

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

**Transposon *Galileo*: uma investigação em espécies neotropicais  
de *Drosophila***

**JULIANA WOLMANN GONÇALVES**

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

Orientadora: Prof. Dra. Vera Lúcia da Silva Valente

Co-orientador: Prof. Dr. Victor Hugo Valiati

Porto Alegre, outubro de 2014.

Este trabalho foi desenvolvido no laboratório de *Drosophila* do Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS) e no laboratório de Biologia Molecular da Universidade do Vale do Rio dos Sinos (UNISINOS), com financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do Programa de Apoio aos Núcleos de Excelência – Fundação de Amparo à Pesquisa do Rio Grande do Sul (PRONEX-FAPERGS – 10/0028-7). Parte também foi desenvolvida no *Departament de Genètica i de Microbiologia* da *Universitat Autònoma de Barcelona* (UAB), com auxílio de recursos da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) e do Ministério de Ciência e Inovação da Espanha.

Dedico este trabalho a minha família,  
pois esta conquista é reflexo de todo  
amor e apoio dado.

## AGRADECIMENTOS

Aos meus orientadores, professora Vera e professor Victor Hugo pela imensa satisfação em trabalhar com pesquisadores exemplares. Verinha, uma palavra a define: entusiasmo! Obrigada por me conceder um projeto tão instigante (e complexo!!!). Victor, agradeço por ter despertado em mim a curiosidade sobre os elementos transponíveis. Sou muito grata pelos inúmeros momentos que desvendamos juntos novos horizontes. Aprendi muito com vocês dois! O carinho, a amizade, o envolvimento constante de vocês fizeram com que este trabalho fosse realizado com muita paixão.

Ao professor Ruiz pela oportunidade oferecida. Agradeço toda a dedicação neste projeto de pesquisa durante o meu período em Barcelona e sua colaboração mesmo a distância.

Às colegas Andrea, Núria e Yolanda do *Departament de Genètica i de Microbiologia* por terem me acolhido com muita alegria, pelas dicas de trabalho que me auxiliaram muito. Em especial à Alejandra, pois além de me oferecer suporte técnico, me amparou nas situações mais diversas enquanto eu estava do outro lado do continente.

Aos colegas de laboratório (atuais, novos e antigos) Brenda, Carolina, Dirleane, Gilberto, Gisele, Hermes, Lucas, Marícia, Mário, Natália, Paula, Rebeca, por tornarem o ambiente de trabalho agradável, pelas conversas, risadas, festinhas, por todos os registros fotográficos e os que ficarão na memória.

Às colegas Juliana e Maríndia pelas trocas de ideias, pelos conselhos técnicos, pelas parcerias nas horas não tão entusiasmantes. Desde meu primeiro dia no laboratório de *Drosophila* carrego muitos ensinamentos que vocês me ofereceram.

À Bibiana pelo suporte técnico dado no Laboratório de *Drosophila* da UFRGS.

À Luciana, pelas conversas nos lugares mais inusitados do departamento. Sorriso estampado no rosto é sua marca registrada!

Ao Elmo, à Ellen e agora ao Victor por todo suporte e disponibilidade constante. Eficiência, tranquilidade e carisma nunca faltaram na secretaria do PPGBM.

À Gisele Passaia, pela amizade desde a graduação. Nossos encontros foram reduzidos durante o doutorado, mas não menos importantes.

Aos meus queridos amigos Carla, Matheus, Clemir, Paulo, Ângela, Viviane, Marcos, Karla e Alex, os quais por inúmeras afinidades, torcem pelo meu sucesso profissional, apoiam todos os processos seletivos, compreendem os distanciamentos e ausências. O carinho de vocês sempre foi muito especial na minha formação.

Ao meu irmão William e cunhadinha Gabriela, pela torcida (de perto e de longe), pelo carinho e boas energias sempre enviadas. Vocês dois sempre tornaram meus dias mais alegres.

Ao Nobu, por ser a definição de companheiro. Compreendeu a importância e apoiou o nosso distanciamento por um longo período. Agradeço pelo amor, carinho e paciência. Pela disposição em editar imagens, pelas dicas de formatação de planilhas, pela leitura deste trabalho. Pelos momentos únicos que este doutorado nos trouxe!

À minha linda filha Luana, que hoje com apenas 1 ano e 9 meses, recarrega minhas energias com um único sorriso. Tornou a execução e finalização deste trabalho muito mais leve e prazeroso.

Aos meus pais, Valdir e Beatriz, pelo apoio sempre oferecido. Obrigada por se orgulharem das minhas conquistas e escolhas acadêmicas. Minhas conquistas são reflexos dos seus posicionamentos desde as minhas primeiras decisões.

Muito obrigada a todos que contribuíram e fizeram parte deste período tão importante na minha vida!

## SUMÁRIO

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES.....	7
RESUMO.....	9
ABSTRACT.....	11

### CAPÍTULO I

INTRODUÇÃO.....	13
Elementos transponíveis – componentes dinâmicos de genoma.....	13
Fonte de variabilidade e plasticidade genética.....	16
Classificando a diversidade móvel do genoma.....	19
Elemento <i>P</i> – o transposons modelo e pioneiro.....	20
Transposon <i>Galileo</i> – agente de reorganização cromossômica.....	22
Modelos polimórficos – grupos <i>willistoni</i> e <i>cardini</i> de <i>Drosophila</i> .....	30
OBJETIVOS.....	38

### CAPÍTULO II

Structural and sequence diversity of the transposon <i>Galileo</i> in the <i>Drosophila willistoni</i> genome.....	39
--	----

### CAPÍTULO II

Transposon <i>Galileo</i> nos grupos <i>willistoni</i> e <i>cardini</i> de <i>Drosophila</i> .....	77
--	----

### CAPÍTULO IV

DISCUSSÃO E CONCLUSÕES.....	113
PERSPECTIVAS.....	116
REFERÊNCIAS BIBLIOGRÁFICAS .....	117

## ABREVIATURAS, SÍMBOLOS E UNIDADES

<		menor
°C		grau Celsius
μL		microlitro
μM		micromolar
<i>Adh</i>		álcool desidrogenase
<i>Amd</i>		alfa metildopa
<i>DSBs</i>	<i>(Double Stranded Breaks)</i>	quebra de dupla fita
<i>HGs</i>	<i>(Host Genes)</i>	genes hospedeiros
<i>HT</i>	<i>(Horizontal Transfer)</i>	transferência horizontal
<i>LINEs</i>	<i>(Long Interspersed Nuclear Elements)</i>	elementos nucleares intercalados longos
<i>LTRs</i>	<i>(Long Terminal Repeats)</i>	repetições terminais longas
M		molar
mM		milimolar
ng		nanograma
<i>ORFs</i>	<i>(Open Reading Frames)</i>	regiões de leituras abertas
<i>PCR</i>	<i>(Polymerase Chain Reaction)</i>	reação em cadeia da polimerase
pmol		picomol
<i>RIP</i>	<i>(Repeat-induced Point Mutation)</i>	mutação de ponto induzida por repetição
<i>subTIRs</i>	<i>(Subterminal Inverted Repeats)</i>	repetições subterminais invertidas

<i>TEs</i>	<i>(Transposable Elements)</i>	elementos transponíveis
<i>THAP</i>	<i>(Thanatos-associated protein)</i>	domínio de ligação ao DNA
<i>TIRs</i>	<i>(Inverted Terminal Repeats)</i>	repetições terminais invertidas
<i>TPase</i>		transposase
<i>TSD</i>	<i>(Target Site Duplication)</i>	duplicação do sítio alvo



## RESUMO

*Galileo* junto aos elementos *P* e *1360* (Hoppel) pertence à superfamília *P* de transposons. Foi originalmente descoberto em *Drosophila buzzatii*, onde três inversões cromossômicas segregantes foram originadas por recombinação ectópica entre cópias de *Galileo*. A investigação de *Galileo* nos 12 primeiros genomas de *Drosophila* demonstrou sua presença em seis espécies, sendo que em *Drosophila willistoni* foi encontrado em abundância. O grupo *cardini* e o grupo *willistoni* de *Drosophila* ainda que pertençam a subgêneros diferentes, são cromossomicamente polimórficos e ecologicamente similares, e assim, apresentam-se como promissores modelos para o estudo deste transposon. Com intuito de elucidar o cenário evolutivo de *Galileo* e seu papel na evolução do genoma de espécies neotropicais, foram caracterizadas detalhadamente as cópias de *Galileo* presentes no genoma de *D. willistoni* e realizou-se uma busca experimental em espécies do grupo *willistoni* e do subgrupo *cardini* de *Drosophila*. Na busca *in silico*, diferentes ferramentas de bioinformática foram utilizadas e como sonda fez-se o uso da maior cópia de *Galileo* descrita para *D. willistoni*, da repetição terminal invertida (*TIR*) e de segmentos da transposase do elemento. No total 191 cópias foram caracterizadas, com variação considerável em comprimento e estrutura. Cópias com a transposase potencialmente funcional não foram encontradas, todas as caracterizadas eram elementos não-autônomos. Duas subfamílias de *Galileo*, aqui nomeadas de *V* e *W*, com uma divergência nucleotídica substancial, foram observadas por meio de análises filogenéticas com segmentos da transposase e com as *TIRs* do elemento. A investigação experimental foi realizada pela técnica de *PCR* com dois pares de *primers* que amplificam segmentos de 470 e 530 pb da transposase. Do grupo *willistoni* sequências correspondentes a *Galileo* foram isoladas de *D. willistoni* (três linhagens, além da sequenciada), *D. tropicalis*, *D. equinoxialis*, *D. insularis* e *D. paulistorum* (seis semiespécies: Amazônica, Andino-Brasileira, Centro-Americana, Interior, Orinocana e Transicional) do subgrupo *willistoni*; *D. nebulosa* e *D. capricorni* do subgrupo *bocainensis*. Do subgrupo *cardini*, *Galileo* foi isolado de *D. cardini*, *D. neocardini* e *D. neomorpha*. O total de 109 sequências, clonadas e sequenciadas foram utilizadas na reconstrução das relações evolutivas de

*Galileo*. Esta análise corroborou com dois grupos altamente divergentes, mostrando a coexistência das subfamílias *V* e *W* de *Galileo* em espécies do grupo *willistoni* e ainda, a presença da subfamília *V*, que divergiu recentemente, em espécies do subgrupo *cardini*. O panorama evolutivo sugere que *Galileo* vem sendo mantido por meio de transferência vertical e horizontal (uma vez que ambas não são mutuamente exclusivas) no grupo *willistoni* e no subgrupo *cardini* de *Drosophila*. Além disso, é possível que eventos de introgressão, tenham desempenhado papel importante na manutenção deste transposon, especialmente em espécies do subgrupo *willistoni*.

## ABSTRACT

*Galileo* along with the *P* and 1360 elements (*Hoppel*) belongs to the *P* superfamily of transposons. It was originally discovered in *Drosophila buzzatii*, in which three segregating chromosomal inversions were generated by ectopic recombination between *Galileo* copies. *Galileo* was identified in six of the first 12 genome of *Drosophila*, being abundant in *Drosophila willistoni*. Although the *cardini* and the *willistoni* groups of *Drosophila* belong to different subgenus, they are chromosomally polymorphic and ecologically similar, which makes them promising models for the study of this transposon. With the aim of explaining the evolutionary scenario of *Galileo* and its role in the evolution of the genome of neotropical species, we carried out a detailed characterization of all *Galileo* copies present in the *D. willistoni* genome and we also carried out an experimental search among species of the *willistoni* group and the *cardini* subgroup of *Drosophila*. For the *in silico* search, we used different bioinformatics tools and we used the nearly-complete copy of *Galileo* described for *D. willistoni*, terminal inverted repeats (TIRs) and segments of the element transposase as query. A total of 191 copies were classified and showed remarkable variation in their length and structure. We did not find any copies with potentially functional transposase; all analyzed copies were non-autonomous. Two *Galileo* subfamilies, which we named subfamilies *V* and *W*, with considerable nucleotide divergence, were observed through phylogenetic analysis of transposase segments and of the element's TIRs. For the experimental research, we used the PCR technic with two pairs of primers that amplify the transposase segments of 470 and 530 bp. In the *willistoni* group, we have isolated *Galileo* sequences from *D. willistoni* (three lineages, besides the sequenced one); *D. tropicalis*, *D. equinoxialis*, *D. insularis* and *D. paulistorum* (six semi-species: Amazonian, Andean-Brazilian, Centroamerican, Interior, Orinocan and Transitional) in the *willistoni* subgroup; *D. nebulosa* and *D. capricorni* in the *bocainensis* subgroup. In the *cardini* subgroup, we have isolated *Galileo* from *D. cardini*, *D. neocardini* and *D. neomorpha*. A total of 109 sequences, cloned and sequenced, were used to analyze the *Galileo* evolutionary relationships. This analysis confirmed the existence of two highly divergent groups, demonstrating the coexistence of the *Galileo V* and *W* subfamilies in species of the *willistoni* group.

Besides that, we noticed the presence of the *V* subfamily, which has recently diverged, in species of the *cardini* subgroup. The evolutionary scenario suggests that *Galileo* has been maintained through vertical and horizontal transfers (considering that they are not mutually exclusive) in the *willistoni* group and in the *cardini* subgroup of *Drosophila*. Furthermore, we believe that introgression events have played an important role on the maintenance of this transposon, especially in species of the *willistoni* subgroup.

# CAPÍTULO I

---

## INTRODUÇÃO

### **Elementos transponíveis – componentes dinâmicos do genoma**

Elementos transponíveis (*TEs*, do inglês *Transposable Elements*) são segmentos de DNA repetitivo, presentes nos mais diversos organismos já estudados, com capacidade de mobilização intrínseca ou não. Como unidades genéticas móveis, exibem uma ampla diversidade estrutural e mecanismos de transposição. Segundo Wicker *et al.* (2007) um elemento seria autônomo simplesmente se ele codificar todos os domínios necessários para a sua transposição, independentemente de estar funcional e não-autônomo ou defectivo, por sua vez, são definidos como qualquer grupo de elementos que tenha perdido algum ou todos os domínios encontrados nos elementos autônomos.

A dinâmica dos *TEs* no genoma de um organismo pode ser dividida em três fases principais: (1) fase invasora, caracterizada pela inserção do elemento e aumento do número de cópias, acompanhada por mutações que originam variantes defectivas; (2) inativação ou adaptação dependente do impacto do elemento sobre o valor adaptativo do organismo, onde a permanência do *TE* é resultado do balanço entre a transposição, mutação e seleção; (3) estágio de senescência, quando não há ocorrência de *TEs* autônomos nem do aumento do número de cópias e as sequências defectivas são alvo de crescente divergência (Kidwell e Lisch, 1997; Brookfield, 2005; Almeida e Carareto, 2005).

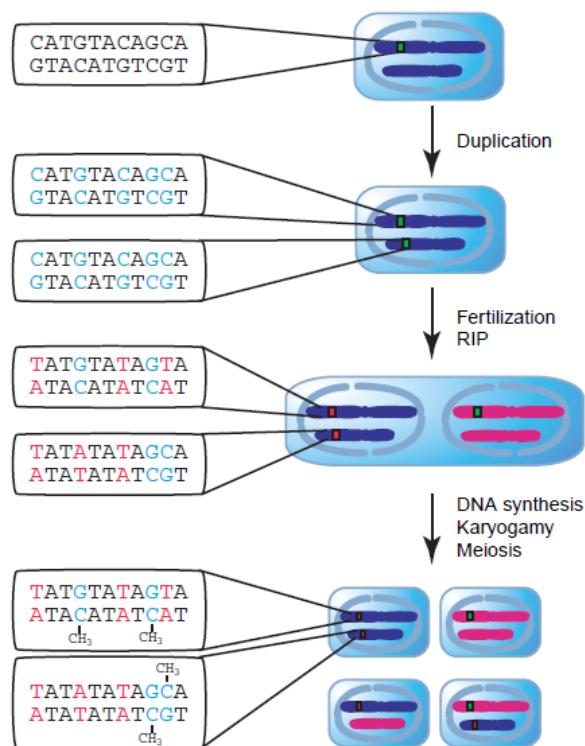
Depois da rápida propagação dos *TEs*, a mobilidade atinge um impasse devido ao acúmulo de cópias defectivas e silenciamento epigenético dirigido pelo genoma do organismo. Subsequentemente, cópias não mais autônomas deixam de ser alvo de pressões seletivas, ficam expostas a mutações aleatórias, sofrem erosão nas suas sequências e são eliminadas a menos que ocorra uma reativação por recombinação entre elementos ou suspensão de mecanismos repressores. Uma rota de fuga da extinção do elemento pelo

organismo é a invasão de um genoma não infectado pelo mecanismo de transferência horizontal (*HT*, do inglês *Horizontal Transfer*), dando início a outro "ciclo de vida" em um novo ambiente permissivo. Essa capacidade de atravessar fronteiras, isolamento reprodutivo entre as espécies e invadir novos genomas é uma característica marcante dos *TEs*. Uma *HT* bem sucedida requer uma transferência do *TE* na linha germinal das espécies receptoras, seguida por uma alta taxa de transposição que conduz a uma rápida propagação no genoma, bem como para a população por transmissão vertical (Le Rouzic e Capy, 2005).

A capacidade inerente dos *TEs* na mobilização e integração em cromossomos aumenta a possibilidade de sua transferência horizontal em comparação com sequências não-móveis, no entanto, essa propensão pode variar em função do mecanismo de transposição de cada grupo (Kidwell, 1992; Schaack *et al.*, 2010). Elementos que possuem intermediário de DNA são aparentemente mais estáveis que retrotransposons que utilizam intermediário de RNA na mobilização e, portanto, é provável que tenham uma maior capacidade de *HT*. Além disso, transposons que são transferidos para táxons amplamente divergentes, por dependerem apenas da transposase e não de fatores específicos do hospedeiro, são mais propensos de manterem a atividade. A estrutura da maioria dos elementos de DNA é bastante simples, geralmente constituída por uma única região codificante, flanqueada por curtas repetições terminais invertidas. Em alguns casos não possuem promotor e utilizam os adjacentes do hospedeiro para serem expressos. Segundo Schaack *et al.* (2010) é tentador interpretar que esta organização mínima seria uma adaptação à *HT*. Alguns retrotransposons com repetições terminais longas (*LTR*, do inglês *Long Terminal Repeats*) codificam a proteína do envelope tal como acontece em retrovírus e essa característica poderia facilitar a capacidade de infecção, embora casos de *HT* envolvendo retrotransposons sem essa capacidade já foram relatados (de Setta *et al.*, 2009). Ainda que elementos autônomos e não autônomos possam ser mobilizados por proteínas que estejam disponíveis, os elementos que não

codificam a proteína requerida para a própria transposição, podem ter uma probabilidade reduzida de se transferirem horizontalmente.

Muitos organismos têm desenvolvido sistemas de defesa para repressão efetiva de *TEs*, evitando assim possíveis efeitos negativos, como o processo de mutação de ponto induzida por repetição (*RIP*, do inglês, *Repeat-induced Point Mutation*), o qual identifica sequências duplicadas e introduz mutações (Figura 1). Para escapar da extinção, análogos aos parasitas, os *TEs* possivelmente utilizam diferentes estratégias (Figura 2), como iniciar com replicação rápida e também *HT*, evoluindo com uma estratégia de baixa “virulência” ao longo do tempo, de forma que fique imperceptível à atuação da seleção no seu hospedeiro (Galagan e Selker, 2004). A persistência pode ser alcançada através de mecanismos de auto-regulação que limitam o número de cópias ou por evolução da preferência para a inserção em regiões “seguras” no genoma, como em genes com alto número de cópias ou na heterocromatina (Bushman, 2003; Zhou e Eickbush, 2009). Por fim, a sobrevivência do *TE* como uma sequência codificante pode ser alcançada pelo processo de domesticação molecular, com a integração estável no genoma e aquisição de uma função nova que beneficia o organismo portador.



Neste caso, embora seja possível identificar a sequência, a capacidade de ser um elemento transponível é perdida e passa a ser considerado como um neogene (Pinsker *et al.*, 2001).

Figura 1: Processo de mutação de ponto induzida por repetição (*RIP*) em *Neurospora crassa*. Duplicações que ocorrem durante a fase vegetativa do organismo são detectadas pelo *RIP* durante o ciclo sexual após a fecundação, mas antes da síntese de DNA e da fusão nuclear. Sequências duplicadas que são maiores que  $\approx 400\text{pb}$  ( $\approx 1\text{kb}$  para duplicações não ligadas) e compartilham identidade maior que 80% são detectadas. Numerosas

mutações pontuais C:G para T:A são introduzidas em ambas as cópias (não mutados são mostrados em azul; mutações são mostradas em vermelho; apenas um pequeno número de pares de bases são mostrados com a finalidade de exemplificação). Sequências mutadas por *RIP* são alvos frequentes para a metilação, o que resulta no silenciamento transcricional em *Neurospora* (Modificada de Galagan e Selker, 2004).

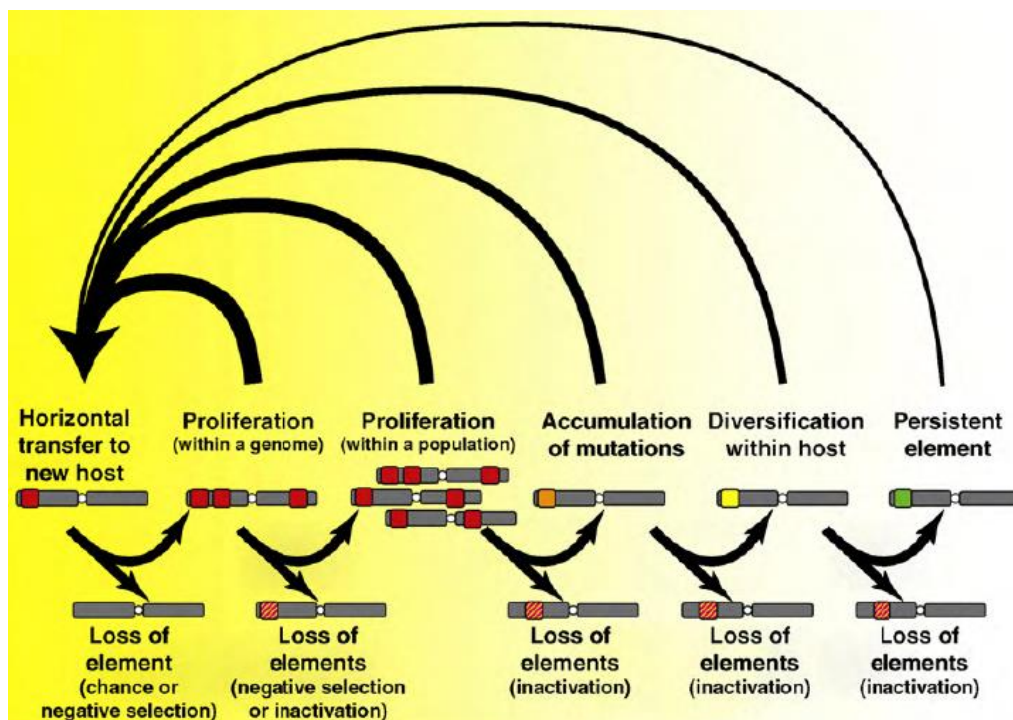


Figura 2: Modelo simplificado da dinâmica de famílias de TEs e a importância da transferência horizontal (Modificada de Schaack *et al.*, 2010).

## Fonte de variabilidade e plasticidade genética

Capazes de se propagarem e de mudarem sua localização dentro de genomas ou entre organismos, as consequências da mobilidade dos *TEs* vão além de possíveis efeitos deletérios. Ocupando uma grande fração do genoma de muitos eucariotos, o movimento e acúmulo dos *TEs* podem contribuir na evolução dos genomas hospedeiros por três vias distintas: alterando a função gênica após uma inserção; induzindo rearranjos cromossômicos; servindo como fonte de material codificante ou não e, conseqüentemente, contribuindo para o aumento do genoma, permitindo a emergência de novos genes e sequências regulatórias (Finnegan, 1989; McDonald, 1993; Feschotte e Pritham, 2007, Chénais *et al.*, 2012).



Como potentes agentes mutagênicos, com taxas de transposição maiores que as taxas de mutações espontâneas, os *TEs* podem produzir um amplo espectro de alterações genéticas. São capazes de gerar desde pequenas mudanças silenciosas, detectáveis apenas em nível de sequência de DNA a efeitos perceptíveis em caracteres fenotípicos, dependendo da localização e instabilidade provocada no genoma do organismo. No processo de inserção e também excisão de *TEs*, a expressão de um gene pode ser afetada de diferentes formas (Kidwell e Lisch, 2001). Um *TE* pode inibir o produto viável de um gene ou ainda interferir em promotores, introns, regiões não codificadoras, provocar perturbações regulatórias epigenéticas e até mesmo a completa perda da função gênica (Wessler *et al.*, 1995; Kidwell e Lisch, 1997; Deininger e Batzer, 1999; Feschotte *et al.*, 2002, Kazazian, 2004, Feschotte, 2008).

Conferindo vantagem adaptativa para o hospedeiro, o *TE* pode ser mantido e domesticado como citado anteriormente. As formas de domesticação podem diferenciar, incluindo a formação de quimeras entre o *TE* e genes do hospedeiro, como o exemplo do gene quimérico *SETMAR* caracterizado em primatas (Cordaux *et al.*, 2006) ou a cooptação de uma função intacta do *TE* beneficiando o hospedeiro, exemplificada com os telômeros de *Drosophila* (Pardue e DeBaryshe, 2003). Além disso, as sequências de *TEs* contêm elementos regulatórios, incluindo sítios de ligação para fatores de transcrição, que podem ser aproveitados em novas redes regulatórias (Feschotte, 2008).

*TEs* ocasionalmente podem carregar fragmentos do genoma do hospedeiro durante o evento de transposição, ocasionando a duplicação maciça de genes ou embaralhamento de éxons (Morgante *et al.*, 2005). Outra maneira pela qual a atividade de um *TE* pode levar à formação de novos genes é por meio do reconhecimento acidental de transcritos celulares pela maquinaria de transposição codificada por retrotransposons. Durante a evolução dos mamíferos, por exemplo, esse processo tem gerado milhares de duplicatas retropostas de genes, muitos dos quais evoluíram para assumir novas funções celulares (Kaessmann *et al.*, 2009).

Inversões cromossômicas podem desempenhar um papel significativo na adaptação, especiação e evolução dos cromossomos como demonstrado por Hoffmann e Rieseberg (2008). A mobilização de *TEs* aumenta o número da dispersão de cópias e conseqüentemente favorece o aumento de rearranjos cromossômicos, uma vez que a recombinação ectópica entre cópias ou até mesmo pequenos fragmentos de *TEs* já foi descrita (Cáceres *et al.*, 2001; McVean 2010). Evidências diretas, implicando recombinação ectópica mediada por *TE*, foram encontradas em cinco inversões polimórficas no gênero *Drosophila*: três em *D. buzzatii*, geradas por *Galileo* (Cáceres *et al.*, 1999; Casals *et al.*, 2003; Delprat *et al.*, 2009), duas envolvendo o elemento *But5*, em *D. uniseta* e *D. mojavensis* (Prada, 2010; Guillén e Ruiz, 2012). Em *D. pseudoobscura* uma sequência repetitiva foi encontrada em junções de inversões e os autores sugerem seu envolvimento na origem das mesmas por meio do mecanismo de recombinação (Stephen *et al.*, 2005). Em *D. melanogaster* e espécies intimamente relacionadas, *D. simulans* e *D. yakuba*, foram analisadas 29 inversões, mas dessas apenas duas apresentavam repetições invertidas em ambos os pontos de quebra, sugerindo um mecanismo diferente daquele observado no envolvimento de *TEs* (Ranz *et al.*, 2007).

A taxa de fixação para rearranjos varia significativamente dentro e entre os grupos de animais (Coghlan *et al.*, 2005), sendo que o gênero *Drosophila* apresenta uma das mais altas taxas em todos os eucariotos, onde mecanismos citológicos especiais, permitem heterozigotos para inversões paracêntricas (Ranz *et al.*, 2001; Bhutkar *et al.*, 2008). A taxa de fixação de rearranjos é maior no subgênero *Sophophora* quando comparada ao *Drosophila* (Papaceit *et al.*, 2006). Tem sido proposto que a variação da atividade de transposição de *TEs*, poderia contribuir na variação das taxas de fixação de rearranjos cromossômicos (González *et al.*, 2007; Guillén e Ruiz, 2012).

## Classificando a diversidade móvel do genoma

Devido à abundância e diversidade de *TEs* e a velocidade com que sequências de DNA estão emergindo a partir do estudo de genomas disponíveis, Wicker *et al.* (2007) reformularam o sistema de classificação para *TEs*, com intuito de tornar simples e eficiente as estratégias de identificação, anotação e agrupamento. Hierarquicamente, inclui os níveis de classe, subclasse, ordem, superfamília, família e subfamília. O nível mais elevado divide os *TEs* pela presença ou ausência de um intermediário de RNA na transposição, como elementos da classe I e da classe II, respectivamente, como proposto por Finnegan (1989).

Em uma breve abordagem, a subclasse distingue elementos da classe II que utilizam o mecanismo replicativo daqueles que se transpõem pelo mecanismo conservativo. A ordem marca diferenças importantes no mecanismo de inserção, conseqüentemente, na organização geral e características enzimáticas do processo. As superfamílias dentro de uma ordem compartilham uma estratégia de replicação, mas são distinguidas por características uniformes e amplamente difundidas, tais como a estrutura protéica e domínios não-codificantes. Além disso, superfamílias diferem na presença e tamanho da duplicação do sítio alvo de inserção. Superfamílias são divididas em famílias, as quais são definidas pela conservação da sequência de DNA, embora essa seja mínima e restrita à região codificante. Subfamílias são definidas com base nos dados filogenéticos e podem, em casos específicos, servir para distinguir elementos autônomos de não-autônomos.

Com ênfase na classe II, onde os elementos são chamados de transposons, a transposase é requerida para que ocorra a mobilização do *TE*. Contém duas subclasses, os quais se distinguem pelo número de fitas de DNA que são cortadas durante a transposição. A subclasse 1 compreende *TEs* do tipo “corta-e-cola” da ordem TIR, caracterizados por conterem repetições terminais invertidas de comprimento variável. São descritas dez superfamílias, diferenciadas pelas *TIRs* (do inglês *Inverted Terminal Repeats*) e *TSD* (do inglês *Target Site Duplication*): *Tc1–Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*,

*PiggyBac*, *PIF-Harbinger* e *CACTA*. Estes *TEs* podem aumentar suas cópias realizando a transposição durante a replicação cromossômica, de uma posição que já foi replicada para outra em que a forquilha de replicação ainda esteja formada. Um mecanismo alternativo é a origem de uma cópia extra durante o reparo da lacuna (*gap*) que foi originada no momento da excisão do *TE*. A transposase reconhece as *TIRs* e realiza o corte em ambas as fitas de DNA. Apesar de pouco conhecida, entre os elementos da subclasse 1, temos a superfamília *Cryptom* (ordem *Cryptom*), que se diferencia principalmente por conter uma tirosina recombinase. Por outro lado, a subclasse 2 de transposons, caracteriza-se por reunir elementos que se mobilizam via o mecanismo “copia-e-cola”, no qual a clivagem é realizada em uma fita e apenas esta é transferida para uma região nova. Nessa subclasse estão as superfamílias *Helitron* (que usa o mecanismo “círculo rolante” na mobilização) e *Maverick*, composta por elementos de até 20 kb (Wicker *et al.*, 2007).

### **Elemento *P* – o transposon modelo e pioneiro**

A superfamília *P* é composta por três transposons: elemento *P* (O’Hare e Rubin, 1983), 1360 (também chamado de *Hoppel* e *PROTOP*) (Kapitonov e Jurka 2003; Reiss *et al.* 2003) e *Galileo* (Marzo *et al.* 2008) (Rius *et al.*, 2013).

O elemento *P* é notavelmente importante na elucidação de mecanismos e efeitos que envolvem a mobilização de *TEs*, uma vez que foi modelo e pioneiro no estudo da chamada disgenesia do híbrido (*HD*), fenômeno onde hipermutabilidade, rearranjos cromossômicos, e esterilidade gonadal são observados em cruzamentos de algumas linhagens de *Drosophila* (Kidwell *et al.*, 1977; Yannopoulos *et al.*, 1987; Crozatier *et al.*, 1988; Lozovskaya *et al.*, 1990; Petrov *et al.*, 1995; Evgen’ev *et al.*, 1997). Por ser um dos primeiros *TEs* a ser descoberto, caracterizado molecularmente e estudado com profundidade, rapidamente se tornou uma ferramenta na pesquisa científica (Rio 1991, 2002; Kidwell 1994; Engels, 1996; Pinsker *et al.* 2001). Serviu como agente mutagênico em estudos de clonagem, transformação e expressão gênica, como vetor para recombinação sítio-específica e substituição de gene (Engels, 1989).

O elemento *P* íntegro, canônico, possui 2,9 kb de comprimento e quatro *ORFs* (do inglês *Open Reading Frames*), com dois genes sobrepostos, um que codifica a transposase (87 kDa) e outro que determina o repressor da transposição (66 kDa) (Craig, 1990). Possui *TIRs* de 31 pb e repetições subterminais invertidas de 11 pb (*subTIRs*, do inglês *Subterminal Inverted Repeats*), as quais atuam como potenciadores para a transposição. No processo de inserção, gera uma duplicação de 8 pb no sítio alvo (Rio, 2002).

Além do elemento *P* canônico, há pelo menos três subfamílias bem caracterizadas designadas como *tipo-M*, *-O*, e *tipo-T* no grupo *obscura* de *Drosophila*. A diferença na sequência entre as três subfamílias em nível de sequência nucleotídica é de cerca de 30%, mas a estrutura básica das quatro *ORFs* e das *TIRs* é a mesma. A subfamília *M* é mais próxima dos elementos *P* canônicos, enquanto que a *O* representa as cópias mais degeneradas, com *TIRs* interrompidas (Hagemann *et al.*, 1992, 1994, 1996a, b). As subfamílias *M* e *O* também foram descritas para os grupos *saltans* e *willistoni* (Setta *et al.* 2007), enquanto Nouaud *et al.*, (2003) descreveram uma subfamília restrita ao subgrupo *montium* do grupo *melanogaster*, a subfamília *K*. Loreto *et al.* (2012), por meio de busca *in silico* e experimental no gênero *Drosophila*, propõem uma reclassificação quanto às subfamílias até então descritas. As subfamílias *M* e *O* poderiam ser ampliadas e a subfamília *T*, subdividida e outras seis subfamílias criadas, devido à formação de clados significativamente divergentes. Com relação à coexistência de subfamílias de *TEs* em um genoma, Loreto *et al.* (2012) observaram que várias espécies possuem elementos *P* de subfamílias distintas. Dois casos foram destacados pelos autores: *D. willistoni*, por apresentar cinco das onze subfamílias observadas na relação evolutiva do elemento e *D. bocqueti*, que possui as sequências das duas mais diversificadas subfamílias, tanto em nível nucleotídico, quanto protéico.

No gênero *Drosophila*, o elemento *P* é amplamente distribuído no subgênero *Sophophora* (Daniels *et al.*, 1990; Hagemann *et al.*, 1992, 1994, 1996a, b; Clark e Kidwell 1997), mas a ocorrência no subgênero *Drosophila* é mais restrita (Loreto *et al.*, 2001, 2012). A sequência nucleotídica do elemento

*P* canônico isolado de *D. willistoni* é quase idêntica (uma única substituição em 2,9 kb) à caracterizada para o elemento *P* de *D. melanogaster* (Daniels et al. 1990). Essa descoberta forneceu a primeira e forte evidência do recente evento de *HT* entre *D. willistoni* e *D. melanogaster* quanto à existência do elemento *P*.

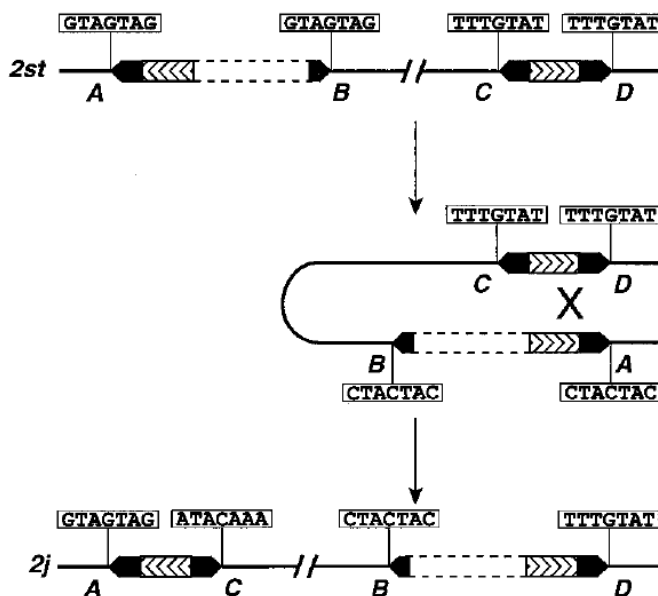
O elemento *P* contém um domínio THAP (do inglês *Thanatos-Associated Protein*) de ligação ao DNA, o qual já foi relacionado com casos de domesticação pelo hospedeiro. É um motivo evolutivamente conservado e compartilhado por diferentes proteínas de origem animal, como reguladores do ciclo celular, fatores pró-apoptóticos, repressores da transcrição e proteínas associadas à cromatina (Clouaire et al., 2005; Quesneville et al., 2005; Roussigne et al., 2003; Sinzelle et al., 2009). No genoma humano 12 genes possuem o domínio THAP e para um deles (THAP9) foi recentemente mostrado que uma transposase ativa, de um elemento *P*, é codificada (Majumdar et al., 2013).

### **Transposon *Galileo* – agente de reorganização cromossômica**

Embora *Galileo* seja caracterizado por conter longas repetições terminais invertidas (*TIRs*), similar às encontradas em elementos *Foldback*, foi classificado como membro da superfamília *P* de transposons com base na inferida sequência da transposase (Marzo et al., 2008). Descoberto por Cáceres et al. (1999) durante o isolamento e caracterização da inversão *2j*, comumente observada em *D. buzzatii*, foi apontado como o responsável pela geração da inversão. As inserções localizadas nos pontos de quebra da inversão, eram similares em aproximadamente 274 pb e apresentavam o sítio alvo duplicado (*TSD*, do inglês *Target Sites Duplication*) trocado, fato que evidenciou o mecanismo de recombinação ectópica utilizado pelas cópias de *Galileo* na origem do rearranjo (Figura 3). Em um trabalho subsequente, Cáceres et al. (2001) estimaram que o arranjo teria surgido há aproximadamente 1 milhão de anos e, de acordo com o tempo de coalescência, os alelos *2j* teriam se diferenciado há menos de 84 mil anos.

Esta foi a primeira evidência direta de que um elemento transponível estaria envolvido em uma inversão cromossômica e representa uma rápida reestruturação genômica nunca vista antes na natureza. As inserções, deleções e os outros rearranjos não estão distribuídos uniformemente ao longo dos cromossomos de *D. buzzatii*, mas parecem ter ocorrido após inserções de *Galileo*, dentro ou muito deles (Figura 4). A partir dessa observação originou-se a ideia de que regiões que continham *Galileo* são propensas a mutações e rearranjos cromossômicos (Cáceres *et al.*, 2001).

Figura 3. Modelo que exemplifica a origem da inversão cromossômica *2j* em *D. buzzatii*. Primeiro duas seqüências homólogas do *TE* se inseriram entre as seqüências A e B e entre C e D do



cromossomo com o arranjo padrão, gerando duplicações de 7 nucleotídeos (destacados nos retângulos). A recombinação ectópica entre as seqüências do *TE* em orientação oposta resultou na inversão do fragmento cromossômico entre as cópias. Como resultado desse processo, os sítios alvos duplicados foram trocados entre as inserções do *TE* adquirindo a atual orientação encontrada nos cromossomos com a inversão *2j*. Linhas pontilhadas referem-se às seqüências nucleotídicas que não estão representadas em escala (Modificado de Cáceres *et al.*, 1999).

Com intuito de esclarecer o mecanismo e verificar a implicação de *TEs* na origem de outras inversões em *D. buzzatii*, os pontos de quebra das inversões  $2q^7$  e  $2z^3$  foram isolados e caracterizados molecularmente. *Galileo* foi novamente o agente indutor das inversões, servindo de substrato para eventos de recombinação ectópica (Casals *et al.* 2003; Delprat *et al.* 2009). Em ambos trabalhos, os autores observaram a troca dos *TSDs* entre as cópias de *Galileo* e ainda constataram que estes elementos eram quiméricos. Com a observação de cópias com extremidades mais divergentes, quando

comparadas com *TIRs* de diferentes cópias de *Galileo*, Delprat *et al.* (2009) fizeram uma busca mais detalhada e encontraram regiões permutadas entre duas variantes do elemento (Figura 5). Com base nas estruturas quiméricas caracterizadas, os autores sustentam que o evento de recombinação que originou a inversão  $2z^3$  ocorreu dentro de apenas 141 pb da região central das cópias. É proposto que as longas *TIRs* de *Galileo* induzem a formação de estruturas secundárias e quebras de dupla fita de DNA (*DSBs*, do inglês *Double Stranded Breaks*) em alta taxa, o que favorece eventos de recombinação e contribui para capacidade de gerar inversões cromossômicas. O fato do evento de recombinação ectópica ter ocorrido entre a região central de duas variantes de *Galileo*, é plausível considerando-se a ação de nuclease, clivando a alça (*loop*) da estrutura.

Com a disponibilidade dos doze primeiros genomas de *Drosophila* (Clark *et al.*, 2007), Marzo *et al.* (2008) expandiram a busca de *Galileo* para verificar se esse *TE* estava presente em espécies distantemente relacionadas ao grupo *repleta* de *Drosophila*. Por meio da análise *in silico* demonstraram que *Galileo* possui uma ampla distribuição, com cópias do elemento em *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. virilis* e *D. mojavensis*. *Galileo* foi encontrado em abundância (com mais de 100 cópias em todas as espécies), sendo que *D. willistoni* apresentou o maior número de cópias. A cópia mais completa do elemento *Galileo* encontrada em *D. willistoni*, apresentou 4386 pb, embora nenhuma transposase funcional fora identificada. O mesmo ocorre nas demais espécies que possuem códons de parada, alterações no quadro de leitura e/ou deleções nas sequências dos elementos encontrados).

Estudos que caracterizaram *Galileo* destacam as notáveis longas *TIRs*, com extremidades de até 1229 pb e 1250 pb identificadas em *D. buzzatii* e *D. mojavensis*, respectivamente (Casals *et al.*, 2005; Marzo *et al.*, 2013). A análise filogenética realizada por Marzo *et al.* (2008), sugere que *Galileo* surgiu a partir de um ancestral com *TIRs* muito mais curtas. *Galileo* aparenta ter uma dinâmica reestruturação de suas *TIRs* com divergência nas sequências nucleotídicas e comprimento variável das *TIRs* observado dentro e entre as



espécies. *TIRs* podem encurtar por deleção, mas é muito provável que possam ser alongadas por duplicações internas, recombinação desigual, conversão gênica (Richardson *et al.*, 1998) ou reparo de quebras de fita simples (Kapitonov e Jurka, 2006). Na análise do genoma de *D. mojavensis*, uma detalhada caracterização das *TIRs* de *Galileo*, revelou a combinação de duas maneiras com as quais as extremidades de um *TE* podem tornarem-se mais longas: a expansão de repetições em *tandem* e o recrutamento de sequências internas para as extremidades (Marzo *et al.*, 2013).

Casals *et al.* (2005) encontraram um alto grau de variabilidade interna nas cópias de *Galileo* em *D. buzzatii* (Figura 6). Para *D. mojavensis* observou-se também desde cópias mais completas, até cópias restritas às *TIRs* (Marzo *et al.*, 2013). De acordo com as estruturas das cópias caracterizadas no genoma de *D. mojavensis*, cinco grupos foram formados: cópias quase completas (*NC*, do inglês *nearly-complete*, derivadas de cópias quase completas (*DD*, do inglês *deletion derivatives*), constituídas por *TIRs* (*T2*, do inglês *two TIR elements*), formadas por *TIRs* estendidas ou recombinantes (*2RT*, do inglês *two extended or recombinant TIR elements*) e compostas apenas por uma *TIR* (*solo-TIR*) (Figuras 7). Além disso, a presença de subfamílias de *Galileo* em um mesmo genoma foi demonstrada para *D. buzzatii* (subfamílias G, K e N), *D. virilis* (subfamílias A e B) e para *D. mojavensis* (subfamílias C, D, E, F e X) (Delprat *et al.*, 2009; Marzo *et al.*, 2008; Marzo *et al.*, 2013). Apesar da hipótese de *HT* não ter sido descartada, ainda que em combinação com a transmissão vertical, os autores concluem que é mais plausível que a subfamílias de *Galileo* teriam divergido de um ancestral comum no genoma.

Apesar de não ter sido encontrado *Galileo* potencialmente ativo nas espécies investigadas, Marzo *et al.* (2008) forneceram evidências de recente (<1 mya) atividade transposicional de *Galileo* nas espécies mais polimórficas do gênero: *D. ananassae*, *D. persimilis*, *D. pseudoobscura*, e *D. willistoni* (Schaeffer *et al.*, 2008). Em *D. mojavensis*, dotada de um menor número de inversões, *Galileo* era funcional há ≈1,5 mya enquanto que em *D. virilis*, cromossomicamente monomórfica, a capacidade de transposição foi inferida

como a mais antiga ( $\approx 3$  mya). Ainda que possam ser coincidência, os dados mostram uma correlação entre o número de inversões e o tempo de atividade mais recente de *Galileo*.

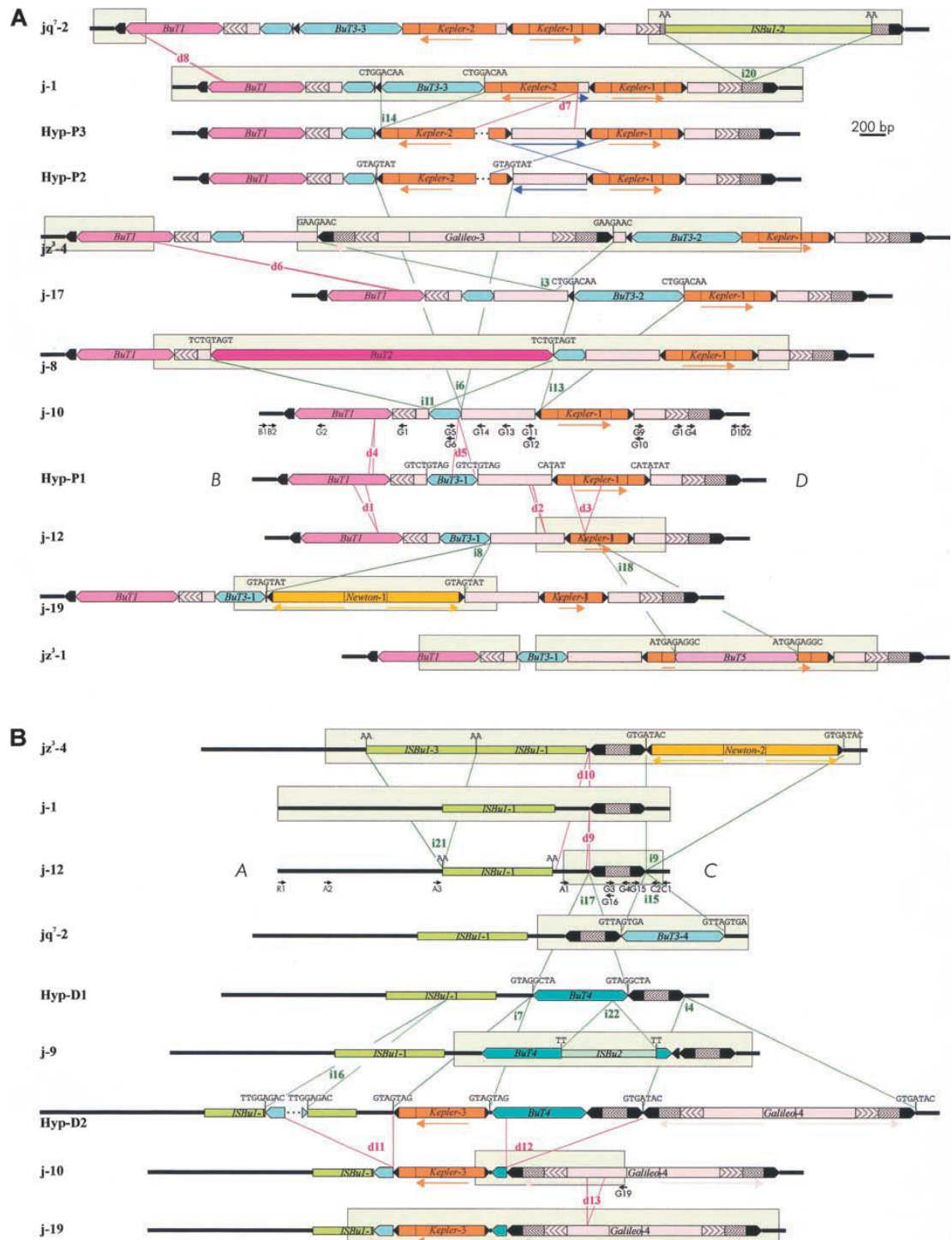


Figura 4. Representação esquemática das estruturas encontradas nos pontos de quebra proximal (A) e distal (B) da inversão 2j em 30 linhagens de *D. buzzatii* estudadas. Todas as estruturas diferentes são apresentadas, com exceção da *j-16* localizada no ponto de quebra proximal, que difere da *jz3-4* pela deleção de 6 nucleotídeos. Linhas grossas representam seqüências de cópia única A, B, C e D. TEs estão ilustrados como retângulos coloridos, sendo que as pontas agudas de cor preta estão representando as TIRs. Inserções e deleções estão delimitadas por linhas verdes e rosas, respectivamente e são identificadas com um i ou d seguido de um número. As



Figura 6: Variantes de *Galileo* caracterizadas em *Drosophila buzzatii*. *Galileo*, *Kepler* e *Newton* foram posteriormente classificados com diferentes subfamílias para *D. buzzatii* por Delprat *et al.* (2009) (Modificada de Casals *et al.*, 2005).

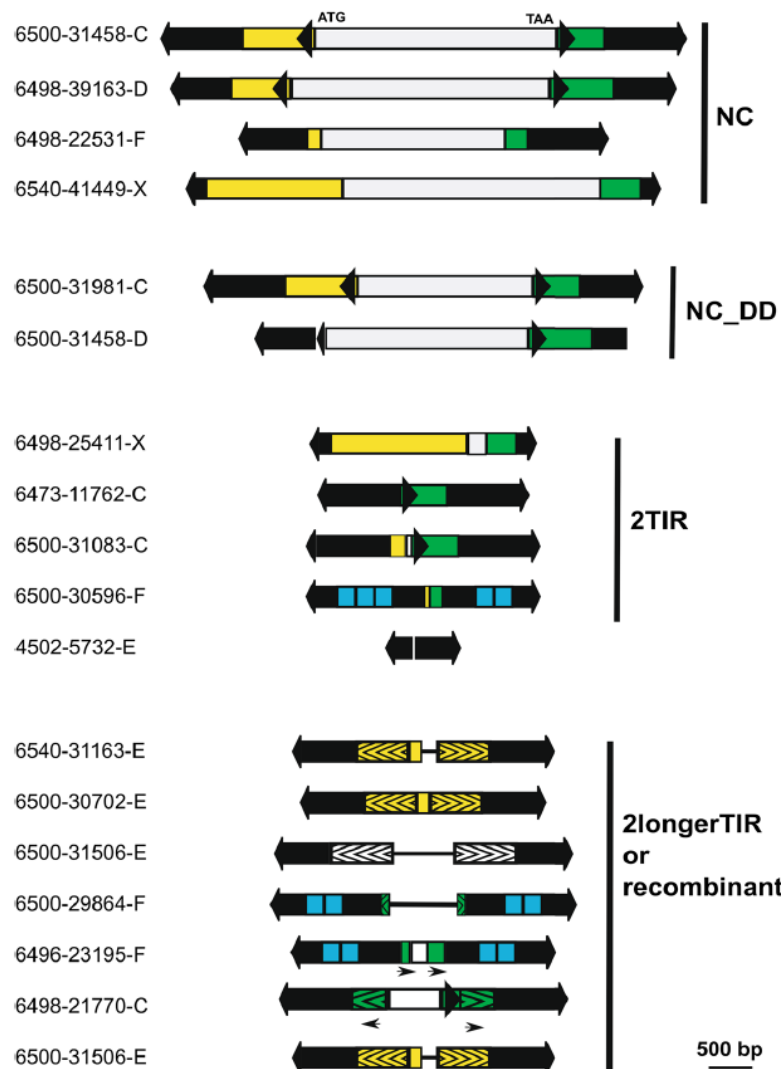


Figura 7: Estruturas representativas de cópias *Galileo* encontradas no genoma de *Drosophila mojavensis*. As setas pretas são as *TIRs*; a região central cinza corresponde à inferida transposase; região amarela é a F1 (sequência de espaçamento entre a *TIR* 5' e o segmento da transposase); a região verde é F2 (sequência de espaçamento entre o segmento da transposase e a *TIR* 3'). Os quadrados azuis são repetições em tandem encontradas no grupo F. A região com padrão entre parênteses (>>>) é a região adicional recrutada nas cópias com *TIRs* estendidas. As setas pretas são repetições internas curtas e invertidas encontradas nos grupos C e D. Cópias NC são cópias quase completas (*nearly-complete*), NC\_DD são derivadas das cópias quase completas (Modificada de Marzo *et al.*, 2013).

## Modelos polimórficos – grupos *willistoni* e *cardini* de *Drosophila*

Numerosas são as inversões cromossômicas descritas em *Drosophila* e outros dípteros. Contudo os mecanismos que originam essas inversões ainda são poucos conhecidos. Um crescente número de estudos aborda elementos transponíveis no gênero *Drosophila*, por meio da investigação e caracterização de genomas ou por análises experimentais de populações naturais. Estudos dessa natureza buscam o esclarecimento de mecanismos e consequências pertinentes à evolução desses segmentos móveis de DNA e das espécies portadoras.

No gênero *Drosophila*, o grupo *willistoni* pertence ao subgênero *Sophophora*, o qual foi estabelecido por Sturtevant (1939). Inclui 23 espécies, agrupadas em três subgrupos: *alagitans* (5 espécies), *bocainensis* (12 espécies), *willistoni* (6 espécies) (Bächli 2013). O subgrupo *willistoni* é composto pela *D. willistoni* Sturtevant (1916) e as suas cinco espécies crípticas: *D. paulistorum* Dobzhansky e Pavan (Burla *et al.*, 1949), *D. pavlovskiana* Kastritsis e Dobzhansky (1967), *D. tropicalis* Burla e da Cunha (Burla *et al.*, 1949), *D. insularis* Dobzhansky (Dobzhansky *et al.*, 1957) e *D. equinoxialis* Dobzhansky (1946). É essencialmente Neotropical, incluindo espécies com grande distribuição geográfica e outras extremamente restritas ou endêmicas (Figuras 9 e 10). Com relação ao subgrupo *bocainensis*, podemos destacar *D. nebulosa* Sturtevant (1916), amplamente distribuída desde os Estados Unidos, Índias Ocidentais, México, até o Brasil e Peru; *D. fumipennis* Duda (1925), com registros desde El Salvador até o Brasil; *D. capricorni* Dobzhansky e Pavan (1943), que se distribui do Panamá até o sul do Brasil; e *D. sucinea* Patterson e Mainland (1944), com ocorrência do México à Colômbia (Val *et al.*, 1981).

Com base em dados moleculares do gene nuclear álcool-desidrogenase (*Adh*), a separação dos subgrupos *willistoni* e *bocainensis* de *Drosophila* teria ocorrido há 15 mya, enquanto o subgrupo *willistoni* teria iniciado sua diversificação há 5,7 mya (Robe *et al.*, 2010a). O subgrupo *willistoni* representado por espécies crípticas constitui um complexo com diferentes níveis taxonômicos, uma vez que é composto por espécies, subespécies,

semiespécies, contendo inclusive espécies incipientes em *statu nascendi* como denominadas por Dobzhansky e Spassky (1959). Estudos com marcadores moleculares e citogenéticos corroboraram que as espécies crípticas ramificaram-se a partir de *D. insularis*, seguindo sucessivamente *D. tropicalis*, *D. willistoni*, *D. equinoxialis*, *D. paulistorum* e *D. pavlovskiana* (Gleason *et al.*,1998; Tarrío *et al.*,2000; Rohde *et al.*, 2006; Robe *et al.*, 2010a).

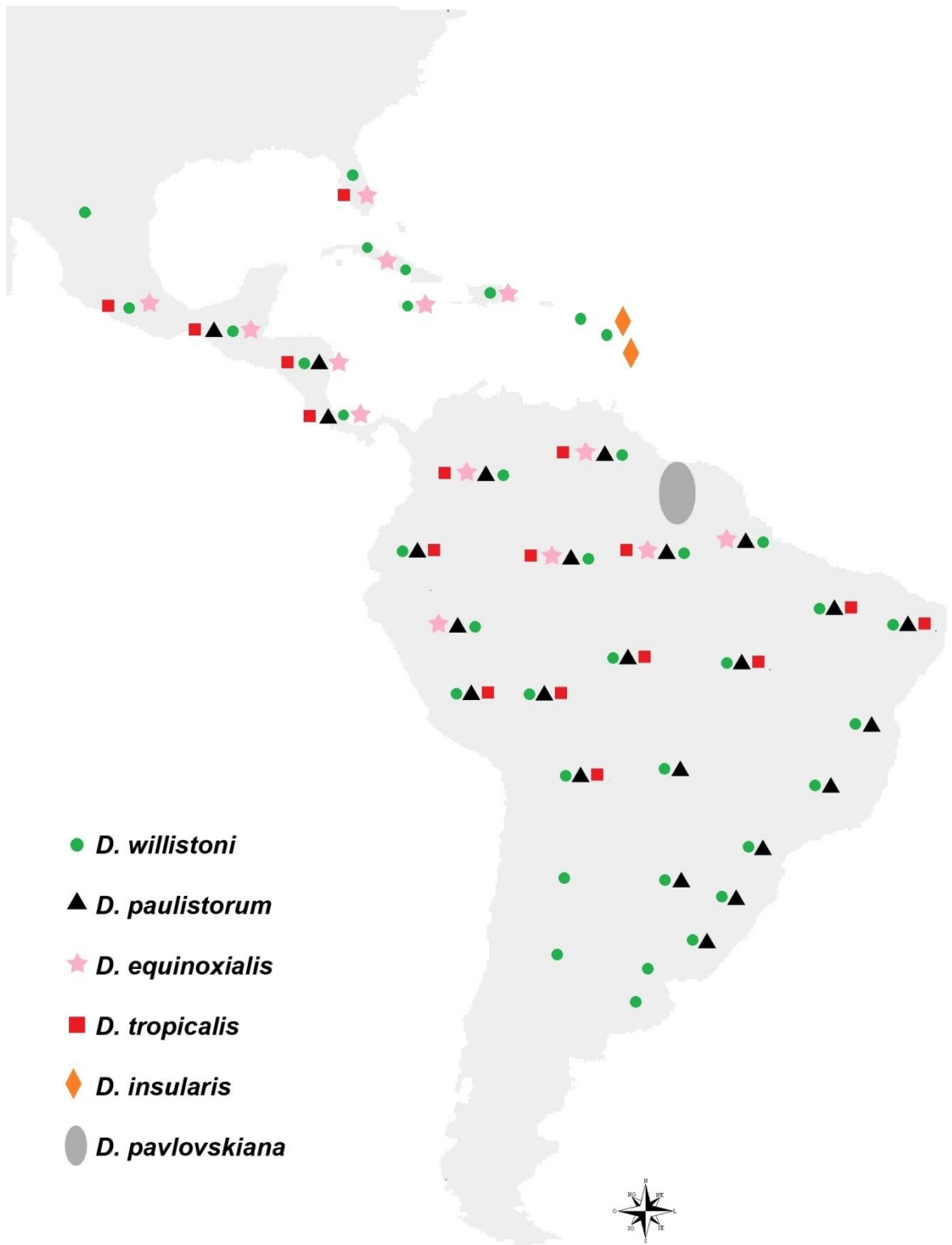


Figura 9. Mapa com o registro da distribuição geográfica do subgrupo *willistoni* de *Drosophila* (Modificada de Robe *et al.*, 2010a).





Figura 10. Mapa com o registro da distribuição geográfica das semiespécies de *Drosophila paulistorum* (Modificada de Robe *et al.*, 2010a).

No que diz respeito à evolução das semiespécies de *D. paulistorum*, a radiação seria mais recente e teria ocorrido no Pleistoceno, em torno de 0,9 mya. A semiespécie mais primitiva, Amazônica, aparece como irmã do clado que se ramifica em duas linhagens independentes: o primeiro agrupa Andino-Brasileira e Orinocana e o segundo, as semiespécies Interior e Centro-Americana, sendo que Transicional se posiciona como táxon-irmão deste último (Robe *et al.*, 2010a). A inferência de Amazônica ser a semiespécie mais primitiva reforça as conclusões de Spassky *et al.* (1971) e de Gleason *et al.* (1998). Os dados citogenéticos de Rohde *et al.* (2006) também apoiam os dois cladogramas independentes formados por Andino-Brasileira/Orinocana e Interior/Centro-Americana. Embora o estudo evolutivo inclua apenas quatro das seis semiespécies, a investigação dos padrões de bandamento encontrados no cromossomo politênico IIR apresenta Andino-Brasileira e Orinocana como homosequências entre si (IIRdg), ao mesmo tempo em que relata o compartilhamento da inversão IIRh encontrada em estado polimórfico na semiespécie Centro-Americana e fixada na semiespécie Interior. Spassky *et al.* (1971) e Ayala *et al.* (1974) além de apresentarem a semiespécie Transicional relacionada à Centro-América, sugerem também a afinidade filogenética entre Transicional e Andino-Brasileira, sendo que ambas as relações estão de acordo com a obtenção de híbridos viáveis e férteis observados (Dobzhansky *et al.*, 1964; Ehrman e Powell, 1982).

Poucos são os trabalhos realizados com as espécies dos subgrupos *alagitans* e *bocainensis* do grupo *willistoni* de *Drosophila*, bem como os estudos que elucidem as relações evolutivas das mesmas. Segundo Tarrío *et al.* (2000) o subgrupo *bocainensis*, representado pelas espécies *D. capricorni*, *D. sucinea* e *D. nebulosa*, apresenta uma composição parafilética em relação ao subgrupo *willistoni*, sendo *D. nebulosa* mais intimamente relacionada às espécies crípticas quando comparada aos demais membros de seu subgrupo. Entretanto, no trabalho de O'Grady e Kidwell (2002), o subgrupo *bocainensis* (com a adição de *D. fumipennis*, quando comparado ao trabalho anterior) é polifileticamente definido, uma vez que *D. nebulosa* é agrupada isoladamente com espécies do grupo *saltans*. O clado *D. capricorni/D. sucinea* constitui o

grupo-irmão do subgrupo *willistoni* e *D. fumipennis* aparece como espécie-irmã do clado formado pelos grupos *saltans/willistoni*. O posicionamento relativo de *D. nebulosa* e *D. willistoni* na análise de Powell *et al.* (2003), apresentou-se variável com relação aos diferentes métodos de reconstrução filogenética, as quais ou são colocadas no mesmo clado ou são colocadas dentro do agrupamento *D. capricorni/D. sucinea*, quer com *D. willistoni* ou *D. nebulosa* sendo o táxon basal. Contudo, o maior corpo de evidências leva à proposição de que *D. nebulosa* é a espécie-irmã do subgrupo *willistoni*.

O grupo *cardini*, denominado por Sturtervant (1942), pertence ao subgênero *Drosophila*. Consiste em 16 espécies neotropicais divididas em dois subgrupos e com a peculiaridade do grande polimorfismo para a pigmentação do corpo (Heed e Russell, 1971; Cordeiro *et al.*, 2009). O subgrupo *dunni* é formado por 7 espécies (*D. antillea* Heed 1962, *D. arawakana* Heed 1962, *D. belladunni* Heed e Krishnamurthy 1959, *D. caribiana* Heed 1962, *D. dunni* Townsend e Wheeler 1959, *D. nigrodunni* Heed e Wheeler 1957, e *D. simillis* Williston 1896), as quais são restritas às ilhas do Caribe. O subgrupo *cardini* é composto por 9 espécies (*D. acutilabella* Stalker 1953, *D. bedichecki* Heed e Wheeler 1971, *D. cardini* Sturtervant 1916, *D. cardinoides* Dobzhansky e Pavan 1943, *D. neocardini* Streisinger 1946, *D. neomorpha* Heed e Wheeler 1957, *D. parthenogenetica* Stalker 1953, *D. polymorpha* Dobzhansky e Pavan 1943 e *D. procardinoides* Frydenburg 1956, distribuídas geograficamente desde o sul dos Estados Unidos ao sul do Brasil (Bächli, 2013). Considerando-se outros drosofilídeos nativos da região Neotropical, as espécies do subgrupo *cardini* são as segundas em abundância, sendo que as espécies do grupo *willistoni* são as mais representativas nas assembleias de drosofilídeos (Gottschalk *et al.*, 2007; Chaves e Tidon, 2005).

Relações filogenéticas têm sido propostas para o grupo *cardini* de *Drosophila*, e segundo Robe *et al.* (2010b) ele teria divergido há 11 mya. Diferentes análises e marcadores foram utilizados com intuito de esclarecer as relações evolutivas dessas espécies, como morfologia da genitália masculina e padrões de isolamento reprodutivo (Futch, 1962; Heed, 1962), inversões

cromossômicas (Heed e Russell 1971), padrão eletroforético de isoenzimas (Napp e Cordeiro 1981) e dados moleculares de genes mitocondriais e nucleares (Brisson *et al.*, 2006; Cenzi de Ré *et al.*, 2010). O panorama comum apresenta *D. polymorpha* e *D. neomorpha* como espécies irmãs e *D. cardinoides*, *D. parthenogenetica*, *D. procardinoides*, assim como o subgrupo *dunni* de espécies insulares, formam grupos monofiléticos independentes.

Quando são comparadas as relações evolutivas obtidas com marcadores mitocôndriais e nucleares, incongruências são observadas, especialmente no que diz respeito à posição de *D. neocardini* (Robe *et al.*, 2005) e a monofilia do subgrupo *cardini*, o qual segundo Brisson *et al.* (2006) seria parafilético e teria divergido há 6.6 mya. Apesar de *D. neocardini* estar relacionada com *D. polymorpha* e *D. neomorpha* segundo Heed (1962) ou próxima do subgrupo *dunni* (Heed e Russell, 1971), de acordo com Brisson *et al.* (2006) *D. neocardini* é mais intimamente relacionada à *D. cardinoides* e *D. parthenogenetica*. Considerando-se os padrões de cruzamentos interespecíficos descritos por Heed e Russell (1971) e os últimos achados de Cenzi de Ré *et al.* (2010), interpreta-se o viés, observado nas estimativas das relações entre as espécies mencionadas, como *incomplete lineage sorting* ou como resultado de introgressão. Além disso, a análise filogenética e de estrutura populacional do grupo *cardini* realizadas pelos autores Heed e Russell (1971), mostrou que dentro do subgrupo *cardini*, composto pelas espécies continentais, havia uma maior divergência genética, quando comparada à divergência encontrada entre espécies de ilhas, do subgrupo *dunni*.

Com exceção de *D. procardinoides*, todas as demais espécies são polimórficas para a pigmentação abdominal (Heed e Krishnamurth, 1959). O padrão monomórfico da pigmentação de *D. procardinoides* deve-se à distribuição geográfica restrita a região de altitude entre a Bolívia, Brasil e Peru (Brisson *et al.*, 2006). Como resultado da seleção atuando nos diferentes *habitats* durante o processo de especiação, o subgrupo *dunni* apresenta um padrão clinal para a pigmentação do corpo, onde as espécies com abdômen mais escuro ocupam ilhas ao sul do Caribe e as espécies com pigmentação

mais claras ocupam as ilhas do norte (Heed e Krisshnamurth, 1959; Hollocher, 1996; Hollocher *et al.*, 2000a,b; Wilder *et al.*, 2004). *D. polymorpha* além de possuir a distribuição clinal para a pigmentação do abdômen (com indivíduos mais escuros ao norte de sua distribuição) (Heed, 1963), é a espécie que possui o maior polimorfismo intraespecífico para este fenótipo, o qual estaria sob a influência de três loci gênicos (Brisson *et al.*, 2005). Segundo Machado *et al.* (2001) esta espécie ainda apresenta um padrão fenotípico sazonal, pois os indivíduos escuros aumentam em número durante períodos de temperaturas mais baixas ( $\approx 13$  °C), mas a explicação fisiológica ainda não foi totalmente compreendida e elucidada.

O grupo *cardini* e o grupo *willistoni* de *Drosophila* ainda que pertençam a subgêneros diferentes, apresentam espécies simpátricas e sincrônicas, que compartilham os mesmos substratos, ao mesmo tempo e ambos possuem algumas espécies bastante restritas (De Toni *et al.*, 2007; Gottschalk *et al.*, 2008; Schmitz *et al.*, 2007). O polimorfismo do grupo *cardini* não se restringe apenas à pigmentação abdominal, pois suas espécies são também polimórficas para inversões cromossômicas (Da Cunha *et al.*, 1953; Heed e Russel, 1971; Rohde e Valente, 1996; De Toni *et al.*, 2001, 2006; Cordeiro *et al.*, 2014). Da mesma forma, *D. willistoni* e suas crípticas são extremamente polimórficas cromossomicamente (Da Cunha *et al.* 1950, 1959; Da Cunha e Dobzhansky 1954; Valente e Araújo 1985, 1986; Valente *et al.* 1993, 2001, 2003; Rohde 2000; Rohde *et al.*, 2005; Bhutkar *et al.*, 2008). Segundo Da Cunha *et al.* (1950, 1959), a quantidade de polimorfismo encontrada em uma espécie reflete a variedade de nichos ecológicos ocupados por suas populações, pois representaria a adaptação dos indivíduos da espécie às diferentes condições ecológicas. A natureza polimórfica de espécies neotropicais de *Drosophila*, portanto, foi forte motivação para investigarmos *Galileo* em seus genomas, como ponto de partida para testarmos seu possível papel na geração de variabilidade genética no neotrópico. Os grupos *willistoni* e *cardini* apresentam-se, pois, como promissores modelos, repletos de informações biológicas para os estudos dos vários níveis de divergência evolutiva.

## OBJETIVOS

### Justificativa e Objetivo geral

O fato de que *Galileo* gerou três inversões independentes em *D. buzzatii* sugere que, ainda que não esteja ativo, não se trata de um substrato passivo. A existência de espécies de dois subgêneros de *Drosophila*, cromossomicamente polimórficas, ecologicamente similares e versáteis, instiga a investigação da presença e história evolutiva deste transposon. Dessa forma, a partir da caracterização de *Galileo* no genoma de *Drosophila willistoni* e da análise deste elemento em diferentes populações naturais de espécies pertencentes ao grupo *willistoni* e da expansão desta busca para representantes do subgrupo *cardini* de *Drosophila*, o presente trabalho pretende traçar um panorama evolutivo do referido transposon.

O estabelecimento da colaboração formalizada com o Dr. Alfredo Ruiz da *Universidad Autònoma de Barcelona* (Espanha), referência na descoberta e descrição de *Galileo* e com o Dr. Victor Hugo Valiati, da UNISINOS, devido ao seu amplo conhecimento das espécies do subgrupo *willistoni*, para um estudo conjunto com *Galileo* em algumas espécies neotropicais de *Drosophila*, favoreceu ainda mais a execução desta tese.

### Objetivos específicos

- Caracterizar molecularmente e *in silico* as cópias de *Galileo* no genoma de *D. willistoni*;
- Identificar a presença de sequências (resíduos) que caracterizariam uma transposase funcional no genoma de *D. willistoni*;
- Avaliar em nível interespecífico a presença de variantes de *Galileo* nas espécies neotropicais dos grupos *willistoni* e *cardini* de *Drosophila*.

## CAPÍTULO II

---

Trabalho publicado na revista *BMC Genomics* 2014, 15:792

### **Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome**

Juliana W Gonçalves<sup>1</sup>, Victor Hugo Valiati<sup>2\*</sup>, Alejandra Delprat<sup>3</sup>,  
Vera L S Valente<sup>1\*</sup> and Alfredo Ruiz<sup>3\*</sup>

<sup>1</sup> Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), CP 15053, Porto Alegre, Rio Grande do Sul, 91501-970, Brazil.

<sup>2</sup> Programa de Pós-Graduação em Biologia: Diversidade e Manejo de Vida Silvestre, Universidade do Vale do Rio dos Sinos (UNISINOS), CP 275, São Leopoldo, Rio Grande do Sul, 93022-000, Brazil.

<sup>3</sup> Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra (Barcelona), Spain.

Correspondence: vera.valente@pq.cnpq.br

\* Equal contributors

RESEARCH ARTICLE

Open Access

# Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome

Juliana W Gonçalves<sup>1</sup>, Victor Hugo Valiati<sup>2†</sup>, Alejandra Delprat<sup>3</sup>, Vera L S Valente<sup>1\*†</sup> and Alfredo Ruiz<sup>3†</sup>

## Abstract

**Background:** *Galileo* is one of three members of the *P* superfamily of DNA transposons. It was originally discovered in *Drosophila buzzatii*, in which three segregating chromosomal inversions were shown to have been generated by ectopic recombination between *Galileo* copies. Subsequently, *Galileo* was identified in six of 12 sequenced *Drosophila* genomes, indicating its widespread distribution within this genus. *Galileo* is strikingly abundant in *Drosophila willistoni*, a neotropical species that is highly polymorphic for chromosomal inversions, suggesting a role for this transposon in the evolution of its genome.

**Results:** We carried out a detailed characterization of all *Galileo* copies present in the *D. willistoni* genome. A total of 191 copies, including 133 with two terminal inverted repeats (TIRs), were classified according to structure in six groups. The TIRs exhibited remarkable variation in their length and structure compared to the most complete copy. Three copies showed extended TIRs due to internal tandem repeats, the insertion of other transposable elements (TEs), or the incorporation of non-TIR sequences into the TIRs. Phylogenetic analyses of the transposase (TPase)-encoding and TIR segments yielded two divergent clades, which we termed *Galileo* subfamilies V and W. Target-site duplications (TSDs) in *D. willistoni Galileo* copies were 7- or 8-bp in length, with the consensus sequence GTATTAC. Analysis of the region around the TSDs revealed a target site motif (TSM) with a 15-bp palindrome that may give rise to a stem-loop secondary structure.

**Conclusions:** There is a remarkable abundance and diversity of *Galileo* copies in the *D. willistoni* genome, although no functional copies were found. The TIRs in particular have a dynamic structure and extend in different ways, but their ends (required for transposition) are more conserved than the rest of the element. The *D. willistoni* genome harbors two *Galileo* subfamilies (V and W) that diverged ~9 million years ago and may have descended from an ancestral element in the genome. *Galileo* shows a significant insertion preference for a 15-bp palindromic TSM.

**Keywords:** Transposable element, *D. willistoni*, Terminal inverted repeats, *P* superfamily, Target site duplications

## Background

Transposable elements (TEs) are part of the middle repetitive portion of DNA that is able to move and replicate within the genome. They comprise a considerable fraction of many eukaryotic genomes and their sequences exhibit broad structural diversity. The wide range of transposition strategies adopted by TEs involve

either RNA (class 1 or retrotransposons) or DNA (class 2 or DNA transposons) intermediates. Selfish and thus in many respects indistinguishable in their behavior from parasites, these mobile genetic units increase in number within the genome because their rates of transposition are higher than those of spontaneous deletion. This evolutionary success of TEs is a major force shaping the genes and genomes of almost all organisms [1,2].

The movement and accumulation of TEs serves as a rich source of genetic material, with a strong impact on the evolutionary reorganization of the genomes of their bearers. However, it is now clear that inactive TEs also play a significant role in macroevolution, because the

\* Correspondence: vera.valente@pq.cnpq.br

†Equal contributors

<sup>1</sup>Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), CP 15053, Porto Alegre, Rio Grande do Sul 91501-970, Brazil  
Full list of author information is available at the end of the article



most influential contributions can arise and persist long after transposition activity has ceased, such that they are manifested as TE by-products. The selfish and parasitic characteristics of TEs ensure their long-standing residence within the host genome and imply their intimate co-evolutionary relationship with it [3].

The specific features of DNA transposons compared to other TEs enhance their influence in shaping eukaryotic genomes, including the capacity to excise imprecisely, jump locally, cause multiple double-strand breaks, and undergo alternative transposition [2]. The transposon *Galileo* was originally discovered in *Drosophila buzzatii*, in which three segregating chromosomal inversions were shown to have been generated by ectopic recombination between *Galileo* copies [4-6]. Although *Galileo* has long terminal inverted repeats (TIRs) similar to those of *Fold-back-like* elements, it is classified as a member of the *P superfamily* of DNA transposons (class II, subclass 1, TIR elements order) based on the sequence of its putative transposase (TPase). Subsequently, *Galileo* was identified in six of the 12 sequenced *Drosophila* genomes of the two subgenera of *Sophophora* and *Drosophila*, indicating its widespread distribution within this genus. Although potentially active *Galileo* copies have not been found, non-autonomous copies are abundant in all species investigated [7]. In addition, two or more *Galileo* subfamilies coexisting within the same genome have been found in several cases: three subfamilies are present in *D. buzzatii* (G, K, and N for *Galileo*, *Kepler* and *Newton*), two in *D. virilis* (A and B), and five in *D. mojavensis* (C, D, E, F, and X) [6-8].

According to *in silico* predictions, *Galileo* is strikingly abundant in *Drosophila willistoni* [7], the most widespread neotropical species of the genus *Drosophila* [9,10], with an extensive gene arrangement polymorphism on all chromosomes [11-20]. This high intraspecific polymorphism for chromosomal inversions and *Galileo* abundance suggest a role for *Galileo* in the generation of inversions in *D. willistoni* and related species. We have an ongoing project to test this hypothesis by identifying and isolating the breakpoints of *D. willistoni* natural polymorphic inversions. As a first step in this in

this project, we carried out an exhaustive search for and characterization of the *Galileo* copies present in the *D. willistoni* genome. A careful and detailed annotation of 191 *Galileo* sequences revealed that they vary considerably in length and structure, ranging from nearly-complete to containing only one TIR. Two *Galileo* subfamilies with a substantial nucleotide divergence were found by phylogenetic analysis of TPase-encoding and TIR segments. In addition, by analyzing the preferred target sequence of *Galileo* in *D. willistoni*, we identified a palindromic target site motif (TSM).

## Results

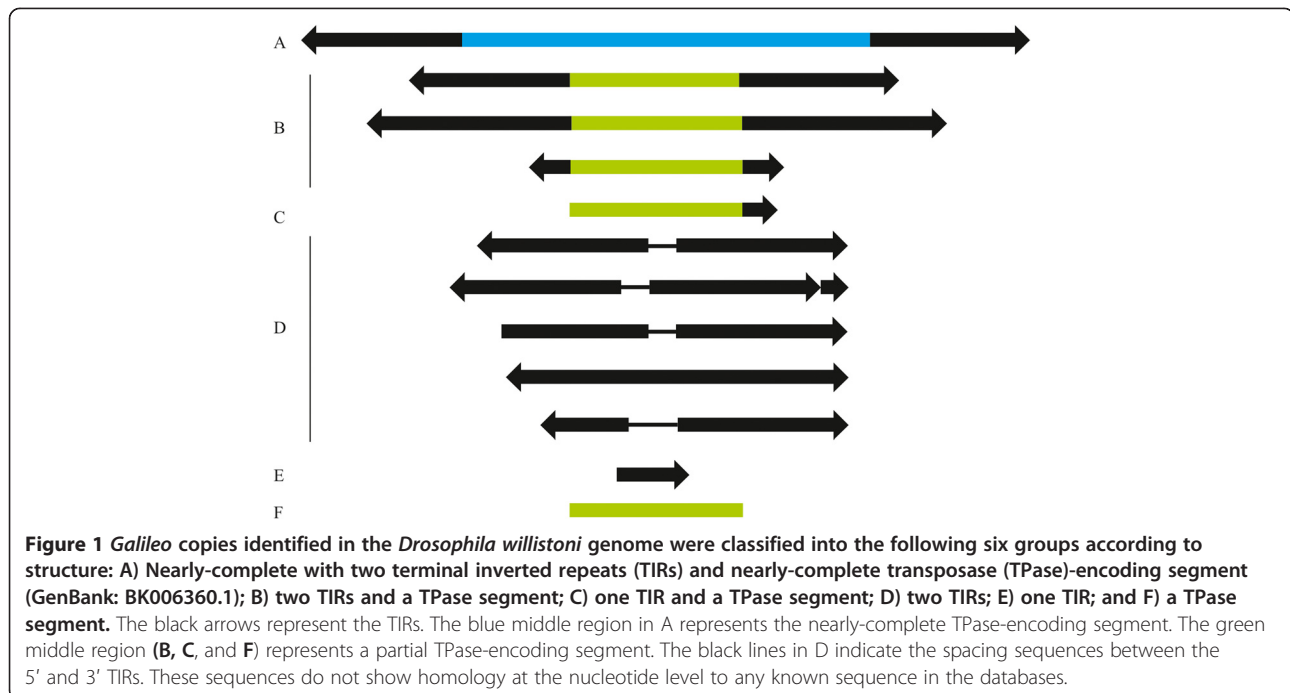
### Characterization of *Galileo* copies in the *D. willistoni* genome

We characterized 191 *Galileo* copies in the *D. willistoni* genome (details are given in Additional file 1), classifying them into six groups according to their structure (Table 1 and Figure 1): (A) nearly-complete; (B) two TIRs and a partial TPase-encoding segment; (C) one TIR and a partial TPase-encoding segment; (D) two TIRs; (E) one TIR only; and (F) a TPase-encoding segment. Only one nearly-complete copy, containing two TIRs and a nearly-complete TPase-encoding segment, was found. This copy, identified in previous work (GenBank: BK006360.1) [7], is 4386-bp long and harbors a long ORF (coordinates 984–3698) encoding a 905-amino-acid TPase. The only mismatch is in the start codon, with ACG = Thr instead of the canonical ATG = Met; thus, this copy cannot be functional. Nonetheless, this putative TPase is similar in size and composition to other *Galileo* elements [7]. Protein functional analysis, performed using InterProScan 4 [21], revealed the presence of a THAP domain (PF05485) in residues 14–93 (2E–12) and a THAP-domain containing a protein 9 domain (PTHR10725) in residues 251–884 (1E–61). THAP is a DNA-binding domain present in TPases of the *P superfamily*; this domain includes a Zn-coordinating C2CH signature and four other invariant residues (P, W, F, and P) that are also required for DNA binding [8]. These eight residues are fully conserved at positions C16, C21, P40, W49, C67, H70, F71, and P87 of the putative *Galileo* TPase. The second conserved domain included the triad

**Table 1** *Galileo* copies characterized in the *Drosophila willistoni* genome

Group	Number of copies	Copies with inserted TE	Copies with flanking TE	Copies with inserted and flanking TE
A	1	0	1	0
B	7	5	0	0
C	26	11	0	3
D	124	11	19	2
E	2	0	1	1
F	31	6	7	3
Total	191	33	28	9

TE, transposable element.



DDE and the motif D(2)H, which is present in the catalytic domain of cut-and-paste TPases of the *P superfamily* [22] at positions D327, D415, E642, and D449(2)H452.

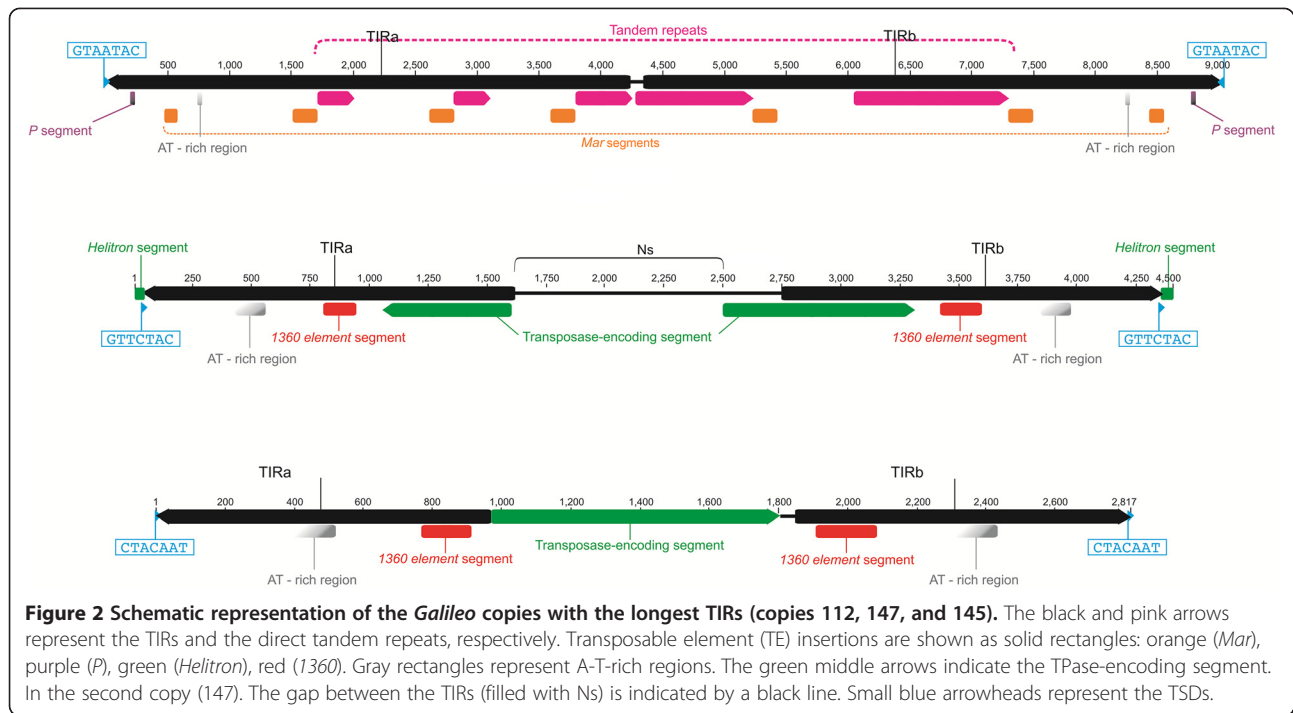
Most (~89%) of the copies with preserved terminal sequences are flanked by identical target-site duplications (TSDs). Approximately 17.3% (33 copies) contain other elements inserted within them, 14.7% (28 copies) have other elements inserted at the *Galileo* termini, and 4.7% (9 copies) have both inserted and flanking elements (Table 1). In one case, we identified a full-length *P element* (99% identity with the *D. melanogaster P element*) [23], possibly with imperfect TSDs (CGCTAGCC/GGCTA GCG) inserted within the *Galileo* copy, that contained only fragments of TIRs and identical TSDs. Of the copies with a TPase-encoding segment only (group F), 58% (18 copies) are located at the ends of short scaffolds ( $\leq 5,598$ -bp); thus, they may be incomplete, either because the rest of the sequence is present somewhere else or it is missing. None of the copies in groups B–F have an intact ORF encoding a putatively functional TPase (i.e., all characterized copies are non-autonomous; with variable portions of the TPase-coding region).

#### TIR structural variation

*Galileo* copies in the *D. willistoni* genome exhibit remarkable structural variation. In particular, the TIRs vary considerably in length and structure compared to the TIRs of the nearly-complete copy (Figure 1), which are 765/757-bp long and have 99% identity (omitting indels). The 3' TIR has a 69-bp overlap with the TPase-coding segment (Figure 1). Thus, the final piece of this segment

is repeated (in reverse orientation) at the 5' TIR. This is a unique trait among the described *Galileo* transposons [7]. Also, there are two AT-rich segments, with the 136-bp segment located in the 5' TIR (coordinates 528–663) and the 137-bp segment located in the 3' TIR (coordinates 3732–3868).

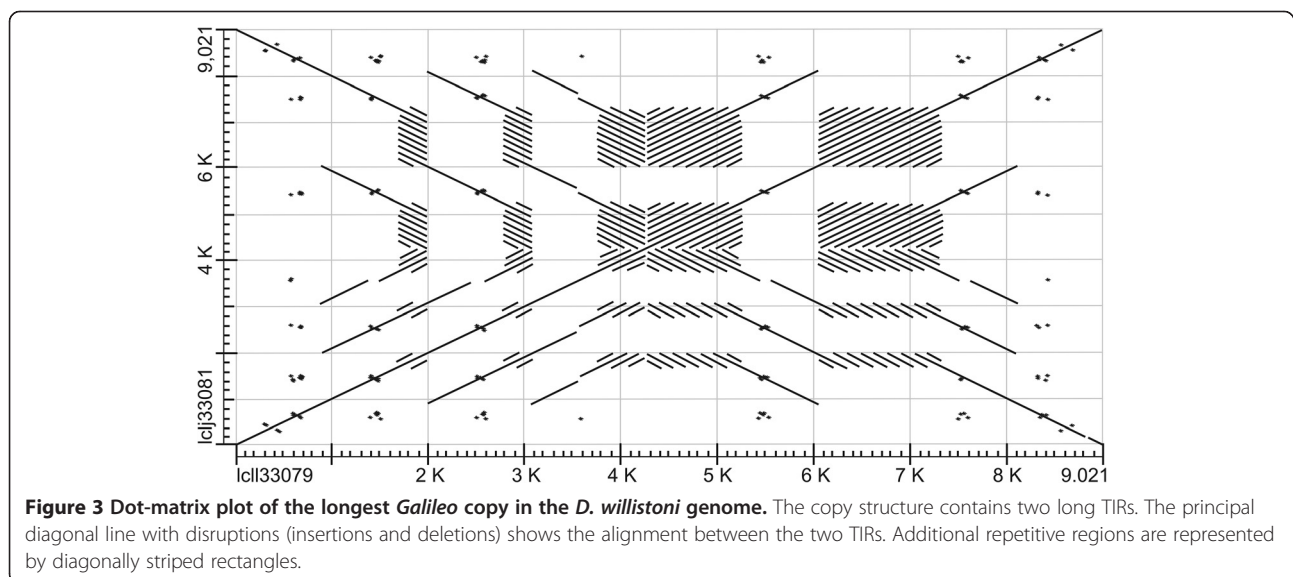
Three *Galileo* copies were found to display significantly extended TIRs and each one is flanked by identical TSDs (Figure 2). The longest copy (112) is 9021-bp long, including TIRs of 4246-bp and 4680-bp (5' and 3', respectively; see Additional file 1). This copy contains only two TIRs and it lacks a TPase-coding segment. However, the TIRs are notable for their striking length and repetitive structure (Figure 3). They contain direct tandem repeats and an insertion of another TE in addition to the AT-rich segments. This longest copy is the only one with direct tandem repeats within the TIRs. The repeats are ~140-bp long and located approximately 1710-bp and 1730-bp from the 5' and 3' ends, respectively. The 5' TIR contains three repeat regions, two that are 275-bp long (2 tandem repeats) and another that is 443-bp long (3.2 repeats). On the 3' end, we annotated two longer repeat regions, a 995-bp region (6.8 repeats) and a 1246-bp region (9 repeats) (Figures 2 and 3). The TIRs of this copy contain fragments of two additional transposons: *P element* and *Mar*. At the 5' and 3' ends, we identified one fragment (36-bp) of a *P element* (86.1% identity with the *D. bifasciata P element*, according to the database; coordinates: 2900–2935) [24]. Of the seven fragments of *Mar* that were annotated, two are 107-bp long (one at each end) and the other five are



202-bp long (98.1% and 94.5% identity with *D. willistoni* *Mar*, respectively; coordinates 491–597 and 299–500) [25].

The copy (147) with the second longest TIRs (1575-bp and 1608-bp; 97.5% identical) is 4306-bp long (Figure 2). The TIRs are composed of: 130-bp of AT-rich sequence, 140-bp and 172-bp stretches (5' and 3' TIR, respectively) similar to transposon *1360* (also known as *Hoppel* or *ProtoP* element; 85.7% identity with coordinates 4020–3869 and 84.9% identity with coordinates 3869–4079, respectively) [26,27], and at least 545-bp of the *Galileo* TPase (coordinates: 3700–3162). Thus, in this

*Galileo* copy, TPase-encoding segments are repeated, forming part of the TIRs (Figure 2). In addition, there is a 872-bp gap between the TIRs (filled with Ns) that may hide a larger TPase fragment. The third copy (145) has TIRs that are 959-bp (99.9% identical) in length, with 132-bp of AT-rich sequence and the same fragments of the *1360* transposon present in the copy previously described. These last two copies are similar in their structure and have 99.8% identity over the first 959-bp. However, they have different TSDs, indicating that they are independent insertions.



### Galileo sequence diversity in the *D. willistoni* genome

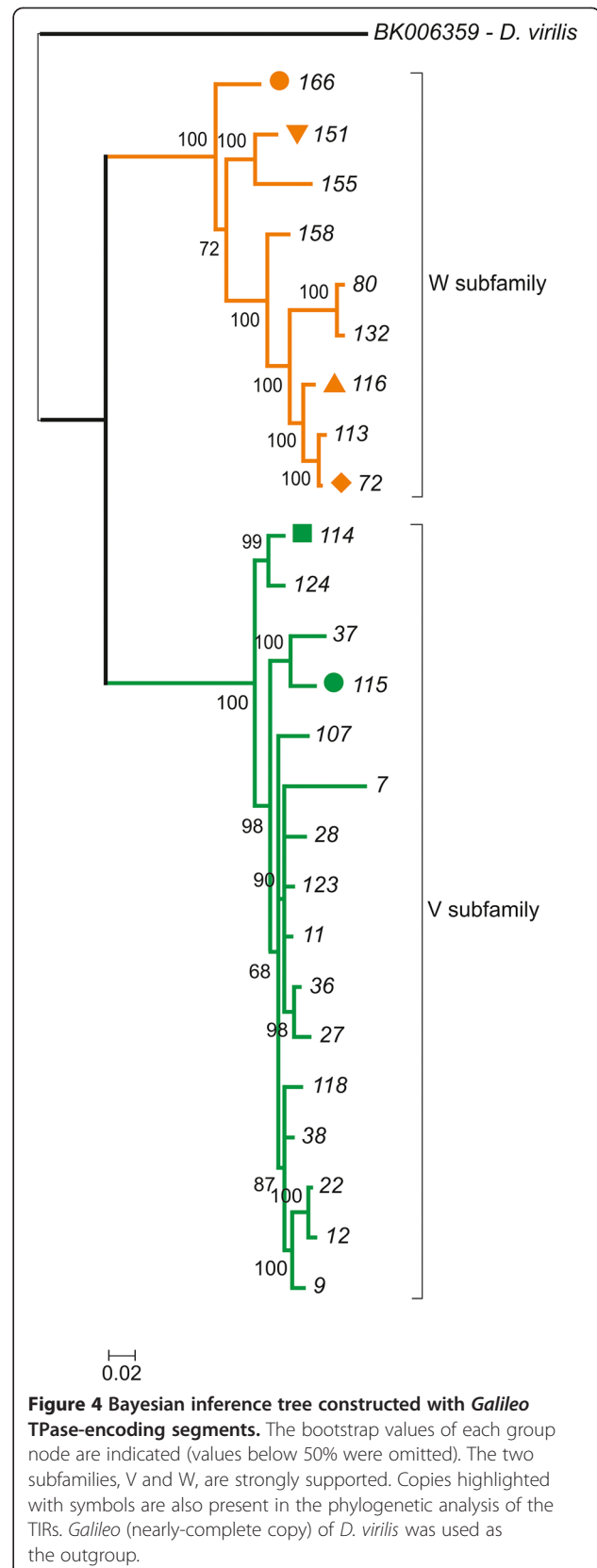
We aligned the TPase-coding segments from 26 *Galileo* copies and used a ~488-bp region (coordinates: 2699–3131) to build a phylogenetic tree using maximum-likelihood (ML) and Bayesian inference (BI) methods. The phylogenetic trees of the TPase-coding sequences created using the two methods were similar and recovered the same two clades with significant statistical support (Figure 4). The two clades showed a substantial nucleotide divergence between them (20%–50.3%) and were termed *Galileo* subfamilies V and W. The analysis placed the nearly-complete copy within the W subfamily. Assuming a *Drosophila* synonymous substitution rate of 0.016 substitutions per nucleotide/myr [28], we estimated the split between the two subfamilies to ~9 million years ago (Mya). The two subfamilies have a modest mean divergence between copies within each one (4.7 and 6.7%, respectively) compared with the mean divergence between them (24%) (Table 2).

We also aligned the TIR regions and built a phylogenetic tree with 238 homologous segments (the first 100-bp). The topologies of the TIR trees also yielded two clades, with 11%–26.6% nucleotide divergence (Figure 5). Several copies (72, 114, 115, 116, 151), in addition to the nearly-complete copy (166), contain both the initial portion of the TIR and a TPase-encoding fragment. The phylogenetic placement of these copies suggests that the two clades in the TIR phylogeny correspond to the above-defined V and W subfamilies (Figure 5 and Additional file 2). Subfamily V was represented by copies with extended TIRs (145 and 147) that have the homologous TIR region (the first 100-bp). There was no significant difference in the lengths of the TIRs among subfamilies. The mean divergence between copies of the two subfamilies was 13.6%, whereas the mean divergence within subfamilies was much smaller (1.3% and 3.1%) (Table 3). The divergence within subfamilies includes the estimates between copies and between the two TIRs within copies.

The consensus sequences for the terminal 40-bp segment in *Galileo* subfamilies V and W differed by 4 bp (10%). A comparison of the 40 terminal bp region conserved in 14 *Galileo* sequences of diverse species and subfamilies showed a total of 17 conserved nucleotides (Figure 6).

### Target site duplication and target site motif

In most *D. willistoni Galileo* copies, the TSDs were 7-bp in length, as similarly reported in *D. buzzatti* [29]. However, we identified three copies (89, 127, and 179) in which the TSDs were 8-bp long (see Additional file 1). Comparison of the 118 flanking sequences of those *Galileo* copies with the 7-bp TSD suggested that the consensus sequence of their preferential insertion site is GTATTAC (Figure 7).



**Figure 4 Bayesian inference tree constructed with *Galileo* TPase-encoding segments.** The bootstrap values of each group node are indicated (values below 50% were omitted). The two subfamilies, V and W, are strongly supported. Copies highlighted with symbols are also present in the phylogenetic analysis of the TIRs. *Galileo* (nearly-complete copy) of *D. virilis* was used as the outgroup.

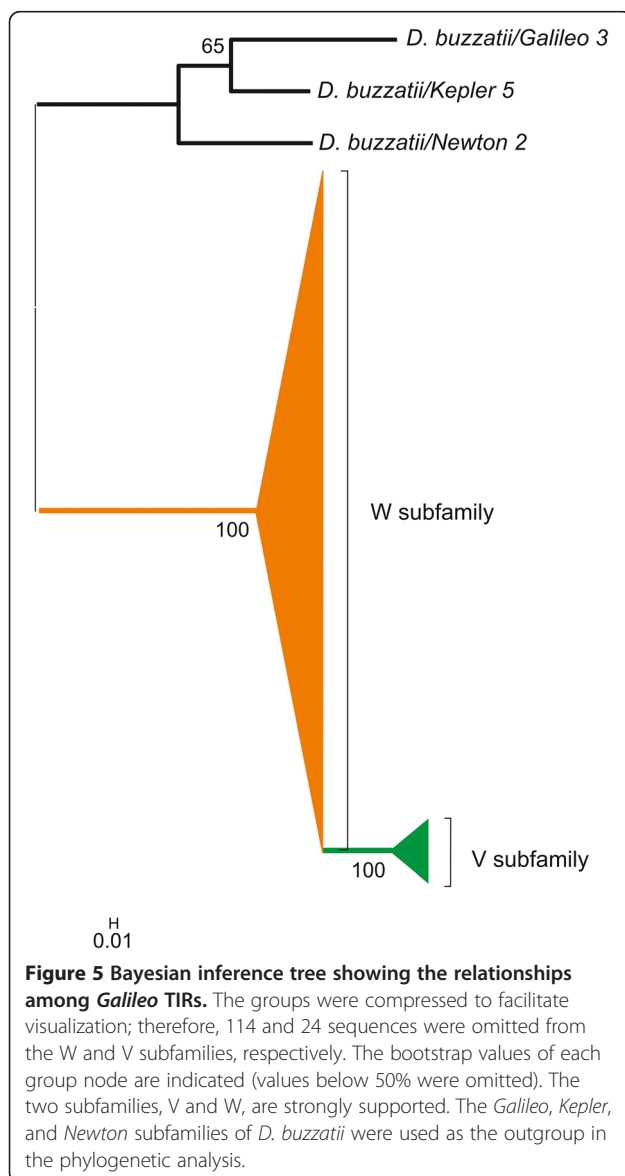


**Table 2 Nucleotide divergence estimates (%) for the Galileo subfamilies using transposase-coding sequences**

	V subfamily	W subfamily	Galileo of <i>D. virilis</i>
V subfamily	4.7 ± 0.60		
W subfamily	24.02 ± 2.19	6.66 ± 0.086	
Galileo of <i>D. virilis</i>	32.34 ± 2.52	30.08 ± 2.60	-

Galileo (nearly-complete copy) of *D. virilis* was used.

The majority-rule consensus sequence for the flanking sequences of *D. willistoni* Galileo copies suggested that the insertion sites are localized at the center of the AT-rich region. Analysis of the 93-bp surrounding the TSDs revealed a target site motif (TSM) with a 15-bp palindromic pattern composed of 7-bp duplicated upon insertion plus 4-bp on either side of Galileo (Figure 7).



Two pentanucleotides, AANGT and ACNTT, were identified on the 5' and 3' halves of the TSM [30]. This motif could adopt a stem-loop secondary structure when denatured.

## Discussion

In the present study, we used different search strategies and a detailed manual annotation to fully characterize Galileo copies in the *Drosophila willistoni* genome. In contrast to previous work [7], which reported on 28 copies, this study presents information on 191 copies. The long term goal of this project is to contrast the hypothesis that Galileo generated some of the *D. willistoni* chromosomal inversions segregating in natural populations. The detailed annotation of all Galileo copies present in the *D. willistoni* genome will greatly assist in the interpretation of the breakpoint sequences.

### Galileo structural variation

Putatively functional copies of Galileo were not found, although one nearly complete copy harbors an ORF coding for a 905-amino-acid TPase (after curating a mismatch in the start codon). Among the non-autonomous copies with TPase segments, the majority (~63.6%) were composed of TIRs and a spacing region. In addition, they exhibit a remarkable structural variation, particularly in the TIRs. Galileo, along with two other transposons, P-element and 1360, are members of the P superfamily [31]. P elements move to a new site through a non-replicative process, i.e., the cut-and-paste mechanism of transposition, in which the excised copy leaves behind a double-strand gap [32]. Because gap repair is not always efficient, whether via homologous recombination or using the sister chromatid strand as a template, defective copies are often generated due to abortion, slippage, or template switching in the course of transposition and repair [2,33]. Furthermore, because transposons are dispersed repeats in the genome, non-allelic homologous recombination or ectopic recombination events are likely, thereby increasing the probability of exchange between two copies and affecting the structure of the sequences. These molecular processes can explain the gradient of Galileo copies found in the *D. willistoni* genome, ranging from an almost-complete copy to defective copies restricted to the TIRs, with various degrees of degeneration.

Moreover, Galileo displays dynamic restructuring. A recent analysis of the *Drosophila mojavensis* genome [8] identified two patterns of extension for Galileo TIRs: (1) expansion of the direct tandem repeats and (2) recruitment of internal sequences (non-TIR segments) into the TIRs. In the *D. willistoni* genome, we identified direct tandem repeats within the TIRs, but in a single copy only (the longest one, Figure 2). We also found evidence of recruitment of non-TIR segments into the TIRs.

**Table 3 Nucleotide divergence estimates (%) for the Galileo subfamilies using the terminal inverted repeat regions**

	V subfamily	W subfamily	Subfamilies of <i>D. buzzatii</i>
V subfamily	3.08 ± 0.09		
W subfamily	13.61 ± 3.57	1.31 ± 0.024	
Subfamilies of <i>D. buzzatii</i>	52.92 ± 3.83	48.51 ± 4.91	28.41 ± 4.02

*Galileo* subfamilies K, G and N of *D. buzzatii* were used.

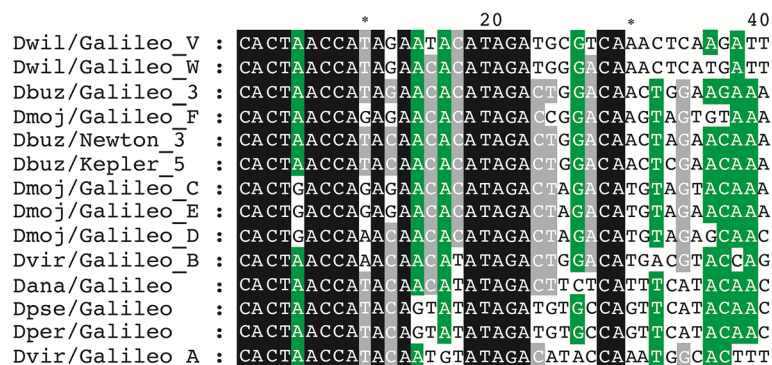
Remarkably, in the nearly-complete copy, a 69-bp final piece of the ORF is now part of the 3' TIR, and it is repeated (in reverse orientation) at the end of the 5' TIR. In other copies, the overlap between the ORF and TIRs is even greater. For instance, in copy 147, the segment of the ORF recruited to the TIRs is more than 500-bp long (Figure 2). This is, so far, a unique trait of *Galileo* transposons [7]. Finally, we found a third pattern of TIR extension: the insertion of another TE into one of the TIRs, which eventually may be transferred to the other TIR, ultimately becoming a part of both. The TE fragments are not occasional insertions in one *Galileo* TIR; rather they are part of the 5' and 3' ends in transposing copies. We detected the insertion of three elements in the longest TIRs, i.e., *P*, *Mar*, and *1360*. The first two were previously studied in *D. willistoni* [34-36], but the origin of the *1360* fragment is obscure because in *D. willistoni* transposon *1360* is missing [7].

### Two Galileo subfamilies in the *D. willistoni* genome

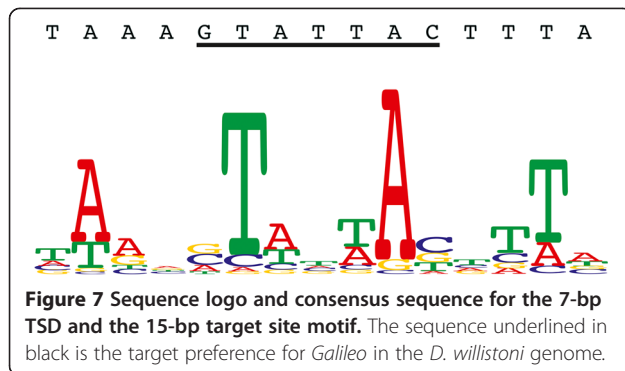
In previous work [7], a limited number of *Galileo* copies were isolated from the *D. willistoni* genome, and a subsequent phylogenetic analysis did not detect a significant structure. Here, our phylogenetic analysis, based on an increased number of copies, revealed two strongly supported clades, which we named subfamilies V and W. The two clades were evident in phylogenetic analyses carried out using either a segment of the ORF or the final 100-bp of the TIRs. Although in our study only six copies were shared between the two phylogenies, two in

subfamily V (114, 115) and four in subfamily W (72, 116, 151 and 166), the results are congruent and suggest the same grouping in the two subfamilies. The presence of *Galileo* subfamilies within the same genome seems to be the rule rather than the exception, as was previously found in *D. buzzatii* (subfamilies G, K, and N), *D. virilis* (subfamilies A and B), and *D. mojavensis* (subfamilies C, D, E, F, and X) [6-8]. Furthermore, the coexistence of different subfamilies, subgroups, or variants of TEs was reported in studies of *Bari* [37] and *Gypsy* [38] in *Drosophila*, *P* in *Anopheles gambiae* and *Drosophila* [39,40], and *mariner* in insects and humans [41,42], among others.

How have these *Galileo* copies differentiated in the genome of *D. willistoni*? Horizontal transfer (HT) and vertical diversification are the two main hypotheses that explain the coexistence of different subfamilies in the same genome [2]. HT would account for the appearance of the two subfamilies, via two independent events of *Galileo* invasion in the *D. willistoni* genome. Several mechanisms and vectors have been proposed to explain HT events. In *Drosophila* parasites and parasitoids, such as mites and wasps, intracellular symbiotic bacteria, such as *Wolbachia* and spiroplasms, are possible vectors of TEs [43]. HT can also result from an introgression, as reported in the *willistoni* subgroup [44-46], and is a potential mechanism for *P* element spreading among this subgroup [47]. Although the HT hypothesis in the case of *Galileo* has yet to be disproven, our data suggest that, based on the landscape of this transposon in *D. willistoni*, the copies instead diverged from an ancestral element in



**Figure 6 Comparison of the TIRs ends.** A consensus sequence was constructed for the V and W subfamilies of the *Galileo* TIRs in *D. willistoni*. Alignments of the 40-bp TIRs of each *Galileo* subfamily and species are shown. Identical positions (17) in all sequences are marked in black, and the 80% and 60% conserved positions in green and gray, respectively.



the genome. Although no complete copies of *Galileo* have been found, its functional differentiation would have had to be driven by specific selective pressures, resulting in the formation of two distinct *Galileo* TPases to overcome cellular repression of transposition. We identified *Galileo* copies composed of TIRs with conserved TPase site affinity in the genome; these could have served as a source for the other defective copies. Furthermore, HT and vertical diversification are not mutually exclusive; thus, successive invasions and structural variations may have occurred during the diversification of TEs. Concerning the preservation of *Galileo* TIRs, the mean divergence for these sequences was only one half (~13.6%) of that for the TPase-encoding segment (~24%). Under a neutral evolution model, the same degree of divergence would be expected; however, in the case of *Galileo*, there are more constraints in the terminal segment (100-bp) of its TIRs than in its TPase-encoding segment because the former are required for transposition.

#### **Galileo insertional preference**

DNA transposons generate, upon insertion, direct duplications of short genome sequences (TSDs). In *D. buzzatii*, a comparison of the 19 flanking sequences suggested that *Galileo* generates 7-bp TSDs with the consensus sequence GTAGTAC [48]. A larger sample (106 *Galileo* copies) in six sequenced *Drosophila* genomes (*D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. virilis*, and *D. mojavensis*) identified the consensus sequence GTANTAC [7]. We found that the *Galileo* TSDs in *D. willistoni* are typically 7-bp but, as occurs in most *P* element insertions, three *Galileo* copies had TSDs of 8-bp. Additionally, by comparing 118 *Galileo* copies flanked by identical 7-bp sequences, we were able to infer that the preferential insertion site has a consensus sequence of GTATTAC, in which the fourth position differs from that occurring in *Galileo* copies in *D. buzzatii*. These findings are in agreement with those of a study of six *Drosophila* genomes. Linheiro and Bergman [30] measured the degree of target specificity for different elements in *D. melanogaster*. They found that *1360* and *P* elements seem to have a

relatively low degree of target specificity. *Galileo* seems to have a higher target specificity than either *1360* or the *P* element. Accordingly, it can be detected with a lower number of insertions [7,29,48].

A previous study identified a 14-bp palindromic pattern centered on the 8-bp TSD generated by *P* element insertion [49]. Sequence motifs at TE target sites are always palindromes that extend beyond the TSD [30]. Here, by analyzing the region around the TSDs, a 15-bp palindrome was identified; in addition, the *Galileo* TSM also had a general tendency to be AT-rich. Although the tendency in the TSMs of both *P* element and *1360* is to have an ANAGT motif in the 5' half and an ACTNT motif in the 3' half, the *Galileo* TSM while palindromic, is not identical in sequence (AANGT and ACNTT, respectively).

#### **Conclusions**

Our detailed analysis of 191 *Galileo* copies revealed an enormous variety in their size and structure. In some copies, there were different forms of TIR extension, including internal duplications, recruitment of the final piece of the TPase-encoding ORF into the TIRs and secondary TE insertions in one TIR that subsequently become part of both TIRs. Two *Galileo* subfamilies (termed V and W) coexist in the *D. willistoni* genome. They are evident in the phylogenetic trees of both the TPase-encoding and the TIR segments. However, phylogenetic analysis showed that the divergence between and within subfamilies is smaller in the TIR segment than in the TPase-encoding segment, presumably because the former is required for transposition. *Galileo* shows a stronger target preference than *1360* or *P*-element, the other two members of the *P* superfamily.

#### **Methods**

##### **Bioinformatic searches**

The *D. willistoni* genome sequence was used for *in silico* analyses. Candidate *Galileo* elements were identified by querying the nearly-complete copy of *Galileo* detected in the *D. willistoni* genome in previous work [7], terminal inverted repeats (TIRs), and segments of the transposase (TPase)-encoding ORF isolated by experimental searches of *Galileo* in *D. willistoni* (Gonçalves et al. in preparation). Blast searches of the *Drosophila willistoni* genome were performed using FlyBase [50], with default parameters and without a low complexity filter, to identify copies with simple and complex repeats. The applied threshold of scores had an e-value of  $<10^{-4}$ . To accept a search hit, we compared previously characterized copies and identified characteristic structures.

When a segment of the *Galileo* TPase was used as the query subject, to identify TIRs and target site duplications (TSDs), pairwise comparisons of upstream and

downstream flanking sequences (up to 5000-bp, if available) were carried out. TSDs were identified by aligning 50-bp upstream and downstream of the TIRs of the TEs. The target site motif (TSM) was constructed by concatenating the flanking sequence upstream of the element insertion, containing the TSD (50-bp), and the flanking sequence (43-bp) downstream from the element insertion, lacking both this element and the TSD. Hits were considered part of the same *Galileo* copy if arranged in the proper orientation at a distance of <5 Kb.

#### Annotation of the *Galileo* copies

The detected sequences were manually annotated using different tools, most of which were implemented using Geneious R6, created by Biomatters [51], and custom Blast searches using specific *Galileo* and *Drosophila* TE databases. To avoid and discard false automatic identifications, all hits from each search were manually curated. Similarity searches were used to identify and annotate the insertions of other TEs inside *Galileo* by Blast searches, carried out using the National Center for Biotechnology Information (NCBI) [52] and RepeatMasker [53] databases. Default parameters were applied, except for basic options, which we set as follows: cross-match in the search engine, slow in speed/sensitivity, and specify *Drosophila* in DNA source. To verify the presence of direct repeats, we used the Tandem Repeat Finder program, with default parameters [54]. Thus, we identified the following regions in each *Galileo* copy: TIRs, the TPase-coding region, and insertions and repeats.

#### Phylogenetic analysis

Phylogenetic trees were built using the TPase-coding sequences (~630-bp) and the homologous TIR region (100-bp). The sequences were aligned with MAFT software [55]. Phylogenetic analyses were conducted with the maximum likelihood (ML) method, using PHYML 2.4.4 [56] and Bayesian inference of phylogeny (BI) using MrBayes 3.1.2 [57], applying default priors and three heated, one cold Markov chains and running each analysis from two random starting points. For TPase segments, the Akaike's information criterion (AIC, Akaike 1974) indicated that the HKY + G model [58] was the best fit-model of sequence evolution ( $-\ln L = 2502.9929$ , AIC = 5015.9858) and of the gamma-distribution shape parameter (5.7262). The Markov chain Monte Carlo search was run with 10,000,000 generations (repeated two times), with sampling conducted every 1,000 generations. The first 25% of the trees were discarded as "burn-in", at which time the chain reached stationarity, ensuring that the average split frequencies between the runs was < 1%. For the TIR sequences, the AIC indicated that the GTR model [59] with an equal gamma distribution rate was the best fit-model of sequence evolution

( $-\ln L = 785.0532$ , AIC = 1586.1063). The Markov chain Monte Carlo search was run with 10,000,000 generations (repeated tree times) and sampled every 1,000 generations. The first 25% of the trees were discarded as "burn-in", at which time the chain reached stationarity. MEGA version 5.1 was used to calculate the average divergence within and between *Galileo* subfamilies, and the p-distance model and 1,000 bootstrap replications were used to date the divergence between the V and W subfamilies, calibrating the tree with the synonymous substitution rate of 0.016 substitutions per site per million years, as calculated for *Drosophila* genes with a low codon usage bias [28,60].

#### Identification of insertion sites

Insertion sites were analyzed by extracting the flanking sequences (50-bp) upstream and downstream of the element insertion, i.e., those lacking the element and TSD. For this analysis, we restricted the data to those from insertions for which the TSD sequence from each end of the element could be independently determined. To examine the potential secondary structure formed at the insertion site, we used the m-fold web server [61] to analyze the majority rule consensus sequence of the sequences around the TSDs; default parameters were applied, except in the case of the "folding temperature", which was set to 23°C.

#### Additional files

**Additional file 1: *Galileo* copies characterized in the *D. willistoni* genome.** The terminal inverted repeats (TIRs) positioned at the 5' and 3' ends are indicated as TIRa and TIRb, respectively. When the two target site duplications (TSDs) are exactly the same, only one sequence is given. Total length is expressed in base pairs for copies flanked by identical TSDs, with the insertions of other transposable elements (TEs) indicated as appropriate. The length was not estimated for fragmented copies or for those copies without preserved TSDs. Insertions with homology to known elements are indicated with the name of the corresponding TE, and their coordinates are provided according to the database and their localization in the *Galileo* copy.

**Additional file 2: Bayesian inference tree showing the relationships among *Galileo* terminal inverted repeats.** The bootstrap values of each group node are indicated (values below 50% were omitted). The two subfamilies, V and W, are strongly supported. Copies highlighted with symbols are also present in the transposase-encoding segments, as determined in the phylogenetic analysis. *Galileo*, *Kepler*, and *Newton* subfamilies of *D. buzzatii* were used as outgroup in the phylogenetic analysis.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JWG carried out the *in silico* searches, the results analysis, and wrote the manuscript. VHV, AD, VLSV, and AR assisted with the analysis and manuscript writing. AR and VLSV provided funding and facilities for the study. AR conceived and coordinated the study. All authors read and approved the final manuscript.



## Acknowledgements

This study was supported by research grants and fellowships from Brazilian CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), PRONEX-FAPERGS (Programa de Apoio aos Núcleos de Excelência - Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - 10/0028-7), and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Additional funds were provided by grant BFU2011-30476 from Ministerio de Ciencia e Innovación (Spain).

## Author details

<sup>1</sup>Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), CP 15053, Porto Alegre, Rio Grande do Sul 91501-970, Brazil.

<sup>2</sup>Programa de Pós-Graduação em Biologia: Diversidade e Manejo de Vida Silvestre, Universidade do Vale do Rio dos Sinos (UNISINOS), CP 275, São Leopoldo, Rio Grande do Sul 93022-000, Brazil. <sup>3</sup>Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Barcelona, Spain.

Received: 26 April 2014 Accepted: 9 September 2014

Published: 13 September 2014

## References

1. Brookfield JFY: The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet* 2005, **6**:128–136.
2. Feschotte C, Pritham EJ: DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 2007, **41**:331–368.
3. Feschotte C: The contribution of transposable elements to the evolution of regulatory networks. *Nat Rev Genet* 2008, **9**:397–405.
4. Cáceres M, Ranz JM, Barbadailla A, Long M, Ruiz A: Generation of a Widespread *Drosophila* Inversion by a Transposable Element. *Science* 1999, **285**:415–418.
5. Casals F, Cáceres M, Ruiz A: The *Foldback*-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol* 2003, **20**:674–685.
6. Delprat A, Negre B, Puig M, Ruiz A: The transposon *Galileo* generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PLoS One* 2009, **4**:13.
7. Marzo M, Puig M, Ruiz A: The *Foldback*-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. *PNAS* 2008, **105**:2957–2962.
8. Marzo M, Bello X, Puig M, Maside X, Ruiz A: Striking structural dynamism and nucleotide sequence variation of the transposon *Galileo* in the genome of *Drosophila mojavensis*. *Mob DNA* 2013, **4**:1.
9. Spassky B, Richmond RC, Perez Salas S, Pavlovsky O, Mourão CA, Hunter AS, Hoenigsberg H, Dobzhansky T, Ayala FJ: Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. *Evolution* 1971, **25**:129–143.
10. Dobzhansky T, Powell JR: The *Willistoni* Group of Sibling Species of *Drosophila*. In *Handbook of Genetics*. Edited by King RC. New York: Plenum Press; 1975:589–622.
11. Da Cunha AB, Burla H, Dobzhansky T: Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution* 1950, **4**:212–235.
12. Da Cunha AB, Dobzhansky T, Pavlovsky O, Spassky B: Genetics on natural populations. XXVIII. Supplementary data on the chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. *Evolution* 1959, **13**:389–404.
13. Da Cunha AB, Dobzhansky T: A further study of chromosomal polymorphism in *Drosophila willistoni* in the relation to the environment. *Evolution* 1954, **8**:119–134.
14. Valente VLS, Morales NB: New inversions and qualitative description of inversion heterozygotes in natural populations of *Drosophila willistoni*. *Rev Bras Genet* 1985, **8**:167–173.
15. Valente VLS, Araújo AM: Chromosomal polymorphism, climatic factors and variation in population size of *Drosophila willistoni*. *Heredity* 1986, **57**:149–160.
16. Valente VLS, Ruzsyczky A, Santos RA: Chromosomal polymorphism in urban *Drosophila willistoni*. *Rev Bras Genet* 1993, **16**:307–319.
17. Valente VLS, Rohde C, Valiati VH, Morales NB, Goñi B: Chromosomal inversions occurring in Uruguayan populations of *Drosophila willistoni*. *Dros Inf Serv* 2001, **84**:55–59.
18. Valente VLS, Goñi B, Valiati VH, Rohde C, Morales NB: Chromosomal polymorphism in *Drosophila willistoni* populations from Uruguay. *Genet Mol Biol* 2003, **26**:163–173.
19. Rohde C, Cristina A, Garcia L, Valiati VH, Valente VLS: Chromosomal evolution of sibling species of the *Drosophila willistoni* group. I. Chromosomal arm IIR (Muller's element B). *Genetica* 2006, **126**:77–88.
20. Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF, Gelbart WM: Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* 2008, **179**:1657–1680.
21. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: InterProScan: protein domains identifier. *Nucleic Acids Res* 2005, **33**:116–120.
22. Yuan Y-W, Wessler SR: The catalytic of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A* 2011, **108**:1–6.
23. O'Hare K, Rubin GM: Structures of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 1983, **34**:25–35.
24. Hagemann S, Miller JW, Pinsker W: Identification of a complete *P*-element in the genome of *Drosophila bifasciata*. *Nucleic Acids Res* 1992, **20**:409–413.
25. Holyoake AJ, Kidwell MG: *Vege* and *Mar*: two novel *hAT* MITE families from *Drosophila willistoni*. *Mol Biol Evol* 2003, **20**:163–167.
26. Kapitonov VV, Jurka J: Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Natl Acad Sci* 2003, **100**:6569–6574.
27. Reiss D, Quesneville H, Nouaud D, Andrieu O, Anxolabehere D: *Hoppel*, a *P*-like element without introns: a *P*-element ancestral structure or a retrotranscription derivative? *Mol Biol Evol* 2003, **20**:869–879.
28. Sharp PM, Li WH: On the rate of DNA sequences evolution in *Drosophila*. *J Mol Evol* 1989, **28**:398–402.
29. Cáceres M, Puig M, Ruiz A: Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 2001, **11**:1353–1364.
30. Linheiro RS, Bergman CM: Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS One* 2012, **7**:e30008.
31. Rius N, Delprat A, Ruiz A: A divergent *P*-element and its associated MITE, *BuT5*, generate chromosomal inversions and are widespread within the *Drosophila repleta* species group. *Genome Biol Evol* 2013, **5**:1127–1141.
32. Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J: High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* 1990, **62**:515–525.
33. Brunet F, Giraud T, Godin F, Capy P: Do deletions of *Mos1*-like elements occur randomly in the *Drosophilidae* family? *J Mol Evol* 2002, **54**:227–234.
34. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A: Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics* 1990, **124**:339–355.
35. Regner LP, Pereira MS, Alonso CE, Abdelhay E, Valente VL: Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. *J Hered* 1996, **87**:191–198.
36. Deprá M, Ludwig A, Valente VLS, Loreto ELS: *Mar*, a MITE family of *hAT* transposons in *Drosophila*. *Mob DNA* 2012, **3**:13.
37. Moschetti R, Chlamydas S, Marsano RM, Caizzi R: Conserved motifs and dynamic aspects of the terminal inverted repeat organization within *Bari*-like transposons. *Mol Genet Genomics* 2008, **279**:451–461.
38. Ludwig A, Valente VLS, Loreto ELS: Multiple invasions of *Errantivirus* in the genus *Drosophila*. *Insect Mol Biol* 2008, **17**:113–124.
39. Quesneville H, Nouaud D, Anxolabehere D: *P* elements and MITE relatives in the whole genome sequence of *Anopheles gambiae*. *BMC Genomics* 2006, **7**:214.
40. Loreto ELS, Zambra FMB, Ortiz MF, Robe LJ: New *Drosophila P*-like elements and reclassification of *Drosophila P*-elements subfamilies. *Mol Genet Genomics* 2012, **287**:531–540.
41. Robertson HM, MacLeod EG: Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. *Insect Mol Biol* 1993, **2**:125–139.
42. Robertson HM, Martos R: Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. *Gene* 1997, **205**:219–228.
43. Loreto ELS, Carareto CM, Capy P: Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 2008, **100**:545–554.
44. Ehrman L, Powell JR: The *Drosophila Willistoni* Species Group. In *The Genetics and Biology of Drosophila*. Edited by Ashburner M, Carson HL, Thompson JN. New York: Academic Press; 1982:193–225.

45. Cordeiro AR, Winge H: **Levels of Evolutionary Divergence of *Drosophila willistoni* Sibling Species.** In *Genetics of Natural Populations: The Continuing Importance of Theodosius Dobzhansky*. Edited by Levine L. New York: Columbia University Press; 1995:262–280.
46. Robe LJ, Cordeiro J, Loreto ELS, Valente VLS: **Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera : Drosophilidae).** *Genetica* 2010, **138**:601–617.
47. Silva JC, Kidwell MG: **Horizontal transfer and selection in the evolution of *P* elements.** *Mol Bio Evol* 2000, **17**:1542–1557.
48. Casals F, Cáceres M, Manfrin MH, González J, Ruiz A: **Molecular characterization and chromosomal distribution of *Galileo*, *Kepler* and *Newton*, three *foldback* transposable elements of the *Drosophila buzzatii* species complex.** *Genetics* 2005, **169**:2047–2059.
49. Liao GC, Rehm EJ, Rubin GM: **Insertion site preferences of the *P* transposable element in *Drosophila melanogaster*.** *PNAS* 2000, **97**:3347–3351.
50. **A database of *drosophila* genes & genomes.** <http://flybase.org/blast>.
51. **Geneious.** <http://www.geneious.com>.
52. **Basic local alignment search tool.** <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.
53. **RepeatMasker.** <http://www.repeatmasker.org>.
54. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic Acids Res* 1999, **27**:573–580.
55. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**:3059–3066.
56. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696–704.
57. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572–1574.
58. Hasegawa M, Kishino H, Yano T: **Dating of the human-ape splitting by a molecular clock of mitochondrial DNA.** *J Mol Evol* 1985, **22**:160–174.
59. Rodríguez FJ, Oliver JL, Marín A, Medina JR: **The general stochastic model of nucleotide substitution.** *J Theor Biol* 1990, **142**:485–501.
60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**:2731–2739.
61. **The mfold Web server.** <http://www.bioinfo.rpi.edu/applications/mfold>.

doi:10.1186/1471-2164-15-792

**Cite this article as:** Gonçalves et al.: Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. *BMC Genomics* 2014 **15**:792.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



**ADDITIONAL MATERIAL**

---

Additional file 1. *Galileo* copies characterized in the *D. willistoni* genome. The terminal inverted repeats (TIRs) positioned at the 5' and 3' ends are indicated as TIRa and TIRb, respectively. When the two target site duplications (TSDs) are exactly the same, only one sequence is given. Total length is expressed in base pairs for copies flanked by identical TSDs, with the insertions of other transposable elements (TEs) indicated as appropriate. The length was not estimated for fragmented copies or for those copies without preserved TSDs. Insertions with homology to known elements are indicated with the name of the corresponding TE, and their coordinates are provided according to the database and their localization in the *Galileo* copy.

Copy	Scaffold (scf2_110000000)	Coordinates	Transposase coordinates	Transposase-encoding segment length	TIRa/TIRb length	TSD	Total length	TEs insertions	TEs insertions coordinates	TEs flanking	Observations
1	0135	15374-16506	–	–	529/518	ATATTAG	1133	–	–	–	–
2	0170	4-1874	4-1874	1871	–	–	1871	–	–	–	Only transposase fragment
3	0234	1514-2588	1514-2588	1075	–	–	1075	–	–	–	–
4	0374	1497-1881	–	–	_/385	_/CCCATAG	385	–	–	Helitron/Helitron (13-135)	Only TIRb fragment
5	0374	2742-4013	2742-2835/2844-4013	119/1170	–	–	–	–	–	–	Only transposase fragmented
6	0598	4925-6180	–	–	483/325	GTAAAAC	1256	–	–	<i>CR1-1/ LINE</i> (2583-2632)	TE flanking at 3' end
7	0641	1140-3631	1140-2646/3631-3693	1507/63	–	–	–	Helitron/Helitron (828-658 , 379-32)	3036-3197, 3296-3612	–	Only transposase fragmented
8	0774	1216-1521	–	–	257/49	ATACAAC/_	306	–	–	–	–
9	1023	256-1447	331-1447	1117	75/_	–	1192	–	–	–	TIRa: end 694-bp and TSD missing
10	1081	284-1501	284-863/874-1501	580/628	–	–	–	<i>Helitron/Helitron</i> (535-827)	5-283	–	Only transposase fragmented
11	1100	1528-3133	1528-1607/1621-3053	80/1433	_/81	–	–	–	–	–	TIRb: end 694-bp and TSD missing
12	1187	702-1816	702-1816	1115	–	–	1115	–	–	–	–
13	1227	47-4569	47-651/1711-2371/2391-2663/2700-2856	605/661/273/157	–	–	–	<i>Helitron/Helitron</i> (32-136)	670 -774	<i>Helitron/Helitron</i> (827-182)	TE flanking at 3' end
14	1413	645-1173	–	–	183/446	GTATAAT/AT TATAC	629	–	–	–	–

15	1533	3827- 4569	3827- 4569	743	–	–	743	BEL/Pao (44-1 , 5745-5444)	3472-3826	–	
16	1536	1-2119	1-494/1426-1746/1953-2119	494/321/167	–	–	–	Helitron/Helitron (379-138 , 137-33)	1064 -1268, 1304 -1407	–	Only transposase fragmented
17	1569	1353-5127	1376-1987/4664-4740/4758-4833/4837-5127	613/77/76/102	23/_	–	–	R1/ LINE (5403-6120)	3993-4643	–	TIRa: end 741-bp and TSD missing
18	1989	4406-5074	4406-4499/4508-5074/5099-5200	94/567/102	–	–	–	–	–	–	Only transposase fragmented
19	2037	47-3806	47-651/1631-1715/1724-2258/2269-2830/3613-3731	605/85/535/562/119	_/75	–	–	Helitron/Helitron (98-136, 137-829, 827-182 , 130-4)	646-684, 716-1236, 2829-3448, 3488-3614	–	TIRb: end 694-bp and TSD missing
20	2104	1562-2952	1637-2952	1316	75/_	–	1391	–	–	–	TIRa: end 694-bp and TSD missing
21	2163	1513- 2050	1513- 2051	539	–	–	538	–	–	–	Only transposase fragment
22	2504	145-1314	145-1082	938	_/75	–	1170	–	–	–	TIRb: end 694-bp and TSD missing
23	2563	1006-3960	1081-2638/3690-3960	1558/271	75/_	–	–	Helitron/Helitron (828-578 , 379-32)	3031-3272, 3339-3671	–	TIRa: end 694-bp and TSD missing
24	2662	5972-6673	–	–	391/139 +36	GTATAAC/_	701	–	–	Gypsy/Gypsy (198-277, 1073-1119)	TIRb: 36-bp end duplication after a Gypsy fragment and unknown 51-bp
25	2890	163-1302	–	–	520/522	GTATAAC	1140	–	–	–	–
26	3010	6251-7503	–	–	589/400	CTATTAC	1253	–	–	–	–
27	3404	2255-3085	2330-3805	1476	75/_	–	1551	–	–	–	TIRa: end 694-bp and TSD missing
28	3426	254-3736	254-1599/3197-3334/3494-3736	1346/138/243	–	–	–	Helitron/Helitron (811-578, 278- 87, 1-139 , 137-422)	1772 1996, 2130 2322, 2393 2531, 2582 2829	–	–
29	3662	1-819	1-697/726-819	697/94	–	–	–	–	–	–	–

30	4109	1-1242	1-549/574-1242	549/668	-	-	-	-	-	-	-
31	4251	1533-5598	1533-1633/3726-3880/4682-5598	98/155/917	-	-	-	Jockey/Jockey (1085-1524 , 2335-2548), Helitron/Helitron( 32-824)	2030-2460, 3495-3700; 3850-4274	-	-
32	4363	3321-6419	3405-5104/5863-5962/6180-6419	1700/100/240	84/_	-	-	Gypsy/Gypsy (1-311); Helitron/Helitron (379-1)	2931-3239; 5511-5900	-	TIRa: end 694-bp and TSD missing; SOBREPOSIÇÃO GYPSY
33	4369	40-2847	40-1197/2144-2416/2604-2847	1158/273/244	-	-	-	Helitron/Helitron (379-2)	1755-2174	-	-
34	4378	51925-52905	-	-	545/245	GTTTTAC	981	-	-	BEL-10/LTR (159-117)	TE flanking at 3' end
35	4389	16293-19305	-	-	72+61/61	GTATGAG	3113	P/P (1-2935)	16438-19344	-	Two TIRa fragments
36	4392	728-2589	728-2203	1476	_/75	-	-	-	-	-	TIRb: end 694-bp and TSD missing
37	4422	5520-7029	5594-7029	1436	75/_	-	-	-	-	-	TIRa: end 694-bp and TSD missing
38	4488	11479-14417	11479-11750/12801-14343	272/1542	_/75	-	-	Helitron/Helitron (32-222)	11769-11959	-	TIRb: end 694-bp and TSD missing
39	4508	376-1519	-	-	558/518	_/CTATTAG	1144	-	-	Penelope2_Dw (1367-1416)	TIRa: 38-bp end duplication and TSD missing; TE flanking at 5' end
40	4510	1366283-1367425	-	-	516/511	CTTAAAG	1143	-	-	-	-
41	4511	10013271-10014261	-	-	225/537	CTTCAAC	1091	-	-	-	-
42	4513	1277255-1278380	-	-	517/517	TGCAAAG	1126	-	-	-	-
43	4513	5754264-5755458	-	-	521/410	GTTTAAG	1195	-	-	-	-
44	4513	5808696-5809706	-	-	41/479	GTAGAAG	1111	-	-	-	-
45	4514	2345714-2346837	-	-	562/562	GTAAAT/GCTAAAT	1124	-	-	-	-

46	4514	2942702-2943048	-	-	387/60	GTATTGT	447	-	-	-	-
47	4515	4092718-4093942	-	-	664/483	CTGTAAC/GCTAAAT	1225	-	-	-	-
48	4515	2242094-2242544	-	-	105/122+75	CAGTTAC/CTACTAC	551	P/P (2897-2935 , 2900-2935)	2242345-2242383, 2242539-2242574	-	Two TIRb fragments separated by a P fragmented
49	4515	3433966-3434259	-	-	118/277	ATATAAT	394	-	-	-	-
50	4516	146752-146761	-	-	55/55	CTGAAAC	110	-	-	-	-
51	4516	1228777-1229171	-	-	104/385	GTAAAAG	495	-	-	-	-
52	4516	1929304-1931141	1930261-1930604	334	620/570	CCAATAC	1838	1360 element/P(4020-3869)	1930064-1930203	-	-
53	4516	2417579-2418728	-	-	504/440	ATGTAAC	1150	-	-	-	-
54	4521	353645-354781	-	-	377/372	GTTAAAC	1237	-	-	-	-
55	4521	1934341-1935448	-	-	519/528	ATACAAC	1108	-	-	-	-
56	4521	2289832-2290961	-	-	484/484	GTTCAGA	1130	-	-	-	-
57	4521	2795387-2796417	-	-	317/598	GCATTAT	1031	-	-	-	-
58	4521	3111906-3113357	-	-	698/639	GTATTAG	1452	-	-	-	-
59	4521	4399725-4400622	-	-	156/291	GTTGAAG	998	-	-	-	-
60	4521	6230004-6231159	-	-	460/570	CTTTTAT/CTCTTAA	1156	-	-	-	-
61	4521	6438143-6439535	-	-	460/570	ATTTTAC	1393	-	-	-	-
62	4521	6898199-6899356	-	-	496/445	GAAGTAG	1158	-	-	-	-
63	4521	8011873-8012853	-	-	449/495	CTTATAT	981	-	-	-	-
64	4521	9497687-9498964	-	-	529/539	ATAAAAC	1278	-	-	-	-
65	4521	9717122-9718253	-	-	497/604	ATATTAG	1132	-	-	-	-

66	4521	9821137-9822278	-	-	516/520	ATCGTTT/ATCGTAT	1142	-	-	-	TIRb: 38-bp end duplication
67	4521	10310941-10312081	-	-	529/518	GTATTAG	1141	-	-	-	-
68	4521	10666195-10667741	-	-	577/577	CTAAAAG	1547	1360 element/ P(4020-3869)	10666933-10667072	-	-
69	4521	10887650-10888742	-	-	521/365	ATAATAC	1093	-	-	-	-
70	4521	11264259-11265225	-	-	367/448	GTTTTAG	967	-	-	-	-
71	4542	115217-118120	115217-116795/117918-118045	1579/127	_/76	-	-	hAT/hAT-Ac (2480-2355 , 93-1); Helitron/Helitron (423-1)	116861-116990, 117129-117221; 117515-117919	-	TIRb: end 694-bp and TSD missing
72	4542	147369-149321	147901-149321	1421	533/_	-	-	-	-	-	TIRa: end 7-bp and TSD missing
73	4558	839588-840704	-	-	524/521	GTTTAAG	1117	-	-	-	-
74	4558	2321513-2322782	-	-	522/737	CTAAAAC	1270	-	-	-	-
75	4577	1321996-1323129	-	-	521/610	GTTTAAT	1134	-	-	-	-
76	4577	3189626-3190769	-	-	530/524	TTAATAG	1144	-	-	-	-
77	4577	370179-371302	-	-	587/587	ATAAGAA	1124	-	-	-	-
78	4577	3841711-3842912	-	-	509/450	CTTTCAG	1202	-	-	-	-
79	4577	661372-662061	-	-	382/291	CTAAAAC	790	-	-	-	-
80	4580	1226-2904	1513-2904	1392	288/_	-	-	-	-	-	TIRa: end 243-bp and TSD missing
81	4585	2138789-2139894	-	-	517/521	ATTCTGC	1106	-	-	-	-
82	4585	3178488-3179617	-	-	519/519	CTTGAAG	1130	-	-	-	-
83	4585	7051799-7052922	-	-	517/517	CCTTTAC	1124	-	-	-	-
84	4585	3213876-3214341	-	-	129/430	GTTCTAT	566	-	-	-	-
85	4585	3350920-3351654	-	-	444/223+57	CTTTAAC	835	-	-	-	Two TIRb fragments



86	4585	3805977-3806606	-	-	582/127	TTATAAC	730	-	-	-	-
87	4585	4805518-4805417	-	-	402/29	ATTTTCAT	432	-	-	-	-
88	4617	5322-6492	-	-	334/536	GTGGTAC	1171	-	-	-	-
89	4620	19691-20889	-	-	517/602	CTTCAAAG	1199	-	-	-	8-bp long TSDs
90	4644	3109-3503	3109-3503	395	-	-	395	-	-	-	-
91	4726	1611-2856	-	-	242/398	GCGTTAG	1247	-	-	Gypsy/Gypsy (2-52, 46-94)	Undetermined region (Ns) of 605-bp between the TIRs
92	4729	1074614-1075753	-	-	529/517	CTATCAA	1140	-	-	-	-
93	4729	2060832-2061899	-	-	470/545	ATAGAAC	1068	-	-	-	-
94	4729	2226616-2228300	-	-	628/628	ATAGAAC	1685	1360 element/ P(4020-3869)	2227376- 2227515	-	-
95	4762	2925436-2926568	-	-	518/529	CTACAAT	1133	-	-	-	-
96	4762	4707860-4708764	-	-	516/368	GTGCTAC	905	-	-	-	-
97	4762	623693-625010	-	-	340/349+178	GTATCAG/G TTTTAC	1318	MAR/hAT (494-411 , 87-9)	624581-624668, 624744-624822	-	Two TIRb fragments
98	4762	3346986-3347546	-	-	324/248	GTATAAG	661	-	-	-	-
99	4813	24579-28069	25822- 26040/27715- 28069	219/ 355	47/_	-	-	Helitron/Helitron (1- 130, 2-613); Harbinger/PIF- Harbinger (3079- 3143); Helitron/Helitron (614-755); Harbinger/PIF- Harbinger (3161- 3287); Helitron/Helitron (756-827 , 12-386)	26041-26170, 26214-26607; 26608-26672; 26673-26764; 26765-26884; 26885-26927, 27034-27393	-	TIRa: end 56-bp and TSD missing
100	4822	1962606-1963795	-	-	552/638	_ATTATAG	1190	-	-	MAR/hAT (561-610)	TIRa: end 6-bp and TSD missing; TE flanking at 5' end

101	4822	618657-619852	-	-	557/639	CTACTGC	1196	-	-	-	-
102	4837	1024740-1025706	-	-	467/458	CTAATAG	967	-	-	-	-
103	4839	65260-66218	-	-	419/395	CTAATAG	959	-	-	Penelope2/LINE (1416-1367, 1373-1336).	-
104	4852	194158-195281	-	-	521/519	ATAATAT	1124	-	-	-	-
105	4854	239191-240025	-	-	332/491	GTAATAT	835	-	-	-	-
106	4854	123511-124043	-	-	182/393	GTACTAC	633	-	-	-	-
107	4880	34589-40789	34589-34785/35759-36079/37093-38571	197/321/1479	_/75+149	-	-	Helitron/Helitron (827-496, 2-268, 137-829, 1-158); Hoppers/hAT-hobo (297-256); Helitron/Helitron (159-259), Hoppers/hAT-hobo (255-99); Helitron/Helitron (658-828); P/P (3575-3714, 3853-4060); Helitron/Helitron (678-612); Harbinger/PIF-Harbinger (3448-3396, 3300-3164); Helitron/Helitron (349-61).	34784-35106, 36079-36213, 36253-36704, 38640-38736/38738-38778/ 38834-38922/ 38923-38993/ 38961-39121/ 39486-39626, 39504-39678/40010-40085/40033-40085, 40268-40400/ 40313-40556	-	TIRa: end 162-bp and TSD missing; due to the similarity in sequence, is observed overlapping of TE insertions

108	4883	168022-170969	-	-	178/63	GTAATAG	3048	Helitron-2N/Helitron (1-47, 429-490 , 549-827 ); Gypsy-57/Gypsy (2420-2355); Jockey-1/Jockey (277-427, 916 1083, 1408 1535, 1811 1928); ISFUN1/Helitron (171-310).	169064-169110, 169096-169148, 169162-169378; 169640-169704; 169707-169851, 169873-170006, 170054-170174, 170180-170297; 170403-170557	Gypsy-60/Gypsy(8338-8290 , 8295-8247)	-
109	4884	187774-188898	-	-	562/563	GTTTAAG	1125	-	-	-	-
110	4884	274791-275931	-	-	644/518	ATAGAAC	1141	-	-	-	-
111	4884	417409-418362	-	-	517/330	GTTGCAC	954	-	-	<i>Helitron-2/Helitron</i> (90-41)	TE flanking at 5' end
112	4884	1182288-1191308	-	-	4246/4680	GTAATAC	9021	<i>P/P (2900-2935); MAR/hAT (491-597, 299-500)</i>	-	-	The <i>P</i> and <i>MAR</i> (491-597) fragments are present twice (TIRa and TIRb ends); the <i>MAR</i> (299-500) is present five times
113	4887	19692-21379	19692-20460	768	_/288	-	-	Helitron/Helitron (63-313); Harbinger/PIF-Harbinger (3161-3294 , 3294-3415)	20513-20756, 20647-20774, 20973-21093	-	TIRa: end 243-bp and TSD missing

114	4887	96173-99756	96732-97988	1257	528/378	CTGGAGC/_	3584	1360/P (3874-4079); Harbinger/PIF-Harbinger (3448-3294 , 3294-3207); Helitron/Helitron (89-51); Harbinger/PIF-Harbinger (3159-3079); Helitron/Helitron (50-1); P/P (2935-2900, 2896-2935)	98010-98180; 98521-98667; 98962-98963; 98973- 98974; 99055-99065; 99178-99213, 99296-99335	-	TIRb: end 7-bp and TSD missing
115	4887	544208-548081	545533-547154/547879-548006	1622/128	112+119/75	-	-	Helitron/Helitron (1-280); Transib/CMC-Transib (2965-2808); Harbinger/PIF-Harbinger (3163-3294 ,3295-3446); Helitron/Helitron(423-1)	544769-545041/544931-545087/544977-545101, 545331-545464/547476-547880	-	-
116	4887	585081-589088	588314-589088	775	532/_	-	-	Harbinger/PIF-Harbinger (3469-3294 , 3294-3161); Helitron/Helitron(281-79); Gypsy/Gypsy (2614-441); Harbinger/PIF-Harbinger (3159-3078)	585589-585763, 585958-586075; 586000-586193; 586194-588235; 588240-588322	R1/LINE-R1 (5991-4256)	TIRa: end 7-bp and TSD missing
117	4895	55008-55625	-	-	316/319	CCCGAAC	718	-	-	-	-
118	4897	28221-29458	28221-29391	1171	_/67	-	1238	-	-	-	TIRb: end 694-bp and TSD missing
119	4902	4476912-4478035	-	-	604/516	GCATTAT	1124	-	-	-	-
120	4902	7913482-7914498	-	-	508/509	CTTTAAC	1017	-	-	-	-

121	4902	8897512-8898628	-	-	558/559	TTGCTAT	1117	-	-	-	-
122	4902	9433010-9434257	-	-	565/283	GTTTCAT	1248	-	-	-	-
123	4904	21027-22893	21459-22893	1434	73/_	-	-	-	-	-	TIRa: end 694-bp and TSD missing
124	4909	515110-516696	515431-516621	1191	291/75	-	1587	-	-	-	TIRa: end 213 -bp and TSD missing; TIRb: end 694-bp and TSD missing
125	4909	980658-982315	-	-	191 + 349/581	ACCGAAG	1658	<i>MAR/ hAT</i> (610-329 , 131-1)	980838-981119, 981279-981409	-	Two TIRa fragments
126	4909	1135641-1137551	-	-	557/533	CTCCTAC	1911	<i>Helitron/Helitron</i> (423-150, 152-2)	1136534-1136781, 1136844-1136994	-	-
127	4909	5499996-5501595	-	-	528/641	CTAATAAC	1600	-	-	-	8-bp long TSDs
128	4921	1076781-1077987	-	-	558/649	GTAATAT	1207	-	-	-	-
129	4921	2430338-2430911	-	-	153/178	ATATAAC	674	<i>P/P</i> (2729-2935)	2430491-2430685	-	-
130	4921	3839719-3840851	-	-	528/518	GTCAAAG	1133	-	-	-	-
131	4927	54747-55067	54747-55067	321	-	-	321	-	-	<i>BEL/Pao</i> (5137-5435)	TE flanking at 5' end
132	4930	105084-105822	105084-105822	739	-	-	739	-	-	<i>Gypsy/Gypsy</i> (4106-3702)	TE flanking at 3' end
133	4936	829429-830464	-	-	256+221/311	ATCGTAC	1036	-	-	-	Two TIRa fragments separated by undetermined region (Ns)
134	4943	190279-191400	-	-	528/527	GTTGTAC	1122	-	-	-	-
135	4943	518032-519115	-	-	507/498	ATACAAC	1084	-	-	<i>ISFN1/Helitron</i> (223-179)	TE flanking at 3' end
136	4943	530255-531418	-	-	137+50+28+17/134+50+28+17	GTAAGAG	1164	-	-	-	Both TIRs fragmented
137	4943	2347896-2348561	-	-	622/123	ATATTAC	766	-	-	<i>DNAREP1/Helitron</i> (70-21)	TE flanking at 5' end

138	4943	3014865-3015318	-	-	439/114	CTAATAG	554	-	-	-	-
139	4943	6207426-6208055	-	-	127/582	GTACAAC	730	-	-	-	-
140	4943	7323401-7324445	-	-	523/522	ATAAAAG	1045	-	-	-	-
141	4943	7553788-7554852	-	-	390/529	ATAAAAC	1065	-	-	-	-
142	4943	9525872-9526991	-	-	485/504	ATAATAG	1120	-	-	-	-
143	4943	9725795-9726903	-	-	525/524	CTTCTAC	1109	-	-	-	-
144	4943	10677524-10678637	-	-	521/532	CTACTAG	1114	-	-	-	-
145	4943	12398110-12400912	12399071-12399904	834	959/959	CTACAAT	2803	1360/P (4020-3869 , 3869-4079)	12398871-12399010, 12400012-12400183	-	The P fragments are part of the TIRs
146	4943	12416983-12418114	-	-	566/566	ATAGTGC	1132	-	-	-	-
147	4949	275156-279461	276167-276711/277604-278418	545/815	1575/1608	GTTCTAG	4306	1360/P (4020-3869 , 3869-4079)	275915-276054, 276054-278700	-	The two transposase-encoding segments and the P fragments are part of the TIRs; between the TIRs there are an undetermined region (with Ns) of 872-bp
148	4949	686836-687299	-	-	137/427	CTAAGAT	564	-	-	-	-
149	4949	1855234-1856739	-	-	359/ 44 + 71	ATATTGC	1506	MAR/hAT (494-411 , 87-9)	1855852-1855939, 1856015-1856093, 1856204-1856291, 1856367-1856445	-	Two TIRb fragments; twice , interleaved the MAR fragments (494-411 , 87-9)
150	4949	2135514-2136029	-	-	120/401+54	GTATAAT	616	-	-	-	Two TIRb fragments

151	4953	99123-101333	99542-99777/100522-101333	236/812	177/_	GCAAAAC/		<i>Helitron/Helitron</i> (224-1, 423-1)	99332-99537, 100110-100526	<i>Helitron/Helitron</i> (1-938)	TE flanking at 3' end
152	4953	160335-161279	-	-	310/541	CTAATAC	945	-	-	BEL-10/LTR (3267-3316 , 3309-3359)	-
153	4953	470690-474122	470839-471593	755	100/85+100	CTTTCAG	-	<i>Gypsy/Gypsy</i> (942-911); <i>Helitron/Helitron</i> (1-316); <i>Harbinger/PIF-Harbinger</i> (3162-3454); <i>Helitron/Helitron</i> (126-1); <i>Helitron/Helitron</i> (1 - 827)	471655-471686; 471663-471976; 471877-472176; 472165-472294	-	-

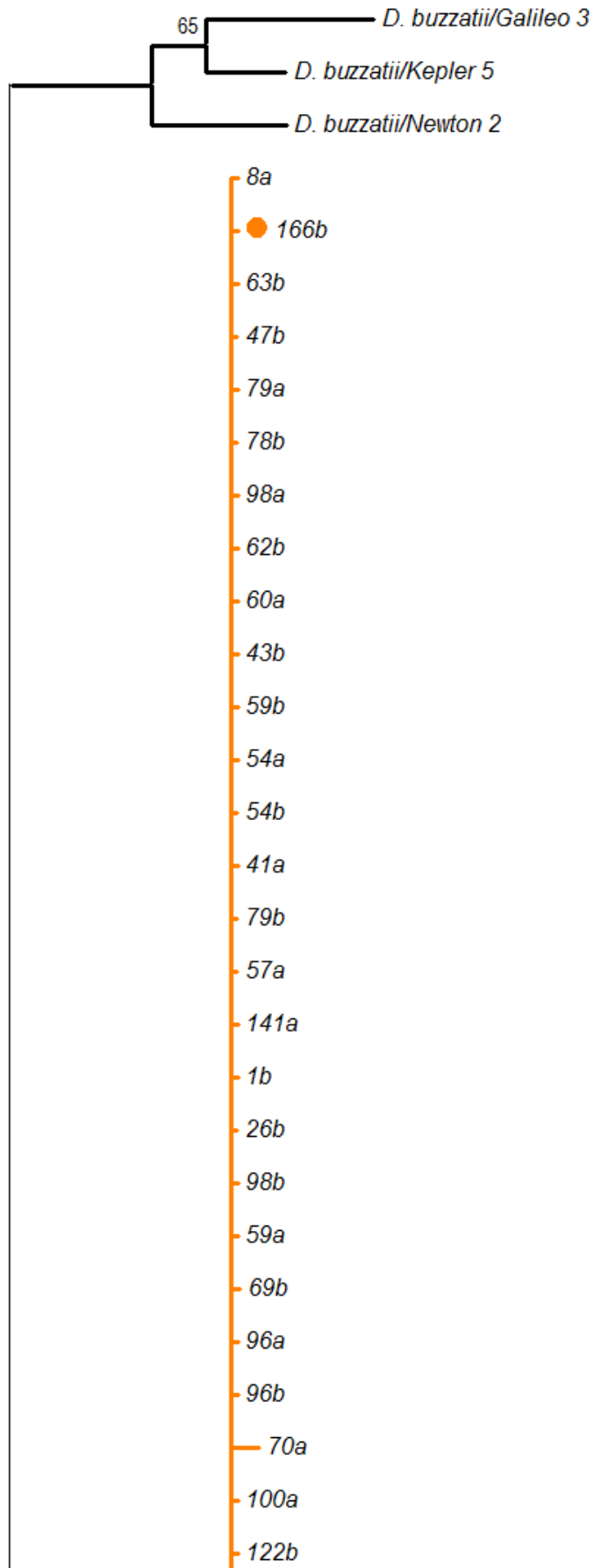
154	4953	1071489-1076232	-	-	100/127+201+100	CTAGAAT	-	<i>Hoana/hAT-Ac</i> (2427-2340); <i>hAT/hAT-Ac</i> (46-8, 2121-1826); <i>Hoana/hAT-Ac</i> (1971-799); <i>hAT/hAT-Ac</i> (560-426, 111-1); <i>MAR/hAT</i> (9-87, 411-494); <i>Gypsy/Gypsy</i> (951-911); <i>Helitron/Helitron</i> (1-272); <i>Transib/CMC-Transib</i> (2965-2808); <i>Harbinger/PIF-Harbinger</i> (3162-3294, 3303-3406); <i>P/P</i> (4060-3888); <i>Helitron/Helitron</i> (5614-5451)	1071683-1071770; 1071741-1071782, 1072035-1072407; 1072517-1073364; 1073669-1073793, 1074129-1074226; 1074372-1074450, 1074526-1074613; 1074938-1074976; 1074953-1075177; 1075098-1075255; 1075143-1075269, 1075493-1075598; 1075943-1076086; 1075952-1076104	<i>CR1/CR1</i> (1793-1744, 1750-1701)	-
155	4953	1126764-1128595	-	-	_/311+100	_/GTAATAA	<i>VEGE/hAT-Tip100</i> (762-884); <i>MAR/hAT</i> (494-411 , 83-9)	1127889-1128013; 1128235-1128322; 1128402-1128476	<i>Harbinger/PIF-Harbinger</i> (3315-3445)	TE flanking at 3' end	
156	4953	1130291-1134090	1130291-1131297/1132303-1132500/1133159-1134090	1007/191/932	-	-	<i>Helitron/Helitron</i> ( 1-303, 172-314, 15-312); <i>Hopers/hAT-hobo</i> (297-175)	1131294-1131655, 1131724-1131996; 1132496-1132688; 1132596-1132700	<i>Transib2/Transib</i> (3032-2903)	Due to the similarity in sequence, is observed overlapping of TE insertions	



157	4953	1278329-1279581	1278329-1279581	1253	–	–	1253			<i>BEL/Pao</i> (5444-5745)	TE flanking at 3' end
158	4953	1286124-1290707	1286124-1286784/1289923-1289996/1290014-1290707	661/74/694	24/_	–	–	<i>Gypsy/Gypsy</i> (164-200, 1-85, 124-353); <i>R1/LINE</i> (247-1506, 4578-6120)	1286806-1286935, 1287135-1287355; 1287352-1289899	–	TIRa: end 741-bp and TSD missing
159	4953	1295987-1297238	1296934-1297007/1297032-1297144/1297162-1297238	74/113/77	330/_	–	–	<i>Harbinger/PIF-Harbinger</i> (3449-3286, 3294-3171); <i>Helitron/Helitron</i> (281-1)	1296312-1296479, 1296662-1296769; 1296704-1296963	–	TIRa: end 205-bp and TSD missing
160	4954	533337-534759	–	–	792/533	GTATTAG	1423	–	–	–	–
161	4954	2040286-2041430	–	–	524/530	CTACTAC	1145	–	–	<i>Howilli5/hAT</i> (152-104, 112-62)	–
162	4954	2296203-2297318	–	–	571/581	ACCGTAT	1116	–	–	<i>BEL-1/LTR</i> (6818-6857, 6854-6904)	–
163	4954	3609317-3609689	–	–	92/379	GTCTTAC	473	–	–	–	TIRb: end 89-bp missing
164	4954	48339341-48340488	–	–	611/513	GTATGAC/_	1148	–	–	–	TIRb: end 23-bp and TSD missing
165	4958	249647-250745	–	–	523/514	CTTGGAT	1099	–	–	–	–
166	4958	317856-322241	318839-321553	2715	765/757	ATTCTAG	4386	–	–	<i>CR1-1/LINE</i> (4517-1744, 1750-1565)	TE flanking at both ends. Nearly complete copy (BK006360.1)
167	4959	72021-72467	–	–	274/273	ATGATAG	547	–	–	<i>Gypsy/Gypsy</i> (5598-5553, 5558-5510)	–
168	4959	180122-181189	–	–	489/523	ATAATAC	1068	–	–	<i>Hoana8/hAT</i> (1995-1945)	TE flanking at 5' end
169	4961	2476375-2477021	–	–	34/621	GTATTAC	747	–	–	–	–
170	4961	3475701-3476111	–	–	438/73	GTATTAC	511	–	–	–	–

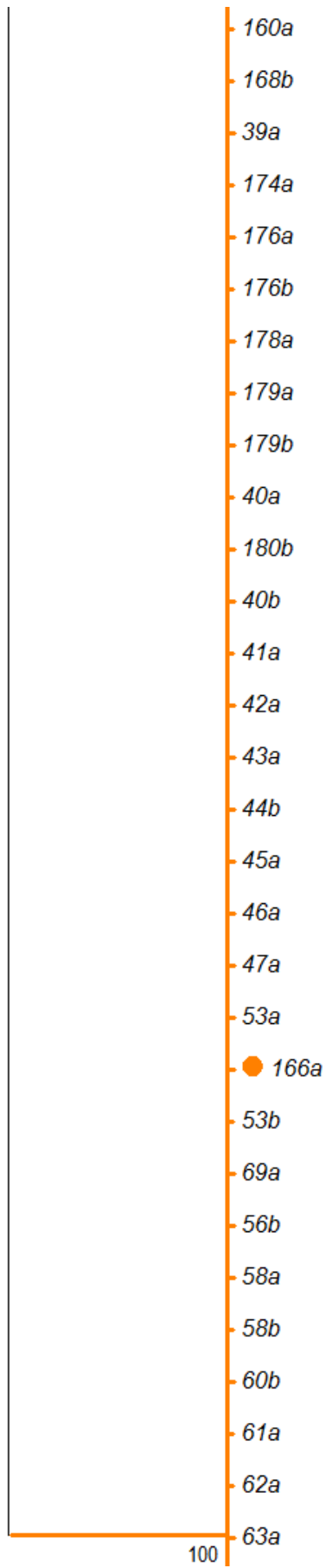
171	4961	45203080-45204085	-	-	334/526	_ATACAAG	1006	<i>Helitron-2N1/Helitron (125-1)</i>	4520733-4520857	TIRa: end 1-bp and TSD missing	-
172	4962	562804-566857	566599-566857	259	352/_	-	-	<i>Helitron/Helitron (205-827), hAT-N1/hAT-Ac (1-94), Helitron/Helitron (303-1)</i>	563161-563687; 564061-564154; 566240-566602	<i>Helitron/Helitron (827-1)</i>	TIRa: end 150-bp and TSD missing
173	4963	203509-204629	-	-	482/391	ATTTAC	1121	-	-	<i>DNAREP1/Helitron(70-21)</i>	TE flanking at 5' end
174	4963	1109997-1111113	-	-	525/520	CTTTGAC	1117	-	-	-	-
175	4963	2065909-2066213	-	-	135/245	CTTATAT	405	-	-	-	TIRb: end 210-bp missing; 17-bp between TIRb and TSD
176	4963	2566468-2567869	-	-	533/522	ATAACAG	1402	-	-	-	-
177	4963	3776642-3777619	-	-	361/453	ATTATAC	978	-	-	-	-
178	4967	2328251-2329287	-	-	482+172/171+287	ACGTTAT	1037	-	-	<i>Gypsy-40/Gypsy (98-65)</i>	Both TIRs fragmented; TE flanking at 3' end
179	4967	2896962-2898053	-	-	546/546	GTATAAAT	1092	-	-	<i>Gypsy-41/Gypsy (286-232)</i>	8-bp long TSDs; TE flanking at 5' end
180	4968	1508443-1508890	-	-	90/458	ATAGCAT	548	-	-	<i>CR1-1_/LINE (4410-4361)</i>	TE flanking at 5' end
181	4973	50976-53103	50976-51220/51408-51663/52367-52453/524690-52692/52696-53103	245/256/87/224/408	-	-	-	<i>Helitron/Helitron (1-271), Hopers2/hAT-hobo (297-100)</i>	51662-51912; 51748-51936	<i>Helitron/Helitron (1-271)</i>	Due to the similarity in sequence, is observed overlapping of TE insertions
182	5111	2-1108	2-479/481-1108	469/628	-	-	-	-	-	-	-
183	7459	34-993	34-918	885	_/75	-	-	-	-	-	TIRb: end 694-bp and TSD missing
184	9598	613-931	613-689/707-819/823-931	77/113/113	-	-	-	-	-	<i>Helitron/Helitron (1-281)</i>	-

185	10020	16-1184	16-548/559-1184	533/626				-	-	-	-
186	11016	1-1388	1-1388	1388	-	-	1388	-	-	-	-
187	11140	277-1137	277-310/379-1137	94/759	-	-	-	-	-	Gypsy/Gyps(479 -840 , 1-361)	-
188	11566	1-655	1-655	665	-	-		-	-	-	-
189	12249	750-947	750-947	198	-	-	198	-	-	Helitron/Helitron (376-178)	-
190	14411	236-1180	236-329/338-1180	94/842	-	-	-	-	-	-	-
191	14656	1-686	1-686	686	-	-	686	-	-	-	-



- 97a
- 39b
- 103a
- 66b
- 173b
- 97b
- 6a
- 6b
- 148b
- 117a
- 177b
- 178b
- 148a
- 51b
- 138b
- 34b
- ▼ 151a
- 14a
- 88a
- 106a
- 133b
- 150a
- 152a
- 117b
- 49a
- 103b
- 66a
- 164a
- 86b
- 139a

24a  
119a  
119b  
120a  
120b  
121a  
126a  
126b  
25a  
127a  
127b  
125b  
128a  
128b  
144a  
144b  
146a  
146b  
134a  
135a  
135b  
140a  
141a  
142a  
142b  
143a  
143b  
152b  
161a  
162a

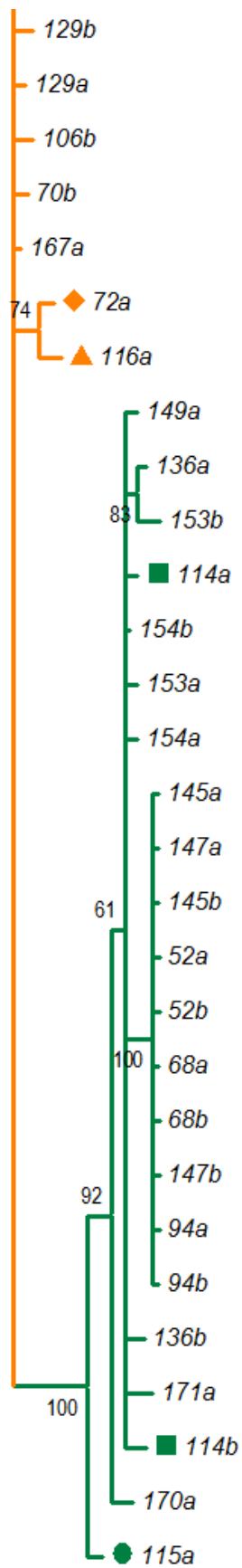


64a  
64b  
74a  
73a  
76b  
75a  
75b  
76a  
77a  
77b  
78a  
82a  
82b  
83a  
83b  
89a  
92b  
93a  
93b  
100b  
101a  
101b  
102a  
104a  
104b  
109b  
110a  
110b  
121b  
34a



168a  
56a  
130a  
130b  
137a  
139a  
26a  
67a  
67b  
81a  
1a  
95a  
95b  
102b  
65a  
65b  
109a  
57b  
140b  
89b  
42b  
161b  
81b  
14b  
74b  
61b  
165a  
122a  
111a  
25b

138a  
162b  
160b  
173a  
174b  
73b  
92a  
86a  
45b  
55a  
55b  
88b  
134b  
169b  
177a  
24b  
125a  
111b  
105a  
51a  
108a  
167b  
49b  
133a  
85a  
175a  
48a  
84a  
171b  
105b



Additional file 2. Bayesian inference tree showing the relationships among *Galileo* terminal inverted repeats. The bootstrap values of each group node are indicated (values below 50% were omitted). The two subfamilies, V and W, are strongly supported. Copies highlighted with symbols are also present in the transposase-encoding segments, as determined in the phylogenetic analysis. *Galileo*, *Kepler*, and *Newton* subfamilies of *D. buzzatii* were used as outgroup in the phylogenetic analysis.

## CAPÍTULO III

---

Trabalho em preparação para ser submetido à revista *Genome Biology and Evolution*

### **Transposon *Galileo* nos grupos *willistoni* e *cardini* de *Drosophila***

Juliana W Gonçalves<sup>1</sup>, Victor Hugo Valiati<sup>2\*</sup>, Alejandra Delprat<sup>3</sup>,  
Alfredo Ruiz<sup>3\*</sup> and Vera L S Valente<sup>1\*</sup>

<sup>1</sup> Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), CP 15053, Porto Alegre, Rio Grande do Sul, 91501-970, Brazil.

<sup>2</sup> Programa de Pós-Graduação em Biologia: Diversidade e Manejo de Vida Silvestre, Universidade do Vale do Rio dos Sinos (UNISINOS), CP 275, São Leopoldo, Rio Grande do Sul, 93022-000, Brazil.

<sup>3</sup> Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona (UAB), 08193, Bellaterra (Barcelona), Spain.

Correspondence: vera.valente@pq.cnpq.br

\* Equal contributors

## Resumo

*Galileo* desde a sua descoberta em *D. buzzatii*, onde pela primeira vez foi apontado como indutor de inversão cromossômica, teve sua história evolutiva investigada apenas na referida espécie, em *D. mojavensis* e em *D. willistoni*. Com o objetivo de compreender o cenário evolutivo de *Galileo* em outras espécies polimórficas cromossomicamente, que pertencem a subgêneros diferentes, mas que compartilham dos mesmos recursos na natureza, realizou-se investigação experimental em 18 linhagens de espécies do grupo *willistoni* e em três espécies do subgrupo *cardini* de *Drosophila*. Foram delineados *primers* específicos para uma região conservada da inferida transposase e por meio das técnicas de amplificação de DNA, clonagem e sequenciamento, verificou-se a presença deste transposon em algumas das espécies destes drosofilídeos. Na reconstrução das relações evolutivas de sequências correspondentes a *Galileo*, hospedadas no genoma de espécies do grupo *willistoni* e do subgrupo *cardini* de *Drosophila*, detectou-se desde incongruências filogenéticas, distribuição descontínua entre espécies do subgrupo *bocainenses* até alta similaridade entre as sequências isoladas. Estimativas promissoras sugerem um panorama evolutivo de transferência vertical e horizontal, perdas estocásticas e possíveis cruzamentos interespecíficos atuando na origem e manutenção de *Galileo* nestas espécies neotropicais de *Drosophila*.

## Introdução

*Galileo*, classificado como membro da superfamília *P* de transposons (Marzo *et al.*, 2008), foi originalmente descoberto em *Drosophila buzzatii*. Evidências diretas o apontaram como indutor de três inversões cromossômicas segregantes, originadas por recombinação ectópica entre cópias encontradas nos respectivos pontos de quebra (Cáceres *et al.*, 1999; Casals *et al.*, 2003; Delprat *et al.*, 2009). Marzo *et al.* (2008) identificaram variantes de *Galileo* não apenas em representantes do grupo *repleta*, mas em 6 espécies (*D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. virilis* e *D. mojavensis*) dos 12 primeiros genomas de *Drosophila* analisados (Clark *et al.*, 2007). *Galileo* apresenta-se como um transposon bastante

variável em composição e comprimento, dotado de *TIRs* com um impressionante dinamismo estrutural (Casals *et al.*, 2005; Marzo *et al.*, 2013; Gonçalves *et al.*, 2014 – presente tese – Cap. II). A partir da investigação *in silico* e detalhada caracterização de cópias de *Galileo* no genoma de *Drosophila willistoni*, duas subfamílias desse transposon foram identificadas com base em nossas análises filogenéticas (Gonçalves *et al.*, 2014).

Há duas possíveis formas dos elementos transponíveis (*TEs*) serem transferidos entre indivíduos. A primeira, transmissão vertical, é estritamente dependente do acasalamento e, a denominada transferência horizontal (*HT*), ocorre entre as espécies que são reprodutivamente isoladas. A transmissão vertical prevê que todos os descendentes de uma espécie ancestral que abriga um *TE*, também devem abrigar variantes homólogas do elemento. Além disso, a conservação da sequência de um *TE* deve ser congruente com a filogenia das espécies hospedeiras. Essas premissas podem ser violadas no caso de polimorfismo ancestral, perdas estocásticas e seleção (Almeida e Carareto, 2005). Por outro lado, a ocorrência de *HT* pode ser apresentada por meio de uma distribuição descontínua onde a conservação da sequência do elemento não está necessariamente relacionada com a filogenia das espécies portadoras (revisado por Silva *et al.*, 2004 e Loreto *et al.*, 2008)

A questão de como e se os *TEs* persistem nos genomas, enquanto não oferecerem benefício direto ao hospedeiro, sempre foi intrigante (Pritham, 2009). No início da década de 90 foi descoberta a capacidade do elemento *P* de *Drosophila* de colonizar novos genomas por meio da *HT*, revelando assim uma forma adicional dos *TEs* se manterem ao longo do tempo (Daniels *et al.*, 1990). Essa via de transmissão permitiria que o elemento escapasse de uma extinção vertical, aparentemente inevitável na linhagem hospedeira original, resultante da eliminação por deriva ou seleção, acúmulo de mutações ou por mecanismos de defesa do hospedeiro inativando-o (Robertson, 2002; revisado por Schaack *et al.*, 2010).

Torna-se difícil detectar *HT* porque é necessário considerar todas as características intrínsecas do dinamismo dos *TEs*, tais como degeneração das sequências, perda estocástica e diferentes taxas evolutivas (Cumplings, 1994;

Capy *et al.*, 1998). Além disso, diferentes fases do ciclo de vida complexo de um *TE*, podem gerar padrões filogenéticos semelhantes aqueles esperados em um evento de *HT* (Wallau *et al.*, 2012). Segundo revisão feita por Schaack *et al.* (2010), 218 são os casos convincentes de *HT*, sendo que destes 103 foram realizadas por transposons, 97 por retrotransposons com *LTR* e 14 dos casos por retrotransposons sem *LTR*.

Além dos requisitos mínimos, tais como sobreposição geográfica, temporal e ecológica entre as espécies doadoras e receptoras, três tipos diferentes de evidências são geralmente usados para inferir *HT* de *TEs*: alta similaridade entre *TEs* de espécies distantemente relacionadas (Daniels *et al.*, 1990; Robertson e Lampe, 1995; Brunet *et al.*, 1999), a incongruência entre as filogenias do *TE* e das espécies (Robertson e Lampe, 1995; Terzian *et al.*, 2000; Almeida e Carareto, 2005 ) e ocorrência descontínua do *TE* em um grupo de espécies (Daniels *et al.*, 1990; Arca e Savakis, 2000; Loreto *et al.*, 2001).

Com intuito de verificar a distribuição de *Galileo* e aprofundar a compreensão da história evolutiva desse transposon, neste trabalho realizou-se a busca experimental de *Galileo* em populações naturais de espécies intimamente relacionadas à *D. willistoni* e também em representantes do grupo *cardini* de *Drosophila*. Representando os subgêneros *Sophophora* e *Drosophila*, o grupo *willistoni* e o grupo *cardini*, respectivamente, são compostos por espécies neotropicais, ecologicamente similares e versáteis, além de apresentarem um polimorfismo cromossômico instigante. Os dados obtidos no presente estudo mostram que a existência das subfamílias *V* e *W* de *Galileo* não se restringem apenas à *D. willistoni* (Gonçalves *et al.*, 2014), mas estão presentes tanto em espécies do subgrupo *willistoni* quanto do subgrupo *bocainensis* de *Drosophila*. Além disso, pela primeira vez, é demonstrado evidências que suportam a ocorrência de eventos de transferência horizontal envolvendo o elemento *Galileo*, entre espécies do subgrupo *willistoni* e *bocainensis*, assim como entre o grupo *willistoni* e o grupo *cardini*.



## **Materiais e Métodos**

### **Manutenção dos estoques de *Drosophila***

As linhagens e espécies utilizadas (Tabela 1) foram mantidas como culturas massais, dentro de câmaras de temperatura e umidade constante (a 17 °C ou a 25 °C  $\pm$  1 grau e 60% de umidade relativa). A manutenção dos estoques foi realizada pela repicagem semanal das linhagens em meio de cultura estabelecido por Marques *et al.* (1966). Em todos os experimentos utilizou-se como controle a linhagem sequenciada de *D. willistoni*, GdH-4 (Clark *et al.*, 2007).

### **Extração de DNA**

O DNA genômico foi extraído a partir de moscas adultas com adaptações ao método de Sassi *et al.* (2005). Modificou-se o volume do tampão de lise (Tris-HCl 0.1 M; EDTA 0.1 M; SDS 1% e NaCl 0.06 M), de 750  $\mu$ L para 650  $\mu$ L, assim como o período de incubação, de 30 minutos para 1 hora a 65 °C. Além disso, ressuspendeu-se o DNA extraído em 30-50  $\mu$ L de H<sub>2</sub>O ultra pura e não em Tris-EDTA (TE), quando o objetivo não era estocar por longo período a amostra.

### **Técnica da Reação em Cadeia da Polimerase (PCR)**

Para a caracterização molecular do elemento *Galileo* utilizou-se dois pares de *primers* denominados de GalF (5'-AGG GGC ATT CAA TAA ACT TGG-3') e GalR "(5'-TGT CTT TCT GGA TCG CTG ATT-3'), Gal11F (5'-ATG TCC CCC ACC TCA TAA AAT Y-3') e Gal12R (5'-ACC TTC TCC TTG ACT CCA AAT ATY-3'). Foram desenhados com base na cópia mais completa disponível (GenBank: BK006360.1) encontrada por Marzo *et al.* (2008) no genoma de *Drosophila willistoni* (Clark *et al.*, 2007) e delineados de forma que flanqueassem uma região conservada da inferida transposase, com fragmentos esperados de 470 e 530 pb (Figura 1).

Na técnica da PCR utilizou-se um volume total de 25-50  $\mu\text{L}$  com 20 ng de DNA genômico, 20 pmol de cada *primer*, 200  $\mu\text{M}$  de dNTPs, 1,0 mM de  $\text{MgCl}_2$  e 1-1,5 unidade de Taq DNA polimerase. Empregou-se uma desnaturação inicial a 94  $^\circ\text{C}$  por 3 minutos, seguida por 30 ciclos de desnaturação a 94  $^\circ\text{C}$  por 1 minuto, anelamento dos *primers* a 52-62  $^\circ\text{C}$  por 45 segundos, extensão a 72  $^\circ\text{C}$  por 3 minutos e extensão final a 72  $^\circ\text{C}$  por 6 minutos, ou por 10 minutos para a ligação do fragmento amplificado ao vetor de clonagem.

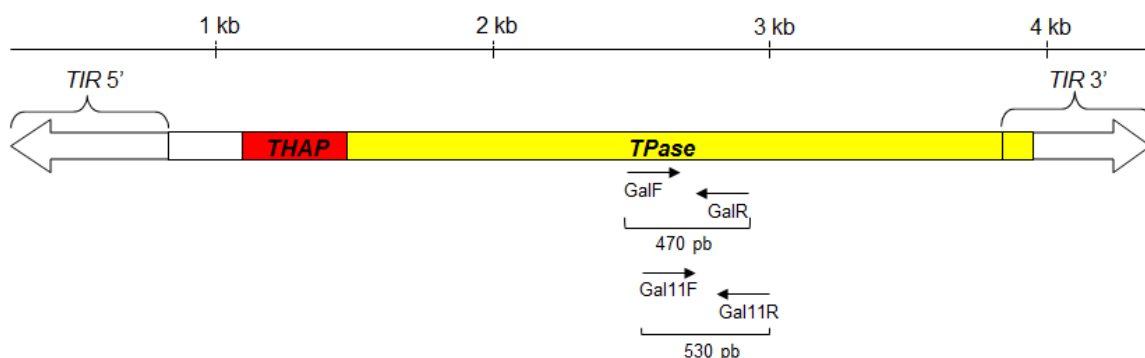


Figura 1: Representação esquemática da maior variante de *Galileo* em *D. willistoni* descrita por Marzo *et al.* (2008), com a localização dos pares de *primers* desenhados e a respectiva região amplificada.

## Clonagem e sequenciamento

Os produtos amplificados foram purificados a partir da excisão da banda do gel de agarose após a eletroforese, utilizando-se o kit *GFX<sup>TM</sup> PCR and Gel Band Purification®* (GE Healthcare) conforme recomendações do fabricante. Os fragmentos isolados foram clonados no vetor *pGEM<sup>®</sup>-T Easy Vector System* (Promega) e posteriormente sequenciados automaticamente com os *primers* universais M13F e M13R-pUC pela *Advancing through Genomics MACROGEN*. Pelo menos três clones (e suas duas fitas) foram sequenciados.

## Definição dos nomes das sequências

Nomes foram atribuídos às sequências segundo Wicker *et al.* (2007). As três primeiras letras correspondem à Classe (*DNA transposons*), à Ordem

(*TIR*) e à superfamília (*P*), resultando em DTP. Em seguida temos a família *Galileo* e a subfamília (*V* ou *W*) (Gonçalves *et al.*, 2014). Na continuação, foram representadas as espécies, onde a letra maiúscula indica o gênero e as três letras minúsculas o epíteto específico ou nos casos de sequências isoladas da espécie *D. paulistorum*, representou-se as subespécies com três letras minúsculas apenas. Sequências consenso das subfamílias *V* e *W* de *Galileo* foram obtidas separadamente para cada espécie investigada e identificadas com a abreviatura “cons”.

### **Análise filogenética**

A sequência consenso para cada clone foi obtida com o uso do *Staden Package* (Staden, 1996). As sequências foram alinhadas utilizando-se o *software* MAFT (Kato *et al.*, 2002) com os parâmetros padrões e analisadas com outras ferramentas do *Geneious* (Drummond *et al.*, 2010). As árvores filogenéticas foram construídas com um fragmento da inserida transposase de aproximadamente 417 pb. As relações evolutivas foram estimadas com o método de Máxima Verossimilhança (MV) usando-se o PHYML 2.4.4 (Guindon e Gascuel, 2003) e pelo método de Inferência Bayesiana (BA) por meio do MrBayes 3.1.2 (Ronquist *et al.*, 2005). O programa Modeltest (Posada e Crandall, 1998) foi utilizado e determinou o modelo GTR+I+G como o melhor modelo de evolução das sequências nucleotídicas baseado no critério de informação AIC (Akaike, 1974), com o auxílio do programa Paup, versão 4 (Swofford, 2002). Para o BA dez milhões de gerações usando duas corridas independentes foram submetidas aos conjuntos de dados. As Cadeias de Markov foram amostradas em intervalos de 1000 gerações, removendo-se 25% das primeiras árvores como *burn-in* para evitar incluir amostras antes de chegar a estacionariedade.

As divergências genéticas nucleotídicas e aminoácidas das sequências consenso de cada espécie e subfamília de *Galileo* foram calculadas utilizando-se a distância p no Mega 5.1 (Tamura *et al.*, 2011), assim como o tempo de divergência dos clados obtidos na reconstrução filogenética considerando-se a taxa de substituições sinônimas de 0,016 por milhão de anos (Sharp e Li, 1989).

## Inferência de transferência horizontal

No programa DNASP 5.10.1 (Librado e Rozas, 2009) foram calculados, para *Galileo* e os genes álcool desidrogenase (*Adh*) e  $\alpha$ -metildopa (*Amd*), o número efetivo de códons (*ENC*) (Wright, 1990) e o índice de viés de códon (*CBI*) (Morton, 1993). Os genes *Adh* e *Amd* foram escolhidos por estarem disponíveis para ambos os grupos *willistoni* e *cardini*. Os códigos de acesso no *Genbank*, assim como a origem geográfica das espécies estão apresentados no Material Suplementar, Tabela 1.

O número de sítios sinônimos (S) e de substituições sinônimas por sítios sinônimos (dS) foram obtidos usando o método Nei-Gojobori (Nei e Gojobori, 1986), com distância p e correção de Junkes-Cantor (1969), implementado pelo Mega 4 (Tamura *et al.*, 2007) e Mega 5.1 (Tamura *et al.*, 2011), respectivamente. Nessa análise, *gaps* foram introduzidos para que o quadro de leitura fosse conservado, enquanto códons de parada foram tratados como informação ausente. Para verificar se as diferenças observadas entre os dS de *Galileo* e do gene *Amd* eram significativamente menores, realizou-se o teste de Fisher, implementado no DNASP 5.10.1, com o número de similaridades (S-(S.dS)) e diferenças (S.dS) observadas nos sítios sinônimos (Ludwig e Loreto 2008).

## Resultados

### Busca experimental

Do grupo *willistoni* de *Drosophila*, foram investigadas *D. willistoni* (três linhagens, além da linhagem GdH-4), *D. tropicalis*, *D. equinoxialis*, *D. insularis* e *D. paulistorum* (seis semiespécies: Amazônica, Andino-Brasileira (duas linhagens), Centro-Americana, Interior, Orinocana e Transicional) do subgrupo *willistoni*; as espécies *D. nebulosa*, *D. capricorni*, *D. succinea* e *D. fumipennis* do subgrupo *bocainensis*. Do grupo *cardini*, as espécies *D. cardini*, *D. neocardini* e *D. neomorpha* (subgrupo *cardini*) foram testadas e seus amplicons sequenciados (Tabela 1). Cabe ressaltar que os resultados, aparentemente negativos quanto

à presença de *Galileo* no genoma de algumas espécies, foram exaustiva e cautelosamente trabalhados, assim como os casos de baixos rendimentos dos produtos de amplificações.

Com auxílio das ferramentas *blastn* e *blastx* identificou-se 109 sequências correspondentes a *Galileo* (identidade nucleotídica de 78 à 100%). A maioria das espécies do grupo *willistoni*, tiveram fragmentos de *Galileo* isolados. Porém todos os produtos amplificados, clonados e sequenciados das espécies *D. succinea* e *D. fumipennis*, do subgrupo *bocainensis*, corresponderam a sequências inespecíficas. No subgrupo *cardini*, pela primeira vez, registra-se sequências nucleotídicas correspondentes à *Galileo* nas espécies *D. neocardini* e *D. neomorpha* e em *D. cardini* apenas um pequeno fragmento de 62 pb com identidade nucleotídica de 83,6% foi isolado.

### **Análise filogenética**

As sequências nucleotídicas, com aproximadamente 417 pb, foram utilizadas na reconstrução das relações evolutivas de *Galileo* no grupo *willistoni* e no subgrupo *cardini* de *Drosophila*. As árvores resultantes dos métodos de *ML* e *BA* apresentaram topologias similares, nas quais dois clados com o máximo de suporte é observado (Figura 2; Material Suplementar, Figuras 1, 2, 3). Os agrupamentos correspondem às subfamílias *V* e *W* de *Galileo* (Gonçalves *et al.*, 2014), que coexistem nos genomas de espécies dos subgrupos *willistoni* e *bocainensis*. A semiespécie Amazônica de *D. paulistorum*, do subgrupo *willistoni*, apresentou apenas a subfamília *V*. As espécies do subgrupo *cardini* hospedam em seus genomas a subfamília *V* de *Galileo*, como bem demonstrado pelas sequências isoladas com altíssima identidade (97,7%) às sequências dessa subfamília primeiramente descrita para a espécie *D. willistoni*.

As divergências genéticas, nucleotídica e aminoacídica, entre e dentro das subfamílias *V* e *W* estão apresentadas na Tabela 1. As distâncias genéticas par a par das subfamílias de *Galileo* nas as espécies investigadas estão apresentadas no Material Suplementar, Tabela 1.

Tabela 1. Lista de espécies e linhagens de *Drosophila* mantidas em culturas massais utilizadas neste trabalho e seus respectivos resultados quanto à presença do transposon *Galileo*. A identificação refere-se à utilizada neste trabalho.

Grupo	Subgrupo	Espécie	Semiespécie	Identificação	Origem geográfica	PCR
<i>willistoni</i>	<i>willistoni</i>	<i>D. willistoni</i>		GdH-4	Guadaloupe, Pequenas Antilhas	+
				17A2	Eldorado do Sul (RS), Brasil	+
				cor	Coronilla, Uruguai	+
				wip4	Salvador (BA), Brasil	+
		<i>D. paulistorum</i>	Amazônica	ama	Bélem (PA), Brasil	+
	and-MLC			Florianópolis (SC), Brasil	+	
	and-Rib			Ribeirão Preto (SP), Brasil	+	
			Centro-Americana	cen	Lancetilla, Honduras	+
			Interior	int	Llanos, Colômbia	+
			Orinocana	ori	Georgetown, Guiana	+
			Transicional	tra	Santa Marta, Colômbia	+
		<i>D. tropicalis</i>		Dtro	San Salvador, El Salvador	+
		<i>D. equinoxialis</i>		Dequ	Cidade do México, México	+
		<i>D. insularis</i>		Dins	Saint Kitts, Pequenas Antilhas	+
	<i>bocainensis</i>	<i>D. nebulosa</i>		Dneb	Porto Alegre (RS), Brasil	+
			Dcap	Porto Alegre (RS), Brasil	+	
			Dsuc	Cidade do México, México	-	
			Dfum	Joinville (SC), Brasil	-	
			<i>D. cardini</i>		Dcar	Itaqui (RS), Brasil
<i>cardini</i>	<i>cardini</i>	<i>D. cardini</i>		Dcar	Itaqui (RS), Brasil	+
		<i>D. neocardini</i>		Dneoc	Florianópolis (SC), Brasil	+
		<i>D. neomorpha</i>		Dneom	-	+

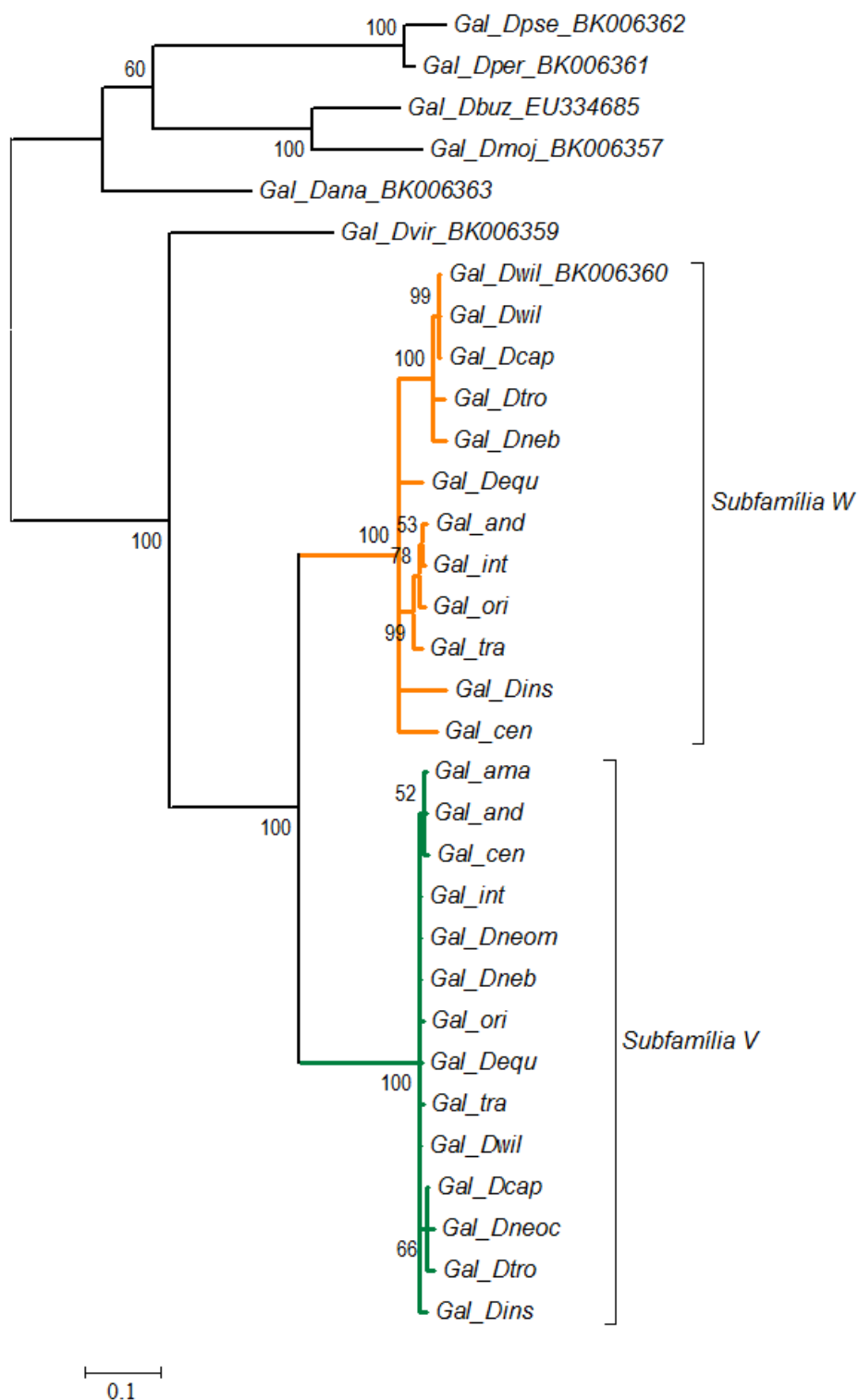


Figura 2. Árvore filogenética construída com seqüências consenso da inferida transposase de *Galileo* em *Drosophila* pela inferência Bayesiana. Probabilidades *a posteriori* estão indicadas em cada nó (valores inferiores a 50% foram omitidos). As duas subfamílias, V e W, são apresentadas com o máximo de suporte. Sequências de *D. ananassae*, *D. buzzatii*, *D. mojavensis*, *D. persimilis*, *D pseudoobscura* e *D. virilis* foram usadas como grupo externo.

Tabela 1. Divergência nucleotídica e aminoacídica (entre parêntesis) estimada (%) para as subfamílias de *Galileo* do grupo *willistoni* e representantes do subgrupo *cardini* de *Drosophila*. O erro padrão está apresentado após cada estimativa.

	Subfamília V	Subfamília W
Subfamília V	1,00 ± 0,20 (2,02 ± 0,067)	
Subfamília W	21,6 ± 1,90 (44,10 ± 4,70)	6,30 ± 0,70 (11,36 ± 1,86)

Considerando-se a taxa de 0,016 substituições por sítio por milhões de anos, a estimativa do tempo de divergência destas subfamílias é de  $\approx 9,9$  milhões de anos, enquanto a subfamília *W* teria iniciado a diversificação há  $\approx 2,9$  milhões de anos e a subfamília *V* há  $\approx 1$  milhão de anos (Material Suplementar, Figura 3).

### Divergência dos sítios sinônimos de *Galileo* e genes hospedeiros

Na ausência de um forte viés para o uso de códons, valores de dS fornecem uma boa medida para evolução neutra. De acordo com a disponibilidade de genes no banco de dados para as espécies em investigação, foi priorizado o uso do gene *Amd*, já que o mesmo não apresentou um considerável viés no uso de códons (Tabela 2) e foi apontado por Vidal *et al.* (2009) como o gene mais adequado para esta estimativa. Ele também foi utilizado em trabalhos anteriores com outros *TEs* e com algumas espécies hospedeiras em comum (Ludwig *et al.*, 2008; Wallau *et al.*, 2011; Loreto *et al.*, 2012). Para *D. capricorni*, especificamente, não há sequências do gene *Amd* disponível, dessa forma utilizou-se o gene *Adh* como melhor alternativa para as comparações. Apesar do gene *Adh* ter um maior viés no uso de códons, para as espécies em estudo, apresentou um *CBI* de valor moderado (Tabela 2), assim como observado para o transposon *Galileo* (Tabela 3).

Para inferir a ocorrência de *HT*, inicialmente considerou-se as incongruências observadas entre as relações filogenéticas de *Galileo* e as relações evolutivas das espécies dos subgrupos *willistoni* e *bocainensis*. Além de não haver a recuperação das relações evolutivas entre as espécies do grupo *willistoni* (Figura 2; Material Suplementar, Figuras 1, 2 e 3), observou-se



que a distribuição de *Galileo* é descontínua no subgrupo *bocainensis*, pois entre as espécies investigadas, *D. succinea* e *D. fumipennis* não apresentaram sequências correspondentes ao elemento *Galileo*. As espécies *D. neocardini* e *D. neomorpha* (subgrupo *cardini*) pertencem ao subgênero *Drosophila*, mas quanto à análise de *Galileo*, apresentam cópias da subfamília *V* intimamente relacionadas às demais cópias caracterizadas no grupo *willistoni* do subgênero *Sophophora*.

Além das incongruências filogenéticas e descontinuidade, baixos valores de dS para *Galileo* foram registrados (Material Suplementar, Tabela 3), quando comparados aos do gene *Amd* e/ou aos do gene *Adh* (Material Suplementar, Tabelas 4 e 5) das espécies de *Drosophila* em estudo. Algumas dessas comparações foram estatisticamente significantes e, conjuntamente com os demais achados, fornecem evidências de *HT* de sequências correspondentes a ambas as subfamílias, *V* e *W*, do subgrupo *willistoni* para o subgrupo *bocainensis*.

Nos casos onde duas ou mais espécies estavam supostamente envolvidas em um mesmo evento de *HT*, considerou-se as espécies que apresentavam as sequências nucleotídica de *Galileo* mais similares. Frente a tal critério, apontou-se para a subfamília *W* que *D. willistoni* pode ter sido a espécie doadora para *D. nebulosa* (dS *Galileo* = 0,028, dS *Amd* = 0,713,  $p = 0,000000$ ) e para *D. capricorni* (dS *Galileo* = 0,000, dS *Adh* = 0,070,  $p = 0,015664$ ). Quanto à existência da subfamília *V* no subgrupo *bocainensis*, *D. paulistorum* (Interior) pode ter doado para *D. nebulosa* (dS *Galileo* = 0,000, dS *Adh* = 0,058,  $p = 0,035261$ ) e *D. willistoni* para *D. capricorni* (dS *Galileo* = 0,000, dS *Adh* = 0,070,  $p = 0,018469$ ).

Comparações realizadas com o gene *Amd* forneceram muitos sinais de *HT* estatisticamente significantes. Dessa forma, não foi possível inferir como *Galileo*, especificamente a subfamília *V*, foi transferido horizontalmente para as espécies *D. neocardini* e *D. neomorpha* (subgrupo *cardini*). A ausência de genes em comum (em banco de dados) para os grupos *cardini* e *willistoni* impossibilitou qualquer análise ou inferência para o fato de espécies de dois subgêneros compartilharem com tal similaridade sequências da subfamília *V*.

Tabela 2. Índice de viés de códon (*CBI*) e número efetivo de códons (*ENC*) para os genes *Amd* e *Adh* em espécies de *Drosophila*. *CBI* varia de 0 (uso uniforme de códons sinônimos) a 1 (viés máximo) e *ENC* varia de 21 (máximo de viés) e 61 (mínimo de viés).

Espécie (subespécie)	<i>Amd</i>		<i>Adh</i>	
	<i>CBI</i>	<i>ENC</i>	<i>CBI</i>	<i>ENC</i>
<i>D. paulistorum</i> (Amazônica)	0,381	52,004	0,585	43,645
<i>D. paulistorum</i> (Orinocana)	0,367	52,269	0,589	43,287
<i>D. paulistorum</i> (Andino Brasileira)	0,366	52,478	0,583	43,314
<i>D. paulistorum</i> (Interior)	0,355	52,980	0,571	44,827
<i>D. paulistorum</i> (Centro Americana)	0,361	52,128	0,576	40,674
<i>D. paulistorum</i> (Transicional)	0,343	53,178	0,589	43,287
<i>D. equinoxialis</i>	0,359	52,282	0,572	46,120
<i>D. willistoni</i>	0,445	46,235	0,631	41,231
<i>D. tropicalis</i>	0,352	52,865	0,572	46,120
<i>D. insularis</i>	0,377	53,094	0,606	43,557
<i>D. nebulosa</i>	0,351	55,445	0,560	43,438
<i>D. capricorni</i>	–	–	0,488	49,001
<i>D. neomorpha</i>	0,454	47,113	–	–
<i>D. neocardini</i>	0,436	49,961	–	–
Média	0,381	51,695	0,577	44,137
Mínimo	0,343	46,235	0,488	40,674
Máximo	0,454	55,445	0,631	49,539
Desvio padrão	0,038	2,527	0,021	2,678
Variância	0,001	6,385	0,004	7,171

Tabela 3. Índice de viés de códon (*CBI*) e número efetivo de códons (*ENC*) para subfamílias *V* e *W* de *Galileo* isoladas. Sequências consensos foram utilizadas. *CBI* varia de 0 (uso uniforme de códons sinônimos) a 1 (viés máximo) e *ENC* varia de 21 (máximo de viés) e 61 (mínimo de viés).

Sufamílias de <i>Galileo</i>	<i>CBI</i>	<i>ENC</i>
Subfamília <i>W</i>	0,490	54,497
Subfamília <i>V</i>	0,417	50,531
Subfamília <i>V</i> ( <i>cardini</i> )	0,412	52,771
Média	0,440	52,600
Mínimo	0,412	50,531
Máximo	0,490	54,497
Desvio padrão	0,044	1,989
Variância	0,002	5,157

## Discussão

A análise evolutiva das subfamílias *V* e *W* de *Galileo* realizada neste trabalho e comparada com as relações filogenéticas das espécies do grupo *willistoni* de *Drosophila* apresentadas por Robe *et al.* (2010), sugere que estas sequências tiveram origem em um ancestral do subgrupo *willistoni* há aproximadamente 9,9 milhões de anos, após a divisão do grupo *willistoni* nos subgrupos das espécies crípticas e *bocainensis* (há  $\approx 15$  mya). Dentro do subgrupo *willistoni Galileo* teria sido mantido por transmissão vertical e a subfamília *W* teria iniciado a diversificação há  $\approx 2,9$  milhões de anos.

Durante a evolução de *TEs* alguns processos, como polimorfismo ancestral, seleção natural e transferência horizontal de sequências entre espécies, podem produzir incongruências filogenéticas (Dias e Carareto, 2012; Wallau *et al.*, 2012). Neste trabalho o cenário das relações evolutivas de *Galileo* é demonstrado por meio de incongruências filogenéticas entre as espécies do grupo *willistoni* e espécies do subgrupo *cardini*. As incongruências para o elemento *Galileo* e a alta conservação de suas sequências, poderiam ser elucidadas se este transposon estivesse sob intensa pressão seletiva, contudo, quando a hipótese de evolução neutra foi rejeitada, tínhamos uma relação de  $dN < dS$ , indicando seleção purificadora.

Uma das metodologias mais promissoras para a detecção de *HT*, segundo Wallau *et al.* (2012), é baseada na comparação entre as espécies, da taxa de evolução neutra (por meio de divergência de substituições sinônimas) para o *TEs* e genes do hospedeiro (*HGs*, do inglês *Host Genes*). Esta abordagem assume que, se os *TEs* foram transmitidos verticalmente e mantidos por meio de processos evolutivos neutros nos genomas de duas espécies diferentes, desde o seu último antepassado comum, o número de substituições sinônimas por sítio sinónimo ( $dS$ ) dos *TEs* deve ser igual ou maior do que a dos *HGs*. No entanto, se o  $dS$  obtido para o *TE* é significativamente menor que o  $dS$  do *HG*, a explicação mais provável é que esses elementos foram transferidos horizontalmente entre as espécies após o seu isolamento reprodutivo. Esse padrão pode ser observado porque o *TE* adquirido via *HT* passou menos tempo no genoma do hospedeiro para

acumular substituições sinônimas que o HGs, oriundos do último ancestral comum das espécies envolvidas no evento de HT. Com base em valores de dS significativamente menores, comparados aos valores para os genes *Amd* e *Adh*, observa-se que *Galileo* não estaria presente desde o ancestral comum dos subgrupos *willistoni* e *bocainensis*. Nossos dados sugerem que *D. willistoni* pode ter transferido horizontalmente para as espécies *D. capricorni* e *D. nebulosa* *Galileo* da subfamília *W*, a qual foi demonstrada por Gonçalves *et al.* (2014) e, neste trabalho, ser a família mais antiga no genoma das espécies crípticas.

Dentro da subfamília *V* de *Galileo* as estimativas estaticamente significantes, seguidas da avaliação da menor distância genética entre as sequências, sugere novamente *D. willistoni* como potencial doadora do elemento para *D. capricorni*. O fato de *D. nebulosa* possuir cópias dessa mesma subfamília de *Galileo*, é possivelmente explicado pela transferência horizontal a partir da semiespécie Interior de *D. paulistorum*. As espécies do subgrupo *cardini*, *D. neocardini* e *D. neomorpha* apresentaram apenas sequências da subfamília *V*, com baixíssimas divergências em nível de nucleotídeos (0-2,3%) e de aminoácidos (0-4,8%). Não foi possível inferir uma potencial espécie doadora de *Galileo* para estas espécies de *cardini* devido à superestimativa de eventos de HT. Contudo, pela ausência de divergência entre suas sequências e espécies distantemente relacionadas do grupo *willistoni* e, considerando que cópias da subfamília *V* de *Galileo* diversificaram há 1 milhão de anos, é sugestivo que os eventos de HT teriam ocorrido recentemente entre espécies do grupo *willistoni* e do subgrupo *cardini*.

A premissa de que as espécies devem apresentar sobreposição geográfica para que ocorra HT é suportada nos eventos inferidos entre *D. willistoni* e as duas espécies do subgrupo *bocainensis* (*D. capricorni* e *D. nebulosa*), assim como entre a semiespécie Interior de *D. paulistorum* e *D. nebulosa*. Além disso, espécies do subgrupo *cardini* também são ecologicamente similares às espécies do grupo *willistoni*, utilizando os mesmos recursos na natureza (De Toni *et al.*, 2007; Gottschalk *et al.*, 2008; Schmitz *et al.*, 2007).

As relações evolutivas das sequências de *Galileo* estabelecidas sugerem que este transposon vem sendo mantido nos grupo *willistoni* e no subgrupo

*cardini* de *Drosophila* por meio de transferência vertical e horizontal (uma vez que ambas não são mutuamente exclusivas) e possivelmente por eventos de introgressão, especialmente entre espécies do subgrupo *willistoni*. Cruzamentos interespecíficos no subgrupo *willistoni* podem ter sido responsáveis pela transferência das sequências de *Galileo*, considerando-se a ausência de divergência nucleotídica entre sequências da superfamília *V* de *Galileo* nessas espécies. A ocorrência de cruzamentos entre espécies do subgrupo *willistoni* tem sido bem documentada (Winge e Cordeiro, 1960, 1963; Cordeiro, 1968; Cordeiro e Winge, 1995; Valiati, 1999) e entre semiespécies de *D. paulistorum* (Ehrman, 1961; Ehrman e Powell, 1982), assim como a introgressão também foi bem fundamentada em estudos recentes com marcadores moleculares para os subgrupos *willistoni* (Gleason *et al.*, 1998; Robe *et al.*, 2010) e *cardini* de *Drosophila* (Cenzi de Ré *et al.*, 2010). A introgressão foi anteriormente proposta por Silva e Kidwell (2000) como mecanismo de transferência do elemento *P*, relacionado à *Galileo* (Marzo *et al.*, 2008) entre espécies do subgrupo críptico *willistoni*. Já o fato da semiespécie Amazônica de *D. paulistorum* não possuir cópias que representam a subfamília *W* de *Galileo* pode ser resultado de perda estocástica ou de divergência dessas sequências durante o dinamismo deste *TE* ou processo evolutivo da semiespécie.

## Referências bibliográficas

Akaike H 1974. A new look at the statistical model identification. IEEE Trans Autom Control 19:716-723.

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

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

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.

- Cáceres M, Ranz JM, Barbadilla A, Long M, Ruiz A 1999. Generation of a widespread *Drosophila* inversion by a transposable element. *Science* 285: 415-418.
- Cáceres M, Puig M and Ruiz A 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353-1364.
- Casals F, Cáceres M, Ruiz A 2003. The foldback-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol.* 20: 674-685.
- Casals F, Cáceres M, Manfrin MH, González J and Ruiz A 2005. Molecular Characterization and chromosomal distribution of *Galileo*, *Kepler* and *Newton*, three *Foldback* transposable elements of the *Drosophila buzzatii* species complex. *Genetics* 169:2047-2059.
- Capy P, Bazin C, Higuete D and Langin T 1998. *Dynamics and Evolution of Transposable Elements*. 1st edition. Landes Bioscience, France, 197p.
- Clark *et al.* 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 480:2003-218.
- Cordeiro AR 1968. Chromosomal pairing variability of interespecific hybrids of *Drosophila willistoni* cryptic group. *Proc Int Congr Genetics, Tokyo*, 1:191.
- Cordeiro AR, Winge H 1995. Levels of evolutionary divergence of *Drosophila willistoni* sibling species. In: Levine L (Ed) *Genetics of natural populations: the continuing importance of Theodosius Dobzhansky*. Columbia University Press, New York, pp 262-280.
- Cummings MP 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. *Trends Ecol Evol.* 9:141-145.
- Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics* 124, 339-355.
- Delprat A, Negre B, Puig M, Ruiz A 2009. The transposon *Galileo* generates natural chromosomal inversions in *Drosophila* by ectopic recombination. *PloS One* 4: 7883.
- De Toni DC, Gottschalk MS, Cordeiro J, Hofmann PRP, Valente VLS 2007. Study of the *Drosophilidae* (Diptera) communities on Atlantic Forest islands of Santa Catarina State, Brazil. *Neotrop Entomol* 36:356-375.

Dias ES, Carareto CMA 2012. Ancestral polymorphism and recent invasion of transposable elements in *Drosophila* species. *BMC Evol. Biol.* 12:119.

Drummond AJ *et al.* 2010. Geneious v6. Available at <http://www.geneious.com>

Ehrman L and Powell JR 1982. The *Drosophila willistoni* species group. In: The genetics and biology of *Drosophila*. p 193-225. Academic Press, New York.

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

Gottschalk MS, Hofmann PRP and Valente VLS 2008. Diptera , Drosophilidae : historical occurrence in Brazil. *Check List.* 4:485-518.

Gonçalves JW, Valiati VH, Delprat A, Valente VL, Ruiz A 2014. Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. *BMC Genomics* 15:792.

Guindon S, Gascuel O 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696-704.

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

Jukes TH and Cantor CR 1969. Evolution of protein molecules. In Munro HN, editor, *Mammalian Protein Metabolism*, pp. 21-132, Academic Press, New York.

Kapitonov V, Jurka J 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U. S. A.* 100:6569-74.

Katoh K, Misawa K, Kuma K, Miyata T 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059-3066.

Loreto EL, Valente VL, Zaha A, Silva JC, Kidwell MG 2001. *Drosophila mediopunctata* P elements: a new example of horizontal transfer. *J Hered.* 92: 375-381.

Loreto ELS, Zambra FMB, Ortiz MF, Robe LJ 2012. New *Drosophila* P-like elements and reclassification of *Drosophila* P-elements subfamilies. *Mol Genet Genomics.* 287:531-540.

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



- Marques EK, Napp M, Winge H and Cordeiro AR 1966. A corn meal, soybean flour, wheat germ medium for *Drosophila*. DIS 41:187.
- Marzo M, Bello X, Puig M, Maside X, Ruiz A 2013. Striking structural dynamism and nucleotide sequence variation of the transposon *Galileo* in the genome of *Drosophila mojavensis*. Mob. DNA. 4:1.
- Marzo M, Puig M, Ruiz A 2008. The *Foldback*-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. Proc. Natl. Acad. Sci. U. S. A. 105:2957-2962.
- Mortan BR 1993. Chloroplast DNA codon use: evidence for selection at the *psb a* locus based on tRNA availability. J. Mol. Evol. 37, 273-280.
- Nei M, Gojobori T 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-426.
- Posada C, Crandall KA 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.
- Pritham EJ 2009. Transposable elements and factors influencing their success in eukaryotes. J Hered. 100: 648-655.
- Robe LJ, Cordeiro J, Loreto ELS, Valente VLS 2010. Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera : Drosophilidae ).
- Robertson HM 2002. Evolution of DNA transposons in eukaryotes. In Mobile DNA II, pp 1093-1110. ASM Press, Washington, DC.
- Robertson, H. M., and D. J. Lampe. 1995. Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol. 12:850-862.
- Rodríguez FJ, Oliver JL, Marín A, Medina JR 1990. The general stochastic model of nucleotide substitution. J Theor Biol. 142:485-501.
- Ronquist F, Huelsenbeck JP: MRBAYES 3 2003. Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
- Sassi AK, Herédia F, Lucio É, Lucia V and Rohde C 2005. Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. Genetics Mol Biol. 28: 734-739.

Schaack S, Gilbert C, Feschotte C 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol.* 25:537-46.

Schmitz HJ, Valente VLS and Hofmann PRP 2007. Taxonomic Survey of Drosophilidae (Diptera) from Mangrove Forests of Santa Catarina Island, Southern Brazil. *Systematics, Morphology and Physiology.* 53-64.

Sharp PM, Li WH 1989. On the rate of DNA sequences evolution in *Drosophila*. *J Mol Evol.* 28:398-402.

Silva JC, Loreto EL and Clark JB 2004. Factors that affect the horizontal transfer of transposables elements. *Curr Issues Mol Biol* 6:57-71.

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

Swofford DL 2002. PAUP: Phylogenetic Analysis Using Parsimony (and other methods), Version 4b10 Sinauer Associates, Sunderland MA.

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.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731-2739.

Terzian C, Ferraz C, Demaille J, Bucheton A 2000. Evolution of the *Gypsy* endogenous retrovirus in the *Drosophila melanogaster* subgroup. *Mol Biol Evol* 17:908-914.

Valiati VH 1999. Genética de Populações Marginais de *Drosophila willistoni* e *D. paulistorum* e Avaliação de Suas Potencialidades para Produção de Híbridos Interespecíficos. PhD thesis. Universidade Federal do Rio Grande do Sul.

Wallau GL, Kaminski VL, Loreto ELS. 2011. The role of vertical and horizontal transfer in the evolution of *Paris*-like elements in drosophilid species. *Genetica.* 139:1487-97.

Wallau GL, Ortiz MF, Loreto ELS 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol. Evol.* 4:689-99.

Wicker T et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973-82.

Winge H and Cordeiro AR 1960. Descoberta de uma ponte biológica entre *Drosophila willistoni* e *paulistorum* (Summ). *Ciência e Cultura* 12:149.

Winge H and Cordeiro AR 1963. Experimental hybrids between *Drosophila willistoni* Stutervant and *Drosophila paulistorum* Dobzhansky and Pavan from Southern marginal populations. *Heredity* 18:215-222.

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

**MATERIAL SUPPLEMENTAR**

---

Tabela 1. Distância genética das subfamílias de *Galileo* para cada espécie de *Drosophila* estudada em nível de nucleotídeos (abaixo da diagonal) e de aminoácidos (acima da diagonal).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
1 DTP_Galileo_BK006363		0,515	0,524	0,505	0,544	0,515	0,602	0,602	0,608	0,602	0,622	0,610	0,624	0,624	0,634	0,640	0,520	0,618	0,569	0,573	0,570	0,577	0,578	0,570	0,592	0,576	0,598	0,573	0,594	0,612	0,602	0,612	
2 DTP_Galileo_BK006357	0,300		0,330	0,583	0,592	0,583	0,612	0,612	0,608	0,602	0,622	0,580	0,564	0,564	0,574	0,600	0,540	0,647	0,598	0,602	0,613	0,598	0,608	0,600	0,612	0,606	0,608	0,612	0,625	0,612	0,612	0,621	
3 DTP_Galileo_EU334685	0,302	0,186		0,553	0,573	0,553	0,583	0,583	0,588	0,573	0,582	0,570	0,554	0,554	0,564	0,590	0,520	0,627	0,569	0,573	0,581	0,557	0,578	0,570	0,583	0,576	0,578	0,583	0,583	0,583	0,583	0,592	
4 DTP_Galileo_BK006361	0,305	0,344	0,333		0,113	0,505	0,602	0,602	0,608	0,592	0,602	0,600	0,594	0,594	0,604	0,620	0,540	0,627	0,588	0,592	0,602	0,588	0,608	0,590	0,612	0,596	0,608	0,602	0,604	0,602	0,612	0,612	
5 DTP_Galileo_BK006362	0,331	0,362	0,349	0,056		0,524	0,621	0,621	0,627	0,612	0,622	0,610	0,604	0,604	0,614	0,630	0,560	0,647	0,598	0,602	0,602	0,588	0,608	0,590	0,612	0,596	0,608	0,612	0,604	0,612	0,612	0,612	
6 DTP_Galileo_BK006359	0,333	0,362	0,370	0,362	0,380		0,438	0,438	0,442	0,457	0,460	0,412	0,437	0,437	0,437	0,451	0,365	0,471	0,452	0,457	0,457	0,459	0,471	0,461	0,476	0,475	0,471	0,467	0,469	0,486	0,486	0,495	
7 DTP_GalileoW_BK006360	0,333	0,372	0,372	0,388	0,408	0,279		0,000	0,000	0,029	0,020	0,137	0,146	0,136	0,155	0,147	0,154	0,202	0,433	0,438	0,436	0,439	0,433	0,412	0,429	0,416	0,423	0,448	0,429	0,438	0,429	0,448	
8 DTP_GalileoW_Dwilcons	0,333	0,372	0,372	0,388	0,408	0,279	0,000		0,000	0,029	0,020	0,137	0,146	0,136	0,155	0,147	0,154	0,202	0,433	0,438	0,436	0,439	0,433	0,412	0,429	0,416	0,423	0,448	0,429	0,438	0,429	0,448	
9 DTP_GalileoW_Dcapcons	0,334	0,370	0,373	0,389	0,409	0,280	0,000	0,000		0,029	0,020	0,139	0,147	0,137	0,157	0,149	0,157	0,204	0,437	0,442	0,441	0,443	0,437	0,416	0,433	0,420	0,427	0,452	0,433	0,442	0,433	0,452	
10 DTP_GalileoW_Dtrocons	0,333	0,370	0,359	0,382	0,408	0,287	0,018	0,018	0,018		0,030	0,137	0,155	0,146	0,165	0,157	0,154	0,212	0,452	0,457	0,447	0,459	0,452	0,431	0,448	0,436	0,442	0,467	0,449	0,457	0,448	0,467	
11 DTP_GalileoW_Dnebcons	0,349	0,384	0,378	0,397	0,413	0,294	0,018	0,018	0,018	0,026		0,144	0,143	0,133	0,153	0,134	0,160	0,202	0,455	0,460	0,449	0,457	0,455	0,433	0,450	0,438	0,444	0,470	0,447	0,450	0,450	0,460	
12 DTP_GalileoW_Dequcons	0,356	0,369	0,369	0,393	0,414	0,293	0,071	0,071	0,071	0,076	0,078		0,070	0,060	0,050	0,071	0,115	0,149	0,455	0,461	0,451	0,463	0,455	0,434	0,451	0,439	0,446	0,471	0,442	0,461	0,451	0,471	
13 DTP_GalileoW_andcons	0,363	0,360	0,358	0,379	0,397	0,288	0,075	0,075	0,075	0,078	0,082	0,053		0,009	0,028	0,039	0,115	0,127	0,441	0,447	0,424	0,427	0,441	0,420	0,437	0,424	0,431	0,456	0,438	0,447	0,437	0,456	
14 DTP_GalileoW_intcons	0,360	0,358	0,355	0,376	0,394	0,285	0,073	0,073	0,073	0,075	0,080	0,050	0,003		0,019	0,029	0,115	0,118	0,441	0,447	0,424	0,427	0,441	0,420	0,437	0,424	0,431	0,456	0,438	0,447	0,437	0,456	
15 DTP_GalileoW_oricons	0,363	0,360	0,358	0,381	0,399	0,285	0,078	0,078	0,078	0,080	0,085	0,048	0,008	0,005		0,029	0,115	0,137	0,451	0,456	0,435	0,438	0,451	0,430	0,447	0,434	0,441	0,466	0,448	0,456	0,447	0,466	
16 DTP_GalileoW_tracons	0,365	0,367	0,370	0,383	0,402	0,289	0,076	0,076	0,076	0,078	0,083	0,053	0,016	0,013	0,013		0,118	0,129	0,446	0,451	0,435	0,432	0,446	0,424	0,441	0,429	0,436	0,461	0,442	0,451	0,441	0,461	
17 DTP_GalileoW_cencons	0,344	0,378	0,394	0,372	0,394	0,339	0,098	0,098	0,099	0,109	0,101	0,077	0,066	0,066	0,066	0,072		0,118	0,392	0,404	0,372	0,383	0,373	0,385	0,385	0,380	0,385	0,423	0,383	0,385	0,385	0,404	
18 DTP_GalileoW_ins	0,381	0,394	0,389	0,407	0,425	0,308	0,090	0,090	0,090	0,098	0,092	0,073	0,070	0,068	0,073	0,065	0,093		0,476	0,481	0,462	0,474	0,476	0,455	0,471	0,460	0,466	0,481	0,464	0,481	0,471	0,490	
19 DTP_GalileoV_amacons	0,339	0,389	0,381	0,394	0,407	0,290	0,216	0,216	0,216	0,226	0,226	0,223	0,210	0,208	0,210	0,209	0,236	0,237		0,000	0,000	0,000	0,010	0,010	0,019	0,010	0,019	0,010	0,021	0,048	0,038	0,048	
20 DTP_GalileoV_andcons	0,341	0,390	0,382	0,395	0,408	0,292	0,218	0,218	0,219	0,228	0,228	0,225	0,212	0,210	0,212	0,211	0,240	0,239	0,000		0,000	0,000	0,010	0,010	0,019	0,020	0,019	0,010	0,020	0,048	0,038	0,048	
21 DTP_GalileoV_Dneomcons	0,338	0,383	0,375	0,394	0,408	0,287	0,205	0,205	0,205	0,213	0,213	0,209	0,196	0,194	0,196	0,194	0,215	0,224	0,000	0,000		0,000	0,000	0,000	0,011	0,011	0,011	0,011	0,011	0,043	0,032	0,043	
22 DTP_GalileoV_intcons	0,346	0,385	0,383	0,396	0,409	0,291	0,209	0,209	0,210	0,220	0,222	0,219	0,206	0,204	0,206	0,205	0,230	0,234	0,000	0,000	0,000		0,000	0,000	0,010	0,011	0,010	0,010	0,010	0,011	0,041	0,020	0,041
23 DTP_GalileoV_cencons	0,345	0,389	0,386	0,394	0,407	0,293	0,213	0,213	0,214	0,224	0,224	0,220	0,208	0,205	0,208	0,206	0,225	0,235	0,005	0,005	0,000	0,000		0,000	0,010	0,010	0,010	0,019	0,010	0,038	0,029	0,038	
24 DTP_GalileoV_Dequcons	0,347	0,386	0,381	0,392	0,405	0,293	0,210	0,210	0,210	0,220	0,220	0,217	0,204	0,202	0,204	0,203	0,230	0,231	0,008	0,008	0,000	0,000	0,003		0,010	0,010	0,010	0,020	0,011	0,039	0,029	0,039	
25 DTP_GalileoV_Dnebcons	0,349	0,388	0,385	0,398	0,411	0,295	0,215	0,215	0,216	0,226	0,226	0,223	0,210	0,207	0,210	0,208	0,230	0,237	0,008	0,008	0,003	0,003	0,003	0,005		0,010	0,000	0,029	0,000	0,029	0,019	0,029	
26 DTP_GalileoV_tracons	0,340	0,382	0,380	0,390	0,403	0,291	0,208	0,208	0,208	0,218	0,218	0,215	0,202	0,199	0,202	0,201	0,227	0,229	0,005	0,008	0,003	0,003	0,003	0,005	0,003		0,010	0,030	0,011	0,040	0,030	0,040	
27 DTP_GalileoV_Dwilcons	0,345	0,382	0,379	0,392	0,408	0,289	0,209	0,209	0,209	0,219	0,222	0,216	0,203	0,201	0,203	0,202	0,230	0,230	0,010	0,010	0,003	0,003	0,005	0,005	0,003	0,003		0,029	0,000	0,019	0,010	0,029	
28 DTP_GalileoV_Doricons	0,341	0,393	0,385	0,398	0,411	0,295	0,221	0,221	0,221	0,231	0,231	0,228	0,215	0,212	0,215	0,214	0,246	0,242	0,003	0,003	0,003	0,003	0,008	0,010	0,010	0,010	0,013		0,031	0,057	0,048	0,057	
29 DTP_GalileoV_Dcapcons	0,340	0,377	0,377	0,388	0,406	0,280	0,204	0,204	0,205	0,215	0,217	0,209	0,196	0,193	0,196	0,194	0,215	0,223	0,013	0,013	0,005	0,005	0,008	0,008	0,005	0,005	0,000	0,016		0,000	0,010	0,020	
30 DTP_GalileoV_Dneoccons	0,349	0,382	0,385	0,393	0,411	0,292	0,215	0,215	0,216	0,226	0,226	0,223	0,210	0,207	0,210	0,208	0,235	0,237	0,028	0,028	0,021	0,021	0,023	0,023	0,021	0,021	0,013	0,031	0,000		0,038	0,038	
31 DTP_GalileoV_Dinscons	0,351	0,385	0,382	0,393	0,408	0,292	0,215	0,215	0,216	0,226	0,228	0,225	0,212	0,210	0,212	0,211	0,240	0,239	0,018	0,018	0,011	0,010	0,013	0,013	0,010	0,010	0,005	0,021	0,008	0,023		0,048	
32 DTP_GalileoV_Dtro	0,354	0,382	0,385	0,393	0,408	0,295	0,218	0,218	0,219	0,228	0,228	0,225	0,212	0,210	0,212	0,211	0,235	0,239	0,026	0,026	0,016	0,018	0,021	0,018	0,018	0,018	0,018	0,013	0,028	0,008	0,018	0,023	

Tabela 2. Lista de espécies, com respectivos códigos de acesso no *Genbank*.

Espécie (semiespécie)	<i>Amd</i>	<i>Adh</i>	Origem geográfica
<i>D. paulistorum</i> (Amazônica)	FJ664498	EU532122	Belém do Pará, PA, Brasil
<i>D. paulistorum</i> (Orinocana)	FJ664503	EU532127	Georgetown, Guiana
<i>D. paulistorum</i> (Andino Brasileira)	FJ664496	EU532129	Ribeirão Preto, SP, Brasil
<i>D. paulistorum</i> (Andino Brasileira)	FJ664497	–	Florianópolis, SC, Brasil
<i>D. paulistorum</i> (Interior)	FJ664502	EU532124	Llanos, Colômbia
<i>D. paulistorum</i> (Centro Americana)	FJ664500	EU532123	Lancetilla, Honduras
<i>D. paulistorum</i> (Transicional)	FJ664501	EU532130	Santa Marta, Colômbia
<i>D. equinoxialis</i>	FJ664506	U95269	Tefé, AM
<i>D. willistoni</i>	FJ664508	EU532119	Belém, PA, Brasil
<i>D. tropicalis</i>	FJ664504	EU532131	San Salvador, El Salvador
<i>D. insularis</i>	FJ664507	EU532120	Saint Kitts, Pequenas Antilhas
<i>D. nebulosa</i>	AF293717	U95275	–
<i>D. capricorni</i>	–	AY335196	–
<i>D. neomorpha</i>	HM006862	–	Joinville, SC, Brasil
<i>D. neocardini</i>	EU444581	–	Porto Alegre, RS, Brasil

Tabela 3. Valores de dS para sequências de *Galileo* em espécies de *Drosophila*. Acima da diagonal está o erro padrão estimado com 1000 réplicas tipo *bootstrap*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1 DTP_GalileoW_BK006360		0,000	0,000	0,026	0,020	0,050	0,058	0,058	0,058	0,051	0,059	0,091	0,223	0,212	0,206	0,174	0,211	0,203	0,206	0,203	0,186	0,193	0,199	0,200	0,207	0,183	
2 DTP_GalileoW_Dcapcons	0,000		0,000	0,027	0,020	0,051	0,058	0,058	0,059	0,052	0,059	0,093	0,212	0,202	0,208	0,166	0,201	0,194	0,197	0,205	0,178	0,181	0,189	0,191	0,196	0,175	
3 DTP_GalileoW_Dwilcons	0,000	0,000		0,026	0,020	0,050	0,058	0,058	0,058	0,051	0,059	0,091	0,223	0,212	0,206	0,174	0,211	0,203	0,206	0,203	0,186	0,193	0,199	0,200	0,207	0,183	
4 DTP_GalileoW_Dtrocons	0,054	0,054	0,054		0,028	0,057	0,059	0,059	0,059	0,052	0,066	0,111	0,240	0,225	0,210	0,190	0,223	0,213	0,217	0,202	0,189	0,195	0,212	0,216	0,210	0,195	
5 DTP_GalileoW_Dnebcons	0,028	0,028	0,028	0,057		0,050	0,059	0,059	0,059	0,051	0,059	0,089	0,230	0,217	0,217	0,183	0,215	0,206	0,209	0,215	0,195	0,196	0,213	0,218	0,218	0,194	
6 DTP_GalileoW_Dequcons	0,157	0,159	0,157	0,193	0,150		0,026	0,026	0,031	0,025	0,026	0,054	0,187	0,181	0,182	0,150	0,180	0,174	0,176	0,165	0,166	0,152	0,168	0,177	0,168	0,166	
7 DTP_GalileoW_andcons	0,197	0,199	0,197	0,199	0,191	0,043		0,000	0,000	0,020	0,039	0,062	0,173	0,182	0,176	0,138	0,181	0,175	0,177	0,164	0,148	0,151	0,147	0,158	0,162	0,147	
8 DTP_GalileoW_intcons	0,197	0,199	0,197	0,199	0,191	0,043	0,000		0,000	0,020	0,039	0,062	0,173	0,182	0,176	0,138	0,181	0,175	0,177	0,164	0,148	0,151	0,147	0,158	0,162	0,147	
9 DTP_GalileoW_oricons	0,198	0,200	0,198	0,199	0,192	0,059	0,000	0,000		0,020	0,040	0,062	0,175	0,185	0,179	0,140	0,184	0,178	0,180	0,166	0,150	0,153	0,149	0,160	0,164	0,149	
10 DTP_GalileoW_tracons	0,165	0,166	0,165	0,166	0,157	0,044	0,028	0,028	0,028		0,031	0,062	0,170	0,161	0,157	0,125	0,161	0,156	0,158	0,147	0,136	0,138	0,135	0,145	0,147	0,135	
11 DTP_GalileoW_cencons	0,194	0,196	0,194	0,231	0,188	0,043	0,087	0,087	0,087	0,057		0,071	0,195	0,189	0,190	0,157	0,189	0,180	0,184	0,173	0,160	0,161	0,162	0,173	0,177	0,158	
12 DTP_GalileoW_ins	0,225	0,230	0,225	0,308	0,201	0,095	0,118	0,118	0,118	0,119	0,144		0,260	0,249	0,268	0,211	0,246	0,238	0,250	0,238	0,238	0,245	0,249	0,270	0,254	0,245	
13 DTP_GalileoV_amacons	0,843	0,818	0,843	0,874	0,864	0,763	0,694	0,694	0,697	0,637	0,782	0,720		0,000	0,000	0,000	0,000	0,012	0,013	0,013	0,018	0,019	0,025	0,036	0,069	0,027	
14 DTP_GalileoV_andcons	0,827	0,802	0,827	0,857	0,846	0,748	0,682	0,682	0,684	0,626	0,768	0,694	0,000		0,000	0,000	0,000	0,012	0,012	0,013	0,018	0,018	0,024	0,036	0,066	0,027	
15 DTP_GalileoV_Dneomcons	0,772	0,746	0,772	0,800	0,808	0,727	0,658	0,658	0,661	0,600	0,748	0,698	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,015	0,032	0,072	0,020
16 DTP_GalileoV_intcons	0,675	0,651	0,675	0,710	0,693	0,604	0,545	0,545	0,547	0,488	0,626	0,570	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,016	0,033	0,068	0,021
17 DTP_GalileoV_cencons	0,824	0,799	0,824	0,853	0,842	0,745	0,679	0,679	0,682	0,623	0,768	0,689	0,000	0,000	0,000	0,000		0,012	0,012	0,013	0,018	0,018	0,024	0,035	0,066	0,027	
18 DTP_GalileoV_Dequcons	0,787	0,763	0,787	0,814	0,803	0,710	0,648	0,648	0,650	0,594	0,730	0,625	0,013	0,013	0,000	0,000	0,013		0,000	0,000	0,013	0,013	0,020	0,033	0,065	0,023	
19 DTP_GalileoV_Dnebcons	0,793	0,768	0,793	0,821	0,809	0,715	0,652	0,652	0,654	0,598	0,735	0,634	0,013	0,013	0,000	0,000	0,013	0,000		0,000	0,013	0,013	0,020	0,033	0,066	0,023	
20 DTP_GalileoV_tracons	0,758	0,734	0,758	0,785	0,812	0,680	0,619	0,619	0,621	0,565	0,701	0,657	0,013	0,013	0,000	0,000	0,013	0,000	0,000		0,000	0,013	0,015	0,030	0,065	0,019	
21 DTP_GalileoV_Dwilcons	0,713	0,690	0,713	0,738	0,760	0,638	0,580	0,580	0,582	0,529	0,658	0,625	0,026	0,026	0,000	0,000	0,026	0,013	0,013	0,000		0,013	0,000	0,022	0,063	0,013	
22 DTP_GalileoV_Doricons	0,715	0,692	0,715	0,741	0,763	0,640	0,582	0,582	0,584	0,531	0,660	0,635	0,026	0,026	0,000	0,000	0,026	0,013	0,013	0,013	0,013		0,020	0,034	0,065	0,024	
23 DTP_GalileoV_Dcapcons	0,725	0,700	0,725	0,752	0,777	0,621	0,554	0,554	0,555	0,501	0,635	0,595	0,042	0,042	0,015	0,015	0,042	0,027	0,028	0,014	0,000	0,028		0,000	0,075	0,000	
24 DTP_GalileoV_Dneoccons	0,773	0,749	0,773	0,800	0,827	0,697	0,635	0,635	0,638	0,582	0,717	0,730	0,095	0,093	0,070	0,073	0,093	0,079	0,079	0,068	0,040	0,081	0,000		0,093	0,022	
25 DTP_GalileoV_Dinscons	0,787	0,763	0,787	0,814	0,842	0,710	0,648	0,648	0,650	0,594	0,730	0,689	0,052	0,052	0,027	0,028	0,052	0,038	0,039	0,026	0,013	0,039	0,027	0,079		0,066	
26 DTP_GalileoV_Dtro	0,718	0,695	0,718	0,743	0,765	0,644	0,588	0,588	0,590	0,537	0,664	0,634	0,052	0,052	0,027	0,028	0,052	0,038	0,039	0,026	0,013	0,040	0,000	0,039	0,052		

Tabela 4. Valores de dS para sequências do gene *Amd* em espécies de *Drosophila*. Acima da diagonal está o erro padrão estimado com 1000 réplicas tipo *bootstrap*. <sup>a</sup> corresponde à Andino Brasileira de Ribeirão Preto (SP) <sup>e</sup> b de Florianópolis (SC).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 <i>D. paulistorum</i> (Amazônica)		0,015	0,013	0,015	0,015	0,015	0,014	0,019	0,028	0,040	0,046	0,090	0,693	0,649
2 <i>D. paulistorum</i> (Orinocana)	0,050		0,006	0,009	0,016	0,017	0,017	0,021	0,031	0,039	0,047	0,104	0,889	0,660
3 <i>D. paulistorum</i> (Andino Brasileira) <sup>a</sup>	0,039	0,010		0,006	0,014	0,015	0,015	0,020	0,030	0,038	0,046	0,100	0,861	0,639
4 <i>D. paulistorum</i> (Andino Brasileira) <sup>b</sup>	0,050	0,019	0,010		0,014	0,015	0,016	0,021	0,031	0,038	0,048	0,102	0,966	0,737
5 <i>D. paulistorum</i> (Interior)	0,046	0,052	0,041	0,041		0,005	0,014	0,020	0,032	0,038	0,050	0,099	0,754	0,603
6 <i>D. paulistorum</i> (Centro Americana)	0,050	0,060	0,050	0,050	0,005		0,013	0,020	0,031	0,038	0,048	0,094	0,985	0,711
7 <i>D. paulistorum</i> (Transicional)	0,045	0,055	0,045	0,055	0,041	0,034		0,022	0,032	0,038	0,050	0,096	0,997	0,649
8 <i>D. equinoxialis</i>	0,076	0,092	0,086	0,086	0,079	0,076	0,092		0,030	0,042	0,048	0,094	0,804	0,635
9 <i>D. willistoni</i>	0,141	0,165	0,159	0,165	0,166	0,159	0,171	0,152		0,040	0,042	0,104	0,748	0,452
10 <i>D. tropicalis</i>	0,241	0,234	0,227	0,227	0,224	0,227	0,227	0,247	0,226		0,047	0,109	0,798	0,530
11 <i>D. insularis</i>	0,297	0,311	0,304	0,318	0,328	0,326	0,334	0,311	0,277	0,314		0,097	0,771	0,773
12 <i>D. nebulosa</i>	0,639	0,720	0,695	0,708	0,647	0,650	0,675	0,659	0,713	0,728	0,674		1,101	1,055
13 <i>D. neomorpha</i>	1,877	2,094	2,056	2,219	1,920	2,241	2,264	1,979	1,911	1,968	1,910	2,506		0,059
14 <i>D. neocardini</i>	1,806	1,820	1,794	1,905	1,735	1,860	1,792	1,788	1,556	1,668	1,923	2,373	0,445	



Tabela 5. Valores de dS para sequências do gene *Adh* em espécies de *Drosophila*. Acima da diagonal está o erro padrão estimado com 1000 réplicas tipo *bootstrap*.

	1	2	3	4	5	6	7	8	9	10	11	12
1 <i>D. paulistorum</i> (Transicional)		0,030	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,010	0,000	0,025
2 <i>D. capricorni</i>	0,074		0,030	0,027	0,030	0,030	0,030	0,030	0,030	0,032	0,029	0,028
3 <i>D. paulistorum</i> (Amazônica)	0,000	0,074		0,010	0,000	0,000	0,000	0,000	0,000	0,010	0,000	0,025
4 <i>D. insularis</i>	0,010	0,063	0,010		0,010	0,010	0,011	0,010	0,010	0,015	0,011	0,027
5 <i>D. paulistorum</i> (Centro Americana)	0,000	0,074	0,000	0,010		0,000	0,000	0,000	0,000	0,010	0,000	0,025
6 <i>D. paulistorum</i> (Interior)	0,000	0,074	0,000	0,010	0,000		0,000	0,000	0,000	0,010	0,000	0,025
7 <i>D. paulistorum</i> (Andino Brasileira)	0,000	0,075	0,000	0,010	0,000	0,000		0,000	0,000	0,010	0,000	0,025
8 <i>D. paulistorum</i> (Orinocana)	0,000	0,074	0,000	0,010	0,000	0,000	0,000		0,000	0,010	0,000	0,025
9 <i>D. equinoxialis</i>	0,000	0,074	0,000	0,010	0,000	0,000	0,000	0,000		0,010	0,000	0,025
10 <i>D. tropicalis</i>	0,010	0,085	0,010	0,021	0,010	0,010	0,010	0,010	0,010		0,010	0,027
11 <i>D. willistoni</i>	0,000	0,070	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,010		0,025
12 <i>D. nebulosa</i>	0,058	0,073	0,058	0,068	0,058	0,058	0,058	0,058	0,058	0,069	0,053	

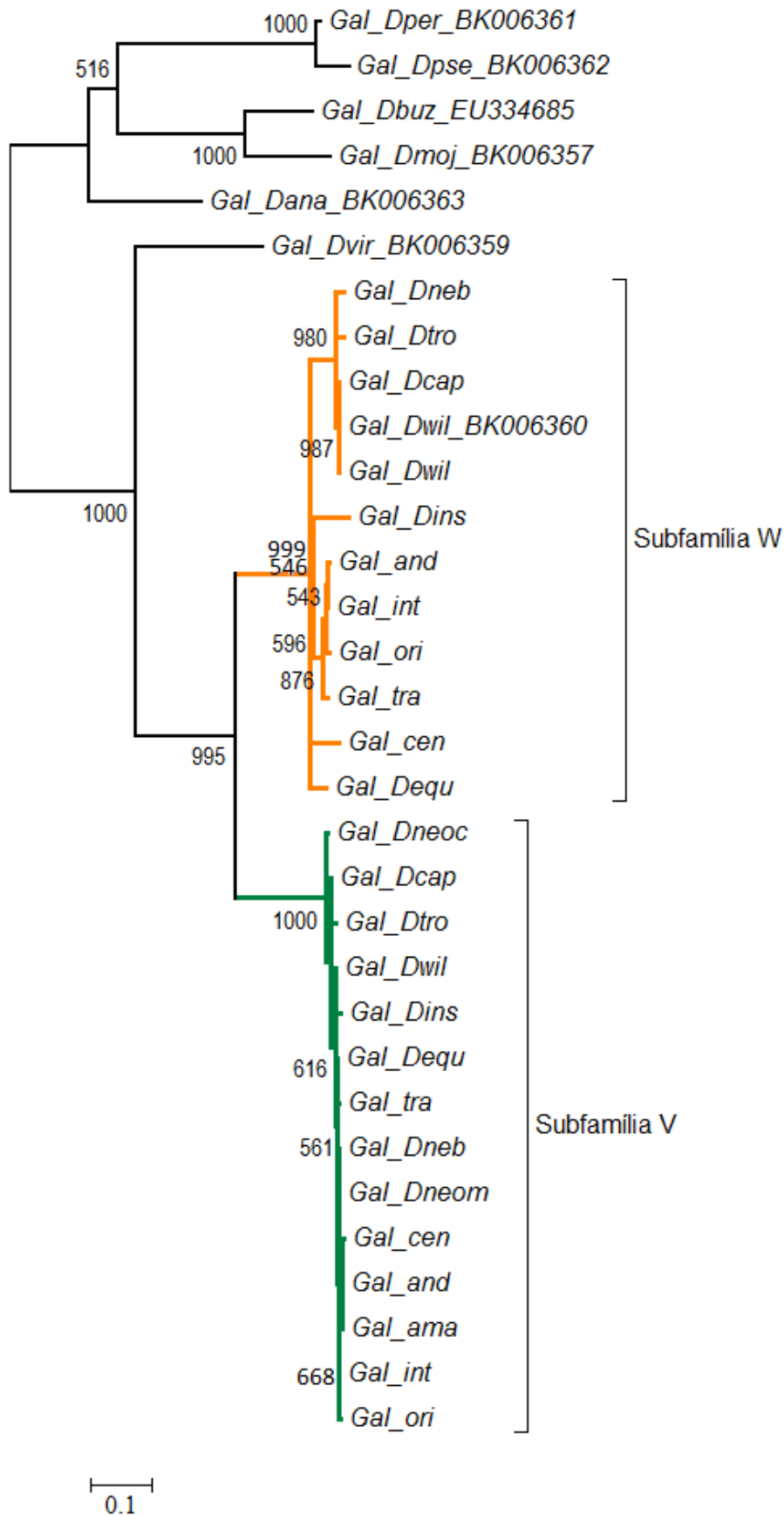
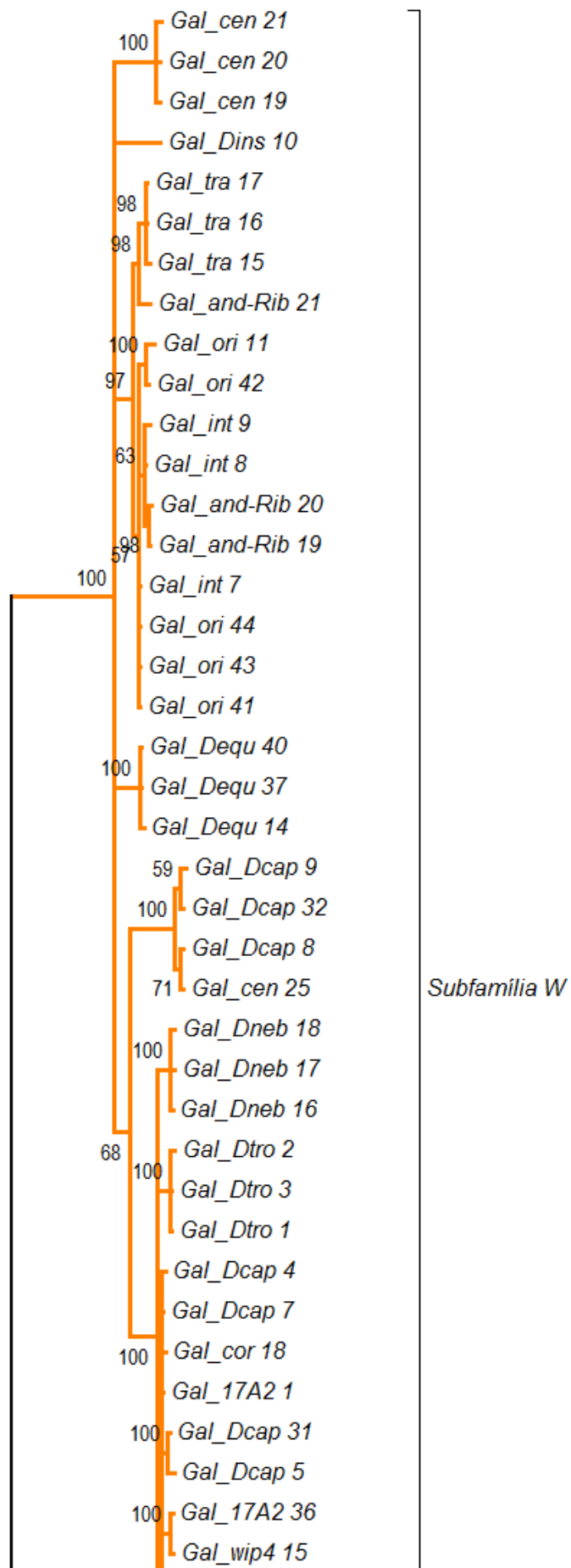
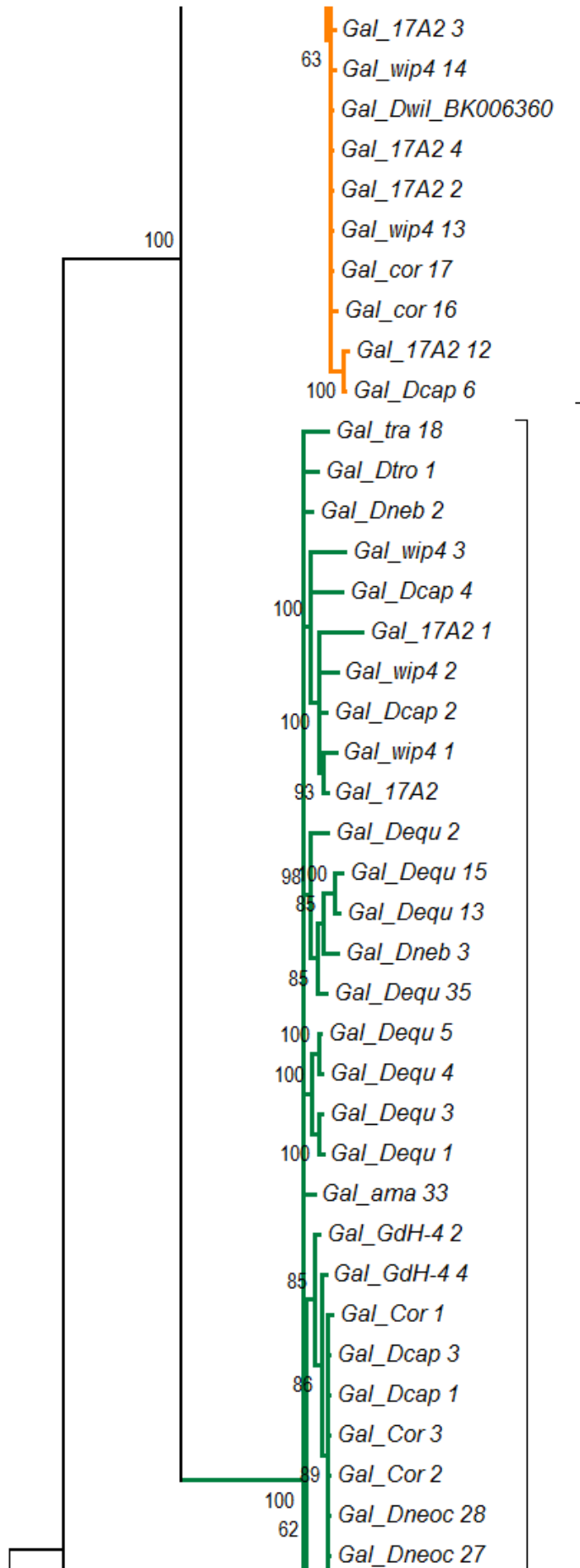


Figura 1. Árvore filogenética construída com seqüências consenso da inferida transposase de *Galileo* em *Drosophila* pela Máxima Verossimilhança. Valores de *bootstrap* estão indicados em cada nó (valores inferiores a 50% foram omitidos). As duas subfamílias, V e W, são apresentadas com o máximo de suporte. Seqüências de *D. ananassae*, *D. buzzatii*, *D. mojavensis*, *D. persimilis*, *D pseudoobscura* e *D. virilis* foram usadas como grupo externo.





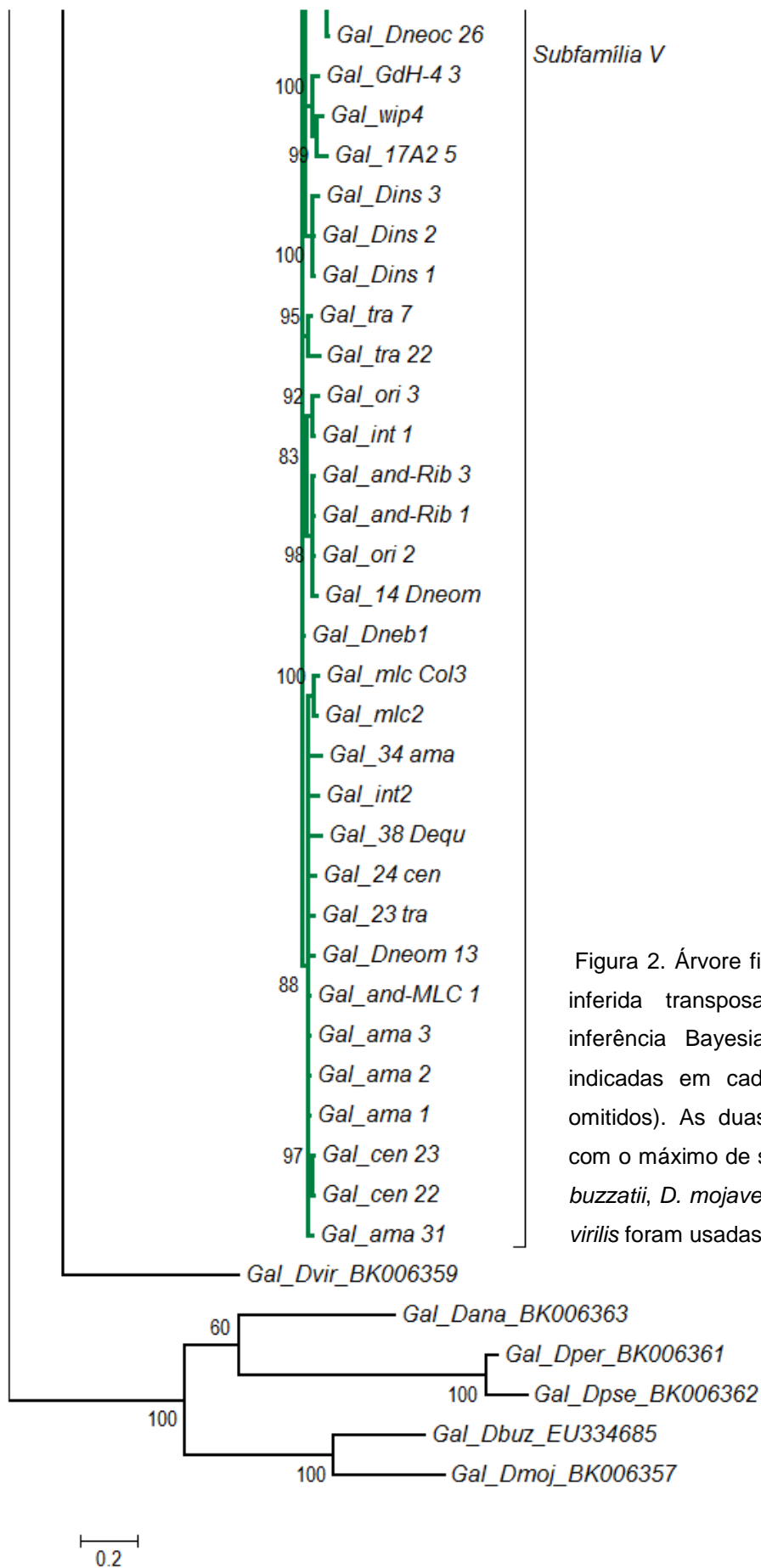
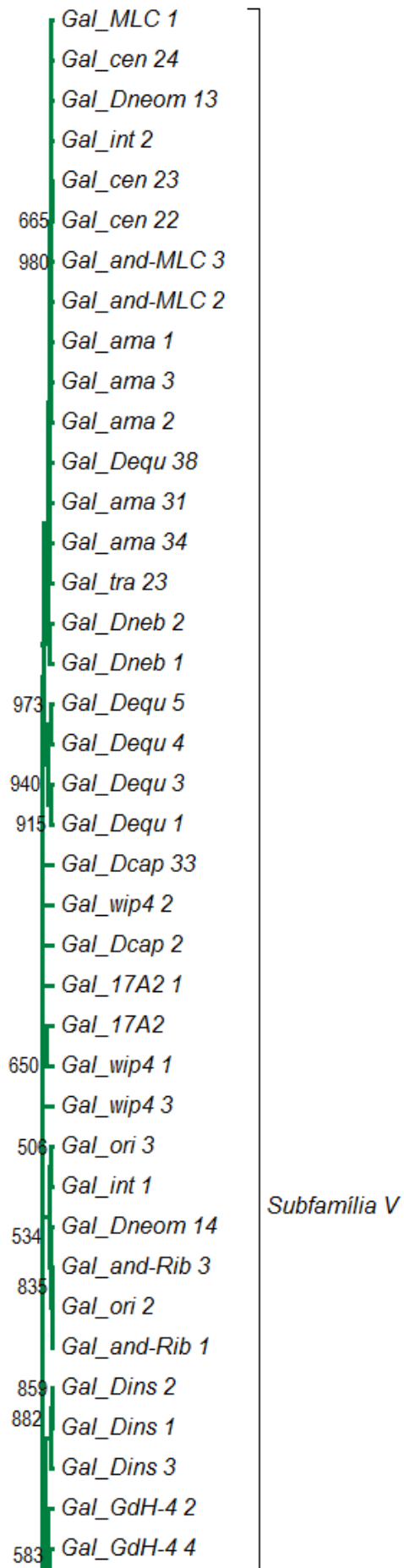
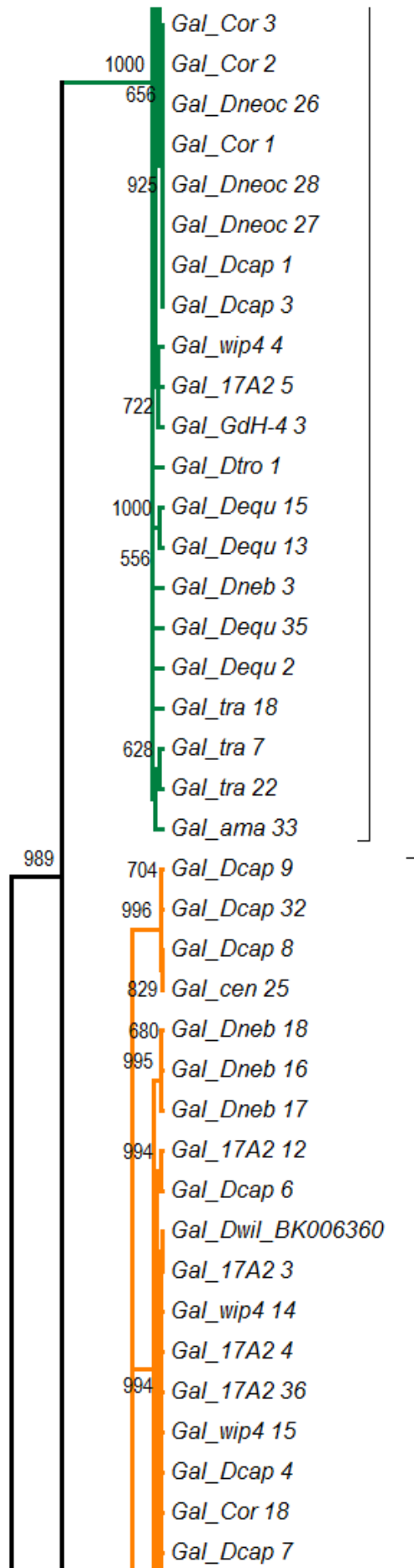


Figura 2. Árvore filogenética construída com segmentos da inferida transposase de *Galileo* em *Drosophila* pela inferência Bayesiana. Probabilidades *a posteriori* estão indicadas em cada nó (valores inferiores a 50% foram omitidos). As duas subfamílias, V e W, são apresentadas com o máximo de suporte. Sequências de *D. ananassae*, *D. buzzatii*, *D. mojavensis*, *D. persimilis*, *D. pseudoobscura* e *D. virilis* foram usadas como grupo externo.





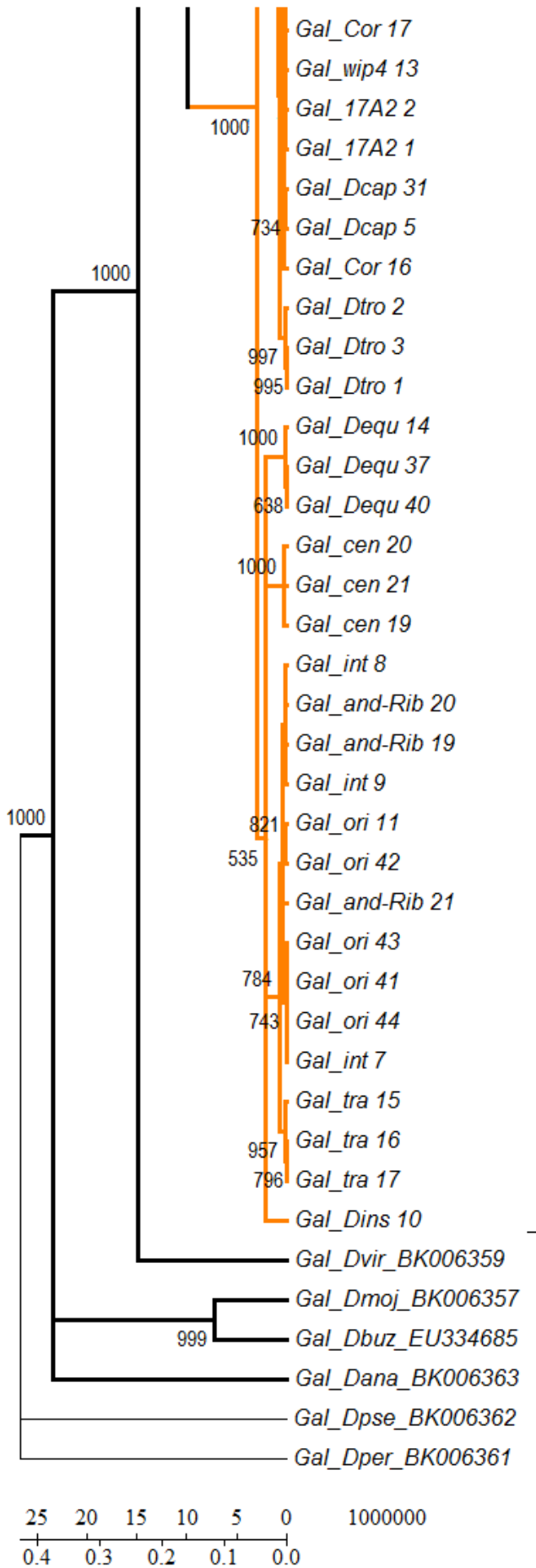


Figura 3. Árvore filogenética construída com segmentos da inserida transposase de *Galileo* em *Drosophila* pela Máxima Verossimilhança. Valores de *bootstrap* estão indicados em cada nó (valores inferiores a 50% foram omitidos). As duas subfamílias, V e W, são apresentadas com o máximo de suporte. Está representado do tempo de divergência das subfamílias, estimado com a taxa de 0,016 substituições por sítio por milhões. Sequências de *D. ananassae*, *D. buzzatii*, *D. mojavensis*, *D. persimilis*, *D pseudoobscura* e *D. virilis* foram usadas como grupo externo.



## CAPÍTULO IV

---

### DISCUSSÃO E CONCLUSÕES

Desde a sua identificação em *Drosophila buzzatii* (Cáceres *et al.*, 1999) e por meio da análise *in silico* realizada por Marzo *et al.* (2008) em mais seis espécies de *Drosophila*, *Galileo* foi caracterizado molecularmente na referida espécie (Cáceres *et al.*, 2001; Casals *et al.*, 2003; Casals *et al.*, 2005; Marzo *et al.*, 2008; Delprat *et al.*, 2009) e em *D. mojavensis* (Marzo *et al.*, 2013). O presente estudo foi o primeiro a caracterizar o transposon *Galileo* em espécies neotropicais do grupo *willistoni* e *cardini* de *Drosophila*.

Usando diferentes estratégias de bioinformática e por meio da anotação manual, caracterizou-se detalhadamente, 191 cópias de *Galileo* no genoma de *Drosophila willistoni* (Capítulo II desta tese). No trabalho de Marzo *et al.* (2008) um número limitado de cópias de *Galileo* foi isolado a partir do genoma de *D. willistoni*, não sendo detectados diferentes agrupamentos na análise filogenética das cópias. Com o aumento do número de cópias identificadas (de 28 para 191 cópias), a análise filogenética revelou dois clados bem suportados, demonstrando a existência de duas subfamílias de *Galileo* no genoma de *D. willistoni*, denominadas de *V* e *W* (Gonçalves *et al.*, 2014). *Galileo* aparenta ter uma dinâmica reestruturação das *TIRs* e, no genoma de *D. willistoni*, observou-se uma forma inédita de expansão das *TIRs*: a inserção de outro *TE* em uma das extremidades, que é transferido e passa a fazer parte de ambas as *TIRs*. Assim, os fragmentos de *TEs* não se constituem em inserções ocasionais em uma *TIR*, mas estão compondo as extremidades 5' e 3' de elementos flanqueados por *TSDs* idênticos, os quais delimitam as cópias em caracterização.

Embora não tenham sido identificadas cópias de *Galileo* com uma *ORF* potencialmente funcional (Gonçalves *et al.*, 2014 – Capítulo II), de acordo com Marzo *et al.* (2008), *Galileo* apresentava atividade no genoma de *D. willistoni* há menos de 1 milhão de anos. Alternativamente, esses elementos poderiam ser mobilizados através de proteínas produzidas por outros *TEs* ou processos celulares (Rebatchouk e Narita, 1997). Na caracterização e análise evolutiva das

sequências de *Galileo* no genoma de *D. willistoni*, observou-se a conservação de certas regiões das *TIRs*, especialmente as extremidades, o que poderia ser um indicativo do possível papel de reconhecimento e ligação da transposase. Sendo *Galileo* classificado como um transposon da superfamília *P* (Marzo *et al.*, 2008), pode-se especular que cópias de *Galileo* poderiam estar sendo alvo de mobilização *in trans*, a partir da disponibilidade de transposases codificadas pelo elemento *P*, bem caracterizado no genoma de *D. willistoni* (Daniels *et al.*, 1990; Clark *et al.*, 1995).

Além de um *TE* ser transferido verticalmente, sua origem em um genoma pode ocorrer das seguintes maneiras: aparecimento *de novo*, pela recombinação de elementos existentes no genoma; transferência horizontal, por um vetor e introgressão, por hibridização entre duas espécies. Já foi demonstrado que cópias de *Galileo* realizam recombinação ectópica dando origem à cópias quiméricas em *D. buzzatii* (Casals *et al.*, 2005; Delprat *et al.*, 2009) e *D. mojavensis* (Marzo *et al.*, 2013), porém eventos de *HT* ou introgressão não foram evidenciados claramente para essas espécies do grupo *repleta* de *Drosophila* e até o momento não faziam parte do processo de origem e manutenção de *Galileo* nos genomas de espécies hospedeiras.

Com o intuito de compreender a história evolutiva de *Galileo* em espécies neotropicais, encontramos pela primeira vez um cenário sugestivo de *HT* (Capítulo III desta tese). Na análise filogenética encontramos a primeira linha de evidências para inferir *HT* (segundo Wallau *et al.*, 2012): incongruências filogenéticas, distribuição descontínua de *Galileo* e alta similaridade de sequências entre espécies distantemente relacionadas. Ao compararmos valores de substituições sinônimas por sítios sinônimos (dS) entre as sequências de *Galileo* isoladas e genes das espécies envolvidas, encontramos diferenças estatisticamente significantes que reforçaram as primeiras evidências. Além da possibilidade de *Galileo* estar sendo transmitido verticalmente e por *HT* entre os subgrupos *willistoni*, *bocainensis* e ainda para espécies do subgrupo *cardini*, distantemente relacionado, a documentação de intercruzamentos dentro do grupo de espécies crípticas de *willistoni* e do subgrupo *cardini* (Gleason *et al.*, 1998; Robe *et al.*, 2010; Cenzi de Ré *et al.*,

2010), sugere também o papel de introgressões na história evolutiva de *Galileo* nestas espécies neotropicais de *Drosophila*.

## PERSPECTIVAS

Ainda que não tenha sido encontrado *Galileo* com características que apontem para a sua atual atividade em *D. willistoni*, cópias defectivas desse transposon, assim como atuam em *D. buzzatii*, podem também estar sendo uma fonte endógena de variação cromossômica para o grupo *willistoni* e grupo *cardini* de *Drosophila*. Além disso, cópias defectivas podem estar sendo mobilizadas via mecanismo *in trans*, com o uso de enzimas codificadas por elementos autônomos da mesma superfamília (*P* de transposons).

Com intuito de dar continuidade à caracterização e compreensão da dinâmica evolutiva de *Galileo*, propõe-se esclarecer questões relacionadas à expressão transcricional do elemento *Galileo* em espécies neotropicais. Quantificar transcritos de *Galileo* pode ser importante no sentido de que alguns estudos desenvolvidos em nosso grupo de pesquisa têm sugerido a atuação de cópias defectivas do elemento *hobo* como reguladores da mobilização de cópias completas (Torres *et al.*, 2006, Deprá *et al.*, 2009). Dessa forma as perspectivas de trabalho são:

- Caracterizar molecularmente as sequências nucleotídicas de cópias do elemento *Galileo* em populações de espécies do grupo *willistoni* e *cardini* de *Drosophila*.
- Detectar a presença de sequências (resíduos) que caracterizariam uma transposase funcional nos genomas das populações naturais destas espécies.
- Avaliar a possível expressão transcricional de cópias de *Galileo*.
- Estimar o número de cópias de *Galileo* nas diferentes espécies dos grupos *willistoni* e *cardini* de *Drosophila*.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida LM, Carareto CMA (2005) Origem, proliferação e extinção de elementos de transponíveis: qual seria a importância da transferência horizontal na manutenção desse ciclo? Monografia SBG, Sociedade Brasileira de Genética, 1ed, 43pp.
- Ayala FJ, Tracey ML, Hedgecock D and Richmond RC (1974) Genetic differentiation during the speciation process in *Drosophila*. *Evolution* 28:576-592.
- Bächli G (2013) TaxoDros: The Database on Taxonomy of Drosophilidae, <http://www.taxodros.uzh.ch/> (August, 2014).
- Bhutkar A, Schaeffer SW, Russo SM, Xu M, Smith TF and Gelbart WM (2008) Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. *Genetics* 179: 1657-1680.
- Brisson JA, De Toni DC, Duncan I, Templeton AR (2005) Abdominal pigmentation variation in *Drosophila polymorpha*: geographic in the trait, and underlying phylogeography. *Evolution* 59(5):104- 1059
- Brisson JA, Wilder J, and Hollocher H (2006). Phylogenetic analysis of the cardini group of *Drosophila* with respect to changes in pigmentation. *Evolution* 60: 1228-1241.
- Brookfield JFY (2005) The ecology of the genome - mobile DNA elements and their hosts. *Nat Rev Genet* 6:128-36.
- Burla H, Da Cunha AB, Cordeiro AR, Dobzhansky T, Malogolowkin C and Pavan C (1949) The *willistoni* group of sibling species of *Drosophila*. *Evolution* 3:300-314.
- Bushman, FD (2003) Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. *Cell* 115, 135-138
- Cáceres M, Ranz JM, Barbadilla A, Long M and Ruiz A (1999) Geration of a widespread *Drosophila* inversion by transposable element. *Science* 285:415-418.
- Cáceres M, Puig M and Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353-1364.
- Casals F, Cáceres M, and Ruiz A (2003) The Foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol* 20:674-685.
- Casals F, Cáceres M, Manfrin MH, González J and Ruiz A (2005) Molecular Characterization and chromosomal distribution of Galileo, Kepler and Newton,

three Foldback transposable elements of the *Drosophila buzzatii* species complex. *Genetics* 169:2047-2059.

Cenzi De Ré F, Loreto ELS and Robe LJ (2010) Gene and species trees reveal mitochondrial and nuclear discordance in the *Drosophila cardini* group (Diptera: Drosophilidae). *Invertebr Biol* 129:353-367.

Chaves NB, Tidon R (2005) Drosophilidae of Brazilian Savanna, the forgotten ecosystem. *Drosoph Inf Serv* 88:25-27

Chénais B, Caruso A, Hiard S and Casse N (2012) The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509:7-15. doi: 10.1016/j.gene.2012.07.042

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

Clark JB, Altheide TK, Schlosser MJ and Kidwell MG (1995) Molecular evolution of *P* transposable elements in the genus *Drosophila*. I. The *saltans* and *willistoni* species groups. *Mol Biol Evol* 12:902-13.

Clark *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 480:2003-218.

Clouaire T, Roussigne M, Ecochard V, Mathe C, Amalric F and Girard JP (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. *Proc Natl Acad Sci USA* 102:6907-6012.

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends in Genetics* 21:673-682.

Cordaux R, Udit S, Batzer MA, Feschotte C (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci USA*. 103:7941-42.

Cordeiro J, Valente VLS, Schmitz HJ (2009) Spontaneous melanic mutant found in a *Drosophila neocardini* natural population. *Drosoph Inf Serv* 92:7-10.

Cordeiro J, De Toni DC, Da Silva GS and Valente VLS (2014) Chromosomal evolution in the *Drosophila cardini* group (Diptera: Drosophilidae): photomaps and inversion analysis. *Genetica* 142: 461-472.

Crozatier M, Vaury C, Busseau I, Pelisson A and Bucheton A (1988) Structure and genomic organization of I elements involved in I-R hybrid dysgenesis in *Drosophila melanogaster*. *Nucleic Acids Res* 16: 9199-9213.

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

Da Cunha AB, Burla H and T. Dobzhansky (1950) Adaptive chromosomal polymorphism in *Drosophila willistoni*. *Evolution* 4: 212-235.

Da Cunha AB, Brncic D, Salzano FM (1953) A comparative study of chromosomal polymorphism in certain South American species of *Drosophila*. *Heredity* 2(7):193-202.

Da Cunha AB and Dobzhansky T (1954) A further study of chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. *Evolution* 8:119-134.

Da Cunha AB, Dobzhansky T, Pavlovsky O and Spassky B (1959) Genetics of natural populations. XXVIII. Supplementary data on the chromosomal polymorphism in *Drosophila willistoni* in its relation to the environment. *Evolution* 13:389-404.

De Setta N, Loreto ELS and Carareto CMA (2007) Is the evolutionary history of the O-type P element in the *saltans* and *willistoni* groups of *Drosophila* similar to that of the canonical P element? *J Mol Evol* 65:715-724.

De Setta N, Van Sluys M-A, Capy P and Carareto CMA (2009) Multiple invasions of Gypsy and Microplasmid retroelements in genus *Zaprionus* and melanogaster subgroup of the genus *Drosophila*. *BMC Evol Biol* 9:279.

De Toni DC, Heredia FO, Valente VLS (2001) Chromosomal variability of *Drosophila polymorpha* populations from Atlantic Forest remnants of continental and insular environments in the State of Santa Catarina, Brazil. *Caryologia G Citol Citogenet* 54(4):329-337.

De Toni DC, Loureiro MA, Hofmann PRP, Valente VLS (2006) Reference photomap of the salivary gland polytene chromosomes of *Drosophila neomorpha* (Heed and Wheeler, 1957). *Drosoph Inf Serv* 89:73-77.

De Toni DC, Gottschalk MS, Cordeiro J, Hofmann PRP, Valente VLS (2007) Study of the *Drosophilidae* (Diptera) communities on Atlantic Forest islands of Santa Catarina State, Brazil. *Neotrop Entomol* 36:356-375.

Deininger PL, Batzer MA. *Alu* repeats and human disease. *Mol Genet Metab* 1999;67:183-93.

Delprat A, Negre B, Puig M and Ruiz A (2009) The Transposon *Galileo* Generates Natural Chromosomal Inversions in *Drosophila* by Ectopic Recombination. *PLoS One* 4: e7883.

Dobzhansky T, Pavan C (1943) Studies on Brazilian species of *Drosophila*. Bolm Fac Filos Cienc S Paulo 36: 7-72.

Dobzhansky T (1946) Complete reproductive isolation between two morphologically similar species of *Drosophila*. Ecology 27:205-211.

Dobzhansky T, Ehrman L and Pavlovsky O (1957) *Drosophila insularis*, a new sibling species of the *willistoni* group. Univ Texas Publ 5721:39-47.

Dobzhansky T and Spassky B (1959) *Drosophila paulistorum* a cluster of species in statu nascendi. Proc Natl Acad Sci USA 45:419-428.

Dobzhansky T, Ehrman L, Pavlovsky O and Spassky B (1964) The superespecies *D. paulistorum*. Proc Natl Acad Sci USA 51:3-9.

Duda O (1925) Die Costaricanischen Drosophiliden (Dipteren) des Ungarischen National Museums zu Budapest. Ann Mus Nat Hung 22:149-229.

Ehrman L and Powell JR (1982) The *Drosophila willistoni* species group. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila*. Academic Press, New York, pp 193-225.

Engels, WR (1989) *P* elements in *Drosophila melanogaster*. In: Berg DE and Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 437-484.

Engels, W. R. (1996) Transposable Elements. Saedler, H and Gierl A (eds), Springer, Berlin pp.103-123.

Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH and Corces VG (1997). *Penelope*, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. Proc Natl Acad Sci USA 94: 196-201.

Feschotte C, Jiang N and Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329-41.

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

Feschotte C (2008) The contribution of transposable elements to the evolution of regulatory networks. Nat Rev Genet 9(5):397-405.

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

Futch DG (1962). Hybridization studies within the *cardini* species group of the genus *Drosophila*. Univ Texas Publ 6205: 539-554.



Galagan JE and Selker EU (2004) RIP: the evolutionary cost of genome defense. *Trends Genet* 20:417-23.

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

Gottschalk MS, Hofmann PRP and Valente VLS (2008) Diptera , Drosophilidae : historical occurrence in Brazil. *Check List*. 4:485-518.

Gonçalves JW, Valiati VH, Delprat A, Valente VL and Ruiz A (2014) Structural and sequence diversity of the transposon *Galileo* in the *Drosophila willistoni* genome. *BMC Genomics* 15:792.

González J, Casals F, Ruiz A: Testing chromosomal phylogenies and inversion breakpoint reuse in *Drosophila*. *Genetics* 2007, 175:167-177.

Gottschalk MS, De Toni DC, Valente VLS, Hofmann PRP (2007) Changes in Brazilian Drosophilidae (Diptera). *Bull AM Mus Natl Hist* 197:1-139.

Guillén Y and Ruiz A (2012) Gene alterations at *Drosophila* inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics* 13:53.

Hagemann S, Miller WJ, Pinsker W (1992) Identification of a complete P element in the genome of *Drosophila bifasciata*. *Nucleic Acids Res* 20:409-413.

Hagemann S, Miller WJ, Pinsker W (1994) Two distinct P element subfamilies in the genome of *Drosophila bifasciata*. *Mol Gen Genet* 244:168-175.

Hagemann S, Haring E, Pinsker W (1996a) Repeated horizontal transfer of P transposons between *Scaptomyza pallida* and *Drosophila bifasciata*. *Genetica* 98:43-51.

Hagemann S, Haring E, Pinsker W (1996b) A new P element subfamily from *Drosophila tristis*, *D. ambigua*, and *D. obscura*. *Genome* 39:978-985.

Heed WB (1962) Genetic characteristics of island populations. *Univ Texas Publ Stud Genet* 6205:173-206.

Heed WB (1963) Density and distribution of *Drosophila polymorpha* and its color alleles in South America. *Evolution* 17:502-518.

Heed WB, Krishnamurthy NB (1959) Genetic studies on the *cardini* group of *Drosophila* in the West Indies. *Univ Texas Publ Stud Genet* 5914:155-179

Heed WB, Russell JS (1971) Phylogeny and population structure in island and continental species of the *cardini* group of *Drosophila* studied by inversion analysis. Univ Texas Publ Stud Gen 6(7103):91-130.

Heed WB, Wheeler MR (1957) Thirteen new species in the genus *Drosophila* from the neotropical region. Univ Texas Publ Stud Gen 5721:17-38.

Hoffmann A, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? Annual review of ecology, evolution, and systematics 39:21-42.

Hollocher H (1996) Island hopping in *Drosophila*: patterns and process. Philos Trans R Soc Lond B Biol 351 (1341):735-743.

Hollocher H, Hatcher JL, Dyreson EG (2000a) Evolution of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. Evolution 54(6):2046-2056.

Hollocher H, Hatcher JL, Dyreson EG (2000b) Genetic and developmental analysis of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. Evolution 54(6):2057-2071.

Kaessmann H. et al. (2009) RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10, 19-31.

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

Kastritsis CD and Dobzhansky (1967) *Drosophila pavlovskiana* a race of a species? Amer Midl Nat 78:244-247.

Kazazian HH (2004) Mobile elements: drivers of genome evolution. Science 303:1626-1632.

Kidwell MG, Kidwell JF and Sved JA (1977). Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86, 813- 833.

Kidwell, M.G. (1992) Horizontal transfer of *P*-elements and other short inverted repeat transposons. Genetica 86, 275-286.

Kidwell MG (1994) The evolutionary history of the *P* family of transposable elements. J Hered 85:339-346.

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

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

Le Rouzic A, Capy P (2005). The first steps of transposable elements invasion: parasitic strategy versus genetic drift. *Genetics* 169: 1033-1043.

Loreto ELS, Carareto CMA and Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 1-10.

Loreto EL, Valente VL, Zaha A, Silva JC, Kidwell MG (2001) *Drosophila mediopunctata* P elements: a new example of horizontal transfer. *J Hered.* 92: 375-381.

Loreto ELS, Zambra FMB, Ortiz MF, Robe LJ: New *Drosophila* P-like elements and reclassification of *Drosophila* P-elements subfamilies. *Mol Genet Genomics* 2012a, 287:531-540.

Lozovskaya ER, Scheinker VS And Evgen'ev MB (1990) A hybrid dysgenesis syndrome in *Drosophila virilis*. *Genetics* 126: 619-623.

Majumdar S, Singh A, Rio DC (2013) The human THAP9 gene encodes an active P- element DNA transposase. *Science* 339: 446-448.

Machado MX, De Toni DC, Hofmann PRP (2001) Abdominal pigmentation polymorphism of *Drosophila polymorpha* (Dobzhansky and Pavan) collected on Ilha de Santa Catarina and neighboring islands. *Biotemas* 14 (1):87-107.

Marzo M, Puig M and Ruiz A (2008) The *Foldback*-like element *Galileo* belongs to the P superfamily of DNA transposons and is widespread within the *Drosophila* genus. *PNAS* 105 (8):2957-2962.

Marzo M, Bello X, Puig M, Maside X and Ruiz A (2013) Striking structural dynamism and nucleotide sequence variation of the transposon *Galileo* in the genome of *Drosophila mojavensis*. *Mob DNA* 4:1. doi: 10.1186/1759-8753-4-6

McDonald JF (1993) Evolution and consequences of transposable elements. *Curr Opin Genet Des* 3:855-864.

McVean G (2010) What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc Lond B Biol Sci.* 365: 1213-1218.

Morgante M et al. (2005) Gene duplication and exon shuffling by *Helitron*-like transposons generate intraspecies diversity in maize. *Nat. Genet.* 37, 997-1002

Napp M and Cordeiro AR (1981) Interspecific relationships in the cardini group of *Drosophila* studied by electrophoresis. *Rev. Bras. Biol.* 4: 537-547.

Nouaud D, Quesneville H and Anxolabehere D. (2003) Recurrent exon shuffling between distant P-element families. *Mol Biol Evol* 20:190-199.

O'Grady PM and Kidwell MG (2002) Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Mol Phylogenet Evol* 22:442-453.

O'Hare K, Rubin GM (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 34: 25-35.

Pardue ML and DeBaryshe PG (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu. Rev. Genet.* 37, 485-511

Patterson JT and Mailand GB (1944) Studies in the genetics of *Drosophila*. IV. Papers dealing with taxonomy, nutrition, cytology and interspecific hybridization in *Drosophila*. I. The Drosophilidae of México. *The Univ Texas Publ* 4445:9-101.

Petrov DA, Schutzman JL, Hartl DL and Lozovskaya ER (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc Natl Acad Sci USA* 92: 8050-8054.

Pinsker W, Haring E, Hagemann S and Miller WJ (2001) The evolutionary life history of P transposons: from horizontal invaders to domesticated neogenes. *Chromosoma* 110:148-158.

Powell JR, Sezzi E, Moriyama EN, Gleason JM and Caccone A (2003) Analysis of a shift in codon usage in *Drosophila*. *J Mol Evol* 57:214-225.

Prada CF (2010). Evolución cromosómica del cluster *Drosophila martensis*: origen de las inversiones y reutilización de los puntos de rotura. PhD thesis. Universitat Autònoma de Barcelona.

Quesneville H, Nouaud D and Anxolabehere D (2005) Recurrent recruitment of the THAP DNA-binding domain and molecular domestication of transposable element. *Mol Biol Evol* 22:741-746.

Ranz JM, Casals F, Ruiz A (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Research* 11:230-239.

Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J, Ashburner M and Bergman CM (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol* 5:e152.

Rebatchouk D and Narita JO (1997) Foldback transposable elements in plants. *Plant Mol Biol* 34:831-835.

Rio DC (1991). Regulation of *Drosophila P* element transposition. Trends Genet. 7: 282- 287.

Rio DC (2002) *P* transposable element in *Drosophila melanogaster*. In: Mobile DNA II. American Society for Microbiology Press, Washington, DC pp. 484-518.

Rius N, Delprat A and Ruiz A (2013) A divergent *P*-element and its associated MITE, *BuT5*, generate chromosomal inversions and are widespread within the *Drosophila repleta* species group. Genome Biol Evol 5:1127-41.

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

Robe LJ, Cordeiro J, Loreto ELS and Valente VLS (2010a) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). Genética doi: 10.1007/s10709-009-9432-5.

Robe LJ, Loreto ELS and Valente VLS (2010b) Radiation of the *Drosophila* subgenus (Drosophilidae, Diptera) in the Neotropics. J Zool Syst Evol Res doi: 10.1007/ s10709-009-9432-5.

Rohde C (2000) Polimorfismo cromossômico e elementos transponíveis em *Drosophila willistoni*. PhD Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Rohde C, Degrandi TH, De Toni DC and Valente VLS (2005). *Drosophila willistoni* polytene chromosomes. I. Pericentric inversion on X chromosome. Caryologia 58(3): 249-254.

Rohde C, Valente VLS (1996) Cytological maps and chromosomal polymorphism of *Drosophila polymorpha* and *Drosophila cardinoides*. Braz J Genet 19:27-32

Rohde C, Garcia ACL, Valiati VH and Valente VLS (2006) Chromosomal evolution of sibling species of the *Drosophila willistoni* group. I. Chromosomal arm IIR (Muller's element B). Genetica 126:77-88.

Roussigne M, Kossida S, Lavigne A-C, Clouaire T, Ecochar V, Glories A, Amalric F and Girard J-P (2003) The THAP domain: a novel protein motif with similarity to the DNA-binding domain of *P* element transposase. Trends Biochem Sci 28:66-69.

Schaack S, Gilbert C and Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537-46.

Schmitz HJ, Valente VLS and Hofmann PRP (2007) Taxonomic Survey of Drosophilidae (Diptera) from Mangrove Forests of Santa Catarina Island, Southern Brazil. *Systematics, Morphology and Physiology*. 53-64.

Sinzelle L, Izsvák Z and Ivics Z (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell Mol Life Sci* 66:1073-93.

Spassky B, Richmond RC, Perez-Salas S, Pavlovsky O, Mourão CA, Hunter AS, Hoentgsberg H, Dobzhansky T and Ayala FJ (1971) Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. *Evolution* 25:129-143.

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP et al. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* 15:1-18.

Stalker HD (1953) Taxonomy and hybridization in the *cardini* group of *Drosophila*. *Ann Entomol Soc Am* 46:343-358.

Sturtevant AH (1916) Notes on North American Drosophilidae with descriptions of twenty-three new species. *Ann Entomol Soc Am* 9:323-343.

Sturtevant AH (1921) The north American species of *Drosophila*. *Carnegie Inst Wash Publi* 301:1-150.

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

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

Val FC, Marques MD, Vilela CR, (1981) Drosophilidae of Neotropical region. In: Ashburner M., Carson HL and Thompson JN (eds.) *The Genetics and Biology of Drosophila*. Academic Press, Orlando, pp. 123-168.

Valente VLS, Sarraveda C, Morales N and Araújo AM (1981) Observations on the attraction of *Drosophila* species for different baits and chromosomal polymorphism in *Drosophila willistoni*. *DIS* 56:147-149.

Valente VLS and Araújo AM (1986) Chromosomal polymorphism, climatic factors and variation in population size of *Drosophila willistoni* in Southern Brazil. *Heredity* 57:149- 160.

Valente VLS and Morales NB (1985) New inversions and qualitative description of inversion heterozygotes in natural populations of *Drosophila willistoni* inhabiting two different regions in the State of Rio Grande do Sul, Brazil. *Revta Bras Genet* 8:167-173.

Valente VLS, Rohde C, Valiati VH, Morales NB and Goñi B (2001) Chromosome inversions occurring in Uruguayan populations of *Drosophila willistoni*. *Drosophila Information Service* 84:55-59.

Valente VLS, Goñi B, Valiati VH, Rohde C and Morales NB (2003). Chromosomal polymorphism in *Drosophila willistoni* populations from Uruguay. *Genet. Molec. Biol.* 26(2): 163-173.

Valente VLS, Ruzsczyk A and dos Santos RA (1993) Chromosomal polymorphism in urban *Drosophila willistoni*. *Rev Bras Genet* 16:307-319.

Wallau GL, Ortiz MF and Loreto ELS (2012) Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol Evol* 4:689-99. doi: 10.1093/gbe/evs055

Wessler SR, Bureau TE and White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Curr Opin Genet Dev* 5:814-821.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al.* (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8:973-982.

Wilder JA, Dyreson EG, O'Neill RJ, Spangler ML, Gupta R, Wilder AS, Hollocher H (2004) Contrasting model of natural selection acting on pigmentation genes in the *Drosophila dunnii* subgroup. *L Exp Zoolog B Mol Dev EVol* 302(5):469-82

Yannopoulos G, Stamatis N, Monastirioti M *et al.* (1987). *hobo* is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23 5.MRF. *Cell* 49:487-495.

Zhou, J. and Eickbush, T.H. (2009) The pattern of R2 retrotransposon activity in natural populations of *Drosophila simulans* reflects the dynamic nature of the rDNA locus. *PLoS Genet.* 5, e1000386