# 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 | ITS2 |

# 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 | 1950 | 5.14 | 35.53 |
| R10 | 1853 | 5.19 | 46.01 |
| R11 | 1682 | 5.49 | 91.42 |
| R12 | 2043 | 5.58 | 78.08 |
| R13 | 1630 | 5.41 | 69.86 |
| R14 | 2235 | 5.75 | 94.51 |
| R2 | 2208 | 5.60 | 81.16 |
| R3 | 2279 | 5.34 | 61.64 |
| R4 | 1743 | 5.34 | 51.64 |
| R5 | 1738 | 5.39 | 71.40 |
| R6 | 1848 | 5.54 | 70.40 |
| R7 | 1790 | 5.47 | 68.61 |
| R8 | 1616 | 5.07 | 34.67 |
| R9 | 1838 | 5.58 | 93.90 |

**Table 2: Biodiversity across samples.** Table showing 3 different biodiversity measures for each sample.

**Observations and notes**

I have also added the biodiversity variables from both fungi and procaryotes (bacteria and archaea, 16S) to an excel file with the metadata variables, where I have applied conditional formatting marking higher values green and lower values red. This helps review the scale of each biodiversity value along with the metadata and field information. This is in file “Metadata_added_microbiome_variables.excl”. Shannon and InvSimpson correlates notably (cor. r between 0.8-0.9) but the InvSimpson can be easier to evaluate manually as the scale is longer so differenes are not in the digits and thus easier to spot.

Difference Between Diversity and Richness

- Inverse Simpson Index: Measures diversity by accounting for both species richness and evenness. A higher value indicates not just many species but a more even distribution of their abundances.
- Observed Richness: Counts the number of unique taxa (e.g., Species, Genera, ASVs). It does not consider evenness.

We see a higher observed fungal diversity (number of different taxa) in the forest farm (Nybordgaard) and in two field of Øm Kloster (the Biomark and Klimamark).

The forest farm also has high prokaryotic diversity and so has Niels Hansens pløjefri farm with high InvSimpson in both R9 and R11, but the highest observed diversity and prokaryotic load in R10 (pløjefri +11 years). R9 and R11 may have more evenly distributed microbial communities, leading to higher diversity scores despite potentially fewer total species.

Within the farm of Anders Knudsen, we see a higher fungal diversity in the bad part of the field than in the good (R6 vs R7). The difference is not big but could be expected to be opposite. In bacteria/Archaea diversity the field has a fairly even diversity and the field with ‘rotationsafgræsning, R8’ has a higher diversity and procaryotic load (qPCR results). The field ‘rotationsafgræsning, R8’ has a low fungal diversity.

## Statistical assessment

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

| Variable | Estimate | std.err | t.value | P |
|---|---|---|---|---|
| **Rt** | **-0.0073852** | **0.025** | **-0.295** | **7.77e-01** |
| **Fosfor** | 0.0141038 | 0.011 | 1.228 | 2.43e-01 |
| **Kalium** | 0.0000151 | 0.004 | 0.003 | 9.97e-01 |
| **Magnesium** | 0.0054527 | 0.007 | 0.760 | 4.63e-01 |
| **Kobber** | 0.0165069 | 0.013 | 1.272 | 2.62e-01 |
| **Organisk_stof** | 0.0427543 | 0.018 | 2.361 | 4.30e-02 |
| **Lerindhold_perc** | -0.0018469 | 0.003 | -0.539 | 6.07e-01 |

| Variable | Estimate | std.err | t.value | P |
|---|---|---|---|---|
| C.N_forhold | 0.0268339 | 0.009 | 2.872 | 1.40e-02 |
| P.afgrøde_lager | 0.0000768 | 0.000 | 0.956 | 3.60e-01 |
| Ca.plante_tilgængelig | -0.0000631 | 0.000 | -0.884 | 3.96e-01 |
| Total_Ca_jordlager | -0.0000002 | 0.000 | -0.035 | 9.72e-01 |
| Ler.humus | 0.0001047 | 0.000 | 0.257 | 8.02e-01 |
| Ombyttelig_CEC | -0.0017018 | 0.003 | -0.584 | 5.72e-01 |
| Mikrobiel_biomasse | 0.0004023 | 0.000 | 1.919 | 7.93e-02 |
| Mikrobiel_aktivitet | 0.0016611 | 0.001 | 1.943 | 7.74e-02 |
| Svampe.bakterie_forhold | 0.1362424 | 0.067 | 2.043 | 7.09e-02 |

**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 metadata variables (rows).

**Observations and notes**

We see that higher organic material associate with a higher fungal diversity and the same does the C/N ratio. There is a positive association between fungal diversity and the 3 microbial measures from Eurofins but it is not significant at $p<0.05$.

# 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 | g__Acremonium | g__Apiotrichum | g__Aspergillus | g__Cheilymenia | g__Cladosporium | g__Clonostachys | g__Exophiala | g__Extre |
|---|---|---|---|---|---|---|---|---|
| R1 | 1.29 | 0.81 | 1.64 | 2.15 | 2.64 | 0.93 | 0.53 | |
| R2 | 0.52 | 0.93 | 0.96 | 1.07 | 2.39 | 1.12 | 0.45 | |
| R3 | 0.92 | 1.20 | 0.78 | 4.41 | 2.63 | 1.08 | 0.39 | |
| R4 | 0.63 | 1.17 | 2.21 | 10.71 | 0.73 | 0.29 | 0.27 | |
| R5 | 0.65 | 0.41 | 1.26 | 4.32 | 0.84 | 1.27 | 0.56 | |
| R6 | 1.46 | 0.16 | 1.26 | 0.09 | 3.08 | 3.06 | 0.61 | |
| R7 | 1.83 | 1.00 | 1.17 | 0.00 | 2.10 | 4.91 | 0.56 | |
| R8 | 2.38 | 1.38 | 0.52 | 0.06 | 2.48 | 2.07 | 0.66 | |
| R9 | 1.41 | 0.48 | 1.15 | 0.09 | 3.90 | 2.81 | 1.82 | |
| R10 | 0.59 | 1.26 | 0.60 | 0.61 | 1.16 | 1.13 | 2.68 | |
| R11 | 0.50 | 1.24 | 1.36 | 0.72 | 2.56 | 1.05 | 2.17 | |
| R12 | 1.55 | 0.24 | 1.28 | 0.15 | 2.53 | 2.63 | 1.47 | |
| R13 | 2.09 | 3.14 | 1.66 | 2.87 | 4.39 | 2.98 | 0.43 | |
| R14 | 2.07 | 0.29 | 1.15 | 0.27 | 4.33 | 2.12 | 0.55 | |

**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.

Just one example is Spizellomyces: A Soil Chytrid Fungus. Within the Agernæs farm, we see this genus high in the good 'kålmark, R4' and low in the new field (R5). When reviewing what is known of this fungi that could make sense:

Spizellomyces belongs to the Chytridiomycota (chytrid fungi), which are primarily saprophytic, meaning they decompose organic matter.
They thrive in soils rich in organic inputs, such as fields with crop residues, compost amendments, or minimal disturbance.

Chytrid fungi play key roles in:

- Decomposing complex organic matter (e.g., cellulose).
- Recycling nutrients (carbon, nitrogen, and phosphorus) that are critical for plant growth.

# 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 fungal 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 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.

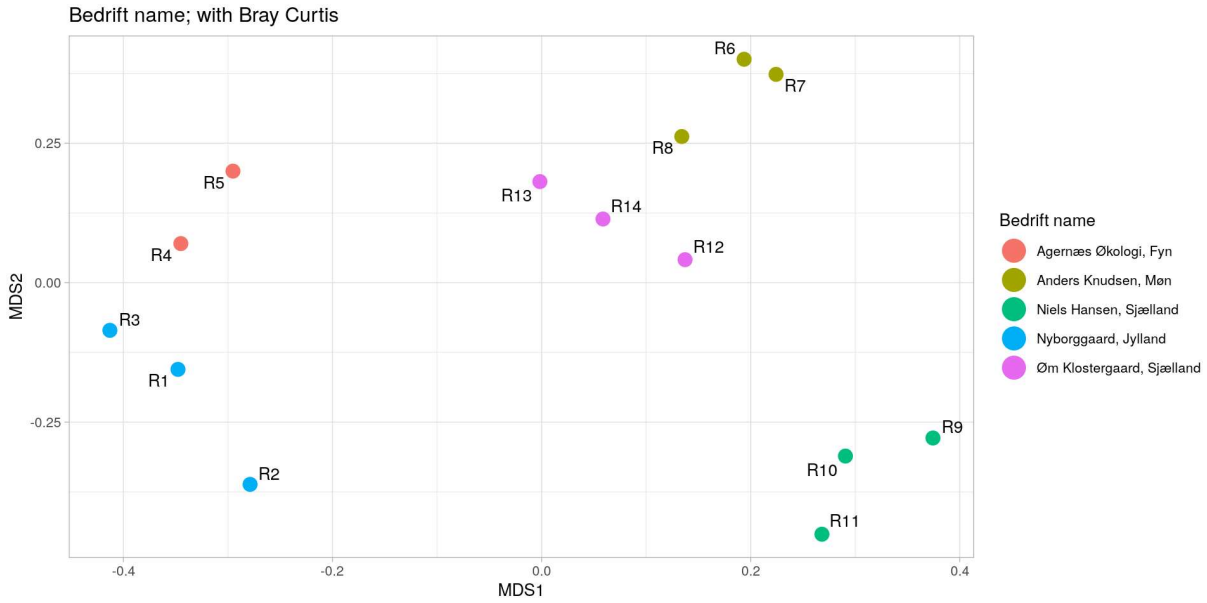


**Figure 1: Visualization of structure of the fungal community between the samples.** Ordination plot using bray-curtis beta-diversity. Dots are colored by farm as seen to the right of the figure panel and each sample is named on the plot.

**Observations and notes**

We see a strong effect of farm on the soil fungal community with samples clustering by farm. We also see that farms seperate along the x-axis with Fyn and Jylland clustering together away from the 3 other more eastern located farms. This the west to east axis is important for soil fungal communities in Denmark and more important than other differences between the fields and farms. When we remove the effect of farm the organisation changes as we can see in Figure 2. We see that R10 is very different from the otehr 2 samples from that farm and that R8 is very different from the other samples from that farm and also from the other fields analyzed.

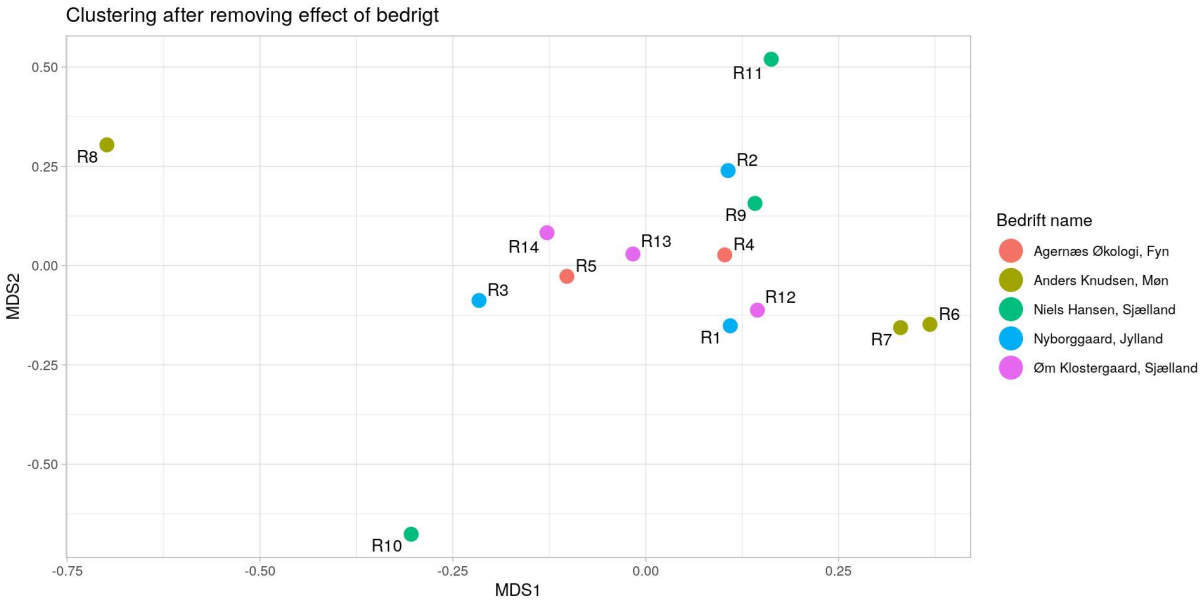


**Figure 2: Visualization of structure of the fungal community between the samples after removing the effect of farm.** Ordination plot using bray-curtis beta-diversity. Dots are colored by farm as seen to the right of the figure panel and each sample is named on the plot.

## 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 fungal community composition and Rt, fosfor, Kalium, organisk stof, lerinhold, and ombyttelig CEC.

We see a trending association between fungal community composition and magnesium, kobber, Total CA jordlager, Ler/humus CEC, and svampe/bakterie forhold.

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 | Svampe/bakterie forhold

| | Bray-Curtis | | Jaccard | | Aitchison | |
|---|---|---|---|---|---|---|
| **Taxa level** | **R2** | **p** | **R2** | **p** | **R2** | **p** |
| Phylum | 0.1846 | 0.164 | 0.0551 | 0.709 | 0.0549 | 0.965 |
| Class | 0.2279 | 0.108 | 0.0978 | 0.071 | 0.1113 | 0.071 |
| Order | 0.1819 | 0.024 | 0.1145 | 0.107 | 0.1315 | 0.051 |
| Family | 0.1823 | 0.003 | 0.1248 | 0.137 | 0.1508 | 0.004 |
| Genus | 0.1774 | 0.003 | 0.1317 | 0.059 | 0.1544 | 0.005 |
| ASV | 0.1758 | 0.002 | 0.1256 | 0.001 | 0.1385 | 0.002 |

**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 |
| **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 |

| Package | Version | Package | Version |
|---|---|---|---|
| **R6** | 2.5.1 | **DEoptimR** | 1.1-3 |
| **estimability** | 1.5.1 | **ape** | 5.8 |
| **colorspace** | 2.1-0 | | |