Genomic and phenotypic evolution of nematode-infecting microsporidia

Lina Wadi[1], Hala Tamim El Jarkass[1*], Tuan D. Tran[2*], Nizar Islah[1], Robert J. Luallen[2], and Aaron W. Reinke[1#]

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

[2] Department of Biology, San Diego State University, San Diego, CA 92182, USA

#Corresponding author

*These authors contributed equally

aaron.reinke@utoronto.ca

Abstract

Microsporidia are a large phylum of intracellular parasites that can infect most types of animals. Species in the *Nematocida* genus can infect nematodes including *Caenorhabditis elegans*, which has become an important model to study mechanisms of microsporidia infection. To understand the genomic properties and evolution of nematode-infecting microsporidia, we sequenced the genomes of nine species of microsporidia, including two genera, *Enteropsectra* and *Pancytospora,* without any previously sequenced genomes. Phylogenetic analysis shows that *Nematocida* is composed of two groups of species. Core cellular processes, including metabolic pathways, are mostly conserved across genera of nematode-infecting microsporidia. Each species encodes unique proteins belonging to large gene families that are likely used to interact with host cells. Most strikingly, we observed one such family, NemLGF1, is present in both *Nematocida* and *Pancytospora* species, suggesting horizontal gene transfer between species from different genera. To understand how *Nematocida* phenotypic traits evolved, we measured the host range, tissue specificity, spore size, and polar tube length of several species in the genus. Our results demonstrate that the ability to infect multiple tissues was likely recently lost in *N. homosporus* and that species with longer polar tubes are able to infect multiple tissues. Together, our work details both genomic and trait evolution between related microsporidia species and provides a useful resource for further understanding microsporidia evolution and infection mechanisms.

Introduction

Microsporidia are a large phylum of obligate intracellular parasites[1]. They were the first eukaryotic parasite to have their genome sequenced, which revealed the smallest known eukaryotic genome[2]. In the last 20 years, genomes from ~35 microsporidian species have been sequenced,

with the smallest genomes belonging to the *Encephalitozoon* which encode as few as ~1800 proteins[3]. Analysis of their genomes has shown that microsporidia are either the earliest diverging group of fungi or a sister group to fungi[4–6]. The evolution of microsporidia coincided with the loss of flagellum and the gain of a novel infection apparatus known as the polar tube[3,7]. Early-diverging microsporidian species retained a mitochondrial genome, but underwent moderate genomic loss[8,9]. Consistent with their evolution as obligate intracellular parasites, however, canonical microsporidia have undergone extensive genomic reduction, which has resulted in the loss of the mitochondrial genome and many other metabolic and regulatory proteins[10]. Additionally, microsporidian ribosomes have lost both proteins and ribosomal RNA expansion segments, resulting in the smallest known eukaryotic ribosomes[11–13]. Microsporidian proteins are also reduced, being on average ~15% shorter than their fungal orthlogs[2,14]. While these larger phylogenetic changes are well studied, it is less clear how individual microsporidia genera have evolved. One interesting example is the *Enterocytozoon,* which have undergone a lineage-specific loss of many genes involved in glycolysis[15].

Microsporidia have evolved to infect a wide range of different hosts[16]. Over 1400 microsporidia species have been characterized and about half of all animal phyla as well as some protists are reported to be infected[17]. Most individual microsporidian species only infect one or two closely related hosts, but ~2% have been reported to infect more than five species. The genomic differences that account for some microsporidia having a broader host range, consistent with being a generalist, is mostly unknown, however a comparison between a pair of mosquito-infecting species showed that the generalist contains a much smaller genome than the specialist[14]. Once inside a host, most microsporidia only infect a single tissue, but ~11% can infect four or more tissues. The mechanisms that restrict microsporidia infection to specific tissues are largely unknown, but longer polar tubes correlate with infecting tissues other than the intestine, suggesting that the longer tube is needed to access these other tissues[17,18]. Interestingly, microsporidia that infect multiple host species are more likely to infect more tissues, suggesting that the selective pressures for these traits may be related[17]. Microsporidia display great diversity in their morphological and infection properties, even within related clades, suggesting these parasites may be particularly plastic in their ability to evolve phenotypic traits[17].

Several species of microsporidia have been found to infect free-living nematodes[18–21]. The first of these to be characterized was *Nematocida parisii,* which naturally infects the model organism *Caenorhabditis elegans*[20]. This species has been used as model to understand microsporidian

infection of a host, host immune responses to infection, and identify inhibitors of microsporidian infection[21–28]. Additional species that infect the intestine, *N. ausubeli* and *N. ironsii,* have also been identified[20,21]. More recently, another species, *N. displodere,* was found to infect other tissues of *C. elegans,* including the neurons, muscle, and epidermis[18]. Several other species infecting both *C. elegans* and other free-living nematodes have also been identified[19]. Besides *Nematocida*, two other genera, *Enteropsectra* and *Pancytospora,* have been reported to infect nematodes and shown to be more closely related to species of microsporidia that infect humans[19].

Four species of nematode-infecting microsporidia from the *Nematocida* genus have been sequenced, revealing moderate-sized genomes encoding between ~2300-2800 proteins. Many of the proteins in these genomes were found to have the potential to directly interact with their hosts, including many proteins that belong to large gene families[5,18,21]. One such large gene family, NemLGF1, contains a signal for secretion and between ~160-240 proteins have been identified in each of the *N. parisii*, *N. ironsii*, and *N. ausubeli* genomes, including many lineage specific expansions[21]. Another family, NemLGF2, is present in only *N. displodere* and contains over 230 proteins, about half of which contain a RING domain, which is thought to involved in protein-protein interactions[19].

Here we present a comprehensive and comparative analyses of multiple genomes of nematode-infecting microsporidia. To elucidate the genomic properties of nematode-infecting microsporidia and to understand evolution within *Nematocida*, we sequenced nine additional microsporidian species, including three newly identified species. By analyzing 13 nematode-infecting microsporidian species, we show that *Nematocida* forms two groups and that *Enteropsectra* and *Pancytospora* form a sister group to the marine invertebrate- and human-infecting *Enterocytozoon*. We find most conserved cellular processes are retained in *Nematocida* and that the metabolic capacity is largely similar between the genera of nematode-infecting microsporidia. We identified many novel large gene families including, surprisingly, members of NemLGF1 in *Pancytospora* species. Finally, we characterized the phenotypic properties of *Nematocida* species and describe the gain and loss of these traits throughout evolution. Our study reveals evolutionary mechanisms of nematode-infecting microsporidia and provides a valuable resource for the use of nematodes to study mechanisms of microsporidian infection.

Results

Sequencing, assembling, and phylogenetic analysis of microsporidia genomes

Previous efforts in discovering microsporidia in free-living terrestrial nematodes has revealed a diversity of species, but so far only four species from a single genus have been sequenced[5,18,19,21]. To further understand the genomic diversity of nematode-infecting microsporidia, we sequenced nine additional species, including two genera that had not been previously sequenced. Six of these microsporidian species have been previously reported[19]. We recently discovered three of them through sampling of wild nematode populations; *Nematocida ferruginous* from France, *Nematocida cider* from the United States, and *Nematocida botruosus* from Canada. To sequence these genomes, we removed contaminating bacteria from microsporidia-infected nematodes using antibiotics, cultured infections in the host animals, isolated spores, and extracted DNA. We sequenced a single isolate for each species, except for *N. ferruginous* for which we sequenced three strains that have high similarity to one another (Fig. S1). Information on each microsporidia species, host, and isolation location is presented in Table S1. Microsporidian genomes were assembled and contaminating DNA were removed (See Methods and Table S2). Proteins from each genome were predicted and annotated (Table S3). These new microsporidian genome assemblies were assessed based on contig continuity and protein conservation and were demonstrated to be of high quality relative to previously assembled microsporidian genomes (Fig. 1D, E, and Table S4).

To determine the relative relationships between microsporidia species we performed phylogenetic analysis. Using OrthoFinder[29], we generated a phylogenetic tree of our nine newly assembled microsporidia species, 35 previously sequenced microsporidia genomes, and the outgroup *Rozella allomycis* (Fig. 1A and Table S5). This tree shows that *Nematocida* genomes form two groups. The first identified *Nematocida* species, *N. parisii*, forms a group with *N. ausubeli*, *N. ironsii*, *N. major*, and *N. minor*; we refer to these species as the "*Parisii"* group[19–21]. The second group contains the more recently identified *N. displodere* along with *N. homosporus*, *N. cider*, *N. ferruginous,* and *N. botruosus;* we refer to these species as the "*Displodere"* group[18,19]. The relative positions of the *Nematocida* species are highly supported by several phylogenetic trees (Fig. 1A and Fig. S2). As described previously from 18s rDNA and β–tubulin sequences, the species *Enteropsectra breve*, *Pancytospora philotis*, and *Pancytospora epiphaga* belong to the large cluster of microsporidia species referred to as the Enterocytozoonida clade[17,19,30,31]. Our whole genome based tree replicates this previously described phylogeny, demonstrating that these three non-*Nematocida* species are most closely related to the human-infecting *Vittaforma corneae*, and together they form a sister group to the *Enterocytozoon*[15,32].

137

138 **Conservation and divergence of proteins in microsporidia genomes**

139 Microsporidia have the smallest known eukaryotic genomes, but genome size and protein content
140 can vary greatly between species[3]. In *Nematocida,* we see consistently small genome sizes that
141 range from ~3.1 to 4.7 megabases. The two *Pancytospora* species genomes are similar to each
142 other, with sizes of 4.2 and 4.6 megabases. In contrast, the ~6.0 megabase genome of *E. breve*
143 is the largest of our nematode-infecting microsporidia species genomes (Fig. 1C). We see similar
144 trends in predicted number of proteins, with *N. botruosus* having the fewest proteins (2159) and
145 *E. breve* having the most predicted proteins (3531) (Fig. 1B). Although microsporidia species can
146 have large amounts of non-coding DNA that make their genomes much larger relative to their
147 protein content[14], we observe a strong correlation between protein content and genome size
148 among nematode-infecting microsporidia (Fig. S3).

149

150 Microsporidia are known to have lost many conserved eukaryotic proteins, including those
151 involved in metabolism[10]. To determine the conservation of metabolic function in nematode-
152 infecting microsporidia species, we analyzed proteins encoded by microsporidia genomes using
153 GO term categories. We first determined the number of proteins from each microsporidia species
154 that belong to a set of GO terms that cover most conserved eukaryotic cellular processes (Fig.
155 2). This analysis reveled that many cellular processes were either lost or greatly reduced in the
156 canonical microsporidia, which has been previously observed[8]. We then determined the proteins
157 from each microsporidian species that belonged to eight metabolic categories (Fig. S4-11,
158 summarized in Fig. 3). As has been previously observed, we see that the *Enterocytozoon* contains
159 many losses in metabolic pathways, and in particular those involved in energy production[15]. As
160 the genera *Pancytospora* and *Enteropsectra,* along with *V. cornea,* form a sister group to the
161 *Enterocytozoon,* we sought to determine how the metabolic capacity of these groups of species
162 compared. We observe that *Pancytospora* and *Enteropsectra* have a similar metabolic capacity
163 compared to most other microsporidia species, including the *Nematocida* (Fig. 3). However, there
164 are several notable differences between the genera, with *Pancytospora* and *Enteropsectra* having
165 losses in glutathione biosynthesis and the *Nematocida* having losses in phosphatidylinositol
166 biosynthesis.

167

168 The RNAi pathway has been lost multiple times from microsporidia genomes[14,33]. *Nematocida*
169 was previously observed to lack this pathway[3]. However, the genomes of the three species from
170 *Pancytospora* and *Enteropsectra* each encode RNAi pathway proteins (Table S6).

171

172 **Nematode-infecting microsporidia contain many large, expanded gene families**

173 A striking feature of microsporidia genomes are abundant large gene families that contain
174 paralogous family members with either signal peptide or transmembrane domains[10,21]. Using a
175 previously described bioinformatic approach[21], we analyzed the large gene family content of
176 nematode-infecting microsporidia, identifying 39 families in 13 species (Fig. 4 and Table S7). We
177 observed that every species had at least one large gene family, with six species having at least
178 five families with more than 10 members each. Some of these families have many members, with
179 5 families (NemLGF1, NemLGF2, NemLGF11, PanLGF1, and EbrLGF1) containing more than
180 100 members in at least one species. Previously, we had observed that almost all large gene
181 families are specific to individual clades of microsporidia species[21]. Most families (30/37) in
182 nematode-infecting microsporidia have at least 10 members in only one species. The families we
183 identified in *Pancytospora* and *Enteropsectra* are also unique to these species and not found in
184 *V. cornea*. Some of the large gene families, such NemLGF11, are present in small numbers in
185 other species, but are only greatly expanded in one of the species. To determine if any of the
186 large gene families identified in nematode-infecting microsporidia were present in other
187 microsporidia species, we searched other microsporidia genomes using family-specific models
188 (Table S8). The only family with more than 10 members present in another species is NemLGF26,
189 which contains the previously described InterB family containing the Pfam domain Duf1609, which
190 is present in NemLGF26 family members[34].

191

192 One of the most abundant *Nematocida* large gene families is NemLGF1[5,21]. This large family was
193 previously found in the genomes of *N. parisii*, *N. ausubeli*, and *N. ironsii*, with each species
194 containing between 160-240 proteins[21]. We also observe 105 members of this this family in *N.*
195 *major*. We do not observe this family in *N. minor* or any of the genomes in the *Displodere* group,
196 suggesting that the family appeared in the *Parisii* group after the divergence of *N. minor*. Strikingly,
197 we find that the NemLGF1 family is also present in *P. philotis* (81 members) and *P. epiphaga* (45
198 members). We generated a phylogenetic tree using all the members from these six species, which
199 shows distinct expansions of NemLGF1 proteins in the *Nematocida* and *Pancytospora* species
200 (Fig. 5A). Using minimal ancestor deviation to root this tree, we find poor support for the root of
201 the tree, suggesting our data is consistent with either group being the origin[35] (Fig. S12). To
202 determine the similarity of NemLGF1 protein from *Nematocida* and *Pancytospora* species, we
203 used AlphaFold to model the protein structures (Fig. 5B and C)[36,37]. We find that NemLGF1
204 proteins consist of three main domains (N-terminal, middle, and C-terminal). The N-terminal and

205 middle domains are similar (RSDM of 1.6 Å for N-terminal domain and 1.2 Å for middle domain)
206 between *N. parisii* member NEPG_02057 and *P. epiphaga* member PAEPH01_0103 (Fig. 5D and
207 E). However, the C-terminal domains form a bundle of alpha helices that are less conserved
208 (RSDM of 18.3 Å) (Fig. 5F). We also compared an *N. ausubeli* NemLGF1 member NERG_01890
209 to *N. parisii* NEPG_02057, and observed similar trends, although the C-terminal domains were
210 more conserved (RSDM 4.5 Å) than with *P. epiphaga* PAEPH01_0103 (Fig. S13A-E).

211

212 To begin to understand the biochemical function of these nematode infecting large gene families,
213 we generated structural models for the four other largest families using Alphafold (Fig. S13F-I).
214 None of these families have apparent structural similarity to other solved proteins, with the
215 exception of NemLGF2, which shares similarity to Leucine Rich Repeat proteins (Fig. S13I). This
216 family was previously described in *N. displodere* and we also observe large expansions of this
217 family in *N. ferruginous* and *N. homosporous,* but not in *N. cider*. As previously shown, although
218 we detect a member of this family in the *Parisii* group species, we only observe expansions of
219 this family within some *Displodere* group species.

220

221 **Evolution of *Nematocida* phenotypic traits**

222 Microsporidia often only infect one or more closely related host species, but generalists that infect
223 many hosts have also been observed[16]. Previous work in *Nematocida* have described species
224 with more restricted host specificity, such as *N. parisii* and *N. ausubeli,* as well as those with a
225 broader host range, like *N. displodere* and *N. homosporus*[19,38]*.* To characterize the host range of
226 the novel species we isolated, we set up experimental infections against seven different nematode
227 hosts. As controls, we also challenged each of these hosts with *N. parisii* and *N. ausubeli*. After
228 4 days of infection, we fixed and stained populations of animals using direct yellow 96[22,39] and
229 quantified the proportion of the population that produced microsporidia spores (Fig. 6A). For *N.*
230 *parisii* and *N. ausubeli*, we observed strong infection of *C. elegans* and *C. briggsae*, as previously
231 reported[19]. *N. ausubeli* also infected a modest percentage of *C. tropicalis* and a small percentage
232 of *C. nigoni*, while *N. parisii* infected a small percentage of *C. tropicalis*. These two species could
233 not infect any of the other hosts. These results are consistent with what has been previously
234 reported[19]. For *N. botruosus*, we only observed infection of its native host (*Panagrellus* sp. 2*)*, and
235 none of the other six host species tested. In contrast, *N. cider* could infect all the hosts tested to
236 some extent (Fig. 6B), with *C. elegans* being the most infected at 85.6% and *Panagrellus* sp. 2
237 being the least infected at 7.4%. For *N. ferruginous*, we were unable to observe more than 50%
238 of animals infected for any nematode species, even when using a high concentration of spores.

239 However, even at this level of infection, we observed infection of *C. elegans, C. briggsae,* and *C.*
240 *tropicalis.* Additionally, we saw infection in *C. remanei* and *Oscheius tipulae*, which were species
241 that none of the other microsporidia besides *N. cider* could infect (Fig. S14). Our data suggests
242 that *N. botruosus* has a specialized host range, whereas *N. cider* and *N. ferruginous* appear to be
243 generalists.
244
245 Microsporidia mostly infect only one tissue, but some microsporidia, including *N. displodere*, have
246 been shown to infect multiple tissues[17]. During our infection experiments, we only ever observed
247 spore formation in the intestine for *N. botruosus,* but observed infection in other tissues for *N.*
248 *cider* and *N. ferruginous* (Fig. S15A-C). Additionally, during prolonged infection with *N. cider* or
249 *N. ferruginous*, the entire animal turns brown, suggesting that infection is occurring throughout
250 the animal (Fig. S15D-E). To further characterize which tissues *N. cider* and *N. ferruginous* infect,
251 we used *C. elegans* strains with GFP specifically expressed in either the intestine, epidermis, or
252 body wall muscle (Fig. 6C). We observed *N. ferruginous* could infect both the muscle and the
253 epidermis but we did not see productive infection in the intestine. In contrast, *N. cider* infection
254 was seen in all three tissues, the muscle, epidermis, and intestine, although at a lower frequency
255 in the intestine (Fig. 6D).
256
257 We recently showed that the *C. elegans* protein AAIM-1 is necessary for efficient invasion by *N.*
258 *parisii*[23]. This protein is secreted and likely functions in the intestinal lumen. Although *N.*
259 *displodere* infects other tissues besides the intestine, feeding is necessary for infection, implying
260 that spores need to first be in the intestinal lumen to infect tissues beyond the intestinal cells. To
261 determine if infection by an epidermal and muscle infecting species was also dependant upon
262 this protein, we infected wild-type and *aaim-1* mutant animals with *N. cider*. Similar to what was
263 observed with *N. parisii* infection, *aaim-1* mutant animals exposed to a high dose of *N. cider* have
264 a reduced prevalence of infection compared to wild-type animals (Fig. S16).
265
266 Microsporidia species have an extracellular spore form of a defined morphology which can be
267 used to identify species[17]. *Nematocida* spores have been shown to have a variety of sizes, with
268 the average size of species ranging from 1.3-2.38 µm long and 0.53-1.03 µm wide. Additionally,
269 some species in the *Parisii* group have been shown to have a second, larger, class of spores[19].
270 To determine spore sizes for the new species we identified, we examined infected animals with
271 stained spores (Fig. S15A-C). This revealed that these species only have a single class of spores,

with *N. botruosus* measuring 1.51 x 0.76 µm, *N. cider* measuring 2.4 x 0.67 µm, and *N. ferruginous* measuring 1.72 x 0.6 µm.

Microsporidia has a characteristic infection apparatus known as the polar tube[40]. The length of this tube was previously shown to correlate with tissues infected, where shorter polar tubes (~ 4 µm for *N. parisii*) are seen in species restricted to the intestine, and longer polar tubes (~ 12 µm for *N. displodere*) seen in species that can infect the muscles, neurons, and epidermis. This correlation was hypothesized to be due to the requirement of longer polar tubes to access more distal tissues from the lumen of the intestine[18]. We measured polar tube lengths of several *Nematocida* species using a previously described freeze-thaw approach to trigger spore firing[18] followed by staining with Nile red[41] (Fig. 6E). We observed a range of lengths throughout *Nematocida*, with the shortest being *N. botruosus* with a length of 2.3 µm and the longest being the previously measured *N. displodere* (Fig. 6F)[18].

To understand how various microsporidian phenotypic traits evolve, we arranged our measured attributes onto the phylogenetic tree of *Nematocida (*Fig. 7*)*. We observe that all of the *Parisii* group species have a specialized host range and show specificity for the intestine. In contrast, all of the species in the *Displodere* group, except for *N. botruosus*, have a more generalist host range, suggesting that this property was either lost in *N. botruosus*, or gained after species divergence. Additionally, the *Displodere* group species, *N. displodere*, *N. cider*, and *N. ferruginous*, infected other tissues besides the intestine, suggesting that the ability to infect multiple tissues was most likely present after the divergence of *N. botruosus*, and then lost in *N. homosporus*.

We observed that spore size is flexible throughout *Nematocida* species, with both groups of *Nematocida* containing species (*N. minor* and *N. botruosus*) exhibiting some of the smallest spores. Additionally, we observed that two classes of spores are present only in the *Parisii* group and large spore clusters are only present in the *Displodere* group. Finally, we observe flexibility of polar tube length ranging from 2.3 - 12.6 µm. Our data suggests that the last common ancestor of *Nematocida* likely had a relatively short polar tube (less than ~ 4 µm). An increase in polar tube length then occurred in the *Displodere* group after the divergence of *N. botruosus*. This increase in polar tube length correlates with tissue specificity, with the three species with the longest polar tubes all infecting tissues besides the intestine. These results provide further support for microsporidian tube length being a factor in determining which tissues are infected[17]. Together,

306 our results suggest that trait evolution in *Nematocida* is flexible, with gains and losses of attributes
307 common throughout speciation.
308
309 **Conclusions**
310 To understand genomic evolution of nematode-infecting microsporidia, we sequenced the
311 genomes of nine additional species. We then analyzed these genomes along with four previously
312 sequenced species. We show that core cellular processes, including metabolism, are largely
313 conserved between genera. Our analysis demonstrates that *Enteropsectra* and *Pancytospora*
314 form a sister group to the *Enterocytozoon*, but have not undergone decay of the glycolytic pathway
315 that has occurred in *Enterocytozoon* species[15]. This conservation of metabolism suggests that
316 host metabolic requirements are similar between species and is in contrast to proteins predicted
317 to interface with hosts, which are quite different between different groups of microsporidia[21].
318 Although we present the largest comparison of protein function between microsporidia species
319 so far, detecting protein function from sequence alone in microsporidia is challenging. Additional
320 analyses taking advantage of advances in structural prediction are likely to detect highly diverged
321 proteins that would be missed based on sequence similarity alone[10,42,43].
322
323 One class of proteins that contributes to diversity between microsporidian species are the large
324 gene families. Previous work identified that all sequenced species of microsporidia contained at
325 least one family, and most families were clade specific[21]. These families were also enriched for
326 secretion signals and transmembrane domains, suggesting that proteins in these families are
327 used to interface with the host[10,21]. Here we show that one of these large gene families, NemLGF1,
328 is present in both *Nematocida* and *Pancytospora* species. As members of this family are not
329 present in any other microsporidia species, the most likely explanation is that this family was
330 horizontally transferred between a species in each genus, though which genus was the recipient,
331 and which was the donor, is not clear. We also identified a family (NemLGF26) present in *N.*
332 *ferruginous* which shares similarity to the previously described InterB family found in
333 *Encephalitozoon* species[34]. Although to our knowledge, horizontal transfer of genes between
334 microsporidia proteins has not been previously reported, co-infections of multiple species are
335 often observed, suggesting that these types of events may be more common that previously
336 appreciated[44].
337
338 Microsporidia have been shown to display much flexibility in their phenotypic traits[17]. However,
339 determining how traits vary between related species has been challenging as most phylogenetic

trees are built from 18S rRNA sequences, resulting in phylogenies that are often of low accuracy[3]. Here, we use genome assemblies to build a high-quality phylogeny for the *Nematocida*. By mapping several phenotypic properties onto this tree, we observe examples of loss and gain of infection properties. As more microsporidian genomes are sequenced, this approach is likely to be useful for further understanding the evolution of microsporidia infection properties and spore characteristics. With our data we show that a generalist host range emerged in the *Displodere* group after the divergence of *N. botruosus*. One potential caveat to our measurements of host range is that they are only done with a single strain for each microsporidia and host species[16]. However, we see similar results to what was reported with other *N. parisii* and *N. ausubeli* strains, and nematode-infecting microsporidia have been observed to mostly infect different host strains at similar levels[19]. Previously, the size of generalist microsporidia genomes were suggested to be smaller than genomes of species that had a more narrow host range[14]. Although the generalist *Nematocida* have some of the smallest genome sizes, they are about the same size as *N. botruosus*, which has a narrow host range. This suggests that other forces can be involved in sculpting microsporidia genomes.

Although *C. elegans* has become an important model in which to study many aspects of microsporidia infection, there are currently no tools for genetically perturbating microsporidia that infect this host[45]. Knockdown of genes using RNAi has been demonstrated in microsporidia that infect honey bees, silkworms, fish, and locusts and is becoming a powerful method to investigate the function of microsporidia proteins, including those that are secreted or surface exposed[46–51]. So far, this approach has not been amenable to species that infect *C. elegans*, as *N. parisii* and the other *Nematocida* don't encode the machinery necessary to carry out RNAi[3]. Here we report that *P. epiphaga,* which infects *C. elegans*[19], encodes this machinery, suggesting RNAi may be possible in this species. *C. elegans* may be particularly useful for microsporidian RNAi knockdown, as RNAi of genes in *C. elegans* is highly efficient and can be carried out by feeding bacteria expressing double stranded RNA against the target transcript, resulting in dozens of successful whole genome screens[52]. Additionally, *P. epiphaga* is more closely related to several human pathogens than *N. parisii* and our genomic data will provide a useful resource for further investigations using this species.

Material and methods

373 **Isolation of infected animals**

374 Sampling of nematodes was done similar to as described previously[18]. Briefly, samples of rotting
375 fruit or vegetation were placed on 6-cm NGM plates seeded with 10x saturated cultures of OP50-
376 1 for 16-48 hours. Rotting matter was then removed. *N. ferruginous* was identified using
377 differential interference contrast microscopy to identify animals displaying meronts and/or spores.
378 *N. cider* and *N. botruosus* were identified by growing animals at 21-23°C, washing animals off
379 plates in M9/0.1% Tween-20, and fixing animals in 800 µl acetone. Samples were then stained
380 with Direct yellow 96 (DY96) staining solution (M9, 0.1% SDS, and 20 µg/ml DY96) for 30 min.
381 Samples were resuspended in EverBrite Mounting Medium (Biotium #23002) with DAPI and
382 placed onto microscope slides. Microsporidia-containing animals were identified as those
383 containing DY96 stained spores. Nematode host species were determined using sequencing
384 primers as described previously[53].

385

386 **Removal of contaminating bacteria from microsporidia-infected nematodes**

387 Infected nematodes isolated in this work and previously[19] were grown on 10-cm NGM plates
388 seeded with 10x OP50-1 until populations of worms had ingested all available bacteria. Animals
389 were then washed off plates with M9/0.1% Tween-20 and frozen at -80°C. 2.0 mm zirconia beads
390 were then added and samples vortexed at 3,000 for 5 min in a bead disrupter. The supernatant
391 was then placed onto 6-cm NGM plates containing antibiotics (50 µg/ml carbenicillin, 25 µg/ml
392 kanamycin, 12.5 µg/ml tetracycline, 37.5 µg/ml chloramphenicol, 200 ug/ml cefotaxime, and 100
393 µg/ml gentamycin) for 24-48 hours. Either the native nematode host strain that the microsporidia
394 species was isolated in or *C. elegans* N2 animals were used to cultivate each species (See Table
395 S1). Mixed populations of nematodes were bleached (~4% sodium hypochlorite and 1 M NaOH)
396 for 2-3 minutes to extract embryos which were hatched at 21°C for 18-24 hours. L1 stage worms
397 were then added to the antibiotic containing plates along with 10X OP50-1. After several days
398 animals were chunked onto new NGM plates containing antibiotics. After several days animals
399 were chunked onto NGM plates without antibiotics. If no contamination was detected, then
400 samples were used to prepare spores. If not, the above procedure was repeated until there was
401 no contamination.

402 **Preparation and DNA sequencing of microsporidian spores**

403 Spores were prepared similar to as previously described[22]. Briefly, contaminate-free, infected
404 populations of worms were chunked onto 12-48 10-cm OP50-1 seeded NGM plates and grown
405 at 21°C for 5-10 days. After worms had consumed all the OP50-1 bacteria, worms were washed

406 off plates with M9 and frozen at -80°C. Spores were extracted from worms as described above
407 and then filtered through a 5 µm filter (Millipore). The concentrations of spore extracts were
408 determined by staining spores with DY96 and counting spores on a sperm counting slide (Cell-
409 VU) using a 20X objective on a Axio Imager.M2 (Zeiss). DNA was extracted using a MasterPure
410 yeast DNA purification kit (Lucigen) according to manufacturer instructions using 10-50 million
411 spores per microsporidia species. DNA libraries were prepared by The Centre for Applied
412 Genomics (TCAG) and each species was sequenced using 2 lanes of a NovaSeq 6000 SP flow
413 cell to produce 250 bp paired-end reads.
414
415 **Genome assembly & annotation**
416 Adapter sequences ('AGATCGGAAGAG') were removed from paired-end raw reads using
417 cutadapt v21[54]. Leading and trailing low quality (quality < 3) or N bases were removed, low quality
418 bases (quality < 30, except for *P. epiphaga* where 36 was used) were cut using a 4-base sliding
419 window and reads less than 36 bases were dropped using Trimmomatic v0.36[55]. Reads were
420 mapped to the corresponding host genome (except for *N. botruosus* which was mapped to *C.*
421 *elegans)* as well as the *Escherichia coli* genome using bowtie2 v2.3.4.1[56]. Reads were assembled
422 into contigs and scaffolds using Abyss v2.02[57] with a Kmer of 128. Only scaffolds greater than
423 500 bp were retained. Assemblies were then filtered for contaminants (bacteria and host) using
424 BlobTools v1.0.1[58]. For some samples (Table S2), subsampling was done prior to removing
425 contaminants with BlobTools. Each assembly was run through Redundans v0.14a[59] with a
426 minimum identity value of 0.85, to remove duplication. Assemblies were also filtered for ERTm1[5]
427 contamination if they met the following criteria: 99% identity and 80% query coverage. Additional
428 filtering was done for size and coverage: scaffolds less than 750 bp were removed and those
429 between 750 bp – 1 kbp were removed if they had less than 5X median coverage of the largest 5
430 scaffolds. NCBI nucleotide blast (BLASTn)[60] was used to filter remaining scaffolds that had >98%
431 identity and >50% query cover to ERTm1/ERTm2 assemblies. Manual filtering of select scaffolds
432 not captured in previous filtering steps was done for six of the samples (See Table S2). Protein
433 coding genes were predicted using Prodigal v2.6.3[61] using translation table 1. Proteins with less
434 than 100 amino acids were kept only if they had a significant (E=0.001) hit to the Uniref90 protein
435 database or Pfam hit using the hmmscan function in HMMER package v3.1b2[62] and E-value cutoff
436 of 0.001. The final list of proteins were annotated with Pfam[63] domains using hmmscan, signal
437 peptides using SignalP v4.1f[64], and transmembrane domains using TMHMM v2.0c[65]. Functional
438 annotation of proteins was also done using BlastKOALA[66] (Table S3).
439

440 **Phylogeny and genome characterization of microsporidia species**

441 The phylogenetic tree (Fig. 1) was generated using 45 species. Species whose proteome was
442 downloaded from NCBI are listed in Table S5. The remaining proteomes were predicted from the
443 genomes assembled in this study. In addition, proteins were predicted from two assemblies
444 available on NCBI for the following species: *Metchnikovella incurvata*, and *Enteropsora canceri*.
445 Orthofinder v2.5.2[29] was used with -M msa option to obtain a MSA species tree with a minimum
446 of 77.8% of species having single copy genes in an orthogroup. Tree was rooted using *Rozella*
447 *Allomycis*[67] in FigTree v1.4.4. Protein content, genome size, as well as genome completeness
448 statistics were computed for each species using BUSCO v3.1.0[68] and microsporidia odb9.

449

450 **Pathway conservation in microsporidia species**

451 InterProScan v5.51-85.0[69] was run on all species with -goterms option. A database was made
452 with POMBE GO-slim terms (https://www.ebi.ac.uk/QuickGO/slimming) and their descendants
453 (found using QuickGO API using the 'find descendants' option,
454 https://www.ebi.ac.uk/QuickGO/api/index.htmllar#!/gene_ontology/). For each species, protein to
455 GO term annotations from InterProScan were mapped to the above-mentioned database while
456 retaining count of proteins mapping to each GO-term. Since multiple GO terms map to a single
457 GO slim, a unique protein count was obtained for each GO-slim. Using *Rozella allomycis* as the
458 base species, a subset of the POMBE GO-slim terms were obtained and the subset was then
459 searched against the remaining species to compare pathway conservation across species.

460

461 **Identifying large gene families**

462 OrthoMCL v2.0.9[70] was run using all proteomes and with the following configuration settings:
463 percentMatchCutoff=50 and evalueExponentCutoff=-5, suggested minimum protein length and
464 maximum percent stop codons, and MCL inflation value of 1.5. Paralogous orthogroups were
465 extracted (at least 10 members in a single genome). The genome with the highest number of
466 paralogs for a specific gene family was used as the seed sequences to expand gene families.
467 Seed sequences were searched against each microsporidia genome iteratively using the
468 jackhmmer function in HMMER and an E-value cutoff of 10-5. All resulting large gene families
469 were assessed for signal peptides, transmembrane domains, and Pfam domains. A family was
470 classified as having signal peptides if >50% of the proteins in the family had a signal peptide. The
471 same criteria were applied for transmembrane domains. Families not classified as having signal
472 peptides or transmembrane domains were removed from the analysis. All families with 90% or
473 greater protein overlap were collapsed, with the exception of NemLGF2 and NemLGF20 which

474 were collapsed with less than 90% overlap in some cases.. We used OrthoMCL to assign
475 membership with families with >5% Pfam representation in either LRR, ZF, or peptidase
476 domains[21].
477
478 **Structural modeling of large gene families**
479 Models of members of large gene families were generated using AlphaFold as implemented in
480 ColabFold[36,37](https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaF
481 old2.ipynb). Default settings were used for NEPG_02057, and for the other proteins PSI-BLAST[60]
482 was used to identify members and multiple sequence alignments were generated using HHblits
483 (https://toolkit.tuebingen.mpg.de/tools/hhblits)[71]. Alignments of NemLGF1 domains were
484 performed using PyMOL v2.5.3 (https://pymol.org). Structural similarity to solved structures was
485 determined using Dali (http://ekhidna2.biocenter.helsinki.fi/dali/)[72].
486
487 **Host specificity assays**
488 The following nematode species (strains) were used for infection assays: *C. elegans* (N2),
489 *Caenorhabditis briggsae* (JU2507), *Caenorhabditis nigoni* (JU1422), *Caenorhabditis remanei*
490 (JU2796), *Caenorhabditis tropicalis* (JU1373), *Oscheius tipulae* (JU1505), and *Panagrellus* sp. 2
491 (AWR79). Mixed populations of nematodes were maintained on 10-cm seeded NGM plates for at
492 least three generations without starvation at 21°C. Animals were then washed off plates and
493 embryos extracted with bleach as described above, except for *Panagrellus* sp. 2. Synchronized
494 animals of *Panagrellus* sp. 2 were prepared by washing animals with M9 three times and then
495 filtering through a 20 µm nylon filter (Millipore SCNY00020). This filtering was then repeated a
496 second time. 400 L1 stage animals of each species were infected with the following microsporidia
497 species (strain [spore amount]): *N. parisii* (JUm1248[3 million]), *N. ausubeli* (ERTm6[0.8 million]),
498 *N. cider* (AWRm77[4 million]), *N. ferruginous* (LUAm3[20 or 40 million]), and *N. botruosus*
499 (AWRm80[28 million]). Animals and spores were mixed with 400 µl 10X OP50-1 and placed onto
500 6-cm NGM plates. Plates were dried in a clean cabinet and incubated at 21°C for 96 hours.
501 Animals were then washed off plates, fixed, stained, and placed onto slides as described above.
502 The percentage of animals infected was determined by counting the number of P0 animals that
503 contained newly formed clusters of microsporidia spores.
504
505 **Infection of *aaim-1* mutants**
506 N2 and *aaim-1* mutants (kea22 and kea28)[23] were infected with *N. cider* (4 million spores) as
507 described above. To quantify pathogen burden within animals (GFP), regions of interest were

508 used to outline individual worms followed by subjection to the “threshold” followed by “measure”
509 tools in FIJI[73].
510
511 **Tissue specificity determination**
512 Strains of *C. elegans* each expressing GFP specifically in the intestine (ERT413), epidermis
513 (ERT446), or muscle (HC46) were grown on standard NGM plates seeded with OP50-1 to the
514 gravid adult stage. Gravid adults were then harvested in M9 to 15 mL conical tubes, washed three
515 times to reduce bacterial excess, and pelleted by centrifuging at 3000 x g. The pellets were
516 resuspended in 1 mL of M9 buffer and added with 1 mL of bleach solution (1:4 volume ratio of 5M
517 NaOH: 6% NaClO; final concentration of 0.5M NaOH, 2.4% NaClO), incubated for 2 min at room
518 temperature, and washed three times with 13 mL of M9. After the final wash, the solutions
519 containing embryos were incubated overnight on a rotator at room temperature to synchronize to
520 the L1 stage.

521 ~1000 L1 stage animals of each strain were plated onto a 6-cm NGM plate with either 1 million
522 *N. ferruginous* or *N. cider* spores. Infection plates were allowed to dry and kept at 20°C. At 3 days
523 post infection, animals were harvested in M9 buffer and fixed with 4% paraformaldehyde in PBST
524 (1x PBS, 0.1% Tween-20) at room temperature for 40 min on a rotator for fluorescent in situ
525 hybridization. Fixed animals were then washed four times, each with 1 mL of PBST, then
526 incubated overnight with a mixture of CAL Fluor Red 610 (CF610)-tagged microsporidian probes,
527 including MicroA-CF610 (CTCTGTCCATCCTCGGCAA), MicroB-CF610
528 (CTCTCGGCACTCCTTCCTG), MicroD-CF610 (CGAAGGTTTCCTCGGATGTC), and MicroF-
529 CF610 (AGACAAATCAGTCCACGAATT), each at a final concentration of 2.5 ng/mL in
530 hybridization buffer (900 mM NaCl, 20 mM Tris pH 7.5, 0.01% SDS) on a thermal shaker at 46°C,
531 1000 rpm. After hybridization, animals were washed with FISH wash buffer (hybridization buffer,
532 5mM EDTA) for three times, each for 30 min, on a thermal shaker at 48°C, 1000 rpm. After
533 washing, animals were rinsed once with PBST, mounted, and imaged with confocal microscope
534 Leica Stellaris 5. Each sample was scanned randomly for 30 infected worms and images were
535 taken in different planes to examine all the infection areas in a single worm. For each strain,
536 infected animals were binned into “yes” or “no” respective to the fluorescent tissue. An infected
537 animal was considered to have a specific tissue type infected if there was at least one infection
538 area colocalized with the tissue marker.

539 ***N. cider* and *N. ferruginous* assays to measure change in host color**

540 ~1000 N2 animals at L1 stage on a 10-cm NGM plate were infected with 3 million spores of
541 either *N. ferruginous* or *N. cider*. As the animals grew, plates were monitored to add more food
542 (1 mL of 10x OP50-1 concentrated from a culture of OD_{600} of 0.5-0.7) to prolong the infection in
543 the population. After 7-10 days post infection, some animals appeared brown compared to other
544 animals on the same plate and uninfected control under standard dissection microscope.
545 Control animals and brown, infected animals were picked and mounted onto a 2% agarose pad
546 and imaged with Eclipse Ni microscope (Nikon). All images were taken at the same light
547 intensity and white balanced.

548 **Microsporidia spore size measurements**

549 Infected, fixed, and stained animals were prepared as described above. Images of these infected
550 animals were taken using an 63X objective on an apotome-equipped Axio Imager.M2 (Zeiss).
551 Images were taken as z-stack maximum intensity projections. These images were analyzed using
552 the straight line tool in ImageJ 1.52[74] to measure the length and width of each spore. Only spores
553 that were not immediately adjacent to other spores were measured. At least 50 spores were
554 measured for each species from a minimum of two infected worms.
555

556 **Measurements of polar tube lengths**

557 Polar tube lengths were measured for *N. parisii* (ERTm1), *N. ausubeli* (ERTm2), N. minor
558 (JUm1510), N. major (JUm2507), *N. homosporus (*JUm1504*)*, *N. cider* (AWRm77), *N. ferruginous*
559 (LUAm3), and *N. botruosus* (AWRm80). To encourage polar tube extrusion, microsporidia spores
560 in microcentrifuge tubes were exposed to two freeze thaw cycles, after the initial thaw from -80°C.
561 This was done by freezing at -80°C for 10 minutes followed by thawing at room temperature for
562 10 minutes. Nile red (Sigma-Aldrich 72485) (1mg/ml in acetone) was prepared, sequentially
563 diluted to 100 µg/ml in M9, and added to a final concentration of 10 µg/ml in freeze-thaw treated
564 spores. Spores were incubated with Nile red for 30 minutes in the dark followed by 2 µl of
565 Calcofluor White (Sigma-Aldrich 18909). Polar tubes were imaged at 63x magnification using an
566 Axio Imager.M2 (Zeiss). Polar tube measurements were performed on FIJI[73], using the freehand
567 line tool to trace polar tubes followed by selecting the Analyze → Measure option.

568 **Data availability**

569 All samples were deposited under NCBI BioProject PRJNA579982, accession numbers are listed
570 in Table S1, and the sequence reads were submitted to the NCBI Sequence Read Archive
571 (SRX13790799).
572

573 **Taxonomic Summaries.**

574 Phylum: Microsporidia Balbiani 1882

575 Species: *Nematocida cider* n. sp. Wadi et al. 2022.

576

577 The type strain AWRm77 was found in 2017 inside of its type host nematode *Caenorhabditis* sp.
578 8 strain AWR77, which was isolated from rotting apples in Stow, Massachusetts, United States.
579 The genome of this microsporidia strain has been sequenced and deposited in Genbank under
580 accession JALPNA000000000. This species has been observed to infect the intestinal, muscle
581 and epidermal tissues of *C. elegans*. Experimental infection has been observed in *C. elegans*
582 (N2), *Caenorhabditis briggsae* (JU2507), *Caenorhabditis nigoni* (JU1422), *Caenorhabditis*
583 *remanei* (JU2796*)*, *Caenorhabditis tropicalis* (JU1373), *Oscheius tipulae* (JU1505), and
584 *Panagrellus* sp. 2 (AWR79). Transmission occurs horizontally. The spores are ovoid and measure
585 as 2.4 x 0.67 µm. The fired polar tube measures 8.3 µm. This species is named after the brown
586 color the animals display after being infected for several days and this color resembles that of
587 cider donuts available in the area where the specimen was found.

588

589 Species: *Nematocida ferruginous* n. sp. Wadi et al. 2022.

590 The type strain LUAm1 was found in 2017 inside of its type host nematode *C. elegans* strain
591 LUA1, which was isolated from a rotting *Heracleum* sp. stem in Santeuil, France (GPS
592 coordinates 49.12138616666667,1.951088333333333). The genome of this strain has been
593 sequenced and deposited in Genbank under accession JALPMW000000000. This species has
594 been observed to infect the muscle and epidermal tissues of *C. elegans*. Experimental infection
595 has been observed in *C. elegans* (N2), *Caenorhabditis briggsae* (JU2507), *Caenorhabditis*
596 *remanei* (JU2796*)*, *Caenorhabditis tropicalis* (JU1373), *Oscheius tipulae* (JU1505). Transmission
597 occurs horizontally and is likely via a fecal-oral route. The spores are ovoid and measure as 1.72
598 x 0.6 µm. The fired polar tube measures 6.2 µm. This species is named after the rust-like brown
599 color the animals display after being infected for several days.

600 Species: *Nematocida botruosus* n. sp. Wadi et al. 2022.

601

602 The type strain AWRm80 was found in 2018 inside of its type host *Panagrellus* sp. 2 strain
603 AWR80, which was isolated from rotting apples in Georgina, Ontario, Canada. The genome of
604 this strain has been sequenced and deposited in Genbank under accession JALPMX000000000.
605 This species has only been observed to infect the intestine of this animal and transmission occurs

horizontally. The spores are ovoid and measure as 1.51 x 0.76 µm. The fired polar tube measures 2.3 µm. This species was named after the prominent large spore clusters formed.

Acknowledgements

We thank Yin Chen Wan, Edward James, and Winnie Zhao for providing helpful comments on the manuscript. We thank Marie-Anne Félix for providing previously published microsporidia species and for providing support to R. L. during isolation of *N. ferruginous*. This work was supported by a Canadian Institutes of Health Research grant no. 400784 (to A.R.) and an Alfred P. Sloan Research Fellowship FG2019-12040 (to A.R.). T.T was supported by SDSU Rees-Stealy Fellowship and H.T.E.J was supported by an University of Toronto Open Fellowship. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

Author contributions: L.W. and A.R. conceived of the project. L.W. performed all genomic analysis. N. I. contributed code to the genomic analysis. T.T., H.T.E.J., R. L, and A. R. performed experiments. R.L. and A. R. provided mentorship and acquisition of funding. L.W and A.R. co-wrote the paper with edits from T.T., H.T.E.J., N.I., and R. L.

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

Figures

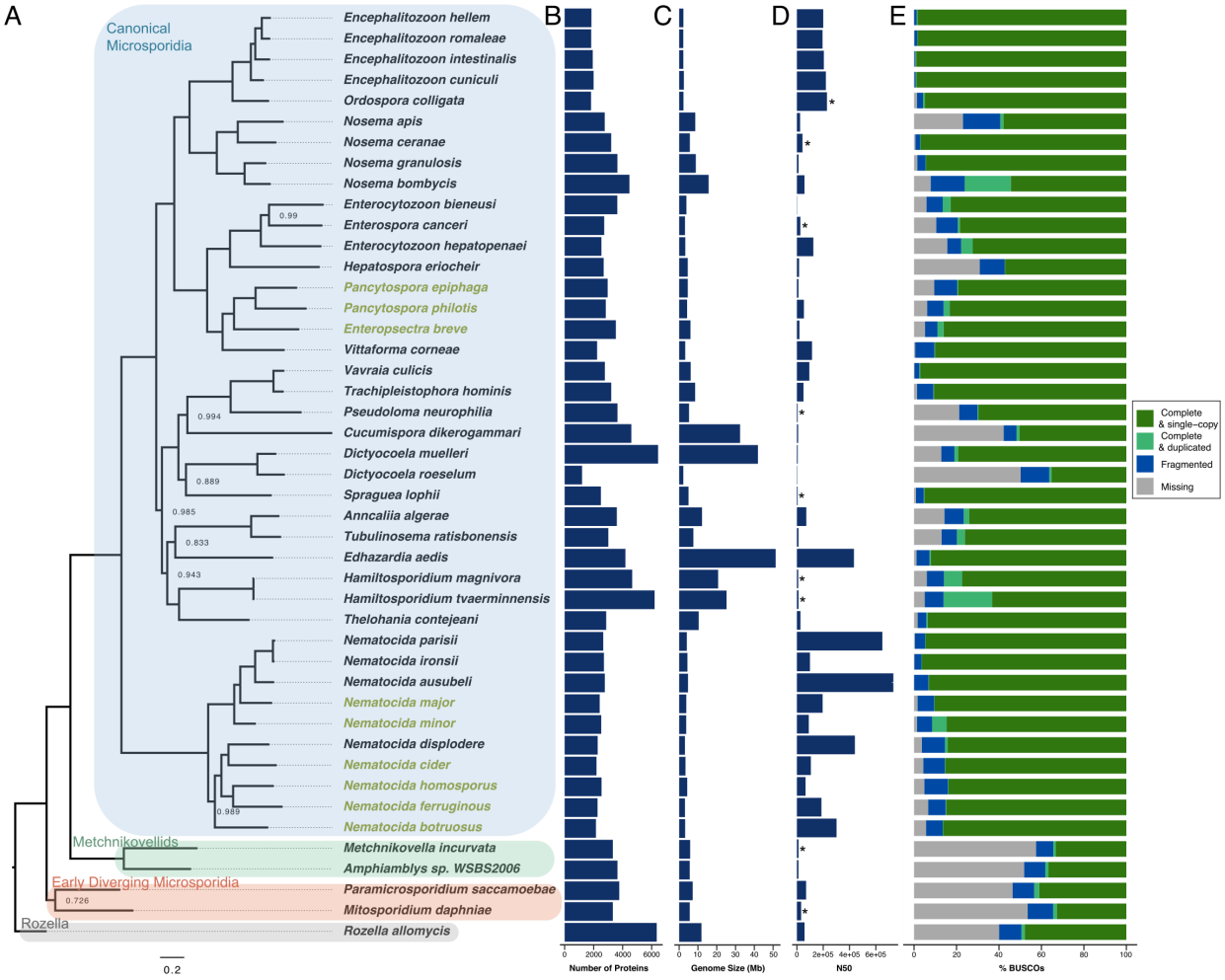
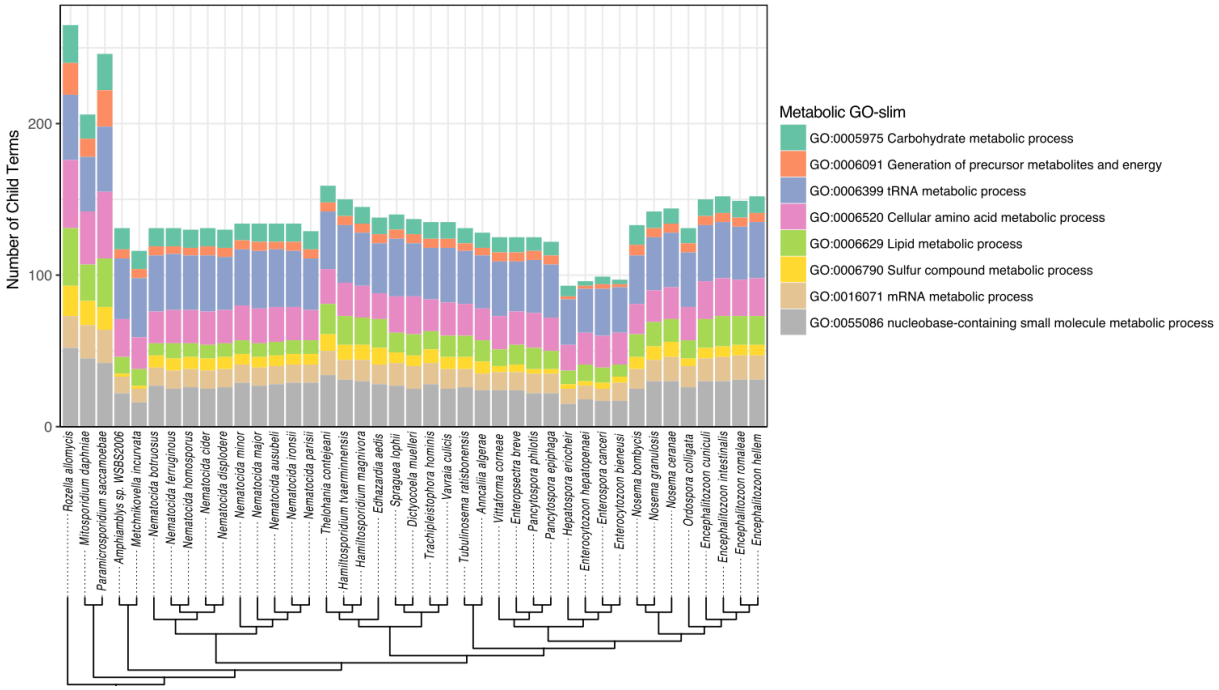


Fig. 1. Phylogeny and properties of microsporidia genomes. (A) Phylogenetic trees of 44 microsporidia species and the outgroup *Rozella allomycis* was constructed using Orthofinder[29]. Bootstrap values less than 1.0 are indicated at each node. Scale indicates changes per site. Previously sequenced species are shown in black and newly sequenced genomes presented in this paper are shown in yellow. Species are grouped into four classes (Rozella, early diverging microsporidia, metchnikovellids, and canonical microsporidia) as previously described[3]. **(B)** number of predicted proteins encoded by each genome. **(C)** Genome assembly size of each species. **(D)** Genome assembly N50 values. The scaffold N50 value is shown, except for those indicated by an * which are the contig N50 values. **(E)** Prescence of conserved microsporidian orthologs measured as the percent BUSCOs present in each genome.

651

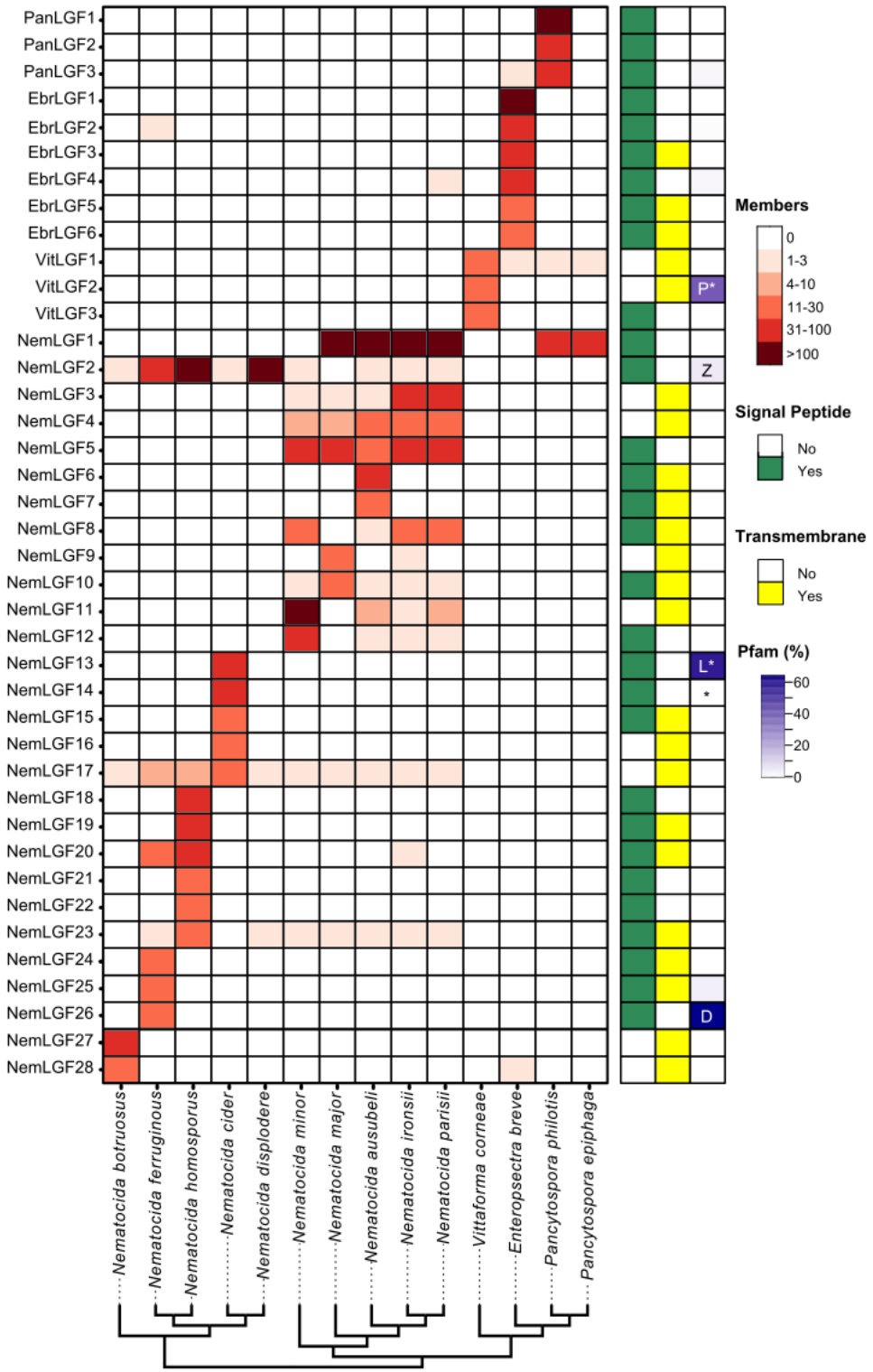
652 **Fig. 2. Conservation of protein function across microsporidian genomes.** Membership of
653 proteins from *R. allomycis* and 40 microsporidia species in Pombe GO-slim categories was
654 determined. The number of proteins from each species determined to belong to each GO slim
655 category is shown as a heatmap with GO-slim categories in rows and species in columns. Only
656 GO slim categories that contain at least one protein from any of these species are shown. Legend
657 for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom,
658 was constructed using Orthofinder[29]. Several species (*Pseudoloma neurophilia, Dictyocoela*

659 *roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality
660 genome assemblies (See Fig. 1).
661



662

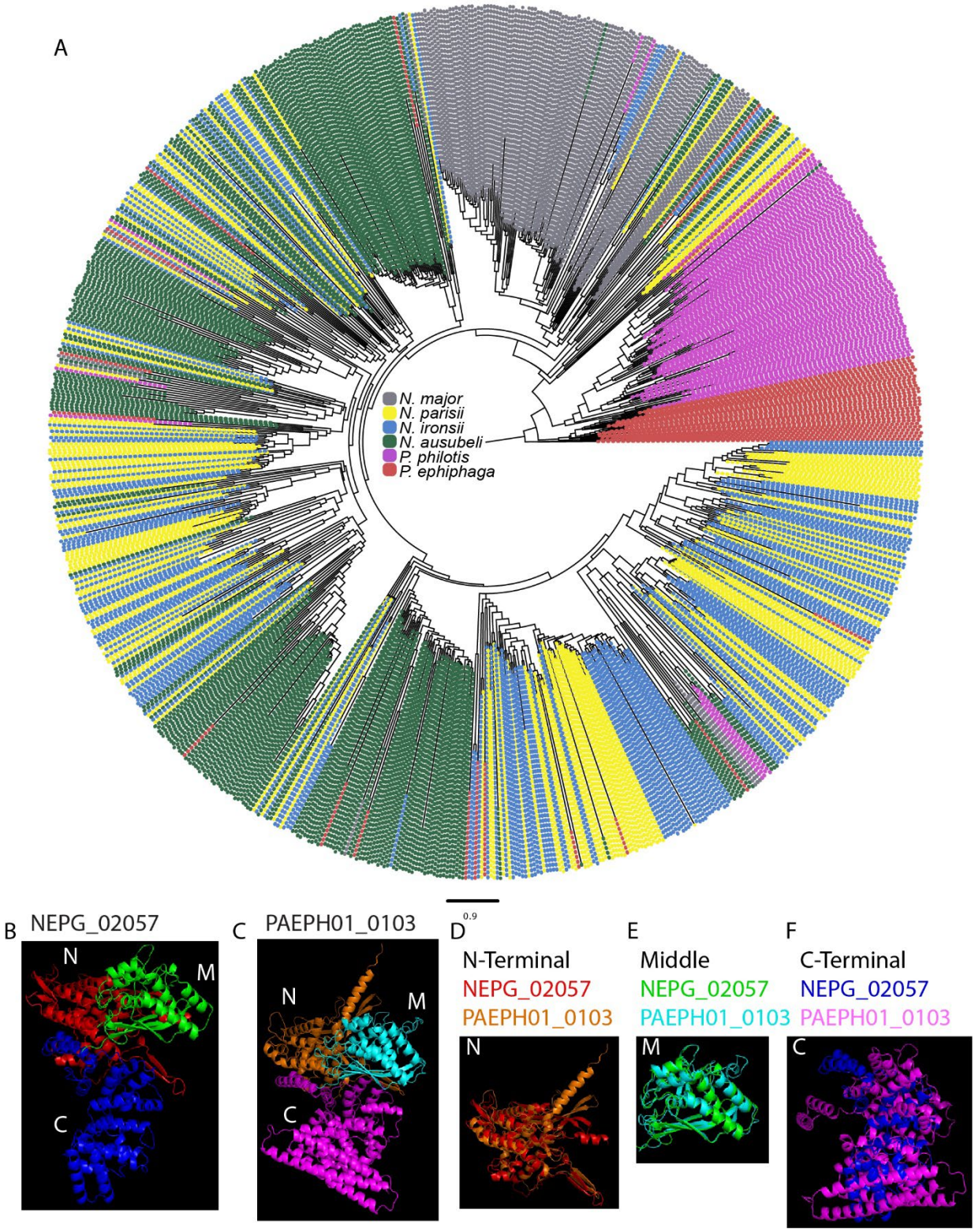
663 **Fig. 3. Prescence of metabolic enzymes in microsporidia species.** Membership of proteins
664 from *R. allomycis* and 40 microsporidia species in eight GO-slim metabolic categories was
665 determined. The number of proteins from each species determined to belong to each category is
666 shown as a stacked bar graph, with the legend for each metabolic Go-slim shown at the right.
667 Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species
668 (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema*
669 *apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).



670

671 **Fig. 4. Nematode infecting microsporidia contain many diverse large gene families.**

672 The number of members of each large gene family (listed in rows) in each species (listed in
673 columns) is shown as a heatmap according to the scale at the top right. The presence of predicted
674 signal peptides (green), transmembrane domains (yellow), or Pfam domains (purple gradient) are
675 shown to the right. Families containing the Pfam domains peptidase (P), Zinc finger (Z), LRR(L)
676 or DUF1609(D) are indicated. Domain types are displayed only for those families with more than
677 one annotated Pfam domain. Families that were determined using OrthoMCL are indicated with
678 * (See methods). Species are arranged according to the phylogenetic tree shown in Figure 1.
679


A
N. major
N. parisii
N. ironsii
N. ausubeli
P. philotis
P. ephiphaga
0.9
B NEPG_02057
N
M
C
C PAEPH01_0103
N
M
C
D
N-Terminal
NEPG_02057
PAEPH01_0103
N
E
Middle
NEPG_02057
PAEPH01_0103
M
F
C-Terminal
NEPG_02057
PAEPH01_0103
C

680

681

682 **Fig. 5. The large gene family NemLGF1 has likely been horizontally transferred between**
683 **genera.** (A) Phylogenetic tree of NemLGF1 family members, colored by species according to the
684 legend in the middle of the tree. Scale indicates changes per site. (B-C) AlphaFold models of *N.*
685 parisii NEPG_02057 (B) and *P. ephiphaga* PAEPH01_0103 (C). (D-F) Aligned structures of the
686 N-terminal (D), middle (E), and C-terminal (F) domains. N, N-terminal. M, middle. C, C-terminal.
687
688
689
690

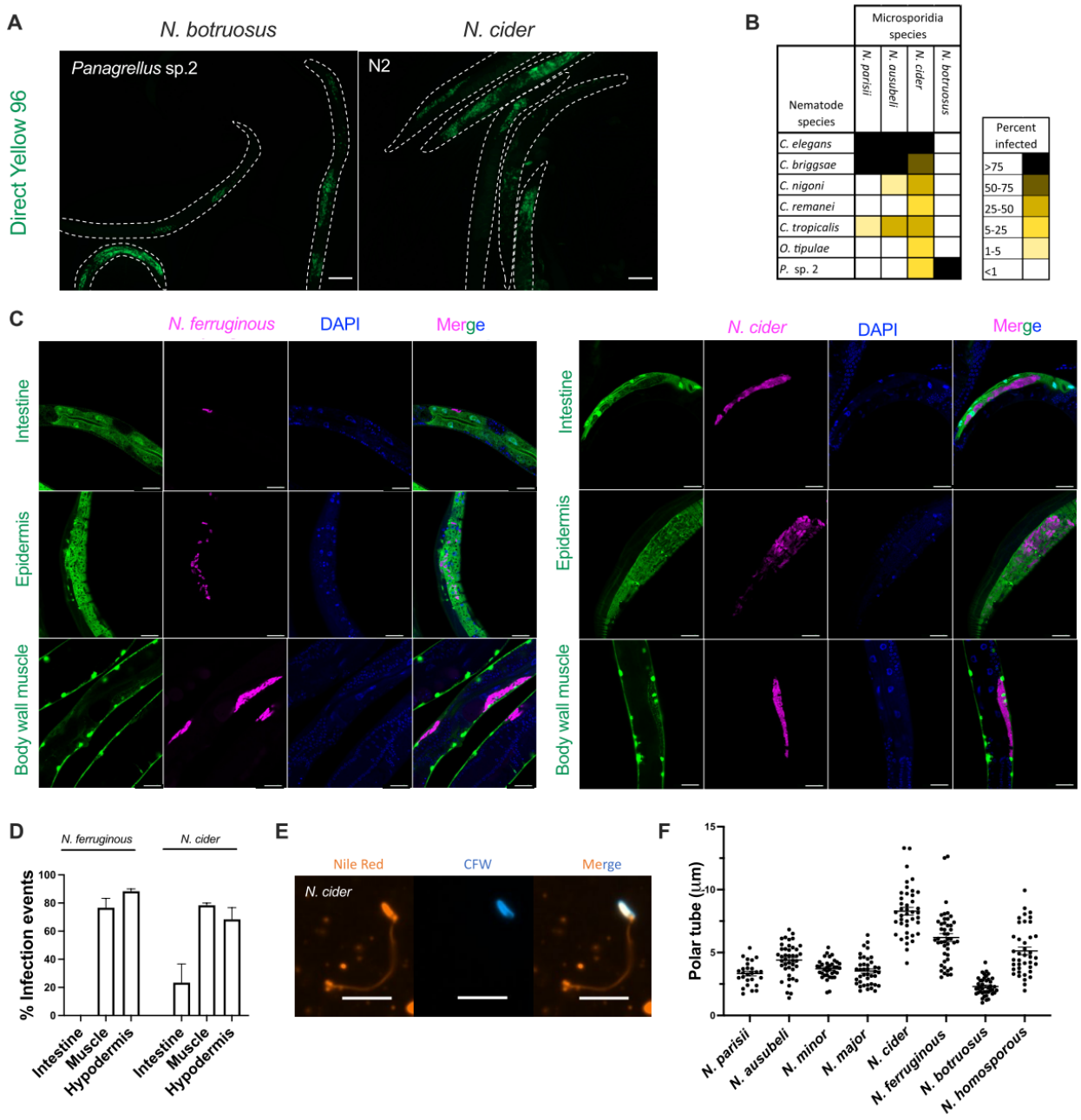
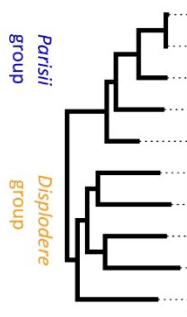


Fig. 6. Determination of host range, tissue specificity, and polar tube length of *Nematocida* species. (A-B) Seven species of L1 stage nematodes were infected with the following species (spore concentration): *N. parisii* (3 million), *N. ausubeli* (0.8 million), *N. cider* (4 million), and *N. botruosus* (28 million). After 96 hours of incubation with spores, animals were fixed and stained with direct yellow 96. **(A)** Representative images of *C. elegans* infected with *N. cider* and *Pangrellus* sp. 2 infected with *N. botruosus*. Scale bars, 100 µm. **(B)** Percentage of each population of animals infected with each species of microsporidia. Data is displayed as a heat map with host species in rows, microsporidia species in columns, and the value of each cell being

700 the percentage of each population that displayed newly formed microsporidia spores. Legend is
701 displayed on the right. Data shown are the average of two biological replicates with 20-391
702 animals counted for each sample. **(C-D)** L1 stage *C. elegans* strains expressing GFP in either the
703 intestine, epidermis, or body wall muscle were infected with either 1 million *N. cider* or *N.*
704 *ferruginous* spores. After 72 hours animals were fixed and stained with an *Nematocida* 18S RNA
705 FISH probe and DAPI. Note that for *N. ferruginous* infection of the GFP-intestinal strain (top), that
706 the infection is seen to lie outside of the GFP indicating a lack of intestinal infection. **(C)**
707 Representative images of infection of each species in each strain. Scale bars, 20 μm. **(D)** The
708 percentage of the population of each strain containing microsporidia infection within the marked
709 tissue. Data are from over 2 biological replicates and a total of 60 animals examined for each
710 condition. Graphs show mean with SD error bars. **(E-F)** Spores of *Nematocida* species were
711 induced to fire through repeated freeze thawing and then stained with Nile red and calcofluor
712 white. **(E)** Representative image of a stained *N. cider* polar tube. Scale bars, 5 μm. **(F)** The length
713 of polar tubes of each species is displayed. 26-41 polar tubes were measured for each species.
714 Mean ± SEM represented by horizontal bars.
715
716
717
718
719



	Host range	Major tissues infected	Spore size classes	Large spore clusters	Spore length (µm)	Spore width (µm)	Estimated spore volume (µm)	Polar tube length (µm)
N. parisii	Specialist	Intestine	2	No	2.18	0.8	0.73	3.3
N. ironsii	Specialist	Intestine	2	No	2.26	0.68	0.55	N/A
N. ausubeli	Specialist	Intestine	2	No	2.04	0.53	0.30	4.4
N. major	Specialist	Intestine	2	No	2.3	0.54	0.35	3.5
N. minor	Specialist	Intestine	2	No	1.3	0.55	0.21	3.7
N. displodere	Generalist	Intestine, Epidermis, Muscle	1	Yes	2.38	1.03	1.32	12.6
N. cider	Generalist	Intestine, Epidermis, Muscle	1	Yes	2.4	0.67	0.56	8.3
N. homosporus	Generalist	Intestine	1	Yes	2	0.72	0.54	5.1
N. ferruginous	Generalist	Epidermis, Muscle	1	Yes	1.72	0.6	0.32	6.2
N. botruosus	Specialist	Intestine	1	Yes	1.51	0.76	0.46	2.3

Fig. 7. *Nematocida* phenotypic properties. Table of phenotypic properties for *Nematocida* species, arranged according to the phylogenetic tree (left) determined in Fig. 1. Host range, tissues infected, spore size classes, presence of large spore clusters, spore size, and polar tube length was determined in this paper and from several previously published studies[18–20]. Estimated spore volume was calculated for spore length and width measurements as previously described[17]. N/A, Not available.

Supplemental material:

Fig. S1. Similarity of *N. ferruginous* assemblies.

Fig. S2. Phylogenetic tree of *Nematocida* species.

Fig. S3. Genome size and number of encoded proteins are correlated in nematode-infecting microsporidia.

Fig. S4. Conservation of carbohydrate metabolic processes across microsporidia genomes.

Fig. S5. Conservation of generation of precursor metabolites and energy pathways across microsporidia genomes.

Fig. S6. Conservation of tRNA metabolic processes across microsporidia genomes.

Fig. S7. Conservation of cellular amino acid metabolic processes across microsporidia genomes.

740 **Fig. S8. Conservation of lipid metabolic processes across microsporidia genomes.**
741 **Fig. S9. Conservation of sulfur compound metabolic processes across microsporidia**
742 **genomes.**
743 **Fig. S10. Conservation of mRNA metabolic processes across microsporidia genomes.**
744 **Fig. S11. Conservation of nucleobase-containing small molecule metabolic processes**
745 **across microsporidia genomes.**
746 **Fig. S12. Minimal ancestral deviation analysis of NEMLGF1 tree root.**
747 **Fig. S13. Structural models of nematode-infecting microsporidia large gene families.**
748 **Fig. S14. *N. ferruginous* infects multiple host species.**
749 **Fig. S15. Spore formation observed in the intestine for *N. botruosus* and throughout the**
750 **animal for *N. cider* and *N. ferruginous*.**
751 **Fig. S16. AAIM-1 is necessary for efficient infection by *N. cider.***
752
753 **Table S1. Summary of microsporidia species and their hosts.**
754 **Table S2. Methods used to assemble each microsporidia genome.**
755 **Table S3. Annotation of microsporidia genomes.**
756 **Table S4. Summary of genome assembly statistics.**
757 **Table S5. List of genome assemblies used.**
758 **Table S6. RNAi pathway proteins in *Pancytospora* and *Enteropsectra.***
759 **Table S7. List of large gene family proteins.**
760 **Table S8. NemLGF26 proteins present in other microsporidia species.**
761
762
763
764

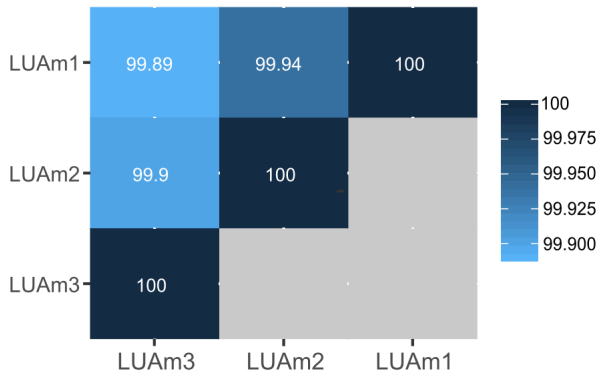
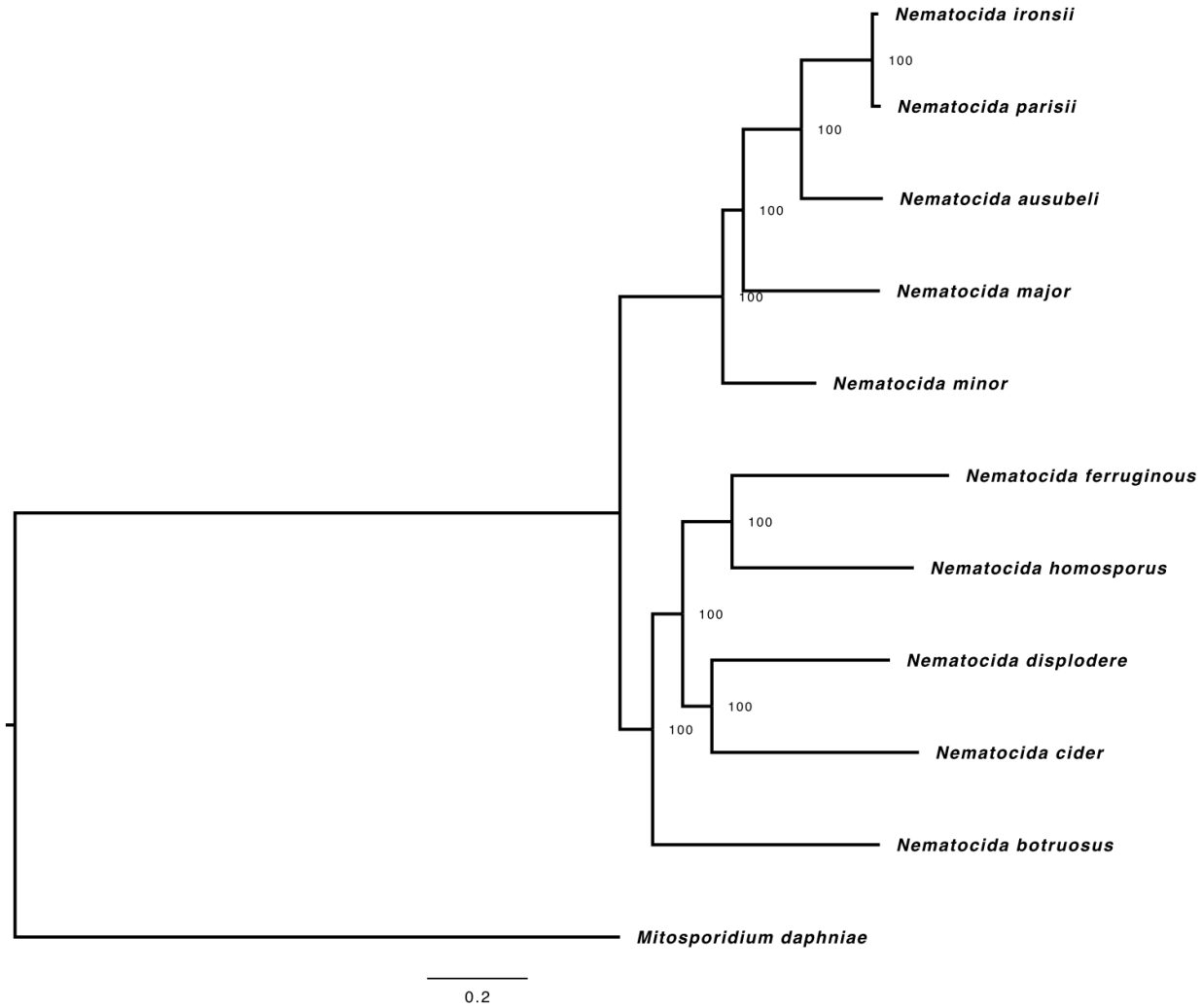
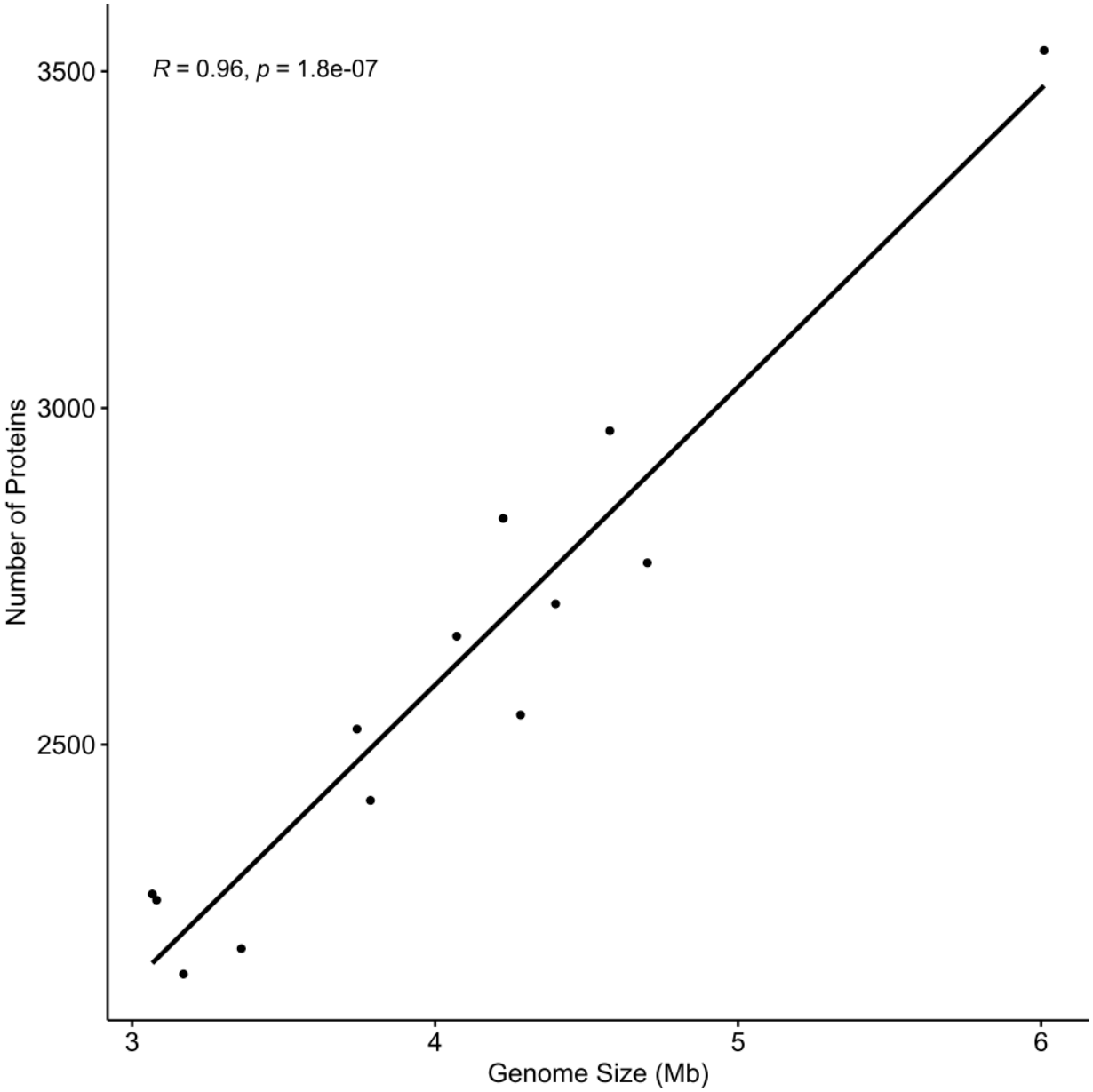


Fig. S1. Similarity of *N. ferruginous* assemblies.

The pairwise nucleotide identify of the three *N. ferruginous* assemblies (LUAm1-3) were calculated and displayed as heat map. The similarity between each assembly is colored according to the legend at the right.



771 **Fig. S2. Phylogenetic tree of *Nematocida* species.** The phylogeny of 10 Nematocida species
772 was determined from single-copy orthologs identified using OrthoMCL. Phylogenetic tree was
773 generated using RaxML. *M. daphniae* is shown as an outgroup. Bootstrap values are indicated at
774 each node. Scale indicates changes per site.
775



776
777 **Fig. S3. Genome size and number of encoded proteins are correlated in nematode-**
778 **infecting microsporidia.** The correlation between genome size and protein number in 13
779 nematode-infecting microsporidia genomes is shown as a scatter plot. Pearson correlation
780 coefficient and p-value are shown in the top left.

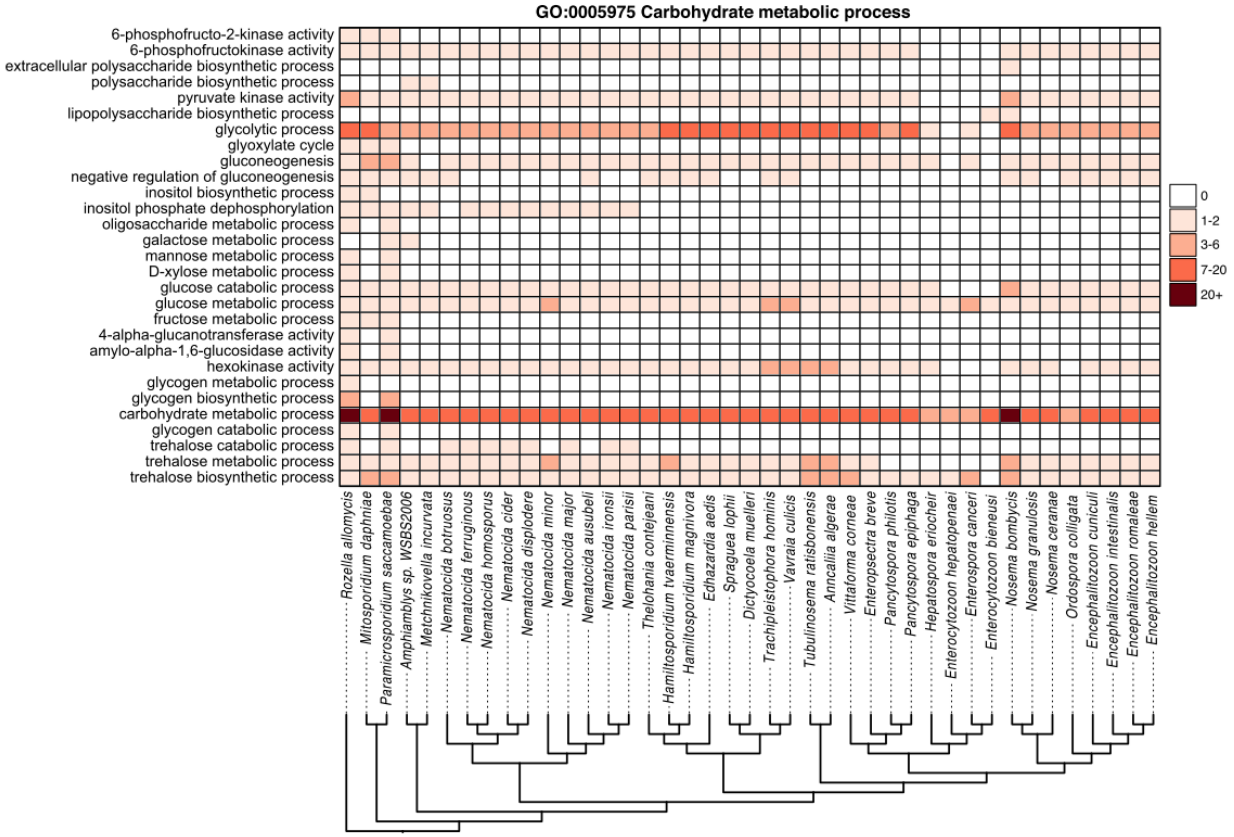


Fig. S4. Conservation of carbohydrate metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "carbohydrate metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

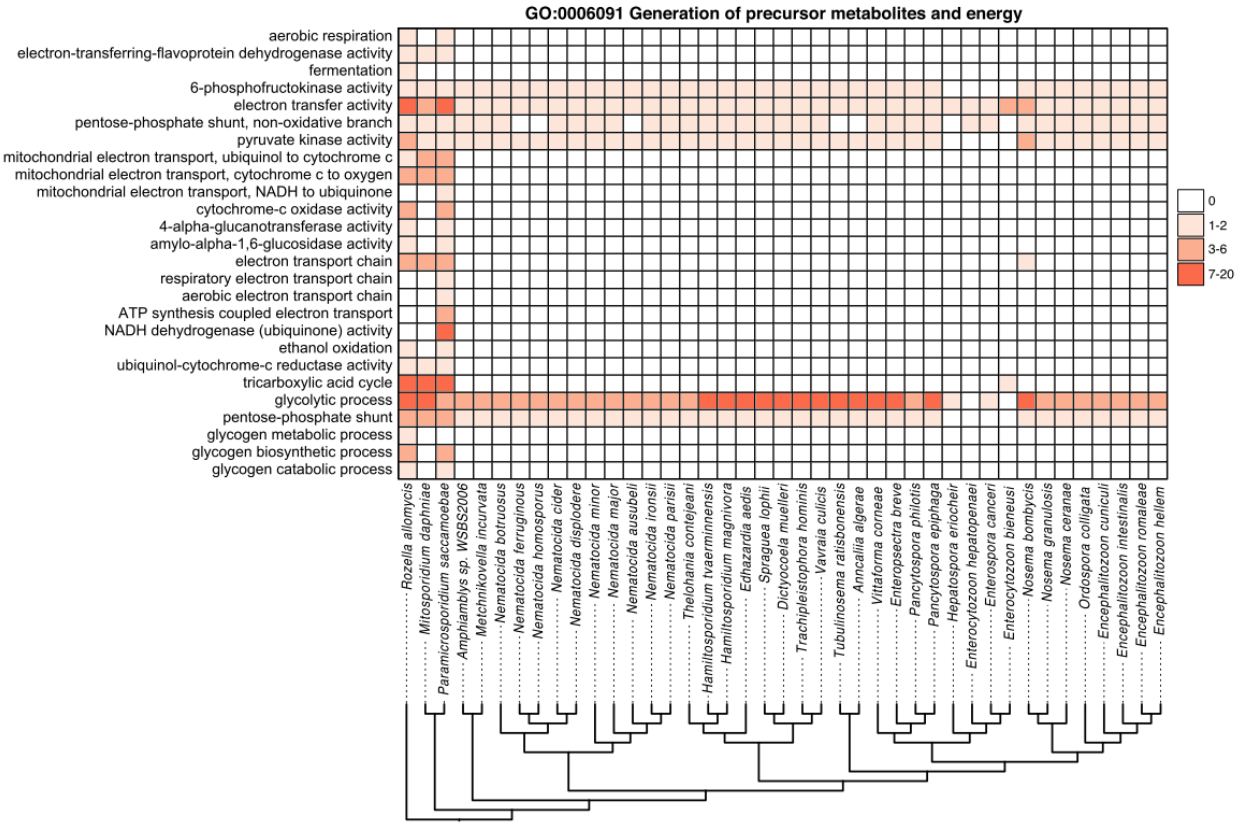


Fig. S5. Conservation of generation of precursor metabolites and energy pathways across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "generation of precursor metabolites and energy" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

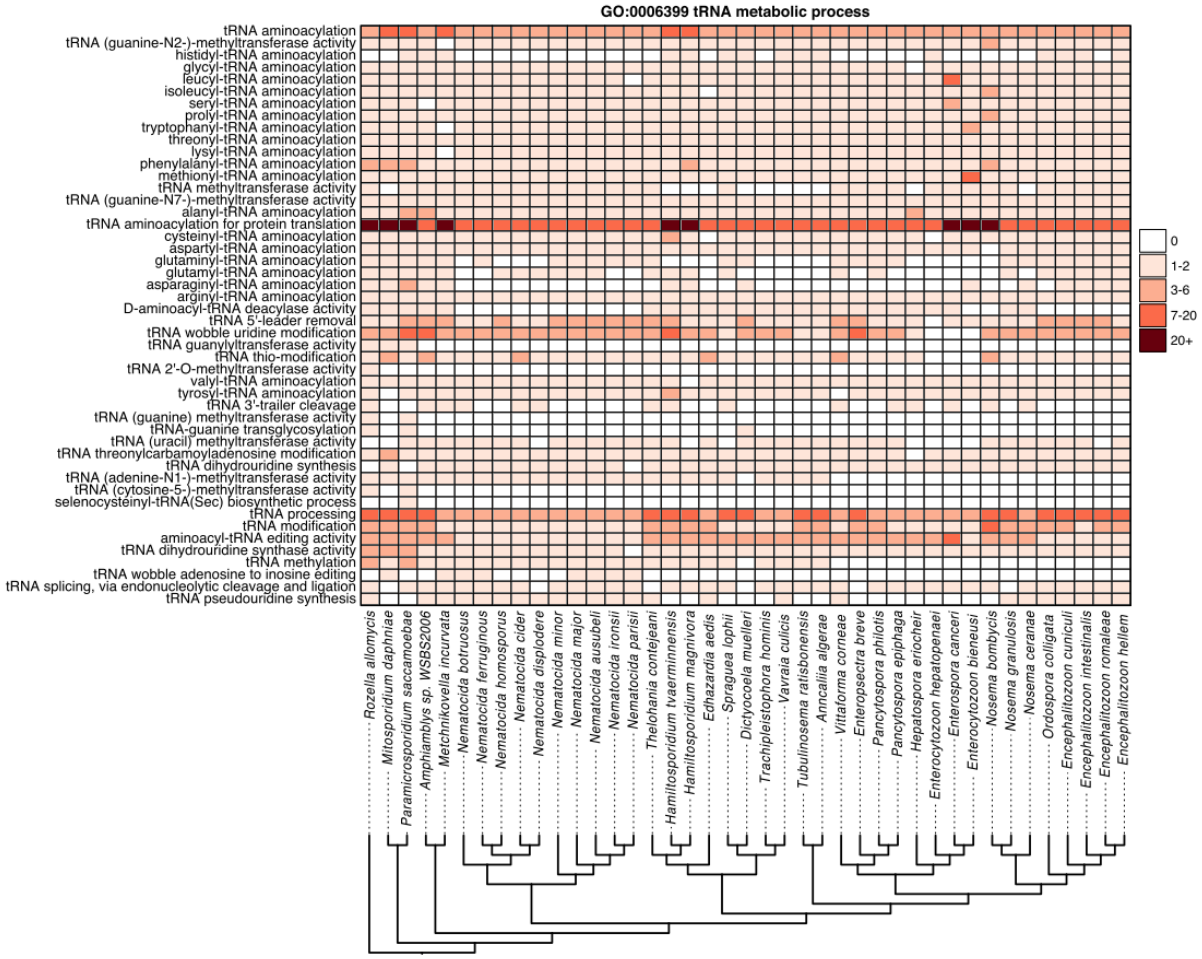


Fig. S6. Conservation of tRNA metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "tRNA metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

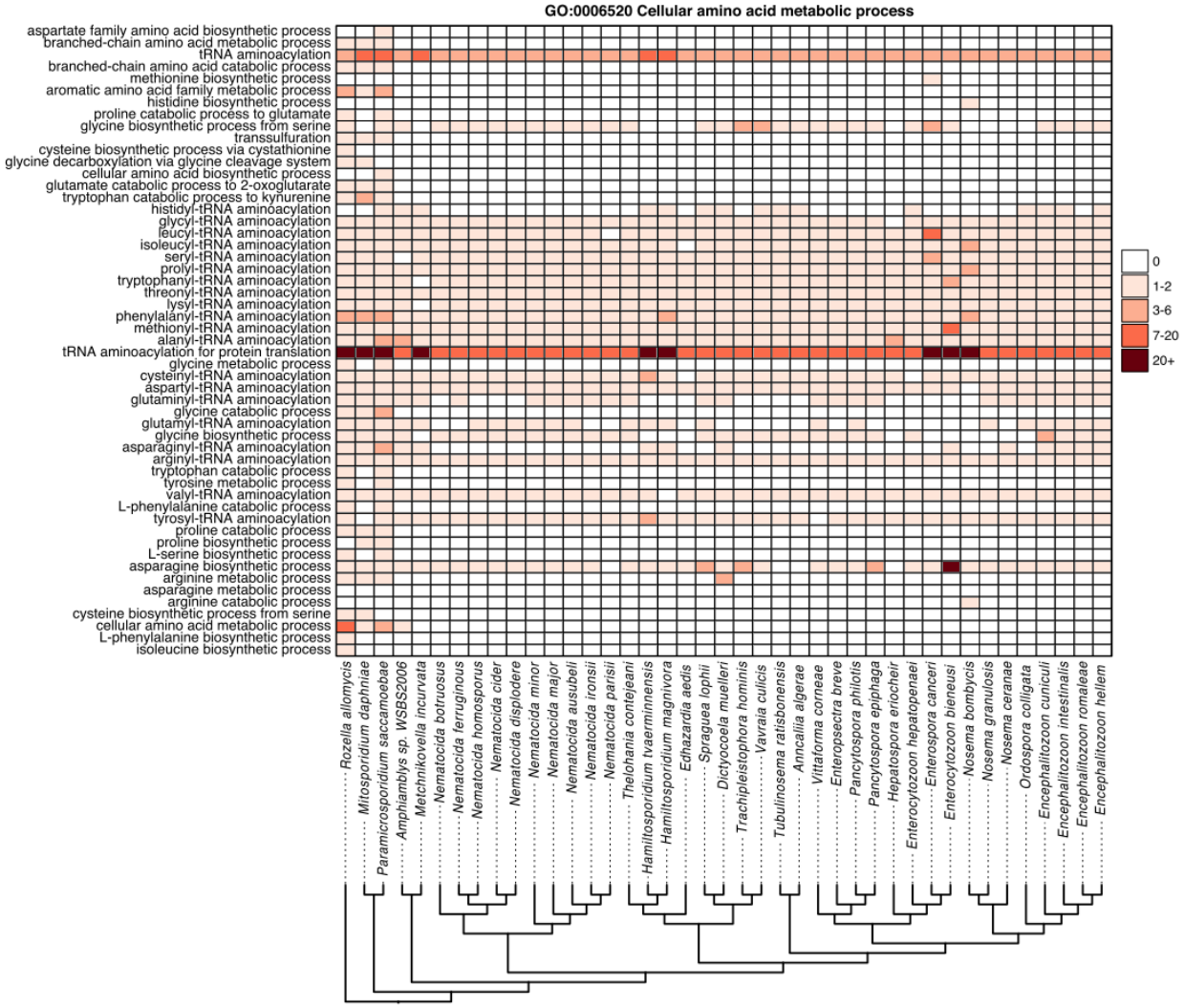


Fig. S7. Conservation of cellular amino acid metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "cellular amino acid metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

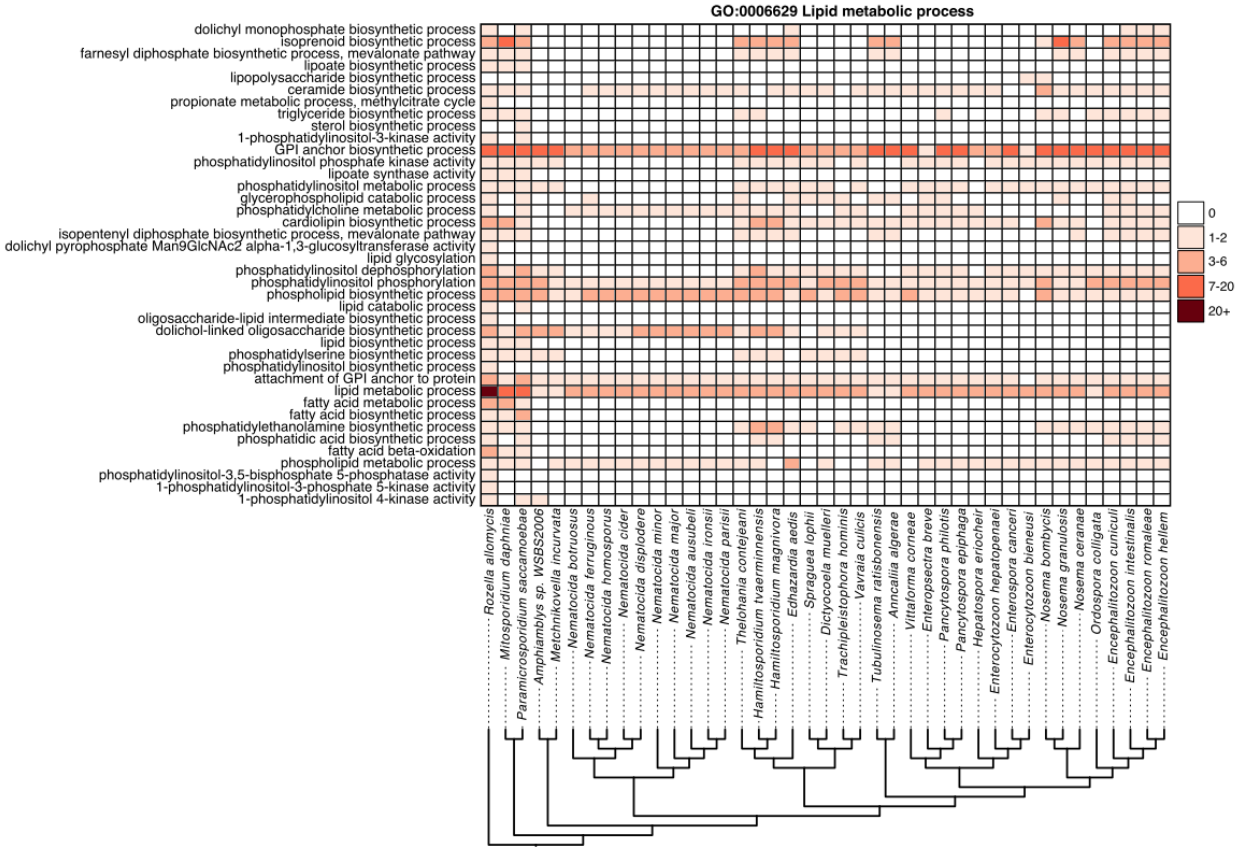


Fig. S8. Conservation of lipid metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "lipid metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

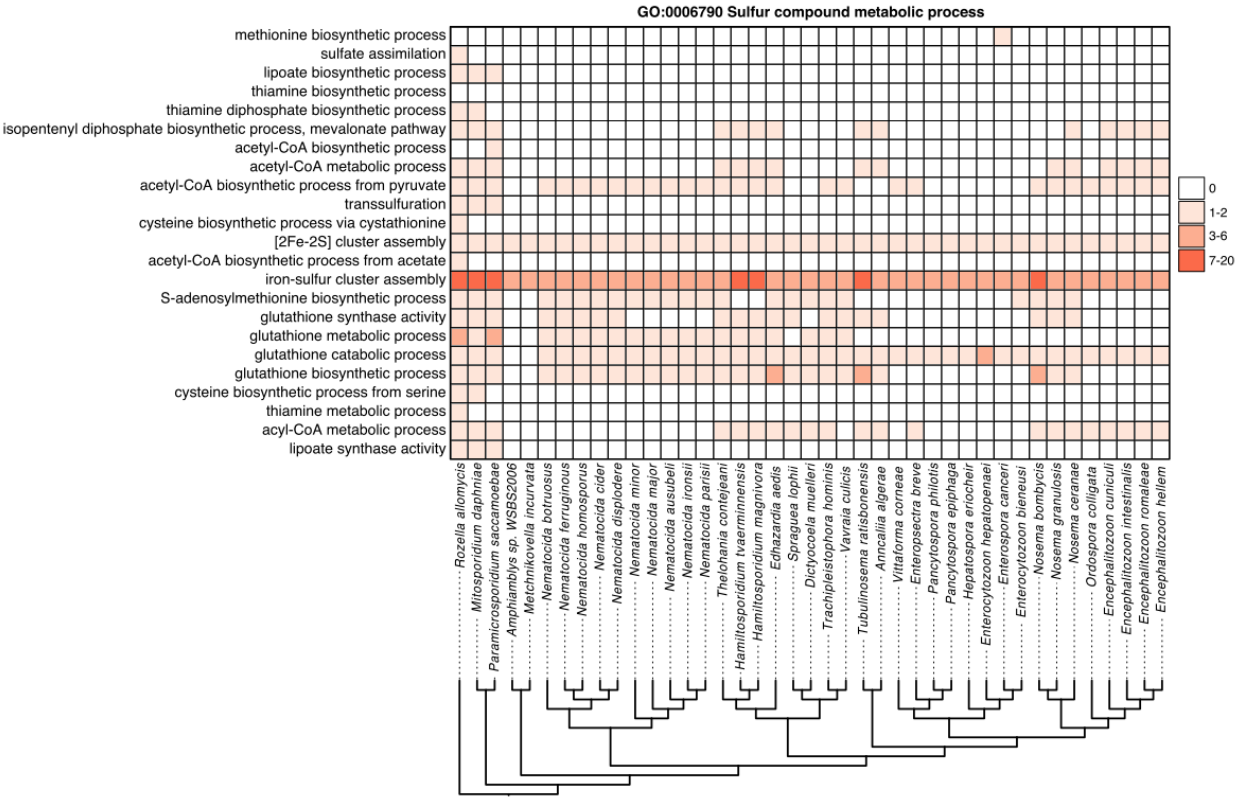


Fig. S9. Conservation of sulfur compound metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "sulfur compound metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

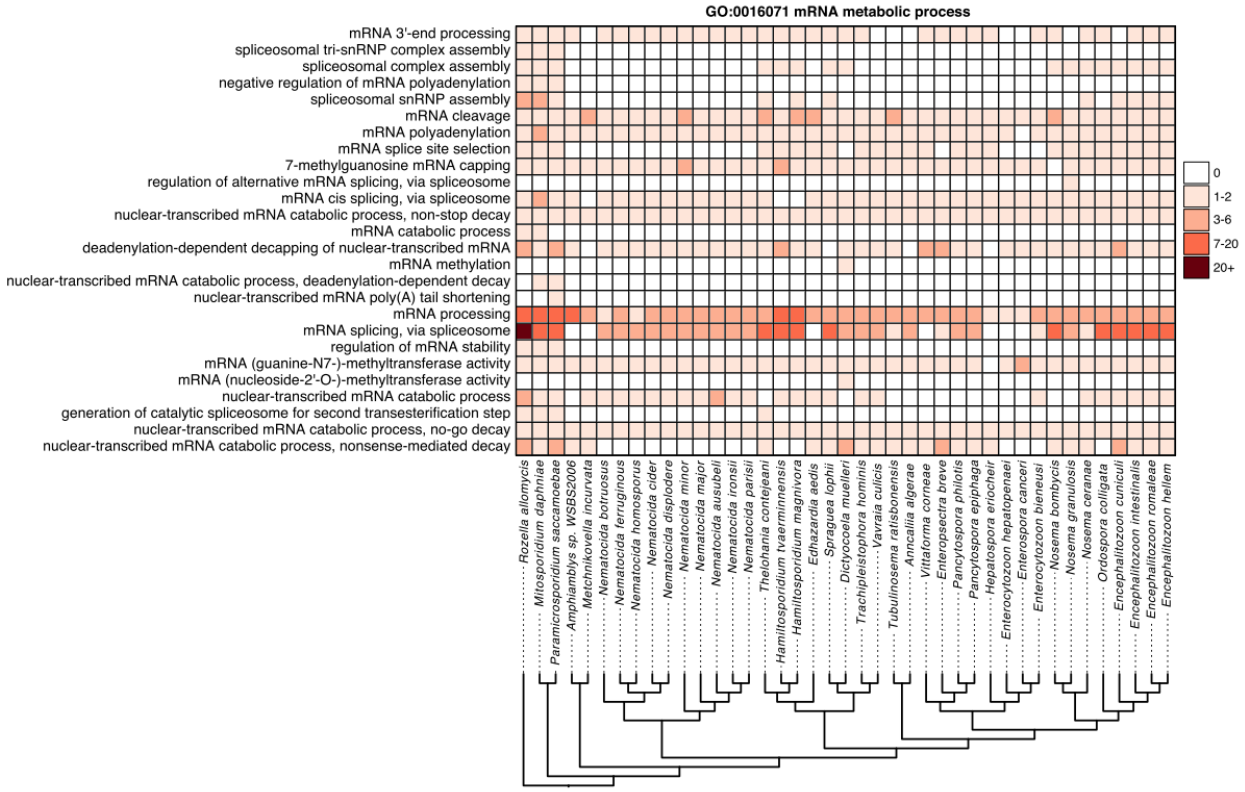


Fig. S10. Conservation of mRNA metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "mRNA metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).

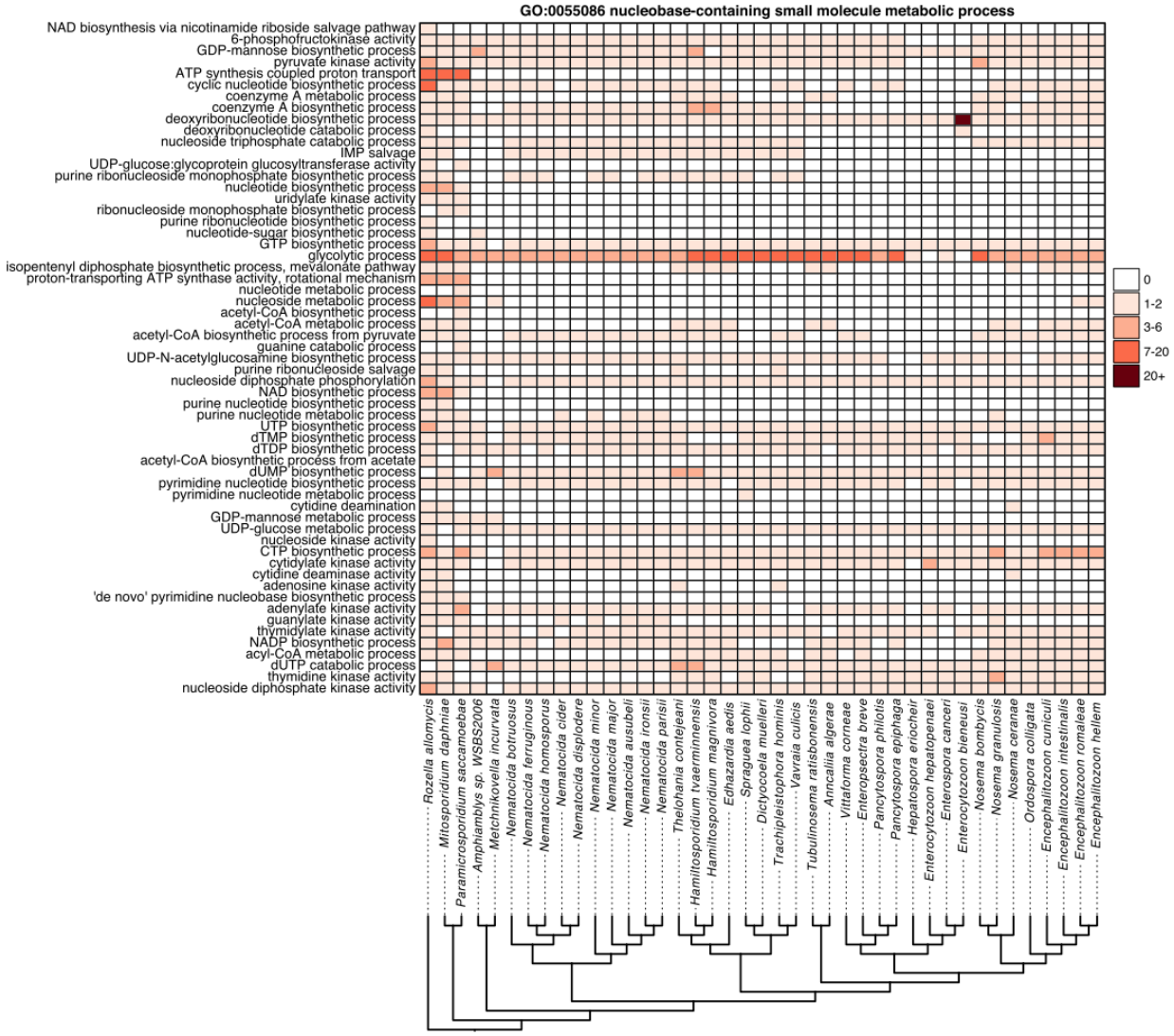
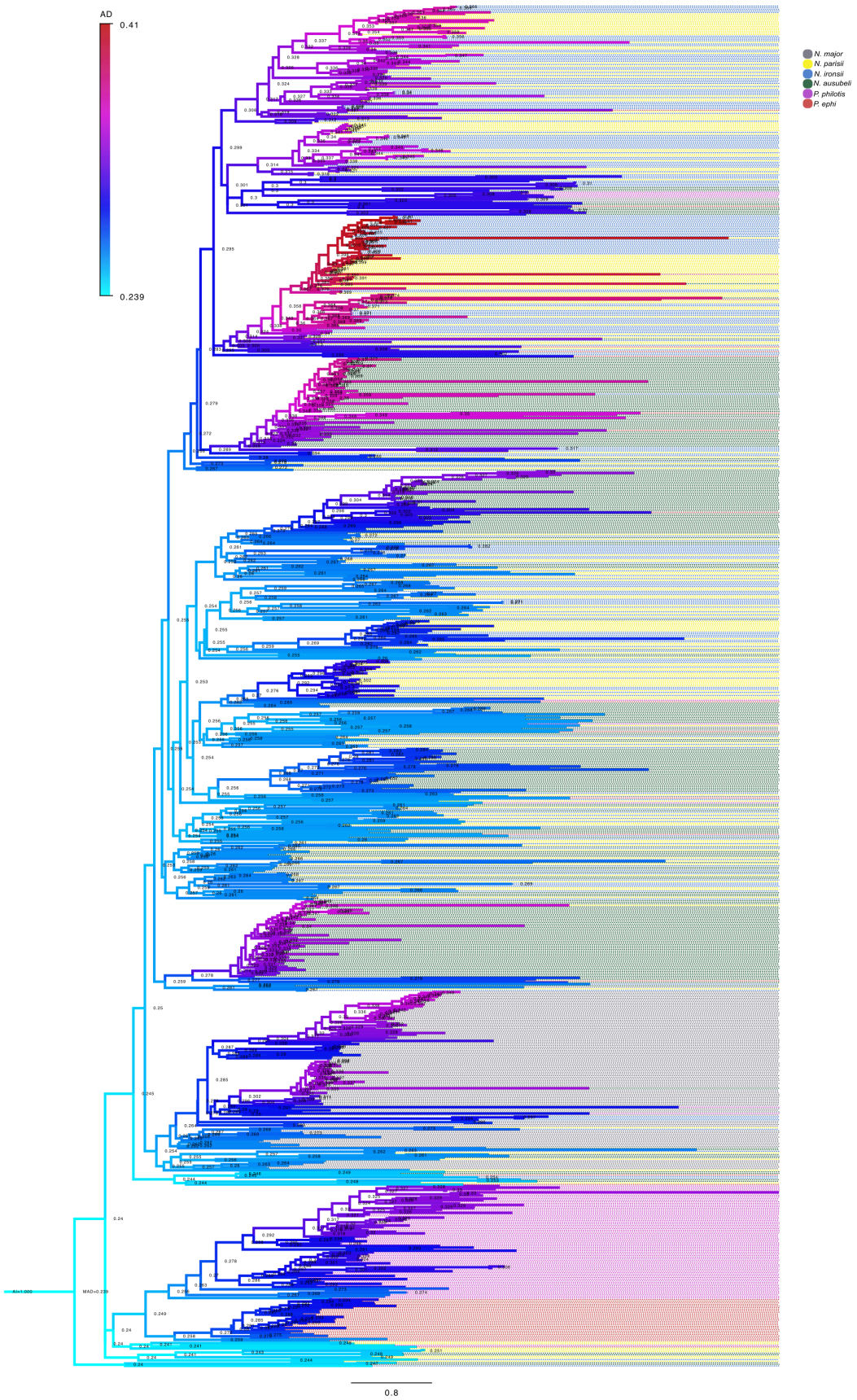


Fig. S11. Conservation of nucleobase-containing small molecule metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "nucleobase-containing small molecule metabolic process" was determined. The number of proteins from each species determined to belong to each Go term is shown as a heatmap with GO-slim categories in rows and microsporidia species in columns. Only descendant GO terms that contain at least one protein from any of these species is shown. Legend for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality genome assemblies (See Fig. 1).


AD
0.41
0.239
N. major
N. parisii
N. ironsii
N. ausubeli
P. philotis
P. ephi
0.8

891

892 **Fig. S12. Minimal ancestral deviation analysis of NEMLGF1 tree root.** Phylogenetic tree of
893 NemLGF1 family members, colored by species according to the legend at the top right. Each
894 branch is coloured according to ancestral deviation (AD) according to the scale at the top left.
895
896
897
898
899
900
901
902

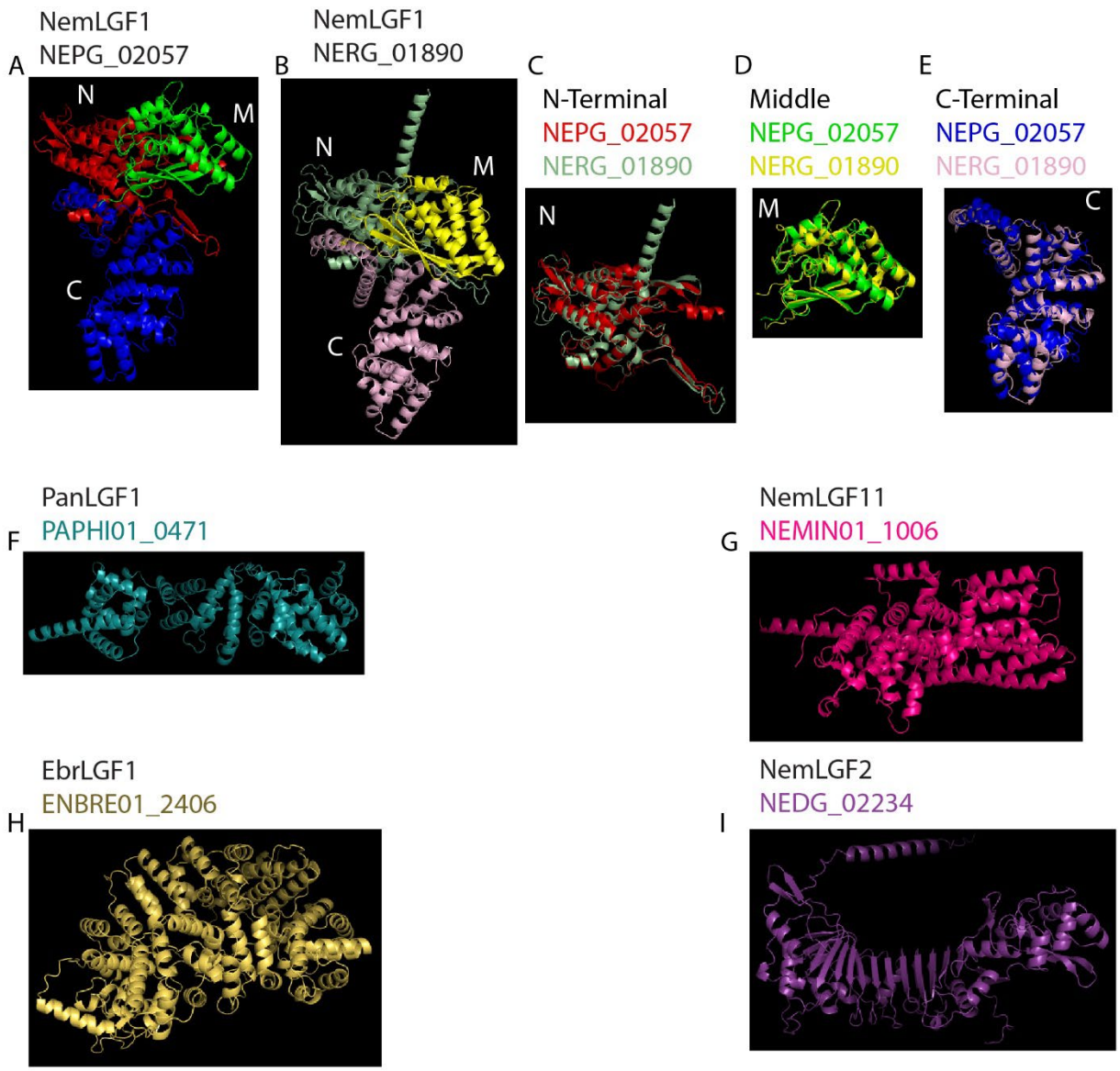


Fig. S13. Structural models of nematode-infecting microsporidia large gene families. The large gene family NemLGF1 has likely been horizontally transferred between genera. (A-B) AlphaFold models of *N.* parisii NEPG_02057 (B) and *N. ausubeli* NERG_01890 (B). (C-E) Aligned structures of the N-terminal (C), middle (D), and C-terminal (E) domains. N, N-terminal. M, middle. C, C-terminal. (F-I) AlphaFold models of PanLGF1 member PAPHI01_0471 (F), NemLGF11 NEMIN01_1006 (G), EbrLGF1 ENBRE01_2406 (H), and NemLGF2 NEDG_02234 (I).

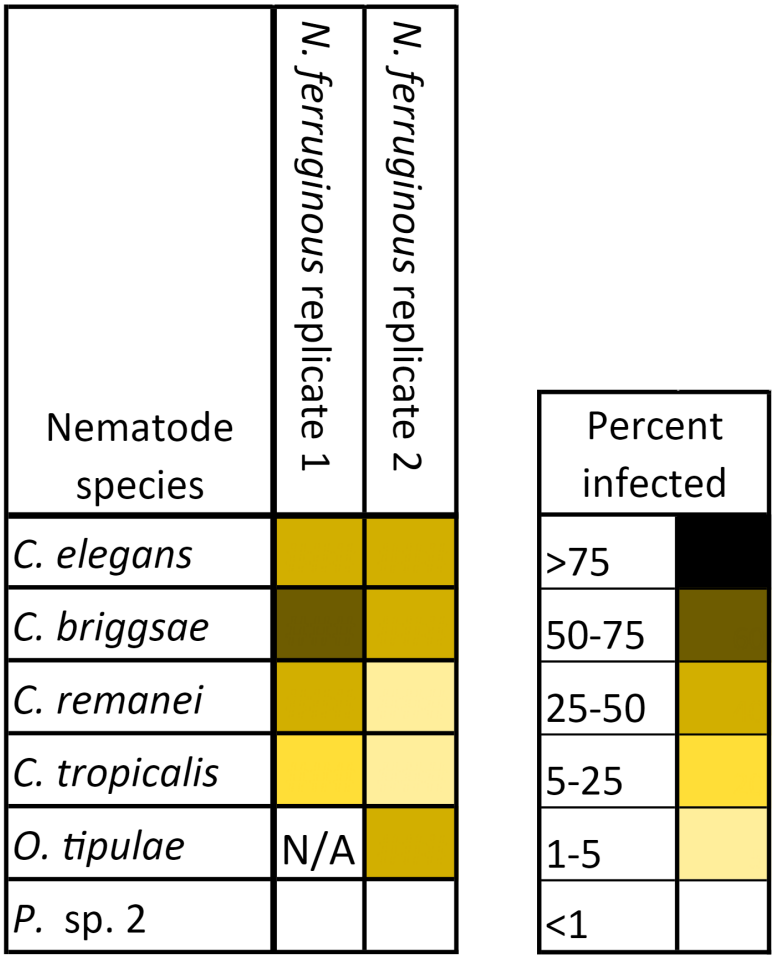


Fig. S14. *N. ferruginous* infects multiple host species. Six species of L1 stage nematodes were infected with either 20 (replicate 1) or 40 million (replicate 2) *N. ferruginous (*LUAm3) spores. After 96 hours of incubation with spores, animals were fixed and stained with direct yellow 96. Percent of each population of animals infected with each species of microsporidia. Data is displayed as a heat map with host species in rows, each *N. ferruginous* replicate in columns, and the value of each cell being the percent of each population that displayed newly formed microsporidia spores. Legend is displayed at the right. 50-196 animals were counted for each sample.

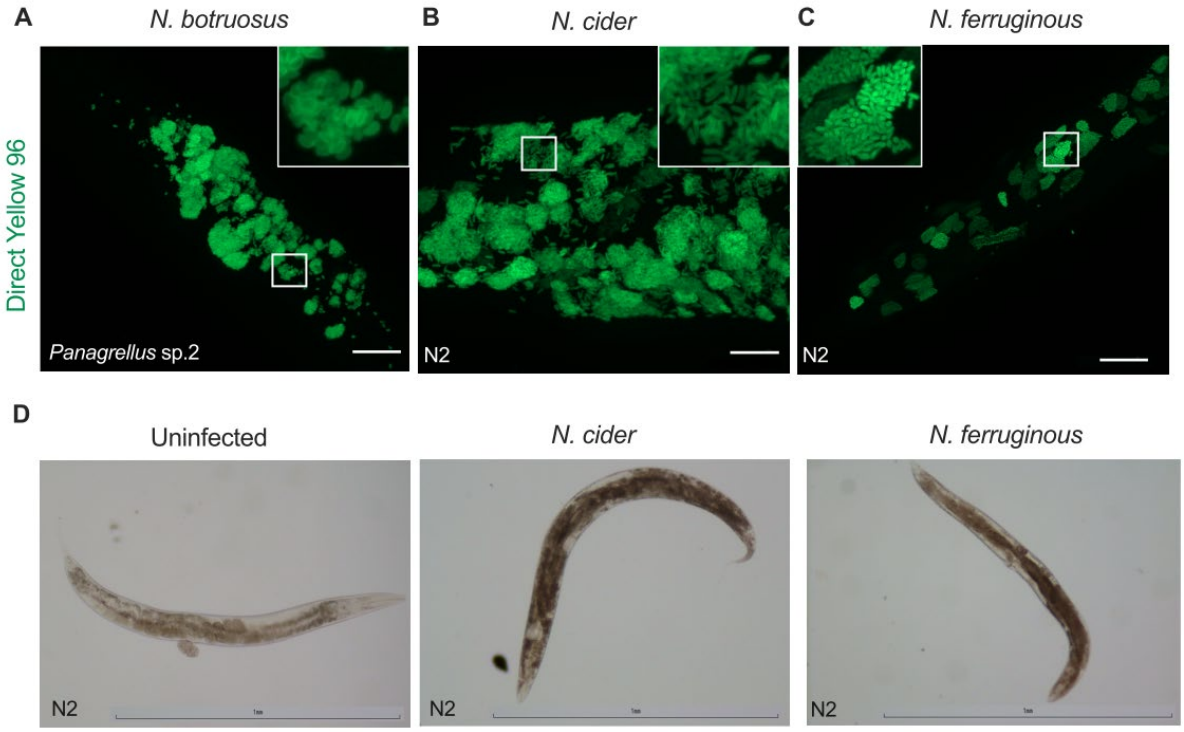
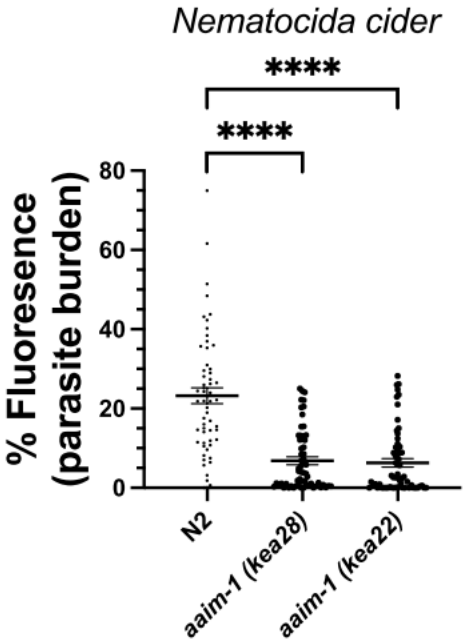


Fig. S15. Spore formation observed in the intestine for *N. botruosus* and throughout the animal for *N. cider* and *N. ferruginous*. (A-C) L1 stage animals were infected for 96 hours, fixed, and stained with DY96. Representative images were taken with the apotome module of a ZEISS Axio Imager at 63x magnification. Multiple z-planes were imaged, and a maximum intensity projection is displayed for each sample. Scale bars, 20 µm. **(A)** *Panagrellus* sp. 2 infected with 28 million *N. botruosus* spores. **(B)** *C. elegans* N2 infected with 4 million *N. cider* spores. **(C)** *C. elegans* N2 infected with 40 million *N. ferruginous* (LUAm3) spores. **(D)** *C. elegans* N2 animals were either infected with 3 million *N. cider* spores or 3 million *N. ferruginous* spores for 5 days. Images taken with Nikon Eclipse Ni at 10 x magnification and displayed.



938

939 **Fig. S16. AAIM-1 is necessary for efficient infection by *N. cider.*** N2 and *aaim-1* mutant
940 animals infected with 4 million *N. cider* spores, fixed at 96 hours, and stained with direct-yellow
941 96 (DY96). 20-30 worms quantified per replicate. % Fluorescence represents percentage of
942 animal carrying DY96 signal, thus reflecting parasite burden, measured via FIJI. Mean ± SEM
943 represented by horizontal bars. P-values determined via One-way Anova with post hoc.
944 Significance defined as **** $p < 0.0001$.

945

946 **References**

947 1. Vavra, J. & Lukes, J. Microsporidia and 'the art of living together'. *Adv Parasitol* **82**, 253–319 (2013).

948 2. Katinka, M. D. *et al.* Genome sequence and gene compaction of the eukaryote parasite
949 Encephalitozoon cuniculi. *Nature* **414**, 450–453 (2001).

950 3. Wadi, L. & Reinke, A. W. Evolution of microsporidia: An extremely successful group of eukaryotic
951 intracellular parasites. *PLoS Pathog* **16**, e1008276 (2020).

952 4. Capella-Gutiérrez, S., Marcet-Houben, M. & Gabaldón, T. Phylogenomics supports microsporidia as
953 the earliest diverging clade of sequenced fungi. *BMC Biology* **10**, 47 (2012).

954 5. Cuomo, C. A. *et al.* Microsporidian genome analysis reveals evolutionary strategies for obligate
955 intracellular growth. *Genome Res.* **22**, 2478–2488 (2012).
956 6. Li, Y. *et al.* A genome-scale phylogeny of the kingdom Fungi. *Current Biology* **31**, 1653-1665.e5
957 (2021).
958 7. Han, B., Takvorian, P. M. & Weiss, L. M. The Function and Structure of the Microsporidia Polar Tube.
959 in *Microsporidia: Current Advances in Biology* (eds. Weiss, L. M. & Reinke, A. W.) 179–213 (Springer
960 International Publishing, 2022). doi:10.1007/978-3-030-93306-7_8.
961 8. Quandt, C. A. *et al.* The genome of an intranuclear parasite, Paramicrosporidium saccamoebae,
962 reveals alternative adaptations to obligate intracellular parasitism. *eLife* **6**, e29594 (2017).
963 9. Haag, K. L. *et al.* Evolution of a morphological novelty occurred before genome compaction in a
964 lineage of extreme parasites. *PNAS* **111**, 15480–15485 (2014).
965 10. Nakjang, S. *et al.* Reduction and expansion in microsporidian genome evolution: new insights from
966 comparative genomics. *Genome Biol Evol* **5**, 2285–2303 (2013).
967 11. Barandun, J., Hunziker, M., Vossbrinck, C. R. & Klinge, S. Evolutionary compaction and adaptation
968 visualized by the structure of the dormant microsporidian ribosome. *Nat Microbiol* **4**, 1798–1804
969 (2019).
970 12. Ehrenbolger, K. *et al.* Differences in structure and hibernation mechanism highlight diversification of
971 the microsporidian ribosome. *PLOS Biology* **18**, e3000958 (2020).
972 13. Nicholson, D. *et al.* Adaptation to genome decay in the structure of the smallest eukaryotic
973 ribosome. *Nat Commun* **13**, 591 (2022).
974 14. Desjardins, C. A. *et al.* Contrasting host–pathogen interactions and genome evolution in two
975 generalist and specialist microsporidian pathogens of mosquitoes. *Nature Communications* **6**, 1–12
976 (2015).

977 15. Wiredu Boakye, D. *et al.* Decay of the glycolytic pathway and adaptation to intranuclear parasitism
978 within Enterocytozoonidae microsporidia. *Environ Microbiol* **19**, 2077–2089 (2017).
979 16. Willis, A. R. & Reinke, A. W. Factors That Determine Microsporidia Infection and Host Specificity. *Exp*
980 *Suppl* **114**, 91–114 (2022).
981 17. Murareanu, B. M. *et al.* Generation of a Microsporidia Species Attribute Database and Analysis of
982 the Extensive Ecological and Phenotypic Diversity of Microsporidia. *mBio* vol. 12 e01490-21 (2021).
983 18. Luallen, R. J. *et al.* Discovery of a Natural Microsporidian Pathogen with a Broad Tissue Tropism in
984 Caenorhabditis elegans. *PLoS Pathog.* **12**, e1005724 (2016).
985 19. Zhang, G. *et al.* A Large Collection of Novel Nematode-Infecting Microsporidia and Their Diverse
986 Interactions with Caenorhabditis elegans and Other Related Nematodes. *PLoS Pathog* **12**, e1006093
987 (2016).
988 20. Troemel, E. R., Felix, M.-A., Whiteman, N. K., Barriere, A. & Ausubel, F. M. Microsporidia are natural
989 intracellular parasites of the nematode Caenorhabditis elegans. *PLoS Biol* **6**, 2736–2752 (2008).
990 21. Reinke, A. W., Balla, K. M., Bennett, E. J. & Troemel, E. R. Identification of microsporidia host-
991 exposed proteins reveals a repertoire of rapidly evolving proteins. *Nat Commun* **8**, 14023 (2017).
992 22. Willis, A. R. *et al.* A parental transcriptional response to microsporidia infection induces inherited
993 immunity in offspring. *Science Advances* **7**, eabf3114 (2021).
994 23. Tamim El Jarkass, H. *et al.* An intestinally secreted host factor promotes microsporidia invasion of C.
995 elegans. *eLife* **11**, e72458 (2022).
996 24. Bakowski, M. A. *et al.* Ubiquitin-mediated response to microsporidia and virus infection in C.
997 elegans. *PLoS Pathog* **10**, e1004200 (2014).
998 25. Szumowski, S. C., Botts, M. R., Popovich, J. J., Smelkinson, M. G. & Troemel, E. R. The small GTPase
999 RAB-11 directs polarized exocytosis of the intracellular pathogen N. parisii for fecal-oral
1000 transmission from C. elegans. *Proc Natl Acad Sci U S A* **111**, 8215–8220 (2014).

1001 26. Murareanu, B. M., Knox, J., Roy, P. J. & Reinke, A. W. High-throughput small molecule screen
1002 identifies inhibitors of microsporidia invasion and proliferation in C. elegans. *bioRxiv*
1003 2021.09.06.459184 (2021) doi:10.1101/2021.09.06.459184.
1004 27. Tecle, E. & Troemel, E. R. Insights from C. elegans into Microsporidia Biology and Host-Pathogen
1005 Relationships. *Exp Suppl* **114**, 115–136 (2022).
1006 28. Jarkass, H. T. E. & Reinke, A. W. The ins and outs of host-microsporidia interactions during invasion,
1007 proliferation and exit. *Cellular Microbiology* **22**, e13247 (2020).
1008 29. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
1009 *Genome Biology* **20**, 238 (2019).
1010 30. Vossbrinck, C. R. & Debrunner-Vossbrinck, B. A. Molecular phylogeny of the Microsporidia:
1011 ecological, ultrastructural and taxonomic considerations. *Folia Parasitologica* **52**, 131–142 (2013).
1012 31. Bojko, J. *et al.* Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends in*
1013 *Parasitology* **38**, 642–659 (2022).
1014 32. Stentiford, G. D., Bass, D. & Williams, B. A. P. Ultimate opportunists—The emergent Enterocytozoon
1015 group Microsporidia. *PLOS Pathogens* **15**, e1007668 (2019).
1016 33. Huang, Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infection,*
1017 *Genetics and Evolution* **65**, 329–332 (2018).
1018 34. Dia, N. *et al.* InterB multigenic family, a gene repertoire associated with subterminal chromosome
1019 regions of Encephalitozoon cuniculi and conserved in several human-infecting microsporidian
1020 species. *Curr Genet* **51**, 171–186 (2007).
1021 35. Tria, F. D. K., Landan, G. & Dagan, T. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol*
1022 *Evol* **1**, 1–7 (2017).
1023 36. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682
1024 (2022).

1025 37. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
1026 (2021).
1027 38. Luallen, R. Infection strategies of related Nematocida microsporidian species in their natural host,
1028 Caenorhabditis elegans. (UC San Diego, 2016).
1029 39. Willis, A. R., Jarkass, H. T. E. & Reinke, A. W. Studying Inherited Immunity in a Caenorhabditis
1030 elegans Model of Microsporidia Infection. *JoVE (Journal of Visualized Experiments)* e63636 (2022)
1031 doi:10.3791/63636.
1032 40. Han, B. & Weiss, L. M. Microsporidia: Obligate Intracellular Pathogens within the Fungal Kingdom.
1033 *Microbiol Spectr* **5**, (2017).
1034 41. Weidner, E., Manale, S. B., Halonen, S. K. & Lynn, J. W. Protein-Membrane Interaction Is Essential to
1035 Normal Assembly of the Microsporidian Spore Invasion Tube. *Biol Bull* **188**, 128–135 (1995).
1036 42. Julian, A. T., Mascarenhas dos Santos, A. C. & Pombert, J.-F. 3DFI: a pipeline to infer protein function
1037 using structural homology. *Bioinformatics Advances* **1**, vbab030 (2021).
1038 43. Mascarenhas dos Santos, A. C., Julian, A. T. & Pombert, J.-F. The Rad9–Rad1–Hus1 DNA Repair
1039 Clamp is Found in Microsporidia. *Genome Biology and Evolution* **14**, evac053 (2022).
1040 44. Trzebny, A., Slodkowicz-Kowalska, A., Becnel, J. J., Sanscrainte, N. & Dabert, M. A new method of
1041 metabarcoding Microsporidia and their hosts reveals high levels of microsporidian infections in
1042 mosquitoes (Culicidae). *Molecular Ecology Resources* **20**, 1486–1504 (2020).
1043 45. Reinke, A. W. & Troemel, E. R. The Development of Genetic Modification Techniques in Intracellular
1044 Parasites and Potential Applications to Microsporidia. *PLoS Pathog* **11**, e1005283 (2015).
1045 46. Huang, Y. *et al.* A secretory hexokinase plays an active role in the proliferation of Nosema bombycis.
1046 *PeerJ* **6**, e5658 (2018).
1047 47. Paldi, N. *et al.* Effective Gene Silencing in a Microsporidian Parasite Associated with Honeybee (*Apis*
1048 *mellifera*) Colony Declines. *Applied and Environmental Microbiology* vol. 76 5960–5964 (2010).

1049 48. Saleh, M. *et al.* In Vitro Gene Silencing of the Fish Microsporidian Heterosporis saurida by RNA
1050 Interference. *Nucleic Acid Therapeutics* **26**, 250–256 (2016).
1051 49. Kim, I.-H., Kim, D.-J., Gwak, W.-S. & Woo, S.-D. Increased survival of the honey bee Apis mellifera
1052 infected with the microsporidian Nosema ceranae by effective gene silencing. *Archives of Insect*
1053 *Biochemistry and Physiology* **105**, e21734 (2020).
1054 50. Zheng, S. *et al.* The role of NbTMP1, a surface protein of sporoplasm, in Nosema bombycis infection.
1055 *Parasites & Vectors* **14**, 81 (2021).
1056 51. Chen, L., Li, R., You, Y., Zhang, K. & Zhang, L. A Novel Spore Wall Protein from Antonospora locustae
1057 (Microsporidia: Nosematidae) Contributes to Sporulation. *Journal of Eukaryotic Microbiology* **64**,
1058 779–791 (2017).
1059 52. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods* **30**,
1060 313–321 (2003).
1061 53. Barriere, A. & Felix, M.-A. Isolation of C. elegans and related nematodes. *WormBook* 1–19 (2014)
1062 doi:10.1895/wormbook.1.115.2.
1063 54. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1064 *EMBnet.journal* vol. 17 3 (2011).
1065 55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
1066 *Bioinformatics* **30**, 2114–2120 (2014).
1067 56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359
1068 (2012).
1069 57. Simpson, J. T. *et al.* ABySS: A parallel assembler for short read sequence data. *Genome Res.* **19**,
1070 1117–1123 (2009).
1071 58. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. Preprint at
1072 https://doi.org/10.12688/f1000research.12232.1 (2017).

1073 59. Pryszcz, L. P. & Gabaldón, T. Redundans: an assembly pipeline for highly heterozygous genomes.
1074 *Nucleic Acids Research* **44**, e113 (2016).
1075 60. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5–W9 (2008).
1076 61. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification.
1077 *BMC Bioinformatics* **11**, 119 (2010).
1078 62. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Computational Biology* **7**, e1002195 (2011).
1079 63. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic*
1080 *Acids Res* **44**, D279-285 (2016).
1081 64. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides
1082 from transmembrane regions. *Nat Methods* **8**, 785–786 (2011).
1083 65. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein
1084 topology with a hidden markov model: application to complete genomes11Edited by F. Cohen.
1085 *Journal of Molecular Biology* **305**, 567–580 (2001).
1086 66. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional
1087 Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**, 726–
1088 731 (2016).
1089 67. James, T. Y. *et al.* Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and
1090 Microsporidia. *Current Biology* **23**, 1548–1553 (2013).
1091 68. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing
1092 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**,
1093 3210–3212 (2015).
1094 69. Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein
1095 sequence annotations. *Nucleic Acids Research* **47**, D351–D360 (2019).

1096 70. Li, L., Stoeckert, C. J. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic
1097 genomes. *Genome Res* **13**, 2178–2189 (2003).
1098 71. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment | Nature
1099 Methods. https://www.nature.com/articles/nmeth.1818.
1100 72. Using Dali for Protein Structure Comparison - PubMed.
1101 https://pubmed.ncbi.nlm.nih.gov/32006276/.
1102 73. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**,
1103 676–682 (2012).
1104 74. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis.
1105 *Nat Methods* **9**, 671–675 (2012).
1106