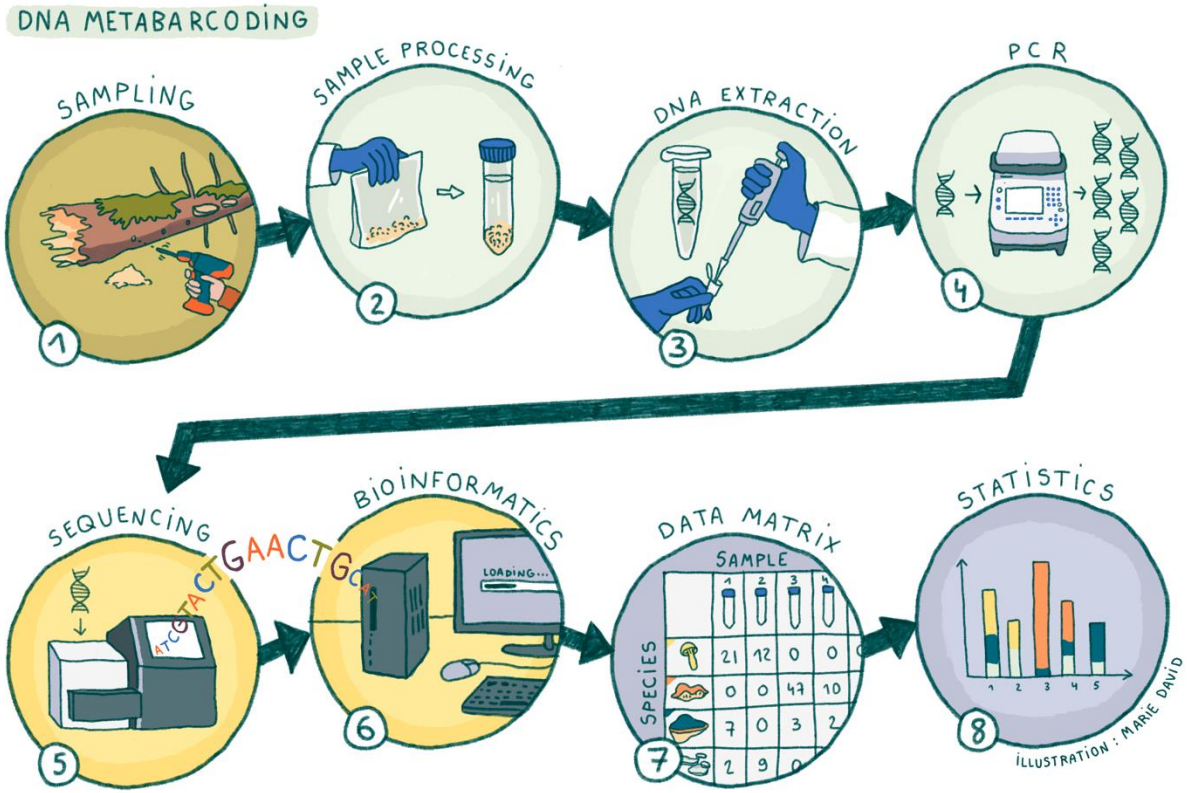
UNIVERSITY OF OSLO

Introduction to long-read metabarcoding



BIO9905MERG1 Spring 2025

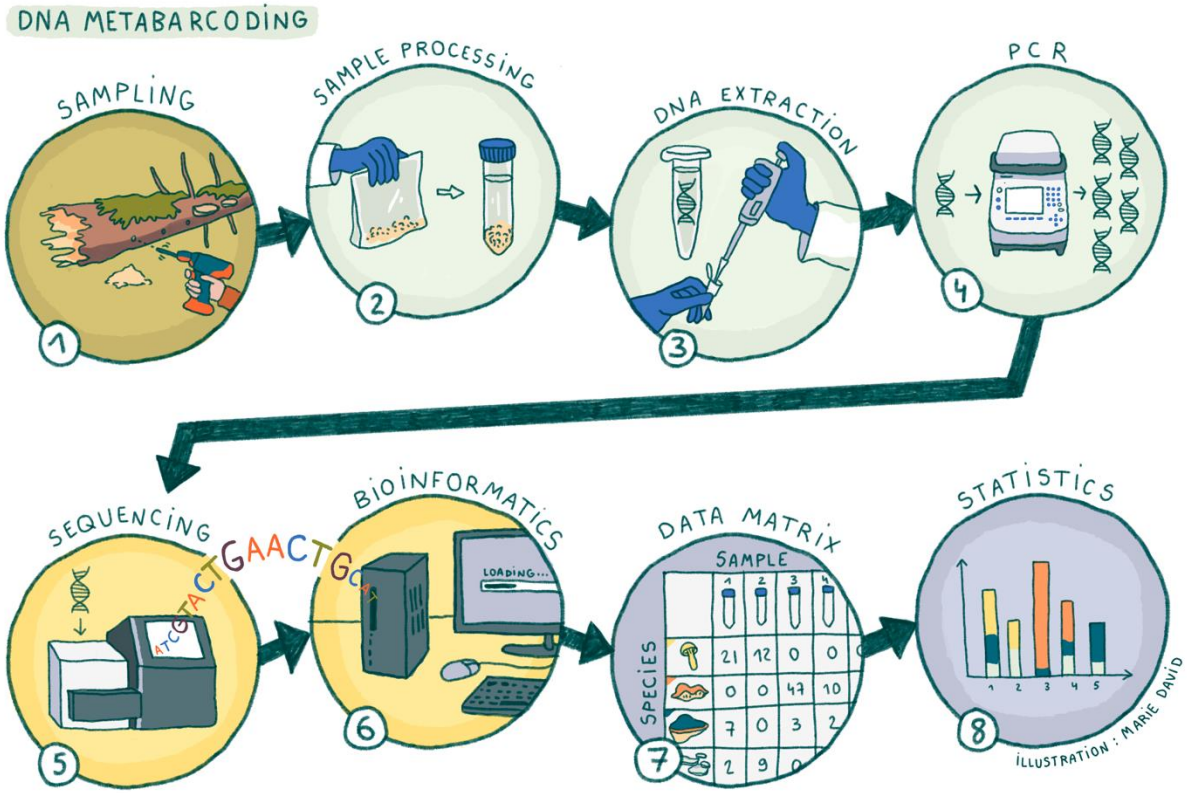
Embla Stokke, UiO

emblaes@student.matnat.uio.no

9/4-2025

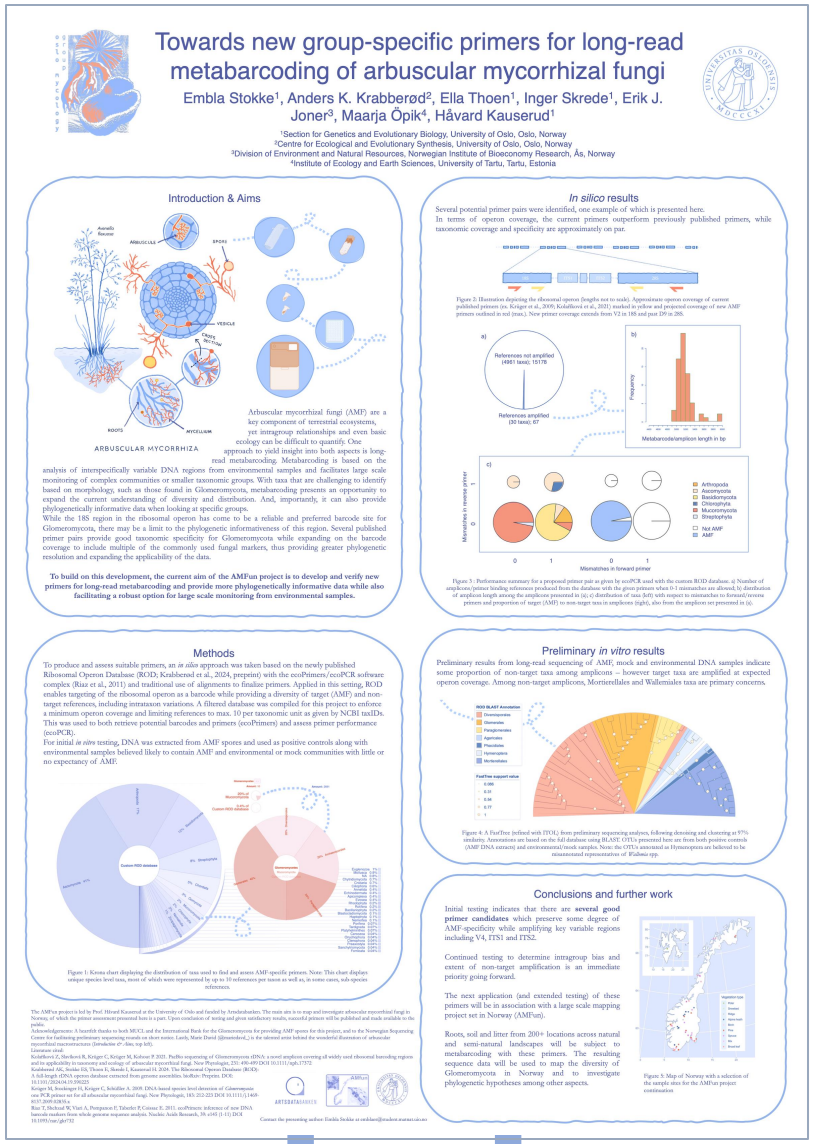
Topics

- Long read metabarcoding versus short read/regular metabarcoding
- Advantages of LRM
- Challenges with LRM
- Long read metabarcoding of soil eukaryotes



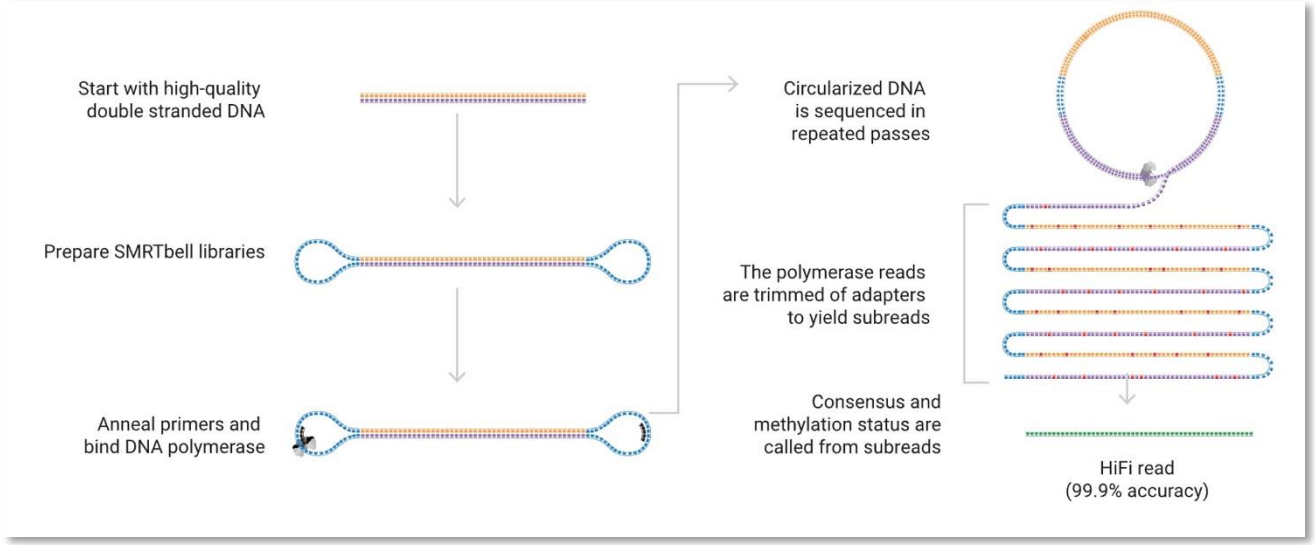
Background

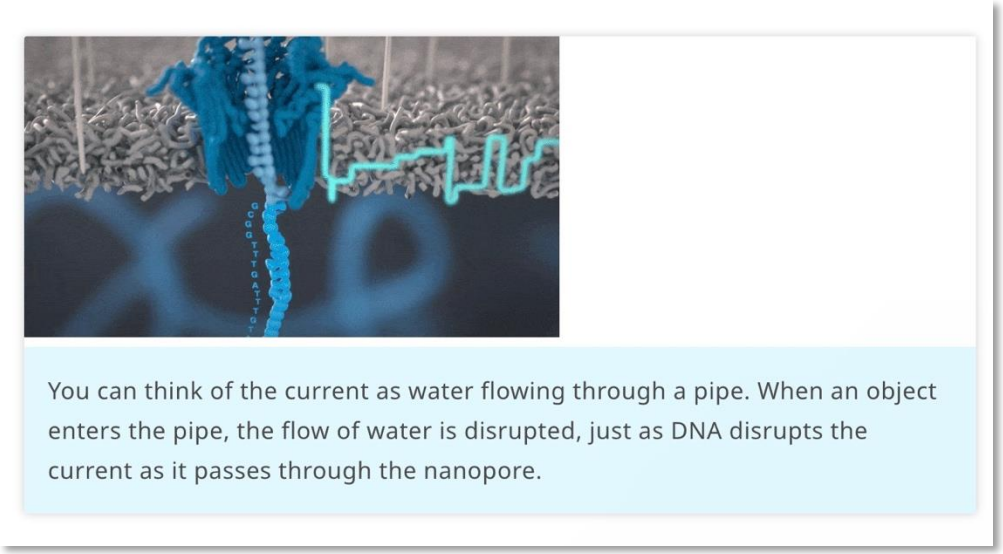
- MSc in Biodiversity & Systematics (NABiS/UiO)
- MSc project with OMG (Håvard Kauserud) – Long read metabarcoding of soil Eukaryotes
- PhD (ongoing) with OMG – Long read metabarcoding of Arbuscular Mycorrhizal Fungi



Long read vs short read

- Generally > 500 bp, 20 000 bp <
- Typically: singular marker vs multimarker
- Pacific Biosciences (PacBio)
- Oxford Nanopore
- Both provide
  - Continuous reads
  - Low sequencing error rate*
  - High yield





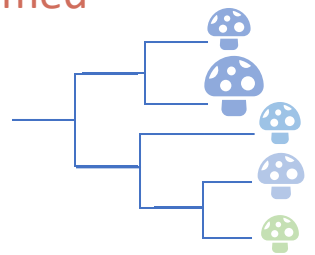
Advantages of LRM

- Longer reads
  - Better lineage resolution
- Consistent quality
- Less noise*
- Cross-database referencing
- Increased multiplex specificity with tagged libraries*

Taxa with tricky barcode gaps

Broad diversity of taxa with differing barcode optima

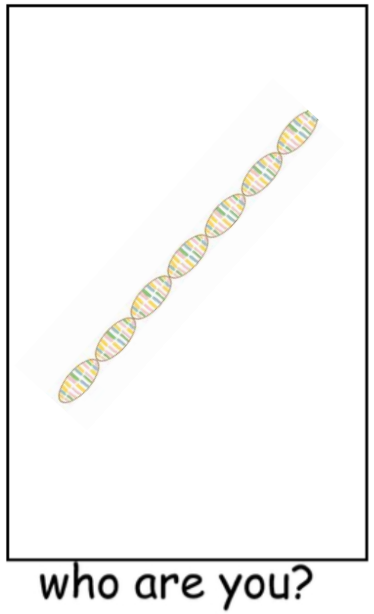
Phylogenetically informed studies

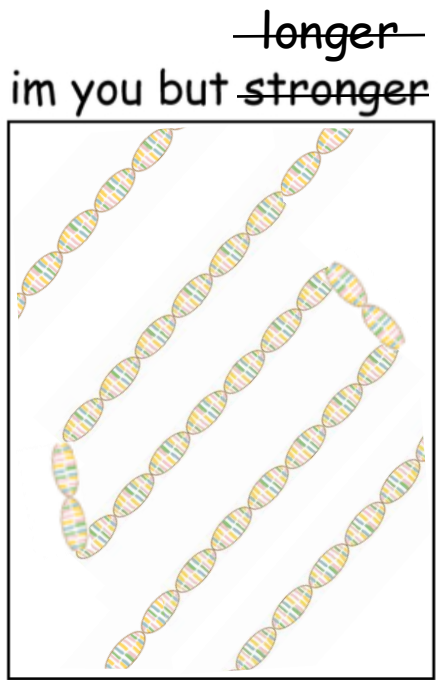


Challenges with LRM

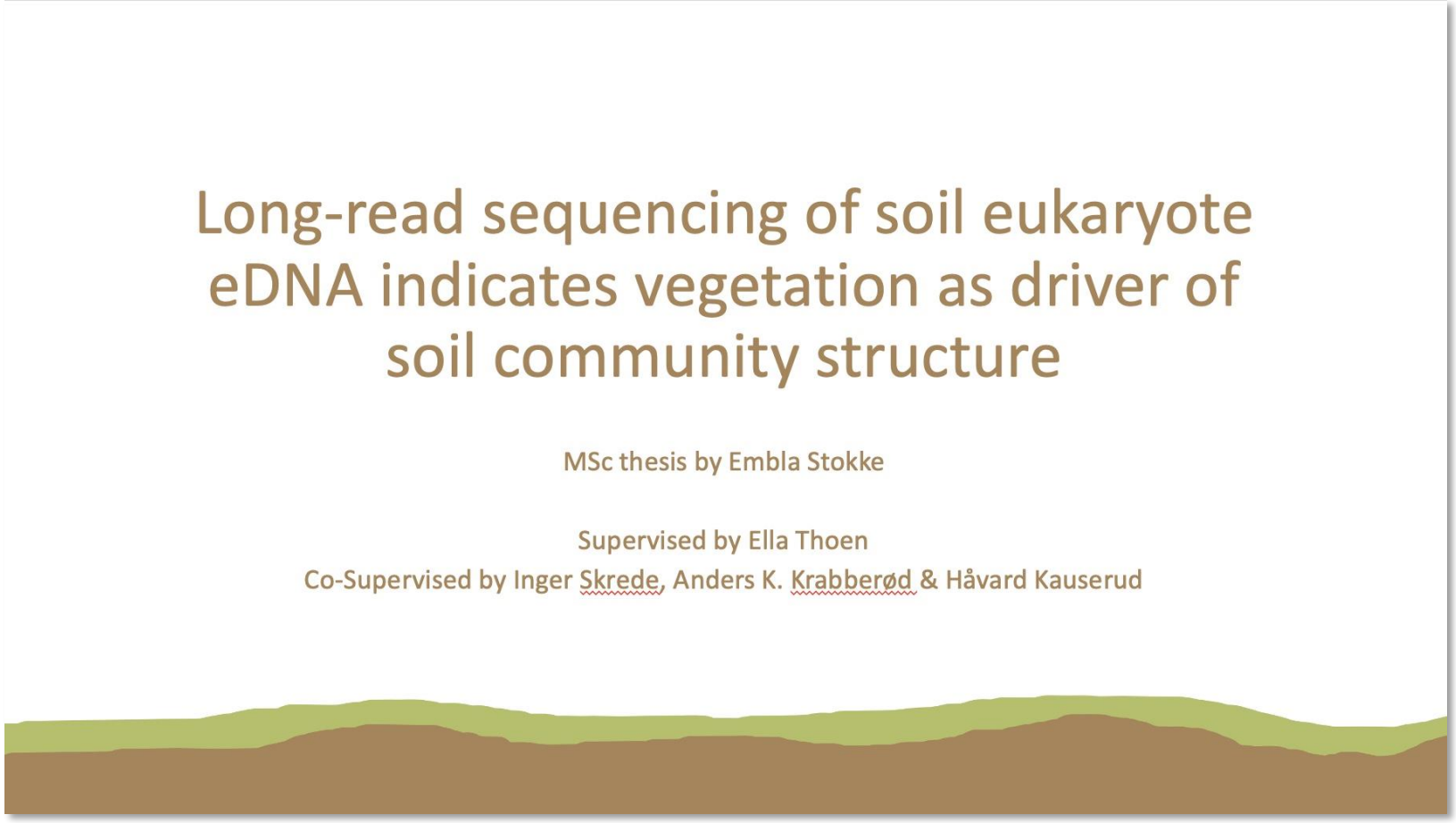
- Not just longer «short» amplicons
- Lab protocols
  - DNA extraction
  - PCR set-up
- Different artefacts/issues to regular amplicons
- Bioinformatics
  - Modified short-read approaches
  - Clustering/OTU synthesis
- Cost

I require a different approach to a number of aspects in the metabarcoding process and data interpretation





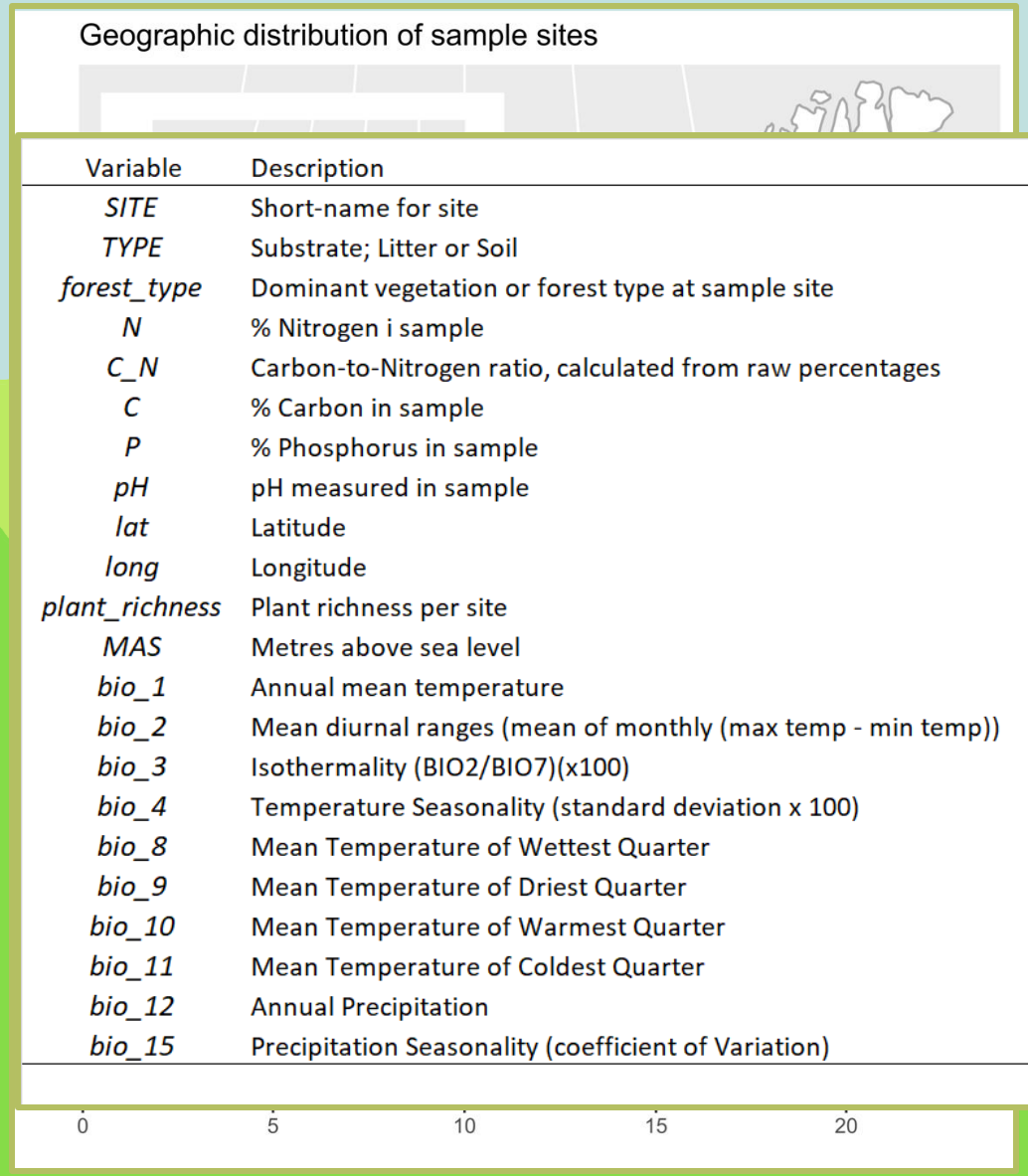
Long read metabarcoding of soil eukaryotes
An example of applied LRM*



Methods

Sample collection

- Archaeorhizomycetes Project
- Norway & Svalbard
- Diverse ecological landscapes
- 90 sample sites



Variable	Description
SITE	Short-name for site
TYPE	Substrate; Litter or Soil
forest_type	Dominant vegetation or forest type at sample site
N	% Nitrogen i sample
C_N	Carbon-to-Nitrogen ratio, calculated from raw percentages
C	% Carbon in sample
P	% Phosphorus in sample
pH	pH measured in sample
lat	Latitude
long	Longitude
plant_richness	Plant richness per site
MAS	Metres above sea level
bio_1	Annual mean temperature
bio_2	Mean diurnal ranges (mean of monthly (max temp - min temp))
bio_3	Isothermality (BIO2/BIO7)(x100)
bio_4	Temperature Seasonality (standard deviation x 100)
bio_8	Mean Temperature of Wettest Quarter
bio_9	Mean Temperature of Driest Quarter
bio_10	Mean Temperature of Warmest Quarter
bio_11	Mean Temperature of Coldest Quarter
bio_12	Annual Precipitation
bio_15	Precipitation Seasonality (coefficient of Variation)

Coring

Separation and pooling

Homogenizing

DNA extraction

Library assembly

Sequencing

180 samples

8 replicates

2 mock community samples

Serpula similis

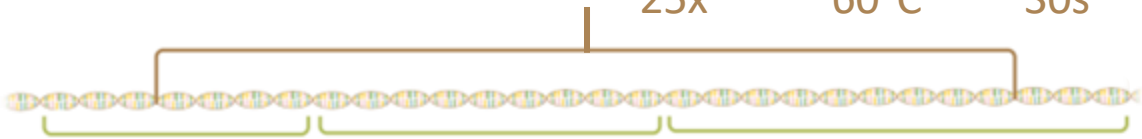
Pleurotus ostreatus

Prymnesium parvum

Amplification

98°C	30s	
	98°C	10s
25x	60°C	30s
	72°C	5 min
	10°C	Hold

Tagged eukaryote primers



SSU + $MgCl_2$ + ITS1 + 5.8S + ITS2 + LSU

Jamy et al., 2019, Mol. Ecol. Resour

cutadapt – linked adapter approach

Demultiplexing

VSEARCH – denovo chimera checking with conservative approach

Chimera checking

PR²

UNITE

NCBI

Taxonomic annotation

Sample inference and denoising

dada2 – filtering, denoising and sample inference

Clustering

VSEARCH – clustering at 97% similarity

Coherence between database annotations & phylogenetic groupings

Phylogenetic annotation check

Phylogenetic inference

SILVA concatenated reference sequences

MAFFT – sequence alignment

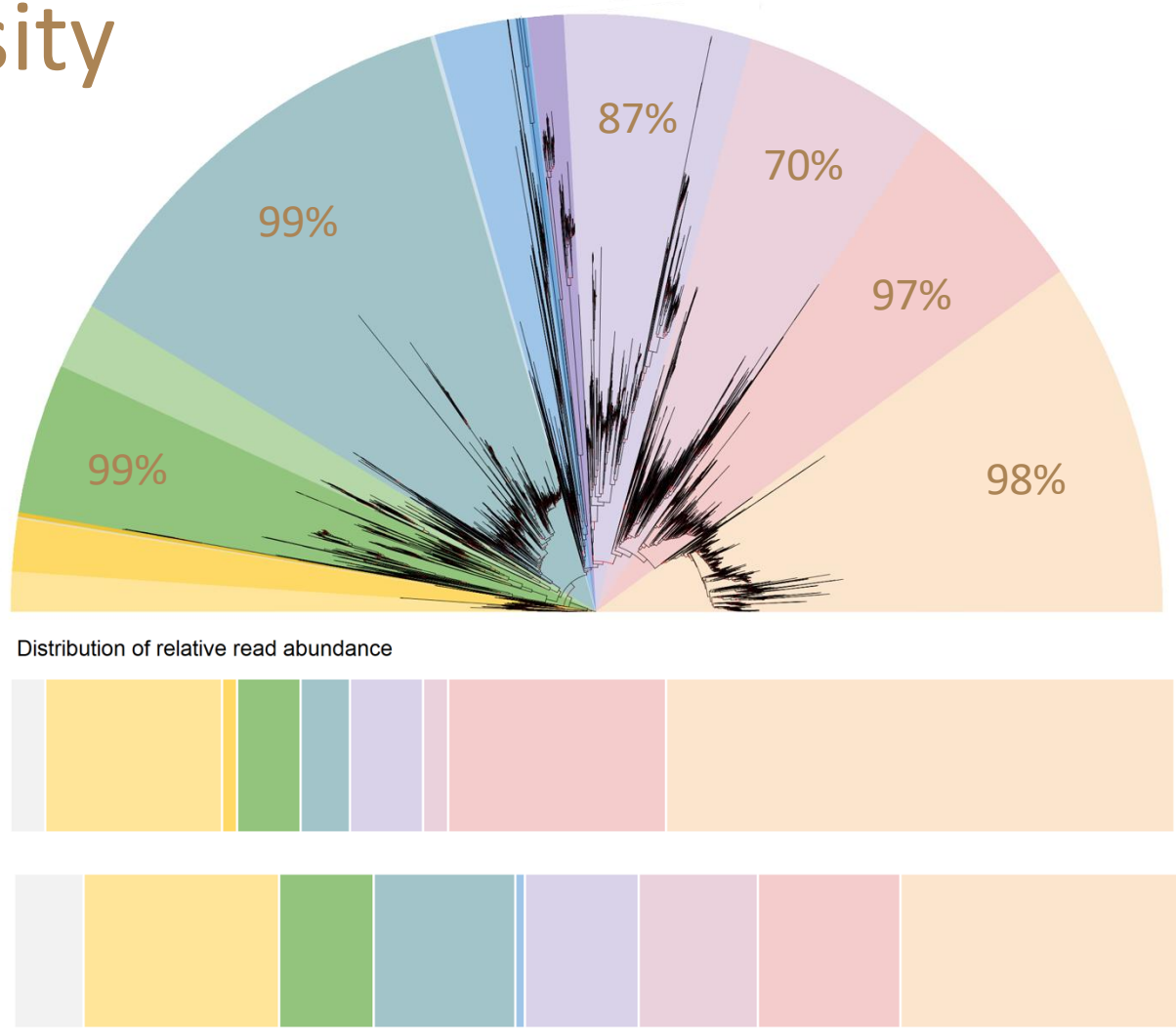
Geneious - 50% gaps & ITS removal

FastTree v.2 - tree building

Statistical analysis

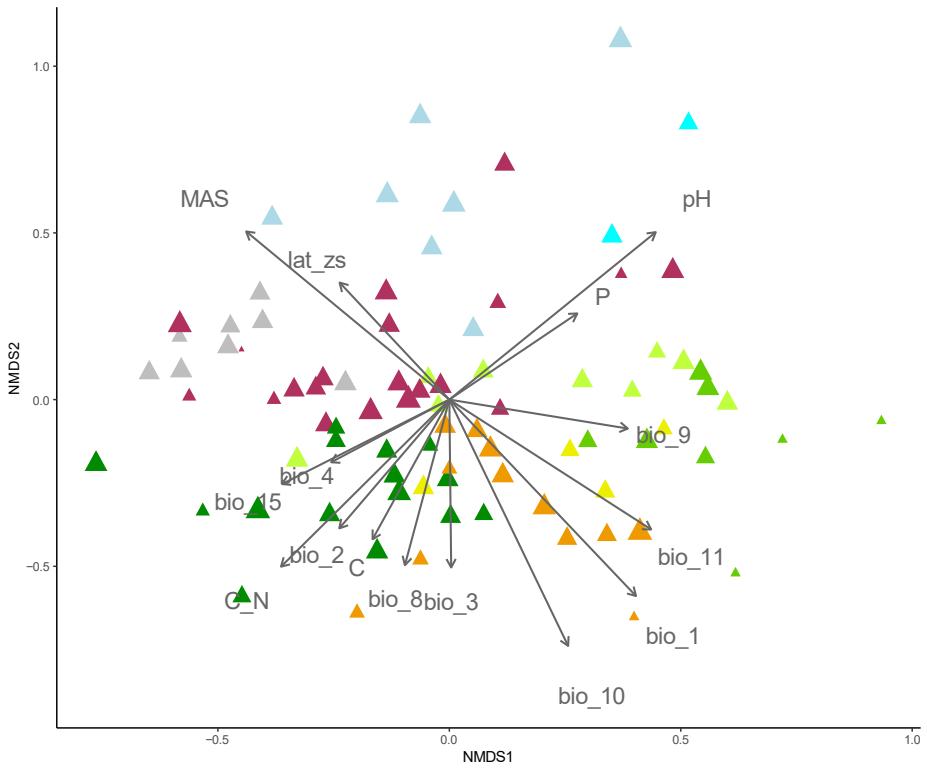
R v. 4.2.2 – various packages

Phylogeny & diversity

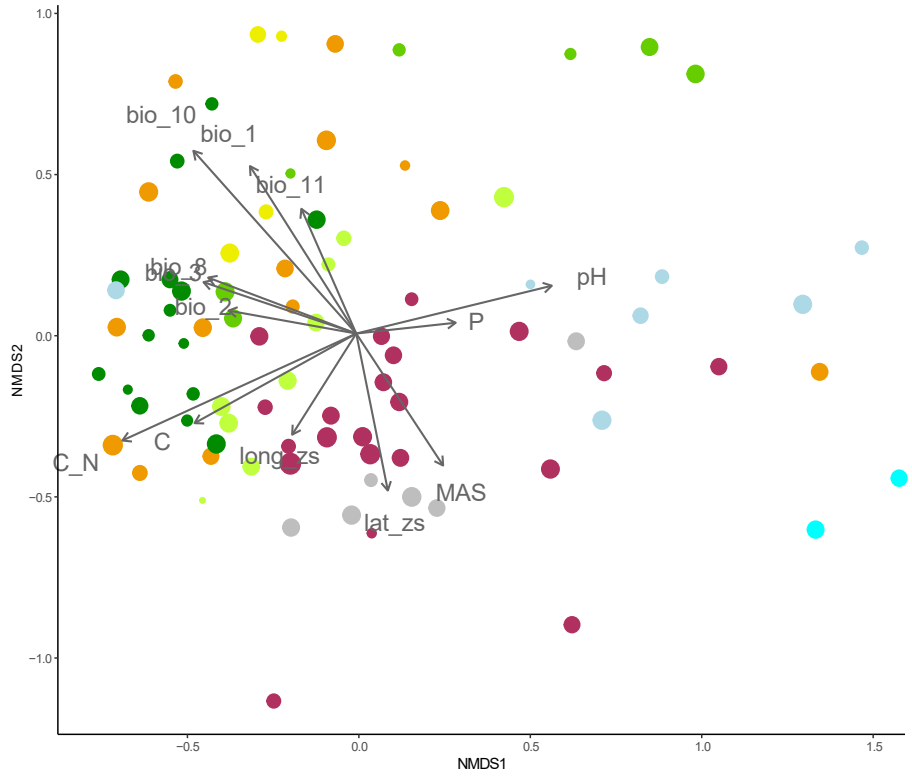


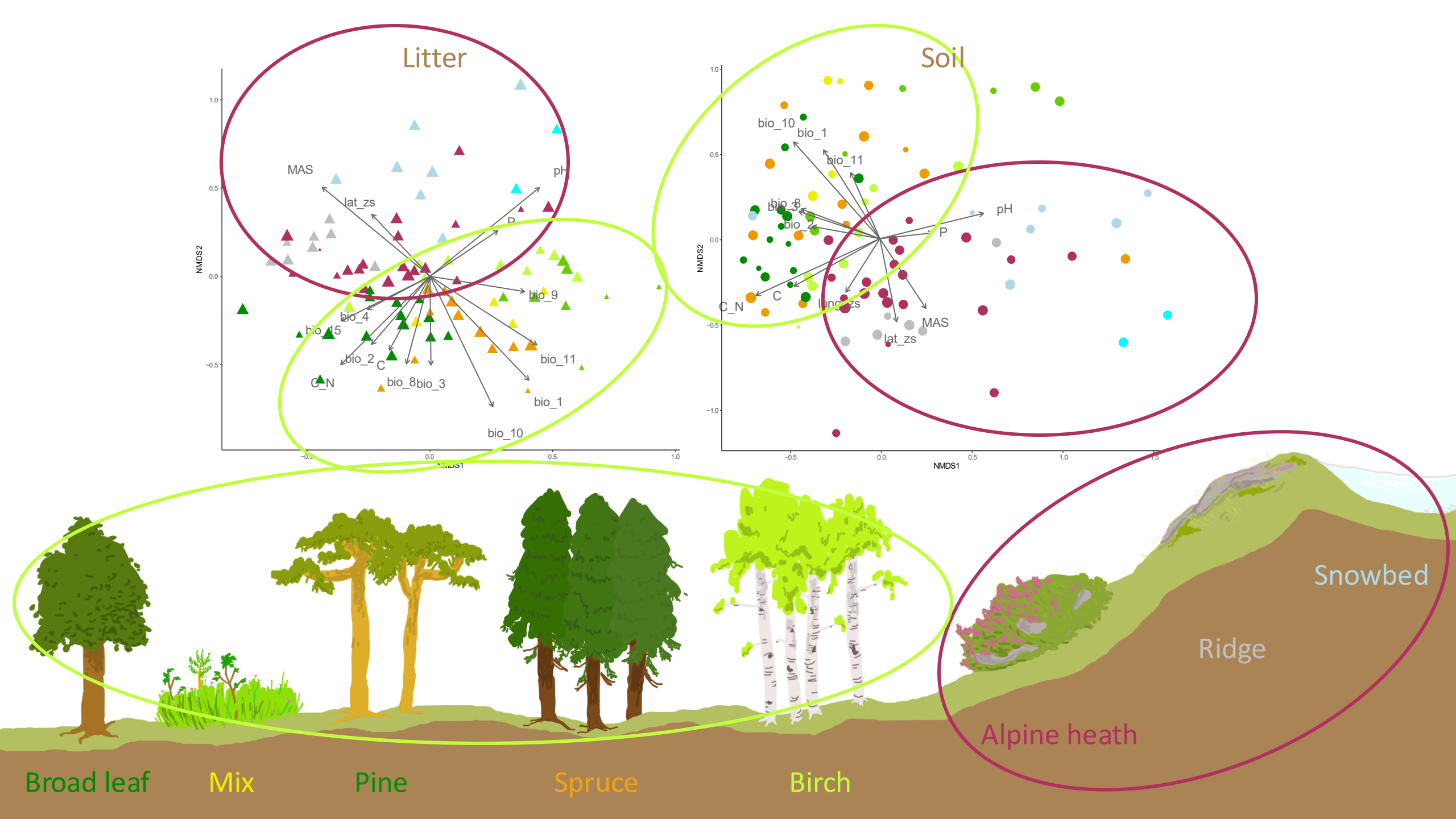
Beta diversity

Litter



Soil




Litter
Soil
MAS
lat_zs
pH
P
bio_9
bio_4
bio_15
bio_2
C
C_N
bio_8
bio_3
bio_11
bio_1
bio_10
NMDS1
NMDS2
Broad leaf
Mix
Pine
Spruce
Birch
Snowbed
Ridge
Alpine heath

Thank you for listening!

Questions?