

Statistical Analysis Report

by BiomCare

04/02/2020

Customer	SEGES
Customer ID	DA00201-19
Project	Differences between bacteria and fungi in plowed and unplowed clay and sand samples
Sample Type	Soil
Number of samples	24 samples
Type of data	Sequencing of bacterial 16S rRNA (V3V4) with spike-in

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 absolute bacterial load
- Differences in alpha-diversity
- Differential abundance of single taxa
 - Relative abundance
 - Absolute abundance
- Version information

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: Microbiome profiles).

Through biostatistical analysis we compared the microbiome profiles between sand and clay, and plowed and unplowed samples. We refer to the status of sand/clay, 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 high effect on the microbiome.

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. Alpha-diversity is a measure of the within-sample diversity and we evaluated whether the alpha-diversity differed between testing conditions using four different measures of alpha-diversity.

4. 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; phylum, class, order, family and genus.

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 described 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.

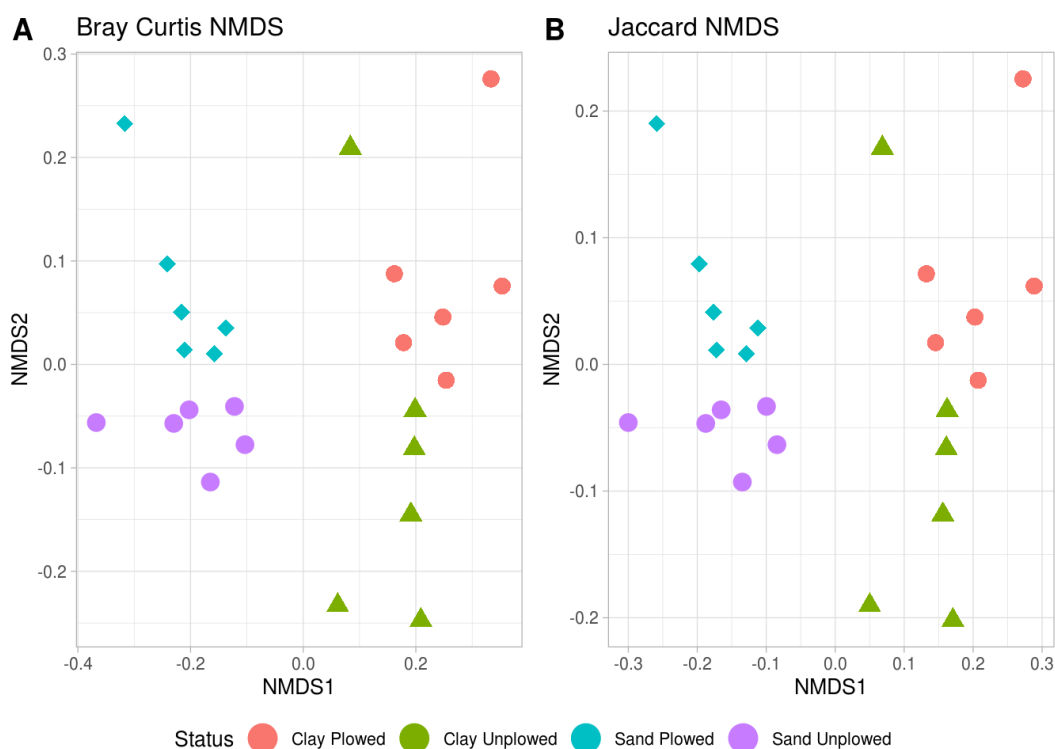


Figure 1: Visualization of the bacterial community in the 24 soil samples. Ordination plots using **A)** Bray distance and **B)** Jaccard distance and relative abundance. The plots were made using nonmetric multidimensional scaling (NMDS) and beta-diversity at the genera level. Samples are colored by 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 up to the phyla level.

Adonis for relative abundance data

Comparison of plowed and unplowed clay samples showed a significant difference using both Bray-Curtis (variance explained $R^2=0.18$, p -value=0.003) and Jaccard (variance explained $R^2=0.15$, p -value=0.007) at genera level. Also for sand, the comparisons were significant with a notable amount of variation in the bacterial community explained by the testing condition (Bray-Curtis $R^2=0.17$, p -value=0.011; Jaccard $R^2=0.15$, p -value=0.017). The patterns were also evident at the higher taxonomic levels (see details in summary statistics below). This is a noteworthy difference considering the small number of samples, the consistent shift between the two soil types (along the second ordination axis) and the consistent pattern across taxonomic levels.

Taxa level	Soil type	Bray		Jaccard	
		R2	p	R2	p
Phylum	Clay	0.1128	0.27	0.1059	0.309
	Sand	0.2846	0.012	0.2712	0.009
Class	Clay	0.184	0.056	0.1693	0.058
	Sand	0.2092	0.014	0.2015	0.007
Order	Clay	0.1742	0.046	0.1582	0.076
	Sand	0.1733	0.018	0.1679	0.027
Family	Clay	0.1461	0.045	0.1319	0.074
	Sand	0.1774	0.016	0.1613	0.007
Genus	Clay	0.1801	0.003	0.1545	0.007
	Sand	0.1695	0.011	0.1484	0.017

Table 1: Results from ADONIS analysis. Results from analysis of the microbiome community association with the testing condition within sand and clay samples. The ADONIS analysis was performed for relative abundance data, for both Bray-Curtis and Jaccard beta-diversity measures. The table shows p-values and variation explained (R^2).

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 ($R^2=0.1$, p -value=0.306) or Jaccard ($R^2=0.1$, p -value=0.282) at the genera level. However, the variance explained reached 9-10% 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, with a decreasing variation explained found as we move towards the phyla level.

Taxa level	Soil type	Bray		Jaccard	
		R2	p	R2	p
Phylum	Clay	0.0429	0.634	0.0503	0.668
	Sand	0.0412	0.648	0.0641	0.516
Class	Clay	0.046	0.615	0.0537	0.67

	Sand	0.0446	0.709	0.0636	0.619
Order	Clay	0.0458	0.65	0.0534	0.724
	Sand	0.0592	0.664	0.075	0.569
Family	Clay	0.0717	0.498	0.0743	0.568
	Sand	0.0811	0.557	0.0799	0.606
Genus	Clay	0.0983	0.306	0.0969	0.282
	Sand	0.096	0.374	0.097	0.311

Table 2: Results from ADONIS analysis. Results from analysis of the microbiome community association with the testing condition within sand and clay samples. The ADONIS analysis was performed for absolute abundance data, for both Bray-Curtis and Jaccard beta-diversity measures. The table shows p-values and variation explained (R²).

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 seen (see boxplots of the distance to centroid). Moving from relative abundance data to absolute abundance data, the analysis of differences in group dispersion showed no significant differences (thereby agreeing with the relative abundance data) (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 0.000418 0.00041799   0.2063 0.6594
## Residuals 10 0.020262 0.00202616
```

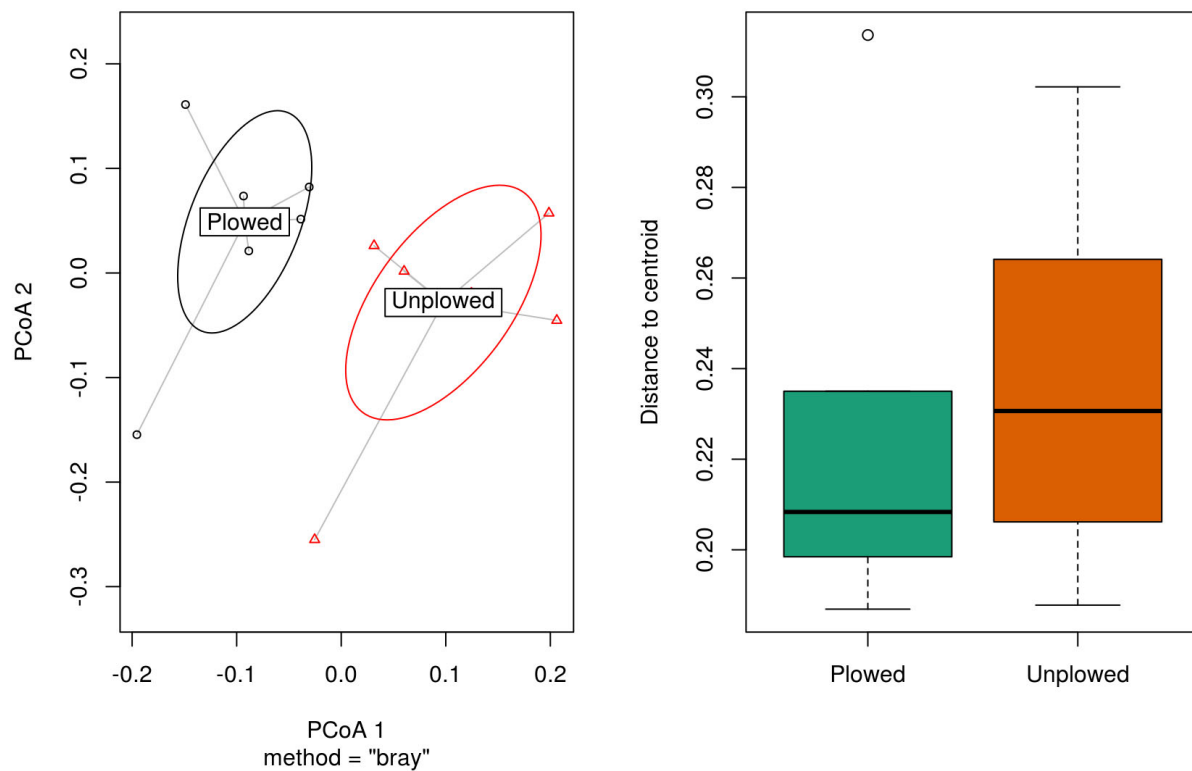


Figure 2: Visualization of the analysis of dispersion for clay samples. The left panel show a PCoA plot over plowed and unplowed clay samples based on relative abundance data, and the right panel show the distance to centroid for these categories.

Sand with relative abundance

```
## Analysis of Variance Table
##
## Response: Distances
##      Df      Sum Sq    Mean Sq F value Pr(>F)
## Groups    1 0.0002113 0.00021134  0.1006 0.7576
## Residuals 10 0.0210082 0.00210082
```

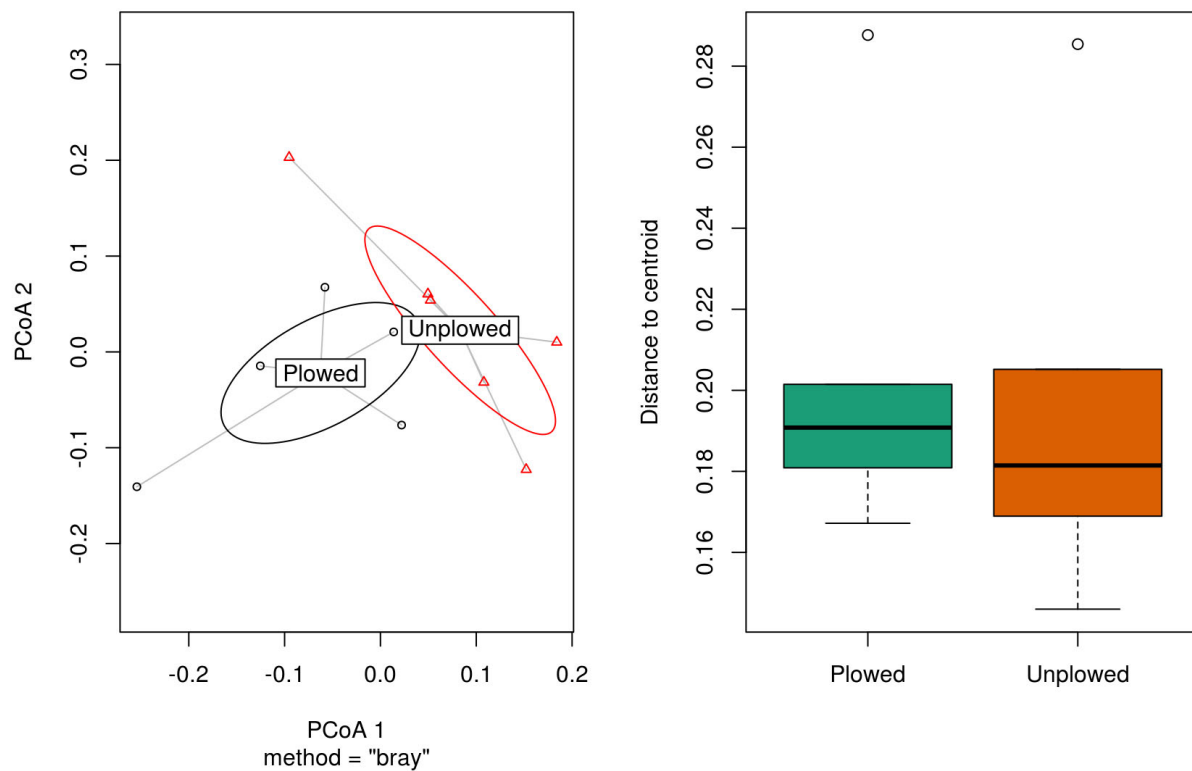


Figure 3: Visualization of the analysis of dispersion for sand samples. The left panel show a PCoA plot over plowed and unplowed sand samples based on relative abundance data and the right panel show the distance to centroid for these categories.

Clay with absolute abundance

```
## Analysis of Variance Table
##
## Response: Distances
##      Df  Sum Sq  Mean Sq F value Pr(>F)
## Groups    1 0.003410  0.0034104   0.4291 0.5272
## Residuals 10 0.079477  0.0079477
```

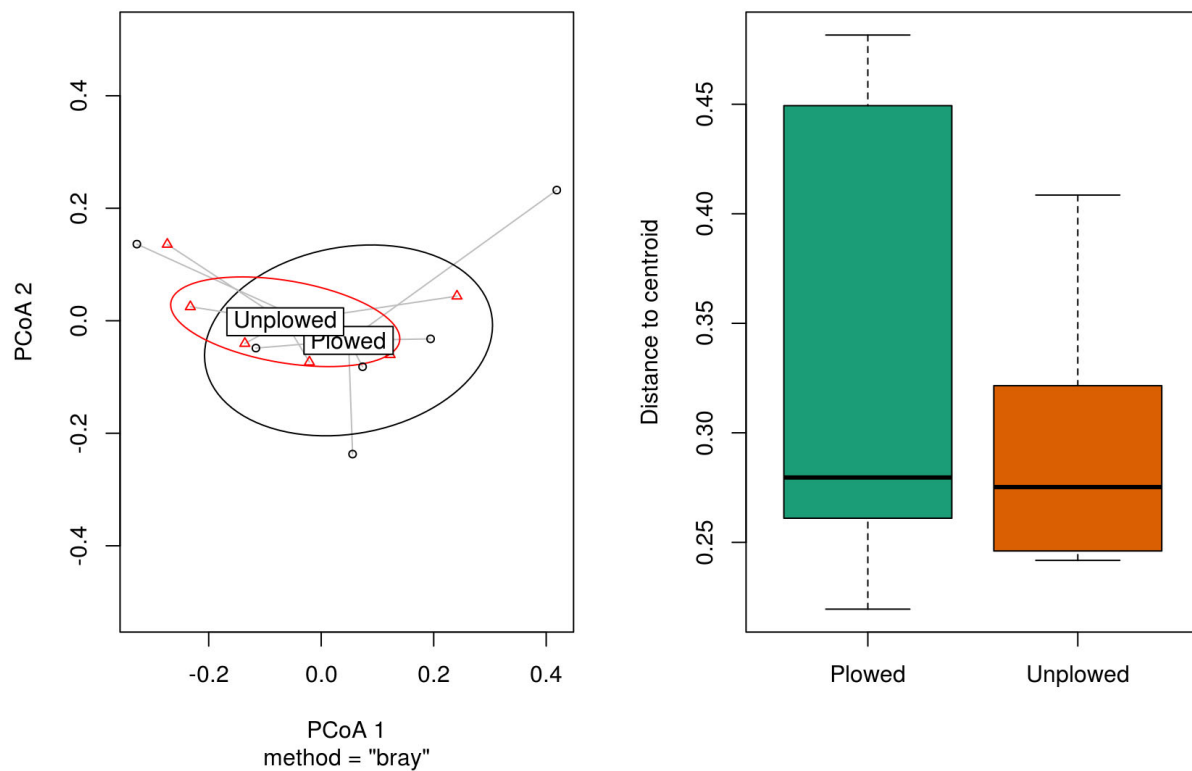


Figure 4: Visualization of the analysis of dispersion for clay samples. The left panel show a PCoA plot over plowed and unplowed clay samples based on absolute abundance data and the right panel show the distance to centroid for these categories.

Sand with absolute abundance

```
## Analysis of Variance Table
##
## Response: Distances
##      Df    Sum Sq   Mean Sq F value Pr(>F)
## Groups    1 0.000974 0.0009745    0.148  0.7085
## Residuals 10 0.065830 0.0065830
```

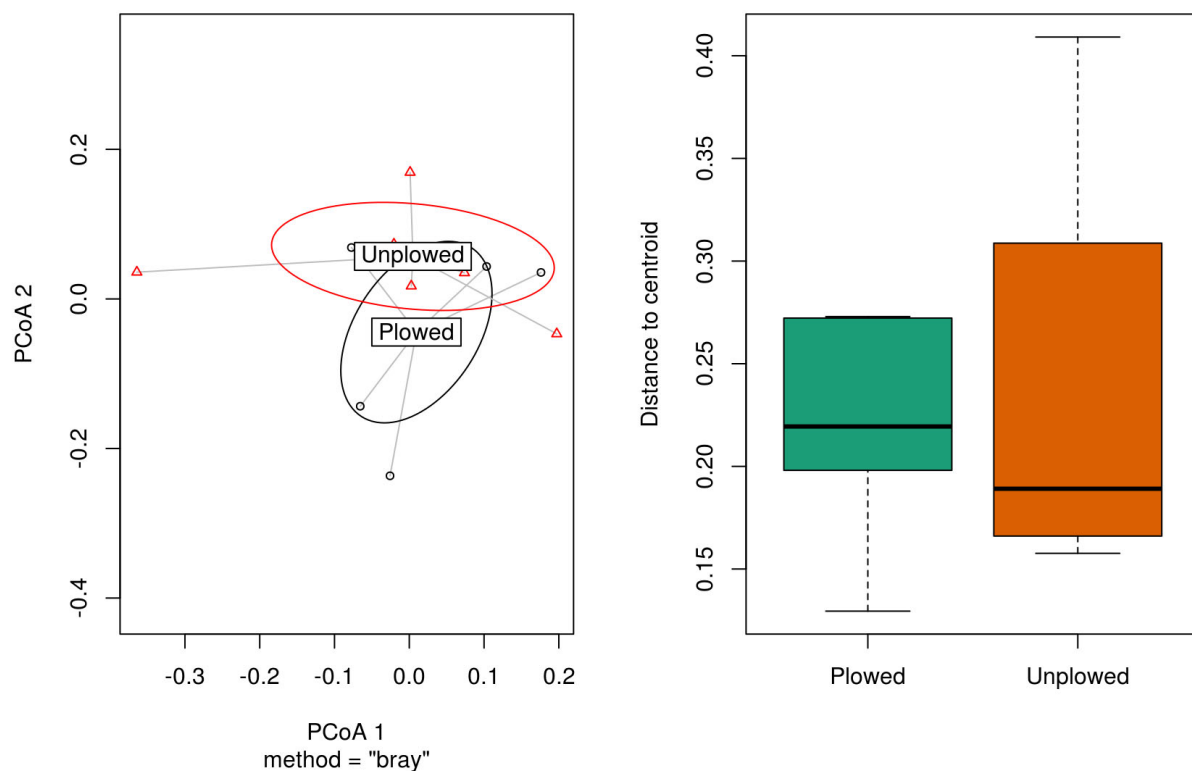


Figure 5: Visualization of the analysis of dispersion for sand samples. The left panel show a PCoA plot over plowed and unplowed sand samples based on absolute abundance data and the right panel show the distance to centroid for these categories.

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 support a difference in the absolute amount of bacteria between the plowed and unplowed clay samples, showing a lower amount of bacteria in the minimally tilled samples (p-value = 0.026). For sand the difference was not significant, but the boxplot indicated a trending increase however with one strong outlier (p-value > 0.05).

```
## CLAY
```

```
##
## Wilcoxon rank sum test
##
## data: asv.meta_clay$Sample_sum by asv.meta_clay$group4
## W = 32, p-value = 0.02597
## 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 = 10, p-value = 0.2403
## alternative hypothesis: true location shift is not equal to 0
```

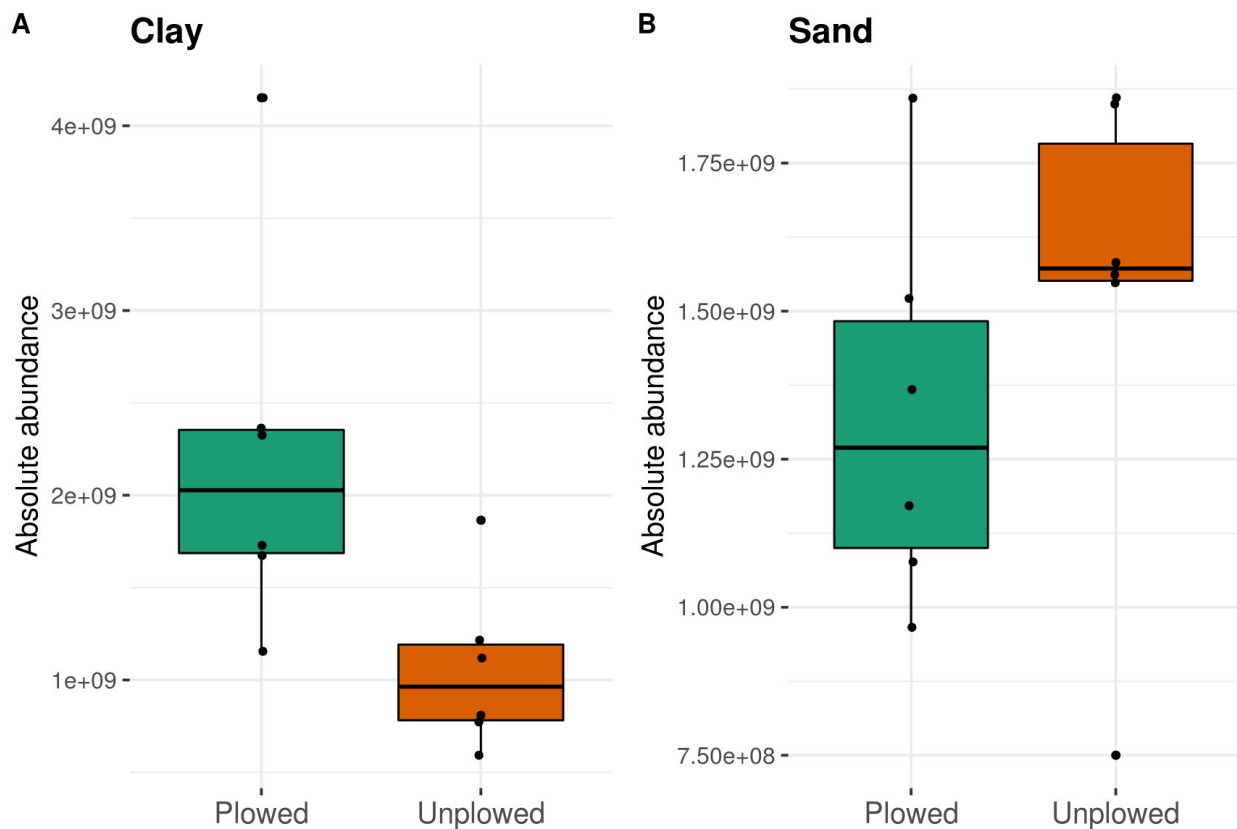


Figure 6: The total absolute abundance for sand and clay samples. The sum of the absolute abundances across plowed and unplowed samples of clay (A) and sand (B).

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 four 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-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 (p-values of 0.18-0.14). The inability 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
```

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

```
##  
## Wilcoxon rank sum test  
##  
## data: ps.sub.relab_clay$InvSimpson by ps.sub.relab_clay$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
```

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

```
##  
## Wilcoxon rank sum test  
##  
## data: ps.sub.relab_sand$InvSimpson by ps.sub.relab_sand$group4  
## W = 10, p-value = 0.2403  
## alternative hypothesis: true location shift is not equal to 0
```

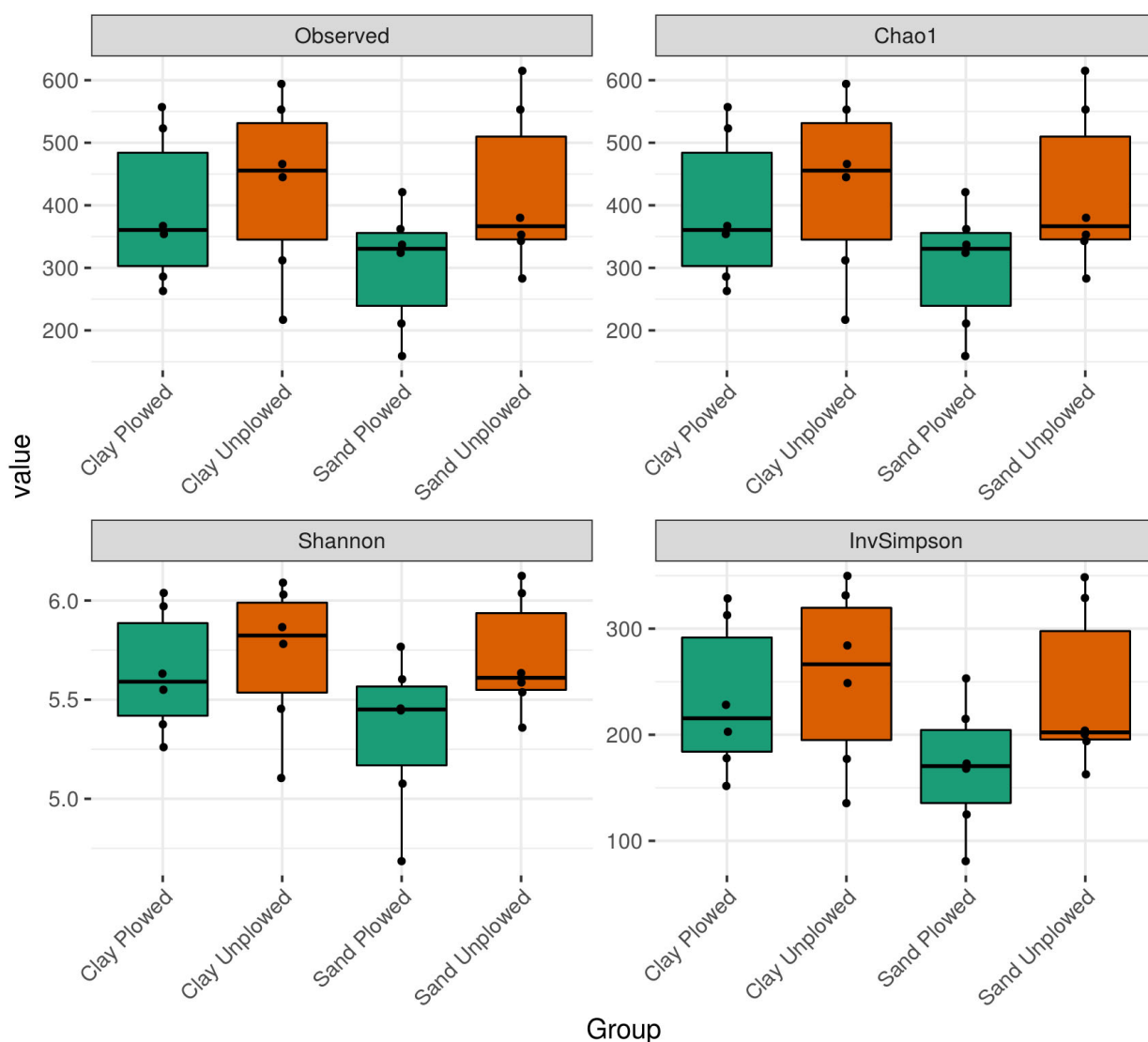


Figure 7: Alpha diversity measures for plowed and unplowed samples. Values of alpha-diversity for all combinations of plowed, unplowed, clay and sand samples are illustrated for four measures of alpha diversity (calculated using absolute abundance data).

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). 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. 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 previously been found increased in soil subjected 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 **Figure 1**. This indicates that it is possible to identify indicator species for tillage-associated soil conditions across soil types and geography.

Relative abundance

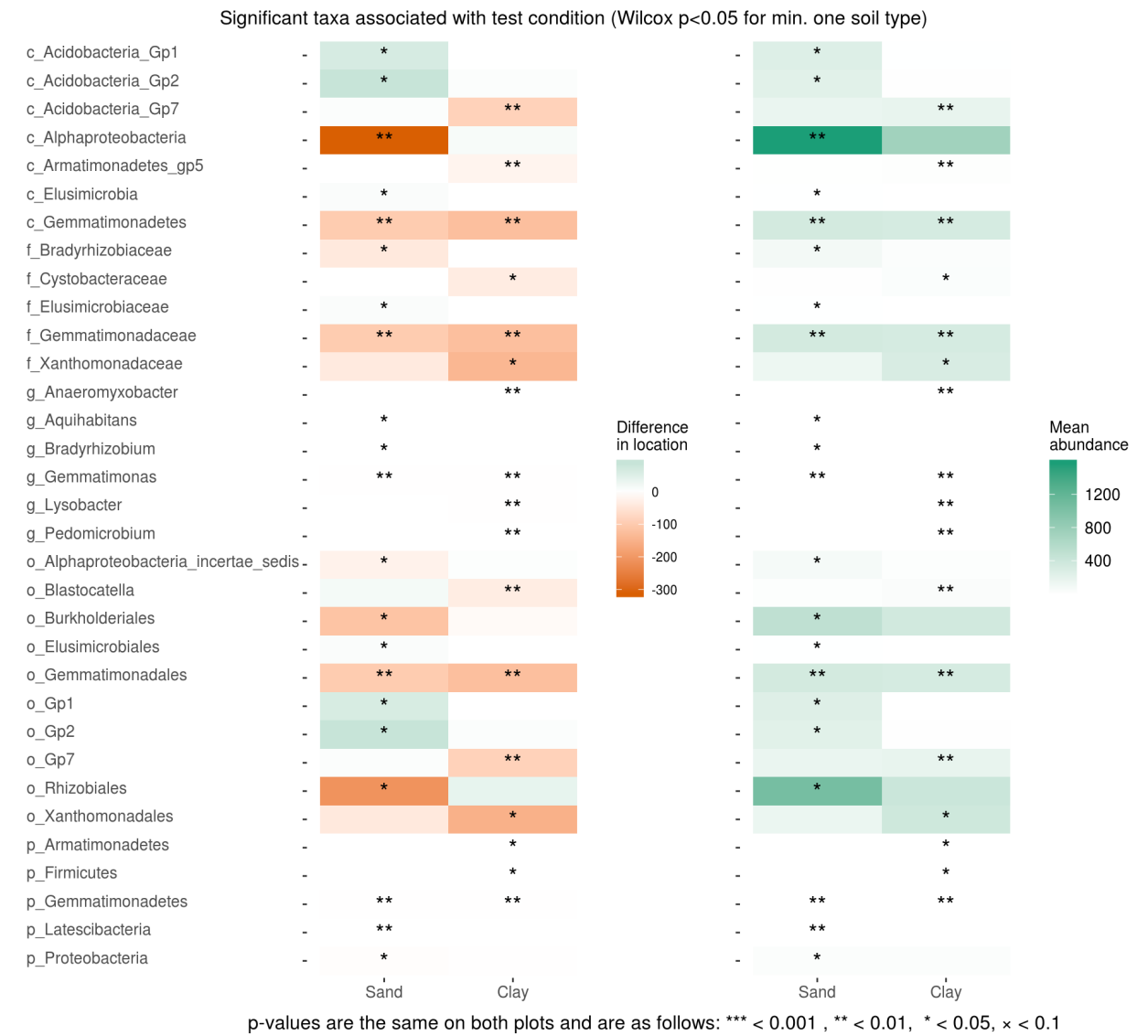


Figure 8: Overview over single taxa associations from relative abundance tests. This plot allows an easy overview of the association patterns observed in clay and sand. Taxa are included if they had a significant ($p < 0.05$) association with plowing status in either sand or clay.

Phylum Sand

Table 3

Table of top indicator phyla

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV69	Latescibacteria	1.012	0.523	4.33e-03	-0.817	-0.242
ASV44	Gemmatimonadetes	3.486	-1.000	8.66e-03	0.196	1.341
ASV14	Proteobacteria	35.050	-3.911	1.52e-02	0.356	8.251
ASV851	Elusimicrobia	0.071	0.075	8.98e-02	-0.232	0.000
ASV10	Acidobacteria	29.136	3.245	1.80e-01	-5.738	0.466

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV470	Planctomycetes	0.039	0.019	2.13e-01	-0.121	0.024
ASV16	Actinobacteria	12.249	-1.930	3.10e-01	-4.555	3.552
ASV1680	Candidatus_Saccharibacteria	0.136	0.073	3.36e-01	-0.320	0.064
ASV20	Nitrospirae	1.918	-0.095	3.94e-01	-0.213	0.407
ASV162	Chloroflexi	0.541	0.112	4.85e-01	-0.322	0.120

Phylum Clay

Table 4

Table of top indicator phyla

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonadetes	3.299	-1.243	8.66e-03	0.176	1.974
ASV486	Armatimonadetes	0.267	-0.271	1.29e-02	0.109	0.590
ASV133	Firmicutes	0.991	0.506	2.60e-02	-0.879	-0.034
ASV2923	BRC1	0.007	0.016	7.40e-02	-0.028	0.000
ASV470	Planctomycetes	0.206	-0.084	1.28e-01	-0.048	0.300
ASV16	Actinobacteria	11.260	2.828	1.32e-01	-6.665	0.863
ASV20	Nitrospirae	1.689	0.481	1.32e-01	-1.079	0.060
ASV69	Latescibacteria	1.996	0.497	2.40e-01	-1.425	0.469
ASV887	Fibrobacteres	0.021	0.000	2.53e-01	0.000	0.063
ASV50	Bacteroidetes	7.556	-1.440	3.10e-01	-1.405	4.698

Phylum Combined

Table 5

Table of top indicator phyla

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV44	Gemmatimonadetes	3.392	-0.285	1.69e-04	0.062
ASV69	Latescibacteria	1.504	0.219	6.57e-03	0.072
ASV133	Firmicutes	0.678	0.135	3.44e-02	0.060
ASV14	Proteobacteria	32.323	-0.205	1.01e-01	0.120
ASV486	Armatimonadetes	0.179	-0.151	1.13e-01	0.091
ASV50	Bacteroidetes	6.134	-0.202	1.76e-01	0.144
ASV20	Nitrospirae	1.804	0.080	2.24e-01	0.064
ASV16	Actinobacteria	11.755	0.177	2.85e-01	0.162
ASV851	Elusimicrobia	0.046	0.073	3.34e-01	0.073

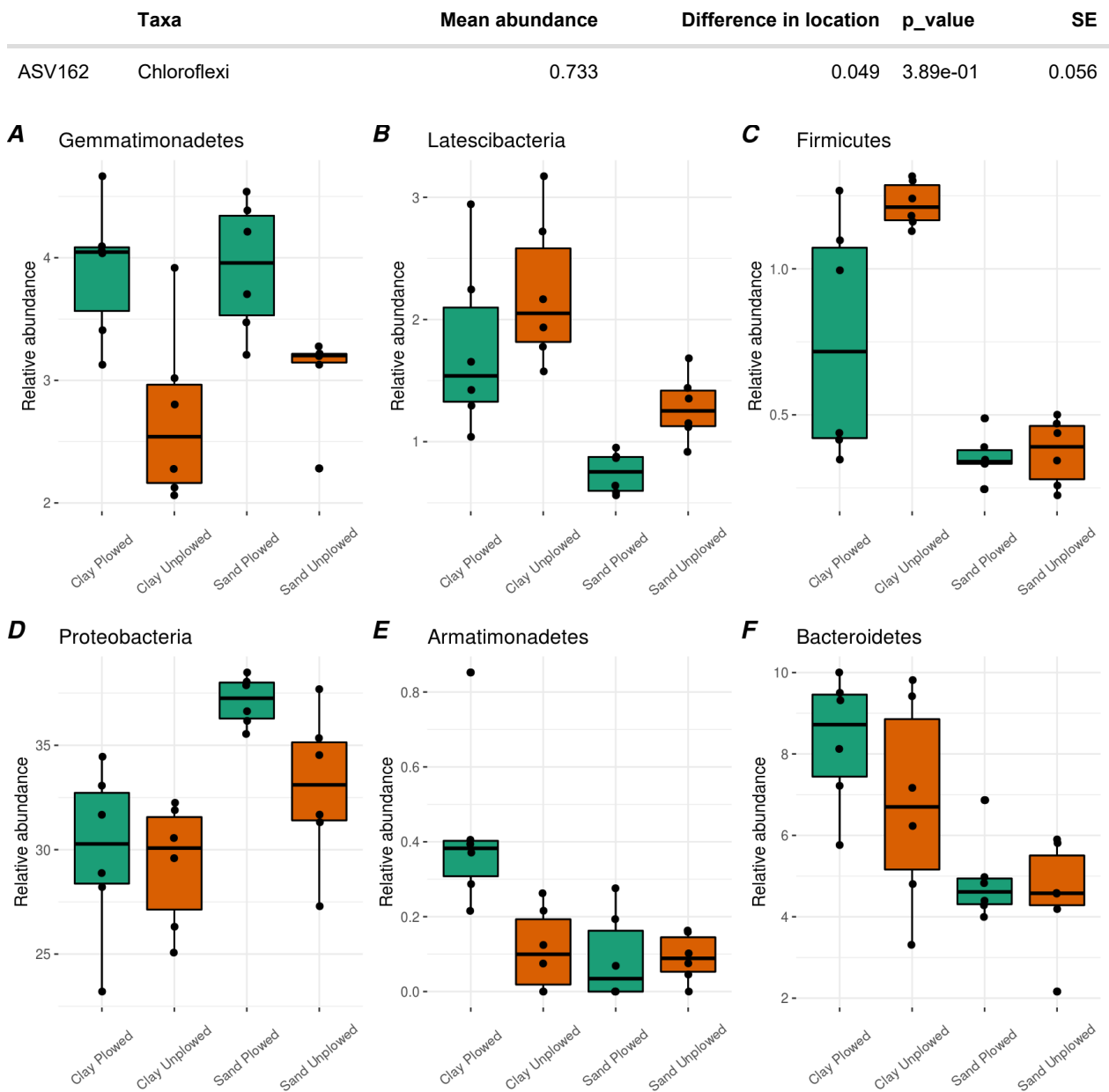


Figure 9: The top phyla with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Class Sand

Table 6

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV14	Alphaproteobacteria	15.773	-3.132	4.33e-03	1.242	5.157
ASV44	Gemmatimonadetes	3.486	-1.000	8.66e-03	0.196	1.341
ASV49	Acidobacteria_Gp1	2.677	0.641	1.52e-02	-1.124	-0.127
ASV134	Acidobacteria_Gp2	2.381	0.869	2.60e-02	-1.793	-0.061
ASV851	Elusimicrobia	0.051	0.084	2.84e-02	-0.173	0.000
ASV1322	Opitutae	0.071	-0.117	6.08e-02	0.000	0.192

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV162	Anaerolineae	0.401	0.139	6.49e-02	-0.291	0.007
ASV58	Thermoleophilia	0.681	0.470	9.21e-02	-0.942	0.097
ASV176	Acidobacteria_Gp11	0.038	0.066	1.29e-01	-0.096	0.000
ASV22	Acidobacteria_Gp3	3.475	-0.742	1.32e-01	-0.073	1.747

Class Clay

Table 7

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV98	Acidobacteria_Gp7	1.824	-0.872	2.16e-03	0.572	1.637
ASV486	Armatimonadetes_gp5	0.181	-0.211	5.00e-03	0.124	0.311
ASV44	Gemmatimonadetes	3.299	-1.243	8.66e-03	0.176	1.974
ASV46	Gammaproteobacteria	5.959	-1.721	6.49e-02	-0.069	3.481
ASV32	Deltaproteobacteria	6.857	1.426	9.31e-02	-2.908	0.182
ASV362	Acidobacteria_Gp17	0.788	0.304	9.31e-02	-0.748	0.020
ASV1715	Chloroflexia	0.023	-0.025	1.04e-01	0.000	0.071
ASV20	Nitrospira	1.689	0.481	1.32e-01	-1.079	0.060
ASV176	Acidobacteria_Gp11	0.353	0.113	1.32e-01	-0.256	0.050
ASV16	Actinobacteria	9.722	2.605	1.80e-01	-6.078	0.926

Class Combined

Table 8

Table of top indicator classes

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV44	Gemmatimonadetes	3.392	-0.285	1.69e-04	0.062
ASV98	Acidobacteria_Gp7	1.732	-0.193	1.10e-02	0.069
ASV176	Acidobacteria_Gp11	0.195	0.131	1.18e-02	0.047
ASV46	Gammaproteobacteria	4.759	-0.267	1.36e-02	0.099
ASV134	Acidobacteria_Gp2	1.242	0.205	1.68e-02	0.079
ASV49	Acidobacteria_Gp1	1.362	0.159	3.16e-02	0.069
ASV851	Elusimicrobia	0.027	0.123	3.25e-02	0.054
ASV14	Alphaproteobacteria	11.580	-0.219	3.27e-02	0.096
ASV362	Acidobacteria_Gp17	0.650	0.104	5.18e-02	0.050
ASV66	Sphingobacteriia	3.367	-0.198	6.96e-02	0.104

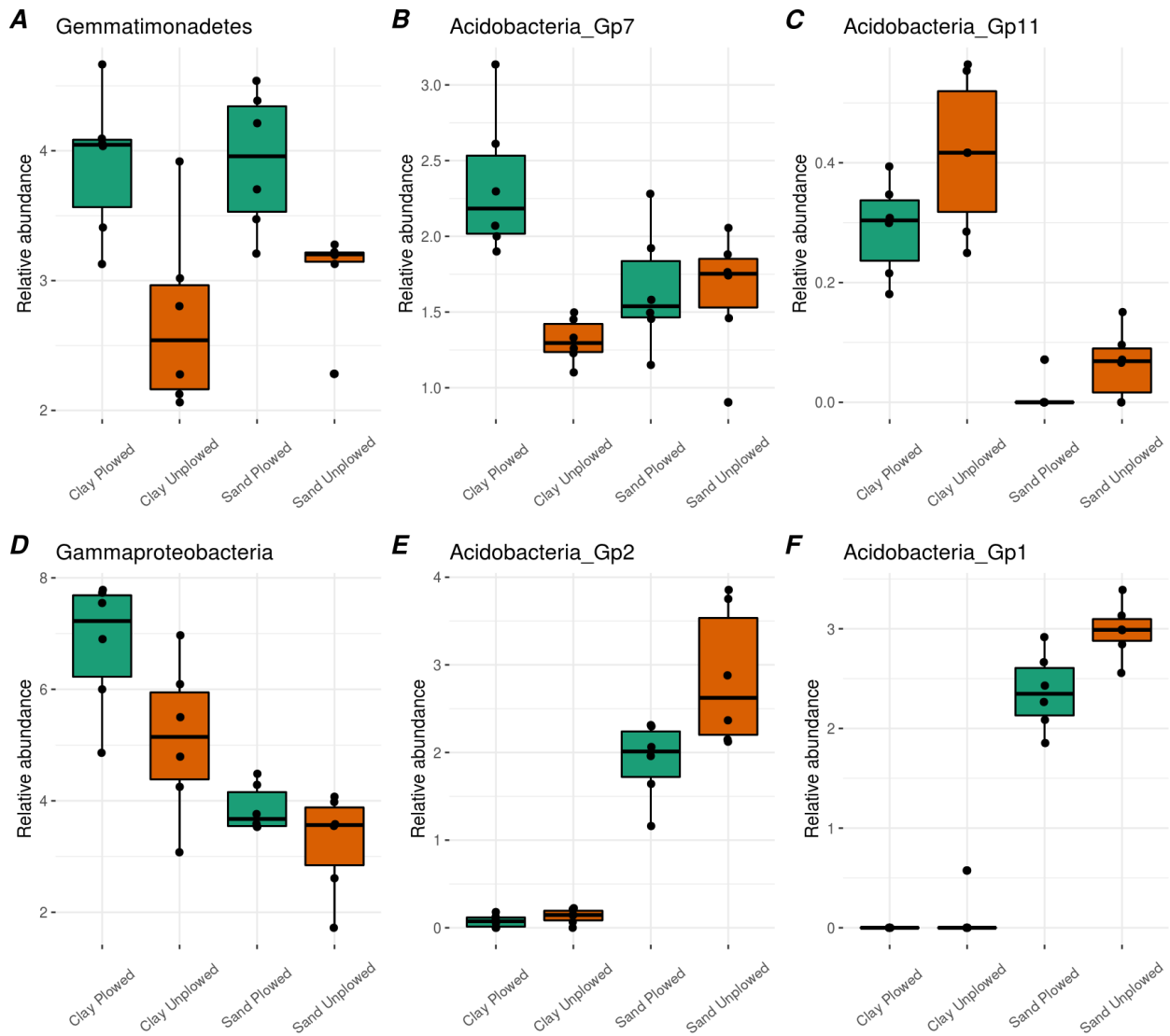


Figure 10: The top classes with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Order Sand

Table 9

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonadales	3.486	-1.000	8.66e-03	0.196	1.341
ASV49	Gp1	2.600	0.631	1.52e-02	-1.124	-0.144
ASV60	Alphaproteobacteria_incertae_sedis	0.786	-0.245	2.60e-02	0.120	0.602
ASV134	Gp2	2.381	0.869	2.60e-02	-1.793	-0.061
ASV851	Elusimicrobiales	0.051	0.084	2.84e-02	-0.173	0.000

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV14	Rhizobiales	10.612	-2.115	4.11e-02	0.006	3.845
ASV18	Burkholderiales	5.041	-1.159	4.11e-02	0.034	2.409
ASV1322	Opitutales	0.071	-0.117	6.08e-02	0.000	0.192
ASV22	Gp3	2.817	-0.734	6.49e-02	-0.235	1.380
ASV162	Anaerolineales	0.401	0.139	6.49e-02	-0.291	0.007

Order Clay

Table 10

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV86	Blastocatella	0.414	-0.345	2.16e-03	0.111	0.601
ASV98	Gp7	1.824	-0.872	2.16e-03	0.572	1.637
ASV44	Gemmatimonadales	3.299	-1.243	8.66e-03	0.176	1.974
ASV46	Xanthomonadales	3.829	-1.522	4.11e-02	0.322	2.445
ASV42	Gaiellales	2.867	0.741	9.31e-02	-2.350	0.204
ASV136	Acidimicrobiales	1.861	0.465	9.31e-02	-1.364	0.216
ASV362	Gp17	0.788	0.304	9.31e-02	-0.748	0.020
ASV20	Nitrospirales	1.689	0.481	1.32e-01	-1.079	0.060
ASV176	Gp11	0.353	0.113	1.32e-01	-0.256	0.050
ASV1802	Bdellovibrionales	0.038	0.052	1.58e-01	-0.060	0.052

Order Combined

Table 11

Table of top indicator orders

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV44	Gemmatimonadales	3.392	-0.285	1.69e-04	0.062
ASV46	Xanthomonadales	2.682	-0.280	4.50e-03	0.088
ASV98	Gp7	1.732	-0.193	1.10e-02	0.069
ASV176	Gp11	0.195	0.131	1.18e-02	0.047
ASV134	Gp2	1.242	0.205	1.68e-02	0.079

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV49	Gp1	1.314	0.146	1.92e-02	0.058
ASV136	Acidimicrobiales	1.712	0.183	2.95e-02	0.078
ASV851	Elusimicrobiales	0.027	0.123	3.25e-02	0.054
ASV362	Gp17	0.650	0.104	5.18e-02	0.050
ASV66	Sphingobacteriales	3.367	-0.198	6.96e-02	0.104

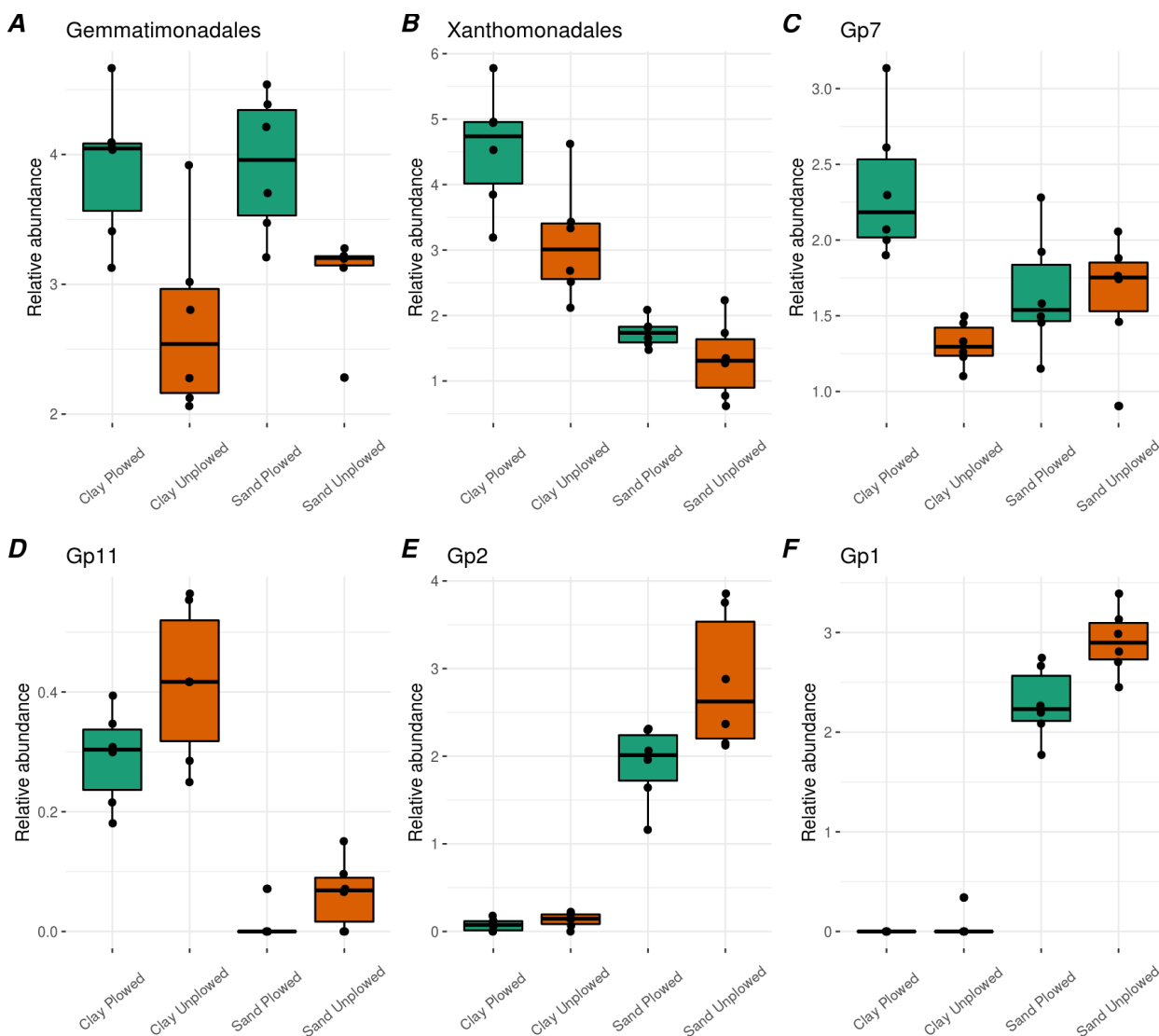


Figure 11: The top orders with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Family Sand

Table 12

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonadaceae	3.486	-1.000	8.66e-03	0.196	1.341
ASV14	Bradyrhizobiaceae	0.936	-0.427	1.52e-02	0.093	0.786

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV60	Rhizomicrobium	0.745	-0.282	2.60e-02	0.095	0.557
ASV851	Elusimicrobiaceae	0.051	0.084	2.84e-02	-0.173	0.000
ASV1265	Nakamurellaceae	0.060	0.170	2.84e-02	-0.183	0.000
ASV166	Kineosporiaceae	0.336	0.509	4.03e-02	-0.795	0.000
ASV1322	Opitutaceae	0.071	-0.117	6.08e-02	0.000	0.192
ASV72	Nocardiodaceae	0.868	-0.293	6.49e-02	-0.010	0.567
ASV43	Comamonadaceae	3.397	-0.987	6.49e-02	-0.063	1.675
ASV162	Anaerolineaceae	0.401	0.139	6.49e-02	-0.291	0.007

Family Clay

Table 13

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonadaceae	3.299	-1.243	8.66e-03	0.176	1.974
ASV118	Cystobacteraceae	0.374	-0.368	2.00e-02	0.132	0.656
ASV46	Xanthomonadaceae	3.076	-1.404	2.60e-02	0.321	2.449
ASV95	Ohtaekwangia	0.601	-0.247	6.49e-02	-0.076	0.520
ASV1115	Geobacteraceae	0.020	-0.044	7.40e-02	0.000	0.083
ASV42	Gaiellaceae	2.867	0.741	9.31e-02	-2.350	0.204
ASV99	Sphingomonadaceae	1.427	-0.583	9.31e-02	-0.070	1.164
ASV1285	Rhodothermaceae	0.032	-0.042	1.03e-01	0.000	0.074
ASV443	Geodermatophilaceae	0.195	-0.230	1.04e-01	0.000	0.477
ASV1738	Micromonosporaceae	0.087	0.129	1.29e-01	-0.246	0.000

Family Combined

Table 14

Table of top indicator families

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV44	Gemmatimonadaceae	3.392	-0.285	1.69e-04	0.062
ASV46	Xanthomonadaceae	2.237	-0.292	2.93e-03	0.087
ASV166	Kineosporiaceae	0.168	0.265	3.08e-02	0.114
ASV851	Elusimicrobiaceae	0.027	0.123	3.25e-02	0.054
ASV174	Iamiaceae	0.488	0.215	3.75e-02	0.097
ASV1285	Rhodothermaceae	0.017	-0.074	3.86e-02	0.034

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV77	Intrasporangiaceae	0.644	-0.221	5.55e-02	0.109
ASV1738	Micromonosporaceae	0.055	0.152	6.00e-02	0.076
ASV60	Rhizomicrobium	0.373	-0.102	6.52e-02	0.052
ASV95	Ohtaekwangia	0.301	-0.079	7.69e-02	0.043

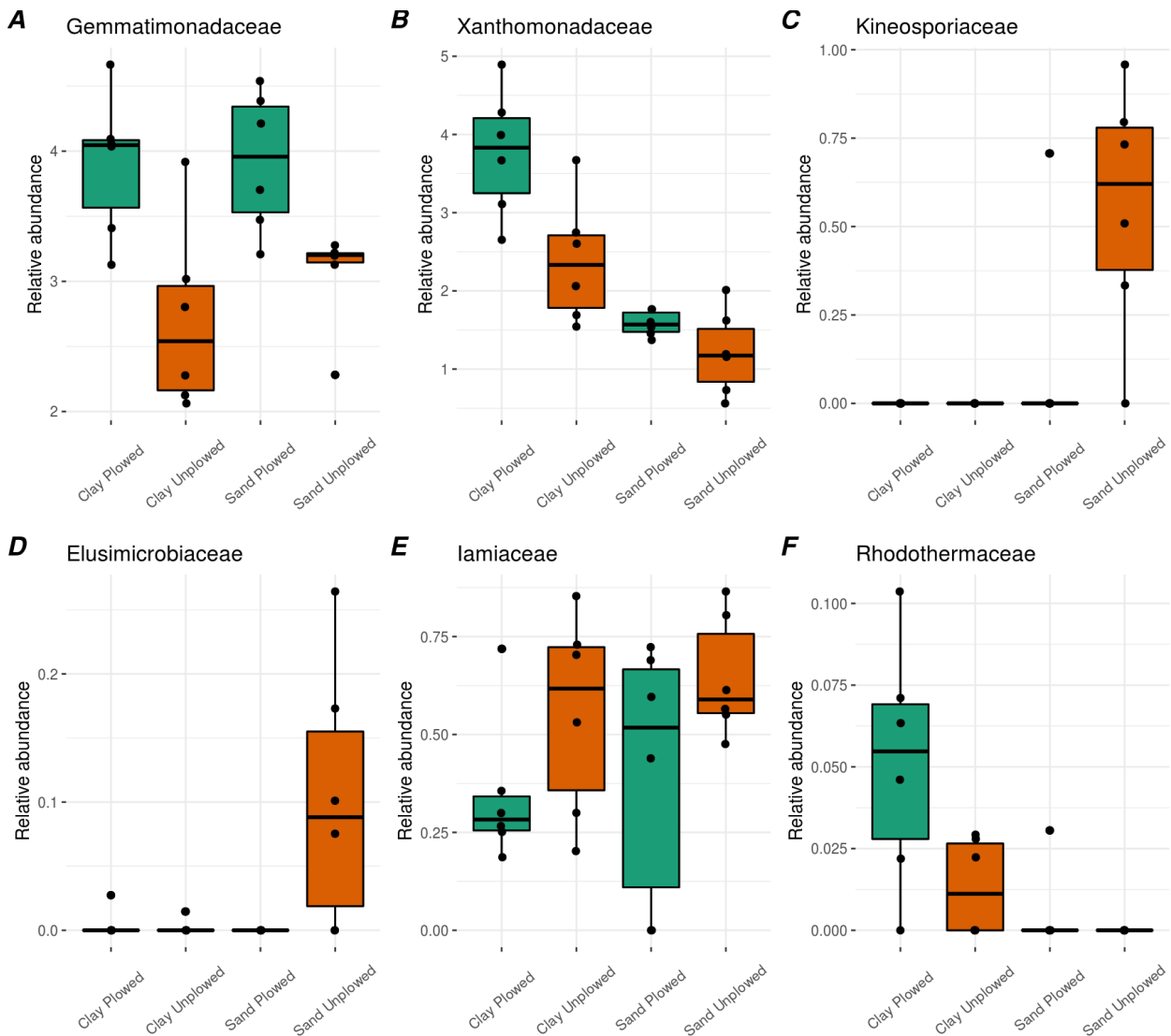


Figure 12: The top families with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Genus Sand

Table 15

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonas	3.486	-1.000	8.66e-03	0.196	1.341
ASV14	Bradyrhizobium	0.936	-0.427	1.52e-02	0.093	0.786
ASV174	Aquihabitans	0.248	0.460	2.63e-02	-0.608	0.000

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV851	Elusimicrobium	0.051	0.084	2.84e-02	-0.173	0.000
ASV1265	Nakamurella	0.060	0.170	2.84e-02	-0.183	0.000
ASV166	Angustibacter	0.320	0.509	4.03e-02	-0.795	0.000
ASV1322	Opitutus	0.071	-0.117	6.08e-02	0.000	0.192
ASV523	Craurococcus	0.103	0.099	6.37e-02	-0.172	0.052
ASV354	Microlunatus	0.017	0.040	7.40e-02	-0.065	0.000
ASV875	Variovorax	0.081	0.156	7.40e-02	-0.334	0.000

Genus Clay

Table 16

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV46	Lysobacter	1.980	-1.288	2.16e-03	0.594	2.102
ASV118	Anaeromyxobacter	0.331	-0.348	5.00e-03	0.127	0.614
ASV73	Pedomicrobium	0.425	0.374	8.13e-03	-0.637	-0.118
ASV44	Gemmatimonas	3.299	-1.243	8.66e-03	0.176	1.974
ASV937	Pontibacter	0.129	-0.153	3.41e-02	0.000	0.346
ASV664	Parasegetibacter	0.091	-0.114	6.01e-02	0.000	0.276
ASV928	Poalibacter	0.109	0.170	6.01e-02	-0.275	0.000
ASV330	Devosia	0.056	-0.123	7.40e-02	0.000	0.187
ASV640	Thermomonas	0.060	0.129	7.40e-02	-0.243	0.000
ASV1115	Geobacter	0.020	-0.044	7.40e-02	0.000	0.083

Genus Combined

Table 17

Table of top indicator genera

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV44	Gemmatimonas	3.392	-0.285	1.69e-04	0.062
ASV928	Poalibacter	0.061	0.196	9.51e-03	0.069
ASV174	Aquihabitans	0.211	0.306	1.33e-02	0.113
ASV166	Angustibacter	0.160	0.275	2.06e-02	0.110
ASV851	Elusimicrobium	0.027	0.123	3.25e-02	0.054
ASV592	Luteolibacter	0.053	0.141	3.33e-02	0.062
ASV330	Devosia	0.141	-0.195	3.98e-02	0.089

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV523	Craurococcus	0.052	0.101	4.31e-02	0.047
ASV77	Terrabacter	0.410	-0.324	4.77e-02	0.154
ASV937	Pontibacter	0.065	-0.128	5.44e-02	0.063

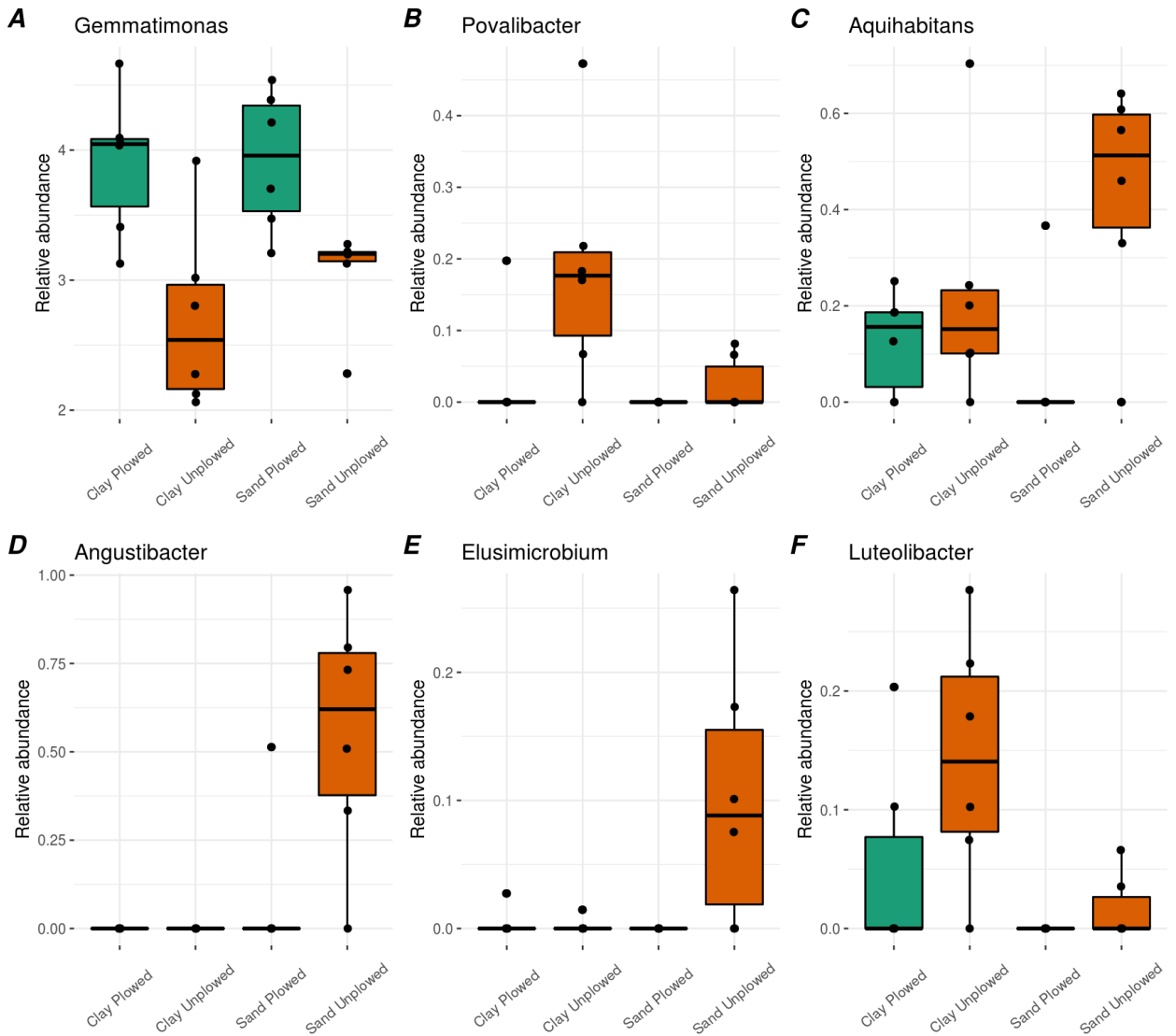


Figure 13: The top genera with deviating relative abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Absolute abundance

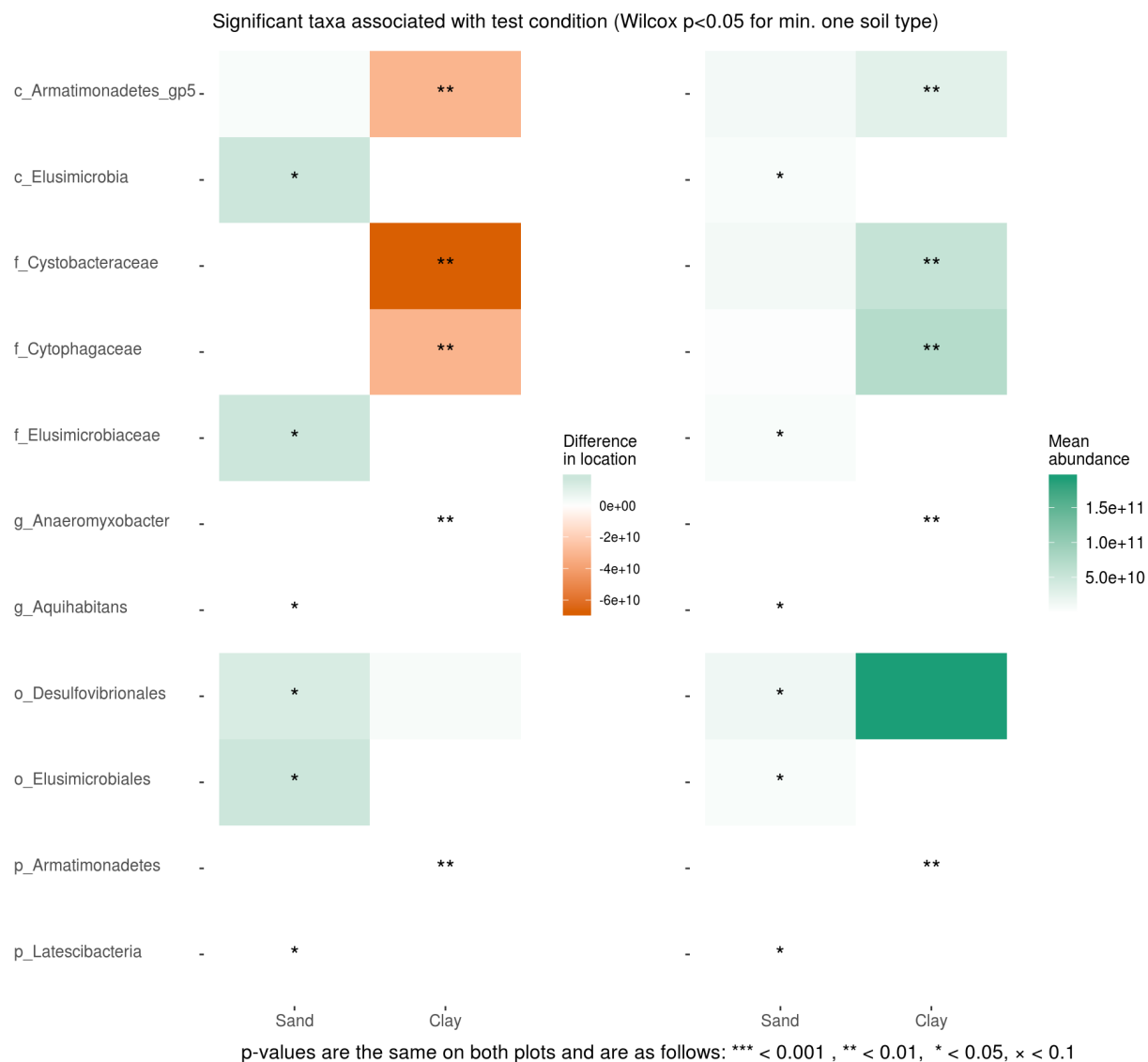


Figure 14: Overview over single taxa associations from absolute abundance tests. This plot allows an easy overview of the association patterns observed in clay and sand. Taxa are included if they had a significant ($p < 0.05$) association with plowing status in either sand or clay.

Phylum Sand

Table 18

Table of top indicator phyla

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV69	Latescibacteria	14717059.7	6863803.9	4.11e-02	-15999735	-259808.9
ASV851	Elusimicrobia	995217.3	1401401.4	8.98e-02	-2405126	0.0
ASV1918	Candidatus_Saccharibacteria	1802076.5	1236046.1	2.61e-01	-2923687	812235.6
ASV470	Planctomycetes	638277.5	279213.8	2.80e-01	-1670146	497020.0
ASV10	Acidobacteria	417672552.4	106712974.4	3.10e-01	-268093987	73093408.0

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV44	Gemmatimonadetes	48588221.6	-7288462.4	3.94e-01	-7449494	19268350.6
ASV162	Chloroflexi	7641760.6	2023098.7	4.85e-01	-5491499	2642176.9
ASV16	Actinobacteria	175984723.9	20211279.3	5.89e-01	-99701406	69288799.2
ASV486	Armatimonadetes	1339494.5	409765.4	6.25e-01	-2226862	1970071.5
ASV20	Nitrospirae	27300332.5	1664151.1	6.99e-01	-10768415	11968587.9

Phylum Clay

Table 19

Table of top indicator phyla

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV486	Armatimonadetes	4030160.28	-4252667.6	5.00e-03	2503498.3	13260657.8
ASV3240	BRC1	83844.40	138754.5	7.40e-02	-304414.0	0.0
ASV470	Planctomycetes	3237776.39	-1605411.3	1.73e-01	-889771.7	8498620.4
ASV44	Gemmatimonadetes	55339387.75	-29439612.3	1.80e-01	-15453189.9	73556753.4
ASV887	Fibrobacteres	327903.97	-119241.7	1.82e-01	0.0	1278116.1
ASV56	Bacteroidetes	116959638.42	-47011708.8	3.10e-01	-39676362.1	124059833.1
ASV1918	Candidatus_Saccharibacteria	1504503.74	-712888.6	3.78e-01	-1164416.2	2020903.9
ASV41	Verrucomicrobia	33070841.55	-6309960.5	3.94e-01	-14729324.7	44168196.0
ASV851	Elusimicrobia	330210.55	0.0	4.62e-01	0.0	3150558.9
ASV1029	Cyanobacteria/Chloroplast	71743.49	0.0	4.62e-01	0.0	341180.5

Phylum Combined

Table 20

Table of top indicator phyla

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

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV486	Armatimonadetes	2684827.4	-680.792	8.00e-02	370.123
ASV44	Gemmatimonadetes	51963804.7	-1212.068	9.08e-02	683.844
ASV56	Bacteroidetes	91635071.3	-1179.798	2.05e-01	902.943
ASV69	Latescibacteria	23063308.2	630.545	2.37e-01	518.320
ASV851	Elusimicrobia	662713.9	244.307	3.97e-01	282.537
ASV470	Planctomycetes	1938027.0	-278.839	4.02e-01	326.087
ASV41	Verrucomicrobia	25590748.3	-466.111	4.30e-01	579.687
ASV133	Firmicutes	11134390.0	317.073	4.95e-01	456.818
ASV14	Proteobacteria	485040577.7	-1265.725	4.96e-01	1824.988
ASV1029	Cyanobacteria/Chloroplast	522699.4	-110.840	6.26e-01	223.962

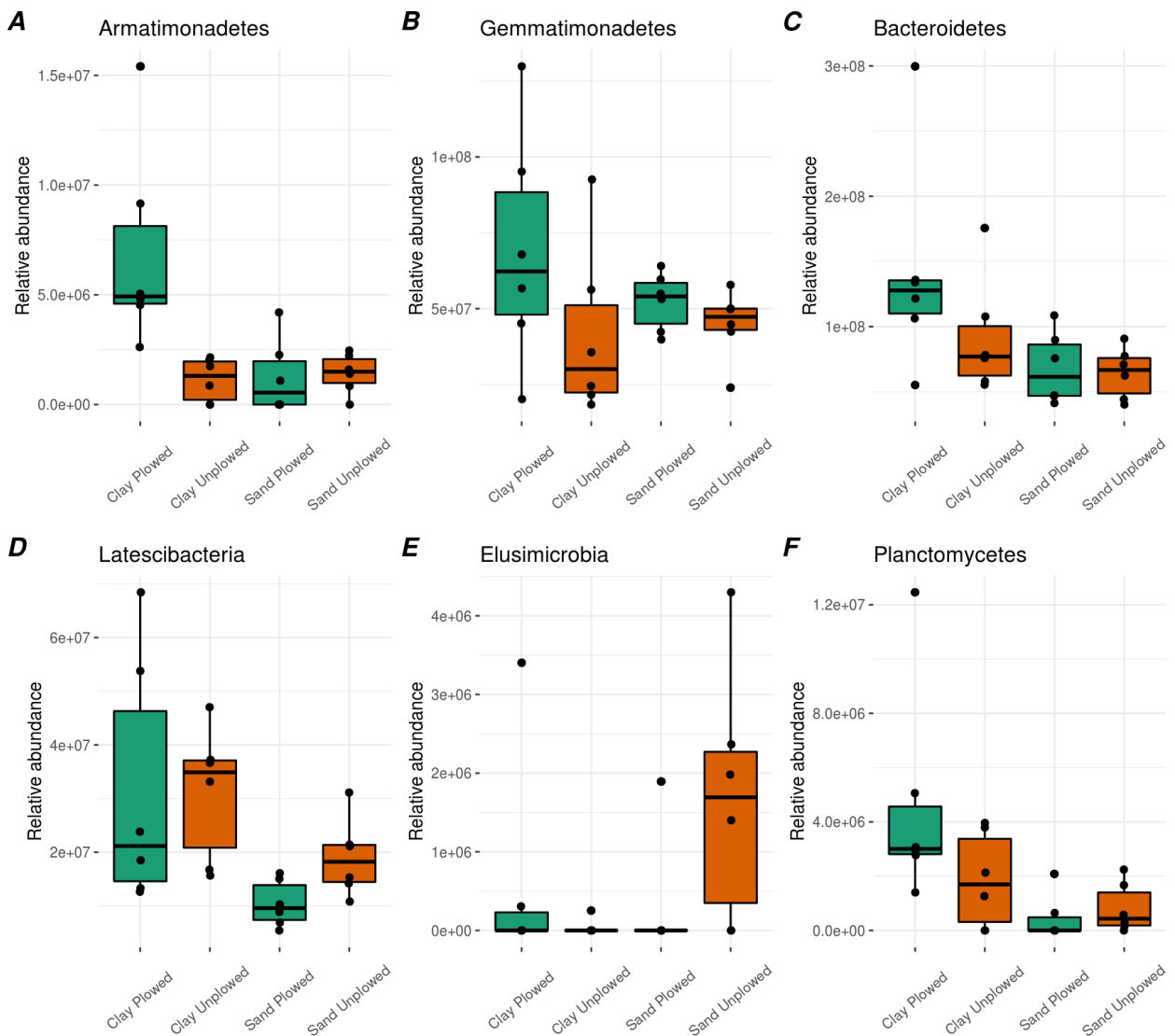


Figure 15: The top phyla with deviating absolute abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Class Sand

Table 21

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV851	Elusimicrobia	634896.8	1727187.8	2.84e-02	-1982160.6	0.0
ASV1409	Opitutae	984011.9	-1775625.5	6.08e-02	0.0	4559720.1
ASV51	Acidobacteria_Gp1	38375641.7	15044904.7	6.49e-02	-28593366.1	2710146.8
ASV134	Acidobacteria_Gp2	34260335.3	12637281.3	6.49e-02	-38670151.9	2314940.6
ASV162	Anaerolineae	5708153.5	2588925.4	6.49e-02	-5262780.0	306492.1
ASV176	Acidobacteria_Gp11	614504.4	495540.1	1.29e-01	-1775700.9	0.0
ASV58	Thermoleophilia	10361908.8	6920900.5	1.73e-01	-15866388.3	2542563.6
ASV470	Phycisphaerae	356484.2	148662.0	1.81e-01	-654205.6	0.0
ASV2009	Acidobacteria_Gp25	539799.3	228702.9	3.47e-01	-1308411.2	832619.4
ASV44	Gemmatimonadetes	48588221.6	-7288462.4	3.94e-01	-7449493.8	19268350.6

Class Clay**Table 22**

Table of top indicator classes

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV486	Armatimonadetes_gp5	2772048.8	-3131995.7	5.00e-03	1488375.155	7228916
ASV98	Acidobacteria_Gp7	33108241.9	-15016560.6	6.49e-02	-4776543.424	98717394
ASV56	Cytophagia	27413010.7	-13565531.0	1.32e-01	-5075183.485	27254679
ASV2615	Armatimonadetes_gp4	882827.5	-1147353.8	1.58e-01	-301023.480	2693315
ASV470	Phycisphaerae	2665588.8	-1553903.2	1.73e-01	-729496.620	5060241
ASV44	Gemmatimonadetes	55339387.8	-29439612.3	1.80e-01	-15453189.908	73556753
ASV887	Fibrobacteria	327904.0	-119241.7	1.82e-01	0.000	1278116
ASV1715	Chloroflexia	332981.1	-308166.4	2.00e-01	-7467.212	1222460
ASV66	Sphingobacteriia	59702725.5	-22836363.8	2.40e-01	-14044214.413	82149173
ASV537	Acidobacteria_Gp15	2584764.7	-961214.2	3.10e-01	-982552.614	8422590

Class Combined**Table 23**

Table of top indicator classes

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV851	Elusimicrobia	340745.2	453.873	2.32e-02	185.394
ASV51	Acidobacteria_Gp1	19753858.6	824.077	5.38e-02	403.294

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV134	Acidobacteria_Gp2	18125591.4	840.261	5.91e-02	420.970
ASV1409	Opitutae	1325770.7	-676.940	6.66e-02	349.864
ASV486	Armatimonadetes_gp5	1950241.5	-601.794	6.68e-02	311.266
ASV44	Gemmatimonadetes	51963804.7	-1212.068	9.08e-02	683.844
ASV1715	Chloroflexia	166490.5	-202.097	1.25e-01	126.397
ASV66	Sphingobacteriia	50621931.4	-1026.594	1.47e-01	681.591
ASV46	Gammaproteobacteria	75193453.9	-1395.064	1.68e-01	977.101
ASV98	Acidobacteria_Gp7	28163049.6	-959.413	1.74e-01	681.435

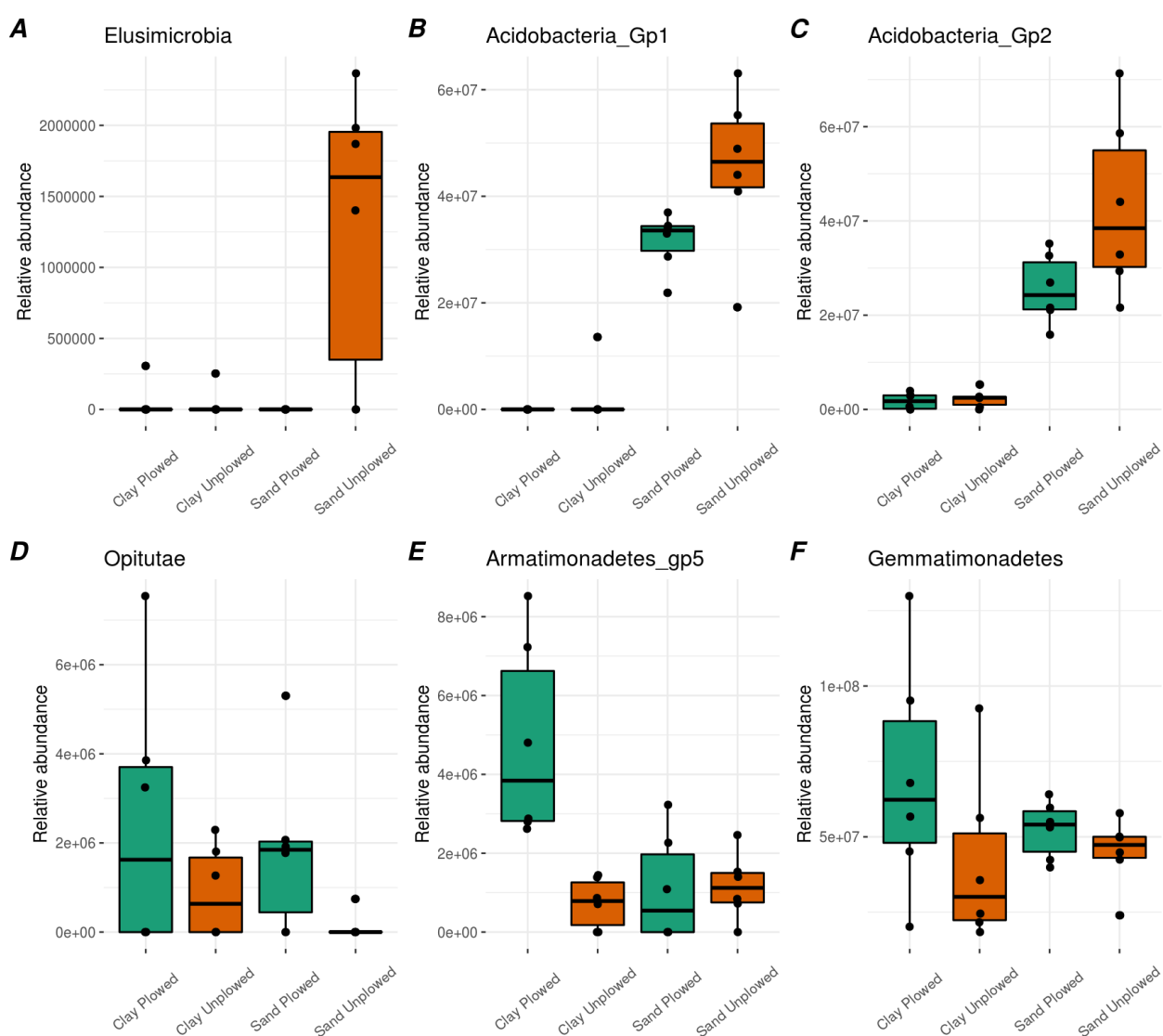


Figure 16: The top classes with deviating absolute abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Order Sand

Table 24

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV851	Elusimicrobiales	634896.8	1727187.8	2.84e-02	-1982160.6	0.00
ASV140	Desulfovibrionales	1539382.7	1343122.4	4.50e-02	-2862989.1	-65093.95
ASV1409	Opitutales	984011.9	-1775625.5	6.08e-02	0.0	4559720.09
ASV51	Gp1	37330543.1	15936890.5	6.49e-02	-30062156.4	3503011.05
ASV134	Gp2	34260335.3	12637281.3	6.49e-02	-38670151.9	2314940.63
ASV162	Anaerolineales	5708153.5	2588925.4	6.49e-02	-5262780.0	306492.10
ASV616	Solirubrobacterales	3930655.6	-3200030.8	9.31e-02	-4176322.4	4063304.41
ASV176	Gp11	614504.4	495540.1	1.29e-01	-1775700.9	0.00
ASV58	Thermoleophilales	10361908.8	6920900.5	1.73e-01	-15866388.3	2542563.55
ASV198	Pseudomonadales	3138808.7	-1986684.6	1.73e-01	-465642.3	4444444.44

Order Clay

Table 25

Table of top indicator orders

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV86	Blastocatella	7389851.8	-7521576.2	6.49e-02	-320644	15403650
ASV98	Gp7	33108241.9	-15016560.6	6.49e-02	-4776543	98717394
ASV56	Cytophagales	27413010.7	-13565531.0	1.32e-01	-5075183	27254679
ASV1078	Tepidisphaerales	911545.5	-426517.8	1.73e-01	-240962	1401401
ASV44	Gemmatimonadales	55339387.8	-29439612.3	1.80e-01	-15453190	73556753
ASV46	Xanthomonadales	65222156.3	-27030792.9	1.80e-01	-11973857	119710915
ASV887	Fibrobacterales	327904.0	-119241.7	1.82e-01	0	1278116
ASV1115	Desulfuromonadales	585102.4	-358072.7	1.82e-01	0	1927711
ASV66	Sphingobacterales	59702725.5	-22836363.8	2.40e-01	-14044214	82149173
ASV616	Solirubrobacterales	3241323.2	0.0	2.53e-01	0	12221640

Order Combined

Table 26

Table of top indicator orders

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV851	Elusimicrobiales	340745.2	453.873	2.32e-02	185.394
ASV51	Gp1	19000701.3	765.688	4.53e-02	359.704
ASV134	Gp2	18125591.4	840.261	5.91e-02	420.970
ASV1078	Tepidisphaerales	487338.4	-279.092	6.23e-02	141.746

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV1409	Opitutales	1325770.7	-676.940	6.66e-02	349.864
ASV616	Solirubrobacterales	3585989.4	-865.281	7.62e-02	463.956
ASV44	Gemmatimonadales	51963804.7	-1212.068	9.08e-02	683.844
ASV46	Xanthomonadales	43546231.2	-1363.951	1.08e-01	813.335
ASV99	Sphingomonadales	20452539.2	-732.980	1.42e-01	480.295
ASV66	Sphingobacterales	50621931.4	-1026.594	1.47e-01	681.591

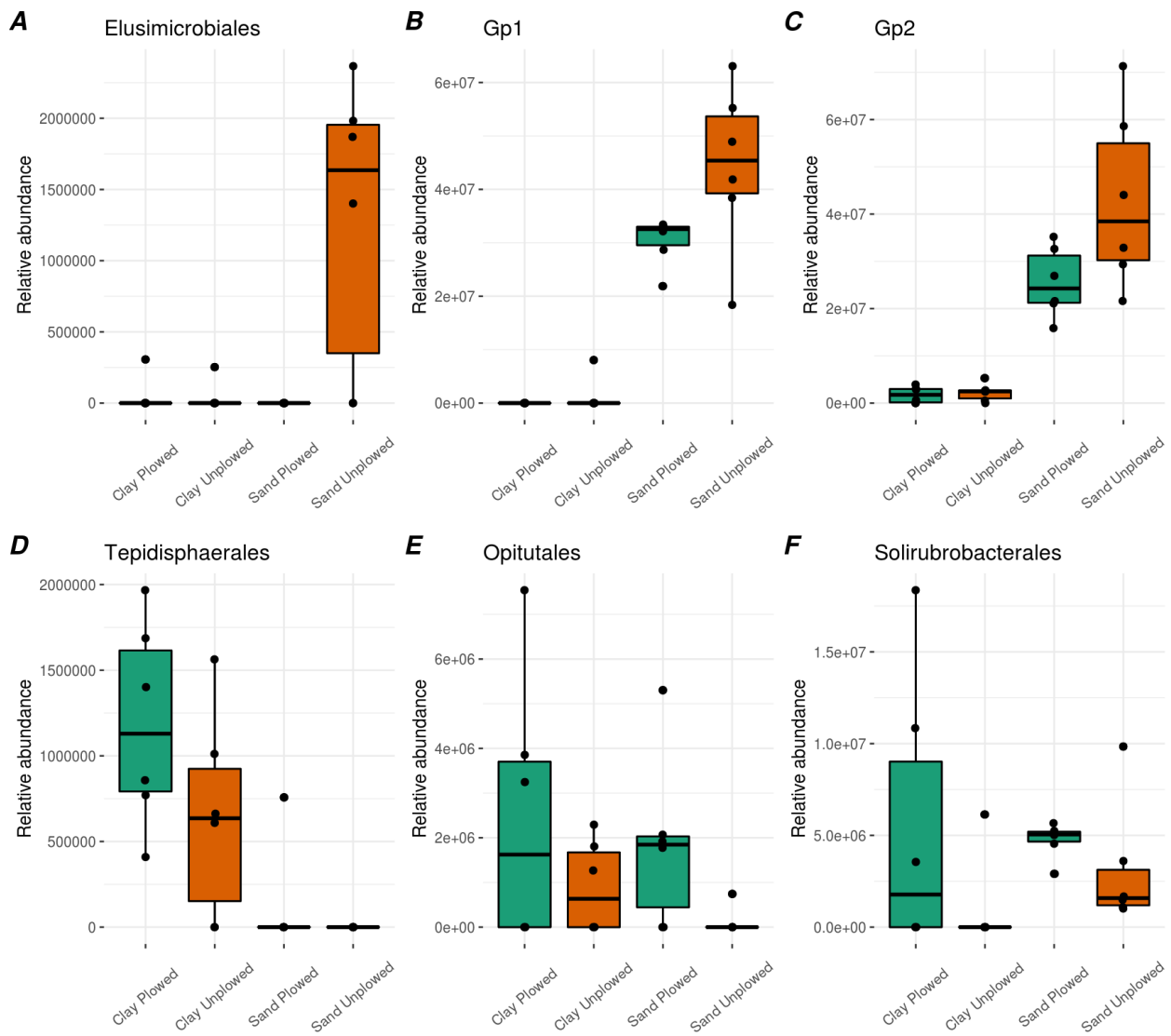


Figure 17: The top orders with deviating absolute abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Family Sand

Table 27

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV851	Elusimicrobiaceae	634896.8	1727188	2.84e-02	-1982161	0.0

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV1265	Nakamurellaceae	815137.4	1398931	2.84e-02	-2897196	0.0
ASV166	Kineosporiaceae	5247206.5	6958942	4.03e-02	-12309920	0.0
ASV1409	Opitutaceae	984011.9	-1775626	6.08e-02	0	4559720.1
ASV162	Anaerolineaceae	5708153.5	2588925	6.49e-02	-5262780	306492.1
ASV189	Desulfohalobiaceae	1418697.2	1326504	6.51e-02	-2304359	201891.3
ASV354	Propionibacteriaceae	296874.2	741972	7.40e-02	-1013025	0.0
ASV14	Bradyrhizobiaceae	13309513.5	-3434791	1.32e-01	-1682069	11770861.8
ASV72	Nocardiodaceae	12356372.1	-3119853	1.32e-01	-9770566	8182571.0
ASV58	Thermoleophilaceae	10361908.8	6920900	1.73e-01	-15866388	2542563.6

Family Clay

Table 28

Table of top indicator families

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV118	Cystobacteraceae	5562239.6	-6758119.7	8.13e-03	1550226.4	11508043
ASV559	Cytophagaceae	7094984.2	-3158015.3	8.66e-03	789428.3	8218591
ASV95	Ohtaekwangia	8770852.9	-5200047.9	4.11e-02	300368.6	9066914
ASV174	Iamiaceae	7321121.8	2324859.9	6.49e-02	-5563640.7	19935075
ASV1115	Geobacteraceae	474727.1	-810087.5	7.40e-02	0.0	1927711
ASV1285	Rhodothermaceae	654687.5	-272944.4	1.42e-01	-232744.6	2429715
ASV1548	Nannocystaceae	592646.5	-402629.8	1.58e-01	-481637.6	1232666
ASV1078	Tepidisphaeraceae	911545.5	-426517.8	1.73e-01	-240962.0	1401401
ASV44	Gemmatimonadaceae	55339387.8	-29439612.3	1.80e-01	-15453189.9	73556753
ASV46	Xanthomonadaceae	53076010.8	-25159553.9	1.80e-01	-6786416.1	109231755

Family Combined

Table 29

Table of top indicator families

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV166	Kineosporiaceae	2623603.3	1112.930	2.11e-02	446.351
ASV851	Elusimicrobiaceae	340745.2	453.873	2.32e-02	185.394
ASV95	Ohtaekwangia	4385426.4	-384.140	4.56e-02	180.734
ASV1285	Rhodothermaceae	351018.0	-338.422	5.77e-02	168.547
ASV1078	Tepidisphaeraceae	487338.4	-279.092	6.23e-02	141.746

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV1409	Opitutaceae	1325770.7	-676.940	6.66e-02	349.864
ASV46	Xanthomonadaceae	36522202.4	-1390.124	8.58e-02	771.060
ASV44	Gemmatimonadaceae	51963804.7	-1212.068	9.08e-02	683.844
ASV77	Intrasporangiaceae	9654542.5	-860.887	9.32e-02	489.488
ASV118	Cystobacteraceae	3363935.3	-711.443	1.05e-01	419.882

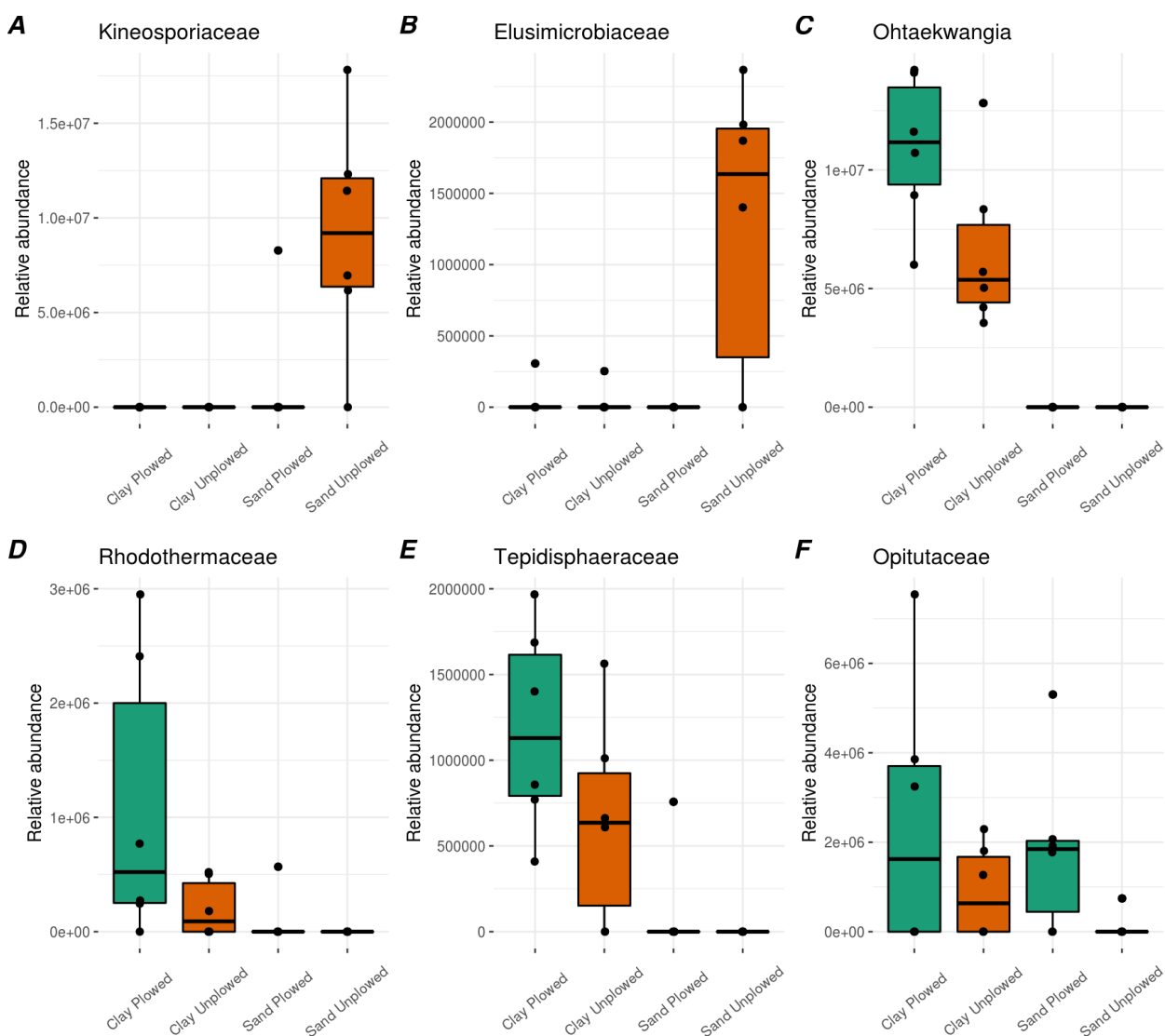


Figure 18: The top families with deviating absolute abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Genus Sand

Table 30

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV166	Angustibacter	5058390.2	6958942	1.67e-02	-12309920	0.0
ASV174	Aquihabitans	3734173.1	8504673	2.63e-02	-8827786	0.0

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV851	Elusimicrobium	634896.8	1727188	2.84e-02	-1982161	0.0
ASV1265	Nakamurella	815137.4	1398931	2.84e-02	-2897196	0.0
ASV523	Craurococcus	1463990.7	1809325	4.38e-02	-3203203	-108326.1
ASV1409	Opitutus	984011.9	-1775626	6.08e-02	0	4559720.1
ASV354	Microlunatus	296874.2	741972	7.40e-02	-1013025	0.0
ASV1585	Variovorax	1179316.1	1866378	7.40e-02	-5209841	0.0
ASV14	Bradyrhizobium	13309513.5	-3434791	1.32e-01	-1682069	11770861.8
ASV307	Pseudonocardia	1856807.2	1530967	1.58e-01	-5650800	647681.9

Genus Clay

Table 31

Table of top indicator genera

	Taxa	Mean abundance	Difference_in_location	p_value	CI_1	CI_2
ASV118	Anaeromyxobacter	5125254.6	-6060374.0	5.00e-03	1707823.8	11777223
ASV937	Pontibacter	2120138.5	-2198478.3	3.41e-02	0.0	8386212
ASV664	Parasegetibacter	1960433.7	-1386748.8	6.01e-02	0.0	7619235
ASV46	Lysobacter	34535606.9	-21068886.0	6.49e-02	-376428.3	90538105
ASV330	Devosia	677744.0	-1154905.6	7.40e-02	0.0	2021440
ASV640	Thermomonas	861315.3	1447484.9	7.40e-02	-3022407.5	0
ASV1115	Geobacter	474727.1	-810087.5	7.40e-02	0.0	1927711
ASV1802	Bacteriovorax	211268.8	429176.1	7.40e-02	-1011378.0	0
ASV3001	Enhygromyxa	325251.6	-694593.1	7.40e-02	0.0	1232666
ASV77	Terrabacter	4258994.0	-5390651.7	1.04e-01	0.0	12135590

Genus Combined

Table 32

Table of top indicator genera

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV166	Angustibacter	2529195.1	1148.359	1.50e-02	433.657
ASV851	Elusimicrobium	340745.2	453.873	2.32e-02	185.394
ASV523	Craurococcus	731995.3	428.964	2.46e-02	177.189
ASV937	Pontibacter	1060069.2	-601.275	3.61e-02	268.551
ASV77	Terrabacter	5914338.5	-1284.144	4.30e-02	596.176
ASV206	Niabella	3706588.7	-1094.428	5.44e-02	537.199

	Taxa	Mean abundance	Difference in location	p_value	SE
ASV174	Aquihabitans	3164721.8	934.040	5.92e-02	468.140
ASV330	Devosia	1994908.8	-674.452	6.03e-02	339.651
ASV1260	Povalibacter	1052118.0	611.974	6.14e-02	309.700
ASV1078	Tepidisphaera	487338.4	-279.092	6.23e-02	141.746

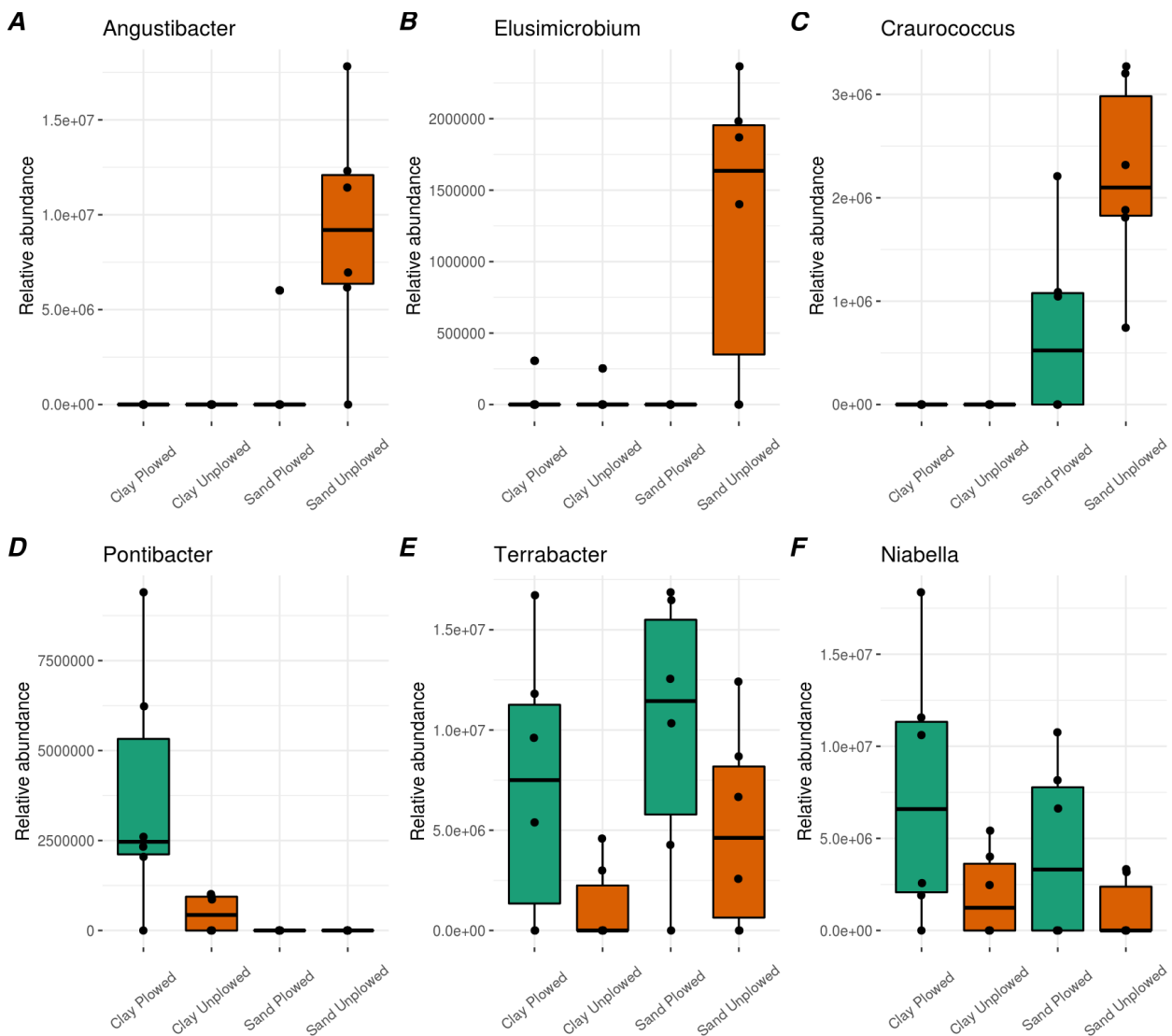


Figure 19: The top genera with deviating absolute abundance between plowed and unplowed samples. The six organisms with the lowest p-values in the linear regression were selected, as the most interesting potential indicator organisms.

Version information

Table 33: Software and package versions

	version
OS	Ubuntu 16.04.6 LTS
R	3.6.1
readxl	1.3.1
backports	1.1.5

	version
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
glue	1.3.1
gtable	0.3.0
ipred	0.9-9
zlibbioc	1.32.0

	version
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
proclim	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
cellranger	1.1.0
tools	3.6.1
cli	1.1.0
generics	0.0.2
ade4	1.7-13
devtools	2.2.1

	version
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
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

	version
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