BMC Genomics

RESEARCH

Open Access

Bioinformatics analysis of the *Microsporidia* sp. MB genome: a malaria transmission-blocking symbiont of the *Anopheles arabiensis* mosquito

Lilian Mbaisi Ang'ang'o[1,2], Jeremy Keith Herren[2*] and Özlem Tastan Bishop[1*]

Abstract

Background The use of microsporidia as a disease-transmission-blocking tool has garnered significant attention. *Microsporidia* sp. MB, known for its ability to block malaria development in mosquitoes, is an optimal candidate for supplementing malaria vector control methods. This symbiont, found in Anopheles mosquitoes, can be transmitted both vertically and horizontally with minimal effects on its mosquito host. Its genome, recently sequenced from *An. arabiensis*, comprises a compact 5.9 Mbp.

Results Here, we analyze the *Microsporidia* sp. MB genome, highlighting its major genomic features, gene content, and protein function. The genome contains 2247 genes, predominantly encoding enzymes. Unlike other members of the Enterocytozoonida group, *Microsporidia* sp. MB has retained most of the genes in the glycolytic pathway. Genes involved in RNA interference (RNAi) were also identified, suggesting a mechanism for host immune suppression. Importantly, meiosis-related genes (MRG) were detected, indicating potential for sexual reproduction in this organism. Comparative analyses revealed similarities with its closest relative, *Vittaforma corneae*, despite key differences in host interactions.

Conclusion This study provides an in-depth analysis of the newly sequenced *Microsporidia* sp. MB genome, uncovering its unique adaptations for intracellular parasitism, including retention of essential metabolic pathways and RNAi machinery. The identification of MRGs suggests the possibility of sexual reproduction, offering insights into the symbiont's evolutionary strategies. Establishing a reference genome for *Microsporidia* sp. MB sets the foundation for future studies on its role in malaria transmission dynamics and host-parasite interactions.

Keywords Microsporidia, Malaria, *Anopheles*, Symbiosis, Genome, Annotation, Transmission-blocking, Biocontrol, Glycolytic pathway, RNA interference

Background

Malaria cases have been on a steady increase since 2016 with a notable 249 million cases reported in 2022 [1]. A total of 608 000 malaria deaths were recorded during this period, with 95.4% of the total global deaths from the WHO Africa region alone [1]. The major impediments to global malaria control have included the development of insecticide resistance in the *Anopheles* vector, *Plasmodium* drug resistance, poor detection of the minor malaria parasites, and the recent invasion of a new vector, *An. stephensi*, in the Horn of Africa and urban

*Correspondence:
Jeremy Keith Herren
jherren@icipe.org
Özlem Tastan Bishop
o.tastanbishop@ru.ac.za
[1] Department of Biochemistry, Microbiology, and Bioinformatics, Research Unit in Bioinformatics (RUBi), Rhodes University, Makhanda 6140, South Africa
[2] International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772-00100, Nairobi, Kenya

environments [1–5]. Due to these growing challenges, innovations in vector control have been the focus to mitigate resistance and improve the effectiveness of the current strategies. This has included the use of biocontrol tools to complement the current vector control tools.

Symbionts, such as bacteria and fungi, have been proposed as potential tools for controlling mosquito-borne diseases, including malaria [6–10]. These symbiotic microorganisms can affect the lifespan, reproductive competence, and vector competence of mosquitoes, which in turn can affect the transmission of mosquito-borne infectious diseases [6–8, 11–15]. Several approaches are being explored in the field of symbiotic control for mosquito-borne diseases. These include disrupting microbial symbionts required by insect pests, manipulating symbionts to express anti-pathogen molecules within the host, and introducing endogenous microbes that affect the lifespan and vector capacity of new hosts in insect populations [16]. Wolbachia is one such symbiont used to successfully suppress dengue transmission in *Aedes* mosquitoes [17–19]. Another unique example is the microsporidia species, which has been highlighted to have the capacity to function effectively as a biocontrol tool [9, 10, 20].

Microsporidia are a set of spore-forming obligate intracellular organisms affecting a diverse variety of vertebrates and invertebrates [21]. Microsporidia have an intricate infection mechanism which is characterized by the formation of a unique polar tube to invade the host cell [22, 23]. They have multiple modes of transmission including horizontal transmission via ingestion of spores, and vertical transmission from an infected mother to their offspring [21]. Microsporidia have the smallest genome sizes among eukaryotes, with the smallest genome of 2.3 Mb recorded from *E. intestinalis* [24, 25]. These enigmatic microorganisms have garnered significant attention due to their unique genomic characteristics, evolutionary adaptations, and varying host-microbe interactions [26].

Over half of the identified microsporidia species infect insects [27, 28]. These microsporidia have varying modes of infection and can have significant impacts on both the economy and the regulation of insect populations [29]. Several microsporidia species are known to be vertically transmitted, where the parasite is passed from parent to offspring. Strictly vertically transmitted microsporidia, like *Nosema granulosis*, *Dictyocoela roeselum* and *Dictyocoela muelleri*, have also displayed feminizing attributes, impacting the population dynamics of their hosts [30]. Other microsporidia display multiple modes of transmission, like *Edhazardia aedis* affects the mosquito *Aedes aegypti*, with both horizontal and vertical transmission pathways. Interestingly, the mode of transmission of *Edhazardia aedis* has been shown to dictate the blood-feeding success of their host with the microorganism being less virulent during its vertical transmission stages [31]. Some microsporidia exhibit relatively low levels of pathogenicity, making them less harmful to their hosts [32]. For example, *Vavraia culicis,* which infects various mosquito species with minimal pathogenic effects [26]. Agriculturally important pathogenic microsporidia, such as *Nosema bombycis* and *Nosema apis* (also known as *Vairimorpha apis*), affect silkworm and honeybee populations, respectively. Conversely, other microsporidia species have been explored as biocontrol tools against agricultural pests such as the use of the pathogenic *Paranosema locustae* to control grasshopper populations in the field and *Nosema pyrausta* in the control of the European corn borer [10, 27].

Mosquitoes are important hosts for microsporidian parasites showing high levels of host specificity and varying effects on their respective hosts [33]. These species belong to the genera Amblyospora, Hazardia, Encephalitozoon, Enterocytospora, Nosema, and Microsporidium [33, 34]. Owing to the extensive effect of microsporidia in mosquitoes, these have the potential for use in the control of vector populations and disease transmission [35]. Pathogenic microsporidia infecting mosquitoes include *Amblyospora connecticus* isolated from the salt-marsh mosquito [36], and *Nosema algerae* infecting *An. stephensi* [37], and *Parathelohania anophelis* affecting several Anopheles mosquitoes including *An. quadrimaculatus* [38]. *Edhazardia aedis* has been isolated from the *Aedes* species [31] while *Vavraia culicis* has been shown to affect plasmodium development in *An. gambiae* [39]. These microsporidia demonstrate significant potential in the control of vector-borne diseases. The most recent of which has been the discovery of *Microsporidia* sp. MB conferring protection against malaria development [40].

Microsporidia sp. MB (often denoted as *Microsporidia MB)* is a microsporidian symbiont that has been found in *Anopheles* mosquitoes in various regions of Africa, including *An. gambiae s.l.* in Kenya, *An. coluzzii* in Niger and both *An. gambiae s.s.* and *An. coluzzii* in Ghana [40–44]. It has been shown to have a strong malaria transmission-blocking phenotype, making it a potential candidate for the development of a symbiont-based malaria transmission-blocking strategy [40, 45]. The symbiont has multiple transmission routes including horizontal transmission between adult mosquitoes, specifically between male and female *An. Arabiensis* and maternal transmission [40, 42, 43]. Female *An. arabiensis* that acquire *Microsporidia* sp. MB horizontally can transmit the symbiont vertically to their offspring [42]. It has also been found that *Microsporidia* sp. MB can infect *Anopheles funestus s.s.*, another primary malaria vector

in Africa [42]. The prevalence of *Microsporidia* sp. MB in *Anopheles* mosquitoes varies across different regions and in different seasons, with a prevalence of 6% in Kenya and 1.8% in Ghana [41]. Further investigations are needed to understand the diversity and range of *Microsporidia* sp. MB across sub-Saharan Africa. Overall, *Microsporidia* sp. MB shows promise as a potential tool for controlling malaria transmission, but more research is needed to fully understand its effectiveness and potential applications [45].

Over the past few decades, researchers have made significant progress in understanding the genomic structures of microsporidia, an organism that is otherwise difficult to tame in laboratory setups owing to its obligatory intracellular nature. Through genome sequencing, scientists have been able to gain valuable insights into the biology, evolution, gene diversity, and pathogenicity of microsporidia [24, 25, 30, 46–71]. These findings have vastly highlighted the highly specialized lifestyle of microsporidia and their dependence on their host for essential biochemical processes. Furthermore, genome size reduction and gene loss in microsporidia species are closely linked to their reliance on host-derived nutrients and biochemical processes, as many genes involved in essential metabolic pathways are no longer needed when these functions are supplemented by the host[24, 46, 48, 49, 51, 57, 58, 72, 73]. This dependence on the host makes microsporidia an interesting model for studying the intricate interactions between microorganisms and their hosts. Challenges in annotating microsporidia genomes arise from their small size, high gene density, absence of introns, and the need for specialized methodologies to accurately predict and validate gene structures, which are influenced by their unique evolutionary adaptations to obligate intracellular parasitism [74].

To understand the mechanism of protection of *Microsporidia* sp. MB against malaria transmission, a high-quality reference genome is necessary. The draft genome of *Microsporidia* sp. MB isolated from *Anopheles arabiensis* in Kenya was recently sequenced using DNA NanoBalls Sequencing (DNBSEQ) short paired-end sequencing technology from dissected mosquito ovaries and de novo assembled [75].

In this study, we describe the structure of the *Microsporidia* sp. MB genome and further highlight its gene composition and key conserved proteins involved in important metabolic pathways within the host (75–80) using several computational tools. Microsporidia, characterized by their highly reduced genomes and lack of mitochondria, have lost most of the components of the glycolytic pathway but utilize host resources, indicating their reliance on host energy sources [76, 77]. This study revealed the presence of 10 of 11 key glycolytic genes in *Microsporidia* sp. MB, that have otherwise been lost in other microsporidia genomes. Additionally, we explored the RNA interference (RNAi) machinery within this symbiont, a complex previously implicated in the regulation of gene expression in microsporidia such as *Nosema ceranae* [78, 79].

Results and discussion

Comparative analysis of *Microsporidia* sp. MB and *Vittaforma corneae ATCC 50505* genome structures

Microsporidia sp. MB and *Vittaforma corneae ATCC 50505* are the closest relatives in Clade IV (Enterocytozoonidae) group based on 18S taxonomic classification [80], yet they differ significantly in their biological roles, transmission dynamics, and effects on their hosts. While *Microsporidia* sp. MB is a symbiotic organism primarily found in mosquitoes with a strong malaria transmission-blocking phenotype, *V. corneae* is known for causing ocular infections in humans, particularly in immunocompromised individuals. Based on these key differences, similarities and the availability of these sequence genomes online, a comparative overview of the two complex microsporidia genomes was conducted.

The de novo assembly of the *Microsporidia* sp. MB genome was based on a total of 3,874,899 reads, with an estimated sequencing depth of 100-fold coverage. The quality of the assembled genome was analyzed using the QUality Assessment Tool (QUAST) revealing that *Microsporidia* sp. MB had an average size of 5.9 Mb spanning across 2335 contigs compared to 3.2 Mb (314 contigs) in *V. corneae ATCC 50505* (GCA_000231115.1; accessed December 14, 2023) (Table 1). Despite having relatively different genome sizes, the number of genes predicted in both genomes was relatively similar (*Microsporidia* sp. MB = 2247*; V. corneae* = 2239).

The statistics above were further visualized on BlobTools as a snail plot (Fig. 1a). The Benchmarking Universal Single-Copy Orthologs (BUSCO) tool suite was used to assess the level of completeness of the assembled genome by checking the presence of core microsporidia genes that were represented in the genome (Fig. 1b).

Several reports suggest that the varying sizes of microsporidia genomes are mostly attributed to the proportion of repeats in their assemblies. To identify whether the size of the assembled genome was due to the presence of such repeats, their composition was analyzed using RepeatMasker v4.0 [81]. The *Microsporidia* sp. MB genome constituted 99.43% non-repeat regions and 0.57% repeats which included: Small RNA repeats (n = 821; 0.01%), low complexity repeats (n = 10,645; 0.18%), and simple repeats (n = 22,870; 0.38%) (Table 2). Short interspersed nuclear elements (SINE), Long interspersed nuclear elements (LINE), long terminal repeat

Table 1 Comparison of *Microsporidia* sp. MB Genome Statistics with its closest relative *V. corneae ATCC 50505*

Metric	*Microsporidia* sp. MB	*V. corneae ATCC 50505*
Assembly size (bp)	5,908,979	3,213,516
Number of contigs	2335	314
Contigs N_{50} (kb)	4	50.9
GC content (%)	31.12	36.5
The proportion of repeats (%)	0.57	Not available
Number of predicted genes	2247	2239
Gene density (genes/kb)	0.363	Not available
Mean CDS length (bp)	1108	Not available
BUSCO ($n = 600$)		
No. (%) of complete genes	486 (81.0)	450 (75)
No. (%) of complete and single-copy genes	485 (80.8)	443 (73.8)
Number (%) of complete and duplicated genes	1 (0.2)	7 (1.2)
Number (%) of fragmented genes	12 (2.0)	22 (3.7)
Number (%) of missing genes	102 (17.0)	128 (21.3)
GenBank accession number	JAVKTW000000000	NZ_AEYK00000000
SRA accession number	SRR25938329	Not available
Sequencing technology	DNBSeq	454

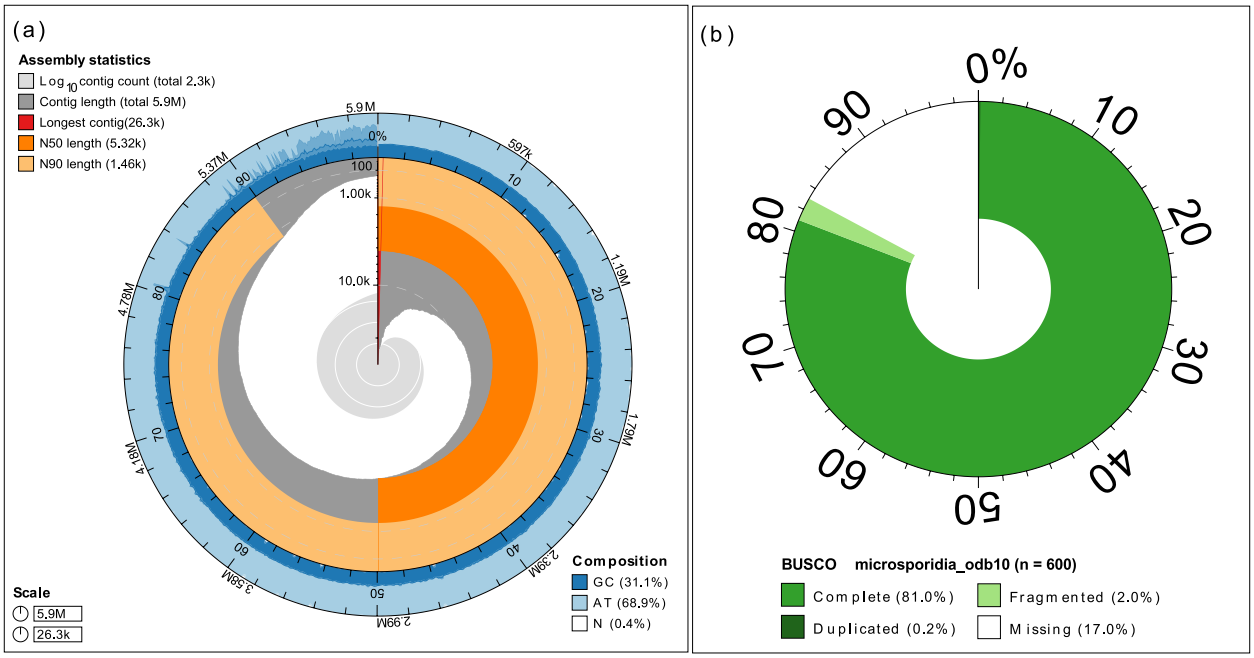

Fig. 1 *Microsporidia* sp. MB genome statistics. **a** The snail plot created using BlobToolKit highlights the genome metrics. A basic legend of the assembly statistics is provided in the top left corner. The outer blue rings highlight the GC content over the assembly, corresponding to each of the contigs in the grey ring. The overall GC content of the assembly is summarized at the bottom left corner. The light grey rings show the total length of the assembly. The red line highlights the longest scaffold, while the dark and light orange rings show the length of the contigs at which 50% and 90% of the total assembly is represented, respectively. **b** A pie-chart representation of the BUSCO gene completeness analysis shows a total of 81% single-copy core microsporidia genes present in *Microsporidia* sp. MB

retrotransposons (LTR), and DNA transposons were notably absent in this assembly. Small RNA repeats refer to repetitive sequences associated with non-coding small RNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs). A list of the different types of

Table 2 Distribution of different types of repeats identified in the *Microsporidia* sp. MB assembly

Repeat Type	Total Number of Elements	Percentage of Sequence covered (%)
SINEs	0	0
LINEs	0	0
LTR elements	0	0
DNA transposons	0	0
Rolling circles	0	0
Unclassified	0	0
Small RNA	821	0.01
Satellites	0	0
Simple repeats	22 870	0.38
Low complexity	10 645	0.18
Non-repetitive	5 874 643	99.43

repeats identified per contig and their positions in the sequences are highlighted in Supplementary Data File 1.

Gene prediction and gene organization

A total of 2247 genes were predicted using GeneMark-ES intronless eukaryote mode and each protein was checked against the Microsporidia taxon (Taxon ID: 6029) using NCBI BLASTp (E-value = 0.0001). The predicted proteins in *Microsporidia* sp. MB range in size from 99 to 263 amino acids. Supplementary Data Files 2–4 contain information on the predicted gene sequences including genbank file, predicted coding sequences and predicted proteins respectively. Moreover, Supplementary Data File 5 contains a highlight of BLAST output results from these sequences.

A set of genes annotated using GeneMark-ES and corresponding to BUSCO identification of single-copy and complete genes were used to plot the genome architecture schematic in Fig. 2 below. The representation was based on the top contigs of lengths greater or equal to 10 kb, ranging from 26.338 – 10.051 kb ($n = 48$).

Analysis of common orthogroups in *Microsporidia* sp. MB and the Enterocytozoonida group of *Microsporidia*

The comparative phylogenomic analysis of *Microsporidia* sp. MB against the Enterocytozoonida group revealed that this symbiont is closely related to *V. corneae ATCC 50505* as has been previously seen using the 18S region [80]. Ninety-two common orthogroups were found in all 8 microsporidia species and were used to infer phylogeny within the group (Fig. 3).

In addition to comparing the genome of *Microsporidia sp. MB* with *V. corneae* ATCC 50505, recently published genomes from the Enterocytozoonida group were also considered, including *Agmasoma sp.* [87], *Enteropsectra* and *Pancytospora* [88] and, *Alternosema astaquatica* [89]. The nematode-infecting microsporidia, *Enteropsectra* possesses a genome size of approximately 6.0 Mb, encoding around 3531 predicted proteins [88]. In contrast, the two *Pancytospora* species exhibit smaller genomes of 4.2 Mb and 4.6 Mb, with predicted protein counts ranging from 2300 to 2800 [88]. When compared to *Microsporidia* sp. MB, which has a genome size of 5.9 Mb and 2247 predicted proteins, *Enteropsectra* displays a larger genome and a substantially higher number of predicted genes, suggesting greater genomic complexity. Conversely, the *Pancytospora* species show reduced genomic size and fewer predicted proteins, highlighting the genomic diversity within this group.

Sequence-based protein analysis

Protein functional analysis

Protein family classification InterProScan [90] was used to classify the superfamilies of the putative *Microsporidia* sp. MB proteins predicted using BUSCO and GeneMark-ES tools. A total of 2,247 proteins were predicted with 2235 of these classified into their respective superfamilies (Supplementary Data File 6). Over 30% ($n = 707$) of these proteins were associated with enzymes and only 12 proteins (0.53%) were left unclassified. The other protein family groups identified included: cell-cycle ($n = 530$; 23.6%), regulatory/cellular ($n = 364$; 16.2%), binding ($n = 353$; 15.7%), transport ($n = 178$; 0.9%), and structural proteins ($n = 103$; 4.6%) (Fig. 4a, Supplementary Data File 2). The large distribution of enzymes in this genome is comparatively similar to a report on the protein family classes of *V. corneae ATCC 50505* [91].

Distribution of enzyme classes Since most of the predicted proteins were classified as enzymes, further analysis of the different groups of enzymes was essential to determine the most common enzyme class represented in the dataset. Seven main enzyme classes were identified, with transferases appearing to be the most common category among the annotated enzymes Fig. 4b**.** These were divided into: hydrolases ($n = 237$; 38.47%), transferases ($n = 228$; 37.01%), ligases ($n = 45$; 7.31%), isomerases ($n = 35$; 5.68%), oxidoreductases, ($n = 33$; 5.36%), translocases ($n = 31$; 5.03%) and lyases ($n = 7$; 1.14%).

Each enzyme class was further grouped into their respective subclasses based on their different specific functions as shown in Fig. 5 below. Hydrolases are involved in the carbohydrate and energy metabolism of intracellular stages of the microsporidian *Nosema grylli* [92]. Furthermore, the localization and functionality of

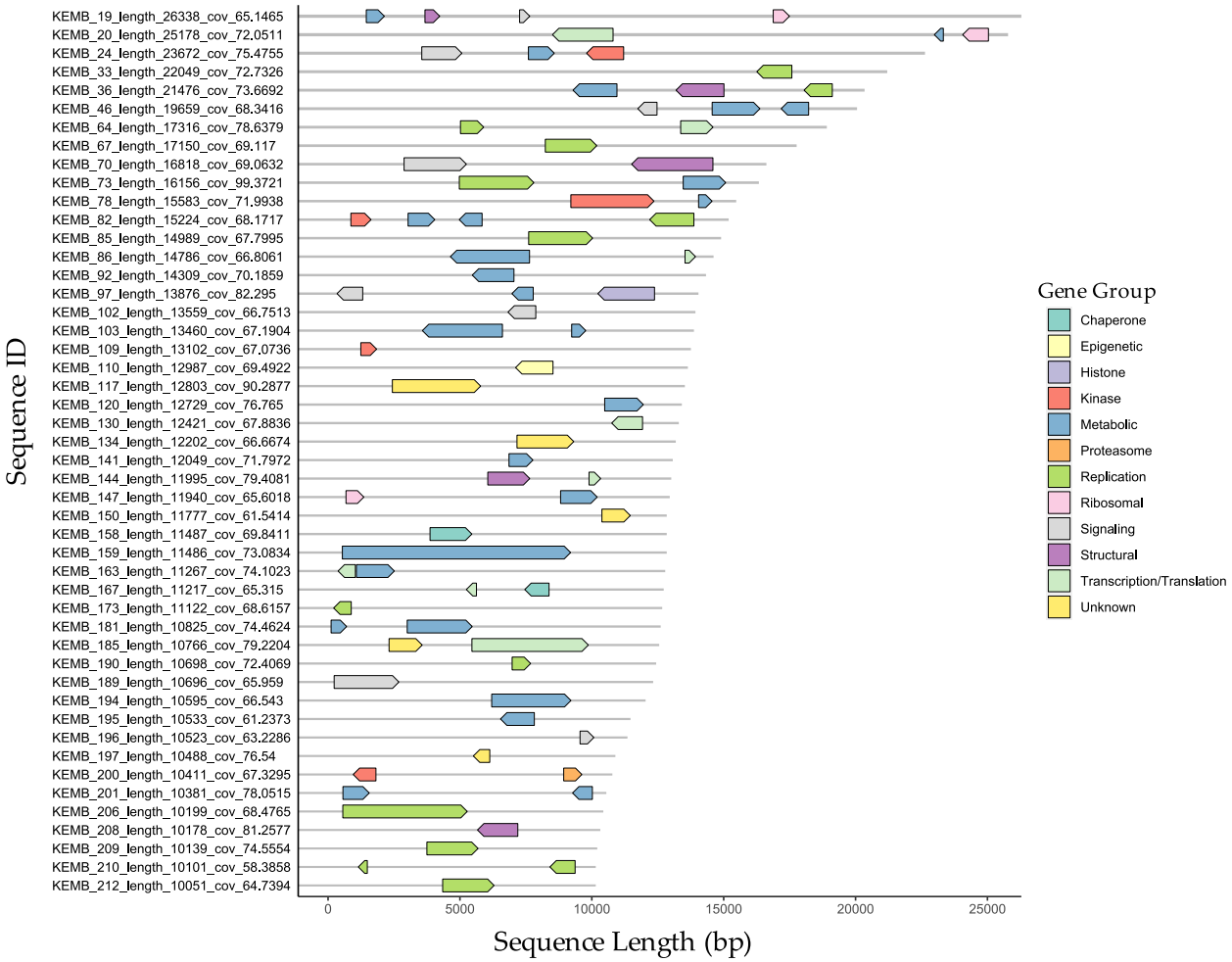

Fig. 2 Genome structure of *Microsporidia* sp. MB, showing contigs larger than 10 kb to focus on the significant portions of the assembly. The figure illustrates gene density and organization across these contigs, with the Y-axis representing the names of the contigs. Each gene is color-coded based on its functional classification, grouped into 12 distinct categories as indicated in the legend. The gene annotations are derived from GeneMark-ES and validated using BUSCO, ensuring high confidence in the predicted gene functions. This schematic helps visualize the distribution and functional diversity of genes across the larger contigs in the genome.The plot was created in RStudio [82] using the gggenes package (https://cran.r-project.org/web/packages/gggenes)

microsporidian iron-sulfur cluster assembly proteins, which are likely to be hydrolases, have been investigated by Goldberg [93]. However, the specific functions of hydrolases in microsporidia, particularly in the context of their pathogenicity and host interactions, remain to be fully elucidated. Here, we identified 38.47% ($n=237$) of the classified enzymes in *Microsporidia* sp. MB as hydrolases with most of these acting on acid anhydrides (Fig. 5).

Transferases play a crucial role in microsporidia intracellular lifestyle, particularly in the acquisition of essential nutrients and energy from their host cells. These enzymes also facilitate the import of nucleotides for nucleic acid biosynthesis. Nine main subclasses of transferases were identified and those involved in transferring phosphorus-containing molecules were the most common (Fig. 5).

The role of ligases in microsporidia is further supported by the identification of several components of the mitochondrial iron-sulfur cluster assembly machinery in these organisms. These components, which participate in the biosynthesis of iron-sulfur proteins, are essential for various cellular processes, including DNA repair [94]. The presence of these components suggests that microsporidia may use ligases to maintain their genome integrity and ensure their survival within the host cell [55, 95].

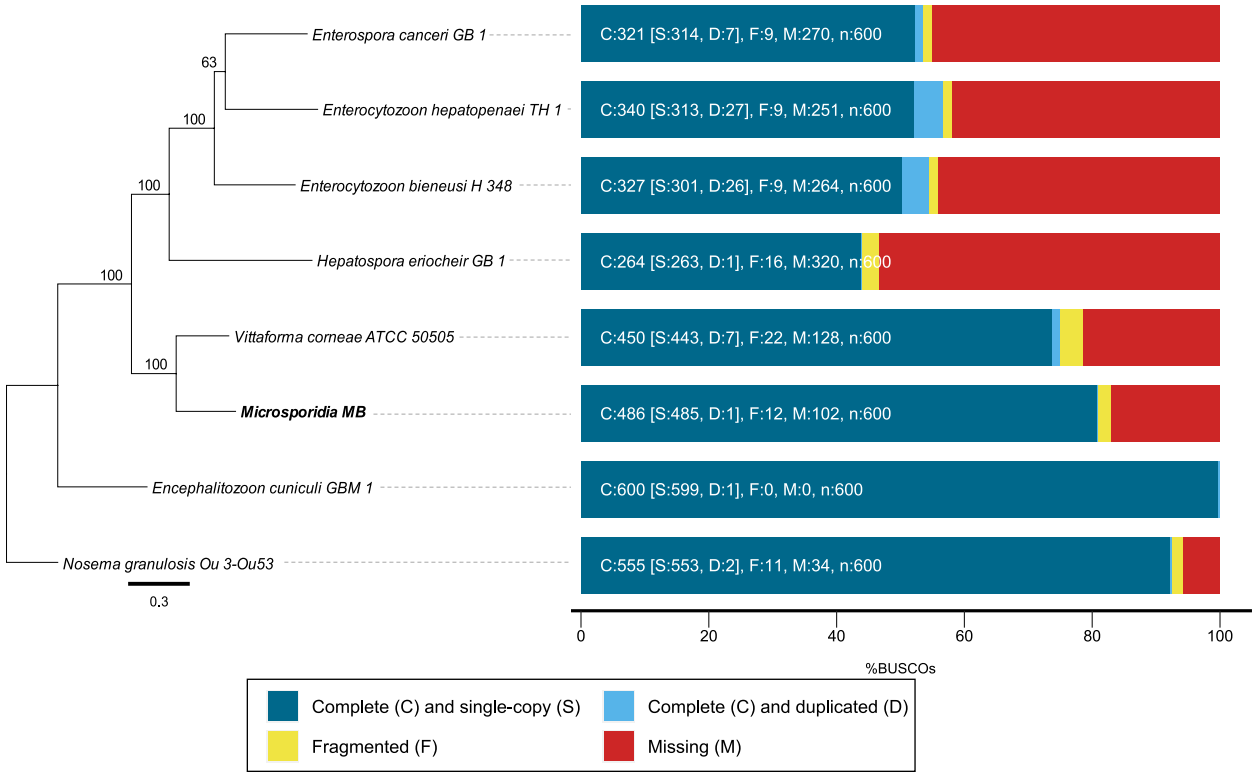

Fig. 3 Phylogenomic analysis of common BUSCO orthogroups found in Enterocytozoonida (left) and their respective genome assembly completeness using BUSCO (right). *Microsporidia* sp. MB is bolded. *Nosema granulosis* was used as an outgroup. Alignments of common orthogroups were performed using MUSCLE and trimAl [83, 84]. The maximum likelihood tree (1000 bootstraps) was constructed using IQ-TREE v1.6.12 [85] based on the 92 single-copy orthologous genes (with the best model selection (LG + F + I + G4) from ModelFinder [86]. Bootstrap percentage values are shown on each node and the scale bar at the bottom of the tree indicates total changes per site

Microsporidia sp. MB contained 7.31% ligases of the predicted enzymes, which were further classified into 5 subclasses depending on the type of bonds they create with different molecules (Fig. 5 above).

A total of eleven oxidoreductase subclasses were observed. Isomerases, oxidases, and thiol-sulfide reductases are important in redox reactions and useful in the survival of the organism in new host environments. Here, 5 isomerase subclasses were identified, some of which had direct links to oxidoreductases and transferases. Lyase proteins were the least-represented enzyme classes in the genome consisting of carbon–carbon, carbon–oxygen, and phosphorus-oxygen subgroups (Fig. 5 above).

Distribution of key domains and repeat regions Domains, repeats, and conserved sites are key features found in proteins. Domains in proteins are distinct regions that have specific structural and functional properties and are typically composed of several secondary structure elements, such as alpha helices and beta sheets, and can fold independently. The term 'key domains' refers to conserved regions within proteins that are essential for their functional properties, such as enzyme activity or interaction with host cells. These domains include well-characterized motifs like reverse transcriptase domains, leucine-rich repeats (LRR), and others identified through InterProScan. Repeats, on the other hand, are segments of protein sequences that are repeated multiple times within the same protein or across different proteins. Repeats can be indicative of functional motifs or structural elements within proteins.

The predicted proteins contained a total of 659 domains and 19 repeats. A total of 1456 domains were identified in the putative proteins with a majority being the reverse-transcriptase domains ($n = 179$) as shown in Fig. 6a. Microsporidia lack nucleotide biosynthetic pathways and rely on "stealing" nucleotides from host cells [96, 97]. Reverse transcriptase domains are crucial components of microsporidia proteins as they allow the proteins to perform reverse transcription, a process in which RNA is converted into DNA [98]. Repeat regions within

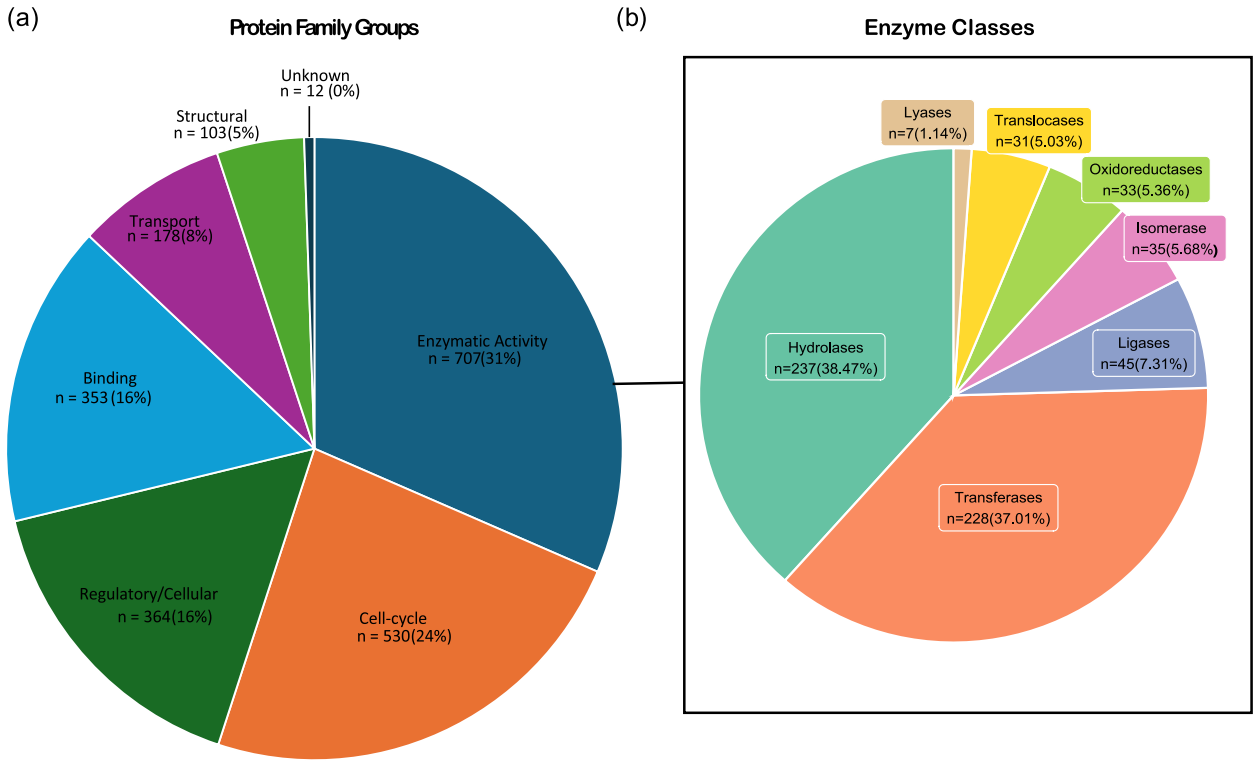

Fig. 4 Protein family classification using InterProScan. **a** Seven main protein superfamily groups were identified, with most of the proteins classified as enzymes. **b** Zooming into the classification of the primary enzyme groups using InterProScan, hydrolases and transferases are highlighted as the most abundant enzyme classes in *Microsporidia* sp. MB. These two enzymes function in energy metabolism and transport of nutrients, respectively

these proteins were also identified through InterProScan as shown in Fig. 6b below.

Most notable was the presence of leucine-rich repeat regions (LRR) with a total of 17 proteins containing these repeats (Fig. 6b above). The LRRs found in microsporidia proteins play a crucial role in the recognition and interaction with host cells and may represent pathogenicity factors. These regions are often involved in mediating protein–protein interactions, allowing microsporidia to bind to and invade host cells [98]. Microsporidian LRR-proteins secreted into the host have the potential to interact with host protein function. LRR-proteins in *S. lophii* were shown to have N-terminal signal peptides indicating potential secretion or targeting to the parasite cell surface for interactions with the host [60]. The presence of similar sequences in *V. corneae* also demonstrates the significance of LRR-proteins in mediating protein–protein interaction [60].

An investigation of the distribution of LRRs in the proteome using MEME Suite showed that 9 of the assembled contigs contained key LRR motifs (Fig. 7). The secretome, which refers to the collection of proteins secreted by an organism, plays an important role in host-parasite interactions by influencing immune responses and cellular processes in the host. To identify whether the LRR-containing proteins could be part of the *Microsporidia* sp. MB secretome, an important feature in host-parasite interactions, further computational assessment of the presence of signal peptides and transmembrane localization prediction was done. Three of the identified contigs contained putative signal peptides suggesting they could be part of the secretome, in direct contact with the host cell (Supplementary Table 2).

One contig (NODE_405) contained 2 main LRR genes (gene_679 and gene_680) with confirmed signal peptides using both TargetP and SignalP prediction tools while contig NODE_1665 contained one gene (gene_1719). Gene_679 was identified to be homologous to the *Anopheles* plasmodium-responsive leucine-rich repeat protein APL1B which has been shown to play a role in *An. gambiae* immune response to invasion by plasmodium [100] and could suggest a probable function of gene_679 in blocking malaria in its host.

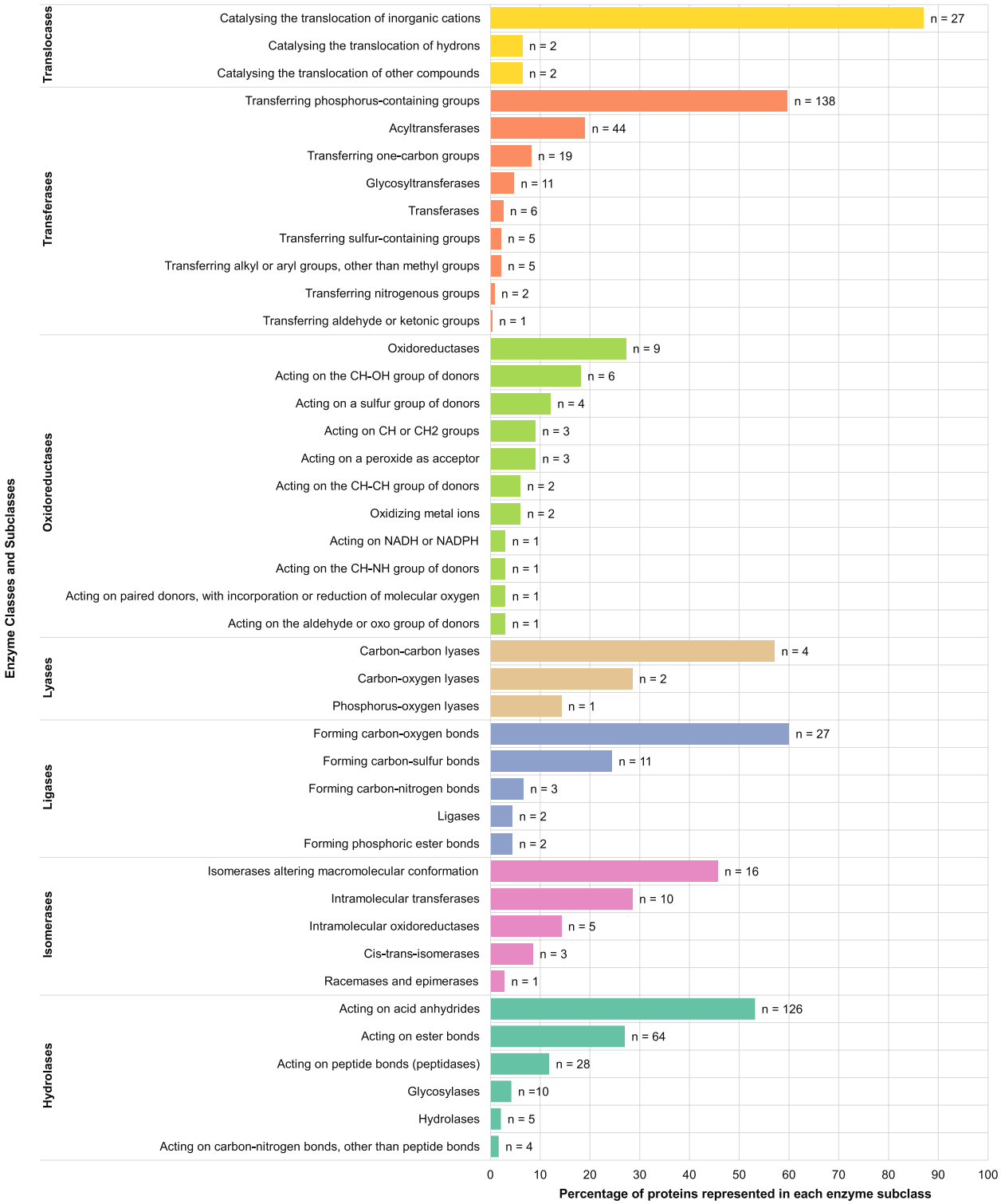

Fig. 5 Distribution of different enzyme classes and subclasses in *Microsporidia* sp. MB as predicted through InterProScan. "*n* = " denotes the number of proteins in each respective subclass while the x -axis represents the proportion of each subclass in their respective enzyme class in percentage

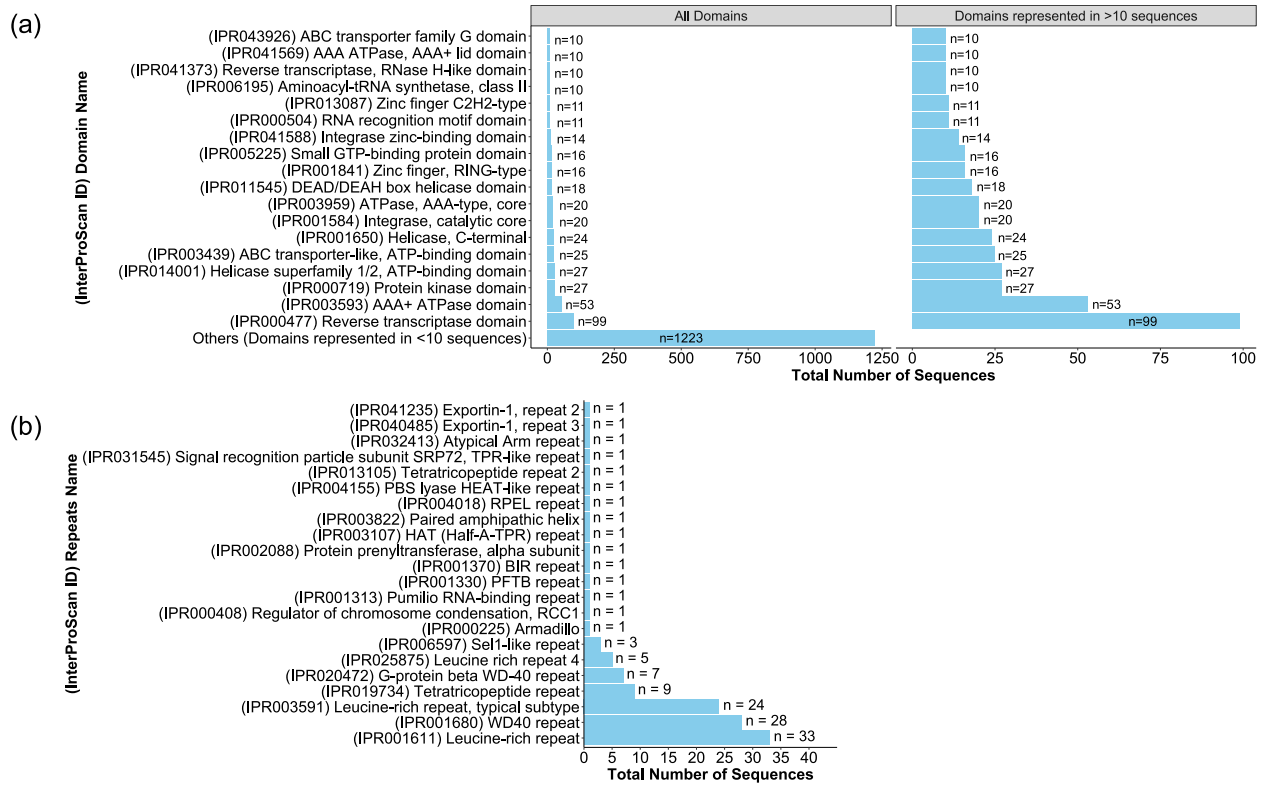

Fig. 6 Sequence distribution of the different domains and repeats predicted using InterProScan. **a** A bar plot of the domains identified from the genome shows a large set of domains involved mainly in enzymatic activities. A zoomed image of the main domains excluding the outlier (where less than 10 sequences were represented in each domain) is shown on the right panel. **b** The distribution of repeat regions is depicted in this bar plot. Leucine-rich repeat regions (LRR) and WD40 were the most abundant. WD40 proteins named for their characteristic WD40 repeat motifs which consist of 40 amino-acids typically ending with tryptophan (W) and aspartic acid (D), are involved in a wide variety of cellular processes due to their ability to mediate protein–protein interactions

The 3D model of gene_679 was designed using template APL1 (PDB ID: 3O6N_A; 1.85 Å) and was characterized with 8 alpha helices, 11 beta sheets, and loops (Fig. 7b, Supplementary Fig. 1). QMEAN score of 0.5, Verify3D overall score of 77.98%, and Z-DOPE score of −3.83. Moreover, gene_679 contained both signal peptides and transmembrane segments. No good-quality template was identified from HHSearch to model gene_680 and gene_1719.

Gene ontology

Gene ontology (GO) analysis is frequently used to annotate genes and gene products and to determine the biological features of large-scale genome or transcriptome data [101]. Here, we illustrate using GO that most of the putative proteins in *Microsporidia* sp. MB are involved in metabolic processes and transport attributed to the bulk of enzymes found in the predicted dataset. Most of these proteins were found in the nucleus and cytoplasm (Fig. 8).

KEGG pathway analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a collective database integrating genomes, biological pathways, diseases, drugs, and chemical substances [102]. A total of 257 pathways were identified with 741 sequences linked to these pathways (E-value = 0.0001). KEGG further categorized the proteins into five distinct groups with most of the proteins involved in metabolism ($n = 83$; 32.3%) and organismal systems ($n = 83$; 32.3%) (Fig. 9a). The other categories included environmental information processing ($n = 35$; 13.6%), cellular processes ($n = 30$; 11.7%), and genetic information processing ($n = 26$; 10.1%).

Sequence distribution in each represented pathway showed that most of the predicted proteins are involved in metabolism (purine, thiamine, nucleotide metabolism and biosynthesis of cofactors), genetic information processing (ribosome biogenesis, protein processing, nucleotide excision repair, RNA degradation and RNA polymerase, among others), and general

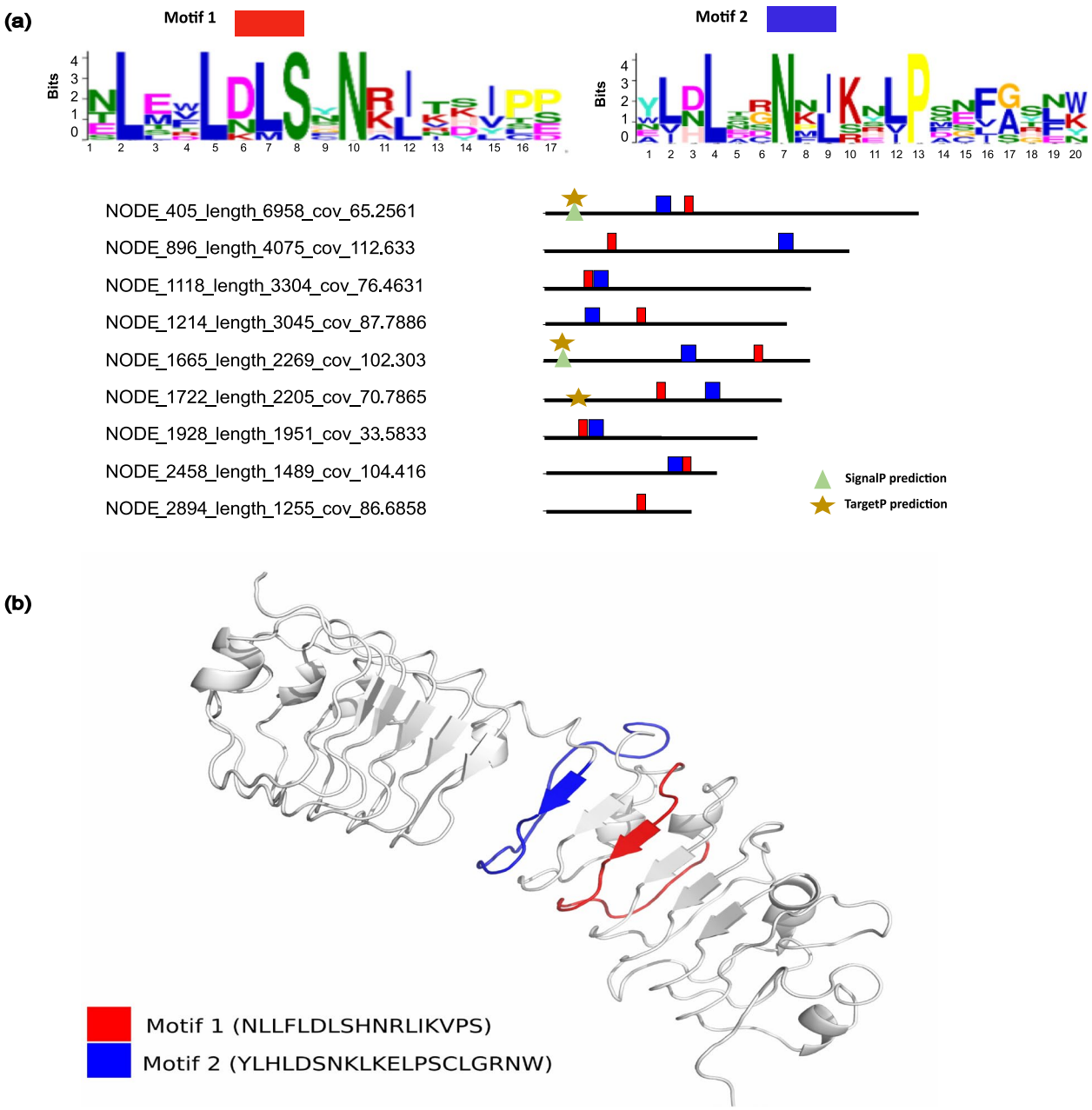

Fig. 7 **a** Organization of leucine-rich repeats (LRR) across 9 contigs. MEME motif output showing conservation levels across the motif. Larger symbols indicate more common amino acids. Y-axis (bits) shows conservation levels—higher bits reflect higher conservation. X-axis denotes positions along the motif, with higher Y values indicating consistent amino acids at specific positions across analyzed sequences. Two LRR motifs, Motif 1 (red, 17 amino acids) and Motif 2 (blue, 20 amino acids), were identified using MEME Suite. Moreover, the prediction of signal peptides using TargetP v2.0 and SignalP v6.0 (green and brown respectively) in 2 of these sequences (NODE_405 and NODE_1665). **b** A 3D model of gene_679 (NODE_405) with motif mapping. The structure was modeled on PRIMO [99]. The identified LRR motifs are numbered based on MEME output

cellular processes (cell cycle, meiosis, cellular senescence and nucleocytoplasmic transport) (Fig. 9b).

Physicochemical properties

The physicochemical properties of proteins, such as surface hydrophobicity, structural stability, and surface charge, play a critical role in determining their biological functions and interactions. These properties are influenced by the protein structure, including the content of α-helices and β-sheets, which in turn affect the flexibility of the protein and functional characteristics. Here, we characterized the physicochemical properties

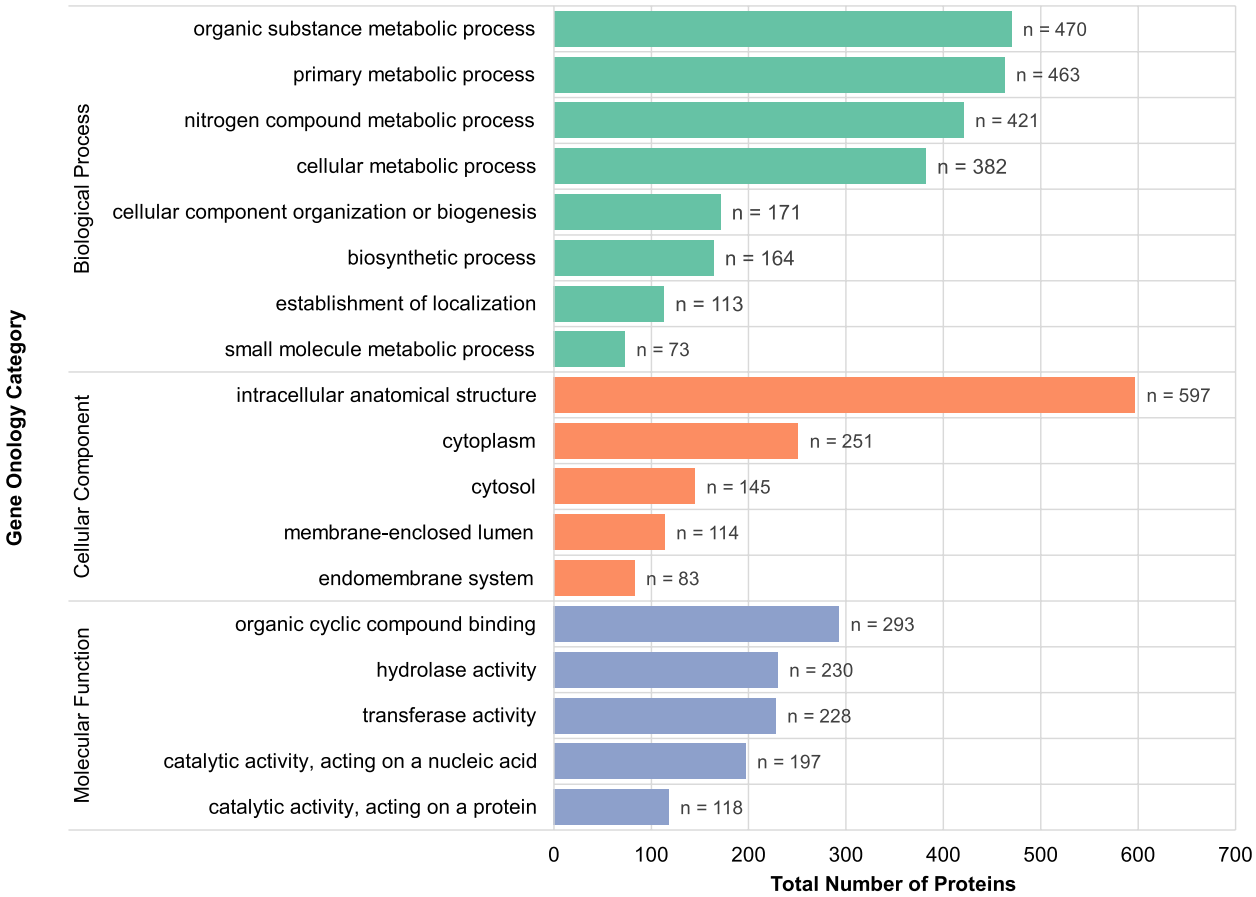

Fig. 8 Gene Ontology Analysis. A highlight of the protein sequence distribution in each primary GO term (Biological Process, Cellular Component, and Molecular Function) shows that the bulk of the putative proteins are involved in metabolic processes, localized within the intracellular anatomical structure (nucleus and cytoplasm), and involved in enzymatic activities

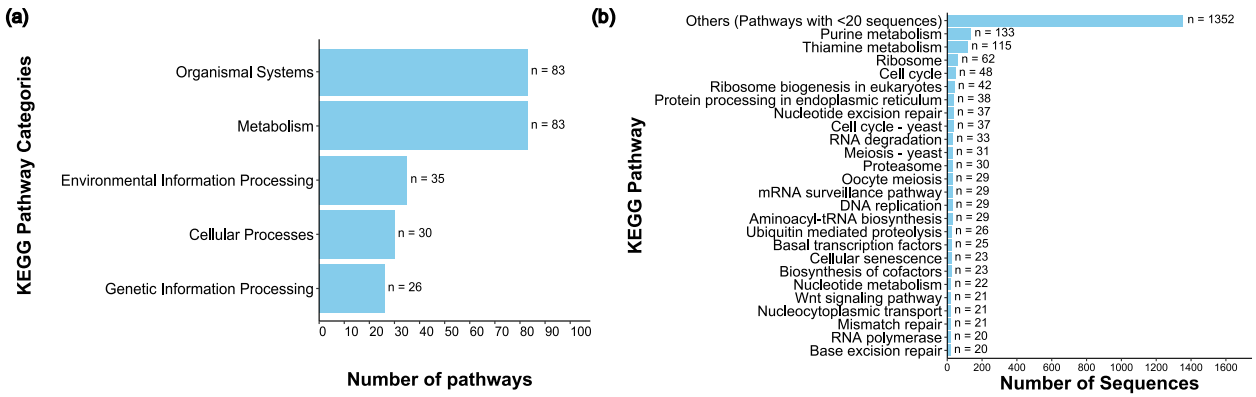

Fig. 9 KEGG metabolic pathway analysis. **a** A barplot representation of the main KEGG categories shows most of the pathways identified within the genome take part in the organismal systems and metabolism. **b** Sequence distribution in each KEGG pathway highlighting the top 25 pathways represented in the dataset. A large portion of the sequences were involved in purine and thiamine metabolism, cellular processes, and genetic information processing. "n" represents the total number of proteins in each category

of *Microsporidia* sp. MB using the EXPASY-ProtParam tool and observed that 1911 (85%) proteins were hydrophilic (Grand Average of hydropathy (GRAVY) score less than zero) while 336 (15%) proteins were hydrophobic (GRAVY score greater or equal to zero). Hydrophilic proteins are more likely to be connected to interactions with the host and thus makes the hydrophobicity test relevant for predicting potential host-symbiont interacting proteins in the dataset. Many of the proteins were identified to be hydrophilic suggesting the heavy reliance of microsporidia on host interactions.

Transmembrane regions and signal peptides prediction

Secretory proteins commonly contain signal peptides or span across membrane regions to allow direct interaction with the host cell. Therefore, identifying signal peptides and transmembrane regions is important in highlighting probable microsporidia-host interactions. The secretion of parasite proteins from different functional categories suggests their active role in manipulating host cell machinery shedding light on the mechanisms employed by microsporidia to modify host metabolic and regulatory pathways [98]. Here, we highlight 84 proteins from different functional groups containing signal peptides using SignalP v6.0 suggesting that these proteins could be part of the *Microsporidia* sp. MB secretome. Moreover, an overview of transmembrane regions using DeepTMHMM [103] identified 328 (14.6%) proteins spanning these regions while 1919 (85.4%) of these proteins did not have any transmembrane segments (Supplementary Table 3).

Eighty-four proteins had signal peptides and 23 of these contained transmembrane regions (Supplementary Table 3). Moreover, out of the 84 proteins, 69 were hydrophilic based on GRAVY scores. Thirteen of these were characterized as hydrophilic and contained both signal peptides and transmembrane segments.

Microsporidia sp. MB has retained most of the components involved in the glycolytic pathway

Microsporidia sp. MB contained 10 of the 11 proteins involved in the glycolytic pathway (Fig. 10). Among the

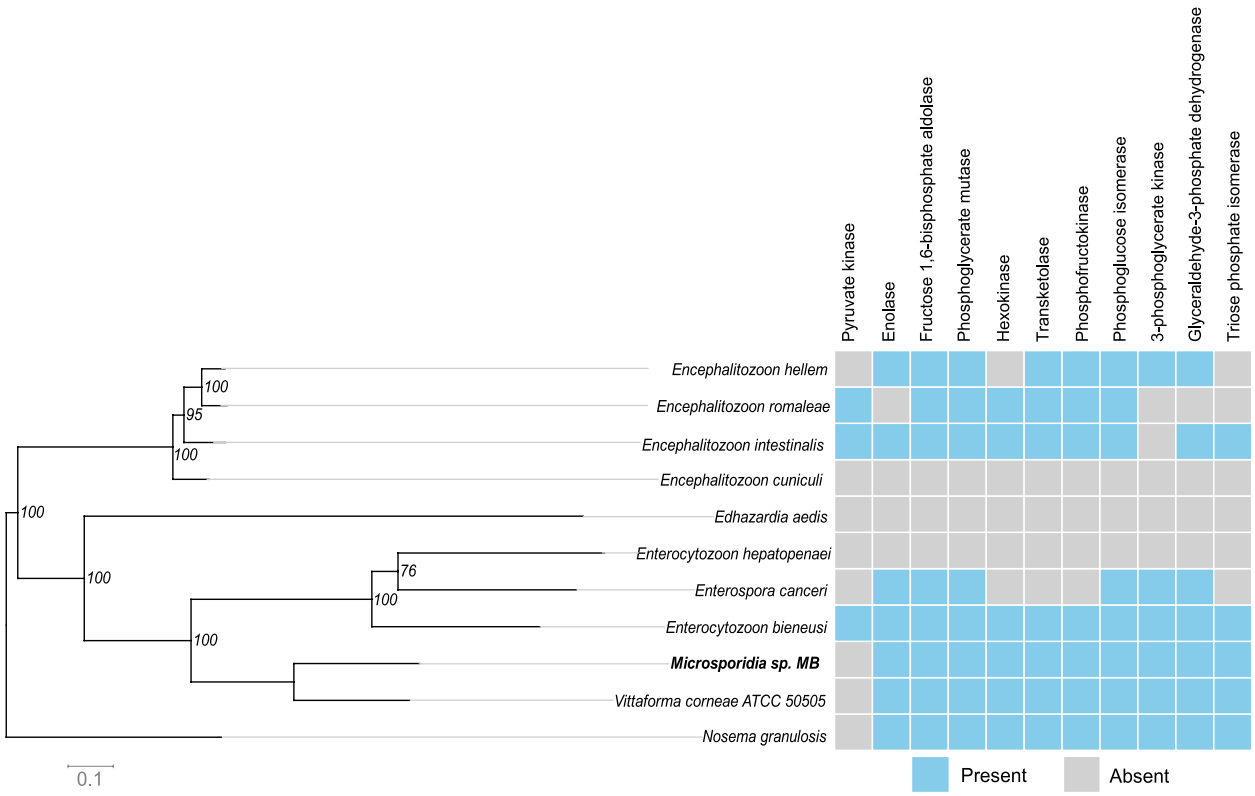

Fig. 10 Overview of the loss and retention of key genes involved in the glycolytic pathway genes across different microsporidia species. Retained genes are highlighted in light blue in the heatmap, while those that have been lost are highlighted in gray. *Microsporidia* sp. MB has retained most of the genes involved in glycolysis except for Pyruvate kinase, similar to *V. corneae* and *N. granulosis*. Comparatively, *E. aedis, E. cuniculi, and E. hepatopenaei* appear to have lost all these genes, suggesting a total reliance on the host for energy production. The maximum likelihood phylogenomic tree on the left was generated using IQ-TREE v1.6.12 [85] based on the 196 common BUSCO single-copy orthologous genes found in all species. Percent bootstrap support values are shown on each node and the branch length scale is bare at the bottom of the tree

Enterocytozoonida species, *E. cuniculi* has been observed to have retained all the proteins in this pathway while *E. bieneusi* seems to have lost all of these. The presence of all glycolytic genes in *E. cuniculi* and their absence in *E. bieneusi* likely reflect biological differences, such as adaptations to varying levels of host dependency and metabolic requirements, rather than being solely due to differences in genome annotation completeness. While *E. cuniculi* has been widely studied and annotated, the retention or loss of glycolytic genes is likely driven by the specific biological needs of the organism. The missing protein in *Microsporidia* sp. MB (3-phosphoglycerate kinase) could also suggest that this gene has not been sequenced yet. Comparably, its closest sister *V. corneae* and *Nosema granulosis* had the same set of proteins involved in the glycolytic pathway (Fig. 10).

The presence of Meiosis-related Genes (MRG) in *Microsporidia sp.* MB indicates possible mode of sexual reproduction

Polyploidy, which involves having multiple sets of chromosomes, is common in microsporidia and shows dynamic changes across different species, suggesting its significant role in their adaptability and evolution [104]. Meiosis is a crucial process in sexual reproduction, leading to the formation of haploid gametes. In microsporidia, meiosis-related genes and their regulation have not been fully studied. However, meiosis plays a pivotal role in the formation of spores, which are essential for their transmission and infection [105]. The identification of meiosis-related genes in microsporidia is essential to understanding the mechanisms underlying their sexual reproduction and transmission.

The process of meiosis occurs in microsporidia during the early phases of sporogony [106]. The merging of haploid nuclei within a diplokaryotic sporont results in the production of a single diploid nucleus. A new diplokaryon with two diploid nuclei is then created by the mitotic division of this single nucleus. These diploid nuclei split, causing the formation of synaptonemal complexes within each nucleus. They then fuse to form a single tetraploid nucleus, where chromosomal mingling occurs. Eight haploid nuclei are finally produced by this tetraploid nucleus by two consecutive reductive divisions, two diploid nuclei, four haploid nuclei, and a final mitotic division [107, 108]. Despite the reduced complexity of microsporidia genomes, investigations have revealed genes that show homology or similarity to known meiotic genes in other eukaryotes [109]. These genes include homologs of meiosis-specific genes found in other organisms, such as Dmc1, Msh4, and Msh5, among others (Table 3) and are frequently involved in basic meiotic processes such as recombination, synapsis, chromosome segregation, and spore production. Their presence and functionality indicate that, despite their reduced genomes, microsporidia have kept critical genetic components required for meiosis, albeit with certain modifications and adaptations peculiar to their parasitic lifestyle and evolutionary history. Identifying and understanding these meiosis-related genes aids in the study of microsporidia reproductive strategies and evolutionary biology.

This study highlights the presence of several MRGs in the newly assembled genome. Some of the genes involved in meiosis and identified in this study are listed in Table 3 below. The presence of these genes in

Table 3 A list of MRGs found in *Microsporidia* sp. MB

MRG Name	Function	*Microsporidia* sp. MB homolog
DNA mismatch repair protein MSH5	Meiotic DNA crossover	gene_684
Meiosis-specific protein HOP2	Meiotic double-strand DNA break repair	gene_731
Recombination protein MND1	Recombination and meiotic nuclear division	gene_1920; gene_1498
Recombination protein RAD52/22	Double-strand DNA break repair	gene_421
DNA mismatch repair protein MSH2	Mismatch repair during meiosis	gene_114; gene_1101
DNA mismatch repair protein MSH6	Mismatch repair during mitosis and meiosis	gene_114
DNA mismatch repair protein MutL homologs (MLH1/PMS1)	Mismatch repair during mitosis and meiosis	gene_478; gene_1355
Structural maintenance of chromosomes protein 2 (SMC2)	Sister chromatid cohesion	gene_89
Structural maintenance of chromosomes protein 4 (SMC4)	Subunit of the cohesion complex	gene_1051; gene_1077
Structural maintenance of chromosomes protein 5—6 complex (SMC5-6)	DNA repair	gene_160; gene_941
DNA repair protein RAD51	DNA repair	gene_1069
RAD18-like recombination and DNA repair protein	Post-replication repair	gene_15
RAD21 / REC8 like protein	Subunit of sister chromatid cohesion complex	gene_1754

Microsporidia sp. MB suggests the putative mode of reproduction of this organism.

Microsporidia sp. MB contains the three main components of the RNAi complex

The RNA-induced silencing complex (RISC) is a conserved molecular machinery responsible for post-transcriptional gene regulation through small RNA molecules, particularly microRNAs (miRNAs) and siRNAs. While the RNAi pathway and RISC are well-studied in many eukaryotes, including animals, plants, and fungi, the presence, and mechanisms of RNAi in microsporidia, have only been recently highlighted [61, 78, 79, 110]. The presence of RNAi machinery in microsporidia varies among species (Table 4). While some microsporidia have retained elements of the RNAi pathway, others have lost or significantly modified these components. This variation is attributed to differences in evolutionary history and the degree of genome reduction among microsporidian lineages. If functional, the RNAi pathway in microsporidian potentially plays various roles, including the regulation of host-parasite interactions, suppression of host immune responses, or even self-regulation of microsporidian gene expression. RNAi could also be involved in the maintenance and evolution of the reduced microsporidian genome [13, 54]. *Microsporidia* sp. MB was observed to contain homologs of key RNA interference proteins. These included the dicer (gene_297), RNA-dependent RNA polymerase (gene_163), and argonaute (gene_271) genes (Table 4). All three of these genes were also present in *V. corneae ATCC 50505* but were noticeably missing in the other microsporidians of the Enterocytozoonida group (Table 4).

Argonaute

The phylogenetic classification of argonaute is shown in Fig. 11a below with *Microsporidia* sp. MB consistently showing high similarity in protein sequence with *V. corneae ATCC 50505.* The argonaute protein is characterized by two main conserved domains: the PIWI-like and PAZ argonaute-like domains. These were also identified in the *Microsporidia* sp. MB homolog (Fig. 11b, Supplementary Fig. 2) and the conservation of the domains across several microsporidia species highlighted through multiple sequence alignment.

Dicer

Phylogenetic analysis of the dicer orthologs found in 15 microsporidia species was done using (Fig. 12a) which showed high similarity between *Microsporidia* sp. MB and *V. corneae ATCC50505*. An analysis of the conserved domains of the dicer gene on CDD revealed that this contained two main conserved domains including the Ribonuclease III C terminal domain (RIBOc) and the dsRNA-specific ribonuclease domain (Rnc) (Fig. 12b). Multiple sequence alignment showed the presence of inserts between the RIBOc and Rnc domains was found in *Microsporidia* sp. MB, *E. breve, P. epiphaga,* and *P. philotis*. Supplementary Fig. 3 highlights the conservation of these genes in the multiple sequence alignment.

RNA-dependent RNA polymerase

RNA-dependent RNA polymerase is involved in post-transcriptional gene silencing, and we observed high similarity between the orthologs of this gene found in *V. corneae* and *Microsporidia* sp. MB (Fig. 13a). RNA-dependent RNA polymerase was identified to include

Table 4 An overview of the retention of the key RNAi components in Enterocytozoonidae and mosquito microsporidia

Microsporidia species	Argonaute	Dicer	RNA-dependent RNA polymerase	Host
***Microsporidia* sp. MB**	✓	✓	✓	**Mosquito**
V. corneae ATCC 50505	✓	✓	✓	Human
Anncaliia algerae	✓	✓	✓	Mosquito
Edhazardia aedis	✓	✓	✓	Mosquito
Encephalitozoon cuniculi	–	–	–	Lemming
Encephalitozoon hellem	–	–	–	Rabbit
Encephalitozoon intestinalis	–	–	–	Rabbit
Encephalitozoon romelae	–	–	–	Grasshopper
Encephalitozoon bieneusi	–	–	–	Human
Enterocytozoon hepatopenaei	–	–	–	Shrimp
Enterospora canceri	–	–	–	Crab
Vavraia culicis floridensis	✓	✓	-	Mosquito

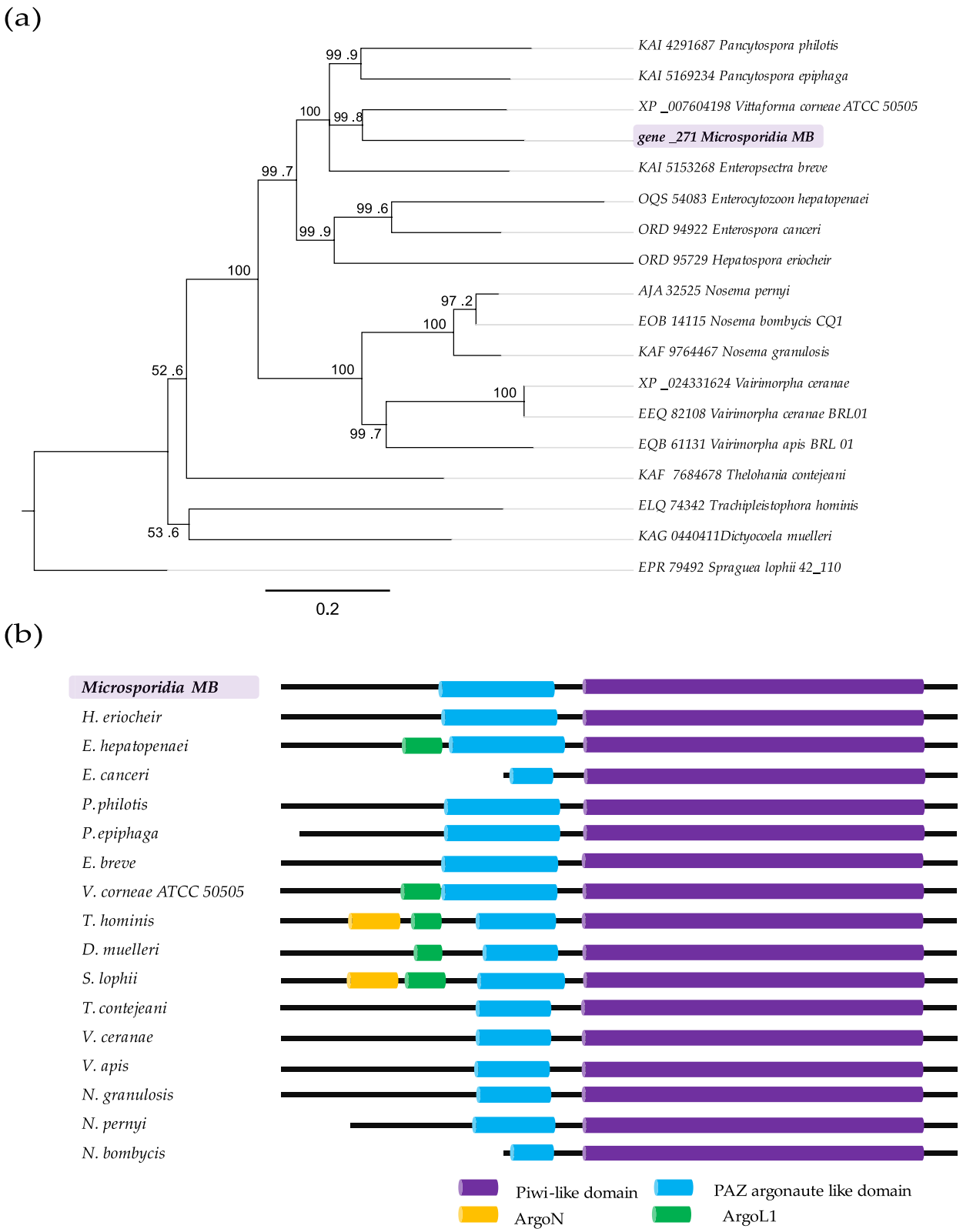

Fig. 11 Analysis of argonaute orthologs. **a** A rooted phylogenetic tree of argonaute orthologs based on 17 microsporidian species. The tree was constructed on protein sequences of the argonaute orthologs based on maximum clade credibility with Mr Bayes. Percent bootstrap values are indicated at the nodes. The scale for the branch length is shown at the bottom. **b** A schematic representation of the position of conserved domains Piwi-like and PAZ are mapped onto the multiple sequence alignment

the RdRP conserved region (Fig. 13b, Supplementary Fig. 4) across all species.

Conclusion

Microsporidia, a group of intracellular parasites, have been found to secrete a range of proteins into infected host cells, potentially influencing host metabolic processes and molecular programs [98]. The genome of the microsporidian parasite *E. cuniculi* contains a smaller number of protein families, reflecting the streamlining of its genome due to its adaptation to an intracellular parasitic lifestyle, with most orthologs being the smallest among eukaryotes [111]. These genomes consist of a set of host-exposed proteins that are rapidly evolving and belong to large, species-specific gene families, suggesting a common strategy for microsporidia to interact with their hosts [112].

Microsporidia sp. MB, a symbiont of the *Anopheles* mosquito, exhibits significant potential to interfere with malaria transmission by effectively inhibiting the development of Plasmodium parasites within its host [45, 113]. This study analyzes the key components of the *Microsporidia* sp. MB genome and focuses on its gene content and structure. This report demonstrates the close relationship between *Microsporidia* sp. MB and *V. corneae ATCC 50505* genomes among Clade IV of microsporidia consistent with previous reports on their taxonomic classification [45, 80, 113]. Notably, the symbiont has a larger genome size of 5.9 Mb compared to the rest of the species in the same clade, with the highest content of core BUSCO genes in the group. In silico methods have been used to annotate hypothetical proteins in *V. corneae*, revealing that many are involved in binding, enzymatic activity, and regulatory functions, despite its compact genome [91]. Here, a total of 2247 putative proteins were predicted and annotated. Most of these were further classified as enzymes involved in metabolic processes within the symbiont, notably similar to its sister *V. corneae*.

Microsporidia have been found to possess a range of metabolic pathways. These include polyamine metabolism, which is essential for their growth and development, and carbohydrate metabolism [114, 115]. Interestingly, some microsporidia have lost the ability to synthesize ATP and rely on host cell metabolism for energy [116]. Despite the absence of a recognizable mitochondrion, elements of mitochondrial metabolism, such as the pyruvate dehydrogenase complex, have been retained in these parasites [117]. Here, we analyzed the glycolytic pathway in *Microsporidia* sp. MB and compared this to other species in the Enterocytozoonida group. The Enterocytozoonida, apart from *V. corneae ATCC 50505*, have largely lost most of the proteins involved in this pathway except for hexokinase. This study demonstrates that *Microsporidia* sp. MB has retained all but one of the proteins involved in the glycolytic pathway complex like its closest relative, *V. corneae*. In addition to identifying a partial glycolytic pathway, which has retained 10 of the 11 key enzymes, we have also uncovered evidence for the functional presence of the pentose phosphate pathway (PPP). The PPP plays a compensatory role in generating NADPH for biosynthesis and redox balance, crucial for the survival of *Microsporidia* sp. MB in the absence of mitochondria. The interconnectedness of the glycolytic and PPP pathways allows this symbiont to adapt to its intracellular, host-dependent environment, supporting essential metabolic processes that contribute to its ability to block malaria transmission.

The *Microsporidia* sp. MB genome also showed evidence of ploidy through the identification of MRGs. Our analysis of the putative proteins shows that *Microsporidia* sp. MB harbors genes linked with a sexual mode of evolution (MRG). To further assert if this microorganism is a diploid or haploid, future analysis of the presence of regions with loss of heterozygosity (LOH) within this genome is necessary.

The study of RNA interference (RNAi) pathways in microsporidia highlights the evolutionary aspects of polyploidy. For instance, the retention or loss of Dicer and Argonaute orthologs across different microsporidian species suggests that these genes are selectively maintained or discarded during lineage divergence. This evolutionary flexibility might be influenced by polyploidy, affecting the parasite's adaptability and interaction with its host [79]. While some microsporidian species appear to retain elements of the RNAi pathway, the functional significance and mechanisms may vary considerably among different species. The presence of the Dicer, Argonaute and RNA-dependent RNA polymerase genes – the three key components of the RNAi pathway, in *Microsporidia* sp. MB is highlighted here. Despite having little information on

(See figure on next page.)

Fig. 12 Analysis of the dicer ortholog in *Microsporidia* sp. MB. **a** A rooted phylogenetic tree of argonaute orthologs based on 20 microsporidian species. The maximum likelihood tree was constructed using IQTREE. The scale for the branch length is shown at the bottom and bootstrap support values shown on each node as a percentage (1000 bootstraps). **b** A schematic representation of the position of conserved domains in the multiple sequence alignment is shown. An insert was observed between the RIBOc and Rnc domains in 4 of the species including *Microsporidia* sp. MB, *E. breve*, *P. epiphaga*, and *P. philotis*

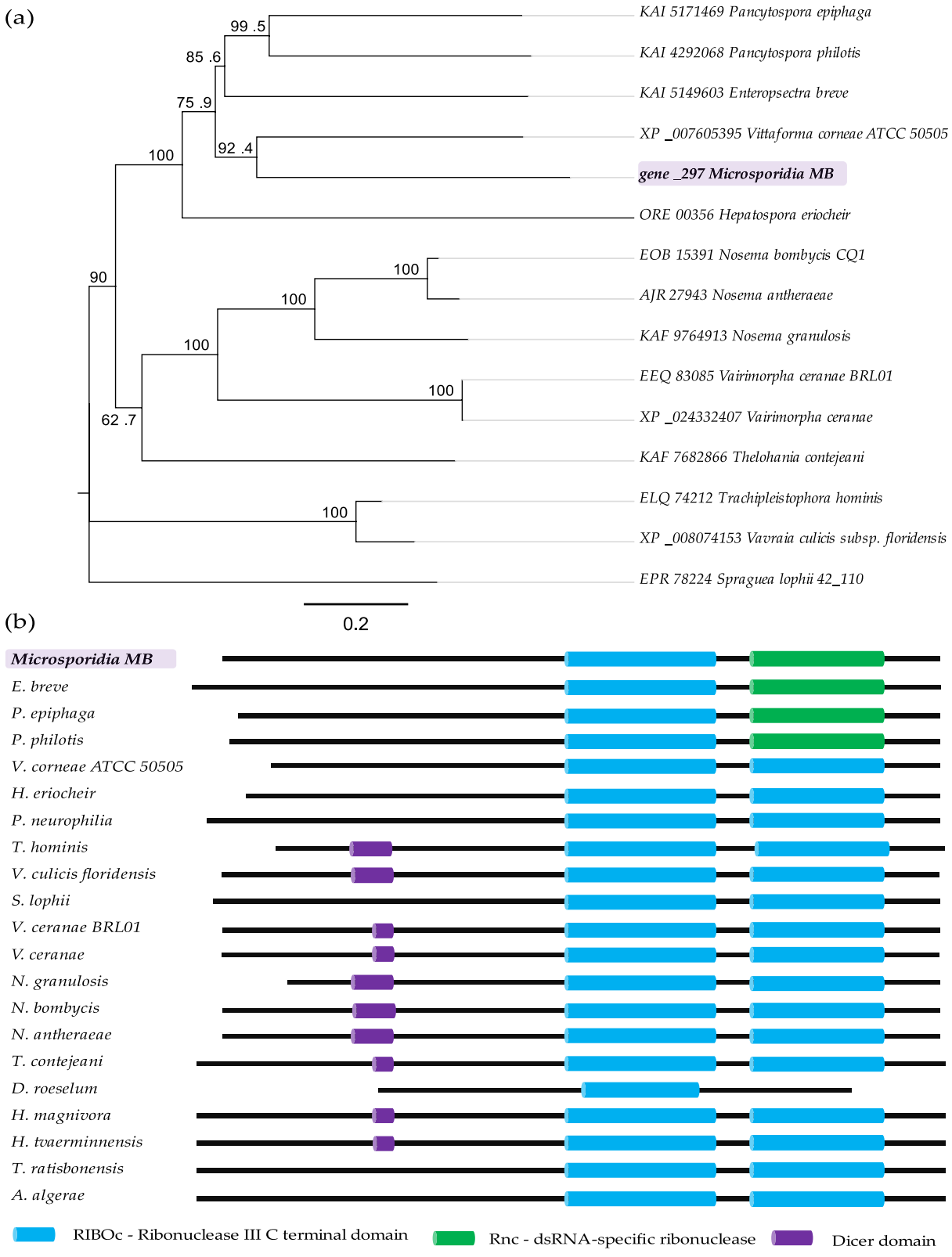

Fig. 12 (See legend on previous page.)

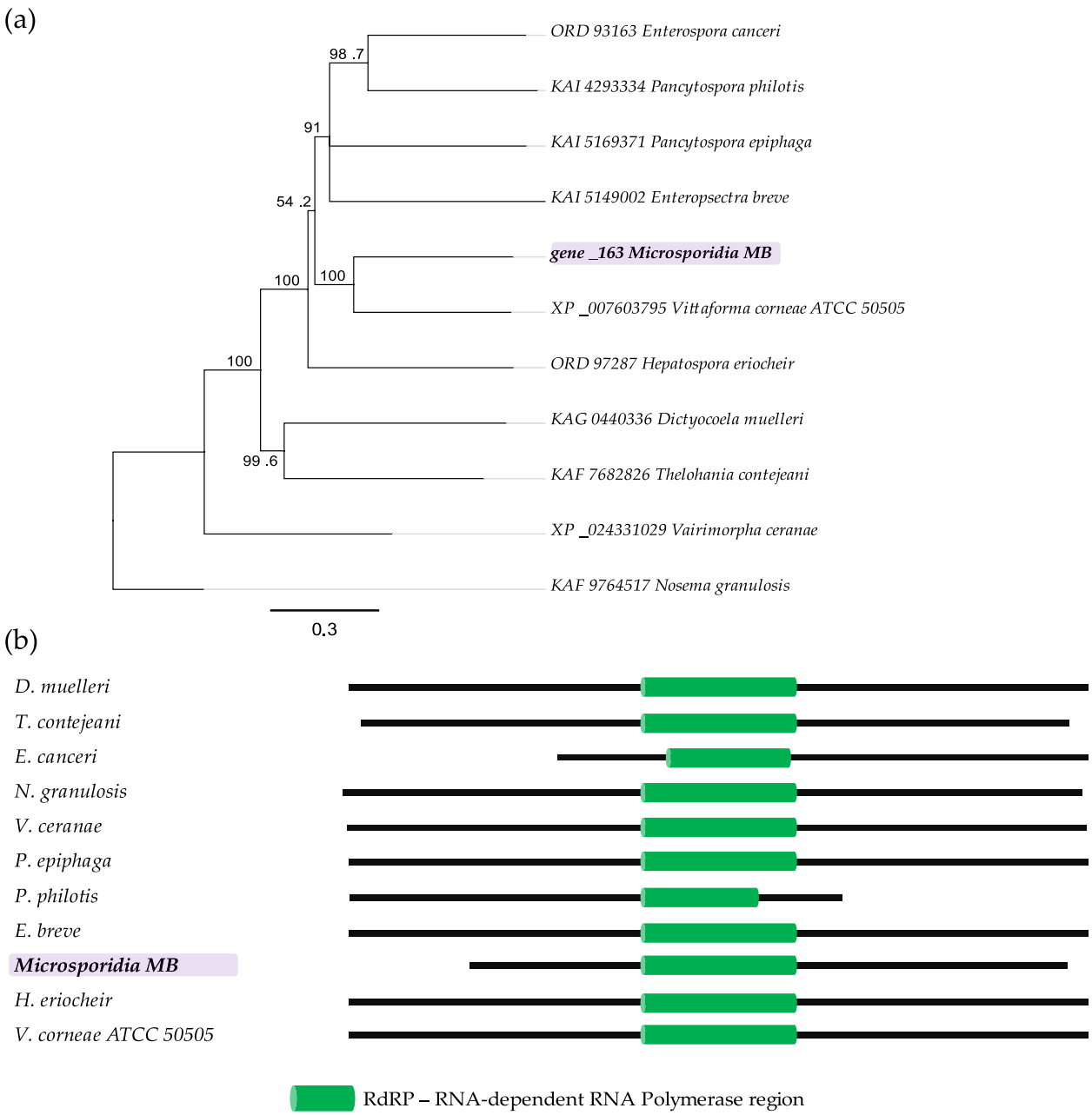

Fig. 13 Analysis of the RNA-dependent RNA polymerase in *Microsporidia* sp. MB. **a** Phylogenetic analysis. Tree construction was done using IQTREE with 1000 bootstraps. Bootstrap support values are shown on each node and the branch length scale is indicated at the bottom of the tree. **b** The identified conserved domain, RdRP, is mapped onto the schematic representation of the multiple sequence alignment

the functioning of this complex in most microsporidia, a study of the bee pathogen *N. ceranae* has in the past confirmed its regulatory functioning. Future studies of the RNAi machinery in *Microsporidia* sp. MB promise to provide more insight into its regulatory role and effect on the survival of the symbiont in the host.

This study provides a foundational analysis of the genome of *Microsporidia* sp. MB, but several limitations should be considered when interpreting the results. First, the absence of specific genes, such as pyruvate kinase (pyk) or 6-phosphogluconolactonase (PGLS), may not necessarily indicate their true absence from the genome. These absences could be due to gaps in the genome assembly, low sequencing depth in certain regions, or sequence divergence that complicates gene identification. Additionally, functional predictions based on homology

(e.g., from InterProScan and BLAST) are valuable for identifying potential protein functions but may not fully capture the biological roles of these proteins. Thus, experimental validation is needed to confirm the functionality of these predicted genes.

Furthermore, while our study highlights key metabolic pathways, such as glycolysis and the PPP, the relationship of these pathways in *Microsporidia* sp. MB remains speculative in some respects. The possibility of alternative metabolic pathways or compensatory mechanisms that were not detected in this genome assembly must be considered. Future studies utilizing more advanced genome sequencing technologies, such as long-read sequencing, and functional assays will be critical to address these gaps and provide a more comprehensive understanding of the biology of *Microsporidia* sp. MB.

Notwithstanding, annotating and analyzing the first reference genome of *Microsporidia* sp. MB is paramount in understanding its biology of infection and symbiosis within the *Anopheles* mosquito. The availability and analysis of these sequences act as a backdrop to important biological studies to understand the mechanism of this symbiont including optimizing the efficiency and stability of parasite transmission, understanding the ecological impact of introducing *Microsporidia* sp. MB into mosquito populations, and assessing any potential negative effects on non-target organisms. Altogether, this study lays the foundation for upcoming studies focusing on host-symbiont molecular interactions, strain variation analysis, and further implementation of RNASeq and other NGS tools in characterizing this important obligate intracellular organism.

Methods

A list of tools and resources used in this study are highlighted in Supplementary Table 1. Figure 14 further outlines the workflow for genome assembly, annotation, and comparative analysis of *Microsporidia* sp. MB adapted in this study.

Genome annotation – quality/completeness/repeats

Sample collection, preparation, genome sequencing, and genome assembly are described in [75]. Highly infected mosquito ovaries were sequenced using BGI DNB technology, followed by stringent quality assessment, de novo genome assembly and gene prediction. The quality of the genome was assessed using QUAST v5.0.2 [118]. Moreover, genome completeness estimation was confirmed using BUSCO v5.4.3 [119] against the microsporidia_odb10 set of 600 core genes. Repeat regions within the genome were identified and annotated using

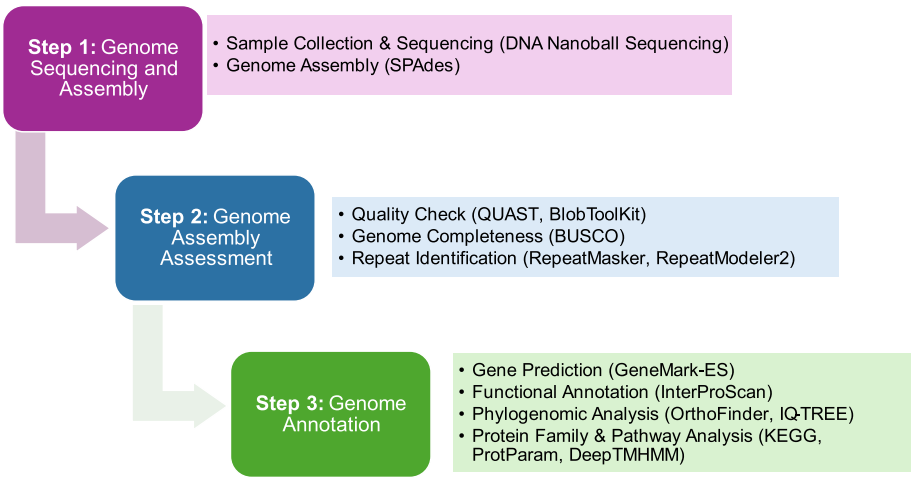

Fig. 14 Workflow for genome assembly, annotation, and comparative analysis of *Microsporidia* sp. MB. This flowchart outlines the step-by-step process used in the genome annotation pipeline. The workflow begins with sample collection and sequencing of Anopheles arabiensis mosquito ovaries using DNBSEQ technology, followed by genome assembly using SPAdes. The quality of the assembled genome was assessed using QUAST and visualized with BlobToolKit to check for contamination. Genome completeness was evaluated using BUSCO by identifying conserved single-copy orthologs specific to microsporidia. Repeat sequences in the genome were detected using RepeatMasker and RepeatModeler2 to annotate known and novel repeats. Gene prediction was performed using GeneMark-ES, optimized for eukaryotes with intronless genomes. Functional annotation of predicted genes was carried out with InterProScan, assigning proteins to families and domains from databases such as Pfam and CDD. Phylogenomic analysis of the genome was conducted using OrthoFinder for ortholog identification and IQ-TREE for maximum likelihood phylogenetic tree construction. Finally, predicted proteins were mapped to biological pathways using KEGG, and physicochemical properties, including transmembrane regions, were characterized using ProtParam and DeepTMHMM. This comprehensive workflow provides a detailed overview of the computational tools used in the experimental design and their roles in the analysis of the *Microsporidia* sp. MB genome

RepeatModeler2 [120] and RepeatMasker v4.0 [81]. BlobToolKit was used to visualize the genome structure [121].

Gene Prediction and Organization

GeneMark-ES v2.5p [122] was used for gene prediction with the intronless eukaryotic option. A comparison of BUSCO and GeneMark-ES output was used to determine common genes across both tools. Gene organizations across the contigs larger than or equal to 10 kb were visualized in RStudio 2023.09.1+494 [82] using the library package gggenes v0.5.1 [123].

Analysis of common orthogroups in *Microsporidia sp.* MB and the Enterocytozoonida group of *Microsporidia*

OrthoFinder v2.5.4 [124] and BUSCOPhylo v1.3 [125] was used to perform a BUSCO-based phylogenomic analysis of *Microsporidia* sp. MB and the clade IV microsporidia genomes. *Nosema granulosis* was used as an outgroup. Protein alignments were conducted using MUSCLE [83]. The maximum likelihood tree was inferred using IQTREE v1.6.12 [85] with the best-fit model LG+F+I+G4 chosen according to the Bayesian Information Criterion (BIC) using ModelFinder [86]. The tree was visualized using Geneious Prime® 2023.2.1 [126].

Protein characterization

Protein function analysis was done using the command line InterProScan tool [90] [accessed 04 October 2023] which includes annotations from multiple freely available tools including Pfam, CDD, Gene3D, HAMAP, MOBIDB, PANTHER, PIRSF, PRINTS, PROSITE, SFLD, SMART, SUPERFAMILY and NCBIFAM [127–138]. Gene Ontology (GO) analysis was performed by Blast2GO. The physicochemical properties of the putative proteins were determined using a batch ProtParam in-house script in Jupyter Lab [139]. Prediction of signal peptides and transmembrane regions was done using TargetP v2.0 [140], SignalP v6.0 [141], and DeepTMHMM v1.0.24 [103] respectively. The annotation of metabolic pathways was conducted using KEGG [102]. Visualization of the biological processes (BP), molecular functions (MF), cellular components (CC), and KEGG pathways was done using RStudio 2023.09.1+494 [82].

Homology modeling and motif analysis

Protein homology modeling was done using PRIMO (accessed 23 November 2023) as described in [99]. Briefly, HHPred [142] was used for the automatic identification of templates. This was followed by template selection where templates with higher resolution and optimal coverage were selected. Pairwise alignment of the target sequence to the template was done using T-COFFEE. Four models using the "Very slow" refinement method were designed. The quality of the models was assessed using multiple tools including the Qualitative Model Energy Analysis (QMEAN) Server, Protein Structure Analysis (ProSA-web) and Verify3D webserver [143–145]. Models with the best Z-dope score and high quality were selected.

LRR motifs were identified from the sequences using MEME and MAST online tools using default parameters [146, 147]. Briefly, MEME was set to identify motifs between 6 and 20 amino acids in length, with a focus on discovering the top three motifs across the sequences. The zoops model (zero or one occurrence per sequence) was applied for motif distribution. MAST was subsequently used to search for these motifs in the sequence database, using an E-value threshold of 10 to report significant matches.

Characterization of the glycolytic pathway

Microsporidia proteins involved in the glycolytic pathway were identified from past literature [46, 49, 76, 77, 148] and 12 *Encephalitozoon cuniculi* genes involved in this pathway were retrieved from NCBI. These included: pyruvate kinase (ECU09_0640), enolase (ECU10_1690), fructose 1,6-bisphosphate aldolase (ECU01_0240), phosphoglycerate mutase (ECU10_1060), hexokinase (ECU11_1540), pyruvate DHase E1, a sub (ODPA SCHPO) (ECU09_1040), pyruvate DHase E1, b sub (ODPB RICPR) (ECU04_1160), phosphofructokinase (ECU03_0680), phosphoglucose isomerase/glucose-6-phosphate isomerase (ECU05_0650), 3-phosphoglycerate kinase (ECU05_0320), glyceraldehyde-3-phosphate dehydrogenase (ECU07_0800) and triose phosphate isomerase (ECU11_0230).

A homology search of these sequences on NCBI BLASTp [149] was conducted (E-value = 0.001; bit-score ≥ 200) and the top 10 hits for each protein were retrieved. A reverse BLAST of the top 10 hits was used to confirm homology and all hits which included the query sequence were retained. A custom local database was created using the BLAST+makeblastdb command [150] and the retained sequences which included proteins from the Enterocytozoon consisting of *V. corneae ATCC 50505, E. bieneusi, E. hepatopenaei, E. canceri, H. eriocheir* and other microsporidia such as *N. granulosis, V. ceranae, A. locustae, O. bayeri*. Putative proteins from the genome assembly were blasted against this local database (E-value = 0.001; bit-score ≥ 200). *Microsporidia* sp. MB genes that were identified to be homologous to any sequence within the custom glycolysis database were selected and a further homology search against the remote BLASTp was done to confirm identity.

Table 5 Pentose phosphate pathway (PPP) genes identified in *Microsporidia* sp. MB

Sequence Name	Top BLAST Hits Description	Length	E-Value	GO IDs	GO Names
NODE_723_length_4854_cov_63.1801	XP_007604969.1hypothetical protein VICG_01523	4854	6.04E-111	P:GO:0005975; P:GO:0006006; P:GO:0006098; F:GO:0004345; F:GO:0016614; F:GO:0050661	P:carbohydrate metabolic process; P:glucose metabolic process; P:pentose-phosphate shunt; F:glucose-6-phosphate dehydrogenase activity; F:oxidoreductase activity, acting on CH-OH group of donors; F:NADP binding
NODE_1037_length_3521_cov_100.817	XP_007605214.1hypothetical protein VICG_01769	3521	4.51E-275	F:GO:0016740	F:transferase activity
NODE_1109_length_3325_cov_75.7963	KAI5175593.1ribose 5-phosphate isomerase A, partial	3325	5.62E-43	P:GO:0006098; P:GO:0009052; F:GO:0004751; F:GO:0016853	P:pentose-phosphate shunt; P:pentose-phosphate shunt, non-oxidative branch; F:ribose-5-phosphate isomerase activity; F:isomerase activity
NODE_1158_length_3204_cov_64.6307	XP_007605146.16-phosphogluconate dehydrogenase, decarboxylating	3204	8.38E-163	P:GO:0006098; P:GO:0019521; F:GO:0004616; F:GO:0016491; F:GO:0050661	P:pentose-phosphate shunt; P:D-gluconate metabolic process; F:phosphogluconate dehydrogenase (decarboxylating) activity; F:oxidoreductase activity; F:NADP binding

Multiple sequence alignment of each set of proteins was done using ClustalO. MEGAX was used to identify the best-fit model for phylogenetic analysis for each protein set. Tree calculations were done using the best model using ModelFinder and visualization was done in Geneious Prime® 2023.2.1.

Key conserved domains in the putative genes predicted to be involved in the glycolytic pathway were identified on the Batch Web CD-Search tool (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) (accessed 03 October 2023) (E-value = 0.001).

In addition to the partial glycolytic pathway identified in *Microsporidia* sp. MB, evidence suggests that the pentose phosphate pathway (PPP) may play a compensatory role in the organism's metabolism. The PPP is closely linked to glycolysis and provides crucial metabolic intermediates while also generating NADPH, which is vital for biosynthetic reactions and maintaining redox balance, particularly in organisms like *Microsporidia* sp. MB that lack mitochondria.

Several key enzymes of the PPP were identified in the genome of *Microsporidia* sp. MB, located on specific scaffolds within the genome assembly (Table 5 below). This table lists the scaffolds (nodes) containing predicted genes involved in the pentose phosphate pathway. The table includes descriptions, sequence lengths, enzyme functions, Gene Ontology (GO) terms, and relevant enzyme activities. Key enzymes such as glucose-6-phosphate dehydrogenase (G6PD), 6-phosphogluconate dehydrogenase (PGD), transketolase (TKT), and ribose-5-phosphate isomerase (RPI), which contribute to both the oxidative and non-oxidative branches of the PPP, are highlighted. These enzymes play essential roles in generating NADPH and intermediates necessary for biosynthesis and redox balance in *Microsporidia* sp. MB.

Notably, glucose-6-phosphate dehydrogenase (G6PD), which catalyzes the first step of the oxidative branch of the PPP, was found on NODE_723 (GO:0004345). This enzyme converts glucose-6-phosphate into 6-phosphoglucono-δ-lactone, producing NADPH in the process, a critical function for maintaining cellular redox balance [151]. Additionally, 6-phosphogluconate dehydrogenase (PGD), responsible for converting 6-phosphogluconate to ribulose-5-phosphate in the oxidative branch, was identified on NODE_1158 (GO:0004616). This enzyme further supports the generation of NADPH for biosynthesis and oxidative stress response.

The non-oxidative branch of the PPP also appears to be functional in *Microsporidia* sp. MB. Transketolase (TKT), found on NODE_1037 (GO:0004802), facilitates the transfer of two-carbon units between sugar phosphates, regenerating glucose-6-phosphate and providing intermediates for nucleotide biosynthesis. Additionally, ribose-5-phosphate isomerase (RPI), located on NODE_1109 (GO:0004751), catalyzes the interconversion of ribose-5-phosphate and ribulose-5-phosphate, a critical step for ensuring the availability of ribose for nucleotide synthesis [152].

However, 6-phosphogluconolactonase (PGLS), which is responsible for converting 6-phosphoglucono-δ-lactone into 6-phosphogluconate, was not identified in the current genome annotation. This gap suggests either incomplete annotation or the possibility that *Microsporidia* sp. MB uses alternative pathways to process this intermediate. Despite this, the presence of key PPP enzymes indicates that the pathway is likely functional and plays a compensatory role in *Microsporidia* sp. MB's metabolism, supplementing the partial glycolytic pathway.

Together, the glycolytic and pentose phosphate pathways form an interconnected system that enables *Microsporidia* sp. MB to adapt to its host-dependent intracellular environment. These pathways provide metabolic flexibility, ensuring the production of essential biosynthetic precursors and redox cofactors that are crucial for survival [95].

Inference of the putative mode of reproduction in *Microsporidia sp.* MB: characterization of Meiotic-Related Genes (MRG)

A set of MRGs were identified from a vast literature search using keywords [MRG + Microsporidia] on Google Scholar (https://scholar.google.com/ accessed 3 October 2023). Nineteen microsporidian MRGs were documented [106] and an NCBI search was used to retrieve these sequences. A custom database was created using the retrieved sequences and a local blast search of these genes against the custom database was conducted as described above (E-value = 0.0001). Multiple sequence alignments and tree calculations were done as described in the section above.

Analysis of the RNA interference mechanism in *Microsporidia sp.* MB

Microsporidia exhibit oxidative stress response mechanisms, including the expression of antioxidant enzymes, which help neutralize reactive oxygen species generated within the host environment. The Dicer, RNA-dependent RNA polymerase, and Argonaute genes have been well-documented to be involved in this system [79]. Their respective orthologs were retrieved from MicrosporidiaDB [153] and homologs of the RNAi complex were retrieved from NCBI (accessed 4 October 2023). Search and analysis of the homologs from the *Microsporidia* sp. MB genome was done as described in the sections above. Moreover, these genes are known to contain unique

conserved domains including PAZ Argonaute-like, DUF, and PIWI-like domains. Therefore, a CD-Search on the *Microsporidia* sp. MB genes homologous to Argonaute and Dicer genes was done using default parameters to determine the presence of these key domains (E-value = 0.001).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-11046-y.

Additional file 1: Supplementary Data File 1. Contains the detailed distribution of repeat sequences identified in the Microsporidia sp. MB genome using RepeatMasker, highlighting the various types of repeats, their positions, and sequence characteristics.

Additional file 2: Supplementary Data File 2. The full GenBank file for Microsporidia sp. MB, including annotated coding sequences (CDS), contig information, and gene features.

Additional file 3: Supplementary Data File 3. A comprehensive file of predicted gene sequences in FASTA format. This includes the nucleotide sequences of all predicted genes identified using GeneMark-ES and BUSCO.

Additional file 4: Supplementary Data File 4. A file containing all predicted protein sequences in FASTA format. These sequences correspond to the predicted coding regions and were used for functional annotation and pathway analysis.

Additional file 5: Supplementary Data File 5. BLAST results and protein annotation summary. This Excel file provides BLASTp results for the predicted proteins, including homologous matches, E-values, and functional descriptions from protein databases.

Additional file 6: Supplementary Data File 6. A detailed classification of protein families based on InterProScan annotations. This file provides the functional categories of the predicted proteins and lists their associated protein domains and superfamilies.

Additional file 7: Supplementary Table 1. A list of bioinformatic tools and resources used in the genome assembly and annotation of Microsporidia MB. Supplementary Table 2. Subcellular localization prediction of MB Leucine-rich repeat proteins using SignalP, TargetP and DeepTMHMM. Supplementary Table 3. Distribution of signal peptides, transmembrane regions and GRAVY scores in selected Microsporidia sp. MB. Supplementary Fig. 1. (a) Multiple sequence alignment of the leucine-rich repeat (LRR) domain of APL1 and Microsporidia sp. MB LRR gene_679. Conservation of amino acids across sequences is visualized with color coding, where green represents highly conserved residues and yellow indicates moderate conservation. This alignment highlights structurally important regions that may contribute to functional similarities between the two proteins. (b) 3D model of APL1 as visualized on the RCSB Protein Data Bank (PDB) website PDB ID: 3O6N , showcasing the spatial arrangement of the LRR domains, which are critical for protein–protein interactions. Supplementary Fig. 2. Multiple sequence alignment of Argonaute orthologs in 18 species and mapping of conserved PAZ and Piwi-like domains. Supplementary Fig. 3. Multiple sequence alignment of the Dicer gene in Microsporidia contains 2 main conserved domains. Supplementary Fig. 4. Multiple sequence alignment of the RNA-dependent RNA Polymerase (RdRP) ortholog in Microsporidia, containing the RdRP domain.

Acknowledgements
We would like to extend our sincere gratitude to the Centre for High Performance Computing (CHPC) for providing the computational resources essential for our research. Special thanks go to Rhodes University for its unwavering support and facilities that made this study possible. We also acknowledge the invaluable contributions of the Symbiovector Group at the International Centre of Insect Physiology and Ecology (ICIPE). Their expertise and collaboration have been instrumental in advancing our understanding of Microsporidia sp. MB and its potential role in malaria transmission dynamics.

Authors' contributions
Conceptualization, O.T.B., J.K.H. and L.M.A.; methodology, O.T.B., J.K.H., and L.M.A.; software, O.T.B., J.K.H. and L.M.A.; validation, O.T.B., J.K.H. and L.M.A; formal analysis, O.T.B., J.K.H. and L.M.A; investigation, O.T.B., J.K.H. and L.M.A; resources, O.T.B. and J.K.H.; data curation, O.T.B., J.K.H., and L.M.A; writing—original draft preparation, O.T.B., J.K.H. and L.M.A; writing—review and editing, O.T.B., J.K.H., and L.M.A; visualization, O.T.B., J.K.H. and L.M.A; supervision, O.T.B. and J.K.H. project administration, O.T.B. and J.K.H. funding acquisition, J.K.H. and O.T.B. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
The authors acknowledge funding from the Organization for Women in Science for the Developing World (OWSD); the Swedish International Development Cooperation Agency (SIDA); Open Philanthropy (SYMBIOVECTOR Track A); the Bill and Melinda Gates Foundation (INV0225840); the Swiss Agency for Development and Cooperation (SDC); the Australian Centre for International Agricultural Research (ACIAR); the Federal Democratic Republic of Ethiopia; and the Government of the Republic of Kenya. The views expressed herein do not necessarily reflect the official opinion of the donors.

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 August 2024 Accepted: 13 November 2024
Published online: 22 November 2024

References

1. World Health Organization. World malaria report 2022. 2023.
2. Whittaker C, Hamlet A, Sherrard-Smith E, Winskill P, Cuomo-Dannenburg G, Walker PGT, et al. Seasonal dynamics of Anopheles stephensi and its implications for mosquito detection and emergent malaria control in the Horn of Africa. Proc Natl Acad Sci. 2023;120:1–9.
3. Ochomo EO, Milanoi S, Abong'o B, Onyango B, Muchoki M, Omoke D, et al. Molecular surveillance leads to the first detection of Anopheles stephensi in Kenya. Res Sq. 2023;v2:1–16. https://doi.org/10.21203/rs.3.rs-2498485/v2.
4. Ojuka P, Boum Y, Denoeud-Ndam L, Nabasumba C, Muller Y, Okia M, et al. Early biting and insecticide resistance in the malaria vector Anopheles might compromise the effectiveness of vector control intervention in Southwestern Uganda. Malar J. 2015. https://doi.org/10.1186/s12936-015-0653-z.
5. Kleinschmidt I, Bradley J, Knox TB, Mnzava AP, Kafy HT, Mbogo C, et al. Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. Lancet Infect Dis. 2018;18:640–9.
6. Martinez J, Longdon B, Bauer S, Chan YS, Miller WJ, Bourtzis K, et al. Symbionts Commonly Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of Wolbachia Strains. PLoS Pathog. 2014;10:e1004369.
7. Harumoto T, Anbutsu H, Lemaitre B, Fukatsu T. Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. Nat Commun. 2016;7:1–12.

8. Lukasik P, Guo H, Van Asch M, Ferrari J, Godfray HCJ. Protection against a fungal pathogen conferred by the aphid facultative endosymbionts Rickettsia and Spiroplasma is expressed in multiple host genotypes and species and is not influenced by co-infection with another symbiont. J Evol Biol. 2013;26:2654–61.
9. Andreadis TG. Polymorphic microsporidia of mosquitoes: potential for biological control. UCLA Symp Mol Cell Biol. 1990;112:177–88.
10. Bjørnson S, Oi D. Microsporidia Biological Control Agents and Pathogens of Beneficial Insects. In: Weiss, L.M., & Becnel, J.J. (Eds.), Microsporidia: Pathogens of Opportunity: First Edition. 2014:635–70. https://doi.org/10.1002/9781118395264.ch25.
11. Dale C, Moran NA. Molecular Interactions between Bacterial Symbionts and Their Hosts. Cell. 2006;126:453–65.
12. McMeniman CJ, Hughes GL, O'Neill SL. A Wolbachia symbiont in Aedes aegypti disrupts mosquito egg development to a greater extent when mosquitoes feed on nonhuman versus human blood. J Med Entomol. 2011;48:76–84.
13. Whitten MMA, Facey PD, Sol R Del, Evans MC, Mitchell JJ, Bodger OG, et al. Symbiont-mediated RNA interference in insects. Proc R Soc B. 2016;283:1–9.
14. Ferguson NM, Kien DTH, Clapham H, Aguas R, Trung VT, Chau TNB, et al. Modeling the impact on virus transmission of Wolbachia-mediated blocking of dengue virus infection of Aedes aegypti. Sci Transl Med. 2015. https://doi.org/10.1126/scitranslmed.3010370.
15. Bourtzis K, Dobson SL, Xi Z, Rasgon JL, Calvitti M, Moreira LA, et al. Harnessing mosquito-Wolbachia symbiosis for vector and disease control. Acta Trop. 2014. https://doi.org/10.1016/j.actatropica.2013.11.004.
16. Ricci I, Valzano M, Ulissi U, Epis S, Cappelli A, Favia G. Symbiotic control of mosquito borne disease. Pathog Glob Health. 2012;106:380–5.
17. Ye YH, Carrasco AM, Frentiu FD, Chenoweth SF, Beebe NW, van den Hurk AF, et al. Wolbachia reduces the transmission potential of dengue-infected Aedes aegypti. PLoS Negl Trop Dis. 2015. https://doi.org/10.1371/journal.pntd.0003894.
18. Hoffmann AA, Montgomery B, Popovici J, Iturbe-Ormaetxe I, Johnson P, Muzzi F, et al. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature. 2011;476:454–7.
19. Rainey SM, Shah P, Kohl A, Dietrich I. Understanding the wolbachia-mediated inhibition of arboviruses in mosquitoes: Progress and challenges. J Gen Virol. 2014;95:517–30.
20. Lewis LC, Bruck DJ, Prasifka JR, Raun ES. Nosema pyrausta: Its biology, history, and potential role in a landscape of transgenic insecticidal crops. Biol Control. 2009;48:223–31.
21. Wittner M, Weiss LM. The Microsporidia and Microsporidiosis. 1999.
22. Keohane EM, Weiss LM. Characterization and function of the microsporidian polar tube: A review. Folia Parasitol (Praha). 1998;45:117–27.
23. Chen Y, Lv Q, Liao H, Xie Z, Hong L, Qi L, et al. The microsporidian polar tube: origin, structure, composition, function, and application. Parasit Vectors. 2023;16:1–11.
24. Corradi N, Slamovits CH. The intriguing nature of microsporidian genomes. Brief Funct Genomics. 2011;10:115–24.
25. Corradi N, Pombert JF, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian Encephalitozoon intestinalis. Nat Commun. 2010;1:1–7.
26. Vávra J, Lukeš J. Microsporidia and 'The Art of Living Together.' Adv Parasitol. 2013;82:253–319.
27. Becnel JJ, Andreadis TG. Microsporidia in Insects. In: Weiss LM, Becnel JJ. (Eds.), Microsporidia: Pathogens of Opportunity: First Edition. 2014:521–70. https://doi.org/10.1002/9781118395264.ch21.
28. Shang Y, Feng P, Wang C. Fungi That Infect Insects: Altering Host Behavior and Beyond. PLoS Pathog. 2015;11:1–6.
29. Pan G, Bao J, Ma Z, Song Y, Han B, Ran M, et al. Invertebrate host responses to microsporidia infections. Dev Comp Immunol. 2018;83:104–13.
30. Cormier A, Chebbi MA, Giraud I, Wattier R, Teixeira M, Gilbert C, et al. Comparative Genomics of Strictly Vertically Transmitted, Feminizing Microsporidia Endosymbionts of Amphipod Crustaceans. Genome Biol Evol. 2021;13:1–13.
31. Koella JC, Agnew P. Blood-feeding success of the mosquito Aedes aegypti depends on the transmission route of its parasite Edhazardia aedis. Oikos. 1997;78:311–6.
32. Agnew P, Becnel JJ, Ebert D, Michalakis Y. Symbiosis of Microsporidia and Insects. In: Insect Symbiosis. 2003. p. 145–63.
33. Andreadis TG. Microsporidian Parasites of Mosquitoes. J Am Mosq Control Assoc. 2007;23:3–29.
34. Trzebny A, Mizera J, Dabert M. Microsporidians (Microsporidia) parasitic on mosquitoes (Culicidae) in central Europe are often multi-host species. J Invertebr Pathol. August 2022;2023(197): 107873.
35. Becnel JJ, White SE, Shapiro AM. Review of microsporidia-mosquito relationships: From the simple to the complex. Folia Parasitol (Praha). 2005;52:41–50.
36. Andreadis TG. Amblyospora connecticus sp. nov. (Microsporida: Amblyosporidae): Horizontal transmission studies in the mosquito Aedes cantator and formal description. J Invertebr Pathol. 1988;52:90–101.
37. Undeen AH, Alger E. The Effect of the Microsporidan, Nosema on Anopheles stephensi. J Invertebr Pathol. 1975;24:19–24.
38. Avery SW, Undeen AH. Horizontal Transmission of Parathelohania anophelis to the Anopheles quadrimaculatus ". J Invertebr Pathol. 1990;56:98–105.
39. Bargielowski I, Koella JC. A possible mechanism for the suppression of Plasmodium berghei development in the mosquito Anopheles gambiae by the Microsporidian Vavraia culicis. PLoS One. 2009;4:e4676.
40. Herren JK, Mbaisi L, Mararo E, Makhulu EE, Mobegi VA, Butungi H, et al. A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. Nat Commun. 2020;11:1–10.
41. Akorli J, Akorli EA, Tetteh SNA, Amlalo GK, Opoku M, Pwalia R, et al. Microsporidia MB is found predominantly associated with Anopheles gambiae s.s and Anopheles coluzzii in Ghana. Sci Rep. 2021;11:1–5.
42. Nattoh G, Maina T, Makhulu EE, Mbaisi L, Mararo E, Otieno FG, et al. Horizontal Transmission of the Symbiont Microsporidia MB in Anopheles arabiensis. Front Microbiol. 2021;12:1–13.
43. Nattoh G, Onyango B, Makhulu E, Omoke D, Mbaisi L, Kamau L, et al. Microsporidia MB in the primary malaria vector Anopheles gambiae s.s. is avirulent and undergoes maternal and horizontal transmission. Parasit Vectors. 2023;16:1–19.
44. Moustapha LM, Sadou IM, Arzika II, Maman LI, Gomgnimbou MK, Konkobo M, et al. First identification of Microsporidia MB in Anopheles coluzzii from Zinder City, Niger. Parasit Vectors. 2024;17:1–5.
45. Bukhari T, Pevsner R, Herren JK. Microsporidia: a promising vector control tool for residual malaria transmission. Front Trop Dis. 2022;3 July:1–18.
46. Vivarès CP, Méténier G. The microsporidia genome: Living with minimal genes as an intracellular eukaryote. Opportunistic Infections: Toxoplasma, Sarcocystis, and Microsporidia. 2004;9:215–42.
47. Xu J, Pan G, Fang L, Li J, Tian X, Li T, et al. The varying microsporidian genome: Existence of long-terminal repeat retrotransposon in domesticated silkworm parasite Nosema bombycis. Int J Parasitol. 2006;36:1049–56.
48. Pombert JF, Haag KL, Beidas S, Ebert D, Keeling PJ. The Ordospora colligata genome: Evolution of extreme reduction in microsporidia and host-to-parasite horizontal gene transfer. mBio. 2015;6:e02400–14.
49. Keeling PJ, Corradi N. Balancing loss of function with shrinking genomes in the microsporidia. Virulence. 2011;2:67–70.
50. Parisot N, Pelin A, Gasc C, Polonais V, Belkorchia A, Panek J, et al. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. Genome Biol Evol. 2014;6:2289–300.
51. 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. Proc Natl Acad Sci U S A. 2012;109:12638–43.
52. Peyretaillade E, Boucher D, Parisot N, Gasc C, Butler R, Pombert JF, et al. Exploiting the architecture and the features of the microsporidian genomes to investigate diversity and impact of these parasites on ecosystems. Heredity (Edinb). 2015;114:441–9.
53. Corradi N, Selman M. Latest progress in microsporidian genome research. J Eukaryot Microbiol. 2013;60:309–12.
54. Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg A V., et al. The Genome of the Obligate Intracellular Parasite Trachipleistophora hominis: New Insights into Microsporidian Genome Dynamics and Reductive Evolution. PLoS Pathog. 2012;8:e1002979.

55. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res. 2012;22:2478–88.
56. Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ. Draft genome sequence of the Daphnia pathogen Octosporea bayeri: Insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. Genome Biol. 2009;10:1–12.
57. Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, et al. Reduction and expansion inmicrosporidian genome evolution: New insights from comparative genomics. Genome Biol Evol. 2013;5:2285–303.
58. Žárský V, Karnkowska A, Boscaro V, Trznadel M, Whelan TA, Hiltunen-Thorén M, et al. Contrasting outcomes of genome reduction in mikrocytids and microsporidians. BMC Biol. 2023;21:1–11.
59. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM, et al. The reduced genome of the parasitic microsporidian Enterocytozoon bieneusi lacks genes for core carbon metabolism. Genome Biol Evol. 2010;2:304–9.
60. Campbell SE, Williams TA, Yousuf A, Soanes DM, Paszkiewicz KH, Williams BAP. The Genome of Spraguea lophii and the Basis of Host-Microsporidian Interactions. PLoS Genet. 2013;9:1–15.
61. Huang Q, Wu ZH, Li WF, Guo R, Xu JS, Dang XQ, et al. Genome and Evolutionary Analysis of Nosema ceranae: A Microsporidian Parasite of Honey Bees. Front Microbiol. 2021;12 June:1–17.
62. Quandt CA, Beaudet D, Corsaro D, Walochnik J, Michel R, Corradi N, et al. The genome of an intranuclear parasite, Paramicrosporidium saccamoebae, reveals alternative adaptations to obligate intracellular parasitism. Elife. 2017;6:1–19.
63. Williams BAP, Lee RCH, Becnel JJ, Weiss LM, Fast NM, Keeling PJ. Genome sequence surveys of Brachiola algerae and Edhazardia aedis reveal microsporidia with low gene densities. BMC Genomics. 2008;9:1–9.
64. Chen L, Gao X, Li R, Zhang L, Huang R, Wang L, et al. Complete genome of a unicellular parasite (Antonospora locustae) and transcriptional interactions with its host locust. Microb Genom. 2020;6:1–12.
65. Corradi N, Akiyoshi DE, Morrison HG, Feng X, Weiss LM, Tzipori S, et al. Patterns of genome evolution among the Microsporidian parasites Encephalitozoon cuniculi, Antonospora locustae and enterocytozoon bieneusi. PLoS One. 2007;2:e1277.
66. Pan G, Xu J, Li T, Xia Q, Liu SL, Zhang G, et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. BMC Genomics. 2013;14:1–14.
67. Peters MJ, Suwannapong G, Pelin A, Corradi N. Genetic and Genome Analyses Reveal Genetically Distinct Populations of the Bee Pathogen Nosema ceranae from Thailand. Microb Ecol. 2019;77:877–89.
68. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nat Commun. 2015;6 May:7121.
69. Pombert JF, Xu J, Smith DR, Heiman D, Young S, Cuomo CA, et al. Complete genome sequences from three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian Encephalitozoon cuniculi. Eukaryot Cell. 2013;12:503–11.
70. Chen Y ping, Pettis JS, Zhao Y, Liu X, Tallon LJ, Sadzewicz LD, et al. Genome sequencing and comparative genomics of honey bee microsporidia, Nosema apis reveal novel insights into host-parasite interactions. BMC Genomics. 2013;14:1–16.
71. Pelin A, Moteshareie H, Sak B, Selman M, Naor A, Eyahpaise ME, et al. The genome of an Encephalitozoon cuniculi type III strain reveals insights into the genetic diversity and mode of reproduction of a ubiquitous vertebrate pathogen. Heredity (Edinb). 2016;116:458–65.
72. Katinka MD, Duprat S, Cornillott E, Méténier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature. 2001;414:450–3.
73. Slamovits CH, Fast NM, Law JS, Keeling PJ. Genome Compaction and Stability in Microsporidian Intracellular Parasites. Curr Biol. 2004;14:891–6.
74. Peyretaillade E, Akossi RF, Tournayre J, Delbac F, Wawrzyniak I. How to overcome constraints imposed by microsporidian genome features to ensure gene prediction? J Eukaryotic Microbiol. 2024;71:e13038–15.
75. Ang'ang'o LM, Waweru JW, Makhulu EE, Wairimu A, Otieno FG, Onchuru T, et al. Draft genome of Microsporidia sp. MB—a malaria-blocking microsporidian symbiont of the Anopheles arabiensis. Microbiol Resour Announc. 2024;13:e00903–23.
76. Wiredu Boakye D, Jaroenlak P, Prachumwat A, Williams TA, Bateman KS, Itsathitphaisarn O, et al. Decay of the glycolytic pathway and adaptation to intranuclear parasitism within Enterocytozoonidae microsporidia. Environ Microbiol. 2017;19:2077–89.
77. Luo J, He Q, Xu JZ, Xu C, Han YZ, Gao HL, et al. Microsporidia infection upregulates host energy metabolism but maintains ATP homeostasis. J Invertebr Pathol. 2021;186 April:107596.
78. Paldi N, Glick E, Oliva M, Zilberberg Y, Aubin L, Pettis J, et al. Effective gene silencing in a microsporidian parasite associated with honeybee (Apis mellifera) colony declines. Appl Environ Microbiol. 2010;76:5960–4.
79. Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. Infection, Genetics and Evolution. 2018;65 August:329–32.
80. Park E, Poulin R. Revisiting the phylogeny of microsporidia. Int J Parasitol. 2021;51 xxxx:855–64.
81. Nishimura D. RepeatMasker. Biotech Software & Internet. Report. 2000;1:1–4.
82. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. https://doi.org/10.1136/bmj.s4-1.148.879-a.
83. Edgar RC, Drive RM, Valley M. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
84. Capella-Gutiérrez S, Silla-Mart\'\inez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
85. Nguyen K, Guo X, Pan Y. Phylogeny in Multiple Sequence Alignments. In: Multiple Biological Sequence Alignment: Scoring Functions, Algorithms and Applications. London: Wiley; 2016. p. 103–12.
86. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
87. Bojko J, Frizzera A, Vázquez N, Taylor G, Rand V, Cremonte F. Comparative genomics for Agmasoma sp (Microsporidia) parasitising invasive Carcinus aestuarii and Carcinus maenas in Argentina. J Invertebr Pathol. 2023;198 February:107908.
88. Wadi L, El Jarkass HT, Tran TD, Islah N, Luallen RJ, Reinke AW. Genomic and phenotypic evolution of nematode-infecting microsporidia. PLoS Pathog. 2023;19:1–29.
89. Stratton CE, Reisinger LS, Behringer DC, Reinke AW, Bojko J. Alternosema astaquatica n. sp. (Microsporidia: Enterocytozoonida), a systemic parasite of the crayfish Faxonius virilis. J Invertebr Pathol. 2023;199:1–10.
90. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49:D344–54.
91. Ang'ang'o LM, Herren JK, Tastan Bishop Ö. Structural and Functional Annotation of Hypothetical Proteins from the Microsporidia Species Vittaforma corneae ATCC 50505 Using in silico Approaches. Int J Mol Sci. 2023;24:1–27.
92. Dolgikh V V. Activities of enzymes of carbohydrate and energy metabolism of the intracellular stages of the microsporidian, Nosema grylli. 2000.
93. Goldberg AV, Molik S, Tsaousis AD, Neumann K, Kuhnke G, Delbac F, et al. Localization and functionality of microsporidian iron-sulphur cluster assembly proteins. Nature. 2008;452:624–8.
94. Garcia PS, Barras F, Gribaldo S. Components of iron–Sulfur cluster assembly machineries are robust phylogenetic markers to trace the origin of mitochondria and plastids. PLoS Biol. 2023;21 11 November.
95. Keeling PJ, Fast NM. Microsporidia: Biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol. 2002;56:93–116.
96. Dean P, Hirt RP, Embley TM. Microsporidia: Why Make Nucleotides if You Can Steal Them? PLoS Pathog. 2016;12:1–13.
97. Alexander WG, Wisecaver JH, Rokas A, Hittinger CT. Horizontally acquired genes in early-diverging pathogenic fungi enable the use of host nucleosides and nucleotides. Proc Natl Acad Sci U S A. 2016;113:4116–21.

98. Senderskiy I V., Timofeev SA, Seliverstova E V., Pavlova OA, Dolgikh VV. Secretion of Antonospora (Paranosema) locustae proteins into infected cells suggests an active role of microsporidia in the control of host programs and metabolic processes. PLoS One. 2014;9:1–12. https://doi.org/10.1371/journal.pone.0093585.
99. Hatherley R, Brown DK, Glenister M, Bishop ÖT. PRIMO: An interactive homology modeling pipeline. PLoS ONE. 2016;11:1–20.
100. Williams M, Summers BJ, Baxter RHG. Biophysical Analysis of Anopheles gambiae Leucine-Rich Repeat Proteins APL1A1, APLB and APL1C and Their Interaction with LRIM1. PLoS ONE. 2015;10:1–20.
101. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
102. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457–62.
103. Hallgren J, Tsirigos KD, Damgaard Pedersen M, Juan J, Armenteros A, Marcatili P, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022;:2022.04.08.487609.
104. Khalaf A, Lawniczak MKN, Blaxter ML, Jaron KS. Polyploidy is widespread in Microsporidia. Microbiol Spectr. 2024;12:e03669–23. https://doi.org/10.1128/spectrum.03669-23.
105. Hazard EI, Andreadis TG, Joslyn DJ, Ellis EA. Meiosis and Its Implications in the Life Cycles of Amblyospora and Parathelohania (Microspora). J Parasitol. 1979;65:117.
106. Lee SC, Heitman J, Ironside JE. Sex and the Microsporidia. In: Weiss, L.M., & Becnel, J.J. (Eds.), Microsporidia: Pathogens of Opportunity: First Edition. 2014:231–43. https://doi.org/10.1002/9781118395264.ch8.
107. Cali A, Takvorian PM. Developmental Morphology and Life Cycles of the Microsporidia. In: Weiss, L.M., & Becnel, J.J. (Eds.), Microsporidia: Pathogens of Opportunity: First Edition. 2014:71–133. https://doi.org/10.1002/9781118395264.ch2.
108. Vossbrinck CR, Debrunner-Vossbrinck BA. Molecular phylogeny of the Microsporidia: Ecological, ultrastructural and taxonomic considerations. Folia Parasitol (Praha). 2005;52:131–42.
109. Ndikumana S, Pelin A, Williot A, Sanders JL, Kent M, Corradi N. Genome Analysis of Pseudoloma neurophilia: A Microsporidian Parasite of Zebrafish (Danio rerio). J Eukaryot Microbiol. 2017;64:18–30.
110. Huang Q, Li W, Chen Y. Dicer regulates Nosema ceranae proliferation in honeybees. 2019;28:74–85.
111. Texier C, Brosson D, El Alaoui H, Méténier G, Vivarès CP. Post-genomics of microsporidia, with emphasis on a model of minimal eukaryotic proteome: A review. Folia Parasitol (Praha). 2005;52:15–22.
112. Reinke AW, Balla KM, Bennett EJ, Troemel ER. Identification of microsporidia host-exposed proteins reveals a repertoire of rapidly evolving proteins. Nat Commun. 2017;8 May 2016.
113. Herren JK, Mbaisi L, Mararo E, Makhulu EE, Mobegi VA, Butungi H, et al. A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. Nat Commun. 2020;11:1–10.
114. Bacchi CJ, Yarlett N, Weiss LM. Polyamine metabolism in the Microsporidia. Biochem Soc Trans. 2003;31:420–3.
115. Bacchi CJ, Lane S, Weiss LM, Yarlett N, Takvorian P, Wittner M. Polyamine synthesis and interconversion by the microsporidian Encephalitozoon cuniculi. J Eukaryot Microbiol. 2001;48:374–81.
116. Timofeev S. Energy metabolism and its evolution in Microsporidia and allied taxa. 2020;119:1433–41.
117. Fast NM, Keeling PJ. Alpha and beta subunits of pyruvate dehydrogenase E1 from the microsporidian Nosema locustae: Mitochondrion-derived carbon metabolism in microsporidia. Mol Biochem Parasitol. 2001;117:201–9.
118. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST : quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
119. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
120. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
121. Laetsch DR, Blaxter ML, Leggett RM. BlobTools : Interrogation of genome assemblies [version 1 ; peer review : 2 approved with reservations]. 2020;1287:1–16.
122. Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 2001;29:2607–18.
123. Wilkins D. gggenes: Draw Gene Arrow Maps in "ggplot2." 2023.
124. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14.
125. Sahbou AE, Iraqi D, Mentag R, Khayi S. BuscoPhylo: a webserver for Busco-based phylogenomic analysis for non-specialists. Sci Rep. 2022;12:1–5.
126. Olsen C. Geneious R7: A Bioinformatics Platform for Biologists. Plant and Animal Genome XXII Conference. 2014;April:20–1.
127. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
128. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res. 2020;48:D265–8.
129. Lees JG, Lee D, Studer RA, Dawson NL, Sillitoe I, Das S, et al. Gene3D: Multi-domain annotations for protein sequence and comparative genome analysis. Nucleic Acids Res. 2014;42:240–5.
130. Bolleman J, de Castro E, Baratin D, Gehant S, Cuche BA, Auchincloss AH, et al. HAMAP as SPARQL rules—A portable annotation pipeline for genomes and proteomes. Gigascience. 2020;9:giaa003.
131. Necci M, Piovesan D, Dosztányi Z, Tosatto SCE. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. Bioinformatics. 2017;33:1402–4.
132. Mi H, Ebert D, Muruganujan A, Mills C, Albou LP, Mushayamaha T, et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021;49:D394-403.
133. Wu CH, Nikolskaya A, Huang H, Yeh LS, Natale DA, Vinayaka CR, et al. PIRSF: Family classification system at the Protein Information Resource. Nucleic Acids Res. 2004;32 DATABASE ISS.:112–4.
134. Attwood TK. The PRINTS database: a resource for identification of protein families. Brief Bioinform. 2002;3:252–63.
135. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, et al. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 2002;3:265–74.
136. Akiva E, Brown S, Almonacid DE, Barber AE 2nd, Custer AF, Hicks MA, et al. The structure–function linkage database. Nucleic Acids Res. 2014;42:D521–30.
137. Letunic I, Khedkar S, Bork P. SMART: Recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49:D458–60.
138. Pandurangan AP, Stahlhacke J, Oates ME, Smithers B, Gough J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. Nucleic Acids Res. 2019;47:D490–4.
139. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM. (Eds) The Proteomics Protocols Handbook. 2005:571–608. https://doi.org/10.1385/1-59259-890-0:571.
140. Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, et al. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2:1–14.
141. Teufel F, Almagro Armenteros JJ, Johansen AR, G\'islason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40:1023–5.
142. Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. Proteins Struct Function Bioinform. 2009;77 SUPPL. 9:128–32.
143. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res. 2009;37(SUPPL):2.
144. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007;35(SUPPL. 2):407–10.
145. Luthy R, Bowei J, Einsenberg D. Verify3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol. 1997;277:396–404.
146. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43:W39-49.

147. Bailey TL, Williams N, Misleh C, Li WW. MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 2006;34 WEB. SERV. ISS.:369–73.
148. Tokarev YS, Timofeev SA, Malysh JM, Tsarev AA, Ignatieva AN, Tomilova OG, et al. Hexokinase as a versatile molecular genetic marker for Microsporidia. Parasitology. 2019;146:472–8.
149. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008;36 Web Server issue:5–9.
150. for Biotechnology Information (US) NC, Camacho C. BLAST (r) command line applications user manual. National Center for Biotechnology Information (US) Bethesda, MD; 2008.
151. Dieni CA, Storey KB. Regulation of glucose-6-phosphate dehydrogenase by reversible phosphorylation in liver of a freeze tolerant frog. J Comp Physiol B. 2010;180:1133–42.
152. Kruger NJ, Von Schaewen A. The oxidative pentose phosphate pathway: Structure and organisation. Current Opinion in Plant Biology. 2003;6:236–46.
153. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, et al. AmoebaDB and MicrosporidiaDB: Functional genomic resources for Amoebozoa and Microsporidia species. Nucleic Acids Res. 2011;39(SUPPL. 1):612–9.

Publisher's Note