

Statistical Analysis Report

Functional capasity

Biomcare ApS

31/01/2025

Customer	Anton Rasmussen and Sidsel Schmidt
Customer ID	DA01102-22
Project	The effect of compost-based fertilizer on the microbial community
Sample Type	Soil
Number of samples	12 samples
Type of data	Shotgun metagenomic sequencing

Introduction to the biostatistical analysis

Objective

The project objective is to evaluate the effect of a de-gassed compost-based fertilizer on the soil microbiome. One field was included in the project, named Field D. The field was split into two blocks, and three replicates were taken from each. One block was included as a control where no fertilizer was applied and in the other block, one of the two fertilizers was added. Sampling was performed at two time points (TP), before fertilizer was applied (baseline) and 2 months after.

We aim to evaluate the effect of fertilizer treatment on the soil microbiome.

To support the analyses, we have named the two blocks per field as G1 and G2, indicating the block that remains untreated and the block that is treated between TP1 and TP2, respectively. These abbreviations, as well as the notation TP for time point, are used in illustrations.

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: Sequencing and data processing report, Report 2: Microbiome profiling report for functions**).

Through biostatistical analysis we compared the microbiome profiles between the different groups of samples. The focus here is to evaluate the effect of treatment and to use the untreated blocks as controls for the changes happening due to other factors.

From among the different types pf functional profiles we have generated as presented in R2, we use KEGG (or KO) ortologous genes in this report. We do so because this data layer holds information on orthologous genes (taking the large Uniref50 profiles of individual genes where one gene ID represent a gene in one organism, and groups the genes by function across organisms) and thus allow us to review functions at a detailed level without the sparsity of the Uniref data layer. Alternatively we can use the GO processes where each process represent the joined function performed by many genes. However the GO processes require a lot of interpretation to understand the function they represent and leave out some detailed infromation gained grom gene groups.

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

 The report initiates with a visual evaluation of the overall KEGG based functional capasity using ordination plots and a statistical analysis to evaluate how much of the compositional variation in the microbiome is explained by the variable(s) of interest and if the compositional differences are statistically significant (using Permutational Multivariate Analysis of Variance (ADONIS)).

- 2. If changes are seen in the overall community composition, it can be due to subtle changes in many genes or more pronounced changes to few KO genes If the latter is causing the observed shifts in the overall microbiome, we can identify the specific genes that have increased or decreased in proportion due to the variable of interest. The analysis thereby allows us to identify indicator genes that may be of interest for further analyses and studies.
- Lastly, we investigate the alpha diversity, which is a measure of the within-sample diversity, often referred to as biodiversity, and we evaluate whether the alpha-diversity differs between groups using two different measures of alphadiversity.

Differences in overall community functional capasity

Visualization by ordination (beta-diversity)

As described in **Report 2**, beta-diversity is a measure of how similar or dissimilar the KEGG based functional capasity 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 and euclidean beta-diversity measures. If not all plots are shown in this report, you can find them in the project folder. We use a combination of measures as each measure highlights different properties of the microbiome. See more details in Report 2. We use the different measures in combination with different microbiome profiles (taxonomic levels and normalization) as follows:

· we use Bray Curtis and euclidean distance calculated from the relative abundance data

In the plot(s), samples are colored by treatment status and time point. When samples are split into one panel per (field)plot, coloring is only used to identify TP.

Bray-Curtis

Euclidean distance

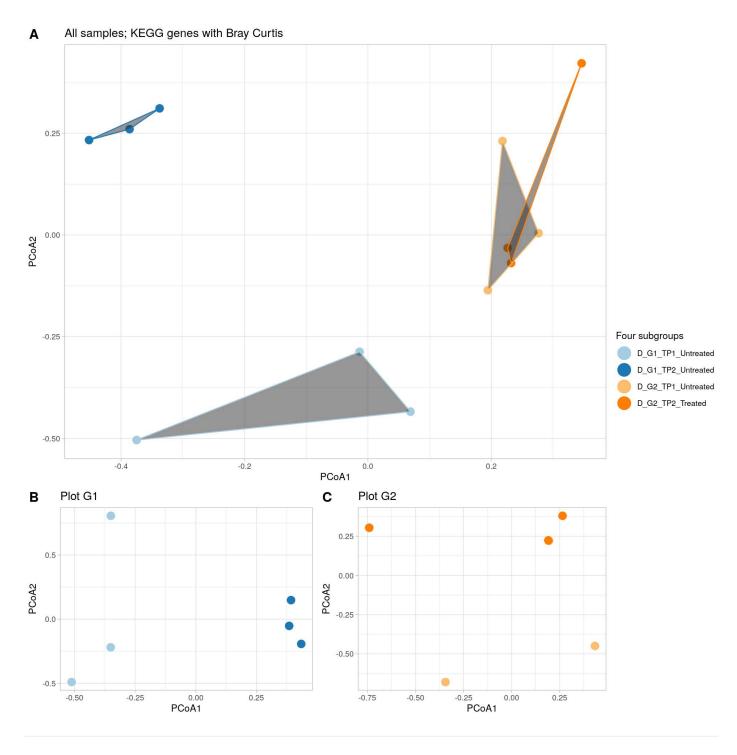


Figure 1: Visualization of structure of the KEGG based functional capasity between the samples. Ordination plots using Bray-Curtis for relative abundance transformed genes and a NMDS ordination method. A) shows community composition for all samples across plots and TP, while samples have been subset to individual plots in B) and C). Dots are colored by treatment status, TP and block as indicated. In A) we added "convex hull shapes" to guide the eye to where the samples of each group are located.

Permutational Multivariate Analysis of Variance

To evaluate if the compositional differences evaluated above using ordination plots explain much of the microbial variation and if they are statistically significant, we perform an analysis named Permutational Multivariate Analysis of Variance (ADONIS). ADONIS partitions sums of squares of a multivariate dataset 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 at the species up to the phylum level.

After evaluating the ADONIS analysis in detail I have decided to not include the results, as the analysis is unstable due to low number of samples (a simple repeated run and comparison gives very fluctuating statistics). Therefore, we must judge by visual inspection the effect of treatment on the microbiome overall composition. I leave the explanation of the analysis model in the report, to show what we can do to add statistical insight to the ordination plots, when we have the power to do so.

Differential abundance of single genes

We now move from the evaluation of overall functional capasity, to evaluate if the abundance of specific genes are affected by the treatment. This analysis provides the first insight into potential indicator functions and a first peek at the single functions that drive compositional differences between the groups of samples.

Due to the small number of samples per group (n=3) we have chosen to analyse a subset of selected genes with functions of interest to the soil microbiome. The list of selected genes include 112 genes from the following functional categories:

- · Antibiotic Production"
- Biocontrol
 - Bactericidal Agents
 - · Fungicidal Agents"
 - Insecticidal Agents"
 - · Nematicidal Agents"
- · Carbon fixation
- Denitrification
- Hormone Production
 - ABA
 - Brassinosteroids
 - CK
 - Ethylene
 - GA
 - IAA
- Nitrogen fixation
- · Nitrogen release
- Nodulation
- Nutrition
 - Ca
 - Cu
 - Fe
 - MgMn
 - S
 - Zn
- · Phosphorus solubilization
- Potassium solubilization
- Stress Adaptations
 - · Abscisic Acid (ABA) Production
 - ACC Deaminase (ACC-d) Activity
 - Exopolysaccharide Production
 - Heavy Metal Solubilization
 - Salicylic Acid (SA) Production
 - Salt Tolerance
 - Siderophore production

From the 112 genes we identify those detected in the current samples and analyse those using the described approach. A overview table of all those genes tested are found below.

A number of statistical models are available for such analysis, and a model must be selected considering the study design and power available. We selected a linear regression with a so-called "change-score model" as the samples are collected from two separate blocks and therefore the groups are not composed of samples taken from the same population at baseline. The model therefore evaluates if the difference in abundance between the two time-points is significantly different between G1 and G2; meaning is the difference that occur over time different when the treatment is applied.

Below you find a table for the analysis of field D. The table provide summary statistics for the analysed genes and results from the statistical analyses. A figure of boxplots is made for the most associated genes (P<0.2). These plots allow for visual inspection of the genes that show significant (p<0.05) or a trending association. Due to the low number of samples per group, we have a low power for obtaining a significant p-value and therefore, it is relevant to also inspect associations that show interesting trends.

	Group	Gene.name	KEGG	Details
1	Antibiotic Production	act	K00626	Involved in the biosynthesis of Actinorhodin, a polyketide antibiotic produced by Streptomyces coelicolor.
3	Antibiotic Production	mupA	K00799	Involved in the biosynthesis of Mupirocin, an antibiotic effective against Gram-positive bacteria, produced by Pseudomonas fluorescens.
5	Antibiotic Production	bmd	K21310	Involved in the biosynthesis of Bacillomycin D, a lipopeptide antibiotic produced by Bacillus subtilis.
7	Biocontrol - Bactericidal Agents	srfAA	K00822	Surfactin synthetase subunit, contributes to antibacterial activity
8	Biocontrol - Fungicidal Agents	prnA	K04100	Pyrrolnitrin biosynthesis protein, involved in antifungal compound production
14	Carbon fixation	rbcL	K01601	Large subunit of RuBisCO, catalyzes CO ₂ fixation in the CBE cycle.
15	Carbon fixation	rbcS	K01602	Small subunit of RuBisCO, essential for enzyme assembly and function.
17	Carbon fixation	frdA	K00244	Fumarate reductase flavoprotein subunit, reduces fumarate to succinate.
19	Carbon fixation	mct	K01847	Methylmalonyl-CoA mutase, rearranges methylmalonyl-CoA to succinyl-CoA.
21	Carbon fixation	fhs	K01938	Formate-tetrahydrofolate ligase, assimilates formate into the reductive acetyl-CoA pathway.
22	Denitrification	narG	K00370	narG, narH, narl: Subunits of membrane-bound nitrate reductase, involved in the reduction of nitrate to nitrite.
23	Denitrification	narH	K00371	narG, narH, narl: Subunits of membrane-bound nitrate reductase, involved in the reduction of nitrate to nitrite.
24	Denitrification	narl	K00374	narG, narH, narI: Subunits of membrane-bound nitrate reductase, involved in the reduction of nitrate to nitrite.
25	Denitrification	napA	K02567	napA: Catalytic subunit of periplasmic nitrate reductase, facilitating nitrate reduction in the periplasm.
26	Denitrification	nirK	K00368	nirK and nirS: Enzymes responsible for the reduction of nitrite to nitric oxide.
27	Denitrification	nirS	K15864	nirK and nirS: Enzymes responsible for the reduction of nitrite to nitric oxide.
28	Denitrification	norB	K04561	norB and norC: Subunits of nitric oxide reductase, catalyzing the reduction of nitric oxide to nitrous oxide.

	Group	Gene.name	KEGG	Details
29	Denitrification	norC	K02305	norB and norC: Subunits of nitric oxide reductase, catalyzing the reduction of nitric oxide to nitrous oxide.
30	Denitrification	nosZ	K00376	nosZ: Catalyzes the final step of denitrification, reducing nitrous oxide to nitrogen gas.
31	Denitrification	nosR	K19339	nosR: Regulates the expression of nosZ, ensuring the proper reduction of nitrous oxide.
32	Hormone Production - ABA	nced	K09809	9-cis-Epoxycarotenoid dioxygenase, key enzyme in ABA biosynthesis
34	Hormone Production - CK	ipt	K13789	Isopentenyltransferase, key enzyme in cytokinin biosynthesis
38	Hormone Production - GA	ks	K01768	ent-Kaurene synthase, involved in early steps of GA biosynthesis
40	Hormone Production - IAA	ami1	K01426	Enzyme converting indole-3-acetamide to IAA
42	Nitrogen fixation	nifD	K02586	Nitrogenase molybdenum-iron protein alpha chain
44	Nitrogen fixation	nifE	K02587	Assembly of the nitrogenase MoFe-cofactor
45	Nitrogen fixation	nifN	K02592	Assembly of the nitrogenase MoFe-cofactor
46	Nitrogen fixation	nifB	K02585	Synthesis of the iron-molybdenum cofactor
47	Nitrogen fixation	nifV	K01951	Homocitrate synthesis
48	Nitrogen fixation	nifQ	K13371	Molybdenum incorporation into FeMo-cofactor
52	Nitrogen fixation	nifS	K04487	Biosynthesis of iron-sulfur clusters
53	Nitrogen fixation	nifU	K04488	Assembly of iron-sulfur clusters
55	Nitrogen fixation	nifA	K02584	Nif-specific regulatory protein
56	Nitrogen fixation	nifT	K02593	nitrogen fixation protein
57	Nitrogen release	narG, narZ, nxrA	K00370	narG: Encodes the alpha subunit of nitrate reductase, converting nitrate to nitrite.
58	Nitrogen release	nirS	K15864	nirS/nirK: Encode nitrite reductases that convert nitrite to nitric oxide.
59	Nitrogen release	nirK	K00368	nirS/nirK: Encode nitrite reductases that convert nitrite to nitric oxide.
60	Nitrogen release	norB	K04561	norB: Encodes nitric oxide reductase, converting nitric oxide to nitrous oxide.
61	Nitrogen release	nosZ	K00376	nosZ: Encodes nitrous oxide reductase, converting nitrous oxide to nitrogen gas.
72	Nutrition - Fe	feoB	K04758	Ferrous iron transport protein B, involved in ferrous iron uptake
73	Nutrition - Fe	fepC	K02014	Ferric enterobactin transport system ATP-binding protein, involved in ferric iron uptake

	Group	Gene.name	KEGG	Details
76	Nutrition - Mn	mntH	K03386	Proton-dependent manganese transporter
77	Nutrition - Mn	mntA	K02010	ABC-type manganese transport system ATP-binding protein
78	Nutrition - S	cysT	K02047	Sulfate transport system permease protein
79	Nutrition - S	cysA	K02042	ABC-type sulfate transport system ATP-binding protein
81	Nutrition - Zn	zitB	K07716	Low-affinity zinc transporter involved in zinc efflux
82	Phosphorus solubilization	gcd	K00111	gcd: Encodes glucose dehydrogenase, which is involved in the direct oxidation pathway and plays a significant role in inorganic phosphate solubilization.
83	Phosphorus solubilization	ppa	K01507	ppa: Encodes inorganic pyrophosphatase, which hydrolyzes pyrophosphate to release phosphate ions.
85	Phosphorus solubilization	pqqC	K06133	pqqC: Part of the pyrroloquinoline quinone (PQQ) biosynthesis gene cluster, essential for the activity of glucose dehydrogenase.
86	Phosphorus solubilization	phoA	K01077	phoA: Encodes alkaline phosphatase, which hydrolyzes organic phosphate compounds to release inorganic phosphate.
87	Phosphorus solubilization	phoD	K01113	phoD: Encodes alkaline phosphatase D, another enzyme involved in the mineralization of organic phosphorus.
88	Phosphorus solubilization	pqqE	K06134	pqqE: Another gene in the PQQ biosynthesis pathway, contributing to the production of cofactors necessary for glucose dehydrogenase activity.
89	Potassium solubilization	ackA	K00925	Encodes acetate kinase, involved in the production of acetic acid, which can solubilize potassium minerals.
90	Potassium solubilization	epsB	K01991	Encodes a protein involved in exopolysaccharide biosynthesis, contributing to biofilm formation and mineral solubilization.
91	Potassium solubilization	gltA	K01647	Encodes citrate synthase, a key enzyme in the citric acid cycle, leading to the production of citric acid, which aids in potassium solubilization.
92	Potassium solubilization	mdh	K00024	Encodes malate dehydrogenase, involved in the production of malic acid, another organic acid that can solubilize potassium.
93	Potassium solubilization	ррс	K01595	Encodes phosphoenolpyruvate carboxylase, which plays a role in the synthesis of organic acids like oxaloacetate and malate, contributing to potassium solubilization.
95	Stress Adaptations - Abscisic Acid (ABA) Production	aba1	K09818	Zeaxanthin epoxidase, involved in the early steps of ABA biosynthesis under stress conditions.
96	Stress Adaptations - ACC Deaminase (ACC-d) Activity	acdS	K01505	Encodes ACC deaminase, facilitating plant growth by breaking down ACC, the precursor to ethylene.

	Group	Gene.name	KEGG	Details
98	Stress Adaptations - Exopolysaccharide Production	epsB	K01991	Involved in EPS polymerization and export, contributing to microbial adhesion and protection.
99	Stress Adaptations - Exopolysaccharide Production	galE	K01785	Encodes UDP-galactose 4-epimerase, involved in the production of sugar nucleotides for EPS synthesis.
103	Stress Adaptations - Salicylic Acid (SA) Production	icsA	K14170	Isochorismate synthase, involved in the biosynthesis of salicylic acid in bacterial plant interactions.
107	Stress Adaptations - Salt Tolerance	ktrB	K03320	Part of the potassium transporter system, helps in maintaining ionic balance under salt stress.
109	Stress Adaptations - Siderophore production	dhbF	K02492	Non-ribosomal peptide synthetase for Bacillibactin production
110	Stress Adaptations - Siderophore production	entB	K02518	Participates in Enterobactin biosynthesis, a catecholate siderophore
111	Stress Adaptations - Siderophore production	fhuF	K03711	Associated with Ferrichrome utilization, an iron-binding siderophore

Table 2: Table of the genes included in the statistical analyses. The table include those genes from among the full list of genes of interest that is found in the current dataset.

Group	Gene name	KEGG ID	Details	Beta	р
Phosphorus solubilization	gcd	K00111	gcd: Encodes glucose dehydrogenase, which is involved in the direct oxidation pathway and plays a significant role in inorganic phosphate solubilization.	0.00016968	0.07070583
Denitrification	narG	K00370	narG, narH, narl: Subunits of membrane-bound nitrate reductase, involved in the reduction of nitrate to nitrite.	-7.186e-05	0.18608956
Denitrification	nosZ	K00376	nosZ: Catalyzes the final step of denitrification, reducing nitrous oxide to nitrogen gas.	0.00014397	0.02017576
Phosphorus solubilization	ppa	K01507	ppa: Encodes inorganic pyrophosphatase, which hydrolyzes pyrophosphate to release phosphate ions.	0.00043267	0.13948496
Potassium solubilization	ppc	K01595	Encodes phosphoenolpyruvate carboxylase, which plays a role in the synthesis of organic acids like oxaloacetate and malate, contributing to potassium solubilization.	-9.596e - 05	0.03636362
Carbon fixation	rbcL	K01601	Large subunit of RuBisCO, catalyzes CO ₂ fixation in the CBB cycle.	5.897e - 05	0.13416088
Carbon fixation	rbcS	K01602	Small subunit of RuBisCO, essential for enzyme assembly and function.	-0.00021062	0.03527401
Hormone Production - GA	ks	K01768	ent-Kaurene synthase, involved in early steps of GA biosynthesis	0.0002719	0.05828934

Group	Gene name	KEGG ID	Details	Beta	p
Carbon fixation	mct	K01847	Methylmalonyl-CoA mutase, rearranges methylmalonyl-CoA to succinyl-CoA.	0.00010607	0.14853472
Nitrogen fixation	nifV	K01951	Homocitrate synthesis	0.00030952	0.03920911
Nutrition - Fe	fepC	K02014	Ferric enterobactin transport system ATP-binding protein, involved in ferric iron uptake	0.00019332	0.14685172
Stress Adaptations - Siderophore production	dhbF	K02492	Non-ribosomal peptide synthetase for Bacillibactin production	7.644e-05	0.08297649
Nitrogen fixation	nifE	K02587	Assembly of the nitrogenase MoFe-cofactor	4.53e-06	0.11759807
Stress Adaptations - Salt Tolerance	ktrB	K03320	Part of the potassium transporter system, helps in maintaining ionic balance under salt stress.	0.00024681	0.13448802
Nitrogen fixation	nifU	K04488	Assembly of iron-sulfur clusters	-0.00041679	0.08862906
Nutrition - Zn	zitB	K07716	Low-affinity zinc transporter involved in zinc efflux	-6.833e-05	0.03085359
Denitrification	nirS	K15864	nirK and nirS: Enzymes responsible for the reduction of nitrite to nitric oxide.	2.286e-05	0.00358355
Antibiotic Production	bmd	K21310	Involved in the biosynthesis of Bacillomycin D, a lipopeptide antibiotic produced by Bacillus subtilis.	8.096e-05	0.04316323

Table 3: Summary table statistical results for analysed genes. The table shows all genes with a p-value < 0.2 and color those qith p-value<0.1. For the statistical analysis, estimates or beta-values are given for the difference between the groups and the corresponding p-values.

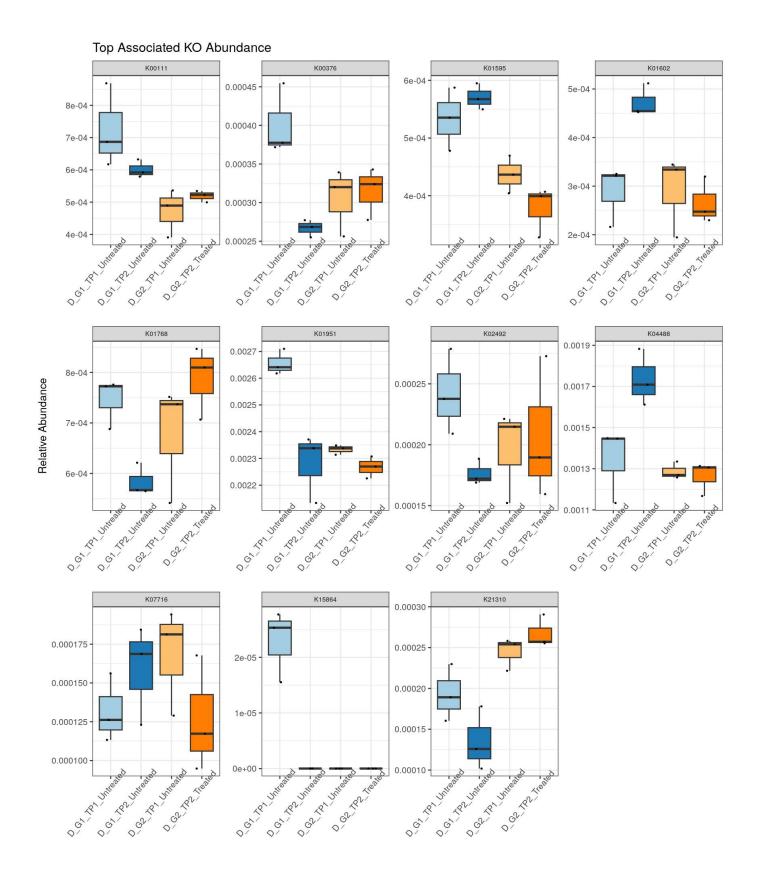


Figure 3: Differential abundant genes. Top differential abundant genes are shown selected at a p-value<0.1.

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 here evaluate the two measures; Shannon and observed computed for KEGG genes. The measures are introduced in Report 2. We start with inspecting the summary statistics of the measures in the different subgroups of samples.

Group	n	median	mean	sd	min	max
D_G1_TP1_Untreated	3	7.448	7.455	0.019	7.441	7.477
D_G1_TP2_Untreated	3	7.417	7.416	0.009	7.407	7.424
D_G2_TP1_Untreated	3	7.429	7.421	0.013	7.406	7.429
D_G2_TP2_Treated	3	7.434	7.420	0.024	7.392	7.434

Table 4: Summary statistics of alpha diversity measure Shannon The table shows the different subgroups of samples, the number of samples in the group (n), the median and mean level in the group, the group standard deviation (sd) and minimum (min) and maximum (max) values.

Group	n	median	mean	sd	min	max
D_G1_TP1_Untreated	3	4225	4288.333	158.774	4171	4469
D_G1_TP2_Untreated	3	4107	4097.333	39.400	4054	4131
D_G2_TP1_Untreated	3	4119	4070.000	93.664	3962	4129
D_G2_TP2_Treated	3	4157	4074.000	169.496	3879	4186

Table 5: Summary statistics of alpha diversity measure Observed. The table shows the different subgroups of samples, the number of samples in the group (n), the median and mean level in the group, the group standard deviation (sd) and minimum (min) and maximum (max) values.

As for the single genes, we use a linear regression with a so-called "change-score model" for the analysis of alpha diversity. The table below shows the effect of treatment within field D.

	Shannon				Observed			
	Estimate	Std. Error	t value	Pr(> t)	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.04	0.014	-2.729	5.25e - 02	-191	106.835	-1.788	1.48e-01
TreatmentTreated	0.038	0.021	1.862	1.36e-01	195	151.087	1.291	2.66e-01

Table 6: Results from analysis of alpha diversity. The effect of treatment on alpha diversity within field D was analysed using a linear regression designed as a "change-score model". Columns 1-4 gives results for Shannon diversity and columns 5-8 gives results for Observed (richness).

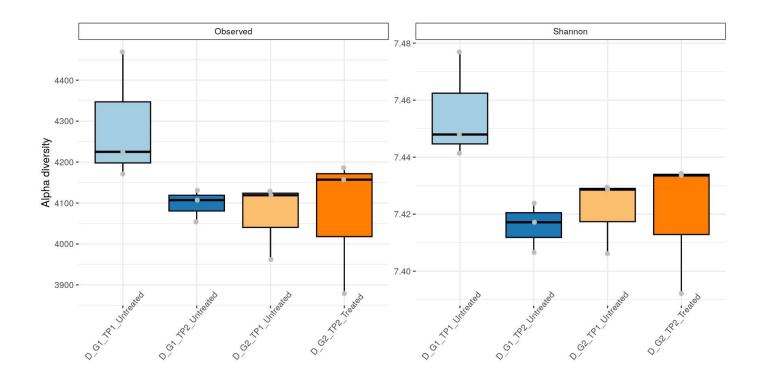


Figure 4: Illustration of the alpha diversity levels in each sample groups. The Observed and Shannon diversity measure is shown and the boxes are colored by sample group.

Version information

Table 7: List of used software including the used R-programming environment packages.

Package	Version	Package	Version
os	Ubuntu 20.04.4 LTS	biomformat	1.30.0
R	4.3.3	tools	4.3.3
rstudioapi	0.16.0	аре	5.8
magrittr	2.0.3	zip	2.3.1
TH.data	1.1-2	glue	1.7.0
estimability	1.5.1	nlme	3.1-165
farver	2.1.2	rhdf5filters	1.14.1
nloptr	2.1.1	cluster	2.1.6
rmarkdown	2.27	reshape2	1.4.4
zlibbioc	1.48.2	ade4	1.7-22
vctrs	0.6.5	generics	0.1.3
multtest	2.58.0	gtable	0.3.5
minqa	1.2.7	tzdb	0.4.0
RCurl	1.98-1.16	hms	1.1.3
rstatix	0.7.2	car	3.1-2
htmltools	0.5.8.1	xml2	1.3.6
S4Arrays	1.2.1	utf8	1.2.4
broom	1.0.6	foreach	1.5.2

Package	Version	Package	Version
Rhdf5lib	1.24.2	pillar	1.9.0
SparseArray	1.2.4	splines	4.3.3
rhdf5	2.46.1	survival	3.7-0
sass	0.4.9	deldir	2.0-4
bslib	0.7.0	tidyselect	1.2.1
sandwich	3.1-0	svglite	2.1.3
z00	1.8-12	xfun	0.46
cachem	1.1.0	stringi	1.8.4
igraph	2.0.3	yaml	2.3.9
lifecycle	1.0.4	boot	1.3-30
iterators	1.0.14	evaluate	0.24.0
pkgconfig	2.0.3	codetools	0.2-20
R6	2.5.1	interp	1.1-6
fastmap	1.2.0	cli	3.6.3
GenomeInfoDbData	1.2.11	RcppParallel	5.1.8
digest	0.6.36	systemfonts	1.1.0
colorspace	2.1-0	xtable	1.8-4
ggpubr	0.6.0	munsell	0.5.1
hwriter	1.3.2.1	jquerylib	0.1.4
labeling	0.4.3	coda	0.19-4.1
fansi	1.0.6	png	0.1-8
timechange	0.3.0	parallel	4.3.3
abind	1.4-5	latticeExtra	0.6-30
mgcv	1.9-1	jpeg	0.1-10
compiler	4.3.3	bitops	1.0-7
withr	3.0.0	viridisLite	0.4.2
backports	1.5.0	mvtnorm	1.2-5
carData	3.0-5	crayon	1.5.3
highr	0.11	rlang	1.1.4
ggsignif	0.6.4	mnormt	2.1.1
MASS	7.3-60.0.1	cowplot	1.1.3
DelayedArray	0.28.0	multcomp	1.4-26