UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIAS BÁSICAS DA SAÚDE
CURSO DE GRADUAÇÃO EM BIOMEDICINA

Nathalia Rammé Medeiros de Albuquerque

AVALIAÇÃO DO CONTEÚDO DE ELEMENTOS GENÉTICOS MÓVEIS EM DUAS ESPÉCIES PRÓXIMAS DE MICROSPORÍDIOS

Porto Alegre
2017

Nathalia Rammé Medeiros de Albuquerque

AVALIAÇÃO DO CONTEÚDO DE ELEMENTOS GENÉTICOS MÓVEIS EM DUAS ESPÉCIES PRÓXIMAS DE MICROSPORÍDIOS

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Biomedicina.

Área de habilitação: Biologia molecular

Orientador: Profª Drª Karen Luisa Haag

Porto Alegre

2017

CIP - Catalogação na Publicação

de Albuquerque, Nathalia Rammé Medeiros
AVALIAÇÃO DO CONTEÚDO DE ELEMENTOS GENÉTICOS MÓVEIS EM DUAS ESPÉCIES PRÓXIMAS DE MICROSPORÍDIOS / Nathalia Rammé Medeiros de Albuquerque. -- 2017.
37 f.
Orientadora: Karen Luisa Haag.

Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciências Básicas da Saúde, Curso de Biomedicina, Porto Alegre, BR-RS, 2017.

1. Elementos transponíveis. 2. Microsporídios. 3. Retrotransposons. I. Haag, Karen Luisa, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Nathalia Rammé Medeiros de Albuquerque

AVALIAÇÃO DO CONTEÚDO DE ELEMENTOS GENÉTICOS MÓVEIS EM DUAS ESPÉCIES PRÓXIMAS DE MICROSPORÍDIOS

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Ciências Básicas da Saúde da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharel(a) em Biomedicina.

Aprovado em: ____ de _____ de _____.

BANCA EXAMINADORA

_____

Lilian Caesar - UFRGS

_____

Gilberto Cavalheiro Vieira - UFRGS

_____

Profª Drª Karen Luisa Haag - UFRGS

RESUMO

Os microsporídios possuem um dos menores genomas entre os eucariotos devido ao parasitismo intracelular obrigatório, que resultou em uma extensiva redução no tamanho e complexidade do genoma. Elementos transponíveis são segmentos de DNA que podem se mover de um *locus* para outro dentro de um genoma. Esses elementos compõe uma grande fração dos genomas eucarióticos, sendo a maioria retrotransposons devido ao seu mecanismo de transposição "copiar-e-colar". São capazes de mutar, alterar a regulação de genes e gerar novos genes. Eventos de transposição podem ser deletérios quando elementos transponíveis se integram em regiões codificantes e interrompem genes importantes. Devido à ausência de recombinação, esses elementos podem acumular no genoma de organismos assexuados e contribuir para a extinção desses organismos.

No presente estudo foi analisado o conteúdo de elementos móveis de duas espécies proximamente relacionadas de microsporídios que diferem no modo de reprodução. *Hamiltosporidium tvaerminnensis* é assexuado e é transmitido horizontal e verticalmente, *H. magnivora* é sexuado transmitido verticalmente ao seu hospedeiro. Foram analisados dois genomas de linhagens de *H. magnivora* (BE-OM-2 e IL-BN-2), e um genoma de uma linhagem de *H. tvaerminnensis* (FI-OER-3-3). O programa RepeatMasker foi usado para a triagem inicial do conteúdo de elementos transponíveis. Um conjunto de domínios de proteínas codificadas por retrotransposons LTR foi utilizado para realizar um BLASTp nos proteomas preditos de BE-OM-2, IL-BN-2 e FI-OER-3-3. Por fim, a ferramenta LTRharvest foi utilizada para detectar retrotransposons LTR inteiros nos genomas.

Foi encontrada uma diferença significativa entre as linhagens, com acúmulo de elementos transponíveis na linhagem sexuada IL-BN-2 em contraste com o baixo número de cópias na linhagem assexuada FI-OER-3-3. Foram encontrados dois retrotransposons LTR potencialmente ativos, "Nanomov" e "Bigshot". Ambos elementos são compartilhados entre as três linhagens, indicando inserções fixas que ocorreram antes da divergência das espécies. A filogenia de "Bigshot" mostrou elementos homólogos distribuídos em diversas espécies de microsporídios. Também está relacionado com elementos encontrados em artrópodes como *D. magna*, indicando uma possível transferência gênica horizontal entre microsporídios e seus hospedeiros artrópodes ancestrais.

Para melhor compreender o papel dos elementos transponíveis na arquitetura e evolução dos genomas e a diferença no conteúdo em organismos sexuados e assexuados, é necessário uma investigação mais profunda em diferentes famílias de retrotransposons e transposons de DNA.

Palavras-chave: Microsporídios. Elementos Trasponíveis. Retrotransposons LTR.

ABSTRACT

Microsporidia have one of the smallest genomes among eukaryotes, due to their obligate intracellular parasitism that resulted in an extensive reduction in genome size and complexity. Transposable elements (TEs) compose a high fraction in eukaryotic genomes, the vast majority being retrotransposons, due to their "copy-paste" transposition mechanism. TEs are able to mutate genes, alter gene regulation and generate new genes. Transposition events can be deleterious when TEs integrate in active coding regions and disrupt important genes. Due to the absence of recombination, TEs may accumulate in asexual genomes and contribute to extinction of asexual organisms.

In this study it was assessed the genomes of three *Hamiltosporidium* lineages for TE content. Two genomes of *H. magnivora* lineages (BE-OM-2 and IL-BN-2) that reproduce sexually and are vertically transmitted, and one genome of an *H. tvaerminnensis* lineage (FI-OER-3-3) that reproduces asexually and is both vertically and horizontally transmitted. The RepeatMasker software was used for initial screening of TE content. A set of core protein domains encoded by LTR retrotransposons was used in BLASTp searches in the proteomes of BEOM2, ILBN2 and FIOER33. Finally, LTRharvest was used to detect full-length elements.

It was found a significant difference in the TE content between asexual/sexual lineages, with excess TE accumulation in the sexual lineage IL-BN-2 in contrast with the low copy number of TEs in the asexual lineage FI-OER-3-3. It was identified two potentially active LTR retrotransons, "Nanomov" and "Bigshot". Both elements are shared between the three lineages, suggesting fixed insertions that occurred before the divergence of the species. The phylogenetic analysis showed elements homologous to “Bigshot” are distributed in the genomes of different species of microsporidia as well as arthropods. A similar is present in *D. magna*, suggesting horizontal gene transfer between parasite and host at some point of their evolution.

To better understand the role of transposable elements in reduced genomes and the impact of different TE loads in sexuals and asexuals, we need further investigation on these retroelements, identifying other superfamilies and include DNA transposons in the analysis.

Keywords: Microsporidia. Transposable Elements. LTR retrotransposons.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AP	Protease aspártica (do inglês, Aspartic Protease)
cDNA	DNA complementar
TE	Elemento Transponível (do inglês, Transposable Element)
ORF	Fase de Leitura Aberta (do inglês, Open Reading Frame)
INT	Integrase
LTR	Longas Repetições Terminais (do inglês, Long Terminal Repeats)
RT	Transcriptase Reversa (do inglês, Reverse Transcriptase)
RH	RNAse H

SUMÁRIO

1. INTRODUÇÃO

1.1 OS MICROSPORÍDIOS

Microsporidia é um filo pertencente ao reino dos fungos, cujos representantes são unicelulares, parasitas intracelulares obrigatórios e com genomas altamente reduzidos entre os eucariotos. Há cerca de 1,500 espécies e 200 gêneros descritos atualmente que infectam uma vasta gama de espécies de animais, inclusive humanos, especialmente pacientes imunodeprimidos (BECNEL; TAKVORIAN; CALI, 2014). Os microsporídios são um grupo ancestral altamente derivado dos fungos. Possuem características resultantes do parasitismo intracelular, refletindo em baixa complexidade genômica e metabólica. Não possuem mitocôndrias, peroxissomos e complexo de Golgi convencionais, mas organelas análogas, como os mitossomos que derivaram das mitocôndrias por degeneração. Os microsporídios foram classificados como leveduras quando foram descobertos, em 1857 por Nägeli. Posteriormente foram reclassificados como protozoários, integrantes do extinto Reino Archezoa, e finalmente, após evidências filogenéticas, novamente no Reino Fungi (THOMARAT; VIVARÈS; GOUY, 2004).

O ciclo de vida desses organismos consiste em uma fase infecciosa com esporo maduro, a primeira fase proliferativa, na qual há produção de merozoitos (merogonia) e na segunda fase proliferativa com a produção de esporos (esporogonia). As formas infectantes e altamente especializadas dos microsporídios, são os esporos, geralmente ovais, que variam de 1 a 40 micrômetros de diâmetro. Cada esporo é composto pela camada mais externa proteica, o exósporo; pela camada interna quitinosa, o endósporo; e pela membrana plasmática envolvendo o esporoplasma. O esporoplasma é rico em ribossomos, contém um ou dois núcleos (diplocarion), e três estruturas responsáveis pela infecção da célula hospedeira: o polaroplasto, o tubo polar e o vacúolo posterior (Figura 1). A função do tubo polar é a perfuração da membrana da célula hospedeira que permite a transferência do esporoplasma diretamente no citoplasma da célula. Uma vez dentro do hospedeiro, se inicia a fase proliferativa (merogonia) com a produção de células-filhas (merozoitos), com potencial de repetir a fase de merogonia ou entrar na segunda fase proliferativa, a de esporogonia.

Figura 1 – Ilustração da anatomia do esporo de microporídeo.

Fonte: adaptado de KEELING; FAST, 2002.

Microsporídios possuem alta variação no tamanho do genoma e o gênero *Hamiltosporidium* tem um genoma comparativamente grande dentro do filo (Tabela 1). Espécies como *Anncaliia algerae*, *Nosema bombycis, Nosema ceranae* e *Edhazardia aedis* possuem grandes porções de regiões não-codificantes, elementos transponíveis e sequências de DNA repetitivo (XU *et al*., 2006; WILLIAMS *et al*., 2008 CORNMAN *et al*., 2009; PARISOT *et al*., 2014). Interessantemente, os genomas mais enxutos do filo são totalmente desprovidos de elementos móveis.

Tabela 1 – Comparação do tamanho do genoma de microsporídios.

Espécie	Tamanho do genoma (Mb)	Hospedeiro	Nº de retrotransposons	Nº de transposons
Anncaliia algerae (Peyretaillade et al. 2012)	23	Mamíferos, insetos	101	139
Edhazardia aedis (Williams et al. 2008)	Não definido	Mosquitos	32	1
Encephalitozoon cuniculi (Katinka et al. 2001)	2.9	Mamíferos	0	0
Encephalitozoon hellem (Pombert et al. 2012)	2.5	Mamíferos	0	0
Enterocytozoon bieneusi (Akiyoshi et al. 2009; Keeling et al. 2010)	>6	Humanos	0	0
Hamiltosporidium tvaerminnensis (Corradi et al. 2009)	>24.2	*Daphnia*	17	8
Nematocida parisii (Cuomo et al. 2012)	>4,1	Nematódeos	4	2
Nosema bombycis (Pan et al. 2013)	15-16	Insetos	25	32
Nosema ceranae (Cornman et al. 2009)	>7.86	Insetos	7	15
Spraguea lophii (Campbell et al. 2013)	6.2-7.3	Peixes	4	8

Fonte: PARISOT *et al*., 2014

1.1.1. Os *Hamiltosporidium*

O gênero *Hamiltosporidium* é composto pelas espécies *Hamiltosporidium tvaerminnensis* e *H. magnivora* que diferem quanto aos modos de transmissão e de reprodução. Enquanto *H. magnivora* é transmitido verticalmente e se reproduz sexuadamente, *H. tvaerminnensis* pode ser

transmitido tanto horizontal como verticalmente ao hospedeiro, e se reproduz assexuadamente. Os *Hamiltosporidium* infectam o microcrustáceo *Daphnia magna.* Essa relação se tornou um bom modelo para estudos de interação e co-evolução de parasita-hospedeiro devido à facilidade de manipulação e manutenção das linhagens de *D. magna* em cultura. Esse microcrustáceo possui ausência de pigmentação no corpo, o que auxilia na identificação de infecções, curto tempo de gerações e a possibilidade de induzir a partenogênese, facilitando a obtenção de linhagens isogênicas (EBERT, 2011).

1.2 ELEMENTOS TRANSPONÍVEIS

Elementos transponíveis (*Transposable Elements,* TEs) são segmentos de DNA que podem se mover de um *locus* para outro dentro de um genoma. Constituem uma grande fração do genoma na maioria dos organismos eucariotos e têm a habilidade de mutar, alterar a regulação e gerar novos genes. São considerados parasíticos, pois seu sucesso replicativo está negativamente relacionado com o valor adaptativo de seus hospedeiros, uma vez que podem se integrar em regiões codificadoras e interromper genes importantes. Porém, TEs têm um papel fundamental na geração de variação intra-específica e na evolução (FESCHOTTE; PRITHAM, 2007). Existem duas grandes classes de TEs: elementos de classe I ou retrotransposons, que utilizam um intermediário de RNA reversamente transcrito para se mover no genoma hospedeiro; e elementos de classe II ou transposons que utilizam intermediário de DNA em um processo mais simples de excisão e inserção em outro *locus* do genoma. O mecanismo utilizado pelos retrotransposons facilita sua disseminação, levando a um elevado número de cópias em grande parte dos genomas eucarióticos nos quais estão inseridos (FESCHOTTE; PRITHAM, 2007). Os retrotransposons são subdivididos em retrotransposons com longas repetições terminais (*Long Terminal Repeats,* LTR) e os retrotransposons não-LTR. As LTR são sequências que flanqueiam a região codificadora, que nos retrotransposons LTR autônomos, consiste nos genes *gag* e *pol*. O gene *gag* codifica proteínas estruturais que constituem uma pseudopartícula viral. O gene *pol* codifica diversas enzimas como protease, que cliva a poliproteína Pol; transcriptase reversa, que polimeriza o DNA complementar (cDNA) a partir do RNA do retrotransposon; ribonuclease H (RNase H), uma endonuclease que cliva a ligação fosfodiéster do RNA no híbrido RNA:DNA; e uma integrase que integra o cDNA no genoma hospedeiro. A estrutura e capacidade de codificação são similares aos retrovírus, porém os retrovírus possuem uma fase de leitura aberta (*Open*

Reading Frame - ORF) adicional que codifica a poliproteína do envelope (*env*) viral, embora alguns retrotransposons LTR também carreguem esse gene. As famílias de retrotransposons LTR mais abundantes e bem descritas na literatura são Ty1/copia e Ty3/gypsy. Ty1/copia é a linhagem mais ancestral e Ty3/gypsy é segunda maior linhagem de retrotransposons LTR, ambas amplamente distribuídas em plantas, fungos e animais. A família Ty3/gypsy é mais próxima aos retrovírus e a mais abundante nos genomas eucarióticos (EICKBUSH; MALIK, 2002). Os retrotransposons LTR encontrados em microsporídios pertencem a família Ty3/gypsy. Ao menos três subgrupos são específicos de microsporídios e diferem de outros elementos Ty3/gypsy por possuírem uma única poliproteína em vez de duas ORFs para *gag* e *pol* (PARISOT *et al*., 2014).

1.2.1 Elementos transponíveis e modo de reprodução

O sucesso replicativo de um elemento transponível dentro de uma população está relacionado não só ao número de cópias por genoma, como ao número de genomas que contém ao menos uma cópia desse elemento. Elementos transponíveis não são capazes de colonizar novos genomas em organismos assexuados com padrão clonal de reprodução. Já em organismos sexuados, a disseminação desses elementos em uma população se dá através da singamia (fusão dos gametas). Portanto, o sucesso replicativo de elementos transponíveis depende da reprodução sexuada (HICKEY, 1982). Entretanto, se uma linhagem eucariótica assexuada herda TEs de seu ancestral sexuado, pode haver um acúmulo desses elementos devido à ausência de recombinação e a perda de alguns ou todos os mecanismos de supressão, como metilação e RNA interferente, dependendo da forma através da qual a assexualidade é adquirida (KRAAIJEVELD *et al*., 2012). O acúmulo de elementos transponíveis deletérios levaria essa linhagem à extinção (DOLGIN; CHARLESWORTH, 2006). Em rotíferos da Classe Bdelloidea, um grupo de organismos considerado ancestralmente assexuado, foi detectado um baixo número de cópias de retrotransposons nas espécies *P. roseola* e *A. vaga*, enquanto transposons de DNA são numerosos em uma grande variedade de genomas de rotíferos bdeloides (ARKHIPOVA; MESELSON, 2000; GLADYSHEV; MESELSON; ARKHIPOVA, 2007). Também foi observado um maior número de TEs em genomas da planta *Arabidopsis lyrata*, que faz polinização cruzada, comparativamente à *A. thaliana*, que é autopolinizadora (DE LA CHAUX *et al*., 2012). Além disso, cinco linhagens independentes de artrópodes assexuados possuem um menor número de TEs nos seus genomas quando

comparadas às espécies relacionadas sexuadas (BAST *et al*., 2015). Em microsporídios já foram detectadas diversas famílias de TEs, inclusive famílias exclusivas desse filo, sendo na sua maioria compostas de transposons (PARISOT *et al*., 2014). A diferença no conteúdo de TEs entre linhagens sexuadas e assexuadas de microsporídios não foi analisada até o momento.

1.3 ESTRATÉGIAS DE BUSCA DE TEs NOS GENOMAS

1.3.1 Busca por homologia de sequências (estratégias *ab initio*)

Na busca por homologia, um conjunto de sequências (p. ex. um genoma) é comparado a uma biblioteca de referência contendo sequências de elementos móveis. Essa biblioteca pode ser construída pelo usuário, ou pode ser obtida em bancos de sequências repetitivas como o Repbase (BAO; KOJIMA; KOHANY, 2015). Esse método não é capaz de detectar novos TEs, pois utiliza sequências previamente descritas, mas pode ser usado em combinação com abordagens *de novo* (LERAT, 2009).

1.3.2 Busca por características estruturais (estratégias *de novo*)

Na busca por características estruturais, isto é, em abordagens *de novo*, os programas procuram por estruturas particulares de um determinado TE, como padrões de repetição. Esse método pode ser empregado na identificação de novos elementos, porém não na identificação de novas classes, pois depende do conhecimento prévio dos padrões estruturais dos TEs (LERAT, 2009).

2. OBJETIVOS

Considerando que não há consenso sobre a acumulação de elementos móveis e seus tipos em organismos assexuados e sexuados, esse trabalho objetiva investigar o conteúdo de elementos móveis utilizando duas espécies proximamente relacionadas de microsporídios que diferem no modo de reprodução como modelo.

3. ARTIGO CIENTÍFICO

Introduction

Microsporidia are unicellular, obligate intracellular parasites with highly reduced genomes belonging to the kingdom of fungi. There are about 1,500 species and 200 genera currently described that infect a wide range of animal species, including humans, especially immunodepressed patients (Becnel et al., 2014). Microsporidia have the smallest genomes among eukaryotes, possibly due to their parasitic mode of life, which resulted in an extreme physiologic simplification. In spite of that, some microsporidian species have large portions of non-coding regions, mobile elements and repetitive DNA sequences in their genomes (Corradi et al., 2009).

Transposable elements (TEs) are selfish DNA sequences that can replicate and integrate themselves to new locations in the genome. It can cause mutations and chromosomal rearrangements and generally induce negative fitness effects in their hosts. TEs are abundant in eukaryotic genomes, comprising about 3-20% of fungal genomes (Daboussi & Capy, 2003). They are divided in two classes: class I includes the retrotransposons, which use a "copy-paste" transposition mechanism with a RNA intermediate reverse-transcribed into complementary DNA (cDNA) that integrates into another locus in the host genome; class II elements are DNA transposons that use a "cut-paste" transposition mechanism with DNA intermediate inserting directly into another locus. Retrotransposons are the most abundant TEs in the eukaryotic genome (Feschotte & Pritham, 2007). They are subdivided in non-LTR retrotransposons and LTR-retrotransposons The long terminal repeats (LTRs) are non-coding sequence repeats flaking the coding region. The coding region contains two open reading frames (ORFs), one for Gag and another for Pol proteins. The Pol polyprotein contains four key protein domains: aspartic protease (AP), RNAse H (RH), reverse transcriptase (RT) and integrase (INT) (Fig. 1). The Ty1/copia and Ty3/gypsy lineages are most abundant families of LTR retrotransposons found in a wide range of eukaryotic organisms. Ty1/copia represents the most ancient lineage, while Ty3/gypsy is the most abundant in eukaryotic genomes and has a closer relationship to retroviruses (Eickbush & Malik, 2002).

Theoretical models and experimental studies propose that spread of TEs depends on sexual reproduction (Hickey, 1982; Arkhipova & Meselson, 2000), on the other hand the absence of recombination could drive asexual population to extinction by an unchecked proliferation of TEs (Dolgin & Charlesworth, 2006). There are previous studies assessing TE loads between

sexual and asexual reproducing organisms, however this comparison is difficult when it comes to microsporidia due to the lack of evidence on their reproduction modes. *Hamiltosporidium tvaerminnensis* and *H. magnivora* are two closely related microsporidian species that differ in their mode of reproduction and transmission, e.g., the former is asexual and transmitted both horizontally and vertically, whereas the latter is sexual and only vertically transmitted (Haag et al., 2013). Here, we propose to investigate the TE load in the genomes of one *H. tvaerminnensis* lineage and two lineages of *H. magnivora*.

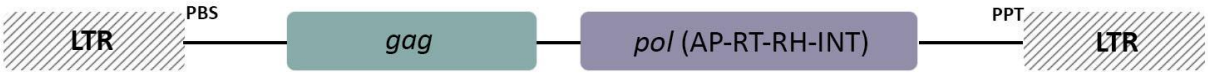

Fig. 1 - Genomic organization of a Ty3/gypsy LTR retrotransposon. PBS: Primer Binding Site; PPT: Polypurine Tract.

Material and Methods

Data source

Spores were isolated from infected specimens of *Daphnia magna* maintained in culture and used for DNA purification and sequencing. The genomic data of three parasite lineages, two of *H. magnivora* (BE-OM-2 from Belgium, and IL-BN-2 from Israel), and one *H. tvaerminnensis* (FI-OER-3-3 from Finland) were obtained by pair-end sequencing using Illumina HiSeq 2000 at University of Basel, Switzerland. The genomes were assembled by Juliano de Oliveira Silveira as part his master's thesis at Federal University of Rio Grande do Sul, Brazil (Silveira, 2017).

Some attributes of the assembled genomes are listed in Table 1.

Table 1 - Genome assembly features.

	BE-OM-2	IL-BN-2	FI-OER-33
Genome size	13.9	13.5	9.8
No. of contigs	746	900	1,171
Contig N50	28,476	22,994	17,299
Contig L50	161	187	145
Genome coverage	55.6	54	39.2
G+C content	26.22	26.67	26.15
Genes predicted	2,224	2,248	2,075

Initial screening of TEs

The genomes of BE-OM-2, IL-BN-2 and FI-OER-3-3 were searched for the presence of repetitive elements using the RepeatMasker software (Smit et al., 2013-2015). RepeatMasker screens genomic sequences and annotates transposable elements, tandem repeats and low complexity DNA sequences by using BLAST as search engine (E-value threshold of 1e-10). Representative transposable elements from fungi obtained from Repbase (Bao et al., 2015) were used as library. Low complexity DNA sequences and simple repeats were masked. The tool "One code to find them all" was used to summarize the number of TE copies and genome coverage for each TE family (Bailly-Bechet et al., 2014).

Domain detection

The predicted proteomes of BE-OM-2, IL-BN-2 and FI-OER-3-3 were searched using BLASTp (E-value threshold of 1e-10 and BLOSUM62 substitution matrix) against 2,636 core elements of proteins encoded by LTR retrotransposons from the Gypsy Database 2.0 (www.gydb.org) (Llorens et al., 2011).

LTR retrotransposon detection

Potential full-length LTR retrotransposons in the *Hamiltosporidium* genomes were detected using the LTRharvest software (Ellinghaus et al., 2008). This tool searches for maximal exact repeats, LTR candidate pairs with target site duplications (TSDs), and palindromic LTR motifs. Elements containing putative Pol ORFs with at least three key protein domains

(reverse transcriptase, RNAseH and integrase) were considered potential full-length LTR retrotransposons.

Phylogeny reconstruction

Protein sequences similar to domains encoded by the *Hamiltosporidium* full-length LTR retrotransposons were selected using BLASTp (E-value threshold of 1e-10) against non-redundant protein sequences database from GenBank. The selected sequences were aligned with MUSCLE version 3.8 (Edgar, 2004) and uninformative columns in the alignments (containing at least 30% of gaps) were trimmed using Geneious version 8.1.7 (Kearse et al., 2012). The best amino acid substitution models fitting our data were searched with MEGA 7 (Kumar et al., 2016).

Phylogenetic analyses were performed using maximum-likelihood with PHYML version 3.0 (Guindon et al., 2010), and confidence values of the internal tree branches were estimated using 100 bootstrap replicates. A phylogeny for the "Bigshot" polyprotein aligned to Pol polyproteins from *Hepatospora eriocheir* GB1 (ORD95944.1), *Enterospora canceri* GB1 (ORD93876.1 and ORD93650.1), *Photinus pyralis* (JAV95998.1), *Daphnia magna* (JAN94511.1 and JAN70235.1), *Triatoma infestans* (JAC14078.1 and JAC14022.1), *Nosema apis* BRL 01 (EQB60916.1), *Nosema bombycis* (ABE26651.1), *Oikopleura dioica* (ATT48675.1) was obtained with the rtREV+G+I+F model. The model selected for the *Hamiltosporidium* "Nanomov" polyprotein aligned to Pol polyproteins from *Pseudoloma neurophilia* (KRH93018.1, KRH93020.1, KRH94229.1, KRH92528.1, KRH92867.1, KRH92640.1, KRH93149.1), *Megachile rotundata* (XP_012147359.1), *Aegilops tauschii* (XP_020156956.1), *Capsaspora owczarzaki* (XP_011270928.1), *Umbilicaria pustulata* (SLM40345.1), *Papilio machaon* (XP_014356853.1), *Maylandia zebra* (XP_014265766.1), *Hepatospora eriocheir* (ORD95944.1 and ORE00292.1), *Enterospora canceri* (ORD93650.1 and ORD93876.1), *Nosema apis* BRL 01 (EQB60916.1) and *Nosema bombycis* (ABE26651.1) was rtREV+G.

Results

Detection and annotation of TEs

For the initial screening of TE content in the genomes of the three *Hamiltosporidium* lineages BE-OM-2 and IL-BN-2 (*H. magnivora*), and FI-OER-33 (*H. tvaerminnensis*) we used

RepeatMasker, a homology-based method. We only considered fragments with >500 bp and >80% similarity to reference sequences. RepeatMasker identified a total of 34 fragmental TEs in BE-OM-2 genome, 196 in IL-BN-2 genome and 20 in FI-OER-33 genome (Table 2). The *H. magnivora* IL-BN-2 lineage shows an increased TE load compared with BE-OM-2 lineage and the *H. tvaerminnensis* FI-OER-33 lineage. Retrotransposons are the most abundant TE class in the IL-BN-2 genome, especially the LTR-retrotransposon from the *gypsy* group. BE-OM-2 and FI-OER-33 lineages presented an increased number of DNA transposons compared with retrotransposons, BE-OM-2 still showed 12 LTR-retrotransposons while FI-OER-33 showed 9 elements from *Penelope* subclass of retroelements.

Table 2 - Summary of TE fragments identified with RepeatMasker.

Class	**Subclass**	**Clades**	**No. of TEs**		
			BE-OM-2	**IL-BN-2**	**FI-OER-3-3**
Retrotransposon	LTR	Gypsy	5	170	0
		Copia	7	2	0
	Penelope	Coprina	0	8	9
Subtotal			*12*	*180*	*9*
DNA transposons	TIR	Mariner	10	0	0
	Helitrons	Helitron	12	16	11
Subtotal			*22*	*16*	*11*
Total			**34**	**196**	**20**

NOTES – Only fragments with >500 bp and >80% similarity to reference sequences.

Identification of potentially active LTR retrotransposons

The putative full-length elements were searched using combined methods. LTRharvest is a *de novo* structure-based method and the protein domain detection is a homology-based method using BLASTp as search tool. Our findings are summarized in Table 3. The sexual lineages showed a similar content of LTR retrotransposons detected by LTRharvest, while the asexual lineage showed a decreased number of those elements. With regard to the protein domains we observe an increased number of retrotransposon core proteins in the IL-BN-2 proteome compared to BE-OM-2, in spite of their similar number of LTR retrotransposons as detected

by LTRharvest. The asexual lineage FI-OER-3-3 shows a decreased number of retrotransposon protein domains. Surprisingly, the number of potentially active LTR retrotransposons was the same (=2) for BE-OM-2, IL-BN-2 and FI-OER-3-3. Both elements are shared between all three lineages, and belong to the Ty3/gypsy clade. One Ty3/gypsy element, named "Nanomov", contains 3,962 bp, 98.8% identical sites and 99.2% of pairwise identity among the three genomes (Fig. 2). In IL-BN-2 and FI-OER-3-3 the "Nanomov" RT, RH and INT domains are encoded as one single polyprotein, while in BE-OM-2 the element has two ORFs, one encoding the INT domain and another encoding the RT and RH domains. The second retrotransposon from the Ty3/gypsy family named "Bigshot", contains 4,043 bp, 98.9% identical sites and 99.3% of pairwise identity among the three genomes (Fig. 3). In "Bigshot" RT and INT domains are encoded by separate ORFs, and although RH is still identified by structural motifs of a RNAse H domain it's not encoded by a ORF. The RT domain in the BE-OM-2 genome has an in frame stop codon shortening the amino acid sequence and probably generating a truncated protein (Fig. 4).

Table 3 - Summary of predicted full-length LTR retrotransposons and ORFs encoding key protein domains.

Lineage	No of LTR retrotransposons[a]	No of predicted proteins			
		AP	RH	RT	INT
BE-OM-2	2 (40)	0	3	9	8
IL-BN-2	2 (41)	0	30	47	54
FI-OER-33	2 (23)	0	1	3	4

AP - aspartic protease; RH - RNase H; RT - reverse transcriptase; INT – integrase.

NOTES - The potentially active LTR retrotransposons are the elements with pair of LTRs 5' and 3', and at least three key protein domains for its transposition.

[a]The numbers of potentially active LTR retrotransposons (total LTR retrotransposons detected by LTRharvest are listed within brackets).

Fig. 2 – The potentially active Ty3/gypsy element "Nanomov", with 3,962 bp, shared between all three lineages.

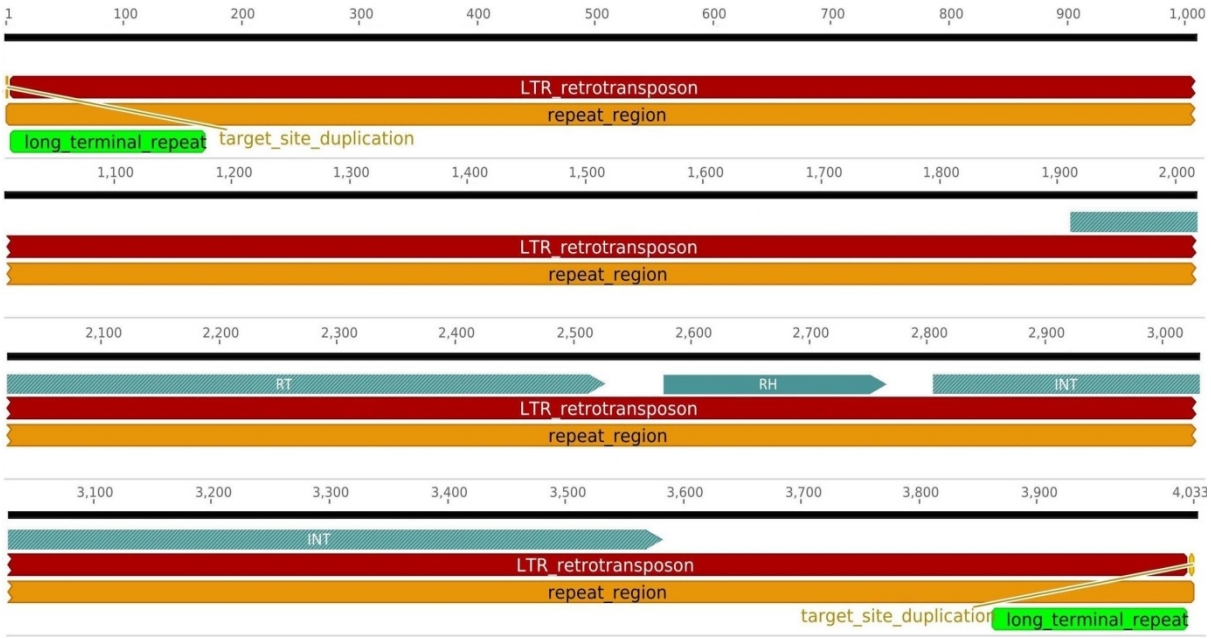

Fig. 3 – The potentially active Ty3/gypsy element "Bigshot", with 4,043 bp, shared between all three lineages.

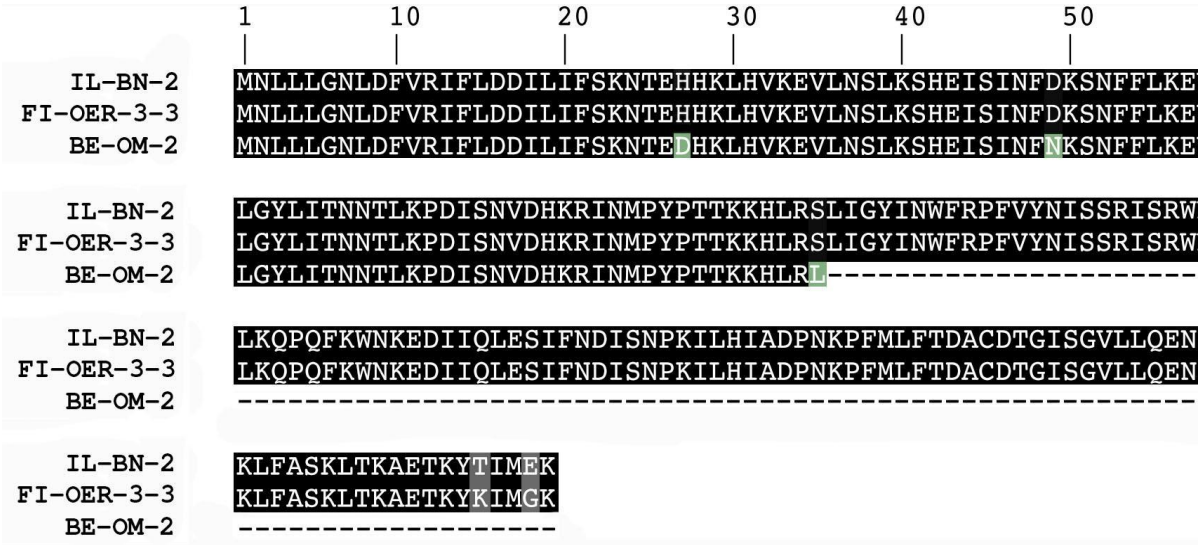

Fig. 4 – RT domain amino acid alignment from "Bigshot" Ty3/gypsy element.

Phylogenetic tree reconstruction

To reconstruct the evolutionary history of our full-legth retroelements we performed maximum likelihood analysis on the amino acid sequences. The "Bigshot" retrotransposon shows high similarity to *gypsy* elements present in other microsporidia, such as *Nosema* (Fig. 5) and is also related to retroelements found in arthropods such as *D. magna*, the only host of

Hamiltosporidium. "Nanomov" groups with other microsporidian retroelements, however with low statistical support (Fig. 6). Similar amino acid sequences belong to different and divergent species such as plants, fungi, arthropods and fish.

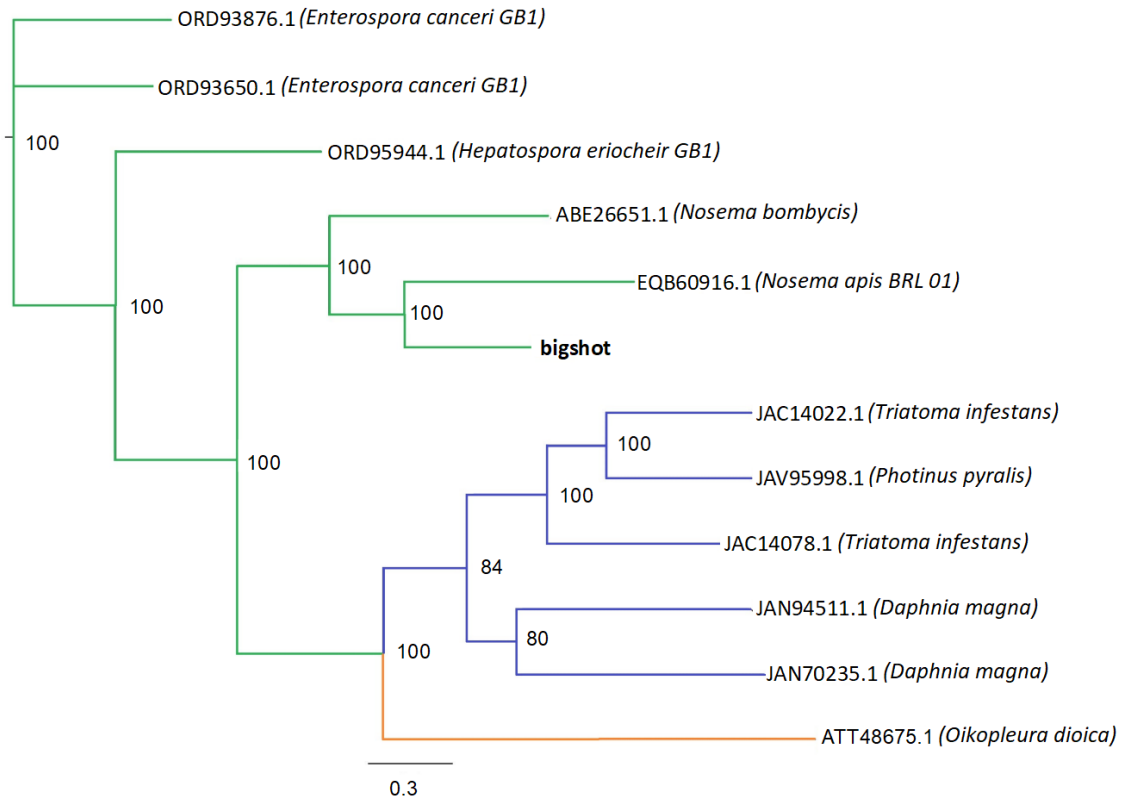

Fig. 5 – Phylogenetic tree of "Bigshot" Ty3/gypsy element. Tree topology obtained using maximum-likelihood analysis on the "Bigshot" polyprotein.

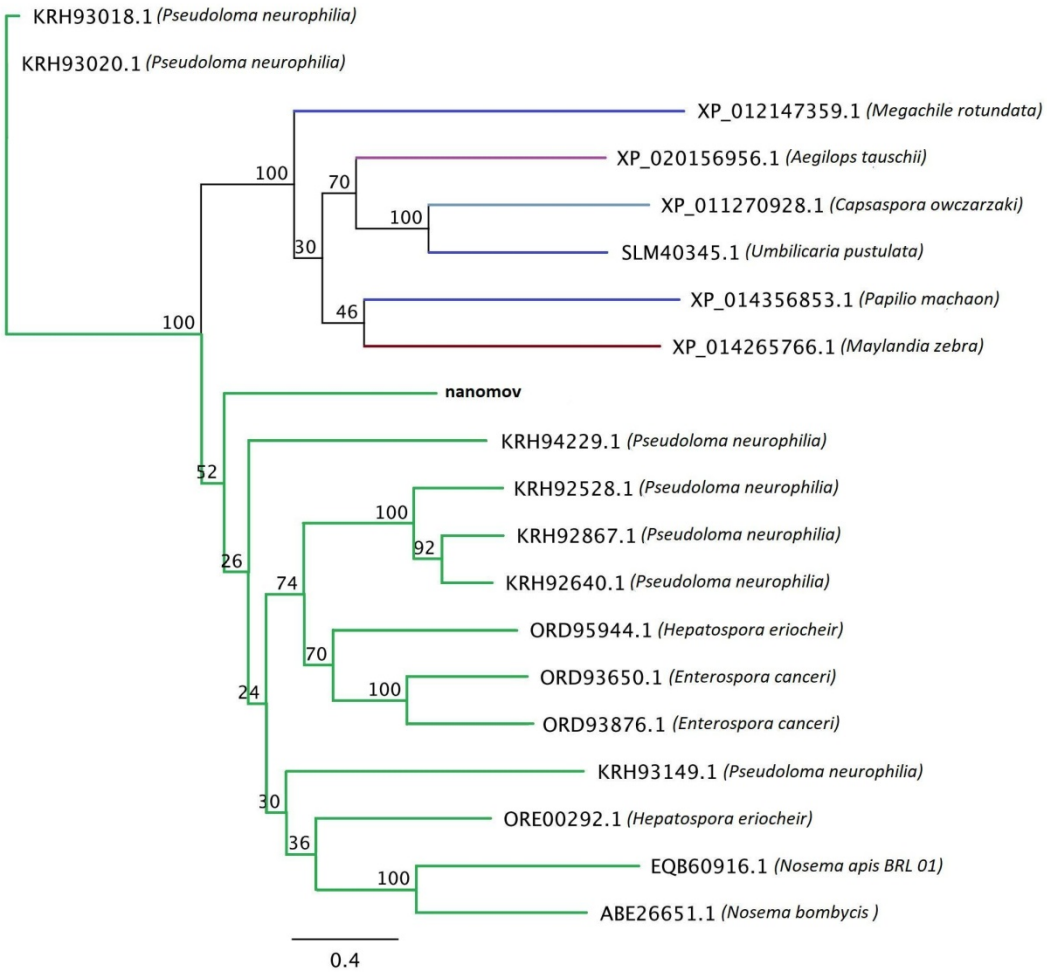

Fig. 6 – Phylogenetic tree of "Nanomov" Ty3/gypsy element. Tree topology obtained using maximum-likelihood analysis on the "Nanomov" polyprotein.

Discussion

LTR retrotransposons in microsporidia genomes

LTR retrotransposons from the Ty3/gypsy family were previously described in microsporidia, with subgroups exclusive to microsporidia species. The Ty3/gypsy elements from these subgroups have a single polyprotein instead of two separate ORFs for gag and pol (Parisot et al., 2014). We identified two potentially active LTR retrotransposons from Ty3/gypsy group, "Nanomov" and "Bigshot". "Nanomov" Ty3/gypsy element has one polyprotein encoding RT, RH and INT domains. Except in BE-OM-2 the INT domain is encoded by a different ORF than RT and RH domains. In the "Bigshot" Ty3/gypsy element there are two ORFs for

RT and INT domains. The RH domain is not encoded by an ORF. In BE-OM-2 genome, the "Bigshot" RT domain has a premature stop codon and probably isn't functional. However, since all genes annotated in the neighborhood of "Bigshot" are identical in our genomes, it is reasonable to assume that they represent a single insertion event in the ancestor of the three *Hamiltosporidium* lineages. As to "Nanomov", the contigs on which it is inserted is exactly its size so we do not know the adjacent genes. In LTR retrotransposons found in fungi, gag and AP (protease) domain are the most variable domains, and the conserved RT, RH and INT domains seem to be inherited together (Muszewska et al., 2011). One possible reason for the absence of these proteins in our analysis is the lack of gag and AP variants available in sequence databases to detect these proteins.

TE content between asexual/sexual lineages

Our results showed significant difference in the TE content between asexual/sexual lineages, with excess TE accumulation in the sexual lineages, particularly in the IL-BN-2 genome. This finding is in accordance with the model proposed by Hickey that the spread of transposable genetic elements depends on sexual reproduction (Hickey, 1982). This result is consistent with previous studies with the long-term asexual bdelloid rotifers in comparison to representatives of a majority of primarily sexual animal phyla, which found a low copy number of retrotransposons in two bdelloid species, *P. roseola* and *A. vaga* in contrast to their sexual relatives. The reason for the low copy of retrotransposons in these bdelloids may be the recent arrival in their present hosts. Furthermore, bdelloid rotifers can repair high levels of DNA breakage, and such repair may favor ectopic crossing-over of TEs, with selection then acting against unviable rearrangements. Ectopic crossing-over could also account for the preferential telomeric localization of these elements and of bdelloid DNA transposons, as rearrangements that occur in gene-poor regions would not be deleterious (Arkhipova & Meselson, 2000, 2005; Gladyshev et al., 2007). In the strictly asexual plant pathogenic fungus *Verticillium dahliae* it was found highly dynamic lineage-specific genomic regions enriched with retrotransposons and DNA transposons. TEs and their activity contribute to genomic diversity by inducing large-scale genomic duplications (De Jonge et al., 2013; Faino et al., 2016). The outcrossing plant *Arabidopsis lyrata* has a higher number of TEs and younger insertions compared to its close selfing relative *Arabidopsis thaliana*, in accordance with the recombinational spreading hypothesis (de la Chaux et al, 2012). The comparison between five independently derived asexual arthropod lineages and their sexual relatives found no support

for the prediction that asexuality results in the accumulation of TEs. (Bast et al., 2016). Although our results seem consistent with the literature, they are not yet final. First, because our assemblies are still in a draft format and need to be completed to allow for a more reliable assessment of TE numbers. Second, our search for TEs was not exhaustive, and mainly focused on LTR retrotransposns. It would be necessary to use other tools to expand the search for new transposable elements and better understand their dynamics in these genomes.

Phylogenetic studies with other microsporidia indicated that the asexual *Vairimorpha* species have evolved recently from sexual ancestors and the loss of sex has occurred on multiple occasions (Ironside, 2007). Phylogeographical studies with *Hamiltosporidium* suggests the recurrent origin of asexual forms (Haag et al., 2013). We identified in our asexual lineage (FI-OER-3-3) proteins related to DNA methylation and RNA interference (RNAi) indicating that with loss of sex the organism does not lose these mechanisms responsible for the suppression of TE spread (Kraaijeveld et al., 2012). Also, TE spread within a population would be restricted to rare horizontal transfers, and selection acting on the host should favor nonfunctional elements. In the absence of any counter-selection at the element level, transposable elements are eventually expected to degenerate over evolutionary time (Wright & Finnegan, 2001).

Evolutionary history of TEs in microsporidia

Our phylogenetic analysis shows that Ty3/gypsy elements homologous to the *Hamiltosporidium* “Bigshot” are distributed in the genomes of different species of microsporidia as well as arthropods. Interestingly a similar Ty3/gypsy element is present in *D. magna*, suggesting horizontal gene transfer between parasite and host at some point of their evolution. The exchange of TEs between microsporidia and various metazoan taxa seems to be more frequent that previously appreciated (Parisot et al. 2014). It was not possible to infer the origin of “Nanomov” with confidence due to high amino acid sequence divergence that led to a low statistical support of its clade in our phylogeny. We failed to find Ty3/gypsy elements with enough similarity to “Nanomov” annotated in sequence databases, suggesting either that they might have been missed during the annotation process of other genomes or a higher rate of evolution of this element in the *Hamiltosporidium* genus.

In conclusion, we showed that LTR retrotransposons are numerous in comparatively large microsporidian genomes such as *Hamiltosporidium*, and could impact in genome size variation through evolution in the presence of sexual reproduction. To better understand the

role of transposable elements in reduced genomes and the impact of different TE loads in sexuals and asexuals, we need further investigation on these retroelements, identifying other superfamilies and include DNA transposons in the analysis. The correct identification and annotation of TEs is the first step to understand the contribution of these selfish elements in genome evolution.

Acknowledgments

We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

Arkhipova I, Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. PNAS. 97(26), 14473–14477.

Arkhipova I, Meselson M. 2005. Deleterious transposable elements and the extinction of asexuals. BioEssays. 27(1), 76–85.

Bailly-Bechet M, Haudry A, Lerat E. 2014. “One code to find them all”: a perl tool to conveniently parse RepeatMasker output files. Mobile DNA. 5(1), 13.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, 6(1), 11.

Bast J, et al. 2016. No accumulation of transposable elements in asexual arthropods. Molecular Biology and Evolution. 33(3), 697–706.

Becnel JJ, Takvorian PM. 2014. Checklist of Available Generic Names for Microsporidia with Type Species and Type Hosts. Microsporidia: Pathogens of Opportunity: First Edition. 671–686.

Corradi N, Haag KL, Pombert JF, Ebert D, Keeling PJ. 2009. Draft genome sequence of the Daphnia pathogen Octosporea bayeri: insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. Genome Biology. 10(10), R106.

Daboussi MJ, Capy P. 2003. Transposable Elements in Filamentous Fungi. Annual Review of Microbiology. 57(1), 275–299.

de Jonge R, et al. 2013. Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Research. 23(8), 1271-1282.

de la Chaux N, et al. 2012. The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. Mobile DNA. 3(1), 2.

Dolgin ES, Charlesworth B. 2006. The fate of transposable elements in asexual populations. Genetics. 174(2), 817–827.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5), 1792–1797.

Eickbush TH, Malik HS. 2002. Origins and Evolution of Retrotransposons. Mobile DNA II. In: Craig NL, et al. (Edt.). Mobile DNA. ASM Press. 1111-1144.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 9(1), 18.

Faino L, et al. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. Genome Research. 26(8), 1091-1100.

Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. Annu Rev Genet. 41: 331–368.

Gladyshev EA, Meselson M, Arkhipova IR. 2007. A deep-branching clade of retrovirus-like retrotransposons in bdelloid rotifers. Gene. 390:136–145.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 59(3), 307–321.

Haag KL, Sheikh-Jabbari E, Ben-Ami F, Ebert D. 2013. Microsatellite and single-nucleotide polymorphisms indicate recurrent transitions to asexuality in a microsporidian parasite. Journal of Evolutionary Biology. 26(5), 1117–1128.

Hickey DA. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. Genetics. 101:519-531.

Ironside JE. 2007. Multiple losses of sex within a single genus of Microsporidia. BMC Evolutionary Biology. 7, 48.

Kearse M, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12), 1647–1649.

Kraaijeveld K, et al. 2012. Transposon proliferation in an asexual parasitoid. Molecular Ecology. 21(16), 3898–3906.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 33(7), 1870–1874.

Llorens C, et al. 2011. The Gypsy Database (GyDB) of Mobile Genetic Elements: Release 2.0. Nucleic Acids Research. 39:70–74.

Muszewska A, Hoffman-Sommer M, Grynberg M. 2011. LTR retrotransposons in fungi. PLoS ONE. 6(12).

Parisot N, et al. 2014. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. Genome Biology and Evolution. 6(9), 2289–2300.

Silveira JO. 2017. Elementos genômicos de reprodução sexuada em microsporídeos. (Master's thesis).

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>.

Wright S, Finnegan D. 2001. Genome evolution: Sex and the transposable element. Current Biology. 11:296–299.

4. CONCLUSÃO

Os resultados desse estudo demonstraram diferenças no conteúdo de TEs entre as linhagens sexuadas e assexuadas de *Hamiltosporidium*. As linhagens sexuadas de *H. magnivora* possuem uma quantidade superior de retrotransposons LTR em comparação com a linhagem assexuada de *H. tvarminnensis*, mostrando que a reprodução sexuada facilita a disseminação de elementos transponíveis no genoma. Houve diferenças também entre as linhagens sexuadas BE-OM-2 e IL-BN-2, esta última com número maior de domínios-chave de proteínas de retrotransposons LTR em comparação com BE-OM-2. Os elementos presentes em BE-OM-2 apresentam mais mutações que as outras linhagens, sugerindo um resultado mais eficiente da seleção natural em prevenir a acumulação de elementos potencialmente deletérios nesta linhagem. É necessário uma busca mais ampla por diferentes famílias e classes de elementos para entender a dinâmica de elementos transponíveis nos genomas de microsporídios.

Quanto aos elementos potencialmente ativos, foram encontrados dois retrotransposons LTR da família Ty3/gypsy que são compartilhados pelas três linhagens, apresentando três dos principais domínios de proteínas presentes na região codificante (RT, RH e INT). Esse achado indica que a inserção desses elementos ocorreu antes da divergência das espécies *H. magnivora* e *H. tvaerminnensis*. Esses elementos foram denominados “Bigshot” e “Nanomov”. O elemento “Bigshot” presente no genoma de BE-OM-2 possui uma sequência de RT com um códon de parada prematuro, provavelmente gerando uma proteína truncada.

Para conhecer a origem desses elementos, foram construídas árvores filogenéticas utilizando sequências de proteínas. O elemento “Bigshot” possui similaridade com outros retrotransposons LTR presentes no genoma de outros microsporídios. Também está relacionado com elementos encontrados em artrópodes como *D. magna*, indicando uma possível transferência gênica horizontal entre microsporídios e seus hospedeiros artrópodes ancestrais. Para o elemento “Nanomov” não foi possível determinar relações filogenéticas confiáveis. Mesmo com a procura de sequências homólogas em banco de dados de sequências de proteínas, esse retrotransposon LTR mostrou-se muito divergente. Sua sequência foi agrupada com outros microsporídios e, também foi relacionado com elementos de organismos eucarióticos variados como plantas, fungos, artrópodes e peixes; porém com um baixo suporte estatístico. A falta de genes e proteínas pertencentes a elementos transponíveis similares ao elemento "Nanomov" indica que a busca por esses elementos em genomas depositados em

bancos de dados não é exaustiva. A correta identificação e anotação desses elementos é essencial para o estudo e compreensão do impacto de elementos transponíveis nos genomas e na evolução dos organismos.

5. REFERÊNCIAS

ARKHIPOVA, Irina R.; MESELSON, Matthew. Transposable elements in sexual and ancient asexual taxa. **PNAS**, v. 97, n. 26, p. 14473-14477, dec. 2000.

BAO, Weidong; KOJIMA, Kenji K.; KOHANY, Oleksiy. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, n. 11, p. 1-6, jun. 2015.

BAST, Jens et al. No accumulation of transposable elements in asexual arthropods. **Molecular Biology and Evolution**, v. 33, n. 3, p. 697-705, dec. 2015.

BECNEL, James J.; TAKVORIAN, Peter M.; CALI, Ann. Checklist of available generic names for microsporidia with type species and type hosts. In: WEISS, Louis M.; BECNEL, James J. (Edt.). **Microsporidia: Pathogens of Opportunity**. John Wiley & Sons; 2014. pp. 671–686.

CORNMAN, Scott R. et al. Genomic analyses of the microsporidian Nosema ceranae, an emergent pathogen of honey bees. **PLoS Pathogens**, v. 5, n. 6, p. 1-14, jun. 2009.

DE LA CHAUX, Nicole et al. The predominantly selfing plant Arabidopsis thaliana experienced a recent reduction in transposable element abundance compared to its outcrossing relative Arabidopsis lyrata. **Mobile DNA**, v. 3, n. 2, p. 1-18, feb. 2012.

DOLGIN, Elie S.; CHARLESWORTH, Brian. The Fate of Transposable Elements in Asexual Populations. **Genetics**, v. 174, n. 2, p. 817-827, jul. 2006.

EBERT, Dieter. A Genome for the Environment. **Science**, v. 331, n. 6017, p. 539–540, feb. 2011.

EICKBUSH, Thomas H.; MALIK, Harmit S. Origins and Evolution of Retrotransposons. In: CRAIG, N. L. et al. (Edt.). **Mobile DNA**. ASM Press; 2002. pp. 1111-1144.

FESCHOTTE, Cédric; PRITHAM, Ellen J. DNA Transposons and the Evolution of Eukaryotic Genomes. **Annual Review of Genetics**, v. 41, n. 1, p. 331-368, dec. 2007.

GLADYSHEV, Eugene A.; MESELSON, Matthew; ARKHIPOVA, Irina R. A deep-branching clade of retrovirus-like retrotransposons in bdelloid rotifers. **Gene**, v. 390, n. 1-2, p. 136-145, april 2007.

HICKEY, Donal A. Selfish DNA: a sexually-transmitted nuclear parasite. **Genetics**, v. 101, n. 3-4, p. 519-531, aug. 1982.

KRAAIJEVELD, Ken et al. Transposon proliferation in an asexual parasitoid. **Molecular Ecology**, v. 21, n. 16, p. 3898-3906, aug. 2012.

LERAT, Emmanuelle. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. **Heredity**, v. 104, n.6, p. 520-533, nov. 2009.

PARISOT, Nicolas et al. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. **Genome Biology and Evolution**, v. 6, n. 9, p. 2289-2300, aug. 2014.

THOMARAT, Fabienne; VIVARÈS, Christian P.; GOUY, Manolo. Phylogenetic analysis of the complete genome sequence of Encephalitozoon cuniculi supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. **Journal of Molecular Evolution**, 2004. v. 59, n. 6, p. 780–791.

WILLIAMS, Bryony A. P. et al. Genome sequence surveys of Brachiola algerae and Edhazardia aedis reveal microsporidia with low gene densities. **BMC Genomics**, v. 9:200, p. 1-9, april 2008.

XU, Jinshan et al. The varying microsporidian genome: existence of long-terminal repeat retrotransposon in domesticated silkworm parasite Nosema bombycis. **International Journal for Parasitology**, v. 36, n. 9, p. 1049-1056, dec. 2006.

ANEXO A – NORMAS DE PUBLICAÇÃO DA REVISTA GENOME BIOLOGY AND EVOLUTION

Instructions to Authors

Upon receipt of accepted manuscripts at Oxford Journals authors will be invited to complete an online copyright license to publish form. Please note that by submitting an article for publication you confirm that you are the corresponding/submitting author and that Oxford University Press ("OUP") may retain your email address for the purpose of communicating with you about the article. You agree to notify OUP immediately if your details change. If your article is accepted for publication OUP will contact you using the email address you have used in the registration process. Please note that OUP does not retain copies of rejected articles.

EMBARGOED DATA

Papers that are accepted at *GBE* are published on Advance Access within 48 hours of acceptance. The data underlying papers submitted to *GBE* must also be made publically accessible at that time. By submitting a paper to *GBE*, authors acknowledge that the data underlying the paper can be released at the time of acceptance. *GBE* follows the Fort Lauderdale Agreement on fair use of community resource data.

SCOPE

Genome Biology and Evolution (*GBE*) publishes leading original research at the interface between evolutionary biology and genomics. Papers considered for publication report novel evolutionary findings that concern natural genome diversity, population genomics, the structure, function, organisation and expression of genomes, comparative genomics, proteomics, and environmental genomic interactions. Major evolutionary insights from the fields of computational biology, structural biology, developmental biology, and cell biology are also considered, as are theoretical advances in the field of genome evolution. *GBE*'s scope embraces genome-wide evolutionary investigations at all taxonomic levels and for all forms of life—within populations or across domains. Its aims are to further the understanding of genomes in their evolutionary context and further the understanding of evolution from a genome-wide perspective.

PUBLICATION CHARGES AND OPEN ACCESS

GBE is an open access journal and depends upon open access publication fees for its operation. For every accepted paper, there will be an open access fee of US$1800 / £1125 / €1463 (plus VAT or other taxes applicable in your country). Perspectives will be charged at $1250 / £781 / €1016. *GBE* does not apply any additional charges, for example, color or excess pages. All accepted papers are freely available to everybody online, upon publication, without the barrier of paid subscription to access them. Oxford Open articles are published under Creative Commons licenses. Authors publishing in *GBE* can use the Creative Commons Attribution licence (CC BY) or the Creative Commons Non-Commercial license for their articles. Please click here for more information about the Creative Commons licenses.

You can pay Open Access charges using our Author Services site. This will enable you to pay online with a credit/debit card, or request an invoice by email or post. Orders from the UK will be subject to the current UK VAT charge. For orders from the rest of the European Union, OUP will assume that the service is provided for business purposes. Please provide a VAT number for yourself or your institution, and ensure you account for your own local VAT correctly.

If you are working in a developing country visit our Developing Countries page for a list of qualifying countries

- Reduced Rate Developing country charge* - £563 / $900 / €732
- Free Developing country charge* - £0 /$0 / €0

If you are working in a developing country and are unable to pay the open access fee, please contact the Editor-in-Chief.

ONLINE SUBMISSION OF MANUSCRIPTS

Manuscripts are submitted electronically. To submit your manuscript to *GBE*, please visit the *GBE* submission portal and follow the instructionsto log in.

Submission of a paper implies that it reports unpublished work and that it is not under consideration for publication elsewhere. If previously published tables, illustrations or more than 200 words of text are included in the submission, then the copyright holder's written permission must be obtained and copies of any such permission letters must be enclosed with the paper. To contact the editorial office, please send an email.

LATEX SUBMISSIONS

While you should prepare your typescript text using a word-processing package (save in .doc or .rtf format), PDF files may be acceptable if you have used LaTeX to write your manuscript. Typescripts should be double-spaced. Please number each page.

LaTeX: When creating the pdf file from standard word processing packages, it is important to select the option "embed all fonts" (or equivalent) to ensure that mathematical symbols appear correctly. Failure to do so may result in an unreadable submission. Instructions for generating correct PDF files can be found at https://gemini.econ.umd.edu/e-editor/pdftips.html.

Manuscripts must be prepared using the TeX class file (available here as a .zip file, accompanied by notes on usage). Once a Tex file is created, it can be submitted to the editorial office.

Please see our LaTeX submissions FAQ page for more detailed instructions for LaTeX submissions.

REVIEW OF MANUSCRIPTS

Manuscripts that are within the scope of the journal are peer reviewed. Those not meeting the journal's scientific standards and those outside the journal's scope are declined and returned to the authors. *GBE* has a strong commitment to the rapid handling of submissions. Authors can expect a first decision on their submission within four to six weeks.

REVISED MANUSCRIPT

Revised manuscripts should be submitted within two months of the author's receipt of the referees' reports. Revised manuscripts returned after two months will be considered as new submissions and may be subject to re-review. Authors should notify the editorial office as soon as possible if revisions will require longer than two months.

AUTHORSHIP

All persons designated as authors should qualify for authorship. The order of authorship should be a joint decision of the co-authors. Each author should have participated sufficiently in the work to take public responsibility for the content. Authorship credit should be based on substantial contribution to conception and design, execution, or analysis and

interpretation of data. All authors should be involved in drafting the article or revising it critically for important intellectual content. All authors must have read and approved the final version of the manuscript, assurance of which should be given in the covering letter.

CONFLICT OF INTEREST

At the stage of online submission, *GBE*'s policy requires that each author reveal any financial interests or connections, direct or indirect, or other situations that might raise the question of bias in the work reported or the conclusions, implications, or opinions stated - including pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organisation(s), personal relationships, or direct academic competition. When considering whether you should declare a conflicting interest or connection please consider the conflict of interest test: Is there any arrangement that would embarrass you or any of your co-authors if it was to emerge after publication and you had not declared it? It is the Corresponding author's responsibility to ensure that all authors adhere to this policy. If the manuscript is published, Conflict of Interest information will be communicated in a statement in the published paper.

PROOFS

For accepted manuscripts, authors are sent page proofs by the publisher. Please provide an email address to enable page proofs to be sent as PDF files via email. To avoid delays in publication, proofs should be checked immediately for typographic errors and returned to the production office by fax or email within 2 working days. Essential changes of an extensive nature may be made only by insertion of a Note Added in Proof. Page charges are not levied. Authors are, however, charged for extensive changes made in proof and for special items such as color plates.

TYPES OF MANUSCRIPTS

GBE considers the following kinds of manuscripts:

- Research Articles
- Letters
- Genome Reports
- Perspectives
- Reviews

- Genome Resources
- **Research Articles** should be divided into the following sequence of headed sections: Abstract, Introduction, Material and Methods, Results, Discussion, Acknowledgments, and References. The total length of the main text (excluding references) should not exceed 10,000 words. Up to 10 items (tables or figures) may be included. Limit the number of references to 100, preferably fewer. Include supplementary material only where necessary and make sure that the supplementary material is supplied in a suitable format that can be readily accessed by all interested parties.
- **Letters** should begin with Abstract, followed by the main text, followed by a Material and Methods section, Acknowledgments, and References. The main text may contain up to four informative sub-headed sections. The total length of the main text (excluding references) should generally not exceed 3,000 words. Up to 6 items (tables or figures) may be included. Limit the number of references to about 50. Include supplementary material only where necessary and make sure that the supplementary material is supplied in a suitable format that can be readily accessed by all interested parties.
- **Genome Reports** are focused papers, usually of about 1500 words, that publish the main evolutionary message of new genome sequences as they become submitted to GenBank. Nearly every new genome sequence harbors an important advance in evolutionary understanding. The Genome Reports section at *GBE* is the premier forum to deliver that information directly to the genome evolutionary community in a rapid and efficient publication of the genome sequence. While the focus of Genome Reports are new genome sequences, submissions to Genome Reports may also contain specifically focused comparative analyses of previously published genomes that contain a substantial and novel insight of broadest evolutionary significance. Genome Reports should be formatted according to the guidelines given for Letters to *GBE*. Of course, *GBE* continues to welcome longer and more detailed manuscripts that report more fully on a new genome sequence and its evolutionary significance. These should be submitted as a standard Research Article.
- **Perspectives** address current topics in genome evolution. Perspectives do not report new data or new results. Instead they provide a forum for authoritative synthesis on issues at the forefront of genome evolution. While authors of Perspectives will have some freedom to explore new ideas and some leeway to express their own views,

submitted manuscripts will be peer reviewed and in order to be accepted they will need to represent an advance in understanding for the field. Perspectives may offer scholarly comment on papers that have appeared recently in *GBE* or in other journals, but submissions offering critique only will not be considered. That is, Perspectives need to convey a positive message moving forward. Perspectives should be terse, with up to 1500 words, two figures or tables and 30 references.

- **Reviews** can be solicited by the editors or can be suggested by prospective authors, who should contact the Reviews Editor with a roughly two-page outline summarizing the scope and content. The total length of the main text (excluding references) should not exceed 15,000 words. Reviews should have an abstract of 250 words or less, followed by a concise introduction, section headers as needed and a concluding summary. Reviews in *GBE* should present a balanced and up-to-date view of developments within their scope. Supporting graphics and tables are encouraged. Textboxes may be used to explain important concepts to the non-specialist reader. Supplementary materials should be avoided where possible. Limit the number of references to 200. Like Research Articles, Letters, and Perspectives, Reviews will be subject to peer review.
- **Genome Resources** report databases, applications, packages and methods that further the study of genome evolution. As data from genomes continue to accumulate, one of the biggest challenges facing the field is keeping up with the data themselves. The Genome Resources section at *GBE* will report tools that allow researchers to compile, extract, and analyze genome data to address evolutionary questions. Genome Resources can either be short papers, formatted according to the guidelines for Letters to *GBE*, or more comprehensive papers corresponding in length to a Research Article, depending on the nature of the resource.

General Format

Prepare your manuscript text using a Word processing package (save in .doc or .rtf format). Remember to number each page. Use double spacing (space between lines of type not less than 6 mm) throughout the manuscript and leave margins of 25 mm (1 inch) at the top, bottom and sides of each page. Please avoid footnotes. Type references in the correct order and in *GBE* style (see below). Type without hyphenation, except for compound words. Type headings in the style of the journal. Use the TAB key once for paragraph indents. Where

possible use Times for the text font and Symbol for the Greek and special characters.

Use the word processing formatting features to indicate strong **Bold**, *Italic*, Greek, Maths, Superscript, and Subscript characters. Clearly identify unusual symbols and Greek letters. Differentiate between the letter O and zero, and the letters I and l and the number 1. Mark the approximate position of each figure and table.

Check the final copy of your paper carefully, as any spelling mistakes and errors may be translated into the typeset version.

Sections of the Manuscript

- **Title**: The title should accurately advertise the paper's content and contain 150 characters or less including spaces.
- **Authors and affiliations:** Provide the name and institutional address of all authors, match addresses to names using superscript numbers.
- **Corresponding author:** The name of the author to whom all correspondence is to be addressed should be indicated with an asterisk in the author line and specified as follows:
    - *Author for Correspondence: John Smith, Department of Science, University of Somewhere, Anytown, USA, telephone number, fax number, email address
- **Data deposition:** Supply all accession numbers for the relevant databases. New sequence data must be deposited in GenBank/DDBJ/EMBL. Any sequence alignments used must be made available as supplemental information or by the corresponding author upon request.
- **Abstract:** The first page of the manuscript should begin with the abstract, which should be a concise summary of the paper. Avoided reference citations in the Abstract; if mentioned, the full reference must be given. The Abstract should contain 250 words or less.
- **Key words:** Up to six key words should be given below the abstract. Key words facilitate retrieval of articles by search engines, web directories and indexes; therefore, terms that are too general should be avoided. The selected key words should not repeat words given in the title. The aim is to assist potential readers to find the article by

clearly and specifically describing its subject matter, including aspects of methodology or the theoretical framework.

- **References:** Published articles and those in press (state the journal that has accepted them, provide a doi where possible) may be included. Do not include any reference cited *only* in Supplementary files.

  In the text citation, a reference should be cited by author and date. Do not place text other than the author and date within the parentheses. No more than two authors may be cited per text citation; if there are more than two authors, use et al. in the text (unless more are necessary to distinguish between references).

  In the reference list, list all authors if the author total is five authors or fewer; with more than five list the first author (only) followed by et al.

  At the end of the manuscript, the references should be typed in alphabetical order, with the authors' names, year, paper title, journal, volume number, inclusive page numbers, and name and address of publisher (for books only). The name of the journal should be abbreviated according to the World List of Scientific Periodicals. References should therefore be listed as follows:

  - Cagan RH, Rhein LD. 1980. Biochemical basis of recognition of taste and olfactory stimuli. In: van der Starre H, editor. *Olfaction and Taste VII.* Oxford: IRL Press. p. 35-44.
  - Marshall DA, Moulton DG. 1981. Olfactory sensitivity to alpha-ionone in humans and dogs. Chem Senses. 6:53-61.
  - van der Starre H, editor. 1980. *Olfaction and Taste VII*. Oxford: IRL Press.
  - Avoid personal communications and mention of unpublished data.
  - References to websites should be avoided, but if they are given, the references should give authors (if known), title of cited page, URL in full, and year of posting in parentheses.

- **Tables:** Tables should be prepared on separate sheets and numbered consecutively with Arabic numerals. They must be supplied in an editable format (.xlsx or .docx, for example) rather than an image. They should be self-explanatory and include a brief descriptive title. They should be of such a size that they fit easily onto a journal page, the type area of which is 234 (height) x 185 mm (double column width) or 89 mm (single column width). Footnotes to tables indicated by lower case letters are acceptable, but they should not include extensive experimental details.

- **Illustrations:** All illustrations (line drawings and photographs) must be referred to in the text (as Figure 1 etc.) and should be abbreviated to 'Fig. 1.' only in the figure legend. At online submission, you will be required to submit images electronically in one of the following formats: .jpg, .gif, .tif, .pdf or .eps.

  Each figure should be on a separate page and should be submitted at roughly final magnification. Use sans serif fonts such as Arial or Helvetica in figures. Use uniform font size in each figure whenever possible and recall that labels should never smaller than 6 pt at final magnification.
- **Electronic submission of figures:** Save figures at a resolution of at least 300 pixels per inch at the final printed size for color figures and photographs, and 600 pixels per inch for black and white line drawings. Color art *must* be submitted in CMYK rather than RGB format. Authors should be satisfied with the colors in CMYK (both on screen and when printed) before submission. Please also keep in mind that colors can appear differently on different screens and printers. Failure to follow these guides could result in complications and delays. For useful information on preparing your figures for publication, click here.
- **Figure legends:** These should be included at the end of the manuscript text. Define all symbols and abbreviations used in the figure. Common abbreviations and others in the preceding text need not be redefined in the legend.
- **Color Figures:** All figures submitted to the journal in color will be published in color online at no cost to authors.
- **Permissions for Illustrations and Figures:** Permission to reproduce copyright material, for print and online publication in perpetuity, must be cleared and if necessary paid for by the author; this includes applications and payments to DACS, ARS, and similar licensing agencies where appropriate. Evidence in writing that such permissions have been secured from the rights-holder must be made available to the editors. It is also the author's responsibility to include acknowledgements as stipulated by the particular institutions. Oxford Journals can offer information and documentation to assist authors in securing print and online permissions: please see the Guidelines for Authors section. Information on permissions contacts for a number of main galleries and museums can also be provided. Should you require copies of

this, please contact the editorial office of the journal in question or the Oxford Journals Rights department.

- **Funding:** Authors who are NIH-funded will have their paper automatically deposited in PubMed Central. Details of all funding sources for the work should be given in the 'Acknowledgements' section. A full list of RIN-approved UK funding agencies may be found here.

  The following convention should be followed:

  - The sentence should begin: 'This work was supported by …'
  - The full official funding agency name should be given, i.e. 'the National Cancer Institute at the National Institutes of Health' or simply 'National Institutes of Health' not 'NCI' (one of the 27 subinstitutions) or 'NCI at NIH (full RIN-approved list of UK funding agencies) Grant numbers should be complete and accurate and provided in brackets as follows: '[grant number ABX CDXXXXXX]'
  - Multiple grant numbers should be separated by a comma as follows: '[grant numbers ABX CDXXXXXX, EFX GHXXXXXX]'
  - Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
  - Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'.
  - Example: This work was supported by the National Institutes of Health [AP50 CA098252 and CA118790 to R.B.S.R.]; and the Education Research Council [hfygr667789].
  - Oxford Journals will deposit all NIH-funded articles in PubMed Central. See this page for details. Authors must ensure that manuscripts are clearly indicated as NIH-funded using the guidelines above.

  Other Information

- **Conventions:** In general, the journal follows the conventions of the *CSE Style Manual* (Council of Science Editors, Reston, VA, 2006, 7th ed.). Follow *Chemical Abstracts* and its indexes for chemical names. For guidance in the use of biochemical terminology follow the recommendations issued by the IUPAC-IUB Commission on

Biochemical Nomenclature, as given in *Biochemical Nomenclature and Related Documents*, published by the Biochemical Society, UK. For enzymes use the recommended name assigned by the IUPAC-IUB Commission on the Biochemical Nomenclature, 1978, as given in *Enzyme Nomenclature*, published by Academic Press, New York, 1980. Where possible, use the recommended SI (Systéme International) units.

Genotypes should be italicized; phenotypes should not be italicized.

- **Abbreviations:** Try to restrict the use of abbreviations to SI symbols and those recommended by the IUPAC-IUB. Abbreviations should be defined in parentheses after their first mention in the text. Standard units of measurements and chemical symbols of elements may be used without definition in the body of the paper.
- **Chemical Formulae and Mathematical Equations:** Wherever possible, write mathematical equations and chemical formulae on a single line. Submit complicated chemical structures as artwork.
- **Human and Animal Experiments:** The editors draw the authors' attention to the *Declaration of Helsinki* for Medical Research involving Human Subjects http://www.wma.net/e/policy/pdf/17c.pdf. In addition, when reporting experiments on animals, authors should indicate whether the institutional and national guidelines for the care and use of laboratory animals were followed.
- **Ethics Guidelines:** In order to guarantee a consistent policy of review and publication, *Genome Biology and Evolution* endorses the Ethics Guidelines offered by the Society for Neuroscience. These guidelines describe the responsibilities and expected conduct not only of authors of scientific articles, but also of the editors and reviewers. We encourage our readers to take a few minutes to download and look over these guidelines at http://www.sfn.org/guidelines/.
- **Crossref Funding Data Registry**: In order to meet your funding requirements authors are required to name their funding sources, or state if there are none, during the submission process. For further information on this process or to find out more about the CHORUS initiative please click here.

LANGUAGE EDITING

Language editing, if your first language is not English, to ensure that the academic content of your paper is fully understood by journal editors and reviewers is optional. Language editing does not guarantee that your manuscript will be accepted for publication. For further information on this service, please click here. Several specialist language editing companies offer similar services and you can also use any of these. Authors are liable for all costs associated with such services.