Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Contrasting roles of horizontal transfer and endogenous transposable element dynamics in microsporidian genome evolution

Anastasia Markelova
Institute of Cytology
Mikhail Rayko
mike.rayko@gmail.com
Institute of Cytology
Elena Nassonova
Institute of Cytology

Research Article

Keywords: Microsporidia, transposable elements, horizontal gene transfer, genome evolution, LINE retrotransposons, Dictyocoela, Nosema, Hamiltosporidium

Posted Date: July 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10422774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Microsporidia are obligate intracellular parasites with highly variable genome sizes, ranging from 2.3 to over 50 Mb, despite a nearly constant proteome size. Transposable elements (TEs) are major contributors to this size variation, yet the dynamics of TEs accumulation, including the roles of horizontal transfer from hosts versus endogenous element activation, remain poorly understood for most lineages. Here we present a comprehensive *de novo* TEs characterization for two microsporidian parasites of the amphipod *Gammarus roeselii* — *Dictyocoela muelleri* and *Nosema granulosis* — and for the distantly related *Hamiltosporidium tvaerminnensis*, a parasite of *Daphnia magna*. We identified a clade of LINE/CR1 elements shared among both *D. muelleri*, *N. granulosis*, and their host *G. roeselii* (> 90% nucleotide identity), representing plausible candidates for host-to-parasite horizontal gene transfer. However, these elements contribute less than 0.15% of each parasite genome and therefore cannot account for genome expansion. In contrast, in *H. tvaerminnensis* a single LINE/Dong-R4 family (rnd-1_family-215) occupies approximately 6% of the genome and shows two distinct divergence peaks on Kimura landscapes, indicative of ancient and recent transpositional bursts. No evidence of HGT from the host *Daphnia magna* was detected for this element. Together, these results demonstrate that microsporidian genome expansion is primarily driven by endogenous TE bursts facilitated by vertical transmission bottlenecks and genetic drift, with horizontal transfer playing a secondary, taxon-specific role.

Introduction

Microsporidia constitute a highly derived, species-rich clade of obligate intracellular eukaryotic parasites placed within the supergroup Holomycota (Nucletmycea), as sister to the rozellids (Bass et al. 2018; Williams et al. 2022). More than 1,700 species infecting representatives of virtually all animal phyla have been described, and environmental sequencing continues to reveal extensive undescribed diversity. Despite an enormous host range, microsporidian cells share a highly derived organization: mitochondria are reduced to mitosomes lacking a genome, no canonical Golgi apparatus is present, ribosomes are prokaryote-sized, and the unique extrusion apparatus — including the polar tube, anchoring disc, polaroplast, and posterior vacuole — is universally conserved as the invasion mechanism (Han and Weiss 2017; Jespersen et al. 2022).

A hallmark of microsporidian biology is extreme genome compaction. The smallest known eukaryotic nuclear genomes occur in *Encephalitozoon* species (2.3–2.9 Mb), achieved through massive gene loss, intron elimination, shortened intergenic regions, and complete absence of transposable elements (Katinka et al. 2001; Williams et al. 2022). Yet genome size varies more than 20-fold across the phylum, while estimated proteome size varies only 2–3-fold (Haag et al. 2020). This paradox is largely explained by differential TE accumulation: species with larger genomes typically harbor diverse and abundant TE populations, whereas compact-genome relatives carry few or none (Nakjang et al. 2013; de Albuquerque et al. 2020).

Transmission strategy has a strong, well-documented effect on TEs abundance. Horizontal transmission generally maintains large effective population sizes (N*e*) and efficient purifying selection, keeping TEs loads low. Vertical (transovarial) transmission imposes repeated population bottlenecks, reducing Ne and allowing mildly deleterious TE insertions to accumulate by genetic drift — a microsporidian instance of Muller's ratchet (Haag et al. 2020; Lynch and Conery 2003). Several vertically transmitted species show dramatically inflated genomes relative to their horizontal counterparts, consistent with this model (Cormier et al. 2021)

A second mechanism of TEs acquisition is horizontal gene transfer (HGT) between host and parasite. HGT of DNA transposons (notably PiggyBac elements) from lepidopteran hosts into *Nosema bombycis* is well documented, and broader surveys have identified metazoan-derived transposons in multiple microsporidian genomes (Pan et al. 2013; Parisot et al. 2014). Evidence for HGT of non-LTR retrotransposons (LINE elements) is considerably rarer, however, and systematic comparisons of TEs repertoires across parasite–host pairs remain scarce.

In this study, we leverage a natural tripartite system to investigate TE dynamics. We analyze the genomes of two microsporidian parasites, *Dictyocoela muelleri* and *Nosema granulosis*, and their shared crustacean host, *Gammarus roeseli* (both parasites are known to be vertically transmitted). For comparison, we include a phylogenetically distant microsporidian with a large genome, *Hamiltosporidium tvaerminnensis*, a parasite of *Daphnia magna*. Our goals were to: (1) characterize and compare *de novo* TE repertoires across all four organisms; (2) test for potential horizontal transfer of TEs between the *G. roeselii* parasites and their host; and (3) analyze the contribution of a specific repeat families to the genome expansion of organisms.

Materials and methods

Genomic data

All genome assemblies were retrieved from NCBI GenBank: *Dictyocoela muelleri* (GCA_016256075.1), *Nosema granulosis* (GCA_015832245.1), *Hamiltosporidium tvaerminnensis* (GCA_022605425.2), and *Gammarus roeseli* (GCA_016164225.1). Assembly statistics relevant to the TEs analysis are summarized in Table 1.

Table 1
Assembly and annotation statistics for the four genomes analyzed

Feature	***D. muelleri***	***N. granulosis***	***H. tvaerminnensis***	***G. roeseli* (host)**
Total length (Mb)	41.9	8.8	21.64	3,239
Number of contigs	11,512	1,754	17	1,130,582
N50 (kb)	6.6	12.7	1,440	4.81
GC content (%)	26.14	31.58	26.60	43
Repeat content (%)	58.80	25.14	39.56	52
Gene count	6,442	3,639	3,573	—
BUSCO completeness (%)	81.0	94.6	94.0	—

De novo repeat annotation

Repeats were identified and annotated using Earl Grey v1.0.8 (Baril et al. 2024), which integrates RepeatModeler2 v2.0.4 and RepeatMasker v4.1.6 with an iterative BEAT (Blast–Extract–Align–Trim) extension procedure. In brief, RepeatModeler2 generated seed consensus libraries, which were then iteratively extended by extracting up to 20 genomic copies per family, padding with 1 kb flanks, and refining boundaries through

MAFFT v7.550 multiple alignment. Extension halted when a consensus failed to grow by at least 50% of the added flank length. Tandem Repeat Finder v4.10 was used to identify and exclude sequences comprising > 90% tandem repeats. LTR_Finder v1.07 was run in parallel to capture full-length LTR retrotransposons, and RepeatCraft was used to defragment overlapping and split hits. All runs used parameters -r eukarya -d yes -e yes; the -e yes flag activated the HELIANO module for Helitron detection. Kimura substitution level (K2P) divergence landscapes were computed with calcDivergenceFromAlign.pl (part of RepeatMasker) and plotted in Python 3.11 (pandas 2.2.2, matplotlib 3.8.4).

Consensus library clustering and domain annotation

Redundancy in *de novo* libraries was reduced with CD-HIT-EST v4.8.1 applying the 80-80-80 rule (≥ 80bp size and ≥ 80% identity over ≥ 80% of the shorter sequence length). Open reading frames were predicted with EMBOSS getorf (minimum 300 nt, internal stop codons excluded) and transeq; predicted proteins were searched against Pfam 37.3 using pfam_scan.pl. BLAST copy-number estimates were produced with blastn (BLAST+ v2.16.0) requiring alignment length ≥ 50% of query length and ≥ 80% identity; overlapping hits were merged to avoid double-counting.

Horizontal transfer analysis

To test for host-to-parasite horizontal transfer, parasite repeat consensus libraries were queried against the host repeat library using blastn, and all four libraries were pooled and clustered with CD-HIT-EST (80-80-80 threshold). Clusters containing sequences from ≥ 2 taxa were treated as HGT candidates. Candidate sequences were aligned with MAFFT v7.526 (--auto) and a maximum likelihood phylogeny was inferred with IQ-TREE2 v2.4.0 (automatic model selection, 1000 ultrafast bootstrap replicates). Trees were visualized using iTOL.

Characterization of the H. tvaerminnensis LINE/Dong-R4 element

For the dominant rnd-1_family-215 element, all blastn hits (≥ 80% identity) were extracted, multiply aligned using MAFFT and refined with T-Coffee v13.46.1 (+ rm_gap 80), and a consensus sequence was derived with EMBOSS cons. An HMM profile was built with HMMER v3.4 to recover highly diverged copies. Structural analysis used TE-Aid (parameters -T -R --min-orf 300). Conserved domain searches employed NCBI CD-Search against Core Nucleotide and Nucleotide Collection (nt) databases, and Dfam 3.9 database. Genomic distribution of copies was visualized on chromosome-level contigs using karyoploteR v1.33.0 in R v4.5.0. Horizontal transfer from *Daphnia magna* was evaluated by (i) blastn search of the microsporidian consensus against the *D. magna* repeat library and (ii) joint CD-HIT-EST clustering of all four TE libraries at 80–90% identity thresholds.

Results

TE landscape overview: composition differs markedly between the three microsporidian species

De novo annotation with Earl Grey revealed substantially different TE compositions among the three microsporidian genomes (Table 2; Fig. 1). In *D. muelleri*, repeats constitute 58.8% of the 41.9-Mb genome, dominated by DNA transposons (32.5% of the genome) and LTR retrotransposons (9.9%), with LINE elements contributing only 0.5%. Unclassified elements account for a further 31.0%. *N. granulosis* has a more compact genome (8.8 Mb) with 25.1% repetitive sequences; here DNA transposons again predominate (13.7%), with LTR elements (8.7%) and LINE elements (3.9%) also present. *H. tvaerminnensis* differs fundamentally: its 21.6-Mb genome carries 39.6% repeats, but LINE elements are the largest classified component (10.0%), with DNA transposons reduced to 6.5% and LTR elements nearly absent (0.2%). The host *G. roeselii* shows a yet different composition, with LINE elements accounting for 22.7% of the genome — the dominant TE class.

Table 2
De novo transposable element annotation summary

TE Classification	***D.muelleri***		***N. granulosis***		***H. tvaerminnensis***		***G. roeselii***	
	% Genome Coverage	**TE Family Count**	**% Genome Coverage**	**TE Family Count**	**% Genome Coverage**	**TE Family Count**	**% Genome Coverage**	**TE Family Count**
DNA	32,54	875	13,73	137	6,50	78	3,89	1673
Rolling Circle	0,00	NA	0,13	5	0,31	26	0,07	212
Penelope	0,00	NA	0,00	NA	0,11	6	1,23	219
LINE	0,52	17	3,90	48	9,98	88	22,69	2352
SINE	0,00	NA	0,00	NA	0,01	7	0,08	114
LTR	9,88	212	8,66	47	0,21	41	3,88	3277
Other (RNA, Simple Repeat, Microsatellite)	3,30	1497	0,88	507	1,62	1096	4,75	26775
Unclassified	30,99	843	9,43	204	24,78	150	16,60	2534
Non-Repeat	22,77	NA	63,26	NA	56,48	NA	46,82	NA

Kimura divergence landscapes (Fig. 1) show that all four genomes share a prominent peak near 5% K2P divergence, indicating a relatively recent transpositional burst shared across lineages or reflecting independent synchronous activity. Both microsporidian parasites of *G. roeselii* additionally display an older peak near 30% K2P divergence attributable primarily to LTR elements, absent in *H. tvaerminnensis*. CD-HIT-EST clustering reduced library redundancy substantially: from 2,020 to 1,446 families in *D. muelleri*, 455 to 314 in *N. granulosis*, and 276 to 153 in *H. tvaerminnensis*. Domain annotation via Pfam identified canonical TE-associated signatures in all three parasites — DDE-type transposases (DDE_1, DDE_3, rve, Tnp families), reverse transcriptases (RVT_1, RT_RNaseH), integrases, and zinc-finger domains — confirming the presence of diverse, at least partially intact TE families.

LINE/CR1 elements bridging host and parasites: candidates for horizontal transfer

Phylogenetic analysis of pooled repeat consensus sequences from all four organisms identified a strongly supported clade of LINE/CR1 elements containing sequences from both microsporidian parasites (*D. muelleri* and *N. granulosis*) and their host *G. roeselii*. Internal branch lengths within this clade are minimal and bootstrap support is high (Fig. 2), consistent with recent shared ancestry. In CD-HIT-EST clustering at 80% identity threshold, the same sequences co-cluster into a single group, with pairwise identities of 90–97% across all taxa (Table 3). The basal position in the clade is occupied by host sequence rnd-1_family-42, suggesting that this *G. roeselii* element is most similar to the putative ancestral form, and that one or more transfer events from host to parasite gave rise to the microsporidian copies.

Table 3
Co-clustering LINE/CR1 sequences from the genomes of *Gammarus roeselii*, *Dictyocoela muelleri*, and *Nosema granulosis*

Organism	Sequence ID	Identity (%)	Pfam domains
D. muelleri	rnd-1_family-1067#LINE/CR1	95.41	RVT_1
D. muelleri	rnd-1_family-281#LINE/CR1	91.03	RVT_1
D. muelleri	rnd-1_family-684#LINE/CR1	96.21	—
N. granulosis	rnd-2_family-322#LINE/CR1	93.69	RVT_1
N. granulosis	rnd-1_family-174#LINE/CR1	92.80	RVT_1
N. granulosis	rnd-1_family-265#LINE/CR1	95.48	RVT_1
G. roeselii (host)*	rnd-1_family-14#LINE/CR1	100 (ref)	Exo_endo_phos_2, RVT_1
G. roeselii (host)	rnd-1_family-11#LINE/CR1	97.44	RVT_1

** Reference (centroid) sequence for this CD-HIT cluster. Identities are relative to rnd-1_family-14.*

All microsporidian copies carry a RVT_1 (reverse transcriptase) domain; the full-length host sequence additionally encodes an Exo_endo_phos_2 (the endonuclease region of retrotransposons) domain required for target-priming reverse transcription (TPRT)-mediated integration. The absence of the endonuclease domain from all microsporidian copies raises uncertainty about their current mobility - these may represent degenerate, non-autonomous derivatives of the host element. Despite the striking sequence similarity, the combined genomic coverage of these LINE/CR1 elements in *D. muelleri* and *N. granulosis* is below 0.15% per genome, making them negligible contributors to genome size. We cannot exclude the alternative that some LINE/CR1 copies in the parasites reflect sequencing contamination from host DNA, a non-trivial concern given the obligate intracellular lifestyle. Targeted long-read re-sequencing and chromosome-level assemblies of parasite genomes would be needed to definitively resolve this question.

LINE/Dong-R4 burst drives genome expansion in Hamiltosporidium tvaerminnensis

Among the *H. tvaerminnensis* LINE elements, a single family — rnd-1_family-215, classified as LINE/Dong-R4 — dominates the landscape (Table 4). Earl Grey estimates 8.3% genome coverage for this family, corresponding to ~ 1,795 kb and approximately 1,263 genomic copies, making it one of the largest single TE contributions reported for a microsporidian genome. Three additional Dong-R4 families contribute modestly (rnd-1_family-230, rnd-2_family-76, rnd-1_family-53), but rnd-1_family-215 alone accounts for the bulk of LINE expansion.

Table 4
Dominant LINE/Dong-R4 families in *Hamiltosporidium tvaerminnensis*

Consensus ID	Coverage (bp)	Copies	Genome %	Pfam domains
rnd-1_family-215	1,795,449	1,263	8.30	RVT_1
rnd-1_family-230	59,172	150	0.27	RVT_1
rnd-2_family-76	13,897	28	0.06	RVT_1
rnd-1_family-53	108,344	203	0.50	RhoGAP (Rho GTPase-activating protein)

The TE-Aid analysis identified five potential domain structures (Fig. 3d), showing weak matches with gag and pol domains of LINE elements from the L1 and CR1 families, as well as with pol domains of Gypsy LTR retroelements. The low similarity, short alignments, and low bit-score values (< 40) indicate a highly degraded, ancient LINE-type retrotransposon whose RT and endonuclease domains are fragmented and may therefore yield signals from different clades. CD-Search against Pfam identified a single conserved domain: RT_nLTR_like (cd01650; $E = 1.34 \times 10^{-33}$), overlapping with RVT_1 (pfam00078; $E = 1.11 \times 10^{-20}$) in the central region of the consensus (approximately positions 1,579–2,334). No endonuclease domain (APE/HUH-type) was detected, consistent with the element being a degenerate or highly derived LINE with only the RT core preserved.

BLAST search against the NCBI Nucleotide Collection (nt) identified the best non-self hits in *Vairimorpha necatrix* chromosomes (57% query coverage, 65% identity) and in *Mitosporidium daphniae* (14% query coverage, 98.9% identity in the RT domain). The Dfam 3.9 match ($E = 6 \times 10^{-29}$) corresponds to the family TE2310_SO2_FAM0374, a LINE/Dong-R4 element annotated from the insect pest *Sitophilus oryzae* (grain weevil) — a host not known to be parasitized by *H. tvaerminnensis*, raising the possibility of a more ancient evolutionary connection that warrants further investigation.

Karyoplots of rnd-1_family-215 insertion distribution across the 17 *H. tvaerminnensis* chromosomes reveal no clustering near telomeres or centromeres (Fig. 4). This uniform, dispersed distribution is consistent with published data on LINE/Dong-R4 elements, which typically lack sequence or chromatin specificity for integration, in contrast to target-site-specific LINEs such as R1/R2. No evidence of horizontal transfer from *D. magna* was detected: blastn of the rnd-1_family-215 consensus against the *D. magna* repeat library produced no hits, and joint CD-HIT-EST clustering yielded exclusively monophyletic, single-taxon clusters for all Dong-R4 families.

Discussion

Two contrasting TE evolutionary scenarios in microsporidian parasites

Our comparative analysis reveals two qualitatively distinct scenarios for TE dynamics in microsporidian parasites. In the system comprising *D. muelleri* and *N. granulosis* on their host *G. roeselii*, we identify LINE/CR1 sequences with > 90% identity shared across both parasites and the host, an identity level difficult to explain by vertical inheritance alone given the expected neutral evolutionary rate. These elements emerge as the strongest candidates for host-to-parasite horizontal gene transfer yet reported for non-LTR retrotransposons in microsporidia. However, their negligible genome coverage (< 0.15%) means they cannot explain the substantial genome size differences between these species and their compact relatives. This is consistent with the general model that HGT of TEs, while evolutionarily significant as a phenomenon, rarely drives genome-wide size changes (Schaack et al. 2010; Parisot et al. 2014).

In *H. tvaerminnensis*, a completely different picture emerges. Here, a single endogenous LINE/Dong-R4 family has achieved extraordinary copy numbers — occupying ~ 6–8% of a 21.6-Mb genome — through at least two distinct transpositional waves. The absence of HGT evidence from *Daphnia magna* and the monophyletic distribution of all Dong-R4 families in phylogenetic clustering strongly suggest that this expansion is entirely endogenous. The bimodal Kimura landscape most parsimoniously indicates reactivation of a dormant element after an initial ancient colonization, analogous to the LINE-1 reactivation events documented in certain mammalian lineages (Grahn et al. 2005). Vertical transmission in *H. tvaerminnensis*, which is transmitted transovarially to *Daphnia* offspring, imposes the bottlenecks and reduced N*e* that make such expansions possible: purifying selection is weakened, and mildly deleterious insertions accumulate through drift (Haag et al. 2020).

Mechanisms and molecular features of the detected elements

The LINE/CR1 candidate HGT elements in *D. muelleri* and *N. granulosis* encode only the reverse transcriptase domain (RVT_1) and lack the endonuclease module present in the full-length host copy. This domain structure characterizes many non-autonomous LINE-derived fragments (SINEs or degenerate derivatives) and raises the question of whether these elements retain any mobility. If they are indeed immobile degenerate copies, they may represent 'dead' transposition events that nonetheless provide a footprint of past HGT. Alternatively, they could exploit the endonuclease encoded by autonomous copies elsewhere in the genome, as documented for certain non-autonomous LINE families.

The rnd-1_family-215 element in *H. tvaerminnensis* similarly retains only an RT core, with no identifiable endonuclease despite exhaustive searching against Pfam, Dfam, and NCBI databases. The TE-Aid 5'-truncation profile and the abundance of partial copies (791 fragments vs. 105 full-length) are both hallmarks of an element that retrotransposes efficiently but generates many 5'-incomplete insertions — typical LINE behavior during a burst. The mystery of how integration occurs without a detectable endonuclease remains open; possibilities include an endonuclease encoded on a different ORF beyond the annotated consensus boundary, reliance on a cellular endonuclease, or extreme sequence divergence that renders the domain unrecognizable by current HMMs. The intriguing match to a *Sitophilus oryzae* Dong-R4 element in Dfam — despite this insect not being a known host — could reflect a wider, phylogenetically ancient distribution of this LINE subfamily among eukaryotes, predating any specific host–parasite associations.

Implications for microsporidian genome evolution

Viewed alongside published data, our results support a nuanced picture of microsporidian genome evolution in which neither genome reduction nor expansion is a single-trajectory process. The dominant narrative of reductive evolution — convergent loss of metabolic pathways, intron elimination, shortening of intergenic regions — applies to many lineages but is punctuated in others by TEs-driven genome inflation (Williams et al. 2022; Nakjang et al. 2013). What determines which trajectory a given lineage follows appears to be primarily a function of (1) transmission strategy and the resulting effective population size, (2) the specific TE repertoire inherited from ancestral holomycotans, and (3) stochastic reactivation events that may be triggered by environmental stress, immune evasion selection, or epigenetic relaxation.

Our finding that *D. muelleri* harbors a genome nearly five times larger than *N. granulosis* (41.9 vs. 8.8 Mb), despite both being vertically transmitted parasites of the same host species, suggests that even within a single host-parasite ecological context, TE dynamics can diverge dramatically. *D. muelleri* carries far more DNA transposon families than *N. granulosis* (875 vs. 137), and the proportion of unclassified repeats — which may include highly diverged, ancient TEs — is also larger. Whether this reflects a more permissive epigenetic TE silencing environment, a historical bottleneck that was more severe in the *D. muelleri* lineage, or differences in the ancestral TE repertoire inherited before divergence from a common ancestor remains to be determined.

Caveats and future directions

Several limitations of this study merit explicit acknowledgment. First, both *D. muelleri* and *N. granulosis* genomes are highly fragmented (N50 of 6.6 and 12.7 kb respectively), which can inflate TE family counts and hamper accurate assessment of full-length element structure. Long-read (PacBio HiFi or Oxford Nanopore) assemblies of these parasites would substantially improve TE characterization. Second, the possibility of host DNA contamination in parasite assemblies cannot be excluded for the LINE/CR1 HGT candidates: because separation of obligate intracellular parasites from host tissue is imperfect, reads from *G. roeselii* could co-assemble into parasite bins. Third, our analysis focused on three microsporidian species and one host; broader taxon sampling within the *Dictyocoela*/*Nosema* parasitic complex and across *Hamiltosporidium* (including the recently described *H. magnivora*) would clarify whether the observed TE patterns are widespread. Finally, the functional significance of ongoing TE activity — for gene regulation, adaptation to new hosts, or immune evasion — represents an important avenue for future experimental investigation.

Conclusions

This study provides the first systematic, genome-wide characterization of transposable elements in *D. muelleri*, *N. granulosis*, and *H. tvaerminnensis*, accompanied by a parallel analysis of their host genomes. We demonstrate that: (i) LINE/CR1 elements shared among two vertically transmitted microsporidian parasites and their amphipod host at > 90% nucleotide identity represent the most compelling evidence to date for non-LTR retrotransposon horizontal transfer in microsporidia, though their negligible genomic footprint precludes a role in genome expansion; (ii) a single LINE/Dong-R4 family in *H. tvaerminnensis* has undergone at least two rounds of expansion and now constitutes ~ 6% of the parasite genome, driven by endogenous reactivation rather than horizontal acquisition from the host; and (iii) microsporidian genome size evolution is shaped predominantly by the interplay of vertical transmission bottlenecks, genetic drift, and endogenous TE burst

dynamics, with horizontal transfer playing a secondary, lineage-specific role. These findings contribute to a growing understanding of Holomycota genome evolution and the diverse molecular strategies underlying the paradox of extreme parasitism.

Declarations

Competing interests.

The authors declare no competing interests.

Funding.

This work was supported by the Russian Science Foundation, grant No. 23-74-00071 ("Integral approach to the study of early evolution of Holomycota").

Author Contribution

A.M. performed analysis and prepared the figures. A.M. and M.R. wrote the manuscript. E.N. conceived and supervised the study and critically revised the manuscript. All authors reviewed and approved the final version.

Data availability.

All genome assemblies are publicly available at NCBI GenBank under the accession numbers listed in Materials and Methods. TE consensus libraries and annotation scripts are deposited at GitHub (https://github.com/natsyam/Microsporidia_repeats/).

References

1. Baril T, Galbraith J, Hayward A (2024) Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. Mol Biol Evol 41:msae068
2. Bass D et al (2018) Clarifying the Relationships between Microsporidia and Cryptomycota. J Eukaryot Microbiol 65:773–782
3. Cormier A et al (2021) Comparative Genomics of Strictly Vertically Transmitted, Feminizing Microsporidia Endosymbionts of Amphipod Crustaceans. Genome Biol Evol 13:evaa245
4. de Albuquerque NRM, Ebert D, Haag KL (2020) Transposable element abundance correlates with mode of transmission in microsporidian parasites. Mob DNA 11:19
5. Grahn RA et al (2005) Extinction of LINE-1 activity coincident with a major mammalian radiation in rodents. Cytogenet Genome Res 110:407–415
6. Haag KL et al (2020) Microsporidia with Vertical Transmission Were Likely Shaped by Nonadaptive Processes. Genome Biol Evol 12:3599–3614

7. Han B, Weiss LM (2017) Microsporidia: Obligate Intracellular Pathogens Within the Fungal Kingdom. Microbiol Spectr 5(2)
8. Jespersen N, Monrroy L, Barandun J (2022) Impact of Genome Reduction in Microsporidia. Exp Suppl 114:1–42
9. Katinka MD et al (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature 414:450–453
10. Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404
11. Nakjang S et al (2013) Reduction and Expansion in Microsporidian Genome Evolution: New Insights from Comparative Genomics. Genome Biol Evol 5:2285–2303
12. Pan G et al (2013) Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. BMC Genomics 14:1–14
13. Parisot N et al (2014) Microsporidian Genomes Harbor a Diverse Array of Transposable Elements that Demonstrate an Ancestry of Horizontal Exchange with Metazoans. Genome Biol Evol 6:2289–2300
14. Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537–546
15. Wells JN, Feschotte C (2020) A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet 54:539–561
16. Williams BAP, Williams TA, Trew J (2022) Comparative genomics of microsporidia. In: Microsporidia: Current Advances in Biology. Springer, Cham, pp 43–69

Figures

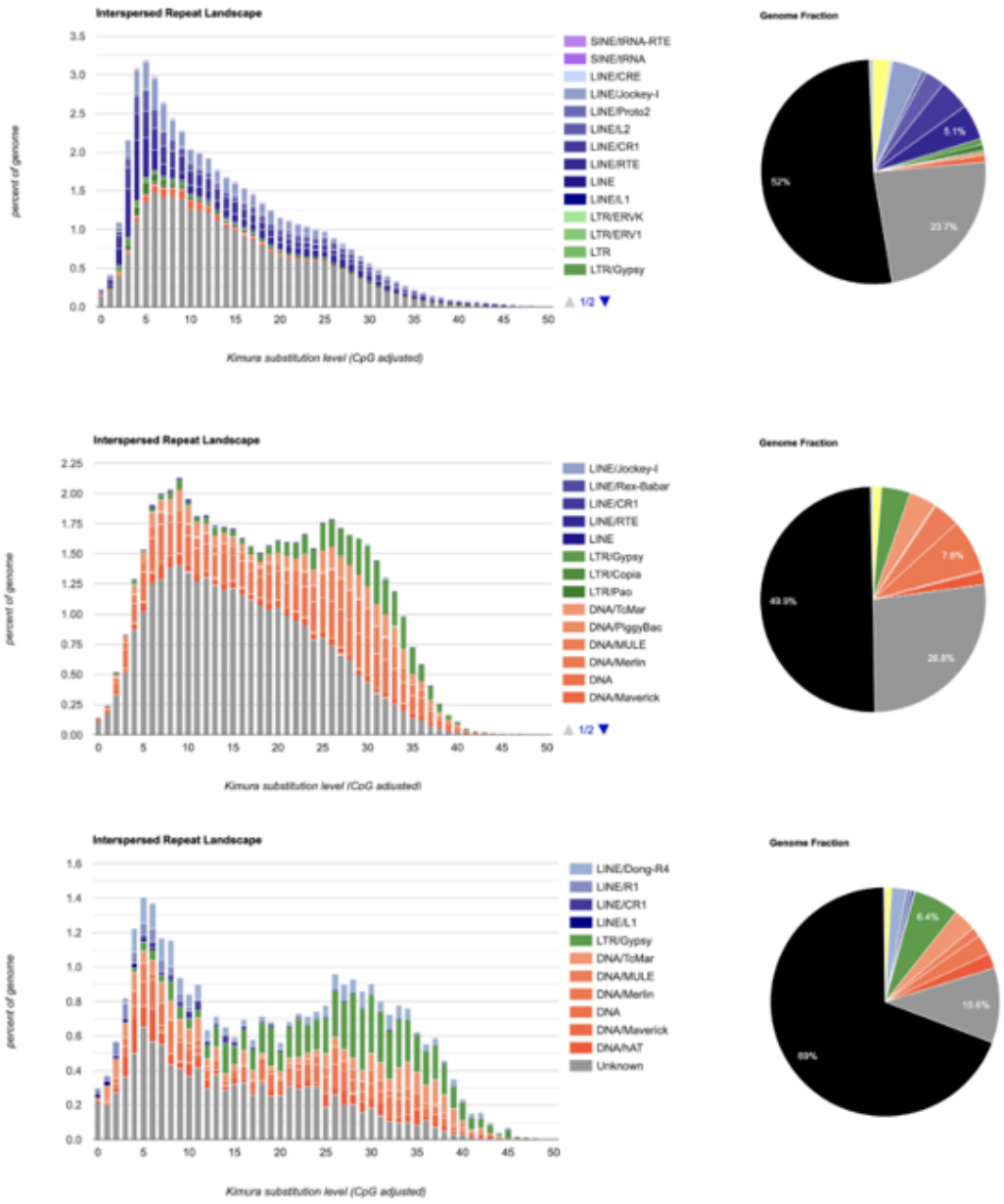
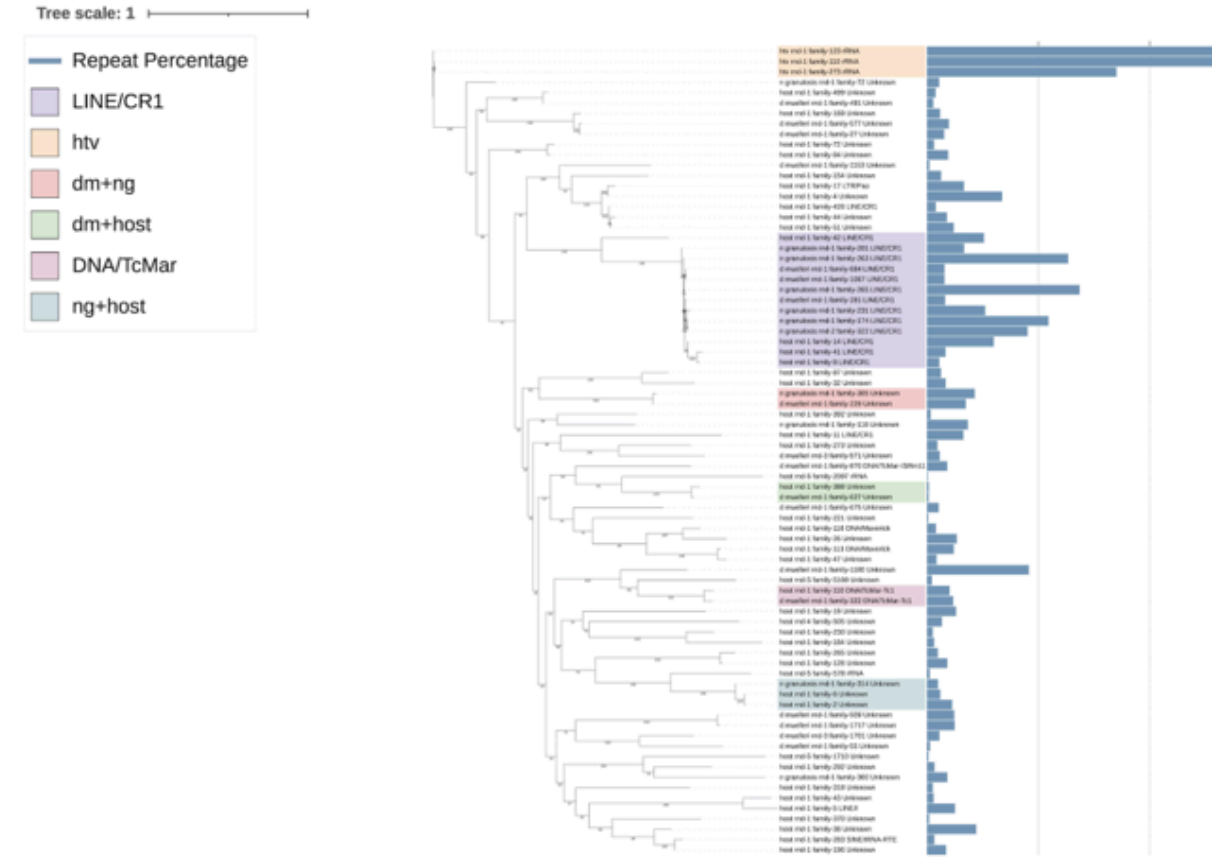
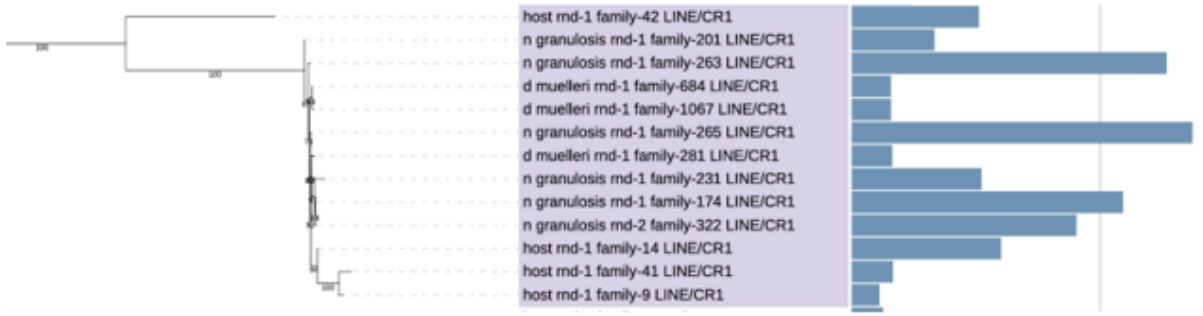


Figure 1

Kimura landscapes for the genomes of *Gammarus roeselii*, *Dictyocoela muelleri*, and *Nosema granulosis*, respectively; the x-axis shows the percentage of divergence (K2P), and the y-axis shows the percentage of the genome covered. The black portion of the pie chart represents the proportion of the genome without repetitive elements; the colors indicate different classes of repeats, and gray represents the proportion of unclassified repeats



A



B

Figure 2

A. Phylogenetic tree of repeats identified in genomes from all four organisms in this study. The tree was constructed using the maximum likelihood method in IQ-TREE (TIM+F+G4 model, 1,000 bootstrap iterations); bootstrap support values are shown next to the branches. The scale on the tree shows the expected number of amino acid substitutions per position. To the right, the percentage of the genome occupied by the given repeat is indicated for each sequence. **B**. A fragment of the phylogenetic tree of repeats showing the clade LINE/CR1, which includes the sequences of *G. roeselii* and both of its parasites, *D. muelleri* and *N. granulosis*; bootstrap support values are given next to the branches

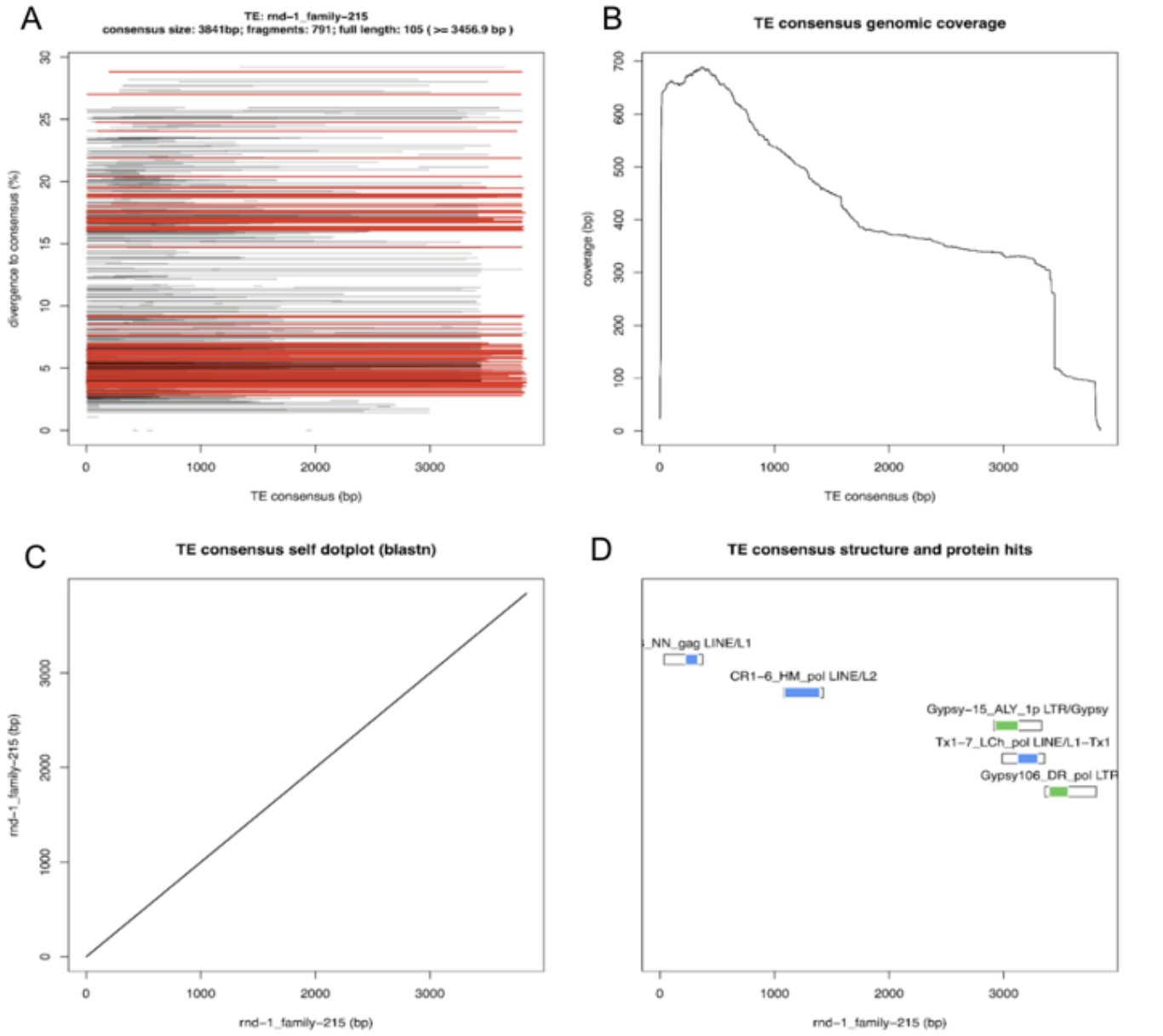


Figure 3

TE-Aid analysis of the dominant LINE/Dong-R4 element rnd-1_family-215 in *Hamiltosporidium tvaerminnensis*. **A.** Copy divergence plot showing two distinct transpositional waves (recent burst at ~5% K2P; ancient insertions at ~20–25% K2P). **B.** Coverage profile across the consensus sequence, illustrating the 5'-truncation pattern characteristic of TPRT-mediated retrotransposition. **C.**Self dot-plot of the consensus; absence of terminal repeats confirms non-LTR status. **D.** Predicted ORFs and matches to transposon protein databases. Weak matches (22–39% identity, E-values of 0.2–3.7) with gag and pol domains of LINE elements (blue)) and with pol domains of Gypsy LTR retroelements (green)

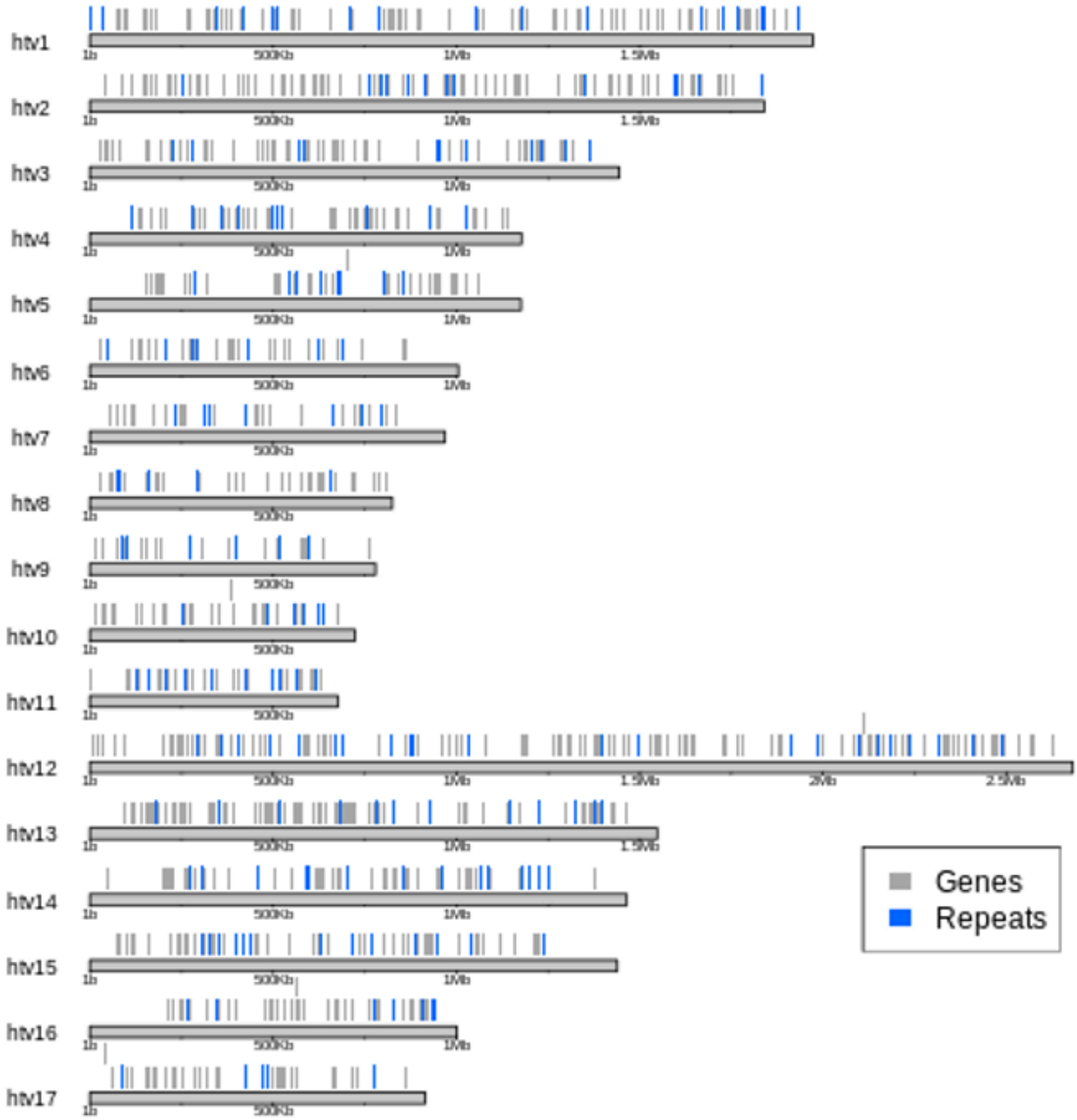


Figure 4

Chromosomal distribution of rnd-1_family-215 copies across the *Hamiltosporidium tvaerminnensis* genome.