

RESEARCH ARTICLE

# The Genome of *Nosema* sp. Isolate YNPr: A Comparative Analysis of Genome Evolution within the *Nosema/Vairimorpha* Clade

Jinshan Xu<sup>1</sup>\*, Qiang He<sup>2</sup>, Zhenggang Ma<sup>1</sup>, Tian Li<sup>2</sup>, Xiaoyan Zhang<sup>1</sup>, Bettina A. Debrunner-Vossbrinck<sup>3</sup>, Zeyang Zhou<sup>1,2</sup>, Charles R. Vossbrinck<sup>4</sup>

**1** College of Life Sciences, Chongqing Normal University, Chongqing, China, **2** State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing, China, **3** Department of Math/Science, Gateway Community College, 20 Church Street, New Haven, CT 06510, United States of America, **4** Department of Environmental Science, Connecticut Agricultural Experiment Station, 123 Huntington Street, New Haven, CT 06511, United States of America

\* These authors contributed equally to this work.

\* [xujinshan2003@aliyun.com](mailto:xujinshan2003@aliyun.com)



**OPEN ACCESS**

**Citation:** Xu J, He Q, Ma Z, Li T, Zhang X, Debrunner-Vossbrinck BA, et al. (2016) The Genome of *Nosema* sp. Isolate YNPr: A Comparative Analysis of Genome Evolution within the *Nosema/Vairimorpha* Clade. PLoS ONE 11(9): e0162336. doi:10.1371/journal.pone.0162336

**Editor:** Marc Robinson-Rechavi, University of Lausanne, SWITZERLAND

**Received:** April 16, 2016

**Accepted:** August 22, 2016

**Published:** September 6, 2016

**Copyright:** © 2016 Xu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All the genomic data that used in this study were produced by the third party and can be freely downloaded from not only public Silkworm Pathogen Database (SilkPathDB, <http://silkpathdb.swu.edu.cn/>), but also GenBank database (accession no.PRJNA325422) without any restriction. All other data have been provided as part of Supporting Information in the paper.

**Funding:** This study is supported by National High-tech R&D Program (863 Program, No.2013AA102507), National Natural Science Foundation of China (No.31270138, 31302037,

## Abstract

The microsporidian parasite designated here as *Nosema* sp. Isolate YNPr was isolated from the cabbage butterfly *Pieris rapae* collected in Honghe Prefecture, Yunnan Province, China. The genome was sequenced by Illumina sequencing and compared to those of two related members of the *Nosema/Vairimorpha* clade, *Nosema ceranae* and *Nosema apis*. Based upon assembly statistics, the *Nosema* sp. YNPr genome is 3.36 x 10<sup>6</sup>bp with a G+C content of 23.18% and 2,075 protein coding sequences. An “ACCCTT” motif is present approximately 50-bp upstream of the start codon, as reported from other members of the clade and from *Encephalitozoon cuniculi*, a sister taxon. Comparative small subunit ribosomal DNA (SSU rDNA) analysis as well as genome-wide phylogenetic analysis confirms a closer relationship between *N. ceranae* and *Nosema* sp. YNPr than between the two honeybee parasites *N. ceranae* and *N. apis*. The more closely related *N. ceranae* and *Nosema* sp. YNPr show similarities in a number of structural characteristics such as gene synteny, gene length, gene number, transposon composition and gene reduction. Based on transposable element content of the assemblies, the transposon content of *Nosema* sp. YNPr is 4.8%, that of *N. ceranae* is 3.7%, and that of *N. apis* is 2.5%, with large differences in the types of transposons present among these 3 species. Gene function annotation indicates that the number of genes participating in most metabolic activities is similar in all three species. However, the number of genes in the transcription, general function, and cysteine protease categories is greater in *N. apis* than in the other two species. Our studies further characterize the evolution of the *Nosema/Vairimorpha* clade of microsporidia. These organisms maintain variable but very reduced genomes. We are interested in understanding the effects of genetic drift versus natural selection on genome size in the microsporidia and in developing a testable hypothesis for further studies on the genomic ecology of this group.

31402142, 31001037), Natural Science Foundation Project of Chongqing Science & technology Commission (No.cstc2015cyjA80020).

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Microsporidia are unicellular eukaryotes that are obligate intracellular parasites, entering host cells by injection of the sporoplasm through a polar filament [1,2,3]. Microsporidial infections in insects are thought to be responsible for naturally occurring low to moderate insect mortality and have promise for biological control [4,5,6]. At present, more than 1,500 microsporidial species belonging to 187 genera have been reported [7,8]. The greatest numbers of species have been described from insects and fish, but microsporidia appear to infect all animal species which have been carefully examined. Microsporidia cause economic damage in sericulture [9,10], apiculture [11,12,13] and aquaculture [14,15], and as opportunistic infections are an important consideration in the AIDS epidemic as numerous microsporidial infections have been reported from immunocompromised humans [16,17]. Some groups of microsporidia, such as the Amblyosporidae, have complex life cycles [18,19] and appear to be relatively specific regarding their definitive host [20,21].

*Nosema* sp. YNPr is considered to be a parasite primarily of the "cabbage butterfly" (*Pieris rapae*), an economic pest of cruciferous crops, such as cabbage, rape, cauliflower and broccoli [4,22]. In addition, there are a number of other microsporidial species from the *Nosema/Vairimorpha* clade that have been isolated from *P. rapae* [23,24,25,26].

Based upon comparative small subunit ribosomal DNA (SSU rDNA) and large subunit ribosomal DNA (LSU rDNA) analyses, the genera *Nosema* and *Vairimorpha* are paraphyletic taxa clustering together in what is now referred to as the *Nosema/Vairimorpha* clade [27,28]. There are over 100 reported species from this clade and they appear to be generalist parasites with simple single-host life cycles, able to switch hosts through cross-infection and adaptation [28,29]. It has been known for some time that, in addition to the domesticated silkworm *Bombyx mori*, *Nosema bombycis* can also infect various other lepidopteran insects [29,30,31]. In fact, it appears that *N. bombycis*, the first described microsporidial species, isolated from *B. mori* [32], was an opportunistic infection originating from other Lepidoptera living in or near the mulberry fields. Host-range virulence feeding studies revealed that many of the members of the *Nosema/Vairimorpha* clade seem to be switching hosts relatively rapidly with varying degrees of infection in various tissues making such host range studies an important part of microsporidial ecology [6,33,34].

The genomes of some microsporidia are extremely compact. The genomes of *Encephalitozoon* species (a sister taxon to the *Nosema/Vairimorpha* clade) range in size from 2.3 to 2.9 x 10<sup>6</sup> bp. The genomes of *Encephalitozoon* species encode roughly 2,000 genes with few introns, very short intergenic regions, and no transposable elements [35,36,37,38]. Phylogenetic analysis of the microsporidia shows that microsporidial genomes expand and contract over relatively short evolutionary time [39]. The highly compact nature of the microsporidial genome has been discussed in terms of minimal genes necessary for the survival of an obligate intracellular parasite and the origins of the microsporidia [40,37]. The reduced genome of *Enterocytozoon bieneusi* lacks some of the genes for core carbon metabolism [41] and contains highly compacted, overlapping genes [42]. The microsporidia are considered to have rapidly evolving genes while maintaining a high degree of genome synteny. The "shrinking" of the genome has been attributed to the obligate intracellular life style which "has permitted the loss of many genes whose functions can be provided by the host cell" and to the loss of introns [43]. Organism complexity is not thought to be related to genome size but perhaps to other factors such as metabolic rate, body size, population size, cell size and nucleus size [44].

The idea that genome size is a limiting factor in the rate of reproduction in prokaryotes is an attractive one, however an analysis of 214 species of bacteria and archaea suggests that in prokaryotes growth optimization is correlated with traits such as codon usage biases and shows no

correlation between genome size and reproductive rate. It has been suggested that generation times may be determined by factors such as environment stability and nutrient availability [45].

Streamlining, the reduction of genome size, has been examined in a number of prokaryotic species. For most species that have been studied, streamlining appears to occur through genetic drift [46]. In thermophilic bacteria, however, it has been shown that the proportion of genomic DNA in intergenic regions decreases with smaller genome size, that there is a correlation between genome size and generation time and that the genome-wide selective constraints (dN/dS measurements) do not decrease with smaller genomes, suggesting that in thermophiles genome reduction is due to natural selection [46]. It has been suggested that cell size, which correlates with genome size, may be the direct target of natural selection; large cells may suffer fitness costs at higher temperatures [46]. Examination of bacterial genome size reduction through experimental selection suggests that genome reduction can occur over a short evolutionary time [47].

Because of their speciose nature and their presence in a wide range of insects, members of the *Nosema/Vairimorpha* clade present a good system in which to study ecological genomics of intracellular parasites. In this study we compare the genome of *Nosema* sp. isolate YNPr from *P. rapae* collected in Honghe Prefecture, Yunnan Province, China with the genomes of two microsporidial honeybee parasites, *Nosema ceranae* and *Nosema apis*. We hypothesize that genome size is a limiting factor in the rate of reproduction of the Microsporidia and that there is an evolutionary tradeoff between having a small genome and reproducing rapidly versus having a larger genome and giving the parasite more genetic options with which to challenge a host.

## Materials and Methods

### Isolation of spores and genomic DNA extraction

Wild-caught cabbage butterflies (*P. rapae*) infected with microsporidia were field collected in Honghe Prefecture, Yunnan Province, China (Fig 1A), and no specific permissions were required for these locations. The wings were removed and the infected *P. rapae* were then disrupted with a glass homogenizer. The homogenates were filtered through three layers of cheesecloth and centrifuged at 10,000 x g for 30 seconds. The pellets were re-suspended in PBS buffer and purified by centrifugation at 1,500 x g for 30 min on a Percoll gradient. The spore band was collected and washed several times with sterile water and checked for purity with a phase contrast microscope (Fig 1B).

Approximately  $1 \times 10^9$  spores of *Nosema* sp. YNPr were extracted directly using the cetyl trimethylammonium bromide (CTAB) method [48]. The concentration and purity of the DNA were determined spectrophotometrically based on absorption readings and ratios at 260 nm and 280 nm by using a NanoDrop spectrophotometer ND-1000

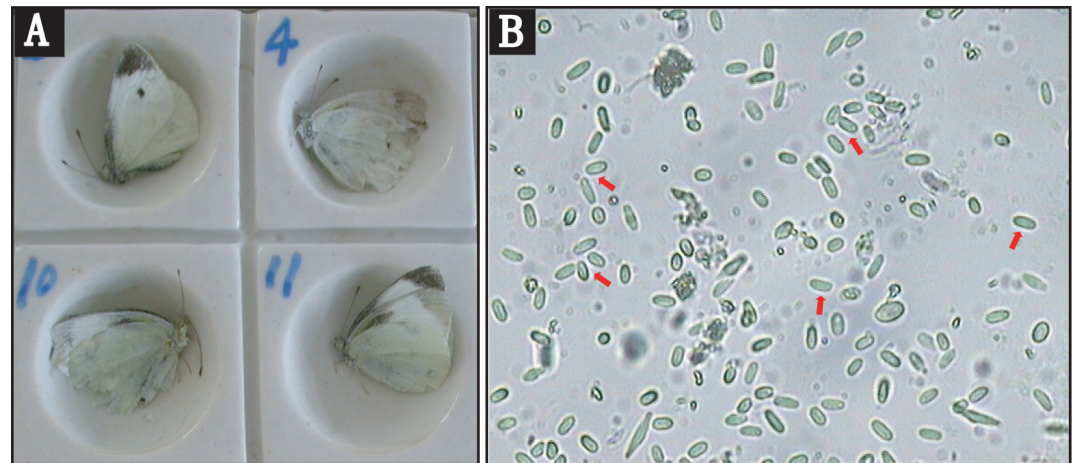
### Ribosomal DNA sequencing

PCR amplification was performed under the following conditions:

Samples were heated to 94°C for 5 min to denature the DNA followed by 30 cycles of: 1 min of denaturation at 94°C, 30 s of annealing at 55–58°C, and 2 min of extension at 72°C. A final extension step was carried out at 72°C for 10 min. Primers used are as follows:

SSU rRNA: 5'-CACCAGGTTGATTCTGCC-3'  
5'-TTATGATCCTGCTAATGGTTC-3'  
ITS: 5'-TGAATGTGTCCCTGTTCTTTG-3'  
5'-GTTAGTTTCTTTTCCTCC-3'

PCR products were run on agarose gels, excised and purified using a gel extraction kit (E.Z. N.A. ® Gel Extraction Kit) following the manufacturer's instructions. The purified PCR products were then cloned into a pMD19-T vector using a TA Cloning Kit (TaKaRa Biotechnology,



**Fig 1. Images of host and parasite.** A. Micrograph of the adult of *Pieris rapae* Linne. B. Spores of *Nosema* sp. YNPr at 400X.

doi:10.1371/journal.pone.0162336.g001

Dalian, China) and bacterial colonies were sent to Beijing Genomics Institute (BGI) Shenzhen, China for Sanger sequencing.

### Comparative rDNA analysis

The SSU rDNA and ITS rDNA sequences from *Nosema* sp. YNPr were compared to those downloaded from GenBank. Sequence alignments were performed by MUSCLE [49]. Phylogenetic trees were constructed using neighbor-joining (NJ) analysis with MEGA version 5 [50]. The Maximum Composite Likelihood model was used and bootstrap support was evaluated based on 1,000 replicates.

### Genome sequencing and assembly

A genomic DNA library was prepared for Illumina HiSeq2500 sequencing following the manufacturer's instructions (Illumina). The genomic DNA was fragmented by nebulization with compressed nitrogen gas at 32psi for 9 minutes. Any overhangs were converted to blunt ends using T4 DNA polymerase and Klenow polymerase, after which an adenine nucleotide was added to the ends of double-stranded DNA using Klenow Exo-Minus polymerase (Qiagen). DNA adaptors with a single "T" base overhang at the 3' end were ligated to the above products. The resulting DNA was then separated on a 2% agarose gel and fragments of approximately 800 bp were excised from the gel and purified (Qiagen Gel Extraction Kit). The adapter-modified DNA fragments were amplified using Illumina primers 1.1 and 2.1. The concentration of the DNA library was determined by measuring the absorbance at 260nm and the DNA was then sequenced (BGI). The paired-end reads were processed by removing adaptor and low quality ( $Q < 30$ ) sequences. As a result, 3,154,248 high-quality reads were obtained for a total of 1.57 million base pairs. Each read was about 100 bp in length. The paired-end reads were assembled de novo using the Velvet1.2.10 assembly program [51], with the standard option as following: insert size of  $800 \pm 160$  bp and a Hash length of 31. The resulting scaffolds were further processed with GapFiller1.11 [52] using the default settings.

### Gene prediction and annotation

The *Nosema* sp. YNPr protein-coding genes were annotated with Glimmer 3.0 software using the lower eukaryote settings [53,54]. We identified 2075 ORFs. To avoid the over-prediction of



small genes, the ORFs were re-annotated by searching for transcriptional signals (CCC or GGG-like motifs) within 50 nucleotides upstream of the first or successive downstream AUG codons within the ORF [48,55]. We identified 1425 CDS by this method. The remaining 650 ORFs were further examined by searching for AT-rich regions (AT content  $\geq 80\%$ ) within 50 nucleotides upstream of the first or successive AUG codons [55] within the ORFs. We identified a further 275 CDS using this method.

The MEME 3.0 software program was used for motif exhibition of putative transcription regulators by searching for over-represented motifs in the 50-bp upstream region of the start codon of protein-coding genes. Gene annotation was accomplished utilizing NCBI BLASTP against the GenBank non-redundant database (nr), Swiss-Prot database with a cut-off e-value of  $1e-6$ . Gene ontology classification of the three *Nosema* species was performed using InterProScan5 software and then visualized through on-line WEGO tools (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>).

Comparison of genes based upon function among *N. ceranae*, *N. apis* and *Nosema* sp. YNPr was accomplished using the Clusters of Orthologous Groups (COG) protein database (<http://www.ncbi.nlm.nih.gov/COG/>). BLAST searches against the local COG database were performed with a cut-off e-value of  $1e-6$ , and the best hit which contained the protein identity was used to assign the functional categories of COG based on the list of COG annotations. Similarly, proteases were identified using a BLAST search of all predicted genes against the MEROPS database ([http://merops.sanger.ac.uk/cgi-bin/batch\\_blast](http://merops.sanger.ac.uk/cgi-bin/batch_blast)) [56]. Genes with e-values less than  $1e-6$  were further verified as proteases by searching against the Genbank database. The best hits in the MEROPS database were used to assign the protease classes. BLASTCLUST 2.2 was used to identify homologous genes among *Nosema* sp. YNPr and three other species (*N. ceranae*, *N. apis* and *E. cuniculi*) with 30% identity and 50% coverage. Differences in gene length among homologous genes and the number of unique genes in each of the species were calculated using custom PERL scripts written in our lab.

The final genome and predicted protein-coding genes for *Nosema* sp. YNPr are deposited in NCBI (Project PRJNA325422) and the Silkworm Pathogen Database (SilkPathDB, <http://silkpathdb.swu.edu.cn/>). The Illumina data generated for *Nosema* sp. YNPr is available in the NCBI SRA (accession number SRR3673305).

## Transposon and signal peptide analysis

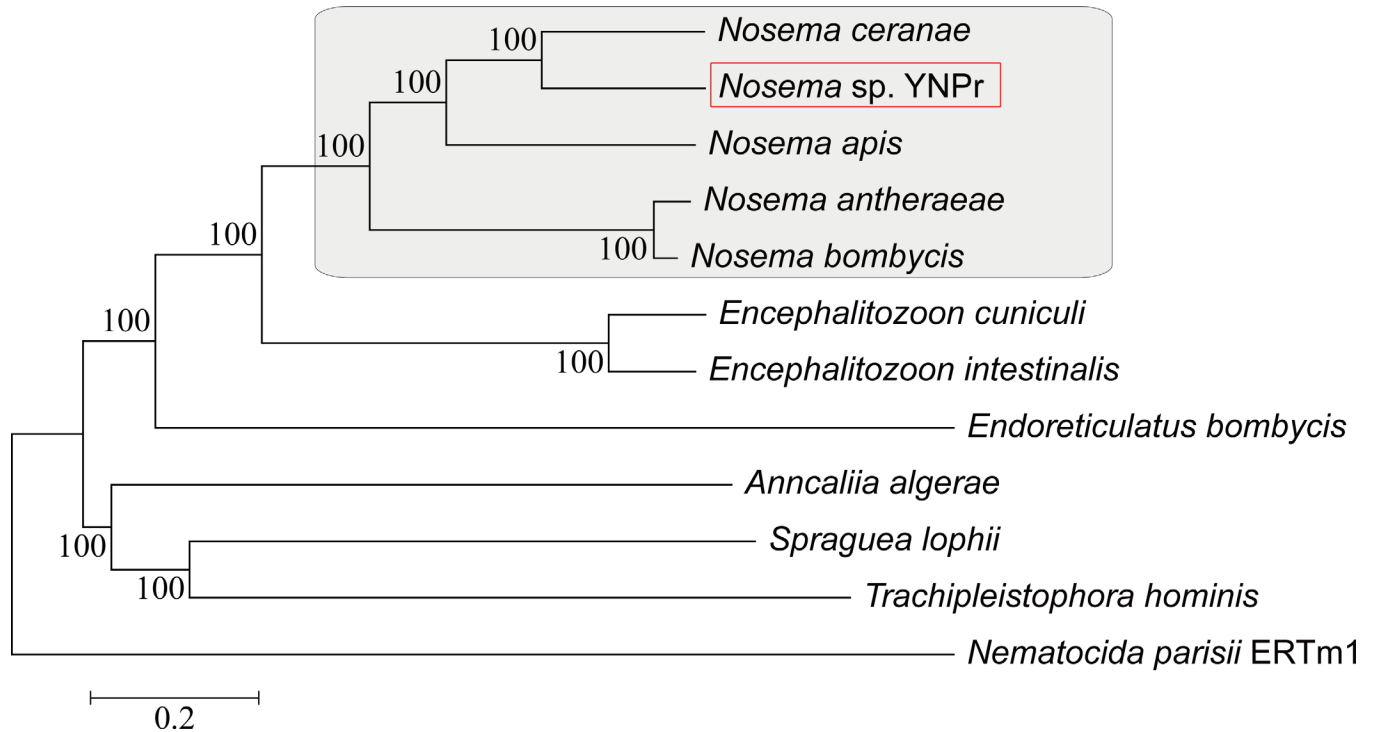
The assembled genomes of *Nosema* sp. YNPr, *N. ceranae*, *N. apis* and *N. bombycis* were checked for interspersed repeats and low complexity DNA sequences using RepeatMasker 4.05 (<http://www.repeatmasker.org>) [57]. The transposons were identified and their lengths were calculated. The conserved reverse transcriptase domain sequences (RVT) from LTR retrotransposons were obtained through the NCBI Conserved Domains search site (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) to predict the conserved domains.

Signal peptide sequences of genes for each microsporidia species were predicted by SignalP 4.1 under the default D-cutoff values (<http://www.cbs.dtu.dk/services/SignalP/>).

## Results

### Phylogenetic analysis

Neighbor Joining analysis using homologous genes shows a closer relationship between *N. ceranae* and *Nosema* sp. YNPr than between *N. ceranae* and *N. apis* (both honeybee parasites) and a close relationship between the two silkworm parasites *N. antheraeae* and *N. bombycis* (Fig 2). The same phylogenetic relationships were obtained using the SSU rDNA and LSU rDNA genes only (S1A and S1B Fig). The arrangement of the functional ribosomal RNA



**Fig 2. The evolution of microsporidia at the genome level.** The maximum likelihood phylogeny of 12 microsporidia species based on a concatenation of common orthologous genes.

doi:10.1371/journal.pone.0162336.g002

operon from *Nosema* sp. YNPr (5'-SSU-ITS-LSU-3') is the same as that found in most members of the *Nosema/Vairimorpha* clade [58,59,60]. The ITS sequence of *Nosema* sp. YNPr from Yunnan Province is identical to that of a previously reported isolate, *Nosema* sp. MPr from the same host species (*P. rapae*) but from Jiangsu province (S1A Fig) [61]. It is clear, as shown in S1A Fig, that genome size changes quite rapidly even between closely related species.

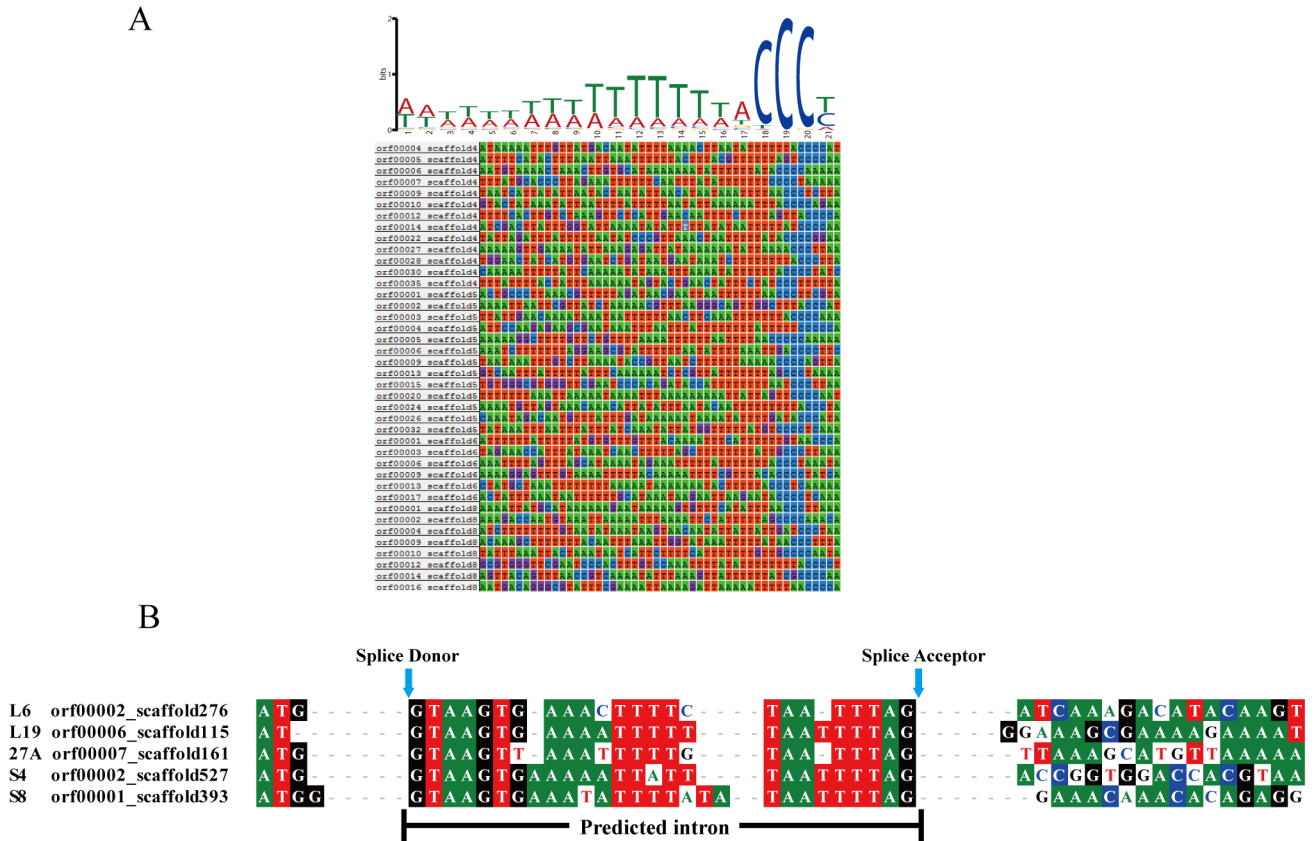
### Genomic architecture of *Nosema* sp. YNPr

Sequencing and assembly statistics are summarized in Table 1. The random genomic *Nosema* sp. YNPr library of 800 bp inserts yielded 3,154,248 reads and was assembled into 462 scaffolds. The combined scaffold length was 3.36 Mb. The transposable element (TE) content of the

**Table 1. Genome assembly statistics for *Nosema* sp. YNPr.**

General characteristics	<i>Nosema</i> sp. YNPr
Clean reads	3,154,248
Assembled (Mb)	3.36
Genome coverage	90X
G+C content (%)	23.19
Total number of scaffold	462
Length of scaffold (bp)	654~45,514
Scaffold N50 (bp)	12,222
No. of CDS	2,075
Mean CDS length (bp)	969
Mean size of scaffold (bp)	7,874

doi:10.1371/journal.pone.0162336.t001



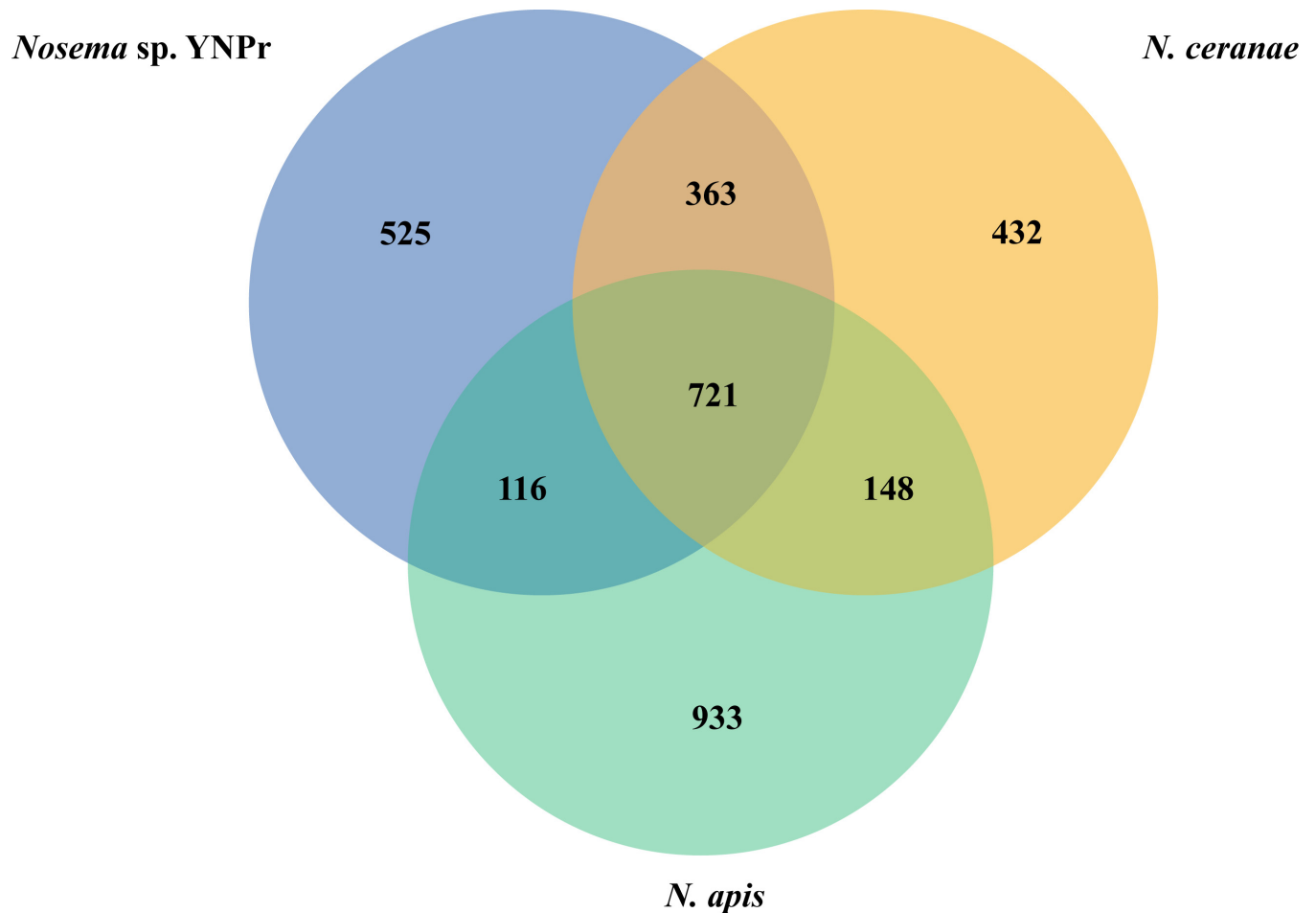
**Fig 3. Noncoding regions of selected *Nosema* sp. YNPr genes.** A. Putative regulatory C-rich motif in the region upstream of the start codon in *Nosema* sp. YNPr. B. Comparison of the regions flanking the introns for five *Nosema* sp. YNPr genes.

doi:10.1371/journal.pone.0162336.g003

assembly was used to approximate the total TE content of the genome and the inferred genome size is contingent on the accuracy of this assumption. The N50 of the *Nosema* sp. YNPr genome is 12,222 bp; in comparison the N50 of *N. apis* is 24,309 bp and that of *N. ceranae* is 2,902 bp. The longest scaffold size is 45,514 bp and the mean scaffold length is 7,874 bp. The mean sequence coverage of scaffolds was 90 X. The G+C content of the *Nosema* sp. YNPr genome is 23.19%, which is lower than those of *N. ceranae* (26%), *N. bombycis* (31%) and *N. antheraeae* (28%) but higher than that of *N. apis* (18.78%) [54,55,56]. The protein coding regions have a significantly higher G+C content (25.53%) than does the overall genome.

We identified 2075 ORFs in the *Nosema* sp. YNPr genome, with a mean length of 969 bp (Table 1). Based on our analysis there are 1,425 ORFs containing the CCC or GGG motifs in the genome, and an additional 275 CDS are predicted based on an AT content >80% in the region 50 nucleotides upstream of an AUG codon within the ORF. (S1 Table, S2 Fig). All of the genes involved in core carbon metabolic pathways previously published from microsporidial genomes have been identified in *Nosema* sp. YNPr. The inability to identify some of the hypothetical genes using BLAST can be explained either by the highly divergent nature of the microsporidia resulting in low similarity with known genes or as genes specific to the microsporidia.

Results of the analysis for the ACCCTT motif approximately 50 bp upstream of the start codon in *Nosema* sp. YNPr is shown in Fig 3A. This motif is conserved in *N. ceranae*, *N. apis* and *E. cuniculi* and was present in 60% of the predicted genes for *Nosema* sp. YNPr [62,63,35]. A search for homology with the genes of *N. ceranae* identified five intron-containing ribosomal



**Fig 4. Venn diagram comparing the genes of three species of microsporidia.** The number of homologous genes and species-specific genes among *Nosema* sp. YNPr, *Nosema ceranae* and *Nosema apis* are shown.

doi:10.1371/journal.pone.0162336.g004

protein genes in *Nosema* sp. YNPr. They all contain **GTAAGT** at the donor site and **TTAG** at the acceptor site (Fig 3B). Two of the proteins (*L19*, *S4*) have orthologous intron-containing genes in *N. apis* and *N. bombycis* while homologues of the other three proteins (*L6*, *27A*, *S8*) do not contain introns in *N. apis* and *N. bombycis*.

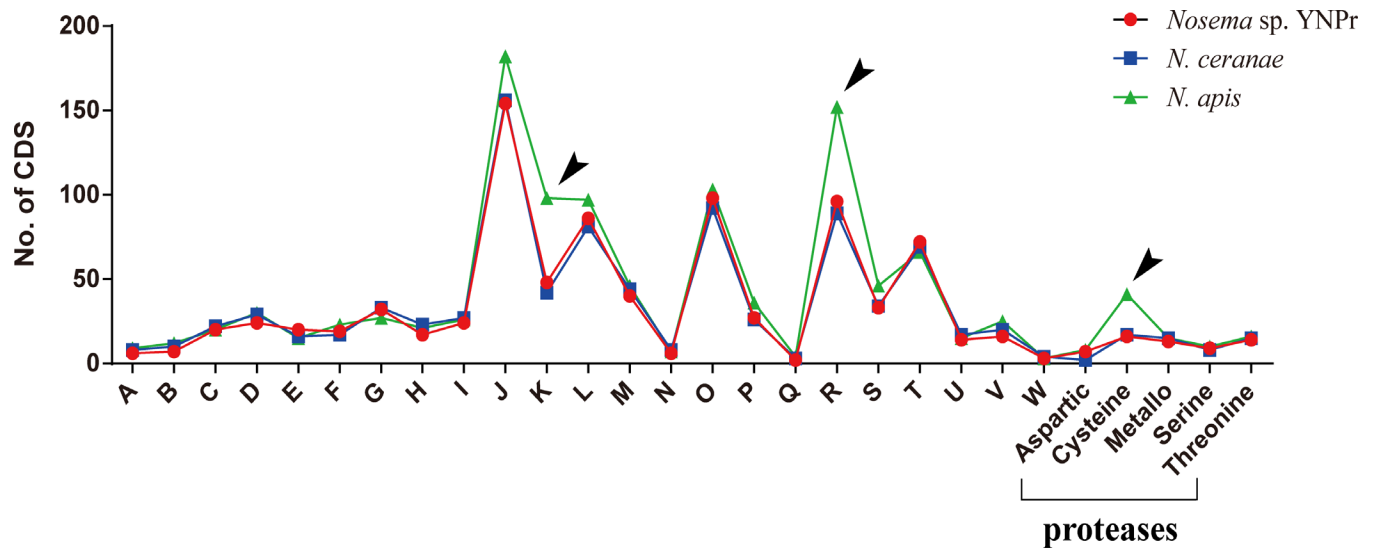
### Comparison of functional genes

Comparison of genes homologous among *Nosema* sp. YNPr, *N. ceranae* and *N. apis* revealed that there are 721 genes shared among all three species, 525 genes that are *Nosema* sp. YNPr-specific, 432 that are *N. ceranae*-specific, and 933 that are *N. apis*-specific (Fig 4). Results of the COG protein database analysis indicate that the various cellular functions are similar in almost all categories except for Transcription (K) and General function prediction (R) which are substantially more numerous in *N. apis* than in the other two species (Fig 5). Cysteine proteases are also more abundant in *N. apis* than in *N. ceranae* and *Nosema* sp. YNPr (Fig 5).

### Variation in Genome Composition

**Decrease in gene length.** We compared the lengths of 1084 genes common to *Nosema* sp. YNPr and *N. ceranae* and found that 75% of the genes were shorter in *Nosema* sp. YNPr (Fig





**Fig 5. Comparison of coding sequence (CDS) numbers based on COG annotation and protease class among *Nosema* sp. YNPr, *N. ceranae* and *N. apis*.** Black arrows indicate gene function categories for which *Nosema* sp. YNPr and *N. ceranae* differ in number of coding sequences from *N. apis*. A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures.

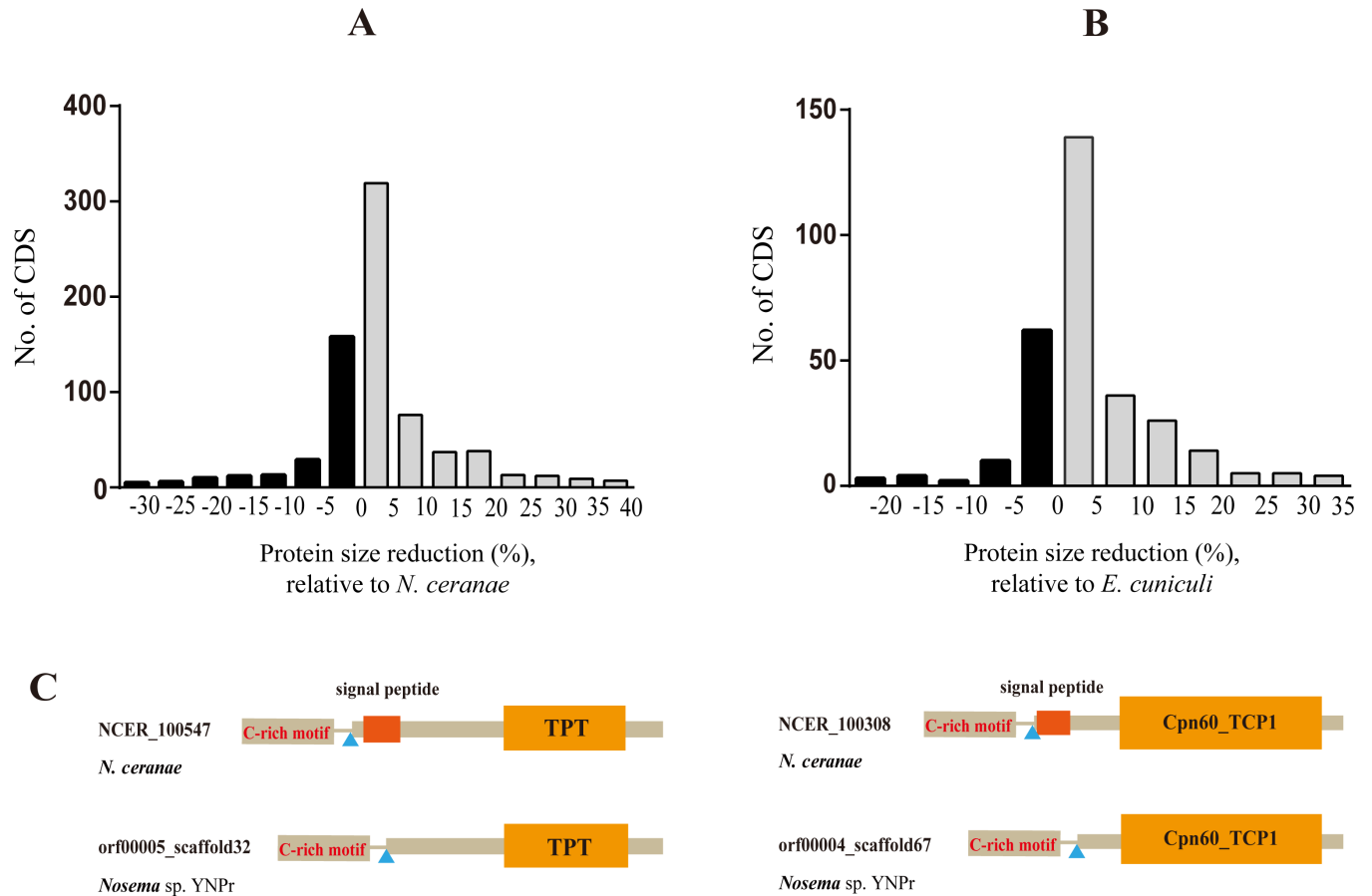
doi:10.1371/journal.pone.0162336.g005

6A). The total coding region of the genome was 364,065 amino acids in *Nosema* sp. YNPr and 368,901 amino acids in *N. ceranae*. Similarly, when 340 genes common to *Nosema* sp. YNPr and *E. cuniculi*, the closest sister taxon to the *Nosema/Vairimorpha* clade, were compared, it was found that 74% of the genes were smaller in *Nosema* sp. YNPr than in *E. cuniculi* (Fig 6B). The total coding region of the genome was 126,156 amino acids in *Nosema* sp. YNPr and 131,659 amino acids in *E. cuniculi*. An analysis of the common genes that are shorter in *Nosema* sp. YNPr than in *N. ceranae* shows a loss of signal sequences in *Nosema* sp. YNPr (Fig 6C) rather than a loss of individual amino acids.

**Signal Peptides.** The number of genes with signal peptides varies widely among species in the *Nosema/Vairimorpha* clade (S2 Table). Closely related species show similar numbers and percentages of genes containing signal peptides. *Nosema bombycis* and *Nosema antheraeae* have 431 and 394 genes with signal peptides respectively, while *Nosema* sp. YNPr and *Nosema ceranae* have 109 and 159 respectively. S3 Table compares homologous genes of *Nosema* sp. YNPr and *Nosema ceranae* that contain predicted signal peptides in one or both species. Twenty-three genes in *Nosema* sp. YNPr lack the signal peptide seen in *Nosema ceranae*, while 5 genes in *Nosema ceranae* lack the signal peptide seen in *Nosema* sp. YNPr.

**Decrease in size of intergenic regions.** Syntenic comparisons among *Nosema* sp. YNPr, *N. ceranae* and *N. apis* indicate that genes are more tightly arranged (compacted) in *Nosema* sp. YNPr and *N. ceranae* than in *N. apis* (Fig 7A). The average length of the intergenic regions for the two scaffolds shown is significantly greater in *N. apis* (1438 nucleotides) than in *Nosema* sp. YNPr (357 nucleotides) and *N. ceranae* (607 nucleotides) (Fig 7B).

**Transposon numbers.** A search of the assembled *Nosema* sp. YNPr genome for repetitive DNA yielded a number of transposon types including LTR, Merlin, Tc1/mariner, LINE and Helitron (Table 2). The transposon content is 4.8% in *Nosema* sp. YNPr, 3.7% in *N. ceranae*,



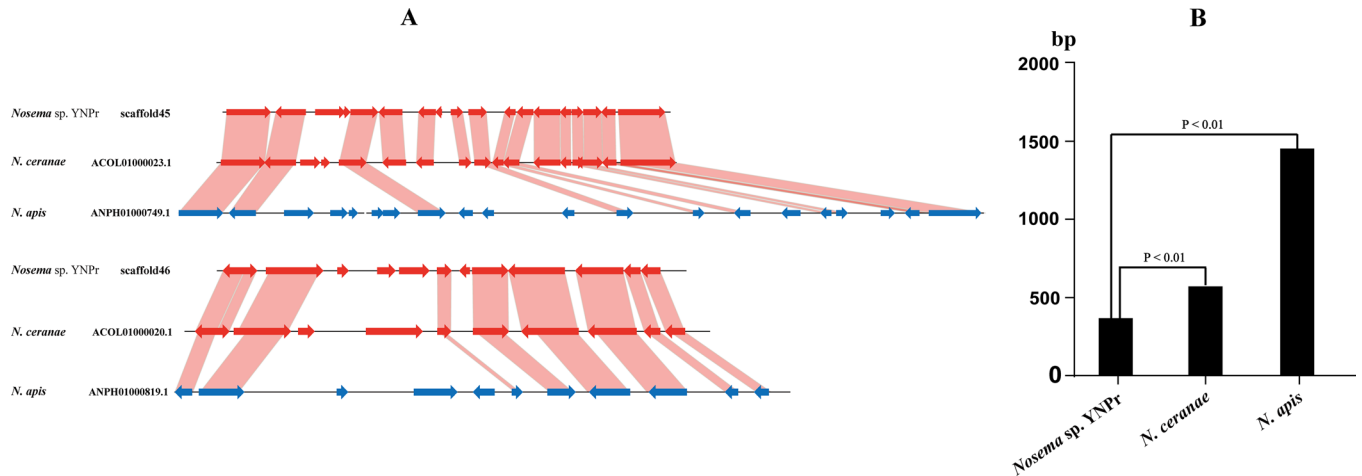
**Fig 6. Comparison of CDS size between *Nosema* sp. YNPr and two other microsporidia.** A. Degrees of reduction in length of *Nosema* sp. YNPr (Np) CDS relative to those of *N. ceranae* (Nc). X-axis value is expressed as percentage:  $100 \times (\text{Nc CDS length} - \text{Np CDS length}) / (\text{Nc CDS length})$ . B. Degrees of reduction in length of *Nosema* sp. YNPr CDS relative to those of *E. cuniculi* (Ec). X-axis value is expressed as percentage:  $100 \times (\text{Ec CDS length} - \text{Np CDS length}) / (\text{Ec protein length})$ . The positive classes representative of shorter *Nosema* sp. YNPr CDS are in grey. C. Two cases showing the signal peptide deletions in *Nosema* sp. YNPr relative to those of *N. ceranae*.

doi:10.1371/journal.pone.0162336.g006

and 2.5% in *N. apis*. *Nosema* sp. YNPr and *N. ceranae* possess all 5 of the above mentioned transposon types but the Merlin, Tc1/mariner and Helitron classes are missing from *N. apis*. For the transposon types shared among these species there is a high variability in copy number. Analysis of the reverse transcriptase of the long terminal repeat (LTR) transposons shows that they cluster into 4 major groups (S3 Fig). The LTR copy number in *N. bombycis*, a sister taxon to the three *Nosema* species analyzed here, is much higher indicating that LTR transposons account for some of the differences in genome size in closely related taxa. Of note also is the absence of Group II LTR sequences in *N. apis*.

## Discussion

Microsporidia have highly reduced genomes [36,39]. These obligate intracellular parasites can use host metabolites for their own cellular processes [41]. Genome compaction in the microsporidia has been studied extensively in terms of biochemical pathways and minimal genome sizes [37,39]. These studies show that the microsporidia have a core set of genes and an expanded set of cell surface transporters which allow them to import metabolic precursors from the host instead of producing these molecules themselves. The highly dynamic nature of



**Fig 7. Syntenic, intergenic and overall size comparisons of *Nosema* sp. YNPr scaffolds with those of *N. ceranae* and *N. apis*.** A. Syntenic and intergenic comparisons of two *Nosema* sp. YNPr scaffolds with those of *N. ceranae* and *N. apis*. B. Comparison of average intergenic lengths in *Nosema* sp. YNPr, *N. ceranae* and *N. apis*.

doi:10.1371/journal.pone.0162336.g007

genome evolution in the microsporidia is discussed in terms of gene compaction, size of intergenic regions, introns and overlapping genes [39].

Microsporidial genomes, though small, change in a dynamic fashion and have been shown in some cases to expand and in others to contract over time (S1 Fig) [39]. The fact that closely related microsporidial species have a number of unique genes (Fig 4) suggests that genes may be gained and lost during the process of host switching. Studies are needed to elucidate the functions of these species-specific genes

Our results show that *Nosema* sp. YNPr has shorter genes (Fig 6), shorter intergenic regions (Fig 7) and fewer transposons than do *N. apis* and *N. ceranae*. Gene size appears to be decreasing through loss of domains, including signal peptides (Fig 6). In a comparison of homologous genes between *Nosema* sp. YNPr and *N. ceranae*, 23 of the *Nosema* sp. YNPr genes lacked the signal peptide present in the *Nosema ceranae* homologue (S3 Table), indicating that the signal peptide content of homologous genes in closely related species can change rapidly. However, the total number of genes (both homologous and unique) containing signal peptides is similar in closely related species (S2 Table).

Phylogenetic analysis shows *Nosema* sp. YNPr and *N. ceranae* to be more closely related than are the two honeybee parasites, *N. ceranae* and *N. apis*. This relationship provides evidence of host switching across the insect orders Lepidoptera and Hymenoptera in the *Nosema/Vairimorpha* clade. A search of the protease database shows that *N. apis* has substantially more cysteine proteases than do either *N. ceranae* or *Nosema* sp. YNPr (Fig 5). Cysteine proteases are among the main proteolytic enzymes found in many protozoan parasites, and cysteine protease inhibitors have been shown to be effective against a variety of protozoans including *Trypanosoma cruzi* [64], *Entamoeba histolytica* [65] and *Plasmodium falciparum* [66].

The transposon makeup of these *Nosema* species varies widely and would seem likely to play a large role in genome evolution. *N. apis* appears to have no Merlin, Tc1/mariner or Helitron transposons, all of which are present in both *Nosema* sp. YNPr and *N. ceranae* (Table 2). These differences in transposon content among closely related species indicate that transposons can move in and out of genomes rapidly over relatively short evolutionary time periods. Determining the roles of these transposons in the adaptation of a microsporidial parasite to its host and in genome expansion and contraction in the microsporidia would be illuminating.

**Table 2. Transposon types in three *Nosema* species.**

Type	<i>Nosema</i> sp. YNPr	<i>N. ceranae</i>	<i>N. apis</i>
LTR	107103	99104	52531
Merlin	829	1165	0
Tc1/mariner	1983	36357	0
LINE	9756	27203	13496
Helitron	1190	453	0
Others	42596	126319	143533
Total	163457(4.8%)	290601(3.7%)	209560(2.5%)

doi:10.1371/journal.pone.0162336.t002

## Conclusion

Microsporidia comprise over 1,300 species and infect hosts from every animal phylum from marine, freshwater and terrestrial habitats. Because of their rich host diversity they are an excellent model system for the study of interactions between obligate single-celled parasites and their hosts at many levels. The *Nosema/Vairimorpha* clade encompasses a wide-ranging group of parasites from a diverse collection of hosts in which host-parasite co-evolutionary principles can be tested. Members of the *Nosema/Vairimorpha* clade have been reported from Lepidoptera, Coleoptera, Hymenoptera and other invertebrates including mites [27,67,68]. From our phylogenetic analyses and those of others [39,69,70] it appears that host switching occurs relatively rapidly over evolutionary time in the microsporidia. We hypothesize that natural selection plays a role in the evolution of the small yet dynamically changing microsporidial genomes and suggest that there may be a trade-off between smaller genomes for rapid reproduction and larger genomes with more genetic options with which to challenge a host. However, in order to examine the role of genetic drift versus natural selection in microsporidial genome evolution it will be necessary to sequence and analyze additional genomes from the *Nosema/Vairimorpha* clade. Analyses would include determining dN/dS ratios, functions of genes unique to closely related species [39] in different hosts, genome size versus host persistence, and the searching for convergence in microsporidia from different genera that infect the same host.

## Supporting Information

**S1 Fig. Phylogenetic analysis for rDNA sequence.** A: Phylogenetic tree of SSU rDNA; B: Phylogenetic tree of LSU rDNA tree; C: Multiple sequence alignments of ITS between *Nosema* sp. YNPr and *Nosema* sp. MPr.

(TIF)

**S2 Fig. GO annotation of *Nosema* sp. YNPr, *N. ceranae* and *Nosema apis* genomes.**

(TIF)

**S3 Fig. The cluster analysis of LTR conservative reverse transcriptase sequences.**

(TIF)

**S1 Table. The annotation of all genes in the *Nosema* sp. YNPr genome.**

(XLS)

**S2 Table. Numbers of genes containing signal peptides in *E. cuniculi* and five members of the *Nosema/Vairimorpha* clade.**

(DOC)

### S3 Table. Prediction of Signal Peptides in homologues from *Nosema* sp. YNPr and *Nosema ceranae*.

(DOC)

## Acknowledgments

This study is supported by National High-tech R&D Program (863 Program, No.2013AA102507), National Natural Science Foundation of China (No.31270138, 31302037, 31402142, 31001037), Natural Science Foundation Project of Chongqing Science&technology Commission (No.cstc2015jcyjA80020).

## Author Contributions

**Conceptualization:** JSX ZYZ CRV.

**Formal analysis:** TL.

**Methodology:** JSX QH.

**Resources:** XYZ.

**Software:** QH ZGM.

**Writing – original draft:** QH JSX CRV.

**Writing – review & editing:** JSX CRV BADV.

## References

1. Lom J, Vavra J. The mode of sporoplasm extrusion in microsporidian spores. *Acta Protozool.* 1963; 1:81–89.
2. Weidner E Ultrastructural study of microsporidian invasion into cells. *Z Parasitenk.* 1972; 40: 227–242. PMID: [4346238](#)
3. Keohane EM, Weiss LM. Characterization and function of the microsporidian polar tube: a review. *Folia Parasitol.* 1998; 45: 117–127. PMID: [9684322](#)
4. Tanada Y. Microbial Control of Insect Pests. *Annu Rev Entomology.* 1959; 4: 277–302.
5. Sweeney AW, Becnel JJ. Potential of microsporidia for the biological control of mosquitoes. *Parasitol Today.* 1991; 7: 217–220. PMID: [15463501](#)
6. Solter LF, Maddox JV. Physiological host specificity of microsporidia as an indicator of ecological host specificity. *J Invertebr Pathol.* 1998; 71: 207–216. PMID: [9538025](#)
7. Sprague V. Systematics of the Microsporidia. In: Bulla LA, Cheng TC, editors. *Comparative Pathobiology*, vol. 2. New York: Plenum Press. 1977.
8. Becnel JJ, Takvorian PM, Cali A. Checklist of available generic names for Microsporidia with type species and type host. In: Weiss LM, Becnel JJ, editors. *Microsporidia: Pathogens of opportunity.* John Wiley & Sons, Inc. 2014. pp. 671–686.
9. Bhat SA, Bashir I, Kamili AF. Microsporidiosis of silkworm, *Bombyx mori* L. (Lepidoptera-Bombycidae): a review. *Afr J Agric Res.* 2009; 4: 1519–1523.
10. Kawarabata T. Biology of microsporidians infecting the silkworm, *Bombyx mori*, in Japan. *J Biotechnol Sericol.* 2003; 72: 1–32.
11. Higes M, Martin R, Meana A. *Nosema ceranae*, a new microsporidian parasite in honeybees in Europe. *J In. Pathol.* 2006; 92: 93–95.
12. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science.* 2007; 318: 283–286. PMID: [17823314](#)
13. Natsopoulou ME, McMahon DP, Doublet V, Bryden J, Paxton RJ. Interspecific competition in honeybee intracellular gut parasites is asymmetric and favours the spread of an emerging infectious disease. *Proceedings of the Royal Society of London B: Biological Sciences.* 2015 Jan 7; 282(1798):20141896.



14. Shaw RW, Kent ML. Fish microsporidia. In: Wittner M, Weiss LM, editors. *The Microsporidia and Microsporidiosis*. ASM Press; 1999. pp. 418–446.
15. Kent ML, Shaw RW, Sanders JL. Microsporidia in fish. In: Weiss LM, Becnel JJ, editors. *Microsporidia: Pathogens of opportunity*. John Wiley & Sons, Inc. 2014. pp. 493–520.
16. Didier ES, Vossbrinck CR, Baker MD, Rogers LB, Bertucci DC, Shaddock JA. Identification and characterization of three *Encephalitozoon cuniculi* strains. *Parasitology*. 1995; 111: 411–421. PMID: [11023405](#)
17. Didier ES. Microsporidiosis: an emerging and opportunistic infection in humans and animals. *Acta Trop*. 2005; 94: 61–76. PMID: [15777637](#)
18. Andreadis TG. Experimental transmission of a microsporidian pathogen from mosquitoes to an alternate copepod host. *Proceedings of the National Academy of Sciences*. 1985; 82 (16): 5574–5577.
19. Sweeney AW, Hazard EI, Graham MF. Intermediate host for an *Amblyospora* sp. (microspora) infecting the mosquito, *Culex annulirostris*. *J Invertebr Pathol*. 1985; 46: 98–102. PMID: [2863312](#)
20. Andreadis TG. Host specificity of *Amblyospora connecticus* (Microsporida: Amblyosporidae), a polymorphic microsporidian parasite of *Aedes cantator* (Diptera: Culicidae). *J Med Entomol*. 1986; 26: 140–145.
21. Andreadis TG. Host range tests with *Edhazardia aedis* (Microsporida: Culicosporidae) against northern Nearctic mosquitoes. *J Invertebr Pathol*. 1994; 64: 46–51. PMID: [7914904](#)
22. Kingsolver JG. Feeding, growth, and the thermal environment of cabbage white caterpillars, *Pieris rapae* L. *Physiol Biochem Zool*. 2000; 73: 621–628. PMID: [11073798](#)
23. Watanabe H. First Report of the Microsporidian Infection of the Cabbageworm, *Pieris rapae crucivora* Boisduval (Lepitoptera: Pieridae) in Japan. *Appl Ent Zool*. 1974; 3: 133–142.
24. Chen D, Shen Z, Zhu F, Guan R, Hou J, Zhang J, et al. Phylogenetic characterization of a microsporidium (*Nosema* sp. MPr) isolated from the *Pieris rapae*. *Parasitol Res*. 2012; 111: 263–269. doi: [10.1007/s00436-012-2829-6](#) PMID: [22314780](#)
25. Choi J, Kim J, Choi Y, Nam SH, Russell J, Kim W, et al. Structure of ribosomal RNA gene and phylogeny of *Nosema* isolates in Korea. *Genes & Genomics*. 2009; 31: 443–450.
26. Choi Y, Lee Y, Cho KS, Lee S, Russell J, Choi J et al. Chimerical nature of the ribosomal RNA gene of a *Nosema* species. *J Invertebr Pathol*. 2011; 107: 86–89. doi: [10.1016/j.jip.2011.02.005](#) PMID: [21376728](#)
27. Vossbrinck CR, Debrunner-Vossbrinck BA. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol*. 2005; 52(1–2): 131–142. PMID: [16004372](#)
28. Kyei-Poku G, Gauthier D, van Frankenhuyzen K. Molecular data and phylogeny of *Nosema* infecting lepidopteran forest defoliators in the genera *Choristoneura* and *Malacosoma*. *Journal of Eukaryot Microbiol*. 2008; 55: 51–58.
29. Hayasaka S, Yonemura N. Infection and development of *Nosema* sp. NIS H5 (Microsporida: Protozoa) in several lepidopteran insects. *JARQ*. 1999; 33: 65–68.
30. Kashkarova L, Khakhanov A. Range of the hosts of the causative agent of pebrine (*Nosema bombycis*) in the mulberry silkworm. *Parazitologija*. 1979; 14: 164–167.
31. Kudo R, DeCoursey J. Experimental infection of *Hyphantria cunea* with *Nosema bombycis*. *The Journal of Parasitology*. 1940; 26: 123–125.
32. Nageli K. Uber die neue Krankheit der Seidenraupe und verwandte Organismen. *Bot Z*. 1857; 15: 760–761.
33. Solter LF, Maddox JV, McManus ML. Host Specificity of Microsporidia (Protista: Microspora) from European Populations of *Lymantria dispar* (Lepidoptera: Lymantriidae) to Indigenous North American Lepidoptera. *J Invertebr Pathol*. 1997; 69(2): 135–150. PMID: [9056464](#)
34. Solter LF, Pilarska DK, Vossbrinck CR. Host specificity of microsporidia pathogenic to forest Lepidoptera. *Biological control*. 2000; 19: 48–56.
35. Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature*. 2001; 414: 450–453. PMID: [11719806](#)
36. Selman M, Pombert JF, Solter L, Farinelli L, Weiss LM, Keeling P, Corradi N Acquisition of an animal gene by microsporidian intracellular parasites. *Curr Biol*. 2011; 21(15):R576–7. doi: [10.1016/j.cub.2011.06.017](#) PMID: [21820617](#)
37. Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nature communications*. 2010; 1: 77. doi: [10.1038/ncomms1082](#) PMID: [20865802](#)

38. Pombert JF, Selman M, Burki F, Bardell FT, Farinelli L, Solter LF, et al. Gain and loss of multiple functionally related, horizontally transferred genes in the reduced genomes of two microsporidian parasites. *PNAS*. 2012; 109: 12638–12643. doi: [10.1073/pnas.1205020109](https://doi.org/10.1073/pnas.1205020109) PMID: [22802648](https://pubmed.ncbi.nlm.nih.gov/22802648/)
39. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol*. 2013; 5: 2285–2303. doi: [10.1093/gbe/evt184](https://doi.org/10.1093/gbe/evt184) PMID: [24259309](https://pubmed.ncbi.nlm.nih.gov/24259309/)
40. Keeling PJ, Fast NM. (2002). Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annual Reviews in Microbiology*, 56(1), 93–116.
41. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM. The reduced genome of the parasitic microsporidian *Enterocytozoon bieneusi* lacks genes for core carbon metabolism. *Genome Biol Evol*. 2010; 2: 304–9. doi: [10.1093/gbe/evq022](https://doi.org/10.1093/gbe/evq022) PMID: [20624735](https://pubmed.ncbi.nlm.nih.gov/20624735/)
42. Williams BAP, Slamovits CH, Patron JJ, Fast NM, Keeling PJ. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *PNAS* (2005) 102:10936–10941. PMID: [16037215](https://pubmed.ncbi.nlm.nih.gov/16037215/)
43. Corradi N, Akiyoshi D E, Morrison HG, Feng X, Weiss LM, Tzipori S, Keeling P J. Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antonosporea locustae* and *Enterocytozoon bieneusi*. *PLoS One*, (2007) 2(12), e1277. PMID: [18060071](https://pubmed.ncbi.nlm.nih.gov/18060071/)
44. Keeling Patrick J, Slamovits Claudio H. "Causes and effects of nuclear genome reduction." *Current opinion in genetics & development* 15.6 (2005: ): 601–608.
45. Vieira-Silva S, Rocha EP. (2010). The systemic imprint of growth and its uses in ecological (meta) genomics. *PLoS Genet*, 6(1), e1000808. doi: [10.1371/journal.pgen.1000808](https://doi.org/10.1371/journal.pgen.1000808) PMID: [20090831](https://pubmed.ncbi.nlm.nih.gov/20090831/)
46. Sabath N, Ferrada E, Barve A, Wagner A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome biology and evolution*. 2013 May 1; 5(5):966–77. doi: [10.1093/gbe/evt050](https://doi.org/10.1093/gbe/evt050) PMID: [23563968](https://pubmed.ncbi.nlm.nih.gov/23563968/)
47. Nilsson AI, Koskiniemi S, Eriksson S, Kugelberg E, Hinton JCD, Andersson DI. (2005). Bacterial genome size reduction by experimental evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34), 12112–12116. PMID: [16099836](https://pubmed.ncbi.nlm.nih.gov/16099836/)
48. He Q, Ma Z, Dang X, Xu J, Zhou Z. Identification, Diversity and Evolution of MITEs in the Genomes of Microsporidian *Nosema* Parasites. *PloS one*. 2015; 10(4): e0123170. doi: [10.1371/journal.pone.0123170](https://doi.org/10.1371/journal.pone.0123170) PMID: [25898273](https://pubmed.ncbi.nlm.nih.gov/25898273/)
49. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*. 2004; 5: 113. PMID: [15318951](https://pubmed.ncbi.nlm.nih.gov/15318951/)
50. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular biology and evolution*. 2007 Aug 1; 24(8): 1596–9. PMID: [17488738](https://pubmed.ncbi.nlm.nih.gov/17488738/)
51. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18: 821–829. doi: [10.1101/gr.074492.107](https://doi.org/10.1101/gr.074492.107) PMID: [18349386](https://pubmed.ncbi.nlm.nih.gov/18349386/)
52. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol*. 2012; 13: R56. doi: [10.1186/gb-2012-13-6-r56](https://doi.org/10.1186/gb-2012-13-6-r56) PMID: [22731987](https://pubmed.ncbi.nlm.nih.gov/22731987/)
53. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res*. 1999; 27: 4636–4641. PMID: [10556321](https://pubmed.ncbi.nlm.nih.gov/10556321/)
54. Delcher AL, Bratke KA, Powers, EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*. 2007; 23: 673–679. PMID: [17237039](https://pubmed.ncbi.nlm.nih.gov/17237039/)
55. Peyretailade E, Parisot N, Polonais V, Terrat S, Denonfoux J, Dugat-Bony E, et al. Annotation of microsporidian genomes using transcriptional signals. *Nat Commun*. 2012; 3: 1137. doi: [10.1038/ncomms2156](https://doi.org/10.1038/ncomms2156) PMID: [23072807](https://pubmed.ncbi.nlm.nih.gov/23072807/)
56. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2014; 42: D503–509. doi: [10.1093/nar/gkt953](https://doi.org/10.1093/nar/gkt953) PMID: [24157837](https://pubmed.ncbi.nlm.nih.gov/24157837/)
57. Smit AF, Hubley R, Green P. Available: <http://www.repeatmasker.org.RepeatMasker> Open.1996; 3: 1996-2004.
58. Gatehouse HS, Malone LA. The Ribosomal RNA Gene Region of *Nosema apis* (Microspora): DNA Sequence for Small and Large Subunit rRNA Genes and Evidence of a Large Tandem Repeat Unit Size. *J Invertebr Pathol*. 1998; 71: 97–105. PMID: [9547137](https://pubmed.ncbi.nlm.nih.gov/9547137/)
59. Huang WF, Tsai SJ, Lo CF, Soichi Y, Wang CH. The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. *Fungal Genetics and Biology*. 2004; 41: 473–481. PMID: [15050536](https://pubmed.ncbi.nlm.nih.gov/15050536/)
60. Tsai SJ, Wei FH, Chung HW. Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene. *J Euk Microbiol*. 2005; 52: 52–54. PMID: [15702980](https://pubmed.ncbi.nlm.nih.gov/15702980/)

61. Chen D, Shen Z, Zhu F, Guan R, Hou J, Zhang J. Phylogenetic characterization of a microsporidium (*Nosema* sp. MPr) isolated from the *Pieris rapae*. *Parasitol Res*. 2012; 111: 263–269. doi: [10.1007/s00436-012-2829-6](https://doi.org/10.1007/s00436-012-2829-6) PMID: [22314780](https://pubmed.ncbi.nlm.nih.gov/22314780/)
62. Chen YP, Pettis J, Zhao Y, Liu X, Tallon L, Sadzewicz L, et al. Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. *BMC genomics*. 2013; 14: 451. doi: [10.1186/1471-2164-14-451](https://doi.org/10.1186/1471-2164-14-451) PMID: [23829473](https://pubmed.ncbi.nlm.nih.gov/23829473/)
63. Comman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS pathogens*. 2009; 5: e1000466. doi: [10.1371/journal.ppat.1000466](https://doi.org/10.1371/journal.ppat.1000466) PMID: [19503607](https://pubmed.ncbi.nlm.nih.gov/19503607/)
64. Eakin AE, Mills AA, Harth G, McKerrow JH, Craik CS. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *J Biol Chem*. 1992; 267(11): 7411–7420. PMID: [1559982](https://pubmed.ncbi.nlm.nih.gov/1559982/)
65. Bruchhaus I, Loftus BJ, Hall N, Tannich E. The intestinal protozoan parasite *Entamoeba histolytica* contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. *Eukaryot cell*. 2003; 2(3): 501–509. PMID: [12796295](https://pubmed.ncbi.nlm.nih.gov/12796295/)
66. Sijwali PS, Rosenthal PJ. Gene disruption confirms a critical role for the cysteine protease falcipain-2 in hemoglobin hydrolysis by *Plasmodium falciparum*. *PNAS*. 2004; 101(13): 4384–4389. PMID: [15070727](https://pubmed.ncbi.nlm.nih.gov/15070727/)
67. Bjørnson S, Keddie BA. Effects of *Microsporidium phytoseiuli* (Microsporidia) on the performance of the predatory mite, *Phytoseiulus persimilis* (Acari: Phytoseiidae). *Biological Control*. 1999; 15(2): 153–161.
68. Becnel JJ, Jeyaprakash A, Hoy MA, Shapiro A. Morphological and molecular characterization of a new microsporidian species from the predatory mite *Metaseiulus occidentalis* (Nesbitt) (Acari, Phytoseiidae). *J Invertebr pathol*, 2002; 79(3): 163–172. PMID: [12133705](https://pubmed.ncbi.nlm.nih.gov/12133705/)
69. 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 research*. 2012; 22(12): 2478–88. doi: [10.1101/gr.142802.112](https://doi.org/10.1101/gr.142802.112) PMID: [22813931](https://pubmed.ncbi.nlm.nih.gov/22813931/)
70. Pan G, Xu J, Li T, Xia Q, Liu SL, Zhang G. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC genomics*. 2013; 14: 186. doi: [10.1186/1471-2164-14-186](https://doi.org/10.1186/1471-2164-14-186) PMID: [23496955](https://pubmed.ncbi.nlm.nih.gov/23496955/)