

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

**História Evolutiva dos LTR-retrotransposons *Tom*,
297, *17.6* e *rover* em *Drosophilidae***

Newton de Medeiros Vidal

Dissertação submetida ao Programa de
Pós-Graduação em Genética e Biologia
Molecular da UFRGS como requisito
parcial para obtenção do título de Mestre

Orientador: Élgion Lúcio da Silva Loreto

Porto Alegre
Março de 2007

Este trabalho foi realizado no LabDros – Laboratório de Biologia Molecular do Departamento de Biologia da Universidade Federal de Santa Maria e no Laboratório de Genômica Estrutural e Funcional do Centro de Biotecnologia do Estado do Rio Grande do Sul – UFRGS, com recursos do CNPq e FAPERGS.

Dedico este trabalho à Isabel, ao Rafael e ao Victor, minha fonte de inspiração; e à Adriana, por fazer parte da minha vida e acreditar em mim.

AGRADECIMENTOS

Ao professor Élgion, pela orientação, amizade e confiança. Por passar aos seus alunos todo seu entusiasmo e amor pela pesquisa. Por ter me dado a oportunidade de seguir meu sonho.

À professora Lenira, pela presença constante no LabDros, pelos ensinamentos e pela amizade.

Ao professor Henrique e ao professor Arnaldo por terem me recebido em seu laboratório e pela ajuda na realização deste trabalho.

À professora Vera pela ajuda e incentivo neste trabalho. E pelo exemplo de pesquisadora que me passa desde que a conheci, ainda na graduação.

Ao Paulo pela ajuda com as clonagens, à Lizandra pela ajuda com as análises filogenéticas e à Paloma pela ajuda com o seqüenciamento.

À Adriana pela ajuda durante todo o trabalho, principalmente pelas discussões sobre a evolução de elementos transponíveis e a metodologia.

Ao antigo LabDros: Ronaldo, Lizandra, Paulo, Marcos, Maríndia, Daniel e Adriana. Por tudo que aprendi e venho aprendendo com vocês, não só no laboratório, mas fora dele também. Pela nossa amizade que resiste ao tempo.

Ao novo LabDros, especialmente à Nina, à Paloma, ao Felipe, ao Mauro e ao Gabriel, com quem pude conviver mais tempo, pela ajuda e pela amizade.

Aos colegas do Centro de Biotecnologia da UFRGS: Melissa, Alice, Cristiano, Paulo, Marcos, Desirée, Gustavo, Larissa, Alessandra, Terimar, Ana Luisa, Martin, Karina, Veridiana, Patrícia, Tanise, Adriane, Luíza, Aline, Anelise, Taíse, Cláudio, Ana Paula, Luciane, Caroline, Eliana, Priscila e Fernanda. Pelo ambiente de trabalho agradável, pela ajuda, pelas conversas, pelos momentos de descontração e pela amizade, que espero que continue.

Ao Marcos e ao Cristiano pelos ensinamentos e discussões sobre ciência e bioinformática.

Ao pessoal do Laboratório de Drosophila da UFRGS, especialmente ao Marco pelas agradáveis conversas e pela disponibilização de linhagens.

Aos amigos do Departamento de Genética da UFRGS: professora Karen, Daniel, Badaraco, Ana, Shaiane, Tábita, Vanderlei, Nice, Jéferson, Priscila e

Valéria. Por me acolherem nesses dois anos, pelas conversas, pelos cafés, pelos momentos de descontração e principalmente pela amizade de vocês, que eu prezo muito.

Ao Elmo e à Helen pela inestimável ajuda em vários momentos e pelo exemplo de profissionalismo.

Ao Felipe e ao Daniel pelos quase dois anos que moramos juntos. Pelo convívio, pela paciência, pela alegria, pelas discussões sobre a ciência e a vida. Vocês são os irmãos que eu escolhi.

Aos amigos: Eduardo, Sandro, Tiago, Cris, Lu, Jack, Rafael, Patrick, André e João Guedes. Pelos momentos de descontração, pela amizade e pelo incentivo durante a realização deste trabalho.

À minha família: vó Geny, pessoa mais bondosa e prestativa que eu já conheci; vô Crescêncio, pessoa mais inteligente que eu já conheci; Daca, exemplo de estudo e dedicação com a família; e Luís, amigo e sempre prestativo. Obrigado pela vida que dedicaram a mim. Pelo incentivo aos estudos e pela compreensão de minha ausência. Vocês são minha referência. Se em toda a minha vida eu conseguir ser parte do que vocês são já me considerarei feliz.

Aos meus irmãos: Isabel, Rafael e Victor. Pelo carinho e o sentimento que nos une. Pela compreensão dos longos anos que já ficamos separados para que eu seguisse meu sonho. Vocês são a coisa mais importante que tenho na vida, o que mais queria era sempre estar perto de vocês.

Aos meus pais Salete e Airton. Pelo estudo que me proporcionaram e pela oportunidade de trabalhar desde cedo, o que me ensinou muitas coisas.

Às Marias, minha avó e minha madrinha, pelo carinho, incentivo e coisas gostosas que fazem para eu comer.

Aos demais familiares, pelo carinho, pelas conversas e pela compreensão de minha ausência.

À Adriana, por ser minha amiga, colega e companheira. Só você mesmo para conseguir me agüentar. Tenho muita sorte de tê-la ao meu lado. Obrigado pelo carinho, pela paciência, pela compreensão, pelos momentos de consolo, mas principalmente, por acreditar em mim. Você foi muito importante para a realização deste trabalho e continua sendo na minha vida.

SUMÁRIO

RESUMO	7
ABSTRACT	8
CAPÍTULO I	9
INTRODUÇÃO GERAL	9
Elementos Transponíveis.....	9
Classificação dos TEs.....	11
LTRs-retrotransposons.....	14
Evolução dos TEs.....	17
Impacto dos TEs na Evolução.....	18
Regulação dos TEs.....	21
Os Retroelementos <i>297</i> , <i>17.6</i> , <i>Tom</i> e <i>rover</i>	25
O Gênero <i>Drosophila</i>	28
CAPÍTULO II	32
História Evolutiva dos LTR-retrotransposons <i>Tom</i>, <i>297</i>, <i>17.6</i> e <i>rover</i> em <i>Drosophilidade</i>	32
CAPÍTULO III	73
DISCUSSÃO GERAL	73
REFERÊNCIAS BIBLIOGRÁFICAS	76
ANEXOS	87

RESUMO

Os elementos de transposição *297*, *17.6* e *rover* de *Drosophila melanogaster* e *Tom* de *D. ananassae* são retrovírus endógenos de insetos filogeneticamente relacionados. Uma análise inicial da distribuição de seqüências homólogas aos retroelementos *297*, *17.6* e *Tom* em 33 espécies Neotropicais de *Drosophila*, *Zaprionus indianus* e *Scaptodrosophila latifasciaeformis*, utilizando Dot Blot e PCR, indicou a distribuição restrita desses elementos ao grupo *melanogaster* e em *Z. indianus*. Essa análise inicial motivou-nos a fazer este trabalho, com o objetivo de entender a história evolutiva destes três retroelementos. Para isso, clonamos e seqüenciamos uma região da transcriptase reversa nessas espécies e buscamos nos doze genomas de *Drosophila* atualmente disponíveis seqüências homólogas a estes elementos. Os dados das seqüências de alguns clones indicaram a amplificação de outro elemento, o retroelemento *rover*, que também foi analisado neste estudo. A dinâmica evolutiva dos retroelementos *Tom*, *297*, *17.6* e *rover* nos genomas de Drosophilidae são coincidentes em muitos aspectos: (1) Em uma visão geral, as filogenias de cada família são discordantes da filogenia das espécies hospedeiras; (2) Os quatro retroelementos estão envolvidos em eventos de transmissão horizontal; (3) As relações entre as seqüências dos elementos das diferentes espécies são complexas, dificultando a compreensão do cenário evolutivo dos elementos; (4) Os quatro retroelementos apresentam um menor viés na utilização de códons se comparados aos genes nucleares; (5) Pelo menos os elementos *297*, *17.6* e *rover* apresentam valores médios de dN/dS equivalentes ao apresentado pelos genes nucleares, o que indica que essas famílias estão evoluindo sob restrição seletiva. Análises filogenéticas e de divergência nos sítios sinônimos das seqüências da transcriptase reversa dos quatro retroelementos indicam que diferentes eventos podem explicar a história evolutiva desses elementos, incluindo polimorfismo ancestral, transmissão vertical, perda estocástica e transmissão horizontal. Os dados apresentados sugerem que no mínimo 16 casos de transmissão horizontal são necessários para explicar o cenário evolutivo apresentado por esses retroelementos.

ABSTRACT

The transposable elements *297*, *17.6* and *rover* from *D. melanogaster* and *Tom* from *D. ananassae* are closely related insect endogenous retroviruses. An initial Dot Blot and PCR analysis of the distribution of sequences homologues to retroelements *297*, *17.6* and *Tom* in 33 *Drosophila* Neotropical species, plus *Zaprionus indianus* and *Scaptodrosophila latifasciaeformis*, indicates that the distribution of these elements is restricted to the *Drosophila melanogaster* subgroup and *Z. indianus*. In order to understand the evolutionary history of these three elements, we have cloned and sequenced a transcriptase reverse region in these species and searched through sequences homologues to these elements in the twelve available *Drosophila* genomes. Some cloned sequences denote another element being amplified, the *rover* retrotransposon, which was included in this work. The evolutionary patterns of the retrotransposons *Tom*, *297*, *17.6* and *rover* in Drosophilidae genomes agree in several aspects: (1) The retrotransposons phylogenies disagree with the host species phylogeny; (2) The four elements are involved in horizontal transfer events; (3) The relationships among sequences of different species are complex; (4) The four retroelements display a lower codon usage bias than nuclear genes; (5) *297*, *17.6* and *rover* showed dN/dS ratio similar to nuclear genes dN/dS ratio, indicating selective constraints. Phylogenetic analysis and divergence in synonymous sites of reverse transcriptase sequences from these four retroelements indicate different events explaining their evolutionary history, including ancestral polymorphism, vertical transmission, stochastic loss and horizontal transfer. Our data suggest sixteen events of horizontal transfer to explain the evolutionary scenario of *297*, *17.6*, *rover* and *Tom*.

CAPÍTULO I

INTRODUÇÃO GERAL

Elementos Transponíveis

Na década de 1940, durante seus estudos com milho, Barbara McClintock observou que os padrões de cor dos grãos mudavam em várias gerações de cruzamentos, não seguindo uma segregação mendeliana. Ela interpretou seus resultados em termos da regulação da atividade gênica por algum tipo de “elemento controlador” que tinham a capacidade de se mover de um ponto a outro nos cromossomos. Como esses elementos se moviam quando as células se dividiam e proliferavam, eles mutavam genes em apenas algumas células e essas mudanças na coloração eram devido ao efeito desses “elementos controladores” nos genes de pigmentação. A principal conclusão de Barbara McClintock foi de que esses elementos influenciariam de maneira importante o desenvolvimento dos organismos.

Na década de 1970, experimentos levaram a descoberta de elementos de DNA do genoma – em *Escherichia coli*, *Saccharomyces cerevisiae* e *Drosophila melanogaster* – com a capacidade de sair de um local e se inserir em outro local do genoma, gerando mutações em células somáticas e germinativas. Esses experimentos, definitivamente identificaram os “elementos controladores” sugeridos por Barbara McClintock, agora chamados de elementos transponíveis (TEs – *Transposable Elements*), como importantes constituintes dos genomas (revisão em Biémont e Vieira, 2006).

Por sua descoberta, Barbara McClintock foi laureada com o Prêmio Nobel de Fisiologia ou Medicina em 1983.

Dessa forma, os TEs podem ser definidos como seqüências de DNA capazes de mudar de lugar no genoma e freqüentemente podem duplicar cópias

de si mesmos no processo (Wessler, 2006). Eles podem possuir em sua seqüência todas as instruções necessárias para transposição. Com o começo do seqüenciamento em larga escala tornou-se aparente que os TEs não são componentes raros de alguns genomas. Eles perfazem pelo menos 45% do genoma humano e 50-90% do genoma de algumas plantas (SanMiguel *et al.*, 1996; Bennetzen, 2000; International Human Genome Sequencing Consortium, 2001; Biémont e Vieira, 2006).

Elementos ativos, os quais mantêm a estrutura necessária para sua mobilização, constituem, porém, uma pequena fração do conteúdo de TEs na maioria dos genomas eucariotos. Em vez disso, esses genomas possuem centenas, mesmo milhares de cópias de elementos aparentemente inativos. Entretanto, tanto TEs ativos como inativos podem influenciar a evolução da estrutura genômica e a regulação da expressão gênica (Feschotte *et al.*, 2002).

Desde a descoberta dos TEs, numerosas famílias têm sido descobertas e caracterizadas em quase todas as espécies examinadas. Cada família pode ser representada por poucas ou até milhares de cópias em um genoma. Dados recentes obtidos da análise do genoma de *D. melanogaster* indicam que 22% desse genoma é formado por TEs e suas seqüências remanescentes (Kapitonov e Jurka, 2003). Já foram identificados nesse genoma mais de 130 famílias diferentes de TEs (Kapitonov e Jurka, 2003), com o número de cópias variando de uma a 146 (Kaminker *et al.*, 2002).

A abundância e a ubiqüidade dessas seqüências origina uma série de questões sobre a natureza das relações entre TEs e seus hospedeiros, bem como sobre as implicações desses elementos na evolução dos genomas (Kidwell e Lisch, 2001). Embora é aceito que os TEs podem comportar-se como parasitas genômicos e produzir mutações neutras e deletérias, algumas de suas mutações e mudanças genômicas têm dado contribuições significativas à evolução dos seus hospedeiros. Um acúmulo de evidências moleculares têm inequivocamente demonstrado uma grande quantidade de maneiras pelas quais os TEs podem afetar a evolução dos organismos nos quais eles residem (McDonald, 1993; Kidwell e Lisch, 2001; Becker e Lönnig, 2001; Volff, 2006; Xing *et al.*, 2006).

Classificação dos TEs

Os TEs eucarióticos são divididos em duas principais classes de acordo com o seu intermediário de transposição. A Classe I é composta pelos retrotransposons, que se movem por um intermediário de RNA que sofre transcrição reversa para reinserção no genoma. A Classe II é composta pelos transposons que codificam uma transposase e usam um intermediário de DNA para se transpor (Finnegan, 1989).

Além da classificação geral, características estruturais e moleculares dos TEs são também consideradas para agrupá-los em subclasses, superfamílias, famílias e subfamílias. Segundo Capy *et al.* (1998) os critérios para classificação nestes níveis são:

- subclasse: estrutura molecular (presença ou ausência de domínios e/ou assinaturas);

- superfamília: existência de motivos ou assinaturas similares;

- família: existência de similaridade em motivos e em seqüência (entre 30 e 50% de similaridade em seqüências de aminoácidos);

- subfamília: similaridade em seqüência (entre 50 e 70% de similaridade em seqüências de aminoácidos).

Em ambas as classes de TEs, retrotransposons e transposons, os elementos ativos ainda podem ser: autônomos ou não autônomos. Os TEs autônomos possuem as ORFs (*Open Reading Frames* – fases abertas de leitura) que codificam os produtos requeridos para transposição. Em contraste, os não autônomos não codificam proteínas para transposição, mas são capazes de se transpor, pois eles retêm seqüências *cis*-regulatórias necessárias para a transposição e dependem da atividade da transcriptase reversa ou transposase fornecida pelos elementos autônomos (Kidwell e Lisch, 1997). A inserção de quase todos os TEs resulta na duplicação de uma curta seqüência genômica no sítio de inserção, a TSD (*Target Site Duplication* – duplicação do sítio alvo).

Os transposons pertencentes à classe II geralmente possuem uma estrutura simples, com uma curta repetição terminal invertida (ITR – *Inverted Terminal Repeat*) de 10 a 500pb e um único gene codificando a transposase, a

enzima que faz a reação de “corta e cola” em que o elemento é excisado do sítio doador e inserido em um novo local. Essa enzima liga-se de uma maneira seqüência específica nas extremidades de seu elemento codificador autônomo e também nas extremidades de membros de famílias não autônomas. O aumento do número de seqüências de um transposon ocorre quando a seqüência do TE é restaurada no sítio doador (Wessler, 2006).

Já os retrotransposons podem ser divididos em dois grupos:

1) LTR-retrotransposons: possuem longas repetições terminais (LTRs – *Long Terminal Repeats*) em orientação direta que variam de 100 a várias centenas de bases. Esses elementos assemelham-se aos retrovírus de vertebrados na sua estrutura e mecanismo de transposição. Elementos autônomos contêm pelo menos dois genes: *gag* e *pol*. O gene *gag* codifica uma poliproteína do tipo capsídio que produz três proteínas maduras: matriz (MA), capsídeo (CA) e nucleocapsídeo (NC), que correspondem aos componentes da partícula viral semelhante a dos vírus (VLP – *Virus-Like Particle*). O gene *pol* codifica uma poliproteína com as funções de protease (PR), transcriptase reversa (RT), RNaseH e integrase (IN) (Capy *et al.*, 1998). O transcrito do retroelemento, que inicia a partir de um promotor na LTR 5’ e termina na LTR 3’, é transportado para o citoplasma, onde ele serve tanto como mRNA quanto como molde para cDNA. Esse cDNA é transportado para dentro do núcleo onde ele pode ser integrado ao genoma, levando a um aumento no número de cópias. O “ciclo de vida” dos LTR-retrotransposons poder ser vista na figura 1 (Havecker *et al.*, 2004). Os LTR-retrotransposons são encontrados em fungos, animais e compõem uma grande fração da maioria dos genomas de plantas, onde eles parecem ser o maior determinante da enorme variação no tamanho dos genomas (Wessler, 2006).

2) Retrotransposons sem LTRs: são divididos em uma superfamília de elementos autônomos: os LINEs (*Long Interspersed Nuclear Elements* – longos elementos nucleares dispersos); e uma superfamília de elementos não autônomos: os SINEs (*Short Interspersed Nuclear Elements* – curtos elementos nucleares dispersos). Os LINEs são compostos por duas ORFs, codificando uma proteína de ligação a RNA (ORF1) e uma proteína com atividade de

endonuclease e RT (ORF2). Seu tamanho pode variar de 5 a 8Kb. Tanto os LINES quanto os SINEs terminam por uma seqüência de repetição simples, geralmente poli-A. Usualmente, os transcritos dos LINES iniciam em um promotor dentro da extremidade 5' do elemento e terminam freqüentemente a jusante da seqüência de repetição simples. Os SINEs formam um grupo heterólogo de elementos que variam de 90 a 300pb e são derivados de uma variedade de genes de tRNAs ou de 7SL RNA. Esses elementos dependem da maquinaria dos LINES necessária para sua amplificação. Esses elementos são os mais abundantes nos genomas de mamíferos (Wessler, 2006).

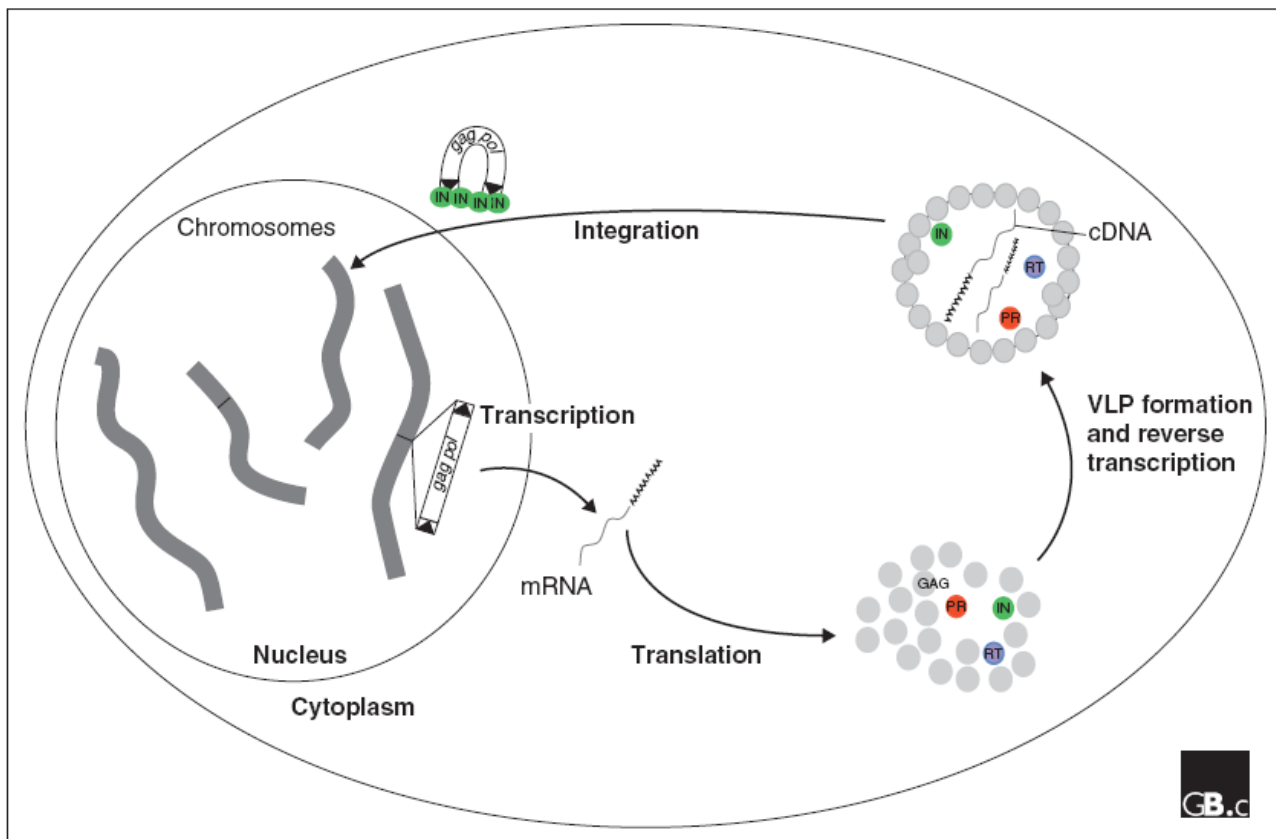


Figura 1: “Ciclo de vida” dos LTR-retrotransposons. IN, integrase; PR, protease; RT, transcriptase reversa; VLP, partícula viral semelhante a dos vírus. Triângulos escuros representam as LTRs (Havecker *et al.*, 2004).

LTRs-retrotransposons

Os LTRs-retrotransposons são subdivididos em dois grandes grupos que diferem na ordem dos domínios enzimáticos codificados pelo gene *pol: Ty1/copia* e *Ty3/gypsy*. No primeiro grupo a ordem dos domínios é 5' PR-INT-RT-RNaseH 3' e no segundo a ordem é 5' PR-RT-RNaseH-INT 3'. O grupo *Ty3/gypsy* pode ainda ser dividido em dois subgrupos (duas superfamílias), aquele que contém famílias de TEs que possuem uma terceira ORF adicional, o gene *env* (superfamília *gypsy-like*) e aquele onde o gene *env* está ausente (superfamília *Ty3*) (Capy *et al.*, 1998). O gene *env* nos retrovírus codifica as proteínas de superfície (SU) e transmembrana (TM) da partícula viral, as quais promovem a interação e fusão da membrana viral com a membrana da célula receptora (Coffin *et al.*, 1997). Na maioria dos elementos da superfamília *gypsy-like*, apesar de apresentarem um gene *env-like*, esse gene não é funcional e ao contrário dos retrovírus seu processo de transposição não é infeccioso (Coffin *et al.*, 1997).

Os LTR-retrotransposons possuem ainda outro sistema de classificação proposto pelo Comitê Internacional em Taxonomia de Vírus (ICTV – *International Committee on Taxonomy of Viruses*). A inclusão dos retrotransposons na classificação dos vírus é baseada nas similaridades da sua organização genômica com os retrovírus e nas relações filogenéticas entre as transcriptases reversas (Hull, 2001). Segundo essa classificação, os retrotransposons com LTRs são subdivididos em duas principais famílias: Pseudoviridae (superfamília *Ty1/copia*) e Metaviridae (superfamília *Ty3* e superfamília *gypsy-like*) (Boeke *et al.*, 1999). Cada uma dessas famílias é formada atualmente por três gêneros.

A família Metaviridae possui os gêneros *Metavirus* (correspondendo à superfamília *Ty3*), *Errantivirus* (correspondendo à superfamília *gypsy-like*) e *Semotivirus*. Cada gênero é composto por diferentes espécies, as quais consistem nos diferentes elementos (Havecker *et al.*, 2004; Mayo, 2005).

O gênero *Errantivirus* é um grupo monofilético de LTR-retrotransposons de insetos. Eles são caracterizados pela aquisição de um gene *env* de baculovírus. Os baculovírus são vírus de DNA dupla fita

patogênicos de insetos, que não possuem estágio de RNA. (Malik *et al.*, 2000). Atualmente existem 10 espécies de *Errantivirus* aprovadas pelo ICTV, as quais compartilham menos de 50% de identidade nas suas seqüências da proteína Gag: *Yoyo* de *Ceratitis capitata*; *Tom* de *D. ananassae*; *17.6*, *297*, *gypsy*, *Idefix*, *Tirant* e *Zam* de *D. melanogaster*; *Tv1* de *D. virilis* e *TED* de *Trichoplusia ni* (ICTVdB – *Universal Virus Database of the International Committee on Taxonomy of Viruses*; Büchen-Osmond, 2003). Existem, porém, inúmeros novos elementos (novas famílias segundo a classificação proposta por Capy *et al.*, 1998) pertencentes a este gênero os quais ainda não foram analisados pelo ICTV para sua classificação como espécie ou subespécie.

Cada um desses elementos citados acima possui pelo menos uma cópia completa com o gene *env* a jusante do gene *pol*. Dessas 10 famílias, em quatro delas, o produto protéico correspondente ao gene *env* foi detectado: *gypsy* e *ZAM* de *D. melanogaster* (Song *et al.*, 1994; Song *et al.*, 1997; Leblanc *et al.*, 2000), *Tom* de *D. ananassae* (Tanda *et al.*, 1994) e *TED* de *T. ni* (Ozers e Friesen, 1996). Entretanto, a presença da proteína Env não implica necessariamente que esses elementos tenham capacidades infecciosas. Até 1994 acreditava-se que os retrovírus fossem restritos aos vertebrados. Porém, Kim *et al.* (1994) demonstraram que o elemento *gypsy* pode ser transmitido por microinjeção de plasma de embriões de *D. melanogaster* de um estoque contendo o elemento ativo em embriões para um estoque desprovido de *gypsy* ativo, mas permissivo à sua transposição. Também foi observada a transferência horizontal do elemento quando indivíduos desse estoque, sem o elemento, são cultivados em meio contendo macerado de pupas do estoque possuidor de elementos capazes de se transpor (Kim *et al.*, 1994). Em outro trabalho, Song *et al.* (1994) demonstraram que a proteína codificada pelo gene *env* é glicosilada, assim como as proteínas do envelope de retrovírus, e é encontrada associada a partículas virais de linhagens de *D. melanogaster* que possuem *gypsy* ativo. Sendo assim, o elemento *gypsy* de *D. melanogaster* é até o momento o único retrotransposon com formação de partículas virais e propriedades infecciosas comprovadas, tendo sido o primeiro retrovírus conhecido de invertebrados. A partir desse momento os

elementos da superfamília *gypsy-like* foram considerados retrovírus endógenos de insetos (IERVs – *Insect Endogenous Retro Viruses* ou *Insect ERantiVirus*) (Terzian *et al.*, 2001). Cabe salientar, que mesmo assim, os *Errantivirus* não são incluídos dentro da família Retroviridae (composta pelos retrovírus de vertebrados), pois são linhagens evolutivas distintas.

Os IERVS são divididos em dois grupos que diferem pela seqüência do sítio de ligação ao primer (PBS – *Primer Binding Site*), que pode ser um tRNA^{Ser} ou tRNA^{Lys}. Essa divisão é suportada por análises filogenéticas das três ORFs. A figura 2 mostra a filogenia dos IERVS baseada nos domínios da RT e RNaseH (Terzian *et al.*, 2001).

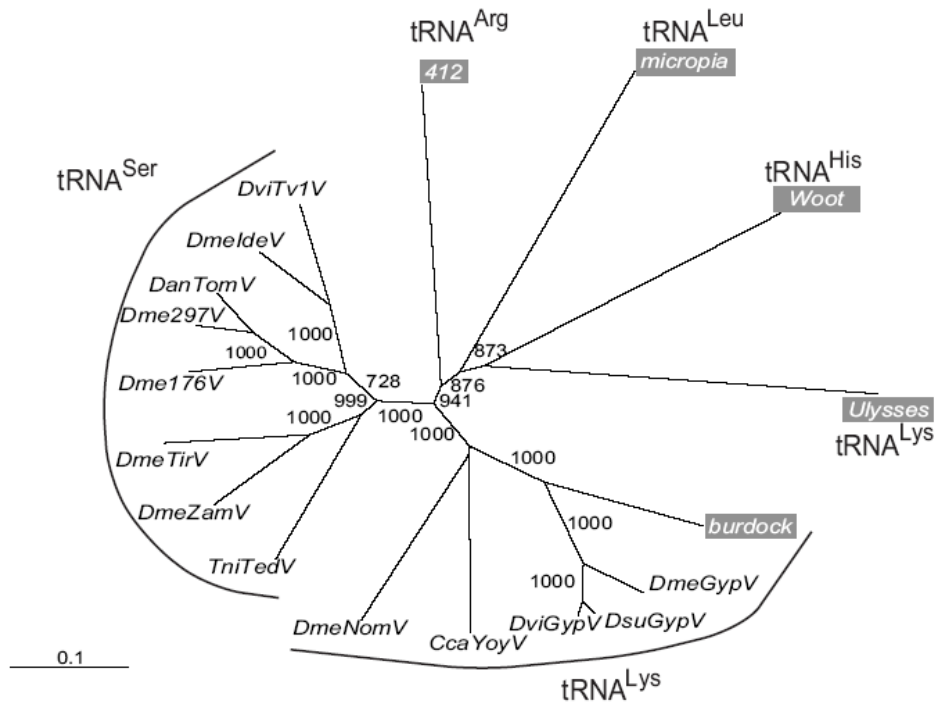


Figura 2: Filogenia não enraizada baseada nos domínios da RT e RNaseH (Terzian *et al.*, 2001).

Evolução dos TEs

A origem dos TEs dentro de uma espécie pode ser resultado de transmissão vertical de cópias presentes em uma espécie ancestral durante o processo de especiação ou de transmissão horizontal, que pode ser definida como o processo pelo qual genes podem se mover entre espécies isoladas reprodutivamente.

Transmissão horizontal tem sido tradicionalmente inferida quando as relações filogenéticas dos TEs não estão de acordo com a história evolutiva das espécies hospedeiras. Segundo Silva *et al.* (2004) existem três métodos utilizados para detectar transmissão horizontal de TEs: (1) detecção de elementos com alto grau de similaridade de seqüência em táxons não relacionados e com divergência entre TEs muito menor do que a divergência entre genes nucleares das suas respectivas espécies hospedeiras; (2) detecção de diferenças topológicas entre a filogenia do TE e das espécies hospedeiras; (3) distribuição descontínua dos elementos entre táxons proximamente relacionados.

Assim, as filogenias obtidas de TEs são freqüentemente comparadas com filogenias de suas espécies hospedeiras, na tentativa de entender a evolução destas seqüências móveis. As congruências entre as duas filogenias são encaradas como indicativos de transmissão vertical dos elementos, enquanto que as discrepâncias podem ser explicadas por transmissão horizontal. Entretanto, explicações alternativas precisam ser testadas, como polimorfismo ancestral com independente distribuição de cópias nas espécies descendentes, taxas diferentes de substituição em TEs nas diferentes espécies e perda estocástica de TEs em alguns táxons (Capy *et al.*, 1994).

Vários casos de transmissão horizontal de elementos entre espécies são relatados tanto para classe I (*copia*, *gypsy*, entre outros) como para classe II (*P*, *mariner*, *hobo*, entre outros). No entanto, o mecanismo pelo qual a transmissão horizontal ocorre permanece obscuro. Alguns vetores potenciais dessa transferência são: vírus, vespas parasitas, ácaros e bactérias endosimbiontes (revisão em Silva *et al.*, 2004).

Para LTRs-retrotransposons, como o *gypsy*, é pensado que os fenômenos de transmissão horizontal são facilitados pela sua natureza viral (Mejlumian *et al.*,

2002). Vários casos de transmissão horizontal do elemento *gypsy* são relatados e associados a possíveis ondas infectivas (Herédia *et al.*, 2004; Ludwig, 2006). Múltiplos casos de transferência horizontal foram sugeridos recentemente para outros elementos *gypsy-like*, como os elementos *gtwin* (Ludwig e Loreto, 2007) e *gypsy2*, *gypsy4*, *gypsy5* e *gypsy6* (Ludwig, 2006).

Impacto dos TEs na Evolução

Por que os TEs predominam na maioria dos genomas e como eles podem influenciar a evolução da vida? Desde a descoberta dos TEs em diferentes organismos, especulações tem centrado no seu possível papel na evolução dos genomas.

No entanto, na década de 1980, duas importantes publicações introduziram fortemente a visão de que os TEs são seqüências de DNA “lixo” e “egoístas” e que seu sucesso evolutivo poderia ser explicado exclusivamente pela sua habilidade de auto-replicação (Doolittle e Sapienza, 1980; Orgel e Crick, 1980). Essa visão tem sido consideravelmente mudada através de uma onda de novas informações demonstrando a importância do DNA móvel na evolução e função dos genes e genomas (Voff, 2006).

Embora o conceito de TEs como parasitas moleculares tenha sido útil para o entendimento da natureza das relações entre o elemento e o hospedeiro, essa abordagem precisa ser expandida e considerar um possível contínuo de interações entre elementos e seus hospedeiros, indo do parasitismo em um extremo ao mutualismo ou simbiose no outro (Kidwell e Lisch, 2001).

No final da década de 1970, François Jacob declarou que a inovação evolutiva acontece primariamente por um processo de “remendos” – do inglês, *tinkering* – que significa essencialmente a criação de novidades através da combinação de formas preexistentes (Jacob, 1977). Pela ubiquidade e potencial mutagênico dos TEs eles foram propostos por alguns autores como uma importante e essencial ferramenta evolutiva, comportando-se como “blocos genéticos” que a evolução pode “remendar”, podendo assim, reestruturar os

genomas e promover variabilidade genética e mudanças na regulação genômica. (Bowen e Jordan, 2002; Jordan, 2006).

Os TEs podem afetar a estruturação genômica de diversas formas. A transdução mediada por retrotransposons de seqüências genômicas a jusante dos elementos é proposta como um importante mecanismo de duplicação gênica (Xing *et al.*, 2006). Os elementos L1 (LINEs) em humanos fornecem a maquinaria molecular necessária para retroposição de genes (Esnault *et al.*, 2000; Wei *et al.*, 2001) e novos genes fusionados podem ser gerados durante o processo (Nisole *et al.*, 2004; Sayah *et al.*, 2004). Transposons de DNA também podem mediar embaralhamento de éxons e a duplicação gênica, como demonstrado para os elementos *mutator-like* e *helitron-like* em plantas (Jiang *et al.*, 2004; Morgante *et al.*, 2005).

Um caso de domesticação molecular de TEs bem conhecido em *Drosophila* é o caso dos elementos Het-A e TART, os quais foram cooptados pelo genoma hospedeiro para manter a integridade dos telômeros (Levis *et al.*, 1993; Pardue e DeBaryshe, 2003). *D. melanogaster* não possui o gene telomerase e assim, não pode utilizar essa enzima para catalisar a regeneração das extremidades dos cromossomos que são progressivamente encurtados depois de ciclos de replicação. Pela alta taxa de transposição e inserção nos telômeros, os elementos Het-A e TART são capazes de estender as extremidades cromossômicas e evitar a perda de DNA associada à replicação.

Além de exercerem papel na estruturação genômica, os TEs por si só podem servir como “matéria-prima” para o surgimento de novas funções pela sua incorporação em genes já existentes (Nebrutenko e Li, 2001; Cordaux *et al.*, 2006). Ainda não está clara qual a real proporção do impacto dos TEs na evolução de regiões codificadoras de proteínas dos genomas. Por exemplo, uma análise de aproximadamente 14 mil genes humanos revelou que 4% continha inserções de TEs (Nekrutenko e Li, 2001). Este mesmo estudo encontrou mais de 100 casos onde as inserções de TEs são responsáveis pela divergência entre genes ortólogos de diferentes espécies. Dessa forma, essas seqüências móveis podem não somente representar um mecanismo para formação de novos genes, mas também desempenhar um papel conspícuo na diversificação das espécies.

Cordaux *et al.* (2006) exploram um exemplo desse tipo de evolução na linhagem humana, onde um gene existente no hospedeiro funde-se com um TE para criar um gene quimérico primata-específico. Os autores mostraram que o gene da transposase foi recrutado como parte do gene SETMAR entre 40 e 58 milhões de anos atrás, após a inserção de um TE *Hsmar1* (elemento *mariner* de humanos) a jusante de um gene SET preexistente.

Em contraste aos genomas de mamíferos, pouco mais de 1% dos genes de *Drosophila* produzem mRNAs que incluem verdadeiramente seqüências de TEs. A grande maioria das inserções de TEs potencialmente quiméricas são provavelmente deletérias. Esses resultados indicam que o processo de domesticação dos TEs pode variar drasticamente entre táxons (Lipatov *et al.*, 2005).

Apesar de muitos estudos descobrirem cassetes de TEs – um fragmento de TE inserido em um mRNA – em regiões codificadoras de muitos genes (Li *et al.*, 2001; Nekrutenko e Li 2001; Lorenc e Makalowski, 2003), a presença dos cassetes nos transcritos não garante a sua tradução. Recentemente, Gotea e Makalowski (2006) realizaram buscas de cassetes de TEs em proteínas humanas bem caracterizadas funcionalmente em bancos de dados de proteínas. Os autores mostraram que cerca de 1% das proteínas analisadas possuíam fragmentos de TEs, muito menos do que a estimativa de 4% (Nekrutenko e Li, 2001) no nível de transcrição. Em todos os casos os cassetes foram derivados de TEs antigos, consistente com a idéia de que a domesticação de um TE em proteínas funcionais requer longos períodos evolutivos. Mas o número significativo de transcritos contendo TEs pode indicar um importante papel dessas seqüências na regulação da expressão gênica (Gotea e Makalowski, 2006).

Existem agora substanciais evidências em uma variedade de organismos que seqüências derivadas de TEs possuem um papel adicional na evolução. Não há dúvidas, porém, de que o que conhecemos sobre o impacto evolutivo da domesticação dos TEs é somente a ponta do iceberg (Voff, 2006).

Regulação dos TEs

Devido a sua capacidade de transposição, os TEs são importantes agentes mutagênicos com o potencial de causar uma variedade de mudanças nos genomas de seus organismos hospedeiros. Embora se reconheça que os TEs contribuam significativamente para a evolução dos genomas, a atividade descontrolada de um TE pode ser potencialmente prejudicial ao “fitness” do hospedeiro. Portanto, mecanismos que silenciam TEs evoluíram para estabilizar os genomas (Kidwell e Lisch, 2001). Para a maioria dos TEs, a transposição é regulada para ocorrer somente nas linhagens germinativas. Esta restrição parece ser uma estratégia altamente eficiente para garantir sua transmissão para a próxima geração, minimizando, ao mesmo tempo, os danos ao hospedeiro (Kidwell e Lisch 2001).

Inicialmente, os próprios TEs possuem suas próprias seqüências *cis*-regulatórias que modulam sua transcrição. Além disso, outros mecanismos são conhecidos. Para elementos de Classe II mecanismos de competição, envolvendo cópias completas e deletadas, são observados, assim como, para o elemento *mariner* observamos uma regulação onde o excesso de transposase inibe sua atividade devido a um fenômeno de multimerização (Hartl *et al.* 1997; Capy *et al.* 1998).

A metilação dos resíduos de citosina em genomas eucarióticos é freqüentemente associada com seqüências repetitivas, incluindo TEs. A metilação dessas seqüências deve refletir a ação de um sistema de defesa genômica que pode silenciar os TEs (Selker, 1999). A metilação está associada no controle de dois potenciais efeitos deletérios dos TEs: a transcrição descontrolada e as mudanças estruturais causadas pela recombinação ectópica. Esse mecanismo parece ser um componente chave de silenciamento de TEs em plantas e vertebrados (Yoder *et al.*, 1997). Por algum tempo, se acreditava que muitas espécies de invertebrados, incluindo *Drosophila*, não possuíam esse mecanismo de metilação. Porém, recentes investigações mostram a existência desse processo em *D. melanogaster* (Liko *et al.*, 2000), *Anopheles gambiae*, *D. pseudoobscura*, *D. simulans*, *D. hydei*, *D. virilis* (Marhold, 2004) e *D. willistoni* (Garcia *et al.*, 2007). Apesar da evidência de nucleotídeos metilados nesses

dípteros, em *D. melanogaster* uma pequena proporção das seqüências metiladas consistem em TEs, argumentando contra o papel da metilação na regulação dos TEs nesses organismos (Mandrioli e Borsatti, 2006).

Recentemente, tem-se demonstrado a contribuição de um importante mecanismo de silenciamento gênico, mediado por RNA dupla fita (dsRNA – *double-stranded RNA*), envolvido na proteção contra proliferação de TEs em diferentes organismos (revisão em Djikeng *et al.*, 2001; Aravin *et al.*, 2003; Almeida e Allshire, 2005; Aravin e Tuschl, 2005).

Esse silenciamento gênico mediado por dsRNA dependente de homologia é denominado interferência de RNA (RNAi – *RNA interference*) (Fire *et al.*, 1998). Simplificadamente, nesse processo, um dsRNA longo é processado pela enzima Dicer (ribonuclease tipo III) em pequenos RNAs interferentes (siRNAs – *small interfering RNAs*) que direcionam a repressão de seqüências homólogas. Os siRNAs guiam a atividade de clivagem de mRNAs complementares pelo complexo silenciador induzido por RNA (RISC – *RNA-Induced Silencing Complex*), ou induz eventos relacionados à cromatina que resultam em silenciamento transcricional. TEs e vírus podem gerar dsRNAs a partir da transcrição da fita anti-senso e da geração de estruturas de grampos no transcrito nascente (revisão em Almeida e Allshire, 2005; Kalmykova *et al.*, 2005).

Mutações em genes necessários para o processo de RNAi, como o gene *piwi*, freqüentemente levam a um aumento na atividade dos TEs. A proteína Piwi é relacionada à família de proteínas altamente conservada Argonauta, que está diretamente relacionada aos mecanismos de RNAi, mais especificamente participando do complexo RISC (revisão em Sarot *et al.*, 2004; Kalmykova *et al.*, 2005).

Existem alguns casos descritos do envolvimento de RNAi no silenciamento de retrotransposons em *Drosophila*:

(1) Um dos primeiros trabalhos indicando o papel desse mecanismo em *Drosophila* foi realizado por Aravin *et al.* (2003). Eles clonaram e seqüenciaram 4000 pequenos RNAs, dos quais 113 corresponderam a TEs, principalmente retrotransposons. Dentre os LTR-retrotransposons com pequenos RNAs identificados estão os elementos *17.6*, *Quasimodo*, *HeT-A* e *TART*.

(2) Vagin *et al.* (2004) revelaram o efeito de genes envolvidos no mecanismo de RNAi na expressão do elemento *I* em *D. melanogaster*.

(3) Em *D. melanogaster* o locus ainda não clonado *flamenco* (*flam*) controla a mobilização do retrotransposon *gypsy* através de silenciamento mediado por RNA. Alelos restritivos (*flamR*) diminuem o acúmulo de transcritos de *gypsy* no epitélio folicular somático dos ovários. De modo inverso, existe um acúmulo de RNAs de *gypsy* no epitélio folicular dos ovários de fêmeas portando dois alelos permissivos (*flamP*) (Prud'homme *et al.*, 1995). Sarot *et al.* (2004) observaram a presença de pequenos RNAs anti-senso de *gypsy* nos ovários restritivos e não nos ovários permissivos. Além disso, esses autores evidenciaram que a repressão de *gypsy* é prejudicada em fêmeas restritivas mutantes para o gene *piwi*. Recentemente, Mevel-Ninio *et al.* (2007) demonstraram que o locus *flam* regula também o retroelemento *Zam*.

(4) Outro retroelemento, *Idefix*, possui sua expressão controlada por um locus próximo ao gene *flam* em *D. melanogaster*. *flam* está localizado em uma região onde numerosas cópias deletadas de TEs são acumuladas (Desset *et al.*, 2003). Esses autores sugerem que *Idefix* e possivelmente outros TEs podem ser potencialmente sintetizados a partir desta região e podem torna-se a base para o mecanismo de repressão envolvendo RNAi.

(5) Kalmykova *et al.* (2005) demonstraram que mutantes *piwi* de *D. melanogaster* possuem expressão dos retrotransposons *mdg1* e *copia* aumentada na linhagem germinativa de machos, alteração da sua expressão padrão e mobilização do elemento *mdg1*, indicando que a perturbação do sistema de silenciamento de RNA nas células germinativas pode causar altas taxas de transposição.

(6) Saito *et al.* (2006) realizaram experimentos de imunopurificação da proteína Piwi de ovários e encontraram pequenos RNAs de 25-29 nucleotídeos associados a essa proteína, os quais foram identificados por clonagem e seqüenciamento como *rasRNAs* (*repeat-associated small interfering RNAs*) derivados de regiões repetitivas, tal como retrotransposons e regiões heterocromáticas no genoma de *Drosophila*. A figura 3 mostra a tabela com os

retroelementos que possuem *ras*/RNAs associados a Piwi, detectados por esses autores.

O possível envolvimento de um mecanismo de silenciamento mediado por RNA também foi encontrado em elementos de DNA como os elementos *1360*, *hopper2* e *transib4* em *D. melanogaster* (Aravin *et al.*, 2003) e o elemento *P* em *D. willistoni* (Blauth, 2005).

Repeat	Name	Total	+	-	
LINE-like	BS	2	0	2	
	F-element	1	0	1	
	G-element	2	0	2	
	HeT-A-element	5	3	2	
	TART-element	3	0	3	
	TAHRE	1	0	1	
	X-element	2	1	1	
	R1Dm	8	4	4	
	R2Dm	2	2	0	
	pilger	5	3	2	
	strider	1	0	1	
	Total (LINE)	32	13	19	
	LTR	1731	4	4	0
		412	9	4	5
blood		3	2	1	
gypsy		4	0	4	
springer		7	1	6	
HMS-Beagle		8	2	6	
Max-element		4	2	2	
mdg1		1	0	1	
mdg3		2	1	1	
opus		4	2	2	
roo		11	1	10	
Stalker		17	0	17	
ninja		2	0	2	
297		5	1	4	
cruiser		3	1	2	
circe		3	2	1	
BEL		3	1	2	
Tirant		1	1	0	
Micropia-Dm11		1	0	1	
rover	3	1	2		
copia	1	0	1		
Total (LTR)	96	26	70		
TIR	Hoppel	1	1	0	
	Bari-1	1	1	0	

(+) Same orientation with the ORF of the LINE-like and LTR retrotransposons.

Figura 3: Tabela mostrando os rasiRNAs identificados associados à proteína Piwi (modificada de Saito *et al.*, 2006).

Os Retroelementos 297, 17.6, Tom e rover

O retroelemento 297 foi descrito por Potter *et al.* (1979) quando procuravam por famílias de seqüências medianamente repetitivas no genoma de *D. melanogaster*. Ele possui 6995pb de comprimento, LTRs com 415pb e produz TSDs de 4pb. Na região eucromática do genoma de *D. melanogaster* foram encontradas 57 cópias deste elemento, sendo 18 delas completas (Kaminker *et al.*, 2002). Esse elemento também foi detectado em *D. simulans* e *D. mauritiana* utilizando Southern Blot (Biémont e Cizeron, 1999). Em *D. simulans* este elemento encontra-se em um menor número de cópias do que em *D. melanogaster*, determinado por hibridização *in situ* (Biémont e Cizeron, 1999). Para *D. simulans* estão disponíveis seqüências parciais das LTRs 5' e 3', da ORF 2 e da ORF 3 (Nuzhdim, 1995).

O retroelemento 17.6 foi encontrado inserido entre genes de histonas de *D. melanogaster*, por hibridização usando como sonda o retroelemento 297. Foi descrito por Saigo *et al.* (1981) e a seqüência obtida por Saigo *et al.* (1984), possui 7439pb de comprimento, LTRs com 512pb e produz TSDs de 4pb. Através da análise *in silico* do genoma de *D. melanogaster* foram encontradas 12 cópias, das quais sete são completas (Kaminker *et al.*, 2002). O elemento 17.6 também foi encontrado em *D. simulans* por hibridização *in situ*. Esse elemento possui um maior número de cópias nessa espécie do que em *D. melanogaster* (Biémont e Cizeron, 1999).

O retroelemento *Tom* possui 7060pb de comprimento, LTRs com 475pb e ao se transpor produz TSDs de 4pb. A denominação *Tom* significa Transposon das mutações OM, pois foi descoberto associado a mutações ópticas morfológicas (Om – *Optic Morfology mutations*) (Tanda *et al.*, 1988; Matsubayashi *et al.*, 1992).

As mutações Om estão relacionadas à pelo menos 22 lócus diferentes espalhados pelo genoma de *D. ananassae* e estão associadas à inserção do retrotransposon *Tom* dentro ou próximo a esses lócus. As LTRs deste elemento carregam seqüências regulatórias de genes tecido-específicos que aumentam a expressão dos genes Om, causando um padrão anormal na morfologia do olho das moscas.

Os retroelementos *297*, *17.6* e *Tom*, são IERVs proximamente relacionados (figura 2) e, portanto, possuem a capacidade de codificar uma provável proteína de envelope. Entretanto, apenas para o elemento *Tom* a presença da proteína Env foi evidenciada (Tanda *et al.*, 1994).

Pouca ênfase tem sido dada ao estudo dessas famílias de retrotransposons, embora sejam boas candidatas a serem infectivas. A maioria dos estudos concentra-se na determinação do número de cópias em diferentes populações de *D. melanogaster* e *D. simulans*.

Para buscar um melhor entendimento dos mecanismos de regulação da transposição e história evolutiva dos IERVs o primeiro passo é conhecer a distribuição dessas seqüências nas diferentes espécies de *Drosophila*.

Os dados da literatura, até o momento, apontam para a presença dos retrotransposons *297*, *17.6* e *Tom* apenas em espécies do grupo *melanogaster*. Entretanto, não podemos afirmar que esses elementos estejam restritos ao grupo *melanogaster* de *Drosophila* sem uma análise mais abrangente da sua distribuição. Para resolver essa dúvida começamos uma investigação da distribuição de seqüências homólogas aos retroelementos pertencentes a essas famílias em 33 espécies Neotropicais de *Drosophila*, *Zaprionus indianus* e *Scaptodrosophila latifasciaeformis*, utilizando Dot Blot e PCR (dados não publicados).

Uma análise inicial por Dot Blot indicou a presença de seqüências homólogas aos retroelementos *297*, *17.6* e *Tom* com um sinal de hibridização forte apenas em espécies do grupo *melanogaster*: *Drosophila melanogaster*, *D. simulans*, *D. ananassae*, *D. malerkotliana* e *D. kikkawai*, concordando com os dados da literatura.

A análise por PCR, utilizando *primers* degenerados projetados para amplificação os três elementos, mostrou resultados concordantes com aqueles encontrados por Dot Blot. Obtivemos a amplificação do fragmento esperado de 1Kb nas espécies do grupo *melanogaster*. No entanto, na análise por PCR, também obtivemos a amplificação de seqüências homólogas a esses retroelementos em *Z. indianus*, o que é um resultado surpreendente, já que se

trata de um outro gênero, embora tenha surgido a partir da radiação *immigrans-Hirtodrosophila* (Throckmorton, 1975).

Este achado desperta nossa atenção, pois já foi sugerida a transferência horizontal do transposon *mariner* entre *Z. indianus* e *D. simulans* (Brunet *et al.*, 1999), e do retrotransposon *gypsy* entre *Z. indianus* e *D. melanogaster* ou *D. simulans* (Herédia *et al.*, 2004).

Embora nossos dados indiquem que os retrotransposons 297, 17.6 e *Tom* sejam restritos a essas espécies do grupo *melanogaster* e *Zaprionus indianus*, não podemos descartar a possibilidade de que eles não tenham sido detectados em nossa análise nos outros grupos de *Drosophila* devido à divergência entre estes elementos, pois as sondas utilizadas e os *primers* projetados foram feitos com base nas seqüências de *D. melanogaster* e *D. ananassae* (ambas do grupo *melanogaster*).

A publicação da seqüência genômica “release 1” de *D. melanogaster* (Adams *et al.* 2000) encorajou muitos estudos de distribuição genômica e abundância dos elementos de transposição nesse organismo. Isso resultou na descoberta e caracterização de novos elementos (Jurka, 2000; Bowen e McDonald, 2001; Bartolome *et al.*, 2002; Kaminker *et al.*, 2002; Rizzon *et al.*, 2002; Kapitonov e Jurka, 2003; Lerat *et al.*, 2003), como é o caso do retroelemento *rover* descrito por Kaminker *et al.* (2002), que possui maior similaridade com 17.6. Este retroelemento possui 7318pb e foram detectadas seis cópias no genoma de *D. melanogaster*, sendo três completas e três parciais. Nenhum estudo adicional da distribuição do retroelemento *rover* foi feito até o momento.

Atualmente, estão sendo seqüenciados outros onze genomas de *Drosophila* (além do genoma de *D. melanogaster* já completado) e assim que as seqüências são geradas tornam-se disponíveis nos bancos de dados públicos. Utilizando ferramentas computacionais podemos utilizar esses bancos de dados para fazer uma busca *in silico* e ampliar as análises de distribuição e evolução dos TEs. Neste contexto, nossa proposta de trabalho consistiu na análise evolutiva dos retrotransposons 297, 17.6, *Tom* e *rover* através de buscas *in silico*

e seqüenciamento desses retroelementos nas espécies onde eles foram encontrados previamente.

O Gênero *Drosophila*

O gênero *Drosophila* é composto por mais de 1400 espécies distribuídas em 15 subgêneros (Wheeler, 1982). Dentre estes subgêneros, pode-se destacar os subgêneros *Drosophila* e *Sophophora* pela diversidade de espécies apresentada.

O subgênero *Sophophora*, estabelecido por Sturtevant (1939), compreende 233 espécies, subdivididas em sete grupos, dentre os quais se destacam os grupos *melanogaster*, *obscura*, *saltans* e *willistoni*.

No grupo *melanogaster* de *Drosophila* (Subgênero: *Sophophora*) atualmente são conhecidas 174 espécies, a maioria distribuída nas regiões Afrotropical ou Oriental. Essas espécies estão distribuídas em 12 subgrupos e são extensivamente utilizadas para diversos estudos em muitos campos da biologia. (Bock e Wheeler, 1972; Ashburner *et al.*, 1984; Lachaise *et al.*, 1988; Schawaroch, 2002). Apesar dos estudos filogenéticos baseados em caracteres morfológicos, biogeografia, estrutura cromossômica e dados moleculares, as relações filogenéticas de alguns subgrupos ainda são controversas, mas todos os trabalhos agrupam as espécies do grupo *melanogaster* em 3 linhagens principais: a) subgrupo *ananassae*; b) subgrupo *montium*; c) subgrupo *melanogaster* e os outros subgrupos orientais (Kastanis *et al.*, 2003; Yang *et al.*, 2004). A figura 4 mostra as relações filogenéticas de várias espécies de diferentes subgrupos do grupo *melanogaster* (Lewis *et al.*, 2005). As espécies que estudamos em nosso trabalho estão grifadas nesta figura.

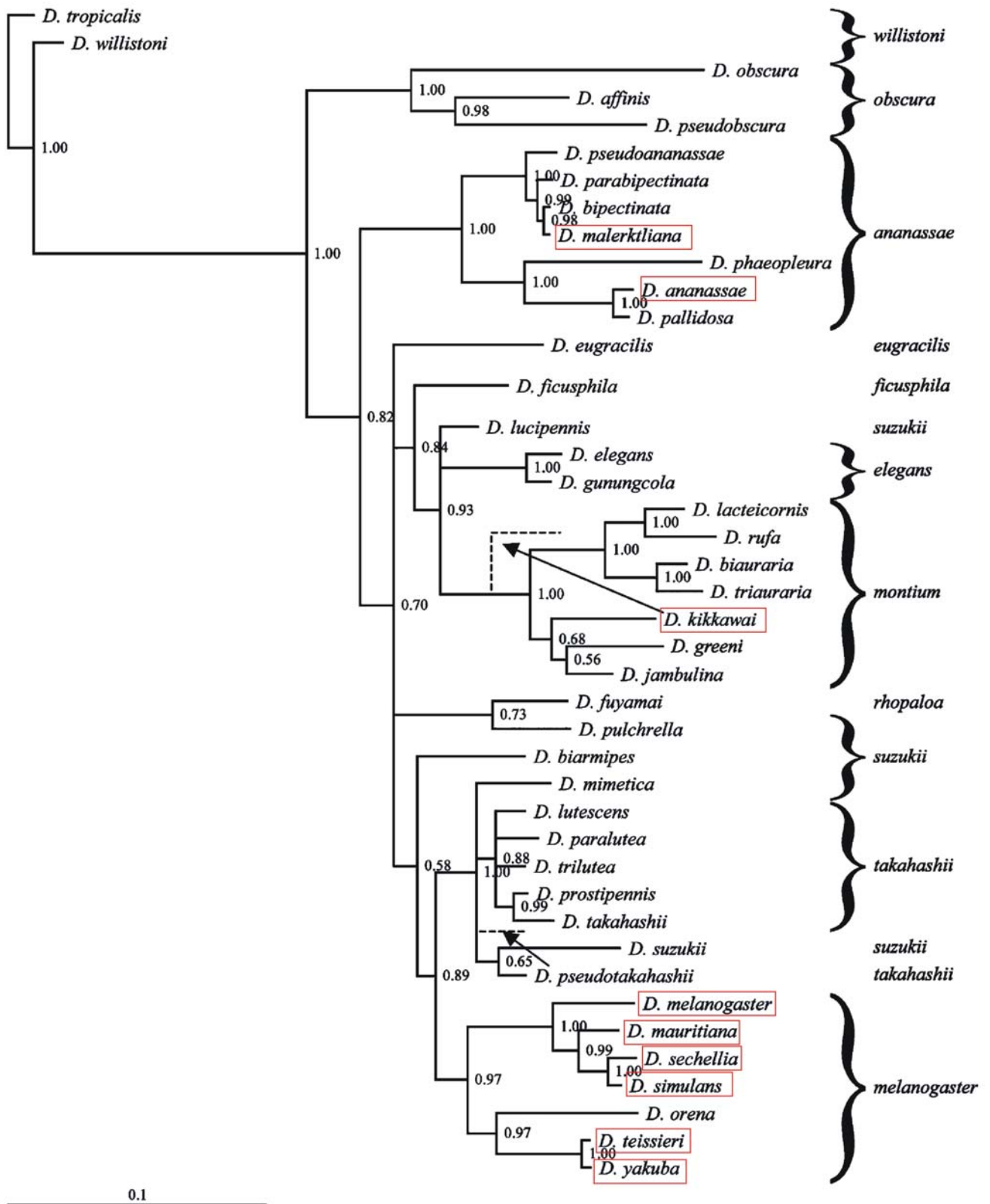


Figura 4: Filogenia por análise Bayesiana com base na análise combinada dos genes citocromo oxidase I e citocromo oxidase II (Lewis *et al.*, 2005). As espécies do grupo *melanogaster* utilizadas neste estudo estão grifadas em vermelho.

Em nosso trabalho estudamos as seguintes espécies do grupo *melanogaster*. 1) *D. ananassae* e *D. malerkotliana* do subgrupo *ananassae*, ambas de origem Oriental. *D. ananassae* é uma espécie de distribuição cosmopolita e *D. malerkotliana* é considerada subcosmopolita; 2) *D. kikkawai* pertence ao subgrupo *montium*, tem origem Oriental e é subcosmopolita; 3) Sete das nove espécies do subgrupo *melanogaster*. *D. yakuba*, *D. teissieri*, *D. erecta*, *D. melanogaster*, *D. sechellia*, *D. mauritiana* e *D. simulans*. Das espécies desse subgrupo estudadas, *D. melanogaster* e *D. simulans* são cosmopolitas e de origem Afrotropical (Bock e Wheeler, 1972; David e Tsacas, 1980).

O subgênero *Drosophila*, por outro lado, é o maior representante da família Drosophilidae em regiões Neotropicais, com 25 grupos de espécies (Val *et al.*, 1981), as quais podem ser divididas em duas linhagens principais (Throckmorton, 1975): a seção virilis-repleta e a seção quinária-tripunctata. A seção virilis-repleta é formada por 15 grupos, dentre os quais, *virilis* e *repleta* possuem espécies utilizadas nesse trabalho, *D. virilis* e *D. mojavensis*, respectivamente.

OBJETIVOS

Esse trabalho visa contribuir para o entendimento da história evolutiva dos retrotransposons com LTRs *297*, *17.6*, *Tom* e *rover* em espécies do gênero *Drosophila* e *Z. Indianus*. Com esta intenção foram traçados os seguintes objetivos específicos:

- 1) Obtenção de seqüências homólogas aos retrotransposons *297*, *17.6*, *Tom* e *rover* nas espécies: *Drosophila malerkotliana*, *D. kikkawai*, *D. mauritiana*, *D. simulans*, *D. sechellia*, *D. teissieri* e *Zaprionus indianus*;
- 2) Busca *in silico* dos retrotransposons *297*, *17.6*, *Tom* e *rover* nos doze genomas de *Drosophila* atualmente disponíveis;
- 3) Inferência das relações evolutivas desses retrotransposons nas diferentes espécies encontradas;
- 4) Análise dos padrões evolutivos dos retrotransposons *297*, *17.6*, *Tom* e *rover* no gênero *Drosophila* e em *Z. Indianus*, comparando-os com genes nucleares.

CAPÍTULO II

História Evolutiva dos LTR-retrotransposons *Tom*, 297, 17.6 e *rover* em *Drosophilidae*

Vidal, Newton de Medeiros¹; Loreto, Élgion Lúcio da Silva^{1,2}

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

2 – Departamento de Biologia, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil.

Manuscrito em preparação para ser enviado à revista **Insect Molecular Biology**.

RESUMO

Os elementos de transposição 297, 17.6 e *rover* de *Drosophila melanogaster* e *Tom* de *D. ananassae* são retrovírus endógenos de insetos filogeneticamente relacionados. Neste trabalho, analisamos a distribuição das seqüências homólogas a estes elementos em 40 espécies de *Drosophila*, *Zaprionus indianus* e *Scaptodrosophila latifasciaeformis*. Estes elementos mostraram distribuição restrita ao grupo *melanogaster*, exceto pela presença do elemento 17.6 em *D. virilis* e *D. mojavensis* e do elemento *rover* em *Z. indianus*. Análises filogenéticas e divergência nos sítios sinônimos de seqüências da transcriptase reversa dos quatro retroelementos indicam que diferentes eventos poderiam explicar a história evolutiva desses elementos, incluindo 16 casos de transmissão horizontal, polimorfismo ancestral e perda estocástica.

INTRODUÇÃO

Os LTR-retrotransposons estão entre os mais abundantes constituintes dos genomas eucarióticos (Havecker *et al.*, 2004). Com base em análises filogenéticas de seqüências da transcriptase reversa esses elementos são divididos em duas linhagens principais: *Ty1/copia* e *Ty3/gypsy* (Xiong e Eickbush, 1990).

A organização genômica dos LTRs-retrotransposons é semelhante a dos retrovírus, consistindo em duas longas repetições terminais (LTRs – *Long Terminal Repeats*) flanqueando uma região interna que contém pelo menos uma das três fases abertas de leitura (ORFs – *Open Reading Frames*): *gag*, *pol* e *env*, que codificam proteínas enzimáticas e estruturais necessárias para o seu ciclo de replicação (Coffin *et al.*, 1997). O gene *pol* codifica uma poliproteína com as funções de protease, transcriptase reversa, RNaseH e integrase (Capy *et al.*, 1998). A transcriptase reversa é o domínio mais conservado entre os retrotransposons, sendo muito utilizada para inferir as relações entre os diferentes elementos.

Os LTR-retrotransposons têm sido formalmente classificados pelo Comitê Internacional em Taxonomia de Vírus (ICTV – *International Committee on Taxonomy of Viruses*) em duas famílias de retroelementos virais: Pseudoviridae (*Ty1/copia*) e Metaviridae (*Ty3/gypsy*). Cada uma dessas famílias possui três gêneros, compostos por diferentes espécies que constituem os diferentes retroelementos (Hull, 2001).

Análises filogenéticas de retroelementos representantes da família Metaviridae, com base nos domínios conservados da transcriptase reversa e RNaseH, demonstraram a existência de linhagens evolutivas distintas. Uma destas linhagens corresponde ao gênero *Errantivirus*, que é composto por diversos elementos encontrados em genomas de insetos (Malik *et al.*, 2000). O *Errantivirus gypsy* de *Drosophila melanogaster* é o mais estudado e possui propriedades infecciosas comprovadas (Kim *et al.*, 1994; Song *et al.*, 1994). Os *Errantivirus*, também chamados de retrovírus endógenos de insetos (IERVs –

Insect Endogenous Retro Viruses ou *Insect ERantiVirus*) (Terzian *et al.*, 2001), são caracterizados pela aquisição de um gene *env* oriundo de baculovírus, que são vírus de DNA de insetos (Malik *et al.*, 2000).

Os IERVs por sua vez podem ser divididos em dois grupos que diferem pela seqüência do sítio de ligação ao primer (PBS – *Primer Binding Site*): IERV-S e IERV-K, que utilizam, respectivamente, tRNA^{Ser} e tRNA^{Lys} para iniciar a transcrição reversa. Essa divisão é suportada por análises filogenéticas dos genes *gag*, *pol* e *env* para treze diferentes IERVs (Terzian *et al.*, 2001).

Dentre os IERVs-S os retroelementos 297 e 17.6 de *D. melanogaster* e *Tom* de *D. ananassae* são proximamente relacionados (Terzian *et al.*, 2001). Pouca ênfase tem sido dada ao estudo da distribuição e evolução desses retrotransposons. A maioria dos estudos dos retroelementos 297 e 17.6 concentra-se na determinação do número de cópias em diferentes populações de *D. melanogaster* e *D. simulans* (Biémont *et al.*, 1999; Vieira *et al.*, 2002) e no genoma seqüenciado de *D. melanogaster* (Bowen e McDonald *et al.*, 2001; Bartolome *et al.*, 2002; Kaminker *et al.*, 2002; Lerat *et al.*, 2003). O retroelemento *Tom* foi estudado em *D. ananassae* como agente causador de mutações Om, que alteram a morfologia do olho (Tanda *et al.*, 1988, 1989, 1994; Juni *et al.*, 1996).

A publicação da seqüência genômica “release 1” de *D. melanogaster* (Adams *et al.* 2000) encorajou muitos estudos de distribuição genômica e abundância dos elementos de transposição nesse organismo. Isso resultou na descoberta e caracterização de novos elementos (Jurka, 2000; Bowen e McDonald, 2001; Bartolome *et al.*, 2002; Kaminker *et al.*, 2002; Rizzon *et al.*, 2002; Kapitonov e Jurka, 2003; Lerat *et al.*, 2003), como é o caso do retroelemento *rover* descrito por Kaminker *et al.* (2002), que possui maior similaridade com 17.6. Nenhum estudo adicional da distribuição do retroelemento *rover* foi feito até o momento.

Neste trabalho, realizamos uma investigação inicial da distribuição de seqüências homólogas aos retroelementos 297, 17.6 e *Tom* em 33 espécies Neotropicais de *Drosophila*, *Zaprionus indianus* e *Scaptodrosophila latifasciaeformis*, utilizando Dot Blot e PCR. A disponibilização das seqüências de

doze genomas de *Drosophila* possibilitou-nos a realização de buscas *in silico* dos retroelementos 297, 17.6, *Tom* e *rover*.

Com o objetivo de contribuir para um melhor entendimento da história evolutiva dos IERVs, realizamos uma análise filogenética de uma região de aproximadamente 1Kb do gene *pol* correspondendo a parte da transcriptase reversa dos elementos 297, 17.6, *Tom* e *rover* em diferentes espécies de *Drosophilidae* pela análise de seqüências obtidas por clonagem e seqüenciamento e também por buscas *in silico* nos genomas.

MATERIAL E MÉTODOS

Manutenção dos estoques de moscas e manipulação de ácidos nucleicos

As moscas foram mantidas em meio de cultura a base de farinha de milho (Klein *et al.*, 1999) e a temperatura constante (20°C). DNA genômico total foi extraído de moscas adultas de acordo com Sassi *et al.* (2005). As manipulações de ácidos nucleicos foram feitas segundo Sambrook *et al.* (1998).

Rastreamento por Dot Blot e PCR

Na busca por seqüências homólogas aos retroelementos 297, 17.6 e *Tom* realizamos um rastreamento em várias espécies de diferentes grupos dos subgêneros *Sophophora*, *Drosophila* e *Dorsilopha* do gênero *Drosophila* e em *Z. indianus* e *S. latifasciaeformis*. Um total de 35 espécies foi analisado, as quais podem ser encontradas na tabela A1 dos anexos.

Um rastreamento inicial foi realizado através de Dot Blot utilizando 1µg de DNA, desnaturado por calor, e aplicado diretamente sobre uma membrana de náilon (Hybond-N⁺ – GE Healthcare). 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 Random Primer Labelling Module (GE Healthcare). As sondas utilizadas foram:

(1) Para o elemento *Tom* a sonda foi o plasmídeo pXTom2, contendo o elemento *Tom*, gentilmente cedido pelo Dr. K. Yoshida.

(2) Para os elementos 297 e 17.6 a sonda foi o fragmento de PCR amplificado com os *primers* T12A e T12B a partir do DNA genômico de *D. melanogaster*. Esses *primers*, os quais são degenerados, foram projetados a partir do alinhamento das seqüências dos elementos canônicos: 297 e 17.6 de *D. melanogaster* (números de acesso no GenBank X03431 e X01472, respectivamente) e *Tom* de *D. ananassae* (número de acesso no GenBank Z24451). O amplicon esperado é de aproximadamente 1Kb (posições 3246-4229

do 297, 4009-4989 do 17.6 e 3185-4168 do *Tom*) e corresponde a uma região da transcriptase reversa. Os *primers* são: T12A 5'AGTWTGGGCSACAAARAC3' e T12B 5'CCGTCYCTRTCYGCCTTT3'. Para cada 50µl de reação da PCR foram utilizados: 5µl de tampão de reação, 100ng de DNA, 1U de Taq Polimerase (Invitrogen), 2,5mM de MgCl₂, 200µM de cada dNTP, 20pmol de cada *primer*. As condições de amplificação foram: desnaturação inicial a 95°C por 5min, seguida por 30 ciclos de 45s a 94°C, 45s a 55°C e 1min a 72°C, e finalmente 5min a 72°C para completar a extensão.

Para as mesmas espécies analisadas por Dot Blot foi realizada também uma análise por PCR utilizando os *primers* T12A e T12B, que pode amplificar os três retroelementos, nas mesmas condições descritas acima.

PCR, clonagem e seqüenciamento

Para as espécies *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. teissieri*, *D. malerkotliana*, *D. kikkawai* e *Zaprionus indianus* (tabela A2 dos anexos) a região amplificada com os *primers* T12A e T12B foi clonada em vetor pUC18/SmaI. O seqüenciamento dos clones foi realizado diretamente dos plasmídeos purificados em seqüenciador automático MegaBace 500. A reação de terminação de cadeia foi implementada usando o kit DYEnamic ET (GE Healthcare). Ambas as fitas foram seqüenciadas. O programa Phred (Ewing *et al.*, 1998) foi utilizado para análise da confiabilidade das seqüências. Para cada clone uma seqüência consenso foi gerada pelo uso do programa GAP4 do Pacote Staden (Staden, 1996). O programa CENSOR (<http://www.girinst.org/censor/index.php>; Kohany *et al.*, 2006) – que faz uma busca das seqüências *query* contra a coleção de seqüências de DNA repetitivo de diferentes espécies eucarióticas na versão mais recente do Repbase Update (<http://www.girinst.org/repbase/index.html>; Jurka *et al.*, 2005) – foi utilizado para confirmar a identidade de todos os clones.

Busca nos genomas

Buscas por seqüências homólogas aos retrotransposons 297, 17.6, *Tom* e *rover* foram feitas nas espécies: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. virilis*, *D. mojavensis* e *D. grimshawi* utilizando a ferramenta BLAT (<http://genome.ucsc.edu/cgi-bin/hgBlat>; Kent, 2002) e em *D. willistoni* utilizando a ferramenta BLAST (<http://flybase.bio.indiana.edu/blast/>; Grumblin e Strelets, 2006). As seqüências utilizadas como *queries* foram uma região de aproximadamente 1Kb da transcriptase reversa dos elementos canônicos 297, 17.6 e *rover* (número de acesso AF492764) de *D. melanogaster* e *Tom* de *D. ananassae*. Essa região corresponde à mesma porção obtida por amplificação pelos *primers* T12A e T12B.

Para cada *query*, todas as seqüências encontradas nos diferentes genomas com score > 200 foram utilizadas para formar um conjunto de dados. O programa CENSOR (Kohany *et al.*, 2006) foi utilizado para confirmar a identidade de todas as seqüências encontradas e evitar que a mesma seqüência fosse obtida com as diferentes *queries*.

Para manipulação e edição das seqüências foi utilizado o programa GeneDoc 2.6.001 (Nicholas e Nicholas, 1997). As seqüências encontradas foram nomeadas da seguinte forma: uma letra ou número inicial que representa a *query* utilizada (2 para 297, 1 para 17.6, r para *rover* e t para *Tom*), seguido por três letras (iniciais da espécie) referentes a espécie em que foi encontrada, acrescida de um número para cada seqüência diferente.

Análise filogenética

Para as análises filogenéticas foram utilizadas as seqüências encontradas nas buscas *in silico* e as seqüências dos clones (as seqüências estão disponíveis em <http://www.ufsm.br/labdros/links/seqstom>).

Para estabelecer as relações entre as diferentes famílias de retrotransposons construímos uma filogenia com base nas seqüências de

aminoácidos, que foram deduzidas utilizando o programa GeneDoc 2.6.001 (Nicholas e Nicholas, 1997). Aquelas seqüências que permaneceram na fase de leitura correta e sem códons de terminação prematuros foram alinhadas pelo ClustalX 1.83 (Jeanmougin *et al.*, 1998). Dois métodos de inferência filogenética foram utilizados: (1) Neighbor Joining (NJ) (Saitou e Nei, 1987) usando distância *p* e com *bootstrap* de 1000 réplicas no programa Mega 3.1 (Kumar *et al.*, 2004); (2) Máxima Parcimônia (MP) com parâmetros *default* e *bootstrap* de 500 réplicas no programa Mega 2.1 (Kumar *et al.*, 2001).

Para as análises filogenéticas das seqüências de nucleotídeos quatro conjuntos de dados foram utilizados, representando os diferentes elementos. Cada conjunto de seqüências foi alinhado separadamente pelo programa ClustalX 1.83 (Jeanmougin *et al.*, 1998). Inicialmente utilizamos o programa MrModelTest 2.2 (Nylander, 2004) seguindo o critério de informação Akaike (AIC) para escolher o modelo evolutivo mais adequado para cada conjunto de seqüências. Para os elementos *297*, *17.6* e *rover* o modelo indicado foi GTR+G e para o elemento *Tom* foi GTR+G+I. Dois métodos foram utilizados para inferência filogenética: (1) NJ (Saitou e Nei, 1987) utilizando o modelo Tamura-Nei com valor de distribuição gama indicado pelo MrModelTest 2.2 (Nylander, 2004), com *bootstrap* de 1000 réplicas usando o programa Mega 3.1 (Kumar *et al.*, 2004); (2) Análise Bayesiana utilizando o programa MrBayes 3.1.2 (Ronquist e Huelsenbeck, 2003), com a avaliação de no mínimo 2.000.000 de gerações e uma região de “burnin” de 500 árvores utilizando o modelo indicado pelo MrModelTest 2.2 (Nylander, 2004) para os diferentes elementos. A probabilidade posterior de cada clado foi obtida pela regra de maioria de 50% do consenso das árvores (Hall, 2001).

Distância *p* entre as seqüências foi calculada pelo programa Mega 3.1 (Kumar *et al.*, 2004) e utilizada para determinar grupos e subfamílias dos elementos.

Tanto para a análise de seqüências nucleotídicas como de aminoácidos foram utilizados como grupo externo os retrotransposons canônicos *Idefix* e *Quasimodo* de *D. melanogaster* (números de acesso AJ009736 e AF364550, respectivamente) e *Tv1* de *D. virilis* (número de acesso AF056940).

Estimativa de dN e dS e uso preferencial de códons

Para cada um dos quatro conjuntos de seqüências um alinhamento de códons foi empregado para estimar o número de substituições não sinônimas por sítio não sinônimo (dN), o número de substituições sinônimas por sítio sinônimo (dS), e o número de sítios sinônimos (S) pelo método de Nei e Gojobori (1986), utilizando o programa Mega 3.1 (Kumar *et al.*, 2004). Para permitir o cálculo de dN e dS em algumas seqüências foram introduzidos indels para preservar o quadro de leitura e os códons de parada foram considerados como informação ausente.

Dois índices foram utilizados para estimar o uso preferencial de códons: o número efetivo de códons (Nc; Wright, 1990); e o índice de uso preferencial de códons (CBI; Morton, 1993), utilizando o programa DnaSP versão 4.0 (Rozas *et al.*, 2003).

Inferência de transmissão horizontal

Quando elementos de transposição (TEs – *Transposable Elements*) são transmitidos verticalmente sua história filogenética deve refletir a história evolutiva das espécies hospedeiras (Silva *et al.*, 2004). Segundo Silva *et al.* (2004) três métodos são geralmente utilizados para detectar transmissão horizontal (TH) de TEs: (1) detecção de elementos com alto grau de similaridade de seqüência, com divergência muito menor do que a divergência entre genes nucleares das suas respectivas espécies hospedeiras; (2) detecção de diferenças topológicas entre a filogenia do TE e das espécies hospedeiras; (3) distribuição descontínua dos elementos entre táxons proximamente relacionados.

Neste trabalho, foram utilizados os critérios acima, sendo que o nível de divergência dos TEs e genes nucleares foi medido com base na divergência encontrada nos sítios sinônimos (dS). Se um TE foi transmitido de uma espécie para outra por TH, é esperado que o dS do TE entre as espécies seja

significativamente menor que os dS encontrado para genes nucleares, refletindo o menor tempo de divergência das seqüências em relação a seqüência ancestral.

Para cada possível evento de TH, empregamos um teste qui-quadrado, com correção de Yates, para verificar se o número observado de similaridades e diferenças para o TE nos sítios sinônimos (calculadas com base nos valores de dS e S do TE) é significativamente menor que o número esperado de similaridades e diferenças para o TE nos sítios sinônimos (calculadas com base no valor de dS do gene nuclear e S do TE).

Levando em conta que a taxa de substituição sinônima em *Drosophila* é negativamente correlacionada com o grau de uso preferencial de códons (Sharp e Li, 1989) consideramos os valores de CBI calculados para os TEs e para os genes nucleares para escolher o melhor gene para fazer as comparações de dS e evitar assim subestimativas e/ou superestimativas do número de similaridades e diferenças esperadas para o TE.

As seqüências de DNA de quatro genes nucleares foram utilizadas neste estudo: *alfa metil-dopa (amd)*, *cdc6*, *álcool desidrogenase (adh)* e *superóxido dismutase (sod)*. Para cada um desses genes foram estimados os valores de dN, dS, Nc e CBI conforme descrito acima. Informações adicionais e número de acesso das seqüências desses genes no GenBank estão na tabela A3 dos anexos.

Estimativa de tempo de ocorrência de TH

Utilizamos a taxa de substituição sinônima de 0.016 substituições por sítio por milhão de anos calculada para genes com baixo uso preferencial de códons de *Drosophila* (Sharp e Li, 1989) e os valores de dS calculados para os elementos para estimar o tempo de ocorrência das THs. Seguindo a fórmula: $T=k/2r$ (Graur e Li, 2000), onde o T é o tempo de divergência entre as seqüências, k é o número de substituições entre os TEs (dS) e r é a taxa de evolução (0.016).

RESULTADOS

Dot Blot e PCR

Encontramos seqüências homólogas ao retroelemento *Tom*, por Dot Blot, com sinal forte em *D. ananassae* e *D. malerkotliana* (subgrupo *ananassae*) e *D. kikkawai* (subgrupo *montium*), espécies do subgênero *Sophophora*. Em duas linhagens de espécies do subgênero *Drosophila*, *D. immigrans* (Porto Alegre) e *D. polymorpha* (Turvo), foi detectado um sinal mais fraco.

Na análise por Dot Blot utilizando como sonda o produto de PCR de *D. melanogaster*, esperando encontrar 297 e/ou 17.6, foi detectado um sinal forte em *D. melanogaster* e nas três linhagens de *D. simulans* (ambas espécies do subgrupo *melanogaster*), e em *D. kikkawai* (subgrupo *montium*). Também foi detectado um sinal fraco em *S. latifasciaeformis*.

Na análise por PCR utilizando os *primers* projetados para amplificar os três retroelementos obtivemos amplicons para aquelas espécies que apresentaram sinal forte no Dot Blot, *D. ananassae*, *D. malerkotliana*, *D. kikkawai*, *D. melanogaster* e *D. simulans*. Além disso, obtivemos amplificação para *Z. indianus*. Os resultados deste rastreamento inicial podem ser vistos na tabela A1 dos anexos.

Clonagem e seqüenciamento

Após a análise inicial da distribuição dos retroelementos 297, 17.6 e *Tom*, os amplicons obtidos em seis espécies do grupo *melanogaster* e em *Z. indianus* (tabela A2 dos anexos) foram clonados e seqüenciados.

Para *D. simulans* quatro clones foram obtidos, *sim_clone1* e *sim_clone2* correspondendo ao retroelemento 17.6 e *sim_clone3* e *sim_clone4* correspondendo ao retroelemento 297.

Dos cinco clones obtidos para *D. sechellia*, *sec_clone2* e *sec_clone5* correspondem ao retroelemento 17.6 e *sec_clone1*, *sec_clone3* e *sec_clone4* correspondem ao retroelemento 297.

Para *D. mauritiana* os cinco clones obtidos correspondem ao retroelemento *rover*: *mau_clone1*, *mau_clone2*, *mau_clone3*, *mau_clone4* e *mau_clone5*.

Da mesma forma, para *D. teissieri*, os seis clones obtidos também correspondem ao retroelemento *rover*: *tei_clone1*, *tei_clone2*, *tei_clone3*, *tei_clone4*, *tei_clone5* e *tei_clone6*.

Para *D. malerkotliana* quatro clones foram obtidos: *mal_clone1*, *mal_clone2*, *mal_clone3* e *mal_clone4*, correspondendo ao retroelemento *Tom*.

Para *D. kikkawai* foram obtidos seis clones, *kik_clone1*, *kik_clone2*, *kik_clone3* e *kik_clone4* correspondendo ao retroelemento *Tom*, *kik_clone5* correspondendo ao retroelemento *rover*, *kik_clone6* correspondendo ao retroelemento 297.

Os quatro clones obtidos para *Z. indianus*, *zap_clone1*, *zap_clone2*, *zap_clone3* e *zap_clone4* correspondem ao retroelemento *rover*.

Buscas nos genomas

Não foram encontradas seqüências homólogas aos retroelementos 297, 17.6, *rover* e *Tom* em *D. pseudoobscura*, *D. persimilis*, *D. willistoni* e *D. grimshawi*. Nas demais espécies, pelo menos uma seqüência foi encontrada. O retroelemento *Tom* foi encontrado apenas em *D. ananassae*. Já o elemento *rover* foi encontrado em todas as espécies analisadas do subgrupo *melanogaster*: *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba* e *D. erecta*, não sendo encontrado nas demais espécies. O elemento 297 também foi encontrado em todas as espécies analisadas do subgrupo *melanogaster* e em *D. ananassae*. Assim como *rover* e 297, o elemento 17.6 foi encontrado em todas as espécies analisadas do subgrupo *melanogaster*. Adicionalmente, seqüências homólogas a 17.6 foram encontradas nas duas espécies do subgênero *Drosophila*: *D. virilis* e *D. mojavensis*.

Análise filogenética

As seqüências nucleotídicas do gene *pol* de 13 espécies (tabela 1) utilizadas para as inferências filogenéticas incluíram: 197 seqüências obtidas com a busca nos genomas, 34 seqüências de clones e sete TEs canônicos, 297, 17.6, *rover*, *Idefix* e *Quasimodo* de *D. melanogaster*, *Tom* de *D. ananassae* e *Tv1* de *D. virilis*.

Tabela 1: Lista de espécies de Drosophilidae utilizadas para análise filogenética.

	Gênero	Subgênero	Grupo	Subgrupo	Espécie	Número de seqüências
1	<i>Drosophila</i>	<i>Drosophila</i>	<i>virilis</i>		<i>D. virilis</i>	10 ^{b,c}
2			<i>repleta</i>		<i>D. mojavensis</i>	12 ^c
3		<i>Sophophora</i>	<i>melanogaster</i>	<i>ananassae</i>	<i>D. ananassae</i>	12 ^{b,c}
4					<i>D. malerkotliana</i>	4 ^a
5				<i>montium</i>	<i>D. kikkawai</i>	6 ^a
6				<i>melanogaster</i>	<i>D. simulans</i>	21 ^{a,c}
7					<i>D. mauritiana</i>	5 ^a
8					<i>D. sechellia</i>	55 ^{a,c}
9					<i>D. melanogaster</i>	53 ^{b,c}
10					<i>D. erecta</i>	38 ^c
11					<i>D. yakuba</i>	12 ^c
12					<i>D. teissieri</i>	6 ^a
13	<i>Zaprionus</i>				<i>Z. indianus</i>	4 ^a
					TOTAL	238

^a espécies cujas seqüências foram obtidas nesse trabalho através de clonagem e seqüenciamento;

^b espécies cujas seqüências foram obtidas pelo GenBank;

^c espécies cujas seqüências foram obtidas nesse trabalho pela busca nos genomas.

A região da transcriptase reversa analisada corresponde a uma região potencialmente capaz de codificar 328 aminoácidos. Um total de 89 seqüências (37% das seqüências analisadas) está na fase de leitura correta e não possui códons de parada prematuros. As filogenias de aminoácidos obtidas por MP e NJ apresentaram topologias similares, diferindo apenas nos valores de *bootstrap* de alguns clados. A figura 1 mostra a árvore de NJ, que pode ser dividida em dois clados principais. Um agrupando com alta confiabilidade (*bootstrap* de 100 para NJ e MP) as seqüências homólogas aos retroelementos *Tom* e 297. E o outro

agrupando as seqüências homólogas aos retroelementos 17.6 e *rover*, porém com menor confiabilidade (bootstrap de 64 para NJ e 49 para MP). Os grupos apresentados na figura foram estabelecidos conforme descrito abaixo.

Para as análises filogenéticas das seqüências nucleotídicas foram feitos quatro conjuntos de dados, um para cada família de retroelementos, 297 (93 seqüências), 17.6 (74 seqüências), *rover* (67 seqüências) e *Tom* (25 seqüências). Para os quatro retroelementos, as filogenias construídas por NJ e análise Bayesiana apresentaram topologia similares, porém alguns clados ficaram melhor resolvidos com a análise Bayesiana. As figuras 2, 3, 4 e 5 mostram as filogenias obtidas por análise Bayesiana dos retroelementos *Tom*, *rover*, 297 e 17.6, respectivamente. As relações entre as diferentes famílias de TEs estabelecidas pelas filogenias de nucleotídeos foram congruentes com as estabelecidas pela filogenia de aminoácidos. Em cada uma das filogenias os clados foram divididos em grupos de acordo com o grau de divergência das seqüências de nucleotídeos.

Calculamos a divergência média de nucleotídeos e aminoácidos dentre e entre todos os grupos, que é mostrada na tabela 2. Como pode ser observada, a divergência média de nucleotídeos entre grupos varia entre 21% (entre *roverA* e *roverB*) e 33.7% (entre *Tom* e *roverB*), e divergência média de aminoácidos entre grupos varia de 12.5% (entre *roverA* e *roverB*) a 34.6% (entre *roverA* e 297B).

Tabela 2: Porcentagem de divergência de nucleotídeos (abaixo e a esquerda na diagonal) e de aminoácidos (acima e a direita na diagonal) encontrada entre e dentre grupos dos retroelementos *Tom*, *rover*, 297 e 17.6.

	<i>Tom</i>	<i>rover A</i>	<i>rover B</i>	297 A	297 B	17.6 A	17.6 B
<i>Tom</i>	4.7/0.2	31.3	27.5	18.8	21.1	30.0	28.4
<i>rover A</i>	33.6	12.1/4.7	12.5	33.0	34.6	28.8	24.6
<i>rover B</i>	33.7	21.0	5.7/4.5	28.4	30.5	26.1	23.3
297 A	26.3	30.8	32.5	4.8/1.4	17.9	29.5	28.5
297B	25.0	33.2	32.8	21.4	1.8/4.2	29.7	30.1
17.6 A	32.6	31.1	32.0	33.5	33.9	11.1/4.6	21.2
17.6B	33.0	28.9	29.6	32.1	31.3	25.5	6.4/0.3

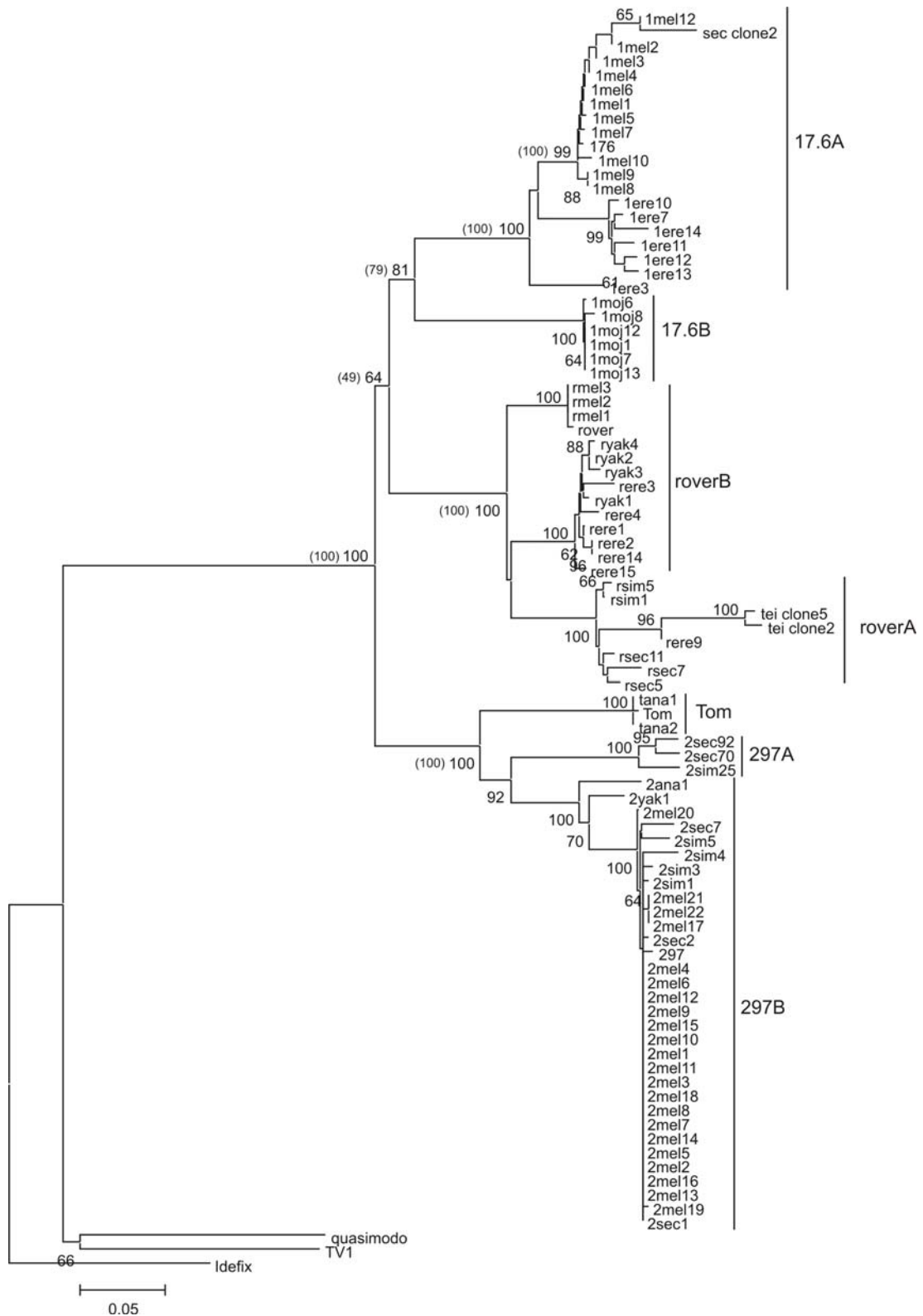


Figura 1: Filogenia de NJ (distância p) das seqüências de aminoácidos de uma porção de aproximadamente 1Kb da transcriptase reversa de seqüências homólogas aos retroelementos 17.6, rover, Tom e 297 apresentando os grupos estabelecidos neste trabalho. Os valores de *bootstrap* (1000 réplicas) menores de 60% não são mostrados. Os valores de *bootstrap* dos principais ramos obtidos por MP (500 réplicas) estão entre parênteses.

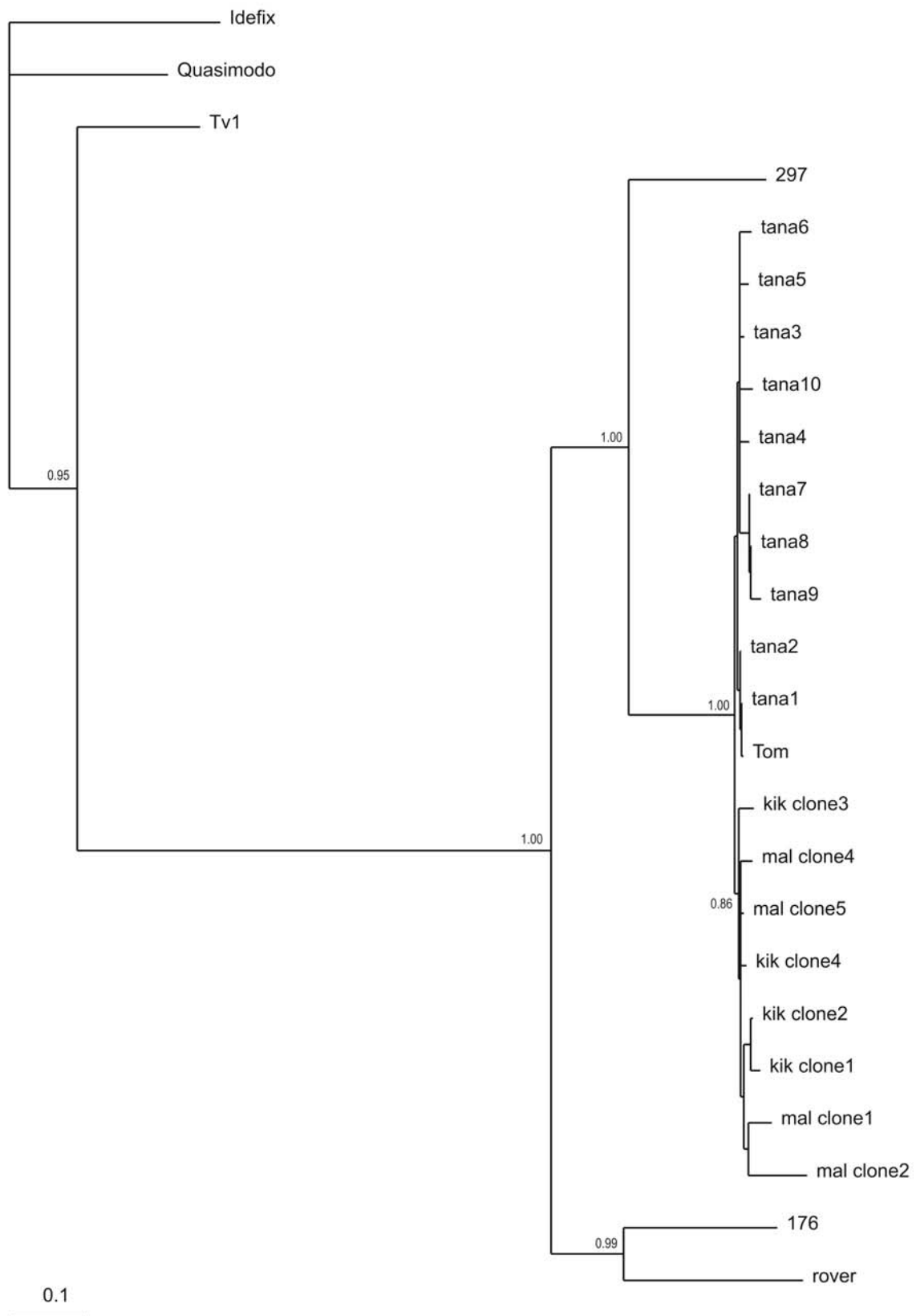


Figura 2: Filogenia obtida por análise Bayesiana e o modelo GTR+G+I para uma porção de aproximadamente 1Kb da transcriptase reversa de seqüências homólogas ao retroelemento *Tom*. A probabilidade posterior dos principais clados está indicada.

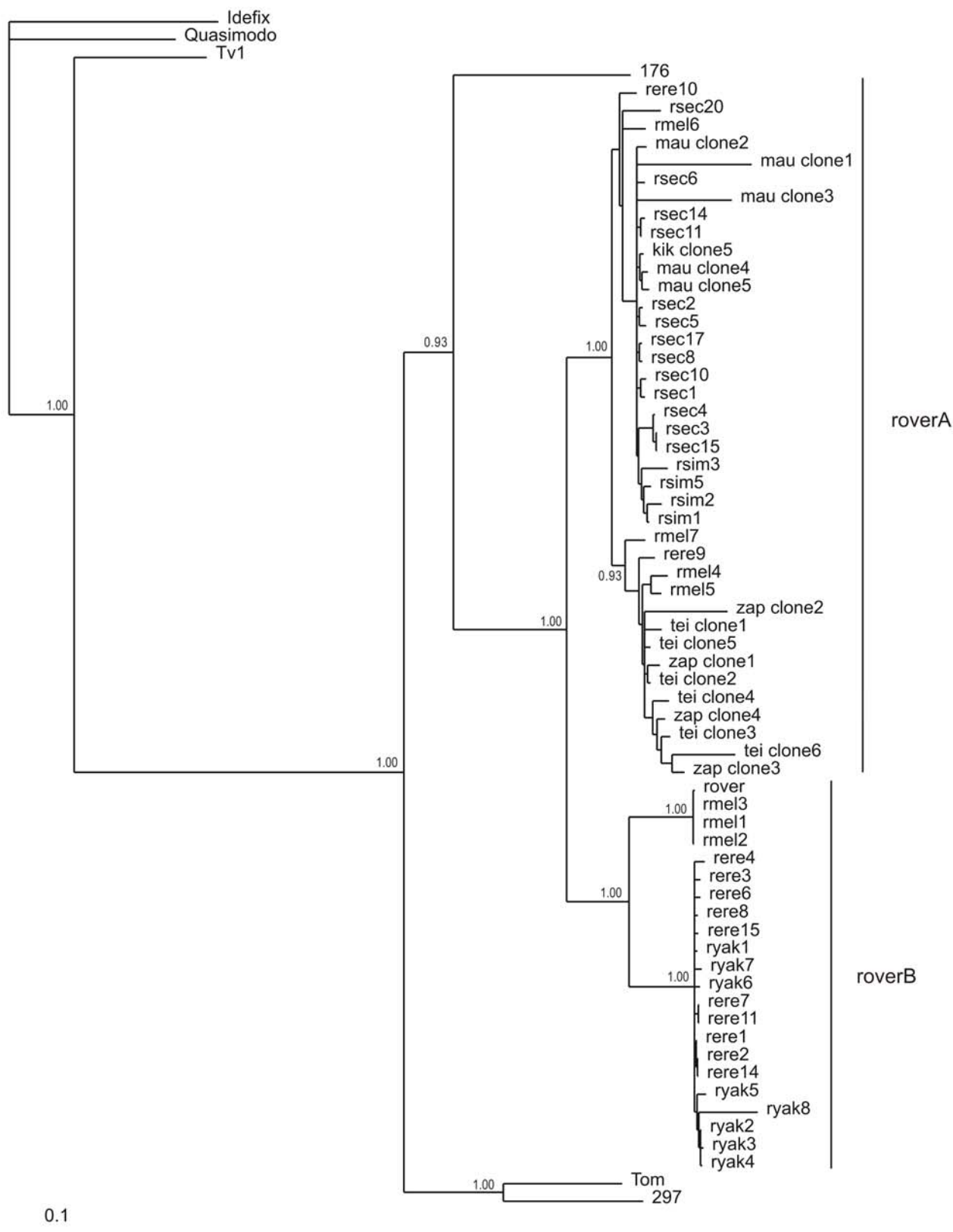


Figura 3: Filogenia obtida por análise Bayesiana e o modelo GTR+G para uma porção de aproximadamente 1Kb da transcriptase reversa de seqüências homólogas ao retroelemento *rover*, indicando os dois grupos estabelecidos. A probabilidade posterior dos principais clados está indicada.

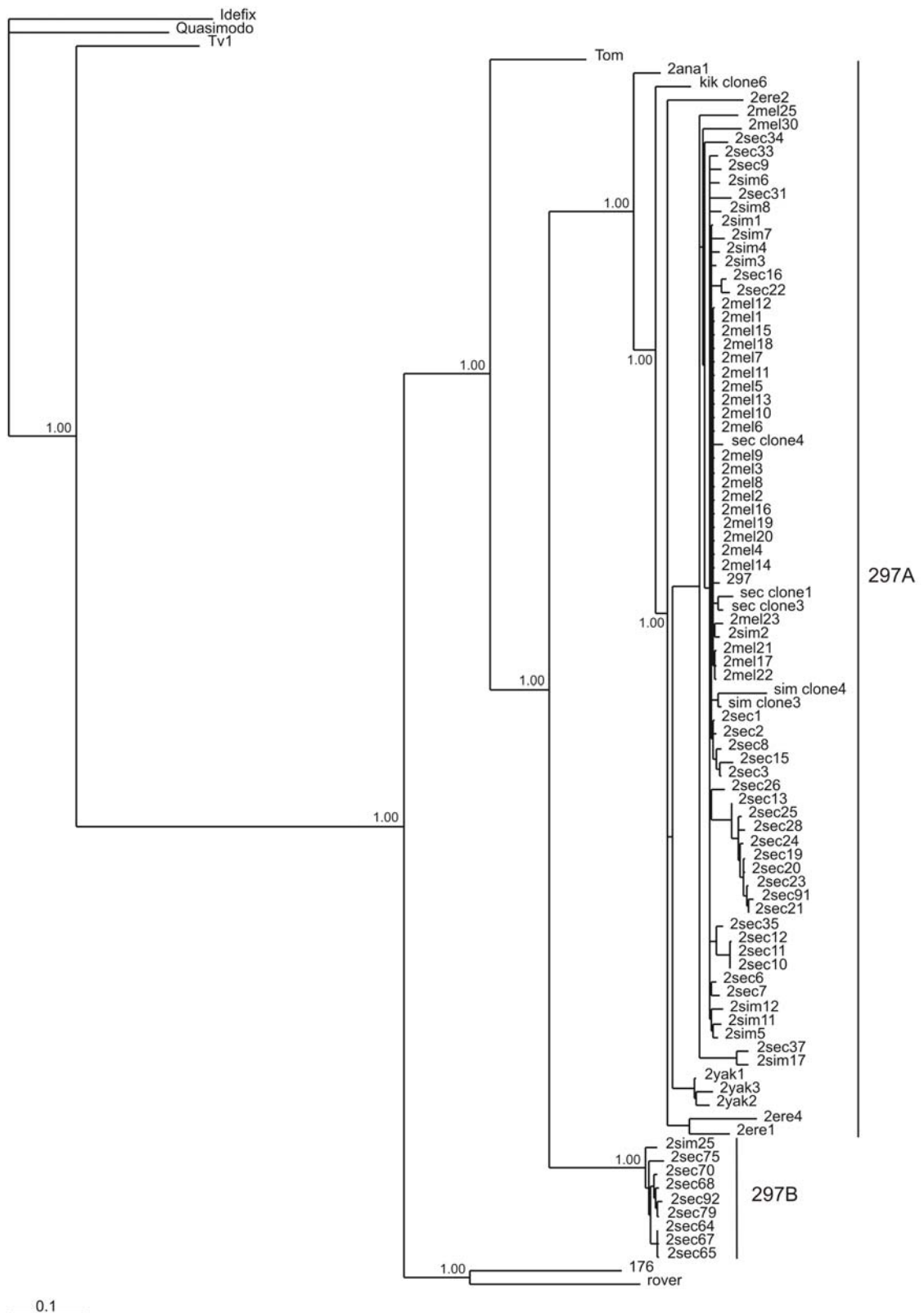


Figura 4: Filogenia obtida por análise Bayesiana e o modelo GTR+G para para uma porção de aproximadamente 1Kb da transcriptase reversa de seqüências homólogas ao retroelemento 297, indicando os dois grupos estabelecidos. A probabilidade posterior dos principais clados está indicada.

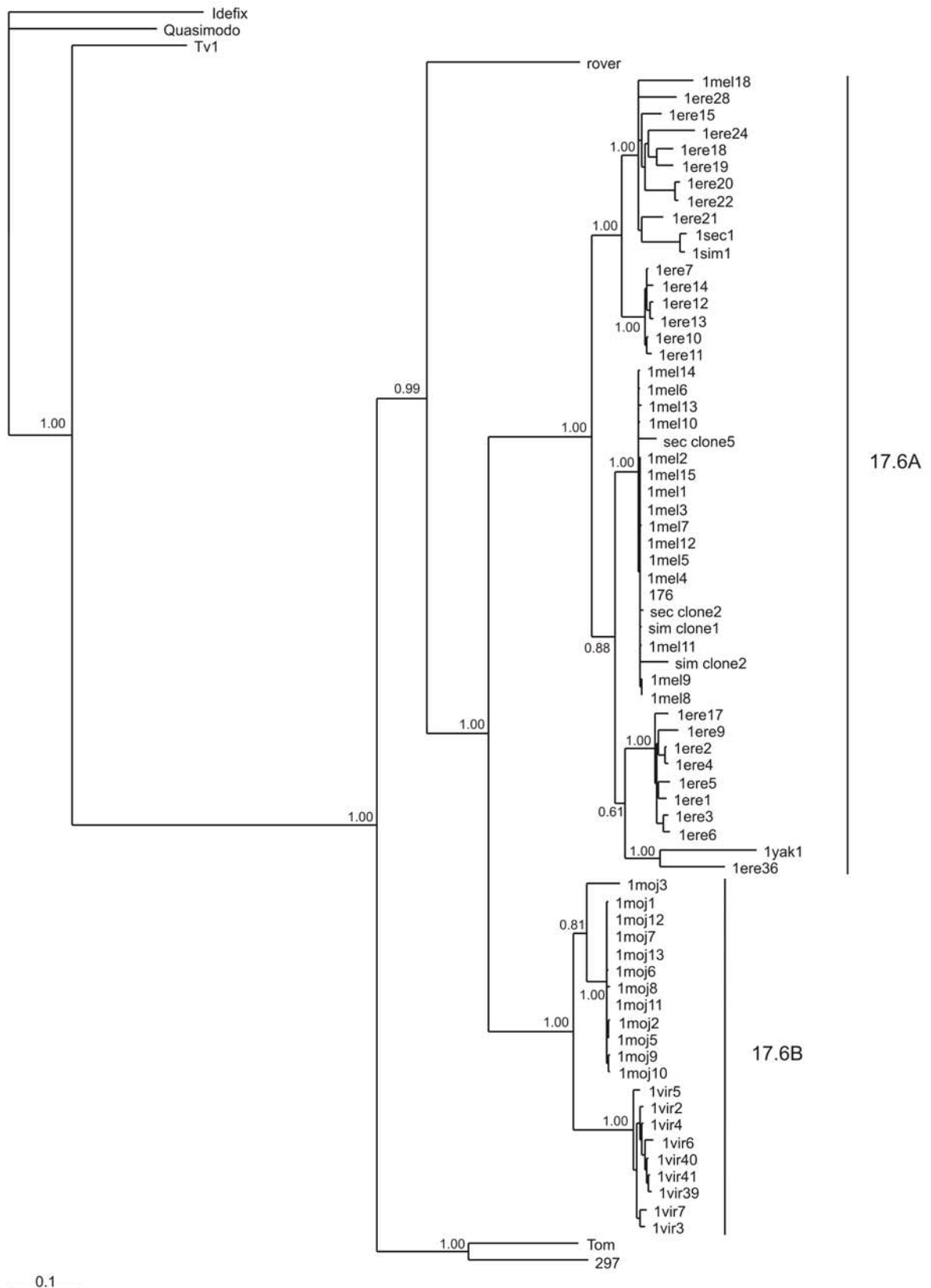


Figura 5: Filogenia obtida por análise Bayesiana e o modelo GTR+G para uma porção de aproximadamente 1Kb da transcriptase reversa de seqüências homólogas ao retroelemento 17.6, indicando os dois grupos estabelecidos. A probabilidade posterior dos principais clados está indicada.

Buscando entender a evolução dos retroelementos *Tom*, *rover*, *297* e *17.6*, investigamos se as filogenias dos TEs estão de acordo com as relações evolutivas das espécies hospedeiras. Um esquema dessas relações está apresentado na figura 6.

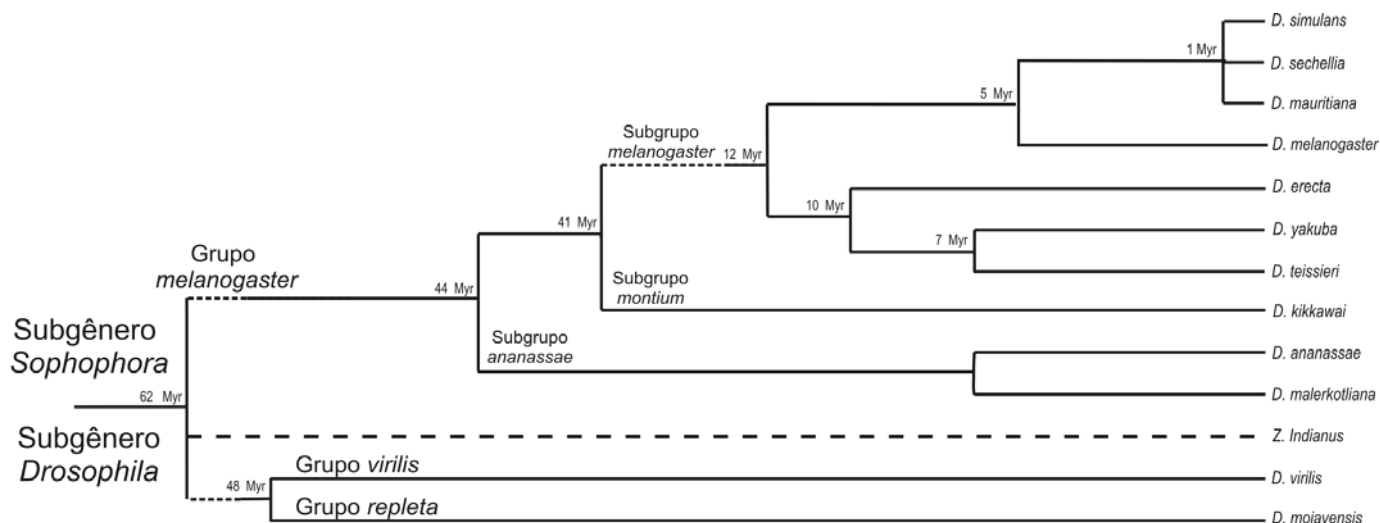


Figura 6: Esquema das relações evolutivas das espécies hospedeiras (com base em Lewis *et al.*, 2005 e Robe *et al.*, 2005) e tempo estimado de divergência das espécies (Ross *et al.*, 2003; Tamura *et al.*, 2004).

A filogenia do retroelemento *Tom* (figura 2) é formada por seqüências muito similares de *D. ananassae*, *D. malerkotliana* e *D. kikkawai* (divergência média de 4.7%), não apresentando mais de um grupo. Como podemos verificar na figura 6, *D. malerkotliana* é mais próxima filogeneticamente a *D. ananassae* (que pertencem ao subgrupo *ananassae*) do que a *D. kikkawai* (subgrupo *montium*). O contrário é observado na filogenia dos TEs. Além disso, o comprimento de todos os ramos é muito curto.

A filogenia do *rover* (figura 3) mostra a presença de dois grupos: *roverA* e *roverB*. O grupo *roverA* ainda pode ser dividido em dois clados principais: (1) um contendo seqüências de *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. melanogaster* e *D. erecta* que são espécies proximamente relacionadas. Porém, transmissão vertical (TV) dessas seqüências foi testada por outra abordagem (ver abaixo), já que o comprimento de alguns ramos é curto. Neste mesmo clado existe uma

grande incongruência, pela presença de uma seqüência de *D. kikkawai* (kik_clone5) proximamente relacionada a seqüências de *D. mauritiana*; (2) o outro clado possui seqüências de *D. teissieri* e *D. erecta* (rere9) cuja divergência com as seqüências do primeiro clado estão de acordo com transmissão vertical. Esse segundo clado também possui seqüências de *D. melanogaster* entre rere9 e as seqüências de *D. teissieri*. O comprimento dos ramos relacionando as seqüências dessas três espécies é curto, levantando a hipótese de TH, a qual foi testada. Outra grande incongruência no grupo roverA é a presença de seqüências de *Z. indianus* próximas as seqüências de *D. teissieri*. Ainda podemos observar a ausência de seqüências desse grupo em *D. yakuba*, que pode estar refletindo sua perda nessa espécie. O grupo roverB também pode ser dividido em dois cladros principais: (1) um formado apenas por seqüências de *D. melanogaster* (incluindo o elemento rover canônico), não estando presente nas espécies relacionadas, *D. simulans* e *D. sechellia*; (2) o outro clado mostra uma forte incongruência, com várias seqüências intercaladas de *D. erecta* e *D. yakuba* com ramos bastante curtos.

A filogenia do 297 (figura 4) pode ser dividida em dois grupos: (1) 297A formado por uma grande quantidade de seqüências intercaladas, muito similares, de *D. melanogaster*, *D. simulans* e *D. sechellia*. Este grupo possui ainda seqüências de *D. erecta* e *D. yakuba*, cujas divergências com as seqüências das outras espécies do subgrupo *melanogaster* estão de acordo com transmissão vertical. Duas outras incongruências são encontradas nesse grupo, a presença de uma seqüência de *D. kikkawai* (kik_clone6) e uma de *D. ananassae* (2ana1) as quais apresentam pequena divergência entre si e entre as seqüências das outras espécies; (2) 297B possui seqüências de *D. sechellia* e de *D. simulans* com comprimento de ramos condizentes com transmissão vertical. Porém seqüências desse grupo não estão presentes nas outras espécies do subgrupo *melanogaster*.

A filogenia do 17.6 (figura 5) também pode ser dividida em dois grupos: 17.6A que incluem seqüências de espécies do subgrupo *melanogaster* do subgênero *Sophophora* e 17.6B com seqüências de *D. virilis* (grupo *virilis*) e *D. mojavensis* (grupo *repleta*), ambas do subgênero *Drosophila*.

O grupo 17.6A pode ser dividido em dois grandes clados: (1) um com várias seqüências de *D. erecta* e uma seqüência de *D. melanogaster* (1mel18), de *D. simulans* (1sim1) e de *D. sechellia* (1sec1). Podemos notar neste clado o curto comprimento dos ramos que agrupam as seqüências 1sim1 e 1sec1. (2) o outro clado agrupa seqüências de *D. melanogaster*, *D. simulans* e *D. sechellia* com ramos bastante curtos, seqüências de *D. erecta* e uma seqüência de *D. yakuba*.

O grupo 17.6B é formado por dois clados: um com seqüências de *D. mojavensis* e outro com seqüências de *D. virilis*. Essas espécies pertencem a diferentes grupos do subgênero *Drosophila*. Desta forma, testamos se estas seqüências evoluíram por TV ou TH.

Apesar da divergência encontrada entre os grupos 17.6B e 17.6A (25.5%) não estar em desacordo com o esperado para TV, testamos a possibilidade de TH, pois a existência do grupo 17.6B mostra uma clara descontinuidade na distribuição das seqüências homólogas ao retroelemento 17.6 nas espécies de *Drosophila*.

Por fim, cabe salientar que incongruências relacionadas à ausência de um elemento em uma espécie cujas seqüências foram obtidas por clonagem não foram consideradas, pois podem estar representando um viés de amostragem. Entretanto, aquelas incongruências relacionadas à ausência de seqüências nas espécies com genoma disponível foram consideradas, pois a busca foi exaustiva, embora essa ausência possa ser característica da linhagem que foi seqüenciada.

Intensidade de seleção nos retroelementos

A intensidade da pressão de seleção nas seqüências dos retroelementos *Tom*, *rover*, *297* e *17.6* foi medida pela razão de dN/dS e comparada com a pressão exercida sobre os quatro genes nucleares analisados neste trabalho. Os genes *adh* (envolvido na metabolização do etanol), *sod* (que codifica uma enzima antioxidante), *amd* (envolvido na pigmentação e esclerotização da cutícula) e *cdc6* (fator de início da replicação) desempenham importantes funções no organismo. Dessa forma é esperada uma forte pressão seletiva atuando sobre esses genes, o que pode ser verificado na figura 7. Os valores de dN/dS variam

de 0.054 para *cdc6* a 0.144 para *adh*. Os retroelementos *rover*, *297* e *17.6* possuem um valor de dN/dS (0.130, 0.149 e 0.150, respectivamente) similar ao dos genes nucleares, indicando que esses elementos, potencialmente codificadores, estão sofrendo uma forte pressão seletiva nos genomas. O retroelemento *Tom* apresentou um valor de dN/dS maior (0.330), porém foram utilizadas para o cálculo apenas três seqüências codificadoras de *D. ananassae*, que diferem em uma substituição sinônima e outra não sinônima.

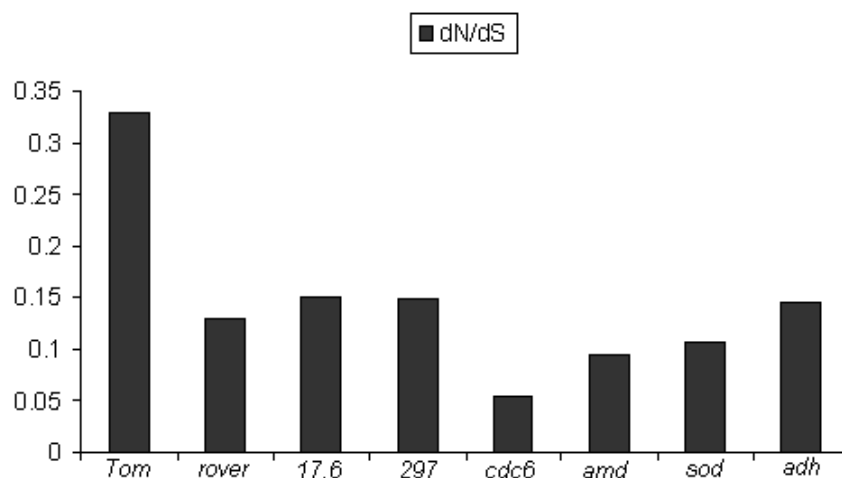


Figura 7: Valores de dN/dS médios calculados para os retroelementos e os genes nucleares.

Uso preferencial de códons dos retrotransposons e dos genes nucleares

Os valores de dN para *adh*, *sod* e *amd* (0.049, 0.048 e 0.050, respectivamente) foram similares e um pouco maiores que o valor de dN estimado para *cdc6* (0.029). Os valores de dS apresentaram-se mais variáveis entre os genes: 0.537 para *cdc6*, 0.524 para *amd*, 0.445 para *sod* e 0.340 para *adh*. Diferenças no valor de dS entre genes estão relacionadas a intensidade do uso preferencial de códons (Shields *et al.*, 1988). A figura 8 mostra a correlação negativa entre os valores médios de CBI e dS para os genes *adh*, *amd*, *sod* e *cdc6*, sugerindo que a seleção na composição dos códons é o principal fator na determinação da taxa de substituição nos sítios sinônimos nesses quatro genes.

Podemos observar que o gene *adh* possui o maior viés na utilização de códons enquanto os genes *amd* e *cdc6* possuem o menor viés.

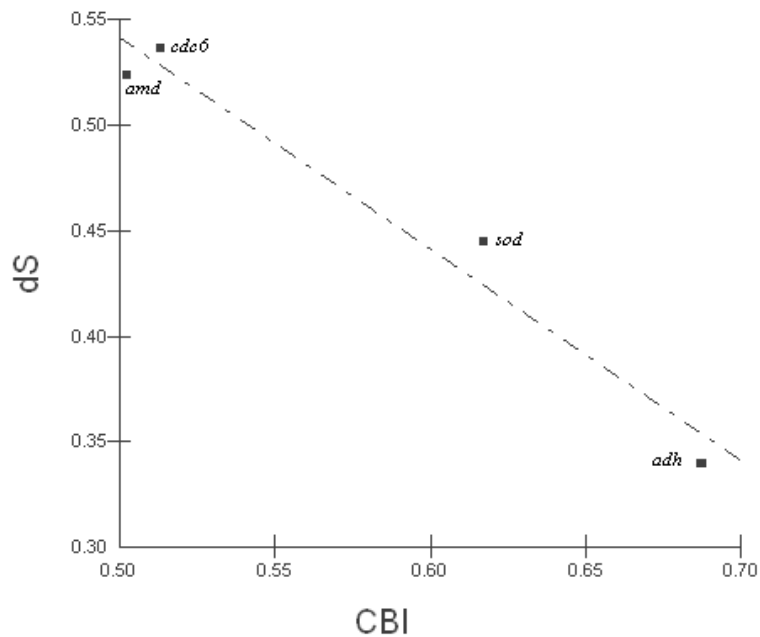


Figura 8: Correlação entre os valores médios de CBI e dS dos genes nucleares ($r^2=0.96$, $p<0.02$).

Para os quatro retroelementos também verificamos o viés na utilização de códons pela estimativa do CBI e N_c , que podem ser vistos na tabela 3. Podemos observar que os quatro elementos possuem um menor uso tendencioso de códons do que todos os genes. Desta forma, valores de dS dos TEs menores do que os valores de dS dos genes nucleares não podem ser explicados pelo uso tendencioso de códons. Comparando os valores de CBI dos retroelementos e dos genes, verificamos que os genes com valores de CBI mais próximos aos dos retroelementos foram os genes *cdc6* e *amd*. Por isso, esses genes foram escolhidos para as análises comparativas dos possíveis casos de TH.

Tabela 3: Valor médio do índice de uso preferencial de códons (CBI) e do número efetivo de códons (Nc) nas seqüências dos retroelementos *Tom*, *rover*, *297* e *17.6* e dos genes *cdc6*, *amd*, *sod* e *adh*.

Elemento, Gene, Grupo de espécies, (número de seqüências)	Nc médio (Variação)	CBI médio (Variação)
Tom		
Grupo <i>melanogaster</i> (5)	51.7 (50.6 – 52.2)	0.288 (0.276 – 0.301)
Média	51.7	0.288
rover		
Grupo <i>melanogaster</i> (26)	51.1 (39.3 – 55.0)	0.362 (0.310 – 0.505)
<i>Z. indianus</i> (1)	51.3	0.402
Média	51.1	0.363
297		
Grupo <i>melanogaster</i> (36)	43.2 (42.5 – 47.2)	0.477 (0.443 – 0.506)
Média	43.2	0.477
17.6		
Grupo <i>melanogaster</i> (23)	51.0 (45.9 – 55.4)	0.362 (0.303 – 0.457)
<i>D. virilis</i> (1)	42.7	0.514
<i>D. mojavensis</i> (6)	41.9 (41.6 – 42.2)	0.453 (0.452 – 0.456)
Média	48.8	0.386
cdc6		
Grupo <i>melanogaster</i> (10)	47.4 (36.4 – 58.1)	0.592 (0.331 – 0.700)
<i>D. virilis</i> (1)	51.0	0.501
<i>D. mojavensis</i> (1)	54.8	0.359
Média	48.3	0.513
amd		
Grupo <i>melanogaster</i> (7)	45.2 (43.4 – 48.2)	0.517 (0.466 – 0.547)
<i>D. virilis</i> (1)	40.2	0.479
<i>D. mojavensis</i> (1)	51.4	0.433
<i>Z. indianus</i> (1)	57.8	0.474
Média	46.0	0.502
sod		
Grupo <i>melanogaster</i> (7)	34.7 (31.1 – 37.4)	0.672 (0.635 – 0.691)
<i>D. virilis</i> (1)	35.1	0.563
<i>D. mojavensis</i> (1)	49.9	0.422
<i>Z. indianus</i> (1)	43.4	47.9
Média	37.1	0.617
adh		
Grupo <i>melanogaster</i> (10)	30.1 (28.5 – 31.6)	0.724 (0.666 – 0.770)
<i>D. virilis</i> (1)	44.3	0.483
<i>D. mojavensis</i> (1)	44.9	0.516
<i>Z. indianus</i> (1)	36.2	0.684
Média	32.8	0.687

Inferências de TH

Os dados apresentados neste trabalho mostram incongruências entre as relações filogenéticas dos TEs e as relações evolutivas das espécies hospedeiras, que são indicativos de possíveis eventos de TH. Entretanto, explicações alternativas, como polimorfismo ancestral com independente distribuição de cópias nas espécies descendentes, taxas diferentes de substituição em TEs nas diferentes espécies e perda estocástica de TEs em alguns táxons também podem explicar algumas dessas incongruências (Capy *et al.*, 1998). Deste modo, testamos a ocorrência de TH comparando a divergência nos sítios sinônimos entre os TEs e os genes nucleares *amd* e *cdc6*. Na figura 9 (A, B, C e D para os retroelementos *Tom*, *rover*, *297* e *17.6*, respectivamente) estão apresentados os valores de dN e dS para cada par de seqüências potencialmente envolvidas em TH e os valores de dN e dS dos genes *amd* e *cdc6* entre as respectivas espécies hospedeiras, onde podemos observar um dS menor para os TEs se comparado aos genes nucleares.

Um total de 18 possíveis eventos de TH foi analisado e a tabela 4 mostra os resultados do teste estatístico para cada caso. Eventos de TH foram inferidos somente quando o resultado do teste estatístico foi significativo para ambos os genes, exceto para as comparações em que apenas um gene foi utilizado. Em 14 comparações, a hipótese de TH é aceita com forte suporte estatístico ($p < 0.001$) para pelo menos um dos genes. Para *tei_clone2* x *rere9* e *1sim1* x *1sec1* a hipótese de TH é suportada com $p < 0.01$ para *cdc6* e $p < 0.05$ para *amd*. A direção da maioria desses 16 possíveis casos de TH não pôde ser definida. A estimativa de tempo da divergência dessas seqüências, mostrada na tabela 4, indica a quanto tempo as THs devem ter ocorrido, variando de 160 mil anos para o elemento *297* canônico entre as espécies *D. melanogaster* e *D. simulans* a 10.9 milhões de anos para o grupo *17.6B* entre *D. mojavensis* e *D. virilis*.

Tabela 4: Resumo dos possíveis eventos de transmissão horizontal identificados no nosso trabalho, resultado do teste estatístico, possível direção e estimativa de tempo de ocorrência dos mesmos.

	Espécie1	Direção	Espécie2	χ^2 <i>cdc6</i>	χ^2 <i>amd</i>	Estimativa de tempo ^a de divergência dos TEs
1	mal_clone1	↔	kik_clone4	***	na	1.31
2	kik_clone4	↔	tana1	***	na	1.59
3	rsec11	→	rere10	***	***	2.9
4	mau_clone4	→	kik_clone5	***	na	0.75
5	rsec11	↔	rsim1	*	NS	1.34
6	rere9	↔	rmel5	**	***	2.84
7	tei_clone5	→	zap_clone4	na	***	0.81
8	tei_clone2	↔	rere9	**	*	1.62
9	rere1	↔	ryak2	***	***	0.31
10	1moj1	↔	1vir5	***	***	10.9
11	1moj1	→	1mel2	**	NS	20.87
12	1sim1	↔	1sec1	**	*	0.44
13	1mel2	→	sec_clone2	***	***	0.25
14	1mel2	→	sim_clone1	***	***	0.19
15	2mel1	↔	2sim1	***	***	0.16
16	2mel1	↔	2sec1	***	***	0.31
17	2ana1	↔	kik_clone6	***	na	8.43
18	kik_clone6	↔	2yak1	***	na	9.19

na - não analisado, pelo menos uma das espécies não possui seqüência do gene disponível;

*** - valor $p < 0.001$;

** - valor $p < 0.01$;

* - valor $p < 0.5$;

NS - não significativo;

^a - estimativa de tempo em milhões de anos.

Os dois casos restantes não mostraram resultado significativo no teste estatístico comparados ao gene *amd*, sendo assim desconsiderados: (1) entre *rsec11* e *rsim1*, que mostraram significância ao nível de 5% em comparação com o gene *cdc6*, mas o caso foi desconsiderado levando em conta também o tempo estimado de divergência dessas seqüências (1.34 milhões de anos) que coincide com o tempo estimado de separação das espécies hospedeiras, *D. sechellia* e *D. simulans*; (2) entre *moj1* de *D. mojavensis* e *mel2* de *D. melanogaster* que mostraram valor significativo ($p < 0.01$) apenas para o gene *cdc6*. Porém, a estimativa de tempo de divergência dessas seqüências é de aproximadamente 20.9 milhões de anos, mais recente do que a estimativa de tempo de separação dos dois subgêneros ao qual pertencem essas espécies. Isso pode indicar que um evento de TH muito antigo possa ter ocorrido, tempo suficiente para impossibilitar a detecção com maior clareza.

Salientamos que quando mais de duas espécies estão envolvidas no mesmo caso de TH, consideramos aquelas espécies que apresentaram a menor divergência para o TE e possivelmente outras espécies não amostradas podem estar relacionadas a esses eventos de TH. Ainda, levando-se em conta o tempo estimado de TH e estimativa de divergência das espécies hospedeiras podemos concluir que alguns casos devem ter ocorrido entre linhagens ancestrais.

DISCUSSÃO

Embora importante, não existem critérios amplamente aceitos para classificação dos TEs, principalmente no que diz respeito ao estabelecimento de famílias e subfamílias. Vários autores têm utilizado diferentes critérios de classificação para elementos de grupos distintos (Lohe *et al.*, 1995; Robertson, 1995; Clark e Kidwell, 1997; Bowen e McDonald, 1999; Herédia *et al.*, 2004). Pinsker *et al.* (2001) sugere que se deve ter cautela ao estabelecer famílias e subfamílias de TEs quando somente uma região do elemento é analisada. Várias comparações entre elementos mostram variações na divergência entre ORFs, como é o próprio caso dos retroelementos estudados nesse trabalho. Os elementos canônicos 297 e 17.6 apresentam aproximadamente 63% de divergência na seqüência de aminoácidos do gene *gag* e 16% na seqüência de aminoácidos do gene *env*. Neste trabalho, para cada TE, as seqüências foram separadas em grupos diferentes de acordo com a divergência de nucleotídeos. Porém, a restrição da análise a apenas uma região dos retroelementos nos impossibilita de classificarmos esses grupos como diferentes famílias. Dessa forma, utilizamos um critério de divergência de nucleotídeos (divergência acima de 20%) estabelecido por Herédia *et al.* (2004) para considerar os grupos 17.6A e 17.6B, 297A e 297B, roverA e roverB como subfamílias dos elementos 17.6, 297 e rover, respectivamente.

A dinâmica evolutiva dos retroelementos *Tom*, 297, 17.6 e rover nos genomas de *Drosophilidae* são coincidentes em muitos aspectos. Em uma visão geral, as filogenias de cada família são discordantes da filogenia das espécies hospedeiras. Os quatro TEs estão envolvidos em eventos de TH. As relações entre as seqüências dos TEs das diferentes espécies são complexas, dificultando a compreensão do cenário evolutivo dos elementos. Os quatro retroelementos apresentam um menor viés na utilização de códons se comparados aos genes nucleares. Pelo menos os elementos 297, 17.6 e rover apresentam valores médios de dN/dS equivalentes ao apresentado pelos genes nucleares, o que indica que essas famílias estão evoluindo sob restrição seletiva.

Podemos imaginar diferentes cenários para as quatro famílias de retrotransposons:

Para o retroelemento *rover* a existência de duas subfamílias e a descontinuidade da presença desse elemento no subgrupo *melanogaster* pode ser explicada por polimorfismo no ancestral do subgrupo seguido de distribuição aleatória de cópias e/ou de perdas em algumas espécies, ou por independentes invasões dessas seqüências nos genomas. O único evento de TV bem claro acontece para a subfamília *roverA* entre as espécies *D. melanogaster*, *D. simulans*, *D. sechellia* e *D. mauritiana*, podendo concluir que essa subfamília estava presente no ancestral dessas espécies a pelo menos 5 milhões de anos. Ao contrário, seis casos de TH devem ter ocorrido para esta família entre 0.31 e 2.9 milhões de anos atrás. Dos seis casos de TH, dois envolvem espécies que não pertencem ao subgrupo *melanogaster*, *D. kikkawai* e *Z. indianus*. Dessa forma, esses dados suportam a idéia de que o surgimento desse elemento nas espécies de *Drosophila* é muito recente.

As seqüências homólogas ao retroelemento 17.6 também foram divididas em dois grupos. A subfamília 17.6A é exclusiva do subgrupo *melanogaster*. Apenas uma seqüência homóloga a esta subfamília foi encontrada nas buscas *in silico* nos genomas de *D. yakuba*, *D. simulans* e *D. sechellia*, que parecem estar perdendo esse elemento. Em contraste, várias seqüências foram encontradas em *D. melanogaster* e *D. erecta*. No entanto, nossas análises de clonagem e seqüenciamento detectaram clones de *D. simulans* e *D. sechellia* com uma pequena divergência em relação ao retroelemento canônico de *D. melanogaster*, sugerindo dois eventos de TH muito recentes bem suportados estatisticamente. Provavelmente, essa subfamília estava presente no ancestral do subgrupo *melanogaster* e foi transmitida por TV para as espécies descendentes, e recentemente casos de TH aconteceram dentro do próprio subgrupo. Porém, a maior incongruência apresentada pelo retroelemento 17.6 é a existência da subfamília 17.6B, com seqüências de espécies de outro subgênero. As análises comparativas e estatísticas do nosso trabalho sugerem TH dessa seqüência entre *D. mojavensis* e *D. virilis*, mas não entre as duas subfamílias. Esse cenário poderia ser explicado pela presença deste elemento no ancestral dos subgêneros

Sophophora e *Drosophila* a aproximadamente 60 milhões de anos atrás. No entanto, precisaríamos supor vários eventos de perda ao longo da evolução desses organismos para explicar sua ausência nas outras espécies de *Drosophila* analisadas neste estudo. Uma explicação mais plausível é que um evento de TH muito antigo aconteceu entre linhagens desses dois subgêneros e as diferenças acumuladas ao longo do tempo impossibilitaram a confirmação estatística em nosso estudo. Porém, o tempo em que ocorreu essa possível TH deveria ser muito antes do início da especiação do subgrupo *melanogaster* (estimado em 13 milhões de anos por Tamura *et al.*, 2004) e deveríamos também pressupor perdas em diferentes genomas. Talvez a análise do elemento inteiro de *D. mojavensis* e *D. virilis* possa ajudar a elucidar esse caso.

O elemento 297 também apresentou uma divisão em duas subfamílias. Uma delas, 297B, é encontrada somente em *D. simulans* e *D. sechellia* com distâncias condizentes com TV. Essas seqüências podem representar vestígios de polimorfismo ancestral ou ainda TH de origem desconhecida, dessa seqüência para o ancestral dessas duas espécies, após a separação de *D. melanogaster*. A subfamília 297A parece estar presente em todas as espécies do subgrupo *melanogaster* por TV e, assim como o elemento 17.6, o retroelemento 297 parece estar envolvido em eventos de TH recentes entre *D. melanogaster*, *D. sechellia* e *D. simulans*. Além disso, dois casos de TH mais antigos devem ter acontecido para essa família envolvendo as espécies *D. ananassae* e *D. kikkawai*.

O retroelemento *Tom* mostrou a distribuição mais restrita, sendo encontrado em apenas três espécies, *D. ananassae*, *D. kikkawai* e *D. malerkotliana*, mas em duas delas por TH. A estreita relação desse elemento com o elemento 297 mostrada pelas diferentes análises filogenéticas, pela pequena divergência apresentada nas seqüências de aminoácidos da transcriptase reversa (18.8%) mostrada na tabela 2 e também pela divergência apresentada pelas outras duas ORFs (aproximadamente 37%) sugerem que esses elementos podem representar a mesma família que estava presente em uma espécie ancestral.

Bowen e McDonald (2001), analisando diversas famílias de LTR-retrotransposons no genoma de *D. melanogaster*, sugerem que muitos destes elementos foram mobilizados muito recentemente, após a separação desta

espécie de *D. simulans*, devido a grande similaridade encontrada comparando as duas LTRs de um mesmo elemento. Esses autores sugerem a possibilidade de que esses elementos foram recentemente introduzidos no genoma de *D. melanogaster* a partir de outras espécies. O trabalho de Sánchez-Gracia *et al.* (2005) reforça a idéia de que grande parte dos TEs em *D. melanogaster* possuem uma origem recente como resultado de movimentos entre espécies. Um alta taxa de TH foi encontrada comparando TEs de *D. melanogaster*, *D. simulans* e *D. yakuba* (Sánchez-Gracia *et al.*, 2005). Em nosso trabalho cinco casos de TH envolveram *D. melanogaster* e outra espécie do subgrupo *melanogaster*, sendo que para os elementos 297 e 17.6 eventos de TH muito recentes aconteceram entre as três espécies relacionadas *D. melanogaster*, *D. simulans* e *D. sechellia*.

O mecanismo pelo qual a TH ocorre ainda é obscuro. Alguns vetores potenciais dessa transmissão são: vírus, vespas parasitas, ácaros e bactérias endossimbiontes (Silva *et al.*, 2004). No caso dos retrotransposons, eles poderiam ser co-empacotados em capsídios de vírus de RNA e então serem carregados para outro genoma. No entanto, com a aquisição de um gene *env* funcional os retrotransposons poderiam passar de uma espécie hospedeira a outra sem a necessidade de um vetor (Malik *et al.*, 2000), já que este gene é responsável pela infecciosidade. Malik *et al.* (2000) mostraram que o gene *env* tem sido adquirido, independentemente, várias vezes na história evolutiva dos retrotransposons e sugerem que essa aquisição poderia aumentar a probabilidade de eventos de TH. Assim, da mesma forma como sugerido para outros IERVs envolvidos em TH (Herédia *et al.*, 2004; Ludwig e Loreto, 2007), é plausível que o grande número de eventos de TH encontrado para os elementos 297, 17.6, *Tom* e *rover* ocorreram em decorrência de sua possível infecciosidade.

No entanto, o aumento no número de casos de TH de TEs entre espécies relacionadas do subgrupo *melanogaster* (Sánchez-Gracia *et al.*, 2005; Ludwig, 2006; Ludwig e Loreto, 2007) levanta a questão sobre o possível efeito da introgressão na aquisição de novas seqüências. Este fenômeno pode ter acontecido no passado principalmente entre *D. melanogaster* e *D. simulans*, as quais podem, ainda hoje, produzir híbridos que dependendo da combinação das linhagens parentais podem ser férteis (Davis *et al.*, 1996; Sawamura *et al.*, 2000).

Dessa forma, embora eventos de TH sejam freqüentemente sugeridos, até o momento é impossível fornecer uma resposta completa sobre os mecanismos envolvidos nesses fenômenos.

Agradecimentos

Os autores agradecem ao Dr. Henrique Bunselmeyer Ferreira, ao Msc. Paulo Marcos Pinto, Msc. Lizandra Jaqueline Robe e a Paloma Rubin pela ajuda durante a realização deste trabalho; a Msc. Adriana Ludwig pelas discussões e comentários neste manuscrito; ao CNPq e a FAPERGS pelos recursos de financiamento à pesquisa; e ao CNPq pela concessão de bolsa.

REFERÊNCIAS BIBLIOGRÁFICAS

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 487:2185-2195.

Bartolome C, Maside X, Charlesworth B. (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol* 19:926-37.

Bowen NJ e McDonald JF (1999) Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. *Genome Res* 9:924-935.

Bowen NJ e McDonald JF (2001) *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* 11:1527-1540.

Capy P, Bazin C, Higuët D e Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197 pp.

- Clark JB e Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc Natl Acad Sci USA* 94:11428-11433.
- Coffin JM, Hughes SH e Varmus HE (1997) *Retroviruses*. Cold Spring Harbor Laboratory Press, Nova York, 843 pp.
- Davis AW, Roote J, Morley T, Sawamura K, Herrmann S, Ashburner M (1996) Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* 380:157-159.
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-185.
- Graur D e Li W (2000) *Fundamentals of molecular evolution*. 2ed. Sinauer Associates, Massachusetts.
- Grumblin G, Strelets V (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Res* 34:D484-488.
- Hall BG (2001) *Phylogenetic Trees Made Easy-a How to Manual for Molecular Biologists*. 2ed. Sinauer Associates, Massachusetts.
- Havecker ER, Gao X e Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biol* 5:225.1-225.6.
- Herédia F, Loreto EL e Valente VL (2004) Complex Evolution of *gypsy* in *Drosophilid* Species. *Mol Biol Evol* 21:1831-1842.
- Hull R (2001) Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. *International Committee on Taxonomy of Viruses. Arch Virol* 146:2255-2261.
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403-405.
- Juni N, Awasaki T, Yoshida K, Hori SH (1996) The Om (1E) mutation in *Drosophila ananassae* causes compound eye overgrowth due to tom retrotransposon-driven overexpression of a novel gene. *Genetics* 143:1257-70.
- Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet* 16:418-420.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462-467.
- Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M e Celniker SE (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* 3:0084.1-0084.20.

Kapitonov VV e Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci U S A 100:6569-6574.

Kent WJ (2002) BLAT-the BLAST-like alignment tool Genome Res 12:656-664.

Kim A, Terzian C, Santamaria P, Pelisson A, Prud'homme N e Bucheton A (1994) Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc Natl Acad Sci USA 91:1285-1289.

Klein CC, Essi L, Golombieski RM e Loreto ELS (1999) Disgenesia do híbrido em populações naturais de *Drosophila melanogaster*. Ciencia e Natura 21: 7-20.

Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 25;7:474.

Kumar S, Tamura K e Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150-163.

Kumar S, Tamura K, Jakobsen IB e Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244-1245.

Lerat E, Rizzon C, Biemont C. (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Res 13:1889-96.

Lewis RL, Beckenbach AT e Mooers AØ (2005) The phylogeny of the subgroups within the *melanogaster* species group: Likelihood tests on *COI* and *COII* sequences and a Bayesian estimate of phylogeny. Mol Phylogenet Evol 37:15-24.

Lohe AR, Moriyama EM, Lidholm DA e Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of mariner like transposable elements. Mol Biol Evol 12:62-72.

Ludwig A (2006) Evolução dos *Errantivirus gtwin* e *gypsy* no gênero *Drosophila*. Dissertação de Mestrado. PPG Genética e Biologia Molecular UFRGS, Porto Alegre, RS.

Ludwig A e Loreto ELS (2007) Evolutionary pattern of the *gtwin* retrotransposon in the *Drosophila melanogaster* subgroup. Genetica (*in press*).

Malik HS, Henikoff S e Eickbush TH (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res 10:1307-1318.

Malik HS, Henikoff S e Eickbush TH (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* 10:1307-1318.

Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability. *J Mol Evol* 37:273-80.

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

Nicholas, KB e Nicholas HB Jr (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. <http://www.psc.edu/biomed/genedoc>

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

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

Rizzon C, Marais G, Gouy M, Biemont C. (2002) Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res* 12:400-7.

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

Robertson HM (1995) The *Tc1-mariner* superfamily of transposons in animals. *J Insect Physiol* 41:99-105.

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

Ross CL, Dyer KA, Erez T, Miller SJ, Jaenike J, Markow TA (2003) Rapid divergence of microsatellite abundance among species of *Drosophila*. *Mol Biol Evol* 20:1143-1157.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.

Sanchez-Gracia A, Maside X, Charlesworth B (2005) High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet* 21:200-203.

- Sassi AK, Herédia F, Loreto ELS, Valente VLS, Rohde C (2005) Transposable elements *P* and *gypsy* in natural populations of *Drosophila willistoni*. *Genet. Mol. Biol* 28: 734-739.
- Sawamura K, Davis AW, Wu CI (2000) Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 97:2652-2655.
- Sharp PM, Li WH (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol* 28:398-402.
- Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol* 5:704-16.
- Silva JC, Loreto EL e Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-71.
- Song SU, Gerasimova T, Kurkulos M, Boeke JD e Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. *Genes Dev* 8:2046-2057.
- Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* 5:233-241.
- Tamura K, Subramanian S e Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21:36-44.
- Tanda S, Corces VG (1991) Retrotransposon-induced overexpression of a homeobox gene causes defects in eye morphogenesis in *Drosophila*. *EMBO J* 10:407-17.
- Tanda S, Mullor JL e Corces VG (1994) The *Drosophila tom* retrotransposon encodes an envelope protein. *Mol Cell Biol* 14:5392-5401.
- Tanda S, Shrimpton AE, Hinton CW, Langley CH. (1989) Analysis of the Om(1D) locus in *Drosophila ananassae*. *Genetics* 123:495-502.
- Tanda S, Shrimpton AE, Ling-Ling C, Itayama H, Matsubayashi H, Saigo K, Tobarí YN e Langley CH (1988) Retrovirus-like features and site specific insertions of a transposable element, *tom*, in *Drosophila ananassae*. *Gen Genet* 214:405-411.
- Terzian C, Pelisson A e Bucheton A (2001) Evolution and phylogeny of insect endogenous retroviruses. *BMC Evol Biol* 1:3.
- Vieira C, Nardon C, Arpin C, Lepetit D, Biemont C (2002) Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? *Mol Biol Evol* 19:1154-61.

Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87:23-29.

Xiong Y e Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353-3362.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning - A Laboratory Manual*. 2ed. Cold Spring Harbour Laboratory Press, New York.

CAPÍTULO III

DISCUSSÃO GERAL

Nesse trabalho realizamos uma análise evolutiva dos retroelementos *Tom*, *297*, *17.6* e *rover* em espécies de *Drosophilidae*.

Anteriormente à realização deste trabalho, realizamos uma investigação da distribuição de seqüências homólogas aos retroelementos pertencentes as famílias *Tom*, *297* e *17.6* em 33 espécies Neotropicais de *Drosophila*, *Zaprionus indianus* e *Scaptodrosophila latifasciaeformis*. A metodologia utilizada foi Dot Blot e PCR, e indicou uma distribuição restrita desses elementos ao grupo *melanogaster* e em *Z. indianus*.

Nossa proposta de trabalho consistiu na análise evolutiva dos retrotransposons *297*, *17.6*, *Tom* e *rover* através de buscas *in silico* e seqüenciamento desses retroelementos nas espécies onde eles foram encontrados previamente.

As análises evolutivas que apresentamos com base nas seqüências da transcriptase reversa mostram que a dinâmica dos retroelementos *Tom*, *297*, *17.6* e *rover* nos genomas de *Drosophilidae* são coincidentes em muitos aspectos: (1) Em uma visão geral, as filogenias de cada família são discordantes da filogenia das espécies hospedeiras. (2) Os quatro TEs estão envolvidos em eventos de TH. (3) As relações entre as seqüências dos TEs das diferentes espécies são complexas, dificultando a compreensão do cenário evolutivo dos elementos. (4) Os quatro retroelementos apresentam um menor viés na utilização de códons se comparados aos genes nucleares. (5) Pelo menos os elementos *297*, *17.6* e *rover* apresentam valores médios de dN/dS equivalentes ao apresentado pelos genes nucleares, o que indica que essas famílias estão evoluindo sob restrição seletiva.

Ao todo, 16 possíveis casos de transferência horizontal (TH) foram necessários para explicar a distribuição desses elementos e as relações das seqüências entre espécies.

Eventos de TH têm sido tradicionalmente inferidos quando o grau de similaridade entre seqüências de TEs é impossível de reconciliar com o longo tempo de divergência de suas respectivas espécies hospedeiras. Segundo Silva e Kidwell (2000), comparando a divergências das seqüências de nucleotídeo do TE com a divergência observada para genes do hospedeiro que estão evoluindo sob similar ou mais forte restrição seletiva, um evento de TH pode ser inferido se a divergência entre as seqüências do TE é significativamente menor do que a observada para os genes nucleares. No entanto, em diversos trabalhos que descrevem eventos de TH em *Drosophilidae*, a grande maioria não apresenta um teste estatístico (Silva e Kidwell, 2000; Loreto *et al.*, 2001; Heredia *et al.*, 2004 de Almeida e Carareto, 2005). Isto se deve principalmente, porque as diferenças são tão óbvias que dispensam qualquer teste. Na maioria das vezes as espécies envolvidas são muito distantes, apresentando uma divergência em genes nucleares bastante grande comparada a pequena divergência dos TEs envolvidos em TH recente.

No entanto, com a possibilidade de realizar estudos de evolução dos TEs nos genomas de espécies de *Drosophila*, nosso grupo têm encontrado um grande número de eventos de TH entre as espécies do subgrupo *melanogaster*. Para o retroelemento *gtwin* três possíveis casos de TH foram propostos (Ludwig *et al.*, 2007) e para os retroelementos *gypsy*, *gypsy2*, *gypsy4*, *gypsy5* e *gypsy6* foram propostos 15 casos no total (Ludwig, 2006), todos estes envolvendo as espécies do subgrupo *melanogaster*. No presente estudo identificamos nove possíveis casos de TH envolvendo essas espécies. Com esse grande número de possíveis eventos de TH entre espécies tão próximas cresceu a preocupação do nosso grupo em utilizar testes estatísticos para estabelecer casos de TH.

Nossa proposta para este trabalho foi a comparação da divergência nos sítios sinônimos entre os TEs e a encontrada para genes nucleares com similares valores de CBI, já que é encontrada uma correlação negativa entre os valores de CBI e os valores de dS para os genes nucleares (Sharp e Li, 1989). Partindo deste princípio é esperada aproximadamente a mesma proporção de dS para genes da mesma espécie, incluindo os TEs, desde que os mesmos estejam evoluindo por transmissão vertical e possuam o mesmo nível de uso preferencial

de códons. Ao contrário, se um TE foi transmitido de uma espécie para outra por TH, é esperado que o dS do TE entre as espécies seja menor que os dS encontrado para genes nucleares, refletindo o menor tempo de divergência das seqüências em relação a seqüência ancestral. Empregamos um teste qui-quadrado, com correção de Yates, para verificar se o número observado de similaridades e diferenças para o TE nos sítios sinônimos (calculadas com base no valor de dS do TE e S do TE) é significativamente menor que o número esperado de similaridades e diferenças para o TE nos sítios sinônimos (calculadas com base no valor de dS do gene nuclear e S do TE).

Ao mesmo tempo em que um maior número de casos de TH estão sendo detectados, surge a questão de como estes eventos, que esperava-se que fossem fenômenos raros, acontecem com tanta freqüência. Ainda não está claro o mecanismo pelo qual a TH acontece. Além dos potenciais vetores para TEs em geral, no caso dos retrotransposons, existe ainda a possibilidade de eles serem empacotados junto com vírus de RNA ou, no caso específico dos IERVS, a infecciosidade do próprio retrotransposon tornaria a TH possível. No entanto, casos de TH entre espécies muito próximas filogeneticamente levanta a questão sobre o possível efeito da introgressão na aquisição de novas seqüências.

Elementos de transposição estão intimamente associados com o genoma hospedeiro, tendo implicações diretas na evolução dos genomas. Deste modo, para melhor compreender o processo de co-evolução hospedeiro-elemento é muito importante conhecer a extensão com que as THs dos retrotransposons têm ocorrido na história evolutiva dos eucariotos, o efeito que essas transferências causam no genoma dos organismos e os mecanismos dessas trocas de seqüências entre espécies.

REFERÊNCIAS BIBLIOGRÁFICAS

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW e Rubin GM, Venter JC. (2000) The genome sequence of *Drosophila melanogaster*. Science 487:2185-2195.

Almeida R e Allshire RC (2005) RNA silencing and genome regulation. Trends Cell Biol 15:251-258.

Aravin A e Tuschl T (2005) Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 579:5830-5840.

Aravin A, Lagos-Quintana M, Yalcin A, Zavalon M, Marks D, Snyder B, Gaasterland T, Meyer J e Tuschl T (2003). The small RNA profile during *Drosophila melanogaster* development. Dev Cell 5:337-350.

Ashburner M, Bodmer M e Lameunier F (1984) On evolutionary relationships of *Drosophila melanogaster*. Dev Genet 4:295-312.

Bartolome C, Maside X e Charlesworth B (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol.* 19:926-937.

Becker H e Lönnig W (2001) Transposons: eukaryotic. In: *Nature encyclopedia of life sciences*, vol. 18. Nature Publishing Group, Londres, Reino Unido, pp 529-539.

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* 42:251-269.

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

Biemont C e Vieira C (2006) Genetics: junk DNA as an evolutionary force. *Nature* 443:521-524.

Blauth ML (2005) Expressão de elementos transponíveis em *Drosophila willistoni*. Tese de Doutorado. PPG Genética e Biologia Molecular UFRGS, Porto Alegre, RS.

Bock IR e Wheeler MR (1972) The *Drosophila melanogaster* species group studies in genetics VII. *Univ Texas Publ* 7213.

Boeke JD, Eickbush TH, Sandmeyer SB e Voytas DF (1999) Metaviridae. In: Murphy FA (ed) *Virus Taxonomy: ICTV VII Report*. Springer-Verlag, Nova York, pp 123-135.

Bowen NJ e Jordan IK (2002) Transposable elements and the evolution of eukaryotic complexity. *Curr Issues Mol Biol* 4:65-76.

Bowen NJ e McDonald JF (1999) Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. *Genome Res* 9:924-935.

Bowen NJ e McDonald JF (2001) *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* 11:1527-1540.

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

Büchen-Osmond C (2003) The Universal Virus Database ICTVdB. *Computing in Science and Engineering* 5:16-25.

Capy P, Anxolabéhère D e Langin T (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? *Trends Genet* 10:7-12.

Capy P, Bazin C, Higuete D e Langin T (1998) Dynamics and evolution of transposable elements. Landes Bioscience, Austin, Texas, 197 pp.

- Clark JB e Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc Natl Acad Sci USA* 94:11428-11433.
- Coffin JM, Hughes SH e Varmus HE (1997) *Retroviruses*. Cold Spring Harbor Laboratory Press, Nova York, 843 pp.
- Cordaux R, Udit S, Batzer MA e Feschotte C (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proc Natl Acad Sci USA* 103:8101-8106.
- David JR e Tsacas L (1980) Cosmopolitan, subcosmopolitan and widespread species: different strategies within the Drosophilid family (Diptera). *C R Soc Biogéogr* 57:11-26.
- Davis AW, Roote J, Morley T, Sawamura K, Herrmann S e Ashburner M (1996) Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* 380:157-159.
- de Almeida LM e Carareto CM (2005) Multiple events of horizontal transfer of the *Minos* transposable element between *Drosophila* species. *Mol Phylogenet Evol* 35:583-594.
- Desset S, Meignin C, Dastugue B e Vaury C (2003) COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164:501-509.
- Djikeng A, Shi H, Tschudi C e Ullu E (2001) RNA interference in *Trypanosoma brucei*: Cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. *RNA* 7:1522-1530.
- Doolittle WF e Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284:601-603.
- Esnault C, Maestre J e Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 24:363-367.
- Ewing B, Hillier L, Wendl MC e Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175-185.
- Feschotte C, Jiang N e Wessler SR (2002) Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3:329-341.
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet* 5:103-107.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE e Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-811.

- Garcia RN, D'Avila MF, Robe LJ, Loreto EL, Panzera Y, de Heredia FO e Valente VL (2007) First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica (in press)*.
- Gotea V e Makalowski W (2006) Do transposable elements really contribute to proteomes? *Trends Genet.* 22:260-267.
- Graur D e Li W (2000) *Fundamentals of molecular evolution*. 2ed. Sinauer Associates, Massachusetts.
- Grumblin G e Strelets V (2006) FlyBase: anatomical data, images and queries. *Nucleic Acids Res* 34:D484-488.
- Hall BG (2001) *Phylogenetic Trees Made Easy-a How to Manual for Molecular Biologists*. 2ed. Sinauer Associates, Massachusetts.
- Hartl DL, Lohe AR e Lozovskaya ER (1997) Regulation of the transposable element *mariner*. *Genetica* 100:177-184.
- Havecker ER, Gao X e Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biol* 5:225.1-225.6.
- Herédia F, Loreto EL e Valente VL (2004) Complex Evolution of *gypsy* in Drosophilid Species. *Mol Biol Evol* 21:1831-1842.
- Hull R (2001) Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. *International Committee on Taxonomy of Viruses. Arch Virol* 146:2255-2261.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- Jacob F (1977) Evolution and tinkering. *Science* 196:1161-1166.
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG e Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23:403-405.
- Jiang N, Bao Z, Zhang X, Eddy SR e Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431:569-573.
- Jordan IK (2006) Evolutionary tinkering with transposable elements. *Proc Natl Acad Sci U S A* 103:7941-7942.
- Juni N, Awasaki T, Yoshida K e Hori SH (1996) The Om (1E) mutation in *Drosophila ananassae* causes compound eye overgrowth due to tom retrotransposon-driven overexpression of a novel gene. *Genetics* 143:1257-70.
- Jurka J (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet* 16:418-420.

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.

Kalmykova AI, Klenov MS e Gvozdev VA (2005) Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res* 33:2052-2059.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M e Celniker SE (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* 3:0084.1-0084.20.

Kapitonov VV e Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proc Natl Acad Sci U S A* 100:6569-6574.

Kastanis P, Eliopoulos E, Goulielmos GN, Tsakas S e Loukas M (2003) Macroevolutionary relationships of species of *Drosophila melanogaster* group based on mtDNA sequences. *Mol Phylogenet Evol* 28:518-528.

Kent WJ (2002) BLAT-the BLAST-like alignment tool *Genome Res* 12:656-664.

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

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

Kim A, Terzian C, Santamaria P, Pelisson A, Prud'homme N e Bucheton A (1994) Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 91:1285-1289.

Klein CC, Essi L, Golombieski RM e Loreto ELS (1999) Disgenesia do híbrido em populações naturais de *Drosophila melanogaster*. *Ciencia e Natura* 21: 7-20.

Kohany O, Gentles AJ, Hankus L e Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 25;7:474.

Kumar S, Tamura K e Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5:150-163.

Kumar S, Tamura K, Jakobsen IB e Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L e Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol Biol* 22:159-226.

Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B e Vaury C (2000) Life cycle of an endogenous retrovirus, *ZAM*, in *Drosophila melanogaster*. *J Virol* 74:10658-10669.

Lerat E, Rizzon C e Biemont C (2003) Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. *Genome Res* 13:1889-1896.

Levis RW, Ganesan R, Houtchens K, Tolar LA e Sheen FM (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell*. Dec 75:1083-1093.

Lewis RL, Beckenbach AT e Mooers AØ (2005) The phylogeny of the subgroups within the *melanogaster* species group: Likelihood tests on *COI* and *COII* sequences and a Bayesian estimate of phylogeny. *Mol Phylogenet Evol* 37:15-24.

Li WH, Gu Z, Wang H e Nekrutenko A (2001) Evolutionary analyses of the human genome. *Nature* 409:847-849.

Lipatov M, Lenkov K, Petrov DA e Bergman CM (2005) Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. *BMC Biol* 12;3:24.

Lohe AR, Moriyama EM, Lidholm DA e Hartl DL (1995) Horizontal transmission, vertical inactivation, and stochastic loss of mariner like transposable elements. *Mol Biol Evol* 12:62-72.

Lorenc A e Makalowski W (2003) Transposable elements and vertebrate protein diversity. *Genetica* 118:183-191.

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

Ludwig A (2006) Evolução dos *Errantivirus gtwin* e *gypsy* no gênero *Drosophila*. Dissertação de Mestrado. PPG Genética e Biologia Molecular UFRGS, Porto Alegre, RS.

Ludwig A e Loreto ELS (2007) Evolutionary pattern of the *gtwin* retrotransposon in the *Drosophila melanogaster* subgroup. *Genetica* (*in press*).

Lyko F, Ramsahoye BH e Jaenisch R (2000) DNA methylation in *Drosophila melanogaster*. *Nature* 408: 538–540.

Malik HS, Henikoff S e Eickbush TH (2000) Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* 10:1307-1318.

- Mandrioli M e Borsatti F (2006) DNA methylation of fly genes and transposons. *Cell Mol Life Sci* 63:1933-1936.
- Marhold J, Rothe N, Pauli A, Mund C, Kuehle K, Brueckner B e Lyko F (2004) Conservation of DNA methylation in dipteran insects. *Insect Mol Biol* 13(2):117-123.
- Matsubayashi H, Matsuda M, Tomimura Y, Shibata M e Tobar YN (1992) Cytological mapping of Om mutants of *Drosophila ananassae*. *Jpn J Genet* 67:259-264.
- Mayo MA (2005) Changes to virus taxonomy 2004. *Arch Virol* 150:189-198.
- McDonald JF (1993) Evolution and consequences of transposable elements. *Curr Opin Genet Dev.* 3:855-864.
- Mejlumian L, Pelisson A, Bucheton A e Terzian C (2002) Comparative and functional studies of *Drosophila* species invasion by the gypsy endogenous retrovirus. *Genetics* 160:201-209.
- Mevel-Ninio MT, Pelisson A, Kinder J, Campos AR e Bucheton A (2007) The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* (*in press*)
- Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A e Rafalski A (2005) Gene duplication and exon shuffling by *helitron-like* transposons generate intraspecies diversity in maize. *Nat Genet* 37:997-1002.
- Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability. *J Mol Evol* 37:273-80.
- Nei M e Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418-426.
- Nekrutenko A e Li WH (2001) Transposable elements are found in a large number of human protein-coding genes. *Trends Genet* 17:619-621.
- Nicholas, KB e Nicholas HB Jr (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. <http://www.psc.edu/biomed/genedoc>
- Nisole S, Lynch C, Stoye JP e Yap MW (2004) A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *Proc Natl Acad Sci U S A.* 101:13324-13328.
- Nuzhdin SV (1995) The distribution of transposable elements on X chromosomes from a natural population of *Drosophila simulans*. *Genet Res* 66:159-166.

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

Orgel LE e Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604-607.

Ozers MS e Friesen PD (1996) The Env-like open reading frame of the baculovirus-integrated retrotransposon *TED* encodes a retrovirus-like envelope protein. *Virology* 226:252-259.

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

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

Potter SS, Brorein WJ, Dunsmuir P e Rubin GM (1979) Transposition of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. *Cell* 17:415-427.

Prud'homme N, Gans M, Masson M, Terzian C e Bucheton A (1995) *Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*. *Genetics* 139:697-711.

Rizzon C, Marais G, Gouy M e Biemont C (2002) Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res* 12:400-407.

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

Robertson HM (1995) The *Tc1-mariner* superfamily of transposons in animals. *J Insect Physiol* 41:99-105.

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

Ross CL, Dyer KA, Erez T, Miller SJ, Jaenike J e Markow TA (2003) Rapid divergence of microsatellite abundance among species of *Drosophila*. *Mol Biol Evol* 20:1143-1157.

Rozas J, Sanchez-DelBarrio JC, Messeguer X e Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.

- Saigo K, Kugimiya W, Matsuo Y, Inouye S, Yoshioka K e Yuki S (1984) Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature* 312:659-661.
- Saigo K, Millstein L e Thomas CA (1981) The organization of *Drosophila melanogaster* histone genes. *Cold Spring Harbor Symp Quant Biol* 45:815-827.
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H e Siomi MC (2006) Specific association of Piwi with *ras*iRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20:2214-2222.
- Saitou N e Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Sambrook J, Fritsch EF e Maniatis T (1989) *Molecular Cloning - A Laboratory Manual*. 2ed. Cold Spring Harbour Laboratory Press, New York.
- Sanchez-Gracia A, Maside X e Charlesworth B (2005) High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends Genet* 21:200-203.
- SanMiguel P, Tikhonov A, Jin Y, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z e Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274:765-768.
- Sarot E, Payen-Groschene G, Bucheton A e Pelisson A (2004) Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* 166:1313-1321.
- Sassi AK, Herédia F, 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.
- Sawamura K, Davis AW e Wu CI (2000) Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 97:2652-2655.
- Sayah DM, Sokolskaja E, Berthoux L e Luban J (2004) Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 430:569-573.
- Schwaroch V (2002) Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera Drosophilidae). *Biol J Linn Soc* 76:21-37.
- Selker EU (1999) Gene silencing: repeats that count. *Cell* 97:157-160.
- Sharp PM e Li WH (1989) On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol* 28:398-402.

- Shields DC, Sharp PM, Higgins DG e Wright F (1988) "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol* 5:704-16.
- Silva JC e Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* elements. *Mol Biol Evol* 17:1542-1557.
- Silva JC, Loreto EL e Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* 6:57-71.
- Song SU, Gerasimova T, Kurkulos M, Boeke JD e Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. *Genes Dev* 8:2046-2057.
- Song SU, Kurkulos M, Boeke JD e Corces VG (1997) Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the *gypsy* retroelement of *Drosophila*. *Development* 124:2789-2798.
- Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* 5:233-241.
- Sturtevant AH (1939) On the subdivision of the genus *Drosophila*. *Proc Natl Acad Sci USA* 25:137-141.
- Tamura K, Subramanian S e Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21:36-44.
- Tanda S, Mullor JL e Corces VG (1994) The *Drosophila tom* retrotransposon encodes an envelope protein. *Mol Cell Biol* 14:5392-5401.
- Tanda S, Shrimpton AE, Hinton CW e Langley CH (1989) Analysis of the Om(1D) locus in *Drosophila ananassae*. *Genetics* 123:495-502.
- Tanda S, Shrimpton AE, Ling-Ling C, Itayama H, Matsubayashi H, Saigo K, Tobar YN e Langley CH (1988) Retrovirus-like features and site specific insertions of a transposable element, *tom*, in *Drosophila ananassae*. *Gen Genet* 214:405-411.
- Terzian C, Pelisson A e Bucheton A (2001) Evolution and phylogeny of insect endogenous retroviruses. *BMC Evol Biol* 1:3.
- Throckmorton LH (1975) The phylogeny, ecology and geography of *Drosophila*. In: King RC (ed) *Handbook of genetics*. V 3: Invertebrates of genetic interest. Plenum, New York, pp 421-469.
- Vagin VV, Klenov MS, Kalmykova AI, Stolyarenko AD, Kotelnikov RN e Gvozdev VA (2004) The RNA interference proteins and *vasa* locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol* 1:51-55.

- Val FC, Marques MD e Vilela CR (1981) *Drosophilidae* of Neotropical region. In: Ashburner M, Carson HL e Thompson JN (eds) *The Genetics and Biology of Drosophila*. V 3. Academic Press, Orlando, pp123-168.
- Vieira C, Nardon C, Arpin C, Lepetit D e Biemont C (2002) Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements?
Mol Biol Evol 19:1154-61.
- Volff JN (2006) Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* 28:913-922.
- Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, Boeke JD e Moran JV (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* 21:1429-1439.
- Wessler SR (2006) Eukaryotic Transposable Elements: Teaching Old Genomes New Tricks. In: Caporale L (ed) *The Implicit Genome* Oxford University Press. USA, pp138-165.
- Wheeler MR (1982) The *Drosophilidae*: a taxonomic overview. In: Ashburner M, Carson HL e Thompson JN (eds) *The Genetics and Biology of Drosophila*. V 3. Academic Press, Orlando, pp 1-97.
- Wright F (1990) The 'effective number of codons' used in a gene. *Gene* 87:23-29.
- Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL e Batzer MA (2006) Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci U S A* 103:17608-17613.
- Xiong Y e Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353-3362.
- Yang Y, Zhang Y, Qian Y e Zeng Q (2004) Phylogenetic relationships of *Drosophila melanogaster* species group deduced from spacer regions of histone gene H2A-H2B. *Mol Phylogenet Evol* 30:336-343.
- Yoder JA, Walsh CP e Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335-340.

ANEXOS

Tabela A1: espécies utilizadas para rastreamento por Dot Blot e PCR com os respectivos resultados.

Gênero	Subgênero	Grupo	Espécie	Linhagem/ Local de coleta	Dot Blot Tom ^a	Dot Blot 297/17.6 ^b	PCR ^c
<i>Drosophila</i>	<i>Sophophora</i>	<i>melanogaster</i>	<i>D. kikkawai</i>	Porto Alegre, Brasil	+	+	+
		<i>melanogaster</i>	<i>D. melanogaster</i>	Harwich, Bowling Green, EUA	-	+	+
		<i>melanogaster</i>	<i>D. simulans</i>	Uruguai	-	+	+
		<i>melanogaster</i>	<i>D. simulans</i>	Florianópolis, Brasil	-	+	+
		<i>melanogaster</i>	<i>D. simulans</i>	Porto Alegre, Brasil	-	+	+
		<i>melanogaster</i>	<i>D. ananassae</i>	Florianópolis, Brasil	+	-	+
		<i>melanogaster</i>	<i>D. malerkotliana</i>	Florianópolis, Brasil	+	-	+
		<i>saltans</i>	<i>D. sturtevanti</i>	Florianópolis, Brasil	-	-	-
		<i>saltans</i>	<i>D. onca</i>	Florianópolis, Brasil	-	-	-
		<i>saltans</i>	<i>D. dimorpha</i>	São José do Rio Preto, Brasil	-	-	-
		<i>willistoni</i>	<i>D. willistoni</i>	Porto Alegre, Brasil	-	-	-
		<i>willistoni</i>	<i>D. willistoni</i>	Pará, Brasil	-	-	-
		<i>willistoni</i>	<i>D. willistoni</i>	Equador	-	-	-
		<i>willistoni</i>	<i>D. paulistorum</i>	Manaus, Brasil	-	-	-
		<i>willistoni</i>	<i>D. paulistorum</i>	Pará, Brasil	-	-	-
		<i>willistoni</i>	<i>D. paulistorum</i>	Florianópolis, Brasil	-	-	-
		<i>willistoni</i>	<i>D. nebulosa</i>	Porto Alegre, Brasil	-	-	-
		<i>willistoni</i>	<i>D. nebulosa</i>	Florianópolis, Brasil	-	-	-
		<i>willistoni</i>	<i>D. capricorni</i>	B9, Florianópolis, Brasil	-	-	-
		<i>willistoni</i>	<i>D. capricorni</i>	G9, Florianópolis, Brasil	-	-	-
		<i>willistoni</i>	<i>D. sucinea</i>	Porto Alegre, Brasil	-	-	-
		<i>willistoni</i>	<i>D. funipennis</i>	Florianópolis, Brasil	-	-	-
		<i>Drosophila</i>	<i>Drosophila</i>	<i>repleta</i>	<i>D. mercatorum</i>	Porto Alegre, Brasil	-
<i>repleta</i>	<i>D. mercatorum</i>			Florianópolis, Brasil	-	-	-
<i>repleta</i>	<i>D. hydei</i>			Porto Alegre, Brasil	-	-	-
<i>flavopilosa</i>	<i>D. incompta</i>			Santa Maria, Brasil	-	-	-
<i>guarani</i>	<i>D. ornatifrons</i>			Bento Gonçalves, Brasil	-	-	-
<i>guarani</i>	<i>D. ornatifrons</i>			Turvo, Brasil	-	-	-
<i>guarani</i>	<i>D. subbadia</i>			Bowling Green, EUA	-	-	-
<i>guarani</i>	<i>D. griseolineata</i>			Florianópolis, Brasil	-	-	-
<i>guarani</i>	<i>D. griseolineata</i>			Porto Alegre, Brasil	-	-	-
<i>guarani</i>	<i>D. maculifrons</i>			Santa Maria, Brasil	-	-	-
	<i>cardini</i>	<i>D. polymorpha</i>	Porto Alegre, Brasil	-	-	-	

	<i>cardini</i>	<i>D. polymorpha</i>	Turvo, Brasil	+	-	-
	<i>cardini</i>	<i>D. polymorpha</i>	Uruguai	-	-	-
	<i>cardini</i>	<i>D. polymorpha</i>	Florianópolis, Brasil	-	-	-
	<i>cardini</i>	<i>D. cardini</i>	Porto Alegre, Brasil	-	-	-
	<i>cardini</i>	<i>D. neocardini</i>	Florianópolis, Brasil	-	-	-
	<i>cardini</i>	<i>D. neocardini</i>	Porto Alegre, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediopunctata</i>	Florianópolis, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediopunctata</i>	Porto Alegre, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediopicta</i>	Florianópolis, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediopicta</i>	Porto Alegre, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. bandeirantorum</i>	Porto Alegre, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. bandeirantorum</i>	Turvo, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. bandeirantorum</i>	Florianópolis, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediotriata</i>	Porto Alegre, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediotriata</i>	Turvo, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. mediotriata</i>	Bowling Green, EUA	-	-	-
	<i>tripunctata</i>	<i>D. mediosignata</i>	Sapucaia do Sul, Brasil	-	-	-
	<i>tripunctata</i>	<i>D. angustibucca</i>	Florianópolis, Brasil	-	-	-
	<i>pallidipennis</i>	<i>D. pallidipennis</i>	Porto Alegre, Brasil	-	-	-
	<i>pallidipennis</i>	<i>D. pallidipennis</i>	Florianópolis, Brasil	-	-	-
	<i>immigrans</i>	<i>D. immigrans</i>	Porto Alegre, Brasil	+	-	-
	<i>immigrans</i>	<i>D. immigrans</i>	Florianópolis, Brasil	-	-	-
<i>Drosophila</i>	<i>Dorsilopha</i>	<i>D. busckii</i>	Uruguai	-	-	-
		<i>D. busckii</i>	Santa Maria, Brasil	-	-	-
<i>Zaprionus</i>		<i>Z. indianus</i>	Florianópolis, Brasil	-	-	+
		<i>Z. indianus</i>	Porto Alegre, Brasil	-	-	+
<i>Scaptodrosophila</i>		<i>S. latifasciaeformis</i>	São José do Rio Preto, Brasil	-	+	-

+ resultado positivo na análise;

- resultado negativo na análise;

^a sonda utilizada: plasmídeo contendo o retroelemento *Tom* de *D. ananassae*;

^b sonda utilizada: produto de PCR de *D. melanogaster* utilizando os *primers* T12A e T12B;

^c PCR utilizando os *primers* T12A e T12B.

Tabela A2: espécies e linhagens utilizadas para clonagem e seqüenciamento.

Gênero	Subgênero	Grupo	Subgrupo	Espécie	Linhagem/ Local de coleta
<i>Drosophila</i>	<i>Sophophora</i>	<i>melanogaster</i>	<i>melanogaster</i>	<i>D. simulans</i>	Santa Maria, Brasil
		<i>melanogaster</i>	<i>melanogaster</i>	<i>D. sechellia</i>	Arquipélago Seychelles
		<i>melanogaster</i>	<i>melanogaster</i>	<i>D. mauritiana</i>	Tucson Stock Center, EUA
		<i>melanogaster</i>	<i>melanogaster</i>	<i>D. teissieri</i>	África
		<i>melanogaster</i>	<i>montium</i>	<i>D. kikkawai</i>	Florianópolis, Brasil
		<i>melanogaster</i>	<i>ananassae</i>	<i>D. malerkotliana</i>	Florianópolis, Brasil
<i>Zaprionus</i>				<i>Z. indianus</i>	Santa Maria, Brasil

Tabela A3: Número de acesso dos genes nucleares *amd*, *cdc6*, *sod* e *adh* para as espécies utilizadas neste estudo.

Gênero	Subgênero	Grupo	Espécie	<i>amd</i>	<i>cdc6</i>	<i>adh</i>	<i>sod</i>		
<i>Drosophila</i>	<i>Drosophila</i>	<i>virilis</i>	<i>D. virilis</i>	BLAT	DQ839742 ^b	BLAT	BLAT		
		<i>repleta</i>	<i>D. mojavensis</i>	BLAT	DQ839743 ^b	BLAT	BLAT		
	<i>Sophophora</i>	<i>melanogaster</i>	<i>D. ananassae</i>	BLAT	DQ839735 ^b	BLAT	BLAT		
			<i>D. malerkotliana</i>	ND	DQ839736 ^b	AB1944231 ^b	ND		
			<i>D. kikkawai</i>	ND	DQ839733 ^b	AB1545621 ^b	ND		
			<i>D. simulans</i>	BLAT	DQ839682 ^b	BLAT	BLAT		
			<i>D. sechellia</i>	BLAT	DQ839694 ^b	BLAT	BLAT		
			<i>D. mauritiana</i>	ND	DQ839713 ^b	X63953 ^b	ND		
			<i>D. melanogaster</i>	BLAT	DQ839693 ^b	BLAT	BLAT		
			<i>D. erecta</i>	BLAT	DQ839717 ^b	BLAT	BLAT		
			<i>D. yakuba</i>	BLAT	DQ839715 ^b	BLAT	BLAT		
			<i>D. teissieri</i>	AF293727 ^b	DQ839714 ^b	X54118 ^b	AF127160 ^b		
			<i>Zaprionus</i>		<i>Z. indianus</i>	AY699263 ^b	ND	X63955 ^{a, b}	AF021823 ^{a, b}
					TOTAL				

a – a espécie utilizada é *Z. tuberculatus*;

b – número de acesso no GenBank (NCBI);

ND – seqüências não disponíveis para esses genes;

BLAT – seqüências dos genes foram encontradas por buscas pela ferramenta BLAT nos genomas (<http://genome.ucsc.edu/cgi-bin/hgBlat>).