UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

Plasticidade genômica de *Drosophila willistoni*, abordagem *in situ* e *in silico*

Natasha Ávila Bertocchi

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.

Profª Drª. Vera Lúcia S. Valente Gaiesky – Orientadora

Profª Drª. Maríndia Deprá – Coorientadora

Porto Alegre, abril de 2022.

Este trabalho foi desenvolvido no Laboratório de *Drosophila* do Departamento de Genética do Instituto de Biociências na Universidade Federal do Rio Grande do Sul, com recursos financeiros da FAPERGS e CNPq, e Bolsa de Doutorado CNPq.

“Cada criatura viva é extraordinariamente improvável”

Alexandre Beck

Dedico este pequeno livro à minha mãe Juliana e ao meu pai Leandro, por todas as coisas imprescindíveis à vida.

AGRADECIMENTOS

Me faltam palavras pra expressar a importância de vocês na minha vida. Eu encarei a universidade, e o meio acadêmico, como uma linda e incrível jornada. Assim, como a ciência, as jornadas são feitas por pessoas, e a minha jornada até a conclusão desta tese foi recheada das pessoas mais incríveis que eu poderia encontrar no meu caminho.

E não poderia começar por outrem, que não as minhas orientadoras, prof. Vera e Maríndia. Prof. Vera, obrigada por tanto incentivo, apoio e presença. Aprendi e me apaixonei por esse universo das *Drosophilas* contigo. Que inspiração tu és profa!

Prof. Maríndia, obrigada pela paciência e carinho sempre. Por todas as vezes que ajudou a me organizar com a quantidade de dados e informações, pela atenção e carinho ao longo de todo esse tempo.

Prof. Fabiano, o primeiro e sempre meu orientador, obrigada por junto com o Gabriel Wallau me apresentar o mundo dos elementos de transposição.

Thays, “the best” coautora que esse mundinho redondo conseguiu produzir, obrigada por tudo. Principalmente, sou imensamente grata por você nunca ter “soltado a minha mão” nessa jornada muito maluca que é fazer uma dissertação sem recursos, e depois uma tese no meio de uma pandemia. Obrigada pelas piadas péssimas e ideias esdrúxulas nas horas mais tensas. Obrigada por rir do meu senso de humor altamente cruel quando estava com fome, sono, frio, sede, enfim tudo que se sente depois de um longo dia fazendo experimentos.

Anelise, Thays, Henrique, Mayara, Carol, Thamiris obrigada pelas tantas vezes que me ouviram reclamar e relatar tudo que não deu certo, e também por todas as comemorações que fizemos juntos. Jean obrigada por todas as vezes que salvou meu computador. Bruna Fernanda por me receber quando cheguei em Porto Alegre.

Rebeca, obrigada por topar minhas ideias, responder as minhas perguntas mais fora de hora possível, me ajudar nas coisas mais inusitadas, e tantas vezes que ouviu eu me lamentar, obrigadão!

Aos meus amigos que carrego com muito carinho desde São Gabriel. Vanusa obrigada por todas as mensagens de carinho e apoio. Jeferson, obrigada por todas as conversas em que dividimos as angústias desse momento e, também por todas as doações de bactérias. Natália, *teacher* obrigada pela paciencia, carinho, atenção sempre.

A todos os integrantes e agregados do laboratório de *Drosophila,* os poucos quando cheguei e aos novos. Paula, que me recebeu incrivelmente bem, tornou meu início muito mais fácil. Carolina Flores, Gilberto, Rebeca, Brenda, Carol Predger, Ane, Henrique, Vitor, Antônio Marco, Marcos, Andrea, Pâmela, Arthur, Leo, Robson, Yasmin, Filipe, Mari e Thays (mesmo sendo de outro lab, são muito nossas), obrigada pelo convívio de sempre. Berenice, Dani, Helena e Pedro obrigada por sempre estarem dispostos a auxiliar no que fosse necessário, o trabalho de vocês foi essencial para realizar essa tese.

Ao Elmo e ao Gabriel por todo o suporte, disponibilidade e ajuda sempre. Vocês tornaram as burocracias durante esses anos menos penosas.

Ao prof. Thales Freitas, Laboratório de citogenética e evolução, por disponibilizar o laboratório para realizações dos experimentos de FISH. Ao Luciano por tanto auxílio, facilitando imensamente a execução dos experimentos. A Lúcia pela paciência impar em todas as vezes que cheguei em cima da hora com amostras para enviar para sequenciamento. Obrigada a vocês e a todos os demais integrantes do laboratório pelos cafés no meio da manhã, em especial o Rafa Kretschmer pela companhia desde o DGA na UNIPAMPA.

Por fim, mas jamais com menor importância, à minha família: minha mãe Juliana, meu pai Leandro, meu padrasto Paulo e o meu irmão Pedro a vida faz sentido com vocês comigo. Obrigada por todo o apoio, incentivo e carinho nessa jornada.

Obrigada a todos por compartilharem essa jornada comigo.

SUMÁRIO

LISTA DE ABREVIATURAS

5S RNA - RNA ribossômico 5S

DIRS- *Sequência repetitiva de Dictyostelium,* do inglês *Dictyostelium repetitive sequence*

DNA - ácido desoxirribonucleico, do inglês *desoxyribonucleic acid*

EN- *Endonuclease*

FISH – hibridização *in situ* fluorescente, do inglês *fluorescence in-situ hybridization*

HTT - transfêrencia horizontal de elementos de transposição, do inglês *Horizontal transposon transfer*

HUH, Rep/Helicase proteína com HUH com atividade de endonuclease,

IN- Integrase

LTR - repetição terminal longa, do inglês *long terminal repeat*

MITEs - elementos transponíveis de repetição invertida em miniatura, do inglês *miniature Inverted-repeat Transposable Elements*

ORF- fase aberta de leitura, do inglês, *open reading frame*

pb- pares de base

PLE – elementos do tipo *Penelope,* do inglês *Penelope*-like *elements*

pPolB – DNA polimerase do tipo B, do inglês *protein-primed type B DNA polymerase*

RH - RNase H- Ribonuclease H

RT - Transcriptase reversa, do inglês *reverse transcriptase*

TE - elementos de transposição, do inglês *Transposable Elements*

TIR- repetição terminal invertida, do inglês *Terminal Inverted Repeat*

TP- *Transposase*

tRNAs – Ácido ribonucleico transportadores, do inglês *transfer ribonucleic acid*

UHU/REP- proteina Replicase/Helicase com domínio HUH

YR- *Tirosina recombinase,* do inglês *tyrosine recombinase*

RESUMO

Os elementos de transposição (TEs do inglês *Transposable elements*) estão ubiquamente presentes nos genomas eucarióticos e são bastante diversos. TEs são elementos genéticos capazes de mobilização cromossômica e replicaticativa em células germinativas, têm papel no tamanho e estrutura dos genomas, e também são uma extensa fonte de mutações e polimorfismos genéticos. Os elementos podem influenciar a expressão de genes, *splicing* alternativos e rearranjos cromossômicos. A origem dos TEs em eucariotos provavelmente ocorreu no ancestral desses organismos, e além disso a origem de superfamílias e famílias de TEs está associada a quimerismos e evolução modular. Os TEs devido a sua variabilidade, foram classificados de acordo com as moléculas responsáveis por sua mobilização e integração ao genoma. Abordamos neste estudo principalmente os elementos: *412* LTR- retrotransposon da Classe I; e os transposons de DNA *hobo, BuT2* e *mar* da Superfamília *hAT* da Classe II. A tese tem como objetivo contribuir para o conhecimento sobre a presença, dinâmica e coevolução dos TEs e os genomas hospedeiros, em espécies predominantemente neotropicais de *Drosophila.* Com esse proposito, revisamos os estudos, principalmente os com abordagens evolutivas, que tiveram como organismo modelo *Drosophila willistoni* e as espécies do grupo *willistoni*. Realizamos uma detalhada pesquisa *in silico* para analisar os TEs *hobo, BuT2 e mar* nas espécies do grupo *willistoni* com genomas sequenciados disponíveis. Além disso, analisamos o número de cópias e a distribuição espacial desses mesmos TEs da superfamília *hAT* nos cromossomos politênicos de algumas linhagens de *D. willistoni* coletadas ao longo da distribuição geográfica da espécie. As linhagens de *D. willistoni* Gd-H4-1, WIP-4 e DSG12.00 foram usadas nas análises *in situ*; *D. willistoni* 00 e L17 foram utilizadas nas análises *in silico*; e *D. willistoni* Gd-H4-1 foi utilizada em ambas as análises. Por fim, por meio de análises *in silico* mostramos a presença e estrutura das cópias do retrotransposon *412* nos genomas sequenciados de dípteros. Nas espécies do grupo *willistoni* mostramos a presença e quantidade de fragmentos homólogos ao domínio da transcriptase reversa do LTR-retrotransposon *412* em espécies com genoma sequenciado e de populações naturais não sequenciadas do grupo *willistoni*. Na espécie *D. willistoni,* utilizando abordagem *in situ,* mostramos

também, a distribuição do elemento *412* nos cromossomos politênicos da linhagem GD-H4-1. Além disso, identificamos a intricada evolução da linhagem *412/mdg1* nos genomas de dípteros.

ABSTRACT

The transposable elements (TEs) are obliquely present in eukaryotic genomes and are quite diverse. TEs are genetic elements capable of chromosomal and replicative mobilization in germ cells, they have a role in the size and structure of genomes, and are also an extensive source of mutations and genetic polymorphisms. The elements influence gene expression, alternative splicing and chromosomal rearrangements. The origin of TEs in eukaryotes probably occurred in the ancestor of these organisms, in addition, the origin of superfamilies and families of TEs is associated with chimerism and modular evolution. Due to their variability, TEs were classified according to the molecules responsible for their mobilization and integration into the genome. In this study, we focus mainly on the elements: *412* LTR- retrotransposon (Class I); and the *hobo, BuT2* and *mar* DNA-transposons of the *hAT* superfamily (Class II). We aim to contribute to the knowledge about the presence, dynamics and coevolution of TEs and their host genomes, in predominantly Neotropical species of *Drosophila*. For this, we reviewed the studies, mainly studies with evolutionary approaches, which had as model organism *Drosophila willistoni* and the species of the *willistoni* group. We conducted an in-silico search to analyze *hobo, BuT2*, and *mar* TEs in available genomes of the *willistoni* group. We also, analyzed the copy number and spatial distribution of these same transposons of the *hAT* superfamily on polytene chromosomes of some *D. willistoni* strains collected along with the geographic distribution of the species. Further, *D. willistoni* Gd-H4-1, WIP-4 and SG12.00 strains were used for *in-situ* analyses; *D. willistoni* 00 and L17 strains were used for *in-silico* analysis; and *D. willistoni* Gd-H4-1 was used for both *in-situ* and *in-silico* analyses. Finally, by using *in-silico* analysis, we showed the presence and structure of copies of retrotransposon *412* in sequenced dipteran genomes. In species of the *willistoni* group, we showed the presence and quantity of homologous fragments to the reverse transcriptase domain of LTR-retrotransposon *412* in species with sequenced genomes and in unsequenced natural populations of the *willistoni* group. In the species *D. willistoni*, using an *in-situ* approach, we also showed the distribution of element *412* in the polytene chromosomes of the lineage GD-H4-1. Furthermore, we identified the intricate evolution of the *412/mdg1* lineage in dipteran genomes.

CAPÍTULO 1

I – Introdução

Elementos de transposição

Os elementos de transposição (do inglês *Transposable Elements*, TEs) são sequências de DNA, encontradas repetidamente nos genomas, que possuem como particularidade a capacidade de mudarem de posição tanto dentro como entre genomas. Os TEs foram conhecidos como “parasitas genômicos”, “genes saltadores”, “DNA lixo” e “DNA egoísta” devido à capacidade de se movimentar de um sítio do genoma a outro, e assim provocar alterações diretas e indiretas nos genomas dos organismos hospedeiros desde sua descoberta por Barbara McClintock. Os estudos na área só aumentam e atualmente os TEs podem ser considerados virtualmente presentes em todos os genomas eucarióticos analisados, com única exceção de *Plasmodium falciparum* e espécies estritamente relacionadas (Wicker *et al*. 2007).

Os TEs têm capacidade de se replicar dentro dos genomas independentemente da célula hospedeira (Wells and Feschotte 2020). Podem ser definidos como: elementos genéticos capazes de mobilização cromossômica e replicativa em células germinativas, consequentemente aumentando em frequência através da herança vertical (Wells and Feschotte 2020). Os TEs são muito variáveis quanto ao seu tamanho ou formas, podendo ter varias: de poucas dezenas de pares de base (pb) a milhares de pares de base, como também ter sequências para a codificação para proteínas com atividades bioquímicas variadas, além de regiões regulatórias não-codificantes que promovem sua transposição (Arkhipova and Yushenova 2019; Wells and Feschotte 2020).

O efeito profundo que os TEs tiveram e tem na evolução dos genomas eucarióticos vem se esclarecendo ao longo do tempo. Os TEs têm um papel crítico no tamanho e estrutura dos genomas, como também na codificação e regulação das proteínas por eles produzidas (Feschotte and Pritham 2007; Ågren and Wright 2011; Fedoroff 2012; Chuong *et al*. 2017; Arkhipova and Yushenova 2019; Bourgeois and Boissinot 2019; Cosby *et al*. 2019). Para compreender como os TEs impactam a diversificação e biologia das espécies é necessário conhecer e entender a diversidade e biologia dos próprios TEs (Wells

and Feschotte 2020). Com base nisso, propomos nesta tese aprofundar o conhecimento sobre três transposons de DNA da superfamília *hAT* (*BuT2, mar* e *hobo*), o retroelemento de *LTR 412,* utilizando como organismo de estudo drosofilídeos, principalmente *Drosophila willistoni* e as espécies do grupo *willistoni.* Por isso, também propomos uma revisão sobre os estudos evolutivos envolvendo as espécies do grupo *willistoni.*

Classificação dos elementos transposição

Dada a ampla distribuição e diversidade de elementos transponíveis se fez necessário um sistema de classificação dos mesmos. Algumas propostas foram desenvolvidas como a de Finnegan (1989), Wicker *et al*. (2007) e Kapitonov and Jurka (2008). Finnegan (1989) propôs a primeira classificação para TEs, onde os divide em duas grandes classes de acordo com a molécula intermediária do mecanismo de transposição. Os elementos de Classe I possuem como molécula intermediária de transposição um RNA e os de Classe II se transpõe diretamente via DNA (Finnegan 1989). Com a descobeta contínua de novos elementos com características distintas, principalmente devido à facilidade e acessibilidade dos sequenciamentos, outras subdivisões foram necessárias. Assim, mantendo a proposta inicial de Finnegan (1989) e incorporando outras subdivisões, as Classe I e II podem ser subdivididas em subclasses ou ordens, que são delimitadas de acordo com o mecanismo de replicação e/ou integração aos cromossomos. Subsequentemente, são divididos em superfamílias e famílias, os quais são determinadas por suas relações filogenéticas (Wicker *et al*. 2007). Podem ainda, em muitos casos, serem subdivididos novamente em subfamílias, de acordo com a identidade nucleotídica e de aminoácidos entre as sequências. Essas divisões estão ilustradas na Figura 1.

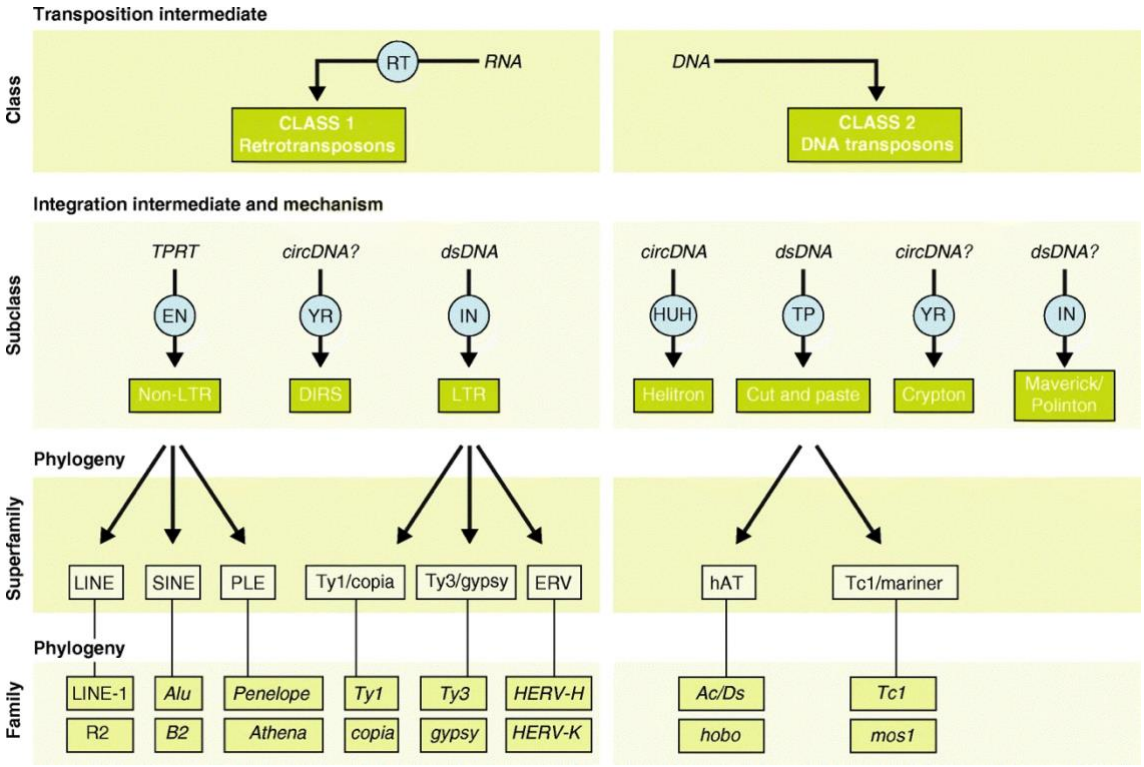


Figura 1: Classificação dos elementos de transposição em eucariotos. Observam-se as características e relações entre as classes, subclasses, superfamília e famílias de TEs. As enzimas codificadas pelos TEs estão representadas nos circulos azuis: EN, endonuclease; YR, tirosina recombinase (do inglês *tyrosine recombinase*); IN, integrase; HUH, proteína Replicase/Helicase com domínio HUH e atividade de endonuclease; TP, transposase. Abreviações: circDNA intermediário por DNA circular, dsDNA intermediário por DNA fita dupla, TPRT transcrição reversa com *target primed* (do inglês, *target primed reverse transcription*) e DIRS sequência repetitiva de *Dictyostelium* (do inglês *Dictyostelium repetitive sequence*). Imagem adaptada de Bourque *et al*. (2018).

Apoiada na divisão de Wicker *et al*. (2007), em Wells and Feschotte (2020) é proposto a divisão da Classe I, os retrotransposons, em três grandes subclasses de acordo com mecanismo de replicação e integração do TE. Os elementos com repetição terminal longa (do inglês *long terminal repeat*, LTR), mobilizados por uma integrase (IN); os elementos não-LTR mobilizados por uma endonuclease (EN); e os elementos mobilizados por uma tirosina recombinase (YR) chamados de sequência repetitiva de *Dictyostelium* - DIRS (do inglês *Dictyostelium repetitive sequence*) (Figura 1).

A classificação de Wicker *et al*. (2007), subdivide os retroelementos da Classe I em cinco ordens de acordo com as características estruturais e filogenia da transcriptase reversa (RT) (Figura 1) – retrotransposons LTR, LINE (do inglês, *Long Interspersed Nuclear Elements*), SINE (do inglês, Short Interspersed Nuclear Element), DIRS e PLE (do inglês *Penelope*-like elements) - com

destaque nos animais para as ordens dos LINES e SINES (Wicker *et al*. 2007). Nos LINES as principais superfamílias são *R2, L1, RTE, I e jockey,* e estruturalmente possuem duas ORFs (do inglês, *open reading frame*) (Wicker *et al*. 2007). A ORF1 codifica uma proteína cuja a função ainda não está esclarecida, e além disso não está presente em alguns grupos desses TEs (Wells and Feschotte 2020). Contudo, a ORF2 tem importante função na integração do TE, assim codifica para uma RT e uma endonuclease nas cópias ativas para transposição (Wicker *et al*. 2007; Wells and Feschotte 2020). Os SINES são elementos não-autônomos, assim necessitam da maquinaria enzimática dos LINES para sua transposição, e têm como origem a retrotransposição acidental de vários transcritos da polimerase III (Kramerov and Vassetzky 2005). Os elementos SINES são pequenos, de 80-500pb, e podem ser encontrados em grande quantidade nos genomas, como por exemplo, o elemento *Alu* com aproximadamente 500 mil cópias no genoma humano (Rowold and Herrera 2000).

Outra ordem relevante é dos retrotransposons com LTR (Figura 1), que possuem longas terminações repetidas diretas nas duas extremidades, podem chegar a 25 mil pares de bases de tamanho, possuem o gene *gag* (codifica para uma proteína similar ao capsídeo viral) e o gene *pol* (codifica uma poliproteína que inclui uma integrase e uma transriptase reversa) (Wicker *et al*. 2005; Carareto *et al*. 2015). Sua origem está associada a retrovírus, existindo duas hipóteses, ou os retrovírus perderam a capacidade de mobilização extracelular, ou os retrotransposons com LTR adquiriram os domínios protéicos para mobilização extracelular (Wicker *et al*. 2005). Um exemplo de retrotransposons com LTR é a superfamília *gypsy,* amplamente distribuída nos metazoários, com destaque para diversos estudos em *Drosophila* (McCullers and Steiniger 2017).

Como cada ciclo de reprodução de elementos de Classe I origina uma cópia nova, distintamente do que é observado nos elementos de Classe II, consequentemente, os retrotransposons contribuem significativamente para fração repetitiva dos genomas (Han and Boeke 2005). Devido à sua transposição via DNA-DNA, os TEs de Classe II são geralmente encontrados em menor quantidade nos genomas dos eucariotos. Contudo, mesmo com diferença

quantitativa, ambos geram diferentes impactos nos genomas de acordo com as interações que foram estabelecidas entre os elementos e o genoma hospedeiro.

Wicker *et al*. (2007), sugere duas subclasses para os elementos de transposons de Classe II, contudo Wells and Feschotte (2020) sugerem quatro subclasses baseadas nas moléculas intermediárias de integração. No entanto, ambas as divisões propostas, mantém os transposons de DNA clássicos, pertencentes a uma subclasse: a subclasse 1 (Wicker *et al*. 2007), ou subclasse "*cut-and-past*" (corta e cola) (Wells and Feschotte 2020). Os transposons de DNA subclasse 1 ou "*cut-and-past*" são o grupo mais estudado da Classe II. São TEs relativamente simples, consistindo de uma única ORF flanqueada por pequenas repetições terminais invertidas (TIRs, do inglês *Terminal Inverted Repeats*). A sua transposição se dá por meio de um intermediário de DNA fita dupla, mecanismo de transposição chamado de "*cut-and-past*" (corta e cola), através da enzima transposase, que reconhece as TIRs (Wicker *et al*. 2007). Estas transposases possuem resíduos catalíticos de ácido aspártico e glutâmico, DD(E/D) (Yuan and Wessler 2011). São os transposons mais diversos e difundidos, o sucesso deste grupo de elementos é tamanho que transposases DDE são, provavelmente, os genes mais antigos antigo e abundantes da Terra (Aziz *et al*. 2010).

O processo de mobilização basicamente é iniciado por uma transposase nucleofílica atacando uma molécula de água próxima das extremidades de cada TIR, eventualmente resultando na excisão direta do transposons e sua realocação em outra região do genoma (Hickman and Dyda 2016). Esse tipo de processo não é replicativo, logo para aumentar o número de cópias esses TEs utilizam outras estratégias beneficiando-se da maquinária do hospedeiro. Os transposons podem aumentar o número de cópias de pelo menos duas maneiras. A primeira, quando a transposição ocorre preferencialmente ao mesmo tempo da síntese de DNA do hospedeiro, ocorrendo de lugares replicados para não replicados, e assim fazendo que os TEs sejam replicados duas vezes (Ros and Kunze 2001; Spradling *et al*. 2011; Fricker and Peters 2014). A segunda, por meio de recombinação homóloga quando há excisão do transposon, quando ocorre o deslizamento da fita e troca de molde pode acarretar a elementos com deleções internas (Engels *et al*. 1990; Hsia and

Schnable 1996; Rubin and Levy 1997). Esses elementos deletados, ou seja, não-autônomos, possivelmente perdem sua capacidade de codificação, contudo muitas vezes conservam as TIRs, podendo assim ser reconhecidos e mobilizados pelas transposases de elementos autônomos. Esses elementos não-autônomos e pequenos são chamados de MITEs (do inglês *Miniature Inverted-repeat Transposable Elements*), e tendem a proliferar-se de forma mais eficiente do que os elementos autônomos, formando assim extensas famílias de MITEs (Feschotte *et al*. 2002; González and Petrov 2009).

A tese tem como foco principalmente nos elementos *hobo, BuT2* e *mar* que fazem parte dos elementos de Classe II, Subclasse I, Superfamília *hAT,* subfamílias *Ac, tip e Buster* respectivamente. E por último, o retroelemento *412* faz parte dos elementos de Classe I, ordem LTR e superfamília retrovírus.

Origem dos elementos de transposição

Quando e como os TEs se orginaram e como estão relacionados entre si? Esta é uma pergunta, que exige uma resposta difícil e complexa, e que muitos pesquisadores vêm ao longo dos anos tentando elucidar desde sua descoberta por Barbara McClintock.

Com o aumento de genomas hospedeiros sequenciados e o desenvolvimento de ferramentas para anotação automática de TEs, as análises filogenômicas, as quais integram distribuição taxonômica dos elementos com análises filogenéticas das proteínas compartilhadas por eles se tornaram muito mais robustas (Xiong and Eickbush 1990; Wicker *et al*. 2007; Yuan and Wessler 2011). Porém, há algumas limitações, que dificultam essas análises como: dificuldade de alinhamentos confiáveis principalmente considerando elementos de diferentes superfamílias, e os inúmeros casos de transferência horizontal, inclusive entre *taxas* muito distantes (Arkhipova 2017).

Contudo, algumas conclusões puderam ser obtidas ao longo do tempo, e foram sumarizadas por Wells and Feschotte (2020). As principais subclasses de elementos são encontradas amplamente distribuídas nas espécies eucarióticas. As topologias filogenéticas obtidas com o "*core*" das proteínas dos TEs são consistentes com a ideia de que as principais subclasses estavam presentes no início da evolução dos genomas eucarióticos. E, a evolução dos TEs é altamente

modular, ou seja, com ganho ou perda de proteínas de um *pool* compartilhado de domínios conservados.

Alguns dados sugerem que as estruturas enzimáticas ou os próprios TEs antecedem a existência dos eucariotos. Essa hipótese é corroborada por alguns indícios, embora os TEs sejam muito diversos em formato e estrutura, apenas um pequeno grupo de domínios catáliticos está envolvido na replicação e transposição dos TEs (como RT, DDE IN, YR, UHU/REP e pPolB). Além disso, alguns desses domínios (HUH, RT, pPolB) presentes em grupos distintos de TEs compartilham a conservação da estrutura de reconhecimento de RNA (Krupovic *et al*. 2019). Embora haja muitas politomias na árvore dos transposons de DNA com DDE catalítico, algumas dessas superfamílias de transposases, como as da superfamília *Tc1/mariner*, estão associadas as transposases codificadas por grupos distintos de sequências de inserção bacteriana (Feschotte and Pritham 2007; Bao *et al*. 2009; Kojima 2019). Mesmo sendo necessário cautela, essas informações sugerem que o surgimento dos transposons de DNA antecede o dos eucariotos (Wells and Feschotte 2020). Nenhum outro grupo de TEs eucariotos tem homólogos em bactérias ou em archea.

Como mostrado no tópico anterior, há profundas relações entre as enzimas de transposição dos TEs, sendo inclusive possível subdividi-los de acordo com essas enzimas. Assim, os grandes grupos de TEs terem origem em comum é esperado, como a origem de superfamílias e famílias de TEs estar associada a quimerismos e evolução modular. Ou seja, os TEs assim como outros elementos móveis como vírus e plasmídeos, conseguem com frequência fazer trocas de unidades proteicas. Essas trocas podem envolver domínios essenciais para transposição como também domínios adquiridos do hospedeiro (Arkhipova 2017; Arkhipova and Yushenova 2019; Krupovic *et al*. 2019). Por exemplo, os retrolementos de LTR e os DIRs “clusterizam” junto com os retrovírus quando as análises filogenéticas se baseiam na RT. Contudo, quando as reconstruções filogenéticas analisam a enzima TY, os elementos DIRs são mais proximamente relacionados com os elementos de Classe II- *Cryptons* (Wells and Feschotte 2020). Outro exemplo da evolução modular, é dos retrotransposons de LTR e os transposons de DNA compartilharem o domínio catalítico DDE recombinase para integração dos TEs nos cromossomos, exceto

os transposons do tipo *Helitrons* (Lerat *et al*. 1999; Koonin and Krupovic 2017; Arkhipova 2017; Arkhipova and Yushenova 2019; Krupovic *et al*. 2019).

O quimerismo é um fator importante na evolução dos elementos SINEs. A maioria dos SINEs originaram-se de transcritos da polimerase III como tRNAS, 7SL RNA ou 5S RNA, que são mobilizados pela maquinaria dos LINEs (Kramerov and Vassetzky 2005; Kramerov and Vassetzky 2011). Cópias individuais de SINES podem ser caracteres “embaralhados” de diferentes subfamílias de SINEs, ou seja, SINEs tem sua evolução intrisicamente relacionada com a fusão e/ou acréscimo de sequências adicionais. Este tipo de evento foi documentado em mamíferos, via fusão e acréscimo de sequências adicionais (Kramerov and Vassetzky 2011).

Outro exemplo muito interessante de quimerismo e evolução modular são dos retroelementos de LTR. Provavelmente os retrotransposons de LTRs originaram-se da fusão quimérica entre os retransposons não-LTR e transposons de DNA com DDE. Alguns indícios corroboram essa hipótese, como a reconstrução filogenética utilizando o domínio da RNase H (RH) mostrou que os domínios da RH de LTR e não-LTR retrotransposon se agrupam, sugerindo origem evolutiva em comum (Malik and Eickbush 2001; Malik 2005). Essa reconstrução filogenética é corroborada pelas filogenias utilizando o domínio da RT. Além disso, o enraizamento da árvore com sequências RH derivadas do hospedeiro sugerem que os elementos não- LTR têm origem anterior aos elementos de LTR, (Malik and Eickbush 2001; Malik 2005). Os dados atualmente apontam para um modelo em que os elementos LTR surgiram através da fusão de um retrotransposons não LTR com um DNA transposon. Analisando a ORF2 (ORF *pol)* dos retrotransposon de LTR, há indícios de que esses elementos se formaram pela fusão dos transposons de DNA e não-LTR retrotransposons. Os transposons de DNA forneceram a integrase (IN), e os não-LTR retrotransposons fornceram os outros domínios da ORF2 (Malik 2005).

Impacto dos Elementos de transposição nos genomas

Os TEs são geralmente encontrados em regiões não codificadoras, contudo também podem ser observados junto a íntrons ou mesmo éxons. Eles podem por exemplo: fornecer promotores alternativos para expressão de genes;

quando inseridos em genes podem inativá-los; viabilizar variantes de proteínas através do *splicing*; ter ação epigenética quando modificam a expressão de um gene sem modificar sua sequência; e ocasionar rearranjos cromossômicos (Carareto *et al*. 2015). Bourque *et al*. (2018) organizaram em um artigo de revisão, chamado de "*Ten things you should know about transposable elements*" (em tradução livre: Dez coisas que você deve saber sobre elementos de transposição"), os dez itens estão reunidos resumidamente na Figura 2. Muitas dessas "coisas" estão relacionadas ao impacto dos TEs nos genomas hospedeiros.

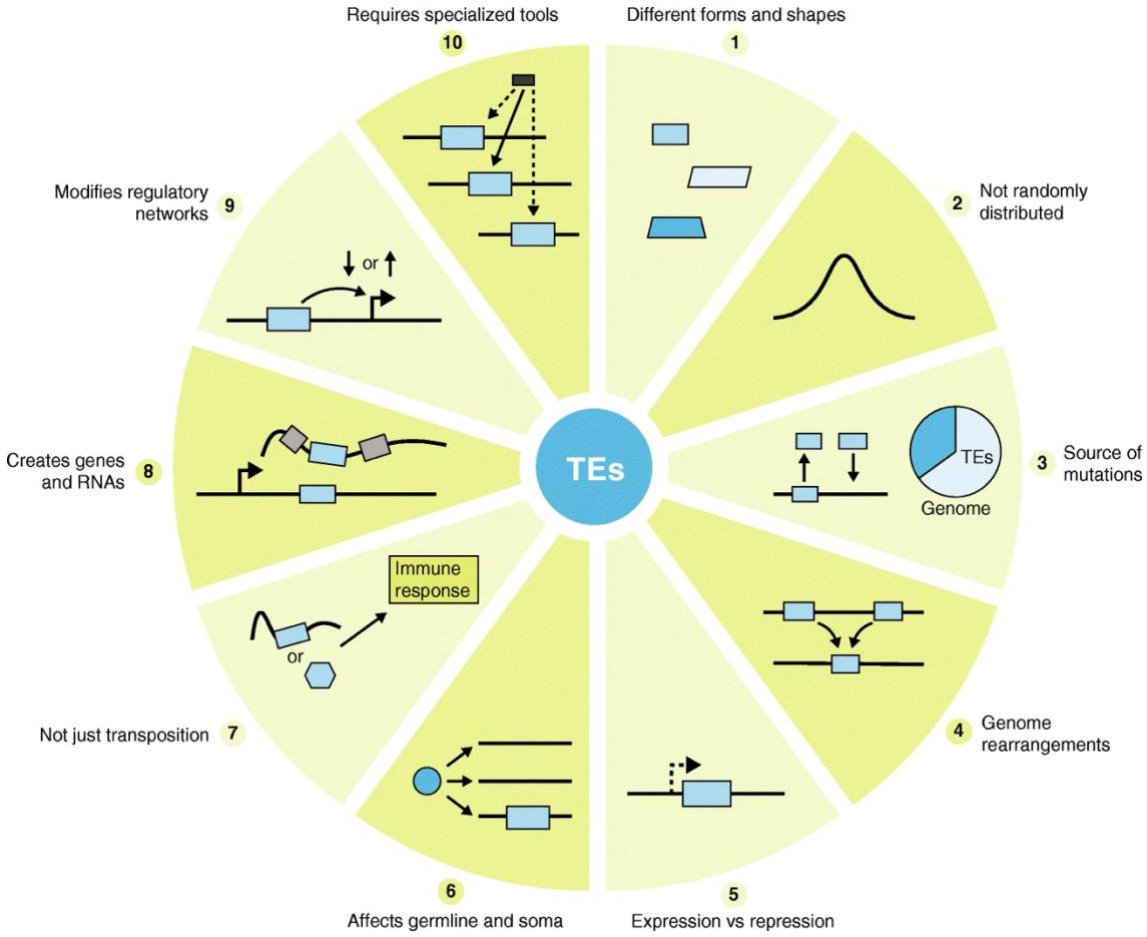


Figura 2: Como os TEs podem impactar os genomas hospedeiros. Os TEs: 1 – possuem formas e formatos diferentes 2- não estão distribuídos aleatoriamente; 3- são fonte de mutações; 4- promovem rearranjos do genoma; 5- influenciam a expressão vs repressão; 6- afetam as células germinativas e as somáticas; 7- não interferem somente por meio da transposição (resposta imune); 8- criam genes e RNAs; 9- modificam as redes regulatórias; e 10- requerem ferramentas especializadas. Adaptado de Bourque *et al*. (2018).

A distribuição dos TEs varia drasticamente entre os genomas eucariotos, e a contínua caracterização de novos genomas demonstra cada vez mais essa variação, e com isso, é possível inferir que os TEs não estão distribuídos randomicamente nos genomas hospedeiros. Como ilustrado na Figura 3, o tamanho do genoma, a quantidade de TEs, e a quantidade de cada tipo de TEs varia absurdamente entre os genomas eucariotos.

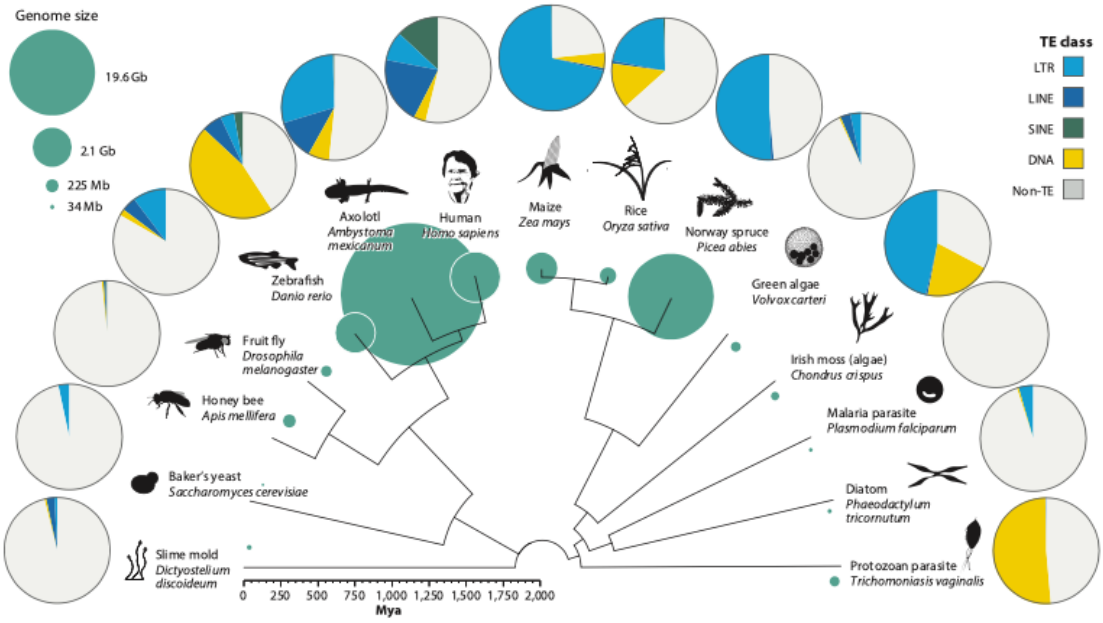


Figura 3: Distribuição dos TEs nos genomas eucarióticos. Círculos verdes representam o tamanho total dos genomas. Adaptado de Wells and Feschotte (2020).

Assim, TEs são uma extensa fonte de mutações e polimorfismos genéticos. Sítios de inserção podem variar drasticamente de um genoma a outro. Essa variação pode ocorrer entre espécies evolutivamente muito próximas como: *D. melanogaster e D. simulans,* onde a maioria dos TEs numa espécie está ausente no sítio ortólogo, e a maioria não está fixada na população (Kofler *et al*. 2015; Rahman *et al*. 2015). Há diferenças substanciais também em diferentes populações de uma mesma espécie, como observado em diferentes linhagens de *D. willistoni* por (Regner *et al*. 1996), como também há melhor descrição dessas variações no Capítulo 3 da tese. Os TEs estão associados a rearranjos do genoma e a características cromossômicas únicas: um exemplo são os telômeros, que em *Drosophila*, são constituídos de retrotransposons repetidos em *tandem* (George *et al*. 2006).

Existe um equilíbrio intrínseco entre a expressão de TE e a repressão, ou seja, é importante para o TEs proliferar-se, contudo, evitando afetar negativamente o hospedeiro. Assim, tanto os TEs têm sistemas de auto-regulação, como também os genomas hospedeiros regulam as populações de TEs (revisado em Carareto *et al*. 2015). Outro efeito resultante deste equilíbrio entre expressão e repressão é que os efeitos dos TEs no hospedeiro podem variar consideravelmente entre os tipos de tecidos e estágios de vida de um organismo (Bourque *et al*. 2018). Além disso, do ponto de vista do TE, o ideal é ser expresso e ativo nas células germinativas, e não nas células somáticas, onde a expressão seria desvantajosa ao TE (Haig 2016). Os TEs podem ser mutagênicos em células germinativas ou somáticas, podendo assim serem causadores de doenças, devido diretamente à sua mobilização no genoma ou indiretamente devido às interações com restante do genoma hospedeiro. Há mais de 100 inserções de TE associadas a doenças humanas (Hancks and Kazazian 2016).

Como geradores de novidades genômicas, os TEs também são fonte de variabilidade de RNAs codificantes ou não-codificantes. As inserções de TEs podem fornecer matéria-prima para o surgimento de genes codificadores de proteínas e RNAs não codificantes, que podem assumir funções celulares importantes e, em alguns casos, essenciais (Jurka *et al*. 2007; Naville *et al*. 2016; Joly-Lopez and Bureau 2018). Assim, os TEs também podem ser cooptados pelo hospedeiro, processo chamado de “domesticação” ou “exaptação” do gene do TE para exercer funções úteis ao hospedeiro. Exemplo de funções profundamente conservadas é o caso das telomerases evolvidas na estruturação dos telômeros, e por consequência dos cromossomos, além de outras características evolutivamente mais recentes (Capy 2021).

Os TEs são agentes capazes de modificar processos biológicos criando novos circuitos reguladores *cis* e ajustando redes pré-existentes (Bourque *et al*. 2018), sendo, assim, fonte de material para a modulação da expressão gênica eucariótica. Os TEs podem fornecer, por exemplo, promotores, *enhancers* e sítios de ligação para fatores de transcrição (Chuong *et al*. 2017). Os TEs podem ter fornecido “os blocos de construção” para a montagem e remodelação de rede

de regulação *cis* durante a evolução, como no desenvolvimento do neocórtex, entre outros exemplos (Revisão em Bourque *et al*. 2018)

Ciclo de vida dos elementos de transposição

O sucesso dos TEs estarem ubiquamente presentes nos genomas dos eucariotos está associada a capacidade dos mesmos de invandir novos genomas hospedeiros, por meio do processo chamado de Transfêrencia Horizontal de Elementos de Transposição (HTT, do inglês *Horizontal transposon transfer*). Na Figure 4, estão ilustradas as possíveis etapas de um TE no genoma hospedeiro.

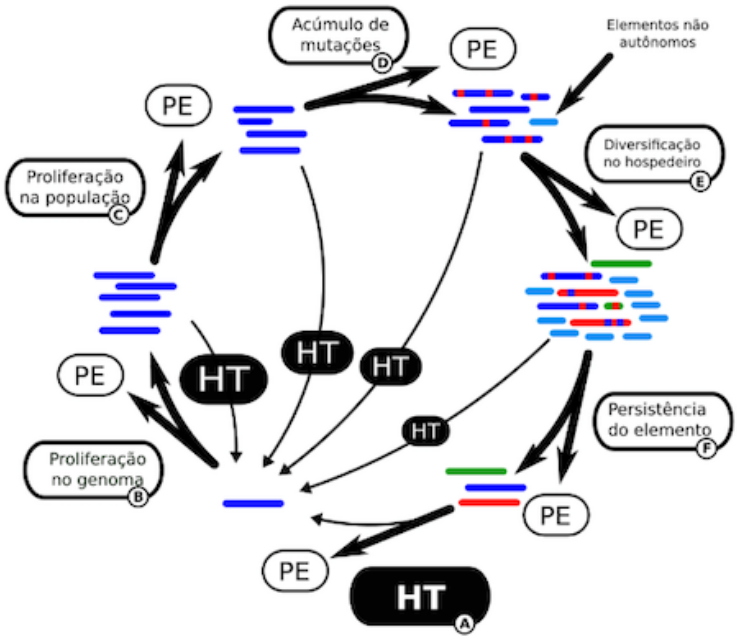


Figure 4: Ciclo de vida de um TE nos genomas eucaróticos. As letras de A-F indicam os principais passos dos TEs nos genomas; PE, significa “perdido” e HT, transferência horizontal. Adaptado de Wallau (2013)

Resumidamente, como descrito por (Schaack *et al*. 2010; Wallau *et al*. 2012), um evento de HTT pode ser considerado o início ou o fim do ciclo de vida do TE no genoma. Então, após a invasão em um novo genoma, há a proliferação do TE no genoma hospedeiro. Durante a proliferação, o TE pode se dispersar pela população ou espécie hospedeira. E, ao longo do tempo esse TE poderá acumular mutações. Nessa etapa, pode ocorrer o surgimento dos elementos MITEs, como consequência, o TE pode diversificar-se, e também aumentar o

número de cópia dos MITEs, e por fim, algumas famílias do TE podem sobreviver no genoma hospedeiro, e/ou sofrerem um evento de HTT. Contudo, em todos os passos do ciclo o TE pode ser perdido ou sofrer HTT.

II – Objetivos

Objetivo Geral

Contribuir para o conhecimento sobre a presença, dinâmica e co-evolução dos elementos de transposição e nos genomas hospedeiros, em espécies neotropicais de *Drosophila.*

Objetivos específicos

- Compilar os principais estudos evolutivos utilizando *Drosophila willistoni* e espécies do grupo como organismo modelo.

- Analisar comparativamente a presença, a estrutura, o número de cópias e localização de três transposons da superfamília *hAT* em diferentes linhagens de *Drosophila willistoni* procedentes de locais distantes ao longo da distribuição geográfica da espécie.

- Contribuir para elucidar a história evolutiva da linhagem *412/mdg1* do grupo *Ty3/gypsy* de retrotransposons com LTR em Diptera e caracterizar as sequências nos genomas, com ênfase nas espécies do grupo *willistoni* de *Drosophila.*

CAPÍTULO 3

Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach

Short title: *hAT* transposons in *D. willistoni*

Natasha Avila Bertocchi[1], Thays Duarte de Oliveira[2], Maríndia Deprá[1,2], Beatriz Goñi[3] and Vera Lúcia S. Valente[1,2*]

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2]Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[3]Universidad de la República, Montevideo, Uruguay

*corresponding author: Universidade Federal do Rio Grande do Sul (UFRGS) – Campus do Vale, Avenida Bento Gonçalves, 9500 Agronomia, CEP 91501-970, Departamento de Genética, prédio 43323 Laboratório de *Drosophila*, sala 210. E-mail: vera.gaiesky@ufrgs.br, ORCID: https://orcid.org/0000-0001-8661-6284

Artigo publicado na revista ***Genetics and Molecular Biology***

Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach

Natasha Ávila Bertocchi[1], Thays Duarte de Oliveira[2], Maríndia Deprá[1,2], Beatriz Goñi[3] and Vera Lúcia S. Valente[1,2]

[1]*Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Genética e Biologia Molecular, Porto Alegre, RS, Brazil.*
[2]*Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Biologia Animal, Porto Alegre, RS, Brazil.*
[3]*Universidad de la República, Facultad de Ciencias, Montevideo, Uruguay.*

Abstract

Transposable elements are abundant and dynamic part of the genome, influencing organisms in different ways through their presence or mobilization, or by acting directly on pre- and post-transcriptional regulatory regions. We compared and evaluated the presence, structure, and copy number of three *hAT* superfamily transposons (*hobo, BuT2*, and *mar*) in five strains of *Drosophila willistoni* species. These *D. willistoni* strains are of different geographical origins, sampled across the north-south occurrence of this species. We used sequenced clones of the *hAT* elements in fluorescence *in-situ* hybridizations in the polytene chromosomes of three strains of *D. willistoni*. We also analyzed the structural characteristics and number of copies of these *hAT* elements in the 10 currently available sequenced genomes of the *willistoni* group. We found that *hobo, BuT2,* and *mar* were widely distributed in *D. willistoni* polytene chromosomes and sequenced genomes of the *willistoni* group, except for *mar*, which is restricted to the subgroup *willistoni.* Furthermore, the elements *hobo, BuT2*, and *mar* have different evolutionary histories. The transposon differences among *D. willistoni* strains, such as variation in the number, structure, and chromosomal distribution of *hAT* transposons, could reflect the genomic and chromosomal plasticity of *D. willistoni* species in adapting to highly variable environments.

Keywords: Transposable elements, *Drosophila willistoni*, *hAT* superfamily, polytene chromosomes.

Received: September 10, 2021; Accepted: January 31, 2022.

Introduction

Transposable elements (TEs) constitute part of the repetitive fraction of the genome and can move within and between host genomes. TEs are thought to be present in virtually all genomes and are best studied in the genus *Drosophila* (Diptera: Drosophilidae) (Wicker *et al.*, 2007). TEs are considered generators of evolutionary novelty, as they can interact with host genomes in a variety of ways, although they were previously characterized as junk DNA. They can be found close to regulatory regions over- or under-expressing genes, as constituents of heterochromatin, and may increase the propensity to chromosomal variations, among other possible roles such as an inducer of cancers (Cáceres *et al.*, 2001; Catania *et al.*, 2004; Bertocchi *et al.*, 2018)

The proposed life cycle of TEs can be summarized as: insertion by Horizontal Transposon Transfer (HTT) or reactivation in the host genome; increase in copy number (proliferation) and dispersal in the host population; and, over time, accumulation of mutations (diversification) (Wallau *et al.*, 2012). Sexual reproduction eventually allows TEs to be distributed in most individuals of a population and/or species. At any stage of the cycle, a TE can be lost by the genome or, to a lesser extent, undergo HTT and restart the cycle (Schaack *et al.*, 2010; Wallau *et al.*, 2012). HTT has been shown to perpetuate TEs in host genomes, and HTT events are increasingly identified in the most varied groups of eukaryotes (Wells and Feschotte, 2020).

As classified by Wicker *et al.* (2007), TEs are divided hierarchically, first into two classes according to the transposition mechanism: class I via intermediary RNA (retrotransposons) and class II via intermediary DNA (transposons). Class II elements, termed transposons, use the enzyme transposase for mobilization; they are subdivided into two subclasses according to the number of DNA strands that are cleaved in the transposition process. Subclass 1 elements cleave the two strands of DNA by a "cut-and-paste" mechanism, and subclass 2 elements cleave only one of the strands, which has other transposition mechanisms. TEs are then classified into orders, superfamilies, families, and subfamilies according to their structural characteristics and conservation of nucleotide and amino-acid sequences.

TEs may also be classified according to their autonomy for mobilization. TEs can be autonomous, that is, possess the entire enzymatic structure needed to carry out their own mobilization; or non-autonomous, when they need the enzymatic machinery of other autonomous TE copies to

Send correspondence to Vera Valente. Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Genética e Biologia Molecular, Avenida Bento Gonçalves, 9500, 91501-970, Prédio 43323, Laboratório de *Drosophila*, sala 210, Porto Alegre, RS, Brazil. E-mail: vera.gaiesky@ufrgs.br.

mobilize. An example of Class II non-autonomous elements are termed miniature inverted-repeat transposable elements (MITEs) (González and Petrov, 2009). MITEs are cross-mobilized by autonomous elements, as they generally conserve the recognition sequences for transposases, the TIRs. They can also be found in high copy numbers in genomes (Deprá *et al.*, 2012; Loreto *et al.*, 2018).

The *hAT* superfamily is present in animals, plants, and fungi. It is subdivided into three families: *Ac*, *buster*, and *tip* (Arensburger *et al.*, 2011; Zhang *et al.*, 2013; Rossato *et al.*, 2014). Elements of the *hAT* superfamily have an 8 bp target site duplication (TSD) and short Terminal Inverted Repeats (TIRs) between 10–25 bp and 2.5–5 kb in size (Feschotte and Pritham, 2007). The elements *hobo*, *BuT2*, and *mar* belong to the *Ac*, *tip*, and *buster* families respectively (Deprá *et al.*, 2012; Rossato *et al.*, 2014). The canonical *hobo* (HFL1) was initially described in *Drosophila melanogaster* and consisted of 2959 bp length, encoding a 1.9-kb transposase gene, and 12 bp of TIRs (Calvi *et al.*, 1991). *Hobo* was originally described to be limited to the *melanogaster* subgroup (Ortiz and Loreto, 2008; reviewed in Loreto *et al.*, 2018). *BuT2* is 2775 bp long and encodes a 643 aa transposase and 12 bp of TIRs (Rossato *et al.*, 2014). *BuT2* was initially described in *Drosophila buzzatii*, in regions of inversion breakpoints, which indicates a recent mobilization, although it is only sparsely present in the genome of this species (Cáceres *et al.*, 2001; Casals *et al.*, 2006). The canonical *mar*-MITE element was originally identified in *D. willistoni* and has 610 bp and 11 bp of TIRs. *Mar* is restricted to the *willistoni* subgroup, and until now partially complete copies have been found only in *Drosophila tropicalis*, with approximately 2600 bp (Holyoake and Kidwell, 2003; Deprá *et al.*, 2012).

Dobzhansky (1950) described the first polytene photomap of *D. willistoni*, this map was further redrawn in Valente and Araújo, 1985, Regner *et al.*, 1996, Bhutkar *et al.*, 2008, and Rohde and Valente, 2012. The chromosome complement of *D. willistoni* consists of two pairs of metacentric chromosomes (IIL, IIR, XL, and XR arms), an acrocentric pair (III arm), and a Y submetacentric chromosome (Dobzhansky and Powell, 1975; Santos-Colares *et al.*, 2003). *D. willistoni* is notable for having multiple chromosomal inversions in every natural population examined (review by Rohde and Valente, 2012).

The first sequenced genome of *D. willistoni was*that of strain Gd-H4-1, the result of several generations of sister-brother crosses to obtain a strain without segregating inversions, i.e., a monokaryotypic strain (Drosophila 12 Genomes Consortium *et al.*, 2007). Strain Gd-H4-1 lacks the high degree of polymorphism and variability found in natural populations of this widely distributed tropical species (review by Zanini *et al.*, 2015). Two additional strains of *D. willistoni* were recently sequenced by Kim *et al.* (2021), who found considerable differences between these strains in the number of repetitive sequences such as transposons and microsatellite elements.

Our research group has been studying several aspects of the chromosomal plasticity of *D. willistoni* (Valente and Araújo, 1985; Valente *et al.*, 1993; Valente *et al.*, 2003; Rohde and Valente, 2012; Garcia *et al.*, 2015). The goal of the present study was to contribute to understanding the high degree of variability of *D. willistoni* over its wide geographical distribution. In view of the significant environmental differences encountered by this species, the chromosomal variations characteristic for *D. willistoni* strains, and the differences found in the number of repetitive fractions in different strains, we compared and characterized the organization and distribution of three transposable elements of the *hAT* superfamily in different *D. willistoni* strains. Studies such as this can clarify how different habitats are capable of promoting evolutionary changes in TEs and hosts.

Material and Methods

Fly stocks and chromosomal preparations

Three strains of *D. willistoni* were used in this study (Table S1). These strains have been maintained in the laboratory by mass crosses and cultivated in cornmeal culture medium (Marques *et al.*, 1966) under controlled temperature (20 ± 1 °C). The polytene chromosome preparations were obtained with third-instar larval salivary glands, squashed, and fixed in 2:1:2 ethanol–lactic acid–acetic acid, v/v.

Probe preparation and fluorescence *in-situ* hybridization (FISH)

TE clones were used as a template for the PCR labeling probe for FISH: *BuT2* in *D. willistoni* (GenBank accession number KF669641.1) obtained from Rossato *et al.* (2014); *mar_trop*: sequence of the *mar* element of *D. tropicalis* (GenBank accession number JQ654772.1) (obtained from Deprá *et al.*, 2012); and *hobo* in *D. willistoni* (submitted GenBank accession number OK032551, this study). For this last, genomic DNA from strain Gd-H4-1 was used to amplify the *hobo* transposon. The primers used were *hobo* CN 991 (5′-ACCGTCGACATGTGGAC-3′) and *hobo* CN 1598 (5′-GGATGGCAATAGGAAGC- 3′) (Deprá *et al.*, 2009). The amplified sample was visualized on 0.8% agarose gel. The bands were purified using the GFX Purification Kit (GE Healthcare) and cloned using the TOPO-TA cloning vector (Invitrogen, Carlsbad, CA, USA). Cloned PCR products were sequenced using the universal primers M13 (forward and reverse) at Macrogen (Korea).

The TE probes *BuT2*, *mar_trop*, and *hobo* were marked directly by PCR, using Biotin-16-dUTP (Jena Bioscience). Slide preparations, hybridizations, and washes were performed according to Deprá *et al.* (2010), with minor modifications. FISH experiments were established in 77% of the stringency. The signal was detected using streptavidin-Cy3 and the chromosomes were counterstained with Fluoroshield with DAPI. The slides were analyzed using the epifluorescence microscope ZEISS Axiophot (Zeiss, Germany). The images were captured using Zeiss ZEN (blue edition) software. The final editing of the images used Adobe Photoshop CS6. The hybridization signals were quantified by visual inspections and using the ImageJ software (Schneider *et al.*, 2012). The following premises were applied to measurements of the hybridization signals: area less than 9.99 μm^2 as non-hybridization borderline, and an area larger than 10 μm^2 for each hybridization on the five chromosome arms (XR, XL, IIR, IIL, and III). In the chromocenter, we considered the presence or absence of a hybridization signal.

Genome searches

Searches for homologous sequences to *BuT2*, *hobo*, and *mar* were carried out in the genomes of the species of the *willistoni* group available in NCBI and Kim *et al.* (2021), last accessed in January 2021. Versions of the assemblies, species, and strains used in this study are available in Table S2.

The queries used were: *mar* sequence from *D. tropicalis* (GenBank accession number JQ654772.1), *BuT2* (GenBank accession number KF669641.1), and *hobo* (GenBank accession number OK032551) from *D. willistoni* available in NCBI. BLASTn searches were performed on the Galaxy platform, using default parameters (Afgan *et al.*, 2016). The sequences with an E-value lower than e^{-10} were extracted for each genome.

Sequence analysis

The sequence alignments were performed using MAFFT (Katoh and Standley, 2013), with default parameter values. AliView (Larsson, 2014) was used for sequence editing and visualization. *Mar* sequences are very variable in copy number, length, and structure, and therefore the alignment for phylogenetic reconstruction of the *mar* copies was submitted to two refinement steps: 1) copies with 100% of the identity in each genome were filtered by the CD-HIT Suite (Huang *et al.*, 2010); and subsequently, 2) the alignment was manually inspected to exclude small and/or very degenerate sequences. All *mar* sequences after refinement (MITEs, relics, complete and partially complete) were used in phylogenetic reconstruction, except: Dwil_Gd_scf2_3; Dins_ctg2309_5, Dins_ctg424, Dins_ctg1175, Dins_ctg1948; Dtro_ctg108_3, Dtro_ctg108_4, Dtro_ctg838, Dtro_ctg804, Dtro_ctg19. All sequences of *hobo* and *BuT2* retrieved were used in the phylogenetic trees.

The phylogenetic trees were inferred by Bayesian Analysis in MrBayes 3.2.6. implemented in the CIPRES gateway (Miller *et al.*, 2010; Ronquist *et al.*, 2012). The evolutionary models GTR+G (*hobo*), HKY+G (*BuT2*), and JC+I (*mar*) were indicated by MrModeltest2 (Nylander, 2004). The analysis was run for at least 10,000,000 generations, sampling trees every 1,000 generations, with 25% of the initial results as burn-in. MEGAX (Kumar *et al.*, 2018) was used to measure the divergence of the sequences by p-distance and Neighbor-Joining phylogenetic reconstruction for the *mar* sequences (data not shown).

Also, we performed a phylogenetic reconstruction of the *D. willistoni hobo* in the *hAT* superfamily, using the Maximum Likelihood method and Le-Gascuel model (LG) (Le and Gascuel, 2008) by MEGA X (Kumar *et al.*, 2018), with the transposase database based on Arensburger *et al.* (2011) and Rossato *et al.* (2014). The transposase sequences were aligned by MUSCLE, implemented in MEGA X.

Results

FISH of the *BuT2*, *hobo*, and *mar* elements in polytene chromosomes of *D. willistoni* strains

For the FISH experiments, we used polytene chromosomes from the three strains of *D. willistoni* from different geographic locations. The strains were: *D. willistoni*-Gd-H4-1, an inbred lineage; *D. willistoni*-WIP-4, descended from a natural population maintained in the laboratory for approximately 60 years and considered by us a standard karyotype for the species; and a natural population, *D. willistoni*-SG12.00, collected in the 2000s in Montevideo (Figure 1 and Table S1). Clear differences were detected in the number and distribution of signals along the chromosomal arms of these strains (Figure 2A-I).

The three probes used were derived from clones of TEs *BuT2*, *hobo*, and *mar*, and were termed *BuT2*, *hobo*, and *mar_trop*, respectively. FISH experiments with the *BuT2* probe revealed differences among the strains in the distribution and number of signals. In *D. willistoni*-Gd-H4-1, visually many strong signals were detected along all chromosomes and the chromocenter (Figure 2B), while in *D. willistoni*-WIP-4 visually strong signals were observed on the IIR and IIL chromosome arms (Figure 2E). In *D. willistoni*-SG12.00, only two stronger signals of *BuT2* hybridization signals were visible on the IIR and IIL arms (Figure 2H), and some signals were detected also in the chromocenter. We noted a pattern in the production of signals according to the geographic origin of the strains; the northernmost strain (from above the Equator; Figure 1) had more signals and more intense signals than the other, more southern strains (Figure 1).

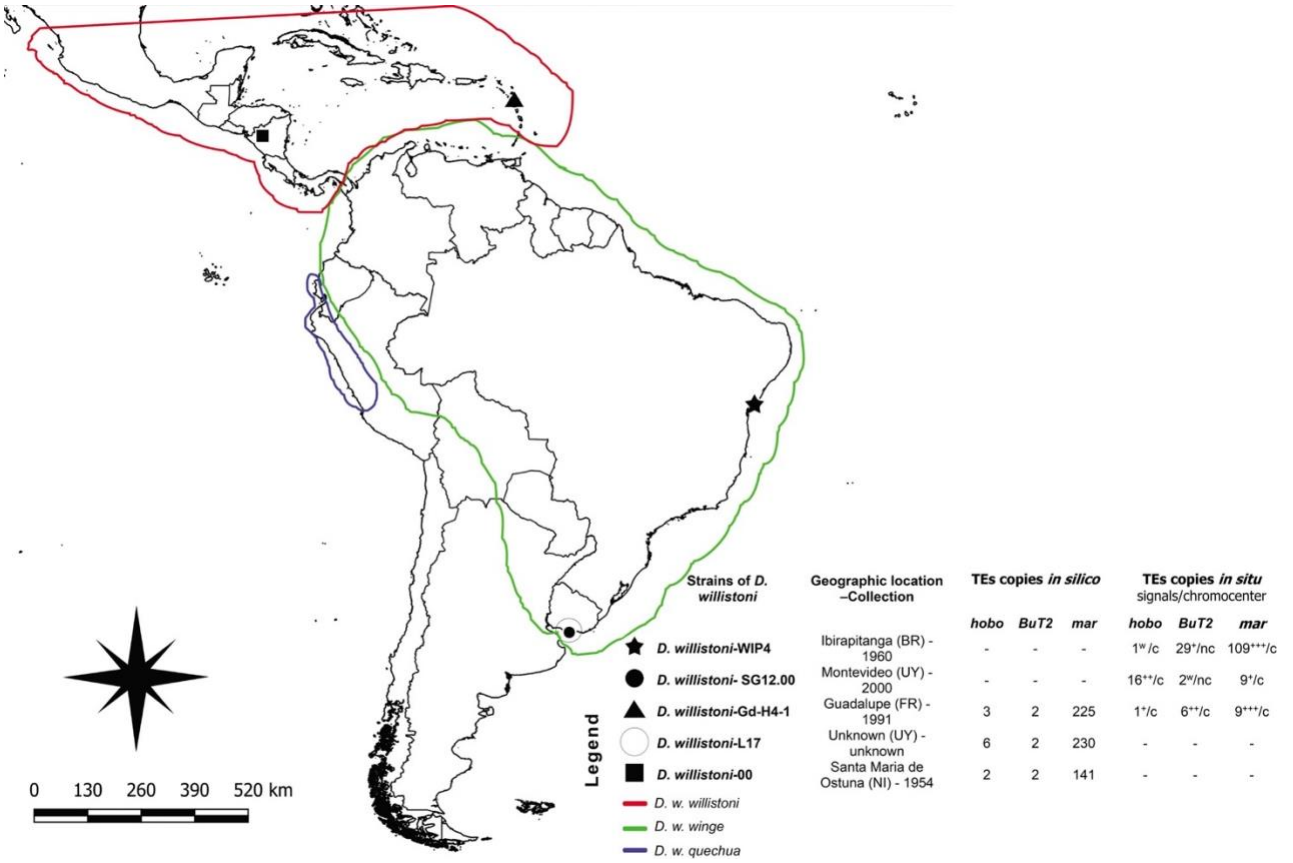
With the *hobo* probe, the pattern was almost the opposite of that seen for *BuT2*: the strain from the extreme southern part of the distribution (*D. willistoni*-SG12.00 - Figure 1) showed many stronger signals on all chromosome arms, mainly in the euchromatin and chromocenter (Figure 2G). *D. willistoni*-Gd-H4-1 and *D. willistoni*-WIP-4 showed one stronger signal on the IIR arm, and we also observed more signals with less intensity in the *D. willistoni*-Gd-H4-1 (Figure 2A, D).

Concerning the *mar_trop* probe, *D. willistoni*-WIP-4 showed many stronger signals in all chromosome arms and the chromocenter (Figure 2F). Although the ImageJ software estimated around the same number of *mar* copies in the strains *D. willistoni*-Gd-H4-1 and *D. willistoni*-SG12.00 (Figure 1), differences between the two strains were apparent (Figure 2C and 2I), mainly concerning the intensity and distribution of the signals along the chromosomal complement. *D. willistoni*-Gd-H4-1 showed stronger signals along the five chromosome arms and the chromocenter, while *D. willistoni*-SG12.00 showed signals on the chromocenter and on the arms near the chromocenter, with no signals observed on the III chromosome.

Transposons *in-silico* search in *Drosophila willistoni* group genomes

hobo search

The cloned fragment of the element *hobo* from *D. willistoni*-Gd-H4-1 contained 439 bp and was 74.7% identical to that of the *D. melanogaster* canonical *hobo* (Calvi *et al.*, 1991). *D. willistoni-hobo* alignments were mainly between nucleotide positions 991 and 1428 of the canonical *hobo* element. The BLASTn search showed that the *D. willistoni-hobo* fragment presented 93% identical to the *hobo* element of the Mediterranean fruit fly *Ceratitis capitata* (Diptera: Tephritidae) (Cc-HRE-GenBank access number U51454.1) (Handler and Gomez, 1996). To establish the relationship between the *D. willistoni-hobo* and *Cc-HRE* (*C. capitata)* putative transposase and the *hAT* superfamily elements, we



Strains of *D. willistoni*	Geographic location –Collection	TEs copies *in silico*			TEs copies *in situ* signals/chromocenter		
		hobo	*BuT2*	*mar*	*hobo*	*BuT2*	*mar*
D. willistoni-WIP4	Ibirapitanga (BR) - 1960	-	-	-	1^{w}/c	29^{+}/nc	109^{+++}/c
D. willistoni- SG12.00	Montevideo (UY) - 2000	-	-	-	16^{++}/c	2^{w}/nc	9^{+}/c
D. willistoni-Gd-H4-1	Guadalupe (FR) - 1991	3	2	225	1^{+}/c	6^{++}/c	9^{+++}/c
D. willistoni-L17	Unknown (UY) - unknown	6	2	230	-	-	-
D. willistoni-00	Santa Maria de Ostuna (NI) - 1954	2	2	141	-	-	-

Figure 1 – Geographical origins of the *Drosophila willistoni* strains analyzed *in silico* and *in situ*, and information about *hAT* TE copies. Lines indicate the approximate geographical distributions of the three subspecies of *Drosophila willistoni* (Mardiros *et al.*, 2016). The numbers of TE copies in polytene chromosomes were measured by ImageJ software (Schneider *et al.*, 2012) and visually. The ordinal number represents stronger signals, and increasing from + to +++ indicate the relative strength of intensity of signals on chromosome arms, as detected visually on polytene chromosome arms. In the table: - indicates absence of information; c and nc indicate presence and absence of signals on the chromocenter, respectively; w indicates weak signals; +,++ and +++, increasing from + to +++ indicate the relative strength of intensity of signals detected visually on polytene chromosomes by FISH.

assembled the transposase sequences described by Arensburger *et al.* (2011) and Rossato *et al.* (2014). A phylogenetic reconstruction of the *hAT* superfamily showed that *D. willistoni*-*hobo* and Cc-HRE (*C. capitata*-*hobo*) were grouped with *Ac* family elements (Figure S1). These formed a clade with *Howilli2* (*D. willistoni*), *Cc-HRE* (*Ceratitis capitata*), *Homo1* (*D. mojavensis*), canonical *hobo* (*D. melanogaster*), *Hermes* (*Musca domestica*; Diptera: Muscidae), *Homer* (*Bactrocera tryoni*; Diptera: Tephritidae), *Hoana1* (*Drosophila ananassae*), *Hoana8* (*D. ananassae*), *Hermit* (*Lucilia cuprina*; Diptera: Calliphoridae), and *Hoana3* (*D. ananassae*).

Using the *hobo* sequence obtained here (cloned element from *D. willistoni* strain Gd-H4-1), we performed BLASTn against the 10 sequenced genomes belonging to seven species of the *willistoni* group (Table S3). Sequences homologous to the *hobo* fragment from *D. willistoni* were identified in the seven species of the *willistoni* group: *D. willistoni* (three strains), *D. paulistorum* (two strains), *D. equinoxialis*, *D. tropicalis*, *D. insularis*, *D. sucinea*, and *D. nebulosa* (Figure 3 and Figure 4A). In a search for complete copies of *hobo* in these genomes, we recovered homologous sequences and added 3000 bp from the *hobo* on each end. However, no complete copies were identified (codifying transposase and TIRs at the ends). A schematic representation of these sequences is shown in Figure 4A.

With respect to the *willistoni* subgroup, in the genomes of *D. willistoni*-Gd-H4-1, *D. willistoni*-L17, *D. willistoni*-00, *D. paulistorum*-L06, and *D. paulistorum*-L12 we identified the most complete copies of *hobo* (≈2850 bp), with small additions in the region of the transposase, 12 bp TIRs conserved and identical to the canonical *hobo* and TSDs (Figure 3, Figure 4A and Table S3). In the genomes of *D. willistoni*-Gd-H4-1, *D. paulistorum*-L06, and *D. paulistorum*-L12 we also observed smaller *hobo*-like fragments without TIRs at both ends (Figure 4A).

The *hobo*-like sequences retrieved from the *D. equinoxialis*, *D. tropicalis*, and *D. insularis* genomes are smaller fragments (Figure 3), conserved mainly in the 520 to 1720 bp region of canonical *hobo* transposase, without TIRs or conserved TSDs (Figure 4A and Table S3).

In the *bocainensis* subgroup, complete sequences of *D. sucinea* and *D. nebulosa* were not identified (Figure 3). Copy Dsuc_ctg141 in *D. sucinea* and copies Dneb_ctg3 and Dneb_ctg46 in *D. nebulosa* had identical canonical TIRs (Figure 4A and Table S4). In both species, TSDs were not present or were variable.

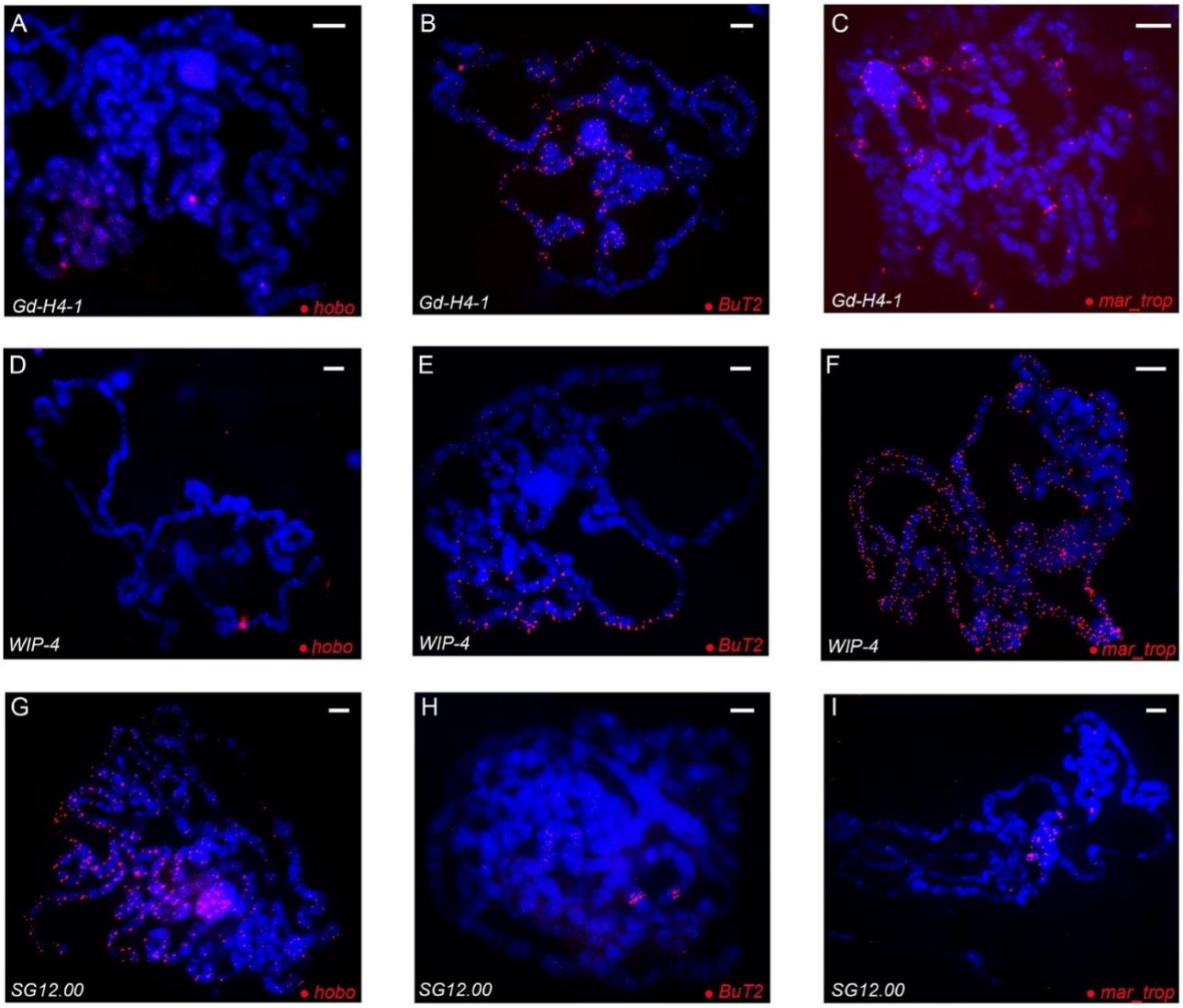


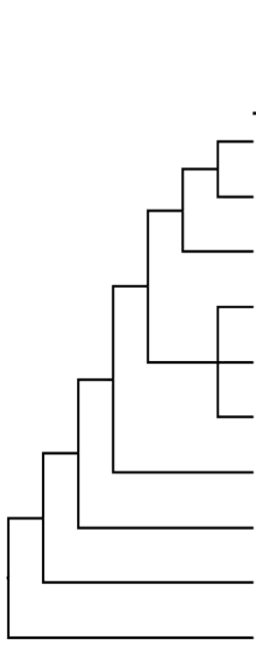
Figure 2 – FISH in polytene chromosomes of *Drosophila willistoni* strains: (A-C) *D. willistoni*-Gd-H4-1; (D-F) *D. willistoni*-WIP-4; and (G-I) *D. willistoni*-SG12.00. The probes used are indicated in the lower right corner and the strains in the lower left corner of the images. Chromosomes were counterstained with DAPI (blue) and transposable element probes were labeled with Cy3 (red). Scale bar=10 µm.

In order to address the average divergence of the *hobo* sequences found within and between species/strains, we evaluated the p-distance (Table S4). The intragenomic divergences in *D. willistoni* strains were 3.91% in *D. willistoni*-L17, 13.64% *in D. willistoni*-00, and 13.88% in *D. willistoni*-Gd-H4-1. Intragenomic divergence of 4.1% was observed in *D. paulistorum*-L12 and 7.98% in *D. paulistorum*-L6. The values of interspecies divergence ranged from 3.74% between *D. paulistorum*-L12 and *D. willistoni*-L17 to 13.05% between *D. tropicalis* and *D. willistoni*-Gd-H4-1. Figure 5 shows the Bayesian tree obtained for all *hobo* copies from the *willistoni* species group identified in this study. The phylogeny showed low resolution in several nodes, groupings with sequences of different species and subgroups, and some polytomies. One group was formed by *D. willistoni* strains, *D. paulistorum* strains, *D. nebulosa*, and *D. sucinea* copies. The recurrent grouping between sequences of *D. nebulosa* and *D. sucinea* was also evidenced. The relationships among the *willistoni* group species together with the branch lengths indicate that these sequences are very similar, likely with recent mobilization.

BuT2 search

Sequences homologous to the element *BuT2* were detected in the 10 genomes of the *willistoni* group analyzed; however, no complete TE copies were identified (Figure 4B). In the subgroup *willistoni*, partially complete copies of *BuT2* were identified in *D. willistoni* strains. However, in the *bocainensis* subgroup (*D. sucinea* and *D. nebulosa*), only one short partial *BuT2* fragment (764 bp) without TIRs was identified in both species (Figure 3).

In *D. willistoni*, two homologous *BuT2* sequences were identified in the sequenced strains. The most complete sequences, i.e., in *D. willistoni*-Gd-H4-1 with 2742 bp (Dwil_scf2_2), *D. willistoni*-L17 with 2695 bp (Dwil_ctg8), and *D. willistoni*-00 with 2737 bp (Dwil_ctg1698), were 91% identical to the *BuT2* element including 12 bp TIRs (Figure 4B and Table S5). Furthermore, the Dwil_scf2_2, Dwil_ctg8, and Dwil_ctg1698 sequences of *BuT2* in *D. willistoni* were flanked



Species/strain	Geographic location	Collection date	*hobo* Copy number	*hobo* Size range (bp)	*hobo* State	*hobo* TIRs	*hobo* TSD	*BuT2* Copy number	*BuT2* Size range (bp)	*BuT2* State	*BuT2* TIRs	*BuT2* TSD	*mar* Copy number	*mar* Size range (bp)	*mar* State	*mar* TIRs	*mar* TSD
D. paulistorum-L06	El Salvador	1955	4	1863	CP, DR	+	+	6	958	DR	+	+	286		DR	+	+
D. paulistorum-L12	Brazil	unknown	5	2542	CP, DR	+	+	4	1179	DR	+	+	232		DR, MITE	+	+
D. equinoxialis	Honduras	unknown	4	1103	DR	-	+	3	889	DR	+	+	183		DR, MITE	+	-
D. willistoni-Gd-H4-1	Guadalupe, Island in the Caribbean	1991	3	2030	CP, DR	+	-	2	1887	CP, DR	+	+	225		CP, DR, MITE	+	+
D. willistoni-L17	Uruguay	unknown	6	2711	CP, DR	+	+	2	1766	CP, DR	+	+	230		CP, DR, MITE	+	+
D. willistoni-00	Santa Maria de Ostuna, Nicaragua	1954	2	2458	CP, DR	+	+	2	1884	CP, DR	+	+	141		CP, DR, MITE	+	+
D. tropicalis	El Salvador	1955	5	1603	DR	-	+	5	1559	DR	+	+	12	1508	CP, DR	+	-
D. insularis	St. Lucia	unknown	4	1706	DR	-	+	1	2546	CP	+	-	24	1344	DR	+	-
D. nebulosa	Costa Rica: San Jose	1955	5	2044	DR	+	+	1	764	DR	-	-	-	-	-	-	-
D. sucinea	Honduras: Monte Ceyuca	1954	4	2016	DR	+	+	1	764	DR	-	-	-	-	-	-	-

Figure 3 – Information on species and evolutionary relationships of sequenced genomes of the *willistoni* group. Schematic evolutionary relationships among species of the *willistoni* group are based on Finet *et al.*, (2021). (-) Absence; (+) presence; State = Structural characteristics of TE; CP = Complete or Partially complete copies; DR = Degenerate or Relic copies; MITE = miniature inverted-repeat transposable elements.

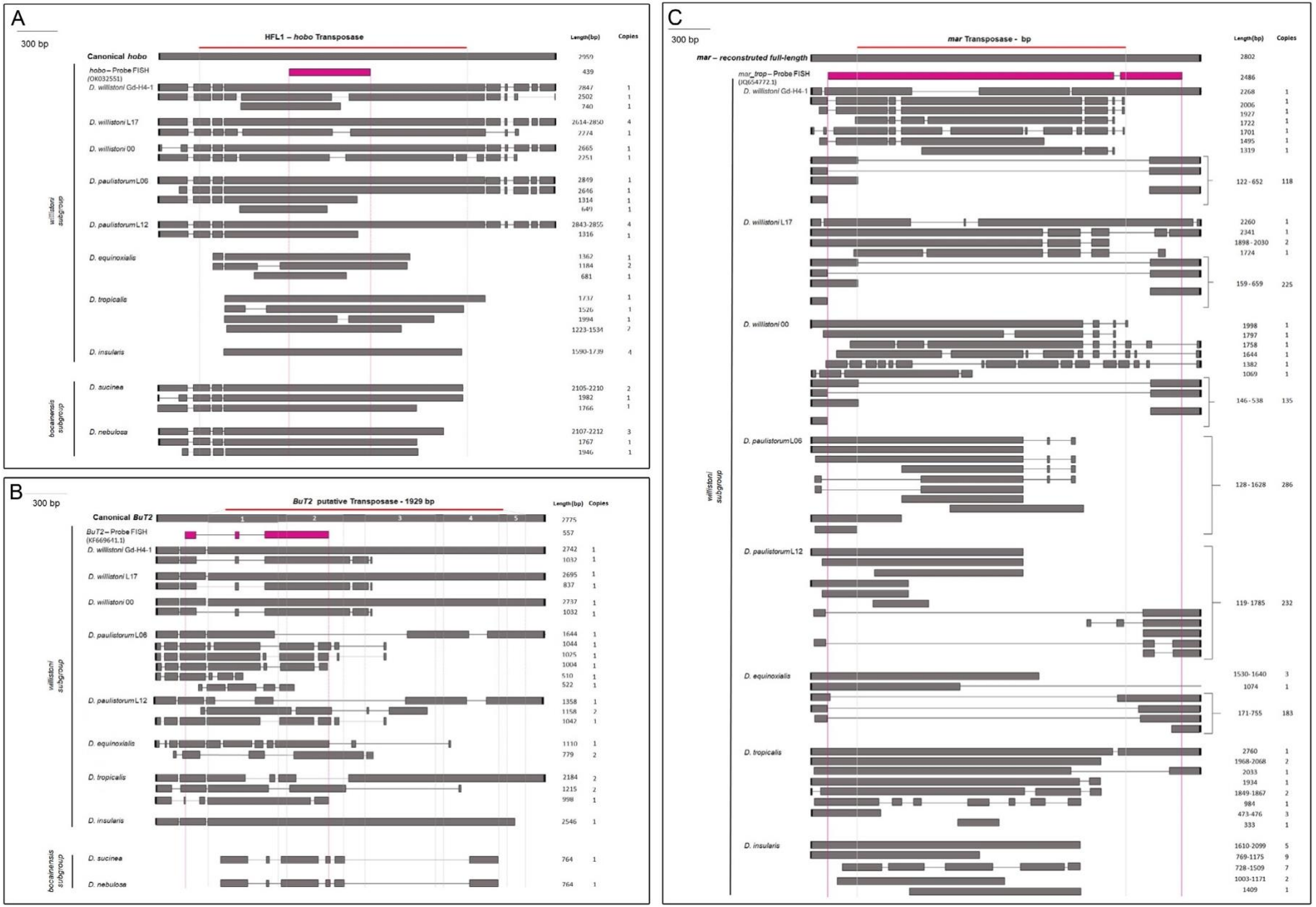


Figure 4 – Schematic representation of reconstructed *hobo, BuT2*, and *mar* copies in the *willistoni* group. (A) *hobo:* all sequences are represented; (B) *BuT2:* all sequences are represented and transposase is formed by 5 exons, indicated by descending ordinal numbers; (C) *mar*: *mar*-MITE and degenerate sequences in *D. willistoni*-Gd-H4-1, *D. willistoni*-L17, *D. willistoni*-00, *D. paulistorum*-L06, *D. paulistorum*-L12, and *D. equinoxialis* were grouped. Regions of terminal inverted repeats shown inside black block, transposase coding region inside red line, and probes used in FISH experiments inside pink block. Only indels and deletions of nucleotides with more than 10 bp are represented**.**

by 8 bp TSDs. In these three *D. willistoni* strains, the TSDs were conserved, with one mismatch in *D. willistoni*-Gd-H4-1 (Table S5). The other copies of *D. willistoni*-Gd-H4-1 (Dwil_scf2), *D. willistoni*-L17 (Dwil_ctg326), and *D. willistoni*-00 (Dwil_ctg675) were incomplete: without the TE initial region, with a size of 837–1032 bp, TIRs conserved in the TE 3' region, and with large deletions in the exon regions 1, 2 and 4 of transposase (Figure 4B).

BuT2 copies in *D. paulistorum*-L06, *D. paulistorum*-L12, *D. equinoxialis*, and *D. tropicalis* genomes were defective, with deletions in the five exons of *BuT2* transposase (Figure 3, Figure 4B, and Table S5). The 12 bp TIRs were conserved in *D. paulistorum*-L06, *D. paulistorum*-L12, and *D. tropicalis* copies (Figure 3 and Table S5). The *D. insularis* genome with one *BuT2* copy lacked the 263 bp 5' end of TE (Figure 4B).

The *BuT2* intragenomic divergence in the different *D. willistoni* strains ranged from 9.11% in *D. willistoni*-Gd-H4-1 and *D. willistoni*-00 to 10.29% in *D. willistoni*-L17. An intragenomic divergence of 13.91% was observed in *D. paulistorum*-L06, and 17.99% in *D. paulistorum*-L12. Interspecies divergence values ranged from 0.13% between *D. nebulosa* and *D. sucinea* to 53.13% between *D. sucinea* and *D. paulistorum-L06*. Table S6 shows the average divergence of the *BuT2* sequences found within and between the species and strains. All copies of *BuT2* retrieved in the sequenced genomes of the *willistoni* group were used to construct a phylogeny (Figure 6). Species of the *bocainensis* and *willistoni* subgroups formed two clusters with well-established relationships. The clade of the *willistoni* subgroup showed different groupings with high probability, with copies of *D. willistoni* strains, *D. paulistorum* strains, and all sequences of *D. tropicalis*. Two groups formed by copies of *D. willistoni* strains showed short branch lengths, which indicates that copies of different strains are very similar.

mar search

We started the search for homologous sequences to the *mar* element in the *willistoni* group genomes by using the query clone_8 from *D. tropicalis* (JQ654772.1), also used in the FISH experiments. *Mar* homologous sequences recovered in the genomes were aligned using the full-length *mar* reconstructed by Deprá *et al.* (2012) in order to obtain putative full copies.

We recovered *mar*-like sequences in *D. willistoni*-Gd-H4-1, *D. willistoni*-L17, *D. willistoni*-00, *D. paulistorum*-L06, *D. paulistorum*-L12, *D. equinoxialis*, *D. tropicalis*, and *D. insularis* genomes (Figure 3). The exact number of copies in *D. willistoni*, *D. paulistorum*, and *D. equinoxialis* strains was difficult to determine because the genome contains some small fragmented copies that were not captured in the searches. Also, the copy number is variable among the species. In the *bocainensis* subgroup (*D. sucinea* and *D. nebulosa*) no *mar*-like sequences were identified (Figure 4C).

Mar full-length copies or putatively active were recovered only from the *D. tropicalis* genome. Partially complete copies were observed in *D. willistoni*-Gd-H4-1 (7 copies), *D. willistoni*-L17 (5 copies), and *D. willistoni*-00 (6 copies); degenerate and *mar* MITE copies were identified also in these strains (Figure 3 and Figure 4C). In *D. tropicalis*, we found the most complete sequence (Dtro_ctg748), with 2760 bp but with small gaps, the largest with a 39 bp base at position 2170-2207 in the reconstructed *mar* (Table S7). In the genome of *D. insularis* we recovered only a few copies of *mar* relics (Figure 3 and Figure 4C) and no full-length or MITE. In *D. insularis*, 6 *mar* sequences (Dins_ctg2309_2, Dins_ctg2309_3, Dins_ctg2309_4, Dins_ctg2309_6, Dins_ctg2309_7, and Dins_ctg2309_8) were flanked by the *BEL-LTR* retrotransposon and the *Transib1* transposon (identified by Censor).

Mar-MITEs, similarly to canonical *mar* sequences, were retrieved in *D. willistoni*-Gd-H4-1, *D. willistoni*-L17, *D. willistoni*-00, *D. equinoxialis*, and *D. paulistorum*-L12 (Figure 4C). The most degenerate copies were found in *D. paulistorum*-L06 and *D. paulistorum*-L12; in these strains, even the largest sequences had many small deletions.

For the *mar* divergence analysis, we used the conserved *mar* region in genomes of the *willistoni* subgroup. The intragenomic divergence in the different *D. willistoni* strains varied by around 10.33% in *D. willistoni*-Gd-H4-1, 12.83% in *D. willistoni*-00, and 10.4% in *D. willistoni*-L17 (Table S8). We found an intragenomic divergence of 8.99% in *D. paulistorum*-L06 and 24.51% in *D. paulistorum*-L12. Interspecies divergence ranged from ~17% between the *D. equinoxialis* and *D. willistoni* strains to 43.29% between *D. insularis* and *D. equinoxialis* (Table S8).

We reconstructed the phylogenetic relationships between *mar* sequences using different methods (for more details see the Material and Methods section) (Figure 7A, 7B, and Figure S2). Figure 7A shows a phylogenetic tree constructed with partially complete sequences obtained from the *willistoni* group genomes, except the degenerate sequences Dtro_ctg838, Dtro_ctg804, and Dins_ctg1175. In Figure 7B, the phylogenetic relationships were generated employing the same sequences from Figure 7A and the representative copies of *mar* MITEs. Degenerate copies were manually selected according to the blocks of the alignments in the genomes of *D. willistoni* (3 strains), *D. paulistorum* (2 strains), and *D. equinoxialis*. In the two phylogenetic reconstructions (Figure 7A and 7B), the potentially complete sequences of *D. tropicalis* were positioned basally in the phylogeny, followed by the partially completed sequences of *D. willistoni*, and degenerate sequences of *D. equinoxialis* (Box I - Fig 7B). *Mar* MITEs and other degenerate sequences (relic sequences) formed a larger cluster composed of a small clade containing two other partially complete sequences of *D. willistoni* (Box II - Fig 7B), and a large clade including the other sequences (Box III - Fig 7B). In box III, sequences from one species usually appeared interspersed among the other species, possibly reflecting a low divergence between some, as well as low posterior probability values. For example, in *D. insularis* there was a clear clustering of the degenerate sequences in one of the well-supported branches; however, some of these sequences are related to MITEs from *D. equinoxialis*, although with low support value (0.57). When analyzing all the sequences recovered in the genomes, it was not possible to clearly identify the relationships established, mainly between MITE and degenerate sequences, probably because of the low sequence divergence (Figure S2).

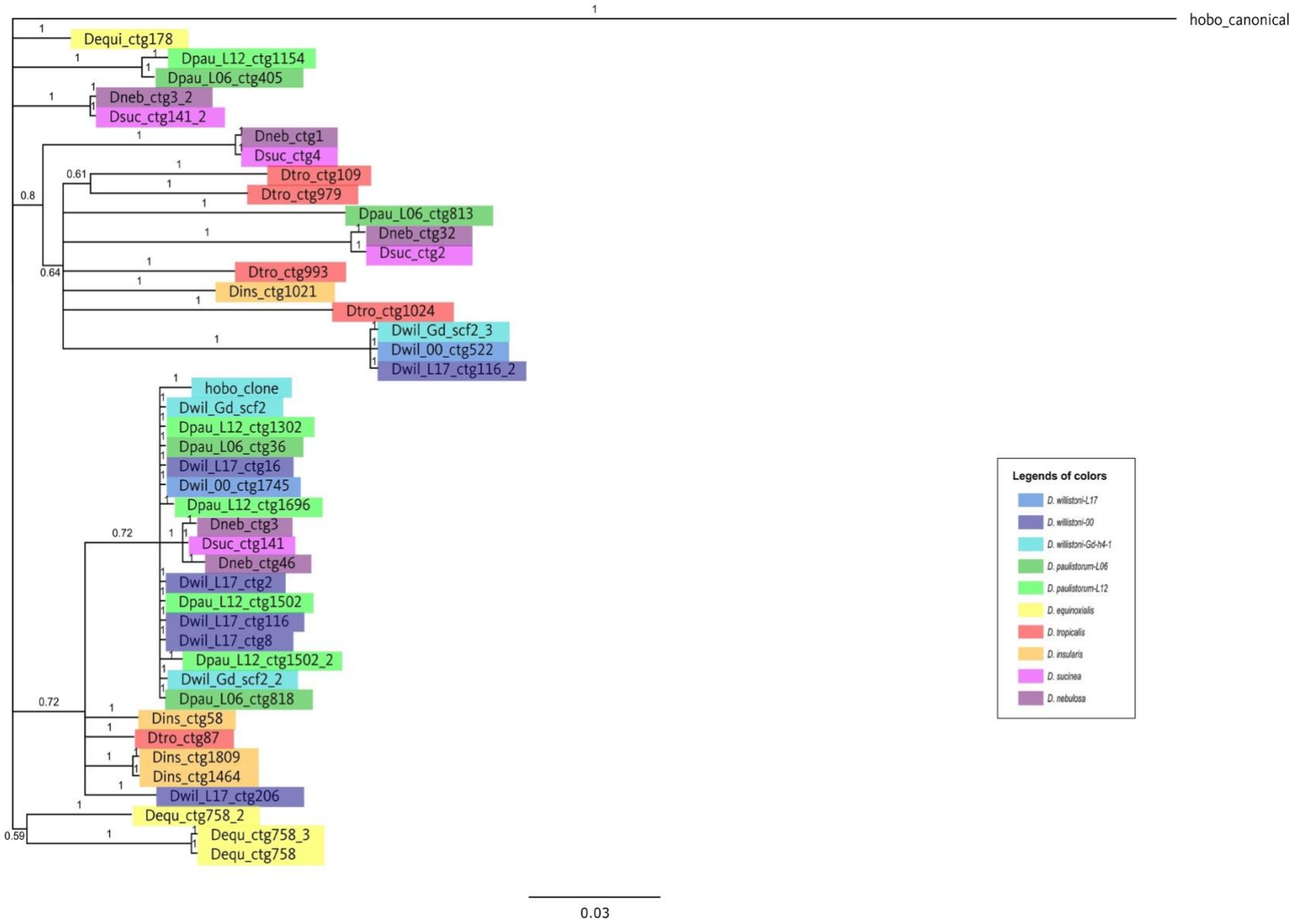


Figure 5 – Phylogenetic relationships of *hobo* copies in the *willistoni* group. Unrooted Bayesian tree (GTR+G) based on nucleotide sequences. Node supports are shown by posterior probability. *Drosophila melanogaster hobo* canonical sequence is shown in black; the *hobo*_clone was used in the FISH experiments. Different strains and species are indicated in different colors, as shown in the legend. Further information on *hobo* sequences

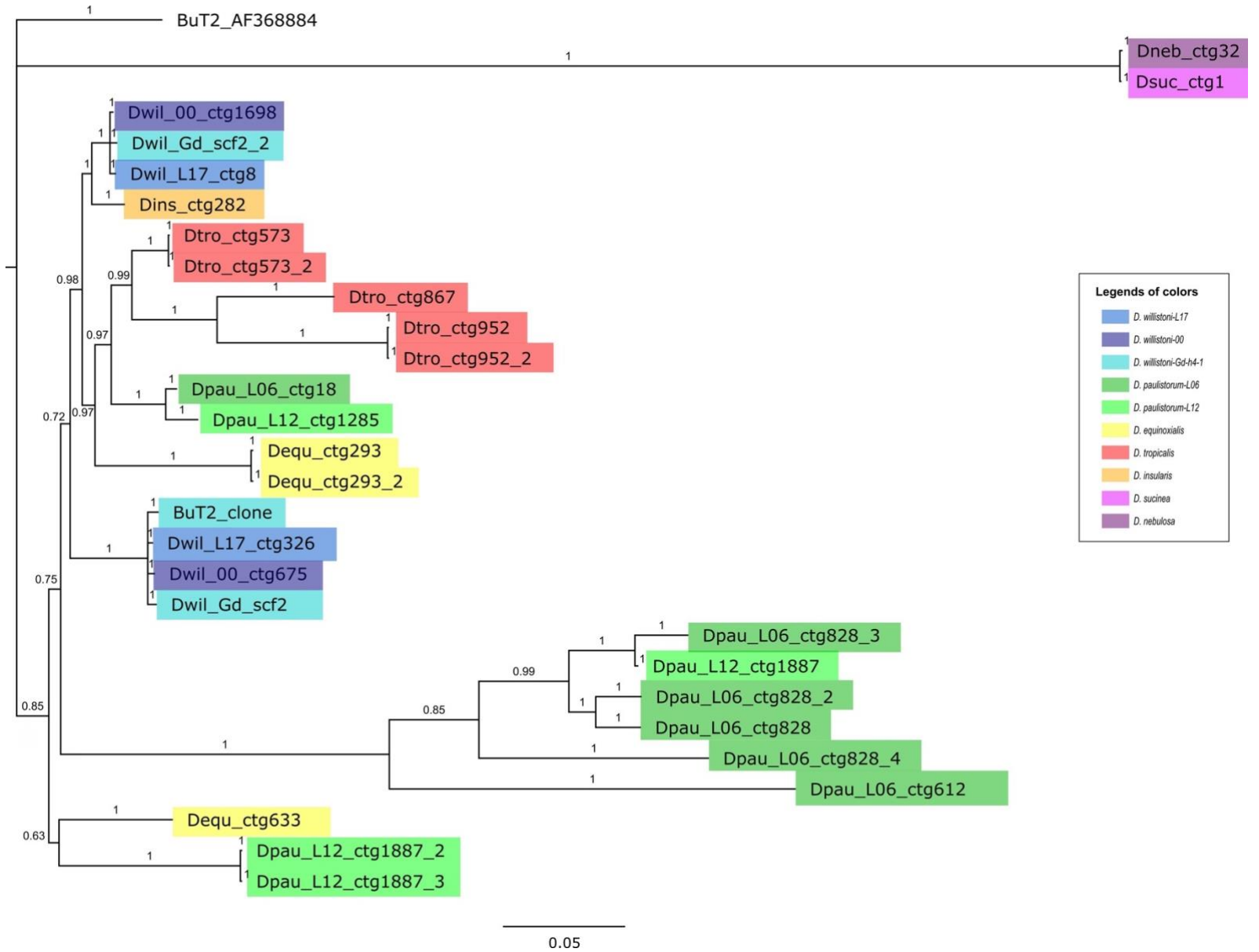


Figure 6 – Phylogenetic relationships of the *BuT2* copies in the *willistoni* group. Unrooted Bayesian tree (HKY+G) based on nucleotide sequences. Node supports are shown by posterior probability. *Drosophila buzzatti BuT2* canonical sequence is shown in black; the *BuT2*_clone was used in the FISH experiments. Different strains and species are indicated in different colors, as shown in the legend. Further information about *BuT2* sequences is available in Table S5.

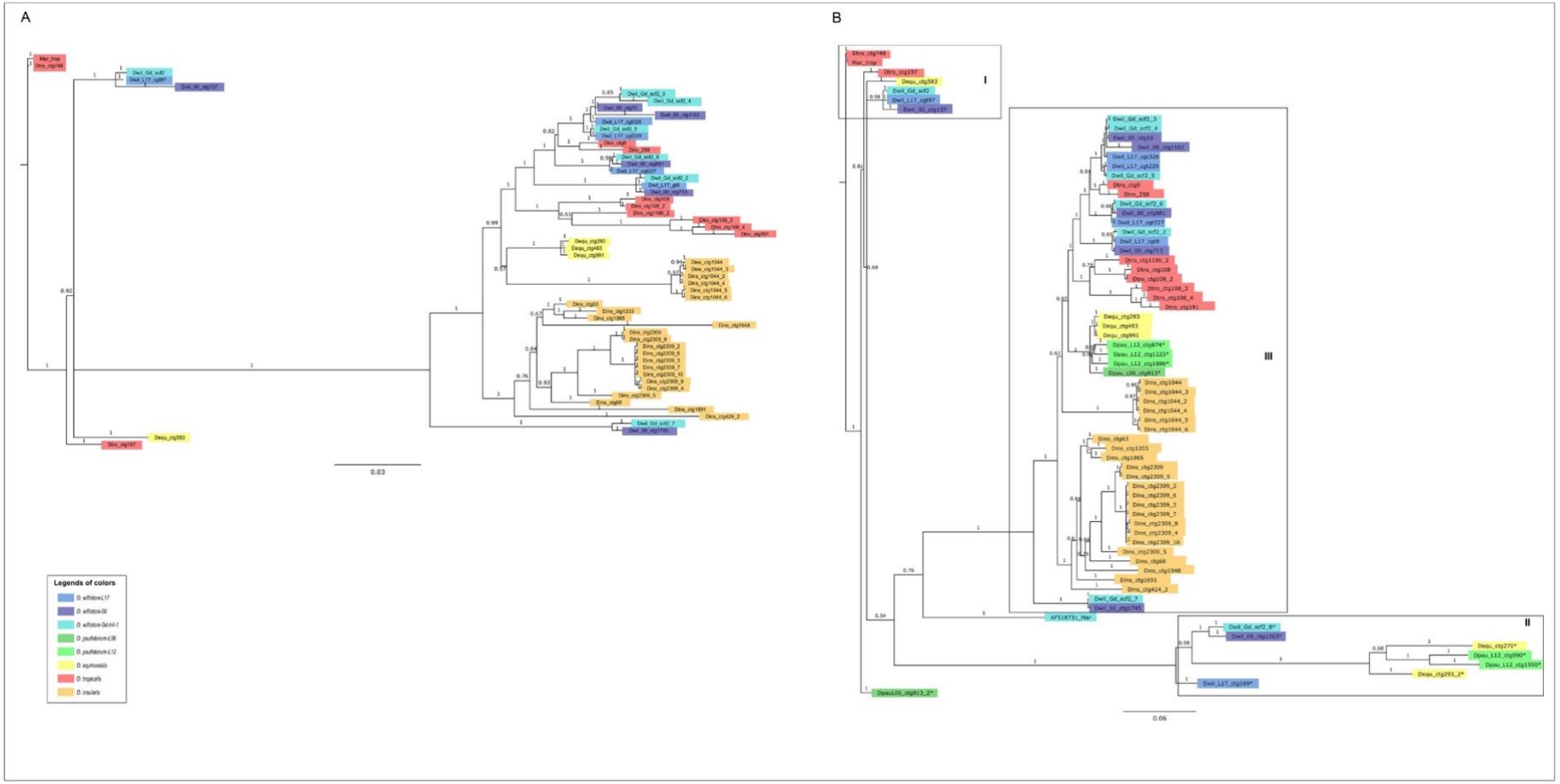


Figure 7 – Phylogenetic relationships of the *mar* copies in the *willistoni* subgroup. **A:** Bayesian tree of partially complete *mar* copies in the sequenced genomes of *D. willistoni* strains, *D. equinoxialis*, *D. tropicalis*, and *D. insularis*. Very degenerate copies of *D. tropicalis* and *D. insularis* were excluded from this analysis. **B**: Bayesian tree of *mar* partially complete, MITEs and relic copies in the sequenced genomes of the *willistoni* subgroup. This tree shows partially complete and relic copies used in A, and representative copies of *mar* MITES and degenerate copies in the *D. willistoni* strains, *D. paulistorum* strains, and *D. equinoxialis* genomes. AF518731_*mar* is the canonical *mar*-MITE of *D. willistoni*; *mar*_trop was used in the FISH experiments. Three different clades are indicated in boxes I, II, and III. Different strains and species are indicated in different colors, as shown in the legend. Degenerate sequences and MITEs are indicated by asterisks. Further information about *mar* sequences is available in Table S7.

Discussion

Drosophila willistoni was the first species in the *willistoni* group to be described, by Samuel Williston in 1896 (Dobzhansky and Powell, 1975). Dobzhansky (1950) described the karyotype of the species, and only in 2007 was the first genome sequenced, by the *Drosophila* 12 Genomes Consortium *et al*. (2007). Currently, there are more than 100 *Drosophila* genomes available (Kim *et al.*, 2021). This allows us to carry out more robust analyses to improve knowledge of the mechanisms involved in the evolution of species, transposons, and host genomes. The availability in our laboratory of strains of different geographical origins and which also have their genome sequenced, allows studies such as this one that are important to deepen the knowledge about the differences in the content and distribution of TE in the same species. Here, we conducted a detailed *in-silico* search to analyze *hobo, BuT2*, and *mar* transposons in available genomes of the *willistoni* group (Kim *et al.*, 2021). In addition, we analyzed the copy number and spatial distribution of these *hAT* transposons on polytene chromosomes of some *D. willistoni* strains. Further, *D. willistoni*-Gd-H4-1, *D. willistoni*-WIP-4, and *D. willistoni*-SG12.00 were used for *in-situ* analyses; *D. willistoni*-L17 and *D. willistoni*-00 were used for *in-silico* analysis; and *D. willistoni*-Gd-H4-1 was used for both *in-situ* and *in-silico* analyses. The available genomes were from the *D. willistoni* species subgroup, represented by *D. willistoni, D. paulistorum, D. equinoxialis, D. tropicais*, and *D. insularis*; and the *bocainensis* subgroup, represented by *D. nebulosa* and *D. sucinea*.

Our results showed that the same TEs (*hobo, BuT2*, and *mar*) varied widely in the copy number and structure of copies, even among the different *Drosophila* strains. Regarding the same TEs (Figure 3 and Figure 4A-C), the number of hybridization signals on the polytene chromosomes varied in the different populations: *D. willistoni*-Gd-H4-1, *D. willistoni*-WIP-4, and *D. willistoni*-SG12.00 (Figure 2A-I). Furthermore, in the strains *D. willistoni*-L17, *D. willistoni*-00, and *D. willistoni*-Gd-H4-1, we identified variations in the number and structure of copies of the same TEs (Figure 3 and Figure 4A-C). This suggests that different populations of *D. willistoni* have undergone changes in the TE content or different selective pressures on TE in that host genome. Differences between insertion sites of the same TE in *D. willistoni* strains have been previously observed. Regner *et al.* (1996) identified by *in-situ* hybridization, in *D. willistoni*-17A2 strain 10 insertion sites of the *P* element coinciding with the breakpoints of inversions, but in *D. willistoni*-WIP-11A observed only hybridization signals on heterochromatin (Regner *et al.*, 1996). Using Southern blot hybridization, Sassi *et al.* (2005) found differences in the number of TE copies of the *P* element, also among *D. willistoni* populations. In *D. mojavensis*, Palazzo *et al.* (2014) also found variability in the distribution and number of copies of the *Bari* element in different subspecies.

We observed different copy numbers for the elements of the *hAT* superfamily in the different *D. willistoni* strainsof different subspecies, both *in-situ* and *in-silico*. A similar situation was reported in *D. willistoni*-L17, from an unknown locality in Uruguay, which proved to have many more repetitive fractions, mainly retrotransposons, than *D. willistoni*-00 from Santa Maria de Ostuna, Nicaragua (Kim *et al.*, 2021). The differences observed between the *in-situ* analysis strains, particularly for the *mar* transposon (Figure 1 and Figure 2C, 2F and 2I), may be related to the chromosomal/genomic characteristics of the different populations of the species. *D. willistoni* can be subdivided into three subspecies: *D. w. willistoni*, *D. w. winge*, and *D. w. quechua* (Ayala and Tracey, 1973; Mardiros *et al.*, 2016), that have different geographic distributions. As shown in Figure 1, *D. willistoni* has a predominantly neotropical distribution, from Mexico and south Florida to the southernmost part of South America and from the Pacific to the Atlantic oceans (Spassky *et al.*, 1971; Zanini *et al.*, 2015). The strains used in the *in-situ* and *in-silico* analyses represent populations arranged along the geographic distribution of the different subspecies (Figure 1): *D. willistoni*-Gd-H4-1 (Guadeloupe Island – *willistoni* subspecies), *D. willistoni*-WIP-4 (Bahia, Brazil – *winge* subspecies), and *D. willistoni*-SG12.00 (Montevideo, Uruguay – *winge* subspecies), used in *in-situ* and *in-silico* analyses; and *D. willistoni-L17* (Uruguay – *winge* subspecies) *and D. willistoni-00* (Santa Maria de Ostuna, Nicaragua – *willistoni* subspecies) used only in *in-silico* analyses.

The differences in copy numbers of the elements of the *hAT* superfamily analyzed here may be related to the chromosomal and genomic plasticity required to allow *D. willistoni* to occupy different habitats within its geographic distribution. The chromosomal and genomic plasticity of *D. willistoni* has been demonstrated in the large number of rearrangements previously found in different populations (Dobzhansky, 1957; Valente and Araújo, 1985; Regner and Valente, 1993; Rohde *et al.*, 2006; Bhutkar *et al.*, 2008; Rohde and Valente, 2012). A characteristic common to all *D. willistoni* populations is paracentric inversions on the five chromosomal arms, although the location and amount of inversions vary among populations - Review in Rohde and Valente (2012). Rohde and Valente (2012) identified and cataloged 50 different rearrangements in 30 populations of polymorphic chromosomes of *D. willistoni* that segregate at different frequencies, with a clear latitudinal cline, from North to South America, along the species' distribution.

Additional evidence to support this hypothesis comes from the records of reproductive isolation between strains: populations found in Central America, North America, and northern Caribbean islands are reproductively isolatedfrom South American and southern Caribbean island strains (Figure 1) (Mardiros *et al.*, 2016). Partial reproductive isolation between populations influences gene exchange and consequently influences the differences of transposable elements in different populations.

For *D. willistoni*-Gd-H4-1 (the only one for which we have on our *Drosophila* Laboratory and whole sequenced genome) we obtained different estimates of copy numbers using different approaches (*in-situ* and *in-silico*). Our results showed that with the *hobo* element the different approaches were in accordance with the presence of low copy numbers (one by *in-situ* and three by *in-silico*). For the *BuT2* and *mar* elements, we observed discrepancies between the analyses (Figure 1 and 2). In *hobo*, we found stronger signals (identified by the ImageJ software) and some weaker ones could be seenin

the FISH picture (Figure 2A). Also, three *hobo* copies were detected in the sequenced genome, only one of which was complete (Figure 4A). However, in both *BuT2* and *mar*, the number of sequences differed between the two approaches; the largest difference was observed in *mar*, copy number estimatedby FISH was higher (by visual analysis) than the number retrieved in the sequenced genome (Figures 1, 2, and 4). The discrepancy between the number of copies found usingFISH and *in-silico* may be related to two factors: limitations of each approach and intrinsic characteristics of the *BuT2* and *mar* TEs that make it difficult to identify an absolute number of copies. In the case of *BuT2* and *mar*, elements are considered MITEs, and share structural characteristics such as small nonautonomous elements, present in high andvariable copy numbers, conservation of TIRs, and rich in ATregion (Bureau and Wessler, 1992; Jiang *et al.*, 2004; Feschotte*et al.*, 2005). Regarding the sequenced genomes, although large amounts of DNA data are available, many genomes are not fully known because of the difficulty in assembling therepetitive fraction, sequences obtained with NGS platformsare short and simply do not span long repetitive sequences,and numerous copies of reads can be nearly identical, leadingto the tendency to group them into single and collapsed contigs (Mascher and Stein, 2014; De Bustos *et al.*, 2016).This type of difference has been observed in other studies using different techniques; for instance, in *D. simulans*, with the *hAT hosimary* element, the number of copies estimatedby *in-silico* and Southern blot was higher than estimated byFISH (Deprá *et al.*, 2010). Maside *et al.* (2005) also reported differences between different techniques (PCR amplification and *in-situ* hybridization) of the *S-element* in *D. melanogaster*,noting that the amplification method can be more biased toward high-frequency elements than the *in-situ* method, which uses to identify the insertion sites.

We also investigated the presence and structure of copies of the *hobo*, *BuT2*, and *mar* elements in the sequenced genomes of the *willistoni* species group. In our analysis, the *hAT* transposase phylogenetic tree revealed three major clusters of related sequences (Figure S1), as previously referred to as the *Buster* family, *Tip* family, and *Ac* family by Rossato *et al.* (2014). The *D. willistoni-hobo* putative transposase fell within the *Ac* family, as did the other *hAT* from *Drosophila*, except for the elements *mar* (*Buster* family) and *BuT2* (*Tip* family) (Deprá *et al.*, 2012; Rossato *et al.*, 2014). The *hobo* element TSD consensus sequence (5`-nTnnnnAn-3`) also indicates that *D. willistoni-hobo* is an *Ac* element (Arensburger *et al.*, 2011; Rossato *et al.*, 2014). The cluster formed by *D. willistoni-hobo* is composed of elements previously described in fly species from different genera: *Drosophila willistoni* (*Howilli2*); *D. melanogaster* (canonical *hobo*), *D. ananassae* (*Hoana1*, *Hoana3*, and *Hoana8*), and *D. mojavensis* (Homo1), as well as *Ceratitis capitata* (*Cc-HRE*), *Bactrocera tryoni* (*Homer*), *Musca domestica* (*Hermes*), and *Lucilia cuprina* (Hermit) (Handler and Gomez, 1996; Ortiz and Loreto, 2009).

Hobo-like elements identified in the *willistoni* group genomes are closely related to the canonical *hobo* (*D. melanogaster)*, as conserved and identical TIRs in *D. willistoni* (three sequenced strains), *D. paulistorum* (two sequenced strains), *D. sucinea*, and *D. nebulosa* genomes (Figure 4A-C) (Calvi *et al.*, 1991). However, there was little divergence between the sequences of species in the *willistoni* group, including *D. sucinea* and *D. nebulosa* belonging to the *bocainensis* subgroup. Furthermore, as seen in the phylogenetic tree, the *hobo* copies do not cluster similarly to the phylogeny of the species in the *willistoni* group, so HTT events cannot be ruled out. Moreover, sequences close to *hobo*, called *hobo-brothers* elements, showed incongruities with the TE and host *Drosophila* species phylogenies, suggesting possible cases of horizontal transfer (Bernardo and Loreto, 2013). The presence of *hobo*-like sequences was previously identified only in some strains of *D. willistoni* collected in Brazil, including *D. willistoni*-WIP-4, but were absent in the Amazon strain and in other species of the *willistoni* group, by Southern and Dot blot hybridization (Loreto *et al.*, 1997). In the *melanogaster* subgroup, *hobo* elements were found in three forms: canonical (complete or deleted, lacking the central part of the sequence), relic (having TIRs and conserved subterminal sequences or defective in one TIR), and elements such as MITEs (review by Loreto *et al.*, 2018). We also identified sequences in canonical and relic form in *willistoni* group genomes, except in *D. equinoxialis*, *D. tropicalis*, and *D. insularis*, since in these genomes we found only degenerate copies (Figure 3).

BuT2 and *mar* were characterized as MITE sequences in *D. willistoni* genomes (Holyoake and Kidwell, 2003; Deprá *et al.*, 2012; Rossato *et al.*, 2014). The *BuT2* MITE elements identified in the *D. willistoni*-Gd-H4-1 genome have conserved TIRs but also the unusually low copy number (24 copies) that is common in MITE elements (Rossato *et al.*, 2014). Our *in-silico* searches were not able to recover *BuT2* MITE sequences in genomes of the *willistoni* species group (Figure 3 and Figure 4B). We also identified more *BuT2* hybridization signals in chromosomes of *D. willistoni*-Gd-H4-1 than in the sequenced genome of *D. willistoni*-Gd-H4-1 (Figure 1, Figure 2B and Figure 4B). The likely reason for the differences observed between the *in-silico* and *in-situ* approaches is that our searches retrieved only full-length and relic *BuT2* copies (Figure 3 and Figure 4B). We identified only *BuT2*-like degenerate sequences in the *bocainensis* subgroup, and one fragment each in *D. sucinea* and *D. nebulosa* (Figure 3 and Figure 4B). We found high rates of divergence between the sequences of species from the *willistoni* subgroup and the *bocainensis* subgroup, reaching 53.13% between *D. sucinea* and *D. paulistorum*-L06. This agrees with the phylogenetic tree, which showed the sequences of the *bocainensis* subgroup grouping separately from the *willistoni* subgroup (Table S6 and Figure 5). These sequences of the *bocainensis* subgroup are degenerate copies and have high divergence rates, which may be due either to a stochastic loss of element *BuT2* in the genomes of the *bocainensis* subgroup, or to retrieval of sequences homologous to other *BuT* elements such as *BuT1* in our searches (Cáceres *et al.*, 2001; Wallau *et al.*, 2012). Furthermore, the phylogenetic tree (Figure 5) has *BuT2* copies of the *willistoni* subgroup with a distribution similar to the evolution of the species in the group, and were probably vertically transmitted during the evolution of these species (Rossato *et al.*, 2014; Zanini *et al.*, 2018; Finet *et al.*, 2021). Our results agree with the findings by Rossato *et al.* (2014),

who hypothesized that *BuT2* was inserted in the ancestor of the neotropical *willistoni/saltans* groups and that MITEsexpanded in the *willistoni* group.

BuT2 showed more signals of hybridization in *D. willistoni*-Gd-H4-1 and *D. willistoni*-WIP4 (Figure 2B, E), whereas only two hybridization signals were identified in *D. willistoni*-SG12.00 (Figure 2H). Assuming that the many hybridization signals in the *D. willistoni*-Gd-H4-1 chromosome are of the *BuT2* MITE sequences described by Rossato *et al.* (2014), we hypothesized that the *BuT2* MITE sequences proliferated in *D. willistoni*-Gd-H4-1 and *D. willistoni*-WIP4 but not in *D. willistoni*-SG12.00. The presence of *BuT2* MITE sequences in the *willistoni* group is not completely clear, and further studies with several other strains are necessary. Interestingly, *BuT2* is associated with inversion breakpoints in *D. buzzatii* chromosomes (Cáceres *et al.*, 2001).

When the *mar* elements were characterized, the only genome of the *willistoni* group sequenced was *D. willistoni*-Gd-H4-1 (Drosophila 12 Genomes Consortium *et al.*, 2007; Deprá *et al.*, 2012). We found *mar* elements only in species of the *willistoni* subgroup (Figure 3 and Figure 4C), reinforcing the idea that this element invaded the genomes after the separation of the *willistoni* and *bocainensis* subgroups, as proposed by Deprá *et al.* (2012), and considering that the *D. willistoni* subgroup diverged approximately 7.3 Mya (review by Zanini *et al.*, 2018).

Mar elements were one of the first MITE families discovered in the *D. willistoni* genome (Holyoake and Kidwell, 2003). The origin of the different MITE families is unclear; one hypothesis is that MITEs originate from deletions of autonomous copies (Deprá *et al.*, 2012; Fattash *et al.*, 2013). Only in *D. tropicalis*, a low number of copies and one potentially complete copy (Dtrop_ctg748) were identified (Figure 3 and Figure 4). This sequence is likely ancestral, as apparent from the phylogenetic reconstruction (Figure 7A and 7B). However, in the genomes of *D. willistoni* (three sequenced strains), *D. paulistorum* (two sequenced strains) and *D. equinoxialis* (Figure 3 and Figure 4), we observed expansion of *mar* sequences, possibly originating from deletion of the TE transposase region (Figure 7A and 7B). MITES can be considered genomic superparasites because they conserve the transposase recognition regions for mobilization and are usually found in high copy numbers (Fattash *et al.*, 2013).

TEs in host genomes tend to survive by horizontal transmission to other hosts. When a TE inserts into a new host it tends to proliferation within a genome and within a population, accumulation of mutations, loss of element by inactivation, diversification within host, and element persistence within host (Schaack *et al.*, 2010). In *D. paulistorum*-L06 we found a large number of relic or degenerate copies (Figure 3 and Figure 4), but did not identify MITEs. One hypothesis is that this genome possibly did not undergo expansion of MITEs, but rather of complete copies that mutated over time. Additionally, the lower diversity of the *mar* sequences observed in *D. paulistorum*-L06 (8.99%) compared to the sequences found in *D. paulistorum*-L12 (24.51%) may be a function of the different geographical origins of the strains. *D. paulistorum*-L12 is Andean-Brazilian, from within the large geographic region of origin (Brazil, Ecuador, Peru, Colombia, and Venezuela) (Zanini *et al.*, 2018). *D. paulistorum*-L06 from San Salvador (El Salvador) has been maintained in the laboratory since 1955 (Kim *et al.*, 2021) (Figure 3), explaining the lower diversity of *mar* in this genome. The genome of *D. insularis* retained a low copy number, highly related but relic or degenerate (Figure 3, Figure 4, Figure 7A and 7B).

The dynamics of the TE and host genome coevolution are complex. In this study we showed the evolutionary history of the elements *hobo*, *BuT2*, and *mar* in the sequenced genomes of the *willistoni* group, as well as the distribution and estimated number of copies in the polytene chromosomes in three strains of *D. willistoni* from different geographic locations. We also compared different approaches (*in-situ* and *in-silico*) in examining the genome of *D. willistoni*-Gd-H4-1. The genome can be viewed as an ecosystem inhabited by diverse communities of TEs that seek to proliferate through interactions with each other TEs and with the genome as a whole and other component of the cell (Venner *et al.*, 2009; Bourque *et al.*, 2018). Evolutionary forces such as natural selection and genetic drift can also shape the distribution and accumulation of TEs in host genomes (Kidwell, 2002; Chénais *et al.*, 2012; Bourque *et al.*, 2018; Moschetti *et al.*, 2020). For example, mobilization in the host genome or colonization of new genomes is necessary to avoid loss by genetic drift, and potentially deleterious inserts will not remain in the population for many generations (Le Rouzic and Capy, 2006; Venner *et al.*, 2009; Bourque *et al.*, 2018). Through a genomic and cytogenetic approach, we reported that different populations (strains) of one species, *D. willistoni*, maintain and share the same transposon differently. Our data also showed that the genetic plasticity enabled by transposable elements can help species such as *D. willistoni* to occupy very different environments over its wide geographic distribution.

Acknowledgements

We thank Dr. Thales R. O. de Freitas for the use of his laboratory for the FISH experiments. We also wish to thank the reviewers for their important comments. This study was supported by research grants and fellowships from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and FAPERGS (Fundação de Amparo à pesquisa do Estado do Rio Grande do Sul).

Conflict of Interest

The authors have no conflicts of interest to declare.

Author Contributions

NAB, TDO, MD, BG and VLSV conceived and designed the study; NAB and TDO performed the experiments; NAB, TDO, MD, VLSV analyzed experiments and wrote the manuscript. All authors read and approved the final version.

References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3-W10.

Arensburger P, Hice RH, Zhou L, Smith RC, Tom AC, Wright JA, Knapp J, O'Brochta DA, Craig NL and Atkinson PW (2011) Phylogenetic and functional characterization of the *hAT* transposon superfamily. Genetics 188:45-57.

Ayala FJ and Tracey ML (1973) Enzyme variability in the *Drosophila willistoni* group. J Hered 64:120-124.

Bernardo LP and Loreto ELS (2013) *hobo-brothers* elements and their time and place for horizontal transfer. Genetica 141:471-478.

Bertocchi NA, de Oliveira TD, del Valle Garnero A, Coan RLB, Gunski RJ, Martins C and Torres FP (2018) Distribution of *CR1-like* transposable element in woodpeckers (Aves Piciformes): Z sex chromosomes can act as a refuge for transposable elements. Chromosome Res 26:333-343.

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.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS *et al.* (2018) Ten things you should know about transposable elements. Genome Biol 19:199.

Bureau TE and Wessler SR (1992) Tourist: A large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283-1294.

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.

Calvi BR, Hong TJ, Findley SD and Gelbart WM (1991) Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *activator*, and *Tam3*. Cell 66:465-471.

Casals F, González J and Ruiz A (2006) Abundance and chromosomal distribution of six *Drosophila buzzatii* transposons: *BuT1, BuT2, BuT3, BuT4, BuT5,* and *BuT6*. Chromosoma 115:403-412.

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH and Schlötterer C (2004) World-wide survey of an Accord insertion and its association with DDT resistance in *Drosophila melanogaster*. Mol Ecol 13:2491-2504.

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.

De Bustos A, Cuadrado A and Jouve N (2016) Sequencing of long stretches of repetitive DNA. Sci Rep 6:36665.

Deprá M, Ludwig A, Valente VL and Loreto EL (2012) *Mar*, a MITE family of *hAT* transposons in *Drosophila*. Mob DNA 3:13.

Deprá M, Panzera Y, Ludwig A, Valente VLS and Loreto ELS (2010) *Hosimary*: A new *hAT* transposon group involved in horizontal transfer. Mol Genet Genomics 283:451-459.

Deprá M, Valente VLS, Margis R and Loreto ELS (2009) The *hobo* transposon and *hobo* -related elements are expressed as developmental genes in *Drosophila*. Gene 448:57-63.

Dobzhansky T (1950) The chromosomes of *Drosophila willistoni*. J Hered 41:156-158.

Dobzhansky T (1957) Genetics of natural populations. XXVI. Chromosomal variability in island and continental populations of *Drosophila willistoni* from Central America and the West Indies. Evolution 11:280-293.

Dobzhansky T and Powell JR (1975) The *willistoni* group of sibling species of *Drosophila.* In: King RC (ed) Handbook of Genetics. Plenum Press, New York, pp 589-622.

Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, William G *et al.* (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203-218.

Fattash I, Rooke R, Wong A, Hui C, Luu T, Bhardwaj P, Yang G and Bainard J (2013) Miniature inverted-repeat transposable elements: Discovery, distribution, and activity. Genome 56:475-486.

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

Feschotte C, Osterlund MT, Peeler R and Wessler SR (2005) DNA-binding specificity of rice *mariner-like* transposases and interactions with Stowaway MITEs. Nucleic Acids Res 33:2153-2165.

Finet C, Kassner VA, Carvalho AB, Chung H, Day JP, Day S, Delaney EK, De Ré FC, Dufour HD, Dupim E *et al.* (2021) DrosoPhyla: Resources for drosophilid phylogeny and systematics. Genome Biol Evol 13:evab179.

Garcia C, Delprat A, Ruiz A and Valente VLS (2015) Reassignment of *Drosophila willistoni* genome scaffolds to chromosome II arms. G3 (Bethesda) 5:2559-2566.

González J and Petrov D (2009) MITEs - the ultimate parasites. Science 325:1352-1353.

Handler AM and Gomez SP (1996) The *hobo* transposable element excises and has related elements in tephritid species. Genetics 143:1339-1347.

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

Huang Y, Niu B, Gao Y, Fu L and Li W (2010) CD-HIT Suite: A web server for clustering and comparing biological sequences. Bioinformatics 26:680-682.

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

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772-780.

Kidwell MG (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica 115:49-63.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'agostino ERR, Pelaez J *et al.* (2021) Highly contiguous assemblies of 101 drosophilid genomes. Elife 10:e66405.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547-1549.

Larsson A (2014) AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30:3276-3278.

Le SQ and Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320.

Le Rouzic A and Capy P (2006) Population genetics models of competition between transposable element subfamilies. Genetics 174:785-793.

Loreto ELS, Basso da Silva L, Zaha A and Valente VLS (1997) Distribution of transposable elements in neotropical species of *Drosophila.* Genetica 101:153-165.

Loreto ELS, Deprá M, Diesel JF, Panzera Y and Valente-Gaiesky VLS (2018) *Drosophila* relics *hobo* and *hobo*-MITEs transposons as raw material for new regulatory networks. Genet Mol Biol 41:198-205.

Mardiros XB, Park R, Clifton B, Grewal G, Khizar AK, Markow TA, Ranz JM and Civetta A (2016) Postmating reproductive isolation between strains of *Drosophila willistoni*. Fly (Austin) 10:162-171.

Marques EK, Napp M, Winge H and Cordeiro AR (1966) A cornmeal, soybean flour, wheat germ medium for *Drosophila*. Drosoph Inf Serv 41:187.

Mascher M and Stein N (2014) Genetic anchoring of whole-genome shotgun assemblies. Front Genet 5:208.

Maside X, Assimacopoulos S and Charlesworth B (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. Genet Res 85:195-203.

Miller MA, Pfeiffer W and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew Comput Environ Work (GCE). DOI: 10.1109/GCE.2010.5676129

Moschetti R, Palazzo A, Lorusso P, Viggiano L and Marsano RM (2020) "What you need, baby, I got it": Transposable elements as suppliers of cis-operating sequences in *Drosophila*. Biology (Basel) 9:25.

Nylander JAA (2004) MrModeltest Version 2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala.

Ortiz M de F and Loreto ELS (2008) The *hobo-related* elements in the *melanogaster* species group. Genet Res (Camb) 90:243-252.

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

Palazzo A, Moschetti R, Caizzi R and Marsano RM (2014) The *Drosophila mojavensis Bari3* transposon: Distribution and functional characterization. Mob DNA 5:21.

Regner LP and Valente VLS (1993) *Drosophila willistoni* mating activity - urbanization effects and a search for its chromosomal basis. Evol Biol 7:327-349.

Regner LP, Pereira MSO, Alonso CEV, Abdelhay E and Valente VLS (1996) Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. J Hered 87:191-198.

Rohde C and Valente VLS (2012) Three decades of studies on chromosomal polymorphism of *Drosophila willistoni* and description of fifty different rearrangements. Genet Mol Biol 35:966-979.

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.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA and Huelsenbeck JP (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539-542.

Rossato DO, Ludwig A, Deprá M, Loreto ELS, Ruiz A and Valente VLS (2014) *BuT2* is a member of the third major group of *hAT* transposons and is involved in horizontal transfer events in the genus *Drosophila*. Genome Biol Evol 6:352-365.

Santos-Colares MC, Valente VL da S and Goñi B (2003) The meiotic chromosomes of male *Drosophila willistoni*. Caryologia 56:431-437.

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

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 25:537-546.

Schneider CA, Rasband WS and Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671-675.

Spassky B, Richmond RC, Perez-Salas S, Pavlovsky O, Mourão CA, Hunter AS, Hoenigsberg 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.

Valente VLS and Araújo A (1985) Observations on the chromosomal polymorphism of natural populations of *Drosophila willistoni* and its association with the choice of feeding and breeding sites. Rev Bras Genet 8:271-284.

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

Valente VLS, Ruszczyk A and Santos RA (1993) Chromosomal polymorphism in urban *Drosophila willistoni*. Rev Bras Genet 16:307-317.

Venner S, Feschotte C and Biémont C (2009) Dynamics of transposable elements: Towards a community ecology of the genome. Trends Genet 25:317-323.

Wallau GL, Ortiz MF and Loreto ELS (2012) Horizontal transposon transfer in Eukarya: Detection, bias, and perspectives. Genome Biol Evol 4:689-699.

Wells JN and Feschotte C (2020) A field guide to eukaryotic transposable elements. Annu Rev Genet 54:539-561.

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

Zanini R, Deprá M and Valente VLS (2015) Can sibling species of the *Drosophila willistoni* subgroup be recognized through combined microscopy techniques? Rev Bras Entomol 59:323-331.

Zanini R, Müller MJ, Vieira GC, Valiati VH, Deprá M and Valente VL da S (2018) Combining morphology and molecular data to improve *Drosophila paulistorum* (Diptera, Drosophilidae) taxonomic status. Fly (Austin) 12:81-94.

Zhang HH, Shen YH, Xu HE, Liang HY, Han MJ and Zhang Z (2013) A novel *hAT* element in *Bombyx mori* and *Rhodnius prolixus*: Its relationship with miniature inverted repeat transposable elements (MITEs) and horizontal transfer. Insect Mol Biol 22:584-596.

Supplementary material

The following online material is available for this article:

Table S1 – Information for the *Drosophila willistoni* strains used in the FISH experiments.

Table S2 – Information on genomes of the *willistoni* group used in this study.

Table S3 – *Hobo* sequences identified in the *willistoni* group genomes.

Table S4 – Nucleotide divergence percentages of *hobo* sequences.

Table S5 – *BuT2* sequences identified in the *willistoni* group genomes.

Table S6 – Nucleotide divergence percentages of *BuT2* sequences.

Table S7 – *Mar* sequences identified in the *willistoni* group genomes.

Table S8 – Nucleotide divergence percentages of *mar* sequences.

Figure S1 – Phylogenetic relationships of the *hAT* superfamily.

Figure S2 – Neighbor-Joining tree of *mar* sequences.

Associate editor: Louis Bernard Klaczko

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S1 - Information for the *Drosophila willistoni* strains used in the FISH experiments.

Strains of *D. willistoni*	Geographic location	Collection date	Geographic coordinates	Collected by
Gd-H4-1	Guadeloupe Islands	1991	16°15' N- 61°35' W	J. Powell
WIP-4	Ibirapitanga, Bahia, Brazil	1960	12°54' S-38°19' W	Helga Winge and A.R. Cordeiro
SG12.00	Montevideo, Uruguay	2000	34°53' S-56°16' W	Beatriz Goñi

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S2 - Species, strains and assemblies of genomes of the *willistoni* group used in this work.

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Strain name[a]
Drosophila	*Sophophora*	*willistoni*	*willistoni*	*D. willistoni*-L17	L-G3
				D. willistoni-00	NA
				D. willistoni-Gd-H4-1	GD-h4-1
				D. paulistorum-L06	(Heed) H66.1C
				D. paulistorum-L12	L12
				D. equinoxialis	NA
				D. tropicalis	(Heed) H65.2
				D. insularis	jp01i
			bocainensis	*D. sucinea*	49.15
				D. nebulosa	H176.10

[a] Strain names along with corresponding NDSSC stock center numbers were provided by Kim *et. al.* 2021.

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S3 - *Hobo* sequences identified in the *willistoni* group genomes with the nomenclature used in this work, length and the location in the contig or scaffolds.

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name
Drosophila	*Sophophora*	*willistoni*	*willistoni*	*D. willistoni*-L17	contig_2:c11570663-11570224	Dwil_L17_ctg2
					contig_116:374204-374643	Dwil_L17_ctg116
					contig_116:1194557-1194841	Dwil_L17_ctg116_2
					contig_8:585969-586408	Dwil_L17_ctg8
					contig_206:264514-264805	Dwil_L17_ctg206
					contig_16:3093225-3099665	Dwil_00_ctg16
				D. willistoni-00	contig_1745:23005-22565	Dwil_00_ctg1745

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

					contig_522:1140-1426	Dwil_00_ctg522
				D. willistoni-Gd-H4-1	scf2_1100000004887	Dwil_Gd_scf2
					scf2_1100000011155	Dwil_Gd_scf2_2
					scf2_1100000004854	Dwil_Gd_scf2_3
				D. paulistorum-L06	contig_36:c1463801-1463361	Dpau_L06_ctg36
					contig_818:c1823481-1823044	Dpau_L06_ctg818
					contig_405:c154382-153957	Dpau_L06_ctg405
					contig_813:c3203276-3202886	Dpau_L06_ctg813
				D. paulistorum-L12	contig_1302:c304983-304542	Dpau_L12_ctg1302
					contig_1696:298949-299389	Dpau_L12_ctg1696
					contig_1502:c92077-91638	Dpau_L12_ctg1502
					contig_1502:83646-84086	Dpau_L12_ctg1502_2

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

		contig_1154:213411-213836	Dpau_L12_ctg1154
	D. equinoxialis	contig_758:c2990-2844	Dequ_ctg758
		contig_758:246449-246884	Dequ_ctg758_2
		contig_758:c4177-3882	Dequ_ctg758_3
		contig_178:c78153-71907	Dequ_ctg178
	D. tropicalis	contig_87:50052-50492	Dtro_ctg87
		contig_109:271215-271655	Dtro_ctg109
		contig_979:1790162-1790453	Dtro_ctg979
		contig_993:c123966-123575	Dtro_ctg993
		contig_1024:c96591-96280	Dtro_ctg1024
	D. insularis	contig_1809:22014-22454	Dins_ctg1809
		contig_1464:c7882-7443	Dins_ctg1464

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

		contig_58:c843557-843113	Dins_ctg58
		contig_1021:c8391183813	Dins_ctg1021
	bocainensis	*D. sucinea*	contig_141:c7473762-7473322	Dsuc_ctg141
			contig_141:c7913447-7913015	Dsuc_ctg141_2
			contig_4:c4456473-4456041	Dsuc_ctg4
			contig_2:1115483-1115920	Dsuc_ctg2
		D. nebulosa	contig_3:1575774-1582214	Dneb_ctg3
			contig_46:14720546-14720983	Dneb_ctg46
			contig_3:1102167-1102610	Dneb_ctg3_2
			contig_1:c3959237-3958802	Dneb_ctg1
			contig_32:276148462-7615283	Dneb_ctg32

[a] TIRs present only in the 5`-*hobo*
(-) absence

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S4 - Nucleotide divergence percentages of *hobo* sequences found within and between species/strains.

Species	*D. willistoni*-Gd-H4-1	*D. willistoni*-L17	*D. willistoni*-00	*D. paulistorum*-L06	*D. paulistorum*-L12	*D. equinoxialis*	*D. tropicalis*	*D. insularis*	*D. sucinea*	*D. nebulosa*
D. willistoni-Gd-H4-1	13.88%									
D. willistoni-L17	8.66%	3.91%								
D. willistoni-00	10.51%	7.00%	13.64%							
D. paulistorum-L06	10.19%	5.78%	8.97%	7.98%						
D. paulistorum-L12	8.74%	3.74%	8.04%	5.45%	4.10%					
D. equinoxialis	11.40%	6.84%	7.88%	7.33%	6.43%	4.48%				
D. tropicalis	13.05%	8.35%	9.84%	9.16%	8.33%	8.76%	9.81%			
D. insularis	10.32%	4.65%	7.25%	6.09%	4.71%	6.80%	8.13%	4.39%		
D. sucinea	12.48%	8.31%	10.87%	9.29%	8.67%	7.59%	9.26%	6.79%	11.41%	
D. nebulosa	11.70%	7.17%	10.21%	8.51%	7.43%	7.45%	9.09%	6.33%	8.40%	9.84%

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S5: *BuT2* sequences identified in the *willistoni* group genomes with the nomenclature used in this work, length and the location in the contig or scaffolds.

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name	TIR	Length (bp)	TSD
Drosophila	*Sophophora*	*willistoni*	*willistoni*	*D. willistoni*-L17	contig_8:c956548-950094	Dwil_L17_ctg8	cagtgctgccaa	2695	ctcccata
					contig_326:10479000-10482557	Dwil_L17_ctg326	-	837	-
				D. willistoni-00	contig_675:c193887-187330	Dwil_00_ctg675	cagtgctgccaa [a]	1032	gtggttag
					contig_1698:54070-63524	Dwil_00_ctg1698	cagtgctgccaa	2737	ctcccata
				D. willistoni-Gd-H4-1	scf2_1100000004958	Dwil_Gd_scf2_2	cagtgctgccaa	2742	c(t/a)cccata
					scf2_1100000004967	Dwil_Gd_scf2	cagtgctgccaa [a]	1032	gtggttag
				D. paulistorum-L06	contig_18:c659779-653712	Dpau_L06_ctg18	cagtgctgccaa	1644	(a/g)gcaaa(a/g)g
					contig_828:33560431-33566507	Dpau_L06_ctg828	cagtgctgccaa [a]	1004	gctaactg
					contig_828:c33557118-33551042	Dpau_L06_ctg828_2	cagtgctgccaa [a]	1044	gcagcagc
					contig_828:35734319-35740395	Dpau_L06_ctg828_3	cagtgctgccaa [a]	1025	gctacctg
					contig_828:c33552056-33545980	Dpau_L06_ctg828_4	cagtgctgccaa [a]	510	gctacctg
					contig_612:347628-353668	Dpau_L06_ctg612	-	522	-
				D. paulistorum-L12	contig_1887:c16021910-16015834	Dpau_L12_ctg1887	cagtgctgccaa [a]	1042	gctacctg
					contig_1887:3773899-3780096	Dpau_L12_ctg1887_2	-	1158	-
					contig_1887:3774096-3780315	Dpau_L12_ctg1887_3	-	1158	-
					contig_1285:664505-670572	Dpau_L12_ctg1285	cagtgctgccaa	1358	(a/g)gcaaa(a/g)(a/g)

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name	TIR	Length (bp)	TSD
				D. equinoxialis	contig_633:2034621-2040974	Dequ_ctg633	cagtgctgccaa [a]	1110	ggtgggga
					contig_293:3234777-3241187	Dequ_ctg293	-	779	-
					contig_293:32345663240643	Dequ_ctg293_2	-	779	-
				D. tropicalis	contig_867:c2136044-2129724	Dtro_ctg867	cagtgctgccaa [a]	998	-
					contig_952:c423594-417411	Dtro_ctg952	cagtgctgccaa [a]	1215	ctaacggc
					contig_952:c424174-418097	Dtro_ctg952_2	cagtgctgccaa [a]	1215	ctaacggc
					contig_573:c1710793-1704667	Dtro_ctg573	cagtgctgccaa	2184	ggcggagc
					contig_573:c1710825-1704789	Dtro_ctg573_2	cagtgctgccaa	2184	ggcggagc
				D. insularis	contig_282:21964482-202902	Dins_ctg282	cagtgctgccaa	2546	-
			bocainensis	*D. sucinea*	contig_1:26154067-26163191	Dsuc_ctg1	-	764	-
				D. nebulosa	contig_32:25989124-25995248	Dneb_ctg32	-	764	-

[a] TIRs present only in the 3`-*BuT2*

(-) absence

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S6 - Nucleotide divergence percentages of *BuT2* sequences found within and between species/strains.

Species	*D. willistoni*-Gd-H4-1	*D. willistoni*-L17	*D. willistoni*-00	*D. paulistorum*-L06	*D. paulistorum*-L12	*D. equinoxialis*	*D. tropicalis*	*D. insularis*	*D. sucinea*	*D. nebulosa*
D. willistoni-Gd-H4-1	9.11%									
D. willistoni-L17	5.03%	10.29%								
D. willistoni-00	4.74%	5.01%	9.11%							
D. paulistorum-L06	22.29%	23.77%	22.11%	13.91%						
D. paulistorum-L12	14.15%	14.36%	13.98%	17.90%	17.99%					
D. equinoxialis	10.00%	9.48%	10.01%	25.68%	16.38%	7.20%				
D. tropicalis	13.20%	15.02%	13.11%	17.20%	15.63%	15.40%	5.30%			
D. insularis	5.88%	6.59%	5.77%	17.98%	12.61%	9.81%	10.04%			
D. sucinea	24.50%	24.63%	24.83%	53.13%	35.92%	33.75%	37.34%	25.99%		
D. nebulosa	24.70%	24.83%	25.02%	53.02%	35.98%	33.97%	37.46%	26.12%	0.13%	

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S7 - *Mar* sequences identified in the *willistoni* group genomes with the nomenclature used in this work, length and the location in the contig or scaffolds.

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name	TIR	Length (bp)	TSD
Drosophila	*Sophophora*	*willistoni*	*willistoni*	*D. willistoni*-L17	contig_229:c66505-74287	Dwil_L17_cgi229	cagggcccggc	2341	g(cta(g/t)tt
					contig_97:c2133965-2140917	Dwil_L17_cgi97	cagggggaggc	2260	ctctac(t/c)c
					contig_326:c7390880-7398393	Dwil_L17_cgi326	cagggcccggca [a]	2030	ct(c/g)t(a/c)(c/a)(a/t)a
					contig_8:c742956-735649	Dwil_L17_cgi8	-	1724	-
					contig_227:c32390491-32398000	Dwil_L17_cgi227	cagggcccggc [a]	1998	gattagaac
					contig_169: c265415-266051	Dwil_L17_ctg169	-	562	-
				D. willistoni-00	contig_713:c8475-1168	Dwil_00_ctg713	-	1754	-
					contig_137:c334978-328626	Dwil_00_cgi137	gtcccggcaatt [a]	1382	-
					contig_1745:c162134-155401	Dwil_00_ctg1745	gcttaacttgt [a]	1644	-
					contig_881:c1057949-1065188	Dwil_00_ctg881	-	1797	-
					contig_1102:c1-1993	Dwil_00_ctg1102	cagggaccgac [a]	1069	ctctacaa
					contig_10:c138987-131462	Dwil_00_ctg10	cagggaccggc [a]	1998	ctctacaa/actggagt
					contig_1263:c81211-81822	Dwil_00_ctg1263	-	519	-
				D. willistoni-Gd-H4-1	scf2_1100000004958:c721448-728037	Dwil_Gd_scf2	cagggggaggc	2268	ctctac(t/c)c
					scf2_1100000004958:c248606-2478755	Dwil_Gd_scf2_2	-	1722	ttcgatga [b]
					scf2_1100000012712:c1372-1	Dwil_Gd_scf2_3	-	1319	-
					scf2_1100000004967:c4275204-4267679	Dwil_Gd_scf2_4	tcccaagcggc [a]	2006	tta aga gt/tta ttc ga
					scf2_1100000011190:c1676-1	Dwil_Gd_cf2_5	-	1495	-
					scf2_1100000004884:c190698-183177	Dwil_Gd_scf2_6	-	1927	-
					scf2_1100000004542:c67336-60603	Dwil_Gd_scf2_7	tcaggacgaat [a]	1701	attctatg [b]
					scf2_1100000004830:c439968-439226	Dwil_Gd_scf2_8	-	646	-
				D. paulistorum-L06	contig_813:c 2236897-2235283	Dpau_L06_ctg813	-	1512	-
					contig_813:c 2180492-2179440	Dpau_L06_ctg813_2	-	897	-
				D. paulistorum-L12	contig_674:c188482-186568	Dpau_L12_ctg674	-	1517	-
					contig_1323:c12301-10250	Dpau_L12_ctg1323	-	1777	-
					contig_1886:c769625-767711	Dpau_L12_ctg1886	-	1516	-

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name	TIR	Length (bp)	TSD
					contig_990:c45776-44933	Dpau_L12_ctg990	-	497	-
					contig_1590:c386924-386165	Dpau_L12_ctg1590	-	458	-
				D. equinoxialis	contig_583:c1253850-1246431	Dequ_ctg583	cagggggaggc [a]	1530	-
					contig_293:c869557-861796	Dequ_ctg293	cagggcccggc [a]	1639	-
					contig_991:c95981-89067	Dequ_ctg991	cagggcccggc [a]	1075	-
					contig_483c:259082-266841	Dequ_ctg483	cagggcccggc [a]	1640	-
					contig_270c:154589-155437	Dequ_ctg270	-	655	-
					contig_293:c1072823-1073495	Dequ_ctg293_2	-	523	-
				D. tropicalis	contig_748:c123411-114941	Dtro_ctg748	cagaggagac	2760	not conserved
					contig_197:16059-24004	Dtro_ctg197	cagggggaggc [a]	2068	-
					contig_9:586622-594136	Dtro_ctg9	cagggcccggc [a]	1966	-
					contig_258:749883-757356	Dtro_ctg258	-	1934	-
					contig_108:c303354-295980	Dtro_ctg108	cagggcccg [a]	1867	-
					contig_108:48387-55723	Dtro_ctg108_2	cagggcccg [a]	1849	-
					contig_108:c467912-464549	Dtro_ctg108_3	cagggaccggc [a]	473	-
					contig_108:191753-198105	Dtro_ctg108_4	cagggaccggc [a]	476	-
					contig_838:c27084-20915	Dtro_ctg838	-	984	-
					contig_1190:30021-30178	Dtro_ctg1190	cagggggaggc [a]	2033	-
					contig_804:1-3233	Dtro_ctg804	-	333	-
					contig_191:c27800-23642	Dtro_ctg191	cagggaccggc [a]	474	-
				D. insularis	contig_1865:1562-6147	Dins_ctg1865	cagggcccggc [a]	1657	-
					contig_1203:17415-24357	Dins_ctg1203	cagggcccggc [a]	1110	-
					contig_2309:37451-44372	Dins_ctg2309	cagggcccggc [a]	2019	-
					contig_2309:3201-9992	Dins_ctg2309_2	cagggcccggc [a]	1175	-
					contig_2309:7064-13855	Dins_ctg2309_3	cagggcccggc [a]	1175	-
					contig_2309:10926-17717	Dins_ctg2309_4	cagggcccggc [a]	1175	-
					contig_2309:20253-26866	Dins_ctg2309_5	cagggcccgcc	769	-
					contig_2309:4000-10443	Dins_ctg2309_6	cagggcccggc [a]	1175	-
					contig_2309:7863-14306	Dins_ctg2309_7	cagggcccggc [a]	1175	-
					contig_2309:11725-18168	Dins_ctg2309_8	cagggcccggc [a]	1175	-
					contig_2309:38380-44823	Dins_ctg2309_9	cagggcccggc [a]	2019	-

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Genus	*Subgenus*	*Group*	*Subgroup*	*Species*	Scaffold/contig position	Name	TIR	Length (bp)	TSD
					contig_2309:136-7443	Dins_ctg2309_10	-	1003	-
					contig_1691:54802-61944	Dins_ctg1691	-	1171	-
					contig_424:c512443-505582	Dins_ctg424	-	1409	-
					contig_1044:75760-82098	Dins_ctg1044	-	1509	-
					contig_1044:88090-93428	Dins_ctg1044_2	-	1508	-
					contig_1044:76350-82617	Dins_ctg1044_3	-	1509	-
					contig_1044:88680-93947	Dins_ctg1044_4	-	1271	-
					contig_1044:c148079-141744	Dins_ctg1044_5	-	1508	-
					contig_1044:c146789-140520	Dins_ctg1044_6	-	1486	-
					contig_1175:27855-33891	Dins_ctg1175	-	728	-
					contig_63:c394755-393231	Dins_ctg63	cagggcccggc [a]	2099	-
					contig_1948:c3029-1043	Dins_ctg1948	cagggcccggc [a]	813	-
					contig_68:c8004-787	Dins_ctg68	cagggcccggc [a]	1610	-

[a] TIRs present only in the 5`-*mar*

[b] Information of several hits representing *D. willistoni* strains, *D. paulistorum* strains, and *D. equinoxialis mar*-MITE and degenerated copies were omitted from this table, except sequences used in the phylogenetic tree in Figure 7.

(-) absence

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

Table S8 - Nucleotide divergence percentages of *mar* sequences found within and between species/strains.

Species	*D. willistoni*-Gd-H4-1	*D. willistoni*-L17	*D. willistoni*-00	*D. paulistorum*-L06	*D. paulistorum*-L12	*D. equinoxialis*	*D. tropicalis*	*D. insularis*
D. willistoni-Gd-H4-1	10.33%							
D. willistoni-L17	10.30%	10.40%						
D. willistoni-00	11.53%	11.59%	12.83%					
D. paulistorum-L06	31.72%	32.03%	31.82%	8.99%				
D. paulistorum-L12	22.63%	22.56%	23.57%	23.51%	24.51%			
D. equinoxialis	17.83%	17.54%	17.54%	38.74%	27.48%	20.91%		
D. tropicalis	31.18%	30.80%	30.99%	24.91%	31.37%	40.42%	22.02%	
D. insularis	35.06%	35.79%	35.61%	20.78%	32.27%	43.29%	19.96%	10.49%

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

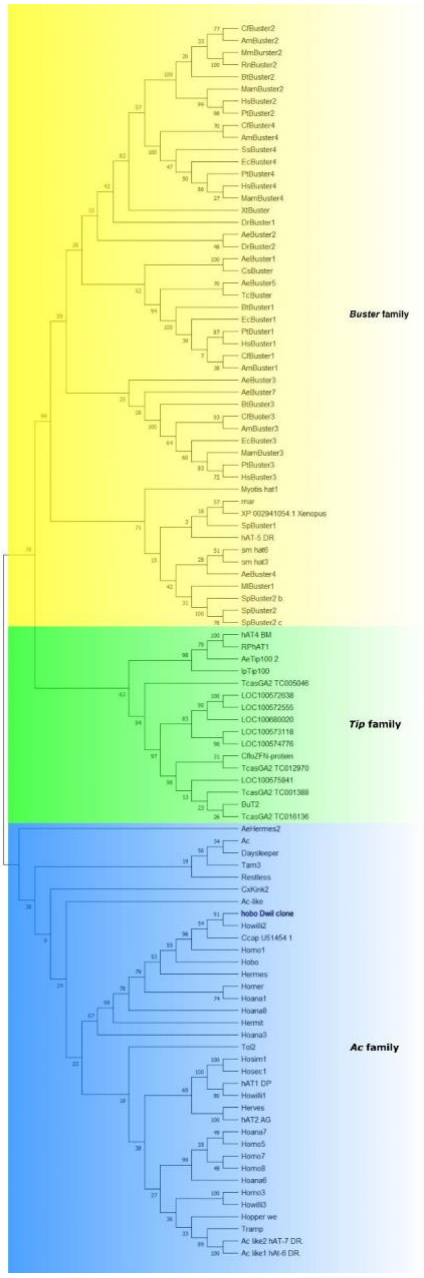


Figure S1 - Phylogenetic relationships of the *hAT* superfamily. Unrooted phylogenetic tree of *hAT* element amino-acid transposase sequences by maximum likelihood, using MEGA X and the Le-Gascuel model (Le and Gascuel 2008; Kumar et al. 2018). Clade colors denote the different families of *hATs* reported by Rossato et al. (2014). Bootstrap values of nodes are percentages for 1000 replicates.

Genetics and Molecular Biology
https://doi.org/10.1590/1678-4685-GMB-2021-0287

Supplementary material to "Interpopulation variation of transposable elements of the *hAT* superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): *in-situ* approach"

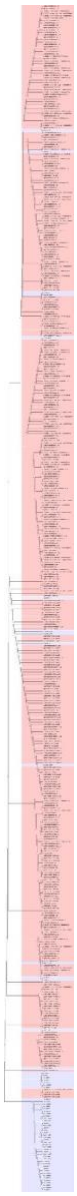


Figure S2 - Neighbor-Joining tree all partially complete *mar*, MITES, and relic copies. Identical, small, and/or very degenerate copies of each genome were excluded from this analysis (see Material and Methods). Node supports are bootstrap values (1000 replicates). Sequences used in Figure 7A and 7B are highlighted in blue, and other sequences are highlighted in red, except Dwil_Gd_scf2_3; Dins_ctg2309_5, Dins_ctg424, Dins_ctg1175, Dins_ctg1948; Dtro_ctg108_3, Dtro_ctg108_4, Dtro_ctg838, Dtro_ctg804, Dtro_ctg191.

CAPÍTULO 5

Considerações finais

Os elementos de transposição são muito diversos e amplamente distribuídos na árvore da vida, e também são fonte de variabilidade genética. Pensando nessa diversidade e distribuição, cada genoma é um ecossistema habitado por sua comunidade de TEs. Esses TEs interagem com o genoma hospedeiro e os outros componentes da célula por meio de interações sofisticadas. Essas interações incluem a multiplicação do TE, o controle dessa propagação pelo TE e pelo genoma hospedeiro, geração de novas famílias de TE devido a mutações na multiplicação ou com acrescimento de sequência do hospedeiro. O hospedeiro por sua vez, também pode utilizar esses TEs para suas funções. Desta forma, podem influenciar a adaptação dos genomas e consequentemente das espécies a diferentes ambientes, como no caso de *D. willistoni* e espécies do grupo.

Por isso, no Capítulo 2, propomos uma revisão onde compilamos sucintamente os principais estudos envolvendo *D. willistoni* e espécies próximas. O intuito foi remontar a “linha do tempo”, envolvendo principalmente os estudos evolutivos de *D. willistoni* e as chamadas espécies irmãs na América do Sul. Assim, demonstrar o que foi descoberto nestas espécies e o que ainda é uma incógnita aos pesquisadores, como por exemplo, a variabilidade cromossômica e consequentemente genômica nas diferentes populações de *D. willistoni*. Com isso, é possível afirmar que estas espécies foram vastamente estudadas e assim estabelecerem uma base muito solida para instigar novas abordagens como a era genômica vem nos mostrando.

A capacidade de *D. willistoni* e outras espécies do grupo habitarem os mais diferentes ambientes requer adapatações genômicas e cromossômicas como sugerimos nos Capítulos 3 e 4 desta tese. Estudando os elementos *hobo, BuT2* e *mar* utilizamos duas abordagens bastante distintas para demonstrar que um mesmo TE evolui diferencialmente em populações diferentes da mesma espécie. Também, como uma família de TEs autônomos podem em uma espécie

ou genoma constituir nova família de elementos MITEs, como no caso do transposon *mar*.

Além disso, também analisamos a presença do retroelemento *412* nos genomas de Dipteros, e também em espécies do grupo de *Drosophila willistoni* que possuem distribuição essencialmente neotropicais. Observamos que o retroelemento *412* está amplamente distribuído nos genomas de moscas. Porém, *412* completo incluindo as LTRs apenas nas espécies próximas a *D. melanogaster*. Nas espécies do grupo *willistoni* foi identificada a presença de *412* em todas as espécies analisadas, tanto por meio *in silico* quanto por meio de hibridização *Dot blot.* Utilizando hibridização *in situ* observamos a distribuição espalhada principalmente em eucromatina do retrotransposon nos cromossomos politênicos de *D. willistoni*.

Além disso, no Capítulo 4, corroborando com a literatura nossos resultados, principalmente baseando-se na reconstrução filogenética, mostram a intrincada relação entre os elementos da linhagem *412/mdg1,* onde provavelmente esses elementos orinam-se de quimerismos entre si. Por fim, demostramos um provável novo elemento da linhagem *412/mdg1* presente em quase todas as espécies das moscas tsé-tsé, do gênero *Glossina*, espécies estas com distribuição apenas no continente africano.

REFERÊNCIAS BIBLIOGRÁFICAS

Ågren JA and Wright SI (2011) Co-evolution between transposable elements and their hosts: A major factor in genome size evolution? Chromosom Res 19:777–786. doi: 10.1007/s10577-011-9229-0

Arkhipova IR (2017) Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mob DNA 8:19. doi: 10.1186/s13100-017-0103-2

Arkhipova IR and Yushenova IA (2019) Giant transposons in Eukaryotes: is bigger better? Genome Biol Evol 11:906–918. doi: 10.1093/gbe/evz041

Aziz RK, Breitbart M and Edwards RA (2010) Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res 38:4207–4217. doi: 10.1093/nar/gkq140

Bao W, Jurka MG, Kapitonov VV and Jurka J (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. Mol Biol Evol 26:983–993. doi: 10.1093/molbev/msp013

Bourgeois Y and Boissinot S (2019) On the population dynamics of junk: A review on the population genomics of transposable elements. Genes (Basel) 10:419. doi: 10.3390/genes10060419

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS *et al*. (2018) Ten things you should know about transposable elements. Genome Biol 19:1–12. doi: 10.1186/s13059-018-1577-z

Capy P (2021) Taming, domestication and exaptation: Trajectories of transposable elements in genomes. Cells 10:3590. doi: 10.3390/cells10123590

Carareto C, Monteiro-Vitorello C and Sluys M-A V (2015) Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos, 1st ed. FIOCRUZ, Rio de Janeiro

Chuong EB, Elde NC and Feschotte C (2017) Regulatory activities of

transposable elements: from conflicts to benefits. Nat Rev Genet 18:71–86. doi: 10.1038/nrg.2016.139

Cosby RL, Chang N-C and Feschotte C (2019) Host–transposon interactions: Conflict, cooperation, and cooption. Genes Dev 33:1098–1116. doi: 10.1101/gad.327312.119

Engels WR, Johnson-Schlitz DM, Eggleston WB and Sved J (1990) High-frequency *P* element loss in *Drosophila* is homolog dependent. Cell 62:515–525. doi: 10.1016/0092-8674(90)90016-8

Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. Science 338:758–767. doi: 10.1126/science.338.6108.758

Feschotte C and Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–68. doi: 10.1146/annurev.genet.40.110405.090448

Feschotte C, Zhang X and Wessler SR (2002) Minature inverted-repeat transposable elements (MITEs) and their relationship with established DNA transposons. In: Craig L, R. Craigie, M. Gellert and A. Lambowitz (eds) Mobile DNA II. American Society for Microbiology Press, Washington, DC, pp 1147–1158

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107. doi: 10.1016/0168-9525(89)90039-5

Fricker AD and Peters JE (2014) Vulnerabilities on the lagging-strand template: Opportunities for mobile elements. Annu Rev Genet 48:167–186. doi: 10.1146/annurev-genet-120213-092046

George JA, DeBaryshe PG, Traverse KL, Celniker SE and Pardue M-L (2006) Genomic organization of the *Drosophila* telomere retrotransposable elements. Genome Res 16:1231–1240. doi: 10.1101/gr.5348806

González J and Petrov D (2009) MITEs— The ultimate parasites. Science 325:1352–1353. doi: 10.1126/science.1179556

Haig D (2016) Transposable elements: Self-seekers of the germline, team-players of the soma. BioEssays 38:1158–1166. doi: 10.1002/bies.201600125

Han JS and Boeke JD (2005) *LINE-1* retrotransposons: Modulators of quantity and quality of mammalian gene expression? BioEssays 27:775–784. doi: 10.1002/bies.20257

Hancks DC and Kazazian HH (2016) Roles for retrotransposon insertions in human disease. Mob DNA. doi: 10.1186/s13100-016-0065-9

Hickman AB and Dyda F (2016) DNA transposition at work. Chem Rev 116:12758–12784. doi: 10.1021/acs.chemrev.6b00003

Hsia A and Schnable PS (1996) DNA sequence analyses support the role of interrupted gap repair in the origin of internal deletions of the maize transposon, MuDR. Genetics 142:603–618. doi: 10.1093/genetics/142.2.603

Joly-Lopez Z and Bureau TE (2018) Exaptation of transposable element coding sequences. Curr Opin Genet Dev 49:34–42. doi: 10.1016/j.gde.2018.02.011

Jurka J, Kapitonov VV, Kohany O and Jurka MV (2007) Repetitive sequences in complex genomes: Structure and evolution. Annu Rev Genomics Hum Genet 8:241–259. doi: 10.1146/annurev.genom.8.080706.092416

Kapitonov VV and Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet 9:411–412; author reply 414. doi: 10.1038/nrg2165-c2

Kofler R, Nolte V and Schlötterer C (2015) Tempo and mode of transposable element activity in *Drosophila.* PLOS Genet 11:e1005406. doi: 10.1371/journal.pgen.1005406

Kojima KK (2019) Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst 94:233–252. doi: 10.1266/ggs.18-00024

Koonin EV and Krupovic M (2017) Polintons, virophages and transpovirons: A tangled web linking viruses, transposons and immunity. Curr Opin Virol 25:7–15. doi: 10.1016/j.coviro.2017.06.008

Kramerov DA and Vassetzky NS (2011) Origin and evolution of SINEs in eukaryotic genomes. Heredity 107:487–495. doi: 10.1038/hdy.2011.43

Kramerov DA and Vassetzky NS (2005) Short retroposons in eukaryotic genomes. Int Rev Cytol 247:165–221. doi: 10.1016/S0074-7696(05)47004-7

Krupovic M, Dolja VV and Koonin EV (2019) Origin of viruses: Primordial replicators recruiting capsids from hosts. Nat Rev Microbiol 17:449–458. doi: 10.1038/s41579-019-0205-6

Lerat E, Brunet F, Bazin C and Capy P (1999) Is the evolution of transposable elements modular? Genetica 107:15–25. doi: 10.1023/A

Malik HS (2005) Evolution of retrotransposable elements Ribonuclease H evolution in retrotransposable elements. Cytogenet Genome Res 110:392–401. doi: 10.1159/000084971

Malik HS and Eickbush TH (2001) Phylogenetic analysis of Ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11:1187–1197. doi: 10.1101/gr.185101.1

McCullers TJ and Steiniger M (2017) Transposable elements in *Drosophila*. Mob Genet Elements 7:1–18. doi: 10.1080/2159256X.2017.1318201

Naville M, Warren IA, Haftek-Terreau Z, Chalopin D, Brunet F, Levin P, Galiana D and Volff J-N (2016) Not so bad after all: Retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. Clin Microbiol Infect 22:312–323. doi: 10.1016/j.cmi.2016.02.001

Rahman R, Chirn G, Kanodia A, Sytnikova YA, Brembs B, Bergman CM and Lau NC (2015) Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes. Nucleic Acids Res 43:10655–10672. doi: 10.1093/nar/gkv1193

Regner LP, Perelra MSO, Alonso CE V, Abdelhay E and Valente VLS (1996) Genomic distribution of *P* elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. J Hered 87:191–198.

Ros F and Kunze R (2001) Regulation of *Activator/Dissociation* transposition by replication and DNA methylation. Genetics 157:1723–1733. doi: 10.1093/genetics/157.4.1723

Rowold DJ and Herrera RJ (2000) *Alu* elements and the human genome. Genetica 108:57–72. doi: 10.1038/35038572

Rubin E and Levy AA (1997) Abortive gap repair: Underlying mechanism for *Ds* element formation. Mol Cell Biol 17:6294–6302. doi: 10.1128/MCB.17.11.6294

Schaack S, Gilbert C and Feschotte C (2010) Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evololution 25:537–546. doi: 10.1016/j.tree.2010.06.001.Promiscuous

Spradling AC, Bellen HJ and Hoskins RA (2011) *Drosophila P* elements preferentially transpose to replication origins. Proc Natl Acad Sci 108:15948–15953. doi: 10.1073/pnas.1112960108

Wallau GL (2013) Evolução de parasitas genômicos: Elementos genéticos móveis e seus superparasitas. Tese. doi: 10.13140/RG.2.1.3045.5442

Wallau GL, Ortiz MF and Loreto ELS (2012) Horizontal transposon transfer in Eukarya: Detection, bias, and perspectives. Genome Biol Evol 4:689–699. doi: 10.1093/gbe/evs055

Wells JN and Feschotte C (2020) A field guide to eukaryotic transposable elements. Annu Rev of Genetics A 54:1–23. doi: 10.1002/9781118500156.ch2

Wicker T, Robertson JS, Schulze SR, Feltus FA, Magrini V, Morrison JA, Mardis ER, Wilson RK, Peterson DG, Paterson AH *et al*. (2005) The repetitive landscape of the chicken genome. Genome Res 15:126–136. doi: 10.1101/gr.2438004

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al*. (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. doi: 10.1038/nrg2165-c4

Xiong Y and Eickbush TH (1990) Origin and evolution of retroelements based their reverse transcriptase sequences. EMBO J 9:3353–3362. doi: 1698615

Yuan Y-W and Wessler SR (2011) The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. Proc Natl Acad Sci 108:7884–7889. doi: 10.1073/pnas.1104208108

APÊNDICE

Outras publicações cientificas realizadas em colaboração com a autora da tese ao longo do doutorado.

Artigo aceito para publicação na revista *Genetics and Molecular Biology* parte da dissertação de mestrado de Anelise Fernandes e Silva pois este foi realizado em colaboraçao com a presente autora dessa tese.

---------- Forwarded message ---------
De: **Genetics and Molecular Biology** <onbehalfof@manuscriptcentral.com>
Date: seg., 31 de jan. de 2022 às 18:11
Subject: Genetics and Molecular Biology - Decision on Manuscript ID GMB-2021-0213.R2
To: <marindiadepra@gmail.com>

31-Jan-2022

Dear Dr. Deprá:

Your manuscript entitled "Study of four Neotropical species of tree crickets Oecanthus Serville, 1831 (Orthoptera, Gryllidae) using cytogenetic and molecular markers" submitted for publication in Genetics and Molecular Biology has been evaluated by the assigned Associate Editor who concluded that you have adequately addressed the concerns raised by the reviewers.

Therefore, I am pleased to inform you that your manuscript is now accepted for publication.

As a next step, your manuscript will be checked by one of our copy editors for correction of language errors and technical inconsistencies.

You will receive an edited manuscript by e-mail for appreciation and approval of eventual corrections before starting the layout process.

From now on, all correspondence will be directly with the Editorial Office by e-mail: editor@gmb.org.br. Please remember to refer to your Manuscript ID number whenever you contact us.

As you were informed in the submission, there is a publication charge for accepted manuscripts. You will be contacted by gestao.editorial@scielo.org or apc@scielo.org in order to proceed with its payment and also for checking the final PDF proofs.

NOTE: If you are a corresponding author who is a Member-in-good-standing of the Sociedade Brasileira de Genética (before the submission date), you are exempt from publication charges. Authors who were granted a waiver under our policies are exempt as well.

On behalf of the Editors of Genetics and Molecular Biology, I would like to take the opportunity to thank you for submitting your interesting contribution to our journal.

Sincerely,

Dr. Carlos Menck
Editor in Chief, Genetics and Molecular Biology
editor@gmb.org.br

Study of four Neotropical species of tree crickets *Oecanthus* Serville, 1831 (Orthoptera, Gryllidae) using cytogenetic and molecular markers

Anelise Fernandes e Silva[1,2], Thays Duarte de Oliveira[1], Natasha Ávila Bertocchi[2], Vera Lúcia da Silva Valente[1,2], Edison Zefa[3] and Maríndia Deprá[1,2]

[1] Programa de Pós-Graduação em Biologia Animal, Departamento de Zoologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

[2] Programa de Pós-Graduação em Genética e Biologia Molecular, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

[3] Programa de Pós-Graduação em Biologia Animal, Departamento de Zoologia, Ecologia e Genética, Instituto de Biologia, Universidade Federal de Pelotas, Capão do Leão, RS, Brazil.

Insect chromosome and phylogenetics

Keywords: Insect, karyotype, Chromosome, Bayesian Inference, Oecanthinae

Corresponding author:

Maríndia Deprá

Laboratório de *Drosophila*, Departamento de Genética, Universidade Federal do Rio Grande do Sul. Avenida Bento Gonçalves, 9500, Bairro Agronomia. CEP: 91509-900, Porto Alegre, RS, Brazil.

E-mail: marindiadepra@gmail.com

ORCID: 0000-0003-4568-1869

ORCID:

Anelise Fernandes e Silva - 0000-0003-2943-4788

Primeira página do artigo submetido e em revisão para a revista *Animals*, parte da tese de Thays Duarte de Oliveira pois este foi realizado em colaboração com a presente autora dessa tese.

animals MDPI

Article

Genomic organization of microsatellites and *LINE-1*-like retrotransposons: evolutionary implications for *Ctenomys minutus* (Rodentia: Ctenomyidae) cytotypes

Thays Duarte de Oliveira[1*], Natasha A Bertocchi [2], Rafael Kretschmer[3], Edivaldo HC de Oliveira[4,5], Marcelo de Bello Cioffi[6], Thomas Liehr[7*] and Thales RO de Freitas[1, 2]

[1]Programa de Pós-Graduação em Biologia Animal, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

[2]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[3] Departamento de Ecologia, Zoologia e Genética, Instituto de Biologia, Universidade Federal de Pelotas, Pelotas 96010-900, RS, Brazil

[4] Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, Pará, Brazil

[5] Laboratório de Cultura de Tecidos e Citogenética, SAMAM, Instituto Evandro Chagas, Ananindeua, Pará, Brazil

[6] Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil

[7] Jena University Hospital, Institute of Human Genetics, Jena, Germany

*Correspondence: Thomas.Liehr@med.uni-jena.de

Simple Summary: http://www.mdpi.com/2076-2615/6/6/40/htm

Abstract: The Neotropical underground rodents of the genus *Ctenomys* (Rodentia: Ctenomyidae) comprises about 65 species, which harbor the most significant chromosomal variation among mammals (2n=10 to 2n=70). Among them, *C. minutus* stands out with 45 different cytotypes already identified, among which seven parental ones, named A to G, are parapatrically distributed in the Coastal Plains of Southern Brazil. Looking for possible causes that led to such extensive karyotype diversification we performed chromosomal mapping of different repetitive DNAs, including microsatellites and long interspersed element-1 (*LINE-1*) retrotransposons in the seven parental cytotypes. Although microsatellites were found mainly in the centromeric and telomeric regions of the chromosomes, different patterns occur for each cytotype, thus revealing specific features. Likewise, the *LINE-1*-like retrotransposons also showed a differential distribution for each cytotype, which may be linked to stochastic loss of *LINE-1* in some populations. The results provided evidence for a correlation between the repetitive genomic (microsatellite motifs $(A)_{30}$, $(C)_{30}$, $(CA)_{15}$, $(CAC)_{10}$, $(CAG)_{10}$, $(CGG)_{10}$, $(GA)_{15}$, and $(GAG)_{10}$) content and of evolutionary breakpoints (fusion of chromosomes 20/17; fission and inversion in the short arm of chromosome 2; fusion of chromosomes 23/19; different combinations of centric and tandem fusions of chromosomes 22/24/16), highlighting their direct impact in promoting chromosomal rearrangements.

Keywords: chromosomal rearrangements; FISH; *LINE-1*; Simple sequence repeats; retrotransposons

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Animals* **2022**, *12*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Last-name

Received: date
Accepted: date
Published: date