Revisiting the phylogeny of microsporidia
Eunji Park*, Robert Poulin
Department of Zoology, University of Otago, 340 Great King Street, Dunedin 9016, New Zealand

A R T I C L E  I N F O
Article history:
Received 16 October 2020
Received in revised form 15 February 2021
Accepted 15 February 2021
Available online 21 April 2021

Keywords:
Microsporidia
Rozellida
Opisthosporidia
Intracellular parasites
Host range
ssrRNA

A B S T R A C T
Canonical microsporidians are a group of obligate intracellular parasites of a wide range of hosts comprising ~1,300 species of >220 genera. Microsporidians are related to fungi, and many characterised and uncharacterized groups closely related to them have been discovered recently, filling the knowledge gaps between them. These groups assigned to the superphylum Opisthosporidia have provided several important insights into the evolution of diverse intracellular parasitic lineages within the tree of eukaryotes. The most studied among opisthospidians, canonical microsporidians, were known to science more than 160 years ago, however, the classification of canonical Microsporidia has been challenging due to common morphological homoplasy, and accelerated evolutionary rates. Instead of morphological characters, ssrRNA sequences have been used as the primary data for the classification of canonical microsporidians. Previous studies have produced a useful backbone of the microsporidian phylogeny, but provided only some nodal support, causing some confusion. Here, we reconstructed phylogenetic trees of canonical microsporidians using Bayesian and Maximum Likelihood inferences. We included rRNA sequences of 126 described/named genera, by far the broadest taxon coverage to date. Overall, our trees show similar topology and recovered four of the five main clades demonstrated in previous studies (Clades 1, 3, 4 and 5). Family level clades were well resolved within each major clade, but many were discordant with the recently revised classification. Therefore, revision and some reshuffling, especially within and between Clades 1 and 3 are required. We also reconstructed phylogenetic trees of Opisthosporidia to better integrate the evolutionary history of canonical microsporidians in a broader context. We discuss several traits shared only by canonical microsporidians that may have contributed to their striking ecological success in diverse metazoans. More targeted studies on the neglected host groups will be of value for a better understanding of the evolutionary history of these interesting intracellular parasites.

1. Introduction

‘Canonical (=classical/derived/higher)’ microsporidians are a monophyletic group of highly specialised intracellular parasites that infect a wide range of hosts. As the name implies, canonical microsporidians include lineages that were first discovered and ones that share common characteristics with those lineages, differentiating them from ‘microsporidia-like organisms’ that are morphologically and genetically similar to them but distinct. All canonical microsporidians share several common characteristics including a compact genome (with some variability; see Wadi and Reinke, 2020), highly reduced mitochondria (mitosomes), the presence of ADP/ATP transporters and well developed polar tubes (Vossbrinck et al., 2014; Dean et al., 2016; Tamim El Jarkass and Reinke, 2020).

The existence of Microsporida has been known since the 19th century due to their devastating impact on animals of economic importance such as silkworms and fishes (Pasteur, 1870; Sandholzer et al., 1945; Naegeli, 1957). During the last >160 years, more than 1,300 species have been described from >220 genera (Franzen, 2008; Becnel et al., 2014), including at least 17 human-infecting species (Stentiford et al., 2016). Microsporidians are common in arthropod and chordate hosts, but they have been found from almost all animal phyla (Stentiford et al., 2013b; Becnel et al., 2014; Snowdon, 2014), including Acanthocephala (de Buron et al., 1990), Annelida (Larsson, 1992), Bryozoa (Desser et al., 2004), Cnidaria (Clausen, 2000), Gastrotricha (Manylov, 1999), Kinorhyncha (Adrianov and Rybakov, 1992), Mesozoa (Czaker, 1997), Mollusca (Sagrìsta et al., 1998), Nematoda (Ardila-Garcia and Fast, 2012), Phoronida (Temereva and Sokolova, 2018), Platyhelminthes (Levron et al., 2005), and Rotifera (Gorbunov and Kosova, 2001). Although this is rare (based on our current knowledge), some canonical microsporidians have been found in Ciliophora, so far the only known non-metazoan host.
group for these parasites (Foissner and Foissner, 1995; Fokin et al., 2008).

Despite their well recognized diversity, resolving the phylogenetic position of canonical microsporidians within the tree of life has been challenging, especially due to fast evolutionary rates in ssrRNA gene(s) that often cause a long-branch attraction (LBA) problem (Lartillot et al., 2007). Their phylogenetic affinity to Fungi has now been widely accepted (Edlin et al., 1996; Keeling and Doolittle, 1996), however their phylogenetic placement in relation to their close relatives and with or within Fungi still remain to be fully resolved. About a decade ago, it was shown that a clade containing Rozella (parasites of Chytridiomycetes, Blastocladiomycetes, and Oomycetes) and many unidentified environmental sequences formed a monophyletic group closely related to Fungi, and this group has been referred to as Rozellida (Lara et al., 2010). Later, the phylum Rozellomycota (=Cryptomycota) was proposed for this group (Jones et al., 2011; Corsaro et al., 2014b). After the characterization of Aphelida (parasites of algae), the Superphylum ‘Opisthosporidia’ was proposed to encompass Aphelida, Rozellida, and Microsporidia (the so-called ARM clade), which are closely related and branched near the base of the fungal radiation (James et al., 2013; Karpov et al., 2014).

Some Microsporidia-like organisms such as Paramicrosporidium, Miitosporidium, and Nucleophagho have been both morphologically and genetically characterised, providing important insights into the specialisation and evolutionary trait reduction within Opisthosporidia (Corsaro et al., 2014a; Haag et al., 2014; Galindo et al., 2018). Bass et al. (2018) proposed the concept of ‘expanded Microsporidia’ to include all these groups together with canonical Microsporidia, which are in a robust monophyletic group sister to Rozella. These Microsporidia-like organisms branched between Rozella and canonical Microsporidia, having short branches in the ssrRNA trees, and therefore they have been referred to as short-branch Microsporidia (SB-Microsporidia) in contrast to canonical Microsporidia, which have long branches (therefore canonical Microsporidia were named ‘LB-Microsporidia’ in Bass et al., 2018). Metchnikovellids and Chytridiopsis, which have also long been known, but only recently genetically characterised, were confirmed as the closest relatives of canonical Microsporidia (Mikhailov et al., 2017; Galindo et al., 2018; Corsaro et al., 2019).

The ribosome is essential to all life and therefore core units are still conserved in canonical Microsporidia, even though they are significantly reduced in size and are highly divergent (Peyretaillade et al., 1998; Bowman et al., 2020). Canonical microsporidians have a prokaryote-like rRNA with a fused lsrRNA-5.8S rRNA (Vossbrinck and Woese, 1986). A comparison of the secondary structure of the rRNA of microsporidians and their relatives shows that extreme reduction occurred only in the lineage of canonical microsporidians (Corsaro et al., 2019). Metchnikovellids and Chytridiopsis also have long branches in ssrRNA trees, but their ribosomal DNAs are similar to that of other eukaryotes in structure and size (Corsaro et al., 2019). Despite the differences in size, these highly conserved orthologous fragments of rRNA provide valuable information on phylogenetic relationships among these divergent groups. Within canonical microsporidians, ssrRNA sequences have been used as primary data for higher classification as morphological characters traditionally used for identification show common homology (Stentiford et al., 2013b; Vossbrinck et al., 2014).

Previous phylogenetic studies based on ssrRNA sequences have played an important role in the classification of canonical Microsporidia. The first ssrRNA tree of Microsporidia was inferred with five species using the maximum parsimony (MP) method (Vossbrinck et al., 1993). Later, Vossbrinck and Debrunner-Vossbrinck (2005) constructed neighbour joining (NJ) and MP trees based on ssrRNA sequences of 125 species from 56 genera, proposing three classes based on the dominant host habitat of each group: Terresporidia, Aquasporidia and Marinosporidia. More recently, Vossbrinck et al. (2014) inferred maximum likelihood (ML) and MP trees with an improved taxon sampling (71 species from 63 genera) showing five main clades; Clades 1–5. Clades 1 and 3 correspond to Aquasporidia, Clades 2 and 4 correspond to Terresporidia, and Clade 5 corresponds to Marinosporidia. Although ecological heterogeneity across major lineages was demonstrated using environmental sequences (Williams et al., 2018), the five clade system has been widely used for the classification of canonical Microsporidia. Indeed, the classification of canonical Microsporidia has recently been revised to accommodate the five major clades (by Tokarev and Issi in Wijayawardene et al., 2020). In the revised classification, orders Neopereziida, Ovavesiculida, and Amblyosporida, were newly established to accommodate Clades 1, 2 and 3, respectively, and orders Nosematida and Glugeida were revised for Clades 4 and 5, respectively.

At present, reporting the ssrRNA sequence (or a sequence of a longer region of rRNA; ssrRNA-internal transcribed spacer (ITS)-lsrRNA) is regarded as an essential part of species description (Stentiford et al., 2013b; Vossbrinck et al., 2014), and the number of rRNA sequences in GenBank has been increasing. Here, we make use of an expanded number of rRNA sequences to better understand the phylogeny of canonical microsporidians. For this purpose, we reconstruct the most up-to-date and most comprehensive genus level phylogeny of canonical Microsporidia with available sequences. Also, we infer the phylogeny of Opisthosporidia to discuss the evolution of canonical Microsporidia within a broader context to provide insights into the origin and diversification of this interesting parasite group. Our Bayesian and ML trees recovered all five of the major clades except for Clade 2 from Vossbrinck et al. (2014). This result suggests the need to revise the current classification of canonical Microsporidia. However, because our current knowledge of the diversity of microsporidians is far from complete, more major lineages might be uncovered and the relationships among them will be better resolved in the future.

2. Materials and methods

2.1. Compiling genetic data

2.1.1. Microsporidia

We aimed to reconstruct phylogenetic trees of canonical Microsporidia that are as complete as possible at the genus level. Bencel et al. (2014) listed 200 formally described generic names within canonical Microsporidia. Among these, rRNA sequences (ssrRNA or ssrRNA-ITS-lsrRNA sequences) of 104 genera were available in GenBank (Supplementary Table S1). For these genera, we included at least one rRNA sequence per genus. Although they have not been formally described, we also included Paramuclespora and Visvesvaria in our analyses because rRNA sequences were also available for these genera in GenBank. In addition, rRNA sequences of 20 newly described genera since Bencel et al. (2014) were also included (Alternosema, Apotaspora, Cambaraspora, Conglomerata, Dictyocoela, Fibrillaspora, Globulispora, Hyperspora, Jirovecia, Myrmecomorba, Nematocenator, Obruspora, Pancytospora, Paradoxium, Parahepatospora, Percutemincola, Pseudobervaldina, Pseudokatiana, Rugispora and Trichotospora; see Supplementary Table S1 for references). Finally, rRNA sequences under the names ‘Microsporidium sp.’ and ‘Microsporidia sp.’ were also added to our dataset since many sequences obtained from diverse hosts are provisionally registered under these names, and these may represent distinct and previously unrecognised lineages within the tree of Microsporidia. For these two provisional groups, sequences that were too short (<500 bp) were excluded and only sequences
with known hosts were included for further analyses. Also, highly similar sequences generated by a single study (e.g., many sequences obtained from amphipods from Lake Baikal) were reduced to one or a few representative sequences. Thus, 60 rRNA sequences of Microsporidium sp. and 16 sequences of Microsporidia sp. were added to the dataset. As a result, a total of 220 rRNA sequences including 126 described/named canonical microsporidians genera were used for further analyses. In addition, 13 rRNA sequences of close relatives within opisthosporigonians were included as outgroups.

2.1.2. Opisthosporidia

ssrRNA sequences were also compiled for the phylogenetic tree of microsporidians and their relatives (Opisthosporidia). Sequences representing canonical Microsporidia, Metchnikovellida, Chytridiopsis, Paramicrosporidium, Mitosporidium, Nucleophaga, and some representative groups of environmental sequences of SB-Microsporidium in Bass et al. (2018), Aphelida, Rozella, and some fungi were included. ssrRNA sequences of Holozooa and Nuclearidae were also included as outgroups. In total, 94 ssrRNA sequences were used in our analyses (Supplementary Table S2).

2.2. Phylogenetic analysis

For canonical microsporidians, both ssrRNA and lsrRNA rRNA sequences were used when the lsrRNA region was available. The ITS region was excluded because this region was too divergent across microsporidians. ssrRNA and lsrRNA sequences were aligned with the E-INS-i algorithm using MAFFT v7.450 implemented in Geneious prime (Katoh and Standley, 2013). Removing ambiguous sites can reduce LBA problems (Qu et al., 2017; Ranwez and Chantret, 2020). Therefore, ambiguous sites were eliminated in groups with the least restrictive setting (Castresana, 2000), and then ssrRNA and lsrRNA were concatenated. For the tree of Opisthosporidia, only the ssrRNA region was used for analyses and the same alignment and refining procedures described above were applied. The best-fitting model of nucleotide evolution for each dataset (canonical Microsporidia and Opisthosporidia) was determined based on the corrected Akaike information criterion (AICc) using jModelTest v2.1.6 (Darriba et al., 2012), which was conducted through the CIPRES and Opisthosporidia) was determined based on the corrected nucleotide evolution for each dataset (canonical Microsporidia genera were used for further analyses. In addition, 13 rRNA sequences of close relatives within opisthosporigonians were included as outgroups.

2.3. Compiling data on host and habitat of canonical microsporidians

Information on host species (and higher taxonomic groups) and habitat of each microsporidian species that was actually used in our analyses was compiled for canonical microsporidians. Notably, some genera include many species and they may infect a different group of hosts and inhabit different habitats (e.g. Encephalitozoon). Some species are generalists infecting distantly related taxa (e.g. Nosema and Vairimirmpora), or alternate between different group of hosts during their life cycle (e.g. Amblyospora). Because uneven taxon sampling could affect our interpretation of the phylogeny (Heath et al., 2008), only one (or a few) representative sequences per genus were included. Habitats were categorised into five groups; Marine (M), Marine-Freshwater (MF), Freshwater (F), Freshwater-Terrestrial (FT), and Terrestrial (T). FT groups include species that spend considerable portions of their lifecycle both in freshwater and on land, such as parasites of many dipteran species, whose larval stage must develop in freshwater (Becnel and Andreadis, 2014).

2.4. Data accessibility

The data alignments used in this study and the supplementary figures and tables are available in MendeleyData (https://doi.org/10.17632/7jrzkbyp63.1).

3. Results and discussion

3.1. The phylogeny of canonical Microsporidia

Our Bayesian and ML trees reconstructed with 220 sequences from 126 named genera represent more than half of the known diversity of canonical microsporidians at the genus level (Fig. 1A–C; full trees in Supplementary Figs. S1 and S2). Overall, both trees were highly congruent and recovered four of the five main clades from Vossbrinck et al. (2014), Clades 1, 3, 4 and 5, although the phylogenetic relationships among them were not well resolved in our analyses. Clade 2 was not recovered as a monophyletic group. Below, we briefly summarise several features of each main clade and discuss patterns and causes of the discrepancy between our trees and those of previous studies.

3.1.1. Clade 5: fish and crustacean hosts from aquatic habitats

Clade 5 comprises mostly aquatic (marine, marine-freshwater, and freshwater) species (Fig. 1A). Also, this clade includes most of the fish-infecting species. Species infecting diverse crustaceans are also included in this clade. All the subclades correspond well to the established families within the order Glugeida (by Tokarev and Issi in Wijayawardene et al., 2020). Only families Spragueidae, Pleistoporidae and Glugeidae include fish-infecting species. The family Unikaryonidae includes Dictyocoeola, the most common microsporidian genus infecting amphipods globally (Bacela-Spychalska et al., 2018; Drozdova et al., 2020; Park et al., 2020). Families Theohaniidae, Pereziidae and Facilisphoridae are largely associated with crustaceans (amphipods, decapods, and copepods). Interestingly, two hyperparasite species (Hyperspora aquatica and Unikaryon legeri) are included in Clade 5. Hyperspora aquatica and U. legeri infect the paramyxid Marteilia cochillia and the digenene trematode Meiglymphallus minutus, respectively, both of which themselves are parasites of cockles (Stentiford et al., 2017). A possible vectoring role of hyperparasitism in the transmission of microsporidian parasites between crustacean and mollusk hosts has been suggested before (Stentiford et al., 2017).

3.1.2. Clade 4: a large clade containing diverse hosts from all realms

Two robust subclades were identified within Clade 4 (Clades 4A and 4B; Fig. 1B). Clade 4A consists mostly of species from terrestrial insects (lepidopterans and coleopterans) but also includes some species infecting freshwater crustaceans. The genus Nosema, which includes many species that infect economically important...
insect species. *Nosema bombycis* in silkworms and *Nosema ceranae*, *Nosema bombi* and *Nosema apis* in bees), belongs to this clade. Clade 4A does not include species from crustacean or dipteran hosts. On the other hand, clade 4B includes species found in diverse crustacean hosts including amphipods, decapods, copepods, anostomats, and cladocerans. Also, Clade 4B includes some parasites of dipterans, suggesting that some lineages within this clade may be largely associated with freshwater environments. The phylogenetic relationships among lineages within Clade 4B were poorly resolved with some polytomies. Many operational taxonomic units (OTUs) belonging to this clade were recovered in a study of environmental sequences, suggesting that Clade 4B may be highly under-sampled (Williams et al., 2018). Notably, the *Enterocytozoon* Group of Microsporidia (EGM) belongs to this clade. This robust monophyletic EGM group includes parasites of many important marine species and human-infecting species; some food- and water-borne microsporidiosis outbreaks were related to this group and possible transmission through the human food-chain has been suggested (Stentiford et al., 2016, 2019).

### 3.1.3. Clades 1 and 3: Parasites with many freshwater and terrestrial hosts

Two well-supported clades corresponding to clades 1 and 3 were recovered in our analyses (Fig. 1C). Species belonging to these clades were mostly obtained from freshwater, freshwater-terrestrial, and terrestrial habitats. Clade 1 includes only species of arthropod hosts (insects and crustaceans), whereas Clade 3 includes species from a broader range of hosts (insects, crustaceans, bryozoans, nematodes, and chordates). *Nematocenator marisprofundi* from a nematode host and *Microsporidium* spp. from an amphipod host were found from marine and brackish environments, respectively.

#### 3.1.4. Unrecovered Clade 2, and other lineages branching near the base of the tree of canonical microsporidians

We also identified some minor lineages that diverged near the base of canonical microsporidian radiations, which do not belong to any of the major groups. In fact, these minor lineages were assigned to either Clade 2 or 3 in previous studies (Vossbrinck and Debrunner-Vossbrinck, 2005; Vossbrinck et al., 2014). To be specific, only four species (*Antonospora locustae* = *Paranosema locustae*, *Antonospora scoticae*, *Nematocida parisii*, *Ovavesicula papillae*) were included in Clade 2 in Vossbrinck et al. (2014). In our analyses, *Paranosema grylli* (96% identical to *A. locustae*) and *A. scoticae* belong to Clade 3, but *N. parisii* and *O. papillae* do not belong to any of the major clades (Fig. 1C). In fact, Clade 2 was poorly supported in the ML tree (Fig. 6.3a in Vossbrinck et al., 2014) in the previous study with a low bootstrap value (BS = 72). Several recent studies with genomic scale data, which included both *A. locustae* and *N. parisii*, also did not recover them as a monophyletic group.

---

**Fig. 1.** A Bayesian phylogenetic tree of canonical microsporidia inferred from concatenated ssrRNA and lsrRNA gene sequences. Nodes with strong support in both Bayesian (posterior probability >95) and maximum likelihood (ML: Bootstrap support >90) analyses were annotated with circles (●). Nodes that were strongly supported only in Bayesian analysis and ML analysis were annotated with triangles (▲) and squares (■), respectively. Major clades of Vossbrinck et al. (2014) and some family names are shown arrows. Names of host species and higher taxonomic classifications are shown next to the tree. Habitats of host-parasites are categorised into five groups and marked with different colours. (A) A part of the Bayesian tree of canonical microsporidia showing clade 5. (B) A part of the Bayesian tree of canonical microsporidia showing clade 4. (C) A part of the Bayesian tree of canonical microsporidia showing clades 1 and 3.
Overall, clades 1–3 were not strongly supported (i.e. low bootstrap values, or bootstrap values were not provided for most of the nodes) and therefore the borders between major clades were not clearly defined in the previous studies (Vossbrinck et al., 2014). For example, Amblyspora bracteata and Caudospora simuli belonged to Clade 1 in their ML tree, but these two species were not grouped with the rest of Clade 1 in the MP tree (Vossbrinck et al., 2014), which shows unstable phylogenetic positions of some lineages. These two species do not belong to any of the major clades in our analyses (Fig. 1C). The reason for the formation of Clade 2 in the previous studies (in contrast to ours where it was not recovered) was probably due to a LBA artifact (Lartillot et al., 2007). Also, our improved taxon sampling (the inclusion of more sequences similar to those of species in Clade 2) may have helped to break up unnatural groupings.

3.1.5. The need for revision of classification within canonical Microsporidia

The four major clades shown in our study have important implications since the current classification is largely based on the ssrRNA phylogeny. Our Bayesian and ML trees support all the suggested orders except for Ovavesiculida from the recently revised classification (Wijayawardene et al., 2020). Several well-supported family-level clades were identified within each main clade (see Fig. 1A–C). Especially, major family-level clades within Clade 5 and Clade 4A correspond well to revised families (Fig. 1A and B, Wijayawardene et al., 2020). There are also well-defined family-level groups within Clades 1, 3 and 4B (Fig. 1B and C), but these are largely discordant with the recently revised classification (Wijayawardene et al., 2020), which was mainly based on the previous phylogenetic studies. Therefore, reshuffling of some families and genera within and among the newly established orders Amblyosporida (Clade 1), Neopereziida (Clade 3), and Nosematida (Clade 4) is needed, and Ovavesiculida (Clade 2) should be dissolved.

Another obvious problem with the current taxonomy is the presence of para- and polyphyletic groups such as Nosema, Vairimorpha, Plesiophora and Amblyospora (Fig. 1). This issue has been continuously discussed and has been mostly attributed to classification based on morphological characters (e.g. number of nuclei, the process of spore formation, the spore shape, and the number of polar filament coils) which could change rapidly and are therefore unreliable for classification (Baker et al., 1994; Stentiford et al., 2013b; Vossbrinck et al., 2014). The presence of numerous para- and polyphyletic lineages illustrates that morphological similarities do not necessarily mean evolutionary relatedness but could be due to the convergent evolution of those morphological
characters (Vossbrinck et al., 2014). With the increasing use of molecular data for classification, many species within these genera were transferred to another genus or new genera were established (Slamovits et al., 2004; Franzen et al., 2006; Vavra et al., 2006; Tokarev et al., 2020). Similarly, even if two microsporidians are morphologically distinct, this does not necessarily mean that they are phylogenetically distantly related. In our trees, some genera with distinct morphological and developmental features are grouped, closely related to each other, suggesting that they could be treated as congeneric species (Fig. 1). These include Agglomerata-Binucleata-Senoma (Sokolova et al., 2016), Larssonia-Berwaldia-Conglomerata (Vávra et al., 2018), and Spraguea-Microgemma-Tetramicra (Casal et al., 2012; Scholz et al., 2017); see referred studies for detailed discussion. Even more strikingly, the extreme polymorphism of a single species shown in Ameson pulvis, which has both Ameson-like and Nadelspora-like lineages infecting the musculature of marine crabs, emphasises the need for classification primarily based on the molecular data (Stentiford et al., 2013a, 2013b; Vossbrinck et al., 2014).

Due to difficulties in describing species, reporting sequences of newly discovered lineages under the names of Microsporidia sp. or Microsporidium sp. is increasingly common. In fact, sequences under these provisional names represent a large portion of diversity within canonical Microsporidia as shown in our tree (Fig. 1). ssrRNA sequences can be useful to assign new species into order and family. The use of ssrRNA together with the additional marker(s), especially fast-evolving ones, could be used for the genus-species level in the future, but which markers can be used needs to be investigated further and general agreement among researchers would be also needed.

3.1.6. Codiversification and host switching

Many canonical microsporidian species are thought to be host-specific and are associated with a single host species or related groups (e.g. con-generic or con-familiar species). Hosts and parasites that are intimately associated may show congruent phylogenies, however, cophylogenetic patterns are rarely seen due to frequent host switching events (Vienne et al., 2013). Within Microsporidia, some studies that focused on a specific host–parasite system have revealed patterns of codiversification (i.e. congruent host–parasite phylogenies at a macroevolutionary scale). Such systems include Amblyospora and other genera in mosquitos, Nosema in bees, and Dictyocaulus in amphipods (Baker et al., 1998; Shafer et al., 2009; Andreadis et al., 2012; Park et al., 2020). These studies also inferred that host switching is more likely among closely related hosts, but also suggested that transmission among

![Fig. 1 (continued)](image-url)
distantly related hosts may also occur. On the other hand, some genera such as Nosema, Encephalitozoon and Enterocytozoon include species that infect distantly related hosts, suggesting common and frequent host switching. In fact, the most extreme example of transmission among distantly related hosts has been shown in human-infecting species (Stentiford et al., 2019). Interestingly, human-infecting taxa emerged within all the major clades. Among them, the EGM group, which contains the most common human-infecting microsporidian species Enterocytozoon bieneusi, was not known until the 1980s but now includes pathogens of diverse companion animals and livestock (Stentiford et al., 2019). Considering their fast evolutionary rate and frequent host switching across distantly related hosts and different habitats, more extensive exploitation of wild animals, habitat destruction, human encroachment into wild habitats, intensive animal farming, and environmental stress may promote these kinds of novel host–parasite associations.

3.2. Phylogenetic relationships among Microsporidia and their relatives

An unrooted Bayesian tree of opisthosporidians has been constructed to highlight the genetic distance between canonical microsporidians and SB-Microsporidia and other relatives (Fig. 2). Both Bayesian and ML trees were similar in overall topology (Fig. 2 and Supplementary Figs. S3–S5). Metchnikovellids and Chytridiopsis also have long branches and diverged before the last common ancestor of canonical Microsporidia. The clade containing canonical Microsporidia, Metchnikovellids and Chytridiopsis is strongly supported in both Bayesian and ML trees. Bass et al. (2018) used the term ‘LB-Microsporidia’ for the first time in contrast to SB-Microsporidia. In their analyses, Metchnikovellids were also shown as a sister group to canonical microsporidians, but whether they should be regarded as LB-Microsporidia was not explicitly discussed. Here, we also include Metchnikovellids and Chytridiopsis within LB-Microsporidia, thus slightly extending the inclusion border because they also have considerably long branches in ssrRNA trees. Although the monophyly of LB-Microsporidia is robust, their sister group is incongruent between our Bayesian and ML trees (Fig. 2 and Supplementary Figs. S3–S5). In our Bayesian tree, the LB-Microsporidia clade is sister to a clade of Mitosporidium and Morellospora, but their sister relationship is poorly supported. Meanwhile, Nucleophaga was grouped with Paramicrosporidium (Fig. 2). Although this is consistent with the early morphological observations that ‘Nucleophaga is similar to Paramicrosporidium in its infective stage, by having non-flagellated walled spores penetrating amoebae through host cell phagocytosis’ (Corsaro et al., 2014a, 2016; Michel et al., 2000), this was also poorly supported. On the other hand, Nucleophaga is sister to LB-Microsporidia in the ML tree (Supplementary Figs. S4 and S5), the same as shown in Bass et al. (2018). In both of our Bayesian and ML analyses, the clade containing LB-Microsporidia, SB-Microsporidia, and Rozellida is sister to Aphelida, and these opisthosporidian clades are sister to the monophyletic Fungi.

In fact, the phylogenetic placement of SB-Microsporidia is far from stable, probably due to heterogeneity in rates of DNA substitutions among taxa and undersampling. In cases such as this, the resulting topology can be greatly affected by taxon sampling, data refinement, and the choice of the model of molecular evolution (Philippe et al., 2011; Lartillot, 2020; Ranwez and Chantret, 2020). We acknowledge our imperfect knowledge of these groups. Genome-scale data from additional representative groups within Opisthosporidia and the use of appropriate models of molecular evolution may allow greater resolution in future studies.

Fig. 2. A Bayesian tree showing the inferred phylogenetic relationships among microsporidians and their relatives. ssrRNA gene sequences of opisthosporidians (canonical Microsporidia, Metchnikovellids, Chytridiopsis, short-branch Microsporidia, Paramicrosporidium, Mitosporidium, Nucleophaga, Rozellida, Aphelida) were included; these names and other group names of environmental sequences are marked. (A) The unrooted tree is shown. Some major traits of each group are also shown in grey boxes. Branches of ‘Expanded microsporidia’ from Bass et al. (2018) are highlighted with blue (grey) colour. (B) The arrow points to an enlarged part of the tree showing ‘short-branch microsporidia’ and other groups.
3.3. What differentiates canonical microsporidians from their relatives?

Although the higher classification of microsporidia has been changing dramatically for reasons similar to those mentioned above, canonical microsporidians show distinct characteristics both morphologically and genetically, making them distinct from other groups. Many comparisons of important traits among microsporidia and their relatives have been conducted recently; these traits include energy metabolism (Timofeev et al., 2020), mechanisms of host invasion, proliferation, and exit (Tamim El Jarkass and Reinke, 2020), the structure of rRNA (Corsaro et al., 2019), and genome architecture (Wadi and Reinke, 2020). These studies highlight that traits important for these intracellular parasites have become specialised or reduced in each lineage differently. The innovative traits that canonical microsporidia acquired are complex polar tubes and ADP/ATP transporters (Tsousis et al., 2008; Vávra and Larsson, 2014; Alexander et al., 2016). Microsporidia have lost many genes involved with DNA repair pathways, which partly explains the accelerated evolutionary rates in canonical microsporidia (Galindo et al., 2018). Traits present only in canonical microsporidia may have contributed to their successful colonisation of diverse metazoans, by promoting efficient host invasion, proliferation, and adaptation to diverse ecological niches (= a wide range of hosts in various habitats).

Although other opisthosporidians shared several traits with canonical microsporidia, none of them have all the traits described above. To be specific, the presence of a well-developed complex polar tube (which commonly consists of three sections including a straight part), which allows efficient penetration of host cells upon infection, is a defining character of canonical microsporidia, although the length and thickness vary among different species (Vávra and Larsson, 2014). Metchnikovellids, sister to canonical Microsporidia, have a rudimentary polar tube (short and thick, without the straight part) and for this reason, they used to be called ‘primitive’ Microsporidia (Larsson, 2014). Paramicrosporidium and Nucleophaga invade the host cell through host phagocytosis even though Paramicrosporidium has a polar filament (Scheid, 2007; Corsaro et al., 2014a). Also, all canonical Microsporidia lack mitochondria, which generate ATP (Tsousis et al., 2008). Microsporidia proliferate (produce spores to complete their life cycle) within the host cell and this is an energy-consuming process (Tamim El Jarkass and Reinke, 2020). Canonical Microsporidia obtain ATP using ADP/ATP transporters, which are believed to have been obtained horizontally from bacteria (Tsousis et al., 2008). The presence of ADP/ATP transporters was also identified in Rozella (Heinz et al., 2014). Lastly, metchnikovellids did not go through genome reduction as extensively as canonical microsporidia did. It is believed that reduction of regulatory genes and non-coding regions resulted in rapid evolutionary rates (Galindo et al., 2018). This may have produced diverse traits that promoted adaptation to diverse hosts and niches.

3.4. Canonical microsporidians in metazoan hosts: due to evolutionary adaptation or biased screening effort?

Canonical microsporidians have successfully colonised metazoan hosts. In fact, based on our current knowledge, canonical microsporidians seem to be exclusive to metazoan hosts. On the other hand, no opisthosporian has been detected from metazoan hosts except for canonical microsporidians, Chytridiopsis and Mitosporidium. Berbee et al. (2017) suggested that the divergence among Aphelida, Rozellida and Microsporidia preceded the major diversification of multicellular organisms based on the fact that Aphelida and Rozellida lack the ability of invasive growth to multicellular tissues. Although canonical Microsporidia form a mono-phyletic group with SB-microsporidia, and they share certain morphological similarities, the genomes of characterised SB-microsporidia are much more similar to those of Rozella and canonical Fungi (Haag et al., 2014; Quandt et al., 2017). Also, considering the fact that most SB-microsporidia parasitize amoebae, the divergence between LB-Microsporidia and SB-Microsporidia may have occurred a long time ago. Multiple lineages of SB-Microsporidia have recently been discovered with environmental sequencing (Bass et al., 2018; Lacerda et al., 2020). According to a recent network analysis investigating potential hosts of SB-Microsporidia, it has been suggested that SB-microsporidia may be associated with the Apicomplexa, Cercozoa, Fungi, as well as some Metazoa (Doliwa et al., 2021). Co-occurrence does not necessarily mean an actual host–parasite relationship, however we are starting to learn more about these long unknown groups. SB-microsporidia are far less known than canonical microsporidia to date, but their actual diversity may be very high. LB- and SB-microsporidia diverged from a common ancestor (probably a long time ago) and may have adapted to different host groups with different specialisation and reduction of traits. However, without more occurrence data and screening efforts from a wide range of hosts and habitats, we cannot exclude the possibility that microsporidia-like organisms were simply not characterised or detected within metazoan hosts. Also, canonical microsporidia may be common in microscopic hosts.

Undoubtedly, further genetic characterization of microsporidia and their relatives from underexplored host groups, environments, or new geographic areas that have never been explored will provide valuable insights into the evolutionary history of these extremely diverse groups. These could result in adding more major groups to the tree that we are presenting here. Although our trees show four major clades, this does not mean that there are only four major clades within canonical Microsporidia. Also, targeted studies on microscopic hosts may recover the hidden diversity of canonical microsporidia as well as SB-microsporidia. Notably, primers targeting canonical microsporidia do not amplify SB-microsporidia, and vice versa (Williams et al., 2018). Considering these factors, the use of different primer sets will reveal more diversity.

4. Conclusion

In this study, we inferred the phylogeny of canonical Microsporidia with an improved dataset and method. The discrepancy between our trees and those from previous studies highlight our imperfect knowledge of the diversity of these parasites. Microsporidia provide an excellent system to study host–parasite associations from the cellular to the ecosystem level. Although it is only a short fragment of the genome, the ssrRNA region is still useful for species identification and classification at the family–order level, and for detecting microsporidia from unknown hosts and environments. In addition to ssrRNA, genomic data could be used to resolve relationships among families and orders, and fast-evolving genes for genus–species level classification. Also, we emphasise that canonical microsporidia are distinct from ‘Microsporidia-like organisms’ or the rest of the ‘expanded Microsporidia’ despite some morphological similarities and genetic affinities among them. Only more data (genetic, morphological and ecological) would fill our knowledge gap and provide insights into the evolutionary relationships among these extremely diverse and successful intracellular parasites.

Acknowledgements

We are grateful to two anonymous reviewers for valuable comments that improved the manuscript. This research was supported...


