

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

ELEMENTO TRANSPONÍVEL *GALILEO* NO GENOMA DE ESPÉCIES DO GRUPO *WILLISTONI* DE
DROSOPHILA (DIPTERA: DROSOPHILIDAE)

JULIANA WOLMANN GONÇALVES

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do grau de Mestre.

VERA LÚCIA DA SILVA VALENTE GAIESKY – Orientadora

Porto Alegre, março de 2010

Este trabalho foi desenvolvido no Laboratório de Drosophila da Universidade Federal do Rio Grande do Sul – UFRGS, sob a orientação da Profa. Dra. Vera Lúcia da Silva Valente Gaiety e no Laboratório de Biologia Molecular do Vale do Rio dos Sinos – UNISINOS, sob a orientação do Prof. Dr. Victor Hugo Valiati, com bolsa e recursos do CNPq, assim como do Programa de Pós Graduação em Genética e Biologia Molecular da UFRGS.

Aos meus pais, Valdir Costa Gonçalves e
Beatriz Wolmann Gonçalves, minhas
estrelas guias.

Ao Nobu, meu esposo

“Mas te vejo e sinto
O brilho desse olhar
Que me acalma
Me traz força pra encarar tudo...”

Dia Especial – Cidadão Quem

AGRADECIMENTOS

Àqueles que me orientaram durante o desenvolvimento deste trabalho: Dr^a. Vera Lúcia da Silva Valente Gaiesky, obrigada pela oportunidade de trabalho que me confiaste, pela orientação sempre muito entusiasmada e contagiante, pelo seu profissionalismo; Dr. Victor Hugo Valiati, agradeço por toda a sua dedicação e seu envolvimento constante no desvendar dos meus dados e, principalmente, pela amizade e carinho.

Aos colegas de laboratório e amigos Carolina Garcia, Cleverton Hochmüller, Dirleane Rossato, Gisele Silva, Hermes Schmitz, Juliana Cordeiro, Marícia Fantinel, Mário Josias Müller, por tornarem o Laboratório de *Drosophila* um ambiente de trabalho tão agradável. Pelas brincadeiras, fotografias, festas e passeios tão cheios de alegria, momentos que ficarão na lembrança para sempre.

À Adriana Ludwig, Adriana Sassi, Ana Cristina Garcia, Fabiano Torres, Maríndia Deprá, Mônica Blauth, Nina Motta e Lizandra Robe, colegas que tive a oportunidade de conhecer no laboratório de *Drosophila* e que de diferentes maneiras contribuíram na execução do meu trabalho.

À Luciana Rodrigues, obrigada pela grande parceria e por me encontrar ao longo dos corredores do departamento sempre assim: extremamente radiante!

À Bibiana Macedo e ao Igor Radamés pelo suporte técnico dado no Laboratório de *Drosophila* da UFRGS e no Laboratório de Biologia Molecular da UNISINOS, respectivamente.

Ao Elmo Cardoso e à Ellen Mezzeck, dupla de imensa eficiência e de um carisma inigualável. Obrigada por todos os esclarecimentos e auxílios ao longo destes dois anos.

Ao Ricardo Weber e à Gisele Passaia do Laboratório de Genética Molecular Vegetal, sempre dispostos a socorrer colegas em situações um tanto complicadas.

À Andresa Cardoso Grandini, Jamile Queiroz e Rosane Giacomini, pela oportunidade que me deram de discutir algumas vezes protocolos e padronizações de técnicas, pelo carinho e pela amizade.

À Carolina Domingues, um encanto de pessoa, impossível não ficar de bem com a vida depois de encontrá-la. Agradeço pelos momentos descontraídos que juntas tivemos no campus.

Aos meus queridos amigos Carla Dorneles, Matheus Levi, Clemir Tarouco, Paulo Renato Duarte, Ângela Souza e Rodrigo Oyarzabal, os quais sempre estiveram ao meu

lado torcendo pelos os melhores resultados desde o processo seletivo até os últimos experimentos. Obrigada pelo constante apoio!

À minha querida amiga Karla Reis, pelo trabalho como linguista na revisão do *abstract*.

À Trícia Kommers Albuquerque e Paulo Eduardo Raimann, hoje grandes amigos, agradeço os primeiros e diversificados ensinamentos práticos em Genética e Biologia Molecular, que despertaram em mim um grande interesse nesta área de estudo.

Ao Nobuhiro Marcos Honda (Nobu), meu esposo, por me socorrer em situações inusitadas, por realizar eventuais consertos urgentes no laboratório, por ouvir alguns apontamentos noturnos (um tanto engraçados) sobre técnicas laboratoriais e ainda por me auxiliar em edição de imagens, assim como na revisão do presente trabalho. Acima de tudo pelo amor e por ser um companheiro com infinita disposição e paciência.

À família Honda e Hanson, pelo total apoio, pelo carinho e pela compreensão de minhas ausências.

À Gabriela Chesim, por dar dicas bem descontraídas (digamos alternativas), aumentando minha disposição no período mais conturbado deste mestrado.

À minha querida e amada família, aos meus pais, Valdir Costa Gonçalves e Beatriz Wolmann Gonçalves, que sempre valorizaram a minha formação acadêmica abdicando muitas vezes de escolhas pessoais e ao meu irmão, William Wolmann Gonçalves, que mesmo à distância, tornou inúmeros momentos de trabalho mais cheios de alegria e entusiasmo, recarregou minhas energias com palavras incentivadoras e animadoras. Vocês dão sentido às minhas escolhas.

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

SUMÁRIO

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES.....	8
RESUMO.....	10
ABSTRACT.....	12
CAPÍTULO I	
INTRODUÇÃO.....	14
Elementos transponíveis.....	14
Evidências da Relação de <i>TEs</i> com Inversões Cromossômicas.....	15
Elementos <i>foldback</i>	17
O Elemento <i>Galileo</i> como o principal agente de reorganização cromossômica em <i>Drosophila</i>	18
Grupo <i>willistoni</i>	34
CAPÍTULO II	
OBJETIVOS.....	44
CAPÍTULO III	
METODOLOGIA.....	45
Manutenção dos estoques de <i>Drosophila</i>	45
Extração de DNA	45
Técnica da Reação em Cadeia da Polimerase (<i>PCR</i>)	45
Medição do comprimento dos fragmentos em número de pares de base (pb).....	47
<i>Dot blot</i>	47
Purificação dos fragmentos amplificados.....	48
Clonagem e sequenciamento.....	48
Análise e edição das sequências.....	48
CAPÍTULO IV	
RESULTADOS.....	50
Isolamento de <i>Galileo</i> no grupo <i>willistoni</i>	50
Caracterização molecular das variantes isoladas.....	51
Análise das sequências de aminoácidos.....	54
CAPÍTULO V	
DISCUSSÃO.....	60
CAPÍTULO VI	
PERSPECTIVAS.....	64
REFERÊNCIAS BIBLIOGRÁFICAS.....	65

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES

<i>Acph1</i>		fosfatase ácida-1
<i>Adh</i>		álcool desidrogenase
<i>Amd</i>		alfa metildopa
<i>BuTs</i>		transposons de <i>D. buzzatii</i>
<i>COI</i>		citocromo oxidase I
<i>COII</i>		citocromo oxidase II
<i>hb</i>		<i>hunchback</i>
<i>HT</i>	(<i>Horizontal Transfer</i>)	transferência horizontal
<i>IPI</i>	(<i>Interpulse interval</i>)	intervalo interpulso
<i>IR-FD</i>	(<i>Inverted repeat-Flanking Domain</i>)	domínio flanqueador da região invertida
<i>IR-ID</i>	(<i>Inverted repeat-Inner Domain</i>)	domínio interno da região invertida
<i>IR-OD</i>	(<i>Inverted repeat-Outer Domain</i>)	domínio externo da região invertida
<i>IRs</i>	(<i>Inverted Repeats</i>)	repetições invertidas
<i>ITRs</i>	(<i>Inverted Terminal Repeats</i>)	repetições terminais invertidas
<i>LINEs</i>	(<i>Long Interspersed Nuclear Elements</i>)	elementos nucleares intercalados longos
<i>LTRs</i>	(<i>Long Terminal Repeats</i>)	repetições terminais longas
<i>M</i>	(<i>Middle Domain</i>)	domínio central
<i>MITEs</i>	(<i>Miniature Inverted Repeat Transposable Elements</i>)	
<i>NAHR</i>	(<i>Nonallelic Homologous Recombination</i>)	recombinação homóloga não-alélica
<i>ORFs</i>	(<i>Open Reading Frames</i>)	regiões de leituras abertas
<i>PCR</i>	(<i>Polymerase Chain Reaction</i>)	reação em cadeia da polimerase
<i>per</i>		<i>period</i>
<i>SINEs</i>	(<i>Short Interspersed Nuclear Elements</i>)	elementos nucleares intercalados curtos
<i>TEs</i>	(<i>Transposable Elements</i>)	elementos transponíveis
<i>THAP</i>	(<i>Thanatos-associated protein</i>)	domínio de ligação ao DNA
<i>Tpase</i>		transposase
<i>TSD</i>	(<i>Target Site Duplication</i>)	duplicação do sítio alvo
<i>UPGMA</i>	(<i>Unweighted Pair Group Method by Arithmetic Averages</i>)	
<i>Xdh</i>		xantina-desidrogenase

μL

M

ng

°C

<

microlitro

molar

nanograma

grau Celsius

menor

RESUMO

Galileo foi identificado em *D. buzzatii* e classificado como transposon do tipo *foldback* devido às suas longas *IRs* (repetições invertidas), que capacitam o elemento a formar uma estrutura secundária que leva a quebras e rearranjos cromossômicos. Evidências indicam que *Galileo* foi responsável pela geração de três inversões naturais segregantes em populações de *D. buzzatii*. A descoberta de *Galileo* em seis das 12 espécies de *Drosophila* sequenciadas, incluindo a *D. willistoni*, através de um estudo *in silico*, foi promissora para a busca de agentes causadores de quebras e rearranjos cromossômicos nesta espécie, que é altamente polimórfica, e em espécies a ela relacionadas. Nesse contexto, o presente estudo foi o primeiro a caracterizar o transposon *Galileo* em populações naturais de espécies do grupo *willistoni* de *Drosophila*. Para a investigação da presença de *Galileo* foram construídos *primers* com base na sequência de *Galileo* encontrada no genoma da linhagem sequenciada (GdH4) de *Drosophila willistoni*, para amplificar um fragmento do domínio central do elemento, assim como da inferida transposase e de seu domínio *THAP*. Os fragmentos polimórficos foram clonados e sequenciados para posterior caracterização molecular. O alinhamento das sequências nucleotídicas e de aminoácidos foi obtido no *software Muscle* com posterior edição no *Bio Edit* e o programa *Gene Doc* foi utilizado na tradução das sequências. Foram investigadas as espécies *Drosophila willistoni*, *D. tropicalis*, *D. equinoxialis*, *D. insularis* e quatro das semiespécies de *D. paulistorum* (Amazônica, Andino-Brasileira, Interior e Orinocana) do subgrupo *willistoni*, e *D. nebulosa*, *D. capricorni* e *D. fumipennis*, subgrupo *bocainensis*, do grupo *willistoni*, pelas técnicas de *PCR* e *Dot blot*. Foi encontrada variabilidade intra e interespecífica no tamanho dos fragmentos amplificados de *Galileo*, sendo que todas as sequências analisadas são indicativas de cópias defectivas. Foram caracterizadas 22 variantes a partir das sequências obtidas, permitindo a construção de um dendrograma de acordo com a presença ou ausência das variantes no genoma de todas as espécies estudadas por meio do *software NTSYSpc 2.1*. Os dados resultaram em dois clados bem-definidos: (1) formado pelas quatro semiespécies de *D. paulistorum* e (2) pelas espécies crípticas *D. tropicalis*, *D. insularis* e *D. equinoxialis*. *Drosophila willistoni* do subgrupo *willistoni*, *D. sucinea*, *D. nebulosa* e *D. capricorni* do subgrupo *bocainensis*, apresentaram variantes exclusivas e, assim, formaram ramos independentes aos dois agrupamentos. Entre todas as espécies

crípticas do subgrupo *willistoni* somente *D. willistoni* não compartilha qualquer uma das suas quatro variantes. O quadro que emerge de nossos achados é sugestivo de mais de um evento evolutivo, implicando tanto a manutenção por ancestralidade de certas variantes quanto a ocorrência de perdas estocásticas. Além disso, o fenômeno de introgressão, já bem-documentado dentro do subgrupo *willistoni*, pode ter sido o mecanismo pelo qual tenham sido transferidas sequências de *Galileo* entre alguns táxons estudados, principalmente entre algumas semiespécies de *D. paulistorum*. A singularidade de *D. willistoni* entre as crípticas quanto às diferentes sequências de *Galileo* em seu genoma, pode ser consequência da presença de elementos quiméricos, uma vez que foi observada a inserção secundária de outros elementos em *Galileo* de *D. buzzatii*. Considerando a peculiaridade de *Galileo*, como gerador de inversões cromossômicas em populações naturais de *D. buzzatii*, nossos resultados sugerem que *Galileo* também pode ter desempenhado um papel importante na evolução genômica no grupo *willistoni* de *Drosophila*, uma vez que *D. willistoni* e *D. paulistorum* apresentam muitos arranjos cromossômicos segregando em suas populações naturais.

ABSTRACT

Galileo was identified in *Drosophila buzzatii* and classified as a *foldback*-like transposon due to its long *IRs* (inverted repeats), which allow the element to form a secondary structure that leads to breaks and chromosomal rearrangements. Several evidences indicate that *Galileo* was responsible for the generation of three natural segregating inversions in populations of *D. buzzatii*. The finding of *Galileo* in six of the 12 species of *Drosophila* that were sequenced, including *D. willistoni*, through an *in silico* study, is promising for the search of agents that cause chromosomal breaks and rearrangements in this species, that is highly polymorphic, and in related species. Based on this context, this was the first work to characterize the *Galileo* transposon in natural populations of the *willistoni* group of *Drosophila*. To detect the presence of *Galileo*, we used primers based on the sequence of *Galileo* found in the genome of the sequenced lineage (GdH4) of *Drosophila willistoni*, so that a fragment of the central domain of the element is amplified, as well as the putative transposase and its THAP domains. The polymorphic fragments were cloned and sequenced for later molecular characterization. The alignment of the nucleotide sequences and amino acids was obtained in the *Muscle* software with later edition in the *Bio Edit*, and the *Gene Doc* program was used to translate the sequences. We investigated, by PCR and Dot blot, the following species: *D. willistoni*, *D. tropicalis*, *D. equinoxialis*, *D. insularis* and four semispecies of *D. paulistorum* (Amazonian, Andean-Brazilian, Interior and Orinocan) of the subgroup *willistoni*, besides *D. nebulosa*, *D. capricorni* and *D. fumipennis*, subgroup *bocainensis*, group *willistoni*. The size variability of the *Galileo* amplified fragments was intra and interspecific; all the analyzed sequences indicate defective copies. Twenty two variants were identified in the obtained sequences, allowing the construction of a dendrogram based in the presence or absence of the variants in the genomes of all the species studied through the NTSYSpc 2.1 software. The data produced two well-defined clades: (1) formed by the four semispecies of *D. paulistorum* and (2) by the sibling species *D. tropicalis*, *D. insularis* and *D. equinoxialis*. *D. willistoni* (subgroup *willistoni*), *D. sucinea*, *D. nebulosa* and *D. capricorni* (subgroup *bocainensis*) have exclusive variants of *Galileo*, forming two independent branches of the two clades. Among all the sibling species of the subgroup *willistoni*, only *D. willistoni* does not share any of its four variants.

The results of our observations suggest more than one evolutionary event, indicating both the maintenance by ancestry of some variants and the occurrence of stochastic loss. Besides, the introgression phenomenon, well-documented in the *willistoni* subgroup, may have been the mechanism through which the *Galileo* sequences were transferred between some taxons that were studied, specially between some semispecies of *D. paulistorum*. The uniqueness of *D. willistoni* among siblings, regarding the different *Galileo* sequences in its genome, can be a consequence of the presence of quimeric elements, since a secondary insertion of other elements in *Galileo* of *D. buzzatii* was observed. Taking into account the *Galileo* element singularities, as a generator of chromosomal inversions in natural populations of *D. buzzatii*, our findings suggest that *Galileo* could also have played an important role in the genomic evolution of the *Drosophila willistoni* group, since *D. willistoni* and *D. paulistorum* have many chromosomal rearrangements segregating in their natural populations.

CAPÍTULO I

INTRODUÇÃO

Elementos Transponíveis

Elementos transponíveis (*TEs*, do inglês *Transposable Elements*) são componentes intrínsecos aos genomas de todos os organismos já investigados, com exceção do *Plasmodium falciparum* e provavelmente espécies intimamente relacionadas. Classificados como DNA medianamente repetido, possuem capacidade de mobilização, causam duplicação do sítio alvo de inserção (*TSD*, do inglês *Target Site Duplication*), são dotados de polimorfismos de sítios de inserção e possuem número variável de cópias dentro e entre espécies. Essas características são importantes para o reconhecimento das diferentes classes de *TEs* (Capy *et al.*, 1998, Wicker *et al.*, 2007).

Wicker *et al.* (2007) propuseram um sistema de classificação para *TEs*, o qual, em uma ordem hierárquica, inclui os níveis de classe, subclasse, ordem, superfamília, família e subfamília. O nível mais elevado divide os *TEs* pela presença ou ausência de um intermediário de RNA na transposição, como elementos da classe I e da classe II, respectivamente, como proposto por Finnegan (1989). A subclasse previamente usada para separar *TEs* da classe I dotados de repetições terminais longas (*LTRs*, do inglês *Long Terminal Repeats*) dos elementos intercalados longos e curtos (*LINEs* e *SINEs*, do inglês *Long and Short Interspersed Nuclear Elements*), distingue agora elementos da classe II que utilizam o mecanismo replicativo daqueles que se transpõem pelo mecanismo conservativo. A ordem marca diferenças importantes no mecanismo de inserção, conseqüentemente, na organização geral e enzimologia, substituindo assim a subclasse para *TEs* da classe I. As superfamílias dentro de uma ordem compartilham uma estratégia de replicação, mas são distinguidas por características uniformes e amplamente difundidas, tais como a estrutura protéica e domínios não-codificantes. Além disso, superfamílias diferem na presença e tamanho da duplicação do sítio alvo de inserção. Superfamílias são divididas em famílias, as quais são definidas pela conservação da sequência de DNA, embora essa seja mínima e restrita à região codificante. Subfamílias são definidas com base nos dados filogenéticos e podem, em casos específicos, servir para distinguir elementos autônomos de não-autônomos.

Em um mesmo genoma podem existir cópias completas de um *TE*, também denominadas de autônomas, e elementos deletados, também referidos como não-autônomos ou defectivos. Recentemente tem sido postulado (Wicker *et al.*, 2007) que as categorias autônomas e não-autônomas, devam ser consideradas em função das seguintes particularidades: um elemento seria autônomo simplesmente se ele codificar todos os domínios necessários para a sua transposição, independentemente de estar funcional ou ativo; não-autônomos, por sua vez, são definidos como qualquer grupo de elementos que tenha perdido algum ou todos os domínios encontrados nos elementos autônomos. Normalmente elementos não-autônomos têm região codificadora altamente degenerada.

A transposição de *TEs* pode ocorrer entre regiões de um mesmo cromossomo, entre cromossomos diferentes (Capy *et al.*, 1998) ou ainda entre genomas de diferentes espécies por meio de vetores, processo chamado de transferência horizontal (*HT*, do inglês *Horizontal Transfer*) (Loreto *et al.*, 2008).

O impacto da presença de *TEs* pode adquirir proporções distintas dependendo de sua localização no genoma do organismo. Enquanto cópias ativas são encontradas, principalmente na eucromatina, cópias defectivas e rearranjadas são observadas na heterocromatina. No entanto, permanece ainda a dúvida se existe uma preferência insercional de *TEs* por determinadas regiões ou se este padrão é resultado de diferentes pressões seletivas (Capy *et al.*, 1998).

Um aumento de evidências sugere que alguns elementos transponíveis podem ter um importante papel na evolução de espécies seja via alterações da função gênica através das inserções e excisões, pela indução de rearranjos cromossômicos, ou ainda permitindo o surgimento de novos genes e sequências regulatórias (McDonald, 1993; Wessler *et al.*, 1995; Kidwell & Lisch, 1997; Kidwell & Holyoake, 2001; Kazazian, 2004; Feschotte & Pritham, 2007; Feschotte, 2008).

Evidências da Relação de *TEs* com Rearranjos Cromossômicos

Inversões cromossômicas são notavelmente abundantes no gênero *Drosophila*, sendo observadas como polimorfismos intra-específicos, assim como diferenças interespecíficas fixadas (Bhutkar *et al.*, 2008).

TEs podem gerar inversões cromossômicas por recombinação homóloga intracromossomal entre duas cópias do mesmo *TE* posicionados em orientações opostas

(Petes & Hill, 1988). Este mecanismo é conhecido como recombinação ectópica ou recombinação homóloga não-alélica (*NAHR*, do inglês *Nonallelic Homologous Recombination*). Podem também induzir inversões, bem como outros tipos de rearranjos, quando duas extremidades oriundas de diferentes variantes de um *TE* participam juntas em um evento aberrante de transposição. O resultado depende da localização e orientação das cópias no cromossomo e na região cromossômica onde se inserem. Se os dois exemplares estão localizados em cromátides irmãs ou em cromossomos homólogos, o processo é chamado de inserção de elemento híbrido (*hybrid element insertion*) (Svoboda *et al.*, 1995; Gray *et al.*, 1996; Gray, 2000). Quando os dois exemplares estão localizados em regiões vizinhas da mesma cromátide, o mecanismo é conhecido como transposição de extremidades invertidas (*reversed ends transposition*) (Huang & Dooner, 2008; Zhang *et al.*, 2009).

Evidências da implicação dos *TEs* na origem de rearranjos cromossômicos, através do mecanismo de recombinação ectópica, foram fornecidas por Daveran-Mingot *et al.* (1998), Parkhill *et al.* (2001), Deng *et al.* (2002), Brinig *et al.* (2006), Redder & Garrett (2006) e Beare *et al.* (2009) para bactérias, sendo que algumas linhagens experimentaram um impressionante grau de rearranjo (Chain *et al.*, 2004; Parkhill *et al.*, 2003; Cho *et al.*, 2007; Reith *et al.*, 2008). Da mesma forma deleções, duplicações, translocações e inversões mediadas pela recombinação de elementos *Ty* em leveduras foram relatadas por Roeder (1983), Petes & Hill (1988), Picologlou *et al.* (1988) e Kim *et al.* (1998). Nos mamíferos elementos intercalados longos e curtos (*LINEs* e *SINEs*) têm sido implicados na geração, por recombinação ectópica, de 50 inversões fixadas entre humanos e chimpanzés (Schwartz *et al.*, 1998; Lee *et al.*, 2008).

A implicação circunstancial de *TEs* na origem de inversões em *Drosophila* foi primeiramente obtida por hibridização *in situ* com sondas de diferentes elementos: *hobo* em *D. melanogaster* (Lim, 1988; Lyttle & Haymer, 1992), *P* em *D. willistoni* (Regner *et al.*, 1996), *Penelope* e *Ulysses* em espécies do grupo *virilis* (Zelentsova *et al.*, 1999; Evgen'ev *et al.*, 2000). A evidência da implicação de *TEs* na geração de inversões é limitada. Em *D. pseudoobscura*, a inversão polimórfica *Arrowhead* e um número de inversões fixadas foram geradas por recombinação ectópica entre repetições de 128 pb e 315 pb, mas a natureza dessas repetições ainda permanece desconhecida (Richards *et al.*, 2005). Cópias de *MITEs* (do inglês *Miniature Inverted Repeat Transposable Elements*)

foram encontradas flanqueando a inversão *In(4)a* em *D. americana* e responsabilizadas pelo rearranjo através da recombinação entre suas repetições (Evans *et al.*, 2007). Em *D. buzzatii*, o elemento *foldback Galileo*, tem sido implicado na geração de três inversões polimórficas ($2j$, $2q^7$ e $2z^3$) através da recombinação entre sequências em orientações opostas (Cáceres *et al.*, 1999; Casals *et al.*, 2003 e Delprat *et al.*, 2009).

Elementos *foldback*

Elementos *foldback* foram descritos pela primeira vez em *Drosophila melanogaster* (Potter *et al.*, 1980), subsequentemente em outros organismos como, por exemplo, em *Strongylocentrotus purpuratus* (Liebermann *et al.*, 1983), *Chiromonus thummi* (Hankeln & Schmidt, 1990), *Caenorhabditis elegans* (Yuan *et al.*, 1991), *Lycopersicon esculentum*, *L. chmielwvskii*, *L. hirsutum*, *Solanum fenderli*, *S. polyadenium* (Rebatchouk & Narita, 1997), *Arabidopsis thaliana* (Adé & Belzile, 1999), *Oryza sativa* (Cheng *et al.*, 2000), *Cionas intestinalis* (Simmen & Bird, 2000); *Drosophila buzzatii* (Cáceres *et al.*, 2001).

Estes elementos, pertencentes à classe II, apresentam características bem peculiares. Entre as principais características que os distinguem são as muito longas repetições invertidas (*IRs*, do inglês *Inverted Repeats*) ao contrário das repetições terminais invertidas (*ITRs* do inglês *Inverted Terminal Repeats*). Estas usualmente se estendem por quase o elemento inteiro e são separadas por um domínio central (*M*, do inglês *Midle Domain*) que varia no seu comprimento e composição. Como consequência de sua estrutura, quando desnaturadas, as duas repetições invertidas (*IRs*) podem dobrar e permanecerem pareadas, dando origem a uma estrutura secundária estável, a qual inspirou o nome do elemento (Potter *et al.*, 1980). As repetições invertidas possuem uma organização modular com três possíveis domínios: o domínio que flanqueia toda a sequência (*IR-FD*, do inglês *Inverted repeat-Flanking Domain*); o domínio externo (*IR-OD*, do inglês *Inverted repeat-Outer Domain*), o qual inclui uma série de repetições imperfeitas em tandem; e o domínio interno (*IR-ID*, do inglês *Inverted repeat-Inner Domain*), o qual habitualmente contém sequências ricas em A-T (Adé & Belzile, 1999; Simmen & Bird, 2000) (Figura 1). A capacidade de codificar não tem sido encontrada na vasta maioria de transposons *foldback* descritos e seu mecanismo de transposição não é plenamente compreendido (Yuan *et al.*, 1991; Rebatchouk & Narita, 1997; Adé & Belzile,

1999; Cheng *et al.*, 2000; Simmen & Bird, 2000; Casals *et al.*, 2005), embora recentemente Marzo *et al.* (2008) sugerem uma alternativa, pelo menos, no que refere ao elemento *Galileo*.

Em nível funcional, elementos *foldback* são caracterizados por sua capacidade de induzir instabilidade e processos de recombinação, conduzindo para a geração espontânea de mutações e rearranjos cromossômicos em *D. melanogaster* (Levis *et al.*, 1982; Bingham & Zachar, 1989; Smith & Corces, 1991) e em *D. buzzatii* (Cáceres *et al.*, 1999; Casals *et al.*, 2003; Delprat *et al.*, 2009). Esses processos parecem ocorrer por recombinação ectópica entre sequências do mesmo elemento (Collins & Rubin, 1984; Cáceres *et al.*, 1999; Delprat *et al.*, 2009) que poderiam também explicar a grande heterogeneidade estrutural dessa família (Hoffman-Liebermann *et al.*, 1985; Cheng *et al.*, 2000; Windsor & Waddel, 2000). Além disso, elementos *foldback*, devido à presença das longas *IRs*, possuem a habilidade de formar estruturas secundárias, as quais estimulam a produção de quebras de dupla fita e recombinação (Lobachev *et al.*, 1998) e, conseqüentemente, sendo uma fonte de instabilidade genômica (Zhou *et al.*, 2001). Adé & Belzile (1999) verificaram que em *Arabidopsis thaliana* os elementos *Hairpin* apresentam um potencial para formar a estrutura secundária muito estável, pois estes autores obtiveram, para esses elementos *foldback*, valores de ΔG entre -54 a -92 kcal mol⁻¹.



Figura 1. Representação esquemática da organização modular dos elementos *foldback* e da sequência duplicada do sítio alvo de inserção (*TSD*). *IR-FD*, domínio flanqueador da repetição invertida; *IR-OD*, domínio externo da repetição invertida; *IR-ID*, domínio interno da repetição invertida; ()_n, o número de repetições do domínio externo pode variar; *M*, domínio central. As diferentes repetições invertidas (*IRs*) estão representadas nas mesmas cores em ambos os lados, como também o *TSD*. Baseada em Casal *et al.* (2005).

O Elemento *Galileo* como o principal agente de reorganização cromossômica em *Drosophila*

Em 1999 Cáceres *et al.* sustentaram que o mecanismo de recombinação ectópica seria o responsável pela geração de uma inversão (2j), comumente observada no segundo cromossomo de *D. buzzatii*. A partir da clonagem e sequenciamento dos pontos de quebra

da inversão, os autores demonstraram que esses pontos possuíam largas inserções, as quais não estão presentes no arranjo *2st* ancestral. A inserção entre A e C (ponto de quebra distal em *2j*) apresenta 392 pb sendo 106 pb correspondentes às repetições terminais invertidas (*ITRs*), que poderiam representar as longas regiões invertidas (*IRs*), características dos elementos *foldback*. A inserção entre B e D (ponto de quebra proximal em *2j*) possui 4319 pb, com uma complexa estrutura interna. Apresenta uma *ITR* proximal como na inserção AC (106 pb) e uma distal com 47 pb correspondentes a sequência inicial da repetição de 106 pb. Os 180 pb centrais da inserção AC e os que acompanham a *ITR* proximal de BD são homólogos, com uma identidade de 95%, mas em orientações opostas. Flanqueando cada inserção, sequências de 7 pb correspondem ao sítio duplicado durante o evento da inserção, contudo, dispostas separadamente e invertidas, o que sugere fortemente que a inversão *2j* decorreu de um pareamento intracromossomal e a recombinação entre as duas sequências homólogas inseridas em sítios distantes, em orientações opostas (Figura 2).

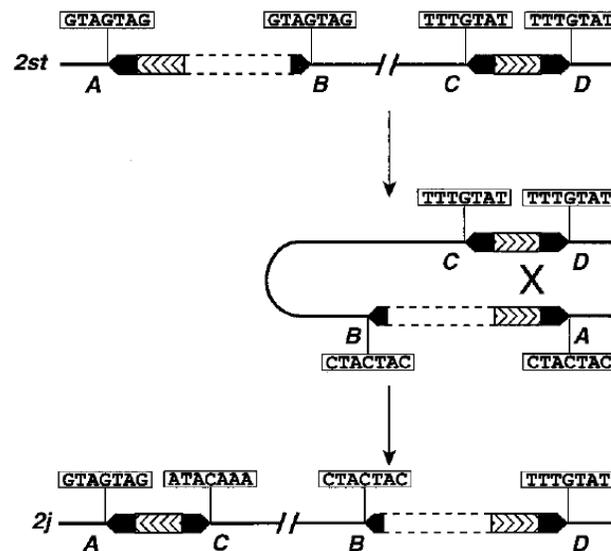


Figura 2. Modelo para a geração da inversão cromossômica *2j* em *D. buzzatii*. Primeiro duas sequências homólogas do TE foram inseridas entre as sequências A e B e entre C e D do cromossomo com o arranjo padrão, gerando duplicações de 7 nucleotídeos (destacados nos retângulos). A recombinação ectópica entre as sequências do TE em orientação oposta resultou na inversão do fragmento cromossômico entre as cópias. Como resultado deste processo, os sítios alvos duplicados foram trocados entre as inserções do TE adquirindo a atual orientação encontrada nos cromossomos com a inversão *2j*. Linhas pontilhadas referem-se às sequências nucleotídicas que não estão representadas em escala (Modificado de Cáceres *et al.*, 1999).

Embora ainda não se conheça com detalhes a estrutura original dessas inserções, elas são homólogas em aproximadamente 274 pb e, mesmo sendo relativamente pequenas, parecem longas suficientes para promover recombinação ectópica em *Drosophila* (Busseau *et al.*, 1989; Nassif & Engels, 1993).

As inserções nos pontos de quebra da inversão *2j* não mostraram homologia nucleotídica com qualquer sequência depositada em bancos de dados e foi designada como *Galileo*. A variação encontrada no número de sinais de hibridização *in situ* e nas suas localizações no cromossomo sugere uma recente atividade transposicional desse *TE*. A inserção *AC* é provavelmente uma cópia defectiva, mas que mantém as sequências terminais necessárias para a transposição. A cópia encontrada entre *B* e *D* pode ter degenerado parcialmente, mas contém regiões de leitura abertas (*ORFs*, do inglês *Open Reading Frames*) que poderiam codificar uma proteína envolvida na transposição. A *ORF* maior prediz uma sequência de 172 aminoácidos e apresenta 32% de identidade com a *ORF* maior do elemento *Gandalf* de *D. koepferae* (Marin & Fontdevila, 1995) espécie críptica de *D. buzzatii* (Cáceres *et al.*, 1999).

Cáceres *et al.* (1999) avaliaram a diversidade nucleotídica entre cinco linhagens com a inversão *2j* e que continham o elemento *Galileo*. Os autores encontraram que a diversidade nucleotídica entre uma sequência de 428 pb, correspondente aos pontos de quebras *A*, *B*, *C* e *D* da inversão *2j*, e 794 pb de *Galileo* era de 0,0053 e virtualmente zero nas regiões da inversão *2j* entre as linhagens. Frente ao exposto e considerando que a divergência nucleotídica (0,0279) entre as linhagens possuidoras do arranjo *standard* (*2st*), é plausível considerar que os arranjos *2j* descendem do produto de um único evento de recombinação ectópica. Cáceres *et al.* (2001) afirmam com base em um alto valor de *bootstrap* que todas as sequências oriundas de *2j* formariam um grupo monofilético e que o arranjo teria ocorrido há aproximadamente 1 milhão de anos. Esse dado é consistente com a difusa distribuição da inversão nas populações de *D. buzzatii* que, de acordo com o tempo de coalescência dos alelos *2j*, teriam se diferenciado há menos de 84 mil anos. Isto representa uma rápida reestruturação genômica nunca vista antes na natureza e qualifica os pontos de quebra *2j* como *hotspots* (pontos quentes) instáveis.

Cáceres *et al.* (2001) realizaram uma exaustiva análise molecular (*Southern blot*, *PCR* de diferentes segmentos, mapa de restrição dos produtos da *PCR* e sequenciamento de DNA) das regiões dos pontos de quebra de *2j* em 9 linhagens de *D. buzzatii* com o

rearranjo ancestral ($2st$) e 30 linhagens com a inversão $2j$. De acordo com a enzima de restrição utilizada, foram observadas diferenças no tamanho das bandas somente nas linhagens $2j$ tanto no ponto de quebra proximal quanto no ponto de quebra distal.

A partir do sequenciamento das regiões e dos fragmentos que apresentaram variação, foram identificados 10 tipos estruturais no ponto de quebra proximal e 7 no ponto de quebra distal, sendo a maioria deles relacionada por uma mudança relativamente simples, como inserções e deleções. Além das inserções, *Galileo-1* e *Galileo-2* implicadas na geração da inversão $2j$ (Cáceres *et al.*, 1999), há mais duas cópias inseridas nos pontos de quebra (*Galileo-3* e *Galileo-4*). As duas últimas cópias são compostas basicamente por longas *IRs* e uma região central relativamente pequena e heterogênea que parece não codificar alguma proteína envolvida na transposição. Cinco outras inserções correspondem aos *TEs Kleper* e *Newton*, intimamente relacionados (com uma média de identidade de 84%), os quais apresentam similaridades com elementos *foldback*. Além disso, sete inserções corresponderam a quatro diferentes *TEs*, os quais têm sido designados como *BuTs* (transposons de *D. buzzatii*) e outras cinco inserções a elementos ainda não classificados (Figura 3).

As inserções, deleções e os outros rearranjos não estão distribuídos uniformemente ao longo dos cromossomos, mas parecem ter ocorrido após as inserções de *Galileo-1* e *Galileo-2*, dentro ou muito próximo deles. Assim, a idéia de que inserções de *Galileo* são *hotspots* mutacionais na geração de rearranjos cromossômicos foi proposta por Cáceres *et al.* (2001).

Junto às múltiplas inserções de *TEs*, foram encontradas por Cáceres *et al.* (2001) 13 deleções com mais de 17 pb sendo que cinco parecem ter sido originadas pelo mecanismo de mal pareamento por deslize (*slipped-strand mispairing*). Um mecanismo similar poderia também ter gerado a duplicação em tandem observada nos 41 pb da extremidade terminal do *Galileo-2* em uma das linhagens estudadas com a inversão $2j$. Em algumas linhagens foi encontrada também uma mudança na orientação de 55 pb no fragmento interno de *Galileo-1*, sugerindo que a inversão teria ocorrido no interior da inserção do ponto de quebra proximal.

Casals *et al.* (2003) caracterizaram molecularmente uma outra inversão polimórfica de *D. buzzatii*, $2q^7$. Enquanto a inversão $2j$ está localizada na posição central

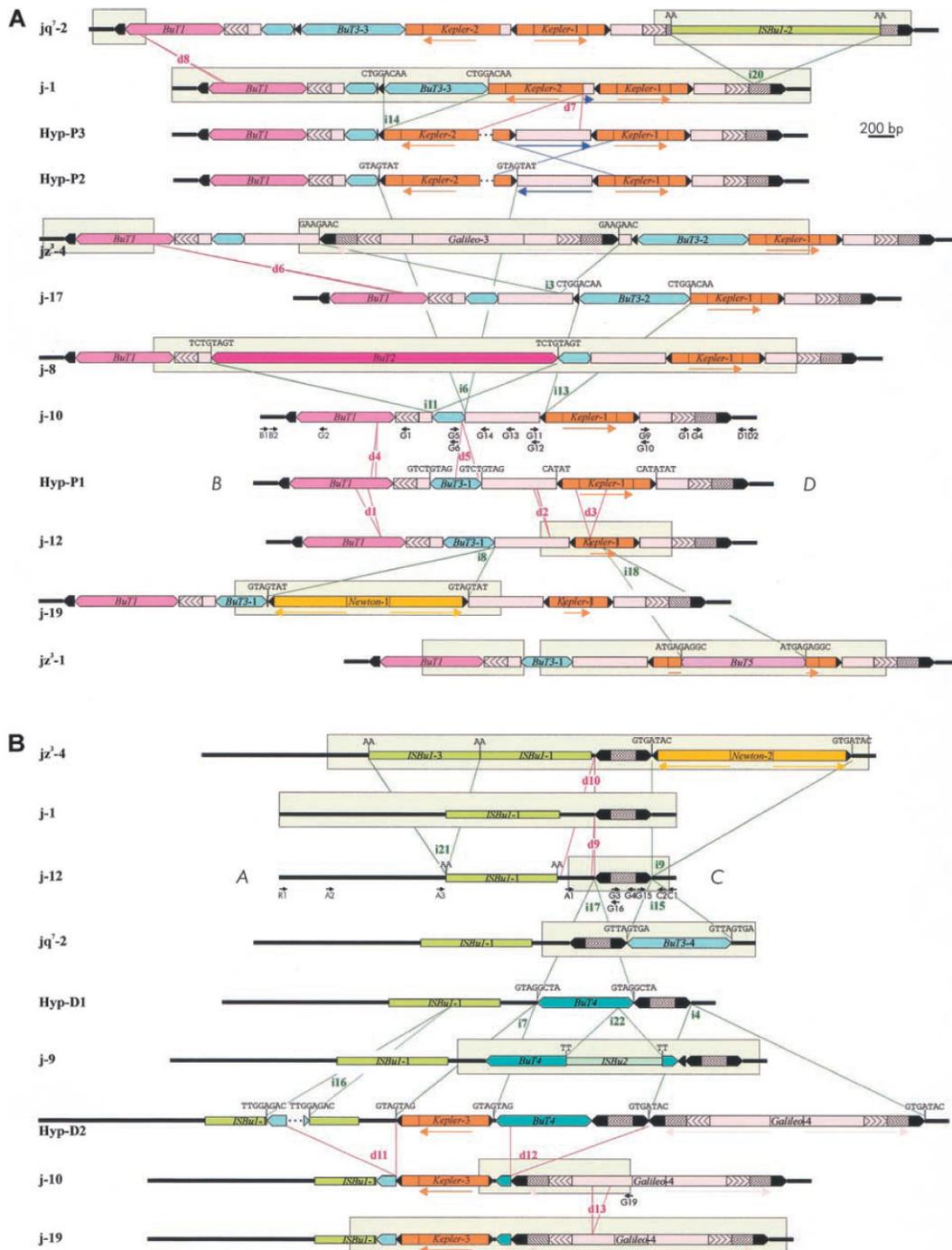


Figura 3. Representação esquemática das estruturas encontradas nos pontos de quebra proximal (A) e distal (B) da inversão $2j$ em 30 linhagens de *D. buzzatii* estudadas. Todas as estruturas diferentes são apresentadas, com exceção da *j-16* localizada no ponto de quebra proximal, que difere da *jq²-4* pela deleção de 6 nucleotídeos. Linhas grossas representam seqüências de cópia única A, B, C e D. TEs estão ilustrados como retângulos coloridos, sendo que as pontas agudas de cor preta estão representando as ITRs. Inserções e deleções estão delimitadas por linhas verdes e rosas, respectivamente e são identificadas com um i ou d

seguido de um número. As sequências dos sítios alvos duplicados são mostradas acima das estruturas. As linhas azuis indicam a inversão de um segmento interno. Setas abaixo dos diagramas informam sobre a orientação de alguns segmentos homólogos. Os fragmentos de cada estrutura que foram sequenciados estão dentro de retângulos claros. Hyp são estruturas hipotéticas não encontradas na amostra, mas representadas como intermediárias daquelas observadas. As pequenas setas pretas são os *primers* utilizados no estudo (Modificada de Cáceres *et al.*, 2001).

do cromossomo, amplamente distribuída em populações de *D. buzzatii* e presente em altas frequências, a inversão $2q^7$ está situada mais proximalmente (com um dos pontos de quebra próximo do centrômero), é mais recente, apresentando uma distribuição geográfica restrita e baixa frequência (< 10%), além de sobrepor a $2j$. Na caracterização fez-se uso de 27 linhagens de *D. buzzatii*, as quais eram homocariótípicas para um dos quatro diferentes arranjos do cromossomo 2: $2st$, $2j$, $2jz^3$ e $2jq^7$.

No trabalho foram encontradas grandes inserções nos pontos de quebra *AC* e *BD* em cromossomos $2jq^7$, as quais apresentaram uma alta identidade com os *TEs* previamente descritos por Cáceres *et al.* (2001). No ponto de quebra distal observou-se uma inserção de 1,8 kb na junção *AC*, que compreende uma cópia truncada do elemento *BuT5-2* e o elemento *Galileo-10*; outra de 387 pb na região *A*, a qual representou a exceção (por não ter sido observada no trabalho anterior), foi designada como sendo um novo elemento da superfamília *hAT*, *BuT6*. No ponto de quebra proximal (*BD*) inserções de 2,4 e 3 kb separadas por apenas 11 pb, os quais correspondem aos primeiros 11 nucleotídeos da sequência *D*, foram identificadas. A inserção de 2,4 kb, localizada na junção *BD*, é constituída por três cópias parciais e uma completa do *TE BuT5*, uma cópia do *Kepler-5* e duas sequências pequenas correspondentes à extremidade terminal do *Galileo-11*; o fragmento de 3 kb continha uma longa cópia do *Galileo-12* inserida dentro de um elemento *BuT3-6*.

A localização em sentido direto das inserções de 1,8 kb e 2,4 kb nas junções *A-C* e *B-D*, respectivamente, aponta para o envolvimento das mesmas na origem da inversão. As outras duas inserções nos pontos de quebra não estão diretamente envolvidas na geração da inversão $2q^7$ e foram aparentemente causadas por invasões secundárias de *TEs*, como mostram os sítios diretos duplicados que flanqueiam *BuT6* na região *A* e *BuT3* na região *D*. Contudo, *Galileo* é apontado como o provável indutor da inversão $2q^7$ com posterior invasão dos demais. A duplicação que flanqueia o *Galileo-10* na região *C* é invertida e

complementar à sequência de 7 pb que flanqueia o *Galileo-11* na região *D*. A sequência de 7 pb da outra extremidade do *Galileo-10* corresponde à sequência alvo consenso (G₁₆T₁₆A₁₀G₁₀T₁₅A₁₈C₉) proposta para *Galileo* e os outros dois elementos *foldback*, *Kepler* e *Newton*, descritos em *D. buzzatii* (Cáceres, 2000). Esta sequência de 7 pb pertence ao *BuT5-2*, o que sugere que *Galileo-10* foi inserido originalmente dentro do elemento *BuT5* pré-existente. Nem a sequência complementar de 7 pb nem o fragmento remanescente de *BuT5-2* foram encontrados na outra extremidade de *Galileo-11* na região *B*, como seria esperado pela recombinação ectópica e pelo modelo de inserção de elemento híbrido (*hybrid element insertion*), ambos modelos propostos para explicar a indução de inversões cromossômicas por *TEs* (Petes & Hill, 1988; Lim & Simmons, 1994). Contudo, devido à região *BD* sustentar um maior número de inserções e mudanças estruturais, a explicação provável é que uma deleção removeu a duplicação de 7 pb junto com o elemento *BuT5-2*. Esta deleção poderia ter também removido os 23 nucleotídeos da sequência *AB* que não estão presentes nos cromossomos com as inversões.

No estudo detalhado realizado Casals *et al.* (2003) uma evidência adicional fornece suporte para o papel do *Galileo* na origem da inversão $2q^7$ e sugere que a recombinação ectópica pode ser o mecanismo envolvido: *Galileo-10* parece ser uma quimera que foi provavelmente gerada por recombinação entre duas cópias de *Galileo* ligeiramente diferentes. Foram comparadas as *IRs* de ambas as extremidades do *Galileo-10* com as sequências correspondentes no *Galileo-12* e com outras duas cópias previamente descritas, *Galileo-3* e *Galileo-4* (Cáceres *et al.*, 2001). As *IRs* de um mesmo elemento sempre exibiram uma menor divergência que aquelas de elementos diferentes. A exceção é o *Galileo-10*, cujas *IRs* mostram um aumento na divergência, similar com aquela observada entre *IRs* de diferentes elementos e seis vezes maior do que entre *IRs* pertencentes a um mesmo elemento. Embora as diferenças entre as duas *IRs* do *Galileo-10* pudessem ter surgido por mutação, a taxa de mutação deveria ser excessivamente alta para dar conta do alto nível de divergência. Dessa forma suporta-se o modelo de recombinação ectópica para a origem do elemento *Galileo-10*.

Casals *et al.* (2003) estimaram que a inversão $2q^7$ ocorreu há 490 mil anos sendo que, ao excluírem a região *A* (a qual apresentou o maior nível de polimorfismo), a estimativa passou para 670 mil anos. Da mesma forma, para as diferenças estruturais entre as linhagens que possuíam o cromossomo $2jq^7$ os autores estimaram entre 20 mil e 25 mil

anos a idade dos alelos. As estimativas são consistentes com aquelas calculadas por Cáceres *et al.* (2001) para inversão $2j$ e suas formas alélicas (aproximadamente 1 milhão de anos e 83 mil anos respectivamente), da qual a inversão $2q^7$ se formou. A ausência de polimorfismo nucleotídico e a pouca idade dos alelos de $2j$, ratificam a baixa frequência da inversão $2q^7$ em populações naturais de *D. buzzatii* (Hasson *et al.*, 1995).

Sequências que formam palíndromos, semelhantes às *ITRs*, tem sido demonstradas como a fonte para a instabilidade genômica em camundongos (Zhou *et al.*, 2001), causando diferentes tipos de rearranjos cromossômicos. O número de rearranjos aumenta com o comprimento das *ITRs* e diminui com o maior tamanho da sequência interna, e ambos estão provavelmente envolvidos na habilidade para formar estruturas secundárias e ocasionar quebras na dupla fita de DNA. O tamanho das *IRs* de cópias completas do elemento *Galileo* variam de 683 pb e 1229 pb (Cáceres *et al.*, 2001; Casals *et al.*, 2003; Marzo *et al.*, 2008) e, juntamente com o comprimento do DNA espaçador (domínio central), são premissas para o aumento no número de rearranjos, o que justificaria a denominação desses como *hotspots* mutacionais.

Além disso, Cáceres *et al.* (2001) e Casals *et al.* (2005) argumentam sobre a capacidade do *Galileo* em formar estruturas secundárias o que poderia explicar as dificuldades encontradas na amplificação deste *TE* por *PCR* e, principalmente, ser sugestivo de um elemento ativo com um importante papel na evolução genômica de *D. buzzatii* e espécies relacionadas.

Casals *et al.* (2005) ao caracterizarem molecularmente os *TEs Galileo*, *Kepler* e *Newton*, analisaram 14 cópias de *Galileo*, as quais apresentaram uma variação em tamanho de 20 a 2304 pb com regiões invertidas (*IRs*) de 8 a 1115 pb (Figura 4). A região flanqueadora das *IRs* de *Galileo* (*IR-FD*) contém 479 pb, os quais são relativamente conservados em todas as cópias. O domínio externo (*IR-OD*) é composto por três repetições imperfeitas em tandem de 136 pb, com uma identidade média de 91%, mais uma repetição menor incompleta de 43 pb. Tipicamente há duas a três repetições por *IR* e esse número difere entre ambos os lados do elemento. O domínio interno (*IR-ID*) além de ser extremamente heterogêneo em comprimento e na composição da sequência, é caracterizado por um alto conteúdo de AT (mais de 75%). Em *Galileo-3* e *-6*, essa região inclui uma região de leitura aberta (*ORF*) de 141 pb e 34% de identidade com a transposase (*TPase*) de *Hoppel*, um elemento *P-like* de *D. melanogaster* (Reiss *et al.*,

2003). Muitos dos elementos contêm uma região central com 81 a 158 pb que não tem sido encontrada formando parte das *IRs* e pode ser considerada como o domínio *M*.

Kepler e *Newton* apresentam uma identidade nucleotídica de 90% e diferem de *Galileo* por aparentemente conter apenas *IR-FD* e o domínio *M*, mas não as repetições em tandem (*IR-OD*) ou domínios *IR-ID*. Em média, os primeiros 600 pb de *Galileo* (*IR-FD* e parte da primeira repetição em tandem), *Kepler* (*IR-FD* e a maior parte da região *M*), e *Newton* (*IR-FD*) têm aproximadamente 75% de identidade nucleotídica (Cáceres *et al.*, 2001). Em particular, os 40 pb terminais são quase idênticos entre eles e a região interna de 200 pb mostra aproximadamente 86% de identidade nucleotídica. Outra característica comum entre *Galileo*, *Kepler* e *Newton* é a geração da duplicação de 7 pb do sítio alvo durante a inserção. As comparações das 19 sequências flanqueadoras desses elementos corroboram com a proposta feita por Cáceres (2000) da preferência por um sítio de inserção que contenha a sequência consenso (G₁₆T₁₆a₁₀g₁₀T₁₅A₁₈c₉). Curiosamente, essa sequência é palindrômica e as primeiras duas bases de cada lado são complementares às extremidades dos elementos (CA...TG).

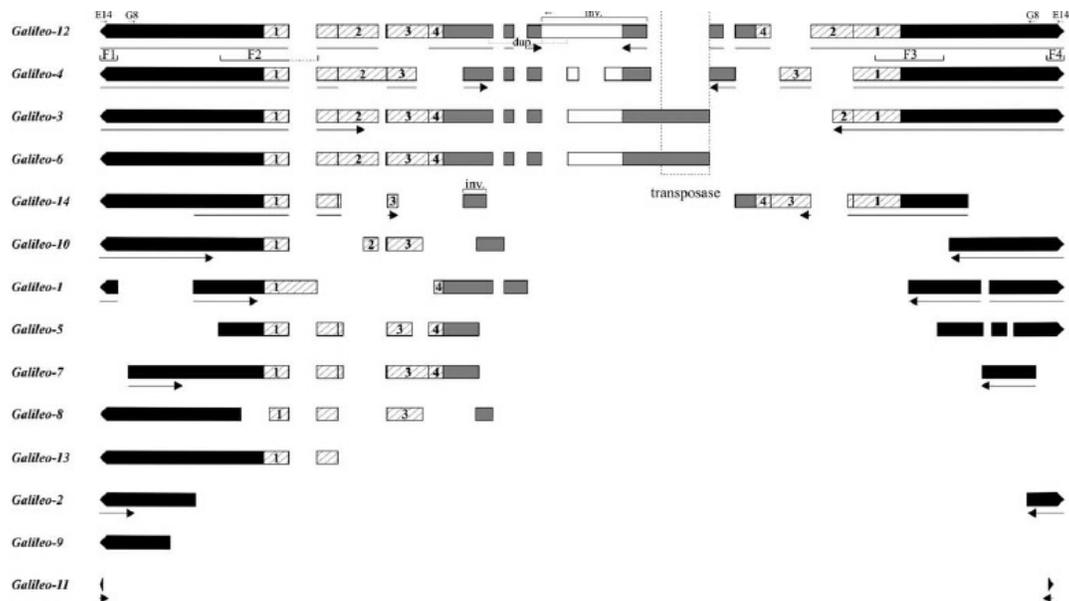


Figura 4. Variantes de *Galileo* caracterizadas em *D. buzzatii* por Casals *et al.* (2005).

Para determinar a relação filogenética entre *Galileo*, *Kepler* e *Newton*, Casals *et al.* (2005) utilizaram duas sequências homólogas das *IRs* (totalizando aproximadamente

250 pb). Apenas os elementos que continham ambas as regiões completas foram incluídos na análise e as duas extremidades dos elementos foram consideradas independentemente. A árvore obtida mostrou dois clados altamente divergentes, um incluindo cópias do *Galileo* e o outro as cópias de *Kepler* e *Newton*. Cópias de *Galileo* apresentaram um alto grau de variação e formaram diferentes subgrupos. Cópias de *Kepler* e *Newton* formaram dois grupos monofiléticos separados, mas estritamente relacionados.

De acordo com Rebatchouk & Narita (1997) e Simmen & Bird (2000), elementos *foldback* podem ser classificados em quatro tipos dependendo da presença de diferentes domínios nas *IRs*. *Galileo* se enquadra no quarto tipo, pois na maioria dos casos ele apresenta os domínios *IR-FD*, *IR-OD* e *IR-ID* (Figura 5). É notável que as repetições internas em tandem do domínio *IR-OD* de *Galileo* são maiores do que as observadas em outros elementos *foldback*, as quais possuem normalmente 7 a 32 pb de comprimento. *Kepler* e *Newton* por apresentarem apenas o domínio *IR-FD* e a região *M* podem ser classificados como um novo tipo de elementos *foldback*, embora seja possível que as cópias analisadas por Casals *et al.* (2005) fossem defectivas (Figura 6).



Figura 5. Representação esquemática dos domínios da maioria das variantes de *Galileo* em *D. buzzatii*. Todas as variantes observadas por Casals *et al.* (2005) apresentaram suas *IRs* na orientação invertida.



Figura 6. Representação esquemática dos domínios observados por Casals *et al.* (2005) nas variantes de *Kepler* e *Newton* em *D. buzzatii*.

As cópias dos três elementos *foldback* caracterizadas por Casals *et al.* (2005) em *D. buzzatii* não mostraram alguma evidência de capacidade de codificação ou *ORFs* com homologia para uma proteína conhecida, com a exceção de duas cópias de *Galileo*, as quais possuem um segmento com baixa similaridade com a transposase de *Hoppel*. O alto nível de heterogeneidade estrutural descrito nesses e outros elementos *foldback* (Hoffman-

Liebermann *et al.*, 1985; Cheng *et al.*, 2000; Cáceres *et al.*, 2001; Casals *et al.*, 2005) sugere que todos os elementos encontrados poderiam ser defectivos e terem perdido a sequência necessária para codificar a transposase. Alternativamente, esses elementos poderiam ser mobilizados através de proteínas produzidas por outros *TEs* ou processos celulares (Rebatchouk & Narita, 1997). Em *D. buzzatii*, as similaridades entre os sítios de inserção alvos dos três elementos *foldback* sugerem que provavelmente eles são mobilizados por um mesmo mecanismo de transposição. Em adição, a conservação de certas regiões das *IRs*, especialmente as extremidades, poderia ser um indicativo do possível papel de reconhecimento e ligação da transposase. Parece que as repetições internas em tandem poderiam ter um importante papel no reconhecimento e na ligação de proteínas. Essas sequências poderiam ser os alvos da transposase que após ligar-se, se deslocaria para o término do elemento. Também, segundo Cheng *et al.* (2000) a maior probabilidade de mobilização de um elemento estaria correlacionada com o aumento do número de repetições em tandem.

A partir da estimativa da idade de *Galileo*, *Kepler* e *Newton* no genoma de *D. buzzatii* realizada por Casals *et al.* (2005), verificou-se um padrão de distribuição interespecífica desses elementos, o qual sugere que estariam presentes antes da radiação do complexo *buzzatii*, sendo transmitidos verticalmente durante subsequentes eventos de especiação. Nenhuma evidência de transmissão horizontal foi encontrada nesse trabalho e em outros estudos da distribuição de elementos *foldback*, como os de *D. melanogaster* (Silber *et al.*, 1989) ou os encontrados nos gêneros *Solanum* e *Lycopersicon* (Rebatchouk & Narita, 1997). Esses resultados diferem consideravelmente dos observados em vários transposons da classe II, os quais frequentemente transmitidos horizontalmente entre espécies distantes (Capy *et al.*, 1998; Loreto *et al.*, 2008).

Com relação à distribuição cromossômica desses elementos Casals *et al.* (2005) ainda observaram que a mesma não é randômica e que há uma tendência clara do acúmulo em regiões proximais dos braços cromossômicos e em cromossomos pontuais. Dois fatores principais têm sido propostos na distribuição de *TEs* no genoma: taxa de recombinação e densidade gênica. Espera-se que a taxa de recombinação afete inversamente a abundância de *TE* de duas maneiras diferentes. Primeiro, a redução da taxa de recombinação em certas regiões irá dificultar a eliminação de inserções deletérias por seleção natural (Hill & Robertson, 1966; Gordo & Charlesworth, 2001). Segundo, *TEs* inseridos em regiões de

baixa recombinação provavelmente evitarão a geração de rearranjos cromossômicos deletérios por recombinação ectópica (Langley *et al.*, 1988; Montgomery *et al.*, 1991; Goldman & Lichten, 1996). Por outro lado, *TEs* poderiam se acumular em regiões de baixa densidade gênica onde suas inserções causariam menores efeitos deletérios. Assim, o acúmulo de elementos *foldback*, em *D. buzzatii*, próximo de regiões centroméricas e em cromossomos pontuais é consistente com o efeito esperado da taxa de recombinação. Contudo, o acúmulo de *TEs* em cromossomos pontuais, cuja densidade gênica é similar a de outras regiões com elevada recombinação, contradiz esta hipótese. Curiosamente, em ambas *Drosophila melanogaster* (Bartolomé *et al.*, 2002; Kaminker *et al.*, 2002) e *D. buzzatii*, *TEs* não se acumulam em regiões teloméricas, onde a taxa de recombinação é também baixa (Charlesworth, 1996). Uma possível explicação é que a recombinação ectópica não é reduzida nessas regiões (Bartolomé *et al.*, 2002), como tem sido encontrado em regiões subteloméricas de leveduras (Haber *et al.*, 1991). Além disso, regiões de baixa recombinação próximas ao telômero são geralmente menores do que as de regiões centroméricas, e a presença de inserções de *TE* em regiões de baixa recombinação parece estar relacionada com a distância da região de elevada recombinação (Maside *et al.*, 2001).

Marzo *et al.* (2008) através de análise *in silico* dos doze genomas de *Drosophila* (*Drosophila* 12 Genomes Consortium, 2007) encontraram cópias do elemento *Galileo* em 6 espécies: *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. virilis* e *D. mojavensis*. Verificaram que esse transposon estava presente em abundância com mais de 100 cópias em todos os casos, sendo que o máximo de 495 foi observado em *D. willistoni*.

A partir de inúmeras *PCRs*, Marzo *et al.* (2008) isolaram a maior cópia de *Galileo* em *D. buzzatii*, a qual apresenta 5406 pb, *IRs* de 1229 pb e um *intronless* de 2738 pb, com a *ORF* posicionada entre os nucleotídeos 1348 e 4087, responsável por uma proteína de 912 aa. Utilizando-se o *BLASTX*, observaram uma similaridade significativa da *TPase* do *Galileo* de *D. buzzatii* com aquelas relacionadas com o elemento *1360* (Hoppel) e elementos *P* de *D. melanogaster* (Laski *et al.*, 1986; Reiss *et al.*, 2003). *Galileo* de *D. buzzatii* possui um domínio *THAP* próximo da extremidade N (27-104 aa) similar ao domínio que se liga ao DNA da *TPase* do elemento *P* (Lee *et al.*, 1998; Roussigne *et al.*, 2003; Clouaire *et al.*, 2005; Quesneville *et al.*, 2005). Uma cópia do *TE 1360*, localizada no cromossomo 4 de *D. melanogaster* (Kapitonov & Jurka, 2003), codifica uma *TPase* de 854 aa e também possui um domínio *THAP* próximo da extremidade N. O alinhamento

global da TPase de *Galileo* (*D. buzzatii*) com as dos elementos *I360* e *P* (*D. melanogaster*) demonstrou identidades de 34,5% e 27,6%, respectivamente. Contudo os autores não encontraram similaridade significativa entre a TPase de *Galileo* (*D. buzzatii*) e proteínas codificadas por elementos *foldback* já descritos em *D. melanogaster* (Templeton & Potter, 1989; Harden & Ashburner, 1990).

Ao caracterizarem as cópias de *Galileo* relativamente longas que continham segmentos da TPase nos 6 genomas, Marzo *et al.* (2008) observaram que eram dotadas de uma ou duas *IRs* longas com similaridade àquela encontrada em *Galileo* de *D. buzzatii*. Com relação à sequência de 7 pb, oriunda da duplicação do sítio alvo, foi calculada a frequência dos nucleotídeos em cada uma das posições para cada espécie separadamente e a sequência consenso resultante foi o palíndromo GTANTAC.

Com exceção de *D. pseudoobscura* e *D. persimilis*, as longas *IRs* apresentam pouca similaridade entre as diferentes espécies tanto em comprimento quanto em composição. A conservação parece ser restrita à extremidade: 17 dos 40 pb terminais são conservados entre treze sequências, entre as quais estavam as cópias de *Galileo* de *D. buzzatii* e de outras 6 espécies do gênero *Drosophila*, além das sequências de *Kepler* e *Newton* de *D. buzzatii*. A partir da sequência consenso da porção terminal de *Galileo* e do elemento *I360*, os autores (Marzo *et al.*, 2008) verificaram que 15 dos 31 pb (que compõem as *ITRs* de *I360*) são idênticos entre os transposons, uma evidência da relação evolutiva entre esses *TEs*. Além disso, a sequência consenso de *Galileo* compartilha 17 pb dos 31 pb das *ITRs* de *DmeNP*, e os três elementos compartilham terminações idênticas: 5'-CA...TG-3'. *Galileo* e *I360* geram a duplicação de 7 pb do sítio alvo de inserção, enquanto o elemento *P* causa a duplicação de 8 pb, os quais também correspondem a um palíndromo (GTCCGGAC), característica essa relacionada com a dimerização da sua transposase. Isso sugere que a transposase funcional do *Galileo* também é um dímero e que esse elemento é membro da superfamília *P* de transposons que utilizam o mecanismo de “corta-e-cola” nas suas mobilizações.

Em *Drosophila willistoni*, além das cópias pequenas não-autônomas, a cópia mais completa do elemento *Galileo* apresentou 4386 pb, embora nenhuma transposase funcional foi identificada, assim como ocorre nas demais espécies (todas carregam códons de parada, alterações no quadro de leitura, e/ou deleções).

A interpretação da árvore filogenética elaborada por de Marzo *et al.* (2008), com a sequência consenso das transposases de *Galileo* e dos elementos *P* e *I360*, sugere que *Galileo* surgiu a partir de um ancestral com *IRs* muito mais curtas. As *IRs* longas de *Galileo* são de tamanho variável dentro e entre as espécies, sugerindo um notável dinamismo estrutural. Por exemplo, entre as cópias mais completas em *D. willistoni* foram encontradas *IRs* com 765 e 959 pb. *ITRs* podem encurtar por deleção, mas é muito provável que possam ser alongadas pela duplicação interna, recombinação desigual e/ou outros mecanismos, tais como conversão gênica (Richardson *et al.*, 1998) ou reparo de quebras de fita simples (Kapitonov & Jurka, 2006). Os autores sugerem que diferentes elementos *foldback* poderiam ter sido originados a partir de linhagens de transposons independentes, que o comprimento e a estrutura de *IRs* não são critérios confiáveis para a classificação de um *TE*, e ainda que elementos *foldback* não constituam um grupo monofilético.

A filogenia de *Galileo* entre as 7 espécies do gênero *Drosophila* é incongruente com as reconstruções filogenéticas das mesmas (*Drosophila* 12 Genomes Consortium, 2007). Os elementos de *D. willistoni* e *D. virilis*, pertencentes a diferentes subgêneros (*Sophophora* e *Drosophila*, respectivamente), são intimamente relacionados. Da mesma forma, o *Galileo* de *D. mojavensis* e *D. buzzatii* (subgênero *Drosophila*) está mais estreitamente relacionado com o de *D. ananassae*, *D. pseudoobscura*, e *D. persimilis* (subgênero *Sophophora*) do que o de *D. virilis*, uma espécie do mesmo subgênero. Há duas explicações plausíveis para estas disparidades topológicas: *lineage sorting* e transferência horizontal (Page & Charleston, 1998). A primeira refere-se à diversificação vertical das linhagens de *TEs* e suas perdas diferenciais ao longo dos ramos da árvore filogenética. Transferência horizontal é o processo de invasão de um novo genoma por um *TE*, que é comum nos transposons e é considerado como uma fase integrante do ciclo de vida do transposon, que permite a sobrevivência em longo prazo (Hartl *et al.*, 1997; Silva *et al.*, 2004, Loreto *et al.*, 2008). Eventos de transferência horizontal têm ocorrido durante a evolução dos elementos *P*, por exemplo, no gênero *Drosophila* (Silva & Kidwell, 2000; Loreto *et al.*, 2008). No entanto, até o momento não há evidências para transferência horizontal de *Galileo* e, portanto, segundo Marzo *et al.* (2008) *lineage sorting* deve ser a provável explicação.

Delprat *et al.* (2009) com o intuito de aumentar a compreensão dos mecanismos e verificar a implicação de *TEs* na geração de inversões em *Drosophila*, isolaram e caracterizaram os pontos de quebra de outra inversão polimórfica de *D. buzzatii*, $2z^3$. Essa inversão surgiu em um cromossomo que já continha a inversão $2j$, dando origem ao arranjo $2jz^3$, o qual abrange cerca de um terço do cromossomo 2 (~ 11Mb) e se sobrepõe ao segmento $2j$ de forma que as duas inversões não podem ser separadas por recombinação (Ruiz & Wasserman, 1993). O arranjo $2jz^3$ tem uma ampla distribuição geográfica, estando presente em populações naturais da Argentina, Brasil, Chile e do Antigo Mundo (Ruiz *et al.* 1984; Hasson *et al.*, 1995). Em populações argentinas a sua frequência relativa varia de 0,5 a 31,5%, com uma média de 8% (Hasson *et al.*, 1995).

A partir do alto grau de similaridade entre *Galileo*, *Newton* e *Kepler* e devido ao fato de todos produzirem a duplicação de 7 pb do sítio alvo de inserção (com a mesma sequência consenso), Delprat *et al.* (2009) passaram a considerá-los como diferentes subfamílias de *Galileo* no genoma de *D. buzzatii* e os denominam como *GalileoG*, *GalileoN* e *GalileoK*, respectivamente.

TEs foram encontrados nas regiões dos pontos de quebra da inversão $2z^3$, contudo apenas *GalileoN* estava presente nas duas junções do ponto de quebra. Dois elementos, *BuT5* e *LINE-like*, foram encontrados inseridos numa cópia de *GalileoN* no ponto de quebra distal e outras duas inserções, *But3* e *But4* dentro da outra cópia de *GalileoN* no ponto de quebra proximal. Segundo Delprat *et al.* (2009) estes quatro *TEs* estão presentes em um único ponto de quebra e cada um deles é flanqueado por *TSDs* diretos e idênticos e, sendo assim, eles não seriam responsáveis pela geração da inversão e são mais bem interpretados como colonizadores secundários do ponto de quebra.

Os dois exemplares de *GalileoN*, *GalileoN-4* e *GalileoN-5*, encontrados na inversão $2z^3$, têm tamanhos e estrutura semelhantes (1541 pb e 1533 pb, respectivamente), com *IRs* relativamente longas, a região central é orientada em posição aposta nas duas cópias e apresentam uma alta similaridade com as outras duas variantes previamente descritas (*GalileoN-1* e *GalileoN-2*) por Cáceres *et al.* (2001) e Casals *et al.* (2005). Assim como no trabalho de Cáceres *et al.* (1999) os 7 pb duplicados, gerados no evento de transposição de *Galileo*, estão trocados entre as cópias de *GalileoN* nos pontos de quebra. Estas observações são consistentes com a geração da inversão $2z^3$ por recombinação ectópica entre cópias de *GalileoN*, mas não exclui o modelo de inserção de elemento híbrido.

Outra evidência, no mesmo trabalho de Delprat *et al.* (2009), foi revelada pela comparação dos nucleotídeos das *IRs* dentro e entre as cópias de *GalileoN*, onde observou-se que as extremidades de cada variante de *GalileoN-4* e *GalileoN-5* apresentam uma divergência de 6%, mas são 99% idênticas entre cópias, sendo dessa forma, elementos quiméricos. Uma análise mais aprofundada revelou um padrão impressionante de organização estrutural e conduziu à mesma conclusão. Em 33 sítios variáveis (posição 1 a 824) os nucleotídeos de *GalileoN-4* são idênticos ao de *GalileoN-1*, sendo que o mesmo é observado entre as cópias *GalileoN-5* e *GalileoN-2*. A situação é completamente invertida para os 20 sítios variáveis, da posição 966 até o final da sequência do elemento, onde os nucleotídeos de *GalileoN-4* são idênticos ao de *GalileoN-2* e a sequência nucleotídica de *GalileoN-5* é idêntica a de *GalileoN-1*. Análises filogenéticas das quatro cópias, realizadas separadamente para as duas porções do elemento e o método do qui-quadrado máximo, corroboraram com as estruturas quiméricas de *GalileoN-4* e *GalileoN-5*.

Através destas evidências Delprat *et al.* (2009) sustentam que o evento de recombinação que originou a inversão $2z^3$ ocorreu dentro de 141 pb da região central entre as posições 825 e 965 de *GalileoN*. Os autores propõem que as longas *IRs* do *Galileo* induzem a formação de estruturas secundárias e quebras de dupla fita de DNA (*DSBs*, do inglês *Double Stranded Breaks*) em alta taxa e isso contribui para a sua capacidade única de gerar inversões cromossômicas. Complementam ainda sugerindo que, pelo fato do evento de recombinação ectópica ter ocorrido entre a região central de duas variantes de *Galileo*, ele parece consistente com a clivagem de nuclease na alça (*loop*) da estrutura.

A origem das numerosas inversões cromossômicas em *Drosophila* e outros dípteros é ainda uma questão em aberto e poucas espécies têm sido investigadas a este respeito. Embora não tenha sido encontrada por Marzo *et al.* (2008) qualquer TPase funcional na análise *in silico*, apenas um genoma de cada espécie foi sequenciado e assim averiguado, sendo necessário o estudo de *Galileo* em outros genomas de populações naturais de *Drosophila*. No referido trabalho os autores estimaram uma recente atividade transposicional de *Galileo* (< 1 Ma) para as quatro espécies mais polimórficas (*D. ananassae*, *D. pseudoobscura*, *D. persimilis* e *D. willistoni*) e para *D. virilis*, cromossomicamente monomórfica, a atividade mais antiga deste elemento de transposição (aproximadamente 3 Ma). A detecção de cópias quiméricas que podem ser o resultado de rearranjos cromossômicos (Casals *et al.*, 2003) indica que de fato *Galileo* poderia estar

envolvido na origem das inversões, pelo menos em algumas outras espécies além *D. buzzatii*.

Grupo *willistoni*

No gênero *Drosophila*, o grupo *willistoni* pertence ao subgênero *Sophophora*, o qual foi estabelecido por Sturtevant (1939). Inclui, segundo Bächli (2008) 23 espécies, agrupadas em três subgrupos: *alagitans* (5 espécies), *bocainensis* (12 espécies), *willistoni* (6 espécies).

O subgrupo *willistoni* é composto pela *D. willistoni* Sturtevant (1916) e as suas cinco espécies crípticas: *D. paulistorum* Dobzhansky & Pavan (Burla *et al.*, 1949), *D. pavlovskiana* Kastritsis & Dobzhansky (1967), *D. tropicalis* Burla & da Cunha (Burla *et al.*, 1949), *D. insularis* Dobzhansky (Dobzhansky *et al.*, 1957) e *D. equinoxialis* Dobzhansky (1946).

Quanto à sua distribuição geográfica, o grupo *willistoni* é essencialmente Neotropical, incluindo espécies com grande distribuição ou mesmo espécies extremamente restritas, endêmicas (Figuras 7 e 8). De acordo com Garcia *et al.* (2006) o subgrupo *willistoni* constitui um dos mais notáveis e abundantes grupos de drosofilídeos nesta região.

D. willistoni apresenta a mais ampla distribuição neotropical e constitui-se na espécie dominante na maioria das suas florestas quentes e úmidas, apesar de apresentar elevadas flutuações sazonais (Hoenigsberg *et al.*, 1977; Sene *et al.*, 1980; Valente *et al.*, 1981; Bizzo & Sene, 1982 e Martins, 1987), existindo em altas densidades desde a Flórida até o sul do Brasil e Argentina, e desde o Oceano Pacífico até o Atlântico (Figura 7) (Spassky *et al.*, 1971 e Ehrman & Powell, 1982). Segundo Ayala & Tracey (1973) *D. willistoni* diferencia-se nas subespécies *willistoni* e *quechua* (restrita ao Peru e Equador) entre as quais o cruzamento não ocorre facilmente e apenas machos híbridos estéreis são produzidos (Dobzhansky, 1965).

A distribuição das demais espécies do subgrupo críptico encontra-se, em geral, dentro da extensão ocupada por *D. willistoni*, sendo que esta espécie, juntamente com *D. paulistorum*, *D. equinoxialis* e *D. tropicalis* são simpátricas desde a Guatemala, através da América Central e do norte da América do Sul, até a região central do Peru e do Brasil (Ayala *et al.*, 1970; Spassky *et al.*, 1971). Segundo Spassky *et al.* (1971), ao sul do estado

de São Paulo, no Brasil, apenas *D. paulistorum* e *D. willistoni* são encontradas. *D. insularis* apresenta uma ocorrência restrita e endêmica em algumas ilhas das Pequenas Antilhas e *D. pavlovskiana*, espécie rara estreitamente relacionada à *D. paulistorum*, ocorre apenas nas Guianas (Figura 7) (Ehrman & Powell, 1982).

D. paulistorum é considerada uma superespécie, constituída por seis semiespécies: Andino-Brasileira, Amazônica, Centro-Americana, Interior, Transicional e Orinocana (Dobzhansky & Spassky, 1959). Distribui-se desde a Guatemala e Trinidad até o sul do Brasil, sendo que as regiões da Guatemala, Honduras, El Salvador, Costa Rica e oeste do Panamá a ocupação é restrita pela semiespécie Centro-Americana; o nordeste, centro e sul do Brasil, juntamente com o Equador, Peru e a Bolívia, apresentam apenas a semiespécie Andino-Brasileira (Spassky *et al.*, 1971). Na região mais ao sul do Panamá e no norte da América do Sul registra-se a ocorrência, isolada ou em simpatria, das outras quatro semiespécies, juntamente com a Andino-Brasileira, que ocupa a maior área de distribuição: a semiespécie Transicional ocorre apenas no norte da Colômbia e Venezuela; a Amazônica estende-se desde o sul do Panamá e Trinidad, até os estuários dos rios Amazonas e Tocantins; Orinocana foi coletada na região sul do Panamá, centro leste e norte da Colômbia, na Venezuela, em Trinidad e nas Guianas; e Interior habita regiões mais ao sul da Colômbia, da Venezuela e ao norte do Brasil (Figura 8) (Spassky *et al.*, 1971; Winge, 1971; Dobzhansky & Powell, 1975; Ehrman & Powell, 1982; Santos & Valente, 1990). Estas semiespécies exibem entre si um isolamento sexual etológico pronunciado, embora não absoluto, sendo que quando cruzamentos foram efetivados em condições laboratoriais, fêmeas férteis e machos estéreis foram produzidos (Ehrman & Powell, 1982). A semiespécie Transicional, entretanto, parece ocupar um *status* intermediário, pois é capaz de produzir híbridos férteis e viáveis, ao menos quando inter-cruzada com representantes das semiespécies Andino-Brasileira e Centro-Americana (Dobzhansky *et al.*, 1964; Ehrman & Powell, 1982).

Drosophila tropicalis subdivide-se nas subespécies *tropicalis* e *cubana*. Há um pequeno isolamento etológico entre elas, mas machos híbridos são estéreis (Townsend, 1954). Winge (1965, 1971) argumenta que as duas subespécies deveriam ser tratadas como espécies separadas, a partir da elevação de *D. tropicalis cubana* para o nível de espécie, contudo Rohde *et al.* (2006) não detectaram diferenças nos arranjos cromossômicos.



Figura 7. Mapa da distribuição geográfica do subgrupo *willistoni* (Robe *et al.*, 2010).



Figura 8. Mapa da distribuição geográfica das semiespécies de *D. paulistorum* (Robe *et al.*, 2010)

Populações de *D. equinoxialis* oriundas da América do Sul ou do Panamá quando cruzadas com populações de Porto Rico e Costa Rica, produzem prole com fêmeas férteis e machos estéreis. Mediante esta observação, Ayala (1973) subdividiu a espécie em *D. equinoxialis equinoxialis* e *D. equinoxialis caribbensis*, a qual se distribui da Costa Rica através da América Central e sul do México.

Ainda compondo o grupo *willistoni*, a linhagem *Carmody* (Carmody, 1965) possui *status* taxonômico indefinido por ser capaz de inter cruzar com algumas linhagens de *D. equinoxialis* e de *D. paulistorum*.

Embora tenham morfologia externa muito similar, as espécies do grupo críptico podem ser identificadas pela forma da genitália dos machos (Burla *et al.*, 1949; Malagolowkin, 1952; Spassky, 1957; Winge & Cordeiro, 1963) e pela forma das espermatecas das fêmeas (Stutervant, 1921; Patterson & Mainland, 1944; Spieth, 1949). Além da análise destas estruturas morfológicas, as espécies crípticas podem ser distinguidas pela caracterização espécie-específica do intervalo interpulso (*IPI*, do inglês *Interpulse interval*) do som produzido pelos machos durante a corte sexual (Ritchie & Gleason, 1995), por meio do estudo detalhado dos cromossomos politênicos (Burla *et al.*, 1949; Rohde *et al.*, 2006) e também pela variação da mobilidade eletroforética da enzima fosfatase ácida-1 (*Acp1*) (Garcia *et al.*, 2006).

Poucos trabalhos tiveram como objetivo a elucidação das relações filogenéticas entre as espécies do grupo *willistoni*. Spassky *et al.* (1971) desenvolveram um diagrama esquemático das relações evolutivas do subgrupo *willistoni*, com base em evidências biogeográficas, genéticas, citogenéticas e bioquímicas. Neste diagrama são apresentadas como espécies-irmãs *D. equinoxialis*, *D. pavlovskiana* e *D. paulistorum*, assim como *D. insularis* e *D. tropicalis*, enquanto *D. willistoni* não aparece agrupada a nenhuma outra espécie. Dentre as semiespécies de *D. paulistorum*, três linhagens principais podem ser distinguidas ainda neste trabalho: a primeira é representada apenas por Amazônica, a mais divergente; a segunda une Interior e Orinocana; e a terceira agrupa as semiespécies Transicional, Centro-Americana e Andino-Brasileira.

Um dendrograma das espécies crípticas foi apresentado no trabalho de Ayala *et al.* (1974), a partir do estudo da diferenciação genética de 36 *loci* para alozimas. *D. insularis* apresentou-se como táxon-irmão dos dois clados formados pelas espécies *D. tropicalis/D. willistoni* e *D. equinoxialis/D. pavlovskiana/D. paulistorum*. Com relação às semiespécies

de *D. paulistorum*, Interior e Orinocana apresentam-se estreitamente relacionadas; Centro-Americana, Transicional e Andino-Brasileira formam um agrupamento, enquanto Amazônica se posiciona como um táxon-irmão deste último clado.

Gleason *et al.* (1998) publicaram o primeiro trabalho de filogenia do subgrupo críptico realizado com base na análise das sequências de três genes: dois nucleares, *period* (*per*) e álcool desidrogenase (*Adh*) e de um gene mitocondrial, *citocromo oxidase I* (*COI*). A topologia melhor suportada por estes dados apresenta *D. insularis* como ramificação basal do subgrupo *willistoni*, seguida por *D. tropicalis*, de forma que *D. willistoni* aparece como espécie-irmã do clado que agrupa *D. equinoxialis*, *D. paulistorum*. Neste trabalho obtêm-se algumas incongruências com a taxonomia tradicional: a linhagem *Carmody* forma o clado composto por diferentes populações de *D. equinoxialis*, dessa forma, sob a perspectiva filogenética, deveria ser classificada como *D. equinoxialis* e *D. pavlovskiana* agrupa-se dentro do *cluster* formado pelas diferentes semiespécies de *D. paulistorum*, intimamente relacionada à semiespécie Centro-Americana, não corroborando com sua identidade biológica como espécie distinta. Quanto às relações entre as semiespécies de *D. paulistorum*, Gleason *et al.* (1998) não consideram muito confiáveis os resultados devido ao fato de terem analisado apenas uma linhagem de cada semiespécie, mas a análise do gene *per*, favoreceu, com forte suporte, a hipótese de que Amazônica foi a primeira semiespécie a divergir, seguida por dois cladogramas, um que agrupa Andino-Brasileira e Orinocana e o segundo que inclui as semiespécies Transicional/Interior e Centro-Americana, além de *D. pavlovskiana* como derivadas e irmãs entre si.

Resultados semelhantes foram recuperados por Tarrío *et al.* (2000), a partir da análise de sequências do gene *xantina-desidrogenase* (*Xdh*). *D. insularis* apresentou-se como a linhagem mais antiga, seguida por *D. tropicalis*, *D. willistoni*, *D. paulistorum* e *D. equinoxialis*.

Mediante a análise das sequências nucleotídicas do gene nuclear álcool desidrogenase (*Adh*) e do gene mitocondrial, *citocromo oxidase II* (*COII*), O'Grady & Kidwell (2002) propuseram um cenário evolutivo para o subgênero *Sophophora*, incluindo no estudo, do grupo *willistoni*, principalmente as espécies crípticas. Observa-se que diferentemente dos trabalhos anteriormente citados, *D. equinoxialis* constitui a base da ramificação, *D. insularis* agrupa-se ao *cluster* formado por *D. willistoni* e *D. tropicalis*, de forma a constituir o grupo-irmão do clado que une *D. paulistorum* e *D. pavlovskiana*.

Em 2006, Rohde *et al.* apresentaram as relações filogenéticas entre cinco espécies crípticas (*D. pavlovskiana* não foi incluída), investigada pelo estudo dos padrões de bandamento encontrados no cromossomo IIR. Esta análise citogenética também confirma a estreita relação evolutiva entre *D. equinoxialis* e *D. paulistorum* (sendo que as semiespécies Andino-Brasileira, Centro-Americana, Orinocana e interior foram incluídas no estudo) uma vez que uma inversão, a IIRd, é compartilhada entre ambas, indicando que este rearranjo de genes antecede a separação destas linhagens das outras espécies do subgrupo *willistoni*. Com relação à *D. willistoni*, *D. tropicalis*, *D. equinoxialis* e *D. paulistorum* os autores destacam o fato de que cada uma das inversões segregantes é restrita a apenas uma das espécies, sendo consistente com o isolamento reprodutivo efetivo entre estas espécies na natureza. Além disso, o trabalho fornece suporte para a idéia de que *D. insularis* seja a espécie mais divergente dentro do subgrupo *willistoni*. Embora o estudo evolutivo inclua apenas quatro das seis semiespécies, mostra-se congruente com o trabalho de Gleason *et al.* (1998) ao apresentar Andino-Brasileira e Orinocana como homossequências entre si (IIRdg), ao mesmo tempo em que relata o compartilhamento da inversão IIRh encontrada em estado polimórfico na semiespécie Centro-Americana e fixada na semiespécie Interior.

Recentemente, Robe *et al.* (2010) através de um exaustivo trabalho de reconstrução filogenética, demonstrou diversos aspectos relacionados à filogenia do subgrupo *willistoni*, com base em dados moleculares do gene mitocondrial citocromo oxidase II (*COII*) e dos genes nucleares *hunchback* (*hb*) e alfa metildopa (*Amd*). Os dados suportam, devido a algumas incongruências obtidas, um cenário no qual os locos, mitocondrial e nuclear, evoluíram através de diferentes histórias evolutivas. Os autores, através da análise das incongruências topológicas, basicamente obtidas a partir do gene mitocondrial, sugerem que estas podem ser resultantes de introgressão assimétrica. Evidências de introgressão foram fornecidas por Winge & Cordeiro (1960, 1963), Cordeiro (1968), Cordeiro & Winge (1995) e Valiati VH (1999) quando reportaram a ocorrência de cruzamentos interespecíficos entre espécies do subgrupo críptico. O trabalho de Valiati VH (1999) apóia os resultados dos autores Cordeiro & Winge (1995) de que em populações sulinas e geograficamente marginais das espécies *D. willistoni* e *D. paulistorum*, segregam genótipos capazes de produzir híbridos interespecíficos, algumas vezes férteis no caso de fêmeas, devido ao forte isolamento sexual entre elas. Ehrman & Powell (1982)

demonstraram a possibilidade de introgressão, como anteriormente citado, para as semiespécies de *D. paulistorum*, de forma que a hibridização mediada por introgressão mitocondrial pode resultar de retrocruzamentos entre as fêmeas híbridas e uma de suas semiespécies parentais.

Com base nos dados nucleares, Robe *et al.* (2010) chegaram à hipótese filogenética, onde as espécies crípticas se ramificam a partir de *D. insularis*, seguindo sucessivamente *D. tropicalis*, *D. willistoni*, *D. equinoxialis*, *D. paulistorum* e *D. pavlovskiana*. Relações evolutivas similares foram corroboradas, como citado anteriormente, com o uso de marcadores moleculares e citogenéticos por Gleason *et al.* (1998), Tarrío *et al.* (2000) e por Rohde *et al.* (2006). No que diz respeito à evolução das semiespécies de *D. paulistorum*, a Amazônica teria divergido primeiro e aparece como irmã do clado que se ramifica em duas linhagens independentes: o primeiro agrupa Andino-Brasileira e Orinocana e o segundo, as semiespécies Interior e Centro-Americana, sendo que Transicional se posiciona como táxon-irmão deste último. A inferência de Amazônica ser a semiespécie mais primitiva é coerente com as conclusões de Spassky *et al.* (1971) e de Gleason *et al.* (1998). Os dados citogenéticos de Rohde *et al.* (2006) também apóiam os dois clados independentes formados por Andino-Brasileira/Orinocana e Interior/Centro-Americana. Spassky *et al.* (1971) e Ayala *et al.* (1974) além de apresentarem a semiespécie Transicional relacionada à Centro-America, sugerem também a afinidade filogenética entre Transicional e Andino-Brasileira, sendo que ambas as relações estão de acordo com híbridos viáveis e férteis observados (Dobzhansky *et al.*, 1964; Ehrman & Powell, 1982).

Com relação ao subgrupo *bocainensis*, podemos destacar *D. nebulosa* Stutervant (1916), amplamente distribuída desde os Estados Unidos, Índias Ocidentais, México, até o Brasil e Peru; *D. fumipennis* Duda (1925), registrada desde El Salvador até o Brasil; *D. capricorni* Dobzhansky & Pavan (1943), que se distribui do Panamá até o sul do Brasil; e *D. sucinea* Patterson & Mainland (1944), com ocorrência do México à Colômbia (Val *et al.*, 1981).

Poucos são os trabalhos realizados com as espécies dos subgrupos *alagitans* e *bocainensis* do grupo *willistoni*, bem como os estudos que auxiliem no esclarecimento das relações entre as mesmas. De um modo geral verifica-se nos trabalhos de Tarrío *et al.* (2000), O'Grady & Kidwell (2002) e Powell *et al.* (2003) a definição de três linhagens dentro do grupo *willistoni*, de acordo com as espécies estudadas: a primeira une *D.*

capricorni e *D. sucinea*; a segunda, mono-específica, formada por *D. nebulosa*; e a terceira formada pelo clado das espécies crípticas. Segundo Tarrío *et al.* (2000) o subgrupo *bocainensis*, representado neste trabalho pelas espécies *D. capricorni*, *D. sucinea* e *D. nebulosa*, apresenta uma composição parafilética em relação ao subgrupo *willistoni*, sendo *D. nebulosa* mais intimamente relacionada às espécies crípticas do que aos demais membros de seu subgrupo. Entretanto, no trabalho de O'Grady & Kidwell (2002), o subgrupo *bocainensis* (com a adição de *D. fumipennis*, quando comparado ao trabalho anterior) é polifileticamente definido, uma vez que *D. nebulosa* é agrupada isoladamente com espécies do grupo *saltans*, o clado *D. capricorni/D. sucinea* constitui o grupo-irmão do subgrupo *willistoni* e *D. fumipennis* aparece como espécie-irmã do clado formado pelos grupos *saltans/willistoni*. O posicionamento relativo de *D. nebulosa* e *D. willistoni* na análise de Powell *et al.* (2003), apresentou-se variável com relação aos diferentes métodos de reconstrução filogenética, as quais ou são colocadas no mesmo clado ou são colocadas dentro do agrupamento *D. capricorni/D. sucinea*, quer com *D. willistoni* ou *D. nebulosa* sendo o táxon basal. Contudo, o maior corpo de evidências leva à proposição de que *D. nebulosa* é a espécie-irmã do subgrupo *willistoni*.

Outra questão relevante e controversa trata a elucidação da monofilia do grupo *willistoni*. O trabalho de Tarrío *et al.* (2000) apóia a monofilia dos grupos *willistoni* e *saltans* em sua filogenia, ligando o fato de terem encontrado uma deleção de um íntron de *Adh* sinapomórfico para o grupo *willistoni*. Entretanto, Pélandakis *et al.* (1991) e O'Grady & Kidwell (2002) definem este grupo como parafilético em relação ao grupo *saltans*.

Diferentes atributos dos genomas de *Drosophila* são muito conservados entre as espécies. Tamanho total do genoma, o número de genes, a distribuição de classes de elementos transponíveis, e os padrões de uso do códon são muito semelhantes em todos os 12 genomas já investigados. *D. willistoni*, como caso excepcional, apresenta o uso de códon enviesado (*codon bias*), a segunda maior quantidade de elementos transponíveis, a ausência potencial de selenoproteínas, assim como um número maior de rearranjos no *cluster* dos genes *Hox* entre as espécies de *Drosophila* (*Drosophila* 12 Genomes Consortium, 2007). Além das evidências que distingue a *D. willistoni* das demais espécies de *Drosophila*, o grupo *willistoni* constitui um complexo em diferentes estágios evolutivos, uma vez que é composto por espécies, subespécies, semiespécies e ainda por espécies crípticas, contendo inclusive diversas espécies incipientes em *statu nascendi* (Dobzhansky

& Spassky, 1959). Tendo muitas espécies simpátricas e ao mesmo tempo algumas bastante restritas, fornece, portanto, um material repleto de informações biológicas para os estudos dos vários níveis de divergência evolutiva.

CAPÍTULO II

OBJETIVOS

Objetivo geral

A partir das evidências que apontam o elemento *Galileo* como o responsável pela geração das inversões naturais em cromossomos de linhagens de *D. buzzatii* e de sua ocorrência em diferentes gêneros de *Drosophila*, inclusive em *D. willistoni* e, considerando que a *D. willistoni* e suas crípticas são extremamente polimórficas cromossomicamente, o achado de *Galileo* em seu genoma torna-se bastante promissor para estudos sobre os agentes causadores das quebras e rearranjos amplamente distribuídos no grupo *willistoni*. Dessa forma o trabalho pretende caracterizar a presença deste transposon no genoma de indivíduos de espécies pertencentes ao grupo *willistoni* provenientes de diferentes populações naturais, dotadas de extenso polimorfismo cromossômico para inversões paracêntricas. Este trabalho é extremamente oportuno e promissor, uma vez que a linhagem sequenciada de *D. willistoni* (GdH4), onde foi detectada a presença de *Galileo*, é cromossomicamente monomórfica, extremamente endocruzada e produto de cruzamentos irmão-irmã. Como tal, ela não representa as populações naturais, cromossomicamente polimórficas.

Objetivos específicos

- Verificar a presença de *Galileo* em diferentes espécies do grupo *willistoni* de *Drosophila*.
- Caracterizar molecularmente as sequências nucleotídicas de variantes possivelmente funcionais do elemento *Galileo* nestas espécies.
- Detectar a presença de sequências que caracterizariam uma transposase funcional nos genomas das populações naturais de espécies do grupo *willistoni*.

CAPÍTULO III

METODOLOGIA

Manutenção dos estoques de *Drosophila*

As linhagens das espécies *D. willistoni*, *D. paulistorum* (Amazônica, Andino-Brasileira, Interior e Orinocana), *D. equinoxialis*, *D. tropicalis*, *D. insularis*, *D. nebulosa*, *D. capricorni*, *D. sucinea*, *D. fumipennis* (Tabela 1), foram mantidas como culturas massais no Laboratório de *Drosophila* da UFRGS, dentro de câmaras de temperatura e umidade constante (a 17 °C ou a 25 °C ± 1 grau e 60% de umidade relativa). A manutenção das populações era realizada semanalmente em meio de cultura estabelecido por Marques *et al.* (1966). Em todos os experimentos, a linhagem já sequenciada de *D. willistoni*, GdH4, será utilizada como controle (Tabela 1). O meio de cultura composto por farinha de centeio (5%), açúcar (10%) e Agar (0,1%), foi utilizado para facilitar a manutenção de linhagens mais sensíveis, como a própria GdH4.

Extração de DNA

O DNA genômico foi extraído a partir de moscas adultas segundo o método de Sassi *et al.* (2005) modificado. Modificou-se o volume de tampão de lise (Tris-HCl 0.1 M; EDTA 0.1 M; SDS 1% e NaCl 0.06 M) utilizado, de 750 µL para 650µL, assim como o período de incubação, de 30 minutos para 1 hora a 65 °C. Além disso, o DNA extraído foi ressuspendido em 30-40 µL de H₂O ultra pura, ao invés de Tris-EDTA (TE).

Técnica da Reação em Cadeia da Polimerase (PCR)

Para a investigação da presença do elemento *Galileo* foram construídos *primers*, denominados de “Gal 846” (5'-TGT TCT TGC TCC TCT GTT CC-3') e “Gal 3247” (5'-CTT ATG TAT CAC GAA TCC CG-3'). O primeiro, senso, anela em uma região supostamente pertencente ao domínio *M* e o anti-senso ancora em uma porção da inferida transposase, para que o domínio *THAP* (motivo de ligação ao DNA) possa ser amplificado (Figura 9). O delineamento dos oligonucleotídeos foi realizado de forma que fosse evitado o anelamento entre eles, assim como a amplificação de um elemento defectivo sem a TPase ou de outro elemento *foldback* que não fosse da subfamília *GalileoG* (Delprat *et al.*,

2009), antes denominados como *Kepler* e *Newton*. Os *primers* foram construídos com base na sequência de *Galileo* encontrada por Marzo *et al.* (2008) no genoma de *Drosophila willistoni* (*Drosophila* 12 Genomes Consortium, 2007).

Para o procedimento utilizou-se um volume total de 25 µL com 80-100 ng de DNA genômico, 20 pmol de cada *primer*, 200 µM de dNTPs, 1,0 mM de MgCl₂ e 1,5 unidade de Taq DNA polimerase. Empregou-se uma desnaturação inicial a 94 °C por 3 minutos, seguida por 32 ciclos de desnaturação a 94 °C por 1 minuto, anelamento dos *primers* a 50 °C por 45 segundos, extensão a 72 °C por 3 minutos e extensão final a 72 °C por 6 minutos, ou por 10 minutos para a ligação do fragmento amplificado ao vetor de clonagem.

Tabela 1. Lista de espécies e linhagens mantidas em culturas massais utilizadas neste trabalho.

Subgrupo	Espécie	Semiespécie	Identificação	Origem geográfica	
<i>willistoni</i>	<i>D. willistoni</i>		GdH4	Guadaloupe, Pequenas Antilhas	
			17A2	Eldorado do Sul (RS), Brasil	
			Cip	Serra do Cipó (MG), Brasil	
			Cor	Coronilla, Uruguai	
			Dla	Dois Lajeados (RS), Brasil	
			Isc	Florianópolis (SC), Brasil	
			Mex	Apazapan, México	
			Rib	Ribeirão Preto (SP), Brasil	
			Wip4	Salvador (BA), Brasil	
		<i>D. paulistorum</i>	Amazônica		Ama
Andino-Brasileira	And-MLC			Florianópolis (SC), Brasil	
	And-Rib			Ribeirão Preto (SP), Brasil	
Interior	Int			Llanos, Colômbia	
<i>D. tropicalis</i>	Orinocana		Ori	Georgetown, Guiana	
			tro-Jal	Jalisco, México	
<i>D. equinoxialis</i>			tro-SS	San Salvador, El Salvador	
			equi	Cidade do México, México	
<i>bocainensis</i>	<i>D. insularis</i>		insu	Saint Kitts, Pequenas Antilhas	
		<i>D. nebulosa</i>	nebu	Porto Alegre (RS), Brasil	
			<i>D. capricorni</i>	capri	Porto Alegre (RS), Brasil
			<i>D. sucinea</i>	suci	Cidade do México, México
	<i>D. fumipennis</i>	fumi	Joinville (SC), Brasil		

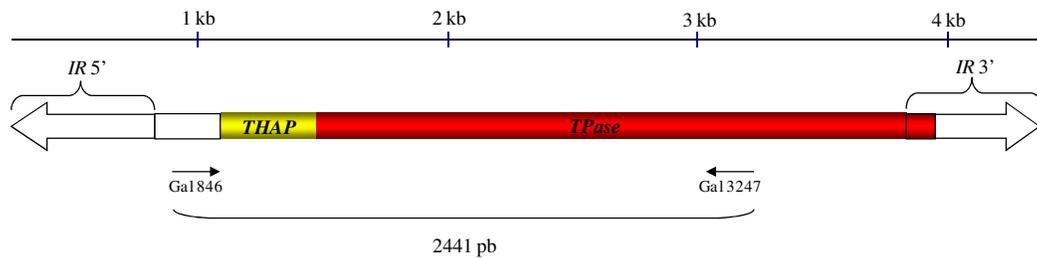


Figura 9. Representação gráfica da maior variante de *Galileo* em *D. willistoni* descrita por Marzo *et al.* (2008), assim como a localização do par de *primers* (Gal 846 e Gal 3247) desenhados e a respectiva região amplificada.

Medição do comprimento dos fragmentos em número de pares de base (pb)

Para estimar as medidas relativas de migração e o tamanho dos fragmentos, após a amplificação e a fotodocumentação, foram utilizados os *softwares*: Adobe® Photoshop® 5.5 e DNASIZE (Raghava, 1994) sendo como controle os fragmentos do marcador de peso molecular *1 Kb Plus DNA Ladder*™ (Invitrogen) e o fragmento de 2441 pb correspondente ao segmento de *Gallileo* amplificado a partir de *D. willistoni*, linhagem GdH4. O DNASIZE é uma ferramenta *online* (imtech.res.in/raghava/dnasize) que estima o tamanho em pares de base considerando a distância de migração (deslocamento do fragmento) do ponto de aplicação até o fragmento (banda no gel) em relação ao marcador de peso molecular. O Adobe® Photoshop® 5.5 foi utilizado para estimar a distância de migração a partir do ponto de aplicação dos produtos amplificados pela *PCR*.

Dot blot

A sonda para hibridizar com as possíveis cópias do elemento *Galileo* dispersos no genoma das espécies em estudo foi produzida a partir do produto da amplificação de *D. willistoni* (GdH4). Utilizou-se um total de 2µg de DNA das espécies submetidas ao método.

Na técnica empregou-se o *kit* de marcação e detecção *AlkPhos*™ (*GE Healthcare*), método não radioativo de detecção e de marcação direta, na qual a sonda é marcada com fosfatase alcalina (AP). Após a hibridização dos ácidos nucléicos e as lavagens com as diferentes estringências, o substrato *CDP-Star* é adicionado. Na etapa de detecção a fosfatase alcalina, ligada à sonda, catalisa quebras dos dioxietanos do substrato *CDP-Star*

produzindo um sinal quimioluminescente que é capturado pelo filme de radiografia. Todo o procedimento foi realizado de acordo com as recomendações do fabricante.

Purificação dos fragmentos amplificados

Os produtos da amplificação do DNA foram purificados, a partir da banda removida do gel de agarose após a eletroforese, utilizando-se o *kit GFXTM PCR and Gel Band Purification*® (GE Healthcare) conforme recomendações do fabricante.

Clonagem e sequenciamento

Os fragmentos polimórficos foram clonados no vetor *pGEM[®]-T Easy Vector System* (Promega) e posteriormente sequenciados automaticamente com os *primers* universais M13F e M13R-pUC pela *Advancing through Genomics MACROGEN*. Dessa forma, com base nas variantes de *Galileo* e de outros elementos *foldback* já depositadas em bancos de dados, foi realizada a caracterização molecular das sequências nucleotídicas, assim como dos domínios das inseridas transposases das variantes isoladas.

Análise e edição das sequências

A análise dos cromatogramas foi realizada no programa *Chromas 2*; o alinhamento das sequências nucleotídicas e de aminoácidos foi obtido no *software Muscle* com posterior edição no *Bio Edit*; o programa *Gene Doc* foi utilizado na tradução das sequências assim como na edição final das mesmas.

A presença (1) ou ausência (0) das diferentes variantes de *Gallileo* em cada uma das espécies e semiespécies de *D. paulistorum* foram convertidas em matriz de dados binários, a partir da qual foi estimada a dissimilaridade com base no complemento do coeficiente de similaridade de Jaccard (Jaccard, 1908). A matriz foi utilizada em uma análise de agrupamento tipo *UPGMA* (do inglês *Unweighted Pair-Group Method by Arithmetic Averages*) (Sneath & Sokal 1973), que produz um arranjo hierárquico de classificação representado por um dendrograma, utilizando-se o *software NTSYSpc 2.1* (Rohlf, 2000). O ajuste entre as matrizes de similaridade e os respectivos dendrogramas foi estimado pelo coeficiente de correlação cofenética (r) (Sokal & Rohlf, 1962) e para a estimativa da significância da correlação (associação) entre as matrizes obtidas foi

empregado o teste de comparação de matrizes de Mantel, com 1.000 permutações (Mantel, 1967), por meio do programa NTSYS pc 2.1 (Rohlf, 2000).

CAPÍTULO IV

RESULTADOS

Isolamento de *Galileo* no grupo *willistoni*

Foram investigadas, do subgrupo *willistoni*, *D. willistoni*, *D. tropicalis*, *D. equinoxialis*, *D. insularis* e quatro semiespécies de *D. paulistorum* (Amazônica, Andino-Brasileira, Interior e Orinocana), e do subgrupo *bocainensis*, *D. nebulosa*, *D. capricorni* e *D. fumipennis* pelas técnicas de *Dot blot* e *PCR* (Figuras 10, 11 e 12). A amplificação dos fragmentos de *Galileo*, como regra, resulta em baixo rendimento, assim as Figuras 11 e 12 ilustram os produtos de *PCR* após terem passado pelo procedimento de clonagem, os quais a partir do vetor eram mais facilmente amplificados, apresentando um rendimento satisfatório. A mesma dificuldade encontrada por Cáceres *et al.* (2001) e Casals *et al.* (2005) na amplificação de *Galileo* por meio da técnica de *PCR* foi constantemente observada no decorrer deste trabalho. Resultados aparentemente negativos quanto à presença do elemento no genoma de algumas espécies, tiveram que ser cautelosamente trabalhados, assim como o baixo rendimento dos produtos quando amplificados. Mediante esta dificuldade realizou-se a técnica de *Dot blot* (Figura 10) com intuito de verificarmos a presença ou ausência deste transposon no grupo *willistoni*.

De acordo com o desenho dos *primers*, isolou-se um produto de 2441 pb da linhagem GdH4 de *D. willistoni* (controle positivo), com o qual foram comparados os fragmentos amplificados das demais espécies. Como regra, foi observada a variabilidade no tamanho dos fragmentos amplificados de *Galileo* tanto intra como interespecificamente e todas as sequências analisadas são indicativas de cópias defectivas de *Galileo*. O fragmento de maior tamanho, de aproximadamente 2198 pb, foi isolado a partir de *D. paulistorum* (Figura 11), mas nenhuma espécie apresentou um produto de igual ou maior tamanho do que a linhagem usada como referência para o delineamento dos *primers*.

Na verificação da presença de *Galileo* em *D. willistoni* por meio da técnica de *PCR*, obteve-se indícios de diferença interpopulacional quanto ao tamanho dos fragmentos amplificados. Mesmo considerando-se a possível dificuldade em amplificar este transposon a partir do DNA genômico, após inúmeras reações, observou-se que o fragmento de maior tamanho (de aproximadamente 1587 pb) para *D. willistoni*, foi isolado apenas a partir de

três populações (Cip, 17A2 e Wip4) entre oito (Cor, Dla, Isc, Mex e Rib, além das três já citadas) submetidas à técnica (dados não apresentados) e os demais fragmentos amplificados foram compartilhados entre as populações.

Ao visualizarmos o resultado da eletroforese dos produtos de *PCR* a partir do DNA genômico das espécies estudadas, observamos o aparente compartilhamento de um único fragmento de aproximadamente 1200 pb entre as espécies *D. tropicalis*, *D. equinoxialis*, *D. insularis* e *D. paulistorum* (Figuras 11 e 12). A partir do isolamento e purificação da única banda em gel de agarose para cada espécie, obteve-se, em alguns casos, mais de uma variante após o procedimento de clonagem e de sequenciamento, de forma que o indício inicial do compartilhamento do fragmento foi confirmado, porém com a caracterização molecular de mais de uma forma de *Galileo*.

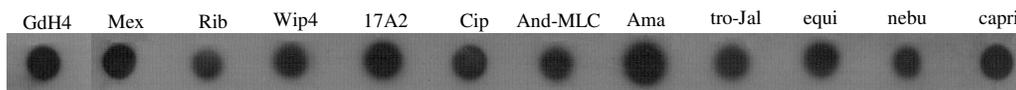


Figura 10. Detecção da presença de *Galileo* por meio da técnica de *Dot blot*. A primeira amostra, oriunda da linhagem GdH4 de *D. willistoni*, corresponde ao controle positivo, a partir da qual a sonda foi produzida. O subgrupo *willistoni* está representado neste procedimento por cinco populações de *D. willistoni* (Mex, Rib, Wip4, 17A2 e Cip, respectivamente), pelas semiespécies Andino-Brasileira (população de Santa Catarina) e Amazônica, pela *D. tropicalis* (de Jalisco) e pela *D. equinoxialis*; do subgrupo *bocainensis* foram submetidas à técnica *D. nebulosa* e *D. capricorni*.

Caracterização molecular das variantes isoladas

A partir da caracterização molecular das sequências obtidas, foi possível identificar 22 variantes de *Galileo* entre as espécies investigadas do grupo *willistoni*, sendo que as sequências que deram origem a cada variante apresentaram uma divergência nucleotídica na ordem de 8%. Com relação ao tamanho das sequências caracterizadas, *D. sucinea* apresentou a menor, com 393 pb e *D. willistoni* a maior, com 1522 pb (fragmentos maiores como aqueles amplificados a partir de *D. paulistorum* e *D. tropicalis* não tiveram até o momento a caracterização totalizada, pois serão necessários *primers* internos para que haja o sequenciamento completo destas variantes). Mediante a comparação das sequências isoladas, observou-se a deleção de vários segmentos, como por exemplo, em *D. tropicalis* quando na análise de duas sequências de aproximadamente 1200 pb, oriundas de

populações diferentes (de Jalisco e de San Salvador), identificou-se a ausência de 7 pb na extremidade 3' do fragmento de *Galileo* (Figura 13).

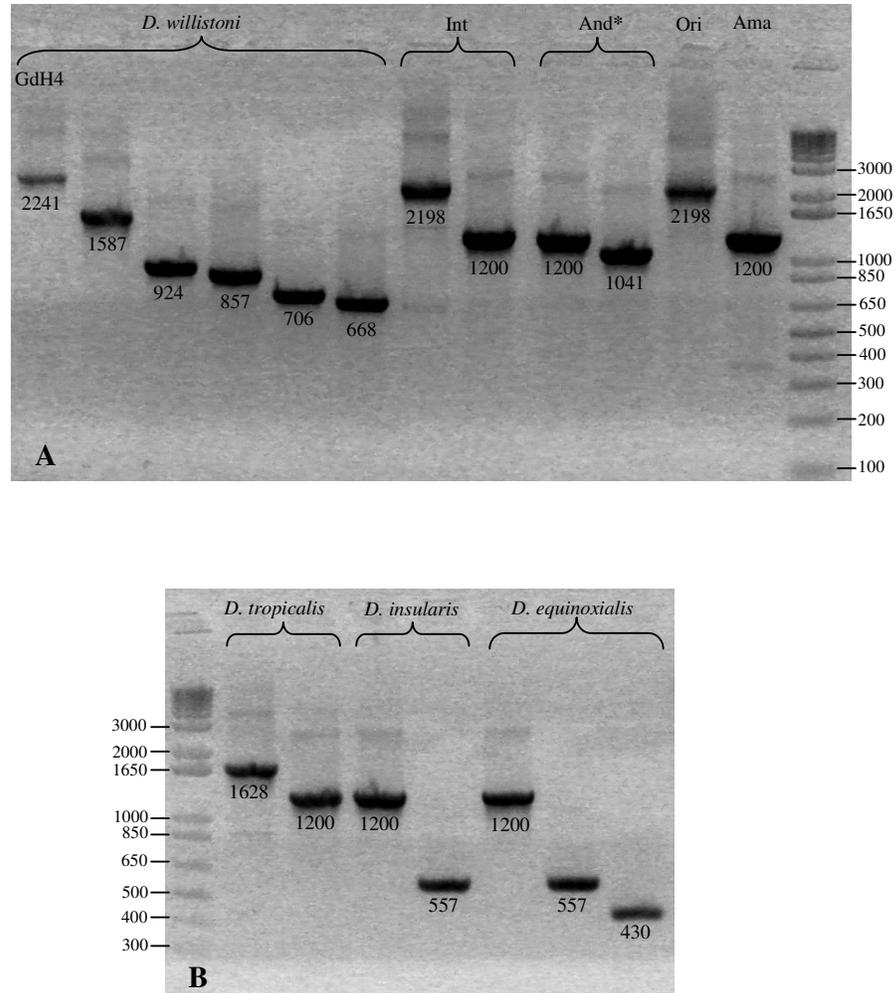


Figura 11. Fragmentos de *Galileo* obtidos a partir da técnica da *PCR* nas espécies do subgrupo *willistoni* após a realização do procedimento de clonagem. A: *D. willistoni* (com cinco fragmentos de comprimentos distintos, além daquele oriundo da linhagem GdH4, o qual é o controle positivo) e *D. paulistorum* (Interior, Andino-Brasileira, Orinocana e Amazônica respectivamente, com um total de três fragmentos de tamanhos diferentes). B: *D. tropicalis*, *D. insularis* (com dois fragmentos cada uma) e *D. equinoxialis* (com três fragmentos isolados). Os valores abaixo das bandas representam o tamanho aproximado em pares de bases (pb) de cada fragmento. * A semiespécie Andino-Brasileira, assim como Interior e Orinocana também apresentou um fragmento de aproximadamente 2198 pb amplificado e posteriormente caracterizado como mostram os resultados a seguir.

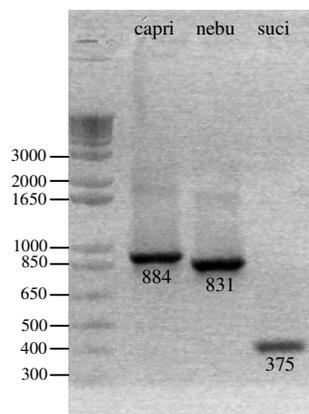


Figura 12. Fragmentos de *Galileo* obtidos a partir da técnica da *PCR* em *D. capricorni*, *D. nebulosa* e *D. sucinea*, respectivamente, do subgrupo *bocainensis* após a realização do procedimento de clonagem. Os valores abaixo das bandas estão representando o tamanho aproximado em pares de base (pb) de cada fragmento. *D. fumipennis* teve um fragmento de aproximadamente 1342 pb amplificado (dado não apresentado).

A Figura 14 mostra o dendrograma resultante da análise de agrupamento pelo método *UPGMA* e a similaridade existente entre as espécies investigadas em decorrência da presença ou ausência das 22 diferentes variantes de *Gallileo* caracterizadas no presente estudo. O coeficiente de correlação cofenética do dendrograma ($r = 0,40$) mostrou baixo ajuste entre a matriz de similaridade e o respectivo dendrograma, sem correlação entre as matrizes ($P = 0,9940$). Contudo, ressalta-se que graficamente é possível observar a formação de dois clados bem definidos: (1) formado pelas quatro semiespécies de *D. paulistorum* (Amazônica, Andino-Brasileira, Interior e Orinocana) submetidas ao estudo e (2) pelas espécies crípticas *D. tropicalis*, *D. insularis* e *D. equinoxialis*, em um total de 11 variantes, sendo seis compartilhadas entre duas a três espécies ou semiespécies. *D. willistoni* do subgrupo *willistoni*, mais *D. sucinea*, *D. nebulosa* e *D. capricorni* do subgrupo *bocainensis*, apresentaram variantes exclusivas e, por consequência, formaram ramos independentes aos dois agrupamentos.

Cabe salientar que entre todas as espécies crípticas do subgrupo *willistoni* somente *D. willistoni* não compartilha alguma de suas quatro variantes. Em *D. tropicalis* e *D. equinoxialis* foram caracterizadas três variantes: uma forma compartilhada entre ambas as espécies e a semiespécie Andino-Brasileira (pertencente ao outro clado); uma segunda variante de *D. equinoxialis* foi também identificada em *D. insularis*, assim como ocorre

com uma forma estrutural caracterizada em *D. tropicalis*, com a diferença que também é compartilhada com a semiespécie Interior (agrupada com as demais semiespécies de *D. paulistorum*); a terceira, é exclusiva de *D. tropicalis* e *D. equinoxialis*. O último agrupamento apresenta inicialmente Andino-Brasileira com cinco formas de *Galileo*, como a semiespécie dotada do maior número de variantes isoladas e caracterizadas, seguido por Interior com três variantes, Amazônica e Orinocana com duas. Dentro deste complexo, além do compartilhamento das variantes já descritas, Andino-Brasileira possui em comum apenas uma forma com Orinocana e outra com Interior (duas formas estruturais são exclusivas), enquanto as duas últimas também compartilham uma variante, ausente em Andino-Brasileira, com a semiespécie Amazônica.

Análise das sequências de aminoácidos

Foram incluídas na análise as sequências de aminoácidos que continham pelo menos 300 resíduos e dessa forma, seis variantes de *Galileo* caracterizadas no grupo *willistoni* foram comparadas com outras variantes de *Galileo* previamente caracterizadas (Marzo *et al.*, 2008). Embora as variantes isoladas neste trabalho sejam estruturalmente mais degeneradas, dentro do motivo *THAP* observou-se a presença de alguns resíduos conservados, assim como o compartilhamento de resíduos entre as inferidas transposases de outras variantes de *Galileo* previamente caracterizadas no trabalho dos autores acima (Figura 15).

Corroborando com a análise de agrupamento, a sequência de aminoácidos da variante 14, isolada a partir de populações de *D. willistoni*, apresenta segmentos exclusivos, uma vez que não são observados nem nas demais espécies do subgrupo *willistoni*, como também naquelas que tiveram o elemento *Galileo* já caracterizado por Marzo *et al.* (2008).

```
trop-Jal : (1189) GTCGGCTGCTTTTCCGTTACACGGGATTTCGTGATACATAACAATCACTAGTGAATTCGCGGCCGCCTGCAGGTC (64)
trop-SS  : (1189) GTCGGCTGCTTTTCCGTTACACGGGATTTCGTGAT-----AATCACTAGTGAATTCGCGGCCGCCTGCAGGTC (64)
```

Figura 13. Sequências nucleotídicas de um dos fragmentos de *Galileo* isolado a partir de duas populações de *D. tropicalis*. Na população de Jalisco (tro-Jal) há 7 nucleotídeos na extremidade 3' (circulados em vermelho), os quais estão ausentes na população de San Salvador (tro-SS). O segmento sublinhado em azul corresponde à sequência nucleotídica do vetor de clonagem. Os valores entre parênteses correspondem ao número de nucleotídeos do transposon omitidos inicialmente, assim como aqueles do vetor no término da sequência. Considerando os 7 nucleotídeos, o fragmento apresenta 1230 pb.

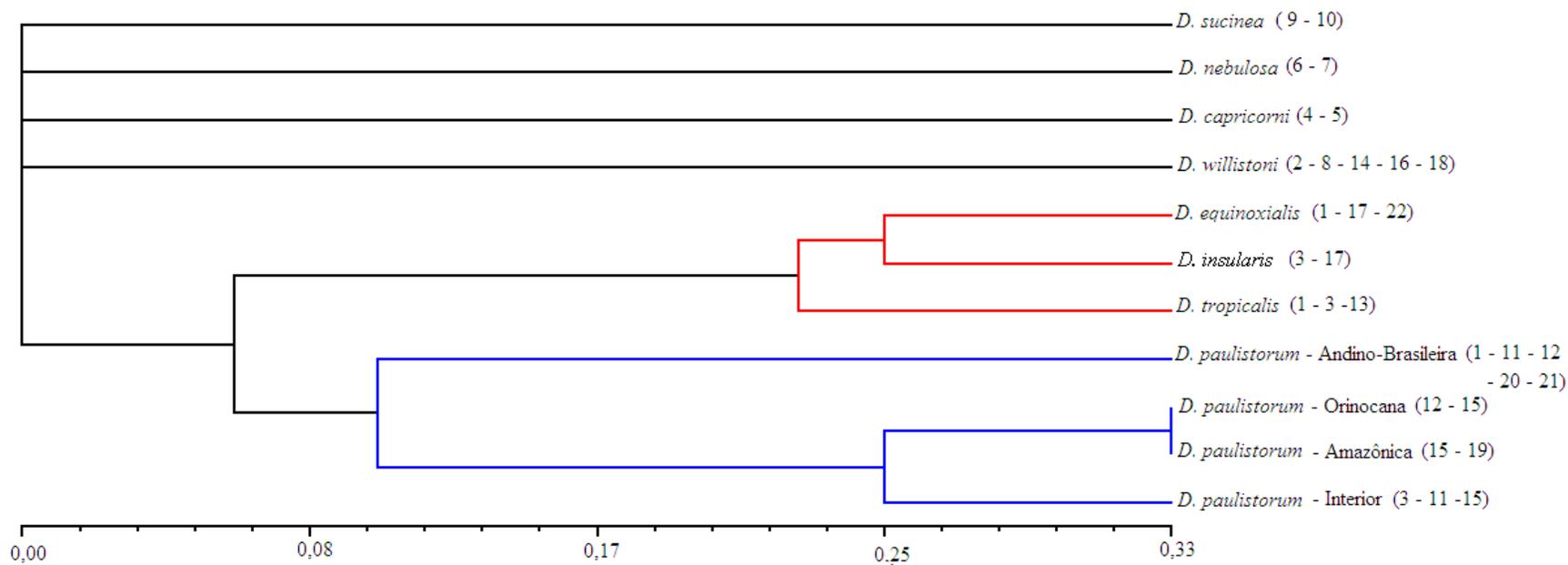


Figura 14. Dendrograma resultante da análise de agrupamento pelo método *UPGMA* de cinco espécies do subgrupo críptico *willistoni* (*D. willistoni*, *D. equinoxialis*, *D. insularis*, *D. tropicalis*, *D. paulistorum*), quatro semiespécies de *D. paulistorum* (Andino-Brasileira, Orinocana, Amazônica e Interior), mais três espécies do subgrupo *bocainensis* (*D. sucinea*, *D. nebulosa* e *D. capricorni*) em relação à presença ou ausência das 22 diferentes variantes de *Gallileo* caracterizadas. Os números entre parênteses correspondem às diferentes variantes caracterizadas neste trabalho.

```

      *          20          *          40          *          60          *          80          *          100          *          120          *          140
V_3   : -----MYHESVIAKQPTISIMMAHT--FWSIMATREQGMRFSAFAQHF----- : 45
V_14  : -----LAPLFL--CFGIILSVPLQICLTGATL--MISLEGLTITGTCNSTWVYQTLLLSGRKE----- : 56
V_20  : -----FDLDSPMSCQINYSYENI-----YISNEAIT--FETHLLCDKNF----- : 38
V_1   : -----LAPLFL--PDEWRCLDILREC-----DILREGCQTEFAKQPTI----- : 36
V_15  : -----DILREC-----DILREDCQTEFAKQPTCN----- : 25
V_19  : -----ILREC-----DILREDCQTEFAKQPTCN----- : 24
Dpse  : (44) MFNFWVEVCGIDSNLLKKTARICDRHHDVKYLGVRRLKAAVEFTLNLSENTLLTAFNALLCNDYEFSE-----NKQDNEFGTFLQAKR----IKLADVHNIMSNELE : 99
Dper  : (44) MFNFWVEVCGIDSNLLKKTARICDRHHDVKYLGVRRLKAAVEFTLNLSENTLLTAFNALLCNDYEFSE-----NKQDNEFGTFLQAKR----IKLADVHNIMSNELE : 99
Dana  : (40) TLKRWVDLCCDDIFNRRATARICNRHFEKAYIGKSRRLRSAAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NDAEFGIYVNPFRTEQFS----- : 89
Dbuz  : (57) EFNWVKICHVPSNFRKKAKICDRHFEKAYIGKSRRLRSAAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NRGAEHQENENVEECD----DELLENLLIFDINSK : 98
Dmoj-C : (53) YLNWWTIAQNFNNFDRARARICNRHFEKAYIGKRYLRRAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NDAEFGIYVNPFRTEQFS----- : 127
Dmoj-D : (53) QLNRWIKACNFNNFDFKSRICNRHFEKAYIGKRYLRRAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NDAEFGIYVNPFRTEQFS----- : 127
Dwil  : (52) VRRQVSKCNLCYDVIDRFLICNLHFEKAYIGKRYLRRAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NDAEFGIYVNPFRTEQFS----- : 100
Dvir  : (36) VRRQVSKCNLCYDVIDRFLICNLHFEKAYIGKRYLRRAVEFTLNLSENTLLCPNSVPVCEKFFDFDI-----NDAEFGIYVNPFRTEQFS----- : 127

```

```

      *          160          *          180          *          200          *          220          *          240          *          260          *          280
V_3   : -----KCGSAAQAVAFVSWIISLMCSLVPGSWIISTVRSYRICLHLLIWI--TSRCSPISRSESLSAANSNRIPLMWCIMFRICRHLRSQPKRKSRRSMHNFHRSRSHIRW----ISCPA : 156
V_14  : -----ARRTLNTPLKIGIEERAISSFAALQC--SMDLIVQDTLLEELVQIGILEVLLIMFH-----FVTKFWLAEWAA-----AIIIRRAIGDRIQVQENYRNSEK-CIS : 148
V_20  : -----PVEQPEPSQFGHGR-----GSESPSRCLLFGLLARGHPEFRICEELPVTWPEPSQPGVGHLSSECDCCKWLR-----SREQPVKLL-----KTSRE : 121
V_1   : -----NEIYFVITMIKRS--SLRNLPSCLARGNNSARFQVHRCOMPITWSVITYSLSTTHMLLSSDIRKYY-----SIRRNALPKKTIINESITY-----HYCRV : 122
V_15  : -----NEIYFVITMIKSYSLRNLPSCLARGNNSARFQVH--CCMPITWSVITYSLSTTHMLLSSDIRKYY-----SIRRNALPKKTIINESITY-----CRV : 114
V_19  : -----NEIYFVITMIKSYSLRNLPSCLARGNNSARFQVH--CCMPITWSVITYSLSTTHMLLSSDIRKYY-----SIRRNALPKKTIINESITY-----CRV : 113
Dpse  : QQCVTAPDKVSLN-KNESIDIRFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SAEARTICIMILKTLSSPFRKRVIAQN : 224
Dper  : QHCVTEEPDKVSLN-KNESINIRFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SAEARTICIMILKTLSSPFRKRVIAQN : 224
Dana  : --NSLDSFNSLETFCAPQLEDSEFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SDEARTLCKMILLKNSYNSAERKRVIAQN : 210
Dbuz  : KEFWQNLAVDR-----FYCNCKIKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SEQSVLCKMILLKNSYNSAERKRVIAQN : 214
Dmoj-C : NQLWQDLSAGRS-----SFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SDQSEVLCMILLKNSYNSAERKRVIAQN : 243
Dmoj-D : NQLWQDLSAGRS-----SFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SENSVLCRMILLKNSYNSAERKRVIAQN : 243
Dwil  : KV-----EDIS--DISTRDCMFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SKNSNTVCKMILLKNSYNSAERKRVIAQN : 217
Dvir  : HVGFPELYDNISLSAEKDSQDIRFCNCLKREQNEIYRKNYYELMADCKKRENGFIRLDRFALRQVY--YRRNRYS--LRTNLKENIIPFARMTV-----SXNSNTVCKMILLKNSYNSAERKRVIAQN : 255

```

```

*      300      *      320      *      340      *      360      *      380      *      400      *      420
V_3   : LHRLCQLVLRCKLPWSFRMAVQCHRRTRRPFPAK-CFRISQHVNT-----CQMMILSIAILINIKLTHKPCNRCFF----- : 231
V_14  : VEPCTQVWVLAHQIQHWVHFSMDTIFIGVFIHNSQSCCSMAIAWSKS-----EHSAAHKAATSSWNEI-LTKVAATASAPLETF----- : 227
V_20  : VSPMTGMGLWITTLIRSRKMGPCGCPCCRRWPKNFSTV-----RHFIFFRSGKIAWASV-WDEIAILLACPKSLKHY--- : 195
V_1   : IIDSFDRLCIFVYI--LFAHCWEAQQP-DRFRHLENYYHRLQGLWVNF-----IKLIKTAIDNSIIWQHV-LT-HCCYHDLGHEI----- : 198
V_15  : IM-ILTCIAYRE-LNILFAHCWEAQQPDRFRHLENYYHRLQGLWVNF-----IKLIKTAIDNSIIWQHV-LT-HCCYHDLGHEI----- : 191
V_19  : IM-ILTCIAYRE-LNILFAHCWEAQQPDRFRHLENYYHRLQGLWVNF-----IKLIKTAIDNSIIWQHV-LT-HCCYHDLGHEI----- : 190
Dpse  : IM-FYSSRTYEV-LRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLHYNIALDEIQGYENDGQNTKSLIGCQVCVEMIRGLFENWYV-LSYTVANGIHEALTKVTA : 361
Dper  : IM-FYSSRTYEV-LRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKR-----IALDEIQGYKNDGQNTKSLIGCQVCVEMIRGLFENWYV-LSYTVNLNGIHEALTKVTA : 355
Dana  : IM-FYSSRTYEV-MRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNTALDEIQGFENDGEKTRFELGQVCVEMIRGLFENWYV-LSYTVSANGIKHDAIMKKVEA : 347
Dbuz  : IM-FYSSRTYEV-LRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNSYLDEIQGFENDGEKTRFELGQVCVEMIRGLFENWYV-LSYTVSANGIRHSLIKSVEA : 351
Dmoj-C : MH-FYSARYDY-LRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNSYLDEIQGFENDGKTRFELGQVCVEMIRGLFENWYV-LSYTVSARGINHTLKKKFEF : 380
Dmoj-D : IM-FYSARYDY-LRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNSYLDEIQGFENDGKTRFELGQVCVEMIRGLFENWYV-LSYTVSARGINHTLKKKFEF : 380
Dwil  : IM-FYSSRTYEV-MRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNSHRCVEVEGFVDDGYEKTDAICKQICVEMVRGLYENWYV-LSYVATSTGLSSHKLTKLDS : 354
Dvir  : IH-FYFAKYDY-MRDVNLKLPKSRITLRWAVLKNMREVENFDLLSKIKTIFDGMSSKGRREAVILFDEIKIKRGLQYNSRRDEIEGHTDDGVKTEPELCKQISVEMVRGLYENWYV-LSYFATSTGLLTLTKKCTES : 392

```

```

*      440      *      460      *      480      *      500      *      520      *      540      *      560
V_3   : -----MCHNRSCGCC-----ASQC-----ANRILRNIAI-----VRII-----LIRQVYYS----- : 270
V_14  : -----SISYFLKTKC--KIKKRRVNRFAHQITEGARVLENTHEWEL--YNNLLKAG-----ICPIWRHTE-----LIRQVYYS----- : 302
V_20  : WLLSFCMFSILVINEIIMMTCISKNTSSCY-----MRIL-----FAFIYVHLQ----- : 238
V_1   : -----AGGLVIRIACCQAIINDHTVGSLSHITSCNHYNLCSAQ-----LSHLMCDFNF-----LCID-----YLFDFHFGGLD----- : 262
V_15  : -----AGGLVIRIACCQAIINDHTVGSLSHITSCNHYNLCSAQ-----LSHLMCDFNGF-----LCID-----YLFDFHFGGLD----- : 257
V_19  : -----AGGLVIRIACCQAIINDHAVGSLHITSCNHYNLCSAGC-----LSHLMCDFNGF-----LCIG-----YLFDFHFGGLD----- : 256
Dpse  : NIEQAQVLGLNVRAIICDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLIKSLRNIVRNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHYPNSFEKMKVKYATQVSSHSVAALRTMIS : 494
Dper  : NIERAQLVGLNVRAIICDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLIKSLRNIVRNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHYPNSFEKMKVKYATQVSSHSVAALRTMIS : 488
Dana  : NIEVSTLGLNVRAIICDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLIKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----VTRKHINPNSFEKMKVKYATQVSSHSVAALRTVVD : 480
Dbuz  : NIGLSQALGLNVKAVVCDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLVKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTSKHVSPNCFEKMVKYATQVSSHSVAALRTVVD : 484
Dmoj-C : NIGLSQALGLNVKAVVCDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLVKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHVSPNCFEKMVKYATQVSSHSVAALRTVVE : 513
Dmoj-D : NIGLSQALGLNVKAVVCDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLVKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHVSPNCFEKMVKYATQVSSHSVAALRTVVE : 513
Dwil  : NIEAARTLGLIRAVVCDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLVKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHVSPNCFEKMVKYATQVSSHSVAALRTVVD : 488
Dvir  : EIRTGYSGLNKAIVCDQGSNNRAVFRRWGLVDIKNPSFKVQCKEIEFAIYVPHLVKSLRNILLKNNINITYTFDGVVSWKIVEELVEIDSR-NSTL-LCPK-----LTKKHVSPNCFEKMVKYATQVSSHSVAALRTVVD : 527

```

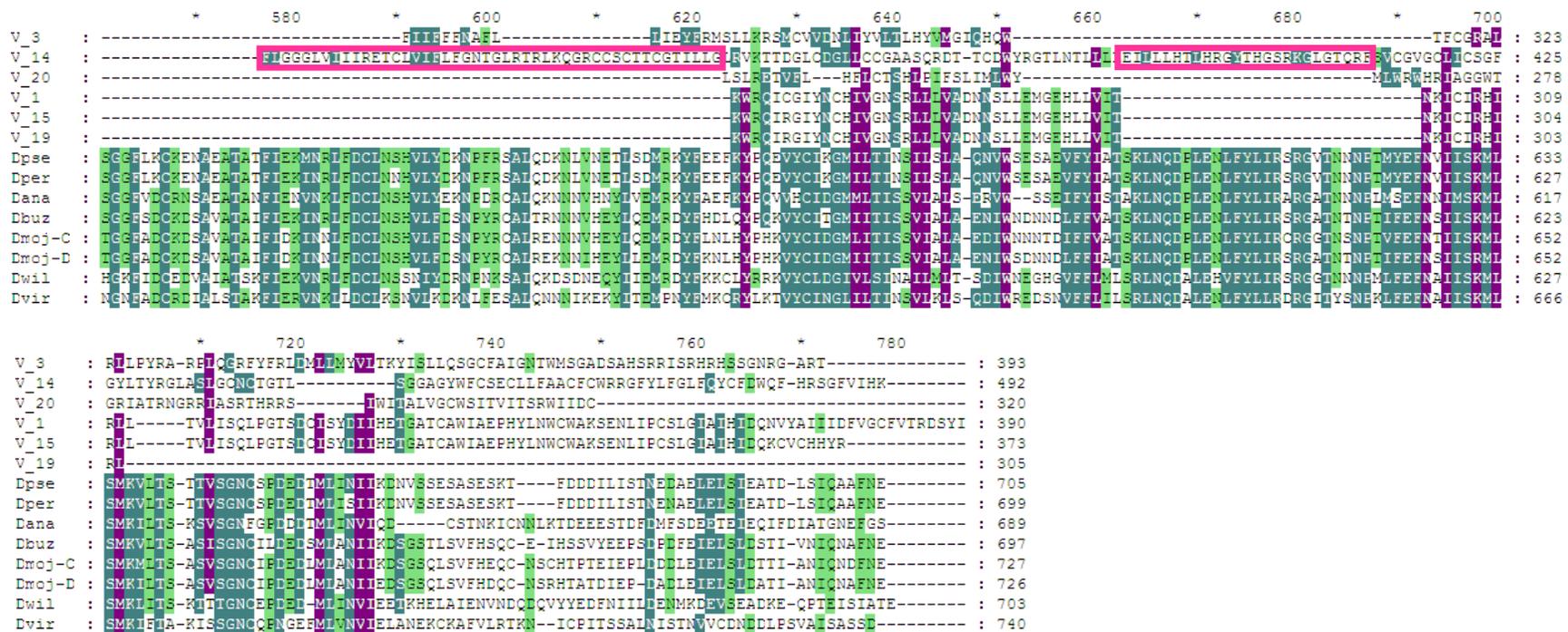


Figura 15. Alinhamento das sequências de aminoácidos entre seis variantes caracterizadas em espécies do subgrupo *willistoni* e sequências identificadas por Marzo *et al.* (2008) em *D. pseudoobscura*, *D. persimilis*, *D. ananassae*, *D. buzzatii*, *D. mojavensis*, *D. willistoni*, e *D. virilis*, respectivamente. V_3: variante 3 (compartilhada por *D. tropicalis*, *D. insularis* e a semiespécie Interior; V_14: variante 14 (isolada a partir de *D. willistoni*); V_20: variante 20 (presente na semiespécie Andino-Brasileira); V_1: variante 1 (compartilhada por *D. tropicalis*, *D. equinoxialis* e a semiespécie Andino-Brasileira; V_15: variante 15 (compartilhada pelas semiespécies Amazônica, Interior e Orinocana); V_19: variante 19 (obtida a partir da semiespécie Amazônica). A linha em vermelho destaca a região correspondente ao motivo THAP, as setas vermelhas os resíduos conservados dentro desta região e as setas pretas os resíduos conservados logo após o término deste motivo de ligação ao DNA. Os retângulos rosas destacam segmentos de DNA exclusivos da variante 14 de *D. willistoni*. Em lilás, azul e verde estão destacados os resíduos com o percentual de 80%, 60% e 40% de conservação, respectivamente. Os valores entre parênteses correspondem ao número de aminoácidos omitidos no início das sequências.

CAPÍTULO V

DISCUSSÃO

Desde a sua identificação em *Drosophila buzzatii* (Cáceres *et al.*, 1999) e depois através da análise *in silico* realizada por Marzo *et al.* (2008) em mais seis espécies de *Drosophila*, *Galileo* vem sendo caracterizado molecularmente apenas na referida espécie (Cáceres *et al.*, 2001; Casals *et al.*, 2003; Casals *et al.*, 2005; Marzo *et al.*, 2008; Delprat *et al.*, 2009), de forma que o presente trabalho foi o primeiro que caracterizou o transposon *Galileo* em populações naturais de espécies neotropicais.

Assim como argumentaram Cáceres *et al.* (2001) e Casals *et al.* (2005), podemos considerar como uma evidência indireta da capacidade de *Galileo* em formar estruturas secundárias quando desnaturado, a dificuldade de isolar um segmento deste transposon, por meio da técnica de *PCR*, no genoma de espécies do grupo *willistoni*. Além disso, o fato de termos encontrado segmentos deletados quando comparamos a região isolada de *Galileo* entre as diferentes linhagens estudadas, pode ser decorrente de eventos de transposição do elemento, mas pode também ser ocasionado pelo deslize da Taq polimerase naquela região durante a síntese de DNA como apontado por Cáceres *et al.* (2001) e Casals *et al.* (2005). Este mecanismo de excisão durante a *PCR* corresponde a deleções espontâneas que podem ocorrer durante a replicação do DNA como consequência do fenômeno conhecido como *slippage* (Farabaugh *et al.*, 1978; Albertini *et al.*, 1982), estimulado pela formação de alças de estruturas secundárias (Egner & Berg, 1981). Desta forma, nossos achados podem ser evidências de que *Galileo* também estaria desempenhando um importante papel na evolução genômica no grupo *willistoni* de *Drosophila*.

Galileo foi isolado a partir de espécies do subgrupo *willistoni*, assim como do subgrupo *bocainensis*, mostrando-se ser antigo no genoma do grupo *willistoni*. É interessante salientar que *D. willistoni* e *D. paulistorum* (Andino-Brasileira) apresentam o maior número de variantes isoladas neste trabalho, sendo que também são as espécies mais polimórficas do subgrupo *willistoni* (Burla *et al.*, 1949; Kastritsis, 1967; Rohde *et al.*, 2006) e que apresentam a maior distribuição geográfica (Spassky *et al.*, 1971 e Ehrman & Powell, 1982). Em contrapartida, a *D. insularis*, que apresenta a distribuição geográfica mais restrita (Ehrman & Powell, 1982) foi a única espécie dentre as investigadas que não tem variante exclusiva, entre as duas encontradas em seu genoma.

A linhagem estudada desta espécie, além disso, é virtualmente monomórfica para inversões paracêntricas (Rohde *et al.*, 2006) e ocupa uma posição basal em relação ao grupo críptico *willistoni* (Gleason *et al.*, 1998; Tarrío *et al.*, 2000; Rohde *et al.*, 2006; Robe *et al.*, 2010).

Estas peculiaridades de *Galileo*, até o momento, não permitem que se estabeleçam hipóteses mais conclusivas sobre a evolução deste elemento nos genomas das espécies do grupo *willistoni*. O quadro que emerge de nossas observações é sugestivo de mais de um evento evolutivo, implicando tanto a manutenção por ancestralidade de certas variantes quanto à ocorrência de perdas estocásticas.

A primeira hipótese pode ser representada pelas variantes 1, 3 e 17. As variantes 3 e 17 de *D. insularis* são compartilhadas ou com *D. equinoxialis* (17) ou com *D. tropicalis* e a semiespécie Interior de *D. paulistorum* (3). Similarmente, a variante 1 é compartilhada por *D. equinoxialis*, *D. tropicalis* e *D. paulistorum* (Andino-Brasileira). O compartilhamento das variantes citadas dentre membros do agrupamento que engloba as crípticas *D. insularis*, *D. tropicalis*, *D. equinoxialis* e *D. paulistorum* é sugestivo de polimorfismo ancestral, que também pode ser evocado para explicar a presença das variantes (1) na semiespécie Andino-Brasileira e (3) na Interior de *D. paulistorum*. Por outro lado, as semiespécies Amazônica e Orinocana de *D. paulistorum* não apresentam nenhuma dessas variantes de *Galileo*, e Andino-Brasileira, não apresenta as variantes 3 e 17, bem como a Interior não possui a 1 e a 17. Considerando que estas variantes estariam presentes nas espécies mais ancestrais, a hipótese de perda estocástica pode ser aventada.

Dentro das semiespécies de *D. paulistorum*, que correspondem à ramificação mais recente, também pode ser proposta ancestralidade, representada pela ocorrência da variante 15, compartilhada pelas semiespécies Amazônica (a mais divergente segundo Spassky *et al.*, 1971; Gleason *et al.*, 1998; Robe *et al.*, 2010) Orinocana e Interior. Também a variante 12 é compartilhada pelas semiespécies Andino-Brasileira e Orinocana, cuja relação filogenética é estritamente mais próxima (Rohde *et al.*, 2006; Robe *et al.*, 2010). Já o compartilhamento da variante 11 por Andino-Brasileira e Interior, não tão próximas filogeneticamente, merece outras considerações.

Se observarmos o mapa da distribuição geográfica (Figura 8) veremos que as semiespécies Amazônica, Andino-Brasileira e Interior são simpátricas em uma ampla faixa territorial que compreende as regiões norte e nordeste do Brasil. Assim, embora mais evidências sejam necessárias, a possibilidade de ocorrência de Transferência

Horizontal (*HT*) não pode ser de todo descartada, uma vez que *Galileo* pertence à superfamília P (Marzo *et al.*, 2008), onde vários eventos de *HT* têm sido registrados (Silva & Kidwell, 2000, Loreto *et al.*, 2008). Da mesma forma, o fenômeno de introgressão tem sido bem-documentado dentro do subgrupo *willistoni* (Ehrman & Powell, 1982; Cordeiro e Winge, 1995; Valiati, 1999; Robe *et al.*, 2010) e pode ter sido o mecanismo pelo qual tenham sido transferidas sequências de *Galileo* entre os táxons considerados.

As espécies não crípticas *D. nebulosa*, *D. sucinea* e *D. capricorni*, apresentam variantes exclusivas de *Galileo*. Seu posicionamento mais basal em relação aos outros membros do grupo críptico pode em parte explicar essa divergência, sem descartar a possibilidade de perdas estocásticas e de surgimento de novas variantes ao longo do tempo, tendo evoluído de forma independente das demais espécies.

A ocorrência de variantes exclusivas de *Galileo* no genoma de *D. willistoni*, nenhuma das quais é compartilhada com suas crípticas, é mais difícil de explicar, já que ela não parece ser a espécie ancestral (Gleason *et al.*, 1998; Tarrío *et al.*, 2000; Rohde *et al.*, 2006; Robe *et al.*, 2010) e que é simpátrica com todas as espécies do subgrupo *willistoni* ao longo da sua distribuição geográfica (Ayala *et al.*, 1970; Spassky *et al.*, 1971; Ehrman & Powell, 1982). A singularidade da *D. willistoni*, entretanto, foi recentemente demonstrada pelos resultados do *Drosophila 12 Genomes Consortium* (2007), que incluem o uso de códon enviesado, a ausência potencial de selenoproteínas, um número maior de rearranjos no *cluster* dos genes *Hox* e inclusive, a segunda maior proporção de elementos transponíveis de várias famílias. Neste contexto, e também considerando que é uma característica do elemento *Galileo* a inserção de outros elementos como, por exemplo, *BuT* (Cáceres *et al.* 2001, ver Figura 3), formando elementos quiméricos, é possível que as diferentes variantes encontradas em *D. willistoni* correspondam a fenômenos deste tipo. A favor desta idéia, podemos ver na Figura 15, que a variante 14 de *D. willistoni*, apresenta claramente segmentos de aminoácidos, não encontrada em nenhuma outra sequência. Seriam eles correspondentes a inserções secundárias de outros elementos dentro de *Galileo*? Análises futuras neste sentido serão importantes para responder tal questionamento e estão entre as perspectivas do grupo de trabalho em curto prazo.

Também merece comentário o achado, ainda que preliminar, de diferença quanto às sequências amplificadas entre populações de *Drosophila willistoni*, representadas pela ocorrência exclusiva, até o momento, do fragmento que gerou a variante 14, nas

populações Cip, Wip e 17A2, não encontrado nas demais 5 populações analisadas desta espécie. Diferenças interpopulacionais também foram registradas para *Galileo* em *D. buzzatii* por Cáceres *et al.* (2001), Casals *et al.* (2003 e 2005). Cabe salientar que essas três populações de *D. willistoni* são muito distantes geograficamente (Cip do Brasil Central, 17A2 do Sul do Brasil e Wip do Nordeste do Brasil), e assim, a variante 14 poderia ser pensada como uma das sequências mais antigas no genoma da *Drosophila willistoni*.

Até o momento nenhum elemento com *ORF* que codificasse uma transposase funcional foi encontrado por nós nem pelos pesquisadores que se dedicaram a estudar *Galileo* em *D. buzzatii* (Cáceres *et al.*, 1999; Cáceres *et al.*, 2001; Casals *et al.*, 2003; Casals *et al.*, 2005; Marzo *et al.*, 2008; Delprat *et al.*, 2009). A estrutura deste transposon vem sendo definida por meio da construção de sequências consensos obtidas a partir de inúmeras variantes ou cópias isoladas de diferentes linhagens de *D. buzzatii*. Os autores mencionados demonstraram que *Galileo* é singular no sentido de que não precisa estar ativo para provocar instabilidade no genoma hospedeiro, mas que sua estrutura molecular permite que ele mesmo como um substrato passivo, seja o indutor de rearranjos cromossômicos. Dessa forma, mesmo que não tenha sido encontradas sequências do segmento da transposase e de seu domínio *THAP* com características que apontassem para a sua atual atividade, cópias defectivas de *Galileo*, assim como atuam em *D. buzzatii*, podem também estar sendo uma fonte endógena de variação para o grupo *willistoni*.

CAPÍTULO VI

PERSPECTIVAS

O panorama de evolução complexa do elemento *Galileo* no grupo *willistoni* de *Drosophila* abre caminho para estudos futuros que devem esclarecer muitas das peculiaridades destas espécies neotropicais, especialmente da singularidade da *Drosophila willistoni* e de sua imensa variabilidade genética e cromossômica.

Durante a realização do presente trabalho, tivemos evidências de que há diferenças em nível interespecífico quanto à presença de variantes de *Galileo*, o que torna promissor o estudo populacional de *Galileo* em espécies do grupo *willistoni*. Também outras espécies como *D. fumipennis* e mais populações de cada uma das espécies, além das semiespécies Transicional e Centro-Americana de *D. paulistorum* deverão ser incluídas em estudo futuro.

Tendo-se evidências indiretas da capacidade de *Galileo* de formar estrutura secundária assim como da participação efetiva na geração de inversões cromossômicas naturais através do mecanismo de recombinação ectópica, será de grande valia a caracterização das IRs das variantes deste transposon encontradas no grupo *willistoni*, que é altamente polimórfico para inversões paracêntricas. A localização *in situ* das variantes de *Galileo* encontradas no presente trabalho está sendo feita pela Mestranda Carolina Flores Garcia, do nosso grupo de pesquisa.

A caracterização molecular das variantes de *Galileo* encontradas neste trabalho deverá ser concluída através da construção de novos *primers*. Neste processo, espera-se detectar a presença de elementos *Galileo* quiméricos se eles forem comuns em *D. willistoni*, como o são em *D. buzzatii*.

REFERÊNCIAS BIBLIOGRÁFICAS

- Adé J and Belzile FJ (1999) Hairpin elements, the first family of foldback transposons (FTs) in *Arabidopsis thaliana*. *Plant J* 19:591-597.
- Albertini AM, Hofer M, Calos MP, and Miller JH (1982) On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell* 29:319-328.
- Ayala FJ (1970) Competition, coexistence and evolution. In: Ashburner M, Carson HL and Thompson Jr JN (eds) *Essays in Evolution and Genetics an Biology of Drosophila*. Academic Press, New York and London, vol 3a, pp 395-429.
- Ayala FJ (1973) Two new subspecies of the *Drosophila willistoni* group. *Pan-Pacific Entom.* 49:273-279.
- Ayala FJ and Tracey ML (1973) Genetic differentiation and reproductive isolation between two subspecies of *Drosophila willistoni*. *J Hered* 64:120-124.
- Ayala FJ, Tracey ML, Hedgecock D and Richmond RC (1974) Genetic differentiation during the speciation process in *Drosophila*. *Evolution* 28:576-592.
- Bächli G (2008) TaxoDros: The Database on Taxonomy of Drosophilidae, <http://www.taxodros.uzh.ch/> (January 13, 2010).
- Bartolomé C, Maside X and Charlesworth B (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol* 19:926-937.
- Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, Williams KP, Sobral BW, Kupko JJ 3rd, Porcella SF, Samuel JE and Heinzen RA (2009) Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus *Coxiella*. *Infect Immun* 77:642-656.
- 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.
- Bingham PM and Zachar Z (1989) Retrotransposons and the FB transposon from *Drosophila melanogaster*. In: Berg DE and Howe MM (eds) *Mobile DNA*. American Society for Microbiology, Washington, pp 485-502.
- Bizzo NMV and Sene FM (1982) Studies on the natural populations of *Drosophila* from Peruipe, SP, Brazil. *Rev Biol* 42:539-544.
- Brinig MM, Cummings CA, Sanden GN, Stefanelli P, Lawrence A and Relman DA. (2006) Significant gene order and expression differences in *Bordetella pertussis* despite limited gene content variation. *J Bacteriol* 188:2375-2382.

- Burla H, DaCunha AB, Cordeiro AR, Dobzhansky T, Malogolowkin C and Pavan C (1949) The *willistoni* group of sibling species of *Drosophila*. *Evolution* 3:300-314.
- Busseau I, Pelisson A and Bucheton A (1989) I elements of *Drosophila melanogaster* generate specific chromosomal rearrangements during transposition. *Mol Gen Genet* 218:222-228.
- Cáceres (2000) Inversiones cromosómicas en *Drosophila*: origen molecular y significado de su tamaño. Ph.D. thesis, Universidad Autònoma de Barcelona, Spain.
- Cáceres M, Ranz JM, Barbadilla A, Long M and Ruiz A (1999) Geration of a widespread *Drosophila* inversion by transposable element. *Science* 285:415-418.
- Cáceres M, Puig M and Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353-1364.
- Capy P, Bazin C, Higuët D and Langin T (1998) *Dinamics and Evolution of Transposables Elements*. 1st edition. Landes Bioscience, France, 197p.
- Carmody G (1965) Two unusual strains of the *D. willistoni* sibling species group. *Drosoph Inf Serv* 40:53.
- Casals F, Cáceres M, and Ruiz A (2003) The Foldback-like transposon Galileo is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. *Mol Biol Evol* 20:674-685.
- Casals F, Cáceres M, Manfrin MH, González J and Ruiz A (2005) Molecular Characterization and chromosomal distribution of Galileo, Kepler and Newton, three Foldback transposable elements of the *Drosophila buzzatii* species complex. *Genetics* 169:2047-2059.
- Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL et al. (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 101:13826-13831.
- Charlesworth B (1996) Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* 68:131-149.
- Cheng CS, Tsuchimoto S, Ohtsubo H and Ohtsubo E (2000) Tnr-8, a foldback transposable element from rice. *Genes Genet Syst* 75:327-333.
- Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius HH, Yin J, et al. (2007) The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci USA* 104:7981-7986.

- Clouaire T, Roussigne M, Ecochard V, Mathe C, Amalric F and Girard JP (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. *Proc Natl Acad Sci USA* 102:6907-6012.
- Collins M and Rubin GM (1984) Structure of chromosomal rearrangements induced by the FB transposable element in *Drosophila*. *Nature* 308:323-327.
- Cordeiro AR (1968) Chromosomal pairing variability of interespecific hybrids of *Drosophila willistoni* cryptic group. *Proc Int Congr Genetics, Tokyo*, 1:191.
- Cordeiro AR, Winge H (1995) Levels of evolutionary divergence of *Drosophila willistoni* sibling species. In: Levine L (Ed) *Genetics of natural populations: the continuing importance of Theodosius Dobzhansky*. Columbia University Press, New York, pp 262-280.
- Daveran-Mingot ML, Campo N, Ritzenthaler P and le Bourgeois P (1998) A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between *Drosophila melanogaster* and *Drosophila subobscura*. *Genetics* 139:321-326.
- Daveran-Mingot ML, Campo N, Ritzenthaler P, Le Bourgeois P (1998) A natural large chromosomal inversion in *Lactococcus lactis* is mediated by homologous recombination between two insertion sequences. *J Bacteriol* 180:4834-4842.
- Delprat A, Negre B, Puig M and Ruiz A (2009) The Transposon *Galileo* Generates Natural Chromosomal Inversions in *Drosophila* by Ectopic Recombination. *PLoS One* 4: e7883.
- Deng W, Burland V, Plunkett G 3rd, Boutin A, Mayhew GF, Liss P, Perna NT, Rose DJ, Mau B, Zhou S et al. (2002) Genome sequence of *Yersinia pestis* KIM. *J Bacteriol* 184:4601-4611.
- Dobzhansky T, Pavan C (1943) Studies on Brazilian species of *Drosophila*. *Bolm Fac Filos Cienc S Paulo* 36: 7-72.
- Dobzhansky T (1946) Complete reproductive isolation between two morphologically similar species of *Drosophila*. *Ecology* 27:205-211.
- Dobzhansky T, Ehrman L and Pavlovsky O (1957) *Drosophila insularis*, a new sibling species of the *willistoni* group. *Univ Texas Publ* 5721:39-47.
- Dobzhansky T and Spassky B (1959) *Drosophila paulistorum* a cluster of species in statu nascendi. *Proc Natl Acad Sci USA* 45:419-428.
- Dobzhansky T (1965). “Wild” and “Domestic” species of *Drosophila*. In Baker HG and Stebbins GL (eds) *The Genetics of Colonizing Species* in. Academic Press, New York pp 533-546.
- Dobzhansky T, Ehrman L, Pavlovsky O and Spassky B (1964) The superespecies *D. paulistorum*. *Proc Natl Acad Sci USA* 51:3-9.

- Dobzhansky T, Pavlovsky O and Ehrman L (1969) Transitional populations of *Drosophila paulistorum*. *Evolution* 23:482-492.
- Dobzhansky T and Powell JR (1975) The *willistoni* group of sibling species of *Drosophila*. In King RC (ed) *Handbook of Genetics* in. Plenum Press, New York pp 589-622.
- Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 480:2003-218.
- Duda O (1925) Die Costaricanischen Drosophiliden (Dipteren) des Ungarischen National Museums zu Budapest. *Ann Mus Nat Hung* 22:149-229.
- Egner C and Berg DE (1981) Excision of transposon Tn5 is dependent on the inverted repeats but not on the transposase function of Tn5. *Proc Natl Acad Sci* 78:459-463.
- Ehrman L and Powell JR (1982) The *Drosophila willistoni* species group. In: Ashburner M, Carson HL, Thompson JN (eds) *The genetics and biology of Drosophila*. Academic Press, New York, pp 193-225.
- Evans AL, Mena PA and McAllister BF (2007) Positive selection near an inversion breakpoint on the neo-X chromosome of *Drosophila americana*. *Genetics* 177: 1303-1319.
- Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikodvorskaja V, Pyatkov KI, Zhivotovsky LA and Kidwell MG (2000). Mobile elements and chromosomal evolution in the virilis group of *Drosophila*. *Proc Natl Acad Sci USA* 97:11337-11342.
- Farabaugh PJ, Schmeissner U, Hofer M, and Miller JH (1978) Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J Mol Biol* 126:847-863.
- Feschotte C and Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 41:331-368.
- Feschotte C (2008) The contribution of transposable elements to the evolution of regulatory networks. *Nat Rev Genet* 9(5):397-405.
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5: 103-107.
- Garcia ACL, Rohde C, Audino GF, Valente VLS, Valiati VH (2006) Identification of the sibling species of the *Drosophila willistoni* subgroup through the electrophoretic mobility of acid phosphatase-1. *J Zool Syst Evol Res* 44:212-216.

- Gleason JM, Griffith EC and Powell JR (1998) A molecular phylogeny of the *Drosophila willistoni* group: conflicts between species concepts? *Evolution* 52:1093-1103.
- Goldman ASH and Lichten M (1996) The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* 144:43-55.
- Gordo I and Charlesworth B (2001) Genetic linkage and molecular evolution. *Curr Biol* 11:R684-686.
- Gray YH, Tanaka MM and Sved JA (1996) P-element-induced recombination in *Drosophila melanogaster*: hybrid element insertion. *Genetics* 144:1601-1610.
- Gray YH (2000) It takes two transposons to tango: transposable-element mediated chromosomal rearrangements. *Trends Genet* 16:461-468.
- Haber JE, Leung WY, Borts RH and Lichten M (1991) The frequency of meiotic recombination in yeast is independent of the number and position of homologous donor sequences: implications for chromosome pairing. *Proc Natl Acad Sci USA* 88:1120-1124.
- Hankeln T and Schmidt ER (1990) New foldback transposable element TFB1 found in histone genes of the midge *Chiromonus thummi*. *J Mol Biol* 215:447-482.
- Harden N and Ashburner M (1990) Characterization of the FB-NOF transposable element of *Drosophila melanogaster*. *Genetics* 126:387-400.
- Hartl DL, Lohe AR and Lozovskaya ER (1997) Modern Thoughts on an ancient marinere: function, evolution, regulation. *Annu Rev Genet* 31:337-358.
- Hasson EC, Rodríguez JJ, Fanara H, Naveira H, Reig OA and Fontdevila A (1995) The evolutionary history of *Drosophila buzzatii*. XXVI. Macrogeographic patterns of inversion polymorphism in New World populations. *J Evol Biol* 8:269-294.
- Hill WG and Robertson A (1966) The effect of linkage on limits to artificial selection. *Gent Res* 8:269-294.
- Hoenigsberg HF, Palomino JJ, Chiappe C, Rojas GG and Cañas BM (1977) Populations genetics in the American tropics. XI Seasonal and temporal variations in relative frequencies of species belonging to the *willistoni* group in Colombia. *Oecologia* 745:1-10.
- Hoffman-Liebermann B, Liebermann D, Kedes LH and Cohen SN (1985) TU elements: a heterogeneous family of modularly structured eukaryotic transposons. *Mol Cell Biol* 5:991-1001.
- Huang JT and Dooner HK (2008) Macrotransposition and other complex chromosomal restructuring in maize by closely linked transposons in direct orientation. *Plant Cell* 20: 2019-2032.

- Jaccard P (1908) Nouvelles recherches sur la distribution florale. v. 44, Bulletin de la Soci t  Vanddoise des Sciences Natureles, 223-270 pp.
- Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M and Celniker SE (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* 3:research0084.1-0084.20
- Kapitonov VV and Jurka J (2003) Molecular paleontology of transposable elements in *Drosophila melanogaster* genome. *Proc Natl Acad Sci USA* 100:6569-6574.
- Kastritsis CD (1967) A comparative study of the chromosomal polymorphs in the incipient of the *Drosophila paulistorum* complex. *Chromosoma* 23:180-202.
- Kastritsis CD and Dobzhansky (1967) *Drosophila pavlovskiana* a race of a species? *Amer Midl Nat* 78:244-247.
- Kazazian HH (2004) Mobile elements: drivers of genome evolution. *Science* 303:1626-1632.
- Kidwell MG and Lisch D (1997) Transposable elements as sources of variation in animals and plants. *Proc Natl Acad Sci USA* 94:7704-7711.
- Kidwell MG and Holyoake AJ (2001) Transposon-Induced Hotspots for Genomic Instability. *Genome Research* 11:1321-1322.
- Kim JM, Vanguri S, Boeke JD, Gabriel A and Voytas DF (1998) Transposables elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8:464-478.
- Langley CH, Montgomery E, Hudson R, Kaplan N and Charlesworth B (1988) On the role of unequal exchange in the containment of transposable element copy number. *Genet Res* 52:233-235.
- Laski FA, Rio DC and Rubin GM (1986) Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell* 44:7-19.
- Lee CC, Beal EL, and Rio DC (1998) DNA binding by the KP repressor protein inhibits P-element transposase activity in vitro. *EMBO J* 17:4166-4174.
- Lee J, Han K, Meyer TJ, Kim HS, Batzer MA (2008) Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE* 3:e4047.
- Levis R, Collins M and Rubin GM (1982) FB elements are the common basis for the instability of the w^{DZL} and w^C *Drosophila* mutations. *Cell* 30:551-565.

- Liebermann D, Hoffman-Liebermann B, Weinthal J, Childs G, Maxson R, Mauron A, Cohen SN and Kedes L (1983) An unusual transposon with long terminal inverted repeats in the sea urchin *Strongylocentrotus purpuratus*. *Nature* 306:342-347. |
- Lim JK (1988) Intrachromosomal rearrangements mediated by hobo transposons in *Drosophila melanogaster*. *Proc Natl Acad Sci* 85:9153-9157.
- Lim JK and Simmons MJ (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. *Bioessays* 16:269-275.
- Lyttle TW and Haymer DS (1992) The role of transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. *Genetica* 86:113-126.
- Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA and Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* 148:1507-1524.
- Loreto ELS, Carareto CMA and Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 1-10.
- Malogolowkin C (1952) Sobre a genitália dos Drosophilidae (Diptera). III. Grupo *willistoni* do gênero *Drosophila*. *Rev Bras Biol* 12: 79-96.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. v.27, n.2, *Cancer Research*, Chestnut, 209-220 pp.
- Marin I and Fontdevila A (1995) Characterization of Gandalf, a new inverted repeat transposable element of *Drosophila koepferae*. *Mol Gen Genet* 248:423-433.
- Marques EK, Winge H and Cordeiro AR (1966) A corn meal, soybean flour, wheat germ medium for *Drosophila*. *DIS* 41:187.
- Martins, M. (1987) Variação espacial e temporal de algumas espécies e grupos de *Drosophila* (Díptera) em duas reservas de mata isoladas, nas vizinhanças de Manaus (Amazonas, Brasil). *Bol Mus Par Emílio Goeldi, Sér. Zool.* 3:107-108.
- Marzo M, Puig M and Ruiz A (2008) The *Foldback*-like element *Galileo* belongs to the *P* superfamily of DNA transposons and is widespread within the *Drosophila* genus. *PNAS* 105 (8):2957-2962.
- Maside X, Bartolomé C, Assimacopoulos S and Charlesworth B (2001) Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: in situ hybridization vs Southern blot data. *Genet Res* 78:121-136.
- McDonald JF (1993) Evolution and consequences of transposable elements. *Curr Opin Genet Des* 3:855-864.

- Montgomery E, Huang S-M, Langley CH and Judd BH (1991) Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*. *Genetics* 129:1085-1098.
- Nassif N and Engels W (1993) DNA homology requirements for mitotic gap repair in *Drosophila*. *Proc Natl Acad Sci USA* 90:1262-1266.
- Negre B, Ranz JM, Casals F, Caceres M and Ruiz A (2003) A new split of the Hox gene complex in *Drosophila*: relocation and evolution of the gene labial. *Mol Biol Evol* 20:2042-2054.
- Negre B, Casillas S, Suzanne M, Sanchez-Herrero E, Akam M, Nefedov M, Barbadilla A, de Jong P and Ruiz A. (2005) Conservation of regulatory sequences and gene expression patterns in the disintegrating *Drosophila* Hox gene complex. *Genome Res* 15:692-700.
- O'Grady PM and Kidwell MG (2002) Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Mol Phylogenet Evol* 22:442-453.
- Page RD and Charleston MA (1998) Trees within trees Phylogeny and historical associations *Trends Ecol Evol* 13:356-359.
- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebahia M, James KD, Churcher C, Mungall KL et al. (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* 413:523-527.
- Parkhill J, Sebahia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT, Churcher CM, Bentley SD, Mungall KL et al. (2003) Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 35:32-40.
- Patterson JT and Mailand GB (1944) Studies in the genetics of *Drosophila*. IV. Papers dealing with taxonomy, nutrition, cytology and interspecific hybridization in *Drosophila*. I. The Drosophilidae of México. *The Univ Texas Publ* 4445:9-101.
- Petes TD and Hill CW (1988) Recombination between repeated genes in microorganisms. *Annu Rev Genet* 22:147-168.
- Picologlou S, Digi ME, Kovarik P, Liebman SW (1988) The same configuration of Ty elements promotes different types and frequencies of rearrangements in different yeast strains. *Mol Gen Genet* 211:272-281.
- Potter SS, Truet M, Phillips M and Maher A (1980) Eukaryotic transposable genetic elements with inverted terminal repeats. *Cell* 20:639-647.
- Powell JR, Sezzi E, Moriyama EN, Gleason JM and Caccione A (2003) Analysis of a shift in codon usage in *Drosophila*. *J Mol Evol* 57:214-225.

- Quesneville H, Nouaud D and Anxolabehere D (2005) Recurrent recruitment of the THAP DNA-binding domain and molecular domestication of transposable element. *Mol Biol Evol* 22:741-746.
- Raghava GPS (1994) Improved estimation of DNA fragment length from gel electrophoresis using a graphical method. *Biotechniques* 17:100-104.
- Rebatchouk D and Narita JO (1997) Foldback transposable elements in plants. *Plant Mol Biol* 34:831-835.
- Redder P, Garrett RA (2006) Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J Bacteriol* 188:4198-4206.
- Regner L, Pereira MSO, Alonso CEV, Abdelhay E and Valente VLS (1996) Genomic distribution of *P* element in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. *J Hered* 87:191-198.
- Reiss D, Quesneville H, Nouaud D, Andrieu O and Anxolabehere D (2003) Hoppel, a P-like element without introns: A P-element ancestral structure or a retrotranscription derivative? *Mol Biol Evol* 20:869-879.
- Reith ME, Singh RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, Munholland J, Murphy C, Sarty D, Williams J et al. (2008) The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics* 9:427.
- Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP et al. (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res* 15:1-18.
- Richardson C, Moynahan ME and Jasin M (1998) Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev* 12:3831-3842.
- Ritchie MG and Gleason JM (1995) Rapid evolution of courtship song pattern in *Drosophila willistoni* sibling species. *J Evol Biol* 8:463-479.
- Robe LJ, Cordeiro J, Loreto ELS and Valente VLS (2010) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genética* DOI 10.1007/s10709-009-9432-5.
- Rohde C, Garcia ACL, Valiati VH and Valente VLS (2006) Chromosomal evolution of sibling species on the *Drosophila willistoni* group. I. Chromosomal arm IIR (Muller's element B). *Genetica* 126:77-88.
- Roeder GS (1983) Unequal crossing-over between yeast transposable elements. *Mol Gen Genet* 190:117-121.

- Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software, New York.
- Roussigne M, Kossida S, Lavigne A-C, Clouaire T, Ecochar V, Glories A, Amalric F and Girard J-P (2003) The THAP domain: a novel protein motif with similarity to the DNA-binding domain of *P* element transposase. *Trends Biochem Sci* 28:66-69.
- Ruiz A, Naveira H and Fontdevila A (1984) La historia evolutiva de *Drosophila buzzatii*. IV. Aspectos citogenéticos de su polimorfismo cromosómico. *Genét Ibér* 36:13-35.
- Ruiz A and Wasserman M (1993) Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. *Heredity* 70 (Pt 6): 582-596.
- Santos RA and VLS Valente (1990). On the occurrence of *Drosophila paulistorum* Dobzhansky & Pavan (Diptera, Drosophilidae) in an urban environment ecological and cytogenetic observations. *Evolución Biológica* 4: 253-268.
- Sassi AK Herádia 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.
- Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Distche C, McGillivray B, de la Chapele A and Page DC (1998) Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum Mol Genet* 7:1-11.
- Sene, FM, Val FC, Vilela CR and Pereira MAQR (1980) Preliminary date on the geographical distribution of *Drosophila* species within morphoclimatic domains of Brazil. *Papéis Avulsos Zool., S.Paulo*, 33:315-326.
- Silber J, Bazin C, Lemeunier F, Aulard S and Volovitch M (1989) Distribution and conservation of the foldback transposable element in *Drosophila*. *J Mol Evol* 28:220-224.
- Silva JC, Loreto EL and Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-71.
- Silva JC and Kidwell MG (2000) Horizontal transfer and selection in the evolution of P elements. *Mol Biol Evol* 17:1542-1557.
- Simmen MW and Bird A (2000) Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. *Mol Biol Evol* 17:1685-1694.
- Smith PA and Corces VG (1991) *Drosophila* transposable elements: mechanisms of mutagenesis and interactions with the host genome. *Adv Genet* 29:229-300.

- Sneath PHA and Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco, 573pp.
- Sokal RR and Rohlf FJ (1962) The comparison of dendrograms by objective methods. v.11, Taxon, Berlin, 30-40 pp.
- Spassky B (1957) Morphological differences between sibling species of *Drosophila*. Univ Texas Publ 5721:48-61.
- Spassky B, Richmond RC, Perez-Salas S, Pavlovsky O, Mourão CA, Hunter AS, Hoentgsberg H, Dobzhansky T and Ayala FJ (1971) Geography of the sibling species related to *Drosophila willistoni*, and of the semispecies of the *Drosophila paulistorum* complex. Evolution 25:129-143.
- Spieth HT (1949) Sexual behavior and isolation in *Drosophila*. II. The interespecific mating behavior of species of the *willistoni* group. Evolution 3:67-81.
- Sturtevant AH (1916) Notes on North American Drosophilidae with descriptions of twenty-three new species. Ann Entomol Soc Am 9:323-343.
- Sturtevant AH (1921) The north American species of *Drosophila*. Canergie Inst Wash Publi 301:1-150.
- Svoboda YH, Robson MK and Sved JA (1995) P-element-induced male recombination can be produced in *Drosophila melanogaster* by combining end-deficient elements in trans. Genetics 139:1601-1610.
- Tarrío R, Rodriguez-Trelles F and Ayala FJ (2000) Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila* saltans and willistoni groups, a case study. Mol Phylogenet Evol 16:344-349.
- Templeton NS and Potter SS (1989) Complete foldback transposable elements encode a novel protein found in *Drosophila melanogaster*. Embo J 8:1887-1894.
- Townsend JL (1954) Cryptic subspeciation in *Drosophila* belonging to the subgenus *Sophophora*. Am Nat 88:339-351.
- Val FC, Marques MD, Vilela CR, (1981) Drosophilidae of Neotropical region. In: Ashburner M., Carson HL and Thompson JN (eds.) The Genetics and Biology of *Drosophila*. Academic Press, Orlando, pp. 123-168.
- Valente VLS, Sarraveda C, Morales N and Araújo AM (1981) Observations on the attraction of *Drosophila* species for different baits and chromosomal polymorphism in *Drosophila willistoni*. DIS 56:147-149.
- Wessler SR, Bureau TE and White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814-821.

- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhou B, Flavell A, Leroy P, Morgante M, Panaud O *et al.* (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8:973-982.
- Windsor AJ and Waddel CS (2000) FARE, a new family of foldback transposons in *Arabidopsis*. *Genetics* 156:1983-1995.
- Winge H (1965) Interspecific hybridisation between the six cryptic species of willistoni group. *Heredity* 20:9-19.
- Winge H (1971) Níveis de Divergência Evolutiva. Ph.D. Thesis, Universidade Federal do Rio Grande do Sul. Porto Alegre, Brazil.
- Winge H and Cordeiro AR (1960) Descoberta de uma ponte biológica entre *Drosophila willistoni* e *paulistorum* (Summ). *Ciência e Cultura* 12:149.
- Winge H and Cordeiro AR (1963) Experimental hybrids between *Drosophila willistoni* Stutervant and *Drosophila paulistorum* Dobzhansky and Pavan from Southern marginal populations. *Heredity* 18:215-222.
- Yuan J, Finney M, Tsung N and Horvitz R (1991) Tc4 *Caenorhabditis elegans* transposable element with an unusual fold-back structure. *Proc Natl Acad Sci USA* 88:3334-3338.
- Zelentsova H, Poluectova H, Mnjoian L, Lyozin G, Veleikodvorkaja, Pyatkov K I, Zhivotovsky LA and Kidwell MG (1999) Distribution and evolution of mobile elements in the *virilis* species group of *Drosophila*. *Chromosoma* 108:443-456.
- Zhang J, Yu C, Pulletikurti V, Lamb J, Danilova T, Weber DF, Birchler J and Peterson T (2009) Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. *Genes Dev* 23:755-765.
- Zhou Z-H, Akgün E and Jasin M (2001) Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. *Proc Natl Acad Sci USA* 98:8326-8333.