

**UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE CIÊNCIAS NATURAIS E EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE ANIMAL**

**Larissa Paim Bernardo**

**CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ELEMENTOS  
TRANSPONÍVEIS (*hAT*) ENCONTRADOS NOS GENOMAS DE  
ESPÉCIES DE *DROSOPHILA***

**Santa Maria, RS  
2017**



Larissa Paim Bernardo

**CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ELEMENTOS  
TRANSPONÍVEIS (*hAT*) ENCONTRADOS NOS GENOMAS DE ESPÉCIES DE  
*DROSOPHILA***

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em Biodiversidade Animal, Área de Concentração em Biologia Evolutiva de Elementos Transponíveis, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial da obtenção do grau de **Doutor em Biodiversidade Animal**.

Orientador: Prof. Élgion Lucio da Silva Loreto

Santa Maria, RS  
2017

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Bernardo, Larissa Paim  
CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ELEMENTOS  
TRANSPONÍVEIS (hAT) ENCONTRADOS NOS GENOMAS DE ESPÉCIES  
DE DROSOPHILA / Larissa Paim Bernardo.- 2017.  
137 p.; 30 cm

Orientador: Élgion Lucio da Silva Loreto  
Tese (doutorado) - Universidade Federal de Santa  
Maria, Centro de Ciências Naturais e Exatas, Programa de  
Pós-Graduação em Biodiversidade Animal, RS, 2017

1. Elementos transponíveis 2. Herves-like 3.  
Minissatélite 4. Transferência horizontal 5. Drosophila  
I. Loreto, Élgion Lucio da Silva II. Título.

---

© 2017

Todos os direitos autorais reservados a Larissa Paim Bernardo. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

E-mail: larissap.bernardo@gmail.com

**Universidade Federal de Santa Maria  
Centro de Ciências Naturais e Exatas  
Programa de Pós-Graduação em Biodiversidade Animal**

---

A comissão Examinadora, abaixo assinada, aprova a tese de Doutorado

**CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ELEMENTOS  
TRANSPONÍVEIS (*hAT*) ENCONTRADOS NOS GENOMAS DE  
ESPÉCIES DE *DROSOPHILA***

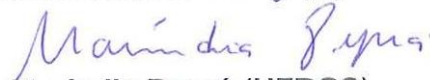
elaborada por  
**Larissa Paim Bernardo**

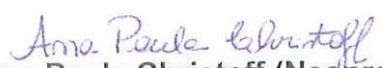
Como requisito parcial para a obtenção do grau de  
**Doutor(a) em Biodiversidade Animal**

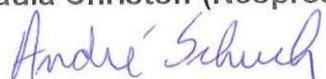
**COMISSÃO EXAMINADORA**

  
Elgion Lúcio da Silva Loreto, Dr.  
(Presidente/Orientador)

  
Vera Lucia Valente Gaiesky (UFRGS)

  
Maríndia Deprá (UFRGS)

  
Ana Paula Christoff (Neoprosecta)

  
André Schuch (UFSM)

**Santa Maria, 31 de Agosto de 2017**



## AGRADECIMENTOS

Difícil falar das pessoas que ajudaram durante minha trajetória na pós-graduação. Em meu longo caminho posso dizer que vivi dois momentos de ciência, um de fatura onde muitos foram valorizados como geradores de conhecimento e estão atualmente empregados, e o atual onde estamos enfrentando muitas dificuldades de financiamento e incertezas quanto ao futuro. Em muitos momentos me questioneei sobre a importância do que vinha fazendo, e aqui estou eu completando esta etapa de formação. Para um pesquisador, só o título não é suficiente. Estamos sempre correndo atrás de um bom currículo, de uma boa colocação, de uma boa instituição que nos possibilite criar. E por isso eu sempre gostei de dizer que: “Nunca termina quando acaba!” Agora é que começa, e algumas pessoas sempre estiveram e estarão comigo. Meu pai sempre incentivador e parceiro, sempre me ajudando com a parte financeira, mas acima de tudo, afetuoso. Sempre acreditando e ensinando-me a ser uma pessoa persistente e a nunca desistir antes do fim. A minha mãe e a Nana, que foram os anjos em forma de avós da Lulu, que possibilitaram que mesmo sendo mãe no meio do doutorado, eu conseguisse chegar ao final da pesquisa. Assim como elas, o Pedro, meu parceiro de vida, de laboratório, de bancada e de soluções de problemas. Hora marido, hora colega, hora “psicólogo”, mas toda hora pai. Realmente, eu não seria capaz, sem essas pessoas eu não teria força física e psicológica para realizar os trabalhos. Colegas de laboratório, todos, sem exceção. Todos me ajudaram muito, seja colocando uma bactéria a crescer, seja fazendo um gel, seja repicando minhas moscas, seja numa conversa. Foram muitos momentos, muitas parcerias. Me ensinaram muito, assim como ensiei, e essa troca fez de mim o que sou hoje. Além dos colegas, os amigos. Os que conheci aqui no LabDros, e os que estão comigo desde a graduação, colégio, desde sempre. Finalmente, agradeço muito o professor Élgion, pela confiança, pelos ensinamentos, pela paciência em orientar e ensinar. Também agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq – Brasil, pelo apoio financeiro e interesse neste estudo.





## RESUMO

### CARACTERIZAÇÃO FUNCIONAL E ESTRUTURAL DE ELEMENTOS TRANSPONÍVEIS (*hAT*) ENCONTRADOS NOS GENOMAS DE ESPÉCIES DE *DROSOPHILA*

AUTORA: Larissa Paim Bernardo

ORIENTADOR: PROF. Élgin Lucio da Silva Loreto

Elementos transponíveis (TEs) são sequências de DNA presentes em praticamente todos os organismos com genoma sequenciado que foram avaliados e apresentam como característica marcante a capacidade de se mobilizarem ou serem mobilizados. Essa mobilização normalmente reflete efeitos negativos para os genomas hospedeiros, o que gera uma tendência de que esses elementos sejam eliminados dos genomas. No entanto, a presença de TEs também podem representar vantagens adaptativas e estes podem ser mantidos nos genomas hospedeiros quando assumem alguma função biológica. A superfamília *hAT* de TEs é amplamente distribuída em plantas, animais e fungos. Apesar de ser uma das superfamílias de TEs mais estudadas, análises que avaliam atividade e interação desses TEs com os genomas onde estão hospedados ainda são escassos. Neste sentido, o objetivo geral desta tese foi contribuir para o entendimento estrutural, funcional e os processos evolutivos dos elementos *hATs* (*hobo* e *Herves-like*) presentes em espécies do gênero *Drosophila*, bem como investigar a influência dos minissatélites na atividade do elemento *howilli1* de *D. willistoni*. A tese será dividida em três capítulos. O primeiro capítulo tem como objetivo específico contextualizar os TEs estudados utilizando uma revisão bibliográfica que aborde aspectos de classificação, evolução e funcionalidade dos TEs, bem como apresentar os objetivos geral e específicos da tese. O segundo capítulo tem como objetivo apresentar os resultados obtidos ao longo dos estudos e é composto por três artigos: o primeiro fornece uma análise do cenário evolutivo do elemento *hobo* em espécies de *Drosophila* utilizando como marcador os TPEs *repeats*; o segundo consiste na caracterização dos minissatélites encontrados no elemento *howilli1* como potenciais formadores de miRNAs; o terceiro artigo consiste na análise de distribuição de elementos *Herves-like* em genomas de *Drosophila*. E o terceiro e último capítulo consiste nas discussões gerais e perspectivas.

**Palavras-chave:** Elementos transponíveis. Herves-like. Minissatélite. Transferência horizontal. *Drosophila*.



## ABSTRACT

### FUNCTIONAL AND STRUCTURAL CHARACTERIZATION OF TRANSPOSABLE ELEMENTS (*hAT*) FOUND IN *DROSOPHILA* GENOMES

AUTHOR: LARISSA PAIM BERNARDO

ADVISOR: PROF. DR. ELGION LÚCIO DA SILVA LORETO

Transposable elements (TEs) are DNA sequences present in almost all organisms with the genome sequenced and its main characteristic is the capability of mobilization. This mobilization usually causes negative effects to host, leading to elimination process as the main defence of hosts. However, the presence of TEs may also represent adaptive advantages, resulting in the maintenance into host genomes with biological functions. The *hAT* superfamily of TEs is widely distributed in plants, animals and fungi. However, even being one of the most studied TEs superfamily, activity and interaction analyses of these TEs with host genomes are still scarce. Thereby, our general objective in this work was to contribute with the knowledge of structural, functional and evolutive processes of *hATs* elements (*hobo* and *Herves-like*) present in *Drosophila* species, as well as investigate the role of minisatellites in the *D. willistoni howilli1* element. This thesis is divided into three chapters. The first chapter has the specific aim of review the studied TEs in the literature covering issues such as classification, evolution and function of TEs, as well as present the thesis general and specific objectives. The second chapter intended to show our results and it is divided into three articles: the first one shows the analysis of *hobo* element evolutive scenario in *Drosophila* species using as markers TPEs repeats; the second article consists in the characterization of minisatellites found in the *howilli1* element as potential generators of miRNAs; the last article analyzes the distribution of *Herves-like* elements into *Drosophila* genomes. The third and last chapter consists in general discussions and perspectives.

**Keywords:** Transposable elements. *Herves-like*. Minisatellite. Horizontal Transfer and *Drosophila*.



## SUMÁRIO

<b>CAPITULO I</b> .....	<b>13</b>
<b>INTRODUÇÃO GERAL</b> .....	<b>15</b>
1.1 ELEMENTOS TRANSPONÍVEIS.....	15
1.2 CLASSIFICAÇÃO .....	18
1.2.1 Superfamília <i>hAT</i> .....	24
1.2.2 Elemento <i>hobo</i> .....	29
1.2.3 Elementos <i>Herves-like</i> .....	31
1.2.4 MITEs .....	31
1.3 CICLO EVOLUTIVO DOS TES E SUA INTERAÇÃO COM OS GENOMAS HOSPEDEIROS .....	<b>34</b>
1.3.1 A fase de invasão .....	35
1.3.2 A fase de expansão, diversificação e de estabelecimento .....	37
1.3.3 Fase de silenciamento.....	38
1.4 REGULAÇÃO TRANSCRICIONAL E PÓS-TRANSCRICIONAL DE TES POR MEIO DE SNCRNAS .....	39
1.5 MINISSATÉLITES E TES .....	42
<b>2 OBJETIVOS</b> .....	<b>45</b>
2.1 OBJETIVO GERAL.....	45
2.2 OBJETIVOS ESPECÍFICOS.....	45
<b>CAPÍTULO II</b> .....	<b>47</b>
<b>3 (ARTIGO 1) UNRAVELING THE EVOLUTIONARY SCENARIO OF THE HOBO ELEMENT IN POPULATIONS OF DROSOPHILA MELANOGASTER AND D. SIMULANS IN SOUTH AMERICA USING THE TPE REPEATS AS MARKERS.....</b>	<b>49</b>
<b>4 (ARTIGO 2) A NEW SOURCE OF MIRNA: MINISATELLITES CARRIED BY A HAT TRANSPOSON IN DROSOPHILA WILLISTONI.....</b>	<b>57</b>
<b>5. (ARTIGO 3) CENÁRIO EVOLUTIVO E CARACTERIZAÇÃO DE ELEMENTOS TRANSPONÍVEIS HERVES-LIKE (HAT) EM ESPÉCIES DE DROSOPHILA.....</b>	<b>93</b>
<b>CAPÍTULO III</b> .....	<b>125</b>
<b>6 DISCUSSÃO GERAL E PERSPECTIVAS</b> .....	<b>127</b>
<b>REFERÊNCIAS</b> .....	<b>131</b>



## CAPITULO I







## INTRODUÇÃO GERAL

### 1.1 ELEMENTOS TRANSPONÍVEIS

Os TEs foram descobertos por Barbara McClintock na década de 40 em estudos de padrões de coloração das sementes de milho. McClintock observou as quebras cromossômicas no cromossomo 9, que ela atribuiu à presença de dois “fatores” que hoje sabemos se tratar de dois elementos transponíveis: *Ds* (do inglês "Dissociation" - localizado no local da quebra) e *Ac* (do inglês "Activator" - porque o sítio *Ds* não quebraria a menos que *Ac* estivesse presente) (McCLINTOCK, 1948; 1950). Na verdade McClintock estava observando a mobilização do elemento *Ds*, um elemento que, sozinho, não possui capacidade de mobilização, mas que pode ser mobilizado pelo elemento *Ac*. Ela observou que esses dois elementos poderiam mudar sua localização no genoma de forma aleatória, e essa capacidade de mobilização, que hoje sabemos ser uma característica marcante desse tipo de sequência repetitiva, faz com que TEs sejam sequências de DNA dinâmicas. A evolução dos TEs se dá por meio das interações que ocorrem quando um elemento surge em um novo genoma, de modo que a sua presença afeta de forma direta a evolução dos organismos onde estão contidos (BIÉMONT, 2010).

Apesar de hoje ainda termos uma baixa representatividade da biodiversidade global em projetos de sequenciamento de genomas, sabe-se que TEs estão presentes em praticamente todos os organismos. Entre os eucariotos com sequenciamento completo do genoma a quantidade de TEs é variável, mas praticamente todos apresentam TEs em seus genomas (SCHNABLE et al., 2009). Alguns organismos apresentam até 85% do seu genoma composto de sequências de TEs, como é o caso do milho (GARDNER et al., 2002). Sabe-se que, em humanos, somente 1,5% do genoma compreende regiões codificadoras de proteínas de cópia única e 43% do genoma é composto por TEs (BIÉMONT, 2010).

É natural assumirmos uma correlação positiva entre o grau de complexidade estrutural de um organismo e o tamanho do genoma, mas essa correlação não reflete a realidade devido à presença de numerosas cópias de TEs presentes nos genomas (BIÉMONT, 2010). Esta ampla distribuição de TEs em genomas tão diferenciados nos remete a uma grande diversidade de sequências e mecanismos de mobilidade, o que gera a necessidade de um sistema de classificação para TEs.

Entretanto, nenhum sistema de classificação proposto até hoje consegue contemplar a ampla complexidade desses elementos (FINNEGAN, 1992; WICKER, 2007; KAPITONOVE; JURKA, 2008; PIÉGU et al., 2015; ARENSBURGER et al., 2016; BAO et al., 2015).

A mobilidade, que normalmente reflete o grau de atividade dos TEs, pode ser favorecida por agentes bióticos e abióticos. Por exemplo: os elementos *copy-like* e *1731* apresentam uma ativação da transposição em culturas celulares; infecções virais estão relacionadas com a excisão do elemento *gypsy*, ativação da transposição do elemento *mdg2* e transposição somática de *msg1*; fatores abióticos e estresse ambiental estão relacionados com ativação de transposição dos elementos  *copia*,  *copia-like*, *412*, *P* (GUERREIRO, 2011) e *mariner* (JARDIM et al., 2015).

A mobilização leva a interações entre os genomas hospedeiros e as sequências de TEs. Essas interações normalmente remetem a consequências deletérias, pois dependendo do lugar onde as cópias de TEs são inseridas, podem ocorrer rearranjos cromossômicos, recombinações desiguais, quebras cromossômicas e mudança nos padrões de expressão e/ou regulação gênica. Entretanto, sabe-se que eles também podem atuar como componentes de variabilidade genética, podendo ser incorporados em sequências gênicas do hospedeiro, adquirindo alguma função biológica (domesticação gênica), e conseqüentemente, gerando inovações biológicas (FESCHOTTE; PHRITHAN, 2007; LEVIN; MORGAN, 2011; HOFFMANN et al., 2015).

Um processo importante de interação é o envolvimento dos TEs com processos de regulação gênica. Essa interação ocorre, por que sequências de TEs também podem atuar como sequências precursoras de pequenos RNAs não codificantes (do inglês, small non-coding RNAs - sncRNAs), tais como pequenos RNAs interferentes (siRNAs), micro RNAs (miRNAs) e (piRNAs). Essas diferentes classes de sncRNAs diferem basicamente em sua biogêneses, mas a sua atuação na regulação gênica ocorre quando essas moléculas servem como guia no reconhecimento de alvos podendo ser incorporados no complexo de silenciamento do genoma hospedeiro. A regulação ocorre dependendo do tipo de tecido ou fase do desenvolvimento do organismo e variam desde a formação de heterochromatina até a desestabilização de mRNA e controle de tradução.

Os TEs podem ser precursores de sncRNAs a partir de transcritos aberrantes de sequências codificadoras de transposases, mas também transcritos que apresentem regiões invertidas terminais repetidas (do inglês, *terminal inverted repeats* TIRs) ou elementos curtos (MITEs – do inglês: *Miniature Inverted–Repeat Transposable Elements*) (PIRIYAPONGSA; JORDAN, 2007; PIRIYAPONGSA et al., 2007). Além de serem precursores de sncRNAs, que participam de silenciamento gênico em diferentes fases do ciclo de vida do genoma hospedeiro, os TEs também podem gerar sncRNAs que atuam na sua própria regulação, como por exemplo em casos de TEs muito invasivos, que aumentam rapidamente em número de cópias e conseqüentemente podem de forma aleatória inserir-se em regiões de “*clusters*” de sncRNAs gerando moldes de silenciamento para suas proteínas de mobilização. Conseqüentemente, um sistema de autorregulação é gerado, levando a redução de mobilidade. Por essa razão, sncRNAs serão abordados em detalhe na seção “Ciclo evolutivo dos TEs e sua interação com os genomas hospedeiros”.

Aliado a todos esses efeitos, TEs são comumente transferidos entre espécies distantemente relacionadas filogeneticamente, processo conhecido como transferência horizontal (do inglês, *Horizontal Transfer* - HT). Embora a HT seja comum em bactérias, nos genomas eucariotos os exemplos de transferência gênica de forma horizontal normalmente remetem a genes bacterianos sendo transferidos para essas espécies. Entretanto, o avanço do conhecimento sobre TEs, seus mecanismos de mobilização, suas formas de invasão e interação com os genomas hospedeiros, sua ampla diversidade e distribuição, demonstra que essas sequências são propensas a eventos de HT. Inclusive, HT é um ponto determinante no ciclo evolutivo de TEs e esse assunto será discutido detalhadamente.

No campo da engenharia genética, o poder de inserção em sequências alvo específicas nos genomas despertou o interesse de alguns pesquisadores para o uso dos TEs como vetores universais, como, por exemplo, o elemento *mariner Himar1* (GRABUNDZIJA et al., 2010; PICARDEAU, 2010). Também se sabe que muitas doenças humanas estão relacionadas à atividade de TEs. Além disso, foi proposto que a atividade de TEs pode ter impactos sobre a biologia de determinadas células somáticas, relacionando as modificações geradas pelos TEs à biogênese de algumas formas de câncer (LEVIN; MORGAN, 2011).

Aqui, focamos em elementos da superfamília *hAT*, de maneira a contribuir com o entendimento da diversidade, evolução e funcionalidade dessa família

amplamente distribuída e diversa em plantas animais e fungos (RUBIN et al., 2001). Em especial, em sequências repetitivas encontradas em alguns desses elementos, buscando desvendar o papel biológico dessas repetições.

## 1.2 CLASSIFICAÇÃO

Quando o genoma humano foi sequenciado pela primeira vez, verificou-se que uma porção considerável do genoma não era responsável por expressar quaisquer características. Essa porção aparentemente sem função foi popularmente chamada de “DNA Lixo” (BIÉMONT, 2010). Atualmente, sabe-se que a maior parte dessas sequências correspondem a TEs, e eles podem estar contidos também em regiões codificadoras, sobrepondo-se e interrompendo exons, ou ainda dentro de introns, nas proximidades de regiões reguladoras de genes, podendo causar impactos na regulação de genes adjacentes (HUA-VAN et al., 2011; GEO et al., 2012). O potencial de interação faz com que os TEs tenham um papel significativo na evolução dos genomas hospedeiros (REBOLLO et al., 2012; BIÉMONT, 2010).

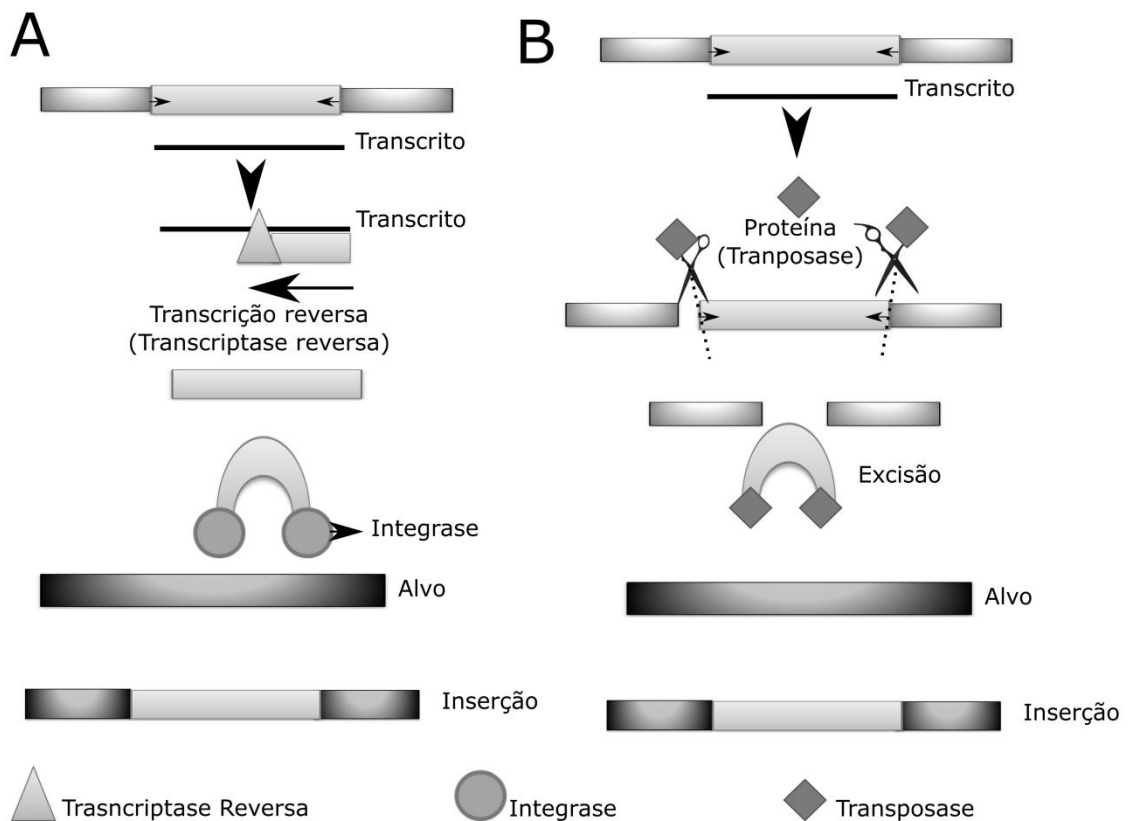
À medida que as tecnologias de sequenciamento de DNA permitiram que os genomas de diversas espécies fossem acessados, a investigação desses genomas mostrou uma diversidade espantosa de TEs (BAO et al., 2015). Sendo assim, a criação de um sistema de classificação e nomenclatura para eles é importante. Entretanto, devido a sua característica parasitária de invasão dos genomas, muitas vezes, cópias de um mesmo elemento podem ocorrer em espécies filogeneticamente distintas, o que gera certa confusão na nomenclatura.

A classificação dos TEs é feita com base na estrutura de suas sequências e no seu mecanismo de transposição. O primeiro sistema de classificação de TEs foi proposto por Finegan, em 1989 e é usado até hoje. Esse autor sugeriu que os TEs fossem primeiramente divididos em duas classes: a classe I, dos Retrotransposons, que é composta por elementos que utilizam o RNA como intermediário de transposição; e a classe II dos Transposons, que é composta por elementos que utilizam DNA como intermediário de transposição.

A transposição de retroelementos (classe I) ocorre com a síntese de RNA mensageiro (mRNA) pela maquinaria de transcrição do genoma hospedeiro. Este mRNA é transportado para o citoplasma onde ocorre a síntese das enzimas relacionadas com a transposição, codificadas pela sequência do elemento transcrito

em mRNA. A Transcriptase Reversa (do inglês, *Reverse Transcriptase - RT*) é uma das enzimas produzidas pelos retroelementos. Ela é responsável por sintetizar uma nova cópia de DNA a partir da sequência do retroelemento, utilizando a própria molécula de mRNA recém transcrita como molde. Essa nova molécula de DNA e todas as enzimas relacionadas com a transposição são direcionadas para o núcleo onde irá acontecer a integração do elemento em alvos do DNA hospedeiro. Esse mecanismo de transposição é conhecido como copia-e-cola (do inglês, *copy and paste*) ou transposição replicativa, e resulta no aumento do número de cópias do TE a cada evento de transposição (CAPY et al., 1997; WICKER et al., 2007) (Figura 1A).

Figura 1 – Mecanismo de transposição de TEs. A. Mecanismo de transposição de elementos de classe I; B. Mecanismo de transposição de elementos de Classe II



Os transposons de classe II se movem por enzimas transposases que são codificadas na sequência do próprio elemento que apresenta autonomia para sua mobilização (Figura 1B). As transposases apresentam domínios conservados que excisam o elemento de um sítio doador, pelo reconhecimento de sequências

específicas das extremidades do elemento, e o inserem em um sítio receptor característico (para uma revisão, VARANI et al., 2015). Esse tipo de transposição é conhecido como corta-e-cola (do inglês, *cut-and-paste*), ou transposição conservativa. As transposases podem mobilizar transposons diferentes (mobilização em *trans*) desde que haja o reconhecimento das sequências das extremidades do elemento. Esse tipo de mobilização é comum em elementos MITEs (CAPY et al., 1997; WICKER et al., 2007). Esses elementos são sequências curtas, normalmente originadas a partir de elementos autônomos que sofreram deleções internas e perderam a sequência codificadora da transposase, mas mantiveram suas extremidades conservadas. Essa característica possibilita o reconhecimento dos MITEs pela transposase de um elemento autônomo derivativo ou relacionado (WICKER et al., 2007; VARANI et al., 2015).

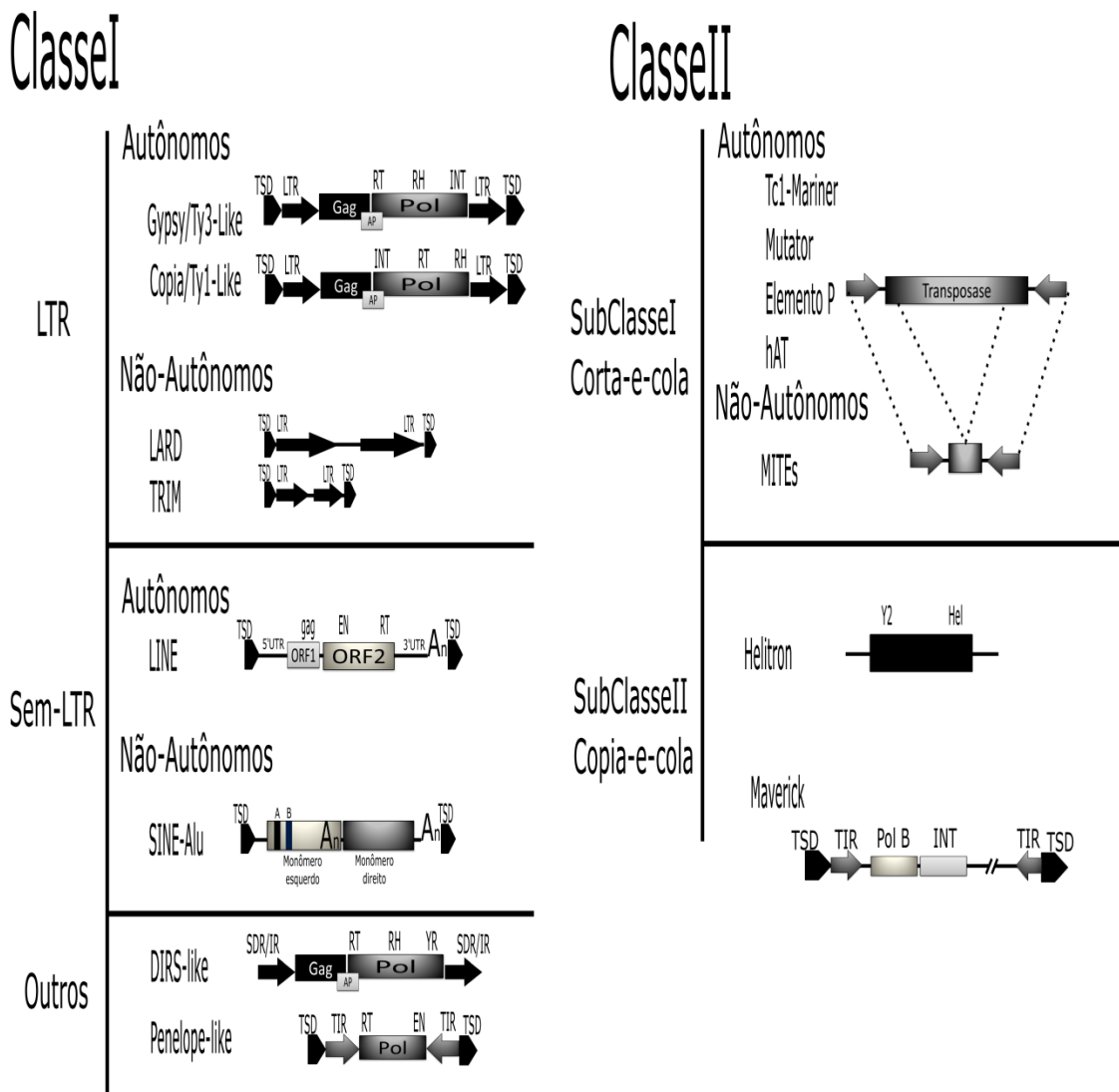
Ambas as classes apresentam elementos autônomos e não autônomos (Figura 2). Elementos autônomos apresentam sequência codificadora das enzimas envolvidas com a transposição, enquanto que os elementos não autônomos não apresentam essas sequências. As classes ainda podem ser subdivididas em subclasses, ordens, superfamílias ou famílias (WICKER et al., 2007). Ambas as classes de TEs são classificadas nas diferentes categorias de classificação com base na estrutura de suas sequências (WICKER et al., 2007).

Os Retroelementos podem ou não apresentar em suas extremidades longas terminações repetidas diretas (do inglês, *Long Terminal repeats* – LTRs) (Figura 2). Essas sequências servem como sítio de reconhecimento e regulação dos elementos de classe I e podem variar de 300 até 5.000 nucleotídeos. Os retroelementos podem chegar até 25 mil pares de base (kpb) e possuem os genes *gag* e *pol*. O gene *gag* codifica uma proteína com domínio semelhante ao do capsídeo viral. Já a expressão do gene *pol* codifica uma poliproteína. A maturação dessa poliproteína é catalisada por um domínio proteico que codifica para uma protease aspártica (do inglês, *Aspartic Protease* - AP) (Figura 2) (WICKER et al., 2007; VARANI et al., 2015). Sendo assim, o domínio AP está envolvido com a transcrição das demais enzimas de transposição, que são: Integrase (INT), Transcriptase Reversa (RT) e RNaseH (RH) (KAZAZIAN, 2011) (Figura 2).

Os retroelementos são categorizados em cinco ordens, com base em suas características de transposição, organização dos domínios proteicos e filogenia da transcriptase reversa (Figura 2) (WICKER et al., 2007; VARANI et al., 2015): (1) LTR

retrotransposons, que incluem, por exemplo, os elementos autônomos *gypsy/Ty3-like*, *copia/Ty1-like* *BEL-Pao*, bem como os não-autônomos *LARD* e *TRIM*; (2) *DIRS-like* (do inglês, *Dictyostelium Intermediate Repeats Sequence*) que apresentam no lugar do domínio proteico para INT, uma tirosina recombinase (YR) gerando um padrão de transposição diferente dos demais retroelementos com LTR, pois esses elementos não apresentam LTR, mas sim repetições diretas parciais semelhantes as *TIRs* dos Transposons de DNA; (3) *Penelope-like* que em sua estrutura apresenta somente regiões codificadoras para RT e endonucleases (EN) e durante a transposição geram TSD (do inglês, *Target Site Duplication*); (4) *LINEs* (do inglês, *Long Interspersed Nuclear Element*), elemento autônomo que também não apresenta LTR, sua transposição ocorre por meio de transcritos codificados por suas duas ORFs (do inglês, *Open Reading Frame*) que codificam uma EN e uma RT, sendo o seu processo de transposição conhecido como *Target-Primed Reverse Transcription* (TPRP) e ambas as extremidades do elemento são flanqueadas por repetições diretas; (5) *SINEs* (*Short Interspersed Nuclear Element*) que são elementos não autônomos que também não apresentam LTR. Esses elementos dependem dos elementos *LINE* para se transpor, apesar de apresentarem uma origem distinta, não sendo originados por deleções internas de elementos autônomos (WICKER et al., 2007; VARANI et al., 2015).

Figura 2 – Classificação e estrutura de TEs de classe I e II. TSD: *Target Sequence Duplication*; TIR: *Terminal Inverted Repeat*; LTR: *Long Terminal repeats*; SDR/IR: repetição Direta Parcial/Inversão Repetida; gag: Capsídeo; AP: Protease aspártica; RT: Transcriptase Reversa; RH: RNaseH; INT: Integrase; YR: Tirosina Recombinase relacionada a Transposase-HUH; EN: Endonucleases; ORF: Open Reading Frame; Y2: Transposase-HUH; Hel: DNA Helicase; Pol-B: DNA polimerase; 5'UTR: Região não traduzida; 3'UTR: Região não traduzida com cauda poli-A; Monômero esquerdo: Região derivada de tRNA com sítio A e B homólogos a promotores de RNA polimerase III



Fonte: Modificado de Varani et al., 2015.

Os transposons de DNA podem apresentar de centenas a milhares de pares de base de tamanho e são representados por alguns elementos simples de procaríotos chamados sequências de inserção (do inglês, *Insertion Sequences - IS*), e ainda são divididos em duas subclasses distintas pelo número de fitas de DNA que



são clivadas durante a transposição. A subclasse I utiliza o mecanismo corta-e-cola das duas fitas de DNA (*double-strand transposition*), e é representada pelas ordens *TIR* e *Crypton* (WICKER et al., 2007). A ordem *TIR* apresenta elementos com repetições terminais invertidas e as nove superfamílias conhecidas são diferenciadas pelas TIRs, pelas sequências alvo duplicadas (do inglês, Target Sites Duplications - TSDs) e pelas características filogenéticas de suas transposases (WICKER et al., 2007). As TIRs (9-40pb) flanqueiam os elementos e são utilizadas como sítio de reconhecimento da transposase, que corta-e-cola o elemento em um sítio doador, e reinsere o TE em um sítio receptor, que apresenta sequência específica e que é duplicado durante a inserção, gerando as TSDs (Figura 1B). Este tipo de transposição pode promover o aumento do número de cópias de TEs. Por exemplo, a quebra da dupla fita de DNA pode ativar o sistema de reparo da célula hospedeira que restaura a fita original, aumentando assim o número de cópias do TE (VARANI et al., 2015), e também se a transposição ocorrer durante a divisão celular, em células germinativas e somáticas, quando o transposon se movimenta para uma região de cromossomo ainda não duplicada (CHEN et al., 1992). Os transposons *Crypton* contêm uma tirosina recombinase como alguns transposons de procariotos IS e retrotransposons *DIRS-like*, mas não apresentam um domínio RT. Isso sugere que eles se transpõem através de um intermediário de DNA, porém esta ordem de TEs ainda é pouco estudada (WICKER et al., 2007).

A subclasse II utiliza o mecanismo corta-e-cola de apenas uma fita de DNA (*single-strand transposition*) e é dividida em duas ordens: *Helitron* e *Maverick*. Os elementos da ordem *Helitron* se transpõem através de um mecanismo conhecido como “círculo rolante” (do Inglês, *rolling-circle*) que corta apenas uma fita de DNA e não gera TSDs (KAPITONOV; JURKA, 2001). Suas extremidades são definidas por domínios CT ou CTRR (onde R representa uma base purina) (WICKER et al., 2007). Os elementos da ordem *Maverick* apresentam TIRs, e em geral, mostram homologia limitada com proteínas de vários vírus de DNA. Eles codificam até 11 proteínas que podem variar em número e ordem, resultando em uma DNA polimerase B e uma integrase relacionada com as integrases de elementos de classe I, mas não possuem sequências codificadoras para transcriptase reversa, sugerindo que os elementos da ordem *Maverick* também sofrem transposição replicativa sem RNA intermediário, por excisão de uma única cadeia de DNA seguida de replicação extra-cromossômica, e depois a integração em um novo local (WICKER et al., 2007).

### 1.2.1 Superfamília *hAT*

A superfamília *hAT* é muito diversa e ocorre em plantas, fungos e animais (RUBIN et al., 2001), o que sugere uma origem antiga (Figura 3). O nome dessa superfamília deriva da primeira letra dos três primeiros elementos descritos: *hobo* de *Drosophila* (McGINNIS et al., 1983), *Ac* do milho (McCLINTOCK 1947; 1948) e *Tam3* de *Antirrhinum majus* (HEHL et al., 1991). Elementos *hATs* são abundantes no genoma humano, compreendendo cerca de 195Mb, que representa 1,55% do genoma, mas nenhum elemento *hAT* apresentou sinais de atividade nos últimos 50 milhões de anos (LANDER et al., 2001). Entretanto, o elemento *TcBuster* (*Tribolium castaneum*) apresentou alta atividade em experimentos com células humanas (LI et al., 2012). Já no genoma dos morcegos *Myotis lucifugus* foi detectado grande diversidade de elementos da superfamília *hAT* com sinais de expansão recente, e apesar de não terem sido encontrados elementos autônomos a abundância desses transposons nesses genomas leva a crer que alguns genomas de mamíferos podem ainda suportar altas atividades de *hATs* (RAY et al., 2008).

Os transposons *SPIN* (do inglês, *Space Invaders*) apresenta sequências altamente conservadas distribuídas nos genomas de vários mamíferos (gambá, pequenos primatas noturnos, porco-espinho da família *Tenrecidae*, roedores da família *Murinae* e pequenos morcegos marrons), bem como no genoma de um sapo africano, um lagarto, e em inseto da família *Reduviidae* (GILBERT et al., 2010; PACE et al., 2008), mas ausentes em gatos, cachorros, bovinos, tatu (mamífero da ordem *Cingulata*), musaranho e humanos (PACE et al., 2008) (Figura 3). No gênero *Drosophila* também observamos elementos conservados distribuídos nos genomas de espécies distantemente relacionadas, como por exemplo, os elementos *harrow* (MOTA et al., 2010), *hosimary* (DEPRÁ et al., 2010), *hobo-brothers* (BERNARDO; LORETO, 2013) e *But2* (ROSSATO et al., 2014). Esta distribuição descontínua de sequências muito similares sugere que eventos de HTT devem ter auxiliado na ampla distribuição dessa superfamília (RUBIN et al., 2001).

A sequência dos elementos da superfamília *hAT* podem chegar a 5kb, apresentam TSDs de 8 bp, TIRs de 5–27 bp e suas ORFs codificam proteínas de 500-800aa, que mediam transposições do tipo corta-e-cola (FESCHOTTE; PRITHAM, 2007). As transposases dos elementos *hATs* apresentam de 20-60% de similaridade entre as suas sequências e quatro domínios catalíticos: (1) Domínio N-

terminal contendo um dedo de zinco (Zn-F-BED), (2) um domínio de ligação de DNA, (3) um domínio catalítico contendo os dois primeiros “Ds” do domínio DDE/DDD<sup>1</sup> que é interrompido pelas sequências do (4) domínio de inserção, seguido do restante das sequências do domínio catalítico com o último “E/D” (ZHOU et al., 2004; HICKMAN et al., 2005). O domínio DDE/DDD é compartilhado por todas as superfamílias de transposases do tipo corta-e-cola em eucariotos, sugerindo uma origem em comum desse mecanismo de transposição (YUAN; WESSLER, 2011).

Arensburger e colaboradores (2011) e Rossato e colaboradores (2014) subdividiram a superfamília *hAT* em três famílias, com base na similaridade de suas transposases e suas TSD e nomearam as famílias conforme o nome dos elementos que foram descritos primeiramente: *Ac* em *Zea mays*, *AeBuster* em *Aedes aegypti* e *Tip100* em *Ipomoea purpurea* (HABU et al., 1998). As variações nas sequências da transposase dos elementos das famílias *Ac* e *Buster* ocorrem principalmente no domínio de ligação de DNA e no domínio de inserção (ARENSBERG et al., 2011).

A família *Ac* é composta por elementos encontrados em animais, plantas e fungos (Figura 3). A maioria dos elementos membros da família *Ac* apresentam um consenso de suas TSDs com o segundo e sétimo nucleotídeos apresentando uma timina e uma adenina, como segue: 5'-nTnnnnAn-3'.

Os elementos *Ac* que compõe essa família são encontrados em vários insetos, tais como: *A. aegypti* (*AeHermes2* e *AeHermes1*), *D. melanogaster* (*hobo*), *D. willistoni* (*harrow*), *Musca domestica* (*Hermes*) *Bactrocera tryoni* (*Homer* e *Hopper\_we*) e *Anopheles gambiae* (*Herves*). Também nas espécies de peixes *Oryzias latipes* (*Tol2*) e *Danio rerio* (*Aclike1* e *Aclike2*); no fungo *Tolypocladium inflatum* (*Restless*); no verme *Caenorhabditis elegans* (*Aclike*); nas plantas *Zea mays* (*Ac*), *Antirrhinum majus* (*Tam3*) e *Arabidopsis thaliana* (*Daysleeper*); e em humanos (*Tramp*) (Figura 3). Sabe-se que os elementos *Ac*, *Tam3*, *Hobo*, *Hermes*, *Herves* e *Tol2* desta família são ativos (ARENSBURGER et al., 2011).

Por sua vez, os elementos que compõe a família *Buster* apresentam uma distribuição complexa, contendo elementos de mamíferos e invertebrados (Figura 3), além de elementos da rã *Xenopus silurana tropicalis* (*XtBuster*); do peixe *D. rerio* (*DrBuster1*, *DrBuster2*, *hAT5\_DR*); do verme *Ciona savignyi* (*CsBuster*); e os dois

---

<sup>1</sup> Transposase-DDE/DDD: apresenta um resíduo catalítico “D” que indica a presença de um resíduo aspartato (Asp) e “E” que representa o resíduo glutamato (Glu). Também podem ser representadas como DD(35)E, onde o segundo D é separado do E/D por aproximadamente 35 aminoácidos.

elementos de ouriço-do-mar *Strongylocentrotus purpuratus* (*SpBuster1* e *SpBuster2*) (ARENSBURGER et al., 2011) (Figura 3). Os elementos dessa família apresentam similaridade e TSD consenso, onde o quarto e o quinto nucleotídeo são uma timina e uma adenina, como segue: 5'-nnnTAnnn-3'. Sabe-se que os elementos *AeBuster1* de *A. aegypti* e *TcBuster* de *T. castaneum* são ativos e Arensburger e colaboradores (2011) ainda sugeriram que mosquitos sejam os vetores potenciais dos vários casos de HT desta família de TEs, já que são vetores conhecidos de agentes patogênicos entre vertebrados e invertebrados. No mesmo trabalho, os autores também demonstraram alto grau de conservação de elementos *Buster* em vários genes de mamíferos, incluindo primatas, cães, cavalos e bovinos. Esses resultados indicam uma função essencial, mas ainda desconhecida para esses genes.

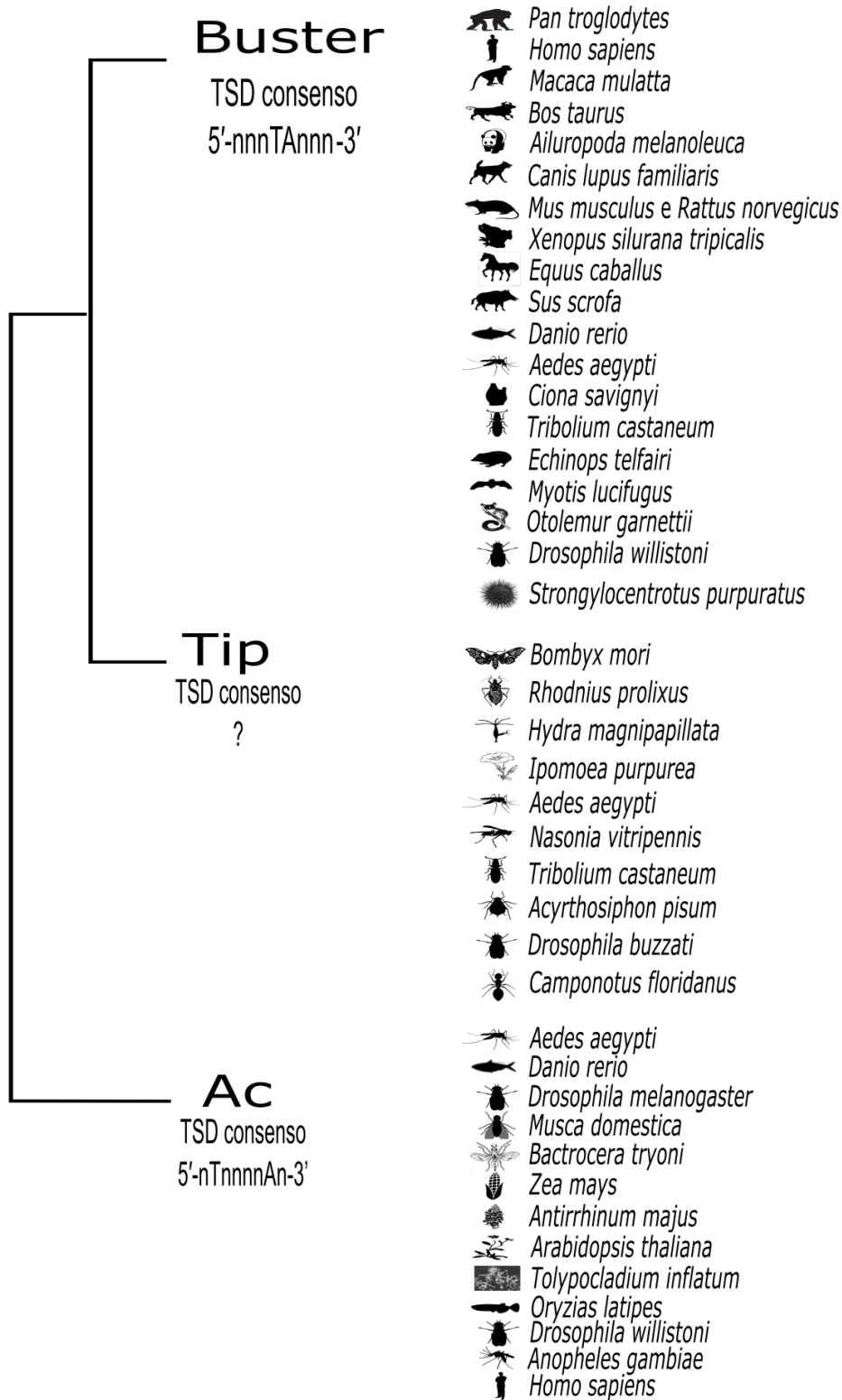
A família *Tip* foi recentemente estabelecida com base em análises filogenéticas das sequências proteicas (ROSSATO et al., 2014). Ela é composta basicamente por TEs descritos em *A. aegypti* (*AeTip100-2*) (ARENSBURGER et al., 2011), três elementos que foram descritos em *Hydra magnipapillata* (*hAT12\_HM*, *hAT29\_HM* e *hAT46\_HM*) (JURKA, 2008), um em *Ipomoea purpurea* (*IpTip100*) (CÁCERES et al., 2001) e um em *Drosophila* (*But2*) (CÁCERES et al., 2001). Entretanto, buscas *in silico* por elementos dessa família mostraram várias outras proteínas hipotéticas de insetos que estão incluídas nesta família, mas não foi possível determinar se estas proteínas são provenientes de sequências de TEs ativos ou de genes domesticados (ROSSATO et al., 2014). Mais estudos são necessários a fim de uma melhor caracterização dessa família.

Com o sequenciamento de novos genomas e caracterização de novos TEs, provavelmente mais famílias de elementos *hATs* poderão ser diagnosticadas. Em 2009, Ortiz e Loreto caracterizaram novos *hATs* de *Drosophila*, ao analisar os 12 genomas disponíveis na época. Os autores descobriram 37 novos elementos com homologia a elementos *hATs*: oito em *D. ananassae*, 11 em *D. mojavensis*, dois em *D. sechelia*, um em *D. simulans*, dois em *D. virilis*, três em *D. yakuba*, três em *D. persimilis*, um em *D. pseudoobscura*, um em *D. grimshawi* e cinco em *D. willistoni*. Posteriormente, uma análise mais detalhada das sequências isoladas nos genomas mostrou que várias cópias apresentam encurtamento de suas sequências, bem como a presença de cópias degeneradas derivadas dessas sequências (ORTIZ et al., 2010). A análise filogenética das cópias potencialmente ativas desses

elementos demonstrou que vários deles apresentam uma distribuição incongruente com a filogenia das espécies onde foram descritos, sugerindo que sua distribuição nos genomas de *Drosophila* pode estar sendo mediada por eventos de HT, assim como sugere que esses elementos podem estar sofrendo ou ter sofrido uma expansão recente nas populações de *Drosophila* estudadas (DEPRÁ et al., 2010; MOTA et al., 2010; BERNARDO; LORETO, 2013).

Em conjunto, essas descobertas mostram que muitos TEs *hATs* podem ser ou ter sido ativos muito recentemente em espécies de *Drosophila*, reforçando a necessidade de estudos com elementos dessa superfamília.

Figura 3 – Cenário atual de distribuição de elementos *hATs* em genomas eucariotos gerado a partir de dados



As três famílias de elementos *hATs* atualmente aceitas estão representadas pelos organismos onde ocorrem os TEs. A sequência consenso de TSDs está indicada para cada família.  
Fonte: Arensburger et al. (2011) e Rossato et al. (2014).

### 1.2.2 Elemento *hobo*

Entre os elementos da superfamília *hAT*, *hobo* é o único para o qual está disponível um conjunto completo de dados em nível genético, fisiológico e populacional, bem como estudos que contam a sua história evolutiva (McGINNIS et al., 1983; DANIELS et al., 1990b; LADEVÈZE et al., 2012).

Em *D. melanogaster* o elemento *hobo* pode ocorrer em duas formas: (1) Elemento autônomo ou cópia canônica e suas cópias com deleções internas; (2) Cópias degeneradas, conhecidas como “Relics”. Ambos os elementos invadiram o genoma dessa espécie quando elas ocorriam apenas no continente europeu, antes de ela se tornar cosmopolita, mas o elemento canônico invadiu o genoma dessa espécie recentemente.

Um estudo com elementos relacionados a *hobo* (*hobo-brothers*) recentemente contextualizou a história de invasão dos *hobo-brothers* nas espécies de *Drosophila* Neotropicais com o elemento *hobo* trazido pelas espécies de *D. melanogaster* e *D. simulans*, quando estas se tornaram cosmopolita. O trabalho sugere que os elementos *hobo-brothers* diversificaram a partir de um ancestral comum entre eles e o elemento *hobo*, e este ancestral provavelmente invadiu o continente americano a aproximadamente 22 a 42 milhões de anos atrás (BERNARDO; LORETO 2013). Porém o elemento *hobo* canônica de *D. simulans* e *D. melanogaster*, provavelmente invadiu esses genomas antes de essas espécies se tornarem cosmopolitas, pois o elemento está presente em espécies do grupo *melanogaster* que são endêmicas da África.

Na América sabe-se que o elemento *hobo* canônico invadiu os genomas de *D. simulans* e *D. melanogaster* após 1950 (PASCUAL; PERIQUET, 1991), e existem evidências de que, apesar de coexistir no genoma de *D. melanogaster*, o elemento *hobo* canônico e suas cópias com deleções internas são aquisições recentes, onde *D. simulans* é o potencial doador do elemento *hobo* canônico para *D. melanogaster*, mas as cópias degeneradas “Relics” são resultados de invasões antigas e mantidas nos genomas por transferência vertical (do inglês, *vertical transfer* – VT) (DANIELS et al., 1990a e b).

O elemento *hobo* canônico foi descrito por McGinnis e colaboradores (1983) apresentando 2959 pb de tamanho, TIRs de 12 pb e uma ORF de 1,9 kb que codifica sua transposase. Esse elemento é ativo em *D. melanogaster* e foi

relacionado a casos de disgenesia do híbrido<sup>2</sup>, várias mutações recessivas letais e quebras cromossômicas (BLACKMAN et al., 1989; YANNOPOULOS et al., 1987; SHEEN et al., 1992; LADEVÈZE et al., 1998). Sua ORF apresenta uma região repetitiva em tandem de 9 bp (ACTCCAGAA) conhecida como repetições TPE, que codificam os aminoácidos Treonina, Prolina e Acido glutâmico. O polimorfismo dessas repetições TPEs pode ser utilizado como marcador molecular e possibilita prover informações sobre a história de invasão do elemento em populações naturais de *D. melanogaster*. Além da possibilidade de prever padrões e cenários evolutivos para o elemento *hobo*, as variações nos números de cópias TPEs foram associadas com efeitos na eficiência da atividade da transposase (BONNIVARD et al., 2002).

Sabendo que *D. melanogaster* possivelmente adquiriu esse elemento após 1950, e com base em estudos de polimorfismos de repetições TPEs, o cenário evolutivo atual para a invasão do elemento *hobo* canônico prevê duas invasões independentes: A primeira invasão de elementos *hobo* em populações de *D. melanogaster*, que se espalharam para o resto do mundo, possivelmente ocorreu por elementos que apresentam 3 TPEs (PERIQUET et al., 1989). Uma segunda invasão de elementos com 5-7 TPEs poderia ter ocorrido em populações da Europa Central, África do Sul e América do Sul, mas elementos com esses polimorfismos também foram encontrados em populações da América do Norte e no Norte da África. Entretanto, esta última região conecta populações da Europa e Sul da África o que gera dúvidas sobre a história e o sentido dessa segunda invasão do elemento *hobo* em populações dessas regiões, tornando a América do Sul a região mais indicada como origem dessa invasão (BONNIVARD et al., 2000 e 2002). Uma amostragem maior das populações de *D. melanogaster* e *D. simulans*, quanto a presença de polimorfismos e diversidade de sequências do elemento *hobo* na América do Sul poderia auxiliar o entendimento da sua história evolutiva.

---

<sup>2</sup> Disgenesia do híbrido (do inglês, *Hybrid dysgenesis*) é uma síndrome causada pelo cruzamento de linhagens de *Drosophila*, onde os machos/fêmeas que apresentam um elemento ativo e uma fêmea/macho que não apresentam o elemento ao cruzarem geram uma incompatibilidade citoplasmática na prole, tornando-os estéreis ou com complicações genéticas.



### 1.2.3 Elementos *Herves-like*

Em 2009, Ortiz e Loreto analisaram os genomas de *Drosophila* disponíveis em busca de elementos *hATs*. Eles descreveram 37 novos *hATs*, e em sua análise filogenética descreveram um grupo de elementos que eles chamaram de família *Herves*, empregando os critérios de divisão de famílias proposto por Wicker e colaboradores (2007). Este grupo de elementos é formado por elementos que ocorrem no genoma do mosquito, elemento *Herves* (*A. gambiae*) e *hAT-2\_AG* (*A. aegypti*); e elementos que ocorrem em moscas, tais como: o elemento *hAT-1\_DSi*, que ocorre em *Zaprionus indianus*, *D. melanogaster*, *D. simulans*, *D. mauritiana* e *D. sechelia*; *howilli1* que ocorre em *D. willistoni*; o elemento *homo2* que ocorre em *D. mojavensis*; e *hAT-1\_DP* que ocorre em *D. persimilis* e *D. pseudoobscura*.

Este grupo de elementos atualmente são membro da família *Ac* de elementos *hATs* descrita por Arensburger e colaboradores em 2011, mas por serem um grupo monofilético aqui serão denominados elementos *Herves-like*. O nome desse clado foi escolhido por que o primeiro elemento descrito do grupo foi o elemento.

Dentre esses elementos, *hAT-1\_DP* e *howilli1* foram caracterizados com uma região de minissatélite na extremidade 5' de ambos os elementos (5'MnRs), e na extremidade 3' do elemento *hAT-1\_DP* (3'MnRs). Esses minissatélites apresentam em torno de 18nt de tamanho e variam de 12-20 em número de cópias (ORTIZ et al., 2010). Apesar de esses elementos apresentarem uma divergência de 0.28 em suas sequências, os minissatélites da porção 5' (5'MnRs) são altamente conservados nessas espécies e esse fato levantou questões sobre uma possível função biológica para essas repetições.

### 1.2.4 MITEs

Os MITEs são um grupo de TEs não autônomos com sequências curtas, que apesar de terem sido descobertos primeiramente em plantas, estão presentes em vários genomas animais, incluindo *Drosophila* (ORTIZ et al., 2010; DEPRÁ et al., 2012). Esses elementos normalmente ocorrem em grande número de cópias e são normalmente associados a genes. Essa associação possibilita a hipótese de que

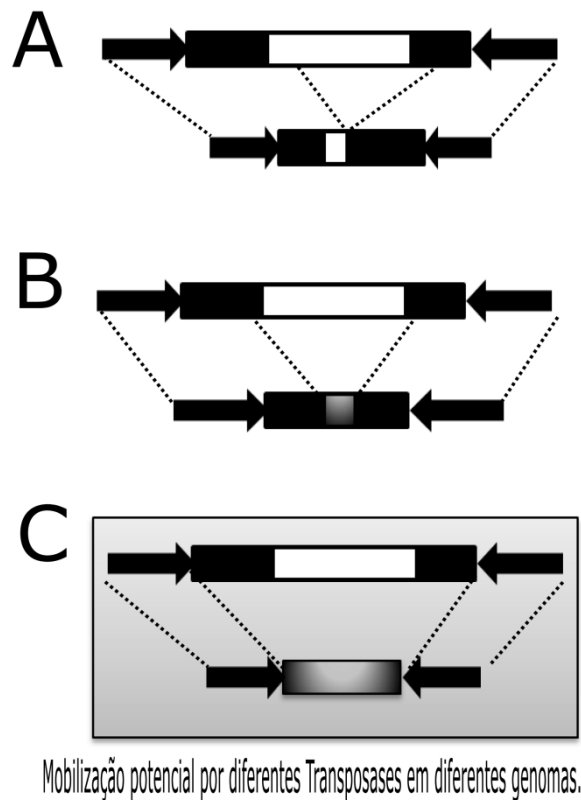
esses elementos podem desempenhar um papel importante na regulação gênica e evolução dos genomas (JIANG et al., 2004; LU et al., 2012).

O papel desses elementos vem sendo cada vez mais associado à biogênese de pequenos RNAs não codificantes (do inglês, small noncoding RNAs – sncRNAs) (PIRIYAPONGSA; JORDAN, 2007; PIRIYAPONGSA et al., 2007), mas para entendermos o papel dos MITEs em processos de regulação gênica, como precursores de sncRNAs, é fundamental entendermos sua origem, bem como investigar pequenos RNAs derivados de MITEs, que possam ter efeitos na regulação da expressão gênica.

Quanto às características das sequências e origem, os MITEs normalmente apresentam em torno de 500pb até 1000pb, e são relacionados a cópias de TEs não autônomos pela similaridade de suas sequências, sendo considerados cópias com deleções internas em relação aos TEs autônomos relacionados a eles (WICKER et al., 2007). Por outro lado, alguns são aparentemente resultantes de eventos de recombinação que geram um par de TIRs iguais ou similares que podem ser reconhecidos por transposase de elementos autônomos e, conseqüentemente, ser mobilizados (JIANG et al., 2004). Como o reconhecimento desses elementos é feito basicamente por similaridade de sequências, sua classificação em famílias pode ser de difícil compreensão. Feschotte e Colaboradores (2002) propuseram critérios para classificar os elementos MITEs em suas diferentes superfamílias. Para isso, utilizaram quatro níveis de padrões de similaridade de sequências (Figura 4): São considerados MITEs: (1) aqueles elementos que apresentarem alta similaridade de toda a sua sequência com o elemento autônomo do qual possivelmente foi derivado; (2) aqueles elementos com regiões terminais ou sub-terminais altamente similares aos autônomos dos quais são possivelmente derivados, ou ainda com regiões terminais ou sub-terminais altamente similares, mas com sequências internas diferentes aos elementos autônomos dos quais são possivelmente derivados; (3) aqueles elementos que apresentam somente TIRs similares ou podendo apresentar pouca variação em seus nucleotídeos em relação às TIRs dos elementos autônomos dos quais possivelmente esse MITE foi derivado e quando analisado o genoma de um organismo; (4) aqueles elementos que apresentam somente TIRs similares ou podendo apresentar pouca variação em seus nucleotídeos em relação às TIRs dos elementos autônomos dos quais possivelmente esse MITE foi derivado quando analisamos elementos de organismos diferentes.

Os níveis 1 e 2 demonstram evidências de que o elemento autônomo pode ter um envolvimento na origem e amplificação do seu MITE derivado. Já os níveis 3 e 4 provêm informações sobre possíveis elementos autônomos que podem estar envolvidos na mobilização do MITE em questão, mas não gera informação sobre sua origem.

Figura 4 – Classificação de MITEs



Onde A, B e C representam os diferentes padrões de similaridade de sequências utilizados para classificar elementos MITEs. A. Toda a sequência do MITE altamente similar a sequência do elemento autônomo derivativo; B. apenas as TIRs e regiões subterminais altamente similares aos elementos autônomos derivativos; C. Apenas as TIRs do elemento semelhantes aos elementos autônomos derivativos, em análises de diferentes genomas. Seta representa as TIRs. Em branco representamos a sequência codificadora da transposase do elemento autônomo derivativo. Em cinza representamos sequências internas que não correspondem a nenhuma porção da sequência do elemento autônomo relacionado. O retângulo cinza ao redor da figura C indica a estrutura de elementos MITEs que não fornecem informações sobre origem, apenas sobre potencial de mobilização.

Fonte: Feschotte e Colaboradores (2002).

Os elementos *hATs* caracterizados por Ortiz e colaboradores (2010) nos genomas de *Drosophila* disponíveis na época apresentaram várias cópias derivadas, muitas das quais possivelmente são MITEs mobilizáveis pelas transposases de

elementos autônomos também descritos. Sabendo da importância dos MITEs na evolução dos genomas, sabendo que eles são bons colonizadores de genomas e sabendo que os elementos *Herves-like*, com minissatélites, também apresentam cópias curtas descritas pelos autores, estudos de caracterização de sequências, distribuição e possíveis funções biológicas para os MITEs de elementos *hATs* são importantes.

### 1.3 CICLO EVOLUTIVO DOS TES E SUA INTERAÇÃO COM OS GENOMAS HOSPEDEIROS

A capacidade de mobilização possibilita que TEs sejam uma fonte de perturbações das organizações estruturais e funcionais dos genomas, e por esse motivo na maioria dos casos a inserção pode ser prejudicial para os hospedeiros. Consequentemente, a manutenção deles nos genomas é um desafio, mas a persistência deles vem sendo demonstrada a medida que novos genomas são sequenciados.

Devido as suas características de invasão que remetem a relações parasitárias, inicialmente os TEs foram considerados “genes egoístas” (BIÉMONT, 2010). Atualmente, sabe-se que TEs estão amplamente distribuídos nos genomas e que sua capacidade de mobilização também pode representar uma fonte de variabilidade genética, e por esta razão, o entendimento das interações entre as sequências de TEs e os genomas nos quais estão contidos são de extrema importância, visto que o surgimento de um novo TE em um genoma pode afetar de forma direta a evolução do próprio TE e a evolução dos genomas hospedeiros.

Jurka e colaboradores (2011) presumem que a maior parte, senão todas, as grandes famílias de TEs de mamíferos originam-se de um pequeno número de cópias que são ativamente expressas, denominadas genes “fonte”, “fundador” ou “mestre”, e essas cópias originam as famílias de elementos repetitivos que se distribuem em populações de organismos. De um ponto de vista bastante amplo, os genomas eucariotos também podem ser vistos como ambientes variáveis e muitas vezes hostis para os TEs (VENNER et al., 2009). O local de inserção de um TE é importante para sua sobrevivência, pois nem todas as cópias inseridas prosperam e a tendência é que suas sequências degenerem progressivamente, acumulando mutações e desaparecendo.

A chance de persistirem nos genomas depende (1) da capacidade invasiva; (2) da capacidade de ser transmitido; (3) e das interações que ocorrem a partir da sua inserção nos genomas hospedeiros. Esses três fatores interagem de forma direta na manutenção dos TEs nos genomas.

A capacidade invasiva reflete diretamente a capacidade de aumentar o número de cópias e essa característica pode possibilitar ao TE um escape de extinção, pois aumentando em número o TE aumenta também as chances de ser transmitido às gerações futuras. Entretanto, também aumenta as chances de sofrer uma forte pressão seletiva.

A capacidade de ser transmitido entre espécies não relacionadas filogeneticamente, por transferência horizontal, possibilita a invasão de genomas diferenciados, o que pode também facilitar o escape do sistema de silenciamento dos genomas.

Interações que sejam mutualística, como no caso de genes de TEs que são incorporados nos genomas hospedeiros por aquisição de alguma função biológica, processo conhecido como domesticação (VOLFF, 2006), possibilitam o escape da seleção natural, e também a sobrevivência dessas sequências por mais tempo nos genomas.

Neste sentido, podemos então assumir que os TEs apresentam um ciclo evolutivo que pode ser resumido em três fases: 1) a fase de invasão, na qual esses elementos surgem nos genomas; 2) a fase de expansão, caracterizada por um aumento rápido no número de cópias; 3) e a fase de silenciamento, na qual o acúmulo de mutações e a diversificação do sistema de defesa do genoma hospedeiro pode levar a extinção dos TEs. Todas as fases são mediadas por várias interações que podem ser positivas ou negativas para os genomas hospedeiros.

### **1.3.1 A fase de invasão**

A invasão de um novo genoma pode ocorrer por eventos de HT ou por recombinação de diferentes elementos presentes no genoma (como elementos antigos ou relíquia) (LE ROUZIC; CAPY, 2009). Eventos de HT vêm sendo bastante documentados (para uma revisão, veja WALLAU et al., 2012), e o sucesso de invasão pode ser influenciado por características estruturais da sequência do próprio

elemento, por mecanismos de mobilização, pela capacidade de aumento rápido do número de cópias e pela capacidade de autorregulação.

As características estruturais de TEs (tais como: o encurtamento de sequências, ser mobilizável, possuir poucos genes e não depender de genes específicos do genoma hospedeiro para sua mobilização) influenciam de forma direta na capacidade do elemento aumentar rapidamente o número de cópias, de maneira que a taxa de transposição se torna maior do que a perda devido à seleção natural ou deriva genética. Entretanto, esses fatores não são determinantes para o estabelecimento dos TEs nos genomas hospedeiros, pois as taxas de transposições excessivas podem gerar mudanças significativas nos genomas invadidos, o que aumenta as chances de repressão pelo genoma hospedeiro e consequente silenciamento.

Quando conseguimos reunir um conjunto de dados que avaliam a distribuição, a estrutura das sequências e o potencial de atividade dos TEs, podemos observar uma relação em que, em uma fase inicial, os TEs aumentam rapidamente o número de cópias, e posteriormente, observa-se a queda da atividade de transposição, o acúmulo de cópias degeneradas seguida de novos processos de invasão e colonização. Podemos citar como exemplo, as famílias de elemento *mariner* que ocorrem em espécies de *Drosophila* (WALLAU et al., 2014).

Em estudos que avaliaram a atividade de elementos em genomas que não apresentam o elemento naturalmente, também foi observado altas taxas de transposição, como no caso de experimentos que avaliaram a atividade de elementos *hATs* em células humanas que não apresentam elementos ativos dessa superfamília de TEs (LI et al., 2011). Entretanto, esse aumento de atividade pode representar apenas o “impulso evolutivo” do elemento nesse novo genoma (LORETO; PEREIRA, 2017), não sendo possível determinar se o elemento seria capaz de se estabelecer, pois outras interações, tais como presença de um sistema de silenciamento para o TE, são fundamentais para o sucesso de colonização.

Os diferentes mecanismos de mobilização também são importantes para o sucesso de invasão, pois a integração da sequência do TE no genoma hospedeiro deve ser estável. Em elementos de classe I podemos perceber uma relação evolutiva forte com os vírus. O elemento *gypsy*, por exemplo, é capaz de produzir um capsídeo viral e essa característica sugere que ele pode promover sua própria HT (revisado por WICKER et al., 2007; VARANI et al., 2015). Elementos de classe II

são mais simples, apresentam apenas uma ORF e normalmente não necessitam utilizar proteínas do hospedeiro para sua transposição. As características das sequências dos elementos MITEs representam uma vantagem no estabelecimento desses elementos em um novo genoma. O encurtamento de suas sequências em relação ao elemento autônomo relacionado e suas extremidades conservadas e reconhecidas por transposases de diferentes elementos possibilitam a persistência dos MITEs por períodos prolongados nos genomas, de maneira que é possível detectar o sinal dos elementos autônomos que já tenham sido extintos dos genomas por seleção natural ou perda estocástica (WALLAU et al., 2014; ZHANG et al., 2013).

Outro fator determinante na fase de invasão, que pode interferir nos eventos de HT, é a diferenciação entre células somáticas e germinativas, que ocorre em muitos eucariotos multicelulares (LORETO; PEREIRA, 2017). Do ponto de vista evolutivo, a mobilização de TEs em células germinativas gera alteração do número de cópias para gerações futuras, mas mobilização em células somáticas pode gerar alterações fenotípicas que podem interferir de forma direta na sobrevivência e adaptação dos organismos hospedeiros (LORETO; PEREIRA, 2017).

### **1.3.2 A fase de expansão, diversificação e de estabelecimento**

A segunda fase (Expansão) do ciclo evolutivo dos TEs é marcada por ondas de expansão e contração no número de cópias, que podem resultar em mudanças profundas na organização dos genomas hospedeiros, gerar rearranjos cromossômicos e alterações gênicas em nível estrutural e regulatório. Esta fase é determinante para o estabelecimento e propagação dos TEs nas populações de genomas. Durante essa fase podem acontecer interações positivas, nas quais os TEs servem como geradores de variabilidade genética para o genoma hospedeiro, e conseqüentemente, para a evolução das espécies (FESCHOTTE, 2008), mas também podem ser negativas, como nos casos em que os TEs se inserem em genes ou regiões regulatórias essenciais para a sobrevivência do hospedeiro, podendo gerar mutações ou inviabilidade da prole (BREGLIANO; KIDWELL, 1983). Neste sentido, muitos TEs evoluíram de maneira que seus mecanismos de transposição foram direcionando as inserções em “portos seguros” nos genomas, e dessa

maneira minimizando seus danos nos genomas hospedeiros (LEVIN; MORAN 2011).

Normalmente a fase de expansão dos TEs ocorre até o surgimento de algum sistema de repressão, regulação ou autorregulação, seguida de inativação ou silenciamento, que pode ser a última fase do ciclo evolutivo dos TEs, mas o ciclo pode recomeçar caso esses sofram HT e invadam novos genomas, ou caso eles sejam ressuscitados através de recombinações ectópicas que podem originar TEs ativos novamente, ou caso o sistema de silenciamento seja reprimido por meio de variações ambientais estressantes, de maneira que os elementos se tornem ativos novamente, ou ainda se forem domesticados e persistirem nos genomas hospedeiros desempenhando alguma função biológica importante (VOLFF, 2006).

### **1.3.3 Fase de silenciamento**

Após a fase de propagação, onde podemos assumir que através do aumento do número de cópias os TEs se estabelecem nos genomas, e dependendo dos locais de inserção e dos tipos de relações de interação que se estabelecem, existe uma tendência de que esses elementos sejam eliminados dos genomas, seja por perda estocástica e acúmulo de mutações ou por estabelecimento de um sistema de silenciamento por parte do genoma hospedeiro. Mas para que TEs sejam silenciados, eles primeiramente precisam ser reconhecidos como agentes causadores de perturbações pelos mecanismos de defesa do genoma hospedeiro. Porém, os TEs não apresentam nenhuma particularidade em suas sequências que os diferenciem dos demais trechos de DNA genômico, e por esse motivo, o silenciamento de TEs acontece quando o seu comportamento nos genomas hospedeiros gera perturbações na integridade dos genomas, a ponto de desencadear o sistema de silenciamento.

A fase de silenciamento visa reduzir os efeitos deletérios associados à taxa elevada de mutação observada na fase de transposição excessiva (LE ROUZIC; CAPY, 2009). Diversos mecanismos de regulação e silenciamento de TEs estão presentes nos genomas eucariotos, e podem ocorrer em níveis transcricionais, tais como, metilação do DNA e modificação de histonas, que geram alterações na estrutura da cromatina, ou mecanismos pós-transcricionais por meio de classes especiais de pequenos RNAs não codificantes (*Small non-coding RNAs* - sncRNAs)



(HUDA et al., 2011; KALMYKOVA et al., 2005; MATZKE et al., 2000). Ambos os mecanismos de regulação são utilizados pelo hospedeiro não só para regular atividade de TEs. Por exemplo, sncRNAs estão envolvidos em muitos processos biológicos, tais como: tempo de desenvolvimento, diferenciação, proliferação e morte celular, controle metabólico e defesa antiviral (KIM; SIOME, 2009).

Em eucariotos, sncRNAs apresentam cerca de 20-30nt e apresentam associação com proteínas da família Argonauta (AGO), podendo ser as da subfamília AGO, que apresentam associação com siRNAs (do inglês, *short interfering RNAs*) e miRNAs (do inglês, *micro RNAs*) ou proteínas da subfamília PIWI, que apresentam associação com os piRNAs (do inglês, *PIWI-interacting RNAs*). Essas três classes de pequenos RNAs são codificadas em genomas eucariotos, mas diferem em biogêneses e associação com as proteínas AGO.

#### 1.4 REGULAÇÃO TRANSCRICIONAL E PÓS-TRANSCRICIONAL DE TES POR MEIO DE SNCRNAS

Os genomas eucariotos apresentam uma organização estrutural que pode ser distinta em regiões de eucromatina e heterocromatina. A região eucromática é representada por porções de DNA onde os nucleossomos são dispersos possibilitando uma acessibilidade maior às regiões gênicas e facilitam os processos transcricionais. Já as regiões de heterocromatina são caracterizadas por porções de DNA fortemente ligadas aos nucleossomos, tornando as regiões de DNA pouco acessíveis para a maquinaria de transcrição gênica (HUISINGA et al., 2006). Mas apesar de apresentar um papel subjetivo na expressão gênica, regiões heterocromáticas podem desempenhar um papel importante em processos biológicos, tais como: reparo de DNA (Hassa e Hottinger, 2005; Greenberg, 2011), manutenção de silenciamento de TEs (MALONE; HANNON, 2009) e, conseqüentemente, mantendo a integridade dos genomas.

Em *Drosophila*, as regiões eucromáticas apresentam acetilação de histonas H3 e lisina 4 (H3K4) que representa um tipo de marcação característico de regiões com atividade transcricional. Já nas regiões heterocromáticas os nucleossomos, que são associados com TEs, são enriquecidos por metilação de histonas H3 e lisina 9 (H3K9), que representa um sinal de marcação para regiões que apresentam

repressão de atividade transcricional (GANDREL et al., 2002; MARTENS et al., 2005).

Em defesa aos potenciais danos causados por elementos móveis ativos, os genomas eucariotos desenvolveram um sistema de defesa epigenética, para suprimir a atividade de TEs que, em análises genômicas, normalmente são encontrados em regiões de heterocromatina. Um TE epigeneticamente inativo, retém o potencial de codificação da proteína responsável por sua mobilização, mas não a produz, devido a sua localização em um ambiente de cromatina repressiva.

siRNAs (do inglês, *short interfering RNAs*) são uma classe de sncRNAs que apresentam cerca de 21 nt, que são originados a partir de múltiplas clivagens de longas cadeias de RNA dupla fita (dsRNA) (do inglês, long double strand RNAs). Essas clivagens são realizadas através de enzimas DICER que geram duplex de RNAs curtos de aproximadamente 21nt, onde cada fita apresenta fosfato 5' e um grupo hidroxila 3', emparelhados de modo a deixar saliências de dois nucleotídeos nas extremidades 3' (GHILDUYAL et al., 2009). Uma das fitas geradas é utilizada como molde no silenciamento gênico, interagindo com proteínas da família AGO, para silenciamento transcricional ou pós-transcricional. O reconhecimento dos alvos de silenciamento dessa classe de RNAs requer pareamento perfeito, e além do silenciamento gênico esses siRNAs também são relacionados com a participação na defesa antiviral (HÁ; KIM, 2014).

O silenciamento pós-transcricional, por meio de siRNAs, ocorre a partir da origem e incorporação de fragmentos curtos e de fita simples de siRNA no complexo de silenciamento (RISC). Essa incorporação ocorre por associação de siRNAs com proteínas da família AGO, que utiliza esses sncRNAs como guia de clivagem de transcritos nascentes em regiões eucromáticas.

O silenciamento transcricional e a regulação epigenética de TEs podem ocorrer quando muitos transcritos são gerados e clivados a partir de regiões de cromatina, levando a um recrutamento de proteínas metiltransferase envolvidas com adição de grupos metil em H3K9, e possivelmente a metilação das bases de citosina do DNA em organismos nos quais o DNA é metilado. As modificações na cromatina onde ocorrem os TEs resultam em heterocromatina condensada e inacessível à transcrição.

siRNAs mais longos (entre 24 e 35 nt) os piRNAs, demonstraram suprimir a atividade de TEs em células germinativas de animais, por meio de associação com

proteínas AGO específicas de células germinativas, chamadas proteínas PIWI. Apesar de nossa compreensão dos mecanismos de biogênese dos piRNAs, pouco se sabe sobre como um *locus* específico no genoma é convertido em um sítio produtor de piRNAs (YAMANAKA et al., 2014). Assim, as sequências de piRNA derivadas desses *clusters* são homólogas não só para TEs nos *clusters*, mas também em TEs relacionados localizados em outro lugar no genoma e, desta maneira, os piRNAs atuam como moléculas guia para reprimir TEs *in trans* (para uma revisão, YAMANAKA et al., 2014).

A origem dos piRNAs ocorre por produção de transcritos de regiões ricas em sequências defeituosas de TEs, conhecidas como *clusters* de piRNAs. Esses *clusters* são evolutivamente conservados entre os organismos e quando transcritos, geram longas cadeias duplas de RNA, que são processadas por meio de RNases, que independem das proteínas DICER. Após processamento por RNases, esses transcritos geram fragmentos curtos e de fita simples de RNA, que se associam as proteínas PIWI e servem como molde para reconhecimento e silenciamento de TEs ativos durante as primeiras fases de desenvolvimento.

O silenciamento pode ser em nível transcricional ou pós-transcricional, tanto por meio de inibição de transcritos que são exportados para o citoplasma, como por meio de modificações específicas da cromatina nos *loci* de TEs no núcleo (CHAPMAN et al., 2007; KIM et al., 2009). Em *Drosophila*, os piRNAs gera um sistema conhecido como “ping-pong”, onde um piRNA primário é gerado a partir dos *clusters* de piRNAs, e após o reconhecimento da sequência de mRNA do TE, novos piRNAs são gerados e estabilizam a atividade do elemento (BRENNECKE et al., 2007).

Os microRNAs ou miRNA são envolvidos especificamente em mecanismos de regulação gênica através de associação com proteínas da família AGO1 (CZECH; HANNON, 2011), e atualmente é crescente a aceitação de que muitas famílias de miRNAs devem ser originados a partir de sequências de TEs (ROBERTS et al., 2014), o que sugere uma influência grande de TEs em processos de regulação gênica.

A origem de miRNAs ocorre também por clivagens de RNAs dupla fita, mas algumas diferenças podem ser percebidas. Os transcritos que antecedem esses sncRNAs são originados a partir de precursores de RNA em forma de grampo (do inglês, *hairpin stem loop*), denominados miRNAs primários (pri-miRNAs). Eles

apresentam aproximadamente 1kb e são transcritos por RNA polimerase II (RNA Pol II). O grampo gerado é caracterizado por um eixo de RNA dupla fita (do inglês, *double strand* – dsRNA) com alguns nucleotídeos não pareados e conectado por uma região terminal de fita simples (alça). Esse transcrito em forma de grampo apresenta uma 7-metilguanossina (CAP) na extremidade 5' e são poliadenilados na extremidade 3' (AXTELL et al., 2011).

Podemos distinguir três fases na biogênese dessa classe de sncRNAs: 1) Transcrição; 2) Maturação; 3) Incorporação. A transcrição é feita pela RNA Pol II e gera pri-miRNAs, capazes de gerar as estruturas secundárias em forma de grampo. A fase de maturação consiste no processamento dos pri-miRNAs em pré-miRNAs com modificações que levam a liberação de grampos simples. Em *Drosophila*, essa maturação envolve as enzimas DROSHA e DGCR8/PASHA. Posteriormente, os grampos simples são exportados para o citoplasma onde sofrem modificações por meio de ação de proteínas DICER, que liberam fragmentos curtos de fita simples de RNA oriundos do eixo de dsRNA. Por fim, na fase de incorporação, esses pequenos RNAs de fita simples são incorporados no complexo RISC, associados a proteínas da família AGO1 e servindo como molde para o silenciamento gênico (KIM et al., 2009; HÁ; KIM, 2014; FABIAN; SONENBERG, 2012).

Os miRNAs regulam a expressão de genes quando se ligam na porção 3'UTR do mRNA alvo (SHEN et al., 2015). Essa ligação não requer pareamento perfeito de todos os nucleotídeos que compõe o miRNA, e a sequência semente, que é uma região conservada normalmente na posição 2-7nt da sequência do miRNA, é utilizada no reconhecimento de alvos de silenciamento. O sentido da ligação do miRNA e seu alvo de mRNA gera uma tendência de manutenção de conservação das extremidades 5' dos miRNAs. Desta forma diferentes miRNAs com variações na extremidade 3', conhecidas como isoformas, podem surgir, mas estas moléculas compartilham o mesmo alvo de silenciamento por conservação da sequência semente (NEWMAN et al., 2011).

## 1.5 MINISSATÉLITES E TES

Os genomas eucariotos são ricos em sequências repetitivas de DNA definidas de acordo com o número de repetições e o tamanho de suas sequências. Repetições de sequências curtas (microsatélites) são unidades de repetições que

variam de 1nt à até 6nt. Já, unidades repetidas longas (minissatélites), podem variar de 10nt à até 60nt. Sequências repetitivas oriundas de TEs já foram relacionadas a diferentes fases do ciclo evolutivo de TEs, como por exemplo: As repetições TPES do elemento *hobo* foram relacionadas com a eficiência dos processos de transposição do elemento; Repetições em tandem (do inglês, *Tandem Repeats* - TRs) presentes nas TIRs do elemento *galileu* e em extremidades 3' do retroelemento *Ty3/Gypsy* estão relacionados com a formação de grampos estáveis que teriam função no sinal de terminação da transposição; As repetições em tandem do *helitron DINE-TR1* foram descritas, mas seu papel biológico ainda não foi completamente entendido, mas foi sugerido que o elemento *DINE-TR1* poderia prover repetições em tandem para os genomas, de maneira que quando esses elementos se inserirem em *clusters* de piRNAs podem ser incorporados no sistema de silenciamento e participarem de processos epigenéticos de regulação gênica (DIAS et al., 2016). Além disso, TEs são uma fonte importante de sequências de miRNAs, não só sendo originados a partir de transcritos da sequência codificadora da transposase, mas também as TIRs e elementos curtos MITEs (PIRIYAPONGSA; JORDAN, 2007; PIRIYAPONGSA et al., 2007; ROBERTS et al., 2014).



## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Contribuir para o entendimento estrutural, funcional e os processos evolutivos dos elementos *hATs* (*hobo* e *Herves-like*) presentes em espécies do gênero *Drosophila*, bem como investigar a influência dos *minissatélites* na atividade do elemento *howilli1* de *D. willistoni*.

### 2.2 OBJETIVOS ESPECÍFICOS

- 1) Analisar o cenário evolutivo de elementos *hATs* em espécies de *D. melanogaster* e *D. simulans* por meio de *TPE repeats* encontrados em elementos *hobo*;
- 2) Investigar a função biológica de minissatélites encontrados no elemento *howilli1* como potenciais formadores de miRNAs;
- 3) Investigar possíveis alvos de silenciamento de miRNAs provenientes de minissatélites repeats (do inglês, *Minisatellites Repeats* - MnRs) do elemento *Herves-like howilli1*;
- 4) Investigar a ocorrência de MnRs em outros elementos *Herves-like* relacionados a *howilli1*;
- 5) Investigar os padrões de distribuição de elementos *Herves-like*, de modo que um cenário evolutivo para esses elementos possa ser proposto.





## CAPÍTULO II





**3. (ARTIGO 1) UNRAVELING THE EVOLUTIONARY SCENARIO OF THE *HOBO* ELEMENT IN POPULATIONS OF *DROSOPHILA MELANOGASTER* AND *D. SIMULANS* IN SOUTH AMERICA USING THE TPE REPEATS AS MARKERS**

Geovani T. Ragagnin<sup>1</sup>, Larissa P. Bernardo<sup>2</sup>; Elgion L. S. Loreto<sup>3</sup>

**Contribuição dos autores:**

1: Revisões Bibliográficas; Montagem e estruturação do manuscrito; Manutenção dos estoques de *Drosophila*; Extração de DNA, PCRs, Eletroforese em gel de poliacrilamida, purificação de fragmentos para posterior sequenciamento;

2. Revisões Bibliográficas; Auxílio na redação do manuscrito; Manutenção dos estoques de *Drosophila*; Extração de DNA, PCRs, Eletroforese em gel de poliacrilamida, purificação de fragmentos para posterior sequenciamento;



Short Communication

## Unraveling the evolutionary scenario of the *hobo* element in populations of *Drosophila melanogaster* and *D. simulans* in South America using the TPE repeats as markers

Geovani T. Ragagnin<sup>1</sup>, Larissa P. Bernardo<sup>2</sup> and Elgion L. S. Loreto<sup>3</sup>

<sup>1</sup>Curso de Ciências Biológicas, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil.

<sup>2</sup>Programa de Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil.

<sup>3</sup>Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil.

### Abstract

Transposable elements (TEs) are nucleotide sequences found in most studied genomes. These elements are highly diversified and have a large variation in nucleotide structure and mechanisms of transposition. *hobo* is a member of class II, belonging to *hAT* superfamily, described in *Drosophila melanogaster*, and it presents in its Open Reading Frame, a repetitive region encoding the amino acids threonine-proline-glutamic acid (TPE), which shows variability in the number of repeats in some regions of the world. Due to this variability some evolutionary scenarios of the *hobo* element are discussed, such as the scenario of the invasion of *hobo* element in populations of *D. melanogaster*. In the present study, we investigated 22 DNA sequences of *D. melanogaster* and seven sequences of *D. simulans*, both from South America, to check the number of repetitions of TPE, in order to clarify the evolutionary scenario of the *hobo* element in these populations. Our results showed a monomorphism in populations of both species in South America, with only three TPE repeats. Hence, we discuss and propose an evolutionary scenario of the invasion of the *hobo* element in populations of *D. melanogaster* and *D. simulans*.

**Keywords:** transposable elements, transposons, hAT, genomic evolution, molecular evolution.

Received: February 17, 2015; Accepted: August 11, 2015.

Transposable Elements (TEs) are sequences of DNA that have the intrinsic ability to move in the host genome (Biémont and Vieira, 2006). This ability gives TEs a wide distribution in the genome of eukaryotes. In humans, 45% of genetic material is composed of TEs, while in *Drosophila melanogaster* about 20% and 80% of genome in maize (International Human Genome Sequencing Consortium, 2001; Meyers *et al.*, 2001; Biémont and Vieira, 2006). Furthermore, the TEs can be involved in events of Horizontal Transfer (HT), a process by which the genetic material is transferred to another who is not his or her descendant (Capy *et al.*, 1994).

TEs are considered mutagenic agents causing structural and regulatory changes in genes. Also, they promote chromosome rearrangements and epigenetic modifications (Biémont 2010; Hua-Van *et al.*, 2011). TEs show a great diversity in nucleotide sequence and structural organization and are classified based in these characteristics. The first

classification was proposed by Finnegan (1989) who divided TEs into two major groups, using the transposition mechanisms and the type of intermediary in the transposition process as criteria. Class I, or retrotransposons, utilize RNA as intermediary and use a mechanism called “copy-and-paste” to replicate. Class II, or transposons, use DNA as intermediary and use a mechanism called “cut-and-paste”. Wicker *et al.* (2007) proposed a new classification system, retaining the two described classes and creating new inclusive categories such as subclass, order, superfamily, family and subfamily based on sequence similarities and structural relationships.

The superfamily *hAT* belongs to class II. Its name is given by the initial of the first elements described and included in this superfamily: *hobo* from *D. melanogaster*, *Ac* from maize and *TAM 3* from *Antirrhinum majus* (Warren *et al.*, 1994). These elements have characteristics in common, such as the presence of terminal inverted repeats (TIRs), generally 12 base pairs (bp) long and produce a duplication of the target site (TSDs) of eight bp. The *hAT* superfamily is old and their wide occurrence in higher taxonomic

Send correspondende to Elgion L.S. Loreto, Departamento de Bioquímica e Biologia Molecular, Universidade Federal de Santa Maria (UFSM), Av. Roraima 1000, Campus Camobi, Predio 16-A, 97105-900 Santa Maria, RS, Brazil. E-mail: [elgion@base.ufsm.br](mailto:elgion@base.ufsm.br)

groups (animals, plants and fungi) suggests the results of many HT events (Rubin *et al.*, 2001).

The *hobo* element described in *D. melanogaster* by McGinnis *et al.* (1983) is observed in three different forms in the host genome: i) as a canonical and autonomous element, ii) canonical elements with internal deletion, and iii) the “relic” elements that are degenerated copies (Ortiz and Loreto, 2008). Evidence has shown that the autonomous elements as well as those sequences showing internal deletions are recent in the genomes, but the “relics” are old and probably the result of ancient invasions (Periquet *et al.*, 1994; Simmons *et al.*, 1998; Galindo *et al.*, 2001; Torres *et al.*, 2006).

*hobo* has in its Open Reading Frame (ORF) a repetitive sequence ACTCCAGAA that encodes the threonine-proline-glutamic acid amino acids (TPE) and presents variability in the number of copies of this repetitions (Calvi *et al.*, 1991; Bazin and Higuete, 1996; Souames *et al.*, 2003a, b). The majority of natural populations of *D. melanogaster* are monomorphic, with three copies of repetitions, while others are polymorphic. These are found in the center of Europe, in equatorial Africa and western South America, with differences in the number of copies, ranging from three to nine repetitions (Bazin and Higuete, 1996; Bonnivard *et al.*, 2000). A temporal and geographic distribution pattern relative to TPE copy number have been observed in the populations of *D. melanogaster*. From this pattern, some evolutionary scenarios are proposed. For example, it is suggested that the first invasion of *hobo* elements in populations of *D. melanogaster*, throughout the world, was made by elements containing three repeats (Periquet *et al.*, 1989), followed by another invasion of *hobo* elements containing five to seven repetitions of TPE (Bonnivard *et al.*, 2000; Souames *et al.*, 2003a, b). The TPE region of *hobo* elements also appears to be related to the invasive ability of the element. Comparative studies with transgenic strains of *D. melanogaster*, containing 3 or 5 TPE repeats suggests that the TPE region has a relationship with transposition or activation of *hobo* elements, such that copies of *hobo* elements containing three TPE increased the transposition activity in the host genome compared to copies containing five repetitions (Souames *et al.*, 2003a).

The *hobo* element has an interesting history in relation to this pattern. Research with old strains of *D. melanogaster* of South America (collected before 1950) showed that the canonical element was absent in these populations, while recent strains present the element. This suggests that the element has been introduced in *D. melanogaster* by HT (Periquet *et al.*, 1989; Boussy and Daniels, 1991; Pascual and Periquet, 1991). Streck *et al.* (1986) and Boussy and Daniels (1991) described the *hobo* element in *D. simulans*, and the evolutionary history of the element in this species is quite similar to the history in *D. melanogaster*. However, these authors suggest that *D. simulans* may have been the donor of the element to *D.*

*melanogaster*, as the *hobo* element is similar between the two species, and yet it apparently was present in *D. simulans* before *D. melanogaster*.

Thus, using the TPE region of the *hobo* element as a marker, the objective of this study was to investigate the polymorphism in populations of *D. melanogaster* in South America, trying to establish a geographic distribution scenario of the element in these populations. We also analyzed the TPE regions in Brazilian populations of *D. simulans* trying to determine the monomorphic condition of three repetitions of TPE, as supposed by the invasion theory of *hobo* elements in *D. melanogaster* from *D. simulans*.

We utilized populations of *D. melanogaster* and *D. simulans* from different locations in South America and Europe (see Table 1 and Figure 2). European populations were used as a control since they are known to possess polymorphisms in the TPE region. The polymorphism analyses were made by Sanger DNA sequencing and polyacrylamide gel electrophoresis (PAGE). DNA extraction was performed in approximately 15 individuals of each population studied according to the protocol of Oliveira *et al.* (2009).

The polymerase chain reaction (PCR) was performed under the following conditions: 35 cycles of 15 s at 94°C, 15 s at 53.5°C, and 15 s at 72°C, using the primers TPEf -5'TGCATTTACAAAATTCAAGTCC, TPEr -5'CTTCCCAAAGCCAGTA. These primers were designed from the sequence of the canonical *hobo* element (GenBank access number X04705), using the Primer3 software. This primer pair amplified the fragment of approximately 240 bp over-passing the TPE region.

The products obtained by the PCR reaction were precipitated in 13% polyethylene glycol (PEG8000) and NaCl (1.6 M) and then sequenced directly using a MegaBACE 500 automated sequencer and the DYEnamic ET® kit (Amersham) according to the manufacturer's protocol.

The sizes of PCR fragments were also estimated by non-denaturing 7% PAGE performed according to the protocol described by Sanguinetti *et al.* (1994). These gels measured 21x23 cm with a thickness of 2 mm and electrophoresis was done at 50V for 12 h. The DNA bands were detected using silver staining following the protocol described by Bassam and Gresshoff (2007).

For 22 populations of *D. melanogaster* we sequenced the PCR product containing the TPE region. This showed that all South American populations had three repeats of TPE, while two European populations, Prunay and La Broche, had five. The other two European populations had three repeats (Table 1). Eleven of the populations that had the TPE region sequenced were also analyzed by PAGE. Two Brazilian populations that were analyzed by PAGE only showed the band characteristic of three TPE repeats (Figure 1). These data suggest that the South American populations of *D. melanogaster* were monomorphic for three TPE repeats in *hobo* transposase. In addition, Table 1



**Table 1** - Table showing the number of repetitions of TPE in populations of *D. melanogaster* and *D. simulans* in South America and the United States, including four European populations.

Population	Species	Collection site; (State / Province); country	TPE repeats	Method used to obtain the number of repetitions of TPE
1	<i>D. melanogaster</i>	Rio de Janeiro; (RJ); Brazil	3	Sequencing
2	<i>D. melanogaster</i>	Erechim; (RS); Brazil	3	Sequencing and PAGE
3	<i>D. melanogaster</i>	São José do Rio Preto; (SP); Brazil	3	PAGE
4	<i>D. melanogaster</i>	São José do Rio Preto; (SP); Brazil	3	PAGE
5	<i>D. melanogaster</i>	São José do Rio Preto; (SP); Brazil	3	Sequencing and PAGE
6	<i>D. melanogaster</i>	Santa Maria; (RS); Brazil	3	Sequencing
7	<i>D. melanogaster</i>	Porto Alegre; (RS); Brazil	3	Sequencing and PAGE
8	<i>D. melanogaster</i>	Rio Grande; (RS); Brazil	3	Sequencing and PAGE
9	<i>D. melanogaster</i>	Tangará da Serra; (MT); Brazil	3	Sequencing and PAGE
10	<i>D. melanogaster</i>	Jari; (RS); Brazil	3	Sequencing and PAGE
11	<i>D. melanogaster</i>	Serra da Capivara; (PI); Brazil	3	Sequencing
12	<i>D. melanogaster</i>	Manaus; (AM); Brazil	3	Sequencing and PAGE
13	<i>D. melanogaster</i>	Fernando de Noronha; (PE); Brazil	3	Sequencing
14	<i>D. melanogaster</i>	Fernando de Noronha; (PE); Brazil	3	Sequencing
15	<i>D. melanogaster</i>	Recife; (PE); Brazil	3	Sequencing
16	<i>D. melanogaster</i>	Ilha Governador; (RJ); Brazil	3	Sequencing
17	<i>D. melanogaster</i>	Santa Teresa; (RJ); Brazil	3	Sequencing
18	<i>D. melanogaster</i>	Venezuela	3	Sequencing
19	<i>D. melanogaster</i>	Vienna; Austria	3	Sequencing and PAGE
20	<i>D. melanogaster</i>	Athens; (Attica); Greece	3	Sequencing and PAGE
21	<i>D. melanogaster</i>	Prunay; (Marne); France	5	Sequencing and PAGE
22	<i>D. melanogaster</i>	La Broche; (Upper Normandy); France	5	Sequencing and PAGE
23	<i>D. simulans</i>	Eldorado do Sul; (RS); Brazil	3	PAGE
24	<i>D. simulans</i>	Farrapos; (Río Negro); Uruguay	3	PAGE
25	<i>D. simulans</i>	Fernando de Noronha; (PE); Brazil	3	PAGE
26	<i>D. simulans</i>	Santa Maria; (RS); Brazil	3	PAGE
27	<i>D. simulans</i>	Lima; (Lima); Peru	3	PAGE
28	<i>D. simulans</i>	Havaí; (HI); United States	3	PAGE
29	<i>D. simulans</i>	Rio de Janeiro; (RJ); Brazil	3	PAGE

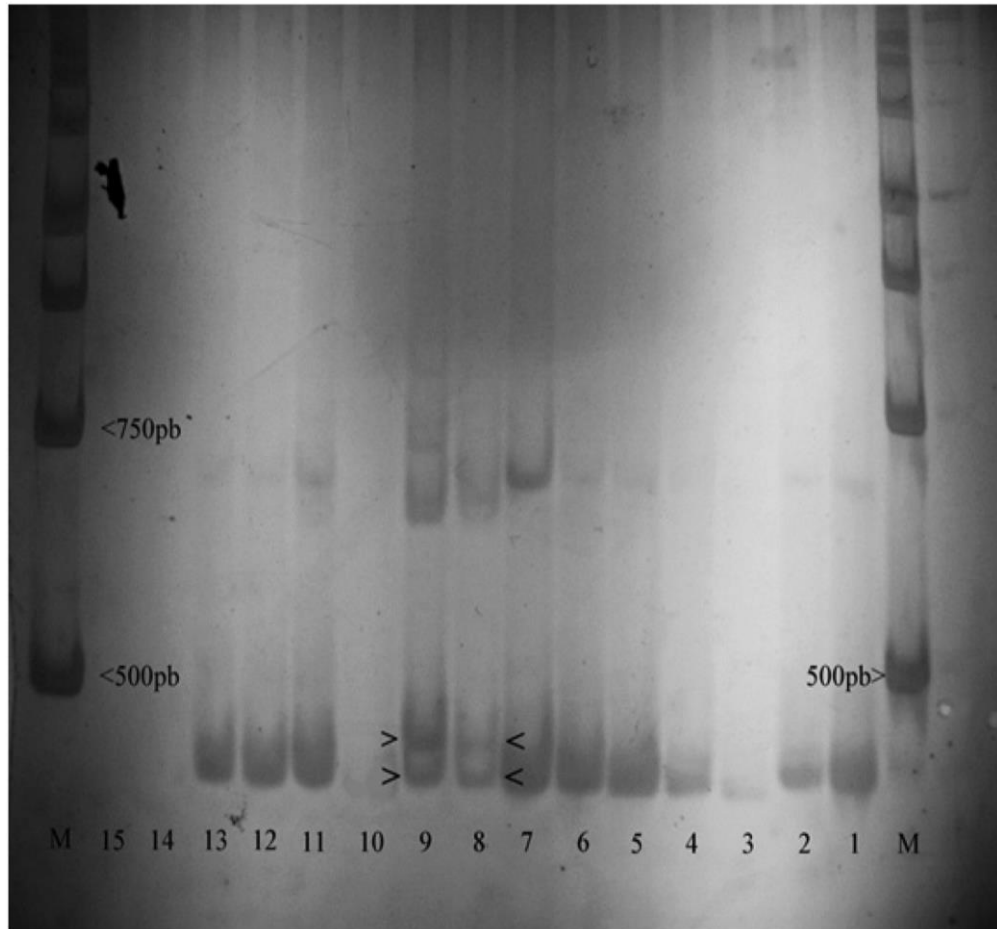
TPE Repeats: Repeats of TPE found in populations; Method used for obtaining the number of repetitions of TPE: Result obtained by 7% Polyacrylamide Gel Electrophoresis and/or Sanger Sequencing.

shows the origins (collection sites) of the studied populations, the number of TPE repetitions and methods that were used for determining the number of repetitions. The geographic distribution of the South American population sampled is depicted in Figure 2.

Seven populations of *D. simulans* were analyzed by PAGE electrophoresis and all showed only the band corresponding to three TPE repetitions (Table 1). The populations of *D. simulans* were similar in their band patterns with Brazilian populations of *D. melanogaster* and showed no polymorphism. The geographic distribution of these studied populations can also be found in Figure 2.

As proposed by Periquet *et al.* (1989), American populations of *D. melanogaster* did not contain the *hobo* element before the 1950s, being characterized as strains E

(empty). Thereafter, the element was found in American populations and these were characterized as H (*hobo*) strains. Bonnivard *et al.* (2000) and Periquet *et al.* (1989) argued that European populations began to present the *hobo* element since the 1960s, proposing a direction for the invasion of *hobo* element. Periquet *et al.* (1989) suggested that the invasion of the *hobo* element began in America, and from it, the populations carrying the *hobo* element spread all over the world. This element contained three TPE repeats and is commonly found around the world. Bonnivard *et al.* (2000, 2002) added to the evolutionary history of the invasive *hobo* element of *D. melanogaster* the hypothesis of a subsequent invasion of the element containing five or seven TPE repeats. They found polymorphic populations in three regions, central Europe, southern Africa and western



**Figure 1** - Polyacrylamide gel photo showing the differences in migration patterns of bands of European and Brazilian populations of *D. melanogaster*. (M) markers; (< 500 bp) marker band representing sequence of 500 bp; (< 750pb) marker band representing sequence 750pb; (>) Marking two bands presented by the two European populations, representing regions with polymorphism in the TPE region; (15) Application of negative sample; (14) empty well; (13) Jari, Rio Grande do Sul; (12) Athens, Greece; (11) Vienna, Austria; (10) empty well; (9) Prunay, France; (8) La Broche, France; (7) Erechim, Rio Grande do Sul; (6) Manaus, Amazonas; (5) Porto Alegre, Rio Grande do Sul; (4) Tangara da Serra, Mato Grosso; (3) São José do Rio Preto, Sao Paulo, collection point two; (2) São José do Rio Preto, Sao Paulo, collection point three; (1) São José do Rio Preto, Sao Paulo, collection point four.

South America. In North America, Bonnivard *et al.* (2000) studied 13 populations and found only 3 repeats in nine populations and 3 and 5 repeats in other four populations. From this, the authors suggested that the possible origin of this invasion happened in South America, since in North Africa only populations containing three regions were found. However, North Africa is the geographic region that promotes a connection between the European continent and southern Africa, thus the authors excluded these two regions as possible origins of invasion, suggesting that South America would be the region of origin for the second invasion.

Our results argue against the supposition for the origin of the second invasion, since, like the region intermediate between Europe and southern Africa, our study area, in the East of America also had no polymorphism, hence excluding South America as a possible area of origin for the second invasion.

Thus, of the three regions containing polymorphisms, none is a candidate for the origin of the invasion. From this standoff we suggest that the second invasion of the *hobo* element may not have occurred in South America or in equatorial Africa, and the polymorphism found in these regions may be due to new mutations occurring in these regions followed by selection or genetic drift.

Souames *et al.* (2003a) analyzed the mutation rate of TPE repetitions of the *hobo* element during 20 generations of *D. melanogaster* and found that the mutation rate in this region was a hundred times higher when compared to other regions of neutral microsatellite genome of *D. melanogaster*. Thus, we suggest that the polymorphism found in western South America and southern Africa could be the result of mutations that were selected or suffered genetic drift, because none of these places has the potential to be the geographic region of origin of the supposed second invasion.





**Figure 2** - Map of South America with the sampling points of the populations of *D. melanogaster* (green circles, represented by letters) and *D. simulans* (blue triangles, represented by numbers), except for the population from Hawaii. (1) Lima, Peru; (2) Fernando de Noronha, Brazil; (3) Rio de Janeiro, Brazil; (4) Eldorado do Sul, Brazil; (5) Santa Maria, Brazil; (6) Farrapos; Uruguay; (A) Jari, Brazil; (B) Santa Maria, Brazil; (C) Rio Grande, Brazil; (D) Porto Alegre, Brazil; (E) Erechim, Brazil; (F) São José do Rio Preto, Brazil; (G) Ilha do Governador, Brazil; (H) Rio de Janeiro, Brazil; (I) Recife, Brazil; (J) Fernando de Noronha, Brazil; (K) Serra da Capivara, Brazil; (L) Tangará da Serra, Brazil; (M) Manaus, Brazil; (N) Santa Rosa, Venezuela.

Unlike these regions, Central Europe has certain peculiarities not shared by South America and Africa. In Europe, the TPE polymorphism frequency is much higher than in any other region and, in addition, presents many kinds of polymorphisms with copies ranging from three repeats and others containing up to ten repetitions (Bonnivard *et al.*, 2002). Thus it is possible that this polymorphism is a local one generated by new mutations.

The monomorphism found in *D. melanogaster* in the TPE regions was shared by *D. simulans*. So far, the TPE region of *D. simulans* has not been widely studied, but for this current study it is important to discuss the supposed donor of the *hobo* element in *D. melanogaster*, the sister species, *D. simulans*. Boussy and Daniels (1991) indicated a direction of the origin of the *hobo* element in *D. melanogaster*, indicating that the donor species of the element would be the sister species *D. simulans*, and this direction of invasion is supported for two reasons: *hobo* would be present in *D. simulans* before *D. melanogaster* and the high similarity of

the element between the two species. Periquet *et al.* (1989) and Bonnivard *et al.* (2000, 2002) argued that the invasion of the *hobo* element in populations of *D. melanogaster* happened in the 1950s in Central America and the *hobo* element had three repetitions of TPE, so for *D. simulans* to be the donor element it would also have to present the *hobo* element with three repetitions. Our results supported the hypothesis proposed by Boussy and Daniels (1991) of *D. simulans* as the donor of the *hobo* element to *D. melanogaster* because all populations studied had three replications of TPE (Figure 2 and Table 1). In this way we can infer with more clarity that *D. simulans* was possibly the donor of the *hobo* element to *D. melanogaster*.

The TPE region has been used by the scientific community because it contains information about the history of the *hobo* element and its dynamics. Based on the TPE repeats, the evolutionary scenario of the *hobo* element was proposed and their history was better understood (Calvi *et al.*, 1991; Bazin and Higuete, 1996; Bonnivard *et al.*, 2000, 2002;). Our results are in accordance with the hypothesis of multiple origins for the second *hobo* invasion, as suggested by Bonnivard *et al.* (2000).

## Acknowledgments

We thank Dr. Jean David, Dr. Claudia Rohde and Dr. Blanche Bitner-Mathe for sending us fly strains. This study was supported by research grants and fellowships from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FAPERGS (Fundação de Amparo a pesquisa do Estado do Rio Grande do Sul).

## References

- Bassam BJ and Gresshoff, PM (2007) Silver staining DNA in polyacrylamide gels. *Nat Protoc* 2:2649-2654.
- Bazin C and Higuete D (1996) Lack of correlation between dysgenic traits in the *hobo* system of hybrid dysgenesis in *Drosophila melanogaster*. *Genet Res* 67:219-226.
- Biémont CA (2010) Brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* 186:1085-1093.
- Biémont C and Vieira C (2006) Junk DNA as an evolutionary force. *Nature* 443:521-524.
- Bonnivard E, Bazin C, Denis B and Higuete D (2000) A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. *Genet Res* 75:13-23.
- Bonnivard E, Bazin C and Higuete D (2002) High polymorphism of TPE repeats within natural populations of *D. melanogaster*: a gradient of the 5TPE *hobo* in Western Europe. *Mol Biol Evol* 19:2277-2284.
- Boussy IA and Daniels SB (1991) *hobo* transposable elements in *Drosophila melanogaster* and *D. simulans*. *Genet Res* 58:27-34.



- Calvi BR, Hong TJ, Findley SD and Gelbart WM (1991) Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator* and *Tam3*. *Cell* 66:465-471.
- Capy P, Anxolabère D and Langin T (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet* 10:7-12.
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5:103-107.
- Galindo MI, Bigot Y, Sánchez MD, Periquet G and Pascual L (2001) Sequences homologous to the *hobo* transposable element in *E* strains of *Drosophila melanogaster*. *Mol Biol Evol* 18:1532-1539.
- Hua-Van A, Le Rouzic A, Boutin TS, File EJ and Capy P (2011). The struggle for life of the genome's selfish architects. *Biol Direct* 6:19.
- International Human Genome Sequencing Consortium (2001) A physical map of the human genome. *Nature* 409:860-921.
- McGinnis W, Shermoen AW and Beckendorf S (1983) A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* 34:75-84.
- Meyers BC, Tingey SV and Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res* 11:1660-1676.
- Oliveira LFV, Wallau GL and Loreto LSV (2009) Isolation of high quality DNA: a protocol combining "rennet" and glass milk. *Electr J Biotechnol* 12:1-6.
- Ortiz MF and Loreto ELS (2008) The *hobo*-related elements in the *melanogaster* species group. *Genet Res* 90:243-252.
- Pascual L and Periquet G (1991) Distribution of *hobo* transposable elements in natural populations of *Drosophila melanogaster*. *Mol Biol Evol* 8:282-296.
- Periquet G, Hamelin MH, Bigot Y and Lepissier A (1989) Geographical and historical patterns of distribution of *hobo* elements in *Drosophila melanogaster* populations. *J Evol Biol* 2:223-229.
- Periquet G, Lemeunier F, Bigot Y, Hamelin MH, Bazin C, Ladevèze V, Eeken J, Galindo MI, Pascual L and Boussy I (1994) The evolutionary genetics of the *hobo* transposable elements in *Drosophila melanogaster* complex. *Genetica* 93:79-90.
- Rubin E, Lithwick G and Levy AA (2001) Structure and evolution of the *hAT* transposon superfamily. *Genetics* 158:949-957.
- Sanguinetti CJ, Dias Neto E and Simpson AJG (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* 17:914-921.
- Simmons GM, Plummer D, Simon A, Boussy IA, Frantsve J and Itoh M (1998) Horizontal and vertical transmission of *hobo*-related sequences between *Drosophila melanogaster* e *Drosophila simulans*. In: Syvanen M and Kado CI (eds), *Horizontal Gene Transfer*. Chapman and Hall, New York, pp 285-294.
- Souames S, Bazin C, Bonnivard E and Higuete D (2003a) Behavior of the *hobo* transposable element with regard to TPE repeats in transgenic lines of *Drosophila melanogaster*. *Mol Biol Evol* 20:2055-2066.
- Souames S, Bonnivard E, Bazin C and Higuete D (2003b) High mutation rate of TPE repeats: a microsatellite in the putative transposase of the *hobo* element in *Drosophila melanogaster*. *Mol Biol Evol* 20:1826-1832.
- Streck RD, MacGaffey JE and Beckendorf SK (1986) The structure of *hobo* transposable elements and their insertion sites. *EMBO J* 5:3615-3623.
- Torres FP, Fonte LFM, Valente VLS and Loreto ELS (2006) Mobilization of a *hobo*-related sequence in the genome of *Drosophila simulans*. *Genetica* 126:101-110.
- Warren WD, Atkinson PW and O'Brochta D (1994) The Hermes transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet Res* 64:87-97.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973-982.

Associate Editor: Louis Bernard Klaczko

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.



#### **4. (ARTIGO 2) A NEW SOURCE OF MIRNA: MINISATELLITES CARRIED BY A HAT TRANSPOSON IN *DROSOPHILA WILLISTONI*.**

Larissa P. Bernardo<sup>1</sup>; Pedro M. Fonseca<sup>2</sup>; Elgion L. S. Loreto<sup>3</sup>

##### **Contribuição dos autores:**

1: Revisões Bibliográficas; Montagem e estruturação do manuscrito; Manutenção dos estoques de *Drosophila*; Extração de DNA e RNAs, incluindo os sncRNAs; PCRs; Southern Blot; Purificação de fragmentos para posterior sequenciamento; Clonagens e sequenciamentos.

2. Auxílio na redação do manuscrito; Manutenção dos estoques de *Drosophila*; Extração de DNA e RNAs; PCRs; Southern Blot; Clonagens e sequenciamentos.

## Submissão

Elsevier Editorial System(tm) for Insect  
Biochemistry and Molecular Biology  
Manuscript Draft

Manuscript Number:

Title: A new source of miRNA: minisatellites carried by a hAT transposon  
in *Drosophila willistoni*

Article Type: Full Length Article

Keywords: MITEs; microRNA; transposable elements; miRNA editing, tandem  
repeats; mir-190.

Corresponding Author: Dr. Elgion da Silva Loreto, Ph.D.

Corresponding Author's Institution: Federal University of Santa Maria

First Author: Elgion da Silva Loreto, Ph.D.

Order of Authors: Elgion da Silva Loreto, Ph.D.; Larissa P Bernardo, Mac

Abstract: The hAT transposable element Howillil of *Drosophila willistoni* possesses long arrays of minisatellites in the subterminal regions. The repeats of the 5' end contain 18 base pairs. Here, we have shown that these repeats are highly polymorphic in different populations of this species, ranging from seven to more than 14 copies. The repeat region is transcribed and produces miRNAs, which have the 18 nucleotides of minisatellites, with the addition of 5 or 9 more by miRNA editing. The miRNA is related to the mir-190 family, and in silico analysis of *Drosophila* genomes suggests seven principal putative genes as targets for this miRNA. This study showed, for the first time, that minisatellites occurring in transposable elements can be a source of miRNA, amplifying the role of transposable elements in the genesis of micro-RNAs.

## Carta ao Editor

Santa Maria, August, 17, 2017.

Dear

**Dr. M. Jindra**

*INSECT BIOCHEMISTRY AND MOLECULAR BIOLOGY*

Dear editor,

We would like to submit the enclosed manuscript entitled "A new source of miRNA: minisatellites carried by a *hAT* transposon in *Drosophila willistoni*", which we wish to be considered for publication in *Insect Biochemistry and Molecular Biology*. Your efforts in the review process of this manuscript are greatly appreciated.

This paper is new. Neither the entire paper nor any part of its content has been published or has been accepted elsewhere. It is not being submitted to any other journal. We believe the paper may be of particular interest to the readers of *Insect Biochemistry and Molecular Biology*. Additionally, inform that this manuscript has been proofread by PRS.

Once again, thank you for your consideration and attention on our manuscript.

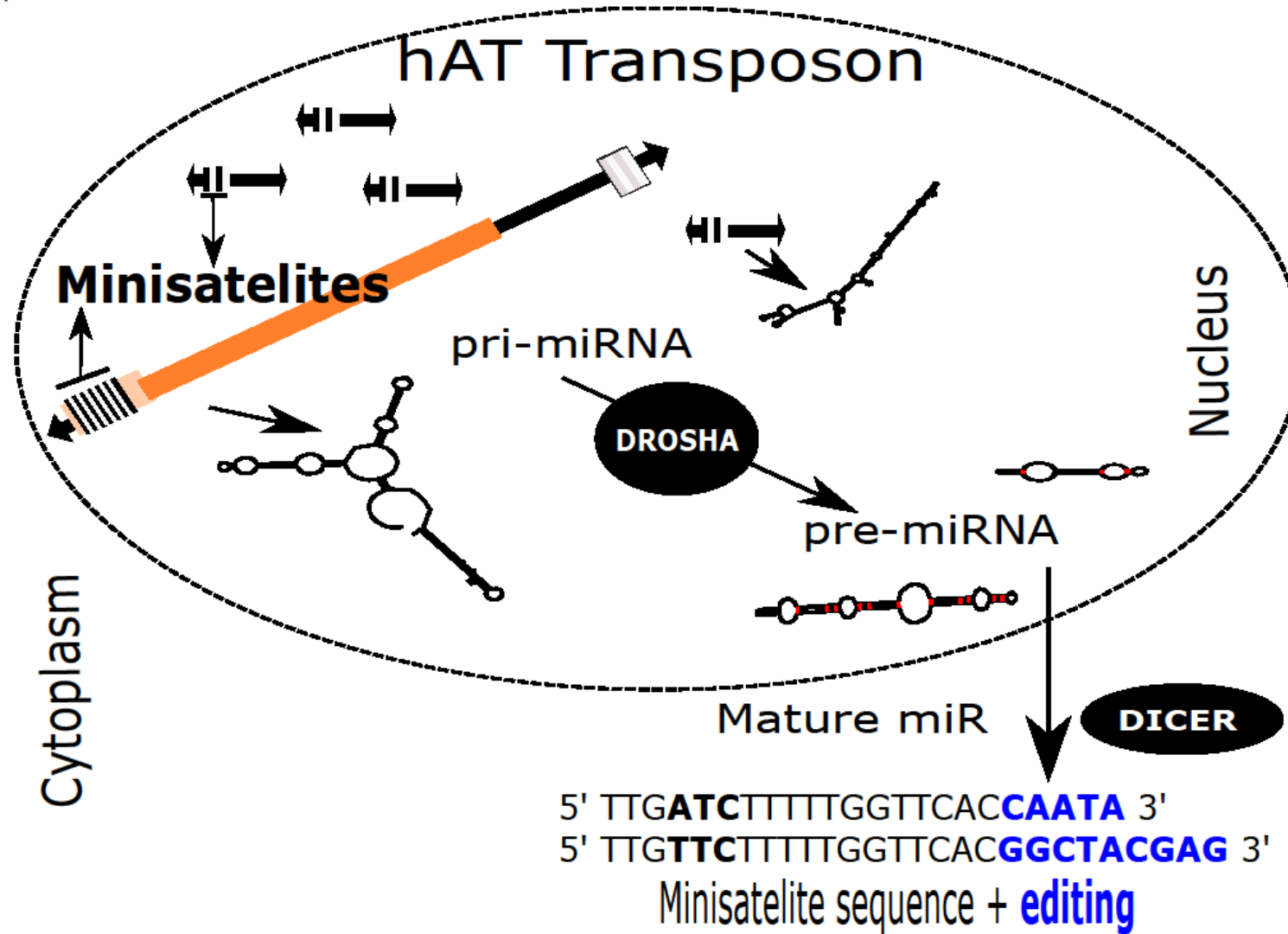
Sincerely yours,

*Elgion L S Loreto*

Dep Biochemistry and Molecular Biology – PPG Animal Biodiversity

## Graphic Abstract

Graphical Abstract



## **A new source of miRNA: minisatellites carried by a *hAT* transposon in *Drosophila willistoni*.**

**Larissa P. Bernardo (1); Pedro M. Fonseca (2); Elgion L. S. Loreto (1,3)\***

1 PPG Biodiversidade Animal, Univ. Fed. de Santa Maria (UFSM), Santa Maria, Brazil;

2 PPG Genética e Biologia Molecular, Univ. Fed. do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil; 3 Department of Biochemistry and Molecular Biology, CCNE, Univ. Fed. de Santa Maria, Santa Maria, Brazil.

\*Correspondence: Elgion L.S. Loreto; Department of Biochemistry and Molecular Biology, CCNE, Univ. Fed. de Santa Maria, Predio 16-A, 3211; CEP 97105-979, Santa Maria, Brazil. Email: elgion@base.ufsm.br.

### **Abstract**

The *hAT* transposable element *howilli1* of *Drosophila willistoni* possesses long arrays of minisatellites in the subterminal regions. The repeats of the 5' end contain 18 base pairs. Here, we have shown that these repeats are highly polymorphic in different populations of this species, ranging from seven to more than 14 copies. The repeat region is transcribed and produces miRNAs, which have the 18 nucleotides of minisatellites, with the addition of 5 or 9 more by miRNA editing. The miRNA is related to the mir-190 family, and *in silico* analysis of *Drosophila* genomes suggests seven principal putative genes as targets for this miRNA. This study showed, for the first time, that minisatellites occurring in transposable elements can be a source of miRNA, amplifying the role of transposable elements in the genesis of micro-RNAs.

### **Key words**

MITEs; microRNA; transposable elements; miRNA editing, tandem repeats; mir-190.

## Introduction

Eukaryotic genomes are replete with repetitive sequences, which can be classified into three main groups depending on their length, structures and replication methods. Microsatellites are short repetitions ranging from two to six nucleotides (nt), whilst longer repetitions of 10 to 60 nt, are called minisatellites. Transposable elements (TEs), the third group, are DNA sequences that are able to change their position within the genomes, in a process termed transposition. TEs are present in almost all organisms, are highly diverse and can promote genetic variability, acting as modulators of structure, function and evolution of host genomes (Biémont, 2010).

TEs are classified into two groups according to their mechanism of transposition (Finnegan, 1989). Class I are retrotransposons that transpose by an RNA intermediate, whereas class II transpose by a DNA intermediate. In both classes, there are autonomous and non-autonomous elements. The former are able to produce the required enzymes for their own transposition, whilst the non-autonomous elements are degenerate copies of TEs that are mobilized by enzymes produced by autonomous elements. A special group of non-autonomous elements are classified as miniature inverted repeat transposable elements (MITEs), that are thought to be cross-activated by autonomous elements (Jiang *et al.*, 2004). However, TEs are very diverse and were traditionally divided into a hierarchical class system of subordinated groups as order, superfamily, family and subfamily (Arensburger *et al.*, 2016; Piégu *et al.*, 2015; Wicker *et al.*, 2007).

The *hAT* superfamily is a large and diverse group of class II elements, found in different eukaryotes such as plants, animals and fungi. They vary in size from approximately 1.5 - 5.0 kb, have short inverted terminal repeats (TIRs) some 10-25 bp in length, and flanked by 8 bp of a target sequence duplication (TSD) (Arensburger *et al.*, 2011). They are divided into three families, *Ac*, *Buster* and *Tip*, based on the sequence of transposase, and differences in the target-site selection (Arensburger *et al.*, 2011; Rossato *et al.*, 2014).

The transposable element *howilli1* was initially identified in *D. willistoni* through *in silico* searches for *hATs* in genomes of *Drosophila* species (Ortiz & Loreto, 2009). The *D. willistoni* sequenced genome contains two copies of a



complete and potentially-active element of 2,954 bp. Ortiz *et al.*, (2010) showed that there are 30 copies of non-autonomous elements, many of which are MITE derivatives. A remarkable characteristic was found in this element, represented by a long array of minisatellite repeats (MnRs) in both extremities of the element. In the *D. willistoni* sequenced genome, the minisatellites in the 5' sub-terminal region are 18 nucleotides long, ranging from 12 to 20 repeat copies, with small variations in sequence. The MnRs in the 3' sub-terminal region have 28 nucleotides of different sequences to those of the 5' repeats. The same array of minisatellites are present in two related *hAT* elements, *hAT1-Dp* of *Drosophila pseudoobscura* and *hoppers1* of *D. persimilis* (Ortiz *et al.*, 2010). However, the TEs of *D. willistoni*, *D. pseudoobscura* and *D. persimilis* show a general divergence of 0.28, while the MnRs are highly conserved. This raises the question of the biological function of these repeats.

Different mechanisms that minimize the deleterious effect of uncontrolled transposition have evolved to regulate TE mobilization, (Hua-Van *et al.*, 2011). They include a special class of RNA referred as small noncoding RNA (sncRNA) (Huda *et al.*, 2011; Kalmykova *et al.*, 2005; Matzke *et al.*, 2000). Two main classes of sncRNA (small interfering RNAs (siRNAs) and PIWI interacting RNA (piRNA)) are known to protect against deleterious transposition or invasion of foreign DNA in the form of TEs or viruses (Ambros *et al.*, 2003; Aravin *et al.*, 2007; Siomi *et al.*, 2011; Slotkin and Martienssen, 2007). A third class of sncRNA, the microRNAs (miRNA), are probably an exaptation of the sncRNA mechanism of defense against TEs and viruses. They have become one of most important mechanisms of gene regulation in eukaryotes, contributing to the post-transcriptional regulation of almost all cellular processes (Jaubert *et al.*, 2007; Ylla *et al.*, 2016).

The miRNAs differ from other classes of small RNAs in their biogenesis. They are derived from transcripts of up to 1 kb which can form a secondary structure known as a hairpin stem loop (Axtell *et al.*, 2011). The biogenesis from miRNA has three stages: (i) transcription, (ii) maturation, and (iii) incorporation. The transcription stage is performed by RNA polymerase II, which generates pri-miRNAs capable of forming secondary structures. The maturation stage is the processing of pri-miRNA by the DROSHA and DICER enzymes to generate pre-miRNAs which are transported to the cytoplasm, where they generate

mature miRNAs, which are incorporated into the RISC complex serving as a mold for gene silencing. (Fabian & Sonenberg, 2012; Ha & Kim, 2014).

TEs are an important source of small RNA and have contributed substantially to the evolution of new regulatory networks (Qin *et al.*, 2015; Roberts *et al.*, 2014). To date, two sources of TEs have been described for the formation of miRNA: retrotransposon insertion in adjacent positions and in opposite directions, and the MITEs. In both cases, TE transcription involves the formation of a stem-loop which can be processed by DROSHA. Insertion of these same TEs into the 3'UTR region of coding genes can be a source of target sequences for gene regulation (Piriyapongsa & Jordan, 2007; Roberts *et al.*, 2014; Shen *et al.*, 2017).

Considering the characteristics of the long array of minisatellites repeats (MnRs) found in *howilli1*, we hypothesized that these sequences have the potential for formation of a stem-loop hairpin secondary structure, which represents a new way in which TEs act as miRNA precursors.

The objective of this study was to characterize the 5' MnR region of the *howilli1* element, and evaluate the contribution of this region to the synthesis of miRNA. The main results obtained were (i) the number of repeats in the 5' MnR region is highly variable in *D. willistoni*, ranging from seven to more than 14 copies, (ii) the 5'MnR region is transcribed and produces miRNA, (iii) the mature miRNA has the 18 nucleotides of minisatellites, with the addition of 5 or 9 nucleotides by miRNA editing, and (iv) *in silico* analysis of *Drosophila* genomes suggests 38 putative genes as targets for this miRNA, seven of which are principal putative target genes.

## **Experimental procedures**

### **Fly stocks and isolines**

The fifteen strains of *D. willistoni* used in this study are indicated in Supplementary Table 1. The flies were maintained at 20°C in cornmeal medium (Marques *et al.*, 1966). Eight different isofemale lines were established from Aratiba, Serra do Cipó and Tangará strains for the detection of variations in the repeat number of elements in individuals of the same population.

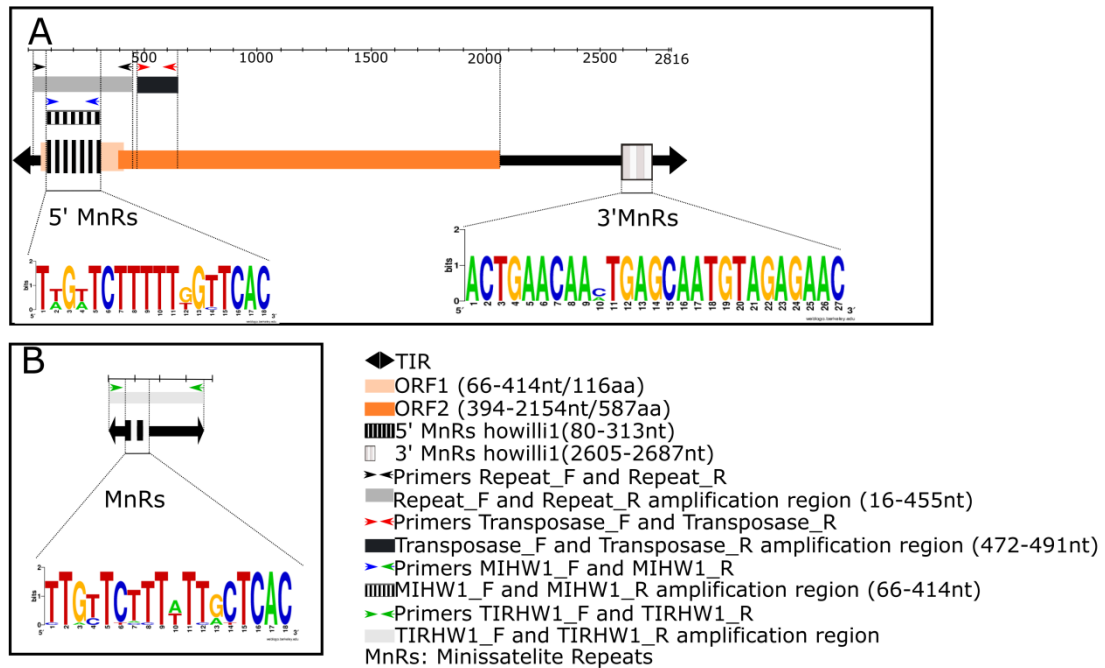
## DNA manipulation and PCR

Genomic DNA was isolated in accordance with the protocol described by Sassi et al., (2005). Polymerase chain reaction (PCR) amplification of *howilli1* elements was performed with primers based on *howilli1* sequences (Ortiz & Loreto, 2009), using Primer3 software (Koressaar & Remm, 2007; Untergasser et al., 2012). Sequences, melting temperature, PCR parameters and annealing positions of the primers are shown in Supplementary Table 2. The concentrations of PCR reagents were: 10 ng genomic DNA; 2.5 mM MgCl<sub>2</sub>, 0.125 mM dNTP, 0.25 pmol primers, 0.2 U Taq DNA polymerase, 1x PCR buffer. PCR amplicons were assessed using agarose (0.8%) gel electrophoresis.

## Polymorphism analysis

Polymorphism analysis of 5'MnRs in *D. willistoni* strains was performed with polyacrylamide gel electrophoresis (PAGE), and by the sequencing of PCR products generated by primers which produce amplicons that overpass the 5' MnRs, and include a small portion of transposase. This avoids the amplification of MITE-*howilli1* elements (Repeats\_F and Transposase\_R primers; Fig1A).

PAGE was performed as described by Sambrook & Russel (2001). The gel dimensions were 21x23cm, with a thickness of 2 mm. An electric potential of 50 V was applied for 20 h. The gels were then stained by silver nitrate in accordance with Bassam & Gresshoff (2007).



**Figure 1** - Schemes of *howilli1* and *MITE-howilli1* elements, with regard to their features and positions of primers used in this study. (A) Characterization of *howilli1* element: ORFs, repeats and TIRs. Primers; Repeats\_F and Repeats\_R=black arrow; Transposase\_F and Transposase\_R=red arrow; MIH1D\_F and MIH1D\_R/ MIH1E\_F and MIH1E\_R primers=blue arrow. (B) *MITE-howilli1* element characterization: TIRH1\_F and TIRH1\_R primers= green arrow. MnRs sequence LOGO observed in the *howilli1* (A) and *MITE-howilli1* element (B). The y-axis shows 2 bit of information, and the x-axis represents the nucleotides in MnRs.

For sequencing, the bands of amplicons of different sizes on the acrylamide gels were excised, and immersed in 100  $\mu$ l water, boiled for 5 min and centrifuged at 13,000 rpm. A new PCR was performed as described above using 1  $\mu$ l of the supernatant of the excised PAGE band as DNA template. The PCR products were precipitated overnight in a solution containing polyethylene glycol 8000 (20%) and NaCl 1.6 M. The samples were then centrifuged, dried at room temperature, resuspended in 20  $\mu$ l of water, and directly sequenced using a MegaBace 500 automatic sequencer. The dideoxy chain-termination reaction was performed using the DYEnamicET kit (GE Healthcare).

### Total RNA extraction, mRNA and miRNA isolation, and cDNA synthesis

Total RNA was extracted from a pool of 20 *D. willistoni* adult flies (Joinville strain), using TRIzol® Reagent (Invitrogen). The samples were treated

with DNase I RNase-Free (Invitrogen) following the manufacturer's instructions, to eliminate DNA contamination. mRNA was isolated from total RNA, using the FastTrack® MAG mRNA Isolation Kit (Invitrogen).

To obtain miRNA, total RNA of *D. willistoni* (Joinville strain) was extracted using TRIzol® Reagent (Invitrogen). The RNA was precipitated overnight at -20°C in sodium acetate (3M; pH 5.2) and absolute ethanol at a volume ratio of 0.1:2.5. The samples were centrifuged at 14,000 g for 30 min at 4°C, air dried at room temperature, and eluted in nuclease-free water. The samples were then quantified using a NanoDrop 2000 spectrophotometer, and total RNA integrity was assessed by gel electrophoresis.

The small RNA (<200 nt) was purified from total RNA using the mirVana miRNA Isolation Kit® (Invitrogen). After isolation, the miRNA was polyadenylated using Poly(A) Polymerase (Invitrogen).

The cDNA synthesis was performed using M-MLV Reverse Transcriptase (Invitrogen) with random primers; miRNA was the exception whereby poly(T) adapter primer was used, as described by Shi & Chiang (2005).

The miRNA amplicons were generated by reaction with the primers MIH1\_F and adapters (Supplementary Table 2). The amplicons were cloned using the TA Cloning Kit (Invitrogen), and sequenced using a MegaBace 500 automatic sequencer (GE Healthcare) and ABI-Prism 3500 Genetic Analyzer (Applied Biosystems). The dideoxy chain-termination reaction was performed using the DYEnamicET kit (GE Healthcare) and BigDye (Applied Biosystems). The obtained clones were named miR-1d and miR-5d.

### **Transcriptional expression of *howilli1***

Transcriptional expression corresponding to repeat region 5'MnRs of *howilli1* element was investigated by RT-PCR, using the following pairs of primers: 1) Repeats\_F and Repeats\_R, 2) Transposase\_F and Transposase\_R, and 3) Repeat\_F and Transposase\_R (Figure 1A and Supplementary Table 2). As a template for RT-PCR, we used cDNA synthesized from total RNA, mRNA and genomic DNA of *D. willistoni* Joinville strains. RT-PCR products were separated by PAGE (0.8%), stained with GelRed® (Biotium), and visualized with Amersham Imager 600 UV. Southern blot was applied to confirm that the

RT-PCR products corresponded to *howilli1* elements. After electrophoresis, the products were transferred to nylon membranes (HybondN+, GE Healthcare). The membranes were hybridized at 60°C with a probe corresponding to the previously cloned and sequenced *howilli1* element. Two probes were used: the first one corresponds mainly to the repeats region and the second one to a section of the transposase region (Figure 4A). Alkphos Direct Labelling Reagents and CPD Star Detection kit (GE Healthcare) were used to label and detect nucleic acids, according to the manufacturer's instructions. Chemiluminescence was detected using the Amersham Imager 600 UV device.

### **Analyses *in silico***

The sequences of putative active *howilli1* and the MITE-*howilli1* elements were analyzed and annotated using UGENE (Okonechnikov *et al.*, 2012). The 5' MnRs were aligned using ClustalW, implemented by MEGA 5.10. The similarities of 5' MnRs were analyzed using WebLogo (Crooks *et al.*, 2004). The MITE-*howilli1* used were *6howilli1*, *7howilli1*, *8howilli1* and *9howilli1*, as described in Ortiz *et al.*, (2010).

Analysis of hairpin formation (required for the production of miRNA precursors) was performed through the evaluation of free energy values of the RNA structures, as per Zhang *et al.*, (2006) and Zuker (2003).

To test the hypothesis that 5' MsR is involved in miRNA formation, three approaches were used: i) search for mature miRNA described in miRbase (Kozomara & Griffiths-Jones, 2014), that shows similarity with the sequences of the 5' MsRs, ii) search for mature miRNA in miRbase that show similar sequences to those isolated and sequenced from *D. willistoni* short RNA, and iii) prediction of putative target-site to putative miRNA. We used sequences generated from the clones of the small RNAs in *D. willistoni* for this analysis. The procedure was implemented in TargetScanFly (Lewis *et al.*, 2005), using the seed sequence (5' UGAUCUU 3') of mir\_1d. The seed sequence of mir\_5d was not used because it is rich in "U". According to Garcia *et al.*, (2011), regions rich in "AU" can generate unstable and imprecise targets.

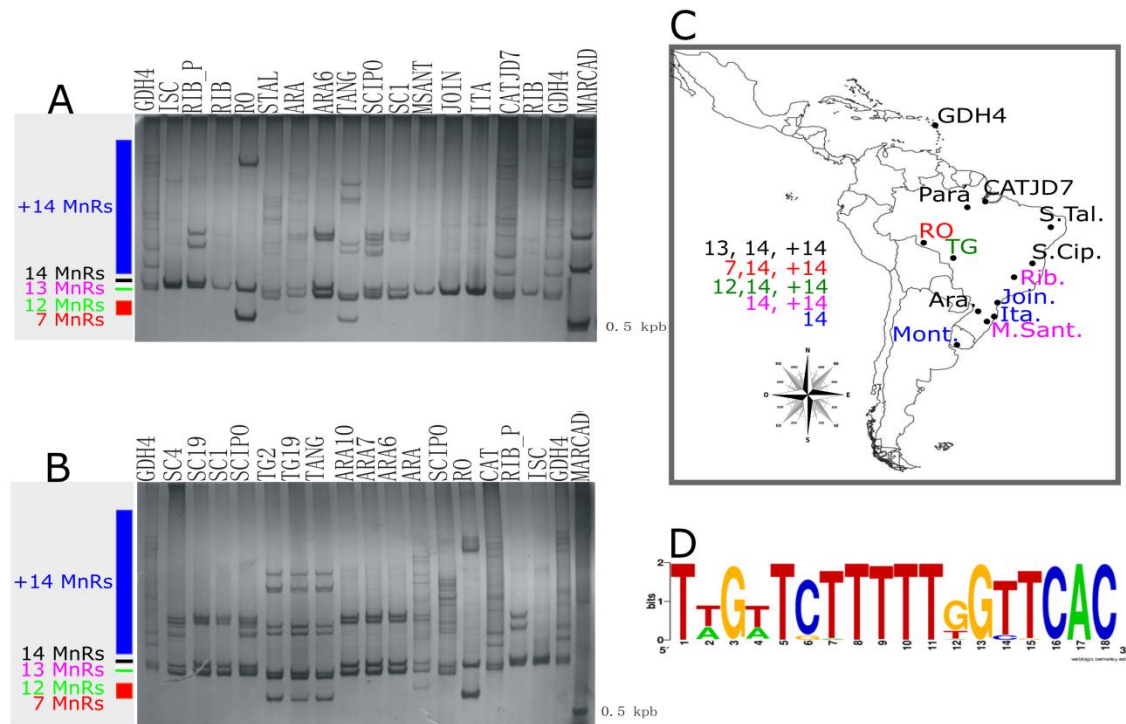
## Results

### Characterization of *D. willistoni howilli1* element and its 5' MnRs polymorphism

The *howilli1* element has two predicted ORFs. ORF1 encompasses the 66–414 nt region, and corresponds mainly to the MnR; this ORF is probably not functional. Although this region is transcribed, translation would result in a protein formed by imperfect repeats of the amino acids C(R)SFW(L)FT. The ORF 2 is in the reading frame +1, and encompasses nucleotides 394 to 2154 encoding the transposase enzyme (Figure 1A). *howilli1* MnRs were found in both extremities of this element. The 5'MnRs encompass nucleotides 80 to 313, whereas the 3'MnRs encompass nucleotides 2605 to 2687. The 5'MnRs are very similar in sequence to minisatellite repeats in the *hAT-1DP* element of *D. persimilis* and *D. pseudoobscura*, but the 3'MnRs are not conserved.

The MITE-*howilli1* element analyzed also contains MnRs, although they show more sequence variability than the MnRs of the *howilli1* element (Figure 1B).

Sequencing and PAGE of PCR amplified segments of *D. willistoni* strains and isofemale lines were used to investigate the polymorphism of repeat number of 5' MnR region (Figure 2A–D). Sequencing showed that *howilli1* elements contained 14 MnRs in four strains: Aratiba, Itapeva, Rondônia and Serra do Cipó; 13 MnRs were found in Aratiba, Erechim, Serra do Cipó, and 6 Aratiba isofemale lines; 12 MnRs were observed only in Tangará strain; and seven MnRs were found in the Rondônia strain. The electropherogram results suggested the presence of sequences greater than 14 repeats. However, we did not obtain reads long enough to assemble these sequences with high confidence.



**Figure 2** - Polyacrylamide gel electrophoresis (PAGE) analysis of polymorphism of number of 5' MnR repeats in *howilli1* element from populations of different geographical origins. Number of repeats ranged from 7 to more than 14. A) Strains sampled: Ilha de Guadalupe (GD-H4), Erechim (Isc), Ribeirão Preto (Rib–Rib\_P), Rondônia (RO), Pernambuco (S\_Tal), Aratiba (Ara), Tangara da Serra\_Mato G. Sul (Tang), Serra Cipó\_Minus Gerais (SC ), Porto Alegre (Msant), Joinville (Join), Itapeva (Ita), CATJD7 (Amazônia), Ribeirão Preto (Rib), MARCAD=0.5 kb ladder. B) Serra Cipó\_Minus Gerais (SC1), Serra Cipó\_Minus Gerais (SC4), Serra Cipó\_Minus Gerais (SC19), Tangara da Serra\_Mato G. Sul (TG2), Tangara da Serra\_Mato G. Sul (TG19), Aratiba (ARA6), Aratiba (ARA7), Aratiba (ARA10) correspond to isofemale lines. C) MnRs sequence LOGO observed in the *howilli1* sequenced in the different populations analyzed. D) Geographic location of collection points of *D. willistoni* strains tested.

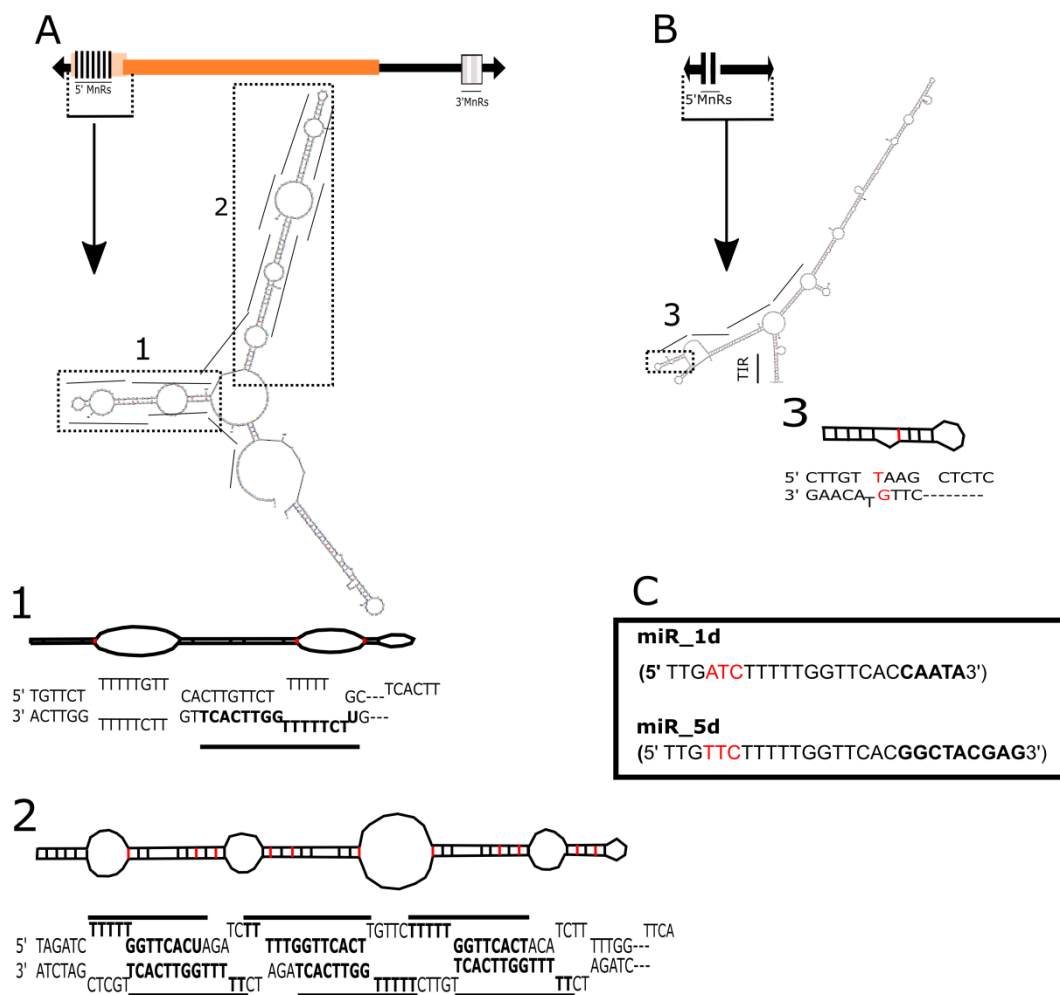
The PAGE analysis revealed MnRs ranging from 7 to more than 14 repeats (Figure 2A and B). Twelve of the strains tested showed MnRs with more than 14 repeats, indicating a very polymorphic nature (Supplementary table 3 and Supplementary archive 1). Only three strains (Itapeva, Montevideo and Joinville) are monomorphic, containing 14 MnR repeats. Polymorphism was also observed in the established isofemale lines. Figure 2B shows that the isofemale lines Serra do Cipó SC1, SC19 and SC4, showed specific bands, that differed from the ancestral strain Serra do Cipó (SCIPO); this suggests a polymorphic population. The same pattern was observed in the isofemale lines Aratiba Ara 10, Ara 7 and Ara 6, and their ancestral strain Aratiba (Ara).

The copy number variations of MnRs are not the unique polymorphism observed. The minisatellites are not perfect in their sequences, showing variations amongst the copies of MnRs (Figure 2D)



## 5'MnRs are transcript and produce miRNAs.

Putative miRNAs were initially identified using criteria based on its expression and biogenesis. These included the prediction of a fold-back precursor structure that contains the ~22 nt miRNA sequence within an arm of a hairpin. We analyzed the secondary structures of portions corresponding to nucleotides 66 to 414 of the *howilli1* element and the MITE-*howilli1*, to determine whether these sequences have stable structures compatible with those observed in miRNA precursors (Figure 3). These characteristics are the long stem-loop structures, higher negative minimal folding free energies (MFEs), and minimal folding free energy indexes (MFEIs), as shown in Table 1.



**Figure 3** - Analysis of the secondary structure using Mfold Web Server. (A) *howilli1* autonomous element. The secondary structure of the potentially transcribed region is shown in detail. The location of cloned miRNAs sequence are highlighted in secondary structure 1 and 2. (B) MITE-*howilli1* secondary structure. 3: The 5'MnRs in the secondary structure are shown in detail.

**Table 1** - MFEI analysis of MITE-*howilli1* sequences and the MnR region of *howilli1* element.

Sequence	Length	% GC	MFE	AMFE	MFEI*
MITE-Herpes-like	335	40.3	-132.9	-39.6716	-0.98441
Sequence 66-414 nt	348	31.61	-46.79	-13.4454	-0.42535

Length in nucleotide

%GC: GC content;

MFE: minimal folding free energy;

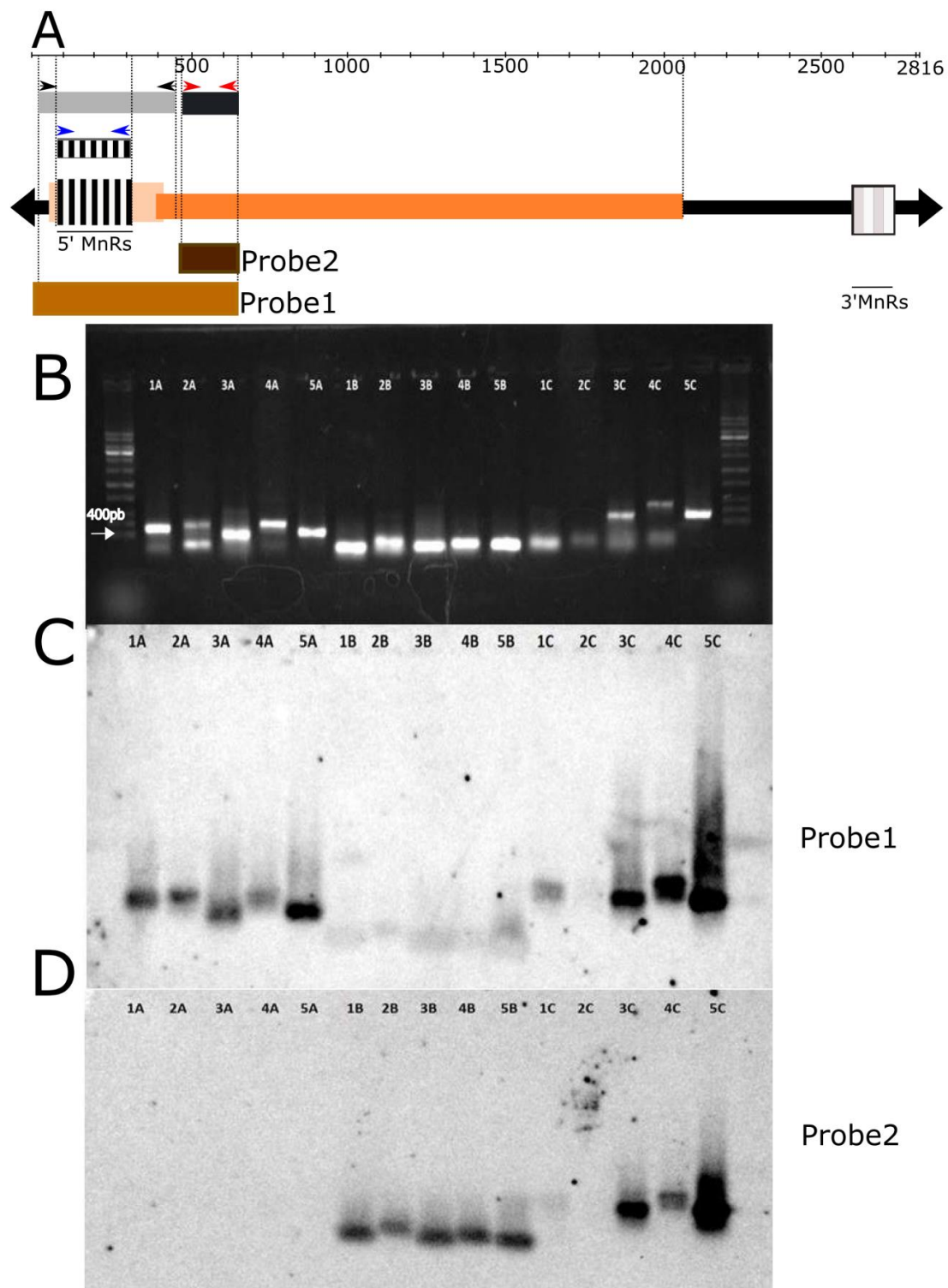
AMFE: adjusted MFE (Representing the MFE of 100 nucleotides);

MFEI: minimal folding free energy index

\*Criterion to distinguish miRNAs from all coding and non-coding RNAs (miRNA was 0.97)

In MITE-*howilli1* elements, the values of these indices were similar to those expected for structures forming hairpins that are necessary for miRNA precursor. In nucleotides 66 to 414 of the *howilli1* element, these indices showed low MFEIs. However, the secondary structure formed from this region generates a fold-back precursor that contains the potential miRNA sequence within one arm of the hairpin.

The second criterion for the prediction of the miRNA was the transcription detection. The RT-PCR analysis showed that *howilli1* is transcriptionally active in *D. willistoni*. Amplicons were obtained from cDNA using three pairs of primers, corresponding to repeat regions, partial region of transposase and the amplification of both regions. Amplifications were obtained for cDNA synthesized from total RNA isolated and mRNA (Fig. 4A–D and Supplementary table 1). No amplification occurred when total RNA or mRNA was used as the template, indicating that the RNA samples were not contaminated with genomic DNA.

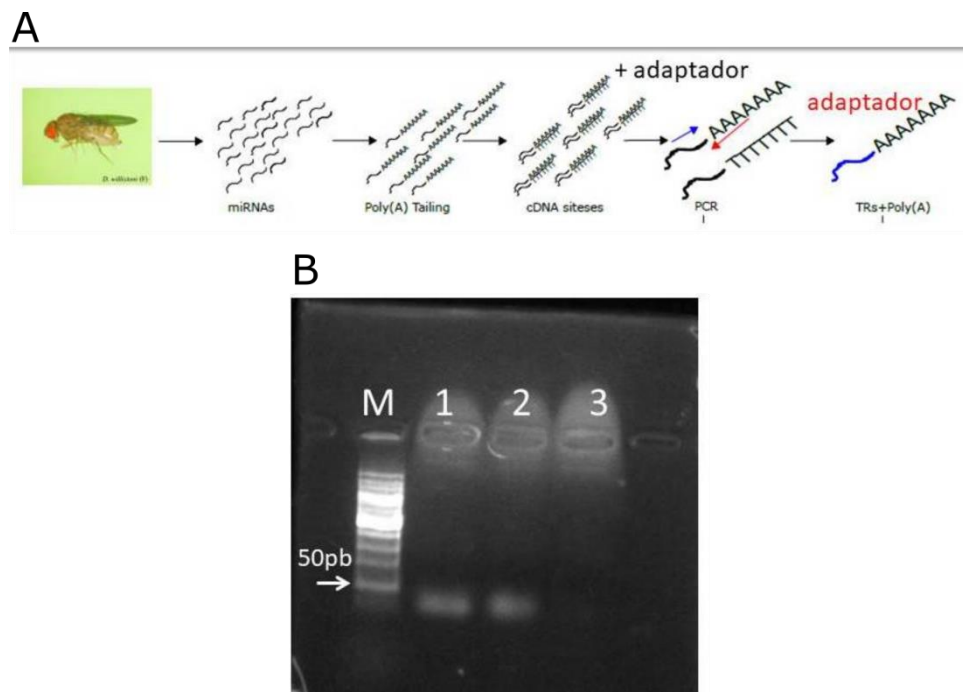


**Figure 4** - Transcriptional analysis of *howilli1* by RT PCR. A) Scheme of *howilli1* element and the regions of alignment of primers used in RT PCR, and the production of the probe for Southern blot analyzes; B) Agarose gel electrophoresis of RT PCR amplicons; (C and D) Southern blot hybridizations using the probe 1 (C) and probe 2 (D). The samples correspond to 1-2 (A-C) cDNAs synthesized from isolated mRNAs; 3-4 (A-C) cDNAs synthesized from total RNA; 5 (A-C) genomic DNA from *D. willistoni*, used as amplification control. Samples identified as A were amplified using the pairs of primers for 5' MnRs region. Samples identified as B were amplified using pairs of primers that amplify part of the transposase region. Samples identified as C correspond to amplification using primers for regions of 5' MnRs, plus the partial transposase sequence.

The amplified RT-PCR products were transferred to membranes and hybridized with probes corresponding to different portions of the element, obtained by PCR from the previously cloned and sequenced *howilli1* element. Southern blotting showed that the amplicons exhibit similarity with *howilli1* and do not correspond to spurious amplifications.

Based on transcription of the portion containing the MnRs, we attempted to isolate and sequence miRNA that contain the MnR sequences, using *D. willistoni* cDNA obtained from isolated small RNAs, and amplified with specific primers (Figure 5). These amplicons were cloned and sequenced, and demonstrated that 5'MnRs sequences are present in the portion of small RNAs studied. Two sequences were obtained in this analysis, one of which is 23 bp, and termed mir\_1d (5' TTGATCTTTTTGGTTCACCAATA 3'). The second has 27 bp, and was termed miR-5d (5' TTGTTCTTTTTGGTTCACGGCTACGAG 3').

The sequences obtained were similar to the MnR sequence hairpins in the predicted secondary structure (Figure 3A). However, nucleotide variations were observed at the 3' ends of the miRNAs as indicated in Figure 3C. These variations were not observed in sequences predicted in the *in silico* analysis of 5' MnRs of the *howilli1* element. In addition, these sequences were not found in the sequenced genome of *D. willistoni*.



**Figure 5** - Result of miRNAs amplification. A) miRNAs isolation pipeline; B) Electrophoresis result. The M is a marker in ladder of 50pb. The number one is the sample of cDNAs synthesized from polyadenylated miRNA samples. The number two is the sample of cDNAs synthesized from non-polyadenylated miRNA samples. The number three is the control sample without cDNA.

We used the mir\_1d and mir\_5d sequences to search for similar mature miRNA in the miRbase. Nineteen miRNAs showing high similarities with the mir\_1d were identified. Of these, thirteen of these were similar to the MnRs belonging to the miR-190 miRNA family (Supplementary table 4). Four other miRNAs in the data bank were similar in seed sequence to mir\_1d: gma-miR9725 (*Glycine max*), mmu-miR-6951-3p (*Mus musculus*), sbi-miR6232a-3p (*Sorghum bicolor*) and stu-miR8011b-3p (*Solanum tuberosum*) (Supplementary table 4). The miR-5d clones were similar to seven miRNAs in the miRbase, but none of them were similar to the seed sequence region of the miRNAs in the database (Supplementary table 4).

### **Prediction of target genes**

The prediction of target genes was performed using the seed sequence of miR\_1d (5 'UGAUCUU 3'). This sequence was elected based on the similarity of seed sequences of pre-existing miRNAs (gma-miR9725 (*Glycine max*), mmu-miR-6951-3p (*Mus musculus*), sbi-miR6232a-3p (*Sorghum bicolor*) and stu-miR8011b-3p (*Solanum tuberosum*) miRNAs (Supplementary table 5). These analyzes revealed 38 conserved targets, with a total of 39 conserved sites (Supplementary table 5). The target predicted genes are mainly constitutive and developmental expression genes. Using a more stringent analysis of restringing to conserved target sites with 8mer, seven genes were found. These genes are involved in very important biological processes such as oogenesis, formation of constitutive proteins of membrane, detection of pheromones, and sensory perception of chemical stimuli (Supplementary table 5).

### **Discussion**

Repetitive sequences in TEs are involved in the regulation of many different steps in transposon "life cycles". For example, TPE repeats from *hobo* are related to its transposition efficiency (Souames et al., 2003). Tandem repeats (TRs) found in TIRs of the *Galileu* element and in the 3' end of retrotransposon *Ty3/Gypsy*, act as transposition termination signals by forming

stable stem-loops (Martinez-Izquierdo *et al.*, 1997; Marzo *et al.*, 2013). The TRs of the *DINE-TR1/Helitron* element is thought to modify gene expression and participate in epigenetic processes (Dias *et al.*, 2016).

The *howilli1* element and its MITE derivatives have numerous arrays of MnRs. These repeats are not usually found in transposable elements (Klein *et al.*, 2015). We have shown, for the first time, that minisatellites carried by these TEs are a source of miRNAs that may be involved in gene regulation, and constitutes a new kind of function for the TR of transposable elements.

Ambros *et al.*, (2003) suggested that miRNAs can be identified using a combination of criteria relating to their expression and biogenesis. We investigated the role of 5' MnRs of the *howilli1* element and their MITEs derivatives, as potential precursors of miRNAs. Using the expression criterion, we showed that the 5'MnRs region is transcriptionally active. We also analyzed certain biogenesis criteria and found that 5'MnRs of *howilli1* and *MITE-howilli1* sequences can generate hairpin and secondary structures, that have lengths and forms that are characteristic of pri-miRNAs (Zhang *et al.*, 2006). Additionally, the MFEI values of the predicted hairpins that are necessary for miRNAs formation, are in agreement with Zhang *et al.*, (2006). The small RNAs of *D. willistoni* were cloned and sequenced after amplification using specific primers for MnRs. The obtained miRNAs consist of 23 and 28 nucleotides, and in both sequences the first 18 nucleotides (5') were identical to the MnRs of the analyzed TEs. The 5 and 9 remaining nucleotides of the 3'end were unknown sequences, and the "complete" sequences of these miRNAs were not identified in the sequenced *D. willistoni* genome. Newman *et al.*, (2011) demonstrated the editing of miRNA, which involves the addition of nontemplate nucleotide to the 3' end, as found in many miRNA families. Similar findings were observed in miRNAs generated by the MnRs of *howilli1*.

Usually, mature miRNAs consist of 21 to 24 nucleotides but shorter or longer elements are sometimes identified (Ambros *et al.*, 2003; Liu *et al.*, 2013). The length may vary depending on the development phase analyzed, the tissue where the expression occurs and the 'age' of the miRNA; young ones undergo more inaccurate excisions, which affects their length (Ma *et al.*, 2010). Also, alternative processing of miRNA and miRNA editing can occur (Kim *et al.*, 2017; Newman *et al.*, 2011). The variations in sequence or length of a mature miRNA

are termed isomiRNAs or isomiRs, and are isoforms of microRNAs associated with differential gene regulation and important physiologic consequences (Kulcheski *et al.*, 2011; Zhang *et al.*, 2016). While the two miRNAs isolated in this study seem to be generated by miRNA editing, we suggest that the high level of polymorphism observed in *howilli1* MnR copy number and sequence variation, is a putative source of isomiRs.

Transposable elements are recognized as an important source of miRNA (Roberts *et al.*, 2014). MITEs are important in this process due to their structural characteristics which, due the proximity of TIRs, allow the generated transcripts to produce the hairpins necessary for miRNA processing (Piriyapongsa and Jordan, 2007; Qin *et al.*, 2015; Shen *et al.*, 2017). However, in *howilli1* and MITE-*howilli1*, the hairpins are generated by the minisatellites of the elements. Nevertheless, a great number of MITE-*howilli1* elements in the *D. willistoni* genome are associated with the spread of target sequences in several genes under the control of miRNAs generated from *howilli1* minisatellites.

miRNAs similar to the MnRs of have been identified in animals and plants. Many of these miRNAs belong to the *miR-190* family, which is widely distributed in animals (Zheng *et al.*, 2010). In *Drosophila*, the miR-190 is involved in hypoxic response. The overexpression of miR-190 enhances hypoxia inducible factors (HIF), and in embryos with loss of function of this miRNA, the hypoxic response was observed to be impaired (De Lella Ezcurra *et al.*, 2016). However, the prediction of target genes suggests that the miRNAs generated by MnRs of *howilli1* are involved in variable cellular processes, and experimental assays are necessary to address this point.

In summary, the *hAT* element *howilli1* and its MITE are abundant and highly polymorphic in *D. willistoni* populations. They have an array of minisatellites in the 5' end of element, and this minisatellite region is transcribed and produces miRNAs, thereby providing a new source of miRNA from transposable elements.

## Acknowledgements

This study was supported by research grants (305330/2014-4) from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and

fellowships from CAPES and CNPq. Bioinformatic analysis was performed using the computational facilities of CPAD/UFSM.

## References

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., 2003. A uniform system for microRNA annotation. *RNA* 9, 277–279.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., Tuschl, T., 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5, 337-350.

Arensburger, P., Hice, H., Zhou., L., Smith, R.C., Tom, A.C., Wright, J.A., Knapp, J., O’Brochta, D.A., Craig, N.L., Atkinson, P.W., 2011. Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* 188, 45–57.

Arensburger, P., Piégu, B., Bigot, Y., 2016. The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine? *Mob Genet Elements* 6(6), e1256852.

Axtell, M.J., Westholm, J.O., Lai, E.C., 2011. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.* 12(4), 221.

Bassam, B.J., Gresshoff, P.M., 2007. Silver staining DNA in polyacrylamide gels. *Nat. Protoc.* 2(11), 2649-2654.

Biemont, C., 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* 186(4), 1085–1093.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: A sequence logo generator. *Genome Research*, 14,1188-1190.

De Lella Ezcurra, A.L., Bertolin, A.P., Kim, K., Katz, M.J., Gándara, L., Misra, T., Luschnig, S., Perrimon, N., Melani, M., Wappner, P., 2016. miR-190 Enhances HIF-Dependent Responses to Hypoxia in *Drosophila* by Inhibiting the Prolyl-4-hydroxylase Fatiga. *PLoS Genet.* 12(5), e1006073.

Dias, G.B., Heringer, P., Kuhn, G.C.S., 2016. Helitrons in *Drosophila*: Chromatin modulation and tandem insertions. *Mob. Genet. Elements* Mar 7;6(2), e1154638.

Fabian, M.R., Sonenberg, N., 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* 19(6), 586-593.

Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5(4), 103-107.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., Bartel, D.P., 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lisy-6* and other microRNAs. *Nat. Struct. Mol. Biol.*, 18 (10), 1139-1146.

Ha, M., Kim, V.N., 2014. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509-524.

Hua-Van, A., Le Rouzic, A., Boutin, T.S., Filée, J., Capy, P., 2011. The struggle for life of the genome’s selfish architects. *Biol. Direct.* 6(19), 19.



Huda, A., Tyagi, E., Mariño-Ramírez, L., Bowen, N.J., Jjingo, D., Jordan, I.K., 2011. Prediction of transposable element derived enhancers using chromatin modification profiles. *PLoS One*. 6(11), e27513.

Jaubert, S., Méreau, A., Antoniewski, C., Tagu, D., 2007. MicroRNAs in *Drosophila*: the magic wand to enter the chamber of secrets? *Biochimie*, 89, 1211-1220.

Jiang, N., Feschotte, C., Zhang, X., Wessler, S.R. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.* 7(2), 115-119.

Kalmykova, A.I., Klenov, M.S., Gvozdev, V.A., 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* 33(6), 2052–2059.

Kim, B., Jeong, K., Kim, V.N., 2017. Genome-wide Mapping of D ROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates, *Molecular Cell* 66, 258–269.

Klein, B.A., Chen, T., Scott, J.C., Koenigsberg, A.L., Duncan, M. J., Hu, L.T., 2015. Identification and characterization of a minisatellite contained within a novel miniature inverted-repeat transposable element (MITE) of *Porphyromonas gingivalis*. *Mobile DNA* 6, 18.

Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23(10),1289-91.

Kozomara, A., Griffiths-Jones, S., 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *NAR* 42, D68-D73.

Kulcheski, F.R., Oliveira L.F.V., Molina, L.G., Almerão, M.P., Rodrigues, F.A., Marcolino, J., Barbosa, J.F., Stolf-Moreira, R., Nepomuceno, A.L., Marcelino-Guimarães, F.C., Abdelnoor, R.V., Nascimento, L.C., Carazzolle, M.F., Pereira, G.A.G., Margis, R., 2011. Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*, 12, 307.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120,15-16.

Liu, Y.P., Schopman, N.C.T., Berkhout, B., 2013. Dicer-independent processing of short hairpin RNAs. *Nucleic Acids Research*, 41(6), 3723–3733.

Ma, Z., Coruh, C., Axtell, M.J., 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. *Plant Cell* 22(4), 1090-1103.

Marques, E.K., Napp, M., Winge, H., 1966. A corn meal, soybean flour, wheat germ medium for *Drosophila*. *Drosoph Inf Serv* 41, 187- 188.

Martínez-Izquierdo, J.A., García-Martínez, J., Vicient, C.M.,1997. What makes *Grande1* retrotransposon different? *Genetica* 100(1), 15-28.

Marzo, M., Liu, D., Ruiz, A., Chalmers, R., 2013. Identification of multiple binding sites for the THAP domain of the Galileo transposase in the long terminal inverted-repeats. *Gene* 525(1), 84-91.

Matzke, M.A., Mette, M.F., Matzke, A.J.M., 2000. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. *Plant Mol. Biol.* 43(2-3), 401-415.

Newman, M.A., Mani, V., Hammond, S.M., 2011. Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA*, 17:1795–1803.

Okonechnikov, K., Golosova, O., Fursov, M., 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 28(8), 1166-1167.

Ortiz, M.F., Lorenzatto, K.R., Corrêa, B.R.S., Loreto E.L.S., 2010. hAT transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica*. 138(6), 649-655.

Ortiz, M.F., Loreto, E.L.S., 2009. Characterization of new hAT transposable elements in 12 *Drosophila* genomes. *Genetica* 135(1), 67-75.

Piégu, B., Bire, S., Arensburger, P., Bigot, Y., 2015. A survey of transposable element classification systems – a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet Evol.* 86, 90-109.

Priyapongsa, J., Jordan, I.K., 2007. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS One* 2(2), e203.

Qin, S., Jin, P., Zhou, X., Chen, L., Ma, F., 2015. The Role of Transposable Elements in the Origin and Evolution of MicroRNAs in Human. *PLoS ONE* 10(6), e0131365.

Roberts, J.T., Cardin, S.E., Borchert, G.M., 2014. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences, *Mob. Genet. Elements* 4, e29255.

Rossato, D.O., Ludwig, A., Deprá, M., Loreto, L.S.E, Ruiz, A., Valente, V.L.S., 2014. BuT2 is a member of the third major group of hATtransposons and is involved in horizontal transfer events in the genus *Drosophila*. *Genome Biol. Evol.* 6(2), 352-365.

Sambrook, J., Russel, D.W., 2001. *Molecular cloning. A laboratory manual*. Third edition. Cold Spring Harbor Laboratory Press, New York.

Sassi, A.K., Heredia, F., Loreto, E.L.S., Valente, V.L.S., Rohde, C., 2005. Transposable elements P and gypsy in natural populations of *Drosophila willistoni*. *Genet. Mol. Biol.* 28(4), 734-739.

Shen, J., Liu, J., Xie, K., Xing, F., Xiong, F., Xiao, J., Li, X., Xiong, L., 2017. Translational repression by a miniature invertedrepeat transposable element in the 30 untranslated region. *Nat. Commun.* 8, 14651.

Shi, R., Chiang, V.L., 2005. Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* 39, 519–524.

Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12(4), 246-258.

Slotkin, R.K., Martienssen, R., 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272-285.

Souames, S., Bazin, C., Bonnivard, E., Higuët, D., 2003. Behavior of the hobo transposable element with regard to TPE repeats in transgenic lines of *Drosophila melanogaster*. *Mol Biol Evol* 20, 2055–2066.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3 - new capabilities and interfaces. *Nucleic Acids Research* 40(15), e115.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8, 973-82.

Ylla, G., Fromm, g., Piulachs M-D., Belles, X., 2016. The microRNA toolkit of insects. *Sci. Rep.* 6, 37736.

Zeng, Y., Zhang, X., Chu, J., Loh, H.H., Law, P-Y., 2010.  $\mu$ -Opioid Receptor Agonists Differentially Regulate the Expression of miR-190 and NeuroD. *Mol Pharmacol.* 77(1), 102–109.

Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P., Anderson, T.A., 2006. Evidence that miRNAs are different from other RNAs. *Cell. Mol. Life Sci.* 63(2), 246-254.

Zhang, Y., Zang, Q., Xu, B., Zheng, W., Ban, R., Zhang, H., Yang, Y., Hao, Q., Iqbal, F., Li, A., Shi, Q., 2016. IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics.* 32(13), 2069-2071.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406-3415.

## Supplementary material

**Supplementary Table 2.** *Drosophila willistoni* strains used in this study and informations about the geographical location of the strains used.

Species	Country	Location	Symbol	Latitude	Longitude
<i>D. willistoni</i>	Brazil	Aratiba	Ara	-52.3186	-27.3908
			Ara 6		
			Ara 7		
	Brazil	Amazônia	Ara 10	-59.6992	-2.8964
			CAX JD7		
	Brazil	Erechim	ISC	-52.26	-27.65
	Brasil	Itapeva	ITA	-49.76	-29.35
	Brasil	Joinvile	JOIN	-48.84	-26.29
	Brasil	Porto Alegre	MSANT	-51.13	-30.06
			MS		
	Brasil	Ribeirão Preto	RIB	-47.77	-21.20
			Rib_P		
	Brasil	Rondônia	RO	-63.59	-11.40
	Brasil	Pernambuco	S_Tal	-38.30	-7.96
	Brasil	Serra Cipó_ Minas Gerais	Scipo	-43.53	-19.36
			SC1		
			SC4		
	Brasil	Tangará da Serra_ Mato Grosso do Sul	SC19	-57.49	-14.62
			Tang		
			TG2		
França	Ilha Guadaloupe (Sotock Center)	TG19	-61.56	16.26	
		GD-H41-0811- 24			
Uruguai	Montevideo	Mon	-56.05	-34.88	

**Supplementary Table 3.** Primers and conditions used in PCRs and RT PCR.

Primers	Cycles	Condition	Sequence
Repeats_F	35	Denaturation at 95 °C for 5 min, denaturation for 30s at 94°C, annealing for 30s at 58°C and annealing for 40s at 72°C, after which the reaction mixture was additionally held at 72 °C for 7 min to complete the extension.	Reverse - 5' TTTGAGGACATTGGCAATCA 3'
Repeats_R			
Transposase_F	35	Denaturation at 95 °C for 5 min, denaturation for 15s at 94°C, annealing for 15s at 58°C and annealing for 15s at 72°C, after which the reaction mixture was additionally held at 72 °C for 4 min to complete the extension.	Forward - 5' TGC AAA TGC CTT CTG AGT TG 3'
Transposase_R			
Repeats_F	35	Denaturation at 95 °C for 5 min, denaturation for 30s at 95 °C, annealing for 30s at 58 °C, and extension for 1min at 72 °C, after which the reaction mixture was additionally held at 72 °C for 7 min to complete the extension.	
Transposase_F			
TIRH1_F	35	Denaturation at 95 °C for 5 min, denaturation for 30s at 94°C, annealing for 30s at 55°C and annealing for 30s at 72°C, after which the reaction mixture was additionally held at 72 °C for 4 min to complete the extension.	Forward - 5' TAG TGT TGG GTA AAG ATT GCT C 3'
TIRH1_R			

MIH1D_F		Denaturation at 95 °C for 5 min.	Forward - 5' TWG WTC TTT TTK GTT CAC 3'
MIH1D_R		<i>Step 1</i> - <b>20</b> cycles of denaturation	Reverse - 5' GTG AAC MAA AAA GAW CWA 3'
Poly(T) Adaptor		for 15s at 94 °C, annealing for 15s	Reverse - 5' GCG AGC ACA GAA TTA ATA CGA 3'
TIRH1_F		at 60 °C, and extension for 5s at 72	Forward - 5' TAG TGT TGG GTA AAG ATT GCT C 3'
	35(20/15)	°C. <i>Step 2</i> - 15 cycles of	
		denaturation for 15s at 94 °C,	
TIRH1_R		annealing for 15s at 60 °C, and	Reverse - 5' TAG TGT TGG GTT GCT CAT GAG 3'
		extension for 10s at 72 °C, after	
		which the reaction mixture was	
		additionally held at 72 °C for 4 min	
		to complete the extension.	
ACTINA_F		Denaturation at 95 °C for 5 min,	Forward - 5' ATG CAG AAG GAG ATC ACC GC 3'
		denaturation for 15s at 94°C,	
ACTINA_R	35	annealing for 15s at 60°C and	Reverse - 5' CTC GTC GTA CTC CTG CTT GG 3'
		annealing for 15s at 72°C, after	
		which the reaction mixture was	
		additionally held at 72 °C for 4 min	
		to complete the extension.	
Adaptor Oligo_DT		cDNA syntese	5' GCGAGCACAGAATTAATACGACTCACTATAGG(T)12VN 3'

---

**Supplementary Table 4** Results of polymorphism analyzes using sequencing and polyacrylamide gel electrophoresis (PAGE). The total number of MnRs found in each strain is informed.

<b>Lineage</b>	<b>Code</b>	<b>Seq</b>	<b>PAGE</b>	<b>Total</b>
CAX JD7	-	-	+14, 14 e 13	+14, 14 e 13
* GD-H41-0811-24	5wi	13	+14, 14 e 13	+14, 14 e 13
ARA	<b>21R</b> <b>22R</b>	14 e 13	+14, 14, 13 e 7	+14, 14 e 13
ARA6	<b>5_11F</b> <b>27R</b>	13	+14, 14, e 13	+14, 14 e 13
ARA7	-	13	+14, 14, e 13	+14, 14, e 13
ARA10	-	13	+14, 14, e 13	+14, 14, e 13
ITA	<b>23</b>	14	14	14
MSANT	-	-	+14 e 14	+14 e 14
JOIN	-	-	14	14
MSANT	-	-	14 e 7	14 e 7
RIB	-	-	+14 e 14	+14 e 14
RIB_P	-	-	+14 e 14	+14 e 14
ISC	<b>4</b>	13	14? E 13	+14, 14 e 13
MONT	-	-	14	14
S_TAL	-	-	+14, 14, e 13	+14, 14, e 13
SCIPO	<b>36</b> <b>3.37</b>	14 e 13	+14, 14, e 13	+14, 14, e 13
SC1	-	-	+14, 14, e 13	+14, 14, e 13
SC4	-	-	+14, 14, e 13	+14, 14, e 13
SC19	-	-	+14, 14, e 13	+14, 14, e 13
TANG	<b>32</b>	12	+14, 14, 13 e 7	+14, 14 e 12
TG2	-	-	+14, 14, 13 e 7	+14, 14 e 13
TG19	-	-	+14, 14, 13 e 7	+14 e 14
RO	<b>14</b> <b>5.8</b>	14 e 7	14 e 7	+14, 14 e 7

**Lineage:** The lineage is correspond of lineage of *D. willistoni* used.

**Code:** The code is correspond of symbol to sequence obtained in sequencing.

**Seq:** The SEQ is correspond of symbol to number of MnRs of *howilli1* element for each lineage.

**Supplementary Table 5** - Results of search for similar mature miRNAs in the miRbase.

Query	Accession	ID	species	GF	QS	QE	SS	SE	Alignment	
									5'MnRs	miRNA
* 5' TTGATCTTTTGGT TCACCAATA 3'	<a href="#">MIMAT0004440</a>	<a href="#">ame-miR-190</a>	<i>Apis mellifera</i>		1	15	9	23		
	<a href="#">MIMAT0005467</a>	<a href="#">dme-miR-190-5p</a>	<i>Drosophila melanogaster</i>		1	15	9	23		
	<a href="#">MIMAT0009153</a>	<a href="#">bmo-miR-190-5p</a>	<i>Bombyx mori</i>		1	15	9	23		
	<a href="#">MIMAT0010100</a>	<a href="#">aga-miR-190</a>	<i>Anopheles gambiae</i>		1	15	9	23		
	<a href="#">MIMAT0012227</a>	<a href="#">dps-miR-190-5p</a>	<i>Drosophila pseudoobscura*</i>		1	15	9	23		
	<a href="#">MIMAT0012449</a>	<a href="#">dsi-miR-190-5p</a>	<i>Drosophila simulans*</i>	mir-190	1	15	9	23	<u>uug auc uuu</u> uug guu	uug aua uuc uug guu
	<a href="#">MIMAT0014127</a>	<a href="#">api-miR-190</a>	<i>Acyrtosiphon pisum</i>		1	15	9	23		
	<a href="#">MIMAT0014243</a>	<a href="#">aae-miR-190</a>	<i>Aedes aegypti</i>		1	15	9	23		
	<a href="#">MIMAT0014402</a>	<a href="#">cqu-miR-190</a>	<i>Culex quinquefasciatus</i>		1	15	9	23		
	<a href="#">MIMAT0015673</a>	<a href="#">nvi-miR-190</a>	<i>Nasonia vitripennis</i>		1	15	9	23		
	<a href="#">MIMAT0023084</a>	<a href="#">tur-miR-190-5p</a>	<i>Tetranychus urticae</i>		1	15	9	23		
	<a href="#">MIMAT0035560</a>	<a href="#">dvi-miR-190-5p</a>	<i>Drosophila virilis</i>		1	15	9	23		
	<a href="#">MIMAT0036340</a>	<a href="#">gma-miR9725</a>	<i>Glycine max</i>		1	20	1	20	<u>uug auc uuu</u> uug guu cac ca	<u>uua auu uuu</u> uug gau cag ca
	<a href="#">MIMAT0027803</a>	<a href="#">mmu-miR-6951-3p</a>	<i>Mus musculus</i>	-	6	23	1	18	cuu uuu ggu uca cca aua	<u>cuu uuu ucu</u> uca caa aua
	<a href="#">MIMAT0013632</a>	<a href="#">bmo-miR-2760-5p</a>	<i>Bombyx mori</i>	-	6	17	12	23	cuu uuu ggu uca	cuu uuu ggu uca
<a href="#">MIMAT0026445</a>	<a href="#">sbi-miR6232a-3p</a>	<i>Sorghum bicolor</i>	-	1	21	1	21	<u>uug auc uuu</u> uug guu cac caa	<u>uug acu uuu</u> uug gua cau cca	
<a href="#">MIMAT0026444</a>	<a href="#">sbi-miR6232a-5p</a>	<i>Sorghum bicolor</i>	-	1	17	7	23	<u>uug auc uuu</u> uug guu ca	<u>uug acu uuu</u> uug gua ca	
<a href="#">MIMAT0030950</a>	<a href="#">stu-miR8011b-3p</a>	<i>Solanum tuberosum</i>	-	5	21	1	17	ucu uuu ugg uuc acc aa	<u>ucu uuu ugu</u> cuc acg aa	
* 5' TTGTTCTTTTGGT TCACGGCTACGAG 3'	<a href="#">MIMAT0023016</a>	<a href="#">bra-miR5718</a>	<i>Brassica rapa</i>	-	1	18	4	21	uug uuc ccu uuu ugg uuc	uug uuc ugu guu ugg uuc
	<a href="#">MIMAT0013632</a>	<a href="#">bmo-miR-2760-5p</a>	<i>Bombyx mori</i>	-	8	19	12	23	cuu uuu ggu uca	cuu uuu ggu uca
	<a href="#">MIMAT0035471</a>	<a href="#">sly-miR9477-5p</a>	<i>Solanum lycopersicum</i>	-	1	12	7	18	uug uuc ccu uuu	uug uuc ccu uuu
	<a href="#">MIMAT0015545</a>	<a href="#">bmo-miR-3359</a>	<i>Bombyx mori</i>	-	14	26	5	17	ggu uca cgg cua c	ggu uca cgg aua c
	<a href="#">MIMAT0026913</a>	<a href="#">mml-miR-890-3p</a>	<i>Macaca mulatta</i>	miR-743	2	14	4	16	ugu ucc cuu uuu g	ugu ucc cuu ucu g
	<a href="#">MIMAT0033117</a>	<a href="#">ppc-miR-8300-5p</a>	<i>Pristionchus pacificus</i>	-	11	23	10	22	uuu ggu uca cgg c	uuu ggu uca cga c
<a href="#">MIMAT0033819</a>	<a href="#">pxy-miR-279b-5p</a>	<i>Plutella xylostella</i>	miR-279 *d	9	21	11	23	uuu uug guu cac g	uuu uua guu cac g	

Query: sequence cloned in transcript of small RNAs; GF: Gene family; QS: Query start; QE: Query end; SS: Subject start; SE: Subject end; \* miR\_1d; \*\* miR\_5d; \*d: Present in *Drosophila*



**Supplementary Table 6- Results of the prediction of target genes.**

Target gene	Transcript	Symbol	8mer	Links to sites in UTRs	Biological Process	Ref.
FBgn0000359	FBtr0071203	Cp36	2	Sites in UTR	Eggshell chorion assembly; Vitelline membrane formation involved in chorion-containing eggshell formation.	Cernilogar et al., 2001; Kim et al., 2002; Perrimon et al., 1986
FBgn0026077	FBtr0078619	Gasp	1	Sites in UTR	Chitin-based embryonic cuticle biosynthetic process; Regulation of multicellular organism growth; Regulation of tube size, open tracheal system.	Tiklová et al., 2013
FBgn0028953	FBtr0086921	CG14478	1	Sites in UTR	Imaginal disc-derived wing morphogenesis; Inter-male aggressive behavior.	Edwards et al, 2009; Carreira et al., 2011
FBgn0036575	FBtr0075490	CG5157	1	Sites in UTR	-	-
FBgn0039688	FBtr0085460	Kul	1	Sites in UTR	Membrane protein ectodomain proteolysis; Regulation of Notch signaling pathway.	Sapir et al., 2005
FBgn0042133	FBtr0081386	CG18810	1	Sites in UTR	Protein palmitoylation.	Bannan et al., 2008
FBgn0066293	FBtr0086192	CheB42b	1	Sites in UTR	Detection of pheromone; Sensory perception of chemical stimulus.	Starostina et al., 2009 Xu et al., 2002

8mer: An exact match to positions 2-8 of the mature miRNA (the seed + position 8) followed by an 'A' (<http://www.targetscan.org/docs/8mer.html>). Accessed 03.05.2017.

Fasta archive from amplicons of MnRs regions of *howilli1* elements to different lineages of *D. willistoni* species.

> GD-H41-0811-24\_5wi

tcatgagtgagtgaacaaaaagagttgttcagtgaagtggagcaactggaacatgttcttccaatt  
gttctttttgttcacttgttcttttggttcacttgttcttttggctcacttgttcttttggttcacttgttctt  
tttggttcactagatcttttggttcacttgttcttttggttcactagatcttttggttcactagatcttttgg  
gttcactagatcttttggttcacttgttcttttggttcacttgttgttttggttcactagatcttttggttc  
acttgctcgatctacgttcttttgtctcaagttcttttcatattagcgatcatatcagttcacttttaa  
aattcagaatgtcgctaataagaaaacatagcgaaattggactcatttgaggacattggcaatcagc  
aggcgaaatgcaattattgcaaatgccttctgagttggaagtcccaaagcaatttaagccgacatttaa  
aatgtaaacaccggcagctatgagcccgtgtccgcaaaatgaggaaagctcggttgtgtgcaaa  
gcgccaaccaaggattacgacatttattcacaggccctgtcccagtgagcaaatctcagcaaagccg  
aattctgcagatatccatcacactggcggccgctcgagcatgcattctagagggccaattcgccctata  
gtgag

> ARA\_21R

ttcttttggttcactggttcttttggttcacttgttcttctggctcacttgttctttttagtgactagctcattt  
ttgttcacttgttcttttggttcacttgttcttttggttcactagatcttttggttcacttgttcttttggtt  
cactagatcttttggttcactagatcttttggttcactagatcttttggttcacttgttcttttggttcac  
ttgttgttttggttcactagatcttttggttcacttgtcctgatctacgttcttttgtctcaagttcttttcc  
atattagcgatcatatcagttcacttttaaaaattcagaatgtcgctaataagaaaacatagcgaaat  
ttggactcatttgaggacattggcaatcagcaggcgaaatgcaattattgcaaatgccttctgagttgg  
aagtcccaaagcaatttaagccgaca

> ARA\_22R

tccgcttaaatgcttgggacttccaactcagaaggcatttgaataattgcatttcgctgctgattgc  
caatgtcctcaaatgagtccaaatttcgctatgtttcttattagcgacattctgaatttttaaagtga  
aactgatatgatcgctaatatgaaaaagaacttgagacaaaagaacgtagatcgagcaagtgaac  
caaaaagatctagtgaacaaaaacaagaatgaacaaaaagaacaagtgaacaaaaagatct  
agtgaacaaaaagatctagtgaacaaaaagatctagtgaacaaaaagaacaagtgaacaaaa

agatctagtgaacaaaaagaacaagtgaacaaaaagaacaagtgagcaaaaaagaacaagtga  
 acaaaaaagaacaagtgaacaaaaagaacaattggaagagaacatgttcagttgctcacttactg  
 acaactcttttt

> ARA6\_5\_11F

taacgggctccattagctgccgggtgttacatatttaaatgttcggcttacattgcttgggactccaac  
 tcagaaggcatttgcaataattgcatttcgctgctgattgccaatgtcctcaaaatgagtccaaatttcg  
 ctatgttttcttattagcgacattctgaatttttaaaagtgaactgatatgatcgctaataatgaaaaa  
 gaacttgagacaaaaagaacgtagatcgagcaagtgaacaaaaagatctagtgaacaaaaaca  
 caagtgaacaaaaagaacaagtgaacaaaaagatctagtgaacaaaaagatctagtgaaccaa  
 aaagatctagtgaacaaaaagaacaagtgaacaaaaagatctagtgaacaaaaagaacaagt  
 gaacaaaaagaacaagtgaacaaaaagaactagtgaactaaaaagaaccagtgaacaaaaa  
 gaacaattggaagagacatgtcg

> ARA6\_27R

cacaaaaagagttgttcagtgaagtgagcaactgaacatgttcttccaattgttctttttgttcactt  
 gttctttttgttcacttgttctttttgctcacttgttctttttgttcacttgttctttttggttcacttagatctt  
 tttggttcacttgttctttttggttcacttagatctttttggttcacttagatctttttggttcacttagatctttttg  
 gttcacttgttctttttggttcacttgttgtttttggttcacttagatctttttggttcacttgtctgatctacgtt  
 tctttgtctcaagttctttttcatattagcgatcatatcagtttcacttttaaaattcagaatgtcgcta  
 ataagaaaacatagcgaatttggtcattttgaggacattggcaatcagcaggcgaatgcaatta  
 ttgcaaatgccttctgagttggaagtccaaagcaatttaagccgacatttaaaatgtaaacaccggc  
 agctatggagcccgttgcgcaaaatgaggaaagctcg

> ITA\_23

cacttgttctttttggttcacttgttctttttggttcacttgttctttttggttcacttgttctttttggttcactt  
 gttctttttggttcacttgttctttttggttcacttagatctttttggttcacttgttctttttggttcacttagatc  
 tttttggttcacttagatctttttggttcacttagatctttttggttcacttgttctttttggttcacttgttgtttt  
 ggttcacttagatctttttggttcacttgtctgatctacggttctttgtctcaagttctttttcatattagcg  
 atcatatcagtttcacttttaaaattcagaatgtcgctaataagaaaacatagcgaatttggtcactat  
 tttgaggacattggcaatcagcaggcgaatgcaattattgcaaatgccttctgagttggaagtcccaa

agcaatttaagccgacattttaaataatgtaaacacccggcagctatggagcccgttgccgcaaaatga  
ggaaaagctcgggttggtgcaaaagcgccaaccaaggattacgacatttattcac

> ISC\_4

tggttcactggggacagccctgtgaataaatgtcgtaatccttggtggcgcttgcacaacaaccgag  
ctttcctcatttgcggacaacgggctccatagctgccgggtgttacattttaaataatgtcggctaaaatt  
gcttgggacttccaactcagaaggcatttgcaataattgcatttcgctgctgattgccaatgtcctcaa  
aatgagtccaaatttcgctatgttttcttattagcgacattctgaatttttaaagtgaactgatgat  
cgctaataatgaaaaagaacttgagacaaaagaacgtagatcgagcaagtgaacaaaaagatct  
agtgaacaaaaacaagaactgaacaaaaagaacaagtgaacaaaaagatctagtgaacaaa  
aagatctagtgaacaaaaagatctagtgaacaaaaagaacaagtgaacaaaaagatctagtga  
acaaaaagaacaagtgaacaaaaagaacaagtgaacaaaaagaacaagtgagcaaaaaaga  
acaagtgaacaaaaagaacaagtgaacaaagaacattggaagagaac

> SCIPO\_36

ttttaaataatgtcggcttaaattgcttgggacttccaactcagaaggcatttgcaataattgcatttcgct  
gctgattgccaatgtcctcaaaatgagtccaaatttcgctatgttttcttattagcgacattctgaatttt  
aaaagtgaactgatgatcgctaataatgaaaaagaacttgagacaaaagaacgtagatcgag  
caagtgaacaaaaagatctagtgaacaaaaacaagaactgaacaaaaagaacaagtgaacca  
aaaagatctagtgaacaaaaagatctagtgaacaaaaagatctagtgaacaaaaagaacaagt  
gaacaaaaagatctagtgaacaaaaagaacaagtgaacaaaaagaacaagtgaacaaaaatg  
agctagtgcactaaaaagaacaagtgagcaaaaaagaacaagtgaacaaaaagaacaagtgaac  
aaaaagaacaattggaagaggaacatgttcagttgctcacgttcactgaacaactctt

> SCIPO\_3.37

caaaaaagagttgttcagtgaggaactgaacatgttcttccaattgttctttttgttcacttgt  
tctttttgttcacttgttctttttgctcacttgttctttttgttcacttgttctttttggttcactagatcttt  
tggttcacttgttctttttggttcactagatctttttggttcactagatctttttggttcactagatctttttg  
ttcacttgttctttttggttcacttgttgtttttggttcactagatctttttggttcacttgtcgcgatctacgtt  
ctttgtctcaagttctttttcatattagcgatcatatcagttcacttttaaataatcagaatgtcgcgtaa  
taagaaaacatagcgaatttggactcattttgaggacattggcaatcagcaggcgaatgcaattat

tgcaaatgccttctgagttggaagtcccaaagcaatttaagccgacatttaaagttaaaccacccggca  
gctatggagcccgttgcgccaatgaggaaagctcggttgttgcaaagcgccaaccaaggatt  
acgacattttattcacaggcctgtcccagtg

> TANG\_32

caactgggacaggcctgtgaataaatgctgtaatcccttgggtggcgctttgcacaacaaccgagcttt  
cctcattttggcggacaacgggctccatagctgccgggtgttacattttaaatgtcagcttaaattgcttt  
gggactccaactcagaaggcatttgcaataattgcatttcgctgctgattgccaatgtcctcaaaatg  
agtccaaatttcgctatgttttcttatttagcgacattctgaatttttaaagtgaaactgatatgatcgct  
aatatgaaaatggcgaactgagacaaaagaacgtagatcgagcaagtgaacaaaaagatctag  
tgaacaaaaacaacaagtgaacaaaaagaaccagtgaacaaaaatgaggtagtgcactaaaaa  
gaacaagtgaacaaaaagatctagtgaacaaaaagaacaagtgaacaaaaatgagctagtgcac  
taaaaagaacaagtgagcaaaaaagaacaagtgaacaaaaagaacaagtaacaaaaaagaac  
aattggaagagaacatgttcagttgctcacttactgaacaactcttt

> RO\_14

ggaataaagtcgtaatccttgggtggcgctttgcacaacaaccgagctttcctcattttggcggacaac  
gggctccatagctgccgggtgttacattttaaatgtcggcttaaattgctttgggactccaactcagaa  
ggcatttgcaataattgcatttcgctgctgattgccaatgtcctcaaaatgagtccaaatttcgctatgt  
tttcttatttagcgacattctgaatttttaaagtgaaactgatatgatcgctaataatgaaaaagaactt  
gagagaaaagaacgtagatcgagcaagtgaacaaaaagatctagtgaacaaaaacaacaagt  
gaacaaaaagaacaagtgaacaaaaagaacaagtgagcaaaaaagaacaagtgaacaaaaa  
gaacaagtgaacaaaaagaacaatgg

> RO\_5.8

caacatgttcttccaattgttcttttggttcacttgttcttttggttcacttgttcttttggctcacttgt  
tcttttggttcacttgttcttttggttcacttgttgttttggttcacttagatcttttggttcacttgtcga  
tctacgtttcttttctcaagttcttttcatattagcgatcatatcagtttacttttaaaaattcagaat  
gtcgctaataagaaaacatagcgaaatttgactcatttgaggacattggcaatcagcaggcgaat  
gcaattattgcaaatgccttctgagttggaagtcccaaagcaatttaagccgacatttaaagttaaaca  
cccggcagctatggagcccgttgcgccaatgaggaaagctcggttgtt



## 5. (ARTIGO 3) CENÁRIO EVOLUTIVO E CARACTERIZAÇÃO DE ELEMENTOS TRANSPONÍVEIS *HERVES-LIKE* (HAT) EM ESPÉCIES DE *DROSOPHILA*.

Larissa Paim Bernardo<sup>1</sup>; Daniela Mombach<sup>2</sup>; Élgion Lucio da Silva Loreto<sup>3</sup>

### Contribuição dos autores:

1: Revisões Bibliográficas; Montagem e estruturação do manuscrito; Reavaliação das sequências dos elementos *Herves-like*; Buscas nos genomas das espécies de *Drosophila*; Análises evolutivas; Caracterização das sequências; Extração de DNA, PCRs, Clonagens e sequenciamentos.

2. Caracterização das sequências; PCRs, Clonagens e sequenciamentos

### Lista de Abreviaturas

TEs Transposable Elementos

TIRs Terminal Inverted Repeats

TSD Target Sequence Duplication

MnRs Minisatelites Repeats

PA Potentially active

PNA Potentially non-active

MITE *Miniature Inverted Repeat Transposable Elements*

HT – Horizontal transfer

VT – Vertical transfer

TPE *threonine-proline-glutamic acid*

PCR - *Polymerase chain reaction*

.

## **Cenário evolutivo e caracterização de elementos transponíveis *Herves-like* (hAT) em espécies de *Drosophila*.**

Larissa Paim Bernardo; Daniela Mombach; Élgion Lucio da Silva Loreto

### Abstract

Transposable elements (TEs) and minisatellites repeats are typically identified as repetitive DNA into eukariotic genomes. TEs have the capability of mobilization and may have repetitive units in its sequences, those who in turn may act in different life stages of TEs. A group of *Drosophila* TEs, here called *Herves-like*, have in their sequences (and of its MITEs too) some minisatellites repeats (MnRs) in the 5'UTR region that contains these repetitive sequences. Recent studies have shown that these MnRs of *D. willistoni Herves-like* elements generates miRNAs that was predicted as regulators of important genes of *Drosophila*. Due to possible influence into regulatory networks as generators of silencers, the distribution of these TEs was evaluated. Through the *in silico* searches on available *Drosophila* genomes, as well as by sequencing clones of Neotropical *Drosophila* species, was possible to establish as actual global distribution evolutionary scenario of *Herves-like* elements. We found that *MITE Herves-like* are being maintained on *Drosophila* genomes and then we can suggest that this held is probably due to domestication since MnRs presents in these elements are putative conserved miRNAs precursors in these organisms.

### Introdução

Elementos transponíveis (do inglês, Transposable elements – TEs) são sequências de DNA que podem mudar sua localização nos genomas, processo conhecido como transposição, o que geralmente promove o aumento do número de cópias do TE no genoma (Biémont, 2010). Os TEs estão presentes em praticamente todos os organismos, são altamente diversos, e em seu processo de mobilização podem gerar rearranjos cromossômicos e mutações. As mutações geradas são, geralmente, deletérias para o organismo. Entretanto, as alterações genômicas e mutações geradas pelos TEs são apontadas como fonte de variabilidade genética e, desta forma, os TEs agem



como modeladores da estrutura, função e evolução dos genomas hospedeiros (Biémont, 2010). Através do mecanismo de transposição empregado é possível diferenciar duas grandes classe de TEs: i) retrotransposons ou elementos de classe I, que utilizam RNA como intermediário de transposição; ii) transposons ou elementos de classe II, que se transpõem via molécula de DNA. Além disso, os TEs são muito variáveis em sequência e estrutura proteica, o que possibilita outros níveis de classificação, tais como Ordem, Superfamília, Família e Subfamília (Wicker et al 2007; Piégu et al 2015). Ambas as classes apresentam elementos curtos, com menos de 1000 pb, originados a partir de deleções internas da sequência de elementos completos. Em elementos de classe II eles são chamados de MITEs e podem ser mobilizados pelas enzimas produzidas pelos elementos ativos dos quais são derivados ou de outros elementos que produzam transposases capazes de reconhecer as regiões terminais conservadas do MITE. Este processo é conhecido como *cross-mobilization* (Jiang et al., 2004).

A superfamília *hAT* é composta por elementos da classe II, e é amplamente distribuída em plantas, animais e fungos. Esses elementos são muito diversos, mas compartilham características em comum, tais como um tamanho aproximado de 2.5-5 kb, repetições terminais invertidas (do inglês, *terminal inverted repeats* – TIRs) de 10-25pb e geram duplicação da sequência alvo (do inglês, *target sequence duplication* - TSD) de 8pb (Arensburger et al., 2011; Feschotte e Pritham 2007). Três famílias de elementos *hATs* são conhecidas (*Ac*, *Buster* e *Tip*) e são caracterizadas pela similaridade da sequência da transposase, das TIRs e TSDs (Arensburger et al., 2011, Rossato et al., 2014).

Os elementos *howilli1* de *D. willistoni* e *hAT-1DP* de *D. persimilis* e *D. pseudoobscura* são elementos *hATs* da família *Ac* que apresentam repetições de minissatélites (do inglês, *Minisatellites Repeats* - MnRs) de 18nt que se repetem de 12-20 vezes na porção 5'UTR do gene da transposase desses elementos, e também na porção 3'UTR da transposase do elemento *hAT1\_DP*. Esses elementos foram inicialmente agrupados com elementos *homo2* de *D. mojavensis*, *hAT-1\_DSi* de *D. simulans*, *D. sechelia*, *D. melanogaster* e *D. mauritiana*, e os elementos *Herves* e *hAT-2\_AG* de *Anopheles gambiae* em um grupo que aqui chamaremos de *Herves-like*. Os elementos *Herves-like*

apresentam uma divergência de suas sequências de 0,28-0,74, e o consenso das sequências das TIRs é TAGTGTTGGGT (Ortiz e Loreto, 2009; Ortiz et al., 2010; Deprá et al., 2010).

Recentemente foi demonstrado que os MnRs presentes, tanto em cópias potencialmente ativas do elemento *howilli1* de *D. willistoni* quanto em cópias curtas (MITEs) provenientes desse elemento, são altamente polimórficos em número de repetições e se encontram em uma região que gera transcritos capazes de formar estruturas secundárias que podem dar origem a pequenos RNAs (Bernardo et al., SUBMETIDO). *D. willistoni* apresenta cerca de 30 cópias do elemento *howilli1* em seu genoma, a maioria delas são elementos MITEs, que apresentam os MnRs que foram encontrados em pequenos RNAs clonados em *D. willistoni*.

Em geral, MITEs são bons candidatos a precursores de miRNAs, pois normalmente estão inseridos em regiões intergênicas, e as sequências das TIRs possibilita a formação de estruturas secundárias estáveis quando esses elementos são transcritos (Piriyapongsa e Jordan, 2008). Os MITEs de elementos *howilli1* apresentam MnRs que foram recentemente associados como fontes de pequenos RNAs, similares os miRNAs conservados em insetos, da família *190-Mir* de miRNAs. Também foi avaliado a possibilidade de interação com genes alvos a partir da região da semente dos pequenos RNAs isolados de *D. willistoni* que apresentam similaridade com miRNAs da família *190-Mir* de miRNAs, e constatouse que os MnRs ao gerar miRNAs poderiam ser capazes de participar da regulação de genes importantes nos genomas de *Drosophila* (BERNARDO, et al, SUBMETIDO).

Diante da importância desses MnRs, o objetivo desse trabalho foi analisar a distribuição e as características de elementos com MnRs. Para isso, as sequências dos elementos *Herves-like* foram reavaliadas, caracterizadas e utilizadas como sonda em buscas *in silico* nos genomas sequenciados de *Drosophila*. Além disso, foram clonados e sequenciados fragmentos amplificados de elementos *Herves-like* de trinta e quatro espécies de *Drosophila* que não apresentam genoma disponível. Descobrimos que os elementos *Herves-like* apresentam uma distribuição ampla, e suas cópias curtas (*MITE Herves-like*) estão presentes em vários genomas sem sinal de presença dos elementos autônomos dos quais são derivados. Sugerindo que

esses elementos podem estar sendo mantidos nesses genomas por aquisição de alguma função biológica, que aqui associamos ao fato de carregarem consigo MnRs que possivelmente servem como fonte de pequenos RNAs.

## Material e Métodos

### Espécies utilizadas, Extração de DNA, PCR, clonagem e sequenciamento

Foram utilizadas 34 espécies do gênero *Drosophila* para a busca de elementos *Herves-like*: *D. bandeirantorum*, *D. brncici*, *D. buzzati*, *D. camargoj*, *D. capricorni*, *D. cardinoides*, *D. crocina*, *D. equinoxialis*, *D. funebris*, *D. gasici*, *D. gaucha*, *D. griseolineata*, *D. guaru*, *D. hydei*, *D. immigrans*, *D. insularis*, *D. maculifrons*, *D. mediodifusa*, *D. mediopunctata*, *D. mercatorum*, *D. ornatifrons*, *D. pallidipennis*, *D. paulistorum*, *D. pavani*, *D. pallidipennis*, *D. phalerata*, *D. robusta*, *D. sturtevantii*, *D. subbadia*, *D. tripunctata*, *D. equinoxialis*, *D. paulistorum*, *D. phalerata*, *Zaprionus indianus*.

O DNA genômico foi extraído de acordo com o protocolo de Sassi et al., (2005). As PCRs foram realizadas utilizando os *primers*: Repeats\_F (5'CAT GAG TGA GTG AAC AAA AAA GAG T 3') e Transposase\_R (5' TGC TGA GAT TTG CTC ACT GG 3') para amplificação da região dos MnRs; Primers TIR\_F (5' TAG TGT TGG GTA AAG ATT GCT C 3') e TIR\_R (5' TAG TGT TGG GTT GCT CAT GAG 3') para amplificação das cópias curtas *MITEs Herves-like*; *Primer* Transposase\_F (5' TGC AAA TGC CTT CTG AGT TG 3') e Transposase\_R2 (Reverse2 - 5'TTC GTC CAT GTC CAC GAT AA 3') para amplificação de cópias PA (Figura suplementar 1 e Tabela 1 do material suplementar). Os parâmetros para a realização da PCR foram: 10 ng de DNA genômico; 2,5 mM de MgCl<sub>2</sub>, 0,125 mM de dNTP, 0,25 pmol de *primers*, 0,2 U de Taq DNA polimerase, 1x de tampão para PCR. Os produtos de das PCRs foram verificados através de gel de agarose 0,8%.

Os *primers* foram desenhados com base no elemento *howilli1* (Ortiz e Loreto 2009) utilizando o software Primer3 (<http://frodo.wi.mit.edu/primer3>).

## **Análises *in silico***

As sequências dos elementos *Herves-like* foram caracterizadas e anotadas através do uso do programa UGENE (Okonechnikov et al 2012). O número de cópias foi inferido através das buscas por BLASTN e a confirmação das buscas foi feito através de BLAST no banco de dados de TEs, RepBase (<http://www.girinst.org/>) . As regiões de 5' MnRs foram alinhadas através do programa ClustalW implementado pelo programa MEGA 7 (Kumar et al., 2016) e a similaridade foi analisada usando WebLogo (<http://weblogo.berkeley.edu/logo.cgi>).

Buscas por elementos *Herves-like* nas espécies de *Drosophila* com genoma sequenciado foram realizadas por BLASTN (NCBI), utilizando algoritmo de alinhamento descontínuo (Megablast) e valor de threshold de 0.001. Os elementos *howilli1* (Ortiz e Loreto 2009), *hAT-1DP* (Kapitonov et al., 2003), *hAT-1\_DSi*, *homo2* (Ortiz e Loreto 2009) foram utilizados como sonda e as espécies *D. albomicans*, *D. ananassae*, *D. arizonae*, *D. biarmipes*, *D. bipectinata*, *D. elegans*, *D. erecta*, *D. eugracilis*, *D. ficusphila*, *D. grimshawi*, *D. kikkawai*, *D. melanogaster*, *D. miranda*, *D. mojavensis*, *D. persimillis*, *D. pseudoobscura*, *D. rhopaloa*, *D. simulans*, *D. sechellia*, *D. suzukii*, *D. takahashii*, *D. virilis*, *D. willistoni* e *D. yakuba* foram analisadas.

Os elementos encontrados nos genomas investigados e caracterizados neste trabalho foram nomeados conforme segue: 1) Prefixo com as iniciais do nome da espécie onde o elemento foi encontrado; 2) Um número para identificação da cópia caracterizada; 3) Sigla PA para elementos com características de sequências potencialmente ativas, PNA para potencialmente não ativas e MITE para cópias curtas. Sendo assim, por exemplo, os MITEs de *D. miranda* são nomeados: Dmir-1\_MITE, Dmir-2\_MITE, Dmir-3\_MITE e assim sucessivamente.

## **Análises evolutivas**

Foram utilizadas 57 sequências de elementos *MITE Herves-like* que apresentam 5'MnRs e que foram encontradas em espécies de *Drosophila* que tiveram os genomas analisados. Dentre essas sequências, 15 eram elementos

de *D. willistoni*, 5 de *D. pseudoobscura*, descritos previamente por Ortiz e Loreto em 2009, 20 de *D. miranda*, 11 de *D. takahashii*, 1 de *D. rhopaloa* e 5 de *D. melanogaster*, resultantes das nossas buscas e caracterizações. Os elementos foram alinhados com o software ClustalW implementado pelo programa MEGA 7 (Kumar et al., 2016). O relacionamento filogenético destes elementos foi estimado por Neighbor-Joining (NJ) (Saitou e Nei 1987) utilizando o programa MEGA 7, e por análise Bayesiana através do programa MrBayes (Huelsenbeck e Ronquist, 2001) utilizando o modelo de substituição nucleotídico GTR+G+I seguindo o modelo proposto por teste AIC (do inglês, *Akaike information criterion*) (Akaike, 1974) usando o programa MrModel Test 2.2 (Darriba et al 2011).

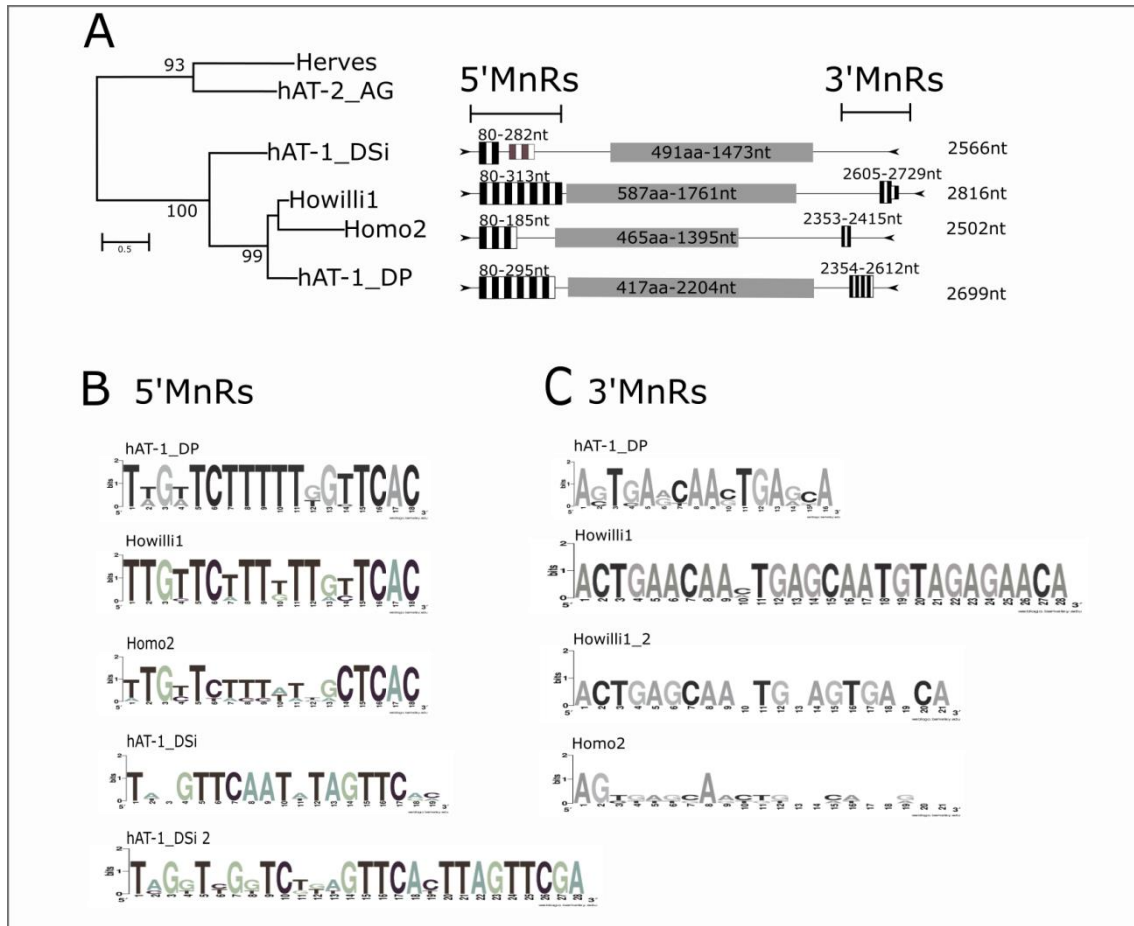
Os tempos de divergência das sequências dos elementos *MITE Herves-like* foram estimados segundo Graur e Li (2000) que estabelecem a equação  $T = k/2r$ , onde  $r$  corresponde a taxa evolutiva de 0.011 mudanças, por base, por milhões de anos e considerando taxa de mutação neutra para *Drosophila* (Tamura et al., 2004). A divergência média dos elementos que formaram os grupos foi utilizada para estimar o tempo de divergência dos grupos mostrados na filogenia.

## Resultados

### ***Elementos Herves-like: Caracterização das sequências, busca e análise dos elementos em outras espécies de Drosophila com genoma sequenciado.***

As análises dos elementos *Herves-like* demonstram que *howilli1* (*D. willistoni*), *hAT-1\_DP* (*D. persimilis* e *D. pseudoobscura*) e *homo2* (*D. mojavensis*) apresentam MnRs na porção 5' e 3' do elemento. O elemento *hAT-1DSi* (encontrado em *D. simulans* e *D. sechelia*, *D. melanogaster*, *D. mauritiana* e *Z. indianus*) apresenta repetições apenas na porção 5' de sua sequência (Fig. 1A-C). A sequência dos elementos *homo2* e *hAT-1\_DSi* foram reavaliadas. *Homo2* apresenta um excedente de 122nt a jusante da TIR 5', e 150nt a montante da TIR 3' do elemento. Excluindo essas regiões, também é possível identificar MnRs à 80 nt da TIR5' do elemento. O elemento *hAT-1\_DSi* também

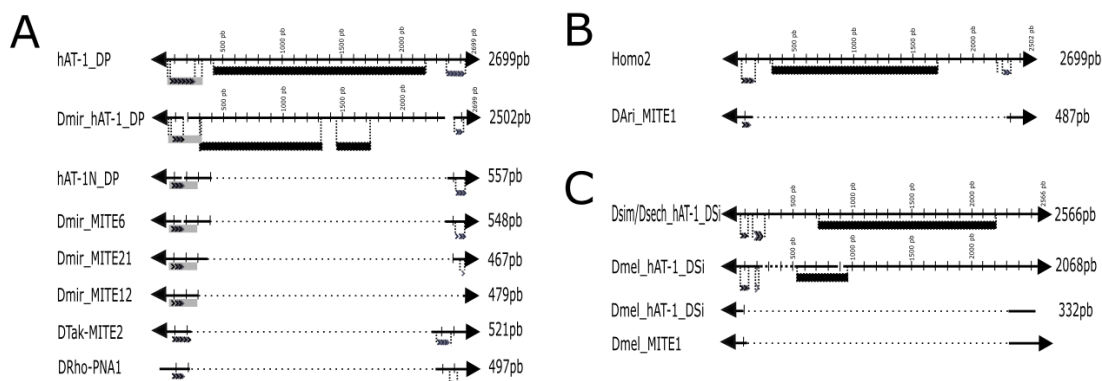
apresenta MnRs à 80nt da TIR 5'. No entanto, *hAT-1\_DSi* tem apenas três MnRs de 18nt na porção 5' do elemento, e quatro repetições de 28 nt imperfeitos, cerca de 36 nt a jusante das repetições MnRs (na posição 171-282nt) (Figura 1A-C).



**Figura 1.** Representação filogenética de elementos *Herves-like*, caracterização das sequências e representação das variações das sequências repetitivas presentes nas extremidades 5' e 3'. A) Neighbor-joining (NJ) utilizando a sequência das proteínas dos elementos *Herves-like*. Caracterização das sequências dos elementos *Herves-like*. Retângulo cinza representa a região da sequência que codifica a proteína; Retângulos listrados representam as regiões 5'MnRs e 3'MnRs. B) Representação das variações nas unidades repetitivas das 5'MnRs do elemento *hAT-1\_DP*, *howilli1*, *Homo2*, *hAT-1\_DSi* e minissatélites de 28nt do elemento *hAT-1\_DSi*. C) Representação das variações nas unidades repetitivas das 3'MnRs do elemento *hAT-1\_DP*, *howilli1*, *howilli1\_2* para representar os minissatélites de 21nt do elemento *howilli1* e *Homo2*.

As buscas por elementos *Herves-like* nas espécies de *Drosophila* com genoma sequenciado mostraram a ocorrência de duas cópias em *D. ananassae* (degeneradas), uma em *D. albomicans* (1MITE), duas em *D.*

*arizonae* (1 MITEs e 1 degenerada), duas em *D. bipectinata* (2PNA), duas em *D. fucusphila* (degeneradas), 33 em *D. miranda* (26 MITEs e 7 degenerados), uma em *D. nasuta* (degenerada), uma em *D. rhopaloa* (1 MITE) e 23 cópias em *D. takahashii* (12 MITEs e 11 degenerados) (Fig.2A-B, Tabela 1 e Tabela suplementar 2). A análise das estruturas das sequências encontradas mostra que os elementos apresentam graus variados de degeneração e rearranjos, desde pequenas deleções como no caso da cópia PNA encontrada em *D. miranda* (Figura 2A), até cópias com alto grau de degeneração de suas sequências, como no caso das sequências encontradas em *D. rhopaloa* e *D. melanogaster* (Figura 2A e 2C), e alguns rearranjos com duplicações das extremidades das sequências dos elementos, como no caso da cópia 16 de *D. miranda* (Fig. 2A).



**Figura 2** Diferentes estruturas de elementos Herves-like encontrados nos genomas de *Drosophila*.

Nas buscas por amplificação de fragmentos dos *Herves-like* nas espécies de *Drosophila*, obtivemos amplicons correspondentes as cópias curtas de elementos *MITE Herves-like* nas espécies: *D. bandeirantorum*, *D. crocina*, *D. equinoxialis*, *D. griseolineata*, *D. guaru*, *D. hydei*, *D. insularis*, *D. pavani* e *D. subbadia*. Encontramos ainda sinal positivo para a região dos 5'MnRs e uma porção da transposase de elementos *Herves-like* nas espécies: *D. equinoxialis*, *D. gaucha*, *D. virilis*, *D. cardinoides*, *D. insularis*, *D. ornatrinons* e *D. pallidipennis* (Tabela 1). Os amplicons das últimas quatro espécies foram sequenciados e a presença dos MnRs foi confirmada. A caracterização das sequências obtidas por sequenciamento feita por BLASTN no banco de dados de TEs (RepBase) e revelou a amplificação de cópias similares a *howilli1* com cerca de 13 MnRs na porção 5' dos elementos.

**Tabela 1** –Tabela com as espécies estudadas, resultado das buscas *in silico* por BLASTN e PCR nos genomas de *Drosophila* investigados.

Gênero	Subgênero	Grupo	Espécie	Positivo	Elemento
			<i>D. ornatifrons</i> (S)	+	Herves-like
		guarani	<i>D. subbadia</i>	+	MITE Herves-like
			<i>D. guaru</i>	+	MITE Herves-like
		guaramuru	<i>D. griseolineata</i>	+	MITE Herves-like
			<i>D. maculifrons</i>	-	-
			<i>D. crocina</i>	+	MITE Herves-like
			<i>D. tripunctata</i>	-	-
		tripunctata	<i>D. mediodifusa</i>	-	-
Drosophila	Drosophila		<i>D. bandeirantorum</i>	+	MITE Herves-like
			<i>D. mediopunctata</i>	-	-
		cardini	<i>D. cardinoides</i> (S)	+	Herves-like
		<b>guinaria</b>	<i>D. phalerata</i>	-	-
		pallidipennis	<i>D. pallidipennis</i> (S)	+	Herves-like
			<i>D. immigrans</i>	-	-
		immigrans	<i>D. albomicans</i> (G)	0	-
			<i>D. nasuta</i> (G)	0	Herves-like (d)
			<i>D. albomicans</i> (G)	0	MITE Herves-like



	<b><u>funnebris</u></b>	<i>D. funnebris</i>	-	-
		<i>D. gasici</i>	-	-
		<i>D. brncici</i>	-	-
	mesophragmatica	<i>D. gaucha</i>	+	Herves-like
		<i>D. pavani</i>	+	MITE Herves-like
		<i>D. hydei</i>	+	MITE Herves-like
		<i>D. buzzati</i>	-	-
		<i>D. mercatorum</i>	-	-
	repleta	<i>D. mojavenis</i> (G)	0	Herves-like
		<i>D. arizonae</i> (G)	0	Herves-like e MITE Herves-like
		<i>D. camargoi</i>	-	-
	virilis	<i>D. virilis</i> (G)	+	Herves-like
	<b><u>grimshawi</u></b>	<i>D. grimshawi</i> , (G)	0	-
	<b><u>robusta</u></b>	<i>D. robusta</i>	-	-
		<i>D. melanogaster</i> (G)	0	
		<i>D. simulans</i> (G)	0	
Sophophora	melanogaster	<i>D. sechelia</i> (G)	0	
		<i>D. mauritiana</i> (G)	0	Herves-like
		<i>D. erecta</i> (G)	0	

	<i>D. yakuba</i> (G)	0	
	<i>D. eugracilis</i> (G)	0	
	<i>D. elegas</i> (G)	0	
	<i>D. ficusphila</i> (G)	0	Herves-like (d)
	<i>D. suzuki</i> (G)	0	
	<i>D. biarmipes</i> (G)	0	
	<i>D. takahashii</i> (G)	0	Herves-like e MITE Herves-like
	<i>D. rhopaloa</i> ,	0	MITE Herves-like
	<i>D. kikawai</i> (G)	0	
	<i>D. ananassae</i> (G)	0	Herves-like (q)
	<i>D. bipectinata</i> (G)	0	Herves-like
	<i>D. miranda</i> (G)	0	Herves-like e MITE Herves-like
obscura	<i>D. persimillis</i> (G)	0	Herves-like e MITE Herves-like
	<i>D. pseudoobscura</i> (G)	0	Herves-like e MITE Herves-like
	<i>D. insularis</i>	+	MITE Herves-like
	<i>D. capricorni</i>	-	
willistoni	<i>D. paulistorum</i>	-	
	<i>D. equinoxialis</i>	+	Herves-like e MITE Herves-like
	<i>D. willistoni</i> (G)	0	Herves-like e MITE Herves-like

	<i>D. sturtevantii</i>	-	
Zaprionus	<i>Zaprionus indianus*</i>	+	Herves-like

---

Sublinhado estão os grupos de espécies em que não encontramos sinal positivo de presença do elemento nas espécies investigadas estão sublinhados.

(G) Representa as espécies em que foi encontrado sinal positivo por meio de buscas *In silico* por BLASTN.

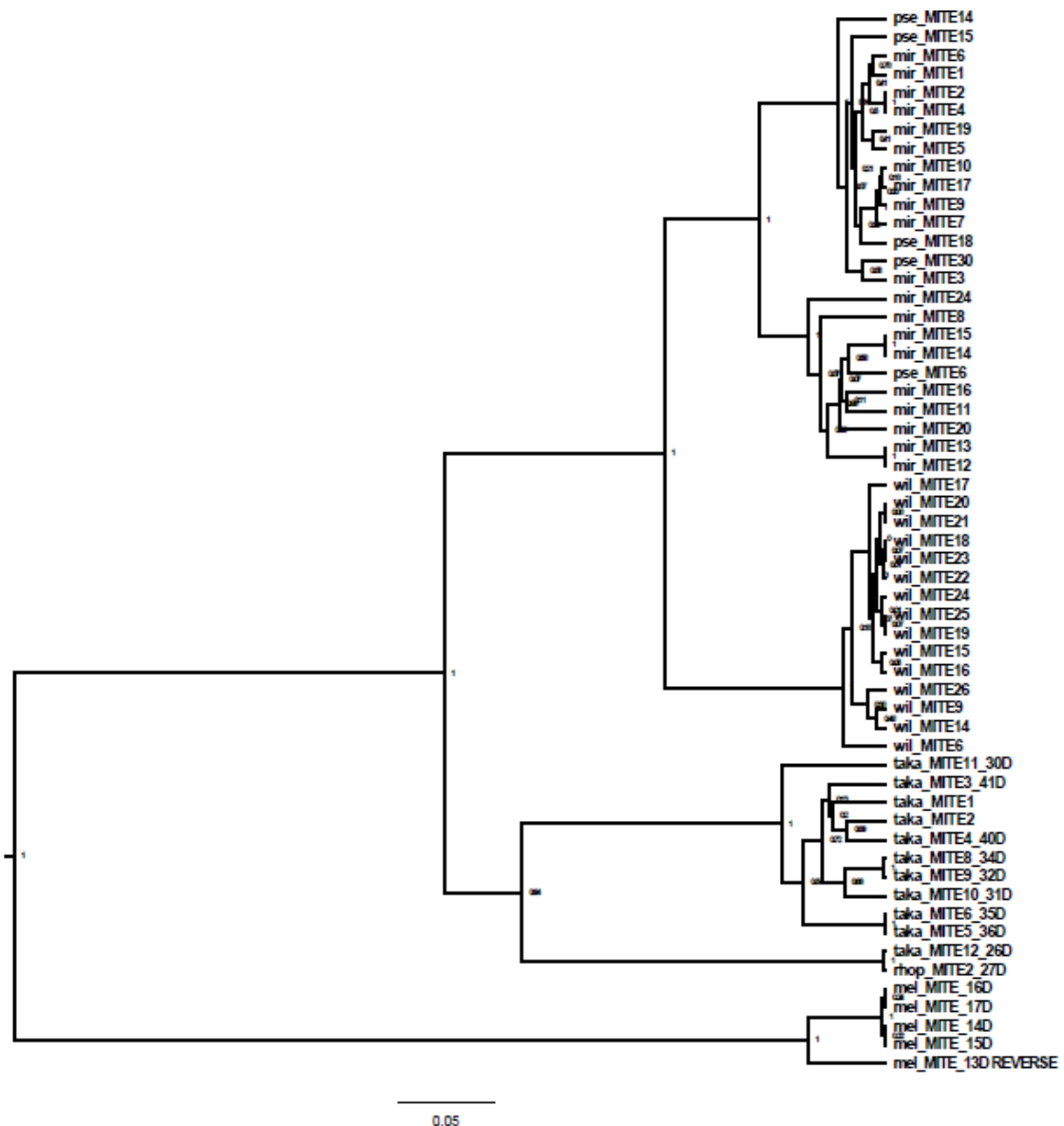
(S) Representa as espécies em que o elemento amplificado foi confirmado por sequenciamento.

(Q) Representa as espécies em que o elemento encontrado apresenta graus elevados de degeneração por meio de inserções de sequências (elementos quimeras).

(d) Representa as espécies em que o elemento encontrado apresenta graus elevados de degeneração por meio de deleções.

## Análise evolutiva dos MITEs

A Análise filogenética dos elementos *MITE-Herves-like* possibilitou a visualização de quatro cladogramas: I) Clado contendo as sequências que ocorrem em *D. pseudoobscura*, *D. persimilis* e *D. miranda*; II) Clado contendo os elementos encontrados no genoma de *D. willistoni*; III) Clado contendo os elementos de *D. rhopaloa* e *D. takahashii*; IV) Clado contendo os elementos de *D. melanogaster* (Figura 3).



**Figura 3** - Filogenia obtida utilizando a análise Bayesiana e o modelo GTR + I + G, conforme proposto pelo teste AIC para as 57 sequências de nucleotídeos dos elementos *MITE Herves-like*.

Os *MITEs Herves-like* que ocorrem em *D. melanogaster* apresentam os maiores valores de divergência nucleotídicas 0,28 entre as sequências de *D. pseudoobscura* (pse\_MITE3) e *D. melanogaster* (mel\_MITE\_13D\_\_REVERSE) até 0,51 entre as sequências de *D. takahashii* (taka\_MITE4) e *D. melanogaster* (mel\_MITE\_1) (Tabela suplementar 3 - Dados não mostrados).

Para as demais sequências de *MITE Herves-like* analisadas as divergências calculadas utilizando a distância-p variaram de 0,00 entre sequências de *D. miranda* (mir\_MITE10) e *D. pseudoobscura* (pse\_MITE15) a até 0,21 entre *D. miranda* (mir\_MITE16) e *D. takahashii* (taka\_MITE6\_35D).

O tempo de divergência estimado para as sequências nucleotídicas dos elementos *MITE Herves-like*, mostrou que os elementos de *D. melanogaster* invadiram esses genomas a cerca de 13-23,4 milhões de anos atrás, e os elementos que ocorrem em *D. willistoni* parecem ter invadido esses genomas mais recentemente, visto que os tempos de divergência estimados variaram de 2,4 milhões de anos (por exemplo, entre *D. willistoni* e *D. takahashii*) e 8,8 milhões de anos (entre *D. willistoni* e *D. pseudoobscura*).

## Discussão

Descobrimos que os 5' MnRs estão presentes em todos os elementos *Herves-like* de *Drosophila*, e não somente em *howilli1* e *hAT-1\_DP*. Os elementos *Herves-like* apresentam uma distribuição ampla e apenas quatro grupos de espécies pobremente amostrados aqui não apresentam sinais positivos para a ocorrência desses elementos, são eles: grupo quinaria (*D. phalerata*), grupo funebris (*D. funebris*), grupo grimshawi (*D. grimshawi*) e grupo robusta (*D. robusta*). Todos os demais grupos analisados apresentam espécies onde encontramos sinal positivo para os elementos *Herves-like*. Entretanto, a distribuição desses elementos ocorre de forma descontínua, e tanto nos resultados de sequenciamento de fragmentos amplificados, quanto por buscas *in silico* nas espécies com genoma sequenciado, em muitos casos só detectamos sinal positivo para a presença de elementos *MITE Herves-like*, como nos grupos guaramuru, tripunctata e virilis.

Quanto à origem, em geral elementos MITEs são considerados cópias com deleções internas em relação aos TEs autônomos relacionados a eles, mas esses elementos podem surgir também por eventos de recombinação nos quais pares de TIRs iguais ou similares são geradas e reconhecidas por transposase de elementos autônomos, podendo então ser mobilizados (Feschotte et al., 2002; Jiang et al., 2004; Ortiz e Loreto, 2010). A classificação de MITEs normalmente é baseada em similaridade de sequências, mas em nossas buscas não encontramos elementos autônomos ou com baixo grau de degeneração de sequências para que pudéssemos indicar com precisão os elementos que deram origem as cópias de *MITEs Herves-like* encontradas, como as sequências encontradas nos genomas de espécies *D. takahashii*, *D. rhopaloa* e *D. bipectinata*. Se pensarmos no ciclo evolutivo das sequências de TEs, após as fases iniciais de invasão e proliferação, os TEs tendem a ser eliminados por deriva genética ou seleção natural em suas linhagens originais (Schaack et al., 2010). Cópias MITEs apresentam vantagens estruturais que possibilitam um aumento rápido do número de cópias, o que poderia fazer com que esses elementos persistissem por mais tempo nos genomas. A HT entre espécies isoladas pode representar uma via de escape, evitando a eliminação

dos TEs por perdas estocásticas, mas a detecção desses eventos é feita por análises evolutivas de cópias autônomas (Schaack et al., 2010).

Alguns aspectos se mostram eficientes e podem ser utilizados na identificação de sinais de HT, como a alta similaridade de sequências de TEs que ocorrem em espécies distantemente relacionadas, a distribuição descontínua de TEs em grupos de espécies e a incongruência entre a filogenia do elemento e das espécies hospedeiras.

Sinais de HT recente para elementos *Herves-like* já haviam sido detectados em espécies do grupo *melanogaster* e em *Z. indianus* (elemento hAT-1-DSi) (Deprá-et al., 2010). Porém, os padrões de distribuição dos elementos *Herves-like* podem ser devido a combinações de eventos evolutivos e cada grupo de elementos deve ser avaliado de forma individual.

De forma geral, nossos resultados demonstram uma distribuição ampla e descontínua, característica que remete a perda estocástica. Porém, as análises evolutivas dos MITE *Herves-like*, apesar de serem pouco informativas para estimar a origem desses elementos, fornecem informações relevantes que podem ser sugestivas para distinguir aquisição por HT ou VT. Por exemplo, a presença de elementos *Herves-like* nas espécies relacionadas *D. persimilis*, *D. pseudoobscura* e *D. miranda* com sequências muito similares sugere que após a invasão em espécies ancestrais desse grupo, a manutenção desses elementos nos genomas deve estar ocorrendo por VT. Já os valores de divergência nucleotídica estimados para as sequências dos elementos de *D. takahashii* e *D. willistoni* foram menores ou iguais a 0,16, sugerindo alta similaridade dessas sequências. Essas espécies são distantemente relacionadas e tão pouco compartilham distribuição geográfica, mas possivelmente esses elementos estão presentes nesses genomas devido a HT dos elementos autônomos ancestrais que originaram esses MITEs e que já foram eliminados por seleção natural ou processos estocásticos. A alta similaridade das sequências também pode ser devido aos padrões de deleções internas das sequências dos elementos autônomos que originaram essas cópias de *MITE Herves-like*. Esta é uma hipótese parcimoniosa que explicaria alta similaridade dessas sequências, visto que, principalmente a extremidade 5' dos elementos *Herves-like* são bastante conservadas, o que inclui as

sequências dos MnRs, que não são afetadas pelas deleções internas que dão origem a esses *MITEs Herves-like* (Figura 2).

A ocorrência apenas das cópias curtas nos genomas sem a identificação de elementos autônomos relacionados, é uma situação que já foi observada em *Bombyx mori* e *Rhodnius prolixus* (Zhang et al., 2013). Estes autores sugerem que a ocorrência dos MITEs sem parceiros autônomos nesses genomas (que divergiram a cerca de ao menos 370 milhões de anos) são provenientes de TH dos elementos que deram origem aos MITEs, mas que já foram eliminados desses genomas por perdas estocásticas. Esses resultados foram argumentados pela identificação dos possíveis elementos autônomos que deram origem a esses MITEs, que ocorrem em espécies relacionadas e que poderiam ainda reter a capacidade de mobilização desses MITEs. Em nossos resultados, encontramos sinal positivo para elemento que podem ainda reter a capacidade de gerar transposase ativas nos grupos: guarani, cardini, pallidipennis, mesophragmatica e virilis (Tabela 1 e Figura 4), além de sinal positivo para presença de elementos degenerados no grupo melanogaster.

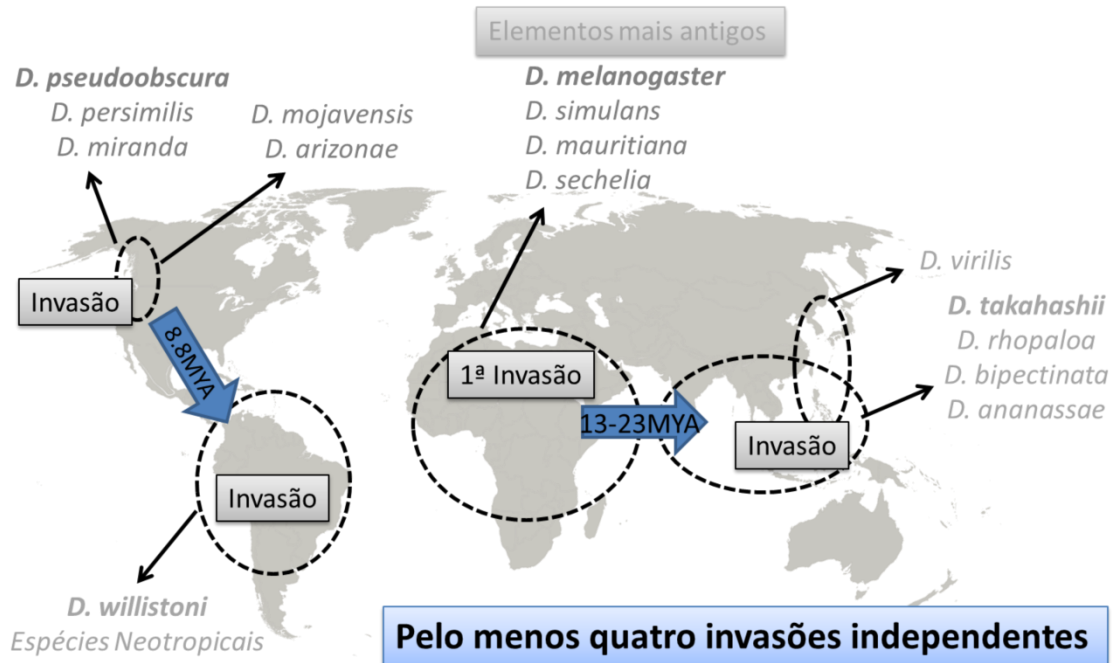
Observando de forma isolada os quatro clados resultantes de nossas análises, percebemos que pelo menos quatro elementos autônomos podem ser os candidatos a parceiros autônomos desses *MITEs Herves-like*, com pelo menos quatro invasões independentes nos diferentes grupos de espécies onde eles ocorrem (Figura 4): 1) elementos que invadiram os genomas das espécies do grupo melanogaster, representam as invasões aparentemente mais antigas, com indicativos de atividades e TH detectada recentemente para os genomas de *Z. indianus*; 2) elementos que invadiram as espécies asiáticas, dos grupo melanogaster e immigrans, onde encontramos sinais de elementos degenerados e cópias de MITEs, sendo que na espécie *D. takahashii*, encontramos cerca de 23 *MITE Herves-like* com sequências altamente similares as encontradas em *D. willistoni*, *D. pseudoobscura* e *D. miranda*; 3) elementos que invadiram as espécies havaianas e podem estar sendo mantidos por VT no grupo obscura e em espécies também havaianas do grupo repleta; 4) Elementos que invadiram as espécies neotropicais e podem estar sendo mantidos por HT e VT.

Sabendo que a distribuição de MITEs nos genomas normalmente é relacionada a genes (Oki et al., 2008), a ocorrência de MITEs Herves-like sem



parceiros autônomos nos genomas, também pode sugerir que eles possam estar sendo mantidos por aquisição de alguma função biológica. Estudos que demonstram as interações desse tipo de TEs com genomas hospedeiros ainda são pouco evidenciados. Uma característica importante dos MITEs Herves-like e dos elementos autônomos que possivelmente deram origem a eles é a presença de MnRs. Recentemente foram descobertos miRNAs provenientes das 5'MnRs encontradas no elemento Herves-like *howilli1*. Essas repetições estão presentes em sequências de miRNAs expresso em *D. willistoni*, e a predição de alvos potenciais para esses miRNAs em *D. melanogaster*, prevê como alvos genes importantes no desenvolvimento do sistema sensorial desses organismos (Bernardo et al., SUBMETIDO). Em *D. willistoni* foi constatado um alto grau de polimorfismo das MnRs (Bernardo et al., SUBMETIDO), e essas variações nas sequências de MnRs possibilitam a formação de diferentes estruturas secundárias podendo gerar diferentes miRNAs com diferentes alvos de regulação.

Por tanto, nossos resultados revelam um cenário evolutivo onde cópias de elementos curtos permanecem nos genomas após a eliminação de suas cópias autônomas relacionadas e sugerimos que essas cópias podem ser mantidas por domesticação (Volff 2006), visto que os MnRs presente em suas sequências são relacionados com a biogênese de miRNAs, e que MITEs em geral são bons precursores de miRNAs (Piriyapongsa and Jordan, 2007).



**Figura 5** Cenário atual de distribuição de elementos *Herves-like* em espécies de *Drosophila*. Elementos *Herves-like* provenientes de invasões mais antigas estão presentes nas espécies do grupo melanogaster: *D. mauritiana*, *D. simulans*, *D. melanogaster* e *D. sechelia*; Pelo menos mais três invasões podem ser sugeridas: (1) em espécies asiáticas do grupo melanogaster (*D. takahashii*, *D. rhopaloa*, *D. ananassae* e *D. bipectinata*) e *virilis* (*D. virilis*); (2) em espécies havaianas do grupo obscura (*D. pseudoobscura*, *D. persimilis*, *D. miranda*) e do grupo repleta (*D. mojavensis* e *D. arizonae*); (3) em espécies neotropicais;

## Referencias

- AKAIKE H. A new look at the statistical model identification. *IEEE Trans Automat Control* 19:716–723. 1974.
- ARENSBURGER, P., HICE, R.H., ZHOU, L., SMITH, R.C., TOM, A.C., WRIGHT, J.A. Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics*. V. 188, p. 45–57. 2011.
- BERNARDO, L.P., FONSECA, P.M., LORETO, E.L.S. A new source of miRNA: minisatellites carried by a hAT transposon in *Drosophila willistoni*. SUBMETIDO.
- BIÉMONT, . A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* v. 186(4), p. 1085-1093. 2010.
- DEPRÁ M., PANZERA Y., LUDWIG A., VALENTE V.L.S., LORETO E.L.S. hosimary: a new hAT transposon group involved in horizontal transfer. *Mol Genet Genomics* v. 283, p. 451–459. 2010.
- FESCHOTTE, C. PRITHAM, E. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* V. 41, p. 331-368. 2007.
- FESCHOTTE, C., ZHANG, X., WESSLER, S. R. Miniature inverted-repeat transposable elements and their relationships to established DNA transposons, pp. 1147–1158 in *Mobile DNA II*, 2002, edited by N. Craig, R. Craigie, M. Gellert and A. Lambowitz. American Society for Microbiology Press, Washington, DC.
- GRAUR, D., LI, W.H. *Fundamentals of molecular evolution*, 2nd edn. Sinauer Associates, Sunderland, MA. 2000.
- HUELSENBECK JP, RONQUIST F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–55
- JIANG, N., FESCHOTTE, C., ZHANG, X., WESSLER, S. R. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant. Biol.* V. 7, p. 115–119. 2004.
- KUMAR S, STECHER G, TAMURA K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33 (7): 1870-1874. 2016.
- NYLANDER, J. ET AL. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67. 2004.
- OKI, N., YANO, K., OKUMOTO, Y., TSUKIYAMA, T., TERAISHI, M., TANISAKA, T., A genome-wide view of miniature inverted-repeat transposable

elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*. *Genes Genet Syst.* 83:321–329. 2008.

OKONECHNIKOV K, GOLOSOVA O, FURSOV M, Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28: 1166-1167. 2012.

ORTIZ, M.F., LORENZATTO, K.R., CORRÊA, B.R.S., LORETO E.L.S., hAT transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica.* 138(6), 649-655. 2010.

ORTIZ, M.F., LORENZATTO, K.R., CORRÊA, B.R.S., LORETO E.L.S., hAT transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. *Genetica.* 138(6), 649-655. 2010.

PIÉGU, B., BIRE, S., ARENSBURGER, P., BIGOT, Y. A survey of transposable element classification systems – a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet Evol.* 86, 90-109. 2015.

PIRIYAPONGSA, J., JORDAN, I.K. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS One* 2(2), e203. 2007.

PIRIYAPONGSA, JITTIMA, AND I. KING JORDAN. Dual coding of siRNAs and miRNAs by plant transposable elements. *Rna* V. 14(5), p. 814-821. 2008.

ROSSATO, D.O., LUDWIG, A., DEPRÁ, M., LORETO, E.L.S., RUIZ, A., VALENTE, V.L.S. *BuT2* is a member of the third major group of *hAT* transposons and is involved in horizontal transfer events in the genus *Drosophila*. *Genome biology and evolution.* V. 6(2), p. 352-365. 2014.

SAITOU, N., NEI, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. 1987.

SASSI AK, HERÉDIA FO, LORETO ELS, VALENTE VLS, ROHDE C. Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. *Genet Mol Biol* 28:734–739. 2005.

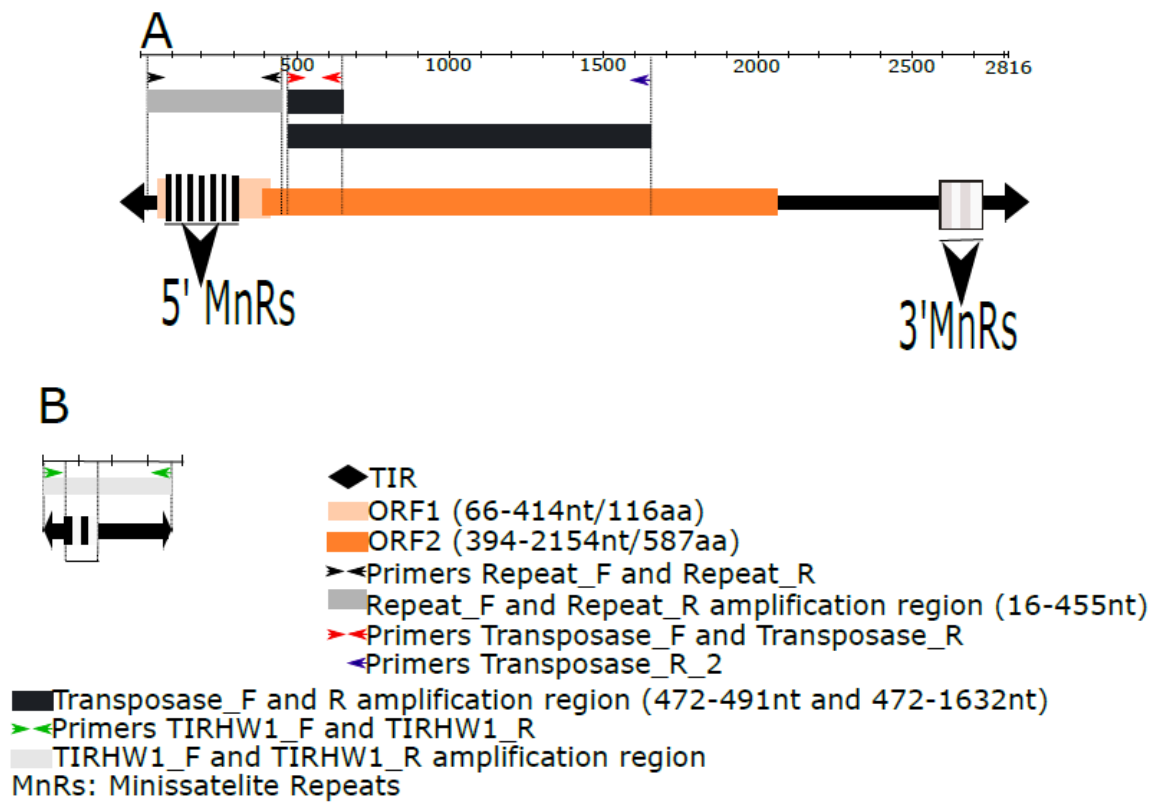
SCHAACK S, GILBERT C., FESCHOTTE C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol.* 25:537- 546. 2010.

VOLFF, J.N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* 28, 913–922. 2006.

WICKER, T, SABOT, F, HUA-VAN, A, ET AL. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973-982. 2007.

ZHANG, HUA-HAO, ET AL. The origin and evolution of six miniature inverted-repeat transposable elements in *Bombyx mori* and *Rhodnius prolixus*. *Genome biology and evolution* 5.11 (2013): 2020-2031.

## Material suplementar



**Figura suplementar 1** - Representação da região de anelamento e amplificação dos *primers*. A) Representação da sequência do elemento autônomo. B) Representação da sequência de um elemento *MITE Herves-like*

**Tabela Suplementar 1** Informações sobre os *primers*.

Primer	Ciclos	Condições	Sequencias
Repeats_F	35	Desnaturação a 95 °C por 5 min, desnaturação por 30s à 94°C, temperatura de anelamento por 30s à 58°C e extensão por 40s à 72°C, extensão final à 72 °C por 7 min.	Forward - 5'CAT GAG TGA GTG AAC AAA AAA GAG T 3'
Repeats_R			Reverse - 5' TTTGAGGACATTGGCAATCA 3'
Transposase_F	35	Desnaturação a 95 °C por 5 min, desnaturação por 15s à 94°C, temperatura de anelamento por 15s à 58°C e extensão por 15s à 72°C, extensão final à 72 °C por 4 min.	Forward - 5' TGC AAA TGC CTT CTG AGT TG 3'
Transposase_R			Reverse - 5' TGC TGA GAT TTG CTC ACT GG 3'
Repeats_F	35	Desnaturação a 95 °C por 5 min, desnaturação por 30s à 94°C, temperatura de anelamento por 30s à 58°C e extensão por 1min à 72°C, extensão final à 72 °C por 7 min.	Forward - 5'CAT GAG TGA GTG AAC AAA AAA GAG T 3'
Transposase_R			Reverse - 5' TGC TGA GAT TTG CTC ACT GG 3'
Transposase_R_2		Desnaturação a 95 °C por 5 min, desnaturação por 30s à 94°C, temperatura de anelamento por 30s à 61°C e extensão por 1,5min à 72°C, extensão final à 72 °C por 7 min.	Reverse2 - 5'TTC GTC CAT GTC CAC GAT AA 3'
TIRH1_F	35	Desnaturação a 95 °C por 5 min, desnaturação por 30s à 94°C, temperatura de anelamento por 30s à 55°C e extensão por 30s à 72°C, extensão final à 72 °C por 4 min..	Forward - 5' TAG TGT TGG GTA AAG ATT GCT C 3'
TIRH1_R			Reverse - 5' TAG TGT TGG GTT GCT CAT GAG 3'

**Tabela Suplementar 2** – Resultado das buscas por elementos *Herper-like* nos genomas de *Drosophila*. As buscas foram realizadas por meio de BLASTN.

Code	Specie	QI	SI	% ID	AL	q. start	q. end	s. start	s. end
1ana	<i>D. ananassae</i>	<i>hAT-1_DP</i>	gi 91861783 gb AAPP01015573.1	67.575	1064	1140	2167	36483	35444
		<i>hAT-1_DP</i>	gi 91861783 gb AAPP01015573.1	67.975	943	188	1112	42550	41627
		<i>hAT-1_DP</i>	gi 91861783 gb AAPP01015573.1	88.652	141	1	141	42629	42489
		<i>hAT-1_DP</i>	gi 91861783 gb AAPP01015573.1	83.099	142	2562	2699	35180	35040
2ana	<i>D. ananassae</i>	<i>homo2</i>	gi 91854186 gb AAPP01018374.1	72.222	324	1743	2065	6747	7069
		<i>homo2</i>	gi 91854186 gb AAPP01018374.1	80	135	2369	2502	7507	7636
1Alb	<i>D. albomicans</i>	<i>homo2</i>	gi 824723177 gb JXOX01007040.1	87.121	132	1	132	2225	2356
		<i>homo2</i>	gi 824723177 gb JXOX01007040.1	74.615	130	2373	2502	2742	2860
1Ari	<i>D. arizonae</i>	<i>homo2</i>	gi 1026678484 gb LSRM01000001.1	95.758	165	1	165	21345913	21345749
		<i>homo2</i>	gi 1026678484 gb LSRM01000001.1	100	119	2384	2502	21345545	21345427
		<i>homo2</i>	gi 1026678410 gb LSRM01000005.1	96.359	412	80	491	5182992	5182582
2Ari	<i>D. arizonae</i>	<i>homo2</i>	gi 1026678410 gb LSRM01000005.1	100	52	2448	2499	5182544	5182493
		<i>homo2</i>	gi 1026678410 gb LSRM01000005.1	80.303	66	6	71	22101328	22101391
		<i>homo2</i>	gi 1026678410 gb LSRM01000005.1	65.566	212	1193	1385	3087856	3088067
1Bip	<i>D. bipectinata</i>	<i>howilli1</i>	gi 449798696 gb AFFE02003354.1	67.362	1391	380	1731	2944	4315
		<i>howilli1</i>	gi 449798696 gb AFFE02003354.1	81.667	180	1	179	2552	2720
		<i>howilli1</i>	gi 449798696 gb AFFE02003354.1	80.108	186	2635	2816	4845	5019
		<i>howilli1</i>	gi 449798696 gb AFFE02003354.1	77.698	139	2019	2157	4312	4450
2Bip	<i>D. bipectinata</i>	<i>hAT-1_DP</i>	gi 1218636909 gb LYTD01025939.1	67.128	1156	359	1490	1816	677
		<i>hAT-1_DP</i>	gi 1218636909 gb LYTD01025939.1	72.519	131	2037	2167	387	257
		<i>hAT-1_DP</i>	gi 1218636909 gb LYTD01025939.1	73.737	99	1659	1757	483	385

1fic	<i>D. ficusphila</i>	<i>hAT-1_DP</i>	gi 449816737 gb AFFG02005623.1	68.162	1360	852	2176	6754	5405
		<i>hAT-1_DP</i>	gi 449816737 gb AFFG02005623.1	70.281	498	354	845	7986	7496
		<i>hAT-1_DP</i>	gi 449816737 gb AFFG02005623.1	79.651	172	1	164	8282	8111
2fic	<i>D. ficusphila</i>	<i>hAT-1_DP</i>	gi 449816667 gb AFFG02005693.1	67.031	1647	152	1762	12093	13727
		<i>hAT-1_DP</i>	gi 449816667 gb AFFG02005693.1	75.49	306	1871	2176	14525	14829
		<i>hAT-1_DP</i>	gi 449816667 gb AFFG02005693.1	94.545	55	1	55	10882	10936
1Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478839145 gb AJMIO2000294.1	89.6	250	207	455	4893	4648
		<i>hAT-1_DP</i>	gi 478839145 gb AJMIO2000294.1	96.203	158	1	158	4973	4816
		<i>hAT-1_DP</i>	gi 478839145 gb AJMIO2000294.1	100	58	2642	2699	4629	4572
2Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478838998 gb AJMIO2000441.1	87.037	216	225	438	378	168
		<i>hAT-1_DP</i>	gi 478838998 gb AJMIO2000441.1	95	140	1	140	458	319
		<i>hAT-1_DP</i>	gi 478838998 gb AJMIO2000441.1	87.5	144	2556	2699	137	1
3Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478838462 gb AJMIO2000977.1	89.787	235	207	440	68729	68958
		<i>hAT-1_DP</i>	gi 478838462 gb AJMIO2000977.1	93.038	158	1	158	68649	68806
		<i>hAT-1_DP</i>	gi 478838462 gb AJMIO2000977.1	93.75	64	2636	2699	69064	69127
4Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478838218 gb AJMIO2001221.1	84.477	277	153	428	22179	21912
		<i>hAT-1_DP</i>	gi 478838218 gb AJMIO2001221.1	90.094	212	1	212	22259	22048
		<i>hAT-1_DP</i>	gi 478838218 gb AJMIO2001221.1	82.051	156	2367	2512	21863	21708
5Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478838218 gb AJMIO2001221.1	92.647	68	2632	2699	20057	19990
		<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	88.764	178	207	383	20976	20803
		<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	91.139	158	1	158	21056	20900
6Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	79.556	225	2367	2578	22171	21952
		<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	78.924	223	2369	2578	20758	20541
		<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	73.77	122	2578	2699	20604	20509
6Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478837978 gb AJMIO2001461.1	73.77	122	2578	2699	22015	21920
		<i>hAT-1_DP</i>	gi 478837849 gb AJMIO2001590.1	85.6	250	207	455	28446	28219
		<i>hAT-1_DP</i>	gi 478837849 gb AJMIO2001590.1	88.021	192	2512	2699	28206	28025
		<i>hAT-1_DP</i>	gi 478837849 gb AJMIO2001590.1	96.429	140	1	140	28526	28387



		<i>hAT-1_DP</i>	gi 478837671 gb AJMIO2001768.1	87.037	216	225	438	381	171
7Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478837671 gb AJMIO2001768.1	95	140	1	140	461	322
		<i>hAT-1_DP</i>	gi 478837671 gb AJMIO2001768.1	87.5	144	2556	2699	140	4
		<i>hAT-1_DP</i>	gi 478837488 gb AJMIO2001951.1	85.595	479	1634	2106	1954	1489
		<i>hAT-1_DP</i>	gi 478837488 gb AJMIO2001951.1	92.33	352	1085	1435	2332	1985
8Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478837488 gb AJMIO2001951.1	80.577	381	225	605	2838	2482
		<i>hAT-1_DP</i>	gi 478837488 gb AJMIO2001951.1	90.141	142	1	142	2918	2777
		<i>hAT-1_DP</i>	gi 478837488 gb AJMIO2001951.1	79.747	158	691	831	2483	2326
		<i>hAT-1_DP</i>	gi 478837485 gb AJMIO2001954.1	87.037	216	225	438	378	168
9Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478837485 gb AJMIO2001954.1	95	140	1	140	458	319
		<i>hAT-1_DP</i>	gi 478837485 gb AJMIO2001954.1	87.5	144	2556	2699	137	1
		<i>hAT-1_DP</i>	gi 478836834 gb AJMIO2002605.1	87.063	286	171	455	2922	3185
10Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836834 gb AJMIO2002605.1	96.023	176	1	176	2842	3017
		<i>hAT-1_DP</i>	gi 478836834 gb AJMIO2002605.1	89.062	192	2512	2699	3198	3381
		<i>hAT-1_DP</i>	gi 478836489 gb AJMIO2002950.1	95.57	158	1	158	53235	53392
11Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836489 gb AJMIO2002950.1	85.38	171	2529	2699	53779	53936
		<i>hAT-1_DP</i>	gi 478836489 gb AJMIO2002950.1	78.947	171	2368	2528	53696	53858
		<i>hAT-1_DP</i>	gi 478836462 gb AJMIO2002977.1	91.958	286	171	455	57405	57686
12Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836462 gb AJMIO2002977.1	96.392	194	1	194	57325	57518
		<i>hAT-1_DP</i>	gi 478836462 gb AJMIO2002977.1	88.172	186	2514	2699	57701	57874
		<i>hAT-1_DP</i>	gi 478836439 gb AJMIO2003000.1	91.958	286	171	455	86613	86894
13Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836439 gb AJMIO2003000.1	94.33	194	1	194	86533	86726
		<i>hAT-1_DP</i>	gi 478836439 gb AJMIO2003000.1	88.298	188	2512	2699	86907	87082
		<i>hAT-1_DP</i>	gi 478836433 gb AJMIO2003006.1	91.259	286	171	455	34122	34403
14Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836433 gb AJMIO2003006.1	93.814	194	2	194	34043	34236
		<i>hAT-1_DP</i>	gi 478836433 gb AJMIO2003006.1	85.784	204	2496	2699	34416	34599
15Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478836315 gb AJMIO2003124.1	90.187	214	243	455	413224	413433
		<i>hAT-1_DP</i>	gi 478836315 gb AJMIO2003124.1	99.18	122	1	122	413144	413265

		<i>hAT-1_DP</i>	gi 478836315 gb AJMIO2003124.1	87.195	164	2536	2699	413446	413597	
		<i>hAT-1_DP</i>	gi 478835854 gb AJMIO2003585.1	90.987	2441	188	2544	1470	3820	
		<i>hAT-1_DP</i>	gi 478835854 gb AJMIO2003585.1	95.381	866	403	1268	1	843	
16	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478835854 gb AJMIO2003585.1	82.91	433	2268	2699	4379	4754
			<i>hAT-1_DP</i>	gi 478835854 gb AJMIO2003585.1	93.182	176	1	176	1394	1566
			<i>hAT-1_DP</i>	gi 478835854 gb AJMIO2003585.1	92.208	154	2546	2699	3745	3895
			<i>hAT-1_DP</i>	gi 478835840 gb AJMIO2003599.1	92.308	286	171	455	18642	18923
17	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478835840 gb AJMIO2003599.1	96.429	196	1	196	18562	18757
			<i>hAT-1_DP</i>	gi 478835840 gb AJMIO2003599.1	86.765	204	2496	2699	18936	19127
			<i>hAT-1_DP</i>	gi 478835500 gb AJMIO2003939.1	86.111	216	225	438	2401	2611
			<i>hAT-1_DP</i>	gi 478835500 gb AJMIO2003939.1	95.714	140	1	140	2321	2460
18	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478835500 gb AJMIO2003939.1	88.194	144	2556	2699	2642	2780
			<i>hAT-1_DP</i>	gi 478835500 gb AJMIO2003939.1	87.5	144	2556	2699	3053	3191
			<i>hAT-1_DP</i>	gi 478835500 gb AJMIO2003939.1	86.806	144	2556	2699	3520	3658
19	Mir	<i>D. miranda</i>	<i>howilli1</i>	gi 478835429 gb AJMIO2004010.1	89.286	140	1	140	18158	18025
			<i>howilli1</i>	gi 478835429 gb AJMIO2004010.1	81.609	174	2653	2816	17735	17563
			<i>hAT-1_DP</i>	gi 478835405 gb AJMIO2004034.1	92.164	268	189	455	57888	57625
20	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 478835405 gb AJMIO2004034.1	96.023	176	1	176	57968	57793
			<i>hAT-1_DP</i>	gi 478835405 gb AJMIO2004034.1	88.235	204	2496	2699	57612	57421
			<i>hAT-1_DP</i>	gi 823758566 gb JXPA01000174.1	92.33	352	1085	1435	639	292
			<i>hAT-1_DP</i>	gi 823758566 gb JXPA01000174.1	75.65	423	225	605	1199	789
21	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823758566 gb JXPA01000174.1	83.212	274	1634	1905	261	1
			<i>hAT-1_DP</i>	gi 823758566 gb JXPA01000174.1	90.141	142	1	142	1279	1138
			<i>hAT-1_DP</i>	gi 823758566 gb JXPA01000174.1	79.747	158	691	831	790	633
			<i>hAT-1_DP</i>	gi 823757871 gb JXPA01000835.1	87.246	2070	344	2341	5534	3530
22	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823757871 gb JXPA01000835.1	87.246	224	2287	2504	2759	2541
			<i>hAT-1_DP</i>	gi 823757871 gb JXPA01000835.1	93.069	101	7	107	5638	5538
23	Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823757198 gb JXPA01001508.1	78.995	438	1	438	2288	2664

		<i>hAT-1_DP</i>	gi 823757198 gb JXPA01001508.1	93.939	99	2601	2699	2795	2893
		<i>hAT-1_DP</i>	gi 823756812 gb JXPA01001894.1	89.787	235	207	440	14910	15139
24Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823756812 gb JXPA01001894.1	93.038	158	1	158	14830	14987
		<i>hAT-1_DP</i>	gi 823756812 gb JXPA01001894.1	93.75	64	2636	2699	15245	15308
		<i>hAT-1_DP</i>	gi 823755158 gb JXPA01002841.1	88.462	234	207	438	2852	3074
25Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823755158 gb JXPA01002841.1	93.671	158	1	158	2772	2928
		<i>hAT-1_DP</i>	gi 823755158 gb JXPA01002841.1	90.909	132	2568	2699	3217	3340
		<i>hAT-1_DP</i>	gi 823755158 gb JXPA01002841.1	81.333	150	2366	2512	3117	3262
		<i>hAT-1_DP</i>	gi 823754198 gb JXPA01003320.1	87.5	216	225	438	789	579
26Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823754198 gb JXPA01003320.1	95.683	139	2	140	868	730
		<i>hAT-1_DP</i>	gi 823754198 gb JXPA01003320.1	88.194	144	2556	2699	548	410
		<i>hAT-1_DP</i>	gi 823753229 gb JXPA01003796.1	90.129	233	207	438	31346	31120
27Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823753229 gb JXPA01003796.1	87.975	158	1	158	31416	31270
		<i>hAT-1_DP</i>	gi 823753229 gb JXPA01003796.1	90.152	132	2568	2699	30989	30866
		<i>hAT-1_DP</i>	gi 823753229 gb JXPA01003796.1	82.639	144	2370	2512	31083	30944
		<i>hAT-1_DP</i>	gi 823752269 gb JXPA01004264.1	93.038	158	1	158	9416	9573
28Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823752269 gb JXPA01004264.1	79.679	187	2368	2544	9749	9931
		<i>hAT-1_DP</i>	gi 823752269 gb JXPA01004264.1	75.51	147	2552	2698	9845	9977
		<i>hAT-1_DP</i>	gi 823751507 gb JXPA01004623.1	91.958	286	171	455	12806	13087
29Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823751507 gb JXPA01004623.1	96.392	194	1	194	12726	12919
		<i>hAT-1_DP</i>	gi 823751507 gb JXPA01004623.1	88.172	186	2514	2699	13102	13275
		<i>hAT-1_DP</i>	gi 823751415 gb JXPA01004667.1	87.554	233	207	438	17700	17926
30Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823751415 gb JXPA01004667.1	91.139	158	1	158	17620	17776
		<i>hAT-1_DP</i>	gi 823751415 gb JXPA01004667.1	78.109	201	2391	2578	18011	18206
		<i>hAT-1_DP</i>	gi 823751415 gb JXPA01004667.1	93.75	48	2652	2699	18191	18238
		<i>hAT-1_DP</i>	gi 823744060 gb JXPA01008136.1	89.6	250	207	455	87431	87676
31Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823744060 gb JXPA01008136.1	96.203	158	1	158	87351	87508
		<i>hAT-1_DP</i>	gi 823744060 gb JXPA01008136.1	100	58	2642	2699	87695	87752

		<i>hAT-1_DP</i>	gi 823742038 gb JXPA01009147.1	89.27	233	207	438	790	563
32Mir	<i>D. miranda</i>	<i>hAT-1_DP</i>	gi 823742038 gb JXPA01009147.1	74.924	331	2369	2699	520	245
		<i>hAT-1_DP</i>	gi 823742038 gb JXPA01009147.1	93.038	158	1	158	870	713
33Mir	<i>D. miranda</i>	<i>howilli1</i>	gi 823742036 gb JXPA01009148.1	89.286	140	1	140	2210	2343
		<i>howilli1</i>	gi 823742036 gb JXPA01009148.1	81.609	174	2653	2816	2633	2805
		<i>hAT-1_DP</i>	gi 1218744469 gb LYTC01003719.1	65.385	1014	373	1378	1331	373
1Nas	<i>D. nasuta</i>	<i>hAT-1_DP</i>	gi 1218744469 gb LYTC01003719.1	85.787	197	1	197	1721	1527
		<i>hAT-1_DP</i>	gi 1218744469 gb LYTC01003719.1	76	175	1993	2167	323	149
1Rho	<i>D. rhopaloa</i>	<i>homo2</i>	gi 449571599 gb AFPP02014479.1	86.31	168	1	165	518	684
		<i>homo2</i>	gi 449571599 gb AFPP02014479.1	81.72	93	2410	2502	1148	1240
1Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449840425 gb AFFI02000282.1	81.452	248	1	240	13251	13495
		<i>hAT-1_DP</i>	gi 449840425 gb AFFI02000282.1	81.905	210	2491	2699	13562	13771
2Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449839927 gb AFFI02000750.1	79.435	248	1	240	14428	14673
		<i>hAT-1_DP</i>	gi 449839927 gb AFFI02000750.1	79.882	169	2531	2699	14719	14881
3Dtak	<i>D. takahashii</i>	<i>homo2</i>	gi 449837134 gb AFFI02003295.1	90.714	140	1	139	9588	9450
		<i>homo2</i>	gi 449837134 gb AFFI02003295.1	80.882	136	2374	2502	9245	9111
4Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449835988 gb AFFI02004441.1	80.242	248	1	240	1349	1104
		<i>hAT-1_DP</i>	gi 449835988 gb AFFI02004441.1	82.822	163	2537	2699	1059	903
5Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449835734 gb AFFI02004695.1	81.452	248	1	240	6349	6594
		<i>hAT-1_DP</i>	gi 449835734 gb AFFI02004695.1	79.048	210	2499	2699	6688	6892
6Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449835240 gb AFFI02005189.1	83.168	202	1	198	7238	7039
		<i>hAT-1_DP</i>	gi 449835240 gb AFFI02005189.1	84.416	154	2546	2699	5483	5335
7Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449835138 gb AFFI02005291.1	80.816	245	1	237	7000	7241
		<i>hAT-1_DP</i>	gi 449835138 gb AFFI02005291.1	82.308	130	2571	2699	7319	7443
8Dtak	<i>D. takahashii</i>	<i>homo2</i>	gi 449834387 gb AFFI02006042.1	89.333	150	1	149	74	222
		<i>homo2</i>	gi 449834387 gb AFFI02006042.1	80.851	141	2369	2502	1836	1976
9Dtak	<i>D. takahashii</i>	<i>hAT-1_DP</i>	gi 449834040 gb AFFI02006389.1	79.435	248	1	240	791	546

	<i>takahashii</i>	<i>hAT-1_DP</i>	gi 449834040 gb AFFI02006389.1	76.735	245	1	240	1589	1362
10Dtak	<i>D.</i>	<i>hAT-1_DP</i>	gi 449834040 gb AFFI02006389.1	80	135	2537	2669	502	373
	<i>takahashii</i>	<i>hAT-1_DP</i>	gi 449834040 gb AFFI02006389.1	79.63	108	2562	2669	1316	1214
11Dtak	<i>D.</i>	<i>howilli1</i>	gi 449834018 gb AFFI02006411.1	81.776	214	2	206	2583	2795
	<i>takahashii</i>	<i>howilli1</i>	gi 449834018 gb AFFI02006411.1	82.906	117	2705	2816	2920	3036
12Dtak	<i>D.</i>	<i>homo2</i>	gi 449833532 gb AFFI02006897.1	83.735	166	1	161	167736	167572
	<i>takahashii</i>	<i>homo2</i>	gi 449833532 gb AFFI02006897.1	78.676	136	2375	2502	167507	167374
13Dtak	<i>D.</i>	<i>homo2</i>	gi 449833451 gb AFFI02006978.1	90.714	140	1	139	455980	456118
	<i>takahashii</i>	<i>homo2</i>	gi 449833451 gb AFFI02006978.1	80.882	136	2374	2502	456321	456455
14Dtak	<i>D.</i>	<i>homo2</i>	gi 449833280 gb AFFI02007149.1	83.929	168	11	172	8236	8402
	<i>takahashii</i>	<i>homo2</i>	gi 449833280 gb AFFI02007149.1	79.259	135	2369	2502	8557	8689
15Dtak	<i>D.</i>	<i>hAT-1_DP</i>	gi 449833106 gb AFFI02007323.1	83.168	202	1	198	272078	272277
	<i>takahashii</i>	<i>hAT-1_DP</i>	gi 449833106 gb AFFI02007323.1	79.333	150	2551	2699	272364	272508
16Dtak	<i>D.</i>	<i>homo2</i>	gi 449833005 gb AFFI02007424.1	89.333	150	1	149	147263	147411
	<i>takahashii</i>	<i>homo2</i>	gi 449833005 gb AFFI02007424.1	80.851	141	2369	2502	148947	149087
17Dtak	<i>D.</i>	<i>homo2</i>	gi 449832405 gb AFFI02008024.1	66.622	1483	2	1439	12010	10565
	<i>takahashii</i>	<i>homo2</i>	gi 449832405 gb AFFI02008024.1	73.585	265	1814	2078	10201	9937
		<i>homo2</i>	gi 449832405 gb AFFI02008024.1	84.783	92	2411	2502	8466	8375
18Dtak	<i>D.</i>	<i>homo2</i>	gi 449832237 gb AFFI02008192.1	86.747	166	1	161	56526	56362
	<i>takahashii</i>	<i>homo2</i>	gi 449832237 gb AFFI02008192.1	82.883	111	2392	2502	55495	55390
19Dtak	<i>D.</i>	<i>homo2</i>	gi 449832014 gb AFFI02008415.1	85.235	149	1	149	1961	1813
	<i>takahashii</i>	<i>homo2</i>	gi 449832014 gb AFFI02008415.1	82.524	103	2374	2472	1639	1543
20Dtak	<i>D.</i>	<i>homo2</i>	gi 449831597 gb AFFI02008832.1	86.145	166	1	161	66545	66381
	<i>takahashii</i>	<i>homo2</i>	gi 449831597 gb AFFI02008832.1	86.022	93	2410	2502	66193	66101
21Dtak	<i>D.</i>	<i>homo2</i>	gi 449830908 gb AFFI02009521.1	89.333	150	1	149	2049	1901
	<i>takahashii</i>	<i>homo2</i>	gi 449830908 gb AFFI02009521.1	80.851	141	2369	2502	282	142
22Dtak	<i>D.</i>	<i>hAT-1_DP</i>	gi 1219029458 gb LYTE01011028.1	67.888	819	150	955	2631	1845
	<i>takahashii</i>	<i>hAT-1_DP</i>	gi 1219029458 gb LYTE01011028.1	69.262	488	1067	1551	1754	1270

		<i>hAT-1_DP</i>	gi 1219029458 gb LYTE01011028.1	78.07	228	2	229	2693	2489
		<i>hAT-1_DP</i>	gi 1219029458 gb LYTE01011028.1	74.118	255	1922	2176	906	652
23Dtak	<i>D.</i>	<i>hAT-1_DP</i>	gi 1219029457 gb LYTE01011029.1	65.033	1979	168	2099	1901	1
	<i>takahashii</i>	<i>hAT-1_DP</i>	gi 1219029457 gb LYTE01011029.1	78.761	226	4	229	1962	1759

---

QI: query id

SI: subject ids

%ID: % identity

AL: alignment length

q. start: query start

q. end: query end

s. start: subject start

s. end: subject end

### CAPÍTULO III







## 6 DISCUSSÃO GERAL E PERSPECTIVAS

Apesar de elementos *hATs* serem amplamente distribuídos e vários estudos demonstrarem interações importantes para evolução dos genomas onde estão contidos, ainda temos poucos dados para avaliar a diversidade de sequências e os processos biológicos em que eles podem estar envolvidos.

Com o avanço das tecnologias de sequenciamento de genomas, e sabendo que praticamente todos os genomas apresentam TEs em grandes proporções, é provável que estejamos subestimando a diversidade de todas as superfamílias já descritas. Os elementos *hATs* descritos estão distribuídos em apenas três grandes famílias com base na similaridade de suas sequências proteicas, das TIRs e TSDs. Entretanto, parece que poucos elementos ativos são descritos, incluindo elementos ativos e bem estudados, como o transposon *Ac* de *Z. mays*, *hobo* de *D. melanogaster*, *Hermes* de *Musca domestica* e *Tol2* de *O. latipes* (ARENSBURGER et al., 2011; ATKINSON, 2015). A caracterização do conjunto de elementos móveis que ocorrem nos genomas das espécies com genoma sequenciado se restringe a estudos que fazem o levantamento da diversidade de famílias de TEs, indicando potencial capacidade de mobilização com base em características de sequências, mas poucos estudos avaliando possíveis efeitos provenientes da interação das sequências dos TEs nos genomas hospedeiros.

Em nossos resultados do primeiro artigo, reunimos dados sobre a história de invasão do elemento *hobo* a partir da observação de polimorfismos de repetições em tandem presentes na sequência da proteína do elemento (repetições TPEs). Essas repetições estão relacionadas com a eficiência da atividade da transposase, de modo que elementos com menor número de repetições (3 TPEs) apresentam uma vantagem na atividade da transposase (SOUAMES et al., 2003a). Sabendo disso e que o elemento *hobo* canônico apresentando as repetições TPEs invadiu recentemente o genoma das espécies cosmopolitas *D. simulans* e *D. melanogaster* (depois de 1950, segundo PERIQUET et al., 1989), um cenário evolutivo para a invasão foi proposto e sugere dois eventos de invasão independentes: a primeira invasão de elementos *hobo* com 3 TPEs e que se espalhou junto com as populações de *D. melanogaster* em todo o mundo (PERIQUET et al., 1989); e uma segunda invasão de elemento com 5 e 7 TPEs que é sugerida para ter ocorrido na América do Sul (BONNIVARD et al., 2000; 2002). Porém, como as populações polimórficas

ocorrem no Oeste da América do Sul, no Norte da África e Europa Equatorial, e estas duas últimas são regiões muito próximas a América do Sul seria a região mais provável de origem da segunda invasão de elementos com 5 e 7 TPE. Entretanto, as populações americanas de espécies de *D. melanogaster* e *D. simulans* investigadas por nós, não apresentam polimorfismo de repetições TPEs, de maneira que a América do Sul também passa a ser uma região excluída como potencial centro da suposta segunda invasão do elemento *hobo*.

Sabendo que as taxas de mutação das regiões de minissatélites TPEs em *D. melanogaster* são cem vezes maiores quando comparados com outras regiões do genoma de microsatélites neutros (SOUAMES et al., 2003a), nós sugerimos que a origem do polimorfismo nas populações da América do Sul, da África e da Europa Equatorial são resultantes de mutações que foram selecionadas ou sofreram deriva genéticas.

Com relação aos MnRs presentes em um grupo de elementos de *Drosophila*, os elementos *Herves-like*, nossos resultados demonstram uma distribuição ampla e antiga em algumas espécies analisadas. Sua persistência nos genomas se dá através de elementos MITEs, que hospedam em suas sequências as repetições em tandem (MnRs). Além disso, descobrimos que os MnRs do elemento *howilli1* estão presentes em pequenos RNAs que foram isolados de *D. willistoni* e que podem ser representantes da família miR-190 de miRNAs, amplamente distribuídos em outras espécies de plantas e animais (ZHENG et al., 2010). Entretanto, análises de frequência com bibliotecas de pequenos RNAs ainda podem ser realizadas para podermos estimar ou validar esses pequenos RNAs.

Em *Drosophila*, sabe-se que a superexpressão de miRNAs da família miR-190 aumenta os fatores induzíveis de hipóxia, e em embriões em desenvolvimento que apresentam perda de função desse miRNAs a resposta à hipóxia é prejudicada (DE LELLA EZCURRA et al., 2016). Uma predição de alvos para o pequenos RNAs isolado de *D. willistoni*, que apresentam as sequências de MnRs do elemento *howilli1*, indicou genes alvos envolvidos em diferentes processo celulares, mas estudos experimentais adicionais são necessários para validar os alvos desses miRNAs.

De qualquer forma, a descoberta de pequenos RNAs provenientes de minissatélites de elementos móveis nos levou a investigar a distribuição e história evolutiva desses elementos com MnRs. Um cenário evolutivo foi proposto, onde

mostramos que, em *Drosophila*, todos os elementos que chamamos de *Herves-like* apresentam MnRs em suas sequências. Esses elementos são amplamente distribuídos em espécies de *Drosophila*, sendo que os elementos *hAT-1\_DSi* presentes em espécies do grupo *melanogaster* são os elementos provenientes de invasões mais antigas, mas que preservam essas repetições MnRs em suas sequências de elementos *MITE Herves-like*. Encontramos sinal de presença de elementos *Herves-like* provenientes de invasões possivelmente mais recentes, com presença de elementos degenerados similares aos elementos *hAT-1DP* e *MITEs Herves-like*, em espécies asiáticas do grupo *melanogaster* (*D. takahashii*, *D. rhopaloa*, *D. bipectinata* e *D. ananassae*), *virilis* (*D. virilis*) e *immigrans* (*D. nasuta* e *D. immigrans*). Também encontramos elementos com sequências altamente similares, tanto a região codificadora da proteína quanto os elementos curtos e regiões de MnRs, em vários grupos de espécies de *Drosophila*, sugerindo atividade recente e a ampla distribuição possivelmente se dá por eventos de HT.

Apesar de sugestões recentes de que os TEs são as fontes principais de miRNAs (ROBERTS et al., 2014), e apesar de sabermos que os efeitos gerados a partir de atividade de TEs podem afetar a diversidade dos genomas hospedeiros, ainda não sabemos como os pequenos RNAs derivados de TE podem afetar os padrões globais de expressão gênica e, conseqüentemente, a diversificação de espécies (HOFFMANN et al., 2015). Nossos resultados reúnem informações sobre diferentes sequências repetitivas curtas, oriundas dos elementos *hATs* investigados, demonstramos que essas sequências podem influenciar no ciclo evolutivo dos TEs, bem como agregar funções biológicas, como potenciais influenciando de forma indireta nas redes regulatórias por meio de pequenosRNAs provenientes de suas sequências. Sugerimos que a persistência de elementos *MITE Herves-like* nos genomas pode estar acontecendo por seleção/domesticação devido as interações que podem acontecer quando estes elementos servem como seuqências precurosras de pequenos RNAs.

A partir dos resultados gerados nesta tese, nossas perspectivas são com relação a investigação de pequenos RNAs nos genomas de *D. willistoni*, através da criação de bibliotecas de pequenos RNAs para análises investigativas sobre TEs como geradores de pequenos RNAs, bem como funcionalidade e previsão de alvos de silenciamento através de ensaios biológicos.



## REFERÊNCIAS

- ARENSBURGER, P.; PIÉEGU, B.; BIGOT, Y. The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine? **Mob. Genet. Elements**. v. 6, n. 6, p. e1256852. 2016.
- ARENSBURGER, P. et al. Phylogenetic and functional characterization of the hAT transposon superfamily. **Genetics**. v. 188, p. 45-57. 2011.
- ATKINSON, P. W. hAT transposable elements. **Microbiol. Spectrum**. v. 3, n. 4, p. MDNA3-0054-2014, 2014.
- AXTELL, M. J.; WESTHOLM, J. O.; LAI, E. C. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. **Genome Biol.** v. 12, n. 4, p. 221, 2011.
- BAO, W.; KOJIMA, K. K.; KOHANY, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. **Mobile DNA**, v. 6, p. 11, 2015.
- BERNARDO, L. P.; LORETO, E. L. S. *hobo*-brothers elements and their time and place for horizontal transfer. **Genetica**, v. 141, p. 471-478. 2013.
- BIÉMONT. A brief history of the status of transposable elements: from junk DNA to major players in evolution. **Genetics**, v. 186, n. 4, p. 1085-1093, 2010.
- BLACKMAN, R. K. et al. Identification of a fully functional hobo transposable element and its use for germ-line transformation of *Drosophila*. **EMBO J.** v. 8, p. 211-217, 1989.
- BONNIVARD, E. et al. A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. **Genet. Res.** v. 75, p. 13-23, 2000.
- BONNIVARD, E.; BAZIN, C.; HIGUET, D. High polymorphism of TPE repeats within natural populations of *D. melanogaster*: a gradient of the 5TPE *hobo* in Western Europe. **Mol. Biol. Evol.** v. 19, p. 2277-2284, 2002.
- BREGLIANO, J.; KIDWELL, M. Hybrid dysgenesis. In: Shapiro J, editor. Mobile genetic elements. New York: **Academic Press**, p. 363-410, 1983.
- BRENNECKE, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. **Cell**, **128**, v. 6, p. 1089-1103, 2007.
- CÁCERES, M.; PUIG, M.; RUIZ, A. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. **Genome research**, v. 11, n. 8, p. 1353-1364, 2001.
- CAPY, P. et al. Dynamics and evolution of transposable elements. **Landes Bioscience**, Austin, Texas. 1997.

CHAPMAN, E. J.; CARRINGTON, J. C. Specialization and evolution of endogenous small RNA pathways. **Nature Rev. Genet.** v. 8, p. 884-896, 2007.

CHEN, J.; GREENBLATT, I. M.; DELLAPORTA, S. L. Molecular analysis of Ac transposition and DNA replication. **Genetics.** v. 130, n. 3, p. 665-676, 1992.

CZECH, B.; HANNON, G. J. Small RNA sorting: matchmaking for Argonautes. **Nat Rev Genet.** v. 12, n. 1, p. 19-31, 2011.

DANIELS, S. B. et al. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics*, v. 124, p. 339-355, 1990a.

\_\_\_\_\_. A. Distribution of *hobo* transposable elements in the genus *Drosophila*. **Mol. Biol. Evol.** v. 7, p. 589-606, 1990b.

DEPRÁ, M. et al. *hosimary*: a new hAT transposon group involved in horizontal transfer. **Mol Genet Genomics.**, v. 283, p. 451-459, 2010.

DEPRÁ, M. et al. *Mar*, a MITE family of *hAT* transposons in *Drosophila*. **Mobile DNA**, v. 3, p. 13, 2012.

DIAS, G. B.; HERINGER, P.; KUHN, G. C. S. *Helitrons* in *Drosophila*: Chromatin modulation and tandem insertions. **Mob. Genet. Elements.**, v. 6, n. 2, p. e1154638, 2016

EZCURRA, A. L. D. L. et al. *miR-190* Enhances HIF-Dependent Responses to Hypoxia in *Drosophila* by Inhibiting the Prolyl-4-hydroxylase Fatiga. **PLoS Genet.** v. 12, n. 5, p. e1006073, 2016.

FABIAN, M. R.; SONENBERG, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct.* **Mol. Biol.** v. 19, p. 586-593, 2012.

FESCHOTTE, C.; PRITHAM, E. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* v. 41, p. 331-368, 2007.

FESCHOTTE, C. The contribution of transposable elements to the evolution of regulatory networks. **Nat Rev Genet**, v. 9, p. 397-405, 2008.

FESCHOTTE, C.; ZHANG, X.; WESSLER, S. R. Miniature inverted-repeat transposable elements and their relationships to established DNA transposons, pp. 1147–1158 in *Mobile DNA II*, 2002, edited by N. Craig, R. Craigie, M. Gellert and A. Lambowitz. **American Society for Microbiology Press**, Washington, DC.

FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends in Genetics.**, v. 5, p. 103-107, 1989.

\_\_\_\_\_. Transposable elements. *Curr. Opin. Genet. Dev.* v. 2, p. 861-867, 1992.

GAO, C. H. et al. Characterization and functional annotation of nested transposable elements in eukaryotic genomes. **Genomics**, v. 100, p. 222-230, 2012.

- GARDNER, J. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. **Nature**. v. 419, n. 6906, p. 498-511, 2002.
- GENDREL, A. V. et al. Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. **Science**. v.297, p. 1871-1873, 2002.
- GHILDIYAL, M.; ZAMORE, P. D. Small silencing RNAs: an expanding universe. **Nature Rev. Genet.** v. 10, 94-108, 2009.
- GILBERT, C. et al. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. **Nature**, v. 464, p. 1347-1350, 2010.
- GRABUNDZIJA, I. et al. Comparative analysis of transposable element vector systems in human cells. **Mol Ther**, v. 18, p. 1200-1209, 2010.
- GREENBERG, R. A. Histone tails: Directing the chromatin response to DNA damage. **FEBS letters**. **585**, v. 18, p. 2883-2890, 2011.
- GUERREIRO, M. P. What makes transposable elements move in the Drosophila genome? **Heredity**, v. 108, p. 461-468, 2011.
- HA, M.; KIM, V. N. Regulation of microRNA biogenesis. **Nature reviews Molecular cell biology**. v. 15, n. 8, p. 509-524, 2014.
- HABU, Y.; HISATOMI, Y.; IIDA, S. Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. **Plant. J.** v. 16, p. 371-376, 1998.
- HASSA, P. O.; MICHAEL, O.; HOTTIGER. An epigenetic code for DNA damage repair pathways?. **Biochemistry and cell biology**. v. 83, n. 3. p. 270-285, 2005.
- HEHL, R. et al. Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveal homologies to the *Ac* element from maize. **Plant. Mol. Biol.** v. 16, p. 369-371, 1991.
- HICKMAN, A. B. et al. Molecular architecture of a eukaryotic DNA transposase. **Nat. Struct. Mol. Biol.** v.12, p. 715-721, 2005.
- HOFFMANN, F. G. et al. Transposable elements and small RNAs: Genomic fuel for species diversity. **Mobile Genetic Elements**. v. 5, n. 5, p. 63-66, 2015.
- HUA-VAN, A. et al. The struggle for life of the genome's selfish architects. **Biol. Direct**. v. 6, p. 19, 2011.
- HUDA, A. A. et al. Epigenetic regulation of transposable element derived human gene promoters. **Gene**. 2011.
- HUISINGA, K. L.; BROWER-TOLAND, B.; ELGIN, S. C. The contradictory definitions of heterochromatin: transcription and silencing. **Chromosoma**. v.115, n. 2, p. 110-122. 2006.

JARDIM, S. S. et al. Effects of heat and UV radiation on the mobilization of transposon *mariner-Mos1*. **Cell Stress and Chaperones**. v.20, p. 843-851, 2015.

JIANG, N. et al. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin. Plant Biol.* v. 7, p. 115-119, 2004.

JURKA, J. hAT-type DNA transposons from *Hydra magnipapillata*, **Rebase Rep.** v. 8 p. 2001, 2008.

JURKA, J.; WEIDONG, B.; KENJI, K. KOJIMA. "Families of transposable elements, population structure and the origin of species. **Biology Direct**. v. 6, n. 1, p. 44, 2011.

KALMYKOVA, A. I.; KLENOV, M. S.; GVOZDEV, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. **Nucleic Acids Res**, v. 33, p. 2052-2059, 2005.

KAPITONOV, V. V.; JURKA, J. A universal classification of eukaryotic transposable elements implemented in Rebase. **Nat Rev Genet**. v. 9, p. 411-412, 2008.

\_\_\_\_\_. Rolling-circle transposons in eukaryotes. **Proc. Natl Acad. Sci. USA**. v. 98, p. 8714-8719, 2001.

KAZAZIAN, H. H. Mobile DNA transposition in somatic cells. **BMC biology**. v. 9, n. 1, p. 62, 2011.

KIM, V. N.; HAN, J.; SIOMI, M. C. Biogenesis of small RNAs in animals. **Nat Rev Mol Cell Biol.**, v. 10, p. 126-139, 2009.

LADEVÉZE, V. et al. *Hobo* transposons causing chromosomal breakpoints. **Proc R Soc Lond B**. v. 265, p. 1157-1159, 1998b.

\_\_\_\_\_. General survey of *hAT* transposon superfamily with highlight on hobo element in *Drosophila*. **Genetica**, v. 140, p. 375-392, 2012.

LE ROUZIC, A.; CAPY, P. Theoretical approaches to the dynamics of transposable elements in genomes, populations, and species. In: LANKENAU, D-H.; VOLFF, J-N. (Eds.). **Transposons and the dynamic genome**. Springer, Berlin, p 1-19, 2009.

LEVIN, H. L.; MORAN, J. V. Dynamic interactions between transposable elements and their hosts. **Nat. Rev. Genet.** v. 12, p. 615-27, 2011.

LI, X. et al. A resurrected mammalian hAT transposable element and a closely related insect element are highly active in human cell culture. **Proc. Natl Acad. Sci. USA**, v. 110, p. E478-E487, 2013.

LORETO, E. L. S.; PEREIRA, C. P. Somatizing the transposons action. **Mobile Genetic Elements**. v. 7, n. 3, p. 1-9, 2017.



LU, C. et al. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. **Mol. Biol. Evol.** v. 29, p.1005-1017, 2012.

MALONE, C. D.; HANNON, G. J. Small RNAs as guardians of the genome. **Cell.** **136**, v. 4, p. 656-668, 2009.

MARTENS, J. H. et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. **The EMBO journal.** **24**, v. 4, p. 800-812, 2005.

MATZKE, M. A.; METTE, M. F.; MATZKE, A. J. Transgene silencing by the host genome 784 defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. **Plant. Mol. Biol.** v. 43, p. 401-415, 2000.

MCCLINTOCK, B. Cytogenetic studies of maize and *Neurospora*. **Carnegie Inst Washington Year Book**, v. 46, p. 146-152, 1947.

\_\_\_\_\_. Mutable loci in maize. **Carnegie Institute of Washington Year Book**, v. 47, p. 155-169, 1948.

\_\_\_\_\_. The origin and behavior of mutable loci in maize. **Proc Natl Acad Sci, USA** v. 36, n. 6, p. 344-355, 1950.

MCGINNIS, W.; SHERMOEN, A. W.; BECKENDORF, S. A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. **Cell**, v. 34, p. 75-84, 1983.

MOTA, N. R. et al. *Harrow*: new *Drosophila hAT* transposons involved in horizontal transfer. **Insect Mol Biol.** v. 19, p. 217-228, 2010.

NEWMAN, M. A.; MANI, V.; HAMMOND, S. M. Deep sequencing of microRNA precursors reveals extensive 3' end modification. **RNA**, v. 17, p. 1795-1803, 2011.

ORTIZ, M. F. et al. *hAT* transposable elements and their derivatives: an analysis in the 12 *Drosophila* genomes. **Genetica.** v. 138, n. 6, p. 649-655, 2010.

PACE, J. K. et al. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. **Proc Natl Acad Sci, USA.** v. 105, p. 17023-17028, 2008.

PASCUAL, L.; PERIQUET, G. Distribution of hobo trans-posable elements in natural populations of *Drosophila melanogaster*. **Mol Biol Evol**, v. 8, p. 282-296, 1991.

PERIQUET, G. et al. Geographical and historical patterns of distribution of hobo elements in *Drosophila melanogaster* populations. **J. Evol. Biol.** v. 2, p. 223-229, 1989.

PICARDEAU, M. Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis. **Genetica**, v. 138, p. 551-558, 2010.

PIÉGU, B. et al. A survey of transposable element classification systems – a call for a fundamental update to meet the challenge of their diversity and complexity. **Mol. Phylogenet Evol.** v. 86, p. 90-109, 2015.

PIRIYAPONGSA, J.; JORDAN, I.K. A family of human microRNA genes from miniature inverted-repeat transposable elements. **PLoS One**, v. 2, n. 2, p. e203, 2007.

PIRIYAPONGSA, J.; MARINO-RAMIREZ, L.; JORDAN, I. K. Origin and evolution of human microRNAs from transposable elements. **Genetics**, v. 176, p. 1323-1337, 2007.

RAY, D. A. C. et al. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. **Genome Research**, v. 18, p. 717-728, 2008.

REBOLLO, R.; ROMANISH, M. T.; MAGER, D. L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. **Annu Rev Genet**, v. 46, p. 21-42, 2012.

ROBERTS, J. T.; CARDIN, S. E.; BORCHERT, G. M. Burgeoning evidence indicates that microRNAs were initially formed from transposable element sequences, **Mob. Genet. Elements**, v. 4, p. e29255, 2014.

ROSSATO, D. O. et al. BuT2 is a member of the third major group of hAT transposons and is involved in horizontal transfer events in the genus *Drosophila*. *Genome biology and evolution*. v. 6, n. 2, p. 352-365, 2014.

RUBIN, E.; LITHWICK, G.; LEVY, A. A. Structure and evolution of the hAT transposon superfamily. **Genetics**, v. 158, p. 949-957, 2001.

SCHNABLE, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. **Science**, v. 326, n. 5956, p. 1112-1115, 2009.

SHEEN, F.; LIM, J. K.; SIMMONS, M. J. Genetic instability in *Drosophila melanogaster* mediated by hobo transposable elements. **Genetics**, v. 133, p. 315-334, 1993.

SHEN, J. et al. Translational repression by a miniature inverted-repeat transposable element in the 3' untranslated region. **Nature Communications**. v. 8, 2017.

SOUAMES, S. et al. Behavior of the hobo transposable element with regard to TPE repeats in transgenic lines of *Drosophila melanogaster*. **Mol Biol Evol.**, v. 20, p. 2055-2066, 2003a.

VARANI, A. M. et al. Elementos de transposição: classificação e mecanismos de mobilização. In: CARARETO, C. M. A.; VITORELLO, C. B. M.; SLUYS, M-A V. **Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos**. 22 ed. Rio de Janeiro. Fiocruz. cap. 1, p. 11-42, 2015.

VENNER, S.; FESCHOTTE, C.; BIE'MONT, C. Dynamics of transposable elements: towards a community ecology of the genome. **Trends Genet.** v. 25, p. 317-323, 2009.

VOLFF, J. N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. **Bioessays**, v. 28, p. 913-922, 2006.

WALLAU, G. et al. Genomic landscape and evolutionary dynamics of mariner transposable elements within the *Drosophila* genus. **BMC Genomics**, v. 15, p. 727, 2014.

WALLAU, G. L.; ORTIZ, M. F.; LORETO, E. L. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. **Genome Biol Evol**, v. 4, p. 689-699, 2012.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nat Rev Genet**, v. 8, p. 973-982, 2007.

YAMANAKA, S.; SIOMI, M. C.; SIOMI, H. piRNA clusters and open chromatin structure. **Mobile DNA**, v. 5, n. 1, p. 22, 2014.

YANNOPOULOS, G. et al. *hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila mela-nogaster* bearing the male recombination factor 23SMRF. **Cell**, v. 49, p. 487-495, 1987.

YUAN, Y-W.; WESSLER, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. **Proc Natl Acad Sci., USA**, v. 108, p. 7884-7889, 2011.

ZENG, Y. et al.  $\mu$ -Opioid Receptor Agonists Differentially Regulate the Expression of miR-190 and NeuroD. **Mol Pharmacol**. v. 77, n. 1, p. 102-109, 2010.

ZHANG, H-H. et al. "The origin and evolution of six miniature inverted-repeat transposable elements in *Bombyx mori* and *Rhodnius prolixus*. **Genome biology and evolution**, v. 5, n. 1), p. 2020-2031, 2013.

ZHOU, L. et al. Transposition of hAT elements links transposable elements and V(D)J recombination. **Nature**. v. 432, p. 995-1001, 2004.