



METABARCODING

METAPR2

Daniel Vaultot

2025-01-17

A world map with a light gray background and a white grid. Numerous blue dots of varying sizes are scattered across the map, representing the geographic distribution of metabarcodes. The dots are most densely clustered in Europe, North America, and parts of Asia, with smaller dots appearing in South America, Africa, and Australia. The text 'metaPR2: a database of metabarcodes' is overlaid in the center of the map.

metaPR2: a database of metabarcodes

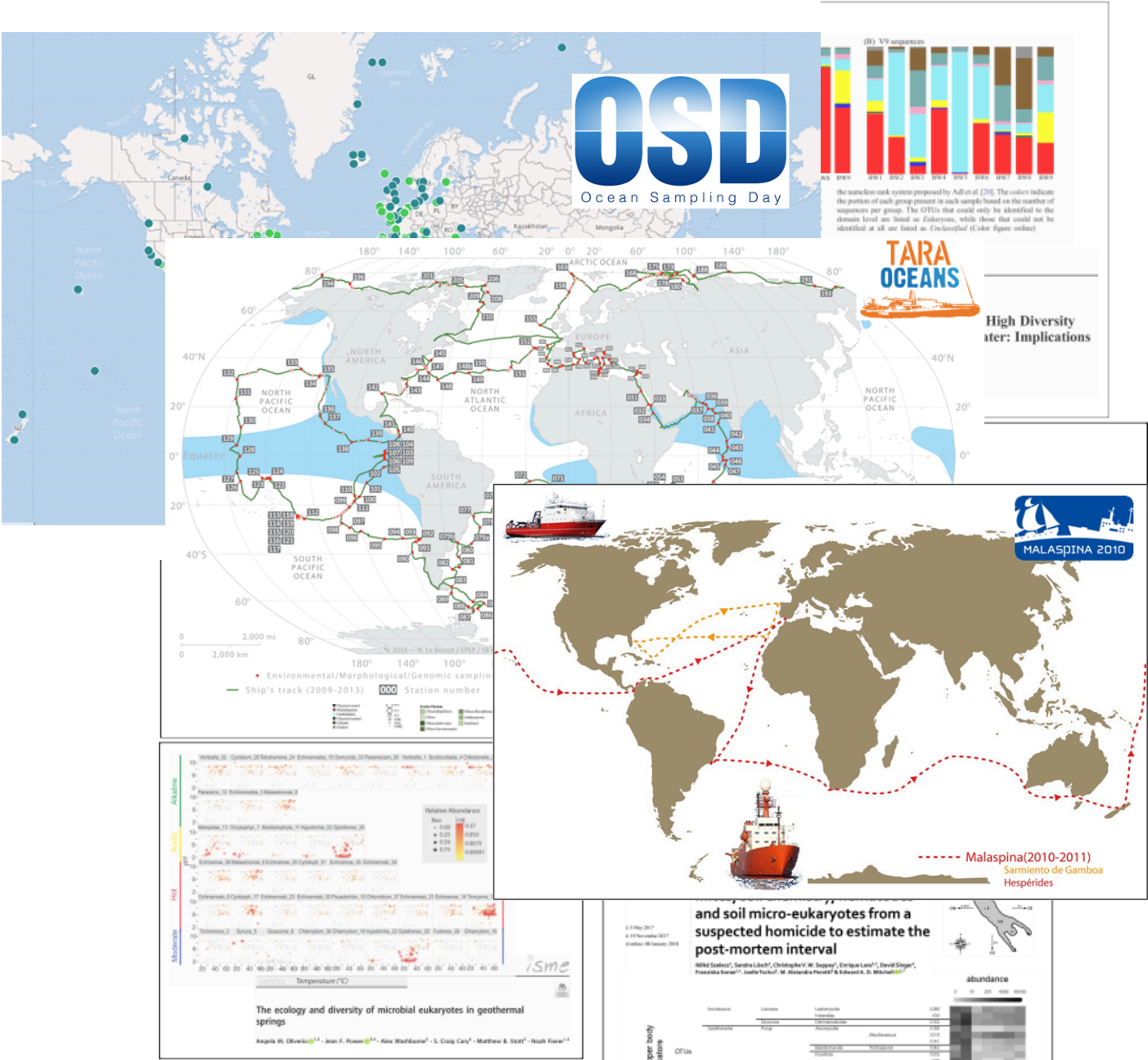
Many metabarcoding studies for eukaryotes

But hard to use...

- Processed with different pipelines
- Different primers
- Different levels of similarity
- Different reference databases
- Metadata lacking

Large datasets

- Ocean Sampling Day
- Tara Oceans
- Malaspina



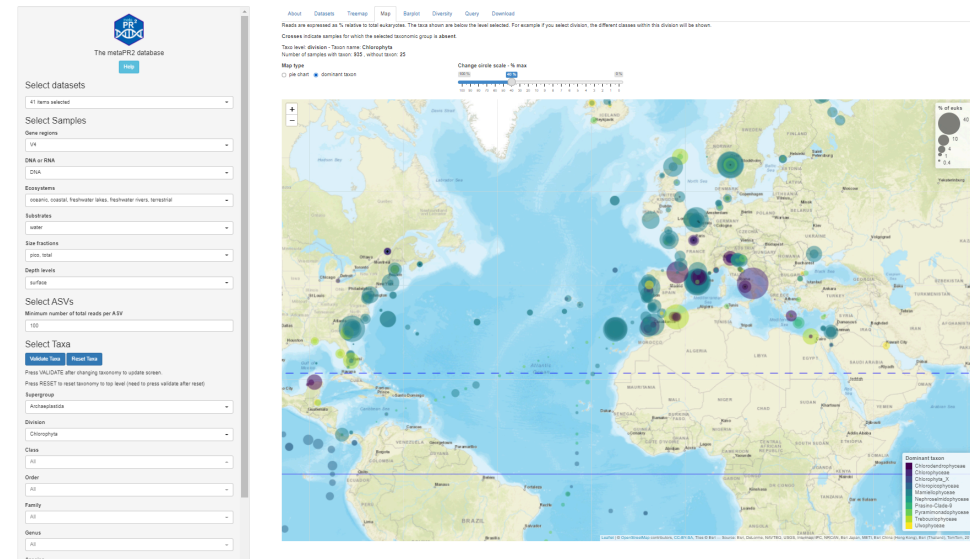
metaPR2 a database of metabarcodes

Reprocess public data

- Download Genbank (SRA) data
 - Raw sequences
 - Metadata
- Reprocess
 - Amplicon Sequence Variant (*dada2*)
 - Merge ASVs with same sequence
- Store in MySQL database
- Develop under R
- Web interface and R package
- <https://app.metapr2.org>

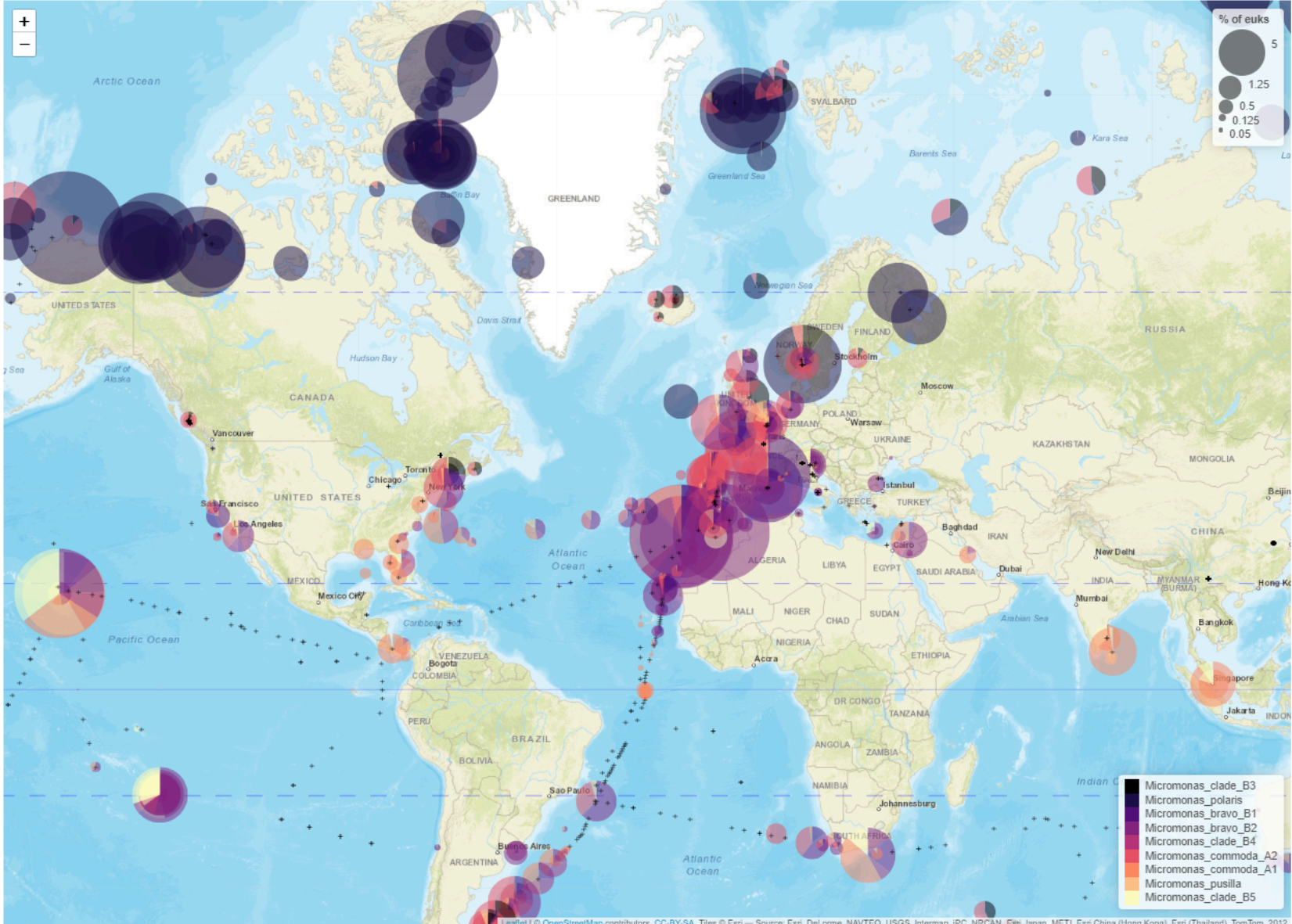
Current status

- Version 2.1
- Datasets: 59
- Samples: 6,202
- Barcodes (ASVs): 93,127



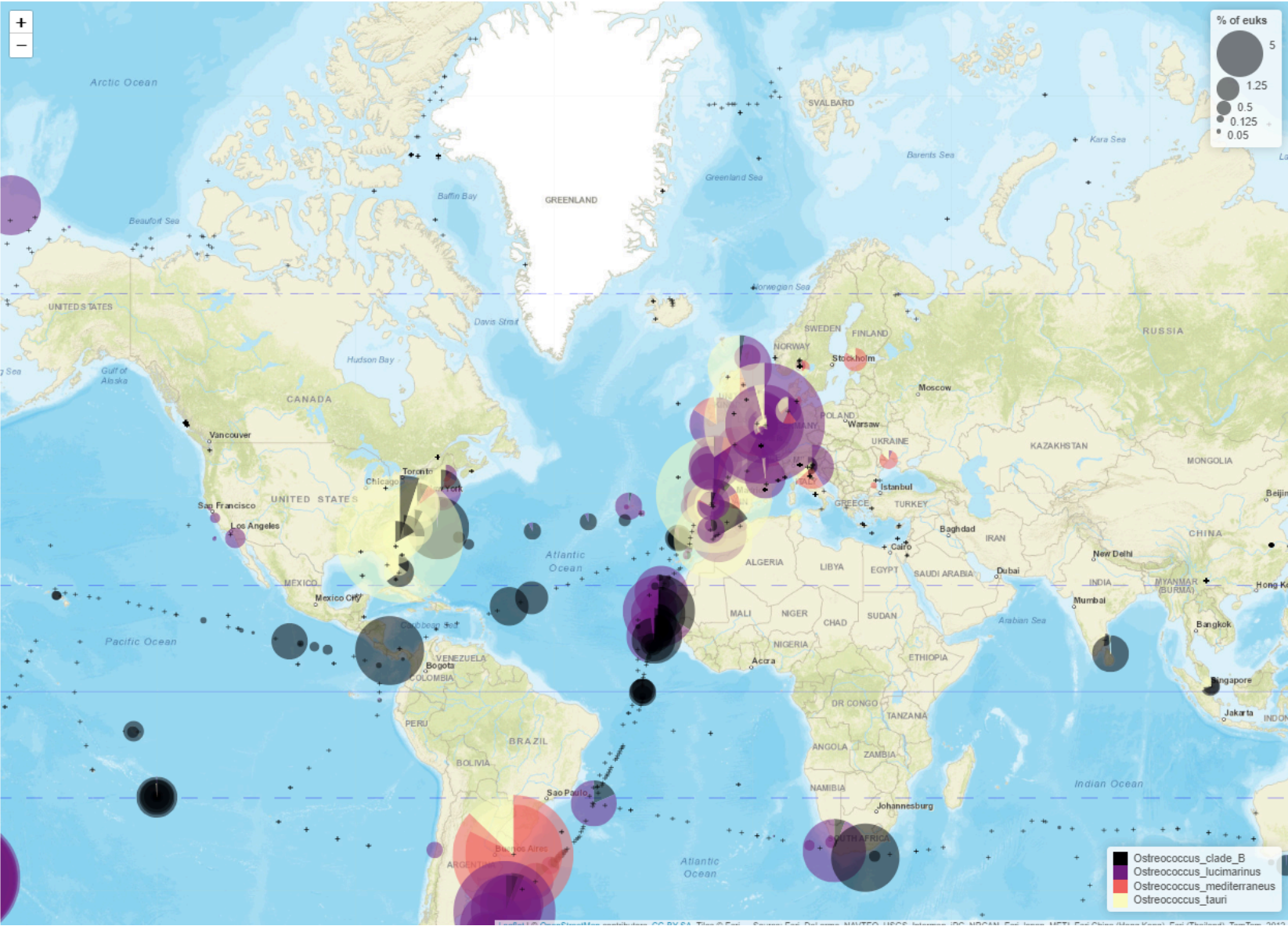
Example: Biogeography of *Micromonas*

Taxo level: genus - Taxon name: *Micromonas*
Number of samples with taxon: 1100, without taxon: 545



Example: Biogeography of *Ostreococcus*

Taxo level: genus - Taxon name: *Ostreococcus*
Number of samples with taxon: 695 , without taxon: 950





The metaPR2 database

Help

Select datasets

41 items selected

Select Samples

Gene regions

V4

DNA or RNA

DNA

Ecosystems

oceanic, coastal, freshwater lakes, freshwater rivers, terrestrial

Substrates

water

Size fractions

pico, total

Depth levels

surface

Select ASVs

Minimum number of total reads per ASV

100

Select Taxa

Validate Taxa Reset Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET to reset taxonomy to top level (need to press validate after reset)

Supergroup

Archaeplastida

Division

Chlorophyta

Class

All

Order

All

Family

All

Genus

All

Species

About Datasets Treemap Map Barplot Diversity Query Download

Reads are expressed as % relative to total eukaryotes. The taxa shown are below the level selected. For example if you select division, the different classes within this division will be shown.

Crosses indicate samples for which the selected taxonomic group is absent.

Taxo level: division - Taxon name: Chlorophyta

Number of samples with taxon: 935 , without taxon: 25

Map type

pie chart dominant taxon

Change circle scale - % max



MetaPR2 - Main functions

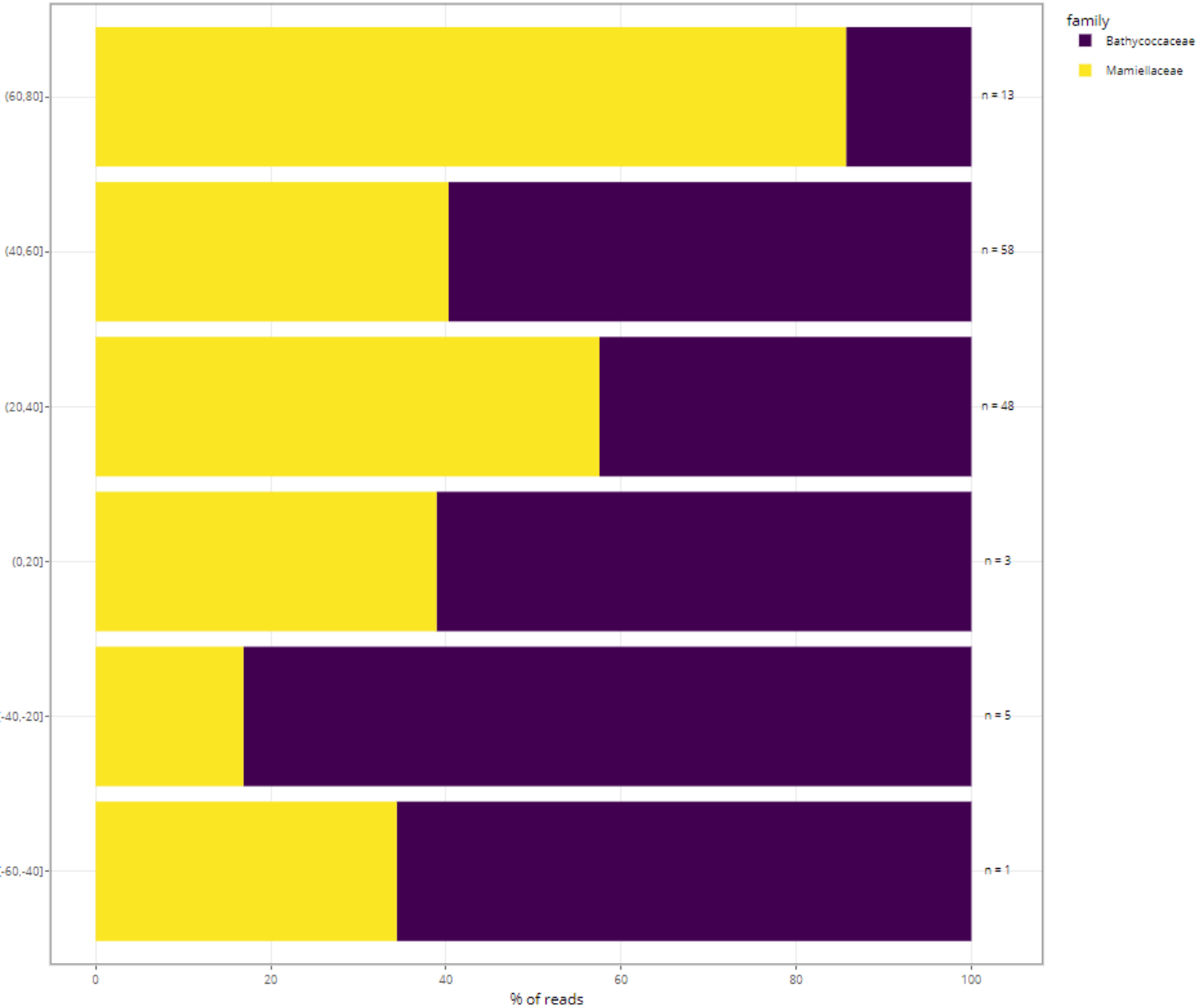
MetaPR2 - Taxonomy

Nine levels:

- Domain: Eukaryota
- Supergroup: Archaeplastida
- Division: Chlorophyta
- Subdivision: Chlorophyta_X
- Class: Mamiellophyceae
- Order: Mamielliales
- Family: Bathycoocaceae
- Genus: *Bathycooccus*
- Species: *B. prasinus*

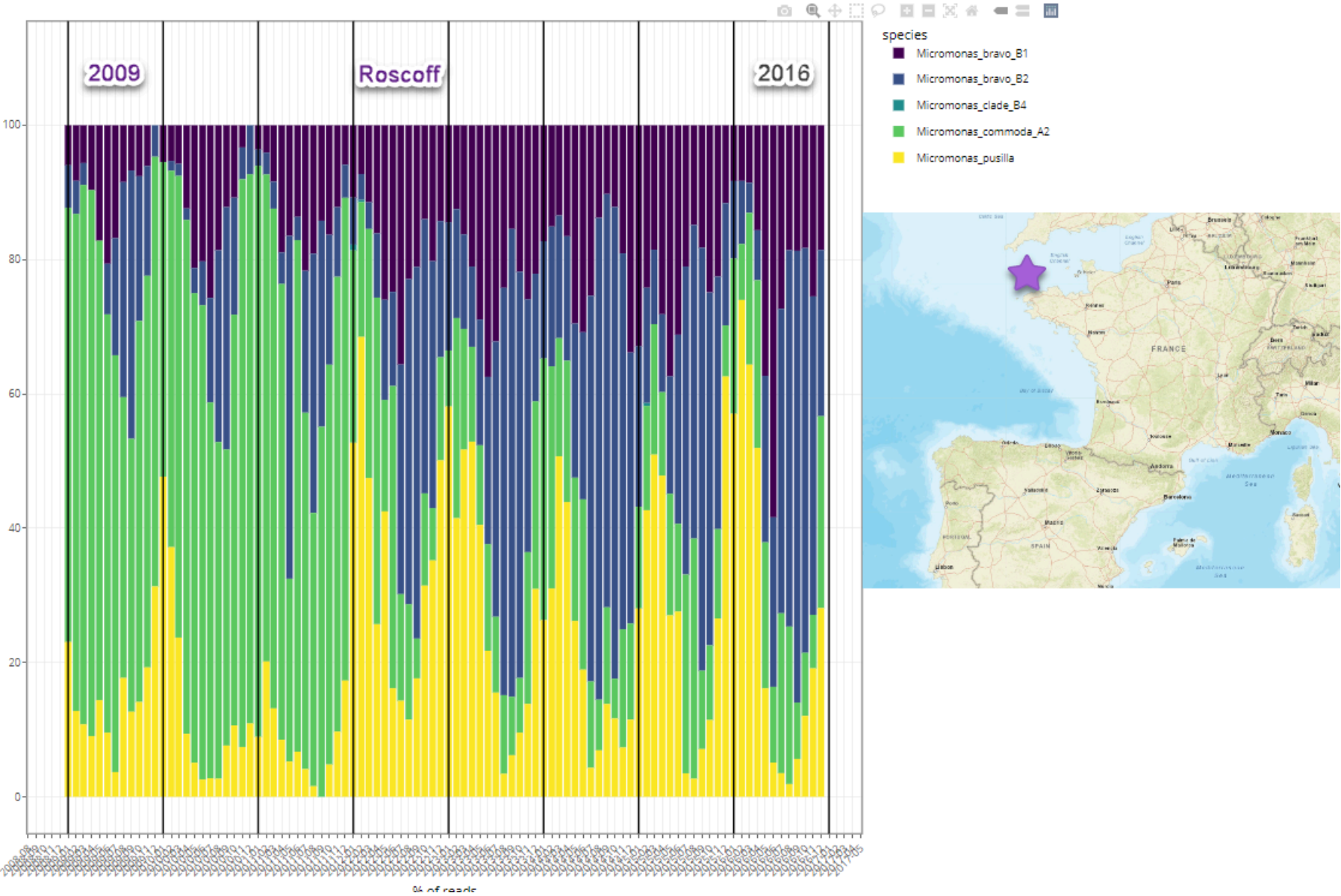


Barplots - Latitude

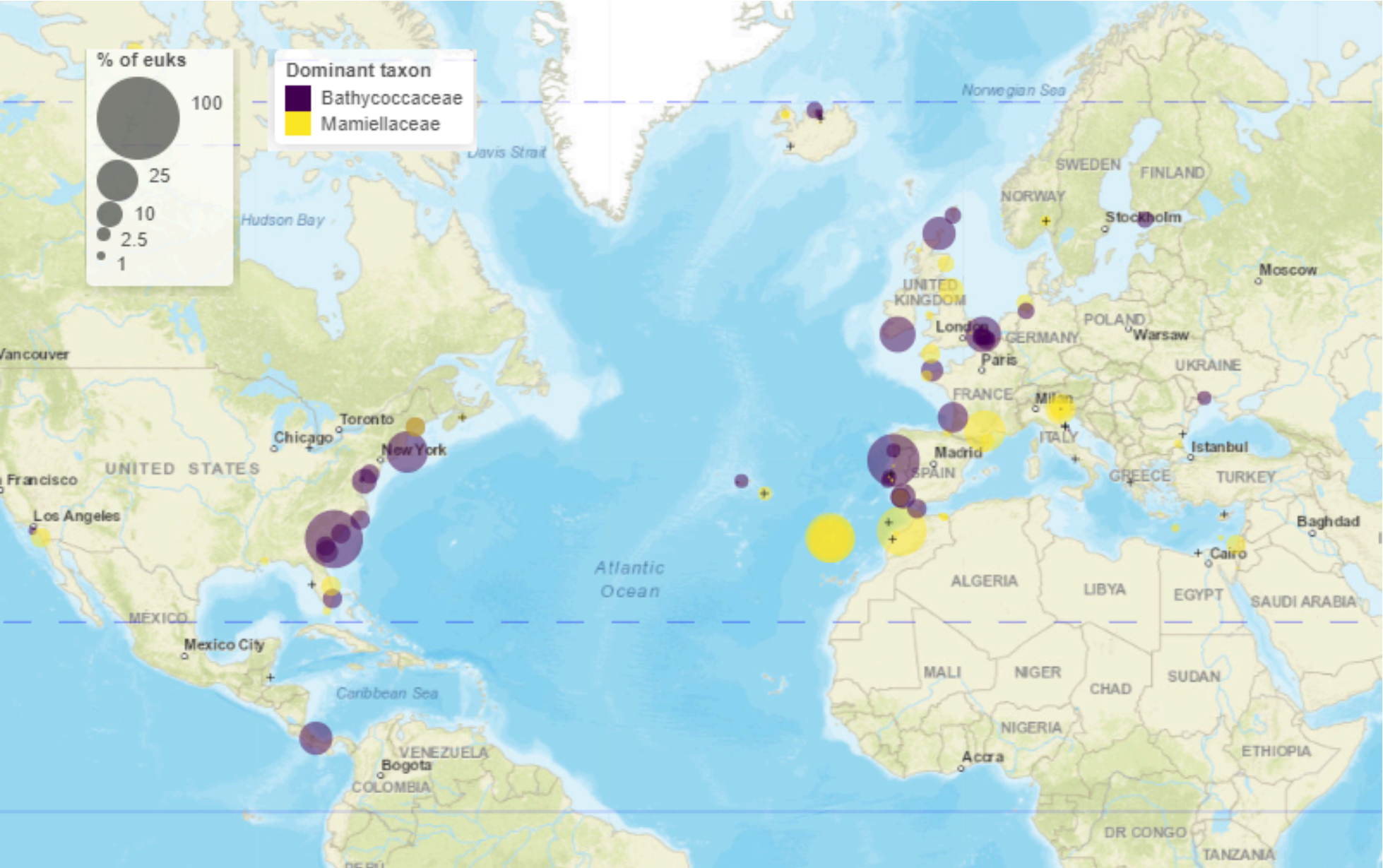


Barplots - Time series

Taxo level: genus - Taxon name: Micromonas



Maps - Dominant



Maps - Pie charts



Diversity

Number of samples: 960

Compute diversity - Press again after updating samples

Alpha diversity **Beta diversity**

Diversity Measure
 Chao1 Shannon Simpson

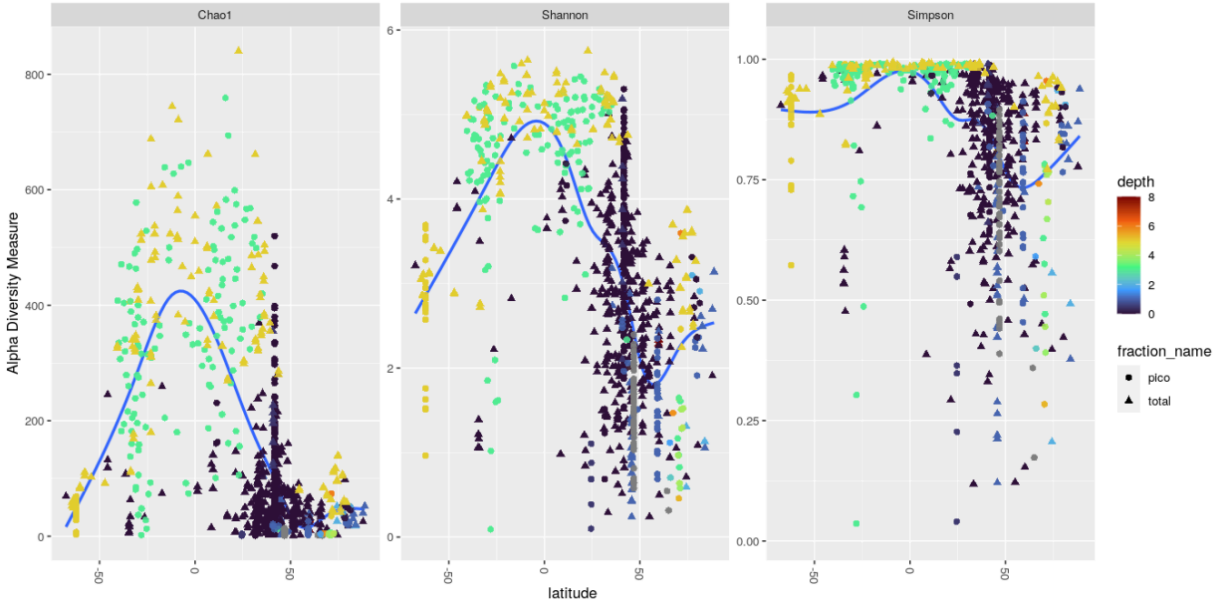
X axis
 latitude ecosystem substrate depth_level depth fraction_name DNA_RNA temperature salinity

Color
 latitude depth temperature salinity

Shape
 fraction_name substrate ecosystem depth_level DNA_RNA

```

phyloseq-class experiment-level object
otu_table() OTU Table: [ 3298 taxa and 908 samples ]
sample_data() Sample Data: [ 908 samples by 9 sample variables ]
tax_table() Taxonomy Table: [ 3298 taxa by 8 taxonomic ranks ]
    
```



Alpha diversity **Beta diversity**

Ordination method
 NMDS CCA RDA MDS PCoA

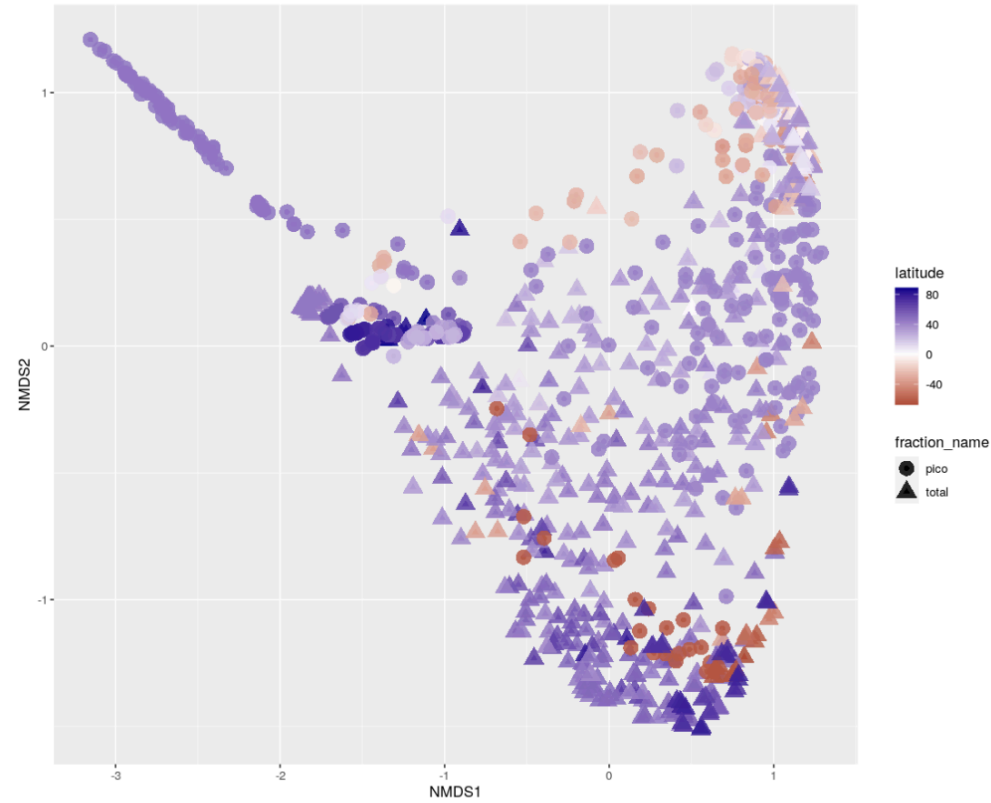
Ordination distance
 Bray-Curtis Gower Jensen-Shannon Divergence Jaccard

Color varies with:
 latitude depth temperature salinity

Shape varies with:
 fraction_name substrate ecosystem depth_level DNA_RNA

```

phyloseq-class experiment-level object
otu_table() OTU Table: [ 3298 taxa and 908 samples ]
sample_data() Sample Data: [ 908 samples by 9 sample variables ]
tax_table() Taxonomy Table: [ 3298 taxa by 8 taxonomic ranks ]
    
```





The metaPR2 database

Help

Select datasets

41 items selected

Select Samples

Gene regions

V4

DNA or RNA

DNA

Ecosystems

oceanic, coastal, freshwater lakes, freshwater rivers, terrestrial

Substrates

water

Size fractions

pico, total

Depth levels

surface

Select ASVs

Minimum number of total reads per ASV

100

Select Taxa

Validate Taxa Reset Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET to reset taxonomy to top level (need to press validate after reset)

Supergroup

Archaeplastida

Division

Chlorophyta

Class

All

Order

All

Family

All

Genus

All

Species

About Datasets Treemap Map Barplot Diversity Query Download

Reads are expressed as % relative to total eukaryotes. The taxa shown are below the level selected. For example if you select division, the different classes within this division will be shown.

Crosses indicate samples for which the selected taxonomic group is absent.

Taxo level: division - Taxon name: Chlorophyta

Number of samples with taxon: 935 , without taxon: 25

Map type

pie chart dominant taxon

Change circle scale - % max



MetaPR2 - In practice

Installation

MetaPR2 package

Install as an R package (You will be asked to install other packages that are required).

There are two options:

1. From Github

```
install.packages("devtools")

devtools::install_github("manutamminen/blaster")
devtools::install_github("pr2database/metapr2-shiny")
```

2. From R-universe

```
options(repos = c(
  pr2database = 'https://pr2database.r-universe.dev',
  CRAN = 'https://cloud.r-project.org'))

install.packages('metapr2')
```

Once the package is installed, you can run it under R or R studio

```
metapr2::run_app()
```

Help and Samples

Help

- Read in detail

Sample table

- dataset_name
- paper (can be useful to read)
- number of samples
- number of ASVs
- number of reads per sample (coverage)

Sample selection

- Major datasets: OSD, Tara, Malaspina
- By habitat: oceanic, coastal etc...
 - Start by “marine global V4”
 - Extend to other habitats/datasets
- V4 vs V9
- DNA vs. RNA
- Ecosystems
- Substrate: water, ice, soil...
- Size fractions: total, pico...
- Depth level: surface, euphotic...
- Minimum ASV: will filter out rare ASVs (e.g. 1000)
- Selection can be saved (yaml file)

Taxonomy

- Can select several taxa within one level
- Press validate every time you need to refresh
- Can exclude taxa to remove fungi, metazoa...
- Can save taxonomy and reload taxonomy (yaml file)

Select Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET + VALIDATE to reset taxonomy to top level

Supergroup

Division

Class

Order

Family

Genus

Micromonas
Mantoniella
Mamiella
RCC391

Exclude Taxa

Save/Load Taxa
 No file selected

Treemaps, Maps and Barplots

Treemaps

- Left panel: abundance (number of reads)
 - Reads are “normalized” to 100
- Right panel: diversity (number of ASVs)

Maps

- Read information at top
 - Taxo level
 - Number of samples with/without taxa
- Crosses where taxa absent
- Map types
 - Dominant
 - Pie chart
- Circle scale
 - Moving right increases size

Barplots

- taxonomy vs. function
- variables to use (but this depends on samples selected !)
 - fraction name
 - ecosystem
 - substrate
 - depth level
 - DNA_RNA
 - latitude
 - temperature
 - salinity
 - year, month, day for time series

Diversity

- Hit “Compute...” after refreshing taxonomy
- Time proportional to N samples and taxa
- Information about
 - Number of samples
 - Number of taxa (ASVs)

Alpha diversity

- X: Chao1, Shannon, Simpson (compare)
- Discretize continuous Y
- Change Y (see barplots)
- Change shape
- Change color

Beta diversity

- Ordination method (difference ?)
- Ordination distance (Bray, Jaccard...)
- Change color and shape

Download

- Download
 - datasets (csv)
 - samples (csv)
 - asv list with taxonomy (csv)
 - asv sequences (FASTA)

You can process these data with R (e.g. dplyr and ggplot2)

MetaPR2 for biogeography

Suggestions

- Diatoms
 - *Coscinodiscus*
 - *Rhizosolenia*
 - *Leptocylindrus*
 - *Pseudo-nitzschia*
- Dinoflagellates
 - *Dinophysis*
 - *Tripos*
 - *Cryptoperidinium*
- Ciliates
- Choanoflagellates

Analysis

Key points

- Look for key papers on this group
- What are the dominant species?
- What is the microdiversity [diversity within dominant species (ASVs)]?
- What is distribution ?
 - Substrate (water, ice...)
 - Ecosystems (marine, freshwater, terrestrial)
 - Size fraction
 - Depth layers (euphotic zone vs. meso and bathypelagic)
 - Latitudinal bands (polar, temperate, tropical)
 - Coastal vs Pelagic
- Alpha diversity
- Beta diversity