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. 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 phylogenetic analysis shows that *Nematocida* is composed of two groups of species with distinct traits and that species with longer polar tubes 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.

Author Summary

Microsporidia are microbial parasites that can infect many animals. Nematodes have become a useful system to study microsporidia as these animals are commonly infected by microsporidia and these infections can be easily studied in a laboratory environment. To better understand how microsporidia evolve and change their properties as they infect different hosts, we sequenced the genomes of nine microsporidia species which infect nematodes. We found that metabolic pathways were mostly conserved between different clades of these species. Surprisingly, we found a family of proteins predicted to facilitate host interactions that is present in two distinct genera, suggesting that these genes were transferred between microsporidia species. We also determined the hosts and tissues that these microsporidia can infect as well as morphological properties of the microsporidia. We show that these properties are quite variable in related microsporidia species. Our results provide insight into the evolution of microsporidia and also provide a resource for studying microsporidia infections in nematodes.

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 of 2.9 megabase [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 orthologs [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 (spore size, spore shape, and polar tube length) and infection properties (hosts and host tissues), even within related clades, suggesting these parasites are quite evolutionarily labile 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 [21–30]. Other species that infect the intestine, *N. ausubeli* and *N. ironsii,* have also been identified [20,21]. Another species, *N. displodere,* was found to infect other tissues of *C. elegans,* including the neurons, muscle, and epidermis [18]. Several other microsporidia species that infect *C. elegans* or other free-living nematodes have also been identified [19]. In addition to *Nematocida*, two other genera, *Enteropsectra* and *Pancytospora,* have been reported

to infect nematodes and these genera have been shown to be 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]. Previous work showed that all sequenced species of microsporidia contained at least one large gene family, and most families were clade specific [21]. These families were also enriched for secretion signals and transmembrane domains, suggesting that proteins in these families are used to interface with the host [10,21]. 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]. This family has only been reported in these *Nematocida* species [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 be involved in protein-protein interactions [19].

Here we present a comprehensive and comparative analysis 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* are related to the human-infecting species *Vittaforma corneae* and together with this species form a sister group to the marine invertebrate- and human-infecting *Enterocytozoon*. We find that 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 relationships of nematode-infecting microsporidia and provides a valuable resource for the use of nematodes to study mechanisms of microsporidian infection.

Results

130 **Sequencing, assembling, and phylogenetic analysis of microsporidian genomes**
131 Previous efforts in discovering microsporidia in free-living terrestrial nematodes have revealed a
132 diversity of species and four species from the genus *Nematocida* have been sequenced
133 [5,18,19,21]. To further understand the genomic properties of nematode-infecting microsporidia,
134 we sequenced nine additional species, including two genera that had not been previously
135 sequenced. Six of these microsporidian species have been previously reported [19]. We recently
136 discovered three novel species through sampling of wild nematode populations; *Nematocida*
137 *ferruginous* from France, *Nematocida cider* from the United States, and *Nematocida botruosus*
138 from Canada. To sequence these genomes, we removed contaminating bacteria from
139 microsporidia-infected nematodes using antibiotics, cultured infections in the host animals,
140 isolated spores, and extracted DNA. We sequenced a single isolate for each species, except for
141 *N. ferruginous* for which we sequenced three strains that have high similarity to one another (S1
142 Fig). Information on each microsporidia species, host, and isolation location is presented in S1
143 Table. Microsporidian genomes were assembled and contaminating DNA were removed (See
144 Methods and S2 Table). Proteins from each genome were predicted and annotated (S3 Table).
145 These new microsporidian genome assemblies were assessed based on contig continuity and
146 protein conservation and were demonstrated to be of high quality relative to previously assembled
147 microsporidian genomes (Fig 1D, E, and S4 Table).
148
149 To determine the relative relationships between microsporidia species we performed phylogenetic
150 analysis. Using OrthoFinder [31], we identified orthologous microsporidia proteins and generated
151 a phylogenetic tree of our nine newly assembled microsporidia species, 35 previously sequenced
152 microsporidia genomes, and the outgroup *Rozella allomycis* (Fig 1A and Table S5). This tree
153 shows that *Nematocida* genomes form two groups. The first identified *Nematocida* species, *N.*
154 *parisii*, forms a group with *N. ausubeli*, *N. ironsii*, *N. major*, and *N. minor*; we refer to these species
155 as the "*Parisii"* group [19–21]. The second group contains the more recently identified *N.*
156 *displodere* along with *N. homosporus*, *N. cider*, *N. ferruginous,* and *N. botruosus;* we refer to these
157 species as the "*Displodere"* group [18,19]. The relative positions of the *Nematocida* species are
158 highly supported by several phylogenetic trees (Fig 1A and S2 Fig). As described previously from
159 18S rDNA and β–tubulin sequences, the species *Enteropsectra breve*, *Pancytospora philotis*, and
160 *Pancytospora epiphaga* belong to the large group of microsporidia species referred to as the
161 Enterocytozoonida clade [17,19,32,33]. Our whole genome based tree replicates this previously
162 described phylogeny, demonstrating that these three non-*Nematocida* species are most closely

163 related to the human-infecting *V. corneae*, and together they form a sister group to the
164 *Enterocytozoon* [15,34].
165
166 **Conservation and divergence of proteins in microsporidia genomes**
167 Microsporidia have the smallest known eukaryotic genomes, but genome size and protein content
168 can vary greatly between species [3]. In *Nematocida,* we see consistently small genome sizes
169 that range from ~3.1 to 4.7 megabases. The two *Pancytospora* species genomes are similar to
170 each other, with sizes of 4.2 and 4.6 megabases. In contrast, the ~6.0 megabase genome of *E.*
171 *breve* is the largest of our nematode-infecting microsporidia species genomes (Fig 1C). We see
172 similar trends in predicted number of proteins, with *N. botruosus* having the fewest proteins (2159)
173 and *E. breve* having the most predicted proteins (3531) (Fig 1B). Although microsporidia species
174 can have large amounts of non-coding DNA that make their genomes much larger relative to their
175 protein content [14], we observe a strong correlation between protein content and genome size
176 among nematode-infecting microsporidia (S3 Fig).
177
178 Microsporidia are known to have lost many conserved eukaryotic proteins, including those
179 involved in metabolism [10]. To determine the conservation of metabolic function in nematode-
180 infecting microsporidia species, we analyzed proteins encoded by microsporidia genomes using
181 GO term categories. We first determined the number of proteins from each microsporidia species
182 that belong to a set of GO terms that cover most conserved eukaryotic cellular processes (Fig 2).
183 This analysis revealed that many cellular processes were either lost or greatly reduced in the
184 canonical microsporidia, which has been previously observed [8]. We then determined the
185 proteins from each microsporidian species that belonged to eight metabolic categories (S4-11
186 Figs, summarized in Fig 3). As has been previously observed, we see that the *Enterocytozoon*
187 contains many losses in metabolic pathways, and in particular those involved in energy production
188 [15]. As the genera *Pancytospora* and *Enteropsectra,* along with *V. cornea,* form a sister group to
189 the *Enterocytozoon,* we sought to determine how the metabolic capacity of these groups of
190 species compared. We observe that *Pancytospora* and *Enteropsectra* have a similar metabolic
191 capacity compared to most other microsporidia species, including the *Nematocida* (Fig 3).
192 However, there are several notable differences between the genera, with *Pancytospora* and
193 *Enteropsectra* having losses in glutathione biosynthesis and the *Nematocida* having losses in
194 phosphatidylinositol biosynthesis.
195

196 The RNAi pathway has been lost multiple times from microsporidia genomes [14,35]. *Nematocida*
197 was previously observed to lack this pathway [3]. However, the genomes of the three species
198 from *Pancytospora* and *Enteropsectra* each encode RNAi pathway proteins (S6 Table).
199
200 **Nematode-infecting microsporidia contain many large, expanded gene families**
201 A striking feature of microsporidia genomes are abundant large gene families that contain
202 paralogous family members with either signal peptide or transmembrane domains [10,21]. Using
203 a previously described bioinformatic approach [21], we analyzed the large gene family content of
204 nematode-infecting microsporidia, identifying 39 families in 13 species, that were enriched in
205 either signal peptide or transmembrane domains [21](Figs 4, S12, and S7 Table). We observed
206 that every species had at least one large gene family, with six species having at least five families
207 with more than 10 members each. Some of these families have many members, with 5 families
208 (NemLGF1, NemLGF2, NemLGF11, PanLGF1, and EbrLGF1) containing more than 100
209 members in at least one species. Previously, we had observed that almost all large gene families
210 are specific to individual clades of microsporidia species [21]. Most families (30/37) in nematode-
211 infecting microsporidia have at least 10 members in only one species. The families we identified
212 in *Pancytospora* and *Enteropsectra* are also unique to these species and not found in *V. cornea*.
213 Some of the large gene families, such as NemLGF11, are present in small numbers in other
214 species, but are only greatly expanded in one of the species. To determine if any of the large
215 gene families identified in nematode-infecting microsporidia were present in other microsporidia
216 species, we searched other microsporidia genomes using family-specific models (S8 Table). The
217 only family with more than 10 members present in microsporidia species that don't infect
218 nematodes is NemLGF26, which includes the previously described InterB family that contains the
219 Pfam domain Duf1609.This domain is also present in NemLGF26 family members [36].
220
221 One of the most abundant *Nematocida* large gene families is NemLGF1 [5,21]. This large family
222 was previously found in the genomes of *N. parisii*, *N. ausubeli*, and *N. ironsii*, with each species
223 containing between 160-240 proteins [21]. We also observe 105 members of this family in *N.*
224 *major*. We do not observe this family in *N. minor* or any of the genomes in the *Displodere* group,
225 suggesting that the family appeared in the *Parisii* group after the divergence of *N. minor*. Strikingly,
226 we find that the NemLGF1 family is also present in *P. philotis* (81 members) and *P. epiphaga* (45
227 members). We generated a phylogenetic tree using all the members from these six species, which
228 shows distinct expansions of NemLGF1 proteins in the *Nematocida* and *Pancytospora* species

229 (Fig 5A). As this family has not been previously reported in any other species of microsporidia,
230 we propose that this family was likely shared between clades via horizontal gene transfer.

231 To determine the similarity of NemLGF1 protein from *Nematocida* and *Pancytospora* species, we
232 used AlphaFold to model the protein structures [37,38] (Fig 5B and C). We find that NemLGF1
233 proteins consist of three main domains (N-terminal, middle, and C-terminal). The N-terminal and
234 middle domains are similar (RSDM of 1.6 Å for N-terminal domain and 1.2 Å for middle domain)
235 between *N. parisii* member NEPG_02057 and *P. epiphaga* member PAEPH01_0103 (Fig 5D and
236 E). However, the C-terminal domains form a bundle of alpha helices that are less conserved
237 (RSDM of 18.3 Å) (Fig 5F). We also compared an *N. ausubeli* NemLGF1 member NERG_01890
238 to *N. parisii* NEPG_02057, and observed similar trends, although the C-terminal domains were
239 more conserved (RSDM 4.5 Å) than with *P. epiphaga* PAEPH01_0103 (13A-E Fig). These high
240 overall structural similarities between NemLGF1 proteins exist despite the sequence identities
241 being only 22.03% between NEPG_02057 and PAEPH01_0103 and 27.37% between
242 NEPG_02057 and NERG_01890 [39].

243

244 To begin to understand the biochemical function of these nematode infecting large gene families,
245 we generated structural models for the four other largest families using Alphafold (S13F-I Fig).
246 None of these families have apparent structural similarity to other solved proteins, with the
247 exception of NemLGF2, which shares similarity to Leucine Rich Repeat proteins (S13I Fig), which
248 is a domain that has been previously observed in microsporidia proteins [21,40]. This family was
249 previously described in *N. displodere* and we also observe large expansions of this family in *N.*
250 *ferruginous* and *N. homosporous,* but not in *N. cider*. As previously shown, although we detect a
251 member of this family in the *Parisii* group species, we only observe expansions of this family within
252 some *Displodere* group species.

253

254 **Evolution of *Nematocida* phenotypic traits**

255 Microsporidia often only infect one or more closely related host species, but generalists that infect
256 many hosts have also been observed [16]. Previous work in *Nematocida* have described species
257 with more restricted host specificity, such as *N. parisii* and *N. ausubeli,* as well as those with a
258 broader host range, like *N. displodere* and *N. homosporus* [19,41]. To characterize the host range
259 of the novel species we isolated, we set up experimental infections against seven different
260 nematode hosts. As controls, we also challenged each of these hosts with *N. parisii* and *N.*
261 *ausubeli*. After 4 days of infection, we fixed and stained populations of animals using direct yellow

262 96 [22,42] and quantified the proportion of the population that produced microsporidia spores (Fig
263 6A). For *N. parisii* and *N. ausubeli*, we observed strong infection of *C. elegans* and *C. briggsae*,
264 as previously reported [19]. *N. ausubeli* also infected a modest percentage of *C. tropicalis* and a
265 small percentage of *C. nigoni*, while *N. parisii* infected a small percentage of *C. tropicalis*. These
266 two species could not infect any of the other hosts. These results are consistent with what has
267 been previously reported [19]. For *N. botruosus*, we only observed infection of its native host
268 (*Panagrellus* sp. 2*)*, and none of the other six host species tested. In contrast, *N. cider* could infect
269 all the hosts tested to some extent (Fig 6B), with *C. elegans* being the most infected at 85.6% and
270 *Panagrellus* sp. 2 being the least infected at 7.4%. For *N. ferruginous*, we were unable to observe
271 more than 50% of animals infected for any nematode species, even when using a high
272 concentration of spores. However, even at this level of infection, we observed infection of *C.*
273 *elegans, C. briggsae,* and *C. tropicalis.* Additionally, we saw infection in *C. remanei* and *Oscheius*
274 *tipulae*, which were species that none of the other microsporidia besides *N. cider* could infect (S14
275 Fig). Our data suggests that *N. botruosus* has a specialized host range, whereas *N. cider* and *N.*
276 *ferruginous* appear to be generalists.
277
278 Microsporidia mostly infect only one tissue, but some microsporidia, including *N. displodere*, have
279 been shown to infect multiple tissues [17]. During our infection experiments, we only ever
280 observed spore formation in the intestine for *N. botruosus,* but observed infection in other tissues
281 for *N. cider* and *N. ferruginous* (S15A-C Fig). Additionally, during prolonged infection with *N. cider*
282 or *N. ferruginous*, the entire animal turns brown, suggesting that infection is occurring throughout
283 the animal (S15D-E Fig). To further characterize which tissues *N. cider* and *N. ferruginous* infect,
284 we used *C. elegans* strains with GFP specifically expressed in either the intestine, epidermis, or
285 body wall muscle (Fig 6C). We observed *N. ferruginous* could infect both the muscle and the
286 epidermis but we did not see productive infection in the intestine. In contrast, *N. cider* infection
287 was seen in all three tissues, the muscle, epidermis, and intestine, although at a lower frequency
288 in the intestine (Fig 6D).
289
290 We recently showed that the *C. elegans* protein AAIM-1 is necessary for efficient invasion by *N.*
291 *parisii* [23]. This protein is secreted and likely functions in the intestinal lumen. Although *N.*
292 *displodere* infects other tissues besides the intestine, feeding is necessary for infection, implying
293 that spores need to first be in the intestinal lumen to infect tissues beyond the intestinal cells. To
294 determine if infection by an epidermal and muscle infecting species was also dependant upon
295 this protein, we infected wild-type and *aaim-1* mutant animals with *N. cider*. Similar to what was

296 observed with *N. parisii* infection, *aaim-1* mutant animals exposed to a high dose of *N. cider* have
297 a reduced prevalence of infection compared to wild-type animals (S16 Fig).
298
299 Microsporidia species have an extracellular spore form of a defined morphology which can be
300 used to identify species [17]. *Nematocida* spores have been shown to have a variety of sizes, with
301 the average size of species ranging from 1.3-2.38 μm long and 0.53-1.03 μm wide. Additionally,
302 some species in the *Parisii* group have been shown to have a second, larger, class of spores [19].
303 To determine spore sizes for the new species we identified, we examined infected animals with
304 stained spores (S15A-C Fig). This revealed that these species only have a single class of spores,
305 with *N. botruosus* measuring 1.51 x 0.76 μm, *N. cider* measuring 2.4 x 0.67 μm, and *N. ferruginous*
306 measuring 1.72 x 0.6 μm.
307
308 Microsporidia has a characteristic infection apparatus known as the polar tube [43]. The length of
309 this tube was previously shown to correlate with tissues infected [17]. For example a species with
310 a shorter polar tube(~ 4 μm for *N. parisii*) infects the intestine, and a species with a longer polar
311 tube (~ 12 μm for *N. displodere*) can infect the muscles, neurons, and epidermis [18]. This
312 correlation is hypothesized to be due to the requirement of longer polar tubes to access more
313 distal tissues from the lumen of the intestine [17,18]. We measured polar tube lengths of several
314 *Nematocida* species using a previously described freeze-thaw approach to trigger spore firing
315 [18] followed by staining with Nile red [44] (Fig 6E). We observed a range of lengths throughout
316 *Nematocida*, with the shortest being *N. botruosus* with a length of 2.3 μm and the longest being
317 the previously measured *N. displodere* [18] (Fig 6F).
318
319 To understand how various microsporidian phenotypic traits evolve, we plotted our measured
320 attributes onto the phylogenetic tree of *Nematocida* (Fig 7). We observe that all of the *Parisii* group
321 species have a specialized host range and show specificity for the intestine. In contrast, all of the
322 species in the *Displodere* group, except for *N. botruosus*, have a more generalist host range,
323 suggesting that this property was either lost in *N. botruosus*, or gained after species divergence.
324 Additionally, the *Displodere* group species, *N. displodere*, *N. cider*, and *N. ferruginous*, infected
325 other tissues besides the intestine, suggesting that the ability to infect multiple tissues was
326 possibly present after the divergence of *N. botruosus*, and then lost in *N. homosporus*.
327
328 We observed that spore size is variable throughout *Nematocida* species, with both *Nematocida*
329 groups containing species (*N. minor* and *N. botruosus*) exhibiting some of the smallest spores.

330 The extent of spore size variation between species exceeds the less than 10% variation that has
331 been reported between strains of the same species [19]. Additionally, we observed that two
332 classes of spores are present only in the *Parisii* group and large spore clusters were only observed
333 in the *Displodere* group. Finally, we observe variability of polar tube length ranging from 2.3 - 12.6
334 µm. Our data suggests that the last common ancestor of *Nematocida* likely had a relatively short
335 polar tube (less than ~ 4 µm). An increase in polar tube length then occurred in the *Displodere*
336 group after the divergence of *N. botruosus*. This increase in polar tube length correlates with
337 tissue specificity, with the three species with the longest polar tubes all infecting tissues besides
338 the intestine. These results provide further support for microsporidian tube length being a factor
339 in determining which tissues are infected and provide a resource to further understand molecular
340 evolution of polar tube length once polar tube proteins in *Nematocida* are identified [17,28].
341 Together, our results support evolutionary change in both morphological features and host
342 infection in *Nematocida*, notably with loss and gain of tissue specificity.
343
344 **Conclusions**
345 To understand genomic evolution of nematode-infecting microsporidia, we sequenced the
346 genomes of nine additional species. We then analyzed these genomes along with four previously
347 sequenced species. We show that core cellular processes, including metabolism, are largely
348 conserved (as determined by presence of proteins) between genera. Our analysis demonstrates
349 that *Enteropsectra* and *Pancytospora* form a sister group to the *Enterocytozoon*, but have not
350 undergone decay of the glycolytic pathway that has occurred in *Enterocytozoon* species [15]. This
351 conservation of metabolism suggests that host metabolic requirements are similar between
352 species and is in contrast to proteins predicted to interface with hosts, which are quite different
353 between different groups of microsporidia [21]. Although we present the largest comparison of
354 protein function between microsporidia species so far, detecting protein function from sequence
355 alone in microsporidia is challenging. Additional analyses taking advantage of advances in
356 structural prediction are likely to detect highly diverged proteins that would be missed based on
357 sequence similarity alone [10,45,46].
358
359 One class of proteins that contributes to diversity between microsporidian species are the large
360 gene families. Here we show that one of these large gene families, NemLGF1, is present in both
361 *Nematocida* and *Pancytospora* species. There are three possibilities that could explain this
362 observation. One, the protein family in the two groups of species could have evolved *de novo*
363 independently of one another. This type of convergent evolution is rare and given the high

364 predicted structural similarities that we observe, this scenario is unlikely [47]. Two, this protein
365 family could have existed in the last common ancestor of *Nematocida* and other microsporidia,
366 and then was lost in all lineages except for *Pancytospora* and part of *Nematocida*. From the
367 genomes we analyzed, this would mean that this family would have been lost 7-9 times
368 (*Nematocida Displodere* group, *N. minor*, *V. corneae*, *E. breve*, Enterocytozoon, Nosematida,
369 Glugeida, Neopereziida, and Amblyosporida) [33]. Although this explanation is also unlikely,
370 additional sequencing of microsporidia genomes could provide more support for or against this
371 scenario. Three, this family was horizontally transferred between a species in each genus, though
372 which genus was the recipient, and which was the donor, is not clear. Horizontal gene transfer is
373 increasingly being appreciated as having a role in the evolution of eukaryotic microbes and we
374 believe this to be the most likely explanation [48]. We also identified a family (NemLGF26) present
375 in *N. ferruginous* which shares similarity to the previously described InterB family found in
376 *Encephalitozoon* species [36]. Although to our knowledge, horizontal transfer of genes between
377 microsporidia species has not been previously reported, co-infections of multiple species are often
378 observed, suggesting that these types of events may be more common that previously
379 appreciated [49].
380
381 Microsporidia have been shown to display much variability in their phenotypic traits [17]. However,
382 determining how traits vary between related species has been challenging as most phylogenetic
383 trees are built from 18S rRNA sequences, resulting in phylogenies that are often of low accuracy
384 [3]. Here, we use genome assemblies to build a high-quality phylogeny for the *Nematocida*. By
385 mapping several phenotypic properties onto this tree, we observe examples of loss and gain of
386 infection properties. As more microsporidian genomes are sequenced, this approach is likely to
387 be useful for further understanding the evolution of microsporidia infection properties and spore
388 characteristics. With our data we show that a generalist host range emerged in the *Displodere*
389 group after the divergence of *N. botruosus*. One potential caveat to our measurements of host
390 range is that they are only done with a single strain for each microsporidia and host species [16].
391 However, we see similar results to what was reported with other *N. parisii* and *N. ausubeli* strains,
392 and nematode-infecting microsporidia have been observed to mostly infect different host strains
393 at similar levels [19]. Previously, the size of generalist microsporidia genomes were suggested to
394 be smaller than genomes of species that had a more narrow host range [14]. Although the
395 generalist *Nematocida* have some of the smallest genome sizes, they are about the same size
396 as *N. botruosus*, which has a narrow host range. This suggests that other forces can be involved
397 in sculpting microsporidia genomes.

398

399 Although *C. elegans* has become an important model in which to study many aspects of
400 microsporidia infection, there are currently no tools for genetically perturbating microsporidia that
401 infect this host [50]. Knockdown of genes using RNAi has been demonstrated in microsporidia
402 that infect honey bees, silkworms, fish, and locusts and is becoming a powerful method to
403 investigate the function of microsporidia proteins, including those that are secreted or surface
404 exposed [51–56]. So far, this approach has not been amenable to species that infect *C. elegans*,
405 as *N. parisii* and the other *Nematocida* don't encode the machinery necessary to carry out RNAi
406 [3]. Here we report that *P. epiphaga,* which infects *C. elegans* [19,26], encodes this machinery,
407 suggesting RNAi may be possible in this species. *C. elegans* may be particularly useful for
408 microsporidian RNAi knockdown, as RNAi of genes in *C. elegans* is highly efficient and can be
409 carried out by feeding bacteria expressing double stranded RNA against the target transcript,
410 resulting in dozens of successful whole genome screens [57]. Additionally, *P. epiphaga* is more
411 closely related to several human pathogens than *N. parisii* and our genomic data will provide a
412 useful resource for further investigations using this species.

413

414 **Material and methods**

415

416 **Isolation of infected animals**

417 Sampling of nematodes was done similar to as described previously [18]. Briefly, samples of
418 rotting fruit or vegetation were placed on 6-cm NGM plates seeded with 10x saturated cultures of
419 OP50-1 for 16-48 hours. Rotting matter was then removed. *N. ferruginous* was identified using
420 differential interference contrast microscopy to identify animals displaying meronts and/or spores.
421 *N. cider* and *N. botruosus* were identified by growing animals at 21-23°C, washing animals off
422 plates in M9/0.1% Tween-20, and fixing animals in 800 µl acetone. Samples were then stained
423 with Direct yellow 96 (DY96) staining solution (M9, 0.1% SDS, and 20 µg/ml DY96) for 30 min.
424 Samples were resuspended in EverBrite Mounting Medium (Biotium #23002) with DAPI and
425 placed onto microscope slides. Microsporidia-containing animals were identified as those
426 containing DY96 stained spores. Nematode host species were determined using sequencing
427 primers as described previously [58].

428

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

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

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

446 Spores were prepared similar to as previously described [22]. Briefly, contaminate-free, infected
447 populations of worms were chunked onto 12-48 10-cm OP50-1 seeded NGM plates and grown
448 at 21°C for 5-10 days. After worms had consumed all the OP50-1 bacteria, worms were washed
449 off plates with M9 and frozen at -80°C. Spores were extracted from worms as described above
450 and then filtered through a 5 µm filter (Millipore). The concentrations of spore extracts were
451 determined by staining spores with DY96 and counting spores on a sperm counting slide (Cell-
452 VU) using a 20X objective on a Axio Imager.M2 (Zeiss). DNA was extracted using a MasterPure
453 yeast DNA purification kit (Lucigen) according to manufacturer instructions using 10-50 million
454 spores per microsporidia species. DNA libraries were prepared by The Centre for Applied
455 Genomics (TCAG) and each species was sequenced using 2 lanes of a NovaSeq 6000 SP flow
456 cell to produce 250 bp paired-end reads.
457

458 **Genome assembly & annotation**

459 Adapter sequences ('AGATCGGAAGAG') were removed from paired-end raw reads using
460 cutadapt v21 [59]. Leading and trailing low quality (quality < 3) or N bases were removed, low
461 quality bases (quality < 30, except for *P. epiphaga* where 36 was used) were cut using a 4-base
462 sliding window and reads less than 36 bases were dropped using Trimmomatic v0.36 [60]. Reads

463 were mapped to the corresponding host genome (except for *N. botruosus* which was mapped to
464 *C. elegans)* as well as the *Escherichia coli* genome using bowtie2 v2.3.4.1 [61]. Reads were
465 assembled into contigs and scaffolds using Abyss v2.02 [62] with a Kmer of 128. Only scaffolds
466 greater than 500 bp were retained. Assemblies were then filtered for contaminants (bacteria and
467 host) using BlobTools v1.0.1 [63]. For some samples (S2 Table), subsampling was done prior to
468 removing contaminants with BlobTools. Each assembly was run through Redundans v0.14a [64]
469 with a minimum identity value of 0.85, to remove duplication. Because the laboratory the
470 microsporidia species were cultured in regularly uses ERTm1 for experiments, there exists the
471 possibility of minor contamination of other microsporidia genomes. To correct for this, assemblies
472 were filtered for ERTm1 [5] contamination if they met the following criteria: 99% identity and 80%
473 query coverage. Additional filtering was done for size and coverage: scaffolds less than 750 bp
474 were removed and those between 750 bp – 1 kbp were removed if they had less than 5X median
475 coverage of the largest 5 scaffolds. NCBI nucleotide blast (BLASTn) [65] was used to filter
476 remaining scaffolds that had >98% identity and >50% query cover to ERTm1/ERTm2 assemblies.
477 Manual filtering of select scaffolds not captured in previous filtering steps was done for six of the
478 samples (See S2 Table). Protein coding genes were predicted using Prodigal v2.6.3 [66] using
479 translation table 1. Proteins with less than 100 amino acids were kept only if they had a significant
480 (E=0.001) hit to the Uniref90 protein database or Pfam hit using the hmmscan function in HMMER
481 package v3.1b2 [67] and E-value cutoff of 0.001. The final list of proteins were annotated with
482 Pfam [68] domains using hmmscan, signal peptides using SignalP v4.1f [69], and transmembrane
483 domains using TMHMM v2.0c [70]. Functional annotation of proteins was also done using
484 BlastKOALA [71] (S3 Table).
485
486 **Phylogeny and genome characterization of microsporidia species**
487 The phylogenetic tree (Fig 1A) was generated using 45 species. Species whose proteome was
488 downloaded from NCBI are listed in S5 Table. The remaining proteomes were predicted from the
489 genomes assembled in this study. In addition, proteins were predicted from two assemblies
490 available on NCBI for the following species: *Metchnikovella incurvata*, and *Enteropsora canceri*.
491 Orthofinder v2.5.2 [31] was used with -M msa option to obtain a MSA species tree with a minimum
492 of 77.8% of species having single copy genes in an orthogroup. Tree was rooted using *Rozella*
493 *Allomycis* [72] in FigTree v1.4.4. Protein content, genome size, as well as genome completeness
494 statistics were computed for each species using BUSCO v3.1.0 [73] and microsporidia odb9.
495
496 **Pathway conservation in microsporidia species**

497 InterProScan v5.51-85.0 [74] was run on all species with -goterms option. A database was made
498 with POMBE GO-slim terms (https://www.ebi.ac.uk/QuickGO/slimming) and their descendants
499 (found using QuickGO API using the 'find descendants' option,
500 https://www.ebi.ac.uk/QuickGO/api/index.htmllar#!/gene_ontology/). For each species, protein to
501 GO term annotations from InterProScan were mapped to the above-mentioned database while
502 retaining count of proteins mapping to each GO-term. Since multiple GO terms map to a single
503 GO slim, a unique protein count was obtained for each GO-slim. Using *Rozella allomycis* as the
504 base species, a subset of the POMBE GO-slim terms were obtained and the subset was then
505 searched against the remaining species to compare pathway conservation across species.
506
507 **Identifying large gene families**
508 OrthoMCL v2.0.9 [75] was run using all proteomes and with the following configuration settings:
509 percentMatchCutoff=50 and evalueExponentCutoff=-5, suggested minimum protein length and
510 maximum percent stop codons, and MCL inflation value of 1.5. Paralogous orthogroups were
511 extracted (at least 10 members in a single genome). The genome with the highest number of
512 paralogs for a specific gene family was used as the seed sequences to expand gene families.
513 Seed sequences were searched against each microsporidia genome iteratively using the
514 jackhmmer function in HMMER and an E-value cutoff of 10-5. All resulting large gene families
515 were assessed for signal peptides, transmembrane domains, and Pfam domains. A family was
516 classified as having signal peptides if >50% of the proteins in the family had a signal peptide. The
517 same criteria were applied for transmembrane domains. Families that did not contain either signal
518 peptides or transmembrane domains in greater than >50% of proteins were not included in our
519 analysis. All families with 90% or greater protein overlap were collapsed, with the exception of
520 NemLGF2 and NemLGF20 which were collapsed with less than 90% overlap in some cases. We
521 used OrthoMCL to assign membership with families with >5% Pfam representation in either LRR,
522 ZF, or peptidase domains [21]. To determine if these large gene families were present in other
523 microsporidia species, we searched the microsporidia genomes in Figure 1 with family-specific
524 HMM models using HMMER. Protein sequences that had at least an E-value of 10-5 were
525 considered as belonging to that family.
526
527 **Structural modeling of large gene families**
528 Models of members of large gene families were generated using AlphaFold as implemented in
529 ColabFold [37,38]
530 (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb).

531 Default settings were used for NEPG_02057, and for the other proteins PSI-BLAST [65] was used
532 to identify members and multiple sequence alignments were generated using HHblits
533 (https://toolkit.tuebingen.mpg.de/tools/hhblits) [76]. All the models analyzed have overall pLDDT
534 scores over 85. Alignments of NemLGF1 domains were performed using PyMOL v2.5.3
535 (https://pymol.org). Structural similarity to solved structures was determined using Dali
536 (http://ekhidna2.biocenter.helsinki.fi/dali/) [77]. The percent identity between proteins was
537 determined using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) [78].
538
539 **Host specificity assays**
540 The following nematode species (strains) were used for infection assays: *C. elegans* (N2),
541 *Caenorhabditis briggsae* (JU2507), *Caenorhabditis nigoni* (JU1422), *Caenorhabditis remanei*
542 (JU2796), *Caenorhabditis tropicalis* (JU1373), *Oscheius tipulae* (JU1505), and *Panagrellus* sp. 2
543 (AWR79). Mixed populations of nematodes were maintained on 10-cm seeded NGM plates for at
544 least three generations without starvation at 21°C. Animals were then washed off plates and
545 embryos extracted with bleach as described above, except for *Panagrellus* sp. 2. Synchronized
546 animals of *Panagrellus* sp. 2 were prepared by washing animals with M9 three times and then
547 filtering through a 20 µm nylon filter (Millipore SCNY00020). This filtering was then repeated a
548 second time. 400 L1 stage animals of each species were infected with the following microsporidia
549 species (strain [spore amount]): *N. parisii* (JUm1248[3 million]), *N. ausubeli* (ERTm6[0.8 million]),
550 *N. cider* (AWRm77[4 million]), *N. ferruginous* (LUAm3[20 or 40 million]), and *N. botruosus*
551 (AWRm80[28 million]). These doses were chosen so that average percentage infection was
552 between 80-99% in either *C. elegans* or the host species the microsporidia was found in, with the
553 exception of *N. ferruginous* where this level of infection was not possible to achieve. Animals and
554 spores were mixed with 400 µl 10X OP50-1 and placed onto 6-cm NGM plates. Plates were dried
555 in a clean cabinet and incubated at 21°C for 96 hours. Animals were then washed off plates, fixed,
556 stained, and placed onto slides as described above. The percentage of animals infected was
557 determined by counting the number of P0 animals that contained newly formed clusters of
558 microsporidia spores.
559
560 **Microsporidia spore size measurements**
561 Infected, fixed, and stained animals were prepared as described above. Images of these infected
562 animals were taken using an 63X objective on an apotome-equipped Axio Imager.M2 (Zeiss).
563 Images were taken as z-stack maximum intensity projections. These images were analyzed using
564 the straight line tool in ImageJ 1.52 [80] to measure the length and width of each spore. Only

565 spores that were not immediately adjacent to other spores were measured. At least 40 spores
566 were measured for each species from a minimum of two infected worms. Presence of large spore
567 clusters was based on visual inspection of either DY96 stained spores in S15 Fig or from
568 previously published images [19].
569
570 **Infection of *aaim-1* mutants**
571 N2 and *aaim-1* mutants (kea22 and kea28) [23] were infected with *N. cider* (4 million spores) as
572 described above. To quantify pathogen burden within animals (GFP), regions of interest were
573 used to outline individual worms followed by subjection to the "threshold" followed by "measure"
574 tools in FIJI [79].
575
576 **Tissue specificity determination**
577 Strains of *C. elegans* each expressing GFP specifically in the intestine (ERT413), epidermis
578 (ERT446), or muscle (HC46) were grown on standard NGM plates seeded with OP50-1 to the
579 gravid adult stage. Gravid adults were then harvested in M9 and transferred to 15 mL conical
580 tubes, washed three times to reduce bacterial excess, and pelleted by centrifuging at 3000 x g.
581 The pellets were resuspended in 1 mL of M9 buffer and added with 1 mL of bleach solution (1:4
582 volume ratio of 5M NaOH: 6% NaClO; final concentration of 0.5M NaOH, 2.4% NaClO), incubated
583 for 2 min at room temperature, and washed three times with 13 mL of M9. After the final wash,
584 the solutions containing embryos were incubated overnight on a rotator at room temperature to
585 synchronize to the L1 stage.

586 ~1000 L1 stage animals of each strain were plated onto a 6-cm NGM plate with either 1 million
587 *N. ferruginous* or *N. cider* spores. Infection plates were allowed to dry and kept at 20°C. At 3 days
588 post infection, animals were harvested in M9 buffer and fixed with 4% paraformaldehyde in PBST
589 (1x PBS, 0.1% Tween-20) at room temperature for 40 min on a rotator for fluorescent in situ
590 hybridization. Fixed animals were then washed four times, each with 1 mL of PBST, then
591 incubated overnight with a mixture of CAL Fluor Red 610 (CF610)-tagged microsporidian probes,
592 including MicroA-CF610 (CTCTGTCCATCCTCGGCAA), MicroB-CF610
593 (CTCTCGGCACTCCTTCCTG), MicroD-CF610 (CGAAGGTTTCCTCGGATGTC), and MicroF-
594 CF610 (AGACAAATCAGTCCACGAATT), each at a final concentration of 2.5 ng/mL in
595 hybridization buffer (900 mM NaCl, 20 mM Tris pH 7.5, 0.01% SDS) on a thermal shaker at 46°C,
596 1000 rpm. After hybridization, animals were washed with FISH wash buffer (hybridization buffer,
597 5mM EDTA) for three times, each for 30 min, on a thermal shaker at 48°C, 1000 rpm. After
598 washing, animals were rinsed once with PBST, mounted, and imaged with confocal microscope

599 Leica Stellaris 5. Each sample was scanned randomly for 30 infected worms and images were
600 taken in different planes to examine all the infection areas in a single worm. For each strain,
601 infected animals were binned into “yes” or “no” respective to the fluorescent tissue. An infected
602 animal was considered to have a specific tissue type infected if there was at least one infection
603 area colocalized with the tissue marker.

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

605 ~1000 N2 animals at L1 stage on a 10-cm NGM plate were infected with 3 million spores of
606 either *N. ferruginous* or *N. cider*. As the animals grew, plates were monitored to add more food
607 (1 mL of 10x OP50-1 concentrated from a culture of OD_{600} of 0.5-0.7) to prolong the infection in
608 the population. After 7-10 days post infection, some animals appeared brown compared to other
609 animals on the same plate and uninfected control under standard dissection microscope.
610 Control animals and brown, infected animals were picked and mounted onto a 2% agarose pad
611 and imaged with Eclipse Ni microscope (Nikon). All images were taken at the same light
612 intensity and white balanced.

613 **Measurements of polar tube lengths**

614 Polar tube lengths were measured for *N. parisii* (ERTm1), *N. ausubeli* (ERTm2), N. minor
615 (JUm1510), N. major (JUm2507), *N. homosporus (*JUm1504*)*, *N. cider* (AWRm77), *N. ferruginous*
616 (LUAm3), and *N. botruosus* (AWRm80). To encourage polar tube extrusion, microsporidia spores
617 in microcentrifuge tubes were exposed to two freeze thaw cycles, after the initial thaw from -80°C.
618 This was done by freezing at -80°C for 10 minutes followed by thawing at room temperature for
619 10 minutes. Nile red (Sigma-Aldrich 72485) (1mg/ml in acetone) was prepared, sequentially
620 diluted to 100 µg/ml in M9, and added to a final concentration of 10 µg/ml in freeze-thaw treated
621 spores. Spores were incubated with Nile red for 30 minutes in the dark followed by 2 µl of
622 Calcofluor White (Sigma-Aldrich 18909). Polar tubes were imaged at 63x magnification using an
623 Axio Imager.M2 (Zeiss). Polar tube measurements were performed on FIJI [79], using the
624 freehand line tool to trace polar tubes followed by selecting the Analyze → Measure option.

625 **Data availability**

626 All samples were deposited under NCBI BioProject PRJNA579982 and genome accession and
627 Sequence Read Archive numbers are listed in S1 Table. Data from all the infection experiments
628 is in S1 Data.

629

630 **Taxonomic Summaries.**

Phylum: Microsporidia Balbiani 1882

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

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

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

The type strain LUAm1 was found in 2017 inside of its type host nematode *C. elegans* strain LUA1, which was isolated from a rotting *Heracleum* sp. stem in Santeuil, France (GPS coordinates 49° 7' 16.99" N, 1° 57' 3.918" E). The type material is deposited in the collection of the corresponding author (AWR). The genome of this strain has been sequenced and deposited in Genbank under accession JALPMW000000000. This is a new species based on the divergence of the genome from other sequenced species. This species has been observed to infect the muscle and epidermal tissues of *C. elegans*. Experimental infection has been observed in *C. elegans* (N2), *Caenorhabditis briggsae* (JU2507), *Caenorhabditis remanei* (JU2796), *Caenorhabditis tropicalis* (JU1373), *Oscheius tipulae* (JU1505). Transmission occurs horizontally and is likely via a fecal-oral route. The spores are ovoid and measure as 1.72 x 0.6 µm. The fired polar tube measures 6.2 µm. This species is named after the rust-like brown color the animals display after being infected for several days.

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

The type strain AWRm80 was found in 2018 inside of its type host *Panagrellus* sp. 2 strain AWR80, which was isolated from rotting apples in Georgina, Ontario, Canada. The type material is deposited in the collection of the corresponding author (AWR). The genome of this strain has been sequenced and deposited in Genbank under accession JALPMX000000000. This is a new species based on the divergence of the genome from other sequenced species. 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.), an Alfred P. Sloan Research Fellowship FG2019-12040 (to A.R.), and an NSF IOS CAREER grant 2143718 (to R.L). T.T was supported by SDSU Rees-Stealy Fellowship and H.T.E.J was supported by a 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

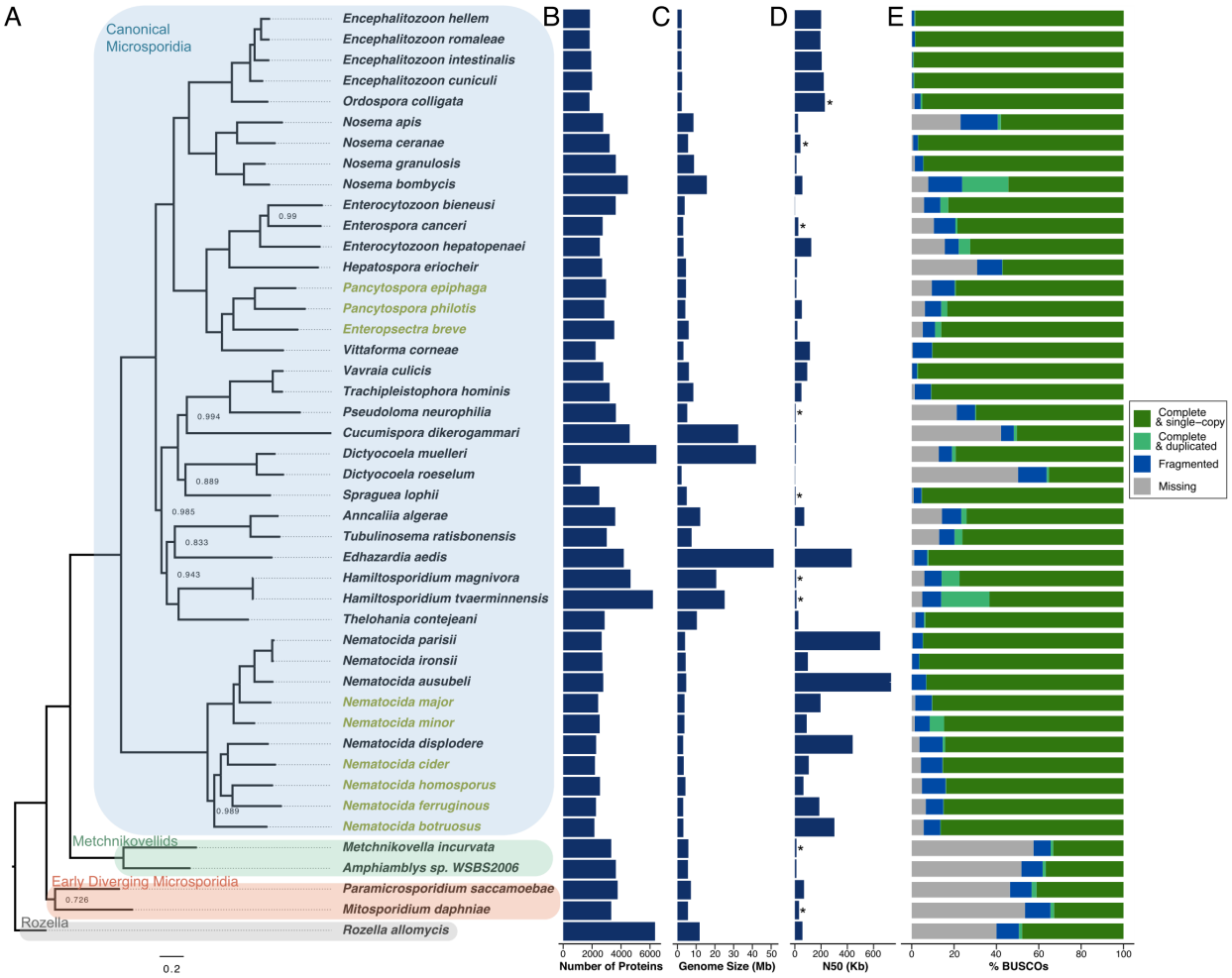


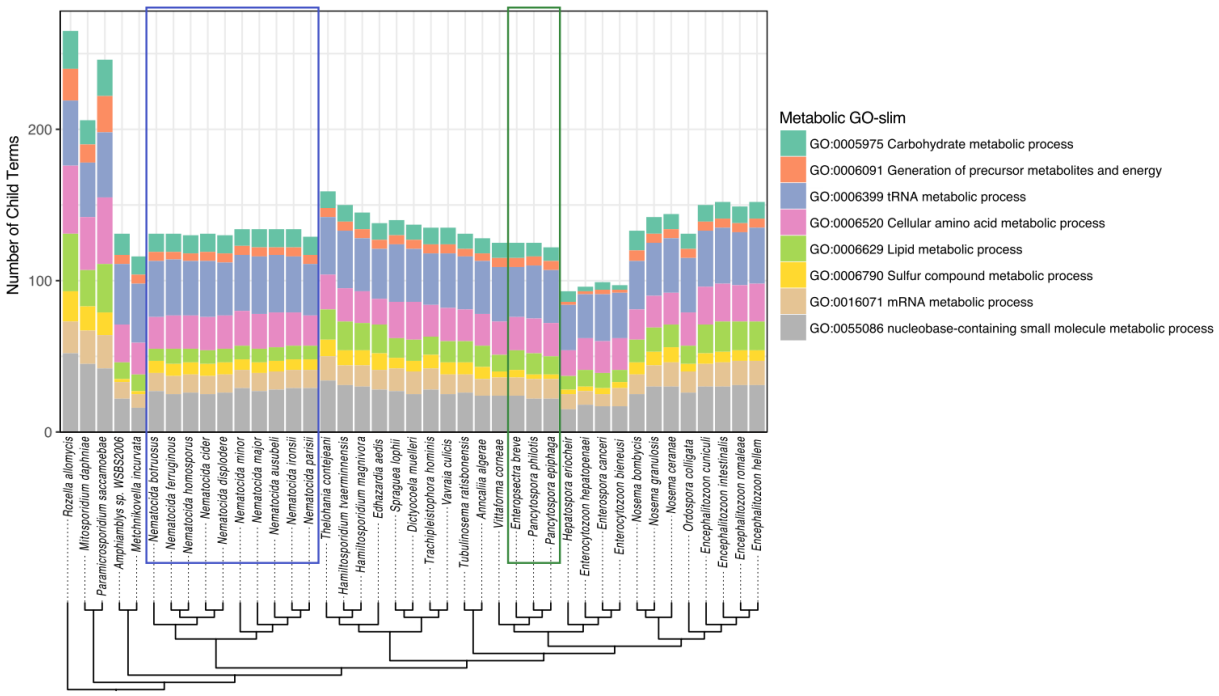
Fig 1. Phylogeny and properties of microsporidia genomes. (A) Phylogenetic tree of 44 microsporidia species and the outgroup *Rozella allomycis* was constructed from orthologous proteins using Orthofinder [31]. Bootstrap values less than 1.0 are indicated at each node. Scale indicates changes per site. 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

712 value is shown, except for those indicated by an * which are the contig N50 values. **(E)** Presence
713 of conserved microsporidian orthologs measured as the percent BUSCOs present in each
714 genome.
715

716

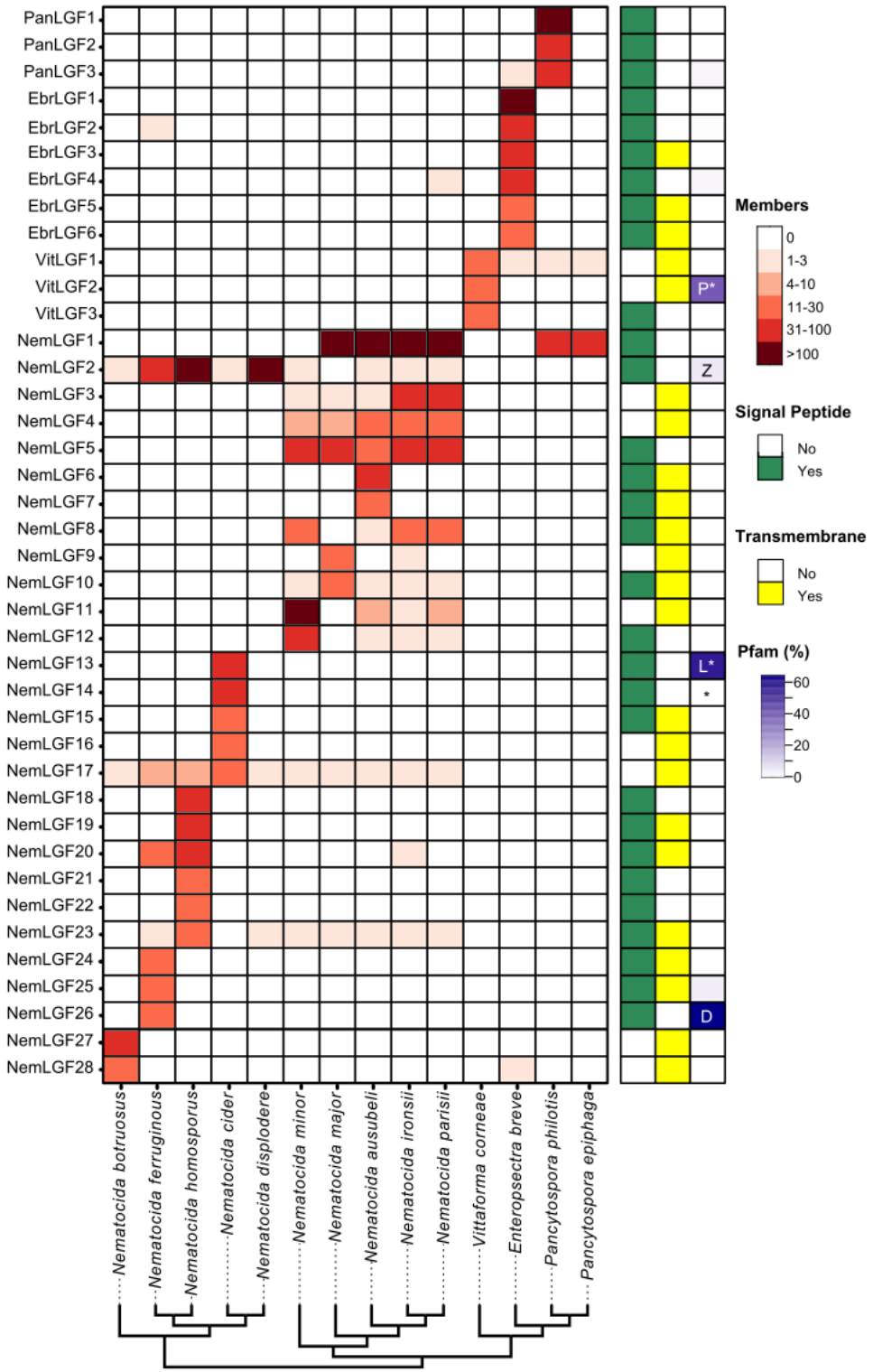
717 **Fig 2. Conservation of protein function across microsporidian genomes.** Membership of
718 proteins from *R. allomycis* and 40 microsporidia species in Pombe GO-slim categories was
719 determined. The number of proteins from each species determined to belong to each GO slim

720 category is shown as a heatmap with GO-slim categories in rows and species in columns. Only
721 GO slim categories that contain at least one protein from any of these species are shown. Legend
722 for the number of proteins in each cell is shown at the right. Phylogenetic tree, shown at bottom,
723 was constructed using Orthofinder [31]. Several species (*Pseudoloma neurophilia, Dictyocoela*
724 *roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were excluded due to poorer quality
725 genome assemblies (See Fig 1). *Nematocida* species are highlighted with a blue box.
726 *Enteropsectra* and *Pancytospora* species are highlighted with a green box.
727
728



729

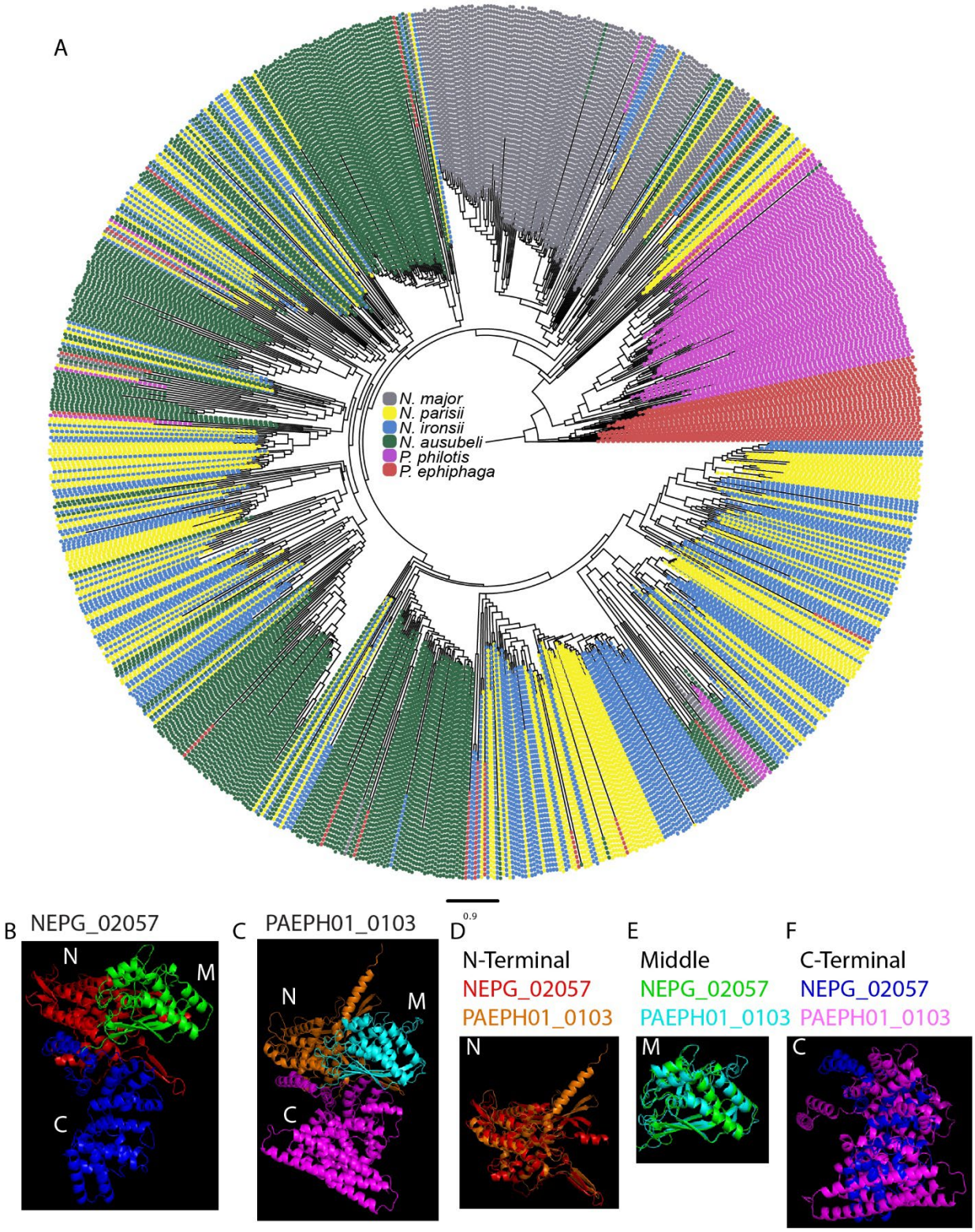
730 **Fig 3. Presence of metabolic enzymes in microsporidia species.** Membership of proteins from
731 *R. allomycis* and 40 microsporidia species in eight GO-slim metabolic categories was determined.
732 The number of proteins from each species determined to belong to each category is shown as a
733 stacked bar graph, with the legend for each metabolic Go-slim shown at the right. Phylogenetic
734 tree, shown at bottom, was constructed using Orthofinder. Several species (*Pseudoloma*
735 *neurophilia, Dictyocoela roeselum*, *Cucumispora dikerogammari*, and *Nosema apis*) were
736 excluded due to poorer quality genome assemblies (See Fig 1). *Nematocida* species are
737 highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green
738 box.



739

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

741 Large gene families and the number of proteins from each species that belongs to each family
742 was determined (See methods). The number of members of each large gene family (listed in
743 rows) in each species (listed in columns) is shown as a heatmap according to the scale at the top
744 right. The presence of predicted signal peptides (green), transmembrane domains (yellow), or
745 Pfam domains (purple gradient) are shown to the right. Families containing the Pfam domains
746 peptidase (P), Zinc finger (Z), LRR(L) or DUF1609(D) are indicated. Domain types are displayed
747 only for those families with more than one annotated Pfam domain. Families that were determined
748 using OrthoMCL are indicated with * (See methods). Species are arranged according to the
749 phylogenetic tree shown in Figure 1. *V. corneae* is included as an outgroup species of
750 *Enteropsectra* and *Pancytospora*
751

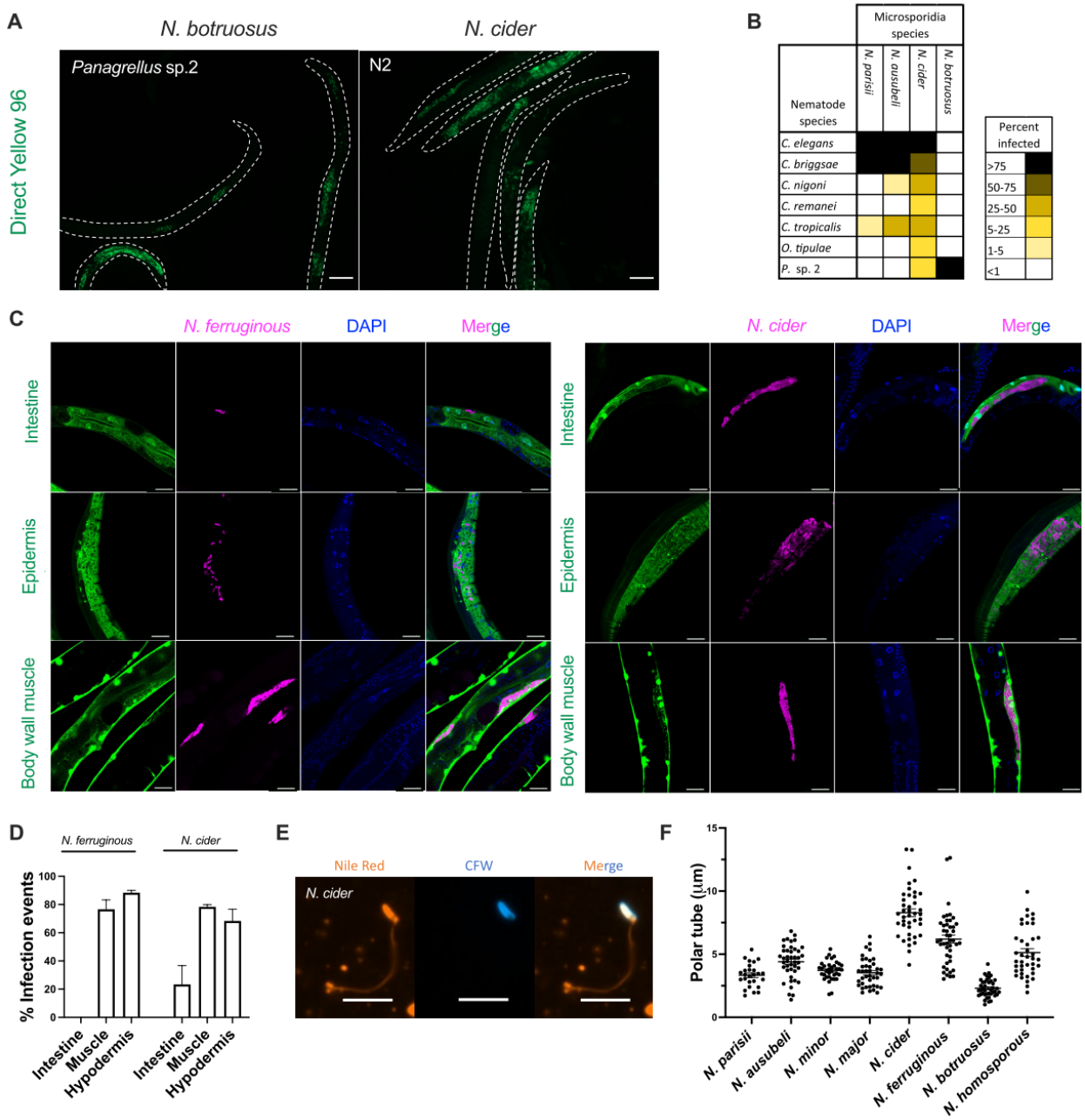

A
N. major
N. parisii
N. ironsii
N. ausubeli
P. philotis
P. ephiphaga
0.9
B NEPG_02057
N
M
C
C PAEPH01_0103
N
M
C
D
N-Terminal
NEPG_02057
PAEPH01_0103
N
E
Middle
NEPG_02057
PAEPH01_0103
M
F
C-Terminal
NEPG_02057
PAEPH01_0103
C

752

753

754 **Fig 5. The large gene family NemLGF1 has likely been horizontally transferred between**
755 **genera.** (A) Phylogenetic tree of NemLGF1 family members, colored by species according to the
756 legend in the middle of the tree. Scale indicates changes per site. (B-C) AlphaFold models of *N.*
757 parisii NEPG_02057 (B) and *P. epiphaga* PAEPH01_0103 (C). (D-F) Aligned structures of the N-
758 terminal (D), middle (E), and C-terminal (F) domains. N, N-terminal. M, middle. C, C-terminal.
759
760
761
762

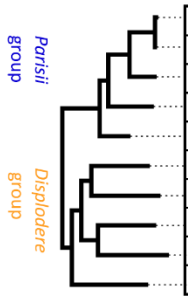
763



764

765 **Fig 6. Determination of host range, tissue specificity, and polar tube length of *Nematocida***
766 **species. (A-B)** Seven species of L1 stage nematodes were infected with the following species
767 (spore concentration): *N. parisii* (3 million), *N. ausubeli* (0.8 million), *N. cider* (4 million), and *N.*
768 *botruosus* (28 million). After 96 hours of incubation with spores, animals were fixed and stained
769 with direct yellow 96. **(A)** Representative images of *C. elegans* infected with *N. cider* and
770 *Panagrellus* sp. 2 infected with *N. botruosus*. Scale bars, 100 µm. **(B)** Percentage of each
771 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 the percentage of each population that displayed newly formed microsporidia spores. Legend is displayed on the right. Data shown are the average of two biological replicates with 20-391 animals counted for each sample. **(C-D)** L1 stage *C. elegans* strains expressing GFP in either the intestine, epidermis, or body wall muscle were infected with either 1 million *N. cider* or *N. ferruginous* spores. After 72 hours animals were fixed and stained with a *Nematocida* 18S RNA FISH probe and DAPI. Note that for *N. ferruginous* infection of the GFP-intestinal strain (top), that the infection is seen to lie outside of the GFP indicating a lack of intestinal infection. **(C)** Representative images of infection of each species in each strain. Scale bars, 20 µm. **(D)** The percentage of the population of each strain containing microsporidia infection within the marked tissue. Data are from over 2 biological replicates and a total of 60 animals examined for each condition. Graphs show mean with SD error bars. **(E-F)** Spores of *Nematocida* species were induced to fire through repeated freeze thawing and then stained with Nile red and calcofluor white. **(E)** Representative image of a stained *N. cider* polar tube. Scale bars, 5 µm. **(F)** The length of polar tubes of each species is displayed. 26-41 polar tubes were measured for each species. Mean ± SEM represented by horizontal bars.



	Host range	Tissues infected	Spore size classes	Large spore clusters	Spore length (µm)	Spore width (µm	Estimated spore volume (µm)	Polar tube length (µm)	Genome size (Mb)
N. parisii	Specialist	Intestine	2	No	2.18	0.8	0.73	3.3	4.1
N. ironsii	Specialist	Intestine	2	No	2.26	0.68	0.55	N/A	4.4
N.ausubeli	Specialist	Intestine	2	No	2.04	0.53	0.30	4.4	4.7
N. major	Specialist	Intestine	2	No	2.3	0.54	0.35	3.5	3.8
N.minor	Specialist	Intestine	2	No	1.3	0.55	0.21	3.7	3.7
N. displodere	Generalist	Epidermis, Muscle	1	Yes	2.38	1.03	1.32	12.6	3.1
N. cider	Generalist	Intestine, Epidermis, Muscle	1	Yes	2.4	0.67	0.56	8.3	3.4
N. homosporus	Generalist	Intestine	1	Yes	2	0.72	0.54	5.1	4.3
N. ferruginous	Generalist	Epidermis, Muscle	1	Yes	1.72	0.6	0.32	6.2	3.1
N. botruosus	Specialist	Intestine	1	Yes	1.51	0.77	0.46	2.3	3.2

795 **Fig 7. *Nematocida* phenotypic properties.** Table of phenotypic properties for *Nematocida*
796 species, arranged according to the phylogenetic tree (left) determined in Fig 1. Host range, tissues
797 infected, spore size classes, presence of large spore clusters, spore size, and polar tube length
798 was determined in this paper and from several previously published studies [18–20]. Estimated
799 spore volume was calculated for spore length and width measurements as previously described
800 [17]. N/A, Not available.
801
802 **Supplemental material:**
803 **S1 Fig. Similarity of *N. ferruginous* assemblies.**
804 **S2 Fig. Phylogenetic tree of *Nematocida* species.**
805 **S3 Fig. Genome size and number of encoded proteins are correlated in nematode-infecting**
806 **microsporidia.**
807 **S4 Fig. Conservation of carbohydrate metabolic processes across microsporidia**
808 **genomes.**
809 **S5 Fig. Conservation of generation of precursor metabolites and energy pathways across**
810 **microsporidia genomes.**
811 **S6 Fig. Conservation of tRNA metabolic processes across microsporidia genomes.**
812 **S7 Fig. Conservation of cellular amino acid metabolic processes across microsporidia**
813 **genomes.**
814 **S8 Fig. Conservation of lipid metabolic processes across microsporidia genomes.**
815 **S9 Fig. Conservation of sulfur compound metabolic processes across microsporidia**
816 **genomes.**
817 **S10 Fig. Conservation of mRNA metabolic processes across microsporidia genomes.**
818 **S11 Fig. Conservation of nucleobase-containing small molecule metabolic processes**
819 **across microsporidia genomes.**
820 **S12 Fig. Signal peptides and transmembrane domains in Nematode-infecting**
821 **microsporidia genomes.**
822 **S13 Fig. Structural models of nematode-infecting microsporidia large gene families.**
823 **S14 Fig. *N. ferruginous* infects multiple host species.**
824 **S15 Fig. Spore formation observed in the intestine for *N. botruosus* and throughout the**
825 **animal for *N. cider* and *N. ferruginous*.**
826 **S16 Fig. AAIM-1 is necessary for efficient infection by *N. cider.***
827
828 **S1 Table. Summary of microsporidia species and their hosts.**

829 **S2 Table. Methods used to assemble each microsporidia genome.**

830 **S3 Table. Annotation of microsporidia genomes.**

831 **S4 Table. Summary of genome assembly statistics.**

832 **S5 Table. List of genome assemblies used.**

833 **S6 Table. RNAi pathway proteins in *Pancytospora* and *Enteropsectra.***

834 **S7 Table. List of large gene family proteins.**

835 **S8 Table. NemLGF26 proteins present in other microsporidia species.**

836 **S1 Data. All infection experiment data.**

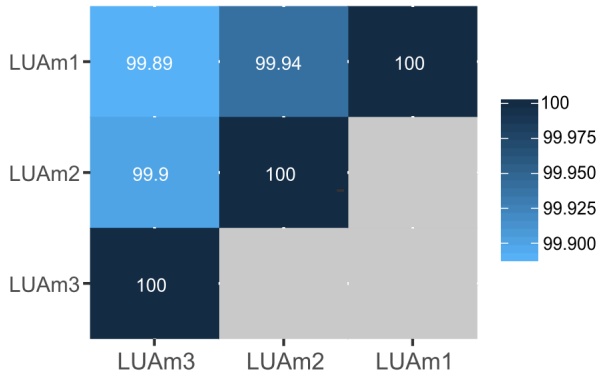
837

838

839

840

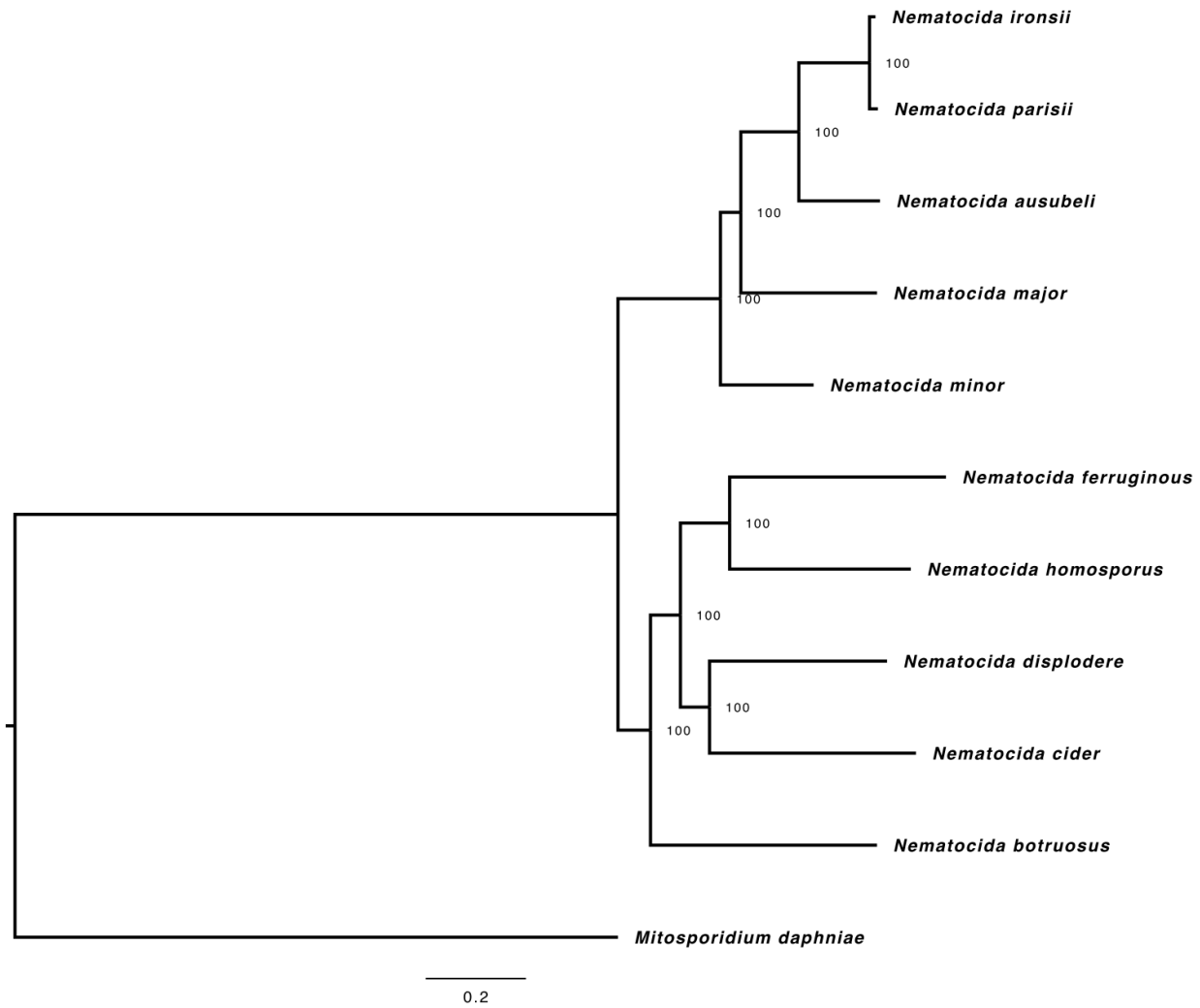
841



842

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

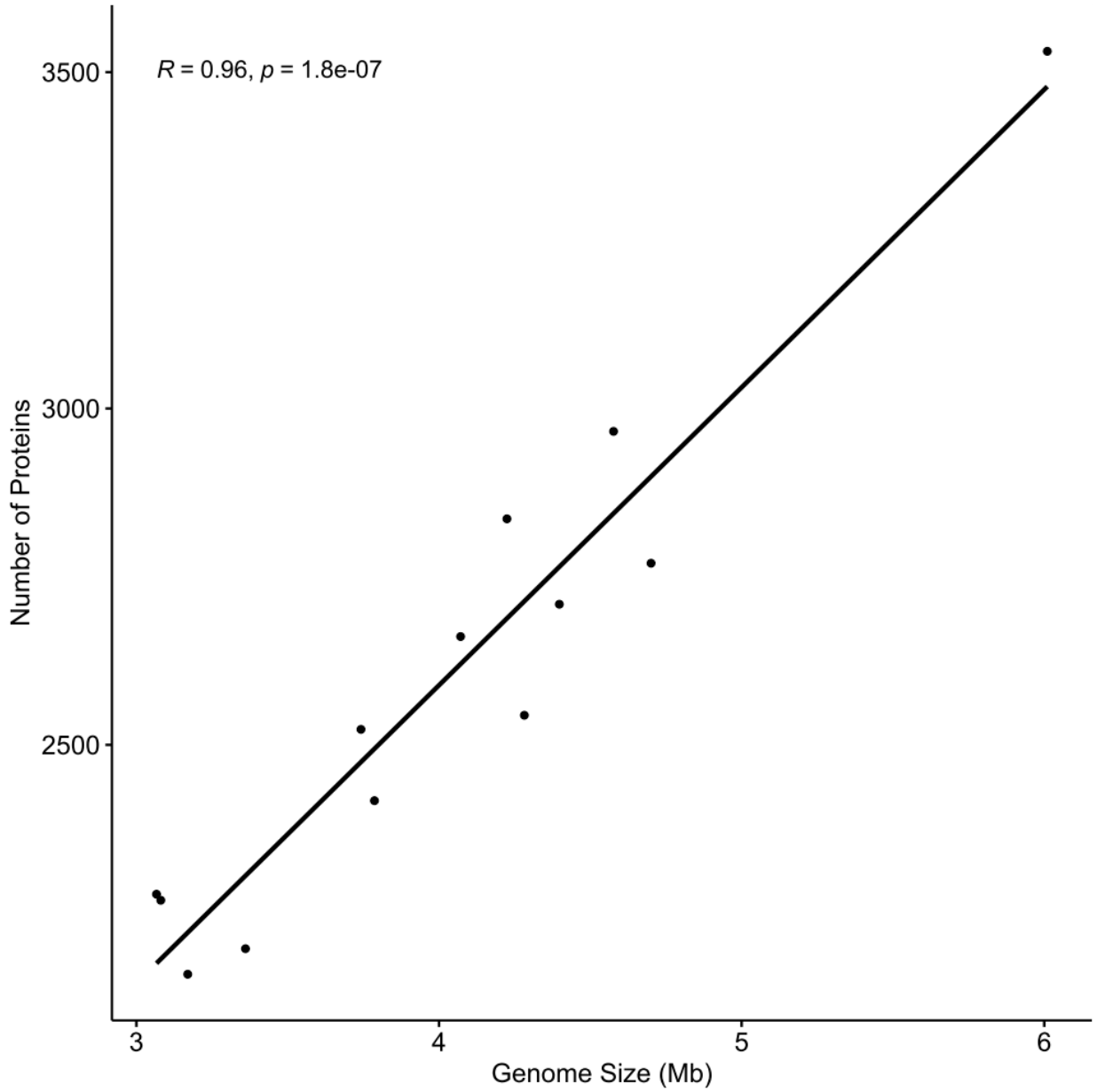
844 The pairwise nucleotide identify of the three *N. ferruginous* assemblies (LUAm1-3) were
845 calculated and displayed as heat map. The similarity between each assembly is colored according
846 to the legend at the right.



847

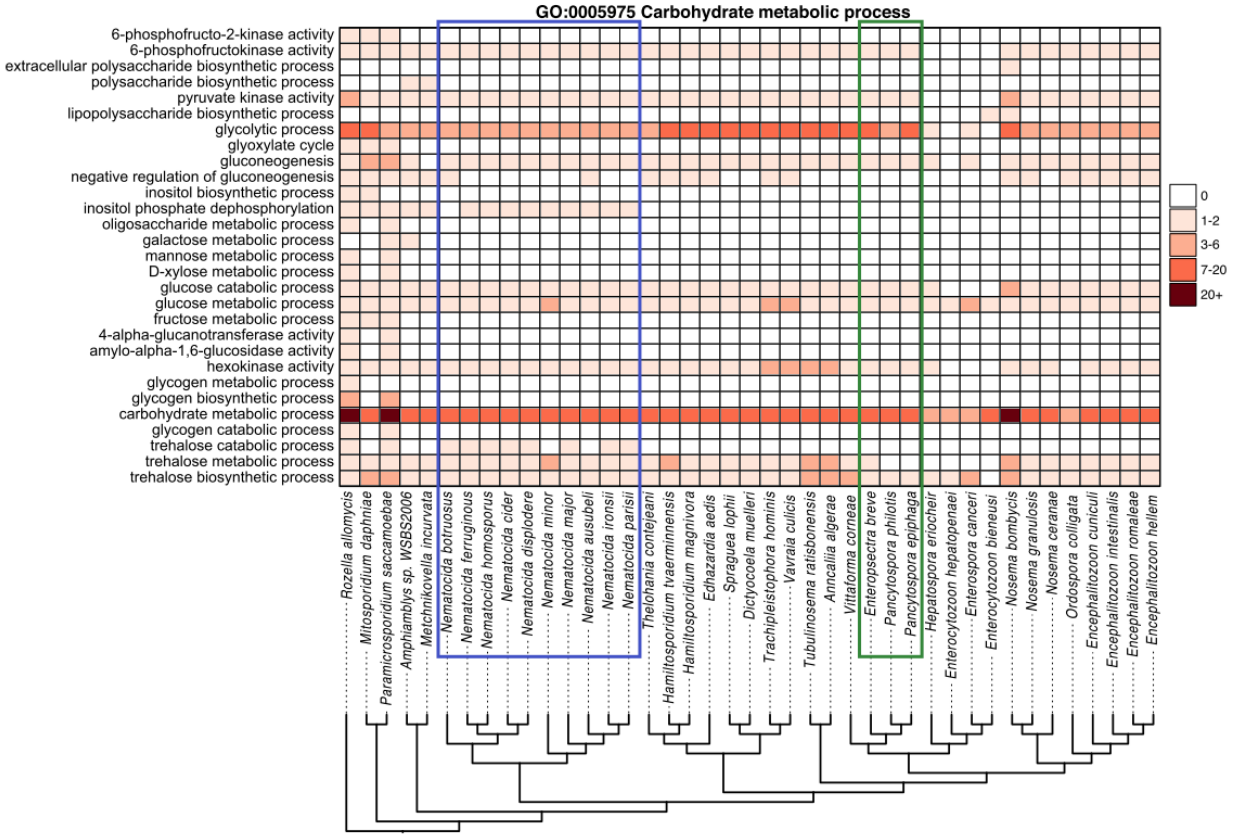
848 **S2 Fig. Phylogenetic tree of *Nematocida* species.** The phylogeny of 10 *Nematocida* species
849 was determined from single-copy orthologs identified using OrthoMCL. Phylogenetic tree was
850 generated using RaxML. *M. daphniae* is shown as an outgroup. Bootstrap values are indicated at
851 each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

852

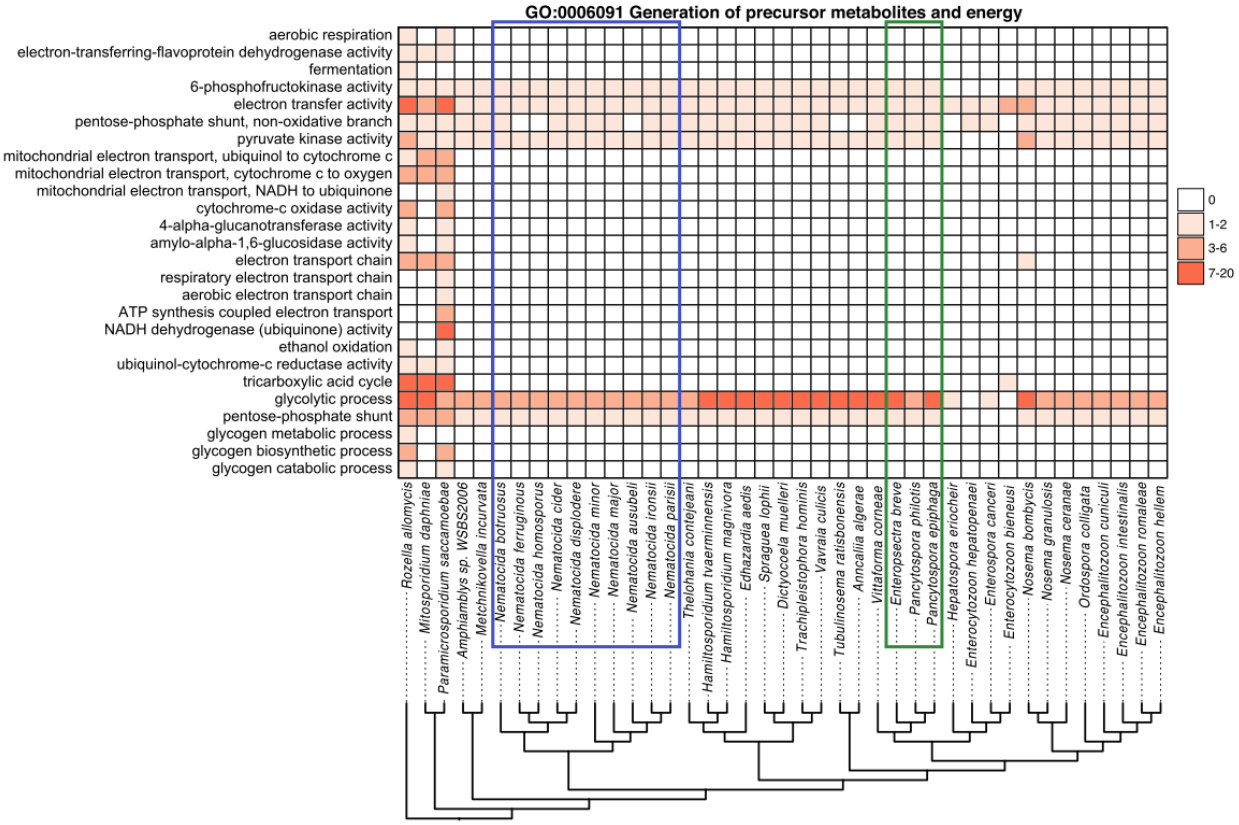


853

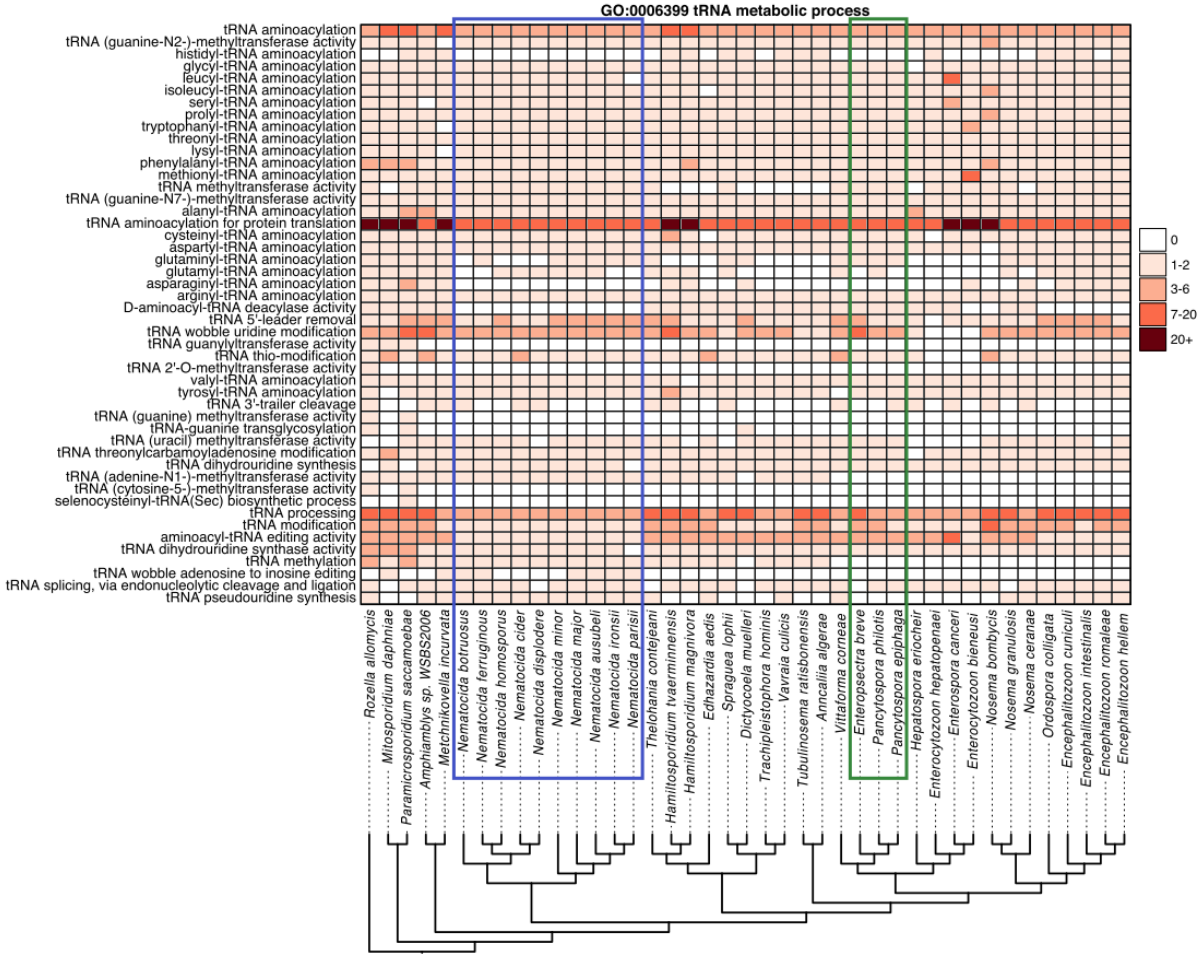
854 **S3 Fig. Genome size and number of encoded proteins are correlated in nematode-infecting**
855 **microsporidia.** The correlation between genome size and protein number in 13 nematode-
856 infecting microsporidia genomes is shown as a scatter plot. Pearson correlation coefficient and
857 p-value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



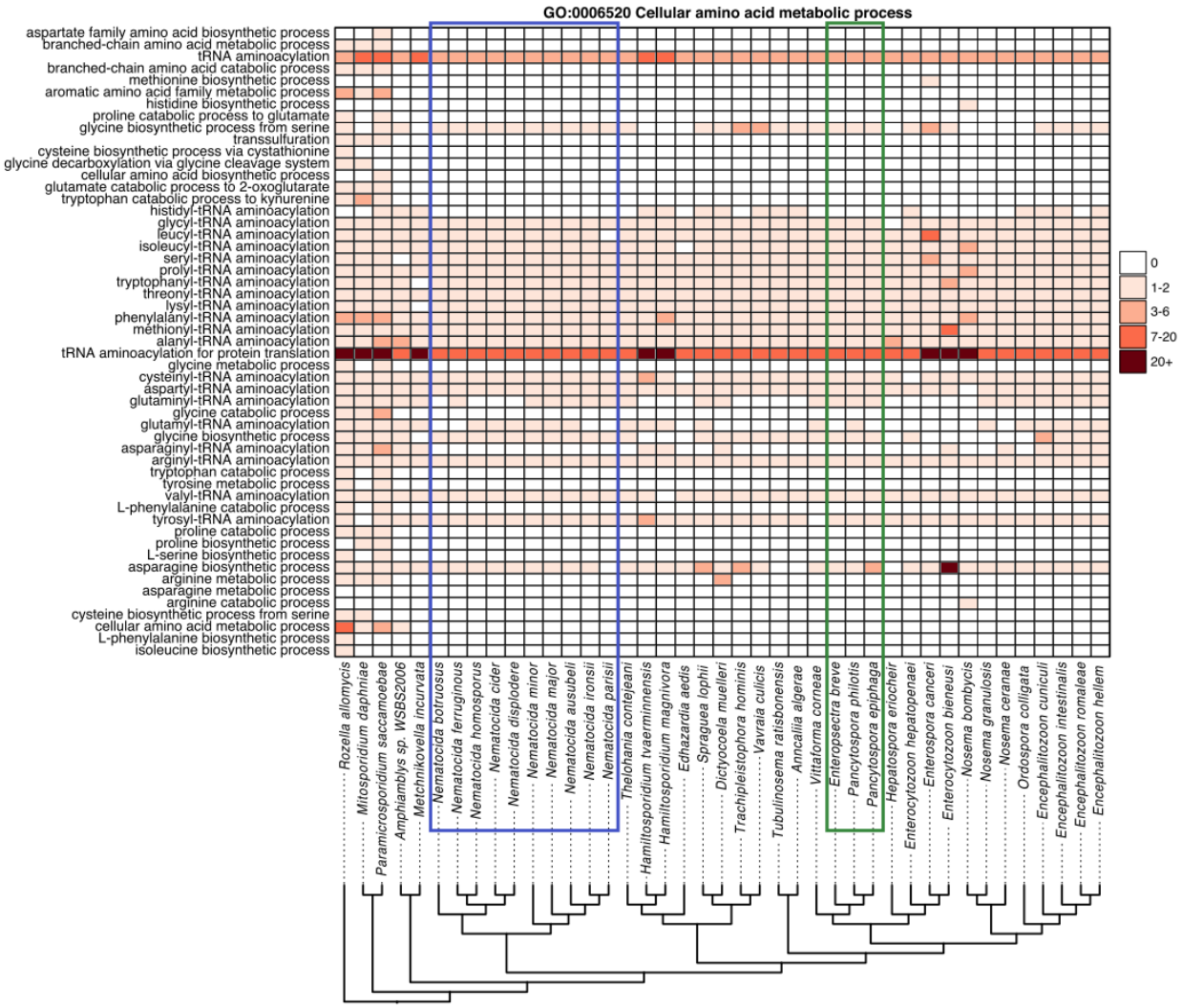
S4 Fig. Conservation of carbohydrate metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "carbohydrate metabolic process" was determined. The number of proteins from each species 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.



S5 Fig. 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.

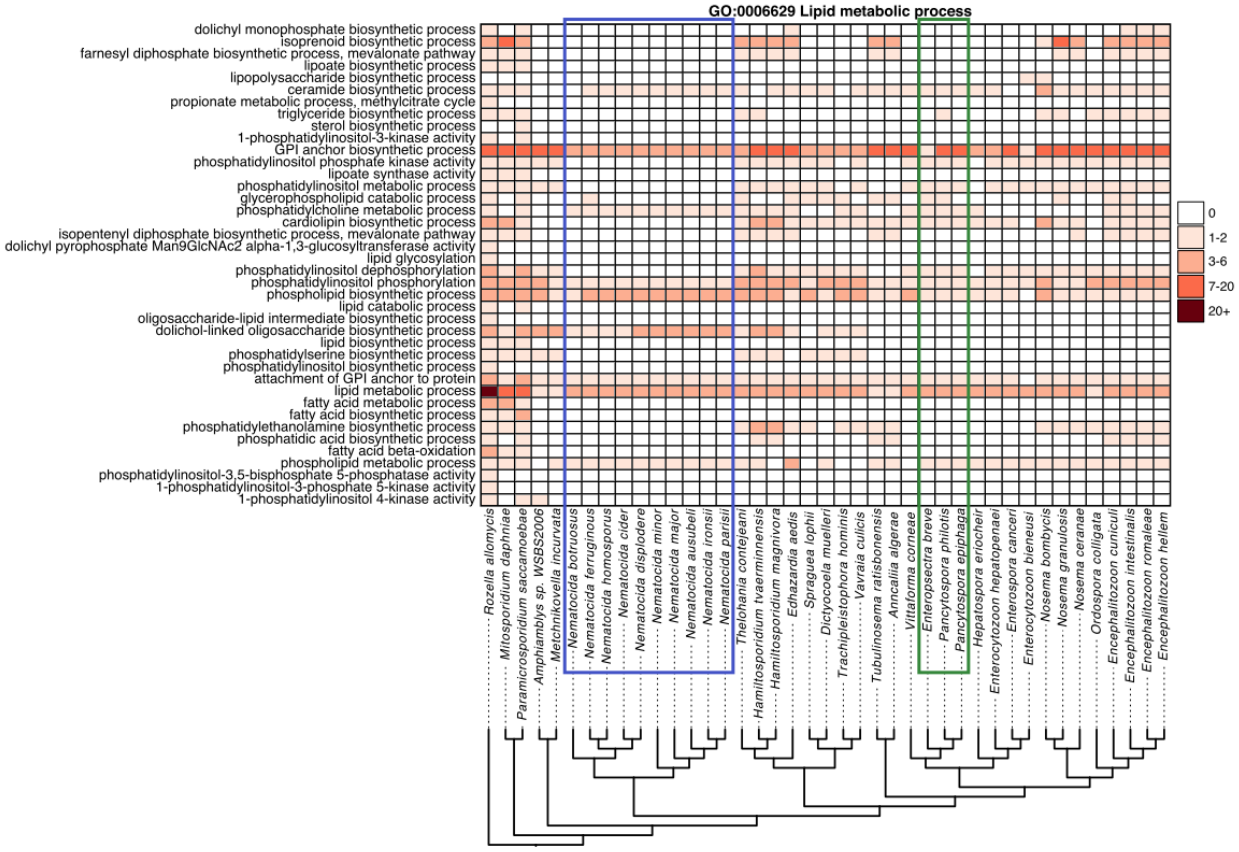


S6 Fig. Conservation of tRNA metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "tRNA metabolic process" was determined. The number of proteins from each species determined to 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.



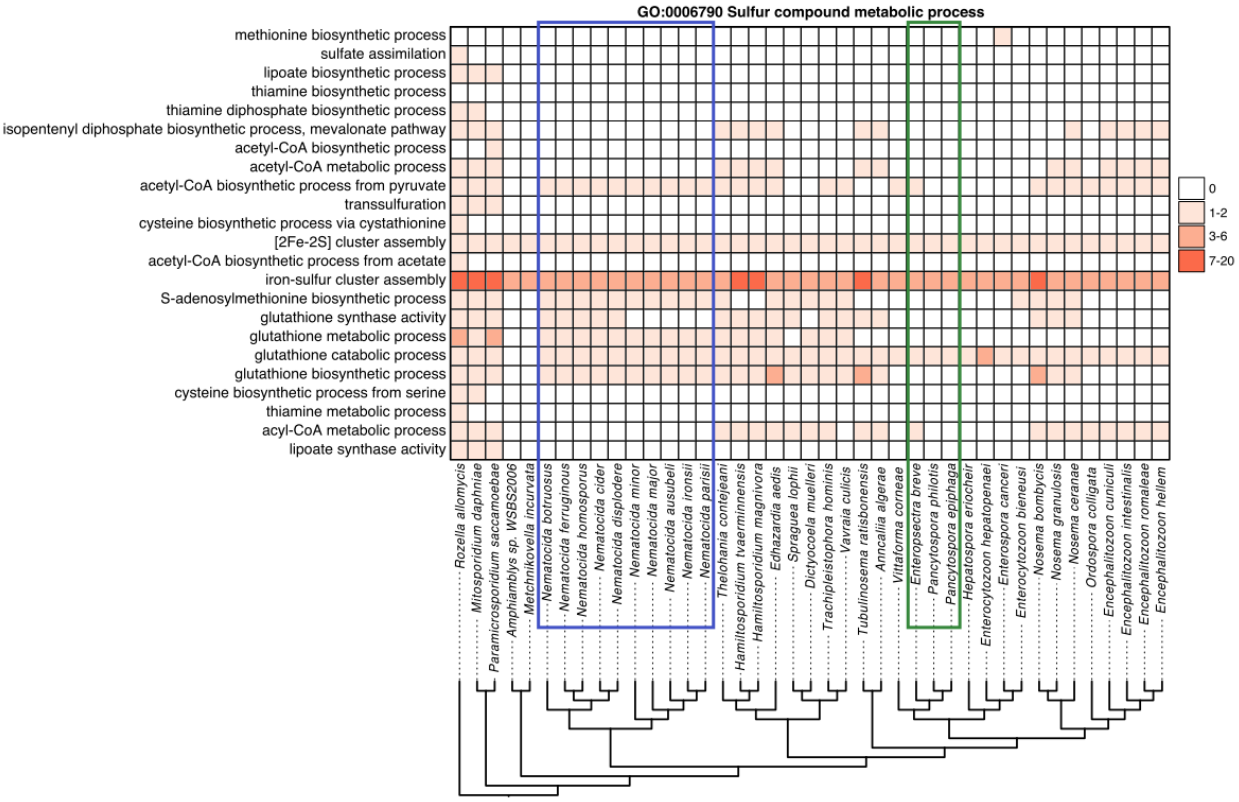
S7 Fig. Conservation of cellular amino acid metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "cellular amino acid metabolic process" was determined. The number of proteins from 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.

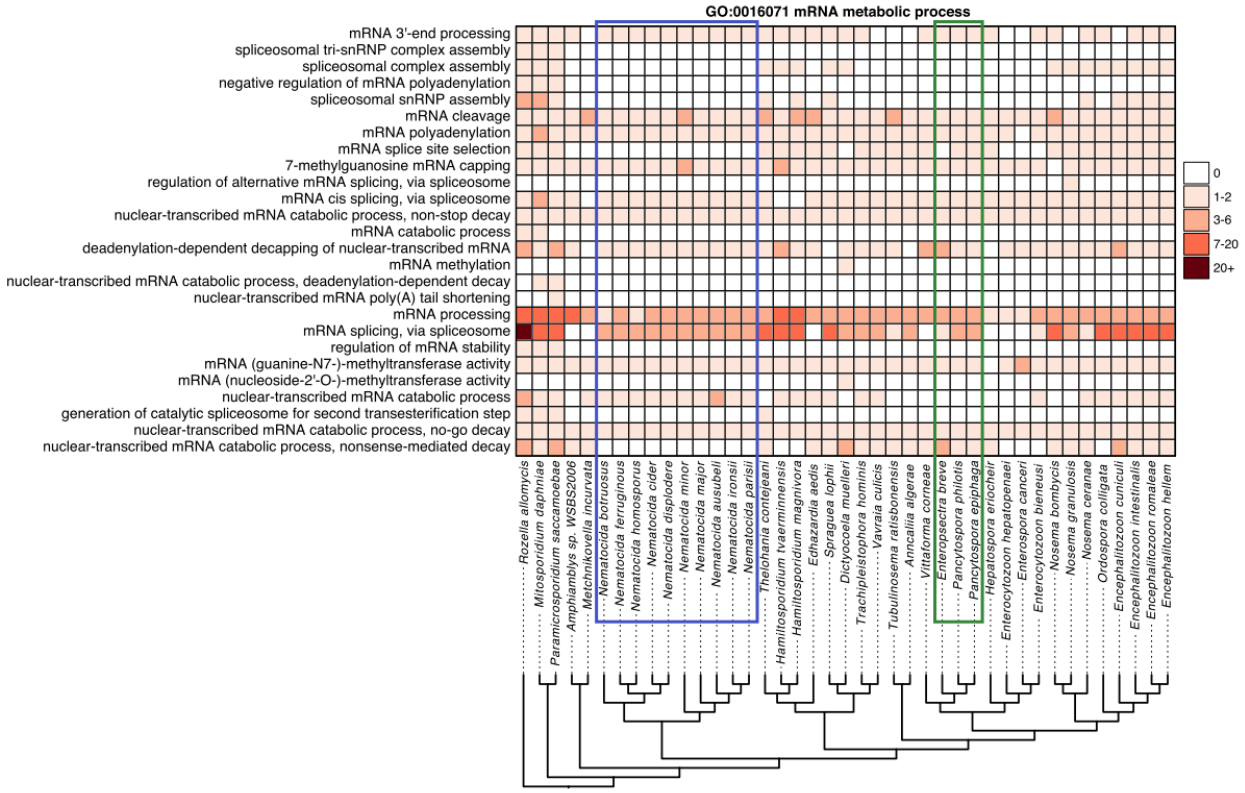


S8 Fig. Conservation of lipid metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "lipid metabolic process" was determined. The number of proteins from each species determined 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

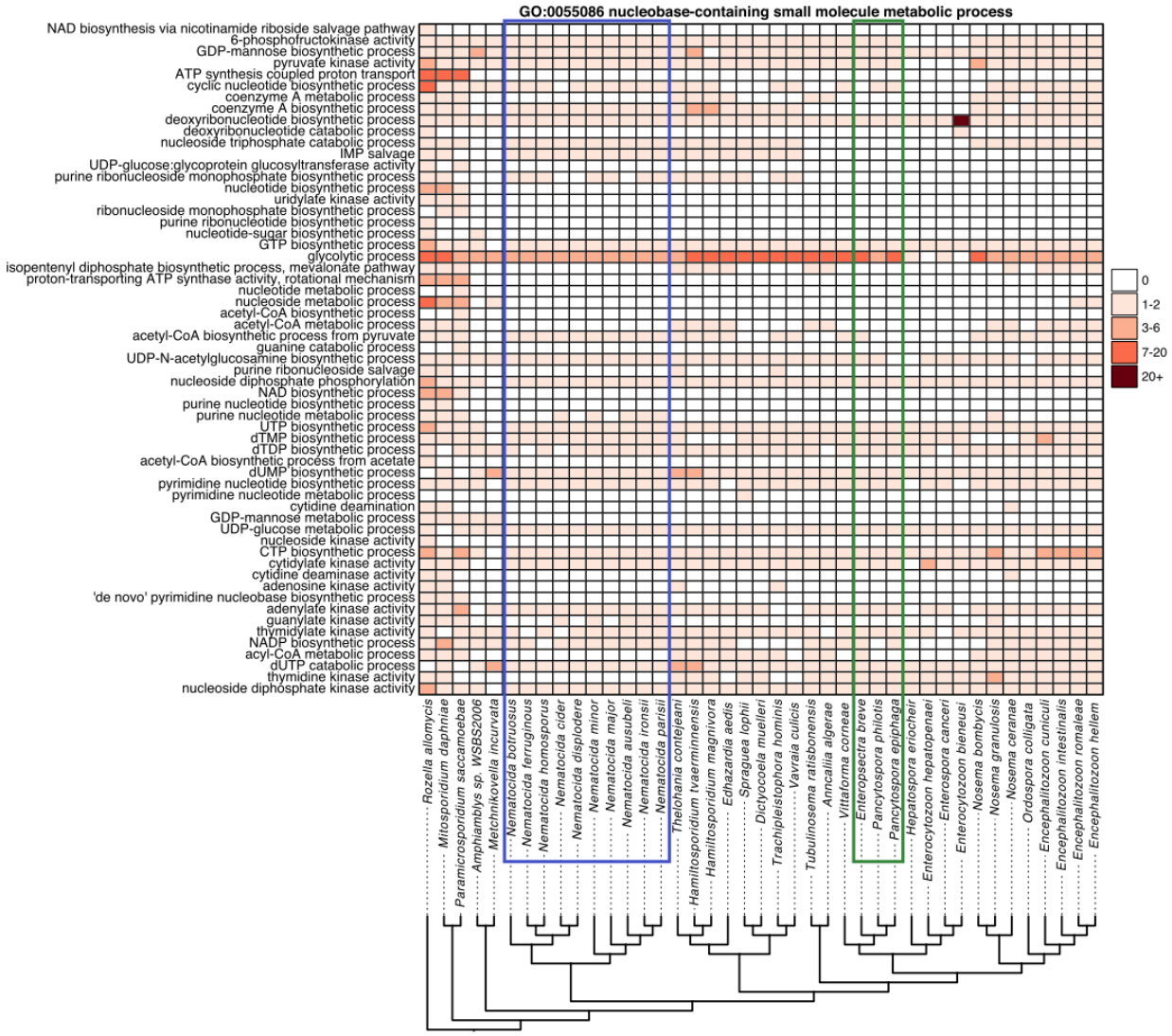
assemblies (See Fig 1). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.



S9 Fig. 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.

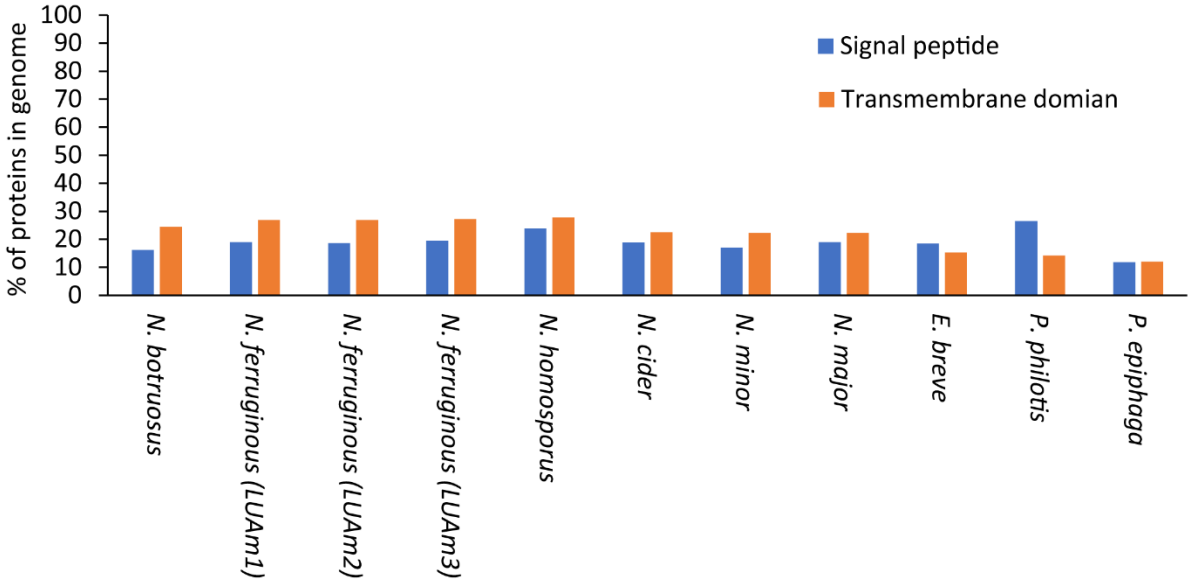


S10 Fig. Conservation of mRNA metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "mRNA metabolic process" was determined. The number of proteins from each species determined to 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). *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species are highlighted with a green box.

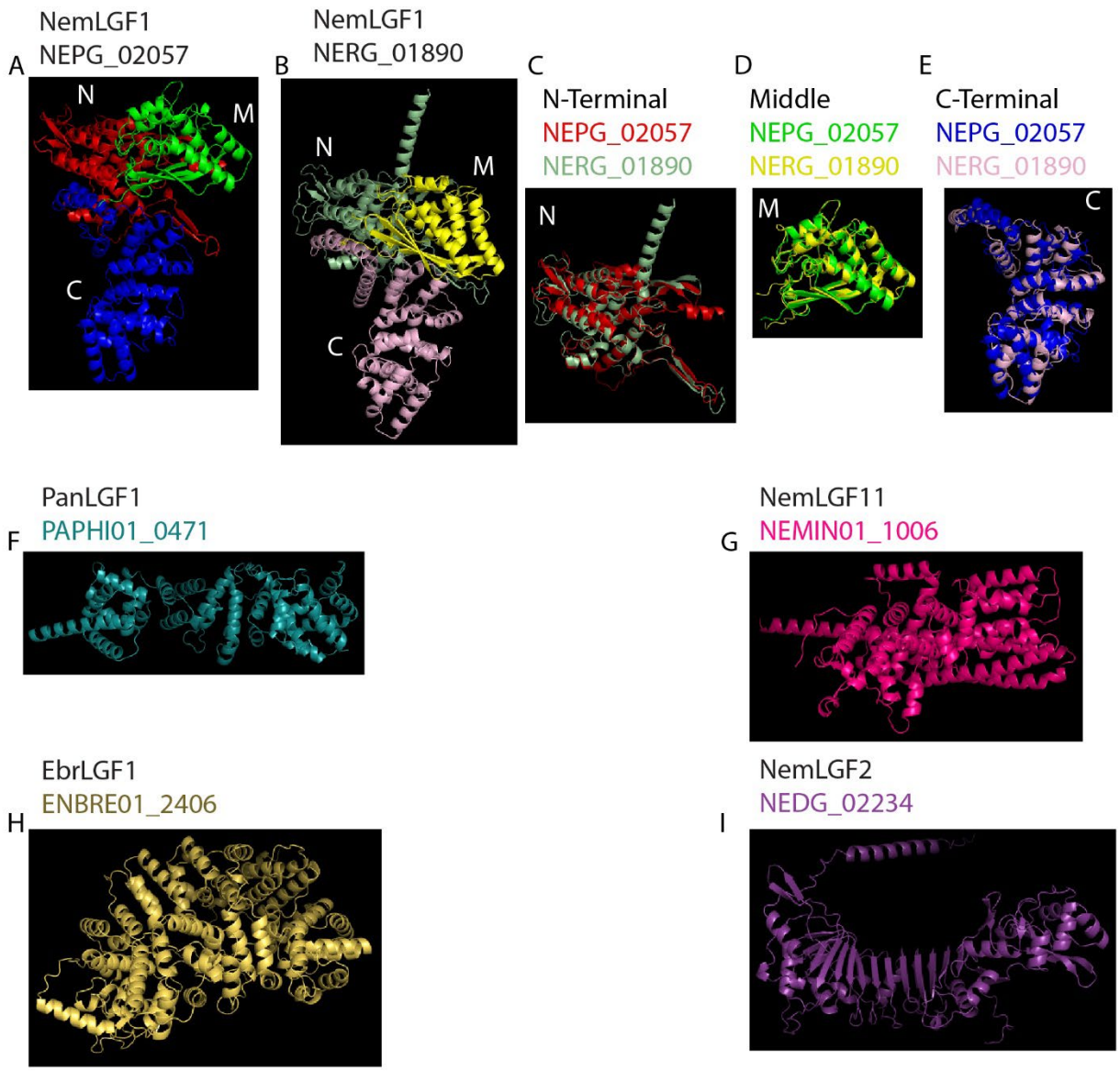


S11 Fig. Conservation of nucleobase-containing small molecule metabolic processes across microsporidia genomes. Membership of proteins from *R. allomycis* and 40 microsporidia species in descendant GO terms from the Pombe GO-slim category "nucleobase-containing small molecule metabolic process" was determined. 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).

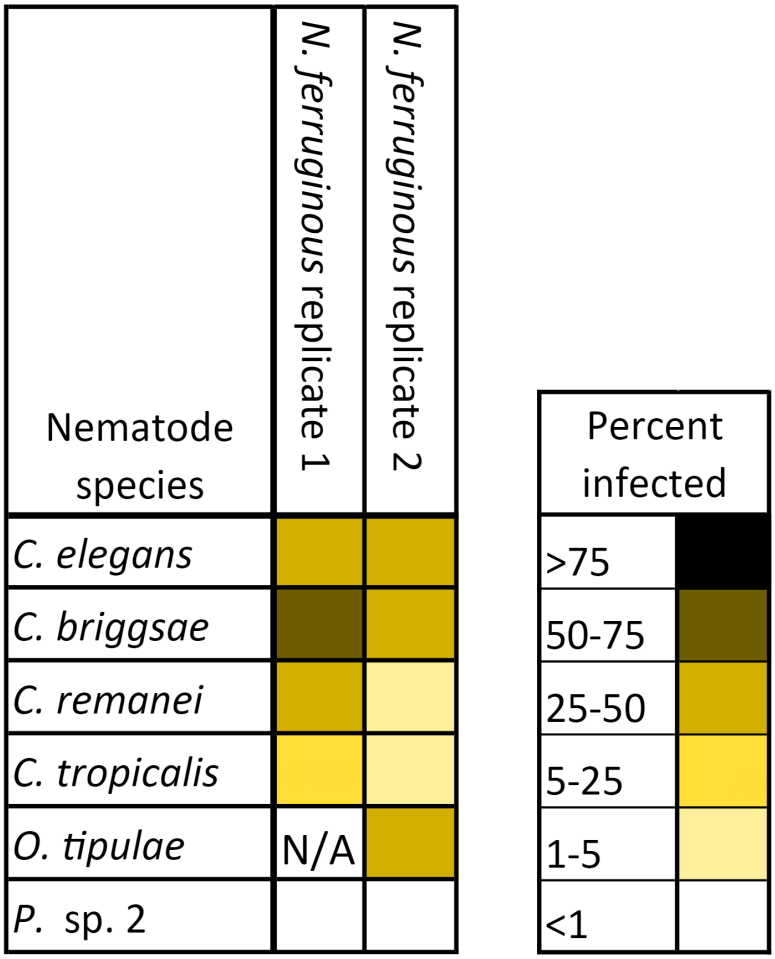
982 *Nematocida* species are highlighted with a blue box. *Enteropsectra* and *Pancytospora* species
983 are highlighted with a green box.
984
985
986
987



988
989
990 **S12 Fig. Signal peptides and transmembrane domains in Nematode-infecting**
991 **microsporidia genomes.** The percentage of proteins in each genome predicted to contain either
992 signal peptides or transmembrane domains is shown.
993
994
995
996
997
998
999
1000

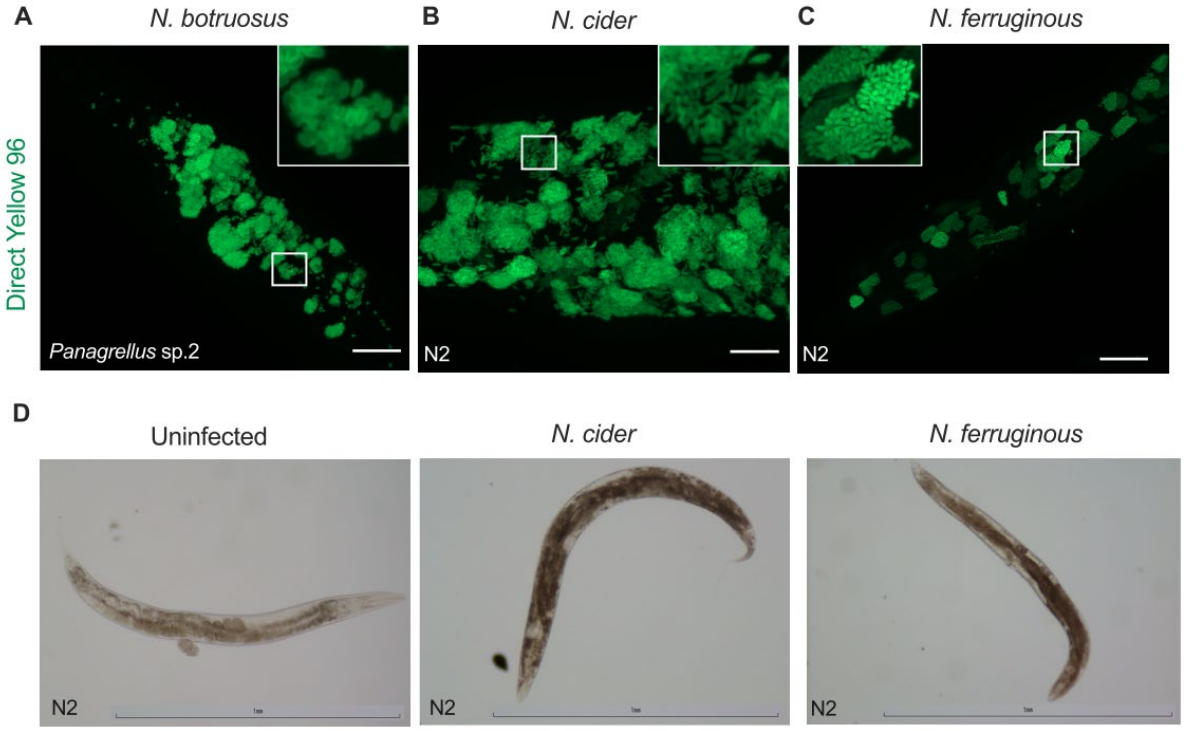


S13 Fig. Structural models of nematode-infecting microsporidia large gene families. (A-B) AlphaFold models of *N.* parisii NEPG_02057 (B) and *N. ausubeli* NERG_01890 (B). (C-E) Aligned structures of the N-terminal (C), middle (D), and C-terminal (E) domains. N, N-terminal. M, middle. C, C-terminal.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



S14 Fig. *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.

1019



1020

1021 **S15 Fig. Spore formation observed in the intestine for *N. botruosus* and throughout the**
1022 **animal for *N. cider* and *N. ferruginous*. (A-C)** L1 stage animals were exposed to spores and
1023 incubated for 96 hours, fixed, and stained with DY96. Representative images were taken with the
1024 apotome module of a ZEISS Axio Imager at 63x magnification. Multiple z-planes were imaged,
1025 and a maximum intensity projection is displayed for each sample. Scale bars, 20 μm. **(A)**
1026 *Panagrellus* sp. 2 infected with 28 million *N. botruosus* spores. **(B)** *C. elegans* N2 infected with 4
1027 million *N. cider* spores. **(C)** *C. elegans* N2 infected with 40 million *N. ferruginous* (LUAm3) spores.
1028 **(D)** *C. elegans* N2 animals were either infected with 3 million *N. cider* spores or 3 million *N.*
1029 *ferruginous* spores for 5 days. Images taken with Nikon Eclipse Ni at 10 x magnification.

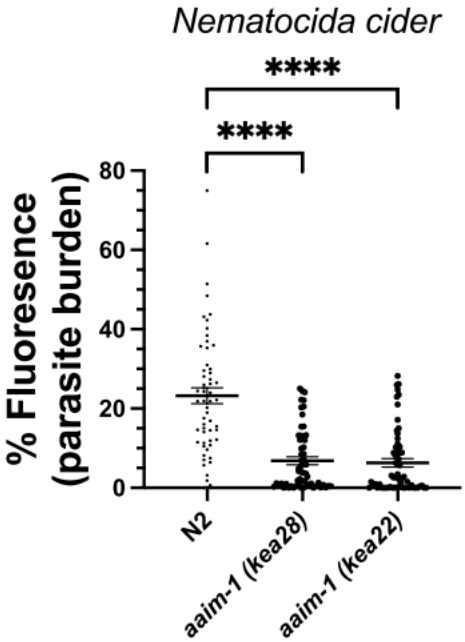
1030

1031

1032

1033

1034



S16 Fig. AAIM-1 is necessary for efficient infection by *N. cider.* N2 and *aaim-1* mutant animals infected with 4 million *N. cider* spores, fixed at 96 hours, and stained with direct-yellow 96 (DY96). 20-30 worms quantified per replicate. The percentage of the animal containing DY96 signal is shown. Mean ± SEM represented by horizontal bars. P-values determined via One-way Anova with post hoc. Significance defined as **** $p < 0.0001$.

References

1. Vavra J, Lukes J. Microsporidia and "the art of living together". Adv Parasitol. 2013;82: 253–319. doi:10.1016/B978-0-12-407706-5.00004-6

2. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature. 2001;414: 450–453. doi:10.1038/35106579

3. Wadi L, Reinke AW. Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. PLoS Pathog. 2020;16: e1008276. doi:10.1371/journal.ppat.1008276

4. Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biology. 2012;10: 47. doi:10.1186/1741-7007-10-47

5. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res. 2012;22: 2478–2488. doi:10.1101/gr.142802.112

1056 6. Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, et al. A genome-scale phylogeny of the
1057 kingdom Fungi. Current Biology. 2021;31: 1653-1665.e5. doi:10.1016/j.cub.2021.01.074

1058 7. Han B, Takvorian PM, Weiss LM. The Function and Structure of the Microsporidia Polar Tube. In:
1059 Weiss LM, Reinke AW, editors. Microsporidia: Current Advances in Biology. Cham: Springer
1060 International Publishing; 2022. pp. 179–213. doi:10.1007/978-3-030-93306-7_8

1061 8. Quandt CA, Beaudet D, Corsaro D, Walochnik J, Michel R, Corradi N, et al. The genome of an
1062 intranuclear parasite, Paramicrosporidium saccamoebae, reveals alternative adaptations to
1063 obligate intracellular parasitism. Rokas A, editor. eLife. 2017;6: e29594. doi:10.7554/eLife.29594

1064 9. Haag KL, James TY, Pombert J-F, Larsson R, Schaer TMM, Refardt D, et al. Evolution of a
1065 morphological novelty occurred before genome compaction in a lineage of extreme parasites.
1066 PNAS. 2014;111: 15480–15485. doi:10.1073/pnas.1410442111

1067 10. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, et al. Reduction and expansion
1068 in microsporidian genome evolution: new insights from comparative genomics. Genome Biol Evol.
1069 2013;5: 2285–2303. doi:10.1093/gbe/evt184

1070 11. Barandun J, Hunziker M, Vossbrinck CR, Klinge S. Evolutionary compaction and adaptation
1071 visualized by the structure of the dormant microsporidian ribosome. Nat Microbiol. 2019;4: 1798–
1072 1804. doi:10.1038/s41564-019-0514-6

1073 12. Ehrenbolger K, Jespersen N, Sharma H, Sokolova YY, Tokarev YS, Vossbrinck CR, et al. Differences in
1074 structure and hibernation mechanism highlight diversification of the microsporidian ribosome.
1075 PLOS Biology. 2020;18: e3000958. doi:10.1371/journal.pbio.3000958

1076 13. Nicholson D, Salamina M, Panek J, Helena-Bueno K, Brown CR, Hirt RP, et al. Adaptation to genome
1077 decay in the structure of the smallest eukaryotic ribosome. Nat Commun. 2022;13: 591.
1078 doi:10.1038/s41467-022-28281-0

1079 14. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host–
1080 pathogen interactions and genome evolution in two generalist and specialist microsporidian
1081 pathogens of mosquitoes. Nature Communications. 2015;6: 1–12. doi:10.1038/ncomms8121

1082 15. Wiredu Boakye D, Jaroenlak P, Prachumwat A, Williams TA, Bateman KS, Itsathitphaisarn O, et al.
1083 Decay of the glycolytic pathway and adaptation to intranuclear parasitism within
1084 Enterocytozoonidae microsporidia. Environ Microbiol. 2017;19: 2077–2089. doi:10.1111/1462-
1085 2920.13734

1086 16. Willis AR, Reinke AW. Factors That Determine Microsporidia Infection and Host Specificity. Exp
1087 Suppl. 2022;114: 91–114. doi:10.1007/978-3-030-93306-7_4

1088 17. Murareanu BM, Sukhdeo R, Qu R, Jiang J, Reinke AW, Weiss LM. Generation of a Microsporidia
1089 Species Attribute Database and Analysis of the Extensive Ecological and Phenotypic Diversity of
1090 Microsporidia. mBio. 2021. pp. e01490-21.

1091 18. Luallen RJ, Reinke AW, Tong L, Botts MR, Félix M-A, Troemel ER. Discovery of a Natural
1092 Microsporidian Pathogen with a Broad Tissue Tropism in Caenorhabditis elegans. PLoS Pathog.
1093 2016;12: e1005724. doi:10.1371/journal.ppat.1005724

1094 19. Zhang G, Sachse M, Prevost M-C, Luallen RJ, Troemel ER, Felix M-A. A Large Collection of Novel
1095 Nematode-Infecting Microsporidia and Their Diverse Interactions with Caenorhabditis elegans and
1096 Other Related Nematodes. PLoS Pathog. 2016;12: e1006093. doi:10.1371/journal.ppat.1006093

1097 20. Troemel ER, Felix M-A, Whiteman NK, Barriere A, Ausubel FM. Microsporidia are natural
1098 intracellular parasites of the nematode Caenorhabditis elegans. PLoS Biol. 2008;6: 2736–2752.
1099 doi:10.1371/journal.pbio.0060309

1100 21. Reinke AW, Balla KM, Bennett EJ, Troemel ER. Identification of microsporidia host-exposed
1101 proteins reveals a repertoire of rapidly evolving proteins. Nat Commun. 2017;8: 14023.
1102 doi:10.1038/ncomms14023

1103 22. Willis AR, Zhao W, Sukhdeo R, Wadi L, Jarkass HTE, Claycomb JM, et al. A parental transcriptional
1104 response to microsporidia infection induces inherited immunity in offspring. Science Advances.
1105 2021;7: eabf3114. doi:10.1126/sciadv.abf3114

1106 23. Tamim El Jarkass H, Mok C, Schertzberg MR, Fraser AG, Troemel ER, Reinke AW. An intestinally
1107 secreted host factor promotes microsporidia invasion of C. elegans. Lourido S, editor. eLife.
1108 2022;11: e72458. doi:10.7554/eLife.72458

1109 24. Bakowski MA, Desjardins CA, Smelkinson MG, Dunbar TL, Lopez-Moyado IF, Rifkin SA, et al.
1110 Ubiquitin-mediated response to microsporidia and virus infection in C. elegans. PLoS Pathog.
1111 2014;10: e1004200. doi:10.1371/journal.ppat.1004200

1112 25. Szumowski SC, Botts MR, Popovich JJ, Smelkinson MG, Troemel ER. The small GTPase RAB-11
1113 directs polarized exocytosis of the intracellular pathogen N. parisii for fecal-oral transmission from
1114 C. elegans. Proc Natl Acad Sci U S A. 2014;111: 8215–8220. doi:10.1073/pnas.1400696111

1115 26. Murareanu BM, Antao NV, Zhao W, Dubuffet A, El Alaoui H, Knox J, et al. High-throughput small
1116 molecule screen identifies inhibitors of microsporidia invasion and proliferation in C. elegans. Nat
1117 Commun. 2022;13: 5653. doi:10.1038/s41467-022-33400-y

1118 27. Tecle E, Troemel ER. Insights from C. elegans into Microsporidia Biology and Host-Pathogen
1119 Relationships. Exp Suppl. 2022;114: 115–136. doi:10.1007/978-3-030-93306-7_5

1120 28. Jarkass HTE, Reinke AW. The ins and outs of host-microsporidia interactions during invasion,
1121 proliferation and exit. Cellular Microbiology. 2020;22: e13247. doi:10.1111/cmi.13247

1122 29. Wan YC, Troemel ER, Reinke AW. Conservation of Nematocida microsporidia gene expression and
1123 host response in Caenorhabditis nematodes. PLOS ONE. 2022;17: e0279103.
1124 doi:10.1371/journal.pone.0279103

1125 30. Mok C, Xiao MA, Wan YC, Zhao W, Ahmed SM, Luallen R, et al. High-throughput phenotyping of C.
1126 elegans wild isolates reveals specific resistance and susceptibility traits to infection by distinct
1127 microsporidia species. bioRxiv; 2022. p. 2022.06.20.496912. doi:10.1101/2022.06.20.496912

1128 31. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics.
1129 Genome Biology. 2019;20: 238. doi:10.1186/s13059-019-1832-y

1130 32. Vossbrinck CR, Debrunner-Vossbrinck BA. Molecular phylogeny of the Microsporidia: ecological,
1131 ultrastructural and taxonomic considerations. Folia Parasitologica. 2013;52: 131–142.
1132 doi:10.14411/fp.2005.017

1133 33. Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new
1134 taxonomic, evolutionary, and ecological synthesis. Trends in Parasitology. 2022;38: 642–659.
1135 doi:10.1016/j.pt.2022.05.007

1136 34. Stentiford GD, Bass D, Williams BAP. Ultimate opportunists—The emergent Enterocytozoon group
1137 Microsporidia. PLOS Pathogens. 2019;15: e1007668. doi:10.1371/journal.ppat.1007668

1138 35. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. Infection,
1139 Genetics and Evolution. 2018;65: 329–332. doi:10.1016/j.meegid.2018.08.011

1140 36. Dia N, Lavie L, Méténier G, Toguebaye BS, Vivarès CP, Cornillot E. InterB multigenic family, a gene
1141 repertoire associated with subterminal chromosome regions of Encephalitozoon cuniculi and
1142 conserved in several human-infecting microsporidian species. Curr Genet. 2007;51: 171–186.
1143 doi:10.1007/s00294-006-0114-x

1144 37. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein
1145 folding accessible to all. Nat Methods. 2022;19: 679–682. doi:10.1038/s41592-022-01488-1

1146 38. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein
1147 structure prediction with AlphaFold. Nature. 2021;596: 583–589. doi:10.1038/s41586-021-03819-2

1148 39. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—
1149 A study of structural response in protein cores. Proteins: Structure, Function, and Bioinformatics.
1150 2009;77: 499–508. doi:10.1002/prot.22458

1151 40. Campbell SE, Williams TA, Yousuf A, Soanes DM, Paszkiewicz KH, Williams BAP. The genome of
1152 Spraguea lophii and the basis of host-microsporidian interactions. PLoS Genet. 2013;9: e1003676.
1153 doi:10.1371/journal.pgen.1003676

1154 41. Luallen R. Infection strategies of related Nematocida microsporidian species in their natural host,
1155 Caenorhabditis elegans. UC San Diego. 2016. Available: https://escholarship.org/uc/item/8r56824f

1156 42. Willis AR, Jarkass HTE, Reinke AW. Studying Inherited Immunity in a Caenorhabditis elegans Model
1157 of Microsporidia Infection. JoVE (Journal of Visualized Experiments). 2022; e63636.
1158 doi:10.3791/63636

1159 43. Han B, Weiss LM. Microsporidia: Obligate Intracellular Pathogens within the Fungal Kingdom.
1160 Microbiol Spectr. 2017;5. doi:10.1128/microbiolspec.FUNK-0018-2016

1161 44. Weidner E, Manale SB, Halonen SK, Lynn JW. Protein-Membrane Interaction Is Essential to Normal
1162 Assembly of the Microsporidian Spore Invasion Tube. Biol Bull. 1995;188: 128–135.
1163 doi:10.2307/1542078

1164 45. Julian AT, Mascarenhas dos Santos AC, Pombert J-F. 3DFI: a pipeline to infer protein function using
1165 structural homology. Bioinformatics Advances. 2021;1: vbab030. doi:10.1093/bioadv/vbab030

1166 46. Mascarenhas dos Santos AC, Julian AT, Pombert J-F. The Rad9–Rad1–Hus1 DNA Repair Clamp is
1167 Found in Microsporidia. Genome Biology and Evolution. 2022;14: evac053.
1168 doi:10.1093/gbe/evac053

1169 47. Gough J. Convergent evolution of domain architectures (is rare). Bioinformatics. 2005;21: 1464–
1170 1471. doi:10.1093/bioinformatics/bti204

1171 48. Van Etten J, Bhattacharya D. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much?
1172 Trends in Genetics. 2020;36: 915–925. doi:10.1016/j.tig.2020.08.006

1173 49. Trzebny A, Slodkowicz-Kowalska A, Becnel JJ, Sanscrainte N, Dabert M. A new method of
1174 metabarcoding Microsporidia and their hosts reveals high levels of microsporidian infections in
1175 mosquitoes (Culicidae). Molecular Ecology Resources. 2020;20: 1486–1504.
1176 doi:https://doi.org/10.1111/1755-0998.13205

1177 50. Reinke AW, Troemel ER. The Development of Genetic Modification Techniques in Intracellular
1178 Parasites and Potential Applications to Microsporidia. PLoS Pathog. 2015;11: e1005283.
1179 doi:10.1371/journal.ppat.1005283

1180 51. Huang Y, Zheng S, Mei X, Yu B, Sun B, Li B, et al. A secretory hexokinase plays an active role in the
1181 proliferation of Nosema bombycis. PeerJ. 2018;6: e5658. doi:10.7717/peerj.5658

1182 52. Paldi N, Glick E, Oliva M, Zilberberg Y, Aubin L, Pettis J, et al. Effective Gene Silencing in a
1183 Microsporidian Parasite Associated with Honeybee (*Apis mellifera*) Colony Declines. Applied and
1184 Environmental Microbiology. 2010. pp. 5960–5964.

1185 53. Saleh M, Kumar G, Abdel-Baki A-A, Dkhil MA, El-Matbouli M, Al-Quraishy S. In Vitro Gene Silencing
1186 of the Fish Microsporidian Heterosporis saurida by RNA Interference. Nucleic Acid Therapeutics.
1187 2016;26: 250–256. doi:10.1089/nat.2016.0613

1188 54. Kim I-H, Kim D-J, Gwak W-S, Woo S-D. Increased survival of the honey bee Apis mellifera infected
1189 with the microsporidian Nosema ceranae by effective gene silencing. Archives of Insect
1190 Biochemistry and Physiology. 2020;105: e21734. doi:10.1002/arch.21734

1191 55. Zheng S, Huang Y, Huang H, Yu B, Zhou N, Wei J, et al. The role of NbTMP1, a surface protein of
1192 sporoplasm, in Nosema bombycis infection. Parasites & Vectors. 2021;14: 81. doi:10.1186/s13071-
1193 021-04595-8

1194 56. Chen L, Li R, You Y, Zhang K, Zhang L. A Novel Spore Wall Protein from Antonospora locustae
1195 (Microsporidia: Nosematidae) Contributes to Sporulation. Journal of Eukaryotic Microbiology.
1196 2017;64: 779–791. doi:10.1111/jeu.12410

1197 57. Kamath RS, Ahringer J. Genome-wide RNAi screening in Caenorhabditis elegans. Methods.
1198 2003;30: 313–321. doi:10.1016/S1046-2023(03)00050-1

1199 58. Barriere A, Felix M-A. Isolation of C. elegans and related nematodes. WormBook. 2014; 1–19.
1200 doi:10.1895/wormbook.1.115.2

1201 59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1202 EMBnet.journal. 2011. p. 3.

1203 60. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
1204 Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170

1205 61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–
1206 359. doi:10.1038/nmeth.1923

1207 62. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ. ABySS: A parallel assembler for
1208 short read sequence data. Genome Res. 2009;19: 1117–1123. doi:10.1101/gr.089532.108

1209 63. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. F1000Research; 2017.
1210 doi:10.12688/f1000research.12232.1

1211 64. Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes.
1212 Nucleic Acids Research. 2016;44: e113. doi:10.1093/nar/gkw294

1213 65. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better
1214 web interface. Nucleic Acids Res. 2008;36: W5–W9. doi:10.1093/nar/gkn201

1215 66. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene
1216 recognition and translation initiation site identification. BMC Bioinformatics. 2010;11: 119.
1217 doi:10.1186/1471-2105-11-119

1218 67. Eddy SR. Accelerated Profile HMM Searches. PLOS Computational Biology. 2011;7: e1002195.
1219 doi:10.1371/journal.pcbi.1002195

1220 68. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families
1221 database: towards a more sustainable future. Nucleic Acids Res. 2016;44: D279-285.
1222 doi:10.1093/nar/gkv1344

1223 69. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from
1224 transmembrane regions. Nat Methods. 2011;8: 785–786. doi:10.1038/nmeth.1701

1225 70. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology
1226 with a hidden markov model: application to complete genomes11Edited by F. Cohen. Journal of
1227 Molecular Biology. 2001;305: 567–580. doi:10.1006/jmbi.2000.4315

1228 71. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional
1229 Characterization of Genome and Metagenome Sequences. Journal of Molecular Biology. 2016;428:
1230 726–731. doi:10.1016/j.jmb.2015.11.006

1231 72. James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared Signatures of Parasitism and
1232 Phylogenomics Unite Cryptomycota and Microsporidia. Current Biology. 2013;23: 1548–1553.
1233 doi:10.1016/j.cub.2013.06.057

1234 73. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome
1235 assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31: 3210–
1236 3212. doi:10.1093/bioinformatics/btv351

1237 74. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving
1238 coverage, classification and access to protein sequence annotations. Nucleic Acids Research.
1239 2019;47: D351–D360. doi:10.1093/nar/gky1100

1240 75. Li L, Stoeckert CJJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes.
1241 Genome Res. 2003;13: 2178–2189. doi:10.1101/gr.1224503

1242 76. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment | Nature
1243 Methods. [cited 12 Aug 2022]. Available: https://www.nature.com/articles/nmeth.1818

1244 77. Using Dali for Protein Structure Comparison - PubMed. [cited 12 Aug 2022]. Available:
1245 https://pubmed.ncbi.nlm.nih.gov/32006276/

1246 78. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic
1247 Acids Res. 2004;32: 1792–1797. doi:10.1093/nar/gkh340

1248 79. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source
1249 platform for biological-image analysis. Nature Methods. 2012;9: 676–682.
1250 doi:10.1038/nmeth.2019

1251 80. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat
1252 Methods. 2012;9: 671–675. doi:10.1038/nmeth.2089

1253