CellPress
OPEN ACCESS

Trends in
Parasitology

Review

Microsporidia: a new taxonomic, evolutionary, and ecological synthesis

Jamie Bojko [1,2,*,@] Aaron W. Reinke [3] Grant D. Stentiford [4,5] Bryony Williams [5] Martin S.J. Rogers [6] and David Bass [4,5,7]

Microsporidian diversity is vast. There is a renewed drive to understand how microsporidian pathological, genomic, and ecological traits relate to their phylogeny. We comprehensively sample and phylogenetically analyse 125 microsporidian genera for which sequence data are available. Comparing these results with existing phylogenomic analyses, we suggest an updated taxonomic framework to replace the inconsistent clade numbering system, using informal taxonomic names: Glugeida (previously clades 5/3), Nosematida (4a), Enterocytozoonida (4b), Amblyosporida (3/5), Neopereziida (1), and Ovavesiculida (2). Cellular, parasitological, and ecological traits for 281 well-defined species are compared with identify clade-specific patterns across long-branch Microsporidia. We suggest that future taxonomic circumscriptions of Microsporidia should involve additional markers (SSU/ITS/LSU), and that a comprehensive suite of phenotypic and ecological traits help to predict broad microsporidian functional and lineage diversity.

Highlights

Microsporidian systematics has entered a genomic era, with ~38 species' genomes available to date, and this number is expected to double by the end of 2022.

Minimal frameworks can unite microsporidian taxonomy, promoting global access to high-quality genetic/pathological data. We propose sequencing of the internal transcribed spacer (ITS), large-subunit (LSU) region, and small-subunit ribosomal RNA gene (SSU), in lieu of genomic data, when describing microsporidian taxa.

Ecological parameters across formally described Microsporidia are synthesised, revealing astonishing ecological diversity, which is mapped to their phylogenetic clade-based higher taxonomy.

We coalesce previous microsporidian taxonomic classifications with new-age clade-based systematics to draw parallels between classical taxonomic approaches for other organisms and the Microsporidia.

Taxonomic and evolutionary history across the Microsporidia

The **Microsporidia** (see Glossary) are a group of human-, animal- and microeukaryote-infecting, obligate, spore-forming parasites, whose systematic framework has been in flux over the past century [1]. Historically, the group has had a morphology/ecology-based taxonomy whereby their complex intracellular life cycle, unique morphological features, and host range (including tissue tropism) were used to provide taxonomic insight [2]. These features remain important, but molecular and genomic technologies are rapidly providing evidence to revise microsporidian systematics and ecological affiliations, becoming the gold standard for species identification and broader phylogenetic placement [1,3]. With these tools, we are beginning to unravel a more complete picture of microsporidian diversity and the role of microsporidians in ecological systems, including the indirect impacts of infection on host populations and their ecosystem services [4,5].

The Microsporidia are classified within the **Opisthosporidia** (Eukaryota: Opisthokonta) [6,7]. Early work using the **small-subunit** (**SSU**) rRNA gene for a large number of species to determine microsporidian phylogenies identified three environmentally defined groups (Aquasporidia, Marinosporidia, and Terresporidia), which were originally classified into five genetically distinct clades, sometimes referenced using Roman numerals (I, II, III, IV, V) and sometimes with Arabic numerals (1, 2, 3, 4, 5) [8] – henceforth, we use Arabic numerals when referring to the clade-based taxonomy. Recently, multiple phylogenetic studies involving the long- and short-branch Microsporidia (including 'Cryptomycota') suggested an alternative configuration of clade numbers [9,10], supported additional smaller clades or 'orphan' lineages [11,12], and presented a somewhat different configuration of the main five well-supported 'clades' (1, 3, 4a, 4b, 5) [12]. Within these clades are multiple microsporidian orders, containing ~45 families, ~218 genera, and an estimated ~1600 species [1]. DNA sequence data are available for 125 (~55%) of the known genera. **Metabarcoding**, **metagenomic**, and other deposited genetic data suggest a

[1]National Horizons Centre, Teesside University, Darlington, DL1 1HG, UK
[2]School of Health and Life Sciences, Teesside University, Middlesbrough, TS1 3BX, UK
[3]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[4]International Centre of Excellence for Aquatic Animal Health, Centre for Environment, Fisheries, and Aquaculture Science, Barrack Road, The Nothe, Weymouth DT4 8UB, UK
[5]Biosciences, University of Exeter, Stocker Road, Exeter, EX4 4HB, UK
[6]Artifical Intelligence Laboratory, British Antarctic Survey, Cambridge, CB3 0ET, UK

CellPress
OPEN ACCESS

much great diversity of microsporidians, predicting thousands of unknown or uncharacterised taxa [5,11,13,14].

The global diversity of microsporidians is only a small part of their story. These parasites play important roles in ecological systems [4,5]. Over evolutionary time they have undergone extensive evolutionary reduction of their genomes and proteomes, resulting in metabolic dependence upon their hosts [15–18]. Their physiologies confer specialised parasitological characteristics, including the potential to remain latent in hosts for many years [19], jumping host species [20], taking advantage of sex, cannibalism, and other host behaviours to transmit [21], persisting in the environment for long periods of time [22], and masking themselves from their hosts' immune defences [23,24]. As pathogens, they pose a significant threat to humans, wildlife, and economically important species [25,26].

In this review, we gathered data from all microsporidian species with a formal taxonomic description that include genetic (partial/full-length SSU rRNA gene), pathological, and ultrastructural data (284 species, 125 genera). We generated a synthesis of multiple physiological and pathological traits and measurements across phylogenetic groups, including host and environmental information, to map microsporidian ecological relationships from across the globe (see Table S1 in the supplemental information online). The data are compared within and between each microsporidian clade to provide an assessment of any clade-scale ecological similarities and provide a discussion on shared traits and putative evolutionary and pathological relationships. We integrated information on evolutionary relationships from recent phylogenomic and SSU rRNA gene phylogenetic analyses, and several recent taxonomic revisions, some informed by intensive microsporidian-targeted environmental sequence diversity studies, to provide a phylogenetically informed name-based taxonomic structure, offering a strong framework for inevitable future discoveries of novel microsporidians infecting hosts from diverse environments.

Taxonomy of canonical microsporidians: past, present, and future

Phylogenetics underpinning microsporidian taxonomy

Our Bayesian SSU rRNA gene analysis includes a well-characterised representative species for all genera for which relevant sequence data exist (Figure 1, Key figure). The Bayesian topology also shows maximum-likelihood (ML) bootstrap values, which together confirm that the canonical Microsporidia comprises several relatively large and strongly supported subclades, and a smaller number of orphan lineages. Also included are sequences from key linages representing previously polyphyletic genera [e.g., *Astathelohania* (= *Thelohania*); Figure 2], whose sequences branched in different clades on the tree (Figure 1 and Table S1), but have recently been redescribed [27]. Such situations require revision by creating new genera for those lineages that do not correspond to the type taxon.

As for many SSU rRNA gene trees, the backbone of the tree is generally poorly resolved. The SSU gene does not provide enough phylogenetic signal to resolve these more ancient divergences. As more genomic datasets are generated for microsporidians, these relationships should become clearer via multigene phylogenomic analyses (Figure 2).

Recent diversity studies and taxonomic revisions have advanced our view of microsporidian systematics [1,5,7,9,11,12,28–31]. These studies have refined the existing clade-numbering system, providing revisions such as the subdivision of clade 4 into 4a and 4b [12]. However, despite these changes, the numbered clade system has become difficult to interpret due to changing phylogenetic tree topologies and the availability of sequenced taxa, compounded by historical inconsistencies in clade numbering (particularly regarding clades 2, 3, and 5;

[7]Department of Life Sciences, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK

*Correspondence: J.Bojko@tees.ac.uk (J. Bojko).
Twitter: @Jamiebojko

Figure 1). We propose the more orthodox and informative use of (informal) taxon names as the basis of microsporidian taxonomy by using order-level names for the major, strongly supported clades (Figure 1 and Table S1). Whether or not these 'orders' are accepted as part of the various taxonomic rankings and nomenclatural schemes currently circulating for Microsporidia is unimportant; what is important is that the group names refer to a key accepted taxonomic entity (i.e., robustly supported clade), and are therefore more informative than an abstract numbering system. The 'orders' we annotate onto Figure 1 are largely concordant with other circumscriptions [28,30], except for some genera that our tree shows to have different phylogenetic affiliations.

We summarise some key findings from recent microsporidian environmental sequencing studies in Figure 1 [5,11,13,14]. Metabarcoding using microsporidian-specific primers on environmental samples is revealing a huge diversity of lineages, much of it related to previously characterised lineages, but also including further lineages that will both expand the size of the known subclades and result in new branches and clades being defined (Figure 1). One drawback to such data is their relatively short amplicon sequences. Robust and informative incorporation of environmentally derived sequences into SSU phylogenies ideally requires (near) full-length SSU sequences rather than short Illumina amplicons. Metabarcoding increasingly employs long-read technologies (e.g., PacBio [32] and Nanopore [33] sequencing), which could be employed for Microsporidia using long-range SSU primers (e.g., CTMicrosp/Microsp1342r combination [34]), or longer amplicons including more taxonomically informative regions (Box 1). Metabarcoding offers insight into microsporidian phylogenetics, taxonomy, and ecology, and we recommend that these data are generated and analysed in robust and consistent ways such that data from different studies are comparable.

Future perspectives on systematics and genomics

SSU sequences are often insufficient to distinguish microsporidian species. For example, an isolate originally identified as a strain of *Nematocida parisii* due to an identical SSU sequence underwent whole-genome sequencing to reveal that it was only 92% identical across the entire genome, leading to reclassification (*Nematocida ironsii*) [24]. In many **microeukaryotic** groups, the SSU region evolves too slowly to be an effective species marker, as demonstrated for example, in oomycetes and paramyxids [32,35].

Other genes besides SSU have been used to help define microsporidian species (Box 1). Further use of reliable marker genes used in addition to the SSU for species discrimination in the Microsporidia is necessary to differentiate between closely related species – guiding our understanding of diversity in environmental sequencing studies [11]. Although additional sequenced genomes will enable a more refined and informative means of distinguishing species, a faster, less expensive, and higher throughput approach is required for rapid assessment of species identity in novel and mixed infections, and for environmental data. The most common alternative marker so far has been the **internal transcribed spacer** (**ITS**) of the ribosomal RNA gene array, among others (Box 1).

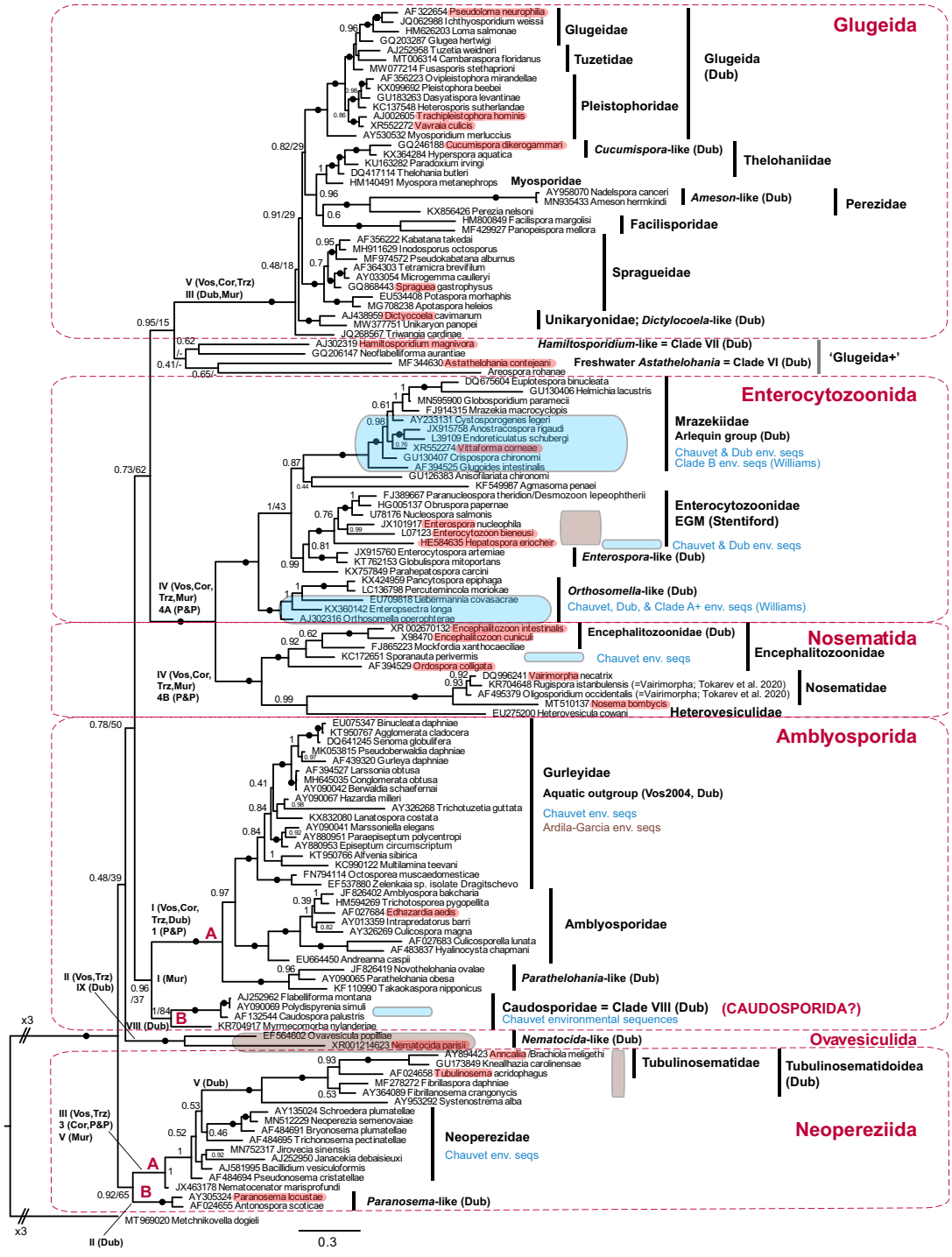
Genomic sequencing is a prerequisite for addressing questions about microsporidian evolutionary history. In addition to clarifying evolutionary relationships, and potentially resolving the backbone of the microsporidian tree, genome sequence data have proven essential for understanding genomic reduction and specialisation, cell evolution, and functionality (e.g., evolution of virulence) [36]. There are 38 microsporidian species genome assemblies available; however, this taxon sampling needs to be much greater to be considered reliable and useful considering the 284 species adequately described to date (Table S1).

Glossary

Internal transcribed spacer (ITS): a genetic region located between the SSU and LSU genes.
Large-subunit (LSU) ribosomal RNA gene: a gene which encodes the RNA that forms the main component of the large subunit of the ribosome.
Metabarcoding: the use of PCR with a general/specific primer set to sequence a short genomic region of a species or multiple species from an environmental sample.
Metagenomics: preparation of a next-generation shotgun sequencing library from the total DNA of a sample to sequence a subset of the entire genetic component of an environmental sample or infected tissue sample.
Microeukaryotic: single-celled, colonial, or syncytial eukaryotic organisms, excluding Metazoa.
Microsporidia: a clade of highly divergent fungi that includes obligate, intracellular, spore-forming parasites.
Opisthosporidia: a proposed high-level taxonomic grouping of Microsporidia, aphelids, and rozellids.
Small-subunit (SSU) ribosomal RNA gene: a gene which encodes the RNA that forms the main component of the small subunit of the ribosome.

Key figure

A Bayesian small-subunit (SSU) rRNA phylogenetic tree of Microsporidia, rooted on *Metchnikovella*



Trends in Parasitology

(See figure legend at the bottom of the next page.)

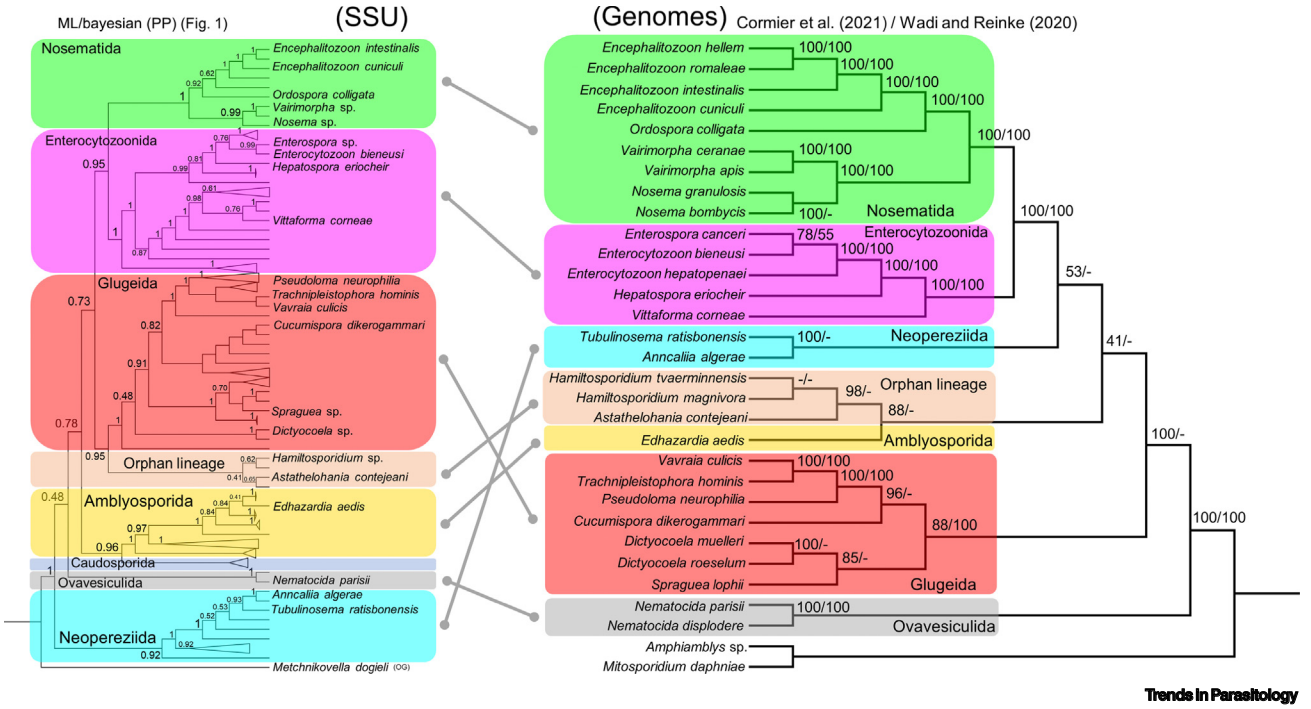


Figure 2. A comparison between the consensus genome tree topology from the most recent/comprehensively sampled phylogenomic analyses to our phylogenetic tree displayed in Figure 1 in the main text. Both trees are represented as cladograms in the figure, sporting a colour-coded system, which is labelled for the higher taxonomic group in which the members sit. The topology of the orthogroup cladogram was developed using Orthofinder (left) [63]; however, the support is based on existing genomic studies (right) [3,54]. The topology shows some complementarity; however, the position of some groups (Ovavesiculida, Glugeida, Amblyosporida, 'orphan lineage', and Neopereziida) branch differently relative to each other. The named clades as defined in this review are robustly recovered by both approaches, in some cases with stronger support in the phylogenomic analyses. Abbreviations: ML, maximum likelihood; SSU, small subunit.

Whole-genome sequences can also address diversity within clusters of closely related species and species complexes. For example, genomes for 12 isolates of *Spraguea*, which are nearly identical at the SSU level, identified variation that correlated with different host species from geographically distinct areas, and distinct spore morphologies and nuclear organisation [37]. Such discoveries raise important questions about our current understanding of microsporidian diversity in ongoing ecological studies (Box 2). Increasingly inexpensive sequencing, and improvements in DNA input requirements and bioinformatic techniques, will likely make whole-genome sequencing of microsporidian species a common approach to resolve questions relating to phylogeny and diversity.

Figure 1. Sequence alignments (MAFFT: G-ins-i algorithm) [59] were masked using TrimAL (–gt 0.5 –w 1 –st 0.01 –cons 70) [60], leaving 1323 alignment positions. Maximum-likelihood (ML) analyses were performed using RaxML BlackBox v.8 [61] using a general time reversible (GTR) model with The CAT-F81 model (CAT) approximation. Bayesian analyses were run using MrBayes v3.2.6 [62]. Two separate MC3 algorithm (MC^3) runs with randomly generated starting trees were carried out for five million generations each with one cold and three heated chains. The evolutionary model included a GTR substitution matrix, a four-category autocorrelated gamma correction, and the covarion model. Bayesian Posterior Probabilities (BPP) are indicated on all but the most terminal nodes. ML bootstrap values are indicated on main nodes. Where both ML bootstrap support >95% and BPP >0.95, black circles are noted. Previous clade numbers are summarised at relevant nodes and to the right of the tree for 'Glugeida+' (an informal label for *Hamiltosporidium*, *Neoflabelliforma*, *Areospora*, and *Astathelohania*). Three-character abbreviations indicate previous numbering systems: Vos = Vossbrinck *et al*. [9], Cor = Corsaro *et al*. [29], Trz = Trzebny *et al*. [31], Dub = Dubuffet *et al*. [11], Mur = Murareanu *et al*. [1], and P&P = Park and Poulin [12]. The names and circumscriptions of the clades proposed in this review are shown in red capitalised text and broken line red boxes. Other family and order level taxa are indicated by vertical black lines, based on the references herein, in consensus with earlier taxonomic treatments [9,30]. Blue- and brown-shaded blocks indicate phylogenetic regions where microsporidian environmental sequence data (from aquatic and terrestrial studies respectively) branch, without showing any of the environmental sequences themselves [5,11,13,14]. Representative genus or species for which genome sequences are available (see Figure 2 in the main text) are highlighted using red shading. Abbreviation: MAFFT, multiple alignment using Fast Fourier Transform.

Box 1. Molecular markers: a 'minimal framework' for microsporidian systematics

Microsporidian taxonomy is moving towards a gold standard, which includes partial/complete genome sequencing and annotation, intracellular development and life cycle information, parasite ultrastructure, and host–parasite pathology. Strains/species can be determined without some of the above; however, we must define a minimal framework to unite ongoing efforts to catalogue microsporidian diversity, and importantly, their associated virulence and health-impacts, which inform epidemiological models and predictive emergence studies.

Microsporidiologists have begun to collect sequence data for the ITS region situated between the SSU and **large-subunit** (**LSU**) ribosomal RNA genes, among other genes [68]. The *Operophtera* [69], *Heterosporis* [70], *Nosema* [71], *Encephalitozoon* [72], and *Enterocytozoon* [73] have sequence data for the LSU region, used to produce more detailed phylogenetic depictions of strains and species complexes. This is especially pertinent for genera where SSU similarity is >99%, such as the *Nosema*, which are rooted in a deep and confusing taxonomic history [74]. For *Enterocytozoon bieneusi*, >1600 ITS sequences from different isolates revealed ~500 unique genotypes [73]. Several genotypes were found only in certain hosts, suggesting strain-level host specificity [73]. Tokarev *et al*. [74] showed that the RNA polymerase II gene provided additional evolutionarily distinction for *Nosema*. Hatjina *et al*. [75] used the polar tube protein gene to derive *Nosema* strains. Bateman *et al*. [76] used the RNA polymerase, arginyl tRNA synthetase, prolyl tRNA synthetase, chitin synthase, beta tubulin, and 'heat shock protein 70' genes to show that *Hepatospora eriocheir* was a host-generalist among crab species.

To achieve a minimal framework for microsporidian descriptions, we propose that ultrastructural and developmental data be collected for the parasite using transmission electron microscopy; histology (or wet-prepared tissue) to define affected organs and broader pathology; and finally, sequence the ITS region and partial/complete LSU region in addition to the SSU. The studies listed earlier provide PCR primers for some microsporidian groups, but it is acknowledged that greater long-read diversity or genome work is necessary for continued primer development (Figure I) [77,78].

Given the expanding availability of genomes, there is a capacity to develop PCR methods for the capture of evolutionarily important genes. Developing such diagnostic methods will benefit the proposed taxonomic framework, increasing the capacity for discovery and providing tools for global application via simple PCR methods. Eventually, it seems plausible that we will follow the International Committee for the Taxonomy of Viruses (ICTV), whereby complete genomes may represent formal species identities. Given their small and low-complexity genomes, Microsporidia are prime candidates to test these advancing technologies.

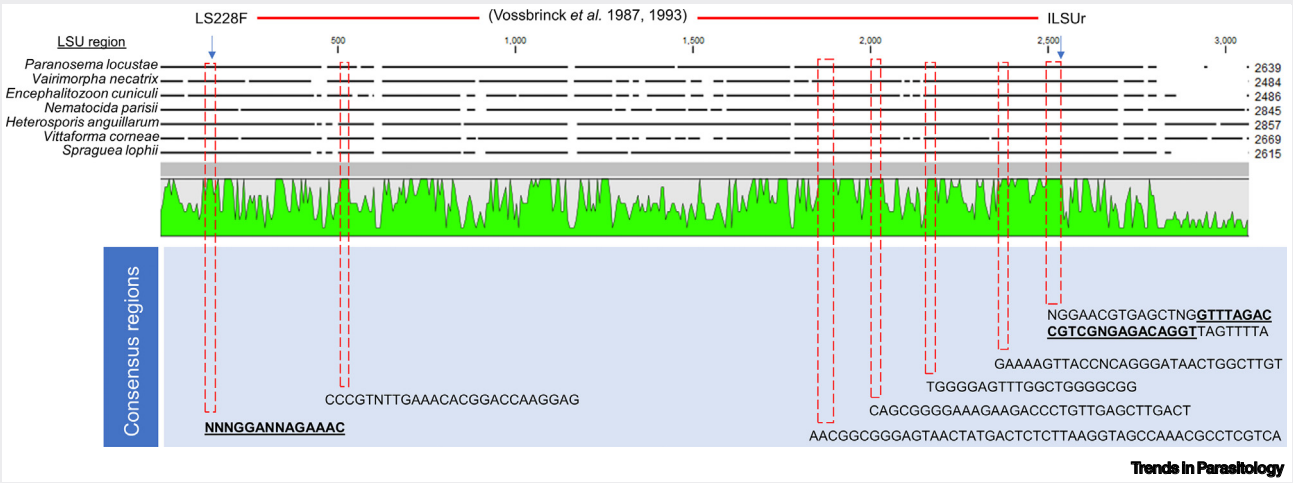


Figure I. A MAFFT alignment of the large subunit rRNA gene of seven microsporidian species. The green line plot indicates regions of high and low conservation. Red broken line boxes accompanied by sequence information highlight regions where primers have been developed [72,78] (bold and underlined) as well as regions where development could be possible. The figure was designed in CLC genomics workbench v.22. See [72,78]. Abbreviations: LSU, large subunit; MAFFT, multiple alignment using Fast Fourier Transform.

Genome sequencing of selected lineages also provides insight into trait evolution in Microsporidia, and diversification of the major clades. As more lineages are characterised phenotypically and phylogenetically it is increasingly clear that generalisations about microsporidian ecology, host interactions, and virulence cannot be made based on the relatively crude resolution offered by SSU rRNA gene trees. Although some patterns (with many exceptions) are apparent, understanding the relatively rapid evolutionary adaptations and ecological shifts of microsporidians requires a focused comparative genomic approach, rather than a wide-scale phylogenetic one, although there is an informative thread that runs between these two extremes.

Box 2. Microsporidian strains, species, and complexes – ecologically relevant issue

Microsporidia are pervasive in most ecosystems (see Figure 3 in the main text), obligately infecting animal and protistan hosts [1]. However, the host ranges of both known and newly discovered microsporidian species are often incompletely known [36]. Environmental sequencing studies [5,11,13] (see Figure 1 in the main text) can provide information about potential hosts that would be very labour-intensive to gather by traditional host-centric sampling approaches, particularly when combined with molecular visualisation techniques such as fluorescent *in situ* hybridisation (FISH) [5]. Understanding host range is essential to understand microsporidian ecology, including their zoonotic potential, or if the parasite may infect economically important species or wildlife. In some cases, studies use a transmission experiment to determine whether a microsporidian parasite can be passaged from one organism to another, either through feeding or cohabitation [79]. In other cases, observation or molecular diagnosis of a microsporidian is used to explore host range [76,80].

Microsporidian environmental diversity studies have been limited to partial SSU amplicons. However, if the functional diversity of Microsporidia is more realistically delineated by faster-evolving gene regions, such as ITS rDNA, individual environmental SSU-based operational taxonomic units (OTU)/amplicon sequence variants (ASVs) may represent multiple, ecologically distinct, lineages (whether we call them species or not) (see Box 1 in the main text). Species diversity is of broad ecological interest, especially in a changing world, where we need to rapidly understand how much biodiversity is being impacted by global change. Environmental sequencing studies can increase known sequence diversity, even locally, many-fold; however, there remains an important need to increase the information derived from such studies for greater taxonomic and ecological insight.

An emerging area of interest includes the process of trophic transmission across highly divergent host species [34,42,80–82]. The *Ovipleistophora diplostomuri* conundrum involves a sequencing effort to look for this parasite in crustacean and fish hosts. Recently, Stratton *et al*. [82] reported a lineage of *O. diplostomuri* infecting multiple crayfish species, multiple fish species, and a trematode species. Given the recommendations of this review, it is important to gather greater sequence evidence for the several *O. diplostomuri* isolates from these hosts to determine whether this parasite is truly trophically transmissible – or if we are seeing a complex of multiple strains/species that are playing different roles within these global ecosystems and each with a specific host range.

Systematic and ecological relationships among and between microsporidian clades

Using one approach (e.g., genomic/pathological) or perspective (e.g., immunological/ecological) over another, or individual biological characteristic, does not provide enough information to reliably inform upon microsporidian systematics. Instead, an interdisciplinary approach (pathology, physiology, and genetics/genomics) is required to relate microsporidian phylogeny and taxonomy to biological traits. The ecological relevance of these parasites can also be inferred from much of the data needed for systematics and provides a critical insight into the ecological niches occupied by microsporidians. To date, 281 microsporidian species have been adequately characterised for these purposes, that is, with available data for a consistent range of traits and characters (Figure 1, Boxes 1 and 3, and Table S1). This is much lower than the 1000s of species predicted/reported to date [1], since formal taxonomic descriptions require a series of detailed information including genetics, pathology, development, and morphology. Below we propose a revised taxonomic approach that also untangles previous uses of the clade-based naming system, while also integrating ecological, physiological, and pathological traits.

Amblyosporida and 'Caudosporida'

In our analyses Amblyosporida comprised a diverse, maximally supported clade including Gurleyidae, Amblyosporidae, and '*Parathelohania*-like' groups [11], branching with moderate Bayesian support as sister to Caudosporidae (= clade 8 [11]), the latter including the genera: *Caudospora*, *Myrmecomorba*, *Flabelliforma* and *Polydispyrenia* (Figure 1). More data are required to confirm that these four genera group robustly together, and that currently unsampled diversity reinforces rather than weaken the exclusive integrity of this clade. If both these requirements are met, there is a strong case for separately recognising Caudosporida as one of the major microsporidian clades. Sixty-five amblyosporid species have been characterised to the extent that they meet the criteria for inclusion in the analyses presented in this review, excluding five species in the Caudosporida. A genome for *Edhazardia aedis* solely represents the

Box 3. Microsporidian databases – a unified approach to microsporidian epidemiology

There is great importance in providing open access data on microsporidian genetics, genomics, ecological association(s), host range, pathology and other experimental studies. Such models may help to identify epidemiological patterns and become powerful enough to one day predict the next emergence of a microsporidian pathogen, or could be used to explore global connectivity and evolutionary history for microsporidian parasites. Some such databases do exist for the Microsporidia, and include: the National Centre for Biotechnology Information (NCBI); The Eukaryotic Pathogen, Vector and Host Informatics Resource Database [VEuPathDB (including the 'MicrosporidiaDB')] [83]; Microsporidia Epidemiology database (MicroEpiDB) [58]; and the silkworm pathogen database (Silkpathdb) [84].

Each database holds various data collections that can benefit the microsporidian research community. NCBI holds most of the available genetic and genomic data for the Microsporidia and holds a useful taxonomy browser tool that includes most microsporidian species. VEPathDB provides a series of tools and data resources that can be valuable for understanding microsporidian systematics, diversity, and pathology, as a part of the MicropsoridiaDB. To promote a better epidemiological understanding of human, cultured animal, and wildlife microsporidiosis, patterns and trends are stored in the MicroEpiDB, which aims to provide global prevalence data on a range of species, but with a focus on *Encephalitozoon* [58]. Finally, Silkpathdb holds data on the pathogens of several silk-producing lepidopteran species, but also holds additional genome data for microsporidia (primarily *Nosema* and *Vairimorpha*) that are particularly damaging to the silk industry [85].

The MicropsoridiaDB currently holds 54 associated datasets, including genome sequence and annotation data and host–pathogen interactome data, as well as providing access to news and general information articles. The database also provides access to population biology datasets, primarily consisting of SSU metabarcoding sequence data, to better understand microsporidian diversity on a global scale. A greater use of this repository to store behavioural data pertaining to parasitised hosts, phylogenetic alignments, geospatial data, and perhaps ecological/environmental data, could allow the database to become a hub for microsporidiologists. Greater use of these databases, and cross-connectivity between them, could promote an increase in the study of microsporidian epidemiology and genomics.

Amblyosporida [38] and is the largest microsporidian genome identified to date (see Figure S1 in the supplemental information online). Caudosporida is currently unrepresented by genomic data (Figure 2).

Amblyosporida are found primarily in hosts from terrestrial and/or freshwater environments. Most are found in parasitic insect hosts that act as pollinators and/or vectors (i.e., mosquitoes). The host distribution of the group includes wild insects with semi-aquatic or terrestrial life cycles as well as freshwater crustaceans. Some members of *Amblyospora* can transmit between these host groups (Figure 3) [39]. No amblyosporids have been identified from marine hosts, but one species (*Takaokaspora nipponicus*), which exhibits two different morphological states, is present in a mosquito from coastal environments [40].

Most Amblyosporida have been discovered and recorded from the USA and Russia; however, their distribution is likely global, with multiple discoveries from Argentina, Africa, and Australia (Figure 4). Within the more clustered countries of Europe, most discoveries are from aquatic environments and laboratory-reared *Daphnia* spp., rather than mosquitoes.

With respect to pathology, this group has been found to develop in seven different host tissue and organ types, as well as systemically (i.e., more than four different tissues in the same host). The fat body of insects is the predominant site of infection (Figure 5A). A principal component analysis (PCA) accounting for Amblyosporida spore size, volume, and polar filament coils (min/max), places this group centrally, meaning that their spore morphology tends not to lean towards extremes (Figure 5B). Their spore morphology is variable, with some species exhibiting the largest spore volumes from across the Microsporidia, as well as presenting almost all known morphological shapes (cordiform, crescent, ellipsoidal, elongate, ovoid, rod-shaped, and spherical), across environments, described from the group to date (Figure 6A–D). Spore shape and size also vary in both host groups (insects and crustaceans) and the majority appear to transmit both horizontally

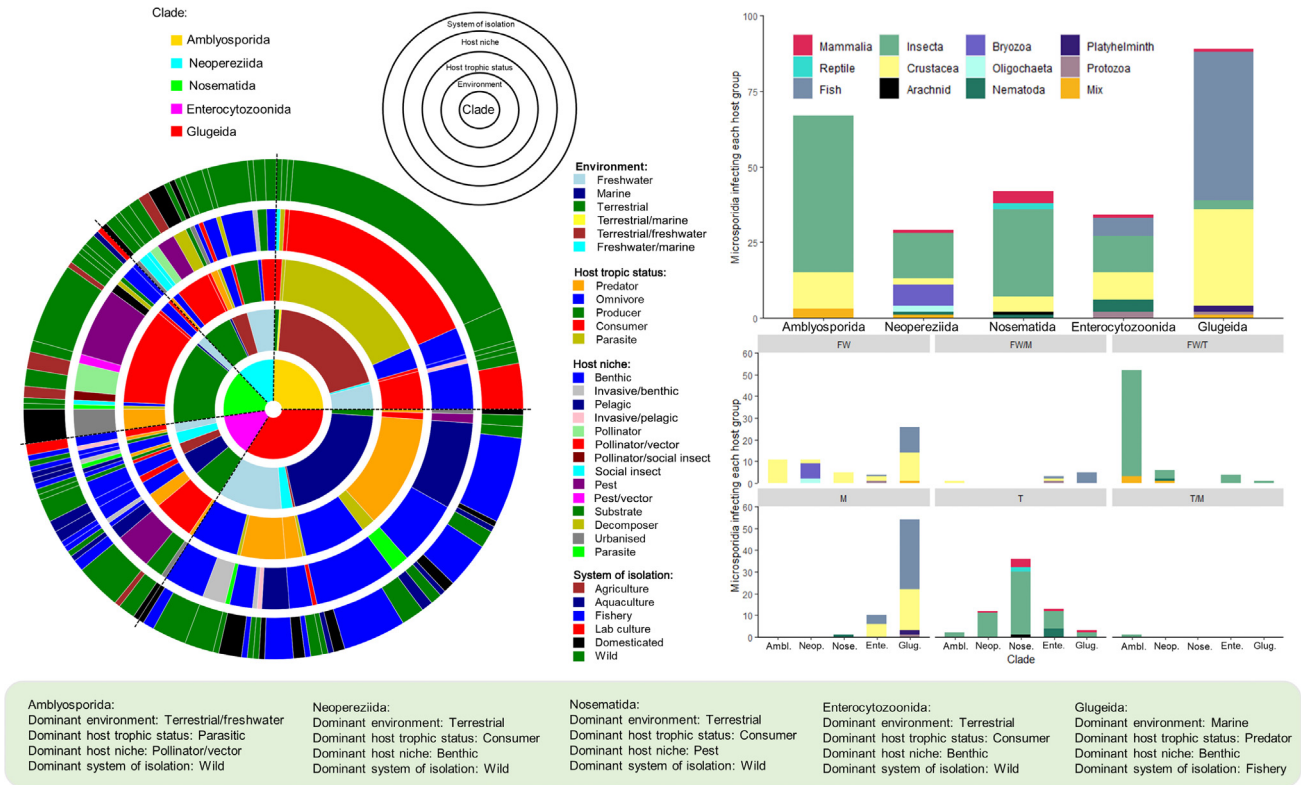


Figure 3. Clade-level comparison of host and ecological traits. The pie chart provides an overview of the environment, host trophic status, host niche, and system in which the microsporidian parasite was isolated using a colour key, with clades represented on the innermost circle. The bar graph plots the host group (represented in colour by the internal key) and frequency. The bar graph is split further, by environment, to provide a more detailed plot using both host group (top right) and environment (M = marine, FW = freshwater, T = terrestrial) (bottom right). The pie chart was developed using Python v3.8 (library: matplotlib [64]) and the bar chart was developed in R v.3.2.2 [65] (library: ggplot2 [66]).

and vertically, or simply horizontally, with only one species transmitting vertically (*Marssoniella elegans*) (Figure 6B,C).

Neopereziida

Neopereziida is represented by 29 characterised species (represented by 17 genera on Figure 1) and is divided into two subclades (Figure 1). Members of Neopereziida are often the earliest diverging branches of long-branch Microsporidia in many SSU phylogenetic analyses (Figure 1). The *Paranosema*-like clade [11] branches with moderate support as sister to a larger clade that has been repeatedly recognised and variably numbered as 3, or 5. This clade comprises Tubulosematidae and robust relatives (labelled clade 5 [11]), and Neoperezidae, which sometimes presents as a sister clade to the latter, but in our Bayesian and ML trees as a paraphyletic assemblage. Due to this ambiguity, we prefer to use the subclade label 'A' to refer to the Neopereziida other than *Paranosema/Antonospora*. Three neopereziid species [*Tubulinosema ratisbonensis*, *Paranosema/Antonospora locustae*, and *Anncaliia* (= *Brachiola*) *algerae*] have sequenced genomes (Figure 2) [38,41–43].

Neopereziid lineages have been isolated from terrestrial and/or freshwater environments, predominantly, with one species from a deep ocean nematode; however, their host trophic status

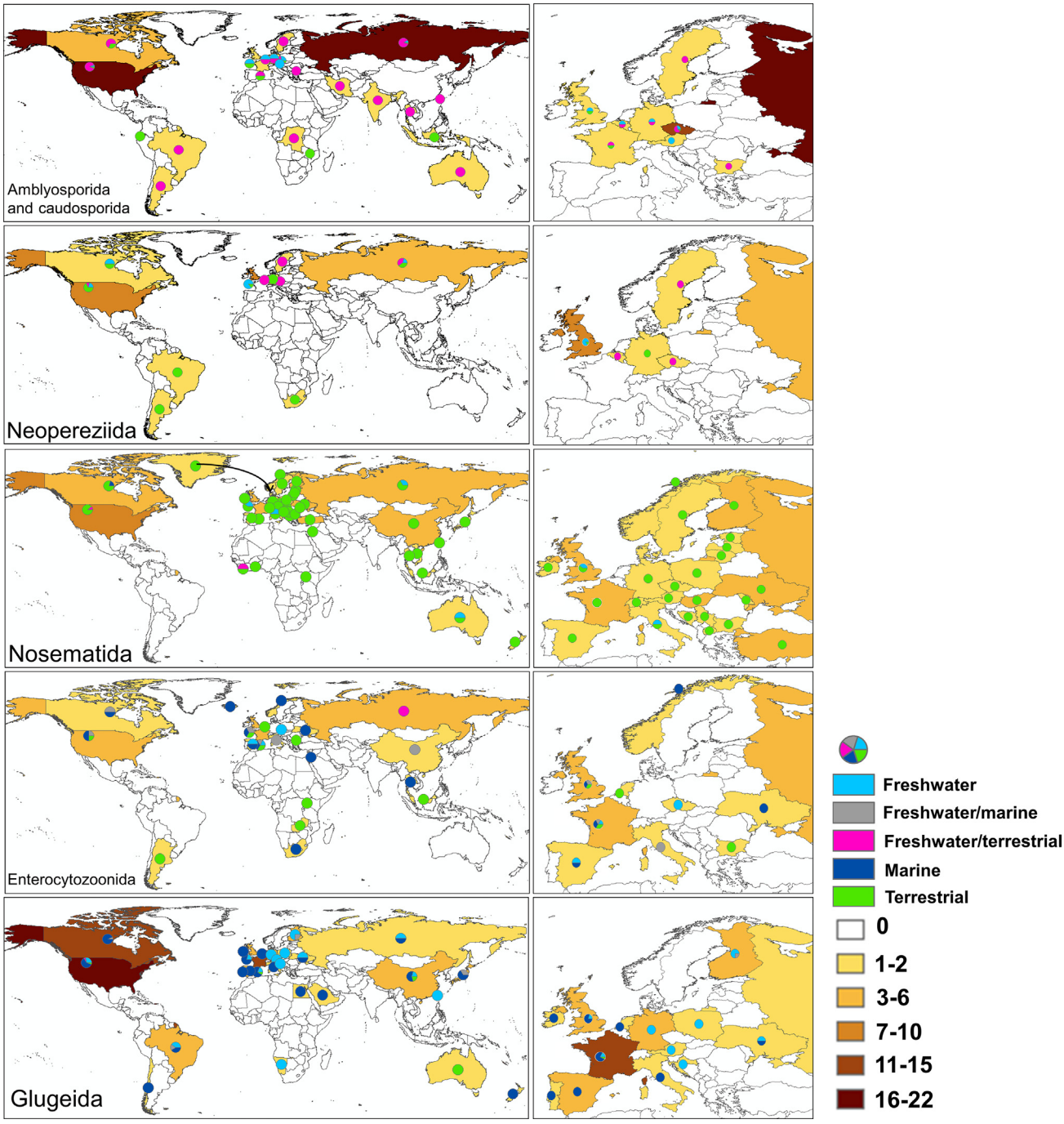


Figure 4. Geographical distribution of microsporidian discoveries, by phylogenetic clade (Amblyosporida and Caudosporida, Neopereziida, Nosematida, Enterocytozoonida, Glugeida). The environment and number of Microsporidia are reflected in the pie chart located on each country with at least one microsporidian discovery. The number of discoveries is reflected in the heat map key, reaching up to 22 novel species finds per clade. The maps include the distribution of 247 microsporidian species, which are published with clear geographical information in available literature. Maps developed and annotated in ArcGIS v.10.4.1.

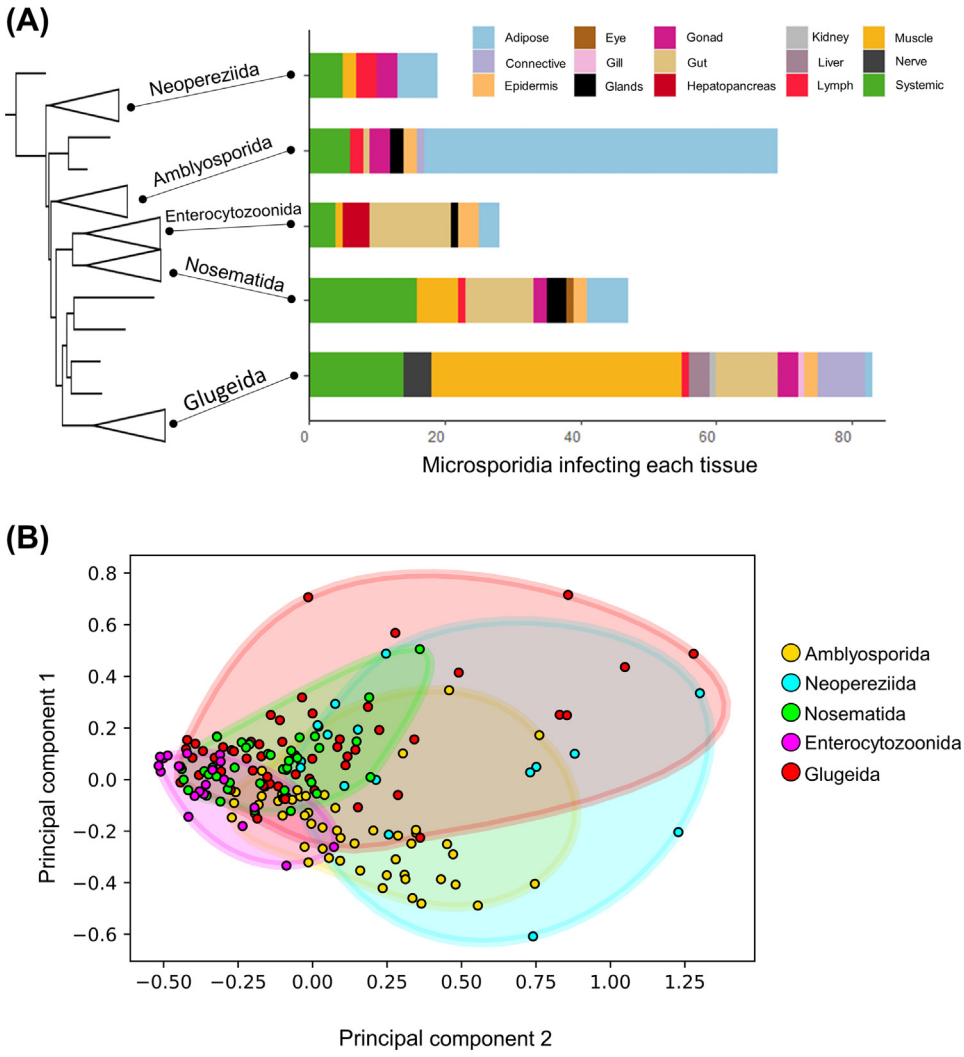


Figure 5. Tissue tropism and spore measurements of mature Microsporidia. (A) Microsporidian tissue tropism plotted by the clade associated with each microsporidian species. The key provides a colour for each tissue/cell type. The stacked bars represent the use of a tissue for microsporidian species in each clade, as well as provide a quantitative capacity to reflect how common the use of certain tissue types is in each clade (see Figure 1 in the main text). Graph developed in R v.3.2.2 [65] (library: ggplot2 [66]). (B) A principal component analysis (PCA) of microsporidian spore length, width, volume, and polar filament minimum and maximum, for each taxonomically categorised microsporidian species with available data for all measurements listed earlier. The colours represent the microsporidian clade (Amblyosporida, Neopereziida, Nosematida, Enterocytozoonida, Glugeida) that the species phylogenetically groups within. Graph developed in Python v.3.8, libraries: 'matplotlib' [64] and 'sklearn' [67].

is variable, including hosts that are predators, omnivores, producers, consumers (predominant), and parasites (Figure 3). These hosts include bryozoans, crustaceans, insects, oligochaetes, and mammals (Figure 3). The niche of their hosts is equally variable, but predominantly includes hosts from the freshwater benthos. This group also infects agriculture pests, social insects, and decomposers, at a lower frequency (Figure 3).

Neopereziida has predominantly been reported from the northern hemisphere, with most discoveries from the UK and USA (Figure 4). Representatives from the southern hemisphere (Brazil,

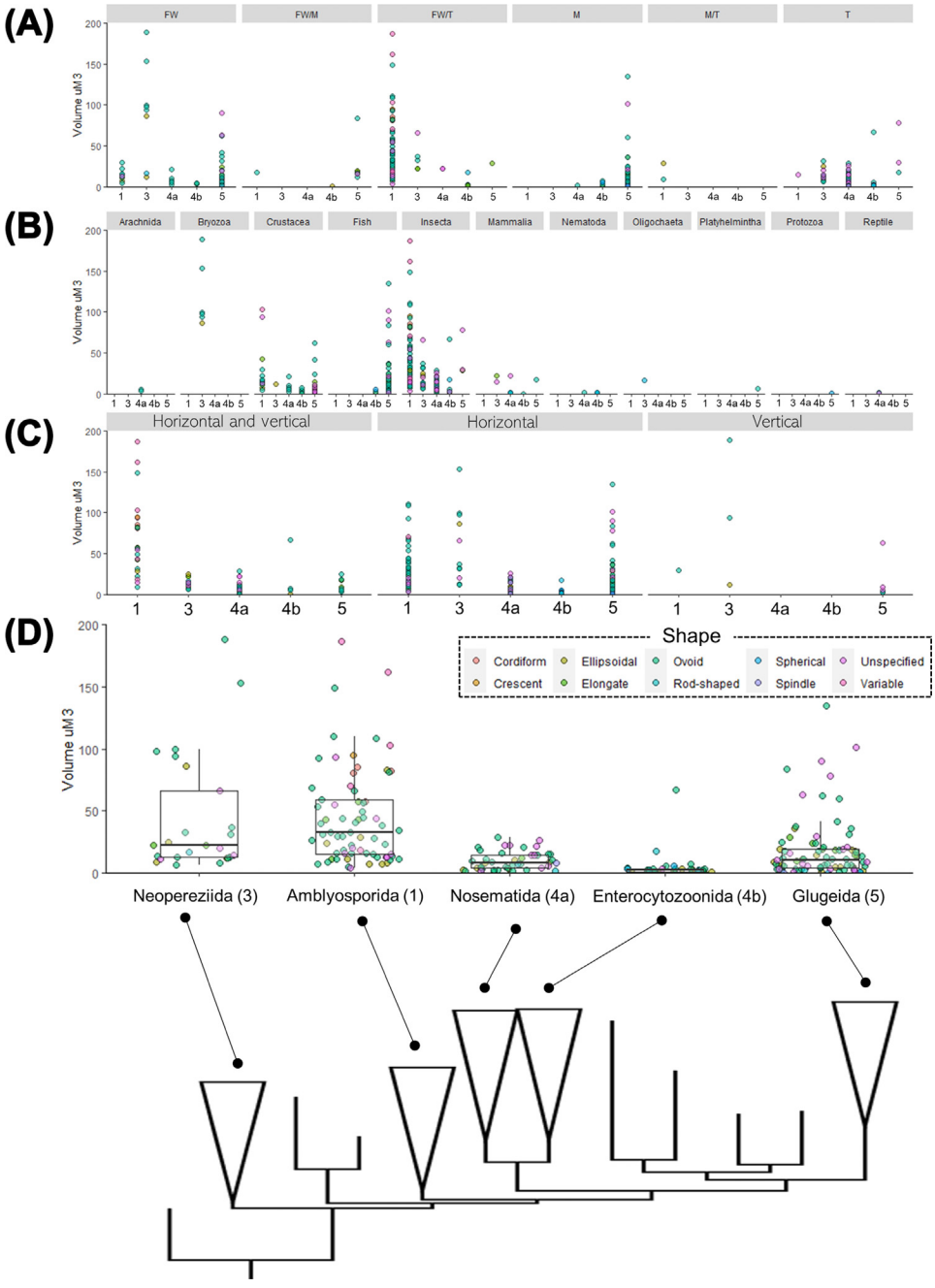


Figure 6. Microsporidian spore volume for each represented species is used as a continuous data element to plot microsporidian clade against other trait factors. (A) Environment (M = marine, FW = freshwater, T = terrestrial). (B) Host group. (C) Transmission (horizontal and vertical; horizontal, vertical). (D) Spore shape. The tree at the base of the graph is a collapsed version of the tree presented in Figure 1 in the main text. The names of the higher taxonomic groups are included on the x axis but are represented by clade numbers on graphs A–C. Graph developed in R v.3.2.2 [65], library: ggplot2 [66].

Argentina, and South Africa) are from terrestrial hosts. This group infects adipose tissues, gonad, lymphoid/blood cells, and muscle, or can infect systemically (Figure 5A). The range of spore measurements of neopereziids is the largest across the Microsporidia (Figure 5B). Spore morphology, transmission method, and volume are also variable, with one species (*Schroedera plumatellae*) having the largest spore volume identified from the Microsporidia to date (Figure 6).

One group within this lineage exemplifies the generalist nature of some microsporidian species: *A. algerae* naturally infects a broad range of mosquito host species [44,45] and the number of insect hosts and cell lines that can be infected by *A. algerae* under laboratory conditions keeps growing [46–48]. This broad host range may in part be facilitated by an ability to survive and grow at a wide range of temperatures [46]. This likely also facilitates *A. algerae* in infecting humans, and this species causes some of the most serious clinical microsporidiosis infections, which penetrate deep into muscle and can be fatal [49,50].

Nosematida

Nosematida (n = 43 species) correspond to clade 4a [12], represented by genomes for 11 species (Figures 1 and 2). Nosematida have been primarily detected in terrestrial hosts, with less representation from marine and freshwater hosts (Figure 3). Hosts include nematodes, arachnids, crustaceans, insects, reptiles, and mammals. Host niche is variable; however, agricultural pest insects occupy the greatest number of hosts (e.g., vast number of *Nosema* and *Vairimorpha* species described to date). Whilst a high proportion of Nosematida are from wild hosts, domesticated hosts and hosts with agricultural importance are also well represented (Figure 3). Geographical distribution is global, due to several species in this group infecting human hosts and honeybees [51,52]; however, in other hosts, the group are found predominantly in the northern hemisphere (primarily the USA) with some cases in Australia, New Zealand, and Africa (Figure 4).

Most infections are systemic, and eight different organ and tissue infection sites have been described (Figure 5A). Spore morphology measurements for Nosematida clusters centrally in the PCA, with some leaning towards the positive extreme of PC1 (Figure 5B). This clade predominantly transmits horizontally (n = 21), and several species can also transmit vertically (n = 15) (Figure 6C).

Enterocytozoonida

Enterocytozoonida (n = 35 species), corresponding to clade 4b [12], derived from the original clade 4 [9], is represented by five species with sequenced genomes (Figures 1 and 2). Enterocytozoonida are distributed among marine, freshwater, and terrestrial environments, and infect producers, consumers, omnivores, and predatory hosts (Figure 3). These include protozoans, nematodes, crustaceans, insects, fish, and mammals. Benthic hosts are commonly infected; however, terrestrial insect pests and nematodes comprise ~one third of known hosts for Enterocytozoonida. Most species have been isolated from animals in anthropogenic systems, such as aquaculture, agriculture, fisheries, and laboratory cultures of model organisms (Figure 3). Members of this group cause high profile infections in human (i.e., *Enterocytozoon bieneusi*) and aquaculture species (i.e., *Enterocytozoon hepatopenaei*, *Hepatospora eriocheir*), resulting in their global distribution, with the majority identified in the northern hemisphere from countries with coastlines (Figure 4) [26].

Enterocytozoonida is the most tissue/organ-specific group; only a low proportion (~15%) appear to cause systemic infections in their host(s) (Figure 5A). Muscle, hepatopancreas, gut (majority), glands, epidermis, and adipose tissues can become infected (Figure 5A). This group has the

smallest average spore size, which associate with the extreme of PC2 (Figures 5B and 6). These smaller spores are particularly associated with aquatic hosts, whilst larger spores from this group are from terrestrial hosts (Figure 6A,B). Enterocytozoonida, like the Nosematida, transmit horizontally and vertically, or horizontally alone, but do not use vertical transmission alone (Figure 6C). Both Enterocytozoonida and Nosematida are maximally supported by Bayesian and ML methods and are similarly maximally supported as mutual sister clades (Figure 1). The Nosematida and Enterocytozoonida clades are moderately to weakly supported as sister to Glugeida.

Glugeida

Glugeida (n = 89 species) represents the largest group of well-described microsporidian species and is represented by seven species with genome sequence data (Figures 1 and 2). Species in this group have been primarily found in aquatic systems (mainly marine), with a small proportion from terrestrial hosts. The group have been found primarily in predatory and omnivorous fish and crustaceans, with some additional isolates from protozoans, platyhelminths, insects, and mammals (Figure 3). Hyperparasitic Microsporidia have been recorded more frequently in Glugeida than other clades, as have parasites that can infect more than two host groups [53]. Most hosts are free-living benthic or pelagic species (Figure 3). Members of this clade have been recorded more often from aquaculture, fisheries (the majority), and domesticated animals or humans, than from wildlife (Figure 3). Glugeida have predominantly been recorded from hosts in the northern hemisphere (USA, Canada, and France); however, their known distribution includes South America, Africa, Australia, and New Zealand (Figure 4).

Glugeida are common in host muscle tissues, especially of crustacean hosts, but the group has been found in 11 different tissue and organ types and can cause systemic infections – particularly in fish (Figure 5A). Some spores of this clade are present at the greatest positive extreme of PC1 and PC2, but overall fall into the central portion of the plot (Figure 5B). The group can transmit horizontally, vertically, or both (Figure 6C). Vertically transmitted spores appear harder to study, with few studies reporting spore size and shape.

Other microsporidian groups

Several microsporidian ‘orphan’ genera group outside of the main SSU-inferred clades and their branching position relative to those clades and other ‘orphan’ lineages is unresolved (Figure 1). The Glugeida clade is strongly supported by SSU Bayesian (but not ML or phylogenomic) analyses as sister to four such genera, which do not branch strongly with each other, or any other group on SSU trees: *Hamiltosporidium* (clade 7 [11]), *Neoflabelliforma*, *Astathelohania* (crayfish-infecting, freshwater) (clade 6 [11]), and *Areospora* (South American crab host). The genomes of *Astathelohania contejeani* and *Hamiltosporidium* spp. have been sequenced and form a maximally supported clade [54] (Figure 2). Genome sequencing of *Neoflabelliforma* and *Areospora* will enable testing of whether all four of these genera group together, as very weakly suggested by the Bayesian SSU analysis.

The SSU (Figure 1) and phylogenomic trees (Figure 2) are concordant in some respects, and importantly all the main clades are recovered by both approaches. As expected, the backbone branching order differs between them. A notable difference is the strong sister relationship between *Edhazardia aedis* (Amblyosporida) and *Hamiltosporidium+Astathelohania* [54]. Analysis of additional Amblyosporida genomes are required to test whether Amblyosporida is sister to *Hamiltosporidium+Astathelohania* [54], a relationship not indicated by the SSU analyses.

Ovavesicula popilliae and *Nematocida* spp. (Ovavesiculida – originally clade 2 [8]; also referred to as clade 9 [11]) diverge before Amblyosporida, Nosematida, Enterocytozoonida and Glugeida in

SSU Bayesian and ML trees with weak support. *Nematocida* alone is the first branching lineage of canonical Microsporidia [54]. *Ovavesicula popilliae* and *Nematocida* spp. are found in terrestrial environments across the globe, most often sampled from agricultural land and grown in the laboratory [55]. The group primarily infect gut tissues of nematode hosts or adipose of insect hosts and are all horizontally transmitted. *Hamiltosporidium* spp., *Neoflabelliforma aurantiae*, *Astahelohania* spp., and *Areospora rohanae* are all isolated from benthic aquatic invertebrates, infecting their muscle, adipose, gut or are systemic. Most discoveries are restricted to Europe; however, two *Astathelohania* have been found in Australia and *Astathelohania rohanae* from Patagonia [56].

Genome sequences are required from more representatives of these 'orphan' lineages, as well as metchnikovellids and short-branch Microsporidia, to provide more comprehensive phylogenomic analyses of the Microsporidia as a whole [10].

Concluding remarks

A robust phylogenetic–taxonomic framework is essential for practical (e.g., nomenclature, consistent referencing) and hypothesis-driven evolutionary and ecological research for any group of organisms, as well as providing a strong basis for health and management policies [57]. This requirement is even greater when data on phenotypic characters are difficult to obtain and/or when phenotypic similarity is not a reliable indicator of evolutionary relationships. For the Microsporidia, rapid evolutionary diversification at the levels of genome, cell biology, and parasitological characters make evolutionary inference, based on phenotype alone, unreliable (see Outstanding questions). Molecular phylogenies based on a single gene (the SSU rRNA gene being sufficiently well sampled for this purpose) are also unable to resolve deeper microsporidian relationships. Balanced and comprehensive taxon sampling, and appropriate phylogenetic methods, provide strong support for major clades of Microsporidia and identify lineages that cannot (yet) be ascribed to higher taxa. Interpreting such phylogenies alongside the most comprehensive available phylogenomic analyses derived from full/partial genome sequences, provides the strongest basis for defining the major clades, helping to prioritise genomes for future sequencing, and provides the most robust taxonomic framework currently possible as a basis for future research.

Additional genetic markers, genomic sequences, and a better understanding of genome and cellular evolution, will provide insight into recent, as well as more ancient, evolutionary history: that is, adaptation and diversification at species level and below, for understanding host preferences and switches, ecological traits, evolution of virulence, which are elusive while restricted to currently available data.

It is essential to gather phenotypic and ecological data in tandem with genomic data to enable comprehensive taxon characterisation. Together, these data facilitate the inference of ancestral characters of robustly supported clades, and to understand more fully the evolutionary trajectories within them. These complementary data can also be used to inform predictive, epidemiological models (perhaps incorporating machine learning and artificial intelligence) of microsporidian epizootics in wildlife, humans, and agricultural/aquacultural industries (see Outstanding questions).

In addition to providing a consensus of a recent spate of taxonomy-related and diversity studies of Microsporidia, and resolving emerging nomenclatural inconsistencies, our taxonomic and phylogenetic synthesis provides a framework for future fundamental and applied microsporidian research, for example elucidating the evolutionary and ecological bases for predicting their emergence as pathogens of concern, including zoonotic diseases, providing an evolutionary

Outstanding questions

Can genomic, morphological, and ecological data be used to develop predictive models pertaining to the virulence and emergence of microsporidian epidemics in animals and humans?

Will a more detailed minimal taxonomic framework, coupled with greater genome availability, inform us upon the microsporidian ecosphere and the diversity within, allowing the elucidation of species complexes?

Given the availability of high-quality genetic, morphological, pathological, and ecological data for >284 microsporidian species, may we be reaching the precipice of where machine learning and artificial intelligence can help us to understand evolutionary patterns, emergence, and epidemiology?

context and robust terms of reference for meta-analyses of microsporidian diversity and host relationships [58], and facilitating a more fundamental understanding of their functional diversity across all ecosystems.

Acknowledgments

This work was supported by the Canadian Institutes of Health Research grant no. 400784 (A.W.R.) and an Alfred P Sloan Research Fellowship FG2019-12040 (A.W.R.). D.B. and G.D.S. were supported by Defra funding (contract FX001) supporting the OIE Collaborating Centre for Emerging Aquatic Animal Diseases.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.pt.2022.05.007.

References

1. Murareanu, B.M. *et al.* (2021) Generation of a Microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of Microsporidia. *mBio* 12, e0149021
2. Sprague, V. (1977) Classification and phylogeny of the Microsporidia. In *Comparative Pathobiology* (Bulla, L.A. and Cheng, T.C., eds), pp. 1–30, Springer
3. Wadi, L. and Reinke, A.W. (2020) Evolution of Microsporidia: an extremely successful group of eukaryotic intracellular parasites. *PLoS Pathog.* 16, e1008276
4. Bojko, J. and Stentiford, G.D. (2022) Microsporidian pathogens of aquatic animals. In *Microsporidia: Current Advances in Biology* (Reinke, A.W. and Weiss, L., eds), pp. 247–284, Springer
5. Chauvet, M. *et al.* (2022) Temporal variations of Microsporidia diversity and discovery of new host-parasite interactions in a lake ecosystem. *Environ. Microbiol.* 24, 1672–1686
6. Karpov, S. *et al.* (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front. Microbiol.* 5, 1–11
7. Adl, S.M. *et al.* (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* 66, 4–119
8. Vossbrinck, C.R. and Debrunner-Vossbrinck, B.A. (2005) Molecular phylogeny of the Microsporidia: ecological, ultrastructural, and taxonomic considerations. *Folia Parasitol.* 52, 131–142
9. Vossbrinck, C.R. *et al.* (2014) Phylogeny of the Microsporidia. In *Microsporidia: Pathogens of Opportunity* (Weiss, L. and Becnel, J., eds), pp. 203–220, Wiley
10. Bass, D. *et al.* (2018) Clarifying the relationships between Microsporidia and Cryptomycota. *J. Eukaryot. Microbiol.* 18, 1282–1298
11. Dubuffet, A. *et al.* (2021) A phylogenetic framework to investigate the microsporidian communities through metabarcoding and its application to lake ecosystems. *Environ. Microbiol.* 23, 4344–4359
12. Park, E. and Poulin, R. (2021) Revisiting the phylogeny of microsporidia. *Int. J. Parasitol.* 51, 855–864
13. Ardila-Garcia, A.M. *et al.* (2013) Microsporidian diversity in soil, sand, and compost of the Pacific Northwest. *J. Eukaryot. Microbiol.* 60, 601–608
14. Williams, B.A.P. *et al.* (2018) Group-specific environmental sequencing reveals high levels of ecological heterogeneity across the microsporidian radiation. *Environ. Microbiol. Rep.* 10, 328–336
15. Nakjang, S. *et al.* (2013) Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol. Evol.* 5, 2285–2303
16. Wiredu-Boakye, D. *et al.* (2017) Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae Microsporidia. *Environ. Microbiol.* 19, 2077–2089
17. Dean, P. *et al.* (2016) Microsporidia: why make nucleotides if you can steal them? *PLoS Pathog.* 12, e1005870
18. Dean, P. *et al.* (2018) Transporter gene acquisition and innovation in the evolution of Microsporidia intracellular parasites. *Nat. Commun.* 9, 1–12
19. Kotkova, M. *et al.* (2013) Latent microsporidiosis caused by *Encephalitozoon cuniculi* in immunocompetent hosts: a murine model demonstrating the ineffectiveness of the immune system and treatment with albendazole. *PLoS One* 8, e60941
20. Ironside, J.E. *et al.* (2008) Distribution and host range of the microsporidian *Pleistophora mulleri*. *J. Eukaryot. Microbiol.* 55, 355–362
21. Bunke, M. *et al.* (2019) Parasites influence cannibalistic and predatory interactions within and between native and invasive amphipods. *Dis. Aquat. Org.* 136, 79–86
22. Collado, J.S. *et al.* (2014) Flow cytometry analysis of *Nosema* species to assess spore viability and longevity. *Parasitol. Res.* 113, 1695–1701
23. Bakowski, M.A. *et al.* (2014) Ubiquitin-mediated response to Microsporidia and virus infection in *C. elegans*. *PLoS Pathog.* 10, e1004200
24. Reinke, A.W. *et al.* (2017) Identification of Microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. *Nat. Commun.* 9, 14023
25. Stentiford, G.D. *et al.* (2016) Microsporidia–emergent pathogens in the global food chain. *Trends Parasitol.* 32, 336–348
26. Stentiford, G.D. *et al.* (2019) Ultimate opportunists – the emergent *Enterocytozoon* group Microsporidia. *PLoS Pathog.* 15, e1007668
27. Stratton, C.E. *et al.* (2022) Revising the freshwater *Thelohania* to *Astathelohania* gen. et comb. nov., and description of two new species. *Microorganisms* 10, 636
28. Wijayawardene, N.N. *et al.* (2018) Notes for genera: basal clades of Fungi (including Aphelidiomycota, Basidiobolomycota, Chytridiomycota, Monoblepharomycota, Olpidiomycota, Rozellomycota and Zoopagomycota). *Fungal Divers.* 92, 43–129
29. Corsaro, D. *et al.* (2019) Filling gaps in the microsporidian tree: rDNA phylogeny of *Chytridiopsis typographi* (Microsporidia: Chytridiopsida). *Parasitol. Res.* 118, 169–180
30. Wijayawardene, N.N. (2020) Outline of Fungi and fungus-like taxa. *Mycosphere* 11, 1060–1456
31. Trzebny, A. *et al.* (2020) A new method of metabarcoding Microsporidia and their hosts reveals high levels of microsporidian infections in mosquitoes (Culicidae). *Mol. Ecol. Resour.* 20, 1486–1504
32. Tedersoo, L. *et al.* (2018) PacBio metabarcoding of Fungi and other eukaryotes: errors, biases, and perspectives. *New Phytol.* 217, 1370–1385
33. Mafune, K.K. *et al.* (2020) A rapid approach to profiling diverse fungal communities using the MinION™ nanopore sequencer. *Biotechniques* 68, 72–78

34. Stentiford, G.D. *et al.* (2018) Evidence for trophic transfer of *Inodosporus octospora* and *Ovipleistophora arlo* n. sp. (Microsporidia) between crustacean and fish hosts. *Parasitology* 145, 1105–1117
35. Kerr, R. *et al.* (2018) *Marteilia refringens* and *Marteilia pararefringens* sp. nov. are distinct parasites of bivalves and have different European distributions. *Parasitology* 145, 1483–1492
36. Willis, A.R. and Reinke, A.W. (2022) Factors That determine Microsporidia infection and host specificity. In *Microsporidia: Current Advances in Biology* (Weiss, L.M. and Reinke, A.W., eds), pp. 91–114, Springer
37. Williams, T.A. *et al.* (2016) A recent whole-genome duplication divides populations of a globally distributed microsporidian. *Mol. Biol. Evol.* 33, 2002–2015
38. Williams, B.A. *et al.* (2008) Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal Microsporidia with low gene densities. *BMC Genomics* 9, 1–9
39. Sweeney, A.W. *et al.* (1985) Intermediate host for an *Amblyospora* sp. (Microspora) infecting the mosquito, *Culex annulirostris*. *J. Invertebr. Pathol.* 46, 98–102
40. Andreadis, T.G. *et al.* (2013) Morphological and molecular characterization of a microsporidian parasite, *Takaokaspora nipponicus* n. gen., n. sp. from the invasive rock pool mosquito, *Ochlerotatus japonicus*. *J. Invertebr. Pathol.* 114, 161–172
41. Peyretaillade, E. *et al.* (2012) Annotation of microsporidian genomes using transcriptional signals. *Nat. Commun.* 3, 1137
42. Polonais, V. *et al.* (2019) Draft genome sequence of *Tubulinosema ratisbonensis*, a microsporidian species infecting the model organism *Drosophila melanogaster*. *Microbiol. Res. Announc.* 8, e00077-19
43. Chen, L. *et al.* (2020) Complete genome of a unicellular parasite (*Antonospora locustae*) and transcriptional interactions with its host locust. *Microb. Genom.* 6, mgen000421
44. Vavra, J. and Undeen, A.H. (1970) *Nosema algerae* n. sp. (Cnidospora, Microsporida) a pathogen in a laboratory colony of *Anopheles stephensi* Liston (Diptera, Culicidae). *J. Protozool.* 17, 240–249
45. Andreadis, T.G. (2007) Microsporidian parasites of mosquitoes. *J. Am. Mosq. Contr. Assoc.* 23, 3–29
46. Lowman, P.M. *et al.* (2000) The effects of elevated temperatures and various time-temperature combinations on the development of *Brachiola* (*Nosema*) *algerae* n. comb. in mammalian cell culture. *J. Eukaryot. Microbiol.* 47, 221–234
47. Monaghan, S.R. *et al.* (2011) *In vitro* growth of Microsporidia *Anncaliia algerae* in cell lines from warm water fish. *Animal* 47, 104–113
48. Sokolova, Y.Y. *et al.* (2020) Development of *Anncaliia algerae* (Microsporidia) in *Drosophila melanogaster*. *J. Eukaryot. Microbiol.* 67, 125–131
49. Coyle, C.M. *et al.* (2004) Fatal myositis due to the microsporidian *Brachiola algerae*, a mosquito pathogen. *N. Engl. J. Med.* 351, 42–47
50. Ziad, F. *et al.* (2021) Fatal disseminated *Anncaliia algerae* myositis mimicking polymyositis in an immunocompromised patient. *Neuromuscul. Disord.* 31, 877–880
51. Grupe, A.C. and Quandt, C.A. (2020) A growing pandemic: A review of *Nosema* parasites in globally distributed domesticated and native bees. *PLoS Pathog.* 16, e1008580
52. Han, B. *et al.* (2021) Microsporidiosis in humans. *Clin. Microbiol. Rev.* 34, e00010–e00020
53. Bojko, J. *et al.* (2020) *Ovipleistophora diplostomuri*, a parasite of fish and their trematodes, also infects the crayfish *Procambarus bivittatus*. *J. Invertebr. Pathol.* 169, 107306
54. Cormier, A. *et al.* (2021) Draft genome sequences of *Thelohania contejeani* and *Cucumispora dikerogammari*, pathogenic microsporidia of freshwater crustaceans. *Microbiol. Res. Announc.* 10, e01346-20
55. Zhang, G. *et al.* (2016) A large collection of novel nematode-infecting Microsporidia and their diverse interactions with *Caenorhabditis elegans* and other related nematodes. *PLoS Pathog.* 12, e1006093
56. Stentiford, G.D. *et al.* (2014) *Areospora rohanae* n. gn. n. sp. (Microsporidia; Areosporiidae n. fam.) elicits multi-nucleate giant-cell formation in crabs. *J. Invertebr. Pathol.* 118, 1–11
57. Stentiford, G.D. *et al.* (2014) Policy, phylogeny, and the parasite. *Trends Parasitol.* 30, 274–281
58. Ruan, Y. *et al.* (2021) The largest meta-analysis on the global prevalence of microsporidia in mammals, avian and water provides insights into the epidemic features of these ubiquitous pathogens. *Parasit. Vectors* 14, 1–14
59. Katoh, K. and Standley, D. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780
60. Capella-Gutiérrez, S. *et al.* (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973
61. Stamatakis, A. (2014) RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313
62. Ronquist, F. *et al.* (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542
63. Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 1–14
64. Hunter, J.D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95
65. R Core Team (2013) *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria
66. Wickham, H. *et al.* (2016) Package 'ggplot2': create elegant data visualisations using the grammar of graphics. *R J.* 5, 1–189
67. Pedregosa, F. *et al.* (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830
68. González-Tortuero, E. *et al.* (2016) Genetic diversity of two *Daphnia*-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region. *Parasit. Vectors* 9, 1–15
69. Donahue, K.L. *et al.* (2019) Using the SSU, ITS, and ribosomal DNA operon arrangement to characterize two Microsporidia infecting Bruce spanworm, *Operophtera bruceata* (Lepidoptera: Geometridae). *J. Eukaryot. Microbiol.* 66, 424–434
70. Tsai, S.J. *et al.* (2002) Complete sequence and structure of ribosomal RNA gene of *Heterosporis anguillarum*. *Dis. Aquat. Org.* 49, 199–206
71. Hajek, A.E. *et al.* (2018) *Nosema maddoxi* sp. nov. (Microsporidia, Nosematidae), a widespread pathogen of the green stink bug *Chinavia hilaris* (Say) and the brown marmorated stink bug *Halyomorpha halys* (Stål). *J. Eukaryot. Microbiol.* 65, 315–330
72. Vossbrinck, C.R. *et al.* (1993) Ribosomal DNA sequences of *Encephalitozoon hellem* and *Encephalitozoon cuniculi*: species identification and phylogenetic construction. *J. Eukaryot. Microbiol.* 40, 354–362
73. Li, W. *et al.* (2019) Host specificity of *Enterocytozoon bieneusi* and public health implications. *Trends Parasitol.* 35, 436–451
74. Tokarev, Y.S. *et al.* (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
75. Hatjina, F. (2011) Polar tube protein gene diversity among *Nosema ceranae* strains derived from a Greek honeybee health study. *J Invert. Pathol.* 108, 131–134
76. Bateman, K.S. *et al.* (2016) Single and multi-gene phylogeny of *Hepatospora* (Microsporidia) – a generalist pathogen of farmed and wild crustacean hosts. *Parasitology* 143, 971–982
77. Jamy, M. *et al.* (2020) Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol. Ecol. Resour.* 20, 429–443
78. Vossbrinck, C.R. *et al.* (1987) Ribosomal RNA sequence suggests Microsporidia are extremely ancient eukaryotes. *Nature* 326, 411–414
79. Imhoff, E.M. *et al.* (2012) Horizontal transmission of *Thelohania contejeani* in the endangered white-clawed (*Austropotamobius pallipes*) and the invasive signal crayfish (*Pacifastacus leniusculus*). *Parasitology* 139, 1471–1477
80. Weng, M. *et al.* (2021) First report of *Ovipleistophora ovariae* and *O. diplostomuri* in China provides new insights into the intraspecific genetic variation and extends their distribution. *Parasitology* 149, 314–324

81. Lovy, J. and Friend, S.E. (2017) Phylogeny and morphology of *Ovipleistophora diplostomuri* n. sp. (Microsporidia) with a unique dual-host tropism for bluegill sunfish and the digenean parasite *Posthodiplostomum minimum* (Strigeatida). *Parasitology* 144, 1898–1911
82. Stratton, C. *et al.* (2022) The plot thickens: *Ovipleistophora diplostomuri* infects two additional species of Florida crayfish. *J. Invertebr. Pathol.* 191, 107766
83. Amos, B. *et al.* (2022) VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.* 50, 898–911
84. Li, T. *et al.* (2017) SilkPathDB: a comprehensive resource for the study of silkworm pathogens. *Database* 2017, 1–6
85. Pan, G. *et al.* (2013) Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics* 14, 1–14