

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

***harrow* : Nova Família de Transposons de *Drosophila*
Envolvida em Transferência Horizontal**

Nina Roth Mota

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Mestre

Orientador: Élgion Lúcio da Silva Loreto

Co-Orientadora: Vera Lúcia da Silva Valente

Porto Alegre
Março de 2009

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

O mundo do trabalho está a exigir profissionais mais bem formados do que simplesmente informados, onde a capacidade de trabalhar em equipe e a preparação para educação permanente ao longo da vida estejam presentes. Enfim, habilidades e competências capazes de fazer com que o profissional não tema o novo e esteja preparado para desafios, sejam eles quais forem. Tais ingredientes serão definidores do sucesso ou insucesso das empreitadas.

(Ronaldo Mota)

Dedico este trabalho aos meus pais, pelo apoio,
pelo incentivo e pelo amor incondicional.

Agradecimentos

Ao Élgiom, meu orientador há seis anos, uma pessoa fantástica, que além de um grande pesquisador, é um exemplo de dedicação, de ética profissional e de responsabilidade com o ensino público. Meu sincero MUITO OBRIGADA por ter proporcionado meu primeiro contato com a pesquisa e a oportunidade de trabalhar na área que sempre adorei, pela orientação, pela grande amizade, por tantos momentos de descontração, pelo ombro amigo, pelos puxões de orelha, pela oportunidade de conhecer e trabalhar com pessoas maravilhosas, pela parceria, pela paciência e pelos ensinamentos, tanto profissionais quanto pessoais, que levarei comigo para o resto da vida.

À Lenira, por ser um exemplo de professora, por sua dedicação ao ensino, sempre inovando, e fazendo o processo de aprendizagem mais leve e o conteúdo aprendido, permanente. À Júlia, juntamente com a Lenira, por terem sempre um comentário inteligente na ponta da língua, na hora certa, que nos faz cair em gargalhadas.

À Verinha por me receber de braços abertos em seu laboratório, propiciar todos os auxílios para o desenvolvimento de um bom trabalho e estar sempre disposta a ajudar a todos.

À Adri, minha orientadora extra-oficial de mestrado, uma grande amiga, pelos incontáveis ensinamentos, pelos conselhos, por ter me acudido sempre que necessário (e não foram poucas vezes), pela parceria, por ter aceitado a mim e a Dharma como “moradoras transitórias” em sua casa, pelas maratonas gastronômicas, pelas risadas, por todo o tempo dedicado e, principalmente, pela maravilhosa amizade. Um grande exemplo, por estar em constante busca de seus sonhos, sem medo, sem titubear, enfrentando e aproveitando tudo que a vida tem para oferecer.

À Liz, um exemplo de dedicação ao estudo, que me “adotou” desde o meu primeiro semestre de faculdade e que entende bem a frase: somos eternamente responsáveis por aqueles que cativamos; pelas intermináveis conversas sobre a profissão, amigos, amores, responsabilidades, enfim, sobre a vida, e por ser uma fonte inesgotável de ensinamentos. Ao Sandro e ao Charles que me aceitaram tantos dias em sua casa.

À Maríndia, por estar sempre pronta a ajudar em qualquer situação, por me puxar as orelhas às vezes, pela amizade, pelo companheirismo, pelas conversas e tantos momentos de diversão que fizeram recarregar nossas forças.

À Paloma, pelo carinho, pelas conversas, pelas festas, pela parceria e por sempre estar com um sorrisão no rosto e conseguir me fazer rir mesmo nos dias mais difíceis. À Paloma e ao Ronaldo, pelas tantas ajudas técnicas que me deram e, principalmente, pela constante alegria e amizade.

Ao pessoal do LabDros da UFSM: Lizandra, Paloma, Ronaldo, Gabriel, Felipe, Ana, Francine, Felipe Ten Caten, Bruna, Francis, Sinara, Marcela, Málvaro

e Gabriela, por sempre me acolherem muito bem nas minhas idas e vindas à Santa Maria, e por manterem o LabDros um ambiente descontraído, aconchegante e, ainda assim, extremamente produtivo.

Ao pessoal do Laboratório de *Drosophila* da UFRGS: Adriana, Maríndia, Hermes, Marícia, Juliana C., Juliana W. G., Mário, Dirleane, Ana Carolina, Gisele, Cleverton, Gilberto, Ferreira e Gilson, por me receberem em seu laboratório, pelas conversas sérias, pelos “papos-furados” e pelos momentos de entretenimento que tornam o ambiente agradável para trabalhar. Ao Hermes, em especial, pela ajuda na identificação de várias mosquinhas e pelos momentos de risadas.

À Dirleane, por sempre me ajudar no laboratório, agüentar minhas implicâncias, pelas conversas amigas e saber que no fundo eu gosto muito dela.

À Paula Angélica, por aceitar eu e a Dharma como “visita permanente” em sua casa no último ano e proporcionar momentos de muita descontração.

Ao Elmo e à Ellen, pelo apoio, paciência, conversas, brincadeiras, e acima de tudo, pelo exemplo de profissionalismo e competência.

Ao PPGBM, por reforçar que é preciso ter orgulho de estar entre os melhores programas de pós-graduação do país e força de vontade e trabalho pra se manter em tal posição, pelo incentivo a todos os alunos e pela receptividade que fez eu me sentir “em casa” desde o início.

Ao meu Pai e minha Mãe, meus maiores exemplos de vida, minha grande fonte de inspiração, impossível colocar em palavras tamanho sentimento de gratidão, de admiração e de amor que eu sinto por vocês; pelo amor incondicional, pelo suporte, pelo apoio, por acreditarem sempre em mim e por terem permitido que eu fosse uma criança questionadora, que não aceitava a resposta mais simples nem o caminho mais fácil como o correto. Talvez na infância estejam os primórdios do meu desejo de ser uma pesquisadora, uma cientista. Agradeço também por agüentarem meus estresses e “faniquitos” no mestrado e, ainda assim, me incentivarem muito a fazer o doutorado.

Ao meu avô Ives e minha avó Marina, meu ninho seguro, pelo carinho, pelo incentivo à busca do meu sonho e por compreenderem que a vida tem épocas mais corridas, e não sobra tanto tempo quanto gostaríamos pra passar junto daqueles que amamos, mas que apesar da falta do convívio do dia-a-dia, os corações estão sempre batendo na mesma sintonia.

À minha dinda Liane, tio Renato, Dudu e Guga, por me acolherem em Porto Alegre, por estarem sempre presentes em minha vida, me dando força, amparo, motivação e momentos de diversão.

Aos tios, primos e agregados Weissheimer, por reforçarem os laços familiares e nunca deixar dúvida de que a base da força para enfrentar o mundo e conquistar nossos objetivos é a família.

Aos meus avós paternos, João e Alice, e à tia Sandra e família, que mesmo de longe, sempre me incentivaram e acreditaram em mim.

Às minhas irmãs-de-coração, Priscila e Geórgia, que me compreenderam, me apoiaram e se fizeram presentes mesmo que distantes fisicamente. Ao Cassiano, um grande amigo, que sempre me deu força nos momentos mais difíceis.

Às amigas de Santa Maria: Priscila, Juju, Nati, Lika, Dai, por compreenderem minha distância e ausência em tantas festas e encontros e torcerem pelas minhas conquistas.

À Dharma Beatriz, minha fiel companheira canina há nove anos, por aceitar o excesso de tempo na frente do computador e estar sempre disposta a brincar e me alegrar nos intervalos da elaboração desta dissertação.

Finalmente, a todos que, embora não estejam citados, de uma forma ou de outra, de perto ou de longe, fazem parte de minha vida e me incentivaram na busca desta conquista.

SUMÁRIO

RESUMO	9
ABSTRACT	10
CAPÍTULO I	11
Introdução	11
1) Elementos transponíveis	11
2) Transferência Horizontal X Transferência Vertical	13
3) Classificação dos TEs	14
4) Superfamília <i>hAT</i>	16
5) Gênero <i>Drosophila</i>	20
5.1) Subgênero <i>Drosophila</i>	20
5.2) Subgênero <i>Sophophora</i>	23
6) Sobre a escolha do tema	25
7) Objetivos	26
CAPÍTULO II	27
<i>harrow</i> : A new <i>Drosophila</i> family of <i>hAT</i> transposons involved in Horizontal Transfer	27
CAPÍTULO III	60
Discussão Geral	60
Referências Bibliográficas	66

RESUMO

Os elementos transponíveis são conhecidos por possuírem a fascinante propriedade de mudarem de local dentro do genoma de suas espécies hospedeiras e mesmo de serem capazes de atravessar barreiras inter-específicas, invadindo novos genomas. Neste estudo, uma nova família de elementos transponíveis, denominada *harrow*, é descrita para a Superfamília *hAT*. Buscas por seqüências *harrow* foram realizadas em 65 espécies de *Drosophilidae*, sendo que a maior parte dessas espécies pertence a grupos Neotropicais do gênero *Drosophila* ou é cosmopolita. As seqüências *harrow* encontradas apresentam distribuições descontínuas ao longo do gênero *Drosophila* e apresentam altas similaridades entre si, inconsistentes com os tempos de divergência das espécies hospedeiras. Além disso, incongruências topológicas foram encontradas entre as filogenias de *harrow* e das espécies hospedeiras. Os resultados obtidos no presente estudo indicam que *harrow* provavelmente invadiu os genomas de seus hospedeiros por múltiplos eventos de transferência horizontal, os quais devem ter ocorrido entre espécies dos Trópicos do Novo Mundo. Estes resultados reforçam a visão de que eventos de transferências horizontais entre eucariotos são mais comuns do que se previa anteriormente.

ABSTRACT

Transposable elements are known for their amazing properties of moving from one place to another inside the genome of their host species, as well as, for their ability to cross species boundaries, invading new genomes. In this study, a new family of transposable elements, *harrow*, is described for the *hAT* Superfamily of DNA transposons. Searches for *harrow* sequences were conducted in 65 *Drosophilidae* species, mainly representing Neotropical and cosmopolitan group species from genus *Drosophila*. The obtained *harrow* sequences show a patchy distribution along the genus *Drosophila* and present high sequence similarities, which are inconsistent with the divergence times of the host species. Additionally, topological incongruities were found between *harrow* and host species phylogenies. The results obtained in this study indicate that *harrow* has probably invaded its host genomes through multiple horizontal transfer events, which should have occurred in Neotropical regions. These results support the growing view that horizontal transfers between eukaryotes are more common than they have been thought before.

CAPÍTULO I

Introdução

1) Elementos transponíveis

Elementos Transponíveis (TEs, do Inglês “Transposable Elements”) são seqüências de DNA que possuem a propriedade de moverem-se de um local para outro dentro do genoma de seu hospedeiro ou mesmo invadirem novos genomas. Esta propriedade pode ser intrínseca, quando o próprio elemento codifica a enzima responsável por sua mobilização, sendo assim denominados elementos autônomos. Outros elementos, chamados não-autônomos, não codificam uma enzima funcional, mas podem vir a ser “trans-mobilizados” com o auxílio de enzimas codificadas por elementos autônomos relacionados.

TEs foram encontrados em virtualmente todos os organismos eucariotos analisados até hoje, perfazendo parte significativa do genoma da maioria das espécies. Por exemplo, estima-se que mais de 50% do genoma do milho (San Miguel *et al.*, 1996), 22% do genoma de *Drosophila melanogaster* (Kapitonov e Jurka, 2003) e 45% do genoma humano (“International Human Genome Sequencing Consortium”, 2001) são constituídos por TEs.

Por muito tempo, a fração do genoma que não possuía uma função conhecida foi chamada de “DNA lixo”, da qual os TEs faziam parte, e acreditava-se que esta região não exercia papel algum na evolução dos genomas. Em seguida, o termo “DNA egoísta” passou a ser usado para descrever a propriedade dos TEs de se multiplicarem no genoma hospedeiro. Finalmente estas denominações, que possuem certa conotação pejorativa, estão caindo no desuso, pois ignoram que, embora a maioria das mutações causadas por TEs (assim como por outros fatores mutagênicos) sejam neutras ou deletérias, eles podem também causar alterações que contribuam significativamente para a evolução dos seus genomas hospedeiros.

As mutações causadas pela mobilização de TEs podem variar desde modificações do tamanho e arranjo de genomas inteiros até substituições, deleções e inserções de um único nucleotídeo. O espectro de mutações induzidas por TEs parece ser maior do que o produzido por qualquer outro fator mutagênico, pois além de ocorrerem no momento de sua inserção, estas alterações podem ocorrer durante excisões não precisas ou por recombinações ectópicas de elementos, por exemplo (Kidwell e Lisch, 2001).

Como mencionado anteriormente, mutações causadas por TEs nem sempre são neutras ou deletérias. Lin *et al.* (1998) propõem que a inserção do elemento *P*, no terceiro íntron do gene *methuselah*, leva a um aumento de 35% do tempo de vida de mutantes homocigotos para essa inserção. Revelam ainda que estes mutantes são mais resistentes a estresses causados por agentes oxidantes (Paraquat), privação de alimento e temperaturas elevadas.

Além disso, TEs podem desempenhar uma função adicional importante na evolução dos organismos ao funcionarem como reservatórios de seqüências para inovações genéticas (Volff, 2006). Em interessante e detalhada revisão, Volff (2006) indica que seqüências derivadas de TEs foram recrutadas repetidas vezes por seus hospedeiros durante a evolução. Um exemplo deste tipo de domesticação de seqüências é apresentado por Muehlbauer *et al.* (2006). Os autores identificam um gene expresso, aparentemente funcional, chamado *gary*, no genoma de milho, cevada e trigo. Este gene, que está sofrendo pressão seletiva positiva, é provavelmente resultado do recrutamento de uma transposase da Superfamília *hAT*, no entanto sua função permanece desconhecida.

Com a publicação do genoma humano, um significativo número de genes aparentemente derivados de recrutamento de TEs foi encontrado. No artigo da publicação do genoma, 47 genes humanos são listados como sendo, provavelmente, originados de transposons ("International Human Genome Sequencing Consortium", 2001). No entanto, elementos transponíveis já foram também reportados como estando envolvidos em casos de algumas doenças humanas, como hemofilia (Kazazian, 1988), distrofia muscular de Duchenne (Narita *et al.*, 1993) e neurofibromatose do tipo I (Wallace *et al.*, 1991), entre outras. Acredita-se que a variação nos níveis de expressão de TEs no genoma humano, principalmente dos elementos *Line-1* e *Alu*, podem estar correlacionados

a variação individual da susceptibilidade de desenvolver câncer e em ter deficiências reprodutivas, assim como de transmitir doenças alélicas geradas *de novo* (Hedges e Deininger, 2007).

Kidwell e Lisch (2001) apresentam uma revisão sobre o papel do TEs na evolução dos genomas e sobre a estreita relação TE-hospedeiro. Nessa revisão eles mostram que a própria evolução dos TEs, com toda a sua rica complexidade, revela a artificialidade da distinção entre “DNA egoísta” e “DNA altruísta”. Segundo esses autores, a relação TE-hospedeiro é tão dinâmica e íntima que qualquer distinção rígida entre os dois é sem sentido.

2) Transferência Horizontal X Transferência Vertical

Além de ter como característica principal a propriedade de mudar de local dentro do genoma, os TEs possuem a impressionante propriedade de atravessar as barreiras inter-específicas para invadir novos genomas hospedeiros. Este processo, denominado Transferência Horizontal (HT, do inglês “Horizontal Transfer”), pode ser definido como a troca de material genético entre espécies isoladas reprodutivamente (Capy *et al.*, 1998). A primeira inferência de HT em *Drosophila* foi a recente invasão do elemento *P* no genoma de *D. melanogaster* (Kidwell, 1983), detectada a partir de estudos sobre a disgenesia do híbrido. Este fenômeno é causado pela mobilização de certas famílias de elementos em um novo genoma, o qual não possui mecanismos para reprimir transposições excessivas, gerando diversos tipos de mutantes. Este caso de HT do elemento *P* foi posteriormente demonstrado por dados moleculares (Daniels *et al.*, 1990a; Clark *et al.*, 1995, 2002; Clark e Kidwell, 1997) e tornou-se um dos casos mais famosos e convincentes de HT em *Drosophila*.

Após esse caso, vários outros eventos de HT no gênero *Drosophila* vêm sendo relatados, envolvendo elementos como: *mariner* (Maruyama e Hartl, 1991; Brunet *et al.*, 1999, entre outros), *Minos* (Arca e Savakis, 2000; Almeida e Carareto, 2005), *hobo* (Daniels *et al.*, 1990b; Simmons, 1992, entre outros), *gypsy* (De Frutos *et al.*, 1992; Herédia *et al.*, 2004; Ludwig e Loreto, 2008), *copla* (Jordan *et al.*, 1999; Almeida e Carareto, 2006) e *gtwin* (Ludwig e Loreto, 2007; Kotnova *et al.*, 2007).

Quando os TEs são transmitidos verticalmente, espera-se que suas histórias filogenéticas retratem, ao menos em termos amplos, aquelas de seus hospedeiros (Silva *et al.*, 2004). Quando inconsistências entre as filogenias dos TEs e de seus hospedeiros são reveladas, deve-se considerar quais possíveis alternativas levam à explicação desses dados. Uma possível explicação para tais inconsistências filogenéticas é a de que estes TEs não tenham sido transmitidos verticalmente e sim por eventos de HT entre as espécies.

Segundo recente revisão de Loreto *et al.* (2008), além de premissas básicas como sobreposição geográfica, ecológica e temporal entre as espécies possivelmente envolvidas, três tipos de distorções filogenéticas são normalmente usadas para inferir eventos de HT de TEs: 1) detecção de elementos com alto grau de similaridade de seqüência entre espécies distantes evolutivamente; 2) diferenças topológicas entre as filogenias dos TEs e de suas espécies hospedeiras; e 3) distribuição descontínua de um TE entre espécies próximas evolutivamente.

No entanto, mesmo que essas situações sejam verdadeiras para um dado caso, antes de inferir eventos de HT, é necessário excluir qualquer possibilidade de transferência vertical, uma vez que polimorfismo ancestral, taxas evolutivas diferentes entre espécies, pressões seletivas fortes em porções do elemento ancestral e perdas estocásticas podem ocorrer (Capy *et al.*, 1998).

3) Classificação dos TEs

A classificação de TEs passou por mudanças a partir de dezembro de 2007, com a publicação de Whicker *et al.* (2007). Os autores reorganizam o sistema de classificação de TEs, mantendo padrões antigos, como a divisão em duas Classes de acordo com o intermediário de transposição, e inovando em outros, como a divisão da Classe II em subclasses de acordo com o número de fitas de DNA clivadas no sítio doador durante o processo de transposição.

Elementos de Classe I, também chamados retrotransposons, utilizam um intermediário de RNA ao transporem-se, não havendo clivagem no sítio doador. O RNA intermediário é transcrito de uma cópia do elemento que está inserido no genoma e em seguida é transcrito reversamente para

cDNA, por uma transcriptase reversa. Na seqüência, este cDNA é inserido em outro local do genoma. Este processo de mobilização é descrito como replicativo, pois aumenta o número de cópias do elemento no genoma a cada evento de transposição.

A Classe I é subdividida em cinco Ordens, com base em características de seu mecanismo, organização e filogenia da transcriptase reversa: retrotransposons com longas repetições terminais (LTRs – do inglês “long terminal repeats”), LINEs (do inglês “long interspersed nuclear elements”, ou seja, longos elementos nucleares dispersos), SINEs (do inglês “short interspersed nuclear elements”, ou seja, curtos elementos nucleares dispersos), elementos relacionados à DIRS (“DIRS-like”) e elementos relacionados à *Penelope* (PLEs - “Penélope-like”)

Nos retrotransposons com LTRs, as longas repetições terminais diretas flanqueiam uma região que tipicamente possui duas fases abertas de leitura, *gag* e *pol*, que codificam as enzimas necessárias para a sua replicação e transposição. LINEs e SINES não possuem LTRs. Os LINEs autônomos codificam pelo menos algumas enzimas requeridas para sua transposição, como a transcriptase reversa e uma nuclease, em sua fase aberta de leitura *pol*. Os SINEs são não-autônomos e as proteínas requeridas para sua mobilização podem ser fornecidas pelos elementos LINEs, embora não sejam derivados de elementos autônomos de Classe I deletados. As Ordens dos DIRs-like e PLEs surgiram com a nova classificação de TEs (Whicker *et al.*, 2007) e, apesar de apresentarem mecanismo/estrutura diferente dos outros elementos de Classe I, foram inseridos nesta classe por também se transporem com intermediário de RNA.

Elementos de Classe II, também chamados transposons de DNA, transpõem-se diretamente via DNA, e são divididos em duas subclasses de acordo com o número de fitas de DNA que são clivadas no processo de transposição (para revisão detalhada da classificação de TEs ver Wicker *et al.*, 2007). A Subclasse 1 compreende os clássicos TEs, os quais clivam ambas as fitas de DNA durante a transposição, por um mecanismo denominado “corta-e-cola”. A Subclasse 2 compreende os TEs que transpõem-se clivando apenas uma

das fitas de DNA, incluindo 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 que não possuem mecanismo de transposição completamente elucidado.

Os elementos da Subclasse 1 podem ser divididos em duas Ordens: a Ordem TIRs (do inglês “Terminal Inverted Repeats”, ou seja, que possuem repetições terminais invertidas) e a Ordem Crypton. Whicker *et al.* (2007) incluem Crypton como uma nova ordem, pois, embora se conheça muito pouco sobre este elemento, acredita-se que ele transponha-se diretamente por DNA a partir da clivagem de ambas as fitas de DNA (sendo posicionado assim na Subclasse 1 da Classe II). No entanto, Crypton não possui TIRs, sendo assim designado como uma segunda Ordem da Subclasse 1.

Membros da Ordem TIRs são elementos da Classe II Subclasse 1 que possuem TIRs flanqueando o gene da transposase. A transposase é a enzima responsável por reconhecer as TIRs e clivar as duas fitas de DNA em ambas as pontas do elemento no processo de transposição. Estes elementos estão distribuídos em nove Superfamílias, de acordo com a estrutura da proteína ou domínios não codificadores, que são: *Tc1-Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger* e *CACTA*. Elementos da mesma Superfamília apresentam limitada similaridade de seqüência de aminoácidos e compartilham características como presença/ausência e tamanho da duplicação do sítio alvo nas extremidades do elemento no momento de sua inserção e baixa similaridade de seqüência das TIRs .

4) Superfamília *hAT*

Desde a caracterização molecular dos primeiros TEs descritos (*Ac* – *Activator* e *Dc* - *Dissociation*), no início da década de 1980, uma quantidade crescente de TEs têm sido isolados e caracterizados. Logo se tornou evidente que alguns transposons compartilhavam características estruturais e funcionais com o elemento *Ac*, formando uma dentre as várias famílias de transposons (Kempken e Windhofer, 2001). Em 1994, Warren *et al.* criaram o nome *hAT* para a Superfamília, fazendo referência aos seus três membros fundadores: *hobo* de

Drosophila melanogaster (Calvi *et al.*, 1991), *Activator* (*Ac*) de *Zea mays* (McClintock, 1948), *Tam3* de *Antirrhinum majus* (Hehl *et al.*, 1991).

Uma característica compartilhada pela maioria dos elementos *hAT* é a presença de TIRs normalmente curtas (5 a 27pb) e com seqüências pouco definidas, além de poucos domínios conservados na transposase (Rubin *et al.*, 2001). Além disso, a inserção desses elementos em determinado local causa a duplicação 8pb do genoma do hospedeiro.

O número e estrutura dos domínios conservados entre seqüências da transposase de elementos *hAT* variam de acordo com os elementos analisados e com os programas utilizados. Calvi *et al.* (1991) localizam três regiões conservadas, situadas da porção mediana à final da transposase, denominadas regiões 1, 2 e 3, sendo que a região 3, localizada na porção C-terminal da transposase, apresenta alta similaridade entre os elementos *hobo*, *Ac* e *Tam3*. Kempken e Windhofer (2001), analisando elementos *hAT* de plantas, animais e fungos e identificaram os mesmos domínios conservados que Calvi *et al.* (1991). Já Rubin *et al.* (2001) identificam seis blocos de aminoácidos conservados, dividindo em dois domínios a região 2 e em três domínios menores a região 3 descrita por Calvi *et al.* (1991). Ortiz e Loreto (2009), analisando outro conjunto de elementos *hAT*, encontram estes mesmos seis domínios, entretanto, com o emprego de diferentes programas computacionais este número pode chegar a 18 domínios conservados.

Hickman *et al.* (2005) apresentam um estudo profundo da arquitetura molecular da proteína da transposase do elemento *Hermes*, membro da superfamília *hAT*. Os autores determinaram a estrutura tridimensional da transposase de *Hermes* e caracterizaram três domínios estruturais: 1) um domínio N-terminal; 2) um domínio catalítico com assinatura DDE e 3) um grande domínio todo em α -hélice inserido no domínio catalítico, constituído pelos aminoácidos da porção mediana à final da transposase.

O domínio N-terminal da transposase identificado por Hickman *et al.* (2005) parece ser o domínio de ligação a DNA sítio-específico, responsável pelo reconhecimento das TIRs, e está localizado adjacente ao domínio BED “finger”. Aravind (2000) caracteriza o domínio BED “finger”, também chamado Zn-BED, (que possui o motivo BED e o motivo “Zinc finger”), presente na região N-terminal

de transposases de vários elementos *hAT*. O domínio BED “finger” é aparentemente responsável pela propriedade de ligação da transposase ao DNA (Aravind, 2000) ou, mais especificamente, ligação ao DNA não específico durante a transposição (Hickman *et al.*, 2005). Uma região N-terminal, de aproximadamente 200 aminoácidos, também foi identificada por Kunze (1996) ao estudar a transposase codificada pelo elemento *Ac*, de milho. Esta região também é apontada como apresentando funções de transporte nuclear e de ligação a DNA (Kunze, 1996).

Hickman *et al.* (2005) identificam um domínio catalítico que possui uma assinatura DDE, comumente encontrada em integrases de retrotransposons e retrovírus e no centro catalítico de transposases (Capy *et al.*, 1998). Entretanto, esta assinatura não havia sido previamente encontrada em transposases de elementos da Superfamília *hAT*.

A região C-terminal da transposase de *Hermes* já havia sido previamente caracterizada, correspondendo, ao menos em termos, à região 3 proposta por Calvi *et al.* (1991) (Michel *et al.*, 2003). Esta porção C-terminal é tida como a mais conservada entre membros da Superfamília *hAT* e contém um domínio de dimerização. Um domínio equivalente foi previamente descrito por Essers *et al.* (2000) para a transposase de *Ac* e é denominado domínio de dimerização da família *hAT*. Essers *et al.* (2000) indicam que vários segmentos do domínio de dimerização apresentam uma estrutura predita de α -hélice e que este domínio pode ser considerado uma seqüência assinatura de transposases de elementos da Superfamília *hAT*.

A Superfamília *hAT* é numerosa e amplamente distribuída, tendo membros descritos nos três reinos eucarióticos (por exemplo, *dTok* em espécies do gênero *Oryza* – Moon *et al.*, 2006; *Restless* em *Tolypocladium inflatum* – Kempken and Kück, 1996; e *Myotis-nhAT* em *Myotis lucifugus* – Ray *et al.*, 2007). No genoma humano, por exemplo, há cerca de 195.000 cópias de membros de *hAT* (“International Human Genome Sequencing Consortium”, 2001).

Várias famílias de elementos *hAT* foram caracterizadas em insetos, como *hobo* em *D. melanogaster* (Calvi *et al.*, 1991) *Hermes* em *Musca domestica* (Warren *et al.*, 1994), *Hermit* em *Lucilia cuprina* (Coates *et al.*, 1996), *Homer* em *Bactrocera tryoni* (Pinkerton *et al.*, 1999), *hopper* em *B. dorsalis* (Handler e

Gomez, 1997) e *Herves* em *Anopheles gambiae* (Arensburger *et al.*, 2005). Recentemente, novas seqüências de *hAT*, pertencentes a quatro famílias diferentes, foram identificadas por Ortiz e Loreto (2009) através de buscas *in silico* no 12 genomas disponíveis de *Drosophila* (“*Drosophila* 12 Genomes Consortium”, 2007).

Esta ampla distribuição de elementos *hAT* parece ser devida à origem muito antiga desta superfamília, provavelmente precedendo a separação de plantas, fungos e animais (Rubin *et al.*, 2001). No entanto, alguns casos de HTs podem explicar a distribuição destes elementos e as relações entre seqüências *hAT* em alguns grupos de espécies. Eventos de HTs parecem fazer parte da história evolutiva de alguns membros da Superfamília *hAT* como, por exemplo, *Herves*, transmitido de um doador desconhecido para o genoma de *Anopheles gambiae* (Subramanian *et al.*, 2007); *Tol2*, o qual invadiu duas espécies de peixes do gênero *Oryzias* (Koga *et al.*, 2000) e *Myotis-hAT1* passado para o genoma de morcegos do gênero *Myotis* a partir de uma fonte desconhecida (Ray *et al.*, 2007). Este último caso é especialmente interessante, visto que, até então, acreditava-se que transposons de DNA se tornaram inativos na linhagem de mamíferos há muito tempo e este é o primeiro registro de atividade recente de transposons de DNA em uma linhagem de mamíferos.

No gênero *Drosophila*, o elemento *hobo*, da Superfamília *hAT*, é um dos TEs mais estudados e aparentemente está envolvido em eventos de HTs recentes entre membros do grupo *melanosgaster*, pertencente ao subgênero *Sophophora* (Boussy e Daniels, 1991; Pascual e Periquet, 1991; Simmons, 1992). Além disso, foi sugerido que *hobo* foi transmitido horizontalmente entre alguma espécie do grupo *melanosgaster* e uma espécie de mosca-das-frutas de um gênero distinto, a *Ceratitis capitata* (Handler e Gomez, 1996; Torti *et al.*, 2005).

5) Gênero *Drosophila*

O gênero *Drosophila* possui uma distribuição ampla e dinâmica, sendo composto por 1.157 espécies, agrupadas em 8 subgêneros (Bächli, 2008), sendo eles: *Chusqueophila*, *Dorsilopha*, *Drosophila*, *Dudaica*, *Phloridosa*, *Psilodorha*, *Siphlodora* e *Sophophora*. Dentre os subgêneros, podemos destacar dois, *Drosophila* e *Sophophora*, pela diversidade de espécies apresentadas.

Segundo Russo *et al.* (1995), estes dois subgêneros divergiram há cerca de 40 milhões de anos. Tamura *et al.* (2004), por outro lado, ao corrigir distâncias sinônimas para os efeitos de viés na utilização de códons, obtém estimativas mais antigas, da ordem de 63 milhões de anos. De acordo com Robe (2008), a partir de extenso estudo com uso de marcadores nucleares e biogeografia, estes dois subgêneros divergiram há cerca de 55 milhões de anos.

5.1) Subgênero *Drosophila*

O subgênero *Drosophila* é o maior representante do gênero *Drosophila*, com pelo menos 739 espécies, subdivididas em 45 grupos (Bächli, 2008): *angor*, *annulimana*, *antioquia*, *aureata*, *bizonata*, *bromeliae*, *calloptera*, *canalina*, *caponei*, *carbonaria*, *cardini*, *carsoni*, *coffeata*, *dreyfusi*, *flavopilosa*, *funnebris*, *guarani*, *guttifera*, *histrion*, *immigrans*, *macroptera*, *melanica*, *mesophragmatica*, *morelia*, *nannoptera*, *nigrosarsa*, *onychophora*, *pallidipennis*, *peruensis*, *peruviana*, *picta*, *pinicola*, *polychaeta*, *quadrisetata*, *quinaria*, *repleta*, *robusta*, *rubrifrons*, *simulivora*, *sticta*, *testacea*, *tripunctata*, *tumiditarsus*, *virilis* e *xanthopallescens*.

Throckmorton (1975) propõe para o subgênero *Drosophila* duas linhagens principais: a radiação *virilis-repleta* e a radiação *quinaria-tripunctata*, além do grupo *funnebris*, basal em relação a ambos. Segundo Russo *et al.* (1995), membros das duas radiações divergiram entre si há aproximadamente 33 milhões de anos. Já Robe (2008) indica que a separação entre essas duas radiações ocorreu há aproximadamente 45 milhões de anos, no Oriente do Velho Mundo. Ainda segundo esses autores, depois de diversificarem-se, as duas radiações invadiram a região Holártica

(há aproximadamente 32-28 milhões de anos), chegaram à região Neártica e, em seguida (há cerca de 25-22 milhões de anos) invadiram a região Neotropical.

A radiação *virilis-repleta* é formada por 19 grupos, dos quais sete são usados no presente estudo, sendo eles: *mesophragmatica*, *repleta*, *annulimana*, *canalinaea*, *flavopilosa*, *virilis* e *robusta*.

O grupo *mesophragmatica*, descrito originalmente por Brncic e Koref-Santibañez (1957), compreende 13 espécies de distribuição exclusivamente Neotropical, sendo praticamente restrito ao complexo de montanhas dos Andes (Bächli, 2008). Quatro espécies deste grupo foram analisadas neste estudo: *D. gasici*, *D. brncici*, *D. gaucha* e *D. pavani*.

O grupo *repleta* é composto por 101 espécies (Bächli, 2008), sendo assim um dos maiores grupos do gênero *Drosophila*. Este grupo apresenta uma distribuição tipicamente Neotropical (Val *et al.*, 1981). A maioria das espécies do grupo *repleta* apresenta distribuição desértica, normalmente associada aos cactos, embora algumas espécies possam ser encontradas em regiões mais úmidas de florestas e em plantações de banana e café da América Central e América do Sul (Throckmorton, 1975). As espécies *D. hydei*, *D. mercatorum* e *D. mojavensis* pertencem ao grupo *repleta* e foram incluídas neste estudo.

O grupo *annulimana* ocupa um posicionamento relativamente basal dentro da radiação *virilis-repleta* (Tatarenkov e Ayala, 2001). Duas espécies (*D. annulimana* e *D. schineri*) das 16 descritas para este grupo (Bächli, 2008) foram analisadas aqui. A maioria das espécies do grupo apresenta distribuição exclusivamente Sul-Americana; somente quatro espécies, não amostradas neste estudo, possuem uma distribuição mais ampla, podendo ser encontradas desde o México até o Brasil e Argentina (Val *et al.*, 1981).

D. canalinea foi a única espécie incluída do grupo *canalinaea*. Este grupo compreende atualmente 11 espécies (Bächli, 2008) e, segundo Throckmorton (1975), possui uma distribuição centralizada na América de Sul e suas espécies são essencialmente habitantes de florestas.

Do grupo *flavopilosa*, as espécies amostradas foram *D. cestri* e *D. incompta*. Este grupo é composto por 17 espécies e possui distribuição essencialmente Neotropical. As espécies desse grupo caracterizam-se por

apresentar uma ecologia restrita, associada às flores do gênero *Cestrum* (Solanaceae).

Apenas uma espécie, *D. virilis*, das 12 descritas para o grupo *virilis* (Bächli 2998/12), foi analisada neste estudo. Este grupo apresenta uma distribuição tipicamente boreal (Markov and O'Grady, 2006), sendo que apenas *D. virilis* apresenta uma distribuição mais abrangente, estendendo-se até a região Neotropical (Val et al., 1981).

O grupo *robusta* também foi amostrado com uma única espécie, *D. robusta*, das 16 descritas (Bächli, 2008). As espécies do grupo são encontradas principalmente em regiões temperadas da América do Norte, da Europa e da Ásia (Markov e O'Grady, 2006).

A radiação *quinaria-tripunctata*, por sua vez, é constituída por pelo menos 13 grupos de espécies, dos quais seis foram amostrados neste estudo: *guarani*, *tripunctata*, *cardini*, *pallidipennis*, *calloptera* e *immigrans*.

O grupo *guarani* é formado por 16 espécies, as quais apresentam distribuição exclusivamente Neotropical, e subdividido em dois subgrupos. Neste estudo foram utilizadas cinco espécies do grupo *guarani*, sendo três do subgrupo *guarani*: *D. ornatifrons* e *D. guaru*, que ocorrem essencialmente em território brasileiro, tendo apenas um registro de coleta de *D. ornatifrons* no Uruguai, e *D. subbadia*, que apresenta distribuição restrita ao México (Bächli, 2008). As outras duas espécies analisadas, pertencentes ao subgrupo *guaramuru*, foram *D. griseolineata*, encontrada em regiões do Brasil, Bolívia, Colômbia, Equador e Venezuela, e *D. maculifrons*, com registros no Brasil, Argentina, Bolívia e Peru (Salzano, 1955; Kastritsis, 1969; Bächli, 2008).

O grupo *tripunctata*, proposto por Sturtevant (1942), engloba atualmente 79 espécies (Bächli, 2008). Segundo Vilela (1992), com exceção de *D. tripunctata* que ocorre em regiões Neárticas, todas as demais espécies são endêmicas da região Neotropical, onde este grupo é superado apenas pelo grupo *repleta* com relação à diversidade de espécies. Deste importante grupo foram incluídas as espécies *D. nappae*, *D. paraguayensis*, *D. crocina*, *D. paramediostriata*, *D. tripunctata*, *D. mediodifusa*, *D. bandeirantium* e *D. mediopictoides*.

O grupo *cardini* é constituído por 16 espécies (Bächli, 2008) com distribuição ao longo da América Neotropical, abrangendo inclusive as ilhas do

Caribe. Deste grupo foram incluídas as seguintes espécies: *D. cardinoides*, *D. neocardini*, *D. polymorpha*, *D. procardinoides* e *D. arawakana*.

Os seguintes grupos foram representados por uma única espécie: grupo *pallidipennis*, o qual é formado unicamente por *D. pallidipennis*, que possui distribuição Neotropical; grupo *calloptera*, formado por oito espécies com distribuição essencialmente Latino-America, foi representado por *D. ornatipennis*, a qual é endêmica das Ilhas do Caribe; grupo *immigrans*, primariamente habitante do Velho Mundo, é constituído por 101 espécies, amostrado por *D. immigrans*, a única espécie cosmopolita do grupo; e grupo *funebri*, atualmente composto por sete espécies, das quais a única representante cosmopolita, *D. funebri*, foi utilizada (Bächli, 2008).

5.2) Subgênero *Sophophora*

O subgênero *Sophophora*, estabelecido por Sturtevant (1939), é constituído atualmente por 330 espécies e dividido em oito grupos (Bächli, 2008), dos quais quatro são amostrados aqui: *melanogaster*, *obscura*, *saltans* e *willistoni*. Os grupos-irmãos *saltans* e *willistoni* são descritos um pouco mais detalhadamente, uma vez que os resultados do nosso estudo envolvem principalmente estes dois grupos.

O grupo *melanogaster* é atualmente constituído por 185 espécies e subdividido em 12 subgrupos (Bächli, 2008). As espécies deste grupo são distribuídas ao longo das regiões tropicais e subtropicais do Velho Mundo, com exceção das quatro espécies do grupo que são consideradas cosmopolitas: *D. ananassae*, *D. kikkawai*, *D. simulans* e *D. melanogaster* (Throckmorton, 1975). Neste estudo as seguintes espécies foram utilizadas: *D. melanogaster*, *D. simulans*, *D. sechelia*, *D. mauritiana*, *D. teissieri*, *D. santomea*, *D. erecta*, *D. yakuba*, *D. kikkawai*, *D. ananassae* e *D. malerkotliana*.

O grupo *obscura*, atualmente formado por 41 espécies (Bächli, 2008), é encontrado principalmente em florestas temperadas Holárticas e em ambientes de altas altitudes, com clima similar ao temperado (Gao *et al.*, 2007). Throckmorton (1975) sugere que este grupo tenha se originado nos trópicos do Velho mundo, a partir de uma linhagem *proto-melanogaster*.

Segundo esse autor, após uma adaptação a ambientes temperados, este grupo teria se diversificado na região Paleártica e se espalhado pela América do Norte através do Estreito de Bering. Apenas uma espécie deste grupo, *D. pseudoobscura*, foi incluída neste trabalho.

O grupo *saltans* é constituído por 20 espécies atualmente, divididas em cinco subgrupos (Bächli, 2008). Espécies de três subgrupos foram incluídas neste estudo: subgrupo *saltans* (*D. prosaltans* e *D. saltans*), subgrupo *elliptica* (*D. neoelliptica*) e subgrupo *sturtevanti* (*D. sturtevanti*). Dentre estes subgrupos amostrados, Trockmorton (1975) e O' Grady *et al.* (1998) postulam o subgrupo *elliptica* como o mais primitivo, o subgrupo *saltans* como o mais derivado e o subgrupo *sturtevanti* como intermediário. No entanto, esta ordem de divergência não é consenso, uma vez que trabalhos recentes têm recuperado histórias evolutivas diferentes para o grupo *saltans* (Rodriguez-Trelles *et al.*, 1999; Tarrío *et al.*, 2000). Como consequência, o local de origem e diversificação do grupo *saltans* volta a ser dúvida, podendo ter ocorrido nos trópicos da América do Norte (Trockmorton, 1975) ou da América do Sul (Rodriguez-Trelles *et al.*, 1999). De qualquer maneira, a distribuição atual das espécies do grupo *saltans* é essencialmente Neotropical, com representantes desde o Sul do Brasil até o Nordeste do México e Sudoeste dos Estados Unidos da América (Bächli, 2008).

O grupo *willistoni* apresenta distribuição essencialmente Neotropical, estando presente desde o México e sul da Flórida (Estados Unidos da América), passando pela América Central e Ilhas do Caribe, até o Sul do Brasil e Argentina (Ayala *et al.*, 1970). As espécies do grupo variam quanto à sua abrangência: algumas espécies estão amplamente distribuídas enquanto outras são restritas a regiões do continente americano ou insulares (Gleason *et al.*, 1998).

O grupo *willistoni* é formado por 23 espécies, distribuídas em três subgrupos (Bächli, 2008). O subgrupo *willistoni*, constituído por seis espécies crípticas, com morfologia praticamente idêntica, distinguíveis por leves diferenças na genitália masculina. Quatro dentre as espécies crípticas (*D. willistoni*, *D. paulistorum*, *D. equinoxialis* e *D. tropicalis*), apresentam uma maior distribuição geográfica, sobrepondo-se em várias regiões, enquanto as outras duas espécies crípticas (*D. insularis* e *D. pavlovskiana*) são endêmicas (Gleason *et al.*, 1998). Neste estudo foram utilizadas todas as espécies do subgrupo *willistoni*, com

exceção de *D. pavlovskiana*. O subgrupo *bocainensis* é formado por 12 espécies, das quais quatro são incluídas no presente estudo, sendo elas: *D. capricorni*, *D. fumipennis*, *D. nebulosa* e *D. sucinea*. Nenhuma das cinco espécies do subgrupo *alagitans* foi analisada.

Além de espécies destes grupos apresentados, incluímos em nossa amostra também uma espécie representante do subgênero *Dorsilopha* (*D. busckii*), duas espécies do gênero *Zaprionus* (*Z. indianus* e *Z. tuberculatus*) e duas do gênero *Scaptodrosophila* (*S. latifasciformis* e *S. lebanonensis*).

6) Sobre a escolha do tema

A idéia deste estudo partiu de resultados obtidos por Ortiz e Loreto (2009). Embora esse artigo tenha sido publicado no início de 2009, tivemos acesso aos resultados com bastante antecedência, no início de 2007, por se tratar do mesmo grupo de pesquisa.

Assim, nosso estudo é baseado em duas seqüências de elementos da Superfamília *hAT*, identificadas por Ortiz e Loreto (2009), a partir de buscas *in silico* nos 12 genomas disponíveis de *Drosophila* (“*Drosophila* 12 Genomes Consortium”, 2007): *Homo3* de *D. mojavensis* e *Howilli3* de *D. willistoni*. Essas seqüências compartilham uma inesperada similaridade de 97% no nível de aminoácido, o que é surpreendente dado que suas espécies hospedeiras são distantemente relacionadas, pertencendo a diferentes subgêneros de *Drosophila*. Esta intrigante descoberta nos levou a tentar compreender melhor o cenário no qual estas seqüências estão evoluindo no gênero *Drosophila*.

Para facilitar a compreensão do item seguinte (Objetivo), esclarecemos desde já que *Homo3* e *Howilli3*, juntamente com as seqüências obtidas no presente estudo, formam uma nova família de elementos transponíveis da Superfamília *hAT*, a qual denominamos *harrow*.

7) Objetivos

O objetivo geral deste trabalho é contribuir para o entendimento sobre a evolução de elementos transponíveis da Superfamília *hAT* em espécies do Gênero *Drosophila*.

Os objetivos específicos são:

1) Caracterização de uma nova família de elementos, pertencente à Superfamília *hAT*, denominada *harrow*;

2) Elucidar a distribuição de seqüências *harrow* no genoma de espécies do gênero *Drosophila*;

3) Compreender as relações filogenéticas entre seqüências da família *harrow* e seus padrões evolutivos;

4) Testar se estas seqüências foram transmitidas vertical ou horizontalmente entre os genomas das espécies hospedeiras;

5) Inferir o cenário responsável pela evolução desta família de elementos transponíveis no gênero *Drosophila*.

CAPÍTULO II

***harrow*: A new *Drosophila* family of *hAT* transposons involved in horizontal transfer**

**Nina Roth Mota¹, Adriana Ludwig¹, Vera Lúcia da Silva Valente¹ and Élgion
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), Campus Universitário, Santa Maria, Rio Grande do Sul, Brazil, CEP
97105-900.

Manuscrito submetido à revista **Insect Molecular Biology**

Abstract

In this study, a new family of transposable elements, *harrow*, belonging to the *hAT* superfamily of DNA transposons, is described. Searches for *harrow* sequences were performed in 65 Drosophilidae species, mainly representing Neotropical and cosmopolitan groups from genus *Drosophila*. The patchy distribution through out genus *Drosophila* and the high similarity presented between all *harrow* sequences indicate that horizontal transfer could play a major role in the evolution of *harrow* family. The results obtained suggest an evolutionary scenario where *harrow* would have gone under multiple horizontal transfer events in the Neotropics, involving *D. tripunctata*, *D. mojavensis* (Subgenus *Drosophila*) and several species of the *willistoni* and *saltans* groups (Subgenus *Sophophora*).

Introduction

Transposable elements (TEs) are the largest component of the genetic material of most eukaryotes, where they are considered to have a profound impact on genome evolution. The Class II elements, which transpose directly via DNA, are characterized by terminal inverted repeats (TIRs) flanking a single gene that encodes a transposase, required for transposition. *hAT* elements comprise a large and prevalent superfamily of DNA transposons which is widely spread and has been found in all eukaryotic kingdoms (e.g. *Activator* from *Zea mays* – McClintock, 1948; *Restless* from *Tolypocladium inflatum* – Kempken & Kück, 1996; *Myotis-nhAT* from *Myotis lucifugus* – Ray *et al.*, 2007). Members of *hAT* superfamily share common features, like the presence of short (5 to 27 bp) and weakly defined TIRs, as well as few conserved domains in the transposase. Moreover, they promote 8 bp duplications in the insertion site.

Several families of *hAT* elements have been characterized in insects, like *hobo* from *D. melanogaster* (Calvi *et al.*, 1991), *Hermes* from *Musca domestica* (Warren *et al.*, 1994), *Homer* from *Bactrocera tryoni* (Pinkerton *et al.*, 1999), *hermit* from *Lucilia cuprina* (Coates *et al.*, 1996), *hopper* from *B. dorsalis* (Handler & Gomez, 1997) and *Herves* from *Anopheles gambiae* (Arensburger *et al.*, 2005). Recently, new *hAT* sequences, belonging to four different families, were identified by Ortiz & Loreto (2009) through *in silico* searches on the 12 *Drosophila* genomes available (*Drosophila* 12 Genomes Consortium, 2007).

This wide distribution of *hAT* superfamily TEs seems to be due to the very ancient origin of these elements, probably predating the plant-fungi-animal separation (Rubin *et al.*, 2001). However, some cases of horizontal transfer (HT) can explain particular distribution and relationships of *hAT* sequences in some species groups, like *Herves*, transmitted from an unknown source to *Anopheles gambiae* (Subramanian *et al.*, 2007), *Tol2*, invading two fish species of the *Oryzias* genus (Koga *et al.*, 2000), *Myotis-hAT1* to the bat genus *Myotis* from an unknown source (Ray *et al.*, 2007). In the *Drosophila* genus, the *hAT* member *hobo* is one of the most studied TEs and is apparently involved in HT events between members of the *D. melanogaster* group (e.g. Simmons, 1992) and even between

these and a member of a distinct genus, *Ceratitis capitata* (Handler & Gomez, 1996; Torti *et al.*, 2005).

Our study is based on two *hAT* sequences, *Homo3* from *D. mojavensis* and *Howilli3* from *D. willistoni*, identified by Ortiz & Loreto (2009). These sequences share an unexpected 97% similarity at amino acid level, which is surprising given that their host species are distantly related, even belonging to different *Drosophila* subgenera. This intriguing finding led us try to better understand the scenario in which these sequences are evolving in the genus *Drosophila*. Ortiz & Loreto (2009) placed *Homo3* and *Howilli3* inside the *homo* family. However, according to a new TE classification proposed by Whicker *et al.* (2007), these sequences, together with the ones obtained in this study, should be grouped on a family of their own. Here we propose to name this new family as *harrow*.

Results

Characterization of a new hAT family

Ortiz & Loreto (2009) have classified the *Drosophila* elements from *hAT* superfamily into four different families based on amino acid similarity, following Capy *et al.* (1998). *Homo3* and *Howilli3* sequences were placed inside their *homo* family. This family was composed additionally for five other sequences from *D. mojavensis* (*Homo5*, 7, 8, 9, 10), three sequences from *D. ananassae* (*Hoana2*, 6 and 7), one from *D. yakuba* (*Hoyak2*), one from *D. sechellia* (*Hosec2*) and the element *hopper* from *B. dorsalis*. Using the 1500 bp most conserved region, we detect an overall mean nucleotide divergence of 46%. The relationships of these sequences are shown in the Figure S1 of Supplementary Material. The phylogeny shows three well supported groups: 1) one clustering *Homo9* sequences together with *Homo3* and *Howilli3* sequences; 2) the second with *hopper* sequences; and 3) a third one comprised by the remaining sequences, which presents high internal divergence. The divergences between and within groups are shown in Table S1 of Supplementary Material. Within the first group, the nucleotide divergence between *Homo9* sequences clade and the one containing *Howilli3* and *Homo3* is 29.7%.

This divergence indicates that these two clusters should be considered as different families according to the criteria proposed by Whicker *et al.* (2007) of 20% divergence to split TE sequences into different families.

Additionally, we used the hierarchical system classification steps, suggested by Whicker *et al.* (2007), to classify the studied TE. First, a BLASTn was performed on NCBI nucleotide collection and on Repbase (Jurka *et al.*, 2005) to ascertain the sequences do not belong to a known family of TEs. In this search, no significant hit was obtained, indicating that the studied sequences could be classified as a new family. In the next step, a BLASTx search validated the classification of these sequences as belonging to *hAT* superfamily, from Order TIR, with *hopper* as the most similar sequence. Once confirming we were indeed dealing with a new family of TEs, we propose to name it as *harrow*, since the resulting phylogeny of the sequences included in this family present a polyphyletic clade, resembling the gardening tool, harrow.

Complete *harrow* sequences were identified from *D. mojavensis* and *D. willistoni* genomes by Ortiz & Loreto (2009). No new complete *harrow* sequence was obtained in our work. *D. mojavensis* genome presents only one complete *harrow* sequence and *D. willistoni* presents two complete *harrow* sequences, however one of them (*Howilli3.2*) has small internal deletions, which eliminates its encoding capacity. At contrast, *Howilli3.1* and *Homo3* were presented as putatively active elements (Ortiz & Loreto, 2009). Here, we identify that *Howilli3.1*, like other *hAT* elements (e.g. *hoho-like*, from *Caenorhabditis elegans* – Bigot *et al.*, 1996; and *Folyt1*, from *Fusarium oxysporum* – Gómez-Gómez *et al.*, 1999) contains an intron, which is 64 bp and splits the BED zinc finger (Zf-BED) and DUF-659 domains from the *hAT* dimerization domain (hATC) (Fig.1). *Homo3*, however, has a 101 bp deletion, which includes the splicing acceptor site, preventing the joining of the two exons. This results in *Homo3* encoding a shorter protein, with no hATC, what probably makes it a non-active element.

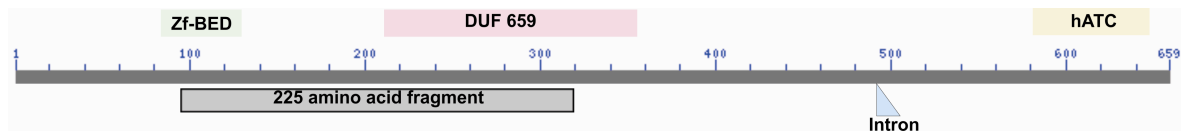


Fig.1. Schematic representation of the *harrow* transposase amino acid sequence, based on *Howilli3.1*. Above the full transposase (659 amino acid), the three domains identified by CDD: Zf-BED domain, the DUF659 domain and the hATC. The rectangle below the transposase represents the analyzed 225 amino acid fragment used in this study, in approximated size and position. The triangle localizes the 64 pb intron of *Howilli3.1*, identified by NetGene2 Server (Brunak *et al.*, 1991; Hebsgaard *et al.*, 1996).

harrow family shows a patchy distribution

A PCR screening for sequences homologous to *harrow* was performed in 61 *Drosophila* species, mainly representing Neotropical and cosmopolitan groups. Among these, 35 species belong to subgenus *Drosophila*, 25 species to subgenus *Sophophora* and one to subgenus *Dorsilopha*. In addition, two species from genus *Zaprionus* and two from genus *Scaptodrosophila* were also screened by PCR. Supplementary Table II lists the species used in this study, together with their PCR results. Besides *D. mojavensis* and *D. willistoni*, we obtained PCR amplification in *D. tripunctata* (subgenus *Drosophila*) and in seven species of the subgenus *Sophophora*: *D. equinoxialis*, *D. paulistorum*, *D. fumipennis*, *D. nebulosa*, *D. prosaltans*, *D. neoelliptica* and *D. sturtevantii*. These Sophophoran species belong to the Neotropical *willistoni* and *saltans* sister groups (Throckmorton, 1975). The phylogenetic relationships among the *harrow* host species are shown in the Fig.2B.

The *harrow* family shows a patchy distribution pattern throughout the genus *Drosophila*. Even in the *willistoni* and *saltans* groups, where *harrow* seems to be more frequent, its distribution is not continuous and would demand several stochastic loss events to be explained by vertical transmission (VT).

harrow sequence present low divergence between species

The amplicons of roughly 680 bp obtained by the PCR screening were cloned and 4-6 clones of each species were sequenced. Here we analyze 37 new *harrow* transposase partial sequences (GenBank accession numbers GQ165474-GQ165510), together with the equivalent fragments of *Homo3* and the two *Howilli3* copies (*Howilli3.1* and *Howilli3.2*). The overall mean nucleotide divergence (p-distance) between sequences is 0.059. The average divergence values between species vary from 0.026 (*D. sturtevanti* x *D. fumipennis*; *D. sturtevanti* x *Homo3*) to 0.093 (*D. tripunctata* x *D. nebulosa*) and the within species values vary from 0.001 (*D. paulistorum*) to 0.020 (*Howilli3* and *D. neoelliptica*). This low *harrow* sequence divergence presented between species is incongruent with the divergence of the host species genomes.

The analyzed fragment potentially encodes 225 amino acids near the N-terminal and center region of the transposase. As can be seen in Fig.1, it includes the last 34 of the 51 residues of the Zf-BED domain and the first 119 of the 163 residues of the DUF 659 domain of *harrow* transposase. The Zf-BED domain seems to be responsible for the DNA-binding ability of the transposase (Aravind, 2000; Hickman *et al.*, 2005) and the DUF 659 is a transposase domain with unknown function. Ortiz & Loreto (2009) indicated that *Howilli3.1* and *Homo3* sequences are putatively encoding. Additionally, we found potentially encoding sequences in *D. tripunctata*, *D. sturtevanti* and *D. fumipennis*. In *D. prosaltans*, *D. nebulosa*, *D. equinoxialis* and *D. paulistorum* we obtained only non-encoding copies of *harrow*, presenting premature stop codons and/or frameshift. *D. neoelliptica* presented both putatively encoding and non-encoding copies. *Howilli3.2* is also a non-encoding copy, from *D. willistoni*, and was identified *in silico* by Ortiz & Loreto (2009).

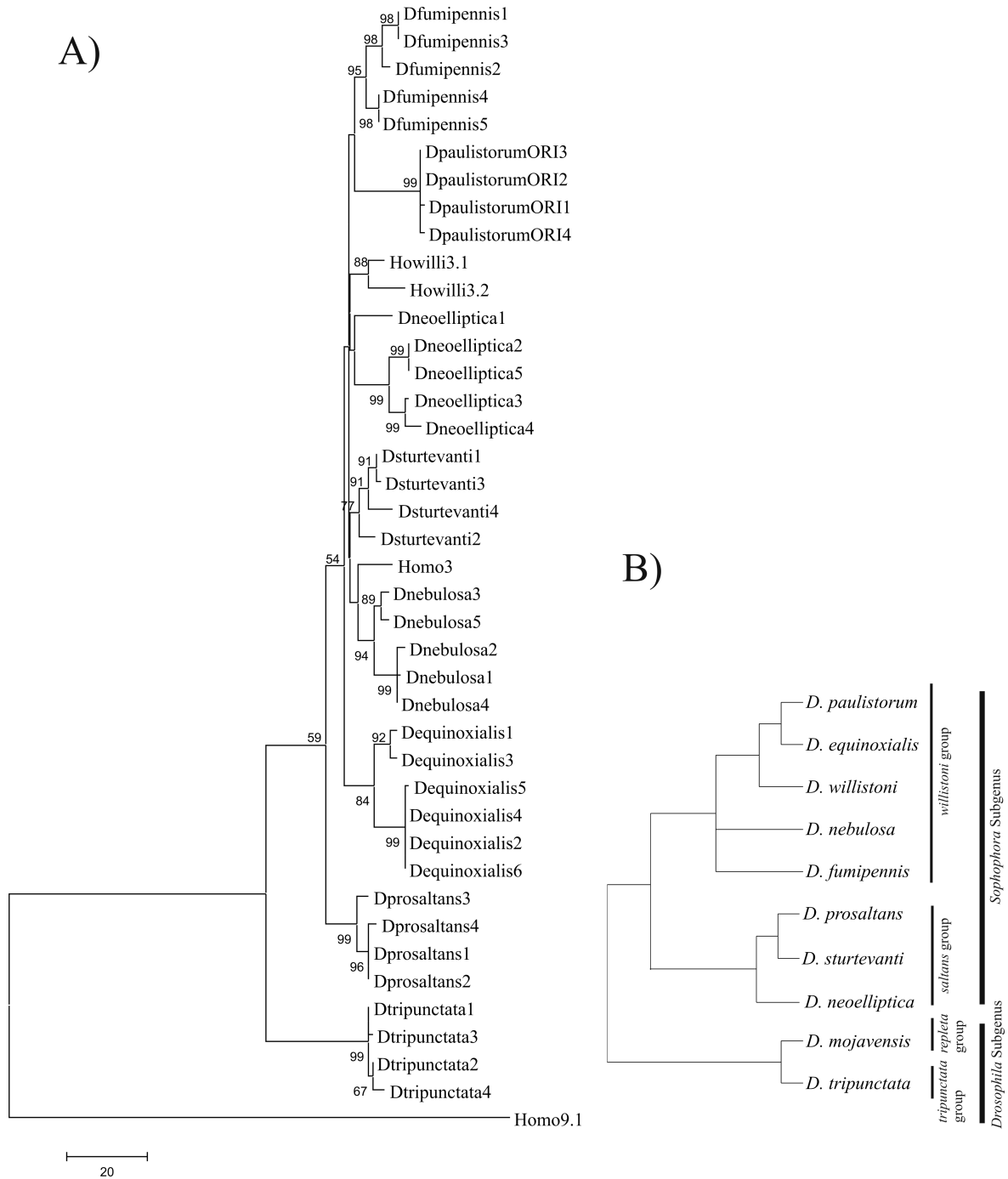


Fig.2. Phylogenetic trees representing A) relationships between *harrow* sequences from different *Drosophila* species, reconstructed through a Maximum Parsimonious bootstrap search. The bootstrap support after 1000 replications is presented near each internal node. Branch lengths are proportional to the scale that reflects the number of substitutions. B) Topology of *harrow*'s host species (species group relationships according to O'Grady *et al.*, 1998; Robe *et al.*, 2005; and L. J. Robe - personal communication).

harrow phylogeny is polytomic

In order to comprehend the evolutionary relationships of these *harrow* sequences, three phylogenetic reconstructing methods were applied: Maximum Parsimony (MP), Neighbor Joining (NJ) and Maximum Likelihood (ML). Fig. 2A shows the MP bootstrap consensus phylogeny, which places as the most basal group of the *harrow* family the *D. tripunctata* sequences clade, although with low bootstrap support. The remaining sequences are divided into two clades: one of them grouping *D. prosaltans* sequences and the other comprising a poorly supported and defined clade, which is almost entirely polytomic. In this polytomy, species specific clades were recovered in all cases, most of them with high bootstrap support. NJ and ML trees can be found in the Figures S2 and S3 of the Supplementary Material. The NJ tree presents a similar topology, positioning the *D. tripunctata* clade as the first to branch off, followed by successive splits of the *D. prosaltans* clade and of the remaining sequences. The ML presents an entire polytomic tree, with the only established relationships being the recovered species specific clades (with the exception of *Dneoelliptica*1 that branches off alone) and the weakly supported clustering of *D. tripunctata* and *D. prosaltans* clades.

harrow dS values are lower than those of host genes

To rule out the chance that the high similarity presented between *harrow* sequences from different species is consequence of high selective constrains, we chose to analyze its divergence on synonymous sites (dS). This approach offers a measure of neutral evolution in the absence of strong codon usage bias (CUB). If a TE was acquired via VT, its dS values are expected to be similar to dS values of host genes. Alternatively, if the TE was horizontally transmitted to the host genome, the TE dS should be significantly lower than host genes dS (Silva & Kidwell, 2000; Ludwig *et al.*, 2008; Bartolomé *et al.*, 2009). Sequences of *Adh*, *Amd* and *Ddc* nuclear genes were obtained in GenBank and used as examples of VT evolution to compare with *harrow* sequences.

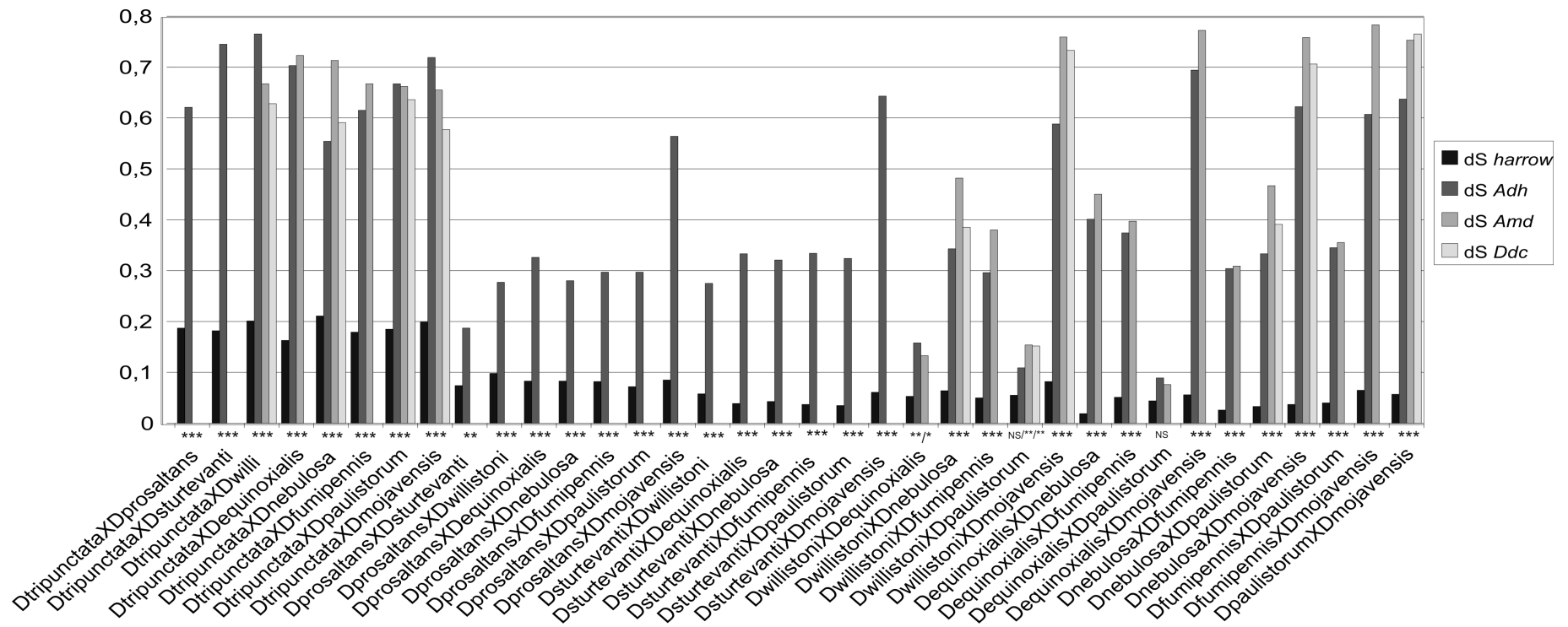


Fig.3. Mean dS values for *harrow*, *Adh*, *Amd* and *Ddc* in species pairwise comparisons. Results of Fisher's exact test: (***) p<0.001; (**) p<0.01; (*) p<0.05; (NS) nonsignificant.

harrow sequences of each species were grouped and the mean dS values were obtained for each pair of species. The differences of *harrow* and host genes dS values were examined using the Fisher's Exact Test. As can be seen in Fig. 3, in all comparisons, the *harrow* dS were significantly lower than those of host genes, except between *D. willistoni* and *D. paulistorum* (for *Adh*), as well as between *D. paulistorum* and *D. equinoxialis* (*Adh* and *Amd*). These data sustain the *harrow* HT hypothesis.

However, as the variation on dS rates among genes has been found to be negatively correlated to the intensity of natural selection on synonymous codon usage (Shields *et al.*, 1988; Sharp & Li, 1989), we took in consideration the CUB level of *harrow* and the three chosen host genes. The CUB was measured by two indices: the effective number of codons (ENC; Wright, 1990) and the codon bias index (CBI; Morton, 1993).

As can be seen in Table 1, the level of *harrow* CUB is lower than those of the host genes, since *harrow* presents the highest ENC and the lowest CBI. Among the host genes, *Amd* exhibits the lowest CUB. In this context, *Amd* is the best fit gene for the dS analyses since it does not present a high CUB (which would lead to low dS values and possibly result in underestimation of HT), and, in the other hand, it presents higher CUB than *harrow* (avoiding overestimation of HT). For these same reasons, *Amd* has been previously used for dS comparisons to infer HT (Ludwig *et al.*, 2008). Nevertheless, *Adh* and *Ddc* were included in dS analyses as well, taking a more conservative position in the HT inference.

harrow sequences are evolving under purifying selection

Evolutionary pressures on proteins are often quantified by the ratio of substitution rates at non-synonymous and synonymous sites (dN/dS). To verify if *harrow* sequences are under evolutionary constrains the mean dN/dS ratio was measured for each pair of species which had potentially encoding sequences. The mean dN/dS ratio of *harrow* sequences is 0.308 and the codon-based Z Test, rules out the neutrality hypothesis, indicating that *harrow* sequences are evolving at some purifying selection to maintain its transposition properties. However, as

can be seen in Fig. 4, *harrow*'s dN/dS ratio is higher than the host genes, indicating that the TE sequences are under lower selective constraints than the other genes.

Table 1. Mean values of ENC and CBI for *harrow* sequences and for three host genes. ENC may vary from 21 for maximum codon bias and 61 for minimum codon bias; CBI value of 0 corresponds to no bias and a value of 1 corresponds to maximum bias. In parenthesis is the number of sequences used in this analysis.

	ENC	CBI
<i>harrow</i> (40)	59.353	0.266
<i>Adh</i> (9)	43.218	0.511
<i>Amd</i> (7)	50.812	0.392
<i>Ddc</i> (5)	46.968	0.455

Note: (ENC) Effective Number of Codons; (CBI) Codon Bias Index.

Estimated divergence time among harrow sequences

We have computed the estimated divergence time between the clades of *harrow* phylogeny. We estimate the divergence time between the *D. tripunctata* clade and the one comprising the rest of the sequences as around 6 Myr. Within this last clade, the estimated time between the *D. prosaltans* sequences and the others is about 2.6 Myr. Within the polytomic clade, the overall divergence time is 1.5 Myr. The divergence times between *harrow* sequences should reveal the times when HT events have occurred. However, as these phylogenetic clades are weakly supported, these estimative do not necessarily provide the direct pathway and direction of the HT events.

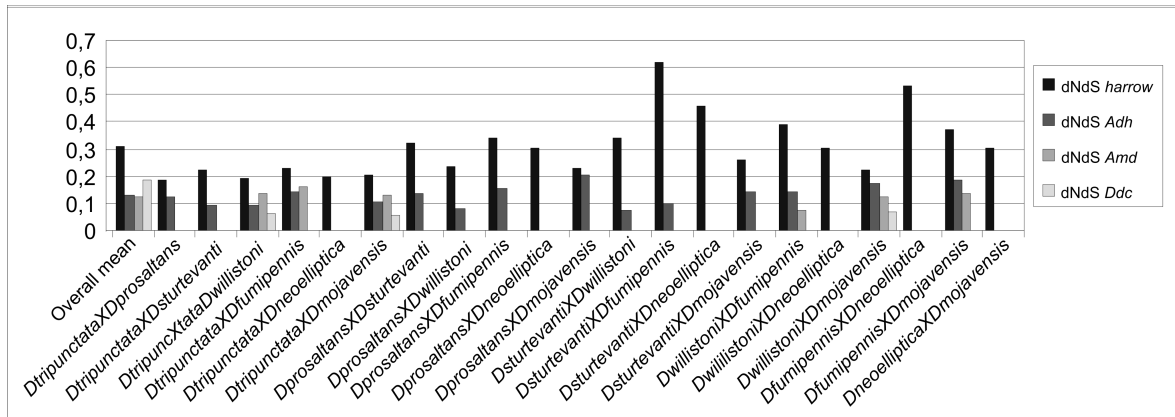


Fig.4. Comparative analyzes of dN/dS values of *harrow* and host genes. Only potentially encoding sequences were used to obtain *harrow*'s pairwise dN/dS mean value (for exception see Material and Methods).

Discussion

In the last two decades, an increasing number of HT events have been identified, showing that this phenomenon is not so rare in eukaryotes as previously thought. HT of a functional element into a new host genome is a way for them to escape from extinction, being an essential step on the life cycle of TEs (Lohe *et al.*, 1995). Recently, Bartolomé *et al.* (2009) compared dS substitutions among 141 TE families and 10150 nuclear genes of *D.melanogaster*, *D. simulans* and *D.yakuba* genomes. They showed that a large number of TE families have dS values lower than expected if acquired by VT, suggesting that HT is a widespread phenomenon across *Drosophila* evolution.

As extensively reviewed by Loreto *et al.* (2008), three different kinds of evidence are usually applied to infer HT of TEs: 1) Patchy distribution of a TE across a group of species; 2) Incongruence between host and TE phylogenies; and 3) High sequence similarity between TEs of distantly related species. Silva *et al.* (2004) proposed that the stronger cases of HT are those confirmed by these three sorts of evidence.

The patchy distribution pattern of *harrow* trough out the *Drosophila* genus is clearly verified by the presence of *harrow* sequences in only 6.5% of the studied species, without phylogenetic pattern. Instead, it was found in only two of 35

species tested from subgenus *Drosophila*, each one belonging to a different group, and in eight of the 25 species tested from subgenus *Sophophora*. A Dot Blot analysis of species from different *Drosophila* groups was performed to check the presence of *harrow* sequences (data not shown). We verify that the species tested that did not present PCR amplification of *harrow* also did not present hybridization signal by Dot Blot, refusing possible PCR amplification absence by miss annealing of the primers

Since the relationships between the species specific clades of *harrow*'s phylogeny are poorly supported, there are no evident and reliable incongruence between host and TE phylogenies. However, this lack of resolution is due to the low divergence presented between *harrow* sequences from different species. As we used different phylogenetic methods, we consider that the brushy pattern observed results from multiple simultaneous HT events.

Unquestionably, the high similarity between *harrow* sequences is the most outstanding evidence that these sequences were spread through events of HT. The high similarity between *harrow* sequences from species belonging to different subgenera is particularly surprising since the split between subgenera *Drosophila* and *Sophophora* has been estimated as approximately 63 and 55 Mya (Tamura *et al.*, 2004, L. J. Robe – personal communication, respectively). There is no indication that this high conservation is due to selection. Moreover, the fragment of the transposase analyzed belongs to its N-terminal and central region, avoiding the C-terminal portion which is the most conserved one between *hAT* transposases (Calvi *et al.*, 1991, Rubin *et al.*, 2001).

An alternative approach implemented, dS comparisons between TE and nuclear genes, also supports the HT inference as *harrow* dS values are much lower than those of the nuclear genes. Bartolomé *et al.*, (2009) recently stated that the best evidence of HT between two species is the detection of much lower level of divergence at neutral sites in the TE sequence than the average neutral divergence between these species genomes. Despite using only three nuclear genes, our results are consistent with the ones obtained by Bartolomé *et al.*, (2009), supporting the notion that HT events between eukaryotes are much more common than previously thought.

The mechanisms for HTs remain obscure, although these transfer events require the occurrence of common premises, such as geographical, temporal, and ecological overlap between donor and recipient species. In the effort of understanding the possible HT scenarios, we have plotted in a map the geographic coordinates where host species have already been collected, also showing where they are probably able to inhabit (Fig. 5).

There is overlap distribution between most of the host species, within the Neotropics. The distribution of members from all the four species groups (*tripunctata*, *mojavensis*, *saltans* and *willistoni*) overlaps near the borders between the United States of America and Mexico, as well in the Central America, allowing species contact and the occurrence of HT events. The overlapping area of members of the *willistoni* and *saltans* species groups is much wider, spreading through most Central and South America.

The patchy distribution of *harrow* suggests previous *harrow* HT event, from an unknown source, to one or more *Drosophila* species. This donor species could be a genus *Drosophila* species that was not included in our study, could be a more distantly related Drosophilidae species or even any fruitfly. More recently, probably recurrent HT events occurred in Neotropical lands, especially between or within *saltans* and *willistoni* species groups, resulting in the observed polytomy. The recently and almost simultaneous transfer of *harrow* prevented the accumulation of substitutions, resulting in short branch lengths, where it is impossible to picture the donor and recipient of each HT event. A similar transfer pattern was observed by Silva & Kidwell (2000) for *P* element, also involving the *saltans* and *willistoni* groups and at approximately the same estimated time.

Although there are well supported evidences sustaining the HT inference of *harrow* between the host species, other possibilities are yet to be considered in some particular situations. For example, for the *willistoni* subgroup species, introgression events could also be invoked as a potential TE spread mechanism, since it has been reported that under some conditions these species could produce viable/fertile hybrids (Cordeiro & Winge, 1995). This is a plausible alternative for this particular species group.

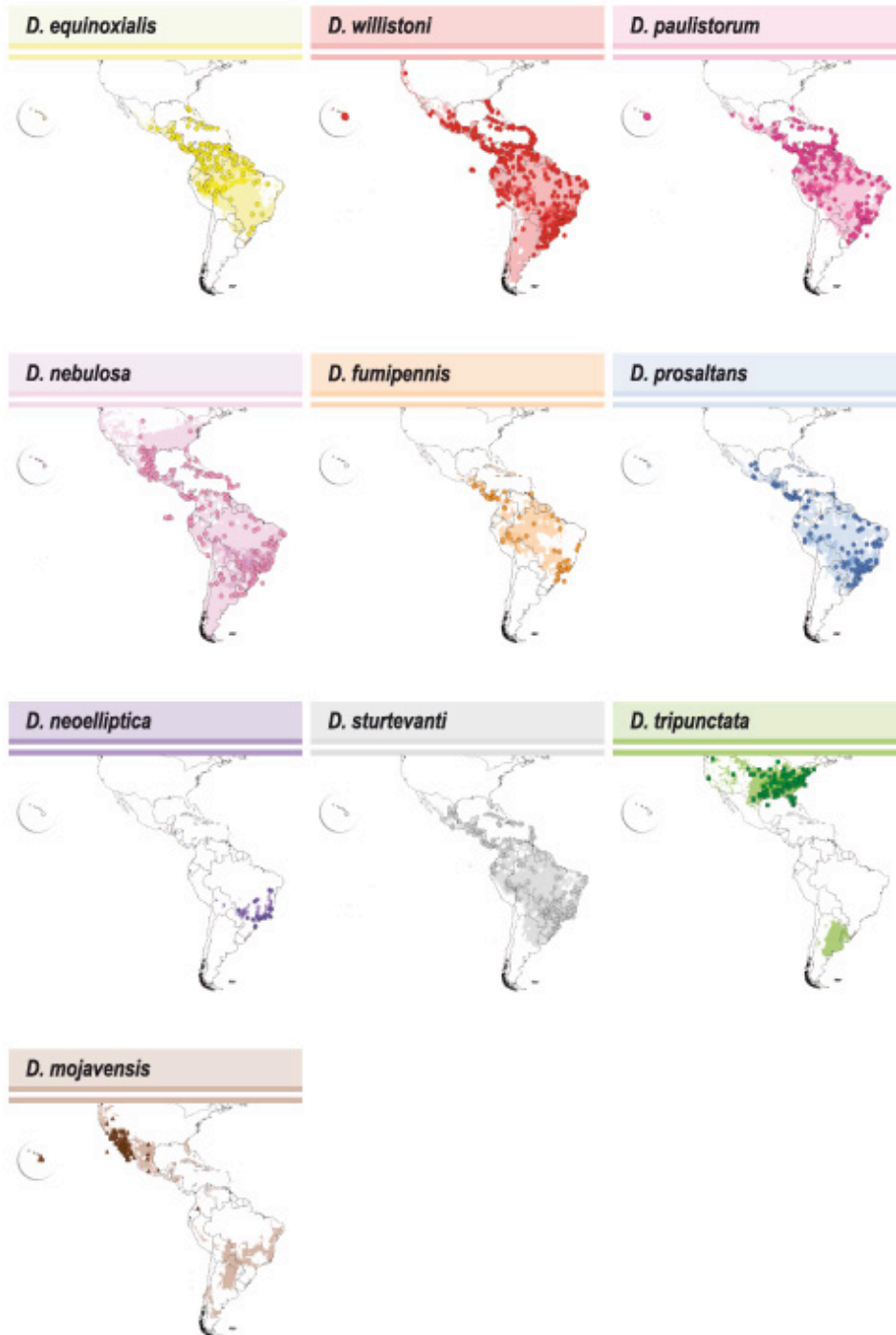


Fig.5. Geographic distribution of *harrowi*'s host species: Geographic coordinates of collection points obtained in TaxoDros (Bächli, 2008) were plotted on maps with Diva-GIS 5.2 software (Hijmans *et al.* 2005).

Another possibility is that *harrow* could have spread throughout genus *Drosophila* through VT, and its patchy distribution could be consequence of subsequent element loss by most species. However, we consider this as a less parsimonious scenario when applied to the entire group of studied species and it does not explain the high similarity among sequences. In the other hand, the stochastic loss should be considered more carefully for the members of *willistoni* subgroup as the difference of dS values form *harrow* and the host genes were not always significant. Alternatively, it is possible that there is an unknown important biological process that leads, throughout the evolution, high dS conservation among TE sequences. However, considering the current biological knowledge and the results presented in this study, we suggest that HT events have indeed played a special role in the evolutionary history of *harrow* family.

Material and Methods

Characterization of harrow family

We used the hierarchical system, suggested by Whicker *et al.* (2007) to classify the studied TE. All nucleotide sequences from *homo* family (proposed by Ortiz & Loreto, 2009) were aligned using MUSCLE 3.7 (Edgar, 2004). The most conserved region was used to access the nucleotide divergence. A phylogenetic tree was inferred by Neighbor-Joining using Kimura 2 Parameters model (Kimura, 1980) executed by Mega 4 software (Tamura *et al.*, 2007).

The NetGene2 Server (Brunak *et al.*, 1991; Hebsgaard *et al.*, 1996) (<http://www.cbs.dtu.dk/services/NetGene2>) was used to detect the presence of introns. Additionally, the transposase ORF was identified by the ORF Finder tool (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>).

The amino acid sequences were deduced by GeneDoc 2.6.001 program (Nicholas & Nicholas, 1997). Conserved domains were identified in the complete *Howilli3.1* transposase aminoacid sequence by the Conserved Domain Database (CDD; Marchler-Bauer *et al.*, 2007), available at NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>).

Fly stocks maintenance and DNA management

The flies were maintained in the laboratory by mass crosses and reared on corn flour culture medium (Klein *et al.*, 1999) at 17 °C. The investigated species are shown in Supplementary Table 1. DNA samples were extracted according to Sassi *et al.* (2005).

PCR screening, cloning and sequencing

Primers were designed with Oligo 4.1 software (Rychlik, 1992) based on the sequences of *Homo3* e *Howilli3.1* described by Ortiz & Loreto (2009). The primers (*hAT1F* sense GCTTTTGYTGGAGGCACTTTAC and *hAT1R* antisense GTAYGTGCTTTTTTCGCCAA) amplify a 728 bp fragment of the transposase gene, of which the core 676 bp was used in the analyses. The components of the 50 µl reaction mixture were 100 ng DNA, 1 U Taq polymerase, 5 µl 10× reaction buffer supplied by the manufacturer (Invitrogen, Carlsbad, CA), 200 µM of each nucleotide, 20 pmol of each primer and 1.5 mM MgCl₂. Amplification parameters were 96 °C for 2 min, 35 cycles at 96 °C for 30 s, 60 °C for 45 s and 72 °C for 45 min, followed by an extension cycle at 72 °C for 7 min. PCR products were cloned into the PCR4.1 cloning vector using the TOPO Cloning® kit (Invitrogen, Carlsbad, CA, USA). DNA sequencing was carried out directly from the purified amplicons in MegaBace 500 automatic sequencer. The dideoxy chain-termination reaction was implemented with the use of the DYEnamicET kit (GE Healthcare, Little Chalfont, Bucks, U.K.). Both DNA chains were sequenced. Consensus sequences for each clone were obtained by the electropherogram analyses using the GAP 4 software of the Staden Package (Staden, 1996).

Evolutionary analyzes

The nucleotide sequences were aligned using the ClustalW software (Higgins *et al.*, 1994) according to the system default parameters. Mega 4 software (Tamura *et al.*, 2007) was used to compute overall and between species average *harrow* nucleotide divergences trough p-distance model (using pairwise deletion).

Phylogenetic trees were inferred by Neighbor-Joining (NJ) and Maximum Parsimony (MP) methods using the Mega 4 software (Tamura *et al.*, 2007) and by Maximum Likelihood (ML) method using PAUP software (Swofford, 2003). In MP analysis, the best tree was searched using 500 random stepwise addition trees followed by close-neighbor interchange rearrangements (search level 1). For the NJ tree, the Tamura-3-Parameters evolutionary model (Tamura, 1992) was followed. In ML analysis, the GTR evolutionary model (Rodriguez *et al.*, 1990) was used as pointed by the AIC Test (Akaike, 1974) executed in ModelTest (Posada & Crandall, 1998). In all analyses, 1000 replications of bootstrap were implemented.

Comparatives analysis of dS and CUB

A codon alignment for *harrow* sequences was used to estimate the mean dS values, the mean number of synonymous sites (S) for each pair of species and the level of CUB for each of the analyzed genes. In order to include all sequences in this analysis, we slightly edited some sequences repairing insertions and premature stop codons.

Sequence of three host genes, *Adh*, *Amd* and *Ddc*, were used to compare with *harrow* dS and CUB with the purpose of testing the HT hypothesis. Sequences obtained in GenBank and their accession numbers are: *Adh* – *D. tripunctata* (AY081440), *D. prosaltans* (AF045119), *D. sturtevantii* (AB026535), *D. equinoxialis* (U95269), *D. willistoni* (U95264), *D. nebulosa* (U95275), *D. fumipennis* (AF264075), *D. paulistorum* (AB026529) and *D. mojavensis* (XM_002002894.1); *Ddc* – *D. tripunctata* (AF324964), *D. nebulosa* (AF293742) and *D. paulistorum* (AF293744); *Amd* – *D. tripunctata* (AF293728), *D. willistoni* (AF293730) and *D. nebulosa* (AF293717). Sequences of *Ddc* gene for *D. willistoni* and *Ddc* and *Amd* genes for *D. mojavensis* were obtained directly on these species available genomes (<http://flybase.org/blast>; Wilson *et al.*, 2008): *Ddc* – *D. willistoni* (Genome scf2_1100000004585: 7091604, 7093603), *D. mojavensis* (scf_6500: 6787407, 6789406); *Amd* – *D. mojavensis* (scf_6500: 6792722, 6794721). Sequences of *Amd* for *D. equinoxialis*, *D. fumipennis* and *D. paulistorum* (semi species *Orinocana*) were obtained by personal communication with L. J. Robe.

The dS and S were estimate using the Nei & Gojobori (1986) method, applied by Mega 4 software (Tamura *et al.*, 2007). The Fisher's Exact Test (1-tailed), applied by DNASP 4.0 software (Rozas *et al.*, 2003), was used to verify if *harrow* dS values were statistically lower than those presented by the host genes.

The CUB level of *harrow* and host genes were checked by the effective number of codons (Nc; Wright, 1990) and the codon bias index (CBI; Morton, 1993) computed by DNASP 4.0 software (Rozas *et al.*, 2003).

Divergence times among harrow sequences

These estimative were conducted according to the formula $T=k/2r$ (Graur & Li, 2000), in which T is the divergence time between species, k is the divergence between TE sequences (dS) in synonymous sites and r is the evolutionary rate. We used a synonymous substitution rate of 0.016 substitutions per site per million years, as calculated for *Drosophila* genes with low codon usage bias (Sharp & Li, 1989).

Estimates of dN/dS

Codon alignments of *harrow* sequences and host genes were used to estimate the dN/dS ratio, using the Nei & Gojobori (1986) method assisted by the Mega 4 software (Tamura *et al.*, 2007). Only *harrow* potentially encoding fragments were used with one exception. None of *D. prosaltans* sequences are potentially encoding, however we considered valuable including at least one of them in this analyzes since they form a distinct branch of the phylogenetic tree. Thus, Dprosaltans3 sequence was used in this analysis, with its stop codon coded as absent data. The mean *harrow* dN/dS value for each pair of species was compared with pairwise dN/dS of the three host species genes. A codon based Z-Test, applied by Mega 4 (Tamura *et al.*, 2007), was performed to verify if *harrow* sequences were under any selective constrains.

Host species distribution

The Diva-GIS 5.2 software (Hijmans *et al.* 2005) was used for plotting in a map the geographic coordinates where host species have already been collected. The information about the species distribution was obtained in TaxoDros database

(Bächli, 2008). This software not only positions the coordinates but also makes a prediction of where else this species is probably able to inhabit.

References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transaction on Automatic Control* **19**: 716–723.

Aravind, L. (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem Sc* **25**: 421-423.

Arensburger, P., Kim, Y.J., Orsetti, J., Aluvihare, C., O'Brochta, D.A. and Atkinson, P.W. (2005) An active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. *Genetics* **169**: 697-708.

Bächli, G. (2008) TaxoDros: The Database on Taxonomy of Drosophilidae, v.1.03, Database 2008/12. <http://taxodros.unizh.ch/>.

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

Bigot, Y.M., Corinne, A. and Georges, P. (1996) Computer analyses reveal a *hobo*-like element in the nematode *Caenorhabditis elegans*, which presents a conserved transposase domain common with the TC1-*Mariner* transposon family. *Gene* **174**: 265-271.

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.

Calvi, B.R., Hong, T.J., Findley, S.D. and 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.

Capy, P., Bazin, C., Higuët, D. and Langin, T. (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197pp.

Coates, C.J., Johnson, K.M., Perkins, H.D., Howells, A.J., O'Brocta, D.A. and Atkinson, P.W. (1996) The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica* **97**: 23-31.

Cordeiro, A.R. and Winge, H. (1995) Levels of evolutionary divergence of *Drosophila willistoni* sibling species. In: Levine L, ed. Genetics of natural populations: The continuing importance of Theodosius Dobzhansky. New York: Columbia University Press, 262–280.

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.

Gómez-Gómez, E., Anaya, N., Roncero, M.I.G. and Hera, C. (1999) *Folyt1*, a New Member of the *hAT* Family, Is Active in the Genome of the Plant Pathogen *Fusarium oxysporum*. *Fungal Genet Biol* **27**: 67-76.

Graur, D. and Li, W.-H. (2000) Fundamentals of Molecular Evolution, 2nd ed. Sinauer Associates, Sunderland, MA.

Handler, A.M. and Gomez, S.P. (1996) The *hobo* Transposable Element Excises and Has Related Elements in Tephritid Species. *Genetics* **143**: 1399-1947.

Handler, A.M. and Gomez, S.P. (1997) A new *hobo*, *Ac*, *Tam3* transposable element, *hopper*, from *Bactrocera dorsalis* is distantly related to *hobo* and *Ac*. *Gene* **185**: 133–135.

Hebsgaard, S.M., Korning, P.G., 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.

Hickman, A.B., Perez, Z.N., Zhou, L.Q., Musingarimi, P., Ghirlando, R., Hinshaw, J.E., Craig, N.L. and Dyda, F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol* **12**: 715–721.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.

Hijmans, R.J., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., Cruz, M., Barrantes, I. and Rojas, E. (2005) Diva-GIS version 5.2. Manual. Available at <http://www.diva-gis.org>.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Res* **110**: 462-467.

Kempken, F. and Kück, U. (1996) *Restless*, an active Ac-like transposon from the fungus *Tolypocladium inflatum*: structure, expression, and alternative RNA splicing. *Mol Cell Biol* **16**: 6563–6572.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111-120.

Klein, C.C., Essi, L., Golombieski, R.M. and Loreto, E.L.S. (1999) Disgenesia do híbrido em populações naturais de *Drosophila melanogaster*. *Ciencia e Natura* **21**: 7-20.

Koga, A., Shimada, A., Shima, A., Sakaizumi, M., Tachida, H. and Hori, H. (2000) Evidence for Recent Invasion of the Medaka Fish Genome by the *Tol2* Transposable Element. *Genetics* **155**: 273-281.

Lohe, A.R., Moriyama, E.M., Lidholm, D.A. and Hartl, D.L. (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner* like transposable elements. *Mol Biol Evol* **12**: 62-72.

Loreto, E.L.S., Carareto, C.M.A. and Capy, P. (2008) Revisiting horizontal transfer of transposable element in *Drosophila*. *Heredity* **100**: 545-554.

Ludwig, A., Valente, V.L.S. and Loreto, E.L.S. (2008) Multiple Invasions of *Errantivirus* in the Genus *Drosophila*. *Insect Molecular Biology* **17**: 113-124.

Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Krylov, D., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Thanki, N., Yamashita, R.A., Yin, J.J., Zhang, D. and Bryant, S.H. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**: D237-240.

McClintock, B. (1948) Mutable loci in maize. Carnegie Inst Wash, Year book, 47: 155-169.

Morton, B.R. (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.

Nicholas, K.B. and Nicholas, H.B.J. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. <http://www.psc.edu/biomed/genedoc>.

O'Grady, P.M., Clark, J.B. and Kidwell, M.G. (1998) Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol Biol Evol* **15**: 656-664.

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

Pinkerton, A.C., Whyard, S., Mende, H.A., Coates, C.J., O'Brochta, D.A. and Atkinson, P.W. (1999) The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol Biol* **8**: 423–434

Posada, C. and Crandall, K.A. (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Ray, A., Pagan, H.J.T., Thompson, M.L. and Stevens, R. (2007) Bats with *hATs*: Evidence for Recent DNA Transposon Activity in Genus *Myotis*. *Mol Biol Evol* **24**: 632-639.

Robe, L.J., Valente, V.L.S., Budnik, M. and Loreto, E.L.S. (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.

Rodriguez, F., Oliver, J.L., Marin, A. and Medina, J.R. (1990) The general stochastic model of nucleotide substitution. *J Theor Biol* **142**: 485–501.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, 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, A.A. (2001) Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158**: 949–957

Rychlik, W. (1992) Oligo 4.1 Primers Analysis Software. National Biosciences, Inc., Plymouth, Minn.

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

Silva, J.C. and Kidwell, M.G. (2000) Horizontal transfer and selection in the evolution of *P* elements. *Mol Biol Evol* **17**: 1542-1557.

Silva, J.C., Loreto, E.L. and Clark, J.B. (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.
- Sharp, P.M. and Li, W.H. (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol* **28**: 398-402.
- Shields, D.C., Sharp, P.M., Higgins, D.G. and Wright, F. (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol* **5**: 704-716.
- Staden, R. (1996) The Staden sequence analysis package. *Mol Biotechnol* **5**: 233-241.
- Subramanian, R.A., Arensburger, P., Atkinson, P. and O'Brochta, D.A. (2007) Transposable Element Dynamics of the *hAT* Element *Herves* in the Human Malaria Vector *Anopheles gambiae* s.s. *Genetics* **176**: 2477-2487.
- Swofford, D.L. (2003) PAUP: Phylogenetic Analysis using Parsimony (and other methods). Version 4. Sinauer Associates, Massachusetts.
- Tamura, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Mol Biol Evol* **9**:678-687.
- Tamura, K., Subramanian, S. and Kumar, S. (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* **21**: 36-44.
- 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.
- Throckmorton, L.H. (1975) The phylogeny, ecology and geography of *Drosophila*. In: King, R. C. (ed) Handbook of Genetics. Plenum, New York, pp 421-469.

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

Warren, W.D., Atkinson, P.W. and O'Brochta, D.A. (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, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. and Schulman, A.H. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973-982.

Wilson, R.J., Goodman, J.L., Strelets, V.B. and The FlyBase Consortium (2008) FlyBase: integration and improvements to query tools. *Nucl Acids Res* **36**: D588-D593.

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

Acknowledgements

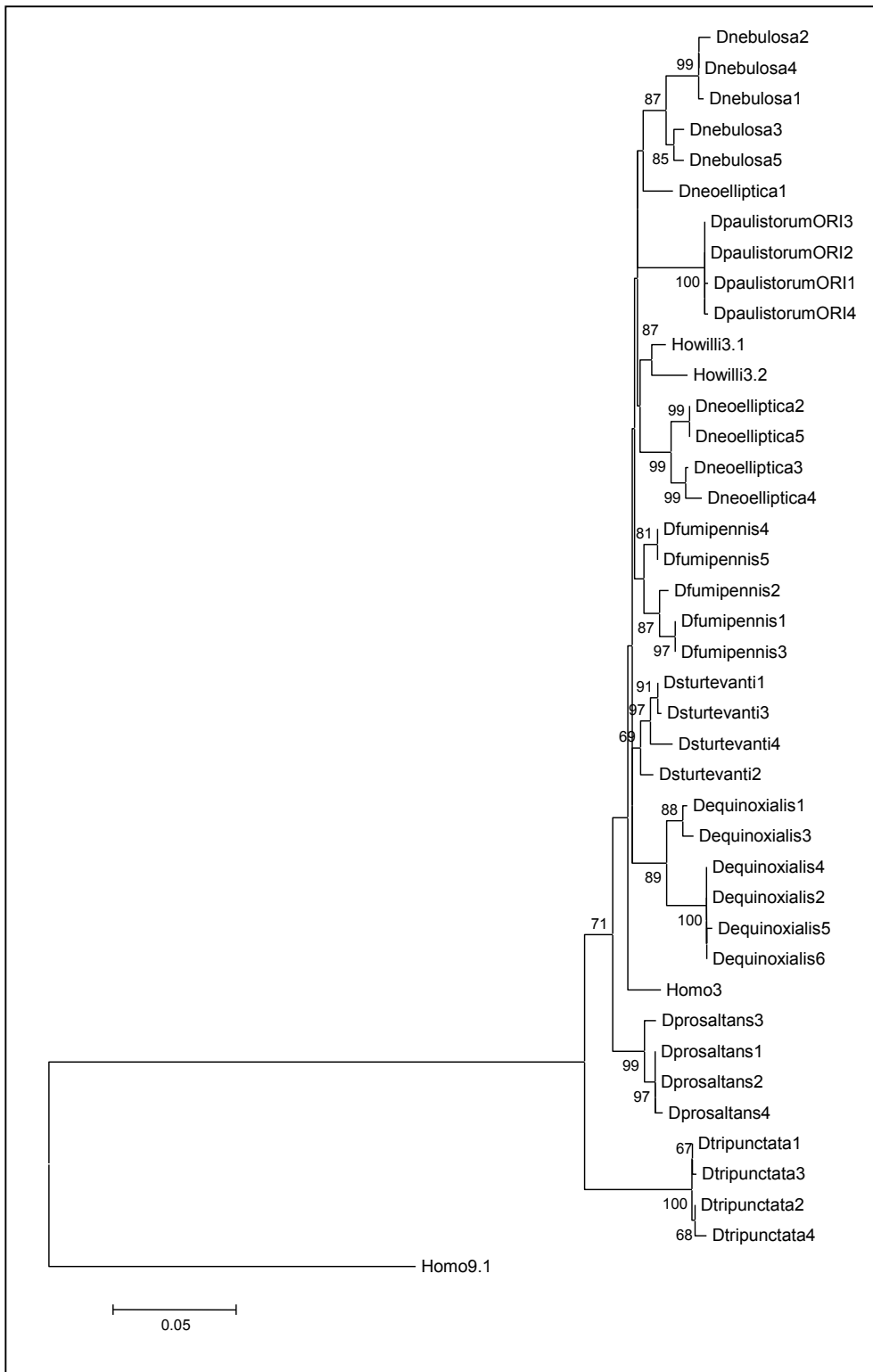
This work was supported by grants and fellowships from CNPq. We are grateful to Dr. Lizandra J. Robe, Dr. Fabiana Herédia, Dr. Cláudia M. Carareto for the valuable help and critical comments which contributed significantly to the improvement of this paper. We also thank Dr. Lizandra J. Robe for providing us with *Amd* sequences and personal communication of unpublished data.

Supplementary Material

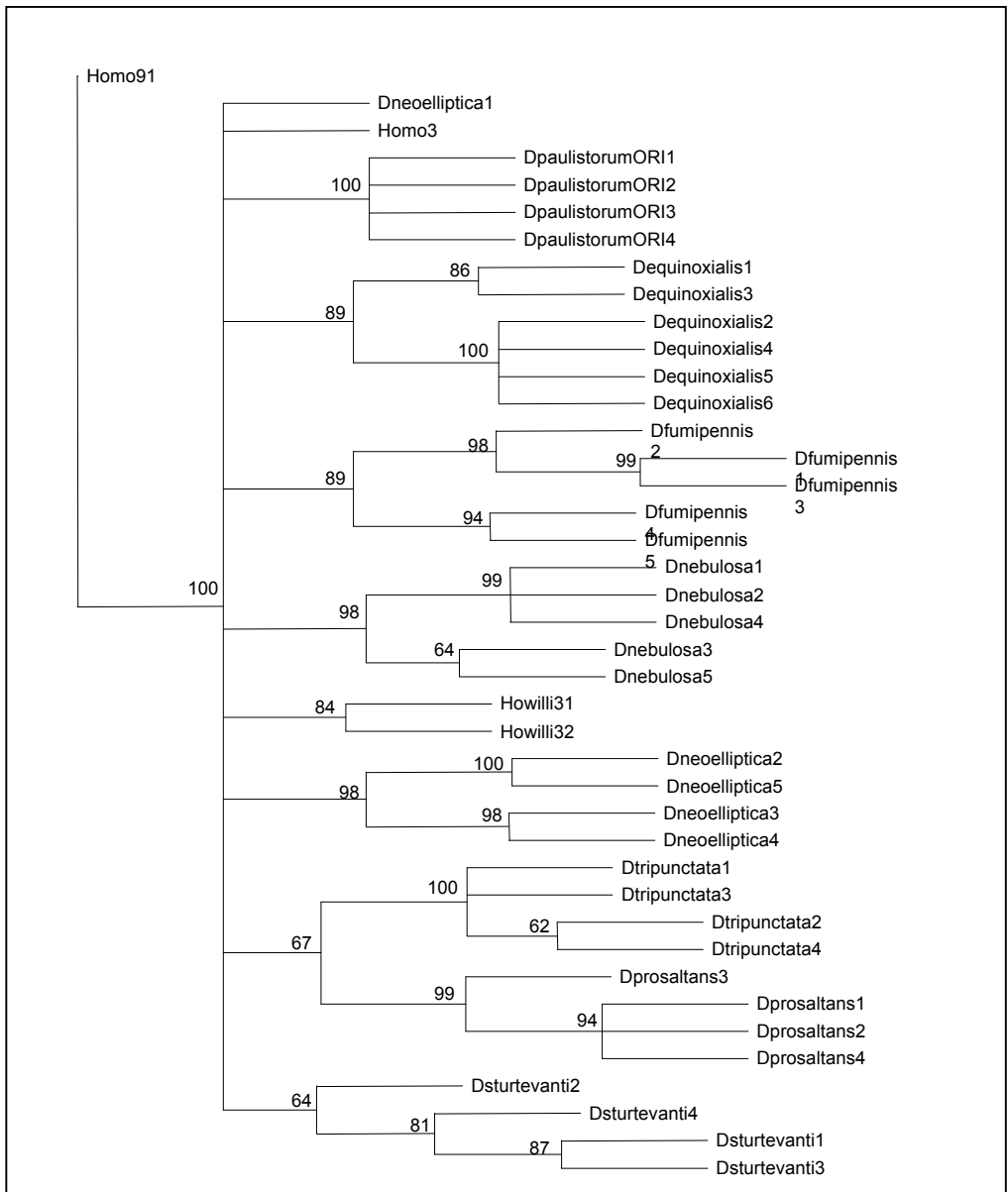
Supplementary Table I: Species studied and their PCR screening results.

Genus	Subgenus	section	Group	Species	PCR amplification	
<i>Drosophila</i>	<i>Drosophila</i>	<i>quinaria</i> <i>tripunctata</i>	<i>guarani</i>	<i>D. ornatifrons</i>	-	
				<i>D. subbadia</i>	-	
				<i>D. guaru</i>	-	
			<i>guaramuru</i>	<i>D. griseolineata</i>	-	
				<i>D. maculifrons</i>	-	
				<i>tripunctata</i>	<i>D. nappae</i>	-
					<i>D. paraguayensis</i>	-
					<i>D. crocina</i>	-
				<i>D. paramediostriata</i>	-	
				<i>D. tripunctata</i>	+	
				<i>D. mediodifusa</i>	-	
				<i>D. bandeirantorum</i>	-	
				<i>D. mediopictoides</i>	-	
			<i>cardini</i>	<i>D. cardinoides</i>	-	
				<i>D. neocardini</i>	-	
				<i>D. polymorpha</i>	-	
				<i>D. procardinoides</i>	-	
				<i>D. arawakana</i>	-	
				<i>D. pallidipennis</i>	-	
				<i>D. ornatipennis</i>	-	
				<i>D. immigrans</i>	-	
		<i>D. funebris</i>		-		
		<i>viriis repleta</i>		<i>D. gasici</i>	-	
			<i>D. brncici</i>	-		
			<i>D. gaucha</i>	-		
			<i>D. pavani</i>	-		
			<i>D. hydei</i>	-		
			<i>D. mercatorum</i>	-		
			<i>D. mojavenensis</i>	GENOME		
			<i>annulimana</i>	<i>D. annulimana</i>	-	
				<i>D. schineri</i>	-	
			<i>canalineae</i> <i>flavopilosa</i>	<i>D. canalinea</i>	-	
				<i>D. cestri</i>	-	
			<i>virilis</i> <i>robusta</i> <i>melanogaster</i>	<i>D. incompta</i>	-	
		<i>D. virilis</i>		-		
		<i>D. robusta</i>		-		
		<i>D. melanogaster</i>		-		
		<i>D. simulans</i>		-		
		<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. malerkotliana</i>	-			
		<i>obscura</i> <i>saltans</i>	<i>D. pseudoobscura</i>	-		
			<i>D. prosaltans</i>	+		
			<i>D. saltans</i>	-		
			<i>D. neoelliptica</i>	+		
			<i>D. sturtevantii</i>	+		
		<i>willistoni</i>	<i>D. sucinea</i>	-		
<i>D. nebulosa</i>	+					
<i>D. willistoni</i>	GENOME					
<i>D. paulistoum</i> (semi-species Orinocana)	+					
<i>D. insularis</i>	-					
<i>D. tropicalis</i>	-					
<i>D. equinoxialis</i>	+					
<i>D. capricorni</i>	-					
<i>D. fumipennis</i>	+					
<i>D. busckii</i>	-					
<i>Z. indianus</i>	-					
<i>Z. tuberculatus</i>	-					
<i>Scaptodrosophila</i>	<i>Dorsilopha</i>	<i>S. latifasciaeformis</i>	-			
		<i>S. lebanonensis</i>	-			

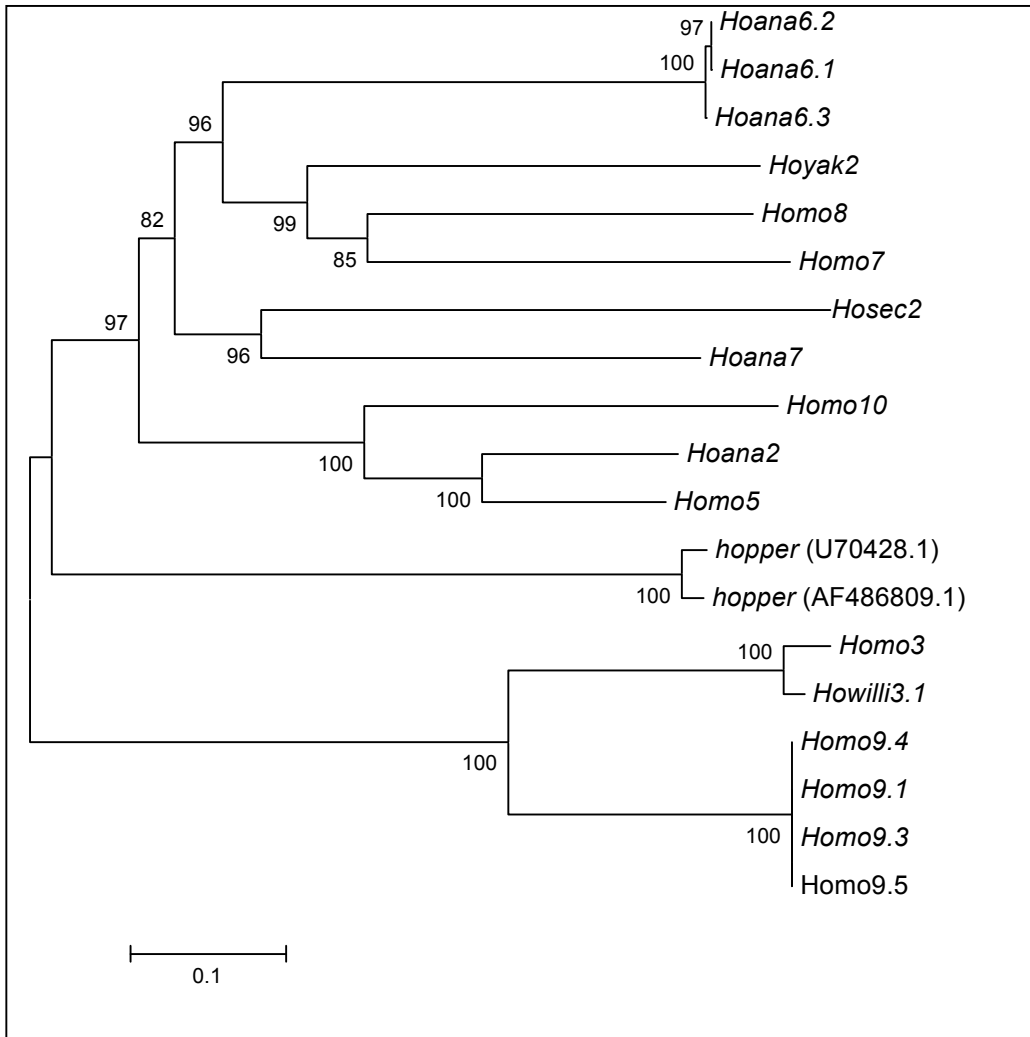
Notes: (-) no amplicons obtained; (+) amplicons obtained; (GENOME) sequence previously found in the available genome by Ortiz and Loreto (2009).



Supplementary Figure S1: Neighbor-Joining bootstrap consensus tree constructed with the used of Tamura 3 Parameters model (Tamura, 1992). The bootstrap support after 1000 replication is presented near each internal node, having been omitted when lower than 50.



Supplementary Figure S2: Maximum Likelihood bootstrap consensus tree following the GTR evolutionary model (Rodriguez *et al.*, 1990), as pointed out by ModelTest (Posada and Crandall, 1998). The bootstrap support after 1000 replication is presented near each internal node.



Supplementary Figure S3: Bootstrap consensus phylogenetic tree of the TEs previously grouped as *homo* family (Ortiz and Loreto, 2009) based on alignment of the most conserved 1500 bp transposase region. Neighbor-Joining method with Kimura 2 Parameters model (Kimura, 1980) and 1000 bootstrap replicates was implemented for tree reconstruction.

Supplementary Table S2: Mean nucleotide divergence (d), computed by p-distance, of sequences previously grouped as *homo* family by Ortiz and Loreto (2009), between and within the three separated groups obtained in our phylogeny of the 1500 bp most conserved transposase alignment region .

	Group 1	Group 2	Group 3
Group 1	0.161		
Group 2	0.531	0.030	
Group 3	0.535	0.509	0.426

CAPÍTULO III

Discussão

Transferência Horizontal

Se transposons de DNA estão condenados à extinção dentro de cada genoma hospedeiro eucarioto, a única maneira de eles terem se perpetuado e espalhado como fizeram por tantos genomas seria através de regularmente passarem por processos de HT para novos genomas hospedeiros (Robertson, 2002). Com essa frase Robertson (2002) inicia um trecho de seu capítulo no livro *Mobile DNA II* (Craig *et al.*, 2002), onde mostra como ele e vários outros pesquisadores chegaram a esta conclusão única. O autor mostra, por exemplo, que estudos do elemento *P* (Clark *et al.*, 1995; Clark e Kidwell, 1997; Silva e Kidwell, 2000; Harring *et al.*, 2000, entre outros), chegam à mesma conclusão: o padrão de distribuição do elemento *P* leva a crer que diversos eventos de TH ocorreram ao longo da evolução. Robertson (2002) argumenta ainda que seus próprios estudos sobre a família *mariner* (Robertson, 1993; Robertson, 1997; Robertson e Martos, 1997), juntamente com trabalhos de outros autores (Garcia-Fernández *et al.*, 1995; Lohe *et al.*, 1995), chegam à conclusão de que muitos eventos de HT ocorreram ao longo da história da família *mariner*, com a invasão de vários novos genomas hospedeiros.

Eventos de TH envolvendo elementos da superfamília *hAT* já foram inferidos (como mencionado na introdução desta Dissertação), assim como para elementos da maioria das outras superfamílias de transposons de DNA descritas. Cada vez mais evidências nos levam a crer que os TEs são transmitidos horizontalmente com mais frequência do que se acreditava anteriormente. Na verdade, os achados atuais apontam para a idéia de que o processo de HT para novos genomas hospedeiros é, de fato, um passo essencial no ciclo de vida dos transposons.

Os resultados obtidos no presente estudo reforçam esta visão de que eventos de HT em eucariotos são mais comuns do que se previa anteriormente. Neste estudo, caracterizamos uma nova família da superfamília *hAT*, no genoma

de espécies do gênero *Drosophila*, e a denominamos *harrow*. Analisamos a distribuição de seqüências desta família ao longo do gênero *Drosophila* e outras espécies de Drosophilidae relacionadas, fizemos uma análise filogenética das seqüências encontradas, computamos as divergências que estas seqüências apresentam entre si e comparamos os dados das análises de *harrow* com os dos genomas das suas espécies hospedeiras.

Nossos resultados permitem inferências robustas de HTs, de acordo com Silva *et al.* (2004), pois possuem os três principais tipos de evidências normalmente utilizados: i) distribuição descontínua do elemento em um grupo de espécies; ii) incongruência entre a filogenia do elemento e das espécies hospedeiras; e iii) alta similaridade de seqüência entre elementos situados em genomas de espécies diferentes. Entretanto, optamos por tomar uma posição mais conservadora ao inferir TH e, assim, incluímos um tipo de evidência complementar: a comparação da divergência em sítios sinônimos (dS) de seqüências de *harrow* e de genes nucleares das espécies hospedeiras.

A magnitude dos valores de dS variam com uma correlação negativa à intensidade da seleção do uso de códons preferenciais. Como as seqüências de *harrow* apresentam menor viés do uso de códons que os genes usados na comparação, seria esperado, caso tivessem sido transferidas verticalmente, que os valores de dS de *harrow* fossem maiores que dos genes das espécies hospedeiras. No entanto, não foi este o resultado que obtivemos. Os valores de dS de *harrow* foram significativamente menores que os dos genes das espécies hospedeiras para a maioria dos casos comparados, indicando que as seqüências de *harrow* devem possuir um ancestral comum mais recente do que os genomas da espécies hospedeiras.

Estes resultados, juntamente com os resultados dos três tipos de evidência tradicionalmente utilizados para inferir HTs (citados anteriormente), reforçam a hipótese de que eventos de HT parecem exercer um importante papel na evolução da família *harrow*. De maneira especulativa, hipotetizamos que *harrow* tenha sido transmitido horizontalmente para *D. tripunctata*, a partir de uma fonte desconhecida. Em seguida, tenha sido transmitido para *D. prosaltans* (podendo ter sido *D. tripunctata* a espécie doadora), e na seqüência tenha havido uma onda de HT, relativamente recente, onde *harrow* invadiria novos genomas de espécies

Neotropicais quase simultaneamente. Situações de ondas de HTs já foram reportadas por outros autores (exemplo, Haring *et al.*, 2000). Além disso, o tempo estimado para a ocorrência desta onda de invasão de *harrow* é similar ao estimado por Silva e Kidwell (2004) para a invasão do elemento *P* nos grupos *saltans* e *willistoni*, o que sugere que processos similares poderiam estar envolvidos na propagação de ambos os elementos nestas espécies.

Vetores e/ou Mecanismos de HT

Em um artigo de revisão sobre casos de HTs em *Drosophila*, Loreto *et al.* (2008) apresentam seis possíveis vetores ou mecanismos pelos quais os TEs poderiam ser transmitidos horizontalmente: 1) no caso de alguns retrotransposons que apresentam capacidade de produzir partículas semelhantes às virais, um vetor externo seria desnecessário; 2) vírus poderiam servir de veículo para a HT, no qual os TEs pegariam “carona”; 3) parasitas e/ou parasitóides de *Drosophila*, sugando material celular de uma espécie e em seguida de outra; 4) bactérias intracelulares simbiotes, transmitidas entre espécies; 5) transferência de bactérias endosimbiontes infectadas por bacteriófagos carregando TEs; e 6) bactérias endosimbiontes infectadas por bacteriófagos carregando TEs sendo transportadas por parasitas e/ou parasitóides de *Drosophila*. Os autores explicam que estes possíveis modelos não são mutuamente exclusivos, uma vez que HTs não parecem ser restritas a um mecanismo único.

Como *harrow* não é um retrotransposon e é incapaz de produzir partículas semelhantes às partículas virais, o primeiro possível modelo apresentado é descartado para o TE em questão. Todos os demais mecanismos são possíveis maneiras pelas quais cópias de *harrow* podem ter invadido os genomas das espécies hospedeiras. Consideramos especialmente interessantes, e talvez um pouco mais prováveis, os processos que envolvem vírus, bacteriófagos e/ou bactérias endosimbiontes. No entanto, não descartamos a possibilidade de que parasitas e parasitóides de *Drosophila* exerçam um papel nos eventos de HTs de *Drosophila*. Contudo, é provável que estes ajam em conjunto com os outros possíveis vetores, uma vez que eles parasitam especialmente ovos e larvas e não há evidências que eles possam, sozinhos, transmitir TEs para a linhagem germinativa das espécies receptoras.

Hipóteses alternativas

Tentamos pensar em hipóteses alternativas para estes resultados, para não parecer que estamos presos somente à idéia de HT. Consideramos que as seqüências de *harrow* poderiam estar sob algum tipo de evolução em concerto, e assim estarem evoluindo muito lentamente dentro de cada espécie. No entanto, nas duas espécies que possuem *harrow* das quais já dispomos dos genomas completos disponíveis, este está presente em um baixo número de cópias. *D. mojavensis* tem uma cópia apenas de *harrow*, impossibilitando essa seqüência de estar sob evolução em concerto. Embora *D. willistoni* apresente duas cópias, uma delas possui pequenas deleções e várias substituições de sítio único ao longo da seqüência do gene de transposase, o que imaginamos ser pouco provável se estivessem sob forte evolução em concerto. As outras espécies estudadas não possuem o genoma completo disponível ainda; assim, não podemos saber o número exato de cópias de *harrow* que elas apresentam. Entretanto, é possível dizer que essas espécies apresentam mais de uma cópia de *harrow* nos seus genomas, com exceção de *D. paulistorum*, pois obtivemos clones com seqüências divergentes.

Este achado (possibilidade de várias cópias) que poderia por um lado suportar a hipótese de evolução em concerto, por outro lado a enfraquece um pouco, pois muitas cópias apresentam divergências razoáveis entre si. Encontramos, inclusive, cópias com e sem grandes deleções ocorrendo no mesmo genoma. A exceção é *D. paulistorum*, em que todos os clones seqüenciados são muito similares, podendo sugerir a ocorrência de convergência gênica. Alternativamente, este resultado também pode refletir a presença de uma cópia de *harrow* no genoma desta espécie, de forma que as duas únicas substituições de nucleotídeos, entre os 676 pb dos quatro clones seqüenciados, podem ser artefatos metodológicos. Assim, embora não possamos descartar a hipótese de evolução em concerto, achamos que esta deve ser insuficiente para explicar este alto grau de similaridade apresentado entre seqüências de *harrow* nos genomas das diferentes espécies hospedeiras.

Outra hipótese alternativa considerada é que as seqüências de *harrow* estariam sob seleção em nível de RNA, por serem alvo de interferência de RNA

(RNAi, do inglês, “RNA interference”). RNAi é um mecanismo de silenciamento de expressão gênica e acredita-se que seja uma importante defesa dos genomas hospedeiros contra invasões e multiplicações de vírus e TEs (Slotkin e Martienssen, 2007). Assim, *harrow* poderia estar sendo mantido conservado para permitir o reconhecimento de seqüências homólogas no processo de RNAi.

Uma descrição simplificada da forma geral como o silenciamento por RNAi ocorre é apresentada a seguir (baseada na revisão de Slotkin e Martienssen, 2007). O processo de RNAi é ativado pela presença de moléculas de RNA dupla fitas homólogas a porções do gene alvo. Estas moléculas são clivadas por proteínas da família *Dicer* em fragmentos menores, de cerca de 21 a 30 pb, chamados siRNAs (do inglês, “small interfering RNAs”). Estes siRNAs ligam-se a várias proteínas formando o complexo de silenciamento RNA-induzido (RISC, do inglês “RNA-induced silencing complex”). Em seguida, RISC é ativado quando estes siRNAs passam a ser fita única. Uma vez ativado, RISC pode reconhecer e se ligar ao mRNA alvo, o qual é clivado e posteriormente degradado por outras proteínas.

Acreditamos que a hipótese da alta similaridade interespecífica das seqüências de *harrow* ser resultado de forte restrição seletiva por estar sob ação de RNAi seja menos provável do que a hipótese *harrow* ter passado HT, pelos seguintes motivos: 1) Primeiramente, uma vez que a ligação ao mRNA é feita por fragmentos tão pequenos de siRNA, não conseguimos visualizar como seria possível manter uma região da transposase grande (quase 700 pb) como a analisada aqui tão conservada; 2) Por outro lado, estudos atuais apontam para um processo inverso à conservação de seqüência devido à RNAi. De fato, uma espécie de “corrida armamentista” é apontada para a relação entre “seqüências parasitas” (como vírus e TEs) e genes responsáveis pelo mecanismo de RNAi, mostrando que estes genes evoluem rapidamente em resposta à também rápida evolução das seqüências invasoras para escapar do silenciamento (Obbard *et al.*, 2006; Aravin *et al.*, 2007; Obbard *et al.*, 2009). Adicionalmente, nem mesmo sabemos se estas seqüências estão envolvidas de qualquer maneira com RNAi.

Além dessas considerações, levamos em conta que, admitindo a hipótese de VT, seguida de evolução em concerto ou restrição seletiva por RNAi, para explicar os resultados obtidos, ainda dependeríamos de inúmeros eventos de

perda estocástica do elemento ao longo da evolução. Acreditamos que se há algum mecanismo para manter seqüências de TEs tão similares, inclusive mantendo baixa divergência sinônima, entre espécies tão distantes (mesmo de subgêneros distintos), este mecanismo é ainda completamente desconhecido. Quem sabe, com descobertas futuras da ciência, o quadro possa mudar, porém, agora, consideramos que a explicação mais razoável para nossos achados seja a ocorrência de múltiplos eventos de HT. Tal interpretação está completamente de acordo com a visão atual sobre a propensão, e aparente necessidade, de transposons de DNA de serem transmitidos horizontalmente, invadindo novos genomas hospedeiros, para não se tornarem extintos.

Referências

Akaike H (1974) A new look at the statistical model identification. IEEE Transaction on Automatic Control 19: 716–723.

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

Almeida LM e Carareto CMA (2006) Sequence heterogeneity and phylogenetic relationships between *copia* retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. Genet Sel Evol 38: 535-550.

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

Aravind L (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci 25: 421-423.

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

Arensburger P, Kim YJ, Orsetti J, Aluvihare C, O’Brochta DA e Atkinson PW (2005) An active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. Genetics 169: 697-708.

Ayala FJ, Mourão CA, Pérez-Salas S, Richmond R e Dobzhansky T (1970) Enzyme variability in the *Drosophila willistoni* group. I – Genetic differentiation among sibling species. Proc Natl Acad Sci 67: 225-232.

Bächli G (2008) TaxoDros: The Database on Taxonomy of Drosophilidae, v.1.03, Database 2008/12. <http://taxodros.unizh.ch/>.

Bigot YM, Corinne A e Georges P (1996) Computer analyses reveal a *hobo*-like element in the nematode *Caenorhabditis elegans*, which presents a conserved transposase domain common with the TC1-*Mariner* transposon family. Gene 174: 265-271.

Boussy e Daniels (1991) *hobo* transposable elements in *Drosophila melanogaster* and *Drosophila simulans*. Genet Res 58: 27-34

Brcic D e Koref-Santibañez S (1957) The *mesophragmatica* species group of *Drosophila*. Evolution 11: 300-310.

Brunak S, Engelbrecht J e 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 e Capy P (1999) Phylogenetic analysis of *Mos1*-like transposable elements in the Drosophilidae *J Molec Evol* 49: 760-768.

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, Anxolabéhère D e Langin T (1994) The strange phylogenies of the transposable elements: are horizontal transfers the only expansion? *Trends Genet* 10: 7-12.

Capy P, Bazin C, Higuete D e Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197pp.

Clark JB, Altheide TK, Schollosser MJ, Kidwell MG (1995) Molecular evolution of *P* transposable elements in genus *Drosophila*. I. The *saltans* and *willistoni* species groups. *Mol Biol Evol* 12: 902–913.

Clark JB e Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc Natl Acad Sci USA* 94: 11428-11433.

Clark JB, Silva JC, Kidwell MG (2002) Evidence for horizontal transfer of *P* transposable elements. *In*: Syvanen M, Kado CI (eds). *Horizontal Gene Transfer*. Academic Press: London, pp 161–171.

Craig NL, Graigie R, Gellert M e Lambowitz AM (2002) *Mobile DNA II*. American Society for Microbiology Press, Washington, DC, 1204pp.

Coates CJ, Johnson KM, Perkins HD, Howells AJ, O'Brocta DA e 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.

Cordeiro AR e Winge H (1995) Levels of evolutionary divergence of *Drosophila willistoni* sibling species. *In*: Levine L, ed. *Genetics of natural populations: The continuing importance of Theodosius Dobzhansky*. New York: Columbia University Press, 262–280.

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG e Chovnick A (1990a) Evidence for horizontal transfer of the *P* element between *Drosophila* species. *Genetics* 124: 339-355.

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.

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.

Essers L, Adolphi RH e Kunze R (2000) A highly conserved domain of the maize *Activator* transposase is involved in dimerization. *Plant Cell* 12: 211-224.

Gao J, Watabe H, Aotsuka T, Pang J e Zhang Y (2007) Molecular phylogeny of the *Drosophila obscura* species group, with emphasis on the Old World species. *BMC Evol Biol* 7:87.

Garcia-Fernandèz J, Bayascas-Ramirez JR, Marfany G, Muñoz-Marmol AM, Casali A, Baguna J e Saló E (1995) High copy number of highly similar *mariner*-like transposons in planarian (*Platyhelminthe*): evidence for a trans-phyla horizontal transfer. *Mol Biol Evol* 12: 421–431.

Gleason JM, Griffith EC e Powell JR (1998) A molecular phylogeny of the *Drosophila willistoni* group: conflicts between species concepts? *Evolution* 52: 1093-1103.

Gómez-Gómez E, Anaya N, Roncero MIG e Hera C (1999) *Folyt1*, a New Member of the *hAT* Family, Is Active in the Genome of the Plant Pathogen *Fusarium oxysporum*. *Fungal Genet Biol* 27: 67-76.

Graur D e Li W-H (2000) *Fundamentals of Molecular Evolution*, 2nd edn. Sinauer Associates, Sunderland, MA.

Handler AM e Gomez SP (1996) The *hobo* Transposable Element Excises and Has Related Elements in Tephritid Species. *Genetics* 143: 1399-1947.

Handler AM e Gomez SP (1997) A new *hobo*, *Ac*, *Tam3* transposable element, *hopper*, from *Bactrocera dorsalis* is distantly related to *hobo* and *Ac*. *Gene* 185: 133-135.

Haring E, Hagemann S e Pinsker W (2000) Ancient and Recent Horizontal Invasions of Drosophilids by *P* Elements. *J Mol Evol* 51: 577-586.

Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht, J, Rouze P e 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 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.

Hedges DJ e Deininger PL (2007) Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 616: 46-59

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

Hickman AB, Perez ZN, Zhou LQ, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL e Dyda F (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat Struct Mol Biol* 12:715–721.

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG e Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.

Hijmans RJ, Guarino L, Jarvis A, O'Brien R, Mathur P, Bussink C, Cruz M, Barrantes I and Rojas E (2005) Diva-GIS version 5.2. Manual. Disponível em <http://www.diva-gis.org>.

Houck MA, Clark JB, Peterson KR e Kidwell MG (1991) Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* 253: 1125-1129.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Jordan IK, Matyunina LV e McDonald FJ (1999) Evidence for recent horizontal transfer of long terminal repeat transposon. *Proc Natl Acad Sci USA* 96: 12621-12625.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O e Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic Genome Res* 110: 462-467.

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

Kastritsis CD (1969) The chromosomes of some species of the *guarani* group of *Drosophila*. *J Heredity* 60: 51-57.

Kazazian HH Jr (1998) Mobile elements and disease. *Curr Opinion in Genetics and Development* 8: 343-350.

Kempken F e Kück, U (1996) *Restless*, an active *Ac*-like transposon from the fungus *Tolytocladium inflatum*: structure, expression, and alternative RNA splicing. *Mol Cell Biol* 16: 6563–6572

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 (1983) Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 80: 1655–1659.

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

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111-120.

Klein CC, Essi L, Golombieski RM e Loreto ELS (1999) Disgenesia do híbrido em populações naturais de *Drosophila melanogaster*. *Ciencia e Natura* 21: 7-20.

Koga A, Shimada A, Shima A, Sakaizumi M, Tachida H e Hori H (2000) Evidence for Recent Invasion of the Medaka Fish Genome by the *To2* Transposable Element. *Genetics* 155: 273-281.

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.

Kunze R (1996) The maize transposable element *Activator (Ac)*. *Curr Top Microbiol Immunol* 204: 161-194.

Lin Y-J, Seroude L e Benzer S (1998) Extended Life-Span and Stress Resistance in the *Drosophila* Mutant Methuselah. *Science* 282: 943-946.

Lohe AR, Moriyama EM, Lidholm DA e Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of *mariner* like transposable elements. *Mol Biol Evol* 12: 62-72.

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

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 Molecular Biology* 17: 113-124.

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D e Bryant SH (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 35: D237-240.

Markow TA e O'Grady PM (2006) Phylogenetic relationships of *Drosophilidae*. In: Markow TA e O'Grady PM, *Drosophila: A guide to species identification and use*. Academic Press, London, pp 3-64.

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 (1948) Mutable loci in maize. *Carnegie Inst Wash, Year book*, 47: 155-169.

Michel K., O'Brochta DA e Atkinson PW (2003) The C-terminus of the *Hermes* transposase contains a protein multimerization domain. *Insect Biochem Mol Biol* 33: 959-970.

Moon S, Jung K-H, Lee D-e, Jiang W-Z, Koh HJ, Heu M-H, Lee DS, Suh HS e An G (2006) Identification of Active Transposon *dTok*, a Member of the hAT Family, in Rice. *Plant Cell Physiol* 47: 1473-1483.

Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on *tRNA* availability, *J Mol Evol* 37: 273-280.

Muehlbauer GJ, Bhau BS, Syed NS, Heinen S, Cho S, Marshall D, Patearon S, Buisine N, Chalhoub B e Flavell AJ (2006) A *hAT* superfamily transposase recruited by the cereal grass genome. *Mol Gen Genomics* 275: 553-563.

Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H e Matsuo M (1993) Insertion of a 5' Truncated L1 Element into the 3' End of Exon 44 of the Dystrophin Gene Resulted in Skipping of the Exon during Splicing in a Case of Duchenne Muscular Dystrophy. *J Clin Invest* 91: 1862-1867.

Nei M e Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418-426.

Nicholas KB e Nicholas HBJ (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Divulgado pelo autor. <http://www.psc.edu/biomed/genedoc>.

Obbard DJ, Jiggins FM, Halligan DL e Little TJ (2006) Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* 16: 580-585

Obbard DJ, Gordon KHJ, Buck AH e Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. *Phil Trans R Soc B* 364: 99-115.

O'Grady PM, Clark JB e Kidwell MG (1998) Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. *Mol Biol Evol* 15: 656-664.

Ortiz MF e Loreto ELS (2009) Characterization of new *hAT* transposable elements in 12 *Drosophila* genomes. *Genetica* 135: 67-75.

Pascual L, Periquet G (1991) Distribution of *hobo* transposable elements in natural populations of *Drosophila melanogaster*. *Mol Biol Evol* 8: 282--296.

Pinkerton AC, Whyard S, Mende HA, Coates CJ, O'Brochta DA e Atkinson PW (1999) The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol Biol* 8: 423-434

Posada C e Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Ray A, Pagan HJT, Thompson ML e Stevens R (2007) Bats with *hAT*s: Evidence for Recent DNA Transposon Activity in Genus *Myotis*. *Mol Biol Evol* 24(3): 632-639.

Robertson HM (1993) The *mariner* transposable element is wide spread in insects. *Nature* 362: 241-245.

Robertson HM (1997) Multiple *mariner* transposons in flatworms and hydras are related to those of insects. *J Hered* 88: 195-201.

Robertson HM e Martos R (1997) Molecular evolution of the second ancient *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genomes lineage. *Gene* 205: 219-228.

Robertson HM (2002) Evolution of DNA Transposons in Eukaryotes *In: Mobile DNA II*. Craig NL, Graigie R, Gellert M e Lambowitz AM. American Society for Microbiology Press, Washington, DC, pp 1093-1110.

Rodriguez F, Oliver JL, Marin A e Medina JR (1990) The general stochastic model of nucleotide substitution. *J Theor Biol* 142: 485-501.

Rodríguez-Trelles F, Tarrío R e Ayala FJ (1999) Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. *Mol Phylogenet Evol* 13: 110-121.

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

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

Russo CAM, Takezaki N e Nei M (1995) Molecular phylogeny and divergence times of *Drosophilid* species. *Mol Biol Evol* 12: 391-404.

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.

Rychlik, W. (1992) Oligo 4.1 Primers Analysis Software. National Biosciences, Inc., Plymouth, Minn.

Salzano FM (1955) Chromosomal polymorphism in two species of the *guarani* group of *Drosophila*. *Chromosoma* 7: 39-50.

SanMiguel P, Tikhonov A, Jin Y, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z e Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765-768.

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

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

Silva JC, Loreto ELS e Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Current Issues in Molecular Biology* 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.

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

Shields DC, Sharp PM, Higgins DG e Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol* 5: 704-716.

Slotkin RK e Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8: 272.

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

Sturtevant AH (1939) On the subdivision of the genus *Drosophila*. *Proc Natl Acad Sci* 25: 137-141.

Sturtevant AH (1942) The classification of the genus *Drosophila* with the description of nine new species. Univ Texas Publ 4213: 5-51.

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.

Swofford DL (2003) PAUP: Phylogenetic Analysis using Parsimony (and other methods). Version 4. Sinauer Associates, Massachusetts.

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

Tamura K, Subramanian S e Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol Biol Evol 21: 36-44.

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

Tarrío R, Rodriguez-Trelles F e Ayala FJ (2000) Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: The *Drosophila saltans* and *willistoni* groups, a case study. Mol Phylogenet Evol 16: 344-349.

Tatarenkov A e Ayala FJ (2001) Phylogenetic relationships among species groups of the *virilis-repleta* radiation of *Drosophila*. Mol Phylogenet Evol 21: 327-331.

Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King, R. C. (ed) Handbook of Genetics. Plenum, New York, pp 421-469.

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.

Val FC, Marques MD e Vilela CR (1981) Drosophilidae of Neotropical Region. In: Ashburner M, Carson HL, Thompson JN (eds) The Genetics and Biology of *Drosophila*. Academic Press, New York, pp 123-167.

Vilela CR 1992. On the *Drosophila tripunctata* species group (Diptera, Drosophilidae). Rev Bras Ent 36: 197-221.

Volff J-N (2006) Turning Junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28: 913-922.

Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW e Collins FS, (1991) A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature* 353: 864-866.

Warren WD, Atkinson PW e 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 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. *Nature Reviews Genetics* 8: 973-982.

Wilson RJ, Goodman JL, Strelets VB and The FlyBase Consortium (2008) FlyBase: integration and improvements to query tools. *Nucl Acids Res* 36: D588-D593.

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

Xiong Y e Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal* 9: 3353-3362.