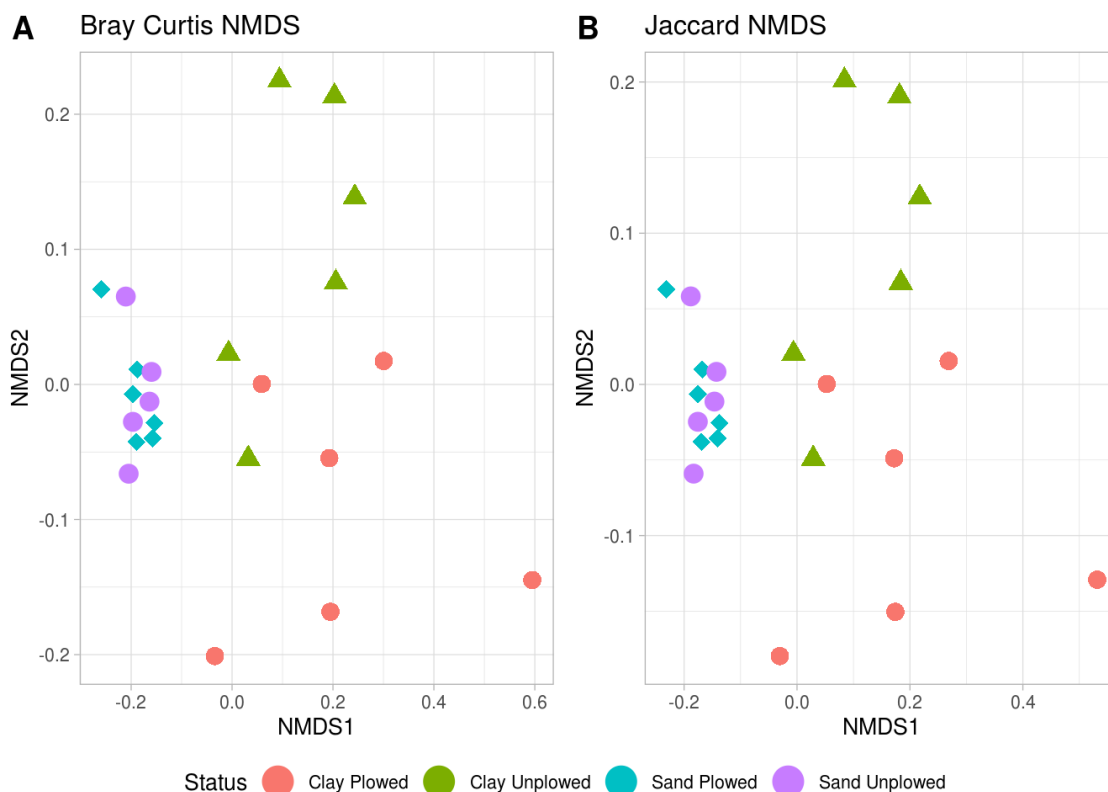


Figure 1: Visualization of the fungal community in the 24 soil samples. Ordination plots using **A)** Bray distance and **B)** Jaccard distance and relative abundance. The plots were made using Non-metric Multi-dimensional Scaling (NMDS) and beta-diversity at the genera level. Samples are coloured by 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 relative abundances, at the five taxonomic levels genus to phylum.

Adonis for relative abundance data

Comparison of plowed and unplowed clay samples showed a significant difference using both Bray-Curtis (variance explained $R^2=0.16$, p -value=0.033) and Jaccard (variance explained $R^2=0.13$, p -value=0.024) at genera level. For sand, the comparisons were not significant however showed a notable variation explained by the testing condition (Bray-Curtis $R^2=0.12$, p -value=0.204; Jaccard $R^2=0.12$, p -value=0.147). These results support the observations from the NMDS plot (Figure 1), but also indicate that there are trending differences in sand that may be evident at other dimensions than the first two axis highlighted in Figure 1. The differences observed are noteworthy considering the small number of samples and highly diverse communities that are in all soil samples. The patterns were also evident at the higher taxonomic level of family (see details in summary statistics below).

		Bray		Jaccard	
Taxa level	Soil type	R^2	p	R^2	p
Phylum	Clay	0.0694	0.489	0.067	0.623
	Sand	0.0858	0.52	0.0889	0.5
Class	Clay	0.1471	0.087	0.1199	0.126
	Sand	0.133	0.256	0.1281	0.258
Order	Clay	0.1404	0.089	0.1151	0.154
	Sand	0.0999	0.468	0.1024	0.429
Family	Clay	0.1425	0.07	0.1226	0.056
	Sand	0.0912	0.572	0.0974	0.486
Genus	Clay	0.1551	0.033	0.1338	0.024
	Sand	0.1225	0.204	0.1213	0.147

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 fungal communities between the plowed and unplowed samples. If plowing disrupts the fungal 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, we calculated the average distance of group members to the group centroid in the multivariate space of the fungal 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 show no significant differences in variance between

groups (p-value > 0.05).

- For clay samples, the analysis again supports a clear shift in the fungal communities between plowed and unplowed samples (see PCoA plots), however we see no support for a difference in variance from the boxplots (see boxplots of the distance to centroid).
- For sand samples, the analysis again do not supports a shift in the fungal communities between plowed and unplowed samples (see PCoA plots), but the boxplot indicate a trending increased dispersion of the unplowed samples (see boxplots of the distance to centroid).

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  0.005728  0.0057275   1.3086  0.2793
## Residuals  10  0.043767  0.0043767
```

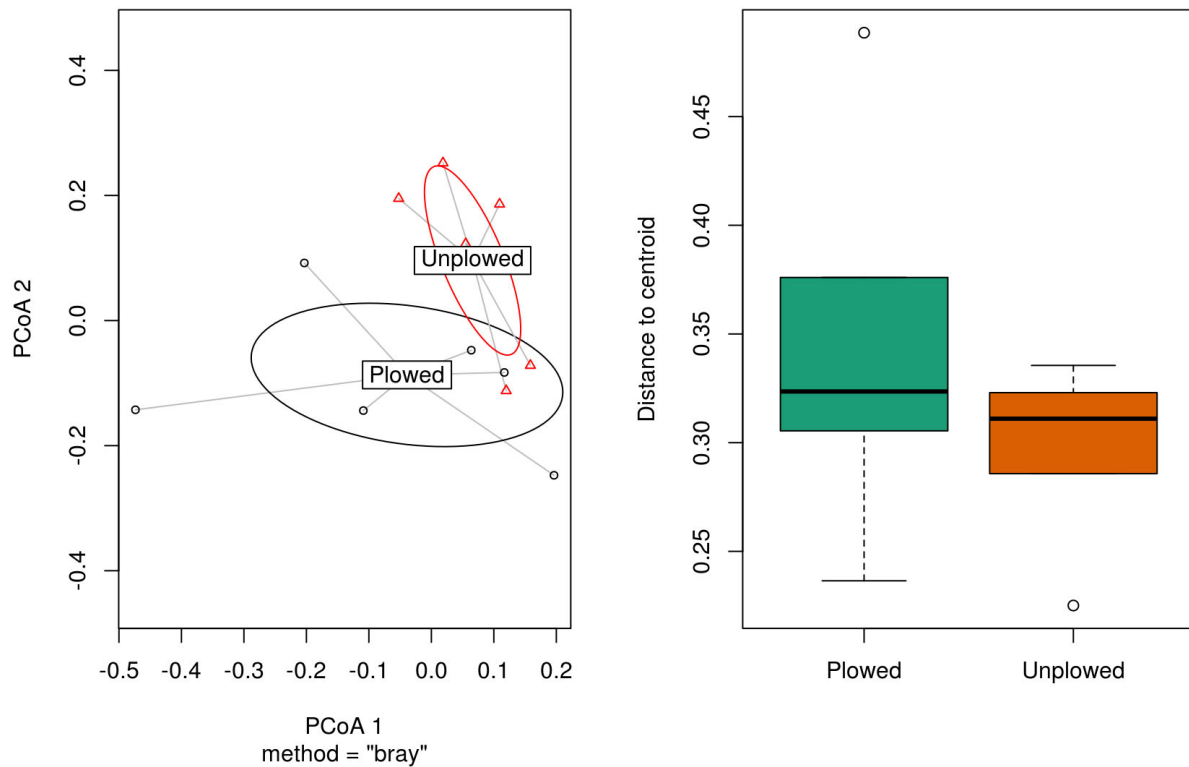


Figure 2: Visualization of the analysis of dispersion for clay samples. The left panel show a PCoA plot over plowed and unplowed clay samples based on relative abundance data, and the right panel show the distance to centroid for these categories.

Sand, with relative abundance

```
## Analysis of Variance Table
##
## Response: Distances
##          Df    Sum Sq   Mean Sq F value Pr(>F)
## Groups     1  0.0021092  0.0021092   0.6389  0.4447
## Residuals   9  0.0297137  0.0033015
```

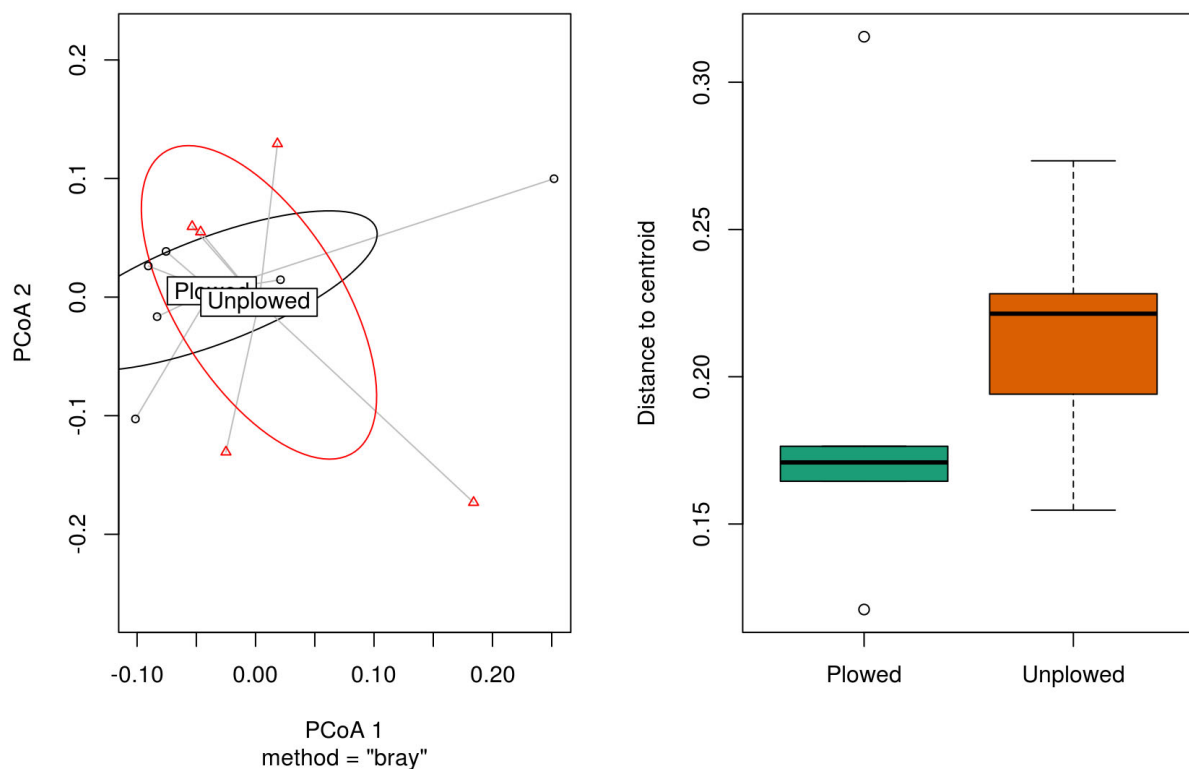


Figure 3: Visualization of the analysis of dispersion for sand samples. The left panel show a PCoA plot over plowed and unplowed sand samples based on relative abundance data and the right panel show the distance to centroid for these categories.

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, Inverse Simpson and Shannon. See **Report 2** for introduction to the different types of alpha diversity. The statistical analysis did not find a significant difference ($p\text{-value} < 0.05$) between plowed and unplowed clay or sand samples. 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 samples

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

```
## Warning in wilcox.test.default(x = c(126L, 176L, 189L, 143L, 176L, 267L), :  
## cannot compute exact p-value with ties
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: ps.sub.relab_clay$Observed by ps.sub.relab_clay$group4  
## W = 16, p-value = 0.8099  
## 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 = 17, p-value = 0.9372
## alternative hypothesis: true location shift is not equal to 0
```

```
## Clay - Inverse Simpson
```

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

Sand samples

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

```
##
## Wilcoxon rank sum test
##
## data: ps.sub.relab_sand$Observed by ps.sub.relab_sand$group4
## W = 7, p-value = 0.1775
## 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 = 8, p-value = 0.2468
## alternative hypothesis: true location shift is not equal to 0
```

```
## Clay - Shannon
```

```
##
## Wilcoxon rank sum test
##
## data: ps.sub.relab_sand$Shannon by ps.sub.relab_sand$group4
## W = 8, p-value = 0.2468
## alternative hypothesis: true location shift is not equal to 0
```

```
## Sand - Inverse Simpson
```

```
##
## Wilcoxon rank sum test
##
## data: ps.sub.relab_sand$InvSimpson by ps.sub.relab_sand$group4
## W = 12, p-value = 0.6623
## alternative hypothesis: true location shift is not equal to 0
```

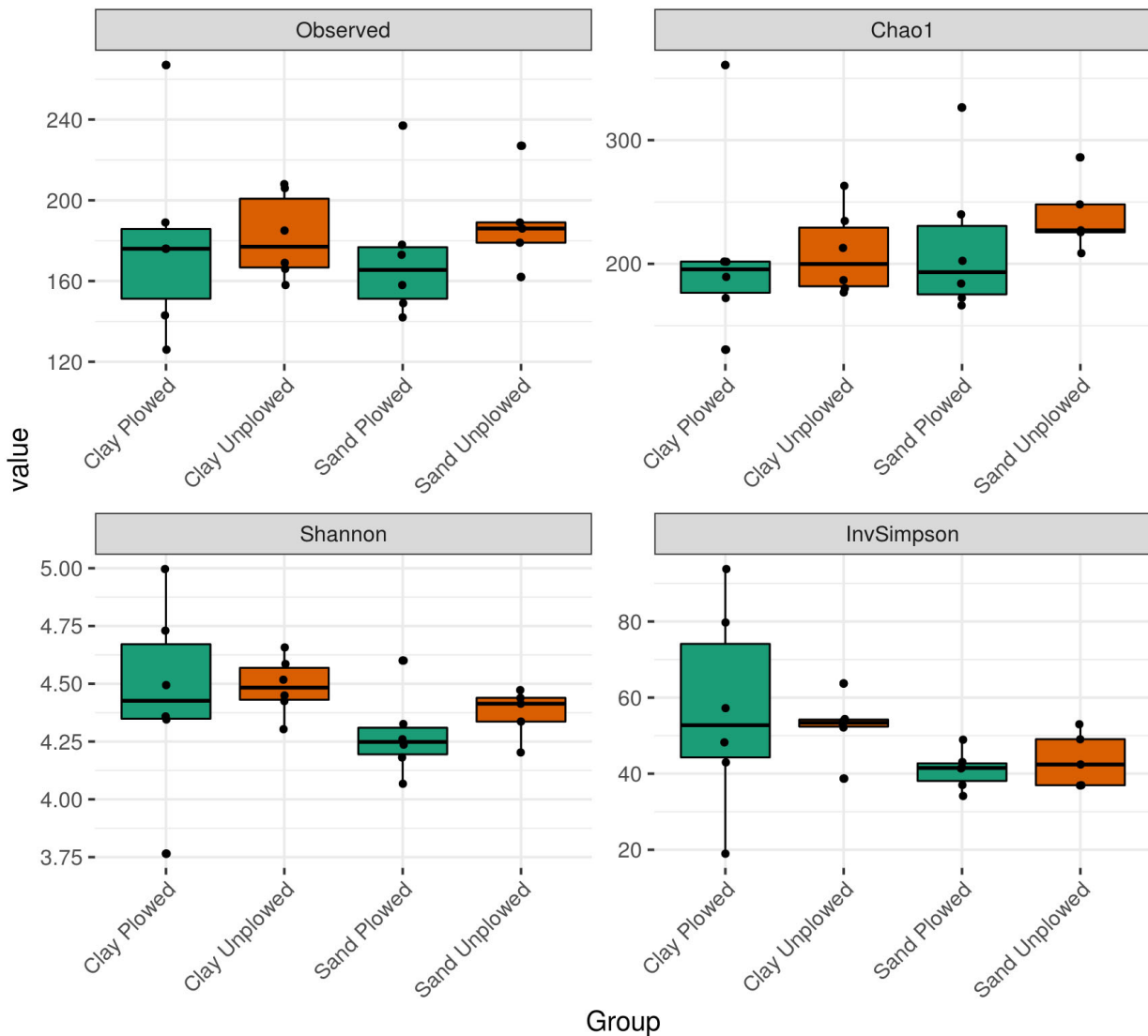


Figure 4: Alpha diversity metrics across clay and sand samples. Values of alpha-diversity for all combinations of plowed, unplowed, clay and sand samples are illustrated for observed species, Chao1, Shannon and Inverse Simpson.

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 peek at the functional differences imposed by reducing plowing activities. We first evaluate the direction and potential significance of a change in each taxon specifically within each soil type (using Wilcoxon rank sum test) and then perform a joined analysis across soil types looking for taxa that associate with plowing independent of soil type (using linear regression with soil type as a confounder and square root transformed taxa abundance). Below are the results for the top associated genera to phyla, while supplementary documents provide tables of summary statistics for all tested taxa. The difference in location information should be understood as the

median of all possible differences between the abundances in one sample group compared to another (here plowed/unplowed) within one type (sand/clay). Note that if the abundance data consists mostly of zeroes, the approximation that R calculates for this data will be close to 0.

Boxplots are used to visualize the top 6 associated taxa (selected as taxa with lowest p-value in the linear regression). The boxplots show a clear difference in the abundance of a number of single organisms despite the low number of samples analyzed in this pilot project. The Wilcoxon rank sum test is used in the analysis within soil type 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 provides a conservative estimate of significance.

It is further worth noting that many top associated fungi correspond between soil and sand, despite the clear differences in the overall microbiome composition observed in the Figure 1. This indicates that it is possible to identify indicator species for tillage-associated soil conditions across soil types and geography.

Relative abundance

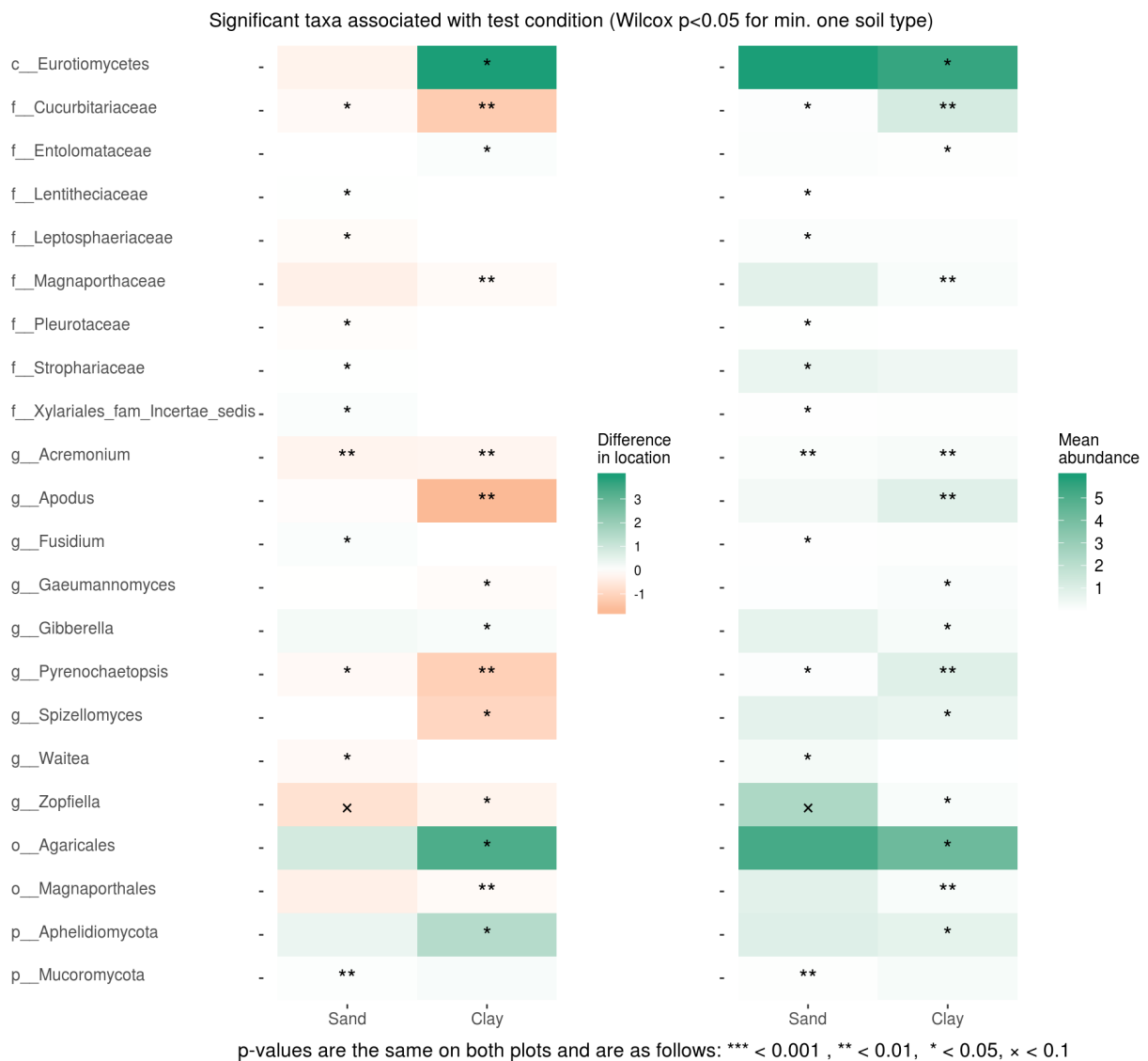


Figure 5: Overview over single taxa relative abundance tests.

Phylum Sand

Table 1

Table of top indicator phyla

Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
------	----------------	------------------------	---------	------	------

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV525	p__Mucoromycota	0.068	0.061	3.89e-03	-0.444	-0.017
ASV3	p__Mortierellomycota	18.344	-1.871	8.23e-02	-5.885	5.044
ASV216	p__Zoopagomycota	0.256	0.244	1.67e-01	-0.459	0.147
ASV53	p__Aphelidiomycota	0.973	0.392	1.70e-01	-1.420	0.204
ASV26	p__Chytridiomycota	7.794	2.596	4.29e-01	-8.022	3.426
ASV433	p__Blastocladiomycota	0.083	0.000	4.86e-01	-0.490	0.000
ASV1	p__Ascomycota	54.290	-2.892	5.37e-01	-5.580	11.087
ASV128	p__Rozellomycota	0.655	-0.086	6.62e-01	-0.944	0.963
ASV8	p__Basidiomycota	13.393	0.553	9.31e-01	-10.455	3.189
ASV218	p__Glomeromycota	0.238	-0.007	1.00e+00	-0.539	0.270

Phylum Clay

Table 2

Table of top indicator phyla

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV53	p__Aphelidiomycota	0.773	1.462	2.63e-02	-2.599	0.000
ASV218	p__Glomeromycota	3.817	2.447	9.31e-02	-6.180	0.440
ASV2949	p__Kickxellomycota	0.014	0.005	1.82e-01	-0.043	0.000
ASV849	p__Olpidiomycota	0.081	-0.107	2.80e-01	-0.025	0.180
ASV525	p__Mucoromycota	0.263	0.150	3.78e-01	-0.429	0.270
ASV1	p__Ascomycota	49.429	2.495	4.85e-01	-13.249	19.435
ASV128	p__Rozellomycota	0.089	0.000	5.99e-01	0.000	0.231
ASV216	p__Zoopagomycota	0.287	0.000	6.55e-01	-0.479	0.809
ASV3	p__Mortierellomycota	12.749	2.355	9.37e-01	-7.545	6.654
ASV8	p__Basidiomycota	14.392	0.462	9.37e-01	-5.601	5.455

Phylum Combined

Table 3

Table of top indicator phyla

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV53	p__Aphelidiomycota	0.869	0.669	2.98e-03	0.198
ASV218	p__Glomeromycota	2.106	0.404	6.61e-02	0.208

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV525	p__Mucoromycota	0.170	0.182	7.87e-02	0.098
ASV849	p__Olpidiomyota	0.042	-0.075	2.29e-01	0.060
ASV216	p__Zoopagomycota	0.272	0.125	4.39e-01	0.158
ASV26	p__Chytridiomycota	9.058	0.257	4.56e-01	0.339
ASV128	p__Rozellomycota	0.360	-0.086	5.44e-01	0.139
ASV8	p__Basidiomycota	13.914	0.115	5.94e-01	0.213
ASV1	p__Ascomycota	51.754	-0.129	6.53e-01	0.284
ASV3	p__Mortierellomycota	15.424	0.033	8.97e-01	0.252

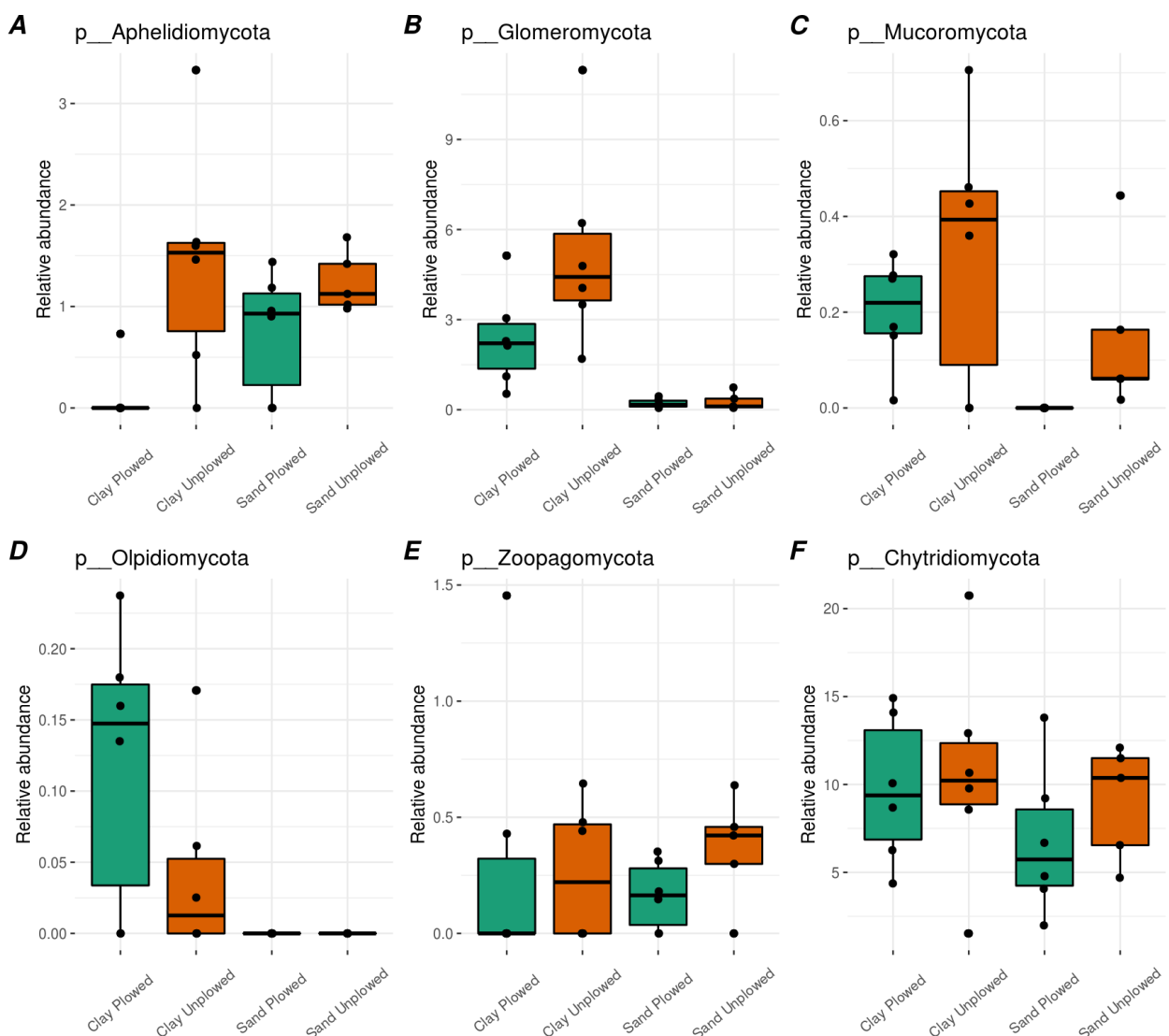


Figure 6: The top phyla with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Class Sand

Table 4

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV1	c__Sordariomycetes	25.304	-3.927	5.19e-02	-0.434	6.170
ASV3	c__Mortierellomycetes	18.329	-1.926	8.23e-02	-5.918	5.153
ASV19	c__Dothideomycetes	6.338	1.752	8.23e-02	-11.262	2.417
ASV53	c__Aphelidiomycetes	0.833	0.347	1.20e-01	-1.170	0.221
ASV216	c__Zoopagomycetes	0.256	0.244	1.67e-01	-0.459	0.147
ASV11	c__Saccharomycetes	5.398	-1.888	1.77e-01	-3.145	4.001
ASV26	c__Lobulomycetes	1.793	0.600	2.47e-01	-6.460	7.415
ASV85	c__Rhizophydiomycetes	2.495	0.611	2.47e-01	-4.391	0.599
ASV34	c__Rhizophlyctidomycetes	0.133	0.034	3.11e-01	-0.728	0.106
ASV285	c__Microbotryomycetes	0.287	-0.031	4.29e-01	-0.134	0.177

Class Clay**Table 5**

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV13	c__Eurotiomycetes	5.545	3.940	4.11e-02	-14.573	-0.184
ASV12	c__Pezizomycetes	6.703	2.922	6.49e-02	-9.177	0.194
ASV11	c__Saccharomycetes	0.288	-0.265	6.51e-02	-0.019	0.487
ASV580	c__Cystobasidiomycetes	0.169	-0.079	8.98e-02	0.000	1.393
ASV218	c__Glomeromycetes	3.732	2.408	9.31e-02	-6.180	0.343
ASV85	c__Rhizophydiomycetes	4.479	2.171	1.32e-01	-7.940	0.330
ASV44	c__Laboulbeniomycetes	0.354	-0.003	1.82e-01	0.000	3.341
ASV2949	c__Kickxellomycetes	0.014	0.005	1.82e-01	-0.043	0.000
ASV285	c__Microbotryomycetes	0.273	-0.053	2.80e-01	-0.075	2.243
ASV849	c__Olpidiomycetes	0.081	-0.107	2.80e-01	-0.025	0.180

Class Combined**Table 6**

Table of top indicator classes

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV85	c__Rhizophydiomycetes	3.530	0.520	3.56e-02	0.231

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV1	c__Sordariomycetes	25.558	-0.575	4.66e-02	0.271
ASV11	c__Saccharomycetes	2.732	-0.315	5.10e-02	0.152
ASV218	c__Glomeromycetes	2.055	0.407	6.76e-02	0.210
ASV13	c__Eurotiomycetes	5.736	0.500	8.90e-02	0.280
ASV12	c__Pezizomycetes	5.061	0.371	9.51e-02	0.212
ASV44	c__Laboulbeniomyces	0.189	-0.232	1.77e-01	0.166
ASV580	c__Cystobasidiomycetes	0.095	-0.151	1.93e-01	0.112
ASV285	c__Microbotryomycetes	0.279	-0.171	2.13e-01	0.133
ASV849	c__Olpidiomyces	0.042	-0.075	2.29e-01	0.060

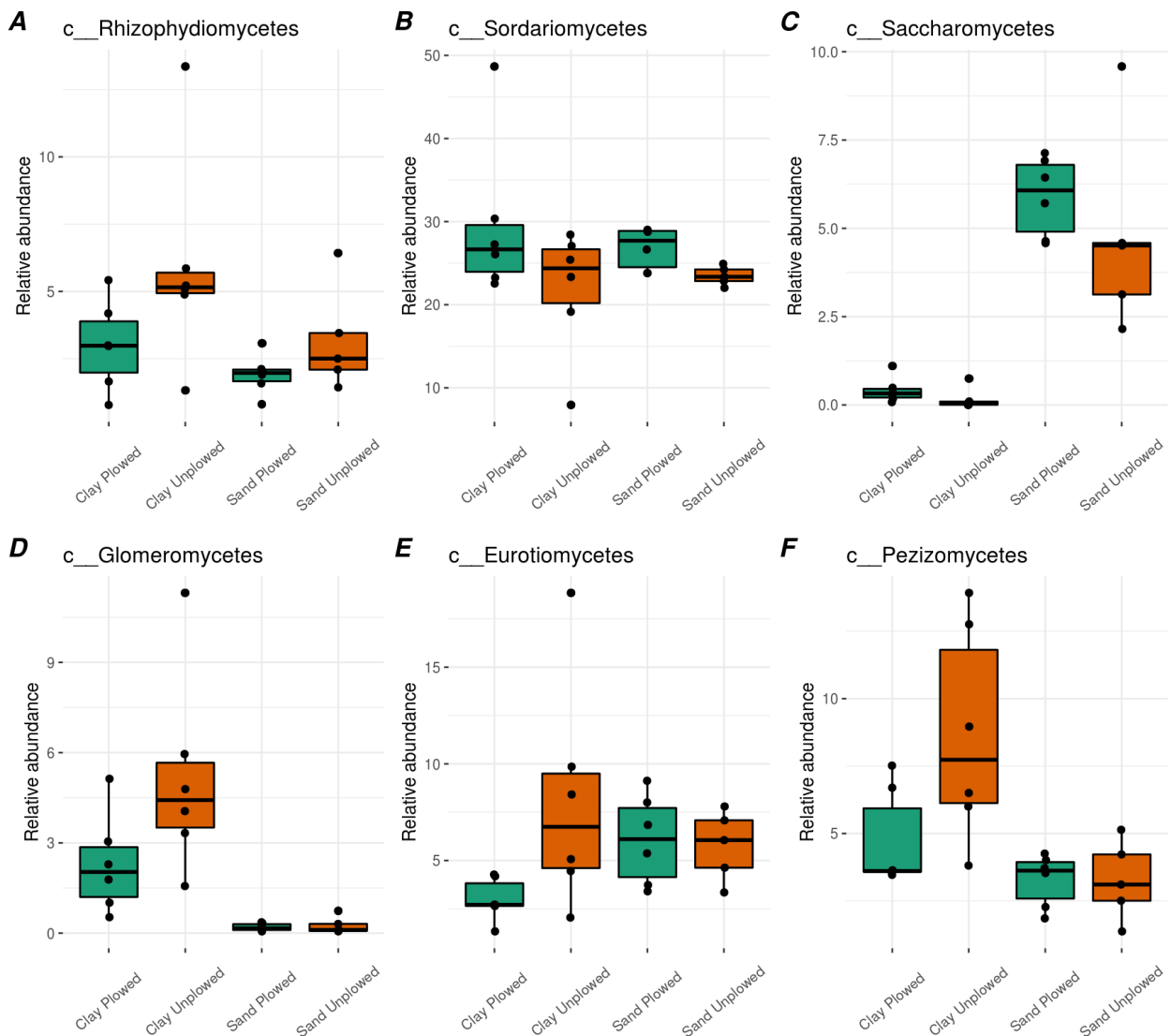


Figure 7: The top classes with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Order Sand

Table 7

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV3	o__Mortierellales	18.329	-1.926	8.23e-02	-5.918	5.153
ASV19	o__Pleosporales	5.347	1.602	8.23e-02	-11.038	2.849
ASV1200	o__Sebacinales	0.016	-0.036	1.04e-01	0.000	0.088
ASV53	o__GS16	0.833	0.347	1.20e-01	-1.170	0.221
ASV10	o__Sordariales	7.126	-2.283	1.26e-01	-0.375	4.665
ASV33	o__Onygenales	0.554	0.222	1.26e-01	-1.914	0.118
ASV77	o__Tremellales	1.006	-0.264	1.26e-01	-0.047	0.948
ASV247	o__Myrmecridiales	0.221	-0.093	1.26e-01	-0.066	0.253
ASV216	o__Zoopagales	0.256	0.244	1.67e-01	-0.459	0.147
ASV11	o__Saccharomycetales	5.398	-1.888	1.77e-01	-3.145	4.001

Order Clay**Table 8**

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV78	o__Magnaporthales	0.200	-0.128	2.78e-03	0.053	0.863
ASV39	o__Agaricales	4.245	3.358	1.52e-02	-6.375	-0.754
ASV12	o__Pezizales	6.703	2.922	6.49e-02	-9.177	0.194
ASV11	o__Saccharomycetales	0.288	-0.265	6.51e-02	-0.019	0.487
ASV218	o__Glomerales	3.422	2.393	9.31e-02	-5.947	0.527
ASV33	o__Onygenales	1.851	0.492	1.26e-01	-12.182	0.101
ASV85	o__Rhizophydiales	4.479	2.171	1.32e-01	-7.940	0.330
ASV44	o__Pyxidiophorales	0.354	-0.003	1.82e-01	0.000	3.341
ASV2949	o__Kickxellales	0.014	0.005	1.82e-01	-0.043	0.000
ASV77	o__Tremellales	0.156	0.123	2.00e-01	-1.019	0.005

Order Combined**Table 9**

Table of top indicator orders

	Taxa	Mean abundance	Difference in location	p_value	SE
--	------	----------------	------------------------	---------	----

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV78	o__Magnaporthales	0.512	-0.359	7.00e-03	0.119
ASV39	o__Agaricales	4.711	0.671	2.51e-02	0.277
ASV1200	o__Sebacinales	0.028	-0.139	3.29e-02	0.061
ASV85	o__Rhizophydiales	3.530	0.520	3.56e-02	0.231
ASV10	o__Sordariales	9.691	-0.518	4.48e-02	0.242
ASV33	o__Onygenales	1.231	0.663	4.76e-02	0.314
ASV11	o__Saccharomycetales	2.732	-0.315	5.10e-02	0.152
ASV574	o__Corticiales	0.080	0.165	7.58e-02	0.088
ASV218	o__Glomerales	1.874	0.395	8.28e-02	0.216
ASV12	o__Pezizales	5.061	0.371	9.51e-02	0.212

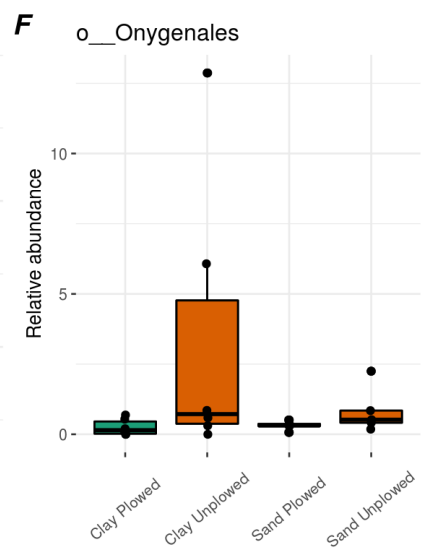
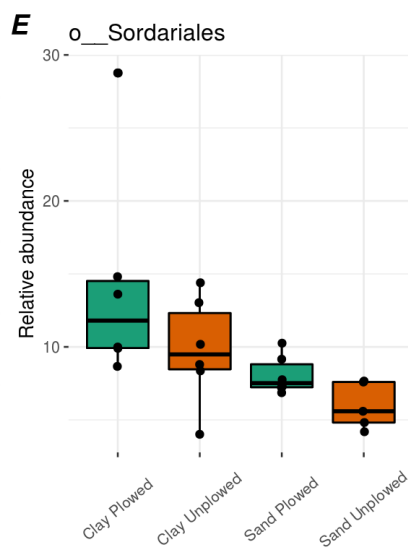
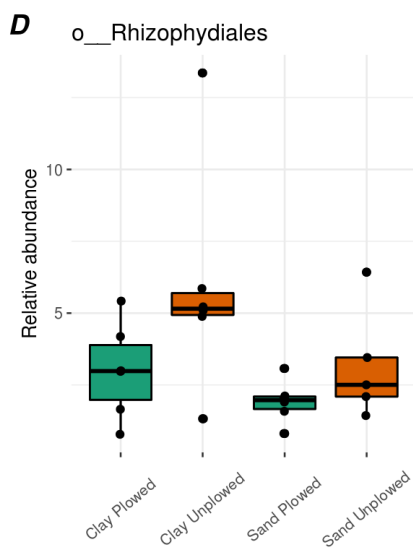
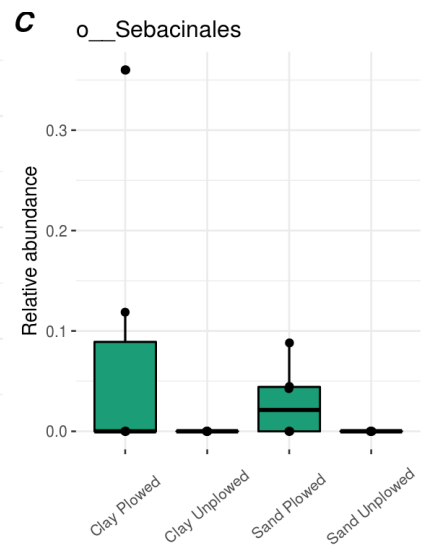
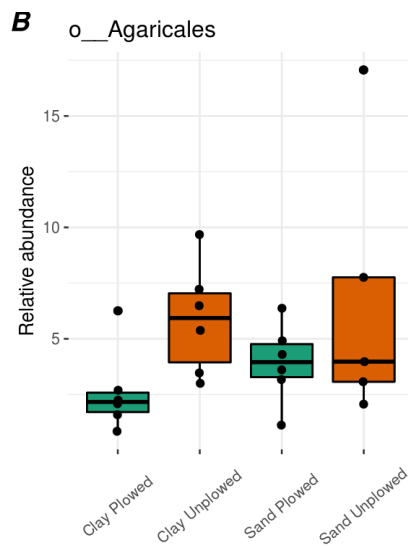
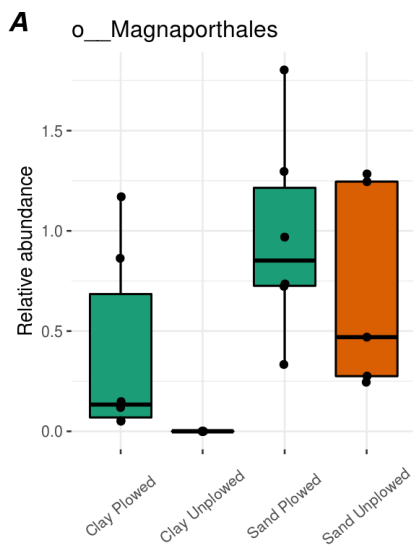


Figure 8: The top orders with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Family Sand

Table 10

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV61	f__Cucurbitariaceae	0.096	-0.147	2.77e-02	0.000	0.301
ASV242	f__Leptosphaeriaceae	0.129	-0.106	3.53e-02	0.008	0.339
ASV1916	f__Pleurotaceae	0.022	-0.058	4.46e-02	0.000	0.067
ASV40	f__Strophariaceae	0.621	0.017	4.84e-02	-6.725	0.000
ASV543	f__Xylariales_fam_Incertae_sedis	0.037	0.092	4.84e-02	-0.213	0.000
ASV1640	f__Lentitheciaceae	0.012	0.017	4.84e-02	-0.061	0.000
ASV9	f__Phaeosphaeriaceae	2.800	1.262	5.19e-02	-11.321	2.023
ASV3	f__Mortierellaceae	18.329	-1.926	8.23e-02	-5.918	5.153
ASV1200	f__Serendipitaceae	0.016	-0.036	1.04e-01	0.000	0.088
ASV247	f__Myrmecridiaceae	0.221	-0.093	1.26e-01	-0.066	0.253

Family Clay

Table 11

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV78	f__Magnaporthaceae	0.200	-0.128	2.78e-03	0.053	0.863
ASV61	f__Cucurbitariaceae	1.232	-1.242	4.33e-03	0.469	1.870
ASV863	f__Entolomataceae	0.063	0.092	1.01e-02	-0.130	-0.050

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV217	f__Clavariaceae	0.640	0.717	2.50e-02	-2.346	0.000
ASV62	f__Hypocreales_fam_Incertae_sedis	1.522	-0.701	6.37e-02	-0.298	10.563
ASV10	f__Lasiosphaeriaceae	7.611	-4.534	6.49e-02	-0.004	22.494
ASV37	f__Pyronemataceae	2.569	1.088	6.49e-02	-3.711	1.136
ASV311	f__Bulleribasidiaceae	0.035	0.062	8.98e-02	-0.098	0.000
ASV218	f__Glomeraceae	2.468	1.517	9.31e-02	-3.191	0.296
ASV516	f__Sordariaceae	0.115	0.126	1.04e-01	-0.425	0.000

Family Combined

Table 12

Table of top indicator families

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV61	f__Cucurbitariaceae	0.689	-0.492	3.18e-04	0.113
ASV78	f__Magnaporthaceae	0.512	-0.359	7.00e-03	0.119
ASV1916	f__Pleurotaceae	0.013	-0.097	1.26e-02	0.035
ASV217	f__Clavariaceae	0.334	0.417	1.91e-02	0.163
ASV85	f__Rhizophydiales_fam_Incertae_sedis	1.574	0.602	2.42e-02	0.247
ASV10	f__Lasiosphaeriaceae	5.263	-0.711	2.55e-02	0.295
ASV1200	f__Serendipitaceae	0.028	-0.139	3.29e-02	0.061
ASV60	f__Hypocreaceae	0.753	-0.283	3.83e-02	0.127
ASV12	f__Ascobolaceae	2.144	0.547	4.40e-02	0.255

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV62	f__Hypocreales_fam_Incertae_sedis	1.004	-0.579	4.75e-02	0.274

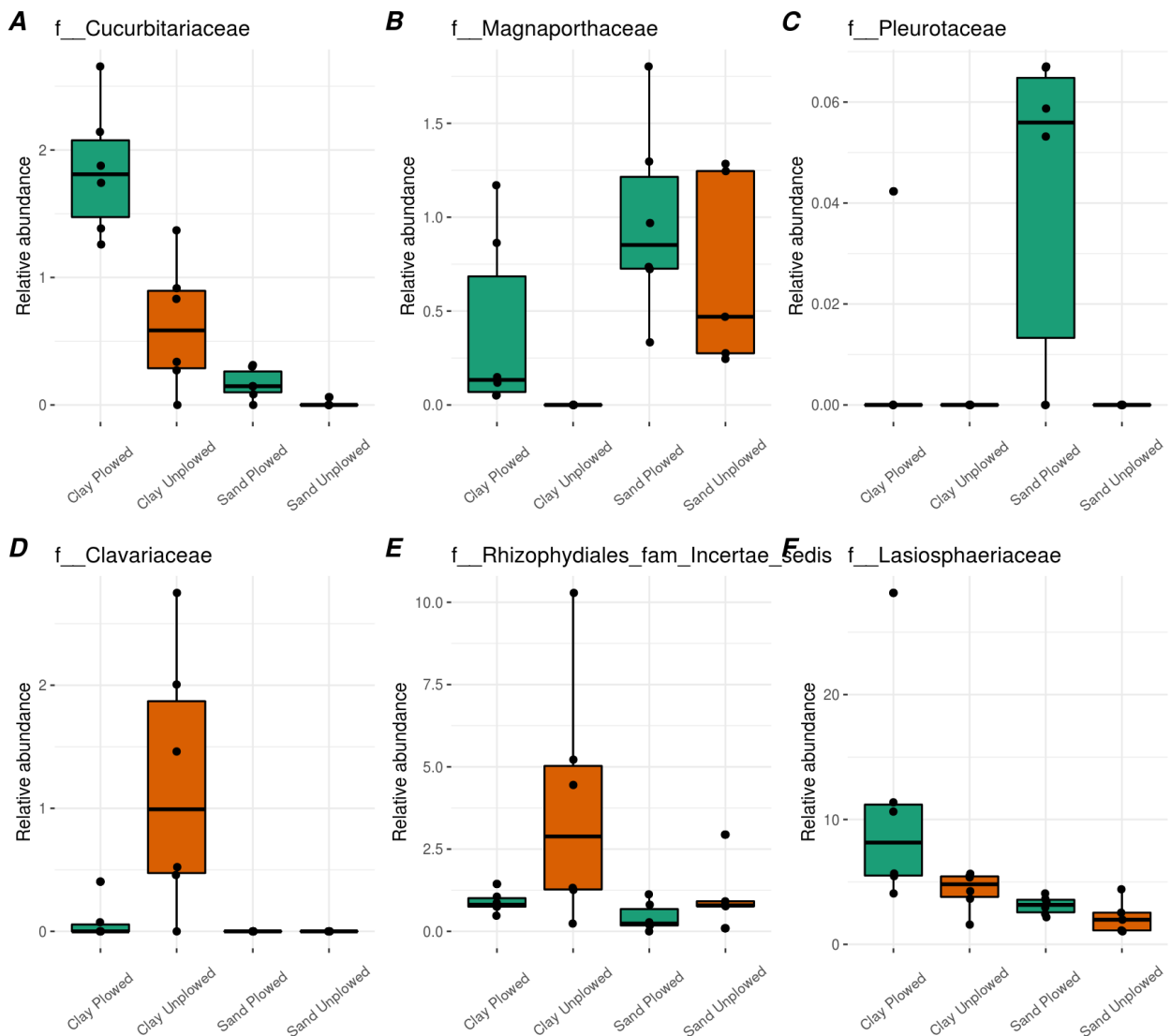


Figure 9: The top families with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Genus Sand

Table 13

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV203	g__Cladorrhinum	0.176	0.313	5.49e-03	-0.800	-0.082
ASV154	g__Acremonium	0.145	-0.266	7.55e-03	0.074	0.444
ASV11	g__Nadsonia	4.472	-2.416	8.66e-03	0.392	3.645
ASV665	g__Septoria	0.045	0.061	1.50e-02	-0.245	0.000

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV644	g__Plenodomus	0.100	-0.093	2.22e-02	0.011	0.189
ASV61	g__Pyrenochaetopsis	0.096	-0.147	2.77e-02	0.000	0.301
ASV114	g__Waitea	0.329	-0.146	4.46e-02	0.000	1.850
ASV543	g__Fusidium	0.037	0.092	4.84e-02	-0.213	0.000
ASV9	g__Parastagonospora	1.641	0.800	5.19e-02	-11.261	1.288
ASV3	g__Mortierella	18.329	-1.926	8.23e-02	-5.918	5.153

Genus Clay

Table 14

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV103	g__Apodus	0.909	-1.696	3.66e-03	0.191	3.283
ASV154	g__Acremonium	0.227	-0.237	3.66e-03	0.124	1.549
ASV61	g__Pyrenochaetopsis	0.952	-1.153	4.33e-03	0.487	1.677
ASV120	g__Spizellomyces	0.632	-0.990	1.24e-02	0.124	1.566
ASV217	g__Clavaria	0.631	0.728	1.67e-02	-2.346	0.000
ASV18	g__Zopfiella	0.199	-0.277	2.84e-02	0.000	0.719
ASV124	g__Gibberella	0.266	0.142	2.84e-02	-1.053	0.000
ASV557	g__Gaeumannomyces	0.184	-0.093	2.84e-02	0.000	0.863
ASV58	g__Chaetomium	1.943	1.970	4.11e-02	-5.480	-0.086
ASV1936	g__Entoloma	0.024	0.053	7.40e-02	-0.092	0.000

Genus Combined

Table 15

Table of top indicator genera

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV154	g__Acremonium	0.187	-0.495	6.62e-06	0.082
ASV61	g__Pyrenochaetopsis	0.542	-0.501	1.67e-04	0.109
ASV103	g__Apodus	0.647	-0.640	2.04e-03	0.181
ASV203	g__Cladorrhinum	0.084	0.261	2.72e-03	0.076
ASV18	g__Zopfiella	1.269	-0.412	5.13e-03	0.131
ASV11	g__Nadsonia	2.139	-0.262	1.14e-02	0.094

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV557	g__Gaeumannomyces	0.125	-0.308	1.57e-02	0.117
ASV217	g__Clavaria	0.329	0.439	1.58e-02	0.166
ASV644	g__Plenodomus	0.048	-0.103	2.04e-02	0.041
ASV2060	g__Udeniomyces	0.015	-0.107	2.25e-02	0.043

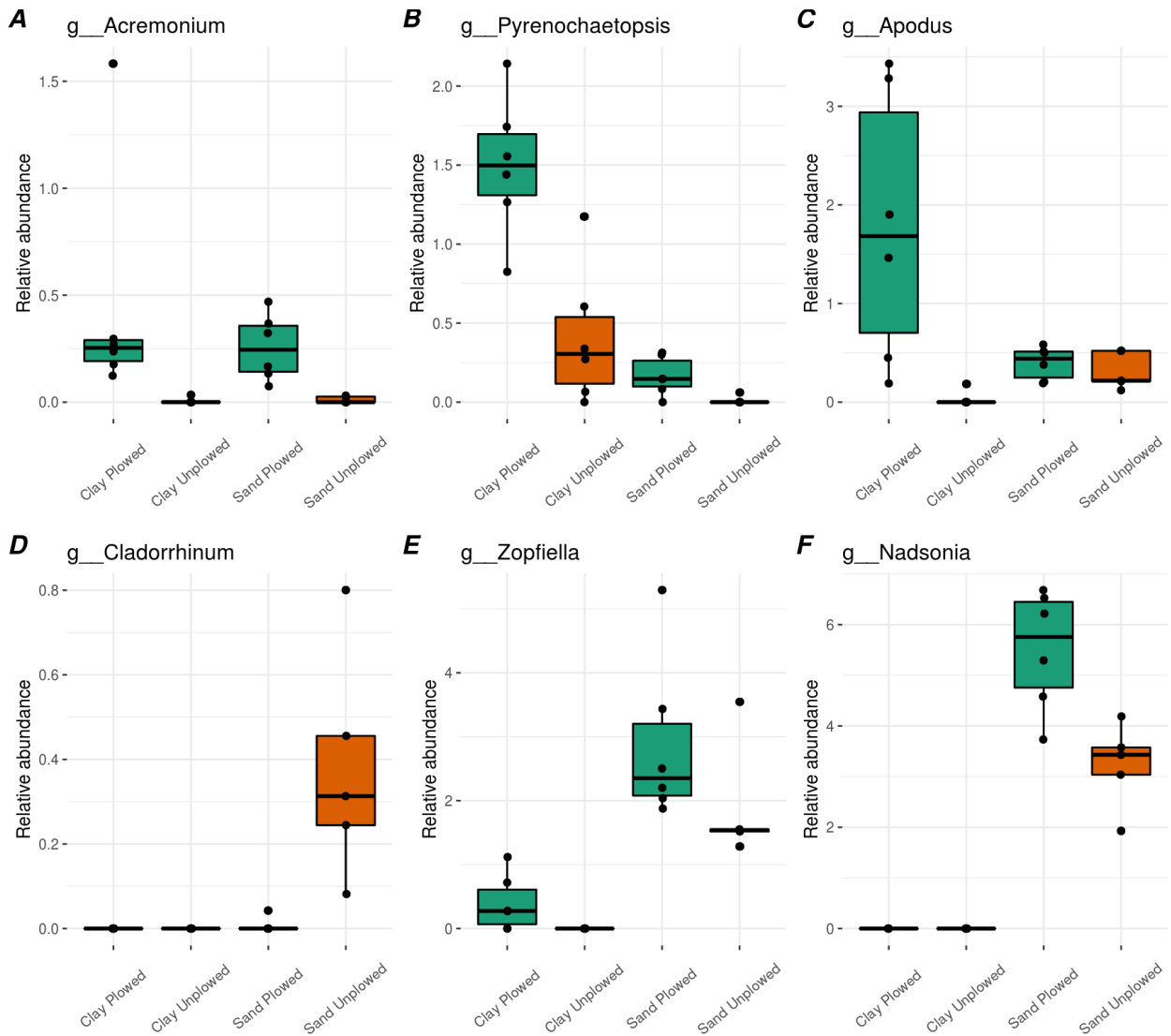


Figure 10: The top genera with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Version information

Table 16: Software and package versions.

	version
OS	Ubuntu 16.04.6 LTS
R	3.6.1

	version
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
openxlsx	4.1.3
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
survival	2.44-1.1
iterators	1.0.12
ape	5.3

	version
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
abind	1.4-5
mvtnorm	1.0-11
DBI	1.0.0
Rcpp	1.0.3
viridisLite	0.3.0
foreign	0.8-71
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
labeling	0.3
tidyselect	0.2.5
rlang	0.4.2
reshape2	1.4.3
munsell	0.5.0

	version
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
zip	2.0.4
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
nloptr	1.2.1
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
rio	0.5.16
IRanges	2.20.1
sessioninfo	1.1.1
codetools	0.2-16
pkgload	1.0.2
boot	1.3-22
MASS	7.3-51.4
assertthat	0.2.1
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
minqa	1.2.4
rmarkdown	1.18
Biobase	2.46.0
lubridate	1.7.4