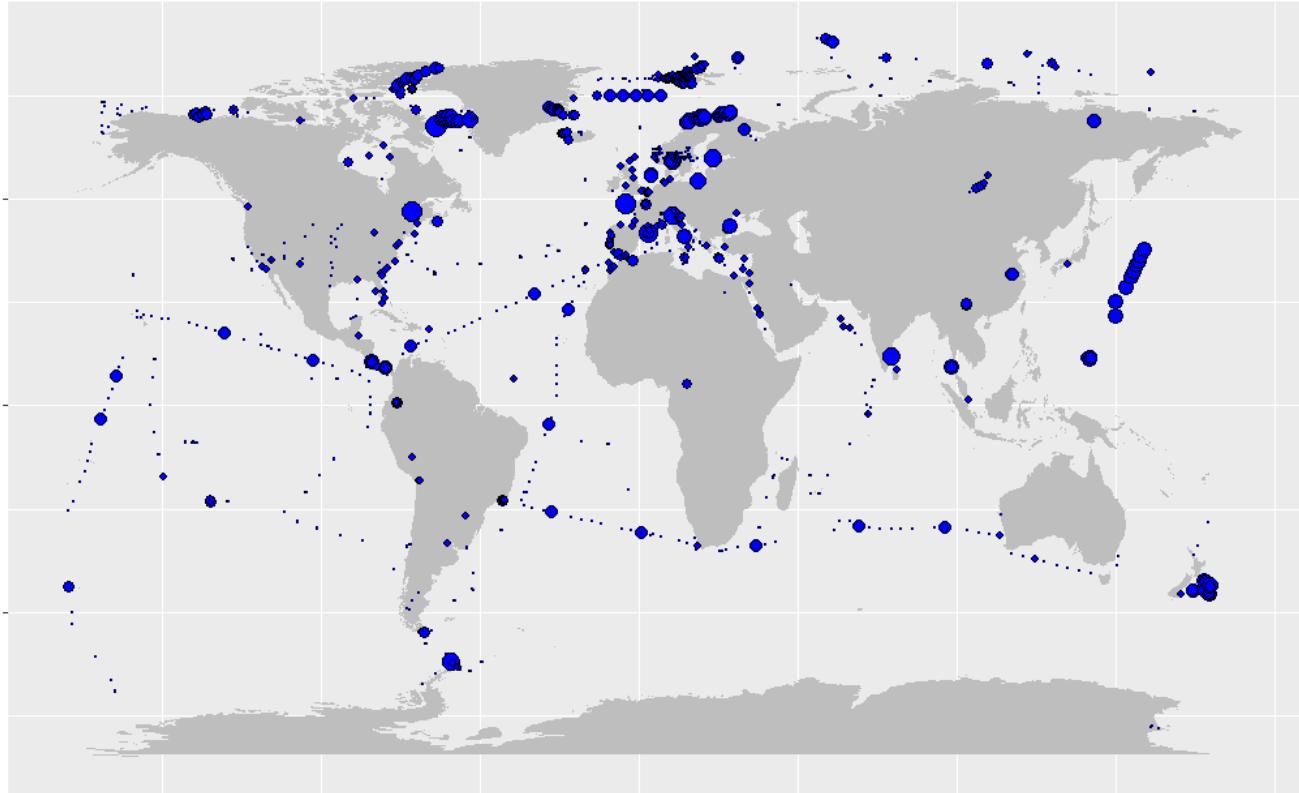


# Microbes

Week 10 - Analyzing protist communities

## A database of metabarcodes



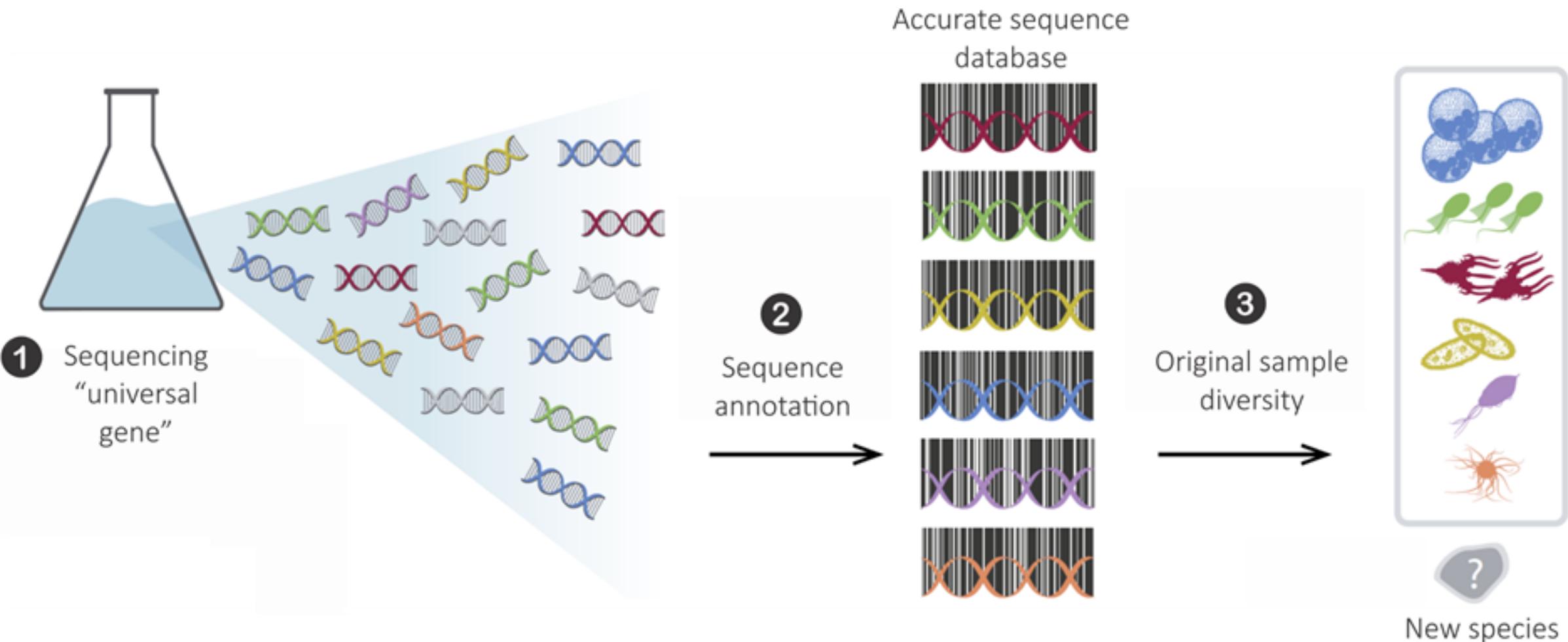
Daniel Vaulot

2022-10-22

# Outline

- Metabarcoding data
- Factors affecting protist communities
- Diversity
- Visualization/Analysis
- MetaPR2 in practice

# Metabarcoding



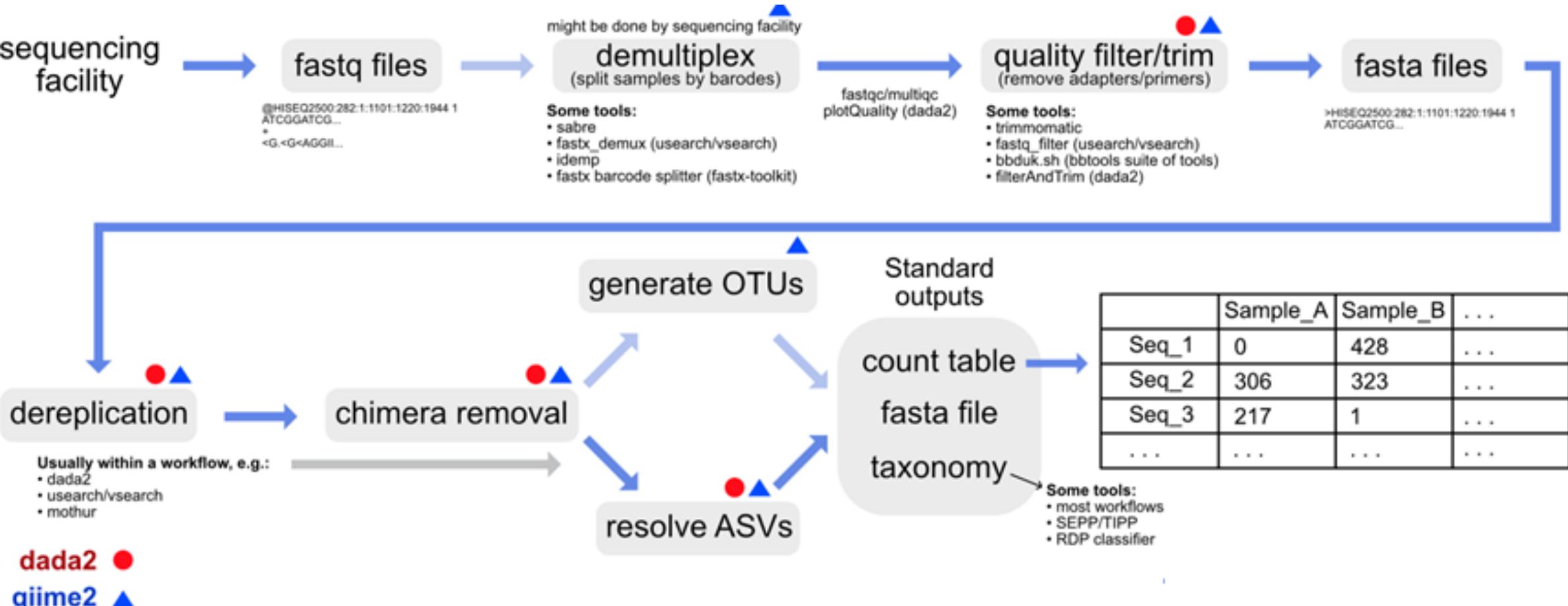
# Metabarcoding

Sequences Cluster Assign

Fastq files

Identifier	• @SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
Sequence	• TTGCCTGCCTATCATTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT
'+' sign	• +
Quality scores	• hhhhhhhhhghhhhhfhhhhfffffe'ee[‘X]b[d[ed‘[Y[^Y
Identifier	• @SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
Sequence	• GATTGTATGAAAGTATAACAACTAAAACGTGCAGGTGGATCAGAGTAAGTC
'+' sign	• +
Quality scores	• hhggfhhcgghggfcffdhfehhhcuhdchhdhahehffffde‘bVd

# Metabarcoding



# Metabarcoding

## Data tables

ASVs   Abundance   Metadata   Merged

	A	B	C	D	E	F	G	H	I
1	otu	Domain	Supergroup	Division	Class	Order	Family	Genus	
2	Otu001	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Ostreococcus	
3	Otu002	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	UCYN_A1_host	
4	Otu003	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	
5	Otu004	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Prorocentrum	
6	Otu005	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Mediophyceae	Mediophyceae_X	Thalassiosira	
7	Otu006	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Pseudo_nitzschia	
8	Otu007	Eukaryota	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagophyceae_X	Pelagophyceae_X	Pelagomonas	
9	Otu008	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X	
10	Otu009	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chryschromulinaceae	Chryschromulina	
11	Otu010	Eukaryota	Opisthokonta	Metazoa	Craniata	Craniata_X	Craniata_XX	Craniata_XX_unclassified	
12	Otu011	Eukaryota	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade_C	Chrysophyceae_Clade_C_X	
13	Otu012	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Gonyaulax	
14	Otu013	Eukaryota	Alveolata	Dinophyta	Syndiniales	Syndiniales_Group_III	Syndiniales_Group_III_X	Syndiniales_Group_III_X	
15	Otu014	Eukaryota	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade_G	Chrysophyceae_Clade_G_X	
16	Otu015	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X	
17	Otu016	Eukaryota	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_X	
18	Otu017	Eukaryota	Opisthokonta	Fungi	Basidiomycota	Agaricomycotina	Agaricomycetes	Hyphodontia	
19	Otu018	Eukaryota	Stramenopiles	Ochrophyta	Dictyochophyceae	Dictyochophyceae_X	Pedinellales	Pedinellales_X	
20	Otu019	Eukaryota	Opisthokonta	Fungi	Basidiomycota	Agaricomycotina	Agaricomycetes	Itersonilia	
21	Otu020	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera	
22	Otu021	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X	
23	Otu022	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Prymnesiophyceae_X	Syracospaera	
24	Otu023	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Bacillariophyceae_X	
25	Otu024	Eukaryota	Archaeplastida	Streptophyta	Klebsormidiophyceae	Klebsormidiophyceae_X	Klebsormidiophyceae_XX	Klebsormidium	
26	Otu025	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas	
27	Otu026	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Cylindrotheca	
28	Otu027	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Suessiales	Suessiales_X	Karlodinium	
29	Otu028	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Isochrysidales	Noelaerhabdaceae	Emiliania	
30	Otu029	Eukaryota	Opisthokonta	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomyces	
31	Otu030	Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptophyceae_X	Cryptomonadales	Teleaulax	
32	Otu031	Eukaryota	Alveolata	Dinophyta	Syndiniales	Syndiniales_Group_I	Syndiniales_Group_I_Clade_1	Syndiniales_Group_I_Clade_1_X	
33	Otu032	Eukaryota	Archaeplastida	Chlorophyta	Prasino_Clade_VII	Prasino_Clade_VII_X	Prasino_Clade_VII_A	Prasino_Clade_VII_A_4_X	

# Factors affecting protist communities

## Substrate

- Water
- Ice
- Sediment
- Soil
- Microbiome

## Size fraction

- Total (0.2 µm -> 100 µm)
- Pico (0.2 µm -> 2-3 µm)
- Nano (2-3 µm -> 20 µm)
- Micro (20 µm -> 100-200 µm)
- Meso (100 µm -> 1000 µm)

## Ecosystem

- Oceanic
- Coastal
- Rivers
- Lakes
- Terrestrial

# Factors affecting protist communities

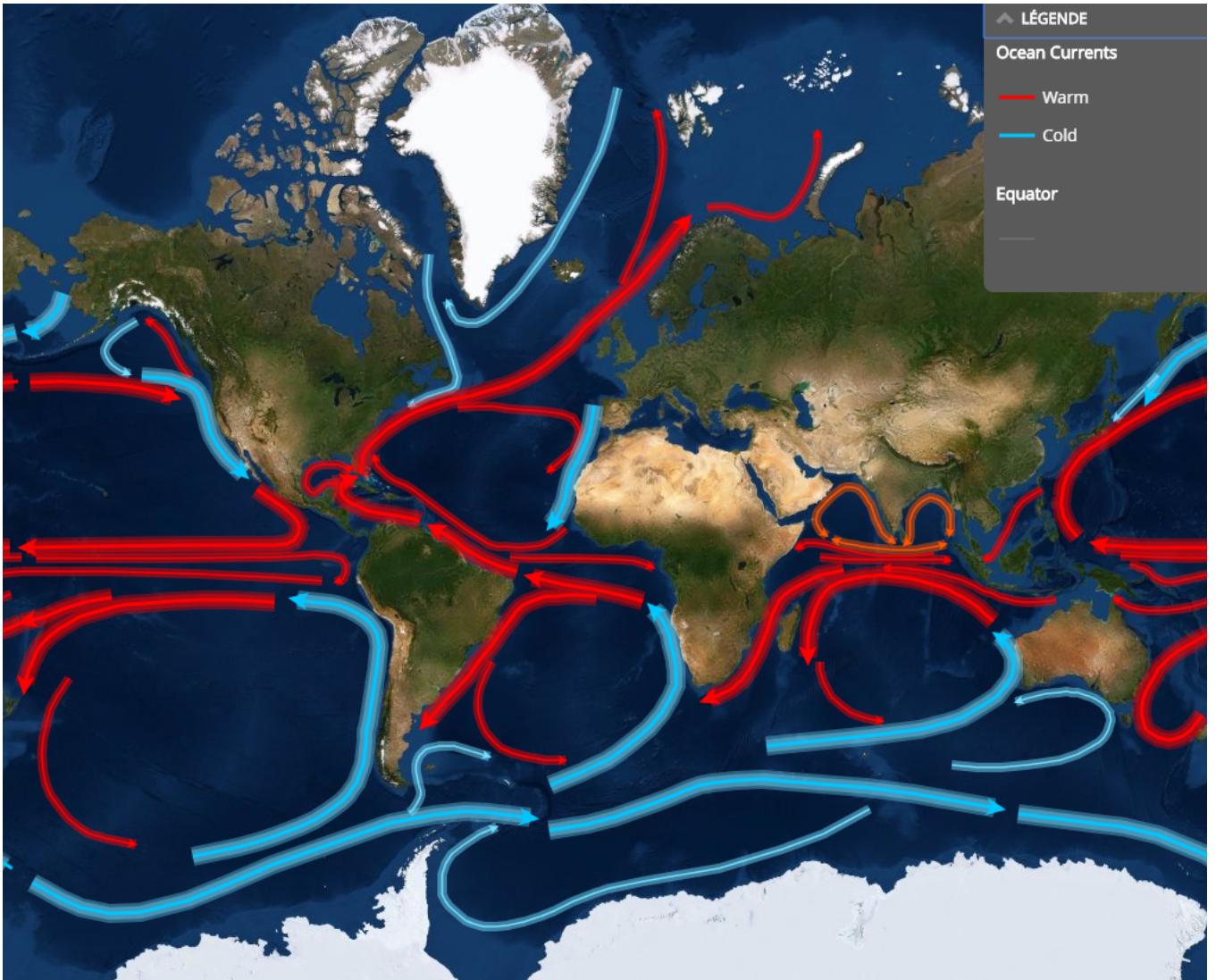
## **Environmental conditions**

In oceanic waters:

- temperature
- salinity
- light
- nutrients

... which depend on:

- substrate (water vs. ice)
- latitude
- time of the year
- depth
- oceanic currents
- proximity of coast

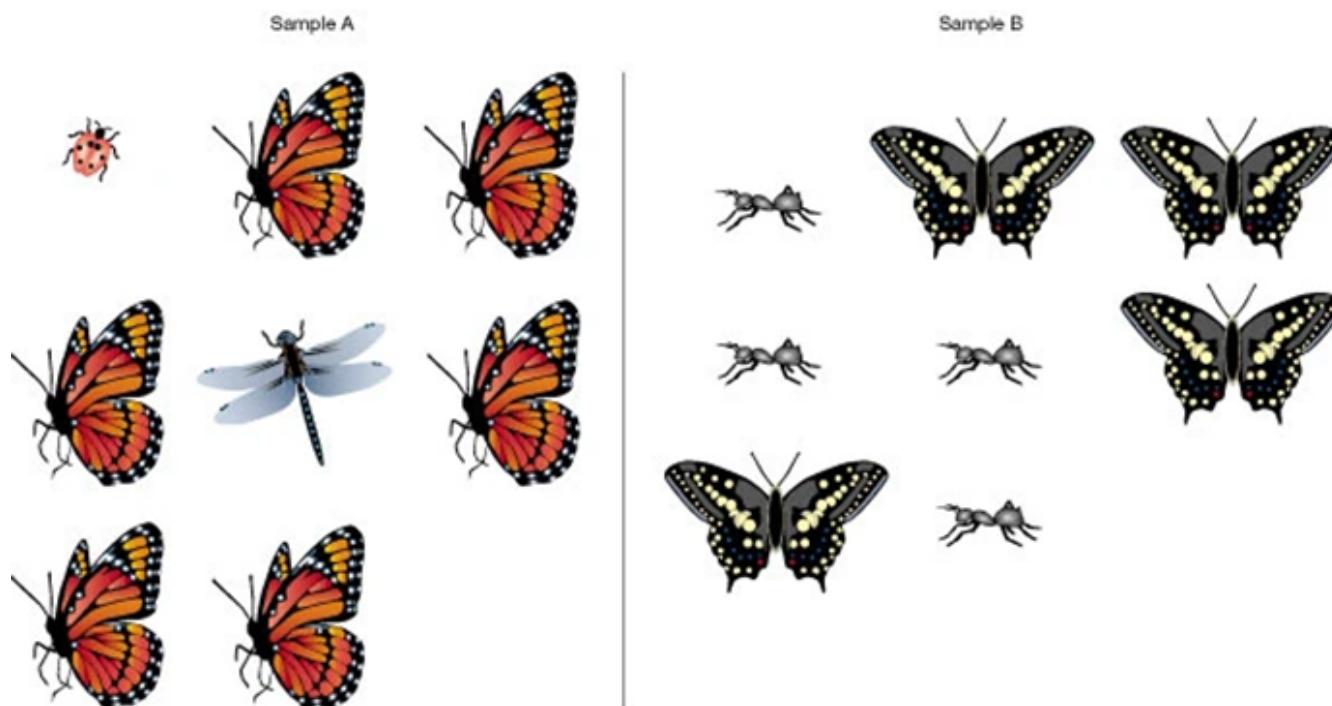


# Diversity

Microbial species in a sample

- **species richness:** total number of species
- **species abundance:** proportion of each species

Richness vs. Evenness



# Diversity

## Alpha diversity - Diversity within a given sample

- **Chao 1** is a non-parametric estimator of the number of species in a community.
- **Shannon index**<sup>1</sup>

$$H = - \sum_{i=1}^S p_i \cdot \log p_i$$

Where:

$p_i$  = fraction of the entire population made up of species  $i$  (proportion of a species  $i$  relative to total number of species present)

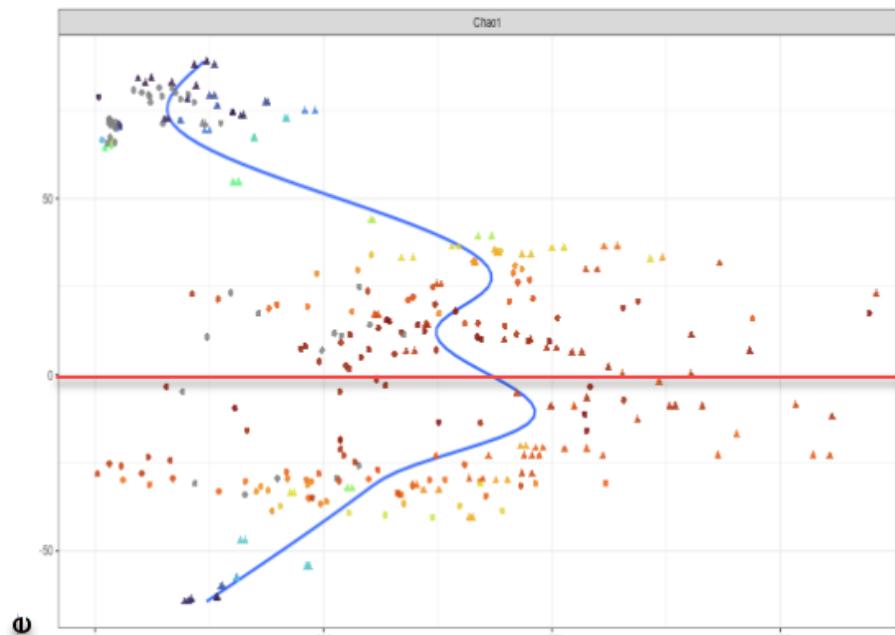
$S$  = numbers of species encountered

A high value of  $H$  would be a representative of a diverse and equally distributed community and lower values represent less diverse community. A value of 0 would represent a community with just one species.

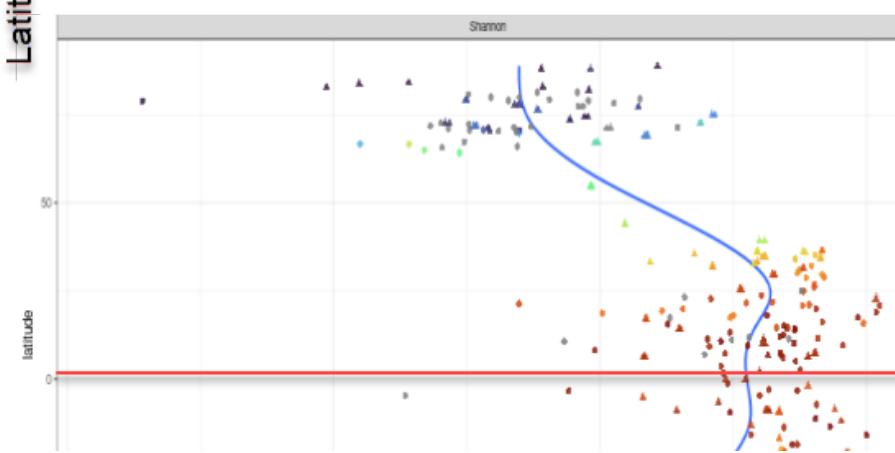
1. Shannon's diversity index quantifies the uncertainty in predicting the species identity of an individual that is taken at random from the dataset.

# Diversity

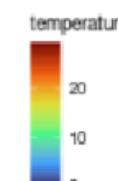
## Alpha diversity - Effect of latitude



Chao1



Shannon



temperature



fraction\_name

# Diversity

## Beta diversity - Compare diversity between samples

- Compute **distance** between samples:

- **Bray-Curtis** dissimilarity: use abundance information

- Varies between 0 and 1:
    - 0 means the two samples have the same composition
    - 1 means the two samples do not share any species

$$BC_{jk} = 1 - \frac{2 \sum_{i=1}^p \min(N_{ij}, N_{ik})}{\sum_{i=1}^p (N_{ij} + N_{ik})}$$

where  $N_{ij}$  is the abundance of species  $i$  in sample  $j$  and  $p$  the total number of species

- **Jaccard similarity** index

- Number of common species between samples divided by total number of species in the two samples

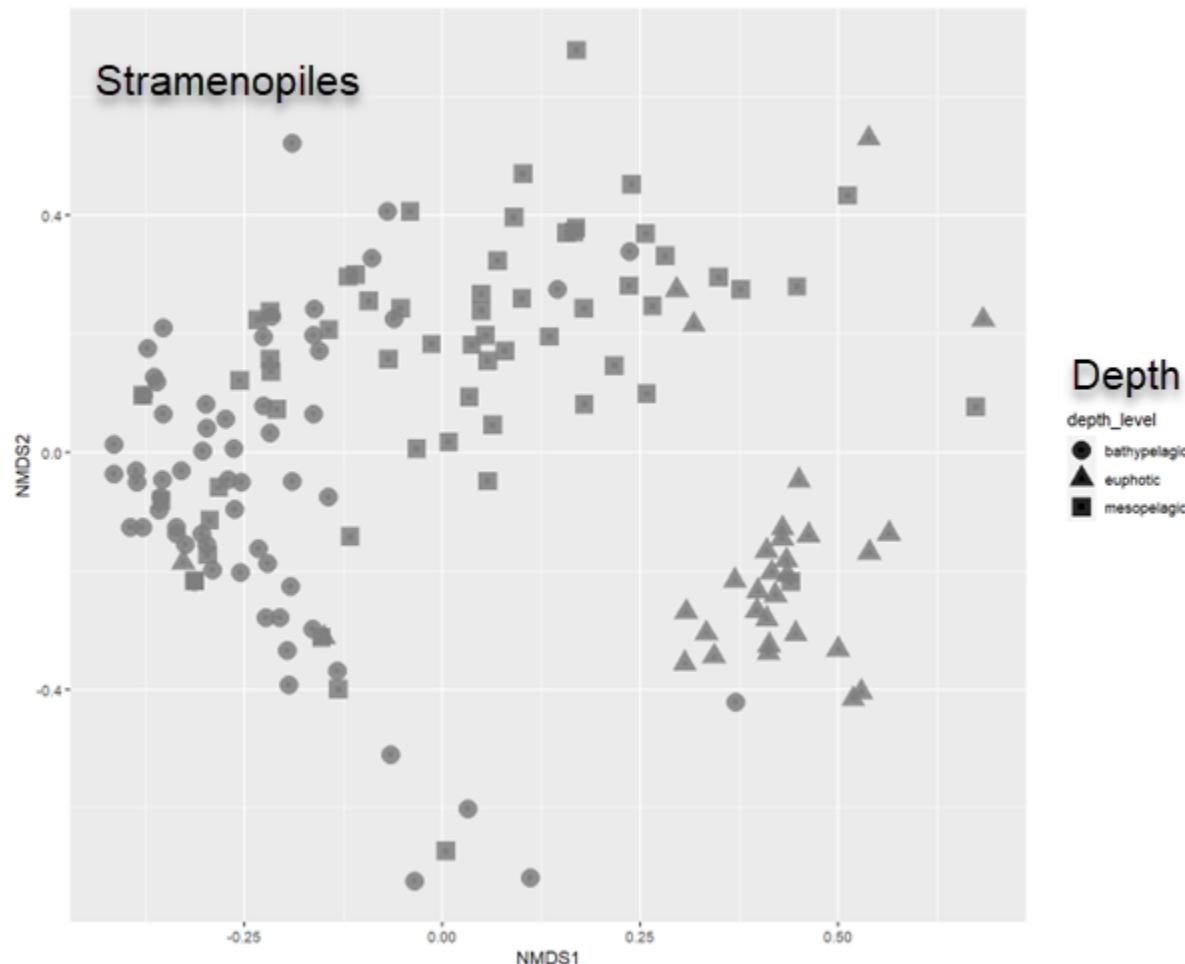
$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- Ordinate the samples

- **NMDS**: Non-Metric Multidimensional Scaling

# Diversity

## Beta diversity - Effect of depth on Stramenopiles communities



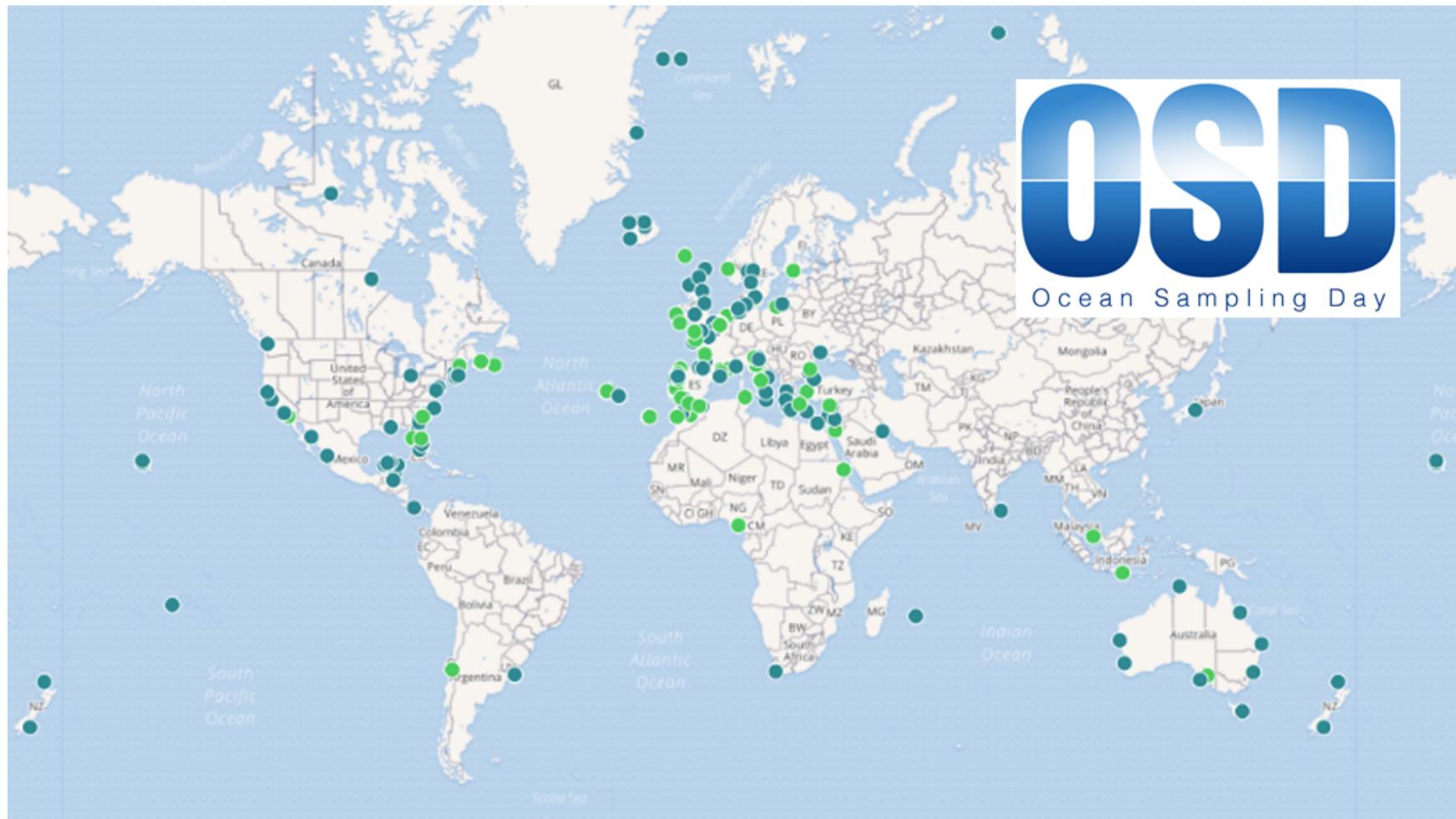
# MetaPR2 - Datasets

OSD

Tara Oceans

Malaspina

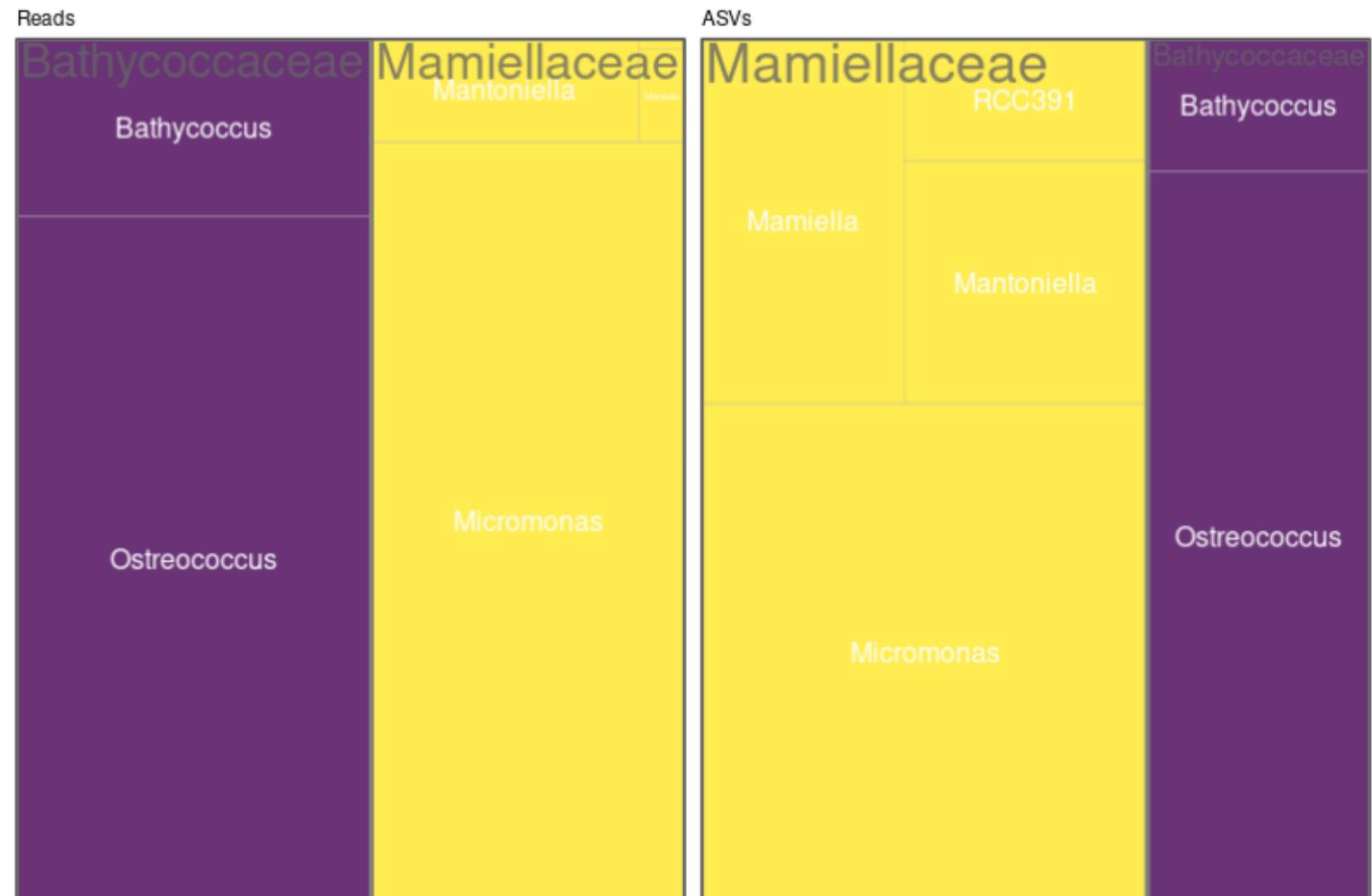
Select



# MetaPR2 - Taxonomy

## Eight levels:

- Kingdom: Eukaryota
- Supergroup: Archaeplastida
- Division: Chlorophyta
- Class: Mamiellophyceae
- Order: Mamiellales
- Family: Bathycoccaceae
- Genus: *Bathycoccus*
- Species: *B. prasinos*



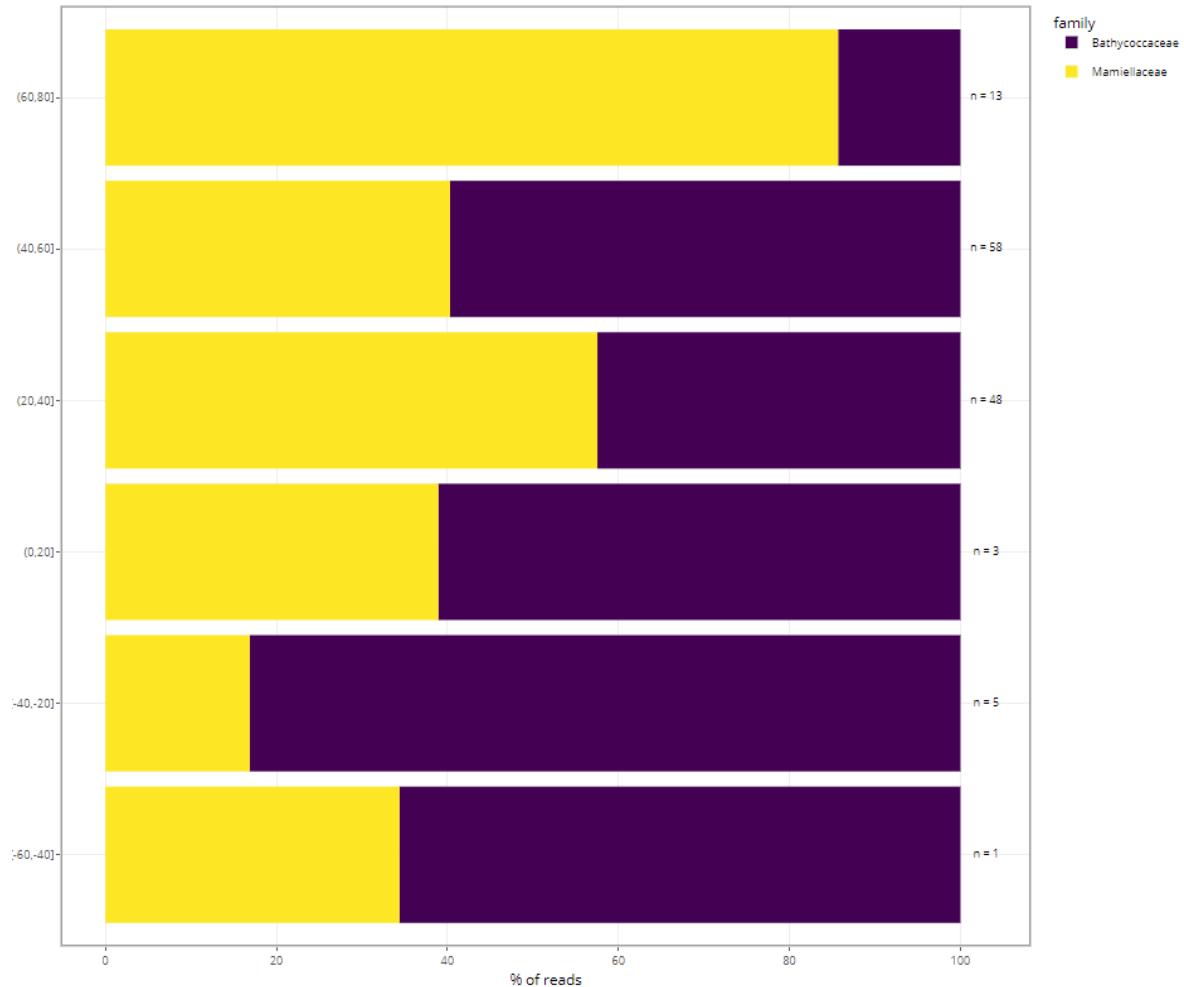
# MetaPR2 - Visualization

Barplots

Maps - Dominant

Maps - Pie charts

Diversity



## 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)

# MetaPR2 - In practice

## 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

Validate Taxa    Reset Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET + VALIDATE to reset taxonomy to top level

Supergroup

Archaeplastida

Division

Chlorophyta

Class

Mamiellophyceae

Order

Mamiellales

Family

Mamiellaceae

Genus

All

All

Micromonas

Mantoniella

Mamiella

RCC391

Exclude Taxa

None

Save/Load Taxa

Save taxonomy

Load taxonomy    No file selected

# MetaPR2 - In practice

## 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

# MetaPR2 - In practice

## 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
  - samples
  - asv list with taxonomy
  - asv sequences

Only for those with extensive experience with data processing.