

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE GENÉTICA

Caracterização e análise evolutiva do elemento *But2* no gênero *Drosophila*

Dirleane Ottonelli Rossato

Trabalho de Conclusão apresentado como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas pelo curso de Ciências Biológicas da Universidade Federal do Rio Grande do Sul

Orientadora: Dra. Vera Lúcia da Silva Valente Gaiesky
Co-orientadora: Dra. Adriana Ludwig

Colaboradores: Maríndia Deprá, Élgion Lúcio da Silva Loreto, Alfredo Ruiz

Porto Alegre
2010/II

Este trabalho foi realizado no Laboratório de *Drosophila* do Departamento de Genética da UFRGS, Porto Alegre, RS, Brasil, com bolsa PIBIC/CNPq e financiamento do CNPq e da PROPESQ-UFRGS.

AGRADECIMENTOS

A minha orientadora, Professora Vera, obrigada por ter me acolhido no seu laboratório e sempre ter me proporcionado as condições necessárias para eu poder realizar esse trabalho. Pela orientação, apoio e incentivo de sempre.

A minha co-orientadora Adriana, obrigada pela enorme dedicação que sempre teve em me orientar, compartilhando comigo o seu conhecimento, mesmo quando estive ausente fisicamente. Pela paciência e empenho de todas as horas. Por confiar em minha capacidade e ser um exemplo de profissional. Tive muita sorte de ter sido orientada por você.

A Maríndia que me ‘adotou’ com bastante responsabilidade. Obrigada pela disposição de sempre, foi um prazer ter trabalhado contigo.

Ao Professor Élgion, que me apresentou o mundo dos elementos de transposição, quando estudei na UFSM. Obrigada por ter colaborado com esse estudo e ter gerado ‘elementos de transposição autônomos’.

Ao Professor Ruiz pelas análises detalhadas e sugestões minuciosas que conferiram a este trabalho ficar mais robusto. Por compartilhar o seu conhecimento de uma forma agradável e com bastante disciplina.

Aos colegas do Laboratório de *Drosophila* da UFRGS, Adriana, Maríndia, Nina, Hermes, Marília, Juliana WG, Mário, Juliana C, Carol, Giseli, Aline, Gilberto, Lucas, Cleverton, Gabi, Bibi, pelo apoio e momentos de descontração que tornaram as horas no lab mais agradáveis.

A UFSM e UFRGS, universidades públicas e de qualidade que contribuíram muito para a construção do meu conhecimento. Aos professores, mestres que me passaram além do conhecimento, valores.

Aos meus ex-colegas da UFSM que dividiram comigo, em um ano, momento inesquecíveis em especial a Marcela, Sinara, Aninha, Elenize, Dani, Marcelinho, as Brunas, Greice e Francis. Ter conhecido vocês fez valer a pena eu ter passado esse tempinho em Santa Maria.

A Kátia e Simone, com quem morei durante dois anos em Santa Maria, ter morado com vocês foi uma experiência muito boa, obrigada pela amizade.

Aos queridos colegas da UFRGS que se tornaram grandes amigos Bibiane, Andressa e Karina, obrigada pelos momentos descontraídos e divertidos que passamos juntas.

As minhas grandes amigas de Porto Alegre, Fernanda, Luciara, Cheila, Anelise, Miche, obrigada meninas pela amizade sem hora e por tornarem os meus dias em Porto Alegre maravilhosos.

Aos meus pais, pelo amor incondicional e pelos valores de vida que me passaram. Em especial a minha mãe pelo exemplo de luta, dedicação e afetividade, você é um exemplo de vida para mim. Obrigada pelo seu alto astral. Ao meu pai, obrigada pela amizade, companheirismo e incentivo de sempre.

Aos meus irmãos, pela amizade incondicional, conversas, brincadeiras e parceria de sempre. Mano, obrigada pelo carinho e atenção que sempre dedica a mim, mesmo estando distante fisicamente. Mana, obrigada pelo seu afeto e pela companhia de todos os dias, é muito bom poder morar contigo.

Aos meus tios, em especial ao meu padrinho Silvestre, pelas boas palavras de sempre. Aos elogios e agrados que mechem de uma forma boa na minha essência.

Às instituições de fomento, em especial ao CNPq pela bolsa PIBIC/CNPq e a PROPESQ-UFRGS pelo auxílio financeiro durante as viagens para eventos científicos, sem os quais seria inviável participar.

Enfim, quero agradecer pela vida. Por ter tido a oportunidade de conhecer pessoas maravilhosas durante toda essa jornada. Por ter a oportunidade de trabalhar com uma equipe eficiente e amiga. Por ter estudado sempre em escola pública e de qualidade. A todos os bons acasos da minha vida, muito obrigada.

Caracterização e análise evolutiva do elemento *But2* no gênero *Drosophila*

Dirleane O. Rossato¹, Adriana Ludwig^{2,3}, Maríndia Deprá², Élgion L. S. Loreto^{2,4},
Alfredo Ruiz⁵ e Vera L. S. Valente^{2,6}

¹ *Curso de Ciências Biológicas, Universidade Federal do Rio Grande do Sul (UFRGS), CEP: 91501-970. Porto Alegre, Rio Grande do Sul, Brasil. Fone: (51) 3308-6713. E-mail: dirleane.ottonelli@gmail.com*

² *Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil.*

³ *Instituto Carlos Chagas (ICC) – Fiocruz - PR, Curitiba, Paraná, Brasil.*

⁴ *Departamento de Biologia, Universidade Federal de Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brasil.*

⁵ *Departament de Genética i de Microbiologia. Facultat de Biociència Universitat Autònoma de Barcelona, Barcelona, Spain.*

⁶ *Departamento de Genética, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil.*

Manuscrito em preparação a ser enviado para a revista “Molecular Genetics and Genomics” com modificações das normas para tornar a leitura mais agradável.

ABSTRACT

The element *But2* belongs to the *hAT* superfamily of DNA transposons discovered and isolated in chromosomal breakpoints in *Drosophila buzzatii*. In this work, we study the distribution and evolution of the element *But2* in 78 insect species (64 species of the genus *Drosophila*) by in silico searches and molecular approaches. Our data revealed that the *But2* is possibly involved in cases of horizontal transfer among species of the *saltans*, *willistoni*, *repleta* groups and the species *D. pallidipennis*. Through in silico PCR, cloning and sequencing approaches, we confirmed the presence of short sequences related to *But2*, that contain similar 5' and 3' regions, including the inverted terminal repeats. These sequences appear to be MITEs (Miniature Inverted Repeat Transposable Elements) and show evidences of recent mobilization. We observed that the complete sequences of *But2* potentially encode a functional transposase. In this work, we obtained the first results on distribution and evolution of *But2* in the genus *Drosophila*. The study of *But2* correspond to a good opportunity to investigate the behavior of transposable elements in the genomes, the evolutionary consequences of the coexistence of complete copies and MITEs, and the impact of the presence of MITEs in the *Drosophila* genomes.

Key words: MITEs, *But2*, Transposable elements, Horizontal transfer, *Drosophila* genus.

RESUMO

O elemento *But2* pertence à superfamília *hAT* de transposons de DNA e foi descoberto em pontos de quebra cromossômica em *Drosophila buzzatii*. Nesse trabalho, nós realizamos um estudo da distribuição e evolução do elemento *But2* em 78 espécies de insetos, das quais 64 pertencem ao gênero *Drosophila*. Nosso estudo consiste na utilização de buscas *in silico* e de abordagens moleculares. Nossos dados revelam que *But2* está possivelmente envolvido em casos de transferência horizontal entre espécies dos grupos *saltans*, *willistoni* e *repleta* e também *D. pallidipennis*. Por *PCR in silico* e clonagem e sequenciamento nós confirmamos a presença de sequências curtas relacionadas ao *But2*, as quais contêm regiões 5' e 3' similares, incluindo as repetições terminais invertidas. Essas sequências parecem ser elementos do tipo MITEs (*Miniature Inverted Repeat Transposable Elements*) e mostram evidências de mobilização recente. Nós observamos que sequências completas de *But2* potencialmente codificam uma transposase funcional. Nesse trabalho, obtivemos os primeiros resultados sobre a distribuição e a evolução do elemento *But2* no gênero *Drosophila*. O estudo do elemento *But2* é uma boa oportunidade para investigar o comportamento dos elementos de transposição nos genomas, as consequências da coexistência de cópias completas e MITEs e o impacto da presença de MITEs nos genomas de *Drosophila*.

Palavras-chave: MITEs, *But2*, Elementos de Transposição, Transferência Horizontal, Gênero *Drosophila*

INTRODUÇÃO

Barbara McClintock nos anos 40 descobriu os elementos de transposição (TEs - *transposable elements*), que são segmentos de DNA que possuem a capacidade de mover-se e replicar-se dentro do genoma. Os TEs estão presentes em praticamente todos os organismos, podendo compreender a maior parte do DNA medianamente repetitivo dos genomas (Capy et al. 1998).

Duas principais classes de TEs são amplamente reconhecidas: os retrotransposons (classe I) que se transpõem por um intermediário de RNA e os transposons (classe II) que se movem pelo uso de DNA como intermediário direto da transposição (Finnegan 1989). Tanto os elementos de classe I quanto os da classe II podem ser elementos autônomos ou não-autônomos. Somente os elementos autônomos possuem todas as sequências que codificam as proteínas essenciais para sua transposição. No entanto, elementos não-autônomos podem ser mobilizados pelas atividades enzimáticas providas por elementos autônomos (Kidwell e Lisch 2001).

Um grupo particular de elementos não autônomos da classe II são os MITEs (*Miniature Inverted Repeat Transposable Elements*), os quais são caracterizados por possuírem um pequeno comprimento, repetições terminais invertidas (TIRs - *Terminal Inverted Repeats*), alto número de cópias e uma região interna rica em AT e são frequentemente encontrados dentro ou próximo a genes (Feschotte et al. 2002). Os MITEs têm sido extensivamente estudados em plantas, em particular no genoma do arroz (González e Petrov 2009).

Os TEs são transmitidos verticalmente de geração para geração juntamente com o restante do genoma. Entretanto, análises da distribuição destes elementos em diferentes espécies e a discrepância entre filogenias baseadas em sequências de TEs e filogenias de acordo com caracteres clássicos sugerem a possível existência de transferência horizontal (TH) de TEs, ou seja, sua transmissão entre diferentes espécies, por meio não sexual (Loreto et al. 2008).

Numerosos casos de TH de TEs entre espécies de *Drosophila* têm sido reportados envolvendo tanto elementos de classe I quanto elementos da classe II (Loreto et al. 2008, Ludwig et al. 2008, Mota et al. 2010, Deprá et al. 2010). Além disso, recentemente, Gilbert et al. (2010) reportaram a ocorrência de TH entre animais de diferentes filos (Chordata, Arthropoda, Mollusca, e Platyhelminthes), para os elementos SPIN, hAT1, OC1 e ET.

Os TEs são apontados como geradores de variabilidade genética e estão associados com a reestruturação dos genomas ao promoverem rearranjos cromossômicos. Diferentes autores têm mostrado que pontos de quebra de inversão cromossômica são frequentemente associados com sítios de inserção de TEs (Lyttle e Haymer 1992; Regner et al. 1996; Andolfatto et al. 1999; Evgen'ev et al. 2000; Cáceres et al. 1999 e 2001). Cáceres et al. (2001) caracterizaram dois pontos de quebra cromossômica em *D. buzzatii*, os quais têm acumulado inserções de diferentes TEs. Dentre esses TEs está o *But2*, que foi incluído na superfamília *hAT* de transposons da classe II baseando-se na similaridade da sequência (Calvi et al. 1991). *But2* mostrou similaridade com o elemento *Gandalf* de *D. koepferae* (Marin e Fontevila 1995).

O elemento *But2* possui 2775 pb, TIRs de 12 pb e gera duplicação no sítio alvo (TSD - *Target Site Duplication*) de 8 pb, pertencendo a ordem TIR de transposons da classe II. Elementos dessa ordem usualmente possuem repetições terminais invertidas curtas e codificam uma transposase que catalisa a sua excisão do sítio original e promove sua reinserção em um novo lugar no genoma, gerando TSDs (Wicker et al. 2007).

Nesse trabalho, nós procuramos contribuir para o conhecimento da distribuição e evolução do elemento *But2* realizando um estudo em 78 espécies de insetos, das quais 64 pertencem ao gênero *Drosophila*. Nosso estudo consiste na utilização de uma abordagem de buscas *in silico* e de abordagens moleculares.

MATERIAIS E MÉTODOS

Manutenção dos estoques de moscas e manipulação de ácidos nucléicos

As moscas estão sendo mantidas em meio de cultura a base de farinha de milho e a temperatura constante (20°C). O DNA genômico total foi extraído de moscas adultas de acordo com Sassi et al. (2005). Um total de 67 espécies (Tabela 1) pertencentes aos gêneros *Drosophila*, *Zaprionus* e *Scaptodrosophila* foram analisadas.

Buscas *in silico*

Inicialmente, nós investigamos a presença de sequências homólogas ao *But2* nos 12 genomas de *Drosophila* sequenciados: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi* e também em 9 genomas de outros insetos: *Culex pipiens*, *Aedes aegypti*, *Anopheles gambiae*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Nasonia vitripennis*, *Acyrtosiphon pisum*, *Pediculus humanus corporis*. Esses genomas estão depositados no banco de dados do Flybase (<http://flybase.bio.indiana.edu/blast/>; Grumblin e Strelets 2006). Nessas buscas utilizamos como sequência de referência (*query*) a sequência nucleotídica de *But2* de *D. buzzatii* de 2775 pb (GenBank AF368884.1) através da ferramenta BLASTn e tBLASTx. Consideramos significativas as sequências (*hits*) que apresentam E-value $\leq 10^{-20}$.

Busca pela região da transposase

A fim de verificarmos se as cópias completas do elemento *But2* teriam uma sequência potencialmente codificadora de uma transposase funcional, utilizamos os programas FGENESH (Yao et al. 2005) e GENESCAN (Burge e Karlin 1997) para prever a existência de possíveis íntrons. As buscas foram realizadas nas sequências completas do elemento *But2* encontradas em *D. buzzatii* (Cáceres et al. 2001), *D. mojavensis* (scaffold_3367:3535-7555) e *D. willistoni* (scf2_1100000004958:2664879-2668034). A ferramenta SMART (Letunic et al. 2009) foi utilizada para buscar domínios nas sequências de proteínas preditas.

Rastreamento da distribuição de *But2* por PCR

Com o objetivo de analisar a distribuição de *But2* em um maior número de espécies pertencentes ao gênero *Drosophila* empregamos a técnica de PCR em 67 espécies (Tabela 1). Quatro diferentes oligonucleotídeos iniciadores (*primers*) foram projetados, os quais quando combinados podem gerar quatro fragmentos com tamanhos distintos (Fig. 1).

Os *primers* But2_S e But2_AS foram desenhados a partir de três sequências: a sequência canônica do elemento *But2* de *D. buzzatii* (GenBank AF368884.1), a sequência de *But2* localizada no scf2_1100000004958:2664879-2668034 de *D. willistoni*, e a sequência de *But2* localizada no scaffold_3367:3535-7555 de *D. mojavensis*. Esses *primers* foram projetados para a amplificação do elemento *But2* completo. A sequência nucleotídica dos *primers* degenerados But2_S e But2_AS são, respectivamente, 5'CAGTGCTGCCAACAWTTYGT3' e 5'CASTGCTGCCAATTTAGCYA3'.

Já os *primers* But2C_S e But2C_AS foram projetados a partir de duas sequências nucleotídicas: a sequência do elemento *But2* canônico de *D. buzzatii* e a sequência de *But2* localizada em scf2_1100000004958:2664879-2668034 de *D. willistoni*. A sequência nucleotídica dos *primers* degenerados de But2C_S e But2C_AS são, respectivamente, 5'AGACYTCGGGRACAGTTTTGC3' e 5'AGCATTAAATGCYAARCTTTC3'.

Usamos o seguinte protocolo para as reações de PCR: 50 ng de DNA genômico adicionado a uma solução de 2.5 mM de MgCl₂, 5 µl de tampão 10X, 200 µM de cada desoxinucleotídeo, 20 picomoles de cada *primer* e 1 U de Taq polimerase em 50 µl de volume total. As condições de reação foram 96°C por 2 minutos, seguidos de 30 ciclos de 96°C por 30 segundos, de 50 a 55,5 °C por 45 segundos, dependendo dos *primers* a serem anelados, 72 °C por 1 a 3 minutos, dependendo do tamanho do fragmento esperado.

PCR *in silico*

PCR *in silico* nos genomas de *D. willistoni* e *D. mojavensis* foi realizado utilizando a ferramenta BLASTn (<http://flybase.bio.indiana.edu/blast/>; Grumbling e Strelets 2006), usando como *query* uma sequência formada a partir da sequência do *primer* But2_S seguida pela complementar reversa da sequência do *primer* But2_AS.

Os *hits* que visivelmente continham ambas as regiões desses *primers*, que correspondem em parte às TIRs do elemento *But2*, foram analisados quanto à presença e conservação de TIRs e TSDs.

Dot blot

Aproximadamente 1µg de DNA, desnaturado por calor foi aplicado diretamente sobre uma membrana de náilon (Hybond-N+ – GE Healthcare). Foram analisadas 60 espécies, as quais podem ser encontradas na Tabela 1. A hibridação e a detecção seguiram o protocolo do *kit* CPD-Star Detection Module (GE Healthcare). A marcação da sonda foi feita com o *kit* Gene Images AlkPhos Direct Labelling Module (GE Healthcare). A sonda utilizada foi o fragmento de PCR amplificado a partir de um clone de *D. willistoni* (Bf2_will1). Este clone corresponde ao fragmento 2 mostrado na Fig. 1. A temperatura de hibridação foi de 55 °C.

Clonagem e sequenciamento

Foram clonados e sequenciados os amplicons correspondentes aos fragmentos 1, 2 e 3 (Fig. 1). Os amplicons clonados e sequenciados de cada fragmento correspondem às seguintes espécies: Fragmento 1: *D. ornatifrons*, *D. subbadia*, *D. pallidipennis*, *D. buzzatii*, *D. saltans*, *D. sucinea*, *D. nebulosa*, *D. paulistorum*, *D. tropicalis*; Fragmento 2: *D. willistoni*, *D. buzzatii* e *D. pallidipennis*; Fragmento 3: *D. incompta*, *D. prosaltans*, *D. saltans*, *D. sturtevanti* e *D. willistoni*.

Os amplicons foram submetidos à reação de clonagem utilizando o Kit TOPO TA[®] (Invitrogen), seguindo protocolo do fabricante. Alguns clones foram sequenciados a partir de produtos de PCR, os quais foram purificados com *Exonuclease I* (USB) e *Shrimp Alkaline Phosphatase* (USB) em sequenciador automático MegaBace 500 e MacroGen Advancing through Genomics (www.macrogen.com).

Análises das sequências

As sequências de nucleotídeos foram alinhadas pelo programa Muscle (Edgar 2004). Distância p foi calculada para *But2* e para os genes nucleares *alpha methyl dopa* (*Amd*) e *alcohol dehydrogenase* (*Adh*) usando o programa MEGA 4 (Tamura et al. 2007). As sequências dos genes *Amd* e *Adh* foram obtidas através do Genbank.

Transmissão vertical (TV) pode ser assumida se a distância de *But2* for maior ou igual à distância do gene nuclear. Por outro lado, se a distância de *But2* for menor do que a do gene nuclear, TH pode ser sugerida. Um teste de qui-quadrado foi empregado para verificar se a divergência observada para *But2* é significativamente diferente da divergência esperada com base nos genes nucleares *Adh* ou *Amd*.

A filogenia de *But2* foi inferida por dois métodos: Neighbor-Joining (NJ), utilizando o modelo evolutivo Kimura2-Parâmetros, e Máxima Parcimônia (MP). Ambos implementados no programa MEGA 4 (Tamura et al. 2007). Bootstrap de 1000 réplicas foi utilizado para acessar a confiabilidade dos ramos em ambos os métodos.

RESULTADOS

Busca *in silico*

Nós investigamos a presença de sequências homólogas ao *But2* nos 12 genomas de *Drosophila* e em genomas de nove outros insetos, disponíveis no Flybase (Tabela 2). Encontramos, por BLASTn, *hits* significativos apenas em *D. mojavensis* e *D. willistoni*.

Em *D. mojavensis*, espécie pertencente ao grupo *repleta* junto com *D. buzzatii*, encontramos apenas três *hits* significativos. Uma cópia está localizada no scaffold_3367:3535-7555 e possui 4020 pb de comprimento. Possui TSDs idênticas de 8 pb e TIRs de 12 pb, as quais apresentam duas substituições. Ao compará-la com a cópia de *D. buzzatii* verificamos que ela possui uma grande inserção de aproximadamente 1720 pb na posição do nucleotídeo 92. Verificamos por BLASTn, nos genomas de *Drosophila*, que essa sequência está repetida muitas vezes no genoma de *D. mojavensis* (47 *hits*) e *D. willistoni* (14 *hits*). Contudo não sabemos a identidade dessa inserção uma vez que, a mesma, não resultou em similaridade significativa com as sequências depositadas no GenBank (Benson et al. 2008). Como trata-se de uma sequência repetida, também procuramos por sequências homólogas a esse fragmento no Repbase Update (Jurka et al. 2005), um banco de dados de sequências repetitivas de eucariotos, mas que também não mostrou sequência com similaridade significativa.

As outras duas cópias em *D. mojavensis* são incompletas e estão localizadas no scaffold_6500:28308678-28313520 e scaffold_6498_3233207-3234328 e possuem respectivamente 5807 pb e 1121 pb. A cópia com 5807 pb apresenta duas grandes

inserções que correspondem a aproximadamente 3500 pb, explicando a diferença de tamanho.

Em *D. willistoni*, espécie pertencente ao subgênero *Sophophora*, obtivemos 28 *hits* significativos, mas apenas duas cópias parecem ter segmentos grandes da transposase; as demais cópias estão incompletas. A cópia localizada no scf2_1100000004958:2664879-2668034 possui 3156 pb de comprimento. Possui TSDs de 8 pb com uma substituição e TIRs de 12 pb as quais também apresentam uma substituição. Ao compará-la com a cópia de *D. buzzatii* observa-se que possui uma inserção de 428 pb. Observamos por BLASTn nos genomas de *Drosophila* que essa inserção está presente apenas em *D. willistoni* e ela representa 381 *hits* significativos dentro do genoma de *D. willistoni*. Contudo não sabemos a identidade dessa inserção uma vez que esta não resultou em similaridade significativa com as sequências depositadas no GenBank e no Repbase.

A outra cópia contendo segmentos grandes de transposase de *D. willistoni* está localizada no scf2_1100000004967:1340263-1346250 e possui 5988 pb de comprimento. Possui uma deleção de 355 pb iniciais e uma inserção de 4224 pb. Essa inserção está presente repetidas vezes no genoma de *D. willistoni*, representando mais de 500 *hits* e também está presente em *D. mojavensis* (13 *hits*), *D. yakuba* (22 *hits*), *D. erecta* e *D. simulans* (2 *hits*) e em *D. melanogaster* e *D. sechelia* (1 *hit*). Ao fazer BLASTn com as sequências depositadas no GenBank e no RepBase verificamos que essa sequência possui uma alta similaridade (aproximadamente 85%) com o transposon *Minos* (Franz e Savakis 1991) de diferentes espécies.

As cópias completas de *But2* de *D. buzzatii*, *D. mojavensis* e *D. willistoni* apresentaram uma identidade surpreendente (>90%). Essa alta similaridade entre *D. willistoni* e as cópias de *D. buzzatii* e *D. mojavensis* é um forte indício que o elemento *But2* sofreu eventos de TH.

Por tBLASTx encontramos *hits* significativos em *D. yakuba*, *D. ananassae*, *D. pseudobscura*, *D. persimilis*, *D. willistoni* e *D. mojavensis* pertencentes ao gênero *Drosophila* e também em outros insetos como: *Aedes aegypti* (mosquito); *Anopheles gambiae* (mosquito); *Bombyx mori* (bicho da seda); *Tribolium castaneum* (besouro vermelho da farinha); *Nasonia vitripennis* (vespa); *Acyrtosiphon pisum* (pulgão da ervilha) (Tabela 2). As sequências obtidas por tBLASTx nesses insetos, bem como nas demais espécies de *Drosophila*, não possuem uma identidade muito alta com *But2* de *D.*

buzzatii. Esse resultado sugere que essas sequências, embora similares a *But2*, pertençam a outras famílias de elementos *hAT*.

Busca pela região da transposase

Investigamos a presença de íntrons e sequências codificadoras nas cópias completas de *But2* de *D. buzzatii*, *D. mojavensis* e *D. willistoni* por meio de dois programas diferentes, FGENESH e GENESCAN. Para *D. buzzatii* e *D. willistoni*, *But2* parece manter a capacidade de codificar uma transposase ativa.

O resultado obtido pelo programa FGENESH para a cópia de *But2* de *D. buzzatii* prediz uma proteína com 643 aminoácidos (aa) (Fig. 2). Esse programa identificou o sítio de início de transcrição, 5 exons e sinal de poliadenilação. A busca por domínios pela ferramenta SMART mostra que essa sequência possui dois domínios importantes da transposase: um domínio dedo de zinco (*zinc finger*) e o domínio hATC, o qual é um domínio de dimerização encontrado em elementos da superfamília *hAT* (Essers et al. 2000). A sequência predita para *But2* de *D. buzzatii* pelo programa GENESCAN apresenta apenas 4 exons e 494 aa, não contendo o domínio hATC.

O programa GENESCAN predisse uma proteína de 386 aa para a cópia de *But2* em *D. mojavensis*, sendo composta por 4 exons. O programa FGENESH predisse uma proteína de 383 aa, porém sugeriu a existência de um exon adicional interno. Ambas as proteínas preditas diferem na região C-terminal e não possuem os domínio hATC, e dessa forma, possivelmente não são ativas.

Para a cópia completa de *But2* de *D. willistoni* o programa GENESCAN predisse uma proteína de 527 aa, enquanto o programa FGENESH predisse uma proteína de 593 aa e sugeriu a existência de 4 exons, esta última se assemelha mais com a proteína predita para *D. buzzatii* e possui os domínios dedo de zinco e hATC. Os programas não identificaram sítio de início de transcrição, o que possivelmente deve ter sido ocasionado por alguma mutação nessa cópia.

Rastreamento da distribuição de *But2* por PCR

Com o objetivo de analisar a distribuição de *But2* no gênero *Drosophila*, empregamos a técnica de PCR em 67 espécies (Tabela 1). Inicialmente, esperávamos amplificar o elemento *But2* completo, de aproximadamente 2775 pb, a partir dos *primers* *But2_S* e *But2_AS*, os quais foram projetados a partir das TIRs (Fragmento 1

da Fig. 1). Inesperadamente, os fragmentos amplificados a partir desses *primers* foram muito menores que o esperado para todas as espécies e apresentaram tamanhos distintos entre si, inclusive num mesmo grupo de espécies. Em nenhuma das espécies analisadas houve amplificação do elemento completo. Os fragmentos menores foram clonados e sequenciados e correspondem à sequências pequenas relacionadas ao *But2* que apresentam regiões homólogas às extremidades 5' e 3' de *But2*, incluindo as TIRs. Possivelmente, os fragmentos menores são amplificados preferencialmente em relação ao elemento completo devido ao menor tamanho e maior frequência nos genomas. Essas sequências menores, relacionadas ao *But2* poderiam ser consideradas MITEs e poderiam ter sido geradas por deleções do elemento completo seguido de amplificação dessas cópias utilizando a transposase codificada por um elemento autônomo.

Posteriormente projetamos novos *primers* a partir da região central do elemento *But2* (But2C_S e But2C_AS), com o objetivo de identificarmos a presença de cópias completas de *But2* nas 67 espécies testadas. O tamanho esperado para sequências obtidas com esses *primers* é de aproximadamente 750 pb. Combinações dos dois pares de *primers* também foram utilizadas para a geração de dois outros fragmentos (3 e 4 na Fig. 1) para obter as sequências completas de *But2*. Os resultados obtidos pelas reações de PCR com todos os conjuntos de *primers* estão mostrados na Tabela 1. Sinais de amplificação foram encontrados em um grande número de espécies e para as quatro combinações possíveis dos *primers*, incluindo espécies controle negativo (espécies cujos genomas sequenciados não apresentam *But2*), sugerindo a existência de amplificação inespecífica.

PCR in silico

Buscas por PCR *in silico* confirmaram a presença de sequências menores, relacionadas ao *But2*, no genoma de *D. willistoni*. Já em *D. mojavensis* nenhuma sequência com *hit* significativo foi encontrada. No genoma de *D. willistoni* encontramos um grande número dessas sequências (18 cópias) com similaridade média de 85% com o elemento *But2* completo de *D. willistoni* e tamanho variando de 490 a 930 pb. Dessas 18 sequências, 15 apresentam TIRs 100% idênticas às do elemento *But2* canônico de *D. buzzatii*, 7 sequências possuem TSDs idênticas e outras 5 sequências apresentam TSDs com diferença de 1 ou 2 nucleotídeos.

As sequências das TIRs e TSD dessas cópias podem ser vistas na Tabela 3. A conservação de TIRs e TSDs sugere recente atividade de mobilização de *But2* em *D. willistoni*.

Dot blot

A fim de direcionarmos as nossas clonagens e sequenciamentos, pois a PCR estava apresentando casos de falso positivo, realizamos um *dot blot* para confirmar a presença de sequências homólogas ao *But2* em 60 espécies. O resultado do rastreamento por *dot blot* pode ser visto na Fig. 3 e Tabela 1. Observamos forte hibridação nas espécies consideradas como controle positivo: *D. buzzatii*, *D. willistoni* e *D. mojavensis*, comprovando os resultados de busca *in silico*. Por outro lado, as espécies utilizadas como controle negativo não apresentaram nenhum sinal de hibridação. Essas espécies são aquelas que na busca *in silico* não apresentaram sequências relacionadas ao *But2*: *D. virilis*, *D. melanogaster*, *D. erecta* e *D. simulans*.

Além disso, observamos sinal de hibridação em todas as espécies do grupo *willistoni* e do seu grupo irmão *saltans*, assim como em *D. pallidipennis* e *D. incompta*, sugerindo a presença de sequências homólogas ao *But2* nessas espécies.

Clonagem e sequenciamento

Foram clonados e sequenciados os amplicons referentes aos fragmentos 1, 2 e 3 das seguintes espécies: Fragmento 1: *D. ornatifrons*, *D. subbadia*, *D. pallidipennis*, *D. buzzatii*, *D. saltans*, *D. sucinea*, *D. nebulosa*, *D. paulistorum* e *D. tropicalis*; Fragmento 2: *D. pallidipennis*, *D. willistoni* e *D. buzzatii*; Fragmento 3: *D. incompta*, *D. prosaltans*, *D. saltans*, *D. sturtevantii*, *D. willistoni* e *D. equinoxialis*. Confirmamos a identidade de todos os amplicons como sequências similares a *But2* de *D. buzzatii* com exceção de *D. ornatifrons*, *D. subbadia*, *D. saltans* e *D. tropicalis* para o fragmento 1 e *D. incompta* para o fragmento 3. Essas sequências apesar de apresentarem uma amplificação positiva pela reação de PCR não são sequências homólogas ao *But2*. Por isso, a análise conjunta com os dados obtidos pelo *dot blot* nos ajudou na escolha das espécies para priorizar o processo de clonagem, devido ao grande número de amplificações positivas pela PCR.

Análises das sequências

Todos os amplicons obtidos a partir da utilização dos *primers* But2_S e But2_AS (fragmento 1 Fig. 1) foram considerados elementos tipo MITEs, pois a utilização destes *primers* não amplificou o elemento *But2* completo. As sequências obtidas pelo sequenciamento destes MITEs juntamente com as sequências obtidas *in silico* foram usadas para a inferência de uma filogenia (Fig. 4). As sequências de MITEs são, em geral, agrupadas por espécie, indicando maior homogeneidade intraespecífica e possível amplificação e diversificação das sequências pós-especiação.

As sequências obtidas por clonagem e sequenciamento dos fragmentos 2 e 3 (Fig. 1) foram utilizadas para inferir duas filogenias (Fig. 5): uma pelo método Neighbor Joining (A) e outra pelo método de Máxima Parcimônia (B). Em ambos os métodos as sequências de *But2* foram agrupadas em clados por espécie. A relação entre os clados, no entanto, não ficou muito clara, pois os ramos não apresentaram valor significativo de *bootstrap*. Na análise de Máxima Parcimônia apresentamos a filogenia condensada, mostrando a existência de uma grande politomia nas relações entre as sequências de *But2*.

A ausência de significância nas relações entre os clados nas duas filogenias construídas para *But2* é incongruente com as relações filogenéticas das espécies hospedeiras (Fig. 6). Essa baixa resolução das filogenias pode ser resultado da alta similaridade entre algumas sequências e possível complexidade nas relações, como existência de TH, que não podem ser demonstradas por simples bifurcações de ramos filogenéticos.

Para investigarmos a ocorrência de TH de *But2*, a divergência do TE foi comparada com a divergência dos genes nucleares *Amd* e *Adh*. Dada a importância desses genes, é esperada uma alta restrição seletiva para os mesmos. Assim, teoricamente, eventos de TH podem ser inferidos quando a divergência do *But2* é menor do que a divergência dos genes nucleares. Neste caso, podemos observar na Fig. 7 que a maioria das comparações das distâncias genéticas entre espécies dos grupos *saltans* e *willistoni* mostram uma divergência do elemento *But2* maior ou igual do que a divergência dos genes nucleares. Porém, quando comparamos espécies dos grupos *saltans* e *willistoni* com *D. buzzatii*, *D. mojavensis* e *D. pallidipennis* encontramos uma menor divergência do elemento *But2* em relação à divergência do gene nuclear, sugerindo a ocorrência de TH.

DISCUSSÃO

Nesse trabalho, obtivemos os primeiros resultados sobre a distribuição e a evolução do elemento *But2*, um transposon da superfamília *hAT*, no gênero *Drosophila*.

Várias famílias de elementos *hAT* foram caracterizadas em insetos, como *hobo* em *D. melanogaster* (Calvi et al. 1991), *Hermes* em *Musca domestica* (Warren et al. 1994), *Hermit* em *Lucilia cuprina* (Coates et al. 1996), *Homer* em *Bactrocera tryoni* (Pinkerton et al. 1999), *hopper* em *B. dorsalis* (Handler 2003) e *Herves* em *Anopheles gambiae* (Arensburger et al. 2005). Recentemente, novas sequências de *hAT*, pertencentes a quatro famílias diferentes, foram identificadas por Ortiz e Loreto (2009) através de buscas *in silico* nos 12 genomas disponíveis de *Drosophila* (Clark et al. 2007). Ao compararmos as sequências relacionadas com o elemento *But2*, obtidas nesse trabalho, com os elementos do tipo *hAT* descritos por Ortiz e Loreto (2009) e também com demais sequências repetitivas de eucariotos depositadas no Repbase, não encontramos similaridade significativa, indicando que, até o momento não foram descritos elementos similares ao *But2*.

Através das sequências obtidas, da análise de divergência do elemento *But2* em relação aos genes nucleares e pelo *dot blot*, nós inferimos um possível cenário evolutivo para o elemento *But2* (Fig. 6). A presença do *But2* em todas as espécies investigadas dos grupos *saltans* e *willistoni*, revelada por *dot blot*, e a análise de divergência entre as sequências é compatível com transferência vertical de *But2* nessas espécies. Cópias autônomas completas de *But2* estariam presentes no ancestral dos grupos *willistoni* e *saltans*, passando verticalmente para as espécies desses grupos.

As sequências de *But2* de *D. mojavensis* e *D. buzzatii* (espécies próximas, pertencentes ao subgrupo *mulleri* do grupo *repleta*) apresentam divergência entre si compatíveis com TV, sugerindo a presença de *But2* no ancestral do subgrupo *mulleri*. Por outro lado, ao comparar essas sequências com as dos grupos *saltans* e *willistoni*, verificamos que *But2* apresenta divergência menor do que a esperada. Esse achado sugere a ocorrência de TH a partir de alguma espécie do grupo *saltans* ou do grupo *willistoni* para o ancestral das espécies *D. buzzatii* e *D. mojavensis*.

Nossos resultados também sugerem a ocorrência de outro evento de transferência horizontal de *But2* para *D. pallidipennis*. Esta hipótese é baseada na alta similaridade da sequência desta espécie com as demais sequências analisadas, como pode se visto na Fig. 7, e a ausência de *But2* em outras espécies próximas de *D.*

pallidipennis. Potencialmente, tanto espécies dos grupos *saltans* e *willistoni* poderiam ter transmitido o elemento *But2* para *D. pallidipennis*.

Nossos dados revelaram outra característica interessante na evolução do elemento *But2*. Em diferentes espécies são encontradas sequências do tipo MITEs relacionadas ao *But2*. Os MITEs de *But2* são elementos de transposição degenerados, mas compartilham regiões similares nas porções 5' e 3', incluindo as TIRs. Em muitos casos, a similaridade da sequência entre MITEs e o elemento autônomo mais próximo está restrita às TIRs (Feschotte e Prithan 2007). Elementos tipo MITEs provavelmente são originados a partir de transposons autônomos de DNA por degeneração (Bureau e Wessler 1992; Jiang et al. 2003; Quesneville et al. 2006; Ortiz e Loreto 2008). A sequência precursora dos MITEs pode ter se originado através de deleção interna do elemento autônomo ancestral seguido de sua amplificação pela transposase codificada por um elemento autônomo. Análises filogenéticas sugerem que os MITEs de *But2* devem ter sido gerados várias vezes de forma independente em algumas espécies.

A conservação das TIRs está fortemente relacionada à capacidade de mobilização de um TE como pode ser observado para elementos tipo CACTA (Wiker et al. 2003). Dessa forma, a conservação das TIRs e TSDs nesses MITEs de *But2* sugere recente mobilização e a presença de transposase ativa capaz de mediar a sua transposição. De fato, os resultados obtidos pelos programas FGENESH sugerem que as sequências do elemento *But2* de *D. buzzatii* e *D. willistoni* possuem a capacidade de codificar uma transposase ativa. O domínio hATC, domínio característico de TEs da superfamília *hAT*, sustenta a possibilidade de que esta sequência ainda pode estar ativa e suporta a atual classificação do elemento *But2* como membro da superfamília *hAT*. Por essa perspectiva, cópias completas do elemento *But2* e os MITEs podem ainda estar sendo mobilizados. Torres et al. (2006) mostraram a capacidade de mobilização de um elemento *hobo* relíquia (*relic*), pertencente à superfamília *hAT*, possivelmente através de uma cópia completa. Além disso, outros elementos ativos podem estar mobilizando tanto os MITEs quanto cópias completas de *But2*. Sundararajan et al. (1999) mostraram, em insetos, a ocorrência de mobilização cruzada, onde o elemento *hobo* foi capaz de mobilizar o transposon *Hermes*, ambos pertencentes à superfamília *hAT* de transposons.

Finalmente, o presente estudo do elemento *But2* se constituiu em uma importante oportunidade para investigar o comportamento dos elementos de transposição nos genomas de eucariotos, a dinâmica evolutiva gerada pela coexistência

de cópias completas e MITEs e as implicações da presença dos MITEs nos genomas hospedeiros.

Agradecimentos

À PROPESQ-UFRGS e ao CNPq que financiaram essa pesquisa.

Referências

Andolfatto P, Wall JD e Kreitman M (1999) Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics* 153:1297-1311.

Arensburger P, Kim YJ, Orsetti J, Aluvihare C, O'Brochta DA e Atkinson PW. (2005) An active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. *Genetics* 169(2):697-708.

Benson DA, Mizrahi IK, Lipman DJ, Ostell J e Wheeler DL (2008) GenBank. *Nucleic Acids Research*. D25-D30.

Bureau TE e Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4:1283-94.

Burge C e Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268, 78-94.

Cáceres M, Puig M e Ruiz A (2001) Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. *Genome Res* 11:1353-1364.

Cáceres M, Ranz JM, Barbadilla A, Long M e Ruiz A (1999) Generation of a widespread *Drosophila* inversion by a transposable element. *Science* 285:415-418.

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

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

Clark et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 201-218.

Coates CJ, Jonhson KM, Perkins HD, Howells AJ, O'Brochta DA e Atkinson PW (1996) The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica* 97:23-31.

Deprá M, Panzera Y, Valente VLS e Loreto ELS (2010) *hosimary*: a new *hAT* transposon group involved in horizontal transfer. *Molecular Genetics and Genomics* 283: 451-459.

Durando CM, Baker RH, Etges WJ, Heed WB, Wasserman M e DeSalle R (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. *Mol Phylogenet Evol* 16:296–307.

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

Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikodvorskaja V, Pyatkov KI, Zhivotovsky LA e Kidwell MG (2000) Mobile elements and chromosomal evolution in the virilis group of *Drosophila*. *Proc Natl Acad Sci USA* 97:11337-11342.

Feschotte C, Zhang X e Wessler SR (2002) Miniature inverted repeat transposable elements and their relationship to established DNA transposons. Em: Craig NL, Craigie R, Gellert M and Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington, DC, pp 1147-1158.

Feschotte C e Pritham EJ (2007) DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet* 41:331-368.

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

Gilbert C, Schaack S, Pace JK 2nd, Brindley PJ e Feschotte C. (2010) A role for host-parasite interactions in the horizontal tranfer of transposns across phyla. *Nature* 464 pp. 1347–1350.

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

Grumblin G e Strelets V (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Research* 34:484-488.

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

Handler AM (2003) Isolation and analysis of a new hopper *hAT* transposon from the *Bactrocera dorsalis* white eye strain. *Genetica* 118:17-24.

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR e Wessler SR (2003) An active DNA transposon family in rice. *Nature* 421:163-167.

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

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

Letunic I, Doerks T e Bork P (2009). SMART 6: recent updates and new developments. *Nucleic Acids Research* D229-D232.

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

Lytle TW e Haymer DS (1992) The role of the transposable element hobo in the origin of endemic inversions in wild populations of *Drosophila melanogaster*. *Genetica* 86:113-126.

Ludwig A, Valente VLA e Loreto ELS (2008) Multiple invasions of *Errantivirus* in the Genus *Drosophila*. *Insect Molecular Biology* 17: 113-124.

Marín I e Fontdevila A (1995) Characterization of *Gandalf*, a new interted-repeat transposable elemento of *Drosophila koepferae*. *Mol Gen Genet* 248:423-433.

Mota NR, Ludwig A, Valente VLS e Loreto ELS (2010) *harrow*: new *Drosophila hAT* transposons involved in horizontal transfer. *Insect Molecular Biology* 19(2): 217-228.

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

O’Grady P M e Kidwell MG (2002) Phylogeny of the Subgenus *Sophophora* (Diptera: Drosophilidae) Based on Combined Analysis of Nuclear and Mitochondrial Sequences. *Molecular Phylogenetics and Evolution*. Vol. 22 No. 3, pp. 442–453.

- Ortiz M, Loreto EL (2009) Characterization of new *hAT* transposable elements in twelve *Drosophila* genomes. *Genetica* 135:67-75.
- Quesneville H, Nouaud D and Anxolabéhère D (2006) P elements and MITE relatives in the whole genome sequence of *Anopheles gambiae*. *BMC Genomics* 7:214.
- Pinkerton AC, Whyard S, Mende HA, Coates CJ, O'Brochta DA e Atkinson PW (1999) The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol Biol* 8:423-34.
- Regner LP, Pereira MS, Alonso CE, Abdelhay E e Valente VL (1996) Genomic distribution of P elements in *Drosophila willistoni* and a search for their relationship with chromosomal inversions. *J Hered.* 87(3):191-8.
- Robe LJ, Cordeiro J, Loreto EL e Valente VL (2010) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). *Genetica* 138:601–617.
- Sassi AK, Herédia FO, Loreto ELS, Valente VLS e Rohde C (2005) Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. *Genet Mol Biol* 28:734-739.
- Silva JC, Loreto EL e Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-71.
- Sundararajan P, Atkinson PW e O'Brochta DA (1999) Transposable element interactions in insects: crossmobilization of *hobo* and *Hermes*. *Insect Mol Biol* 8:359-368.
- Tamura K, Dudley J, Nei M e Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
- Torres FP, Fonte LF, Valente VL, Loreto EL (2006) Mobilization of a *hobo*-related sequence in the genome of *Drosophila simulans*. *Genetica* 126:101-10.
- Yao H, Guo L, Fu Y, Borsuk LA, Wen TJ, Skibbe DS, Cui X, Scheffler BE, Cao J, Emrich SJ, Ashlock DA, Schnable PS (2005) Evaluation of five ab initio gene prediction programs for the discovery of maize genes. *Plant Mol Biol* 57:445–60.
- Warren WD, Atkinson PW, O'Brochta DA (1994) The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet Res* 64:87-97.
- Wicker T, Guyot R, Yahiaoui N e Keller B (2003) CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements1. *Plant Physiology* 132:52–63.

FIGURAS

Figura 1. Desenho esquemático do elemento *But2* com a posição dos *primers* (setas) e os 4 fragmentos diferentes que podem ser amplificados e seus tamanhos.

Figura 2. Organização da região codificadora da transposase de *But2* de *Drosophila buzzatii* predita pelo programa FGENESH 2.6. TSS – início da transcrição; CDSf – primeiro éxon; CDSi – éxons internos; CDSl – éxon final; PolA – sinal de poliadenilação.

Figura 3. Rastreamento da presença de *But2* por *dot blot* em espécies de *Drosophila*. A sonda utilizada foi o *amplicon* do clone Bf2willi1 de *But2* de *D. willistoni*. Os números no *blot* correspondem às espécies listadas abaixo. 1 – *D. ornatifrons*; 2 – *D. subbadia*; 3 – *D. guaru*; 4 – *D. griseolineata*; 5 – *D. maculifrons*; 6 – *D. nappae*; 7 – *D. paramediostriata*; 8 – *D. tripunctata*; 9 – *D. mediodiffusa*; 10 – *D. mediopictoides*; 11 – *D. cardinoides*; 12 – *D. neocardini*; 13 – *D. polymorpha*; 14 – *D. procardinoides*; 15 – *D. arawakana*; 16 – *D. ornatipennis*; 17 – *D. immigrans*; 18 – *D. funebris*; 19 – *D. gasic*; 20 – *D. brncici*; 21 – *D. gaúcha*; 22 – *D. buzzati*; 23 – *D. mercatorum*; 24 – *D. incompta*; 25 – *D. virilis*; 26 – *D. robusta*; 27 – *D. melanogaster*; 28 – *D. mauritiana*; 29 – *D. teissieri*; 30 – *D. santomea*; 31 – *D. erecta*; 32 – *D. yakuba*; 33 – *D. kikkawai*; 34 – *D. ananassae*; 35 – *D. malerkotliana*; 36 – *D. orena*; 37 – *D. pseudoobscura*; 38 – *D. prosaltans*; 39 – *D. saltans*; 40 – *D. neoelliptica*; 41 – *D. sturtevantii*; 42 – *D. sucinea*; 43 – *D. nebulosa*; 44 – *D. paulistorum*; 45 – *D. willistoni*; 46 – *D. equinoxialis*; 47 – *D. tropicalis*; 48 – *D. insularis*; 49 – *D. capricorni*; 50 – *D. busckii*; 51 – *Z. indianus*; 52 – *Z. tuberculatus*; 53 – *S. latifasciaeformis*; 54 – *S. lebanonensis*; 55 – *D. crocina*; 56 – *D. pallidipennis*; 57 – *D. pavani*; 58 – *D. hydei*; 59 – *D. mojavensis*; 60 – *D. simulans*; Bf1_willi – clone do fragmento 1 de *D. willistoni* (MITE); Bf2_willi – clone do fragmento 2 de *D. willistoni* (sonda).

Figura 4. Árvore filogenética construída pelo método de Neighbor-Joining (Kimura2-Parâmetros) no programa MEGA 4.0. A confiabilidade dos ramos foi acessada por *bootstrap* (1000 réplicas). A filogenia foi obtida a partir das sequências completas do elemento *But2* e dos seus MITEs (*in silico* e clones). A localização das sequências do tipo MITEs obtidas *in silico* pode ser observada na Tabela 3.

Figura 5. Filogenia do elemento *But2* a partir dos elementos completos e dos fragmentos Bf2 obtidos a partir dos clones Bf2 e Bf4. A) Árvore filogenética construída pelo método de Neighbor-Joining (Kimura2-Parâmetros) no programa MEGA 4.0. B) Máxima parcimônia. A confiabilidade dos ramos foi acessada por *bootstrap* (1000 réplicas).

Figura 6. Esquema das relações filogenéticas dos grupos de espécies de *Drosophila* empregados nesse estudo - entre parênteses o número de espécies analisadas. A distribuição de cópias de *But2* do tipo MITEs e completo pode ser observada pela presença das barras vermelha e azul, respectivamente, ao lado do nome da espécie. As barras nos ramos internos indicam possível presença desse tipo de cópia na espécie ancestral.

Figura 7. Análise comparativa da divergência encontrada entre as sequências de *But2* e as sequências dos genes nucleares *Adh* e *Amd*. Teste de qui-quadrado foi utilizado para verificar se a divergência de *But2* observada é significativamente diferente da esperada de acordo com os genes *Adh* ou *Amd*. *** valor $p < 0,001$; ** valor $p < 0,01$; NS não significativo.

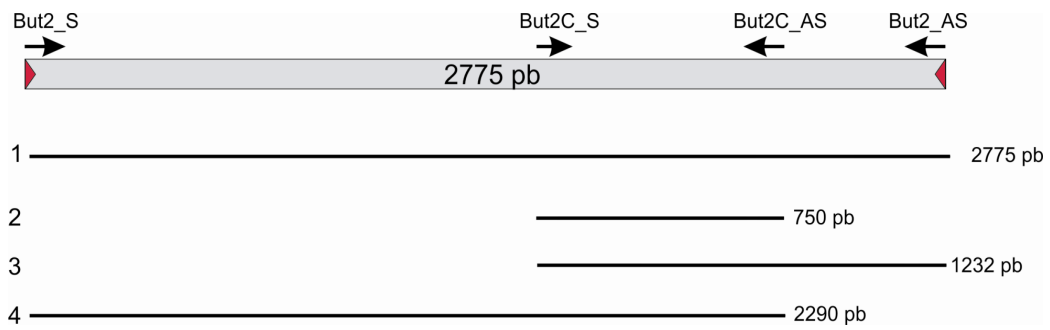


Figura 1. Desenho esquemático do elemento *But2* com a posição dos *primers* (setas) e os quatros fragmentos diferentes que podem ser amplificados e seus tamanhos.

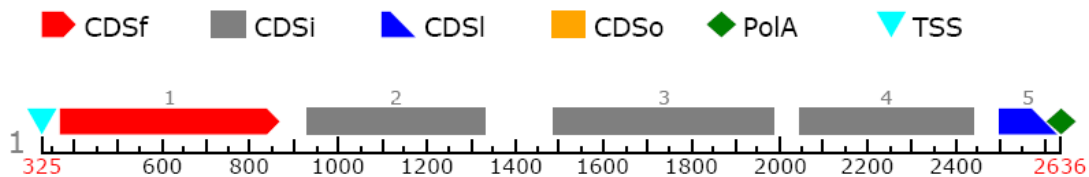


Figura 2. Organização da região codificadora da transposase de *But2* de *D. buzzatii* predita pelo programa FGENESH 2.6. TSS – início da transcrição; CDSf – primeiro éxon; CDSi – éxons internos; CDSl – éxon final; PoA – sinal de poliadenilação.

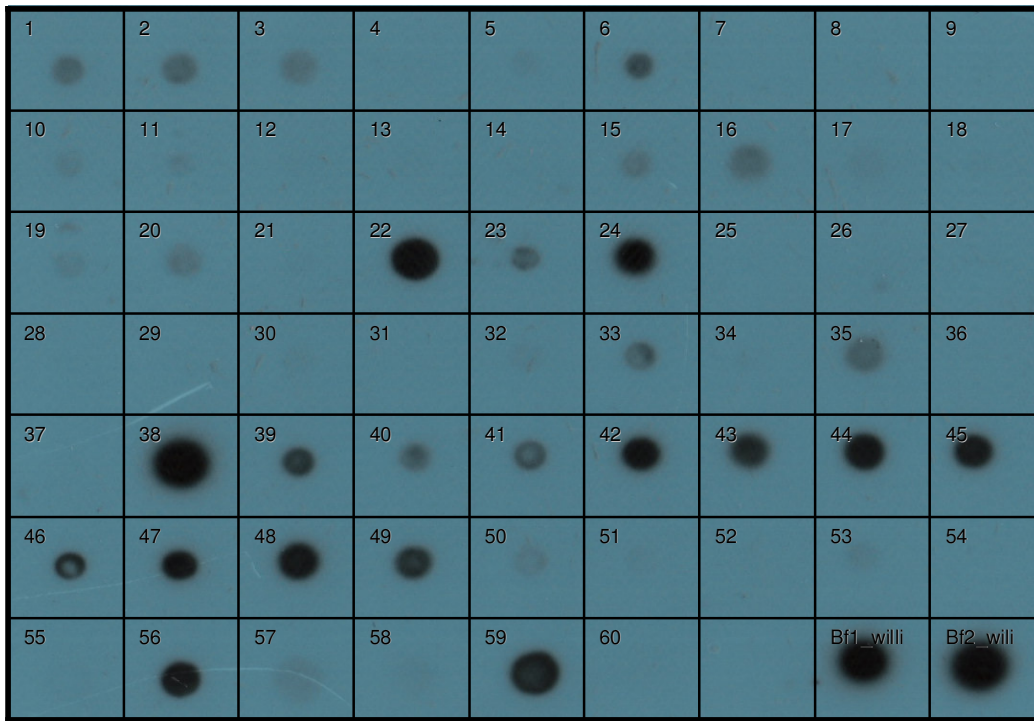


Figura 3. Rastreamento da presença de *But2* por *dot blot* em espécies de *Drosophila*. A sonda utilizada foi o *amplicon* do clone Bf2willi1 de *But2* de *D. willistoni*. Os números no *blot* correspondem às espécies listadas abaixo. 1 – *D. ornatifrons*; 2 – *D. subbadia*; 3 – *D. guaru*; 4 – *D. griseolineata*; 5 - *D. maculifrons*; 6 -*D. nappae*; 7 - *D. paramediostriata*; 8 - *D. tripunctata*; 9 - *D. mediodiffusa*; 10 - *D. mediopictoides*; 11- *D. cardinoides*; 12 - *D. neocardini*; 13 - *D. polymorpha*; 14 – *D. procardinoides*; 15 - *D. arawakana*; 16 - *D. ornatipennis*; 17 - *D. immigrans*; 18 - *D. funebris*; 19 - *D. gasici*; 20 - *D. brncici*; 21 - *D. gaúcha*; 22 - *D. buzzati*; 23 - *D. mercatorum*; 24 - *D. incompta*; 25 - *D. virilis*; 26 - *D. robusta*; 27 - *D. melanogaster*; 28 - *D. mauritiana*; 29 - *D. teissieri*; 30 - *D. santomea*; 31 - *D. erecta*; 32 - *D. yakuba*; 33 - *D. kikkawai*; 34 - *D. ananassae*; 35 - *D. malerkotliana*; 36 - *D. orena*; 37 - *D. pseudoobscura*; 38 - *D. prosaltans*; 39 - *D. saltans*; 40 - *D. neoelliptica*; 41 - *D. sturtevantii*; 42 - *D. sucinea*; 43 - *D. nebulosa*; 44 - *D. paulistorum*; 45 - *D. willistoni*; 46 - *D. equinoxialis*; 47 - *D. tropicalis*; 48 - *D. insularis*; 49 – *D. capricorni*; 50 - *D. busckii*; 51 - *Z. indianus*; 52 – *Z. tuberculatus*; 53 - *S. latifasciaeformis*; 54 - *S. lebanonensis*; 55 – *D. crocina*; 56 – *D. pallidipennis*; 57 – *D. pavani*; 58 – *D. hydei*; 59 – *D. mojavenensis*; 60 – *D. simulans*; Bf1_willi – clone do fragmento 1 de *D. willistoni* (MITE); Bf2_willi – clone do fragmento 2 de *D. willistoni* (sonda).

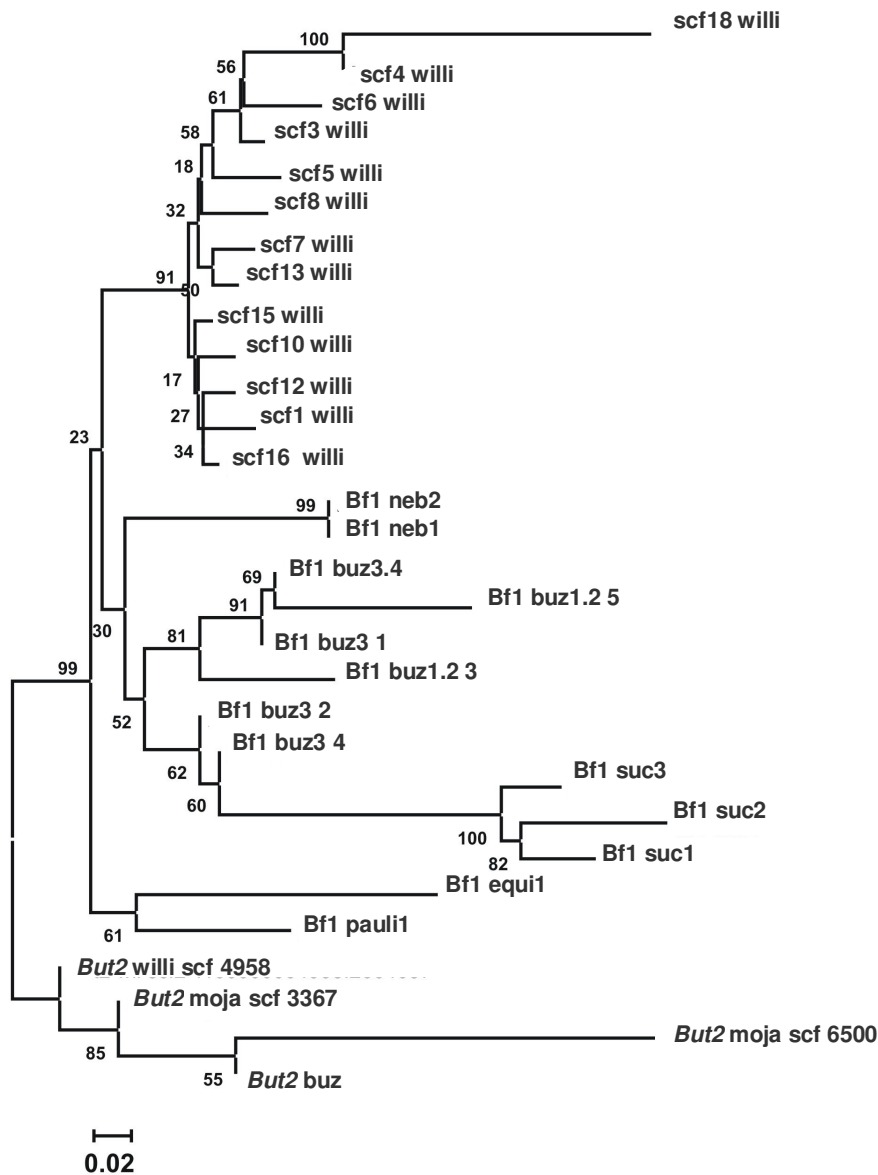


Figura 4. Árvore filogenética construída pelo método de Neighbor-Joining (Kimura2-Parâmetros) no programa MEGA 4.0. A confiabilidade dos ramos foi acessada por *bootstrap* (1000 réplicas). A filogenia foi obtida a partir das sequências completas do elemento *But2* de *D. buzzatii* (But2 buz), *D. willistoni* (But2 willi) e *D. mojavensis* (But2 moja) e dos MITEs clonados encontrados em *D. buzzatii* (Bf1 buz), *D. equinoxialis* (Bf1 equi), *D. nebulosa* (Bf1 neb), *D. paulistorum* (Bf1 pauli), *D. sucinea* (Bf1 suc) e dos MITEs obtidos *in silico* de *D. willistoni* (scfx willi). A localização das sequências do tipo MITEs obtidas *in silico* pode ser observada na Tabela 3.

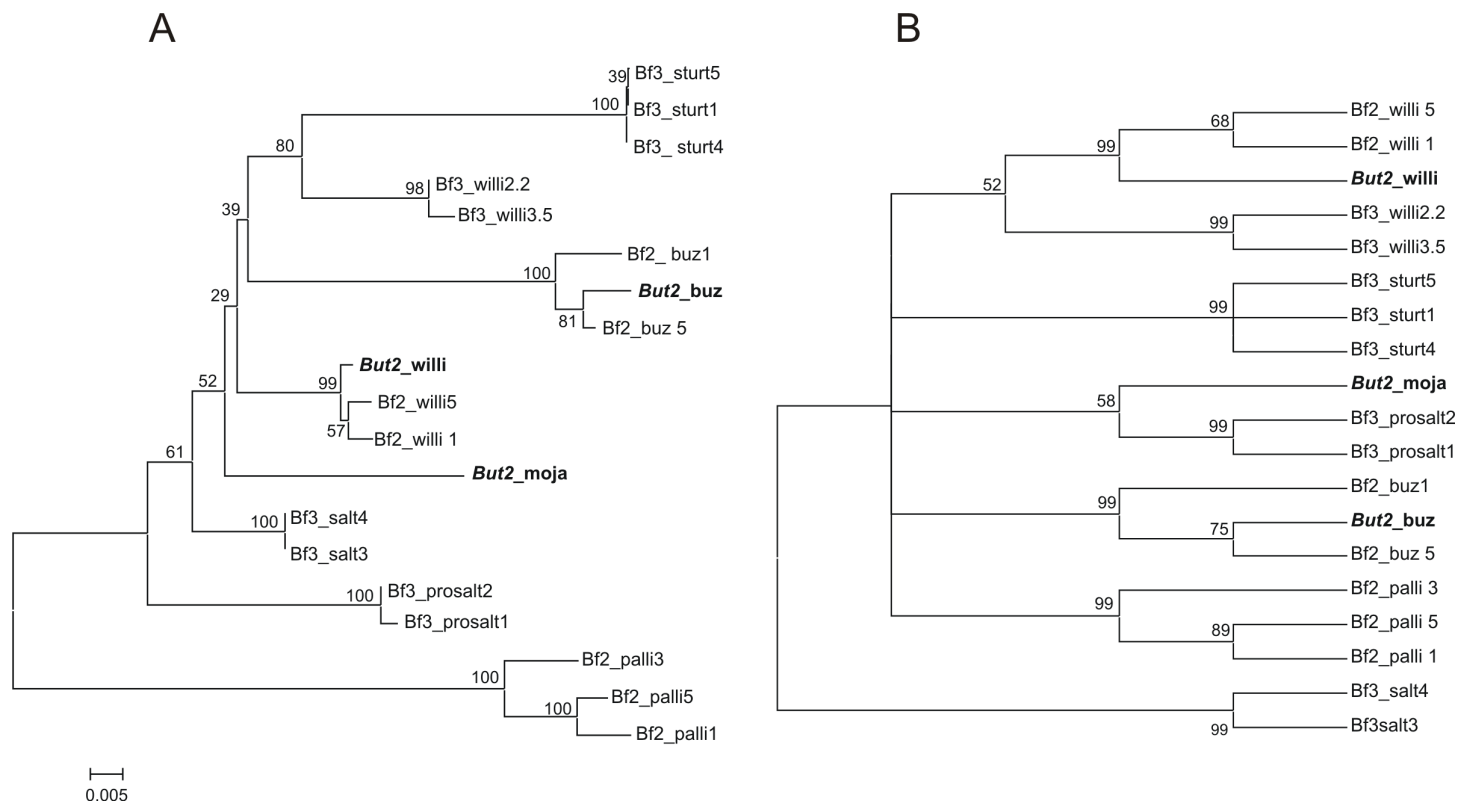


Figura 5. Filogenia do elemento *But2* a partir dos elementos completos e dos fragmentos Bf2 obtidos a partir dos clones Bf2 e Bf4. A) Árvore filogenética construída pelo método de Neighbor-Joining (Kimura2-Parâmetros) no programa MEGA 4.0. B) Árvore filogenética construída pelo método de Máxima parcimônia no programa MEGA 4.0. A confiabilidade dos ramos foi acessada por *bootstrap* (1000 réplicas).

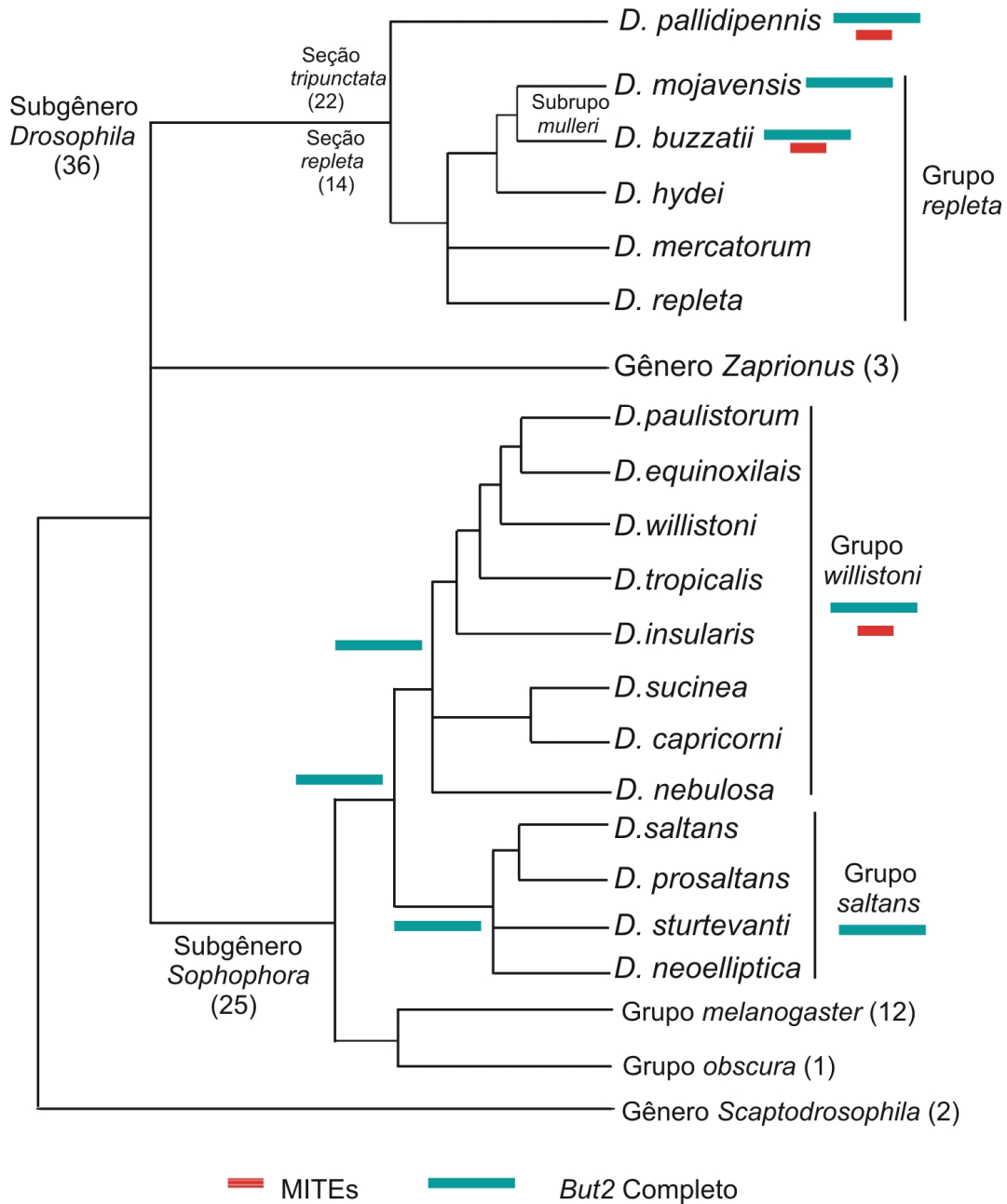


Figura 6. Esquema das relações filogenéticas dos grupos de espécies de *Drosophila* (baseado em Robe et al. 2010; O'Grady e Kidwell 2002 e Durando et al. 2000) empregados nesse estudo - entre parênteses o número de espécies analisadas. A distribuição de cópias de *But2* do tipo MITEs e completo pode ser observada pela presença das barras vermelha e azul, respectivamente, ao lado do nome da espécie. As barras nos ramos internos indicam possível presença desse tipo de cópia na espécie ancestral.

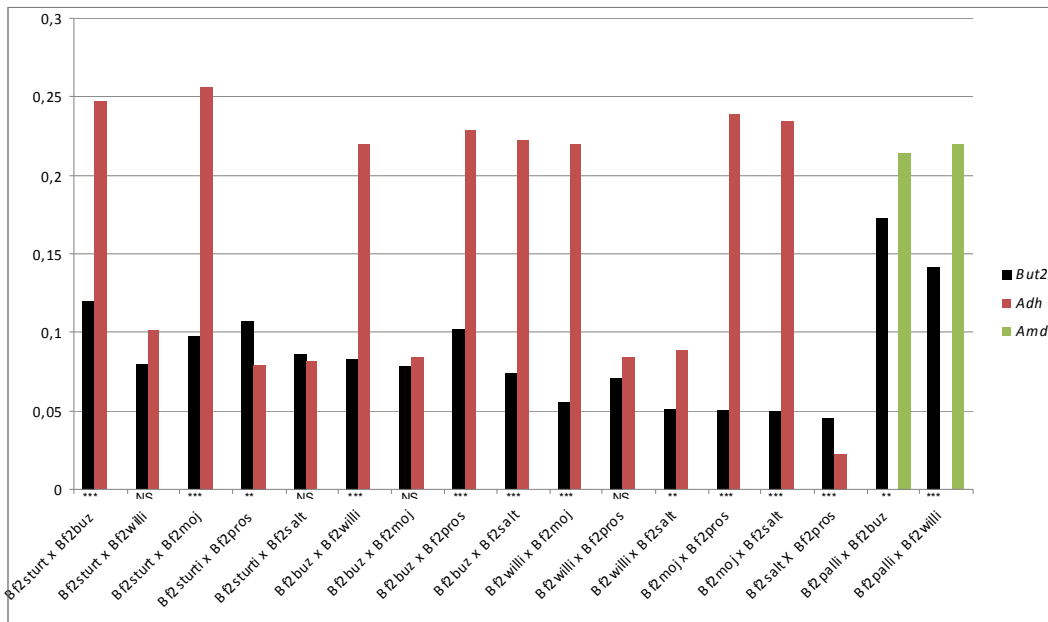


Figura 7. Análise comparativa da divergência encontrada entre as sequências de *But2* e as sequências dos genes nucleares *Adh* e *Amd* para cada um dos pares de espécies mencionados. Teste de qui-quadrado foi utilizado para verificar se a divergência de *But2* observada é significativamente diferente da esperada de acordo com os genes *Adh* ou *Amd*. *** valor $p < 0,001$; ** valor $p < 0,01$; NS não significativo.

Tabela 1. Espécies analisadas pela técnica de PCR com seus respectivos resultados quanto à amplificação de fragmentos do elemento *But2* e *dot blot*.

Gênero	Subgênero	Grupo	Espécie	PCR				Dot	blot	
				1	2	3	4	Resultado	Número no	
<i>Drosophila</i>	<i>Drosophila</i>	<i>guarani</i>	<i>D. ornatifrons</i>	+	+	+	-	w	1	
			<i>D. subbadia</i>	+	+	+	-	w	2	
			<i>D. guaru</i>	+	+	+	-	w	3	
		<i>guaramuru</i>	<i>D. griseolineata</i>	+	+	-	-	-	4	
			<i>D. maculifrons</i>	+	+	+	-	w	5	
		<i>tripunctata</i>	<i>D. nappae</i>	+	+	+	-	w	6	
			<i>D. paraguayensis</i>	+	+	+	-	?		
			<i>D. crocina</i>	+	+	-	-	-	55	
			<i>D. paramediostrata</i>	+	+	+	-	-	7	
			<i>D. tripunctata</i>	-	+	+	-	-	8	
			<i>D. mediodiffusa</i>	+	+	+	+	-	9	
			<i>D. mediopictoides</i>	+	+	+	+	-	10	
			<i>cardini</i>	<i>D. cardini</i>	+	+	-	-	?	
				<i>D. cardinoides</i>	+	+	+	+	-	11
				<i>D. neocardini</i>	+	+	+	-	-	12
		<i>D. polymorpha</i>		-	+	+	-	-	13	
		<i>D. procardinoides</i>		-	-	-	-	-	14	
		<i>D. arawakana</i>		+	+	+	-	-	15	
		<i>D. pallidipennis</i>		+	+	+	-	+	56	
		<i>calloptera</i>	<i>D. ornatipennis</i>	+	+	+	-	w	16	
		<i>immigrans</i>	<i>D. immigrans</i>	-	-	-	-	-	17	
		<i>funnebris</i>	<i>D. funnebris</i>	-	-	+	-	-	18	
		<i>mesophragmatica</i>	<i>D. gasci</i>	-	+	+	-	-	19	
			<i>D. brncici</i>	-	-	-	-	w	20	
			<i>D. gaucha</i>	-	+	+	-	-	21	
			<i>D. pavani</i>	-	-	+	+	w	57	
			<i>repleta</i>	<i>D. hydei</i>	-	-	+	+	-	58
				<i>D. mojavensis</i>	-	+	+	-	+	59
				<i>D. buzzati</i>	+	+	+	+	+	22
				<i>D. mercatorum</i>	-	-	-	-	+	23
			<i>D. repleta</i>	+	+	-	+	?		
			<i>canalinaea</i>	<i>D. canalinea</i>	-	+	-	-	?	
		<i>flavopilosa</i>	<i>D. cestri</i>	+	+	+	-	?		
			<i>D. incompta</i>	-	+	+	-	+	24	
		<i>virilis</i>	<i>D. virilis</i>	-	+	+	-	-	25	
		<i>robusta</i>	<i>D. robusta</i>	+	+	+	-	-	26	
		<i>Sophophora</i>	<i>melanogaster</i>	<i>D. melanogaster</i>	-	+	+	+	-	27
				<i>D. simulans</i>	-	+	-	+	-	60
				<i>D. sechellia</i>	-	+	+	+	?	
				<i>D. mauritiana</i>	-	+	-	-	-	28
				<i>D. teissieri</i>	-	+	-	+	-	29
				<i>D. santomea</i>	-	+	+	+	-	30
				<i>D. erecta</i>	-	+	+	+	-	31
				<i>D. yakuba</i>	-	-	+	+	-	32
				<i>D. kikkawai</i>	-	+	+	+	w	33
				<i>D. ananassae</i>	-	+	+	+	-	34
				<i>D. malerkotliana</i>	-	+	-	+	w	35
<i>D. orena</i>	-			+	-	-	-	36		
<i>obscura</i>	<i>D. pseudoobscura</i>			-	+	-	+	-	37	
<i>saltans</i>	<i>D. prosaltans</i>			+	+	+	+	+	38	
	<i>D. saltans</i>			+	+	+	+	+	39	
<i>neoelliptica</i>	<i>D. neoelliptica</i>			+	+	+	+	w	40	
<i>sturtevantii</i>	<i>D. sturtevantii</i>			+	+	+	+	w	41	
<i>willistoni</i>	<i>D. sucinea</i>			+	-	+	+	+	42	
	<i>D. nebulosa</i>			+	+	+	+	+	43	
	<i>D. paulistorum</i>			+	-	+	+	+	44	
	<i>D. willistoni</i>	+	+	+	+	+	45			
	<i>D. equinoxialis</i>	+	-	+	-	+	46			
	<i>D. insularis</i>	+	-	-	+	+	47			

		<i>D. tropicalis</i>	+	-	+	+	+	48
		<i>D. capricorni</i>	+	+	+	+	+	49
	<i>Dorsilopha</i>	<i>D. busckii</i>	+	+	+	+	w	50
<i>Zaprionus</i>		<i>Z. indianus</i>	-	+	-	+	-	51
		<i>Z. tuberculatus</i>	-	-	+	-	-	52
		<i>Z. sepsoide</i>	-	-	-	-	?	
<i>Scaptodrosophila</i>		<i>S. latifasciaeformis</i>	-	-	-	-	w	53
		<i>S. lebanonensis</i>	-	+	-	+	-	54

(-) ausência de amplificação ou sinal de hibridação; (+) amplificação ou hibridação positiva; (w) sinal de hibridação fraco; (?) não testado. Os números indicam a posição da referida espécie no *dot blot* mostrado na figura 3.

Tabela 2. Número de *hits* encontrados, com valor $E \leq e^{-20}$, em cada espécie utilizando as ferramentas BLASTn e tBLASTx, no Flybase. A *query* utilizada foi a sequência nucleotídica de *But2* de *D. buzzatii*.

Espécies	BLASTn	tBLASTx
<i>Drosophila simulans</i>	0	0
<i>Drosophila sechellia</i>	0	0
<i>Drosophila melanogaster</i>	0	0
<i>Drosophila yakuba</i>	0	8
<i>Drosophila erecta</i>	0	0
<i>Drosophila ananassae</i>	0	11
<i>Drosophila pseudoobscura</i>	0	6
<i>Drosophila persimilis</i>	0	10
<i>Drosophil willistoni</i>	28	51
<i>Drosophila mojavensis</i>	3	26
<i>Drosophila virilis</i>	0	0
<i>Drosophila grimshawi</i>	0	0
<i>Culex pipiens</i>	0	0
<i>Anopheles gambiae</i>	0	1
<i>Aedes aegypti</i>	0	10
<i>Acyrtosiphon pisum</i>	0	59
<i>Nasonia vitripennis</i>	0	26
<i>Apis mellifera</i>	0	0
<i>Tribolium castaneum</i>	0	32
<i>Bombyx mori</i>	0	5

Tabela 3. Localização das seqüências do tipo MITEs obtidas a partir do PCR *in silico* em *D. willistoni* e as respectivas seqüências nucleotídicas das TSDs e TIRs.

Nome da seqüência	Posição no genoma	TSDs	TIRs
scf1_willi	scf2_1100000004943:8134656-8135168	CTC(G/A)AAGT	CAGTGCTGCCAA
scf2_willi	scf2_1100000004521:10175847-10176460	Não conservada	CAGTGCTGCCAA
scf3_willi	scf2_1100000004902:5686245-5686846	Não conservada	CAGTGCTGCCAA
scf4_willi	scf2_1100000004962:321475-322145	GCTCTCGC	CAGTGCTGCCAA
scf5_willi	scf2_1100000004953:192719-193643	CTTC(A/G)AAA	CAGTGCTGCCAA
scf6_willi	scf2_1100000004953:1127087-1128013	GTAGCGAC	CAGTGCTGCCAA
scf7_willi	scf2_1100000004949:1397471-1397893	GCTGCCGG	CAGTGCTGCCAA
scf8_willi	scf2_1100000004954:3708901-3709527	CTTCTAGT	CAGTGCTGCCAA
scf9_willi	scf2_1100000004943:1982535-1983257	Não conservada	CAGTGCTGCCAA
scf10_willi	scf2_1100000004943:8078800-8079574	Não conservada	CAGTGCTGCCAA
scf11_willi	scf2_1100000004943:1484483-1485186	GCGA(G/A)AGG	CAGTGCTGCCAA
scf12_willi	scf2_1100000004962:232665-233445	(C/G)TTTCGT(A/C)	CAGTGCTGCCAA
scf13_willi	scf2_1100000004511:9081204-9081994	CTTCCACT	CAGTGCTGCCAA
scf14_willi	scf2_1100000004839:28485-29274	CGTCGCGG	CAGTGCTGCCAA
scf15_willi	scf2_1100000004830:149935-15707	(G/C)CGGTGAG	C(A/G)GTGCTGCCAA
scf16_willi	scf2_1100000004954:3908919-3909661	CTACTAGA	CAGTGCTGCCAA
scf17_willi	scf2_1100000004943:5612842-5613666	Não conservada	CAGTGCTGC(C/A)AA
scf18_willi	scf2_1100000004521:8051069-8050310	(G/T)TGATAGT	CAGTGCTG(C/T)CAA

Nota: nucleotídeos entre parênteses representam mutações encontradas