



A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems

Aurore Dubuffet, Marina Chauvet, Anne Moné, Didier Debroas, Cécile Lepère

► To cite this version:

Aurore Dubuffet, Marina Chauvet, Anne Moné, Didier Debroas, Cécile Lepère. A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environmental Microbiology*, 2021, 23 (8), pp.4344-4359. 10.1111/1462-2920.15618 . hal-03252987

HAL Id: hal-03252987

<https://hal.science/hal-03252987>

Submitted on 27 Oct 2021

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.

**A phylogenetic framework to investigate the microsporidian communities through
metabarcoding and its application to lake ecosystems**

Running title: Microsporidian diversity in freshwater lakes

Aurore Dubuffet[°], Marina Chauvet, Anne Moné, Didier Debroas, Cécile Lepère.

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

[°]Corresponding author: Aurore Dubuffet, Université Clermont Auvergne, CNRS, Laboratoire Microorganismes: Génome et Environnement, F-63000 Clermont-Ferrand, France. E-mail: aurore.dubuffet@uca.fr

ORIGINALITY-SIGNIFICANCE STATEMENT

We introduce here a methodology which enables the study of environmental Microsporidia diversity (a phylogenetically diverse phylum of unicellular parasites), through metabarcoding.

We provide a comprehensive reference database which positions microsporidian SSU rRNA gene sequences within a coherent ranked phylogenetic framework. Using the developed methodology on lacustrine samples, we highlighted a considerable diversity, with many unknown genera and families.

SUMMARY

Microsporidia are obligate intracellular eukaryotic parasites known to parasitize many species of the animal kingdom as well as some protists. However, their diversity is underestimated, in part as a consequence of the failure of “universal” primers to detect them in metabarcoding studies. Besides, due to the inconsistency between taxonomy and phylogenetic data, available databases may assign incorrectly sequences obtained with high-throughput sequencing. In this work, we developed a comprehensive reference database which positions microsporidian SSU rRNA gene sequences within a coherent ranked phylogenetic framework. We used this phylogenetic framework to study the microsporidian diversity in lacustrine ecosystems, focusing on < 150 µm planktonic size fractions. Our analysis shows a high diversity of Microsporidia, with the identification of 1531 OTUs distributed within 7 clades, of which 76% were affiliated to clade IV2 and 20% to clade I (nomenclature presented hereby). About a quarter of the obtained sequences shared less than 85% identity to the closest known species, which might represent undescribed genera or families infecting small hosts. Variations in the abundance of Microsporidia were recorded between the two lakes sampled and across the sampling period, which might be explained by spatio-temporal variations of their potential hosts such as microeukaryotes and metazooplankton.

INTRODUCTION

For long underestimated, the diversity of parasites and their central role in the functioning of ecosystems are nowadays recognized (Lafferty et al., 2006; Dobson et al., 2008; Lafferty et al., 2008). For example, it is estimated that about 40% of species, all ecosystems combined, are parasites (Dobson et al., 2008) and that more than 75% of the trophic links include parasites (Lafferty et al., 2006). However, these parasites only represent 5% of the described species (Costello, 2016) and further studies are needed to reveal the extent of their diversity and to clarify their role in ecosystems. The advent of molecular approaches, in particular metabarcoding, has revealed a considerable diversity of microeukaryotes (<200 μm) in different types of ecosystems, including parasite groups which can represent up to 50% of the total diversity (Cleary and Durbin, 2016; Lepère et al., 2016; Mahé et al., 2017). In lacustrine ecosystems these studies highlighted the presence of several group of putative parasites, such as chytrids (Fungi), Apicomplexa and Perkinsea (Alveolates) (Debroas et al., 2017). Some groups of parasites, however, are largely absent from these studies (see Bass et al., 2015 for review). This is the case of Microsporidia, which are obligate unicellular intracellular parasites related to Fungi (Vavra and Lukes, 2013). Because of their rapid evolution and significant reduction of their genomes (Keeling and Fast, 2002), they are genetically divergent from other eukaryotes, making the primers conventionally targeting the conserved regions of the rRNA 18S gene ineffective (Marande et al., 2009; Hadziavdic et al., 2014; Hugerth et al., 2014). Some studies attempted to characterize the diversity of this phylum using Microsporidia-specific primers, mainly in particular hosts such as daphnids, amphipods or mosquitoes (Weigl et al., 2012; Quiles et al., 2019; Trzebny et al., 2020) but more rarely in environmental samples (Ardila-Garcia et al., 2013; Williams et al., 2018).

Based on ultrastructural, ecological (*e.g.* host species, habitat, transmission route) and tissue tropism criteria, more than 1300 Microsporidia species have been described so far (Vossbrinck et al., 2014), most infecting one or several closely related hosts. This may nevertheless represent only a minor fraction of their real diversity, as the number of microsporidian species is estimated to approach the number of all animal species (Keeling and Fast, 2002; Larsen et al., 2017). Microsporidia are found in all ecosystems, and particularly in aquatic ecosystems which harbour about 50% of the described genera (Stentiford et al., 2013a). Vertebrates, crustaceans and insects alone account for nearly 80% of known

microsporidian hosts (Stentiford et al., 2013a; Weiss and Becnel, 2014; Stentiford et al., 2016). Such proportion however is likely due to biases towards the search of infectious agents in hosts of sanitary (human, disease vector insects such as mosquitoes), veterinarian (domestic and farm animals), agronomic (auxiliary insects and pests) or ecological importance (daphnids, amphipods). As a result, the global diversity of Microsporidia is largely unknown, particularly in small sized hosts, such as protists which represent only 10 of the 7366 microsporidian SSU sequences listed in the databases (NCBI consultation 18/11/2020 with the key words ‘small subunit ribosomal’+ microsporidia).

In this work we used a targeted metabarcoding approach to specifically study the microsporidian diversity in lake ecosystems, focusing on planktonic fractions containing small sized hosts. High-throughput sequencing classically produces thousands of sequences, which need to be assigned taxonomically in order to perform comparisons between samples. Such assignments are based on databases that link sequences to taxonomy. We first checked the consistency between available databases and phylogenetic data, and found that they are either weakly detailed or contain several errors. We thus built a comprehensive reference database which positions microsporidian SSU rRNA gene sequences within a coherent ranked phylogenetic framework covering the microsporidian diversity. Moreover, we addressed the question of OTU (Operational Taxonomic Unit) definition in Microsporidia. Indeed, sequences of taxonomic marker genes with high similarity scores are classically clustered into OTUs, themselves considered as proxies for species. We therefore determined the clustering threshold that maximizes the recovery of distinct species and/or genera for microsporidian SSU sequences. Using this framework, we highlighted a considerable uncovered microsporidian diversity in our samples, and evidenced spatio-temporal variations in the abundance of many Microsporidia species.

MATERIALS AND METHODS

Microsporidia database

An exhaustive search of all Microsporidia species for which a SSU sequence is available was conducted (last update: march 2020), using hundreds of publications and the NCBI database (nucleotides and WGS sequences). We focused on the canonical Microsporidia, characterized by a specific invasion apparatus known as the polar tube, an extreme genome compaction, a lack of true mitochondria, and a rRNA operon of prokaryote-like size and structure (Wadi and Reinke, 2020). We also included sequences from Chytridiopsida and Metchnikovellida, considered as unusual Microsporidia as they lack some of the typical features of Microsporidia (Mikhailov et al., 2016; Corsaro et al., 2019). When several sequences were found for a species, we kept the longest or the most consensual one. We found 337 microsporidian species with a SSU sequence. We then selected the species for which the full length of the SSU rRNA gene was available (at least 90% of the V1-V9 region, 265 sequences, Supplementary Information 1). Their taxonomic assignments were searched in public databases (NCBI, SILVA, PR², PHYMYCO-DB) and compared with several microsporidian phylogenies (Vossbrinck and Debrunner-Vossbrinck, 2005; Smith, 2009; Vossbrinck et al., 2014; Corsaro et al., 2019) and a recent taxonomic review (Wijayawardene et al., 2018).

As none of the public databases was satisfying (see results), we built an extensive phylogeny of Microsporidia, based on their full-length SSU gene sequences. These sequences, as well as 6 outgroup sequences (from *Rozella allomycis*, *Paramicrosporidium* sp., *Nucleophaga* sp. and *Mitosporidium daphnia*), were aligned using the online version of MAFFT (ver. 7; Computational Biology Research Consortium [<http://mafft.cbrc.jp/alignment/server/index.html>]; Katoh et al., 2005) with the E-INS-i multiple alignment method, leaving gappy regions, a gap opening penalty of 1.0 and gap extension penalty extension of 0.0. The alignment was checked using MEGA 7 (Kumar et al., 2016) and both ends were trimmed. The resultant alignment was analysed by Maximum Likelihood (ML) and Neighbor joining (NJ) using MEGA 7.0 (Kumar et al., 2016). A General Time-Reversible (GTR) model of evolution with gamma-distributed rate heterogeneity (G) and a proportion of invariable sites (I) (GTR

+ G + I) was used to create the ML tree, with 500 bootstrap replications and partial deletion (85%). The NJ tree was obtained using the Kimura-2-parameter model with a gamma distribution of 4, partial deletion (85%) and 1000 bootstraps. Metadata (habitat and hosts) were reported on the ML tree using iTOL (Letunic and Bork, 2016). Similar phylogenetic analyses were done using the region of the SSU gene which was chose for the metabarcoding analysis described below (V1-V3, corresponding to the first three variables regions of the SSU gene). These trees were used to define three phylogenetic ranks inside the Microsporidia phylum: clade, group and genus. This classification (see results and discussion section) was subsequently used to assign the Operational Taxonomic Units (OTUs) obtained from environmental samples.

OTU definition

We tested different clustering thresholds on selected SSU sequences (species with a clear genus affiliation, and for which the full length of the SSU gene was available, 255 sequences) in order to determine which one could maximize the recovery of distinct genera and species. Full-length and V1-V3 sequences were aligned with the Needleman-alignment algorithm and a distance matrix was obtained with the pairwise.seqs command in Mothur (Schloss et al., 2009). Then, the average neighbor algorithm was used with clustering thresholds ranging from 100% (unique sequences) to 90% in steps of 1%. The number of obtained OTUs was then compared to the number of species (255) and genera (104) in the dataset (% recovery).

Study sites and sampling procedure

The study was conducted in two lakes (Massif Central, France). Lake Pavin (45°55' N and 2°54' E) is an oligo-mesotrophic lake with an area of 0.44 km² and a maximum depth of 92 m, at an altitude of 1197 m above sea level. It is a meromictic lake characterized by the presence of two permanent stratified layers. Lake Aydat (45°39' 50" N, 2° 59' 04" E) is a dimictic and eutrophic lake with a maximum depth

of 15 m, situated at an altitude of 825 m. In April, May and June 2016, water samples were collected in the euphotic zone of both lakes at a permanent station located at the deepest zone of the water column. 6-10 L of water were filtered successively through 150 µm, 50 µm, and 10 µm pore-size filters in order to concentrate Microsporidia spores (<10 µm fraction) and planktonic species susceptible to host Microsporidia (10-50 and 50-150 µm fractions). The microbial biomass of each size-fraction was collected on 0.8µm pore-size (pressure, <10 kPa) polycarbonate filters (Millipore, Molsheim, France) and stored at -80°C until further analysis.

Nucleic acids extraction, amplification and sequencing

Filters were covered with the extraction buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris). 0.1g of 0.1 mm glass beads were added and samples were homogenized in a bead beater (3 pulses for 30s at 30 Hz). Tubes were then centrifuged during 1 min at 600 g. EDTA pH8 and a lysozyme solution (final concentrations 0.1 M and 1.5 mg.ml⁻¹ respectively) were added to the supernatant and incubated at 37°C for 30 min. Then sodium dodecyl sulfate (1.8%, final), proteinase K (100 µg.ml⁻¹, final) and RNaseA (50 µg.ml⁻¹, final) were added, and samples were incubated at 37°C for at least 60 min. A cetyltrimethylammonium bromide (CTAB) solution (final concentration 1% in a 0.7 M NaCl solution) was added, and samples were incubated at 65°C for 10 min. Nucleic acids were extracted with phenol-chloroform-isoamyl alcohol (25:24:1); the aqueous phase containing nucleic acids was kept and purified by adding chloroform-isoamyl alcohol (24:1). After isopropanol (0.6 vol) addition, the nucleic acids were precipitated at -20°C for 12 h. After centrifugation, the DNA pellet was ethanol rinsed and resuspended in 20 µL of water.

Amplification of the V1-V3 region of the Microsporidia small subunit rDNA was performed using the universal microsporidian V1F (CACCAGGTTGATTCTGCCTGAC) and 530R (CCGCGGCTGCTGGCAC) primers set (Baker et al., 1994; Weiss et al., 1994) modified by adding barcodes. Each PCR was performed in a total volume of 30 µL containing 3 µL of 10× NH₄ reaction buffer, 1.2 µL of 50 mM MgCl₂, 0.15 µL of EurobioTaqII (Eurobio, Les Ulis, France, 5 U/µL), 0.6 µL

of 10 mM each dNTP, 0.3 μ L of 50 mg/mL BSA and 1.2 μ L of each 5 μ M primer. The amplification conditions consisted of initial denaturation at 94 °C for 5 min followed by 35 cycles for 1 min at 94 °C, 1 min 10 s at 61 °C and 1min at 72 °C, and a final elongation of 7 min at 72 °C.

The PCR products were run on a 1.2 % agarose gel electrophoresis. The amplicons were purified and concentrated using the MinElute gel extraction kit (Qiagen®, Courtaboeuf, France), and quantified using the Agilent 2200 TapeStation system and the D1000 ScreenTape kit (Agilent Technologies, Les Ulis, France). Finally, the purified amplicons for each sample were pooled at equimolar concentrations and sent to GATC Biotech (Konstanz, Germany) for library preparation and paired-end (2x300 bp) sequencing on a MiSeq Illumina instrument (San Diego, CA).

Amplicons analysis & phylogenetic affiliation

The MiSEQ data were assembled with the VSEARCH tool (Rognes et al., 2016). These sequences were then subjected to four cleaning procedures: (i) no undefined bases (Ns), (ii) a minimum sequence length of 300 bp, (iii) no sequencing error in the forward and reverse primers and (iv) no chimeras (Edgar et al., 2011). After this cleaning step and demultiplexing process, the remaining DNA sequences were clustered at 99%. The representative OTUs were then affiliated by similarity and phylogeny using a custom curated reference database (Freshwater-database: <https://github.com/panammeb/Freshwater-database>). This database contains eukaryote references extracted from the SSURef SILVA database (Pruesse et al., 2007), typical freshwater lineages identified in previous studies (e.g. Debroas et al., 2015), and 528 sequences from the 265 microsporidian species (107 genera) we analysed in the microsporidian phylogenetic tree (full length SSU sequences). After a comparison of the OTUs with the Freshwater-database by a similarity approach (VSEARCH tool), trees of OTUs with their closest reference sequences were built in FastTree (Price et al., 2010) (see the detailed pipeline in Debroas et al., 2017). Taxonomic assignment was conducted according to two methods: nearest neighbour (NN) and last common ancestor (LCA) affiliations. This process was implemented in the pipeline PANAM (Phylogenetic Analysis of Next-generation Amplicons; <https://github.com/panammeb/>) and is described

in more detail in Taib et al., 2013. Affiliations to groups and genera were considered as valid when identity scores were $\geq 85\%$ and $\geq 94.5\%$, respectively (empirical thresholds based on comparisons between affiliations and phylogenetic analyses, not shown).

Since the clustering threshold we used (99%) may generate OTUs resulting from sequencing errors, we only kept for the final analysis OTUs representing more than 0.05% of the total reads in at least one sample. Two samples were removed from the analysis since they contained less than 20 000 reads. Data were finally normalized using rarefaction with the phyloseq package on R (R Development Core Team, 2011; McMurdie and Holmes, 2013).

Diversity indices

Community richness and diversity indices (Richness observed, Chao1, Shannon and Simpson) were calculated using vegan package on R (R Development Core Team, 2011; Oksanen et al., 2017). Non-metric multidimensional scaling (NMDS) ordinations, based on a Bray-Curtis dissimilarity matrix (Bray and Curtis, 1957; Paliy and Shankar, 2016), were performed to depict the difference in Microsporidia community compositions among the samples using the R vegan package (R Development Core Team, 2011; Oksanen et al., 2017).

Phylogenetic analysis of OTUs

All OTUs obtained in this analysis were assigned at least to a clade, but 24% of them were not assigned to a group. We therefore performed phylogenetic analyses (for each clade) on OTUs not assigned to a group in order to determine whether new groups could be highlighted. Besides, only 48% of the OTUs were assigned to a genus. We checked the accuracy of affiliations to genera and highlighted potential new genera through phylogenetic analyses in each identified group. We focused our analyses on abundant OTUs ($>1\%$ of reads in at least one sample, 95 OTUs), as they represent between 53 and 92% of the total reads in each sample. We included in these analyses several sequences from undescribed

microsporidian species (found through BLAST search and/or bibliographic review), in addition to species from our database. After alignment using the MAFFT algorithm (Katoh et al., 2005), ML trees were obtained using a General Time-Reversible (GTR) model of evolution with gamma-distributed rate heterogeneity (G) and a proportion of invariable sites (I) (GTR + G + I), with 200 bootstrap replications and partial deletion (85%), using MEGA 7 (Kumar et al., 2016). Relative abundances were plotted on these trees using iTOL (Letunic and Bork, 2016).

RESULTS AND DISCUSSION

OTU definition for Microsporidia

Metabarcoding analyses usually produce thousands of sequences, which are then clustered into OTUs to produce ‘species equivalents’ (Eren et al., 2011; Alberdi et al., 2018). For eukaryotic SSU sequences, a similarity threshold of 95% (Caron et al., 2009; Debroas et al., 2015) or alternatively 97% (Cleary and Durbin, 2016; Bista et al., 2017) is usually employed. However, an incorrect clustering threshold could cause oversplitting or undersplitting of sequences into species equivalents, resulting in either inflated or underestimated species counts (Alberdi et al., 2018). We therefore addressed this issue by applying different clustering thresholds on the full-length and V1-V3 sequences of the Microsporidia species we selected. Such analysis indicated that an important proportion of microsporidian species has highly similar SSU sequences (about 13% of species clustered together with a 99% clustering threshold) or even identical sequences (0.4% (full-length) and 2.7% (V1-V3) of species clustered together with a 100% clustering threshold) (Supplementary Information 2). None of the clustering thresholds may in this way generate OTUs that could be identified with high accuracy at the species level, thus highlighting the limits of this gene marker for species discrimination. Besides, thresholds for species delineation may vary according to the lineages, since some species belonging to the same genera have sometimes less than 90% identity (Zheng et al 2016), while some share more than of 99.5% identity in other genera (Tokarev et al., 2020). For the later genera, combination of different molecular markers is then sometimes required to improve phylogenetic resolution (Bacela-Spychalska et al 2018; Tokarev et al

2020). As 100% of genera could nevertheless be distinguished using a 99% threshold on V1-V3 sequences, we chose to keep this threshold and only consider affiliations to genera as valid in further analyses.

Available databases are not reliable for microsporidian classification

OTUs assignment to phylogenetic clades requires the use of a reliable reference database, in which species are classified into several taxonomic ranks. As assignment lies on the comparison between OTUs and reference sequences, the efficiency of OTU classification depends on the degree to which taxonomic classification of reference sequences fits their evolutionary relationships. However, it is largely recognized that taxonomic designation of Microsporidia, based primarily on morphological characters, is inconsistent with phylogenetic analyses (Vossbrinck and Debrunner-Vossbrinck, 2005; Stentiford et al., 2013b; Tokarev et al., 2020). We thus primarily checked whether taxonomic information contained in various public databases was consistent with microsporidian phylogenetic data, focusing on the species for which the full length of the SSU gene was available (265 sequences). We found that taxonomic data found in GenBank (Federhen, 2012) are largely in disagreement with the phylogenetic clusters identified in the literature (Vossbrinck and Debrunner-Vossbrinck, 2005; Smith, 2009; Vossbrinck et al., 2014; Corsaro et al., 2019; Supplementary Information 1). In SILVA (Quast et al., 2013), many of these microsporidian species are lacking (12%, Supplementary Information 1) or when present, often assigned incorrectly to phyla which are not Microsporidia (21%, Supplementary Information 1). Besides, SILVA, as well as PR² (a database dedicated to protists; Guillou et al., 2013), do not include any taxonomic rank between phyla and genera (see Supplementary Information 1). PHYMYCO-DB (Mahé et al., 2012), a database dedicated to Fungi, does not contain microsporidian sequences. The existing databases being inappropriate, another one based on the phylogeny rather than on morphological characters, was thus required. Although the microsporidian taxonomy should probably be revised (as discussed in Vossbrinck and Debrunner-Vossbrinck, 2005 and Stentiford et al.,

2013b), our objective was simply here to provide a practical classification of microsporidian species which could be used to analyse metabarcoding data.

Extensive phylogeny of Microsporidia

We built a phylogenetic tree which is twice more exhaustive than the previously published reference trees (Vossbrinck and Debrunner-Vossbrinck, 2005; Vossbrinck et al., 2014). NJ and ML phylogenetic trees obtained on the full-length sequences placed canonical Microsporidia in a monophyletic group, with Metchnikovellids and Chytridiopsidae at the basal position of the microsporidian phylogenetic tree (Figure 1). A total of 10 clades were identified (9 for canonical Microsporidia numbered with Latin names, and one for Metchnikovellidae), all being supported by high bootstrap values (>85%).

Former reference trees identified five major clades in canonical Microsporidia (Vossbrinck and Debrunner-Vossbrinck, 2005; Vossbrinck et al., 2014). It should however be noted that clade numbering between these two reference publications differs, as “clade 5” in Vossbrinck et al., 2014 corresponds to “clade III” in Vossbrinck and Debrunner-Vossbrinck, 2005, and vice-versa. As a result, clade numbering for Microsporidia is currently confusing, since some publications follow the former and others the later (including in recent publications, see Zhang et al., 2016 or Quiles et al., 2019 for example). We chose here to keep the oldest clade numbering, as in proportion more publications refer to it.

Among the ten clades we highlighted, four of them (clades I, III, IV and V) were largely in agreement with the ones formerly described (Vossbrinck and Debrunner-Vossbrinck, 2005; Smith 2009; Vossbrinck et al., 2014; Corsaro et al., 2019, see comparison in Supplementary Information 4). The biggest clade (clade IV, thus considered as a “super-clade”, 86 species) was here subdivided into 2 clades (IV1 and IV2). Such division was supported by high bootstrap values, a clear genetic divergence, and literature data (Vossbrinck and Debrunner-Vossbrinck, 2005; Vossbrinck et al., 2014; Trzebný et al., 2020). However, we found only a weak overlap for the clade II, which species composition already differed markedly between the former reference phylogenies (Supplementary Information 4). As we included several new species in the current phylogenies, it appears that former clade II could be split

into three different clades (here II, VIII and IX). Eventually, we introduce two distinct novel clades (VI and VII), containing species which were either unclassified (Vossbrinck and Debrunner-Vossbrinck, 2005; Smith, 2009) or included in clade 5 (Vossbrinck et al., 2014) probably due to long branch attraction.

Within these clades, we furthermore identified 22 ‘groups’ of species, also supported by high bootstrap values ($\geq 95\%$) and previous phylogenetic studies (Figure 1). We named these groups according to the names given in these studies when available, or according to a representative genus (i.e. *genera name – like* group) for the others. Again, a strong discrepancy between these phylogenetic data and taxonomy provided in NCBI was found (Supplementary Information 5). However, our data were mostly in agreement with the microsporidian taxonomy, since most of the phylogenetic ‘groups’ we identified corresponded to families reviewed in Wijayawardene et al., 2018 (Supplementary Information 6). Other groups included species classified in various families (such as the ‘Aquatic group’ (Vossbrinck et al., 2004), which contains representatives of the Caudosporidae, Gurleyidae and Agglomeratidae families (Wijayawardene et al., 2018)), or species not assigned yet to any family (such as representatives of the Orthosomella-like or Nematocida-like groups). All these groups were recovered in the phylogeny based on the V1-V3 region, highlighting the relevance of this region (Supplementary Information 7).

Some trends in terms of habitats and to a lesser extent of hosts are characterizing these clades or groups (Figure 1). Clades I, VI, VII, VIII, as well as Neopereziiidae (clade V) mostly contain species infecting freshwater hosts, such as Crustacea, aquatic Diptera or Bryozoa, while clades II, IV1, IX and Tubulinosematidoidea (clade V) contain mostly species infecting terrestrial hosts. Clade III contain aquatic (freshwater and marine) species. One group within clade IV2 however, contains microsporidian species infecting very different hosts (ciliates, insects, crustacean, human) found in diverse habitats (terrestrial, freshwater, brackish). We named this group ‘Arlequin group’ to reflect this variability. Clades and groups highlighted by this phylogenetic analysis and literature review were thereafter used to classify OTUs in our metabarcoding analysis. Sequences included in the phylogeny, as well as other full-length SSU sequences of the same species (total of 528 sequences), were included in a database used to study eukaryotic communities in freshwater environments (Debroas et al., 2017).

Application of the proposed methodology to study the lacustrine microsporidian diversity

So far, only two studies attempted to characterize the diversity of Microsporidia in environmental samples, focusing mostly on terrestrial and marine habitats (Ardila-Garcia et al., 2013; Williams et al., 2018). Both studies highlighted new diversity across all areas of the microsporidian tree. However, due to the low number of sequences obtained in these studies (25 and 745 respectively, with a maximum of 20 OTUs per sample), comparisons of microsporidian communities between habitats were either not done or could only uncover a general variability between samples. In this study, we use our expertise database combined with the use of specific primers to characterize the diversity of Microsporidia in two lake ecosystems (planktonic samples) as well as its spatio-temporal variations through metabarcoding.

Diversity

After quality filtering and clustering at 99%, 1 135 287 reads were obtained, and 91% of them were assigned to Microsporidia. After removal of the two samples which contained < 20 000 reads (Pavin, April, fraction 50-150 µm and Pavin, June, fraction 10-50µm) and normalisation, 323 904 reads were assigned to the microsporidian classification we defined above. This analysis revealed a considerable diversity, since 1531 OTUs were obtained in total. However, the Microsporidia community was dominated by few highly abundant OTUs, since 95 “abundant” OTUs (>1% in at least one sample) covered more than 75% of the total reads. These results highlight, together with those of Ardila-Garcia et al., 2013 and Williams et al., 2018, the necessity to sample more environments in order to reveal the extent of the environmental microsporidian diversity.

All OTUs obtained in this study belonged to the canonical Microsporidia, and were distributed within 7 clades (over the 10 clades identified for Microsporidia). The microsporidian community was clearly dominated by clades IV2 and I, which represent respectively 76% and 20% of the OTUs (Figure 2). The

remaining OTUs were classified into clades IV1 (<2%), V (<1.5%), II (<0.2%), VIII (<0.2%) and IX (<0.2%).

Nine groups (over the 21 belonging to the canonical Microsporidia) were represented. Three of them accounted for more than 74% of the OTUs: Aquatic outgroup (clade I, 19%), Arlequin and *Orthosomella*-like groups (clade IV2, 45% and 10% respectively) (Figure 2). 47% of OTUs were assigned to a genus, of which *Crispospora*, *Pseudoberwaldia* and *Conglomerata* were the most represented (35%, 5% and 4%, respectively). Phylogenetic analyses confirmed these affiliations (Figure 3 and Supplementary Information 8), which supports the accuracy of the developed methodology. As expected, most of these genera are known to infect hosts having at least one aquatic stage across their full development (crustaceans: daphnids or cyclops, insects: chironomids, Figure 2).

Besides the identified genera, many OTUs were not assigned to a genus (53%) nor sometimes even to a group (23%). OTUs not assigned to any genus within the Arlequin group (Figure 3), Aquatic outgroup (Supplementary Information 8) and *Orthosomella*-like group (Supplementary Information 9) were often related to undescribed Microsporidia found in freshwater hosts, such as amphipods, cladocerans, copepods (crustacean), caddisflies, mayflies and mosquitoes (insects). OTUs not assigned to any group mostly belonged to clade IV2 (21%), and few to clades IV1 and V (1% each) (Figure 2). Phylogenetic analysis of these OTUs confirmed that they do not belong to any of the groups we formerly identified (clade IV2: Figure 4; clades IV1 and V: Supplementary Information 10-11). Their closest relatives were often Microsporidia amplified from environmental samples (Ardila-Garcia et al., 2013; Williams et al., 2018) or from amphipods sampled in the lake Baikal (Ironside and Wilkinson, 2018). In clade IV2, they notably form 5 new groups (Van9, BFAS6, Van10, BPAR5 and Van8 groups, Figure 4), each having less than 85% similarity with groups identified in this clade through phylogenetic analysis of known species (Supplementary Information 12). These results suggest that for freshwater habitats, a large proportion of genera and/or families remains to be described. As our analysis focused on small sized fractions (< 150 μ m), this novel diversity may correspond at least partly to not yet described microsporidian species infecting small hosts. Some studies report infections of ciliates and rotifers by Microsporidia in freshwater environments (Fokin et al., 2008; Wolska and Mazurkiewicz-Zapalowicz,

2013; Stentiford et al., 2013a; Yakovleva et al., 2020), but only two sequences from ciliates, and none from rotifers are currently available in the databases. Further sampling efforts aiming to identify host-microsporidian interactions in these small size organisms will be necessary in order to confirm this hypothesis.

Spatio-temporal variations

α -diversity of microsporidian community varied among the different samples (Supplementary Information 13). When focusing on community richness (Richness observed, Chao1 index, number of OTUs), the two lakes seem to present an opposite pattern, as richness increased from April to June in lake Aydat while it decreased in Lake Pavin. Similar patterns were observed for Shannon and Simpson indexes (Supplementary Information 13 B).

While no trend for the different size fractions was detected, variations in the relative abundances of some taxa were observed between sampling sites, and to a lesser extent across time (in lake Aydat notably), as suggested by the non-metric multidimensional scaling (Figure 5A). For example, while OTUs identified as *Crispospora* (IV2-Arlequin group), *Conglomerata* and *Pseudoberwaldia* (I-Aquatic outgroup) sp. are very abundant in lake Aydat (up to 91, 17 and 5%, respectively, shown in black, light-blue and deep blue in Figure 5B), their relative abundances were in comparison low in lake Pavin (maximum 11, 0.2 and 0.3%, respectively, except in one sample where they reached 19, 2 and 3%). OTUs belonging to clade IV1 from unknown group(s) had similar pattern, reaching up to 13% in Aydat, and being nearly absent in Pavin (<0.4%) (striped green on Figure 5B). These trends were also observed at the OTU level, some being clearly more abundant in one lake than in the other (Figure 3 and Supplementary Information 8 and 10). At the opposite, lake Pavin was clearly dominated by Microsporidia belonging to unknown group(s) in the clade IV2 or to unknown genera in the Aquatic outgroup (clade I) (striped orange and dotted blue respectively on Figure 5B). Phylogenetic analysis of these unassigned OTUs within clade IV2 revealed that they belong to 3 *de novo* groups (Van 9, BFAS6 and Van8) (Figure 4). In particular, a cluster of closely related OTUs in group Van8 was far more

abundant in lake Pavin (some OTU reaching up to 54% of the total reads) than in lake Aydat (<1.3%) (Figure 4). Similarly, phylogenetic analysis of abundant OTUs belonging to the Aquatic outgroup revealed that several OTUs clustering with *Octosporea muscadomesticae* and unknown Microsporidia found in freshwater hosts were far more abundant in lake Pavin (up to 21%) than in lake Aydat (<0.5%) (Supplementary Information 8).

Relative abundances of some taxa can also significantly vary across the sampling period (Figure 5B). For example, abundances of *Pseudoberwaldia* sp. (Aquatic outgroup, clade I) increased over time in lake Aydat (deep blue in Figure 5B, Supplementary Information 8), as did Microsporidia in the Van8 group (clade IV2) in lake Pavin, which clearly dominates in June (>84% of the reads) (Figure 4). Besides, Microsporidia belonging to the *Orthosomella* group were much more abundant in April compared to May and June (dotted green in Figure 5B, Supplementary Information 9). Although these general trends can be observed at the group or genus level, striking differences can also be observed at the OTU level: in OTUs assigned to the genus *Crispospora* for example, OTU3989 and OTU7498 were more abundant in April (up to 85 and 29%), while OTU19444 was more abundant in May (up to 45%), and OTU24464 in June (up to 21%), suggesting these OTUs might belong to different species (Figure 3 and Supplementary Information 14A). As genetic distances between these OTUs can be lower than 2% (Supplementary Information 14B), such data support the relevance of a clustering threshold fixed at 99%.

These spatio-temporal variations might be related to the abundance of the hosts present in these lakes at the time of sampling, as Microsporidia have generally a narrow host specificity. Indeed, these two lakes differ in their zooplankton species compositions (Lair, 1978; Devaux, 1980; Lair, 1990; Quiblier-Lloberas et al., 1996), which could themselves be influenced by differences in the nutrients content, altitude and mixing type recorded in these two lakes (Lepère et al., 2010). Considerable temporal variations in the abundances of these species have also been recorded, which could in part explain the temporal variations observed in some Microsporidia lineages. For example, increases of the Cladoceran population are repeatedly recorded in lake Aydat in spring, which could explain the increase of Agglomeratidae (*Pseudoberwaldia* and *Conglomerata* species) across our sampling period (Lair, 1990).

Monitoring the presence and abundances of both the microsporidian and other eukaryotic species across a significant period of time could help, for example through co-occurrence network analyses, to highlight potential new host-microsporidian interactions. This could be of particular interest for the Van8 group, which clearly dominates in lake Pavin, and which hosts are totally unknown. This group, as well as the other dominant lineages (*Crispospora* sp. and Agglomeratidae species) might play a significant role in the regulation of hosts populations, that further work may help to determine.

Conclusion

In this work, we propose a methodology which aims at studying the microsporidian diversity through metabarcoding. We notably provide a robust database, based on an extensive phylogenetic and bibliographic analysis, which classify microsporidian species in a coherent framework. Using this database, we explored the microsporidian diversity in lacustrine samples through a metabarcoding approach, and highlighted a substantial amount of Microsporidia phylogenetically distant from any described microsporidian species, which might represent undescribed genera and families. Further studies, including larger datasets (in terms of geography and/or ecology) will help to reveal the extent of Microsporidia diversity, which is obviously far from being fully uncovered. Besides, further analyses will be necessary to determine which hosts these Microsporidia are able to infect, a necessary step to understand the role of these parasites in the ecosystem functioning.

ACKNOWLEDGMENTS

The authors are very grateful to Laetitia Juillard for her priceless tenacity to devise the right PCR conditions during her Master traineeship, and for her help with sampling and DNA extractions. We also gratefully acknowledge constructive comments by two anonymous referees. Authors acknowledge the financial support from the “Fédération des Recherches en Environnement” through the CPER “Environment” founded by the Region Auvergne, the French government and the FEDER (European Community).

451

452 **CONFLICT OF INTEREST**

453 The authors declare they have no competing interests

454

REFERENCES

- Alberdi, A., Aizpurua, O., Gilbert, M.T.P., and Bohmann, K. (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods Ecol Evol* **9**: 134–147.
- Ardila-Garcia, A., Raghuram, N., Sihota, P., and Fast, N. (2013) Microsporidian diversity in soil, sand and compost of the pacific Northwest. *J. Eukaryot. Microbiol* **60**: 601-608.
- Baker, M.D., Vossbrinck, C.R., Maddox, J.V., and Undeen, A.H. (1994) Phylogenetic Relationships among *Vairimorpha* and *Nosema* Species (Microspora) Based on Ribosomal RNA Sequence Data. *J Invertebr Pathol* **64**: 100-106.
- Bass, D., Stentiford, G.D., Littlewood, D.T.J., and Hartikainen, H. (2015) Diverse Applications of Environmental DNA Methods in Parasitology. *Trends Parasitol* **31**: 499-513.
- Bista, I., Carvalho, G.R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D. et al. (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat Commun* **8**: 14087.
- Bray, J.R., and Curtis, J.T. (1957) An ordination of upland forest communities of southern Wisconsin. *Ecol Monogr* **27**: 325-349.
- 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. *Appl Environ Microbiol* **75**: 5797–5808.
- Cleary, A.C., and Durbin, E.G. (2016) Unexpected prevalence of parasite 18S rDNA sequences in winter among Antarctic marine protists. *J Plankton Res* **38**: 401–417.
- Corsaro, D., Wylezich, C., Venditti, D., Michel, R., Walochnik, J., and Wegensteiner, R. (2019) Filling gaps in the microsporidian tree: rDNA phylogeny of *Chytridiopsis typographi* (Microsporida: Chytridiopsida). *Parasitol Res* **118**: 169-180.
- Costello, M.J. (2016) Parasite rates of discovery, global species richness and host specificity. *Integr Comp Biol* **56**: 588-599.
- Debroas, D., Hugoni, M., and Domaizon, I. (2015) Evidence for an active rare biosphere within freshwater protists community. *Mol Ecol* **24**: 1236-1247.

482 Debroas, D., Domaizon, I., Humbert, J.F., Jardillier, L., Lepere, C., Oudart, A., and Taib, N. (2017)
 483 Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available
 484 metabarcoding data. *FEMS Microbiol Ecol* **93**.
 485 Devaux, J. (1980) Contribution à l'étude limnologique du lac Pavin (France). II. Relations entre
 486 paramètres abiotiques - Phyto et zooplancton dans la zone 0-20 mètres. *Hydrobiologia* **68**: 17-
 487 34.
 488 Dobson, A., Lafferty, K.D., Kuris, A.M., Hechinger, R.F., and Jetz, W. (2008) Colloquium paper:
 489 homage to Linnaeus: how many parasites? How many hosts? *Proc Natl Acad Sci U S A* **105**
 490 **Suppl 1**: 11482-11489.
 491 Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves
 492 sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194-2200.
 493 Eren, A.M., Ferris, M.J., and Taylor, C.M. (2011) A framework for analysis of metagenomic sequencing
 494 data. *Pac Symp Biocomput*: 131-141.
 495 Federhen, S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res* **40**: D136-D143.
 496 Fokin, S.I., Di Giuseppe, G., Erra, F., and Dini, F. (2008) *Euplotespora binucleata* n. gen., n. sp.
 497 (Protozoa: Microsporidia), a parasite infecting the hypotrichous ciliate *Euplotes woodruffi*, with
 498 observations on microsporidian infections in ciliophora. *J Eukaryot Microbiol* **55**: 214-228.
 499 Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L. et al. (2013) The Protist Ribosomal
 500 Reference database (PR²): a catalog of unicellular eukaryote small sub-unit rRNA sequences
 501 with curated taxonomy. *Nucleic Acids Res* **41**: D597–D604.
 502 Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M., and Troedsson, C. (2014)
 503 Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers.
 504 *PLoS ONE* **9**: e87624.
 505 Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D. et al. (2014)
 506 Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in
 507 Microbial Consortia. *PLoS ONE* **9**: e95567.
 508 Ironside, J.E., and Wilkinson, T.J. (2018) Accumulation and exchange of parasites during adaptive
 509 radiation in an ancient lake. *Int J Parasitol* **48**: 297-307.

510 Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of
511 multiple sequence alignment. *Nucleic Acids Res* **33**: 511-518.

512 Keeling, P.J., and Fast, N. (2002) MICROSPORIDIA : Biology and Evolution of Highly Reduced
513 Intracellular Parasites. *Annu Rev Microbiol* **56**: 93–116.

514 Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis
515 Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**: 1870-1874.

516 Lafferty, K.D., Dobson, A.P., and Kuris, A.M. (2006) Parasites dominate food web links. *Proc Natl*
517 *Acad Sci U S A* **103**: 11211-11216.

518 Lafferty, K.D., Allesina, S., Arim, M., Briggs, C.J., De Leo, G., Dobson, A.P. et al. (2008) Parasites in
519 food webs: the ultimate missing links. *Ecol Lett* **11**: 533-546.

520 Lair, N. (1978) Répartition spatio-temporelle, biomasse et production des populations zooplanctoniques
521 du lac d'Aydat en période estivale. *Hydrobiologia* **61**: 237-256.

522 Lair, N. (1990) Effects of invertebrate predation on the seasonal succession of a zooplankton
523 community: a two year study in Lake Aydat, France. *Hydrobiologia* **198**: 1-12.

524 Larsen, B.B., Miller, E.C., Rhodes, M.K., and Wiens, J.J. (2017) Inordinate Fondness Multiplied and
525 Redistributed: the Number of Species on Earth and the New Pie of Life. *Q Rev Biol* **92**: 229-
526 265.

527 Lepère, C., Masquelier, S., Mangot, J.F., Debroas, D., and Domaizon, I. (2010). Vertical structure of
528 small eukaryotes in three lakes that differ by their trophic status: a quantitative approach. *Isme*
529 *J* **4**, 1509-1519.

530 Lepère, C., Domaizon, I., and Debroas, D. (2016) Diversity and biogeography of picoeukaryotes: new
531 insights into the rare biosphere. In *Lake Pavin*. Sime-Ngando, T., Boivin, P., Chapron, E.,
532 Jezequel, D., and Meybeck, M. (eds): Springer, Cham.

533 Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and
534 annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242-245.

535 Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E. et al. (2017) Parasites dominate
536 hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol* **1**: 91.

537 Mahé, S., Duhamel, M., Calvez, T.L., Guillot, L., Sarbu, L., Bretaudeau, A. et al. (2012) PHYMYCO-
538 DB: A Curated Database for Analyses of Fungal Diversity and Evolution. *PLoS ONE* **7**: e43117.

539 Marande, W., Lopez-Garcia, P., and Moreira, D. (2009) Eukaryotic diversity and phylogeny using
540 small- and large-subunit ribosomal RNA genes from environmental samples. *Environ Microbiol*
541 **11**: 3179-3188.

542 McMurdie, P.J., and Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis
543 and graphics of microbiome census data. *PLoS ONE* **8**: e61217.

544 Mikhailov, K.V., Simdyanov, T.G., and Aleoshin, V.V. (2016) Genomic survey of a hyperparasitic
545 microsporidian *Amphiamblys* sp. (Metchnikovellidae). *Genome Biol Evol* **9**: 454-467.

546 Oksanen, F.J., et al., Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P. et al. (2017)
547 Vegan: Community Ecology Package. R package Version 2.4-3. In, pp. [https://CRAN.R-](https://CRAN.R-project.org/package=vegan)
548 [project.org/package=vegan](https://CRAN.R-project.org/package=vegan).

549 Paliy, O., and Shankar, V. (2016) Application of multivariate statistical techniques in microbial ecology.
550 *Mol Ecol* **25**: 1032-1057.

551 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees
552 for large alignments. *PLoS One* **5**: e9490.

553 Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007)
554 SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA
555 sequence data compatible with ARB. *Nucleic Acids Res.* **35**: 7188–7196.

556 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal
557 RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*
558 **41**: D590–D596.

559 Quiblier-Lloberas, C., Bourdier, G., Amblard, C., and Pepin, D. (1996) Impact of grazing on
560 phytoplankton in Lake Pavin (France): Contribution of different zooplankton groups. *J Plankton*
561 *Res* **18**: 305-322.

562 Quiles, A., Bacela-Spychalska, K., Teixeira, M., Lambin, N., Grabowski, M., Rigaud, T., and Wattier,
563 R.A. (2019) Microsporidian infections in the species complex *Gammarus roeselii* (Amphipoda)

over its geographical range: evidence for both host-parasite co-diversification and recent host shifts. *Parasit Vectors* **12**: 327.

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. In: Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at <http://www.R-project.org/>.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Smith, J.E. (2009) The ecology and evolution of microsporidian parasites. *Parasitology* **136**: 1901-1914.

Stentiford, G.D., Feist, S.W., Stone, D.M., Bateman, K.S., and Dunn, A.M. (2013a) Microsporidia: diverse, dynamic, and emergent pathogens in aquatic systems. *Trends Parasitol* **29**: 567-578.

Stentiford, G.D., Bateman, K.S., Feist, S.W., Chambers, E., and Stone, D.M. (2013b) Plastic parasites: extreme dimorphism creates a taxonomic conundrum in the phylum Microsporidia. *Int J Parasitol* **43**: 339-352.

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.

Stentiford, G.D., Becnel, J.J., Weiss, L.M., Keeling, P.J., Didier, E.S., Williams, B.A. et al. (2016) Microsporidia - Emergent Pathogens in the Global Food Chain. *Trends Parasitol* **32**: 336-348.

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., Huang, W.-F., Solter, L.F., Malysh, J.M., Becnel, J.J., and Vossbrinck, C.R. (2020) A formal redefinition of the genera *Nosema* and *Vairimorpha* (Microsporidia: Nosematidae) and reassignment of species based on molecular phylogenetics. *J Invertebr Pathol* **169**: 107279.

591 Trzebný, A., Slodkiewicz-Kowalska, A., Becnel, J.J., Sanscrainte, N., and Dabert, M. (2020) A new
592 method of metabarcoding Microsporidia and their hosts reveals high levels of microsporidian
593 infections in mosquitoes (Culicidae) *Mol Ecol Resour* **00**: 1-19.

594 Vavra, J., and Lukes, J. (2013) Microsporidia and 'the art of living together'. *Adv Parasitol* **82**: 253-319.

595 Vossbrinck, C.R., and Debrunner-Vossbrinck, B.A. (2005) Molecular phylogeny of the Microsporidia:
596 ecological, ultrastructural and taxonomic considerations. *Folia Parasitol* **52**: 131-142.

597 Vossbrinck, C.R., Debrunner-Vossbrinck, B.A., and Weiss, L.M. (2014) Phylogeny of the
598 Microsporidia. In *Microsporidia: Pathogens of Opportunity, First Edition*. Weiss, L.M., and
599 Becnel, J.J. (eds).

600 Vossbrinck, C.R., Andreadis, T.G., Vavra, J., and Becnel, J.J. (2004) Molecular phylogeny and
601 evolution of mosquito parasitic Microsporidia (Microsporidia: Amblyosporidae). *J Eukaryot*
602 *Microbiol* **51**: 88-95.

603 Wadi, L., and Reinke, A.W. (2020) Evolution of microsporidia: An extremely successful group of
604 eukaryotic intracellular parasites. *Plos Pathog* **16**: e1008276.

605 Weigl, S., Korner, H., Petrusek, A., Seda, J., and Wolinska, J. (2012) Natural distribution and co-
606 infection patterns of microsporidia parasites in the *Daphnia longispina* complex. *Parasitology*
607 **139**: 870-880.

608 Weiss, L.M., and Becnel, J.J. (2014) *Microsporidia: Pathogens of Opportunity*: John Wiley and Sons.

609 Weiss, L.M., Zhu, X., Cali, A., Tanowitz, H.B., and Wittner, M. (1994) Utility of microsporidian rRNA
610 in diagnosis and phylogeny: a review. *Folia Parasitol* **41**: 81-90.

611 Wijayawardene, N.N., Pawłowska, J., Letcher, P.M., Kirk, P.M., Humber, R.A., Schüßler, A. et al.
612 (2018) Notes for genera: basal clades of Fungi (including Aphelidiomycota,
613 Basidiobolomycota, Blastocladiomycota, Calcarisporiellomycota, Caulochytriomycota,
614 Chytridiomycota, Entomophthoromycota, Glomeromycota, Kickxellomycota,
615 Monoblepharomycota, Mortierellomycota, Mucoromycota, Neocallimastigomycota,
616 Olpidiomycota, Rozellomycota and Zoopagomycota). *Fungal Divers* **92**: 43-129.

617 Williams, B.A.P., Hamilton, K.M., Jones, M.D., and Bass, D. (2018) Group-specific environmental
618 sequencing reveals high levels of ecological heterogeneity across the microsporidian radiation.
619 *Environ Microbiol Rep* **10**: 328-336.

620 Wolska, M., and Mazurkiewicz-Zapalowicz, K. (2013) Parasites of zooplankton and periphyton
621 assemblages in the littoral zone of lakes in Drawa National Park, Poland. *Acta Mycol* **48**: 51-59.

622 Yakovleva, Y., Nassonova, E., Lebedeva, N., Lanzoni, O., Petroni, G., Potekhin, A., and Sabaneyeva,
623 E. (2020) The first case of microsporidiosis in *Paramecium*. *Parasitology* **147**: 957-971.

624

FIGURE LEGENDS

Figure 1: Microsporidian phylogenetic tree based on full length SSU rDNA of 265 microsporidian species and six outgroup taxa (1127 positions). Values above branches indicate the ML and Neighbour Joining bootstrap supports (the same topology has been obtained, at least at the clade and group levels). Bootstraps below the group level are not shown to lighten the figure. Nine clades (highlighted with background colours, and Latin numbers I to IX) are identified for canonical Microsporidia and one for unusual Microsporidia (Metchnikovellidae). Correspondence between different clade numbering systems is provided for the four major clades recovered in all phylogenetic analyses (Latin numbers: this study and Vossbrinck and Debrunner-Vossbrinck, 2005; Arabic numbers: Vossbrinck et al., 2014). Within these clades, 1 to 5 groups are identified (highlighted with branches and species names colours), for a total of 22 groups. References to other studies that support these groups are provided, and listed in Bibliography for Supplementary Information. Colour strips indicate hosts (different colours can be given for a single species when different hosts are known) and habitat (different colours can be given for a single species when different habitats are known). Tree annotated using the interactive Tree of Life (iTOL) server (Letunic & Bork, 2016). The alignment used to build the phylogenetic tree is provided in Supplementary Information 3.

Figure 2: Krona chart representing the affiliation of the 1531 OTUs found in the analysis. Inner circles represent clades, and deeper taxonomic ranks (groups and genera) are presented in outer circles. Group names are given in capital letters, and genera names in italic. For identified genera, habitats are shown with coloured squares and host species are given in brackets. Seven clades are represented, although most OTUs fall in clades I and IV2. A large proportion of OTUs are unidentified at the group (stripe patterns) or genus (dot patterns) level. Among OTUs identified at the genera level, most are assigned to genera known to parasitize freshwater hosts.

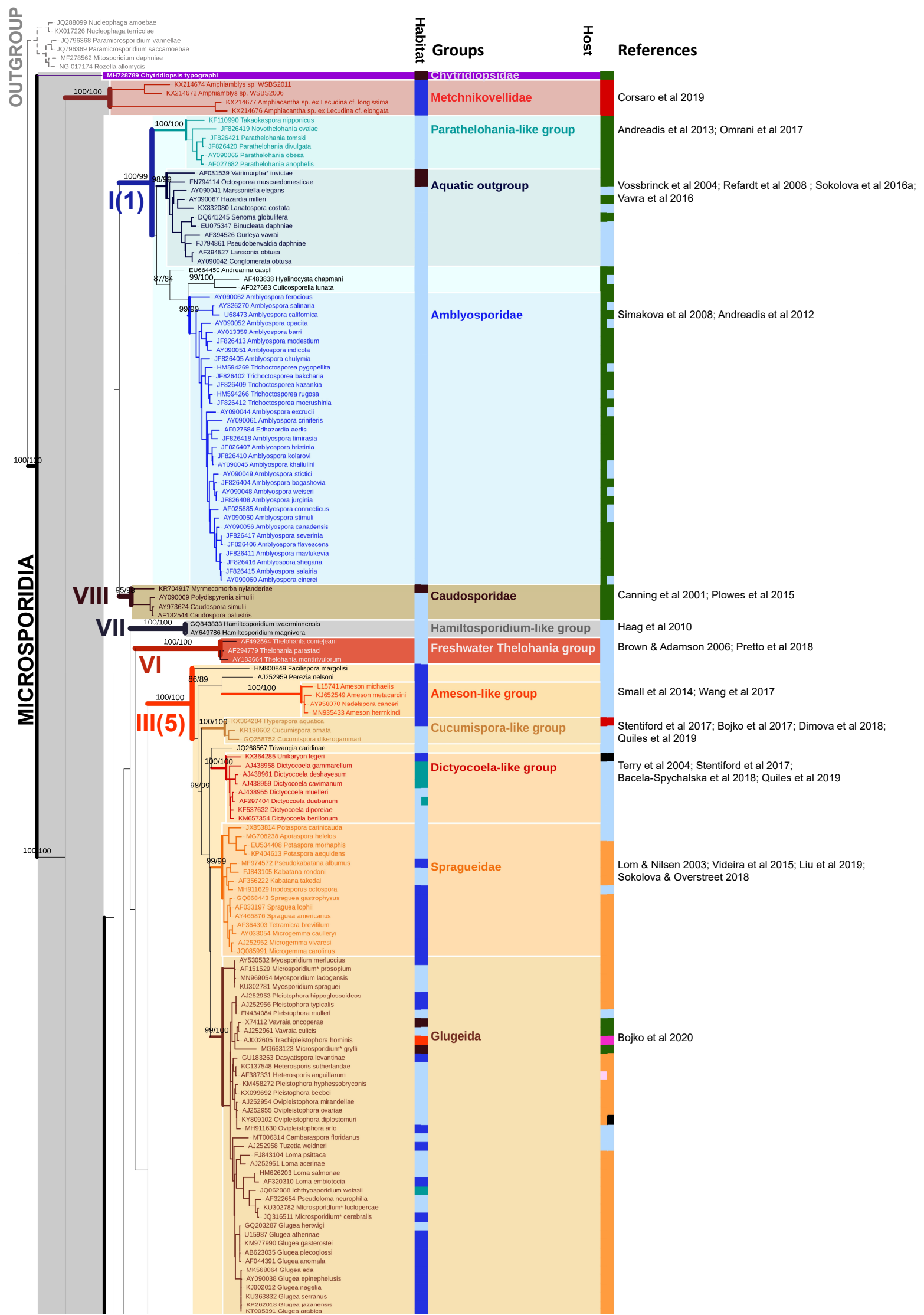
Figure 3: Maximum Likelihood phylogeny of abundant OTUs (>1% of reads in at least one sample) affiliated to the Arlequin group (clade IV2) with other sequences belonging to this group.

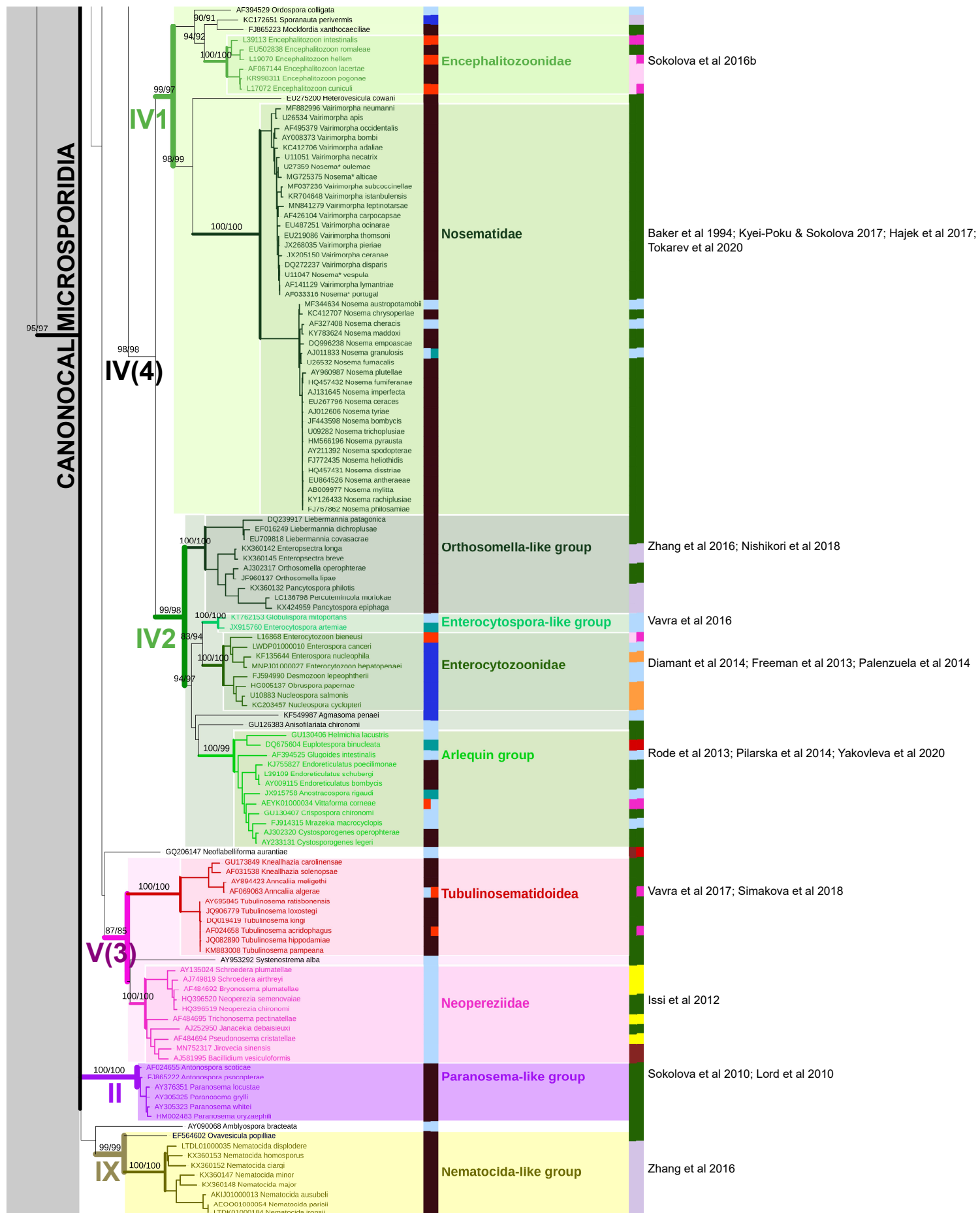
Numbers next to branches indicate the ML bootstrap support (200 replicates). Node supports <70 are not shown. Outgroup species are representative species of other groups belonging to clade IV2 (details in Supplementary Information 15). Affiliation of each OTU following the bioinformatics analysis is given in grey text. Described species are given in bold-italic, and sequences from undescribed species in normal text. Hosts are given in brackets. C: Crustacea; I: Insecta. Relative abundances of each OTU across samples are reported with a heatmap.

Figure 4: Maximum Likelihood phylogeny of abundant OTUs (>1% of reads in at last one sample) not assigned to any group, within clade IV2, with other sequences belonging to this clade.

Numbers next to branches indicate the ML bootstrap support (200 replicates). Node supports <70 are not shown. Outgroup species are representative species of clade IV1 (details in Supplementary Information 15). Affiliation of each OTU following the bioinformatics analysis is given in grey text. Described species are given in bold-italic, and sequences from undescribed species in normal text. Hosts are given in brackets. C: Crustacea; I: Insecta. Relative abundances of each OTU across samples are reported with a heatmap.

Figure 5: A. Non-metric multidimensional scaling (NMDS) ordination of the microsporidian OTU relative abundances across samples based on Bray–Curtis distances (stress: 0.09). Ellipses in the plots denote 95% confidence intervals, points that are not overlapped by the two ellipses are considered significantly different. B. Relative abundance of identified microsporidian taxa across samples.



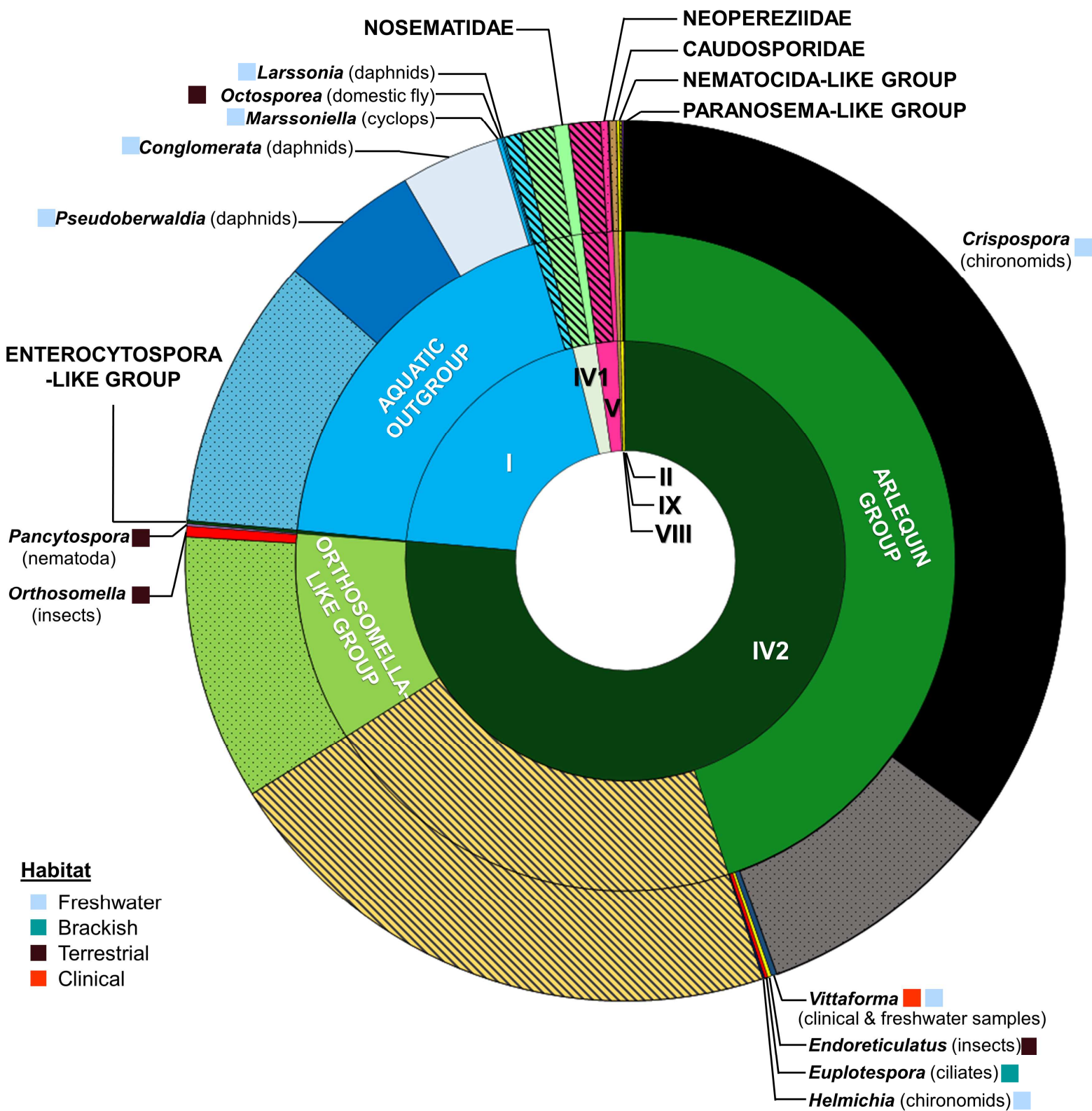


Habitat

- Freshwater
- Brackish or saltern
- Marine
- Terrestrial
- Clinical

Hosts

- Crustacea
- Insects
- Fishes
- Homo sapiens*
- Other vertebrates
- Annelida
- Bryozoa
- Trematoda
- Nematoda
- Unicellular hosts



Aydat

Pavin



