

SHORT REPORT

Open Access



# Transposable element abundance correlates with mode of transmission in microsporidian parasites

Nathalia Rammé Medeiros de Albuquerque<sup>1</sup>, Dieter Ebert<sup>2</sup> and Karen Luisa Haag<sup>1\*</sup> 

## Abstract

The extreme genome reduction and physiological simplicity of some microsporidia has been attributed to their intracellular, obligate parasitic lifestyle. Although not all microsporidian genomes are small (size range from about 2 to 50 MB), it is suggested that the size of their genomes has been streamlined by natural selection. We explore the hypothesis that vertical transmission in microsporidia produces population bottlenecks, and thus reduces the effectiveness of natural selection. Here we compare the transposable element (TE) content of 47 microsporidian genomes, and show that genome size is positively correlated with the amount of TEs, and that species that experience vertical transmission have larger genomes with higher proportion of TEs. Our findings are consistent with earlier studies inferring that nonadaptive processes play an important role in microsporidian evolution.

## Background

Microsporidia, a phylum of fast evolving intracellular parasites related to fungi [1–3], are interesting models for studying the mechanisms shaping eukaryotic genome evolution. Their genomes are small and highly diverse. Two contrasting genome architectures are found in microsporidians. The extremely compact, gene-dense genomes of the encephalitozoonids ( $\cong$  2 Mb), which include human parasites, are the smallest known eukaryote parasites [4, 5]. At the other extreme are the large, gene-sparse genomes of several phylogenetically distant species, such as *Edhazardia aedis* ( $\cong$  50 Mb), *Hamiltosporidium* spp. ( $\cong$  17–25 Mb), *Nosema bombycis* ( $\cong$  15 Mb) and *Anncaliia algerae* ( $\cong$  12–17 Mb) [6–10]. Despite showing an about 25-fold difference in size, the gene-sparse and gene-dense genomes only differ by a factor of 2–3 in the number of proteins they encode, and a large fraction of the noncoding regions in the

gene-sparse genomes is populated by repetitive sequences, most of which have been identified as transposable elements [11]. It is suggested that the majority of transposable element (TE) insertions cause mildly deleterious effects in eukaryotes [12, 13]. Therefore, their spread and accumulation depend on the cellular mechanisms of TE silencing [14] as well as on the efficacy of natural selection to eliminate or retain them [15]. TEs are also known to escape cellular defense mechanisms by transferring themselves horizontally to different genomes [16, 17]. When a TE enters a genome for the first time, it can often duplicate freely before becoming epigenetically silenced [14]. Symbiotic organisms living in close proximity are more likely to exchange TEs [16, 18]; not surprisingly, a large number of horizontal gene transfer (HGT) events involving TEs have been identified in the evolutionary history of microsporidia [11].

We have recently proposed that nonadaptive processes have a major role in the evolution of microsporidian species that are predominantly vertically transmitted [8]. In such species genetic drift might override selection because the translocation of parasites from the host somatic cells to gametes reduces the parasite effective

\* Correspondence: [karen.haag@ufrgs.br](mailto:karen.haag@ufrgs.br)

<sup>1</sup>Department of Genetics and Post-Graduation Program of Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, RS 91501-970, Brazil

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

population size ( $N_e$ ) [19]. Moreover, vertical transmission might be predominant in host species with a meta-population structure, where local extinction and re-colonization is common and where the parasite enters new populations with the hosts dispersal stage and thus experiences severe bottlenecks [20].

Natural selection is efficient only if selection coefficients are greater than about  $1/2N_e$  [21]. In our previous study we showed that vertical transmission in a subset of microsporidian genomes is associated with a higher ratio of non-synonymous to synonymous substitutions [8]. Furthermore, microsporidia that experience vertical transmission usually have gene-sparse genomes [8]. Here we expand our analyses by characterizing the mobilomes from a larger number of genomes, and explore TE evolutionary dynamics in the light of microsporidian modes of transmission.

## Results and discussion

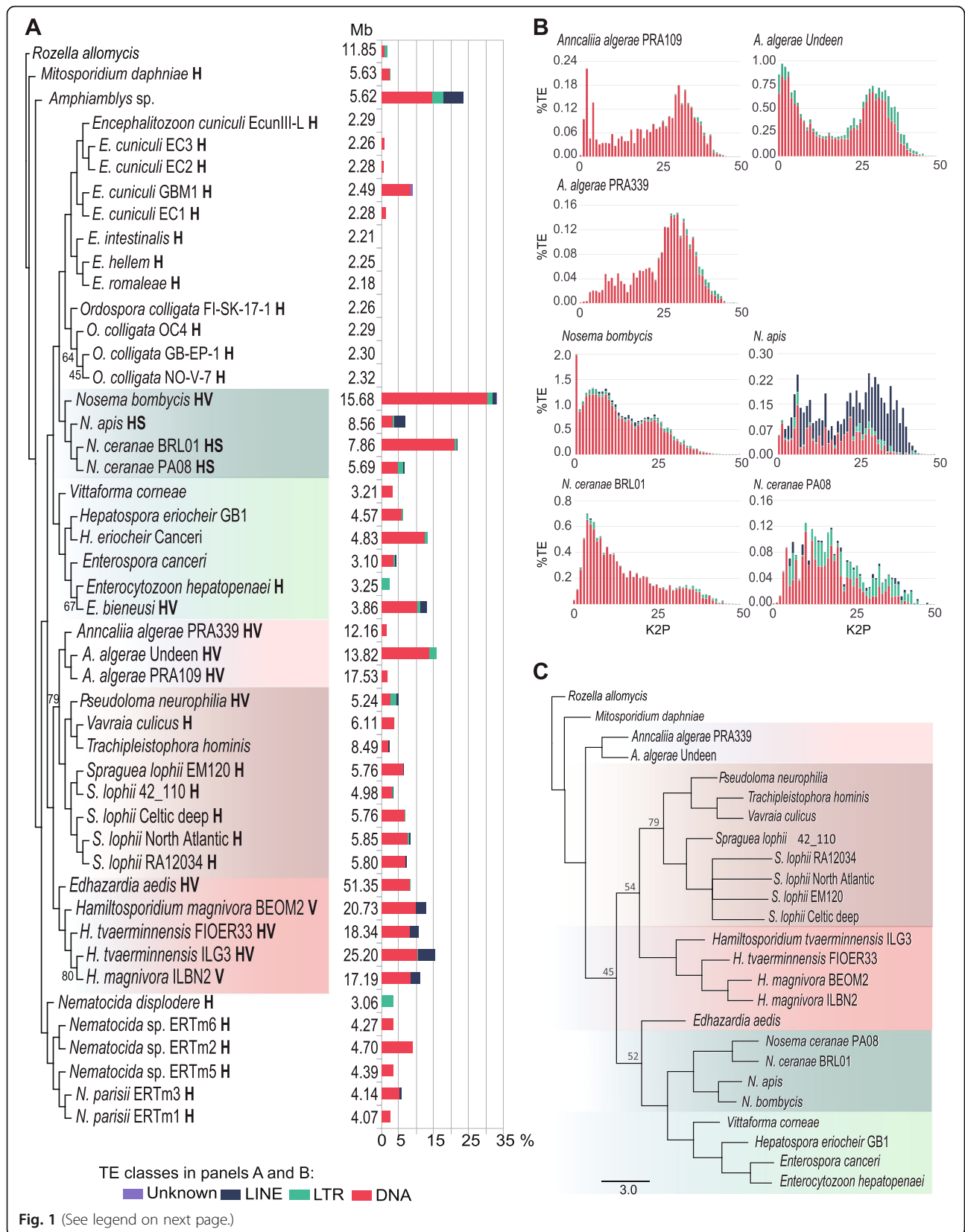
All 47 microsporidian genome assemblies available as of March 2019 were retrieved from GenBank (Additional file 1) and used for the identification of TEs. DNA transposons are the dominant class of TEs in microsporidian genomes, but their abundance varies greatly between species, and is not phylogenetically conserved (Fig. 1a; Additional file 2). The tree presented in Fig. 1a, built from a dataset of 37 highly conserved proteins (Additional file 3) established to reduce long-branch-attraction artifacts in microsporidian phylogenies [1, 2], is congruent with previously published microsporidian phylogenetic inferences based on other data [22, 23].

To evaluate whether TE accumulation in microsporidian lineages represents ancient or recent events, we used the Kimura 2-parameter distance (K2P) to generate age distributions of TEs within each genome (Additional file 4). The evolutionary dynamics of TEs shows very distinct patterns, even within species. For example, the three *A. algerae* genomes show variable amounts of TE content (1.64% in strain PRA109, 1.34% in PRA339 and 15.75% in Undeen; Additional file 1; Fig. 1a), and the accumulation of TEs in *A. algerae* Undeen apparently results from a recent spread of DNA transposons, mainly from the Merlin superfamily (Additional Files 5 and 6). A sharply bimodal distribution of TE abundance per age class suggests two successive expansions in *A. algerae* Undeen that are not observed in the remaining two strains (Fig. 1b). Similarly, the *Nosema* genomes from *N. bombycis* (33.41% of TEs) and *N. ceranae* BRL01 (21.64% of TEs) show an excess of young elements, which also seems to result from the recent spread of Merlin (Fig. 1b; Additional Files 5 and 6).

We reasoned that these recent events of TE expansion in some evolutionary lineages could result from the loss

of genes encoding components from cellular defense mechanisms, such as members of the RNAi pathway. We test this idea using the example of *Dicer* and *Argonaute*, two genes essential for the RNAi pathway. It was suggested that *Dicer* and *Argonaute* genes are selected as a unit in microsporidia, and that they were selectively maintained or lost during lineage divergence [24]. Therefore, we searched for orthologs of proteins related to cellular defense against TEs in the 47 proteomes (Additional file 6). No RIP/MIP-related proteins, which silence TEs by inducing mutation and/or cytosine methylation within repetitive sequences, were found in any lineage. Most strains carrying *Dicer* and *Argonaute* orthologs also carry the protein sets necessary for somatic quelling and meiotic silencing, but this is still no evidence of the activity of these mechanisms in microsporidia. The number of *Dicer* and *Argonaute* gene copies varies between 0 and 5 and 0–7, respectively, and in one lineage (*A. algerae* PRA 109) we found 2 copies of *Dicer*, but no *Argonaute* (Additional file 6). Thus, *Dicer* and *Argonaute* seem not necessarily selected as a unit. Interestingly, although there is an overall congruence between the *Dicer* + *Argonaute* tree (Fig. 1c) and the one built with 37 conserved proteins (Fig. 1a), there are two inconsistencies, i.e., the *A. algerae*, and the *E. aedis* clades, though with low statistical support.

A strong negative correlation was found between the percentage of total coding sequences (CDS; including both TE and host derived sequences) and genome size ( $\rho = -0.92$ ,  $p$ -value =  $2.2e-16$ ; Fig. 2a), indicating that non-coding regions are responsible for variations in genome size. On the other hand, there is a positive correlation between the fraction of genomes occupied by TEs and genome size ( $\rho = 0.66$ ,  $p$ -value =  $4.375e-07$ ; Fig. 2a). Regression analysis between TE fraction and genome size is also significant (R-squared = 0.33,  $p$ -value =  $1.501e-05$ ; Additional file 7). We further applied phylogenetic independent contrasts to the regression analysis [25] in order to correct for the non-independency of traits among taxa, which showed that the positive correlation between genome size and the amount of TEs in microsporidia does not result exclusively from phylogenetic inertia (R-squared = 0.16,  $p$ -value = 0.003; Additional file 7). However, some lineages with comparatively large genomes such as *A. algerae* PRA 109, PRA339 and *Edhazardia aedis* have TE fractions similar to small genomes such as *Encephalitozoon cuniculi* GBM1 and *Enterocytozoon bieneusi*. Both *E. cuniculi* GBM1 and *E. bieneusi* genomes suggest recent expansions of miniature inverted-repeat transposable elements according to the K2P age distributions of TEs (MITEs; Additional files 3 and 5). MITEs are deletion derivatives of DNA transposons from the *Tc1/Mariner* superfamily [26]. *Tc1/Mariner* elements are absent in *Encephalitozoon cuniculi* and



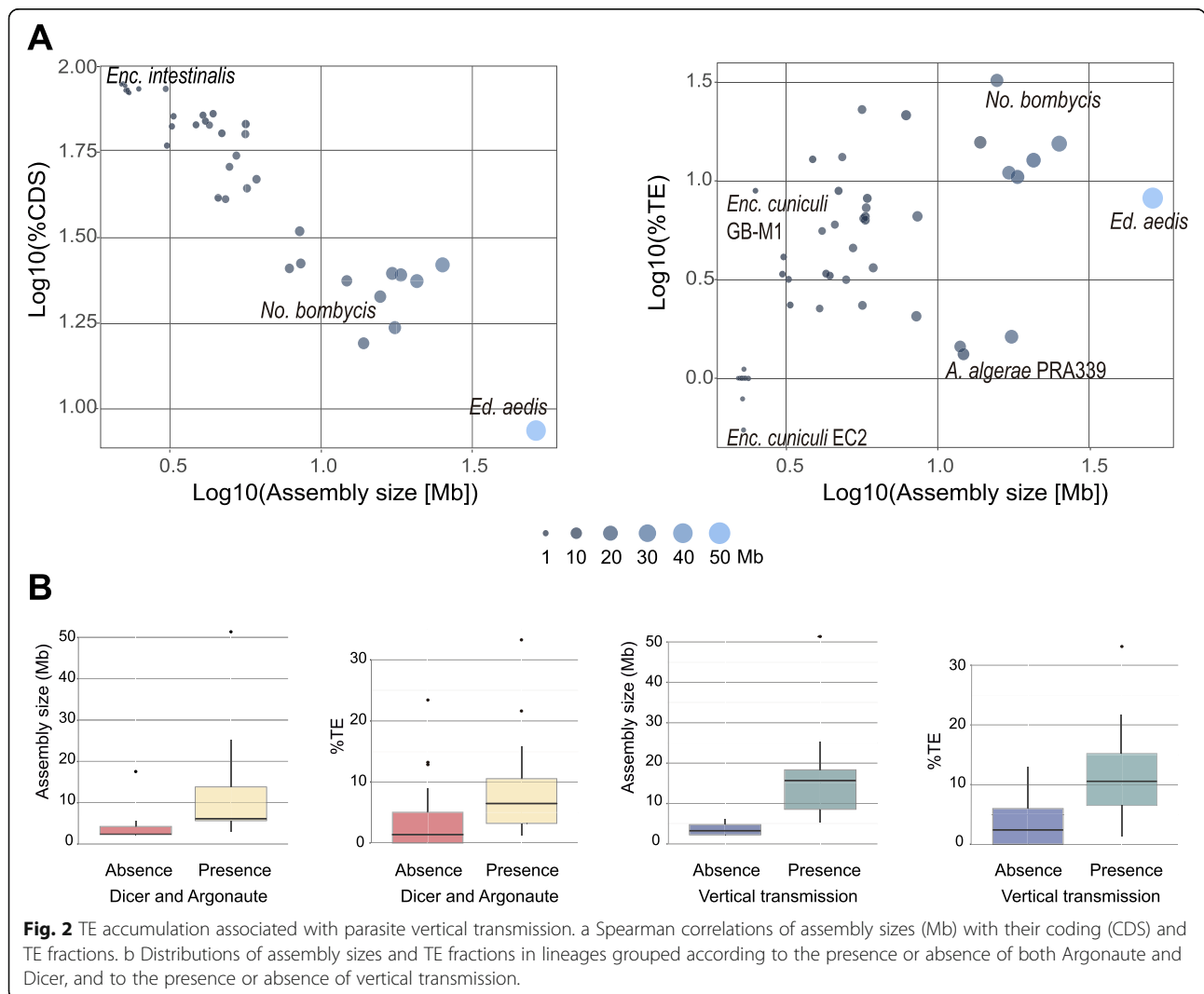
(See figure on previous page.)

**Fig. 1** TE dynamics in the history of microsporidia. a Assembly sizes (Mb) and their percentages of distinct TE classes are displayed next to the phylogeny of 47 microsporidia. The tree was inferred using 37 highly conserved proteins; only bootstrap support values lower than 90% are shown. Known modes of transmission are indicated in bold letters: H = horizontal; HV = horizontal and vertical; HS=Horizontal sexual. b Age distribution of different classes of TEs; larger Kimura distances (K2P) represent older elements. c Phylogeny inferred using concatenated Argonaute and Dicer proteins from 25 lineages of microsporidia. Only lineages where both proteins were found have been included in the analysis. *Rozella allomycis*, an early diverging fungus [3], is used as root in both (a) and (b) phylogenies. The major clades of microsporidia bearing both Argonaute and Dicer are highlighted in different colors

*Enterocytozoon bienersi* but are found in closely related microsporidia (*Nosema* spp. and *Hepatospora eriocheir*; Additional files 3 and 5).

We identified Dicer and Argonaute orthologs in 27 proteomes (Additional file 6). Strains with putative *Dicer* and *Argonaute* genes have, on average, a higher proportion of TEs (8.2%), and larger genome sizes (11.05 Mb) compared to strains in which these proteins are absent (4.06% and 3.81 Mb, respectively; Wilcoxon rank sum;  $p$ -

value  $< 0.004$ , and  $p$ -value  $< 2.2e-16$ , respectively; Fig. 2b). Although the association between the presence of both *Dicer* and *Argonaute* with the accumulation of TEs seems to support the idea that a RNAi machinery is being selectively maintained in microsporidian genomes attacked by selfish elements, we think that random events may have had an equivalent or perhaps more important role. First, there is only partial phylogenetic congruence between *Dicer* + *Argonaute* proteins (Fig. 1c)



and the history of microsporidia inferred from 37 highly conserved proteins (Fig. 1a), suggesting rate heterogeneity in the evolution of the microsporidian RNAi machinery. Second, genes involved in the RNAi pathway may have been lost by deletion or gained by duplications and/or HGT. For example, although it cannot be ruled out that genes might be lacking in the assemblies due to genome incompleteness, and that microsporidia may have mechanisms to eliminate TEs other than RNAi, *Argonaute* was apparently lost in *A. algerae* PRA109 but not in *A. algerae* PRA339, in spite of both strains having a similar TE age structure (Fig. 1b). We have also found evidence suggestive of *Argonaute* HGT involving *N. bombycis*; its *Argonaute* protein sequence is identical to *Papilio xuthus*, the Asian swallowtail butterfly (Additional file 8). The ortholog protein in *N. apis* has only 62.3% identity with the *P. xuthus* *Argonaute*. Previous comparative studies focusing on the enlargement of the *N. bombycis* genome suggested a major implication of HGT in the acquisition of new genes and TEs from its lepidopteran host *Bombyx mori* [10]. However, in the particular case of *Argonaute*, HGT may have occurred in the opposite direction. The genome of the butterfly *Papilio xuthus* encodes orthologs of *Argonaute* that cluster with microsporidian proteins (Additional file 8). Two out of three *P. xuthus* microsporidian-like *Argonaute* genes are found on large contigs (NW\_013531057.1, length = 195,660 bp; and NW\_013530778.1, length = 550,741 bp), weakening the hypothesis that they represent assembly contamination with microsporidian DNA.

Horizontal transfer of different genes has been reported in microsporidia [23, 27–30]. Among TEs, DNA transposons are the most often horizontally transferred in eukaryotes [16], and in *A. algerae*, several *Merlin* and *PiggyBac* transposon insertions were shown to result from HGT [11]. We hypothesize that the invasion and accumulation of TEs in microsporidian genomes is facilitated by their intracellular mode of life and by the low power of purifying selection. We further hypothesize that the latter is caused by low  $N_e$  in association with vertical transmission. Microsporidia use vertical transmission as a survival strategy when opportunities to transmit horizontally are reduced [31], thus being of particular importance during times of low density or the colonization of new host population, both of which cause population bottlenecks. Therefore, this mode of transmission may reduce  $N_e$  and thus the strength of natural selection [8]. Species with larger mobilomes, such as *N. bombycis* and *Hamiltosporidium* spp. show mixed mode transmission (a combination of horizontal and vertical transmission) [32, 33]. Microsporidia with mixed-mode transmission have, on average, a higher proportion of TEs than species with horizontal transmission. The difference in TE percentage between species

capable of vertical transmission (11.48%) and those showing only horizontal transmission (3.37%) is statistically significant (Wilcoxon rank sum test,  $p$ -value < 0.0006; Fig. 2b). The difference in genome size between these two groups is also significant (Wilcoxon rank sum test,  $p$ -value = 1.77e-05; Fig. 2b), with vertically transmitted species having larger genome sizes, on average (16.87 Mb, against 3.67 Mb, from species exclusively horizontally transmitted).

On closer inspection, one can see that some microsporidia that use vertical transmission do exhibit a low proportion of TEs (such as *A. algerae* PRA109, PRA339 and *E. aedis*; Fig. 1; Additional files 2 and 4). While these examples are few and do not reverse the overall statistical signal, they seemingly contradict our hypothesis. It is unclear if these low abundances are genuine or artefacts resulting from current methodological approaches. TEs are a challenge for genome assembly and annotation, and some sequencing pipelines may discard TEs during the assembly process and/or later during the filtering of contaminants [34]. Different sequencing platforms and pipelines were used to sequence and assemble the genomes used in our study, which might have interfered with subsequent TE annotation. Most TE annotation methods rely on homology-based approaches that limit the discovery of novel families of TEs, or require previous knowledge of structural features of TEs, being prone to false negatives [35]. Moreover, highly degenerated TEs without recognizable structural features can be missed from annotation [36, 37].

Taken together with our previous study [8] the here presented results show that vertical transmission is correlated with genome expansions and the spread of TEs in microsporidia. Thus, we suggest that nonadaptive forces play a strong role in shaping the evolution of genome size in populations of microsporidian parasites [13].

## Materials and methods

### TE annotation

For de novo identification of TE in all genomes we used RepeatModeler software with default parameters. RepeatModeler creates consensus sequences for each TE family and classifies based in homology. The output is TE lineage-specific libraries for each genome. We combined these TE lineage-specific libraries with all TE from fungi obtained from Repbase to be used in RepeatMasker software [38, 39]. RepeatMasker is a homology-based method that screens genomic sequences and annotates TEs by using BLAST (e-value threshold of 1e-10) as search engine. Low complexity DNA sequences and simple repeats were not masked. The annotation of Miniature Inverted-Repeat Transposable Elements (MITEs) insertions was performed with MITE Tracker [40]. The software identifies elements with valid Terminal Inverted Repeats (TIRs)

between 50 and 800 nt, and Target Site Duplications (TSDs). MITE candidates are filtered by flanking sequence (sequences outside the TSDs with 50 nt in length) similarity and grouped into families. MITEs undetected or classified as “unknown” by RepeatModeler were added or reclassified in the TE lineage-specific libraries. The tool “One code to find them all” [41] was used to filtrate retrotransposons longer than 80 bp and DNA transposons longer than 50 bp, with >80% of identity with the reference sequences. The output summarizes the number of TE copies and genome coverage for each TE family.

### Kimura distance and genome coverage

The Kimura 2-Parameter (K2P) divergence metric was calculated for TE sequences present in all genomes analyzed with RepeatMasker package RepeatLandscape [39, 42]. RepeatLandscape creates graphs depicting the contribution of repeat classes for each genome assembly and K2P distance from the consensus TE sequences contained in the RepeatMasker libraries.

### Highly conserved proteins for phylogenetic analyses

We searched the 47 genome assemblies for highly conserved proteins [1, 2] using BLASTp (e-value threshold of  $1e-15$ ). The conserved proteins needed to be present in at least half of the analyzed proteomes for not being removed in the multiple alignment trimming phase. We identified 37 conserved proteins in at least 50% of the proteomes (Additional file 3).

### Genome defense mechanisms against TEs

We searched for proteins related to genome defense mechanisms against TE [43]. BLASTp searches with e-value threshold of  $1e-5$  were performed for using *N. crassa* and other ascomycete proteins as queries: RNA-dependent RNA polymerase (Q1K6C4), DNA repair protein RAD51 (Q1K929), DNA repair protein RAD52 (Q9HGI2), DNA repair protein Rad54 (Q9P978), Replication protein A (Q1K7R9), RNA helicase (Q7S3A3), Msc1 (Methyltransferase from *Ascobolus 1*) (O13369), RID (RIP defective) (V51NW3). BLASTp searches with e-value threshold of  $1e-8$  were performed using *Hamiltosporidium tvaerminnensis* FI-OER-3-3 Dicer (A0A4Q9LBM7) and Argonaute (A0A4Q9L0B4) as queries (Additional file 6).

### Phylogenies

Protein or nucleotide sequences were aligned with MUSCLE and uninformative columns in the alignments, containing at least 50% of gaps, were trimmed using Geneious Prime 2019.0.4 [44, 45]. The best amino acid substitution models fitting our data were selected with ProtTest 3.4.2 [46]. The 37 conserved proteins were concatenated, and the final alignment contained 12,467 sites. The best model according with the AIC and BIC

criteria was VT + I + G. Argonaute and Dicer proteins were concatenated and the final alignment contained 1822 sites. The best model was LG + I + G. In order to investigate the direction of Argonaute HGT, we selected and aligned representative nucleotide sequences from *Nosema* spp. and Lepidoptera. The best model according with the AIC criteria was GTR + G. The maximum likelihood phylogenies were estimated using in RAXML 8.2.10 [47], and the statistical support for each branch was estimated by 1000 bootstrap replicates.

### Statistical tests

In order to test correlations between TE content and genome size all data were logarithmically transformed. We applied Shapiro-Wilks method [48] to test the normality of our data with the function `shapiro.test()` in R version 3.6.2. The data were not normally distributed therefore nonparametric tests were applied. The Spearman correlation between TE percentage in the genome and genome size, as well as total CDS length and genome size, was determined by using the function `cor.test()` in R [49]. The package `ape` in R [50] was used to perform a regression analysis between the amount of TEs and genome size using the function `fit.pic()`. We performed a Wilcoxon rank sum test with the function `wilcox.test()` in R to test for a difference in TE percentage between lineages horizontally transmitted vs. lineages with mixed-mode transmission, and genome size between lineages horizontally transmitted and lineages with mixed-mode transmission (Additional file 9). We also tested the difference in TE percentage between genomes encoding for both Dicer and Argonaute proteins vs. genomes where these proteins are absent. The plots were constructed using `ggplot2` package in R [51].

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13100-020-00218-8>.

**Additional file 1.** Microsporidian assemblies analyzed in our study.

**Additional file 2.** Content of distinct TE classes in the microsporidian assemblies.

**Additional file 3.** List of the 37 highly conserved proteins employed for phylogenetic reconstruction.

**Additional file 4.** Distribution of the Kimura 2-Parameter (K2P) divergence metric calculated for all TE sequences present in all genomes analyzed in our study.

**Additional file 5.** Frequency of subtypes of TEs in microsporidian mobilomes.

**Additional file 6.** Results of BLASTp searches for proteins involved in genome defense mechanisms against TEs.

**Additional file 7.** Regression analyses of the amount of TEs in relation to genome size with and without phylogenetic independent contrasts.

**Additional file 8.** Protein alignment of representative microsporidian and lepidopteran Argonaute sequences; Argonaute phylogeny testing the direction of the HGT event.

**Additional file 9.** Known transmission modes of the microsporidia included in our study.

### Abbreviations

HGT: Horizontal Gene Transfer; K2P: Kimura 2-Parameter distance; RIP: Repeat-induced Point mutation; MIP: Methylation Induced Premeiotically; MITE: Miniature Inverted-repeat Transposable Element; TIR: Terminal Inverted Repeat; TSD: Target Site Duplication; CDS: Coding sequence; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion

### Acknowledgements

We especially thank to Sidia Maria Callegari Jacques for her help with the statistical analyses, and to Jean-François Pombert for the critical reading of this manuscript.

### Authors' contributions

N.R.M.A. acquired and analyzed the data, designed the study, interpreted the data, and wrote the manuscript. D.E. interpreted the data, and wrote the manuscript. K.L.H. designed the study, interpreted the data, and wrote the manuscript. All the authors approved the final manuscript.

### Funding

This work was supported by the Swiss National Science Foundation grant numbers 31003A\_146462 and 310030B\_166677 to D.E., Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant PQ 302121/2017-0 to K.L.H., and fellowship to N.R.M.A.

### Availability of data and materials

The datasets generated as part of the study are available as supplementary information. Please contact the authors for further data request.

### 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 Genetics and Post-Graduation Program of Genetics and Molecular Biology, Federal University of Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, RS 91501-970, Brazil. <sup>2</sup>Department of Environmental Sciences, Zoology, Basel University, Vesalgasse 1, 4051 Basel, Switzerland.

Received: 16 April 2020 Accepted: 18 June 2020

Published online: 23 June 2020

### References

- Capella-Gutierrez S, Marcet-Houben M, Gabaldon T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biol.* 2012;10:47.
- Haag KL, James TY, Pombert J-F, Larsson R, Schaer TMM, Refardt D, et al. Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *P Natl Acad Sci USA.* 2014;111:15480–5.
- James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. *Curr Biol.* 2013;23:1548–53.
- 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 Commun.* 2010;1:77.
- 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.
- Williams BA, Lee RC, 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;9:200.
- Keeling P, Fast NM, Corradi N. Microsporidian genome structure and function. In: Louis M, Weiss BJJ, editors. *Microsporidia: Pathogens of Opportunity*. Ames: Wiley; 2014. p. 221–9.
- 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.
- Corradi N, Haag KL, Pombert J-F, Ebert D, Keeling PJ. Draft genome sequence of the *Daphnia* pathogen *Octospora bayeri*: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. *Genome Biol.* 2009;10:R106.
- Pan G, Xu J, Li T, Xia Q, Liu S-L, Zhang G, et al. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics.* 2013;14:186.
- 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.
- Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. *Biol Direct.* 2011;6:19.
- Lynch M, Conery JS. The origins of genome complexity. *Science.* 2003;302:1401–4.
- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8:272–85.
- Song MJ, Schaack S. Evolutionary conflict between Mobile DNA and host genomes. *The Am Nat.* 2018;192:263–73.
- Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. *P Natl Acad Sci USA.* 2017;114:4721–6.
- Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol.* 2010;25:537–46.
- Panaud O. Horizontal transfers of transposable elements in eukaryotes: the flying genes. *Comptes Rendus Biologies.* 2016;339:296–9.
- Mira A, Moran NA. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb Ecol.* 2002;44:137–43.
- Sheikh-Jabbari E, Hall MD, Ben-Ami F, Ebert D. The expression of virulence for a mixed-mode transmitted parasite in a diapausing host. *Parasitology.* 2014;141:1097–107.
- Ohta T. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst.* 1992;23:263–86.
- Vossbrinck CR, Debrunner-Vossbrinck BA. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol.* 2005;52:131–42.
- Pombert J-F, 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.
- Huang Q. Evolution of dicer and Argonaute orthologs in microsporidian parasites. *Infect Genet Evol.* 2018;65:329–32.
- Felsenstein J. Phylogenies and the comparative method. *Am Nat.* 1985;125:1–15.
- Feschotte C, Mouchès C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol Biol Evol.* 2000;17:730–7.
- Pombert J-F, 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. *P Natl Acad Sci USA.* 2012;109:12638–43.
- Corradi N. Microsporidia: Eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. *Annu Rev Microbiol.* 2015;69:167–83 null.
- Selman M, Pombert J-F, Solter L, Farinelli L, Weiss LM, Keeling P, et al. Acquisition of an animal gene by microsporidian intracellular parasites. *Curr Biol.* 2011;21:R576–7.
- 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.
- Turner PE, Cooper VS, Lenski RE. Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution.* 1998;52:315–29.
- Han M-S, Watanabe H. Transovarial transmission of two microsporidia in the silkworm, *Bombyx mori*, and disease occurrence in the progeny population. *J Invertebr Pathol.* 1988;51:41–5.
- Vizoso DB, Ebert D. Phenotypic plasticity of host–parasite interactions in response to the route of infection. *J Evol Biol.* 2005;18:911–21.

34. Goerner-Potvin P, Bourque G. Computational tools to unmask transposable elements. *Nat Rev Genet.* 2018;19:688–704.
35. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 2019;20:275.
36. Juretic N, Bureau TE, Bruskewich RM. Transposable element annotation of the rice genome. *Bioinformatics.* 2004;20:155–60.
37. Bergman CM, Quesneville H. Discovering and detecting transposable elements in genome sequences. *Brief Bioinform.* 2007;8:382–92.
38. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6:11.
39. Smit A, Hubley R, Green P. RepeatMasker. Institute for Systems Biology; 2013. Available from: <http://www.repeatmasker.org/RepeatModeler/>.
40. Crescente JM, Zavallo D, Helguera M, Vanzetti LS. MITE tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics.* 2018;19:348.
41. Bailly-Bechet M, Haudry A, Lerat E. "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. *Mob DNA.* 2014;5:13.
42. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16:111–20.
43. Gladyshev E. Repeat-induced point mutation and other genome defense mechanisms in fungi. *The Fungal Kingdom: Wiley;* 2017. p. 687–99. Available from: <https://doi.org/10.1128/9781555819583.ch33>.
44. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
46. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164–5.
47. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
48. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika.* 1965;52:591–611.
49. R Core Team. R: The R Project for Statistical Computing. 2019 [cited 2020 Feb 22]. Available from: <https://www.r-project.org/>.
50. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019;35:526–8.
51. Wickham H. ggplot2: elegant graphics for data analysis. 1st ed. 2009. Corr. 3rd printing 2010 edition. New York: Springer; 2010.

## 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)

