Conservation of *Nematocida* microsporidia gene expression and host response in *Caenorhabditis* nematodes

Yin Chen Wan[1], Emily R. Troemel[2], Aaron W. Reinke[1#]

[1] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.
[2] School of Biological Sciences, University of California, San Diego, La Jolla, California, United States of America.

#Author for Correspondence: Aaron W. Reinke, Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, 416-946-0889, aaron.reinke@utoronto.ca

Abstract

Microsporidia are obligate intracellular parasites that are known to infect most types of animals. Many species of microsporidia can infect multiple related hosts, but it is not known if microsporidia express different genes depending upon which host species is infected or if the host response to infection is specific to each microsporidia species. To address these questions, we took advantage of two species of *Nematocida* microsporidia, *N. parisii* and *N. ausubeli*, that infect two species of *Caenorhabditis* nematodes, *C. elegans* and *C. briggsae*. We performed RNA-seq at several time points for each host infected with either microsporidia species. We observed that *Nematocida* transcription was largely independent of its host. We also observed that the host transcriptional response was similar when infected with either microsporidia species. Finally, we analyzed if the host response to microsporidia infection was conserved across host species. We observed that although many of the genes upregulated in response to infection are not direct orthologs, the same expanded gene families are upregulated in both *Caenorhabditis* hosts. Together our results describe the transcriptional interactions of *Nematocida* infection in *Caenorhabditis* hosts and demonstrate that these responses are evolutionarily conserved.

Key words: *Nematocida* microsporidia, *Caenorhabditis* nematodes, immunity, transcriptional response, host-pathogen interactions

Significance statement

Microsporidia are a powerful model to study pathogen evolution, but much is still unknown about how these pathogens have evolved to infect multiple host species. We found that microsporidia express most of their genes similarly even when they are infecting different host species and that related host species respond similarly to different microsporidia. Our results suggests that there are conserved transcriptional responses during microsporidia infection.

Introduction

Microsporidia are obligate eukaryotic intracellular pathogens (Vávra and Lukeš, 2013). This fungal-related phylum contains over 1400 described species that infect a wide range of animal hosts including invertebrates, vertebrates, and protists (Stentiford *et al.*, 2016; Bojko *et al.*, 2022). Although as a phylum microsporidia infect a wide range of hosts, most species only infect one or several closely related hosts (Murareanu *et al.*, 2021; Willis and Reinke, 2022). Throughout evolution, microsporidia have lost many metabolic and biosynthesis genes which are present in other eukaryotes (Nakjang *et al.*, 2013). These adaptations to survive within their hosts have resulted in microsporidia having the smallest known eukaryotic genomes (Katinka *et al.*, 2001). These features of microsporidia being able to specifically infect most types of animals with a limited coding capacity have made microsporidia a powerful model for understanding the evolution of intracellular parasites (James *et al.*, 2013; Haag *et al.*, 2014; Quandt *et al.*, 2017; Wadi and Reinke, 2020).

The model nematode *Caenorhabditis elegans* has become an important system for studying microsporidia infections (Tecle and Troemel, 2022). *Nematocida parisii* and *Nematocida ausubeli* (referred to as *Nematocia* sp. 1 in some publications) are the two microsporidia species most commonly observed to infect *Caenorhabditis elegans* (Zhang *et al.*, 2016). These two species of *Nematocida* are also commonly found to infect *Caenorhabditis briggsae,* which has been developed as a comparative species to *C. elegans* (Stein *et al.*, 2003). These two nematode species live in distinct, but overlapping geographical locations (Barrière and Félix, 2005). *N. parisii* was first found infecting *C. elegans* outside of Paris and *N. ausubeli* was originally found infecting *C. briggsae* in India (Troemel *et al.*, 2008). Since then, both microsporidia species have been found infecting both hosts in Europe (Zhang *et al.*, 2016). There are some similarities of infection shared between these two *Nematocida* species; they both exclusively infect the intestine, cause intestinal cells to fuse, and have similar life cycles (Balla *et al.*, 2016). Both species also use similar types of secreted and membrane bound proteins to interface with host proteins (Reinke *et al.*, 2017). There are also differences in their infection characteristics, such as *N. ausubeli* displaying faster growth and increased impairment of host fitness (Balla *et al.*, 2016).

In response to microsporidia infection, hosts often display large transcriptional changes. The gene expression of *C. elegans* in response to *N. parisii* infection is reported to be distinct from responses to extracellular pathogens, but similar to nodavirus infection (Bakowski *et al.*, 2014). This transcriptional response has been termed the intracellular pathogen response (IPR) (Reddy *et al.*, 2017). Among upregulated IPR genes, many contain F-box, FTH, and MATH domains that are implicated in substrate recognition during ubiquitin-mediated degradation (Thomas, 2006; Bakowski *et al.*, 2014). The IPR also includes upregulation of several *Caenorhabditis* specific families that are still poorly understood, such as the PALS family (Leyva-Díaz *et al.*, 2017). Mutants that cause the IPR to be upregulated are resistant to infection and can clear *N. parisii* infections (Reddy *et al.*, 2019; Tecle *et al.*, 2021; Willis *et al.*, 2021).

Studies looking at infection of nematodes by different bacterial and fungal pathogens showed both significant overlap between pathogen infections, as well as species-specific responses (Wong *et al.*, 2007; Engelmann *et al.*, 2011; Lansdon, Carlson and Ackley, 2022). In addition, recognition of two species of oomycete displayed a shared transcriptional response (Grover *et al.*, 2021). Similarities of responses between different host species have also been observed. For example the response to nodavirus infection is conserved between *C. elegans* and *C. briggsae* (Chen *et al.*, 2017). An intergenerational transcriptional response to *Pseudomonas vranovensis* was conserved between some, but not all species of *Caenorhabditis* (Burton *et al.*, 2021). A study looking at different species of bacterial infection in the nematode *Pristionchus pacificus* also showed both similarities and differences between the transcriptional responses in the two hosts (Sinha *et al.*, 2012). Although conservation of responses between pathogen and hosts in *C. elegans* is observed, different strains of bacteria can elicit different responses, and different strains of *C. elegans* can have distinct responses (Zárate-Potes *et al.*, 2020; Lansdon, Carlson and Ackley, 2022).

To determine the extent that microsporidia gene expression is influenced by its host and how conserved the host response is to microsporidia infection, we generated and analysed transcriptional data of *N. parisii* and *N. ausubeli* infecting *C. elegans* and *C. briggsae*. We observe similar transcriptional patterns of each *Nematocida* species in the two species of hosts with only a small set of differentially regulated genes. The host response to either of the two microsporidia species was also similar. This transcriptional response is conserved across host species. Altogether, our results suggest that different *Nematocida* species do not have distinct expression programs depending on the host, and transcriptional responses of *Caenorhabditis* hosts to *Nematocida* infection are conserved.

Results

Gene expression of *N. parisii* and *N. ausubeli* is similar in *C. elegans* and *C. briggsae*

To understand how gene expression of related microsporidia and hosts species is conserved, we infected *Caenorhabditis* nematodes with *Nematocida* microsporidia (fig. 1*A*). We designed our infection experiment to compare transcriptional differences between related microsporidia and host species (fig. 1*B*). Unlike previous transcriptional profiling experiments of *Nematocida* infection which were done as continuous infections (Bakowski *et al.*, 2014; Chen *et al.*, 2017), we infected the worms for a short period of time to synchronize the infection. We pulse-infected *C. elegans* and *C. briggsae* with either *N. parisii*, *N. ausubeli*, or a mock treatment for 2.5 hours, washed to remove spores from outside the worms, and then replated the animals for a total of 10, 20, or 28 hours of infection. Each condition was done once, except for the 10-hour time point which was performed in duplicate (fig. 1*A*). Samples were stained using a probe specific to the *Nematocida* 18S rRNA and we observed that greater than 75% of each population was infected (figs. 2*A* and *B*). Compared to *N. parisii* infected animals, we observed more parasite in *N. ausubeli* infected hosts at 28 hours, consistent with a previous report that *N. ausubeli* grows faster (Balla *et al.*, 2016). RNA from infected and uninfected animals was extracted and sequenced. To determine the expression of microsporidia genes, we mapped reads of *N. parisii* and *N. ausubeli* to their respective genomes (Supplementary table S1). At 10 hours post infection, the overall percentage of reads from either *N. parisii* or *N. ausubeli* were <1%, which increased to ~3-5% at 28 hours post infection (fig. 2*C*).

To define the transcriptional patterns of *N. parisii* and *N. ausubeli* during infection, we analysed RNA-seq results of each microsporidia species between *C. elegans* and *C. briggsae* at 10, 20, and 28 hours post infection. First, we used principal component analysis to compare microsporidia gene expression in each of these samples. We observed that at the 10-hour time point, there is a larger difference between expression in the two hosts, and at the later time points, expression in the two hosts are similar (figs. 3*A* and *B*). The 20- and 28-hour time points in *N. parisii* cluster closely together, but there is a larger difference between these time points in *N. ausubeli*, likely due to the accelerated growth of this species (Balla *et al.*, 2016). Moreover, we observed the expression of genes in each microsporidia species to be similar between the two *Caenorhabditis* hosts across the timepoints (figs. 3*C* and *D*, S1*A* and S1*B*). Strong correlation of *Nematocida* gene expression between *C. elegans* and *C. briggsae* was also observed, with similar levels of correlation of each microsporidia species in either host than between replicates in the same host (figs. 3*E* and *F*, S1*C* and S1*D*). These similarities between expression in different host species also did not change across time points as the parasite had gone through larger amounts of replication. At 10 hours post infection, over 45% of *Nematocida* genes are within 2-fold of each other, and over 60% are within four-fold of each other. As the time post infection increased to 28 hours, over 75% of *Nematocida* genes are within two-fold of each other, and more than 89% are within four-fold of each other. The similarity is even more pronounced in highly expressed genes (genes with greater than 100 FPKM in either host) at the 20- and 28-hour time point where 95-99% are within four-fold of each other.

To determine genes that are significantly differentially regulated depending upon the host species infected, we compared the 10-hour time points of each *Nematocida* species between the two hosts at 10 hours post infection (fig. 4*A* and *B*). We identified 34 differentially regulated genes in *N. parisii* and 11 in *N. ausubeli* (alpha< 0.05) (Supplementary table S1). Most of these differentially regulated genes had higher expression in *C. elegans* than *C. briggsae* (31/34 in *N. parisii* and 7/11 in *N. ausubeli*). Notably, differentially regulated genes were significantly enriched for ribosomal protein genes in both *N. parisii* (17/34, Fisher's exact test p-value=2.2×10^{-18}) and *N. ausubeli* (6/11, Fisher's exact test p-value=1.2×10^{-8}). Taken together, our results indicate that the transcriptional programs of *N. parisii* and *N. ausubeli* are largely independent of which *Caenorhabditis* hosts they infect.

Shared and unique transcriptional responses of each *Caenorhabditis* host to *N. parisii* and *N. ausubeli* infection

Different microsporidia species may induce similar or distinct transcriptional responses in the same host. To address this question, we compared how either *C. elegans* or *C. briggsae* responded to infection with either *N. parisii* or *N. ausubeli* at the 10-hour timepoint for which we had replicate data. We first mapped the reads from each sample to the corresponding host genome (Supplementary table S2). In *C. elegans*, we identified 875 genes that are significantly differentially regulated in *N. parisii* infected samples and 735 genes that are significantly differentially regulated in *N. ausubeli* infected samples (fig. 5*A-B* and Supplementary table S3). For *C. briggsae,* 1091 genes are significant in *N. parisii* infected samples while 449 were significant in *N. ausubeli* infected samples (fig. 5*C-D* and Supplementary table S3). There are more downregulated than upregulated genes in each sample. Next, we directly compared shared genes which are upregulated in each of the *Caenorhabditis* hosts when infected by *N. parisii* or *N. ausubeli*. We observed significant overlap in the transcriptional response to these infections with 222 upregulated genes shared in *C. elegans* and 117 upregulated genes shared in *C. briggsae.*

We next compared both the overlapping and unique genes between samples to determine if they are enriched in known biological functions. Using statistical enrichment tests, we observed that the shared upregulated genes in *N. parisii* and *N. ausubeli* infected *C. elegans* are enriched for GO Biological Process associated with metabolism; as well as GO Molecular Process associated with catalytic activity (FDR<0.05, Supplementary table S4). We did not observe any enrichment in the shared upregulated genes in *C. briggsae*, however the 267 shared downregulated *C. briggsae* genes showed enrichment for small molecule binding and structural constituent of cuticle (Supplementary table S4). Of the non-shared genes, only the 415 downregulated genes in *C. briggsae* infected with *N. parisii* showed enrichment for GO Molecular Process associated with structural constituent of cuticle and protein binding (Supplementary table S4).

We then compared our RNA-seq data to previously published studies of the transcriptional response to *N. parisii* infection (Supplementary table S5). Bakowski *et al*. performed RNA-seq at multiple time points of a germline-deficient mutant of *C. elegans* continuously infected with spores at the L3/L4 stage at 25°C. We compared our 10-hour time point to the five time points in the Bakowski *et al.* study and observed strong overlap especially at the 8-, 16-, and 30-hour time points (Supplementary fig. S2*A*). We observed significant overlap for all the time points with both our *N. parisii* and *N. ausubeli* 10-hour post infection samples (fig. S2*B-C*). We also compared our data to Chen *et al.* where *C. elegans* was infected with *N. parisii* at the L3 stage at 20°C. We observed significant overlap with both our 10-hour *N. parisii* and *N. ausubeli* data (fig. S2*D-E*). Chen *et al.* also measured Orsay virus infected *C. elegans* and Santeuil virus infected *C. briggsae* and we saw significant similarities to these as well (fig. S2*E-D*). Together our analysis demonstrates that despite the experimental differences between studies, a largely similar upregulated host response is consistently observed.

Evolutionary conserved host response to *Nematocida* infection

To investigate the evolutionary conserved response of *Caenorhabditis* to *Nematocida* infection, we studied the expression levels of *C. elegans* and *C. briggsae* orthologs. We identified orthologs using Orthofinder (Emms and Kelly, 2019). Additionally, we determined the subset of orthologs that only had a single copy present in each nematode species. Of the orthogroups shared between *C. elegans* and *C. briggsae*, about 76% are single-copy orthologs (10826/14183). We first clustered the expression of these one-to-one orthologs, which shows a smaller cluster of upregulated genes and a larger cluster of downregulated genes (fig. 6*A*). When we compared the host response to infection between the two host species, we observed

a modest but significant corelation of response to infection between *C. elegans* and *C. briggsae* (fig. 6*B*). We then compared the statistical significance to the fold change of differentially expressed genes from the different infection conditions (fig. 6*C-D*) and investigated which of these differentially expressed genes were one-to-one orthologs (fig. 6*E-F*). We observed that only 8-17% of the differentially upregulated genes were one-to-one orthologs. Furthermore, of the most highly expressed genes (greater than four-fold), only 0-2% were one-to-one orthologs. In contrast to less that 20% of significantly upregulated genes not having one-to-one orthologs, between 29-42% of significantly downregulated genes did.

As much of the upregulated response is in the non-single copy orthologs, we analysed these genes for commonalties between samples. We performed domain enrichment analysis to determine the types of proteins induced in the non-single copy orthologs by microsporidia infection. Previous transcriptional analysis of *C. elegans* infected with either *N. parisii* or Orsay virus has identified serval types of domains enriched in expression, such as F-box, MATH/BATH, PALS, DUF713, DUF684 and C-type lectins (Bakowski *et al.*, 2014; Chen *et al.*, 2017). In addition to these domains, we also included other types of domains that are enriched in the significantly upregulated and downregulated non-single copy orthologs, for a total of thirteen domains that we analysed (Supplementary table S6). Finally, the proportion of each type of domain or gene family in the upregulated and downregulated non-single copy orthologs was determined (fig. 7*A-B*). For *C. elegans* infected with *N. parisii*, we observed that ~29% of the significantly expressed non-single copy orthologs contained at least one of these thirteen domains, with genes containing PALS domains being the most common. For genes upregulated after infection with *N. ausubeli*, we observed similar trends with ~35% containing one of these domains. The downregulated genes showed similar domain patterns for both microsporidia species, with cuticle collagen and glycosyltransferase being the two most common domains. Similar patterns regarding types of domains in regulated genes induced by the two microsporidia species was also observed in *C. briggsae* (fig. 7*B*).

As some of the same expanded gene families are upregulated in both *C. elegans* and *C. briggsae*, we sought to determine the relationships between expression of these families in response to infection. We first constructed gene trees of PALS and DUF713 families (figs. *S3-S4*). Next, we compared the expression levels of these orthologs in our 10-hour samples. This analysis shows that some phylogenetically related PALS and DUF713 genes are upregulated in both *C. elegans* and *C. briggsae* in response to *Nematocida* infection (fig. 8*A-B*).

Discussion

To understand the conservation of transcriptional responses during microsporidia infection, we took the approach of comparing interactions between related microsporidia and host species. To do this we, used a pair of *Nematocida* species that share ~66% amino acid identity and have distinct growth and phenotypic characteristics during infection in *C. elegans* (Cuomo *et al.*, 2012; Balla *et al.*, 2016; Luallen *et al.*, 2016). We used a pair of *Caenorhabditis* species, for which about 60% of their genes are orthologous and these proteins are 80% identical, which is approximately the same extent of divergence between humans and mice (Stein *et al.*, 2003). We found that responses between *Caenorhabditis* and *Nematocida* species are largely conserved. However, there are several limitations to our study. Although we monitored infection at multiple time points, we only performed duplicates at 10 hours post infection. Although a small number of replicates limit the ability to detect differentially regulated genes (Conesa *et al.*, 2016), we saw a large overlap in significantly regulated genes compared to other *N. parisii* infection datasets (Bakowski *et al.*, 2014; Chen *et al.*, 2017). Our study is also limited to only two pairs of hosts and microsporidia species. Additionally, differences in transcriptional responses between species could be dependent upon particular host and microsporidia strains or environmental conditions.

Our results suggest that *Nematocida* microsporidia species do not sense and respond differently depending upon their host environment. The main difference we observed in *Nematocida* gene expression between the two hosts was an upregulation of both small and large ribosomal subunit genes in *C. elegans* at 10 hours post infection. We also observed about three-fold more ribosomal genes being differentially regulated between *N. parisii* and *N. ausubeli*. After invasion both species undergo a lag period before starting replication. This lag period is slightly shorter in *N. ausubeli* and the larger increase of ribosomal genes in *N. parisii* is potentially related to these differences in lag time (Balla *et al.*, 2016). The lack of large differences in the host response also suggests that these *Nematocida* microsporidia are not differentially regulating host genes for their own benefit.

How hosts sense microsporidia infection is not fully understood. In mammals, several toll-like receptors are necessary for immune activation (Tamim El Jarkass and Reinke, 2020), but the proteins that *C. elegans* uses to sense and respond to microsporidia infection are mostly unknown. Several negative regulators of the IPR are known, including PALS-22, LIN-35, and PNP-1, but none of these are known to be necessary for the response to infection (Reddy *et al.*, 2017, 2019; Tecle *et al.*, 2021; Willis *et al.*, 2021). Recently, a basic-region leucine-zipper transcription factor, ZIP-1, was found to positively regulate a subset of IPR genes (Lažetić *et al.*, 2022). A viral RNA receptor, DRH-1, is necessary for activation of the IPR by the Orsay virus, but not for microsporidia infection (Sowa *et al.*, 2019). Whether there is an equivalent protein that is specifically involved in detecting microsporidia infection in *C. elegans* is unknown. Although a previous study showed that three IPR genes were induced to a lesser extent by *N. ausubeli* than *N. parisii* (Zhang *et al.*, 2016) by comparing the full transcriptional profile we see a largely similar set of genes being induced. As these different microsporidia species induce a similar transcriptional response, this suggests that some common feature, such as invasion, is detected by the host.

Nematodes, insects, and vertebrates all have strong transcriptional responses to infection by microsporidia, though the responses appear to be quite diverse (Szumowski and Troemel, 2015; Midttun *et al.*, 2020). For example, antimicrobial peptides are observed to be upregulated in silkworms and cytokines are induced in infected human cells (Ma *et al.*, 2013; Flores *et al.*, 2021). The responses seen in these other animals are quite different than what is observed in *Caenorhabditis*. One reason for this is that many IPR genes are part of large gene families that are not conserved outside of *Caenorhabditis*, such as PALS (Leyva-Díaz *et al.*, 2017). Little is known about how a host responds to different species of microsporidia. A study examining two species of microsporidia that infect the same mosquito found that a horizontal transmitted species elicited a strong transcriptional immune response, but this response was not enriched during infection with a vertically transmitted species (Desjardins *et al.*, 2015). Further studies will be necessary to know if the similar immune responses we observe to infection by two horizontally transmitted microsporidia species are common in other hosts infected by different microsporidia.

Microsporidia that infect free-living terrestrial nematodes appear to be common, and there are other genera and families of nematodes that can be cultured in the laboratory and infected with microsporidia. Additionally, there are other genera of microsporidia that have been shown to infect *C. elegans* and other nematodes. Some nematode-infecting microsporidia also infect multiple genera of nematodes, which could facilitate cross-species host expression (Zhang *et al.*, 2016). These types of broader comparisons would allow for a fuller view of how animals evolve transcriptional responses to microsporidia infection.

Material and Methods

Infection of nematodes with microsporidia

C. elegans strain N2 and *C. briggsae* strain AF16 were maintained at 21°C on 10-cm nematode growth medium (NGM) plates seeded with *Escherichia coli* OP50-1 for at least three generations without starvation. Three plates each of mixed stage populations of animals were washed with M9 and embryos extracted by treating with sodium hypochlorite/1 M NaOH for 2.5 minutes. Embryos were hatched by incubating for 18 hours at 21°C. ~30,000 L1s of each species were placed on a 10-cm plates along with 1 ml M9 and 10 µl of 10X OP50-1. Three plates of each species were prepared. Animals were infected with either 20 million *N. parisii* (ERTm1) or *N. ausubeli* (ERTm2) spores. These spores were prepared as described previously (Troemel *et al.*, 2008). Uninfected plates were used as the control. Plates were dried for 1 hour in a clean cabinet and incubated an additional 1.5 hours at 21°C. Animals were then removed from plates using two 5 ml washes of M9. After washing samples twice with 10 ml M9, all but 1 ml of M9 was removed from the samples. An additional 3 ml of M9 and 1 ml of 10X OP50-1 was added to each sample, and 1 ml of this solution was plated onto 4 10-cm plates per sample, resulting in ~6,000 animals per plate. At either 10, 20, or 28 hours post infection, plates were harvested by washing off worms 3 times with 725 µl M9. Samples were then washed 3 times with 1 ml M9/0.1% Tween-20.

To determine the extent of infection, ~1,000 animals from each sample were removed, fixed, and stained using an *N. parisii* 18S rRNA fluorescent in situ hybridization probe as previously described (Reinke *et al.*, 2017). Vectashield containing DAPI (Vector Labs) was added and samples imaged using an Zeiss Axioimager 2.

RNA extraction and sequencing

After the last wash, 1 ml of TRIzol was added to each sample and RNA was extracted similar to as described previously (Bakowski *et al.*, 2014). mRNA libraries were prepared by the UCSD IGM Genomics Center and sequenced on a single lane of an Illumina HiSeq 4000, using 100 bp paired-end reads.

Microsporidia gene expression analysis

Reads from each sample were mapped to either *Nematocida parisii* strain ERTm1 (Genbank: GCF_000250985.1) or *Nematocida ausubeli* strain ERTm2 (Genbank: GCA_000250695.1) using TopHat v2.1.2 (Trapnell, Pachter and Salzberg, 2009). The respective *N. parisii* and *N. ausubeli* genome annotation files were converted into .gtf format using Gffread v0.12.3. Transcriptome assemblies were then generated using Cufflinks v2.2.1 (Trapnell *et al.*, 2010) and Cuffmerge in the Cufflinks package. Differentially expressed genes were determined using Cuffdiff v2.2.1 (Trapnell *et al.*, 2013) and were visualised using the R package cummeRbund v2.36.0 (L. Goff, 2017).

The Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values of *N. parisii* and *N. ausubeli* transcripts in each *Caenorhabditis* host were calculated by Cuffdiff v2.2.1 (Trapnell *et al.*, 2013). The linear correlation between gene expression in each host was determined using the R packages ggplot2 v3.3.5 and ggpubr v0.4.0; the ratio of FPKM values between the two hosts was calculated by dividing the microsporidia's FPKM values in *C. elegans* by that of *C. briggsae*'s. Hierarchical clustering of log10-transformed FPKM+1 values was done using the hclust function from the R package stats v4.1.1; heatmaps were generated by ggplot2 v3.3.5 in R.

***Caenorhabditis* RNA-seq analysis**

The paired end reads of each sample 10 hours post infection sample were submitted to Alaska v1.7.2 (http://alaska.caltech.edu). Briefly, Bowtie2 (Langmead and Salzberg, 2012), Samtools (Li *et al.*, 2009), RSeQC (Wang, Wang and Li, 2012), FastQC, and MultiQC (Ewels *et al.*, 2016) were used for quality control of the input files. Then, Kallisto (Bray *et al.*, 2016) was used for read alignment and quantification, followed by differential analysis using Sleuth (Pimentel *et al.*, 2017). For *C. elegans* samples, the reads were aligned to the N2 reference of the WS268

release (accession: PRJNA13758); the reads from *C. briggsae* samples were aligned to reference genome of the WS268 release (accession: PRJNA10731).

Log2 fold change values of duplicated genes in each of the eight samples were averaged using the R package dplyr v1.0.8. Genes with FDR-adjusted p-value of <0.05 were regarded as significant. Differentially upregulated genes were defined as those with an FDR-adjusted p-value<0.05 and log2 fold change ≥ 0 (infected vs. control); differentially downregulated genes were defined as those with an FDR-adjusted p-value <0.05 and log2 fold change ≤ 0 (infected vs. control).

Principal component analysis

The normalised abundance measurements in *C. elegans* and *C. briggsae* generated by Alaska, were read by the readRDS() function of the R package base v4.1.1. Normalized counts for each gene in each host species were generated via sleuth_to_matrix() with the "obs_norm" data and "tpm" units. PCA of gene expression was performed using samples from each species via the R package pcaexplorer v2.20.2 (Marini and Binder, 2019).

Determination of gene expression overlap

Genes with an FDR<0.05 were used to compare expression overlap. Values of duplicated genes in each sample were averaged using the R package dplyr v1.0.8. Overlap between samples was determined using the R package gplots v3.1.1. The p-values were calculated using the Fisher exact test. Statistical enrichment tests of shared and non-shared genes were performed using PANTHER on pantherdb.org (Mi *et al.*, 2021). Each input list was statistically tested for enrichment against the annotation sets: "GO biological process complete", "GO cellular component complete", "GO molecular function complete", "PANTHER GO-Slim Biological Process", "PANTHER GO-Slim Cellular Component", "PANTHER GO-Slim Molecular Function", "PANTHER Pathways", "Reactome pathways" and "PANTHER Protein Class". Results from these tests with p-value<0.05 are in Supplementary table S4.

To determine overlap in our samples compared to differentially regulated genes from Bakowski *et al*., we first used Alaska to process the read files of *C. elegans* infected by *N. parisii* samples at 8, 16, 30, 40 and 64 hours. To compare our samples' expression with data from Bakowski *et al*., genes expressed in at least four of those samples were used for hierarchical clustering by hclust function in the R package stats v4.1.1. The dendrogram in the was produced using the R package ggdendrogram v0.1.23; the heatmap was generated by ggplot2 v3.3.5. Overlap between our samples, the Bakowski *et al.,* (2014) samples, and the Chen *et al.* (2017) samples containing differentially regulated genes for *C. elegans* N2 infected by Orsay virus or *N. parisii*, and *C. briggsae* infected by Santeuil virus were calculated as described above.

Domain enrichment analysis

The type of gene classes and domains investigated for the enrichment analyses are listed in Supplementary table S6. Among these, we actively chose to look at F-box, MATH/BATH, PALS, DUF684, DUF713, C-type lectins, and *skr* domains containing genes, implicated from differentially regulated genes of *N. parisii* infected *C. elegans* in previous publications (Bakowski *et al.*, 2014; Chen *et al.*, 2017). We also examined additional types of genes and domains enriched in our datasets using DAVID Bioinformatics Resources (2021) (Huang, Brad T. Sherman and Lempicki, 2009; Huang, Brad T Sherman and Lempicki, 2009) for *C. elegans* and the WormBase Simple Gene Queries tool for *C. briggsae*. From this analysis, we found additional domains and genes enriched with greater than three domain-containing proteins in any of our samples. These identified domains are chitinase-like (*chil*) proteins, CUB and CUB-like domains, cytochrome P450, glucosyltransferase family 92, nematode cuticle collagen N-terminal domain, and UDP-glucuronosyltransferase. A list of gene names from respective gene classes were downloaded separately from Wormbase (http://wormbase.org/) and the corresponding protein-coding sequences were extracted from Wormbase ParaSite

(https://parasite.wormbase.org/). Alignments of other domains or classes were downloaded directly from Pfam (http://pfam.xfam.org/) separately. After deleting duplicated sequences, protein sequences of these classes of genes were converted to Stockholm format using Clustal Omega Multiple Sequence Alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). Hmmbuild of HMMER v3.3.2 (Eddy, 2011) was used to build respective profile HMMs, which was then searched against *C. elegans* (accession: PRJNA13758) proteome of WS268 release using hmmsearch. Output genes with E-value <1e-5 were used as the gene list for the enrichment analyses. In the output genes, MATH and BATH genes were combined into one list; *fbxa*, *fbxb* and *fbxc* genes were merged into the gene list for F-box while genes with CUB and CUB-like domains were categorised together (Supplementary table S6). Using the R package gplots v3.1.1, the number of genes overlapping between the lists and samples were computed, then plotted using R package ggplot2 v3.3.5. Genes that do not fall into any of these gene classes or have any of these domains were categorized as "other".

Determination of orthologs between *C. elegans* and *C. briggsae*

Orthologous genes were determined between *C. elegans* (accession: PRJNA13758) and *C. briggsae* (accession: PRJNA13758) proteomes of the WS268 release using Orthofinder v2.5.2 with the default settings (Emms and Kelly, 2019).

Single copy orthologs analysis

Genes that are single copy orthologs were extracted, based on the single copy orthologs list computed by Orthofinder. Single copy orthologs expressed in at least one out of the four samples were used for hierarchical clustering via the function hclust() from the R package stats v4.1.1, then plotted using R packages ggdendrogram and ggplot2 v3.3.5. The linear correlation of each single copy ortholog expression between the two host was calculated using ggpubr v0.4.0. The scatterplots and volcano plots were generated by ggplot2 v3.3.5.

Comparison of gene expression in expanded gene families

The *C. elegans* and *C. briggsae* genes in PALS and DUF713 families were obtained from the output of HMMER v3.3.2 (Eddy, 2011) using a E-value threshold of <1e-5. The protein sequences of the genes were obtained from Wormbase ParaSite (https://parasite.wormbase.org/) and aligned by M-Coffee using default settings (Notredame, Higgins and Heringa, 2000; Di Tommaso *et al.*, 2011). A phylogenetic tree for each family was generated using MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) with the following settings: lset nst = 1, rates = invgamma, nruns = 2, stoprule = YES, stopval =0.05 and mcmcdiagn=YES. A dendrogram of the phylogenetic tree of each gene family was created using the function ReadDendrogram from the R package DECIPHER v2.22.0. For each gene family, the orthologs' log2 fold change values across the *C. elegans* and *C. briggsae* samples were plotted as a heatmap using ggplot2 v3.3.5.

Data availability

All samples were deposited under NCBI BioProject PRJNA841614 and the sequence reads for all samples were submitted to the NCBI Sequence Read Archive.

Acknowledgements

We thank Hala Tamim El Jarkass and Meng A. Xiao for providing helpful comments on the manuscript. We thank Winnie Zhao for technical help for the comparisons to the Bakowski data. 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 NIH R01's AG052622 and GM114139 to E.T. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440) and we thank WormBase.

Author contributions: Y.C.W. and A.R. conceived of the project. Y.C.W. performed all computational analyses. A. R. performed all experiments. E.T. and A. R. provided mentorship and acquisition of funding. Y.C.W. and A.R. co-wrote the paper with edits from E.T.

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

Figures and legends

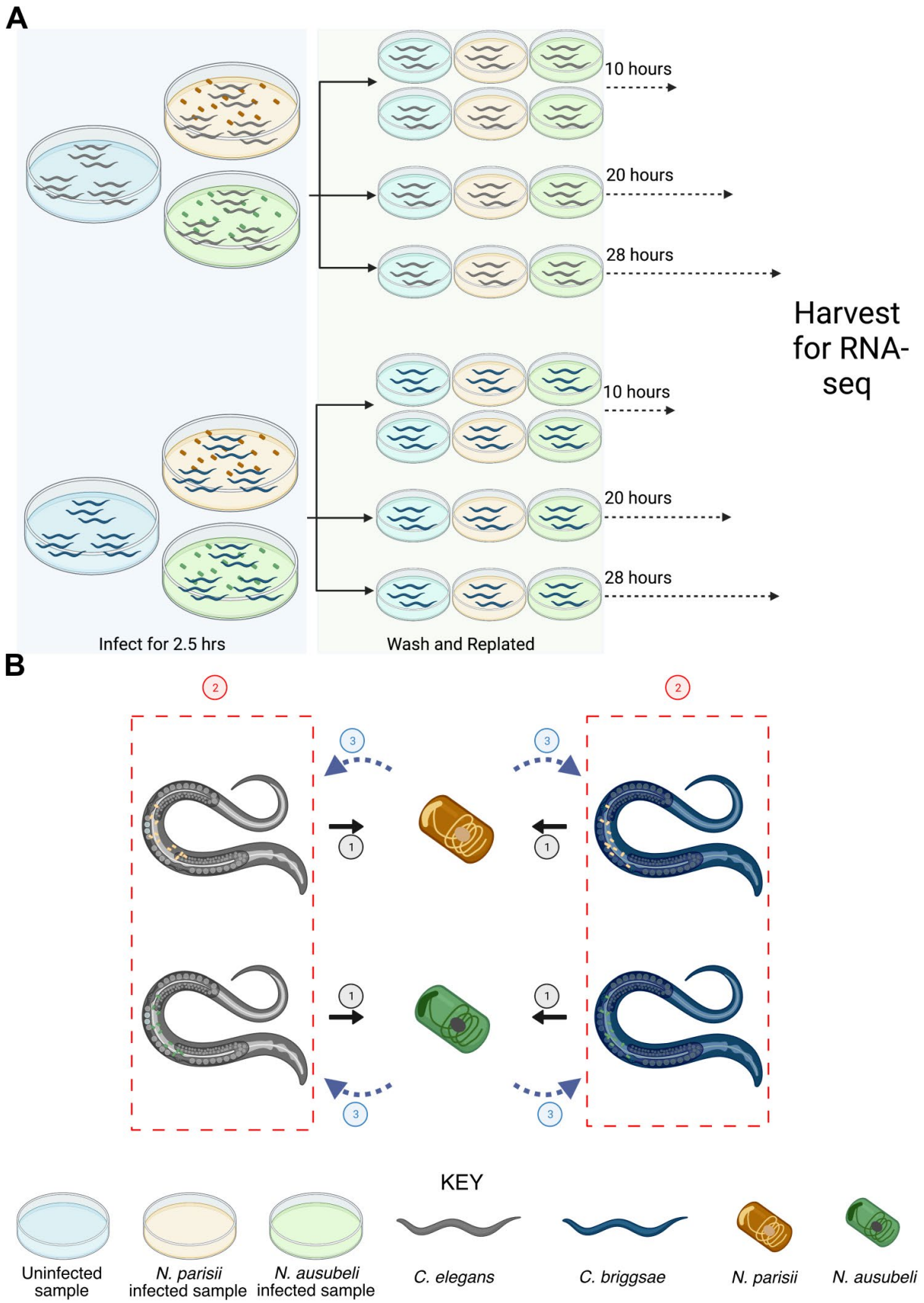


Fig. 1. Overview of study design. (*A*) Schematic diagram of the RNA-seq experiment. *C. elegans* and *C. briggsae* were either not infected or infected with *N. parisii* or *N. ausubeli* separately at the L1 stage for two and a half hours. Next, animals were washed to remove any remaining microsporidia spores and replated. Worms were harvested for RNA-sequencing at

498 10, 20 and 28 hours after infection. (*B*) The three comparisons of transcriptional responses
499 made in this study are represented. Comparison 1 is microsporidia gene expression between
500 hosts, comparison 2 is the response of each host species to different microsporidia species,
501 and comparison 3 is the conserved response between different hosts to each microsporidia
502 species. Figure was created with biorender (www.biorender.com).
503

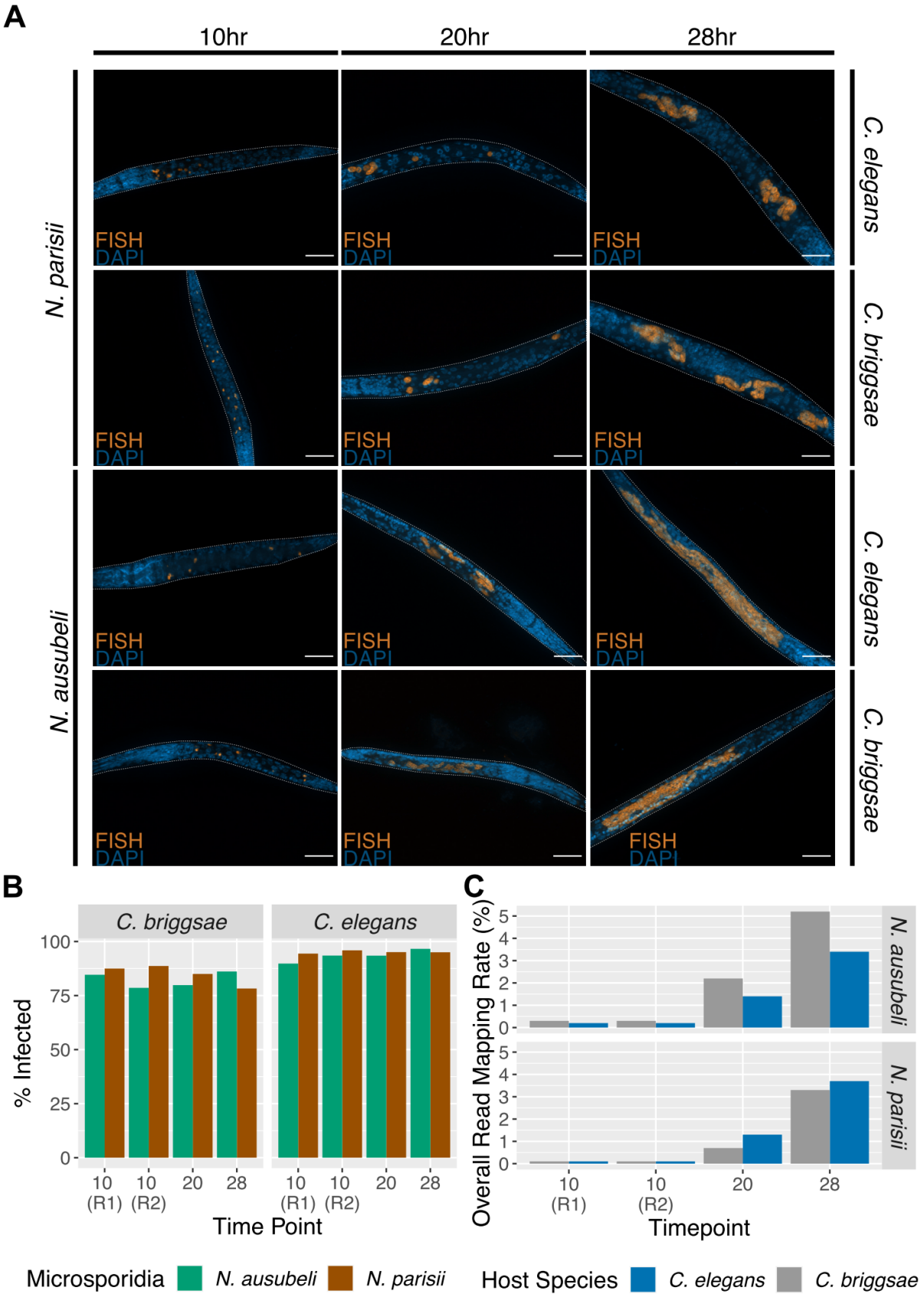


Fig. 2. *Nematocida* infection of *Caenorhabditis* hosts. (A) Pulse-infected *C. elegans* or *C. briggsae* were fixed and then stained with DAPI to detect nuclei and a FISH probe specific to *Nematocida* 18S rRNA to detect either *N. parisii* or *N. ausubeli* at 10, 20 and 28 hours after

infection. Representative images of infection are shown. Scale bars are 23 μm. (B) Percentage of infected animals at different timepoints. Between 39 to 220 animals were counted for each sample. (C) Overall read mapping rate of *Nematocida* in each host at different time points. R1 or R2 indicates the replicate samples at the 10-hour timepoint.

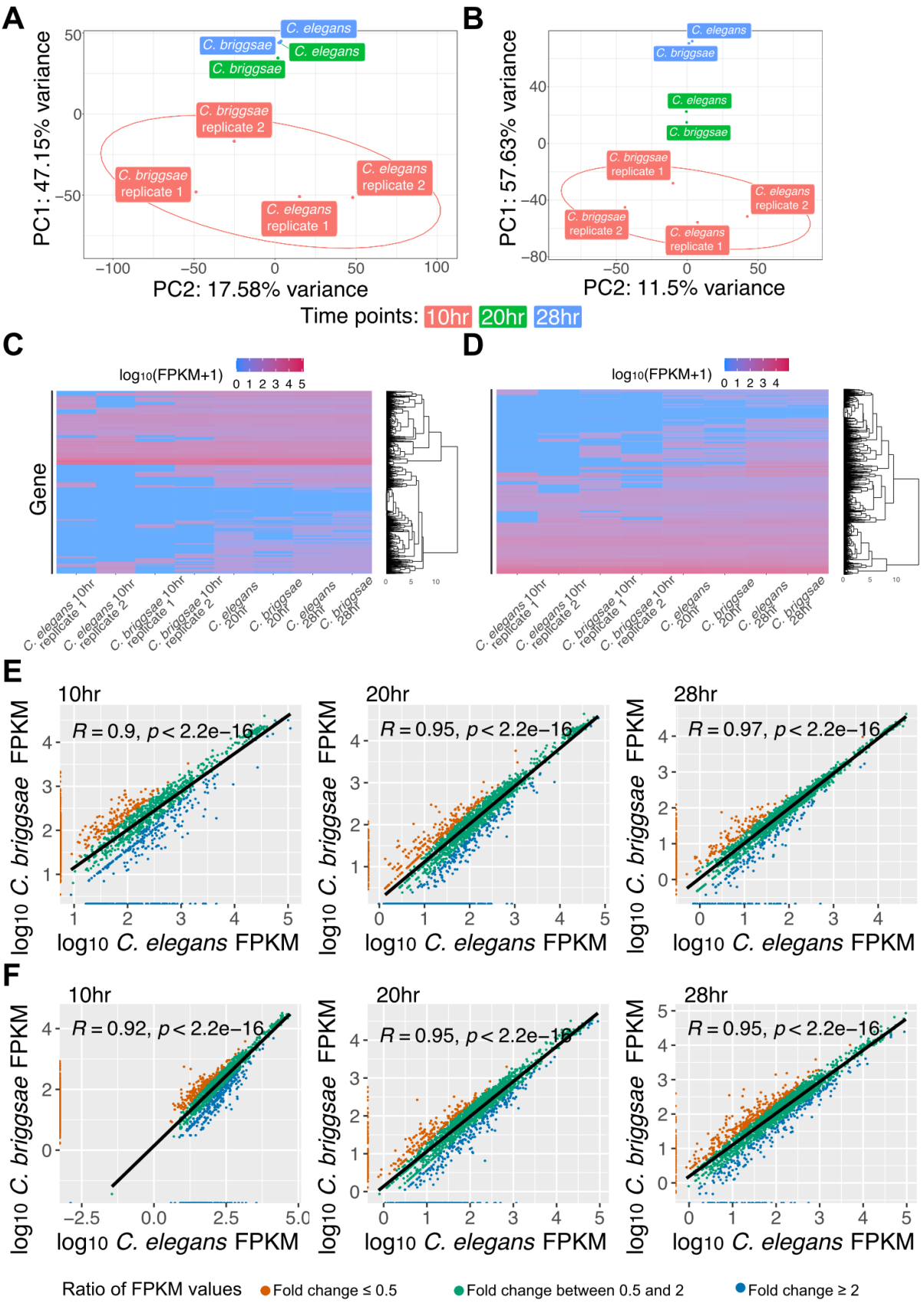
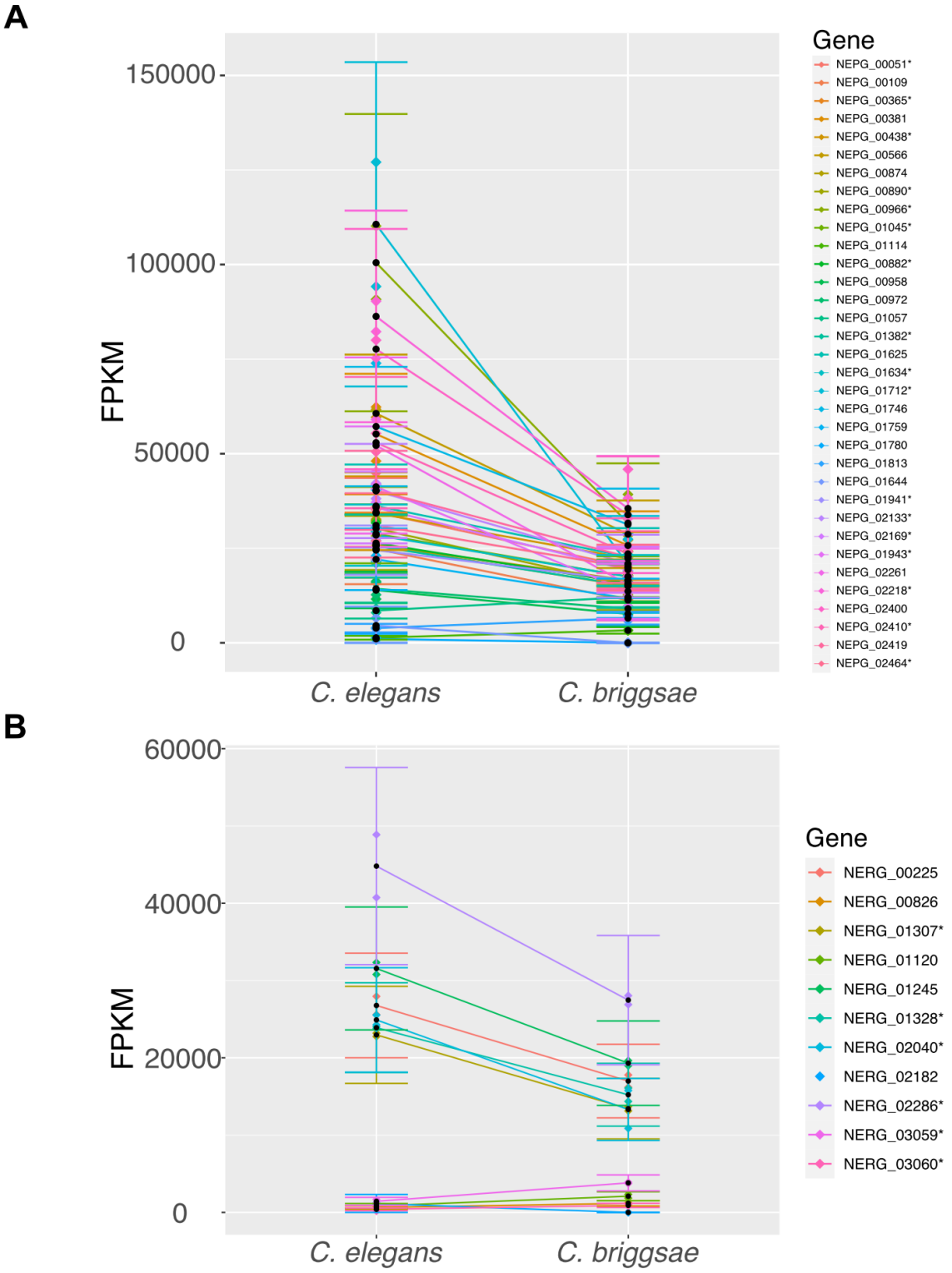
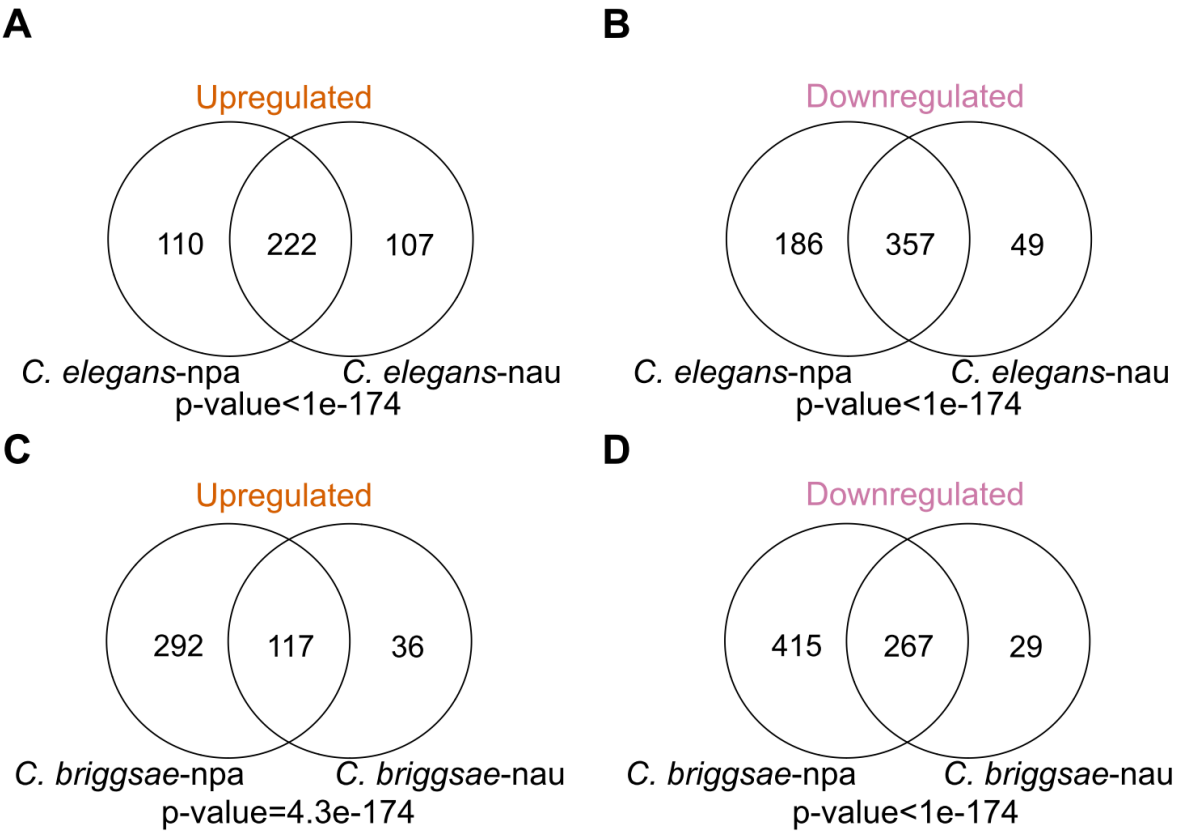


Fig. 3. Microsporidia gene expression in different host species. Principal component analysis of (A) *N. parisii* and (B) *N. ausubeli* in *C. elegans* and *C. briggsae* at 10, 20 and 28 hours. Circles represent confidence ellipses around each strain at 95% confidence interval.

(C-D) Heatmap of transcriptional profiles of genes expressed in *N. parisii* (C) and *N. ausubeli* (D). Rows represent gene clustered hierarchically. Scale is of differential regulation of infected compared to uninfected samples. (E-F) Scatterplot of log10 FPKM values in *N. parisii* (E) and *N. ausubeli* (F) when infecting *C. briggsae* and *C. elegans*. Each point represents a microsporidian gene. Pearson correlation value and p-value are indicated on the top left of each plot. The ratio of FPKM values for each gene between *C. elegans* and *C. briggsae* is demonstrated by the colour of the points, which is described in the legend at the bottom.



529 **Fig. 4. Differentially regulated genes of microsporidia.** (A-B) Line plots of the FPKM values
530 with confidence intervals, calculated for the differentially regulated genes (FDR<0.05) in *N.*
531 *parisii* (A) and *N. ausubeli* (B) at 10 hours post infection. * indicates ribosomal protein-
532 encoding genes. The aggregate sample value of the two replicates is noted as a black dot
533 within the confidence interval, while the value for each replicate is shown as a dot of matching
534 colour along the confidence interval. Error bars of each gene show the variability of the
535 distribution of FPKM values.
536
537



538
539
540 **Fig. 5. Transcriptional responses of *Caenorhabditis* hosts to *Nematocida* infection.** (A-
541 D) Venn diagrams of genes in *C. elegans* (A-B) or *C. briggsae* (C-D) that are upregulated (A
542 and C) or downregulated (B and D) when infected by *N. parisii* or *N. ausubeli*. npa (*N. parisii)*
543 and nau (*N. ausubeli*).
544
545

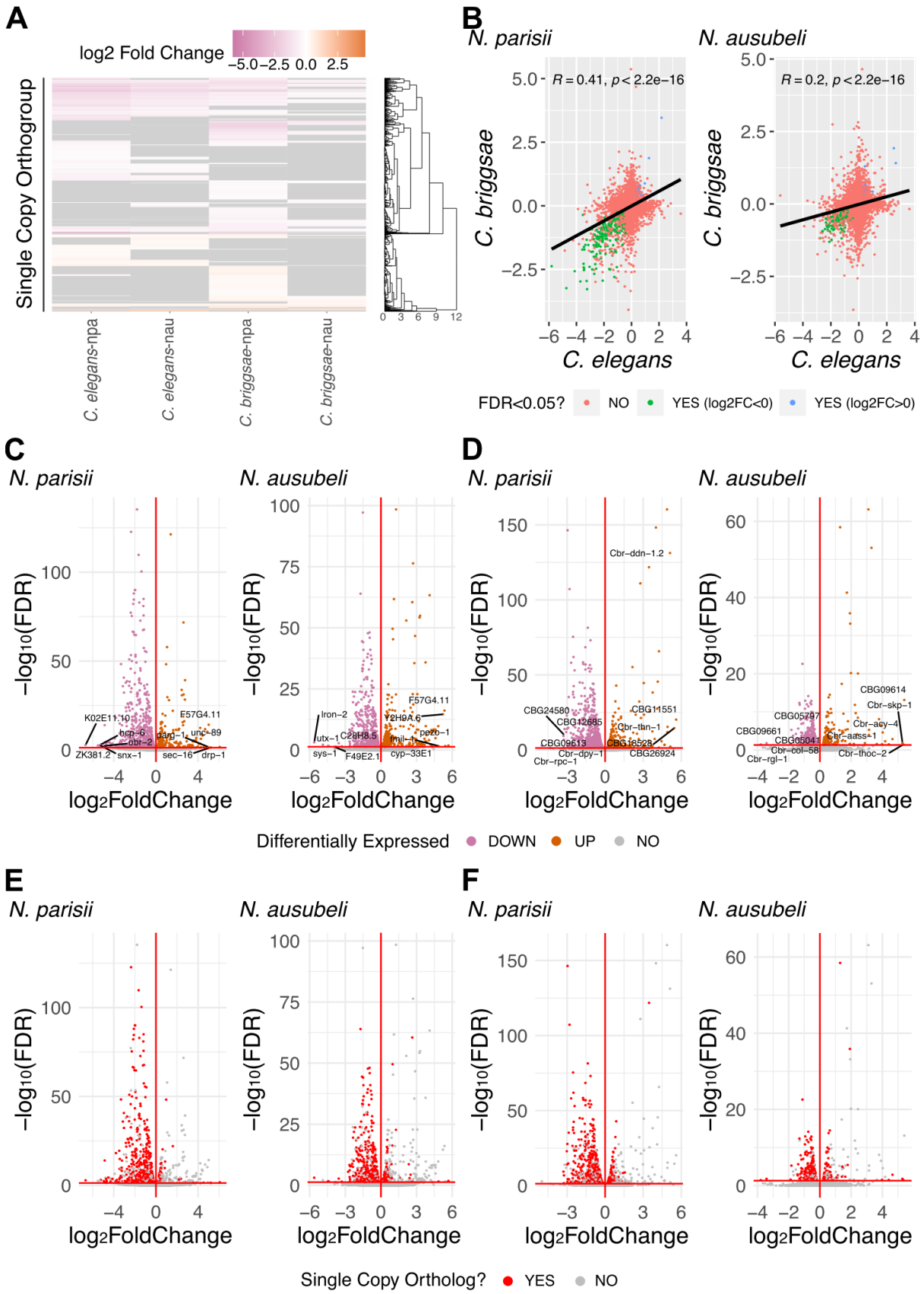


Fig. 6. Transcriptional responses of single copy orthologs in *Caenorhabditis* hosts to *Nematocida* infection. (A) Heatmap of transcriptional profiles of single-copy orthologs in *C. elegans* and *C. briggsae* after infection by *N. parisii* or *N. ausubeli*. Rows represent genes clustered hierarchically. Scale is of differential regulation of infected compared to uninfected

samples. **(B)** Scatterplot of log2 fold change values of single-copy orthologs expressed in *C. elegans* and *C. briggsae*. Each point represents a single-copy ortholog between the two host species. Pearson correlation values and p-values are indicated on top left of each plot. The ratio of log2 fold change values for each single copy ortholog between *C. elegans* and *C. briggsae* is demonstrated by the colour of the points. (C-D) Volcano plots of genes expressed in *Nematocida* infected *C. elegans* (C) and *C. briggsae* (D). Upregulated genes are represented with orange points; downregulated genes are represented with pink points. The top 5 upregulated and downregulated genes are labelled. (E-F) Volcano plots of genes expressed in *Nematocida* infected *C. elegans* (E) and *C. briggsae* (F). Single-copy orthologs are represented with red points and non-single copy orthologs represented with grey points. npa (*N. parisii)* and nau (*N. ausubeli*).

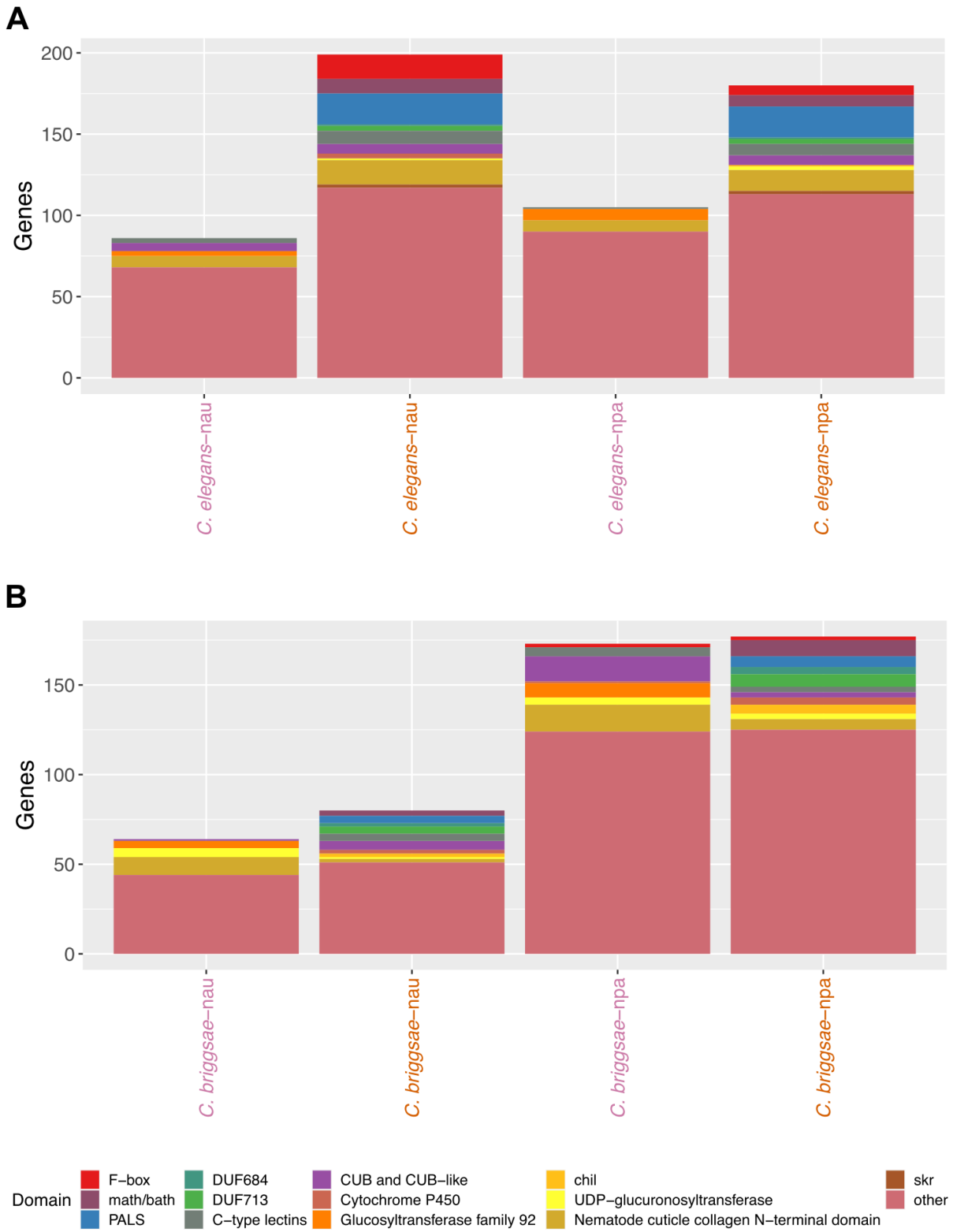


Fig. 7. Transcriptional responses of non-single copy orthologs in *Caenorhabditis* hosts to *Nematocida* infection. Domain enrichment analysis of significantly upregulated and downregulated non-single copy ortholog genes in *C. elegans* (A) and *C. briggsae* (B). npa (*N. parisii)* and nau (*N. ausubeli*). Upregulated samples are represented with orange text; downregulated samples are represented with pink text.

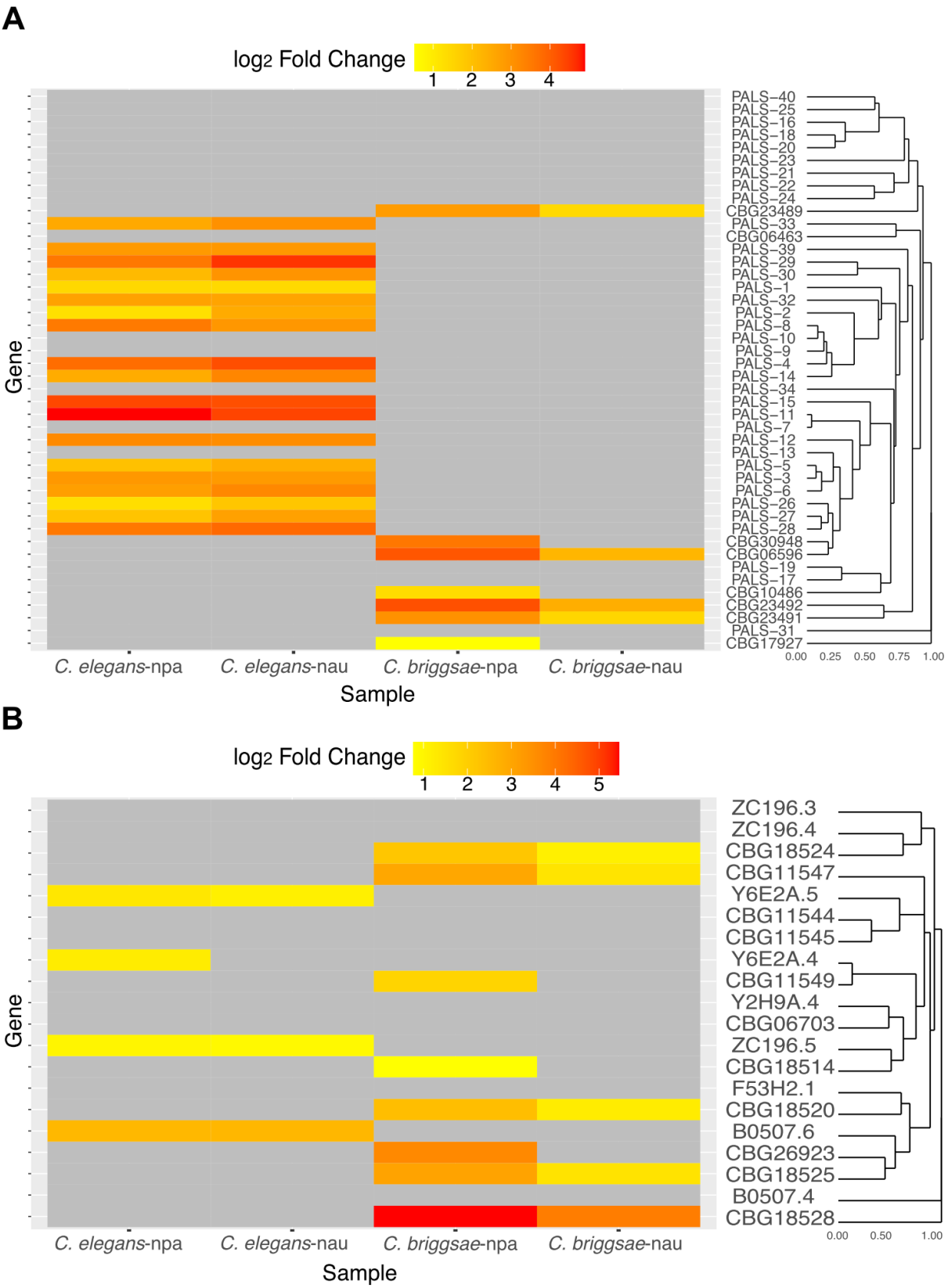
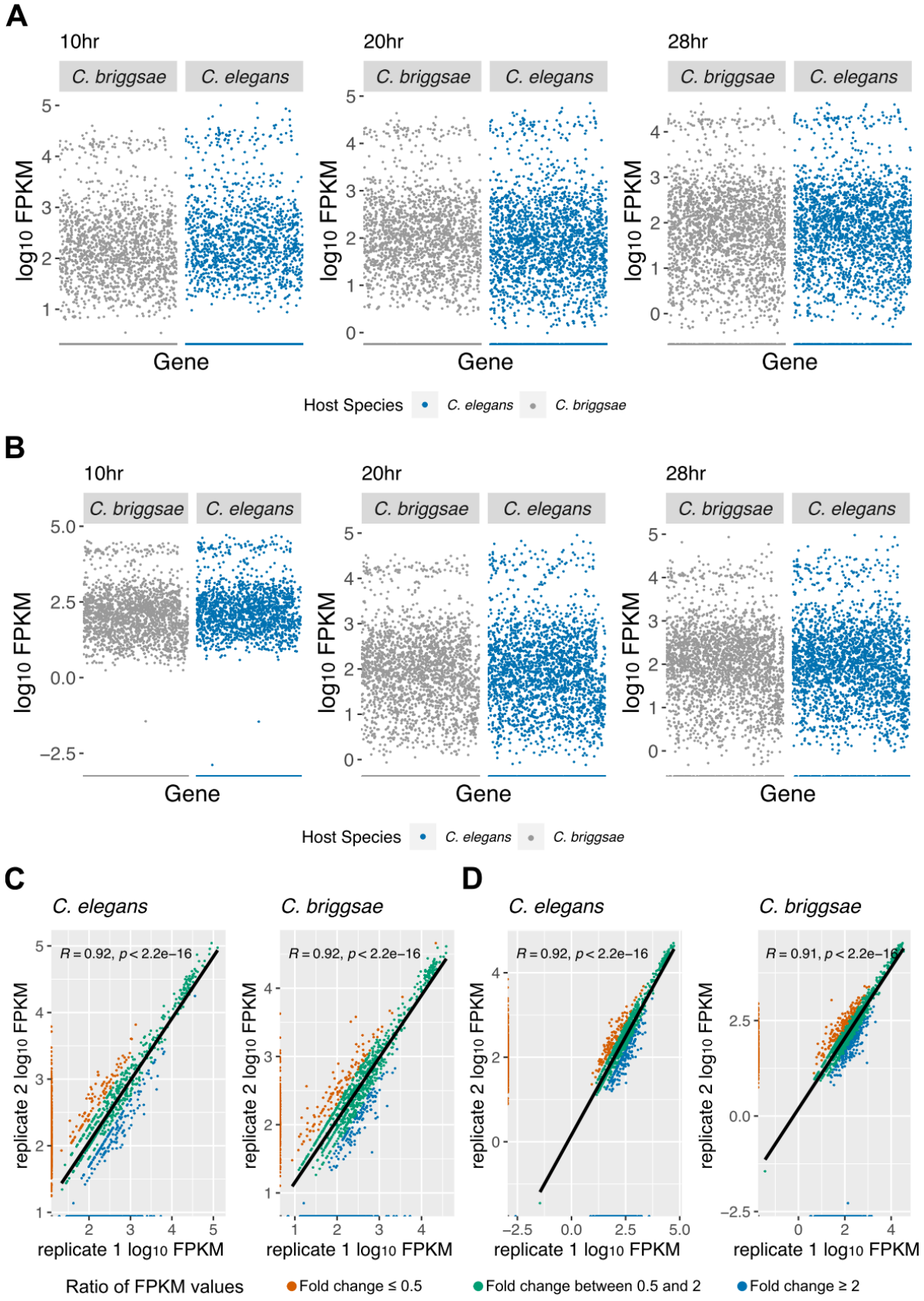


Fig. 8. Conserved response of *Caenorhabditis* hosts to *Nematocida* infection. (A-B) Heatmap of transcriptional profiles of differentially regulated genes (FDR<0.05) in the PALS (A) and DUF713 (B) gene families. Scale is of differential regulation of infected compared to uninfected samples. npa (*N. parisii)* and nau (*N. ausubeli*).

578



579 580 **Fig. S1. Transcriptional response of *Nematocida* species in *Caenorhabditis* hosts.** *(A-*
581 *B)* Scatterplot of *N. parisii* (*A*) and *N. ausubeli* (*B*) log10 FPKM values between *C. briggsae*
582 and *C. elegans* replicates at 10 hours. Pearson correlation values and p-values are indicated

583 on the top left of each plot. The ratio of FPKM values for each gene between the replicates is
584 demonstrated by the colour of the points. (*C-D*) Gene expression pattern of *N. parisii (C)* and
585 *N. ausubeli (D)* in *C. elegans* and *C. briggsae*.
586
587

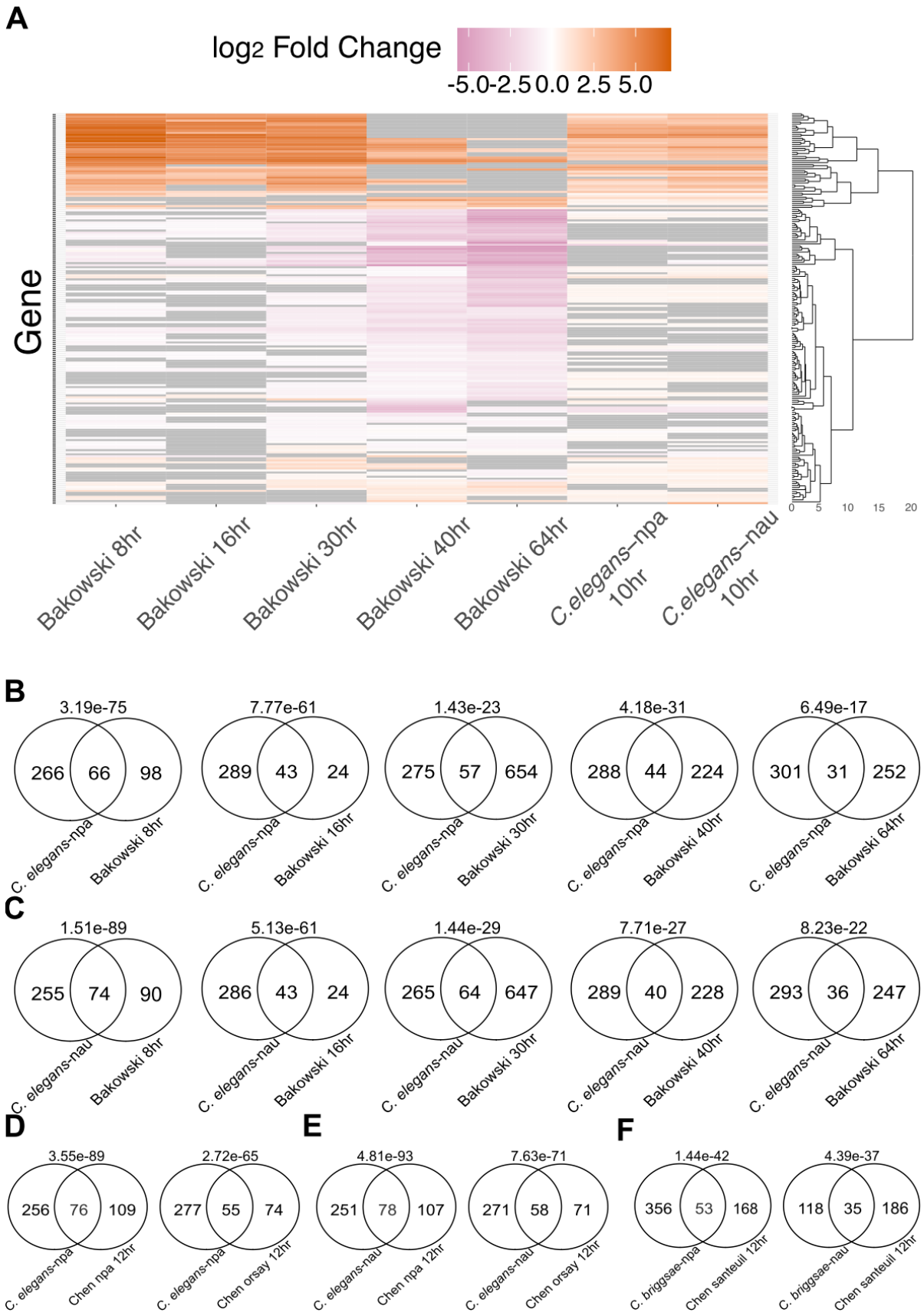


Fig. S2. Data from this study has high overlap with previously published intracellular infection samples. (*A*) Heatmap of transcriptional profiles of differentially regulated genes (FDR<0.05) across five samples from Bakowski *et al*. and our 10hr *N. parisii* and *N. ausubeli* infected *C. elegans* samples. Only genes expressed in at least four out of the eight total

samples are included. *(B-C*) Statistical overlap of significant genes between respective *N. parisii* or *N. ausubeli* infected *C. elegans* samples and Bakowski *et al.* samples across five timepoints. (*D-E*) Statistical overlap of significant genes between *N. parisii (D)* or *N. ausubeli (E)* infected *C. elegans* samples and Chen *et al*. Orsay virus infected *C. elegans* sample at 12 hours. (*F*) Statistical overlap of significant genes between respective *N. parisii* or *N. ausubeli* infected *C. briggsae* sample and Chen *et al*. Santeuil virus infected *C. briggsae* sample at 12 hours. The p-value of each comparison is indicated on top of each Venn diagram. npa (*N. parisii)* and nau (*N. ausubeli*).

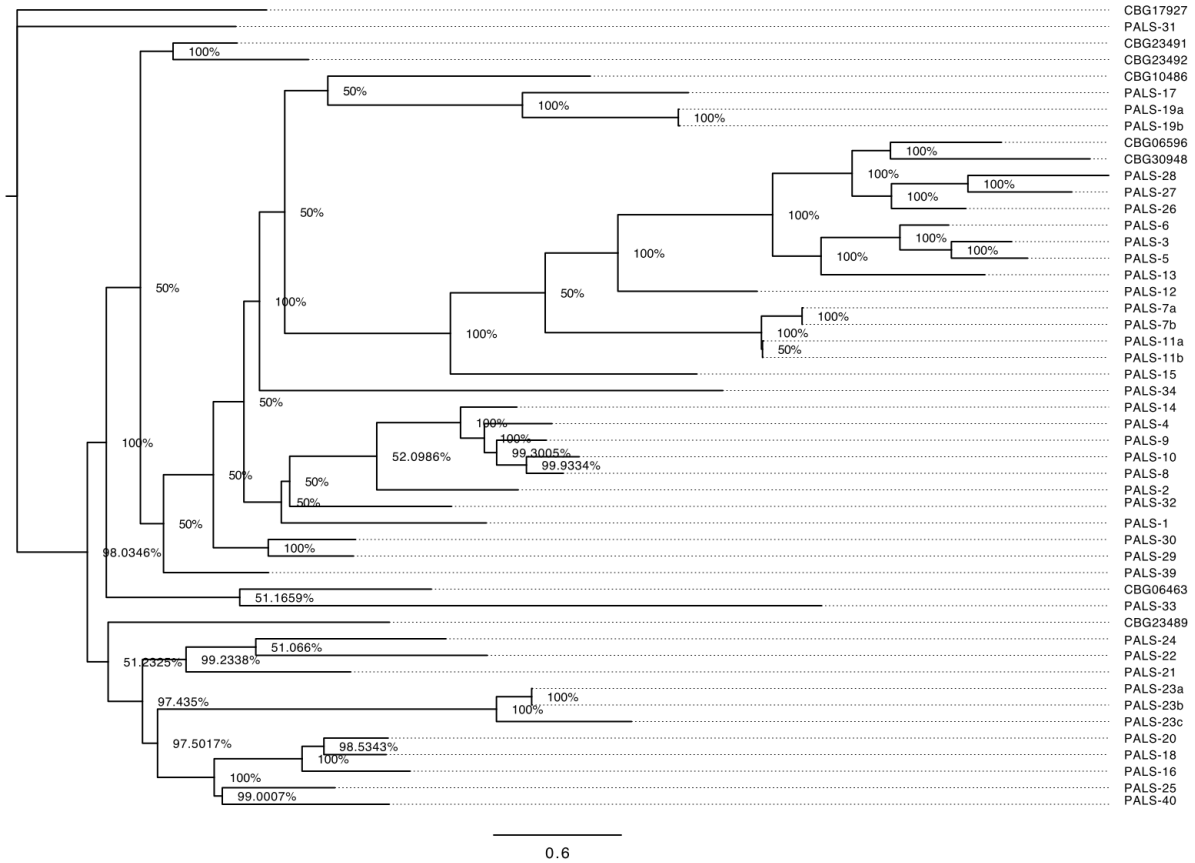


Fig. S3. Phylogram of *C. elegans* and *C. briggsae* PALS family genes. Genes with the same name that end with different letters indicate protein isoforms. Node values indicate posterior probabilities for each split as percentage. The scale bar indicates average branch length measured in expected substitutions per site.

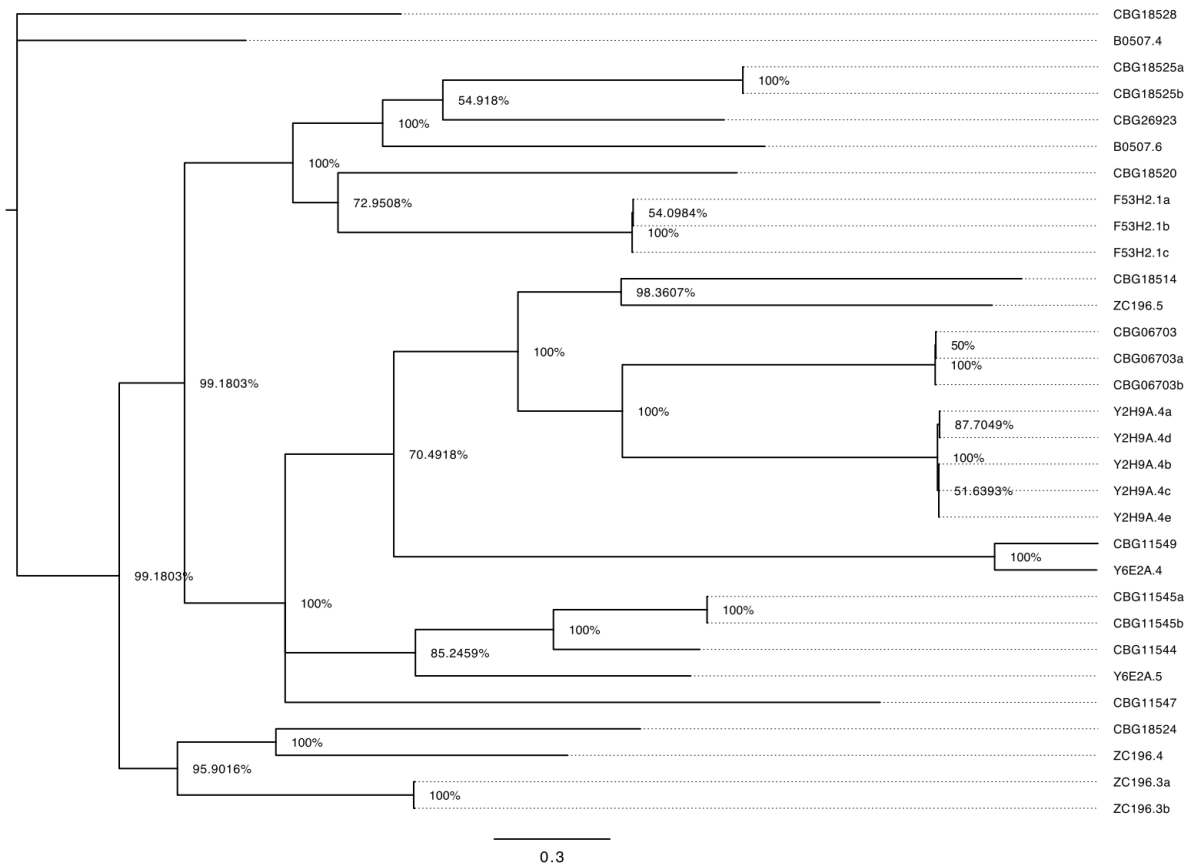


Fig. S4. Phylogram of *C. elegans* and *C. briggsae* DUF713 family genes. Genes with the same name that end with different letters indicate protein isoforms. Node values indicate posterior probabilities for each split as percentage. The scale bar indicates average branch length measured in expected substitutions per site.

Supplementary Table and Materials

Supplementary Table S1. FPKM values of gene expression and differentially expressed microsporidia genes.

Supplementary Table S2. Normalised RNA-seq counts of the *C. elegans* and *C. briggsae* for the transcriptional analysis generated by Alaska.

Supplementary Table S3. Differentially expressed genes of *C. elegans* and *C. briggsae* exposed to the two species of microsporidia generated by Alaska.

Supplementary Table S4. PANTHER GO enrichment analyses. Sheet S1 contains results of the statistical enrichment tests of the 222 significantly upregulated genes shared between *N. parisii* and *N. ausubeli* infected *C. elegans*. Sheet S2 contains results of the statistical enrichment tests of the 267 significantly downregulated genes shared between *N. parisii* and *N. ausubeli* infected *C. briggsae*. Sheet S3 contains results of the statistical enrichment tests of the 415 significantly downregulated genes specifically in *N. parisii* infected *C. briggsae*.

Supplementary Table S5. List of gene set overlaps with previously published infection samples.

640
641
642 **Supplementary Table S6. Gene classes and domains used for enrichment analyses.**
643

Domain/ Gene Family	Databse	Wormbase gene class OR Pfam ID
F-box	Wormbase	fbxa, fbxb, fbxc
MATH (meprin or Traf homology) or BATH (BTB and MATH domain-containing)	Wormbase	math, bath
PALS (protein containing ALS2CR12 signature)	Wormbase	pals
C-type lectins	Wormbase	clec
DUF713	Pfam	PF015218
DUF684	Pfam	PF05075
chil	Wormbase	chil
Nematode cuticle collagen N-terminal domain	Pfam	PF01484
CUB and CUB-like	Pfam	PF00431, PF02408
Cytochrome P450	Wormbase	cyp
Glucosyltransferase family 92	Pfam	PF00201
UDP-glucuronosyltransferase	Wormbase	ugt
skr (Skp1-related)	Wormbase	Skr
Other (Any genes not in the domain or gene family above)	-	-

644
645 **References**

646 Bakowski, M.A. *et al.* (2014) 'Ubiquitin-Mediated Response to Microsporidia and Virus
647 Infection in C. elegans', *PLoS Pathogens*. Edited by D.S. Schneider, 10(6), p. e1004200.
648 doi:10.1371/journal.ppat.1004200.

649 Balla, K.M. *et al.* (2016) 'Cell-to-cell spread of microsporidia causes Caenorhabditis elegans
650 organs to form syncytia', *Nature Microbiology*, 1(11), p. 16144.
651 doi:10.1038/nmicrobiol.2016.144.

652 Barrière, A. and Félix, M.-A. (2005) 'High Local Genetic Diversity and Low Outcrossing Rate in
653 Caenorhabditis elegans Natural Populations', *Current Biology*, 15(13), pp. 1176–1184.
654 doi:10.1016/j.cub.2005.06.022.

655 Bojko, J. *et al.* (2022) 'Microsporidia: a new taxonomic, evolutionary, and ecological
656 synthesis', *Trends in Parasitology*, 0(0). doi:10.1016/j.pt.2022.05.007.

657 Bray, N.L. *et al.* (2016) 'Near-optimal probabilistic RNA-seq quantification', *Nature*
658 *Biotechnology*, 34(5), pp. 525–527. doi:10.1038/nbt.3519.

659 Burton, N.O. *et al.* (2021) 'Intergenerational adaptations to stress are evolutionarily
660 conserved, stress-specific, and have deleterious trade-offs.', *eLife*, 10.
661 doi:10.7554/eLife.73425.

662 Chen, K. *et al.* (2017) 'An evolutionarily conserved transcriptional response to viral infection
663 in Caenorhabditis nematodes', *BMC Genomics*, 18(1), p. 303. doi:10.1186/s12864-017-3689-
664 3.

665 Conesa, A. *et al.* (2016) 'A survey of best practices for RNA-seq data analysis', *Genome*
666 *Biology*, 17(1), p. 13. doi:10.1186/s13059-016-0881-8.

667 Cuomo, C.A. *et al.* (2012) 'Microsporidian genome analysis reveals evolutionary strategies
668 for obligate intracellular growth', *Genome Research*, 22(12), pp. 2478–2488.
669 doi:10.1101/gr.142802.112.

670 Desjardins, C.A. *et al.* (2015) 'Contrasting host-pathogen interactions and genome evolution
671 in two generalist and specialist microsporidian pathogens of mosquitoes.', *Nature*
672 *communications*, 6, p. 7121. doi:10.1038/ncomms8121.

673 Di Tommaso, P. *et al.* (2011) 'T-Coffee: a web server for the multiple sequence alignment of
674 protein and RNA sequences using structural information and homology extension', *Nucleic*
675 *Acids Research*, 39(suppl), pp. W13–W17. doi:10.1093/nar/gkr245.

676 Eddy, S.R. (2011) 'Accelerated Profile HMM Searches', *PLOS Computational Biology*, 7(10), p.
677 e1002195. doi:10.1371/journal.pcbi.1002195.

678 Emms, D.M. and Kelly, S. (2019) 'OrthoFinder: phylogenetic orthology inference for
679 comparative genomics', *Genome Biology*, 20(1), p. 238. doi:10.1186/s13059-019-1832-y.

680 Engelmann, I. *et al.* (2011) 'A comprehensive analysis of gene expression changes provoked
681 by bacterial and fungal infection in C. elegans.', *PloS one*, 6(5), p. e19055.
682 doi:10.1371/journal.pone.0019055.

683 Ewels, P. *et al.* (2016) 'MultiQC: summarize analysis results for multiple tools and samples in
684 a single report', *Bioinformatics*, 32(19), pp. 3047–3048. doi:10.1093/bioinformatics/btw354.

685 Flores, J. *et al.* (2021) 'Human microsporidian pathogen Encephalitozoon intestinalis
686 impinges on enterocyte membrane trafficking and signaling', *Journal of Cell Science*, 134(5),
687 p. jcs253757. doi:10.1242/jcs.253757.

688 Grover, M. *et al.* (2021) 'Infection of C. elegans by Haptoglossa Species Reveals Shared
689 Features in the Host Response to Oomycete Detection', *Frontiers in Cellular and Infection*
690 *Microbiology*, 11, p. 973. doi:10.3389/fcimb.2021.733094.

691 Haag, K.L. *et al.* (2014) 'Evolution of a morphological novelty occurred before genome
692 compaction in a lineage of extreme parasites', *Proceedings of the National Academy of*
693 *Sciences*, 111(43), pp. 15480–15485. doi:10.1073/pnas.1410442111.

694 Huang, D.W., Sherman, Brad T. and Lempicki, R.A. (2009) 'Bioinformatics enrichment tools:
695 paths toward the comprehensive functional analysis of large gene lists', *Nucleic Acids*
696 *Research*, 37(1), pp. 1–13. doi:10.1093/nar/gkn923.

697 Huang, D.W., Sherman, Brad T and Lempicki, R.A. (2009) 'Systematic and integrative analysis
698 of large gene lists using DAVID bioinformatics resources', *Nature Protocols*, 4(1), pp. 44–57.
699 doi:10.1038/nprot.2008.211.

700 Huelsenbeck, J.P. and Ronquist, F. (2001) 'MRBAYES: Bayesian inference of phylogenetic
701 trees', *Bioinformatics*, 17(8), pp. 754–755. doi:10.1093/bioinformatics/17.8.754.

702 James, T.Y. *et al.* (2013) 'Shared Signatures of Parasitism and Phylogenomics Unite
703 Cryptomycota and Microsporidia', *Current Biology*, 23(16), pp. 1548–1553.
704 doi:10.1016/j.cub.2013.06.057.

705 Katinka, M.D. *et al.* (2001) 'Genome sequence and gene compaction of the eukaryote
706 parasite Encephalitozoon cuniculi.', *Nature*, 414(6862), pp. 450–453. doi:10.1038/35106579.

707 L. Goff, C.T. (2017) *Cummerbund: analysis, exploration, manipulation, and visualization of*
708 *cufflinks high-throughput sequencing data.* Bioconductor.
709 doi:10.18129/B9.BIOC.CUMMERBUND.

710 Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature*
711 *Methods*, 9(4), pp. 357–359. doi:10.1038/nmeth.1923.

712 Lansdon, P., Carlson, M. and Ackley, B.D. (2022) 'Wild-type Caenorhabditis elegans isolates
713 exhibit distinct gene expression profiles in response to microbial infection', *BMC Genomics*,
714 23(1), p. 229. doi:10.1186/s12864-022-08455-2.

715 Lažetić, V. *et al.* (2022) 'The transcription factor ZIP-1 promotes resistance to intracellular
716 infection in Caenorhabditis elegans', *Nature Communications*, 13(1), p. 17.
717 doi:10.1038/s41467-021-27621-w.

718 Leyva-Díaz, E. *et al.* (2017) 'Silencing of Repetitive DNA Is Controlled by a Member of an
719 Unusual Caenorhabditis elegans Gene Family', *Genetics*, 207(2), pp. 529–545.
720 doi:10.1534/genetics.117.300134.

721 Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*,
722 25(16), pp. 2078–2079. doi:10.1093/bioinformatics/btp352.

723 Luallen, R.J. *et al.* (2016) 'Discovery of a Natural Microsporidian Pathogen with a Broad
724 Tissue Tropism in Caenorhabditis elegans', *PLoS pathogens*, 12(6), p. e1005724.
725 doi:10.1371/journal.ppat.1005724.

726 Ma, Z. *et al.* (2013) 'Genome-Wide Transcriptional Response of Silkworm (Bombyx mori) to
727 Infection by the Microsporidian Nosema bombycis', *PLOS ONE*, 8(12), p. e84137.
728 doi:10.1371/journal.pone.0084137.

729 Marini, F. and Binder, H. (2019) 'pcaExplorer: an R/Bioconductor package for interacting
730 with RNA-seq principal components', *BMC Bioinformatics*, 20(1), p. 331.
731 doi:10.1186/s12859-019-2879-1.

732 Mi, H. *et al.* (2021) 'PANTHER version 16: a revised family classification, tree-based
733 classification tool, enhancer regions and extensive API', *Nucleic Acids Research*, 49(D1), pp.
734 D394–D403. doi:10.1093/nar/gkaa1106.

735 Midttun, H.L.E. *et al.* (2020) 'Effects of Pseudoloma neurophilia infection on the brain
736 transcriptome in zebrafish (Danio rerio)', *Journal of Fish Diseases*, 43(8), pp. 863–875.
737 doi:10.1111/jfd.13198.

738 Murareanu, B.M. *et al.* (2021) 'Generation of a Microsporidia Species Attribute Database
739 and Analysis of the Extensive Ecological and Phenotypic Diversity of Microsporidia', *mBio*.

740 Nakjang, S. *et al.* (2013) 'Reduction and expansion in microsporidian genome evolution: new
741 insights from comparative genomics.', *Genome biology and evolution*, 5(12), pp. 2285–2303.
742 doi:10.1093/gbe/evt184.

743 Notredame, C., Higgins, D.G. and Heringa, J. (2000) 'T-coffee: a novel method for fast and
744 accurate multiple sequence alignment 1 1Edited by J. Thornton', *Journal of Molecular*
745 *Biology*, 302(1), pp. 205–217. doi:10.1006/jmbi.2000.4042.

746 Pimentel, H. *et al.* (2017) 'Differential analysis of RNA-seq incorporating quantification
747 uncertainty', *Nature Methods*, 14(7), pp. 687–690. doi:10.1038/nmeth.4324.

748 Quandt, C.A. *et al.* (2017) 'The genome of an intranuclear parasite, Paramicrosporidium
749 saccamoebae, reveals alternative adaptations to obligate intracellular parasitism', *eLife*.
750 Edited by A. Rokas, 6, p. e29594. doi:10.7554/eLife.29594.

751 Reddy, K.C. *et al.* (2017) 'An Intracellular Pathogen Response Pathway Promotes
752 Proteostasis in C. elegans', *Current biology: CB*, 27(22), pp. 3544-3553.e5.
753 doi:10.1016/j.cub.2017.10.009.

754 Reddy, K.C. *et al.* (2019) 'Antagonistic paralogs control a switch between growth and
755 pathogen resistance in C. elegans', *PLoS pathogens*, 15(1), p. e1007528.
756 doi:10.1371/journal.ppat.1007528.

757 Reinke, A.W. *et al.* (2017) 'Identification of microsporidia host-exposed proteins reveals a
758 repertoire of rapidly evolving proteins.', *Nature communications*, 8, p. 14023.
759 doi:10.1038/ncomms14023.

760 Ronquist, F. and Huelsenbeck, J.P. (2003) 'MrBayes 3: Bayesian phylogenetic inference
761 under mixed models', *Bioinformatics*, 19(12), pp. 1572–1574.
762 doi:10.1093/bioinformatics/btg180.

763 Sinha, A. *et al.* (2012) 'System Wide Analysis of the Evolution of Innate Immunity in the
764 Nematode Model Species Caenorhabditis elegans and Pristionchus pacificus', *PLOS ONE*,
765 7(9), p. e44255. doi:10.1371/journal.pone.0044255.

766 Sowa, J.N. *et al.* (2019) 'The Caenorhabditis elegans RIG-I Homolog DRH-1 Mediates the
767 Intracellular Pathogen Response upon Viral Infection', *Journal of Virology* [Preprint].
768 doi:10.1128/JVI.01173-19.

769 Stein, L.D. *et al.* (2003) 'The genome sequence of Caenorhabditis briggsae: a platform for
770 comparative genomics.', *PLoS biology*, 1(2), p. E45. doi:10.1371/journal.pbio.0000045.

771 Stentiford, G.D. *et al.* (2016) 'Microsporidia – Emergent Pathogens in the Global Food
772 Chain', *Trends in Parasitology*, 32(4), pp. 336–348. doi:10.1016/j.pt.2015.12.004.

773 Szumowski, S.C. and Troemel, E.R. (2015) 'Microsporidia–host interactions', *Current Opinion*
774 *in Microbiology*, 26, pp. 10–16. doi:10.1016/j.mib.2015.03.006.

775 Tamim El Jarkass, H. and Reinke, A.W. (2020) 'The ins and outs of host-microsporidia
776 interactions during invasion, proliferation and exit', *Cellular Microbiology*, 22(11).
777 doi:10.1111/cmi.13247.

778 Tecle, E. *et al.* (2021) 'The purine nucleoside phosphorylase pnp-1 regulates epithelial cell
779 resistance to infection in C. elegans', *PLOS Pathogens*, 17(4), p. e1009350.
780 doi:10.1371/journal.ppat.1009350.

781 Tecle, E. and Troemel, E.R. (2022) 'Insights from C. elegans into Microsporidia Biology and
782 Host-Pathogen Relationships.', *Experientia supplementum (2012)*, 114, pp. 115–136.
783 doi:10.1007/978-3-030-93306-7_5.

784 Thomas, J.H. (2006) 'Adaptive evolution in two large families of ubiquitin-ligase adapters in
785 nematodes and plants', *Genome Research*, 16(8), pp. 1017–1030. doi:10.1101/gr.5089806.

786 Trapnell, C. *et al.* (2010) 'Transcript assembly and quantification by RNA-Seq reveals
787 unannotated transcripts and isoform switching during cell differentiation', *Nature*
788 *Biotechnology*, 28(5), pp. 511–515. doi:10.1038/nbt.1621.

789 Trapnell, C. *et al.* (2013) 'Differential analysis of gene regulation at transcript resolution with
790 RNA-seq', *Nature Biotechnology*, 31(1), pp. 46–53. doi:10.1038/nbt.2450.

791 Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) 'TopHat: discovering splice junctions with
792 RNA-Seq', *Bioinformatics*, 25(9), pp. 1105–1111. doi:10.1093/bioinformatics/btp120.

793 Troemel, E.R. *et al.* (2008) 'Microsporidia are natural intracellular parasites of the nematode
794 Caenorhabditis elegans.', *PLoS biology*, 6(12), pp. 2736–2752.
795 doi:10.1371/journal.pbio.0060309.

796 Vávra, J. and Lukeš, J. (2013) 'Chapter Four - Microsporidia and "The Art of Living Together"',
797 in Rollinson, D. (ed.) *Advances in Parasitology*. Academic Press, pp. 253–319.
798 doi:10.1016/B978-0-12-407706-5.00004-6.

799 Wadi, L. and Reinke, A.W. (2020) 'Evolution of microsporidia: An extremely successful group
800 of eukaryotic intracellular parasites', *PLOS Pathogens*. Edited by L.J. Knoll, 16(2), p.
801 e1008276. doi:10.1371/journal.ppat.1008276.

802 Wang, L., Wang, S. and Li, W. (2012) 'RSeQC: quality control of RNA-seq experiments',
803 *Bioinformatics*, 28(16), pp. 2184–2185. doi:10.1093/bioinformatics/bts356.

804 Willis, A.R. *et al.* (2021) 'A parental transcriptional response to microsporidia infection
805 induces inherited immunity in offspring', *Science Advances*, 7(19), p. eabf3114.
806 doi:10.1126/sciadv.abf3114.

807 Willis, A.R. and Reinke, A.W. (2022) 'Factors That Determine Microsporidia Infection and
808 Host Specificity.', *Experientia supplementum (2012)*, 114, pp. 91–114. doi:10.1007/978-3-
809 030-93306-7_4.

810 Wong, D. *et al.* (2007) 'Genome-wide investigation reveals pathogen-specific and shared
811 signatures in the response of Caenorhabditis elegansto infection', *Genome Biology*, 8(9), p.
812 R194. doi:10.1186/gb-2007-8-9-r194.

813 Zárate-Potes, A. *et al.* (2020) 'The C. elegans GATA transcription factor elt-2 mediates
814 distinct transcriptional responses and opposite infection outcomes towards different
815 Bacillus thuringiensis strains', *PLOS Pathogens*, 16(9), p. e1008826.
816 doi:10.1371/journal.ppat.1008826.

817 Zhang, G. *et al.* (2016) 'A Large Collection of Novel Nematode-Infecting Microsporidia and
818 Their Diverse Interactions with Caenorhabditis elegans and Other Related Nematodes', *PLOS*
819 *Pathogens*. Edited by J.B. Lok, 12(12), p. e1006093. doi:10.1371/journal.ppat.1006093.

820
821