

RESEARCH ARTICLE

Open Access



# Contrasting outcomes of genome reduction in mikrocytids and microsporidians

Vojtěch Žárský<sup>1†</sup>, Anna Karnkowska<sup>1,2†</sup>, Vittorio Boscaro<sup>1\*</sup> , Morelia Trznadel<sup>1</sup>, Thomas A. Whelan<sup>1</sup>, Markus Hiltunen-Thorén<sup>3,4</sup>, Ioana Onut-Brännström<sup>3,5,6</sup>, Cathryn L. Abbott<sup>7</sup>, Naomi M. Fast<sup>1</sup>, Fabien Burki<sup>3</sup> and Patrick J. Keeling<sup>1\*</sup>

## Abstract

**Background** Intracellular symbionts often undergo genome reduction, losing both coding and non-coding DNA in a process that ultimately produces small, gene-dense genomes with few genes. Among eukaryotes, an extreme example is found in microsporidians, which are anaerobic, obligate intracellular parasites related to fungi that have the smallest nuclear genomes known (except for the relic nucleomorphs of some secondary plastids). Mikrocytids are superficially similar to microsporidians: they are also small, reduced, obligate parasites; however, as they belong to a very different branch of the tree of eukaryotes, the rhizarians, such similarities must have evolved in parallel. Since little genomic data are available from mikrocytids, we assembled a draft genome of the type species, *Mikrocytos mackini*, and compared the genomic architecture and content of microsporidians and mikrocytids to identify common characteristics of reduction and possible convergent evolution.

**Results** At the coarsest level, the genome of *M. mackini* does not exhibit signs of extreme genome reduction; at 49.7 Mbp with 14,372 genes, the assembly is much larger and gene-rich than those of microsporidians. However, much of the genomic sequence and most (8075) of the protein-coding genes code for transposons, and may not contribute much of functional relevance to the parasite. Indeed, the energy and carbon metabolism of *M. mackini* share several similarities with those of microsporidians. Overall, the predicted proteome involved in cellular functions is quite reduced and gene sequences are extremely divergent. Microsporidians and mikrocytids also share highly reduced spliceosomes that have retained a strikingly similar subset of proteins despite having reduced independently. In contrast, the spliceosomal introns in mikrocytids are very different from those of microsporidians in that they are numerous, conserved in sequence, and constrained to an exceptionally narrow size range (all 16 or 17 nucleotides long) at the shortest extreme of known intron lengths.

**Conclusions** Nuclear genome reduction has taken place many times and has proceeded along different routes in different lineages. Mikrocytids show a mix of similarities and differences with other extreme cases, including uncoupling the actual size of a genome with its functional reduction.

**Keywords** Genomics, Evolution, Parasite, Reduction, Anaerobic, Intron, Splicing, Metabolism, *Mikrocytos*, Microsporidians

<sup>†</sup>Vojtěch Žárský and Anna Karnkowska shared first authorship.

\*Correspondence:

Vittorio Boscaro  
vittorio.boscaro@botany.ubc.ca

Patrick J. Keeling  
pkeeling@mail.ubc.ca

Full list of author information is available at the end of the article



## Background

One of the most consistent trends observed in the evolution of intracellular parasites and symbionts more broadly is that obligate intracellular organisms undergo genome reduction [1–3]. Genome size, gene number, and even non-coding sequence length decrease in endosymbionts compared to their free-living ancestors, while sequence substitution rates often increase. Sometimes these changes can be drastic [4–6]. Alternatively attributed to adaptive streamlining of non-essential functions or neutral loss due to weakened selection [2, 7], there are few exceptions to this outcome. While genome reduction is mostly studied in specialized pathogenic or ancient mutualistic bacteria [8, 9], it occurs in a wide variety of contexts, and not just in prokaryotes but also in eukaryotes. A notable case is found in Microsporidia [10, 11], unicellular protists related to fungi with several unusual features [12, 13]. Microsporidians are obligate intracellular parasites that can be pathogenic in immunocompromised humans and cause widespread diseases in other animals, including economically important species such as bees and silkworms [14–16]. While microsporidian spores possess a complex infection mechanism, the cells are highly reduced in nearly every other way. Their metabolism is so limited that microsporidians have lost most or in some cases all ATP production pathways and steal ATP directly from their hosts [17–19]. Their mitochondria evolved into anaerobic “mitosomes” lacking a genome and seemingly having the only function of synthesizing Fe-S clusters [20]. Microsporidian nuclear genomes are typically also highly reduced and include the smallest nuclear genome known in any cell: 2.3 Mbp and 1800 protein-coding genes in *Encephalitozoon intestinalis* [4]. As models for nuclear genome reduction and compaction, microsporidian genome content, gene density, and introns have all been studied in some detail [4, 21, 22].

An interesting potential lineage to compare and contrast with microsporidians are the mikrocytids, a more

recently discovered group of parasites of marine invertebrates currently comprising only a few described species [5, 23–25]. Mikrocytids belong to the understudied eukaryotic “supergroup” Rhizaria [5] and are therefore only distantly related to Microsporidia. However, the two lineages share a host-dependent, intracellular lifestyle and some convergent features including the reduction of mitochondria to mitosomes [5, 26]. The transcriptome of the mikrocytid *Mikrocytos mackini*, the causative agent of Denman Island disease in oysters [27], showed the fastest sequence substitution rate of any known eukaryote [5], suggesting once again a marked effect of endosymbiosis on molecular evolution.

While genome reduction is common to intracellular organisms, its extent, underlying mechanisms, and the order of events leading to it are not the same in different lineages, especially among eukaryotes. To examine some of the similarities and differences in the process, here we sequenced the genome of the mikrocytid *M. mackini* and compared its overall characteristics (as well as those of the recently reported genome of its closest known relative, *Paramikrocytos canceri* [26]) with the more thoroughly studied genomes of microsporidians.

## Results and discussion

### The *Mikrocytos* genome is large and gene-rich

*Mikrocytos mackini* is a tiny (<5 μm), strictly intracellular parasite that cannot be cultured outside its host. To obtain as clean an assembly as possible in such circumstances, *M. mackini* cells were isolated from the tissue of the host, the Pacific oyster *Crassostrea gigas*, and libraries constructed from the inevitably low DNA yield. The final 49.7 Mbp assembly appears to be largely complete, albeit very fragmented (16,018 contigs; N50 = 4547 bp; Table 1), which is likely at least in part due to a high number of repetitive sequences (see below). The relative completeness of the assembly is evidenced by the high percentages of RNA-Seq reads mapping against the genome draft (96%) and of detected orthologs from a dataset [28] of 263

**Table 1** Assembly statistics for genomes of mikrocytids (*Mikrocytos mackini* and *Paramikrocytos canceri*) and other available rhizarians

	<i>Mikrocytos mackini</i> (parasite)	<i>Paramikrocytos canceri</i> (parasite)	<i>Plasmodiophora brassicae</i> (parasite)	<i>Bigelowiella natans</i> (free-living)	<i>Reticulomyxa filosa</i> (free-living)
Assembly size (Mbp)	49.7	12.7	24.0	94.7	101.9
No. contigs	16,018	3,113	165	302	50,809
N50	4547	6806	472,887	819,951	3609
Max. contig length (bp)	76,988	40,223	1,189,627	3,030,241	48,547
GC content	33.60%	30.06%	59.40%	44.90%	35.00%
No. of protein-coding genes	14,372	8201	9,913	22,320	40,160
Reference	This paper	[26]	[31]	[32]	[33]

highly conserved eukaryotic genes (80%). Differences in metabolic gene sets were further inspected between the *M. mackini* transcriptome and genome, and only three genes were found in the former but not the latter: two were part of the assembly, but split across separate contigs, and one was entirely missing, although its predicted function was represented by other paralogs. Additionally, all rRNA, tRNA, and tRNA synthetase genes were present, as well as most ribosomal protein genes (75%). Low BUSCO scores (36%) have been recovered before for genomic data from protists belonging to undersampled groups [26, 29, 30] and are likely the consequence of sequence divergence and poor representation in reference databases. The *M. mackini* assembly is considerably larger than those obtained from other rhizarian parasites like *Plasmodiophora brassicae* (24 Mbp [31]) and *Paramikrocytos canceri* (13 Mbp [26]).

Although the sample size is limited by the scarcity of data on Rhizaria overall, the genome reduction trend is confirmed in this eukaryotic supergroup, with much smaller genomes observed for parasitic than free-living organisms (Table 1). Its extent is however not as dramatic as in microsporidians, where genome sizes vary widely (2–50 Mbp) but are usually well below 10 Mbp [22]. The difference is more prominent in the number of protein-coding genes, uniformly low in microsporidians (2000–5000) and higher in mikrocytids (14,372 predicted putative genes in *M. mackini*, >8000 in *P. canceri*). This cannot simply be attributed to a more recent origin of parasitism or slower progression of DNA loss, since *M. mackini* and *P. canceri* undoubtedly share an already parasitic common ancestor but have considerably different degrees of genome reduction, suggesting a more complex dynamic, possibly linked to the surprisingly high number of transposons in *M. mackini*.

#### The *M. mackini* genome encodes abundant and diverse transposable elements

According to models developed in bacteria, genome reduction usually goes through an early, chaotic stage characterized by the uncontrolled spread of mobile elements (due to relaxed purifying selection), which in turn facilitates chromosome rearrangements and pseudogenization [34], and a later more stable stage involving loss of non-essential sequences (including mobile elements) and compaction [2]. There are exceptions to this rule [35], and the progression has not been established in eukaryotes. We did, however, observe a large number of mobile elements in the genome of *M. mackini*: 8075, or 56%, of the predicted protein-coding genes show signatures of transposon origin, as do many other regions of the genome (Table 2). About half of the predicted transposable elements (TE) could be assigned to

**Table 2** Number and classification of transposable elements in mikrocytid genomes

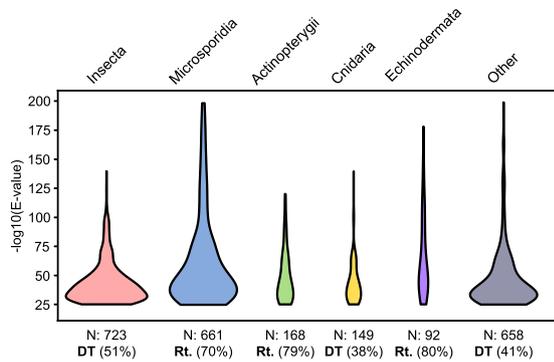
		<i>M. mackini</i>	<i>P. canceri</i>
<b>Total</b>		<b>32,821</b>	<b>13,633</b>
<b>Retrotransposons</b>	<b>LTR</b>	4228	298
	<b>LINE</b>	1730	1148
<b>DNA transposons</b>	<b>TIR</b>	8228	1877
	<b>Helitrons</b>	792	0
<b>Unclassified</b>		17,843	10,310

*LTR* long terminal repeats, *LINE* long interspersed nuclear element, *TIR* terminal inverted repeat

known families, especially long terminal repeats, long interspersed nuclear elements, terminal inverted repeat-containing DNA transposons, and helitrons. The *P. canceri* genome assembly also encodes TEs, albeit to a much lower extent (Table 2). If this is an accurate reflection of both genomes, as seems to be the case, the rates of TE spread and/or loss in the two lineages must have been highly dynamic.

It is also possible that the *M. mackini* genome is the result of more recent “invasions” of transposons in an already reduced genome under weak purifying selection. This seems to be the case for many microsporidians, where species with larger genomes have more TEs than species with the most reduced genomes, which have few or none [36, 37]. Since the last common ancestor of extant microsporidians was already an intracellular parasite with a highly reduced gene content, this diversity is unlikely to reflect the ancestral state, but rather later TE invasions that produced secondary genome bloat. Further evidence in microsporidians comes from the sources of TEs, which were seemingly acquired from a variety of animals, probably reflecting host shifts over their evolutionary history [38, 39]. Similarly, TE sequences in *Mikrocytos* show relatively high similarities with homologs in, among others, ray-finned fishes, echinoderms, insects, cnidarians, and even microsporidians (the latter likely indicating an exchange between co-occurring parasites) (Fig. 1). More genomes from mikrocytids are required to confirm the observed pattern and exclude any influence from undetected contaminant sequences, which are always a possibility when working with intracellular organisms. However, the current data lend more support to a differential transposon acquisition rather than the unchecked multiplication of ancestral TEs scenario.

Interestingly, another strong correlation found in microsporidians is the presence of Argonaute and Dicer components of the RNAi machinery in all TE-rich genomes [39, 40], which is not the case in *M. mackini*, where orthologs of these genes could not be identified.



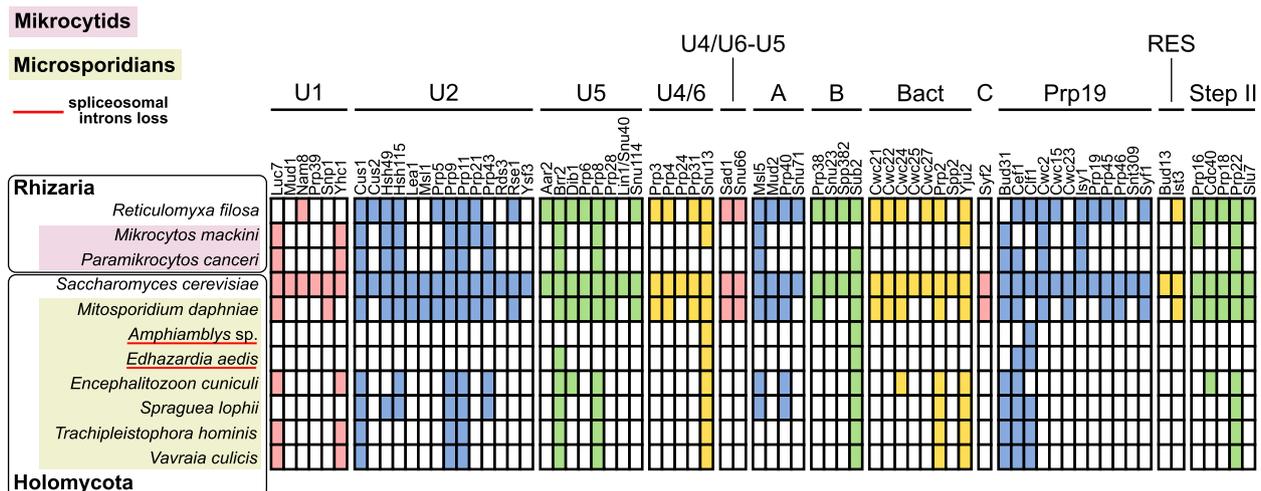
**Fig. 1** Sources of transposable elements in the genome of *M. mackini*. Each graph shows the distribution of BLAST hits against selected taxa of animals and protists. The number of reliable ( $e\text{-value} < 1e^{-20}$ ) hits shown for each target group is reported (N). The percentage of hits belonging to either DNA transposons (DT) or retrotransposons (Rt.), whichever is higher, is also shown

**Microcytids have many, extremely short introns of highly uniform length**

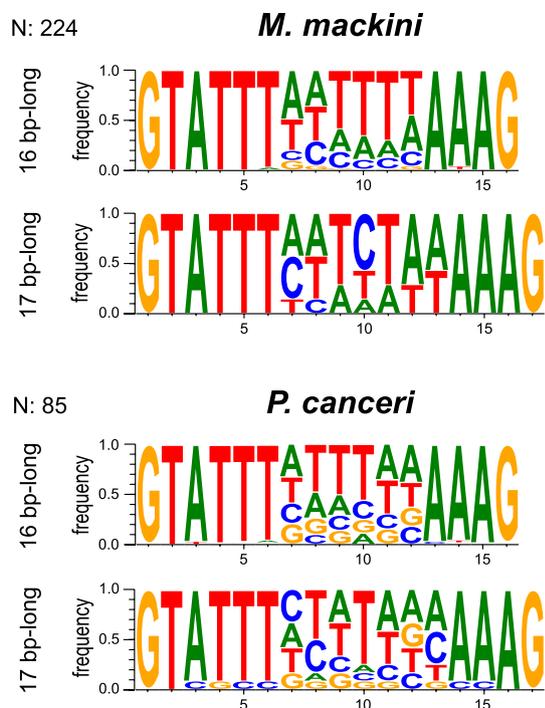
In obligately symbiotic bacteria, the shortening of non-coding sequences during genome reduction generally means short intergenic regions. In intracellular eukaryotes, the trend can also extend to introns, either due to loss, length reduction, or both. Microsporidians generally have few introns that are relatively short and retain a higher-than-average sequence similarity to one another [18, 21, 41, 42], as well as a reduced spliceosomal machinery [41, 43, 44]. A few microsporidians have independently lost introns altogether [18, 45, 46]. The spliceosome in mikrocytids seems to be almost as small (only 17–19 proteins plus the U2, U4, and U6 snRNAs

were identified), and there is a striking degree of overlap in the proteins that have been retained in the two groups, despite their independent spliceosome reduction (Fig. 2). In contrast to the intron-poor genomes of microsporidians, however, our annotation predicted 224 introns in 179 genes in the genome of *M. mackini*. These introns are incredibly small and very uniform in length: nearly all were 16 bp long (217/224), and the rest were 17 bp long (7/224). Comparing the genome to transcriptomic data showed that 16 bp introns spliced twice as frequently as 17 bp introns (63% vs. 37%, respectively). Moreover, all introns shared highly conserved sequences (Fig. 3). About half of the intron-containing *M. mackini* genes were functionally annotated, revealing that most are involved in essential functions related to gene expression (including DNA damage repair, RNA transcription, splicing, etc.) and cell-cycle regulation (Additional file 1: Fig. S1). No intron was found in metabolic enzyme or transporter genes. Examining the *P. canceri* assembly revealed that it too contains introns with these same characteristics (Fig. 3).

It is generally unclear why introns vary so much in length and number, from more than 90,000 in the ciliate *Paramecium* [47] to few or none in microsporidians, trypanosomes, and other protists [46, 48, 49]. The yeast *Saccharomyces cerevisiae* has relatively few (282) and long (~400 bp) introns, which play an important role in gene expression regulation [50, 51]. There is no strong evidence for the same function in microsporidians, despite some similarities in intron distribution and localization (as in yeasts, they are often found at the 5' end of ribosomal protein genes). Another reduced



**Fig. 2** Spliceosome convergent reduction in mikrocytids and microsporidians. The table shows the presence/absence of spliceosomal protein components in the genomes of the two mikrocytids, selected microsporidians, and two non-reduced relatives for reference: the free-living *Reticulomyxa filosa* (Rhizaria) for mikrocytids and the yeast *Saccharomyces cerevisiae* (Holomycota) for microsporidians. Microsporidians that have completely lost spliceosomal introns are underlined in red



**Fig. 3** Conserved sequences of the extremely short spliceosomal introns detected in both mikrocytids, *M. mackini* (top) and *P. canceri* (bottom). N stands for the total number of spliceosomal introns found in each genome

genome rich in introns (more than 800 in approximately 300 genes) is found in the chlorarachniophyte nucleomorph, a remnant nucleus of secondary plastids derived from an ancient symbiosis with a green alga [52]. The nucleomorph genome is another example of extreme genome reduction in an intracellular symbiosis and, like those of mikrocytids, its introns are not only short, but also fall into a narrow size range: 18 to 21 bp in this case. The smaller and more narrowly constrained introns of mikrocytids are matched only by the 15–16 bp introns of heterotrich ciliates [53, 54], which, seemingly against the trend, are free-living organisms with very large cells, nuclei, and genomes.

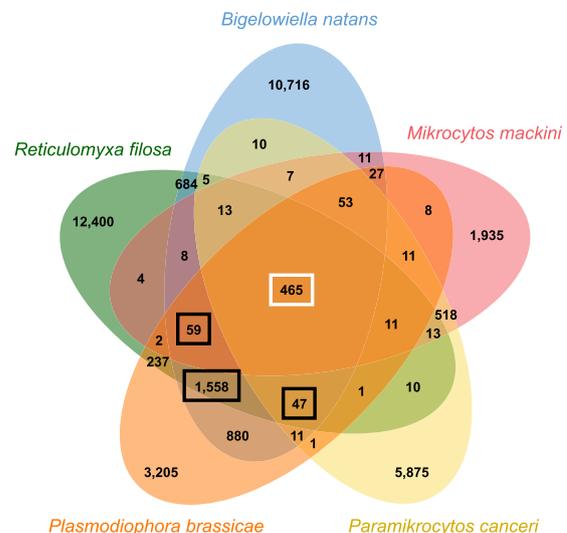
Intron reduction is likely occurring in different systems for different reasons, so seeking a single unifying explanation may be fruitless. In yeasts, many introns are hypothesized to be maintained for functional reasons [55], but alternative neutral explanations are also possible. For instance, like other non-coding sequences, introns in endosymbionts might simply gradually shrink in size due to genome erosion, where reduced DNA repair mechanisms lead to a bias for deletions over insertions. This could presumably continue until a functional threshold is hit, below which the introns might be too short to be efficiently spliced and further gradual reductions would be

strongly deleterious [54]. This threshold could be slightly different in systems evolving independently, for instance because introns in organisms with lower intron densities and reduced spliceosomes also tend to evolve greater dependence on sequence conservation for spliceosomal recognition and base-pairing with the snRNAs [56, 57]—the longer the recognition sequence, the longer the minimal intron size. A balance between this threshold and the deletion bias would lead introns to fall into a narrower and narrower size range, bounded on one side by their functional minimal length and eroded on the other by the strength of the deletion-bias.

**Divergent genes and reduced metabolism of *M. mackini***

Relatively few protein-coding genes unrelated to transposable elements (2072, or 33%, out of 6297) in the genome draft of *M. mackini* could be functionally annotated. While this is in part due to the paucity of data on close relatives of mikrocytids, an even larger effect is probably played by the sequence divergence characterizing this protist [5]. Indicative of this is the fact that rhizarians share only 465 gene orthogroups if mikrocytids are included, but 2129 if they are not (Fig. 4).

As in other parasites, many metabolic pathways that are considered essential in free-living eukaryotes are absent from *M. mackini*. Significantly, both *M. mackini* and *P. canceri* share a rare trait with microsporidians: the absence of the ATP synthase complex, as well as associated pathways like the carboxylic acid cycle and beta



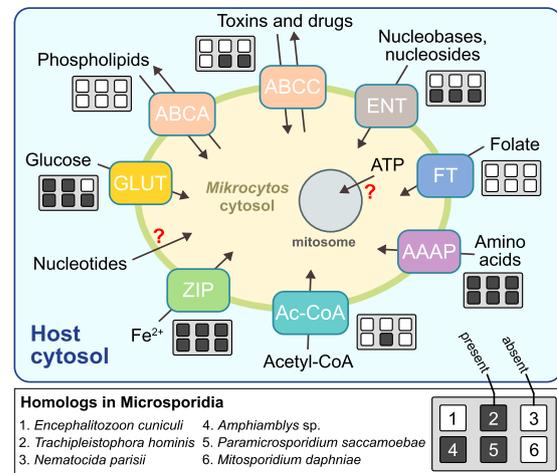
**Fig. 4** Shared orthogroups in rhizarian genomes. The Venn diagram shows the numbers of orthologous groups of genes inferred by Orthofinder in the five available rhizarian genomes. Highlighted with a white rectangle are the orthogroups shared by all rhizarians, including mikrocytids (465). Highlighted in black rectangles are additional orthogroups shared by all non-mikrocytid rhizarians (for a total of 2129)

oxidation. Genes for a full glycolysis pathway are present in the *M. mackini* genome, suggesting that some ATP can be produced by substrate-level phosphorylation. Another parallel with microsporidians [58] is the preservation of trehalose metabolism genes in an otherwise depleted carbon metabolism (Additional file 2: Fig. S2). Trehalose plays a role in carbohydrate storage in many invertebrates [59], and this together with its retention in *M. mackini* indicate that this compound might be important to the interactions between microcytids and their hosts. We additionally detected a putative trehalase gene with a signal peptide, suggesting it is secreted from the parasite cell, possibly to modulate and redirect the flow of carbohydrates in the host's cytoplasm. A similar use of trehalase has been predicted in microsporidians [60].

Overall, gene content supports a metabolic convergence between microsporidians and mikrocytids to energy parasitism, or the direct acquisition of some or all of the parasite's ATP from the host. This prediction is consistent with the close association to the host cell's mitochondria that is observed both in mikrocytids [25, 61] and microsporidians [62, 63]. However, it should be noted that among the 61 transporter genes, representing 17 families, annotated in *M. mackini* (a more reduced repertoire than that of microsporidians [64]), we did not find a clear candidate ATP transporter (Fig. 5). The same was true for *P. cankeri* [26]. The bacterial-derived nucleotide transporter (NTT) microsporidians use to import ATP [17] was not present in *M. mackini*, although we did identify the more common equilibrative nucleoside transporter (ENT). Also notably absent from mikrocytids are the mitochondria carrier family (MCF) genes, responsible for the transport of metabolites in mitochondria and mitosomes, which have also been replaced by bacterial transporters in some microsporidians [17, 64]. Considering how common horizontal gene transfers are, and inherent difficulties in transporter annotation, we cannot state that a particular transport function is missing in mikrocytids, but it seems likely that when it comes to transporters, these parasites often rely on different protein families than microsporidians to perform similar, key functions (Fig. 5).

**Conclusions**

Superficially, microsporidians and mikrocytids have a lot in common: they are intracellular parasites of other eukaryotes with tiny cells, mitosomes, and peculiar genomic traits. In fact, we have shown here that these two lineages have also converged to a similar form of very reduced metabolism with shared, rare features (Fig. 6). However, microsporidians and mikrocytids provide very different examples of how the process of



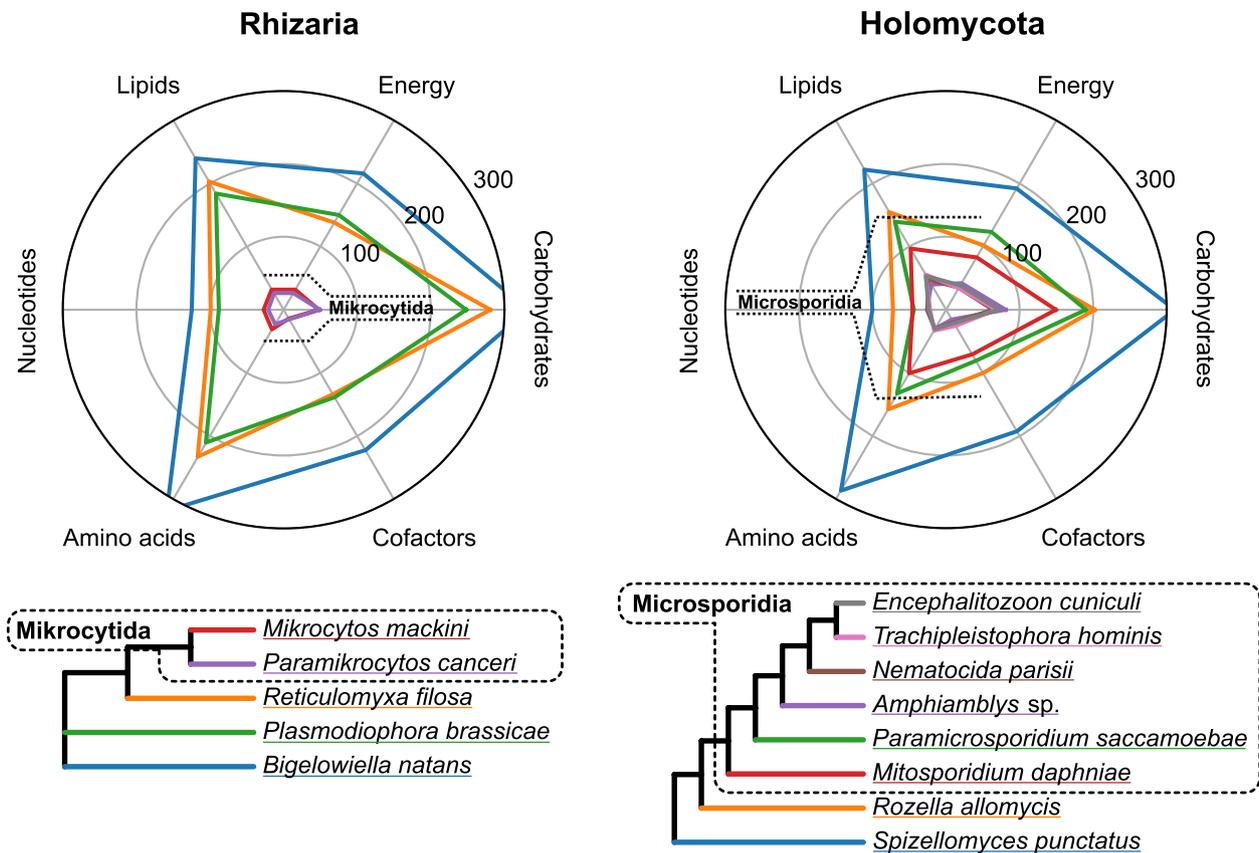
**Fig. 5** Predicted types of transporters identified in the *M. mackini* genome and differences with the microsporidian sets. The red question marks stand for metabolite exchanges that are supposed to happen, but for which no good candidate gene was detected. Next to each transporter predicted for *M. mackini*, a small plot shows which of the representative microsporidians have corresponding homologous genes. The set differs in many details between mikrocytids and microsporidians, as well as within microsporidians (as a further note, while microsporidians lack ABCA transporters, they possess the functionally related ABCG, which is missing in *Mikrocytos*). AAAP, amino acid/auxin permease; ABC, ATP-binding cassette transporter (types A and C); Ac-CoA, acetyl-CoA transporter; ENT, equilibrative nucleoside transporter; FT, folate transporter; GLUT, glucose transporter; ZIP, zinc/iron permease

genome reduction can develop. *Mikrocytos mackini* is an unusual case study for extensive, multiple transposon invasions in the context of an otherwise reduced genome, as well as extreme intron length reduction without outright loss.

**Methods**

**Cell isolation, library preparation, and sequencing**

*Mikrocytos mackini* were collected from parasitic lesions on the adductor muscle tissues of wild *Crassostrea gigas* harvested from Deep Bay (Vancouver Island, British Columbia, Canada), then used to infect oysters in the lab in order to generate sufficient material for nucleic acid extractions. Parasites were concentrated and isolated from the lab-infected hosts as described in [5]. DNA was extracted with the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. About 4 µg of DNA was submitted to the G enome Qu ebec sequencing center for library preparation and sequencing. TruSeq paired-end libraries were sequenced on the Illumina MiSeq (2×250 bp and 2×300 bp) and HiSeq (2×100 bp) platforms.



**Fig. 6** Convergent minimal metabolism of mikrocytids and the most reduced microsporidians. Plots are shown for all sequenced rhizarians including mikrocytids (left), and selected holomycotes (fungi and relatives, such as microsporidians and rozellids). Cladograms depict the phylogenetic relationships of the analyzed taxa. On each radial axis, representing a major KEGG metabolic category, the number of unique enzymes is plotted, showing a convergence of the most genome-reduced representatives of each group to similarly depleted enzyme sets

**Genomic and transcriptomic assemblies**

Adaptor sequences were removed and low-quality sequences trimmed from genomic reads using fastq-mcf [65]. Host contaminant reads were identified through mapping against a *Crassostrea gigas* reference genome using Megablast as implemented in the BLAST+ package [66], then removed (thresholds: >90% identity and >40% hit coverage). This first filter culled about 30% of the data. A preliminary assembly was built using Ray [67] and the contigs were aligned using BLAST against the NCBI nt database. Four potential *C. gigas* contigs were flagged and reads mapping to those contigs were removed. Remaining redundant reads were discarded using the normalize-by-median.py script of the khmer package [68].

Three assemblies were built using Ray (v.2.3.1), SPAdes (v.3.6.1) [69], and MIRA (an iterative assembler; three passes were used) [70]. The assemblies were first compared by mapping transcripts against each of them with gmap (v.2020-04-08) [71], which produced values of 91.9%, 92.7%, and 95.8% for the outputs of Ray, SPAdes,

and MIRA, respectively. Then, ALE [72] was run to estimate likelihood values, with the MIRA assembly obtaining the highest score. The final genome draft was then created by selecting large contigs from the MIRA assembly (>550 bp) and adding shorter contigs that did not have a match against the large contigs. A final decontamination step was performed by searching against the NCBI nt database using Megablast [73] and removing contigs matching *C. gigas*.

Transcriptomic reads from a previously reported study [5] were also re-assembled to examine genome completeness. Raw reads were trimmed using Trimmomatic [74] and assembled de novo using Trinity [75]. Common contaminants were detected using blobology [76] and the reads were filtered through mapping against database of identified contaminants with bwa [77]. De novo and genome-guided assemblies using only decontaminated reads were built again on Trinity, and a comprehensive set of transcripts was generated using the build\_comprehensive\_transcriptome.dbi script from the PASA pipeline (v.2) [78].

## Genome annotation

Preliminary gene predictions were performed using the PASA pipeline (v.2) to align transcripts to the assembly, Genemark [79], and Augustus (v.3.0.3) [80]. The final set of predicted genes was generated using EVM (v.r2012-06-25) [81] with inputs from all three other programs. BUSCO (v.5) [82] was run using the *alveolata\_odb10* dataset to obtain a completeness estimate.

Predicted protein-coding genes were annotated using eggNOG-mapper (v.2) [83] and Interproscan (v.5.50) [84]. The KEGG database of orthologs [85] was searched using HMMER [86] and served as the basis for the classification of enzymes and metabolic pathways. Orthologous protein-coding gene groups (orthogroups) were predicted for rhizarian genomes using OrthoFinder (v.2.5.2) [87] with default settings. The Venn diagram of shared orthogroups was created using OrthoVenn2 [88]. rRNA, tRNA, and snRNA genes were predicted using Infernal cmscan (v.1.1.3) [89] against the Rfam database (v.14) [90]. Metabolic graphs were built by counting the number of unique enzymes annotated in major KEGG Pathway Families (energy metabolism, carbohydrate metabolism, metabolism of cofactors and vitamins, amino acid metabolism, nucleotide metabolism, lipid metabolism), then plotting the numbers on radial axes using the polar plot projection as implemented in Matplotlib (v.3.7) [91].

Transposable elements in the *M. mackini* genome were detected and classified using RepeatModeler (v.2.0.3) [92] with the `-LTRStruct` option. TEs were then compared against the NCBI nt database using diamond blastx (v.2.0.7) [93] and best hits with  $e$ -values  $< 1e^{-20}$  (amino acid similarity values ranged from 74.8% to 21.4%; average: 35.2%) were collected and sorted by taxonomic group.

Putative introns were first pinpointed by mapping transcripts from *M. mackini* onto the genome draft using gmap with the `-min-intron-length 10` option. RNA-Seq reads were also mapped against the genome using the splice-aware aligner TopHat (v.2.1.1) [94] with the same length restriction. Mapped reads were then used to assess the exon coverage, count intron-spanning reads, and estimate splicing efficiency of putative intron–exon boundaries. The conservation of intron sequences was visualized using WebLogo (v.3) [95]. The gene ontology enrichment analysis of genes with spliceosomal introns was performed using Ontologizer (v.2.1) [96] and visualized with GO-Figure! (v.1.0) [97]. Spliceosomal proteins were detected using reciprocal BLAST against the human and yeast proteomes and candidates were checked using the HHpred server [98]. Sm and Lsm proteins were not included in the analysis as they are very short and unreliably differentiated based on sequence similarity alone.

## Abbreviations

bp	Base pairs
ENT	Equilibrative nucleoside transporter
MCF	Mitochondria carrier family
NTT	Nucleotide transporter
TE	Transposable elements

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-023-01635-w>.

**Additional file 1: Figure S1.** Gene Ontology enrichment analysis of genes with spliceosomal introns in the genome of *M. mackini* plotted in the GO semantic space using GO-Figure!. The colour scale represents the significance of the enrichment and the size of the circles stands for the number of genes with spliceosomal introns with that particular annotation. Where available, functional categories are shown in the legend.

**Additional file 2: Figure S2.** Predicted carbohydrate metabolism of *M. mackini*. The presence of a putative lactate dehydrogenase-enzyme has been deduced from the comparison with LDH / malate dehydrogenase homologs.

## Acknowledgements

We thank Gary R. Meyer (Fisheries and Oceans Canada) for helping with the preparation of the *M. mackini* sample used here.

## Authors' contributions

PJK, FB, and NMF conceived and planned the project. CLA isolated the organism from the host tissue and performed DNA extractions. VZ, AK, TW, MH, and IOB performed informatic analyses. VZ, VB, MT, and PJK wrote the first draft of the paper. All authors read and approved the final manuscript.

## Funding

This work was supported by a grant to PJK from the Gordon and Betty Moore Foundation (<https://doi.org/10.37807/GBMF9201>) and a NSERC Discovery Grant (262988) to NMF. FB wishes to thank the support for this work from the Swedish Research Council VR (2017-04563), Formas (2017-01197), and SciLifeLab.

## Availability of data and materials

The datasets generated and analyzed during the current study are available in the GenBank database at the following link: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA940158> [99].

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Botany, University of British Columbia, V6T 1Z4, Vancouver, 3529-6270 University Boulevard, BC, Canada. <sup>2</sup>Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, 02-089 Warsaw, Poland. <sup>3</sup>Department of Organismal Biology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden. <sup>4</sup>Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91, Stockholm, Sweden. <sup>5</sup>Department of Ecology and Genetics, Uppsala University, 752 36 Uppsala, Sweden. <sup>6</sup>Natural History Museum, University of Oslo, 0562 Oslo, Norway. <sup>7</sup>Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC V9T 6N7, Canada.

Received: 1 March 2023 Accepted: 26 May 2023  
Published online: 06 June 2023

## References

- Andersson SGE, Kurland CG. Reductive evolution of resident genomes. *Trends Microbiol.* 1998;6:263–8.
- McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 2011;10:13–26.
- Husnik F, Keeling PJ. The fate of obligate endosymbionts: reduction, integration, or extinction. *Curr Opin Genet Dev.* 2019;58–59:1–8.
- Corradi N, Pombert J-F, Farinelli L, Didier ES, Keeling PJ. The complete sequence of the smallest known nuclear genome from the microsporidian *Encephalitozoon intestinalis*. *Nat Comm.* 2010;1:77.
- Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, Keeling PJ. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. *Curr Biol.* 2013;23:1541–7.
- Moran NA, Bennett GM. The tiniest tiny genomes. *Annu Rev Microbiol.* 2014;68:195–215.
- Wernegreen JJ. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann NY Acad Sci.* 2015;1360:16–35.
- McCutcheon JP, Boyd BM, Dale C. The life of an insect endosymbiont from the cradle to the grave. *Curr Biol.* 2019;29:R485–95.
- Perreau J, Moran NA. Genetic innovations in animal-microbe symbioses. *Nat Rev Genet.* 2022;23:23–39.
- Pombert J-F, Haag KL, Beidas S, Ebert D, Keeling PJ. The *Ordospora col-ligata* genome: evolution of extreme reduction in microsporidian and host-to-parasite horizontal gene transfer. *mBi.* 2015;6:e02400.
- Melnikov SV, Manakongtreecheep K, Rivera KD, Makarenko A, Pappin DJ, Söll D. Muller's ratchet and ribosome degeneration in the obligate intracellular parasites Microsporidia. *Int J Mol Sci.* 2018;19:4125.
- Keeling P. Five questions about Microsporidia. *PLoS Pathog.* 2009;5:e1000489.
- Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol.* 2022;38:642–59.
- Klee J, Besana AM, Genersch E, Gisder S, Nanetti A, Tam DQ, et al. Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee *Apis mellifera*. *J Invertebr Pathol.* 2007;96:1–10.
- Stentiford GD, Becnel JJ, Weiss LM, Keeling PJ, Didier ES, Williams BAP, et al. Microsporidia – emergent pathogens in the global food chain. *Trends Parasitol.* 2016;32:336–48.
- Bhat IA, Buhroo ZI, Bhat MA. Microsporidiosis in silkworms with particular reference to mulberry silkworm (*Bombyx mori* L.). *Int J Entomol Res.* 2017;2:1–9.
- Tsaousis AD, Kunji ERS, Goldberg AV, Lucocq JM, Hirt RP, Embley TM. A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature.* 2008;453:553–6.
- 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.
- Dean P, Hirt RP, Embley M. Microsporidia: why make nucleotides if you can steal them? *PLoS Pathog.* 2016;12:e1005870.
- Freibert S-A, Goldberg AV, Hacker C, Molik S, Dean P, Williams TA, et al. Evolutionary conservation and *in vitro* reconstitution of microsporidian iron-sulfur cluster biosynthesis. *Nat Comm.* 2017;8:13932.
- Whelan TA, Lee NT, Lee RCH, Fast NM. Microsporidian introns retained against a background of genome reduction: characterization of an unusual set of introns. *Genome Biol Evol.* 2019;11:263–9.
- Wadi L, Reinke AW. Evolution of microsporidia: an extremely successful group of eukaryotic intracellular parasites. *PLoS Pathog.* 2020;16:e1008276.
- Hervio D, Bower SM, Meyer GR. Detection, isolation, and experimental transmission of *Mikrocytos mackini*, a microcell parasite of Pacific oyster *Crassostrea gigas* (Thunberg). *J Invertebr Pathol.* 1996;67:72–9.
- Abbott CL, Meyer GR. Review of *Mikrocytos* microcell parasites at the dawn of a new age of scientific discovery. *Dis Aquat Organ.* 2014;110:25–32.
- Hartikainen H, Stentiford GD, Bateman KS, Berney C, Feist SW, Longshaw M, et al. Mikrocytids are a broadly distributed and divergent radiation of parasites in aquatic invertebrates. *Curr Biol.* 2014;24:807–12.
- Onu-Brännström I, Stairs CW, Campos KIA, Ettema TJG, Keeling PJ, Bass D, Burki F. A mitosome with distinct metabolism in the uncultured protist parasite *Paramikrocytos canceri* (Rhizaria, Ascetosporea). *Genome Biol Evol.* 2023;15:evad022.
- Polinski MP, Meyer GR, Lowe GJ, Abbott CL. Seawater detection and biological assessments regarding transmission of the oyster parasite *Mikrocytos mackini* using qPCR. *Dis Aquat Organ.* 2017;126:143–53.
- Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc R Soc B.* 2016;283:20152802.
- Karnkowska A, Vacek V, Zubáčová Z, Treitl SC, Petrželková Z, Eme L, et al. A eukaryote without a mitochondrial organelle. *Curr Biol.* 2016;26:1274–84.
- Salas-Leiva DE, Tromer EC, Curtis BA, Jerlström-Hultqvist J, Kolisko M, Yi Z, Salas-Leiva JS, et al. Genomic analysis finds no evidence of canonical eukaryotic DNA processing complexes in a free-living protist. *Nat Commun.* 2021;12:6003.
- Rolfe SA, Strelkov SE, Links MG, Clarke WE, Robinson SJ, Djavaheri M, et al. The compact genome of the plant pathogen *Plasmodiophora brassicae* is adapted to intracellular interactions with host *Brassica* spp. *BMC Genomics.* 2016;17:272.
- Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, et al. Algal genomes reveal evolutionary mosaicism and the fate of the nucleomorph. *Nature.* 2012;492:59–65.
- Glöckner G, Hülsmann N, Schleicher M, Noegel AA, Eichinger L, Gallinger C, et al. The genome of the foraminiferan *Reticulomyxa filosa*. *Curr Biol.* 2014;24:11–8.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. *Genome Biol.* 2018;19:199.
- Boscaro V, Kolisko M, Felletti M, Vannini C, Lynn DH, Keeling PJ. Parallel genome reduction in symbionts descended from closely related free-living bacteria. *Nat Ecol Evol.* 2017;1:1160–7.
- de Albuquerque NRM, Ebert D, Haag KL. Transposable element abundance correlates with mode of transmission in microsporidian parasites. *Mob DNA.* 2020;11:19.
- Haag KL, Pombert J-F, Sun Y, de Albuquerque NRM, Batliner B, Fields P, et al. Microsporidia with vertical transmission were likely shaped by nonadaptive processes. *Genome Biol Evol.* 2020;12:3599–614.
- 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.
- Corradi N. Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Ann Rev Microbiol.* 2015;69:167–83.
- Huang Q. Evolution of Dicer and Argonaute orthologs in microsporidian parasites. *Infect Genet Evol.* 2018;65:329–32.
- Katinka MD, Duprat S, Cornillot E, Méténier G, Thomarat F, Prensier G, et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature.* 2001;414:450–3.
- Lee RCH, Gill EE, Roy SW, Fast NM. Constrained intron structures in a microsporidian. *Mol Biol Evol.* 2010;27:1979–82.
- Grisdale CJ, Bowers LC, Didier ES, Fast NM. Transcriptome analysis of the parasite *Encephalitozoon cuniculi*: an in-depth examination of pre-mRNA splicing in a reduced eukaryote. *BMC Genomics.* 2013;14:207.
- Black CS, Whelan TA, Garside EL, MacMillan AM, Fast NM, Rader SD. Spliceosomal assembly and regulation: insights from analysis of highly reduced spliceosomes. *RNA.* 2023;29:531–50.
- Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, Didier ES, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res.* 2012;22:2478–88.
- 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;6:7121.
- Chen C-L, Zhou H, Liao J-Y, Qu L-H, Amar L. Genome-wide evolutionary analysis of the noncoding RNA genes and noncoding DNA of *Paramecium tetraurelia*. *RNA.* 2009;15:503–14.

48. Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, Bowman S, Archibald JM. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci USA*. 2007;104:19908–13.
49. Maslov DA, Opperdoes FR, Kostygov AY, Hashimi H, Lukeš J, Yurchenko V. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology*. 2018;146:1–27.
50. Roy B, Granas D, Bragg F, Cher JAY, White MA, Stormo GD. Autoregulation of yeast ribosomal proteins discovered by efficient search for feedback regulation. *Commun Biol*. 2020;3:761.
51. Lim CS, Weinstein BN, Roy SW, Brown CM. Analysis of fungal genomes reveals commonalities of intron gain or loss and functions in intron-poor species. *Mol Biol Evol*. 2021;38:4166–86.
52. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, McFadden GI. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci USA*. 2006;103:9566–71.
53. Slabodnick MM, Graham Ruby J, Reiff SB, Swart EC, Gosai S, Prabakaran S, et al. The macronuclear genome of *Stentor coeruleus* reveals tiny introns in a giant cell. *Curr Biol*. 2017;27:569–75.
54. Nuadthaisong J, Phetruen T, Techawisutthinan C, Chanarat S. Insights into the mechanism of pre-mRNA splicing of tiny introns from the genome of a giant ciliate *Stentor coeruleus*. *Int J Mol Sci*. 2022;23:10973.
55. Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Elela SA. Introns are mediators of cell response to starvation. *Nature*. 2019;565:612–7.
56. Irimia M, Penny D, Roy SW. Coevolution of genomic intron number and splice site. *Trends Genet*. 2007;23:321–5.
57. Hudson AJ, McWatters DC, Bowser BA, Moore AN, Larue GE, Roy SW, Russell AG. Patterns of conservation of spliceosomal intron structures and spliceosome divergence in representatives of the diplomonad and parabasalid lineages. *BMC Evol Biol*. 2019;19:162.
58. Undeen AH, Solter LF. The sugar content and density of living and dead microsporidian (Protozoa: Microsporida) spores. *J Invertebr Pathol*. 1996;67:80–91.
59. Elbein AD, Pan YT, Pstuszek I, Carroll D. New insights on trehalose: a multifunctional molecule. *Glycobiology*. 2003;13:17R–27R.
60. Senderskiy IV, Timofeev SA, Seliverstova EV, Pavlova OA, Dolgikh VV. Secretion of *Antonospora (Paranosema) locustae* proteins into infected cells suggests an active role of Microsporida in the control of host programs and metabolic processes. *PLoS ONE*. 2014;9: e93585.
61. Hine PM, Bower SM, Meyer GR, Cochenne-Laureau N, Berthe FCJ. Ultrastructure of *Mikrocytos mackini*, the cause of Denman Island disease in oysters *Crassostrea* spp. and *Ostrea* spp. in British Columbia, Canada. *Dis Aquat Organ*. 2001;45:215–27.
62. Scanlon M, Leitch GJ, Visvesvara GS, Shaw AP. Relationship between the host cell mitochondria and the parasitophorous vacuole in cells infected with *Encephalitozoon microsporidia*. *J Eukaryot Microbiol*. 2004;51:81–7.
63. Hacker C, Howell M, Bhella D, Lucocq J. Strategies for maximizing ATP supply in the microsporidian *Encephalitozoon cuniculi*: direct binding of mitochondria to the parasitophorous vacuole and clustering of the mitochondrial porin VDAC. *Cell Microbiol*. 2014;16:565–79.
64. Heinz E, Williams TA, Nakjang S, Noël CJ, Swan DC, Goldberg AV, 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.
65. Aronesty E. Comparison of sequencing utility programs. *Open Bioinform J*. 2013;7:1–8.
66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
67. Boisvert S, Lavolette F, Corbeil J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol*. 2010;17:1519–33.
68. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, et al. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res*. 2015;4:900.
69. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
70. Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. Using the MiraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res*. 2004;14:1147–59.
71. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005;21:1859–75.
72. Clark SC, Egan R, Frazier PI, Wang Z. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics*. 2013;29:435–43.
73. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res*. 2015;43:7762–8.
74. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
75. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8:1494–512.
76. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. Blobology: exploring raw genome data for contaminants, symbionts, and parasites using taxon-annotated GC-coverage plots. *Front Genet*. 2013;4:237.
77. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
78. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–66.
79. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom Bioinform*. 2020;2:lqaa026.
80. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 2006;34:W435–9.
81. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9:R7.
82. 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.
83. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. EggNOG-Mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38:5825–9.
84. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
85. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45:D353–61.
86. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–37.
87. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238.
88. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2019;47:W52–8.
89. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29:2933–5.
90. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, et al. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res*. 2021;49:D192–200.
91. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Engin*. 2007;9:90–5.
92. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;117:9451–7.
93. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18(4):366–8.
94. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25:1105–11.
95. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188–90.

96. Bauer S, Grossmann S, Vingron M, Robinson PN. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*. 2008;24:1650–1.
97. Reijnders MJMF, Waterhouse RM. Summary visualizations of gene ontology terms with GO-Figure! *Front Bioinform*. 2021;1: 638255.
98. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*. 2005;33:W244–8.
99. Zarsky V, Karnkowska A, Abbott CL, Burki F, Keeling PJ. *Mikrocytos mackini* genome sequencing and assembly. GenBank. (2023). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA940158>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

