Statistical Analysis Report

Taxonomic profile from shotgun metagenomes for full project (version May 2024)

Biomcare ApS

22/05/2024

Customer	Innovation Centre for Organic Farming, Tove Mariegaard Pedersen
Customer ID	DA00204
Project	The microbial community of the field (2021, 2022 and 2023).
Sample Type	Soil
Number of samples	102 samples
Type of data	shotgun metagenomics

Introduction to the biostatistical analysis

The Project

The current report describes microbiome profiles of 100 samples collected across 3 project years (2021-2023), across a corresponding number of fields in Denmark. For each field, one sample was collected to represent the field, corresponding to the 'main' samples collected for each field from 2021. These samples were taken for each field based on 16 subsamples taken in a w-pattern throughout the field.

In this report we analyse the full project dataset. We split many analysis by JB values into 2 groups based on initial analysis of data from 2021-2022 where we saw a strong association between JB and microbiome profiles. We split the further analysis based on JB as we find the effect of JB overshadoves the associations that may be between the microbime and other variables of interest. Different from the prior years reports we also added geography information and some weather information like rainfall and temeratures. We will relate these informations to the microbiome as well. Where feasiable, we will both look at the cross-year effect where we adjust for the effect of year and look for effects that differ between years thus looking for association between the microbiome and metadata that differ (or depend on) other factors that associate with a specific year. The aim is to evaluate how the microbiome of the fields associate with other field parameters of both agricultural practices and soil indicators of nutrients, type and structure.

The JB groups are:

- JB 1 + JB 2
- JB 5 + JB 6 + JB 7

The one sample with JB4 is removed. (one sample was removed for stating JB 2-4)

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 48 samples in the below table.

Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
year	99						
… 2021	13	13%					
… 2022	48	48%					
… 2023	38	38%					
JB_groups	99						
… JB1_JB2	53	54%					
… JB5_JB6_JB7	46	46%					
JB_value	99						
… 1	32	32%					
… 2	21	21%					
… 5	4	4%					
… 6	39	39%					
… 7	3	3%					
Geographic_location_letters	99						
… B	4	4%					
… F	4	4%					
… MJ	32	32%					
… NJ	11	11%					
… S	15	15%					
… SJ	18	18%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
… V	15	15%					
Rainfall	99	352	49	256	313	388	442
Average_drought_index	99	6.5	1.6	3.8	4.9	7.9	8.7
Average_temp.	99	13	0.45	13	13	14	14
Earthworm_status	99						
… 0	22	22%					
… 1	77	78%					
Cold_soil	99						
… 0	73	74%					
… 1	26	26%					
Compact_soil	99						
… 0	85	86%					
… 1	14	14%					
field_well_drained	99						
… 0	10	10%					
… 1	89	90%					
Mulching_of_straw	99						
… 0	50	51%					
… 1	49	49%					
Clovergrass_within_3_years	99						
… 0	69	70%					
… 1	30	30%					
No_plough	99						
… 0	77	78%					
… 1	22	22%					
ConservationAgriculture	99						
… 0	93	94%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
… 1	6	6%					
Years_since_plowing	99	3.8	3.5	1	1	5	14
Rt	99	6.4	0.52	5.4	6	6.7	7.6
Phosphorus	99	3.1	1.3	0.7	2.1	3.9	8.2
Potassium	99	11	7	1.5	6.5	13	50
Magnesium	99	6.5	2.6	1.9	5	7.3	16
Cobber	99	2.5	1.1	1	1.8	3.1	7.5
Organic_material_perc	99	3.2	1.4	1.2	2.3	3.8	8.7
Clay_perc	99	8.8	4.3	2.4	4.8	12	20
Nitrogen_perc	99	0.15	0.054	0.07	0.12	0.18	0.35
Organic_farm	99						
… 0	42	42%					
… 1	57	58%					
Years_since_turning_organic	57	15	6.2	1	10	20	22
Livestock	99						
… 0	40	40%					
… 1	59	60%					
Livestock_manure	99						
… 0	23	23%					
… 1	76	77%					
Commercial.fertilizer	99						
… 0	56	57%					
… 1	43	43%					
Vinasse	99						
… 0	92	93%					
… 1	7	7%					
Cast	99						

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
… 0	97	98%					
… 1	2	2%					
Degassed.fertilizer	99						
… 0	76	77%					
… 1	23	23%					
Haveparkaffald	99	0.051	0.22	0	0	0	1
Chalked	98						
… 0	76	78%					
… 1	22	22%					

Table 1: Summary statistics of the key variables selected for evaluation in relation to the fields microbiome profiles in year 2022.

Geography and JB values

As a new feature we include information on the geographic location of the fields collected in the project. Se use coordinates to show the location of the fields on a map, and color the data points by JB value and year of sampling to get an understanding of the relationship. We then identify single species that differ in abundance between geographic reagions, using an anova model adjusting for year.

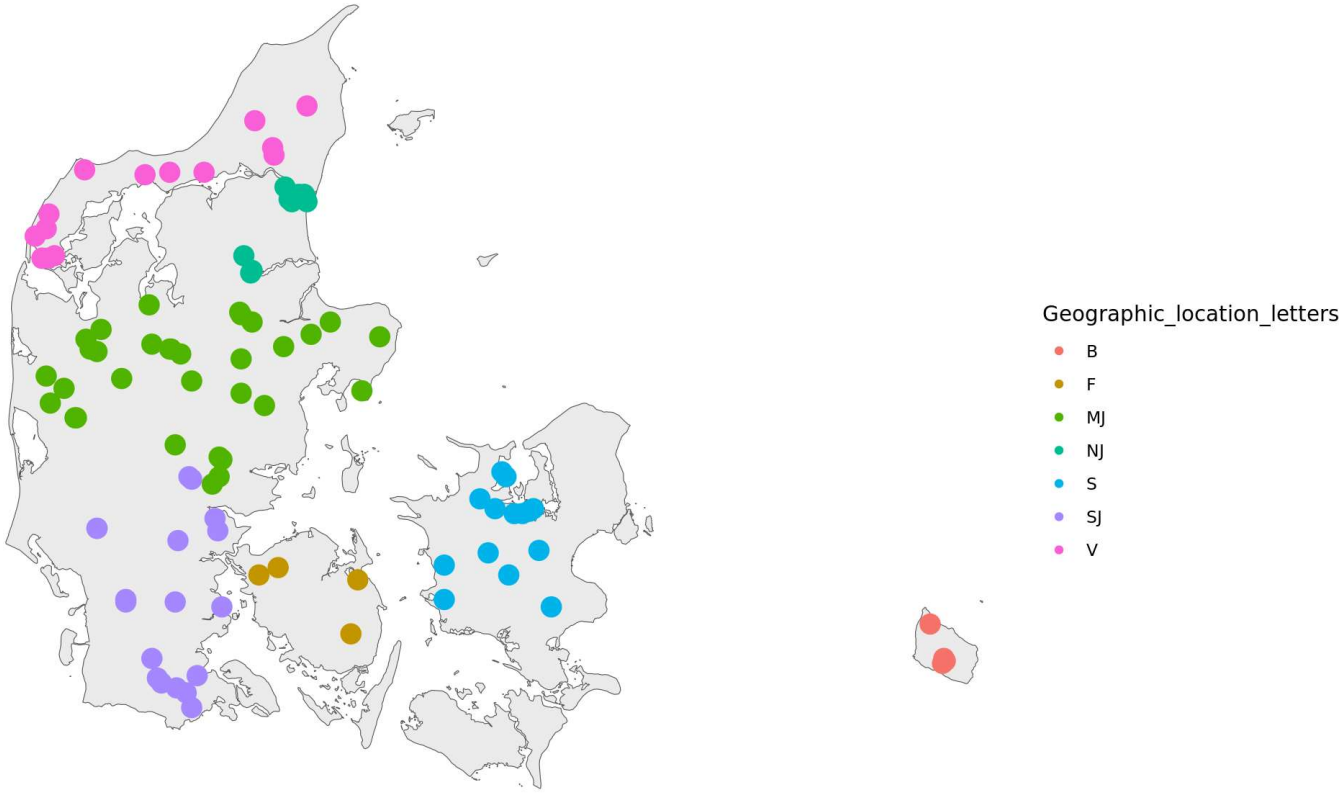


Figure 1: Visualization of the geographic location of the samples. Using the coordinates of each field in the project, we show the locations on a map and color the samples by region to illustrate the regional distribution of samples in the project.

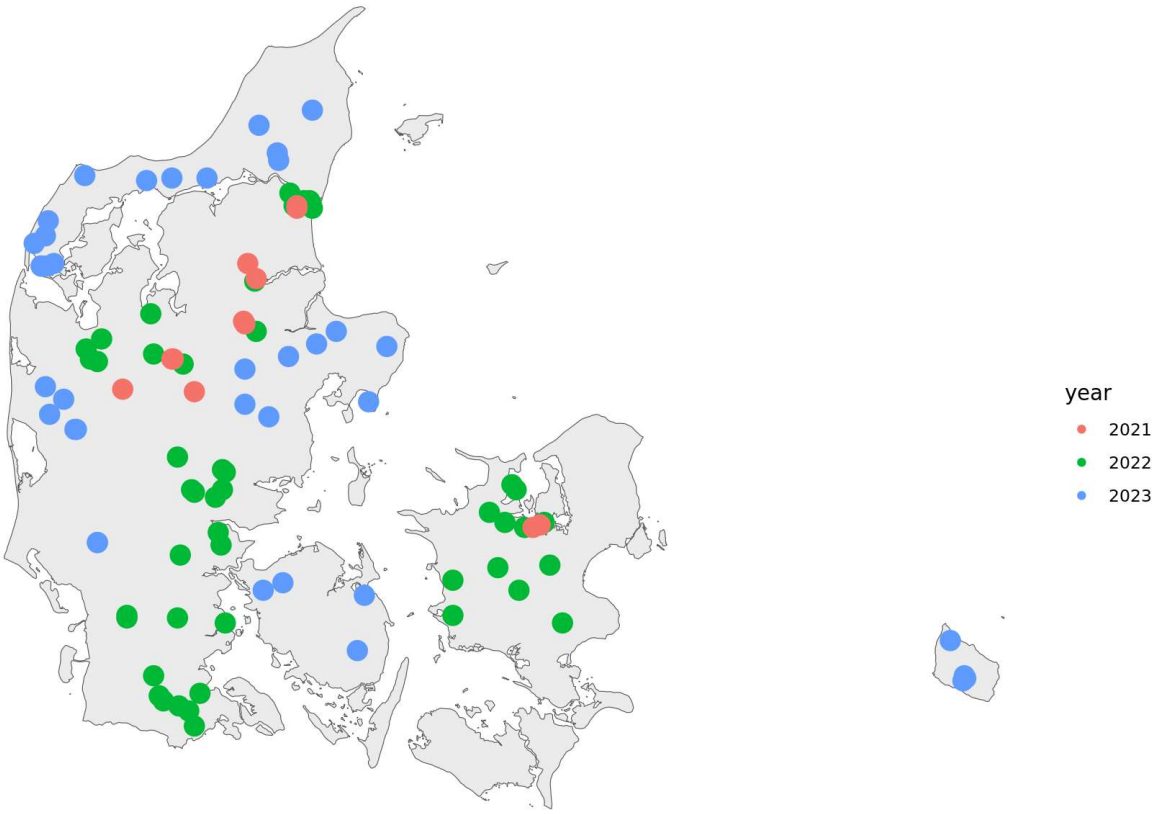


Figure 2: Visualization of the geographic distribution of samples collected each year. Using the coordinates of each field in the project, we show the locations on a map and color the samples by year of sampling.

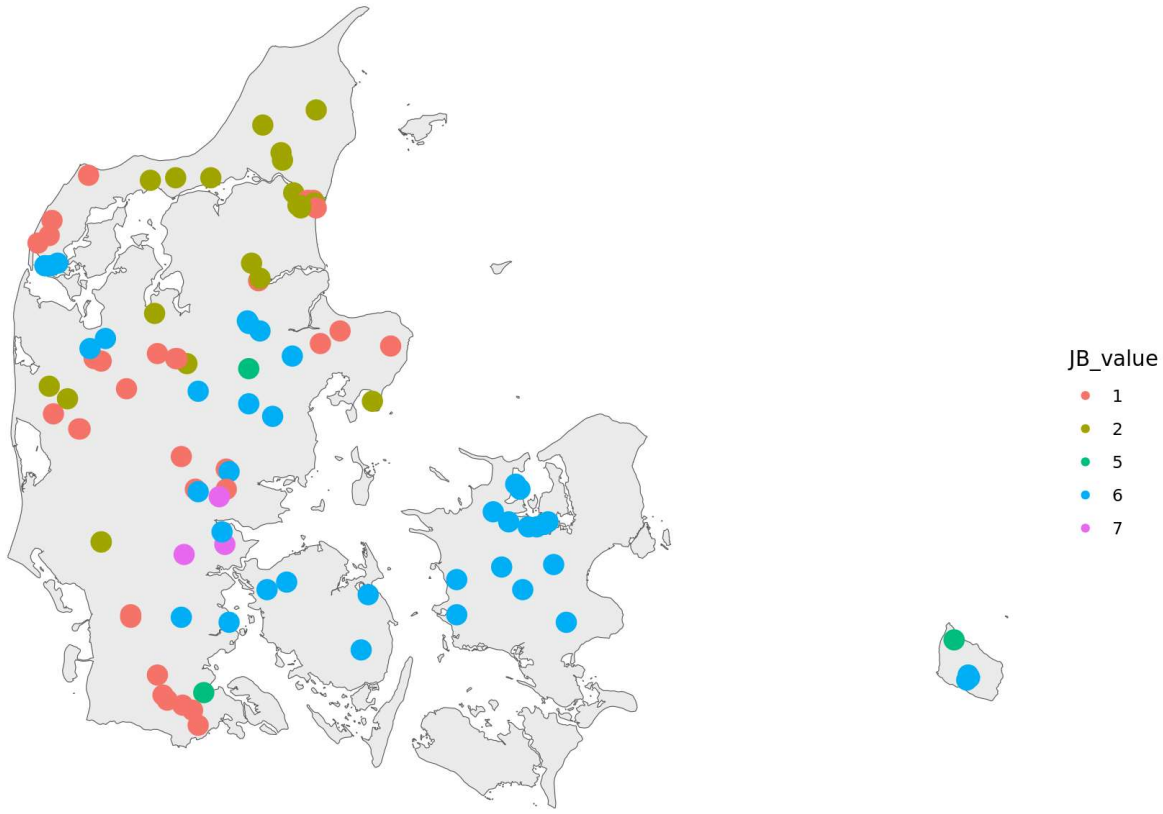


Figure 3: Visualization of the geographic distribution of JB types for the included samples. Using the coordinates of each field in the project, we show the locations on a map and color the samples by year of JB value

The abundance of individual species may also deviate between regions driven by the differences in soil properties, farming practices and climate tbetween the regions. We here identify species that differ in abundance between any two regions and for top deviating species show their abundance on a map.

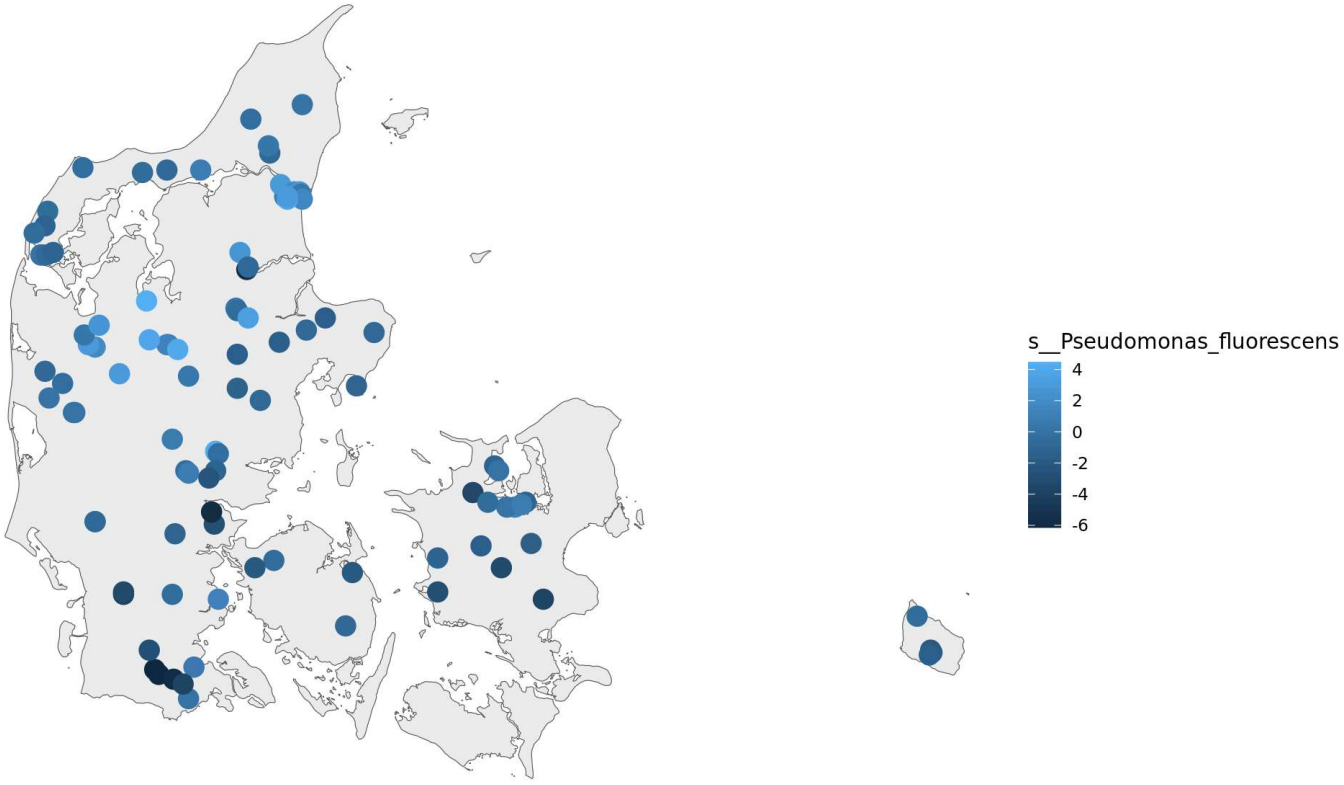
	Sum.Sq	Df	F.value	Pr..F.	p.adj
s__Mycolicibacterium_frederiksbergense	5.495	6	5.278	0.000108	0.00537
s__Pseudomonas_fluorescens	112.057	6	5.186	0.000129	0.00537
s__Streptomyces_vinaceus	19.073	6	4.590	0.000420	0.01160
s__Streptomyces_sp._RPA4_2	11.113	6	3.666	0.002670	0.03690
s__Nocardioides_euryhalodurans	11.123	6	3.769	0.002170	0.03690
s__Peribacillus_butanolivorans	18.044	6	3.699	0.002500	0.03690
s__Micromonospora_echinofusca	24.770	6	3.340	0.005160	0.04760
s__Turicibacter_sp._H121	15.736	6	3.389	0.004670	0.04760

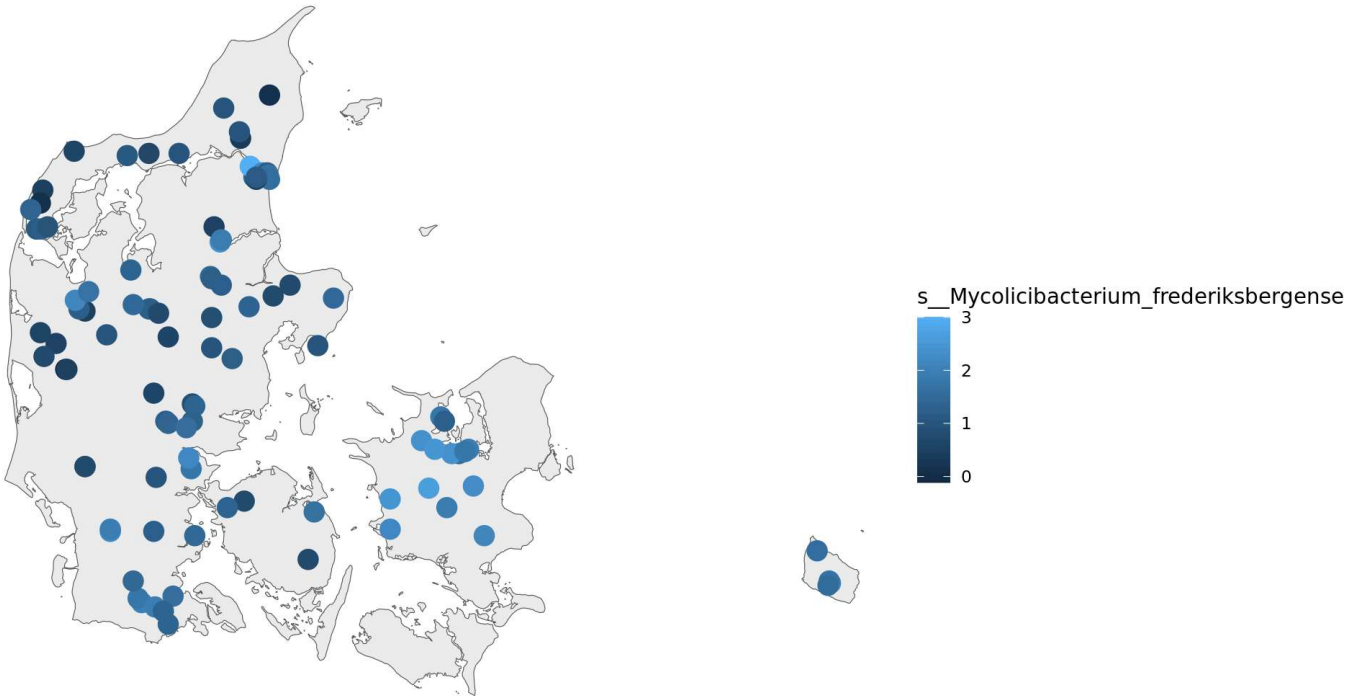
	Sum.Sq	Df	F.value	Pr..F.	p.adj
s__Paraburkholderia_hospita	31.014	6	3.404	0.004530	0.04760
s__Ensifer_adhaerens	24.209	6	3.210	0.006710	0.05570
s__Arthrobacter_sp._24S4_2	5.385	6	2.861	0.013500	0.09200
s__Microbacterium_foliorum	10.988	6	2.765	0.016400	0.09200
s__Nocardioides_ungokensis	7.492	6	2.874	0.013200	0.09200
s__Nakamurella_multipartita	6.886	6	2.758	0.016600	0.09200
s__Pseudomonas_putida	64.339	6	2.818	0.014700	0.09200
s__Nocardioides_sp._S5	5.245	6	2.550	0.025200	0.13100

Table 2: Results from ANOVA analysis of single species differential abundance between regions. A anova model was used to identify single species that wary in clr-transformed abundance between any two regions (adjusting for year and JB groups). Species to be analyzed where pre-selected as the top abundant and most varying species in the dataset. The table show species with adjusted p<0.15 and rows are colored red (p.adj<0.05) and salmon (p.adj<0.1) based on the adjusted p values.

Consider if we do want to adjust the analysis above for JB values. We know there is a srong correspondance between location and JB values, and one could argue that we expect a difference between location largely due to soil type differences. When we adjust for JB, we thus look for any location/region differences not driven by JB, and leave direct JB differences to the below analyses.

Maps reflecting species abundance of the top 3 species showing strongest region differences The clr-transformed abundance of the top 3 species from the above table is shown on individual maps.


s__Pseudomonas_fluorescens
4
2
0
-2
-4
-6


s__Mycolicibacterium_frederiksbergense
3
2
1
0

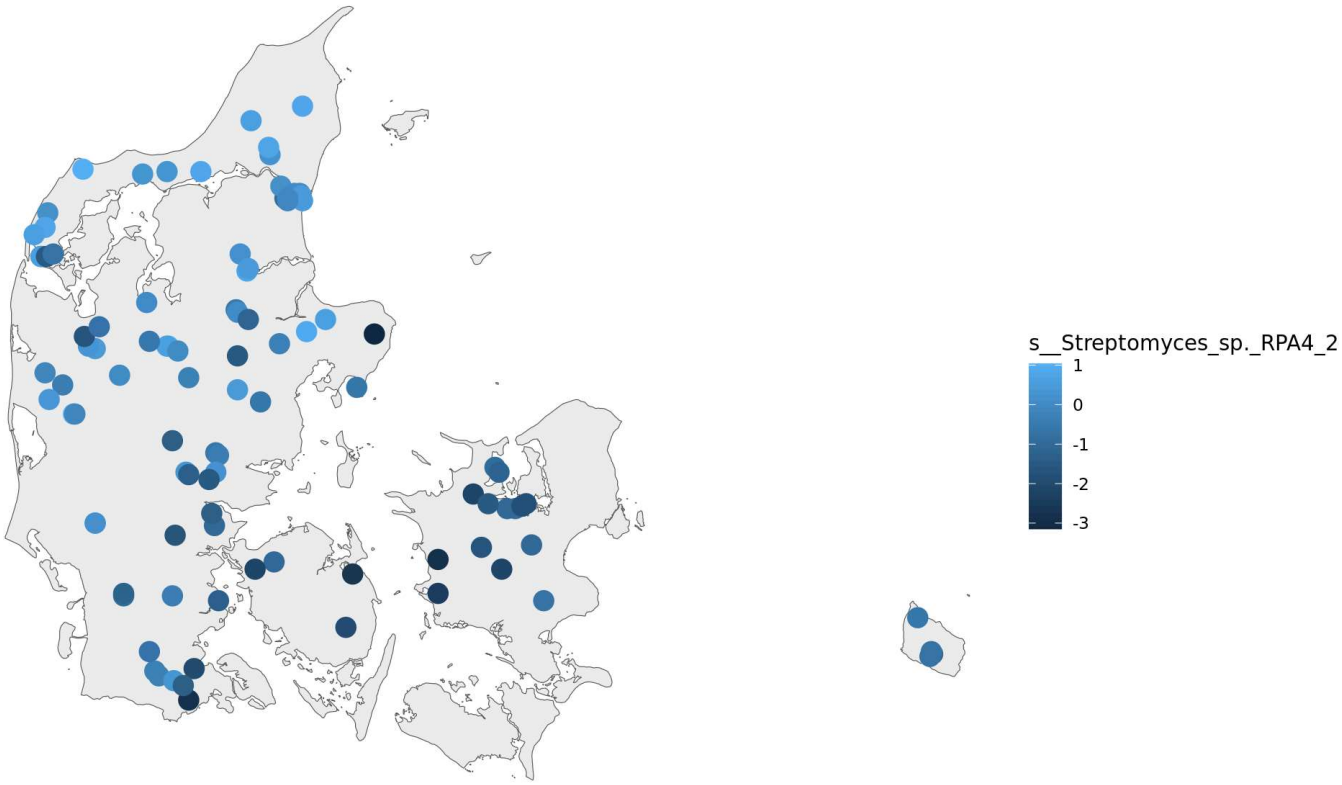


Figure 4: Visualization of the geographic distribution of samples colored by the clr-transformed abundance of one specific species.

The species *Streptomyces sp. RPA4 2* show an interesting pattern on the above plot and we thus zoom in to review the species abundance across the location using a boxplot, and review the association with JB groups and year of sampling.

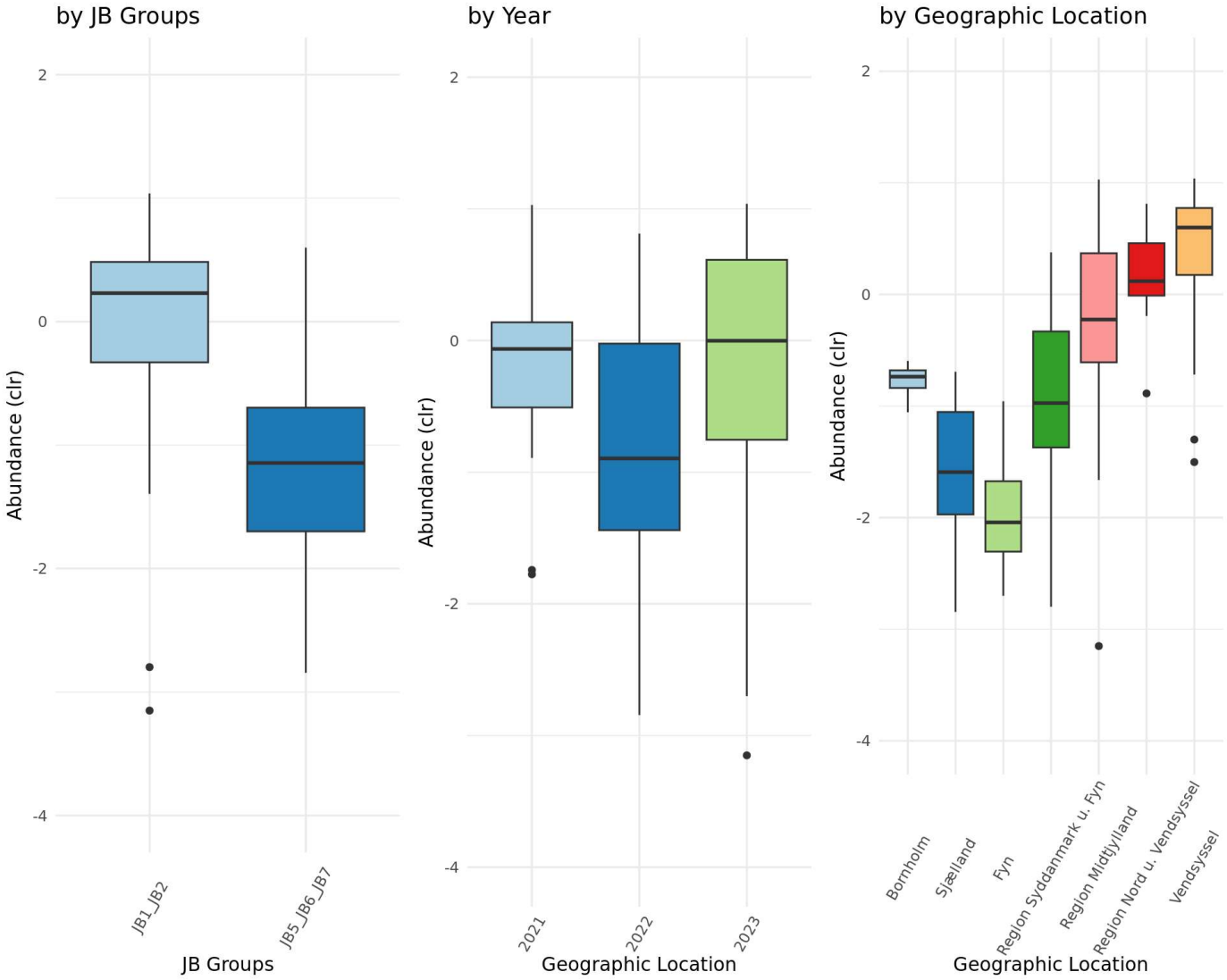


Figure 5: Boxplots showing the clr-trannformed species abundance between JB groups, year and region.

Evaluation of overall microbiome profiles

We initiate the evaluation of the 10 samples (1 per field) with a stacked barplot of the microbiome profiles in each sample. This allows us to make a first evaluation of the extent of difference in the taxonomic profiles between the fields.

Note that in order to show the organisms with a color scheme that is interpretable, it is necessary to filter the profiles and select a subset of the most abundant clades to be included in the plots. The filtering used is specified in the axis labels of each plot (e.g. >2% in the relative abundance plots mean that a clade must have a relative abundance across samples of more than 2% in order to be included in the plot).

Stacked barplots

The stacked barplots allow us to visually access the stability of the taxonomic profile across the fields, and get a feeling of the level to which individual clades are found across field or more sporadic. Compared to the bacterial part of the microbiome, the fungi show a large deviation between fields, with both large variation in some, and others that are dominated by a few clades. And we see how the dominating clade is also different between many fields.

Phylum | Class | Order

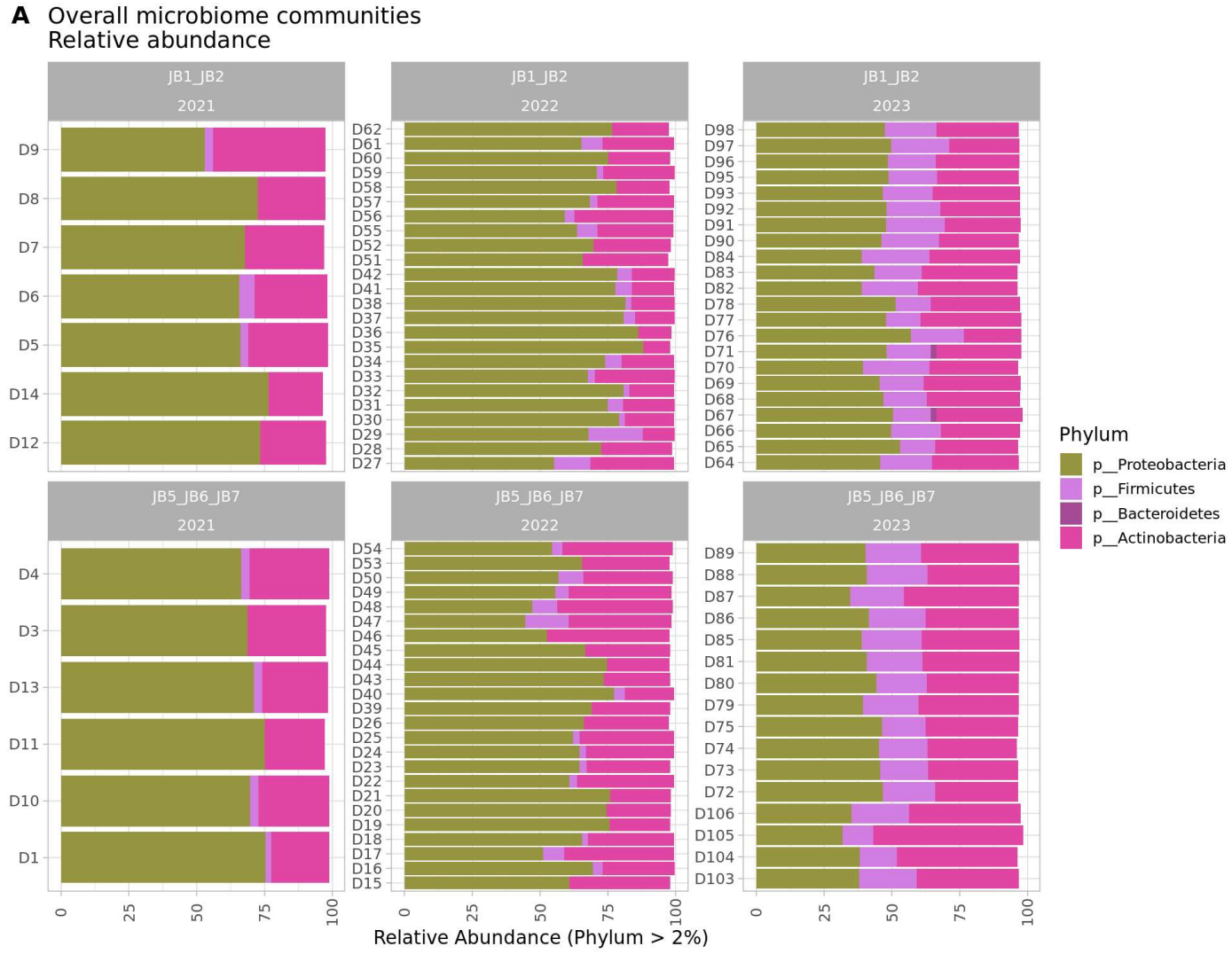


Figure 6: Visualization of the microbial community in the samples. Stacked barplots of taxonomic clades in each of the evaluated samples. Clade abundance was transformed to relative abundance to sum to 100% in each sample. The plot is seperated into two parts - one per group of samples based on JB values.

Microbiome communities related to factor variables

We use the overall microbiome profiles presented in the stacked barplots above, 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 the 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. For these evaluations we focus on the variables that can be viewed as binary or generating groups (e.g. JB value, Earthworm status, Crops, Mulching of straw and Years since plowing). These variables are “grouping variables” that allow us to group samples into subgroups. The remaining variables constitute continuous values (concentrations or percentages) and therefore, we use a different set of tools to evaluate how they relate to the overall microbiome composition (see below).

Note that “years since plowing” and “Years since turning organic” are analyzed as integers with increasing values (0,1,2, etc) and could be analyzed both with ADONIS and the model used for continuous variables (ENVFIT, see below). However, we find it interesting to visualize the shift in the composition according to “years” in an ordination plot, and the variable is therefore included here.

Visualization by heatmaps (for single organisms)

We show the relative abundance (0-100) for the most abundant species in samples groups. The samples are grouped first by JB group and then the factor variable. This allow us to inspect JB specific and general effects on the taxa for each factor variable.

Note, we show here plots for one taxonomic level, but plots are made for all levels, and should be reviewed. See these in the project folder under Illustrations (DA00204/data/ITS/4_Statistical_analysis/Illustrations). There heatmaps are foudn both with samples grouped by factor variables and with abundance shown for each sample seperated just by JB value (e.g. Heatmap_Class_byJB.png).

Year and JB | Earthworm_status | Cold_soil | Compact_soil | field_well_drained | Mulching_of_straw | Clovergrass_within_3_years | No_plough | ConservationAgriculture | Years_since_plowing | Organic_farm | Years_since_turning_organic | Livestock | Livestock_manure | Commercial.fertilizer | Vinasse | Cast | Degassed.fertilizer | Chalked

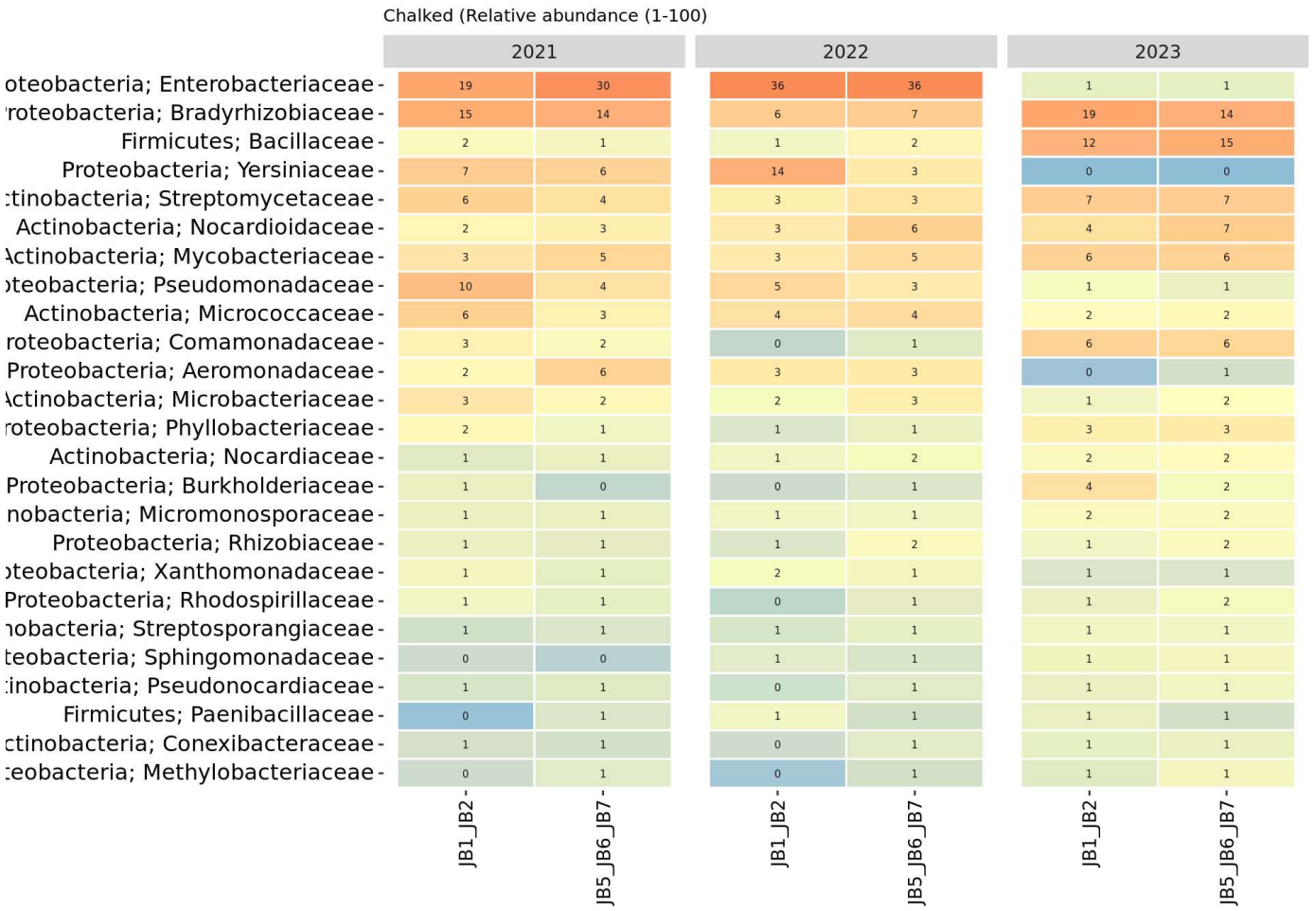


Figure 9: Visualization of abundance of the microbial organism between samples grouped by JB and factor variable.

Visualization by ordination (beta-diversity)

As described in **Report 2**, beta-diversity is a measure of how similar or dissimilar the microbial 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. If not all plots are shown in this report, you can find them in the project folder. We use a combination of beta-diversity 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:

- Bray Curtis and Jaccard are computed from the relative abundance data, at the the genus level
- Aitchison is computed from the total 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.

As we move from the per-year analysis to a cross-years analysis, I have changed the ordination method from an NMDS method (metaMDS in package vegan) to MDS/dbRDA (capscale in package vegan), as the capscale approach allows me to condition (or partial out, aka. remove), the effect of year from the data allowing us to better inspect effects of other variables in the ordination plots. In the statistical analysis e.g. the ADONIS model, I do this by including a term for year in the model.

Perhaps when we have data from 2023 as well, we can try to look if any effects are specific to one year, and then discuss what was different between the years to possible cause this. But with the few samples form 2021 we cannot relably extract such information at this point.

Permutational Multivariate Analysis of Variance

To evaluate if the compositional differences evaluated below using ordination plots 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.

In each plot, samples are colored by the variable being assessed.

JB value | Earthworm status | Cold_soil | Compact soil | Field well drained

Mulching of straw | Clovergrass within 3 years | No plough | Conservation Agriculture

Years since plowing Organic farm Years since turning organic Livestock

Commercial fertilizer Vinasse Cast Degassed fertilizer Chalked

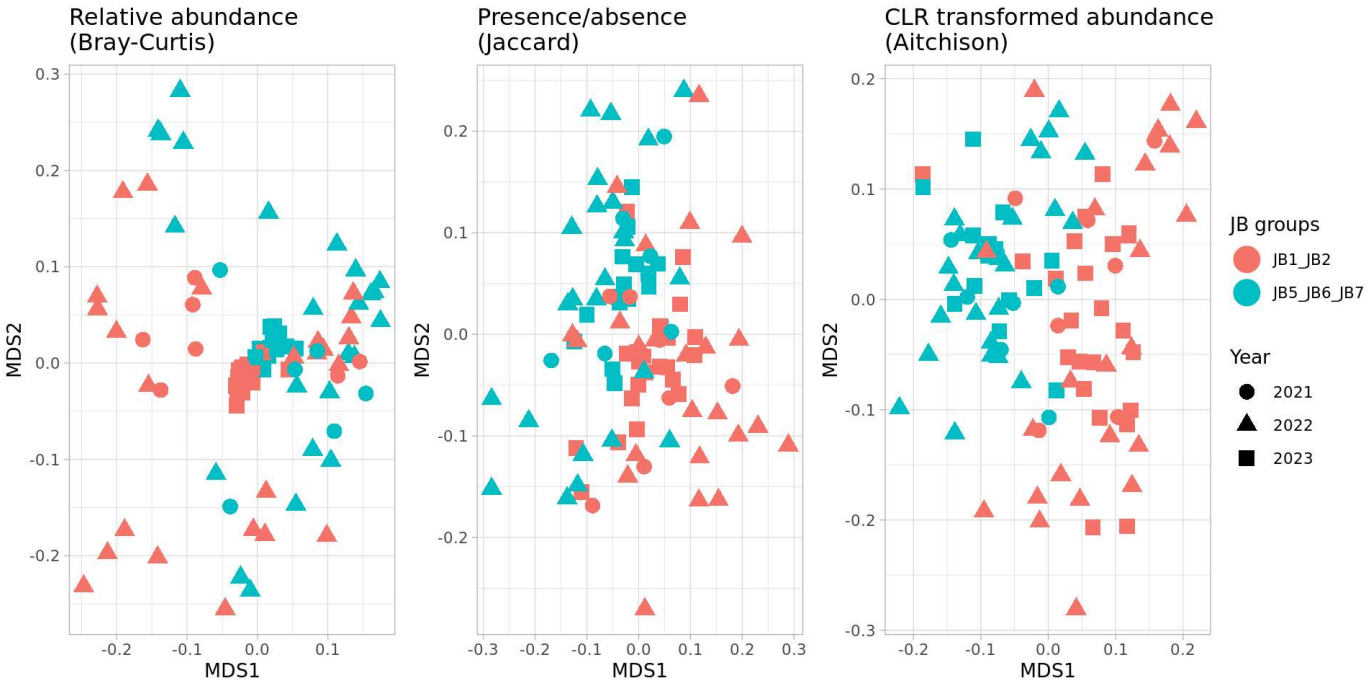


Figure 28: Visualization of structure of the microbial community between the samples. Ordination plots using different beta-diversity measures and data transformations as stated in the plot titles. Dots are colored by the variable of interest as seen to the right of the figure panels.

	Bray-Curtis		Jaccard		Aitchison	
Taxa level	**R2**	**p**	**R2**	**p**	**R2**	**p**
Phylum	0.0653	0.001	0.016	0.033	0.0303	0.002
Class	0.0244	0.001	0.0249	0.004	0.0351	0.001
Order	0.0246	0.002	0.0226	0.007	0.0324	0.001
Family	0.0349	0.001	0.0236	0.006	0.0356	0.001
Genus	0.0413	0.001	0.0241	0.003	0.0344	0.001
ASV	0.0495	0.001	0.0365	0.001	0.0368	0.001

Table 3: Results from ADONIS analysis. The table shows results from ADONIS analyses including samples from all fields. The analysis was performed using 999 permutations 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.

With the striking effect of JB groups on microbiome community when viewed in respect to the composition properties (clr data), we look further into the differences. First we apply an ANOVA model to identify single species differing in abundance between the two JB groups.

	Sum.Sq	Df	F.value	Pr..F.	p.adj
s__Streptomyces_sp._RLB1_33	42.083	1	99.560	0.0000000	0.0000000
s__Bradyrhizobium_diazoefficiens	26.351	1	97.504	0.0000000	0.0000000
s__Luteitalea_pratensis	28.893	1	89.742	0.0000000	0.0000000
s__Ensifer_adhaerens	121.557	1	84.854	0.0000000	0.0000000
s__Phycicoccus_sp._HDW14	30.312	1	77.809	0.0000000	0.0000000
s__Arthrobacter_sp._PAMC25564	25.942	1	63.976	0.0000000	0.0000000
s__Microvirga_ossetica	68.529	1	61.034	0.0000000	0.0000000
s__Paraburkholderia_hospita	99.690	1	56.994	0.0000000	0.0000000
s__Streptomyces_sp._RPA4_2	32.771	1	55.521	0.0000000	0.0000000
s__Bradyrhizobium_betae	65.914	1	40.710	0.0000000	0.0000001
s__Afipia_sp._GAS231	10.468	1	39.088	0.0000000	0.0000001
s__Streptomyces_vinaceus	29.807	1	35.082	0.0000001	0.0000003
s__Agromyces_aureus	29.973	1	30.428	0.0000003	0.0000019
s__Skermanella_sp._TT6	10.540	1	30.190	0.0000003	0.0000019
s__Bradyrhizobium_japonicum	10.854	1	28.985	0.0000005	0.0000029
s__Micromonospora_zamorensis	18.445	1	25.144	0.0000025	0.0000128
s__Peribacillus_butanolivorans	22.894	1	24.058	0.0000039	0.0000188
s__Nocardioides_sp._zg_1228	10.456	1	21.695	0.0000104	0.0000479
s__Nocardioides_sp._S5	7.752	1	20.593	0.0000166	0.0000726
s__Oerskovia_sp._KBS0722	16.464	1	19.483	0.0000268	0.0001110
s__Mycolicibacterium_frederiksbergense	4.196	1	19.038	0.0000326	0.0001290
s__Mycobacterium_sp._JS623	7.843	1	18.898	0.0000346	0.0001310
s__Brevibacterium_sp._PAMC23299	80.045	1	18.385	0.0000433	0.0001560
s__Arthrobacter_sp._U41	7.069	1	16.893	0.0000839	0.0002900
s__Rhodococcus_erythropolis	6.108	1	15.523	0.0001560	0.0005170
s__Nocardioides_euryhalodurans	8.594	1	14.873	0.0002090	0.0006680
s__Aeromonas_encheleia	64.813	1	14.194	0.0002860	0.0008790

	Sum.Sq	Df	F.value	Pr..F.	p.adj
s__Ureibacillus_thermosphaericus	19.860	1	13.408	0.0004120	0.0012200
s__Nocardioides_ungokensis	6.399	1	13.170	0.0004600	0.0013200
s__Streptomyces_sp._P3	7.401	1	12.767	0.0005560	0.0015400
s__Herbinix_luporum	8.688	1	11.303	0.0011200	0.0029900
s__Rhodococcus_sp._MTM3W5.2	8.579	1	11.209	0.0011700	0.0030300
s__Bradyrhizobium_lablabi	2.282	1	10.281	0.0018300	0.0046100
s__Bradyrhizobium_sp._SK17	6.130	1	10.188	0.0019200	0.0046800
s__Streptomyces_hygroscopicus	4.683	1	8.842	0.0037300	0.0088500
s__Arthrobacter_sp._FB24	2.314	1	7.769	0.0064200	0.0148000
s__Arthrobacter_sp._24S4_2	2.596	1	7.405	0.0077400	0.0174000
s__Micromonospora_sp._B006	14.112	1	7.086	0.0091200	0.0196000
s__Bradyrhizobium_sp._LCT2	3.937	1	7.068	0.0092100	0.0196000
s__Pseudarthrobacter_equi	13.843	1	6.997	0.0095600	0.0198000
s__Streptomyces_niveus	7.469	1	6.816	0.0105000	0.0213000
s__Streptomyces_sp._3214.6	5.938	1	6.639	0.0115000	0.0228000
s__Aminobacter_sp._MSH1	5.005	1	6.593	0.0118000	0.0228000
s__Nocardioides_cynanchi	2.082	1	6.500	0.0124000	0.0234000
s__Conexibacter_sp._SYSU_D00693	1.059	1	5.747	0.0185000	0.0333000
s__Rhodoplanes_sp._Z2_YC6860	1.542	1	5.756	0.0184000	0.0333000
s__Variovorax_sp._WDL1	4.406	1	5.246	0.0242000	0.0427000
s__Microbacterium_foliorum	3.778	1	5.131	0.0258000	0.0437000
s__Turicibacter_sp._H121	4.598	1	5.162	0.0253000	0.0437000
s__Aeromicrobium_choanae	9.903	1	4.234	0.0424000	0.0703000
s__Pseudomonas_fluorescens	18.086	1	3.972	0.0491000	0.0800000
s__Priestia_megaterium	2.906	1	3.930	0.0503000	0.0803000
s__Microterricola_viridarii	1.699	1	3.815	0.0537000	0.0841000
s__Bradyrhizobium_ottawaense	0.760	1	3.353	0.0702000	0.1080000

	Sum.Sq	Df	F.value	Pr..F.	p.adj
s__Bradyrhizobium_erythrophlei	1.027	1	2.840	0.0952000	0.1440000

Table 4: Results from ANOVA analysis of single species differential abundance between JB groups A anova model was used to identify single species that wary in clr-transformed abundance between any two JB groups (adjusting for year). Species to be analyzed where pre-selected as the top abundant and most varying species in the dataset. The table show species with adjusted p<0.15 and rows are colored red (p.adj<0.05) and salmon (p.adj<0.1) based on the adjusted p values.

The species *Luteitalea pratensis* is among the top associated species, and is a species previously found to displays a narrower pH, possible leading to the observed differences. This possibility was supported by a significant association between the species abundance and Rt (correlation test r=0.38, p=0.00009). We thus zoom in to review the species abundance across the location using a boxplot, and review the association with JB groups and year of sampling.

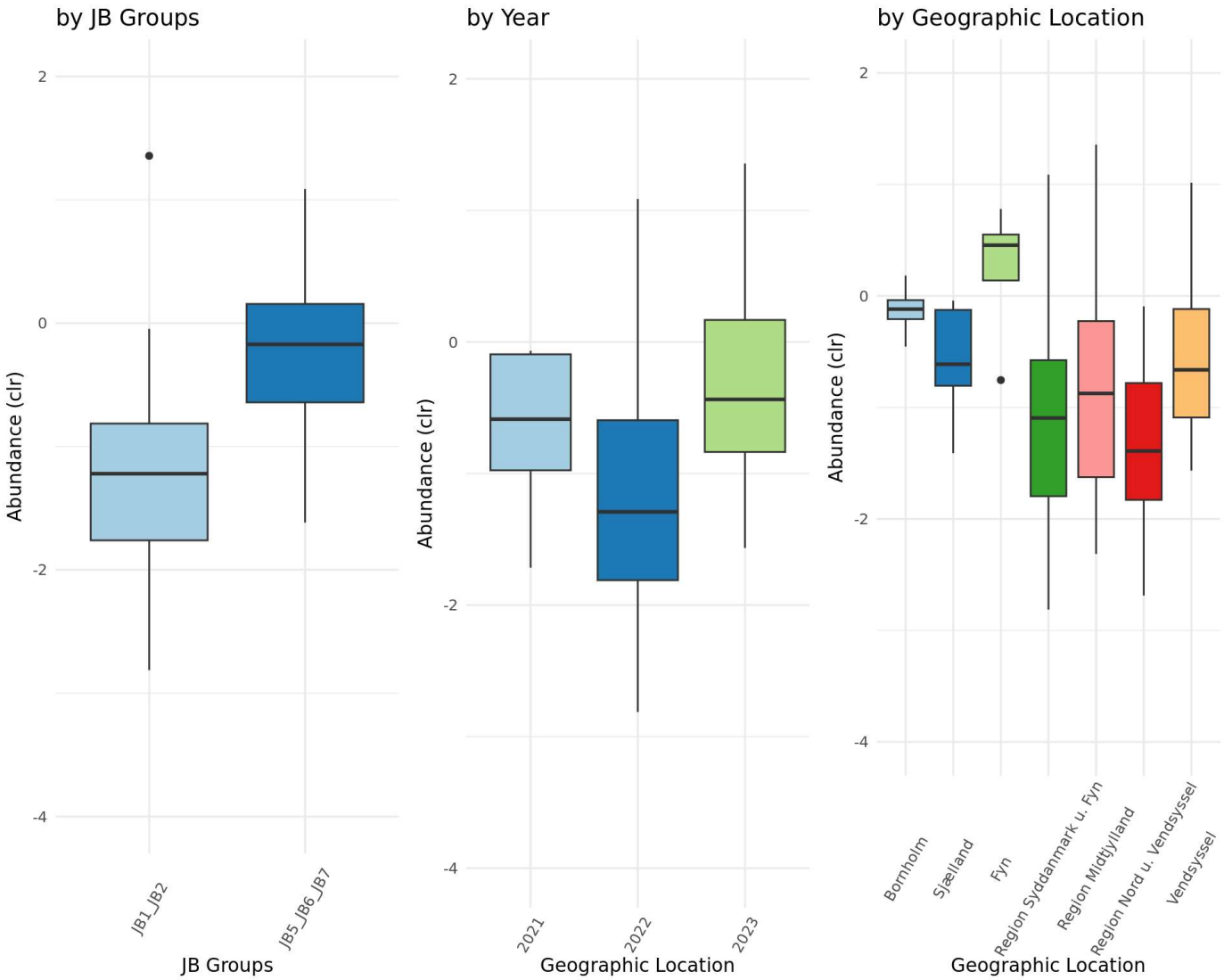


Figure 29: Boxplots showing the clr-trannformed species abundance between JB groups, year and region.

Overall microbiome communities related to continuous variables

We now turn to the continuous variables. We continue to use the beta-diversity measure from the microbiome, but use a different analysis model named ENVFIT. We again display the samples in ordination plots but add to these plots arrows representing the association of the variables with the microbiome.

ENVFIT

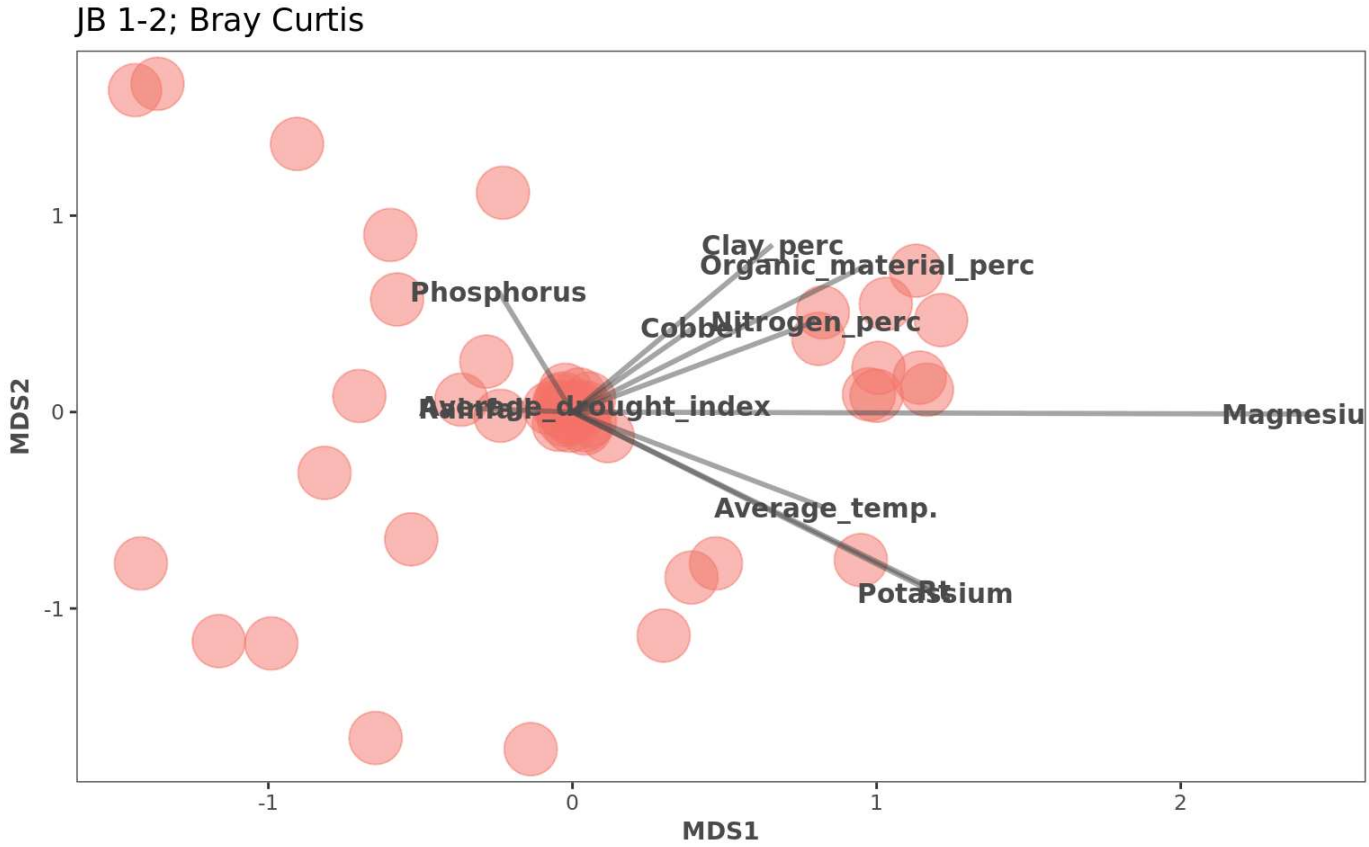
The function fits environmental variables onto an ordination using multiple regression and thus looks for linear relationships between the axis that drive the main variation in the microbiome and the environmental variables.

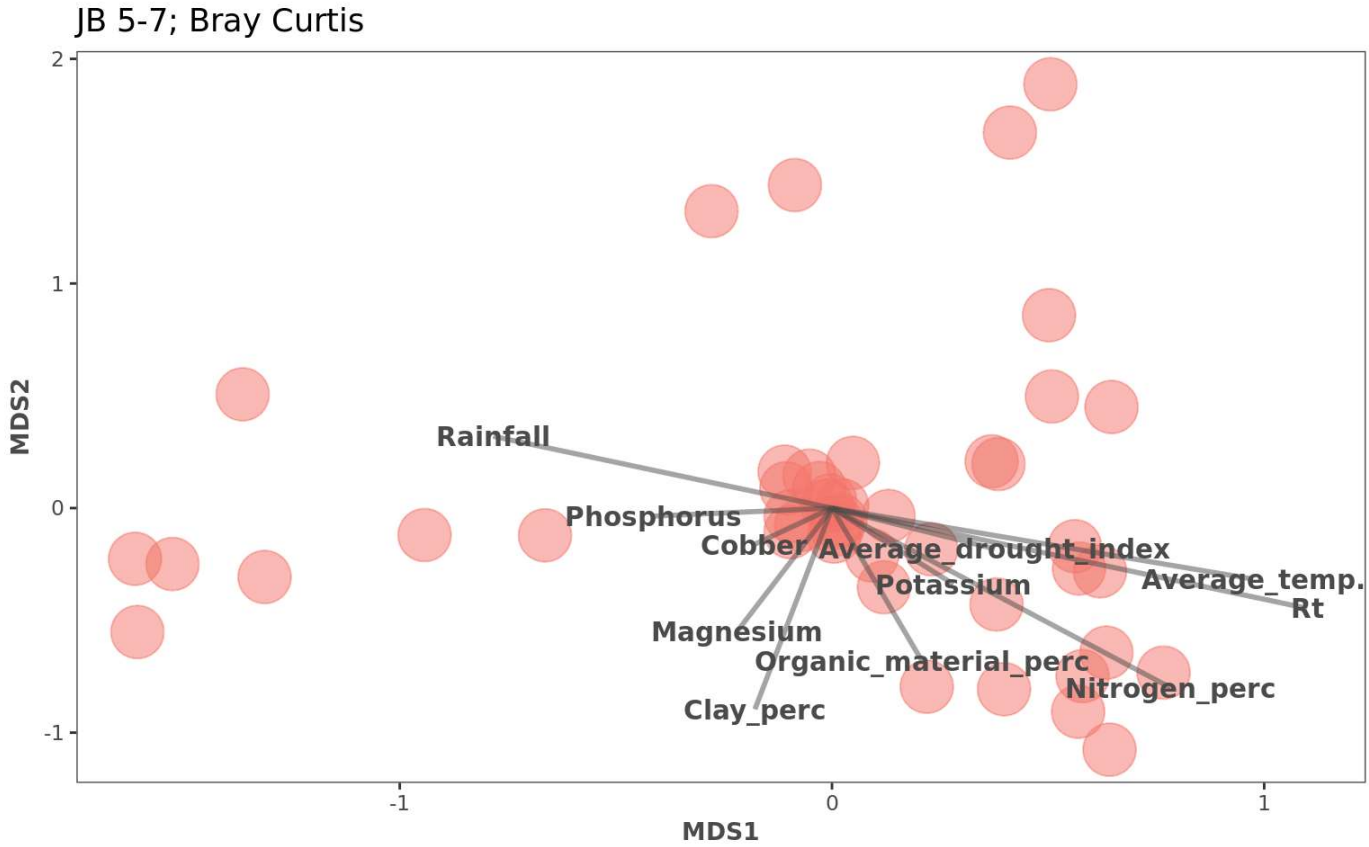
The output of the analysis gives the direction of the association in terms of coordinates in the ordination plot (the arrow head). In the plots, the arrows are scaled by their correlation (square root of R2) so that “weak” variables have shorter arrows than “strong” variables. The arrows point in the direction where they have maximal correlations with the ordination axis. The lengths of arrows for fitted vectors are automatically adjusted for the physical size of the plot, and the arrow lengths can thus not be compared across plots, only the arrows within a plot can be compared.

Note that the ENVFIT analysis uses the ordination exes and therefore is restricted to relate the environmental factors to the part of the microbiome that is represented by these axes (and thus not the full microbiome variation). This fact is also relevant when comparing R2 (the squared correlation coefficient) from ADONIS and envfit: ADONIS decomposes the entire dissimilarity matrix into “variance” explained by each covariate. In ENVFIT, you have reduced the entire dissimilarity matrix to two dimensions and then look for correlations in those three dimensions with the covariates. The “discrepancy” is therefore due to the fact that different amounts of variation are assessed by the two models. Therefore, if ADONIS gives higher R2 values, this suggests that the effect of the variable is on the parts of the dissimilarity matrix not represented well by the 3-d NMDS solution. We are using capscale, and not the standard NMDS method, for the ENVFIT analysis, as capscale allow us to condition out the effect of year befroe performing the correlation analyses. * R2 - variation explained by the model of multiple regression; the square-root of this value is used to scale lengths of vectors (arrows) in the ordination diagrams (variables with higher squared (R2) are represented by longer arrows). * Pr(>r) - the significance of the multiple regression, calculated by permutation test. Indicates whether the variable is related to ordination axes more than if it is a randomly generated one.

Ordination plots

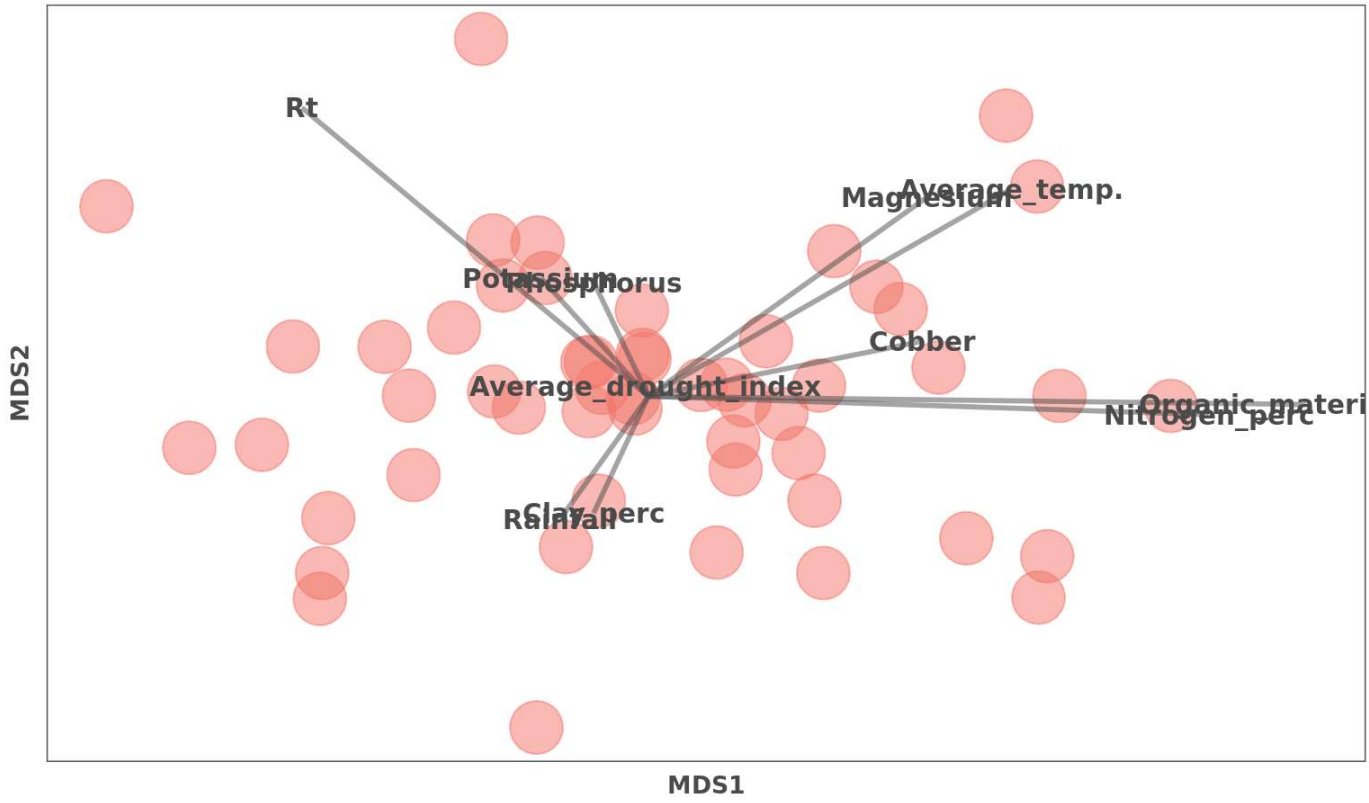
Bray Curtis


JB 1-2; Bray Curtis
Clay_perc
Organic_material_perc
Phosphorus
Nitrogen_perc
Cobber
Average_drought_index
Magnesiu
Average_temp.
Potassium
Rt
MDS2
MDS1
1
0
-1
2

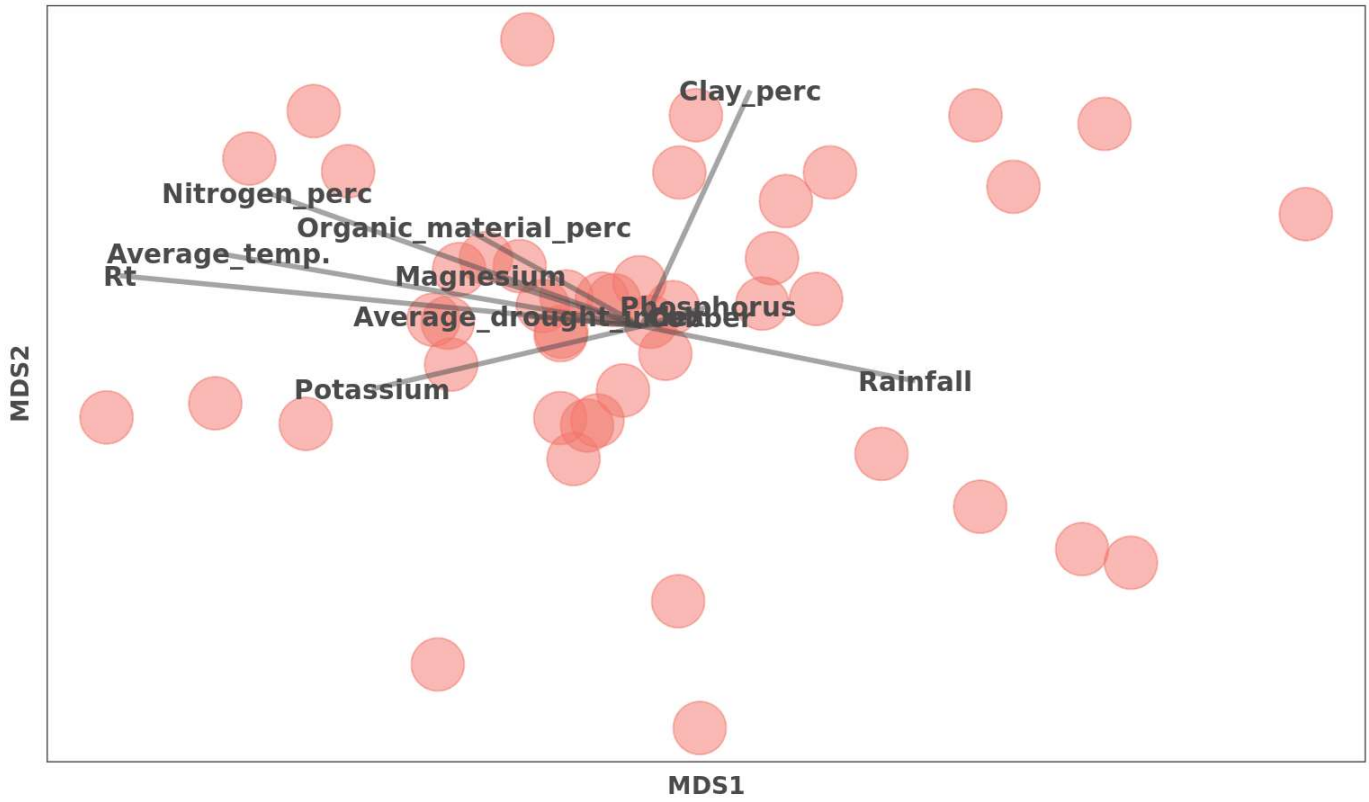

JB 5-7; Bray Curtis
Rainfall
Phosphorus
Cobber
Average_drought_index
Potassium
Average_temp.
Rt
Magnesium
Organic_material_perc
Nitrogen_perc
Clay_perc
MDS2
MDS1
2
1
0
-1

Jaccard

JB 1-2; Jaccard



JB 5-7; Jaccard



Aitchison

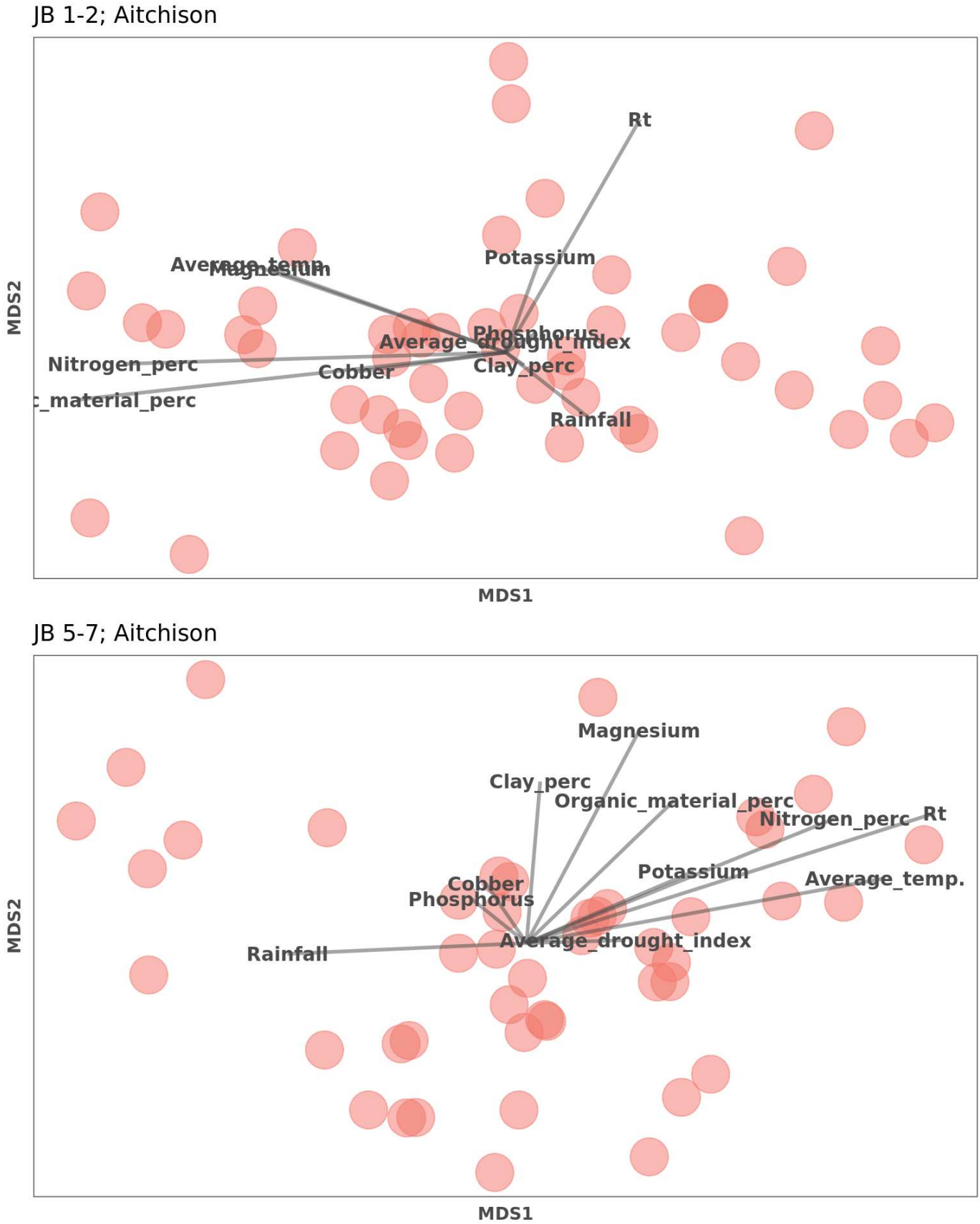


Figure 48: Visualization of the ENVFIT analysis. The association of the evaluated environmental variables with the ordination axis is displayed by arrows, with longer arrows indicating a stronger association.

ENVFIT summary statistics tables

ENVFIT for JB 1 & JB2

.	Bray		Jaccard		Euclidean	
Variable	**R2**	**p**	**R2**	**p**	**R2**	**p**
Rainfall	0.005	0.878	0.05	0.282	0.062	0.201
Average drought index	0	0.994	0	0.993	0.001	0.978
Average temp.	0.05	0.285	0.23	0.003	0.217	0.002
Rt	0.119	0.05	0.338	0.001	0.61	0.001
Phosphorus	0.023	0.52	0.04	0.391	0.006	0.859
Potassium	0.121	0.035	0.05	0.277	0.098	0.073
Magnesium	0.312	0.001	0.177	0.004	0.19	0.006
Cobber	0.018	0.628	0.069	0.188	0.05	0.305
Organic material (%)	0.08	0.12	0.346	0.001	0.406	0.001
Clay (%)	0.062	0.202	0.042	0.357	0.002	0.942
Nitrogen (%)	0.045	0.318	0.252	0.002	0.305	0.003

Table 32: Results from ENVFIT analysis across JB1 and JB2 fields. The table shows results from ENVFIT analyses including samples from all fields. The table shows the obtained R-squared values and the corresponding p-values from the ENVFIT analysis, for each of the assessed variables in each of the three beta-diversity ordinations.

ENVFIT for JB 1 & JB2

.	Bray		Jaccard		Euclidean	
Variable	**R2**	**p**	**R2**	**p**	**R2**	**p**
Rainfall	0.091	0.133	0.104	0.095	0.127	0.047
Average drought index	0.022	0.63	0.016	0.722	0.022	0.646
Average temp.	0.133	0.046	0.242	0.004	0.31	0.001
Rt	0.179	0.015	0.35	0.002	0.475	0.001
Phosphorus	0.022	0.631	0.007	0.847	0.019	0.641
Potassium	0.025	0.556	0.105	0.113	0.094	0.114
Magnesium	0.044	0.369	0.042	0.397	0.314	0.003

Cobber	0.008	0.839	0.005	0.923	0.026	0.561
Organic material (%)	0.065	0.262	0.074	0.2	0.177	0.023
Clay (%)	0.106	0.108	0.212	0.006	0.166	0.028
Nitrogen (%)	0.159	0.025	0.238	0.003	0.308	0.001

Table 33: Results from ENVFIT analysis across JB1 and JB2 fields. The table shows results from ENVFIT analyses including samples from all fields. The table shows the obtained R-squared values and the corresponding p-values from the ENVFIT analysis, for each of the assessed variables in each of the three beta-diversity ordinations.

ENVFIT for climate

Note above we have adjusted for year in all ENVFIT analyses. But this is not ideal for the climate variables as they often are a direct consequence of yearly differences and thus we remvoe the effect we are looking the study. Therefore, here we eprform ENVFIT analyses again but only for climate variables and this time without adjustign for year.

Ordination plots | ENVFIT summary statistics tables

Bray Curtis | Jaccard | Aitchison

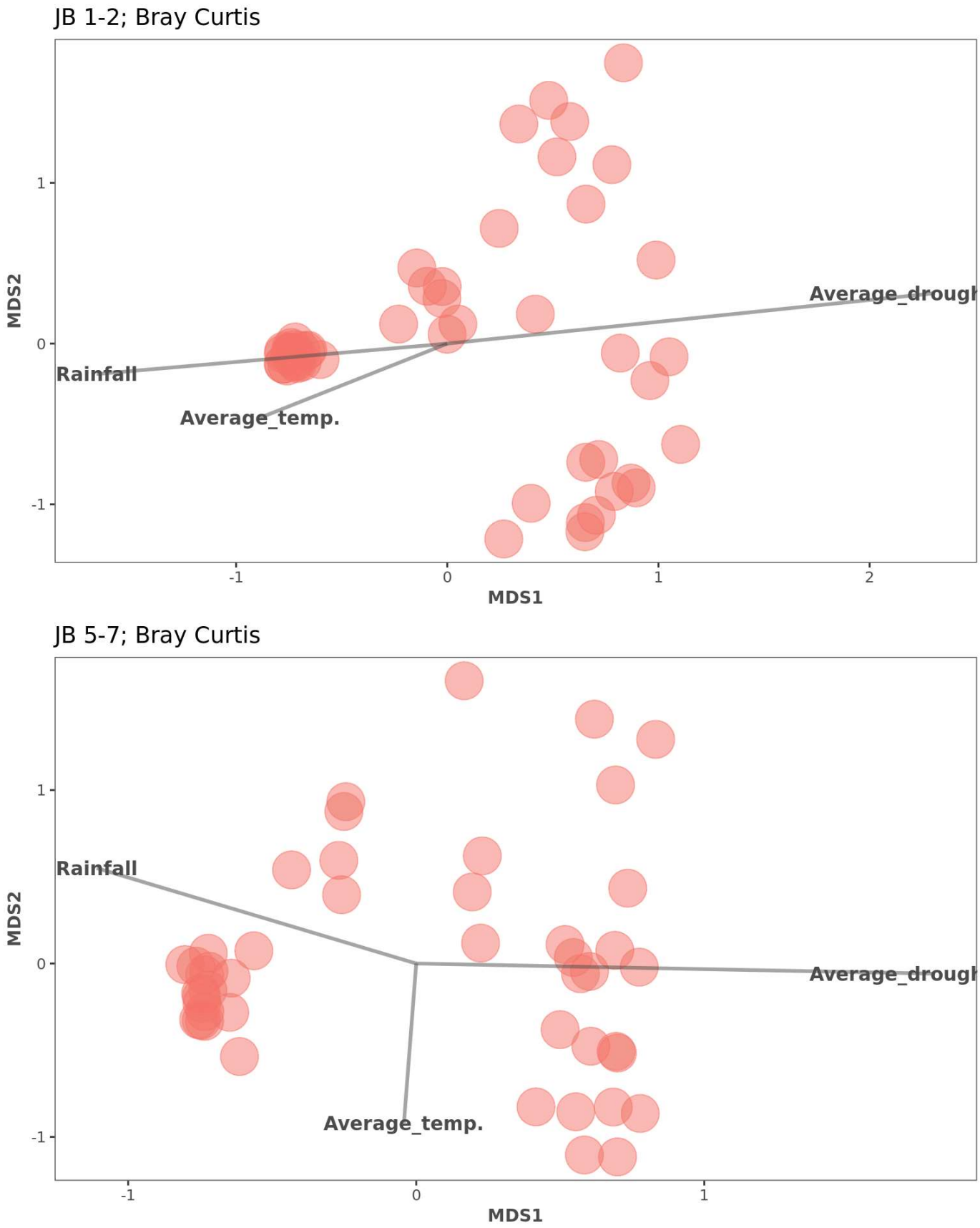


Figure 49: Visualization of the ENVFIT analysis for 3 climate variables. The association of the evaluated environmental variables with the ordination axis is displayed by arrows, with longer arrows indicating a stronger association.

Non-linear relationships for continuous environmental variables

The ENVFIT model looks for linear relationships between the main variation of the microbiome (the separation of the samples along the first ordination axes) and the environmental variables. It can also be relevant to look for non-linear relationships - so called smooth surfaces - where the environmental variable for example is

highest in the centrally located samples on the ordination plot and lower in samples located towards the edges. We use a mode called Ordisurf to look for such non-linear patterns. Ordisurf internally uses generalized additive models with integrated smoothness estimation, so-called GAMs, to fit the variable. Below is a table of the results from the analysis looking for smooth surfaces for the selected variables and further down are the surfaces illustrated on ordination plots.

Euclidean		
	F	**P**
Rainfall	1.249	1.75e-02
Average_drought_index	1.656	5.37e-03
Average_temp.	1.324	2.15e-03
Rt	9.404	0.00e+00
Phosphorus	0	8.68e-01
Potassium	0.39	7.29e-02
Magnesium	2.583	4.75e-04
Cobber	0.085	2.74e-01
Organic_material_perc	4.102	9.25e-07
Clay_perc	0	9.90e-01
Nitrogen_perc	2.627	6.41e-05

Table 36: Results from non-linear analysis. The table shows results from the non-linear analyses performed using Ordisurf for samples JB1 and JB 2. The table shows the obtained F values and the corresponding p-values, for each of the assessed variables based on the Aitchison beta-diversity.

Euclidean		
	F	**P**
Rainfall	2.269	1.56e-03
Average_drought_index	4.567	1.64e-05
Average_temp.	1.918	3.49e-04
Rt	4.104	1.51e-06
Phosphorus	0	5.87e-01
Potassium	0.34	1.10e-01

Magnesium	2.108	2.02e-04
Cobber	0.132	3.38e-01
Organic_material_perc	0.805	1.59e-02
Clay_perc	0.695	2.16e-02
Nitrogen_perc	1.879	3.91e-04

Table 37: Results from non-linear analysis. The table shows results from the non-linear analyses performed using Ordisurf for samples JB5, JB6 and JB 7. The table shows the obtained F values and the corresponding p-values, for each of the assessed variables based on the Aitchison beta-diversity.

Version information

Table 38: List of used software including the used R-programming environment packages.

Package	Version	Package	Version
OS	Ubuntu 20.04.4 LTS	**utf8**	1.2.4
R	4.3.3	**generics**	0.1.3
splines	4.3.3	**robustbase**	0.99-2
bitops	1.0-7	**class**	7.3-22
lifecycle	1.0.4	**httr**	1.4.7
rstatix	0.7.2	**htmlwidgets**	1.6.4
sf	1.0-16	**S4Arrays**	1.2.1
MASS	7.3-60.0.1	**pkgconfig**	2.0.3
insight	0.19.10	**gtable**	0.3.5
backports	1.4.1	**hwriter**	1.3.2.1
magrittr	2.0.3	**pcaPP**	2.0-4
plotly	4.10.4	**htmltools**	0.5.8.1
sass	0.4.9	**biomformat**	1.30.0
rmarkdown	2.26	**png**	0.1-8
jquerylib	0.1.4	**rstudioapi**	0.16.0
yaml	2.3.8	**tzdb**	0.4.0
zip	2.3.1	**reshape2**	1.4.4
cowplot	1.1.3	**coda**	0.19-4.1
DBI	1.2.2	**nlme**	3.1-164
minqa	1.2.6	**curl**	5.2.1

Package	Version	Package	Version
ade4	1.7-22	**nloptr**	2.0.3
multcomp	1.4-25	**proxy**	0.4-27
abind	1.4-5	**cachem**	1.0.8
zlibbioc	1.48.2	**zoo**	1.8-12
Rtsne	0.17	**rhdf5**	2.46.1
RCurl	1.98-1.14	**sjlabelled**	1.2.0
TH.data	1.1-2	**KernSmooth**	2.23-22
sandwich	3.1-0	**parallel**	4.3.3
GenomeInfoDbData	1.2.11	**pillar**	1.9.0
ggrepel	0.9.5	**vctrs**	0.6.5
units	0.8-5	**ggpubr**	0.6.0
svglite	2.1.3	**xtable**	1.8-4
codetools	0.2-20	**cluster**	2.1.6
DelayedArray	0.28.0	**evaluate**	0.23
xml2	1.3.6	**mvtnorm**	1.2-4
tidyselect	1.2.1	**cli**	3.6.2
farver	2.1.1	**compiler**	4.3.3
multtest	2.58.0	**rlang**	1.1.3
e1071	1.7-14	**crayon**	1.5.2
survival	3.5-8	**ggsignif**	0.6.4
iterators	1.0.14	**rrcov**	1.7-5
systemfonts	1.0.6	**labeling**	0.4.3
foreach	1.5.2	**interp**	1.1-6
tools	4.3.3	**classInt**	0.4-10
glue	1.7.0	**plyr**	1.8.9
SparseArray	1.2.4	**stringi**	1.8.4
xfun	0.43	**viridisLite**	0.4.2
mgcv	1.9-1	**deldir**	2.0-4
withr	3.0.0	**munsell**	0.5.1
fastmap	1.1.1	**lazyeval**	0.2.2
latticeExtra	0.6-30	**V8**	4.4.2
boot	1.3-30	**hms**	1.1.3

Package	Version	Package	Version
rhdf5filters	1.14.1	**Rhdf5lib**	1.24.2
fansi	1.0.6	**highr**	0.10
digest	0.6.35	**broom**	1.0.5
timechange	0.3.0	**igraph**	2.0.3
R6	2.5.1	**RcppParallel**	5.1.7
estimability	1.5	**bslib**	0.7.0
colorspace	2.1-0	**DEoptimR**	1.1-3
jpeg	0.1-10	**ape**	5.8