**#READ.ME**

This ensemble of scripts aims to compare eukaryotic soil biodiversity in six vegetation cover types and to determine the main drivers (soil properties, climate, vegetation).

Analyses are performed on eukaryotic 18S DNA sequences from 787 soil samples collected all over Europe as part of LUCAS 2018 (Land Use/Cover Area frame Survey), the largest European soil survey coordinated by the European Commission.

Beyond LUCAS data, the WordClim2 and CGIAR3 databases were used to obtain climatic data for each LUCAS site, and the NASA-USDA Enhanced SMAP Global Soil Moisture Data4 to obtain soil moisture information for each LUCAS sample.

Data analysis was conducted in r, using the following packages:

**LIST OF r PACKAGES AND VERSION**

Packages without specified installation are available on CRAN repository.

library(phyloseq) v1.34.0 – installation requires the package BiocManager (see here : https://www.bioconductor.org/packages/release/bioc/html/phyloseq.html)

library(vegan) v2.6.2

library(ranacapa) v0.1.0 - instructions of installation here: https://github.com/gauravsk/ranacapa

library(ggpubr) v0.4.0

library(dplyr) v1.0.9

library(SRS) v0.2.3

library(ggplot2) v3.3.6

library(forcats) v1.0.0

library(stats) v4.2.1

library(ade4) v1.7.19

library(ggord) v1.1.7 – instructions of installation here: https://fawda123.github.io/ggord/

library(adespatial) v0.3.20

library(sf) v1.0.7

library(readxl) v1.4.2

library(imputeTS) v3.3

library(betapart) v1.6

library(randomForest) v4.7.1.1

library(caret) v6.0.92

library(rfPermute) v2.5.

library(dttr2) v0.4.2

library(agricolae) v1.3.5

library(ranger) v0.14.1

library(ggrepel) v0.9.2

library(PMCMRplus) v1.9.4

library(dunn.test) v1.3.5

library(steprf) v1.0.2

library(VennDiagram) v1.7.3

library(ggtext) v0.1.2

library("rnaturalearth") v0.3.2

library("rnaturalearthdata") v0.1.0

library(heatmaply) v1.4.2

library(corrplot) v0.92

This folder includes the following scripts:

**1: ASV\_tables**

**2: Plots\_Fig1\_2\_3\_SRScurves**

**3: Tax\_analysis**

**4: Alpha diversity**

**5: RandomForest\_alpha\_div\_analysis**

**6: Beta diversity**

**1: ASV\_tables**

For the downstream analysis, a phyloseq file was created combining the ASV table, taxonomy table and the sample data.

This script permits to obtain the normalized ASV\_tables, once the raw data have been filtered based on the main land covers and biogeographical regions (shapefile of the regions is downloadable here:

<https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3>

ASVs that were not found at least in one site were filtered out. In the file, normalization was conducted with the Scaling by ranked subsampling (SRS).

**2: Fig1\_2\_3\_SRScurves**

The script comprises the code to produce Fig 1-3 + SRS curves (Fig. S12)

Figure 1: Distribution of the sampling sites along ecosystem types and sampling season including a spatio-temporal overview of the sampling campaigns performed at the 787 sites included in this study (CL1 = annual croplands, CL2 = permanent croplands, GL1 = managed grasslands, GL2 = unmanaged grasslands, WL1 = broadleaved woodlands, WL2 = coniferous woodlands), the distribution of sampling sites among the different ecosystem types investigated and the distribution of sampling sites according to sampling season.

Figure 2: 18S-DNA sequencing read results A Proportion of ASVs assigned to different domains, with the majority of sequence reads found to belong to eukaryotes. B Proportion of taxa assigned to the different eukaryotic kingdoms. C Proportion of taxa assigned to the different phyla.

Figure 3: Proportion of taxa (ASVs) shared between different ecosystem types.

**3: Taxonomy analysis**

The script comprises the code to produce Fig S1-S6:

Figure S1: Distribution of 10 most abundant phyla based on ASVs occurrences for (a) fungi and (b) protists.

Figure S2: Distribution of 10 most abundant classes based on ASVs occurrences for (a) fungi, (b) protists, (c) rotifers, (d) tardigrades, (e) nematodes, (f) arthropods and (g) annelids.

Figure S3: Distribution of 10 most abundant orders based on ASVs occurrences for for (a) fungi, (b) protists, (c) rotifers, (d) tardigrades, (e) nematodes, (f) arthropods and (g) annelids.

Figure S4: Distribution of 10 most abundant families based on ASVs occurrences for (a) fungi, (b) protists, (c) rotifers, (d) tardigrades, (e) nematodes, (f) arthropods and (g) annelids.

Figure S5: Distribution of 10 most abundant genera based on ASVs occurrences for (a) fungi, (b) protists, (c) rotifers, (d) tardigrades, (e) nematodes, (f) arthropods and (g) annelids.

Figure S6: Distribution of 10 most abundant species based on ASVs occurrences for (a) fungi, (b) protists, (c) rotifers, (d) tardigrades, (e) nematodes, (f) arthropods and (g) annelids.

**4: Alpha diversity**

The script permits to produce Fig 4, Fig S7 producing alpha diversity plots showing the diversity for different ecosystem types along a gradient of land-use intensity. Also, Kruskal-Wallis and Dunn post-hoc tests are conducted.

Figure 4: α-diversity of eukaryotic groups in different ecosystem types. Observed ASV-richness for fungi and protists and for animals (rotifers, tardigrades, nematodes, arthropods and annelids). Shannon index for fungi and protists, and for animals.

Figure S7: Results of Dunn post-hoc test for the effect of environmental variables on observed richness and Shannon diversity.

**5: Random forest analysis**

This script permits to determine the main drivers (soil properties, climate, vegetation, microbial, topography) and their effects on soil eukaryotic diversity (observed ASV richness, Shannon index). The script comprises the code to produce Fig 5, S8, S9:

Figure 5: Variable importance plots for observed richness and alpha diversity

Figure S8: Variable importance of (a) observed ASV-richness and (b) Shannon diversity for fungi, protists, rotifers, tardigrades and nematodes.

Figure S9: Variable importance (backwards selection) for the observed ASV diversity (Shannon diversity) for fungi, protists, tardigrades, and nematodes (from left to right).

**6: Beta diversity**

This script permits to determine the main drivers (as single-effects) of eukaryotic-diversity (community composition). It also explores effects of land-use perturbation on beta-diversity. The script comprises the code to produce Fig 6, S10, conducting Anova forward selection, test for VIF/AIC, producing Venn diagrams, conducting betadispers analysis, producing the adonis table

Figure 6: β-diversity of different eukaryotic groups explained by ecosystem properties. (a) dbRDA analysis showing the influence of ecosystem type and environmental variables on eukaryotic community structure. (b) Variation partitioning showing the explained variance of unique and shared effects of ecosystem properties on β-diversity at the sampled sites

Figure S10: Homogeneity of multivariate dispersion for the different ecosystem features, showing the highest dispersion for nematodes.

Figure S11: Venn diagrams for the different eukaryotic groups showing the variation of the β-diversity

Table S5: Betadisper ANOVA

Table S6: Betadisper distances

Table S7: Adonis permanova, permutest dispersion

**References**

1. Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. & Fernández‐Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur J Soil Sci 69, 140–153 (2018).
2. WordClim. Historical monthly weather data. (2022). <https://worldclim.org/data/monthlywth.html>
3. NASA-USDA Enhanced SMAP Global Soil Moisture Data. Accessed on 24 November 2022: https://developers.google.com/earth-engine/datasets/catalog/NASA\_USDA\_HSL\_SMAP10KM\_soil\_moisture#description