Statistical Analysis Report

Samples from 2022

Biomcare ApS

06/01/2025

Customer	Innovation Centre for Organic Farming, Tove Mariegaard Pedersen
Customer ID	DA00206-23
Project	Regenerativt landbrug.
Sample Type	Soil
Number of samples	14 samples
Type of data	16S rRNA gene

Introduction to the biostatistical analysis

The Project

The current report describes microbiome profiles of 14 samples collected from different field or location at five different productions.

Analysis

In “Report 3”, biostatistical analyses are performed and the results presented, building on the data generated and evaluated in the 2 prior reports (**Report 1: Sequencing and data processing report, Report 2: Microbiome profiling report**).
Through biostatistical analysis we relate the microbiome profiles to the key variables selected for year 2022. The focus here is to evaluate how and to what extent the variables shape and relate to the soil microbiome composition and diversity. We therefore focus on the overall structure of the microbiome also called the microbiome composition and the diversity.

The key variables assessed in this report are summarized with summary statistics across the 14 samples in the below table.

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Rt	14	6.3	0.68	5.2	5.8	6.8	7.2
Fosfor	14	2.7	1.2	0.5	1.9	3.6	4.1
Kalium	14	11	3.1	6.7	8	13	16
Magnesium	14	5.6	2.2	2.8	3.9	6.6	9.6
Kobber	14	2.7	1.1	0.8	2.1	3	4.9
Organisk_stof	14	2.7	0.66	1.6	2.3	3.1	3.9
Lerindhold_perc	14	13	4.9	7.9	10	16	25
C.N_forhold	14	11	1.2	9	10	11	13
P.afgrøde_lager	14	341	191	60	228	440	700
Ca.plante_tilgængelig	14	364	178	115	192	454	725
Total_Ca_jordlager	14	6209	2651	3695	4288	9021	10715
Ler.humus_.CEC.	14	102	39	64	72	140	174
Ombyttelig_CEC_.perc.	14	96	5.3	84	95	100	100
Mikrobiel_biomasse	14	265	58	160	229	297	381
Mikrobiel_aktivitet	14	44	14	24	36	54	73
Svampe.bakterie_forhold	14	0.79	0.16	0.5	0.7	0.9	1.1

Table 1: Summary statistics of the key variables selected for evaluation in relation to the fields microbiome profiles.

Differences in biodiversity (alpha-diversity)

As described in **Report 2**, alpha diversity is a measure of the diversity within (or complexity within) one microbiome community (or sample). We here evaluate the one measures of alpha diversity; Shannon. The measures are introduced in **Report 2**.

Samples	Observed	Shannon	InvSimpson
R1	8930	7.86	299.47
R10	8907	7.79	271.23
R11	5815	7.47	439.75
R12	6666	7.27	162.52
R13	5378	6.75	49.31
R14	6816	7.28	128.68
R2	7467	7.73	381.86
R3	7927	7.77	366.38
R4	7549	7.31	92.03
R5	7312	7.24	80.18
R6	5934	6.70	29.78
R7	5777	6.73	35.99
R8	8862	7.19	35.74
R9	6631	7.56	391.88

Table 2: Biodiversity across samples. Table showign the biodiversity for each sample, showing 3 different biodiversity measures.

Observations and notes

See comments on biodiversity for both fungi and prokaryotic (16S) diversity in the ITS report.

Statistical assessment

A linear mixed effect model (lmer in R) was used to evaluate if the biodiversity associated significantly with each environmental variable. The mixed model was used to control for the data astructure of different farms by setting 'farm' as a random effect.

Variable	Estimate	std.err	t.value	P
Rt	**-0.3791264**	**0.130**	**-2.914**	**3.11e-02**
Fosfor	-0.0137895	0.071	-0.195	8.49e-01
Kalium	-0.0166174	0.024	-0.706	4.98e-01
Magnesium	-0.0399682	0.044	-0.912	3.81e-01
Kobber	-0.0655053	0.100	-0.658	5.23e-01
Organisk_stof	0.2408049	0.120	1.999	7.09e-02
Lerindhold_perc	-0.0680279	0.015	-4.583	7.35e-04
C.N_forhold	0.1253643	0.082	1.535	1.51e-01
P.afgrøde_lager	-0.0001099	0.001	-0.214	8.34e-01
Ca.plante_tilgængelig	-0.0001198	0.000	-0.302	7.69e-01
Total_Ca_jordlager	-0.0000703	0.000	-1.983	7.50e-02
Ler.humus	-0.0036186	0.002	-1.487	1.63e-01
Ombyttelig_CEC	-0.0309806	0.017	-1.804	9.65e-02
Mikrobiel_biomasse	0.0008999	0.001	0.699	5.01e-01
Mikrobiel_aktivitet	0.0065492	0.005	1.281	2.32e-01
Svampe.bakterie_forhold	0.4367876	0.412	1.060	3.18e-01

Table 3: Results from LMER analysis across all samples. The table shows results from LMER analyses including samples from all fields. The table shows the obtained statistical values for each of the environmental variables (rows).

Observations and notes

We see that higher Rt and percent clay associate with a lower prokaryotic diversity. There is no association between prokaryotic diversity and the 3 microbial measures from Eurofins.

Evaluation of the top 20 genera

Here is a table of the top 20 most abundant genera in the dataset. These can be inspected individually to look for any interesting patterns.

Sample	Acidibacter	Acidothermus	Bacillus	Bradyrhizobium	Bryobacter	Candidatus Nitrocosmicus	Candidatus Solibacter	Ferruginibacter	Flavobacterium
R1	0.18	0.10	0.34	0.20	0.21	1.23	0.25	0.13	0.51
R2	0.24	0.85	0.37	0.27	0.70	1.00	0.76	0.07	0.08
R3	0.40	0.07	0.45	0.68	0.30	1.80	0.45	0.50	0.81
R4	0.12	0.00	0.40	0.03	0.09	0.79	0.06	0.08	0.23
R5	0.25	0.00	0.53	0.10	0.17	1.11	0.11	0.12	0.27
R6	0.23	0.00	0.27	0.06	0.07	0.99	0.03	0.14	0.14
R7	0.28	0.01	0.19	0.11	0.10	0.88	0.04	0.16	0.13
R8	0.28	0.02	0.29	0.26	0.17	0.60	0.20	0.26	0.41
R9	0.20	0.45	0.80	0.07	0.31	1.01	0.27	0.05	1.24
R10	0.37	0.25	0.82	0.42	0.53	2.99	0.56	0.28	1.14
R11	0.13	0.34	0.66	0.03	0.20	0.50	0.19	0.04	0.71
R12	0.16	0.18	0.50	0.28	0.28	0.46	0.36	0.15	0.18
R13	0.22	0.01	0.38	0.16	0.13	0.56	0.12	0.15	0.19
R14	0.21	0.05	0.45	0.42	0.21	0.72	0.32	0.30	0.24

Table 4: Abundance of top 20 most abundant genera. The values are the abundance re-scaled to qPCR results where each sample total abundance correspond to the result of the qPCR analyses insetad of summing to 100 (where each taxa is percentage of sample community).

Observations and notes

I have also saved this table to an excel file that can be colored by taxa to help review the differences in abundance. It is likely most informative to review within farm with the known big regional differences in specific microbes abundance.

A first look at Acidothermus I see it correlates with pH level differences when reviewed within farm. That could reflecta direct relationship as: Acidibacter species are associated with acidophilic environments and play a role in breaking down organic matter under low pH conditions. This can contribute to nutrient cycling and potentially enhance soil organic content in acidic soils

Overall microbiome communities

We use the overall microbiome profiles to calculate a measure of difference in the microbiome composition between samples (beta-diversity). The calculated beta-diversity measures are used for visual inspection of the relationship between the microbiome profiles in so called ordination plots (see below), and in a statistical model named ADONIS (or PERMANOVA, see details below) to evaluate if the overall microbiome composition associates with the selected variables.

Visualization by ordination (beta-diversity)

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. In ordination plots, each sample is a point and the distance between the points increases with increasing dissimilarity in the microbiome communities.

Here we evaluate the microbiome communities using the Bray-Curtis, Aitchison and Jaccard beta-diversity measures. We use Bray-curtis for the ordination plots to visualize the inter-sample relationshipts, and all 3 measures in statistial analyses (ADONIS).

We use the different measures in combination with different microbiome profiles (taxonomic levels and normalization) as follows:

- Bray Curtis and Jaccard are computed from the absolute abundance data, at the the genus level
- Aitchison is computed from the absolute abundance data transformed with central-log-ratio (CLR), at the genus level

The Aitchison distance is a simple euclidean distance calculated using CLR transformed microbiome profiles. An analysis of CLR transformed data will reveal how the organisms behave relative to the per-sample average microbiome. Values for a microbe can therefore be negative after CLR transformation - meaning that it makes up a smaller amount of the microbiome than the average abundant microbe. This is a very different way to view the microbiome than Bray-curtis and Jaccard that uses the data as relative proportions (i.e. how big a proportion of the sample's microbiome does the individual microbe comprise). This might appear unnecessarily mathematical and unrelated to agrobiology but the CLR transformation has proved to be able to pinpoint patterns in microbiomes that are driven by environmental factors such as nutrient content or treatment applied to the samples. We therefore evaluate structures in the dataset using all three measures.

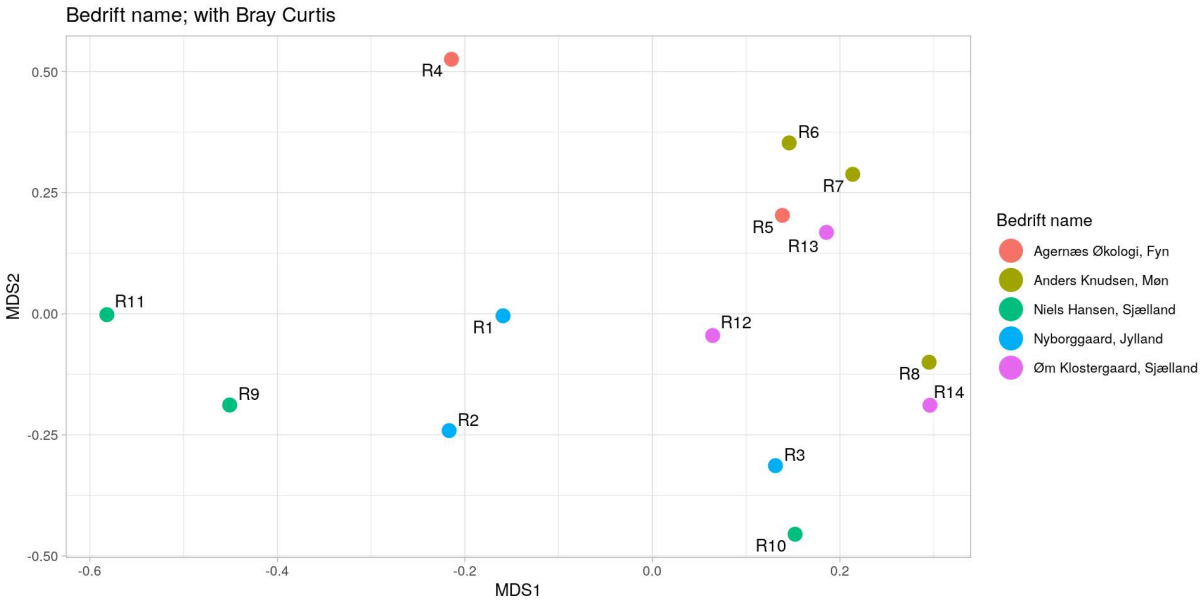

Figure 1: Visualization of structure of the bacterial community between the samples. Ordination plot using bray-curtis beta-diversity. Dots are colored by farm as seen to the right of the figure panel and each sample is named on the plot.

Observations and notes

We see some clustering by farm but much less than what we observed for fungal communities, with other factors playing a role for how the prokaryotic community differ between samples. Again is R10 very different from the other fields both from same farm and the other farms. R6 and R7 has similar communities and the 3 samples from Nyborggaard are very different from each other.

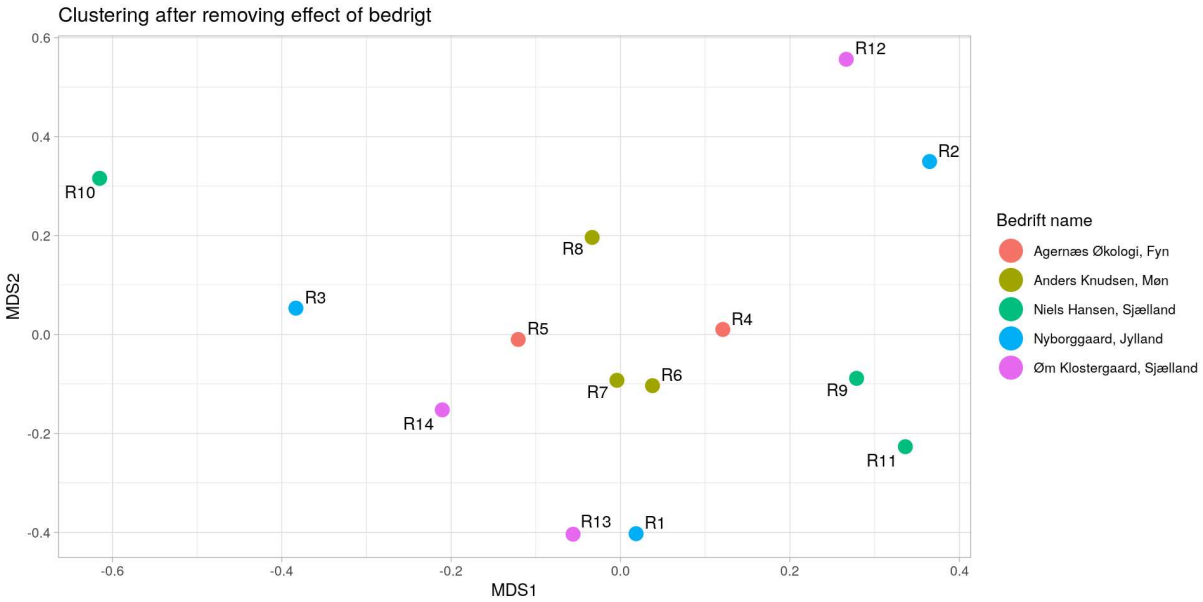

Figure 2: Visualization of structure of the bacterial community between the samples after removing the effect of farm. Ordination plot using bray-curtis beta-diversity. Dots are colored by farm as seen to the right of the figure panel and each sample is named on the plot.

Permutational Multivariate Analysis of Variance

To evaluate if the metadata variables explain a notable amount of the variation in the microbial composition, and if the amount of explained variation is statistically significant, we perform an analysis named Permutational Multivariate Analysis of Variance (ADONIS). ADONIS uses sums of squares of a multivariate dataset and is analogous to MANOVA (multivariate analysis of variance) using beta-diversity measures. It uses 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, Jaccard and Aitchison beta-diversity measures and perform the analysis at the phylum level down to the ASV level. The latter is used in amplicon sequencing in which a group of exact sequences is referred to as an amplicon sequence variant (ASV).

Each table shows results from evaluation of the effect of one variable and there is thus one table per variable.

Observations and notes

We see a strong association between prokaryotic community composition and Rt, fosfor, lerindhold, and ‘svampe/bakterie forhold’.

We see a trending association between fungal community composition and ‘mikrobiel biomasse’ and ‘Total CA jordlager’.

Rt | Fosfor | Kalium | Magnesium | Kobber | Organisk stof | Lerindhold (perc) | C/N forhold | P afgrøde lager

Ca plante tilgængelig | Total Ca jordlager | Ler/humus (CEC) | Ombyttelig CEC | Mikrobiel biomasse | Mikrobiel aktivitet

	Bray-Curtis		Jaccard		Aitchison	
Taxa level	**R2**	**p**	**R2**	**p**	**R2**	**p**
Phylum	0.533	0.092	0.1787	0.067	0.2673	0.004
Class	0.5263	0.008	0.1847	0.004	0.2834	0.004
Order	0.5467	0.003	0.2389	0.004	0.3409	0.004
Family	0.5387	0.003	0.2153	0.003	0.3132	0.004
Genus	0.5326	0.003	0.2094	0.002	0.2892	0.004
ASV	0.3043	0.003	0.1188	0.004	0.1584	0.004

Table 5: Results from ADONIS analysis. The table shows results from ADONIS analyses including samples from all farms The analysis was performed using 999 permutations constrained within farm to robustly calculate significance. The table shows the obtained R-squared values that indicate the percentage of variation that the variable could explain and the corresponding p-values.

Version information

Table 21: List of used software including the used R-programming environment packages.

Package	Version	Package	Version
OS	Ubuntu 20.04.4 LTS	**jpeg**	0.1-10
R	4.3.3	**utf8**	1.2.4
splines	4.3.3	**generics**	0.1.3
bitops	1.0-7	**robustbase**	0.99-3
lifecycle	1.0.4	**S4Arrays**	1.2.1
MASS	7.3-60.0.1	**pkgconfig**	2.0.3
insight	0.20.2	**gtable**	0.3.5
magrittr	2.0.3	**hwriter**	1.3.2.1
sass	0.4.9	**pcaPP**	2.0-4
rmarkdown	2.27	**htmltools**	0.5.8.1
jquerylib	0.1.4	**biomformat**	1.30.0
yaml	2.3.9	**png**	0.1-8
zip	2.3.1	**rstudioapi**	0.16.0
minqa	1.2.7	**tzdb**	0.4.0
ade4	1.7-22	**reshape2**	1.4.4
multcomp	1.4-26	**coda**	0.19-4.1
abind	1.4-5	**nlme**	3.1-165
zlibbioc	1.48.2	**curl**	5.2.1
Rtsne	0.17	**nloptr**	2.1.1
RCurl	1.98-1.16	**cachem**	1.1.0
TH.data	1.1-2	**zoo**	1.8-12
sandwich	3.1-0	**rhdf5**	2.46.1
GenomeInfoDbData	1.2.11	**sjlabelled**	1.2.0
svglite	2.1.3	**parallel**	4.3.3
codetools	0.2-20	**pillar**	1.9.0
DelayedArray	0.28.0	**vctrs**	0.6.5
xml2	1.3.6	**xtable**	1.8-4
tidyselect	1.2.1	**cluster**	2.1.6
farver	2.1.2	**evaluate**	0.24.0
multtest	2.58.0	**mvtnorm**	1.2-5

Package	Version	Package	Version
survival	3.7-0	**cli**	3.6.3
iterators	1.0.14	**compiler**	4.3.3
systemfonts	1.1.0	**rlang**	1.1.4
foreach	1.5.2	**crayon**	1.5.3
tools	4.3.3	**rrcov**	1.7-5
glue	1.7.0	**labeling**	0.4.3
mnormt	2.1.1	**interp**	1.1-6
SparseArray	1.2.4	**plyr**	1.8.9
xfun	0.46	**stringi**	1.8.4
mgcv	1.9-1	**viridisLite**	0.4.2
withr	3.0.0	**deldir**	2.0-4
numDeriv	2016.8-1.1	**munsell**	0.5.1
fastmap	1.2.0	**V8**	4.4.2
latticeExtra	0.6-30	**hms**	1.1.3
boot	1.3-30	**Rhdf5lib**	1.24.2
rhdf5filters	1.14.1	**highr**	0.11
fansi	1.0.6	**igraph**	2.0.3
digest	0.6.36	**RcppParallel**	5.1.8
timechange	0.3.0	**bslib**	0.7.0
R6	2.5.1	**DEoptimR**	1.1-3
estimability	1.5.1	**ape**	5.8
colorspace	2.1-0		