



HAL
open science

Les eucaryotes unicellulaires dans les écosystèmes lacustres : de la diversité fonctionnelle aux interactions hôte-parasites

Arthur Monjot

► **To cite this version:**

Arthur Monjot. Les eucaryotes unicellulaires dans les écosystèmes lacustres : de la diversité fonctionnelle aux interactions hôte-parasites. Microbiologie et Parasitologie. Université Clermont Auvergne, 2023. Français. NNT : 2023UCFA0109 . tel-04573892

HAL Id: tel-04573892

<https://theses.hal.science/tel-04573892>

Submitted on 13 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Doctorale n°65 : Science de la Vie, Santé, Agronomie, Environnement

THÈSE

Pour obtenir le grade de
Docteur de l'Université Clermont Auvergne
Spécialité Microbiologie (Écologie microbienne)

Présentée et soutenue publiquement par
Arthur Monjot

Les eucaryotes unicellulaires dans les écosystèmes lacustres : de la diversité fonctionnelle aux interactions hôte-parasites

Devant le jury composé de :

Isabelle Domaizon	Rapporteuse	DR, UMR CARRTEL
Aurélie Chambouvet	Rapporteuse	CR, Station biologique de Roscoff
Éric Pelletier	Examineur	DR, Genoscope - CEA
Jean-François Briand	Examineur	MCF, Université de Toulon
Jean-François Carrias	Président du jury	PR, Université Clermont Auvergne
Cécile Lepère	Directrice de thèse	PR, Université Clermont Auvergne

Laboratoire Microorganismes : Génome et Environnement
UMR CNRS 6023, F-63000, Clermont-Ferrand

Soutenue le 19 décembre 2023

Remerciements

Je tiens tout d'abord à remercier Aurélie Chambouvet et Isabelle Domaizon d'avoir accepté d'être rapportrices de ce manuscrit. Merci également à Jean-François Briand, Jean-François Carrias et Éric Pelletier de faire partie de ce jury. Mes remerciements s'adressent aussi à Fabrice Not et Frédéric Delbac qui, avec Éric Pelletier, m'ont apporté de précieux conseils durant mes comités de thèse.

Cécile, je ne te remercierai jamais suffisamment pour tout ce que tu m'as appris et pour le temps que tu m'as consacré au cours de ces trois années (et demie). Tu m'as fait confiance, poussé jusqu'à mes limites et même si cela a été difficile parfois, c'était nécessaire. Si je peux être fier de mon parcours, c'est grâce à toi. Merci !

Mes remerciements s'adressent bien évidemment à ma Maman, mon Papa et Clément. Vous m'avez supporté toutes ces années et apporté tellement. Merci Maman pour ta bienveillance, les nombreux conseils, ta tendresse (et ta cuisine !!). Merci Papa (et la glycine) de m'avoir écouté avec tant d'attention ... ne soit pas surpris, mon titre de thèse a un peu changé depuis le moment où je t'ai dit qu'elle porterait sur l'effet du cycle nyctéméral sur la sécrétion de somatropine lors de la période de reproduction de l'argiope fasciée en milieu anthropisé. Merci Clément d'être le meilleur des petits frères, tu me connais par cœur et j'aime tellement nos délires. Mille fois merci.

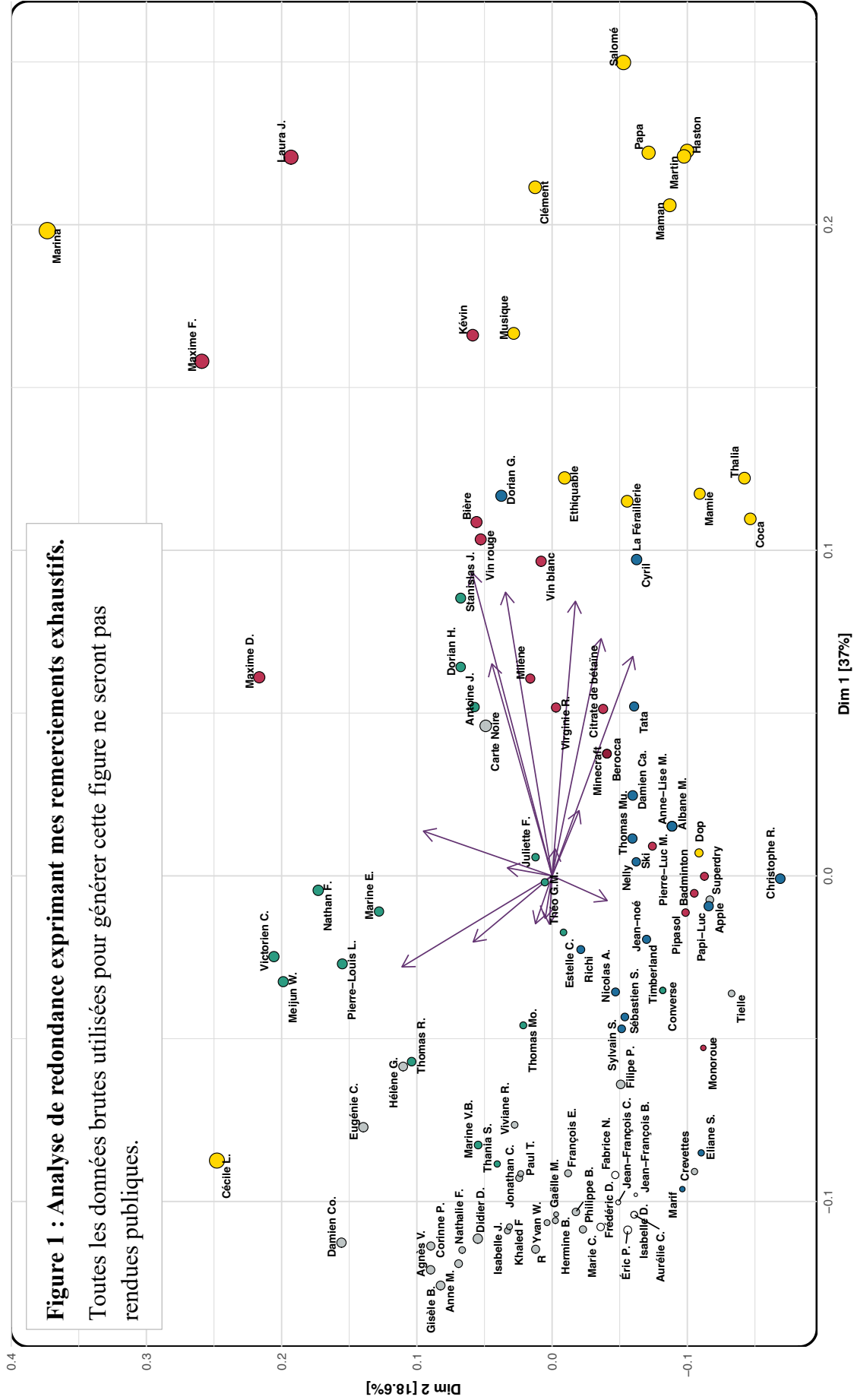
Je remercie également Marina, Maxime et Laura. Vous m'avez accueilli dans votre petite famille et m'avez apporté tellement de bonheur et de réconfort. Il y a peu de gens comme vous ... aussi bienveillants, fondamentalement gentils et avec qui je me sens si bien. Merci d'être là, je vous aime.

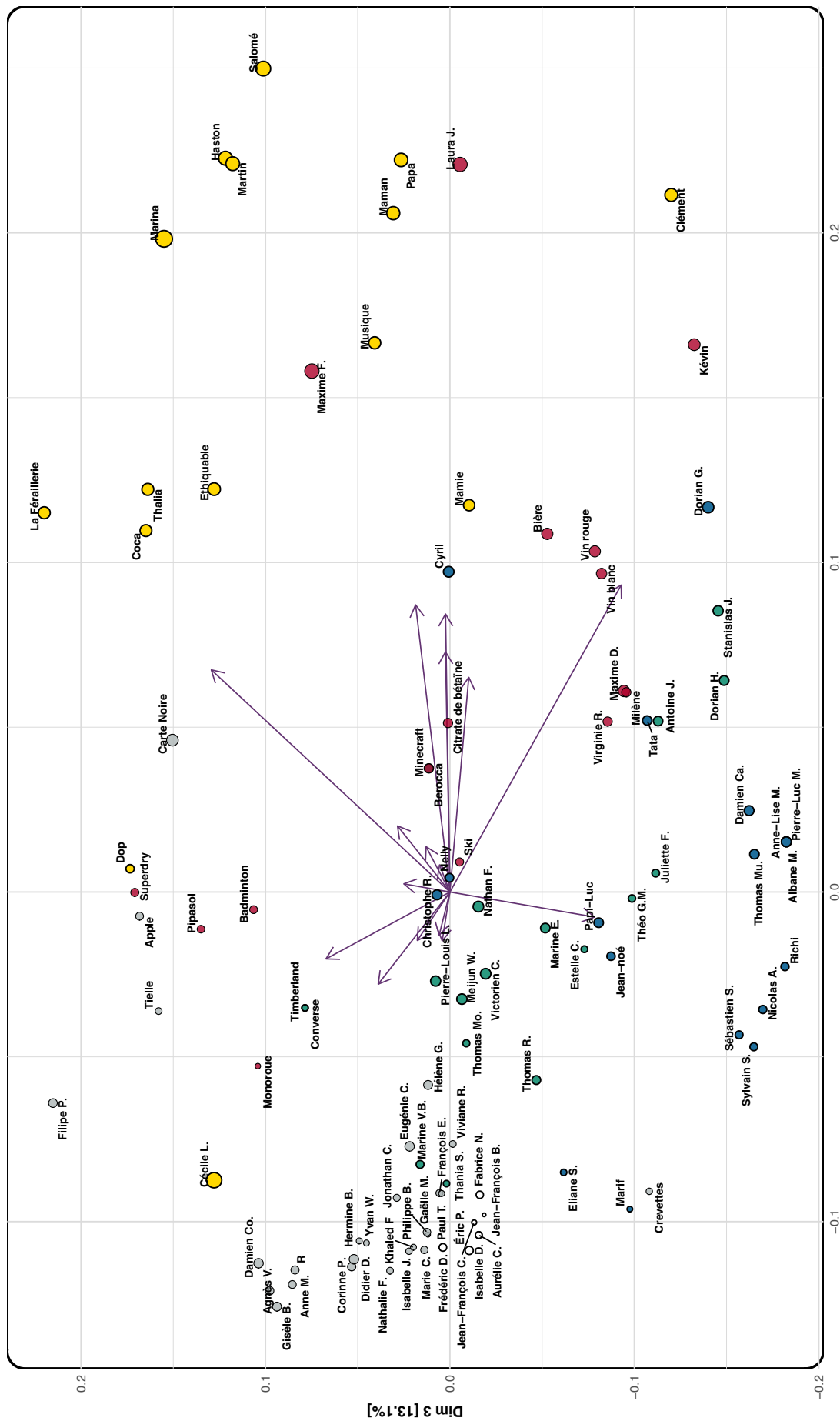
Enfin, merci ma Salamèche de m'avoir soutenu toutes ces années, d'avoir supporté tous mes caprices, mes tares, mes excès (de colère et d'amour) ... je sais que je ne suis pas facile. Je te remercie d'être là ... encore et toujours ... ☺. Je t'aime.

Évidemment merci à tous, qui avez contribué à ce travail et à rendre ces trois années tellement agréables : Mamie, Kévin, Maxime, Milène, Nelly, Cyril, les Tata, Tonton, Corinne. J'aurais pu m'étendre et « couler » sur plusieurs pages encore mais j'ai choisi quelque chose d'un peu plus ... original disons (**Fig. 1**).

Figure 1 : Analyse de redondance exprimant mes remerciements exhaustifs.

Toutes les données brutes utilisées pour générer cette figure ne seront pas rendues publiques.





Dim 1 [37%]

Dim 3 [13.1%]

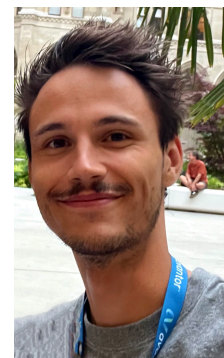
Arthur Monjot

Doctorant - Novembre 2023

Laboratoire Microorganismes : Génome et Environnement
(LMGE - UMR 6023, Aubière)

Arthur.Monjot.pro@gmail.com

+33 (0)607023779



Expériences de recherche

- Depuis 2020** **Doctorat**, Laboratoire Microorganismes : Génome et Environnement, UMR 6023 – Encadrement : C. Lepère
Sujet : Les eucaryotes unicellulaires dans les écosystèmes lacustres : de la diversité fonctionnelle aux interactions hôte-parasites
Description : Étude de la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires et des interactions hôte-parasites dans les écosystèmes lacustres par des approches -omiques (métabarcoding, métatranscriptomique et réseaux de similarité de séquences) et de culture (microscopie et microcosmes)
- 2020** **Stage Master 2**, Laboratoire Microorganismes : Génome et Environnement, UMR 6023 - Encadrement : C. Lepère - 28 semaines
Sujet : Diversité et rôle écologique des eucaryotes microbiens parasites dans les écosystème lacustres
Description : Étude de la diversité taxonomique des parasites lacustres et recherche de nouveaux hôtes par mise en place de microcosmes et analyse de données de métabarcoding
- 2019** **Stage Master 1**, Centre Armand Frappier de l'Institut National de la Recherche pour la Santé (INRS) - Encadrement : F. Veyrier - 12 semaines
Sujet : Identification des facteurs d'interaction entre *Neisseria meningitidis* et d'autres commensaux du nasopharynx
Description : Étude du pouvoir antibiotique de certaines souches de *Neisseria* par une approche de génomique comparative et de co-culture
- 2018** **Stage d'observation**, Laboratoire Microorganismes : Génome et Environnement, UMR 6023 - Encadrement : H. El Alaoui - 2 semaines
Sujet : Décryptage des mécanismes d'adhésion des microsporidies

Parcours Universitaire

- 2019 - 2020** **Master 2**, mention Microbiologie, parcours : Génome, Écologie et Physiologie Microbienne - UCA Clermont-Ferrand - Mention Bien (15,01/20), classement 2/18
- 2018 - 2019** **Master 1**, mention Microbiologie, parcours : Génome, Écologie et Physiologie Microbienne - UCA Clermont-Ferrand - Mention Bien (13,30/20), classement 3/30
- 2015 - 2018** **Licence**, mention Sciences de la Vie, parcours : Biologie Cellulaire et Physiologie - UBP Clermont-Ferrand - Mention Assez Bien (12,34/20 sur la L3)

Compétences

- Informatique**
- Bureautique** : Suite Microsoft et Apple
 - Edition d'image** : Photoshop, Illustrator, Fiji
 - Programmation** : python, bash, awk, slurm, git
 - Système** : Linux (toutes distributions), Mac OS, Windows
 - Modélisation 3D** : Fusion 360
 - Acquisition d'image** : Zeiss Zen
 - Analyses de données** : R
 - Autres** : Administration serveur et réseau
- Techniques**
- Microbiologie** : microscopie optique et en épifluorescence, FISH, TSA-FISH, micromanipulation/isolation cellule unique, culture de microorganismes (eucaryotes et procaryotes (en laboratoire P2)), marquages multiples, préparation de filtres pour l'étude du plancton, ultra-filtration à flux tangentiel
 - Biologie moléculaire** : PCR, extractions ADN, électrophorèse, WGA, clonage, RFLP
 - Bioinformatique** : traitement de données issues de métabarcoding (OTU et ASV), de métatranscriptomique (DESeq2), analyse de diversité microbienne (α et β -diversité), développement de chaînes de traitement et d'outils d'analyse, réseaux de similarité de séquences
 - Sorties terrain** : échantillonnage en lac et en rivière, maîtrise du matériel d'échantillonnage (bouteille fermante, intégratrice, sondes, disque de secchi, etc.)
- Langues**
- Français** : langue maternelle — **Anglais** : niveau B1 — **Espagnol** : niveau A2

Communications scientifiques

Articles

Monjot A, Bittner L, Lepère C. Metatranscriptomes-based sequence similarity networks uncover genetic signature within parasites and other functional groups of freshwater microbial eukaryotes. *En préparation*.

Bouazizi S, Montevecchi G, Masino F, Biderre-Petit C, **Monjot A**, Antonelli A, Hamdi M : Alcoholic beverages from prickly pear juice. Dynamics and kinetics of fermentation processes and chemical characterization of finished products. *En préparation*.

Biderre-Petit C, Courtine D, Hennequin C, Galand PE, Bertilsson S, Debroas D, **Monjot A**, Lepère C, Divne AM, Hochart C : A pan-genomic approach reveals novel Sulfurimonas clade in the ferruginous meromictic Lake Pavin. En révision. *Molecular Ecology Resources*

Monjot A, Bronner G, Courtine D, Cruaud C, Da Silva C, Aury JM, Moné A, Vellet A, Wawrzyniak I, Colombet J, Billard H, Debroas D, Lepère C : Functional diversity of microbial eukaryotes in a meromictic lake ecosystem: coupling between metatranscriptomic and a trait-based approach. 2023. *Environmental Microbiology*. <https://doi.org/10.1111/1462-2920.16531>

Chauvet M, **Monjot A**, Lepère C : High diversity of microsporidian parasites and new planktonic hosts in freshwater and marine ecosystems. 2023. *Limnology and Oceanography*. <https://doi.org/10.1002/lno.12321>

Monjot A & Chauvet M, Moné A, Lepère C : Single cell analysis reveals a new Microsporidia-host association in a freshwater lake. 2022. *Microbial Ecology*. <https://doi.org/10.1007/s00248-022-02023-7>

Biderre-Petit C, Charvy J, Bronner G, Chauvet M, Debroas D, Gardon H, Hennequin C, Jouan-Dufournel I, Moné A, **Monjot A**, Ravet V, Vellet A, Lepère C : FreshOmics: a manually curated and standardized -omics database for investigating freshwater microbiomes. 2022. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.13692>

Com. Orales

Monjot A et al. : Functional diversity of microbial eukaryotes in a meromictic lake ecosystem: coupling between metatranscriptomic and a trait-based approach. 2023. *IX European Congress of Protistology & Annual Congress of the International Society of Protistologists (ECOP-ISOP) : The Century of Protists, Vienne, AUTRICHE*.

Monjot A et al. : Métabarcoding et étude de traits, vers une description de la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires dans le lac Pavin. 2022. *Journée thématique de l'Association Francophone d'Écologie Microbienne (AFEM) : Les microeucaryotes : le chaînon manquant des études en écologie microbienne, Clermont Ferrand, FRANCE*. Prix de la 2ème meilleure communication

Monjot A et al. : Coupler NGS et bibliographie, une méthode pour étudier la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires dans le lac Pavin. 2022. *18ème rencontre des microbiologistes du pôle clermontois, en ligne*.

Monjot A & Lepère C : Étude de traits et apprentissage non supervisé, concepts pour l'étude de la diversité fonctionnelle des eucaryotes unicellulaires dans les écosystèmes aquatiques. 2021. *9ème Journée Des Doctorants du LMGE, Clermont Ferrand, FRANCE*.

Posters

Monjot A et al. : Functional diversity of microbial eukaryotes in a meromictic lake ecosystem: coupling between metatranscriptomic and a trait-based approach. 2023. *Journée de l'École Doctorale SVSAE, Clermont Ferrand, FRANCE*.

Monjot A et al. : Métabarcoding et étude de traits, vers une description de la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires dans le lac Pavin. 2022. *Semaine écologie environnement biodiversité du CNRS, Clermont Ferrand, FRANCE*.

Monjot A et al. : Métabarcoding et étude de traits, vers une description de la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires dans le lac Pavin. 2021. *10ème colloque de l'Association Francophone d'Écologie Microbienne (AFEM), en ligne*.

Formations

- 2022 **Training School** : « Identification and detection methods for aquatic zoosporic parasites » (Germany) dans le cadre du COST PARAQUA
- 2021 **Formation au Mésocentre de Clermont Ferrand** : Utilisation d'un cluster de calcul
Formations en ligne (MOOC) : « Traitement de données NGS » ; « Introduction à la statistique avec R » ; « Intégrité scientifique dans les métiers de la recherche »
Formation du Master de bioinformatique de Clermont-Ferrand : Programmation en Python
Formation en langue anglaise : Dispensée par l'école doctorale Sciences Pour l'Ingénieur à Clermont-Ferrand
- 2020 **Formation en bioinformatique** : Initiation à la ligne de commande (en interne)

Enseignements

- 2020 - 2023
- 54 heures - Unité d'Enseignement : Animal et milieu de vie**
Travaux Pratiques en 3ème année de **Licence** Sciences de la Vie
Correction de comptes rendus et préparation du matériel
 - 18 heures - Unité d'Enseignement : Apparition et diversification de la vie**
Travaux Pratiques en 1ère année de **Licence** Sciences de la Vie
Correction de comptes rendus, préparation du matériel et entretien des cultures de microorganismes
 - 21 heures - Unité d'Enseignement : Génomique des communautés microbiennes et bio-analyse**
Travaux Dirigés et Pratiques en 1ère année de **Master** de Microbiologie et de Master de Bioinformatique
Création du support de cours et du plan de séance - Élaboration du pipeline d'analyse de donnée de métabarcoding étudié et reproduit en cours - Participation au jury d'oraux pour l'évaluation terminale
 - 8 heures - Unité d'Enseignement : Découverte de l'Océanographie**
Participation au jury d'oraux pour l'évaluation terminale - 3ème année de **Licence** Sciences de la Vie

Encadrement de stage

- 2021
- Encadrement d'un stage de Master 1**
Master mention Microbiologie, parcours : Génome, Écologie et Physiologie Microbienne
Sujet : Étude du parasitisme et de son rôle dans le réseau trophique lacustre

Divers

- Depuis 2020
- Co-crédation et implication dans l'**Organisation des Jeunes Chercheurs en Ecologie** (OJCE) - Association loi 1901
 - Membre élu représentant des doctorants/post-doctorants au **conseil de laboratoire**
 - Membre du **comité d'animation scientifique et de communication au laboratoire** - Organisation des animations scientifiques
- 2022
- Président de session en congrès** : Journée thématique de l'Association Francophone d'Écologie Microbienne (AFEM) : *Les microeucaryotes : le chaînon manquant des études en écologie microbienne*
- 2022
- Organisation de congrès** : Journée thématique de l'Association Francophone d'Écologie Microbienne (AFEM) : *Les microeucaryotes : le chaînon manquant des études en écologie microbienne*
- 2021
- Organisation d'un atelier et encadrement de collégiens à la **journée de la science** - « Découverte du réseau trophique aquatique »

Résumé

Au cours des dernières décennies, notre compréhension de la diversité microbienne dans l'environnement a beaucoup progressé notamment avec l'avènement des méthodes de séquençage de nouvelle génération et les approches « -omics ». Leur application a notamment permis de s'affranchir des approches de mise en culture ne contribuant qu'à une évaluation partielle et orientée de la diversité microbienne. Le métabarcoding, basé sur l'étude d'un marqueur unique et ubiquiste, est la méthode la plus utilisée pour l'analyse de la diversité car elle permet une évaluation assez exhaustive des espèces présentes dans un écosystème. En milieu aquatique, elle a en effet permis la mise en évidence d'une diversité considérable et insoupçonnée d'eucaryotes unicellulaires. Toutefois cette approche, essentiellement descriptive, ne permet pas de déterminer la physiologie ni de comprendre le rôle de ces microorganismes dans les écosystèmes. D'autres méthodes, telles que la métatranscriptomique, offrent la possibilité d'étudier leur potentiel métabolique et d'en évaluer les variations en fonction de paramètres environnementaux. Néanmoins, ces approches de séquençage haut débit conduisent à la production d'une quantité colossale de séquences environnementales dont la majeure partie reste inconnue. Afin d'étudier le lien diversité-fonction au sein des eucaryotes unicellulaires et ainsi mieux comprendre leur rôle dans les réseaux trophiques lacustres, largement sous-étudiés en comparaison aux écosystèmes marins, plusieurs approches ont été utilisées.

Du métabarcoding couplé à une étude de traits morpho-physio-phénotypiques, de la métatranscriptomique ainsi qu'une méthodologie basée sur l'isolement et la caractérisation de couples hôte-parasites (séquençage et hybridation *in situ*), ont été réalisés à partir d'échantillons lacustres (Pavin, méromictique ; Aydat, dimictique). Ces analyses ont révélé l'importante diversité des photo-osmo-phago-mixotrophes et des parasites tout en mettant en évidence les fortes variations saisonnières qu'ils subissent dans le mixolimnion du lac Pavin. Il semblerait, par exemple, que les périodes de brassage profitant aux communautés hôtes photosynthétiques favorisent le développement et la dissémination de champignons parasites notamment par la surexpression de gènes impliqués dans la phototaxie des zoospores et dans les métabolismes lipidiques. Parmi ces champignons parasites, les microsporidies sont des acteurs nouvellement identifiés dans les réseaux trophiques aquatiques. Nous avons en effet découvert avec une forte prévalence (42,5%) dans le lac d'Aydat, une association hôte-parasites entre une potentielle nouvelle espèce de microsporidie et une espèce de rotifère.

Une importante biosphère rare a également été mise en évidence dans le monimolimnion anoxique du lac Pavin, caractérisée par de nombreux saprotrophes surexprimant des gènes liés aux métabolismes du soufre, du nitrate et de dégradation de la matière organique. Les métabolismes caractéristiques d'organismes de différents modes trophiques ont également été étudiés en réalisant une étude basée sur la construction de réseaux de similarité de séquences protéiques. Tout en caractérisant pour la première fois une majorité de séquences inconnues (>40%), nous avons révélé la proximité génétique de protéines entre microorganismes hétérotrophes et photo-osmo-phago-mixotrophes et entre saprotrophes et parasites, ainsi qu'une redondance fonctionnelle relative des métabolismes primaires. En revanche, nous avons identifié près d'un million de protéines caractéristiques d'un seul groupe fonctionnel qui, pour certaines, représentent de réelles perspectives d'étude des voies métaboliques impliquées dans les interactions hôte-parasites.

Abstract

Over the last few decades, our understanding of microbial diversity in the environment has advanced considerably, particularly with the advent of next-generation sequencing methods and -omics approaches. These methods have allowed for a more comprehensive evaluation of microbial diversity compared to traditional culture-based approaches. The most commonly used method for analyzing diversity is metabarcoding, which is based on the study of a unique and ubiquitous marker. This method has revealed a considerable and unsuspected diversity of microbial eukaryotes in aquatic environments. However, this approach is mainly descriptive and does not allow for the determination of the physiology or understanding of the role of these microorganisms in ecosystems. Other methods, such as metatranscriptomics, offer the possibility of studying their metabolic potential in relation to environmental parameters. Nevertheless, these high-throughput sequencing approaches lead to the production of a vast quantity of environmental sequences, most of which remain unknown. To better understand the diversity-function link within microbial eukaryotes and their role in lacustrine trophic networks, several approaches have been used.

Metabarcoding coupled with a study of morpho-physio-phenotypic traits, metatranscriptomics and a methodology based on the isolation and characterization of host-parasite pairs (sequencing and *in situ* hybridization), were carried out on lake samples (Pavin, meromictic; Aydat, dimictic). These analyses revealed the high diversity of photo-osmo-phago-mixotrophs and parasites, while also highlighting the strong seasonal variations they undergo in the mixolimnion of lake Pavin. For example, periods of mixing benefiting photosynthetic host communities favor the development and dissemination of parasitic fungi, notably through the overexpression of genes involved in zoospore phototaxis and lipid metabolism. Among these parasitic fungi, Microsporidia are newly identified players in aquatic food webs. Indeed, we discovered a high prevalence (42.5%) host-parasite association between a potential new species of Microsporidia and a species of rotifer in lake Aydat. An important rare biosphere has also been highlighted in the anoxic monimolimnion of Lake Pavin, characterized by numerous saprotrophs overexpressing genes related to sulfur, nitrate, and organic matter degradation metabolisms. The characteristic metabolisms of organisms of different trophic modes have also been studied by constructing protein sequences similarity networks.

While characterizing the majority of unknown sequences for the first time (>40%), we have revealed the genetic proximity of proteins between heterotrophic and photo-osmo-phago-mixotrophic microorganisms and between saprotrophs and parasites, as well as a relative functional redundancy of primary metabolisms. On the other hand, we have identified nearly one million proteins characteristic of a single functional group, which, for some, represent real prospects for studying the metabolic pathways involved in host-parasite interactions.

Table des matières

CHAPITRE I : Contexte bibliographique

I. Les eucaryotes unicellulaires	5
A. Découverte, histoire et évolution de la phylogénie des eucaryotes unicellulaires	5
B. Structure des communautés d'eucaryotes unicellulaires au sein des réseaux trophiques : de la taxonomie vers l'établissement de groupes fonctionnels	15
i. Du concept d'écosystème au réseau trophique	15
ii. Rôle majeur des parasites	17
iii. La classification en groupes fonctionnels	25
II. Les approches « -omics » dans l'étude du lien diversité-fonction des eucaryotes unicellulaires	29
A. Métabarcoding.....	31
i. Principe.....	31
ii. De la diversité taxonomique aux fonctions	33
B. Métatranscriptomique.....	37
i. Principe.....	37
ii. Mesure de la diversité fonctionnelle	41
iii. Réseaux de similarité de séquences protéiques.....	43
C. Métagénomique.....	43
i. Principe.....	43
ii. Exploration des métagénomomes.....	45
III. Objectifs de la thèse et organisation du mémoire.....	47

CHAPITRE II : Étude de la diversité fonctionnelle des eucaryotes unicellulaires dans un lac méromictique par une approche de métatranscriptomique couplée à une étude basée sur le référencement de traits morpho-physio-phénotypiques

Contexte de l'étude.....	53
I. Abstract	55
II. Introduction	57
III. Materials and Methods	59
A. Study site and sampling.....	61
B. DNA and RNA extraction, amplification and sequencing procedure	61
C. Amplicons analysis and taxonomic affiliation	63

D. Taxonomic diversity analysis	65
E. Trait-based analysis and functional groups definition.....	65
F. Unigene catalog and proteins prediction and annotation	69
G. Differential gene expression analysis.....	71
IV. Results	71
A. Eukaryotic diversity	71
B. Functional groups distribution	73
C. Metabolic potential of functional groups across sampling conditions.....	77
V. Discussion	81
A. Methodological considerations	81
B. Functional groups distribution and metabolic potential of microbial eukaryotes in lake Pavin	83
i. Conditions effect	83
ii. Seasonal variation in the mixolimnion.....	85
iii. Specificity of the monimolimnion.....	87
VI. Conclusion.....	89
VII. Supplementary Materials.....	103

CHAPITRE III : Découverte de signatures génétiques au sein des parasites et d'autres groupes fonctionnels d'eucaryotes unicellulaires d'eau douce par la construction de réseaux de similarité de séquences protéiques

Contexte de l'étude.....	121
I. Abstract	123
II. Introduction	125
III. Materials and Methods	129
A. Dataset description	129
B. Unigene catalog proteins prediction and annotations	129
C. SSN Building.....	131
D. CCs analysis	133
E. SPCs related to parasitism.....	135
IV. Results	135
A. Sequences similarity network statistics.....	135
B. CCs distribution among functional groups.....	137

C. Taxonomic and metabolic specificity of functional groups	137
D. Using SSN to consider the unknown.....	141
E. SPCs related to parasitism.....	141
V. Discussion	149
A. Identify the unknown	149
B. SSN mining to study the microbial functional diversity	151
C. Toward a functional marker of parasitism.....	153
VI. Supplementary Materials.....	175

CHAPITRE IV : Découverte d’une nouvelle association hôte-microsporidie dans un lac d’eau douce par « single cell »

Contexte de l’étude.....	185
I. Abstract	187
II. Note.....	189
III. Supplementary Materials and Methods.....	201

CHAPITRE V : Discussion générale et perspectives

I. Les approches « -omics » dans l’étude du lien diversité-fonction des eucaryotes unicellulaires	209
A. Limite méthodologique	209
B. Limite analytique.....	213
C. Étude du lien diversité-fonction	219
i. Classification en groupes fonctionnels.....	219
ii. Des transcrits aux stratégies trophiques	223
II. Les interactions hôte-parasites au centre du réseau trophique	227
A. Apport des approches « -omics »	227
B. Apport des approches <i>in-vitro</i>	233

ANNEXE I : Microsporidies dans les écosystèmes lacustres et marins : importante diversité et nouveaux hôtes planctoniques

I. Abstract	245
II. Introduction	247
III. Materials and methods	249

A.	Study sites and sampling procedure	249
B.	Nucleic acids extraction, amplification and sequencing	251
C.	Amplicon analysis and taxonomic affiliation.....	251
D.	Diversity and correlation analyses	253
E.	TSA-FISH (Tyramide Signal Amplification-Fluorescent <i>In Situ</i> Hybridization) ..	255
IV.	Results and discussion.....	257
A.	Microsporidian diversity across freshwater environments.....	257
B.	Extent of the marine microsporidian diversity	261
C.	Microsporidian community composition in aquatic ecosystems in relation to their hosts.....	265
V.	Conclusion.....	273
VI.	Supplementary Materials.....	285

Références bibliographiques des chapitres I et V

CHAPITRE I

CONTEXTE BIBLIOGRAPHIQUE

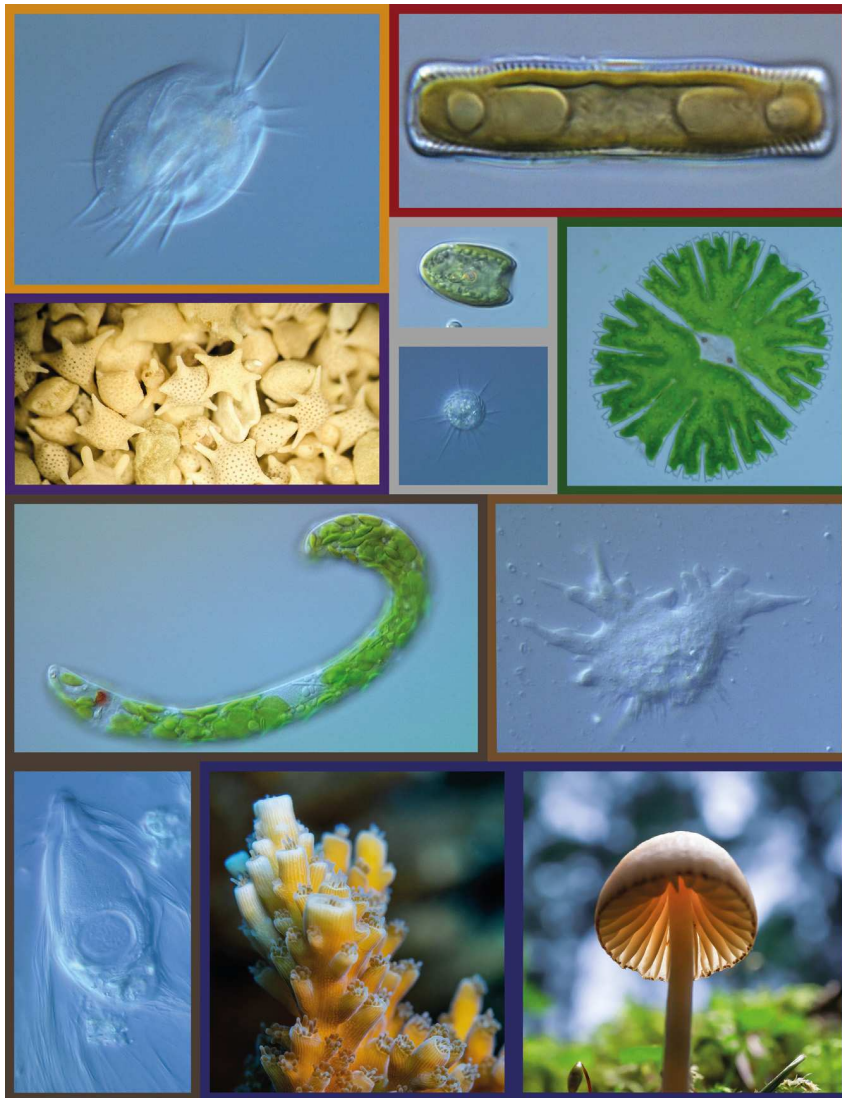


Figure 1 : Diversité des morphologies des organismes eucaryotes d’après Keeling et Burki, 2019.

De gauche à droite, la rangée supérieure montre un Ciliophora du genre *Euplotes* (Alvéolés) et un Bacillariophyta du genre *Pinnularia* (**Stramenopiles**). La deuxième rangée montre des squelettes de Foraminifera (**Rhizaria**), un Cryptomonada (**Cryptista**) au-dessus d'un Centrohelida (**Haptista**), et l'algue verte *Micrasterias* (**Archaeplastida**). La troisième rangée montre un Euglenozoa du genre *Euglena* (**Discoba**) et un Tubulinea du genre *Rhizamoeba* (**Amoebozoa**). La rangée du bas montre un Parabasalia du genre *Trichonympha* (**Metamonada**), l'animal Cnidaires *Acropora* (**Opisthokonta**) et un champignon Dikarya (**Opisthokonta**). Toutes les photos sont de P. Keeling.

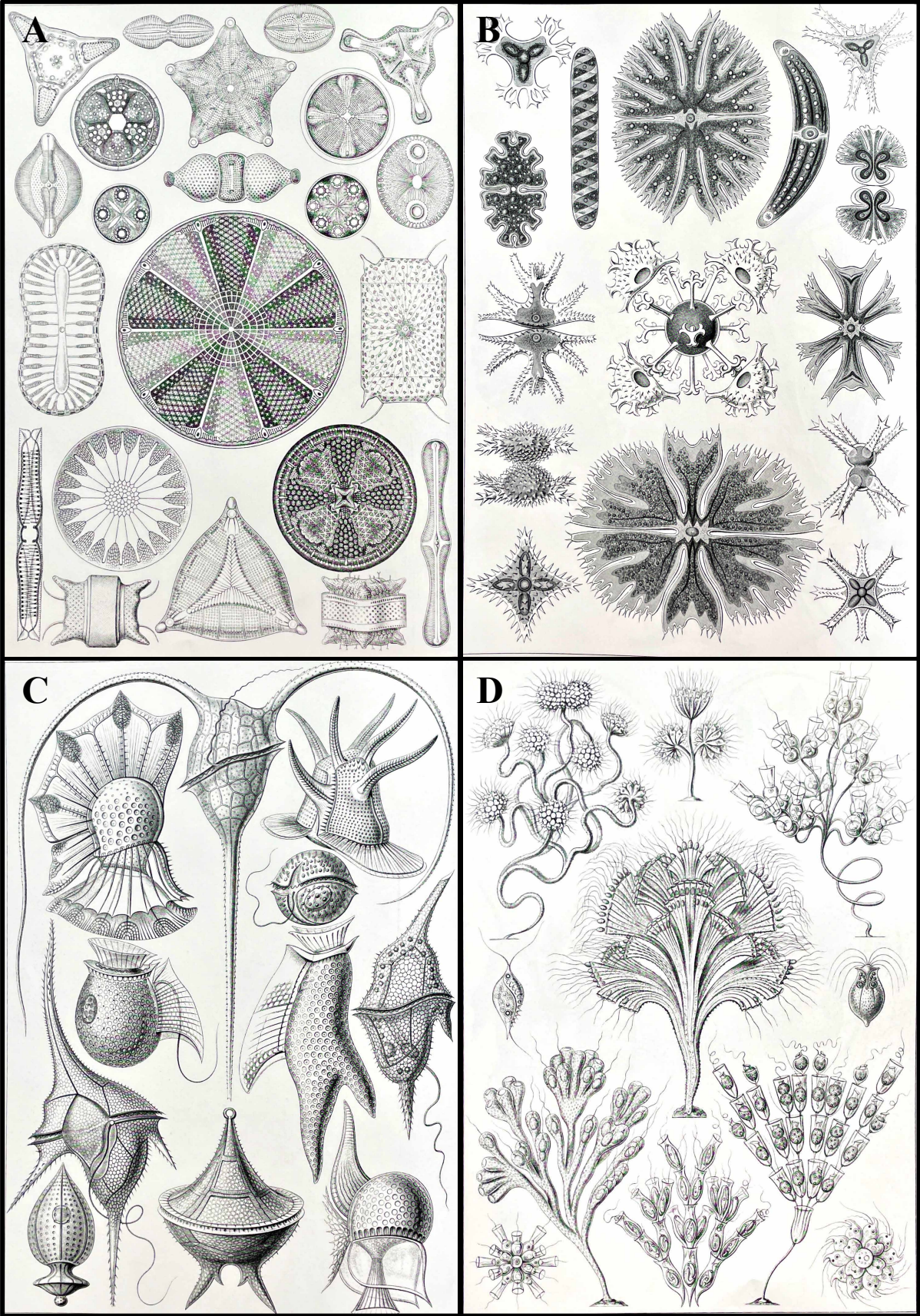


Figure 2 : Planche représentant une variété de protistes d’après Haeckel, 1899.

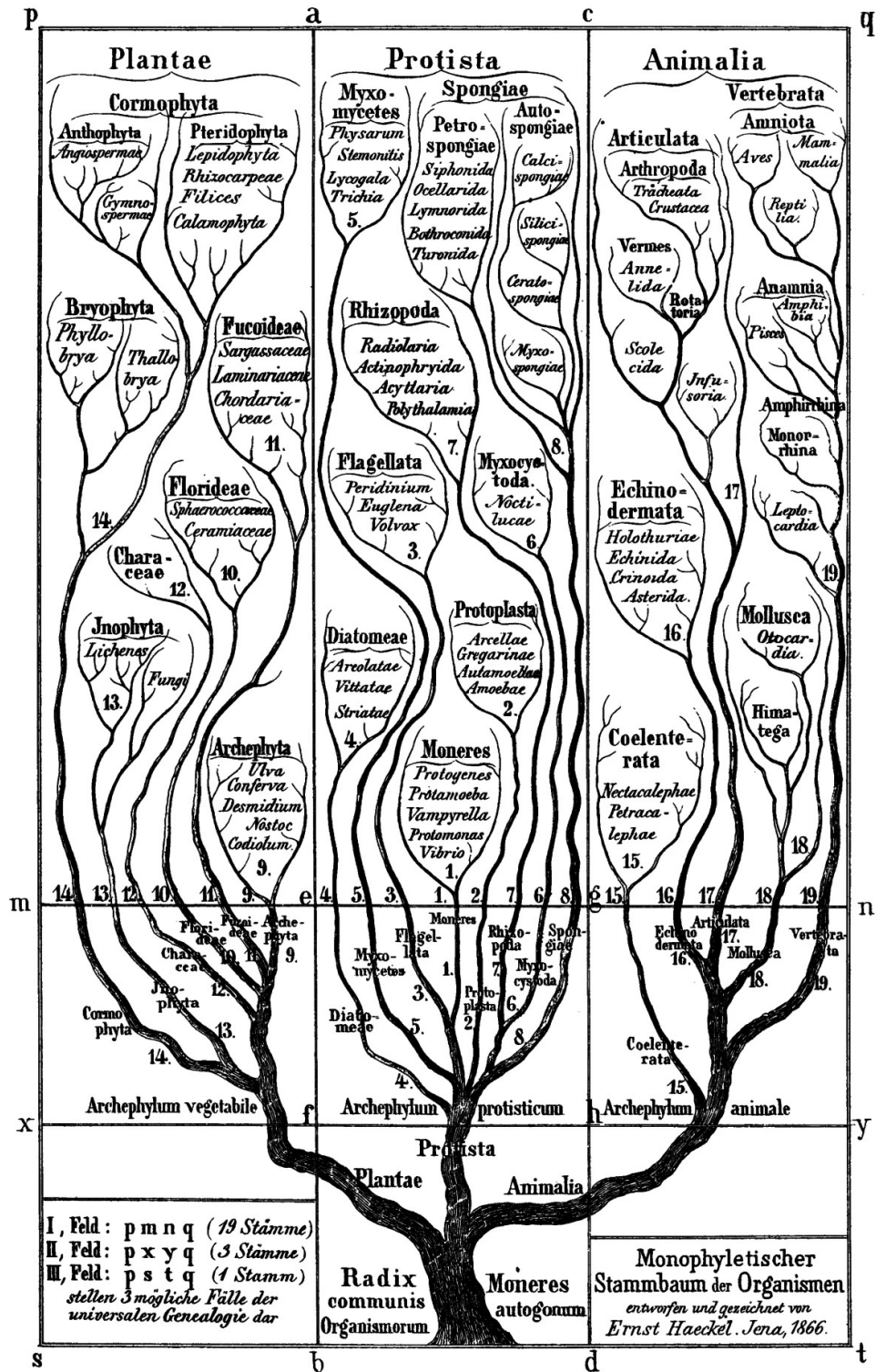


Figure 3 : Arbre du vivant d'après Haeckel, 1866.

I. Les eucaryotes unicellulaires

A. Découverte, histoire et évolution de la phylogénie des eucaryotes unicellulaires

Les eucaryotes unicellulaires sont des organismes peuplant la quasi-totalité des environnements de notre planète. On les retrouve dans les sols et les nuages, mais aussi, et tout particulièrement, dans les écosystèmes aquatiques tels que les lacs, les rivières et les océans. Ils possèdent une multitude de tailles, de formes et de fonctions (Caron *et al.*, 2012) (**Fig. 1**) et leur biomasse à l'échelle mondiale est considérable, estimée à deux fois celle de tous les animaux (Keeling et Burki, 2019). Ils présentent des modes trophiques variés tels que l'autotrophie (réalisant la photosynthèse), l'hétérotrophie (consommateurs primaires de bactéries et/ou d'organismes autotrophes, prédateurs et parasites) ou encore la mixotrophie (alternant entre l'autotrophie et l'hétérotrophie (Caron, 2016)). Les eucaryotes unicellulaires participent donc activement à la production primaire globale et aux cycles biogéochimiques *via* leur implication dans l'ensemble des niveaux des réseaux trophiques.

Les eucaryotes unicellulaires, dont la diversité a pendant longtemps été sous-estimée, représentent en réalité la grande majorité des eucaryotes (Sogin, 1989 ; del Campo *et al.*, 2014) et regroupent des espèces caractérisées par des histoires évolutives variées. Initialement, les eucaryotes unicellulaires étaient regroupés sous le terme « protistes », excluant les champignons unicellulaires. Les protistes ont été parmi les premiers taxa microbiens à être observés et décrits par Anton van Leeuwenhoek en 1670 avec la naissance des premiers microscopes et de la microbiologie. A cette époque, les protistes ont été décrits comme des organismes microscopiques assimilés à de petites plantes (les protophytes) ou de petits animaux (les protozoaires). La description d'une grande diversité de formes et de fonctions chez ces organismes s'est ensuite développée au cours des siècles suivants. Parmi ces descriptions, les illustrations d'Ernst Haeckel sont parmi les plus belles jamais réalisées (Haeckel, 1899) (**Fig. 2**). Ernst Haeckel, conscient du problème de classification posé par ces organismes ne présentant pas d'affinités plus convaincantes avec les animaux qu'avec les plantes, a proposé en 1866 un troisième règne. Cette branche comprenait à l'époque la plupart des microorganismes eucaryotes (*e.g.*, les diatomées, les radiolaires) mais aussi les monères (bactéries) et les éponges (**Fig. 3**).

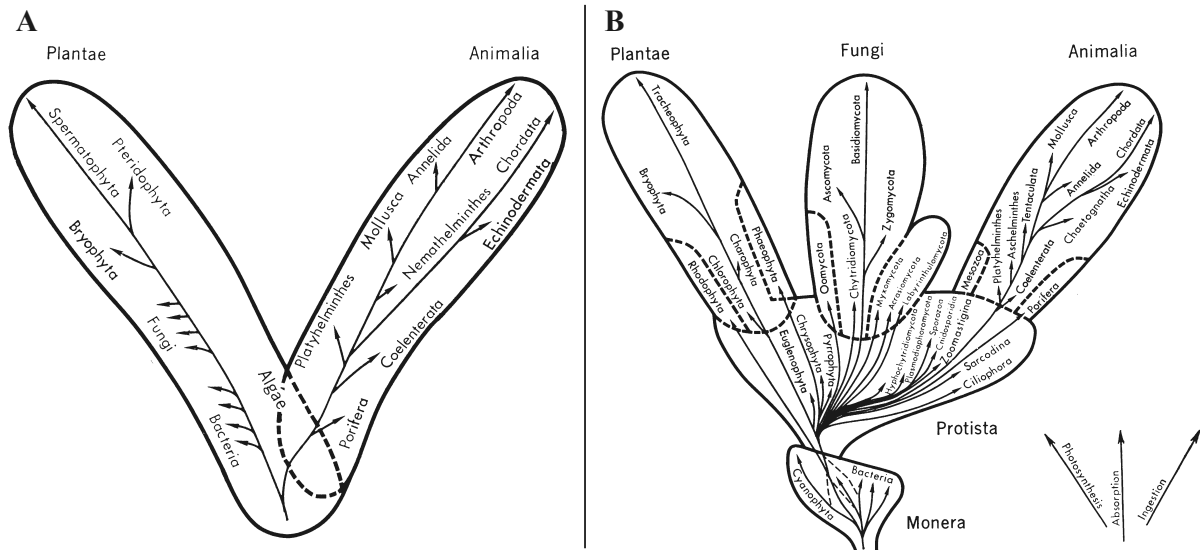


Figure 4 : Représentations d’arbres du vivant d’après Whittaker, 1969.

A : Schéma évolutif simplifié de l’arbre du vivant à deux royaumes tel qu’il aurait pu apparaître au début du XX^{ème} siècle.

B : Schéma évolutif de l’arbre du vivant à cinq royaumes, basé sur trois niveaux d’organisation : procaryote (règne des Monera), eucaryote unicellulaire (règne des Protista) et eucaryote pluricellulaire et multinucléé.

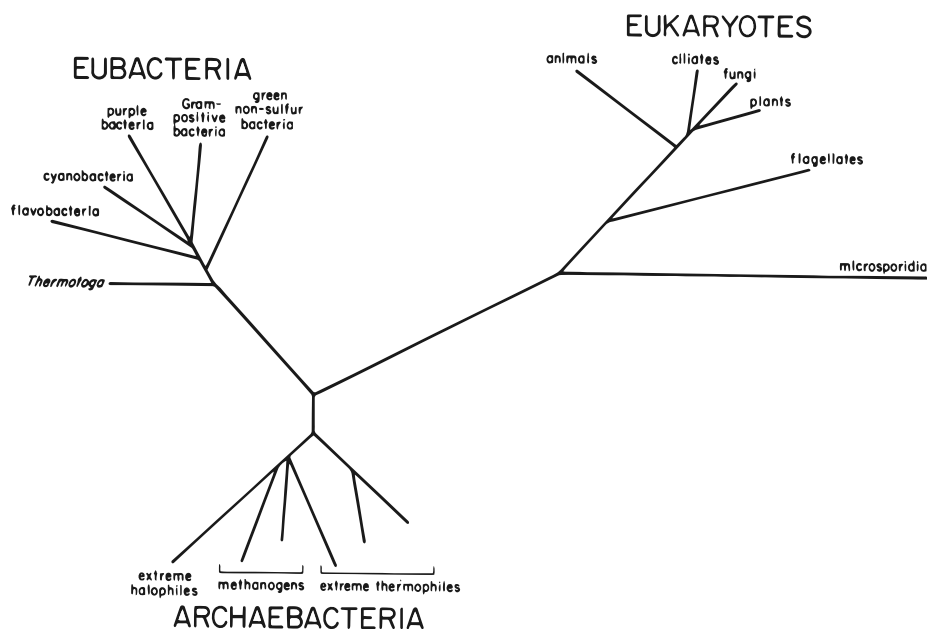


Figure 5 : Arbre représentant les trois règnes d’après Woese, 1987.

En 1911, un article fondateur de Dobell montrait que les protistes ne possèdent pas une organisation simple mais plutôt alternative, très différente de celle des plantes ou des animaux (Corliss, 1999). Pour Dobell, il était évident que les protistes actuels n'étaient en aucun cas les ancêtres des métazoaires contemporains. Les protistes ne pouvaient ainsi pas être appelés avec des termes trompeurs tels que « simples », « inférieurs » ou « primitifs », et devaient, en raison de leurs organisations particulières, être étudiés spécifiquement.

Les premiers arbres d'eucaryotes n'étaient évidemment pas basés sur des données moléculaires, mais utilisaient des caractéristiques morphologiques ou trophiques pour décrypter les relations évolutives. Les travaux de Williams et Wyckoff sur l'utilisation des premiers microscopes électroniques à transmission (MET) pour l'observation des petits objets (*i.e.*, virus influenza et de la mosaïque de tabac (Williams et Wyckoff, 1945)), appliqués à l'observation des protistes, ont permis l'acquisition des premières images de cils et de trichocystes de paramécies (Jakus et Hall, 1946). Ces observations, couplées aux nombreuses études des organes cinétiques des protistes menées à cette période (Jakus, 1945 ; Vazart, 1956 ; Faure-Fremiet, 1961) révèlent la complexité et la diversité des organismes unicellulaires, jusqu'alors décrits comme simples. Les approches basées sur la morphologie ont ainsi permis l'identification d'un grand nombre de lignées que nous reconnaissons encore aujourd'hui. Ces considérations ont alors abouti à l'établissement d'un nouvel arbre du vivant par Whittaker (1969), cette fois-ci divisé en 5 royaumes (**Fig. 4**). Les observations de multiples organismes unicellulaires en microscopie électronique ont par la suite conduit à distinguer deux types cellulaires caractérisés par des différences notables d'organisation (Stanier et Niel, 1962). Ces travaux proposent alors une classification binaire en distinguant les procaryotes, équivalents à cette époque au terme « bactérie », et les eucaryotes.

Cette classification est ensuite complétée majoritairement par les travaux de Carl R. Woese au début des années 1970 qui, en plus d'ajouter un troisième règne (celui des archées (**Fig. 5**), initialement associé aux bactéries (Woese et Fox, 1977)), est l'un des premiers à avoir l'idée d'utiliser la petite sous-unité de l'ADN ribosomal (SSU rRNA) comme marqueur moléculaire pour reconstituer l'arbre universel du vivant (Sogin *et al.*, 1986 ; Woese, 1987 ; Paolozzi et Liébart, 2015 ; Forterre, 2022).

La diversité morphologique des organismes étant si vaste, et les caractères si souvent contradictoires, qu'il était impossible de distinguer l'ensemble des détails et des relations entre la majorité des lignées décrites. Ces phylogénies moléculaires, se développant au cours des années 2000, se sont alors montrées essentielles dans la compréhension des relations évolutives entre les eucaryotes unicellulaires (*e.g.*, Pawlowski et Holzmann, 2002).

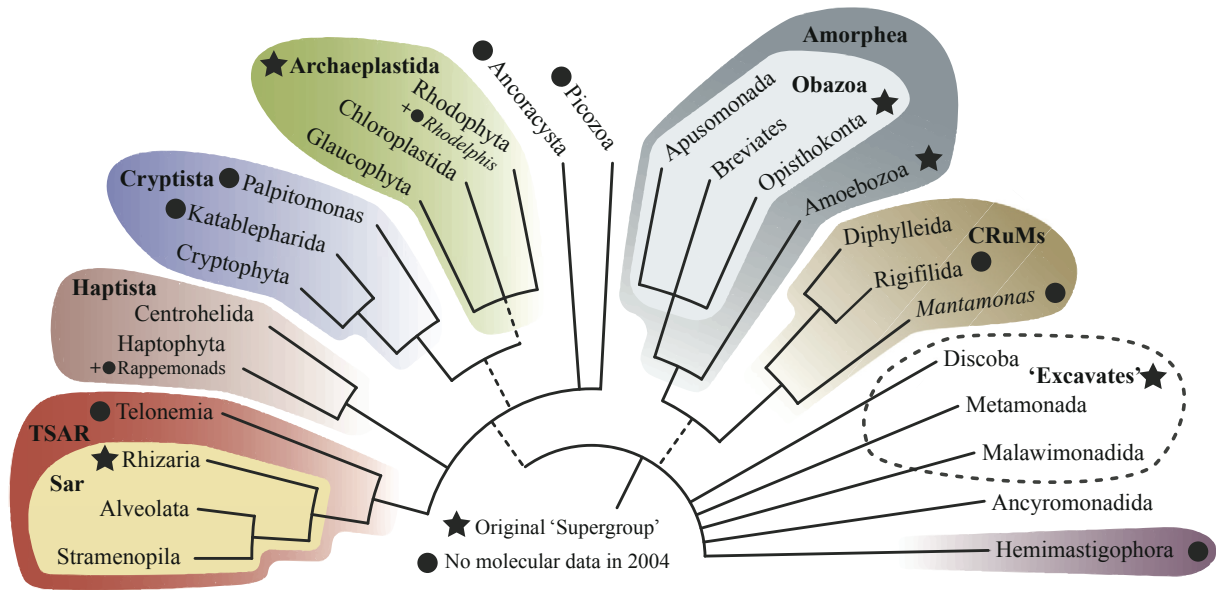


Figure 6 : Le nouvel arbre des eucaryotes ou « eToL » (i.e., eukaryotic Tree of Life) d’après Burki *et al.*, 2020.

Cet arbre synthétique est basé sur un consensus d’études phylogénomiques récentes. Les groupes colorés correspondent aux « supergroupes » actuels. Les ordres de ramification non résolus entre les lignées sont représentés par des multifurcations. Les lignes discontinues reflètent des incertitudes quant à la monophylie de certains groupes.

Box 1 : Description des principaux groupes d’eucaryotes actuellement définis.

Synthèse des descriptions de Keeling et Burki, 2019 ; Burki *et al.*, 2020 ; Tikhonenkov *et al.*, 2022). *supergroupes

***TSAR**

Le supergroupe des TSAR est constitué des Telonemia, Stramenopiles, Alveolata et Rhizaria. Les trois derniers groupes forment le clade « SAR », apparu relativement tôt dans l’ère phylogénomique et régulièrement considéré comme un supergroupe (remplaçant en partie les Chromalveolata). On estime que les SAR représentent jusqu’à la moitié de la diversité des espèces d’eucaryotes (del Campo *et al.*, 2014).

[...]

Au cours des dernières décennies, l'avènement des techniques de séquençage à haut débit a contribué à faire avancer le domaine de la systématique en permettant l'accès à une quantité massive de données de séquences. Ceci a ainsi permis aux scientifiques d'aborder en profondeur des hypothèses phylogénétiques en analysant des milliers de gènes issus de données génomiques (Philippe *et al.*, 2011 ; Chan et Ragan, 2013 ; Burki, 2014 ; Burki *et al.*, 2020). Ces analyses phylogénomiques basées sur un nombre considérable et croissant de gènes concaténés en ensembles de données ont convergé vers une image fondamentalement différente des relations entre les eucaryotes, dans laquelle la diversité est répartie en un petit nombre d'assemblages extrêmement vastes.

Ces assemblages sont appelés informellement « supergroupes » pour refléter leur nature « plus grande que le royaume ». A l'origine, cinq à six supergroupes ont été proposés, le chiffre variant en fonction de l'unification ou non des Opisthokonta et des Amoebozoa au sein des Unikonta (Baldauf, 2003 ; Simpson et Roger, 2004). Adl *et al.* (2005) proposent un schéma de classification basé sur des données structurales et moléculaires comprenant six supergroupes : les Opisthokonta, les Amoebozoa, les Excavata, les Rhizaria, les Archaeplastida et les Chromalveolata. Des réorganisations ont ensuite mené, en 2007, à la combinaison des Stramenopiles, des Alveolata (deux groupes de Chromalveolata) ainsi que des Rhizaria en un supergroupe désigné sous le nom de « SAR » (Burki *et al.*, 2007). C'est en 2019 que le groupe orphelin des Telonemia est intégré au « SAR » formant ainsi le nouveau supergroupe « TSAR » (Strassert *et al.*, 2019). Burki *et al.* (2020) synthétisent les résultats d'analyses phylogénomiques et les principales lignées ajoutées au cours des 15 dernières années en un nouvel arbre des eucaryotes. Cet arbre consensus actuel a été remanié au point que la plupart des supergroupes d'origine ont été soit intégrés dans de nouveaux taxa, soit ont complètement disparu (**Fig. 6, Box 1**).

[...]

Les **Stramenopiles** comprennent des eucaryotes unicellulaires photosynthétiques bien connus (*e.g.*, les diatomées), mais aussi des macroalgues (*e.g.*, les laminaires) et une vaste diversité de protistes hétérotrophes ou mixotrophes libres ainsi que d'importants agents pathogènes d'animaux et de plantes (*e.g.*, *Blastocystis* ou les oomycètes). Les Stramenopiles contiennent également certains des taxa les plus abondants dans l'océan (MAST : MARine STRamenopiles), dont la plupart n'ont pas été caractérisés au niveau cellulaire (Labarre *et al.*, 2021).

Les **Alveolata** comprennent trois groupes d'eucaryotes unicellulaires bien étudiés (Ciliophora, Dinoflagellata et Apicomplexa), ainsi que plusieurs groupes de parasites et de flagellés tels que les Perkinsea, les Chrompodellids et Colponemida. Les Ciliophora constituent un groupe majeur de prédateurs dans tous les environnements aquatiques (Weisse, 2017). Les dinoflagellés sont également extrêmement abondants et peuvent être photosynthétiques, prédateurs et parasites (Taylor *et al.*, 2007). Les Apicomplexa, quant à eux, sont des parasites intracellulaires (del Campo *et al.*, 2019).

Le groupe des **Rhizaria** comprend une grande diversité de protistes essentiellement amiboïdes, dotés de pseudopodes utilisés davantage pour l'alimentation que pour la locomotion. Le groupe comprend également des amiboflagellés et des flagellés, parasites de plantes cultivées et d'invertébrés (*e.g.*, *Plasmodiophora* et *Bonamia*, respectivement), ainsi que des algues amiboïdes (Chlorarachniophyta).

Les **Telonemia** sont morphologiquement complexes et ne contiennent que huit espèces décrites (Klaveness *et al.*, 2005 ; Yabuki *et al.*, 2013 ; Tikhonenkov *et al.*, 2022). Les observations microscopiques et l'analyse des séquences environnementales indiquent que les Telonemia sont largement distribués. Ils jouent probablement un rôle écologique important en tant que consommateurs de bactéries et d'eucaryotes unicellulaires aussi bien en milieu marin que dans les eaux douces.

*Haptista

Ce supergroupe comprend les Haptophyta (précédemment assignés aux Chromalveolata) et les Centrohelida (Burki *et al.*, 2016). Les Haptophyta sont des espèces photosynthétiques, principalement marines, pouvant se développer jusqu'à de fortes densités (*e.g.*, le coccolithophore *Emiliania*). Les Centrohelida, en revanche, sont des hétérotrophes libres caractérisés par des pseudopodes radiants distinctifs (axopodes) et se trouvant le plus souvent dans des environnements d'eau douce.

[...]

[...]

***Cryptista**

Le supergroupe des Cryptista contient les Cryptomonada (anciennement Chromalveolata), une lignée qui a été centrale dans l'étude de l'origine des plastes et de leur propagation à travers les eucaryotes. Il comprend également les Katablepharids et les *Palpitomonas* plus récemment découverts et qui restent tous deux des flagellés hétérotrophes énigmatiques (Yabuki *et al.*, 2014).

***Archaeplastida**

Les trois taxa qui composent les Archaeplastida sont les Chloroplastida (algues vertes et plantes terrestres), les Rhodophyta (algues rouges) et les Glaucophyta. Ils sont définis par la présence de plastides primaires directement dérivés de l'endosymbiose avec une cyanobactérie.

***Amorphea**

Ce supergroupe renferme les Opisthokonta et les Amoebozoa. Les Amoebozoa possèdent des pseudopodes pour l'alimentation et la locomotion (Blandenier, 2020). Ce groupe comprend également des flagellés, ainsi que certains pathogènes importants (*e.g.*, *Entamoeba*). Les Opisthokonta comprennent des eucaryotes unicellulaires, mais également les animaux et les Fungi. Les unicellulaires sont composés d'une variété de flagellés hétérotrophes, d'amibes et de parasites. Le supergroupe des Amorphea comprend maintenant également deux petites lignées de flagellés hétérotrophes, les Breviata et les Apusomonada.

***CRuMs**

CRuMs représente un nouveau supergroupe nommé à partir des noms de ses membres : collodictyonids (syn. diphylleids), Rigifilida et Mantamonas. Ces trois taxa d'eucaryotes unicellulaires libres possèdent des morphologies très différentes (flagellés nageurs, cellules amiboïdes filosees et petites cellules glissantes, respectivement) et étaient précédemment des « groupes orphelins », mais se sont solidement coalisés dans des analyses phylogénomiques récentes (Brown *et al.*, 2018).

***Hemimastigophora**

Les Hemimastigophora sont des eucaryotes unicellulaires prédateurs libres. Ils ont été observés pour la première fois au XIXe siècle mais n'ont jamais été cultivés (Foissner *et al.*, 1988). Les analyses phylogénomiques récentes, basées sur des transcriptomes de « single cell », montrent que les Hemimastigophora représentent l'une des branches les plus profondes des eucaryotes. Ne pouvant pas être reliés à l'un des supergroupes « établis » (ou « orphelins »), il a été proposé de les considérer comme un nouveau supergroupe.

[...]

[...]

Les Discoba et les Metamonada sont deux groupes précédemment classés (avec les Malawimonadida) dans le supergroupe des **Excavata** en raison d'une morphologie distinctive. Les phylogénies moléculaires ont pour la plupart échoué à soutenir cela, de sorte qu'ils sont considérés comme des groupes distincts et peut-être liés. Les Discoba comprennent les Euglenozoa photosynthétiques (*e.g.*, *Euglena*), des parasites (*e.g.*, *Trypanosoma*) et de nombreux flagellés hétérotrophes libres (Pánek *et al.*, 2014). Les Metamonada contiennent des protistes anaérobies, dont plusieurs pathogènes (*e.g.*, *Giardia*, *Trichomonas*) et des symbiotes intestinaux de nombreux animaux (*e.g.*, *Trichonympha*).

Groupes orphelins

Il s'agit de plusieurs groupes apparemment pauvres en espèces pour lesquels les analyses phylogénomiques n'ont jusqu'à présent pas fourni de placement phylogénétique convaincant. Ces groupes comprennent les Ancoracysta, Picozoa, Malawimonadida et Ancyromonadida (aussi appelés Planomonadida), qui sont exclusivement des eucaryotes unicellulaires libres (Burki *et al.*, 2020).

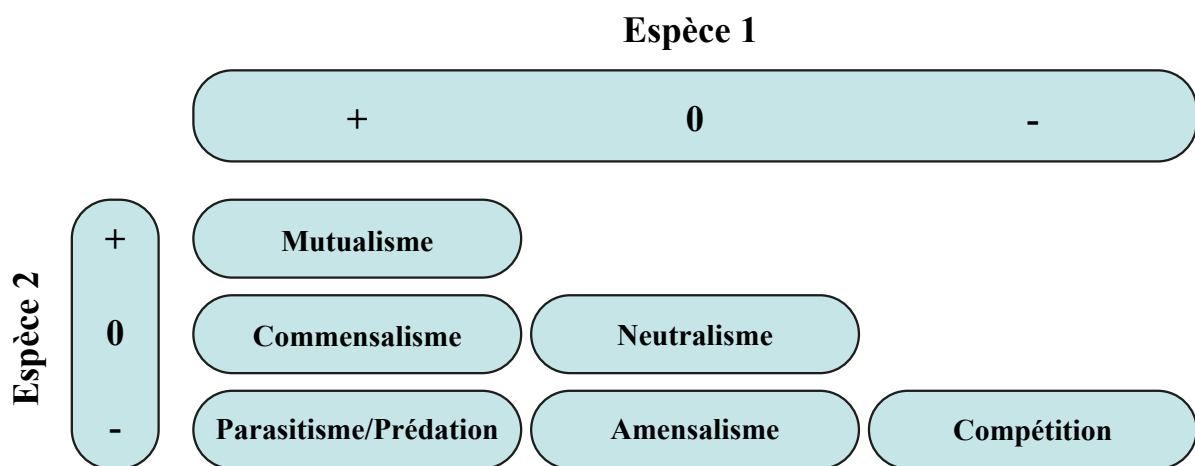


Figure 7 : Représentation des mécanismes d'interactions possibles entre deux espèces (modifiée d'après Bronstein, 1994).

Les « + », « 0 » et « - » représentent les effets de l'interaction sur chacune des deux espèces.



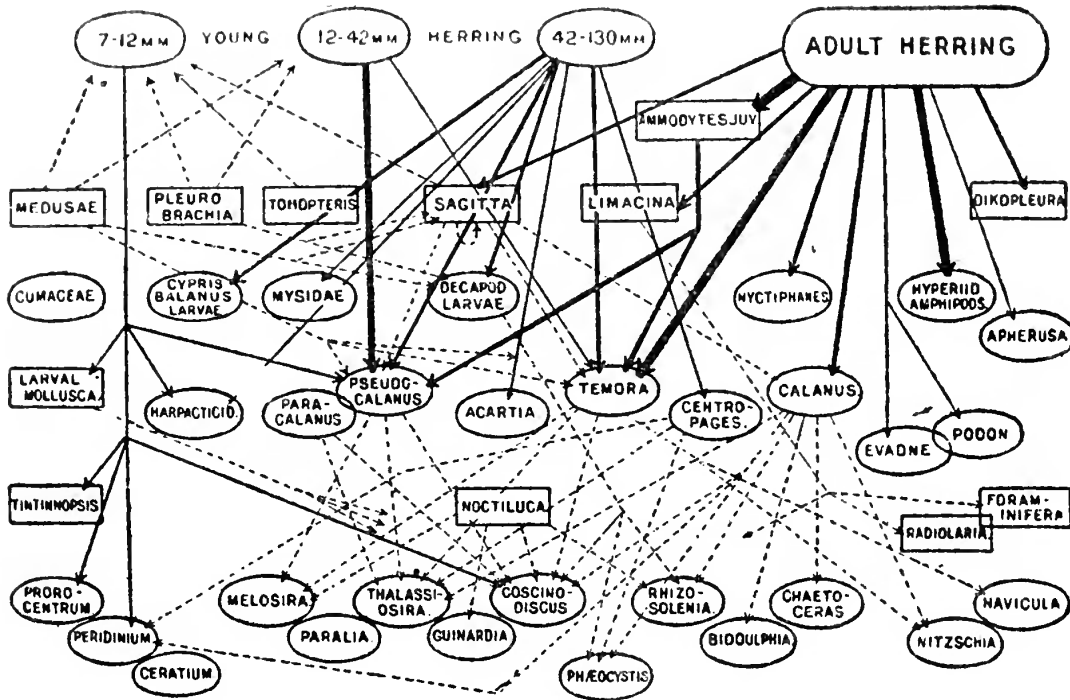


Figure 8 : Réseau trophique marin d'après Hardy, 1924.

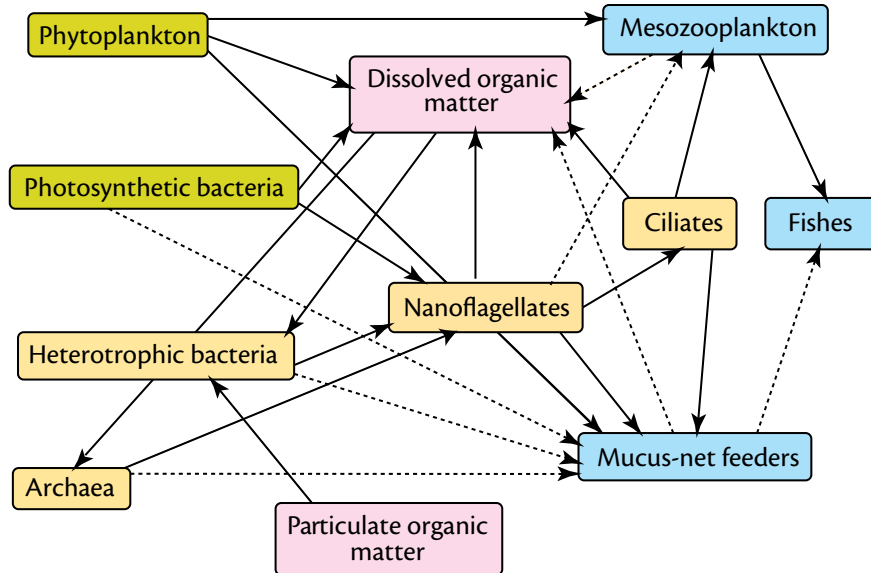


Figure 9 : Réseau trophique marin simplifié montrant le rôle dominant de la boucle microbienne d'après Pomeroy *et al.*, 2007.

Les principaux flux de carbone et d'énergie sont délimités par des lignes continues ; les flux généralement moins importants sont délimités par des lignes discontinues. Les cases représentant les organismes faisant partie de la boucle microbienne sont indiquées en vert (photosynthétique) et en jaune (hétérotrophe).

B. Structure des communautés d'eucaryotes unicellulaires au sein des réseaux trophiques : de la taxonomie vers l'établissement de groupes fonctionnels

i. Du concept d'écosystème au réseau trophique

Le concept d'écosystème, introduit dans les années 1930 (Tansley, 1935), associe pour la première fois les caractéristiques physiques et chimiques d'un milieu à l'ensemble des entités biologiques qui l'habitent. Il unifie ainsi deux éléments indissociables : une partie biotique, la biocénose, décrite plusieurs années auparavant par Möbius (1877), et une partie abiotique : le biotope, plus ou moins bien délimité, renfermant les ressources nécessaires au maintien de la vie. Un écosystème désigne alors un système à deux composantes, caractérisé par de multiples interactions. On distingue les interactions entre biotope et biocénose symbolisées, entre autres, par le processus d'adaptation entre les espèces et leur milieu, et également les interactions biologiques entre les êtres vivants constitutifs de la biocénose. De cette manière, la composition, la structure, la stabilité et la résilience des écosystèmes s'établissent sur la base d'un réseau complexe d'interactions parmi lesquelles les interactions biologiques sont essentielles (**Fig. 7**) (Bronstein, 1994).

C'est au début des années 1920, et sur la base de ces interactions, que le biologiste Charles S. Elton théorise le premier réseau trophique en décrivant la structure d'une biocénose animale en fonction de la taille des organismes, qu'il représente sous la forme d'une pyramide (Elton, 1927). Un peu plus tard, Hardy (1924) décrit l'un des premiers réseaux trophiques aquatiques en intégrant certains dinoflagellés, chlorophycées et diatomées et en évoquant les relations de prédation comme composante centrale du réseau (**Fig. 8**). A partir de là, et jusqu'aux années 1970, les premiers modèles de fonctionnement des réseaux trophiques se structurent autour d'une chaîne linéaire reposant sur l'assimilation photosynthétique par le phytoplancton suivie du transfert d'énergie et de matière aux maillons trophiques supérieurs (zooplancton et poissons) (Wetzel et Manny, 1972). Cette chaîne linéaire n'est complétée que plusieurs dizaines d'années plus tard, par la re-description de deux sous-systèmes du réseau trophique associant un système « herbivore » à un système « saprophage », majoritairement composé de microorganismes, et rétablissant l'équilibre des flux de matière et d'énergie (Heal et MacLean, 1975). Ces recherches fondent ainsi les bases du concept de boucle microbienne introduite en 1983, mettant ainsi en évidence l'importance du compartiment microbien dans la structure des réseaux trophiques (Azam *et al.*, 1983). Le modèle classique est alors peu à peu remplacé par un modèle plus complexe dans lequel les microorganismes ont un rôle central (**Fig. 9**) (Pomeroy *et al.*, 2007).

Box 2 : Définition générale du parasitisme.

Le parasitisme est une relation biologique entre deux êtres vivants au cours de laquelle l'un, le parasite, tire bénéfice de l'interaction, au détriment de l'autre, l'hôte, qui en est désavantagé. D'après cette définition, parasitisme et prédation sont indissociables (**Fig. 7**). Pourtant, ils diffèrent par de nombreux aspects : le parasite tire parti de son hôte pour sa reproduction, sa dispersion, sa nutrition et son développement (Combes, 2001). De plus, le parasite, généralement plus petit, ne tue que très rarement son hôte, ce qui est inévitable dans le cas de la prédation. Finalement, le terme parasite définit un être vivant qui partiellement, ou tout au long de son existence, vit au dépend d'un ou de plusieurs hôtes (Combes, 2018). On distingue les ectoparasites, vivant à la surface, des endoparasites qui quant à eux vivent à l'intérieur de leur hôte.

ii. Rôle majeur des parasites

Le compartiment microbien, central dans les réseaux trophiques, est très diversifié. Il comprend notamment un grand nombre de parasites (**Box 2**), dont le rôle a longtemps été sous-estimé (Marcogliese et Cone, 1997) et non pris en compte dans les modèles de fonctionnement des réseaux trophiques aquatiques. En effet, de manière générale, les parasites n'étaient perçus que comme des organismes de faible biomasse dont l'effet sur les flux de matière et d'énergie était dérisoire. Certaines théories généralistes estimaient, par exemple, que les proportions des groupes trophiques au sein des écosystèmes étaient constantes (Havens, 1992) et que le nombre d'interactions ne dépendait que du nombre d'espèces du réseau (« the species scaling law » (Cohen et Briand, 1984)). Toutefois, ces théories peu ou difficilement applicables aux microorganismes (Locey et Lennon, 2016), n'intégraient aucunement l'effet des parasites sur le nombre et la longueur des chaînes trophiques.

Ce n'est qu'au début des années 2000 que les parasites sont intégrés dans les modèles de réseaux trophiques, notamment du fait de réévaluations de leur abondance. Kuris *et al.* (2008) rapportent, par exemple, des biomasses équivalentes des espèces parasites et prédatrices dans différents estuaires. Des estimations de la richesse et de la diversité réalisées pour l'ensemble des organismes vivants partagent la même conclusion : les espèces parasites représentent une proportion élevée et non négligeable du monde vivant. Il a été montré que 30 (de Meeûs et Renaud, 2002) à plus de 50% (Morand, 2015) des espèces pouvaient être parasites. D'autres études encore décrivent le parasitisme comme étant le mode de vie le plus courant sur terre (Lafferty *et al.*, 2006 ; Dobson *et al.*, 2008 ; Lafferty, 2010) et mettent en évidence l'influence majeure des parasites, tant sur le plan biologique à l'échelle d'un hôte (Dajoz, 2019), que sur le plan écologique à l'échelle des populations et des communautés (Combes, 2001 ; Hudson *et al.*, 2002, 2006).

En modulant le comportement ainsi que la physiologie de son hôte (Park *et al.*, 2004), le parasite peut favoriser sa propre dispersion et faciliter sa transmission. Il est également capable de moduler le système immunitaire de l'individu parasité (Alvarez-Pellitero, 2008), afin d'en limiter l'effet, et ainsi favoriser l'établissement d'une interaction durable (Dajoz, 2019). Ces mécanismes impliquent que l'expression des gènes d'un des partenaires influence des cellules et des tissus qu'il n'a pas lui-même élaborés. Il est question de manipulation parasitaire, on parle même de phénotype étendu (Dawkins, 1982 ; Hunter, 2009).

Enfin, un parasite est également susceptible de modifier le sex-ratio d'une espèce hôte (Dajoz, 2019). C'est le cas, par exemple, d'espèces de microsporidies capables d'induire une modification du sex-ratio chez des espèces de crustacés par féminisation des mâles (Ironsides et Alexander, 2015).

Le parasitisme influence également l'écologie des communautés en modulant l'effectif et la composition des populations qui les constituent. En effet, l'effectif d'une population dépend de facteurs tels que le succès reproducteur (« fitness » en anglais) ou la mortalité, lesquels étant tous deux sensiblement affectés par le parasitisme (Hudson *et al.*, 2006). Même si provoquer la mort de son hôte est en contradiction avec ce qu'implique le mode de vie parasitaire, ces événements peuvent survenir, notamment si l'hôte est trop faible ou le parasite trop virulent. Ce phénomène a pour conséquence de moduler la compétition intraspécifique. En effet, diminuer l'effectif d'une espèce revient à améliorer son accès à davantage de ressources et éventuellement favoriser sa survie face à d'autres stress (Combes, 2018). Ces modifications d'effectif peuvent toutefois s'avérer irréversibles. Bien que l'extinction d'une espèce due uniquement au parasitisme soit un événement rare (Ebert *et al.*, 2000), ce phénomène a déjà été constaté (Cunningham et Daszak, 1998).

Enfin, le parasitisme influence la composition des communautés en modulant les compétitions interspécifiques par le biais d'un phénomène d'arbitrage parasitaire (Thomas *et al.*, 2000 ; Combes, 2018). Ainsi, le devenir de deux espèces en compétition dépend de leurs parasites. Dans certains cas, un parasite peut équilibrer une interaction s'il désavantage l'espèce la plus adaptée à son environnement et, de surcroît, la plus compétitive. Dans d'autres, il la déséquilibre en parasitant l'espèce la moins compétitive. De cette manière, le parasitisme participe activement à la dynamique des écosystèmes en modifiant leurs degrés de biodiversité (Hudson *et al.*, 2006). Par ailleurs, la plupart des parasites sont caractérisés par des cycles de vie relativement complexes au cours desquels ils infectent différents hôtes, eux-mêmes de niveaux trophiques différents (Arias-González et Morand, 2006).

En conclusion, le parasitisme représente une forme d'interaction entre organismes qui augmente grandement la complexité des réseaux trophiques (Dunne *et al.*, 2013) en ayant des effets directs sur leurs hôtes, de par leur virulence, ainsi que des effets indirects sur les populations et les communautés.

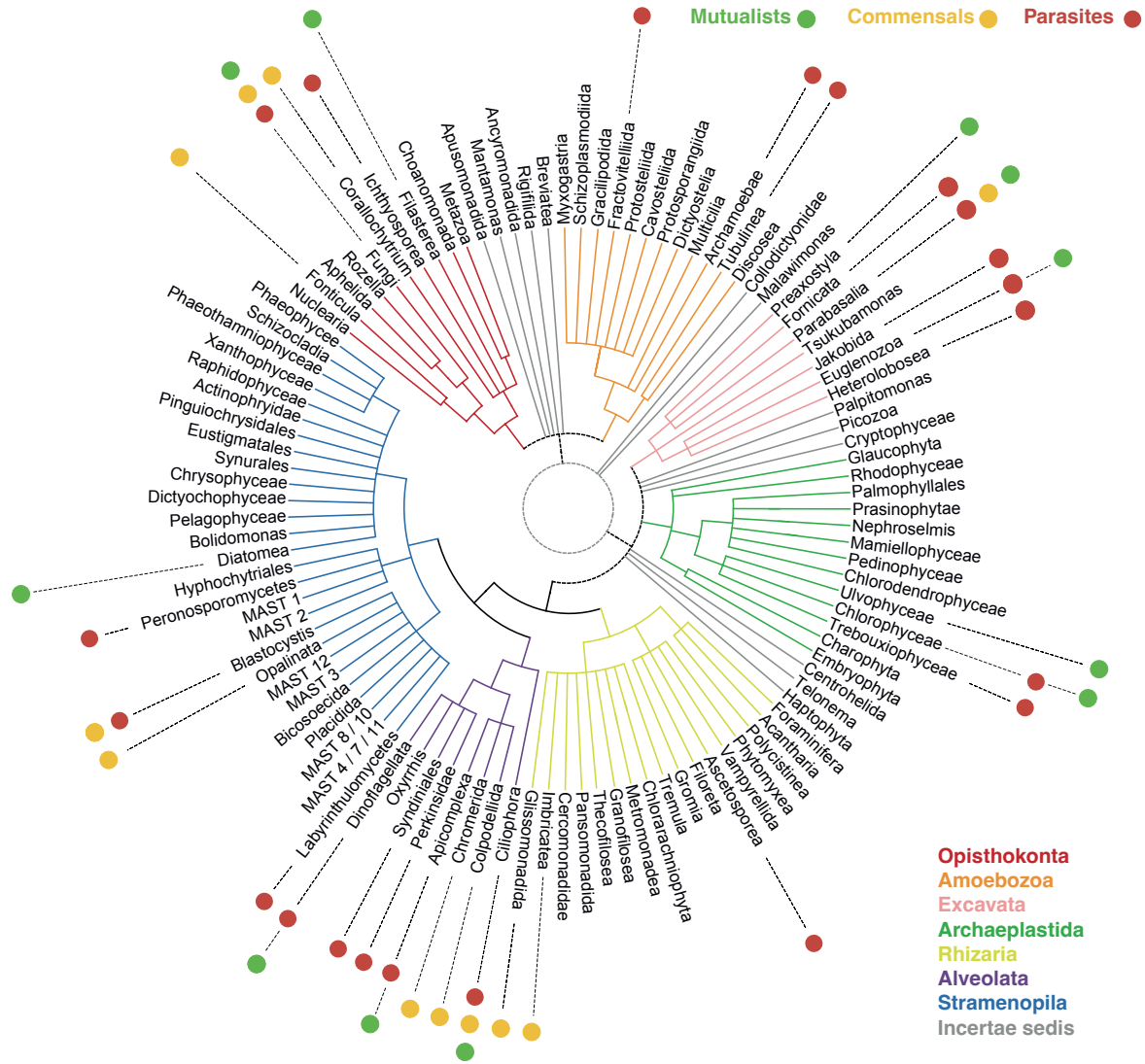


Figure 10 : Arbre phylogénétique des eucaryotes identifiant certains types d’interactions biologiques d’après del Campo *et al.*, 2020.

Les points de couleurs indiquent les groupes dans lesquels sont retrouvés des organismes mutualistes (vert), commensalistes (jaune) et parasites (rouge). Seuls les groupes pour lesquels une relation symbiotique avec les animaux a été documentée sont indiqués dans l'arbre.

En milieu aquatique, et notamment dans les écosystèmes lacustres, l'avènement du séquençage haut débit, ainsi que les importants progrès techniques afférents aux méthodes d'analyse de séquences environnementales, ont permis une meilleure description de la diversité des communautés microbiennes eucaryotes (estimée à $20 \cdot 10^4$ espèces (Debroas *et al.*, 2017)) et de leur structure (Lepère *et al.*, 2016). Ces études ont également mis en évidence une proportion d'eucaryotes microbiens parasites étonnamment élevée (de Vargas *et al.*, 2015 ; Debroas *et al.*, 2017). Parmi les phyla habituellement retrouvés dans les écosystèmes aquatiques, nombreux sont ceux qui contiennent des organismes parasites. Il s'agit notamment d'organismes appartenant aux Fungi (Opisthokonta), Alveolata, Rhizaria et Stramenopiles (**Fig. 10**) (Lefranc *et al.*, 2005 ; Lefèvre *et al.*, 2008 ; Jobard *et al.*, 2010 ; Lepère *et al.*, 2019 ; del Campo *et al.*, 2020 ; Singer *et al.*, 2021).

Parmi ces différents groupes, les Fungi sont ubiquistes et jouent un rôle crucial dans le fonctionnement des écosystèmes aquatiques (Wurzbacher *et al.*, 2010, 2014 ; Sime-Ngando *et al.*, 2011 ; Lepère *et al.*, 2019). Certains, définis comme saprophytes, participent notamment à la décomposition de la matière organique. D'autres, définis comme parasites, sont susceptibles d'affecter des espèces aquatiques présentes à tous les niveaux de la chaîne trophique (phytoplancton, zooplancton, vertébrés planctonophages) (**Fig. 11**). Dans les lacs, beaucoup d'organismes parasites sont affiliés aux branches divergentes les plus anciennes des Fungi et sont regroupés sous le nom de Dark Matter Fungi (DMF) (Grossart *et al.*, 2015). Cette DMF reste largement méconnue, les microorganismes affiliés à celle-ci ayant été identifiés principalement sur la base d'analyses moléculaires. Par ailleurs, bien que ces analyses aient mis en évidence une importante diversité (Karpov *et al.*, 2014 ; Lazarus et James, 2015), très peu d'espèces sont à ce jour cultivées. Leur physiologie, leur rôle écologique ainsi que leur cycle de développement restent par conséquent incertains. C'est le cas par exemple d'organismes appartenant aux phyla des Aphelida (Karpov *et al.*, 2014) et des Cryptomycota (James et Berbee, 2012), dont le mode de vie présumé est le parasitisme.

D'autres Fungi, les Chytridiomycota, sont également omniprésents en milieu lacustre. Pourtant décrits depuis plusieurs dizaines d'années (Mackin *et al.*, 1950), les approches moléculaires ont permis de les « redécouvrir », de décrypter l'étendue de leur diversité et de mieux comprendre leur rôle écologique. Dans les écosystèmes lacustres, en plus d'avoir une importante activité de décomposeurs, ils sont également parasites de nombreuses espèces phytoplanctoniques (Kagami *et al.*, 2012 ; Agha *et al.*, 2016 ; Van den Wyngaert *et al.*, 2022). En effet, ils sont capables d'infecter des diatomées du genre *Asterionella* (Kagami *et al.*, 2007 ; Rad-Menéndez *et al.*, 2018) ou encore des chlorophycées du genre *Pediastrum* (Kagami *et al.*, 2012).

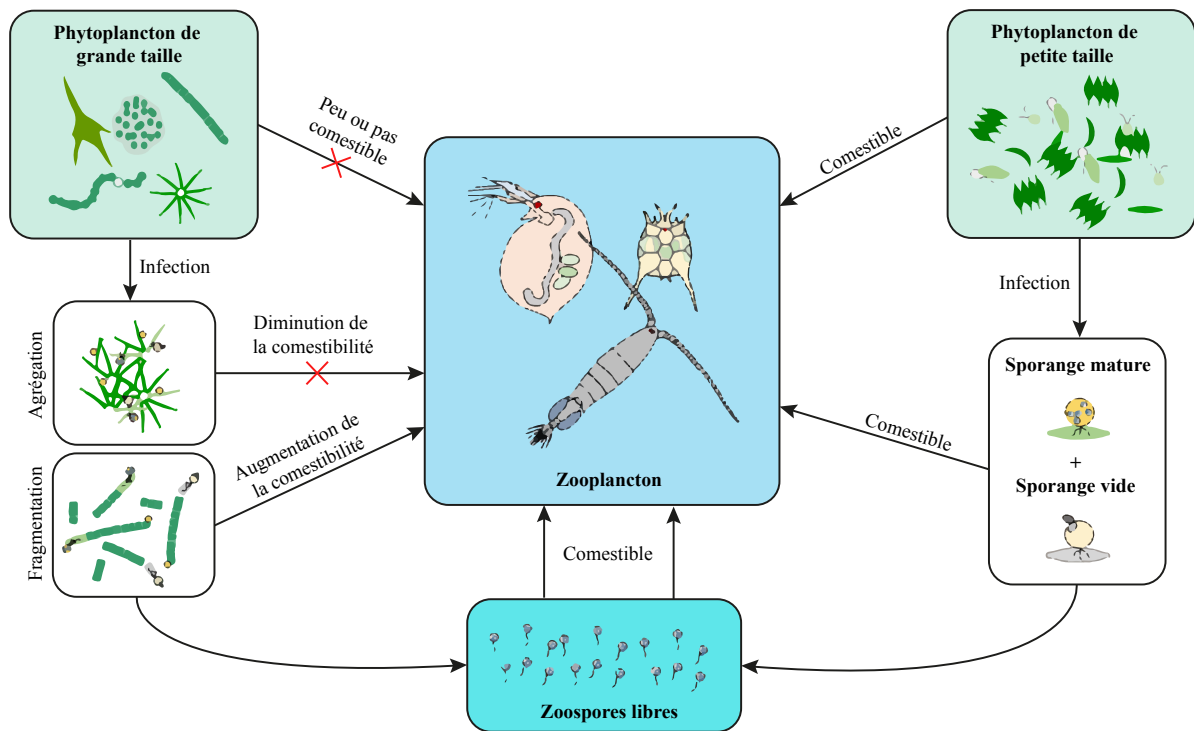


Figure 11 : Impact des chytrides sur le fonctionnement des réseaux trophiques (modifiée d'après Frenken *et al.*, 2017).

Les petites espèces phytoplanctoniques sont facilement assimilables par le zooplancton contrairement aux grandes constituant des proies peu, voire non comestibles. Les infections des grandes espèces de phytoplancton par les chytrides peuvent induire des changements dans leur comestibilité *via* une agrégation des cellules hôtes (non comestibles pour le zooplancton) ou leur fragmentation (comestibles). Les chytrides représentent également eux-mêmes une source d'alimentation puisque les sporanges (attachés aux hôtes phytoplanctoniques) et les zoospores libres peuvent également être consommés par le zooplancton.

Par ailleurs, leurs stades de dissémination (*i.e.*, les zoospores, taille $\leq 10 \mu\text{m}$) constituent une importante source de matière organique, notamment pour le zooplancton (**Fig. 11**), alimentant ainsi les réseaux trophiques (Sime-Ngando *et al.*, 2011 ; Rasconi *et al.*, 2014 ; Frenken *et al.*, 2017).

Les Perkinsea constituent un autre exemple de parasites eucaryotes mis en évidence en milieu lacustre pour la première fois en 2005 (Lefranc *et al.*, 2005). Il s'agit de microorganismes appartenant au groupe des Alveolata et pour lesquels le mode de vie parasitaire est supposé prédominant (Moreira, 2003). Initialement découverts en milieu marin, leur distribution ainsi que leur mode de vie parasitaire ont permis d'identifier leur rôle majeur dans la régulation des populations phytoplanctoniques marines ainsi que des populations de bivalves (Mackin *et al.*, 1950 ; Norén *et al.*, 1999 ; Itoiz *et al.*, 2021, 2022 ; Reñé *et al.*, 2021). Le rôle écologique ainsi que les hôtes des Perkinsea dans les milieux lacustres restent cependant difficiles à identifier du fait de l'absence d'espèce cultivée (Jobard *et al.*, 2020).

Enfin, d'autres espèces suspectées parasites, affiliées aux Cercozoa (Rhizaria) et aux Oomycota (Stramenopiles), ont également été détectées en milieu lacustre (Lefèvre *et al.*, 2008 ; Lepère *et al.*, 2008 ; Debroas *et al.*, 2017). A l'instar des Perkinsea, le mode de vie parasitaire d'espèces rattachées à ces groupes a été mis en évidence en milieu marin (Scholz *et al.*, 2016). Les Oomycota sont notamment connus pour parasiter certains genres de diatomées tels que les *Pseudo-nitzschia* (Garcia *et al.*, 2018 ; Garvetto *et al.*, 2018). Les Cercozoa, quant à eux, incluent entre autres des espèces parasites de bivalves (Azevedo et Hine, 2016). L'observation du mode de vie parasitaire de ces organismes demeure toutefois rare en milieu lacustre.

Par ailleurs, malgré la découverte croissante de séquences environnementales affiliées à des parasites, leur diversité reste sous-estimée. En effet, de par leurs divergences génétiques, certains groupes sont absents des études visant à décrire la diversité globale des eucaryotes, ou plus spécifiquement des parasites. C'est le cas, par exemple, des microsporidies (Bass *et al.*, 2015). Ces dernières constituent un vaste groupe de parasites intracellulaires obligatoires apparentés aux champignons, connus principalement comme parasites de vertébrés et d'invertébrés. En effet, les microsporidies sont couramment décrites comme parasites d'organismes d'intérêt (*e.g.*, insectes auxiliaires, bioindicateurs, humains, poissons et crustacés comestibles). Les microsporidies sont génétiquement divergentes des autres eucaryotes en raison de leur évolution rapide et de la réduction importante de leurs génomes (Keeling et Fast, 2002).

Box 3 : Description de quelques modes de vie actuellement décrits chez les eucaryotes unicellulaires.

Synthèse des descriptions de Worden *et al.* (2015) et Faure (2020).

Autotrophe : organisme qui synthétise sa propre matière organique à partir de matière minérale.

Phototrophe : type d'organisme autotrophe qui génère sa matière organique par le biais de la photosynthèse en utilisant la lumière comme source d'énergie et en fixant le CO₂. Le terme **photo-autotrophe** est également utilisé.

Hétérotrophe : organisme dépendant de matière organique provenant d'autres organismes comme source de nutriments pour survivre. Les termes **prédateur** et **consommateur** sont généralement utilisés comme équivalents. Ils caractérisent tous les trois le **zooplancton** mais également d'autres organismes tels que certains eucaryotes unicellulaires ciliés et flagellés (Fig. 13).

Mixotrophe : organisme capable à la fois d'autotrophie et d'hétérotrophie.

Saprotrophe : organisme utilisant de la matière organique morte ou en décomposition comme élément nutritif. Le terme **saprophyte** est utilisé dans certaines études comme équivalent (Fig. 13).

Phagotrophie : mode de nutrition impliquant l'ingestion et la digestion de particules de matière organique plus ou moins grosses. On distingue la pinocytose qui implique les petites particules de la phagocytose qui est le terme utilisé pour l'absorption des grosses particules. Ce mode de nutrition implique le plus souvent une structure cellulaire particulière servant d'embouchure (*e.g.*, cytostome chez les Ciliophora). Certaines espèces photosynthétiques ont la capacité de réaliser la phagotrophie, on parle alors de **photo-osmo-phago-mixotrophes** ou de **mixoplancton** (Fig. 13).

Osmotrophie : phénomène d'absorption de la matière organique dissoute présente dans l'environnement sous forme de petites molécules ou de macromolécules. Tous les eucaryotes unicellulaires seraient capables d'osmotrophie, ce qui suggère que toutes les espèces phytoplanctoniques photosynthétiques soient mixotrophes et non strictement phototrophes. On parle de **photo-osmo-mixotrophe** ou de **phytoplancton** (Fig. 13).

En conséquence, les amorces ciblant classiquement les régions conservées du gène de l'ARNr 18S des eucaryotes ne permettent pas d'amplifier les microsporidies (Marande *et al.*, 2009 ; Hadziavdic *et al.*, 2014 ; Hugerth *et al.*, 2014). Grâce à une approche de métabarcoding ciblé, des études récentes (Dubuffet *et al.*, 2021 ; Chauvet *et al.*, 2022) révèlent une diversité insoupçonnée de microsporidies dans les environnements aquatiques (lacs et écosystèmes marins). Cependant l'étendue de leur spectre d'hôtes reste encore à découvrir. Il n'existe en effet que très peu de descriptions de microsporidies parasites d'organismes à la base du fonctionnement des réseaux trophiques (*e.g.*, protistes, microplancton). Leur impact sur la structure et le fonctionnement des écosystèmes aquatiques est donc largement méconnu et non pris en compte.

iii. La classification en groupes fonctionnels

Historiquement, la diversité fonctionnelle est une composante de la biodiversité répertoriant le rôle des espèces au sein des communautés. Appréhender et mesurer la diversité fonctionnelle constitue depuis le début des années 1950 une préoccupation majeure en microbiologie environnementale (Petchey et Gaston, 2006 ; Violle *et al.*, 2007). Elle repose ainsi sur le rôle fondamental des espèces au sein des écosystèmes sans tenir compte de leur histoire évolutive variée. La mesure de cette diversité fonctionnelle passe traditionnellement par l'établissement de classifications fonctionnelles en catégorisant les espèces au sein de groupes délimités par des caractéristiques définies (Petchey et Gaston, 2006). La première classification fonctionnelle générale réalisée pour l'étude du monde procaryote classe ces organismes au sein de quatre groupes fonctionnels sur la base de caractéristiques décrivant la manière dont ils assimilent l'énergie et le carbone (Kinzig *et al.*, 2001 ; Naeem, 2002). L'énergie provient soit de la transformation de la lumière (photo), soit de la matière (chimio), tandis que le carbone provient de sources organiques (hétérotrophie) ou inorganiques (autotrophie).

Ces classifications n'étant que peu applicables aux eucaryotes planctoniques, d'autres, fondamentalement différentes, ont été développées et co-existent encore aujourd'hui. Certaines se basent uniquement sur des critères simples, tels que la taille (Armstrong *et al.*, 1993) comme seul descripteur de la contribution des espèces au fonctionnement des écosystèmes.

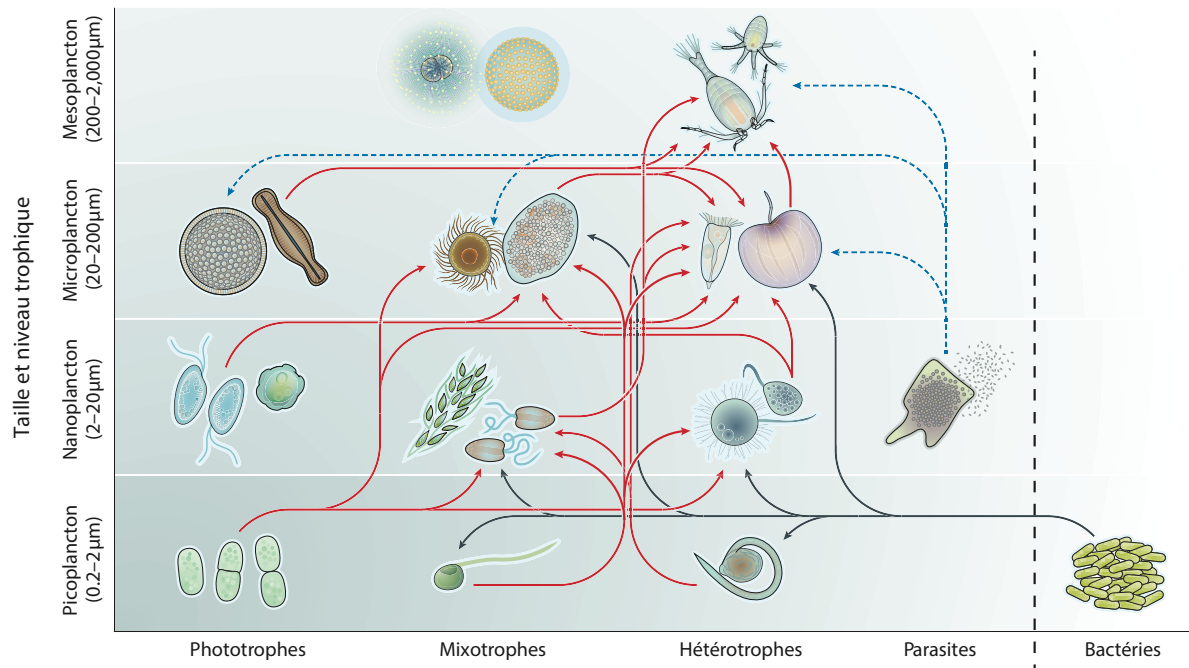


Figure 12 : Rôles écologiques et biogéochimiques des protistes en milieu marin (modifiée d’après Caron *et al.*, 2017).

Les relations prédateur-proie complexes qui impliquent les protistes rendent la matière organique produite par photosynthèse (flèches rouges) et la biomasse bactérienne (flèches noires) disponibles pour les niveaux trophiques supérieurs.

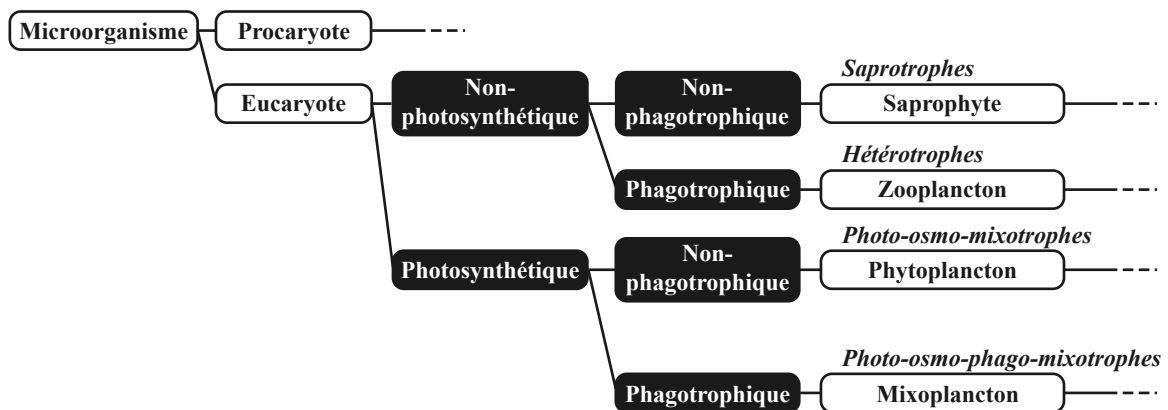


Figure 13 : Exemple de classification fonctionnelle pour les eucaryotes unicellulaires (modifiée d’après Mitra *et al.*, 2023).

Seuls deux critères sont nécessaires pour faire émerger quatre types trophiques différents : la capacité à réaliser la photosynthèse ainsi que la phagocytose.

Elles permettent ainsi de distinguer le picoplancton (0,2 μm – 2 μm), le nanoplancton (2 μm – 20 μm), le microplancton (20 μm – 200 μm) et le mésoplancton (200 μm – 2 000 μm) (**Fig. 12**). Ce type de classification suggère qu'un organisme de grande taille ait une contribution plus importante au sein des flux de matière et d'énergie qu'un organisme plus petit (Armstrong *et al.*, 1993).

D'autres classifications, bien plus complexes, proposent de nombreux groupes dont la contribution au sein de chaque cycle biogéochimique est clairement identifiée (Moore, 1982 ; Anderson, 2005 ; Le Quéré *et al.*, 2005). Le modèle proposé par Le Quéré *et al.* (2005), par exemple, a révolutionné la manière dont la diversité du plancton était représentée dans les modèles biogéochimiques en introduisant pas moins de 10 groupes fonctionnels différents (*i.e.*, les pico-hétérotrophes, les pico-autotrophes, les diazotrophes, les calcificateurs, les producteurs de DMS (diméthyle sulfide), les silicificateurs, le phytoplancton mixte, le proto-zooplancton, le méso-zooplancton et le macro-zooplancton) (**Box 3**). Toutefois, ce genre de classification nécessite un grand nombre de données métaboliques souvent inaccessibles dans les études de données environnementales (*e.g.*, approches « -omics »). Ainsi, le type de classification le plus utilisé dans ce genre d'études infère, à partir de la taxonomie, le mode référencé dans la littérature qu'occupent les espèces au sein des réseaux trophiques. Les groupes classiquement identifiés sont les phototrophes, les consommateurs (ou hétérotrophes) et les saprotrophes (de Vargas *et al.*, 2015 ; Singer *et al.*, 2021) (**Box 3**). Néanmoins ces modèles, relativement simples, ne semblent pas satisfaisants en occultant le rôle majeur des nombreuses espèces parasites au sein des réseaux ainsi que la capacité de certains eucaryotes unicellulaires à combiner phototrophie et phagotrophie (**Box 3**). En effet, la distinction fréquente entre photo- et hétérotrophe, basée uniquement sur la division historique entre le phytoplancton (phototrophe) et le zooplancton (phago-hétérotrophe) va à l'encontre de récentes recherches montrant que beaucoup d'eucaryotes unicellulaires combinent en réalité la phototrophie et la phagotrophie (Hartmann *et al.*, 2013 ; Caron, 2016 ; Caron *et al.*, 2017 ; Edwards, 2019 ; Faure *et al.*, 2019 ; Mitra *et al.*, 2023) ; ils sont regroupés sous le terme « mixoplancton » (**Fig. 13** ; en référence à la mixotrophie) (Mitra *et al.*, 2023) (**Box 3**). Ces découvertes ont complètement modifié notre perception de la structure des réseaux trophiques en les rendant bien plus complexes. En effet, le mixoplancton comprend des espèces fonctionnellement diverses : certaines possédant la capacité innée de photosynthèse (« mixoplancton constitutif »), et d'autres (« mixoplancton non-constitutif ») pouvant acquérir cette capacité par des moyens tels que la kleptoplastie ou l'endosymbiose (Gomaa *et al.*, 2014 ; Maselli *et al.*, 2021).

II. Les approches « -omics » dans l'étude du lien diversité-fonction des eucaryotes unicellulaires

L'observation au microscope couplée à la mise en culture des microorganismes eucaryotes unicellulaires est une approche cruciale pour l'exploration de la diversité des espèces, des fonctions et des rôles au sein des écosystèmes. Toutefois, longue et fastidieuse, elle ne permet actuellement pas l'étude exhaustive de la diversité de ces organismes. En effet, l'ensemble des cultures disponibles ne représente qu'une proportion faible et biaisée de la totalité de la diversité des eucaryotes unicellulaires estimée, et ce, en faveur d'organismes phototrophes (*e.g.*, les Archaeplastida et certains Stramenopiles), d'intérêt économique et/ou sanitaire (*e.g.*, certains pathogènes affiliés aux Apicomplexa) (del Campo *et al.*, 2014 ; Debroas *et al.*, 2017 ; Vitorino et Bessa, 2018 ; Singer *et al.*, 2021). Ce n'est qu'au début des années 1990 que le besoin d'approches indépendantes de la mise en culture commence à être évoqué par les scientifiques pour l'étude de la diversité des communautés microbiennes. La technique décrite par Sanger (1977), utilisée pour le séquençage de la petite sous-unité de l'ADN ribosomal comme marqueur moléculaire de la diversité (Woese et Fox, 1977) est la première approche adoptée par les scientifiques. Des variantes de cette méthode ont généré un ensemble de techniques pour (i) reconstruire les phylogénies (Woese et Fox, 1977 ; Sogin, 1991), (ii) évaluer la diversité microbienne d'échantillons naturels par la construction de banque de clones (Lefranc *et al.*, 2005 ; Lepère *et al.*, 2006), et (iii) quantifier l'abondance relative de chaque groupe taxonomique par l'utilisation de méthodes d'hybridation *in situ* (*i.e.*, FISH) (Not *et al.*, 2002 ; Riesenfeld *et al.*, 2004). C'est en 2004 que la méthode Sanger est appliquée pour la première fois à grande échelle pour l'analyse des communautés planctoniques par séquençage massif de l'ADN échantillonné (*i.e.*, naissance de la métagénomique) lors de l'expédition Global Ocean Sampling (Venter *et al.*, 2004).

Le séquençage de l'ADN provenant d'échantillons environnementaux est devenu une méthode très répandue en écologie microbienne au milieu des années 2000, notamment grâce à l'avènement des méthodes de séquençage à haut débit (HTS). Ces nouvelles technologies ont considérablement augmenté la quantité de matériel séquencé tout en diminuant fortement le coût.

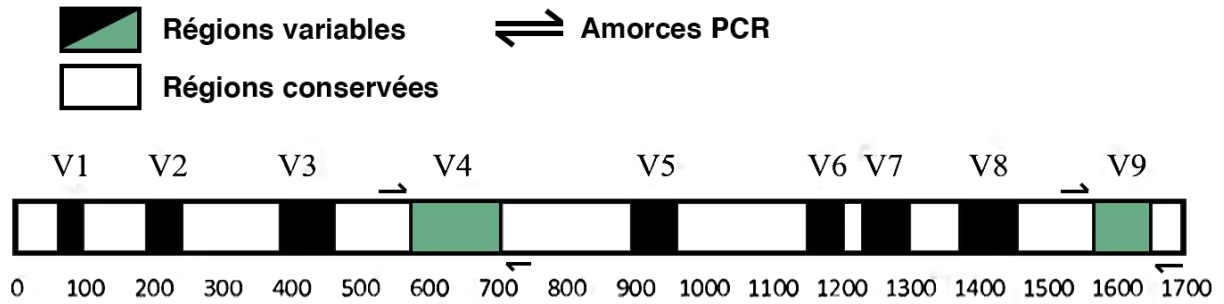


Figure 14 : Représentation schématique de la structure du gène codant pour la petite sous-unité de l'ARNr eucaryote (ARNr 18S) (modifiée d'après Ishaq et Wright, 2014).

Il présente une alternance de régions hypervariables (9 au total, V1 à V9) et de régions conservées. Des couples d'amorces s'hybridant aux régions conservées et permettant d'amplifier les régions V4 et V9 sont symbolisés par des flèches (*e.g.*, de Vargas *et al.*, 2015 ; Debroas *et al.*, 2017).

Ainsi, en multipliant 500 000 fois le nombre de séquences produites par jour et diminuant d'au moins 250 000 fois leurs prix, entre 1996 et 2015 (Reuter *et al.*, 2015), l'utilisation du HTS est devenue une référence pour l'exploration des eucaryotes unicellulaires par la mise en place d'approches variées (*i.e.*, approches méta « -omics » telles que le métabarcoding (Taib *et al.*, 2013 ; de Vargas *et al.*, 2015 ; Debros *et al.*, 2017 ; Lepère *et al.*, 2019), la métatranscriptomique (Poretsky *et al.*, 2009 ; Grossmann *et al.*, 2016 ; Carradec *et al.*, 2018 ; Hu *et al.*, 2018 ; Caputi *et al.*, 2019 ; Trench-Fiol et Fink, 2020) et la métagénomique (Delmont *et al.*, 2022)).

A. Métabarcoding

i. Principe

Le métabarcoding est l'approche la plus répandue pour l'analyse de la diversité microbienne car elle permet une évaluation plus exhaustive des espèces présentes dans un écosystème grâce à l'étude d'un marqueur unique (Geisen *et al.*, 2019 ; Jamy *et al.*, 2019).

Malgré l'existence de plusieurs marqueurs phylogénétiques, le gène codant pour la petite sous-unité de l'ARNr (18S pour les eucaryotes) est resté le plus utilisé pour les études de phylogénie moléculaire du fait de son ubiquité (présence chez tous les organismes eucaryotes), de sa structure (alternance de régions conservées et variables (Gutell *et al.*, 1994)) (**Fig. 14**) et du nombre important de séquences dans les bases de données. Son amplification est possible par une réaction en chaîne par polymérase (PCR) grâce à l'utilisation d'amorces universelles, spécifiques et généralement positionnées dans les régions conservées (**Fig. 14**). L'amplicon ainsi obtenu contient par conséquent en théorie un mélange de toutes les séquences représentatives des espèces présentes dans l'écosystème. Après séquençage, l'assignation taxonomique des espèces se fait par similarité de séquences par rapport à des bases de données de référence telles que PR² (Guillou *et al.*, 2013) et Silva (Pruesse *et al.*, 2007). Les évolutions et optimisations récentes du métabarcoding et le développement d'outils de bio-informatique pour l'analyse des données de séquençage ont permis une plus large utilisation de cette méthode pour l'étude de la diversité des eucaryotes unicellulaires (Geisen *et al.*, 2019).

Les gènes codant pour l'ARNr contiennent neuf régions variables (V1 à V9) (**Fig. 14**). Par rapport aux régions conservées, celles-ci montrent un grand nombre de variations entre espèces du fait de l'accumulation de mutations au cours du temps (Zuckerandl et Pauling, 1965 ; Douzery *et al.*, 2006).

Pour les eucaryotes, la comparaison des séquences d'ADNr 18S a permis d'identifier les régions variables les plus propices aux études phylogénétiques, c'est à dire avec suffisamment de variations pour différencier les espèces, mais aussi encadrées de régions très conservées au niveau de leurs extrémités pour la détermination d'amorces universelles. Ainsi, cette démarche a révélé deux régions préférentielles, la V9 longue d'environ 200 pb (de Vargas *et al.*, 2015) et la V4, longue d'environ 400 pb (Debroas *et al.*, 2017). La région V4, de plus grande taille, a comme intérêt d'augmenter la résolution phylogénétique. Ces régions peuvent être séquencées par la technique Illumina (*via* les méthodes MiSeq et HiSeq) qui est la technologie HTS la plus utilisée actuellement en métabarcoding du fait de sa robustesse (peu d'erreurs de séquençage : taux proche de 0,1 %), sa capacité (milliards de bases) et son coût (Geisen *et al.*, 2019).

Plus récemment, les technologies HTS de troisième génération, telles que celles proposées par Pacific Biosciences (PacBio) et Oxford Nanopore, ont rendu possible le séquençage de fragments d'ADN de 10 kb à 20 kb en moyenne mais de faible qualité initiale. Leur utilisation en métabarcoding pourrait toutefois s'avérer prometteuse. En effet, le principe développé par PacBio, se distinguant par une circularisation des molécules d'ADN suivie de leur séquençage à plusieurs reprises, génère des séquences consensus de bonne qualité (Jamy *et al.*, 2022). Son utilisation pour l'étude de la diversité des eucaryotes unicellulaires s'est récemment révélée pertinente, l'augmentation de la taille des séquences générant un plus fort signal phylogénétique et donc permettant une meilleure identification (Orr *et al.*, 2018 ; Tedersoo et Anslan, 2019 ; Jamy *et al.*, 2022).

ii. De la diversité taxonomique aux fonctions

Les approches de métabarcoding pour l'étude de la diversité taxonomique des eucaryotes unicellulaires se sont révélées incontournables en milieu aquatique. En milieu marin, des expéditions telles que TARA Océans ou Malaspina-2010 ont permis l'obtention de plus d'un millier d'échantillons, entraînant ainsi l'émergence de nombreuses études utilisant le métabarcoding comme moyen d'explorer la diversité des eucaryotes unicellulaires (Bork *et al.*, 2015 ; Massana *et al.*, 2015 ; Vernet *et al.*, 2021).

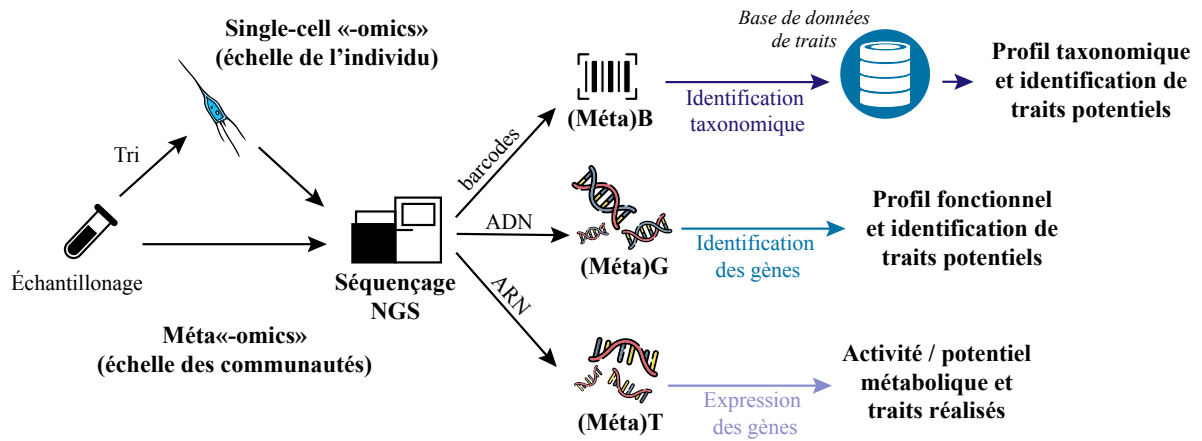


Figure 15 : Utilisation des approches «-omics» pour l’exploration de la diversité fonctionnelle (modifiée d’après Martini *et al.*, 2021).

De tels efforts ont permis, en ciblant le gène de la petite sous-unité ribosomique, la découverte d'une diversité considérable et insoupçonnée d'eucaryotes unicellulaires ainsi que la re-description de la phylogénie des eucaryotes, en découvrant de nombreux groupes dont nous ne soupçonnions pas l'existence (del Campo *et al.*, 2014 ; de Vargas *et al.*, 2015 ; Le Bescot *et al.*, 2015). L'intégration des données environnementales a, par ailleurs, permis d'ajouter une dimension fonctionnelle à l'appréciation de cette diversité en décrivant le rôle déterminant de la température et de la salinité dans la distribution des espèces (Ibarbalz *et al.*, 2019 ; Singer *et al.*, 2021 ; Jamy *et al.*, 2022). L'analyse de nombreux échantillons collectés dans l'intégralité des océans du globe a, par exemple, permis l'identification des facteurs abiotiques influençant la biogéographie de certains eucaryotes unicellulaires mixotrophes (Faure *et al.*, 2019).

En milieu lacustre, les approches de métabarcoding ont également contribué à l'exploration de la diversité des eucaryotes unicellulaires (Debroas *et al.*, 2017), apparaissant bien différente de celle préalablement décrite en milieu marin et variant d'un lac à l'autre (Singer *et al.*, 2021 ; Jamy *et al.*, 2022). L'utilisation d'amorces spécifiques ciblant des marqueurs phylogénétiques plus adaptés à l'étude de certains groupes d'eucaryotes a également permis une meilleure description de leur diversité. C'est le cas par exemple du gène *rbcL* pour l'exploration de la diversité des diatomées (Kezlya *et al.*, 2023 ; Turk Dermastia *et al.*, 2023) ou de la région V1-V3 permettant de cibler les microsporidies (Dubuffet *et al.*, 2021 ; Chauvet *et al.*, 2022).

Bien que ces approches aient permis de s'affranchir des processus de mise en culture pour l'étude de la diversité taxonomique, elles ne permettent pas, à elles seules, d'évaluer le rôle des espèces au sein des écosystèmes. Dans cette perspective, certains chercheurs combinent l'affiliation taxonomique, propre au métabarcoding, à une seconde annotation à l'aide de bases de données de traits reliant ainsi les espèces identifiées à leurs traits fonctionnels potentiels. D'après Violle *et al.* (2007), le terme « trait », dans ce genre de méthode, correspond à une caractéristique morphologique, physiologique ou phénotypique mesurable au niveau individuel (*i.e.*, de la cellule à l'organisme entier), sans référence à l'environnement ou à tout autre niveau d'organisation. Le terme « fonctionnel » caractérise, quant à lui, tout trait impactant la « fitness » d'un individu indirectement par ses effets sur la croissance, la reproduction et/ou la survie (Violle *et al.*, 2007 ; Martini *et al.*, 2021). En outre, on distingue les traits dit « potentiels », décrits dans la littérature généralement au niveau de l'espèce et qui couvrent idéalement une grande variété de conditions environnementales, et les traits « réalisés », effectivement mesurés *in situ* ou en laboratoire (**Fig. 15**) (Reu *et al.*, 2011).

Initialement développées pour l'exploration de la diversité fonctionnelle des organismes pluricellulaires en milieu terrestre (Violle *et al.*, 2007 ; Lamanna *et al.*, 2014), les études de traits morpho-physio-phénotypiques ont, en effet, progressivement été adaptées au monde microbien. Leur réalisation consiste à référencer, pour chaque espèce identifiée en métabarcoding, des traits, avec pour objectif de caractériser leur rôle au sein de l'écosystème. Ces traits peuvent alors servir à la classification en groupes fonctionnels, en se basant par exemple sur la capacité des microorganismes à réaliser la phagotrophie et/ou la photosynthèse (**Fig. 13** ; Mitra *et al.*, 2023). Par ailleurs, différents traits peuvent varier de manière conjointe, faisant alors émerger le concept de « compromis » de traits. En effet, les traits liés à la croissance, à la reproduction et à la survie sont souvent quantitativement corrélés, et ces relations, définies comme des « compromis », peuvent fournir une vision globale des stratégies écologiques des organismes étudiés (Litchman *et al.*, 2007 ; Litchman et Klausmeier, 2008 ; Maire *et al.*, 2015). En couplant cette étude à l'utilisation d'algorithmes de clusterisation non supervisée, certains chercheurs ont alors réussi à définir des groupes fonctionnels de manière automatique (Maire *et al.*, 2015), recouvrant la diversité des stratégies trophiques d'eucaryotes unicellulaires en milieu marin (Ramond *et al.*, 2019). Toutefois, ces méthodes nécessitent des bases de traits exhaustives et leur mise en place se révèle souvent laborieuse (Escalas *et al.*, 2019 ; Martini *et al.*, 2021). Enfin, peu d'entre elles sont réalisées pour l'exploration de la diversité fonctionnelle des eucaryotes unicellulaires et encore moins en milieu lacustre.

B. Métatranscriptomique

i. Principe

L'exploration des communautés microbiennes à travers l'étude des ARN est utilisée depuis plusieurs décennies, notamment *via* l'utilisation de méthodes ciblées telles que la RT-PCR quantitative ou les technologies « microarray » (Foley *et al.*, 1993 ; Schena *et al.*, 1995 ; Zhu *et al.*, 2005). En ciblant la petite sous-unité de l'ARNr 18S, ces méthodes ont notamment permis l'évaluation de l'activité des communautés microbiennes eucaryotes dans divers écosystèmes aquatiques (Zhu *et al.*, 2005 ; Lepère *et al.*, 2016). Par ailleurs, appliquées à l'étude des ARNm, elles permettent de mesurer avec succès les variations de certains métabolismes tels que ceux impliqués dans la réponse à la carence en fer ainsi que dans la photosynthèse en milieu océanique (Parro *et al.*, 2007 ; Lommer *et al.*, 2012 ; Islam *et al.*, 2020 ; Caputi *et al.*, 2019).

Toutefois, ces approches, fortement limitées et/ou coûteuses, permettent uniquement d'évaluer l'expression de gènes spécifiques, sans permettre l'identification de nouvelles fonctions. La métatranscriptomique, basée sur le séquençage de l'ensemble des ARN extraits d'un échantillon, surmonte cette limitation en permettant l'évaluation sans a priori des métabolismes des communautés microbiennes présentes à un moment précis.

A l'instar du métabarcoding, c'est la technologie de séquençage Illumina qui est utilisée en majorité pour la métatranscriptomique notamment du fait de la quantité de lectures massives générées, permettant ainsi une profondeur de séquençage importante (paramètre crucial en métatranscriptomique). Les lectures brutes subissent ensuite un prétraitement, un contrôle qualité rigoureux (*e.g.*, suppression des adaptateurs, des bases erronées et des séquences de mauvaise qualité ou trop courtes) et une étape de ribodépletion. En effet, bien que les ARN ribosomaux soient éliminés avant le séquençage, leur abondance est si importante qu'une portion significative est tout de même séquencée.

Lors de l'étude d'environnements complexes (*e.g.*, échantillons environnementaux), les lectures nettoyées, caractérisées par un score de qualité suffisant, sont ensuite assemblées en contigs par chevauchement (assemblage *de novo*) afin de reconstruire au mieux l'ensemble des transcrits du jeu de données (Shakya *et al.*, 2019). Le métatranscriptome ainsi obtenu rassemble alors des transcrits pouvant être traduits en une ou plusieurs protéines (*e.g.*, cas des structures en opéron).

Ces transcrits et/ou protéines peuvent ensuite être annotés fonctionnellement par comparaison à des bases de données de référence telles que KEGG, EggNOG ou UniProt (Boutet *et al.*, 2007 ; Kanehisa *et al.*, 2016 ; Huerta-Cepas *et al.*, 2019). Ces bases de données jouent un rôle clef dans l'étude de la diversité fonctionnelle des communautés microbiennes car elles permettent d'établir des liens directs entre les gènes et les voies métaboliques voire même entre les gènes et certaines fonctions ou stratégies trophiques (il s'agit alors de traits fonctionnels réalisés ; **Fig. 15**).

L'affiliation taxonomique des transcrits est également possible mais demeure bien moins fiable que celle réalisée à partir des ARN ribosomaux. Tout comme pour le processus d'annotation fonctionnelle, les bases de données et outils utilisés pour l'affiliation des eucaryotes unicellulaires sont identiques à ceux utilisés lors d'études de métagénomique (*e.g.*, MetaEuk (Keeling *et al.*, 2014 ; L. K. Johnson *et al.*, 2019 ; Levy Karin *et al.*, 2020)).

Enfin, la comparaison des quantités de transcrits à travers les différents échantillons, aussi appelée analyse d'expression différentielle, permet d'examiner les différents niveaux d'expression sous différentes conditions et/ou d'observer les dynamiques fonctionnelles au cours du temps (Robinson *et al.*, 2010 ; Love *et al.*, 2014).

ii. Mesure de la diversité fonctionnelle

Les études de métatranscriptomique ont gagné en popularité ces dernières années (Shakya *et al.*, 2019) notamment grâce aux avancées en matière de capacité de séquençage et à l'amélioration des bases de données de référence (Keeling *et al.*, 2014 ; Kanehisa *et al.*, 2016 ; Carradec *et al.*, 2018 ; Huerta-Cepas *et al.*, 2019). Récemment, Carradec *et al.* (2018) ont capturé, à travers une étude globale de métatranscriptomique, une quantité colossale de transcrits représentant les profils d'expression du plancton océanique (*i.e.*, eucaryotes unicellulaires et zooplancton). Regroupant environ 116 millions d'unigènes, ce catalogue représente, à ce jour, la collection la plus exhaustive de transcrits issus d'écosystèmes aquatiques. L'apport significatif de telles recherches permet ainsi de dévoiler les fonctions exprimées par le plancton marin eucaryote et d'évaluer leur biogéographie fonctionnelle (Carradec *et al.*, 2018).

Les approches de métatranscriptomique permettent également d'explorer la diversité des gènes transcrits dans plusieurs échantillons simultanément et ainsi de comparer leurs expressions différentielles en fonction des conditions échantillonnées (Robinson *et al.*, 2010 ; Love *et al.*, 2014). Cette approche s'est avérée particulièrement pertinente en milieu marin en permettant de décrire les variations significatives des métabolismes des eucaryotes unicellulaires en fonction de la profondeur d'échantillonnage (Hu *et al.*, 2018) et du cycle jour/nuit (Poretsky *et al.*, 2009).

Enfin, bien que les études utilisant les approches de métatranscriptomique pour l'exploration de la diversité fonctionnelle des eucaryotes unicellulaires soient restées rares en milieu lacustre (Grossmann *et al.*, 2016 ; Trench-Fiol et Fink, 2020 ; Li *et al.*, 2022), certaines ont sensiblement enrichi les descriptions du fonctionnement de ces écosystèmes. Par exemple, en examinant la diversité des transcrits dans des échantillons provenant de 21 lacs européens, Grossmann *et al.* (2016) ont constaté qu'en dépit d'une β -diversité élevée, les voies métaboliques enregistrées semblaient bien plus stables entre les lacs, suggérant ainsi l'existence d'une potentielle redondance fonctionnelle préservant le fonctionnement des écosystèmes lacustres.

iii. Réseaux de similarité de séquences protéiques

Bien que les approches de métatranscriptomique se soient révélées convaincantes pour l'étude de stratégies trophiques distinctes, telles que la phototrophie (Trench-Fiol et Fink, 2020) et la phagotrophie (Labarre *et al.*, 2020 ; Massana *et al.*, 2021), ainsi que dans l'exploration des voies métaboliques induites lors d'épidémies d'espèces parasites (Li *et al.*, 2022), la quantité colossale de données générée demeure extrêmement difficile à analyser. Par ailleurs, établir les connexions entre les données d'expression de gènes et les fonctions au sein des écosystèmes reste une tâche tout aussi délicate. En effet, la majorité des études de métatranscriptomique enregistre des proportions d'affiliations relativement faibles (~40 à 70%) ne permettant qu'une utilisation partielle des données séquencées (Grossmann *et al.*, 2016 ; Carradec *et al.*, 2018 ; Trench-Fiol et Fink, 2020). De plus, il n'existe que très peu de gènes marqueurs de différents métabolismes décrivant la biologie des eucaryotes unicellulaires.

L'utilisation d'approches basées sur la construction de réseaux de similarité semble être une piste intéressante pour l'exploration des données de métatranscriptomique (Cheng *et al.*, 2014 ; Forster *et al.*, 2015 ; Bittner, 2018). En regroupant les séquences protéiques en fonction de leur similarité, ces réseaux permettent d'étudier les caractéristiques métaboliques de différents organismes partageant des fonctions similaires, et ce, tout en intégrant dans l'analyse globale des séquences non annotées fonctionnellement et/ou non affiliées taxonomiquement. Seules quelques études utilisant cette stratégie ont déjà été conduites dans un contexte fonctionnel. Elles ont, par exemple, révélé des ensembles de gènes caractéristiques de traits fonctionnels spécifiques, tels que la toxicité et la capacité de symbiose des Dinoflagellata (Meng *et al.*, 2018), ou adaptés à des conditions environnementales particulières (Faure *et al.*, 2021).

C. Métagénomique

i. Principe

Les études basées sur la métagénomique, définie ici comme le séquençage de l'ADN total extrait d'un échantillon environnemental, représentent également une valeur ajoutée pour l'investigation du lien diversité-fonction (**Fig. 15**). En effet, l'avènement récent des génomes assemblés à partir de métagénomiques (Metagenome-Assembled Genomes : MAGs) a largement contribué à accroître les connaissances quant à la diversité taxonomique et fonctionnelle des communautés microbiennes.

Les lectures obtenues après séquençage et filtrage qualité sont, dans un premier temps, assemblées en contigs. Ces contigs (rassemblant généralement quelques milliers de paires de bases), sont ensuite regroupés selon des critères tels que leurs profils d'abondance à travers les métagénomomes (*i.e.*, les contigs ayant une distribution similaire dans les échantillons) et leur teneur en bases guanine et cytosine (*i.e.*, GC%). Ce regroupement permet alors d'identifier des groupes de contigs (on parle de bins) provenant de la même population et donc probablement de la même lignée (Nielsen *et al.*, 2014 ; Parks *et al.*, 2015 ; Delmont *et al.*, 2018 ; Tully *et al.*, 2018). L'assemblage de chaque bin peut ensuite donner un MAG plus ou moins complet pouvant être annoté taxonomiquement et/ou fonctionnellement par comparaison à des génomes de référence. La complétude d'un MAG est estimée sur la base de la présence d'un ensemble de gènes marqueurs ubiquistes. Un MAG dont la complétude est évaluée en dessous du seuil généralement établi à 50% est exclu dans la plupart des études (Parks *et al.*, 2015).

ii. Exploration des métagénomomes

Les approches de métagénomique ont été utilisées, jusqu'à récemment, en majorité pour l'étude des procaryotes. Elles ont, par exemple, montré lors d'une étude sur les eaux souterraines polycontaminées, que les Burkholderiales étaient capables de dégrader divers contaminants, ce qui pourrait être important pour des aspects de bioremédiation (Hauptfeld *et al.*, 2022). Plus récemment, d'autres résultats obtenus à l'aide de la métagénomique ont permis de découvrir de nouveaux microorganismes et diverses stratégies métaboliques au sein des communautés bathypélagiques océaniques (Acinas *et al.*, 2021).

L'ensemble des données de métagénomique marine accumulé jusqu'à aujourd'hui représente une quantité colossale de références s'avérant extrêmement utile pour l'exploration des génomes eucaryotes. Elles ont d'ailleurs récemment été combinées afin d'établir une base de données de référence (« MarineMetagenomeDB »), constituant ainsi une ressource précieuse pour la mise en œuvre de nouvelles études de métagénomique (Nata'ala *et al.*, 2022).

Bien que la taille et la complexité des génomes eucaryotes représentent un défi majeur pour leur assemblage *de novo*, la métagénomique appliquée aux eucaryotes unicellulaires est progressivement facilitée. En effet, le développement récent de nouvelles méthodologies et l'accès à la quantité colossale de données multi « -omics », notamment *via* le jeu de données de l'expédition TARA Océans, ont permis de lever certains verrous techniques liés aux génomes eucaryotes.

Ces progrès, ouvrant ainsi de nouvelles perspectives d'études, démontrent l'intérêt et la faisabilité de cette approche (Delmont *et al.*, 2015, 2022 ; Duncan *et al.*, 2022). Delmont *et al.* (2022) ont, par exemple, récemment réussi à reconstruire et valider manuellement plus de 700 génomes environnementaux d'eucaryotes abondants et répandus en milieu océanique (variant de 10 Mbp à 1,3 Gbp).

III. Objectifs de la thèse et organisation du mémoire

Ce premier chapitre de synthèse bibliographique fait état des connaissances afférentes à la diversité taxonomique des eucaryotes unicellulaires et introduit l'intérêt porté à l'étude de la diversité des rôles et des fonctions qu'ils occupent au sein des écosystèmes aquatiques. En effet, appréhender la diversité fonctionnelle des eucaryotes unicellulaires constitue un défi majeur en écologie aquatique.

Dans ce contexte, l'avènement des technologies de séquençage haut débit a fait émerger ces dernières années plusieurs approches, dites « -omics », permettant l'exploration des communautés d'eucaryotes unicellulaires, et ce, indépendamment des approches de mise en culture. Elles ont ainsi permis de mettre en évidence une diversité considérable d'eucaryotes unicellulaires tout en contribuant significativement à la compréhension du lien diversité-fonction essentiellement en milieu marin. Toutefois, l'utilisation de ce genre d'approches est restée peu fréquente pour l'étude des écosystèmes lacustres nonobstant le fait qu'ils abritent une diversité colossale d'eucaryotes unicellulaires.

Par ailleurs, les parasites représentent un maillon de la diversité des eucaryotes unicellulaires encore mal appréhendé dans ces écosystèmes. En effet, bien que certains soient relativement bien documentés, le rôle de la majorité d'entre eux dans la structuration des communautés et le fonctionnement des écosystèmes ainsi que les différents métabolismes mis en place lors de l'établissement des relations hôte-parasites restent à être élucidés.

Dans ce contexte, plusieurs études ont été menées au cours de cette thèse et sont exposées dans ce manuscrit selon l'organisation suivante :

- Le **chapitre II** est une étude publiée en 2023 dans **Environmental Microbiology** couplant une approche basée sur les traits morpho-physio-phénotypiques à de la métatranscriptomique afin d'explorer la diversité taxonomique et fonctionnelle de différents groupes d'eucaryotes unicellulaires de stratégies trophiques distinctes au sein d'un lac méromictique. L'effet de différents paramètres de forçage tels que la présence ou non d'oxygène, les fractions de taille ou les saisons sur la distribution de ces groupes ainsi que sur l'expression de leurs gènes est également exploré.
- Le **chapitre III** expose les résultats d'une étude explorant les spécificités métaboliques de différents groupes fonctionnels d'eucaryotes. En explorant le jeu de données de métatranscriptomique introduit dans le chapitre II par le biais d'une approche basée sur la construction de réseaux de similarité de séquences protéiques, nous avons étudié les caractéristiques métaboliques de phyla éloignés partageant des fonctions écologiques similaires. Cette étude nous a ainsi permis d'explorer plus spécifiquement les signatures protéiques de différents groupes de parasites lacustres. Ce chapitre sera prochainement soumis à **ISME Journal**.
- Le **chapitre IV** correspond à une étude publiée en 2022 dans **Microbial Ecology** mettant en évidence, au travers de deux approches, une nouvelle association hôte-parasites caractérisée par une prévalence importante entre une microsporidie et un rotifère en milieu lacustre.

Enfin, une discussion générale (**chapitre V**) permet d'aborder et d'intégrer l'ensemble des résultats obtenus au cours de ce travail de thèse tout en introduisant différents travaux annexes et futures perspectives de recherches pour l'exploration de la diversité fonctionnelle des eucaryotes unicellulaires et des interactions hôte-parasites dans les réseaux trophiques lacustres.

Une étude supplémentaire, publiée dans **Limnology and Oceanography** en 2023 est également disponible en **annexe I**, présentant et comparant la diversité des microsporidies ainsi que leurs hôtes potentiels dans différents environnements marins et d'eau douce.

CHAPITRE II

ÉTUDE DE LA DIVERSITÉ FONCTIONNELLE DES
EUCARYOTES UNICELLULAIRES DANS UN LAC
MÉROMICTIQUE PAR UNE APPROCHE DE
MÉTATRANSCRIPTOMIQUE COUPLÉE À UNE ÉTUDE
BASÉE SUR LE RÉFÉRENCIEMENT DE TRAITS
MORPHO-PHYSIO-PHÉNOTYPIQUES

Contexte de l'étude

La diversité des eucaryotes unicellulaires, longtemps sous-estimée, a fait l'objet d'une réévaluation majeure au cours des deux dernières décennies, notamment grâce aux avancées techniques qu'ont connues les méthodes de séquençage haut-débit. Initialement mise en lumière en milieu marin, cette diversité est progressivement dévoilée dans les écosystèmes d'eau douce où elle rassemble un nombre considérable d'espèces réparties dans l'ensemble des branches de l'arbre des eucaryotes (Debroas *et al.*, 2017). Toutefois, bien que la taxonomie des eucaryotes unicellulaires soit de mieux en mieux documentée, leur rôle au sein des écosystèmes aquatiques demeure largement méconnu. Ainsi, l'étude de leur diversité fonctionnelle apparaît essentielle afin de mieux évaluer leur contribution aux flux de matière et d'énergie et finalement au fonctionnement des écosystèmes lacustres.

Le premier axe développé au cours de cette thèse concerne l'étude du lien diversité-fonction des eucaryotes unicellulaires dans un lac présentant des caractéristiques particulières : le lac Pavin. L'objectif général était de déchiffrer la structure et les stratégies trophiques des communautés eucaryotes en milieu lacustre en combinant des données de métabarcoding et de métatranscriptomique ainsi qu'une approche basée sur le référencement de traits morpho-physio-phénotypiques. Ainsi, nos objectifs principaux étaient de (i) décrire l'étendue de la diversité taxonomique ainsi que le potentiel métabolique des eucaryotes unicellulaires, (ii) élaborer une classification en groupes fonctionnels basée sur l'évaluation de la performance individuelle des espèces et (iii) étudier la distribution de ces groupes fonctionnels, ainsi que les expressions différentielles des gènes, en relation avec différents paramètres de forçage tels que la présence ou non d'oxygène, les fractions de tailles ou les saisons.

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

Arthur Monjot^{1°}, Gisèle Bronner¹, Damien Courtine¹, Corinne Cruaud², Corinne Da Silva³, Jean-Marc Aury³, Frederick Gavory³, Anne Moné¹, Agnès Vellet¹, Ivan Wawrzyniak¹, Jonathan Colombet¹, Hermine Billard¹, Didier Debroas¹, Cécile Lepère¹

1- CNRS, Laboratoire Microorganismes : Génome et Environnement, Université Clermont Auvergne, Clermont-Ferrand, F-63000, 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

Publié dans Environmental Microbiology – Novembre 2023

doi: 10.1111/1462-2920.16531

I. 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.

II. 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.

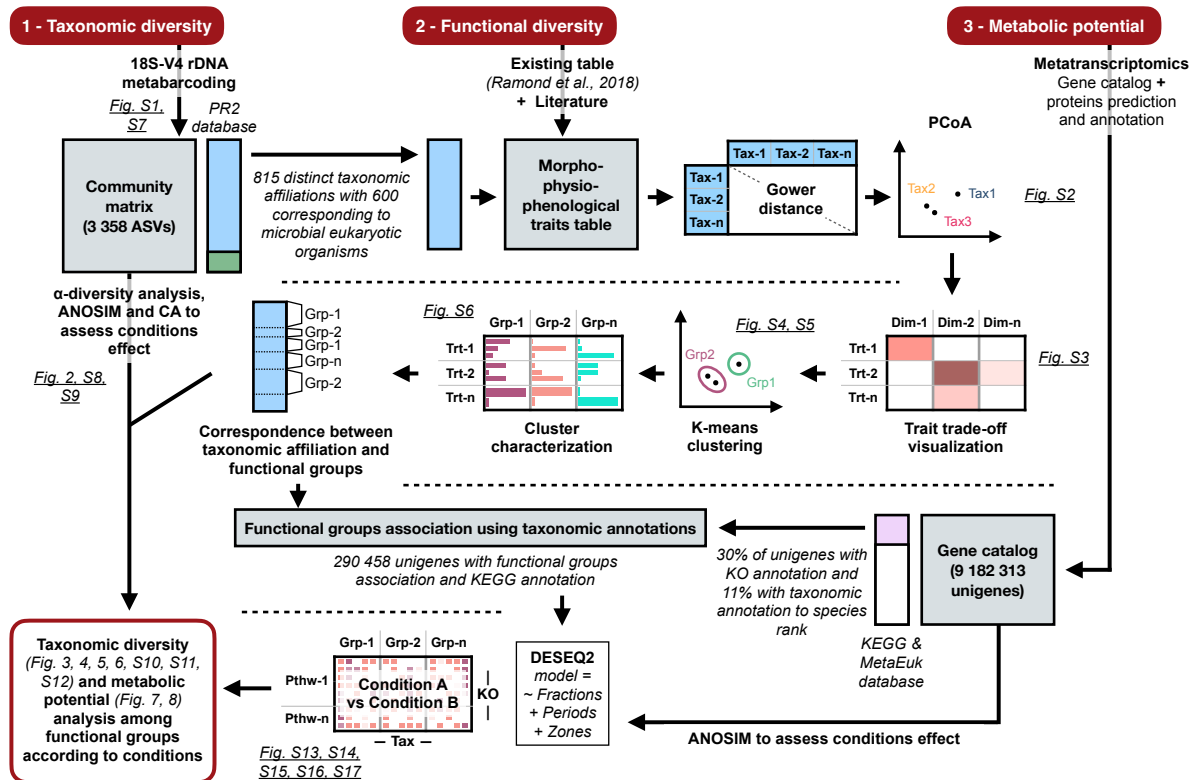


Figure 1: Workflow of the trait-based approach coupled to metatranscriptomic analysis.

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).

III. Materials and Methods

Figure 1 presents the workflow used in this study.

A. 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.

B. 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).

C. 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).

D. 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.

E. 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 reference labels.

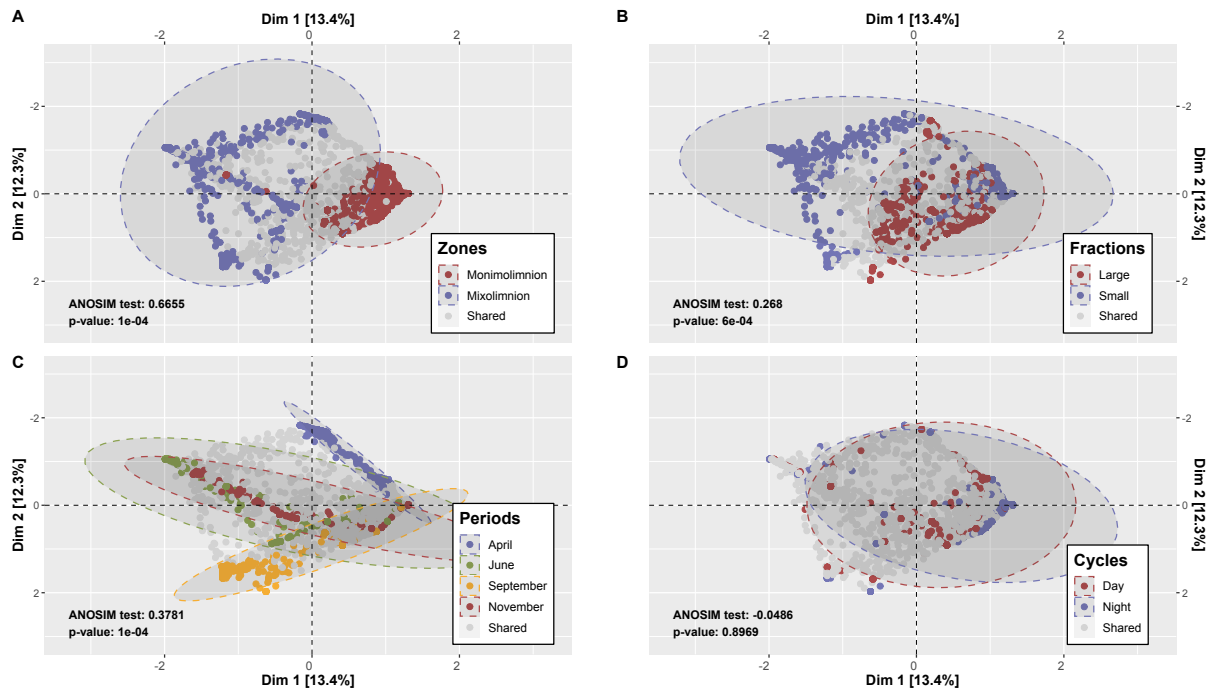


Figure 2: Two-dimensional Correspondence Analysis (CA) testing the correspondence between sampling conditions and the abundance of each ASV.

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 $\alpha = 5\%$).

F. 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.

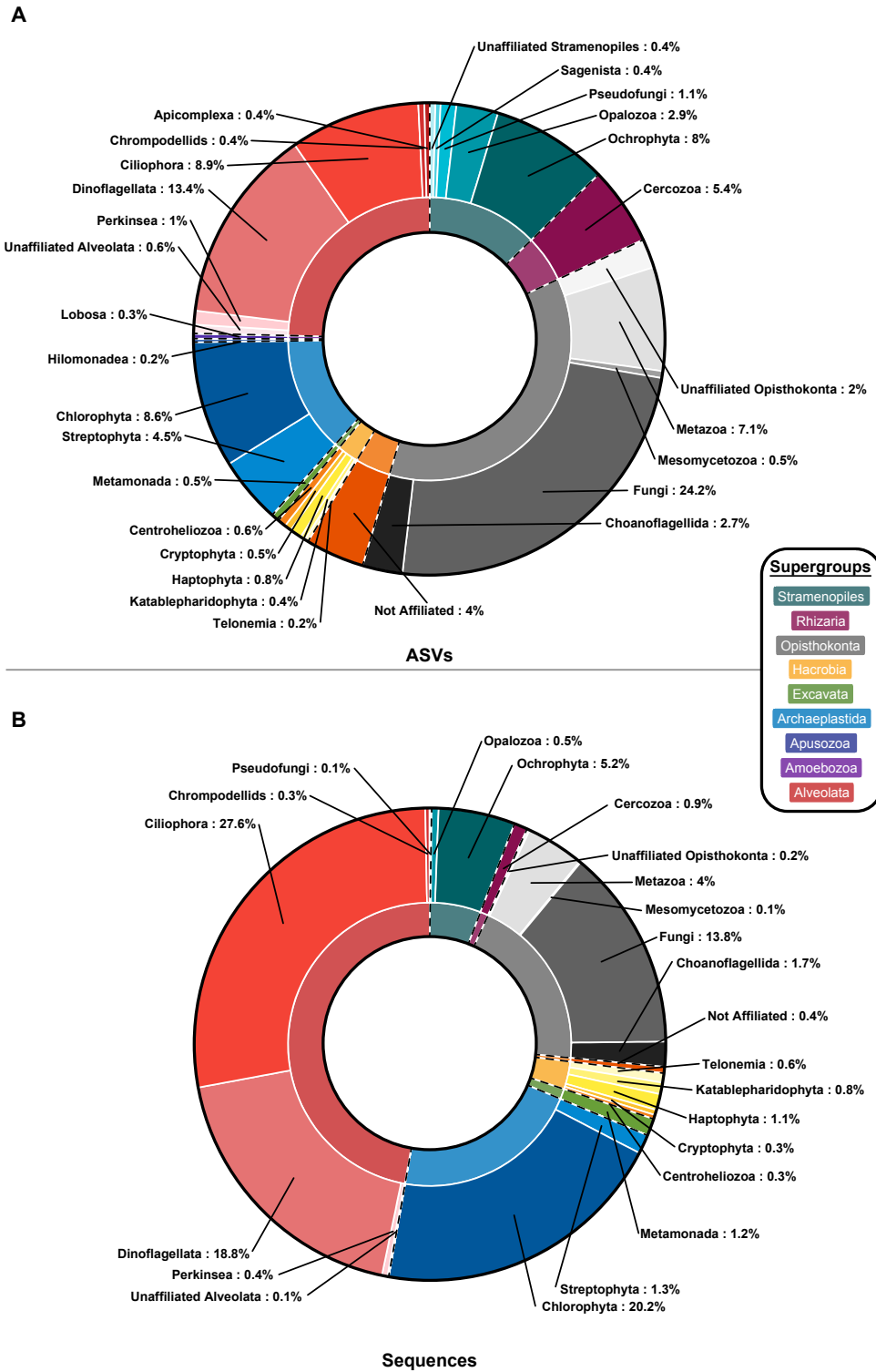


Figure 3: ASVs richness (A) and sequences abundance (B) of microbial eukaryotes in lake Pavin.

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

G. 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: \sim 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 .

IV. Results

A. 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.

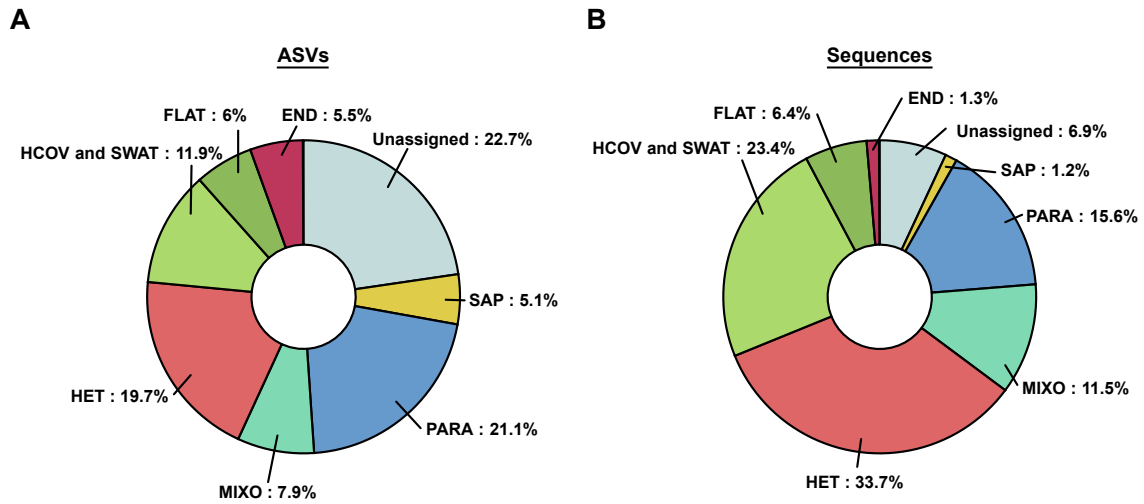


Figure 4: ASVs richness (A) and relative abundance (B) of the functional groups in lake Pavin.

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**).

B. Functional groups distribution

Using modalities of the 14 traits, 7 functional groups are defined and characterized (**Fig. S6**): Saprotophs (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**).

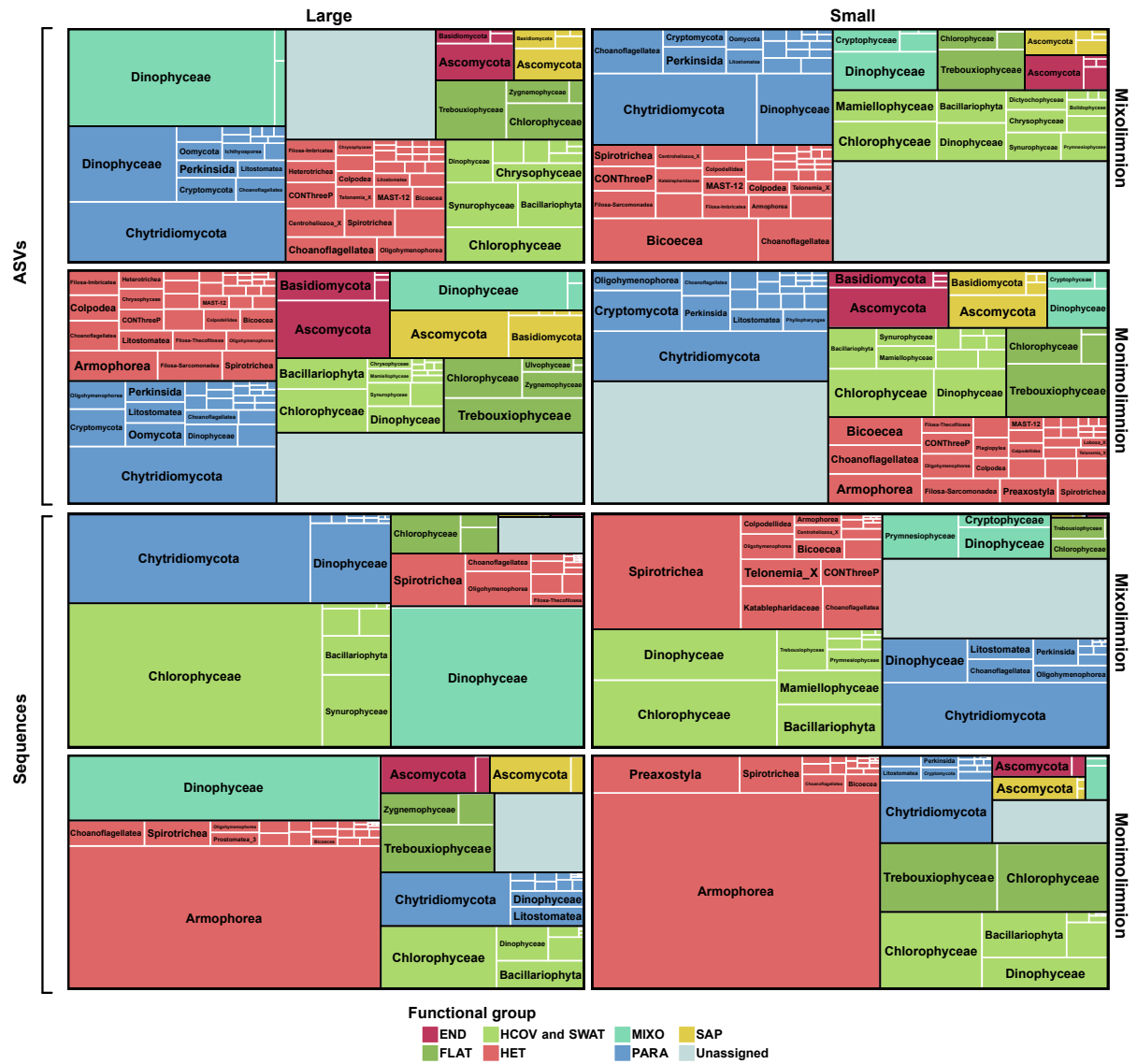


Figure 5: ASVs richness and relative abundance of the eukaryotic classes across each functional group according to sampling zone and size fraction in lake Pavin.

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.

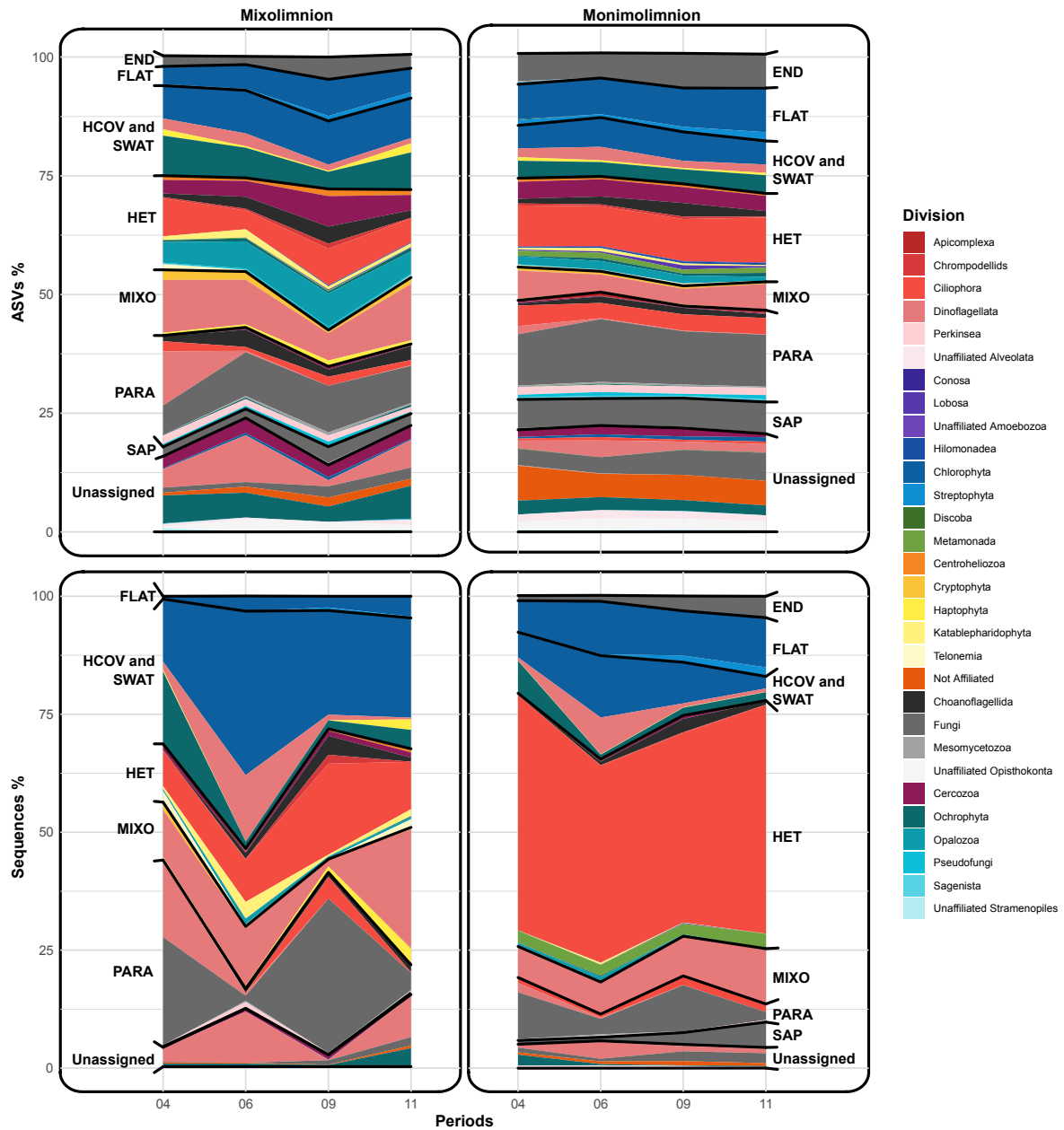


Figure 6: ASVs richness and relative abundance of the eukaryotic classes across each functional group according to sampling periods and zone in lake Pavin.

Richness and abundance of eukaryotic classes (4th 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.

C. 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**).

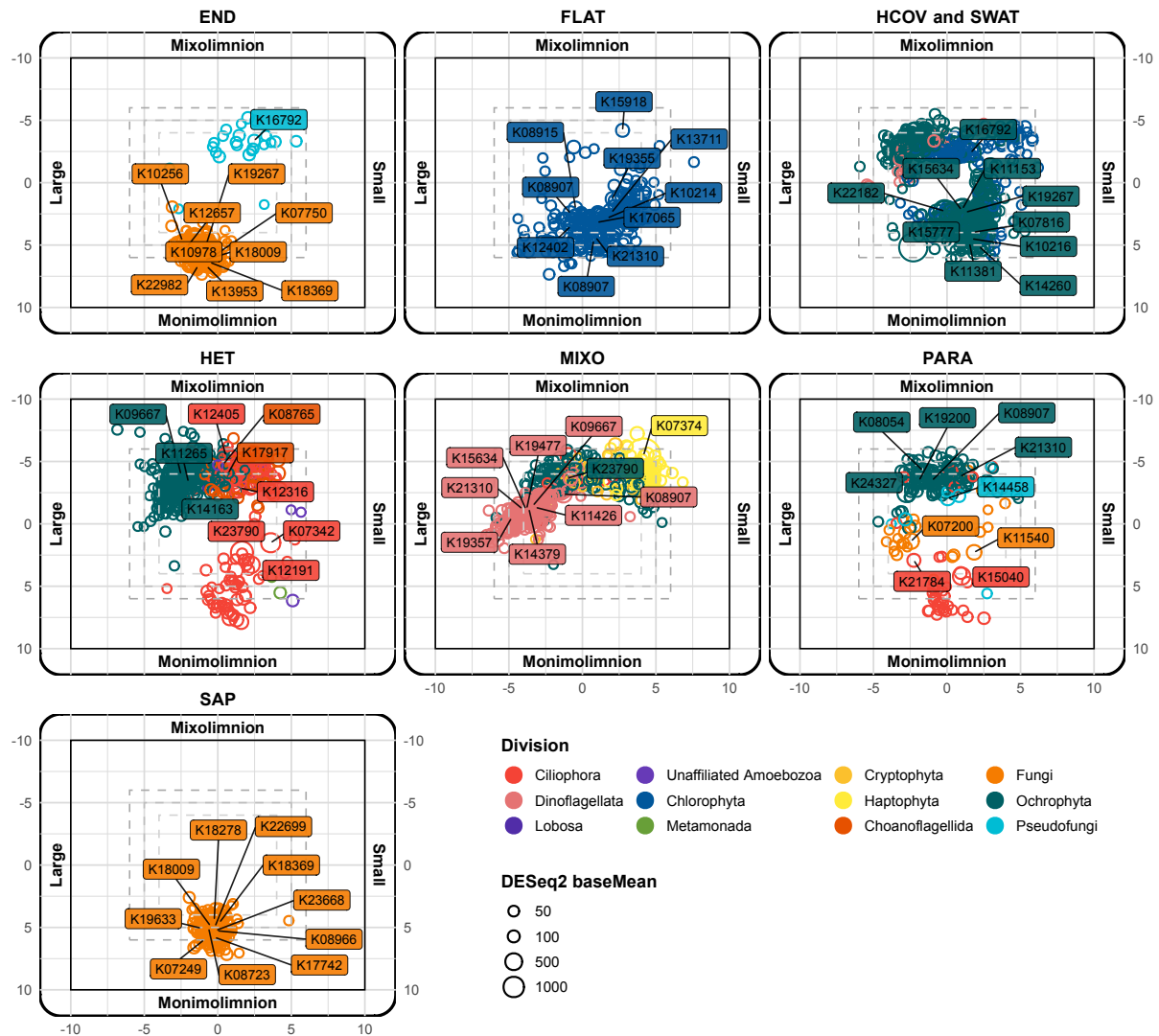


Figure 7: Differential expression analysis of genes across size fraction and sampling zone for each eukaryotic division and functional group.

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 Choanoflagellata (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 $>10 \mu\text{m}$ size fraction while glycolipids anabolisms pathways -related transcripts (*e.g.*, K09881 and K22831) are enriched in the $<10 \mu\text{m}$ size fraction.

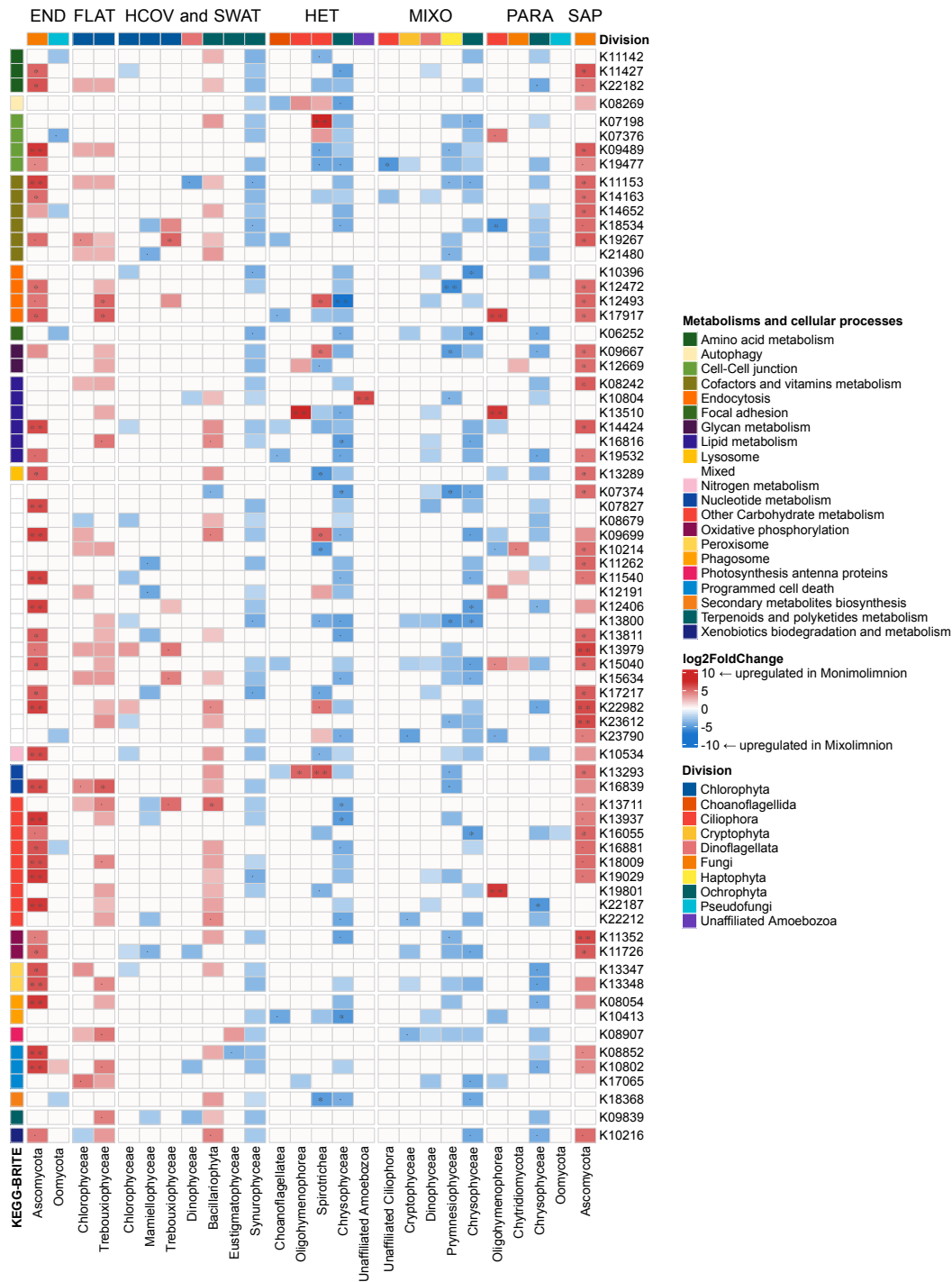


Figure 8: Heatmap representing differential expression of genes across sampling zones for each eukaryotic classes and functional group.

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 \log_2 -FoldChange > 2 are represented. Furthermore, only differentially expressed KO in at least 5 taxonomic classes among functional groups are displayed. Values of \log_2 -FoldChange > 4 are notified with « · », \log_2 -FoldChange > 5 with « * » while \log_2 -FoldChange > 6 with « ** ».

V. Discussion

A. 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 influenced 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.

B. 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 *Perididium* 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.

VI. 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* AtPI4K γ 3, 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 Communications*, **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. *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., Christmas, 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 zoospore 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.

VII. 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.

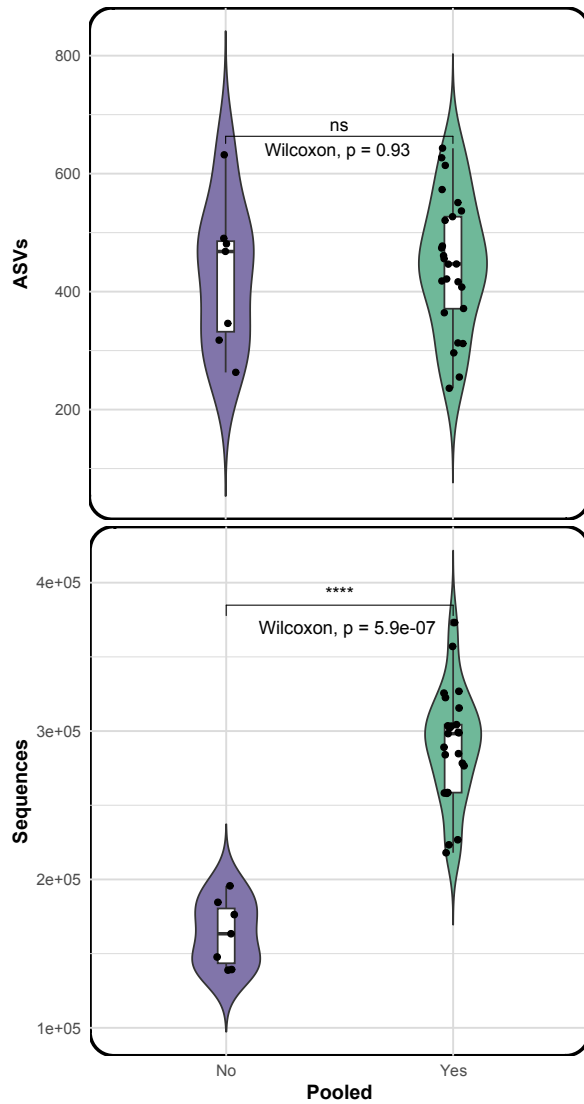


Figure S1: Assessment of the effect of the pooling procedure before rarefaction steps.

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).

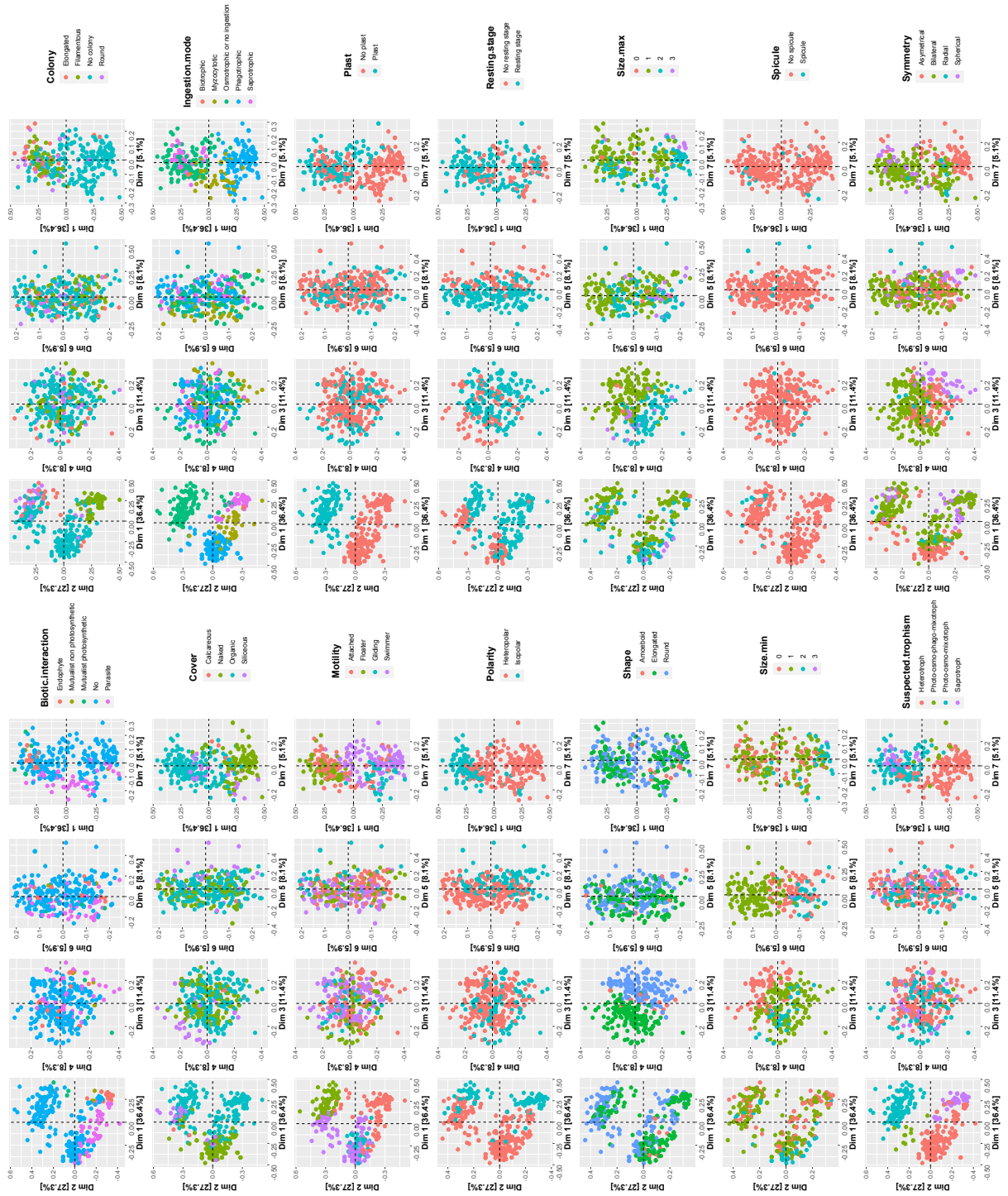


Figure S2: Principal Coordinate Analysis (PCoA) using the Gower distance between all taxonomic references of the functional traits table.

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.

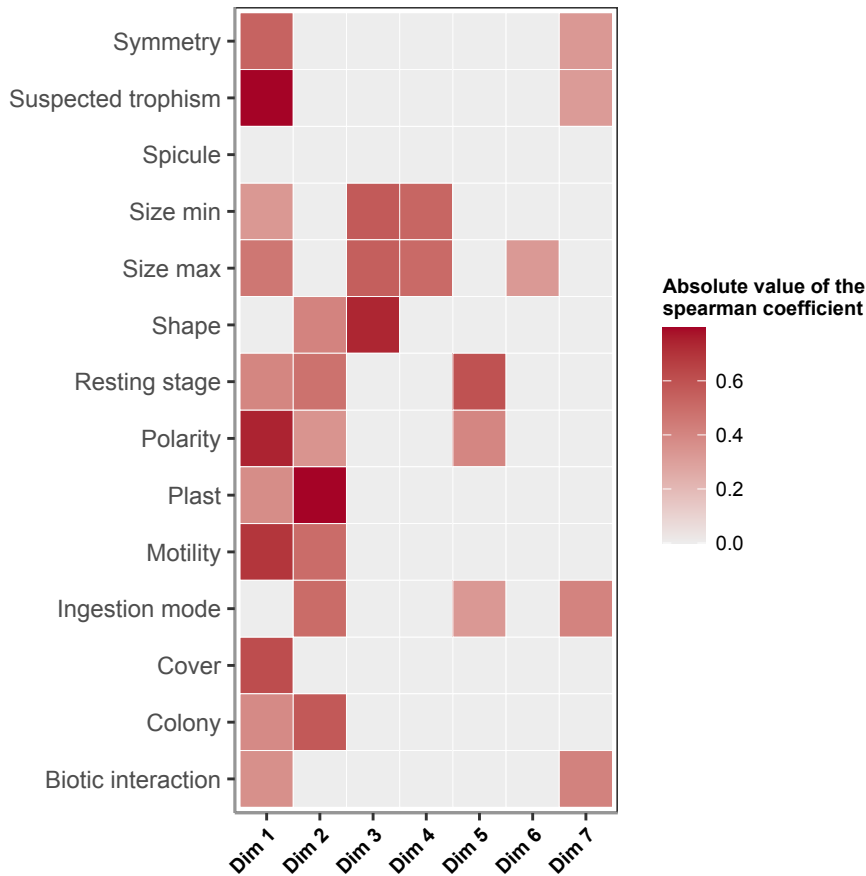


Figure S3: Heatmap representing trade-off between traits using PCoA scores.

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.

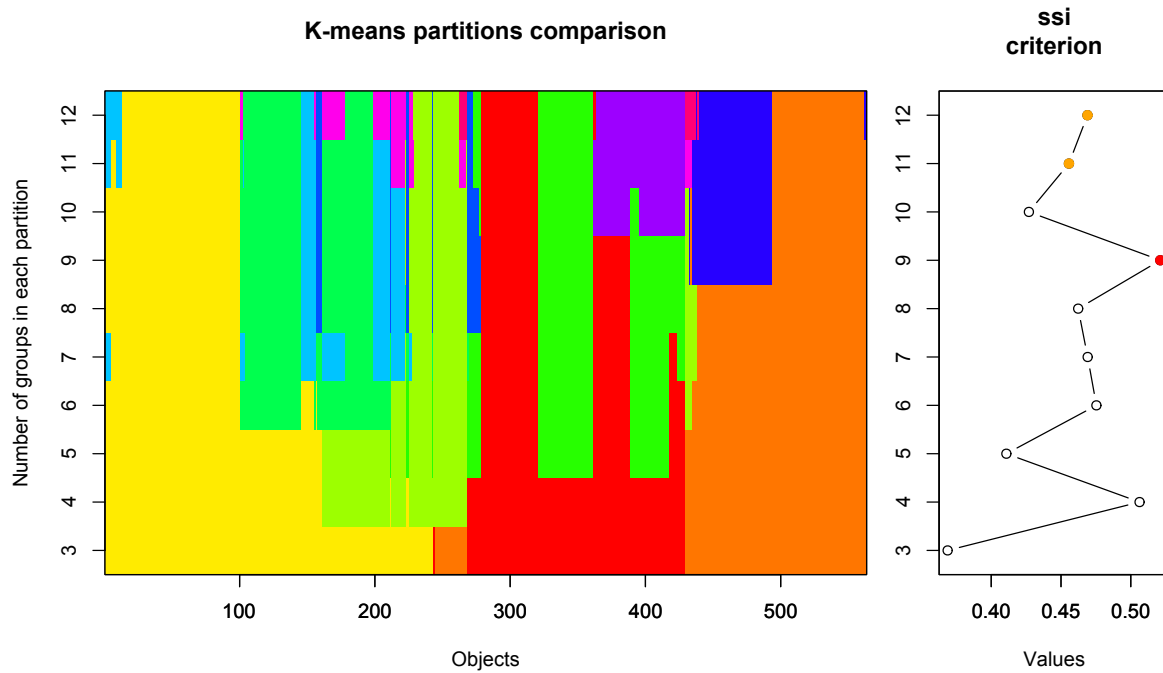


Figure S4: Best partitioning result from the Simple Structure Index (SSI) based on the two first dimensions of the PCoA.

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).

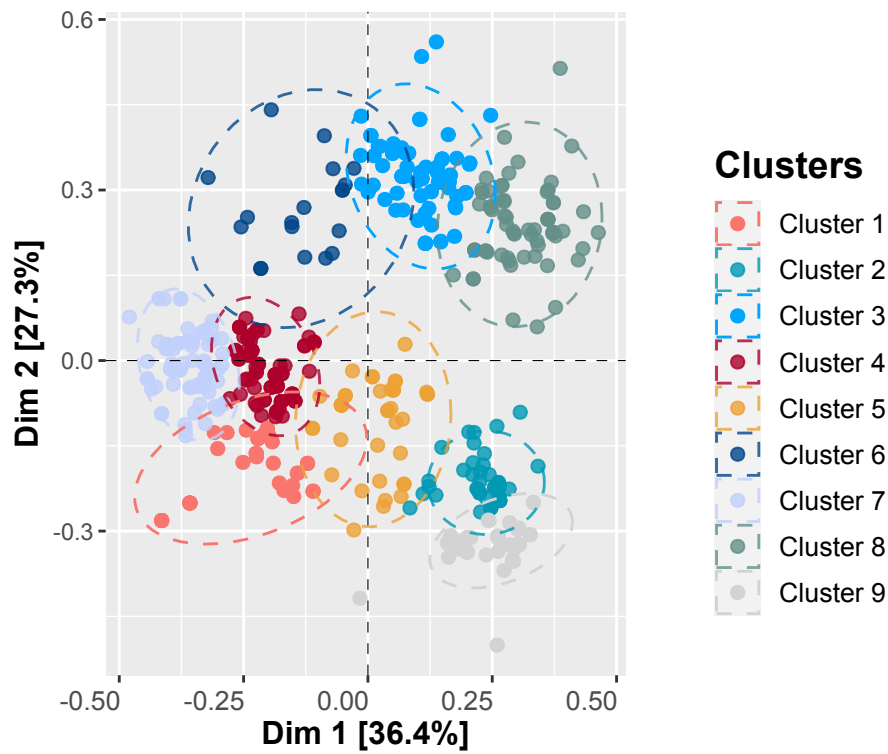


Figure S5: Principal Coordinate Analysis (PCoA) using the Gower distance between all taxonomic references of the functional traits table (groups visualization after K-means clustering).

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.

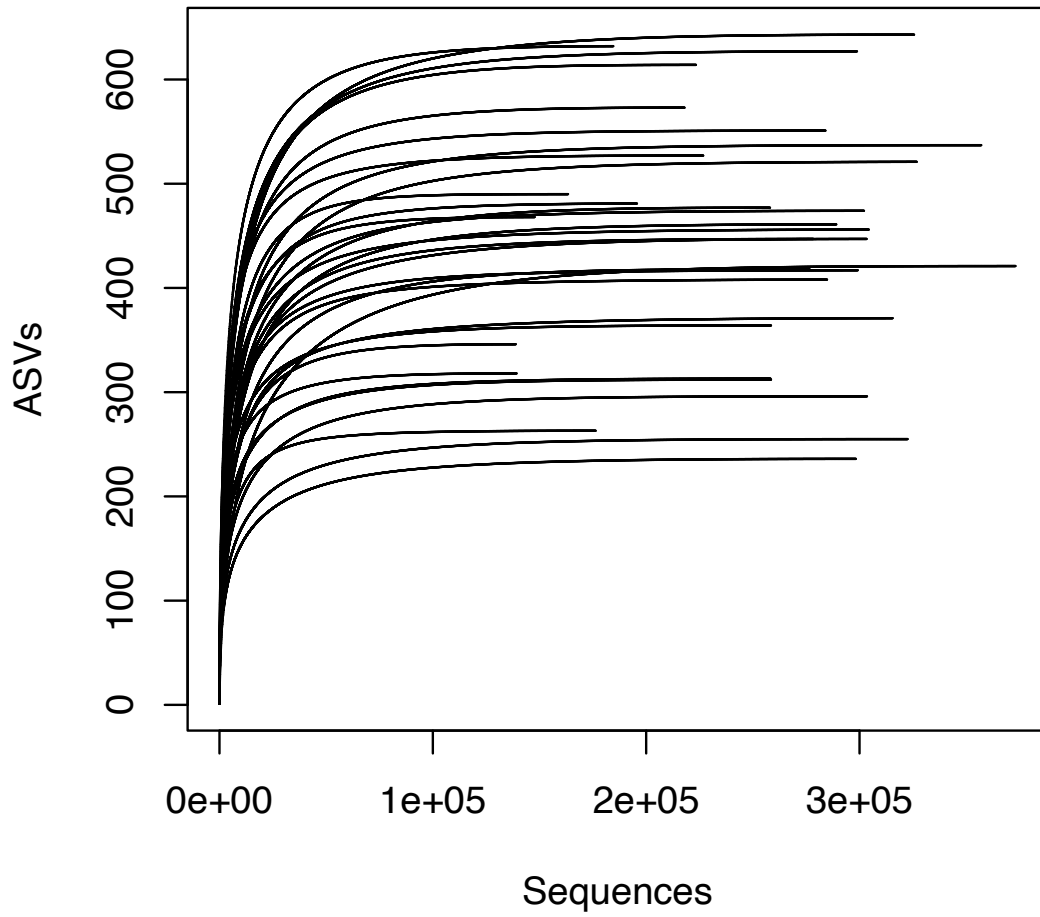


Figure S7: Rarefaction curves.

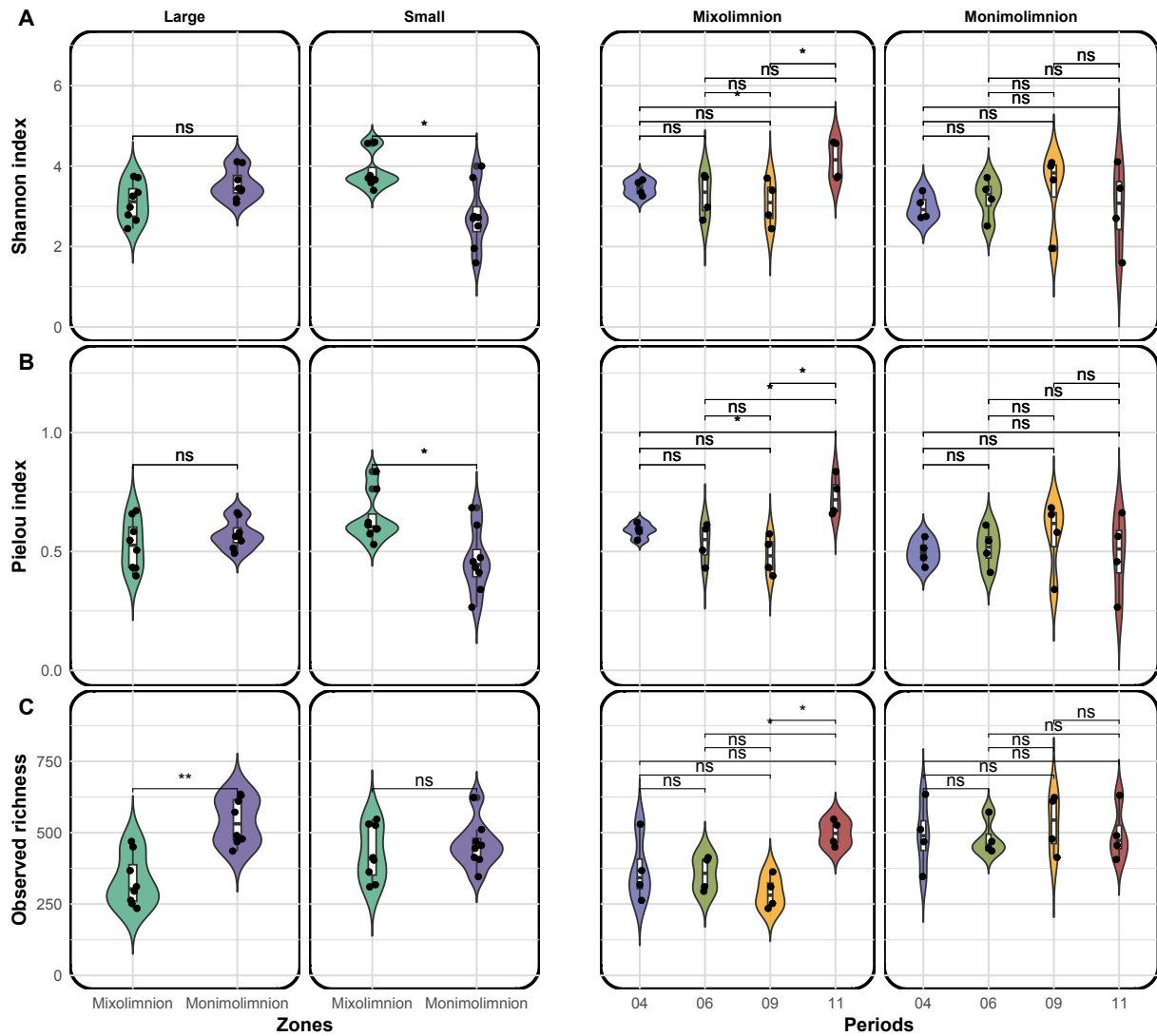


Figure S8: Diversity and equitability index (Shannon (A), Pielou (B)) and observed richness (C) of amplicons across sampling zones, size fractions and periods.

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

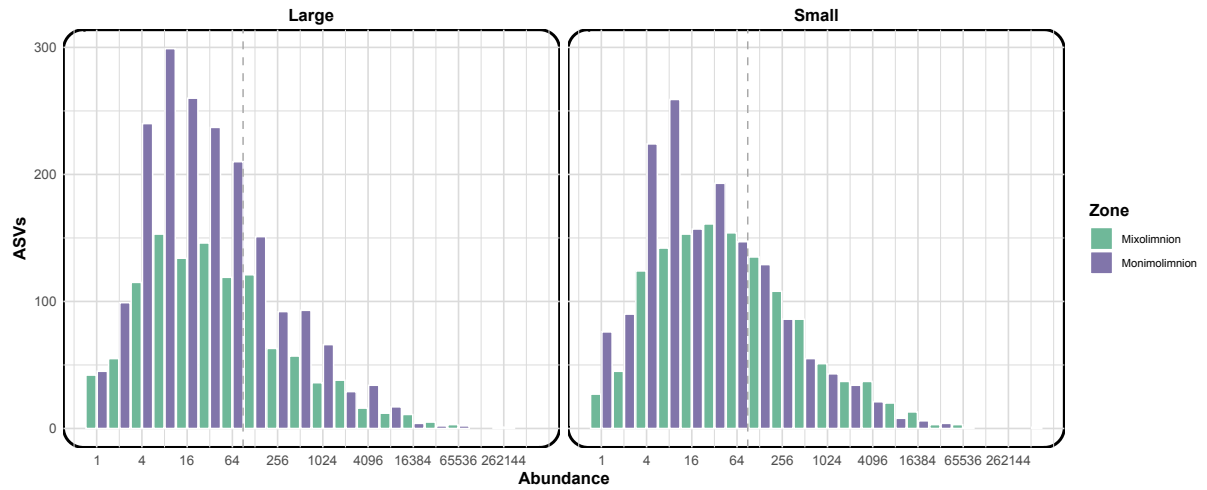


Figure S9: Barplots representing the number of ASVs (y-axis) by sequences abundance (x-axis) across sampling zones and size fractions.

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

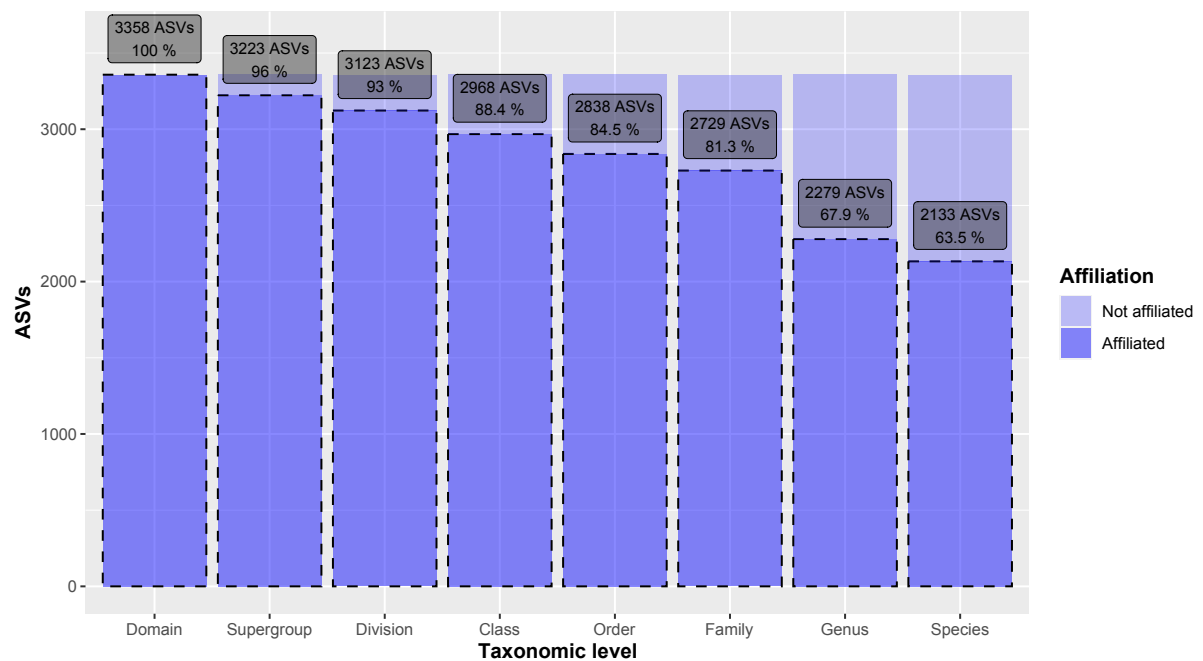


Figure S10: Taxonomic affiliation statistics.

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)).

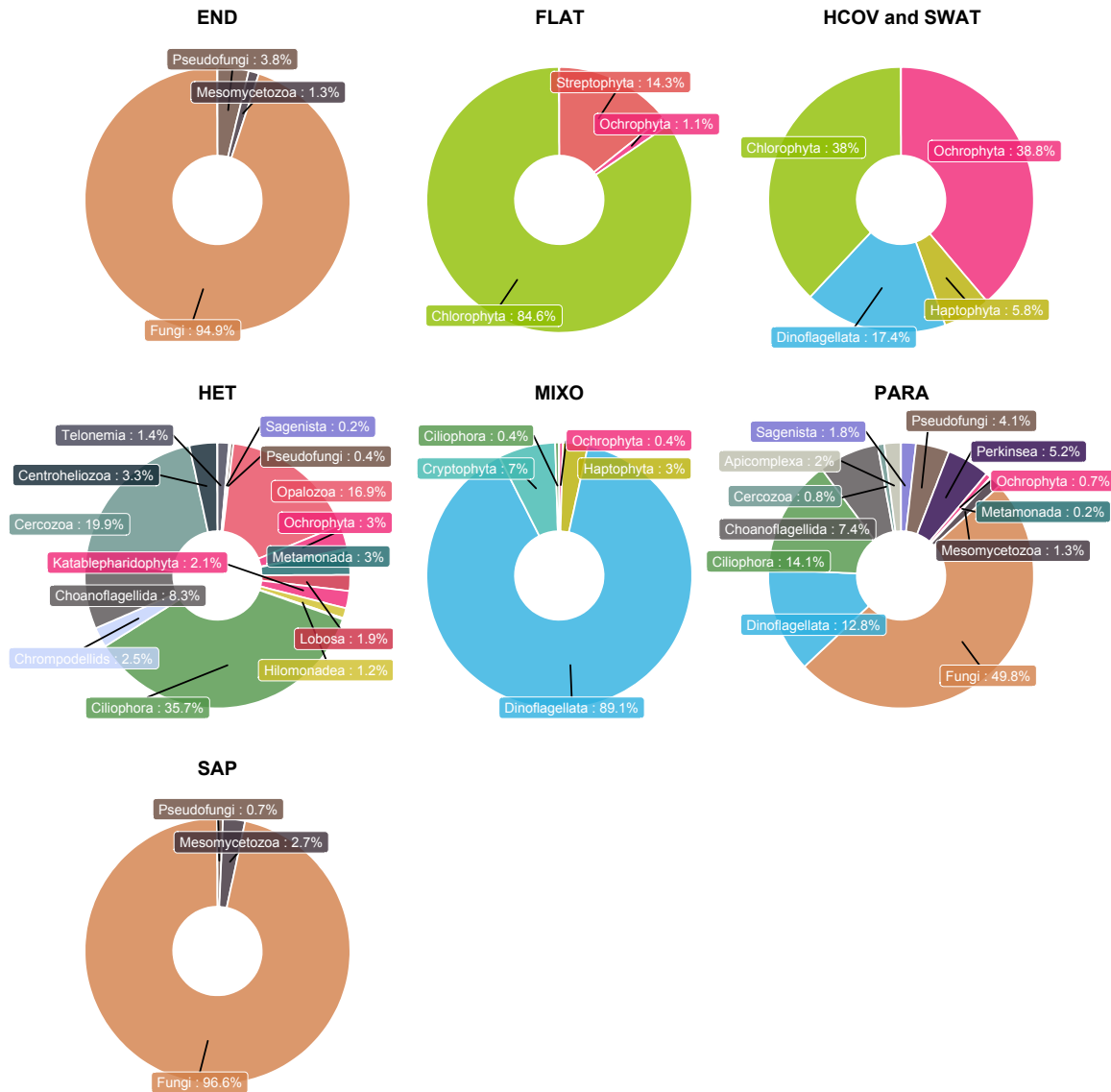


Figure S11: ASVs richness of the eukaryotic divisions across each functional group. Functional groups are named as described in Figure 5.

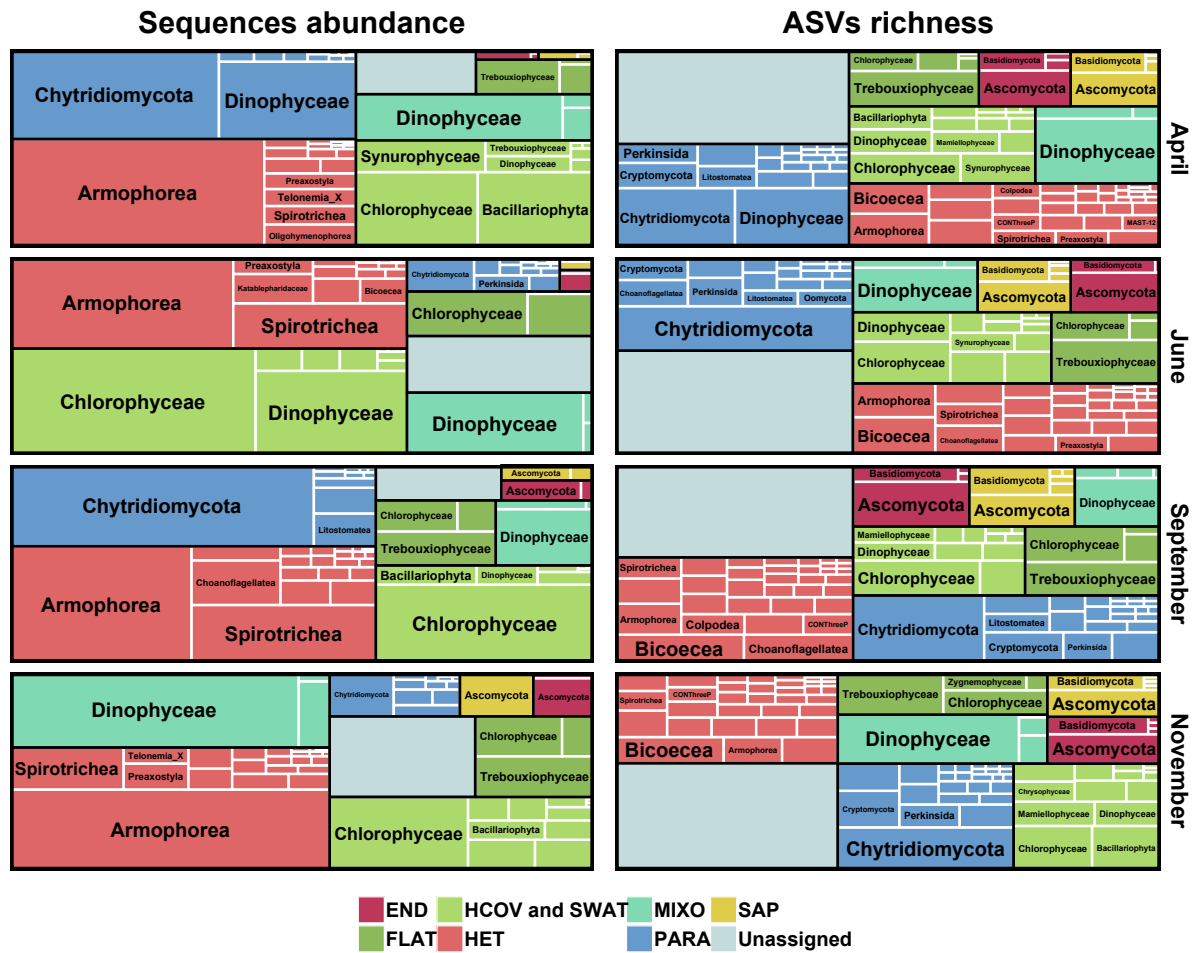


Figure S12: Relative abundance and ASVs richness of the eukaryotic classes across each functional group according to the sampling period.

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).

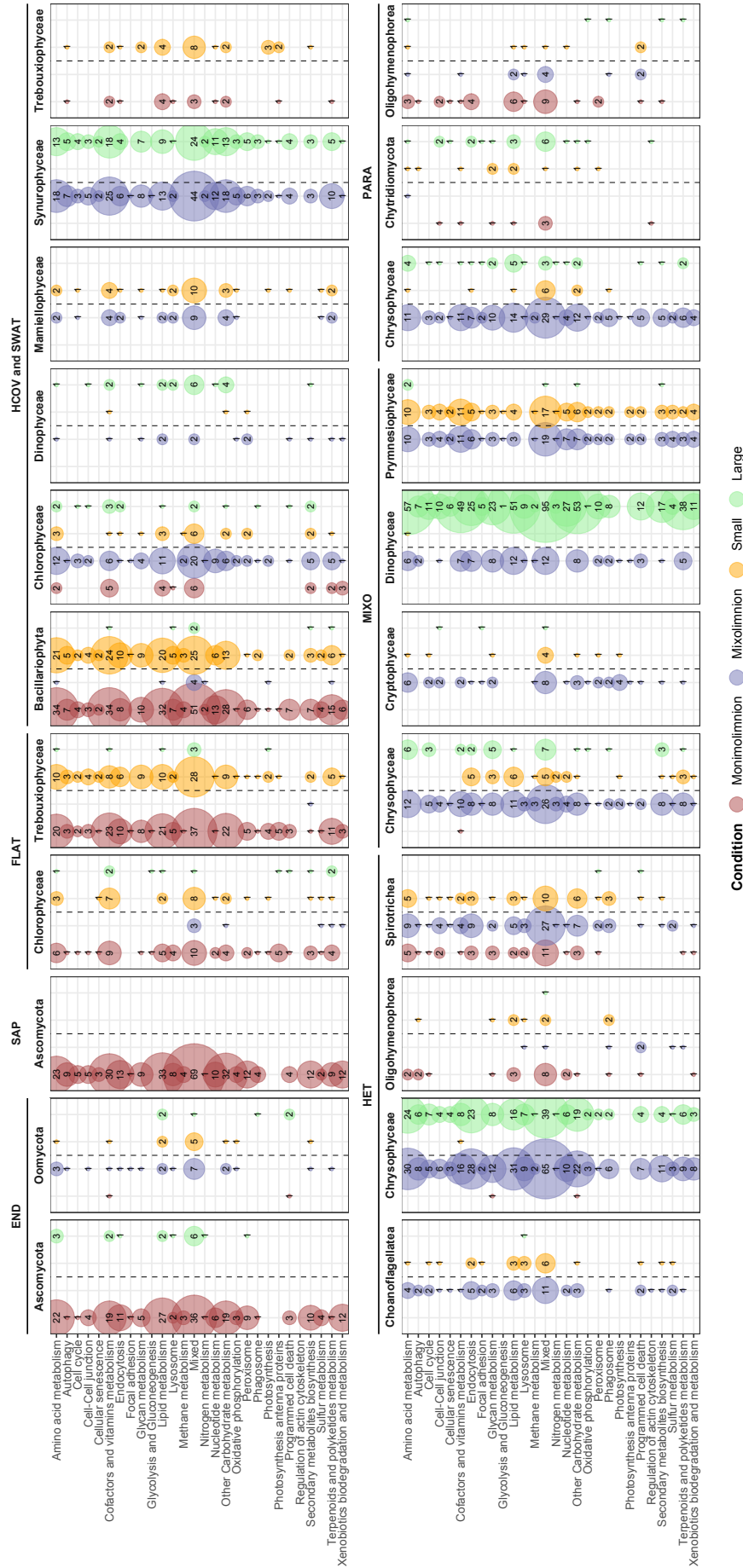
Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across zone is characterized by an adjusted p -value <0.05 and an absolute value of \log_2 -FoldChange >2 are represented. Values of \log_2 -FoldChange >4 are notified with « · », \log_2 -FoldChange >5 with « * » while \log_2 -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 \log_2 -FoldChange >2 are represented. Values of \log_2 -FoldChange >4 are notified with « · », \log_2 -FoldChange >5 with « * » while \log_2 -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 \log_2 -FoldChange >2 are represented. Values of \log_2 -FoldChange >4 are notified with « · », \log_2 -FoldChange >5 with « * » while \log_2 -FoldChange >6 with « ** ».



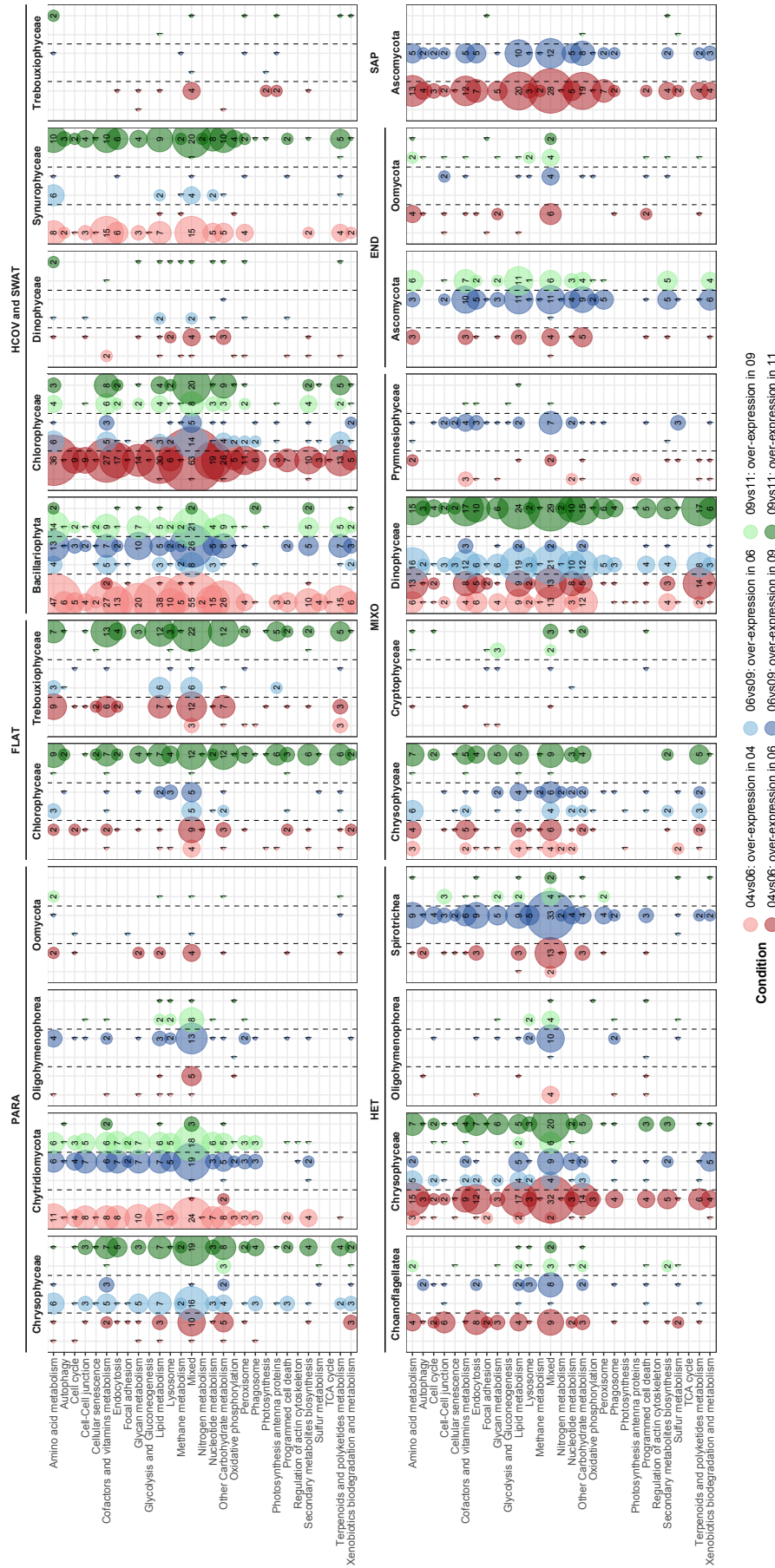


Figure S17: Bubble plot representing the number of over-expressed genes involved in eukaryotic metabolisms and cellular processes across periods for each eukaryotic class and functional group. Only KO identifiers whose DESeq2 baseMean are upper than 10 and differential expression analysis across periods is characterized by an adjusted p -value < 0.05 and an absolute value of \log_2 -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 pluricellular 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., McGlenn, 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.

CHAPITRE III

DÉCOUVERTE DE SIGNATURES GÉNÉTIQUES AU SEIN
DES PARASITES ET D'AUTRES GROUPES
FONCTIONNELS D'EUCARYOTES UNICELLULAIRES
D'EAU DOUCE PAR LA CONSTRUCTION DE RÉSEAUX
DE SIMILARITÉ DE SÉQUENCES PROTÉIQUES

Contexte de l'étude

Au cours de l'étude précédente, nous avons utilisé une approche originale afin d'étudier la diversité taxonomique et fonctionnelle de différents groupes d'eucaryotes microbiens lacustres. Les données de métatranscriptomique, couplées à l'étude de traits, ont notamment permis de décrire le potentiel métabolique de différents groupes fonctionnels, caractérisés par des stratégies trophiques distinctes, tout en évaluant leurs dynamiques en fonction de différents paramètres. Toutefois, le jeu de données finalement utilisé dans le chapitre II était relativement réduit en raison des différents filtres d'analyses. En effet, seulement 30% et 7,6% des protéines ont pu, respectivement, être affiliés taxonomiquement et assignés à un groupe fonctionnel. En outre, les eucaryotes renferment un vaste réservoir de gènes dont 40 à 70% sont de fonction inconnue rendant les études fonctionnelles généralement limitées aux fonctions « d'entretien » (Grossman *et al.*, 2016 ; Carradec *et al.*, 2018 ; Trench-Fiol et Fink, 2023). De plus, les spécificités métaboliques d'organismes de différentes stratégies trophiques sont encore méconnues ce qui complexifie grandement l'interprétation des données fonctionnelles acquises. En effet il n'est décrit, à ce jour, que peu de marqueurs protéiques partagés par un groupe d'organismes reflétant par exemple leur mode trophique.

Dans ce contexte, l'objectif principal de ce chapitre était, à partir des données de métatranscriptomique préalablement obtenues (chapitre II), d'étudier plus spécifiquement le lien entre protéines et stratégies trophiques et notamment d'en définir de potentielles signatures protéiques. Du fait de leur forte diversité et du rôle important qu'ils jouent dans la structuration des communautés et le fonctionnement des écosystèmes lacustres, nous avons choisi de nous focaliser sur les parasites. Pour cela, nous avons utilisé une approche basée sur la construction de réseaux de similarité de séquences protéiques permettant d'étudier facilement les caractéristiques métaboliques de phyla éloignés pouvant partager des fonctions écologiques similaires tout en considérant à la fois les protéines connues et inconnues.

Metatranscriptomes-based sequence similarity networks uncover genetic signature within parasites and other functional groups of freshwater microbial eukaryotes

Arthur Monjot¹, Lucie Bittner², Jérémy Rousseau², Cécile Lepère^{1°}

1- CNRS, Laboratoire Microorganismes : Génome et Environnement, Université Clermont Auvergne, Clermont-Ferrand, F-63000, France.

2- Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Paris, France

°corresponding author: Cecile.Lepere@uca.fr

En préparation

I. Abstract

Taxonomic and functional diversity of microbial eukaryotes in aquatic ecosystems is increasingly well described due to recent advances in sequencing technologies. However, the vast amount of data produced by -omics approaches require data-driven methodologies to make predictions about role of these organisms within ecosystems. From metatranscriptomics data we use sequence similarity network-based approach to explore metabolic specificities of microbial eukaryotes of different trophic modes in an understudied freshwater ecosystem (lake Pavin, France). Both known and unknown sequences of 2 165 106 proteins were clustered in connected components (CCs) enabling to link 64.6% of proteins to a functional group while only 18.5% were assigned before extrapolation. By investigating functional group assignment as well as proteins functional annotations within CCs, we observed genetic proximity of proteins of different groups (between heterotrophs and photo-osmo-phago-mixotrophs or between saprotrophs and parasites), as well as a relative functional redundancy of primary metabolisms. On the contrary, we identified that almost 1 million of the proteins appear to be group-specific.

By applying our strategy to study the abundant, diverse and yet enigmatic parasitic organisms, we identified shared genetic signatures unveiling genes likely involved in host-parasites interactions and representing a few interesting lines in the comprehension of their roles within environment.

Keywords: Sequence Similarity Network, Metatranscriptomic, Microbial eukaryotes, Freshwater ecosystems, Functional diversity, Parasites

II. Introduction

Aquatic ecosystems, such as oceans, lakes, and rivers, host a wide array of microbial eukaryotes that play essential roles in nutrient cycling, energy transfer, and ecosystem functioning. However, the functional diversity and ecological roles of these microbial eukaryotes remain largely unexplored. Traditional methods for studying microbial eukaryotes have limitations in capturing their functional potential and interactions within complex ecosystems. To overcome these limitations, researchers have turned to the use of -omics approaches to predict the diversity and functional composition of microbial communities based on environmental data (Keeling *et al.*, 2014; de Vargas *et al.*, 2015; Faure *et al.*, 2021; Delmont *et al.*, 2022). Metatranscriptomic is currently the best available approach to obtain large amount of genomic information from uncultured organisms isolated from the environment. While freshwater microbial eukaryotes are rarely studied through RNA seq analysis (Grossmann *et al.*, 2016; Trench-Fiol and Fink, 2020; Li *et al.*, 2022), numerous marine metatranscriptomic studies have been conducted across spatial and temporal scales leading to the increase of our knowledge on physiology (Alexander *et al.*, 2015; Pearson *et al.*, 2015; Lampe *et al.*, 2018; Caputi *et al.*, 2019; Kolody *et al.*, 2019), nutritional modes (Hu *et al.*, 2018; Lambert *et al.*, 2021) and contributions to ocean biogeochemistry (Carradec *et al.*, 2018; Cohen *et al.*, 2021).

As most current analyses depend on a species or function name, a considerable amount of newly generated sequences in environmental samples are overlooked. Microbial eukaryotes have indeed diverse and complex genomes, and this vast genes reservoir is for 40% to 60% with unknown function (Bernard *et al.*, 2018; Carradec *et al.*, 2018; Almeida *et al.*, 2021; Vanni *et al.*, 2022).

Current analytical approaches (Huerta-Cepas *et al.*, 2017; Quince *et al.*, 2017; Franzosa *et al.*, 2018; Chen *et al.*, 2019; Mitchell *et al.*, 2019) generally do not include this uncharacterized fraction in downstream analyses, constraining their results to conserved pathways and housekeeping functions (Quince *et al.*, 2017). Sequence similarity network (SSN) approaches are powerful tools for analyzing the large amount of data produced by high throughput sequencing and allow to study relationships between and within protein families (*e.g.*, (Atkinson *et al.*, 2009; Bittner *et al.*, 2010; Alvarez-Ponce *et al.*, 2013; Cheng *et al.*, 2014; Forster *et al.*, 2015; Corel *et al.*, 2016; Méheust *et al.*, 2016; Copp *et al.*, 2018)). By clustering sequences with high similarity, these networks reveal potential functional associations and ecological interactions among microbial communities and enable the inclusion of functionally unannotated sequences in the global analysis. Previous studies using SSN, have been already conducted in a functional context, for example, in order to unveil gene sets characteristic of specific functional traits (*e.g.*, toxicity and symbiosis capability of Dinoflagellata (Meng *et al.*, 2018)) or specific environmental conditions (Faure *et al.*, 2021).

This approach can also help in the definition of genetic marker (Cheng *et al.*, 2014). Some genetic markers are already known and used to target for example phototrophic organisms in metagenomic context (*e.g.*, gene *psbO* (Pierella Karlusich *et al.*, 2023)). This knowledge is linked with the fact that photosynthesis is probably the best described cellular mechanism. Markers for other type of nutritional mode are more complicated to find however phagotrophy in a variety of phyla is increasingly studied and peptidases, proton pumps and lysosome enzymes such as cathepsin and rhodopsin could be candidate to target this ecologically important feeding strategy (Labarre *et al.*, 2020; Massana *et al.*, 2021). Moreover, phagotrophy used by photosynthetic organisms (*i.e.*, phago-mixotrophic organisms) has been the subject of many recent studies using experimental assays as well as in-silico prediction models (Bock *et al.*, 2021; Lambert *et al.*, 2021; Mitra *et al.*, 2023). Although there are no smoking guns for parasitism, parasites might be identified through the presence of various proteins involved in pathogenesis. The CRN domain or Crinkler proteins are for example a class of effectors (proteins that are secreted into a host and modify their behavior) that are known only from parasites (James *et al.*, 2013; Quiroz Velasquez *et al.*, 2014).

In this study, the metabolic specificities of microbial eukaryotes of different trophic modes were explored in an understudied freshwater ecosystem using an original Sequence Similarity Network (SSN)-based approach.

Metatranscriptomic dataset obtained from the meromictic lake Pavin (France) was deeply investigated, enabling the study of the abundant but understudied microbial dark matter, as well as the shared genetic signatures of relatively little-known freshwater parasitic microorganisms.

III. Materials and Methods

A. Dataset description

The dataset analyzed in this study is publicly available under ENA accession number PRJEB61515. It corresponds to metatranscriptomic data obtained through the sampling of the meromictic lake Pavin (Massif Central, France, 45°29'45" N, 2°53'17" E) (Monjot *et al.*, 2023). This lake is characterized by two permanently stratified water layer: an upper oxygenated layer (mixolimnion) extending from the surface to 60 m, and an anoxic lower layer (monimolimnion) extending from 60 to 92 m depth. Metatranscriptomes were retrieved from sampling carried out in both zones (*i.e.*, at 9m and 80m) at the center of the lake, by day and night and at four contrasted periods in 2018: April, June, September and November (Monjot *et al.*, 2023).

B. Unigene catalog proteins prediction and annotations

The metatranscriptome derived unigene catalog were obtained 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% of 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 unigene catalog. Proteins prediction was realized from all unigenes using *Transdecoder.LongOrfs* followed by *TransDecoder.Predict* (v5.5.0) with the default parameters. A second run with a minimum protein length of 70 (*-m*) was achieved for unigenes without predicted protein.

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 MetaEuk database (Levy Karin *et al.*, 2020) using *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 Bacteria, Archaea, Virus and multicellular organisms (*i.e.*, Metazoan, some Basidiomycota and Ascomycota, Rhodophyta, Magnoliopsida, Pinopsida and Polypodiopsida).

Manually curated association table was then developed, using bibliography (362 scientific articles), to link, when possible, each taxonomic affiliation (*i.e.*, 1052 distinct affiliations) to a functional group (*i.e.*, phototrophs (considered as photo-osmo-mixotrophic (identified by PHOTO)), mixotrophs (considered as photo-osmo-phago-mixotrophic (MIXO)), heterotrophic (HET), saprotrophic (SAP) and parasitic (PARA) (facultative and obligate) microorganisms (Table S1)).

C. SSN Building

A sequence similarity network (SSN) where vertices correspond to sequences and edges represent the similarity and coverage between pairs of sequences was built. Diamond v2.1.7 (Buchfink *et al.*, 2015) was used in *blastp* mode to compute the percentage of similarity between every pair of proteins detected in the metatranscriptomic dataset, with the options *-e 1e-5 --sensitive*. Diamond output was filtered using 80% identity and 80% coverage threshold. This coverage threshold is commonly used in SSN studies (Meng *et al.*, 2018) and identity threshold is usually determined on the basis of maximum functional homogeneity between linked proteins (Faure *et al.*, 2021). The filtered output was used to build SSN with the *igraph-python* library (v0.10.4) (Csardi and Nepusz, 2006).

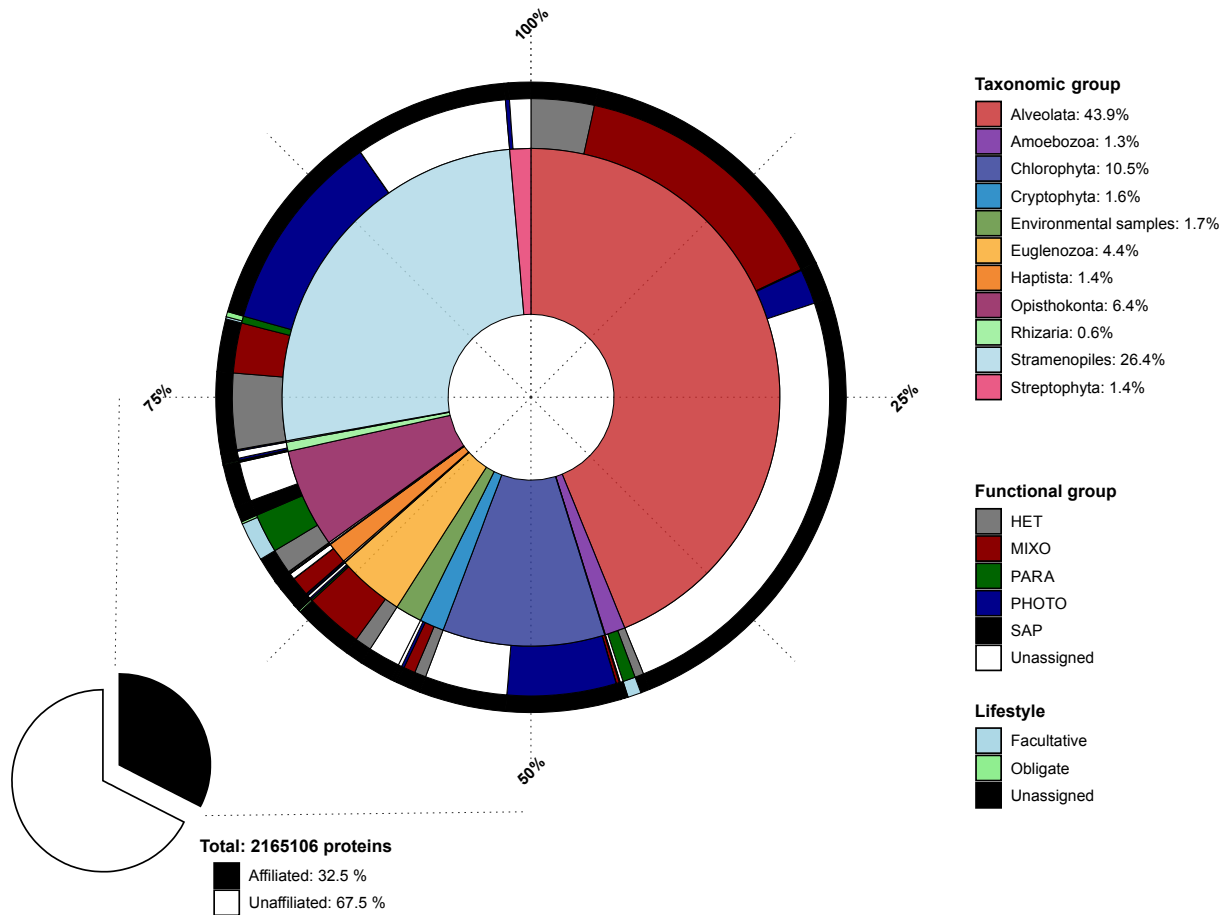


Figure 1: Taxonomic diversity and functional group assignment of proteins integrated in the sequence similarity network.

From inner to outer circles, colors represent taxonomic affiliation, functional group assignment and lifestyle. Functional groups were identified using the following codes: PHOTO: photo-osmo-mixotrophs, MIXO: photo-osmo-phago-mixotrophs, HET: heterotrophs, SAP: saprotrophs and PARA: parasites (facultative and obligate).

The resulted SSN gathered singleton (*i.e.*, vertices without any homology with other sequences) and CCs (connected components: subgraphs composed of at least two vertices disconnected from the rest of the network). Only CCs characterized by at least three vertices were kept for the rest of the analysis. The resulting SSN was finally composed of 302 304 CCs, including 2 165 106 proteins (5 317 694 proteins clustered in CCs characterized by less than 3 proteins were excluded from the analysis).

D. CCs analysis

Statistical and CCs analysis were performed using R (R Core Team, 2022). Each CCs was characterized by its sequences, which were taxonomically affiliated, assigned to a functional group and functionally annotated against KEGG database. Using these annotations, functional homogeneity of each CCs was assessed by computing a homogeneity score (Fhom) as described in Faure *et al.* (2021). The means of Fhom scores as well as CCs number and their annotations were compared for eight similarity thresholds (*i.e.*, 65%, 70%, 75%, 80% ,85%, 90%, 95% and 100%) (**Fig. S1**). The intermediary 80% identity threshold was selected to maximize the functional homogeneity between linked proteins relative to the total number of CCs while minimizing the number of unannotated components.

CCs were linked to a functional group when a majority of their proteins was associated to the same functional group. Considering potential taxonomic affiliation errors leading to mis-association of proteins, the proportion of 75% among the total assigned proteins of the component was chosen. These specific components were called SPCs (Specific Protein Clusters) hereafter.

Correspondence Analysis (CA) was realized to describe the relationship between CCs and functional groups with the FactoMineR package (v2.6) (Lê *et al.*, 2008) and using the number of proteins associated to each functional groups for each CC. Kruskal-Wallis test was realized to measure the significance of the difference between functional groups using the number of each proteins within each CCs.

CCs which clustered both unknown proteins and known proteins characterized by the same functional annotation, taxonomic affiliation and functional group assignment were used to estimate the number of unknown proteins that could be in fact characterized.

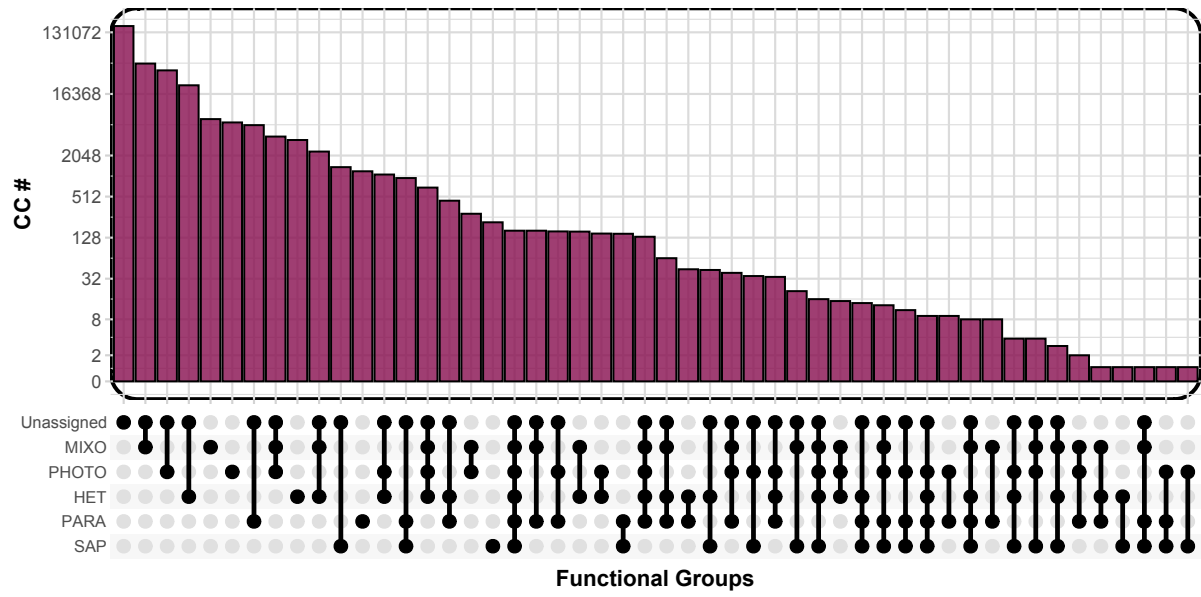


Figure 2: Upset plot displaying shared and unique connected components across functional groups.

Functional groups are identified using the same code as in Figure 1.

E. SPCs related to parasitism

This study focused on the SPCs related to parasitism. Among them, SPCs which contained at least one protein associated to an obligated parasite (*i.e.*, $n=1\ 679$) were kept in order to select proteins potentially involved in host-parasite relationships. Using the positions of all CCs in the two first dimensions of the CA space (44.4% of variance), barycenter positions were computed for each KO id (KEGG Orthology identifier) metabolic pathway (KEGG_{PATHWAY}=Metabolism) detected among obligate parasites-related SPCs. These SPCs were graphically represented using *ggraph* (v2.1.0) (Lin Pedersen, 2022), *tydigraph* (v1.2.3) (Lin Pedersen, 2023) and the *igraph* R package (v1.5.0) (Csardi and Nepusz, 2006). Upset plot were generated using the *ggupset* R package (v0.3.0) (<https://github.com/const-ae/ggupset>).

IV. Results

A. Sequences similarity network statistics

Almost 10 million transcripts were retrieved from the lake Pavin metatranscriptomic dataset. From these transcripts, 7 511 376 proteins have been predicted and compared to each other (*i.e.*, 107 650 304 diamond alignments) to build sequences similarity network (SSN). After filtration steps, 302 304 connected components (CCs) were obtained, ranging from 3 to 6 203 proteins (**Fig. S2**) and gathered a total of 2 165 106 proteins. Among them, 32.5% were taxonomically affiliated and 18.5% were assigned to a functional group (**Fig. 1**). Functionally unassigned proteins were predominantly affiliated to Alveolata, Stramenopiles, Chlorophyta and Opisthokonta, accounting, for example, for more than half of the Alveolata-affiliated proteins. Among assigned proteins, mixotrophs dominated ($n = 156\ 485$) and were mainly affiliated to Alveolata, Euglenozoa, Stramenopiles, Haptista and Cryptophyta. Phototrophs affiliated to Chlorophyta, Stramenopiles and Dinoflagellata comprised 136 946 assigned proteins. Heterotrophs ($n = 77\ 955$) presented half the amount of mixotrophs assigned proteins and clustered a wide variety of organisms: Stramenopiles, Alveolata, Opisthokonta, Euglenozoa, Cryptophyta and Amoebozoa. Assigned proteins within parasites and saprotrophs were low (24 232 and 5 952, respectively) and the majority was affiliated to Opisthokonta. Proteins assigned to parasites were also found among Stramenopiles, Amoebozoa, Alveolata, Heterolobosea, Metamonada and Rhizaria.

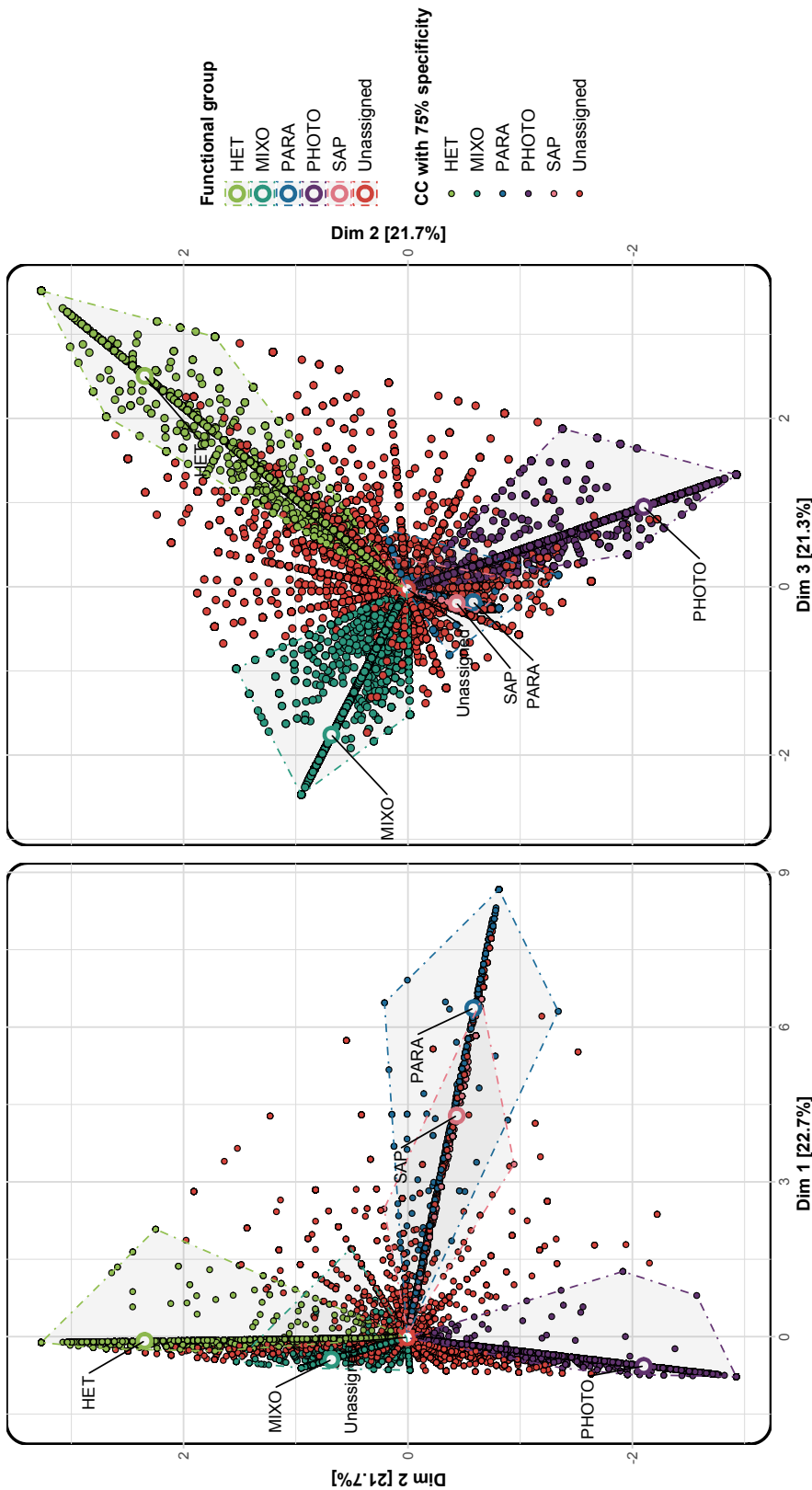


Figure 3: Correspondence analysis testing the relationship between CCs and functional groups.

This analysis is processed using the number of proteins associated to each functional group for each CC. Functional groups are identified using the same code as in Figure 1.

Facultative parasites mainly affiliated to Opisthokonta, Stramenopiles and Amoebozoa were represented by 20 152 proteins, while those with an obligatory lifestyle accounted for only 4 080 proteins and were found in majority among Alveolata, Rhizaria, Euglenozoa, Opisthokonta and Stramenopiles (**Fig. 1**).

B. CCs distribution among functional groups

More than half of CCs grouped proteins without any functional group assignment (*i.e.*, 162 001 CCs). The rest of the CCs clustered proteins (*i.e.*, a total of 1 400 022 sequences) assigned to only one functional group (129 151) and up to the five functionals groups (162) defined in this study. 18 960 CCs were only composed with assigned proteins while 121 343 CCs were composed of both assigned and unassigned proteins (**Fig. 2**). Among the specific CCs, (i) 52 682 (including 7 007 CCs for which all the proteins were assigned) were linked to mixotrophs, (ii) 42 561 (6 258) to phototrophs, (iii) 25 389 (3 458) to heterotrophs, (iv) 6 920 (1 203) to parasites, (v) and 1 599 (215) to saprotrophs.

CCs linked to multiple functional groups suggest they share common features. Thus, mixotrophs shared 4 171 CCs with phototrophs and 2 498 with heterotrophs. Parasites shared a significant number of CCs with saprotrophs (*i.e.*, 1 103) and heterotrophs (490) (**Fig. 2**).

Using the number of proteins associated to each functional groups within each CCs, correspondence analysis (CA) showed significant variability (Kruskal-Wallis p -value <0.001) between almost all of groups (the three first dimensions explaining 65.7% of variability) (**Fig. 3**). The first dimension dissociated parasites and saprotrophs from the other functional group while second- and third-dimensions split phototrophs, heterotrophs and mixotrophs.

C. Taxonomic and metabolic specificity of functional groups

Mixotrophs, photo-osmo-mixotrophs and heterotrophs were represented by numerous SPCs (Specific Protein Clusters: CCs for which at least 75% of assigned proteins were assigned to the same functional group) (*i.e.*, respectively 53 962, 43 291 and 26 063 gathering 403 174, 334 474 and 181 251 proteins), almost half of their sequences were taxonomically affiliated (**Fig. 4**).

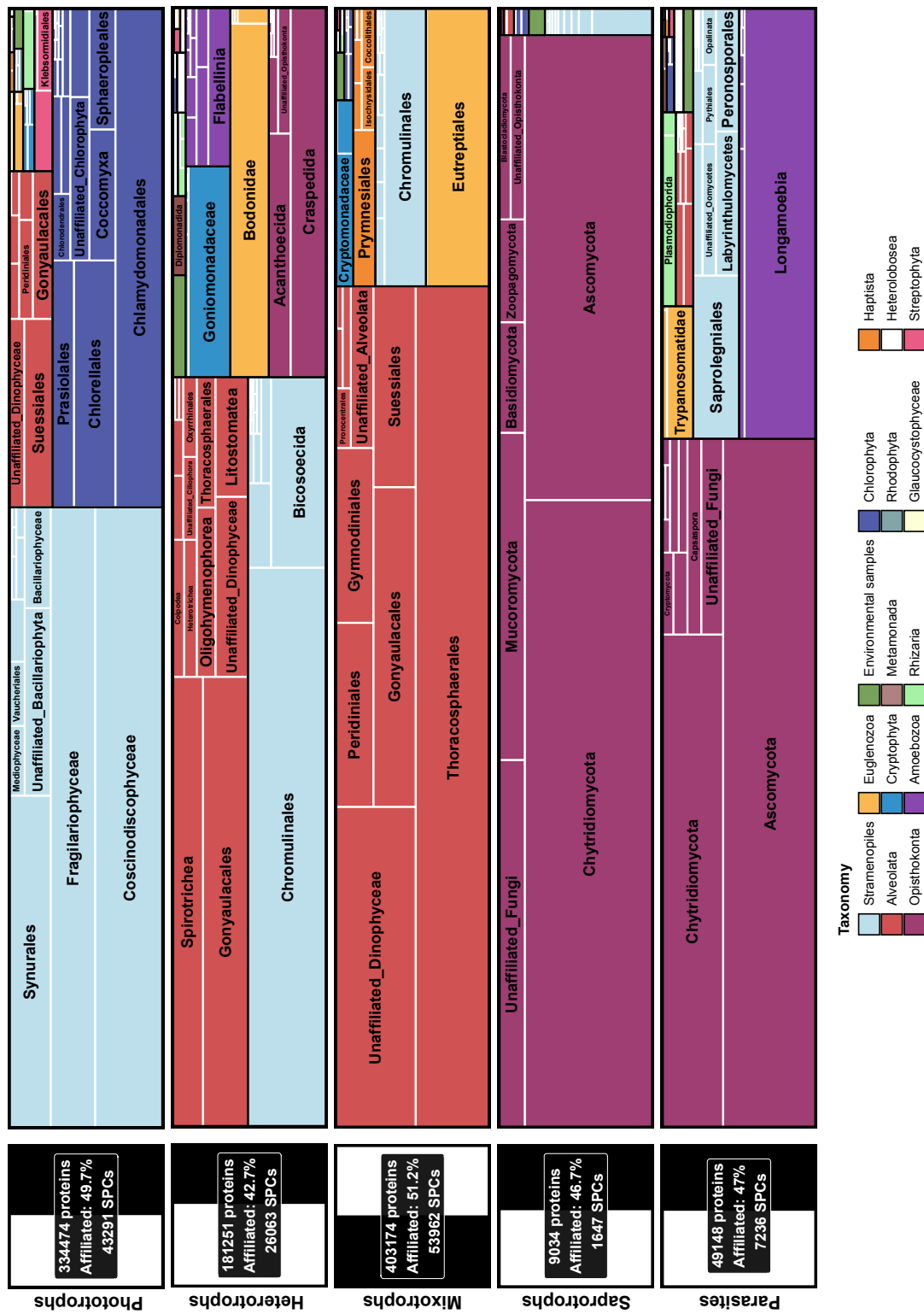


Figure 4: Taxonomic affiliation of proteins gathered in Specific Proteins Clusters (SPCs) of each functional group.

Number of SPCs, total number of proteins within SPCs as well as affiliation rate are displayed on the horizontal histograms for each functional group. Functional groups are identified using the same code as in Figure 1.

Although some taxa appeared in more than one of these three functional groups (*e.g.*, Dinophyceae, Chromulinales, etc.), taxonomic diversity greatly differed at low taxonomic levels (<Class) (**Fig. 4**). Parasites and saprotrophs were characterized by a low number of SPCs: 7 236 and 1 647, respectively, grouping only 49 148 and 9 034 proteins with an average taxonomic affiliation rate of 46.9%. Proteins clustered in SPCs of both groups mainly belonged to Chytridiomycota and Ascomycota (Opisthokonta) but parasites also clustered numerous proteins affiliated to Longamoebia (Amoebozoa), Saprolegniales, Labyrinthulomycetes and Peronosporales (Stramenopiles), Trypanosomatidae (Euglenozoa) and Plasmodiophorida (Rhizaria).

Functional annotation against the KEGG database allowed the identification of more than half of photo-osmo-mixotrophs, heterotrophs, saprotrophs and parasites proteins clustered in SPCs (**Fig. 5**) (*i.e.*, 53.9% of the total proteins were annotated with a KO Id (KEGG Orthology identifier)). Mixotrophs were the least annotated groups with only 46% of annotated proteins. Metabolisms-related proteins accounted for almost a third of annotations, ranging from 11.4% (mixotrophs) to 21.5% (saprotrophs).

Contrary to the high variability of proteins taxonomic affiliations between the SPCs of each functional groups, their functional annotations varied less considering the B levels of KEGG metabolism annotations (**Fig. 5**). Amino acids, carbohydrates, lipids and energy metabolism were the most found across all functional groups.

However, more variations between functional groups were recorded regarding the C level (**Fig. S4**). Indeed, although heterotrophs and parasites shared an important proportion of proteins related to lipid metabolisms such as those involved in fatty acid degradation, some specificities can be detected: (i) heterotrophs were characterized by numerous proteins involved in sphingolipid and glycerophospholipid metabolisms; (ii) parasites revealed a majority of their proteins linked to biosynthesis of unsaturated fatty acids and glycerophospholipid metabolism. Saprotrophs SPCs also clustered proteins involved in glycerophospholipid as well as sulfur metabolism and prodigiosin biosynthesis. Finally, a great proportion of proteins linked to photosynthetic organisms (phototrophs and mixotrophs) are involved in photosynthesis (*e.g.*, antenna proteins) and carotenoid biosynthesis (**Fig. S4**).

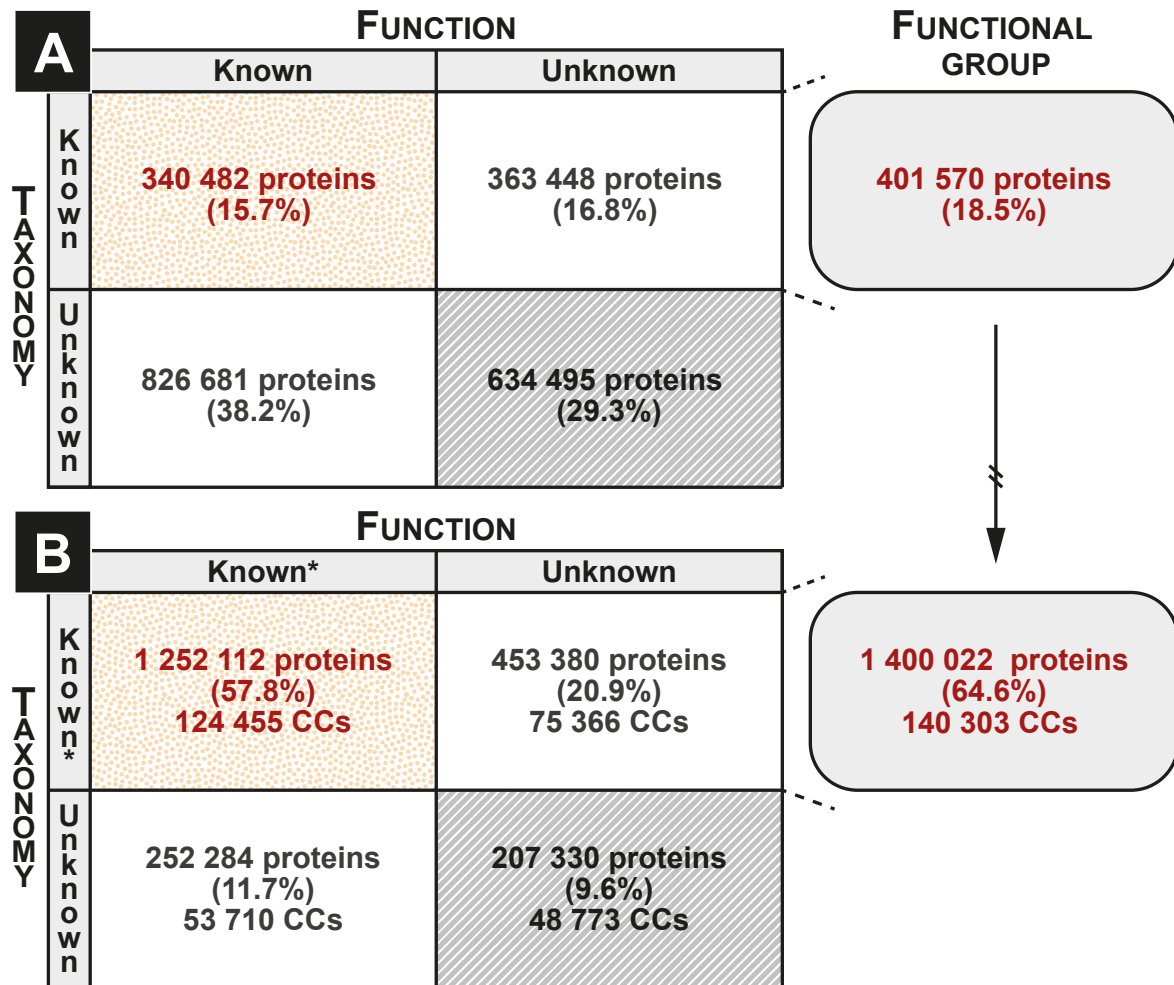


Figure 5: Taxonomic affiliation, functional annotation and functional group assignment of proteins with metatranscriptomic approach (A) and with the implementation of SSN-based approach (B).

*: proteins are defined as “Known” when they clustered within the same CCs of proteins that are taxonomically affiliated, functionally annotated and assigned to a functional group.

D. Using SSN to consider the unknown

The use of SSN evaluating homology between proteins allows to consider unknown sequences in the analysis by clustering them with both annotated and affiliated sequences within CCs. Thereby, 57.8% of total sequences were taken into account in the analysis (1 252 112 distributed in 124 455 CCs) while only 15.7% (340 482 proteins) were really annotated and affiliated (**Fig. 5**). Likewise, 1 400 022 sequences (64.4% of the dataset) were linked to functional groups through CCs while a minority were initially assigned (*i.e.*, 18.5%) (**Fig. 5**). Furthermore, making the hypothesis that unknown proteins found in CCs which grouped known proteins characterized by the same taxonomic and functional information, were close to each other, we estimate the number of unknown proteins that can be characterized through the use of SSN. This re-evaluation allowed the increase of the taxonomic affiliation rate of the proteins by an average of 6% at all taxonomic levels (*i.e.*, from 25.09% to 31.96% (Class), from 29.95% to 36.32% (Phylum) and from 32.51% to 38.74% (Division)), as well as the functional annotation which evolved from 53.91% to 57.38% and the functional group assignment which present an 8% improvement (from 18.54% to 26.18%).

E. SPCs related to parasitism

Parasite SPCs were reduced to 1 679 when selecting those that grouped at least one protein affiliated to an obligate parasite. They clustered 12 946 proteins among which 62% were assigned to obligate parasites, 9.2% to facultative parasites (including Amoebozoa, Opisthokonta and Stramenopiles), 1.6% to saprotrophs (*e.g.*, Opisthokonta and Stramenopiles) and 26.8% remained unassigned (**Fig. S5**). This selection also reduced distinct functional annotations from 3 268 to 1 143 KO Id which were likely to be involved in host-parasite relationships. Among the selected SPCs, 496 had no protein functional annotation (*e.g.*, n°18 (**Fig. 6**)), 837 were annotated to a unique KO Id (*e.g.*, n°33 938) and 346 had multiple annotation (*e.g.*, n°32 689).

Although some SPCs grouped proteins with different taxonomic affiliations (at the class level such as the component n°32 689 and n°18 grouping respectively Oomycetes and Fungi, or at the phylum level such as the n°33 938 which clustered Cercozoa, Discosea and Fungi) the majority clustered few proteins (~ 7.3 proteins/CC) with similar taxonomic affiliations (*e.g.*, n°24 890 and n°29 238 (**Fig. 6**)).

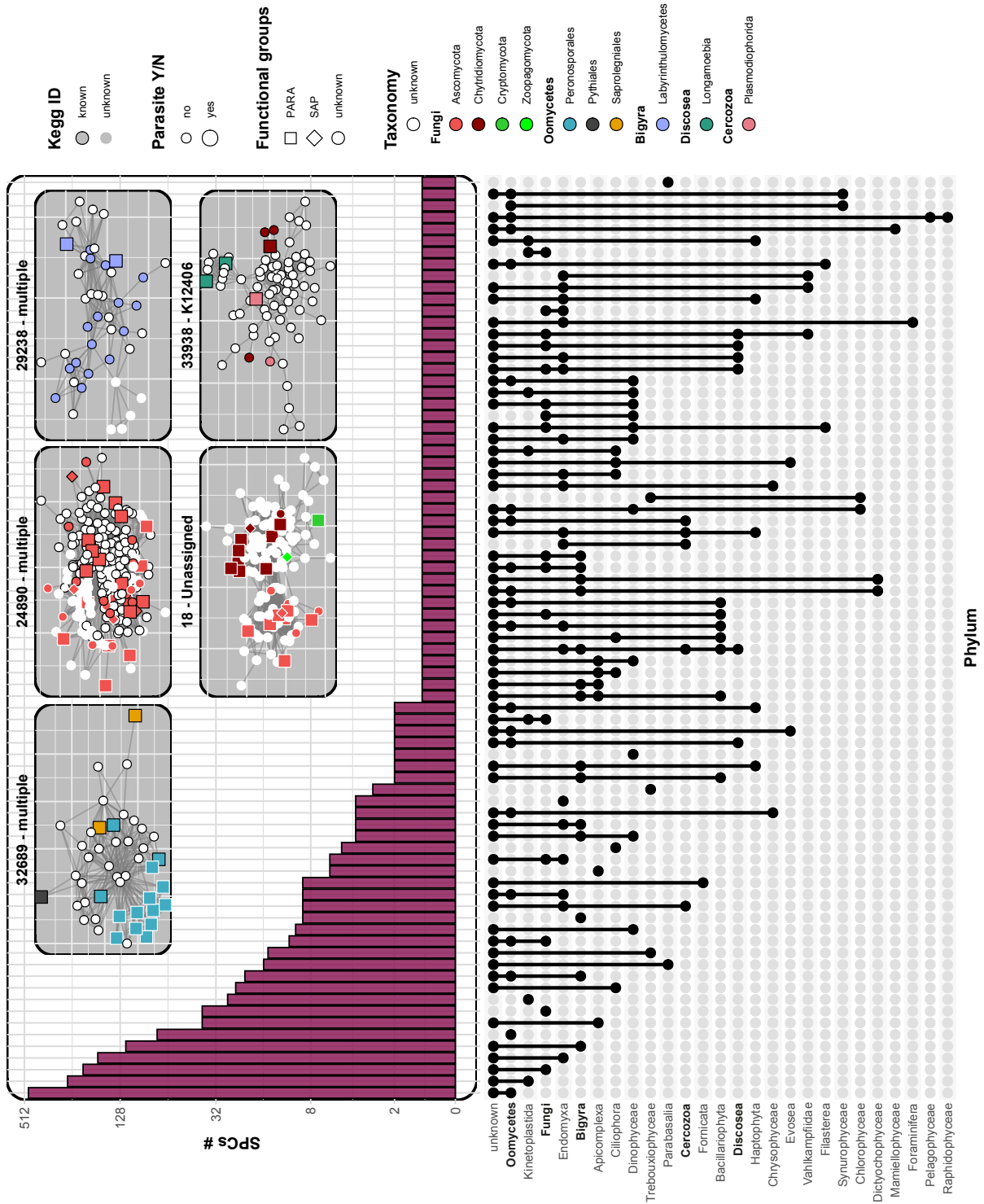


Figure 6: Upset plot displaying shared and unique Parasite-related SPCs across taxonomic phyla.

Examples of SPCs are represented on the top of the figure by sub-graphs gathering proteins (colored points) which are characterized by taxonomic affiliation, functional group assignment and functional annotation (KEGG Id). The mention “multiple” means that proteins of the SPC are annotated with multiple KO Id. The identification number of SPCs are referenced on the top of each sub-graph. Black outlined points represent a protein annotated with the KO Id referenced at the top of the sub-graph (*e.g.*, SPC n°33 938). Bold phyla are those represented within example SPCs. Functional groups are identified using the following codes: SAP: saprotrophs and PARA: parasites (facultative and obligate).

Figure 7: Correspondence analysis testing the relationship between CCs and functional groups with the addition of barycenter of KO Id correlated with parasitism.

Barycenter with coordinates on the dimension 1 (differentiate CCs with a majority (>75%) of their proteins assigned to parasite organisms and the others) exceeding 2 are defined as correlated with parasitism and represented with stars. Functional groups are identified using the same code as in Figure 1.

In order to identify KO Id with a clear correspondence with parasitism, barycenter of each KO id characterizing obligate parasites-related SPCs were computed using the coordinate of all CCs clustering proteins annotated with these identifiers (**Fig. 7**). This strategy enabled (i) to overcome protein filtration bias, (ii) to consider unannotated proteins as well as proteins clustered in SPCs of other functional groups. By selecting all KO Id whose barycenter coordinates on axis 1 were among the most important (>2), 40 identifiers likely to be involved in host-parasite interactions were obtained. SPCs which contained these identifiers (*i.e.*, $n = 74$, **Fig. S6**) clustered very few proteins (~ 6.7 proteins/SPC) characterized by similar affiliations suggesting that there is probably no overall functional pattern shared by all parasitic microbial eukaryotes.

Among these KO Id (complete list in **Table S2**), two involved uncharacterized proteins (*i.e.*, K09795/K09983), 12 were not clearly related to parasitism in literature (*e.g.*, K13514, K10971, K14263) and 26 were linked to host-parasite interaction (already described (12) or probably linked but not yet described (14)).

Among those already described, we found proteins affiliated to Trypanosomatidae and involved in the well-known mechanism of immunogenicity and extracellular survival of these microorganisms (*i.e.*, K20656/K12167 (Villas Bôas *et al.*, 1999; López-Farfán *et al.*, 2014)). Others affiliated to Fungi were related in the resistance against toxic compounds (*e.g.*, K06141, K09043 (Costa *et al.*, 2014; So *et al.*, 2019)) or described as stimulating parasite metabolism (*e.g.*, K09240, K18278) (**Table S2**).

Finally, the others KO Id, although not yet described in a host-parasite interaction, were characterized by functions which may contribute to establish this type of interaction or by features correlated with parasitism. We found, among others, some (i) related to anti-apoptotic agents and affiliated to Plasmodiophorida and Ascomycota (K20637/K17968 (Potting *et al.*, 2013; He *et al.*, 2021)), (ii) involved in DNA repair process (K19465, K10791) (Smolarz *et al.*, 2014) with multiple affiliations (Conoidasida, Ascomycota and Oomycetes), (iii) corresponding to genes whose repartition within fungi-like organism diversity (*e.g.*, Oomycetes, Fungi, Labyrinthulomycetes) seems to depend on their trophic capacity (saprotrophic or parasitic) (*i.e.*, K07414, K21200 and K15629 which belonged cytochrome P450 family (Sello *et al.*, 2015)) (**Table S2**).

V. Discussion

A. Identify the unknown

The proportion of unknown is high when examining functional diversity of eukaryotes in environmental datasets and increases rapidly due to the deluge of sequences characteristics of the post genomic era (Ellens *et al.*, 2017; Wyman *et al.*, 2018). Recent extensive sequencing initiatives such as the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP, <https://www.ncbi.nlm.nih.gov/bioproject/248394>, (Keeling *et al.*, 2014)) as well as the global ocean atlas of eukaryotic genes retrieved from TARA Ocean expeditions (Carradec *et al.*, 2018) report that an average of almost 50% of sequences remain without relative in public databases (Meng *et al.*, 2018; Levy Karin *et al.*, 2020). Estimates of the unknown are even higher when considering the few large-scale metatranscriptomic studies conducted in freshwater ecosystems (Grossmann *et al.*, 2016; Li *et al.*, 2022), indicating that 85% of all predicted proteins lack both known functions and assigned taxonomies (Trench-Fiol and Fink, 2020).

Sequence Similarity Network (SSN) offers an alternative to classical methods to study dataset in their globality (Cheng *et al.*, 2014; Méheust *et al.*, 2016; Copp *et al.*, 2018) and overcome several biases of de novo assembly processes (Meng *et al.*, 2018). By taking into account both known and unknown proteins, our SSN-based workflow enables to improve the number of sequences analyzed by 42.1% while traditional approaches would have used only 15.7% of the dataset by excluding proteins without functional annotation and/or taxonomic affiliation (**Fig. 5**). Furthermore, it allows to consider significantly more sequences which were neither annotated nor affiliated (decreasing from 29.3% to 9.6%) in accordance with previous SSN-based study which link 15% more proteins to taxonomy or function (Faure *et al.*, 2021).

Assuming that functional homogeneity is strictly maintained within connected components (*i.e.*, each CC would group proteins sharing the same functional characteristics), we've also estimate that an average of 6% of unknown proteins function could be identified by extrapolation of known proteins information's. Transitive transfer of annotations remains however debatable (Promponas *et al.*, 2015; de Crécy-Lagard *et al.*, 2022) and could be completed by further SSN-based prediction tools to substantially improve the consistency and the accuracy of annotations (*e.g.*, based on connectedness of network to refine annotations (Hornung and Terrapon, 2023) or on the construction of a probabilistic graphical model to assess the relevance of annotations (Yunes and Babbitt, 2019)).

B. SSN mining to study the microbial functional diversity

While highly permissive parameters are usually set to infer eukaryotic core genes (*i.e.*, at least 10% of identity and 75% of coverage (Parra *et al.*, 2007)), our SSN-based workflow use more stringent criterion (identity and coverage threshold of 80%) to study specificity of genes at a finer scale: the functional group. Only 162 CCs encompass all functional groups (**Fig. 2**) and may be recognized as constituting the core proteome of all eukaryotes in line with previous studies that reported however slightly higher observations due to less stringent parameters. For example, Meng *et al.* (2018) stated that core genome of Dinoflagellata was composed of 252 CCs while study at the eukaryotic domain scale report 429 single copy orthologs (Simão *et al.*, 2015). Although representing a small proportion of the network (~ 3.69%), CCs connecting multiple functional groups suggest metabolic similarity, as observed among phototrophs, mixotrophs and heterotrophs. Mixotrophs exhibit many CCs shared with both phototrophs and heterotrophs (~ 2000 CCs (**Fig. 2**)) in accordance with their metabolism versatility being able to combine ability of both specialists (*e.g.*, photosynthesis, osmotrophy and phagotrophy (Edwards, 2019; Ward, 2019)). In contrast, heterotrophs and phototrophs have fewer common CCs (*i.e.*, ~ 1000 CCs). Parasites and saprotrophs also exhibit proteins that are clustered in common CCs. This clustering may be consistent with their potential shared metabolic pathways, as well as their closely related taxonomy (*e.g.*, Fungi). Indeed, the lifestyles of Fungi affiliated to Chytridiomycota and Ascomycota in freshwater environments span a large spectrum, ranging from obligate parasite to saprotrophs (**Table S1**), and exhibit high variability from one species to another within the same genus (Shearer, 1993; Frenken *et al.*, 2017; Van den Wingaert *et al.*, 2022).

The high quantity of SPCs (Specific Protein Clusters, (~ 43% of the CCs)) suggest that functional groups own numerous specific genes involved in distinct metabolic pathways and/or are affiliated to highly divergent taxonomy. Within these SPCs, an average of 47.5% of proteins are taxonomically affiliated and 55.6% are functionally annotated (**Figs. 4, S3**). While taxonomy appeared highly divergent depending on the functional group assignment (*e.g.*, Opisthokonta and Amoebozoa are absent from photosynthetic groups, while Haptista, on the contrary, exclusively comprise photosynthetic members (phototrophs or mixotrophs)), functional annotations varied much less, at least at a high level (*i.e.*, level A: KEGG_{PATHWAY} = Metabolism).

Categories taking part of primary metabolisms such as amino-acids, carbohydrate or lipid metabolic pathways remain stable across all functional groups, indicating a relative functional redundancy which guarantee the resilience and stability of ecosystem functioning.

These relatively low differences have to be considered taking into account that we neither examined relative expression of genes nor considered all genes but analyzed the richness of those mapped in metabolic pathways categories in KEGG database. Nevertheless, some functional variations turned out significant at finer level (*i.e.*, C: KEGG_{PATHWAY} = Metabolism) and appear to reflect the functional specificities of functional groups (**Fig. S4**). For instance, in line with their trophic abilities, phototrophs and mixotrophs are characterized by numerous genes involved in photosynthesis as well as in carotenoid biosynthesis. In contrast, strict heterotrophs exhibit a substantial portion of their genes dedicated to fatty acid processes, including sphingolipid and glycerolipid metabolism, along with fatty acid degradation which are likely associated with their phagocytosis activity (Lima *et al.*, 2017) and the synthesis of biological membranes (de Carvalho and Caramujo, 2018). Numerous genes involved in sulfur metabolism were found within saprotrophs suggesting their implication in the sulfur cycle. Sulfur cycle is an important biochemical cycle in lakes Pavin however the majority of researches involve bacterial species (Biderre-Petit *et al.*, 2011; Berg *et al.*, 2019). Yet, assimilation of conventional sources of sulfur by eukaryotes such as Fungi (almost exclusively) is already demonstrated (*e.g.*, in association with plant or involve in human pathologies) (Linder, 2018; Amich, 2022) but their real contribution within biogeochemical cycles in freshwater ecosystems remained to be explored (Krauss *et al.*, 2011). Finally, parasites have extensive genes involved in fatty acid metabolism, which is in line with the expectation that lipid metabolism is highly developed within parasites of highly divergent taxonomy (Ramakrishnan *et al.*, 2013; Bi *et al.*, 2016; Laundon *et al.*, 2022; Shunmugam *et al.*, 2022) in order to established host-parasite interactions and involving high diversity of genes.

C. Toward a functional marker of parasitism

Parasitic microorganisms are abundant and diverse in aquatic ecosystems however they remain enigmatic (role, life cycle) and understudied. This is primarily due to the challenges associated with culturing parasites in association with their hosts, their relatively small size, as well as with the challenge of isolating rare parasites with very few free-living forms (Bass *et al.*, 2015).

The use of environmental DNA (eDNA) offers therefore an alternative for assessing parasites diversities and abundance as well as potential function (Bass *et al.*, 2015). Here we took advantage of a metatranscriptomic dataset and the use of SSN to further explore parasitic microbial eukaryotes. Considering that SPCs enable the investigation of functional specificities within functional groups, it appears relevant to incorporate them into the process of identifying potential genetic markers associated to trophic modes. In this context, we focused on the 1 679 SPCs clustering at least one protein assigned to an obligate parasite, with the aim of identifying genes that are likely playing a role in host-parasite interactions and proposing potential marker candidates. Among the 12 946 proteins gathered in the selected SPCs, a majority are assigned to parasite (~ 71.2%) and a limited number are characterized by alternative assignments (*i.e.*, ~ 2% with 1.6% corresponding to saprotrophs). This suggest that parasite-assigned proteins may be relatively specific and thus that derived functions are not shared by other trophic modes. Facultative parasite-assigned proteins represent a small proportion (~ 9.2%) and are affiliated, mainly to Opisthokonta and Stramenopiles, as well as for a minority to Amoebozoa. Amoebozoa only include facultative representatives, which explains the low number of proteins in SPCs (**Figs. 1, S5**). Opisthokonta and Stramenopiles group together both obligate and facultative parasites indicating a relative genetic similarity between proteins of each trophic mode. Indeed, there is often a fine line between obligate, facultative and saprotrophic organisms and dissimilarity could be slight reporting multiple degrees depending on the nature of interaction (*e.g.*, Oomycetes and Fungi may be characterized as facultative and obligate parasites, biotrophic, necrotrophic, hemi-biotrophic or saprophytic) (Shearer, 1993; Zhao *et al.*, 2013; Frenken *et al.*, 2017; Fiore-Donno and Bonkowski, 2021; Van den Wyngaert *et al.*, 2022). In contrast, the others divisions (*i.e.*, mainly Alveolata, Euglenozoa, Metamonada and Rhizaria (**Fig. S5**)) are only represented in SPCs by proteins assigned to obligate parasites. This suggests that selected parasite-related proteins are involved in host-parasite interactions and have no homologue in free-living organisms of similar divisions which is consistent with the fact that these division gather distinct phylogenetic levels of organisms known to be exclusively obligate parasite such as Apicomplexa and Parabasalia phyla or Syndiniales, Plasmodiophorida and Trypanosomatida orders.

None SPCs grouping together all phyla with parasite relatives has been identified and therefore prevent from identifying a universal marker for parasitism. While it might be argued that a less stringent identity threshold could be used to enhance the identification of potential parasitism marker genes, there is uncertainty regarding the possibility of other trophic modes also being included in parasite-related cluster, undermining the effectiveness of the strategy.

Nevertheless, by examining link between KOs centroids and parasite-related SPCs (**Fig. 7**), we identified candidate genes that could be referenced as parasitism marker for specific phyla (**Figs. 7, S6, Table S2**).

We for example reported genes linked with the well-known mechanism of immunogenicity and extracellular survival of *Trypanosoma* coding for plasmanylethanolamine desaturase (involved in the biosynthesis of glycosylphosphatidylinositol which is known as variant surface glycoproteins (VSG) of Kinetoplastida (Menon *et al.*, 1993; Villas Bôas *et al.*, 1999; Hong and Kinoshita, 2009)) as well as encoding E3 ubiquitin-protein ligase (involved in SUMOylation process which positively regulate VSG expression (López-Farfán *et al.*, 2014)). We furthermore reported several genes linked to pathogenicity of fungi-like organisms such as MFS multidrug transporters or glyoxalase protein associated to toxic compounds detoxification (Ascomycota and Labyrinthulomycetes, respectively) (Chauhan and Madhubala, 2009; Costa *et al.*, 2014), as well as others corresponding to virulence factors in Ascomycota: zinc-cluster proteins (John *et al.*, 2022), sodium/hydrogen antiporter (Jimenez and Mesones, 2022), thiamine precursor (Jin *et al.*, 2021); or proline-specific permease in Oomycetes (Berg *et al.*, 2021; Garbe *et al.*, 2022) (**Table S2**).

Other genes, not yet described in a host-parasite interaction but coding for proteins whose function may coincide with parasitic lifestyle, such as anti-apoptotic agents have been highlighted for Ascomycota, Plasmodiophorida and Oomycetes. Genes involved in DNA repair process are found within Labyrinthulomycetes, Conoidasida, Ascomycota and Oomycetes and genes encoded for cytochrome P450 family, which is described as distributed within fungi-like organism depending on their trophic capacity (Sello *et al.*, 2015)), are found in Labyrinthulomycetes and Oomycetes. All the genes that are reported within parasite species that can cause diseases in humans or affect species of economic interest could represent a few interesting lines for consideration in a health or economic context for the development of antiparasitic treatment (Stijlemans *et al.*, 2007; Chauhan and Madhubala, 2009; Hong and Kinoshita, 2009; Warrilow *et al.*, 2014).

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 PRJEB61515.

Competing Interests

The authors declare no competing interests.

References

- Alexander, H., Jenkins, B.D., Rynearson, T.A., and Dyhrman, S.T. (2015) Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, E2182-E2190.
- Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., et al. (2021) A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology*, **39**, 105–114.
- Alvarez-Ponce, D., Lopez, P., Bapteste, E., and McInerney, J.O. (2013) Gene similarity networks provide tools for understanding eukaryote origins and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E1594-E1603.
- Amich, J. (2022) Sulfur metabolism as a promising source of new antifungal targets. *J Fungi (Basel)*, **8**, 295.
- Atkinson, H.J., Morris, J.H., Ferrin, T.E., and Babbitt, P.C. (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One*, **4**, e4345.
- Bass, D., Stentiford, G.D., Littlewood, D.T.J., and Hartikainen, H. (2015) Diverse applications of environmental DNA methods in parasitology. *Trends in Parasitology*, **31**, 499–513.
- Berg, J.A., Hermans, F.W.K., Beenders, F., Abedinpour, H., Vriezen, W.H., Visser, R.G.F., et al. (2021) The amino acid permease (AAP) genes CsAAP2A and SlAAP5A / B are required for oomycete susceptibility in cucumber and tomato. *Molecular Plant Pathology*, **22**, 658–672.
- Berg, J.S., Jézéquel, D., Duverger, A., Lamy, D., Laberty-Robert, C., and Miot, J. (2019) Microbial diversity involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of lake Pavin. *PLoS One*, **14**, e0212787.
- Bernard, G., Pathmanathan, J.S., Lannes, R., Lopez, P., and Bapteste, E. (2018) Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery. *Genome Biology and Evolution*, **10**, 707–715.
- Bi, K., He, Z., Gao, Z., Zhao, Y., Fu, Y., Cheng, J., et al. (2016) Integrated omics study of lipid droplets from *Plasmodiophora brassicae*. *Scientific Report*, **6**, 36965.
- Biderre-Petit, C., Boucher, D., Kuever, J., Alberic, P., Jézéquel, D., Chebance, B., et al. (2011) Identification of sulfur-cycle prokaryotes in a low-sulfate lake (lake Pavin) using *aprA* and 16S rRNA gene markers. *Microbial Ecology*, **61**, 313–327.
- Bittner, L., Halary, S., Payri, C., Cruaud, C., De Reviers, B., Lopez, P., and Bapteste, E. (2010) Some considerations for analyzing biodiversity using integrative metagenomics and gene networks. *Biology Direct*, **5**, 47.
- Bock, N.A., Charvet, S., Burns, J., Gyaltsen, Y., Rozenberg, A., Duhamel, S., and Kim, E. (2021) Experimental identification and in silico prediction of bacterivory in green algae. *The ISME Journal*, **15**, 1987–2000.
- Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**, 59–60.

- Callahan, H.A., Litaker, R.W., and Noga, E.J. (2002) Molecular taxonomy of the suborder bodonina (order Kinetoplastida), including the important fish parasite, *Ichthyobodo necator*. *The Journal of Eukaryotic Microbiology*, **49**, 119–128.
- Caputi, L., Carradec, Q., Eveillard, D., Kirilovsky, A., Pelletier, E., Pierella Karlusich, J.J., et al. (2019) Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochemical Cycles*, **33**, 391–419.
- 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*, **9**, 373.
- de Carvalho, C.C.C.R. and Caramujo, M.J. (2018) The various roles of fatty acids. *Molecules*, **23**, 2583.
- Chauhan, S.C. and Madhubala, R. (2009) Glyoxalase I gene deletion mutants of *Leishmania donovani* exhibit reduced methylglyoxal detoxification. *PLoS One*, **4**, e6805.
- 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, I.-M.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, **47**, D666–D677.
- Cheng, S., Karkar, S., Baptiste, E., Yee, N., Falkowski, P., and Bhattacharya, D. (2014) Sequence similarity network reveals the imprints of major diversification events in the evolution of microbial life. *Frontiers in Ecology and Evolution*, **2**, 72.
- Cohen, N.R., McIlvin, M.R., Moran, D.M., Held, N.A., Saunders, J.K., Hawco, N.J., et al. (2021) Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nature Microbiology*, **6**, 173–186.
- Copp, J.N., Akiva, E., Babbitt, P.C., and Tokuriki, N. (2018) Revealing unexplored sequence-function space using sequence similarity networks. *Biochemistry*, **57**, 4651–4662.
- Corel, E., Lopez, P., Méheust, R., and Baptiste, E. (2016) Network-thinking: graphs to analyze microbial complexity and evolution. *Trends in Microbiology*, **24**, 224–237.
- Costa, C., Dias, P.J., SÃ¡-Correia, I., and Teixeira, M.C. (2014) MFS multidrug transporters in pathogenic fungi: do they have real clinical impact? *Frontiers in Physiology*, **5**, 197.
- de Crécy-Lagard, V., Amorin de Hegedus, R., Arighi, C., Babor, J., Bateman, A., Blaby, I., et al. (2022) A roadmap for the functional annotation of protein families: a community perspective. *Database (Oxford)*, **2022**, baac062.
- Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, **1695**, 1–9.
- Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Fernandez-Guerra, A., et al. (2022) Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*, **2**, 100123.
- Dobson, A., Lafferty, K.D., Kuris, A.M., Hechinger, R.F., and Jetz, W. (2008) Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11482–11489.

- Dunne, J.A., Lafferty, K.D., Dobson, A.P., Hechinger, R.F., Kuris, A.M., Martinez, N.D., et al. (2013) Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biology*, **11**, e1001579.
- 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. (2019) Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 6211–6220.
- Ellens, K.W., Christian, N., Singh, C., Satagopam, V.P., May, P., and Linster, C.L. (2017) Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic Acids Research*, **45**, 11495–11514.
- Faure, E., Ayata, S.-D., and Bittner, L. (2021) Towards omics-based predictions of planktonic functional composition from environmental data. *Nature Communication*, **12**, 4361.
- Fiore-Donno, A.M. and Bonkowski, M. (2021) Different community compositions between obligate and facultative oomycete plant parasites in a landscape-scale metabarcoding survey. *Biology and Fertility of Soils*, **57**, 245–256.
- Forster, D., Bittner, L., Karkar, S., Dunthorn, M., Romac, S., Audic, S., et al. (2015) Testing ecological theories with sequence similarity networks: marine ciliates exhibit similar geographic dispersal patterns as multicellular organisms. *BMC Biology*, **13**, 16.
- Franzosa, E.A., McIver, L.J., Rahnava, G., Thompson, L.R., Schirmer, M., Weingart, G., et al. (2018) Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, **15**, 962–968.
- Frenken, T., Alacid, E., Berger, S.A., Bourne, E.C., Gerphagnon, M., Grossart, H.-P., et al. (2017) Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. *Environmental Microbiology*, **19**, 3802–3822.
- Garbe, E., Miramón, P., Gerwien, F., Ueberschaar, N., Hansske-Braun, L., Brandt, P., et al. (2022) GNP2 encodes a high-specificity proline permease in *Candida albicans*. *mBio*, **13**, e03142-21.
- Garvetto, A., Nézan, E., Badis, Y., Bilien, G., Arce, P., Bresnan, E., et al. (2018) Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudonitzschia*: genetic, morphological, and ecological characterisation. *Frontiers in Microbiology*, **9**, 2918.
- Grossmann, L., Beisser, D., Bock, C., Chatzinotas, A., Jensen, M., Preisfeld, A., et al. (2016) Trade-off between taxon diversity and functional diversity in European lake ecosystems. *Molecular Ecology*, **25**, 5876–5888.
- He, H., Huang, J., Wu, S., Jiang, S., Liang, L., Liu, Y., et al. (2021) The roles of GTPase-activating proteins in regulated cell death and tumor immunity. *Journal of Hematology & Oncology*, **14**, 171.
- Hong, Y. and Kinoshita, T. (2009) Trypanosome glycosylphosphatidylinositol biosynthesis. *The Korean Journal of Parasitology*, **47**, 197.

- Hornung, B.V.H. and Terrapon, N. (2023) An objective criterion to evaluate sequence-similarity networks helps in dividing the protein family sequence space. *PLoS Computational Biology*, **19**, e1010881.
- 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.
- Hudson, P.J., Dobson, A.P., and Lafferty, K.D. (2006) Is a healthy ecosystem one that is rich in parasites? *Trends in Ecology & Evolution*, **21**, 381–385.
- Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., and Bork, P. (2017) Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, **34**, 2115–2122.
- James, T.Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., and Stajich, J.E. (2013) Shared signatures of parasitism and phylogenomics unite Cryptomycota and Microsporidia. *Current Biology*, **23**, 1548–1553.
- Jimenez, V. and Mesones, S. (2022) Down the membrane hole: Ion channels in protozoan parasites. *PLoS Pathogens*, **18**, e1011004.
- Jin, D., Sun, B., Zhao, W., Ma, J., Zhou, Q., Han, X., et al. (2021) Thiamine-biosynthesis genes Bbpyr and Bbthi are required for conidial production and cell wall integrity of the entomopathogenic fungus *Beauveria bassiana*. *Journal of Invertebrate Pathology*, **184**, 107639.
- Jobard, M., Wawrzyniak, I., Bronner, G., Marie, D., Vellet, A., Sime-Ngando, T., et al. (2019) Freshwater Perkinsea: diversity, ecology and genomic information. *Journal of Plankton Research*, **42**, 3–17.
- John, E., Singh, K.B., Oliver, R.P., and Tan, K.-C. (2022) Transcription factor lineages in plant-pathogenic fungi, connecting diversity with fungal virulence. *Fungal Genetics and Biology*, **161**, 103712.
- 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.
- Kolody, B.C., McCrow, J.P., Allen, L.Z., Aylward, F.O., Fontanez, K.M., Moustafa, A., et al. (2019) Diel transcriptional response of a California Current plankton microbiome to light, low iron, and enduring viral infection. *The ISME Journal*, **13**, 2817–2833.
- Krauss, G.-J., Solé, M., Krauss, G., Schlosser, D., Wesenberg, D., and Bärlocher, F. (2011) Fungi in freshwaters: ecology, physiology and biochemical potential. *FEMS Microbiology Reviews*, **35**, 620–651.
- Labarre, A., Obiol, A., Wilken, S., Forn, I., and Massana, R. (2020) Expression of genes involved in phagocytosis in uncultured heterotrophic flagellates. *Limnology and Oceanography*, **65**, 149–160.
- Lafferty, K.D., Allesina, S., Arim, M., Briggs, C.J., De Leo, G., Dobson, A.P., et al. (2008) Parasites in food webs: the ultimate missing links. *Ecology Letters*, **11**, 533–546.

- Lambert, B.S., Groussman, R.D., Schatz, M.J., Coesel, S.N., Durham, B.P., Alverson, A.J., et al. (2021) The dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *Proceedings of the National Academy of Sciences of the United States of America*, **119**, e2100916119.
- Lampe, R.H., Mann, E.L., Cohen, N.R., Till, C.P., Thamatrakoln, K., Brzezinski, M.A., et al. (2018) Different iron storage strategies among bloom-forming diatoms. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E12275-E12284.
- Laundon, D., Christmas, 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.
- 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, L., Delgado-Viscogliosi, P., Gerphagnon, M., Viscogliosi, E., Christaki, U., Sime-Ngando, T., and Monchy, S. (2022) Taxonomic and functional dynamics during chytrid epidemics in an aquatic ecosystem. *Molecular Ecology*, **31**, 5618–5634.
- Lima, S., Milstien, S., and Spiegel, S. (2017) Sphingosine and sphingosine kinase 1 involvement in endocytic membrane trafficking. *Journal of Biological Chemistry*, **292**, 3074–3088.
- Lin Pedersen, T. (2022) ggraph: An implementation of grammar of graphics for graphs and networks.
- Lin Pedersen, T. (2023) tidygraph: A tidy API for graph manipulation.
- Linder, T. (2018) Assimilation of alternative sulfur sources in fungi. *World Journal of Microbiology and Biotechnology*, **34**, 51.
- López-Farfán, D., Bart, J.-M., Rojas-Barros, D.I., and Navarro, M. (2014) SUMOylation by the E3 ligase TbSIZ1/PIAS1 positively regulates VSG expression in *Trypanosoma brucei*. *PLoS Pathogens*, **10**, e1004545.
- Massana, R., Labarre, A., López-Escardó, D., Obiol, A., Bucchini, F., Hackl, T., et al. (2021) Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate. *The ISME Journal*, **15**, 154–167.
- de Meeûs, T. and Renaud, F. (2002) Parasites within the new phylogeny of eukaryotes. *Trends in Parasitology*, **18**, 247–251.
- Méheust, R., Zelzion, E., Bhattacharya, D., Lopez, P., and Bapteste, E. (2016) Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 3579–3584.
- Meng, A., Corre, E., Probert, I., Gutierrez-Rodriguez, A., Siano, R., Annamale, A., et al. (2018) Analysis of the genomic basis of functional diversity in dinoflagellates using a transcriptome-based sequence similarity network. *Molecular Ecology*, **27**, 2365–2380.
- Menon, A.K., Eppinger, M., Mayor, S., and Schwarz, R.T. (1993) Phosphatidylethanolamine is the donor of the terminal phosphoethanolamine group in trypanosome glycosylphosphatidylinositols. *The EMBO Journal*, **12**, 1907–1914.

- Mitchell, A.L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., et al. (2019) MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Research*, **48**, D570–D578.
- 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.
- Monjot, A., Bronner, G., Courtine, D., Cruaud, C., Da Silva, C., Aury, J.-M., et al. (2023) Functional diversity of microbial eukaryotes in a meromictic lake ecosystem: coupling between metatranscriptomic and a trait-based approach. *Environmental Microbiology*.
- Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061–1067.
- Pearson, G.A., Lago-Leston, A., Cánovas, F., Cox, C.J., Verret, F., Lasternas, S., et al. (2015) Metatranscriptomes reveal functional variation in diatom communities from the Antarctic Peninsula. *The ISME Journal*, **9**, 2275–2289.
- Pierella Karlusich, J.J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., et al. (2023) A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Molecular Ecology Resources*, **23**, 16–40.
- Popova, O.V., Belevich, T.A., Golyshev, S.A., Kireev, I.I., and Aleoshin, V.V. (2020) *Labyrinthula diatomea* n. sp. - A labyrinthulid associated with marine diatoms. *The Journal of Eukaryotic Microbiology*, **67**, 393–402.
- Potting, C., Tatsuta, T., König, T., Haag, M., Wai, T., Aaltonen, M.J., and Langer, T. (2013) TRIAP1/PRELI complexes prevent apoptosis by mediating intramitochondrial transport of phosphatidic acid. *Cell Metabolism*, **18**, 287–295.
- Promponas, V.J., Iliopoulos, I., and Ouzounis, C.A. (2015) Annotation inconsistencies beyond sequence similarity-based function prediction – phylogeny and genome structure. *Standards in Genomic Sciences*, **10**, 108.
- Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017) Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, **35**, 833–844.
- Quiroz Velasquez, P.F., Abiff, S.K., Fins, K.C., Conway, Q.B., Salazar, N.C., Delgado, A.P., et al. (2014) Transcriptome analysis of the entomopathogenic oomycete *Lagenidium giganteum* reveals putative virulence factors. *Applied and Environmental Microbiology*, **80**, 6427–6436.
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing.
- Ramakrishnan, S., Serricchio, M., Striepen, B., and Bütikofer, P. (2013) Lipid synthesis in protozoan parasites: a comparison between kinetoplastids and apicomplexans. *Progress in Lipid Research*, **52**, 488–512.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.-H., Parvez, M., et al. (2015) Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific Reports*, **5**, 11572.
- Shearer, C. (1993) The freshwater Ascomycetes. *Nova Hedwigia*, **56**, 1–33.
- Shunmugam, S., Arnold, C.-S., Dass, S., Katris, N.J., and Botté, C.Y. (2022) The flexibility of Apicomplexa parasites in lipid metabolism. *PLoS Pathogens*, **18**, e1010313.

- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.
- Smolarz, B., Wilczyński, J., and Nowakowska, D. (2014) DNA repair mechanisms and *Toxoplasma gondii* infection. *Archives of Microbiology*, **196**, 1–8.
- So, Y.-S., Maeng, S., Yang, D.-H., Kim, H., Lee, K.-T., Yu, S.-R., et al. (2019) Regulatory mechanism of the atypical AP-1-like transcription factor Yap1 in *Cryptococcus neoformans*. *mSphere*, **4**, e00785-19.
- Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35**, 1026–1028.
- Stijlemans, B., Baral, T.N., Guilliams, M., Brys, L., Korf, J., Drennan, M., et al. (2007) A glycosylphosphatidylinositol-based treatment alleviates trypanosomiasis-associated immuno-pathology. *The Journal of Immunology*, **179**, 4003–4014.
- 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 zoospore fungi: linking sequence data to ecological traits. *The ISME Journal*, **16**, 2242–2254.
- Vanni, C., Schechter, M.S., Acinas, S.G., Barberán, A., Buttigieg, P.L., Casamayor, E.O., et al. (2022) Unifying the known and unknown microbial coding sequence space. *eLife*, **11**, e67667.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**, 1261605–1261605.
- Villas Bôas, M.H.S., Lara, L.S., Wait, R., and Barreto Bergter, E. (1999) Identification of plasmenylethanolamine as a major component of the phospholipids of strain DM 28c of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology*, **98**, 175–186.
- Ward, B.A. (2019) Mixotroph ecology: More than the sum of its parts. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 5846–5848.
- Warrilow, A.G.S., Hull, C.M., Rolley, N.J., Parker, J.E., Nes, W.D., Smith, S.N., et al. (2014) Clotrimazole as a potent agent for treating the oomycete fish pathogen *Saprolegnia parasitica* through inhibition of sterol 14 α -demethylase (CYP51). *Applied and Environmental Microbiology*, **80**, 6154–6166.
- Wyman, S.K., Avila-Herrera, A., Nayfach, S., and Pollard, K.S. (2018) A most wanted list of conserved microbial protein families with no known domains. *PLoS One*, **13**, e0205749.
- Yunes, J.M. and Babbitt, P.C. (2019) Effusion: prediction of protein function from sequence similarity networks. *Bioinformatics*, **35**, 442–451.
- Zhao, Z., Liu, H., Wang, C., and Xu, J.-R. (2014) Correction: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, **15**, 6.

VI. Supplementary Materials

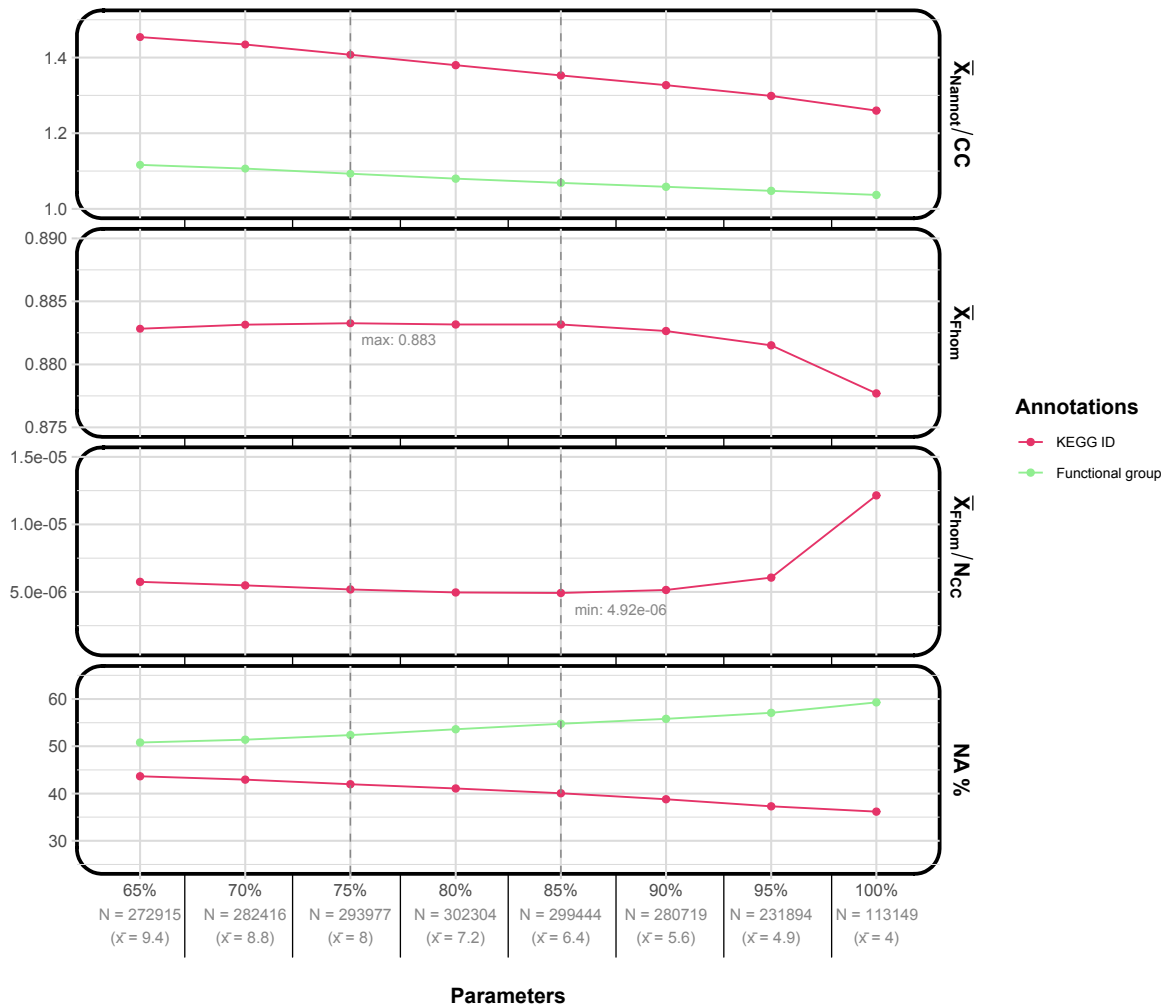


Figure S1: Statistics of sequence similarity clusterisation using different similarity threshold for each type of annotations (*i.e.*, functional annotation or functional group assignment).

On the right, from the top to the bottom: $\bar{X}_{Nannot/CC}$ = Mean of distinct annotations number by CC; \bar{X}_{Fhom} = Mean of annotation scores; $\bar{X}_{Fhom/NCC}$ = Mean of annotation scores divided by the total number of CCs; NA% = proportion of CCs clustering only unannotated proteins.

At the bottom: N=total number of CCs; \bar{X} = mean number of proteins by CC.

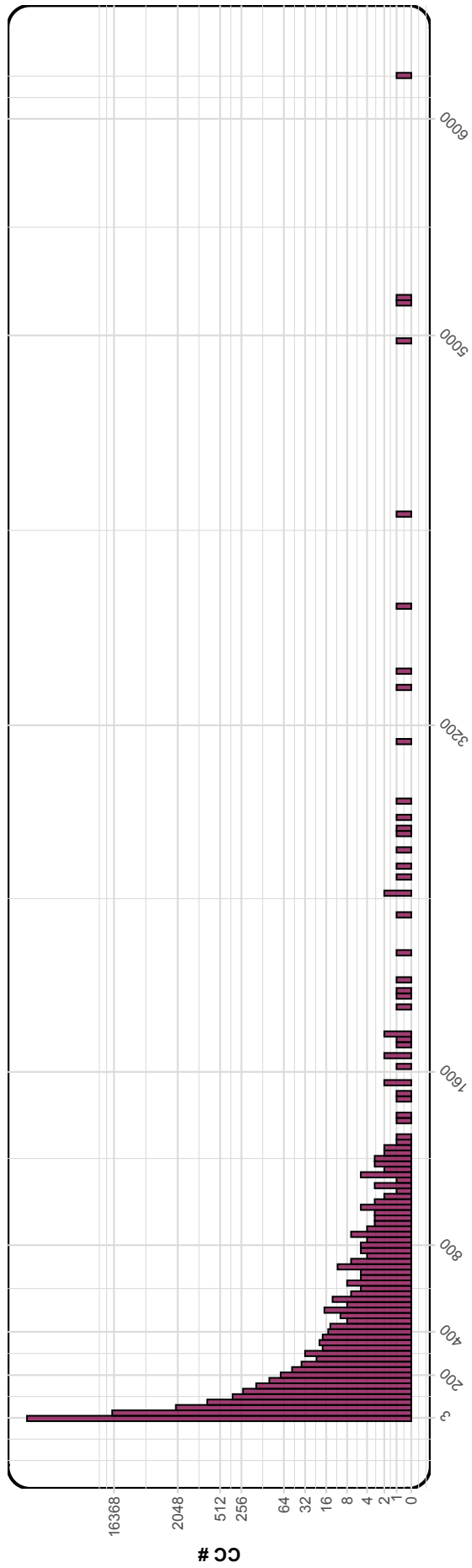


Figure S2: Histogram reporting number of CCs in relation with their protein's abundance.

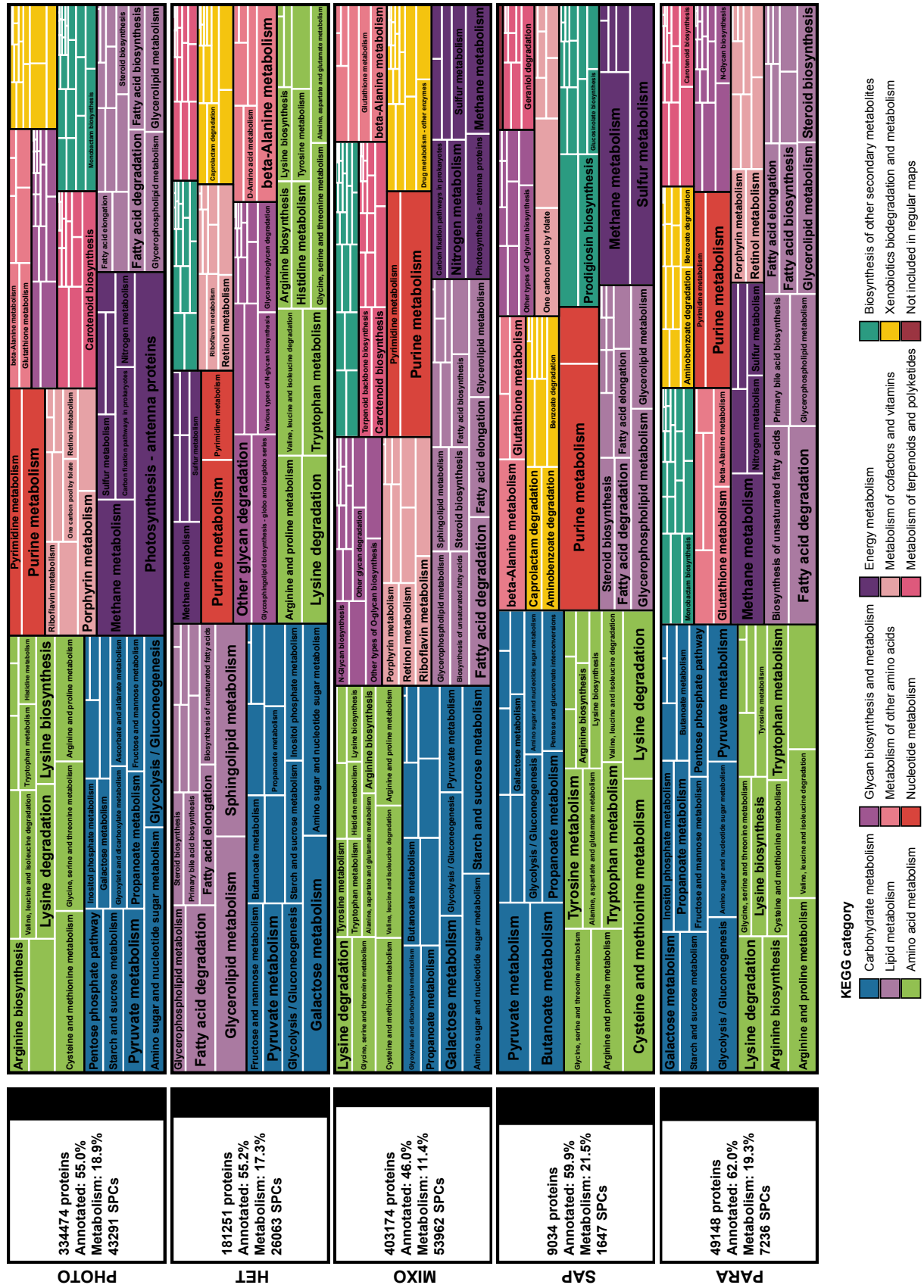


Figure S3: Functional annotation of proteins gathered in Specific Proteins Clusters of each functional group.

The number of SPCs, number of proteins within SPCs, annotation rate against KEGG database as well as proportion of proteins annotated with KEGG metabolism category ($KEGG_{PATHWAY} = \text{Metabolism}$) are displayed on the horizontal histograms for each functional group. Functional groups are identified using the same code as in Figure 1.

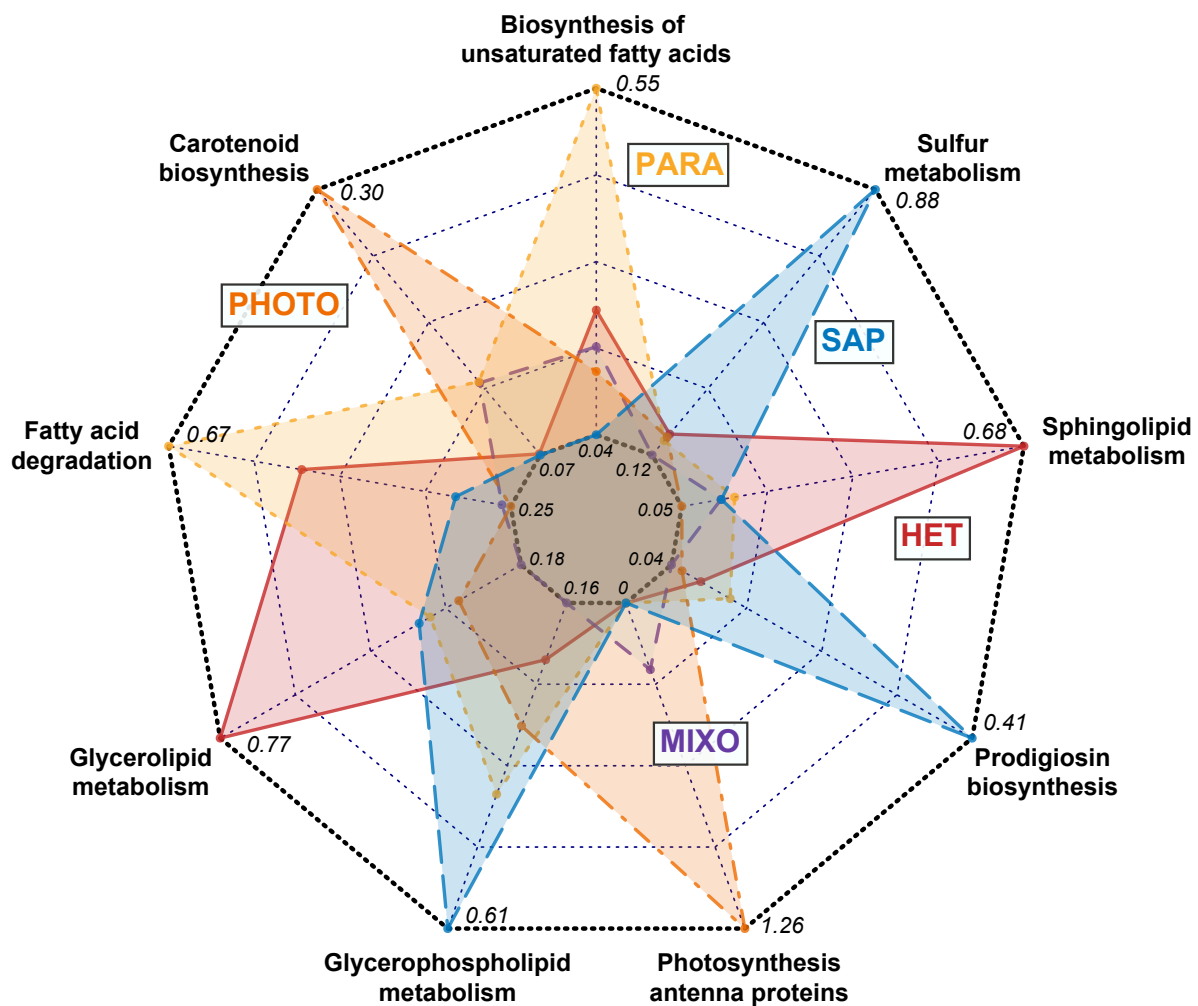


Figure S4: Proportion of proteins annotated with the most fluctuating KEGG metabolic pathways for each functional group.

Values at the center are the minimum proportion value for this KEGG category while those at the exterior represent maximum proportion value. Functional groups are identified using the same code as in Figure 1.

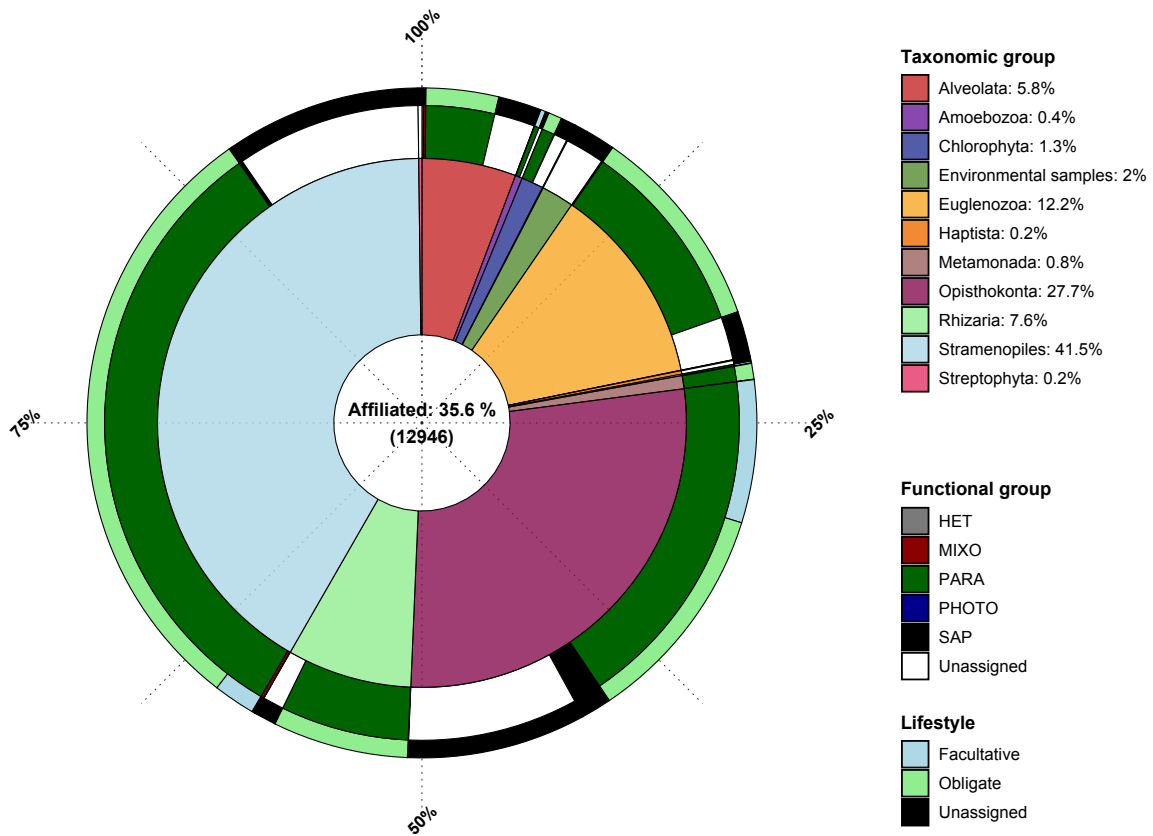


Figure S5: Taxonomic diversity and functional group assignment of proteins gathered in SPCs characterized by at least one protein assigned to obligate parasite.

From inner to outer circles, colors represent taxonomic affiliation, functional group assignment and lifestyle. Functional groups are identified using the same code as in Figure 1.

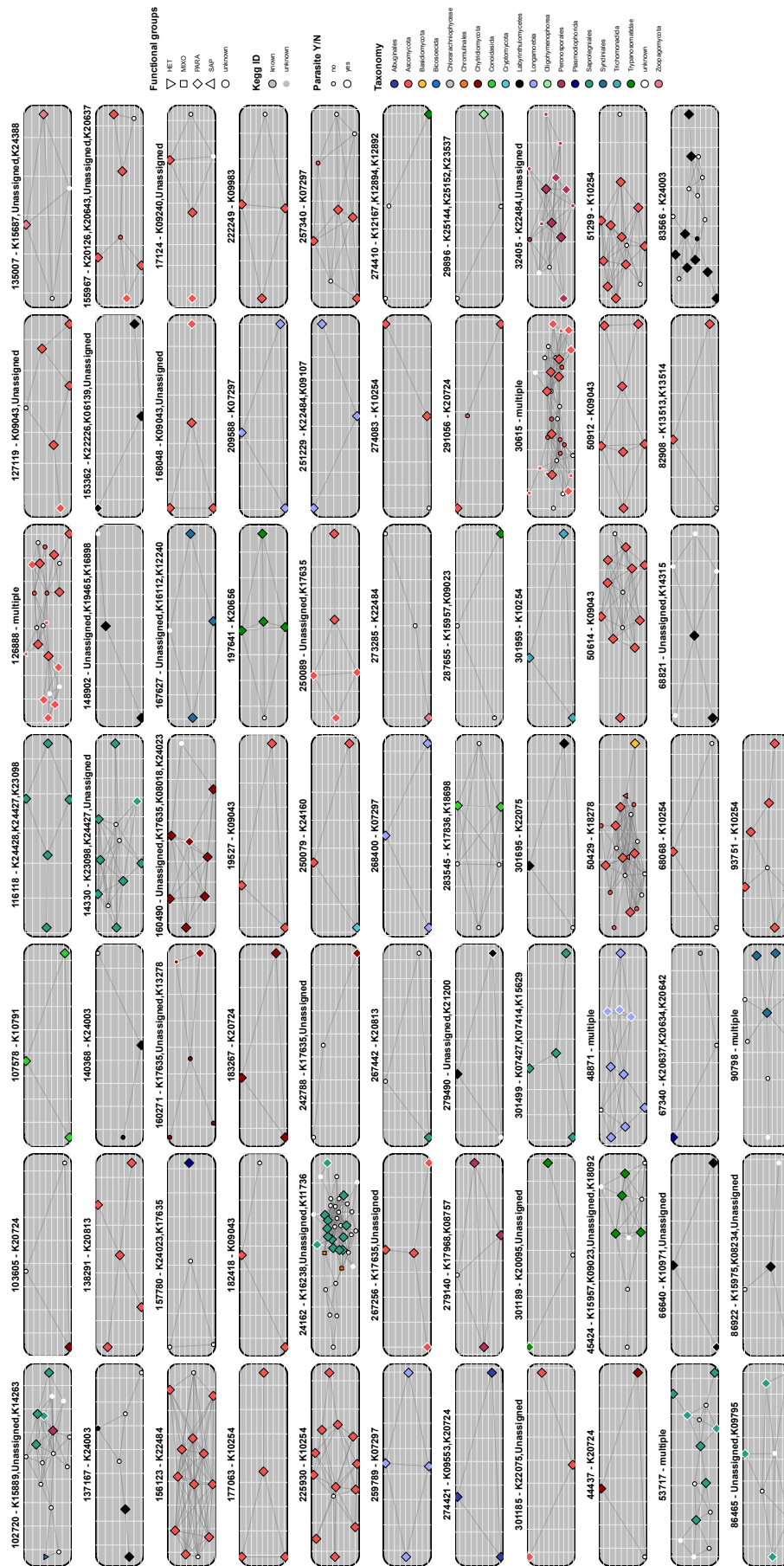


Figure S6: SPCs representation gathering proteins annotated with KO Id which centroids are highly correlated to parasitism.

The mention “multiple” means that proteins of the SPC are annotated with more than 3 KO Id. The identification number of SPCs are referenced on the top of each sub-graph. Black outlined points represent a protein annotated with the KO Id referenced at the top of the sub-graph. Functional groups are identified using the same code as in Figure 1.

Table S1: Association table linking taxonomy to functional group (accessible online at: <https://zenodo.org/doi/10.5281/zenodo.10035567>).

Taxonomic ranks are specified with the following code: d_(Division), k_(Kingdom), p_(Phylum), c_(Class), o_(Order), f_(Family) and g_(Genus). Functional groups are identified using the same code as in Figure 1.

Table S2: Function and potential link with parasitic lifestyle of KO Id which coordinates of centroids are highly correlated to parasitism (accessible online at: <https://zenodo.org/doi/10.5281/zenodo.10035567>).

CHAPITRE IV

DÉCOUVERTE D'UNE NOUVELLE ASSOCIATION
HÔTE-MICROSPORIDIE DANS UN LAC D'EAU DOUCE
PAR « SINGLE CELL »

Contexte de l'étude

Les chapitres II et III confirment l'importance des parasites (notamment affiliés aux Fungi) dans les écosystèmes lacustres, aussi bien en termes de diversité fonctionnelle que d'abondance relative. Cependant, la diversité des parasites obtenue dans le chapitre II est bien sous-évaluée essentiellement en raison de biais méthodologiques (amorces utilisées, extraction de spores difficile, etc.). En effet, de par leurs divergences génétiques, certains groupes de parasites sont régulièrement absents des études visant à décrire la diversité des eucaryotes microbiens (Bass *et al.*, 2015). C'est par exemple le cas des microsporidies, que nous n'avons pas pu identifier dans le chapitre II, malgré une importante diversité de microorganismes eucaryotes mise en évidence dans le lac Pavin (3 358 ASVs). Ces parasites obligatoires apparentés aux Fungi renferment pourtant une diversité importante récemment mise en lumière dans les écosystèmes marins et lacustres, et sont impliqués dans des associations hôte-parasites pouvant impacter grandement les dynamiques des communautés planctoniques (annexe 1, Chauvet *et al.*, 2022). L'étude présentée en annexe 1 (dont je suis second auteur) ainsi que le travail publié par Chauvet *et al.* (2022) montrent en effet, grâce à l'utilisation d'une sonde spécifique des microsporidies, que celles-ci sont capables de parasiter des rotifères et des dinoflagellés avec des prévalences importantes (38% de rotifères infectés en moyenne au lac d'Aydat, Chauvet *et al.* (2022) ; 20 à 49% de dinoflagellés infectés dans différents écosystèmes lacustres et marins, annexe 1). Suite à ces résultats, nos objectifs étaient de développer d'autres types d'approches ayant pour but (i) la mise en évidence ou la confirmation de nouvelles associations hôte-microsporidie et (ii) la caractérisation morphologique et moléculaire des deux partenaires. Nous avons pour cela prélevé puis isolé manuellement un à un des rotifères (lac Aydat) et des dinoflagellés (lac Léman) de genres différents. Les individus uniques ont ensuite été criblés pour détecter la présence de microsporidies grâce à des amorces spécifiques. Les associations hôte-parasites ont également été confirmées par l'approche d'hybridation *in situ* (TSA-FISH). En l'absence de résultats concernant les dinoflagellés, seuls ceux concernant les rotifères seront présentés dans ce chapitre.

Single cell analysis reveals a new Microsporidia-host association in a freshwater lake

Arthur Monjot^{1*}, Marina Chauvet^{1*}, Anne Moné¹, Cécile Lepère^{1°}

1- CNRS, Laboratoire Microorganismes : Génome et Environnement, Université Clermont Auvergne, Clermont-Ferrand, F-63000, France.

* These authors contributed equally to this work

°corresponding author: Cecile.Lepere@uca.fr

Publié dans *Microbial Ecology* – Mai 2022

doi: 10.1007/s00248-022-02023-7

I. Abstract

Microsporidia are a large group of obligate intracellular eukaryotic parasites. Recent studies suggest that their diversity can be huge in freshwater lake ecosystems especially in the <150 µm size fraction. However, little is known about their hosts and therefore their impact on the trophic food web functioning. In this study, single cell analysis and fluorescence microscopy were used to detect new host-parasite association within rotifer communities in lake Aydat (France). Our analysis showed the existence of a potential new species belonging to the *Crispospora* genus able of infecting the rotifer *Kellicottia* with a high prevalence (42.5%) suggesting that Microsporidia could have a great impact on the rotifer populations' regulation in lakes.

Keywords: Microsporidia, Host-parasite association, Single cell, TSA-FISH, Lake, Microbial ecology

II. Note

Microsporidia are obligate intracellular eukaryotic parasites related to Fungi. They are mostly known to parasitize a broad range of hosts among the animal kingdom in different environments (Murareanu *et al.*, 2021). Recent metabarcoding studies revealed an unexpected diversity in lake ecosystems especially in the <150 µm size fraction suggesting that Microsporidia could have an impact on the planktonic communities at the base of the food web (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022). However, they have rarely been described as parasites of protists and microzooplankton. In marine environments, they are mainly known to parasitize gregarines (*e.g.*, (Mikhailov *et al.*, 2016)), but they were also shown in association with Paramyxea (Stentiford *et al.*, 2017). In freshwaters, only a few observations showed Microsporidia parasiting ciliates with only two microsporidian species characterized by morphological and molecular approaches (*Euplotespora binucleata* and *Globosporidium paramecii* parasitizing the ciliates *Euplotes woodruffi* and *Paramecium aurelia*, respectively (Fokin *et al.*, 2008; Yakovleva *et al.*, 2020)). Among microzooplankton, Microsporidia have been observed in the rotifers *Brachionus calicyflorus* as well as in *Asplanchna priodonta* and *A. brightwellii* in both river and freshwater lakes (Gorbunov and Kosova, 2001; Wolska and Mazurkiewicz-Zapalowicz, 2013); however these observations (no molecular data) remain sporadic and for most at low prevalence. Moreover, with co-occurrence network and microscopic observations, a more recent study suggested that Microsporidia could infect a broader range of rotifers (Chauvet *et al.*, 2022). We therefore used, in this study, single cell analysis as well as Tyramide Signal Amplification – Fluorescent *In Situ* Hybridization (TSA-FISH) to detect new host-parasite association within microzooplankton communities in a lake ecosystem (lake Aydat, Massif Central, France) and especially (i) confirm the hypothesis raised about the broader range of rotifers parasitized by Microsporidia and (ii) characterize by molecular information and microscopic observation both partners.

Zooplankton was sampled with vertical tows of 20 µm-mesh plankton net at the central point of the lake (depth = 14 m) on the 18/05/2021 (see Supplementary Material). 17 rotifers, identified morphologically as probably belonging to the genera *Kellicottia* (n = 13) and *Keratella* (n = 4), were isolated by micromanipulation and individually placed in a 96 wells plate before Whole Genome Amplification (WGA) (see Supplementary Material).

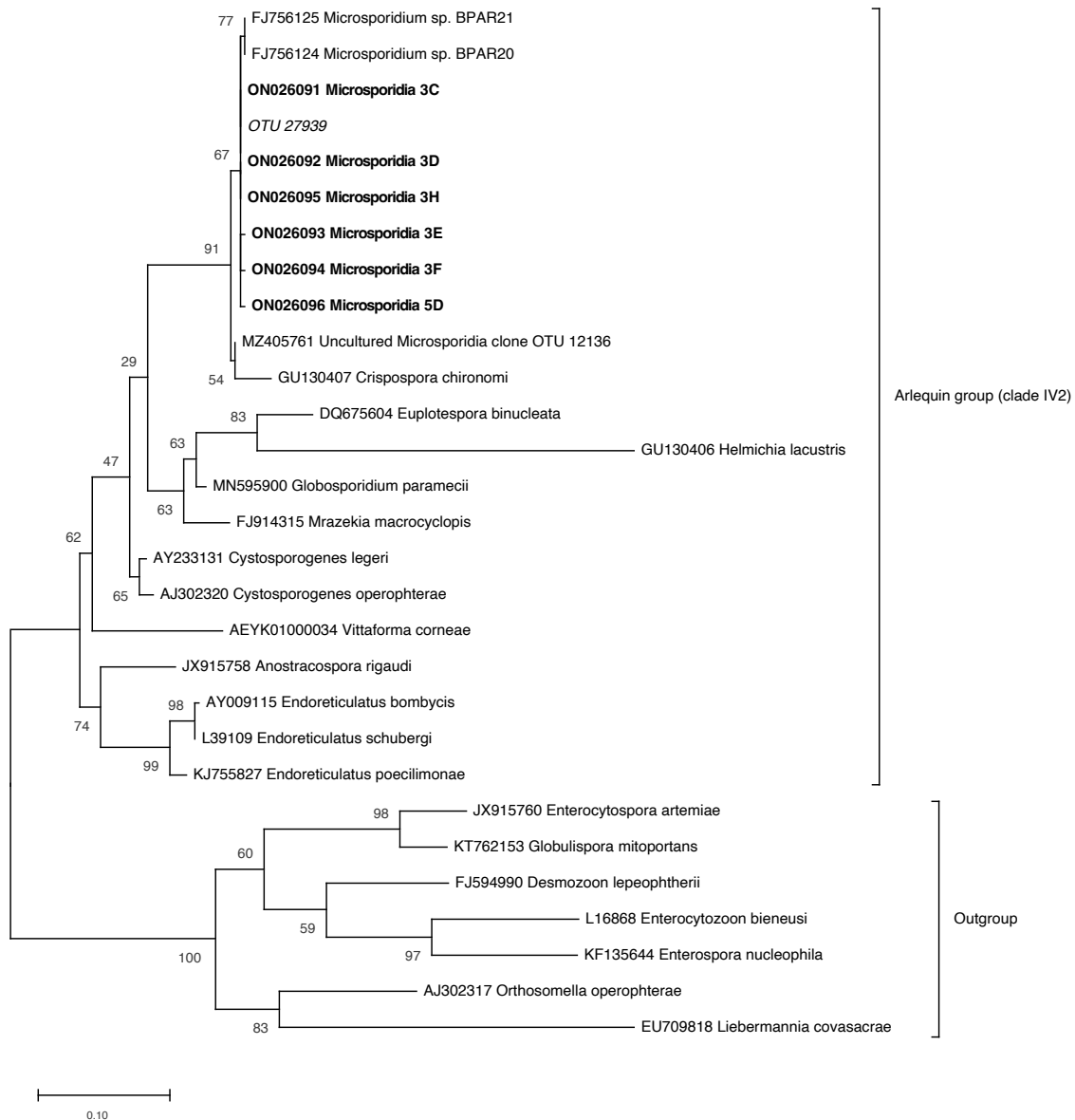


Figure 1: Maximum likelihood phylogeny of microsporidian sequences.

We included in this phylogeny several sequences from undescribed microsporidian species (found through BLAST search) and a sequence found extensively in a previous study (in italic, (Chauvet *et al.*, 2022)), in addition to described species, all belonging to the Arlequin group (clade IV2). Sequences from this study are given in bold. Outgroup species are representative species of other groups belonging to clade IV2. Numbers next to branches indicate the ML bootstrap support (100 replicates).

The Single Amplified Genomes (SAGs) were then screened for the presence of Microsporidia by the amplification of the V1-V3 region of the microsporidian gene coding for the SSU rRNA according to Chauvet *et al.* (2022). The sequences obtained after cloning and sequencing were compared to those available in public databases and included in a phylogenetic tree (see Supplementary Material). This analysis showed that on the 17 isolated rotifers, 6 *Kellicottia* had a positive PCR amplification signal with Microsporidia primers however no Microsporidia amplification was recorded in *Keratella*. All Microsporidia sequences were affiliated to the genus *Crispospora* (**Fig. 1**) which is affiliated to the clade IV2; Arlequin group (defined in Dubuffet *et al.* (2021)), and well represented (~ 50% of the total OTUs) in metabarcoding studies conducted in lake Aydat (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022). With 97% identity with *Crispospora chironomi* (the only known species of the genus so far) these sequences could represent a new species of the *Crispospora* genus (**Fig. S2A**). Indeed, although thresholds for species delineation may vary, an identity close to 100% is suggested for taxa separation at the species level for Microsporidia (Dubuffet *et al.*, 2021; Tokarev *et al.*, 2020). *Crispospora chironomi* is known to parasitize chironomids eggs (Tokarev *et al.*, 2010) that can be found in aquatic environments especially when they are eutrophic and anthropized (Langdon *et al.*, 2006).

On the Microsporidia positive SAGs, rotifers were characterized by amplification of the COI using the 30F and 885R primers set according to Zhang *et al.* (2021a) with some modifications (see Supplementary Material). The rotifer affiliation analysis revealed that this potential new species of *Crispospora* is able to parasitize the genus *Kellicottia* (**Fig. S1**). Even though *Kellicottia* sequences grouped into two clades in the phylogenetic analysis, species determination is challenging due to high intraspecific variation of identity percent which can be higher than 20% for rotifers (**Fig. S2B**) (Zhang *et al.*, 2021b).

To visualize and quantify Microsporidia-rotifer associations, TSA-FISH was performed on the rest of the plankton net sample, using a Microsporidia specific probe (USP01) according to Chauvet *et al.* (2022) (see Supplementary Material). Positive signals were obtained in *Kellicottia* rotifers with a surprisingly high prevalence (42.5%, total rotifers observed = 73) (**Fig. 2**). Microsporidia were so far recorded in a small number of rotifers (Gorbunov and Kosova, 2001; Wolska and Mazurkiewicz-Zapalowicz, 2013).

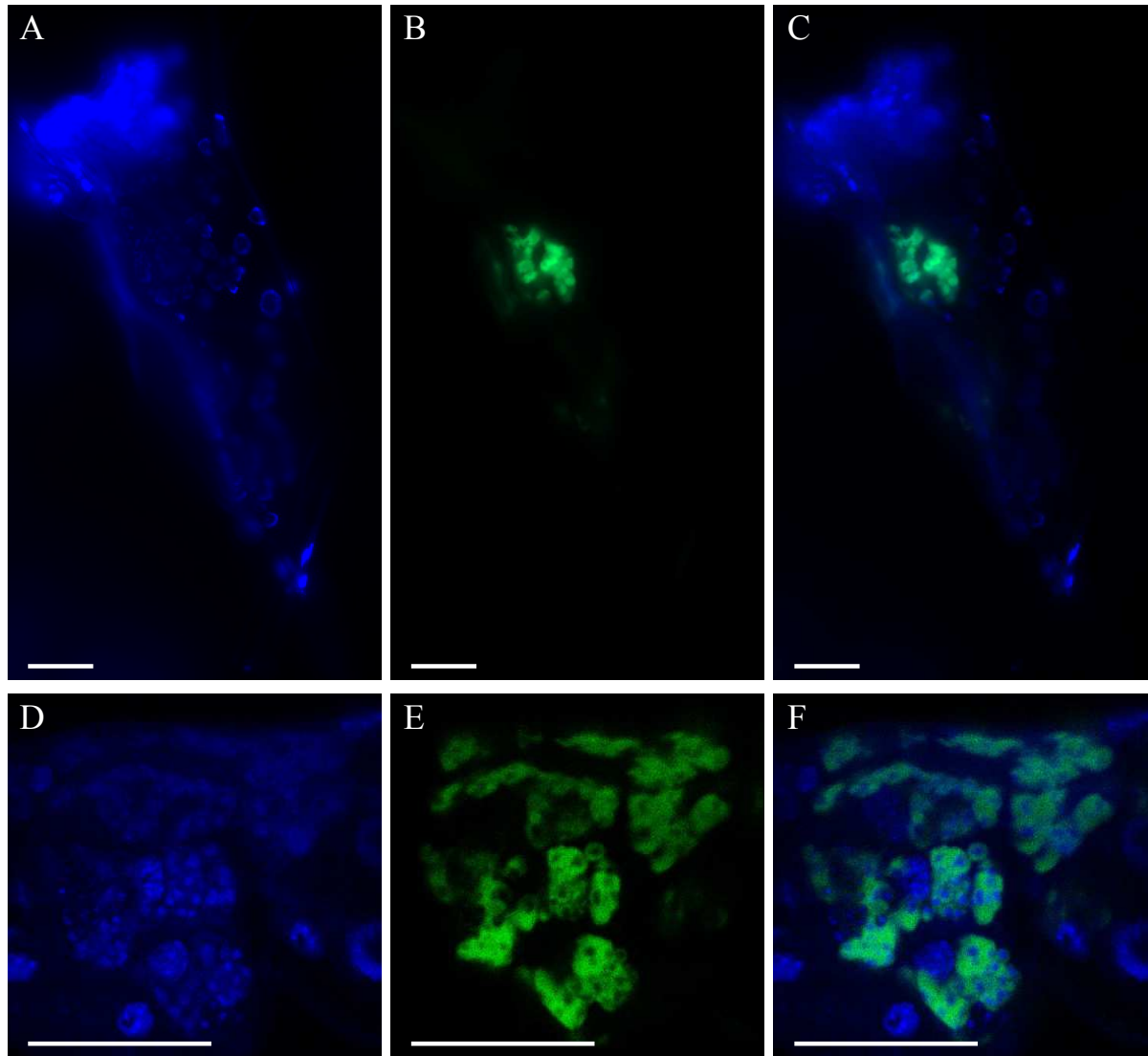


Figure 2: Microsporidia (targeted by TSA-FISH with the USP01 probe, (Chauvet *et al.*, 2022)) infecting a rotifer (genus *Kellicottia*).

Under UV showing nucleus after DAPI staining (A), under blue light showing Microsporidia in green fluorescence (B), overlay (C). Magnification on microsporidian cells inside a rotifer with a confocal microscope (UV (D), blue light (E), overlay (F)). Scale bars, 20 μm

Microsporidium sp. and *Microsporidium asperospora* have been observed in *Brachionus calyciflorus*, and *Microsporidium polygona* in *Asplanchna priodonta* and *A. brightwellii* in freshwater ecosystems most of the time at low prevalence (e.g., *Brachionus calyciflorus* by *Microsporidium* sp. on 1.2% of the host population (Wolska and Mazurkiewicz-Zapalowicz, 2013)). However, these descriptions were only morphological and neither the site of infection nor the microsporidian spore or intracellular form were described. Here, we identified an infection which could be in the rotifer's stomach or pseudocoelom. The intracellular forms of the Microsporidia were on average 1.53 μm long, most being oval, some spherical. These findings were consistent with the described shape of the only known *Crispospora* species (i.e., *Crispospora chironomi*), with two shapes identified: short oval and spherical, with a size of 1.5x2.5 μm and 1.5x2 μm , respectively (Tokarev *et al.*, 2010). This finding is highly relevant since Microsporidia, along with hyphomycetes and oomycetes (Barron and Szijarto, 1990; Thomas *et al.*, 2011) could be an important regulator of these populations in lakes. Rotifers are indeed an important component of plankton in aquatic ecosystems playing fundamental role in both microbial loop and classical food web and as a link in energy flow between primary producers (algae, cyanobacteria) and higher trophic levels (zooplankton, shellfish, fish) (Rico-Martínez *et al.*, 2016). *Kellicottia* can reach 600 cell/l in lake Aydat and is the most abundant rotifers along with *Keratella* (Lair and Oulad Ali, 1990). Microsporidia sequences retrieved during this study had an identity >99% with a highly abundant OTU identified in the same lake in a previous study where the temporal dynamic of the Microsporidia diversity was analyzed over one year (Chauvet *et al.*, 2022) (Fig. S2A). This *Crispospora* OTU and the only OTU affiliated to *Kellicottia* presented the same pattern of temporal variations (data not shown). This co-variation and the high prevalence of this association suggest an impact on the dynamic of *Kellicottia* in the lake however this needs to be confirmed for example by the culture of both partners which will also allow the characterization of the developmental cycle of this potential new species of Microsporidia.

Acknowledgements

The authors are very grateful to Jonathan Colombet and Eugénie Carriere for their help with sampling. We also gratefully acknowledge the platform CLIC (CLermont-ferrand Imagerie Confocale, Université Clermont Auvergne) and especially Caroline Vachias for her assistance with the confocal microscopy. The work of MC and AM was supported by two PhD fellowships funded by the 'Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation'.

Data Availability

All data and material are available in this publication and additional information online (supplementary material and supplementary figures). Sequences are available on GenBank, accession numbers from ON026091 to ON026096 for the microsporidian sequences and from ON033621 to ON033624 for the rotifers.

Competing Interests

The authors declare no competing interests.

References

- Barron, G.L., Szijarto, E. (1990) A new genus of the hyphomycetes endoparasitic in rotifers. *Mycologia*, **82**,134–137.
- Chauvet, M., Debroas, D., Moné, A., Dubuffet, A., Lepère, C. (2022) Temporal variations of Microsporidia diversity and discovery of new host-parasite interactions in a lake ecosystem. *Environmental Microbiology*, **24**, 1672–1686.
- Dubuffet, A., Chauvet, M., Moné, A., Debroas, D., Lepère, C. (2021) A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environmental Microbiology*, **23**, 4344–4359.
- Fokin, S.I., Di Giuseppe, G., Erra, F., Dini, F. (2008) *Euplotespora binucleata* n. gen., n. sp. (Protozoa: Microsporidia), a parasite infecting the hypotrichous ciliate *Euplotes woodruffi*, with observations on microsporidian infections in ciliophora. *The Journal of Eukaryotic Microbiology*, **55**, 214–228.
- Gorbunov, A.K., Kosova, A.A. (2001) Parasites in rotifers from the Volga delta. *Hydrobiologia*, **446**, 51–55.
- Lair, N., Oulad Ali, H. (1990) Grazing and assimilation rates of natural populations of planktonic rotifers *Keratella cochlearis*, *Keratella quadrata* and *Kellicottia longispina* in a eutrophic lake (Aydat, France). *Hydrobiologia*, **194**, 119–131.
- Langdon, P.G., Ruiz, Z., Brodersen, K.P., Foster, I.D.L. (2006) Assessing lake eutrophication using chironomids: understanding the nature of community response in different lake types. *Freshwater Biology*, **51**, 562–577.
- Mikhailov, K.V., Simdyanov, T.G., Aleoshin, V.V. (2016) Genomic survey of a hyperparasitic Microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biology and Evolution*, **9**, 454–467.
- Murareanu, B.M., Sukhdeo, R., Qu, R., Jiang, J., Reinke, A.W. (2021) Generation of a Microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of microsporidia. *mBio*, **12**, e01490-21.
- Rico-Martínez, R., Arzate-Cardenas, M.A., Robles, D., Pérez-Legaspi, I., Alvarado-Flores, J., Santos Medrano, G.E. (2016) Rotifers as models in toxicity screening of chemicals and environmental samples. In: *Invertebrates – experimental models in toxicity screening*. InTech, London, pp 57–99.
- Stentiford, G.D., Ramilo, A., Abollo, E., Kerr, R., Bateman, K.S., Feist, S.W., et al. (2017) *Hyperspora aquatica* n.gen., n.sp. (Microsporidia), hyperparasitic in *Marteilia cochillia* (Paramyxida), is closely related to crustacean-infecting microsporidian taxa. *Parasitology*, **144**,186–199.
- Thomas, S.H., Housley, J.M., Reynolds, A.N., Penczykowski, R.M., Kenline, K.H., Hardegree, N., et al. (2011) The ecology and phylogeny of oomycete infections in *Asplanchna* rotifers: oomycete infections in *Asplanchna*. *Freshwater Biology*, **56**, 384–394.
- Tokarev, Y.S., Voronin, V.N., Seliverstova, E.V., Pavlova, O.A., Issi, I.V. (2010) Life cycle, ultrastructure, and molecular phylogeny of *Crispospora chironomi* g.n. sp.n. (Microsporidia: Terresporidia), a parasite of *Chironomus plumosus* L. (Diptera: Chironomidae). *Parasitology Research*, **107**, 1381–1389.

- Tokarev, Y.S., Huang, W.F., Solter, L.F., Malysch, J.M., Becnel, J.J., Vossbrinck, C.R. (2020) A formal redefinition of the genera *Nosema* and *Vairimorpha* (Microsporidia: Nosematidae) and reassignment of species based on molecular phylogenetics. *Journal of Invertebrate Pathology*, **169**, 107279.
- Wolska, M., Mazurkiewicz-Zapalowicz, K. (2013) Parasites of zooplankton and periphyton assemblages in the littoral zone of lakes in Drawa National Park, Poland. *Acta Mycologica*, **48**, 51–59.
- Yakovleva, Y., Nassonova, E., Lebedeva, N., Lanzoni, O., Petroni, G., Potekhin, A., et al. (2020) The first case of microsporidiosis in *Paramecium*. *Parasitology*, **147**, 957–971.
- Zhang, Y.N., Xu, S.L., Sun, C., Dumont, H., Han, B.P. (2021a) A new set of highly efficient primers for COI amplification in rotifers. *Mitochondrial DNA part B*, **6**, 636–640.
- Zhang, Y.N., Xu, S.L., Huang, Q., Liu, P., Han, B.P. (2021b) Application of COI primers 30F/885R in rotifers to regional species diversity in (sub)tropical China. *Diversity*, **13**, 390.

III. Supplementary Materials and Methods

Sampling

This study was conducted in lake Aydat (Massif Central, France, 45°39' 50" N, 2° 59' 04" E). It is a dimictic and eutrophic lake with a maximum depth of 15 m, situated at an altitude of 825 m. 150 ml of water concentrated in zooplankton was sampled using vertical tows of 20 µm-mesh plankton net and kept at RT until the next day.

Single cell isolation

17 rotifers were isolated by micromanipulation. A drop of the plankton net sample diluted with distilled water was placed in the center of a medium sized petri dish. Using an inverted microscope (Axio Vert.A1, Carl Zeiss, Oberkochen, Germany) and a stretched capillary, rotifers were individually collected. Each rotifer was then placed in a 96 wells plate in which 1 µl of PBS was previously deposited. After centrifugation, the 96 wells plate was stored at -80°C until further processing.

Host-Parasite characterization

Whole Genome Amplification (WGA) was carried out using the GenomiPhi™ V2 DNA Amplification Kit (GE Healthcare, Chicago, IL). Briefly, the cells underwent 5 cycles of thermic lysis (alternatively, bowled water and liquid nitrogen). The sorted cells were chemically lysed with the addition of 1 µl of an alkaline solution (400 mM KOH, 100 mM DTT and 10 mM EDTA) and incubated at RT for 10 min. Lysed samples were neutralized with 1 µl of neutralization buffer (HCL 400 mM, Tris-HCl 600 mM pH 7.5). 7 µl of sample buffer was added and incubated 3 min at 95°C. A total of 10 µl of master mix was added (9 µL of reaction buffer and 1 µl of Phi29 DNA polymerase) and the plate was incubated at 30°C for 16 h. A final incubation at 65°C for 10 min inactivated the Phi29 DNA polymerase. The amplification reactions were stored at -20°C.

After the WGA, the Single Amplified Genomes (SAGs) were screened for the presence of Microsporidia by the amplification of the V1-V3 region of the microsporidian gene coding for the SSU rRNA according to Chauvet *et al.* (2022). Microsporidia positive PCR products were then cloned into pCR 2.1-TOPOs vectors (TA cloning kit, Invitrogen, Carlsbad, CA) and transformed into *Escherichia coli* competent cells following the manufacturer's instructions.

The presence of the target small-subunit rRNA gene insert in positive colonies was checked by PCR amplification using flanking vector primers (M13f and M13r). PCR products purified with QIAquick PCR purification kit (Qiagen, Hilden, Germany) were sequenced using Sanger technology (Eurofins Genomics, Luxembourg, Luxembourg). The sequences were then compared with those available in public databases with the NCBI BLAST web application (<http://www.ncbi.nih.gov/BLAST/>) and included in a phylogenetic tree after an alignment using the MAFFT algorithm (Katoh *et al.*, 2005). Maximum Likelihood tree was obtained using IQ-TREE (Nguyen *et al.*, 2015) with the ModelFinder (Kalyaanomorthy *et al.*, 2017) tool to find the best predicted model of evolution (TVM+F+I+G4) according to the BIC score, with 100 bootstrap replications. Average Nucleotide Identities (ANIs) between sequences were determined using PYANI tool v0.2.11 (Pritchard *et al.*, 2015).

On the Microsporidia positive SAGs, rotifers were characterized by amplification of the COI using the 30F and 885R primers set according to Zhang *et al.* (2021) with the following modifications. Each PCR was performed in a total volume of 20 μ L including 4 μ L of 5x Green GoTaq Flexi buffer, 1.6 μ L of 25 mM MgCl₂, 0.4 μ L of 10 mM each dNTP, 0.4 μ L of each 10 μ M primer, 0.2 μ L of 50 mg/mL BSA, 0.12 μ L of GoTaqG2 Flexi DNA Polymerase (Promega, Madison, WI, 5 U/ μ L) and 1 μ L of DNA sample. The amplification conditions consisted of an initial denaturation at 95°C for 10 min, followed by 35 cycles for 30 s at 95°C, 40 s at 51°C and 30 s at 72°C, and a final elongation at 72 °C for 10 min. PCR products underwent the same protocol previously detailed for Microsporidia characterization. The Maximum Likelihood tree was obtained using the best predicted model of evolution (GTR+G+I) according to the BIC score, with 200 bootstrap replications and partial deletion (85%), using MEGA11 (Tamura *et al.*, 2021).

Tyramide Signal Amplification – Fluorescent *In Situ* Hybridization (TSA-FISH)

TSA-FISH was performed using a Microsporidia specific probe (USP01) according to Chauvet *et al.* (2022). Hybridized cells were visualized under a Zeiss Axio Imager 2 microscope and confocal images were acquired with a Zeiss LSM 980 microscope.

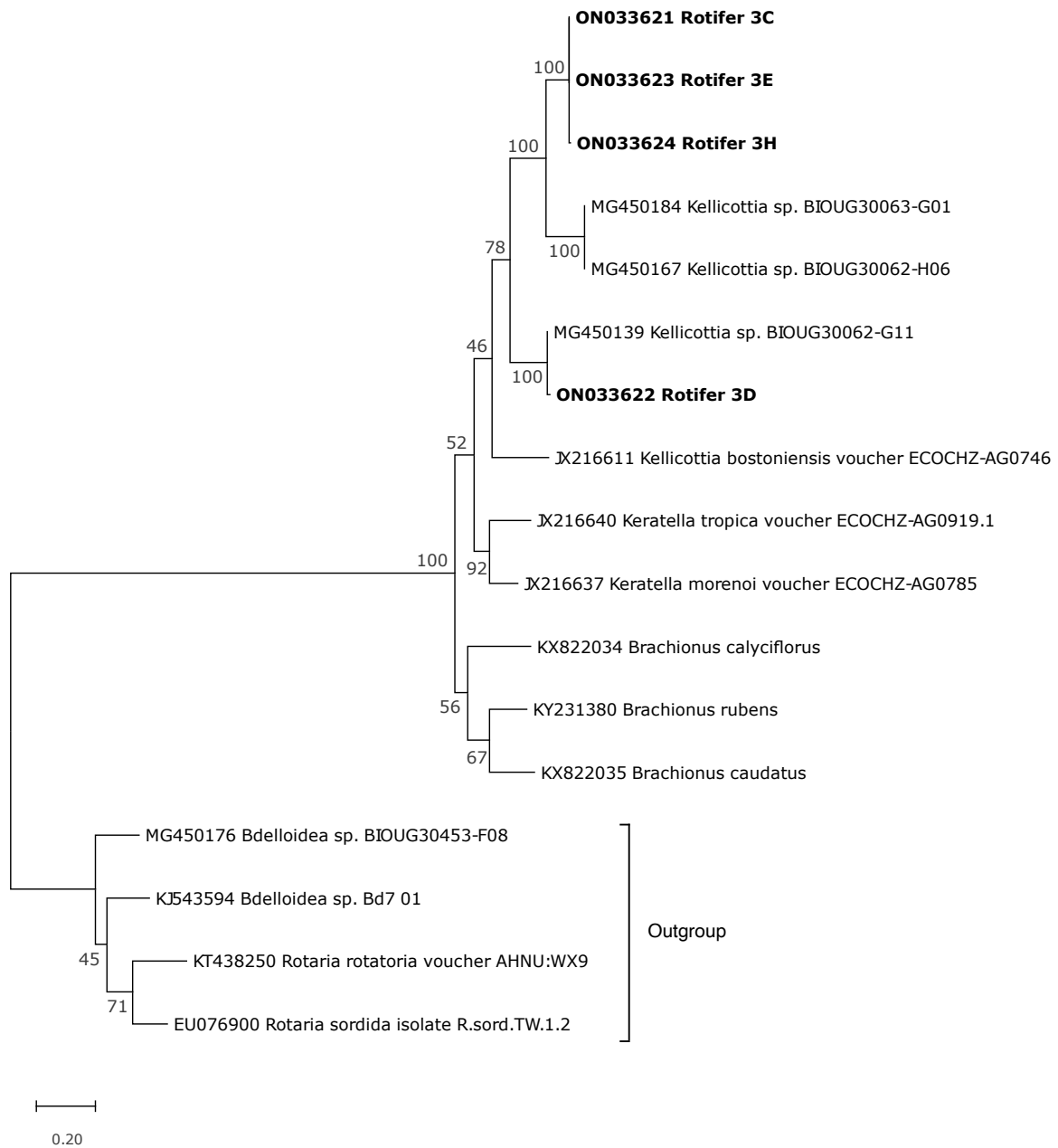
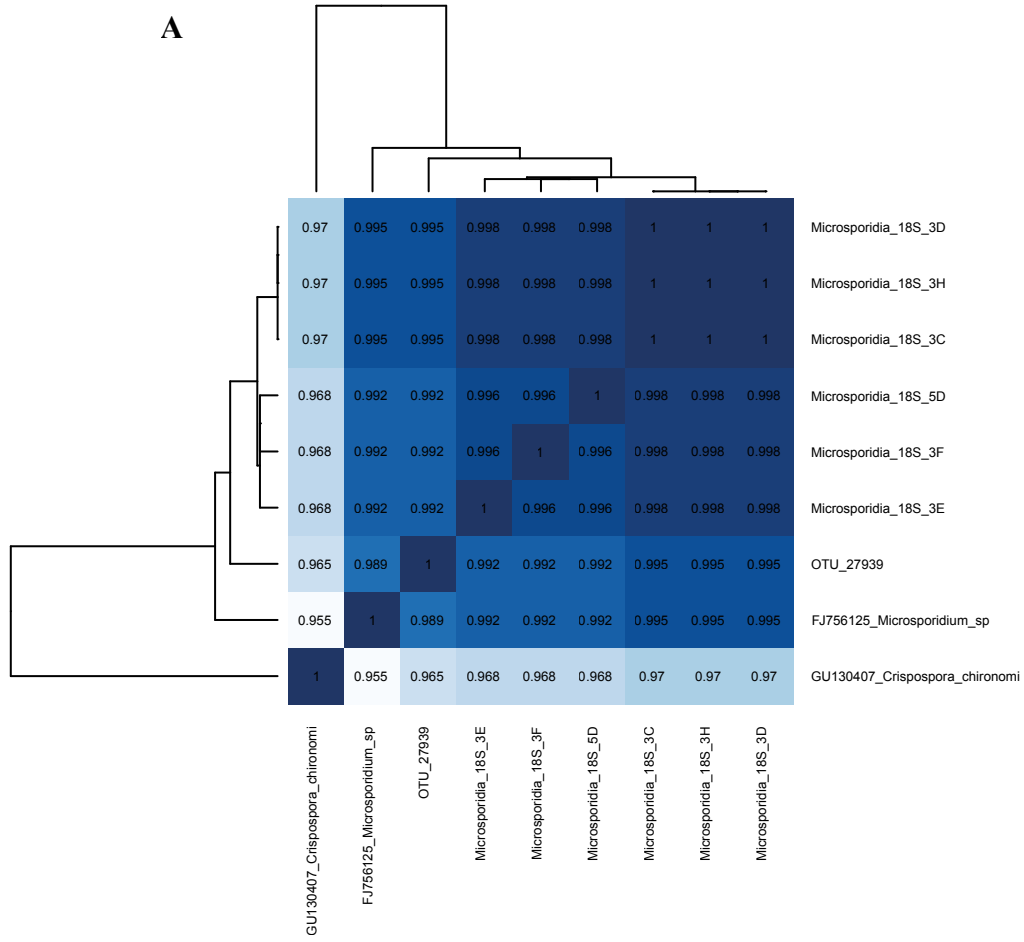


Figure S1: Maximum Likelihood phylogeny of rotifer sequences.

We included in this phylogeny several sequences of environmental *Kellicottia* (Chauvet *et al.*, 2022) and sequences of identified rotifer species. Sequences from this study are given in bold. Outgroup species are rotifers belonging to the Bdelloidea subclass. Numbers next to branches indicate the ML bootstrap support (200 replicates).



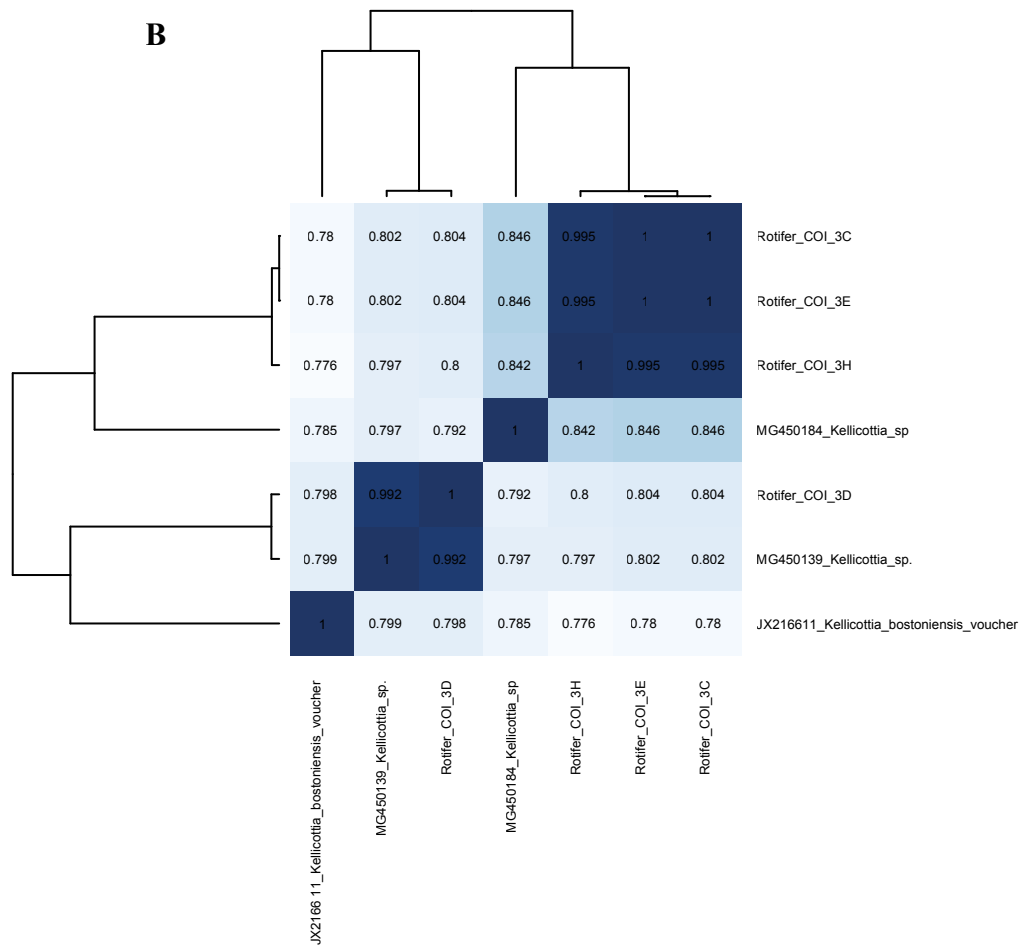


Figure S2: Heatmap of the Average Nucleotide Identity (ANI) of Microsporidia (A) and *Kellicottia* (B) sequences included in the phylogenetic analyses.

References

- 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.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S. (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589.
- Katoh, K., Kuma, K., Toh, H., Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.
- Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

- Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., Toth, I.K. (2015) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*, **8**, 12–24.
- Tamura, K., Stecher, G., Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, **38**, 3022–3027.
- Zhang, Y., Xu, S., Sun, C., Dumont, H., Han, B.P. (2021) A new set of highly efficient primers for COI amplification in rotifers. *Mitochondrial DNA part B*, **6**, 636–640.

CHAPITRE V

DISCUSSION GÉNÉRALE ET PERSPECTIVES

I. Les approches « -omics » dans l'étude du lien diversité-fonction des eucaryotes unicellulaires

Les connaissances quant à la diversité des eucaryotes unicellulaires ont connu une avancée majeure au cours des dernières décennies notamment avec l'avènement des méthodes de séquençage haut débit. En effet, les approches « -omics » telles que le métabarcoding ont permis de mettre en évidence une diversité insoupçonnée d'eucaryotes unicellulaires surpassant largement les approches de mise en culture (del Campo *et al.*, 2014 ; de Vargas *et al.*, 2015 ; Debroas *et al.*, 2017 ; Singer *et al.*, 2021). D'autres approches, comme la métatranscriptomique ou la métagénomique offrent la possibilité d'évaluer les capacités fonctionnelles et métaboliques des communautés microbiennes, tout en permettant d'étendre les connaissances de l'arbre des eucaryotes en fournissant des informations évolutives pertinentes (Grossmann *et al.*, 2016 ; Carradec *et al.*, 2018 ; Trench-Fiol et Fink, 2020 ; Delmont *et al.*, 2022). Toutefois, même si elles offrent des avantages considérables, les approches « -omics » présentent certains biais qu'il est nécessaire de connaître et d'appréhender pour une analyse optimale des résultats.

A. Limite méthodologique

D'un point de vue méthodologique, la quasi-totalité des étapes allant de l'extraction des acides nucléiques au séquençage peut introduire des erreurs. Pour l'étude des organismes parasites, par exemple, l'extraction peut s'avérer délicate du fait de la résistance de certaines formes de dissémination (Bass *et al.*, 2015). C'est le cas notamment des spores produites par les microsporidies, les Cryptomycota et les Oomycota qui ne sont pas dégradées par les agents lytiques utilisés dans les protocoles d'extraction (Felczykowska *et al.*, 2015 ; Brauer et Bengtsson, 2022). Au cours de nos études, nous avons combiné, en conséquence, plusieurs techniques d'extraction (*i.e.*, chocs thermiques et/ou physiques ainsi qu'un traitement enzymatique) permettant, par exemple dans le chapitre IV, d'accéder au contenu génétique de parasites affiliés aux microsporidies au sein des rotifères. Toutefois, il est envisageable que la diversité des parasites, bien que nous l'ayons identifiée dans le chapitre II comme dominante au sein du lac Pavin, soit encore sous-évaluée.

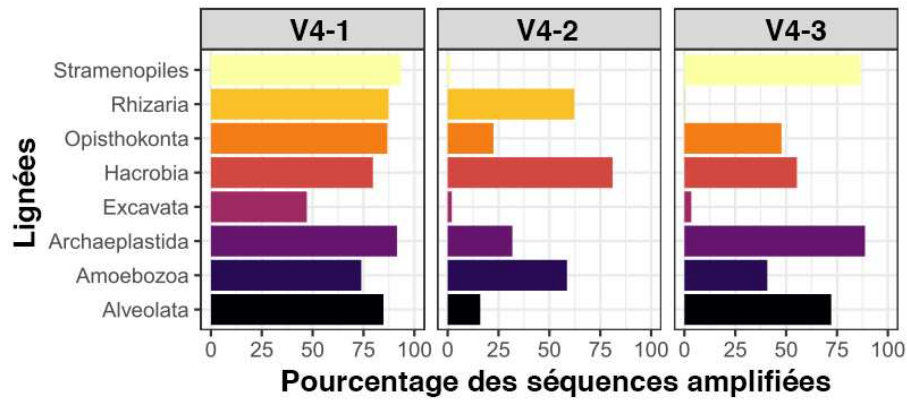


Figure 1 : Représentation graphique de la proportion de séquences affiliées à chaque lignée en fonction du couple d'amorces utilisé (V4-1, V4-2 ou V4-3) ciblant la région V4 du gène codant pour l'ARNr 18S (modifiée d'après Geisen *et al.*, 2019).

Le nombre restreint de transcrits liés aux parasites, rapporté par notre étude de métatranscriptomique, associé à l'éventuelle (Novodvorska *et al.*, 2016 ; Huang *et al.*, 2023), mais pas systématique (Laundon *et al.*, 2022) faible activité métabolique des stades de dissémination renforce, par ailleurs, cette hypothèse. Le temps de demi-vie intrinsèquement court des ARNm influant sur leur stabilité, et donc sur leur séquençage, est également un élément à prendre en compte. Nos protocoles ont été réalisés en tenant compte de ce paramètre en congelant les ARNm moins de deux heures après l'échantillonnage.

Les approches de metabarcoding comme celle que nous avons réalisée dans le chapitre II, nécessitent l'utilisation d'un marqueur génétique ubiquiste. Nous avons choisi le gène codant pour la petite sous-unité de l'ARNr 18S car il offre une structure avantageuse ainsi qu'un nombre de références important dans les bases de données. Toutefois, bien qu'il soit le plus couramment utilisé, il ne permet pas d'évaluer la diversité de l'ensemble des lignées eucaryotes avec le même degré de précision. L'utilisation des régions ITS, par exemple, est recommandée par certains chercheurs pour l'étude de la diversité fongique (Tedersoo *et al.*, 2022) car elle permettrait une meilleure résolution phylogénétique et ainsi une meilleure distinction des espèces (Schoch *et al.*, 2012 ; Heeger *et al.*, 2018). De plus, les variations de structure de l'ADNr 18S entre les différentes lignées d'eucaryotes (Geisen *et al.*, 2019), complexifient grandement la définition d'amorces réellement universelles (Vaulot *et al.*, 2022). Certains chercheurs utilisent même plusieurs couples d'amorces simultanément pour espérer couvrir la totalité de la diversité (**Fig. 1**) (Stoeck *et al.*, 2006 ; Geisen *et al.*, 2019). Dans notre cas, bien que nous ayons dévoilé une diversité importante de microorganismes eucaryotes dans le lac Pavin (3 358 ASVs), nous n'avons pas ciblé l'entièreté de la diversité présente. C'est le cas par exemple d'un groupe de parasites obligatoires apparentés aux Fungi : les microsporidies, absentes de notre premier jeu de données (chapitre II), et dont la diversité a nouvellement été mise en lumière dans les écosystèmes marins et lacustres par l'utilisation d'amorces adaptées (cf. annexe 1 ; Chauvet *et al.*, 2022).

L'utilisation de séquences courtes afférentes aux techniques de séquençage dites « short reads », freine également l'identification précise de nos séquences. En metabarcoding, nous n'avons ciblé que des régions partielles du gène codant pour la petite sous-unité de l'ARNr : la région hypervariable V4. Longue de 400 pb, elle n'est parfois pas suffisante pour différencier deux espèces proches. En métatranscriptomique, tous les ARNm ne sont également pas séquencés dans leur intégralité ce qui rend bien plus difficile l'étape d'annotation fonctionnelle et d'affiliation taxonomique (Pollard *et al.*, 2018 ; Shakya *et al.*, 2019).

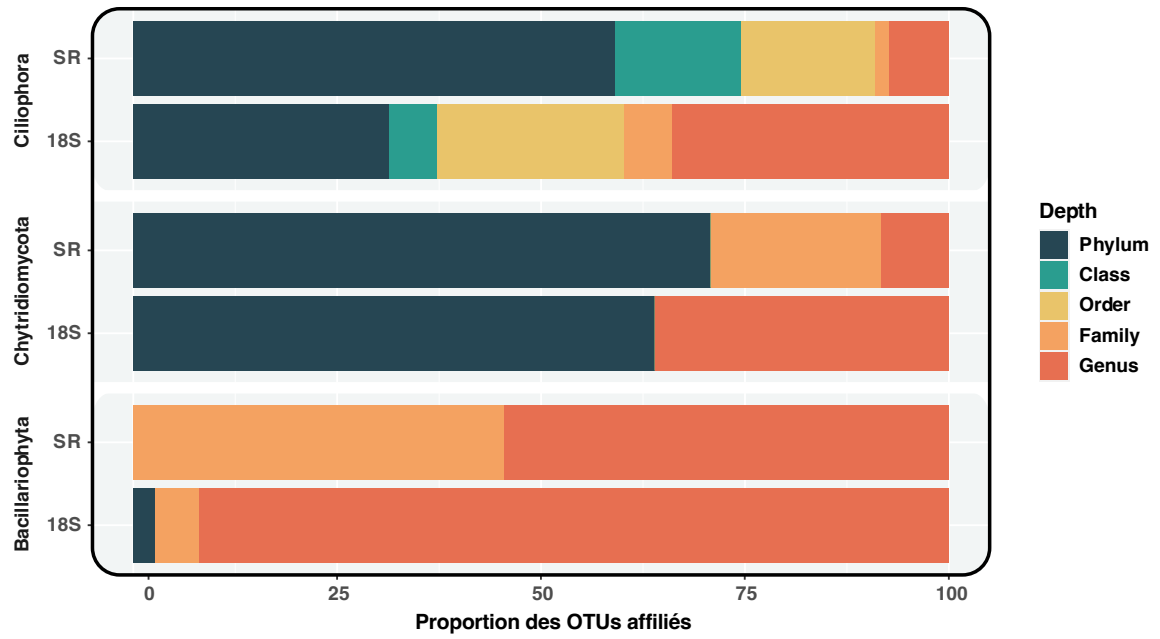


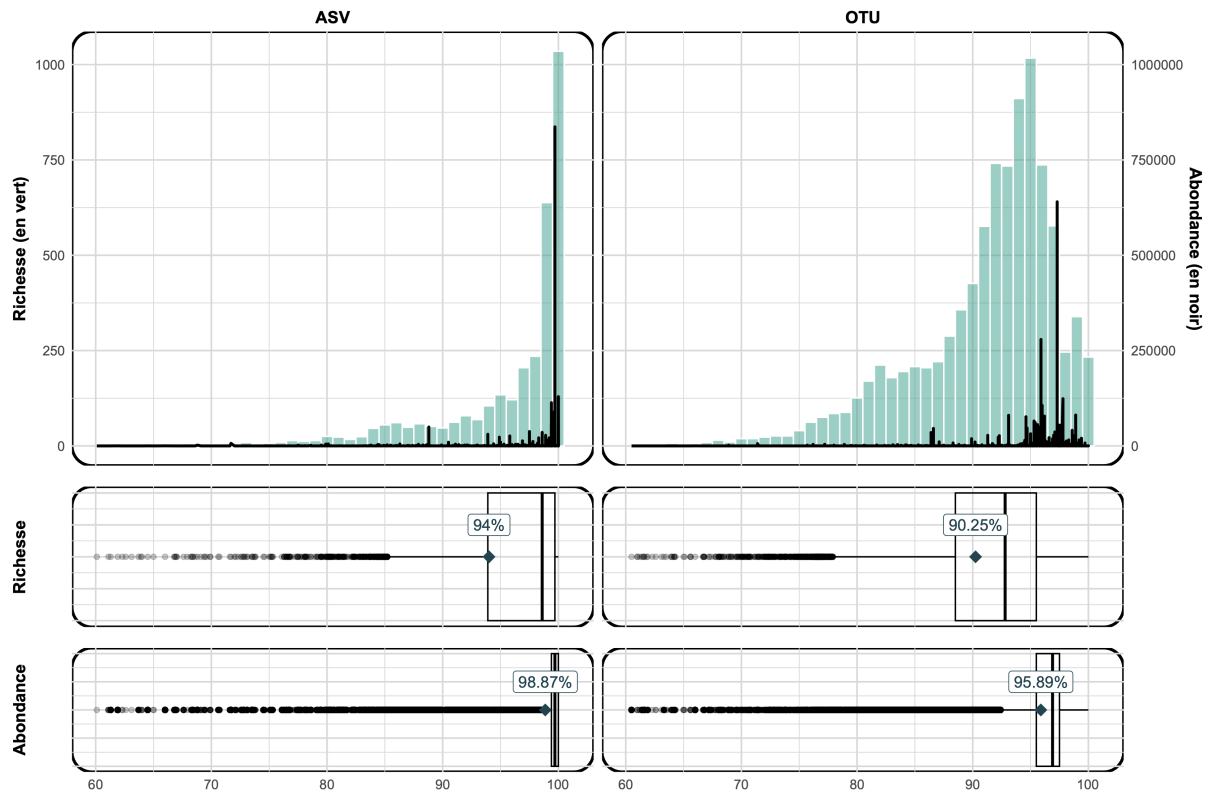
Figure 2 : Représentation graphique de la proportion des OTUs affiliés à chaque rang taxonomique au sein de trois lignées eucaryotes, en fonction de la région séquencée (modifiée d'après Chauvet *et al.*, en prep.).

L'utilisation de méthodes de séquençage dites « long-reads » serait pertinente afin d'agrandir significativement le cadre phylogénétique en métabarcoding, en ciblant l'intégralité du gène de l'ARNr par exemple (Jamy *et al.*, 2019, 2022 ; Amarasinghe *et al.*, 2020). Des travaux utilisant ces approches ont d'ailleurs été initiés dans notre équipe à partir des échantillons décrits dans le chapitre II. Le séquençage de la totalité de la petite sous-unité ribosomique 18S de l'ADN a permis la production d'OTUs pouvant être affiliés plus profondément qu'en séquençant uniquement la région hypervariable V4 (Chauvet *et al.*, en prep.), et ce, pour plusieurs lignées eucaryotes affiliées au Stramenopiles, Alveolata et Opisthokonta (**Fig. 2**). Toutefois, l'application de ces approches de séquençage (*e.g.*, PacBio ou Oxford Nanopore Technologies) n'est pas systématique ou alors utilisée en complément des approches « short reads » (Pollard *et al.*, 2018 ; Jamy *et al.*, 2019 ; Shakya *et al.*, 2019) en raison de leur fort taux d'erreur. Les innovations techniques récentes basées sur la combinaison de processus correctifs et de contrôles qualité rigoureux, abaissant ce taux (<1%), permettront certainement de démocratiser ces approches (Heeger *et al.*, 2018 ; Tedersoo *et al.*, 2018 ; Amarasinghe *et al.*, 2020).

B. Limite analytique

L'énorme quantité de données générées par les techniques HTS (High Throughput Sequencing) nécessite le développement d'outils de traitement bio-informatique adaptés afin de pouvoir contrôler leur qualité et faciliter leur interprétation tout en détectant et supprimant les séquences résultantes d'erreurs méthodologiques. En métabarcoding, le choix d'une unité de base de la biodiversité est nécessaire afin d'estimer le plus fidèlement possible la diversité de nos échantillons. Dans le chapitre II, nous avons choisi d'utiliser les variants de séquence d'amplicon (ASV) par rapport à la méthode de clusterisation en unités taxonomiques opérationnelles (OTU). Le principal reproche adressé aux OTUs est leur tendance à regrouper au-delà du seuil établi (défini historiquement à 95% pour les eucaryotes) des séquences qui pourraient correspondre à des unités de diversité proches, telles que des espèces, voire des genres, pourtant distincts. De plus, ils ne sont pas très efficaces pour agréger les séquences erronées résultant d'erreurs d'amplification et de séquençage ce qui a pour effet d'augmenter considérablement le nombre d'OTUs générés, tout en masquant la micro-diversité qui pourrait exister (Geisen *et al.*, 2019 ; Johnson *et al.*, 2019). D'un autre côté, les ASVs permettent de réduire de manière plus efficace cette diversité aléatoire (Callahan *et al.*, 2017 ; Johnson *et al.*, 2019).

A



B

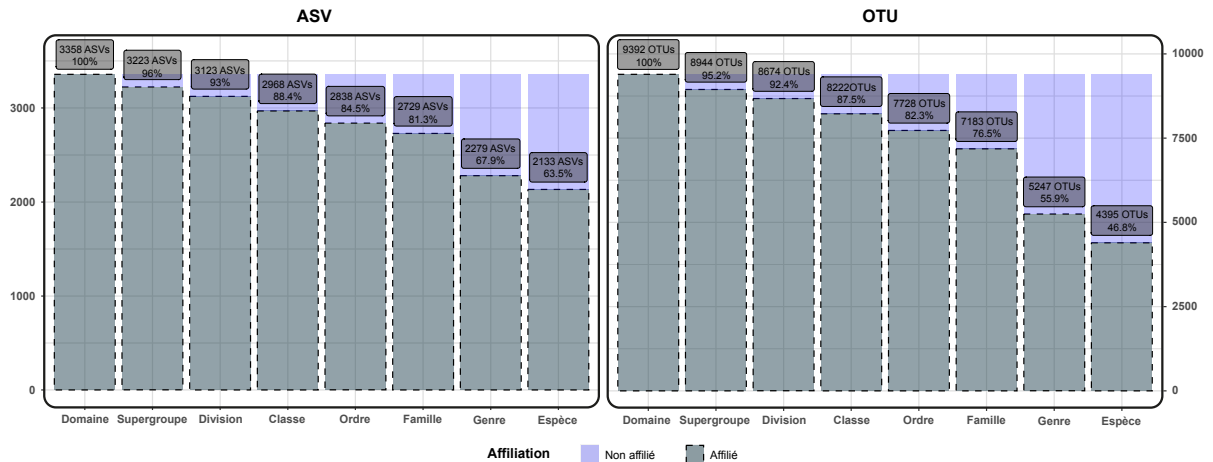


Figure 3 : Pourcentage d'identité (A) et profondeur d'affiliation (B) des ASVs et des OTUs contre la base de données PR2.

L'affiliation taxonomique des ASVs et des OTUs a été réalisée par recherche de similarité avec l'outil VSEARCH en A (Rognes *et al.*, 2016) et en utilisant le classificateur naïf bayésien RDP intégré au pipeline DADA² en B (Callahan *et al.*, 2016). Les paramètres et les filtres utilisés pour l'obtention des OTUs sont rigoureusement identiques à ceux décrits dans le chapitre II. Une suppression des singletons a été effectuée pour les deux types d'unité et la clusterisation en OTUs a été réalisée au seuil historique de 95% (Caron *et al.*, 2009).

Il est à noter toutefois que des ASVs distincts peuvent, *in fine*, correspondre à une même espèce ou sous-espèce mettant alors en évidence les multiples loci de l'opéron d'ADNr co-existant au sein des génomes des organismes détectés (Callahan *et al.*, 2019 ; Schloss, 2021). C'est le cas, par exemple, des Dinophyceae et des Ciliophora dont le nombres de copies de l'ADNr est décrit comme important dans la littérature (Massana, 2011).

Préalablement à l'étude décrite dans le chapitre II et à partir du même jeu de données, nous avons réalisé une étude comparative de l'efficacité de l'affiliation taxonomique des OTUs et des ASVs, obtenus respectivement avec les outils VSEARCH et DADA² (**Fig. 3**). Bien qu'il ait été difficile de mettre en évidence les multiples copies de l'opéron de l'ADNr restituées par les ASVs, nous avons clairement identifié un nombre bien supérieur d'OTUs pour lesquels l'affiliation était moins précise. En effet, les séquences centroïdes des OTUs présentaient une identité bien plus faible avec les séquences de référence (90,25% en moyenne) comparativement au ASVs (94%). De plus, l'affiliation taxonomique implémentée dans DADA² (RDP) permettait une identification plus profonde pour les ASVs surpassant largement celle réalisée sur les OTUs (*i.e.*, 63,5% vs 46,8% au rang de l'espèce). Même si ces résultats devraient être complétés par l'analyse d'un jeu de données synthétique ou « mock », il semble que les ASVs présentent de bien meilleures statistiques d'affiliation pour nos séquences.

Afin d'écartier toutes les erreurs méthodologiques liées au métabarcoding, nous avons appliqué, en plus des paramètres recommandés, des filtres stricts (*i.e.*, ne tolérant aucun mismatch lors de l'étape d'assemblage et aucune erreur dans les primers en début et en fin de séquences), un processus de suppression des chimères ainsi qu'une suppression des ASVs peu abondants (*i.e.*, singleton). Bien que ces étapes aient considérablement réduit notre jeu de données (plus de 16% des séquences), elles ont quand même permis l'analyse, *in fine*, de la majorité de la diversité, et ce, pour l'intégralité de nos échantillons (**chapitre II : Fig. S7**). Nos données ont également été normalisées par raréfaction, en suivant les recommandations de Weiss *et al.* (2017), afin de corriger les différences importantes de profondeur (variant du simple au double) induites lors du regroupement des duplicats lorsqu'ils existaient (Aguirre De Cárcer *et al.*, 2011 ; Wong *et al.*, 2016). Ceci a permis la comparaison de nos échantillons sans modification de richesse (**Fig. 4**). Toutefois, cette stratégie, bien que réalisée depuis des années dans la majorité des études de diversité, est largement critiquée, induisant une perte significative de la sensibilité et de la spécificité du jeu de données tout en ajoutant une incertitude inutile liée au sous-échantillonnage aléatoire (McMurdie et Holmes, 2014). L'utilisation de données compositionnelles ne devrait, en effet, pas être systématique, et la méthode pour les obtenir (*e.g.*, raréfaction) devrait dépendre de la structure du jeu de données (Weiss *et al.*, 2017).

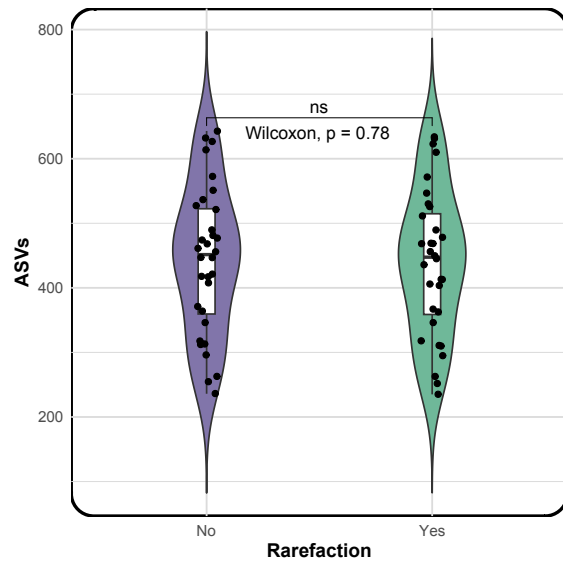


Figure 4 : Effet de la raréfaction sur la richesse des ASVs.

Les méthodes de normalisation implémentées dans les packages d'analyse d'expression différentielle edgeR et DESeq2, que nous avons appliquées dans le chapitre II, semblent davantage satisfaisantes en limitant considérablement la quantité de faux positifs et en permettant une meilleure détection des abondances différentielles, et ce, quelles que soient les différences de taille de bibliothèques (McMurdie et Holmes, 2014 ; Weiss *et al.*, 2017). Enfin, il est à noter que l'intégralité des données acquises par les techniques de séquençage haut-débit sont *de facto* des données compositionnelles (Gloor *et al.*, 2017). Les séquenceurs ne pouvant délivrer des lectures que dans la limite de leur capacité, le nombre total de lectures observées est un échantillon aléatoire, de taille fixe, de l'abondance relative des molécules de l'échantillon d'entrée. Ainsi, même si nous avons choisi de ne pas normaliser nos données, nous aurions dû, tout de même, les considérer comme compositionnelles. Tous ces paramètres influent grandement sur notre interprétation des données d'abondance, que ce soit en métabarcoding ou en métatranscriptomique.

Les étapes d'affiliation taxonomique et d'annotation fonctionnelle sont également cruciales pour l'évaluation exhaustive et qualitative des données de métabarcoding et de métatranscriptomique. Elles dépendent en effet grandement des bases de données de référence utilisées et de leur richesse (Geisen *et al.*, 2019). Dans les chapitres II et III, faute d'équivalents lacustres, nous avons dû utiliser des bases de données taxonomiques développées majoritairement à partir de données marines (*i.e.*, PR2 (Guillou *et al.*, 2013) et MetaEuk (Keeling *et al.*, 2014 ; Levy Karin *et al.*, 2020)). Bien que spécialisées pour l'affiliation des organismes eucaryotes, elles n'ont permis d'affilier que 63,5% des ASVs et seulement 11% des transcrits au niveau de l'espèce et suggèrent un manque crucial de données relatives aux écosystèmes d'eau douce. Ces résultats étayaient d'ailleurs d'autres observations obtenues par Singer *et al.* (2021) décrivant que les séquences obtenues lors du séquençage d'écosystème d'eau douce ont un pourcentage de similarité avec des séquences de références plus faible que celles obtenues en milieu marin. Dans le cadre de l'étude évaluant la diversité taxonomique des microsporidies (annexe 1), nous avons fait le choix d'utiliser une base de données sur mesure développée au laboratoire et reposant sur l'analyse phylogénétique de 265 espèces de microsporidies issues de la littérature et provenant d'écosystèmes variés (Dubuffet *et al.*, 2021). L'utilisation de bases de données « faites maison » n'est pas rare dans le cadre d'étude de la diversité de microorganismes peu représentés dans les bases de références, mais ne représente pas une stratégie viable à long terme tant leur entretien et leurs mises à jour sont chronophages.

La centralisation des données au sein de bases publiques devrait être systématique en permettant l'établissement d'une classification universelle facilitant la comparaison des études ainsi que leur contextualisation.

Des observations similaires sont rapportées pour les bases de données fonctionnelles dont la majorité des références sont bactériennes (*e.g.*, KEGG et eggNOG comptabilisent respectivement 959 et 477 génomes eucaryotes contre 4 430 et 4 613 procaryotes (Kanehisa *et al.*, 2016 ; Huerta-Cepas *et al.*, 2019). De plus, des biais afférents aux difficultés de mise en culture et d'étude de certains eucaryotes s'ajoutent à cette problématique. La représentation des différentes lignées eucaryotes à travers les bases de données est en effet très inégale, et ce, en faveur d'organismes macroscopiques, pluricellulaires, photosynthétiques, pathogènes ou encore d'intérêt économique (del Campo *et al.*, 2014 ; Richter *et al.*, 2022). Dans les chapitres II et III, seuls 30,7% de l'intégralité des protéines prédites ont pu être associés à un identifiant KEGG ne permettant qu'une étude partielle de la diversité fonctionnelle.

C. Étude du lien diversité-fonction

i. Classification en groupes fonctionnels

Dans la majorité des études explorant le lien diversité-fonction, le rôle des espèces au sein d'un écosystème est déduit à partir de leurs taxonomies. Les chercheurs utilisent le plus souvent une classification fonctionnelle permettant de catégoriser les espèces, les genres, voire les classes lorsque cela est possible, au sein de groupes fonctionnels sur la base de caractéristiques décrites dans la littérature (Petchey et Gaston, 2006). Le choix de cette classification, ainsi que les caractéristiques sur lesquelles elle repose, joue grandement sur l'évaluation et la mesure finale de la diversité fonctionnelle. Les caractéristiques classiquement définies dans les études ciblant les eucaryotes unicellulaires aquatiques sont basées sur leurs positions trophiques au sein d'un réseau simplifié. Les groupes ainsi définis dans ces études sont les phototrophes, les hétérotrophes et plus rarement les parasites (de Vargas *et al.*, 2015 ; Singer *et al.*, 2021).

Dans le chapitre III, nous avons choisi une classification plus détaillée, basée sur les capacités des organismes à réaliser la photosynthèse et/ou la phagocytose ainsi que sur leur possible mode de vie parasitaire.

Celle-ci permet ainsi la caractérisation de la presque totalité des stratégies trophiques d'organismes eucaryotes unicellulaires distribués, cette fois-ci, au sein d'un réseau complet, en les répartissant parmi 5 groupes fonctionnels : les hétérotrophes, les saprotrophes, les photo-osmo-mixotrophes, les photo-osmo-phago-mixotrophes et les parasites (obligatoires ou facultatifs). Toutefois, cette approche nécessite de classer les organismes parmi des groupes préalablement établis.

Dans le chapitre II, nous avons choisi d'utiliser davantage de caractéristiques pour effectuer la classification, ainsi qu'une méthode de clusterisation non supervisée prenant alors en compte la structure réelle du jeu de données. Cette seconde approche est basée sur le référencement de traits morpho-physio-phénotypiques afin d'évaluer la performance individuelle des espèces au sein de l'écosystème. Ces traits, issus du processus de sélection naturelle, représentent un proxy efficace pour caractériser la « fitness » des espèces *via* leurs effets sur la croissance, la reproduction, l'acquisition des ressources et la survie (Violle *et al.*, 2007 ; Litchman et Klausmeier, 2008 ; Ramond *et al.*, 2019). Les eucaryotes unicellulaires identifiés ont alors été regroupés sur la base de leurs traits. Cette méthode s'est révélée efficace par la définition, non supervisée, de groupes fonctionnels, tout en apportant des informations quant aux traits orientant le plus significativement la clusterisation. Elle a ainsi permis de mettre en évidence des divergences au sein de groupes trophiques traditionnellement décrits dans la littérature en différenciant, par exemple, les endophytes des saprotrophes et en identifiant différents types de photo-osmo-mixotrophes : des flotteurs, des coloniaux, des mobiles et/ou caractérisés par une couverture cellulaire minérale (silice ou calcaire). Nous avons ainsi obtenu des clusters représentant les véritables groupes écologiques caractéristiques du milieu étudié.

Certains groupes n'ont pu être identifiés dans cette étude que par la prise en compte de différentes conditions d'échantillonnage (*e.g.*, les saprotrophes majoritairement retrouvés dans le monimolimnion). En effet, contrairement à la méthode de regroupement effectuée dans le chapitre III, la performance de l'étude basée sur les traits dépend intégralement du jeu de données et de la multiplicité des conditions étudiées. Toutefois, l'importante diversité des modes de vie des eucaryotes unicellulaires s'est révélée extrêmement difficile à catégoriser suggérant ainsi que le nombre de traits utilisés n'est pas suffisant. De plus, les traits influençant fortement la mise en évidence de certains groupes (*e.g.*, la présence ou non de plaste induisant l'identification de groupes photosynthétiques), il est probable que ceux utilisés dans notre étude orientent et limitent notre évaluation.

Cette deuxième approche nécessite également des descriptions détaillées de la morphologie, de la physiologie et du mode de vie des organismes étudiés. Par conséquent, les données d'observation en microscopie acquises depuis des décennies ainsi que les caractérisations métaboliques à partir de cultures semblent essentielles. Bien qu'il soit aisé de retrouver quelques-unes de ces informations dans la littérature, il n'existe que très peu de bases référençant l'ensemble de ces données (Ramond *et al.*, 2018 ; Martini *et al.*, 2021 ; Mitra *et al.*, 2023). Ainsi, même si nous avons fixé un minimum de 5 traits pour la classification en groupes fonctionnels, les données manquantes dans la littérature pour certaines espèces peu étudiées, ont comme conséquence leur non, ou pire, mauvaise classification.

Bien que l'utilisation de méthodes de clusterisation douce permettrait d'éviter ces associations erronées, en déterminant pour chaque espèce à classifier une valeur de pondération associée à chaque groupe (Ferraro *et al.*, 2019), elles ne semblent pas satisfaisantes. En effet, en fixant des seuils de confiance afin d'exclure les classifications douteuses, cette manipulation a comme effet la réduction significative du nombre d'espèces caractérisées.

ii. Des transcrits aux stratégies trophiques

Dans le chapitre II, nous avons choisi de coupler l'approche basée sur les traits morpho-physiophénotypiques à une analyse de métatranscriptomique afin d'étudier la diversité des transcrits associés aux eucaryotes unicellulaires en fonction de leur groupe fonctionnel, et ce, tout en prenant en compte les différentes conditions d'échantillonnage. La complémentarité de ces approches a permis d'améliorer la description de la diversité fonctionnelle en établissant un lien entre potentiel métabolique, groupes fonctionnels, taxonomies et paramètres environnementaux. Ce lien nous a permis de contextualiser l'expression des gènes évaluée en métatranscriptomique, laquelle, à elle seule, reste très difficilement interprétable pour l'évaluation du rôle écologique (Keeling et del Campo, 2017). En effet, les analyses de métatranscriptomique permettent d'évaluer la diversité des transcrits sans garantir qu'ils seront effectivement traduits en protéines fonctionnelles, *a fortiori* lorsqu'il s'agit de transcrits eucaryotes, pour lesquels les processus post-traductionnels sont nombreux et complexes. De plus, même si les bases de données actuelles tendent à renforcer les descriptions des protéines ainsi que leurs implications dans les voies métaboliques (Boutet *et al.*, 2007 ; Kanehisa *et al.*, 2016 ; Huerta-Cepas *et al.*, 2019), il n'est décrit, à ce jour, que très peu de marqueurs protéiques reflétant le rôle écologique des organismes qui les portent.

Enfin, la très faible proportion de transcrits identifiés taxonomiquement et associés à un groupe fonctionnel, n'a permis de décrire qu'une proportion réduite de la diversité fonctionnelle des eucaryotes unicellulaires. Dans ce contexte, afin de permettre une étude plus exhaustive des données de métatranscriptomique et d'approfondir notre analyse du lien entre protéines et stratégie trophique, nous avons conduit une étude basée sur la construction de réseaux de similarité de séquences protéiques (SSN) (chapitre III). Cette approche nous a permis de considérer à la fois les protéines connues et « inconnues » (*i.e.*, sans affiliation taxonomique, annotation fonctionnelle ou assignation à un groupe fonctionnel), en les regroupant au sein de composantes connexes (CCs). Cette approche s'est révélée particulièrement efficace pour étudier les caractéristiques métaboliques des groupes fonctionnels en mettant en évidence des similitudes entre protéines de lignées eucaryotes éloignées, pouvant partager des fonctions écologiques identiques. Ainsi, en étudiant plus spécifiquement l'ensemble des informations caractéristiques des protéines de chaque CCs, il nous a été possible d'identifier des voies métaboliques préférentielles de certains modes trophiques ainsi que des signatures protéiques propres à certains groupes de parasites. En parallèle, nos résultats ont démontré que l'étude de SSN représentait une approche prometteuse pour l'exploration de la matière noire microbienne (Bernard *et al.*, 2018). En effet, nous avons pu étudier la majorité des protéines de notre jeu de données en considérant plus de 45% de protéines supplémentaires par rapport à une étude traditionnelle (**chapitre III : Fig. 5**) (Grossmann *et al.*, 2016 ; Trench-Fiol et Fink, 2020). Toutefois, au cours de cette étude préliminaire (chapitre III), nous n'avons pas pris en compte les conditions environnementales qui auraient pourtant contextualisé les spécificités métaboliques enregistrées. Finalement, notre étude s'appuie sur la similarité des séquences calculée entre chaque paire de protéines, et ne tient pas compte des structures tridimensionnelles de celles-ci. La prédiction et l'analyse de ces structures, facilitées par les progrès considérables dans les domaines de l'intelligence artificielle, constitueraient une contribution majeure à l'évaluation des fonctions protéiques (Jumper *et al.*, 2021 ; Ma *et al.*, 2022). À l'avenir, l'étude du lien entre la diversité microbienne environnementale et ses fonctions dans l'écosystème sera certainement facilitée par le développement des approches de métagénomique et des bases de données adaptées.

Cependant, la métagénomique est jusqu'à présent peu utilisée pour l'étude des eucaryotes unicellulaires, notamment en raison de la complexité (*e.g.*, une densité élevée de répétitions) et de la taille de leurs génomes. Pourtant, Delmont *et al.* (2022) montrent la faisabilité de cette approche en reconstruisant plus de 700 génomes d'eucaryotes planctoniques à partir de centaines de milliards de lectures métagénomiques générées à travers l'expédition Tara Océans. D'autre part, le développement de base de données intégratives mêlant les informations taxonomiques, morpho-physio-phénotypiques et fonctionnelles est à privilégier pour l'exploration du lien diversité-fonction (Escalas *et al.*, 2019). Certaines initiatives tendent en effet à développer ce type de bases de données (del Campo *et al.*, 2018). C'est le cas par exemple, de la nouvelle version de PR² (v5.0.0) intégrant les données de MDB « The Mixoplankton Database » référençant des données de physiologie et de stratégie trophique relatives aux organismes mixotrophes (Mitra *et al.*, 2023).

II. Les interactions hôte-parasites au centre du réseau trophique

A. Apport des approches « -omics »

La diversité et le potentiel fonctionnel des eucaryotes unicellulaires ont pu être explorés au cours de cette thèse notamment par l'intermédiaire d'une classification en groupes fonctionnels (chapitre II et III). Le groupe des parasites s'est révélé être le plus divers en représentant près de 20% des ASVs obtenus (chapitre II), en adéquation avec des résultats précédemment acquis en milieu lacustre (*e.g.*, Lepère *et al.*, 2016). Une majorité d'entre eux était affiliée aux Fungi, parmi lesquels les Chytridiomycota, relativement bien décrits dans les écosystèmes d'eau douce (Ibelings *et al.*, 2004 ; Kagami *et al.*, 2012 ; Rasconi *et al.*, 2014 ; Van den Wyngaert *et al.*, 2022), étaient les plus abondants. Leur abondance relative s'est avérée être deux fois supérieure dans le mixolimnion suggérant la présence de nombreux stades de dissémination et/ou stades associés à des hôtes photosynthétiques abondant à la surface du lac. Des variations encore plus importantes ont pu être enregistrées en fonction des saisons. En effet, les périodes de brassage, au cours desquelles les communautés hôtes telles que les diatomées (Ochrophyta) étaient particulièrement abondantes (*e.g.*, au printemps et en automne), semblaient également profiter aux Chytridiomycota dont l'abondance relative était maximale en avril.

Les données d'expression différentielle ont montré des variations similaires en rapportant l'abondance de nombreux transcrits au cours de ces mêmes périodes de brassage. Certains, impliqués dans la phototaxie des zoospores, joueraient un rôle significatif dans l'établissement des interactions hôte-parasites. Les stades de dissémination de certains genres de Chytridiomycota sont en effet décrits comme possédant une organelle, sensible à la lumière, qui leur permettrait de cibler plus efficacement leurs hôtes (Kazama, 1972 ; Medina et Buchler, 2020).

Par ailleurs, bien qu'aucune différence d'abondance n'ait pu être enregistrée entre les deux fractions de taille, nous avons tout de même identifié des profils d'expression de gènes bien différents susceptibles de refléter divers stades du cycle parasitaire des Chytridiomycota. Ainsi, nous avons rapporté dans la grande fraction taille la surexpression de gènes potentiellement impliqués dans le mode de vie attaché, au cours duquel ces parasites doivent dégrader les membranes et parois de leurs hôtes et détourner les réserves lipidiques (Gerphagnon *et al.*, 2019). Au contraire, la surexpression de gènes impliqués dans le métabolisme des lipides polaires, qui a récemment été décrite comme ayant lieu au cours de la maturation des zoospores (Laundon *et al.*, 2022), a été rapportée dans la petite fraction de taille. Le métabolisme des lipides chez les Chytridiomycota, comme pour la plupart des parasites, s'est avéré extrêmement développé, reposant sur une diversité de protéines importantes (chapitre III). Cette diversité pourrait potentiellement être la conséquence de phénomènes de coévolution entre hôtes et parasites, suggérant que certaines de ces fonctions soient directement impliquées dans les interactions parasitaires. Ainsi, le succès de l'établissement d'une interaction résiderait en partie sur la performance de ces gènes (Ramakrishnan *et al.*, 2013 ; Shunmugam *et al.*, 2022).

D'autre part, l'abondance relative des Chytridiomycota enregistrée dans le mixolimnion en septembre coïncidait avec l'augmentation de l'abondance relative d'espèces hétérotrophes affiliées aux Ciliophora. Il est, en effet, probable que ces derniers utilisent les nombreux stades de dissémination disponibles en surface comme source de nourriture. Par ailleurs, il est aussi probable que la fragmentation des colonies d'hôtes photosynthétiques (*e.g.*, diatomées, chlorophycées ou cyanobactéries) induite lors des épidémies de Chytridiomycota (Rasconi *et al.*, 2014 ; Agha *et al.*, 2016 ; Haraldsson *et al.*, 2018) favorise leur comestibilité, ceci en faveur des Ciliophora. Ces hypothèses sont toutes deux en accord avec d'autres résultats récents, obtenus en métatranscriptomique, décrivant ces ciliés hétérotrophes comme « bénéficiaires » des « épidémies » de Chytridiomycota (Li *et al.*, 2022).

Bien que d'autres lignées d'eucaryotes unicellulaires parasites aient pu être identifiées au cours de ces études, le lien entre leur diversité taxonomique et leur potentiel métabolique n'a pu être exploré. C'est le cas notamment du genre *Chytriodinium* (affilié aux Dinoflagellata) dont l'abondance relative importante a été enregistrée en métabarcoding en avril et pour lequel aucun transcrit n'a pu être identifié (chapitre II). Décrit uniquement en milieu marin comme ectoparasite d'œufs de crustacés (Gómez *et al.*, 2009), il est probable qu'il infecte les œufs de copépodes, abondants à cette période dans le Pavin (Lair, 1975). Les microsporidies sont également des parasites absents de nos études utilisant des approches « -omics » (chapitres II et III), même si l'étendue de leur diversité a pourtant été récemment montrée en milieu lacustre (Chauvet *et al.*, 2022 ; annexe 1). Les résultats obtenus dans le chapitre IV, mettant en évidence une nouvelle espèce de microsporidie parasitant (avec une forte prévalence) un genre de rotifère abondant en milieu lacustre, suggèrent, au côté d'autres données de la littérature (Ebert *et al.*, 2000 ; Gorbunov et Kosova, 2001 ; Wolska et Mazurkiewicz-Zapałowicz, 2013), leur rôle prépondérant dans la régulation d'espèces prédatrices. Bien que des études supplémentaires soient nécessaires pour confirmer ces hypothèses, l'impact de ces deux types de parasites sur les populations de copépodes et de rotifères, régulant eux-mêmes les espèces phytoplanctoniques, pourrait effectivement jouer un rôle important dans la structuration du réseau trophique.

Par ailleurs, l'exploration approfondie de la diversité fonctionnelle des eucaryotes unicellulaires parasites à travers l'étude de réseaux de similarité de séquences (SSN ; chapitre III), a confirmé le rôle dominant et hautement complexe des Fungi dans les écosystèmes lacustres. En effet, les protéines affiliées aux Fungi représentaient la moitié de l'ensemble des protéines associées aux parasites. De plus, contrairement aux Alveolata, Euglenozoa, Metamonada et Rhizaria, dont les représentant parasites, strictement obligatoires (*e.g.*, Apicomplexa, Parabasalia, Syndiniales, Plasmodiophorida et Trypanosomatida) possédaient des protéines spécifiques de leur stratégie trophique, les Fungi tout comme les Oomycota comptaient une variabilité importante de modes de vie avec bien moins de signatures protéiques apparentes. Le cas des Chytridiomycota illustre tout à fait ce phénomène en regroupant des espèces parasites obligatoires et facultatives, voir exclusivement saprotrophes (Wurzbacher *et al.*, 2014 ; Van den Wyngaert *et al.*, 2022).

A défaut de mettre en évidence des marqueurs protéiques généralistes du parasitisme nous avons pu identifier des protéines dont la présence était corrélée au mode de vie parasite de certaines lignées eucaryotes.

Nous avons par exemple mis en évidence des protéines dont le rôle était déjà bien décrit, impliquées dans l'immunogénicité des trypanosomes ou représentant des facteurs de virulence des Labyrinthulomycetes et des Fungi. Toutefois, la majorité des protéines identifiées manquait de descriptions dans la littérature, complexifiant ainsi grandement la compréhension de leur rôle dans un contexte de parasitisme. L'absence de ces descriptions est la conséquence du manque important de cultures de parasites unicellulaires et d'études des mécanismes impliqués dans les relations hôte-parasites (del Campo *et al.*, 2014 ; Bass *et al.*, 2015).

Les analyses de données « -omics » que nous avons réalisées dans les chapitre II et III pourraient être complétées par des approches de cytométrie en flux avec imagerie couplées à des approches de machine learning afin de permettre la caractérisation morphologique des associations hôte-parasites et le tri entre microorganismes parasités et sains (Luo *et al.*, 2021). Ainsi, des études de transcriptomique (« single holobiont transcriptomic ») permettraient, à la fois, de caractériser et confirmer la diversité taxonomique de l'holobionte mais aussi d'analyser les cycles de vie, les états cellulaires, la physiologie et les interactions écologiques à travers l'expression des gènes (Ku et Sebé-Pedrós, 2019). D'autre part, nos analyses de réseaux de similarité pourraient également être améliorées par la prise en compte des variables environnementales permettant ainsi d'intégrer les résultats d'expression différentielle et de contextualiser la diversité des protéines décrites dans chacune des conditions. A titre d'exemple, l'étude de notre jeu de données en ciblant spécifiquement les Chytridiomycota, couplée à l'ajout des données d'expression différentielle afférentes aux zones d'échantillonnage (mixolimnion vs monimolimnion), permettrait une meilleure compréhension des mécanismes et des protéines impliquées dans leurs différentes stratégies trophiques. En effet, il est probable que nous mettions en évidence des protéines signatures différentes entre les Chytridiomycota saprotrophes et parasites et dont l'expression dépendrait de leur zone de résidence.

B. Apport des approches *in-vitro*

Bien que les approches « -omics » se soient révélées incontournables pour étudier la diversité taxonomique et fonctionnelle des eucaryotes unicellulaires parasites (chapitres II et III), tout en révélant de potentiels hôtes par la mise en place de réseaux de cooccurrence (annexe I), elles ne permettent pas, à elles seules, d'évaluer l'impact de ces parasites sur les communautés planctoniques ni d'étudier les mécanismes cellulaires sous-jacents.

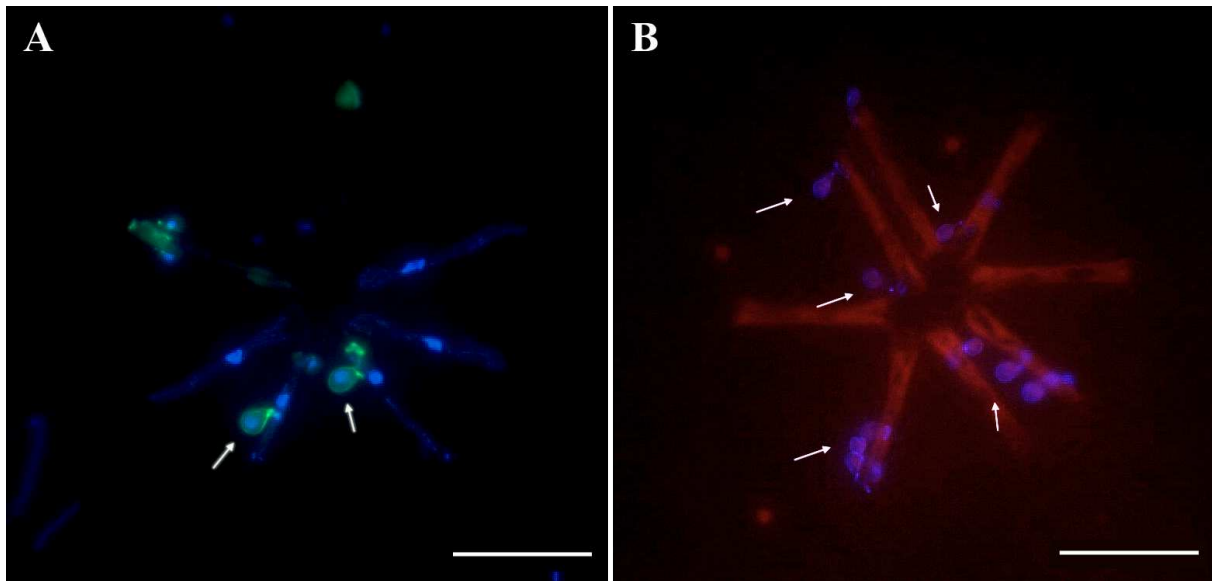


Figure 5 : Observations au microscope à épifluorescence de diatomées *Asterionella formosa* issues du microcosme infecté et marquées au DAPI, au DY96 et au calcofluor blanc (x100).

L'observation est réalisée sept jours après mise en contact avec le concentrat.

A : Superposition des observations des marquages au DAPI et au DY96. Le DAPI (intercalant de l'ADN) fluoresce en bleu tandis que le DY96 (marqueur de la chitine, molécule présente dans la paroi de certains parasites) fluoresce en vert. Le noyau des diatomées est visible au centre des cellules.

B : Superposition des observations de l'auto-fluorescence de la diatomée (en rouge, présence de chlorophylle dans les chloroplastes) et de la fluorescence bleue (marquage au calcofluor blanc).

Les potentiels parasites sont indiqués par des flèches. Barres d'échelle : 20 μm .

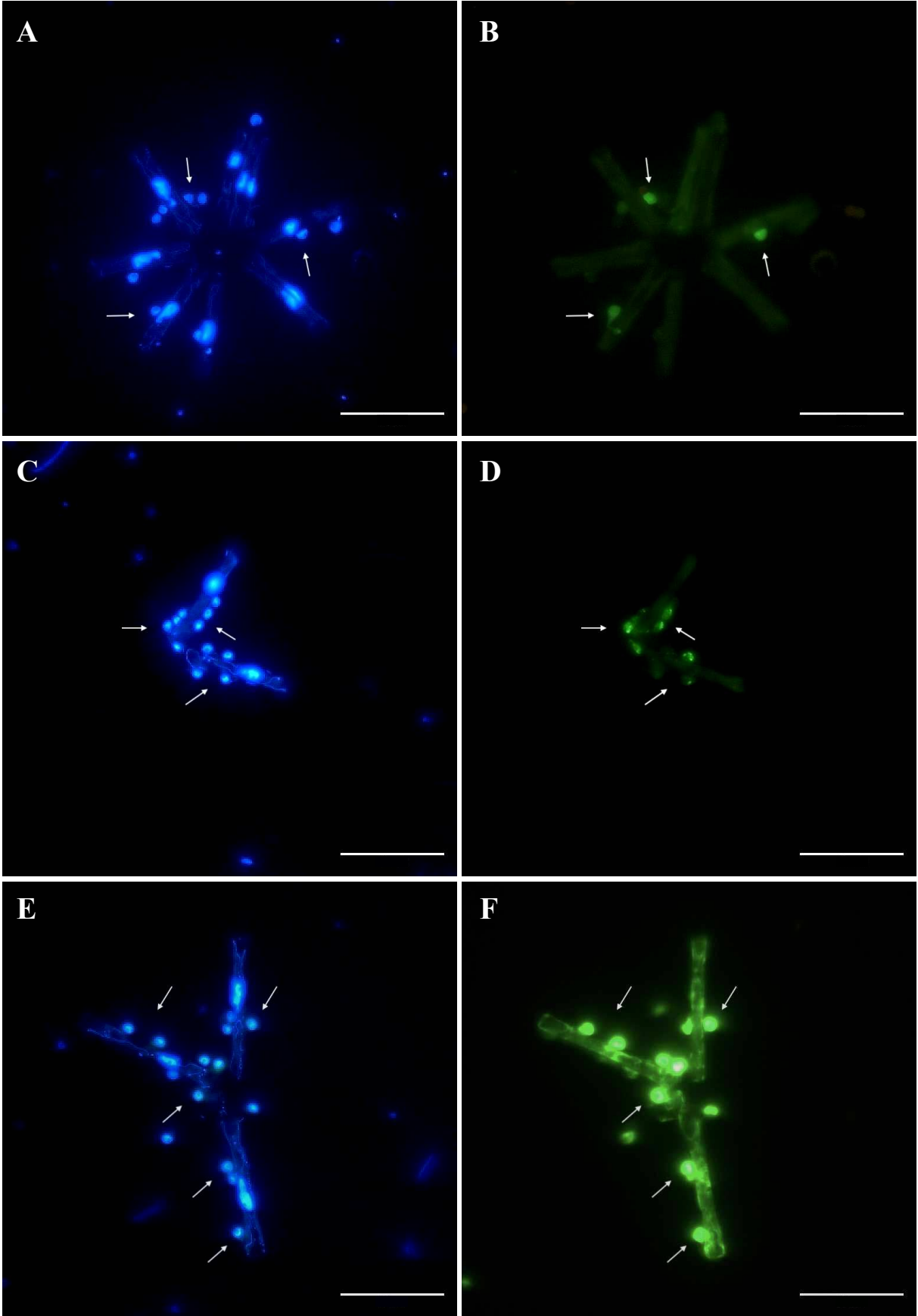


Figure 6 : Observations au microscope à épifluorescence de diatomées *Asterionella formosa* issues du microcosme infecté et marquées au DAPI et en TSA-FISH avec trois sondes ciblant différents microorganismes affiliés aux Fungi (x100).

L'observation est réalisée sept jours après mise en contact avec le concentrat.

A, C et E : Marquage au DAPI des noyaux des cellules.

B : Marquage en TSA-FISH avec la sonde MY1574 qui s'hybride spécifiquement aux ARNr 18S des Fungi.

D et F : Marquage en TSA-FISH avec les sondes CM1.1 et LKM11_01 qui s'hybrident spécifiquement aux ARNr 18S des différents clades de *Cryptomycota*.

Le marquage TSA-FISH est visible en vert. Les potentiels parasites sont indiqués par des flèches. Barres d'échelle : 20 μm .

Dans ce contexte, afin (i) de mettre en évidence des associations parasitaires qu'il est très difficile de repérer en milieu naturel, (ii) de permettre la caractérisation morphologique et moléculaire des deux partenaires et (iii) d'obtenir de nouveaux couples hôte-parasites en culture, plusieurs expériences de tests d'infection en microcosmes ont été réalisées au cours de cette thèse.

Différentes cultures d'hôtes potentiels ont été exposées à du concentrat d'eau du lac d'Aydat ainsi qu'à du sédiment, tous deux susceptibles de contenir des stades de dissémination de divers parasites (*e.g.*, spores et zoospores). L'objectif était alors de provoquer la rencontre entre un hôte et un parasite afin d'induire des « infections » et d'en suivre l'évolution en milieu contrôlé. Des cultures de ciliés (*Vorticella similis*, *Euplotes aediculatus* et *Paramecium tetraurelia*), de zooplanctons (*Daphnia magna*), de chlorophycées (*Coelastrum reticulatum*) et de diatomées (*Asterionella formosa*) ont été utilisées afin d'étudier des organismes de différents niveaux trophiques. Nous avons effectué des prélèvements réguliers, sur une période de 20 jours, afin de réaliser un suivi de croissance et de détecter de potentielles associations hôte-parasites en microscopie à l'aide de différents marquages (*i.e.*, marquage de paroi et hybridation *in situ*). De tels tests d'infection ont déjà donné lieu à des résultats positifs, notamment chez *Daphnia magna* parasitée par des microsporidies (Ebert, 1995 ; Decaestecker *et al.*, 2004 ; Jansen *et al.*, 2010), ou plus récemment chez des hôtes dinoflagellés parasités par des Chytridiomycota puis maintenus en culture à partir d'inocula (Fernández-Valero *et al.*, 2022).

Dans notre cas, cette approche ne s'est pas révélée être la plus efficace en présentant un taux de réussite relativement faible et des résultats non reproductibles. Sur les quatre essais réalisés, un seul a permis, pour un des réplicats d'expérience, l'observation d'associations diatomées-parasite. En effet, les prélèvements réguliers ont permis de mettre en évidence la présence de cellules eucaryotes de forme ovoïde fixées aux diatomées (**Fig. 5**). Parmi les cellules fixées, respectivement 26,5% et 45,5% étaient marquées après coloration au DY96 (**Fig. 5A**) et au calcofluor blanc (**Fig. 5B**) suggérant ainsi que la paroi de la majorité d'entre elles était constituée de chitine (Hoch *et al.*, 2005 ; Rasconi *et al.*, 2009), ce qui est caractéristique de certains stades de développement d'organismes affiliés aux Fungi (James et Berbee, 2012). En parallèle, des marquages TSA-FISH ont été réalisés en utilisant des sondes ciblant la séquence du gène de l'ARNr 18S de différents clades affiliés aux Fungi. Parmi elles, trois sondes ont permis l'acquisition d'un signal positif (**Fig. 6**).

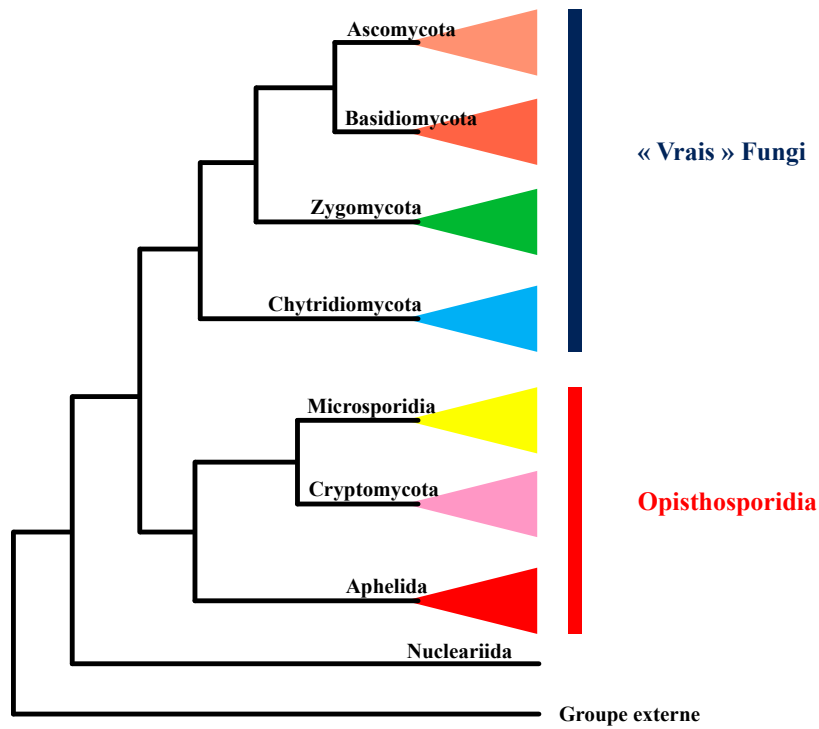


Figure 7 : Arbre phylogénétique simplifié des Fungi construit sur la base d'analyses phylogénétiques du gène codant la petite sous-unité de l'ARNr (modifiée d'après Karpov *et al.*, 2014).

La sonde MY1574 (**Fig. 6A-B**), spécifique des Fungi (Baschien, 2003), marquait 12,5% des cellules fixées aux diatomées tandis les sondes CM1.1 (**Fig. 6C-D**) et LKM11_01 (**Fig. 6E-F**), spécifiques des Cryptomycota (sous embranchement des Fungi (Mangot *et al.*, 2009 ; Jones *et al.*, 2011)) (**Fig. 7**), en marquaient respectivement 18,5% et 39%. En considérant que la sonde MY1574 spécifique des « vrais » Fungi (**Fig. 7**) exclut les Cryptomycota, il est probable qu'une minorité des organismes marqués (12,5 % des cellules fixées) soit affiliée aux Chytridiomycota qui sont connus pour parasiter différentes espèces de diatomées, et notamment la diatomée *Asterionella* dans les écosystèmes lacustres (Rasconi *et al.*, 2012 ; Rad-Menéndez *et al.*, 2018 ; Van den Wyngaert *et al.*, 2022). Les marquages TSA-FISH réalisés avec les sondes CM1.1 et LKM11_01 ont, quant à eux, permis d'identifier qu'une majorité des organismes fixés aux diatomées était affiliée aux Cryptomycota. Dans la littérature, seules deux études ont montré, à partir d'échantillons naturels, des Cryptomycota associés à des diatomées dans les écosystèmes marins et d'eau douce (Jones *et al.*, 2011 ; Lepère *et al.*, 2019). Malheureusement, nous n'avons pas pu identifier plus précisément ces parasites et ces observations n'ont pu être reproduites. L'échec relatif de cette méthodologie nous a amené à utiliser une approche différente basée sur la micromanipulation d'individus potentiellement parasités à partir de l'environnement afin de caractériser moléculairement les deux partenaires. Ainsi, dans le chapitre IV, après avoir isolé manuellement et un à un, différents genres de rotifères du lac d'Aydat, nous avons découvert, au sein d'individus appartenant au genre *Kellicottia*, la présence d'une potentielle nouvelle espèce de microsporidie appartenant au genre *Crispospora*. Le genre *Crispospora* n'est, à l'heure actuelle, défini que par une espèce (*Crispospora chironomi*) parasite de larve de diptère (Tokarev *et al.*, 2010). Cette nouvelle association, que nous avons également confirmée en hybridation *in situ* (**Fig. 6**), est non négligeable puisque la prévalence était de 42%. Bien que chronophage et nécessitant une extrême précision, cette approche de micromanipulation couplée à de l'amplification du génome complet, a prouvé qu'elle permettait la mise en évidence d'associations hôte-parasites et leur caractérisation moléculaire tout en s'affranchissant des biais inhérents à la culture de microorganismes eucaryotes. En théorie, cette approche permettrait également de séquencer les génomes complets de l'hôte et de son parasite. Étant donné le nombre réduit de couples hôte-parasites unicellulaire en culture, ces approches représentent de véritables perspectives pour la caractérisation morphologique et génétique de ces parasites ainsi que pour la description de leur cycle de développement et de leur effet sur la survie et la « fitness » des populations hôtes.

Table 1 : Récapitulatif des clones séquencés issus du tri en cytométrie en flux.

Clone	Définition courte	Accession	e-value	%Id	Taxonomy
Clone 1	<i>Cryptomonas curvata</i>	KF907370	0	100%	Cryptophyta
Clone 2	<i>Plagioselmis nannoplanctica</i>	FM876311	0	100%	Cryptophyta
Clone 3	Uncultured Rhizophyidium	LS986819	3.0858e-57	85.31%	Chytridiomycota
Clone 4	Uncultured Chytridiomycota	HQ219421	5.85495e-49	97.90%	Chytridiomycota
Clone 5	Uncultured Zygothlyctidales	OL869112	2.72876e-70	97.56%	Chytridiomycota
Clone 6	<i>Zygorhizidium affluens</i>	MH626496	5.32999e-93	100%	Chytridiomycota
Clone 7	Fungal sp.	MH625123	3.45378e-12	97.96%	Fungi
Clone 8	Fungal sp.	MT587149	4.24761e-68	98.06%	Fungi
Clone 9	Uncultured fungus	MF793694	1.12237e-21	98.51%	Fungi
Clone 10	Uncultured fungus	LS993366	4.09783e-60	87.22%	Fungi
Clone 11	Uncultured fungus	MF791570	1.59575e-63	90.0%	Fungi
Clone 12	<i>Neoxylomyces multiseptatus</i>	MW260332	1.01351e-35	87.94%	Ascomycota
Clone 13	<i>Simplicillium aogashimaense</i>	MK579181	1.63907e-103	100%	Ascomycota
Clone 14	Uncultured Microsporidia	MZ405751	3.48643e-177	98.88%	Microsporidia
Clone 15	Uncultured Microsporidia	ON026095	1.16848e-166	99.72%	Microsporidia
Clone 16	Uncultured Microsporidia	MG241423	9.99472e-143	93.47%	Microsporidia
Clone 17	Uncultured Microsporidia	MZ405795	1.26962e-176	98.33%	Microsporidia

La littérature (Jobard *et al.*, 2010 ; Frenken *et al.*, 2017 ; Chambouvet *et al.*, 2019 ; Van den Wyngaert *et al.*, 2022) ainsi que les résultats obtenus lors de ces différents travaux ont montré qu'une majorité de parasites lacustres était affiliée aux Fungi et qu'ils étaient impliqués dans de nombreuses associations (*i.e.*, représentant près de 50% de la diversité des parasites (chapitre II) et parasitant des hôtes de divers niveaux trophiques, avec une prévalence importante (**Figs. 5, 6** ; chapitre IV et annexe 1)). Dans ce contexte, le développement d'une méthode de tri en cytométrie en flux couplée à un marquage de la chitine (au DY96 (Hoch *et al.*, 2005)), caractéristique de la paroi de certains stades de développement des Fungi (James et Berbee, 2012), a été initiée afin de cibler spécifiquement ces organismes et ainsi faciliter leur étude.

Les premiers tests réalisés au cours de cette thèse nous ont permis, à partir d'un concentrat d'eau du lac d'Aydat (obtenu par filtration tangentielle), de trier de nombreux microorganismes (~ 510 000 cellules) marqués au DY96. L'extraction d'ADN couplée au criblage par PCR, clonage, séquençage et comparaison par BLAST contre la base de données NR (<https://blast.ncbi.nlm.nih.gov>) a permis d'identifier certains d'entre eux. Sur 17 clones séquencés, 2 étaient associés aux Cryptophyta et 15 aux Fungi (**Table 1**). Les Cryptophyta ne disposant pas de paroi constituée de chitine, il est probable qu'ils aient été triés à cause de leur auto-fluorescence. Un redimensionnement de la fenêtre de tri en cytométrie devrait donc corriger cela. Parmi les clones affiliés au Fungi, 4 étaient apparentés aux microsporidies, 4 autres aux Chytridiomycota, 2 aux Ascomycota et, finalement, 5 n'ont pas pu être identifiés, n'étant associés qu'à des séquences environnementales du type « fungal sp. » ou « uncultured fungus » (**Table 1**). Parmi les 10 clones associés aux microsporidies, Chytridiomycota ou Ascomycota, une majorité était clairement affiliée à des espèces parasites. Le clone n°15, par exemple, était associé, avec un pourcentage de similarité de 99,72%, à la potentielle nouvelle espèce parasitant les rotifères du genre *Kellicottia* retrouvée dans le chapitre IV. La séquence du clone n°6, quant à elle, était identique à celle de référence appartenant à l'espèce *Zygorhizidium affluens*, décrite comme parasite de la diatomée *Asterionella formosa* (Rad-Menéndez *et al.*, 2018) et disponible en culture (CCAP 4086/1). Enfin, le clone n°13 était associé à l'espèce *Simplicillium aogashimaense*, décrite comme hyper-parasite de l'espèce phytopathogène *Blumeria graminis* également affiliée aux Ascomycota (Zhu *et al.*, 2022).

Ces tests préliminaires semblent donc prometteurs pour l'analyse et l'isolement d'espèces parasites affiliées aux Fungi. Ainsi, en combinant cette méthodologie aux approches de microcosme, il semble envisageable de maximiser nos chances de réussite et ainsi faciliter la mise en culture de ces parasites en association avec leurs hôtes.

ANNEXE I

MICROSPORIDIES DANS LES ÉCOSYSTÈMES
LACUSTRES ET MARINS : IMPORTANTE DIVERSITÉ
ET NOUVEAUX HÔTES PLANCTONIQUES

High diversity of microsporidian parasites and new planktonic hosts in freshwater and marine ecosystems

Marina Chauvet¹°, Arthur Monjot¹, Cécile Lepère¹

1- CNRS, Laboratoire Microorganismes : Génome et Environnement, Université Clermont Auvergne, Clermont-Ferrand, F-63000, France.

°corresponding author: Chauvet.Marina.pro@gmail.com

Publié dans *Limnology and Oceanography* – Février 2023

doi: 10.1002/lno.12321

I. Abstract

Microsporidia are a large group of obligate intracellular eukaryotic parasites related to fungi primarily known as parasites of vertebrates and invertebrates. They are well described as parasites of organisms of interest (*e.g.*, edible fish and crustaceans, honeybees, bioindicators such as daphnia, humans) on which they can have an important impact (*e.g.*, reduced survival or fecundity and sex ratio distortion). However, their diversity in aquatic environments, especially in marine ecosystems, has been greatly understudied since they are not targeted by classical eukaryotic primers used in metabarcoding studies. Moreover, little is known about their hosts among protists or microzooplankton and therefore about their impact on the trophic food web functioning. In this work, we sampled 15 different sites across marine and freshwater environments, size-fractionated the samples, and used microsporidian specific primers associated with metabarcoding to study the microsporidian diversity (and the associated spatial variation). Co-occurrence networks as well as tyramide signal amplification fluorescent *in situ* hybridization were used to link potential hosts (planktonic eukaryotes <150 µm) and Microsporidia diversity. Our analysis unraveled a large microsporidian diversity which was widely divergent between the two environments studied. In both of them, an important part of this diversity was not affiliated to a genus, suggesting an important reservoir of new microsporidian species and thus new hosts among planktonic eukaryotes. Co-occurrence networks and fluorescence microscopy showed for the first-time associations between Microsporidia and dinoflagellates in freshwater and marine environments.

II. Introduction

Parasitism has been described as one of the oldest and most ubiquitous interactions in nature (Cavalier-Smith, 1993); however, the diversity of parasites and their effects on the aquatic trophic food web functioning have long been neglected (Marcogliese and Cone, 1997). In the last decades, high throughput sequencing allowed the description of an increasing number of parasites (Singer *et al.*, 2021) and revolutionized our understanding of aquatic food webs in term of trophic interactions as well as of energy and biomass flows (Lafferty *et al.*, 2008; Kagami *et al.*, 2014). Among these parasites, Microsporidia (obligate intracellular eukaryotic parasites related to fungi; Karpov *et al.*, 2014) were known in aquatic ecosystems only through the search of infectious agent in organisms of interest (*e.g.*, edible fish and crustaceans, bioindicators, humans, etc.). These host-targeted studies show that Microsporidia can have a considerable impact on host populations that are of great importance in freshwater ecosystems, by causing a decrease in the survival and fecundity rates of daphnia for example (Ebert, 2005). However, they were missing from metabarcoding studies focusing on microbial eukaryotes diversity (Bass *et al.*, 2015) especially due to their divergence from other eukaryotes (rapid evolution and compaction of their genome; Keeling and Fast, 2002) making the primers unable to detect them (Hadziavdic *et al.*, 2014; Hugerth *et al.*, 2014). As a result, the overall microsporidian diversity in aquatic (freshwater and marine) ecosystems is largely unknown and thus their impact on ecosystem functioning is neglected.

In 2018, a targeted metabarcoding study (*i.e.*, using specific primer for Microsporidia) presented the first description of the microsporidian diversity in various environments including freshwater and marine samples (Williams *et al.*, 2018), but this allowed only a limited analysis due to the low number of operational taxonomic units (OTUs) obtained. This study nevertheless allowed the identification of a new diversity across the microsporidian tree and encouraged further work. Indeed, two others metabarcoding studies focusing on lacustrine ecosystems followed and allowed the discovery of an unsuspected diversity of Microsporidia (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022a). In these studies, between 40% and 50% of the uncovered diversity could not be affiliated to a genus suggesting that the microsporidian phylum hide a large amount of undiscovered diversity. These studies also provided temporal variation data and highlighted new microsporidian planktonic hosts in a small size fraction (*i.e.*, <150 µm; Chauvet *et al.*, 2022a,b).

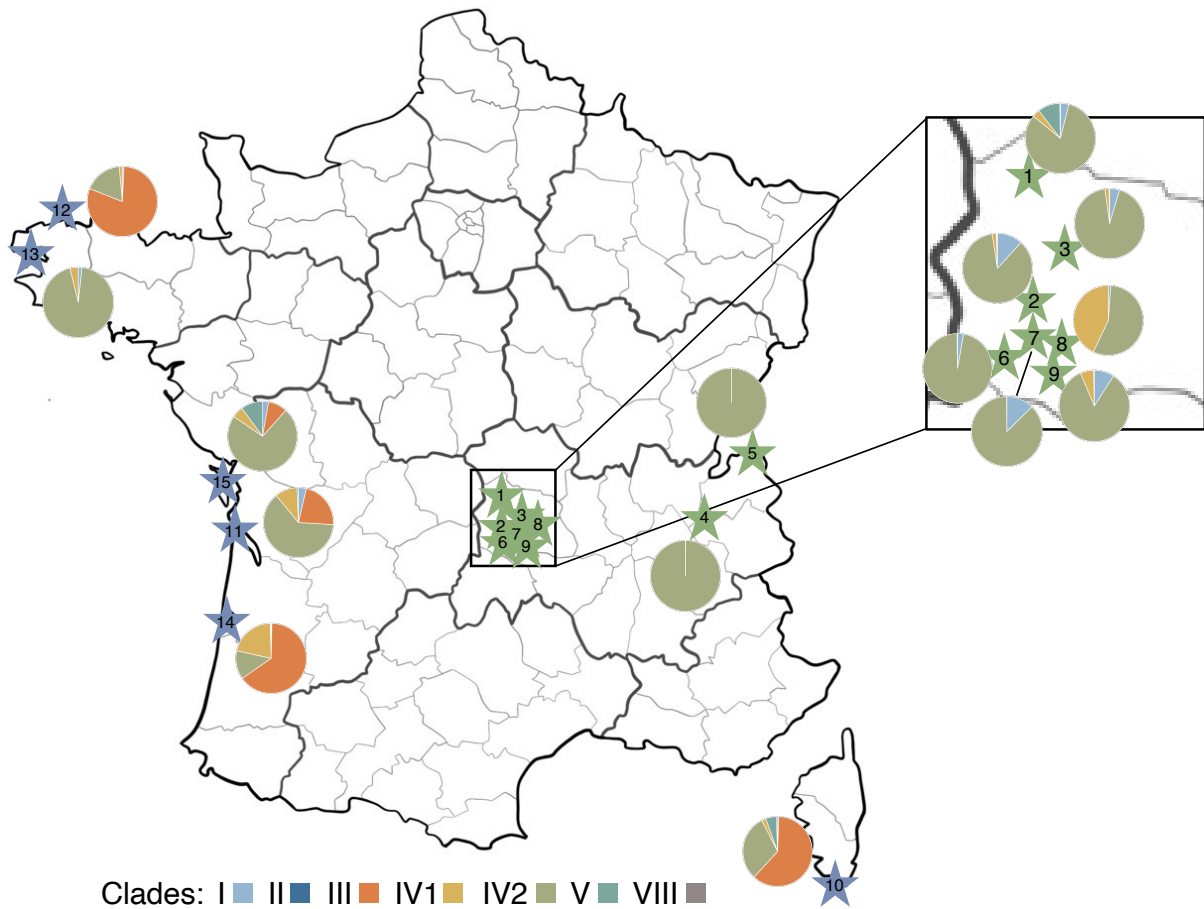


Figure 1: Sample collection site.

Samples were collected in France from freshwater lakes (green stars): 1, Saint-Gervais (hypereutrophic); 2, Chambon (eutrophic); 3, Aydat (eutrophic); 4, Bourget (oligomesotrophic); 5, Lemans (mesotrophic); 6, Chauvet (oligotrophic); 7, Pavin (oligomesotrophic); 8, Estivadoux (temporary marsh); 9, Bourdouze (oligotrophic); and marine environments (blue stars): 10, Corsica (bay); 11, Royan (estuary); 12, Roscoff (coastal); 13, Brest (bay); 14, Arcachon (bay); 15, La Rochelle (coastal). Pie charts represent the proportions of microsporidian clades identified in each sample.

Moreover, few Microsporidia are currently described as parasites of protists or microzooplankton (Murareanu *et al.*, 2021), even though they are major components of food webs.

Altogether, the previous studies did not allow a broad comparison of the diversity of Microsporidia across aquatic ecosystems and thus lead to an incomplete picture of their diversity. Although Dubuffet *et al.* (2021) and Chauvet *et al.* (2022a,b) showed an in-depth description of the microsporidian diversity, these studies were limited to two lakes (Aydat and Pavin, Massif Central, France). Moreover, the microsporidian diversity in marine ecosystems is probably still underestimated as it has been neglected so far. Here, spatial variation of the microsporidian diversity (canonical [or long-branch] Microsporidia; Bass *et al.*, 2018) was assessed through targeted metabarcoding in nine different freshwater and six marine ecosystems. Co-occurrence networks and fluorescence microscopy were eventually used to identify potential new microsporidian hosts among microeukaryotes.

III. Materials and methods

A. Study sites and sampling procedure

Water samples were collected in France from different freshwater lakes ($n = 9$) and marine environments ($n = 6$; Mediterranean Sea and Atlantic Ocean; **Fig. 1**) between 18 June 2021 and 27 September 2021 (**Table S1**). Then, 2.5-22 liters of water were sampled at the surface with a Niskin bottle or along the euphotic zone using an integrated sampler IWS (HYDRO-BIOS; **Table S1**).

For most of the samples, we collected the 0.8-150 μm fraction to concentrate potentially infected planktonic hosts except for the Brest and Roscoff samples where we were able to collect the $>3 \mu\text{m}$ fraction (**Table S1**). Part of the microbial biomass of each size fraction was collected in 0.8 μm -pore-size polycarbonate filters (Millipore) at a very low vacuum to prevent cell damage ($<15 \text{ kPa}$) and stored at -80°C until nucleic acids extraction. For some samples (**Table S1**), the rest of the microbial biomass was fixed with paraformaldehyde (final concentration, 2%), incubated at 4°C for 24 h and collected in 0.8 μm pore-size polycarbonate filters (Millipore; pressure $<15 \text{ kPa}$).

Filters were preserved by dehydration in an ethanol series (50%, 80%, and 100% ethanol for 3 min each) and stored at -20°C until tyramide signal amplification-fluorescent *in situ* hybridization (TSA-FISH) staining.

B. Nucleic acids extraction, amplification and sequencing

Filters prepared for DNA extraction were processed as described by Dubuffet *et al.* (2021).

Amplification of the V1-V3 region of the microsporidian gene coding for the small subunit (SSU) rRNA was performed using the universal microsporidian V1F and 530R primers set (Zhu *et al.*, 1993; Baker *et al.*, 1995). Even though other recently developed primers have shown their efficiency (*i.e.*, CM-V5F CM-V5R, ~ 200 bp; Trzebny *et al.*, 2020), this set of primers amplify a fragment of ~ 400 bp and has shown its efficiency in the detection of aquatic Microsporidia (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022a,b). Each polymerase chain reaction (PCR) was performed in a total volume of 30 µL containing 6 µL of 5x Green GoTaq Flexi buffer, 2.4 µL of 25 mM MgCl₂, 0.18 µL of GoTaqG2 Flexi DNA Polymerase (Promega; 5 U/µL), 0.6 µL of 10 mM each deoxyribonucleotide triphosphates (dNTP), 0.3 µL of 50 mg mL⁻¹ bovine serum albumin (BSA), and 0.6 µL of each 10 µM primer. The amplification conditions consisted of initial denaturation at 95°C for 5 min followed by 35 cycles for 1 min at 95°C, 1 min 30 s at 62°C and 1 min 30 s at 72°C, and a final elongation of 7 min at 72°C.

Amplification of the V4 region of total eukaryotes (excluding Microsporidia) was performed using the universal primer 515F (Caporaso *et al.*, 2011) and the eukaryotic primer 951R (Mangot *et al.*, 2013). Each PCR was performed in the mix described above. The amplification conditions consisted of initial denaturation at 95°C for 5 min followed by 33 cycles for 1 min at 95°C, 1 min at 55°C and 1 min 30 s at 72°C, and a final elongation of 10 min at 72°C.

The amplicons were purified and concentrated using the QIAquick Gel Extraction kit (Qiagen). Samples were then sequenced on a MiSeq Illumina instrument (2x250 bp; Microsynth AG).

C. Amplicon analysis and taxonomic affiliation

Sequencing data were processed using the PANAM2 (Phylogenetic Analysis of Next-generation AMplicons v2, [https:// github.com/meb-team/PANAM2](https://github.com/meb-team/PANAM2)) pipeline (Taib *et al.*, 2013).

Sequences with ambiguous ‘N’ bases, those below 200 bp, chimeras or sequencing errors in the primers were eliminated. The sequences were then assembled (overlap of 50 bp minimum between forward and reverse reads, maximum two mismatches in the overlap) and clustered into OTUs with a similarity threshold of 99% for Microsporidia (Dubuffet *et al.*, 2021) and 95% for other eukaryotes (Mangot *et al.*, 2013; Debroas *et al.*, 2017). The representative sequences of each OTU were affiliated using a custom curated reference database (Freshwater-database: <https://github.com/panammeb/Freshwater-database>). Affiliations to microsporidian groups and genera were considered as valid when the identity scores were above 85% and 94.5%, respectively (Dubuffet *et al.*, 2021). As clustering threshold used here (especially for Microsporidia) may generate OTUs resulting from sequencing errors, we only kept for the final analysis OTUs representing more than 0.005% (Bokulich *et al.*, 2013) of the total reads. Data were normalized using the scaling with ranked subsampling method (Beule and Karlovsky, 2020).

Phylogenetic analyses were performed to show the placement of the sequences obtained in this study among the different microsporidian clades (**Figs. S1, S2**). We focused our analyses on abundant OTUs (100 most abundant of each clade, when more than 100 OTUs were identified) and on the five major clades of this study (I, III, IV1, IV2, and V). After an alignment using multiple alignment using fast Fourier transform (MAFFT) algorithm (Katoch *et al.*, 2005), maximum likelihood (ML) trees were obtained using the best predicted model of evolution according to the Bayesian information criterion (BIC) score (general time-reversible model with gamma-distributed rate heterogeneity (G)) with 200 bootstrap replications and partial deletion (85%), using molecular evolutionary genetics analysis (MEGA) 7 (Kumar *et al.*, 2016).

D. Diversity and correlation analyses

Diversity analyses were performed using R (version 3.6.3; R Development Core Team, 2012). Community richness and diversity indices (Richness observed, Chao1, Shannon and Simpson) were used to infer the taxa richness of Microsporidia. These estimators were computed with the phyloseq package (version 1.30.0; McMurdie and Holmes, 2013). Principal coordinates analysis (PCoA) ordination based on a Bray-Curtis dissimilarity matrix (Bray and Curtis, 1957; Paliy and Shankar, 2016) was computed using ape (version 5.6-2; Paradis and Schliep, 2019) and vegan (version 2.5-7; Oksanen *et al.*, 2020) packages.

An UpSet plot was done to show the sets and intersections of OTUs between samples with the UpSetR package (version 1.4.0; Conway *et al.*, 2017).

As microsporidian and eukaryotic communities' composition was largely different between marine and freshwater samples, the correlation analysis was done on each dataset. To reduce the number of tested hypotheses and focus the analysis on “abundant OTUs” only those representing more than 0.05% of the reads were retained, representing 302 eukaryotic OTUs and 253 microsporidian OTUs in freshwater and 237 eukaryotic OTUs and 268 microsporidian OTUs in marine ecosystems. Spearman's ρ correlations were calculated, and to decrease false discovery rate due to multiple testing we used Benjamini-Hochberg corrections (Benjamini and Hochberg, 1995). A correlation was considered as valid when $\rho > 0.8$ and the associated p -value < 0.01 . Correlations were then visualized with chord diagrams using the circlize package (version 0.4.14) on R (R Development Core Team, 2012; Gu *et al.*, 2014).

E. TSA-FISH (Tyramide Signal Amplification-Fluorescent *In Situ* Hybridization)

TSA-FISH was performed using a Microsporidia specific probe (USP01) as described by Chauvet *et al.* (2022a). This probe is suitable for the detection and observation of aquatic Microsporidia, being able to hybridize (*in silico*) with 100% of the microsporidian sequences obtained during recent metabarcoding studies (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022a) as well as all the sequences obtained in the current study. *In silico*, this probe does not hybridize with any sequence other than microsporidian. Hybridization conditions were determined on a target species (*Encephalitozoon cuniculi*) and a nontarget species (*Anncaliia algerae*) having one mismatch with the probe sequence. The probe was also tested on two other Microsporidia in culture: *Nosema ceranae* and *Nematocida parisii*. Observations were performed using a Zeiss Axio Imager 2 microscope (Carl Zeiss).

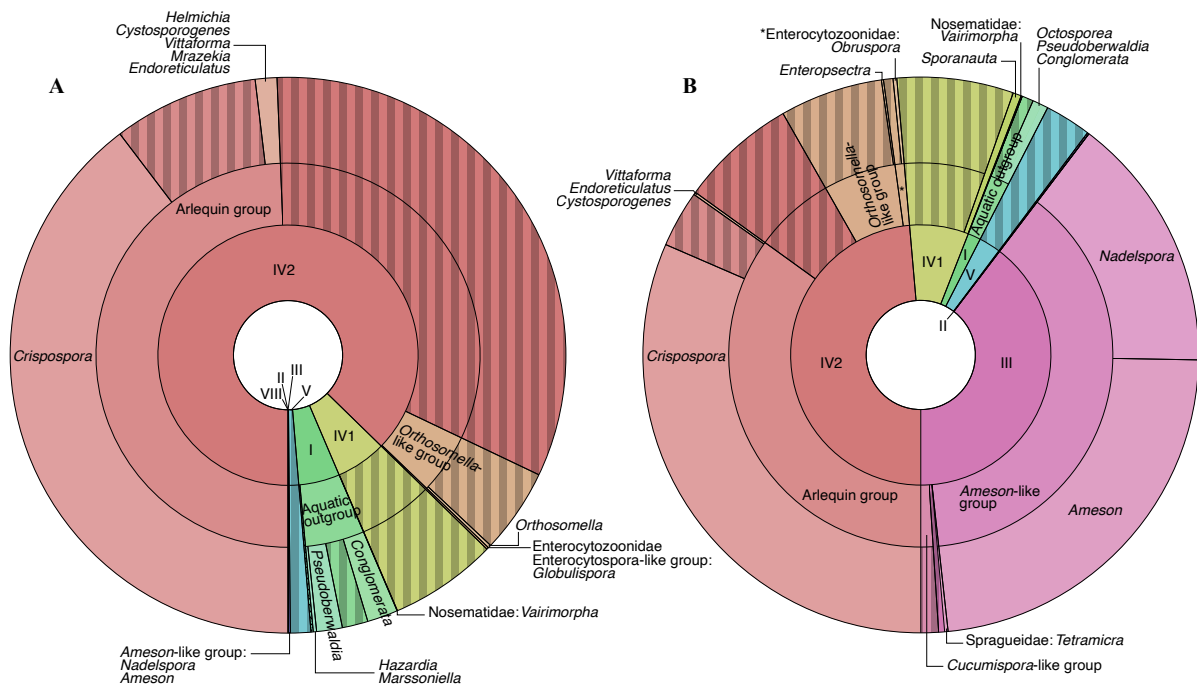


Figure 2: Krona charts representing the taxonomic affiliation of the microsporidian OTUs retrieved in freshwater (A) and marine (B) samples.

Inner circle represents clades. Deeper taxonomic ranks (groups and genera) are presented in outer circles. Genera names are given in italic. Hatched patterns represent OTUs unaffiliated beyond the group/clade. Genera listed are ranked from most to least abundant.

IV. Results and discussion

In this work, we sampled 15 different sites across freshwater ($n = 9$) and marine ($n = 6$) environments (**Fig. 1, Table S1**), size-fractionated the samples and used microsporidian specific primers to unravel microsporidian diversity which once associated with total eukaryotic diversity allow the identification of potential host-parasite associations.

After quality filtering, clustering at 99% and cleaning steps, a total of 1 853 microsporidian OTUs were obtained from 2 944 084 SSU rRNA gene sequencing reads (1 045 560 reads after normalization). OTUs obtained in this study belonged to the canonical (or long-branch; Bass *et al.*, 2018) Microsporidia were all affiliated at least to a clade. Among the 10 clades defined for canonical Microsporidia (Dubuffet *et al.*, 2021), our OTUs were distributed in seven clades: I, II, III, IV1, IV2, V, and VIII (**Figs. 1, S1, S2**); 63.9% of the OTUs were affiliated to a group and 9 groups were identified. Less than half (45.5%) of the OTUs were affiliated to a genus among the 20 genera identified, demonstrating that a large part of the microsporidian diversity remains to be described.

A. Microsporidian diversity across freshwater environments

So far, few studies have investigated the global diversity of Microsporidia in freshwater ecosystems using high throughput sequencing (*i.e.*, metabarcoding). However, these studies (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022a) allowed the identification of a huge and new microsporidian diversity but did not allow a broad comparison of this diversity across different freshwater ecosystems.

Here, we studied this diversity in nine different freshwater sites: eight lakes and one temporary marsh (**Fig. 1, Table S1**). The freshwater microsporidian diversity was largely represented by clade IV2, accounting for over 87% of the total reads (**Fig. 2A**). This clade is highly diverse and is known to parasitize a wide range of hosts such as insects, nematodes, crustaceans and fish, as well as humans, and its members have been identified in terrestrial, freshwater, and marine environments (Dubuffet *et al.*, 2021). *Euplotespora binucleata* and *Globosporidium paramecii* (Fokin *et al.*, 2008; Yakovleva *et al.*, 2020), which are among the few microsporidian species known (SSU sequences available) to parasitize protists (*i.e.*, ciliates), are also found in this clade.

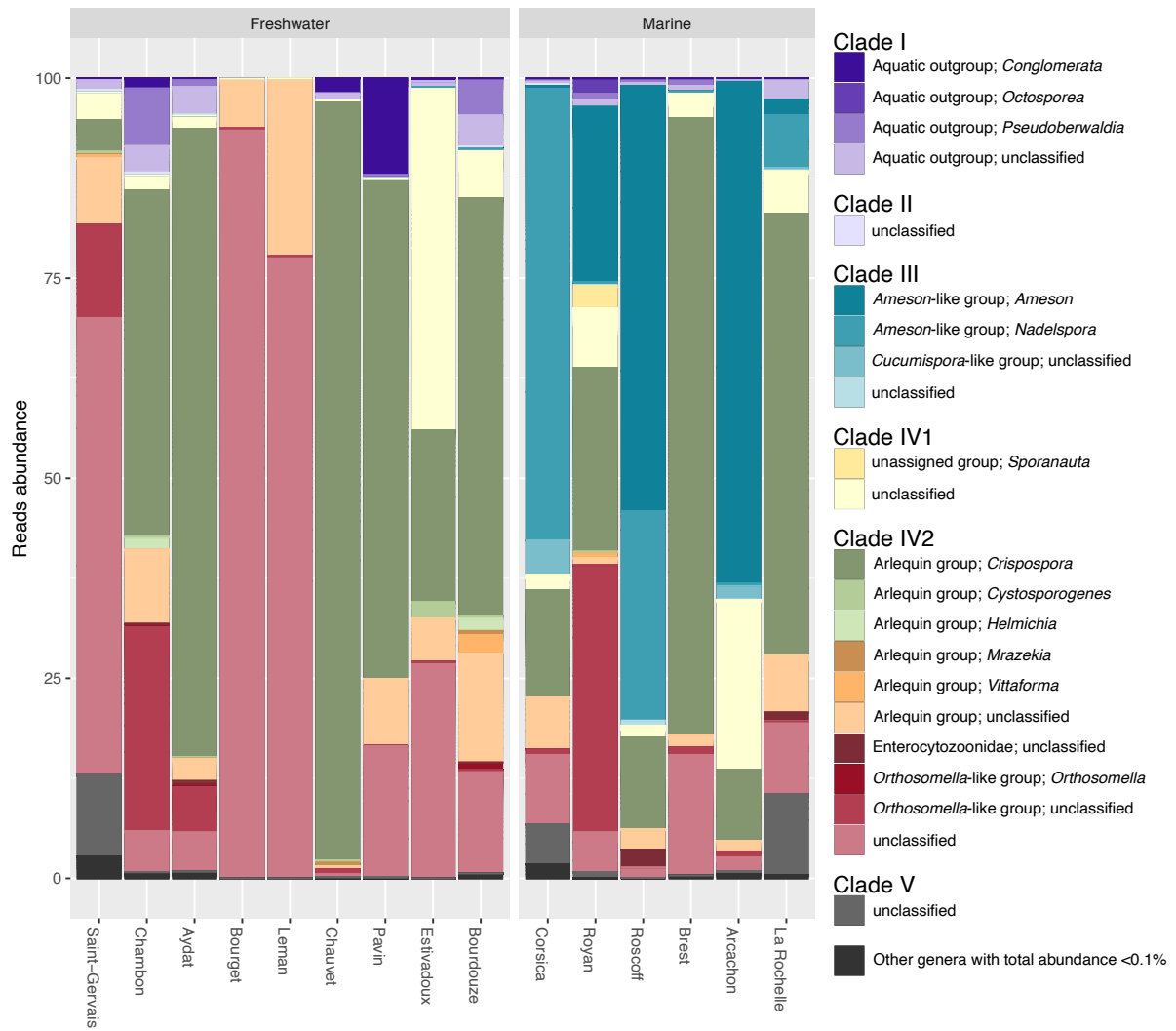


Figure 3: Relative abundances of the most abundant (>0.1% of the total reads) microsporidian genera across freshwater (left) and marine (right) samples.

The black zone at the bottom of each bar represents the cumulative proportion of other genera with low abundances.

Clades IV1 and I were the next most abundant representing 6% and 5% of the OTUs, respectively. The first one mainly includes species described as insects' parasites in terrestrial environments (Park and Poulin, 2021), including also human-infecting species. Clade I, on the other hand, is commonly described as a freshwater clade, since it is mainly found in freshwater crustaceans and insects in various environments such as lakes, rivers, ponds, and so on (Andreadis *et al.*, 2012; Vossbrinck *et al.*, 2014). The other clades were less than 1% abundant (**Fig. 2A**). These findings partly overlap those from Dubuffet *et al.* (2021) and Chauvet *et al.* (2022a) in which clades IV2 and I were the most abundant. However, in these two studies, clade IV1 was not well represented (~ 2%), whereas it was the second most abundant clade in the present dataset, largely due to its overrepresentation in the Estivadoux sample (**Figs. 1, 3**). Among the sampled freshwater ecosystems which are mostly lakes, the Estivadoux is quite different, it is indeed a temporary marsh that can be completely dried during summer. The typical hosts of clade IV1 are mainly insects of terrestrial environments, it therefore seems consistent to retrieve this clade with a higher relative abundance in this sample.

The Arlequin group (clade IV2) was the most abundant (49% of the reads), this group being largely represented by the *Crispospora* genus (40% of the reads), as already found in previous studies (Dubuffet *et al.*, 2021; Chauvet *et al.*, 2022a; **Fig. 1**). It is noteworthy that the only one OTU shared by the nine freshwater samples was affiliated to the genus *Crispospora* and was actually the most abundant OTU in the whole dataset. In this genus, only one species is described (*Crispospora chironomi*) and known to parasitize chironomid larvae (*Chironomus plumosus*) in freshwater ecosystems (Tokarev *et al.*, 2010). Moreover, a recent study suggests that a new *Crispospora* species might be able to parasitize *Kellicottia* rotifers in lake environments (Chauvet *et al.*, 2022b).

Also, in agreement with previous studies, *Pseudoberwaldia* and *Conglomerata* (Aquatic outgroup; clade I) were the next most abundant genera (**Fig. 2A**). These two genera are, each, described by only one species, both of which are known to be parasites of cladocerans in freshwater (Vávra *et al.*, 2018, 2019).

Some groups and/or genera were recorded with a significant relative abundance only in two or three samples as the *Orthosomella*-like group (clade IV2), mainly found in samples from lakes Chambon, Saint-Gervais, and Aydat. This group is so far only described as parasite of insects and nematodes in terrestrial environments (Nishikori *et al.*, 2018). *Conglomerata* (Aquatic outgroup; clade I) was mainly found in lakes Chambon, Pavin and Chauvet, and *Pseudoberwaldia* (Aquatic outgroup; clade I) in lakes Bourdouze and Chambon (**Fig. 3**).

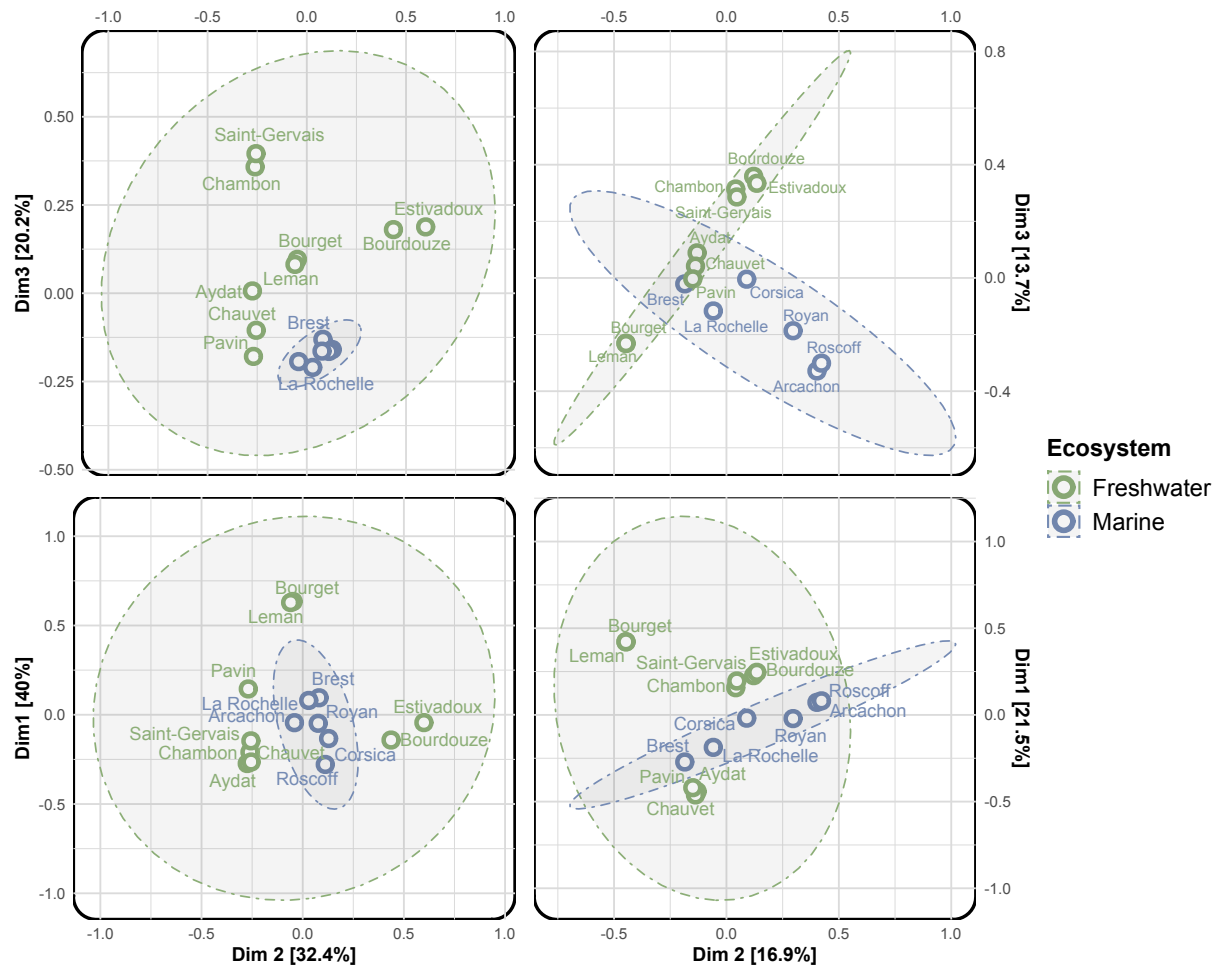


Figure 4: PCoA based on Bray–Curtis distances of the microsporidian OTUs richness (left) and relative abundances (right) across samples (first three PCoA axes shown).

Colors of the data points indicate the ecosystem (freshwater or marine). Ellipses in the plots denote 95% confidence intervals, points that are not overlapped by the two ellipses are considered significantly different.

These sites may be particularly rich in daphnia (cladoceran), which is the only described host of the latter two genera. Surprisingly, the *Vittaforma* genus (Arlequin group; clade IV2) has been mainly found in lake Bourdouze and not in the recreative lakes (Aydat, Chambon, Leman, Bourget), whereas it is currently described through a single species that is a human parasite (Silveira and Canning, 1995).

It therefore appeared that the microsporidian diversity varied across freshwater sampling sites even at a high taxonomic rank (*i.e.*, clade level; **Figs. 1, 3, 4**). In addition, a surprising result compared to previous studies was the significant proportion of clade IV2 OTUs not affiliated to any group (33%), especially in samples from lakes Bourget and Leman where these OTUs accounted for 83.2% and 75%, respectively (**Fig. 3**). These two lakes could represent an important reservoir of new microsporidian genera and species and thus potential new hosts. Altogether, these results illustrated the importance to increase the sampling effort in freshwater to unravel the full extent of the microsporidian diversity in these environments.

B. Extent of the marine microsporidian diversity

To date, the overall diversity of Microsporidia in marine ecosystems has received little attention. Indeed, to our best knowledge, only Williams *et al.* (2018) have initiated the study of the global microsporidian diversity in the marine environment, through coastal, estuary and rock pool samples; however, with a low number of OTUs. Here, we studied this diversity in six different marine sites close to the coast: three bays, one estuary, and two coastal ecosystems (**Fig. 1, Table S1**). The marine microsporidian diversity was largely dominated by clade IV2 and III, accounting for 49% and 40% of the total reads (**Fig. 2B**), the latter being a clade commonly described as “marine”, but which includes species parasite of crustaceans and fish, in both marine and freshwater environments (Vossbrinck *et al.*, 2014; Park and Poulin, 2021; clade V in these studies = clade III here). This clade also includes a protist parasite, the species *Hyperspora aquatica* hyperparasite of paramyxids (Stentiford *et al.*, 2017). Clades IV1, V, and I followed with 7%, 3% and 2% of the reads, respectively. Clade II was the least abundant with 0.1% of the reads and was identified only in Corsica (**Fig. 2B**). Due to the low number of OTUs (50) obtained by Williams *et al.* (2018), the comparison with our data is limited. However, we can mention that higher proportion of clade III was recorded in our samples but lower proportions of clades I, V, and especially II and no OTUs belonging to clade IX was found.

The latter clades (I, II, V, and IX) are known to be parasites of terrestrial and/or freshwater hosts including crustaceans, insects, nematodes, annelids, bryozoans, and humans (Vossbrinck *et al.*, 2014; Dubuffet *et al.*, 2021), it is thus not surprising to find them with a low relative abundance here. Clades IV2 and III on the other hand are the ones with the most members found in marine ecosystems (Park and Poulin, 2021).

The *Ameson*-like group (clade III) and Arlequin group (clade IV2) were the most abundant with 38% and 35% of the total reads, respectively (**Fig. 2B**). Three samples were mostly represented by the *Ameson*-like group (Roscoff, Corsica, and Arcachon), while two other samples were dominated by the Arlequin group (Brest and La Rochelle; **Fig. 3**).

Three genera were particularly abundant: *Crispospora* (Arlequin group), *Ameson* and *Nadelspora* (*Ameson*-like group), accounting for 31%, 23%, and 15% of the reads, respectively (**Fig. 2B**). *Crispospora* has never been identified in hosts living in marine environments, and neither chironomids nor rotifers are abundant in these systems, thus raising the hypothesis of one or many other hosts for the *Crispospora* genus. Moreover, the proportion of *Crispospora* identified here in the two ecosystems studied, both in terms of OTUs and abundance, suggests that a significant diversity remains to be discovered within this genus where only one species is currently described. This relative abundance may also reveal significant host diversity, as ~80% of microsporidian species are specialists (Murareanu *et al.*, 2021).

The *Ameson* genus includes different species all described as parasitizing arthropods in the marine environment. The hosts described are mainly crustaceans (crab, lobster, shrimp) and one species is described as an insect parasite (Murareanu *et al.*, 2021). *Nadelspora* is known to parasitize the crab *Metacarcinus magister* through the only described species: *Nadelspora canceri* (Olson *et al.*, 1994). Other genera identified in these samples were much less abundant such as *Sporanauta* (clade IV1), *Octosporea*, and *Pseudoberwaldia* (Aquatic outgroup; clade I), each representing about 0.5% of the reads (**Fig. 2B**). *Sporanauta* and *Pseudoberwaldia* are described by one species each as parasites of nematodes and cladocerans, respectively (Ardila-Garcia and Fast, 2012; Vávra *et al.*, 2019). *Octosporea* genus, on the other hand, includes many species that are parasitic on different hosts, all of which belong to the arthropods. *Octosporea* species are mainly described as insect parasites (Diptera and Hemiptera) and are also found in some cladocerans (*Daphnia*) and amphipods (*Gammarus*; Murareanu *et al.*, 2021).

In these marine samples, a higher proportion of the diversity was common between the samples compared to freshwater ecosystems as 402 OTUs were shared by the six marine samples; 84% of these OTUs were affiliated to *Crispospora*. Despite these shared OTUs, some samples could be quite different from each other considering some clades/groups/genera (**Fig. 3**).

The sample from Brest, for example, had only one OTU belonging to clade III. This was surprising since this was the second most abundant clade in our study and the clade in which the highest number of marine Microsporidia are found.

C. Microsporidian community composition in aquatic ecosystems in relation to their hosts

Although marine samples contained less reads than freshwater samples, all diversity index showed a more diverse community in marine sites (**Figs. S3, S4**). The marine sample from La Rochelle, for example, was the one with the highest number of OTUs (1 157) but the fewest reads (71 122); 63.4% of the OTUs were common to both ecosystem types (the majority belonging to the *Crispospora* genus, or being unaffiliated beyond clade IV2), 16.1% and 20.5% were specific to marine and freshwater ecosystems, respectively. Considering only OTUs richness, ecosystem types were indistinguishable, whereas differences were identified when considering the relative abundance of OTUs (**Fig. 4**). Most of the OTUs (82%) found only in freshwater samples belong to the clade IV2, whereas OTUs specific to marine samples were mostly found in clade III (72%; **Fig. S2B-D**). Only one OTU (the most abundant of the whole dataset), affiliated to *Crispospora* (clade IV2; Arlequin group), was shared by all the samples. Some OTUs were shared by different samples (**Fig. S5**), but all the 15 samples showed specific OTUs (on average 1% of the OTUs). The affiliation was better in marine samples, as 58.4% of OTUs were affiliated to a group and 38.9% to a genus in freshwater vs 62.0% and 53.6% in marine samples suggesting that a large part of the diversity is unknown. It therefore seems essential to continue the sampling effort across aquatic ecosystems to unravel the global microsporidian diversity (without the bias of targeting hosts).

At the clade level, the major differences were the absence of clade VIII in the marine samples and the very low relative abundance (0.003%) of clade III in freshwater whereas it represents 40% of the microsporidian abundance in marine samples (**Fig. 2**). Species belonging to clade VIII are described as parasites of insects (mainly Diptera), affecting larvae that are present in freshwater (Murareanu *et al.*, 2021). Clade III is mostly described as parasitizing fishes and crustaceans in marine and freshwater environments and is sometimes described as the “marinosporidia” clade (Vossbrinck *et al.*, 2014) even if Williams *et al.* (2018) showed the inconsistency in the classification based on the ecological niche that Microsporidia and/or their hosts inhabit.

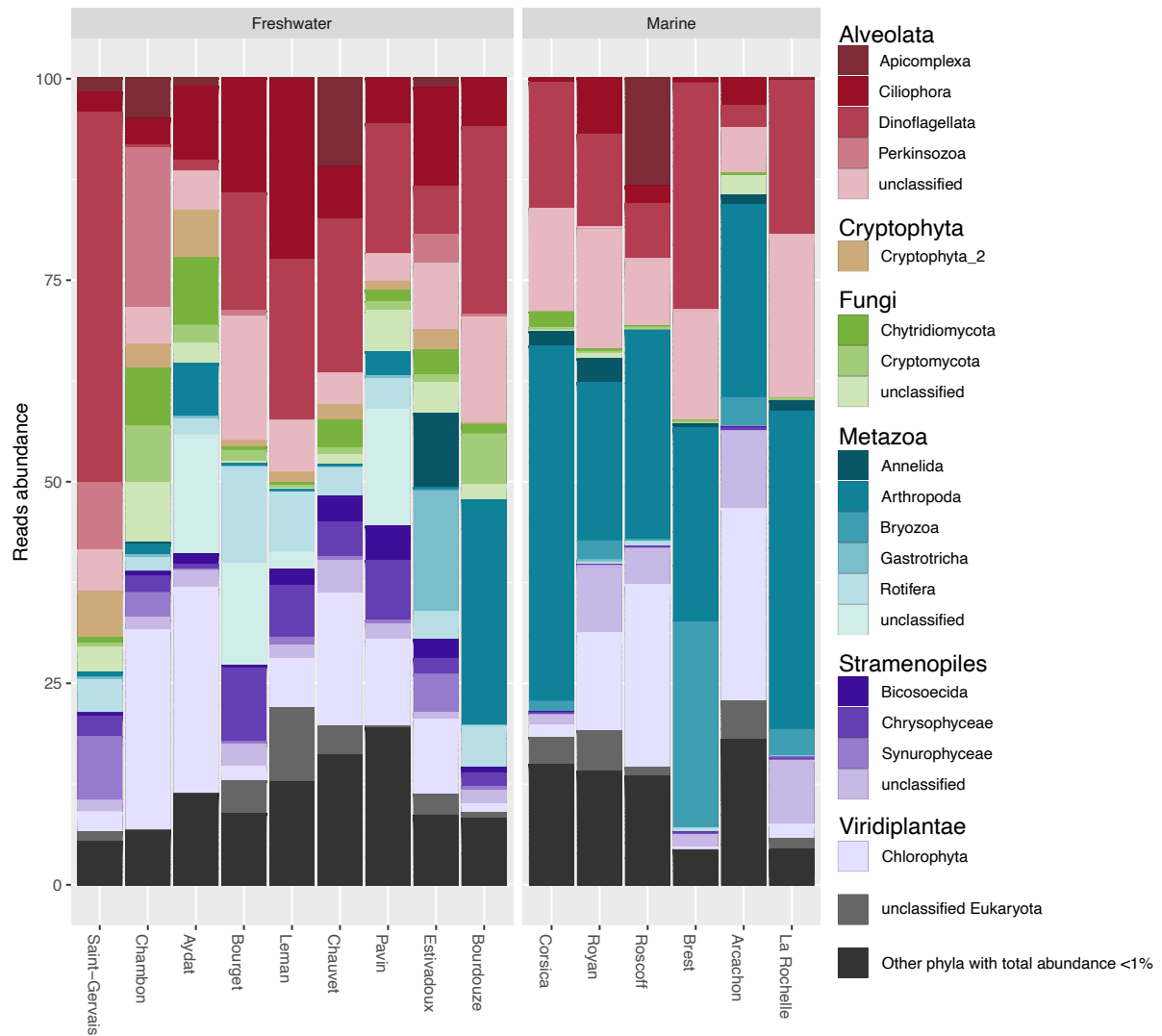


Figure 5: Relative abundances of the most abundant (>1% of the total reads) eukaryotic phyla across freshwater (left) and marine (right) samples.

The black zone at the bottom of each bar represents the cumulative proportion of other phyla with low abundances.

However, the primers used in this study target a wide microsporidian diversity, except in clade III (data not shown). The environmental diversity of Microsporidia affiliated to clade III is therefore likely underestimated in the two types of ecosystems.

Other differences at the genus level (apart from the genera belonging to clade III) were notable between the two types of environments such as the significantly higher relative abundance of *Conglomerata* (Aquatic outgroup; clade I) in freshwater (**Fig. S6**) and the absence of *Octosporea* (Aquatic outgroup; clade I) as well as *Mrazekia* and *Helmichia* (both in Arlequin group; clade IV2) in freshwater and marine samples, respectively (**Figs. 2, 3**). Some of these results are consistent with the type of host described for these genera, such as *Conglomerata* which is currently described as parasitizing only the freshwater *Daphnia pulex* (Vávra *et al.*, 2018). These differences nevertheless involve genera found at low relative abundance, not allowing us to clearly establish their real absence in the second ecosystem and to draw conclusions on these differences.

The environmental variables studied here (salinity and temperature), trophic states, ecosystem type, size fraction, and geographical location, do not seem to explain the distribution of the microsporidian communities (**Fig. 4, Table S1**). Although the number of parameters considered here are low, it has also never been shown that abiotic parameter can directly influence Microsporidia in their environment. The diversity observed was rather in agreement with the hosts described so far, at least in terms of habitat, even though some genera have been surprisingly identified in an environment that does not harbor their (often single) described hosts. Ultimately, it appears that microsporidian communities are probably driven by the dynamic of their hosts and not directly influenced by environmental variables. Most of the Microsporidia identified in this study are so far known to parasite large hosts such as fish, crustaceans, and insects. It, therefore, appears that the Microsporidia found in our study can infect smaller and so far, unknown hosts and/or that their free-living forms, the spores, actively circulate in the water column via ingestion/filtration by planktonic organisms of different sizes. Approximately, 80% of the microsporidian species described (with host data) are specialists (*i.e.*, described as infecting only a single host), the remaining species are generalists with some capable of infecting more than 10 hosts (Murareanu *et al.*, 2021), some having the ability to infect both vertebrate and invertebrate (Stentiford *et al.*, 2016). The hypothesis of a second host is already suggested for *Pseudoberwaldia* (Vávra *et al.*, 2019). Our results thus also suggest that a significant proportion of Microsporidia could infect more than one host, especially in the overlooked microeukaryotic compartment.

We studied the diversity of Microsporidia and planktonic eukaryotes (<150 µm) in parallel (**Fig. 5**). The eukaryotic communities identified were quite different between the two types of ecosystems studied (**Figs. 5, S7**). The freshwater diversity was mainly represented by dinoflagellates (16%), Chlorophyta (11%), ciliates (9%), and rotifers (5%), whereas marine diversity was mainly composed of arthropods (29%), dinoflagellates (14%), Chlorophyta (10%), and Bryozoa (6%; **Fig. 5**).

As for Microsporidia, some sites showed a specificity in their eukaryotic diversity that could explain the distribution of the microsporidian communities. Lakes Lemans and Bourget, for example, were particularly rich in rotifers compared to other lakes (**Fig. 5**), these rotifers belonging mainly to the Brachionidae and Notommatidae. Similarly, dinoflagellates were most abundant in lake Saint Gervais and Bryozoa in the Brest sampling point. To identify new host-parasite associations, OTUs representing more than 0.05% of reads in both datasets were used to reveal co-occurrences. Only positive associations were considered as it has already been showed that parasitism is retrieved as a positive link between the partners in co-occurrence network despite the detrimental impact of the parasite on the host (Weiss *et al.*, 2016). Moreover, potential associations were searched by ecosystem types since it is better to have homogeneous environments to identify biotic interactions (Faust, 2021).

In freshwater, 205 potential associations were identified between 32 microsporidian OTUs and 14 eukaryotic OTUs, mostly involving metazoans (platyhelminths, gastrotrichs, and arthropods) and alveolates (dinoflagellates, apicomplexans, and ciliates; **Fig. 6A**). These correlations largely involved OTUs unaffiliated beyond the clade (IV1, IV2, and I) or the group (Arlequin and Aquatic outgroup), as well as two genera: *Cystosporogenes* and *Crispospora* (both in the Arlequin group; clade IV2). In marine environments, 128 potential associations were identified between 69 microsporidian OTUs and 35 eukaryotic OTUs (**Fig. 6B**). As well as for freshwater, a lot of potential associations involved Microsporidia OTUs unaffiliated beyond the clade (V, IV2, IV1, II, and III) or the group (Arlequin, *Orthosomella*-like, Enterocytozoonidae, *Cucumispora*-like group, and Aquatic outgroup), and some genera: *Crispospora*, *Ameson*, *Obruspora*, *Nadelspora*, *Vittaforma*, *Sporanauta*, and *Octosporea*. Eukaryotic OTUs involved belong mostly to the alveolates (dinoflagellates, apicomplexans, and ciliates). Metazoans (cnidarians, arthropods, annelids, mollusks, and Bryozoa) and fungi (Ascomycota, Basidiomycota, chytrids) were also involved in these co-occurrences (**Fig. 6B**). Metazoans represent most of the described microsporidian host. Among them, arthropods, platyhelminths, gastrotrichs, cnidarians, annelids, mollusks, and Bryozoa (identified here in the co-occurrences) have already been described as hosts (Murareanu *et al.*, 2021).

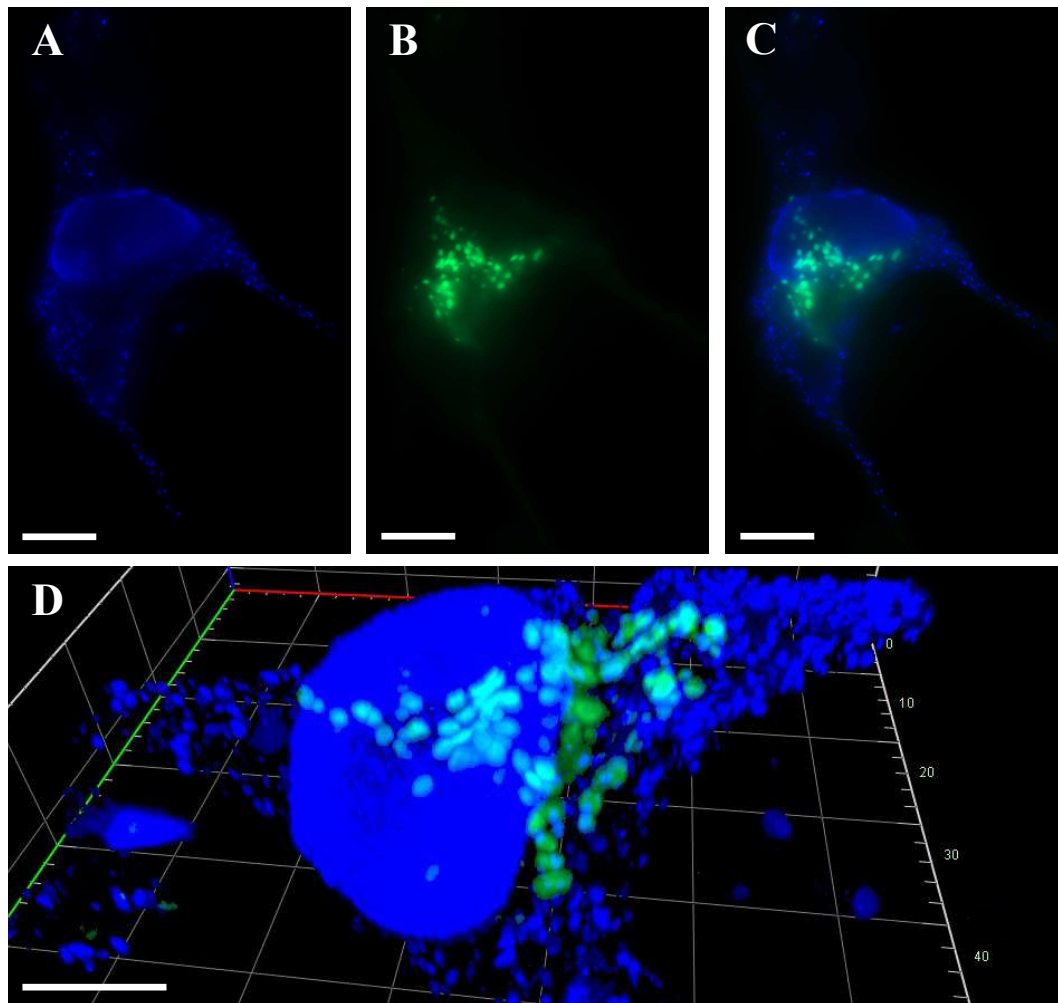


Figure 7: Microsporidia (targeted by TSA-FISH with the USP01 probe) infecting dinoflagellates in freshwater samples.

Under UV showing nucleus after DAPI staining (A), under blue light showing Microsporidia in green fluorescence (B), overlay (C, D). Magnification on the microsporidian cells inside the dinoflagellate with 3D reconstruction (D). Scale bars, 20 μm.

Similarly, the potential associations identified with alveolates were previously described for some, as Microsporidia have been shown to parasitize apicomplexans and ciliates although these descriptions are anecdotal and generally only morphological (*e.g.*, only two microsporidian SSU sequences parasites of ciliates; Fokin *et al.*, 2008; Yakovleva *et al.*, 2020). In contrast, there are currently no descriptions of Microsporidia parasitizing dinoflagellates. Fungi, also retrieved in these co-occurrences in potential association with Microsporidia, have also never been described as their hosts. In both ecosystems, some correlations were also found with phytoplanktonic organisms (**Fig. 6**), whereas to date, infection by Microsporidia in phytoplankton has never been recorded. Interestingly, some of the potential associations identified here were also retrieved in co-occurrence networks involving short-branch Microsporidia (Doliwa *et al.*, 2021).

Although co-occurrences network analyses should be interpreted with caution (Carr *et al.*, 2019; Blanchet *et al.*, 2020), they are increasingly used to identify potential associations within microorganisms. The method used here has its limitations, as other tools are more efficient and adapted to such datasets (Kurtz *et al.*, 2015) but often require more samples. However, this correlation-based method can show comparable efficiency to other more sophisticated methods (Hirano and Takemoto, 2019; Faust, 2021). However, associations between Microsporidia and eukaryotes suggested by co-occurrences network should be considered as working hypothesis and need to be confirmed by other approaches allowing the observation and the identification of both partners (*e.g.*, fluorescence microscopy or host/parasite genomic analysis).

Using TSA-FISH with a universal microsporidian probe (USP01; Chauvet *et al.*, 2022a) we confirmed the Microsporidia-*Kellicottia* (rotifer) association already observed in another lake (Chauvet *et al.*, 2022a,b), here in lake Chauvet. We also identified microsporidian infections in dinoflagellates (probably belonging the genus *Ceratium*) both in freshwater and marine ecosystems (**Figs. 7, S8**). Indeed, these infections were observed in lakes Lemans (prevalence, 49% of infected individuals), Bourget (53%) and Saint Gervais (20%), and in the La Rochelle (38%) marine sample. These observed Microsporidia-dinoflagellates associations in both ecosystems significantly expand the host range of Microsporidia and their impact on food web functioning. Indeed, dinoflagellates are relatively abundant in freshwater and marine environments and show a diversity of trophic modes (autotrophy, heterotrophy, or mixotrophy). Thanks to this diversity, dinoflagellates are among the major primary producers (with diatoms), they can as well be grazers, and have symbiotic or parasitic relationships, while others are known to be important toxin producers (Carty and Parrow, 2015).

V. Conclusion

In conclusion, this study highlighted for the first time a huge microsporidian diversity in marine samples, with an important part (46%) of the OTUs not affiliated to a genus suggesting an important reservoir of new microsporidian species and thus potential new hosts, especially among clade IV2. This diversity was also widely divergent from the diversity identified in freshwater. Interestingly, most of the identified genera are described as parasites of organisms much larger than the studied size fraction (<150 µm). It therefore appears that these Microsporidia infect smaller and, so far, unknown hosts such as dinoflagellates both in marine and freshwater ecosystems. These findings also greatly expand the host range of Microsporidia which were never described in association with phytoplanktonic hosts. Also, the proportion of the *Crispospora* genus raises questions about their hosts and impact especially in marine environments and should be the subject of further research.

Altogether, these results indicate that the microsporidian diversity and host spectrum could be therefore more important than previously described, and their role in the functioning of aquatic ecosystems and especially marine ecosystems is undoubtedly largely unknown.

Acknowledgements

The authors are very grateful towards the LIENSs (LIttoral ENvironnement et Sociétés), EPOC (Environnements et Paléoenvironnements Océaniques et Continentaux) and CARRTEL (Centre Alpin de Recherche sur les Réseaux Trophiques et Ecosystèmes Limniques) labs for their hospitality and help with the sampling. We especially want to thank Hélène Agogué, Philippe Pineau, Yves Brizard, Yolanda Del Amo, Laurent Letord, Anne Bohi, Isabelle Domaizon, Jean-Christophe Hustache and the French Oceanographic Fleet. We also gratefully acknowledge Raffaele Siano (Ifremer), Fabienne Rigaut-Jalabert (Station Biologique de Roscoff) and the captain and crew of the Neomysis research ship for the sampling in Brest and Roscoff. We also thank Simon Lafarge for facilitating the access of the private lake Chauvet.

We thank the SOMLIT program (Service d'Observation en Milieu Littoral, CNRS/INSU, a monitoring network of the Research Infrastructure ILICO), ROME project (Réseau d'Observatoires de Microbiologie Environnementale intégré) and OLA observatory (Observation au long terme et expérimentation - écosystèmes LAcustres, OLA-IS, AnaEE-France, INRAE of Thonon-les-Bains, CIPEL, Rimet *et al.*, 2020) for collecting samples and granting access to some metadata.

Finally, we would like to thank Jonathan Colombet, Marie Diogon, Maxime Fuster, Karine Massé, Jean-Christophe Pointud and Agnès Vellet for their help with sampling and filtration.

The work of MC and AM was supported by two PhD fellowships funded by the 'Ministère de l'Enseignement Supérieur de la Recherche et de l'Innovation'.

Competing interest

The authors declare they have no competing interests.

Data availability

All data and material are available in this publication and additional information online (Supporting Information). The data have been deposited with links to BioProject accession number PRJNA922474 in the NCBI BioProject database.

References

- Andreadis, T.G., Simakova, A.V., Vossbrinck, C.R., Shepard, J.J., and Yurchenko, Y.A. (2012) Ultrastructural characterization and comparative phylogenetic analysis of new Microsporidia from Siberian mosquitoes: Evidence for coevolution and host switching. *Journal of Invertebrate Pathology*, **109**, 59–75.
- Ardila-Garcia, A.M. and Fast, N.M. (2012) Microsporidian infection in a free-living marine nematode. *Eukaryotic Cell*, **11**, 1544–1551.
- Baker, M.D., Vossbrinck, C.R., Didier, E.S., Maddox, J.V., and Shadduck, J.A. (1995) Small subunit ribosomal DNA phylogeny of various Microsporidia with emphasis on AIDS related forms. *The Journal of Eukaryotic Microbiology*, **42**, 564–570.
- Bass, D., Czech, L., Williams, B.A.P., Berney, C., Dunthorn, M., Mahé, F., et al. (2018) Clarifying the relationships between Microsporidia and Cryptomycota. *The Journal of Eukaryotic Microbiology*, **65**, 773–782.
- Bass, D., Stentiford, G.D., Littlewood, D.T.J., and Hartikainen, H. (2015) Diverse applications of environmental DNA methods in parasitology. *Trends in Parasitology*, **31**, 499–513.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 289–300.
- Beule, L. and Karlovsky, P. (2020) Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. *PeerJ*, **8**, e9593.
- Blanchet, F.G., Cazelles, K., and Gravel, D. (2020) Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, **23**, 1050–1063.
- Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., et al. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 57–59.
- Bray, J.R. and Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.
- 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.
- Carr, A., Diener, C., Baliga, N.S., and Gibbons, S.M. (2019) Use and abuse of correlation analyses in microbial ecology. *The ISME Journal*, **13**, 2647–2655.
- Carty, S. and Parrow, M.W. (2015) Dinoflagellates. In *Freshwater Algae of North America*. Elsevier, pp. 773–807.
- Cavalier-Smith, T. (1993) Kingdom protozoa and its 18 phyla. *Microbiological Reviews*, **57**, 42.
- Chauvet, M., Debroas, D., Moné, A., Dubuffet, A., and Lepère, C. (2022a) Temporal variations of Microsporidia diversity and discovery of new host–parasite interactions in a lake ecosystem. *Environmental Microbiology*, **24**, 1672–1686.
- Chauvet, M., Monjot, A., Moné, A., and Lepère, C. (2022b) Single cell analysis reveals a new microsporidia-host association in a freshwater lake. *Microbial Ecology*, **85**, 1630–1633.

- Conway, J.R., Lex, A., and Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.
- 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.
- Doliwa, A., Dunthorn, M., Rassoshanska, E., Mahé, F., Bass, D., and Duarte Ritter, C. (2021) Identifying potential hosts of short-branch Microsporidia. *Microbial Ecology*, **82**, 549–553.
- Dubuffet, A., Chauvet, M., Moné, A., Debroas, D., and Lepère, C. (2021) A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environmental Microbiology*, **23**, 4344–4359.
- Ebert, D. (2005) Some parasites of *Daphnia*. In *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*. National Center for Biotechnology Information.
- Faust, K. (2021) Open challenges for microbial network construction and analysis. *The ISME Journal*, **15**, 3111–3118.
- Fokin, S.I., Di Giuseppe, G., Erra, F., and Dini, F. (2008) *Euplotespora binucleata* n. gen., n. sp. (Protozoa: Microsporidia), a parasite infecting the hypotrichous ciliate *Euplotes woodruffi*, with observations on microsporidian infections in Ciliophora. *The Journal of Eukaryotic Microbiology*, **55**, 214–228.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014) Circlize implements and enhances circular visualization in R. *Bioinformatics*, **30**, 2811–2812.
- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., and Troedsson, C. (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One*, **9**, e87624.
- Hirano, H. and Takemoto, K. (2019) Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics*, **20**, 329.
- Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., et al. (2014) Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS One*, **9**, e95567.
- Kagami, M., Miki, T., and Takimoto, G. (2014) Mycoloop: chytrids in aquatic food webs. *Frontiers in Microbiology*, **5**.
- Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*, **5**, 112.
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.
- Keeling, P.J. and Fast, N.M. (2002) Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annual Review of Microbiology*, **56**, 93–116.
- Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.

- Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., and Bonneau, R.A. (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biology*, **11**, e1004226.
- Lafferty, K.D., Allesina, S., Arim, M., Briggs, C.J., Leo, G.D., Dobson, A.P., et al. (2008) Parasites in food webs: the ultimate missing links. *Ecology Letters*, **11**, 533–546.
- 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. *Environmental Microbiology*, **15**, 1745–1758.
- Marcogliese, D.J. and Cone, D.K. (1997) Food webs: a plea for parasites. *Trends in Ecology & Evolution*, **12**, 320–325.
- McMurdie, P.J. and Holmes, S. (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, **8**, e61217.
- Murareanu, B.M., Sukhdeo, R., Qu, R., Jiang, J., and Reinke, A.W. (2021) Generation of a Microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of microsporidia. *mBio*, **12**, e01490-21.
- Nishikori, K., Setiamarga, D.H.E., Tanji, T., Kuroda, E., Shiraishi, H., and Ohashi-Kobayashi, A. (2018) A new microsporidium *Percutemincola moriokae* gen. nov., sp. nov. from *Oscheius tipulae* : A novel model of microsporidia–nematode associations. *Parasitology*, **145**, 1853–1864.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P., O’Hara, B., et al. (2015) Vegan: Community Ecology Package. *R Package Version 22-1*, **2**, 1–2.
- Olson, R.E., Tiekotter, K.L., and Reno, P.W. (1994) *Nadelspora canceri* N. G., N. Sp., an Unusual Microsporidian Parasite of the Dungeness Crab, Cancer Magister. *The Journal of Eukaryotic Microbiology*, **41**, 349–359.
- Paliy, O. and Shankar, V. (2016) Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology*, **25**, 1032–1057.
- Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.
- Park, E. and Poulin, R. (2021) Revisiting the phylogeny of microsporidia. *International Journal for Parasitology*, **51**, 855–864.
- R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Rimet, F., Anneville, O., Barbet, D., Chardon, C., Crépin, L., Domaizon, I., et al. (2020) The Observatory on LAkes (OLA) database: sixty years of environmental data accessible to the public. *Journal of Limnology*, **79**.
- Silveira, H. and Canning, E.U. (1995) *Vittaforma corneae* N. Comb. for the Human Microsporidium *Nosema corneum* Shadduck, Meccoli, Davis & Font, 1990, Based on its Ultrastructure in the Liver of Experimentally Infected Athymic Mice. *The Journal of Eukaryotic Microbiology*, **42**, 158–165.
- Singer, D., Sepey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., et al. (2021) Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environment International*, **146**, 106262.

- Stentiford, G.D., Becnel, J.J., Weiss, L.M., Keeling, P.J., Didier, E.S., Williams, B.A.P., et al. (2016) Microsporidia – Emergent Pathogens in the Global Food Chain. *Trends in Parasitology*, **32**, 336–348.
- Stentiford, G.D., Ramilo, A., Abollo, E., Kerr, R., Bateman, K.S., Feist, S.W., et al. (2017) *Hyperspora aquatica* n.g.n., n.sp. (Microsporidia), hyperparasitic in *Marteilia cochillia* (Paramyxida), is closely related to crustacean-infecting microsporidian taxa. *Parasitology*, **144**, 186–199.
- Taib, N., Mangot, J.-F., Domaizon, I., Bronner, G., and Debroas, D. (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. *PLoS One*, **8**, e58950.
- Tokarev, Y.S., Voronin, V.N., Seliverstova, E.V., Pavlova, O.A., and Issi, I.V. (2010) Life cycle, ultrastructure, and molecular phylogeny of *Crispospora chironomi* g.n. sp.n. (Microsporidia: Terresporidia), a parasite of *Chironomus plumosus* L. (Diptera: Chironomidae). *Parasitol Research*, **107**, 1381–1389.
- Trzebny, A., Slodkiewicz-Kowalska, A., Becnel, J.J., Sanscrainte, N., and Dabert, M. (2020) A new method of metabarcoding Microsporidia and their hosts reveals high levels of microsporidian infections in mosquitoes (Culicidae). *Molecular Ecology Resources*, **20**, 1486–1504.
- Vávra, J., Fiala, I., Krylová, P., Petrussek, A., and Hyliš, M. (2019) Establishment of a new microsporidian genus and species, *Pseudoberwaldia daphniae* (Microsporidia, Opisthosporidia), a common parasite of the *Daphnia longispina* complex in Europe. *Journal of Invertebrate Pathology*, **162**, 43–54.
- Vávra, J., Fiala, I., Krylová, P., Petrussek, A., and Hyliš, M. (2018) Molecular and structural assessment of Microsporidia infecting daphnids: The “obtusa-like” Microsporidia, a branch of the monophyletic Agglomeratidae clade, with the establishment of a new genus *Conglomerata*. *Journal of Invertebrate Pathology*, **159**, 95–104.
- Vossbrinck, C.R., Debrunner-Vossbrinck, B.A., and Weiss, L.M. (2014) Phylogeny of the Microsporidia. In *Microsporidia*. Weiss, L.M. and Becnel, J.J. (eds). Chichester, UK: John Wiley & Sons, Inc., pp. 203–220.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016) Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal*, **10**, 1669–1681.
- Williams, B.A.P., Hamilton, K.M., Jones, M.D., and Bass, D. (2018) Group-specific environmental sequencing reveals high levels of ecological heterogeneity across the microsporidian radiation. *Environmental Microbiology Reports*, **10**, 328–336.
- Yakovleva, Y., Nassonova, E., Lebedeva, N., Lanzoni, O., Petroni, G., Potekhin, A., and Sabaneyeva, E. (2020) The first case of microsporidiosis in *Paramecium*. *Parasitology*, **147**, 957–971.
- Zhu, X., Wittner, M., Tanowitz, H.B., Kotler, D., Cali, A., and Weiss, L.M. (1993) Small subunit rRNA sequence of *Enterocytozoon bieneusi* and its potential diagnostic role with use of the polymerase chain reaction. *Journal of Infectious Diseases*, **168**, 1570–1575.

VI. Supplementary Materials

Site	Coordinates	Ecosystem	Date	Temperature (°C)	Salinity (PSU)	Volume (litres)	Depth	Size fraction	Filters	TSA-FISH filters
Saint-gervais	45°02' 17" N, 2° 48' 54" E	Freshwater - Hypereutrophic	30/06/2021	19.8	NA	10	Surface	0.8-150 µm	4	X
Chambon	45°34' 10" N, 2° 55' 26" E	Freshwater - Eutrophic	01/07/2021	15.7	0.03	15	Surface	0.8-150 µm	4	X
Aydat	45°39' 50" N, 2° 59' 04" E	Freshwater - Eutrophic	01/07/2021	17.1	0.06	15	IEZ 5.75 m	0.8-150 µm	4	X
Bouquet	45°44' 48" N, 5° 51' 34" E	Freshwater - Oligomesotrophic	06/07/2021	21.7	NA	9.5	IEZ 10 m	0.8-150 µm	4	X
Leman	46°27' 09" N, 6° 35' 19" E	Freshwater - Mesotrophic	07/07/2021	20.0	NA	9.5	IEZ 10 m	0.8-150 µm	4	X
Chauvet	45°27' 36" N, 2° 49' 50" E	Freshwater - Oligotrophic	22/07/2021	17.7	0.01	17	IEZ 9 m	0.8-150 µm	4	X
Pavin	45°29' 45" N, 2° 53' 16" E	Freshwater - Oligomesotrophic	22/07/2021	17.6	0.02	14.5	IEZ 11 m	0.8-150 µm	4	X
Estivadoux	45°29' 31" N, 2° 54' 05" E	Freshwater - Temporary marsh	22/07/2021	NA	NA	2.5	Surface	0.8-150 µm	4	X
Bourdouze	45°28' 11" N, 2° 55' 31" E	Freshwater - Oligotrophic	22/07/2021	21.3	0.02	16	IEZ 2 m	0.8-150 µm	4	X
Corsica	41°28' 13" N, 9° 04' 03" E	Marine - Bay	31/08/2021	24.7	NA	22	Surface	0.8-150 µm	2	
Royan (Vaux-Sur-Mer)	45°37' 42" N, 1° 03' 55" W	Marine - Estuary	15/08/2021	18.5	NA	5	Surface	0.8-150 µm	3	
Roscoff	48°46' 40" N, 3° 56' 15" W	Marine - Coastal	18/06/2021	13.9	35.1	9	Surface	>3 µm	2	
Brest	48°18' 33" N, 4° 26' 54" W	Marine - Bay	22/06/2021	16.5	36.7	20	Surface	>3 µm	4	
Arcachon	44°40' 08" N, 1° 09' 52" W	Marine - Bay	23/09/2021	19.9	33.8	10	Surface	0.8-150 µm	4	X
La Rochelle	46°05' 03" N, 1° 18' 29" W	Marine - Coastal	27/09/2021	16.9	34.8	11.5	Surface	0.8-150 µm	4	X

Table S1: Information and metadata from the sampling sites.

NA = Not Available.

Figure S1: Microsporidian phylogenetic tree based on full length SSU rDNA of 140 microsporidian species (1 319 positions) (accessible online).

Numbers next to branches indicate the ML bootstrap support (200 replicates), bootstrap supports <70 are not shown. We included species from the main microsporidian clades, which were also identified in this study (I, II, III, IV1, IV2, V and VIII). Other microsporidian clades (VI, VII and IX) considered as ‘orphan’ lineages by Bojko *et al.* (2022) were not included.

The names and circumscriptions of the clades defined by Dubuffet *et al.* (2021) and used in this study are shown in red capitalized text and broken line red boxes. Below, in black and capital letters, are indicated the clade names proposed by Bojko *et al.* (2022), which is the most recent microsporidian nomenclature proposed. Circumscriptions are the same between the two nomenclatures for clades I, III, IV1, IV2 and VIII. Clades II and V are grouped together according to Bojko *et al.* (2022) in a clade named Neopereziida. Group (proxy for family) names defined in Dubuffet *et al.* (2021) are indicated in red italic text. Family names proposed by Bojko *et al.* (2022) are indicated in black italic text.

Blue-shaded blocks indicate phylogenetic regions where the microsporidian environmental sequences found in this study branch, without showing any of the environmental sequences themselves.

Figure S2: Maximum Likelihood phylogenies of the 100 most abundant OTUs affiliated to (A) Clade I (only 72 OTUs), (B) Clade III, (C) Clade IV1, (D) Clade IV2, (E) Clade V (only 12 OTUs) (accessible online).

Numbers next to branches indicate the ML bootstrap support (200 replicates), bootstrap supports <70 are not shown. Outgroup species are representative species from other clades. OTUs identified in this study are presented in bold. OTUs specific to freshwater or marine samples are given in green and blue text, respectively. When several OTUs are grouped together, numbers in brackets display the amount of OTUs specific to each ecosystem. Annotations on the right of each phylogeny display the group names (in italics) and clade names (in capital letters) defined by Dubuffet *et al.* (2021) and used in this study (in red), as well as the clade and family names proposed by Bojko *et al.* (2022) (in black).

Figure S3: Rarefaction curves for the microsporidian samples (accessible online).

Figure S4: Boxplot showing α -diversity indices (Richness observed, Chao1, Shannon and Simpson) of Microsporidia community grouped according to the ecosystem (accessible online).

Wilcoxon test: ***, p -value <0.001.

Figure S5: UpSet plot showing shared and unique OTUs across samples (green: freshwater, blue: marine) (accessible online).

A minimum of 10 shared OTUs is shown to lighten the figure.

Figure S6: Differential abundance of the 9 most abundant microsporidian genera between the two ecosystems studied (accessible online).

Wilcoxon test: *, p -value <0.05.

Figure S7: Principal Coordinates Analysis (PCoA) based on Bray-Curtis distances of the eukaryotic OTUs relative abundances across samples (accessible online).

Colors of the data points indicate the ecosystem (freshwater or marine). Ellipses in the plots denote 95% confidence intervals, points that are not overlapped by the two ellipses are considered significantly different.

Figure S8: Microsporidia (targeted by TSA-FISH with the USP01 probe) infecting dinoflagellates in marine samples (accessible online).

Under UV showing nucleus after DAPI staining (a), under blue light showing Microsporidia in green fluorescence (b), overlay (c). Scale bars, 20 μ m.

References

- Dubuffet, A., Chauvet, M., Moné, A., Debroas, D. and Lepère C. (2021) A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environmental Microbiology*, **23**, 4344–4359.
- Bojko, J., Reinke, A. W., Stentiford, G. D., Williams, B., Rogers, M. S. J. and Bass, D. (2022) Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends in Parasitology*, **38**, 642–659.

RÉFÉRENCES BIBLIOGRAPHIQUES
DES CHAPITRES I ET V



A

- Acinas, S.G., Sánchez, P., Salazar, G., Cornejo-Castillo, F.M., Sebastián, M., Logares, R., et al. (2021) Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Communications Biology*, **4**, 1–15.
- Adl, S.M., Simpson, A.G.B., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., et al. (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *The Journal of Eukaryotic Microbiology*, **52**, 399–451.
- Agha, R., Saebelfeld, M., Manthey, C., Rohrlack, T., and Wolinska, J. (2016) Chytrid parasitism facilitates trophic transfer between bloom-forming cyanobacteria and zooplankton (*Daphnia*). *Scientific Report*, **6**, 35039.
- Aguirre De Cárcer, D., Denman, S.E., McSweeney, C., and Morrison, M. (2011) Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. *Applied Environmental Microbiology*, **77**, 8795–8798.
- Alvarez-Pellitero, P. (2008) Fish immunity and parasite infections: from innate immunity to immunoprophylactic prospects. *Veterinary Immunology and Immunopathology*, **126**, 171–198.
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., and Gouil, Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, **21**, 30.
- Anderson, T.R. (2005) Plankton functional type modelling: running before we can walk? *Journal of Plankton Research*, **27**, 1073–1081.
- Arias-González, J. and Morand, S. (2006) Trophic functioning with parasites: a new insight for ecosystem analysis. *Marine Ecology Progress Series*, **320**, 43–53.
- Armstrong, R.A., Drange, H., Parslow, J.S., Powell, T.M., Taylor, A.H., and Totterdell, I.J. (1993) Trophic resolution. In *Towards a Model of Ocean Biogeochemical Processes*. Evans, G.T. and Fasham, M.J.R. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 71–92.
- Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., and Thingstad, F. (1983) The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, **10**, 257–263.
- Azevedo, C. and Hine, P.M. (2016) Haplosporidia. In *Handbook of the Protists*. Archibald, J.M., Simpson, A.G.B., Slamovits, C.H., Margulis, L., Melkonian, M., Chapman, D.J., and Corliss, J.O. (eds). Cham: Springer International Publishing, pp. 1–29.

B

- Baldauf, S.L. (2003) The deep roots of eukaryotes. *Science*, **300**, 1703–1706.
- Baschien, C. (2003) Development and evaluation of rRNA targeted in situ probes and phylogenetic relationships of freshwater Fungi. Thèse : Berlin, DE. Technische Universität Berlin.

- Bass, D., Stentiford, G.D., Littlewood, D.T.J., and Hartikainen, H. (2015) Diverse applications of environmental DNA methods in parasitology. *Trends in Parasitology*, **31**, 499–513.
- Bernard, G., Pathmanathan, J.S., Lannes, R., Lopez, P., and Baptiste, E. (2018) Microbial dark matter investigations: how microbial studies transform biological knowledge and empirically sketch a logic of scientific discovery. *Genome Biology and Evolution*, **10**, 707–715.
- Bittner, L. (2018) On the use of networks to link molecules to ecosystems: Investigating the uncultured lineages and/or the unknown functions from protists for a better understanding of ecosystems. Habilitation à Diriger des Recherches : Paris, FR. Sorbonne Université.
- Blandenier, Q. (2020) Exploration and characterization of “Amoebozoa” diversity and investigation of their diversity patterns at regional and global scales. Thèse : Neuchâtel, CH. University of Neuchâtel.
- Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., and Wincker, P. (2015) Tara Oceans studies plankton at planetary scale. *Science*, **348**, 873–873.
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., and Bairoch, A. (2007) UniProtKB/Swiss-Prot. *Methods in Molecular Biology*, **406**, 89–112.
- Brauer, A. and Bengtsson, M.M. (2022) DNA extraction bias is more pronounced for microbial eukaryotes than for prokaryotes. *MicrobiologyOpen*, **11**, e1323.
- Bronstein, J.L. (1994) Our current understanding of mutualism. *The Quarterly Review of Biology*, **69**, 31–51.
- Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., et al. (2018) Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. *Genome Biology and Evolution*, **10**, 427–433.
- Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjæveland, Å., Nikolaev, S.I., Jakobsen, K.S., and Pawlowski, J. (2007) Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One*, **2**, e790.
- Burki, F. (2014) The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology*, **6**, a016147.
- Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., et al. (2016) Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B*, **283**, 20152802.
- Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020) The new tree of eukaryotes. *Trends in Ecology & Evolution*, **35**, 43–55.

C

- 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.

- Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, **11**, 2639–2643.
- Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., et al. (2019) High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, **47**, e103.
- del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014) The others: our biased perspective of eukaryotic genomes. *Trends in Ecology & Evolution*, **29**, 252–259.
- del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., Massana, R., et al. (2018) EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biology*, **16**, e2005849.
- del Campo, J., Heger, T.J., Rodríguez-Martínez, R., Worden, A.Z., Richards, T.A., Massana, R., and Keeling, P.J. (2019) Assessing the diversity and distribution of apicomplexans in host and free-living environments using high-throughput amplicon data and a phylogenetically informed reference framework. *Frontiers in Microbiology*, **10**, 2373.
- del Campo, J., Bass, D., and Keeling, P.J. (2020) The eukaryome: diversity and role of microeukaryotic organisms associated with animal hosts. *Functional Ecology*, **34**, 2045–2054.
- Caputi, L., Carradec, Q., Eveillard, D., Kirilovsky, A., Pelletier, E., Pierella Karlusich, J.J., et al. (2019) Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochemical Cycles*, **33**, 391–419.
- Caron, D.A., Countway, P.D., Savai, P., Gast, R.J., Schnetzer, A., Moorthi, S.D., et al. (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Applied and Environmental Microbiology*, **75**, 5797–5808.
- Caron, D.A., Countway, P.D., Jones, A.C., Kim, D.Y., and Schnetzer, A. (2012) Marine protistan diversity. *Annual Review of Marine Science*, **4**, 467–493.
- Caron, D.A. (2016) Mixotrophy stirs up our understanding of marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 2806–2808.
- Caron, D.A., Alexander, H., Allen, A.E., Archibald, J.M., Armbrust, E.V., Bachy, C., et al. (2017) Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Review Microbiology*, **15**, 6–20.
- 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*, **9**, 373.
- Chambouvet, A., Monier, A., Maguire, F., Itoiz, S., del Campo, J., Elies, P., et al. (2019) Intracellular infection of diverse diatoms by an evolutionary distinct relative of the Fungi. *Current Biology*, **29**, 4093–4101.
- Chan, C.X. and Ragan, M.A. (2013) Next-generation phylogenomics. *Biology Direct*, **8**, 3.
- 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.

- Cheng, S., Karkar, S., Baptiste, E., Yee, N., Falkowski, P., and Bhattacharya, D. (2014) Sequence similarity network reveals the imprints of major diversification events in the evolution of microbial life. *Frontiers in Ecology and Evolution*, **2**, 72.
- Cohen, J.E. and Briand, F. (1984) Trophic links of community food webs. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 4105–4109.
- Combes, C. (2001) L'art d'être parasite : les associations du vivant. Edition revue et augmentée 2010, Paris, FR, Flammarion.
- Combes, C. (2018) Parasitisme : Écologie et évolution des interactions durables, Paris, FR, Dunod.
- Corliss, J.O. (1999) Annotated excerpts from Clifford Dobell's 88-year-old insightful classic paper, "The principles of protistology". *Protist*, **150**, 85–98.
- Costa, C., Dias, P.J., Sãl-Correia, I., and Teixeira, M.C. (2014) MFS multidrug transporters in pathogenic Fungi: do they have real clinical impact? *Frontiers in Physiology*, **5**, 197.
- Cunningham, A.A. and Daszak, P. (1998) Extinction of a species of land snail due to infection with a microsporidian parasite. *Conservation Biology*, **12**, 1139–1141.

D

- Dajoz, R. (2019) Précis d'écologie, 8ème Edition. Domont, FR, Dunod.
- Dawkins, R. (1982) The extended phenotype: the long reach of the gene, Oxford, UK, Oxford University Press.
- 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.
- Decaestecker, E., Lefever, C., De Meester, L., and Ebert, D. (2004) Haunted by the past: evidence for dormant stage banks of microparasites and epibionts of *Daphnia*. *Limnology and Oceanography*, **49**, 1355–1364.
- Delmont, T.O., Eren, A.M., Vineis, J.H., and Post, A.F. (2015) Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica. *Frontiers in Microbiology*, **6**, 1090.
- Delmont, T.O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S.T., Rappé, M.S., et al. (2018) Nitrogen-fixing populations of *Planctomycetes* and *Proteobacteria* are abundant in surface ocean metagenomes. *Nature Microbiology*, **3**, 804–813.
- Delmont, T.O., Gaia, M., Hinsinger, D.D., Frémont, P., Vanni, C., Fernandez-Guerra, A., et al. (2022) Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*, **2**, 100123.
- Dobson, A., Lafferty, K.D., Kuris, A.M., Hechinger, R.F., and Jetz, W. (2008) Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11482–11489.
- Douzery, E.J.P., Delsuc, F., and Philippe, H. (2006) Les datations moléculaires à l'heure de la génomique. *Medecine/Sciences*, **22**, 374–380.

- Dubuffet, A., Chauvet, M., Moné, A., Debroas, D., and Lepère, C. (2021) A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environmental Microbiology*, **23**, 4344–4359.
- Duncan, A., Barry, K., Daum, C., Eloë-Fadrosch, E., Roux, S., Schmidt, K., et al. (2022) Metagenome-assembled genomes of phytoplankton microbiomes from the Arctic and Atlantic Oceans. *Microbiome*, **10**, 67.
- Dunne, J.A., Lafferty, K.D., Dobson, A.P., Hechinger, R.F., Kuris, A.M., Martinez, N.D., et al. (2013) Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biology*, **11**, e1001579.

E

- Ebert, D. (1995) The ecological interactions between a microsporidian parasite and its host *Daphnia magna*. *The Journal of Animal Ecology*, **64**, 361.
- Ebert, D., Lipsitch, M., and Mangin, K.L. (2000) The effect of parasites on host population density and extinction: experimental epidemiology with *Daphnia* and six microparasites. *The American Naturalist*, **156**, 459–477.
- Edwards, K.F. (2019) Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 6211–6220.
- Elton, C.S. (1927) *Animal ecology*, New York, NY: The Macmillan Company.
- Escalas, A., Hale, L., Voordeckers, J.W., Yang, Y., Firestone, M.K., Alvarez-Cohen, L., and Zhou, J. (2019) Microbial functional diversity: from concepts to applications. *Ecology & Evolution*, **9**, 12000–12016.

F

- Faure, E., Not, F., Benoiston, A.-S., Labadie, K., Bittner, L., and Ayata, S.-D. (2019) Mixotrophic protists display contrasted biogeographies in the global ocean. *The ISME Journal*, **13**, 1072–1083.
- Faure, E. (2020) Contribution of meta-omics data in the detection and quantification of functional traits in planktonic ecosystems. Thèse : Paris, FR. Sorbonne Université.
- Faure, E., Ayata, S.-D., and Bittner, L. (2021) Towards omics-based predictions of planktonic functional composition from environmental data. *Nature Communication*, **12**, 4361.
- Faure-Fremiet, P.E. (1961) Cils vibratiles et flagelles. *Biological Reviews*, **36**, 464–533.
- Felczykowska, A., Krajewska, A., Zielińska, S., and Łoś, J.M. (2015) Sampling, metadata and DNA extraction - important steps in metagenomic studies. *Acta Biochimica Polonica*, **62**, 151–160.

- Fernández-Valero, A.D., Reñé, A., Timoneda, N., Sampedro, N., and Garcés, E. (2022) Dinoflagellate hosts determine the community structure of marine Chytridiomycota: demonstration of their prominent interactions. *Environmental Microbiology*, **24**, 5951–5965.
- Ferraro, M., Brigida, Giordani, P., and Serafini, A. (2019) fclust: an R package for fuzzy clustering. *The R Journal*, **11**, 198.
- Foissner, W., Blatterer, H., and Foissner, I. (1988) The hemimastigophora (*Hemimastix amphikineta* nov. gen., nov. spec.), a new protistan phylum from gondwanian soils. *European Journal of Protistology*, **23**, 361–383.
- Foley, K.P., Leonard, M.W., and Engel, J.D. (1993) Quantitation of RNA using the polymerase chain reaction. *Trends in Genetics*, **9**, 380–385.
- Forster, D., Bittner, L., Karkar, S., Dunthorn, M., Romac, S., Audic, S., et al. (2015) Testing ecological theories with sequence similarity networks: marine ciliates exhibit similar geographic dispersal patterns as multicellular organisms. *BMC Biology*, **13**, 16.
- Forterre, P. (2022) Archaea: a goldmine for molecular biologists and evolutionists. In *Archaea. Methods in Molecular Biology*. Ferreira-Cerca, S. (ed). New York, NY, Springer US, pp. 1–21.
- Frenken, T., Alacid, E., Berger, S.A., Bourne, E.C., Gerphagnon, M., Grossart, H.-P., et al. (2017) Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. *Environmental Microbiology*, **19**, 3802–3822.

G

- Garcia, R., Jara, F., Steciow, M., and Reissig, M. (2018) Oomycete parasites in freshwater copepods of Patagonia: effects on survival and recruitment. *Diseases of Aquatic Organisms*, **129**, 123–134.
- Garvetto, A., Nézan, E., Badis, Y., Bilien, G., Arce, P., Bresnan, E., et al. (2018) Novel widespread marine oomycetes parasitising diatoms, including the toxic genus *Pseudonitzschia*: genetic, morphological, and ecological characterisation. *Frontiers in Microbiology*, **9**, 2918.
- Geisen, S., Vaulot, D., Mahé, F., Lara, E., de Vargas, C., and Bass, D. (2019) A user guide to environmental protistology: primers, metabarcoding, sequencing, and analyses. *bioRxiv* (preprint).
- 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.
- Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome datasets are compositional: and this is not optional. *Frontiers in Microbiology*, **8**, 2224.
- Gomaa, F., Kosakyan, A., Heger, T.J., Corsaro, D., Mitchell, E.A.D., and Lara, E. (2014) One alga to rule them all: unrelated mixotrophic testate Amoebae (Amoebozoa, Rhizaria and Stramenopiles) share the same symbiont (Trebouxiophyceae). *Protist*, **165**, 161–176.

- 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.
- Gorbunov, A.K. and Kosova, A.A. (2001) Parasites in rotifers from the Volga delta. *Hydrobiologia*, **446**, 51–55.
- 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.
- Grossmann, L., Beisser, D., Bock, C., Chatzinotas, A., Jensen, M., Preisfeld, A., et al. (2016) Trade-off between taxon diversity and functional diversity in European lake ecosystems. *Molecular Ecology*, **25**, 5876–5888.
- 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.
- Gutell, R.R., Larsen, N., and Woese, C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews*, **58**, 10–26.

H

- Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., and Troedsson, C. (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One*, **9**, e87624.
- Haeckel, E. (1866) *Generelle Morphologie der Organismen*, Reimer, G. (ed.). Berlin, Germany.
- Haeckel, E. (1899) *Kunstformen der Natur: die einhundert Farbtafeln*, Vienna, Austria.
- Haraldsson, M., Gerphagnon, M., Bazin, P., Colombet, J., Tecchio, S., Sime-Ngando, T., and Niquil, N. (2018) Microbial parasites make cyanobacteria blooms less of a trophic dead end than commonly assumed. *The ISME Journal*, **12**, 1008–1020.
- Hardy, A.C. (1924) *The herring in relation to its animate environment: the food and feeding habits of the herring with special reference to the east coast of England*, H.M. Stationery Office.
- 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. *Environmental Microbiology Reports*, **5**, 835–840.
- Hauptfeld, E., Pelkmans, J., Huisman, T.T., Anocic, A., Snoek, B.L., von Meijenfildt, F.A.B., et al. (2022) A metagenomic portrait of the microbial community responsible for two decades of bioremediation of poly-contaminated groundwater. *Water Research*, **221**, 118767.
- Havens, K. (1992) Scale and structure in natural food webs. *Science*, **257**, 1107–1109.
- Heal, O.W. and MacLean, S.F. (1975) Comparative productivity in ecosystems-secondary productivity. In *Unifying Concepts in Ecology*. Van Dobben, W.H. and Lowe-McConnell, R.H. (eds). Dordrecht: Springer Netherlands, pp. 89–108.

- Heeger, F., Bourne, E.C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., et al. (2018) Long-read DNA metabarcoding of ribosomal RNA in the analysis of Fungi from aquatic environments. *Molecular Ecology Resources*, **18**, 1500–1514.
- Hoch, H.C., Galvani, C.D., Szarowski, D.H., and Turner, J.N. (2005) Two new fluorescent dyes applicable for visualization of fungal cell walls. *Mycologia*, **97**, 580–588.
- 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.
- Huang, Q., Chen, J., Lv, Q., Long, M., Pan, G., and Zhou, Z. (2023) Germination of microsporidian spores: the known and unknown. *Journal of Fungi*, **9**, 774.
- Hudson, P.J., Dobson, A.P., Cattadori, I.M., Newborn, D., Haydon, D.T., Shaw, D.J., et al. (2002) Trophic interactions and population growth rates: describing patterns and identifying mechanisms. *Philosophical Transactions of the Royal Society B*, **357**, 1259–1271.
- Hudson, P.J., Dobson, A.P., and Lafferty, K.D. (2006) Is a healthy ecosystem one that is rich in parasites? *Trends in Ecology & Evolution*, **21**, 381–385.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., et al. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, **47**, D309–D314.
- Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., et al. (2014) Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS One*, **9**, e95567.
- Hunter, P. (2009) Extended phenotype redux: how far can the reach of genes extend in manipulating the environment of an organism? *EMBO Reports*, **10**, 212–215.

I

- Ibarbalz, F.M., Henry, N., Brandão, M.C., Martini, S., Busseni, G., Byrne, H., et al. (2019) Global trends in marine plankton diversity across kingdoms of life. *Cell*, **179**, 1084–1097.
- 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.
- Ironside, J.E. and Alexander, J. (2015) Microsporidian parasites feminise hosts without paramyxean co-infection: support for convergent evolution of parasitic feminisation. *International Journal for Parasitology*, **45**, 427–433.
- Ishaq, S.L. and Wright, A.-D.G. (2014) Design and validation of four new primers for next-generation sequencing to target the 18S rRNA genes of gastrointestinal ciliate protozoa. *Applied Environmental Microbiology*, **80**, 5515–5521.
- Islam, S., Sabharwal, T., Wu, S., Bullock, T.J., and Mehdy, M.C. (2020) Early dynamics of photosynthetic Lhcf2 and Lhcf15 transcription and mRNA stabilities in response to herbivory-related decadienal in *Phaeodactylum tricornutum*. *Scientific Reports*, **10**, 2029.

- Itoiz, S., Perennou, M., Mouronvalle, C., Derelle, E., Le Goïc, N., Bidault, A., et al. (2021) Development of duplex TaqMan-based real-time PCR assay for the simultaneous detection of *Perkinsus olseni* and *P. chesapeaki* in host Manila clam tissue samples. *Journal of Invertebrate Pathology*, **184**, 107603.
- Itoiz, S., Metz, S., Derelle, E., Reñé, A., Garcés, E., Bass, D., et al. (2022) Emerging parasitic protists: the case of Perkinsea. *Frontiers in Microbiology*, **12**, 735815.

J

- Jakus, M.A. (1945) The structure and properties of the trichocysts of *Paramecium*. *Journal of Experimental Zoology*, **100**, 457–485.
- Jakus, M.A. and Hall, C.E. (1946) Electron microscope observations of the trichocysts and cilia of *paramecium*. *The Biological Bulletin*, **91**, 141–144.
- James, T.Y. and Berbee, M.L. (2012) No jacket required - new fungal lineage defies dress code: recently described zoosporic fungi lack a cell wall during trophic phase. *Bioessays*, **34**, 94–102.
- Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., et al. (2019) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Molecular Ecology Resources*, **20**, 429–443.
- Jamy, M., Biwer, C., Vaultot, D., Obiol, A., Jing, H., Peura, S., et al. (2022) Global patterns and rates of habitat transitions across the eukaryotic tree of life. *Nature Ecology and Evolution*, **6**, 1458–1470.
- Jansen, M., Stoks, R., Decaestecker, E., Coors, A., Van De Meutter, F., and De Meester, L. (2010) Local exposure shapes spatial patterns in infectivity and community structure of *Daphnia* parasites. *Journal of Animal Ecology*, **79**: 1023–1033.
- Jobard, M., Rasconi, S., and Sime-Ngando, T. (2010) Diversity and functions of microscopic fungi: a missing component in pelagic food webs. *Aquatic Sciences*, **72**, 255–268.
- Jobard, M., Wawrzyniak, I., Bronner, G., Marie, D., Vellet, A., Sime-Ngando, T., et al. (2020) Freshwater Perkinsea: diversity, ecology and genomic information. *Journal of Plankton Research*, **42**, 3–17.
- Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., et al. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communication*, **10**, 5029.
- 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.
- Jones, M.D.M., Forn, I., Gadelha, C., Egan, M.J., Bass, D., Massana, R., and Richards, T.A. (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*, **474**, 200–203.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

K

- Kagami, M., de Bruin, A., Ibelings, B.W., and Van Donk, E. (2007) Parasitic chytrids: their effects on phytoplankton communities and food-web dynamics. *Hydrobiologia*, **578**, 113–129.
- Kagami, M., Amano, Y., and Ishii, N. (2012) Community structure of planktonic Fungi and the impact of parasitic chytrids on phytoplankton in lake Inba, Japan. *Microbial Ecology*, **63**, 358–368.
- 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.
- Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*, **5**, 112.
- Kazama, F.Y. (1972) Ultrastructure and phototaxis of the zoospores of *Phlyctochytrium* sp., an estuarine chytrid. *Journal of General Microbiology*, **71**, 555–566.
- Keeling, P.J. and Fast, N.M. (2002) Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annual Review of Microbiology*, **56**, 93–116.
- 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.
- Keeling, P.J. and del Campo, J. (2017) Marine protists are not just big Bacteria. *Current Biology*, **27**, R541–R549.
- Keeling, P.J. and Burki, F. (2019) Progress towards the tree of eukaryotes. *Current Biology*, **29**, R808–R817.
- Kezlya, E., Tseplik, N., and Kulikovskiy, M. (2023) Genetic markers for metabarcoding of freshwater microalgae: review. *Biology*, **12**, 1038.
- Kinzig, A.P., Pacala, S.W., and Tilman, D. (2001) The functional consequences of biodiversity: empirical progress and theoretical extensions (MPB-33), Princeton, N.J., Princeton University Press.
- Klaveness, D., Shalchian-Tabrizi, K., Thomsen, H.A., Eikrem, W., and Jakobsen, K.S. (2005) *Telonema antarcticum* sp. nov., a common marine phagotrophic flagellate. *International Journal of Systematic and Evolutionary Microbiology*, **55**, 2595–2604.
- Ku, C. and Sebé-Pedrós, A. (2019) Using single-cell transcriptomics to understand functional states and interactions in microbial eukaryotes. *Philosophical Transactions of the Royal Society B*, **374**, 20190098.
- Kuris, A.M., Hechinger, R.F., Shaw, J.C., Whitney, K.L., Aguirre-Macedo, L., Boch, C.A., et al. (2008) Ecosystem energetic implications of parasite and free-living biomass in three estuaries. *Nature*, **454**, 515–518.

L

- Labarre, A., Obiol, A., Wilken, S., Forn, I., and Massana, R. (2020) Expression of genes involved in phagocytosis in uncultured heterotrophic flagellates. *Limnology and Oceanography*, **65**, 149–160.
- Labarre, A., López-Escardó, D., Latorre, F., Leonard, G., Bucchini, F., Obiol, A., et al. (2021) Comparative genomics reveals new functional insights in uncultured MAST species. *The ISME Journal*, **15**, 1767–1781.
- Lafferty, K.D., Dobson, A.P., and Kuris, A.M. (2006) Parasites dominate food web links. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11211–11216.
- Lafferty, K.D. (2010) Interacting Parasites. *Science*, **330**, 187–188.
- Lair, N. (1975) Sur la production des copépodes dans deux lacs du Massif Central français. *SIL Proceedings 1922-2010*, **19**, 3204–3211.
- Lamanna, C., Blonder, B., Violle, C., Kraft, N.J.B., Sandel, B., Šimová, I., et al. (2014) Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 13745–13750.
- Laundon, D., Christmas, 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.
- Lazarus, K.L. and James, T.Y. (2015) Surveying the biodiversity of the Cryptomycota using a targeted PCR approach. *Fungal Ecology*, **14**, 62–70.
- Le Bescot, N., Mahé, F., Audic, S., Dimier, C., Garet, M.-J., Poulain, J., et al. (2015) Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environmental Microbiology*, **18**, 609–626.
- Le Quéré, C., Harrison, S.P., Colin Prentice, I., Buitenhuis, E.T., Aumont, O., Bopp, L., et al. (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, **11**, 2016–2040.
- Lefèvre, E., Roussel, B., Amblard, C., and Sime-Ngando, T. (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS One*, **3**, e2324.
- Lefranc, M., Thénot, A., Lepère, C., and Debroas, D. (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Applied and Environmental Microbiology*, **71**, 5935–5942.
- 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., and Debroas, D. (2008) Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Applied and Environmental Microbiology*, **74**, 2940–2949.
- Lepère, C., Domaizon, I., Hugoni, M., Vellet, A., and Debroas, D. (2016) Diversity and dynamics of active small microbial eukaryotes in the anoxic zone of a freshwater meromictic lake (Pavin, France). *Frontiers in Microbiology*, **7**, 130.

- 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, L., Delgado-Viscogliosi, P., Gerphagnon, M., Viscogliosi, E., Christaki, U., Sime-Ngando, T., and Monchy, S. (2022) Taxonomic and functional dynamics during chytrid epidemics in an aquatic ecosystem. *Molecular Ecology*, **31**, 5618–5634.
- Litchman, E., Klausmeier, C.A., Schofield, O.M., and Falkowski, P.G. (2007) The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecology Letters*, **10**, 1170–1181.
- Litchman, E. and Klausmeier, C.A. (2008) Trait-based community ecology of phytoplankton. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 615–639.
- Locey, K.J. and Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 5970–5975.
- Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M.A., et al. (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology*, **13**, R66.
- 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.
- Luo, S., Shi, Y., Chin, L.K., Hutchinson, P.E., Zhang, Y., Chierchia, G., et al. (2021) Machine-learning-assisted intelligent imaging flow cytometry: a review. *Advanced Intelligent Systems*, **3**, 2100073.

M

- Ma, W., Zhang, S., Li, Z., Jiang, M., Wang, S., Lu, W., et al. (2022) Enhancing protein function prediction performance by utilizing AlphaFold-predicted protein structures. *Journal of Chemical Information and Modeling*, **62**, 4008–4017.
- Mackin, J.G., Owen, H.M., and Collier, A. (1950) Preliminary note on the occurrence of a new protistan parasite, *Dermocystidium marinum* n. sp. in *Crassostrea virginica* (Gmelin). *Science*, **111**, 328–329.
- 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. *Global Ecology and Biogeography*, **24**, 728–740.
- Mangot, J.-F., Lepère, C., Bouvier, C., Debroas, D., and Domaizon, I. (2009) Community structure and dynamics of small eukaryotes targeted by new oligonucleotide probes: new insight into the lacustrine microbial food web. *Applied and Environmental Microbiology*, **75**, 6373–6381.

- Marande, W., López-García, P., and Moreira, D. (2009) Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. *Environmental Microbiology*, **11**, 3179–3188.
- Marcogliese, D.J. and Cone, D.K. (1997) Food webs: a plea for parasites. *Trends in Ecology & Evolution*, **12**, 320–325.
- Martini, S., Larras, F., Boyé, A., Faure, E., Aberle, N., Archambault, P., et al. (2021) Functional trait-based approaches as a common framework for aquatic ecologists. *Limnology and Oceanography*, **66**, 965–994.
- Maselli, M., Anestis, K., Klemm, K., Hansen, P.J., and John, U. (2021) Retention of prey genetic material by the kleptoplastidic ciliate *Strombidium* cf. *basimorphum*. *Frontiers in Microbiology*, **12**, 694508.
- Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology*, **65**, 91–110.
- Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., et al. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environmental Microbiology*, **17**, 4035–4049.
- Massana, R., Labarre, A., López-Escardó, D., Obiol, A., Bucchini, F., Hackl, T., et al. (2021) Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate. *The ISME Journal*, **15**, 154–167.
- McMurdie, P.J. and Holmes, S. (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, **10**, e1003531.
- Medina, E.M. and Buchler, N.E. (2020) Chytrid fungi. *Current Biology*, **30**, R516–R520.
- de Meeûs, T. and Renaud, F. (2002) Parasites within the new phylogeny of eukaryotes. *Trends in Parasitology*, **18**, 247–251.
- Meng, A., Corre, E., Probert, I., Gutierrez-Rodriguez, A., Siano, R., Annamale, A., et al. (2018) Analysis of the genomic basis of functional diversity in dinoflagellates using a transcriptome-based sequence similarity network. *Molecular Ecology*, **27**, 2365–2380.
- 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.
- Möbius, K. (1877) *Die Auster und die Austernwirtschaft*, Berlin: Verlag Van Wiegandt, Hempel and Parey.
- Moore, C.W. (1982) Modulation of bleomycin cytotoxicity. *Antimicrobial Agents and Chemotherapy*, **21**, 595–600.
- Morand, S. (2015) (macro-) Evolutionary ecology of parasite diversity: from determinants of parasite species richness to host diversification. *International Journal for Parasitology: Parasites and Wildlife*, **4**, 80–87.
- Moreira, D. (2003) Are hydrothermal vents oases for parasitic protists? *Trends in Parasitology*, **19**, 556–558.

N

- Naeem, S. (2002) Autotrophic-heterotrophic interactions and their impacts on biodiversity and ecosystem functioning. In *The Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions (MPB-33)*. Kinzig, A.P., Pacala, S.W., and Tilman, D. (eds). Princeton University Press, pp. 96–119.
- Nata'ala, M.K., Avila Santos, A.P., Coelho Kasmanas, J., Bartholomäus, A., Saraiva, J.P., Godinho Silva, S., et al. (2022) MarineMetagenomeDB: a public repository for curated and standardized metadata for marine metagenomes. *Environmental Microbiome*, **17**, 57.
- Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., et al. (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, **32**, 822–828.
- Norén, F., Moestrup, Ø., and Rehnstam-Holm, A.-S. (1999) *Parvilucifera infectans* norén et moestrup gen. et sp. nov. (perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *European Journal of Protistology*, **35**, 233–254.
- Not, F., Simon, N., Biegala, I., and Vaultot, D. (2002) Application of fluorescent in situ hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. *Aquatic Microbial Ecology*, **28**, 157–166.
- Novodvorska, M., Stratford, M., Blythe, M.J., Wilson, R., Beniston, R.G., and Archer, D.B. (2016) Metabolic activity in dormant conidia of *Aspergillus niger* and developmental changes during conidial outgrowth. *Fungal Genetics and Biology*, **94**, 23–31.

O

- Orr, R.J.S., Zhao, S., Klaveness, D., Yabuki, A., Ikeda, K., Watanabe, M.M., and Shalchian-Tabrizi, K. (2018) Enigmatic Diphyllozoa eukaryotes: culturing and targeted PacBio RS amplicon sequencing reveals a higher order taxonomic diversity and global distribution. *BMC Evolutionary Biology*, **18**, 115.

P

- Pánek, T., Ptáčková, E., and Čepička, I. (2014) Survey on diversity of marine/saline anaerobic Heterolobosea (Excavata: Discoba) with description of seven new species. *International Journal of Systematic and Evolutionary Microbiology*, **64**, 2280–2304.
- Paolozzi, L. and Liébart, J.-C. (2015) *Microbiologie: biologie des procaryotes et de leurs virus*, Paris: Dunod.
- Park, M.G., Yih, W., and Coats, D.W. (2004) Parasites and phytoplankton, with special emphasis on dinoflagellate infections. *The Journal of Eukaryotic Microbiology*, **51**, 145–155.

- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, **25**, 1043–1055.
- Parro, V., Moreno-Paz, M., and González-Toril, E. (2007) Analysis of environmental transcriptomes by DNA microarrays. *Environmental Microbiology*, **9**, 453–464.
- Pawlowski, J. and Holzmann, M. (2002) Molecular phylogeny of Foraminifera a review. *European Journal of Protistology*, **38**, 1–10.
- Petchey, O.L. and Gaston, K.J. (2006) Functional diversity: back to basics and looking forward. *Ecology Letters*, **9**, 741–758.
- Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., and Baurain, D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, e1000602.
- Pollard, M.O., Gurdasani, D., Mentzer, A.J., Porter, T., and Sandhu, M.S. (2018) Long reads: their purpose and place. *Human Molecular Genetics*, **27**, R234–R241.
- Pomeroy, L., Williams, P.J.leB., Azam, F., and Hobbie, J. (2007) The Microbial Loop. *Oceanography*, **20**, 28–33.
- Poretzky, 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.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

R

- Rad-Menéndez, C., Gerphagnon, M., Garvetto, A., Arce, P., Badis, Y., Sime-Ngando, T., and Gachon, C.M.M. (2018) Rediscovering *Zygorhizidium affluens* Canter: molecular taxonomy, infectious cycle, and cryopreservation of a chytrid infecting the bloom-forming diatom *Asterionella formosa*. *Applied and Environmental Microbiology*, **84**, e01826-18.
- Ramakrishnan, S., Serricchio, M., Striepen, B., and Bütikofer, P. (2013) Lipid synthesis in protozoan parasites: a comparison between kinetoplastids and apicomplexans. *Progress in Lipid Research*, **52**, 488–512.
- 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.
- Rasconi, S., Jobard, M., Jouve, L., and Sime-Ngando, T. (2009) Use of calcofluor white for detection, identification, and quantification of phytoplanktonic fungal parasites. *Applied and Environmental Microbiology*, **75**, 2545–2553.

- Rasconi, S., Niquil, N., and Sime-Ngando, T. (2012) Phytoplankton chytridiomycosis: community structure and infectivity of fungal parasites in aquatic ecosystems: fungal epidemics in the plankton. *Environmental Microbiology*, **14**, 2151–2170.
- Rasconi, S., Grami, B., Niquil, N., Jobard, M., and Sime-Ngando, T. (2014) Parasitic chytrids sustain zooplankton growth during inedible algal bloom. *Frontiers in Microbiology*, **5**, 229.
- Reñé, A., Alacid, E., Gallisai, R., Chambouvet, A., Fernández-Valero, A.D., and Garcés, E. (2021) New Perkinsea parasitoids of dinoflagellates distantly related to Parviluciferaceae members. *Frontiers in Microbiology*, **12**, 701196.
- Reu, B., Proulx, R., Bohn, K., Dyke, J.G., Kleidon, A., Pavlick, R., and Schmidtlein, S. (2011) The role of climate and plant functional trade-offs in shaping global biome and biodiversity patterns: climate, biome and biodiversity patterns. *Global Ecology and Biogeography*, **20**, 570–581.
- Reuter, J.A., Spacek, D.V., and Snyder, M.P. (2015) High-throughput sequencing technologies. *Molecular Cell*, **58**, 586–597.
- Richter, D.J., Berney, C., Strassert, J.F.H., Poh, Y.-P., Herman, E.K., Muñoz-Gómez, S.A., et al. (2022) EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal*, **2**, e56.
- Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004) Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, **38**, 525–552.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, **4**, e2584.

S

- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
- Schloss, P.D. (2021) Amplicon sequence variants artificially split bacterial genomes into separate clusters. *mSphere*, **6**, e0019121.
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6241–6246.
- Scholz, B., Guillou, L., Marano, A.V., Neuhauser, S., Sullivan, B.K., Karsten, U., et al. (2016) Zoosporic parasites infecting marine diatoms – A black box that needs to be opened. *Fungal Ecology*, **19**, 59–76.

- Shakya, M., Lo, C.-C., and Chain, P.S.G. (2019) Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics*, **10**, 904.
- Shunmugam, S., Arnold, C.-S., Dass, S., Katris, N.J., and Botté, C.Y. (2022) The flexibility of Apicomplexa parasites in lipid metabolism. *PLoS Pathogens*, **18**, e1010313.
- Sime-Ngando, T., Lefèvre, E., and Gleason, F.H. (2011) Hidden diversity among aquatic heterotrophic flagellates: ecological potentials of zoosporic fungi. *Hydrobiologia*, **659**, 5–22.
- Simpson, A.G.B. and Roger, A.J. (2004) The real ‘kingdoms’ of eukaryotes. *Current Biology*, **14**, R693–R696.
- Singer, D., Seppey, C.V.W., Lentendu, G., Dunthorn, M., Bass, D., Belbahri, L., et al. (2021) Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environment International*, **146**, 106262.
- So, Y.-S., Maeng, S., Yang, D.-H., Kim, H., Lee, K.-T., Yu, S.-R., et al. (2019) Regulatory mechanism of the atypical AP-1-like transcription factor YAP1 in *Cryptococcus neoformans*. *mSphere*, **4**, e00785-19.
- Sogin, M.L., Elwood, H.J., and Gunderson, J.H. (1986) Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 1383–1387.
- Sogin, M.L. (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *American Zoologist*, **29**, 487–499.
- Sogin, M.L. (1991) Early evolution and the origin of eukaryotes. *Current Opinion in Genetics & Development*, **1**, 457–463.
- Stanier, R.Y. and Niel, C.B. (1962) The concept of a bacterium. *Archiv für Mikrobiologie*, **42**, 17–35.
- Stoeck, T., Hayward, B., Taylor, G.T., Varela, R., and Epstein, S.S. (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist*, **157**, 31–43.
- Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D.V., and Burki, F. (2019) New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Molecular Biology and Evolution*, **36**, 757–765.

T

- Taib, N., Mangot, J.-F., Domaizon, I., Bronner, G., and Debroas, D. (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. *PLoS One*, **8**, e58950.
- Tansley, A.G. (1935) The use and abuse of vegetational concepts and terms. *Ecology*, **16**, 284–307.
- Taylor, F.J.R., Hoppenrath, M., and Saldarriaga, J.F. (2007) Dinoflagellate diversity and distribution. In *Protist Diversity and Geographical Distribution. Topics in Biodiversity and Conservation*. Foissner, W. and Hawksworth, D.L. (eds). Dordrecht: Springer Netherlands, pp. 173–184.

- Tedersoo, L., Tooming-Klunderud, A., and Anslan, S. (2018) PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist*, **217**, 1370–1385.
- Tedersoo, L. and Anslan, S. (2019) Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. *Environmental Microbiology Reports*, **11**, 659–668.
- Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., et al. (2022) Best practices in metabarcoding of fungi: from experimental design to results. *Molecular Ecology*, **31**, 2769–2795.
- Thomas, F., Guégan, J.-F., Michalakakis, Y., and Renaud, F. (2000) Parasites and host life-history traits: implications for community ecology and species co-existence. *International Journal for Parasitology*, **30**, 669–674.
- Tikhonenkov, D.V., Mikhailov, K.V., Gawryluk, R.M.R., Belyaev, A.O., Mathur, V., Karpov, S.A., et al. (2022) Microbial predators form a new supergroup of eukaryotes. *Nature*, **612**, 714–719.
- Tokarev, Y.S., Voronin, V.N., Seliverstova, E.V., Pavlova, O.A., and Issi, I.V. (2010) Life cycle, ultrastructure, and molecular phylogeny of *Crispospora chironomi* g.n. sp.n. (Microsporidia: Terresporidia), a parasite of *Chironomus plumosus* L. (Diptera: Chironomidae). *Parasitology Research*, **107**, 1381–1389.
- 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.
- Tully, B.J., Graham, E.D., and Heidelberg, J.F. (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data*, **5**, 170203.
- Turk Dermastia, T., Vascotto, I., Francé, J., Stanković, D., and Mozetič, P. (2023) Evaluation of the rbcL marker for metabarcoding of marine diatoms and inference of population structure of selected genera. *Frontiers in Microbiology*, **14**, 1071379.

V

- Van den Wyngaert, S., Ganzert, L., Seto, K., Rojas-Jimenez, K., Agha, R., Berger, S.A., et al. (2022) Seasonality of parasitic and saprotrophic zoospore fungi: linking sequence data to ecological traits. *The ISME Journal*, **16**, 2242–2254.
- Vanni, C., Schechter, M.S., Acinas, S.G., Barberán, A., Buttigieg, P.L., Casamayor, E.O., et al. (2022) Unifying the known and unknown microbial coding sequence space. *eLife*, **11**, e67667.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science*, **348**, 1261605.
- Vaulot, D., Geisen, S., Mahé, F., and Bass, D. (2022) pr2-primers: an 18S rRNA primer database for protists. *Molecular Ecology Resources*, **22**, 168–179.
- Vazart, J. (1956) L'appareil cinétique chez les protistes et dans le règne végétal. *Bulletin de la Société Botanique de France*, **103**, 162–221.

- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., et al. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.
- Vernette, C., Henry, N., Lecubin, J., De Vargas, C., Hingamp, P., and Lescot, M. (2021) The ocean barcode atlas: a web service to explore the biodiversity and biogeography of marine organisms. *Molecular Ecology Resources*, **21**, 1347–1358.
- 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.
- Vitorino, L. and Bessa, L. (2018) Microbial diversity: the gap between the estimated and the known. *Diversity*, **10**, 46.

W

- 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.
- Wetzel, R.G. and Manny, B.A. (1972) Decomposition of dissolved organic carbon and nitrogen compounds from leaves in an experimental hard-water stream. *Limnology and Oceanography*, **17**, 927–931.
- Whittaker, R.H. (1969) New concepts of kingdoms of organisms: evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science*, **163**, 150–160.
- Williams, R.C. and Wyckoff, R.W.G. (1945) Electron shadow-micrography of virus particles. *Experimental Biology and Medicine*, **58**, 265–270.
- Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5088–5090.
- Woese, C.R. (1987) Bacterial evolution. *Microbiological Reviews*, **51**, 221–271.
- Wolska, M. and Mazurkiewicz-Zapałowicz, K. (2013) Parasites of zooplankton and periphyton assemblages in the littoral zone of lakes in Drawa national park, Poland. *Acta Mycologica*, **48**, 51–59.
- Wong, R.G., Wu, J.R., and Gloor, G.B. (2016) Expanding the UniFrac toolbox. *PLoS One*, **11**, e0161196.
- 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.
- Wurzbacher, C., Bärlocher, F., and Grossart, H. (2010) Fungi in lake ecosystems. *Aquatic Microbial Ecology*, **59**, 125–149.
- Wurzbacher, C., Rösel, S., Rychła, A., and Grossart, H.-P. (2014) Importance of saprotrophic freshwater fungi for pollen degradation. *PLoS One*, **9**, e94643.

Y

- Yabuki, A., Eikrem, W., Takishita, K., and Patterson, D.J. (2013) Fine structure of *Telonema subtilis* Griessmann, 1913: a flagellate with a unique cytoskeletal structure among eukaryotes. *Protist*, **164**, 556–569.
- Yabuki, A., Kamikawa, R., Ishikawa, S.A., Kolisko, M., Kim, E., Tanabe, A.S., et al. (2014) *Palpitomonas bilix* represents a basal cryptist lineage: insight into the character evolution in Cryptista. *Scientific Reports*, **4**, 4641.

Z

- Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, **52**, 79–92.
- Zhu, M., Duan, X., Cai, P., Li, Y., and Qiu, Z. (2022) Deciphering the genome of *Simplicillium aogashimaense* to understand its mechanisms against the wheat powdery mildew fungus *Blumeria graminis* f. sp. *tritici*. *Phytopathology Research*, **4**, 16.
- Zuckerkindl, E. and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*. Elsevier, pp. 97–166.

