

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

ELEMENTOS *hAT* DE *Drosophila*:
Análise de expressão e distribuição

MARÍNDIA DEPRÁ

Tese de Doutorado submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para obtenção do grau de Doutor em Ciências.

Orientador: Prof. Dr. Élgion Lúcio Loreto

Co-orientadora: Prof. Dra. Vera Lúcia da Silva Valente

Porto Alegre, abril de 2009.

Este trabalho foi realizado no laboratório de *Drosophila* do Departamento de Genética da Universidade Federal do Rio Grande do Sul e no laboratório de *Drosophila* do Departamento de Biologia da Universidade Federal de Santa Maria, com o auxílio de recursos do CNPq.

Dedico esta tese a toda minha
família, especialmente ao Tiago e
à minha mãe.

AGRADECIMENTOS

Ao Prof. Elgion, pela oportunidade de iniciação científica, pela orientação constante, por ser um exemplo de profissional, pela ajuda e paciência que sempre teve comigo – Obrigada por tudo prof.!

À Verinha, pela orientação, pelo incentivo e carinho dedicados durante esta etapa, e por também ser um exemplo de profissional – uma “imortal”.

Aos colegas, meus grandes e queridos amigos, do eterno LabDros que sempre fizeram deste local uma segunda casa para mim. Ao Paulo Marcos Pinto, por estar sempre disposto a ajudar. Ao Ronaldo M. Golombieski, por ser sempre um ombro amigo em todos os momentos. À Lizandra J. Robe, por me aguentar como um muro das lamentações quando tudo dava errado, e por tudo o que me ensinou e ajudou incondicionalmente. Ao Daniel Graichen, Newton Vidal e Marcos O. de Carvalho, por estarem sempre dispostos a ajudar. Meus amigos, muito obrigada a todos vocês!

Aos antigos colegas do Laboratório de *Drosophila* da UFRGS: Adri Sassi, Ana, Claudia, Fabiano, Marco, Mônica e Rosane; por me ensinarem muitas das coisas que sei hoje. À Mary, que acompanha essa caminhada desde o mestrado e que também me ensinou muito. Aos novos colegas do Laboratório de *Drosophila* da UFRGS: Carol, Cleverton, Dirleane, Gilberto, Gisele, Hermes, Ju Cordeiro, Ju WG e Mario, pelas risadas, pelas festas, pela alegria, ajuda e parceria, que fazem deste local um bom ambiente de trabalho.

À Adri Ludwig e à Nina por tudo o que me ajudaram e me ensinaram, por estarem sempre presentes e disponíveis e acima de tudo pela amizade.

À Lari juntamente com os colegas da Imuno e da Vegetal, por estarem sempre dispostos a dar aquela força necessária em muitas horas.

Às amigas: Ana Paula, Renata e Louise, pela amizade, parceria, compreensão, e por esse grande coleguismo.

Aos funcionários do Laboratório de *Drosophila* pelo grande auxílio.

Aos Professores do Programa de Pós-Graduação em Genética e Biologia Molecular.

Ao Elmo e à Elen, por serem o ombro amigo de todos os momentos.

Aos meus amigos erechinenses, por me fazerem esquecer por alguns instantes das tensões presentes nesta etapa da minha formação.

À minha família que eu amo tanto, pela confiança depositada, pelo conforto, segurança e alegria que sempre me proporcionaram.

Ao Tiago pelo carinho, amor, paciência, compreensão, auxílio, enfim, por ser meu companheiro sempre, em todos os momentos. Obrigada por tudo!

Todos vocês, de uma forma ou de outra, contribuíram para a realização deste trabalho.

SUMÁRIO

Abreviaturas, símbolos e unidades.....	7
Resumo.....	8
Abstract.....	9
CAPÍTULO 1.....	10
I – Introdução.....	10
Transposons.....	10
Classificação dos TEs.....	10
Dinâmica dos TEs.....	11
Implicações evolutivas dos TEs nos genomas.....	13
Silenciamento de Transposons.....	15
A Superfamília <i>hAT</i>	18
O Elemento Transponível <i>hobo</i>	19
O elemento <i>hobo</i> ^{VA}	21
O Elemento <i>hobo</i> em <i>D. simulans</i> e Hipermutabilidade.....	22
Expressão de TEs durante o desenvolvimento de <i>Drosophila</i>	23
II - Objetivos.....	25
CAPÍTULO 2.....	26
<i>hosimary</i> : a new <i>hAT</i> transposon family of <i>Zaprionus indianus</i> and <i>Drosophila melanogaster</i>	26
CAPÍTULO 3.....	50
The <i>hobo</i> Transposon and <i>hobo</i> -Related Elements are Expressed as Developmental Genes in <i>Drosophila</i>	50
CAPÍTULO 4.....	70
III - Discussão.....	70
Família <i>hosimary</i>	70
Família <i>hobo</i>	74
Referências Bibliográficas.....	79
Anexo 1.....	86

Abreviaturas, símbolos e unidades.

aa	aminoácido
bp	pares de base
cDNA	DNA complementar
DNA	ácido desoxirribonucleico
kb	kilobases
kDa	kilodalton
µg	micrograma
µl	microlitro
mM	milimolar
min	minuto
mRNA	RNA mensageiro
MYA	milhões de anos
ng	nanograma
ORF	módulo aberto de leitura
pb	pares de base
PCR	reação de polimerização em cadeia
RNA	ácido ribonucléico
RT-PCR	PCR com transcriptase reversa
s	segundos
TE	elemento de transposição
TIR	repetição terminal invertida
TFBS	sítio de ligação de fator de transcrição
TSD	duplicação do sítio alvo
U	unidade
UTR	região não traduzida

Resumo

Os elementos de transposição (TEs) são ubíquos e abundantes, presentes nos mais diversos genomas analisados até então. Neste trabalho foram estudadas duas famílias de elementos (*hosimary* e *hobo*) pertencentes à superfamília *hAT* de transposons que possui membros amplamente distribuídos. Descrevemos aqui uma nova família de TEs denominada *hosimary* que surgiu da análise de sequências inicialmente descritas por nosso grupo de pesquisa, e que foram denominadas *hosim* e *hosec*. Através de análise de PCR sequências homólogas a *hosim* e *hosec* foram buscadas no genoma de 51 espécies de drosofilídeos e foram detectadas apenas em espécies do subgrupo *melanogaster* de *Drosophila* e em *Zaprionus indianus*. Entre essas espécies o número de cópias observadas mostrou-se variável, sendo que a maioria das sequências mostrou-se potencialmente codificadora. A alta similaridade entre essas sequências de dois gêneros distintos, juntamente com inconsistências observadas entre a filogenia da espécie hospedeira e dos TEs, pode ser um indicativo de transferência horizontal desses elementos. Adicionalmente, as sequências apresentaram uma similaridade maior que 90%, e baixa similaridade com outros elementos descritos, o que nos levou a sugerir que essas sequências devam constituir uma nova família, a qual denominamos *hosimary*. A outra família analisada neste trabalho inclui os elementos *hobo* e *hobo*^{VA} (ou *hobo*^{VAHS}), que podem estar envolvidos em um fenômeno de hipermutabilidade observado por nosso grupo de pesquisa. Através da análise de expressão transcricional dos elementos foram detectados transcritos senso e antisenso, sugerindo a atuação de um mecanismo de interferência por RNA (iRNA) como controlador da atividade destes elementos. Através de hibridização *in situ* de embriões inteiros o padrão de expressão de *hobo* e *hobo*^{VAHS} revelou similaridades com padrões observados para genes do desenvolvimento. Isso nos levou a sugerir a existência de sequências *cis*-reguladoras nos elementos que podem estar interagindo com genes do hospedeiro. A presença destas sequências *cis*-reguladoras em TEs pode ter implicações evolutivas, uma vez que podem ser geradoras de variabilidade genética.

Abstract

Transposable elements (TEs) are ubiquitous and abundant in several analyzed genomes. In this work two TE families (*hosimary* and *hobo*), belonging to the *hAT* superfamily, were studied. Here is described a new family, called *hosimary*, that was arised of analyzes of two sequences initially described by our research group (*hosim* e *hosec*). By performing PCR analysis in 50 drosophilidae species, sequences homologous to *hosim* and *hosec* were detected in *Drosophila* species of the *melanogaster* subgroup and in *Zaprionus indianus*. The number of copies observed among these species showed to be variable and most sequences presented coding potential. High similarity between these sequences, which belong to two distinct genera, as well as some inconsistencies among the phylogeny of the host species and the TEs were observed. This can be an indicative of horizontal transfer of these elements. Additionally, the sequences presented more than 90% of mutual similarity and no significant similarity with other already described element. This led us to suggest this new transposon family, that we called *hosimary*. The other family analyzed in this work includes the *hobo* and *hobo*^{VA} (or *hobo*^{VAHS}) elements. These elements may be involved in a hypermutability phenomenon observed by our research group. Sense and antisense transcripts were detected by transcriptional expression analysis, which suggests an RNA interference mechanism (iRNA) controlling the activity of these elements. *in situ* hybridization of whole mount embryo analysis showed similarity of the *hobo* and *hobo*^{VAHS} expression pattern with patterns observed in developmental genes. This led us to suggest the existence of *cis*-regulatory sequences into the elements that can be interacting with the host genes. The presence of these *cis*-regulatory sequences in TEs can be related to evolutionary issues since they can be responsible for genetic variability.

CAPÍTULO 1

I – INTRODUÇÃO

TRANSPOSONS

Elementos de transposição (TEs – do inglês *transposable elements*) são sequências de DNA que possuem a capacidade, intrínseca ou não, de mudar sua localização dentro dos genomas. Eles compartilham as seguintes características: causam duplicação dos sítios de inserção, apresentam polimorfismo de sítios de inserção e apresentam uma variabilidade no número de cópias dentro e/ou entre espécies (Capy *et al.*, 1998).

Os TEs possuem uma ampla distribuição, representando cerca de 77% do genoma de *Rana esculenta*, 60% do genoma de *Zea mays* e 45% do genoma de humanos. Na bactéria *Escherichia coli* cerca de 0,3% do reduzido genoma parece ser composto por elementos de transposição (Biémont e Vieira, 2006). Em *Drosophila* a quantidade de transposons é variável, representando cerca de 2,7% dos genomas de *D. simulans* e *D. grimshawi* até 25% no genoma de *D. ananassae* (Clark *et al.*, 2007).

A presença dos TEs nos mais diversos genomas e suas relações com os genomas hospedeiros pode gerar algumas implicações evolutivas. Estes elementos são apontados como um mecanismo gerador de variabilidade, provável sistema gerador de isolamento reprodutivo e de transmissão horizontal de informação genética entre espécies (Kidwell e Lisch, 2001).

CLASSIFICAÇÃO DOS TEs

Na proposta de classificação dos TEs sugerida por Wicker *et al.* (2007) estes são agrupados hierarquicamente em: classes, subclasses, ordens, superfamílias, famílias e subfamílias, de acordo com algumas características. Primeiramente os TEs são divididos em duas grandes classes de acordo com o mecanismo pelo qual ocorre a transposição: por meio de DNA ou através de um intermediário de RNA. A Classe I contém os elementos que se transpõem via um intermediário de RNA, sem a clivagem do elemento do sítio doador – são os chamados retrotransposons. O intermediário de RNA é transcrito reversamente em DNA pela

enzima transcriptase reversa. Cada ciclo completo de replicação produz uma nova cópia. Esta classe pode ser dividida em cinco ordens, com base em características de seu mecanismo, organização e filogenia da transcriptase reversa. Uma destas ordens é representada pelos retrotransposons com LTRs (do inglês *long terminal repeats*), que apresenta como exemplo os elementos *gypsy* e *copia*. Outras duas ordens são representadas pelos retrotransposons sem LTRs ou retroposons, que são LINEs (do inglês *long interspersed nuclear elements*) e SINEs (do inglês *short interspersed nuclear elements*). E as duas últimas ordens incluem os elementos relacionados a *DIRS* (*DIRS-like*) e elementos relacionados a *Penelope* (PLEs - *Penelope-like*).

A Classe II compreende os elementos que se transpõem diretamente de DNA a DNA; são os transposons de DNA propriamente ditos. Estes elementos utilizam uma enzima denominada transposase para se mobilizarem e são divididos em duas subclasses, de acordo com o número de fitas de DNA que são clivadas no processo de transposição. Os elementos da Subclasse 1 clivam ambas as fitas de DNA durante a transposição por um mecanismo denominado “corta-e-cola” (do inglês *cut-and-paste*). Esta subclasse é composta pela ordem TIRs (do inglês *Terminal Inverted Repeats*, ou seja, que possuem repetições terminais invertidas) representada por superfamílias como *P*, *hAT*, *Transib* e *piggyBac*; e pela ordem *Crypton* (que não possui TIRs). Já a Subclasse 2 compreende os TEs que se transpõem clivando apenas uma das fitas de DNA, e inclui a Ordem dos *Helitrons*, que aparentemente replicam-se pelo mecanismo de “círculo-rolante”, e a Ordem dos *Mavericks*, que são elementos longos (até 20 kb) e não possuem mecanismo de transposição completamente elucidado.

DINÂMICA DOS TES

Os elementos de transposição podem ter duas possíveis origens. Eles podem ser derivados de cópias presentes em uma espécie ancestral por transmissão vertical, ou podem ser oriundos de outra espécie distante evolutivamente, através de transferência horizontal (HT – do inglês *Horizontal Transfer*). Alternativamente, elementos móveis podem se originar *de*

novo por eventos de mutação e de recombinação, que alteram ou reorganizam sequências já presentes no genoma (Capy *et al.*, 1998).

O “ciclo de vida” de um TE pode ser considerado como uma sucessão de três fases: replicação dinâmica, inativação/adaptação e degradação (revisão em Kidwell e Lisch, 1997; Almeida e Carareto, 2005a). O primeiro passo na introdução de um TE em um genoma é o estabelecimento de uma fase invasora. A partir do momento que uma sequência se insere em um novo genoma (seja por origem *de novo* ou transferência horizontal), o número de cópias tende a aumentar, sendo o número final dependente de algumas características do genoma hospedeiro. A interação estabelecida entre elemento e genoma hospedeiro define a sobrevivência de ambos. A segunda “fase da vida” de um TE é caracterizada por sua inativação, e depende, em grande parte, do impacto do elemento sobre o valor adaptativo do hospedeiro. Finalmente, ocorre um estágio de senescência que pode se estender possivelmente por milhões de anos. Neste estágio todos os elementos autônomos são perdidos, não ocorre aumento do número de cópias, e as sequências não-autônomas são perdidas, deletadas ou divergem (Pinsker *et al.*, 2001; Brookfield, 2005).

Neste cenário, vários autores têm sugerido o mecanismo de transferência horizontal como um meio dos transposons escaparem da fase de degradação, que pode levar à completa extinção do transposon.

Transferência horizontal pode ser definida como a transferência de material genético entre espécies isoladas reprodutivamente (Capy *et al.*, 1998). Diversos casos de HT vêm sendo reportados no gênero *Drosophila* para elementos de transposição como: *mariner* (Maruyama e Hartl, 1991; Brunet *et al.*, 1999), *Minos* (Arca e Savakis, 2000; Almeida e Carareto, 2005b), *hobo* (Daniels *et al.*, 1990; Simmons, 1992), *gypsy* (De Frutos *et al.*, 1992; Herédia *et al.*, 2004; Ludwig *et al.*, 2008), e *gtwin* (Kotnova *et al.*, 2005; Ludwig e Loreto, 2007) – revisão em Loreto *et al.* (2008).

Quando os TEs são transmitidos verticalmente espera-se que suas histórias filogenéticas retratem, ao menos em parte, aquelas de seus hospedeiros (Silva *et al.*, 2004). Assim, além de sobreposição geográfica, ecológica e temporal entre as espécies possivelmente envolvidas, três tipos de evidências são normalmente usadas para inferir eventos de HT de TEs: 1) a detecção de elementos de transposição com alto grau de similaridade de sequência

entre espécies distantes evolutivamente; 2) a observação de diferenças topológicas entre as filogenias dos TEs e de suas espécies hospedeiras; e 3) a distribuição descontínua de um TE entre espécies próximas evolutivamente (Silva *et al.*, 2004; Loreto *et al.*, 2008).

Eventos de HT com forte sustentação são aqueles em que todas estas evidências são observadas. No entanto, hipóteses alternativas devem ser consideradas, como polimorfismo ancestral, taxas evolutivas diferentes entre espécies, pressões seletivas fortes em porções do elemento ancestral, perdas estocásticas ou mesmo introgressão (Capy *et al.*, 1998).

Diferentes vetores têm sido sugeridos para os casos de HT em *Drosophila* (revisão em Loreto *et al.*, 2008), sendo que alguns desses vetores também podem ser funcionais para outros organismos. Alguns TEs, como os retrotransposons com LTRs, têm a capacidade de produzir partículas virais (VLPs), as quais podem funcionar como um vetor na transmissão de material genético. Outros prováveis vetores incluem os vírus de DNA, parasitas ou parasitóides de *Drosophila*, como ácaros e vespas, além de bactérias endossimbiontes como *Wolbachia*.

IMPLICAÇÕES EVOLUTIVAS DOS TÊS NOS GENOMAS

O mecanismo de transposição utilizado pelos TEs é uma reação de recombinação que intermedia o movimento desses segmentos de DNA entre sítios não homólogos. Assim, uma vez que são elementos móveis, os TEs têm a capacidade de mudar informações genéticas do hospedeiro, alterando a estrutura dos cromossomos ou a organização dos genes (Craig *et al.*, 2002). De uma maneira geral, os TEs podem influenciar a trajetória evolutiva de seus hospedeiros de três distintos modos: (1) alterando a função de um gene através de sua inserção, (2) através de rearranjos cromossômicos, e (3) como uma fonte de material codificante ou não que permite a emergência de novidades genéticas como novos genes e sequências regulatórias (Feschotte e Pritham, 2007).

A inserção de um TE na região codificadora de determinado gene pode levar a alteração da sequência, inibindo a geração de produtos gênicos viáveis. Já inserções em regiões promotoras, introns, ou regiões não codificadoras de um gene, podem gerar os mais diversos fenótipos, desde alterações regulatórias até a perda completa da função gênica. A

excisão de um transposon pode levar, algumas vezes, à reversão de fenótipos mutantes, além de deleções (Kidwell e Lisch, 2001).

Fenótipos mutantes em *Drosophila* são frequentemente causados pela inserção de TEs. Por exemplo: uma série de mutações nulas no loco *white* levou à descoberta do elemento *P* como agente causal da disgenesia gonadal P-M (Rubin *et al.*, 1982). Já a inserção do retroelemento *gypsy* na região regulatória 5' do gene *yellow* causou a perda de expressão deste gene em tecidos específicos (Corces e Geyer, 1991).

Em decorrência do seu mecanismo de transposição, o qual pode envolver múltiplas quebras da dupla-fita de DNA e eventos de reparo, os TEs estão envolvidos na indução em larga escala de rearranjos cromossômicos. Múltiplas cópias de um elemento podem também atuar como substrato para recombinação homóloga, o que pode levar a alterações na estrutura do genoma (Craig *et al.*, 2002).

Apesar dos efeitos às vezes deletérios, a presença de TEs nos genomas hospedeiros constitui uma fonte de variabilidade genética. Assim, os TEs podem atuar como geradores de novos alelos e novas redes regulatórias, que constituem as maiores forças que dirigem a diversificação das espécies (Feschotte e Pritham, 2007). Exemplos de TEs com aparente função adaptativa nos genomas hospedeiros vêm sendo descritos. Feschotte (2008), em um trabalho de revisão, lista uma série de fatores de transcrição e proteínas de ligação ao DNA que são derivados de transposases. Um exemplo é o fator de transcrição BEAF-32 em *Drosophila* que parece ser derivado de transposases da superfamília *hAT*. Bundock e Hooykaas (2005) também descrevem o gene *Daysleeper*, com funções essenciais para o crescimento e desenvolvimento de *Arabidopsis*, como derivado de transposases da superfamília *hAT*. Jordan *et al.* (2003), em um estudo sobre o genoma humano, reportam que aproximadamente 25% das regiões promotoras contêm sequências derivadas de TEs, incluindo elementos *cis*-reguladores

Outras evidências de que os TEs podem adquirir funções regulatórias nos genomas vêm da observação de que esses elementos tendem a se agrupar ao redor de genes envolvidos no desenvolvimento e regulação transcricional (Lowe *et al.*, 2007). Eles também são super representados dentro de sítios preditos como regiões *cis*-reguladoras (Gentles *et al.*, 2007).

Assim, parece que a acumulação de transposons nos genomas pode funcionar como matéria-prima para a geração de novos elementos regulatórios, novos genes, podendo levar à emergência de novas redes regulatórias.

SILENCIAMENTO DE TRANSPOSONS

Algumas vezes os TEs podem produzir muitas mutações letais, podendo causar a sua extinção, bem como a de seu hospedeiro. Sequências móveis podem sobreviver em um genoma somente se houver um equilíbrio no qual ambos adquiram alguma vantagem genética (Capy *et al.*, 1998). Assim, devido aos seus efeitos deletérios e à grande quantidade de TEs existentes nos mais diversos genomas, a manutenção da integridade dos genomas hospedeiros exige regulação da transposição ou mesmo o silenciamento dos TEs.

Mecanismos gerais de silenciamento dos TEs incluem a metilação, inibição por múltiplas cópias, formação de complexos oligoméricos, síntese de repressores e outros mecanismos globais de repressão como o mecanismo de interferência por RNA (iRNA).

O silenciamento de TEs mediado por RNA tem sido sugerido como um mecanismo chave para regular a atividade de expressão e transposição de TEs em *Drosophila*. De uma maneira geral, o silenciamento por RNA se refere a uma série de mecanismos mediados por pequenos RNAs que controlam uma variedade de processos biológicos, tais como: desenvolvimento, estabilidade do genoma, dispersão de viroses e elementos de transposição (Girard e Hannon, 2008; Chambeyron *et al.*, 2008).

Três maiores classes de pequenos RNAs têm sido definidas em plantas e animais (Birchler e Kavi, 2008; Girard e Hannon, 2008):

- miRNAs (microRNAs) são codificados pelo genoma e processados em produtos de 22 pb que interagem com mRNAs homólogos, modulando sua tradução e afetando processos do desenvolvimento. Em alguns organismos miRNAs podem guiar a clivagem de mRNA complementares (Bartel, 2004).
- siRNAs (do inglês *small interfering RNAs*) possuem 21 pb e são gerados a partir de RNAs dupla fita precursores, como de viroses e transposons endógenos (Ghildiyal *et al.*, 2008).

- piRNAs (do inglês *Piwi-interacting RNAs*) possuem aproximadamente 25 pb de tamanho e apresentam homologia com elementos de transposição e outras sequências. Eles também podem ser denominados rasiRNAs (do inglês *repeat-associated siRNAs*) (Brennecke *et al.*, 2007).

As principais enzimas envolvidas no processamento destes pequenos RNAs e na clivagem dos mRNA alvos incluem Dicer e Argonata. Dicer é uma endonuclease da família da RNaseIII, específica para RNA dupla fita. Já Argonata é uma família de proteínas com papel essencial no mecanismo de silenciamento. Proteínas Argonata possuem dois domínios de reconhecimento: o domínio PAZ, o qual liga extremidades 3' de pequenos RNAs, e o domínio Piwi que confere a atividade catalítica de clivagem à proteína. A figura 1 representa a ação dessas enzimas em dirigir a degradação de RNAs alvo.

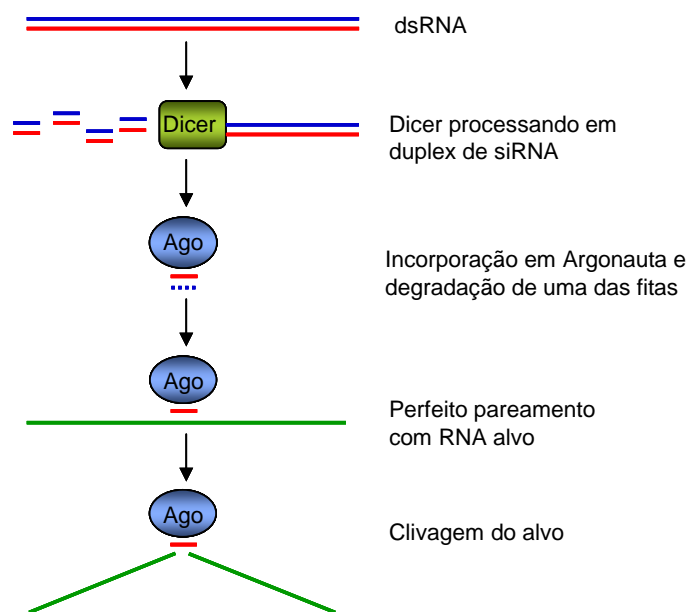


Figura 1: Padrão de silenciamento mediado por siRNA quimicamente caracterizado (adaptado de Girard & Hannon, 2008). O processo é iniciado pela detecção de dsRNA (RNA dupla fita – do inglês *double strand RNA*) através do mecanismo de iRNA. Esses dsRNAs são então processados pela enzima Dicer em siRNA (small RNA). Esses siRNA são incorporados a um complexo RISC (do inglês *RNA-induced silencing complex*) contendo a proteína Argonata. Após a incorporação uma das fitas é clivada pelo sítio catalítico da proteína Argonata e degradada. O pareamento perfeito entre a fita restante à moléculas de RNA do genoma hospedeiro dirigem a clivagem deste RNA alvo.

Em *D. melanogaster* a enzima Dicer-1 cliva miRNAs dupla fita precursores, e os produtos resultantes ligam-se à proteína Argonauta1 (Ago1). siRNAs são gerados pela enzima Dicer-2 a partir de RNA dupla fita precursor e ligam-se à Argonauta2 (Ago2), clivando RNAs homólogos. piRNAs são gerados por um mecanismo independente de Dicer, sendo a função de clivagem realizada pelo conjunto de proteínas Argonauta, formado por Piwi, Aubergine, e Argonauta3. Estas proteínas usam transcritos antisense (complementares ao mRNA) codificados no genoma por *clusters* piRNA que abrigam fragmentos de elementos de transposição e direcionam a destruição de transcritos senso do transposon (Aravin *et al.*, 2007; Brennecke *et al.*, 2007).

Tem sido demonstrado que piRNAs antisense são produzidos por loci heterocromáticos que transcrevem um *cluster* de elementos de transposição defectivos na orientação antisense (Brennecke *et al.*, 2007; Mevel-Ninio *et al.*, 2007). Adicionalmente, os piRNAs antisense associados a Piwi e Aub podem gerar piRNAs senso por marcar como alvo os transcritos de elementos funcionais presentes na eucromatina. piRNAs senso podem ser usados novamente para produzir piRNAs antisense adicionais, direcionando a clivagem de transcritos antisense sintetizados a partir do *cluster* heterocromático (Brennecke *et al.*, 2007). Esse ciclo de amplificação, denominado “ping-pong”, parece ser um dos principais mecanismos que regulam a atividade de transposons.

Estudos em *Drosophila* têm demonstrado o importante papel de pelo menos dois loci piRNA como principais reguladores do silenciamento de transposons. O primeiro é o locus *flamenco* (Brennecke *et al.*, 2007), o qual parece ser requerido para a repressão de diversas famílias de transposons, como *gypsy*, *idexis* e *Zam. flamenco* foi originalmente identificado como um locus controlador da atividade de *gypsy* (Pelisson *et al.*, 1994). O outro locus piRNA geneticamente caracterizado é o locus X-TAS, localizado na região pericentromérica do cromossomo X. X-TAS está envolvido na repressão do elemento *P* através da interação com Aubergine (Stuart *et al.*, 2002; Reiss *et al.*, 2004).

Adicionalmente, pequenos RNAs e proteínas Argonauta parecem ter um papel no silenciamento de TEs através de modificações na cromatina, por metilação do DNA e/ou modificação de histonas (Slotkin e Martienssen, 2007; Girard e Hannon, 2008).

A SUPERFAMÍLIA *hAT*

A superfamília *hAT* de transposons é composta por elementos presentes nos mais diversos organismos. Transposons desta superfamília são conhecidos em fungos, nematódeos, peixes, insetos, plantas e até mesmo no homem (Kempken e Windhofer, 2001). O nome da superfamília *hAT* é devido a três elementos que a compõem: *hobo*, descrito originalmente em *Drosophila melanogaster* (Calvi *et al.*, 1991), *Activator (Ac)* de *Zea mays* (McClintock, 1947) e *Tam3* de *Antirrhinum majus* (Hehl *et al.*, 1991). Elementos de insetos e plantas foram classificados em uma mesma superfamília por apresentarem características estruturais e funcionais comuns.

De acordo com a classificação de TEs sugerida por Wicker *et al.* (2007) os elementos que fazem parte da superfamília *hAT* são pertencentes à Classe II, como sugerido por Finnegan (1989), por se moverem diretamente através do DNA, sem a necessidade de um intermediário de RNA para a transposição. Eles são chamados de transposons propriamente ditos. Dentro da Classe II os membros da superfamília *hAT* fazem parte da Subclasse 1, caracterizada por elementos que utilizam o mecanismo clássico de *cut-and-paste* para sua transposição. Na superfamília *hAT* os elementos que a compõem compartilham algumas características como: (1) duplicação do sítio alvo (TSD) de 8 pb, (2) repetições terminais invertidas (TIRs) relativamente curtas (entre 5 e 27 pb), e (3) geralmente possuem menos de 4 kb de tamanho. Uma exceção é o elemento *gulliver* de *Chlamydomonas reinhardtii* (Ferris, 1989), que possui até 12 kb de tamanho. Outra característica comum entre os membros desta superfamília é a similaridade na sequência de aminoácidos da transposase (Calvi *et al.*, 1991), que é limitada a três motivos de sequência de aminoácidos (Kempken e Windhofer, 2001).

Membros da superfamília *hAT* podem ainda ser agrupados em famílias que são definidas pelo grau de similaridade em suas sequências de DNA. Dentro de cada família, podem existir as subfamílias, definidas com base em dados filogenéticos, que podem servir para distinguir entre elementos autônomos e não-autônomos.

O ELEMENTO TRANSPONÍVEL *hobo*

O elemento transponível *hobo* foi originalmente descrito por McGinnis *et al.* (1983) como uma inserção no gene *Sgs-4* de *Drosophila melanogaster* próximo da região TATA box. Essa inserção de 1273 pb de tamanho provocou a redução na expressão do gene *Sgs-4* e levou à identificação de uma família de sequências de DNA de tamanho e padrão de restrição não conservados em linhagens de *Drosophila melanogaster*.

O elemento *hobo* canônico (HFL1) possui 2959 pb de tamanho, com TIRs de 12 pb, e contém uma ORF que codifica a transposase. Os sítios de inserção de *hobo* são tipicamente flanqueados por uma duplicação de 8 pb na sequência hospedeira, o que é uma característica de membros da superfamília *hAT* (Calvi *et al.*, 1991). Outra característica importante da estrutura do elemento canônico é a presença de dois sítios de clivagem para *XhoI*, que geram um fragmento de 2,6 Kb presente na maioria dos elementos *hobo* analisados até agora. Além disso, a sequência completa inclui sítios TATA e CAAT (Streck *et al.*, 1986), potenciais promotores basais da atividade transcricional.

Dentro da família *hobo* podem ser encontrados elementos de diferentes formas, como demonstrado na Figura 2. A primeira forma é de elementos com sequências completas, como o elemento *hobo* canônico, com potencial para codificar a transposase. A segunda forma é descrita como elementos *hobo* defectivos, que são menores, de tamanho variável e parecem derivar de elementos completos a partir de deleções internas (Streck *et al.*, 1986). Os elementos defectivos tem sido descritos como sequências não funcionais e, portanto, incapazes de codificar sua própria transposase (Simmons *et al.*, 1998; Galindo *et al.*, 2001). A terceira forma é descrita como *hobo* “reliquia” (*relic*), ou sequências relacionadas a *hobo* (*hRs* – do inglês *hobo-related sequences*). Essas *hRs* apresentam alta similaridade com o elemento *hobo* canônico, múltiplos rearranjos na região interna e são potencialmente inativas (Simmons *et al.*, 1998).

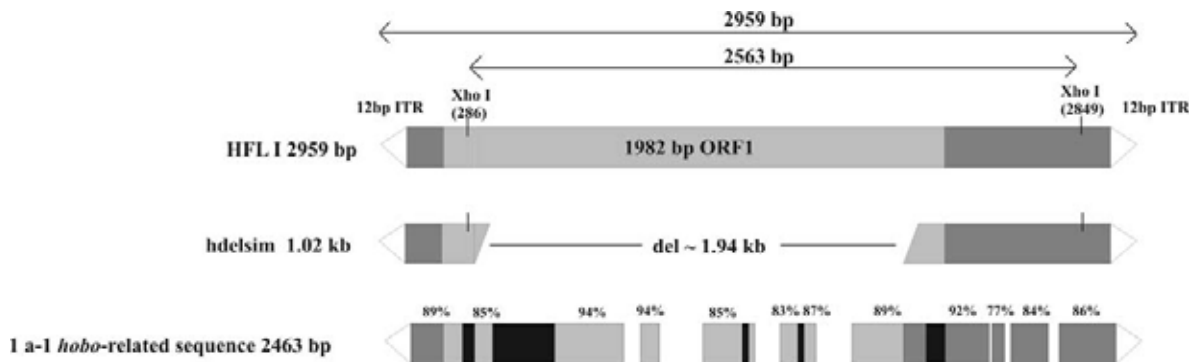


Figura 2. Três possíveis formas do elemento *hobo*. O elemento *hobo* canônico (HFL1), cópias internamente deletadas como *hdelsim* de *D. simulans*, e uma sequência relacionada a *hobo* de *D. melanogaster*. São demonstradas as TIRs observadas em todos os elementos, a ORF1 no elemento completo, e os sítios de clivagem para *XhoI* que podem estar ausentes nas *hRs*. A identidade entre *hRS* e *HflI* é indicada acima da sequência, e inserções são representadas por boxes pretos. Adaptado de Torres *et al.* (2006).

Em *Drosophila* o padrão de distribuição de elementos *hobo* parece ser restrito. Através de Southern blot, Daniels *et al.* (1990) analisaram 134 espécies e demonstraram que no gênero *Drosophila* somente as espécies do subgrupo *melanogaster* e *montium* contêm elementos *hobo* suficientemente conservados que permitem sua detecção. Dentre estes subgrupos somente *D. melanogaster*, *D. simulans* e *D. mauritiana* parecem conter cópias potencialmente completas. Simmons (1992) demonstrou que muitos dos elementos *hobo* completos de *D. melanogaster* são praticamente idênticos, em sequência de DNA, aos elementos canônicos encontrados em *D. simulans* e *D. mauritiana*. Fora do gênero *Drosophila* elementos similares a *hobo* (*hobo-like*) têm sido encontrados em algumas espécies de Diptera, como *Musca domestica* (Atkinson *et al.*, 1993) e diferentes tefritídeos (Handler e Gomez, 1996; Torti *et al.*, 2005) e também em algumas espécies de lepidóptera (DeVault e Narang, 1994; Borsati *et al.*, 2003).

O número de cópias de elementos *hobo* nos genomas é variável e, além disso, a posição de inserção dessas cópias também varia entre linhagens (Blackman *et al.*, 1989). Elementos *hobo* deletados representam a forma predominante no genoma de *D. melanogaster* (cerca de 30 a 75 cópias por genoma), enquanto elementos completos parecem estar presentes

em menor número de cópias (de duas a dez cópias) (Streck *et al.*, 1986; Blackman *et al.*, 1989).

Em relação ao padrão temporal e geográfico de distribuição do elemento *hobo*, análises das linhagens mais antigas de *D. melanogaster*, coletadas antes de 1950, sugerem que elementos *hobo* completos eram ausentes nessas linhagens, mas que eles foram recentemente introduzidos nesta espécie por transferência horizontal (Daniels *et al.*, 1990; Simmons, 1992; Periquet *et al.*, 1994). Considerando a variabilidade de elementos *hobo* observada nos genomas, dois eventos independentes de introdução de *hobo* são sugeridos: um inicial antes da separação dos subgrupos *montium* e *melanogaster*, e outro mais recente para o progenitor do complexo *melanogaster* (Periquet *et al.*, 1994).

As linhagens de *Drosophila* podem ser divididas em linhagens H, que contêm elementos *hobo*, e linhagens E (do inglês *empty*) que não possuem elementos *hobo* em seu genoma. Linhagens H apresentam elementos *hobo* completos e deletados. Contudo, tanto linhagens E quanto H demonstram sequências relacionadas a *hobo*. Essas *hRs* são observadas como uma fraca hibridização por Southern Blot utilizando o elemento canônico como sonda (Periquet *et al.*, 1994; Galindo *et al.*, 2001).

O sistema *hobo* parece ser o agente causal de instabilidades genéticas como hipermutabilidade, aberrações cromossômicas e disgenesia gonadal (Blackman *et al.*, 1987; Yannopoulos *et al.*, 1987; Lim, 1988). Essa instabilidade genética parece ser devida à mobilização de elementos *hobo* na linhagem germinativa (Galindo *et al.*, 1995).

O ELEMENTO *HOB*^{VA}

O elemento *hobo*^{VA} foi caracterizado como uma inserção no locus *white* de *D. simulans*, a qual levou à geração *de novo* de um mutante *white*. A análise da sequência deste elemento mostrou tratar-se de uma sequência relacionada a *hobo* (*hRS*). *hobo*^{VA} apresenta 81,2% de similaridade com a sequência de DNA do elemento *hobo* canônico. Possui 1220 pb de tamanho e não apresenta ORF funcional. No entanto, suas TIRs são virtualmente iguais às do elemento canônico, com substituição de apenas um nucleotídeo na TIR 3'. Adicionalmente,

cerca de 300 pb em cada região subterminal são muito similares ao elemento canônico (Torres *et al.*, 2006).

Apesar do elemento *hobo*^{VA} não possuir as sequências necessárias para codificar a transposase, a presença de TIRs praticamente intactas e a similaridade com o elemento canônico, torna *hobo*^{VA} um elemento potencialmente mobilizável, provavelmente pelo fornecimento *in trans* de transposase a partir de elementos *hobo* completos.

O padrão de distribuição do elemento *hobo*^{VA} é muito similar ao elemento canônico. Através de buscas *in silico* nos 12 genomas de *Drosophila* disponíveis, Ortiz e Loreto (2008) demonstraram a presença de sequências homólogas a *hobo*^{VA} (*hobo*^{VAHS} – do inglês *hobo*^{VA} *homologous sequences*) somente no grupo *melanogaster*. Os autores também observaram que o número de cópias dessas *hobo*^{VAHS} nos genomas é bastante variável. Foram observadas 147 cópias de *hobo*^{VAHS} no genoma de *D. simulans*, 12 cópias no genoma de *D. melanogaster* e apenas uma cópia no genoma de *D. erecta* (Ortiz e Loreto, 2008).

Considerando a estrutura dos *hRs*, têm sido sugerido que esses elementos são muito ancestrais, representando sequências *relic* dentro dos genomas hospedeiros (Periquet *et al.*, 1994). No entanto, Ortiz e Loreto (2008) apontam para o fato de que estas sequências, embora não autônomas, vêm sendo mantidas mobilizáveis por um tempo evolutivo considerável (entre 0,8 a 12 milhões de anos).

O ELEMENTO *hobo* EM *D. simulans* E HIPERMUTABILIDADE

Loreto e colaboradores (1998) caracterizaram uma linhagem hipermutável de *D. simulans* e, ao longo de aproximadamente 100 gerações de observação, foram isoladas sete diferentes mutações, tanto autossômicas como ligadas ao X. As mutações caracterizadas por análises genéticas demonstraram duas mutações no cromossomo X, uma no locus *lozenge* e outra no gene *ruby*. Já as mutações autossômicas caracterizadas foram um mutante *dpp*-like, um mutante *blistered*-like e um mutante homeótico que apresentava a antena transformada em pata e olhos ectópicos. Outro mutante isolado foi um mutante com olhos de cor branca, sendo que com essa mutação no loco *white* foi possível estabelecer-se uma linhagem *white* estável.

As propriedades biológicas da linhagem hipermutável de *D. simulans* descritas por Loreto *et al.* (1998) sugeriram ser um elemento transponível muito ativo o agente causal da hipermutabilidade desta linhagem. Uma vez que eventos de mutações espontâneas são raros, o encontro de uma série de mutações espontâneas em uma linhagem de *D. simulans*, caracterizando um fenômeno de hipermutabilidade, é extremamente importante. Este achado também fornece uma excelente oportunidade de se estudar a relação entre elementos transponíveis, mutações espontâneas e o genoma hospedeiro. Posteriormente, Torres *et al.* (2006) demonstraram que o elemento transponível envolvido na mutação *white*, isolado nesta linhagem, foi causado pela inserção de um elemento *hobo* relíquia, que foi denominado *hobo*^{VA} (descrito anteriormente).

EXPRESSÃO DE TES DURANTE O DESENVOLVIMENTO DE *DROSOPHILA*

O padrão de expressão transcricional de TEs tem sido descrito por alguns autores, principalmente para retrotransposons. A localização desses transcritos vem sendo observada nos mais diversos tecidos e padrões em embriões de *Drosophila*. Ding e Lipshits (1994) descreveram o padrão de expressão transcricional de quinze famílias de retrotransposons (*17.6*, *blood*, *297*, *412*, *copia*, *gypsy*, *1731*, *3S18*, *HSM Beagle*, *Kermit/flea*, *mdg1*, *mdg3*, *opus*, *roo/B104* e *springer*) durante a embriogênese de *Drosophila*. Devido ao padrão espacial de expressão desses retrotransposons ser conservado entre as diferentes linhagens de *D. melanogaster* testadas, os autores sugerem a existência de sequências *cis*-reguladoras controlando a expressão desses elementos. Como os sítios de inserção podem variar de uma linhagem para outra, é sugerido que as sequências que modulam a expressão são intrínsecas aos retrotransposons. Dentre os padrões observados através de hibridização *in situ* de embriões inteiros, alguns parecem ser compartilhados entre as famílias de retrotransposons, por exemplo: a expressão nas gônadas é observada para *17.6*, *412*, *mdg1*, *297* e *gypsy*; já a expressão observada no sistema nervoso central é compartilhada por *297*, *1731*, *mdg3*, *copia*, *HMS Beagle* e *opus*. Isso sugere a existência de sequências conservadas também entre os elementos analisados.

Frommer *et al.* (1994) descreveram a expressão restrita do elemento *blastopia* como um “anel circunferencial” na porção anterior de embriões de *D. melanogaster* no estágio de blastoderme. Essa expressão localizada de *blastopia* é sugerida como dependente do morfógeno de origem materna *bicoid*. O trabalho de Brönner *et al.* (1995) demonstra que a expressão de B104 parece ser direcionada por sequências internas a esse retrotransposon que são capazes de atuar como um elemento *cis*-regulador frente a promotores heterólogos de *Drosophila*. Os autores sugerem que a inserção de retrotransposons pode afetar o padrão de expressão de genes endógenos por adicionar elementos controladores que atuam em *cis* através do genoma hospedeiro. Já Mozer e Benzer (1994) demonstraram que as sequências *cis*-reguladoras que controlam a expressão do elemento 17.6 residem dentro deste retrotransposon e pode ser dependente de genes adjacentes ao seu sítio de inserção.

As regiões *cis*-reguladoras são segmentos de DNA que controlam a transcrição gênica. Estas unidades reguladoras, algumas vezes referidas como módulos *cis*-reguladores (CRMs, do inglês *cis-regulatory modules*), podem conter sítios de ligação para diversos fatores de transcrição (TFBS, do inglês *transcription factor binding sites*) e normalmente são localizadas antes do sítio de início da transcrição. Segmentos *cis*-reguladores individuais podem residir dentro de introns, em regiões 5' e 3' não traduzidas (UTRs, do inglês *untranslated regions*) ou mesmo dezenas de kb em qualquer um dos dois lados do gene que regulam (Berman *et al.*, 2004). Mutações nessas regiões podem ter consequências na morfologia, fisiologia e comportamento (Wray, 2007).

Os padrões de expressão transcricional de TEs observados até agora sugerem que esses elementos podem interagir com genes do desenvolvimento ou mesmo estar sob controle de algum destes.

II - OBJETIVOS

O presente trabalho tem como objetivo principal contribuir para o entendimento das consequências evolutivas da presença e expressão de TEs da superfamília *hAT* nos genomas de espécies de *Drosophila*.

Objetivos específicos:

1. Verificar a distribuição das sequências de *hAT* (homólogas aos elementos *hosim* e *hosec* previamente descritos por nosso grupo) nos genomas de espécies de *Drosophilidae*.
2. Obter um panorama geral das relações evolutivas dessas sequências de TEs.
3. Analisar e comparar o padrão de expressão transcricional de *hobo* canônico e *hobo*^{VA} em espécies do grupo *melanogaster* de *Drosophila*.

CAPÍTULO 2

HOSIMARY: A NEW HAT TRANSPOSON FAMILY OF ZAPRIONUS INDIANUS AND DROSOPHILA

MELANOGASTER

Maríndia Deprá¹, Yanina Panzera², Vera Lúcia da Silva Valente¹, Elgion Lúcio da Silva Loreto^{1,3}

¹ Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

² Sección Genética Evolutiva, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

³ Departamento de Biologia, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brazil

Manuscrito em preparação para ser enviado à revista **Cytogenetic and genome research**

Abstract

A PCR screening approach was used to search for sequences homologous to the previously described *hAT* transposon found in *D. simulans* and *D. sechellia*. In this study, 52 Drosophilidae species were analyzed and these sequences seem to be restricted to some species of the *melanogaster* group and *Z. indianus*. These species presented variable number of copies and most of these appear to be putatively encoding. The high sequence similarities among fragments of different species and the incongruence presented between the transposon and host species phylogenies strongly support the hypothesis that this element was horizontally transferred between the *melanogaster* group and *Z. indianus*. Additionally, considering that the nucleotide similarity shared between species was higher than 90% and that no high sequence similarity was detected with other transposable elements, we propose a new *hAT* transposon family called *hosimary*.

Key words: Transposable elements, *hAT*, lateral gene transfer, genomic evolution.

Introduction

Transposable elements (TE) have been found in almost all genomes studied and it is known that they play an important role in eukaryotic evolution by promoting genetic variability (Kidwell and Lisch, 1997; 2001). TEs may comprise a large part of species' genomes. In *Drosophila*, for example, the amount of TEs among the twelve genomes available ranged from 2.7% (in *D. simulans* and *D. grimshawi*) to 25% (in *D. ananassae*) (Clark *et al.*, 2007).

TEs are subdivided in two classes based on the transposition intermediary, DNA or RNA. These classes are further divided into order, superfamilies and families. The *hAT* superfamily belongs to the class II elements (also called DNA transposons), which are mobilized by a DNA intermediary through a “cut-and-paste” mechanism (Finnegan, 1987; Wicker *et al.*, 2007). Furthermore, the transposase encoded by *hAT* transposons share amino

acid sequence similarity (Calvi *et al.*, 1991; Essers and Kunze, 1995), which is limited mainly to three motifs (Kempken and Windhofer, 2001).

The *hAT* superfamily of TEs includes the *hobo* element, reported originally in *Drosophila melanogaster* (Calvi *et al.*, 1991), *Activator (Ac)* from *Zea mays* (McClintock, 1947) and *Tam3* from *Antirrhinum majus* (Hehl *et al.*, 1991). Other elements, such as *Herves* from *Anopheles gambiae* (Arensburger *et al.*, 2005), *Hermit* from *Lucilia cuprina* (Coates *et al.*, 1996) and *Hermes* from *M. domestica* (Warren *et al.*, 1994), were also shown to belong to the *hAT* superfamily.

The transposons of this superfamily are widespread, being present in fungi, nematodes, fish, plants, insects and even in humans (Kempken and Windhofer, 2001). Due to its wide distribution, it has been suggested that the *hAT* superfamily is probably very ancient, predating the separation of plants, animals, and fungi, once no evidence of trans-kingdom horizontal transfer have been shown until now (Rubin *et al.*, 2001).

Ortiz and Loreto (2009) have performed *in silico* searches in order to identify *hAT* elements in the twelve available *Drosophila* genomes. They have identified 37 new *hAT* elements. Two of these elements, labeled *hosim* e *hosec*, are remarkably similar and were found in the genome of *D. simulans* and *D. sechellia*, respectively. These are putatively active 2566-bp-long elements with a 491-aa transposase gene, 11-bp terminal inverted repeats (TIRs) as well as an 8-bp target site duplication (TSD), which is a typical trait of the *hAT* family.

In this work, we performed analyses to identify the presence and distribution of homologous sequences to *hosim* and *hosec* in 52 Drosophilidae species. In the genus *Drosophila*, among the 18 groups analyzed, these sequences were found only in the *melanogaster* group. Nevertheless, this TE was found in *Zaprionus indianus*, a species that although taxonomically assigned to a different genus, was shown to be intimately related to the genus *Drosophila* (Remsen and O'Grady, 2002; Robe *et al.*, 2005). The high sequence similarities presented (more than 90 % at the nucleotide level) suggest a new probable case of horizontal transfer (HT) involving some species from *melanogaster* group and *Z. indianus*. Additionally, we propose that these sequences, together with *hosim* and *hosec* (described by Ortiz and Loreto, 2009), should be clustered in a new family of transposons. Here we propose to name this new family *hosimary*.

Material and methods

Fly stocks

Forty-six *Drosophila* species were used in this study, along with *Zaprionus indianus*, *Z. sepsoides*, *Z. tuberculatus*, *Z. multistriatus*, *Scaptodrosophila latifasciaeformis* and *S. lebanonensis* (Table 1 – supplementary material). Three strains were used for *D. simulans*: *D. simulans* white (stock number 14021-0251.195, obtained from Tucson Stock Center; used for genome library to sequencing project; Clark *et al.*, 2007), yellow (Loreto *et al.*, 1998) and the Águas de Lindóia strain (collected in Águas de Lindóia, SP, Brazil in 2007). As for *D. melanogaster*, two strains were used: Oregon R and Zola (Souames *et al.*, 2003). All these strains have been maintained in laboratory by means of mass crosses and cultivated in corn flour culture medium, in a controlled chamber ($17 \pm 1^\circ\text{C}$, 60% r.h.).

PCR, cloning and sequencing

Genomic DNA was extracted from adult flies according to Sassi *et al.* (2005). The primers were designed specifically to TEs *hosimary* elements based on the sequences of *hosec* and *hosim* (Ortiz and Loreto, 2009). A fragment of approximately 735 bp was amplified with primers hATss-F (5' GGA CCC CAA CAG CAA GTA ATA A 3') (forward) and hATss-R (5' ACG AGT TGA CAC ATC TTG AAT C 3') (reverse). The Figure 1 - supplementary material) shown the domains found in *hosimary* family elements as well as the fragment amplified that it is located mainly in the DUF 659 domain.

The PCRs were performed in 25 μl using 20 ng of Genomic DNA, 1 U Taq DNA Polymerase (Invitrogen), 1X reaction buffer, 200 μM of each nucleotide, 20 pmol of each primer and 1.5 mM MgCl_2 . The employed amplification conditions were 95°C for 5 min, 30 cycles at 95°C for 40 s, 55°C for 40 s and 72°C for 1 min, followed by a final extension cycle at 72°C for 5 min. Different conditions were tested in an attempt to obtain amplification in a larger number of species, but were unfruitful. The amplicons obtained were cloned into pCR-TOPO plasmid (Invitrogen). DNA sequencing was performed directly from the purified plasmids in the MegaBace 500 automatic sequencer. The dideoxy chain-termination reaction was implemented by using the DYEnamicET kit (GE Healthcare).

Southern blot analyses

For the Southern blot analyses, approximately 6 µg of genomic DNA of each sample was digested with *EcoRV* (Invitrogen) following the conditions suggested by the manufacturer. The *EcoRV* restriction enzyme was chosen because it has no cleavage site within the *hosimary* sequence. DNA fragments were separated by electrophoresis in 0.8% agarose gels and transferred to nylon membranes (HybondN+, GE Healthcare). The PCR fragment of *D. simulans* white clone (white3) was used as probe. The labeling and hybridization were performed using a Gene Images kit following the manufacturer's instructions for a high stringency hybridization (GE Healthcare). The detection was made with a CPD Star Detection kit (GE Healthcare).

FISH

The FISH analysis was carried out with squashed preparations of larval salivary glands of *D. simulans* fixed in 3:1:2 (ethanol-acid lactic-acetic acid, v/v). The chromosome preparation was pre-treated with RNAase, proteinase and post-fixed in 4% paraformaldehyde. The DNA probe (clone of *D. simulans* white - white3), were labeled through the Nick Translation System (Invitrogen) with Cy3-dUTP (GE Healthcare). Before *in situ* hybridization, the probe (50 ng per slide) was mixed with *in situ* hybridization solution as follows: 50% formamide, 10% dextran sulfate and 2xSSC. The hybridization mixture was added to the slides and denatured in 70% formamide and 2xSSC at 70°C for six minutes. After hybridization, the squashed preparations were washed five times with 2x SSC, for 5 min. Vectashield with 4', 6-diamidino-2-phenylindole (DAPI; Vector Laboratories, Inc., Burlingame, CA, USA) was used as an antifade solution. The location of the probe was observed in at least 10 interphase chromosomes per slide. The preparations were analyzed on Axioplan 2 imaging HAL 100 of Zeiss. The images were processed with the ISIS program (METSYSTEMS GmbH).

Sequence analyses

Consensus sequences for each clone were obtained using the GAP 4 software of the Staden Package (Staden, 1996). Then, the sequences were aligned using the Muscle tool (Robert, 2004) with default parameter values.

The phylogenetic relationship between the TE sequences were estimated using: (1) Neighbor-joining (NJ) method (Saitou and Nei, 1987), in the Mega 4 program (Tamura *et al.*, 2007), with the Tamura three-parameter nucleotide substitution model (Tamura, 1992); (2) Bayesian analysis implemented in the MrBayes (Ronquist and Huelsenbeck, 2003) program, with the HKY nucleotide substitution model (Hasegawa *et al.*, 1985), as suggested in the MrModel Test 2.2 (Nylander, 2004).

In the divergence analyses, the sequences were clustered for each species to perform a p-distance analysis using the Mega 4 program (Tamura *et al.*, 2007).

Inference of horizontal transmission.

To infer the occurrence of horizontal transfer (HT), two aspects were considered: (1) incongruence in the TE phylogeny and (2) the low dS (synonymous sites) values of the TE as compared to the host genes. The host genes chosen for this analysis were: *amd* (*alpha methyl dopa-resistant*), *amy* (*amyrel*), *per* (*period*) and *sod* (*superoxide dismutase*). In the case of *sod* gene, *Z. tuberculatus* sequences were utilized as a placeholder for *Z. indianus*. Accession numbers of the host genes sequences are available in Supplementary material Table 2.

The values of synonymous substitution per synonymous sites (dS) were obtained using the Nei and Gojobori method (Nei and Gojobori, 1986) with p-distance implemented in Mega 4 program (Tamura *et al.*, 2007). In this analysis, the sequences were also clustered in groups and gaps were introduced to conserve the reading frame, while stop codons were considered as absent data.

To verify whether the differences observed in the synonymous sites of the TE are significantly lower than those observed for host genes, we carried out a Fisher's test, assisted by DNASP 4.0 software. In this case, to avoid under- and over-estimation of horizontal transfer events (Ludwig *et al.*, 2008), the codon bias usage was calculated in the DnaSP 4.0 software (Rozas *et al.*, 2003) by the codon bias index (CBI; Morton, 1993) and the effective number of codons (Nc; Wright, 1990).

In order to estimate the divergence time of TEs sequences we employed the equation $T=K/2r$ (Graur and Li, 2000), in which T is the divergence time between species, K is the divergence among TE sequences in synonymous sites (dS) and r is the evolutionary rate. The

rate of 0.016 synonymous substitutions per site per million years is suggested for *Drosophila* genes with low codon usage bias (Sharp and Li, 1989).

Results

Detection and distribution of *hosimary* sequences in the analyzed species.

We have analyzed the detection of *hosimary* by PCR in 52 species belonging to three Drosophilidae genera (Supplementary Table 1). DNA amplification was observed in four of 45 analyzed *Drosophila* species, all of them belonging to the *melanogaster* subgroup. Fragments of nearly 735 bp were amplified in *D. simulans*, *D. sechellia* and *D. mauritiana*. In *D. melanogaster* a smaller fragment was detected. Besides the *melanogaster* group, amplificons were obtained only in *Z. indianus*. Southern blot analysis with genomic DNA confirmed the PCR results (Figure 1). By using PCR fragment of *D. simulans* white strain as probe, several hybridization bands were observed in *D. simulans*, *D. mauritiana* and *Z. indianus*, while a single signal was observed in *D. melanogaster*. In other species no hybridization signal was detected.

Number of copies

Ortiz and Loreto (2009) have found four copies of *hosim* and 8 copies of *hosec* through *in silico* search in *D. simulans* and *D. sechellia*, respectively. In order to validate our sequences we performed a BLAST search in *Drosophila* genomes (<http://flybase.bio.indiana.edu/blast/>; Grumbling and Strelets, 2006). Our results recovered one additional copy of the *hosec* from *D. sechellia* (scaffold 82: 62715 - 67257), with 100% of similarity with our clones, and a copy in the *D. melanogaster* genome (2R: 666706 - 670810). This unique copy of *D. melanogaster* was confirmed by Southern blot analysis (Figure 1 – lane 3). In the sequenced genome it is located at the section 41 - 42 proximal to the 2R chromosome. It shows an internal deletion when compared to consensus sequences of *hosim* and *hosec*, and contains TIR just in the 3' end. The 5' TIR is degenerated and followed by a long insertion with no similarity to consensus sequence.

For *D. simulans* white, the Southern blot analysis suggested the occurrence of at least seven copies (Figure 1 – lane 1) and in a strain collected recently (Águas de Lindóia) a similar number of bands was observed (Figure 1 – lane 11), suggesting a similar copy number. By performing a FISH on polytene chromosomes we found three hybridization signals in the chromosomal arms 3L (section 78- proximal) and 2L (section 22- distal and 38- proximal) of *D. simulans* (Figure 2). Besides, some signals were detected also in the chromocenter. Research in the FlyBase, (in the *D. simulans* genome) confirmed the presence of three copies in the chromosomal arms and four copies in a non-determined chromosome region.

D. mauritiana presents five hybridization bands suggesting the occurrence of roughly the same number of copies present in *D. simulans*. *Z. indianus* showed fewer copies, two heavy and two faint bands, as detected in lane 6 of Figure 1.

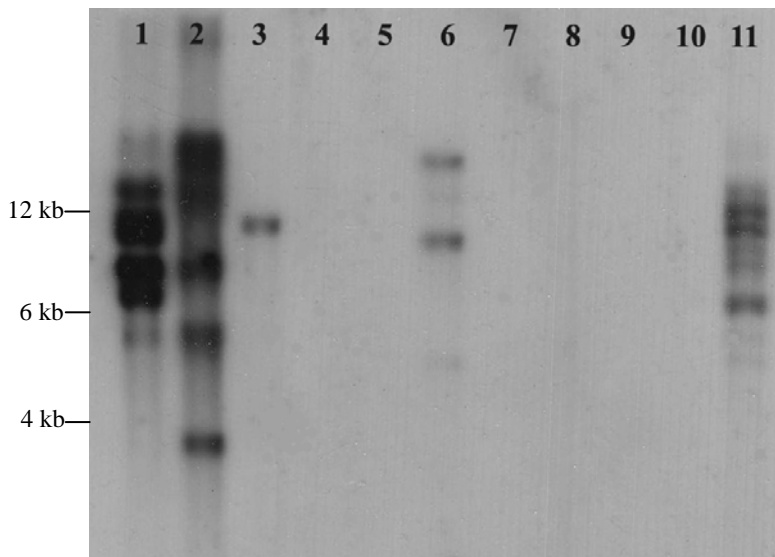


Figure 1. Southern blot showing hybridization bands of *hosimary* sequences in the genome of different species. The genomic DNA was digested with *EcoRV* and hybridized using the *D. simulans* white sequence as probe. The samples are: 1- *D. simulans* white, 2- *D. mauritiana*, 3- *D. melanogaster* Oregon, 4- *D. kikkawai*, 5- *D. santomea*, 6- *Z. indianus*, 7- *D. arawakana*, 8- *D. cardinoides*, 9- *D. procardinoides*, 10- *D. cardini*, 11- *D. simulans* Águas de Lindóia strain.

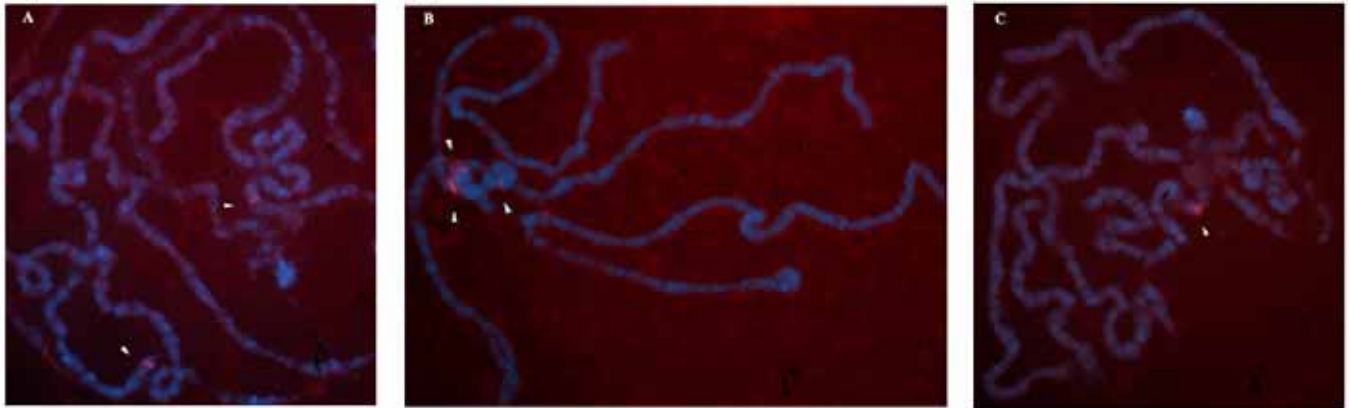
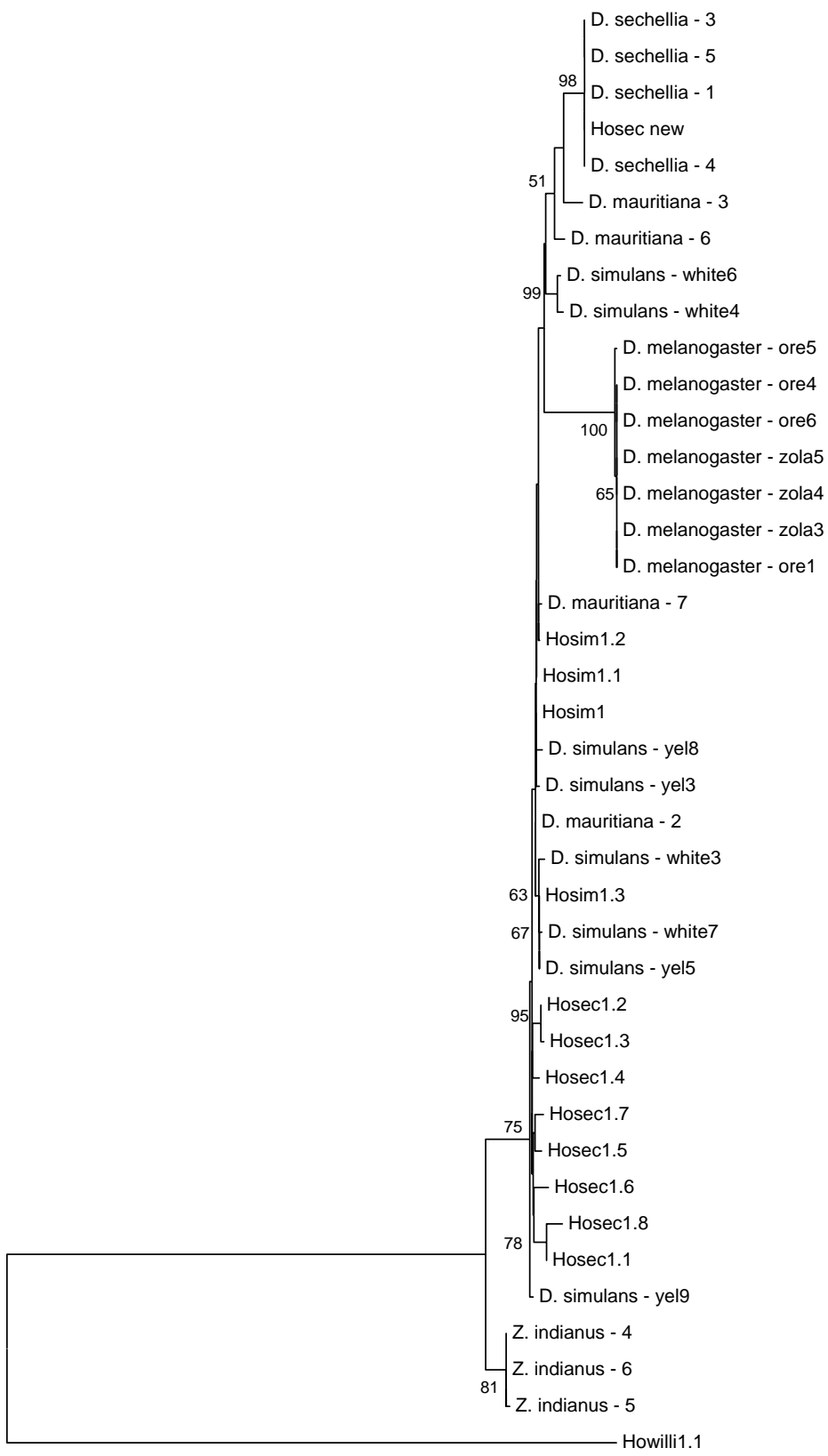


Figure 2: Localization of members of the *hosimary* family of transposable elements in the *D. simulans* white strain polytene chromosomes. (A) arrowheads shows euchromatic insertion sites in 2L and 3L chromosomes, (B,C) insertion sites of *hosimary* family in the heterochromatin or in the transition between heterochromatin – euchromatin.

Evolutionary analyses

The sequences obtained were aligned and approximately 700 nucleotides were used to build a phylogenetic matrix, from 39 clones sequences analyzed. The two phylogenetic inference methods used (Neighbor- Joining and Bayesian analysis) revealed similar topologies. The tree obtained by the Neighbor-Joining method (Figure 3) consists of two main clades (the *melanogaster* group sequences and the *Z. indianus* species sequences). Within both of these clusters there is no defined topology, given that all sequences showed low nucleotide divergence values, resulting in short branch lengths and low bootstrap values. Besides that, with the exception of those from *Z. indianus* and *D. melanogaster*, the sequences of the other species usually appear interspersed among each other.



H
0.002

Figure 3: Phylogenetic analysis of the *hosimary* family using nucleotide sequences alignment by Neighbor-Joining distance method with the Tamura three-parameters model. Numbers above branches are bootstrap values based on 1000 replications in the consensus tree. Bootstrap values lower than 50 were omitted. The species name abbreviations are followed by clone number and the sequences called *hosim* and *hosec* are available in supplementary material. The *howilli1* sequence (Ortiz and Loreto, 2009) was used as outgroup.

In order to perform the analyses of the nucleotide divergences, the sequences were clustered by species. In general, the comparison of all nucleotide sequences between species shows that the sequences are very similar to each other (Table 1). The highest divergence can be seen between *D. melanogaster* and *Z. indianus*, which a p-distance value of 7.7%, while the lowest divergence (p-distance value of 1.3%) was detected between *D. simulans* and *D. mauritiana*. The sequences of *D. melanogaster* (Oregon and Zola strains) and one sequence of *D. mauritiana* presented a premature stop codon. The remaining sequences did not show premature stop codons, which suggest that they are putatively encoding.

Table 1. Nucleotide sequences divergence percentage observed between TE sequences.

	<i>D. melanogaster</i>	<i>Z. indianus</i>	<i>D. sechellia</i>	<i>D. mauritiana</i>	<i>D. simulans</i>
<i>D. melanogaster</i>					
<i>Z. indianus</i>	7.7				
<i>D. sechellia</i>	4.7	4.9			
<i>D. mauritiana</i>	4.2	4.3	2		
<i>D. simulans</i>	4.1	3.6	1.8	1.3	

Inferring horizontal transfer

We have estimated the dS values to verify the similarity between TEs and host genes sequences. The dS value provides a measurement of neutral evolution in the absence of a strong codon usage bias. The analysis of the codon bias usage for TE and the nuclear genes were performed, the results are presented in Table 2. The values observed in Nc and CBI showed a lower codon usage bias for the TE when comparing to host genes, which is important to avoid over-estimation of horizontal transfer events. Besides, no host genes showed a very high bias in the codon usage, which is needed to avoid under-estimation of horizontal transfer events.

Table 2: Mean number of effective codons (Nc) and codon bias index (CBI) for *amd*, *amyrel*, *sod*, *cdc6* and TE sequences. The Nc varies between 21 for maximum codon bias and 61 for minimum codon bias. A CBI value of 0 corresponds to no bias, and a value of 1 corresponds to maximum bias.

	<i>amd</i>		<i>amy</i>		<i>sod</i>		<i>per</i>		TE	
	Nc	CBI	Nc	CBI	Nc	CBI	Nc	CBI	Nc	CBI
<i>D. simulans</i>	43.454	0.53	36.318	0.55	36.26	0.668	41.845	0.432	61	0.36
<i>D. sechellia</i>	44.859	0.508	43.732	0.542	35.585	0.643	43,223	0,426	58.94	0.372
<i>D. melano</i>	48.189	0.466	40.919	0.501	35.69	0.626	48,912	0,443	60.706	0.397
<i>Z. indianus</i>	57.835	0.474	47.156	0.462	42.601	0.451			59.01	0.358
<i>D. maurit</i>			39.558	0.511	36.252	0.639	45,902	0,403	53.032	0.365

The comparison between dS values of TE and host genes is presented in Figure 3. In this figure, six comparison of dS values were considered significant when $P < 0.001$. Of these, four are between *Z. indianus* and the other species (*D. melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana*), and two are between *D. melanogaster* and *D. simulans* and between *D. melanogaster* and *D. sechellia*. For all these comparisons, the dS values observed for the transposon were significantly lower than the dS values for the host genes.

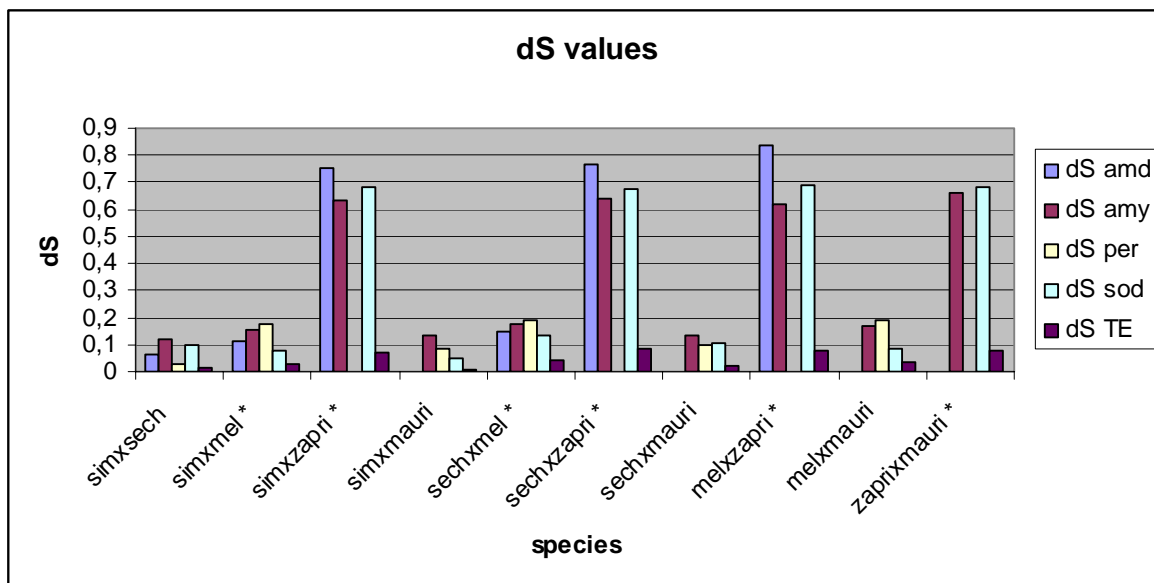


Figure 4: Comparative analysis between dS values for *hosimary* and host genes (*amd*, *amyrel*, *per* and *sod*). Missing columns indicate that sequences are not available. The dS divergence values comparisons between TE and host genes were marked with an asterisk (*) when differences considered statistically significant were found

with two or more host genes. * $P < 0.001$; sim = *D. simulans*; sech = *D. sechellia*; mel = *D. melanogaster*; mauri = *D. mauritiana*; zapri = *Z. indianus*.

Discussion

One of the questions that remain unanswered in evolutionary genomics is about the origin, role and dynamics of the significant amount of TEs that generally occur in the eukaryotes genomes. Apart from being numerous, TEs are very diversified and the different “kinds” of transposable elements may establish origin and relationship strategies with the host genomes that are also variable (Le Rouzic *et al.*, 2007). In this perspective, the characterization of diversity of TEs and their phylogenetic histories is important to understand evolutionary genomics. In this work we have described a new TE family, *hosimary*, which belongs to the *hAT* superfamily. The elements *hosim* and *hosec*, which are now placed inside the *hosimary* family, were originally described by Ortiz and Loreto (2009) using *in silico* search. The authors showed that both share some amino acid sequence similarity to the *herves* element from *Anopheles gambiae* (Arensburger *et al.*, 2005), which led them to propose a new cluster of the *hAT* superfamily, including sequences from *D. mojavensis*, *D. pseudoobscura* and *D. willistoni* as well, which they called the *herves* family. However, Wicker *et al.* (2007) suggest a unified classification system for TEs. In this new classification system, a TE family can be established when no strong hit is observed in the BLASTn search. In addition, the elements of a given family must present at least 80% of nucleotide sequence similarity to each other. The elements analyzed in this work showed no strong hit in the nucleotide sequence with any transposon. Therefore, considering that the nucleotide similarity shared between species was higher than 90% and that no high sequence similarity was detected with other transposable elements, we propose a new transposon family called *hosimary*.

hosimary sequences showed a variable number of copies in the species in which they were found. Our results suggest the presence of one copy in the *D. melanogaster* genome. Comparing the *in silico* localization sites of *hobo* in *D. melanogaster* by FISH analysis, Zakharenko *et al.* (2007) observed that the number of FISH *hobo* sites is almost twice as high as in the annotated genome sequence. Kaminker *et al.* (2002) have found 15 copies of the

hopper element and 24 copies of *hobo* in the *D. melanogaster* genome. Both elements belong to the *hAT* superfamily. Ortiz and Loreto (2009) have found four copies of *hosim* in the *D. simulans* through *in silico* search. However, we have found at least seven copies. According to our FISH analysis, three signals were observed in euchromatic region in the *D. simulans*, and some signals were detected also in the chromocenter. So, it is possible to suggest that other copies of the *hosim* element are located in the heterochromatin region. The preferential insertion in heterochromatin is probably because selection has less impact on heterochromatin than on euchromatin, difficulting the removing of TEs from host genome by selection (Kapitonov and Jurka, 2003; Lerat *et al.*, 2003).

In the sequence analysis no premature stop codon was observed in any of the analyzed sequences, except in the *D. melanogaster* sequences and in one sequence from *D. mauritiana*, which suggests that the elements could be active. The *D. melanogaster* copy shows degeneration of its nucleotide sequences. Its location near the centromeric region (2R chromosome, section 41 – 42) suggests an inactive element (Maside *et al.*, 2005). Active forms of the TEs can be mobilized in the host genomes and thus lead to variability, while inactive forms tend to be extinct by stochastic losses.

Incongruence found in evolutionary analysis of *hosimary* led us to suggest the occurrence of horizontal transfer for this element. First, there is no defined topology in the phylogenetic tree, with some species interspersed to each other. Second, *hosimary* phylogeny shows very short branch lengths, reflecting the high sequence similarity among the species. Third, the discontinuous occurrence (patchy distribution) observed for *hosimary* among species (it was found only in the *melanogaster* subgroup and *Z. indianus* among the 51 *Drosophilidae* species analyzed). These distortions may indicate horizontal transfer. Nevertheless, other explanations should be considered: the high sequence similarity among the species can be explained by different evolution rates of the TE between species, and the patchy distribution of the TE among species can be due to stochastic losses, in which the element is vertically transmitted with losses along the branches. Additionally, ancestral polymorphism could justify the incongruence observed in the phylogeny (Capy *et al.*, 1998; Silva *et al.*, 2004; Loreto *et al.*, 2008). However, the high similarity of the elements and the time of divergence of their host species (55 MYA between *Zaprionus* genus and the

melanogaster group; Robe, 2008), practically excludes the possibility of *hosimary* acquisition from a common ancestor.

Horizontal transfer appears to be a more common phenomenon than previously thought (Silva and Kidwell, 2000; Sanchez-Gracia *et al.*, 2005). Recently, Bartolomé *et al.* (2009) performed a comprehensive genomic analyses comparing nucleotide divergence in the neutrally evolving sites for TEs and host genes to *Drosophila melanogaster*, *D. simulans* and *D. yakuba*. Synonymous substitutions (dS) are significantly lower in number for several TEs that were compared to the 10150 host genes analyzed. The authors produced cogent evidence that HT is very frequent, mainly among LTR retrotransposons and DNA transposons, and that the rate of HT is about 0.035 event per TE family per MY. Others studies also have shown that TE horizontal transfer has occurred in a wide range of organisms, like mammals and other tetrapods (Pace *et al.*, 2008); fishes (Boer *et al.*, 2007), bacteria to insects (Hotopp *et al.*, 2007; Klasson *et al.*, 2009); and plants (Diao *et al.*, 2006; Roulin *et al.*, 2008).

In order to reinforce the occurrence of HT we compared the dS values between TE and the host genes. HT is strongly suggested when dS values are significantly lower for the TE if compared to dS of the host genes in the absence of a strong codon usage bias (Silva and Kidwell, 2000). The nuclear genes employed for the comparisons show variable N_c and CBI values, but not a very high bias in the codon usage, which is needed to avoid under-estimation of HT events. The comparison between dS values of TE and host genes dS values suggests the occurrence of at least three HT events. These events occurred between *Z. indianus* and other species of the *melanogaster* subgroup, between *D. melanogaster* and *D. simulans*, and also between *D. melanogaster* and *D. sechellia*. However, in the *melanogaster* subgroup, introgression cannot be excluded.

D. melanogaster and *D. simulans* are human commensal species, while *D. mauritiana* is a semi-domestic species. *D. sechellia* is specialized and geographically confined. Despite that, the four species of the subgroup *melanogaster* investigated in this paper are of Afrotropical origin (Lachaise and Silvain 2004; David *et al.*, 2007). In addition, the *Zaprionus* genus seems to share the same age and geographic origin as the *melanogaster* subgroup (Remsen and O'Grady, 2002; Yassin *et al.*, 2008). Therefore, there is a physical opportunity of horizontal transfer between *melanogaster* group and *Z. indianus*.

Several works have shown that many transposable elements have been transferred horizontally among the *Zaprionus* genus and the *melanogaster* subgroup. Maruyama and Hartl (1991) suggest HT of the *mariner* element between *Z. tuberculatus* and *D. mauritiana*, since the elements in these species are 97% identical to each other. This event was confirmed by Brunet (1999) and extended to *Z. verruca*. Herédia *et al.* (2004) also report the horizontal transfer of the *gypsy* retroelement between *D. simulans* and *Z. indianus*. Almeida and Carareto (2006) showed the HT of the *copia* element between *Z. tuberculatus* and *melanogaster* subgroup species.

So, it is possible to conclude that the horizontal transfer is a phenomenon that occurs more frequently than previously imagined (Silva and Kidwell, 2000; Sanchez-Gracia *et al.*, 2005; Bartolomé *et al.*, 2009). Furthermore, our results could represent an underestimation of the possible number of HT cases that have occurred in the *melanogaster* subgroup. Alternatively, considering the similarity between sequences, only a yet unknown mechanism, acting in the TE and not in host genes, could explain the lower dS values in the TE than in host genes.

Acknowledgments

To Dra. Claude Bazin for gently providing the Zola strain. To M.Sc. Adriana Ludwig, M.Sc. Nina Roth Mota and Lizandra Robe for the valuable help and critical comments which contributed significantly to the improvement of this manuscript. To M.Sc. Paloma Rubin for sequencing the DNA samples. This work was supported by grants from CNPq.

Supplementary material.

Table 1. List of Drosophilidae species used in this study.

Genus	Subgenus	Section	Group	Species	Strain
<i>Drosophila</i>	<i>Drosophila</i>	<i>quinaria/tripunctata</i>	<i>guarani</i>	<i>D. ornatifrons</i>	
				<i>D. subbadia</i>	
			<i>guaramunu</i>	<i>D. griseolineata</i>	
				<i>D. maculifrons</i>	
			<i>tripunctata</i>	<i>D. tripunctata</i>	
				<i>D. mediodiffusa</i>	
			<i>cardini</i>	<i>D. cardini</i>	
				<i>D. cardinoides</i>	
				<i>D. neocardini</i>	
				<i>D. polymorpha</i>	
				<i>D. procardinoides</i>	
				<i>D. arawakana</i>	
			<i>pallidipennis</i>	<i>D. pallidipennis</i>	
			<i>calloptera</i>	<i>D. ornatipennis</i>	
			<i>immigrans</i>	<i>D. immigrans</i>	
			<i>funebri</i>	<i>D. funebri</i>	
		<i>virilis- repleta</i>	<i>mesophragmatica</i>	<i>D. gasici</i>	
				<i>D. gaucha</i>	
			<i>repleta</i>	<i>D. hydei</i>	
				<i>D. mercatorum</i>	
			<i>annulimana</i>	<i>D. annulimana</i>	
			<i>canalina</i>	<i>D. canalina</i>	
			<i>flavopilosa</i>	<i>D. cestri</i>	
				<i>D. incompta</i>	
			<i>virilis</i>	<i>D. virilis</i>	
			<i>robusta</i>	<i>D. robusta</i>	
	<i>Sophophora</i>		<i>melanogaster</i>	<i>D. melanogaster</i>	Oregon
					Zola
				<i>D. simulans</i>	white
					yellow
					Águas de Lindóia
				<i>D. sechellia</i>	
				<i>D. mauritiana</i>	
				<i>D. teissieri</i>	
				<i>D. santomea</i>	

				<i>D. erecta</i>	
				<i>D. yakuba</i>	
				<i>D. kikkawai</i>	
				<i>D. ananassae</i>	
				<i>D. malerkotiana</i>	
				<i>D. orena</i>	
			<i>obscura</i>	<i>D. pseudoobscura</i>	
			<i>saltans</i>	<i>D. sturtevanti</i>	
				<i>D. saltans</i>	
			<i>willistoni</i>	<i>D. sucinea</i>	
				<i>D. nebulosa</i>	
				<i>D. paulistorum</i>	
				<i>D. willistoni</i>	
	<i>Dorsilopha</i>			<i>D. busckii</i>	
<i>Zaprionus</i>				<i>Z. indianus</i>	
				<i>Z. tuberculatus</i>	
				<i>Z. sepsoides</i>	
				<i>Z. multistriatus</i>	
<i>Scaptodrosophila</i>				<i>S. latifasciaeformis</i>	
				<i>S. lebanonensis</i>	

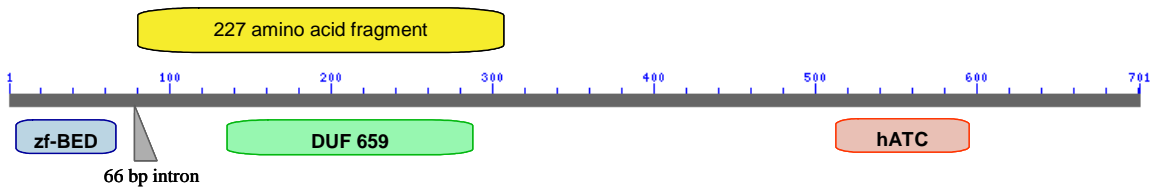


Figure 2: Schematic representation of the transposase amino acid fragment *hosimary* family elements, based on *Hosec1.2*. Below the putative transposase (701 amino acid), the three domains identified by CDD are: Zf-BED, DUF659 and the *hAT* dimerization domain (hATC). The rectangle above the transposase represents the analyzed 227 amino acid fragment used in this study (approximated size and position). The triangle localizes the 66-bp intron of *Hosec1.2*, identified by NetGene2 Server (Brunak *et al.*, 1991; Hebsgaard *et al.*, 1996). These softwares as well as BLASTn and BLASTx are available at NCBI (<http://www.ncbi.nlm.nih.gov>).

Table 2. Accession numbers of the nuclear genes sequences

	<i>amd</i>	<i>amyrel</i>	<i>per</i>	<i>sod</i>
<i>D. melanogaster</i>	NM165278.1	NM057914.3	AF033029.1	NM057387.3
<i>D. simulans</i>	XM002079852.1	XM002081767	AF251249.1	XM00208442.1
<i>D. sechellia</i>	XM002039039.1	XM002034125.1	AF251250.1	XM002030033.1
<i>D. mauritiana</i>		U96157	AF251240.1	AF127158.1
<i>Z. indianus</i>	AY699263.1	EF458322		
<i>Z. tuberculatus</i>				AF021823

Table 3. Time of divergence estimate of the TE between the species. sim = *D. simulans*; sech = *D. sechellia*; mel = *D. melanogaster*; mauri = *D. mauritiana*; zapri = *Z. indianus*.

Pairwise comparison	T
melxzapri	2.3125
melxsech	1.3125
melxmauri	1.0625
melxsim	0.9375
zaprixsech	2.6875
zaprixmauri	2.40625
zaprixsim	2.09375
sechxmauri	0.59375
sechxsim	0.46875
simxmauri	0.28125

Sequences analyzed in this work.

- um arquivo com as sequências será enviado para a revista juntamente com o artigo.

References

- Almeida L and Carareto CM (2006) Sequence heterogeneity and phylogenetic relationships between the *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. *Genet Sel Evol* 38:535-550.
- Arensburger P, Kim YJ, Orsetti J, Aluvihare C, O'Brochta DA and Atkinson PW (2005) An active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. *Genetics* 169(2):697-708.
- Bartolomé C, Bello X and Maside (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol* 10(2):R22.
- Boer JG, Yazawa R, Davidson WS and Koop BF (2007) Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:422.
- Brunak S, Engelbrecht J and Knudsen S (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* 220:49-65.
- Brunet F, Godin F, Bazin C and Capy P (1999) Phylogenetic analysis of *Mos1*-Like transposable elements in the *Drosophilidae*. *J Mol Evol* 49:760-768.
- 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, Bazin C, Higuete D and Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197pp.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Coates CJ, Johnson KM, Perkins HD, Howells AJ, O'Brochta DA and Atkinson PW (1996) The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica* 97:23-31.
- David J, Lemeunier F, Tsacas L and Yassin A (2007) The historical discovery of the nine species in the *Drosophila melanogaster* species subgroup. *Genetics* 177:1969-1973.
- Diao X, Freeling M and Lisch D (2006) Horizontal transfer of a plant transposon. *PLoS Biology* 4:e5.
- Essers L and Kunze R (1995) Transposable elements *Bg* (*Zea mays*) and *Tag1* (*Arabidopsis thaliana*) encode protein sequences with homology to *Ac* like transposases. *Maize Genet Coop Newslett* 69:38-41.

Graur D and Li W (2000) Fundamentals of molecular evolution, 2ed. Sinauer Associates, Sunderland, MA.

Grumbling G and Strelets V (2006) FlyBase: anatomical data, images and queries. Nucleic Acids Research 34:484-488.

Hasegawa M, Kishino H and Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht, J, Rouze P and Brunak S (1996) Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. Nucleic Acids Res 24: 3439-3452.

Hehl R, Nacken WK, Krause A, Saedler H and Sommer H (1991) Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveal homologies to the *Ac* element from maize. Plant Mol Biol 16 (2):369-71.

Herédia FO, Loreto ELS and Valente VLS (2004) Complex evolution of *gypsy* in Drosophilid species. Mol Biol Evol 21:1-12.

Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H and Werren JH (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753-1756.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J *et al.* (2002) The transposable elements on the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol 3:0084.1-0084.20.

Kapitonov VV and Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci USA 100:6569-6574.

Kempken F and Windhofer F (2001) The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma 110:1-9.

Kidwell MG and Lisch DR (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704-7711.

Kidwell MG and Lisch DR (2001) Transposable elements, parasitic DNA, and genome evolution. Evolution 55:1-24.

Klasson L, Kambris Z, Cook PE, Walker T and Sinkins SP (2009) Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. BMC Genomics 10:33.

- Lachaise D and Silvain JF (2004) How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*-*D. simulans* paleogeographic riddle. *Genetica* 120:17-39.
- Lerat E, Rizzon C and Biémont C (2003) Sequence divergence of transposable elements in the *D. melanogaster* genome. *Genome Res* 13:1889-1896.
- Le Rouzic A, Boutin TS and Capy P (2007) Long-term evolution of transposable elements. *Proc Natl Acad Sci USA* 104:19375-19380.
- Loreto EL, Zaha A, Nichols C, Pollock JA and Valente VLS (1998) Characterization of a hypermutable strain of *Drosophila simulans*. *Cell Mol Life Sci* 54:1283-1290.
- Loreto EL, Carareto CMA and Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 100:545-554.
- Ludwig A, Valente VLS and Loreto ELS (2008) Multiple invasions of *Errantivirus* in the genus *Drosophila*. *Insect Mol Biol* 17:113-124.
- Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L *et al.* (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35:237-240.
- Maruyama K and Hartl DL (1991) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J Mol Evol* 33:514-524.
- Maside X, Assimacopoulos S and Charlesworth B (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. *Genet Res* 85:195-203.
- McClintock B (1947) Cytogenetic studies of maize and *Neurospora*. *Carnegie Inst Washington Year Book* 46:146-152.
- Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on *tRNA* availability. *J Mol Evol* 37:273-280.
- Nei M and Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418-426.
- Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala.
- Ortiz M and Loreto EL (2009) Characterization of new *hAT* transposable elements in twelve *Drosophila* genomes. *Genetica* 135:67-75.

Pace II JK, Gilbert C, Clark MS and Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA* 105:17023-17028.

Remsen J and O'Grady P (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol Phylogenet Evol* 24:249-264.

Robe LJ, Valente VLS, Budnik M and Loreto ELS (2005) Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: A nuclear versus mitochondrial gene approach. *Mol Phylogenet Evol* 36:623-640.

Robe LJ (2008) Relações filogenéticas no gênero *Drosophila* (Diptera, Drosophilidae): uma abordagem molecular. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. pp 303.

Robert CE (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

Roulin A, Piegu B, Wing RA and Panaud O (2008) Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon *RIRE1* within the genus *Oryza*. *Plant J* 53:950-959.

Rozas J, Sanchez-DelBarrio JC, Messenguer X and Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.

Rubin E, Lithwick G and Levy AA (2001) Structure and evolution of the *hAT* transposon superfamily. *Genetics* 158:949-957.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.

Sanchez-Gracia A, Maside X and Charlesworth B (2005) High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet* 21:200-203.

Sassi AK, Herédia FO, Loreto ELS, Valente VLS and Rohde C (2005) Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. *Genet Mol Biol* 28:734-739.

Silva JC and Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* element. *Mol Biol Evol* 17:1542-1557.

Silva JC, Loreto EL and Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-72.

Sharp PM and Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol 28:398-402.

Souames S, Bazin C, Bonnivard E and Higuët D (2003) Behavior of the *hobo* Transposable Element with Regard to TPE Repeats in Transgenic Lines of *Drosophila*. Mol Biol Evol 20(12):2055-2066.

Staden R (1996) The Staden sequence analyses package. Mol Biotechnol 5:233-241.

Tamura K (1992) Estimation of the number of nucleotide substitution when there are strong transition-transversion and G + C biases. Mol Biol Evol 9:678-687.

Tamura K, Dudley J, Nei M and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

Warren WD, Atkinson PW and O'Brochta DA (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, Paux E, SanMiguel P and Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973-982.

Wright F (1990) The "effective number of codons" used in a gene. Gene 87:23-29

Yassin A, Araripe LO, Capy P, Da Lage JL, Klaczko LB, Maisonhaute C, Ogereau D and David JR (2008) Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Díptera: Drosophilidae). Mol Phylogenet Evol 47:903-15.

Zakharenko LP, Kovalenko LV and Mai S (2007) Fluorescence *in situ* hybridization analysis of *hobo*, *mdg1* and *Dm412* transposable elements reveals genomic instability following the *Drosophila melanogaster* genome sequencing. Heredity 99(5):525-530.

CAPÍTULO 3

THE *HOBO* TRANSPOSON AND *HOBO*-RELATED ELEMENTS ARE EXPRESSED AS DEVELOPMENTAL GENES IN *DROSOPHILA*

Maríndia Deprá¹, Vera Lúcia da Silva Valente¹, Elgion Lúcio da Silva Loreto^{1,2}

¹ Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil

² Departamento de Biologia, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brazil

Manuscrito submetido à revista **GENE**

Abstract

hobo is an active class II transposable element in *D. melanogaster* that is able to induce gonadal dysgenesis. This element was recently introduced into the genomes of the *melanogaster* subgroup by horizontal transfer. *hobo*-related sequences (*hRSs*) are thought to be relics of old “*hobo*” invasions, and are therefore ancient genomic constituents. However, some of these *hRSs* are still mobile. In the present study, we analyzed the expression pattern of *hobo* and one particular class of *hRSs*, *hobo*^{VAHS}. We show that both elements are expressed as sense and antisense mRNA transcripts. It has been suggested that this expression may be involved in the regulation of *hobo* transposition by iRNA. Expression analysis in whole mount embryos revealed a pattern similar to that of some developmental regulatory genes. It has been suggested that *cis*-regulatory sequences similar to those in developmental genes exist in *hobo* sequences. Therefore, *hobo* mobilization may contribute to the development of new regulatory networks during genomic evolution.

Key words: *hAT* superfamily, relic *hobo*, transposable elements, regulatory genes, homeotic genes.

Introduction

When Barbara McClintock discovered transposable elements (TEs), she referred to them as “controlling elements.” Later, many researchers defined TEs as selfish or parasitic sequences (Doolittle and Sapienza 1980; Orgel and Crick 1980), or even as “junk DNA.” Despite their parasitic nature, TEs have a strong influence on the evolutionary trajectory of their hosts by creating the genetic variability that is necessary for evolution (Kidwell and Lisch, 1997; 2001). Advantageous genetic variability in the host may be obtained by the insertion of a TE into a gene regulatory region, which generates a new regulatory pattern by adding additional TE regulatory sequences. Individual examples of selectively beneficial TE insertions with apparent regulatory functions have been described in *Drosophila* and other species (Feschotte, 2008). Chung et al. (2007) showed that the insertion of the *Accord* TE is responsible for the increased expression of an insecticide resistance gene. Additionally, they

demonstrated that the *Accord* LTR carries regulatory sequences that change the spatial expression of the gene, resulting in an insecticide resistance phenotype. Thornburg et al. (2006) showed that some mobile elements may be rich in transcription factor binding sites. Jordan et al. (2003) reported that nearly 25% of experimentally characterized human promoters, including *cis*-regulatory elements, contain TE-derived sequences.

In contrast, the mobilization of TEs can have negative effects on the fitness of their hosts, and, therefore demand strategies for transposition control. A recent review (Girard and Hannon, 2008) suggested that mechanisms related to RNA interference act as key mediators of transposition suppression. In particular, piwi-interacting RNAs (piRNAs) are thought to be major regulators of transposon activity (Brennecke et al., 2007). Nevertheless, the extension in that TE parasite the genomes or the actual contribution that they do for the hosts' evolution is already an open question.

The *hobo* element of *Drosophila melanogaster* is a Class II TE that belongs to the *hAT* superfamily and transposes via a DNA intermediary (Wicker et al., 2007). A significant amount of information about this TE has been obtained in *D. melanogaster* and relative species, and this information may help answer the unresolved issues mentioned above. In the genomes of different species of the *melanogaster* subgroup, the *hobo* element is present in three forms: i) complete and autonomous elements, ii) internally deleted elements and iii) *hobo*-related sequences (*hRS*), which are thought to be relics of older *hobo* elements (Anxolabéhère et al., 1988). We have previously characterized one type of *hobo*-related sequence, referred to as *hobo*^{VAHS}, which includes nonautonomous elements that have maintained mobility for a significantly long evolutionary time in the genomes of subgroup *melanogaster* flies (Torres et al., 2006; Ortiz and Loreto, 2008). These elements have intact terminal inverted repeats (TIRs) and conserved subterminal regions that are significantly similar to the canonical *hobo*. It has been suggested that these relic *hobo* elements may be mobilized *in trans* by the canonical *hobo* element (Torres et al., 2006; Ortiz and Loreto, 2008).

In the present study, we describe the transcriptional expression patterns of both canonical *hobo* and *hobo*^{VAHS} in whole mount embryos. In the first phase of development, the expression patterns of these transposable elements are similar to some developmental regulatory genes. In posterior developmental phases, TEs are expressed in the nerve chord.

This expression pattern suggests that these TEs may have *cis*-regulatory sequences that are recognized by transcription factors, as ones that switching on some homeotic genes. Additionally, antisense transcripts were detected, suggesting the existence of a silencing mechanism based on interference RNA (iRNA).

Material and methods

Drosophila strains

Two strains of *D. simulans* were used in this study: white (originated from a *D. simulans* hypermutable strain (Loreto et al., 1998a) and yellow (Loreto et al., 1998b). Two *D. melanogaster* strains were also used: Oregon R and Zola (Souames et al., 2003). These strains were chosen because they have different quantities of the canonical *hobo* elements in their genomes (see cited bibliography and results).

DNA extraction and PCR

Genomic DNA was extracted from adult flies according to Sassi et al. (2005). To detect the presence of *hobo* and *hobo*^{VAHS} in the fly genomes, the following primers were used: hoboCN 991 – 5' ACCGTCGACATGTGGAC 3' and hoboCN 1598 – 5' GGATGGCAATAGGAAGC 3' for the canonical *hobo*, and hoboVA 23 – 5' TAGCCATCGATCAACCAACA 3' and hoboVA 431 – 5' AGAGCCACGCCGAAAGAATAGT 3' for the *hobo*^{VAHS}. PCR analyses were performed in 25 µl reactions using 20 ng of genomic DNA, 1 U Taq DNA Polymerase (Invitrogen), 1X reaction buffer, 200 µM dNTPs, 20 pmol of each primer and 1.5 mM MgCl₂. The amplification conditions were 95°C for 5 min, 30 cycles at 95°C for 40 s, 55°C for 40 s and 72°C for 1 min, followed by a final extension cycle at 72°C for 5 min.

Southern Blot

Approximately 6 µg of DNA samples were digested with *XhoI* (Invitrogen), separated by electrophoresis on 1% agarose gels and transferred to nylon membranes (HybondN+, GE Healthcare). The membranes were hybridized with a probe corresponding to the *XhoI* fragment obtained by cleavage of the pHFL1 plasmid (Calvi et al., 1991). In order to label and

detect nucleic acids, Gene Images and a CPD Star Detection kit (GE Healthcare) were used according to manufacturer's instructions.

in situ hybridization in whole mount embryos

For canonical *hobo* and *hobo*^{VAHS} riboprobe synthesis, the transposons were amplified using the template plasmids pHFL1 and pHVA as previously described. The resulting PCR products were cloned into pCR4-TOPO (Invitrogen). Resulting plasmids were then linearized and digoxigenin labeled using a DIG-labeled dNTP mix (Roche). For the canonical *hobo*, the antisense probe was labeled by T3 *in vitro* transcription of the *NotI*-linearized plasmid. The sense *hobo* element riboprobe was labeled by T7 *in vitro* transcription of the *PstI*-digested pCR4-TOPO plasmid. The *hobo*^{VAHS} sense and antisense probes were created by cleavage with *NotI* and *PstI* respectively, because the orientation of the inserted PCR fragment into each plasmid is in the inverse orientation to that of the canonical *hobo*. *in situ* hybridization was performed in embryos collected in different developmental stages as previously described (Tautz and Pfeifle, 1989). The embryonic stages were identified according to the criteria outlined by Campos-Ortega and Hartenstein (1985).

RT-PCR

RNA was obtained from adult flies following Trizol protocols (Invitrogen) and treated with *DNaseI* (Promega) to eliminate DNA contamination. cDNA synthesis was performed using the M-MLV enzyme (Invitrogen). We synthesized two cDNAs using either random primers or specific primers designed to detect sense and antisense transcripts separately. The primers hoboCN991 and hoboCN1597 were used to detect canonical *hobo* sense and antisense transcripts, respectively, while hoboVA23 and hoboVA431 primers were used to detect the sense and antisense *hobo*^{VAHS} transcripts, respectively. After the synthesis of the first cDNA strand, PCR for the *hobo* and *hobo*^{VAHS} transcripts was performed as previously described. The same specific primer combination was used in the PCR, except for the canonical *hobo*. For this element, another reverse primer was utilized, producing a small, 438 bp fragment (hoboCN 1407 – 5' GTAGTTGGAGTTCCATCTAGT 3'). Additionally, the *Gpdh* (*Glycerol 3 phosphate dehydrogenase*) housekeeping gene was used as a control of amplification and

DNA contamination. The primers for *Gpdh* amplification were designed using two exons, generating a 703 bp fragment in DNA amplification, and 374 bp fragment in cDNA amplification. The *Gpdh* primers used were: GPDH 40 – 5' GAGGTGGCTGAGGGCAACTT 3' and GPDH 723 – 5' GAAGTCACAAACGCCTCAGA 3'. Furthermore, other control reactions were performed alongside RNA samples to test for possible DNA contamination.

in silico search for regulatory sequences

To identify potential transcription factor binding sites in canonical *hobo* and *hobo*^{VAHS}, their sequences (accession number: M69216 and AY764286, respectively) were examined using Patch 1.0 software (Authors: Jochen Striepe and Ellen Goessling) with the pattern search program using TRANSFAC 6.0 public sites. Default parameters were used for this search. This program is available at <http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi>.

Results

The presence of *hobo* in fly genomes

We have performed a PCR analysis of genomic DNA to verify the presence of *hobo* and *hobo*^{VAHS} in the different *Drosophila* strains. As can be seen in Figure 1, the expected 607 bp fragment of the canonical *hobo* was detected in the *D. simulans* (white and yellow) and *D. melanogaster* Oregon R strains. A smaller fragment was observed in the *D. melanogaster* Zola strain. This strain was previously described as empty of canonical *hobo* element (Souames et al., 2003). The 408 bp *hobo*^{VAHS} fragment was readily amplified in *D. simulans* white and yellow strains, while a smaller fragment was observed in the other species (Figure 1C). Southern blot analysis confirmed the presence of canonical *hobo* in the *D. simulans* white and *D. melanogaster* Oregon R strains, as a 2.6 kb band was observed in the *XhoI* digested genomic DNA (Figure 2). In the *D. melanogaster* Oregon R strain, smaller fragments, likely corresponding to truncated *hobo* elements, were seen. Additionally, longer fragments were observed at the top of the blot. These fragments correspond to *hobo*-related sequences and

were seen in all of the strains. The *D. simulans* yellow strain did not show the expected 2.6 kb band of canonical *hobo*.

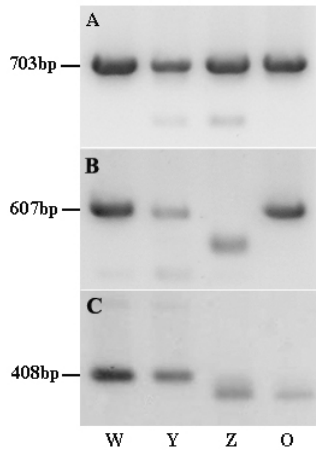


Figure 1: Products of PCR amplifications for (A) Gpdh gene (control), (B) canonical *hobo* and (C) *hobo*^{VAHS}. W= *D. simulans* white strain; Y= *D. simulans* yellow strain; Z= *D. melanogaster* Zola; O= *D. melanogaster* Oregon R.

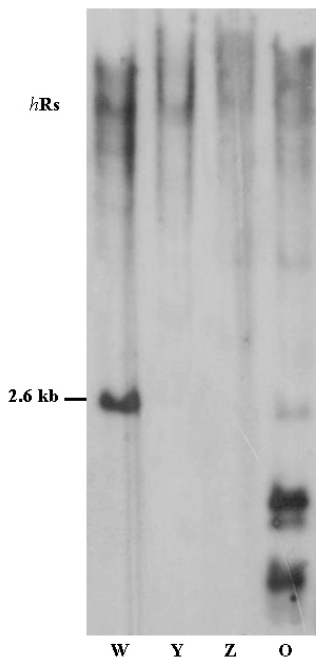


Figure 2: Southern blot of genomic DNA digested with *XhoI*. The membrane was hybridized with a probe to canonical *hobo* that corresponded to a 2.6 kb fragment. The *hRSs* (*hobo* related sequences) are sequences that are

similar to the canonical *hobo*. W= *D. simulans* white strain; Y= *D. simulans* yellow strain; Z= *D. melanogaster* Zola; O= *D. melanogaster* Oregon R.

Transcriptional expression of *hobo* and *hobo*^{VAHS}

RT-PCR analysis was performed to verify the transcriptional expression of *hobo* and *hobo*^{VAHS} in each strain. Using specific primers for cDNA synthesis, expression of canonical *hobo* transcripts, both sense and antisense, were detected in the *D. simulans* white and *D. melanogaster* Oregon R strains (Figure 3B). Canonical *hobo* antisense transcripts of unexpected size were detected in the *D. simulans* yellow and *D. melanogaster* Zola strains. Sense and antisense transcripts of the *hobo*^{VAHS} element were observed in all strains (Figure 3C), but antisense transcripts showed some bands of unexpected size. Similar results were obtained when random primers were used for cDNA synthesis of the first strand (data not shown). The amplification of only a 374 bp fragment of *Gpdh*, corresponding to processed mRNA, showed the absence of genomic DNA in all the samples.

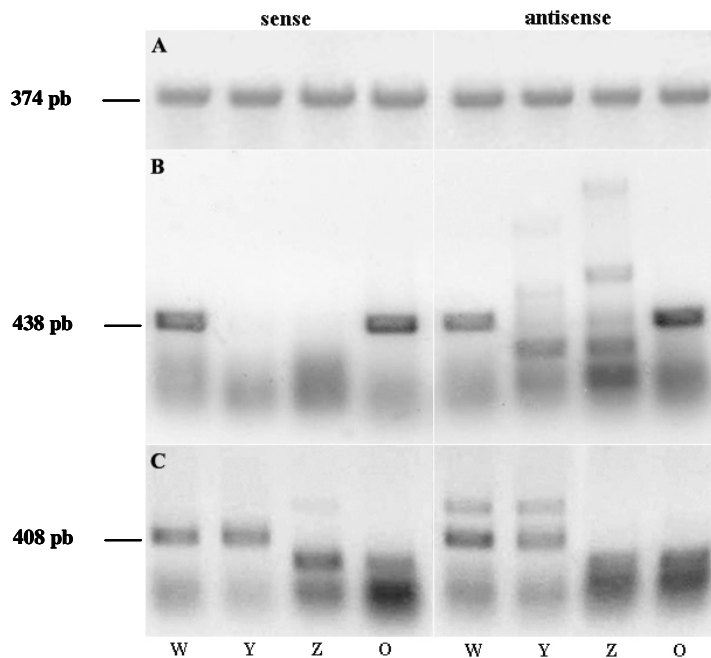


Figure 3: RT- PCR of canonical *hobo* and *hobo*^{VAHS} in different strains show sense and antisense transcripts. In (A), *gpdh* transcripts were used as a control. In (B), the products of canonical *hobo* transcript amplification are shown and in (C), products of *hobo*^{VAHS} transcript amplification are shown. W= *D. simulans* white strain; Y= *D. simulans* yellow strain; Z= *D. melanogaster* Zola; O= *D. melanogaster* Oregon R.

in situ hybridization of whole mount embryos

To determine the spatial and temporal distribution of the transcriptional expression of *hobo* and *hobo*^{VAHS}, *in situ* hybridization experiments using DIG-labeled sense and antisense probes were conducted at different developmental embryonic stages. The *D. simulans* and *D. melanogaster* strains have different quantities of these elements in their genomes. Thus, distinct patterns of expression can be seen during embryogenesis. We describe below the expression pattern of *hobo*^{VAHS}, which is virtually identical to that of the canonical *hobo* element.

hobo^{VAHS} antisense probe

Expression of *hobo*^{VAHS} was initially detected in the syncytial blastoderm during early embryogenesis, in a uniform and widespread pattern (Figure 4A). In the cellular blastoderm, an intense accumulation of *hobo*^{VAHS} transcripts occurs at the periphery of the embryos (Figure 4B). These patterns were detected in all strains examined. In *D. melanogaster* Zola, however, these were the only signals observed and subsequent expression was not detected. At later developmental stages, expression was detected in a strip just anterior to where the cephalic furrow will later form (Figure 4C). An additional strip was also detected at the posterior embryo pole (Figure 4D). The anterior and posterior strips persist as the embryo develops and the cells in between them begin to express *hobo*^{VAHS}. A total of seven strips can be seen in the late cellular blastoderm (Figure 4E). Around the time of the germ band extension stage, staining of the seven stripes decreased and another signal was detected near the head of the embryo (Figure 4F). Later, an accumulation of *hobo*^{VAHS} transcripts occurs in the central nervous system (CNS) along the anterior-posterior axis, with the highest levels in the brain and nerve cord (Figure 4G). During late embryogenesis, transcript accumulation remains in the CNS and is also present in the tracheal networks (Figure 4H-I). Although all of these patterns were observed in the yellow and Oregon R strains, only the stripe staining was detected in the white strain.

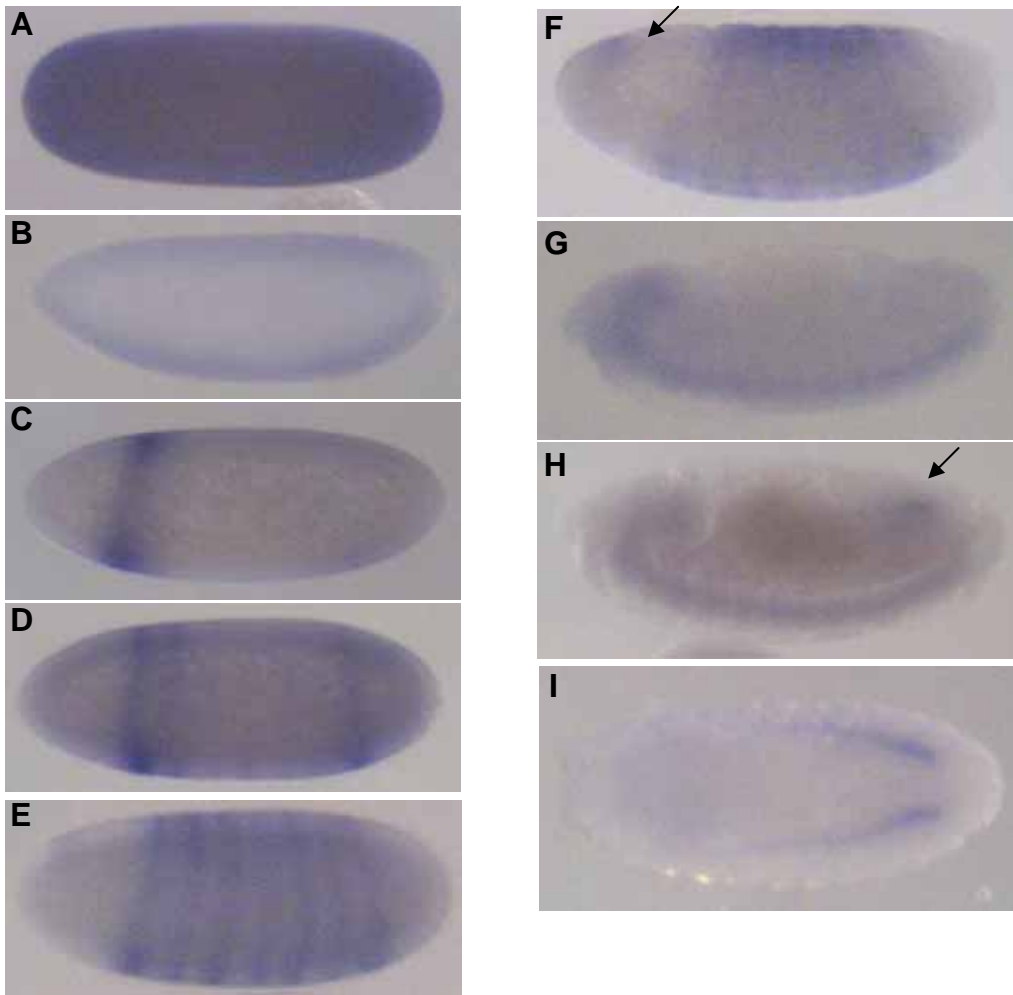


Figure 4: The spatial expression patterns of *hobo* and *hobo*^{VAHS} during embryogenesis. (A) Syncytial blastoderm stage. (B) Cellular blastoderm stage, when an accumulation of *hobo*^{VAHS} transcripts is present at the periphery of the embryos. (C - E) Late cellular blastoderm stage, showing the formation of seven stripes. (F) Germ band extension stage; the arrow indicates the signal in the head region. (G - H) Late embryogenesis, with staining in the CNS; the arrow indicates the tracheal system, which can also be seen in (I). Orientation of embryos: anterior is to the left. Lateral (A - G) and dorsal (I) views are shown.

hobo^{VAHS} and *hobo* sense probes

To detect antisense transcripts of *hobo* and *hobo*^{VAHS} elements, we used a sense probe. The observed pattern is similar for both elements and transcriptional expression can be seen primarily in the first developmental stage. Using a *hobo*^{VAHS} sense riboprobe, expression was detected in the syncytial blastoderm in a uniform distribution, similar to that of maternal

transcripts (Figure 5A). At the cellular blastoderm stage, *hobo*^{VAHS} antisense transcripts accumulated at the periphery of the embryos (Figure 5B). These patterns were detected in all strain examined, but were the only signals observed in the Zola strain. In the *D. simulans* and Oregon R strains, accumulation of *hobo*^{VAHS} antisense transcripts were also detected in the CNS (Figure 5C) and in the tracheal networks (Figure 5D).

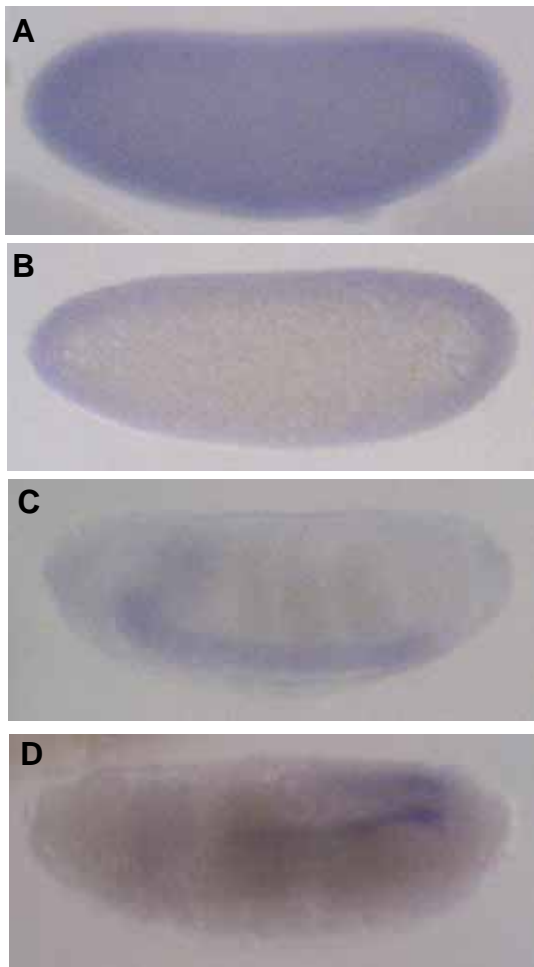


Figure 5: *hobo* and *hobo*^{VAHS} spatial expression during embryogenesis using sense probes. (A) Syncytial blastoderm stage. (B) Cellular blastoderm stage. (C) Germ band retraction stage, with staining in the CNS. (D) Transcriptional expression in the tracheal system. Orientation of embryos: anterior is to the left. Lateral (A - D) views are shown.

in silico search for regulatory sequences

The *in silico* search for putative transcription factor binding sites has revealed 235 segments of canonical *hobo* sequence as potential binding sites. In the *hobo*^{VAHS} sequence, 110 segments were identified as potential binding sites.

Transcription factor binding sites such as Ftz (*Fushi tarazu*), Eve (*Even-skipped*), D1 (*Dorsal*), Hb (*Hunchback*), Zeste, BEAF-32A/B (boundary element-associated factor of 32 kDa), EcR (ecdysone receptor) and others were identified for both elements.

Since we have detected similar spatial expression patterns for both elements, we compared the sequence region that is the same in the two elements. As shown in Figure 5A, in the 5' region, *hobo* has putative binding sites for Zeste, EcR, D1, Cad (caudal) and GAGA-factor upstream of the start codon. In contrast, *hobo*^{VA} (Figure 5B) has putative binding sites for Bcd, EcR, Cad and Zeste in its 5' region. In the 3' terminal regions, however, both elements have the same putative binding sites, with the exception of a GAGA-factor binding site, which is only present in the canonical *hobo* sequence. The GAGA protein is associated with the transcription of numerous *Drosophila* genes, including *hsp70*, *hsp26* and *Ubx*, and can also act as a potent transcriptional activator (O'Brien et al., 1995).

In the inner sequence region, which is more variable between the elements, the putative binding sites are virtually the same. Although binding sites for *Drosophila* heat shock factor (HSF) were only detected in the canonical *hobo* sequence, binding sites for the EcR transcription factor, which is involved in the developmental regulation of heat shock genes (Luo et al., 1991), and BEAF-32A/B, a boundary element of *hsp70* genes, (Hart et al., 1997) were detected.

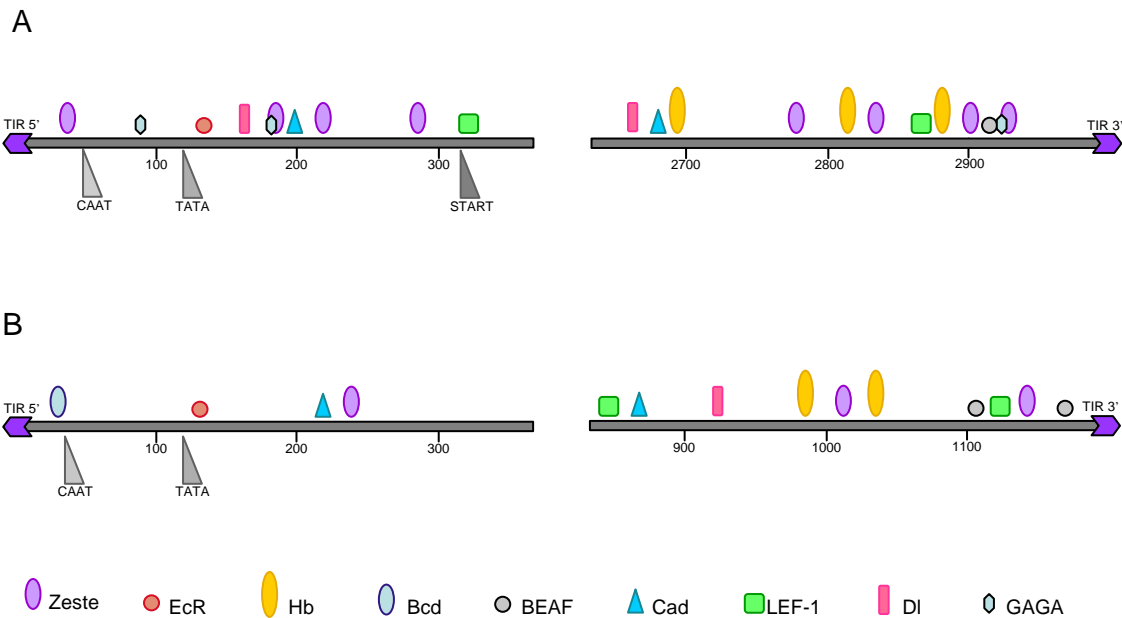


Figure 5: Predicted transcription factor binding sites in homologous regions of the canonical *hobo* and *hobo*^{VAHS} elements, implemented by the software Patch 1.0. (A) binding sites in the canonical *hobo* sequence and (B) binding sites in the *hobo*^{VAHS} sequence. The putatively CAAT and TATA box in the *hobo*^{VAHS} sequence are also represented.

Discussion

Several studies have focused on the populational and evolutionary aspects of the *hobo* element in *D. melanogaster* and relative species. Some evidence indicates that the canonical *hobo* element is a recent acquisition to the *D. melanogaster* and *D. simulans* genomes through horizontal transfer (Daniels et al., 1990; Periquet et al., 1990, 1994; Simmons, 1992). In contrast, *hobo* relics or *hobo*-related sequences (*hRSs*) are thought as related to ancient *hobo* elements present in the ancestral of *melanogaster* ancestral group (Lim, 1988; Daniels et al., 1990; Galindo et al., 2001). In *D. melanogaster*, the complete *hobo* element is active and is capable of producing hybrid dysgenesis syndrome (Blackman et al., 1989; Yannopoulos et al., 1987). Furthermore, some *hRSs* have remained mobile for a long evolutionary period, likely using the canonical *hobo* as a source of transposase (Ortiz and Loreto, 2008). However, the dynamic relationship between these sequences and the mechanisms behind their control are

poorly understood. The strains studied in this work possess varying quantities of canonical *hobo* and *hobo*^{VAHS} elements. Therefore, expression analysis can be useful for understanding these relationships and mechanisms.

The Zola strain of *D. melanogaster* was previously shown to not have a canonical *hobo* element (Souames et al., 2003) and the *D. simulans* yellow showed positive PCR amplification to inner region of canonical *hobo*. However, in our studies, *hobo* was not detected by Southern blot or RT-PCR analysis in either of these strains. When examining the expression of antisense mRNA, only bands of sizes different than that expected were observed for these strains. However, the white strain of *D. simulans* and the Oregon R strain of *D. melanogaster* had both sense and antisense transcripts. Recent work has suggested that RNA interference mechanisms act as mediators of transposition repression (review in Girard and Hannon, 2008). Shpiz et al. (2009) detected sense and antisense HeT-A transcripts in *Drosophila*, and suggest that antisense transcripts of transposable elements serve as a source for rasiRNA generation. Chambeyron et al. (2008) also detected antisense RNAs transcribed by defective *I* elements. These antisense RNAs are likely used to produce antisense piRNAs. Furthermore, efficient production of these antisense small RNAs requires the presence of functional *I* elements in the genome. Interestingly, the Zola strain is highly permissive to *hobo* mobilization, resulting in an increased occurrence of gonadal dysgenesis (GD) symptoms (Souames et al., 2003). However, GD has not been described in *D. simulans*. We have tested the yellow strain for GD in two temperatures (20°C and 28°C, data not shown) and have not observed this phenomenon. *D. melanogaster* and *D. simulans* are apparently affected differently by this trait.

It has been suggested that deleted elements may be involved in regulating the transposition of autonomous elements. Therefore, it is possible that *hobo*^{VAHS} is involved in the control of canonical *hobo* transposition (Ortiz and Loreto, 2008).

Natural populations show significant variation in *hobo* copy number (Vieira et al., 1999; Ruiz and Carareto, 2003). However, little information is available concerning the number of different “kind of *hobo*-” elements in the genome. In *D. melanogaster*, Kaminker et al. (2002) found only one full length canonical copy and 23 internally deleted copies. Ortiz and Loreto (2008) found 11 copies of *hobo*^{VAHS}. In the sequenced genome of *D. simulans* we found only one copy of full length canonical *hobo* and a few copies of internally deleted *hobo* are present

(data not shown). However, 147 copies of *hobo*^{VAHS} were previously found in the sequenced genome (Ortiz and Loreto, 2008). The expression of *hobo*^{VAHS} in the *D. simulans* yellow strain is significant, with as much sense and antisense mRNA as can be seen in the whole mount embryo analysis. However, the *D. melanogaster* Zola strain does not show significant *hobo*^{VAHS} expression. This may suggest that *hobo*^{VAHS} is involved in the control of gonadal dysgenesis in *D. simulans*, and that the high number of *hobo*^{VAHS} elements observed in the *D. simulans* genome may be responsible for the differences in *D. simulans* and *D. melanogaster* regarding hybrid dysgenesis.

hobo and *hobo*^{VAHS} are primarily expressed in early embryogenesis and in the central nervous system. These patterns are similar to those previously reported for some retrotransposons. Ding and Lipshitz (1994) detected expression of both *297* and *opus* in the yolk nuclei of early embryos and in the central nervous system. In addition, they detected *opus* expression in seven stripes during the cell blastoderm and early gastrulation stages. As this pattern evolves into 14 stripes, the authors suggest that this pattern may be characteristic of some segment polarity genes (e.g. *engrailed*). However, expression in the tracheal system has not been previously reported for transposable elements. The developing tracheal system appears to be dependent on expression of the *Hsp60C* gene (Sarkar and Lakhotia, 2005).

The conserved expression patterns and their similarity to that of developmental genes indicate that *hobo* and *hobo*^{VAHS} expression are linked, directly or indirectly, to the same host genome factor. Since TEs contain their own *cis*-regulatory sequences, they may be capable of changing the transcriptional start sites of adjacent genes. Several studies have demonstrated that TEs can directly influence the regulation of nearby gene expression, both at the transcriptional and post-transcriptional level (review in Feschotte, 2008). Brönnner et al. (1995) found that the *B104* transposable element is capable of acting in conjunction with a promoter of the host genome.

We have searched for putative transcription factor binding sites in the *hobo* and *hobo*^{VAHS} sequences. A large number of transcription factors regulate the genes involved in *Drosophila* development. For example, the products of the *Drosophila* gap genes, *hunchback* and *krüppel*, bind to the *hunchback* promoters (Treisman and Desplan 1989). However, *krüppel* is also a target for the redundant activator/repressor system regulated by the anterior

morphogens bicoid and hb (Hoch et al., 1991). Furthermore, *hunchback*, together with the segmentation genes *tailless* and *ftz*, regulates the bx region enhancer, a distant *cis*-element of the *Drosophila Ubx* gene (Qian et al., 1991). The *ftz* gene is located within a cluster of homeotic genes that comprise the Antp complex (Wakimoto and Kaufman 1981). The *Drosophila fushi tarazu* upstream element is an enhancer-like element that is required for the correct expression of *ftz* in developing embryos and directs transcription from a minimal promoter in a *ftz*-like seven-stripped pattern (Pick et al., 1990). However, the seven striped expression pattern can also be detected for other pair-rule genes, such as engrailed and even-skipped (Akam, 1987). In this way, the regulatory genes act in a cascade during embryonic development using a myriad of *cis*-regulatory sequences. Some of these sequences are present in *hobo* and *hobo*^{VAHS}. Interestingly, the *hobo*^{VAHS} element was identified in a hypermutable strain from which many homeotic mutants were isolated (Loreto et al., 1998), and some of these strains were temperature sensitive for phenotypic expression. Interestingly, binding sites for EcR and BEAF-32A/B, which are associated with the transcriptional activity of heat shock genes, were found in the *hobo*^{VAHS} sequence.

Although sequence scans can identify candidate binding sites, confirmation that a particular sequence motif can actually act in transcriptional regulation requires direct experimental examination (Wray 2003). As the regulatory network that drives *Drosophila* development is very complicated, it is difficult to know in practice when all the binding sites that regulate transcription of a particular locus have been identified.

Our results suggest that *hobo* and *hobo*^{VAHS} contain *cis*-regulatory elements that are able to respond to some developmental genes. There is growing evidence that changes in *cis*-regulatory elements are important in determining key phenotypic differences (Bourque et al., 2008). Many TEs that can add a variety of functions to their targeted genes have been described. These additions include polyadenylation sites, promoters, enhancers and silencers (Trhonburg et al., 2006). The TEs may also have evolved such that they have acquired a novel function that serves their host (Bejerano et al., 2006). In addition to their deleterious effects, TEs may provide potentially beneficial developmental and evolutionary effects by acting as “control elements” in eukaryotic genomes, as initially suggested by Barbara McClintock.

References

- Akam M., 1987. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* 101, 1 – 22.
- Anxolabéhère, D., Kidwell, M.G., Periquet, G., 1988. Molecular characteristics of diverse population are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. *Mol. Biol. Evol.* 5, 252 – 269.
- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, J.W., Haussler, D., 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441, 87 – 90.
- Blackman, R.K., Koehler, M.M., Grimaila, R., Gelbart, W.M., 1989. Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J.* 8, 211–217.
- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., Liu, E.T., 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752 – 1762.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., Hannon, G.J., 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089 – 1103.
- Brönner, G., Taubert, H., Jäckle, H., 1995. Mesoderm-specific B104 expression in the *drosophila* embryo is mediated by internal *cis*-acting elements of the transposon. *Chromosoma* 103, 669 – 675.
- Calvi, B.R., Hong, T.J., Findley, S.D., Gelbart, W.M., 1991. Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator* and *Tam3*. *Cell* 66, 465 – 471.
- Campos-Ortega, J.A., Hartenstein, V., 1985. The embryonic development of *Drosophila melanogaster*. Springer- Verlag Berlin Heidelberg, Germany, 227p.
- Chambeyron, S., Popkova, A., Payen-Groschêne, G., Brun, C., Laouini, D., Pelisson, A., Bucheton, A., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14964–14969.
- Chung, H., Bogwitz, M.R., McCart, C., Andrianopoulos, A., Ffrench-Constant, R.H., Batterham, P., Daborn, P.J., 2007. *Cis*-Regulatory Elements in the *Accord* Retrotransposon Result in Tissue-Specific Expression of the *Drosophila melanogaster* Insecticide Resistance Gene *Cyp6g1*. *Genetics* 175, 1071 – 1077.

- Daniels, S.B., Chovnick, A., Boussy, I.A., 1990. Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol. Biol. Evol.* 7, 589 – 606.
- Ding, D., Lipshitz, H.D., 1994. Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. *Genet. Res.* 64, 167 – 181.
- Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603.
- Feshotte, C., 2008 Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397 – 405.
- Galindo, M.I., Bigot, Y., Sánchez, M.D., Periquet, G., Pascual, L., 2001. Sequences homologous to the *hobo* transposable element in E strains of *Drosophila melanogaster*. *Mol. Biol. Evol.* 18, 1532–1539.
- Girard, A., Hannon, G.J., 2008. Conserved themes in small-RNA mediated transposon control. *Trends Cell Biol.* 18, 136 – 148.
- Hart, C., Zhao, K., Laemmli, U.K., 1997. The scs' Boundary Element: Characterization of Boundary Element-Associated Factors. *Mol. Cell. Biol.* 17, 999 – 1009.
- Hoch, M., Seifert, E., Jäckle, H., 1991. Gene expression mediated by *cis*-acting sequences of the *Krüppel* gene in response to the *Drosophila* morphogens bicoid and hunchback. *EMBO J.* 10, 2267 – 2278.
- Jordan, I.K., Rogozin, I.B., Glazko, G.V., Koonin, E.V., 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 19, 68–72.
- Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M., Celniker, S.E., 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3, 1 – 20.
- Kidwell, M.G., Lisch, D.R., 1997. Transposable elements as sources of variation in animals and plants. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7704 – 7711.
- Kidwell, M.G., Lisch, D.R., 2001. Transposable elements, parasitic DNA, and genome evolution. *Evolution* 55, 1 – 24.
- Lim, J.K., 1988. Intrachromosomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 85, 9153–9157.
- Loreto, E.L.S., Zaha, A., Nichols, C., Pollock, J.A., Valente, V.L.S., 1998a. Characterization of a hypermutable strain of *Drosophila simulans*. *Cell. Mol. Life Sci.* 54, 1283 – 1290.

- Loreto, E.L.S., Zaha, A., Valente, V.L.S., 1998b. Transposable elements in South American populations of *Drosophila simulans*. *Genet. Sel. Evol.* 30, 171 – 180.
- Luo, Y., Amin, J., Voellmy, R., 1991. Ecdysterone Receptor Is a Sequence-Specific Transcription Factor Involved in the Developmental Regulation of Heat Shock Genes. *Mol. Cell. Biol.* 11, 3660 – 3675.
- O'Brien, T., Wilkins, R.C., Giardina, C., Lis J.T., 1995. Distribution of GAGA protein on *Drosophila* genes *in vivo*. *Genes Dev.* 9, 1098 – 1110.
- Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. *Nature* 284, 604–607.
- Ortiz, M.F., Loreto, E.L.S., 2008. The *hobo*-related elements in the melanogaster species group. *Genet. Res.* 90, 243 – 252.
- Periquet, G., Hamelin, M. H., Kalmes, R., Eeken, J., 1990. *hobo* elements and their deletion-derivative sequences in *Drosophila melanogaster* and its sibling species *D. simulans*, *D. mauritiana* and *D. sechellia*. *Genet. Sel. Evol.* 22, 393–402.
- Periquet, G., Lemeunier, F., Bigot, Y., Hamelin, M. H., Bazin, C., Ladevèze, V., Eeken, J., Galindo, M. I., Pascual, L., Boussy, I., 1994. The evolutionary genetics of the *hobo* transposable element in the *Drosophila melanogaster* complex. *Genetica* 93, 79–90.
- Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T., Gehring, W.J., 1990. Analysis of the *ftz* upstream element: germ layer-specific enhancers are independently autoregulated. *Genes Dev.* 4, 1224 – 1239.
- Qian, S., Capovilla, M., Pirrotta, V., 1991. The *bx* region enhancer, a distant *cis*-control element of the *Drosophila Ubx* gene and its regulation by hunchback and other segmentation genes. *EMBO J.* 10, 1415 – 1425.
- Ruiz, M.T., Carareto, C.M., 2003. Characterization of *hobo* element copy number and integrity in Brazilian populations of *Drosophila melanogaster*. *Hereditas* 138, 154 – 157.
- Sarkar, S., Lakhota, S.C., 2005. The *Hsp60C* gene in the 25F cytogenetic region in *Drosophila melanogaster* is essential for tracheal development and fertility. *J. Genet.* 84, 265–281.
- Sassi, A.K., Herédia, F.O., 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, 734 – 739.

Shpiz, S., Kwon, D., Rozovsky, Y., Kalmykova, A., 2009. rasiRNA pathway controls antisense expression of *Drosophila* telomeric retrotransposons in the nucleus. *Nucleic Acids Res.* 37, 268–278.

Simmons, G.M., 1992. Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Mol. Biol. Evol.* 9, 1050–1060.

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*. *Mol. Biol. Evol.* 20, 2055 – 2066.

Tautz, D., Pfeifle, C., 1989. A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma* 98, 81–85.

Thornburg, B.G., Gotea, V., Makayowski, W., 2006. Transposable elements as a significant source of transcription regulating signals. *Gene* 365, 104–110.

Torres, F.P., Fonte, L.F.M., Valente, V.L.S., Loreto, E.L.S., 2006. Mobilization of a *hobo*-related sequence in the genome of *Drosophila simulans*. *Genetica* 126, 101–110.

Treisman, J., Desplan, C., 1989. The products of the *Drosophila* gap genes *hunchback* and *Krüppel* bind to the *hunchback* promoters. *Nature* 341, 335 – 337.

Vieira, C., Lepetit, D., Dumont, S., Biémont, C., 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol.* 16, 1251 – 1255.

Wakimoto, B.T., Kaufman, T.C., 1981. Analysis of larval segmentation in lethal genotypes associated with the antennapedia gene complex in *Drosophila melanogaster*. *Dev Biol.* 81, 51 – 64.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973 – 982.

Wray, G.A., 2003. Transcriptional regulation and the evolution of development. *Int. J. Dev. Biol.* 47, 675 – 684.

Yannopoulos, G., Stamatis, N., Monastrioti, M., Hatzopoulos, P., Louis, C., 1987. *Hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell* 49, 487–495.

CAPÍTULO 4

III - DISCUSSÃO

Os elementos de transposição (TEs) são ubíquos e abundantes, presentes nos mais diversos genomas analisados até agora. Desta forma, estudos envolvendo essas sequências podem ser de grande importância para a ciência. Neste trabalho nós analisamos alguns elementos da superfamília *hAT* de transposons, com a intenção de contribuir para o entendimento da evolução desses TEs, bem como das suas relações com os genomas hospedeiros.

Dentro da superfamília *hAT*, que possuem membros amplamente distribuídos ao longo dos genomas (Kempken e Windhofer, 2001), nós estudamos duas famílias de transposons: a família *hosimary* e a família *hobo*.

FAMÍLIA *HOSIMARY*

A descrição desta nova família de TEs surgiu da análise de sequências originalmente descritas por nosso grupo de pesquisa (Ortiz e Loreto, 2009). A partir de *hosim* e *hosec* foi caracterizada a família *hosimary*, com representantes presentes no subgrupo *melanogaster* de *Drosophila* e em *Zaprionus indianus*, apresentando mais de 90% de similaridade nucleotídica entre si. Nestas espécies o número de cópias do elemento mostrou-se variável: apenas uma cópia foi observada no genoma de *D. melanogaster* enquanto aproximadamente sete cópias foram observadas em *D. simulans*. Apesar de algumas sequências, como as de *D. melanogaster*, apresentarem mutações sem sentido, a maioria das sequências representam cópias completas, indicando um elemento potencialmente ativo e capaz de se mobilizar. O baixo número de cópias observado, juntamente com o fato de estarem completas pode ser indicativo do quão recente este elemento encontra-se no genoma.

Analisando uma das espécies, através da técnica de FISH, observamos que a maioria das cópias de *hosimary* está inserida em regiões heterocromáticas, com poucas cópias nas regiões da eucromatina. Utilizando essa mesma técnica de FISH Pimpinelli *et al.* (1995) analisaram a distribuição de diversos TEs e observaram que a maioria dos elementos são

preferencialmente agrupados em uma ou mais regiões discretas heterocromáticas nos cromossomos de Oregon R. Eles demonstraram ainda que a localização desses grupos de TEs são altamente conservadas, mesmo em linhagens geograficamente distantes. Por outro lado, Vieira *et al.* (1999) observaram uma grande variação no número de sítios de inserção entre populações de *D. simulans*. Os nossos resultados de Southern blot demonstram que o padrão de inserção de *hosimary* não parece ser conservado, ao menos entre as duas linhagens de *D. simulans* analisadas.

Adicionalmente, apesar das cópias localizadas na heterocromatina serem frequentemente incompletas, elas podem ser *trans* mobilizadas e algumas delas podem estar intactas e completas e podem ser reativadas (Biémont e Vieira, 2005). A presença de cópias potencialmente ativas de *hosimary* no genoma de *D. simulans* sugere que as inserções observadas na heterocromatina são um evento recente.

Alguns autores sugerem que a atividade transcricional pode favorecer a inserção de transposons. A transcrição é associada com a descondensação da cromatina, o que permite a acessibilidade do mecanismo de transcrição ao DNA, e também das enzimas envolvidas na transposição. Fontanillas *et al.* (2007) observaram que a inserção de transposons na eucromatina é moldada pela interação da expressão gênica e organização do genoma. No genoma de *D. melanogaster* os autores verificaram que a taxa de inserção de transposons é maior próximo a genes que são expressos na linhagem germinativa que serão transmitidos para a próxima geração. Entretanto, esse efeito é parcialmente compensado devido à forte seleção contra essas inserções. Assim, as regiões heterocromáticas que apresentam reduzida atividade transcricional, parecem funcionar como um “depósito” de transposons, que podem ser extintos, reativados, domesticados ou mesmo atuar no mecanismo de silenciamento de cópias ativas.

Na tentativa de estabelecer as relações evolutivas dos elementos da família *hosimary* inconsistências na reconstrução filogenética foram observadas. Essas inconsistências sugeriram que a aquisição da família *hosimary* nos genomas pode ter sido através de transferência horizontal (HT).

Os principais argumentos em favor da HT dos elementos entre espécies baseiam-se no cumprimento de três requisitos classicamente usados para inferências de casos de transmissão

horizontal, como descrito anteriormente. Dentre as hipóteses alternativas à HT todas elas assumem a existência de um ancestral comum entre as espécies envolvidas (Capy *et al.*, 1998).

Apesar de todas as premissas para a inferência de HT terem sido observadas, alguns aspectos merecem consideração. No que diz respeito à distribuição descontínua do elemento entre as espécies hospedeiras eventos de perda estocástica poderiam estar envolvidos. No entanto, se o elemento tivesse sido transferido verticalmente, muitos casos de perda dos elementos ao longo dos ramos filogenéticos teriam que ser considerados, uma vez que o tempo de divergência entre o grupo *melanogaster* e o gênero *Zaprionus* é de aproximadamente 55 milhões de anos (Robe, 2008). No grupo *obscura*, grupo irmão ao *melanogaster*, não foi observada amplificação de fragmentos da família *hosimary*. Assim, é improvável que esses TEs estivessem presentes no ancestral comum aos dois *clados* de espécies. Por outro lado, apesar dos elementos presentes nas espécies terem sido confirmados através de Southern blot, a presença em outras espécies não pode ser completamente descartada, uma vez que pode ser resultado da falta de amplificação com os *primers* usados. Uma alternativa para confirmar a distribuição dos elementos seria a utilização de outros *primers* ou Dot blot com todas as espécies analisadas neste trabalho.

Outro aspecto a ser considerado é a falta de resolução nos ramos da árvore filogenética. Entre *D. simulans*, *D. sechellia* e *D. mauritiana* não há resolução nos ramos que separam as espécies. Esta falta de resolução parece representar a alta similaridade entre as sequências, porém, pode estar relacionada à falta de sítios informativos nesta região. Por outro lado, a grande similaridade entre estas sequências pode ser atribuída a uma forte pressão seletiva ao invés de eventos de transmissão horizontal.

A hipótese de seleção mantendo a similaridade das sequências de *hosimary* foi considerada através da análise de divergência apenas em nível de substituições sinônimas, que, na ausência de um alto índice de uso preferencial de códons, fornece uma medida de evolução neutra. Por nossas análises pode-se descartar a hipótese de que a alta similaridade encontrada entre as sequências de *hosimary* seja efeito de alta pressão seletiva, ao menos em nível de códons. Esta afirmação justifica-se pois, além de utilizarmos apenas as substituições sinônimas em nossa análise, demonstramos previamente que o uso preferencial de códons do elemento é menor que o dos genes nucleares com os quais o mesmo foi comparado.

Os valores de dS estão inversamente relacionados ao uso preferencial de códons. Assim, como as análises de CBI e N_c mostraram menor tendenciosidade no uso de códons para *hosimary* do que para genes do hospedeiro, o esperado seria o dS do elemento ser maior do que o dos genes nucleares. Entretanto, não foi o que observamos. Os valores de dS de *hosimary* foram significativamente menores que os dos genes das espécies hospedeiras para a maioria dos casos comparados, indicando que as sequências de *hosimary* devem possuir um ancestral comum mais recente do que os genes das espécies hospedeiras.

Explicações alternativas para a alta similaridade observada entre as sequências de *hosimary* ainda incluem pressão seletiva em nível de RNA. Uma vez que o mecanismo de interferência por RNA (iRNA) tem sido proposto como um dos principais reguladores da atividade de transposons (Girard e Hannon, 2008), as sequências de *hosimary* poderiam estar sendo mantidas conservadas por gerarem mRNAs alvo para a degradação através do reconhecimento de pequenos RNAs homólogos a essas sequências. Porém, esta hipótese é pouco provável, pois além de alguns desses pequenos RNAs serem geralmente gerados por sequências defectivas presentes nos genomas, estas sequências são muito pequenas e não haveria necessidade de manter um elemento inteiro com tal similaridade. Ainda, uma grande pressão seletiva sofrida por estas sequências, por estarem envolvidas em algum tipo de domesticação, poderia explicar a grande conservação observada. Apesar de todas essas alternativas serem possíveis, são eventos pouco prováveis, e a aquisição dessas sequências através de transferência horizontal nos parece ser a hipótese mais parcimoniosa.

Transferência horizontal tem sido um mecanismo proposto para os TEs escaparem do processo de extinção dentro dos genomas, e parece ser um evento mais comum do que se acreditava (Bartolomé *et al.*, 2009). Muitos eventos de HT já foram descritos, inclusive para elementos da superfamília *hAT*, como sugerido para o elemento *herve*, que parece ter sido transmitido de um doador desconhecido para o genoma de *Anopheles gambiae* (Subramanian *et al.*, 2007), e para o elemento *hobo* (Daniels *et al.*, 1990). Para muitos transposons a origem através desse mecanismo parece ser a única explicação para justificar a história evolutiva do elemento.

Desta forma as sequências de *hosimary* aqui analisadas podem ser um exemplo de TEs adquiridos muito recentemente pelo genoma destas espécies hospedeiras e a transmissão horizontal pode ser a provável geradora deste cenário.

FAMÍLIA *HOBO*

O elemento *hobo* canônico (HFL1) é potencialmente ativo, apontado como causador de disgenesia do híbrido com altas taxas de mutabilidade e esterilidade gonadal (Yannopoulos *et al.*, 1987; Blackman *et al.*, 1987). Já o elemento *hobo*^{VA} (ou *hobo*^{VAHS}), embora não possua capacidade para codificar uma transposase funcional, foi apontado como gerador da mutação *white* em *D. simulans*. As características desses elementos podem ter profundo impacto nos genomas hospedeiros, como descrito na introdução desta tese. Dessa forma, mecanismos de silenciamento de TEs podem ser essenciais para a manutenção da integridade dos genomas.

Analisando a expressão dos elementos da família *hobo*, transcritos antisenso foram detectados, sugerindo que o mecanismo de silenciamento mediado por RNA controla a atividade desses TEs. A interferência por RNA parece ser um dos mecanismos principais para regular transposons e surgiu como uma resposta adaptativa contra a invasão de viroses (Waterhouse *et al.*, 2001). Neste mecanismo, os transcritos antisenso de *hobo* podem estar guiando a clivagem de transcritos senso através de enzimas da família Argonauta. Outra alternativa é que essas sequências antisenso, que reconhecem sequências senso homólogas, estejam atuando em nível pré-transcricional, através de modificações na cromatina que impedem a transcrição dos TEs alvo. Um mecanismo adicional de silenciamento do elemento *hobo* foi descrito por Handler e Gomes (1995, 1996), o qual sugere a possibilidade de uma interação negativa entre *hobo* e elementos relacionados a ele, o que poderia impedir sua mobilização.

O padrão de expressão de *hobo* e *hobo*^{VAHS} durante a embriogênese de *Drosophila* foi observado através de hibridização *in situ* de embriões inteiros. As diferentes linhagens analisadas apresentam um padrão de expressão transcricional similar, porém um dos padrões observados só foi detectado na linhagem *white* de *D. simulans*. A linhagem Zola, usada mais como um controle negativo da técnica de hibridização, apresentou sinais de hibridização apenas nos estágios iniciais de desenvolvimento, refletindo provavelmente a ação de cópias de

elementos relacionados a *hobo*. Dentre as outras linhagens estudadas foi detectada expressão transcricional essencialmente no sistema nervoso central e na região de formação da traquéia. Padrões similares já foram reportados para outros elementos, principalmente no sistema nervoso central (Ding e Lipshitz, 1994; O’Kane e Gehring, 1987). O padrão único sugestivo dos genes de segmentação de *Drosophila* foi observado apenas em *D. simulans* white.

Os trabalhos que descrevem padrões de expressão transcricional de TEs em embriões de *Drosophila* normalmente sugerem que a observação dos mesmos padrões em diferentes espécies ou linhagens é uma sugestão de que sequências intrínsecas aos elementos estão regulando a transcrição. Isso é observado na maioria dos casos descritos na literatura. No trabalho de Ding e Lipshitz (1994), dos 15 retroelementos investigados, apenas um deles apresentou padrão de expressão distinto entre as quatro linhagens analisadas. Isso sugere uma provável interação com outras sequências, regulando a atividade transcricional do TE relacionada à posição de inserção dentro do genoma do hospedeiro. Por outro lado, analisando o padrão de expressão do elemento *P* fusionado a lacZ, O’Kane e Gehring (1987) sugeriram que a expressão induzida pelo promotor do elemento *P* é muito dependente das sequências adjacentes ao sítio de inserção. Em análise similar com o fator *I* Lachaume *et al.* (1992) observaram que o efeito de posição pode afetar o nível de expressão deste elemento, mas não modifica a base do padrão de expressão durante o desenvolvimento de *Drosophila*.

Dessa forma, sugerimos que o padrão de expressão transcricional observado entre as linhagens pode ser dependente de sequências *cis*-reguladoras presentes em *hobo* e *hobo*^{VAHS}, que por sua vez podem ser dependentes de interações com genes do hospedeiro. O trabalho de Parkhurst e Corces (1987) revela a existência de relações evolutivas entre diferentes TEs em *Drosophila* e o envolvimento de diferentes genes celulares no controle da expressão destes elementos.

Como foi descrito anteriormente, existem diversas evidências de que os TEs podem servir como uma fonte de novas sequências regulatórias através da evolução. Uma explicação para isso é que a “simples” presença dos TEs nos genomas serviria como uma nova fonte de sequências, através das quais os elementos *cis*-reguladores poderiam emergir *de novo*, por meio de mutações pontuais (Feschotte, 2008). Outra alternativa é que elementos *cis*-reguladores existentes dentro dos TEs estariam diretamente interagindo com genes do

hospedeiro. Uma grande variedade de elementos regulatórios tem sido identificada em TEs ativos. Normalmente são sinais que os TEs usam para controlar sua própria expressão (como promotores basais e sinais de poliadenilação) e uma grande quantidade de sítios de ligação para fatores de transcrição (TFBSs), que parecem ser super representados em TEs (Thornburg *et al.*, 2006).

Mudanças em regiões *cis*-reguladoras com efeitos em seus hospedeiros vêm sendo descritas. Wittkopp *et al.* (2002) sugeriram que as diferenças na pigmentação abdominal de *D. melanogaster*, *D. subobscura* e *D. virilis* são parcialmente devidas às mutações em regiões *cis*-reguladoras do gene *yellow*. Ainda em relação ao gene *yellow*, mudanças dentro de porções *cis*-reguladoras específicas deste gene parecem relacionadas à evolução da pigmentação das asas no gênero *Drosophila* (revisão em Wray, 2007).

Para tentar entender as possíveis relações entre o padrão de expressão observado e os elementos *hobo* e *hobo*^{VAHS} foram realizadas buscas *in silico* por possíveis sítios de ligação para fatores de transcrição nas sequências desses TEs. Nesta análise diversos prováveis TFBSs foram observados em ambas as sequências. As implicações da presença desses TFBSs nesses elementos podem ser muitas. Os TEs podem alterar a expressão de genes vizinhos, gerando variabilidade, ou mesmo interagir com genes que são “ligados” por esses fatores de transcrição. A linhagem hipermutável de *D. simulans* descrita por Loreto *et al.* (1998) pode ser um bom exemplo da implicação desses elementos no genoma hospedeiro.

Buscas *in silico* por possíveis sítios para fatores de transcrição não representam a melhor ferramenta para identificar essas regiões, mas podem fornecer alguns indicativos das relações existentes entre a sequência dos TEs e o padrão de transcrição observado nos embriões. Um dos fatores de transcrição com maior número de sítios detectados em ambos elementos foi Zeste. Isso é bastante interessante, uma vez que este fator de transcrição parece interagir com o locus *white* e também com o complexo *bithorax* (Mansukhani *et al.*, 1988), e mutantes homeóticos foram observados na linhagem hipermutável anteriormente descrita. Além de Zeste, outros fatores de transcrição são sugestivos, tais como LEF-1 (*lymphocyte enhancer binding factor 1*), que parece regular a expressão do gene homeótico *Ultrabithorax* em *Drosophila* (Riese *et al.*, 1997), e *fushi tarazu*, que poderia explicar o padrão em 7 faixas observadas nos embriões de *D. simulans white*. Apesar de serem muito sugestivos a real

interação destes fatores de transcrição com *hobo* e *hobo*^{VAHS} pode ser confirmada em trabalhos futuros, provavelmente através da construção de uma linhagem transgênica.

PERSPECTIVAS

Como forma de continuação deste trabalho, delineamos algumas perspectivas e análises a serem empregadas com a finalidade de investigar os aspectos que não foram totalmente elucidados nesta tese.

- Com relação ao trabalho intitulado “*hosimary: a new transposon family of Zaprionus indianus and Drosophila melanogaster*”, algumas análises ainda não foram completamente finalizadas. Pretendemos confirmar a distribuição de membros da família *hosimary* através de Dot Blot. Adicionalmente, considerando que além do subgrupo *melanogaster* de *Drosophila* este elemento foi detectado somente no gênero *Zaprionus*, gostaríamos de ampliar o número de espécies deste gênero em nossas análises.

Além disso, pretendemos concluir outros dois trabalhos não apresentados nesta tese mas realizados durante o trabalho de doutorado:

- Um deles refere-se à análise do elemento *Mar*, um MITE da superfamília *hAT* de transposons descrito por Holyoake e Kidwell (2003). Um resumo deste trabalho, com os principais resultados é apresentado no Anexo 1. Este trabalho tem a colaboração da MSc. Adriana Ludwig. Para a finalização deste trabalho realizaremos a clonagem e análise do elemento em uma espécie adicional.

- Outro trabalho envolve a quantificação do número de cópias dos elementos *hobo* e *hobo*^{VA} através de hibridização *in situ* de cromossomos politênicos (FISH) e PCR quantitativo (qPCR) em linhagens de *D. simulans*. Este último trabalho tem a colaboração do Dr. Rogério Margis da UFRGS e da Dra. Yanina Panzera da Universidad de la República, Montevideo, Uruguay. Os resultados obtidos até o momento sugerem a presença de 6 cópias de *hobo* no genoma de *Drosophila simulans* white detectado através de FISH e uma centena de sinais de *hobo*^{VA}. Esses dados foram confirmados por qPCR apenas para o elemento *hobo* canônico. Adicionalmente, outras três linhagens de *Drosophila simulans* foram utilizadas para quantificação de *hobo* canônico através de qPCR. Para a finalização deste trabalho análises de FISH deverão ser empregadas em uma linhagem adicional de *D. simulans*, e a quantificação através de PCR quantitativo do elemento *hobo*^{VA} deverá ser concluída.

REFERÊNCIAS BIBLIOGRÁFICAS.

Almeida LM e Carareto CMA (2005a) Origem, proliferação e extinção de elementos transponíveis: qual seria a importância horizontal na manutenção desse ciclo? - Monografia SBG, Sociedade Brasileira de Genética, 1 ed, 43pp.

Almeida LM e Carareto CMA (2005b) Multiple events of horizontal transfer of the *Minos* transposable element between *Drosophila* species. *Mol Phylogenet Evol* 35:583-594.

Aravin AA, Hannon GJ e Brennecke J (2007) The Piwi piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318:761-764.

Arca B e Savakis C (2000) Distribution of the transposable element *Minos* in the genus *Drosophila*. *Genetica* 108:263-267.

Atkinson PW, Warren WD e O'Brochta DA (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilized in house-flies and excises like the *Ac* element of maize. *Proc Natl Acad Sci USA* 90:9693-9697.

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism and function. *Cell* 116:281-297.

Bartolomé C, Bello X e Maside (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol* 10(2):R22.

Berman BP, Pfeiffer BD, Lavery TR, Salzberg SL, Rubin GM, Eisen MB e Celniker SE (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol* 5:R61.

Biémont C e Vieira C (2005) What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. *Cytogenet Genome Res* 110:25-34.

Biémont C e Vieira C (2006) Junk DNA as an evolutionary force. *Nat Genet* 443:521-524.

Blackman RK, Grimaila R, Koehler MMD e Gelbart WM (1987) Mobilization of *hobo* elements residing within the *decapentaplegic* gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell* 49:497-505.

Blackman RK, Koehler MM, Grimaila R e Gelbart WM (1989) Identification of a fully functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J* 8:211-217.

- Borsatti F, Azzoni P e Mandrioli M (2003) Identification of a new *hobo* element in the cabbage moth, *Mamestra brassicae* (Lepidoptera). *Hereditas* 139:151-155.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R e Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089-1103.
- Birchler JA e Kavi HH (2008) Slicing and dicing for small RNAs. *Science* 320:1023-1024.
- Brönner G, Taubert H e Jäckle H (1995) Mesoderm-specific B104 expression in the *Drosophila* embryo is mediated by internal *cis*-acting elements of the transposon. *Chromosoma* 103:669-675.
- Brookfield JF (2005) The ecology of the genome – mobile DNA elements and their hosts. *Nat Rev Genet* 6:128-136.
- Brunet F, Godin F, Bazin C e Capy P (1999) Phylogenetic analysis of *Mos1*-like transposable elements in the Drosophilidae. *J Mol Evol* 49:760-768.
- Bundock P e Hooykaas P (2005) An *Arabidopsis* *hAT*-like transposase is essential for plant development. *Nature* 436:282-284.
- Calvi BR, Hong TJ, Findley SD e 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, Bazin C, Higuete D e Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197pp.
- Chambeyron S, Popkova A, Payen-Groschêne G, Brun C, Laouini D, Pelisson A e Bucheton A (2008) piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci USA* 105:14964-14969.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218.
- Corces VG e Geyer PK (1991) Interactions of retrotransposons with the host genome: the case of the *gypsy* element of *Drosophila*. *Trends Genet* 7: 86 – 90.
- Craig NL, Graigie R, Gellert M e Lambowitz AM (2002) Mobile DNA II. American Society for Microbiology Press, Washington, DC, 1204pp.
- Daniels SB, Chovnick A e Boussy IA (1990) Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol Biol Evol* 7:589-606.

De Frutos R, Peterson KR e Kidwell MG (1992) Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. *Chromosoma* 101:293-300.

DeVault JD e Narang SK (1994) Transposable elements in Lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem Biophys Res Commun* 203:169-175.

Ding D e Lipshitz HD (1994) Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. *Genet Res* 64:167-181.

Ferris PJ (1989) Characterization of a *Chlamydomonas* transposon, *Gulliver*, resembling those in higher plants. *Genetics* 122:363-377.

Feshotte C e Pritham EJ (2007) DNA Transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331-368.

Feshotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9:397-405.

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5:103-107.

Fontanillas P, Hartl DL e Reuter M (2007) Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *PLoS Genetics* 3:2256-2267.

Frommer G, Schuh R e Jäckle H (1994) Localized expression of a novel micropia-like element in the blastoderm of *Drosophila melanogaster* is dependent on the anterior morphogen *bicoid*. *Chromosoma* 103:82-89.

Galindo MI, Ladeveze V, Lemeunier F, Kalmes R, Pèriquet G e Pascual L (1995) Spread of the autonomous transposable element *hobo* in the genome of *Drosophila melanogaster*. *Mol Biol Evol* 12:723-734.

Galindo MI, Bigot Y, Sánchez MD, Periquet G e Pascual L (2001) Sequences homologous to the *hobo* transposable element in E strains of *Drosophila melanogaster*. *Mol Biol Evol* 18:1532-1539.

Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD e Jurka J (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. *Genome Res* 17:992-1004.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320:1077-1081.

Girard A e Hannon GJ (2008) Conserved themes in small-RNA mediated transposon control. Trends Cell Biol 18:136-148.

Handler AM e Gomez SP (1995) The *hobo* transposable elements has transposase-dependent and -independent excision activity in drosophilid species. Mol Gen Genet 247:399-408.

Handler AM e Gomez SP (1996) The *hobo* transposable element excises and has related elements in tephritid species. Genetics 143:1399-1947.

Hehl R, Nacken WK, Krause A, Saedler H e Sommer H (1991) Structural analysis of *Tam3*, a transposable element from *Antirrhinum majus*, reveals homologies to the *Ac* element from maize. Plant Mol Biol 16:369-371.

Herédia F, Loreto ELS e Valente VLS (2004) Complex evolution of *gypsy* in Drosophilid species. Mol Biol Evol 21:1831-1842.

Holyoake A e Kidwell MG (2003) *Vege* and *Mar*: two novel *hAT* MITE families from *Drosophila willistoni*. Mol Biol Evol 20:163-167.

Jordan IK, Rogozin IB, Glazko GV e Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68-72.

Kempken F e Windhofer F (2001) The *hAT* family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma 110:1-9.

Kidwell MG e Lisch DR (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94:7704-7711.

Kidwell MG e Lisch DR (2001) Perspectiva: transposable elements, parasitic DNA and genome evolution. Evolution 55:1-24.

Kotnova AP, Karpova NN, Feoktistova MA, Liubomirskaia NV, Kim AI e Ilyin IuV (2005) Retrotransposon *gtwin*: structural analysis and distribution in *Drosophila melanogaster* strains. Genetica 41:23-29.

Lachaume P, Bouhidel K, Mesure M e Pinon H (1992) Spatial and temporal expression of the *I* factor during oogenesis in *Drosophila melanogaster*. Development 115: 729 – 735.

Lim JK (1988) Intracromossomal rearrangements mediated by *hobo* transposons in *Drosophila melanogaster*. Proc Natl Acad Sci USA 85:9153-9157.

Loreto ELS, Zaha A, Nichols C, Pollock JA e Valente VLS (1998) Characterization of a hypermutable strain of *Drosophila simulans*. Cell Mol Life Sci 54:1283-1290.

Loreto ELS, Carareto CMA e Capy P (2008) Revisiting horizontal transfer of transposable element in *Drosophila*. *Heredity* 100:545-554.

Lowe CB, Bejerano G e Haussler D (2007) Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci USA* 104:8005-8010.

Ludwig A e Loreto ELS (2007) Evolutionary pattern of the *gtwin* retrotransposon in the *Drosophila melanogaster* subgroup. *Genetica* 130:161-168.

Ludwig A, Valente VLS e Loreto ELS (2008). Multiple invasions of *Errantivirus* in the genus *Drosophila*. *Insect Mol Biol* 17:113-124.

Mansukhani A, Crickmore A, Sherwood PW e Goldberg ML (1988) DNA-binding properties of the *Drosophila melanogaster* *zeste* gene product. *Mol Cell Biol* 8:615-623.

Maruyama K e Hartl DL (1991) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J Mol Evol* 33:514-524.

McClintock B (1947) Cytogenetic studies of maize and *Neurospora*. *Carnegie Inst Washington Year Book* 46:146-152.

McGinnis W, Shermoen AW e Beckendorf SK (1983) A transposable element insert just 5' to *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* 34:75-84.

Mevel-Ninio M, Pelisson A, Kinder J, Campos AR e Bucheton A (2007) The *flamenco* locus controls the *gypsy* and *ZAM* retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175:1615-1624.

Mozer BA e Benzer S (1994) Ingrowth by photoreceptor axons induces transcription of a retrotransposon in the developing *Drosophila* brain. *Development* 120:1049-1058.

O'Kane CJ e Gehring WJ (1987) Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci USA* 84:9123-9127.

Ortiz MF e Loreto ELS (2008) The *hobo*-related elements in the *melanogaster* species group. *Genet Res* 90:243-252.

Ortiz MF e Loreto EL (2009) Characterization of new *hAT* transposable elements in twelve *Drosophila* genomes. *Genetica* 135:67-75.

Parkhurstl SM e Corces VG (1987) Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. *EMBO J* 6:419-424.

- Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A e Corces VG (1994) *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissuespecific control of the *Drosophila flamenco* gene. EMBO J 13:4401-4411.
- Periquet G, Lemeunier F, Bigot Y, Hamelin MH, Bazin C, Ladevèze V, Eeken J, Galindo MI, Pascual L e Boussy I (1994) The evolutionary genetics of the *hobo* transposable element in the *Drosophila melanogaster* complex. Genetica 93:79-90.
- Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C e Gatti M (1995) Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc Natl Acad Sci USA 92:3804-3808.
- Pinsker W, Haring E, Hagemann S e Miller WJ (2001) The evolutionary life history of *P* transposons: from horizontal invaders to domesticated neogenes. Chromosoma 110:148-158.
- Reiss D, Josse T, Anxolabéhère D e Ronsseray S (2004) *aubergine* mutations in *Drosophila melanogaster* impair *P* cytotype determination by telomeric *P* elements inserted in heterochromatin. Mol Genet Genomics 272:336-343.
- Riese J, Yu X, Munnerlyn A, Eresh S, Hsu SC, Grosschedl R e Bienz M (1997) LEF-1, a nuclear factor coordinating signaling inputs from *wingless* and *decapentaplegic*. Cell 88:777-787.
- Robe LJ (2008) Relações filogenéticas no gênero *Drosophila* (Diptera, Drosophilidae): uma abordagem molecular. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. pp 303.
- Rubin GM, Kidwell MG e Bingham PM (1982) The molecular basis of *Drosophila melanogaster* hybrid dysgenesis: The nature of induced mutations. Cell 29:987-994.
- Silva JC, Loreto ELS e Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol 6:57-71.
- Simmons G (1992) Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. Mol Biol Evol 9:1050-1060.
- Simmons GM, Plummer D, Simon A, Boussy IA, Frantsve J e Itoh M (1998) Horizontal and vertical transmission of *hobo*-related sequences between *Drosophila melanogaster* and *Drosophila simulans*, pp. 285-294 in Horizontal Gene Transfer, edited by M. Syvanen & C.I. Kado. Chapman & Hall, New York.
- Slotkin RK e Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8:272-285.

Streck RD, MacGaffey JE e Beckendorf SK (1986) The structure of *hobo* transposable elements and their insertion sites. *EMBO J* 5:3615-3623.

Stuart JR, Haley KJ, Swedzinski D, Lockner S, Kocian PE, Merriman PJ *et al.* (2002). Telomeric *P* elements associated with cytotype regulation of the *P* transposon family in *Drosophila melanogaster*. *Genetics* 162:1641-1654.

Subramanian RA, Arensburger P, Atkinson P e O'Brochta DA (2007) Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector *Anopheles gambiae* s.s. *Genetics* 176:2477-2487.

Thornburg BG, Gotea V e Makayowski W (2006) Transposable elements as a significant source of transcription regulating signals. *Gene* 365:104-110.

Torres FP, Fonte LFM, Valente VLS e Loreto ELS (2006) Mobilization of a *hobo*-related sequence in the genome of *Drosophila simulans*. *Genetica* 126:101-110.

Torti C, Gomulski LM, Bonizzoni M, Murelli V, Moralli D, Guglielmino CR, Raimondi E, Crisafulli D, Capy P, Gasperi G e Malacrida AR (2005) *Cchobo*, a *hobo*-related sequence in *Ceratitis capitata*. *Genetica* 123:313-325.

Vieira C, Lepetit D, Dumont S e Biémont C (1999) Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol* 16:1251-1255.

Waterhouse PM, Wang MB e Lough T (2001) Gene silencing as an adaptive defence against viruses. *Nature* 411:834-842.

Wicker T, Sabot F, Hua-Van A, Bennetzen J L, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P e Schulman AH (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 8:973-982.

Wittkopp PJ, Vaccaro K e Carroll SB (2002) Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr Biol* 12:1547-1556.

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206-216.

Yannopoulos G, Stamatis N, Monastrioti M, Hatzopoulos P e Louis C (1987) *hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF. *Cell* 49:487-495.

ANEXO 1.

Estudo do elemento *Mar*, um MITE, no gênero *Drosophila*.

Maríndia Deprá^{1*}, Adriana Ludwig^{1*}, Vera L. S. Valente^{1,3} e Elgion L. S. Loreto^{1,2}

1- PPGBM – UFRGS; 2-Departamento de Biologia – UFSM; 3-Departamento de Genética – UFRGS

*- Esses autores contribuíram igualmente para a realização deste trabalho.

INTRODUÇÃO: MITEs (*Miniature inverted repeat transposable elements*) são elementos de transposição curtos, não-autônomos e frequentemente presentes em alto número de cópias nos genomas. Vários trabalhos têm sugerido que esses elementos originam-se a partir de transposons autônomos que sofreram deleções internas, mas mantêm as repetições terminais invertidas (TIRs) conservadas, possibilitando mobilização *in trans*. O elemento *Mar* de *Drosophila willistoni* possui 610 pb, sítio alvo de duplicação (TSDs) de 8 pb, TIRs perfeitas de 11 pb e não apresenta capacidade codificadora. Esse elemento foi classificado como um MITE da superfamília *hAT* com base na similaridade das TIRs e TSDs, e foi encontrado em apenas uma linhagem de *D. willistoni*, analisada em um trabalho anterior de outros autores.

RESULTADOS: Na tentativa de caracterizar a representatividade deste elemento e suas implicações evolutivas nos genomas de *Drosophila*, iniciamos um trabalho de análise da distribuição do elemento *Mar* em um grande número de espécies. Um total de 58 espécies de *Drosophila* de diferentes grupos dos dois principais subgêneros, *Sophophora* e *Drosophila*, juntamente com *Zaprionus indianus*, *Z. sepsoides*, *Scaptodrosophila latifasciaeormis* e *S. lebanonensis* foram testadas por PCR para a presença do elemento *Mar*. Obtivemos amplificação de fragmentos somente nas espécies do subgrupo *willistoni* investigadas: 3 linhagens de *D. willistoni*, *D. paulistorum*, *D. equinoxialis*, *D. tropicalis* e *D. insularis*. Os amplicons de cada espécie foram clonados e sequenciados para a realização de análises evolutivas do elemento. No entanto, ainda não foram obtidas as sequências dos clones de *D. tropicallis*. Adicionalmente às análises por PCR, realizamos buscas *in silico* por sequências homólogas ao *Mar*, nos 12 genomas de *Drosophila* atualmente disponíveis, através da ferramenta BLAST. Como esperado, o elemento *Mar* foi encontrado no genoma de *D.*

willistoni, porém, nenhuma cópia foi encontrada nos outros genomas. As cópias de *D. willistoni* foram analisadas quanto a conservação de TIRs e TSDs, além da caracterização da localização dessas cópias no genoma. Encontramos cerca de 80 cópias completas (TIRs intactas) deste elemento no genoma de *D. willistoni*, sendo que a maior parte dessas sequências também apresenta TSDs intactas indicando possível mobilização recente desse elemento. A divergência média entre as cópias é de 9 %, sendo essa alta divergência encontrada tanto entre como dentro espécies. Consistente com prévios trabalhos mostrando associação entre MITEs e genes, nós encontramos 19 cópias do elemento *Mar* localizados perto de genes e 16 cópias inseridas dentro de introns. Adicionalmente, realizamos diversas buscas *in silico* no genoma de *D. willistoni* para tentar localizar cópias do elemento *Mar* codificadoras de transposase. Encontramos apenas uma sequência de 2314 pb, que possui sequências homólogas ao elemento *Mar* nas extremidades 5' e 3' e uma região potencialmente codificadora de apenas 239 aminoácidos com baixa similaridade com transposase. Essa cópia precisa ser melhor analisada, pois sua sequência pode apresentar algum íntron, o que pode aumentar a região codificadora, aumento as chances de ser uma transposase funcional. Ainda, esta sequência pode ser apenas uma inserção de um transposon dentro do elemento *Mar*. Adicionalmente às análises *in silico*, uma busca por essa cópia nas outras linhagens de *D. willistoni* poderia ajudar a elucidar se essa cópia trata-se de um transposon codificador de transposase e potencialmente mobilizador do elemento *Mar*.

CONCLUSÕES: Com base nos dados obtidos até o momento podemos concluir que o elemento *Mar* está ainda ativo ou esteve ativo recentemente no genoma de *D. willistoni*. Ainda, pela comparação da filogenia obtida para o elemento *Mar* com a filogenia das espécies hospedeiras e a observação de alta divergência do elemento, podemos sugerir que a origem deste MITE ocorreu após a separação dos subgrupos *willistoni* e *bocainensis* (desde que espécies deste último subgrupo não possuem o elemento), porém antes do início da diversificação do subgrupo *willistoni* estimado em 5,6 Mya.