# Functional diversity of microbial eukaryotes in a meromictic lake: Coupling between metatranscriptomic and a trait-based approach

Arthur Monjot, Gisèle Bronner, Damien Courtine, Corinne Cruaud, Corinne da Silva, Jean-marc Aury, Frederick Gavory, Anne Moné, Agnès Vellet, Ivan Wawrzyniak, et al.

# Functional diversity of microbial eukaryotes in a meromictic lake: coupling between metatranscriptomic and a trait-based approach

**Arthur Monjot[1°], Gisèle Bronner[1], Damien Courtine[1], Corinne Cruaud[2], Corinne Da Silva[3], Jean-Marc Aury[3], Frederick Gavory[3], Anne Moné[1], Agnès Vellet[1], Ivan Wawrzyniak[1], Jonathan Colombet[1], Hermine Billard[1], Didier Debroas[1], Cécile Lepère[1]**

1- CNRS, Laboratoire Microorganismes : Génome et Environnement, Université Clermont Auvergne, Clermont-Ferrand, France.

2- Genoscope, Institut de Biologie François Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, Evry, France.

3- Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France.

°corresponding author: Arthur.Monjot.pro@gmail.com

## Abstract

The advent of high-throughput sequencing has led to the discovery of a considerable diversity of microbial eukaryotes in aquatic ecosystems, nevertheless their function and contribution to the trophic food web functioning remain poorly characterized especially in freshwater ecosystems. Based on metabarcoding data obtained from a meromictic lake ecosystem (Pavin, France), we performed a morpho-physio-phenological traits-based approach to infer functional groups of microbial eukaryotes. Metatranscriptomic data were also analyzed to assess the metabolic potential of these groups across diel cycle, size fraction, sampling depth and periods. Our analysis highlights a huge microbial eukaryotic diversity in the monimolimnion characterized by numerous saprotrophs expressing transcripts related to sulfur and nitrate metabolism as well as dissolved and particulate organic matter degradation. We also describe strong seasonal variations of microbial eukaryotes in the mixolimnion, especially for parasites and mixoplankton.

It appears that the water mixing (occurring during spring and autumn) which benefits photosynthetic host communities also promotes parasitic fungi dissemination and over-expression of genes involved in the zoospore phototaxis and stage transition in parasitic cycle. Mixoplanktonic haptophytes over-expressing photosynthesis-, endocytosis- and phagosome-linked genes under nutrient limitation also suggest that phagotrophy may provide them an advantage over non-phagotrophic phytoplankton.

*Keywords: Trait-based approach, Metatranscriptomic, Microbial eukaryotes, Functional diversity.*

## Introduction

In the last decade, the advent of high-throughput sequencing technologies has provided insight into the breadth of the microbial eukaryotic diversity. In aquatic environments, metabarcoding surveys revealing an unsuspected diversity distributed among the entire tree of eukaryotic life. However, these data are limited to richness and diversity, and functions are deduced from taxonomies. This significant bias diminishes our understanding of the whole eukaryotes' diversity, and leads to skewed views of what microbial eukaryotes even are, as well as their role in the environment. For example, molecular analyses of environmental DNA samples have recently revealed an unexpectedly large diversity of undescribed fungi ("dark matter fungi" (DMF)) (Grossart *et al.*, 2015). Members of the DMF are ubiquitous and abundant in lakes, but have a small number of representatives in culture. These fungi could be saprotrophic and/or parasitic but very little is known about their ecological functions, such as their role in food web dynamics and biogeochemical cycling of organic matter, nutrients and energy. Metabolic plasticity (*e.g.*, of mixoplankton: photo-osmo-phago-mixotroph) is also difficult to comprehend with short 18S ribosomal RNA sequences. Even though phagotrophy by photosynthetic microbial eukaryotes is certainly underestimated due to experimentation difficulties, evidence is accumulating regarding the versatility of microbial eukaryotes in their nutritional mode (Flynn *et al.*, 2019; Hartmann *et al.*, 2013; Leles *et al.*, 2018). The environmental conditions that favor one nutritional mode over another (*e.g.*, phototrophy *vs* phagotrophy, parasitism *vs* saprotrophy) and the metabolic priorities that shift among important microbial eukaryotic taxa are not yet well understood.

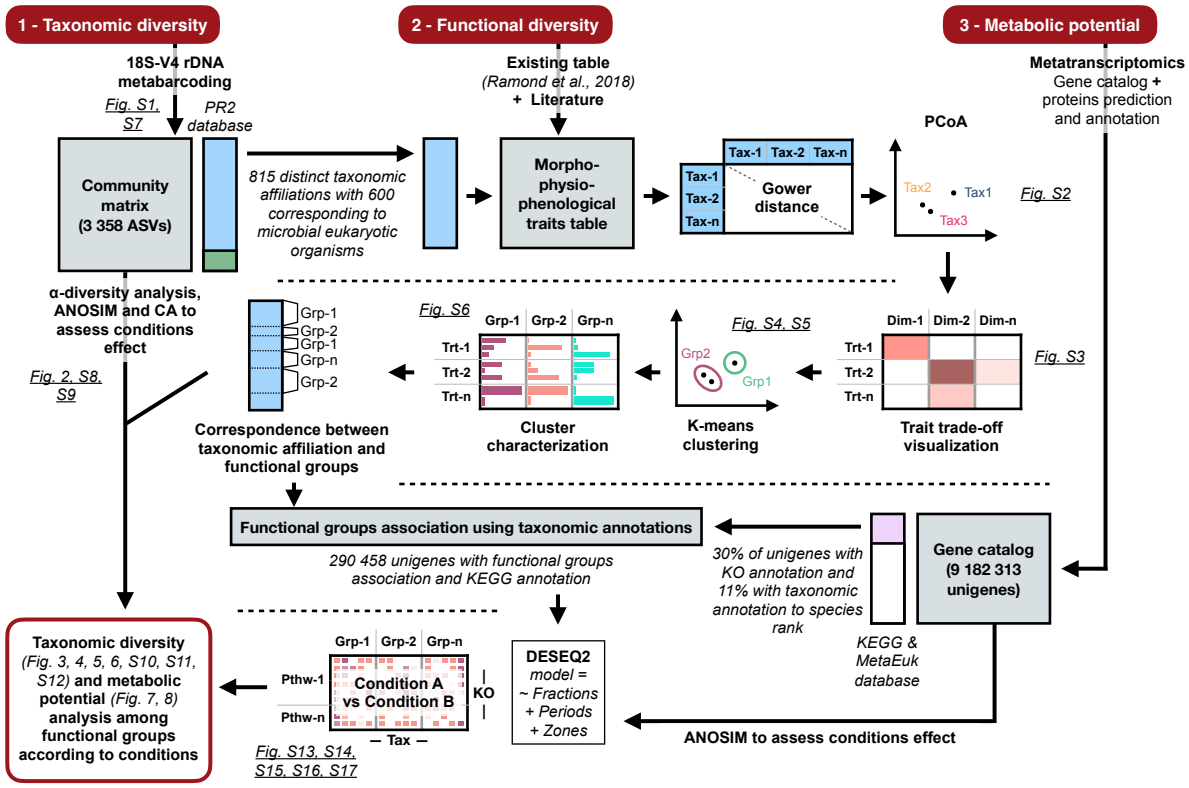



**Figure 1: Workflow of the trait-based approach coupled to metatranscriptomic analysis.** The following abbreviations are used: ASV: Amplicon Sequence Variant; CA: Correspondence analysis; PCoA: Principal Coordinate Analysis; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG Orthology; Tax: Taxonomic affiliation; Dim: Dimension; Trt: Morpho-physio-phenological trait; Grp: Functional group; Pthw: Metabolic pathway and/or cellular process.

The implementation of a trophic mode is multifactorial and depends on both innate factors (*e.g.*, evolution, stage of the organism's life cycle (juvenile, cyst), size, etc.) and environmental factors (*e.g.*, light, temperature, availability of nutrients and/or prey, oxygen gradient, etc.) (Hu *et al.*, 2018; Ramond *et al.*, 2019; Wilken *et al.*, 2020).

The challenge is therefore to link taxonomic diversity, functional role and forcing environmental parameters. While working with uncultured microorganisms, an alternative can be to couple metabarcoding data with a trait-based approach (Ramond *et al.*, 2019). Trait-based approaches involve the use of morpho-physio-phenological traits to assess the individual performance of species within ecosystems (*e.g.*, growth, reproduction, resource acquisition and survival (Litchman & Klausmeier, 2008; Ramond *et al.*, 2019; Violle *et al.*, 2007). On the other hand, metatranscriptomic provide an assessment of species functional activity with access of differential expression of genes under specific environmental conditions. Metatranscriptomic is however not widely applied in community ecology studies and has been used essentially to analyze marine microbial eukaryotes especially in the first meters of the water column (Dupont *et al.*, 2015; Hu *et al.*, 2018, Caputi *et al.*, 2019; Carradec *et al.*, 2018). In comparison to marine environments, freshwater ecosystems have globally received less attention in term of functional diversity although they also contain a great microbial diversity (Biderre-Petit *et al.*, 2022; Debroas *et al.*, 2017).

In our study functional diversity of understudied freshwater microbial eukaryotes was investigated with an original approach which couples a trait-based methodology and a metatranscriptomic analysis. Expressed functions and ecological strategies of microbial eukaryotes were studied for the first time in relation to oxygen presence/absence, diel cycle, size fractions and seasons in a meromictic lake (Pavin, France).

## Materials and Methods

Figure 1 presents the workflow used in this study.

## Study site and sampling

This study was conducted in the meromictic lake Pavin (Massif Central, France, 45°29'45" N, 2°53'17" E). This lake represents an excellent opportunity to study microbial communities with multiple biotic and abiotic variables found in one ecosystem. The upper oxygenated layer (mixolimnion) extends from the surface to 60 m whereas the lower layer (monimolimnion), permanently anoxic, extends from 60 to 92 m depth. Water sampling was carried out in both zones (*i.e.*, at 9 m and 80 m) using a Van Dorn bottle, at the center of the lake, by day and night and at four contrasted periods in 2018: April, June, September and November. Water samples (60 L each) were pre-filtered sequentially on filters of decreasing porosity (150 µm, 50 µm, 10 µm). Microorganisms that are retained on the 10 µm filter corresponding to the large size fraction (10-50 µm) were resuspended in a final volume of 300 ml of sterile water. The others corresponding to the small size fraction (<10 µm) were concentrated by tangential filtration on a Spectrum KrosFlo PES 0.65 µm column (Spectrum Labs, Rancho Dominguez, CA) and recovered in a final volume of 1 L. Samples were treated with 0.01% pluronic acid and centrifuged at 10 000 g, 4°C for 30 min. The pellets were then divided for DNA and RNA extraction with the aim of sequencing all samples in duplicates. The pellets were resuspended in 250 µL of RLT buffer for DNA extraction and with the addition of β-mercaptoethanol for RNA extraction. Samples were immediately frozen in liquid nitrogen.

## DNA and RNA extraction, amplification and sequencing procedure

After thawing, DNA samples were ground in a bead beater (3 pulses for 30 s at 30 Hz) with 0.1 g of 0.1 mm glass beads (Sigma-Aldrich, Saint-Louis, MO). Following 1 min of centrifugation at 800 g, supernatants, adjusted to 300 µL with extraction buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris), were incubated with 150 µl of sodium dodecyl sulfate 10% (SDS) for 5 min. Then, 5 µL of proteinase K (10 mg/ml) and 5 µL of RNAse A (10 U/µL) were added, and mixtures were incubated for 2 hours at 37°C. 80 µL of cetyltrimethylammonium bromide (CTAB 10%, NaCl 0.7M) solution and 100 µL of NaCl 5M were added before the nucleic acid extraction by phenol-chloroform-isoamyl alcohol and isopropanol precipitation. DNA pellets were washed with 70% ethanol and resuspended in 30 µL of water and stored in LoBinding DNA tubes at 4°C.

PCR amplification of the V4 region of the gene coding for the 18S rRNA was performed using the universal primer 515F (5'-GTGYCAGCMGCCGCGGTA-3') (Caporaso *et al.*, 2011) and the eukaryotic primer 951R (5'-TTGGYRAATGCTTTCGC-3') (Mangot *et al.*, 2013) according to Chauvet *et al.* (2022). Unfortunately, we had to pool some samples to get enough material especially for the samples originating from the monimolimnion. We therefore succeeded to sequence only 25 samples in duplicates (resulting in 57 amplicon datasets). RNA extraction was processed according to Chauvet *et al.* (2022). RNA was quantified using the Qubit RNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA) and quality was assessed with an Agilent RNA 6000 Pico Kit (Agilent, Santa Clara, CA) on the Agilent 2100 Bioanalyzer (Agilent). RINs (RNA integrity numbers) varied from 7.3 to 9.4. Illumina RNAseq libraries were prepared using the NEBNext Ultra II directional RNA library prep kit for Illumina (New England Biolabs, Ipswich, MA), according to the manufacturer's protocol starting with 100 ng total RNA. Ready-to-sequence Illumina libraries were then quantified by qPCR and library profiles evaluated with an Agilent 2100 Bioanalyzer.
Sequencing was carried out by the Genoscope (Evry, France) using Illumina Hiseq 2500 technology (2x250 bp) for metabarcoding and Illumina NovaSeq 6000 (2x150 bp) for metatranscriptomics (Illumina, San Diego, CA).

**Amplicons analysis and taxonomic affiliation**

Sequences datasets of the same conditions were pooled and processed jointly using DADA2 pipeline (v1.14.1) (Callahan *et al.*, 2016) to infer Amplicon Sequence Variants (ASVs). First, the clean-up procedures consisted of eliminating sequences that were under 200 bp in length, those with more than 2 expected errors and those containing Ns or sequencing errors in the primers. The remaining sequences were then trimmed to eliminate primer sequences using the cutadapt tool (Martin, 2011) and ASVs inference was carried out following the default parameters. The merging step was performed with no mismatch and 50 bp minimum overlap; and the chimeras were removed using the 'consensus' identification method, as well as singletons. Finally, each ASV was affiliated with RDP Naïve Bayesian Classifier algorithm (Wang *et al.*, 2007) with kmer size 8 and 100 bootstrap replicates using the PR2 database as reference (v4.14.0) (Guillou *et al.*, 2013). The effect of the pooling procedure on specific richness and sequence abundance was assessed using the Wilcoxon-Mann-Whitney test.

Comparisons between datasets resulting from the duplicates grouping and the others (without duplicates) showed no significant difference in richness ($p$-value >0.5) and a significant effect on sequence abundance ($p$-value <0.001) (**Fig. S1**). To overcome this effect, datasets were normalized using rarefaction strategy (Weiss *et al.*, 2017) with the GUniFrac R package (v1.3) (Chen *et al.*, 2012).

### Taxonomic diversity analysis

Diversity analyses were performed using R (R Core Team, 2022). Specific richness, Pielou equitability and Shannon α-diversity index as well as rarefaction curves were computed to infer eukaryotic richness and sequencing efficiency, respectively, using the vegan package (v2.5-7) (Oksanen *et al.*, 2020). Correspondence Analysis (CA) was used to assess the influence of the conditions on the composition and abundance of eukaryotic taxa with the FactoMineR package (v2.4) (Lê *et al.*, 2008) and significance was evaluated by ANOSIM tests on Bray-Curtis distance using the vegan package (v2.5-7) (Oksanen *et al.*, 2020). Abundance refers to relative abundance inferred on the number of reads throughout the paper.

### Trait-based analysis and functional groups definition

Trait-based analysis was carried out using R (R Core Team, 2022), following the procedure and using the same traits described in Ramond *et al.* (2019) with the addition of the 'suspected trophism' trait recording trophic modes described in the literature. Single-celled eukaryotes were the target of this study, Metazoa, Embryophyceae, Lecanoromycetes and Lichinomycetes, and some Ascomycota and Basidiomycota (**Table S1**) identified as pluricellular organisms were therefore discarded. Briefly, trait annotation was performed on the 600 unique taxonomic affiliation of ASVs (called taxonomic references) using the existing trait database (Ramond *et al.*, 2018) (allowing the annotation of 30% of taxonomic references) and biological descriptions from the literature (**Table S1**). The high taxonomic rank as well as the lack of trait features in literature for few taxonomic references prevented the annotation of the totality of the morpho-physio-phenological traits.

Therefore, taxonomic references annotated with less than 5 traits (n = 37) have been set aside and identified as "Unassigned" in order to enable the clustering of those poorly described without miss-associating to functional groups those with too high taxonomic rank. We finally obtained a trait table composed of 563 taxonomic references and 14 morpho-physio-phenological traits (**Table S1**).

The Gower distance, which manages both qualitative and quantitative variables, was then calculated between each taxonomic reference on the basis of their trait characteristics. The resulting distance matrix was analyzed by Principal Coordinate Analysis (PCoA) (**Fig. S2**). According to Maire *et al.* (2015), and taking into account the trait table characteristics, 7 Euclidean dimensions were used to ordinate the 563 taxonomic references (**Fig. S2**). The structure of the PCoA showed the disparity between taxonomic references according to their trait characteristics and therefore suggested trait trade-offs following several PCoA dimensions. As an example, representing trade-off between cell cover, suspected trophism, motility and cell polarity; motility is an important feature of heterotrophic organisms to reach their preys or hosts, but it involves a special cell polarity and a light, non-rigid cell wall. Consequently, trait trade-offs investigation was carried out, following the procedure described in (Ramond *et al.*, 2019), by computing the correlation between trait features and each PCoA dimensions using a Spearman Rank test (**Fig. S3**). The most explanatory dimensions (*i.e.*, dimension 1 (36.4%) and 2 (27.3%)) highlighting trait trade-offs were taken into account for functional groups definition. The Simple-Structure-Index (SSI) criterion of the K-means clustering method supplied in the vegan package (v2.5-7) (Oksanen *et al.*, 2020) was used to determine the optimal number of cluster and thus partition our taxonomic references (**Fig. S4**). These clusters were displayed on the previous PCoA to confirm the dimension selection (**Fig. S5**) and were characterized by their composition in morpho-physio-phenological traits (**Fig. S6**). Since some clusters mainly differed by their morphometric traits (cell size, cover, shape or symmetry, etc.) rather than physiology (ability to form resting stage or colony, the presence of plast, the mode of ingestion, etc.), we chose to treat them together (**Fig. S6**). Thus, the two clusters of microbial eukaryotes reported as strict-heterotrophs (Cluster 4&7, **Fig. S6**) and the two clusters of microorganisms described as having a parasitic lifestyle called hereafter "parasites" (Cluster 1&5, **Fig. S6**) were merged to define 7 functional groups. Finally, each ASVs was linked to functional group using the taxonomic refence labels.

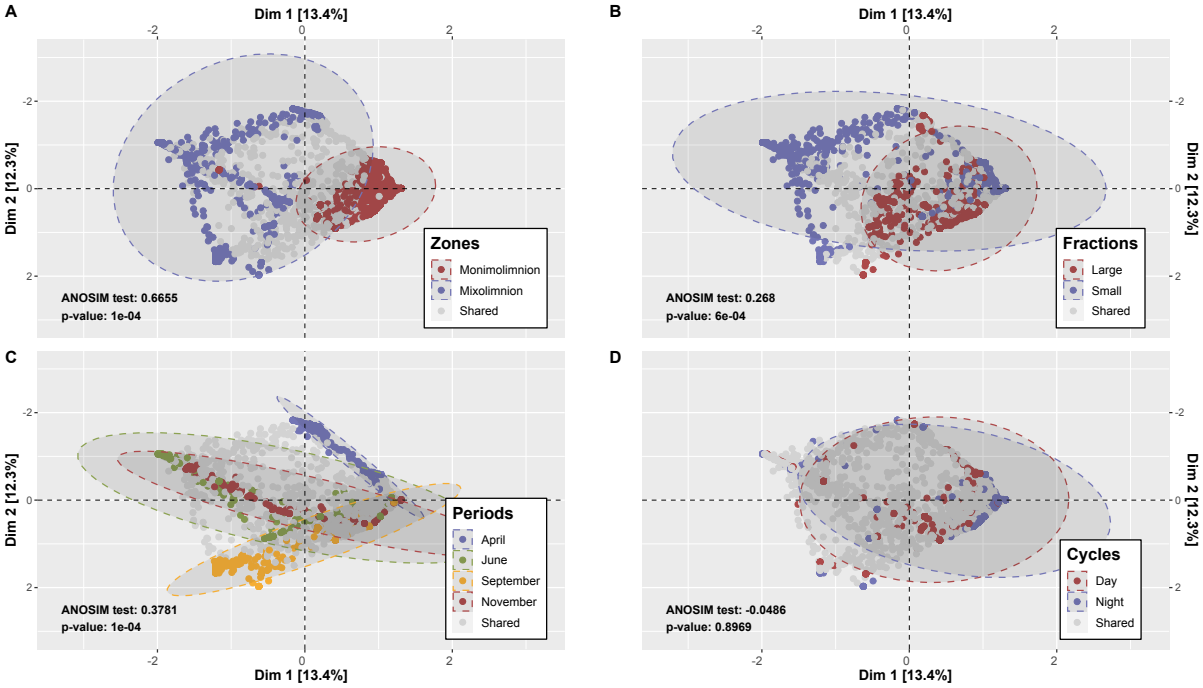



**Figure 2: Two-dimensional Correspondence Analysis (CA) testing the correspondence between sampling conditions and the abundance of each ASV.**

ASVs are represented by dots. The colored one are those whose abundance in a condition represents at least 90% of its total abundance. The ellipses represent 95% confidence intervals. ASVs that are not in the intersection of the two (A, B and D) or four (C) ellipses are characterized by an abundance influenced by the studied conditions (associated risk $\alpha = 5\%$).

**Unigene catalog and proteins prediction and annotation**

The assessment of unigenes expression was done as described in Carradec *et al.* (2018). Paired-ends reads from each metatranscriptomic sample were assembled using velvet (v1.2.07) with a kmer size of 89 as described in Carradec *et al.* (2018). Isoform detection was performed using Oases (v0.2.08). Contigs smaller than 150 bp were removed from further analysis. Contig redundancy was removed using CD-HIT-EST (v4.6.1), with the following parameters: *-id 95 -aS 90* (95% nucleic identity over 90% of the length of the smallest sequence) as described in Carradec *et al.* (2018). For each cluster of contigs, the longest sequence was kept as reference for the unigenes catalog. To estimate the expression of each unigene in each sample, cleaned reads were mapped against the reference catalog using the bwa tool (v0.7.15). The following parameters were used: *bwa aln -l 30 -O 11 -R 1; bwa sampe -a 20000 -n 1 –N; samtools; rmdup*. Low complexity reads were removed. Reads covering at least 80% of read length with at least 95% identity were retained for further analysis. In the case of several possible best matches, a random one was picked.

Proteins were predicted from all unigenes with *Transdecoder.LongOrfs* followed by *TransDecoder.Predict* (v5.5.0) using the default parameters. Then, unigenes without predicted protein were used for a second run with a minimum protein length of 70 *(-m*). Finally, the predicted proteins were tested against the AntiFam database (v7.0) (Eberhardt *et al.*, 2012) with *hmmsearch* using the *--cut_ga* parameter (Eddy, 2011).

The KEGG Orthology (KO) identifiers were assigned by KoFamScan (v1.3.0) with the KO's HMM profiles (2022-01-03 release). For proteins without significant hit, the best hit with an e-value <1e-5 was retained as described in Hu *et al.* (2018).

Taxonomic affiliation was performed on proteins with the MMseqs2 suite (v407b315) (Steinegger and Söding, 2017), against the pre-formatted MetaEuk database (Levy Karin *et al.*, 2020) summarizing Uniclust90 and MMETSP (Keeling *et al.*, 2014; Johnson *et al.*, 2019). Taxonomy was assigned with mmseqs taxonomy and the parameters *--tax-lineage 1 --lca-mode 2 --max-seqs 100 -e 0.00001 -s 6 --max-accept 100*. The unigene catalog was cleaned of contaminants by excluding proteins and unigenes affiliated to Human, Bacteria, Archaea, Virus and Metazoans.

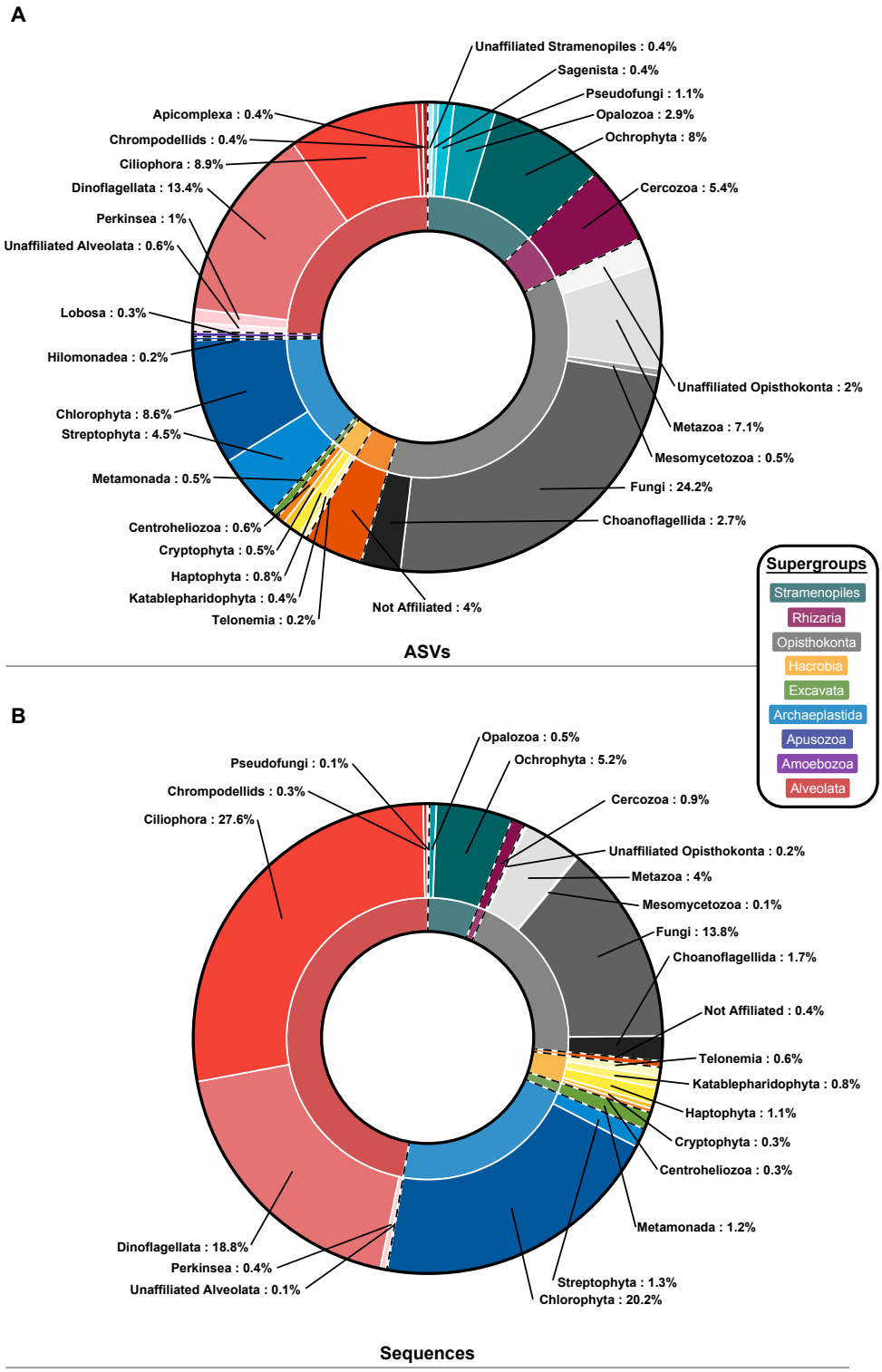



**Figure 3: ASVs richness (A) and sequences abundance (B) of microbial eukaryotes in lake Pavin.**

PR2 taxonomy is used to classify the entire eukaryotic diversity within 8 ranks (*i.e.*, Domain / Supergroup / Division / Class / Order / Family / Genus / Species). Supergroups ($2^{nd}$ rank) are displayed in the center of the round histogram while the divisions ($3^{rd}$ rank) are placed on the outside.

### Differential gene expression analysis

Conditions effects were first assessed by ANOSIM tests (vegan v2.5-7) (Oksanen *et al.*, 2020) on the Bray-Curtis distance using the full raw unigene counts. No dissimilarities are highlighted between day and night (R value: -0.03; ANOSIM *p*-value >0.5), while significant differences are recorded (ANOSIM *p*-value <0.001) across size fractions (R = 0.34), sampling zones (0.41) and periods (0.69), thus day/night samples were considered as biological replicates. Unigenes were then associated to functional groups using their precise taxonomic affiliations (**Fig. 1**) when match with taxonomic references exists. Only unigenes associated with a functional group and which could be assigned to a KEGG-BRITE 'cellular process' and/or 'metabolism' (**Table S2**) were retained. Differential Gene Expression (DGE) analysis was performed on KO associated to functional groups and taxonomic classes with the DESeq2 package (v1.38.2) (Love *et al.*, 2014) and the following model: *~size Fractions+Periods+Zones*. DGE were further filtered using the parameters: baseMean (*i.e.*, the average normalized count of transcripts, dividing by size factors, taken over all samples (Love *et al.*, 2014)) >10, adjusted *p*-value <0.05 and absolute value of log2FoldChange value >2.

## Results

### Eukaryotic diversity

Thirty-two datasets corresponding to the V4 region of the 18S rRNA gene are analyzed (*i.e.*, 10 002 064 paired-end reads with an average of 312 564 sequences per datasets). Among these sequences, 16.1% are removed after clean-up procedures. The remaining sequences are split into 3 727 ASVs among which 331 non-eukaryotic ASVs and 5 singletons are removed. Rarefaction curves show that a plateau is reached for all samples (**Fig. S7**). The diversity of eukaryotic communities is analyzed after a rarefaction step reducing the samples depth to 138 888 sequences without significant richness reduction (*p*-value = 0.78): samples range from 235 to 634 ASVs with a total of 3 358 ASVs.

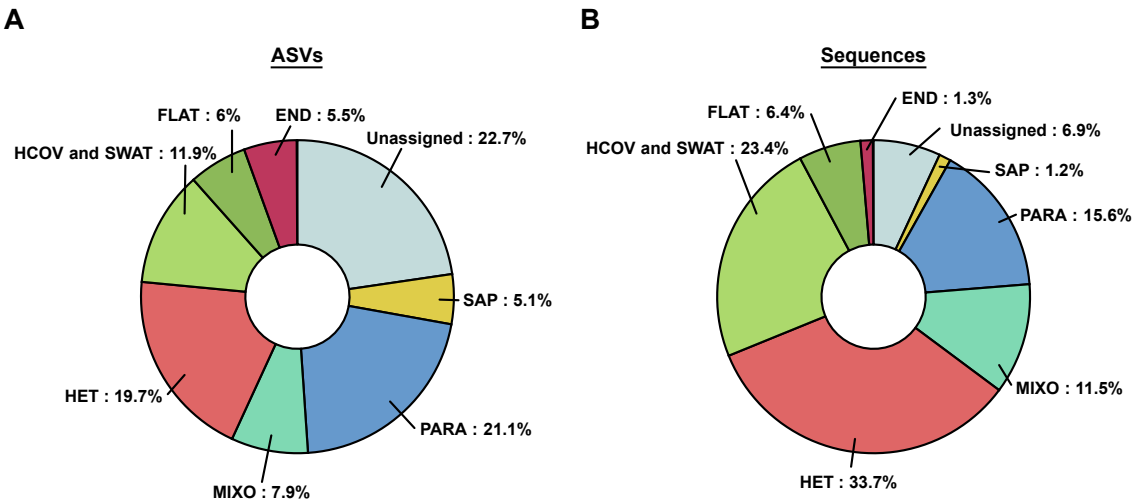


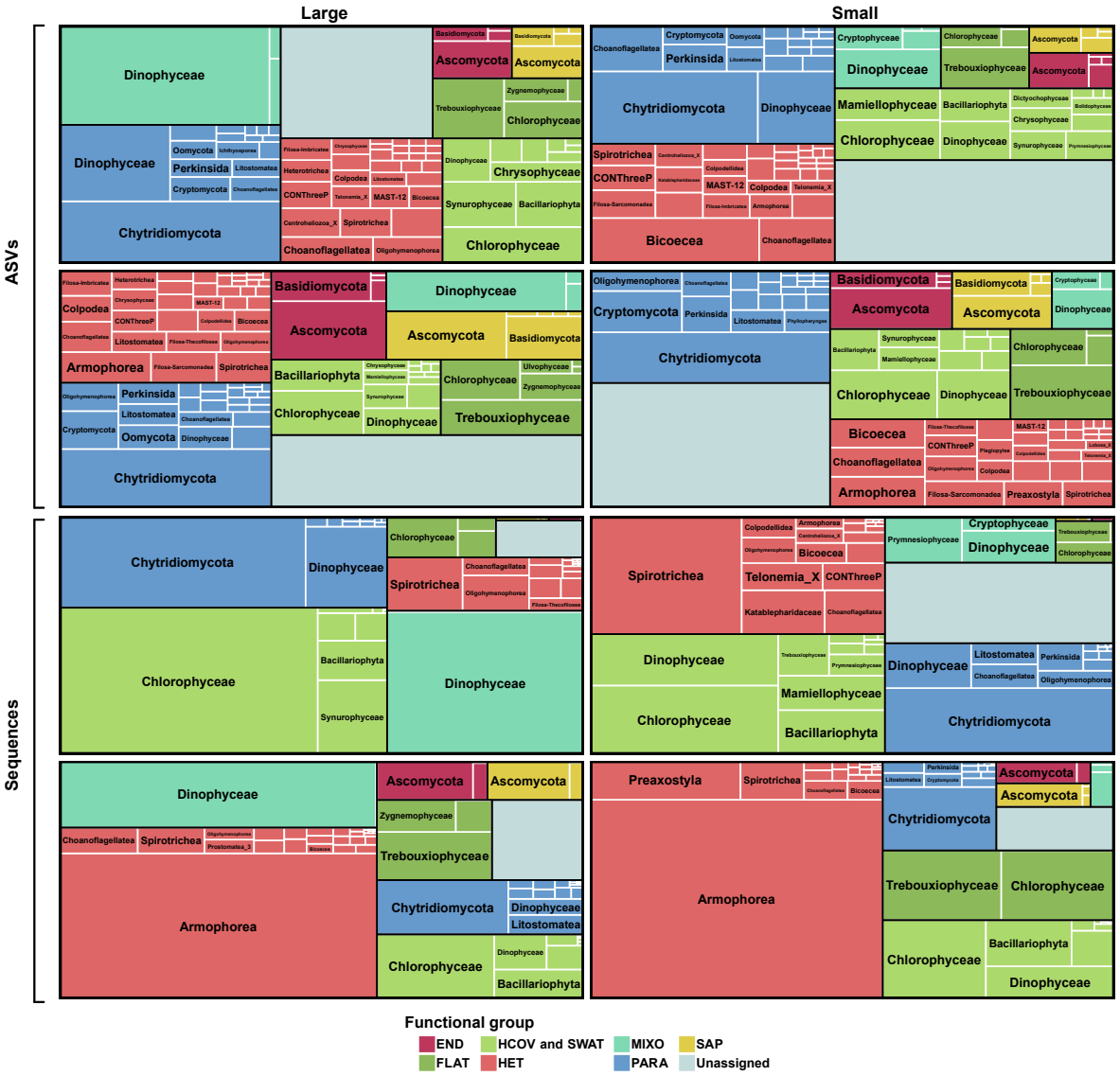

**Figure 4: ASVs richness (A) and relative abundance (B) of the functional groups in lake Pavin.**

Functional groups are named as follows: MIXO: Mixoplankton; HCOV&SWAT: Heavy-cover- and Swimmer-photoautotrophs; PARA: Parasites; HET: Strict-heterotrophs; SAP: Saprotrophs; END: Endophytes; FLAT: Floater- and Colonial-photoautotrophs. The group 'unassigned' corresponds to ASVs (A) or sequences (B) affiliated to taxonomic references characterized by less than 5 traits.

Correspondence analysis and ANOSIM tests reveal significant discrimination of the eukaryotic diversity between monimolimnion and mixolimnion as well as between small and large size fraction on the first dimension (ANOSIM *p*-value <0.001) (**Fig. 2A-B**) while the second dimension differentiates the sampling periods (**Fig. 2C**); dissimilarities between periods being significant only in the mixolimnion (R = 0.76; *p*-value <0.001 *vs* R = 0.20; *p*-value >0.01). However, this analysis shows no discrimination between day and night samples (**Fig. 2D**).
Eukaryotic α-diversity varied among samples (**Fig. S8**), the highest diversity being recorded in the mixolimnion for small sized microbial eukaryotes. For both size fraction samples, an important rare biosphere characterized by numerous low abundance ASVs (<0.002% of sample abundance) is observed in particular in the monimolimnion (**Figs. S8, S9**).
Taxonomic affiliation results in 96% of the ASVs are affiliated at the 'Supergroup' level and 63.5% at the 'Species' level (**Fig. S10**). ASVs are spread into 9 supergroups (**Fig. 3A**). Opisthokonta (1 226 ASVs), Alveolata (827), Archaeplastida (442) and Stramenopiles (427) accounting for almost 87% of all ASVs (94% of the total sequences). In term of abundance (*i.e.*, sequence numbers), Alveolata, represented in majority by Ciliophora (27.6% of the sequences) and Dinoflagellata (18.8%), dominates the eukaryotic community. Chlorophyta, Fungi, and Ochrophyta are also abundant with 20.2%, 13.8% and 5.2% of the sequences, respectively (**Fig. 3B**).

**Functional groups distribution**

Using modalities of the 14 traits, 7 functional groups are defined and characterized (Fig. S6): Saprotrophs (SAP), Strict-heterotrophs (HET), Parasites (PARA), Mixoplankton (MIXO) (here considered as photo-osmo-phago-mixotrophs (Mitra *et al.*, 2023)), Floater- and colonial-photoautotrophs (FLAT), Heavy-cover- and Swimmer-photoautotrophs (HCOV&SWAT) and Endophyte (END) (see supplementary material 1 for the characterization of the functional group). Among ASVs affiliated to unicellular organisms, 77.3% are associated to functional groups: SAP (149 ASVs), HET (569), PARA (611), MIXO (230), FLAT (175), HCOV&SWAT (345) and END (158) (**Fig. 4A**). The relative abundance of each functional group shows the dominance of HET (33.7%), HCOV&SWAT (23.4%) and PARA (15.6%) (**Fig. 4B**).




Functional group

END HCOV and SWAT MIXO SAP

FLAT HET PARA Unassigned

**Figure 5: ASVs richness and relative abundance of the eukaryotic classes across each functional group according to sampling zone and size fraction in lake Pavin.**

Richness and abundance of eukaryotic classes (4th rank according to PR2 taxonomy) combined to the richness and abundance of functional groups is displayed for each sampling zone and size fraction. Functional groups are named as described in Figure 4.

The diversity of microbial eukaryotes, assessed for each functional group, highlights their polyphyletic nature (**Fig. S11**). SAP and END are mostly composed of Fungi and Pseudofungi while HET are composed by many divisions, among which Ciliophora, Cercozoa, Opalozoa and Choanoflagellida, altogether representing the majority (*i.e.*, 80.8% of HET-related ASVs). PARA include many divisions such as Fungi, Ciliophora and Dinoflagellata. ASVs related to Cryptophyta, Haptophyta and Dinoflagellata group together and form the MIXO group; Dinoflagellata accounts for almost 90% of MIXO-related ASVs. FLAT, are mainly composed of Chlorophyta, Streptophyta and some Ochrophyta. Finally, HCOV&SWAT ASVs are affiliated to Dinoflagellata, Chlorophyta, Ochrophyta and Haptophyta.

Richness and relative abundance of each functional group show variations through the sampling conditions (**Figs. 5, 6**). PARA dominate microbial diversity across zones, representing more than 20% of total richness and are twice as abundant in the mixolimnion (**Fig. 5**). They are mainly composed by Chytridiomycota and Dinophyceae whose abundance vary across seasons, with highest Chytridiomycota abundance during water mixing of lake Pavin occurring in April and Autumn (**Figs. 6, S12**). Regardless of the size fraction, although FLAT richness is comparable between the two zones, they are more abundant in the monimolimnion (**Fig. 5**) and are mainly affiliated to Trebouxiophyceae, Chlorophyceae (Chlorophyta) and Zygnemophyceae (Streptophyta). Monimolimnion is also characterized by higher SAP and END diversity and abundance (**Fig. 5**), with a majority of Ascomycota and Basidiomycota (Fungi). On the other hand, the mixolimnion displays a higher diversity of HCOV&SWAT (*i.e.*, 15.2% *vs* 11.7% of the richness and 33.5 *vs* 12.7% of the total abundance), among which numerous ASVs are affiliated to Bacillariophyta, Dinophyceae and Chlorophyceae. Being abundant and rich in the mixolimnion, MIXO also show important temporal variations with a maximum abundance in November (**Figs. 6, S12**). Furthermore, their composition differs greatly across size: MIXO >10 µm being mainly composed by Dinophyceae, while those <10 µm are affiliated to Prymnesiophyceae, Dinophyceae and Cryptophyceae.

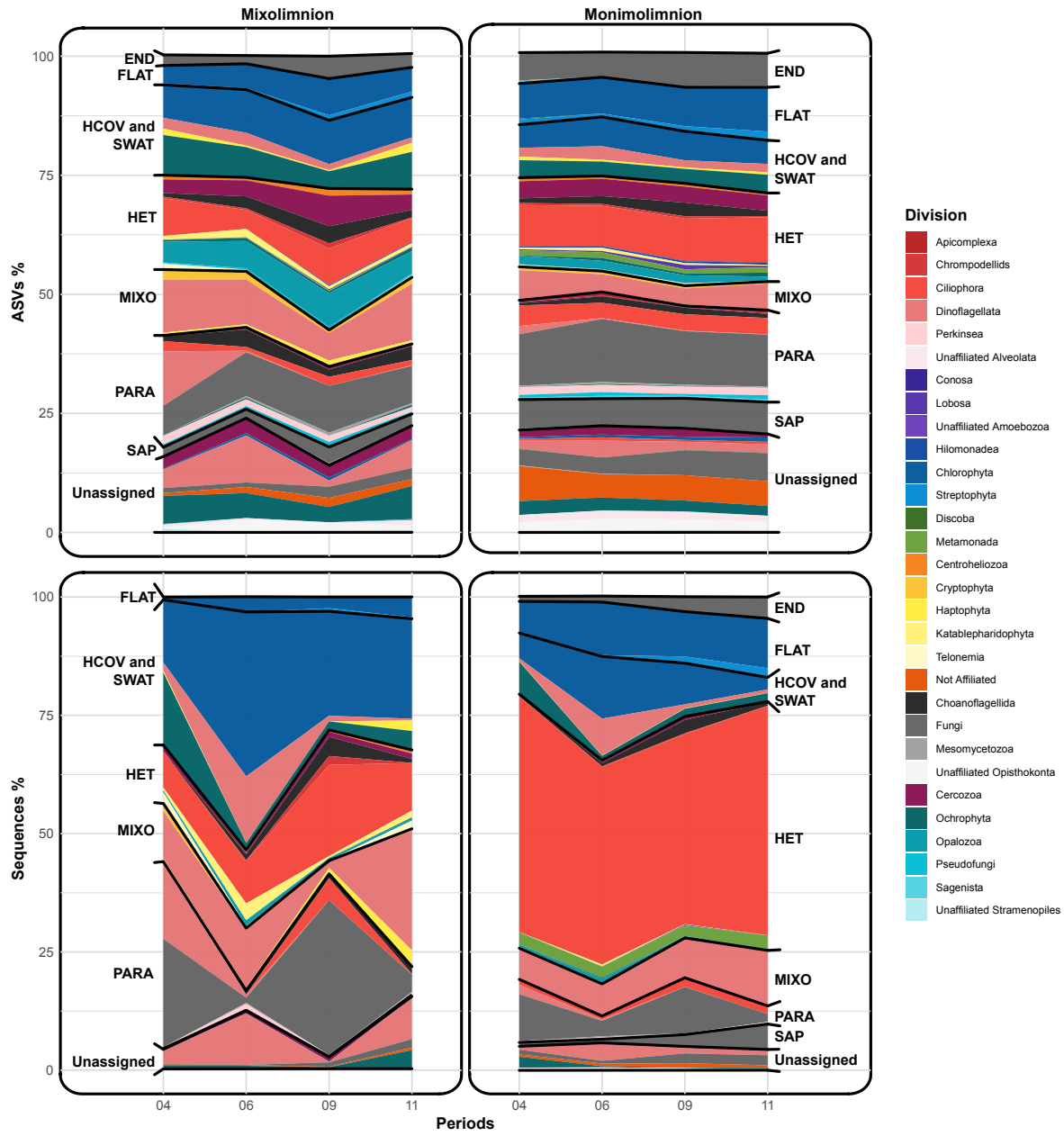



**Figure 6: ASVs richness and relative abundance of the eukaryotic classes across each functional group according to sampling periods and zone in lake Pavin.**

Richness and abundance of eukaryotic classes (4$^{th}$ rank according to PR2 taxonomy) combined to the richness and abundance of functional groups is displayed for each sampling periods and zone. Functional groups are named as described in Figure 4.

**Metabolic potential of functional groups across sampling conditions**

Annotation of assembled unigenes from all samples provide 9 182 313 proteins from which 30.7% are assigned to a KO identifier, 30.8% are taxonomically affiliated to the eukaryotic domain, 23.9% to the 'Supergroup' level and 11% to the 'Species' level allowing for the assignment of 7.6% of proteins to a functional group. Therefore, 57 324 402 unigenes coding for 290 458 proteins are assigned to both a KO identifier and a functional group.

Regardless of size fraction, Ascomycota (Fungi) related to END and SAP of the monimolimnion (**Figs. 7, 8, S13, S16**), over-express genes coding for proteins related to fungal development (*e.g.*, K11427: histone-H3-lysine79-N-trimethyltransferase involve in conidiation (Li *et al.*, 2019), K19029: fructose-2,6-biphosphatase which have a role in spore germination (Van Laere *et al.*, 1983)) and catabolism enzymes such as sulfide-quinone-oxidoreductase, nitrate-reductases, sorbose-reductases, α-ketoglutaric-semialdehyde-dehydrogenase and several alcohol dehydrogenases involve in xylose and lignin degradation and fermentation pathways (respectively, K22470, K10534, K17742+K19633, K22187 and K13979+K13953+K18369) (Kanehisa *et al.*, 2016).

Autotrophs (*i.e.*, FLAT, HCOV&SWAT) affiliated to Trebouxiophyceae (Chlorophyta) and Bacillariophyta (Ochrophyta) also over-express numerous genes in the monimolimnion (**Figs. 8, S13, S16**) (with clear over-expression of Bacillariophyta's genes in April and September compared to others periods (**Figs. S15, S17**)). Among these, HCOV&SWAT-related genes are involved in programmed cell death under environmental stress (*e.g.*, K13711: NAD(P)H-dehydrogenase-(quinone) and K19267: phosphatidylinositol-4-kinase) (Akhter *et al.*, 2016; Herrmann and Riemer, 2021), or encode for proteins with antioxidant activity (*e.g.*, K15777: 4,5-DOPA-dioxygenase-extradiol); while those related to FLAT encode for light-harvesting complexes I&II chlorophyll binding proteins (K08907, K08916), photosystems subunits (K08901, K08905, K14332) and proteins involved in endocytosis such as sorting-nexin and ADP-ribosylation-factor (K12493, K17917) (Kanehisa *et al.*, 2016). On the other hand, over-expressed genes affiliated to Mamiellophyceae (Chlorophyta), Synurophyceae (Ochrophyta) as well as Chlorophyceae (Chlorophyta) linked to HCOV&SWAT are mostly found in the mixolimnion (**Figs. 8, S13, S16**).

Although MIXO-related over-expressed-genes are essentially found in the mixolimnion, their occurrence seems to depend on size fraction and taxonomic affiliation (**Figs. 7, 8, S13, S14, S16**).

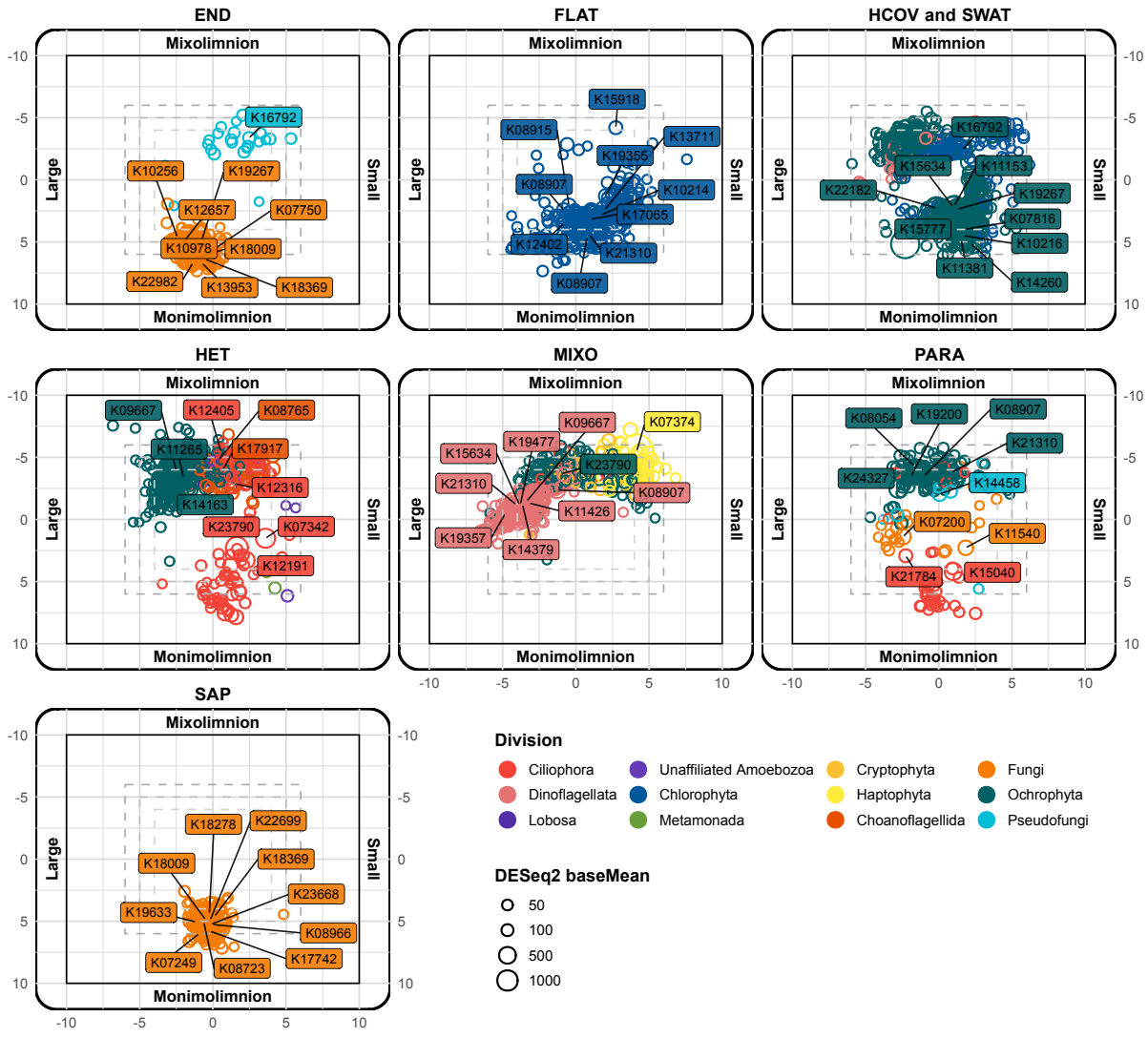



**Figure 7: Differential expression analysis of genes across size fraction and sampling zone for each eukaryotic division and functional group.**

Genes categorized in KO identifiers are represented by dots that are distributed along x-axis corresponding to the size fraction and y-axis corresponding to the sampling zone. Each axis displays the log2-FoldChange of each KO identifiers along conditions. The dots size represents the average normalized count of transcripts, dividing by size factors, taken over all samples (= DESeq2 baseMean) while their colors represent eukaryotic divisions (3rd rank according to PR2 taxonomy). Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across at least one of the two conditions is characterized by an adjusted *p*-value >0.05 and an absolute value of log2-FoldChange >2 are represented. For each functional group, only the 10 most abundant (DESeq2 baseMean) KO identifiers are labeled. Functional groups are named as described in Figure 4.

Prymnesiophyceae (Haptophyta) genes over-expression occur in the small size fraction in September and November, while transcript affiliated to Dinophyceae (Dinoflagellata) are enriched in the large size fraction, in November and April (**Figs. S14, S15, S16, S17**). Chrysophyceae (Ochrophyta) seem to over-express genes whatever the size fraction and preferentially in November (**Figs. S14, S15, S17**). Despite these differences, most of them seem to be involve in nutrient acquisition such as α-tubulin (K07374, which may be linked to the Haptophyta's haptonema structure (Lechtreck, 2004) and are characterized by very high transcripts abundance (**Fig. 7**)), photosynthesis (K08907, K14332 and K08902) or encode for proteins associated to clathrin-mediated endocytosis (*e.g.*, K12472: Epidermal-growth-factor-receptor-substrate-15 in Haptophyta; K11824: AP-2-complex in Dinoflagellata and Cryptophyta; K10396: Kinesin in Ochrophyta) (Kanehisa *et al.*, 2016).

Spirotricheae (Ciliophora) related to HET displays genes over-expressed in the mixolimnion as well as in the monimolimnion whatever the size fractions (**Figs. 7, 8, S13, S14, S16**) suggesting the high adaptability of their metabolisms across conditions while Chrysophyceae (Ochrophyta) and Choanoflagellatea (Choanoflagellida) related to HET seem to be more specific to the mixolimnion, over-expressing endocytosis and phagotrophic-related genes (*e.g.*, K12493: ADP-ribosylation-factor-GTPase-activating-protein, K17917: sorting-nexin, K10413: dynein-cytoplasmic-1, K13711: phosphatidylinositol-4-kinase) (Kanehisa *et al.*, 2016).

Regarding PARA affiliated to Fungi (*i.e.*, Chytridiomycota and Cryptomycota), very few genes are differentially expressed according to the sampling zone. However, high variation of gene expression between stirring (spring and autumn) and stratification (summer and winter) periods (**Figs. 8, S13, S14, S15, S16, S17**) as well as across large and small size fraction are recorded. Indeed, genes potentially involved in parasitic lifestyle (*e.g.*, K19477: cGMP dependent protein kinase) (Baker *et al.*, 2020) or in phototaxis of zoospore (K12319: guanylate cyclase) (Medina and Buchler, 2020) are over-expressed in April and September. Moreover, genes involved in glycan degradation (*e.g.*, K12373: Hexosaminidase), endocytosis vesicle formation (K12471: Epsin, K12562: Amphiphysin) and coding for lipase (*i.e.*, K14452 and K17900) are significantly more expressed in the >10 µm size fraction while glycolipids anabolisms pathways -related transcripts (*e.g.*, K09881 and K22831) are enriched in the <10 µm size fraction.


**Figure 8: Heatmap representing differential expression of genes across sampling zones for each eukaryotic classes and functional group.**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across zone is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are represented. Furthermore, only differentially expressed KO in at least 5 taxonomic classes among functional groups are displayed. Values of log2-FoldChange >4 are notified with « · », log2-FoldChange >5 with « ∗ » while log2-FoldChange >6 with « ∗∗ ».

# Discussion

## Methodological considerations

Our analysis provides interesting insight into functional diversity and metabolic activities of microbial eukaryotes. However, some methodological limitations need to be kept in mind. 1) Assessment of the taxonomic and functional diversity could be influence by affiliation and annotation methods as well as by databases that have been mainly developed from non-freshwater sequences (*e.g.*, PR2 and MetaEuk). 2) Traits/features are referenced from described organisms and therefore does not fit the unknown diversity. 3) The intrinsic limitations of the trait-based approach such as the number, the proportion and the quality of annotations along with the nature of traits could influence drastically functional groups inference. To deal with these biases, we used a trade-off approach to select only the traits well-annotated and well-correlated as shown in Ramond *et al.* (2019). We also added supplemental cautions compared to other studies: (i) a 'suspected trophism' trait was used to buffer mis-annotations and help to identify clusters; (ii) taxonomic references were characterized with at least 5 traits.

Despite these precautions some biases remain and influence functional groups definition. References annotated at a high taxonomic level (*i.e.*, division) can be associated with the wrong functional group (*e.g.*, unaffiliated Ciliophora are associated to PARA while they could include heterotrophs and non-consecutive mixoplankton (*e.g.*, Mesodinium genus which has kleptoplasty capacity)). The complexity of some functional group such as parasites and mixoplankton will remain difficult to deal with. For example, characteristic features of parasites such as the 'attached lifestyle' and 'ability to produce resting stage' influence the clustering of other organisms with the same traits (*e.g.*, Dinobryon (Ochrophyta) associated to PARA whereas it belongs to mixoplankton (Caron *et al.*, 1993)). Although these issues do not have major consequences in the metabarcoding analysis as they only affect a few ASVs with low abundance, they may have great impact when the trait table is used to annotate other dataset such as metatranscriptomic where differential expression is analyzed. To prevent these spurious associations, other clustering method could be explored to establish functional groups such as fuzzy clustering which determines for each taxonomic reference a weighting value associated with each cluster (Ferraro *et al.*, 2019). By setting confidence thresholds, we can exclude doubtful affiliations. However, this approach has an unintended consequence: it significantly reduces the number of references that form a functional group.

## Functional groups distribution and metabolic potential of microbial eukaryotes in lake Pavin

### i. Conditions effect

Among the studied conditions, day/night cycle is the only one that provide no difference on richness, relative abundances (**Fig. 2**), and metabolic potentials of all functional groups. While these results were predictable for DNA based analysis these findings are surprising when considering the metatranscriptome, as microbial eukaryotes in surface waters are known to alternate between energy acquisition metabolisms (*i.e.*, oxidative phosphorylation and photosynthesis) during the day and housekeeping activities (*e.g.*, biosynthesis of amino acids, vitamins and membrane) at night (Poretsky *et al.*, 2009; Trench-Fiol and Fink, 2020). It is likely that ANOSIM tests taking into account the entire dataset does not allow the detection of metabolism variations which could affect the expression level of only a few genes in potentially only a few organisms, especially when these variations are assumed to occur only in the mixolimnion. Thus, further elements, such as the interaction of conditions remain to be studied to confirm this hypothesis.

Size fraction impacts functional groups. Two patterns can be observed, (1) size affects taxonomic diversity of functional groups, leading to metabolic variations. This is the case for mixoplankton, which switch from Dinophyceae in the large to Prymnesiophyceae in the small size fraction. (2) Metabolic variations were also recorded event though size fraction is not linked with change in taxonomic diversity. This is the case for parasites, whose members remain predominantly affiliated to Chytridiomycota regardless of the size fraction. Those of the large size fractions over-express genes which may be likely involved in attached lifestyle during which they have to degrade hosts' cell wall and hijack its lipid reserves (Gerphagnon *et al.*, 2019). On the contrary, those of the small size fraction over-express genes involved in glycolipids anabolisms, which has recently been described to take place during zoospore germination by Laundon *et al.* (2022). Finally, some groups do not seem to be affected by size fraction such as saprotrophs for which neither taxonomic diversity nor metabolic potential variations are detected.

### ii. Seasonal variation in the mixolimnion

Significant diversity temporal variations are recorded only in the mixolimnion of lake Pavin (**Figs. 2, 6, S8**) as already reported by Lepère *et al.* (2006). The turnover between seasonal stirring occurring in spring and autumn and stratification periods in the upper water layer appears to affect eukaryotic communities, especially Fungi and Ochrophyta (**Figs. 6, S12**). Chytridiomycota belonging to the parasites functional group display huge sequence abundance during stirring periods suggesting water mixing may promote germination of resting spore as described in Ibelings *et al.* (2004). This is also supported by the metatranscriptomic data with two unigenes highly expressed in spring and autumn and potentially involved in zoospore phototaxis (*i.e.*, guanylate cyclase (Medina and Buchler, 2020)) and in stage transition within parasite cycle (*i.e.*, cGMP-dependent protein kinase (Baker *et al.*, 2020)). These observations can also be related to the high abundance of Bacillariophyta in April, which are known as hosts of Chytridiomycota (Van den Wyngaert *et al.*, 2022).

Except in September, Dinoflagellata are highly abundant and diverse, especially in the large size fraction in the mixolimnion (**Figs. 6, S12**) and their taxonomy coupled to functional modes appear to vary greatly across time (**Fig. 6**): in April, most of them are related to parasites and mixoplankton respectively affiliated to *Chytriodinium* and *Apocalathium+Gymnodinium* genera while June is characterized by a majority of heavy-cover- and swimmer-photoautotrophs (*Prorocentrum*) and mixoplankton (*Peridinium* and *Ceratium*). As *Chytriodinium* has only been described in marine ecosystems as parasite of crustacean egg (Bass *et al.*, 2021; Gómez *et al.*, 2009), it is likely that they parasite copepods eggs which are abundant in the lake Pavin during spring stirring (Lair, 1975). Mixoplankton are also well represented in term of abundance and diversity in the dataset, ingestion of preys that are rich in nutrients, relative to the dissolved phase, may provide them an advantage over strictly autotrophs especially when nutrient concentration is low in lake Pavin (Lepère *et al.*, 2006). Mixoplanktonic Dinoflagellates are abundant across the year while Haptophyta (*Chrysochromulina* genus) are found preferentially in November. These temporal variations could reflect gradient in phago-mixotrophy observed across classes and even genera (Fischer *et al.*, 2022; Wilken *et al.*, 2020). Dinoflagellates (*Apocalathium* and *Gymnodinium* genera) could be obligate mixoplankton favored by high light availability and stable stratified conditions in summer while Haptophyta and dinoflagellates belonging to *Peridiidium* and *Ceratium* genera found in November when the environmental conditions are more variable could be facultative mixoplankton (Edwards *et al.*, 2023).

In addition to photosynthesis and light harvesting pathways, Haptophyta over-express endocytosis- and phagosome-linked genes as well as α-tubulin coding gene (K07374 in **Figs. 7, 8**) which is involved in haptonema structure, confirming their grazing activity (Kanehisa *et al.*, 2016; Lechtreck, 2004). Furthermore, although metabarcoding data show the presence of mixoplankton in both zones, expression of genes involved in phagotrophy seem to take place preferentially in the mixolimnion. This might be due to the combined effects of low temperature and darkness in the monimolimnion since phagotrophy seem to be dependent of light availability and temperature in numerous mixoplankton (Edwards *et al.*, 2023; Wilken *et al.*, 2020). Greater irradiance could make mixoplankton more competitive in the mixolimnion against specialist heterotrophs which are more abundant in the monimolimnion of lake Pavin.

### iii. Specificity of the monimolimnion

Even though the monimolimnion is not affected by temporal variation, its community composition and metabolic potential is extremely different from the oxic layer (**Figs. 5, 6, 8**). Some groups appear to be zone-specific such as saprotrophs and endophytes which are found almost exclusively in the monimolimnion (**Figs. 6, S12**). These functional groups are mainly composed by Ascomycota and Basidiomycota, essentially known as terrestrial fungi (*e.g.*, associated or not with plants) which can be introduced in lakes through rainwater and wind events (Lepère *et al.*, 2019; Voronin, 2014). Although they often fail to establish a stable population in aquatic environment, as described by Graupner *et al.* (2017), their ASVs richness remains stable across the year. Moreover, some of them are already described as active in freshwater ecosystems (Lepère *et al.*, 2019). Here, saprotrophs and endophytes highly express several genes in the monimolimnion related to lignin degradation, nitrate and sulfur metabolism as well as L-sorbose catabolism (Kanehisa *et al.*, 2016), which is toxic for Fungi, suggesting their saprotrophic activity at the bottom of the lake and their contribution to the trophic food web functioning *via* the biological pump.

Surprisingly, floater, colonial and heavy-cover photoautotrophs affiliated to Chlorophyceae, Trebouxiophyceae, Zygnemophyceae and Bacillariophyta are substantially abundant in the monimolimnion (**Fig. 5**). These groups are likely to accumulate in the monimolimnion due to the sedimentation process suggesting their inactivity in this zone.

Nevertheless, we also detect important over-expression of their genes (*e.g.*, some linked to photosynthesis and pigment biosynthesis (K15777: 4,5-DOPA dioxygenase extradiol) (**Figs.7, 8**). As floater and colonial photoautotrophs are more abundant in the monimolimnion, it seems difficult to conclude on their transcript's over-expression, however, this is different regarding heavy-cover photoautotrophs affiliated to Bacillariophyta whose abundance is maximum in the mixolimnion. Being unable to resist the sedimentation process, they might try to capture as much light as possible by overexpressing light-harvesting pathway when sinking (Gandía-Herrero and García-Carmona, 2012). However, at 80m, without light, it is unlikely that these genes are expressed for photosynthesis but are rather involved in antioxidant activity to manage the redox balance in order to inhibit apoptosis and support autophagy process (Gandía-Herrero and García-Carmona, 2012; Nagakannan *et al.*, 2016). Furthermore, metabolisms such as cell-wall auto-fermentation induce under starvation may be used to survive, thus maintaining a metabolic activity (Halim *et al.*, 2019). Nevertheless, these hypotheses may not be sufficient to explain the over-expression of a majority of Bacillariophyta-related genes in the monimolimnion. The presence of these groups in this zone could therefore be linked to anaerobic metabolism. Such metabolism has been highlighted to date in only a few green algae and diatoms (Atteia *et al.*, 2013) being mostly based on mixed-acid type fermentation. However, we did not find any gene linked to this pathway in our reduced dataset.

## Conclusion

While our knowledge about microorganisms' roles and diversity in the ocean has increased tremendously due to recent advances of -omics approaches, microbial eukaryotes and especially those thriving in freshwater are largely unexplored. In this context, we developed an original workflow to study taxonomic and functional diversity of microbial eukaryotes communities as well as their distribution across size fractions, oxygen gradient and seasons. This approach allows to highlight microbial eukaryote ecological strategies according to these variables. Parasites (*i.e.*, fungi) and mixoplankton are strongly impacted by seasonal variations, parasites in relation to their phytoplankton hosts and mixoplankton in relation to genera trophic abilities (*e.g.*, obligate/facultative mixoplankton). We for example show the small haptophytes could be facultative mixoplankton over-expressing photosynthesis-, endocytosis- and phagosome-linked genes under nutrients limitations conditions.

Numerous microbial eukaryotes are also able to thrive in absence of oxygen such as saprotrophs expressing transcripts related to sulfur and nitrate metabolism as well as organic matter degradation in the monimolimnion, photosynthetic microbial eukaryotes are also present in this zone raising the hypothesis of unknown anaerobic metabolisms.

Altogether, this study provides also a morpho-physio-phenological traits database for freshwater microbial eukaryotes as well as a significant amount of reference data for metatranscriptomes annotation.

This work carried out in a single temperate meromictic lake, should be of course complete by additional study in different freshwater ecosystems in order to cover the microbial eukaryotes global diversity. Furthermore, substantial efforts in microbial eukaryotes cultivation are required to best described their metabolisms and therefore complete protein databases enabling more detailed exploration of large environmental datasets.

**Acknowledgements**

The authors are grateful to the Mésocentre Clermont Auvergne University as well as AuBi platform for providing help, computing and storage resources. The work of AM was supported by a PhD fellowship funded by the 'Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation'.

**Funding statement**

This work was supported by France Génomique (ANR-10-INBS-09-08).

**Data Availability**

All data and material are available in this publication and additional information online (supplementary material and supplementary figures). Sequencing data are archived at ENA under accession number PRJEB61527 for metabarcoding and metatranscriptomic data.

**Code availability**

All scripts and procedure necessary to reproduce our analysis are publicly available at https://github.com/meb-team/MICROSTORE_MetaT-MetaB_Monjot_2023.

**Competing Interests**

The authors declare no competing interests.

## References


Akhter, S., Uddin, M.N., Jeong, I.S., Kim, D.W., Liu, X.-M., and Bahk, J.D. (2016) Role of *Arabidopsis* AtPI4Kγ3, a type II phosphoinositide 4-kinase, in abiotic stress responses and floral transition. *Plant Biotechnology Journal*, **14**, 215–230.

Atteia, A., van Lis, R., Tielens, A.G.M., and Martin, W.F. (2013) Anaerobic energy metabolism in unicellular photosynthetic eukaryotes. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, **1827**, 210–223.

Baker, D.A., Matralis, A.N., Osborne, S.A., Large, J.M., and Penzo, M. (2020) Targeting the malaria parasite cGMP-dependent protein kinase to develop new drugs. *Frontiers in Microbiology*, **11**, 602803.

Bass, D., Rueckert, S., Stern, R., Cleary, A.C., Taylor, J.D., Ward, G.M., and Huys, R. (2021) Parasites, pathogens, and other symbionts of copepods. *Trends in Parasitology*, **37**, 875–889.

Biderre-Petit, C., Charvy, J., Bronner, G., Chauvet, M., Debroas, D., Gardon, H., et al. (2022) FRESHOMICS: A manually curated and standardized –omics database for investigating freshwater microbiomes. *Molecular Ecology Resources*, **23**, 222–232.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**, 581–583.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4516–4522.

Caron, D.A., Sanders, R.W., Lim, E.L., Marras, C., Amaral, L.A., Whitney, S., et al. (1993) Light-dependent phagotrophy in the freshwater mixotrophic chrysophyte *Dinobryon cylindricum*. **Microbial Ecology**, **25**, 93–111.

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018) A global ocean atlas of eukaryotic genes. *Nature Communication*s, **9**, 373.

Chauvet, M., Debroas, D., Moné, A., Dubuffet, A., and Lepère, C. (2022) Temporal variations of Microsporidia diversity and discovery of new host–parasite interactions in a lake ecosystem. *Environmental Microbiology*, **24**, 1672–1686.

Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., et al. (2012) Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*, **28**, 2106–2113.

Debroas, D., Domaizon, I., Humbert, J.-F., Jardillier, L., Lepère, C., Oudart, A., and Taïb, N. (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiology Ecology*, **93**, fix023.

Dupont, C.L., McCrow, J.P., Valas, R., Moustafa, A., Walworth, N., Goodenough, U., et al. (2015) Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *The ISME Journal*, **9**, 1076–1092.

Eberhardt, R.Y., Haft, D.H., Punta, M., Martin, M., O'Donovan, C., and Bateman, A. (2012) AntiFam: a tool to help identify spurious ORFs in protein annotation. *Database (Oxford)*, **2012**, 1–5.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Computational Biology*, **7**, e1002195.

Edwards, K.F., Li, Q., McBeain, K.A., Schvarcz, C.R., and Steward, G.F. (2023) Trophic strategies explain the ocean niches of small eukaryotic phytoplankton. *Proceedings of the Royal Society B*, **290**, 20222021.

Ferraro, M., Brigida, Giordani, P., and Serafini, A. (2019) fclust: an R package for fuzzy clustering. *The R Journal*, **11**, 198.

Fischer, R., Kitzwögerer, J., and Ptacnik, R. (2022) Light-dependent niche differentiation in two mixotrophic bacterivores. *Environmental Microbiology Reports*, **14**, 530–537.

Flynn, K.J., Mitra, A., Anestis, K., Anschütz, A.A., Calbet, A., Ferreira, G.D., et al. (2019) Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? *Journal of Plankton Research*, **41**, 375–391.

Gandía-Herrero, F. and García-Carmona, F. (2012) Characterization of recombinant *Beta vulgaris* 4,5-DOPA-extradiol-dioxygenase active in the biosynthesis of betalains. *Planta*, **236**, 91–100.

Gerphagnon, M., Agha, R., Martin-Creuzburg, D., Bec, A., Perriere, F., Rad-Menéndez, C., et al. (2019) Comparison of sterol and fatty acid profiles of chytrids and their hosts reveals trophic upgrading of nutritionally inadequate phytoplankton by fungal parasites. *Environmental Microbiology*, **21**, 949–958.

Gómez, F., Moreira, D., and López-García, P. (2009) Life cycle and molecular phylogeny of the dinoflagellates *Chytriodinium* and *Dissodinium*, ectoparasites of copepod eggs. *European Journal of Protistology*, **45**, 260–270.

Graupner, N., Röhl, O., Jensen, M., Beisser, D., Begerow, D., and Boenigk, J. (2017) Effects of short-term flooding on aquatic and terrestrial microeukaryotic communities: a mesocosm approach. *Aquatic Microbial Ecology*, **80**, 257–272.

Grossart, H.-P., Wurzbacher, C., James, T.Y., and Kagami, M. (2015) Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi. *Fungal Ecology*, **19**, 28–38.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2013) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, **41**, D597–D604.

Halim, R., Hill, D.R.A., Hanssen, E., Webley, P.A., and Martin, G.J.O. (2019) Thermally coupled dark-anoxia incubation: a platform technology to induce auto-fermentation and thus cell-wall thinning in both nitrogen-replete and nitrogen-deplete *Nannochloropsis* slurries. *Bioresource Technology*, **290**, 121769.

Hartmann, M., Zubkov, M.V., Scanlan, D.J., and Lepère, C. (2013) In situ interactions between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean: evidence for mixotrophy: bacterivory by marine phototrophic picoeukaryotes. *Environmental Microbiology Reports*, **5**, 835–840.

Herrmann, J.M. and Riemer, J. (2021) Apoptosis inducing factor and mitochondrial NADH dehydrogenases: redox-controlled gear boxes to switch between mitochondrial biogenesis and cell death. *Biological Chemistry*, **402**, 289–297.

Hu, S.K., Liu, Z., Alexander, H., Campbell, V., Connell, P.E., Dyhrman, S.T., et al. (2018) Shifting metabolic priorities among key protistan taxa within and below the euphotic zone. *Environmental Microbiology*, **20**, 2865–2879.

Ibelings, B.W., De Bruin, A., Kagami, M., Rijkeboer, M., Brehm, M., and Donk, E.V. (2004) Host parasite interactions between freshwater phytoplankton and chytrid fungi (chytridiomycota). *Journal of Phycology*, **40**, 437–453.

Johnson, L.K., Alexander, H., and Brown, C.T. (2019) Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. *GigaScience*, **8**, giy158.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, **44**, D457–D462.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., et al. (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biology*, **12**, e1001889.

Lair, N. (1975) Sur la production des Copépodes dans deux lacs du Massif Central français. *SIL Proceedings*, **1922-2010 19**, 3204–3211.

Laundon, D., Chrismas, N., Bird, K., Thomas, S., Mock, T., and Cunliffe, M. (2022) A cellular and molecular atlas reveals the basis of chytrid development. *eLife*, **11**, e73933.

Lê, S., Josse, J., and Husson, F. (2008) FactoMineR an R package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–18.

Lechtreck, K.-F. (2004) An immunofluorescence study of the haptonema of *Chrysochromulina parva* (Prymnesiophyceae). *Phycologia*, **43**, 635–640.

Leles, S.G., Polimene, L., Bruggeman, J., Blackford, J., Ciavatta, S., Mitra, A., and Flynn, K.J. (2018) Modelling mixotrophic functional diversity and implications for ecosystem function. *Journal of Plankton Research*, **40**, 627–642.

Lepère, C., Boucher, D., Jardillier, L., Domaizon, I., and Debroas, D. (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (lake Pavin). *Applied and Environmental Microbiology*, **72**, 2971–2981.

Lepère, C., Domaizon, I., Humbert, J.-F., Jardillier, L., Hugoni, M., and Debroas, D. (2019) Diversity, spatial distribution and activity of fungi in freshwater ecosystems. *PeerJ*, **7**, e6247.

Levy Karin, E., Mirdita, M., and Söding, J. (2020) MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*, **8**, 48.

Li, Y., Hu, Y., Zhao, K., Pan, Y., Qu, Y., Zhao, J., and Qin, Y. (2019) The indispensable role of histone methyltransferase PoDot1 in extracellular glycoside hydrolase biosynthesis of *Penicillium oxalicum*. *Frontiers in Microbiology*, **10**, 2566.

Litchman, E. and Klausmeier, C.A. (2008) Trait-based community ecology of phytoplankton. *Annual Review of Ecology, Evolution and Systematic*, **39**, 615–639.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Maire, E., Grenouillet, G., Brosse, S., and Villéger, S. (2015) How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces: assessing functional space quality. *Global Ecology and Biogeography*, **24**, 728–740.

Mangot, J.-F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., and Debroas, D. (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes: short-term dynamics of small eukaryotes. *Environmental Microbiology*, **15**, 1745–1758.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j*, **17**, 10.

Medina, E.M. and Buchler, N.E. (2020) Chytrid fungi. *Current Biology*, **30**, R516–R520.

Mitra, A., Caron, D.A., Faure, E., Flynn, K.J., Leles, S.G., Hansen, P.J., et al. (2023) The Mixoplankton Database (MDB): diversity of photo-phago-trophic plankton in form, function, and distribution across the global ocean. *Journal of Eukaryotic Microbiology*, **70**, e12972.

Nagakannan, P., Iqbal, M.A., Yeung, A., Thliveris, J.A., Rastegar, M., Ghavami, S., and Eftekharpour, E. (2016) Perturbation of redox balance after thioredoxin reductase deficiency interrupts autophagy-lysosomal degradation pathway and enhances cell death in nutritionally stressed SH-SY5Y cells. *Free Radical Biology and Medicine*, **101**, 53–70.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020) vegan: Community Ecology Package. R package version 2.5–7.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., and Moran, M.A. (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology*, **11**, 1358–1375.

R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing.

Ramond, P., Siano, R., and Sourisseau, M. (2018) Functional traits of marine protists. *SEANOE*.

Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., et al. (2019) Coupling between taxonomic and functional diversity in protistan coastal communities: functional diversity of marine protists. *Environmental Microbiology*, **21**, 730–749.

Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35**, 1026–1028.

Trench-Fiol, S. and Fink, P. (2020) Metatranscriptomics from a small aquatic system: microeukaryotic community functions through the diurnal cycle. *Frontiers in Microbiology*, **11**, 1006.

Van den Wyngaert, S., Ganzert, L., Seto, K., Rojas-Jimenez, K., Agha, R., Berger, S.A., et al. (2022) Seasonality of parasitic and saprotrophic zoosporic fungi: linking sequence data to ecological traits. *The ISME Journal*, **16**, 2242–2254.

Van Laere, A., Van Schaftingen, E., and Hers, H.-G. (1983) Fructose 2,6-bisphosphate and germination of fungal spores. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 6601–6605.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., and Garnier, E. (2007) Let the concept of trait be functional! *Oikos*, **116**, 882–892.

Voronin, L.V. (2014) Terrigenous micromycetes in freshwater ecosystems (review). *Inland Water Biology*, **7**, 352–356.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied Environmental Microbiology*, **73**, 5261–5267.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, **5**, 27.

Wilken, S., Choi, C.J., and Worden, A.Z. (2020) Contrasting mixotrophic lifestyles reveal different ecological niches in two closely related marine protists. *Journal of Phycology*, **56**, 52–67.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*, **347**, 735–745.

# Supplementary Materials

**Supp. Mat. 1**

Clusters were characterized using modalities of each morpho-physio-phenological trait (Fig. S6). The functional groups have been named consequently: 1&5) Parasites (PARA): characterized by their feeding strategy, symbiosis type (the majority has been described as having a parasitic lifestyle) and organic covers or naked; 2) Saprotrophs (SAP): characterized by saprotrophic feeding strategy, attached lifestyle and mainly absence of biotic interaction reported in the literature; 3) Heavy-cover- and Swimmer-photoautotrophs (HCOV and SWAT): characterized by plastids presence and osmotrophic ingestion mode with either mineral cover (*e.g.*, siliceous) and non-swimming abilities or organic cover and swimming abilities; 4&7) Strict-heterotrophs (HET): characterized by phagotrophic feeding strategy and the absence of plastids; 6) Mixoplankton (MIXO): considered as photo-osmo-phago-mixotrophs according to Mitra *et al.* (2023) and characterized by the presence of chloroplast, motility and their feeding strategy (mainly phagotrophic); 8) Floater- and colonial-photoautotrophs (FLAT): characterized by non-swimming abilities, plastids presence and osmotrophic ingestion mode; 9) Endophyte (END): characterized by their attached life-style, feeding strategy, biotic interaction with plants and mostly organic covers.

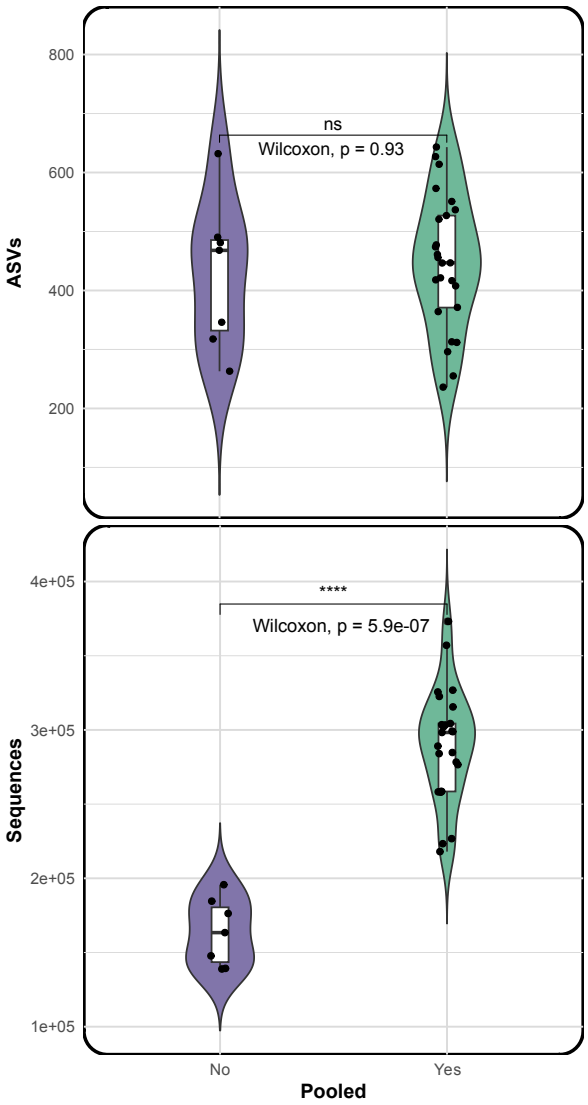



**Figure S1: Assessment of the effect of the pooling procedure before rarefaction steps.**

The effect on specific richness and sequence abundance of the pooling procedure were assessed using Wilcoxon-Mann-Whitney test by comparing datasets resulting from the duplicates grouping (on the left) and the others (without duplicates) (on the right).

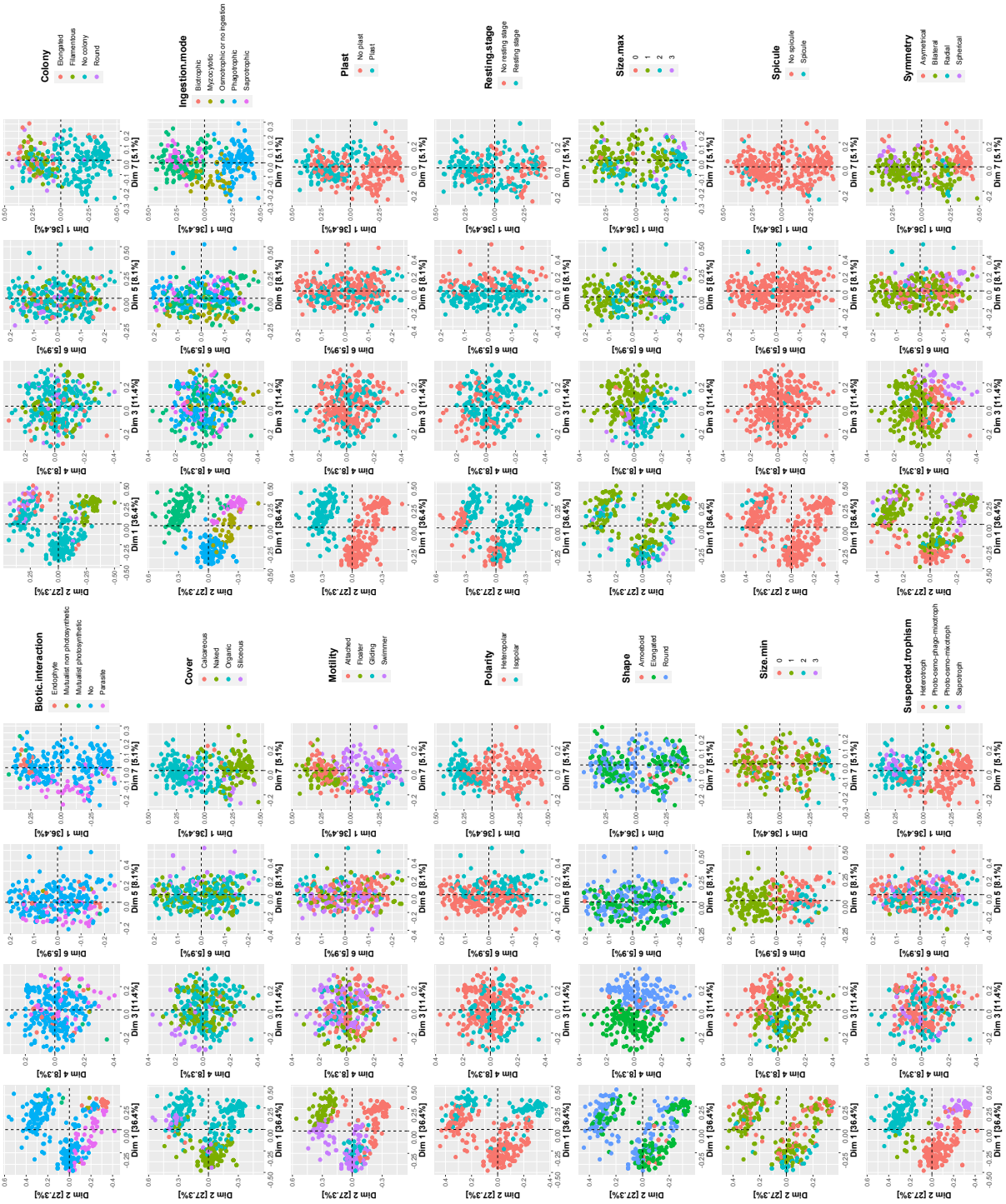



**Figure S2: Principal Coordinate Analysis (PCoA) using the Gower distance between all taxonomic references of the functional traits table.**

Different colors represent traits modalities. Seven Euclidean dimensions are used to ordinate all taxonomic references as described in Maire *et al.* (2015). Sizes (min and max) are expressed as log10 values.

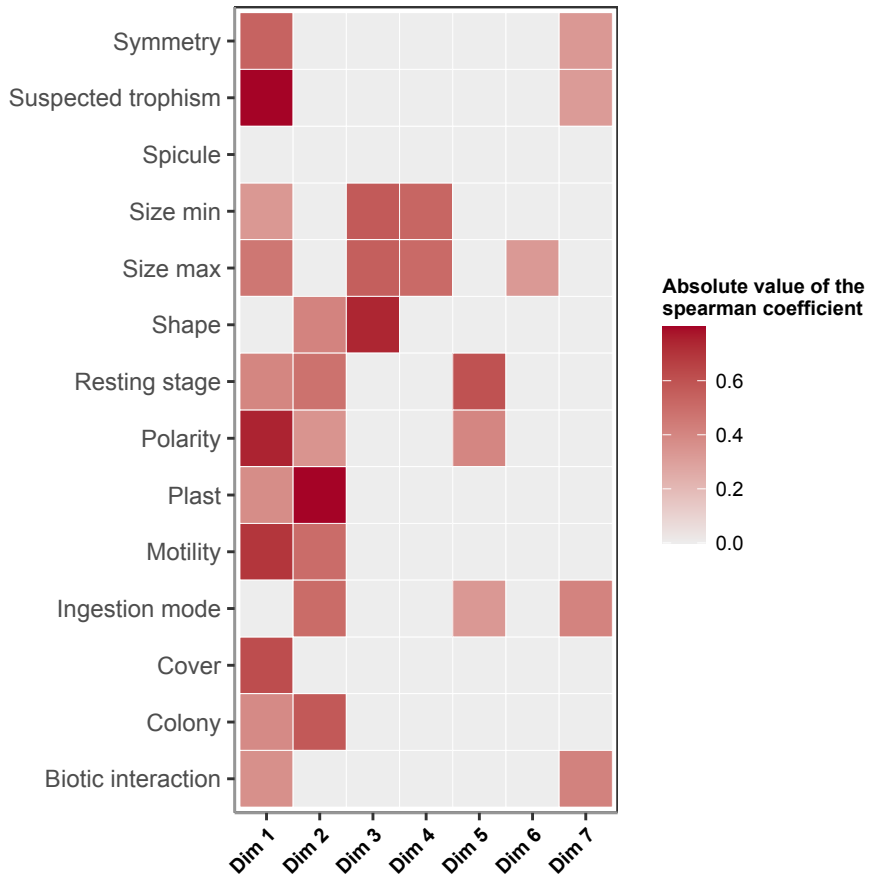



**Figure S3: Heatmap representing trade-off between traits using PCoA scores.**

Correlation of traits annotation and PCoA scores for each dimension highlights trade-offs. The correlation is represented using the absolute value of the Spearman's rank test coefficient. As described in Ramond *et al.* (2019), absolute values <0.3 and/or with a *p*-value >0.05 are not displayed. The trade-offs are visible on the two first dimensions of the PCoA, showing clear correlations between almost all traits. The other dimensions are only correlated with a maximum of 3 traits already described by the dimensions 1 and/or 2.

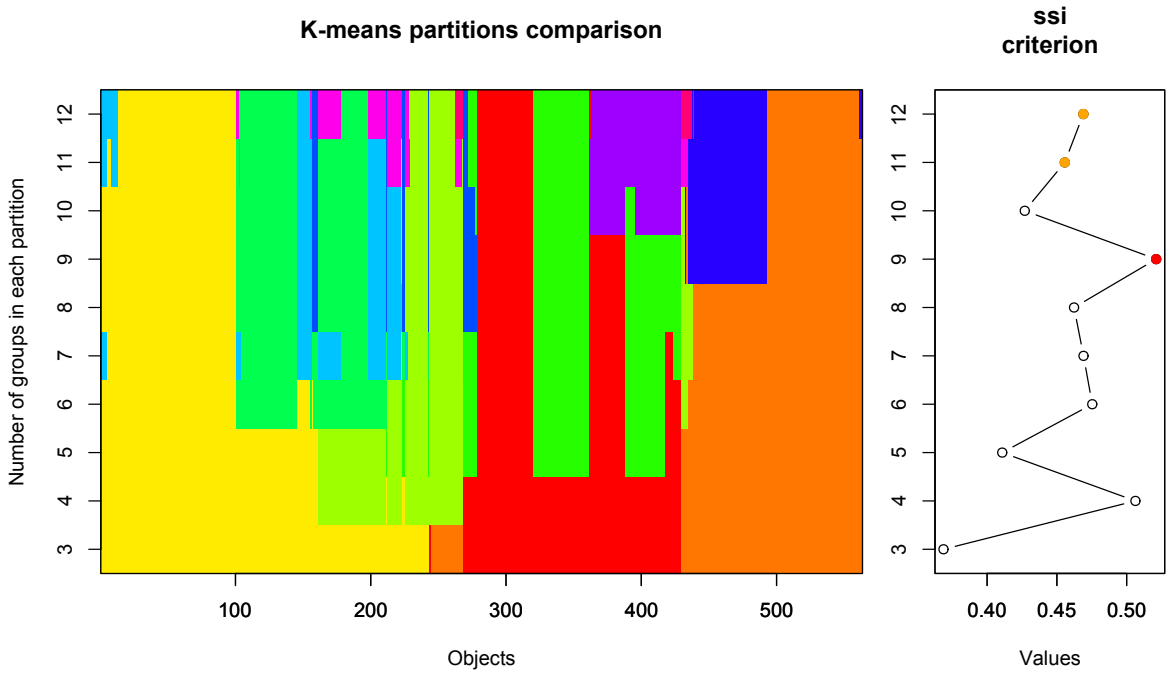



**Figure S4: Best partitioning result from the Simple Structure Index (SSI) based on the two first dimensions of the PCoA.**

The left-hand graph represents the distribution of the 563 taxonomic references (objects) within partitioning (y axis) of increasing number of groups. The right-hand graph displays results from the SSI criterion, the highest value indicated with a red dot shows the number of groups with the best partitioning of the functional traits. This graph is obtained using the *cascadeKM*() function provided in the R package vegan (Oksanen *et al.*, 2020).

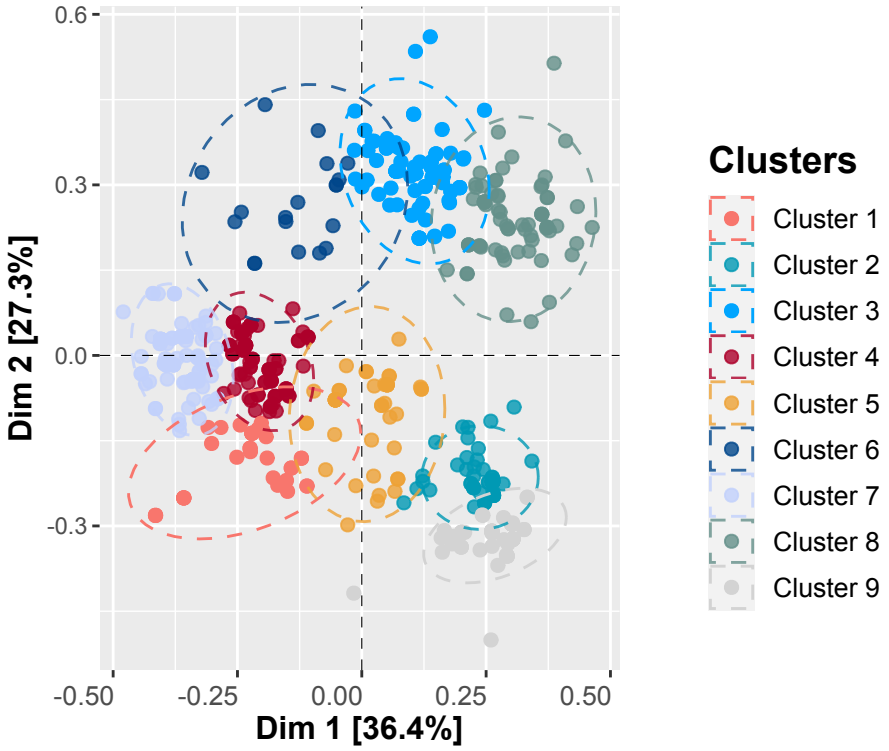



**Figure S5: Principal Coordinate Analysis (PCoA) using the Gower distance between all taxonomic references of the functional traits table (groups visualization after K-means clustering).**

Different colors represent functional groups. Seven Euclidean dimensions are used to ordinate all taxonomic references as described in Maire *et al.* (2015).

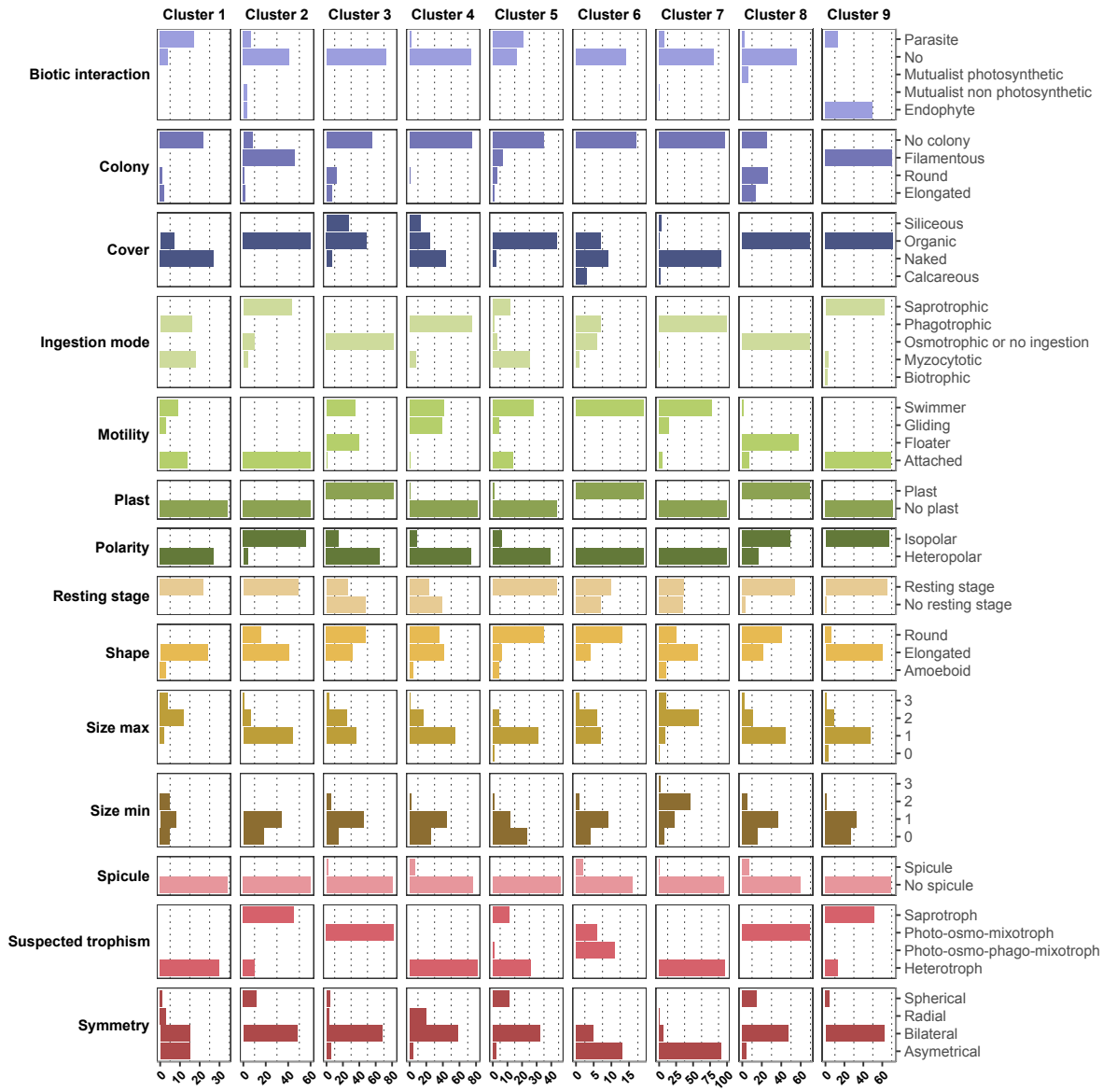



**Figure S6: Barplots representing traits modalities of each functional group.**

Functional groups are defined and characterized using dominant modalities of each trait. Sizes (min and max) are expressed as log10 values.

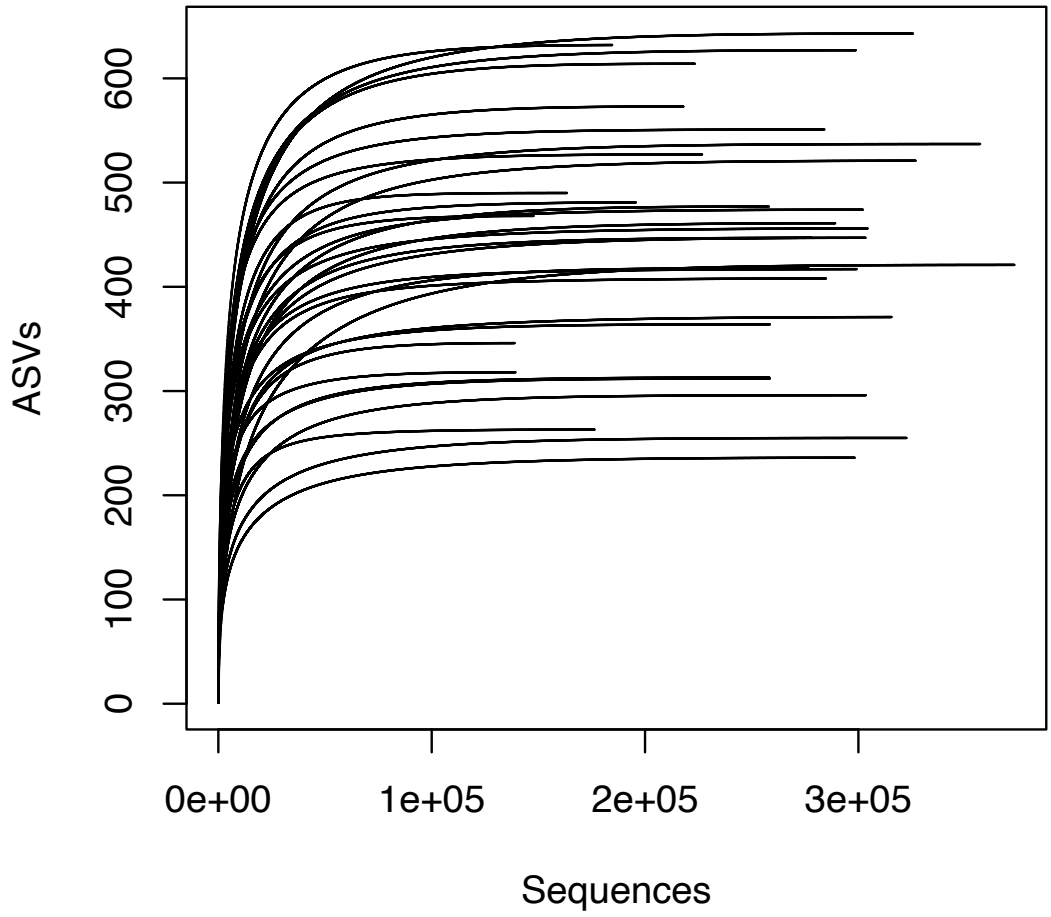



**Figure S7: Rarefaction curves.**

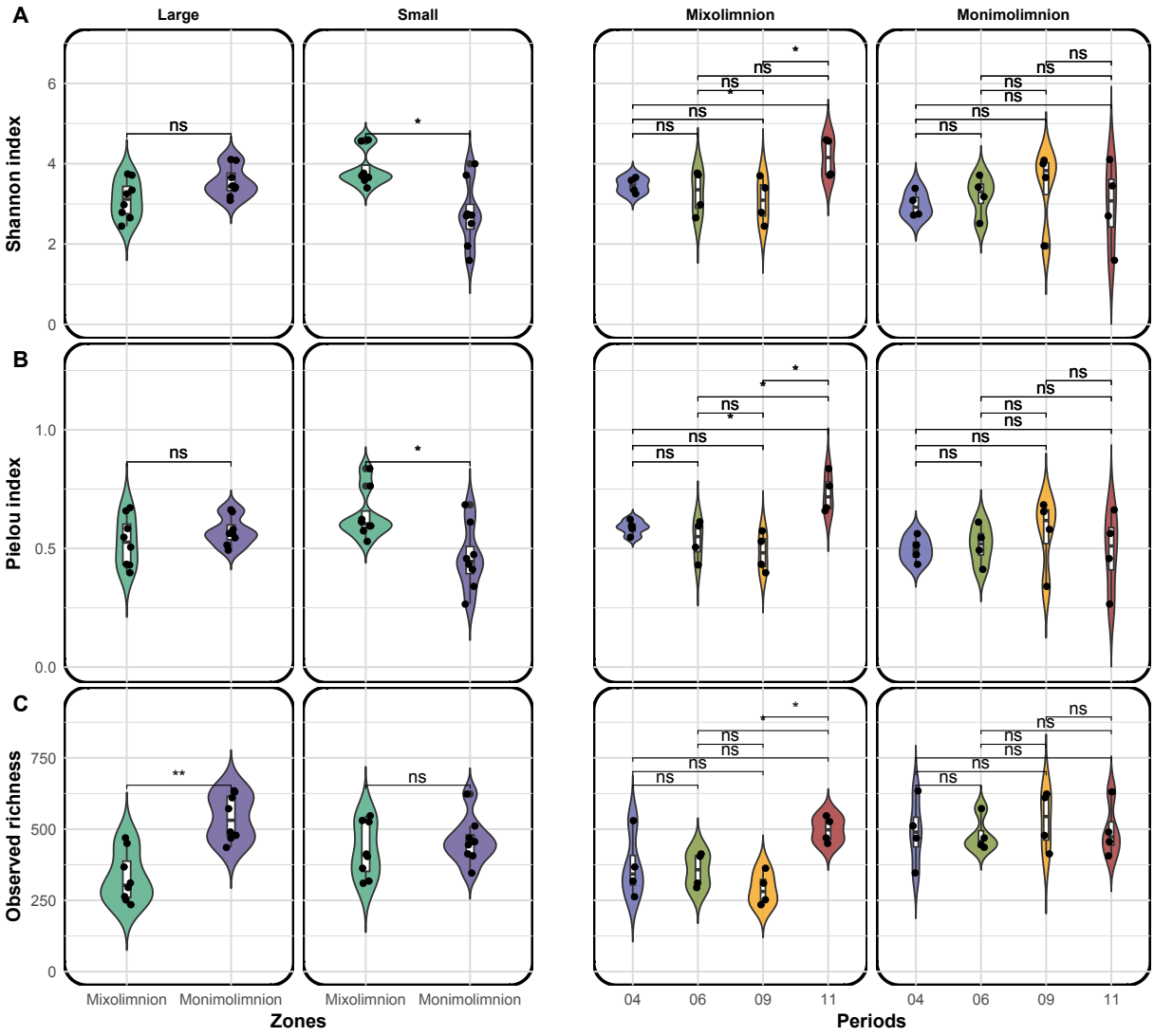



**Figure S8: Diversity and equitability index (Shannon (A), Pielou (B)) and observed richness (C) of amplicons across sampling zones, size fractions and periods.**

Wilcoxon-Mann-Whitney test is used to assessed differences between conditions.

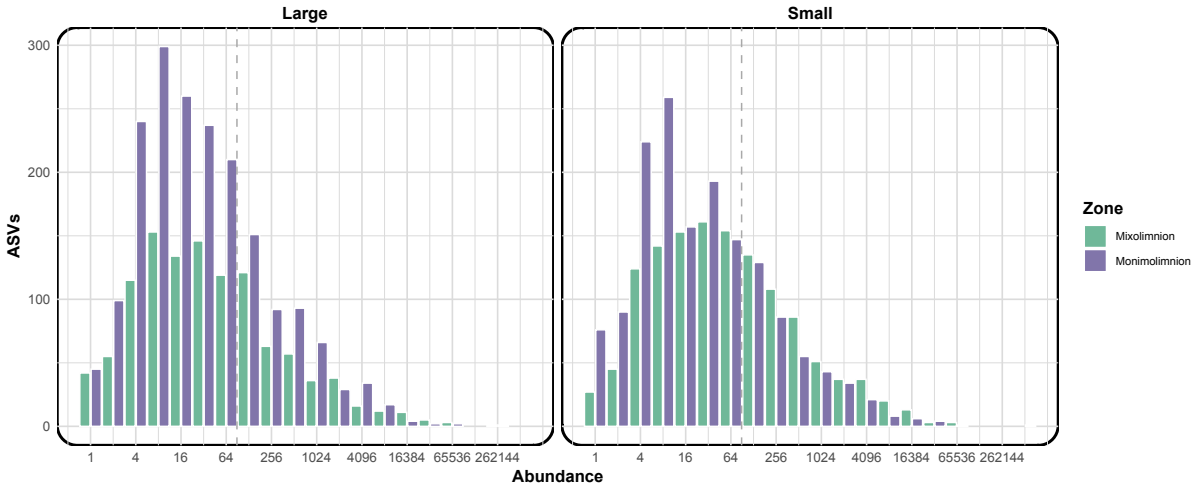



**Figure S9: Barplots representing the number of ASVs (y-axis) by sequences abundance (x-axis) across sampling zones and size fractions.**

Dashed lines represent delimitation between 0.002% (left) and 99.998% (right) of the total sequences' abundance.

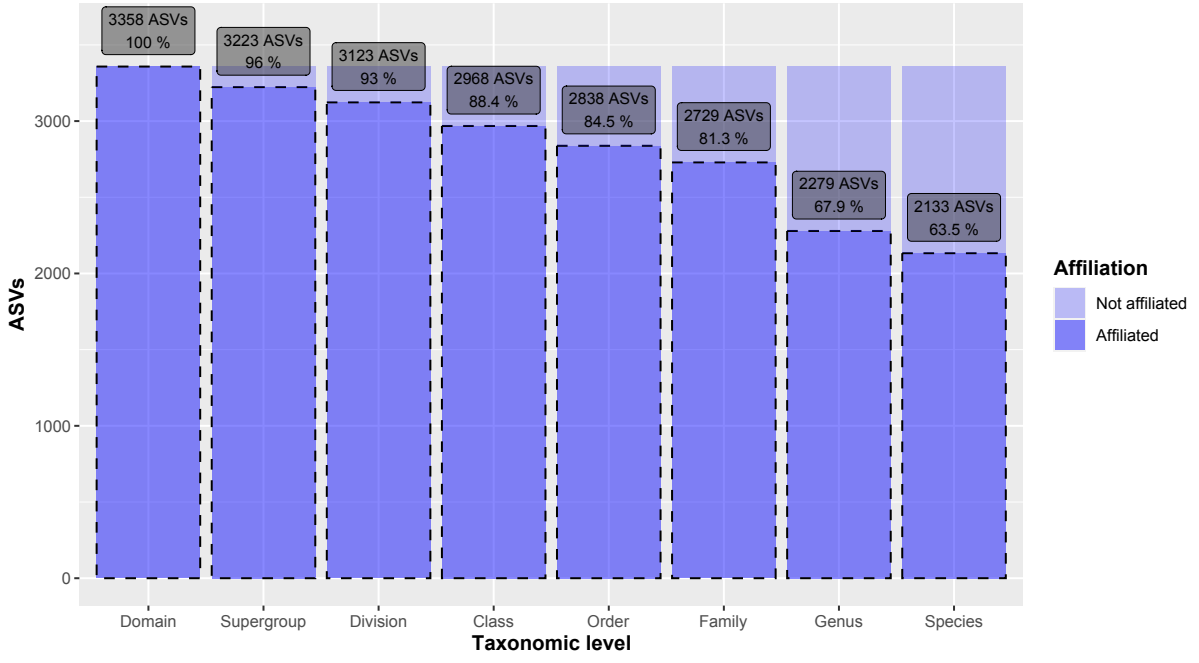



**Figure S10: Taxonomic affiliation statistics.**

The proportion of annotated ASVs is displayed for each taxonomic level (PR2 taxonomy is used to classify the entire eukaryotic diversity within 8 ranks (*i.e.*, Domain / Supergroup / Division / Class / Order / Family / Genus / Species)).

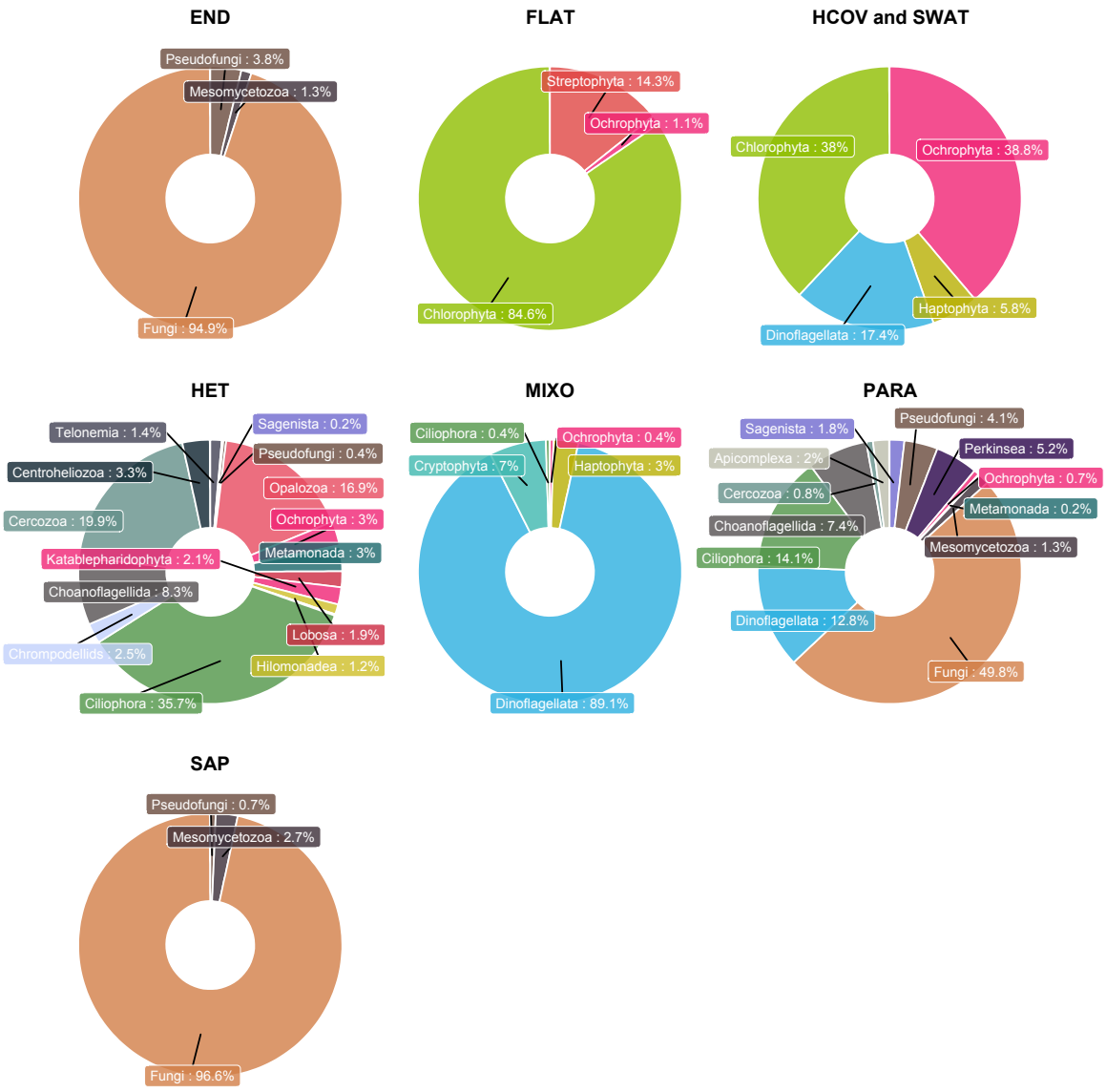



**Figure S11: ASVs richness of the eukaryotic divisions across each functional group.** Functional groups are named as described in Figure 5.

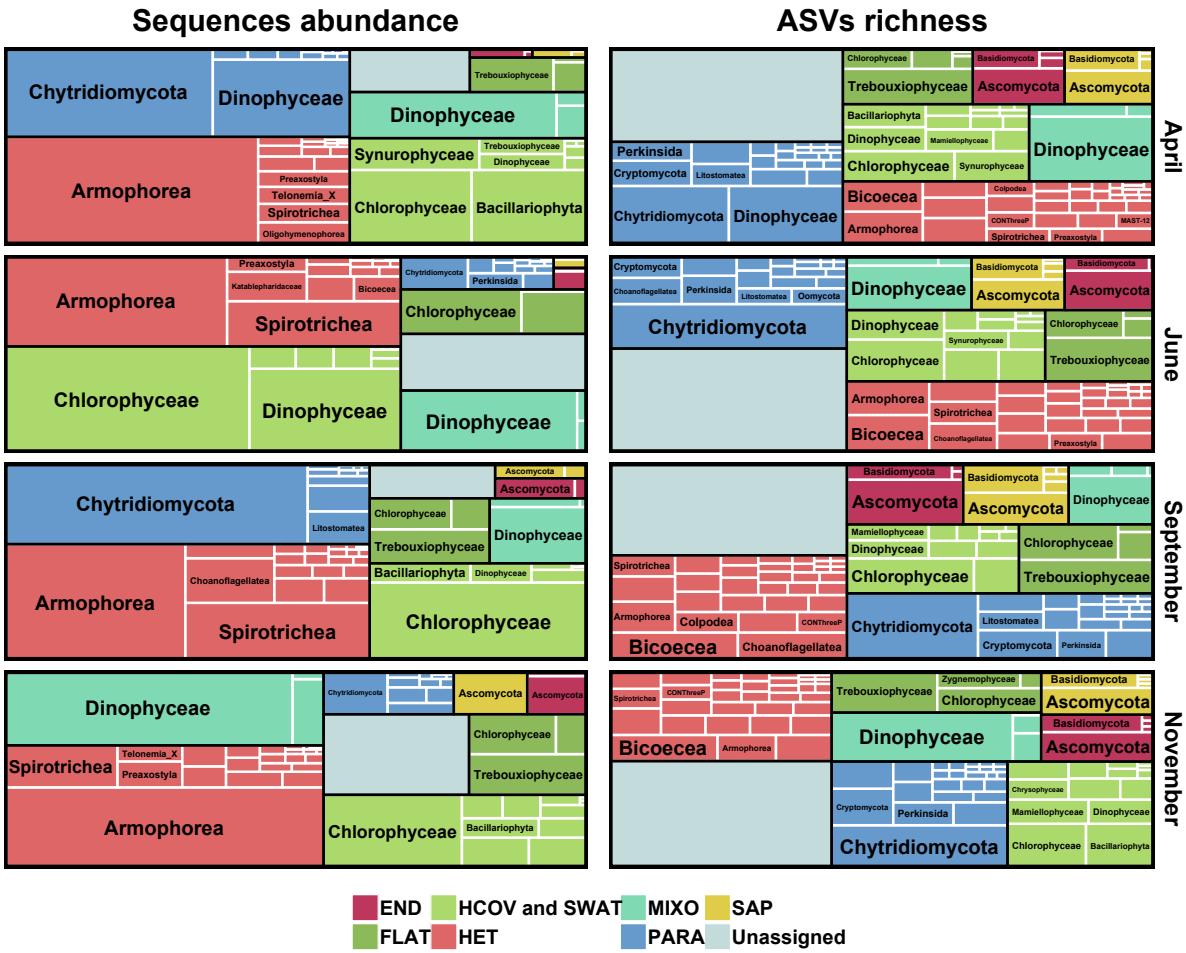



**Figure S12: Relative abundance and ASVs richness of the eukaryotic classes across each functional group according to the sampling period.**

Abundance and richness of eukaryotic classes (4th rank according to PR2 taxonomy) combined to those of functional groups is displayed for each sampling periods. Functional groups are named as described in Figure 5.

**Figure S13: Heatmap representing differential expression of genes across sampling zones for each eukaryotic class and functional group (accessible online).**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across zone is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are represented. Values of log2-FoldChange >4 are notified with « · », log2-FoldChange >5 with « ∗ » while log2-FoldChange >6 with « ∗∗ ».

**Figure S14: Heatmap representing differential expression of genes across size fractions for each eukaryotic class and functional group (accessible online).**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across size fraction is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are represented. Values of log2-FoldChange >4 are notified with « · », log2-FoldChange >5 with « ∗ » while log2-FoldChange >6 with « ∗∗ ».

**Figure S15: Heatmap representing differential expression of genes across periods for each eukaryotic class and functional group (accessible online).**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across periods is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are represented. Values of log2-FoldChange >4 are notified with « · », log2-FoldChange >5 with « ∗ » while log2-FoldChange >6 with « ∗∗ ».

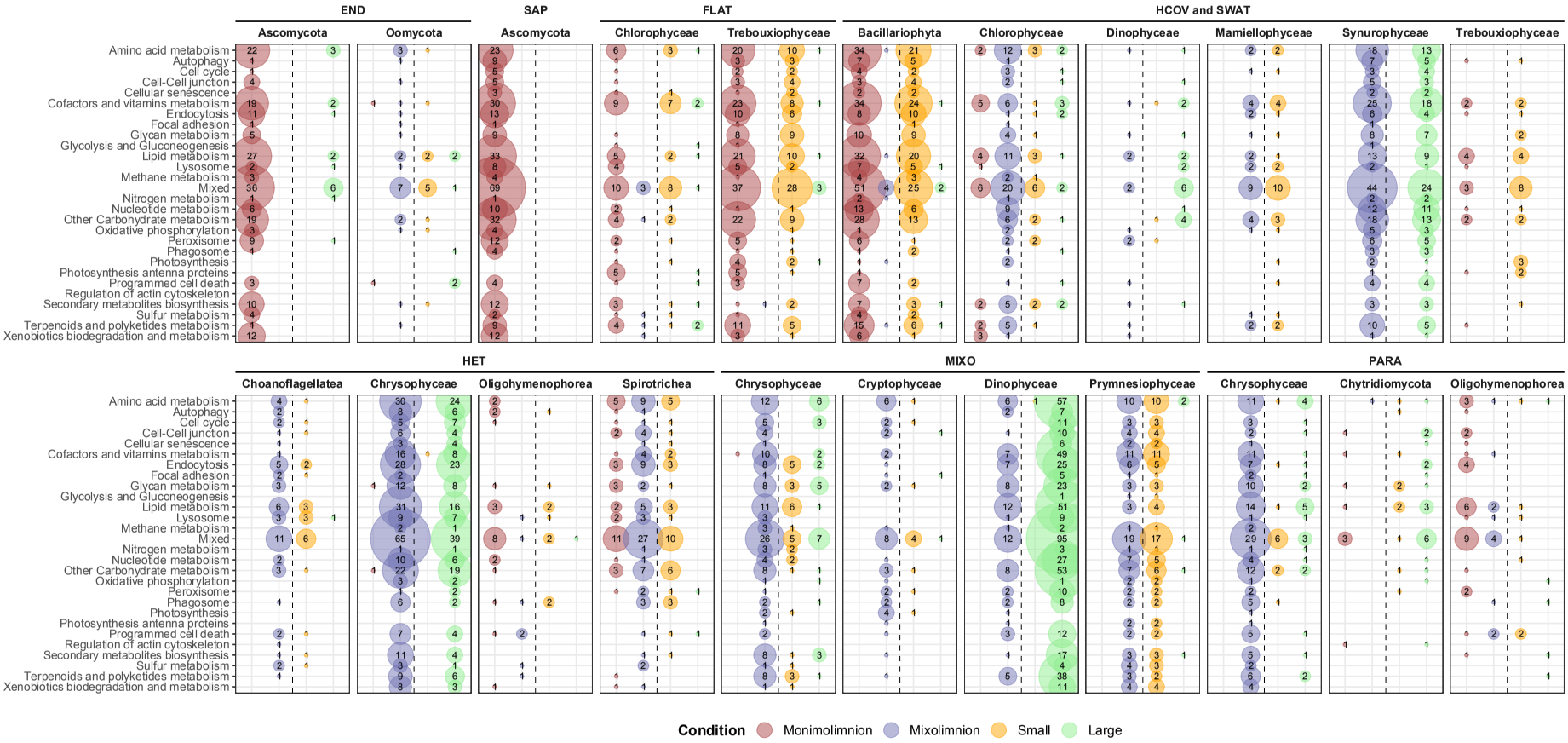



**Figure S16: Bubble plot representing the number of over-expressed genes involved in eukaryotic metabolisms and cellular processes across size fractions and sampling zones for each eukaryotic class and functional group.**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across size fraction and sampling zones is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are counted.

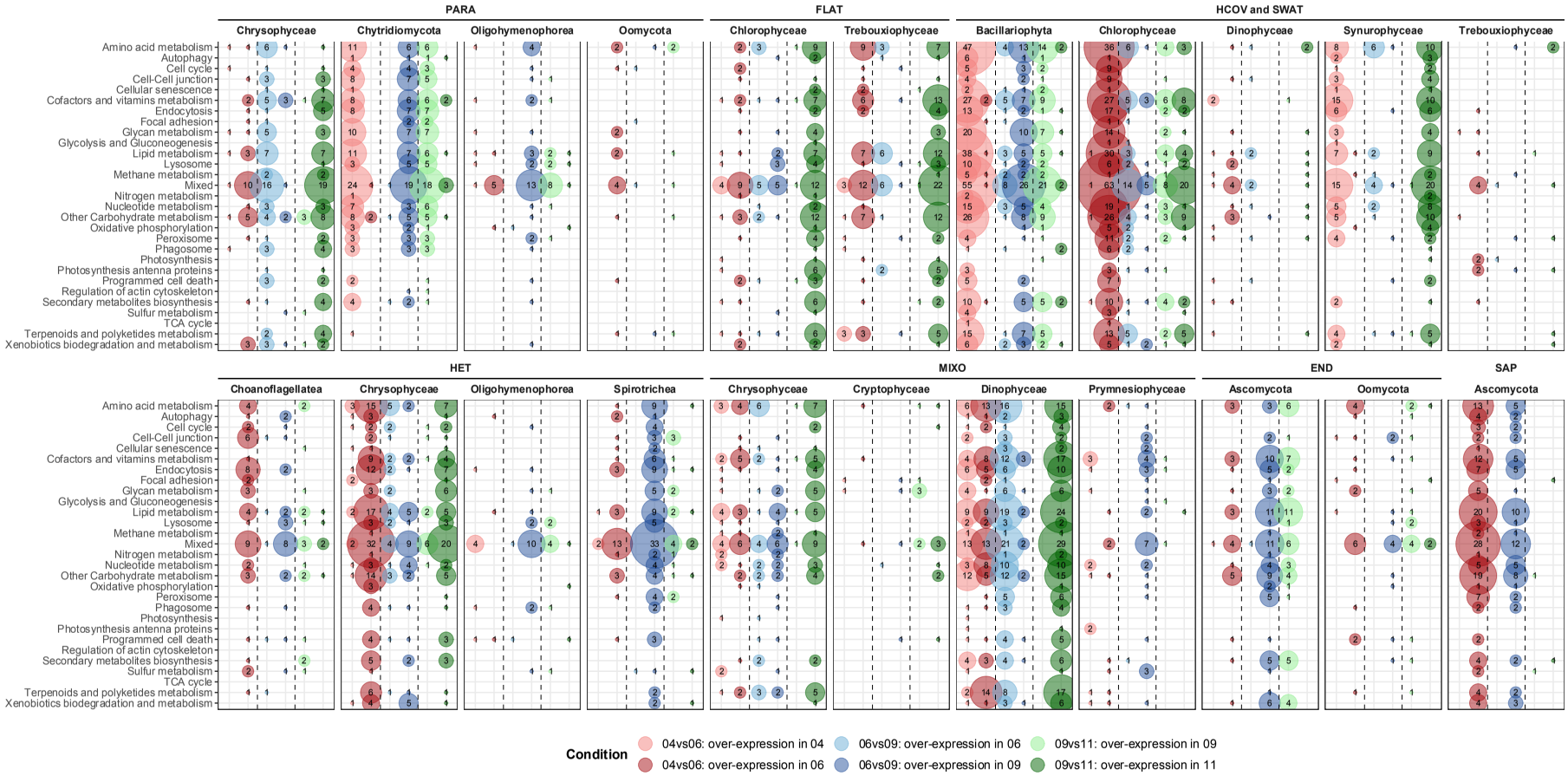



**Figure S17: Bubble plot representing the number of over-expressed genes involved in eukaryotic metabolisms and cellular processes across periods for each eukaryotic class and functional group.**

Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across periods is characterized by an adjusted *p*-value <0.05 and an absolute value of log2-FoldChange >2 are counted.

**Table S1: Traits table (accessible online).**

Table of 815 taxonomic references corresponding to the metabarcoding dataset associated to 14 morpho-physio-phenological traits. The traits selected for this analysis are the same as those used in Ramond *et al.* (2019) with the addition of the 'suspected trophism' trait. Multi- or pluri-cellular eukaryotes are not considered in the trait annotation process (*i.e.*, Metazoa, Embryophyceae, Lecanoromycetes and Lichinomycetes, and some Ascomycota and Basidiomycota).

**Table S2: KO selection corresponding to genes related to KEGG-BRITE 'cellular process' and/or 'metabolism' (accessible online).**

## References


Maire, E., Grenouillet, G., Brosse, S., and Villéger, S. (2015) How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces: Assessing functional space quality. *Global Ecology and Biogeography*, **24**, 728–740.

Mitra, A., Caron, D.A., Faure, E., Flynn, K.J., Leles, S.G., Hansen, P.J., et al. (2023) The Mixoplankton Database (MDB): Diversity of photo-phago-trophic plankton in form, function, and distribution across the global ocean. *Journal of Eukaryotic Microbiology*, **70**, e12972.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020) vegan: Community Ecology Package. R package version 2.5–7.

Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., et al. (2019) Coupling between taxonomic and functional diversity in protistan coastal communities: functional diversity of marine protists. *Environmental Microbiology*, **21**, 730–749