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

HISTÓRIA EVOLUTIVA DE ELEMENTOS TRANSPONÍVEIS DA SUPERFAMÍLIA *Tc1-Mariner* EM DROSOFILÍDEOS

DISSERTAÇÃO DE MESTRADO

Gabriel da Luz Wallau

Santa Maria, RS, Brasil.
2010

2010 Mestre WALLAU, Gabriel da Luz PPGBA/UFSM, RS

HISTÓRIA EVOLUTIVA DE ELEMENTOS TRANSPONÍVEIS DA SUPERFAMÍLIA *Tc1-Mariner* EM DROSOFILÍDEOS

por

Gabriel da Luz Wallau

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Biodiversidade Animal, Área de Concentração em Biologia Evolutiva de Insetos, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial da obtenção do grau de
Mestre em Ciencias Biológicas
Área Biodiversidade Animal

Orientador: Prof. Dr. Elgion Lucio da Silva Loreto

Santa Maria, RS, Brasil.
2010

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

A Comissão Examinadora, abaixo assinada,
aprova a dissertação de Mestrado

HISTÓRIA EVOLUTIVA DE ELEMENTOS TRANSPONÍVEIS DA SUPERFAMÍLIA *Tc1-Mariner* EM DROSOFILÍDEOS

elaborada por
Gabriel da Luz Wallau

como requisito parcial para a obtenção do grau de
Mestre em Biodiversidade Animal

COMISÃO EXAMINADORA

____
Elgion Lucio da Silva Loreto, Drº.
(Presidente/Orientador)

____
Monica Laner Blauth, Drª. (Universidade Federal de Pelotas)

____
Yanina Panzera Crespo, Drª. (Facultad de Ciencias)

Santa Maria, 26 de fevereiro de 2010.

À minha família
por toda a confiança e tranqüilidade
passada durante essa longa jornada.

Agradecimentos

Ao Prof. Elgion, primeiro pela oportunidade de trabalhar no LabDros, pelo apoio e dedicação com minha formação como cientista e por sua amizade e incontáveis conselhos durante os últimos 6 anos.

À Profª Lenira pelos ensinamentos sobre ciência e pelos muitos conselhos a um pai de primeira viagem e por proporcionar um ótimo ambiente de trabalho, com muito bom humor.

À Profª Sonia Cechin e ao Sr. Paulo pelos inúmeros esclarecimentos e ajuda com os problemas relacionados à Pós-Graduação.

À minha querida professora de Francês, dona Elisa, pelos ensinamentos e amizade.

À CAPES e CNPq, pelo apoio financeiro para a execução do trabalho.

Aos meus colegas do PPGBA, pela parceria e amizade conquistada durante esses anos de estudo.

A toda a "gangue, tribo, tigrada" do LabDros, pelas incontáveis ajudas. Obrigado à Maríndia e a Nina pela amizade e por conseguirem algumas espécies de *Drosophila* para esse trabalho. Obrigado a Adriana, Marcos, Fabiano e ao Paulo, pela amizade e pelo orgulho que eu sinto em saber que vocês também passaram pelo LabDros. Obrigado Ana, Ronaldo, Paloma, Felipe ten Caten, pelas muitas vezes que tiveram que terminar alguma extração de plasmídeo quando a Malu adoecia. À Paloma e ao Ronaldo, pelas muitas placas de sequenciamento montadas para esse trabalho. À Lizandra, pelos conselhos e pelas muitas vezes que eu tomei seu tempo pedindo uma ajuda aqui e ali. Queria agradecer em especial ao Felipe (meu cumpadre), a Paloma (minha cumadre), a Ana e o Ronaldo, pela incalculável ajuda com a Maluzita e pelos conselhos, companheirismo, amizade sincera e muitas gargalhadas durante o tempo que passamos juntos. A todo o pessoal do Lab: Andreza, Bruna, Francis, Francine, Sinara, Ana, Felipe, Gabriela, Marcela, Málvaro, Vanessa, Micheli, Pedro e Valéria que propiciaram um ótimo ambiente de trabalho, com discussão científica de qualidade e muitas risadas.

Ao Franchesco, Anderson, Cristian, Juliano, Mauro, Maicon (mais conhecido como Brusque) e João, meus amigos de todas as horas que, apesar de alguns estarem longe, nunca deixaram de me apoiar.

Ao Josué (meu compadre), pelas incontáveis ajudas com a Maluzita e por sempre estar com o ombro amigo por perto. Obrigado ao Josué e Diogo, pela amizade e pelas interessantíssimas conversas filosóficas sobre os acontecimentos do mundo.

Ao meu grande amigo de infância Faber (Fabinho), à Cris e ao Caua que, apesar da distância, tenho certeza de que, quando retornamos ao pago, nossa amizade só aumenta.

Aos meus amigos do futebol da semana: Paulo, Cassiano, Franchesco, Rodrigo, Douglas, Matheus, Diego, Bruno, Tiago, Jonas, Jeferson, entre outros, pelos "canelaços" e lances de "bola murcha" que propiciaram muita risada, além dos churrascos no fim do semestre.

E por último, um agradecimento muito especial às pessoas mais importantes da minha vida, a minha família. Obrigado a família da Luz e Wallau, por estar sempre presente quando precisei de ajuda. Obrigado à minha mãe, meu pai e minha irmã (e a Sofia) por todo o carinho, suporte e confiança que me passaram. Obrigado em especial a minha amada Cristiane e a minha princesinha Malu por me aguentarem no dia-a-dia, não deixarem de me amar em um instante sequer e por me esperarem sempre com um sorriso lindo, quando eu chegava em casa.

Resumo

Dissertação de Mestrado
Programa de Pós-Graduação em Biodiversidade Animal
Universidade Federal de Santa Maria

História Evolutiva de Elementos Transponíveis da Superfamília *Tc1-Mariner* em Drosofilídeos

Autor: Gabriel da Luz Wallau
Orientador: Elgion Lucio da Silva Loreto
Data e Local da Defesa: Santa Maria, 26 de fevereiro de 2010

Elementos transponíveis (TEs) são regiões do DNA que podem se mover dentro e entre genomas, causando grande impacto na evolução dos organismos. A Superfamília *Tc1-Mariner* se destaca por ser, provavelmente, a superfamília de transposons de DNA com maior distribuição na natureza, sendo ubíqua em eucariotos. Em parte desse trabalho, caracterizamos elementos da família *mariner* em drosofilideos Neotropicais para os quais obtivemos amplificação com primers degenerados. Os primers foram construídos na região do domínio catalítico da transposase de *mariner* o que permite amplificar uma ampla gama de sequências relacionadas à *mariner*. Um total de 23 espécies apresentou sequências relacionadas à *mariner* pertencentes a três subfamílias (*mellifera*, *mauritiana* e *irritans*). Esses elementos apresentaram uma distribuição descontínua e incongruências com a filogenia das espécies hospedeiras, o que sugere eventos de transmissão horizontal entre drosofilideos e, até mesmo entre drosofilideos e espécies de outra família, superfamília e ordem. Além disso, algumas sequências apresentaram um quadro aberto de leitura, os motivos catalíticos conservados e uma forte seleção purificadora atuando, o que sugere que esses elementos sejam provenientes de elementos ativos. Em outra parte do trabalho, caracterizamos as sequências relacionadas ao elemento *Paris* (pertencentes à família *Tc1*), com buscas nos doze genomas de *Drosophila* disponíveis. Nessas buscas,foram encontrados cinco novos elementos relacionados à *Paris* (um em *D. ananassae*, um em *D. pseudoobscura*, um em *D. persimilis*, um em *D. mojavensis* e um em *D. willistoni*), com um número de cópias variando de dois a sete. Três espécies apresentaram elementos potencialmente ativos. A análise evolutiva desses elementos sugere que estão sendo mantidos por transmissão vertical, com alguns eventos de perda estocástica nas espécies analisadas.

Palavras-chave: transposon, *Tc1-Mariner*, drosofilídeos, evolução genômica

Abstract

Dissertação de Mestrado
Programa de Pós-Graduação em Biodiversidade Animal
Universidade Federal de Santa Maria

Evolutionary History of Transposable Elements of Superfamily *Tc1-mariner* in Drosophilids

Autor: Gabriel da Luz Wallau
Orientador: Elgion Lucio da Silva Loreto
Data e Local da Defesa: Santa Maria, 26 de fevereiro de 2010

Transposable elements (TEs) are DNA regions that can move within and between genomes, causing great impact on the host organisms. The *Tc1-Mariner* superfamily stands out for being, probably, the DNA transposons superfamily with greater distribution in nature, being ubiquitous in eukaryotes. In part of this work, we characterize elements of the *mariner* family in Neotropical drosophilids, which were obtained through amplification with degenerated primers. The primers were designed for the catalytic domain region of *mariner* transposase allowing amplification of a wide range of *mariner*-like sequences. A sum of twenty-three species have *mariner*-like sequences belonging to three subfamilies (*mellifera*, *mauritiana* and *irritans*). These elements present a patchy distribution and incongruences with the host phylogeny, suggesting horizontal transmission events between drosophilids and even between drosophilids and species of other families, subfamilies and orders. Moreover, some sequences present open reading frames, conserved catalytic motifs and evidence for the action of a strong purifying selection, which suggest yhat they originated from active elements. In another part of the work, we characterize *Paris*-like elements (belonging to *Tc1* family), through searches in the twelve *Drosophila* genomes available. These searches, enabled us to find five new *Paris*-like elements (one in *D. ananassae*, one in *D. pseudoobscura*, one in *D. persimilis*, one in *D. mojavensis* and one in *D. willistoni*), with a copy number ranging from two to seven. Three species have putatively active elements. The evolutionary analysis of these elements suggests that they have envolved through vertical transmission associated with some events of stochastic loss in the analysed species.

Key-words: transposon, *Tc1-Mariner*, drosophilids, genome evolution

SUMÁRIO

CAPÍTULO I 11

1. Introdução Geral 12

1.1 Elementos de Transponíveis 12

1.1.1 Classificação 13

1.1.2 Evolução dos TEs 14

1.1.3 Superfamília *Tc1-Mariner* 16

1.2 A Família Drosophilidae 18

1.2.1 O Gênero *Drosophila* 18

1.3 Objetivos 19

1.3.1 Objetivos específicos 19

CAPÍTULO II

The evolutionary history of *mariner*-like elements in neotropical drosophilids 21

1. Abstract 22

2. Introduction 22

3. Materials and methods 25

4. Results 29

5. Discussion 34

6. References 37

7. Supplementary material 45

CAPÍTULO III

Evolutionary history of *Paris*-like elements in 12 *Drosophila* genomes 53

1. Abstract 54

2. Introduction 55

3. Materials and methods 57

4. Results and Discussion 60

5. Acknowledgments 68

6. References 69

7. Supplementary material 74

CAPÍTULO IV

Conclusões gerais e perspectivas 80

REFERÊNCIAS BIBLIOGRÁFICAS 82

CAPÍTULO I

1. Introdução Geral

1.1 Elementos de Transposição

Os elementos transponíveis (TEs - Transposable Elements) são fragmentos de DNA que possuem a capacidade de se mover entre regiões de um genoma. A grande maioria dos genomas investigados até hoje apresentaram elementos transponíveis (BIÉMONT E VIEIRA, 2006), com apenas algumas exceções (PRITHAM, 2009).

Na década de 40, Bárbara McClintok descobriu os TEs no genoma de milho, definindo-os como elementos controladores, pois estes aparentemente controlavam a coloração dos grãos de milho. Nos anos subsequentes, a idéia de que sequências de DNA podiam se mover dentro dos genomas não foi aceita pela comunidade científica. Somente na década de 80 os elementos controladores, atualmente chamados de elementos transponíveis, foram redescobertos, gerando várias mutações em *Drosophila* (50% das mutações) e em milho (10% das mutações) (FINNEGAN, 1992, KAZAZIAN, 1998). A partir do reconhecimento dos TEs, muitos estudos mostraram que estes elementos poderiam influenciar profundamente o processo evolutivo do seu hospedeiro. A mobilização desses elementos pode promover mudanças estruturais de larga escala, como: rearranjos cromossômicos, modificações nos padrões epigenéticos de regulação, além da geração de variabilidade genética, novos genes e, consequentemente, inovações biológicas (FESCHOTTE E PHRITHAN, 2007).

Uma constatação decorrente do aumento do número de genomas sequenciados, foi que grande parte de alguns genomas são constituídos de TEs. Em alguns casos extremos, como no genoma de: *Rana esculenta, Zea mays*, *Drosophila melanogaster* e *Homo sapiens*, os TEs representam, respectivamente, 77, 60, 15-22 e 45% do genoma (BIÉMONT E VIEIRA, 2006). Além disso, existe uma grande variação entre número de cópias, distribuição e tipos de TEs, entre diferentes espécies. Um indivíduo de *D. melanogaster* pode possuir, aproximadamente, 50 famílias e uma variação de 1 a 100 cópias de alguns elementos (FINNEGAN, 1986). As variações no número de cópias e o tipo de TEs entre diferentes espécies são decorrentes de fatores como características intrínsecas dos TEs e as diferentes forças evolutivas que atuam sobre esses. A grande distribuição e variabilidade dos TEs presentes nos seres vivos sugere que sua origem se

deu em eventos antigos, com íntimas ligações com os primeiros genomas (CAPY *et al.*, 1998).

Muitos autores consideram os TEs como parasitas genômicos, pois esses utilizam o material genético do hospedeiro como seu ambiente para a replicação. Porém, as relações com o genoma hospedeiro podem variar continuamente entre parasitismo e mutualismo (simbiose) (KIDWELL E LISCH, 2001). Além das relações entre os elementos e o genoma hospedeiro, existem relações entre diferentes elementos. Os diferentes TEs podem ser comparados a espécies que vivem e interagem em um ecossistema genômico (LE ROUZIC *et al.*, 2007, VENNER *et al.*, 2009).

1.1.1. Classificação dos TEs

Devido à diversidade de TEs descobertos, a partir da década de 80, fez-se necessária a criação de um sistema de classificação. Em 1989, Finnegan propôs o primeiro sistema de classificação dos TEs que discriminava duas classes (Classe I ou retrotransposons e Classe II ou transposons) que diferem basicamente no mecanismo de mobilização. Os elementos da Classe I possuem um mecanismo de mobilização baseado em um intermediário de RNA que utiliza duas enzimas transcritas pelos elementos: transcriptase reversa (RT, do inglês Reverse Transcriptase) e integrase (IN). Primeiro o RNA menssageiro (mRNA) do TE é transcrito pela maquinaria de transcrição do hospedeiro, após, o mRNA é revertido para DNA pela enzima transcriptase reversa e, em seguida, o fragmento de DNA é inserido em outra região do genoma pela enzima integrase. Este sistema de mobilização é conhecido como "copiar e colar". Já nos elementos da Classe II, o mecanismo de mobilização é baseado em um intermediário de DNA, onde somente uma enzima chamada transposase retira o elemento de uma região do genoma e insere em outra. Este sistema de mobilização é conhecido como "cortar e colar". Apesar das diferenças nos mecanismos de transposição, a integrase de alguns elementos de RNA e a transposase de alguns elementos de DNA provavelmente têm uma origem comum (CAPY *et al.*, 1997).

Com a descrição de novos TEs decorrentes dos sequenciamentos de vários genomas, alguns elementos com novas características no processo de mobilização foram descobertos. Em vista disso, um novo sistema de classificação foi proposto por Wicker *et al.* (2007). Esse novo sistema manteve as duas Classes propostas por Finnegan, no

entanto, adicionou novas características do mecanismo de mobilização, similaridade de sequências e relações estruturais, seguindo uma ordem hierárquica de taxonomia.

- Subclasse: distingue os elementos de Classe II que fazem cópias de si mesmo para uma nova inserção, cortando apenas uma das fitas do DNA e os que saem totalmente do sítio onde estavam inseridos, para se inserirem em outro local do genoma, cortando as duas fitas do DNA.
- Ordem: distingue os grupos baseados nas diferenças do mecanismo de inserção e, consequentemente, a organização geral e enzimologia dos TEs.
- Superfamília: as superfamílias, dentro de uma ordem, compartilham a estratégia de replicação, mas distinguem-se umas das outras por características como: estrutura das proteínas, domínios não codificantes, presença e tamanho de uma repetição gerada nas extremidades dos TEs, quando estes são inseridos nos genomas, o sítio de duplicação (TSD, do inglês Target Site Duplication).
- Família: é definida pela conservação da sequência de aminoácidos. Entre as famílias de uma mesma Superfamília, a similaridade das proteínas geralmente é grande, porém, em nível de DNA, a conservação é mínima e restrita a partes da região codificante.
- Subfamília: são definidas com análises filogenéticas e, em casos específicos, pode distinguir populações homogêneas de elementos autônomos e não autônomos.

1.1.2. Evolução dos TEs

Os TEs são parasitas genômicos que têm um papel importante na formação e mudança desses, ao longo de milhões de anos de evolução (ARKHIPOVA, 2001). A presença de TEs em uma determinada espécie pode ser decorrente de dois processos: transmissão vertical (TV) ou transmissão horizontal (TH) (Figura 1). O primeiro processo é caracterizado pela herança do material genético que ocorre entre espécies ancestrais e seus descendentes. Segundo essa premissa, se os TEs são herdados por TV, é esperado que sua história filogenética reflita a filogenia dos hospedeiros. Já o segundo processo é caracterizado pela transferência de material genético, entre espécies isoladas reprodutivamente. Esses eventos, geralmente são inferidos quando as relações filogenéticas dos TEs não estão de acordo com a história evolutiva do hospedeiro. Vários fatores podem influenciar o sucesso de eventos de TH, como: estabilidade da integração do TE em um novo genoma, distribuição dentro do novo hospedeiro, fatores de regulação

cis e *trans* do próprio TE, regulação do hospedeiro sobre o TE, seleção natural e deriva genética (revisão em SILVA *et al.* 2004). Nos últimos anos, muitos casos de TH foram relatados para o gênero *Drosophila* (BARTOLOMÉ *et al.,* 2009, SÁNCHEZ-GARCIA *et al.*, 2005), com os elementos de Classe II apresentando uma maior frequência de TH (LORETO *et al.*, 2008); porém, ainda se conhece muito pouco sobre os mecanismos e vetores que promovem eventos de TH. Algumas hipóteses potenciais são: alguns retrotransposons podem ser os próprios vetores, pois são capazes de produzir partícula viral, vírus de DNA, bactérias simbiontes intracelulares e parasitas (ácaros e vespas), ou pode ocorrer a transferência de transposon de classe II, na forma de DNA episomal (O'BROCHTA *et al.* 2009).

.

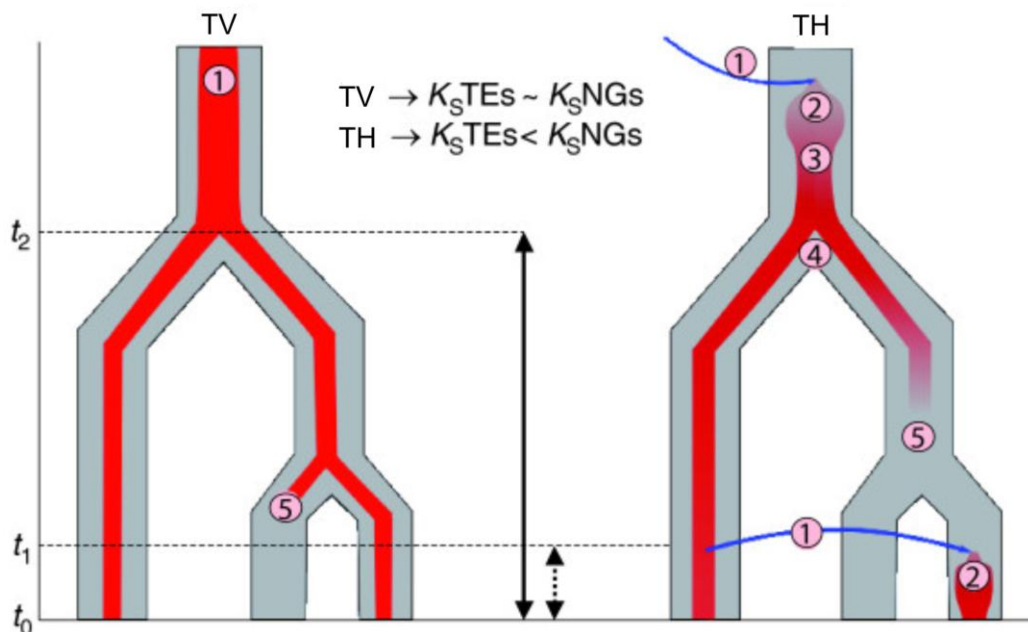


Figura 1: História evolutiva dos TEs nos seus hospedeiros. Barras vermelhas - processo de transmissão vertical, linhas azuis - eventos de trasmissão horizontal, eixo *y* descreve o tempo evolutivo. Se o TE é herdado por TV é esperado que as *Ks* (substituições sinônimas) sejam semelhantes às dos NGs (genes nucleares), porém se ocorreu um evento de TH é esperado que a *Ks* do TE seja significativamente menor que a dos NGs. (1) A chegada de um TE em um novo hospedeiro por TH (2) é seguido por um período de diminuição no número de cópias (3) até ser alcançado o equilíbrio entre transposição e seleção. (4) Após especiação e concomitante diversificação de hospedeiro e TEs, (5) a perda estocástica de uma família em uma dada linhagem pode ser revertido por TH. Modificada de Bartolomé *et al.*, 2009.

Existem três metodologias básicas para inferir eventos de TH: (1) baseado na distribuição descontínua dos TEs: presença de um TE específico em uma ou em poucas

espécies dentro de um clado filogenético que não possui o TE; (2) baseado em incongruências filogenéticas: comparação direta entre a filogenia dos TEs e dos hospedeiros e discordância filogenética entre hospedeiros e TEs avaliado por algoritmos que buscam associações plausíveis entre eles; (3) baseado na similaridade de sequências: comparando-se as distâncias entre TEs de diferentes táxons e, em alguns casos, esta distância com a dos genes do hospedeiro, analisando-se a razão entre sítios sinônimos e não sinônimos entre TEs e genes do hospedeiro entre diferentes táxons (Figura 1) e, comparando o padrão de bandas e/ou a intensidade do sinal de hibridização entre TEs de diferentes táxons. Alguns desses padrões, entretanto, também podem ser explicados por hipóteses alternativas como: polimorfismo ancestral, diferentes taxas de evolução em TEs de diferentes espécies e perdas estocásticas em alguns táxons (Revisão Loreto *et al.* 2008).

1.1.3. Superfamília *Tc1-Mariner*

Essa superfamília pertence aos elementos de Classe II, Subclasse I e ordem TIR (Figura 2) e se destaca por ser, provavelmente, a superfamília de transposons de DNA com maior distribuição na natureza, sendo ubíqua em eucariotos (PLASTERK *et al.*, 1999, WICKER *et al.*, 2007). Os elementos dessa superfamília possuem aproximadamente 1300-2400 pb e se caracterizam por possuírem uma estrutura simples com duas repetições terminais invertidas (TIRs, do inglês Inverted Terminal Repeat), um quadro aberto de leitura (ORF, do inglês Openning Reading Framme) que codifica uma transposase (Figura 3). A transposase dos elementos dessa superfamília possue dois domínios: N-terminal, onde está o domínio de ligação ao DNA e C-terminal, onde está o domínio catalítico (Figura 2). Os TEs da superfamília *Tc1-mariner* têm preferência por inserções, em regiões ricas em TA, gerando um sítio de duplicação (TSDs, do inglês Target Site Duplication) TA. As suas famílias são distinguidas basicamente por uma assinatura de três aminoácidos: dois resíduos de ácido aspártico (DD) separados 34 ou 35 aminoácidos de um resíduo de ácido aspártico ou glutâmico (D/E).

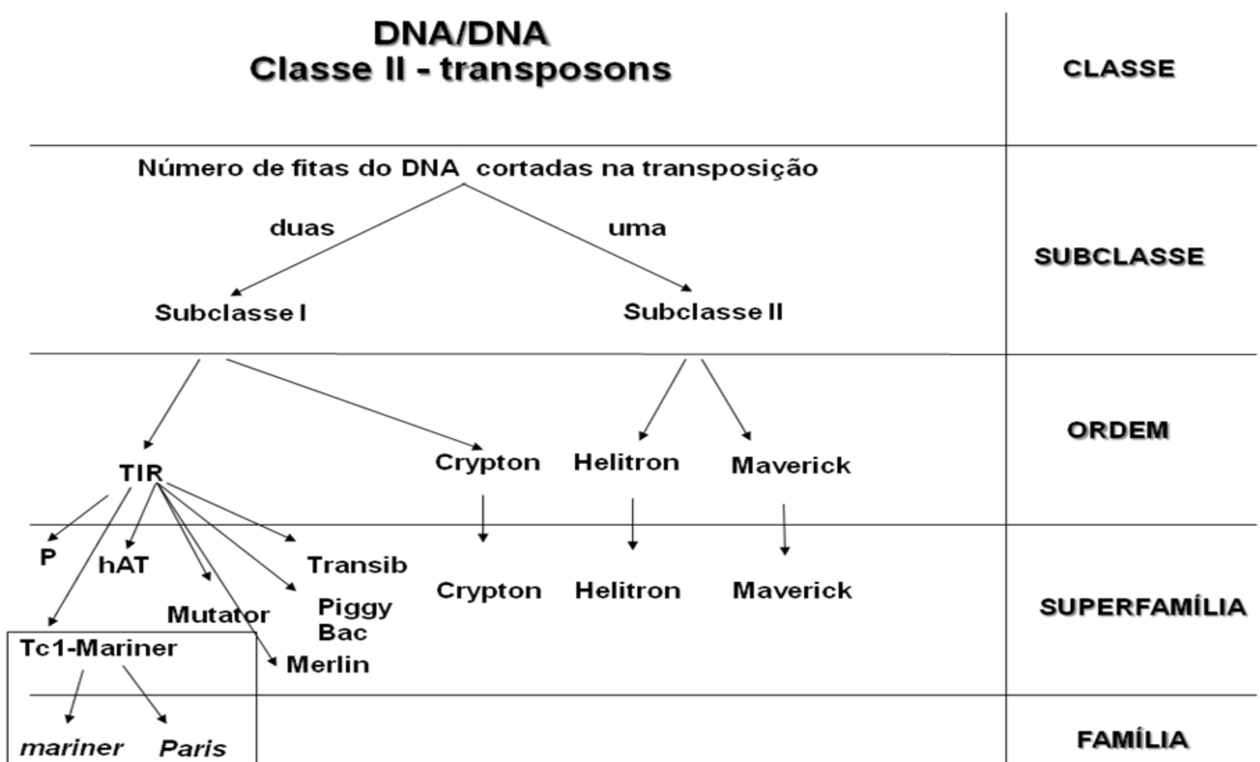


Figura 2: Classificação da Superfamília *Tc1-Mariner* e os respectivos elementos estudados nesse trabalho.

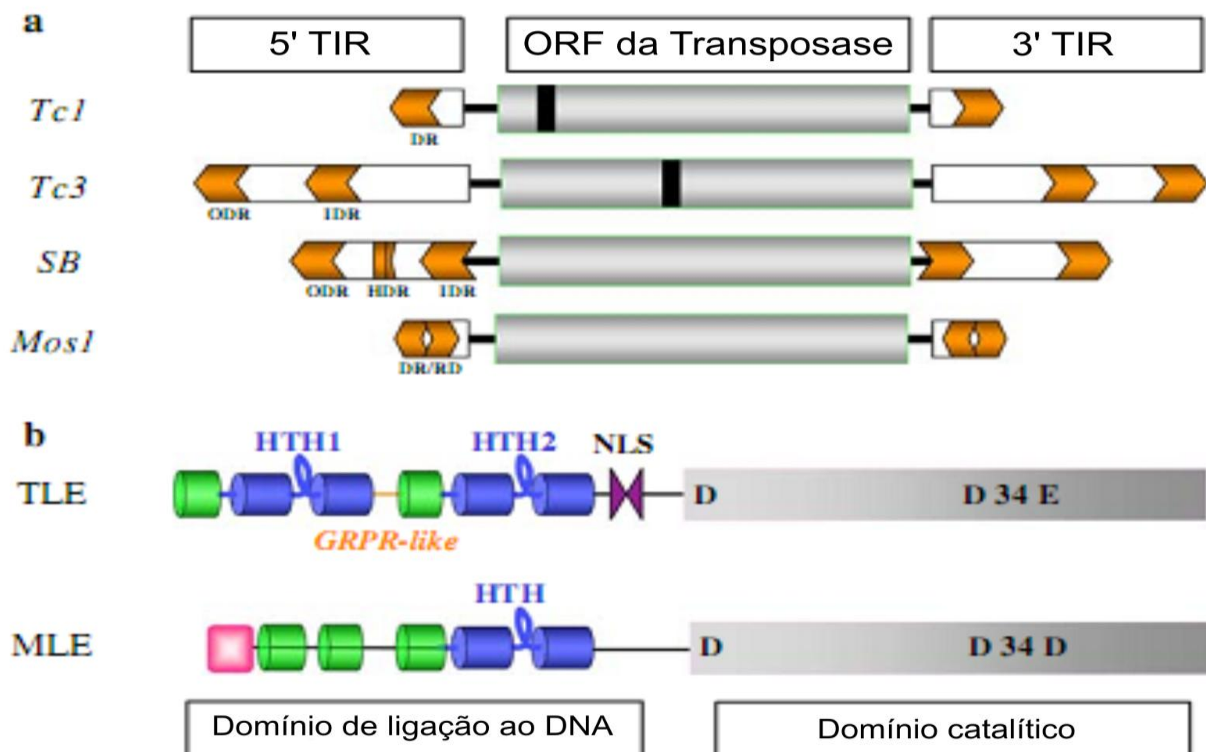


Figura 3: Transposons e transposases de elementos da superfamília *Tc1-Mariner*. a) Genes que codificam a transposase (caixas cinza) com ou sem um "*intron"* (caixas pretas) são flanqueados por TIRs (flechas brancas). Dentro das TIRs existem um ou mais sítios de ligação para a transposase (flechas laranja). b) Estrutura modular da transposase. HTH - Domínio de ligação da transposase ao DNA (caixas azuis); NLS – Sinal de localização nuclear; O domínio catalitico (caixas cinzas) contém a triade DDD e DDE características da superfamília *Tc1-Mariner*. Modificado de Brillet *et al.*, 2007.

Muitos elementos dessa superfamília já foram descritos. Como exemplos, podemos citar: *minos, S, Paris, mariner, Bari1, Sleep-Beauty*, *Tc1,* entre outros. O fato desses elementos poderem se transpor usando somente a sua transposase e não necessitarem de fatores específicos do hospedeiro para a sua transposição, faz com que sejam uma ferramenta genética atrativa, podendo ser utilizados como vetores de transformação em uma ampla gama de espécies de procariotos e eucariotos (LAMPE, 2009, ROBERTSON *et al.*, 2002). Esses fatores também podem facilitar o processo de invasão desses elementos em novos genomas, explicando, em parte, os vários eventos de TH já descritos para essa superfamília. Os elementos *minos* e *mariner* são exemplos clássicos de elementos que se transferem horizontalmente, entre espécies próximas e distantemente relacionadas (ALMEIDA E CARARETO, 2005, BRUNET *et al.,* 1994, 1999, MARUYAMA E HARTL, 1991, LAHA *et al.*, 2007, LOHE *et al.,* 1995, ROBERTSON E LAMPE, 1995).

A estrutura simples dos elementos dessa superfamília permite que sejam determinados muitos detalhes de seus mecanismos de transposição como: a estrutura tridimensional da transposase de alguns TEs ativos (VAN POUDEROYEN *et al.*, 1997, RICHARDSON *et al.*, 2006, 2009) e quais fatores influenciam a transposição *in vitro* (LAMPE *et al.*, 1998, MUÑOS-LOPEZ *et al.*, 2008, SINZELLE *et al.*, 2008). Além disso, linhagens de *Drosophila*, com uma mutação na coloração do olho causadas pelo elemento *mariner*, permite sugerir quais fatores evolutivos influenciam a atividade desse elemento em populações naturais (PICOT *et al.*, 2008).

1.2. A Família Drosophilidae

A família Drosophilidae pertence à superfamília Ephydroidea da seção Acalyptratae (THOMPSON, 2005). Esta família é uma das mais diversas e amplamente distribuídas entre as mais de 150 famílias da ordem Diptera (YEATES E WIEGMANN, 2005), possuindo atualmente mais de 3900 espécies distribuídas em 76 gêneros (BÄCHLI, 2008). A origem da família Drosophilidae se deu em regiões tropicais do Velho Mundo a partir de ancestrais que utilizavam fungos e folhiço, como recurso alimentar. Contudo, a datação da dessa origem ainda é uma questão controversa. Segundo Throckmorton (1975), isso se deu há cerca de 50 milhões de anos (Eoceno), porém em um estudo mais recente, Grimaldi (1988) sugere uma datação de 70 milhões de anos.

1.2.1. O Gênero *Drosophila*

Entre os vários gêneros da família Drosophilidae, o gênero *Drosophila* se destaca pela diversidade de espécies. Esse gênero atualmente é constituído por aproximadamente 1149 espécies, subdivididas em oito subgêneros (BÄCHLI, 2008). Dois subgêneros (*Drosophila e Sophophora*) se destacam dentro do gênero *Drosophila* por sua diversidade de espécies e pela sua importância nas ciências naturais. Segundo Tamura *et al.* 2004, esses subgêneros divergiram há aproximadamente 63 milhões de anos.

O subgênero Sophophora é atualmente constituído por 332 espécies subdivididas em oito grupos. Os grupos *saltans* e *willistoni* representam as espécies do subgênero *Sophophora* presentes na região neotropical (Novo Mundo) (THROCKMORTON, 1975). As evidências acerca das relações filogenéticas do subgênero *Sophophora* sugerem que ele é monofilético e basal dentro do gênero *Drosophila* (DALAGE *et al.*, 2007, KATOH *et al.*, 2007).

Já o subgênero *Drosophila* é composto atualmente por 721 espécies subdivididas em 43 grupos. Esse subgênero constitui o maior representante da família Drosophilidae na região neotropical, onde é representado por 25 grupos de espécies. Diversos trabalhos têm apontado evidências sobre a merofilia do subgênero *Drosophila*.

Muitas espécies do gênero *Drosophila* vêm sendo utilizadas, há mais de um século, como um excelente modelo para estudos genéticos e evolutivos. A disponibilidade de 12 genomas de espécies deste gênero (CLARK *et al.*, 2007) e mais recentemente os vários genomas populacionais (http://www.dpgp.org/) tornaram estes organismos ainda mais atrativos para responder às questões sobre evolução de TEs. Além da disponibilidade dos genomas, o grande número de espécies com uma relação filogenética razoavelmente esclarecida (MARKOW E O`GRADY, 2006, ROBE *et al.,* 2005, TATARENKOV *et al.*, 2001), possibilita um amplo campo de estudos visando esclarecer os mecanismos gerais de evolução genômica.

1.3. Objetivos

O presente trabalho teve como objetivos gerais:

- Caracterizar a história evolutiva dos elementos relacionados à *mariner* em drosofilídeos neotropicais.

- Caracterizar a história evolutiva dos elementos relacionados à *Paris* nos doze genomas de *Drosophila*.

1.3.1 Objetivos específicos:

- Caracterizar quais subfamílias de elementos relacionados à *mariner* estão presentes em drosofilídeos neotropicais.

- Compreender os processos evolutivos que podem ter influenciado a distribuição atual dos elementos relacionados à *mariner* em drosofilídeos neotropicais.

- Caracterizar estruturalmente os elementos relacionados à *Paris* nos doze genomas de *Drosophila*.

- Descrever os elementos relacionados à *Paris,* potencialmente ativos.

- Compreender os processos evolutivos que podem ter influenciado a distribuição atual dos elementos relacionados à *Paris,* nos doze genomas de *Drosophila.*

CAPÍTULO II

The evolutionary history of *mariner*-like elements in Neotropical drosophilids

1. Abstract

The evolutionary history of *mariner*-like elements (MELs) in 49 drosophilids species, mainly Neotropical, is here described. Twenty-three species presented MELs of three different subfamilies in their genomes: eighteen species have MELs of subfamily *mellífera*, fifteen of subfamily *mauritiana* and three of subfamily *irritans*. Eleven of these species exhibited more than one subfamily in their genome. Two subfamilies showed sequences with intact ORF, conserved catalytic motifs and strong evidence of purifying selection acting on their sequences. The species with these putative active elements are *D. mediopunctata* and *D. busckii*, for the *mauritiana* subfamily and *D. paramediostriata* for the *mellifera* subfamily. These characteristics suggest that these sequences are probably derived from active elements. In phylogenetics analysis of MELs we found a complex evolutionary pattern with vertical transfer, stochastic loss and evidences of putative events of horizontal transmissions occurring between different Drosophilidae species and even Drosophilidae and species belonging to taxa more distantly related as: *Bactrocera tryoni* (Tephritidae family), *Sphyracephala europaea* (Diopsoidea superfamily) and *Buenoa sp.* (Hemiptera order).

Key Words: *mariner*-like elements, vertical transfer, horizontal transfer, Neotropical, *Drosophila*, Drosophilidae

2. Introduction

Almost all organisms investigated so far have a fraction of their genomes composed by transposable elements (TEs). Generally these elements are considered genome parasites, nevertheless there may exist a continuum between extreme parasitism to mutualism for the relationship between the TEs and their hosts (Volff 2006). A recognized characteristic of TEs is that they can move within and between genomes causing great impact to the host, as for example, large-scale structural changes, epigenetic regulatory modifications, production of allelic diversity, new genes and biological innovations (Feschotte and Pritham 2007).

TEs are divided in two main classes depending on their transposition mechanism (Finnegan 1989). Class I or retrotransposons are elements that use a transposition mechanism called 'copy-and-past' with a RNA intermediate and Class II or transposons are elements that transpose by mechanism called 'cut-and-past' with a DNA intermediate (Wicker et al. 2007).

mariner family, also called *mariner*-like elements (MELs) or *ItmD34D* is one of the best known Class II elements family. They belongs to the order TIR, superfamily *Tc1-mariner* and is found in a wide variety of organism as plants (Feschotte and Wessler 2002, Feschotte et al. 2003) and metazoan (Augé-Gouillou et al. 1995, Garcia-Fernàndez et al. 1995, Ren et al. 2006, Robertson and MacLeod 1993, Silva et al. 2005, Sinzelle et al. 2006). The broad distribution of MELs can be explained in part for their incredible characteristic to move between genomes, a phenomenon known as Horizontal Transfer (HT) (Brunet et al. 1999; Maruyama and Hartl 1991b; Laha et al. 2007, Lampe et al. 2003, Robertson 1997, Robertson and Lampe 1995; Yoshiyama et al. 2001).

The *mariner* elements are generally 1.3 Kb long and have terminal inverted repeats (TIRs) with around 28 bp. They contain a unique ORF with a protein aspartic triad DD34D. Due to the great variability of MELs, they are classified in several subfamilies according to phylogenetic analysis. The major subfamilies are *mauritiana*, *cecropia*, *irritans*, *mellifera* and *capitata* (Robertson and MacLeod 1993). However there are other minor subfamilies with more limited distribution (Robertson et al. 2002, Rouault et al. 2009). Some of these subfamilies encompass some autonomous elements (as the *mauritiana* subfamily with the *Mos1* element and the *mellifera* subfamily with the *Famar1* element), although most *mariner* elements found to date harbor inactivating point mutations or framesshifts (Muñoz-López et al. 2008).

The first described *mariner* element (*Dromar*) and the first described active element (*Mos1*), were isolated from *Drosophila mauritiana* (Jacobson et al. 1986, Medhora et al. 1988) and belong to the *mauritiana* subfamily. The subsequent studies for related sequences in Drosophilidae species were performed mainly by hybridization assays using *Dromar* or *Mos1* as probes (Biémont and Cizeron 1999, Brunet et al. 1994, 1999, Capy et al. 1992, Castro et al. 2006, Germanos et al. 2006, Loreto et al. 1998 and Maruyama and Hartl 1991a). These studies emphasized mainly the *Drosophila* genus, although, some species of *Zaprionus*, *Scaptomyza, Zygothrica*, *Chymomyza* and *Scaptodrosophila* were also analyzed. Within the *Drosophila* genus, the four *Sophophora* subgenus species group (*melanogaster*, *obscura*, *willistoni* and *saltans*) contain elements of the *mauritiana* subfamily. Moreover, for the *melanogaster* group two more subfamilies were found: *irritans* in *D. ananassea* and *D. melanogaster* (Augé-Gouillou et al. 1999, Robertson and Lampe 1995) and *mellifera* in some species of *melanogaster* subgroup (Lohe et al. 1995). The elements of these subfamilies present patchy distributions and incongruences between the phylogeny of TEs and the phylogeny of their host species, suggesting some events of HT. Within the *Drosophila* subgenus only the *mauritiana* subfamily was found in 5 species: two

from the *virilis* group, one from the *repleta* group and two from the *immigrans* group (Biémont and Cizeron 1999). Besides, some *Zaprionus* and *Scaptomyza* species, also present elements of the *mauritiana* subfamily (Brunet et al. 1994, Maruyama and Hartl 1991a, 1991b).

Based on the distribution of the *mauritiana* subfamily elements in endemic species of Drosophilidae, Brunet et al. (1994) described that MELs are mainly present in Asia and Africa and proposed two hypotheses, not mutually excluded, to explain this pattern: (1) it may result from the introduction of this element through HT in species endemic to these geographical regions. In this case the HT could have occurred between *Drosophila* species and/or between *Drosophila* species and one or several donor species outside the Drosophilidae family; (2) it may correspond to the evolution of the *mariner* element from an ancestral copy which was present in the Drosophilidae ancestor being lost in some lineages.

In the present paper, we have studied the evolutionary history of *mariner* family in the 49 Neotropical Drosophilidae species aiming to increase our comprehension about the evolution of these TEs. Twenty-three species have MELs in their genomes, and these belong to three different subfamilies: eighteen species have MELs from the *mellifera* subfamily, fifteen from the *mauritiana* subfamily and three from the *irritans* subfamily. Three species have putatively active elements: *D. mediopunctata* and *D. busckii,* for the *mauritiana* subfamily and *D. paramediostriata* for the *mellifera* subfamily. The phylogenetics analysis showed evidence of vertical transfer (VT) with stochastic losses and some putative HT events.

3. Materials and Methods

PCR, cloning and sequencing

Total DNA from 49 drosophilids species (Suppl. Table 1) was prepared according Oliveira *et al.* (2009). Degenerated *primers* designed by Robertson and MacLeod (1993) on the conserved motifs of the transposase catalitic domain were used to amplify a fragment with approximately 491 bp. The components of the 50µl reaction mixture were 100ng DNA, 1 U Taq polymerase, 5ul 10x reaction buffer supplied by the manufacturer (Invitrogen, Carlsbad, CA, USA), 200µM of each nucleotide, 20 pmol of each primer and 1,5 mM $MgCl_2$. Amplification parameters were 95 ºC for 1 min, 40 cycles at 95 ºC for 30 s, 50ºC for 30 s, 75 ºC for 30 s, followed by an extension cycle at 72 ºC for 7 min. PCR products were fractioned by agarose gel 0.8%, when the "491-bp" fragment was removed from the gel using the "glass milk" protocol described by Oliveira et al. (2009). The purified bands were cloned using the PCR2.1-TA Cloning Kit (Invitrogen, Carlsbad, CA, USA). DNA sequencing was carried out directly from the purified plasmids in the 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 and assembled using the GAP 4 software of the Staden Package (Staden 1996).

Dot Blot

To analyze if degenerated primers could have been underestimating the number of species that have MELs, about 1µg of genomic DNA from different Drosophilidae species was transferred to Hybond N+ nylon membranes (GE Healthcare), fixed at 80ºC for 2 h and hybridized with probes for different *mariner* subfamilies. The probes used composed

clones, described in this paper, representing the three subfamilies (*mellifera*, *mauritiana* and *irritans*). For each subfamily a pool of probes representative of the subfamilies variability was used: *mellifera* probes (DTTMarMel_Dparam7, DTTMarMel_Dcardi8, DTTMarMel_Slatif18 and DTTMarMel_neoca15), *mauritiana* probes (DTTMarMau_Dmedpu20, DTTMarMau_Dbusck3 and DTTMarMau_Dparam11) and *irritans* probes (DTTMarIrr_Dtripu7 and DTTMarIrr_Dbuski21). For hybridization and detection the Alkphos Direct Labeling Reagent kit (GE Healthcare) was used at high stringency conditions, according to the manufacturer's instruction. The membranes were washed twice with SSC 0.2X and 0.5% SDS (w/v) for 15 min at 65 °C.

Distance, dS, dN/dS estimates and codons bias

The softwares ClustalX2 (Thompson et al. 1997) and Muscle 3.2 (Edgar 2004) were used to align the nucleotide and amino acids sequences, respectively. The Cons software from the EMBOSS suite (http://emboss.sourceforge.net/) available in Mobyle Portal (Neron et al. 2009) was used to obtain the consensus of sequences from the same species that showed less than eight percent of divergence and were monophyletic in a previous Bayesian analysis. The amino acid sequences were deduced using GeneDoc 2.6.001 software (Nicholas et al. 1997).

The Hasegawa-Kishano-Yano evolutionary model with a gamma correction (HKY+G) was selected by ModelTest 3.7 (Posada and Crandall 1998) and used in the nucleotide distance analysis, as performed in Paup 4.0b10 (Swofford, 2003). The amino acid distance analysis was performed in MEGA 4.0.1 (Tamura et al. 2007) with the Jones-Taylor-Thornton model with a gamma correction (JTT+G) substitution model, as appointed by ProtTest 1.3 (Abascal et al. 2005).

The codon alignment of each subfamily was used to estimate the synonymous distance values (dS) and the number of synonymous sites (S) using the Nei and Gojobori (1986) model assisted by the MEGA 4.0.1 software (Tamura et al. 2007). In order to allow these calculations, gaps were introduced in order to conserve the reading frames of some sequences while stop codons were considered as absent information. The TEs dS values were compared with these presented for other host genes through a Fisher exact test (Ludwig et al 2008). Three genes were used for these comparisons: α-metildopa (Amd), cytochrome oxidase II (COII) and dopa-decarboxylase (Ddc).

To analyze the evolutionary force that is acting on the MELs we performed a dN/dS Codon-Based Z Test, as implemented in MEGA 4.0.1 (Tamura et al. 2007). The effective number of codons (Nc; Wright, 1990), GC content and the codon bias index (CBI; Morton, 1993) were calculated by DnaSP 5.0 software (Librado and Rozas 2009)

Phylogenetics analysis

Bayesian analysis was performed in MrBayes 3.2 software (Ronquist and Huelsenbeck 2003) using amino acid sequences and the JTT+G evolutionary model selected by ProtTest 1.3 (Abascal et al. 2005). A total of 1,000,000 generations were evaluated, with sampling every 100 generations and burning 25% of the trees. As outgroup we used the sequence DTTmaT_Bmori (AB006196.1) of *Bombix mori*, that belongs to the *maT* family (Claudianos et al. 2002). Moreover, some sequences that showed a high nucleotide identity in Blast search with the sequences obtained in this work, were used to identify the different *mariner* subfamilies: *mellifera* (DTTMarMel_Derect (U08094.1), DTTMarMel_Btryon (AF346541), DTTMarMel_Danana (U91359.1), *mauritiana* (DTTMarMau_Seurop (EF407579), DTTMarMau_Dsimul (AF037052.1),

DTTMarMau_Dmauri (M14653), DTTMarMau_Danana (U91360.1), DTTMarMau_Tdalma (DQ197022.1), *capitata* (DTTMarCap_Ccapit (U04467.1), DTTMarCap_Hvulga (U51183.1) *irritans* (DTTMarIrr_Danana (U11647.1), DTTMarIrr_Cplora (U11653.1), DTTMarIrr_Bsp. (U91352.1), *vertumana* (DTTMarVer_Evertu (U04458.1), DTTMarVer_Bneohu (AF348438.1), *lineata* (DTTMarLin_Nlinea (U91362.1) and *cecropia* (DTTMarCec_Hsapie (U52077.1)).

The clone names were attributed following Wicker *et al.* (2007) with some modifications proposed by Rouault *et al.* (2009). The name for each sequences are 3 initial letters code DTT that characterizes the *Tc1-Mariner* superfamily, followed by the 3 letters characterize the *mariner* family (Mar) and the 3 letters designates the subfamilies (Mel – *mellifera*, Mau – *mauritiana*, Cap – *capitata,* Irr – *irritans*, Ver – *vertumnana*, Lin – *lineata* and Cec - *cecropia*). After this, the name further encompasses a capital word, five lower case words and a number, representing the genera, specific epithet and order for which the clone was obtained, respectively (e.g. DTTMarMel_Dparam7). When consensus sequences were constructed the clone number was substituted by the consensus number representing the order for which is was made (e.g. DTTMarMel_Dparam1cons).

4. Results:

Species with mariner sequences and putative actives elements

Forty nine mainly Neotropical Drosophilidae species were analyzed by PCR. Fourteen of these showed MELs and for these species, an amount of 62 clones were sequenced for them. These elements belong to three *mariner* subfamilies (*mellífera,*

mauritiana and *irritans*). The name of different clones, the consensus building using similar elements and the GenBank accession number for those sequences can be found in the Supplementary Table 1. Additionally, nine other species showed MELs when analyzed by Dot Blot. Thus, an amount of 23 species (48%) showed MELs.

The more representative subfamily found among the analyzed species was the *mellifera* (in 18 species). The second was the *mauritiana* (15 species) and the *irritans* subfamily was present in three species (Table 1).

Three species, *D. mediopunctata, D. busckii* and *D. paramediostriata,* showed sequences with intact ORF and the catalytic motifs in the analyzed region, suggesting these elements can be active in these species. In *D. busckii* there are one amino acid substitution in the catalytic motif (WVPHEL/YS(T)PDLAP), however the amino acids S (Serine) and T (Threonine) have similar physical chemistry properties. The putatively active sequences of *D. mediopunctata* and *D. busckii* belong to *mauritiana* subfamily. Nevertheless, the *D. paramediostriata* sequence belongs to *mellífera* subfamily. The sequences having intact ORFs correspond to clones DTTMarMau_Dmedpu10, DTTMarMau_Dmedpu20, DTTMarMau_Dmedpu23, DTTMarMau_Dbuski3, DTTMarMel_Dparam7 and DTTMarMel_Dparam29 (Suppl. Table 2).

To verify if the putatively active sequences are under purifying selection, these sequences were aligned with known active elements belonging to *mauritiana* (*Mos1* element) and *mellifera* (*Famar1* from *Forlicula auriculata)* subfamilies, being the analyzed for the ratio between the non synonymous rate and the synonymous rate (dN/dS). All comparisons showed that the dS are significantly higher than dN (p = 0,000), suggesting these sequences are under purifying selection.

Phylogenetics analysis and putative horizontal transfer events

The obtained sequences subdivided into three major clades which correspond to the *mariner* subfamilies *mellifera*, *mauritiana* and *irritans* (Fig. 1).Within *mellifera*

Table 1. Distribution of MELs in species of family Drosophilidae.

Genus Subgenus	Section	Group	Species	Hybridization Analysis					Sequences
				Literature data of hybridization analysis[a]	*maurtiana*[b]	*mellifera*[c]	*irritans*[d]	PCR[e]	Nº clones[f]
Drosophila *Drosophila*	*quinaria-tripunctata*	*guarani*	*D. ornatifrons*		+	+	-	+	2 (*mellífera*)
			D. guaru		+	+	-	-	
			D. subbadia					-	
		guaramuru	*D. maculifrons*		+	-	-	+	8 (*mellífera*) 1 (*mauritiana*)
			D. griseolineata		-	+	-	+	
		tripunctata	*D. nappae*		-	+	-	-	
			D. paraguayensis					+	2 (*mellífera*)
			D. mediopunctata	-[*]				+	7 (*mauritiana*) 1 (*irritans*)
			D. crocina		+	-	-	-	
			D. mediostriata	-[*]				+	4 (*mellífera*)
			D. paramediostriata					+	3 (*mellífera*) 2 (*mauritiana*)
			D. mediopictoides					-	
			D. bandeirantorum					+	
			D. tripunctata					+	6 (*mellífera*) 7 (*irritans*)
			D. mediodifusa		+	+	-	-	
			D. mediopicta	-[*]				-	
		cardini	*D. cardini*					+	1 (*mellífera*)
			D. neocardini					+	3 (*mellífera*) 1 (*mauritiana*)
			D. cardinoides		+	-	-	-	
			D. parthenogenética		+	-	-	-	
			D. procardinoides					-	
			D. arawakana	-[&]				+	2 (*mellífera*)
			D. dunni		+	+	-	-	
			D. polymorpha	-[*]	+	-	-	-	
		pallidpennis	*D. pallidpennis*					+	1 (*mellífera*)
	virilis-repleta	*mesophragmatica*	*D. gasisi*	-[¢]				+	
			D. brncici	-[¢]	-	-	-	-	
			D. gaucha	-[*]				-	
			D. pavani	-[¢]	-	-	-	-	

				Hybridization Analysis					Sequences
Genus Subgenus	Section	Group	Species	Literature data of hybridization analysis[a]	*maurtiana*[b]	*mellifera*[c]	*irritans*[d]	PCR[e]	Nº clones[f]
		repleta	*D. hydei*	-[*]	-	-	-	-	
			D. buzzati	-[£]	-	-	-	-	
			D. onca	+[&]				-	
			D. repleta		-	-	-	-	
			D. mercatorum	-[£] -[*]	-	-	-	-	
		immigrans	*D. immigrans*	+[#] -[*]	-	-	-	-	
		flavopilosa	*D. incompta*					+	
Dorsilopha			*D. buskii*	-[£] +[#]				+	2 (*mauritiana*) 5 (*irritans*)
Sophophora		*willistoni*	*D. nebulosa*	-[£] -[*] -[&]	-	-	-	+	
			D. willistoni	-[*]	-	-	-	+	
			D. fumipennis					+	
			D. paulistorum	-[£] -[*] -[&]	-	-	-	+	1 (*mauritiana*)
			D. tropicalis					-	
			D. equinoxialis	-[*]	-	-	-	-	
			D. succinea					-	
			D. insularis	-[*]	-	-	-	+	
			D. capricorni	-[£] -[*] -[&]				-	
		saltans	*D. sturtevant*	-[£] +[#] -[*] -[&]				-	
Scaptodrosophila		*rulifrons*	*S. galloi*					-	
			S. latifasciformes					+	5 (*mellífera*)

[a] Literature data of Hybridization analysis: (&) Bièmont e Cizeron 1999, (#) Brunet et al. 1994, (§) Castro et al. 2006, (¢) Germanos et al. 2006, (*) Loreto et al. 1998 e (£) Maruyama and Hartl 1991 using as probe the *Mos1* element of *D. mauritiana.* (+) evidence of hybridizing signal, (-) no hybridizing signal.

[b] Dot Blot of *mauritiana* subfamily using as probes the sequences (DTTMarMau_Dparam11, DTTMarMau_Dbusck3 and DTTMarMau_Dmedpu20).

[c] Dot Blot of *mellífer* subfamily *a* using as probes the sequences (DTTMarMel_Dparam7, DTTMarMel_Dcardi8, DTTMarMel_Slatif18 and DTTMarMel_Dneoca15).

[d] Dot Blot of *irritans* subfamily using as probes the sequences (DTTMarIrr_Dtripu7 and DTTMarIrr_Dbusck21).

[e] Amplification of fragment with approximately 491 bp by PC: (+) positive amplification, (-) negative amplification.

[f] Numbers of clones sequenced for each subfamily.

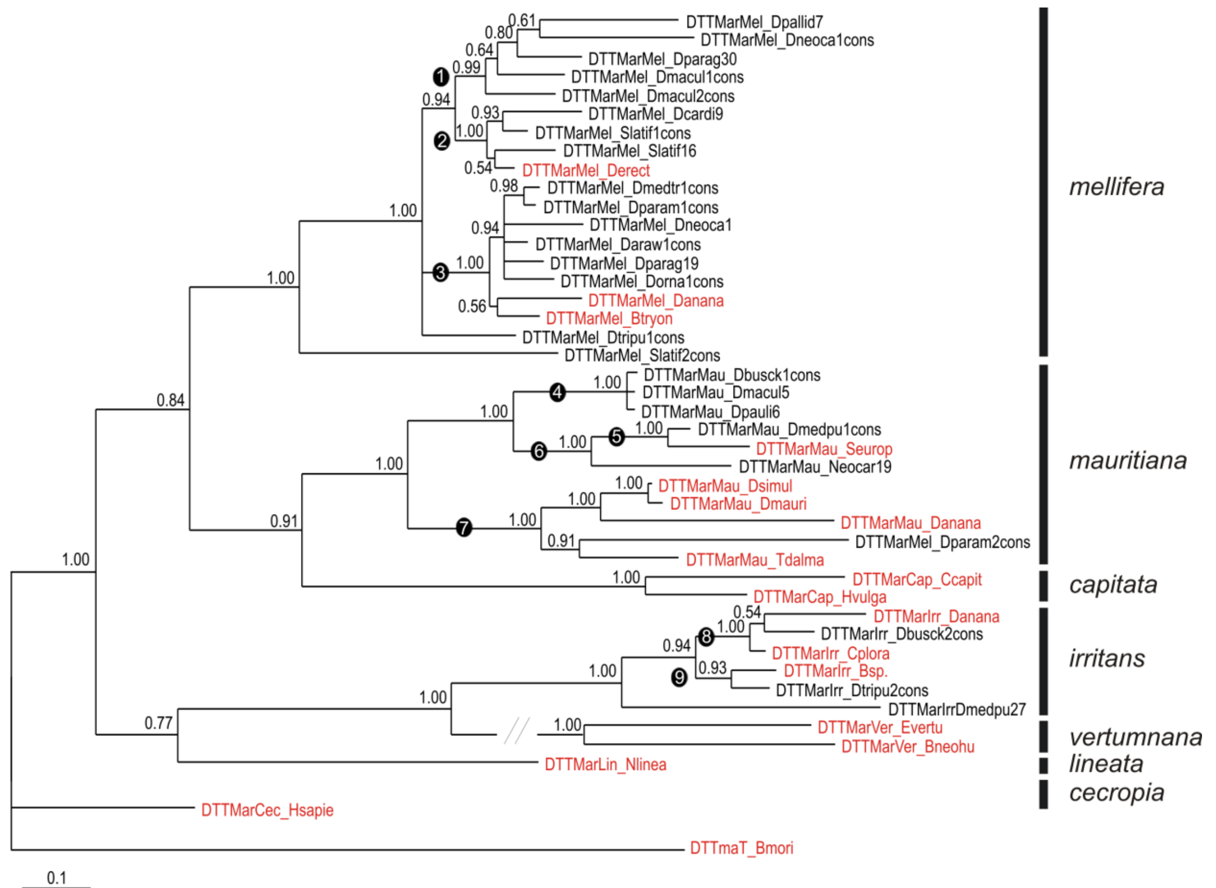


Figure 1: Phylogeny of *mariner*-like sequences using a Bayesian analysis with the JTT+G model. Posterior probabilities of clades are show next to the branches. The names in red are sequences obtained in databank (NCBI) and in black are sequences produced in this study. The black bars in right of figure are the subfamilies of *mariner* Family. The black circles with a white number are the clusters cited in the text.

subfamily, nucleotide distances ranged from 0.025 to 0.926, Whereas the amino acid distances ranged from 0.091 to 1.959 (Suppl. Table 3). The elements included in the *mauritiana* subfamily showed a nucleotide divergence varying from 0.016 to 1.199 and an amino acid divergence 0.018 to 1.495. The *irritans* subfamily showed sequences which the nucleotide divergence fluctuating from 0.044 to 0.594 and for the amino acid divergence from 0.280 to 2.036.

Some incongruences can be observed in the comparison of *mariner* phylogeny (Fig. 1) with the host species phylogeny (Robe et al. 2005, 2010; Hatadani et al. 2009). For example: in the *mellifera* subfamily the cluster 2 with *S. latifasciformis, D. erecta* and *D. cardini* sequences and cluster 3 with *D. ananassae* (Diptera; Drosophilidae family) and

Bactrocera tryoni (Diptera, Tephritidae family); in the *mauritiana* subfamily the cluster 7 with *D. paramediostriata* (Diptera, Drosophilidae family) and *Teleopsis dalmanni* (Diptera, Diopsidae family) and cluster 5 with *D. mediopunctata* (Diptera, Ephydroidea superfamily) and *Sphyracephala europaea* (Diptera, Diopsoidea superfamily); in the *irritans* subfamily the cluster 9 shows TE sequences of *D. tripunctata* (Diptera order) and *Buenoa sp.* (Hemiptera order). Incongruences between TE´s and hosts species phylogenies can be indicative of horizontal transfer (Robertson and Lampe 1995, Loreto et al 2008). In last two cases, from *mauritiana* and *irritans* subfamilies, these species do not have gene sequences available for comparison, however the low distance (0,152 and 0,193) and dS value (0,309 and 0,422) found for TEs when contrasted with the estimate of last common ancestor shared between these species, in early Senonian (80-85 Mya) (Labandeira 2005) and late Devonian (373.0-382.9 Mya) (Gaunt and Miles 2002) are difficult to explain by vertical transmission (VT).

Other analyses that can be used to infer HT is the comparison of dS values between TEs and host genes. dS values offer a measurement of neutral evolution when a strong codon usage bias is absent. Consequently, when two species are compared for this parameter we expect the same proportion of divergence in TEs and host genes. Alternatively, if the dS values of TEs are significantly lower than dS for the host genes, reflecting the lower divergence time of the TE sequences when compared to the species divergence time, it can be seen as an indicative of HT (Ludwig et al. 2008). The host genes chosen for comparison were the Amd, Ddc and COII. These genes do not shown higher codon usage bias much larger than TEs which is a condition to use these genes to infer HT (Suppl. Table 4).

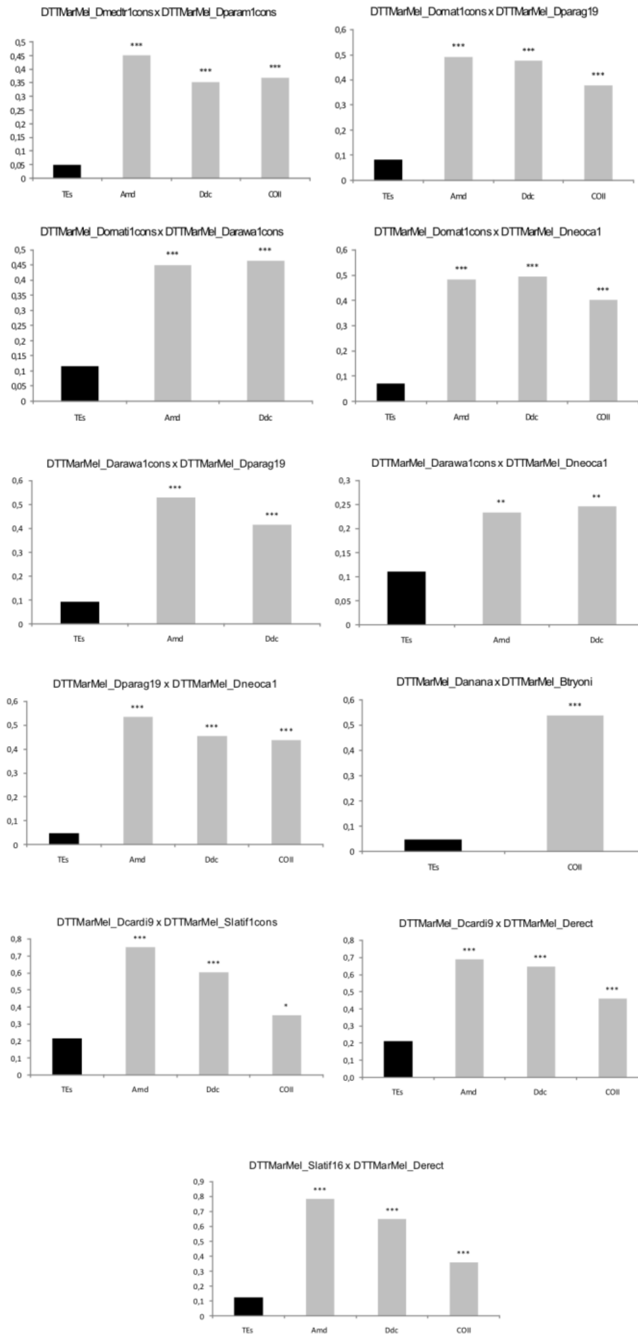
With the dS value analysis many significant comparisons were found, suggesting HT, most of them occurring between Drosophilidae species. However one case occurred between Drosophilidae and distantly related taxa.

The *mellífera* subfamily comprises eleven putative HT significant comparisons: in the cluster 2 one between the *cardini* group of *Drosophila* subgenus and *melanogaster* group of *sophophora* subgenus (*D. cardini* and *D. erecta*) and two events between the *Drosophila* genus and *Scaptodrosophila* genus (*D. cardini* and *Scaptodrosophila latifasciformes, D. erecta* and *S. latifasciformes*); in the cluster 3 one between two species of *tripunctata* group (*D. mediostriata* and *D. paramediostriata*), one between species of *guarani* group and *tripunctata* group (*D. ornatifrons* and *D. paraguayensis*), two events between the *guarani* group and *cardini* group (*D. ornatifrons* and *D. arawakana*, *D. ornatifrons* and *D. neocardini*), two events between *cardini* group and *tripunctata* group (*D. arawakana* and *D. paraguayensis*, *D. paraguayensis* and *D. neocardini*), one event between species of *cardini* group (*D. arawakana* and *D. neocardini*) and a event between *D. ananassae* (Family Drosophilidae) and *Bactrocera tryoni* (Family Tephritidae) (Figure 2A). In this last case the families shared the last common ancestor in early Senonian (80-85 Mya) (Labandeira 2005), however the MELs these species show a low distance (0,095) and dS value (0,047).

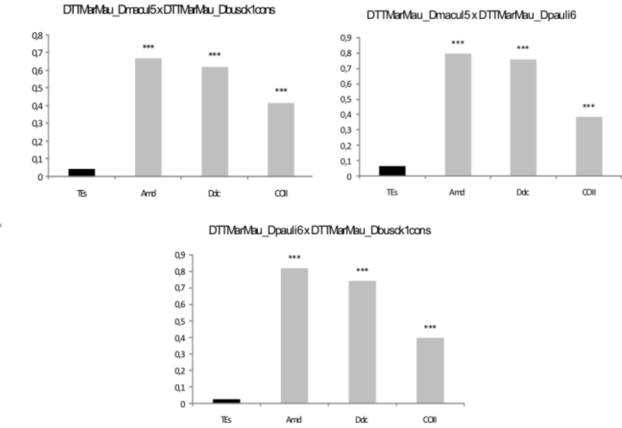
The *mauritiana* subfamily presents three significant comparisons: in the cluster 4, one between the *guaramuru* group and *willistoni* group (*D. maculifrons* and *D.* paulistorum), and two events between the *Drosophila* subgenus and *Dorsilopha* subgenus (*D. maculifrons* and *D. busckii*, *D. paulistorum* and *D. busckii*) (Figure 2B).

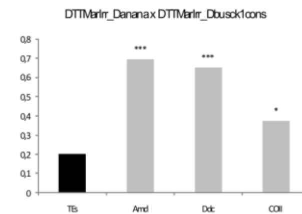
The *irritans* subfamily comprises one significant comparison in the cluster 8 between *Drosophila* subgenus and *Dorsilopha* subgenus (*D. ananassae* and *D. busckii*) (Figure 2C).

A



B





C

DTTMarIrr_Danana x DTTMarIrr_Dbusck1cons

Figure 2: Significant dS comparisons between TEs and host genes as tested by Fisher exact test. (*) p < 0,05; (**) p < 0,01; (***) p < 0,001; A) Significant dS comparision present in *mellifera* subfamily: right column: *D. mediostriata/D. paramediostriata*, *D. ornatifrons/D. arawakana*, *D. arawakana/D. paraguayensis, D. paraguayensis/D. neocardini, D. cardini/S. latifasciformis,* left colomn: *D. ornatifrons/D. paraguayensis, D. ornatifrons/D. neocardini, D. arawakana/D. neocardini, D. ananassae/B. tryoni, D. cardini/D. erecta* and *S. latifasciformis/D. erecta*; B) Significant dS comparision present in *mauritiana* subfamily: *D. maculifrons/D. busckii*, *D. maculifrons/D. paulistorum* and *D. paulistorum/D. busckii;* , C) Significant dS comparision present in *irritans* subfamily: *D. ananassae/D. busckii*.

5. Discussion:

The broad distribution of MELs is generally attributed to their great capacity to perform horizontal transfer (Lohe et al. 1995, Robertson 1993, Robertson et al. 2002). This property can explain the presence of similar elements in the genomes of distantly related species (Laha et al. 2007, Robertson 1997). Nevertheless, the relative contribution of VT and HT for maintenance and evolution of MELs in the genomes is not completely understood. The Drosophilidae family is a model for evolutionary studies and the phylogenetic relationship among the species is now well known (Markow and O'Grady 2006). The colonization of American territory for the Drosophilidae should have occurred around 22 to 28 Mya and the major lineages of the *Drosophila* genus diversified in the Neotropical region between 15-25 Mya (Robe et al. 2010, in press). The analysis of MELs evolutionary history in Neotropical Drosophilidae can be useful to understand the role of VT and HT of *mariner* in the genomes.

The evolutionary analysis of MELs sequences present in the Neotropical Drosophilidae show the many cases are compatible with VT. For example, in the cluster 1 of *mellifera* subfamily elements found in the groups *guarani*, *cardini*, *tripunctata* and *pallidipenis* are probably transmitted vertically. These species diversified in the Neotropical region after Miocen (Robe et al, 2010, in press). Therefore, these elements probably are an old component of the genome of these species. In the cluster 6 of *mauritiana* subfamily, the sequences found in the *tripunctata* and *cardini* groups are also an example of VT to evolutionary time like the previously described. This pattern suggests that VT is important for maintenance and diversification of these TEs in genomes.

Putatives cases of HT are also found in the evolutionary history of MELs in Neotropical Drosophilidae. These cases are new related HT events between Drosophilidae

species and between these and distantly related species. These events were tested comparing the evolutionary divergence, based in dS comparison, between MELs and nuclear and mitochondrial genes. Both comparisons suggest the HT events. The ability of colonizing new genomes by HT is a well recognized characteristic of MELs (Garcia-Fernàndez et al 1995, Laha et al. 2007, Lampe et al. 2003, Robertson 1997, Robertson and Lampe 1995, Robertson and MacLeod 1993).

As described by Robertson and MacLeod (1993), the use of degenerated primer to search MELs can result in an underestimation once some divergent elements cannot be amplified by the primers. The hybridization analysis can solve, at least partially, this problem. In agreement with this our hybridization assays have shown the presence of some MELs in species with negative signal in our PCR screen. However, even considering as possessing MELs the species that have been positive by PCR or by Dot Blot, the distribution of these elements in Neotropical Drosophilidae is greatly patchy. At least for the *Drosophila* groups in that we have evidences of ancient VT, the patchy distribution is better explained by stochastic loss.

The analysis concerning distribuition of MELs in Neotropical species of Drosophilidae showed that the subfamily *mellifera* has a wider distribution than found by Lohe et al. 1995. The subfamily *mauritiana* is also widely distributed in Neotropical Drosophilidae in contrast with previous studies (Castro et al. 2006, Loreto et al. 1998), however the subfamily *irritans* presents a restricted distribution. The coexistence of two or more subfamilies in a same genome is a phenomenon that has been previously observed for the *mariner* family (Robertson 1993, Carr 2008). However as Drosophilidae species are concerned, so far this phenomenon was found only in some species of *melanogaster* species group (Lohe et al. 1995). Nevertheless, our data showed that co-existence of multiple *mariner* subfamilies within a same genome appears to be a common pattern in Drosophilidae.

The subfamilies *mellifera* and *mauritiana* have some putatively active elements. From our knowledge, the *mellifera* subfamily sequence found in *D. paramediostriata* may be the first active element of this subfamily described in *Drosophila*. In the *mauritiana* subfamily the only described *Drosophila* active element, the *Mos1*, was found in *D. mauritiana* (*sophophora* subgenus). In this study, putatively active elements for the *mauritiana* subfamily were also found in *D. mediopunctata* (*Drosophila* subgenus) and *D. busckii* (*Dorsilopha* subgenus). The coexistence of multiples MELs in the same genome and the existence of putatively active elements arise the question about possible cross mobilization of these TEs. Further characterizations of the putatively active elements here described are interesting for testing the hypothesis concerning interaction within and between subfamilies in related taxa.

So in general, our data shows that MELs are widely distributed within the *Drosophila* subgenus and corroborate the hypothesis of Brunet et al. (1994), which suggest that MELs probably evolved from ancient elements that were present in the ancestor of the Drosophilidae family with eventual loss in some lineages and reacquirement through HT.

6. References:

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21(9): 2104-2105.

Augé-Gouillou C, Bigot Y, Periquet G (1999) Mariner-like sequences are present in the genome of the fruitfly, Drosophila melanogaster. J.Evol. Biol. 12: 742-745.

Augé-Gouillou C, Bigot Y, Pollet N, Hamelin M, Meunierrotival M, Periquet G (1995) Human and other mammalian genomes contain transposons of the *mariner* family. FEBS Letters 368: 541-546.

Carr M (2008) Multiple subfamilies of mariner transposable elements are present in stalk-eyed flies (Diptera: Diopsidae). Genetica 132: 113-122.

Claudianos C, Brownlie R, Russel R, Oakeshott J, Whyard S (2002) *maT* – a clade of transposon between *mariner* and *Tc1*. Mol. Biol. Evol. 19: 2101-2109.

Biémont C, Cizeron G (1999) Distribution of transposable elements in Drosophila species. Genetica 105: 43-62.

Brunet F, Godin F, Bazin C, Capy P 1999. Phylogenetic analysis of Mos1-like transposable elements in the Drosophilidae. J. Mol. Evol. 49: 760-768.

Brunet F, Godin F, David J, Capy P (1994) The mariner transposable element in the Drosophilidae family. Heredity 73: 377-385.

Capy P, David, J.R., Hartl, D.L. (1992) Evolution of transposable element *mariner* in the *Drosophila melanogaster* species group. Genetica 86:37-46.

Castro J, Setta N, Carareto CM (2006) Distribution and insertion numbers of transposable elements in species of the *Drosophila* saltans group. Gen. Mol. Biol. 29: 384-390.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5): 1792-1797.

Feschotte C, Pritham E (2007) DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41: 331-368.

Feschotte C, Swamy L, Wessler S (2003) Genome-wide analysis of mariner-Like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). Genetics 163: 747-758.

Feschotte C, Wessler S (2002) Mariner-like transposases are widespread and diverse in flowering plants. Proc. Nat. Acad. Sci. 99: 280-285.

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

Garcia-Fernàndez J, Bayascas-Ramírez J, Marfany G, Muñoz-Mármol A, Casali A, Baguñà J, 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.

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. 19(5): 748-761.

Germanos E, Mota NR, Loreto EL (2006) Transposable elements from the mesophragmatica group of Drosophila. Gen. Mol. Biol. 746: 741-746.

Hatadani LM, McInerney JO, Medeiros HF, Junqueira ACM, Azeredo-Espin AM, Klaczko LB (2009) Molecular phylogeny of the Drosophila tripunctata and closely related species groups (Diptera, Drosophilidae). Mol. Phyl. Evol. 51: 595-600.

Jacobson JW, Medhora MM, Hartl DL (1986) Molecular structure of a somatically unstable transposable element in Drosophila. PNAS 83:8684-8688.

Labandeira CC (2005) Fossil history and evolutionary ecology of Diptera and their assciations with plants. The Evolutionary Biology of Flies. Columbia University Press, New York. pp. 217-274.

Laha T, Loukas A, Wattanasatitarpa S, Somprakhon J, Kewgrai N, Sithithaworn P, Kaewkes S, Mitreva M, Brindley PJ (2007) The bandit, a new DNA Transposon from a hookworm—possible horizontal genetic transfer between host and parasite. PLoS Negl Trop Dis 1(1): e35.

Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson H (2003) Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages

representing four different orders shows that selection acts only during horizontal transfer. Mol. Biol. Evol. 20: 554-562.

Librado P, Rojas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. doi: 10.1093/bioinformatics/btp187.

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

Loreto E, Basso Da Silva L, Zaha A, Valente V (1998) Distribution of transposable elements in neotropical species of *Drosophila*. Genetica 101: 153-165.

Ludwig A, Loreto ELS (2008) Multiple invasions of *Errantivirus* in the genus *Drosophila.* Insect Mol. Biol. 17(2): 113-124.

Markow TA, O`Grady PM (2006) DROSOPHILA A guide to species identification and use. Elservier, UK, 259p.

Maruyama K, Hartl DL (1991a). Evolution of the Transposable Element mariner in Drosophila Species. Genetics 128: 319-329.

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

Medhora MM, MacPeek AH, Hartl DL (1988) Excision of the Drosophila transposable element *mariner*: identification and characterization of the Mos factor. EMBO J. 7: 2185–2189.

Moriyama EN, Powell JR (1997) Synonymous substituition rates in *Drosophila*: Mitochondrial versus Nuclear Genes. J. Mol. Evol. 45: 378-391.

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

Muños-Lopes M, Siddique A, Bischerour J, Lorite P, Chalmers R, Palomeque T (2008) Transposition of Mboumar-9: Identification of a new naturally active mariner-family transposon. J. Mol. Biol. 382: 567-572.

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

Neron B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C (2009) Mobyle: a new full web bioinformatics framework. Bioinformatics btp493+.

Nicholas KB, Nicholas HB (1997) GeneDoc: a tool for editing and annotating multiple sequences alignments. Distributed by the author.

Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Oliveira LVF, Wallau GL, Loreto ELS (2009) Isolation of high quality DNA: a protocol combining “rennet” and glass milk. Electronic J. Biotechnol. 12(2): 1-6.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14(9): 817-818.

Ren X, Park Y, Miller T (2006) Intact mariner-like element in tobacco budworm,Heliothis virescens (Lepidoptera: Noctuidae). Insect Mol. Biol. 15: 743-748.

Robe LJ, Loreto ELS, Valente VLS (2010) Radiation of the Drosophila subgenus (Drosophilidae, Diptera) in the neotropics. J. Zool. Syst. Evol. Res.

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

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

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

Robertson HM, Lampe DJ (1995) Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol. 12: 850-862.

Robertson HM, MacLeod EG (1993) Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. 2: 125-139.

Robertson HM, Soto-Adames FN, Walden KKO, Avancini RMP, Lampe DJ (2002) The *mariner* transposons of animals: horizontally jumping genes. Horizontal Gene Transfer, second edition, Academic Press, San Diego. pp. 173-187.

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

Rouault J, Casse N, Chénais B, Hua-Van A, Capy P (2009) Automatic classification within families of transposable elements: Application to the *mariner* Family. Gene. 448: 227-232.

Silva JC, Bastida F, Bidwell SL, Johnson PJ, Carlton JM (2005) A potentially functional *mariner* transposable element in the protist *Trichomonas vaginalis*. Mol. Biol. Evol. 22: 126-134.

Sinzelle L, Chesneau A, Bigot Y, Mazabraud A, Pollet N (2006) The *mariner* transposons belonging to the *irritans* subfamily were maintained in chordate genomes by vertical transmission. J. Mol. Evol. 62: 53-65.

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

Swofford DL (2003) PAUP: phylogenetic analysis using parsimony (and others methods). Version 4. Sinauer Associates. Massachusetts.

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.

Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T, Kimura K (2001) Possible horizontal transfer of a transposable element from host to parasitoid. Mol. Biol. Evol. 18: 1952-1958.

Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28**:**913–922.

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

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

7. Supplementary material

Suppl. Table 1. *Drosophila* taxa used in this study, with their source.

Genus Subgenus	Section	Group	Species	Origin
Drosophila *Drosophila*	*quinaria-tripunctata*	*guarani*	*D. ornatifrons*	Bento Gonçalves/RS, Brazil
			D. guaru	Joinville/SC, Brazil
			D. subbadia	El Narango, México
		guaramuru	*D. maculifrons*	Morro Santana/RS, Brazil
			D. griseolineata	Morro Santana/RS, Brazil
		tripunctata	*D. nappae*	Itapuã/RS, Brazil
			D. paraguayensis	Porto Alegre/RS, Brazil
			D. mediopunctata	Morro Santana/RS, Brazil
			D. crocina	Marko, Brazil
			D. mediostriata	Porto Alegre/RS, Brazil
			D. paramediostriata	Porto Alegre/RS, Brazil
			D. mediopictoides	Boquete, Panamá
			D. bandeirantorum	Morro Santana/RS, Brazil
			D. tripunctata	Iowa River, Iowa, EUA
			D. mediodiffusa	Maricão, Porto Rico
			D. mediopicta	Itapuã/RS, Brazil
		cardini	*D. cardini*	Morro Santana/RS, Brazil
			D. neocardini	Florionópolis/SC, Brazil
			D. cardinoides	Porto Alegre/RS, Brazil
			D. parthenogenética	Florianópolis/SC, Brazil
			D. procardinoides	Coroico, Bolivia
			D. arawakana	St Kitts, Caribbe
			D. dunni	St Thomas, Caribbe
			D. polymorpha	Uruguay
		pallidpennis	*D. pallidpennis*	Joinville, SC, Brazil
	virilis-repleta	*mesophragmatica*	*D. gasici*	
			D. brncici	Bogotá, Colombia
			D. gaucha	Campos do Jordão/SP, Brazil
			D. pavani	La Florida, Chile
		repleta	*D. hydei*	Morro Santana/RS, Brazil
			D. buzzati	Carboneras, Spain
			D. onca	
			D. repleta	
			D. mercatorum	Florianópolis/SC, Brazil
		immigrans	*D. immigrans*	Uruguay
		flavopilosa	*D. incompta*	Santa Maria/RS, Brazil
Dorsilopha			*D. busckii*	Santa Maria/RS, Brazil
Sophophora		*willistoni*	*D. nebulosa*	Ilha do Arvoredo/SC, Brazil
			D. willistoni	Belém/PA, Brazil
			D. fumipennis	Florianópolis/SC, Brazil
			D. paulistorum	Pará, Brazil
			D. tropicalis	Palma/GO, Brazil
			D. equinoxialis	México
			D. sucinea	México
			D. insularis	St Kitts, Caribbe
			D. capricorni	Florianópolis/SC, Brazil
		saltans	*D. sturtevant*	Maquiné/RS, Brazil
Scaptodrosophila		*rulifrons*	*S. galloi*	S. J. Rio Preto/SP, Brazil
			S. latifasciformes	S. J. Rio Preto/SP, Brazil

Suppl. Table 2: Consensus names, clone names and GeneBank accession number of MELs.

mellífera subfamily		
Species	Consensus Name	Clones Name
D. ornatifrons	DTTMarMel_Dornat1cons	DTTMarMel_Dornat6 (GU229940)
		DTTMarMel_Dornat17 (GU229941)
D. maculifrons	DTTMarMel_Dmacul1cons	DTTMarMel_Dmacul2 (GU229977)
		DTTMarMel_Dmacul6 (GU229979)
		DTTMarMel_Dmacul11 (GU229980)
		DTTMarMel_Dmacul16 (GU229982)
		DTTMarMel_Dmacul20 (GU229985)
	DTTMarMel_Dmacul2cons	DTTMarMel_Dmacul14 (GU229981)
		DTTMarMel_Dmacul17 (GU229983)
		DTTMarMel_Dmacul19 (GU229984)
D. paraguayensis		DTTMarMel_Dparag19 (GU229942)
		DTTMarMel_Dparag30 (GU229943)
D. mediostriata	DTTMarMel_Dmedtr1cons	DTTMarMel_Dmedtr2 (GU229948)
		DTTMarMel_Dmedtr5 (GU229949)
		DTTMarMel_Dmedtr12 (GU229950)
		DTTMarMel_Dmedtr16 (GU229951)
D. paramediostriata	DTTMarMel_Dparam1cons	DTTMarMel_Dparam7 (GU229952)
		DTTMarMel_Dparam25 (GU229955)
		DTTMarMel_Dparam29 (GU229956)
D. tripunctata	DTTMarMel_Dtripu1cons	DTTMarMel_Dtripu2 (GU225844)
		DTTMarMel_Dtripu3 (GU225845)
		DTTMarMel_Dtripu4 (GU225846)
		DTTMarMel_Dtripu9 (GU225849)
		DTTMarMel_Dtripu10 (GU225850)
		DTTMarMel_Dtripu20 (GU225856)
D. cardini		DTTMarMel_Dcardi9 (GU229938)
D. neocardini		DTTMarMel_Dneoca1 (GU229944)
	DTTMarMel_Dneoca1cons	DTTMarMel_Dneoca6 (GU229945)
		DTTMarMel_Dneoca15 (GU229946)
D. arawakana	DTTMarMel_Darawa1cons	DTTMarMel_Darawa1 (GU225842)
		DTTMarMel_Darawa13 (GU225843)
D. pallidpennis		DTTMarMel_Dpallid7 (GU229939)
S. latifasciformis		DTTMarMel_Slatif16 (GU229960)
	DTTMarMel_Slatif1cons	DTTMarMel_Slatif5 (GU229957)
		DTTMarMel_Slatif15 (GU229959)
	DTTMarMel_Slatif2cons	DTTMarMel_Slatif10 (GU229958)
		DTTMarMel_Slatif18 (GU229961)

mauritiana subfamily		
Species	Consensus	Clones
D. maculifrons		DTTMarMau_Dmacul5 (GU229978)
D. mediopunctata	DTTMarMau_Dmedpu1cons	DTTMarMau_Dmedpu2 (GU229969)
		DTTMarMau_Dmedpu6 (GU229970)
		DTTMarMau_Dmedpu10 (GU229971)
		DTTMarMau_Dmedpu11 (GU229972)
		DTTMarMau_Dmedpu20 (GU229973)
		DTTMarMau_Dmedpu23 (GU229974)
		DTTMarMau_Dmedpu26 (GU229975)
D. paramediostriata	DTTMarMau_Dparam2cons	DTTMarMau_Dparam11 (GU229953)
		DTTMarMau_Dparam15 (GU229954)

D. neocardini		DTTMarMau_Dneoca19 (GU229947)
D. busckii	DTTMarMau_Dbusck1cons	DTTMarMau_Dbusck3 (GU229962)
		DTTMarMau_Dbusck9 (GU229963)
D. paulistorum		DTTMarMau_Dpauli6 (GU229937)
	irritans subfamily	
Species	Consensus	Clones
D. mediopunctata		DTTMarIrr_Dmedpu27 (GU229976)
D. tripunctata	DTTMarIrr_Dtripu2cons	DTTMarIrr_Dtripu5 (GU225847)
		DTTMarIrr_Dtripu7 (GU225848)
		DTTMarIrr_Dtripu11 (GU225851)
		DTTMarIrr_Dtripu14 (GU225852)
		DTTMarIrr_Dtripu17 (GU225853)
		DTTMarIrr_Dtripu18 (GU225854)
		DTTMarIrr_Dtripu19 (GU225855)
D.busckii	DTTMarIrr_Dbuski2cons	DTTMarIrr_Dbusck16 (GU229964)
		DTTMarIrr_Dbusck19 (GU229965)
		DTTMarIrr_Dbusck20 (GU229966)
		DTTMarIrr_Dbusck21 (GU229967)
		DTTMarIrr_Dbusck23 (GU229968)

Suppl. Table 3. Maximun, minimum and mean of average number of substitution per site of nucleotide and amino acid calculed for each subfamily.

		Máximum	Minimum	Mean
mellifera	Nucleotide	0,926	0,025	0,353
	Amino acid	1,959	0,091	0,530
mauritiana	Nucleotide	1,199	0,016	0,665
	Amino acid	1,495	0,018	0,782
irritans	Nucleotide	0,594	0,044	0,300
	Amino acid	2,036	0,280	0,879

Suppl. Table 4: Mean codon bias index (CBI), GC content (G+Cc) at coding positions and number of effective codons (Nc) for five nuclear genes and transposon sequences. CBI varies between 0 (uniform use of synonymous codons) and 1 (maximum bias) and Nc varies between 21 (for maximum bias) and 61 (for no bias).

	CBI	G+Cc	Nc
Amd	0,472 (0,059)	0,559 (0,040)	46,28 (3,761)
COII	0,698 (0,056)	0,277 (0,020)	35,33 (3,341)
Ddc	0,504 (0,078)	0,562 (0,035)	43,08 (4,059)
mellifera	0,375 (0,042)	0,507 (0,034)	54,43 (4,119)
mauritiana	0,378 (0,015)	0,480 (0,021)	53,26 (3,562)
irritans	0,347 (0,053)	0,438 (0,025)	56,70 (3,738)

CAPÍTULO III

Evolutionary history of *Paris*-like elements in 12 *Drosophila* genomes

Artigo submetido à revista Genome

1. Abstract

Searches in 12 *Drosophila* genomes for related sequences of the transposable element *Paris* showed the existence of five new *Paris*-like elements (one in *D. ananassae*, one in *D. pseudoobscura,* one in *D. persimilis,* one in *D. mojavensis* and one in *D. willistoni*) whose copy numbers vary between two and seven. Three species (*D. ananassae, D. peseudoobscura* and *D.* mojavensis) present putatively active elements. They have terminal inverted repeats with the direct repeats conserved, target site duplications and intact ORFs. Analysis of presumed protein of *Paris* element of *D. virilis* and the *Paris*-like putatively active elements revealed some motifs normally found in *Tc*1-*mariner* transposases: a helix-turn-helix domain, a bipartite nuclear localization motive and the amino acid aspartic triad (DD34E). The *Paris*-like elements showed, in the 5' and 3' boundaries, an approximately 242-bp-long terminal inverted repeat (called LIR: long inverted repeat), with two 28-bp-long direct repeats in each LIR. Some of these elements have degeneration in the internal region of terminal inverted repeat. Despite the LIR's degeneration in some elements, the distance of 185 bp between the direct repeats was always kept. This suggests that the spacing between direct repeats is important to transposase binding. The evolutionary analysis of elements suggests that they are maintained by vertical transmission with some events of stochastic loss in the analyzed *Drosophila* species.

Key words: paris-like, transposable element, transposase motifs, *Drosophila,* inverted terminal repeat

2. Introduction

Transposable elements (TEs) correspond to significant part of the numerous genomes and have played a role in shaping these over millions of years of evolution (Arkhipova 2001; Biémont et al. 2006). The movement of these elements can promote large-scale structural changes, epigenetics modifications and contribute to the production of allelic diversity, new genes and biological innovations (Feschotte and Pritham 2007). In 12 *Drosophila* genomes sequenced transposable elements varies by over one order of magnitude, ranging from 2.7% in *D. simulans* and *D. grimshawi* to 25% in *D. ananassae* (Clark et al. 2007). However, a more detailed analysis of the structural and evolutionary relationships among these transposable elements and its related elements remains an open question for *Drosophila* genomics.

The *Paris* transposable element was originally described in a system of hybrid dysgenesis in *Drosophila virilis* associated with the mobilization of at three unrelated transposable elements designated *Helena*, *Penelope* and *Ulysses* (Petrov et al. 1995; Vieira et al. 1998). *Paris* is approximately 1,730 bp long and has an atypical long terminal inverted repeat (TIRs) of 242 bp (Petrov et al. 1995). This transposon belongs to class II elements, which transpose via a DNA intermediate, order TIR and superfamily *Tc1-mariner*, following the TE classification proposed by Wicker et al. 2007. The characteristics of this superfamily are: (i) the generation of TA target site duplication (TSDs), (ii) presence of TIRs and (iii) presence of only one ORF.

The member of the *Tc1-mariner* superfamily can be classified based on the organization of the catalytic residues or on TIRs size and complexity. Based on catalytic

residues criteria, this superfamily is divided in three lineages. The first is composed by two families (*ITmD37E* and *ITmD37D*), the second lineage corresponds to the *Tc1*-like elements (TELs, known as *ITmD34E*), and the third is composed by three families: *Gambol* (known as *ItmD34E-bis*), plant-*D39D* and *mariner*-like elements (MELs, known as *ItmD34D*) (Shau and Tu 2001). Based on the sequence properties of TIRs this superfamily is divided in four groups. The first group comprises TIRs with size ranging from 24 bp to nearly 100 bp, the second comprises TIRs that are larger than 300 bp and contains two similarly oriented transposase binding sites called direct repeats (DRs). The third group, known as IR-DR, contains TIRs of 200-250 bp in length and two DRs, while the fourth group has TIRs of 20-40 bp (Brillet et al. 2007).

The transposable element *Paris* belongs to the lineage of *Tc1*-like elements and the group IR-DR, which also includes other *Drosophila* elements like *S*, *Bari1, minos* and *SB.* These elements encode transposases that contain, in the N-terminal region, a functional bipartite nuclear localization signal (NLS_BP) and two HTH motifs (Brillet et al. 2007). In C-terminal region they have an acidic DD34E triad (Brillet et al. 2007). However, so far all motifs that are present in *Paris* transposase have not been fully characterized. The TIRs of this element have two direct repeats (DRs) called outer DR (ODR) and inner DR (IDR) located at opposite ends of the TIR (Brillet et al. 2007). These repeats are the binding sites of transposase necessary for the mechanism of transposition of autonomous elements (Cui et al. 2002; Fischer et al. 1999; Izsvak et al. 2002). In a recent study Moschetti et al. 2008 analyzing the TIRs of three elements (*Paris*, *S*, and *minos*), found a third conserved DR located between ODR and IDR, with only 9 bp.

With the availability of 12 *Drosophila* genomes some works have focused on the search and characterization of new transposable elements (Ludwig and Loreto 2008; Ortiz

and Loreto 2009; Yang and Barbash 2008). Here, we describe the search for *Paris* homologous sequences in the 12 *Drosophila* genomes and performed a structural characterization and evolutionary analysis to five new *Paris*-like elements found in six different species.

3. Material and Methods

Genome search

Searches for *Paris* homologous elements were carried out in the genomes of the following *Drosophila* species: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis* and *D. grimshawi*. The searches were performed using Blat (Kent 2002) with default parameters. However, the criteria for sequence retrieving were: homology score >100, for the translated DNA query, and score >1000, for the DNA query. Searches were also performed in the FlyBase BLAST Service (http://flybase.bio.indiana.edu/blast/) (Altschul et al. 1990) using the default parameters. The sequence used as initial query was the *Paris* element described in *D. virilis* (Petrov et al. 1995) named in this paper as *DviriPA*. Further rounds of search were done using all sequences retrieved in the previous search as probe, with the same parameters described above, until no other new sequences were obtained.

Sequences analysis

The sequence alignments were carried out using the ClustalX2 software (Larkin et al. 2007), with default parameter values. MEGA 4.0.2 (Tamura et al. 2007) was used for sequence editing and visualization. The Einverted software was used for TIRs identification and the Cons program was used to obtain the consensus sequences that showed less than 9 percent of divergence from the EMBOSS suite available in Mobyle Portal (Neron et al. 2009). WebLogo (http://www.bio.cam.ac.uk/seqlogo/) was used for TSD and protein motifs analysis. Gene structure was inferred from the results of ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The similarity with *Paris* elements was established using CENSOR (Kohany et al. 2006).

The phylogenetics analysis was performed using amino acids sequences of *Paris*-like elements aligned with the previously described *Paris* elements of *D. buzzatii* (AY756169) (Casals et al. 2005) and *D. virilis* (U26938) (Petrov et al. 1995). Also, two elements of the *Tc1-mariner* superfamily, the *S* element (U33463) found in *D. melanogaster* and *Tc*1 (X01005) in *Ceanorabiditis elegans* were used as outgroups. The JTT+G evolutionary model was suggested by ProtTest 2.2 (Abascal et al. 2005) according to the Akaike information criterion (Akaike 1974). Maximum Parsimony (MP) and Neighbor Joining (NJ) analyses were performed using MEGA 4.0.2 (Tamura et al. 2007) with 5,000 bootstrap replicates. The MP was performed with CNI (level = 1) with initial tree by Random addition (10 reps) and NJ was performed with the JTT model. Bayesian analysis was performed using the MrBayes 3.1.2 software (Ronquist and Huelsenbeck 2003), with evaluation of at least 1,000,000 generations and a burning region of 2,500 trees, using the model suggested by ProtTest 2.2 (Abascal et al. 2005).

Codon alignment was used to estimate the *dS* values (divergence in the synonymous sites) and the number of synonymous sites (S) using the Nei and Gojobori 1986 method assisted by the MEGA 4.0.2 software (Tamura et al. 2007). Five genes were used in the *dS* analysis: *alpha methyldopa hypersensitive protein* (Amd), *amyrel* (Amy), *cell division cycle* (Cdc), *dopa decarboxylase* (Ddc) and *superoxide dismutase* (Sod). In order to allow the calculation of some sequences, gaps were introduced to conserve the reading frame while stop codons were considered as absent information.

The effective number of codons (Nc) (Wright 1990), the codon bias index (CBI) (Morton 1993) and GC content were calculated by DnaSP 4.90 software (Rozas et al. 2003).

Analysis of protein motifs

The putative domains in the transposase proteins were determined using two different softwares: Helix-turn-helix DNA-binding motifs prediction (Dodd and Egan 1990) and Motif Scan tool (http://hits.isb-sib.ch/cgi-bin/PFSCAN) (Hulo et al. 2008). The physical-chemical analysis of amino acids in the primary structure of enzyme was performed using GeneDoc 2.7.000 (Nicholas and Nicholas 1997).

4. Results and Discussion

Characterization of *Paris*-like elements

DNA analysis

The BLASTn searches in the *Drosophila* genomes have found *Paris*-like sequences in six of 12 genomes analyzed. However, the sequence *DviriPA* described by Petrov et al. 1995 was not found in the *D. virilis* genome. In total, five new elements similar to *Paris* were identified, one in *D. ananassae,* one in *D. pseudoobscura*, one in *D. persimilis*, one in *D. mojavensis* and one in *D. willistoni*. The copy number of each element varied between two and seven (two in *D. ananassae*, five in *D. pseudoobscura,* seven in *D. persimilis,* four in *D. mojavensis* and three in *D. willistoni*). The genome of *D. virilis* presented eight copies of the *Paris* element, though these copies are putatively non-active (supplementary material Table 1). The names of new sequences were attributed following Robertson and Asplund 1996 with the inclusion of tree other letters to avoid misunderstanding when species have similar names. The characteristics of these sequences, the contig names and genomic coordinates obtained in the Blat and Blast analyses are summarized in supplementary material Table 1.

The *Paris*-like elements presented lengths varying from 1,722 to 2,235 bp, while lengths of TIRs ranged from 237 to 246 bp. The TIRs were found in 21 of 29 analyzed sequences. The TA dinucleotide TSDs, which is a characteristic of Tc1-mariner superfamily, was present in eight of 29 sequences. The sequences *Dananpar*1*, Dmojapar*1*, Dmojapar*2 and *Dpseupar*1 have a complete ORF, while the other 27 showed nonsense mutations, deletions or insertions, suggesting that these sequences are not able to produce an active transposase. The length of predicted proteins varied from 322 to 350 amino acids.

Those elements that show TIRs, DRs, TSDs and complete ORFs were classified as putatively active (PA) in the present work. We have named as putatively mobilizable (PM) those elements that have conserved TIRs, DRs and TSDs but do not have complete ORFs. The elements that show nonsense mutations and indels, primarily in the sequences presumably used for mobilization (TIRs and DRs), were denominated putatively non-active (PNA) elements.

In the class II elements, some families show variation in the TIRs that can be useful to classify these TEs in lower levels, as subfamilies. Moschetti et al. 2008 described that the *Bari*-like family has some elements with long TIRs called LIRs (long inverted repeats) with about 250 bp, as well as other elements that contain short SIRs (short inverted repeats), with about 26 bp. Also, these authors argue that the elements that have SIRs have derived from LIRs elements due to a decrease in TIR size. Those authors described that the *Paris*, *S* and *minos* elements have LIRs with three DRs, two located in the TIRs boundaries with 18 bp and one internal DR with 9 bp.

We have performed an analysis of TIRs structure of *Paris*-like sequences. These elements showed a similar structural organization of TIRs described for *Bari*-like TEs. The elements have approximately 242-bp-long LIRs and two DRs. However, some elements showed an internal region of 182 bp in length and not conserved between the DRs (Figure 1). Nevertheless, they maintain intact DRs at the same position (expressed in bp) as that observed in those elements that have “perfect” LIRs. Table 1 lists the elements that present LIRs with non-conserved sequences between DRs as LIRs-ID (LIRs -Interned Degenerated). Table 1 also shows the sequences that showed “perfect” LIRs and the sequences that showed degenerated LIRs or in those in which LIRs were absent.

Table 1: Characteristics of *Paris-like* elements.

Consensus name[a]	Number of PA copies[b]	Number of PM copies[c]	Number of PNA copies[d]	TIRs structure[e]
Danancons	1	0	1	LIRs
Dpseucons	1	0	4	LIRsID
Dperscons	0	1	6	LIRsID
Dwillcons	0	0	3	DIR
Dmojacons	1	0	3	LIRs
Dviricons	0	3	5	LIRs

[a] Consensus name: nomenclature of *Paris*-like elements consensus.
[b] PA (putatively active) are element copies that have TIRs, DRs, TSDs and complete ORFs.
[c] PM (putatively mobilizable) are element copies that have conserved TIRs, DRs and TSDs.
[d] PNA (putatively not active are element copies with nonsense mutations and indels, primarily in the sequences presumably used for mobilization (TIRs and DRs)).

[e] TIRs structure of consensus: LIRs = Long inverted repeats; LIRsID = Long inverted repeats internally degenerated; DIR = Degenerated or absent inverted repeat.

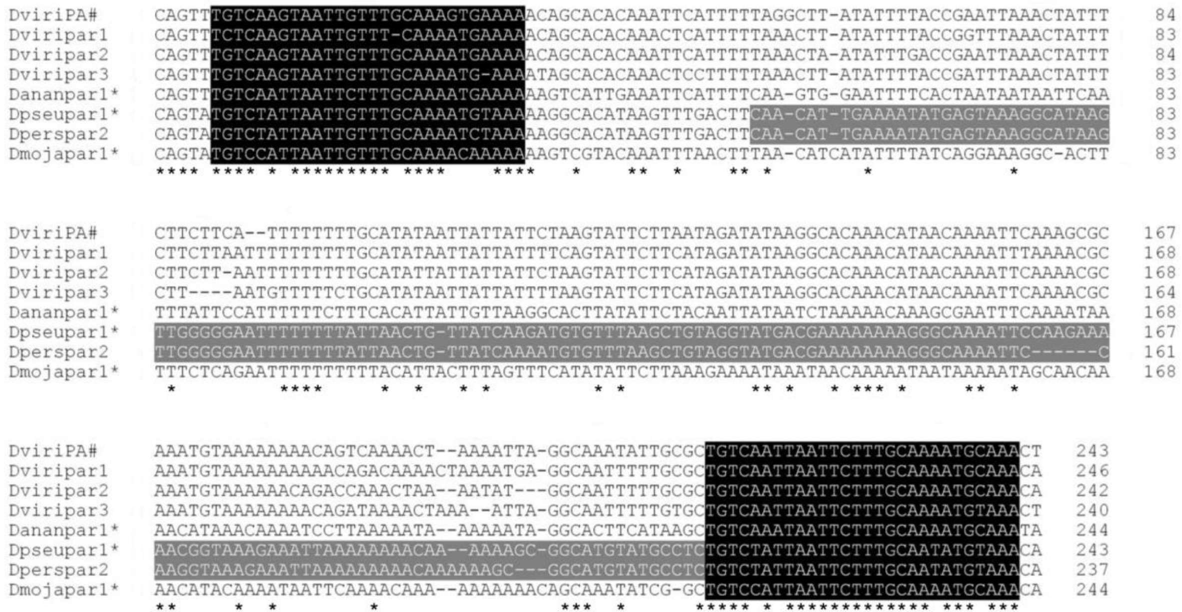


Figure 1: Alignment of TIRs from different copies of *Paris*-like elements. The areas in black represent the DRs sequences and those highlighted in gray represent regions whose TIRs are slightly similar. Sequences marked with (*) conserve the correct reading frame and did not have premature stop codons (are putatively active sequences). The sequence marked with (#) is the first active *Paris* element, described by Petrov et al. 1995.

The *Paris*-like elements present PA sequences with LIRs and LIRsID. This suggests that some elements suffered degeneration in the internal region of TIRs which are not important for transposase recognition. Analysis of the putatively active elements *DviriPA, Dananpar1*, *Dmojapar*1*, Dpseupar*1 in separate or adding four PM elements (*Dviripar*1*, Dviripar*2*, Dviripar*3 and *Dperspar*2) have shown the presence of only two 28-bp-long DRs and 28 DRs in TIRs boundaries. These DRs are separated by an approximately 185-bp-long variable sequence (Figure 1). This suggests that the 9-bp-long internal direct repeat found by Moschetti et al. 2008 has little or no importance in the transposition process of *Paris*-like elements. Remarkable is the fact that the distance of approximately 185 bp between DRs is maintained in both LIRs and LIRsID elements, suggesting that this spacing is important to transposase binding.

Some copies (*Dperspar2*, *Dviripar1*, *Dviripar2* and *Dviripar3*) that present defective ORFs maintain TIRs with conserved DRs, suggesting that they could be mobilized by transposases supplied by an autonomous element.

Analysis of TSDs revealed the TA nucleotide duplication characteristic of superfamily *Tc1-mariner* and the preference these transposons exhibit with regard to being inserted into a hotspot TA, where the fifth nucleotide would very likely be adenine (Figure 2C). This TSD pattern (AYANATA/TATATRN) has been described in the *S* element, another member of the *Tc1-mariner* superfamily, by Merriman et al. 1995.

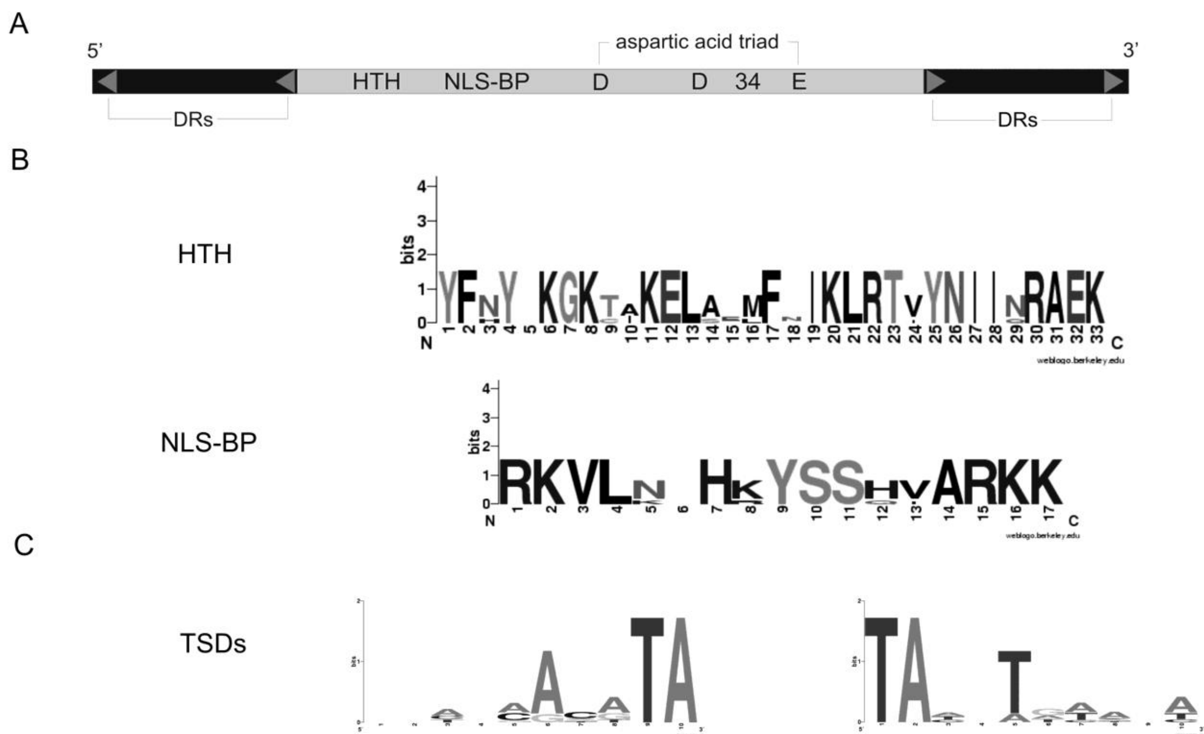


Figure 2: Structural characterization of *Paris*-like elements. A – 28-bp direct repeats are signaled by blue triangles and location of motifs present in *DvirPA* and PA sequences, B – LOGO scheme for HTH (helix-turn-helix) and NLS_BP (bipartite nuclear localization) motifs, C – LOGO scheme for target sequences duplication (TSD) showing the conserved TA TSD.

Proteins Analysis

The motifs present in putative protein of *DviriPA* were analyzed using the Motif Scan software. A helix-turn-helix (HTH) DNA-binding domain present in the N-terminal region was found comprehending amino acids 15 to 47 and a bipartite nuclear localization signal profile (NLS_BP) was observed between the amino acids 104 to 120. Flanking the motif NLS_BP two putative CK-II phosphorylation sites were found between amino acids 64-67 (SRRD) and 174-177 (TALE). A Transposase_5 domain was detected between amino acids 68 to 140 in the C-terminal region. The characteristic aspartic acid DD34E triad was located at positions 153, 251 and 286, respectively (Fig 2).

In a study that analyzed multiple protein sequences of *Tc1-mariner* elements, Avancini et al. 1996 found four conserved motifs ((WF)SDESKF, GGSVMVW, FQQDN(DA) and PDLNPIE). We found similar motifs in the hypothetical *DviriPA* transposase and in the putatively active *Paris*-like elements. However, some amino acids are not the same as those reported by Avancini et al. 1996. In *Paris*-like elements the motifs found were: FCDETKI, KLSVMVW, LYQDND and PDLNPIE. Nevertheless, the different amino acids we observed in each motif and those described by Avancini et al. 1996 have similar physical-chemical properties, suggesting functional importance for transposase activity.

Phylogenetic analyses

Phylogenetic analyses using the consensus of the *Paris*-like sequences obtained revealed the existence of two clades (Figure 3). Clade A is composed by *D. mojavensis*, *D. buzzatii, D. pseudoobscura* and *D. persimilis* sequences and showed a nucleotide

divergence that ranged from 1.2 to 21.4%. This clade is subdivided in two clusters, one with the sequences from the *obscura* group (*D. pseudoobscura* and *D. persimilis)* and the other with *repleta* group sequences (*D. mojavensis* and *D. buzzati*). Clade B is formed by *D. ananassae, D. virilis* and *D. willistoni* sequences, with a nucleotide divergence between 6.4 and 27.5%. The overall divergence between Clade A and Clade B varied from 1.2 to 33.4% (supplementary material Table 2). Both clades are formed by species from the *Drosophila* and *Sophophora* subgenera that show incongruence with the species phylogeny.

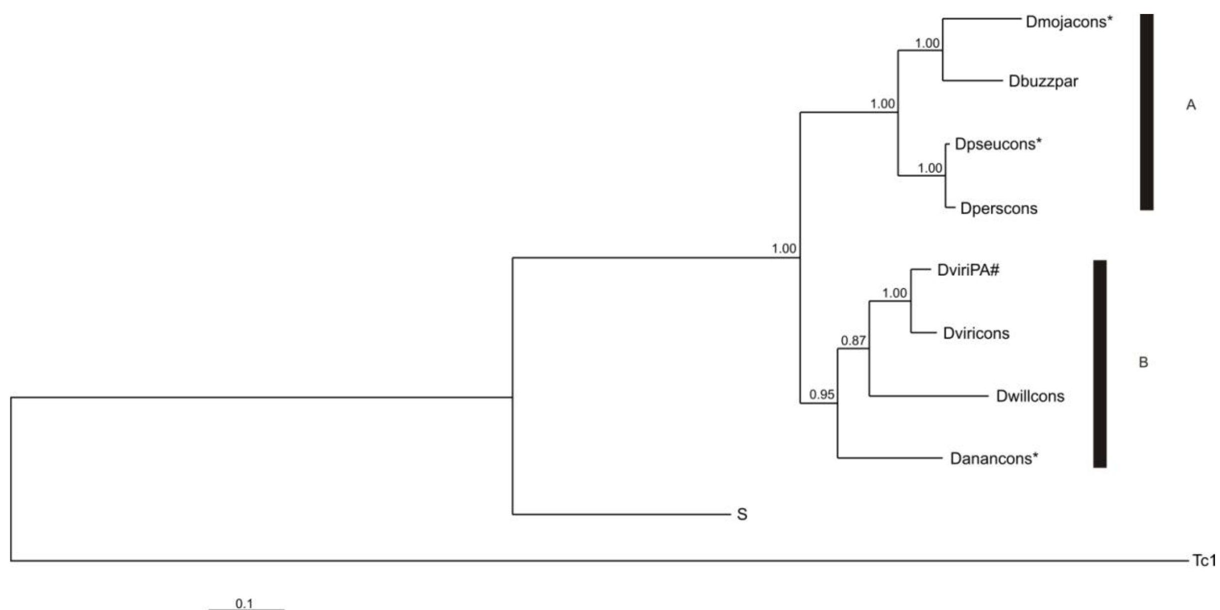


Figure 3: Phylogeny of *Paris*-like elements using the Bayesian analysis with the JTT+G model for 10 sequences. Posterior probabilities of clades are shown next to the branches. Asterisks (*) are putatively active elements (PA) and the hash (#) indicates the first active *Paris* element, described by Petrov et al. 1995.

In order to gain more insights as to the probability that the incongruence observed in the phylogenetic analyses is due to horizontal transmission (HT) of *Paris*-like elements among the species, the divergence at synonymous sites (*dS*) of these sequences was compared with levels of the five nuclear genes of *Drosophila* species (Amd, Cdc, Amy, Ddc,

Sod) (supplementary material Table 3). The comparison of *dS* values was chosen to infer HT because *dS* values offer a measurement of neutral evolution in the absence of a strong codon usage bias. The divergences were compared using the Fisher exact test to verify whether the *dS* of the *Paris*-like elements are significantly lower than the *dS* of nuclear genes (Ludwig and Loreto 2008). The comparisons showed that the *dS* observed among genes and TEs was not significant, suggesting that these elements are vertically transferred. The incongruence observed in the phylogenetic analyses can be explained by ancestral polymorphism and/or different evolutionary rates of *Paris*-like elements in the ancestral of the subgenus *Sophophora* and *Drosophila* and with some events of stochastic loss.

5. Acknowledgments

This study was supported by research grants and fellowships from CNPq and CAPES.

6. References:

Abascal, F., Zardoya, R. and Posada, D. 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. **21**: 2104-2105. doi:10.1093/bioinformatics/bti263.

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

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang Z, Miller W and Lipman D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.

Arkhipova, I.R. 2001. Transposable elements in the animal kingdom. Mol. Biol. **35**: 157-167.

Avancini, R.M.P., Walden, K.K.O. and Robertson, H.M. 1996. The genomes of most animals have multiple members of the Tc1 family of transposable elements. Genetica. **98**: 131-140.

Biémont, C. and Vieira, C. 2006. Junk DNA as an evolutionary force. Nature. **443**: 521-524. doi:10.1038/443521a.

Brillet, B., Bigot, Y. and Augé-Gouillou, C. 2007. Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica. **130**: 105-120. doi:10.1007/s10709-006-0025-2.

Casals, F., Cáceres, M., Manfrin, M.H., Gonzáles, J. and Ruiz A. 2005. Molecular characterization and chromosomal distribution of Galileo, Kepler and Newton, three foldback transposable elements of the *Drosophila buzzati* species complex. Genetics. **169**(4): 2047-2059. doi:10.1534/genetics.104.035048.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M. and Gelbart, W. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature. **450**: 203-218. doi:10.1038/nature06341.

Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D. and Hackett, P.B. 2002. Structure-function analysis of the inverted terminal repeats of the *Sleeping Beauty* transposon. J. Mol. Biol. **318**: 1221-1235. doi:10.1016/S0022-2836(02)00237-1.

Dodd, I.B. and Egan, J.B. 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18**: 5019-5026.

Feschotte, C. and Pritham, E. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. **41**: 331-368. doi:10.1146/annurev.genet.40.110405.090448.

Fischer, S.E., van Luenen, H.G. and Plasterk, R.H. 1999. Cis requirement for transposition fo *Tc1*-like transposon in *C. elegans*. Mol. Gen. Genet. **262**: 268-274.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A. de Castro, E., Lachaize, C., Langendijk-Genevaux, P.S. and Sigrist, C.J. 2008. The 20 years of PROSITE. Nucleic Acids Res. **36**(Database issue): D245-9. doi:10.1093/nar/gkm977.

Izsvak, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R.H. and Ivics, Z. 2002. Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition. J. Biol. Chem. **277**: 34581-34588. doi: 10.1074/jbc.M204001200.

Kent, W.J. 2002. BLAT - The BLAST-Like Alignment Tool. *Genome Res.* **12**(4): 656-664. doi:10.1101/gr.229202.

Kohany, O., Gentles, A.J., Hankus, L. and Jurka, J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. **7**: 474. doi:10.1186/1471-2105-7-474.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. **23**: 2947-2948. doi:10.1093/**bioinformatics**/btm404.

Ludwig, A. and Loreto, E.L.S. 2008. Multiple invasions of *Errantivirus* in the genus *Drosophila.* Insect Mol. Biol. **17**(2): 113-124. doi:10.1111/j.1365-2583.2007.00787.x.

Merriman, P.J., Grimes, C.D., Ambroziak, J., Hackett, D.A., Skinner, P. and Simmons, M.J. 1995. *S* elements: a family of *Tc1*-like transposons in the genome of *Drosophila melanogaster*. Genetics. **141**: 1425-1438.

Morton, B.R. 1993. Chloroplast DNA codons use: evidence for selection at the psb A locus based on tRNA availability. J. Mol. Evol. **37**: 273-280.

Moschetti, R., Chlamydas, S., Marsano, R.M. and Caizzi, R. 2008. Conserved motifs and dynamic aspects of the terminal inverted repeat organization within *Bari*-like transposons. Mol. Genet. Genomics. **279**: 451-461. doi:10.1007/s00438-008-0324-7.

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

Neron, B., Ménager, H., Maufrais, C., Joly, N., Maupetit, J., Letort, S., Carrene, S., Tuffery, P. and Letondal, C. 2009. Mobyle: a new full web bioinformatics framework. *Bioinformatics.* **25**(22): 3005-3011. doi:10.1093/**bioinformatics**/btp493.

Nicholas, K.B. and Nicholas, H.B. 1997. GeneDoc: a tool for editing and annotating multiple sequences alignments. Distributed by the author.

Ortiz, M.F. and Loreto, E.L.S. 2009. Characterization of new *hAT* transposable elements in 12 *Drosophila* genomes. Genetica. **135**(1): 67-75. doi:10.1007/s10709-008-9259-5.

Petrov, D.A., Schutzman, J.L., Hartl, D.L. and Lozovskaya, E.R. 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proc. Natl. Acad. Sci. **92**: 8050-8054.

Robertson, H.M. and Asplund, M.L. 1996. Bmmar1: a basal lineage of the mariner family of transposable elements in the silkworm moth, *Bombyx mori*. Insect Biochem. Mol. Biol. **26**: 945-954.

Ronquist, F. and Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. **19**: 1572-1574. doi:10.1093/bioinformatics/btg180.

Rozas, J., Sánchez-DelBarrio, J.C., Messenger, X. and Rozas, S. 2003. DnaSP, Dna polymorphism analyses by the coalescent and other methods. Bioinformatics. **19**: 2496-2497. doi:10.1093/bioinformatics/btg359.

Shau, H. and Tu, Z. 2001. Expanding the diversity of the *IS630-Tc1-mariner* superfamily: dicovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics. **159**: 1103-1115.

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. doi:10.1093/molbev/msm092.

Vieira, J., Vieira, C., Hartl, D. and Lozovskaya, E.R. 1998. Factors contributing to the hybrid dysgenesis syndrome in *Drosophila virilis*. Genet. Res. **71**: 109-117.

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. doi:10.1038/nrg2165.

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

Yang, H. and Barbash, D.A. 2008. Abundant and species-specific *DINE-1* transposable elements in 12 *Drosophila* genomes. Genome Biol. **9**(2): R39. doi:10.1186/gb-2008-9-2-r39

7. Supplementary material

Suppl. Table 1. Characterization of the 29 new Paris-like sequences and their Blat and Blast hits genomic coordinates. The Setember 2009 version of genomes assembly was used in BLAST searches.

								BLAT Positon			BLAST Position		
Consensus Name	Name[a]	Size (pb)[b]	Protein (aa)[c]	TIRs structure[d]	TIRs size (pb)	TSD[e]	State[f]	Contig	Star	End	Contig	Star	End
Danancons	*Dananpar1*	1722	348	LIR	238	+	pa	13045	350850	352849	CH902641	350934	350934
	Dananpar2	1147(ps)	251§	-	-	-	pna	12966	149774	150920	CH902666	149839	150985
Dpseucons	*Dpseupar1*	1728	348	LIRID	238	+	pa	U	12392714	12394457	CH674952	6186	7913
	Dpseupar2	1628(ps)	325§	-	-	-	pna	U	5738226	5739862	CH475521	14365	15992
	Dpseupar3	837(ps)	178§	-	-	-	pna	XL_group1a	1576739	1579385	CH674952	7077	7707
	Dpseupar4	1195(ps)	322§	-	-	-	pna	U	9029788	9030978	CH476176	11100	9943
	Dpseupar5	951	107	DIR	59	-	pna	XL_group1a	1198141	1199075	CH379063	1220586	1221516
Dperscons	*Dperspar*1	1243	306	DIR	63	-	pna	super_5722	761	1677	CH693729	452	1669
	*Dperspar*2	1614	306	LIRID	54	+	pm	super_867	673	1925	CH689800	494	2101
	*Dperspar*3	1188	306	DIR	36	-	pna	super_7691	231	1456	CH695692	268	1456
	*Dperspar*4	1366	347	DIR	59	-	pna	super_59	133861	135689	CH479238	134323	135677
	*Dperspar*5	1477	347	DIR	28	-	pna	super_26	757179	758661	CH479205	757203	758658
	*Dperspar*6	1477	347	DIR	28	-	pna	super_26	769175	770657	CH479205	769198	770448
	*Dperspar*7	1260(ps)	342§	-	-	-	pna	super_173	63304	64563	CH479351	63313	64553

Dwillcons	*Dwillpar*1	1489	287	DIR	198	+	pna	-	-	-	CH964281	1108229	1109725
	*Dwillpar*2	1039(ps)	232§	-	-	-	pna	-	-	-	CH964290	480685	481723
	*Dwillpar*3	1057(ps)	246§	-	-	-	pna	-	-	-	CH964145	35227	34171
Dmojacons	*Dmojapar*1	1729	348	LIR	247	+	pa	6482	236498	238226	CH933815	236473	238201
	*Dmojapar*2	1552	348	DIR	158	-	pna	6541	1740226	1741900	CH933817	1737267	1738584
	*Dmojapar*3	1306	332	DIR	52	-	pna	5812	4144	5550	CH934891	5550	4245
	*Dmojapar*4	1247(ps)	343§	-	-	-	pna	2111	2472	3812	CH934530	2556	3802
Dviricons	*Dviripar*1	2239	348	LIR	245	+	pm	12970	273358	275083	CH940651	272818	275076
	*Dviripar*2	1685	335	LIR	246	+	pm	13324	2246758	2248430	CH940659	2242990	2244694
	*Dviripar*3	1632	331	LIR	239	+	pm	12954	1618821	1620448	CH940666	1614671	1616322
	*Dviripar*4	1050 (ps)	300§	-	-	-	pna	13052	1083736	1085970	CH940665	1080346	1081415
	*Dviripar*5	1407	342	DIR	82	-	pna	12958	128142	129698	CH940657	128480	129906
	*Dviripar*6	1387	178	DIR	219	-	pna	13324	1721003	1722410	CH940659	1718680	1720086
	*Dviripar*7	1348	344	DIR	27	-	pna	12855	8762071	8763413	CH940652	8760693	8761854
	*Dviripar*8	1078	166	DIR	75	-	pna	13045	1977785	1979059	CH940663	1968220	1969292

[a] Element name: nomenclature proposed for Robertson and Asplund modified, even a number that corresponds to the order in which sequences were identified.
[b] Nucleotide size of elements: (ps) partial sequences.
[c] Aminoacids size of deduced proteins: (§) probable proteins of partial sequences.
[d] TIRs structure: LIR = Long inverted repeated; LIRID = Long inverted repeated internally degenerated; DIR = Degenerated inverted repeat.
[e] TSD: (-) absence and (+) presence of target site duplication.
[f] State = Putative mobilization state of TE: pa = Putatively active; pm = Putatively mobilizable; pna = Putatively not active.

Suppl. Table 2. Nucleotide divergence percentages found between of the Paris-like elements.

	Dperscons	Dpseucons	Dbuzzapar	Dmojacons	Danancons	DviriPA	Dviricons	Dwillcon
Dperscons								
Dpseucons	1,2							
Dbuzzapar	18,6	18,6						
Dmojacons	21,3	21,4	14,8					
Danancons	29	28,8	32,7	30,8				
DviriPA	28,7	28,8	29	29,6	23,6			
Dviricons	27,9	27,5	29,2	30,6	24,5	6,4		
Dwillcons	31,1	31	32	33,4	27,5	23,8	24	

Suppl. Table 3. dS values between Paris-like consensus elements and five nuclear genes.

Paris-like consensus elements

	Dperscons	Dpseucons	Dbuzzapar	Dmojacons	Danancons	DviriPA	Dviricons	Dwillcons
Dperscons								
Dpseucons	0,039							
Dbuzzapar	0,514	0,531						
Dmojacons	0,606	0,616	0,400					
Danancons	0,768	0,758	0,830	0,758				
DviriPA	0,768	0,788	0,714	0,720	0,679			
Dviricons	0,738	0,745	0,715	0,752	0,699	0,196		
Dwillcons	0,833	0,850	0,828	0,870	0,775	0,717	0,706	

Amd

	D. persimilis	*D. pseudoobscura*	*D. buzzatii*	*D. mojavensis*	*D. ananassae*	*D. virilis*	*D. willistoni*
D. persimilis							
D. pseudoobscura	0,011						
D. buzzatii	0,750	0,753					
D. mojavensis	0,724	0,725	0,344				
D. ananassae	0,669	0,681	0,747	0,717			

D. virilis	0,701	0,697	0,656	0,544	0,707		
D. willistoni	0,831	0,844	0,750	0,757	0,766	0,825	

Ddc

	D. persimilis	*D. pseudoobscura*	*D. buzzatii*	*D. mojavensis*	*D. ananassae*	*D. virilis*	*D. willistoni*
D. persimilis							
D. pseudoobscura	0,022						
D. buzzatii	0,628	0,628					
D. mojavensis	0,631	0,631	0,281				
D. ananassae	0,595	0,599	0,634	0,612			
D. virilis	0,694	0,690	0,560	0,566	0,694		
D. willistoni	0,793	0,793	0,702	0,733	0,701	0,726	

Amy

	D. persimilis	*D. pseudoobscura*	*D. mojavensis*	*D. ananassae*	*D. virilis*	*D. willistoni*
D. persimilis						
D. pseudoobscura	0,016					
D. mojavensis	0,517	0,517				
D. ananassae	0,527	0,527	0.534			
D. virilis	0,580	0,580	0,417	0,649		
D. willistoni	0,722	0,776	0,789	0,860	0,698	

Cdc

	D. persimilis	*D. pseudoobscura*	*D. mojavensis*	*D. ananassae*	*D. virilis*	*D. willistoni*
D. persimilis						
D. pseudoobscura	0,038					
D. mojavensis	0,644	0,646				
D. ananassae	0,698	0,699	0,750			

D. virilis	0,570	0,578	0,523	0,734		
D. willistoni	0,656	0,643	0,656	0,735	0,683	

	Sod					
	D. persimilis	*D. pseudoobscura*	*D. mojavensis*	*D. ananassae*	*D. virilis*	*D. willistoni*
D. persimilis						
D. pseudoobscura	0,036					
D. mojavensis	0,831	0,831				
D. ananassae	0,453	0,453	0,742			
D. virilis	0,627	0,627	0,633	0,590		
D. willistoni	0,751	0,715	0,746	0,633	0,612	

Conclusões gerais e Perspectivas

Os resultados obtidos nesta pesquisa demonstram que os elementos estudados possuem uma distribuição maior do que a conhecida até o momento. Os elementos relacionados à *mariner* são amplamente distribuídos em Drosophilidae e os elementos relacionados à *Paris* são distribuídos em espécies do subgênero *Sophophora* e *Drosophila,* do gênero *Drosophila*. Outra constatação é que foram detectados elementos potencialmente ativos, para elementos relacionados à *Paris* e *mariner*. Além disso, foram caracterizados eventos de transmissão horizontal, o que pode influenciar nos processos de evolução genômica dos organismos envolvidos.

A família *mariner* se mostrou amplamente distribuída nos drosofilídeos neotropicais, sendo representada por três subfamílias. Cada subfamília apresentou um complexo padrão de evolução, com alguns eventos de TH, entre espécies próximas e distantemente relacionadas. Apesar dessa constatação, esses eventos provavelmente foram subestimados, dado que os *primers* usados não amplificaram ou os fragmentos não foram clonados em algumas espécies. As incongruências existentes entre algumas espécies que possuem sequências e não foram hibridizadas no *Dot Blot*, podem ser decorrentes das diferentes linhagens das espécies utilizadas entre as diferentes metodologias, ou algum problema de hibridização das sondas.

Corroborando os dados da literatura (Robertson e MacLeod 1993), a grande maioria dos MELs encontrados neste trabalho também possuem mutações na região da ORF analisada, indicando que são provenientes de elementos inativos. Três espécies, entretando, apresentaram sequências de duas subfamílias de *mariner* (*mauritiana* e *mellifera*) que possuem uma ORF evoluindo sob uma forte seleção purificadora, sugerindo que essas sequências podem ser provenientes de elementos ativos.

Análises adicionais serão necessárias para esclarecer os eventos de TH e testar a possível atividade dos MELs em drosofilídeos neotropicais.

Os elementos relacionados à *Paris* estão presentes em 6 das 12 espécies de *Drosophila* com genoma sequenciado. Dentre estes, existem elementos potencialmente ativos, potencialmente mobilizáveis e degenerados, que estão evoluindo por transmissão vertical, com alguns eventos de perda estocástica, em algumas espécies.

Os elementos relacionados à *Paris* potencialmente ativos, apresentaram sequências com uma degeneração na região interna das TIRs e elementos que não possuem a degeneração, porém ambos os tipos de elementos apresentaram duas repetições diretas dentro das TIRs separadas por aproximadamente 185 pb. Isso sugere que o espaçamento entre as repetições é mais importante para a ligação da transposase do que a conservação de toda a TIR.

Análises adicionais sobre o processo de transposição dos elementos relacionados à *Paris, in vitro* ou *in vivo,* serão necessárias para testar as inferências geradas com dados *in silico*.

A caracterização funcional de sequências naturalmente ativas é essencial para responder a questões sobre o impacto dos TEs nos genomas. Como constatado existem elementos relacionados à *Paris* e *mariner* que são potencialmente ativos. Em meu projeto de doutorado pretendo caracterizar alguns desses elementos potencialmente ativos dos genomas de *Drosophila* com ensaios de transposição, *in vivo,* e análises populacionais e assim poder contribuir sobre o entendimento do real impacto dos elementos transponíveis no processo de evolução genômica.

REFERÊNCIAS BIBLIOGRÁFICAS

Almeida, L.M.; Carareto, C.M.A. Multiple events of horizontal transfer of the *Minos* transposable element between *Drosophila* species. Mol Phyl. Evol., v. 35, P. 583-594. 2005.

Arkhipova, I.R. Transposable elements in the animal kingdom. Mol. Biol., v. 35, p. 157-167. 2001.

Bartolomé, C.; Bello, X.; Maside, X. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genon. Biol, v. 10, p. R22. 2009.

Biémont, C.; Vieira, C. Junk DNA as an evolutionary force. Nature. v. 443, p. 521-524. 2006.

Bächli, G. (2008) Taxodros: The database on taxonomy of Drosophilidae, v1, 03. Database 2008/01. Disponível em < http://www.taxodros.uzh.ch/>. Acesso em: 8 dezembro. 2009.

Brillet, B.; Yves, B.; Corinne, A. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica. v. 130, p. 105-120. 2007.

Brunet, F.; Godin, F.; David, J.R.; Capy, P. The *mariner* transposable element in the Drosophilidae family. Heredity. V. 73, p. 377-385. 1994.

Brunet, F.; Godin, F.; Bazin, C.; Capy, P. Phylogenetic analysis of *Mos1*-like transposable elements in the Drosophilidae. J. Mol. Evol., v. 49, p. 760-768. 1999.

Capy, P.; Langin, T.; Anxolabehere, D.; Bazin, C. Dynamics and evolution of transposable elements. Landes Biosciense, Austin, Texas, 1998.

Capy, P.; Langin, T.; Higuet, D.; Maurer, P.; Bazin, C. Do the intregases of the LTR-retrotransposons and class II elements transposases have a common ancestor? Genetica. v. 100, p. 63-72. 1997.

Clark, A.G.; Eisen, M.B.; Smith, D.R.; Bergman, C.M.; Oliver, B.; Markow, T.A.; Kaufman, T.C.; Kellis, M; Gelbart, W. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature. v. 450, p. 203-218. 2007.

DaLage, J.L.; Dergoat, G.J.; Maczkowiak, F.; Silvain, J.F.; Cariou, M.L.; Lachaise, D. A phylogeny of Drosophilidae using the Amyrel gene: questioning the Drosophila melanogaster species group boundaries. J. Zool. Syst. Evol. Res., v. 45, p. 47-63. 2007.

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

Finnegam, D. J.; Fawcett, D. H. Transposable elements in Drosophila melanogaster. Oxford surveys on eukaryotic genes, v.3, p.1-62, 1986.

Finnegan, D.J. Eukaryotic transposable elements and genome evolution. Trends Genet., v. 5, p. 103-107. 1989.

Finnegan, D. J. The Genome of *Drosophila melanogaster*, pp. 1096–1107. Academic Press, San Diego. 1992.

Grimaldi, D.A.

Katoh, T.; Nakaya, D.; Tamura, T.; Aotsuka, T. Phylogeny of the Drosophila immigrans species group (Diptera: Drosophilidae) based on Adh and Gpdh sequences. Zool. Science, v. 24, p. 913-921. 2007.

Kazazian, Jr. H.H. Mobile elements and disease. Curr. Opin. Genet. Dev. v. 8, p. 343-350. 1998.

Kidwell, M.G.; Lisch, D.R. Perspective: Transposable elements, parasitic dna, and genome evolution. Evolution. v. 55, n. 1, p. 1-24. 2001.

Lampe, D.J. Bacterial genetic methods to explore the biology of *mariner* transposons. Genetica. 2009.

Lampe, D.J.; Grant, T.E.; Robertson, H.M. Factors affecting transposition of the *Himar1mariner* transposon *in vitro*. Genetics. v. 149, p. 179-187. 1998.

Laha, T.; Loukas, A.; Wattanasatitarpa, S.; Somprakhon, J.; Kewgrai, N.; Sithithaworn, P.; Kaewkes, S.; Mitreva, M.; Brindley, P.J. The *bandit*, a new DNA transposon from a Hookworm - Possible horizontal genetic transfer between host and parasite. Plos Neglected Tropical Diseases. v. 1, n. 1, p. e35. 2007.

Le Rouzic, A.; Dupas, S.; Capy, P. Genome ecosystem and transposable elements species. Gene. v. 390, p. 214-220. 2007.

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

Loreto, E.L.S.; Carareto, C.M.A.; Capy, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity. v. 100, p. 545-554. 2008.

Markow, T.A.; O`Grady, P.M. DROSOPHILA A guide to species identification and use. Elservier, UK, 259p. 2006.

Maruyama, K.; Hartl, D.L. Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus.* J. Mol. Evol. v. 33, p. 514-524. 1991.

Muños-Lopez, M.; Siddique, A.; Bischerour, J.; Lorite, P.; Chalmers, R.; Palomeque, T. Transposition of *Mboumar-9*: Identification of a new naturally active *mariner*-Family transposon. J. Mol. Biol., v. 382, p. 567-572. 2008.

O'Brochta, D.A.; Stosic, C.D.; Pilitt, K.; Subramanian, R.A.; Hice, R.H.; Atkinson, P.W. Transpositionally active episomal *hAT* elements. BMC Mol. Biol., v. 10, p. 108. 2009. doi:10.1186/1471-2199-10-108.

Plasterk, R.H.A.; Izsvák, Z.; Ivics, Z. Resident Aliens the Tc1/mariner superfamily of transposable elements. Trends Genet., v. 15, n. 8, p. 326-332. 1999.

Picot, S.; Wallau, G.L.; Loreto, E.L.S.; Heredia, F.O.; Hua-Van, A.; Capy, P. The *mariner* transposable element in natural populations of *Drosophila simulans*. Heredity. v. 101, p. 53-59. 2008.

Pritham, E.J. Transposable elements and factors influencing their success in eukaryotes. J Heredity, v. 100, n. 5, p. 648-655. 2009.

Richardson, J.M.; Dawson, A.; O'Hagan, N.; Taylor, P.; Finnegan, D.J.; Walkinshaw, M.D. Mechanism of Mos1 transposition: insights from structural analysis. The EMBO Journal. v. 25, p. 1324-1334. 2006

Richardson, J.M.; Colloms, S.D.; Finnegan, D.J.; Walkinshaw, M.D. Molecular architecture of the Mos1 paired-end complex: the structural basis of dna transposition in a eukaryote. Cell., v. 138, p. 1096-1108. 2009.

Robe, L.J.; Valente, V.L.S.; Budnik, M.; Loreto E.L.S. Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical spacies and groups: A nuclear versus mitochondrial gene approach. Mol. Phylogenet. Evol., v. 36, p. 623-540. 2005.

Robertson, H.M.; Lampe, D.J. Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol., v. 12, n. 5, p. 850-862. 1995.

Robertson, H.M.; MacLeod, E. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol., v. 2, n. 3, p. 125-139. 1993.

Sánchez-Garcia, A.; Maside, X.; Charlesworth, B. High rate of horizontal transfer of transposable elements in *Drosophila*. Trends Genet., v. 21, n. 4, p. 200-203. 2005.

Silva, J.C.; Loreto, E.L.S.; Clark, J.B. Factors that affect the horizontal transfer of transposable elements. Curr. Issues Mol. Biol., v. 6, p. 57-72. 2004.

Sinzelle, L.; Jégot, G.; Brillet, B.; Rouleux-Bonnin, F.; Bigot, Y.; Augé-Gouillou, C. Factors acting on *Mos1* transposition efficiency. BMC Mol. Biol. v. 9, p. 106. 2008.

Tamura, K.; Subramanian, S.; Kumar, S. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol. Biol. Evol. v. 21, p. 36−44. 2004.

Tatarenkov, A.; Zurovcová, M.; Ayala, F.J. *Ddc* and *Amd* sequences resolve phylogenetic relationships of *Drosophila*. Mol. Phylogenet. Evol., v. 20, p. 321-325. 2001.

Thompson , F.C. Biosystematic database of world diptera. http://www.diptera.org/biosys.htm. Acessado em 15 dezembro. 2009.

Throckmorton, L.H. The phylogeny, ecology and geography of Drosophila. In:King RC (ed) Handbook of genetics. V 3: Invertabrates of genetic interest. Plenum, New York, pp 421-469. 1975.

van Pouderoyen, G.; Ketting, R.; Perrakis, A.; Plas-Terk, R.; Sixma, T. Crystal structure of the specific DNA-binding domain of Tc3 transposase of *C. elegans* in complex with transposon DNA. EMBO J. v. 16, p. 6044–6054. 1997.

Venner, S.; Feschotte, C.; Biémont, C. Dynamics of transposable elements: towards a comunity ecology of the genome. Trends Genet., v. 25, n. 7, p. 317-323. 2009.

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

Yeates, D.K.; Wiegmann, B.M. Congruence and controversy: Towards a higher-level phylogeny of Diptera. Annu. Rev. Entomol., v. 44, p. 397-428. 1999.