Generation of a microsporidia species attribute database and analysis of the extensive ecological and phenotypic diversity of microsporidia

Brandon M. Murareanu[1*], Ronesh Sukhdeo[1*], Rui Qu[1*], Jason Jiang[1*], and Aaron W. Reinke[1#]

[1] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

*These authors contributed equally

#Corresponding author

aaron.reinke@utoronto.ca

Abstract

Microsporidia are a large group of fungal-related obligate intracellular parasites. Though many microsporidia species have been identified over the past 160 years, there is a lacking depiction of the full diversity of this phylum. To systematically describe the characteristics of these parasites, we created a database of 1,440 species and their attributes, including the hosts they infect and spore characteristics. We find that microsporidia have been reported to infect 16 metazoan and four protozoan phyla, with smaller phyla being underrepresented. Most species are only reported to infect a single host, but those that are generalists are also more likely to infect a broader set of host tissues. Strikingly, polar tubes are 3-fold longer in species that infect tissues besides the intestine, suggesting that polar tube length is a determinant of tissue specificity. Phylogenetic analysis revealed four clades which each contain microsporidia infecting hosts from all major habitats. Although related species are more likely to infect similar hosts, we observe examples of changes in host specificity and convergent evolution. Taken together, our results show that microsporidia display vast diversity in their morphology and the hosts they infect, illustrating the flexibility of these parasites to evolve new traits.

Introduction

Microsporidia are a large phylum of obligate intracellular parasites. Although they were once thought to be protists, it is now clear they belong to a lineage of early diverging fungi(1). Microsporidia was the first parasite to

have its genome sequenced, revealing the smallest eukaryotic genome with as few as ~2000 protein coding genes(2, 3). Genome reduction results from the loss of many metabolic and regulatory genes, reduced non-coding DNA, and reduced gene sizes that encode for proteins smaller than their yeast orthologs(4–7). Microsporidia have also lost mitochondria and instead depend on reduced organelles known as mitosomes(8). As a result, microsporidia rely heavily on their hosts for nutrients and have evolved many novel proteins for interacting with host cells(9, 10). Paradoxically, although microsporidia are amongst the simplest eukaryotes, there are at least 1500 known species that as a phylum can infect many different animals and most major tissues(11, 12).

Microsporidia cause death and disease and are a threat to both animal and human health. They were first discovered in 1857 in silkworms and microsporidia infections have triggered historical collapses of the sericulture industry(13). Microsporidia infections in animals are very common. For example, 50% of honey bee colonies and 60% of field-collected mosquitos have been reported to be infected with microsporidia(14, 15). These parasites are an emerging threat, with the most common species that infect honey bee colonies and shrimp farms only being discovered within the last two decades(16, 17). Microsporidia also cause problems in laboratory research, commonly infecting zebrafish facilities and causing fruit fly colonies to collapse(18, 19). Infections often result in delayed animal growth, reduction in offspring, and premature mortality(20–23). Over a dozen species of microsporidia can infect humans and several of these species are also found in a large variety of different mammals and birds, leading to concern over zoonotic transfer(24–26). In addition to the threats posed by microsporidia, they also have potential benefits, including being used as biocontrol agents against insect pests or preventing the spread of mosquito-vectored diseases(27, 28).

All microsporidia produce environmentally resistant spores. These spores contain both the infectious sporoplasm and the unique polar tube infectious apparatus(29). In the presence of a host, microsporidia fire their polar tube which is thought to either pierce the host cell or fuse with the host plasma membrane to form an invasion synapse(30). Each species has spores of a characteristic size, but some species have multiple classes of spores(31, 32). While many infections are thought to result from the oral consumption of spores, other types of infections e.g. ocular, can also occur(26).

Although many microsporidia species have been characterized, their properties have not been examined systematically. Thus, the true diversity of these medically and agriculturally important parasites is unknown. Here, we curate a database of microsporidia attributes for 1440 species. Using this data, we describe the history and geography of microsporidia discovery. We show that microsporidia can infect a diverse range of host species from all major habitats. We find that microsporidia that infect more hosts tend to infect more tissues. We observe large ranges for spore size, which correlates with the number of nuclei and polar tube length. We also find that the length of the polar tube is longer in species that infect other tissues besides the intestine. Phylogenetic analysis shows that microsporidia group into clades with predominant environmental habitats, but the ability to infect in other environments within a clade is also common. Related species mostly infect similar hosts, but dramatic changes in host specificity and convergent evolution also occur. Together, this work describes the diverse properties of microsporidia species and provides a resource for further understanding these enigmatic parasites.

Results

Generating a database of microsporidia species' characteristics

To determine the reported diversity of microsporidia species, we comprehensively examined the literature. We began with Victor Sprague's 1977 book, the most complete compendium of microsporidia published thus far, which contains information on 692 species(33). We then identified 704 journal articles containing reports of identified microsporidia species. We recorded the described attributes of each species including the date and location of discovery, host species, infected tissues, spore size, and polar tube length. Overall, we generated a database containing information on 1,440 microsporidia species (Table S1).

We first examined the history of microsporidia species discovery. The original article describing the first species of microsporidia was published in 1857(34). Since 1881, on average ~100 novel species have been reported per decade. The number of discovered species peaked in the 1970s with 246 species being reported. Although in recent years fewer species have been reported than at this peak, during the last decade, on average 14 novel species have been described every year (Fig. 1A).

We next determined the geographical locations from which microsporidia have been isolated. Although the majority of described species have been found in Europe and North America, microsporidia have been found in all continents except Antarctica (Fig. 1B). They have also been reported in 11 oceans or seas, including a species found in the Weddell sea off the coast of Antarctica, and from methane seeps on the floor of the pacific ocean(35, 36). While the United States has reported the most species of microsporidia (283), 123 species have also been found in the much smaller Czech Republic (Fig. 1C). We also see a distribution of reported species within a country, with microsporidia being reported in the majority of states and provinces in the United States and Canada (Fig. 1D and E).

Hosts and tissues reported to be infected by microsporidia

Microsporidia have often been described as infecting nearly all animal taxa, but a full description of which host phyla these parasites infect is lacking(26, 37, 38). Using several databases describing animal and protozoan diversity, we determined the taxonomical placement of 1436 microsporidia hosts (Table S2). We find that infections have been reported in 16 metazoan and four protozoan phyla. The majority of the hosts in six of these phyla are themselves parasites, demonstrating that hyperparasitism is a widespread microsporidian trait (Fig. 2A). There are 16 metazoan phyla that have not been reported to be infected by microsporidia. These phyla have on average 26-fold fewer known species than phyla that have been reported to be infected with microsporidia. This suggests that the absence of known microsporidia infections in these phyla may be due to a lack of sampling from these hosts (Fig. 2B). We also observed a diversity of hosts within a phylum; the two phyla with the most hosts, Chordates and Arthropods, contain hosts in 39 out of 155 orders and 42 out of 162 orders, respectively (Fig. 2C and S1).

Because microsporidia are obligate intracellular parasites, the habitat of the host informs that of the microsporidia species(12). We determined the ecological environments of the different hosts in our database. Of the four major habitat types, 59.3% of hosts are terrestrial, 36.8% of hosts are freshwater, 25.1% of hosts are marine, and 13.6% of hosts are brackish (of hosts with environment data). Interestingly, 27.4% of hosts with environment assignments are classified as living in more than one environment. This is exemplified by eight hosts, including

the mosquito (*Anopheles pseudopunctipennis)* and parasitic flatworm (*Heterophyes heterophyes*), which inhabit all four environment types (Fig. 2D and Table S2).

Microsporidia are known to infect a range of different tissues within their hosts(12). To investigate the extent of this diversity, we classified infections into 23 tissue types (Table S3A). The tissues infected by microsporidia encompass all embryonic germ layers and most major organ systems(39). The adipose tissue, intestine, muscles, germline, and excretory system are the most commonly infected tissues. Infections have also been reported in specialized tissues such as the gills of fish and oenocytes, which are lipid-processing cells in insects(40) (Fig. 3A). Most microsporidia infections appear to be restricted to a single tissue, but ~11% of species can infect four or more tissues (Fig. 3B).

Microsporidia can be specialists that infect only one or a few hosts, or generalists that can infect many hosts(5, 37). Over 80% of reported species (with host data) only infect a single host, but 2.2% can infect five or more hosts (Fig. 3C). To determine if there is a relationship between host range and tissue specificity, we examined if species that infect more hosts also infect more tissues. Indeed, 34% of species that infect one host infect multiple tissues, compared to 57% of species that infect four or more hosts infecting multiple tissues (Fig. 3D). This correlation suggests that species with larger host ranges tend to infect a broader set of tissues.

Diversity of microsporidia spore morphology

All microsporidia have an extracellular spore form and the morphological properties of these spores can be used to distinguish species(31). To examine the diversity of microsporidia spore morphology, we first classified spores into 20 unique shape descriptors (Table S3B). Over 93% of classified spores belong to just six shape classes (Fig. 4A). Examination of spore size revealed a large diversity in both spore length (1 – 25 µm) and width (0.15 – 10 µm). Although there is overlap between shape description and spore dimensions, there are differences in the average length/width ratio of the main shape classes with rod-shaped being 3.7, ovocylindrical being 2.7, oval, pyriform, and elliptical being 1.8-1.9, and spherical being 1 (Fig. 4B and S2).

Using length and width measurements of spores, we estimated spore volumes, which range from $3.5 \times 10^{-2}\ \mu m^3 - 1.3 \times 10^3\ \mu m^3$. We compared our volume estimates with 3-dimentionsal measurements using

the two species for which such data is available(41). For *Anncaliia algerae*, we estimated a volume of 22.7 μm^3 and the volume was calculated to be 8.8 μm^3. For *Encephalitozoon hellem,* we estimated a volume of 2.1 μm^3, and the volume was calculated to be 3.6 μm^3. We also used the same volume formula, but the length and width measurements from (41), and estimated *A. algerae* to be 12.8 μm^3 and *E. hellem* to be 4.1 μm^3. These comparisons demonstrate, at least for this limited set, that our estimated volumes are similar to those calculated using three-dimensional methods and that the majority of the differences in volume arise from differences in spore length and width measurements.

Microsporidia spores contain polar tubes that fire in the presence of a host signal and are used to infect host cells(30). The length of the polar tube can be measured by artificially triggering spore firing by treating with different chemicals or mechanical pressure(30, 42). The polar tube is coiled inside of the spore and the number of coils can be determined using electron microscopy(29). Using the spore diameter and the number of polar tube coils, we estimated the length of the unfired polar tube. We observe a large range in both experimentally measured (2.5 µm – 420 µm) and estimated (1.5 µm – 752.1 µm) polar tube lengths. We compared estimated polar tube length to calculated polar tube length and observed a significant correlation (Fig. 4C).

Microsporidia spores need to be able to accommodate both the polar tube and the sporoplasm. Sporoplasm size measurements do not exist for most species, so we used the number of nuclei (either one or two) as a proxy for the size of the sporoplasm. We examined the relationship between spore size and nuclei number and observed that spores with two nuclei are on average ~22% larger than those with one nucleus (Fig. 4D). We next examined the relationship between polar tube length and spore size. We observed that experimentally measured polar tube lengths were positively and significantly corelated with spore volume (Fig. 4E and S3). Taken together, these results suggest that although the number of nuclei and the length of the polar tube do not explain all of the variation in spore volume, the size of intracellular contents is correlated with spore size.

One possible explanation for why polar tube length varies so widely, is that the polar tube specifies which tissues are infected. This idea was proposed by Luallen and colleagues studying microsporidia infection in the nematode *Caenorhabditis elegans*(6). The authors observed that *Nematocida parisii*, which infects the intestine, has a polar tube length of 4.03 µm, in contrast to that of *Nematocida displodere*, which infects the muscle and epidermis

from the intestinal lumen, and has a polar tube length of 12.55 μm. To determine if polar tube length correlates with infected tissue type, we examined 11 hosts that were infected by at least one microsporidia species that exclusively infects the digestive system and at least one species that infects other tissues. In 82% of microsporidia species pairs, the polar tube was longer in the microsporidia species that infected other tissues besides the digestive system. On average, polar tubes of species that infect other tissues were 3-fold longer than those of species restricted to the digestive system. This suggests that one function of polar tube length is determining host-tissue specificity (Fig. 4F and Table S4).

Microsporidia can either be transmitted horizontality, from host to host, or vertically, from parent to offspring(20, 43). We examined whether species were reported to infect the germline (vertical) or somatic tissue (horizontal) and observed that ~18% of species are potentially vertically transmitted (Fig. S4A). We also observed that transmission mode is not dependent on host environment (Fig. S4B). Furthermore, spore volume and polar tube length are not correlated with transmission mode (Fig. S4C-D).

Evolutionary properties of microsporidia

Microsporidia have previously been reported to group into five clades, classified based on host environmental habitat(44). Using our large collection of host and species information, we sought to revisit these groupings. We performed Bayesian phylogenetic analysis on 273 18S rRNA sequences belonging to microsporidia species for which we also had attribute data (Table S5A and B). The resulting tree recovers four clades (I,III,IV, and V) and is largely in agreement with recently published microsporidia phylogenetic trees(45, 46) (Fig. 5A and S5). We also assembled a Maximum Likelihood phylogenetic tree using the same sequences, and the majority of species (~95%) were placed in the same clade in both trees, further supporting our clade groupings (see methods). All subsequent analysis was done using the Bayesian tree. We then determined the percentage of microsporidia species in each clade that inhabit each of the four major environments. We mostly observed the previously determined environmental groupings, with clade IV being terrestrial and clade III being marine. We also observed that clades I and V have a higher percentage of microsporidia with freshwater hosts than the other clades, as has been previously shown. However, clades I and V have a higher percentage of microsporidia with terrestrial than freshwater hosts (Fig. 5B). Although there is a predominant habitat in each clade, between 29-44% of

microsporidia species in each clade do not infect a host in the predominant habitat. This suggests that evolution of microsporidia within clades is not strictly restricted to hosts from a single environmental habitat.

We next examined the diversity of spore size and polar tube length amongst clades (Fig. 5C and D). We observe that the clades with the highest percentage of microsporidia infecting freshwater hosts (I and V) have on average larger spores and more variation in spore size than the other two clades (Fig. 5C). This same trend exists in polar tube length, but to a lesser extent (Fig. 5D).

Related species of microsporidia are known to infect very different hosts, but it is unclear how common of an occurrence this is(7). To determine the relationship between microsporidia species similarity and host similarity, we first calculated the pairwise distances between 270 microsporidia 18S rRNA sequences (Fig. 6A). We then compared how often related species infect the same host family, order, class, or phylum. We observed that while 43% of microsporidia species pairs with greater than 65% 18S similarity infect hosts from the same family, this was true for only 2% of pairs with less than 65% 18S similarity. These strong associations between microsporidia and host relatedness were also observed at the other taxonomic ranks (Fig. 6B). Although very closely related microsporidia almost always infect the same phylum, we observed 28 pairs of species with greater than 80% similarity infecting different phyla (Table S5C). This data includes known examples of highly similar species infecting different phyla such as between microsporidia that infect ciliates and those that infect aquatic arthropods and between microsporidia infecting aquatic crustaceans and those infecting parasitic trematodes and paramyxids (47, 48). The species pairs we identified also included eight such instances between species that infect arthropods and the known human infecting species *Vittaforma corneae* and *Trachipleistophora hominis* (26). To identify examples of convergent evolution, we determined how often a host species is infected by microsporidia from more than one clade and found 22 such cases (Fig. 5E). Taken together, our results indicate that microsporidia mostly evolve to infect similar hosts but can undergo dramatic hosts shifts.

Finally, we determined the relationship between microsporidia sequence relatedness and infected tissues, habitat, and spore similarity. We find that closely related species mostly infect the same tissues and share the same environmental habitats (Fig. 6C-D). Spore size and polar tube length are also more similar in closely related species, but to a lesser extent than that of host similarity (Fig. 6E-F). Even closely related species show less

than 75% similarity in the tissues they infect and in their spore properties. These results suggest that although related species are likely to have shared characteristics, microsporidia are flexible in their ability to evolve new traits.

Conclusions

Here we report the largest systemic collection of microsporidia species characteristics assembled since Sprague in 1977(33). Although the data we present is comprehensive, there are several limitations. First, we only curated named species that were reported with attribute data. There are at least 93 identified microsporidia species that we could not find information on and therefore did not include in our analysis(49). Second, many microsporidia species names have changed throughout the years and this likely led to missing information(50). Third, we do not include complete attribute data for every species as authors do not always report on the same properties. Fourth, the data we collected is from a variety of different sources and some characteristics such as spore size were determined through multiple techniques with differing accuracy, lending some uncertainty to the data and our analyses. Finally, for species that are the subject of many publications, we are unlikely to have captured all of the known data including the full breadth of hosts these species infect.

About 1500 microsporidia species have been identified, which is likely only a small fraction of the true number of species(51). We find that identification of novel microsporidia species peaked in the 1970s, which is consistent with a previous report showing identification of microsporidia has lagged behind other parasites(52). Sampling for microsporidia species has been biased with ~65% of the species identified so far being from Europe and North America. Africa, South America, and Oceania have been shown to have large levels of parasite-host diversity, suggesting that these parts of the world could contain many undiscovered microsporidia species(53).

Although the pace of microsporidia discovery has slowed, many new and diverse species are still being discovered. Since the 1990s, microsporidia were reported for the first time in seven phyla (Cephalorhyncha, Porifera, Amoebozoa, Dicyemida, Gastrotricha, Phoronida, and Nemertea), suggesting that most types of animals may be infected by microsporidia. In support of this, although there are no named microsporidia species infecting tardigrades, animals in this phylum has been observed to be infected by microsporidia(54). Environmental sampling has also revealed the existence of many species of unnamed microsporidia, especially

a diversity of short-branched microsporidia(45, 55, 56). Echinoderms comprise the largest phylum without any reported microsporidia infections and represent an interesting target for future efforts to uncover novel microsporidia diversity.

The molecular mechanisms that determine host and tissue specificity in microsporidia are mostly unknown. Here we show that microsporidia polar tube length is longer in species that infect tissues other than the intestine. In bilaterian animals, the intestine is in the middle of the animal and access to most other tissues requires going through the intestine(57). This has been shown in *C. elegans*, where infection of the epidermis is dependent upon oral consumption of spores(6). Microsporidia have been shown to be more selective for which types of cells they infect inside of animals than in cell culture (58, 59). Microsporidia have also been shown to infect tissues via injection that were not infected during feeding of spores to mosquito hosts(60, 61). These results suggest that access to tissues could determine specificity and that longer polar tubes could be used to access these other tissues. Microsporidia infections can be disseminated within a host both through spread of meronts between cells and the production of spores that infect other tissues(21, 62). Thus, it is possible that in some cases spores with shorter polar tubes could infect more distal tissues after an initial round of infection. We also show that there is a tendency for microsporidia species that infect more hosts to also infect more tissues. This is likely even more common than our data indicates, as it only takes a single infected animal to determine tissue distribution, but many identifications to determine host range(31). Our data suggest that the evolutionary forces that drive species to be generalists also tends to generate species with broad tissue specificity.

In 2005 Vossbrinck and Debrunner-Vossbrinck proposed that microsporidia compose five major clades that group according to the environmental habitat of each species host(44). Although these environmental groupings provide an attractive explanation of microsporidia evolution, none of the clades only contained hosts from a single environment(63). Clade II was not well supported in this original phylogeny and has not been consistently recovered in more recent phylogenetic trees(20, 31, 45, 55). The position of *Nematocida* in these phylogenies has not been constant and often differs from the basal position recovered from trees based on whole genome sequencing(46, 64). Environmental sampling from diverse environments recovered microsporidia from many different clades, demonstrating that there is not a strict correspondence between clades and host environment. For example, many samples from marine environments were found in clade IV (Terresporidia). Intriguingly,

several OTUs (operational taxonomy unit) were found in multiple environments, providing support for some species having multiple hosts from different environments or the same host being infected in multiple environments(45). Our analysis suggests that although clades do belong predominantly to a single environment, microsporidia can evolve to infect hosts in other habitats. A host often exists in more than one habitat and some microsporidia have multiple hosts that exist in distinct habitats. Because of this, many species of microsporidia are found in multiple environments, likely aiding their ability to evolve to infect new hosts.

In conclusion, we have assembled a collection of microsporidia attributes for 1,440 species. This data will be a useful comparative resource which will support further species identification. In addition to the questions that we have examined, this database will aid in addressing other systematic inquires about microsporidia diversity. The community of microsporidia researchers can add to the microsporidia species attributes database by submitting newly discovered species and updating existing species at https://www.reinkelab.org/microsporidiaspecies.

Materials and methods

Construction of microsporidia species properties database

Each microsporidia species described from Victor Sprague's 1977 book was manually entered into an Excel table(33). Each row represented a microsporidium, and columns were created for each species attribute recorded. Parenthesized text was used to denote which spore class or host an attribute corresponded to. The spore conditions (ex: fixed, fresh, etc.) under which morphological attributes were recorded were also denoted by parenthesized text. Species without a specific name (ex: Nosema sp.) were numbered to give each entry a unique name.

To identify additional species, we searched the journal databases Scopus, Web Of Science and JSTOR for papers published between 1977 and April 2021. The search terms ("microsporidia" OR "microsporidium" OR "microsporida") were used for all searches. Papers describing new microsporidia species or further describing existing species were used. Moreover, we looked at citing articles for papers describing new species, to find papers further describing those species. Non-English papers were translated into English using Google Translate. An additional 134 species were identified from a database of compiled by Kirk(49). Papers describing

these additional species were found by searching in the aforementioned databases, or with Google. Microsporidia attributes for species described in journal articles were recorded as described above. For a subset of papers where the full article could not be accessed, information was recorded from the abstract.

Analysis of microsporidia discovery date and geographical location

For each species, the earliest year from the "Date Identified" column of Table S1 was extracted. Species without identification year data were excluded from this analysis. Data of the countries where each species had been reported was extracted from the "Locality" column of the database. A species found in multiple countries was counted for all of those countries. State and province distribution maps for the United States and Canada were created using the same method. With the exception of oceans, species found in bodies of water were counted as the country located closest to it. If a species was reported either in a country that no longer exists or a general region, it was counted as the largest country that occupied its lands. For example, species found in the U.S.S.R. were counted as species found in Russia. Species reported from each continent were determined according to the categorization of worldpopulationreview.com.

Determination of host taxonomy and environment

The taxonomy for each host species was determined using Catalogue of Life (http://www.catalogueoflife.org/annual-checklist/2019/), GBIF (https://www.gbif.org/), WoRMS (http://www.marinespecies.org/), NCBI Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) or Wikipedia (https://en.wikipedia.org/). Using this approach, we annotated the taxonomy for 1422 out of 1449 host species. Catalogue of Life was used to determine the number of extant species in each phyla and the number of metazoan phyla, with the exception that Myxozoa was considered part of Cnidaria.

Host environment was determined using Catalogue of Life, GBIF and WoRMS. Host environment data was also extracted from other online databases, papers describing the host, or the original papers describing infection of the host by microsporidia (Table S2). To classify moths, all non-Pyralidae moths were classified as "terrestrial". Pyralidae moths were classified as freshwater and terrestrial, since they have aquatic larval stages(65). If environment data was not available at the species level, we used environment data from the closest taxonomic

level, the highest being order. Some species without environment data in the sources mentioned above, were assigned environments by common knowledge, i.e., parakeets inhabiting a terrestrial environment. The environment of 1338 out of 1449 species was annotated. Experimental hosts for microsporidia were excluded from analysis, as they may not reflect the natural host range of a microsporidia species(66).

Categorizing and counting tissues infected

All unique infected-tissue names were extracted from the “Site” column of Table S1. These names were sorted into 23 broader tissue categories (Table S3A). Each unique tissue name was restricted to a single tissue category. Tissues that could not be clearly assigned to a category were classified as “Ambiguous”. Infections of protist species were classified as “Protist”. To count the number of tissues infected by each microsporidium, we counted how many tissues were listed for each microsporidium. Microsporidia species with systemic or ambiguous infections were excluded, as tissues in these categories could not be associated with a single tissue. The overall infection frequencies within tissue categories were determined using a custom R script. Each species was counted as infecting a tissue category one or zero times. Ambiguous and protist infections were not counted.

Comparing infected hosts to infected tissues

To compare the number of hosts infected to the number of tissues infected, microsporidia species were grouped based on the number of hosts they infect. The proportion of species infecting one tissue was calculated within each host bin. Species with systemic infections were counted as infecting multiple tissues. Experimental hosts were excluded from analysis.

Determining microsporidia transmission mode

Microsporidia’s mode of transmission was either determined from literature transmission data or inferred from the tissue infected, as has been done previously(43). (“Site” and “Transmission (described in literature)” columns in Table S1). Species without literature transmission data were classified as vertically transmitted if infecting germline, horizontally transmitted if infecting somatic tissue, or both. Transmission mode was not determined for species with systemic infections, as it is unclear what tissues they infect. When available, literature transmission data supplemented transmission mode determination. Species only infecting somatic tissues but with vertical transmission data from literature were classified as both. Species only infecting germline with horizontal

transmission from literature were also classified as both. For species without tissue infection data, mode of transmission was determined solely through literature transmission data (horizontal, vertical, or both). Transmission mode was not determined for species with no tissue data and only autoinfection as literature transmission data, due to uncertainty over its mode of transmission. Moreover, transmission mode was not determined when transmission was explicitly described as ambiguous in literature. When comparing rates of horizontal and vertical transmission across environments, transmission modes were described for individual microsporidia host species. Only hosts in single environments were considered. Host species that could not be matched to specific tissues or modes of transmission were excluded from analysis.

Spore shape classification

Spore shapes were classified using spore shapes reported in each publication and extracted from the "Spore Shape" column in Table S1. These shapes were grouped and placed into 20 shape categories (Table S3B). Shapes that could not be categorized as a defined grouped shape were labelled "ambiguous" and excluded from analysis. If a given spore was documented to be oval in addition to another shape, the latter shape classification was kept and used for analysis. However, if a spore was documented to be oval in addition to at least two other shapes, it was excluded from analysis.

Analysis of spore dimensions

Spore lengths and widths were extracted from "Spore Length Average" and "Spore Width Average" columns in Table S1. If a given species had multiple measurements for a spore class, the "fresh" measurement was used and the remaining "fixed" measurements were excluded. If a spore class only had "fixed" measurements available, they were included in the analysis.

Calculating spore volume

Spore volume was estimated for all types of documented spores using the formula for an ovoid, as this shape accounts for the majority of spores (Fig. 4A). The following formula was used: Volume= $4/3\pi abc$, where a, b, and c are the 3 semi-axes of the ovoid. Width/2 was used for both a and b as width and depth of spores have been shown to be similar(41). Length/2 was used for c. When possible, priority was given to the "fresh" measurements of documented spore for calculating volume.

Calculating polar tube length

Polar tube length was calculated by using the spore width and number of observed polar filament coils documented. If a range of coils was provided instead of a single value, the average number of coils was used. Spore width typically came from light microscopy observations, while coils came from transmission electron microscopy. Fresh width values were prioritized for calculations. The following formula was used: $Calculated\ polar\ tubule\ length = \pi \times \text{width of spore} \times \text{number of polar filament coils}$. Polar tube length was not calculated for spores with uncoiled polar tubes.

Microsporidia tissue specificity

To assess if microsporidia polar tube length affects which tissues are infected, hosts infected by multiple microsporidia species were identified. Calculated polar tube lengths were organized into “Intestinal” infection if the species only infected the gut/digestive system as determined in Table S3A. If a species infecting the same host infected non-intestinal tissues, it was classified as “Other Tissues”. Every pair of microsporidia species between these two classes was compared for every host (Table S4).

Comparing polar tube length to volume

To compare polar tube length to spore volume, only species with experimental polar tube values were considered. Species with multiple spore classes were represented multiple times. Entries where spore class could not be clearly matched to volume, polar tube length, or nucleus data were excluded. Least squares regression was performed for all spores (Fig. S4), as well as uninucleate and binucleate spores individually to control for the confounding effects of nuclei number. Correlation was calculated with Spearman r_s to limit the effect of outliers on correlation values.

18S phylogenetic analysis of microsporidia species

Using a modified version of the *read.Genbank* function from the *ape* package for R, microsporidia 18S sequences and associated data were collected from the NCBI GenBank nucleotide database: https://www.ncbi.nlm.nih.gov/genbank/. The search terms *("Microsporidia"[Organism] AND (rRNA[Title] OR ribosomal[Title])) AND (small[Title] OR SSU[Title] OR 18s[Title] OR 16s[Title])* were used to collect 7830 unique

18S sequences representing 1107 unique microsporidia species identifiers. The longest sequence for each unique species identifier was extracted from this collection for subsequent analysis. Only the sequences with species identifiers corresponding to species in our database (Table S1) were included in subsequent steps. This left us with 18S sequences for 273 unique microsporidia species. We also included a single 18S sequence for *Rozella allomycis*, a commonly used outgroup in microsporidia phylogenies(67). A list of accession numbers for all 274 18S sequences and corresponding species can be found in Table S5A.

The 274 18S sequences were aligned with MAFFT, using the default parameters. The alignment was masked using the *MaskAlignment* function from the *DECIPHER* package for R. Bayesian phylogenetic analysis of the masked alignment was performed with MrBayes v 3.2.7, using two MCMC runs, each with one cold and three heated chains, for 10 million generations, sampling every 1000 generations, with the first 2.5 million generations being discarded as burn-in. For this inference, we used the GTR model of substitution, with an *"invgamma"* model for rate variation. Maximum Likelihood phylogenetic analysis of the masked alignment was performed with RAxML-HPC2 using the default parameters and 1000 bootstrap iterations. 50% majority rule consensus trees were generated for both Bayesian and Maximum Likelihood analyses and were annotated using the *ggtree* and *treeio* packages for R. Clade groupings from both trees were compared by computing the fraction of species placed in the same clade in both trees. A complete list of parameters for the MrBayes and RAxML phylogenetic analyses can be found in Table S5B. MAFFT, MrBayes and RAxML were all run on XSEDE accessed via the CIPRES Science Gateway: *https://www.phylo.org/*(68).

Analysis of microsporidia evolutionary relationships

All analysis of evolutionary relationships between the four major clades was done in R. The distribution of host environments was determined by counting the number of species in each clade that are found in each environment, expressed as a percentage of the total number of species in each clade. The distributions of spore volume and polar tube length were determined using the calculated spore volume and calculated polar tube length data (Table S1). In cases where species had multiple spore classes, all were included as separate data points in the analysis. To determine if any hosts were infected by multiple microsporidia species from different

clades, the assortment of hosts for each clade were compared with each other to see if any hosts appeared in multiple clades.

Using the *dist.alignment* function from the *seqinr* package for R, a pairwise sequence similarity matrix for the 36,315 unique species pairs encompassing 270 microsporidia species from the phylogenetic analysis was computed based on the unmasked MAFFT alignment. (The unnamed Pleistophora sp. 1, 2 and 3 were excluded due to ambiguity in species identity between our database and GenBank). The fractions of species pairs within a given sequence similarity range that had at least one shared host taxonym, environment, and infected tissue were computed separately and displayed as distributions with 20 sequence similarity bins. The percentage similarities in spore volume and polar tube length in each species pair were computed separately according to the formula *(larger value / smaller value)*100,* using the calculated spore volume and calculated polar tube length data (Table S1). In cases where species had multiple spore classes, values for each class were averaged prior to the percentage similarity computation. The average percentage similarities in spore volume as well as polar tube length within a given sequence similarity range were then computed and displayed as distributions with 20 sequence similarity bins. For each distribution, species pairs that had at least one species with missing data for the given comparison criterion were excluded.

Data and code availability

All microsporidia species and host data are available in Tables S1 and S2. The most recent version of the microsporidia species data can be at www.placeholder.com. All data analysis was performed using Microsoft Excel and R version 3.6.1 or later accessed via RStudio version 1.2.5019 or later(69). All code used for data analysis can be found on the Microsporidia-Database GitHub repository *https://github.com/bmurareanu/Microsporidia-Database*.

Acknowledgements

We thank Hala Tamim El Jarkass, Alexandra R. Willis, JiHae Jeon, and Robert J. Luallen for providing helpful comments on the manuscript. We thank Alexandra R. Willis for providing the illustrations for figure 2. This work was supported by the Natural Sciences and Engineering Research Council of Canada (Grant #522691522691

and Undergraduate Student Research Awards to B.M., R.S., R.Q., and J.J.) and an Alfred P. Sloan Research Fellowship FG2019-12040 (to A.W.R.).

Author contributions: B.M., R.S., R.Q., J.J, and A. R. designed the project, planned the data analysis, and co-wrote the paper. B.M., R.S., R.Q., and J.J. carried out all of the data curation and analysis.

Competing Interests: The authors declare that they have no competing interests.

References

1. Keeling P. 2009. Five Questions about Microsporidia. PLOS Pathogens 5:e1000489.

2. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivarès CP. 2001. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature 414:450–453.

3. Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. 2010. The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. Nat Commun 1:77.

4. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, Heaps SE, Hirt RP, Martin Embley T. 2013. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. Genome Biol Evol 5:2285–2303.

5. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, Madhani HD, Becnel JJ, Cuomo CA. 2015. Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. 1. Nature Communications 6:1–12.

6. Luallen RJ, Reinke AW, Tong L, Botts MR, Felix M-A, Troemel ER. 2016. Discovery of a Natural Microsporidian Pathogen with a Broad Tissue Tropism in Caenorhabditis elegans. PLoS Pathog 12:e1005724.

7. Wadi L, Reinke AW. 2020. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. PLoS Pathog 16:e1008276.

8. Williams BAP, Hirt RP, Lucocq JM, Embley TM. 2002. A mitochondrial remnant in the microsporidian Trachipleistophora hominis. Nature 418:865–869.

9. Reinke AW, Balla KM, Bennett EJ, Troemel ER. 2017. Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. Nat Commun 8:14023.

10. Franchet A, Niehus S, Caravello G, Ferrandon D. 2019. Phosphatidic acid as a limiting host metabolite for the proliferation of the microsporidium Tubulinosema ratisbonensis in Drosophila flies. 4. Nature Microbiology 4:645–655.

11. Vávra J, Lukeš J. 2013. Chapter Four - Microsporidia and 'The Art of Living Together,' p. 253–319. *In* Rollinson, D (ed.), Advances in Parasitology. Academic Press.

12. Stentiford GD, Feist SW, Stone DM, Bateman KS, Dunn AM. 2013. Microsporidia: diverse, dynamic, and emergent pathogens in aquatic systems. Trends in Parasitology 29:567–578.

13. Buhroo Z. 2018. Microsporidiosis in silkworms with particular reference to mulberry silkworm (Bombyx Mori L.).

14. Trzebny A, Slodkowicz-Kowalska A, Becnel JJ, Sanscrainte N, Dabert M. 2020. A new method of metabarcoding Microsporidia and their hosts reveals high levels of microsporidian infections in mosquitoes (Culicidae). Molecular Ecology Resources 20:1486–1504.

15. Emsen B, Guzman-Novoa E, Hamiduzzaman M, Eccles L, Lacey B, Ruiz-Pérez R, Nasr M. 2015. Higher prevalence and levels of Nosema ceranae than Nosema apis infections in Canadian honey bee colonies. Parasitology research 115.

16. Martín-Hernández R, Bartolomé C, Chejanovsky N, Conte YL, Dalmon A, Dussaubat C, García-Palencia P, Meana A, Pinto MA, Soroker V, Higes M. 2018. Nosema ceranae in Apis mellifera: a 12 years postdetection perspective. Environmental Microbiology 20:1302–1329.

17. Chayaburakul K, Nash G, Pratanpipat P, Sriurairatana S, Withyachumnarnkul B. 2004. Multiple pathogens found in growth-retarded black tiger shrimp Penaeus monodon cultivated in Thailand. Diseases of Aquatic Organisms 60:89–96.

18. Niehus S, Giammarinaro P, Liégeois S, Quintin J, Ferrandon D. 2012. Fly culture collapse disorder: detection, prophylaxis and eradication of the microsporidian parasite Tubulinosema ratisbonensis infecting Drosophila melanogaster. Fly 6:193–204.

19. Sanders JL, Watral V, Kent ML. 2012. Microsporidiosis in Zebrafish Research Facilities. ILAR J 53:106–113.

20. Troemel ER, Félix M-A, Whiteman NK, Barrière A, Ausubel FM. 2008. Microsporidia Are Natural Intracellular Parasites of the Nematode Caenorhabditis elegans. PLoS Biol 6:e309.

21. Balla KM, Luallen RJ, Bakowski MA, Troemel ER. 2016. Cell-to-cell spread of microsporidia causes Caenorhabditis elegans organs to form syncytia. Nat Microbiol 1:16144.

22. Botías C, Martín-Hernández R, Barrios L, Meana A, Higes M. 2013. Nosema spp. infection and its negative effects on honey bees (Apis mellifera iberiensis) at the colony level. Veterinary research 44:25.

23. Gupta SK, Hossain Z, Nanu MM, Mondal K. 2016. Impact of microsporidian infection on growth and development of silkworm Bombyx mori L. (Lepidoptera: Bombycidae). Agriculture and Natural Resources 50:388–395.

24. Li W, Feng Y, Santin M. 2019. Host Specificity of Enterocytozoon bieneusi and Public Health Implications. Trends in Parasitology 35:436–451.

25. Hinney B, Sak B, Joachim A, Kváč M. 2016. More than a rabbit's tale – Encephalitozoon spp. in wild mammals and birds. International Journal for Parasitology: Parasites and Wildlife 5:76–87.

26. Han B, Weiss LM. 2017. Microsporidia: Obligate Intracellular Pathogens within the Fungal Kingdom. Microbiol Spectr 5.

27. Herren JK, Mbaisi L, Mararo E, Makhulu EE, Mobegi VA, Butungi H, Mancini MV, Oundo JW, Teal ET, Pinaud S, Lawniczak MKN, Jabara J, Nattoh G, Sinkins SP. 2020. A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. 1. Nature Communications 11:1–10.

28. Bjørnson S, Oi D. 2014. Microsporidia Biological Control Agents and Pathogens of Beneficial Insects, p. 635–670. *In* Microsporidia. John Wiley & Sons, Ltd.

29. Vávra J, Larsson JIR. 2014. Structure of Microsporidia, p. 1–70. *In* Microsporidia. John Wiley & Sons, Ltd.

30. Han B, Takvorian PM, Weiss LM. 2020. Invasion of Host Cells by Microsporidia. Front Microbiol 11.

31. Zhang G, Sachse M, Prevost M-C, Luallen RJ, Troemel ER, Felix M-A. 2016. A Large Collection of Novel Nematode-Infecting Microsporidia and Their Diverse Interactions with Caenorhabditis elegans and Other Related Nematodes. PLoS Pathog 12:e1006093.

32. Vizoso DB, Ebert D. 2004. Within-host dynamics of a microsporidium with horizontal and vertical transmission: Octosporea bayeri in Daphnia magna. Parasitology 128:31–38.

33. Sprague V. 1977. Annotated List of Species of Microsporidia, p. 31–334. *In* Bulla, LA, Cheng, TC (eds.), Comparative Pathobiology: Volume 2 Systematics of the Microsporidia. Springer US, Boston, MA.

34. Franzen C. 2008. Microsporidia: A Review of 150 Years of Research. TOPARAJ 2:1–34.

35. Sapir A, Dillman AR, Connon SA, Grupe BM, Ingels J, Mundo-Ocampo M, Levin LA, Baldwin JG, Orphan VJ, Sternberg PW. 2014. Microsporidia-nematode associations in methane seeps reveal basal fungal parasitism in the deep sea. Front Microbiol 5.

36. Czaker R. 1997. Wittmannia antarctica N. G., N. Sp. (Nosematidae), a New Hyperparasite in the Antarctic Dicyemid Mesozoan Kantharella antarctica. Journal of Eukaryotic Microbiology 44:438–446.

37. Stentiford GD, Becnel JJ, Weiss LM, Keeling PJ, Didier ES, Williams BAP, Bjornson S, Kent ML, Freeman MA, Brown MJF, Troemel ER, Roesel K, Sokolova Y, Snowden KF, Solter L. 2016. Microsporidia – Emergent Pathogens in the Global Food Chain. Trends Parasitol 32:336–348.

38. Jarkass HTE, Reinke AW. 2020. The ins and outs of host-microsporidia interactions during invasion, proliferation and exit. Cellular Microbiology 22:e13247.

39. Technau U, Scholz CB. 2003. Origin and evolution of endoderm and mesoderm. 7–8. Int J Dev Biol 47:531–539.

40. Oenocytes in insects | Invertebrate Survival Journal.

41. Jaroenlak P, Cammer M, Davydov A, Sall J, Usmani M, Liang F-X, Ekiert DC, Bhabha G. 2020. 3-Dimensional organization and dynamics of the microsporidian polar tube invasion machinery. PLOS Pathogens 16:e1008738.

42. Kudo R. 1920. Notes on Nosema apis Zander. The Journal of Parasitology 7:85–90.

43. Terry RS, Smith JE, Sharpe RG, Rigaud T, Littlewood DTJ, Ironside JE, Rollinson D, Bouchon D, MacNeil C, Dick JTA, Dunn AM. 2004. Widespread vertical transmission and associated host sex–ratio distortion within the eukaryotic phylum Microspora. Proceedings of the Royal Society of London Series B: Biological Sciences 271:1783–1789.

44. Vossbrinck CR, Debrunner-Vossbrinck BA. 2013. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. Folia Parasitologica 52:131–142.

45. Williams BAP, Hamilton KM, Jones MD, Bass D. 2018. Group-specific environmental sequencing reveals high levels of ecological heterogeneity across the microsporidian radiation. Environmental Microbiology Reports 10:328–336.

46. Cormier A, Wattier R, Giraud I, Teixeira M, Grandjean F, Rigaud T, Cordaux R. 2021. Draft Genome Sequences of Thelohania contejeani and Cucumispora dikerogammari, Pathogenic Microsporidia of Freshwater Crustaceans. Microbiol Resour Announc 10.

47. Yakovleva Y, Nassonova E, Lebedeva N, Lanzoni O, Petroni G, Potekhin A, Sabaneyeva E. 2020. The first case of microsporidiosis in Paramecium. Parasitology 147:957–971.

48. Stentiford GD, Ramilo A, Abollo E, Kerr R, Bateman KS, Feist SW, Bass D, Villalba A. 2017. Hyperspora aquatica n.gn., n.sp. (Microsporidia), hyperparasitic in Marteilia cochillia (Paramyxida), is closely related to crustacean-infecting microspordian taxa. Parasitology 144:186–199.

49. Kirk PM. Microsporidia: Unicellular spore-forming protozoan parasites (version Nov 2015). In: Species 2000 & ITIS Catalogue of Life, 2020-09-01 Beta. (Roskov Y.; Ower G.; Orrell T.; Nicolson D.; Bailly N.; Kirk P.M.; Bourgoin T.; DeWalt R.E.; Decock W.; Nieukerken E. van; Penev L.; eds.).

50. Tokarev YS, Huang W-F, Solter LF, Malysh JM, Becnel JJ, Vossbrinck CR. 2020. A formal redefinition of the genera Nosema and Vairimorpha (Microsporidia: Nosematidae) and reassignment of species based on molecular phylogenetics. Journal of Invertebrate Pathology 169:107279.

51. Larsen BB, Miller EC, Rhodes MK, Wiens JJ. 2017. Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life. The Quarterly Review of Biology 92:229–265.

52. Costello MJ. 2016. Parasite Rates of Discovery, Global Species Richness and Host Specificity. Integr Comp Biol 56:588–599.

53. Jorge F, Poulin R. 2018. Poor geographical match between the distributions of host diversity and parasite discovery effort. Proceedings of the Royal Society B: Biological Sciences 285:20180072.

54. Rost-Roszkowska MM, Poprawa I, Kaczmarek Ł. 2013. Autophagy as the cell survival in response to a microsporidian infection of the midgut epithelium of Isohypsibius granulifer granulifer (Eutardigrada: Hypsibiidae). Acta Zoologica 94:273–279.

55. Ardila-Garcia AM, Fast NM. 2012. Microsporidian Infection in a Free-Living Marine Nematode. Eukaryotic Cell 11:1544–1551.

56. Bass D, Czech L, Williams BAP, Berney C, Dunthorn M, Mahé F, Torruella G, Stentiford GD, Williams TA. 2018. Clarifying the Relationships between Microsporidia and Cryptomycota. Journal of Eukaryotic Microbiology 65:773–782.

57. Hejnol A, Martín-Durán JM. 2015. Getting to the bottom of anal evolution. Zoologischer Anzeiger - A Journal of Comparative Zoology 256:61–74.

58. Franzen C, Hösl M, Salzberger B, Hartmann P. 2005. UPTAKE OF ENCEPHALITOZOON SPP. AND VITTAFORMA CORNEAE (MICROSPORIDIA) BY DIFFERENT CELLS. para 91:745–749.

59. Monaghan SR, Kent ML, Watral VG, Kaufman RJ, Lee LEJ, Bols NC. 2009. Animal cell cultures in microsporidial research: their general roles and their specific use for fish microsporidia. In Vitro CellDevBiol-Animal 45:135.

60. 1984. Factors controlling in vitro hatching of Vairimorpha plodiae (Microspora) spores and their infectivity to Plodia interpunctella, Heliothis virescens, and Pieris brassicae. Journal of Invertebrate Pathology 44:192–197.

61. 1973. The infection of nonmosquito hosts by injection with spores of the microsporidan Nosema algerae. Journal of Invertebrate Pathology 22:258–265.

62. Becnel JJ, White SE, Shapiro AM. 2013. Review of microsporidia-mosquito relationships: from the simple to the complex. Folia Parasitologica 52:41–50.

63. Larsson JI. 2013. Molecular versus morphological approach to microsporidian classification. Folia Parasitologica 52:143–144.

64. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, Didier ES, Fan L, Heiman DI, Levin JZ, Young S, Zeng Q, Troemel ER. 2012. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res 22:2478–2488.

65. Pabis K. 2018. What is a moth doing under water? Ecology of aquatic and semi-aquatic Lepidoptera. 419. Knowl Manag Aquat Ecosyst 42.

66. Solter LF, Maddox JV. 1998. Physiological Host Specificity of Microsporidia as an Indicator of Ecological Host Specificity. Journal of Invertebrate Pathology 71:207–216.

67. James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, Stajich JE. 2013. Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. Current Biology 23:1548–1553.

68. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, p. 1–8. *In* 2010 Gateway Computing Environments Workshop (GCE).

69. R Development Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Figure Legends

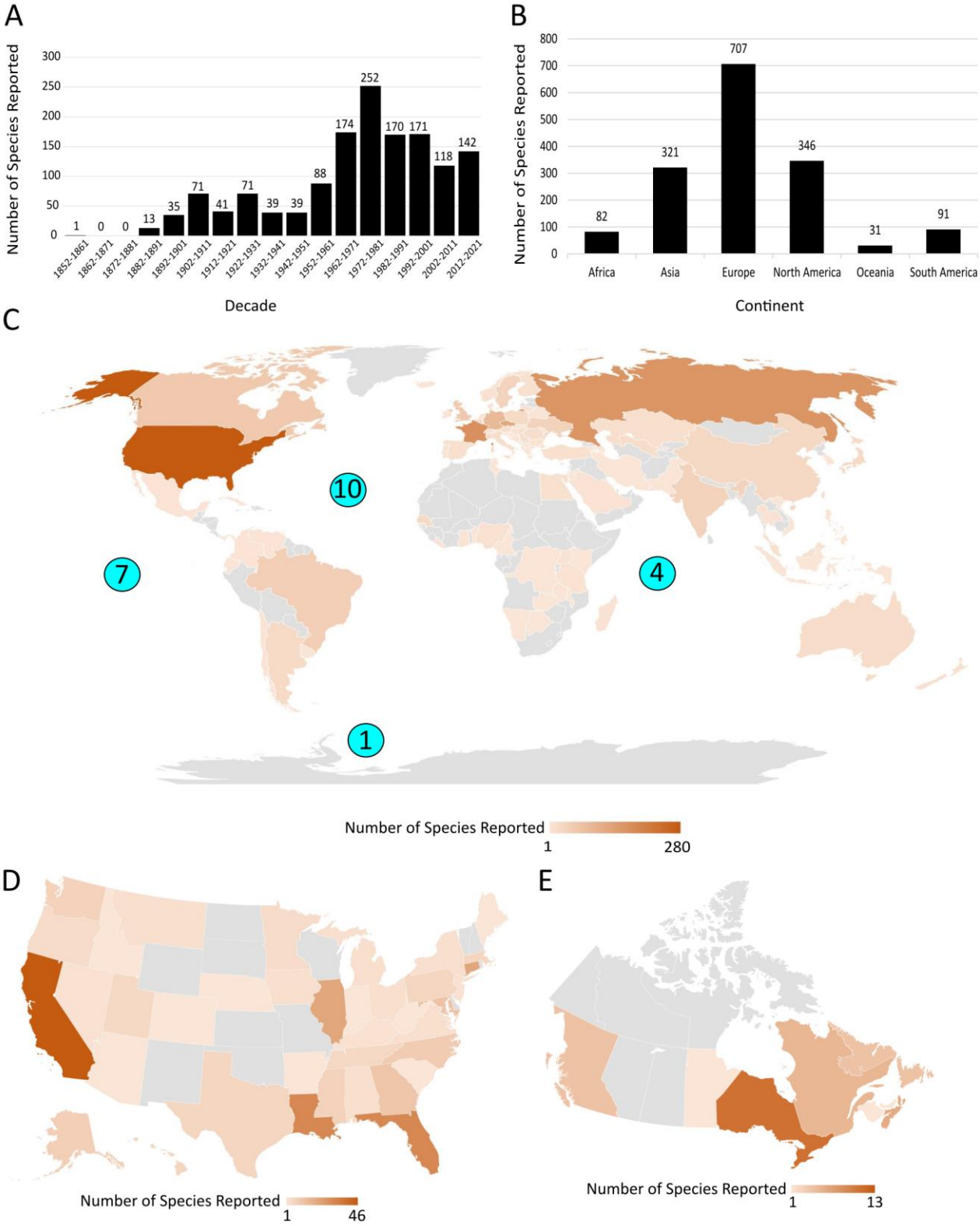
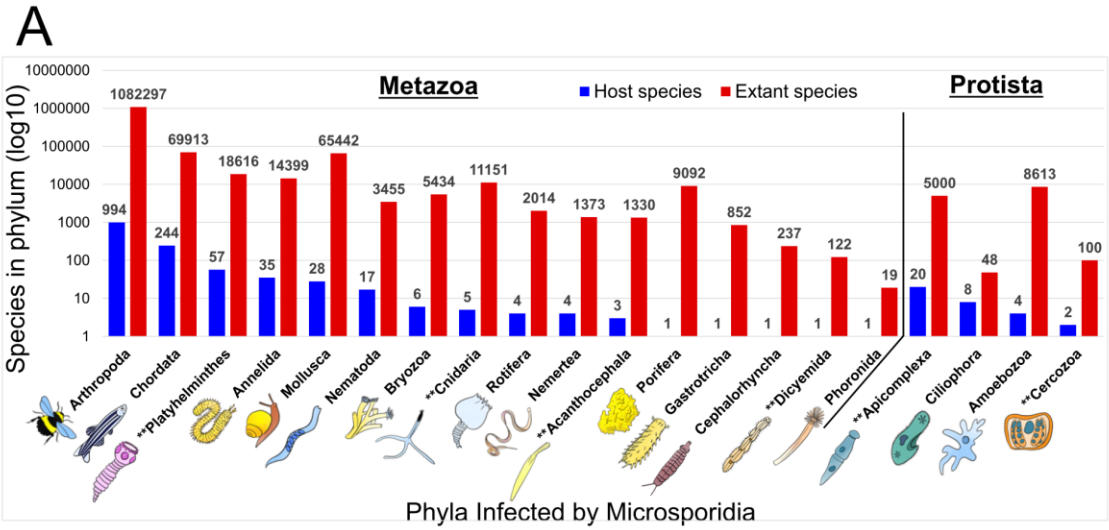
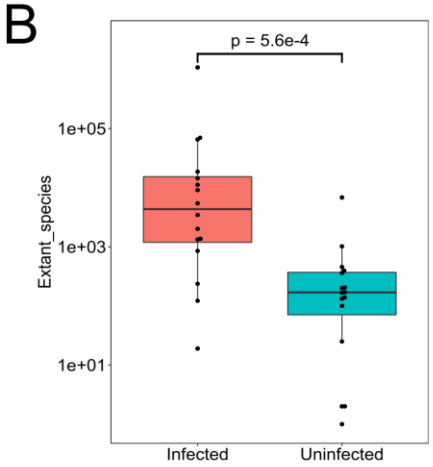
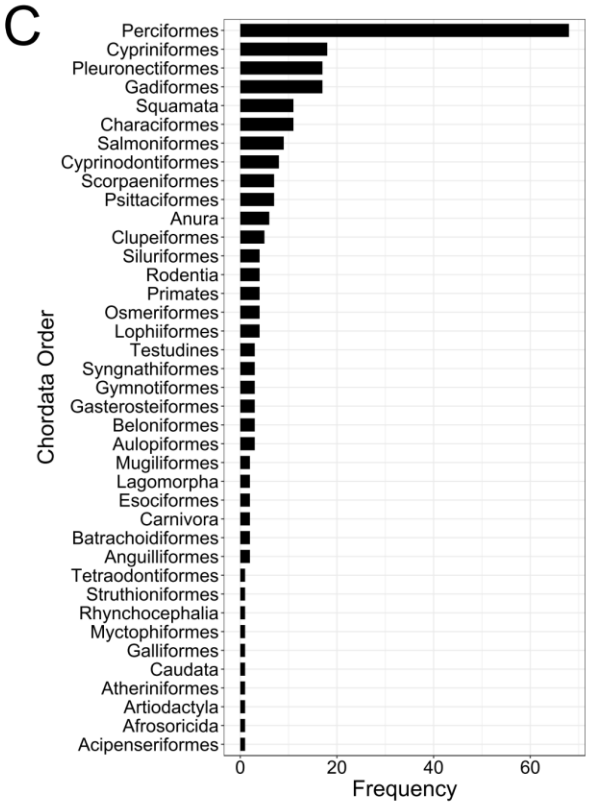
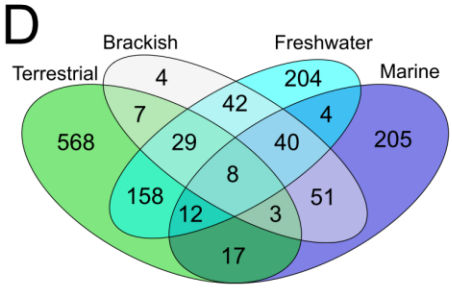

A
Number of Species Reported
300
250
200
150
100
50
0
1
0
0
13
35
71
41
71
39
39
88
174
252
170
171
118
142
1852-1861
1862-1871
1872-1881
1882-1891
1892-1901
1902-1911
1912-1921
1922-1931
1932-1941
1942-1951
1952-1961
1962-1971
1972-1981
1982-1991
1992-2001
2002-2011
2012-2021
Decade
B
Number of Species Reported
800
700
600
500
400
300
200
100
0
82
321
707
346
31
91
Africa
Asia
Europe
North America
Oceania
South America
Continent
C
10
7
4
1
Number of Species Reported
1
280
D
Number of Species Reported
1
46
E
Number of Species Reported
1
13

Fig. 1. History and geography of microsporidia species discovery. (A) Number of new microsporidia species described per decade (n = 1425 microsporidia species). **(B)** Number of microsporidia species found on each continent (n = 1310 microsporidia species). **(C-E)** Geographical location of microsporidia, with each species depicted once for each locality where they have been reported. The number of species found in each locality is depicted as a heatmap, according to the scale at the bottom. Areas in grey indicate that no species have been reported. **(C)** Location of microsporidia species identified, by country and ocean. Numbers in blue circles describe species found in respective oceans (n = 1310 microsporidia species). **(D)** Location of microsporidia species reported by state within the United States (n = 249 microsporidia species). **(E)** Location of microsporidia species reported by province and territory within Canada (n = 45 microsporidia species).


A
Metazoa
Protista
Host species
Extant species
Species in phylum (log10)
10000000
1000000
100000
10000
1000
100
10
1
994
1082297
244
69913
57
18616
35
14399
28
65442
17
3455
6
5434
5
11151
4
2014
4
1373
3
1330
1
9092
1
852
1
237
1
122
1
19
20
5000
8
48
4
8613
2
100
Arthropoda
Chordata
**Platyhelminthes
Annelida
Mollusca
Nematoda
Bryozoa
**Cnidaria
Rotifera
Nemertea
**Acanthocephala
Porifera
Gastrotricha
Cephalorhyncha
**Dicyemida
Phoronida
**Apicomplexa
Ciliophora
Amoebozoa
**Cercozoa
Phyla Infected by Microsporidia


B
p = 5.6e-4
Extant_species
1e+05
1e+03
1e+01
Infected
Uninfected


C
Chordata Order
Perciformes
Cypriniformes
Pleuronectiformes
Gadiformes
Squamata
Characiformes
Salmoniformes
Cyprinodontiformes
Scorpaeniformes
Psittaciformes
Anura
Clupeiformes
Siluriformes
Rodentia
Primates
Osmeriformes
Lophiiformes
Testudines
Syngnathiformes
Gymnotiformes
Gasterosteiformes
Beloniformes
Aulopiformes
Mugiliformes
Lagomorpha
Esociformes
Carnivora
Batrachoidiformes
Anguilliformes
Tetraodontiformes
Struthioniformes
Rhynchocephalia
Myctophiformes
Galliformes
Caudata
Atheriniformes
Artiodactyla
Afrosoricida
Acipenseriformes
0
20
40
60
Frequency


D
Brackish
Freshwater
Terrestrial
Marine
4
204
42
7
4
568
29
40
205
8
158
51
12
3
17

623

Fig 2. Microsporidia infect a diverse range of hosts from all major environments. Host taxonomy and environment was determined for each reported microsporidia species (see methods). **(A)** Distribution of microsporidia hosts by metazoan (left) and protist (right) phyla. The number of host species in each phylum (blue) and the total number of extant species in each phylum (red) are shown. Phyla where the majority of microsporidia hosts are parasites are indicated (**)(n = 1436 host species). **(B)** Number of described species from phyla reported to be infected by microsporidia (red) (n = 16 phyla) and those that have not been reported to be infected by microsporidia (blue) (n = 16 phyla). The Mann-Whitney U test was used to compare mean ranks between the groups. **(C)** Frequency of Chordate orders reported to be infected by microsporidia (n = 39 orders). **(D)** Venn diagram of the major environments that microsporidia hosts dwell in (n = 1352 host species).

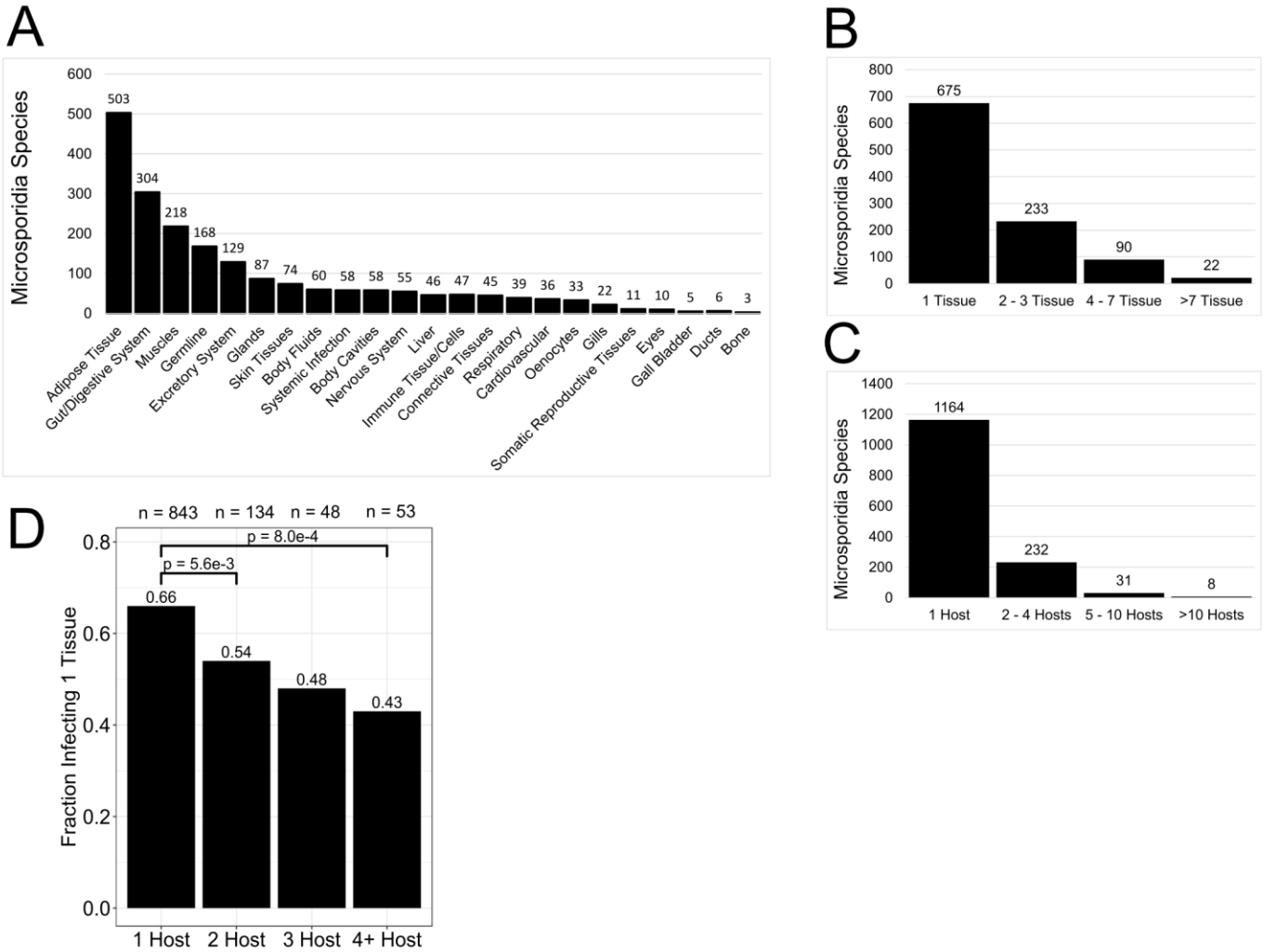
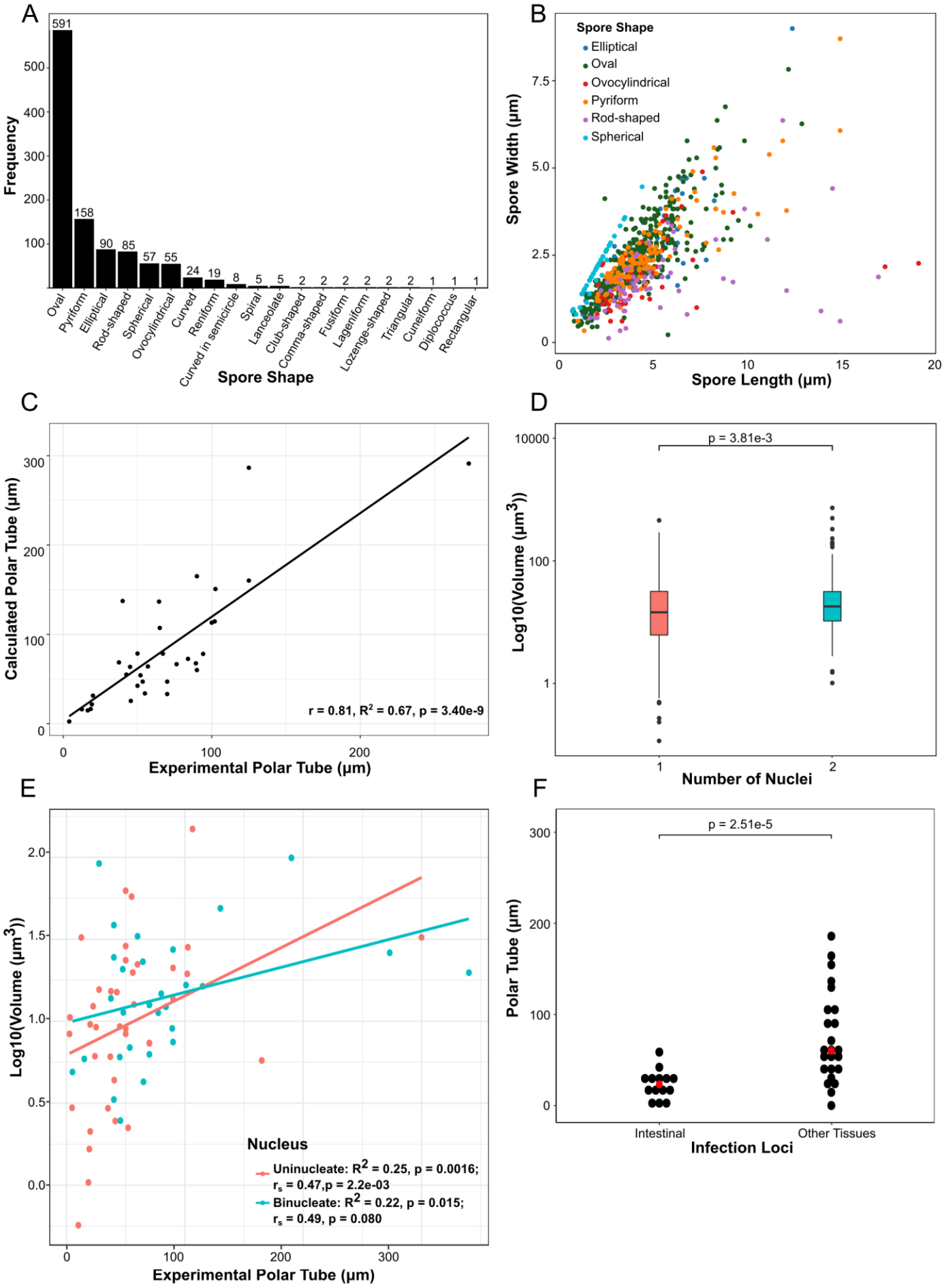


Fig. 3. Microsporidia infect a diversity of tissues and species that infect more hosts tend to infect a broader set of tissues. (A-B) The tissues described to be infected by microsporidia were placed into 23 tissue-

types (see methods). **(A)** The number of microsporidia species reported to infect each tissue type (n = 1123 species). **(B)** Distribution of the number of tissue types infected by each microsporidia species (n = 1020 species). **(C)** Distribution of the number of hosts infected by each microsporidia species (n = 1435 species). **(D)** Proportion of species that infect more than one tissue compared to how many hosts are infected by that species (n = 1078). Species that are described to have general or systemic infections are considered to infect more than one tissue. Significant difference between proportions were determined with pairwise chi-square tests, with Bonferroni p-value correction (α = 0.05/6 comparisons = 8.3×10^{-3}).


A
Frequency
Spore Shape
Oval
Pyriform
Elliptical
Rod-shaped
Spherical
Ovocylindrical
Curved
Reniform
Curved in semicircle
Spiral
Lanceolate
Club-shaped
Comma-shaped
Fusiform
Lageniform
Lozenge-shaped
Triangular
Cuneiform
Diplococcus
Rectangular
B
Spore Shape
Elliptical
Oval
Ovocylindrical
Pyriform
Rod-shaped
Spherical
Spore Width (µm)
Spore Length (µm)
C
Calculated Polar Tube (µm)
Experimental Polar Tube (µm)
r = 0.81, R² = 0.67, p = 3.40e-9
D
p = 3.81e-3
Log10(Volume (µm³))
Number of Nuclei
E
Log10(Volume (µm³))
Experimental Polar Tube (µm)
Nucleus
Uninucleate: R² = 0.25, p = 0.0016;
rs = 0.47,p = 2.2e-03
Binucleate: R² = 0.22, p = 0.015;
rs = 0.49, p = 0.080
F
p = 2.51e-5
Polar Tube (µm)
Intestinal
Other Tissues
Infection Loci

643

Fig. 4. Microsporidia spores display a diversity of shapes, sizes, and polar tube lengths. (A-B) Spore shapes were placed into 20 morphology classes (see methods and Table S3B). **(A)** The frequency of each type of spore shape type (n = 1112 spores). **(B)** Spore length and width dimensions plotted for the six most common spore shapes (n = 816 spores). **(C)** Comparison of experimentally determined and calculated polar tubule lengths (n = 34 spores). **(D)** Median spore volume for uninucleate (n = 496) and binucleate (n = 287) spores. The p-value was determined by the Mann–Whitney U test. **(E)** Comparison between experimental polar filament lengths and spore volume (n = 37 spores for uninucleate and 26 species for binucleate). Correlation was calculated with spearman r_s. **(F)** Relationship between polar tubule length and tissues infected. Median polar tube values are indicated in red. Significance was calculated using a paired T-test between “Intestinal” species and “Other Tissues” species infecting the same host.

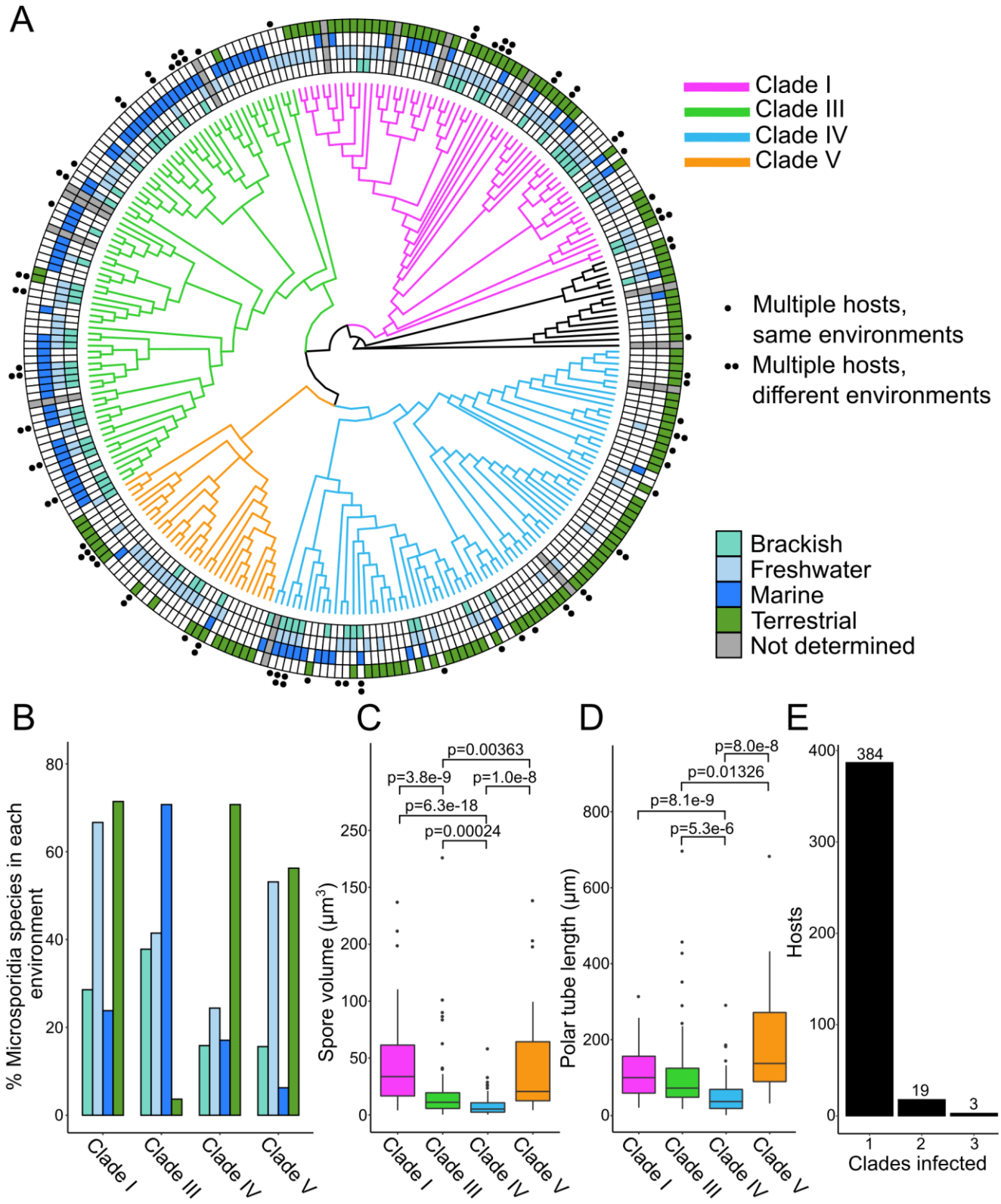


Fig. 5. Bayesian 18S Phylogeny of Microsporidia Species Reveals Four Phenotypically Diverse Clades. **(A)** 18S phylogeny of 273 microsporidia species, inferred under a Bayesian framework (see methods). Clades are colored and number of hosts indicated according to the legends at the right. Species not belonging to any of

the four main clades are colored black. **(B-D)** Analysis of microsporidia phenotypes in each of the four major clades. **(B)** Frequency of host environments by clade. **(C)** Spore volume distribution by clade. **(D)** Calculated polar tube length distribution by clade. Significance evaluated using pairwise Mann-Whitney U-tests, with Bonferroni p-value adjustments applied to correct for multiple testing ($\alpha = 0.05/6 = 8.3 \times 10^{-3}$) **(E)** Number of hosts infected by microsporidia species from one, two or three different clades (n = 406 microsporidia hosts).

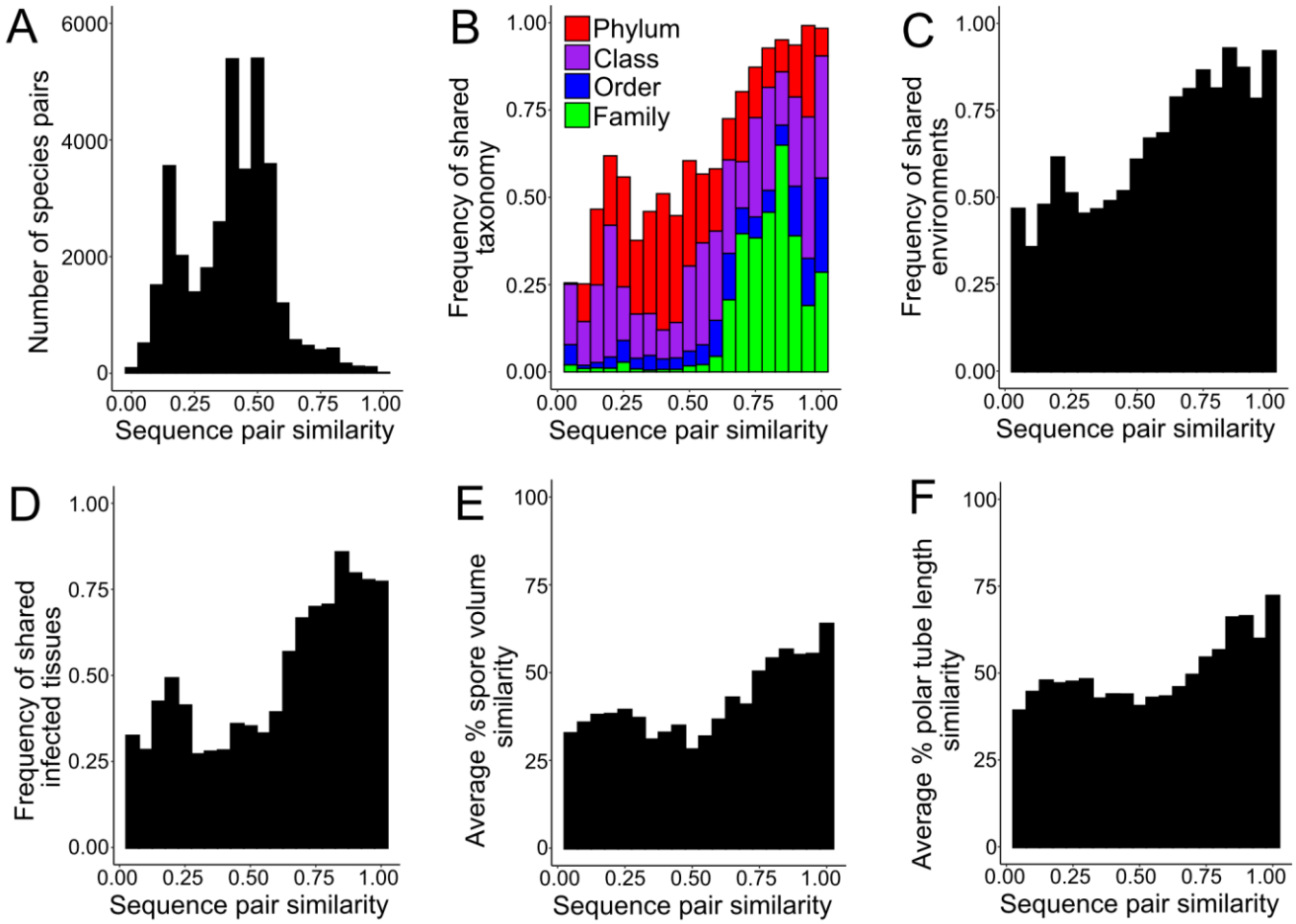


Fig. 6. Evolutionary relatedness of microsporidia phenotypes. (A) Distribution of species pairs according to sequence similarity (n = 36,315 comparisons). **(B-F)** Distribution of sequence pair similarity and frequency of shared **(B)** host taxonomy (n = 34,980 comparisons), **(C)** environments (n = 33,670 comparisons), **(D)** infected tissues (n = 27,028 comparisons), **(E)** difference in spore volume (n = 31,878 comparisons), **(F)** difference in calculated polar tube length (n = 17,955 comparisons).

Supplemental material

Fig. S1. Orders of Arthropods infected by microsporidia.

Fig. S2. Relationship between microsporidia spore dimensions and shape classification.

Fig. S3. Relationship between polar tube length and spore volume.

Fig. S4. Microsporidia host environment and spore morphology are independent of infection transmission mode.

Fig. S5. Bayesian 18S phylogeny of microsporidia species.

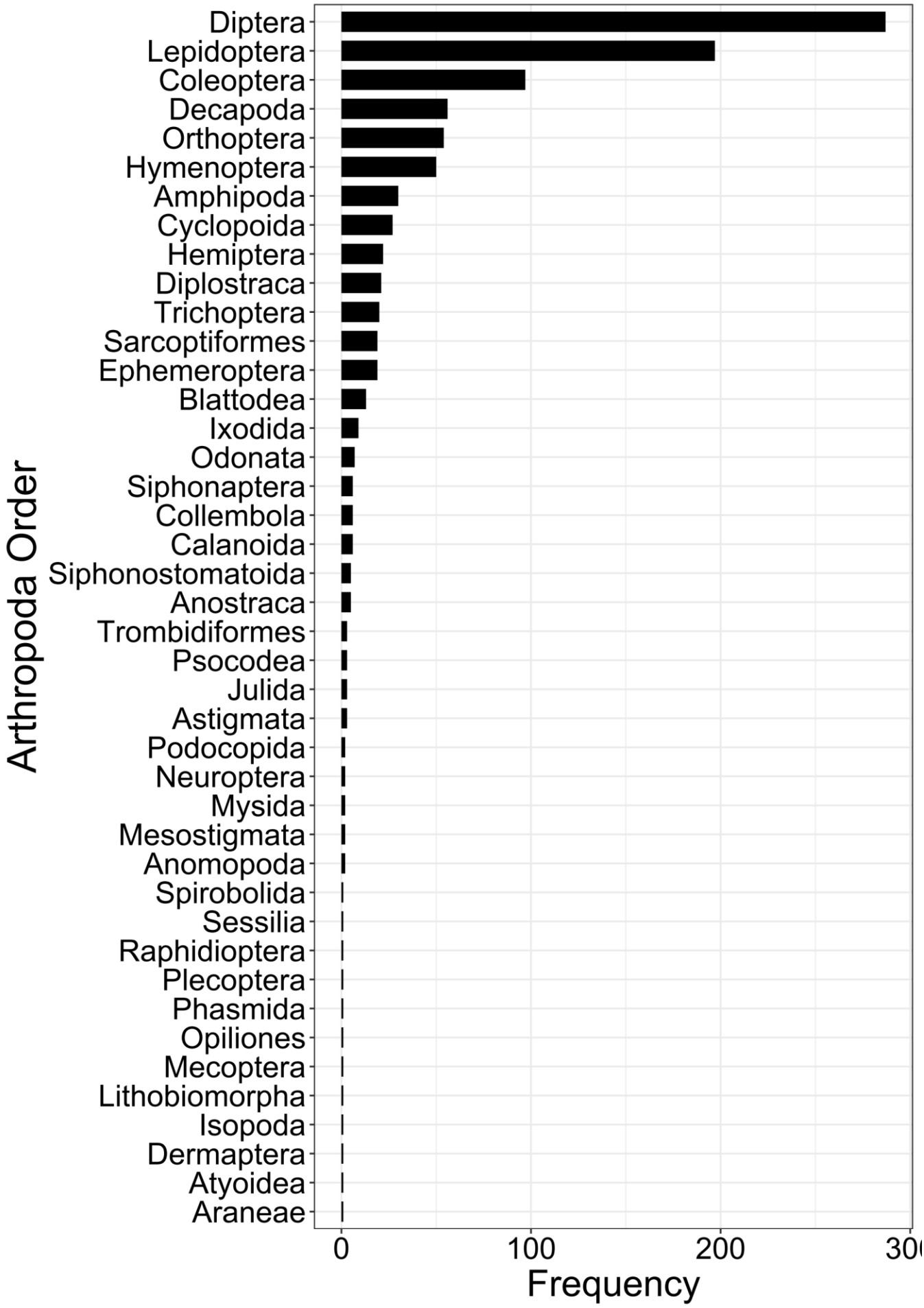
Table S1. Database of microsporidia species attributes.

Table S2. Taxonomical descriptions of microsporidia hosts.

Table S3. Microsporidia host-tissues and spore-shape classifications.

Table S4. Polar tube lengths of species pairs that infect only the intestine and also other tissues in a common host.

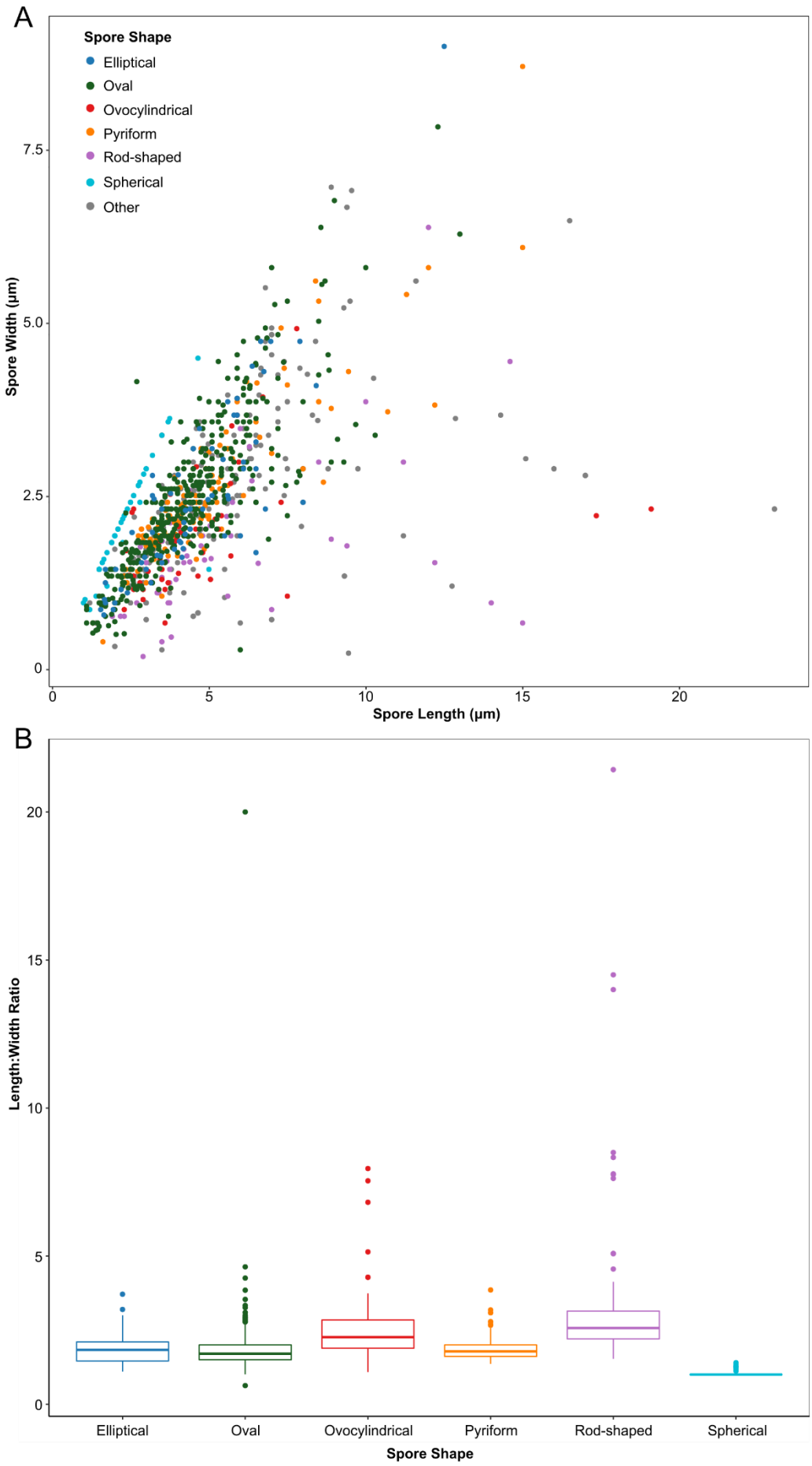
Table S5. Accession numbers for microsporidia DNA sequences and parameters for MrBayes phylogenetic analysis.


Arthropoda Order
Diptera
Lepidoptera
Coleoptera
Decapoda
Orthoptera
Hymenoptera
Amphipoda
Cyclopoida
Hemiptera
Diplostraca
Trichoptera
Sarcoptiformes
Ephemeroptera
Blattodea
Ixodida
Odonata
Siphonaptera
Collembola
Calanoida
Siphonostomatoida
Anostraca
Trombidiformes
Psocodea
Julida
Astigmata
Podocopida
Neuroptera
Mysida
Mesostigmata
Anomopoda
Spirobolida
Sessilia
Raphidioptera
Plecoptera
Phasmida
Opiliones
Mecoptera
Lithobiomorpha
Isopoda
Dermaptera
Atyoidea
Araneae
0
100
200
300
Frequency

687

Fig. S1. Frequency of Arthropod order infections.

Frequency of Arthropod orders reported to be infected by microsporidia (n = 46 orders).

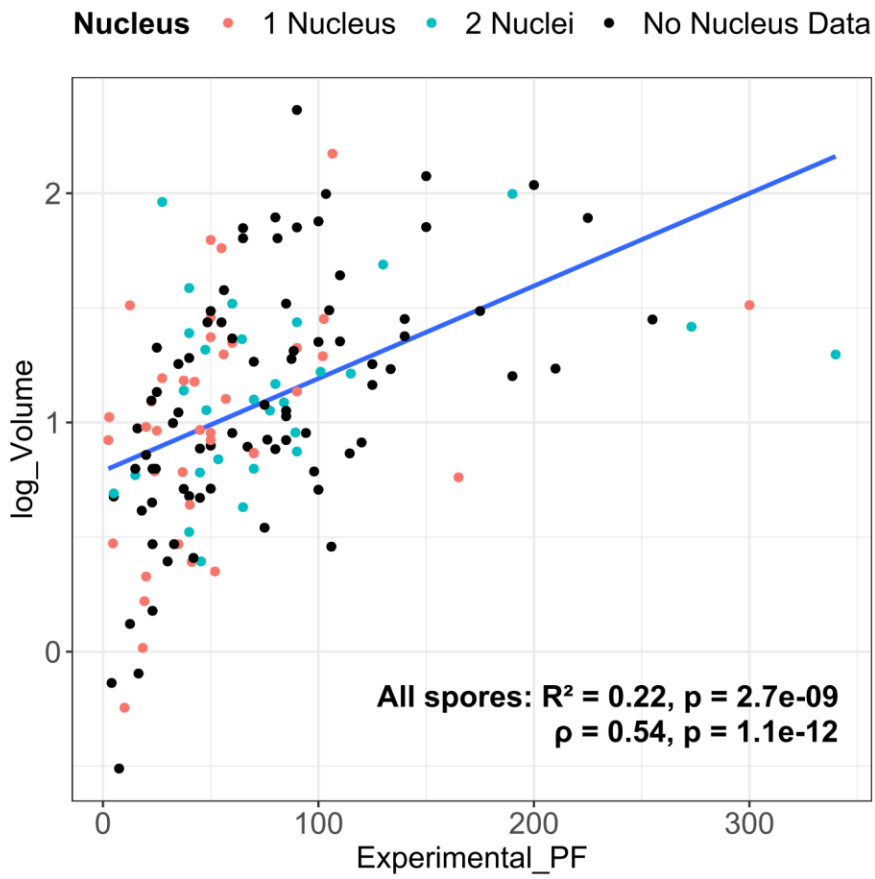

A
Spore Shape
Elliptical
Oval
Ovocylindrical
Pyriform
Rod-shaped
Spherical
Other
Spore Width (µm)
Spore Length (µm)
B
Length:Width Ratio
Elliptical
Oval
Ovocylindrical
Pyriform
Rod-shaped
Spherical
Spore Shape

692

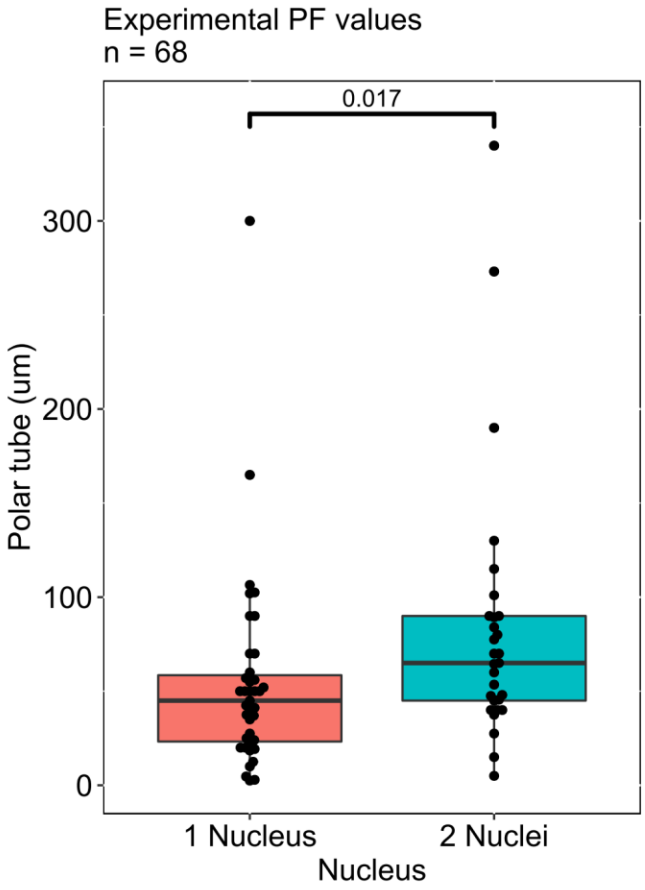
Fig. S2. Relationship between microsporidia spore dimensions and shape classification.

(A) Spore length and width dimensions for the six most common shape groups (Table S4) with the 16 remaining shapes classified as other (n=1032 spores). **(B)** Spore length to width ratios for the six most common shape groups.

A


Nucleus
1 Nucleus
2 Nuclei
No Nucleus Data
log_Volume
Experimental_PF
All spores: R² = 0.22, p = 2.7e-09
ρ = 0.54, p = 1.1e-12

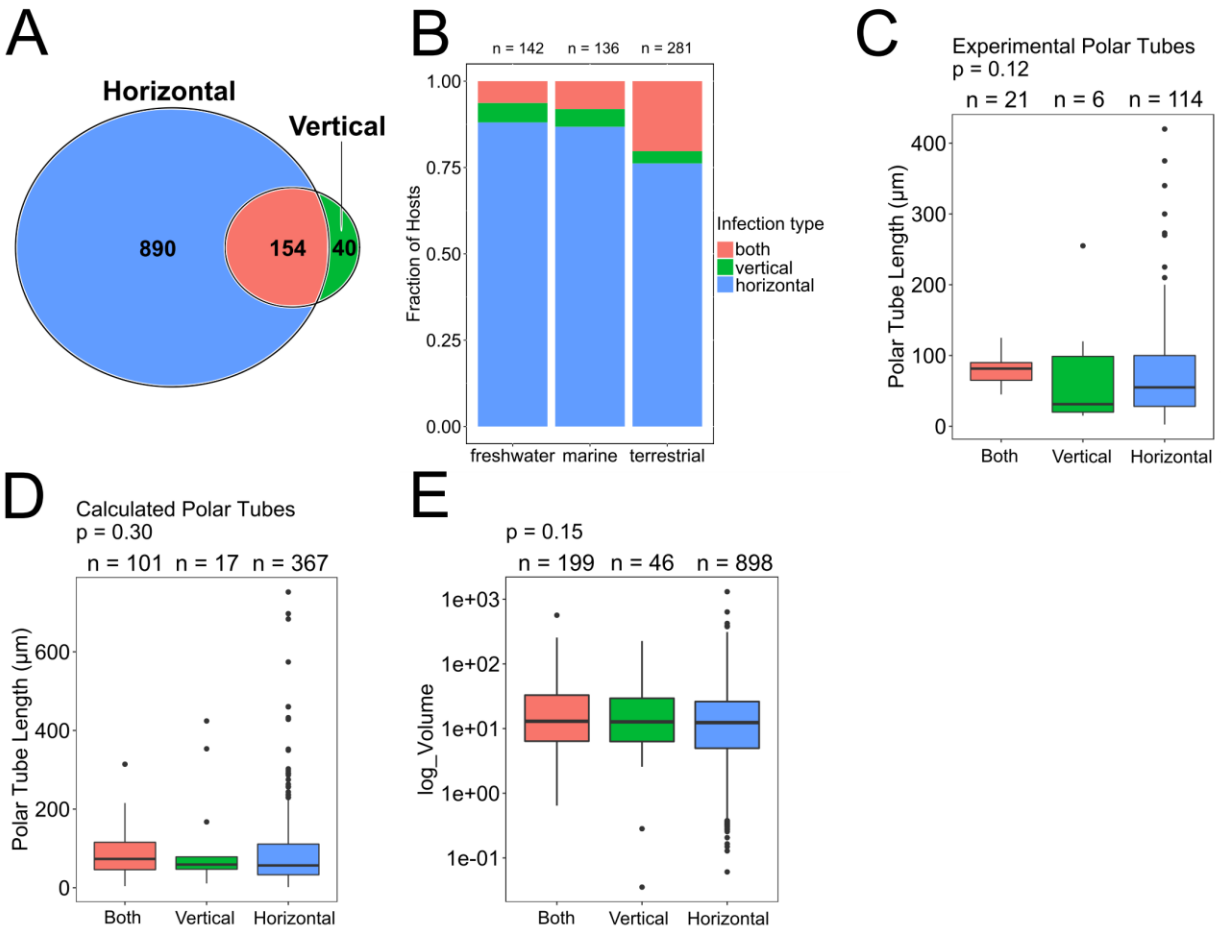
B


Experimental PF values
n = 68
0.017
Polar tube (um)
1 Nucleus
2 Nuclei
Nucleus

698

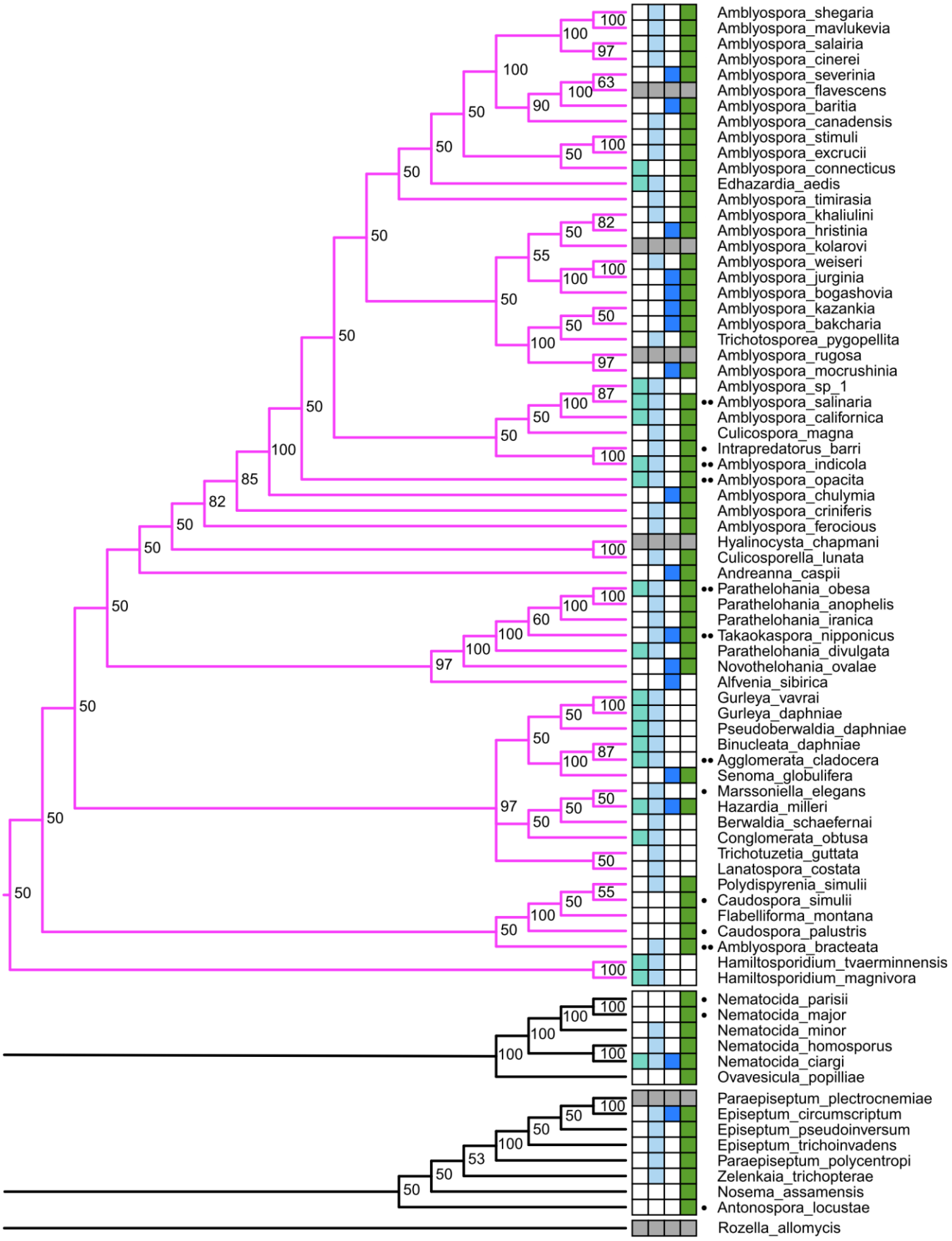
Fig. S3. Relationship between polar tube length, nuclei number and spore volume.

(A) Comparison between experimental polar filament lengths and spore volume for spores of any nuclei number (n = 147). Correlation was calculated with spearman r_s. **(B)** Median spore experimental tube length for uninucleate (n = 39) and binucleate (n = 29) spores. The p-value was determined by the Mann–Whitney U test.

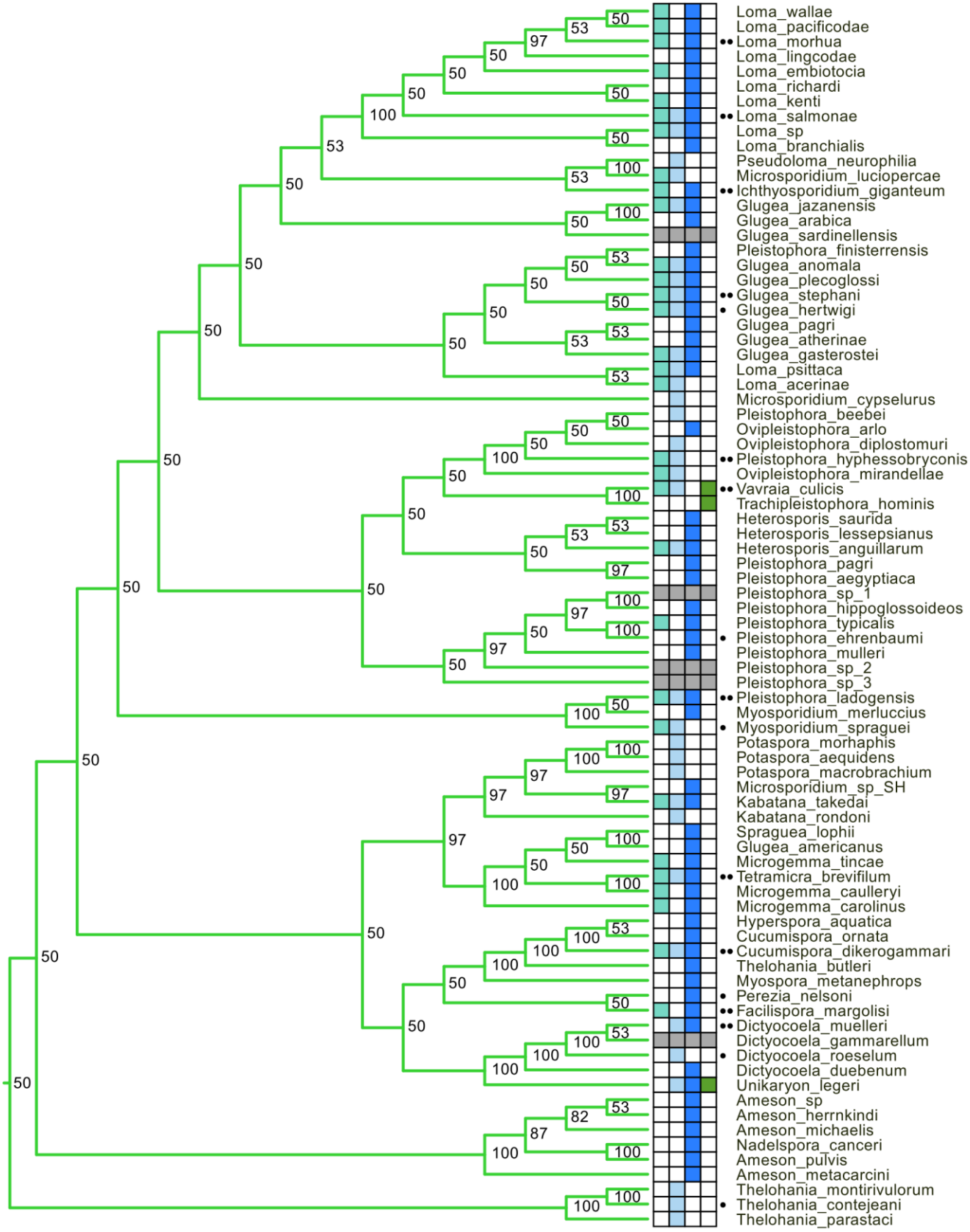

A
Horizontal
Vertical
890
154
40
B
n = 142
n = 136
n = 281
Fraction of Hosts
1.00
0.75
0.50
0.25
0.00
freshwater
marine
terrestrial
Infection type
both
vertical
horizontal
C
Experimental Polar Tubes
p = 0.12
n = 21
n = 6
n = 114
Polar Tube Length (µm)
400
300
200
100
0
Both
Vertical
Horizontal
D
Calculated Polar Tubes
p = 0.30
n = 101
n = 17
n = 367
Polar Tube Length (µm)
600
400
200
0
Both
Vertical
Horizontal
E
p = 0.15
n = 199
n = 46
n = 898
log_Volume
1e+03
1e+02
1e+01
1e+00
1e-01
Both
Vertical
Horizontal

704

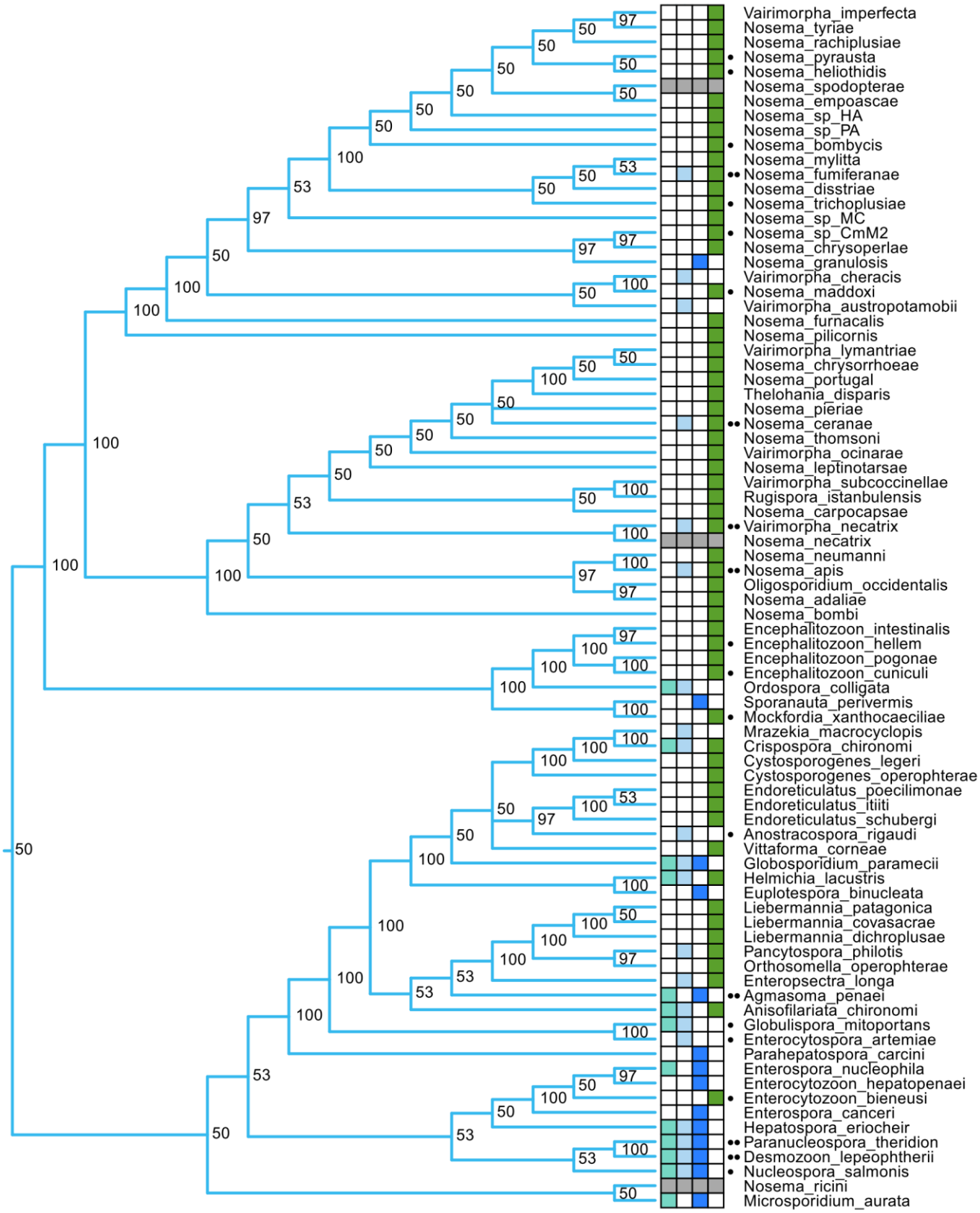
Fig. S4. Microsporidia host environment and spore morphology are independent of infection transmission mode. The transmission mode of each species of microsporidia was inferred from tissues infected, with those that only infect somatic tissues being classified as horizontally transmitted, those that only infect germline tissues being classified as vertically transmitted, and those that infect both types of tissues being classified as both (see methods). **(A)** Venn diagram of transmission mode of microsporidia species (n = 1084 microsporidia species). **(B)** Comparison of infection transmission mode to host environment (n = 559 hosts). Only hosts classified as belonging to a single environment were analyzed. **(C)** Comparison of infection transmission mode to experimentally measured polar tube values (n = 141 spores) and **(D)** calculated polar tube values (n = 485 spores). Species with multiple polar tube values were represented multiple times. **(E)** Comparison of infection transmission mode to spore volume (n = 1143 spores). Species with multiple spore sizes were represented multiple times.


Amblyospora_shegaria
Amblyospora_mavlukevia
Amblyospora_salairia
Amblyospora_cinerei
Amblyospora_severinia
Amblyospora_flavescens
Amblyospora_baritia
Amblyospora_canadensis
Amblyospora_stimuli
Amblyospora_excrucii
Amblyospora_connecticus
Edhazardia_aedis
Amblyospora_timirasia
Amblyospora_khaliulini
Amblyospora_hristinia
Amblyospora_kolarovi
Amblyospora_weiseri
Amblyospora_jurginia
Amblyospora_bogashovia
Amblyospora_kazankia
Amblyospora_bakcharia
Trichotosporea_pygopellita
Amblyospora_rugosa
Amblyospora_mocrushinia
Amblyospora_sp_1
•• Amblyospora_salinaria
Amblyospora_californica
Culicospora_magna
• Intrapredatorus_barri
•• Amblyospora_indicola
•• Amblyospora_opacita
Amblyospora_chulymia
Amblyospora_criniferis
Amblyospora_ferocious
Hyalinocysta_chapmani
Culicosporella_lunata
Andreanna_caspii
•• Parathelohania_obesa
Parathelohania_anophelis
Parathelohania_iranica
•• Takaokaspora_nipponicus
Parathelohania_divulgata
Novothelohania_ovalae
Alfvenia_sibirica
Gurleya_vavrai
Gurleya_daphniae
Pseudoberwaldia_daphniae
Binucleata_daphniae
•• Agglomerata_cladocera
Senoma_globulifera
• Marssoniella_elegans
Hazardia_milleri
Berwaldia_schaefernai
Conglomerata_obtusa
Trichotuzetia_guttata
Lanatospora_costata
Polydispyrenia_simulii
• Caudospora_simulii
Flabelliforma_montana
• Caudospora_palustris
•• Amblyospora_bracteata
Hamiltosporidium_tvaerminnensis
Hamiltosporidium_magnivora
• Nematocida_parisii
• Nematocida_major
Nematocida_minor
Nematocida_homosporus
Nematocida_ciargi
Ovavesicula_popilliae
Paraepiseptum_plectrocnemiae
Episeptum_circumscriptum
Episeptum_pseudoinversum
Episeptum_trichoinvadens
Paraepiseptum_polycentropi
Zelenkaia_trichopterae
Nosema_assamensis
• Antonospora_locustae
Rozella_allomycis

716


Loma_wallae
Loma_pacificodae
•• Loma_morhua
Loma_lingcodae
Loma_embiotocia
Loma_richardi
Loma_kenti
•• Loma_salmonae
Loma_sp
Loma_branchialis
Pseudoloma_neurophilia
Microsporidium_luciopercae
•• Ichthyosporidium_giganteum
Glugea_jazanensis
Glugea_arabica
Glugea_sardinellensis
Pleistophora_finisterrensis
Glugea_anomala
Glugea_plecoglossi
•• Glugea_stephani
• Glugea_hertwigi
Glugea_pagri
Glugea_atherinae
Glugea_gasterostei
Loma_psittaca
Loma_acerinae
Microsporidium_cypselurus
Pleistophora_beebei
Ovipleistophora_arlo
Ovipleistophora_diplostomuri
•• Pleistophora_hyphessobryconis
Ovipleistophora_mirandellae
•• Vavraia_culicis
Trachipleistophora_hominis
Heterosporis_saurida
Heterosporis_lessepsianus
Heterosporis_anguillarum
Pleistophora_pagri
Pleistophora_aegyptiaca
Pleistophora_sp_1
Pleistophora_hippoglossoideos
Pleistophora_typicalis
• Pleistophora_ehrenbaumi
Pleistophora_mulleri
Pleistophora_sp_2
Pleistophora_sp_3
•• Pleistophora_ladogensis
Myosporidium_merluccius
• Myosporidium_spraguei
Potaspora_morhaphis
Potaspora_aequidens
Potaspora_macrobrachium
Microsporidium_sp_SH
Kabatana_takedai
Kabatana_rondoni
Spraguea_lophii
Glugea_americanus
Microgemma_tincae
•• Tetramicra_brevifilum
Microgemma_caulleryi
Microgemma_carolinus
Hyperspora_aquatica
Cucumispora_ornata
•• Cucumispora_dikerogammari
Thelohania_butleri
Myospora_metanephrops
• Perezia_nelsoni
•• Facilispora_margolisi
•• Dictyocoela_muelleri
Dictyocoela_gammarellum
• Dictyocoela_roeselum
Dictyocoela_duebenum
Unikaryon_legeri
Ameson_sp
Ameson_herrnkindi
Ameson_michaelis
Nadelspora_canceri
Ameson_pulvis
Ameson_metacarcini
Thelohania_montirivulorum
• Thelohania_contejeani
Thelohania_parastaci

717


Vairimorpha_imperfecta
Nosema_tyriae
Nosema_rachiplusiae
Nosema_pyrausta
Nosema_heliothidis
Nosema_spodopterae
Nosema_empoascae
Nosema_sp_HA
Nosema_sp_PA
Nosema_bombycis
Nosema_mylitta
Nosema_fumiferanae
Nosema_disstriae
Nosema_trichoplusiae
Nosema_sp_MC
Nosema_sp_CmM2
Nosema_chrysoperlae
Nosema_granulosis
Vairimorpha_cheracis
Nosema_maddoxi
Vairimorpha_austropotamobii
Nosema_furnacalis
Nosema_pilicornis
Vairimorpha_lymantriae
Nosema_chrysorrhoeae
Nosema_portugal
Thelohania_disparis
Nosema_pieriae
Nosema_ceranae
Nosema_thomsoni
Vairimorpha_ocinarae
Nosema_leptinotarsae
Vairimorpha_subcoccinellae
Rugispora_istanbulensis
Nosema_carpocapsae
Vairimorpha_necatrix
Nosema_necatrix
Nosema_neumanni
Nosema_apis
Oligosporidium_occidentalis
Nosema_adaliae
Nosema_bombi
Encephalitozoon_intestinalis
Encephalitozoon_hellem
Encephalitozoon_pogonae
Encephalitozoon_cuniculi
Ordospora_colligata
Sporanauta_perivermis
Mockfordia_xanthocaeciliae
Mrazekia_macrocyclopis
Crispospora_chironomi
Cystosporogenes_legeri
Cystosporogenes_operophterae
Endoreticulatus_poecilimonae
Endoreticulatus_itiiti
Endoreticulatus_schubergi
Anostracospora_rigaudi
Vittaforma_corneae
Globosporidium_paramecii
Helmichia_lacustris
Euplotespora_binucleata
Liebermannia_patagonica
Liebermannia_covasacrae
Liebermannia_dichroplusae
Pancytospora_philotis
Orthosomella_operophterae
Enteropsectra_longa
Agmasoma_penaei
Anisofilariata_chironomi
Globulispora_mitoportans
Enterocytospora_artemiae
Parahepatospora_carcini
Enterospora_nucleophila
Enterocytozoon_hepatopenaei
Enterocytozoon_bieneusi
Enterospora_canceri
Hepatospora_eriocheir
Paranucleospora_theridion
Desmozoon_lepeophtherii
Nucleospora_salmonis
Nosema_ricini
Microsporidium_aurata

718

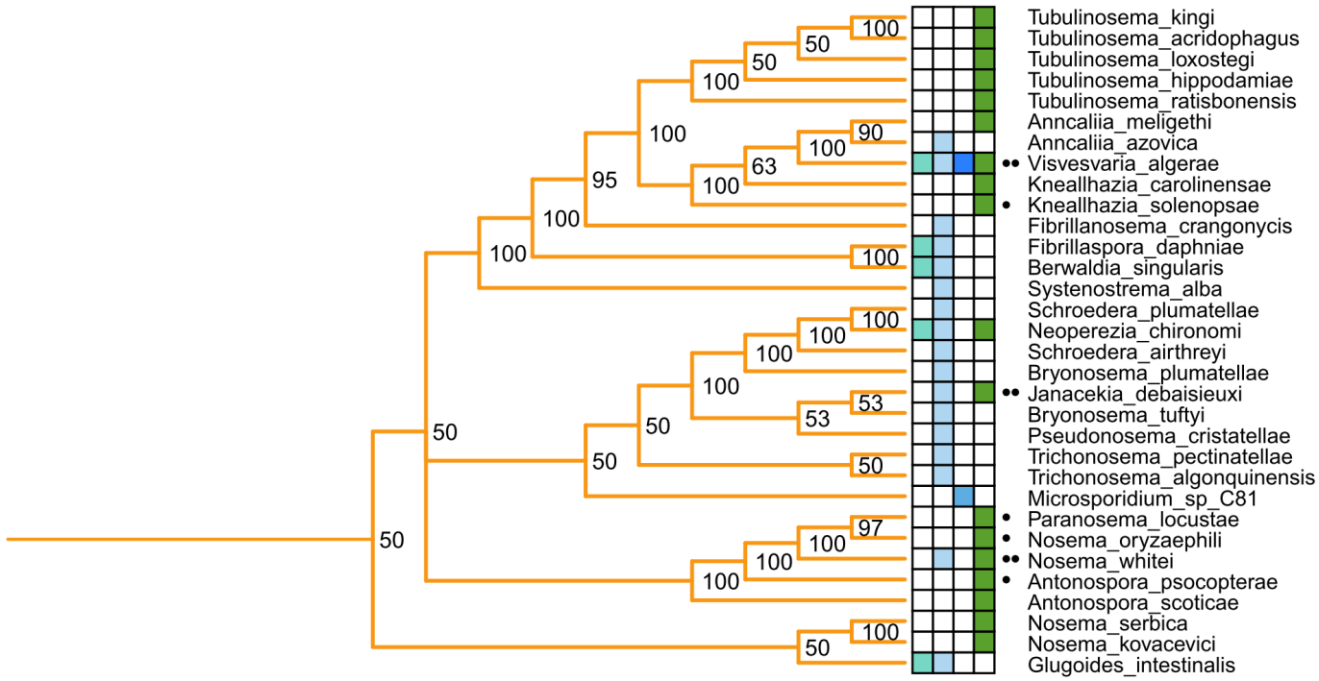


Fig. S5. Bayesian 18S phylogeny of microsporidia species. Phylogenetic tree of 273 microsporidia species and *Rozella allomycis*. The tree was generated and labelled for clades and environments the same as in Fig 5. The species name of each branch is shown on the right, and the Bayesian posterior probability support value is shown at each node.