Universidade Federal do Rio Grande do Sul

Evolução da promiscuidade genômica: transferência horizontal de transposons

Mauro de Freitas Ortiz

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

Orientador: Elgion Lucio da Silva Loreto

Porto Alegre, Novembro de 2014.

Este trabalho foi desenvolvido junto ao Laboratório de Biologia Molecular (Labdros) do Departamento de Biologia da Universidade Federal de Santa Maria e junto ao Feschotte Lab da University of Utah. Também contou com apoio do Laboratório de Drosophila da Universidade Federal do Rio Grande do Sul. Contou com o financiamento do CNPq (bolsa), CAPES (bolsa-sanduíche) e UFRGS (missão de curta duração).

Dedico esta tese ao meu pai e à minha mãe.

AGRADECIMENTOS

Ao Dr. Elgion Loreto pela orientação durante esse período e por todo incentivo nesse início da carreira científica.

Aos colegas e amigos do Labdros, Gabriel, Paloma, Lizandra, Daniel, Lenira, Ronaldo, André, Francine, Larissa, Sinara, Stela, Pedro, Valéria, Tailini, Camila, Geovani, Marcos, José, Carla e Crisley pela companhia e por todo o apoio nos vários momentos compartilhados.

À Dra. Vera Valente e integrantes de seu laboratório pela acolhida em parte dessa etapa do meu doutorado.

Ao PPGBM, em especial ao Elmo pelo trabalho de excelência.

Ao Dr. Cedric Feschotte pela orientação durante meu estágio em seu laboratório.

Aos colegas e amigos do Feschotte Lab, Aurelie, Ray, John, Xiaoyu, Ellen, Claudia, Jainy e Rachel pela companhia e por toda a ajuda.

Aos amigos de SLC, Bia, Ricardo, Camille, Leandro, Diana, Oscar, Cris, Fabinho, Dennis, Marina e Osman pela ótima companhia.

À Luciana pela companhia e apoio especiais.

À minha família pelo apoio constante e incondicional.

“Ou o século XXI é dedicado aos valores humanos, morais e éticos... ou de nada valeram os avanços tecnológicos conquistados até aqui.”

Gilson Volpato

Sumário

LISTA DE ABREVIATURAS

DNA	Ácido Desoxirribonucleico - do inglês Deoxyribonucleic Acid
HT	Transferência Horizontal - do inglês Horizontal Transfer
HTT	Transferência Horizontal de Elementos de Transposição ou Transferência Horizontal de Transposons - do inglês Horizontal Transposon Transfer
LTR	Terminação Longa Repetida - do inglês Long terminal repeat
ORF	Quadro Aberto de leitura - do inglês Open Reading Frame
piRNA	Do inglês Piwi-interacting RNA
RNA	Ácido Ribonucleico - do inglês Ribonucleic Acid
TE	Elemento de Transposição - do inglês Transposable Elements
TIR	Terminação Invertida Repetida - do inglês Terminal Inverted Repeats

RESUMO

Os elementos de transposição (TEs) são elementos repetitivos ubíquos dos genomas. São capazes de transpor-se e multiplicar-se, podendo atingir um número de milhares de cópias. A quantidade de TEs nos genomas varia imensamente mas podem atingir até mais de 50% de todo o DNA das espécies. Os TEs comportam-se como parasitas genômicos e eventualmente são domesticados adquirindo alguma função no hospedeiro. Os TEs têm um ciclo de vida nos genomas que começa no momento de sua entrada (invasão), seguido pela amplificação e decaimento (degeneração). O único meio de sobrevivência dos TEs é uma "fuga" para outra espécie. Esse processo é denominado transferência horizontal (HTT, do inglês *Horizontal Transposon Transfer*). Inúmeros casos de HTT já foram descritos, inclusive entre espécies de níveis taxonômicos muito distantes como ordens, classes e até mesmo filos. Neste trabalho utilizamos diferentes abordagens na compreensão dos TEs e do processo da HTT. Na primeira parte parte desse trabalho tratamos acerca da história evolutiva de elementos *P* de *Drosophila*. O grupo foi reclassificado e inúmeros potenciais eventos de HTT foram detectados. Os resultados também confirmam a monofilia dos elementos *P* em *Drosophila* e mostram a existência de 11 subfamílias. Em seguida, testamos possíveis vetores para HTT entre *Drosophila* e vespas parasitóides. Não encontramos evidência de que estes parasitóides sejam vetores de HTT, mas foram detectados eventos de HTT entre diferentes espécies de *Drosophila*. E por fim, analisamos 82 genomas de mamíferos para caracterização do padrão de HTT recentes nestas espécies. Detectamos um grande viés na distribuição das HTT em mamíferos e grupos de morcegos apresentaram uma alta incidência de invasões de TEs nos últimos 50 milhões de anos.

ABSTRACT

The transposable elements (TEs) are repetitive elements ubiquitous of the genome. They are capable to transpose and multiply in the genomes and may reach a number thousands of copies. The amount of TEs in the genomes varies tremendously but can reach up to more than 50% of all DNA from species. The TEs behave as genomic parasites and eventually they are domesticated and acquired some function in the host. The TEs have a life cycle that begins in the genomes invasion, followed by amplification and decay (degeneration). The only way to the TEs survive is to escape to other species. This process is called horizontal transposon transfer (HTT). A lot of cases of HTT have been described, including among very distant species as from different orders, classes and even phyla. In this work we use different approaches in understanding the TEs and the process of HTT. Initially we work on evolutionary history of *P* elements in the *Drosophila* species. The group was reclassified and a lot of potential events HTT were detected. The results also confirm the monophyly of *P* elements in the *Drosophila* species and the existence of 11 subfamilies. Then we tested for potential vectors to HTT between *Drosophila* and parasitoid wasps. We not found evidence that these parasites are vectors of HTT but some HTT events between different *Drosophila* species were detected. Finally, we analyzed 82 mammalian genomes for characterization of the recent HTT pattern in this group. We detected a large bias in the distribution of HTT in mammals and some bats showed a high incidence of TE invasions in the last 50 million years.

CAPÍTULO I

Introdução

Os elementos de transposição (TEs) são elementos repetitivos ubíquos dos genomas. Possuem a capacidade de transpor-se e multiplicar-se levando algumas famílias de TEs a atingirem milhares de cópias. A quantidade de TEs pode variar imensamente nos genomas, como por exemplo em espécies de *Drosophila* onde os TEs representam de 3% a 25% dependendo da espécie (Clark 2011). Em outros organismos podem atingir proporções ainda mais significativas, como no genoma humano onde chegam a cerca de 45% do seu total (Lander et al 2001) ou ainda de forma mais extrema, como 85% no genoma do milho (Schnable et al 2009). O fato é que os TEs são amplamente presentes nos organismos eucariotos, desde os mais simples até os mais complexos. Na maioria das vezes correspondem a porções significativas dos genomas e em muitos casos são até mesmo a maior parte destes.

A descoberta dos TEs deu-se entre 1944 e 1947 por Barbara McClintock durante seus estudos citogenéticos com milho. Durante sua investigação sobre os padrões de cores que levavam ao mosaicismo nos grãos de milho ela identificou recorrentes quebras cromossômicas as quais atribuiu a elementos que mudavam de lugar no genoma. Ela identificou dois elementos, *Dissociator* (*Ds*) e *Activator* (*Ac*), que denominou de elementos controladores (McClintock 1950). McClintock verificou que os elementos *Ds* causavam quebras cromossômicas e deixavam algumas mutações instáveis quando *Ac* estivesse presente. E em seguida inferiu que *Ds* e *Ac* poderiam mudar de posição nos cromossomos.

A classificação dos TEs divide-os tradicionalmente em duas grandes classes com base nos seus mecanismos de transposição. Os elementos da classe I, chamados retrotransposons, são aqueles que possuem um intermediário de RNA na transposição, enquanto os elementos da classe II (transposons de DNA) transpõe-se diretamente por excisão e inserção em um novo sítio (Capy et al 1998). Os retrotransposons codificam a enzima transcriptase reversa

necessária para copiar seu RNA em DNA durante sua replicação entre outras enzimas também necessárias nesse processo. Os elementos de classe I ainda se dividem em dois grupos, um grupo com longas terminações repetidas (LTRs) e outro sem (non-LTRs). Já os transposons de DNA tipicamente possuem uma estrutura mais simples, possuem apenas uma ORF que codifica uma transposase, enzima necessária para sua remoção de um sítio e inserção em outro, são flanqueados por uma repetição terminal invertida (TIRs) e geralmente são menores que 5Kb (Wicker et al 2007, Pritham 2009).

Desde sua descoberta, os TEs passaram três décadas até serem redescobertos e aceitos como elementos móveis. Apesar disso, permaneceram sendo vistos como componentes deletérios ou sem função nos genomas e adquirindo inicialmente a alcunha de DNA lixo. De fato, hoje sabemos que os TEs possuem na maior parte das vezes um comportamento egoísta, atuando como parasitas genômicos (Palazzo e Gregory 2014). Mas uma crescente literatura tem mostrado que além da face egoísta, os TEs também têm o potencial de contribuírem na evolução, regulação gênica e na criação de genes estruturais nos hospedeiros (Feschotte 2008, Hua-Van et al 2011, Cowley e Oakey 2013). Nesse processo, chamado domesticação, os TEs mudam de uma relação parasitaria para uma relação benéfica ao seu hospedeiro (Kidwell e Lisch 2000).

A maioria dos TEs são inativos. São eles que compõem a maior parte da heterocromatina, uma porção do genoma que possui poucos genes, mas muitos TEs, geralmente, apresentam uma baixa atividade de expressão gênica (Lippman et al 2004). No geral, os TEs são muito ativos quando estes são recentes em um genoma e tornam-se menos ativos ou inativos ao longo do tempo (Rouzic e Capy 2005). Mas estes também podem ser ativados em diferentes situações como estresse ou eventos de hibridização. Entre os tipos de estresse que podem mobilizar TEs estão o estresse térmico, químico entre outros fatores ambientais (Capy et al 2000). Os mais numerosos e melhores casos documentados de mobilização de TEs após hibridização foram reportados em plantas (Rieseberg et al 1995, Josefsson et al 2006, Ungerer et al 2006, Michalak 2009). Mas também há evidências de atividades de TEs relacionadas à hibridização em outros grupos

(Petrov et al 1995, O'Neil et al 1998, Fontdevila 2005, Carnelossi et al 2014).

Como dito anteriormente, os TEs são muito ativos quando recentes nos genomas. Isso acontece porque logo após a invasão do genoma por um novo TE, o genoma ainda não possui mecanismos de controle específicos para o novo invasor. No entanto, os genomas possuem alguns mecanismos de reconhecimento e repressão de elementos genéticos invasores, incluindo TEs, que serão ativados para tentar conter uma replicação potencialmente danosa do novo TE. Entre estes estão distintos classes de pequenos RNAs chamados Piwi-interacting RNAs (piRNAs) (Aravin et al 2007, Siomi et al 2011). Além disso outras formas de controle atuam na regulação dos TEs, incluindo outros processos de controle epigenético, como a metilação e modificações de histonas (Hua-Van et al 2011).

Paralelamente aos seres vivos, os TEs também possuem um ciclo de vida. Este processo começa quando uma família de TEs coloniza um novo genoma (invasão) e termina quando todas as cópias de elementos desse família são perdidos ou inativados. De modo geral todos os passos deste processo são invasão, proliferação no genoma, proliferação na população, diversificação e inativação, com a consequente eliminação do genoma. Há duas possibilidades para os TEs escaparem da extinção, a primeira é a transferência horizontal do TE para um novo hospedeiro antes da inativação e a segunda é gerando o mínimo possível de efeitos deletérios de modo a passar desapercebido pela seleção natural (Schaack et al 2010).

Uma vez controlados dentro do genoma, os TEs inativos podem ser perdidos de forma estocástica mas, eventualmente, podem ser aproveitados pelo seu hospedeiro por via de um processo de domesticação molecular ou também chamado de exaptação (Gould e Vrba 1982, Sinzelle et al 2009). A degradação de TEs no genoma é um caminho frequente após sua inativação. Tanto que porções significativas dos genomas de eucariotos é composta por esse tipo de material, fragmentos de TEs. Mas os TEs também possuem o potencial de fornecer para seu hospedeiros componentes de sua própria estrutura como domínios enzimáticos e regiões regulatórias (Feschotte 2008, Rebollo et al 2012). Alguns

casos de domesticação são os casos de TEs reparadores de telômeros em *Drosophila* (George et al 2010) ou TEs geradores da recombinação somática em vertebrados (Kapitonov e Jurka 2005).

Uma característica marcante dos TEs, é que além de serem transmitidos verticalmente nas linhagens, eles também podem transferir-se horizontalmente nas espécies. Hoje isso parece ser muito mais comum do que pensava-se (Schaack et al 2010). Uma última revisão contabilizou 330 casos de transferência horizontal de TEs (HTT) em eucariotos (Wallau et al 2012). Mas isso possivelmente ainda é apenas a ponta do iceberg em termos do número de HTTs. A origem da maioria das TEs ainda é desconhecida e possivelmente a maioria das colonizações de novos genomas pelos TEs seja resultado de uma invasão por HTT. Cada espécie tem um padrão da sua composição de elementos repetitivos assim como também variam quanto ao seu histórico de invasões do seu genoma. A análise dos genomas mostra que algumas espécies possuem pouca ou nenhuma atividade recente de TEs e invasões, como é o caso dos grandes primatas (Pace and Feschotte 2007), enquanto outras apresentam ondas de invasões recentes e muita atividade de TEs, como é o caso dos morcegos Vespertilionidae (Ray et al 2008). Os padrões de invasões dos genomas ainda mostram que um mesmo TE pode invadir múltiplas espécies em um mesmo período, assim como espécies muito distantes, até mesmo de Classes ou Filos diferentes (Robertson 1997, Gilbert et al 2010).

Já foram contabilizados eventos de HTT em diversos grupos de eucariotos (Schaack et al 2010, Wallau et al 2012). A grande maioria dos casos foram descritos em animais, no entanto tudo indica que o número de casos de em plantas ainda seja subestimado (Baidouri et al 2014). A maior parte dos casos relatados em animais são em artrópodes, mais especificamente em *Drosophila*. Ainda não é claro se estes organismos possuem alguma propensão a este fenômeno ou se isso é apenas um viés amostral, visto a quantidade de estudos com estes organismos. Mas independente disso, o fato é que a literatura têm mostrado cada vez mais casos nos mais variados organismos. Dentre os grupos de grupos de TEs que destacam-se, são mais numerosos casos nas superfamílias

de elementos de RNA, *gypsy* e entre os elementos de DNA, nas superfamílias *mariner* e *hAT* (Wallau et al 2012).

A transferência horizontal pode ser inferida principalmente com base em três tipos de evidências que são a alta similaridade entre TEs de espécies diferentes, as incongruências entre a filogenia dos TEs e a dos hospedeiros e a distribuição descontínua dos TEs (Silva et al 2004). Estas evidências ainda podem ser usadas em conjunto para a detecção de eventos de HTT. Também cabe ressaltar que outros processos evolutivos do ciclo de vida dos TEs, como perda estocástica, variação de taxas evolutivas e o polimorfismo ancestral (Capy, 1994; Cummings, 1994) podem dificultar a identificação destes padrões usados na detecção de HTT e em alguns casos até serem a explicação desses padrões.

É considerada uma evidência de HTT quando as sequências de TEs de duas espécies são muito similares e estes valores não correspondem com o tempo de divergência esperado para estas espécies. Neste caso podem ser comparados os valores de divergência dos TEs com os de genes das espécies hospedeiras. Caso os TEs sejam herdados por forma vertical, ou seja, estão sendo mantidos no genoma das espécies junto com o genes constitutivos, desde o último ancestral comum até os descendentes atuais, é esperado que a divergência seja semelhante a dos genes. Mas caso o TE tenha sido adquirido por HT é esperado que a divergência significativamente menor que a dos genes constitutivos. Esse padrão é esperado, porque os TEs que foram compartilhados após o isolamento reprodutivo das espécies, acumularam diferenças somente depois do evento de HTT (Silva e Kidwell, 2000; para revisão ver Wallau et al 2012).

Embora fique evidente que há esse fluxo de TEs entre as espécies, alguns pontos ainda são pouco claros. Isso inclui: i) como ocorrem os eventos de HTT e quais seus vetores; ii) com que frequência ocorrem e, iii) quais as implicações biológicas. Dentre os possíveis meios de ocorrem eventos de HTT sugere-se uma diversidade de potenciais vetores. Entre estes estão os próprios TEs circularizados (O'Brochta et al 2009), vírus (Piskurek e Okada 2007; Routh et al 2012), bactérias (Hotopp et al 2007; Klasson et al 2009), parasitas (Gilbert et al

2010) e parasitóides (Yoshiyama et al 2001). A relação entre parasita e hospedeiro aparentemente tem um forte potencial em promover a HTT. É o que mostra a relação do triatomíneo parasita *Rhodnius prolixus* com alguns de seus hospedeiros, em que em alguns casos alguns TEs desse inseto chegam a compartilhar TEs 98% similares com o hospedeiro (Gilbert et al 2010). Cabe também salientar que os vetores não precisam, necessariamente, que os TEs sejam integrados em seus genomas para estes sirvam como intermediários de uma HTT.

Já a algum tempo sabemos que os TEs são onipresentes dos genomas dos eucariotos. E hoje sabemos que além da transmissão vertical muitos TEs também acabam sendo transferidos de um genoma para outro. E esse fluxo de TEs pelos genomas têm um potencial impacto na evolução dos genomas. No entanto, ainda sabe-se muito pouco sobre a extensão exata desse fenômeno, assim como acerca dos mecanismos e potenciais vetores que levam os TEs de um genoma para outro. É com o intuito de esclarecer pontos como estes que foi proposta esta tese. Pontos desde mais específicos, como a evolução da superfamília de elementos *P* em *Drosophila* (Capítulo 1) ou sobre o potencial de vespas parasitoides como vetores também em *Drosophila* (Capítulo 2) ou então mais amplos como o entendimento da distribuição dos eventos de HTT em mamíferos (Capítulo 3).

Objetivos

O objetivo geral desta tese é contribuir para o entendimento do processo biológico transferência horizontal de elementos de transposição de eucariotos.

Capítulo II - Reconstruir a história evolutiva dos elementos *P* em espécies de *Drosophila* neotropicais.

Capítulo III - Testar o papel de vespas parasitóides como potenciais vetores de transferência horizontal de transposons em *Drosophila*.

Capítulo IV - Descrever os padrões de distribuição de transferência horizontal de transposons em mamíferos.

CAPÍTULO II

Trabalho publicado na revista *Molecular Genetics and Genomics*[#]

New *Drosophila P*-like elements and reclassification of *Drosophila P*-elements subfamilies

Elgion L. S. Loreto* [1,2]; Francis M.B. Zambra [2]; Mauro F. Ortiz[2]; Lizandra J. Robe [3].

1- Dep de Biologia, CCNE, Santa Maria Federal University, Av. Roraima 1000, 97105-900, Santa Maria, RS, Brazil. eMail: elgion.loreto@pq.cnpq.br

2- PPGGBM, UFRGS; Univ. Fed do Rio Grande do Sul, Porto Alegre, RS, Brazil

3- ICB, Univ. Fed. do Rio Grande- FURG; Rio Grande, RS; Brazil.

* Corresponding author

Este trabalho fez parte do Trabalho de Conclusão de Curso da aluna Francis M. B. Zambra junto ao Curso de Ciências Biológicas da UFSM.

Mol Genet Genomics
DOI 10.1007/s00438-012-0691-y

ORIGINAL PAPER

New *Drosophila P*-like elements and reclassification of *Drosophila P*-elements subfamilies

Elgion L. S. Loreto · Francis M. B. Zambra · Mauro F. Ortiz · Lizandra J. Robe

Received: 1 November 2011 / Accepted: 12 April 2012

Abstract Genomic searches for *P*-like transposable elements were performed (1) in silico in the 12 available *Drosophila* genomes and (2) by PCR using degenerate primers in 21 Neotropical *Drosophila* species. In silico searches revealed *P*-like sequences only in *Drosophila persimilis* and *Drosophila willistoni*. Sixteen new *P*-like elements were obtained by PCR. These sequences were added to sequences of previously described *P*-like elements, and a phylogenetic analysis was performed. The subfamilies of *P*-elements described in the literature (Canonical, M, O, T, and K) were included in the reconstructed tree, and all were monophyletic. However, we suggest that some subfamilies can be enlarged, other subdivided, and some new subfamilies may be proposed, totalizing eleven subfamilies, most of which contain new *P*-like sequences. Our analyses support the monophyly of *P*-like elements in Drosophilidae. We suggest that, once these elements need host-specific factors to be mobilizable, the horizontal transfer (HT) of *P*-like elements may be inhibited among more distant taxa. Nevertheless, HT among Drosophilidae species appears to be a common phenomenon.

Keywords Lateral transfer · *P* superfamily · Horizontal transfer · *Drosophila* · Transposable elements

Introduction

The *Drosophila P*-element has been a model for pioneering studies of eukaryotic transposable elements (TE). These studies began with the study of a now well-characterized biological phenomenon, called hybrid dysgenesis (HD), in which hypermutability, chromosome rearrangements, and gonadal sterility occur in crosses of some strains (Kidwell et al. 1977). Not long after, the agent of HD was cloned and named the *P*-element. The *P*-element was one of the first eukaryotic TE to be molecularly characterized (Bingham et al. 1982; Rubin et al. 1982). It quickly became a revolutionary tool for genetic research. For example, it was used as a mutagenic agent, as a tool for gene cloning by transposon tagging, as a vector for genetic transformation and transgenic expression, and as a vector for site-specific recombination and gene replacement (review in Engels 1989).

The *P*-element has also been important in evolutionary studies and was the subject of the first well-documented example of horizontal transfer (HT) of eukaryotic TE (Daniels et al. 1990). After the molecular characterization of the *P*-element of *Drosophila melanogaster*, similar sequences were also found in other *Drosophila* species, but not in species closely related to *D. melanogaster* (Lansman

Communicated by G. Reuter.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-012-0691-y) contains supplementary material, which is available to authorized users.

E. L. S. Loreto (✉)
Dep de Biologia, CCNE, Santa Maria Federal University, Av. Roraima 1000, Santa Maria, RS 97105-900, Brazil
e-mail: elgion.loreto@pq.cnpq.br

E. L. S. Loreto · F. M. B. Zambra · M. F. Ortiz
PPGBM, UFRGS, Univ. Fed do Rio Grande do Sul, Porto Alegre, RS, Brazil

L. J. Robe
ICB, Univ. Fed. do Rio Grande—FURG, Rio Grande, RS, Brazil

Published online: 20 May 2012

et al. 1985; Brookfield et al. 1984). The presence of significant divergences among *P* sequences from the *willistoni* group (even though the *P*-elements of *Drosophila willistoni* and *D. melanogaster* are almost identical) in concert with the fact that old laboratory strains of *D. melanogaster* do not contain this TE in their genomes indicate that *D. melanogaster* obtained its *P*-element by HT from *D. willistoni* (Daniels et al. 1990). In only 30 or 40 years, this TE spread into the genomes of *D. melanogaster* worldwide (Bregliano and Kidwell 1983).

Phylogenetic analyses of *P*-like sequences from different *Drosophila* species reveal high levels of diversification, which suggests the possibility of additional subdivision. In addition to canonical elements, Hagemann et al. (1992, 1994, 1996a, b) described three *P*-element subfamilies (M-, O-, and T-type) in the Old World *obscura* species group. The M- and O-types also occur in the *saltans* and *willistoni* groups (Setta et al. 2007). Nouaud et al. (2003) described the K-subfamily, which is restricted to the *montium* subgroup of the *melanogaster* group.

Incongruence between the *P*-element phylogeny and those of *Drosophila* indicates that HT is a recurrent phenomenon in the evolutionary history of the *P*-element (Clark and Kidwell 1997; Clark et al. 1994, 1995). Comparison of the ratio of the number of synonymous substitutions per synonymous site (*d*S) to the number of nonsynonymous substitutions per nonsynonymous site (*d*N) between *P* sequences and nuclear host genes suggest that multiple HT events within the *Drosophila* genus have occurred during *P*-element evolution (Silva and Kidwell 2000, 2004).

Other dipteran insects are also hosts to *P*-like elements, for example, the Australian sheep blowfly *Lucilia cuprina* (Perkins and Howells 1992), the house fly *Musca domestica* (Lee et al. 1999), and the Drosophilidae *Scaptomyza pallida* (Simonelig and Anxolabéhère 1991). The mosquitoes *Anopheles gambiae* and *Aedes aegypti* have a high copy number of highly diversified *P*-like sequences in their genomes (Sarkar et al. 2003; Carvalho et al. 2004; Nene et al. 2007). In vertebrates, *P*-like elements have been found in humans and chickens as a stationary, single-copy gene (Hagemann and Pinsker 2001; Hammer et al. 2005); however, there are typical and putatively active *P*-elements in the fish *Danio rerio* and in the urochordate *Ciona intestinalis* (Kimbacher et al. 2009).

The availability of the sequences of 12 *Drosophila* genomes allowed us to search for new *P*-like elements. This search was complemented by PCR searches for *P*-like sequences in the genomic DNA of Neotropical *Drosophila* species that have not been previously tested. Using an analysis of both the existing and newly obtained sequence data, we were able to reconstruct the putative evolution of *P*-like elements within the genus *Drosophila*.

Materials and methods

Identification of *P*-like transposable element sequences in Neotropical *Drosophila*

Genomic DNA from 21 Neotropical *Drosophila* species (Supplementary spreadsheet 1) was prepared following Oliveira et al. (2009). *P*-like transposable element sequences were amplified by PCR using degenerated primers and the PCR conditions described by Lee et al. (1999). Amplified fragments were approximately 800 bp long, corresponding to parts of exons 1 and 2 of the transposase gene. Fragments were cloned using the *TA Cloning® Kit* (Invitrogen). Sequencing was performed using the DYEnamic ET Dye Terminator Cycle Sequencing Kit (GE Healthcare) and a Megabace 500 automatic sequencer. Both DNA chains were sequenced and assembled using Gap4 software of the Staden Package (Staden 1996). For each cloned sequence, runs were included in the assemblage until sequence confidence was higher than 30. Seven to ten clones were sequenced per species. Similarity with *P*-elements was confirmed using BLASTN on the NCBI website.

Genomic searches

We searched for *P*-like elements in the ''nr/nt'' database of GenBank (http://www.ncbi.nih.gov) using the program BLASTN with default parameters (Altschul et al. 1997). As a query, we used sequences representative of the 16 subfamilies previously described for *Drosophila* (Clark and Kidwell 1997; Hagemann et al. 1994, 1996a, b; Silva and Kidwell 2004; Castro and Carareto 2004; Setta et al. 2007). The accession numbers of these sequences are presented in Supplementary spreadsheet 1. The BLASTN outputs were parsed with Perl scripts.

Additional searches for *P*-like transposable elements were conducted in the *D. melanogaster*, *Drosophila simulans*, *Drosophila sechellia*, *Drosophila yakuba*, *Drosophila erecta*, *Drosophila ananassae*, *Drosophila pseudoobscura*, *D. persimilis*, *Drosophila virilis*, *Drosophila mojavensis*, *Drosophila grimshawi*, *D. willistoni*, *Culex pipiens*, *A. aegypti*, *A. gambiae*, *Bombyx mori*, *Tribolium castaneum*, *Apis mellifera*, *Nasonia vitripennis*, *Acyrthosiphon pisum,* and *Pediculus humanus corporis* genomes using the BLAST tool as available in Flybase (http://flybase.bio.indiana.edu/blast/; Grumbling and Strelets 2006). Also, searches were done in the Trace Archive Nucleotide BLAST (NCBI). All retrieved sequences were also used as queries until no new sequences were obtained. For each search, all sequences with a score >50 and a percent identity >50 were retrieved.

Phylogenetic analysis

Sequences were first aligned using the Clustal W algorithm of the MEGA5 program (Tamura et al. 2011). Once the intervening intron was removed, the resulting 116 × 841 matrix was used to perform an initial Bayesian analysis (BA) in MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001). The analyses used the GTR + G model, as selected by the AIC test (Akaike 1974) implemented in MrModelTest 2.2.1 (Nylander 2004). The search was conducted with uninformed priors for 20,000,000 generations, with trees saved every 1,000 generations and a 25 % burn in.

To reduce computation time, intraspecific sequences with less than 3 % *p*-distance with a monophyletic origin in the resulting Bayesian phylogeny were collapsed in the form of a majority-rule consensus sequence. These consensus sequences were built using Cons software from the EMBOSS suite (http://emboss.sourceforge.net/), available from the Mobyle Portal (Nerón et al. 2009).

The resulting 52 × 819 matrix was further subjected to the following phylogenetic analyses: (1) neighbor-joining (NJ) (Saitou and Nei 1987) performed in PAUP 4.0b10 (Swofford 2003), specifying the transitional model with a gamma distribution as indicated by the AIC test performed in ModelTest3.7 (Posada and Crandall 1998); (2) maximum parsimony (MP) performed in PAUP 4.0b10 through a heuristic search with Tree-Bisection-Reconnection (TBR) rearrangements applied to 100 random stepwise addition trees; (3) maximum likelihood (ML) in PhyML2.4.4 (Guindon and Gascuel 2003), using the AIC test best-fit model (ModelTest 3.7); (4) BA implemented in MrBayes 3.1.2, employing the GTR + G model selected by the AIC Test performed in MrModelTest 2.2.1, with the above-described MCMC search parameters. Confidence was measured through 1,000 bootstrap replications in the NJ, MP, and ML searches; for the BA, the posterior probability (PP) of each clade was evaluated.

Divergence estimates

The mean nucleotide distance between clades/lineages reconstructed by the Bayesian 50 % majority-rule tree was estimated with the transitional model with a gamma distribution (as indicated by the AIC test performed in ModelTest3.7) using the PAUP 4.0b10 software. Conversely, the mean amino acid sequence distance between clades/lineages was measured with the Jones–Taylor–Thornton model (Jones et al. 1992) with a gamma distribution [as indicated by the AIC test performed by ProtTest 2.4 (Abascal et al. 2005) using the MEGA5 software.

A codon alignment of the 116 × 841 matrix was used to estimate the number of synonymous substitutions per synonymous site (*d*S) and the number of synonymous sites (S), which were both estimated with the Nei–Gojobori equation (1986) computed with MEGA5 software using p-distances. In this case, indels relative to the canonical *P* sequence of *D. melanogaster* were removed to conserve the open reading frames of some sequences, whereas stop codons were considered as missing information. The α-methyldopa (*amd*), alcohol dehydrogenase (*adh*), and/or amyrel (*amy*) nuclear genes were used in dS comparisons to test for horizontal transfer (using one-tailed Fisher's exact tests; Ludwig et al. 2008). R scripts (R Development Core Team 2011) were used to conduct multiple *dS* tests.

The estimated time since HT events occurred was done according to the equation $T = k/2r$ (T = divergent time between species, $k = d$S between TE sequences; r = evolutionary rate) (Graur and Li 2000). We used the synonymous substitution rate of 0.016 substitutions per site per million years (Sharp and Li 1989).

Results

Searching new *P*-like elements in *Drosophila*

Twenty-one species were screened by PCR for *P*-like transposable elements, and amplicons of the expected size were identified in 16 of them (Table 1). However, some amplicons showed no sequence similarity with *P*-elements. These results suggest that PCR screening is insufficient for estimating the presence or absence of TEs because of the potential for false positives or negatives. In total, *P*-like elements were confirmed in eight species and identified for the first time in the *cardini* group (*Drosophila procardinoides*, *Drosophila cardini,* and *Drosophila cardinoides*); in previously untested species of the *tripunctata* group (*Drosophila angustibucca*, *Drosophila bandeirantorum,* and *Drosophila mediopicta*); and in two species of the *willistoni* group (*Drosophila sucinea* and *Drosophila fumipennis*) (Table 1). The sequences were deposited in GenBank:JQ915140-JQ915155 and are also available in supplementary material spreadsheet 1.

Searches for new *P*-like elements in the 12 *Drosophila* genomes resulted in the detection of one sequence in *D. persimilis* and 19 different sequences in *D. willistoni*. These 19 sequences were clustered into five different consensus and two ungrouped sequences, of which five have not been formerly described (see Supplementary spreadsheet 1). *P*-like sequences were not identified in the remaining ten *Drosophila* genomes or in any of the non-*Drosophila* genomes.

Diversity and phylogenetic relationships of *P*-like elements in *Drosophila*

P-like transposable elements represent a diversified group of TEs in the genomes of *Drosophila* species. The phylogenetic analyses using the *Lucilia cuprina P*-element as an

Table 1 List of species that were screened by PCR and Blast for *P*-like elements

Family	Genus	Species group	Species	Amplicon[a]	*P*-like[b]	Genome
Drosophilidae	*Drosophila*	*melanogaster*	*D. melanogaster*			−
			D. simulans			−
			D. sechellia			−
			D. yakuba			−
			D. erecta			−
			D. ananassae			−
		obscura	*D. pseudoobscura*			−
			D. persimilis			+
		willistoni	*D. willistoni*			+
			D. sucinea	+	+	
			D. fumipennis	+	+	
		repleta	*D. mojavensis*			−
		virilis	*D. virilis*			−
		grimshawi	*D. grimshawi*			−
		mesophragmatica	*D. brncici*	+	−	
			D. gasici	+	−	
			D. gaucha	+	−	
			D. pavani	+	−	
		cardini	*D. polymorpha*	+	−	
			D. procardinoides	+	+	
			D. dunni	+	−	
			D. cardini	+	+	
			D. cardinoides	+	+	
			D. neocardini	−	−	
		tripunctata	*D. mediostriata*	+	−	
			D. angustibucca	+	+	
			D. bandeirantorum	+	+	
			D. mediopicta	+	+	
		guaramunu	*D. griseolineata*	+	−	
		guarani	*D. ornatifrons*	−	−	
		immigrans	*D. immigrans*	−	−	
		pallidipenis	*D. pallidipenis*	−	−	
	Scaptodrosophila		*S. galloi*	−	−	
Culicidae	*Culex*		*C. pipiens*			−
	Aedes		*A. aegypti*			−
	Anopheles		*A. gambiae*			−
Bombycidae	*Bombyx*		*B. mori*			−
Tenebrionidae	*Tribolium*		*T. castaneum*			−
Apidae	*Apis*		*A. mellifera*			−
Pteromalidae	*Nasonia*		*N. vitripennis*			−
Aphididae	*Acyrthosiphon*		*A. pisum*			−
Pediculidae	*Pediculus*		*P. humanus*			−

[a] Amplicons with expected size (±800 bp) obtained (+) or not (−)

[b] Sequenced amplicons showing similarity with *P*-elements (+) or not (−)

outgroup under the BA, ML, MP, and NJ methods held similar results, suggesting that *P*-elements can be subdivided into eleven groups or subfamilies (Fig. 1) based on the criteria of recurrent monophyly according to all phylogenetic analyses and internal divergences less than 0.3 at the amino acid level (Capy et al. 1998):

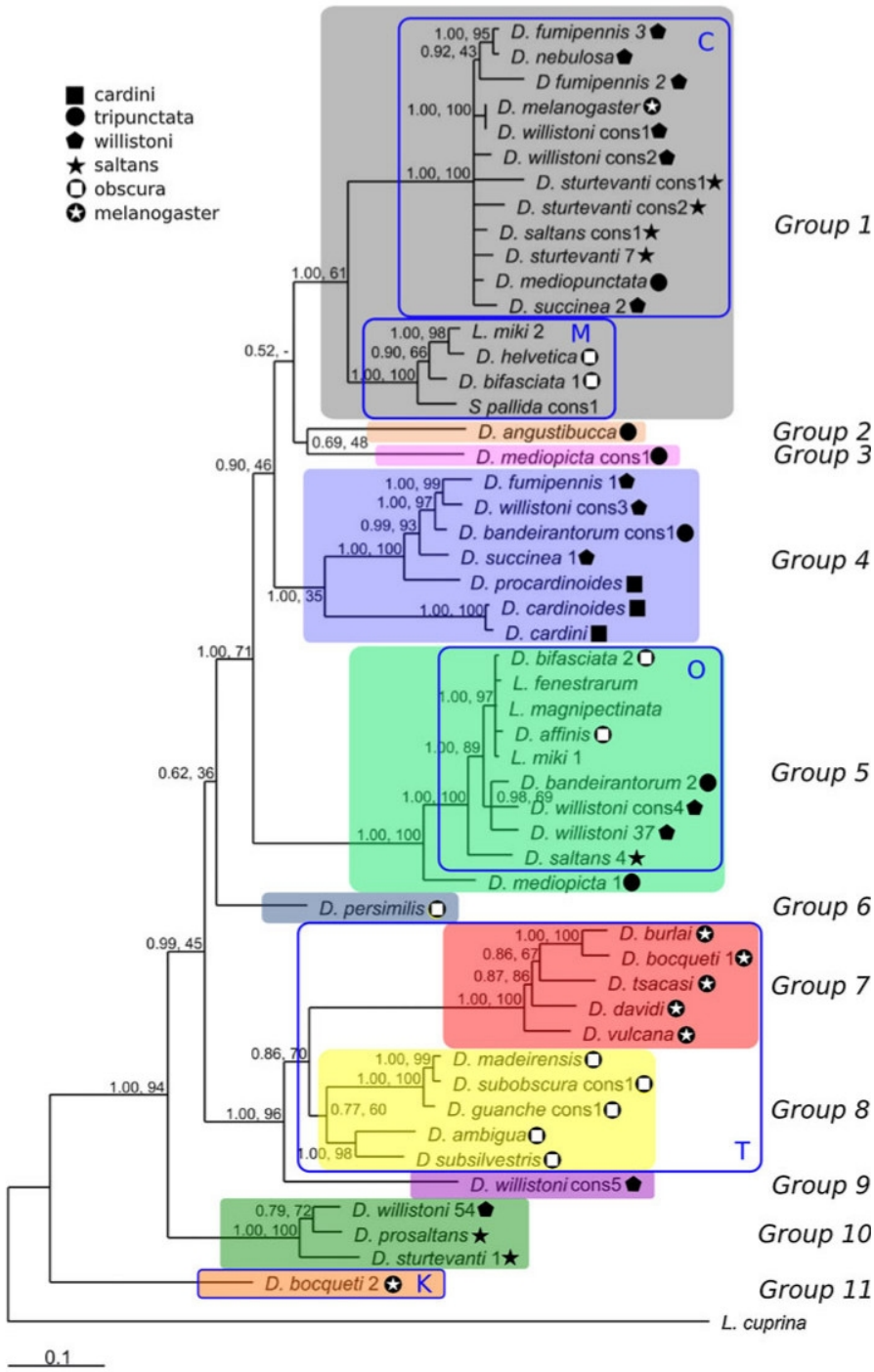

Fig. 1 Phylogenetic analysis of *P*-like sequences using nucleotide sequences (Bayesian Analysis, GTR + G model selected according to the AIC test). Numbers at nodes are the posterior probability estimates for each clade, followed by the bootstrap values attained for the same clades in the ML analysis. Groups identified by this analysis are numbered 1 through 11. Subfamilies described in the literature are indicated inside the *boxes* (*C* for canonical, *M*, *O*, *T*, and *K*). *Drosophila* species groups are indicated by *colored dots*. *P* sequence of *Lucilia cuprina* was used as outgroup (colour figure online)

1. Group 1 comprises the canonical *P*-element of *D. melanogaster* and related sequences (i.e., the canonical subfamily), including the "M" subfamily (Hagemann et al. 1992, 1994, 1996a, b). Sequences of this clade are widely distributed within the *Drosophila* genus, occurring in the four main groups of the *Sophophora* subgenus (*willistoni*, *saltans*, *melanogaster*, and *obscura*) and in one species of the *Drosophila* subgenus (*D. mediopunctata*, a member of the *tripunctata* group). Sequences of this clade were also encountered in some other Drosophilidae genera (*Lordiphosa* and *Scaptomyza*).
2. Group 2 is for now a monotypic subfamily, found only in *D. angustibucca*, a member of the *tripunctata* group (*Drosophila* subgenus).

3. Group 3 encompasses two similar sequences isolated from *D. mediopicta*, another member of the *tripunctata* group (*Drosophila* subgenus).
4. Group 4 includes a set of sequences heterogeneously distributed in three species of the *willistoni* group of the *Sophophora* subgenus (*D. fumipennis*, *D willistoni,* and *D. succinea*) and in four species of the *Drosophila* subgenus (one of which belongs to the *tripunctata* group—*D. bandeirantorum*—whereas the other three are part of the *cardini* group—*D. procardinoides*, *D. cardinoides,* and *D. cardini*).
5. Group 5 consists of *P*-elements previously described as members of the "O" subfamily, present in species of the *obscura*, *willistoni,* and *saltans* groups of the *Sophophora* subgenus and in the genus *Lordiphosa* (Hagemann et al. 1992, 1994, 1996a, b; Setta et al. 2007). We also found representatives of this group in two *tripunctata* group species (*D. bandeirantorum* and *D. mediopicta*), both of which also possess *P*-element sequences related to other groups.
6. Group 6 is formed by the unique *P*-like element found in *D. persimilis*, a member of the *obscura* group of the subgenus *Sophophora* of *Drosophila*.
7, 8. Groups 7 and 8 consist of *P*-like sequences restricted to the *Sophophora* subgenus species of the *montium* subgroup (*melanogaster* group) and the *obscura* group, respectively. Both clades were described previously as the "T" subfamily (Hagemann et al. 1996b).
9. Group 9 consists of two sequences found only in the *D. willistoni* genome.
10. Group 10 consists of sequences found in the *Sophophora* subgenus species *D. willistoni* and in two species of the *saltans* group (*D. prosaltans* and *D. sturtevanti*).
11. Group 11 comprises a sequence found in *Drosophila bocqueti* (*melanogaster* group—*Sophophora* subgenus), previously assigned to the K-subfamily.

The greatest nucleotide distance observed between different groups was 0.72 (between groups 7 and 11) and the smallest was 0.28 (between groups 6 and 10) (Table 2). Analyses of amino acid distances also revealed the greatest distance between groups 7 and 11 (0.67) and the shortest between groups 1 and 4 and 6 and 10 (0.26).

Comparisons of synonymous site divergence (*d*S) between transposable elements and host genes

The divergence time among different *P*-elements can be estimated for *d*S pairwise comparisons. Some features of the evolution of TEs, such as the occurrence of horizontal transfer, can be inferred by comparing the *d*S values of TEs with those of host genes. In the absence of strong codon-usage bias, *d*S values provide a measure of neutral evolution. When transmitted vertically during evolution, *d*S values should be similar among TE and host genes, especially considering host genes that present low codon-usage bias, since this is also a common TE feature (Vidal et al. 2009). In contrast, when HT occurs, values of *d*S should be lower between TEs than those observed between host genes (Silva and Kidwell 2000; Ludwig et al. 2008).

Pairwise comparisons between the *d*S values of 52 *P*-like sequences and those of nuclear host genes were performed, rendering a total of 1326 comparisons (Supplementary spreadsheet 2). In this case, in order to avoid HT underestimation, the *Amd* gene was chosen whenever possible, since it was previously shown that this gene presents codon-usage bias more similar to those presented by TEs (Vidal et al. 2009). Nevertheless, when *Amd* sequences were not available, *Adh* and *Amyrel* sequences

Table 2 Genetic distance between each of the inferred *P*-like sequence groups or subfamilies at the nucleotide (below diagonal) and at amino acid levels (above diagonal)

	Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	Gr 9	Gr 10	Gr 11
Gr 1		0.32	0.29	0.26	0.35	0.28	0.56	0.35	0.33	0.32	0.39
Gr 2	0.36		0.30	0.33	0.41	0.33	0.61	0.39	0.43	0.43	0.52
Gr 3	0.38	0.32		0.32	0.40	0.29	0.62	0.41	0.39	0.43	0.47
Gr 4	0.34	0.35	0.37		0.36	0.29	0.54	0.33	0.37	0.33	0.37
Gr 5	0.43	0.40	0.39	0.42		0.40	0.65	0.42	0.44	0.42	0.43
Gr 6	0.33	0.31	0.31	0.33	0.35		0.54	0.30	0.34	0.26	0.42
Gr 7	0.65	0.61	0.62	0.57	0.61	0.52		0.48	0.52	0.49	0.67
Gr 8	0.46	0.44	0.44	0.41	0.47	0.31	0.38		0.31	0.37	0.48
Gr 9	0.42	0.46	0.41	0.43	0.52	0.35	0.49	0.30		0.35	0.42
Gr 10	0.40	0.43	0.42	0.39	0.43	0.28	0.49	0.42	0.42		0.42
Gr 11	0.53	0.55	0.58	0.52	0.54	0.51	0.72	0.55	0.51	0.50	

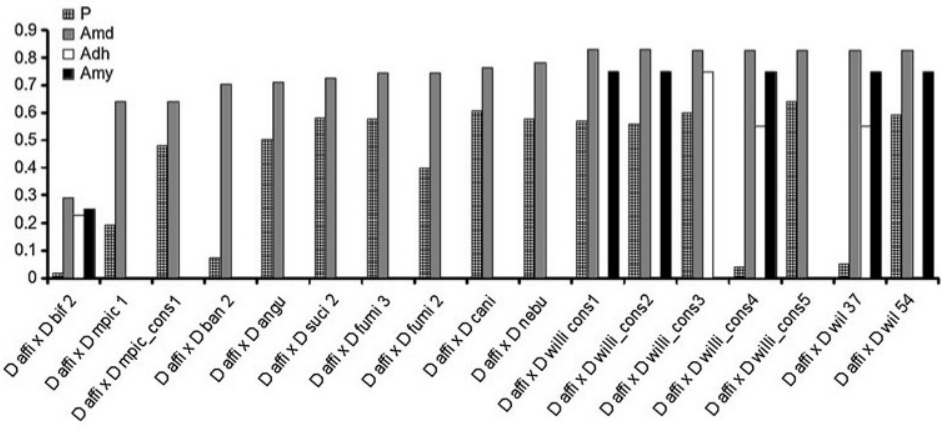

Fig. 2 Example of the results of pairwise comparisons between the *dS* values of *P*-like sequences and those of host genes (*Amd*; *Adh* and *Amy*). D affi = *Drosophila affinis*; D mpic = *D. mediopicta*; D ban = *D. banderantorum*; D. angu = *D. angustibucca*; D suci = *D. sucinea*; D fumi = *D. fumipennis*; D cani = *D. cardini*; D nebu = *D. nebulosa*; D will = *D. willistoni*. Cons = consensus of sequences with less than 3 % divergence

were used as a best conservative alternative. These genes show higher codon-usage bias than *Amd*. Of these comparisons, 238 exhibited *d*S values of *P*-like sequences that were significantly lower than those observed for at least one of the three nuclear host genes (of which 141 were obtained in comparisons with *Amd*). For example, the comparisons involving *D. affinis* yielded significantly lower *d*S values for the element (Fig. 2). From the 51 pairwise comparisons performed with the *D. affinis* sequence, 17 showed significantly lower *d*S values for the TE than for the alpha methyl dopa (*Amd*), alcohol dehydrogenase (*Adh*), and/or Amyrel (*Amy*) genes. Comparisons between the "D. affi" and "D.bif2" sequences and between the "D. affi" and "D.willi_cons4" sequences suggest a recent HT event, as suggested by low *d*S values. The estimated divergence times of these sequences are 0.59 and 1.21 million of years ago (MYA), respectively. Other comparisons in which the *d*S values of TEs are greater but yet significantly lower than those of host genes may suggest older HT events, as suggested by the divergence time estimates (12–20 MYA).

Supplementary spreadsheet 2 presents the 237 comparisons in which the *d*S values obtained for the *P* sequences were significantly lower than those obtained for the nuclear host genes. Divergence time estimates are also presented.

Discussion

The *Drosophila P*-element subfamilies

The inclusion of new *P-like* sequences obtained in this study did not change the clustering of the canonical, M, O, T, and K subfamilies previously described (Hagemann et al. 1992, 1994, 1996a, b; Setta et al. 2007; Nouaud et al. 2003), which remain monophyletic. However, using the criterion for subfamilies classification proposed by Capy et al. (1998), we suggest that the "M" and "O" subfamilies could be enlarged whereas the T subfamily could be subdivided, so that only the "K" subfamily would remain the same. Additionally, according to our data, six entirely new subfamilies could be created.

The subfamily "M", originally composed only of Old World Drosophilidae species (Hagemann et al. 1992), now appears as a cluster within group 1. The other clade in this group encompasses sequences that are closely related to the canonical *D. melanogaster* sequence but occur mostly in Neotropical *Drosophila* (excluding *D. melanogaster*, which received its sequence from *D. willistoni*; Daniels et al. 1990). These results illustrate that *P*-like elements of group 1 are widely spread in Drosophilidae, although these elements are until now more extensively represented among host flies of the New World.

Our results regarding the "O" subfamily are consistent with the previously described pattern (Hagemann et al. 1992, 1994). We did find, however, the new result that these sequences also occur in Neotropical species of the *Drosophila* subgenus. Conversely, the "T" subfamily is formed by two clusters that are sufficiently divergent to be split in two "subfamilies" according to the Capy et al. (1998) criterion: one consisting of host species of the *montium* subgroup of the *melanogaster* group (group 7) and the other encompassing five species of the *obscura* group (group 8). Our results indicate that the "K" subfamily proposed by Nouaud et al. (2003) is an independent branch and is sufficiently divergent to constitute an independent subfamily (group 11).

Six other branches/clusters were significantly divergent and may constitute new subfamilies (groups 2, 3, 4, 6, 9, and 10). Five of the six new subfamilies encountered here present distributions currently restricted to Neotropical species: groups 2 and 3 were found in a sole species of the *tripunctata* group (*D. angustibucca* and *D. mediopicta*, respectively); group 4 currently presents a wider taxonomic distribution, being found in species of two different *Drosophila* subgenera, encompassing the Neotropical *willistoni* (*Sophophora* subgenus), *tripunctata,* and *cardini* groups (*Drosophila* subgenus); group 9 is solely found in

D. willistoni; and group 10 is found in Neotropical *Sophophora* subgenus species included in the *willistoni* and *saltans* groups. The only non-Neotropical new subfamily presented here is the Nearctic *D. persimilis* group 6.

It is interesting to note that several species possess *P*-element sequences related to more than one subfamily. In this scenario, two cases deserve special distinction: (1) the case of *D. willistoni*, which presents sequences related to five of the eleven studied subfamilies, and (2) the case of *D. bocqueti*, which embraces sequences from the two most diversified subfamilies at both the nucleotide and the protein levels (groups 7 and 11).

Drosophila P-elements are monophyletic

Drosophila P-like sequences are monophyletic. The inclusion of other *P*-like sequences (such as those found in *Musca*, blowflies, mosquitoes, and vertebrates) in phylogenetic analyses reveals a topology clustered out of the *Drosophila P*-like elements clade (data not shown). Conversely, the phylogenies of *hATs* (Arensburger et al. 2011) and *mariner* elements (Rouault et al. 2009) reveal *Drosophila* TEs mixed with elements from other taxa, indicating a polyphyletic origin. Horizontal transfer is usually inferred from these phylogenetic inconsistencies, with *mariner* elements capable of "large jumps" between different phyla (Garcia-Fernàndez et al. 1995).

However, if *P*-elements can perform HT and phylogenetic incongruence is so common for these elements (relative to *Drosophila* species phylogeny), why are these elements of host *Drosophila* species monophyletic? Are they unable to do "large jumps"? The finding that the canonical *P*-element is mobilizable in other Drosophilidae (O'Brochta and Handler 1988) but not outside this family (Miller et al. 1987; Rio et al. 1988) suggests that there are specific host factors that are required for transposition. These host factors could limit HT between distantly related taxa.

As mentioned above, *d*S values of TEs that are significantly lower than those of host genes (*d*S TE < *d*S HG) can be indicative of HT (Silva and Kidwell 2000; Ludwig et al. 2008; Bartolomé et al. 2009; Lerat et al. 2011). The rational for this inference is that if *d*S values evolve neutrally, similar values should be expected for genes separated for the same amount of evolutionary time. There is presently no known evolutionary force that can systematically maintain lower *d*S values in TEs that in other genes. The most likely explanation for lower *d*S values in TEs is that TEs are evolving independently over a smaller evolutionary time period than other genes, possibly due to genomic acquisition by HT. However, not every case in which *d*S TE < *d*S HG is observed can be interpreted as an independent HT event. An HT event that occurred in the ancestor of two or more species, when evaluated by *d*S analysis, can be interpreted as an HT event in all pairwise analyses involving the descendant species, producing an overestimate of the true number of HT events. Moreover, even when recent, an HT event between species of two different groups can result in lower *d*S values for the TE in all comparisons involving the HT receptor species and any other member of the donor species group that inherited their TEs by vertical transmission.

Our analyses showed *P*-like elements in Drosophilidae are monophyletic and more variable than previously described. Furthermore, these data suggest that horizontal transfer involving *P*-element among Drosophilidae species is a common phenomenon.

Acknowledgments We are grateful to Dr. Luciano Basso da Silva and to Dr. Marco Silva Gottschalk for their assistance in *Drosophila* species identification; M.Sc. Bruna Corrêa for help with the R scripts and two anonymous reviewers for suggestions. This work was supported by CNPq, CAPES (PNPD), and Pronex-FAPERGS (10/0028-7).

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

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

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arensburger P, Hice RH, Zhou L, Smith RC, Tom AC, Wright JA, Knapp J, O'Brochta DA, Craig NL, Atkinson PW (2011) Phylogenetic and functional characterization of the *hAT* transposon superfamily. Genetics 188:45–57. doi:10.1534/genetics.111.126813

Bartolomé C, Bello X, Maside X (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genome Biol 10:R22. doi:10.1186/gb-2009-10-2-r22

Bingham PM, Kidwell MG, Rubin GM (1982) The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a *P* strain-specific transposon family. Cell 29:995–1004

Bregliano J, Kidwell MG (1983) Hybrid dysgenesis determinants. In: Shapiro J (ed) Mobile genetic elements. Academic Press, New York, pp 363–410

Brookfield JFK, Montgomery E, Langley CH (1984) Apparent absence of transposable elements related to the P elements of *Drosophila melanogaster* in other species of *Drosophila*. Nature 310:330–332

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

Carvalho MO, Silva JC, Loreto ELS (2004) Analyses of *P*-like transposable element sequences from the genome of *Anopheles gambiae*. Insect Mol Biol 13:55–63

Castro JP, Carareto CMA (2004) Characterization of two full-sized *P* elements from *Drosophila sturtevanti* and *Drosophila prosaltans*. Genet Mol Biol 27:373–377. doi:10.1590/S1415-47572004000300011

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

Clark JB, Maddison WP, Kidwell MG (1994) Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol Biol Evol 11:40–50

Clark JB, Altheide TK, Schlosser MJ, Kidwell MG (1995) Molecular evolution of *P* transposable elements in the genus *Drosophila*. I. The *saltans* and *willistoni* species groups. Mol Biol Evol 12:902–913

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

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

Garcia-Fernàndez J, Bayascas-Ramírez JR, Marfany G, Muñoz-Mármol AM, Casali A, Baguñà J, Saló E (1995) High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthe): evidence for a trans-phyla horizontal transfer. Mol Biol Evol 12:421–431

Graur D, Li W-H (2000) Fundamentals of molecular evolution, 2nd edn. Sinauer Associates, Sunderland

Grumbling G, Strelets V (2006) FlyBase: anatomical data, images and queries. Nucleic Acids Res 34:D484–D488. doi:10.1093/nar/gkj068

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52(5):696–704. doi:10.1080/10635150390235520

Hagemann S, Pinsker W (2001) *Drosophila P* transposons in the human genome? Mol Biol Evol 18:1979–1982

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

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

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

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

Hammer SE, Strehl S, Hagemann S (2005) Homologs of *Drosophila P* transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. Mol Biol Evol 22:833–844. doi:10.1093/molbev/msi068

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

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

Kimbacher S, Gerstl I, Velimirov B, Hagemann S (2009) Drosophila P transposons of the urochordata *Ciona intestinalis*. Mol Genet Genomics 282:165–172. doi:10.1007/s00438-009-0453-7

Lansman RA, Stacey SN, Grigliatti TA, Brock HW (1985) Sequences homologous to the P mobile element of *Drosophila melanogaster* are widely distributed in the subgenus Sophophora. Nature 318:561–563

Lee SH, Clark JB, Kidwell MG (1999) A *P* element-homologous sequence in the house fly, *Musca domestica*. Insect Mol Biol 8:491–500

Lerat E, Burlet N, Biémont C, Vieira C (2011) Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473(2):100–109. doi:10.1016/j.gene.2010.11.009

Ludwig A, Valente VLS, Loreto ELS (2008) Multiple invasions of Errantivirus in the genus *Drosophila*. Insect Mol Biol 17(2):113–124. doi:10.1111/j.1365-2583.2007.00787.x

Miller LH, Sakai RK, Romans P, Gwadz RW, Kantoff P, Coon HG (1987) Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. Science 237:779–781

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3(5):418–426

Nene V, Wortman JR, Lawson D, Haas B, Kodira C et al (2007) Genome sequence of *Aedes aegypt*, a major arbovirus vector. Science 316:1718–1723. doi:10.1126/science.1138878

Nerón B, Ménager H, Maufrais C, Joly N, Maupetit J, Letort S, Carrere S, Tuffery P, Letondal C (2009) Mobyle: a new full web bioinformatics framework. Bioinformatics 25(22):3005–3011. doi:10.1093/bioinformatics/btp493

Nouaud D, Quesneville H, Anxolabéhère D (2003) Recurrent exon shuffling between distant *P*-element families. Mol Biol Evol 20(2):190–199. doi:10.1093/molbev/msg036

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

O'Brochta DA, Handler AM (1988) Mobility of *P* elements in drosophilids and nondrosophilids. Proc Natl Acad Sci USA 85(16):6052–6056

Oliveira LFV, Wallau GL, Loreto ELS (2009) Isolation of high quality DNA: a protocol combining rennet and glass milk. Electron J Biotechnol 12:1–6. doi:10.2225/vol12-issue2-fulltext-4

Perkins HD, Howells AJ (1992) Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina*. Proc Natl Acad Sci USA 89:10753–10757

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14(9):817–818

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (Austria)

Rio DC, Barnes G, Laski FA, Rine J, Rubin GM (1988) Evidence for *Drosophila P* element transposase activity in mammalian cells and yeast. J Mol Biol 200(2):411–415

Rouault J, Casse N, Chénais B, Hua-Van A, Capy P (2009) Automatic classification within families of transposable elements: application to the *mariner* family. Gene 448:227–232. doi:10.1016/j.gene.2009.08.009

Rubin GM, Kidwell MG, Bingham PM (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29:987–994

Saitou N, Nei M (1987) The Neighbor-Joining method—a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sarkar A, Sengupta R, Krzywinski J, Wang X, Roth C, Collins FH (2003) *P* elements are found in the genomes of nematoceran insects of the genus *Anopheles*. Insect Biochem Mol Biol 33:381–387. doi:10.1016/S0965-1748(03)00004-3

Setta N, Loreto EL, Carareto CMA (2007) Is the evolutionary history of the O-type *P* element in the saltans and willistoni groups of *Drosophila* similar to that of the canonical *P* element? J Mol Evol 65:715–724. doi:10.1007/s00239-007-9051-7

Sharp PM, Li WH (1989) On the rate of DNA sequences evolution in *Drosophila*. J Mol Evol 28:398–402. doi:10.1007/BF02603075

Silva JC, Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* elements. Mol Biol Evol 17:1542–1557

Silva JC, Kidwell MG (2004) Evolution of *P* elements in natural populations of *Drosophila willistoni* and *D. sturtevanti*. Genetics 168:1323–1335. doi:10.1534/genetics.103.025775

Simonelig M, Anxolabéhère D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc Natl Acad Sci USA 88:6102–6106

Staden R (1996) The Staden sequence analyses package. Mol Biotechnol 5:233–241

Swofford DL (2003) PAUP: Phylogenetic Analysis Using Parsimony (and other methods), Version 4. Sinauer Associates, Massachusetts

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi: 10.1093/molbev/msr121

Vidal NM, Ludwig A, Loreto EL (2009) Evolution of *Tom*, *297*, *17.6* and *rover* retrotransposons in Drosophilidae species. Mol Genet Genomics 282:351–362

CAPÍTULO III

Trabalho publicado na revista *Molecular Genetics and Genomics*

An evaluation of the ecological relationship between *Drosophila* species and their parasitoid wasps as an opportunity for horizontal transposon transfer

Mauro Freitas Ortiz[1#], Gabriel Luz Wallau[2#], Daniel Ângelo Sganzela Graichen[3], Elgion Lucio Silva Loreto[124*]

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

[2] Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria, Brazil

[3] Departamento de Zootecnia e Ciências Biológicas, Universidade Federal de Santa Maria, Brazil

[4] Departamento de Biologia, Universidade Federal de Santa Maria, Brazil

* Corresponding author: E-mail: elgionl@gmail.com.

[#] These authors contributed equally to this work.

Mol Genet Genomics
DOI 10.1007/s00438-014-0900-y

ORIGINAL PAPER

An evaluation of the ecological relationship between *Drosophila* species and their parasitoid wasps as an opportunity for horizontal transposon transfer

Mauro Freitas Ortiz · Gabriel Luz Wallau · Daniel Ângelo Sganzela Graichen · Elgion Lucio Silva Loreto

Received: 1 May 2014 / Accepted: 12 August 2014

Abstract Evidences of horizontal transfer, the exchange of genetic material between reproductively isolated species, have accumulated over the last decades, including for multicellular eukaryotic organisms. However, the mechanisms and ecological relationships that promote such phenomenon is still poorly known. Host–parasite interaction is one type of relationship usually pointed in the literature that could potentially increase the probability of the horizontal transfer between species, because the species involved in such relationships are generally in close contact. Transposable elements, which are well-known genomic parasites, are DNA entities that tend to be involved in horizontal transfer due to their ability to mobilize between different genomic locations. Using *Drosophila* species and their parasitoid wasps as a host–parasite model, we evaluated the hypothesis that horizontal transposon transfers (HTTs) are more frequent in this set of species than in species that do not exhibit a close ecological and phylogenetic relationship. For this purpose, we sequenced two sets of species using a metagenomic and single-species genomic sampling approach through next-generation DNA sequencing. The first set was composed of five generalist *Drosophila* (*D. maculifrons*, *D. bandeirantorum*, *D. polymorpha*, *D. mercatorum* and *D. willistoni*) species and their associated parasitoid wasps, whereas the second set was composed of *D. incompta*, which is a flower specialist species, and its parasitoid wasp. We did not find strong evidence of HTT in the two sets of *Drosophila* and wasp parasites. However, at least five cases of HTT were observed between the generalist and specialist *Drosophila* species. Moreover, we detected an HT event involving a *Wolbachia* lineage between generalist and specialist species, indicating that these endosymbiotic bacteria could play a role as HTT vectors. In summary, our results do not support the hypothesis of prevalent HTT between species with a host–parasite relationship, at least for the studied wasp–*Drosophila* pairs. Moreover, it suggests that other mechanisms or parasites are involved in promoting HTT between Drosophila species as the *Wolbachia* endosymbiotic bacteria.

Keywords Transposon · Horizontal transfer · Parasitoid wasps · Wolbachia · High-throughput sequencing

Communicated by Q. Xia.

M. F. Ortiz and G. L. Wallau contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-014-0900-y) contains supplementary material, which is available to authorized users.

M. F. Ortiz · E. L. S. Loreto (✉)
Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
e-mail: elgionl@gmail.com

G. L. Wallau · E. L. S. Loreto
Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria, Santa Maria, Brazil

D. Â. S. Graichen
Departamento de Zootecnia e Ciências Biológicas, Universidade Federal de Santa Maria, Frederico Westfalen, Brazil

E. L. S. Loreto
Departamento de Biologia, Universidade Federal de Santa Maria, Santa Maria, Brazil

Introduction

The inheritance of genetic material in multicellular eukaryotes is primarily driven by vertical transmission from parents to offspring and, hence, from ancestor to descendant

Published online: 22 August 2014

species. However, increasing evidence obtained over the past few years suggests that the horizontal transfer (HT) of genetic material between reproductively isolated species is also an important phenomenon in the evolution of multicellular eukaryotic species (Schaack et al. 2010; Wallau et al. 2012). Based on current knowledge about the distribution of HT in many taxa, such events can essentially be divided into two categories: horizontal gene transfer (HGT) and horizontal transposon transfer (HTT). The basis of this distinction is most likely due to selfishness and the mobile nature of transposable elements (TEs), which are known genomic parasites and differ from non-mobile genes that are involved more directly in the fitness of the host organism. It is currently accepted that HTT is more frequent than previously thought. However, the mechanism by which TEs are transferred from one species to another remains a matter of debate, primarily because it is difficult to reproduce such events in the laboratory.

With the recent advent of new-generation sequencing technology, ecological approaches can be used to address questions about the types of species relationships that can promote HTT events and about the vectors that are involved. Several authors have suggested possible conditions for the occurrence of HTT, including the potential activity of the TE, the dependence, or lack thereof, of host factors, such as specific enzymes, or the existence of host TE silencing or repressing mechanisms (see review Silva et al. 2004). Another necessary condition for the occurrence of HTT is the spatiotemporal superposition of the involved species (Casse et al. 2006; Mota et al. 2009; Carareto 2011). Once these prerequisites are met, particular ecological relationships among species can facilitate the occurrence of HTT to various degrees. For example, species that present close ecological relationships such as parasitism, commensalism or mutualism likely display a higher probability of sharing TEs via HT compared with species that only exhibit temporal/spatial overlap (Yoshiyama et al. 2001; Silva et al. 2004; Laha et al. 2007). Another factor that may influence such rates is the presence of potential vectors. Viruses, bacteria and certain arthropods may serve as vectors carrying TEs between species (Gilbert et al. 2010, 2014; Dupuy et al. 2011; Routh et al. 2012).

The genus *Drosophila* has been used as one of the principal models for HTT studies. Among all of the HTT events described to date, more than half have occurred in *Drosophila* (Wallau et al. 2012). Similar to most multicellular eukaryotes, *Drosophila* species exhibit intricate host–parasite interactions with many types of parasites, which include viruses, bacteria, trypanosomes, nematodes, mites and other invertebrates (Carton et al. 1986; Perlman and Jaenike 2003; Klasson et al. 2009; Stefanov et al. 2012; Chandler and James 2013). A special category of host–parasite relationships involves parasitic wasp species that use *Drosophila* larvae as a site for oviposition and early development (Kacsoh and Schlenke 2012), ultimately killing the host. *Drosophila* species serve as the hosts of parasitoid wasps belonging to at least four families: Braconidae, Figitidae, Diapriidae and Pteromalidae (Carton et al. 1986). In the larvae of some *Drosophila* species, a strong immune response can be triggered in response to oviposition by the parasitoid. This response allows the larva to kill a number of the parasitoid eggs (Nappi and Vass 1993; Russo et al. 1996). To avoid this defense mechanism, some species of wasps inject venom (Carton et al. 1986) that suppresses the immunity of the host and allows the wasp eggs to evade the host's immune response (Carton and Nappi 1997). Many viral-like particles (VLPs) are generally injected into *Drosophila* larvae as a component of the venom of the parasitoid wasps (Rizki and Rizki 1990).

The injection of VLPs that may carry double-stranded DNA and TE transcripts suggests that they can serve as a potential HTT vector (Routh et al. 2012). Several TE fragments (*mariner*-like, *gypsy*-like, *Maverick*-like and *DIRS*-like sequences; Dupuy et al. 2011) have been found integrated in the genomes of polydnaviruses (PDVs), which are the viruses injected as VLPs by certain parasitoid wasps of butterflies (Herniou et al. 2013). Although VLPs injected into butterflies are not homologous to the VLPs injected by the parasitoid wasps of *Drosophila*, these findings suggest that VLPs can promote the movement of TEs between *Drosophila* and its parasitoid wasps, acting as a true HTT vector. In addition to VLPs, the exchange of other viruses, endosymbiont bacteria or even direct cell exchange, due to the very close contact between *Drosophila* species and their parasitoid wasps, make these parasites good candidates for HTT vectors.

Using an ecological approach involving two sets of host–parasitoid pairs (*Drosophila*/wasp species), we attempted to address several open questions about HTTs. We investigated the relative occurrence of these events between (1) *Drosophila* species and their parasitoid wasps; (2) *Drosophila* species that are ecologically separated and, thus, only share temporal and spatial overlap; and (3) wasps that parasitize ecologically separated *Drosophila*. For this purpose, we sampled two sets of Neotropical *Drosophila* species. The first set was composed of a metagenomic sample including five generalist *Drosophila* species that oviposit on decaying bananas (one belonging to the *Sophophora* subgenus and the remainder coming from the *Drosophila* subgenus) and their parasitoid wasp *(Leptopilina boulardi*). The second consisted of a specialist *Drosophila* species, *D. incompta*, and its parasitoid wasp. This *Drosophila* species belongs to the flavopilosa species group, which is a taxon that has evolved to use the flowers of genus *Cestrum* (Solanaceae) as unique sites for oviposition and larval

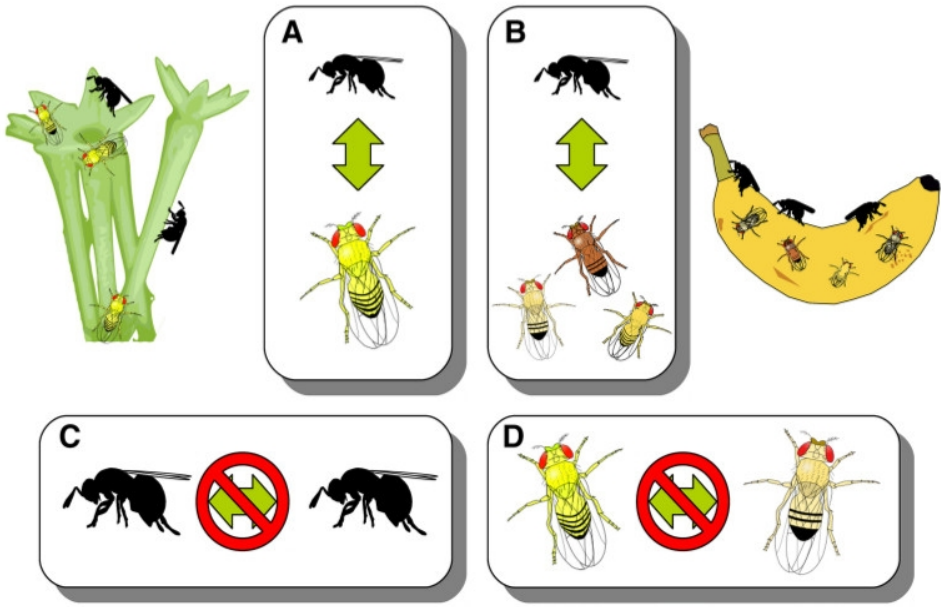

Fig. 1 The *squares* represent our working hypothesis: **a** the genome of the flower-breeding species *Drosophila incompta* (DiG) was expected to share TEs with the genome of its parasitoid wasp (DiWG) via HTT, as were **b** the fruit-breeding *Drosophila* species (DsP) with their parasitoid wasps (DsWP). The *marks* over the *arrows* in "**c**" and "**d**" represent comparisons between DsP–DiG and DsWP–DiWG, which were not expected to show significant evidence of HTT because these species are ecologically divergent (color figure online)

development (Wheeler and Takada 1962; Brncic 1983; Robe et al. 2013). Flies of the *flavopilosa* group have never been collected using resources other than *Cestrum* flowers, and only other drosophilidae *(Zygothrica vittimaculosa*) have been collected while using *Cestrum* flowers as a developmental site (Brncic 1983; Robe et al. 2013). For these reasons, *D. incompta* and the generalist *Drosophila* that we employed in this study have been ecologically separated for a considerable evolutionary time.

Our working hypotheses were that if the wasp parasitoids of *Drosophila* species are strongly involved in the transfer of TEs between the wasp and *Drosophila* genomes, TEs with high similarity should be shared by *Drosophila* and their specific wasp parasites. On other hand, if a close ecological relationship is an important factor for HTT, flies and parasitoid wasps that are ecologically separated should share fewer highly similar TEs (Fig. 1).

Materials and methods

Samples

We sampled generalist *Drosophila* species (known to oviposit in many different types of decaying fruit substrates; Markow and O'Grady 2005) and parasitoid wasps that emerge in the same sources (Fig. 1). To collect fruit flies, we placed ripe bananas in the wild for 4 days to allow oviposition. The samples were maintained in the laboratory until complete emergence of the flies and their parasitoid wasps. In the second experiment, we sampled a specialist *Drosophila* species with a restricted ecology (*D. incompta*) and its parasitoid wasp. To collect this species, *Cestrum* flowers were collected and maintained in the laboratory until the complete emergence of flies and wasps. Both sets of species were collected at the same site at Santa Maria city, south of Brazil, at latitude 34.95303 and longitude −120.43572. In summary, four samples were collected: (1) the generalist *Drosophila* species pool sample (DsP); (2) the generalist *Drosophila* species parasitoid wasps (DsWP); (3) the *D. incompta* sample (DiG); and (4) *D. incompta* wasps (DiWG). As there are complete available genomes for *D. melanogaster* and *D. simulans,* we removed all individuals of these species that emerged in our collections from our DNA sample pool. This procedure was applied to increase the genome coverage for the other sampled species and, hence, the probability of detecting HTT. The available genomes of *D. melanogaster* and *D. simulans* were also used in our analyses.

DNA sequencing and assembly

Total DNA was isolated from 20 individuals from each fly and wasp species using NucleoSpin Tissue XS (MACHEREY-NAGEL) following the manufacturer's protocol. For the *D. incompta* parasitoid wasps, only three specimens were obtained, and their genomic DNA was amplified using a whole-genome amplification protocol with the GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences), as there was not sufficient DNA necessary for Illumina sequencing. Genomic DNA was sequenced by the Fasteris DNA Sequencing Service in a Solexa-Illumina HiSeq 2000 New Generation Sequencing (NGS) device according to the manufacturer's instructions. A single-end

Table 1 Details of the draft genome assemblies and their TE contents

Assembly (code)	Assembled (Gb)	N50 (bp)	Contig length (bp)			Number of contigs	Number of TEs >80 % identity and >100 score	Content of TEs in Kbs	Average size of TEs (bp)
			Min.	Mean	Max.				
DsP (75)	1.55	185	64	180	15,928	863,894	5,914	879	148.66
DsWP (76)	3.0	305	64	216	18,248	1,394,480	381	39	105.66
DiG (77)	1.95	181	64	169	5,044	1,154,472	3,970	529	133.60
DiWG (78)	1.77	304	64	226	11,564	781,780	476	52	114.42

approach with a read size of ~100 bp was employed. Sequence assembly was performed with AbySS (MPI) software (Simpson et al. 2009) with the following parameters: $K = 64$, adjlist = 30, popbubbles = 0.9, distanceest $s = 200$ and $n = 10$, generating four genomic datasets. A range of kmer (31 until 64) were tested and the kmer 64 presented the best results in terms of the assembly metrics shown in Table 1; then we used this parameter to generate our final draft for each sample.

Species identification

Two methods were used to identify *Drosophila* species: a priori and a posteriori identification. However, for the wasps, only a posteriori identification was conducted, as we could not find experts on Neotropical *Drosophila* parasitoid wasps. Nevertheless, because we used a metagenomic approach, we were able to identify the potential species in our datasets using barcodes. For the a priori identification of *Drosophila*, we followed Markow and O'Grady (2005). For the a posteriori identification of all four datasets, we retrieved *Drosophila* (3,669 sequences) and Hymenoptera (4,007 sequences) cytochrome oxidase I (COI) and cytochrome oxidase II (COII) mitochondrial genes from the BOLD databank (Ratnasingham and Hebert 2007) and GenBank (accessed August 2012). We employed nuclear gene sequences (obtained from Robe et al. 2005, 2010a, b) to confirm the *Drosophila* identifications using mitochondrial genes.

To establish a threshold for identity and subject size from the BLAST output used in the a posteriori species identification, we applied a BLASTn procedure with a Lepidoptera gene dataset obtained from GenBank (2,422 sequences). We used this approach because it can be assumed that no Lepidoptera genes are included in our dataset. Accordingly, the results of this analysis would represent the greatest identity and subject size possible by chance, from which we established a threshold for biologically irrelevant data. The application of the Lepidoptera dataset together with all of our datasets yielded maximum identity values ranging from 95.04 to 97.47 for fragments that varied between 101 and 183 bp. Accordingly, we employed a minimum identity value of 95 and a subject size of >200 bp for species identification using mitochondrial genes. We then compared our a priori results with our a posteriori results to filter out the correct *Drosophila* species.

Additionally, we performed BLASTn searches using 409 Acari COI genes to exclude any possible contamination from mites and 8,135 *Wolbachia* sequences to analyze the infection status of our datasets. Finally, we performed BLASTn searches using 2,674 virus sequences from RefSeq and 498 polydnavirus (PDV) sequences to investigate the presence of polydnaviruses in the studied species.

Phylogenetic analysis

A phylogenetic analysis based on COI sequences was performed to evaluate the relationships of the parasitoid wasps sampled in this study. For the wasp phylogeny, we used all of the COI sequences obtained from the samples and added all recognized *Drosophila* parasitoid wasp sequences from four families acquired from www.biology.emory.edu/research/schlenke/resources.html and several very similar sequences acquired from BLAST-NCBI using our COI sequences as a query. As an outgroup, we employed *Apis mellifera* COI sequences (accession number JQ350734.1). For tree reconstruction, we used Mr. Bayes 3.2 software (Ronquist et al. 2012), with nucleotide sequences and a GTR + G nucleotide evolutionary model selected by jModelTest (Darriba et al. 2012). A total of 1,000,000 generations were evaluated. Sampling was performed every 100 generations, and 25 % of the trees were used for burn-in. For the *Drosophila* phylogeny, we employed all four genes available for the studied species: the nuclear genes *α-methyldopa* (*Amd*) and *dopa decarboxylase* (*Ddc*) and the mitochondrial genes *cytochrome oxidase I and II* (*COI* and *COII*). We also used various species from the *Drosophila* genus and *Scaptodrosophila latifasciaeformis* to validate the topology of our tree. For the accession numbers of all of the sequences used in the phylogenetic analysis, see Supplementary spreadsheet 2d.

To evaluate the relationships of the *Wolbachia* genomes, we performed phylogenetic analysis using PhyML (Guindon et al. 2010) based on the genome alignment obtained

with LAGAN software (Brudno et al. 2003) available at the mVISTA web server (Frazer et al. 2004). Branch support was determined via the aLRT method described by Anisimova and Gascuel (2006). We also used the draft *Wolbachia* genomes from our samples as queries against the *Wolbachia* MLST database to validate our phylogenetic analysis based on the most closely related lineage (Baldo et al. 2006).

Description of TEs

To identify the TEs in the genomic datasets, we performed screening through BLASTn (Altschul et al. 1997) homology searches using the TE database from Repbase 16.03 (Jurka et al. 2005) as a query. The default parameters were employed in the BLAST searches, with the exception of applying a value of 0.01 for the *e*-value parameter. Due to the short contigs in our datasets, we retrieved all of the contigs for which BLAST hits with TEs were observed (identity >80 % or score >100), see Supplementary spreadsheet 1a–d. We then obtained the TE content of each sample (Table 1). We also used RepeatMasker 4.0.3 (Smit and Hubley 1996–2010) with the commands *no_is* and *nolow* to estimate the TE content of each sample. In addition, we used an ab initio method implemented in RepeatScout v1.0.5 (Price et al. 2005). However, it did not improve our TE description once the transposable elements of insect are well known and due also to the fragmentary nature of our assembly. At the end of this step we obtained four datasets of TE content to each genomic dataset (DsP_TEs, DsWP_TEs, DiG_TEs, DiWG_TEs).

HTT searches and TE reconstruction

To perform HTT detection, we applied BLAST to the TE content from each sample against all of the contigs from all other samples. First, we used BLAST on the TE dataset from generalist *Drosophila* species (DsP_TEs) against its parasitoid wasp contigs (DsWP). We then applied reciprocal BLAST to the TE dataset from the parasitoid wasps (DsWP_TEs) against the DsP contigs. A second round of BLAST was performed with the TE dataset from *D. incompta* (DiG_TEs) against its parasitoid wasp contigs and also in an inverse sense, using the TE dataset from the parasitoid wasps (DiWG_TEs) against the DiG contigs. Moreover, we determined whether we could detect any HTT events occurring between the sequenced *Drosophila* species and between the wasp species. We then blasted the DsP_TEs against the DiG contigs and the DiG_TEs against the DsP contigs. The same procedure was performed using the DsWP_TEs against the DiWP wasp contigs and the DiWP_TEs against the DsWP wasp contigs.

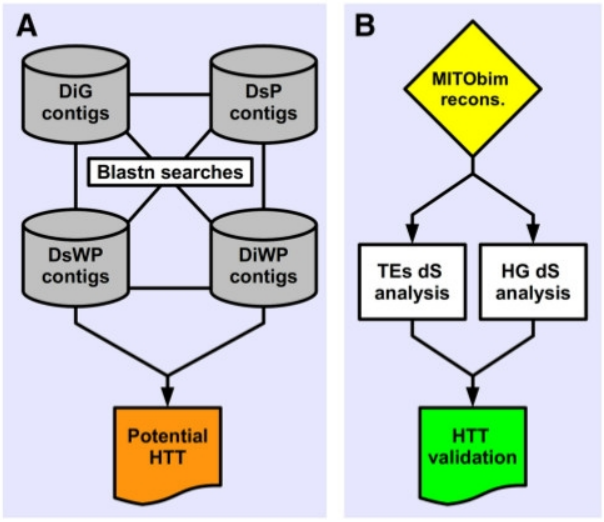

Fig. 2 Experimental design used to detect HTT events between our samples. We performed all-against-all BLAST searches using our genomic and metagenomic datasets. **a** We performed the four datasets to BLAST searches against each other. **b** Then, we reconstructed these TEs (MITObin) to perform dS analysis in comparison with host genes to validate the HTT

Additionally, we subjected each draft genome or metagenomic sample to BLASTn searches against each other sample, as described in Fig. 2. We also included the *D. melanogaster* and *D. simulans* genomes available at the FlyBase website in these analyses, as the larvae of these two species are well-known hosts for *L. boulardi* wasps and these *Drosophila* species also emerged in our collection. We did this last step to prevent potential missing elements from previously described TEs contents.

We considered a potential HTT case to have been retrieved if the TEs from different samples showed more than 90 % identity in a fragment of at least 100 bp. This is a conservative threshold, considering that the Diptera and Hymenoptera orders diverged approximately 366 Mya (Hedges et al. 2006), and it could underestimate older HTT events. However, if this host–parasite relationship promotes frequent HTTs, we should still observe a significant HTT signal.

After we retrieved all of the contigs above our threshold in the BLAST analysis, we attempted to reconstruct the full element of each TE from reads with MITObim v 1.5 (Hahn et al. 2013), using the best BLAST hit as a seed in each of our datasets separately (Fig. 2f). We were able to reconstruct sequences of at least 800 bp from the coding region of each of the analyzed TEs. These reconstruction procedures were applied because our fragmented TE dataset could preclude making dS estimates. Thus, the larger fragments obtained from the reconstruction of the coding regions of these TEs allowed us to perform statistical

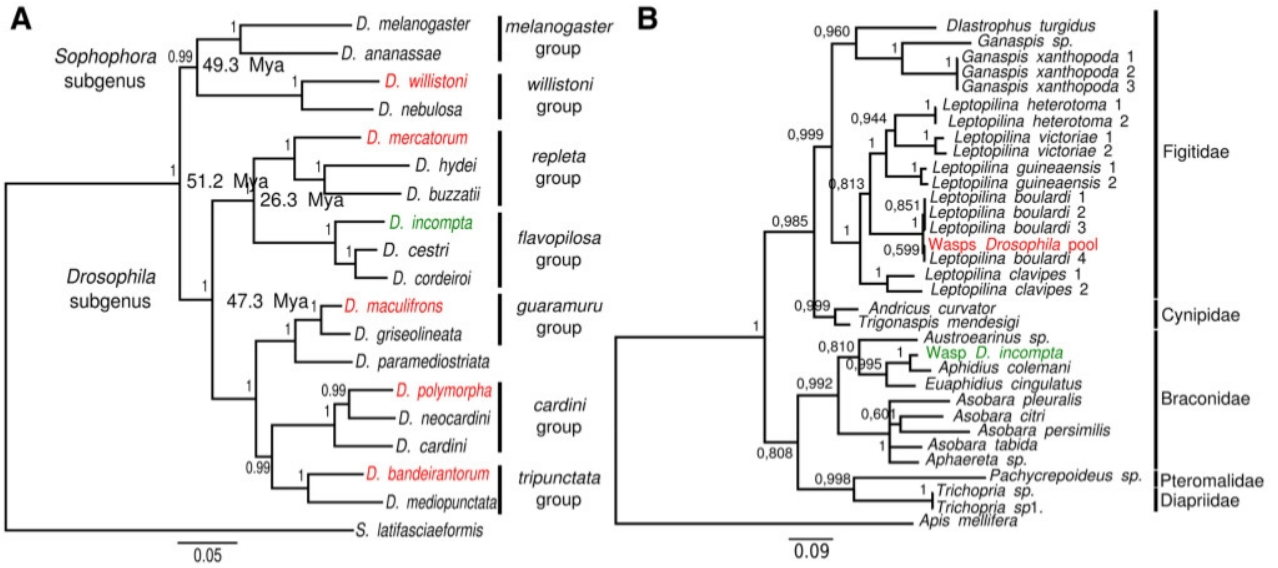

Fig. 3 **a** Phylogenetic reconstruction, using two nuclear and two mitochondrial genes, of the species from the pool of *Drosophila* species (DsP species—in *red*), *D. incompta* (in *green*) and two other species, *D. ananassae* and *D. melanogaster* (in *black*). **b** Phylogenetic relationships obtained using COI sequences for all known parasitoid wasp species and other wasp species that showed a high similarity with the COI sequences obtained from our datasets (in *black*), including both the wasp species from the wasp pool (in *red*) and the wasp species that emerged from the *D. incompta* larvae (in *green*). The *black box* on the right represents the wasp families. The posterior probabilities obtained from the Bayesian analysis appear on the left side of each node in both phylogenies. The average time estimates in the *Drosophila* phylogeny appear on the right side of each node of interest and are based on the TimeTree website (Hedges et al. 2006) and on Robe et al. (2010a) (color figure online)

analysis on the dataset. Reconstructed TEs could be formed from the species pool with reads from several related elements, which may render the elements more divergent than they really are and, hence, underestimate HTT events. Although HTT detection could be weakened by this bias, HTT events could still be readily detected based on low divergence estimates between TEs because *D. incompta* and the most closely related species sampled in this study (*D. mercatorum*) diverged at an approximate time of at least 26.3 Mya (Robe et al. 2010b).

After we obtained the reconstructed TE, the coding region was employed to estimate the *p*-distance and synonymous substitutions rates (dS) using MEGA 5.0 (Tamura et al. 2011). The dS values were estimated from the TEs and from four genes available for these *Drosophila* species, which included two mitochondrial (cytochrome oxidase I and II) and two nuclear genes [alpha methyldopa (Amd) and dopa decarboxylase (Ddc)]. We only used these four genes to test potential HTT events in DsP and DiG, because we were not able to reconstruct a large number of single-copy genes from our metagenomic dataset. Comparison of the synonymous substitutions rates (dS) of host genes with the dS of TEs is a robust method for the inference of HTT (Silva and Kidwell 2000; Loreto et al. 2008; Wallau et al. 2012).

To better identify the species from our *Drosophila* pool involved in cases of HTT, we performed additional PCR searches with primers designed based on three reconstructed LTR retrotransposon sequences (*Copia, MINOS* and *Tabor*) that showed a strong HTT signal in the dS analysis. The expected amplicon sizes and PCR conditions for these primers are presented in Supplementary spreadsheet 2a.

Results

Species identification

The dataset corresponding to the *Drosophila* species pool (DsP) contained five Neotropical species: four species of the *Drosophila* subgenus (*D. maculifrons, D. bandeirantorum, D. polymorpha* and *D. mercatorum*) and one species of the *Sophophora* subgenus (*D. willistoni*), all of which were in agreement with our a priori identifications (Fig. 3a; Supplementary spreadsheet 2b). As can be seen in Fig. 3a, the phylogenetic distribution of the studied species is spread throughout the *Drosophila* tree, showing that they are phylogenetically distantly related species. The parasitoid wasps that emerged from these generalist *Drosophila* species (DsWP) all came from one species, *Leptopilina boulardi,* which belongs to the Figitidae family (Fig. 3b; Supplementary spreadsheet 2b).

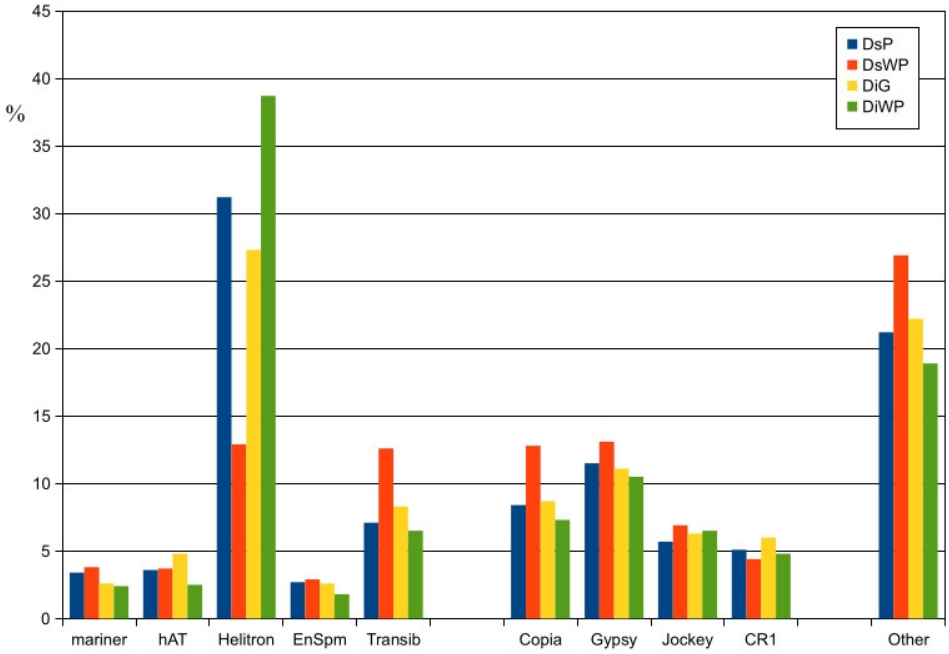

Fig. 4 Abundance and diversity of TEs in the DsP, DsWP, DiG and DiWP datasets. The different TE superfamilies are represented as the % of TE content found in each dataset

The dataset corresponding to the specialist *Drosophila* species *D. incompta* (DiG) confirmed the identification of this species. We also detected a wasp belonging to the family Braconidae in the DiWP dataset. This wasp was grouped with the wasp *Aphidius colemani* (Fig. 3b; Supplementary spreadsheet 2b).

Assembly and TE characterization

A total of 8.3 Gb of assembled sequences were obtained. The four datasets, DsP, DsWP, DiG and DiWG, showed variation in N50 values ranging from 181 bp for DsP to 305 bp for DsWP in the final draft assembly. Additional details on the reads and the assembly dataset are provided in Table 1. When the four datasets were subjected to BLAST searches against the Repbase dataset to describe the TE contents, it was found to be quite different between the *Drosophila* and wasp datasets. In *Drosophila*, we identified 5,914 TEs hits in DsP and 3,970 in DiG. In the wasps, only 381 TEs hits were found in DsWP and 476 in DiWP (Table 1). In general, approximately half or more of these TEs were Class II transposons, while the proportion of Class I elements ranged from 28.4 to 51.8 % (Supplementary Figure 1). The representations of different superfamilies in the analyzed datasets were similar (Fig. 4). The most representative superfamily was *Helitron*, showing proportions ranging from 38.7 % in DiWP to 26.9 % in DsWP. The second most abundant Class II superfamily in our dataset was *Transib*, which contributed 6.5 to 12.6 % of the identified TEs. The *mariner* and *hAT* superfamilies accounted for 3 or 4 % of the TEs. The *Gypsy* and *Copia* retrotransposons showed proportions of approximately 8–12 % in different datasets. The superfamilies *CR1* and *Jockey* contributed 5–7 % of the identified TEs (Fig. 4).

HTT searches and validation

Contradictory to our working hypothesis, we identified only one potential HTT event between the *Drosophila* pool (DsP) and the parasitoid wasps emerging from the same sources (DsWP). This putative event involved a *Helitron* element related to *Helitron-2_DVir* from Repbase. The identified elements showed a divergence of only 8.3 % between sequences from DsP and DsWP. No HTT events were detected between the flower specialist *Drosophila* species and its parasitoid wasp (DiG and DiWG). Moreover, when we analyzed the BLAST results between *Drosophila* species (DsP and DiG), we found five potential HTT events. These events primarily involved LTR retrotransposons, with four of the five TEs coming from this group. These LTR retrotransposons included a *Copia*-like element, related to the *Copia-1_Dper-I* consensus sequence available in Repbase (Jurka et al. 2005), which presented a homologous region of 3,514 bp between the DiG and DsP datasets, showing a *p*-distance of only 0.031; two *Gypsy*-like sequences, related to *Gypsy-6B* and the *Gypsy-31_DWill* consensus sequence, which presented two homologous regions (of 2,165 and of 898 bp), and these regions exhibit *p*-distances of only 0.039 and 0.124, respectively; and a *Tabor*-like sequence, related to the *Tabor_DA-I* consensus sequence, showing a 2,625 bp homologous region and a *p*-distance of 0.065 between the DiG and DsP

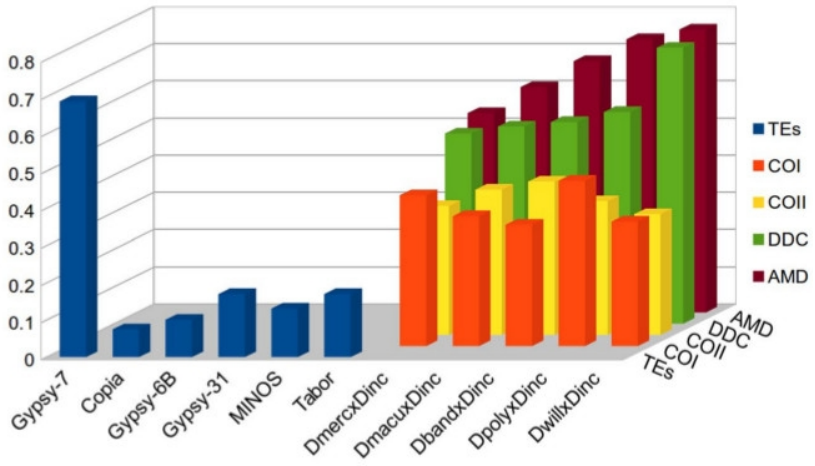

Fig. 5 dS comparisons of TEs and host genes from *D. incompta* and all of the species from the *Drosophila* pool. The first *bar* represents a Gypsy-7 element with a dS that did not differ significantly from those of the host genes (vertical transfer). Five TEs showed a significantly lower dS than the host genes (horizontal transfer) based on the comparison of DsP and DiG (*blue bars*) and the dS of the host genes (*orange* COI, *yellow* COII, *green* Ddc and *red* Amd) (color figure online)

datasets. Only one DNA transposon was found, the *MINOS* element. This transposon presented a homologous region of 1,151 bp and a *p*-distance of 0.066 (Supplementary spreadsheet 2c). These five sets of TEs shared by the *Drosophila* samples also showed significantly lower dS values compared with the host genes of all species from the DsP dataset, providing additional support that these very similar TEs in these species were shared via HTT (Fig. 5; Supplementary spreadsheet 3a–b). The reconstruction of the elements did not appear to affect the dS estimates, as all potential TEs showing an HTT signal in the BLAST analysis presented a significantly lower dS than the vertically transmitted genes when evaluated through dS analysis. See Supplementary Spreadsheet 4a–f for the genome-against-genome BLAST output. Searches for potential HTT events between wasps (DsWP and DiWG) did not reveal any significant cases.

PCR screening was performed on the genomic DNA of each *Drosophila* species present in our pool of generalist *Drosophila*. The *Copia* and *Tabor* elements were amplified in virtually all species present in the DsP dataset, with the exception of the *Copia* element in *D. maculifrons*. The *mariner* element was amplified in *D. mercatorum* and *D. polymorpha*. The *MINOS* element was amplified in *D. willistoni* and *D. polymorpha*. All of these elements were amplified in *D. incompta* (Supplementary spreadsheet 2a).

Using the Repbase TE dataset to characterize the TE content of each sample (Supplementary Figure 1) also allowed us to evaluate potential HTT events involving the previously described TEs. Based on the output of the BLAST analysis, we detected a high level of similarity among certain TEs from the Repbase dataset and our samples. We then carefully analyzed these putative HTT events. We found only one puzzling case, one *mariner*-like element with a length of 2,608 bp in the *D. incompta* genome that showed a similarity of 98.6 % to the planarian *Schmidtea mediterranea mariner* element (*SMAR25*). Using primers to this element, we got PCR amplicons from *D. incompta* and two species from our DsP dataset, *D. mercatorum* and *D. polymorpha* (Supplementary spreadsheet 2a). However, no *mariner* HTT signal was detected between the DiG and DsP datasets, suggesting that the amplification signal came from more divergent elements.

Wolbachia and PDV sequences

Wolbachia has also been suggested as a putative vector for HTT, including as a possible intermediate in a wasp–*Wolbachia*–*Drosophila* transfer. For this reason, we decided to check whether our studied species were infected with *Wolbachia* strains. To this end, we subjected an extensive *Wolbachia* dataset composed of genomes and partial sequences obtained from NCBI to BLAST searches against our datasets. From this analysis, we found that both our DsP and DiG datasets were infected with a lineage that is very closely related to *Wolbachia* wMel, which was originally identified from a *D. melanogaster* infection (Fig. 6). Based on the high identity observed in the BLASTn results, we used the entire genome of the lineage from *D. melanogaster* to comparatively reconstruct the draft genome of the *Wolbachia* lineage infecting *D. incompta* and the five generalist *Drosophila* species. Thus, we were able to assemble a draft *Wolbachia* genome, present in the *D. incompta* sample, with a size of approximately 1.1 Mb. Additionally, a *Wolbachia* draft genome of approximately 208 Kbs was assembled from the DsP dataset. These two genomes presented 95 and 85 % similarity, respectively, to the wRI strain from *D. simulans* and 96.6 and 95.3 % similarity to the wMel strain from *D. melanogaster*.

Although PDVs are not homologous to the VLPs injected in the venom of the parasitoid wasps of *Drosophila*, we decided to subject the sequences from our samples to BLAST searches against the PDV sequences available in NCBI to evaluate whether there were any PDV-related

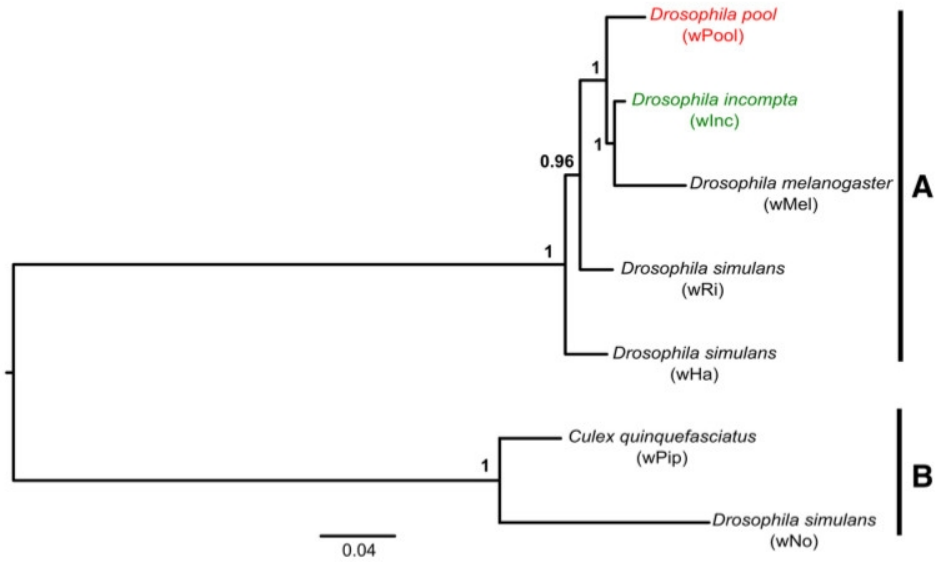

Fig. 6 Phylogenomic analysis of five *Wolbachia* genomes obtained from NCBI (*black* names) and the two strains found in our DsP (wPool) and DiG (wInc) datasets, with the names shown in *red* and *green*, respectively. On the left side of each node, the aLTR branch support values are provided, and the *black bars* on the right represent the two main groups of *Wolbachia* that infect arthropods (color figure online)

sequences in our datasets. Indeed, in both the DsP and DiG samples, we identified sequences of 200–300 bp that shared more than 97 % identity with a helicase from the PDV genome of *Cotesia vestalis,* a parasitoid wasp of the diamondback moth (Supplementary spreadsheet 5a–d). In contrast, in the wasp samples, we found fragments with an identity above 80 % with sizes that were never greater than 159 bp in the DsWP dataset and fragments reaching 253 bp showing an identity of 81 % with the *Cotesia vestalis* PDV genome (Supplementary spreadsheet 5a–d).

Discussion

Horizontal transposon transfer is a highly intriguing phenomenon. HTT has likely occurred frequently during the evolution of life. However, little is known about the mechanisms and features of TEs, hosts and their interactions that can influence such events. Discussions about these topics have generally been primarily hypothetical, involving multiple hypotheses, without any evaluation of these possibilities through experiments designed to test them. It has been suggested that close ecological relationships can increase the probability of HTT events, such as the HTTs observed among mammal species mediated by the blood-feeding insect *Rhodnius prolixus* (Gilbert et al. 2010). *Drosophila* species offer excellent opportunities for testing several of these hypotheses, as the biology of these species is well understood, and more than half of all HTT events reported to date involve *Drosophila* species (Wallau et al. 2012). One hypothesis that is occasionally invoked to explain these cases is the existence of *Drosophila* parasitoid wasps that exhibit close contact with *Drosophila* larvae of various species (Silva et al. 2004; Loreto et al. 2008; Schaack et al. 2010). Under this hypothesis, we can visualize several possibilities: (1) transfer between the *Drosophila* species and the parasitoid wasps; (2) transfer from the parasitoid wasps to another *Drosophila* species; (3) transfer among wasp species if two or more different wasp species parasitize the same *Drosophila* species; and (4) transfer among *Drosophila* species if two or more species share the same wasp parasites. However, in our experiment, we did not expect to observe HTT among *D. incompta* and the other *Drosophila* species or between their parasitoid wasps because they are ecologically separated, assuming that ecological proximity is an important factor for HTT. In an attempt to evaluate this host–parasite HTT route, we sequenced two sets of *Drosophila* species with their associated parasitoid wasps using both a genomic approach (1 *Drosophila* species and 1 wasp species) and a metagenomic approach (5 *Drosophila* species and 1 wasp species).

In our analyses, we identified a significant number of TEs in the *Drosophila* genomes, but few in the wasps. In contrast, the three available genomes of parasitoid wasps, from the *Nasonia* genus, exhibit a high TE content (The Nasonia Genome working group 2010). However, the TEs described in *Nasonia* are deposited in the Repbase dataset and were used in our searches. The TEs found in the *Drosophila* parasitoid wasps in our samples were expected to be divergent from those of *Nasonia* found via BLAST searches. Another limitation observed regarding the characterization of TEs in our dataset was the predominance of Class II elements, in contrast with previous findings in *Drosophila*. In the previously characterized genomes, Class I TEs are more abundant than Class II (Clark et al. 2007). These limitations may be related to the small sizes of the contigs obtained in our assemblage. To bypass these limitations, two approaches were used: (1) an ab initio search

was conducted in our dataset using RepeatScout, though this tentative approach did not allow us to obtain new TEs compared with our BLAST searches; and (2) the complete assemblies of each dataset were subjected to BLAST searches against each other, and all sequences showing similarity above our cutoff were carefully analyzed. We presumed that the second procedure would provide good support for finding very similar sequences, allowing us to search for characteristics of known TEs and, thus, to find new TEs shared among the analyzed datasets. Additionally, the small size of the contigs can be at least partially bypassed by this approach, keeping in mind that our purpose was to find HTTs, rather than a full characterization of TEs.

The experimental data obtained in this study do not provide support for our initial hypothesis, indicating *Drosophila* parasitoid wasps as an important HTT vector. Sequences with high similarity were not found to be shared between the examined *Drosophila* species and their parasitoid wasps. The unique potential HTT identified between our pool of *Drosophila* species and their parasitoid wasp, involving a *Helitron* element, showed a level of similarity that suggested that it was not a recent event.

Our experimental design did not permit us to make inferences about the possibility of wasps serving as a vector among generalist *Drosophila* species due the fact that we used a metagenomic approach. This strategy was employed to increase our chances of detecting HTT between the flies and wasps. As this is a unique dataset, it was not possible to determine which species share similar transposons in our "generalist" sample. This limitation could, in principle, be overcome by cloning and sequencing PCR amplicons obtained from DNA samples from each species present in our metagenomic sample. However, we did not implement these analyses because they would not answer the question of whether the wasps serve as a vector among these flies. At present, although we know that *Leptopilina boulardi* parasitizes some of these species, it is not known exactly which flies are parasitized. Hence, we do not know which Neotropical *Drosophila* species share the same parasite. This information will be imperative to design an experiment to test whether wasps serve as a vector among *Drosophila* species.

Contradictory to our primary hypotheses that recent HTTs should not be observed among ecologically separated *Drosophila* species (e.g., the *Drosophila* pool and *D. incompta*), five TEs showing strong evidence of HTT were found. Four of these TEs corresponded to LTR retrotransposons, two of which came from the *Gypsy* family. Some elements in this family are known to encode a functional *env* gene and, hence, to be able to carry out infections similar to viruses, suggesting that they can promote their own transmission (Mejlumian et al. 2002) and therefore do not require a vector to promote their horizontal transfer.

Some TEs require host factors for transposition. For example, *P* elements use the host's IRBP enzyme (Beall and Rio 1997). For this reason, it has been suggested that the canonical *P* element and related TEs are restricted to *Drosophilidae* and do not undergo transposition into distant species (Silva et al. 2004). The dependence on host factors could at least partly explain the greater number of HTTs found among *Drosophila* species, even those that are ecologically separated, than between *Drosophila* and their parasitoids.

To obtain further insights about HTTs involving species from our metagenomic sample, we tested three elements via PCR: *Copia, MINOS* and *Tabor*. These elements were amplified in almost all of the examined species, suggesting that multiple HTT events have taken place among these distantly related species. *D. incompta* and the other *Drosophila* species do not use the same resources in nature. Nevertheless, they have overlaps of their home range (Markow and O'Grady 2005; Bächli 2009), which may increase the chance of HTT among these species, likely due to other, still-unrecognized vectors.

Wolbachia lineages are another type of *Drosophila* parasite that has been hypothesized to promote HT between reproductively separated species (Silva et al. 2004; Hotopp et al. 2007; Loreto et al. 2008; Schaack et al. 2010; Dupeyron et al. 2014). Some instances of HT of *Wolbachia* genes (and genomes) to insect genomes and of HT of some genes in the opposite direction have been described (Hotopp, 2011). It is well known that lineages of these endosymbionts commonly undergo host shifts (Heath et al. 1999; Nikoh et al. 2008; Le Clec'h et al. 2013). In our two *Drosophila* samples (DsP and DiG), we found highly similar *Wolbachia* lineages. These results are very suggestive that there were, or are, contacts allowing *D. incompta* to share *Wolbachia* endosymbionts with generalist *Drosophila*, despite the fact that these species have been ecologically separated from generalist *Drosophila* for quite some time (Robe et al. 2013). *Wolbachia* can be suggested as putative HTT vectors for the events observed between the DsP and DiG samples. However, other vectors that were not considered in this work, such as viruses, could also be candidate HTT vectors.

Additionally, the similar elements shared by *Drosophila* and a freshwater planarian species (*Schimidea mediterranea*) deserve attention. We cannot hypothesize that a direct HTT event occurred between these species, which have been separated geographically and ecologically since approximately 625 Mya. However, it is evident that this element has not been maintained via vertical transmission from the latest common ancestor of these two species. Robertson (1997) previously reported the occurrence of HT of *mariner* elements between insects and planarians. However, the potential vector that is able to transport these TEs

between species that appear to have nothing in common is unknown. Most likely, other unknown donors are involved in these *mariner* HTT events. For example, we can hypothesize a more complex explanation in which an HTT event could have occurred between a freshwater planarian and an arthropod species that spends part of its life cycle in the water, with a second transfer then taking place between this species and *Drosophila*.

In summary, (1) our findings suggest that HTTs do not seem to be frequent between *Drosophila* species and their parasitoid wasps, as we did not find any TEs with high similarity between the two examined sets of *Drosophila*–parasitoid pairs; (2) despite an exhaustive search for PDV-related sequences, only one PDV gene was found in this study, indicating that if a similar system of HTT is present in *Drosophila*, it is divergent from PDV described in Lepidoptera; (3) well-supported HTT events were identified between generalist and flower specialist *Drosophila*, indicating that some vectors are able to bypass the ecological barriers observed in these species; (4) the similar *Wolbachia* lineages shared by ecologically separated *Drosophila* suggest that these endosymbionts may be potential vectors between different *Drosophila* species; and (5) the existence of similar TEs indicating HTT between ecologically separated *Drosophila* and the virtual absence of these events among *Drosophila* and their parasitoid wasps reinforce the notion that host factors can be important parameters for HTT.

Data deposition

NCBI SRA: All raw data from the Illumina sequencing dataset is archived in the SRA database at the NCBI website. BioProject ID: PRJNA215814—Title: Metagenome of *Drosophila* and their parasitoid wasp species. BioSample accessions SAMN02319231. For the Wolbachia sequences the aligment in .fasta file can be found in Supplementary file 1.fasta.

Acknowledgments This study was supported by research grants and fellowships from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), PRONEX-FAPERGS (10/0028-7) and FAPERGS (11/0938-0). We thank Julien Varaldi for his critical reading and very useful comments.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389 3402

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55:539 552

Bächli G (2009) TaxoDros: The Database on Taxonomy of Drosophilidae, v. 1.03, Database 2009/04. http://taxodros.unizh.ch/. Accessed 05 May 2013

Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MCJ, Tettelin H, Werren JH (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol 72:7098 7110

Beall EL, Rio DC (1997) Drosophila *P* element transposase is a novel site-specific endonuclease. Genes Dev 11:2137 2151

Bmcic D (1983) The flavopilosa group of species as an example of flower-breeding species. In: Ashburner M, Carson HL, Thompson JN (eds) The genetics and biology of *Drosophila*. Academic Press, New York, pp 360 377

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S (2003) LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res 13:721 731

Carareto CMA (2011) Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. Mob Genet Elem 1 8

Carton Y, Nappi AJ (1997) Drosophila cellular immunity against parasitoids. Parasitol Today 13:218 227

Carton Y, Bouletreau M, van Alphen JJM, van Lenteren JC (1986) The *Drosophila* parasitic wasps. In: Ashburner M, Carson L, Thompson JN (eds) The genetics and biology of *Drosophila*. Academic Press, London, pp 347 394

Casse N, Bui QT, Nicolas V, Renault S, Bigot Y, Laulier M (2006) Species sympatry and horizontal transfers of Mariner transposons in marine crustacean genomes. Mol Phylogenet Evol 40:609 619

Chandler JA, James PM (2013) Discovery of trypanosomatid parasites in globally distributed *Drosophila* species. PLoS One 8:e61937

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman T, Kellis M, Gelbart W et al (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203 218

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

Dupeyron M, Leclercq S, Cerveau N, Bouchon D, Gilbert C (2014) Horizontal transfer of transposons between and within crustaceans and insects. Mob DNA 5:4

Dupuy C, Periquet G, Serbielle C, Bezier A, Louis F, Drezen JM (2011) Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. Genetica 139:489 496

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res. 1:32(Web Server issue):W273-9

Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C (2010) A role for host parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347 1350

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou EA, Cordaux R (2014) Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun 5:3348

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307 321

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads a baiting and iterative mapping approach. Nucleic Acids Res 41:e129

Heath B, Butcher R, Whitfield W, Hubbard S (1999) Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Curr Biol 9:313 316

Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971 2972

Herniou EA, Huguet E, Thézé J, Bézier A, Periquet G, Drezen JM (2013) When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Philos Trans R Soc B 12:368

Hotopp JCD (2011) Horizontal gene transfer between bacteria and animals. Trends Genet 27(4):157 163

Hotopp JCD et al (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753 1756

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

Kacsoh BZ, Schlenke TA (2012) High hemocyte load is associated with increased resistance against parasitoids in *Drosophila suzukii*, a relative of *D. melanogaster*. PLoS One 7:e34721

Klasson L, Westberg J, Sapountzis P, Näslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG (2009) The mosaic genome structure of the Wolbachia wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci USA 106:5725 5730

Laha T, Loukas A, Wattanasatitarpa S, Somprakhon J, Kewgrai N, Sithithaworn P, Kaewkes S, Mitreva M, Brindley PJ (2007) The bandit, a New DNA Transposon from a Hookworm-possible horizontal genetic transfer between host and parasite. PLoS Negl Trop Dis 1:e35

Le Clec'h W, Chevalier FD, Genty L, Bertaux J, Bouchon D, Sicard M (2013) Cannibalism and predation as paths for horizontal passage of *Wolbachia* between terrestrial isopods. PLoS One 8:e60232

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

Markow TA, O'Grady PM (2005) Drosophila: a guide to species identification and use. Academic Press, San Diego

Mejlumian L, Pelisson A, Bucheton A, Terzian C (2002) Comparative and functional studies of *Drosophila* species invasion by the gypsy endogenous retrovirus. Genetics 160:201 209

Mota NR, Ludwig A, Valente VLS, Loreto ELS (2009) Harrow: new Drosophila hAT transposons involved in horizontal transfer. Insect Mol Biol 19:217 228

Nappi AJ, Vass E (1993) Melanogenesis and the generation of cytotoxic molecules during insect cellular immune reactions. Pigment Cell Res 6:117 126

Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T (2008) *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. Genome Res 18:272 280

Perlman SJ, Jaenike J (2003) Evolution of multiple components of virulence in *Drosophila*-nematode associations. Evolution 57(7):1543 1551

Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. Bioinformatics 21(suppl 1):i351 i358

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life data System (www.barcodinglife.org). Mol Ecol Notes 7:355 364

Rizki RM, Rizki TM (1990) Parasitoid virus-like particles destroy *Drosophila* cellular immunity. Proc Natl Acad Sci USA 87:8388 8392

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

Robe LJ, Cordeiro J, Loreto ELS, Valente VLS (2010a) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). Genetica 138:601 617

Robe LJ, Loreto ELS, Valente VLS (2010b) Radiation of the, *Drosophila* "subgenus (Drosophilidae, Diptera) in the Neotropics. J Zool Syst Evol Res 48:310 321

Robe LJ, Cenzi de Ré F, Ludwig A, Loreto ELS (2013) The *Drosophila flavopilosa* species group (Diptera, Drosophilidae): an array of exciting questions. Fly 7:59 69

Robertson HM (1997) Multiple Mariner transposons in flatworms and hydras are related to those of insects. J Hered 88:195 201

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

Routh A, Domitrovic T, Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci USA 109:1907 1912

Russo J, Dupas S, Frey F, Carton Y, Brehelin M (1996) Insect immunity: early events in the encapsulation process of parasitoid (*Leptopilina boulardi*) eggs in resistant and susceptible strains of Drosophila. Parasitology 112:135 142

Schaack S, Gilbert C, Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537 546

Silva JC, Kidwell MG (2000) Horizontal transfer and selection in the evolution of P elements. Mol Biol Evol 17:1542 1557

Silva JC, Loreto E, Clark J (2004) Factor that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol 6:57 72

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117 1123

Smit AFA, Hubley R Green P (1996 2010) RepeatMasker Open 3.0 http://www.repeatmasker.org

Stefanov Y, Salenko V, Glukhov I (2012) *Drosophila* errantiviruses. Mob Genet Elem 2:36 45

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

The Nasonia Genome working group (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* Species. Science 327:343 348

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

Wheeler MR, Takada H e Brncic D (1962) The *flavopilosa* species group of *Drosophila*. Studies in Genetic II. Univ Texas Publ pp 395 413

Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T, Kimura K (2001) Possible horizontal transfer of a transposable element from host to parasitoid. Mol Biol Evol 18:1952 1958

CAPÍTULO IV

Artigo em preparação a ser submetido para revista *Genome Research*

Insectivorous bats are the more prone mammals to DNA transposon invasions

Mauro F. Ortiz*[1], Elgion L. S. Loreto[2]; Cedric Feschotte[3]

1 - PGGBM, UFRGS; Univiversidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

2 - Departamento de Biologia, CCNE, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil.

3 – University of Utah, Salt lake City, Utah, EUA

* Corresponding author

ABSTRACT

Almost all species have transposable elements (TEs) and they can account for significant portions of the genomes. Most of the time the TEs “live” inside of the genome but eventually they “jump” from one genome to another, a process called horizontal transposon transfer (HTT). Unlike arthropod genomes, the mammalian genomes generally lack recent DNA transposon activity or TEs invasions. As an exception, the vesper bats have show some recent DNA transposon activity and multiples waves of TEs invasions. We performed a *in silico* searches for horizontal transfer of DNA transposon families among 82 mammal genomes. We found that at least 82 different TE families invaded mammalian lineages in the last 50 Myr. The highest number occurred in the insectivorous bats lineages, where we identified 62 (75%) different TE families invasions. However, there are a strikingly contrast with groups that have no TE invasions in the last 50 Myr, such as orders Cetacea and Carnivora or few sporadic invasions, like Rodentia. We also determined that 82 families that invaded mammals, at least 21 are shared with arthropods species. These results show that insectivorous bats are a special group among mammals in regards to horizontal transfers. Additionally, because insectivorous bats are very diverse and their ability to fly, allows them to occupy wide range of environments, we suggest that this may increase the exposure to parasites, prey and vector. Because host-parasite interactions have been implicated in increased rates in HT, it is possible that this expose could have accelerated the abundant HT events seen in the insectivorous bats.

INTRODUCTION

Transposable elements (TEs) are ubiquitous repetitive components of the genome and can account for a significant fraction of eukaryotic genomes. They may be classified in two classes based upon their method of transposition. Class I is represented by the retrotransposons and Class II includes the “cut-and-paste” DNA transposons. The TE content of human genome reach 45% of all DNA and Class II TEs match 3% of the genome (Lander et al 2001). Most of the time, TEs propagate within the genome but, on occasion, they are capable to “jump” from one genome to another. This process, called horizontal transposon transfer (HTT), explains the presence of nearly identical TE in species separated by large evolutionary distances (Silva et al 2004). Although the mechanism by which HT of TE occurs remains mysterious, it seems to be a common and widespread phenomenon (Schaack et al 2010, Wallau et al 2012).

Horizontal transposon transfer has been described in several organism, from plants (Baidouri et al 2014) to animals (Gilbert et al 2010). Several striking cases of DNA transposon transfers have been reported in mammals but it is still not clear whether these are exceptional events and if they are restricted to a particular time frame or phylogenetic groups. Theoretical models have indicated that for a successful genome invasion is necessary be followed by a TE copy number burst, indicated by a intense transpositional activity (Rouzic and Capy 2005). After the phases of increase of copy number, mechanisms of silencing or control of transposition is developed by the host genome. This include a range of epigenetic mechanisms as methylation, histone modifications and RNA

interference (Hua-Van et al 2011). Normally, after the establishment of mechanisms for control the transposition, TEs became inactive and degenerated. For example, some evidences show a intense ancient activity of DNA transposons at primates lineages but a cessation in the transpositional activity of DNA transposons in a recent period, without evidence for any human DNA transposon families younger than 37 My (Pace and Feschotte 2007). Otherwise, have been recorded 23 families of DNA transpons younger than 40 My in some bat lineages and at least half are more recent than 10 My (Ray et al 2008). This suggest us that HTT may be not a uniform process over time and across the phylogeny of mammals.

To shed light on these questions, we conducted a systematic survey of HT of DNA transposons over the last 50 million years (My) and across 82 mammalian genomes. We perform the first systematic and comprehensive search for HTT at mammals looking for frequency and widespread patterns. We found that at least 82 different TE families invaded mammalian lineages which 62 (75%) occurred particularly in the insectivorous bats lineages. Furthermore 21 families are shared with arthropods species, showing us potential sources to these TEs or at least species that share the common TE sources. Finally, our results suggest the insectivorous bats are a stand out group among mammals regarding to HTT seems a group of mammals striking prone to TEs invasions.

RESULTS

We conducted a search for HTT among 82 mammal genomes belonging to

24 orders. We look for know DNA transposons (Class II) using homology based searches. We established as cut-off to analyze TE genome invasions that have occurred in the last 50 My at mammals lineages. In this periods almost all mammals order had already clearly diverged (Meredith et al 2011, Springer et al 2011), furthermore determine the more ancient HTT could be imprecise. First, we retrieved a TE database Repbase (Jurka et al 2005) and the genome sequences assemblies (www.ncbi.nlm.nih.gov). As initial criteria to find the potential HTT events we considered the elements that share at least 80% of identity in at least 80% of your size or when the elements are unique in the lineage. Using this approach we detected 143 invasions (HTT) of 82 different TE families in the mammal lineages (Figure 1). We highlight that some TE families invaded independently more than one lineage, like *SPIN* family that invaded seven times the mammals lineages or *HSMAR1* with eleven invasions. This explain the higher value of invasions than the TE families. These 82 TE families are represented primarily just by three DNA transposons superfamilies, *mariner*, *hAT* and *piggyBac*. The mammal species shown to be more prone to invasions TE from *mariner* superfamily with 38 cases, followed by *hAT* with 23 and *piggyBac* with 16 cases. Five TE families are from unknown superfamilies. There are a concordance between the families more involved in HTT and numerical abundance in genomes since the *mariner* and *hAT* are the two most abundant superfamilies of DNA transposons described in all species and *piggyBac* one of most abundant, especially in animals (Jurka et al 2005). Despite the HTT distribution bias discussed below, the *mariner* and *hAT* supefamilies are not apparently restricted to

specific lineages, otherwise the *piggyBac* seems to be restricted to Chiroptera and the primate mouse lemur (*Microcebus murinus*).

We can see some strong differences among the lineages in regards the frequencies of HTT. Some mammals present virtually no invasions in the last 50 My, like in Cetartiodactyla or Carnivora orders (Figure 1) otherwise the representative number of species with available genomes. Altogether there are 31 mammal lineages without DNA transposons invasions in the last 50 My (Figure 1). This contrast with some orders with sporadic invasions like Rodents (8 invasions/6 families) or Simiiformes primates (6/5), but this strikingly contrast with some lineages that seems to be more propensity to DNA transposons invasions. The Chiroptera order proved to be extremely prone to receive new TEs, with 83/63 invasions at the last 50 My (Figure 1). But even within bats there are great bias, especially between the insectivorous and non-insectivourous bats. In this case the non-insectivourous or frugivorous bats have only 2/2 invasions while the insectivorous bats count 81/62 invasions of Chiroptera order (Figure 1). At this first analysis, specialty a family Vespertilionidae of bats lineage (*Eptesicus fuscus*, *Myotis brandtii*, *M. davidii*, *M. lucifugus*) seems to be a special group with 40/40 invasions. The number of invasions could be underestimated in this group, especially because *Myotis* group have close related species making difficult to count independently invasions in this branch. Four other mammals species, the marsupial opossum (14 invasions), the primates lemur (10) and the mouse lemur (7) and the tenrec (7) from Afrosoricida order, are also highlighted by the number of HTT, however they account a lower invasions each one and they are not a

phylogenetic cluster.

We identified that 21(25%) from this 82 DNA transposons families what invaded mammals recently are share with arthropods (Figure 1). We considered a HTT elements that share at least 70% of identity in at least 70% of your size. Any TEs with this identity level are, much probably, share by HTT once any arthropods and mammals diverged at least 700 My ago. Among these arthropods species we detected diverse kind of organisms. These include 27 arthropods species, from 10 orders. Two groups in particular stand out among this arthropods, arachnids and hymenopterans. We have identified five arachnids species that share 15 TEs with mammals while 13 hymenopterans species share seven different TEs (Table 2). Among species that most share TEs are the scorpions, *Mesobuthus martensii* (11) and *Centruroides exilicauda* (8), the spider, *Stegodyphus mimosarum* (6) and the ant, *Harpegnathos saltator* (4).

DISCUSSION

Our results show that a biological phenomenon horizontal transposon transfers is unequally distributed in mammals lineages, suggesting that HTT is an evolutionary force impacting unevenly the mammal lineages. We concluded this because among 24 mammals orders analyzed there are a large discrepancy among invasions frequency and distribution. We detected zero invasions in 11 orders and a significant amount of invasions in others (83 in Chiroptera), especially in some insectivorous lineages, highlighting the vesper bats. We verified a special volume of invasions in insectivorous bats (81) represented by seven

species, especially when compared with megabats (2), a frugivorous lineage of bats, represented by three species. This lead us to think that a feeding behavior of bats could promote the expose to DNA transposons invasions leading the insectivorous bats to be potentially more prone to DNA transposable invasions. The HTT could happen through vectors but could occur without them too. Our hypothesis is reinforced when we look the secondary group prone to invasions, all four species have insectivorous behavior and looking some orders with few or sporadic invasions, particularly among Laurasiatheria group, orders like Cetacea, Perissodactyla, Artiodactyla and Carnivora, any of them exhibit the insectivorous habit. Finally, our results clearly show that some species are more susceptible to invasion by genomic parasites than others.

This idea of insectivore diet of mammals (potentially the same thing as arthropods diet) to explain the HTT is supported when we searched for potential sources to this TEs or at least species who share the same TEs. At least 25% of all TEs are shared with arthropods, mainly arachnids (spiders and scorpions) and insects(especially wasp and ants). It is direct the predation relationship between insects and insectivorous mammals, especially bats, although this is not so evident between spiders and mammals. But typically arachnids are insectivorous which could explain the sharing TEs (8) with some insects. In that case mammals and arachnids are sharing TEs indirectly, by share the same potentially source, the food. The TEs horizontally transferred some times seems to be omnipresents in the species (see review, Wallau et al 2012). It seems to be evident that biological relationship is the way of promotions of promiscuity of TEs by species, but now we

see that the foods chains could be a explanations to this omnipresence of TEs. The relationship of vertebrates and insect has been explored to explain the HTT events (Gilbert et al 2010), in that case, the host-parasite interactions. The biological interactions, as parasitism and the predation (insectivory in this case) in any way are mutually exclusive forms to promote HTT, but could be a associated forms of biological interactions promoting the TEs spread in the species. Because this seems plausible the idea of insectivore diet of mammals is a way of promoting the exposure to the new DNA transposons invasions.

Intrinsics characteristics of species could too explain why some groups are more prone to successful HTT. These include differences into epigenetics mechanisms of TEs control (Hua-Van et al 2011). Also differences of viromes to which the species are expose once this infections agents are a potential vectors to spread TEs (Gilbert et al 2014). We can also mentioning that DNA transposons are simples DNA strands that eventually not even need the vector since they could form autonomous particulates like episomes (O'Brochta et al 2009) or yet use the transport system (microvesicles) of organism as exosomes (Balaj et al 2011).

METHODS

We performed searches for HTT at 82 available mammals genomes comprising 24 different mammals orders (see Supplementary Table 1 to complete list and informations) download from NCBI Genome Browser (http://www.ncbi.nlm.nih.gov/genome/browse/). At the first step to detect HTT we perform homology based searches, using BLAST software (Altschul et al 1990),

querying a total of 7417 Class II transposable elements from Repbase18.10 database (Jurka et al 2005) against all mammals genomes. We exclude the *Helitron* superfamily because of its distinct characteristics of TIR DNA transposons. Initially, we considered potential HTT events, the TEs with hits with sequence identity >80% in at least 80% of your size. Moreover, we performed the *de novo* searches at select species because some available genomes are relative new generated and don't have your TE content fully inquired and to discard the potential bias from initial query datasets. We used RepeatScout software to *de novo* with default parameters (Price et al 2005). *Ab initio* searches was conducted in the 20 following species: Chiroptera (*Rhinolophus ferrumequinum*, *Pteronotus parnellii*, *Megaderma lyra*, *Pteropus alecto*, *Eidolon helvum*); Didelphimorphia (*Monodelphis domestica*); Afrosoricida (*Echinops telfairi*, *Chrysochloris asiatica*); Primates (*Microcebus murinus*, *Otolemur garnettii*, *Daubentonia madagascariensis*, *Tarsius syrichta*, *Saimiri boliviensis*); Cetartiodactyla (*Sus scrofa*, *Bos taurus*); Carnivora (*Canis lupus*); Rodentia (*Chinchilla lanigera*, *Microtus ochrogaster*), Scandentia (*Tupaia belangeri*) and Hyracoidea (*Procavia capensis*).

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403–410

Baidouri ME, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O. 2014. Widespread and frequent

horizontal transfers of transposable elements in plants. *Genome Res* **24:** 831–838.

Balaj L, Lessard R, Dai L, Cho YJ, Pomeroy SL, Breakefield XO, Skog J. 2011. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat Commun* **2:** 180.

Gilbert C, Schaack S, Pace II JK, Brindley PJ, Feschotte C, Pace JK. 2010. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. *Nature* **464:** 1347–1350.

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou E, Cordaux R. 2014. Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat Commun* **5:** 3348.

Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011. The struggle for life of the genome's selfish architects. *Biol Direct* **6:** 19.

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.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334:** 521–524.

O'Brochta D, Stosic CD, Pilitt K, Subramanian RA, Hice RH and Atkinson PW. 2009. Transpositionally active episomal *hAT* elements. *BMC Mol Biol* **10:** 108.

Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res* **17:** 422–432.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. To appear in Proceedings of the 13 Annual International conference on Intelligent Systems for Molecular Biology (ISMB-05). Detroit, Michigan.

Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P, Atkinson PW and Craig NL. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* **18:** 717–728.

Rouzic AL, Capy P. 2005. The First Steps of Transposable Elements Invasion: Parasitic Strategy vs. Genetic Drift. *Genetics* **1043:** 1033–1043.

Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* **25:** 537–546.

Silva JC, Loreto ELS, Clark JB. 2004. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol* **6:** 57–71.

Springer MS, Meredith RW, Janecka JE, Murphy WJ. 2011. The historical biogeography of Mammalia. *Philos Trans R Soc Lond B Biol Sci* **366:** 2478–502.

Wallau GL, Ortiz MF, Loreto ELS. 2012. Horizontal Transposon Transfer in Eukarya: detection, bias and perspectives. *Genome Biol Evol* **4:** 689–699.

FIGURE LEGENDS

Figure 1. Distribution of HTT events in 82 mammals species. Each cycle represent a HTT. In the left there are all TEs divided in the superfamilies *mariner* (yellow), *hAT* (green), *piggybac* (blue) and unknown families (grey). The black cycles (A) indicate the TEs with are shared with arthropods species. This is a schematic representation of phylogeny with no branches support modified from Springer et al 2011 and Meredith et al 2011.

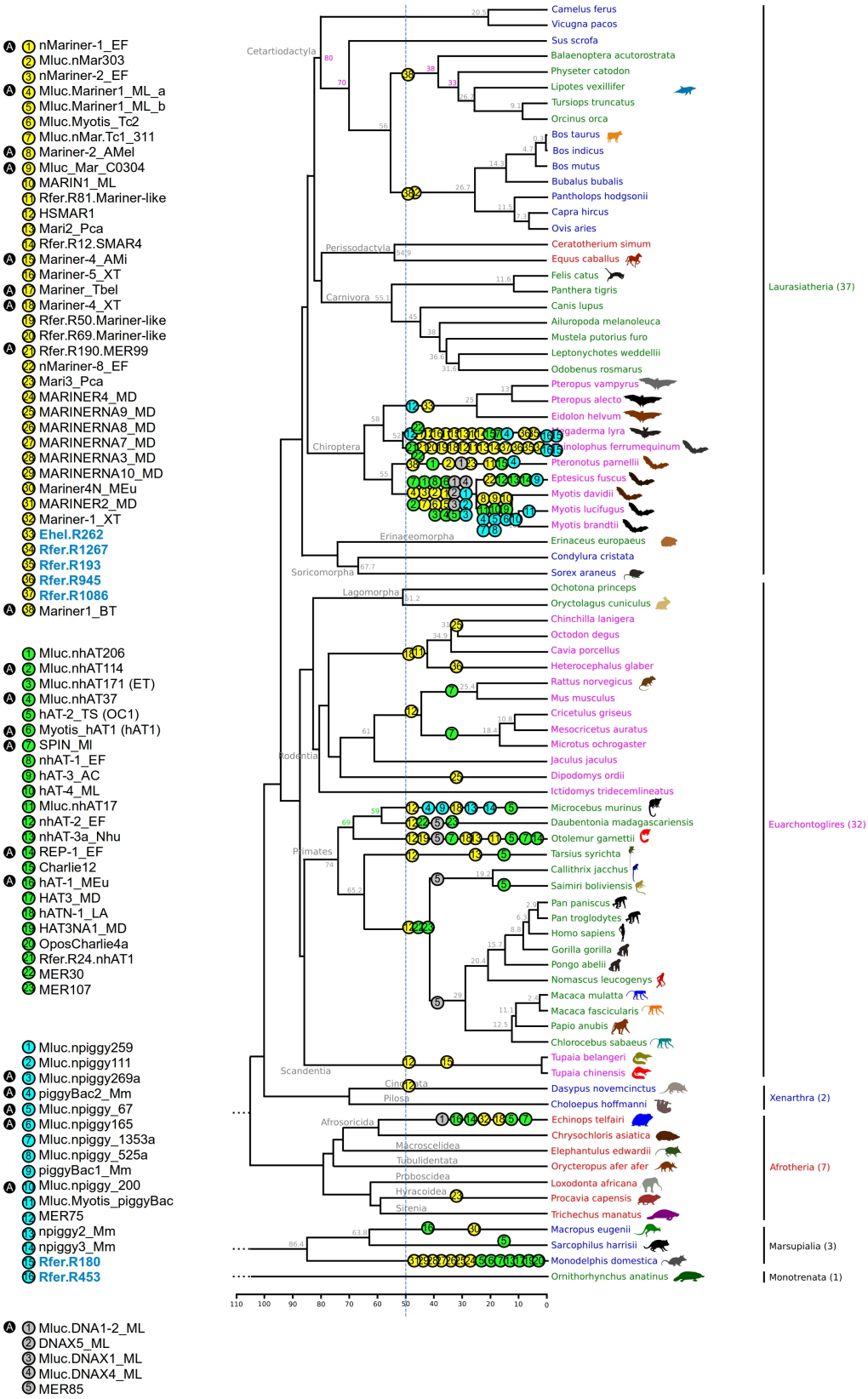
A 1 nMariner-1_EF
2 Mluc.nMar303
3 nMariner-2_EF
A 4 Mluc.Mariner1_ML_a
5 Mluc.Mariner1_ML_b
6 Mluc.Myotis_Tc2
7 Mluc.nMar.Tc1_311
A 8 Mariner-2_AMel
A 9 Mluc_Mar_C0304
10 MARIN1_ML
11 Rfer.R81.Mariner-like
12 HSMAR1
13 Mari2_Pca
14 Rfer.R12.SMAR4
A 15 Mariner-4_AMi
16 Mariner-5_XT
A 17 Mariner_Tbel
A 18 Mariner-4_XT
19 Rfer.R50.Mariner-like
20 Rfer.R69.Mariner-like
A 21 Rfer.R190.MER99
22 nMariner-8_EF
23 Mari3_Pca
24 MARINER4_MD
25 MARINERNA9_MD
26 MARINERNA8_MD
27 MARINERNA7_MD
28 MARINERNA3_MD
29 MARINERNA10_MD
30 Mariner4N_MEu
31 MARINER2_MD
32 Mariner-1_XT
33 Ehel.R262
34 Rfer.R1267
35 Rfer.R193
36 Rfer.R945
37 Rfer.R1086
A 38 Mariner1_BT
1 Mluc.nhAT206
A 2 Mluc.nhAT114
3 Mluc.nhAT171 (ET)
A 4 Mluc.nhAT37
5 hAT-2_TS (OC1)
A 6 Myotis_hAT1 (hAT1)
A 7 SPIN_MI
8 nhAT-1_EF
9 hAT-3_AC
10 hAT-4_ML
11 Mluc.nhAT17
12 nhAT-2_EF
13 nhAT-3a_Nhu
A 14 REP-1_EF
15 Charlie12
A 16 hAT-1_MEu
17 HAT3_MD
18 hATN-1_LA
19 HAT3NA1_MD
20 OposCharlie4a
21 Rfer.R24.nhAT1
22 MER30
23 MER107
1 Mluc.npiggy259
2 Mluc.npiggy111
A 3 Mluc.npiggy269a
A 4 piggyBac2_Mm
A 5 Mluc.npiggy_67
A 6 Mluc.npiggy165
7 Mluc.npiggy_1353a
8 Mluc.npiggy_525a
9 piggyBac1_Mm
A 10 Mluc.npiggy_200
11 Mluc.Myotis_piggyBac
12 MER75
13 npiggy2_Mm
14 npiggy3_Mm
15 Rfer.R180
16 Rfer.R453
A 1 Mluc.DNA1-2_ML
2 DNAX5_ML
3 Mluc.DNAX1_ML
4 Mluc.DNAX4_ML
5 MER85
Camelus ferus
Vicugna pacos
Sus scrofa
Balaenoptera acutorostrata
Physeter catodon
Lipotes vexillifer
Tursiops truncatus
Orcinus orca
Bos taurus
Bos indicus
Bos mutus
Bubalus bubalis
Pantholops hodgsonii
Capra hircus
Ovis aries
Ceratotherium simum
Equus caballus
Felis catus
Panthera tigris
Canis lupus
Ailuropoda melanoleuca
Mustela putorius furo
Leptonychotes weddellii
Odobenus rosmarus
Pteropus vampyrus
Pteropus alecto
Eidolon helvum
Megaderma lyra
Rhinolophus ferrumequinum
Pteronotus parnellii
Eptesicus fuscus
Myotis davidii
Myotis lucifugus
Myotis brandtii
Erinaceus europaeus
Condylura cristata
Sorex araneus
Ochotona princeps
Oryctolagus cuniculus
Chinchilla lanigera
Octodon degus
Cavia porcellus
Heterocephalus glaber
Rattus norvegicus
Mus musculus
Cricetulus griseus
Mesocricetus auratus
Microtus ochrogaster
Jaculus jaculus
Dipodomys ordii
Ictidomys tridecemlineatus
Microcebus murinus
Daubentonia madagascariensis
Otolemur garnettii
Tarsius syrichta
Callithrix jacchus
Saimiri boliviensis
Pan paniscus
Pan troglodytes
Homo sapiens
Gorilla gorilla
Pongo abelii
Nomascus leucogenys
Macaca mulatta
Macaca fascicularis
Papio anubis
Chlorocebus sabaeus
Tupaia belangeri
Tupaia chinensis
Dasypus novemcinctus
Choloepus hoffmanni
Echinops telfairi
Chrysochloris asiatica
Elephantulus edwardii
Orycteropus afer afer
Loxodonta africana
Procavia capensis
Trichechus manatus
Macropus eugenii
Sarcophilus harrisii
Monodelphis domestica
Ornithorhynchus anatinus
Cetartiodactyla
Perissodactyla
Carnivora
Chiroptera
Erinaceomorpha
Soricomorpha
Lagomorpha
Rodentia
Primates
Scandentia
Cingulata
Pilosa
Afrosoricida
Macroscelidea
Tubulidentata
Proboscidea
Hyracoidea
Sirenia
Laurasiatheria (37)
Euarchontoglires (32)
Xenarthra (2)
Afrotheria (7)
Marsupialia (3)
Monotrenata (1)
110
100
90
80
70
60
50
40
30
20
10
0

Table 1. Description of HTTs identified in the mammal genomes.

HTT	DNA TE Family	Identity (%)	Invasions
1	Mluc.npiggy_67	99	1
2	Mluc.npiggy165	99	1
3	Mluc.Myotis_piggyBac	99	1
4	Mluc.nhAT17	99	1
5	nhAT-3a_Nhu	99	2
6	Rfer.R24.nhAT1	99	1
7	Mluc.npiggy_200	99	1
8	Mluc.npiggy_525a	99	1
9	MARIN1_ML	99	1
10	Myotis_hAT1 (hAT1)	98	2
11	hAT-3_AC	98	1
12	Mluc.npiggy_1353a	98	1
13	nMariner-8_EF	98	1
14	Mluc_Mar_C0304	97	1
15	REP-1_EF	97	3
16	SPIN_MI	97	7
17	Mluc.npiggy111	97	1
18	Mluc.npiggy269a	97	1
19	MER85	97	4
20	DNAX5_ML	96	1
21	Mluc.nhAT171 (ET)	96	1
22	Mluc.Mariner1_ML_a	96	1
23	Mluc.Mariner1_ML_b	96	1
24	piggyBac1_Mm	96	2
25	piggyBac2_Mm	96	5
26	npiggy2_Mm	96	1
27	nhAT-2_EF	96	1
28	Mluc.nhAT37	95	1
29	Mluc.npiggy259	95	1
30	hAT-4_ML	95	1
31	Mluc.DNAX1_ML	95	1
32	nhAT-1_EF	95	1
33	MER107	95	2
34	Mluc.nhAT114	94	1
35	nMariner-1_EF	94	2
36	Rfer.R69.Mariner-like	94	1
37	Rfer.R1267	94	1
38	Rfer.R453	94	2
39	Mluc.Myotis_Tc2	93	1
40	Rfer.R190.MER99	93	1
41	hATN-1_LA	93	1
42	Rfer.R50.Mariner-like	93	2
43	Rfer.R180	93	2
44	hAT-2_TS (OC1)	92	8
45	nMariner-2_EF	92	1
46	Rfer.R81.Mariner-like	92	4
47	npiggy3_Mm	92	1
48	Mariner_Tbel	92	1
49	Mluc.DNA1-2_ML	92	3
50	Rfer.R193	92	2
51	HSMAR1	91	5
52	Rfer.R1086	91	1
53	Mluc.nhAT206	90	2
54	Mluc.nMar303	90	2
55	Mari2_Pca	90	5
56	Rfer.R12.SMAR4	90	2
57	MARINERNA8_MD	90	1
58	MARINERNA10_MD	90	1
59	MARINER2_MD	90	1
60	Rfer.R945	90	2
61	MER30	90	4
62	Mari3_Pca	89	1
63	Mariner-1_XT	89	1
64	Ehel.R262	89	1
65	Mluc.nMar.Tc1_311	88	1
66	Mariner-4_AMi	88	1
67	Mariner1_BT	88	1
68	MARINERNA9_MD	88	2
69	Mariner-2_AMel	87	1
70	Mariner-5_XT	87	1
71	MER75	87	3
72	hAT-1_MEu	87	2
73	MARINERNA7_MD	87	1
74	Mariner4N_MEu	87	1
75	Mariner-4_XT	86	5
76	HAT3_MD	86	1
77	OposCharlie4a	86	1
78	MARINERNA3_MD	85	1
79	Mluc.DNAX4_ML	84	1
80	Charlie12	83	3
81	HAT3NA1_MD	83	1
82	MARINER4_MD	82	1

Table 2. Distribution of TEs shared between arthropods and mammals species. First column include the TEs name and the following columns the most representative arthropods species.

TEs	*Mesobuthus martensii* 	*Centruroides exilicauda* 	*Stegodyphus mimosarum* 	*Harpegnathos saltator* 	Others
Mluc_Mar_C0304	+	+	+		+
Mariner-4_XT	+	+	+		
Mluc.Mariner1_ML_a	+	+		+	+
Mariner-4_AMi	+	+			
Mariner1_BT	+	+			
Mariner-2_AMel	+		+	+	+
Mariner_Tbel	+		+	+	+
REP-1_EF	+		+		
Mluc.nhAT114	+				
Rfer.R190.MER99	+				
hAT-1_MEu	+				
Mluc.nhAT37		+			+
SPIN_Ml		+			+
Myotis_hAT1 (hAT1)		+			
piggyBac2_Mm			+		+
Mluc.npiggy_200				+	+
Mluc.npiggy_67					+
Mluc.npiggy165					+
nMariner-1_EF					+
Mluc.DNA1-2_ML					+
Mluc.npiggy269a					+

CAPÍTULO V

Discussão

É cada vez mais evidente a importância evolutiva dos elementos de transposição junto aos genomas eucariotos. Hoje os TEs possuem cada vez menos a alcunha de DNA lixo. No entanto, ainda na maioria das vezes acabam comportando-se como parasitas genômicos e em muitas situações acabam gerando detritos genômicos que podem corresponder a porções significativas dos genomas e de fato, sem função, tais como DNA lixo (Palazzo e Gregory 2014). Os TEs também podem tornarem-se úteis para seu hospedeiro, processo denominado domesticação. O fato é que o estudo da evolução dos TEs além de ajudar na compreensão desses elementos que possuem uma "vida" paralela nos genomas dos eucariotos, também traz contribuições no entendimento da evolução dos genomas dos eucariotos. Estudos específicos de um grupo particular de TEs ou mais amplos como a busca por padrões gerais dos TEs certamente vem a trazer contribuições nesse sentido.

O primeiro capítulo desta tese traz novas informações acerca de história evolutiva de uma grupo de TEs em particular, os elementos *P* de *Drosophila*. O elemento *P* é um transposon com TIRs descoberto primeiramente em *Drosophila melanogaster* (Bingham et al 1982; Rubin et at 1982). É desse grupo também, o primeiro caso de HTT descrito, entre *D. melanogaster* e *D. willistoni* (Daniels et al 1990). Outros transposons dessa mesma superfamília estão presentes em outros grupos de seres vivos, mas são distantemente relacionados aos elementos de *Drosophila* e de algumas outras espécies de insetos onde também são encontrados. Além de suportar a monofilia dos elementos *P* em *Drosophila* propomos a reclassificação das subfamílias já conhecidas e a criação de novas, totalizando 11 subfamílias. O fato dos elementos *P* de *Drosophila* serem monofiléticos mostra-se muito interessante por ao mesmo tempo mostrar-se a

existência de muitas incongruências filogenéticas entre a filogenia das diferentes subfamílias e a das suas espécies hospedeiras. Isso mostra que diferentes subfamílias podem coabitar um mesmo genoma hospedeiro mas também sugere inúmeras transferências horizontais, mas todas restritas as espécies de *Drosophila*. Esse cenário é marcadamente diferente de outras superfamílias de TEs como *mariner*, *hAT* ou *piggybac* em que os eventos de HTT não são restritos a um grupo filogenético (Wallau et al 2012). Esse fato confirma a existência de fatores específicos dos hospedeiros que restringem a HT de elementos *P* para outros taxa distantes. *Drosophila willistoni* também mostrou-se possuir representantes de cinco diferentes subfamílias. Essa diversidade de elementos nessa espécie pode refletir, em parte, sua ampla distribuição geográfica na América do Sul, o que propiciou o seu contato com outras espécies e consequentes invasões no seu genoma.

Diante do complexo cenário evolutivo de elementos *P* em *Drosophila* não foi estimado o número de eventos de HT entre estas espécies. Esse cenário deve-se em parte pelo alto número de HTT, mas também por muitas das espécies analisadas serem muito próximas evolutivamente. Isso porque uma HTT pode gerar uma série de falsos positivos quando comparados os valores de substituições sinônimas usados para inferir os eventos de HTT nesse tipo de caso. Essa discussão sobre subestimativas e superestimativas de HTT, além de outros pontos, foram tratados na revisão Wallau et al 2012 (Apêndice).

Embora o fenômeno da HTT seja cada vez mais evidente, ainda são pouco claros os mecanismos que levam a esses eventos. Uma série de possíveis vetores já foi levantada como meios de promoção da HTT (Schaack et al 2010). Mas ainda carecem evidências acerca dos mecanismos e possíveis vetores que levam a HTT. Na tentativa de tentar esclarecer sobre um potencial vetor é que propomos testar o papel de vespas parasitóides de *Drosophila* como potenciais vetores de HTT, por meio da infecção de alguns vírus do tipo polidnavírus. Segundo nossos resultados nossa hipótese mostrou-se falsa. No entanto, dados mais recentes apontam o potencial de alguns vírus do tipo baculovírus como vetores para HTT entre animais, especialmente entre insetos (Gilbert et al 2014).

Ainda mostram que esses eventos de HTT ocorrem em uma frequência baixíssima, o que poderia explicar parte desse resultado negativo dos experimentos com vespas parasitóides de *Drosophila*. Além disso, nem todas as espécies estão necessariamente igualmente propensas a invasões de TEs.

Por fim, no terceiro capítulo tratamos dos padrões de HTT em mamíferos. Nessa abordagem foram buscadas por todas as potenciais invasões de TEs nos genomas de mamíferos nos últimos 50 milhões de anos. Diferenças nos padrões de HTT entre as linhagens já eram apontados na literatura. Entre os grandes primatas, por exemplo, já tinha sido visto que essa linhagem sofreu uma grande quantidade de HT por transposons de DNA no passado mas que não haviam quase invasões recentemente (Pace e Feschotte 2007), enquanto que algumas linhagens de morcegos mostravam o contrário, uma intensa onda de invasões extremamente recente em suas genomas (Ray et al 2008). Os nossos dados confirmaram exatamente isso e ainda mais. Ao analisarmos 82 genomas, verificamos um grande viés entre as diferentes linhagens de mamíferos. Especialmente as linhagens de morcegos insetívoros apresentaram um grande volume de HTT nos últimos 50 milhões de anos. Outras linhagens também apresentaram um número significativo de HTT, embora em menor escala.

De forma geral, estes trabalhos contribuíram para o entendimento do fenômeno Transferência Horizontal de Transposons. E ao final, contribuímos para reforçar as evidências de que este é um fenômeno amplo e de grande impacto na história evolutiva das espécies. Os meios pelos quais este processo evolutivo se dá ainda precisam ser melhor esclarecidos. Mas já é evidente que esse processo é fundamental para sobrevivência dos TEs e que tem um forte impacto nas espécies hospedeiras dos TEs.

Perspectivas

A crescente quantidade de dados genômicos abre possibilidades imensuráveis para o estudo evolutivos dos TEs. E consequentemente de todos os processos biológicos inerentes a eles, incluindo suas transferências entre as espécies. Tais informações irão permitir a detecção de padrões evolutivos acerca das populações de TEs nos genomas. Chegaremos a uma definição melhor acerca de seu papel nos genomas, papel este que transita entre parasita e material potencial para seleção natural. Estes elementos de DNA já foram considerados sem importância e agora cada vez mais assumem um papel central na evolução das espécies. No entanto, ainda enfrenta-se muitas dificuldades para estudar os TEs. Apesar de sua crescente importância, muitos cientistas ainda consideram os TEs apenas material intergênico e pouco relevante. Ainda carecem programas para análise de TEs, assim como análises padronizadas destes elementos, mas a maior dificuldade da área talvez ainda seja outra. Necessita-se de uma urgente revisão da classificação dos TEs e de sua nomenclatura. A padronização nesses pontos é fundamental para o desenvolvimento dessa área do conhecimento.

REFERÊNCIAS BIBLIOGRÁFICAS

Aravin AA, Hannon GJ and Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318:761-764.

Baidouri ME, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA and Panaud O (2014) Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24:831–838.

Bingham PM, Kidwell MG, Rubin GM (1982) The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell 29:995–1004.

Capy P (1994) The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet 10:7-12.

Capy P, Bazin C, Higuet D, Langin T (eds.) (1998) Dynamics and evolution of transposable elements (Library of Congress, Austin, Texas).

Capy P, Gasperi G, Biemont C and Bazin C (2000) Stress and transposable elements: co-evolution or useful parasites? Heredity 85:101–106.

Carnelossi EAG, Lerat E, Henri H, Martinez S, Carareto CMA, Vieira C (2014) Specific activation of an I-like element in *Drosophila* interspecific hybrids. Genome Biol Evol 6:1806-1817.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN et al (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450: 203–18.

Cowley M and Oakey RJ (2013) Transposable Elements Re-Wire and Fine-Tune the Transcriptome. PLoS Genet 9:e1003234.

Cummings M (1994) Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol Evol 9: 141-145.

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

Daniels SB, Strausbaugh LD, Ehrman L and Armstrong R (1984)

Sequences homologous to *P* elements occur in *Drosophila paulistorum*. Proc Natl Acad Sci U S A 81:6794–6797.

Dunning Hotopp JC, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Muñoz TMC, Giebel JD, Kumar N, Ishmael N, Wang S et al (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753–1756.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9:397–405.

Fontdevila A (2005) Hybrid genome evolution by transposition. Cytogenet Genome Res 110:49–55.

George JA, Traverse KL, Debaryshe PG, Kelley KJ, Pardue ML (2010) Evolution of diverse mechanisms for protecting chromosome ends by Drosophila TART telomere retrotransposons. Proc Natl Acad Sci U S A 107:21052-21057.

Gilbert C, Schaack S, Pace II JK, Brindley PJ, Feschotte C, Pace JK (2010) A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.

Gilbert C, Chateigner A, Ernenwein L, Barbe V, Bézier A, Herniou E, Cordaux R (2014) Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. Nat Commun 5:3348.

Gould SJ and Vrba ES (1982). Exaptation - a missing term in the science of form. Paleobiology 8:4–15.

Hua-Van A, Le Rouzic A, Boutin TS, Filée J and Capy P (2011) The struggle for life of the genome's selfish architects. Biol Direct 6:19.

Josefsson C, Dilkes B and Comai L (2006) Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol 16:1322–1328.

Kapitonov VV, Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biology 3:e181.

Kidwell MG, Lisch D (2000) Transposable elements and host genome evolution. Trends Ecol Evol 15:95–99.

Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP (2009) Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. BMC

Genomics 10:33.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lippman Z, Vaughn MW, Dedhia N, Mccombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R et al (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471-476.

McClintock B (1950) The Origin and Behavior of Mutable Loci in Maize. Proc Natl Acad Sci U S A. 36:344–355.

Michalak P (2009) Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity 102:45–50.

O'Brochta D, Stosic CD, Pilitt K, Subramanian RA, Hice RH and Atkinson PW (2009) Transpositionally active episomal hAT elements. BMC Mol Biol 10:108.

O'Neill RJ, O'Neill MJ and Graves JA (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–72.

Pace JK and Feschotte C (2007) The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. Genome Res 17:422–432.

Palazzo AF and Gregory TR (2014) The Case for Junk DNA. PLoS Genetics 10:e1004351.

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

Piskurek O and Okada N (2007) Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc Natl Acad Sci USA 104:12046–12051.

Pritham EJ (2009) Transposable elements and factors influencing their success in eukaryotes. The Journal of Heredity 100:648–55.

Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P,

Atkinson PW and Craig NL (2008) Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. Genome Res 18:717–728.

Rebollo R, Romanish MT and Mager DL (2012) Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet 46:21–42.

Rieseberg LH, Van Fossen C and Desrochers AM (1995) Hybrid speciation accompanied by genomic reorganization in wild sunflowers. Nature 375:313–316.

Robertson HM (1997) Multiple Mariner Transposons in Flatworms and Hydras Are Related to Those of Insects. J Hered 88:195-201.

Rubin GM, Kidwell MG, Bingham PM (1982) The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29:987–994.

Routh A, Domitrovic T and Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci USA 109:1907-1912.

Rouzic AL and Capy P (2005) The First Steps of Transposable Elements Invasion: Parasitic Strategy vs. Genetic Drift. Genetics 1043:1033–1043.

Schaack S, Gilbert C and Feschotte C (2010) Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537–546.

Schnable PS, Ware D, Fulton RS, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115.

Silva JC and Kidwell MG (2000) Horizontal transfer and selection in the evolution of P elements. Mol Biol Evol 17:1542–1557.

Silva JC, Loreto ELS, Clark JB (2004) Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol 6:57–71.

Sinzelle L, Izsvák Z and Ivics Z (2009) Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cell Mol Life Sci 66:1073–1093.

Siomi MC, Sato K, Pezic D and Aravin AA (2011) PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 12:246–258.

Ungerer MC, Suzanne C and Zhen Y (2006) Genome expansion in three

hybrid sunflower species is associated with retrotransposon proliferation. Curr Biol 16:872–873.

Wallau GL, Ortiz MF and Loreto EL (2012) Horizontal Transposon Transfer in Eukarya: detection, bias and perspectives. Genome Biol Evol 4:689–699.

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

Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T and Kimura K (2001) Possible Horizontal Transfer of a Transposable Element from Host to Parasitoid. Mol Biol Evol 18:1952-1958.

APÊNDICE

Trabalho publicado na revista *Genome Biology and Evolution*[#]

Horizontal Transposon Transfer in Eukarya: Detection, Bias, and Perspectives

*[#]Gabriel Luz Wallau – Pós Graduation in Animal Biodiversity, Universidade Federal de Santa Maria – RS – Brazil; Av. Roraima nº 1000 Santa Maria - RS - CEP: 97105-900. gabriel.wallau@gmail.com

[#]Mauro Freitas Ortiz – Pós Graduation in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul – RS – Brazil; Av. Bento Gonçalves, 9500 - Prédio 43323M Porto Alegre – RS – CEP: 91501-970. maurobios@gmail.com

Elgion Lucio Silva Loreto – Biology Department - Universidade Federal de Santa Maria - RS – Brazil; Av. Roraima nº 1000 Santa Maria - RS - CEP: 97105-900. elgion.loreto@pq.cnpq.br

***Corresponding author**

[#] These two authors contributed equally to this work.

Este trabalho fez parte da tese de doutorado do aluno Gabriel da Luz Wallau junto ao Programa de Pós-Graduação em Biodiversidade Animal da UFSM.

GBE

Horizontal Transposon Transfer in Eukarya: Detection, Bias, and Perspectives

Gabriel Luz Wallau[1,*,†], Mauro Freitas Ortiz[2,†], and Elgion Lucio Silva Loreto[3]

[1]Programa de Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria, Brazil

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

[3]Departamento de Biologia, Universidade Federal de Santa Maria, Brazil

*Corresponding author: E-mail: gabriel.wallau@gmail.com.

[†]These authors contributed equally to this work.

Accepted: July 5, 2012

Abstract

The genetic similarity observed among species is normally attributed to the existence of a common ancestor. However, a growing body of evidence suggests that the exchange of genetic material is not limited to the transfer from parent to offspring but can also occur through horizontal transfer (HT). Transposable elements (TEs) are DNA fragments with an innate propensity for HT; they are mobile and possess parasitic characteristics that allow them to exist and proliferate within host genomes. However, horizontal transposon transfer (HTT) is not easily detected, primarily because the complex TE life cycle can generate phylogenetic patterns similar to those expected for HTT events. The increasingly large number of new genome projects, in all branches of life, has provided an unprecedented opportunity to evaluate the TE content and HTT events in these species, although a standardized method of HTT detection is required before trends in the HTT rates can be evaluated in a wide range of eukaryotic taxa and predictions about these events can be made. Thus, we propose a straightforward hypothesis test that can be used by TE specialists and nonspecialists alike to discriminate between HTT events and natural TE life cycle patterns. We also discuss several plausible explanations and predictions for the distribution and frequency of HTT and for the inherent biases of HTT detection. Finally, we discuss some of the methodological concerns for HTT detection that may result in the underestimation and overestimation of HTT rates during eukaryotic genome evolution.

Key words: horizontal transfer, horizontal transmission, transposable elements, genome, eukaryote evolution.

Introduction

Since the discovery of DNA as the molecule that stores genetic information and governs trait inheritance from parents to their offspring, no biologist doubts that the vertical transfer of genetic material between ancestral and extant species has occurred. However, there is now growing evidence suggesting that another process also promotes the sharing of genetic material among species: horizontal transfer (HT) (Keeling and Palmer 2008).

HT events are characterized by the exchange of genetic material between species by methods other than ancestral to descendant inheritance (Schaack et al. 2010). These events are quite common among bacterial species (Gogarten and Townsend 2005), and as a result, sets of bacterial species are now being called genetic exchange communities (Skippington and Ragan 2011). In multicellular eukaryotes, HT is thought to be a rare event (Kidwell 1993; Anderson 2005). However, a growing body of evidence suggests that a particular type of HT, horizontal transposon transfer (HTT), could be a widespread process during eukaryote evolution (Schaack et al. 2010).

Transposable elements (TEs) are prone to HT compared with other coding and noncoding DNA sequences because of their parasitic characteristics and their intrinsic capacity to mobilize and reintegrate into chromosomes (Schaack et al. 2010). HT is a key step in the TE life cycle, allowing these parasites to immigrate to and colonize new genomes and escape loss by genetic drift (Le Rouzic and Capy 2006; Venner et al. 2009; Hua-Van et al. 2011). The arrival of a new TE in a host genome can have detrimental consequences

because TE mobility may induce mutation. Moreover, transposition activity increases the TE copy number and generates chromosomal rearrangement hotspots (Cáceres et al. 2001; McVean 2010). However, HTT can also introduce new genetic material into a genome and promote the shuffling of genes and TE domains among hosts, which can be co-opted by the host genome to perform new functions (Pace et al. 2008; Thomas et al. 2010).

HTT is difficult to detect because it is necessary to consider all the intrinsic features of the TE life cycle, such as sequence degeneration, stochastic loss, and any different evolutionary rates (Cummings 1994; Capy et al. 1998). In addition, the same patterns found in HTT can be observed at various stages during the natural TE life cycle, or they can be generated by the hybridization of closely related species. Since HTT was first described, many authors have suggested different approaches to obtain evidence of these events (Loreto et al. 2008). These methodologies involve looking for phylogenetic incongruence (PI) between the host and TE phylogenies, patchy TE distributions (PD), or a high similarity (HS) between TEs from different species (Silva et al. 2004).

In the last decade, new methodological approaches based on comparisons between host genes (HGs) and TEs were developed, allowing a broader evaluation of HTT events (Silva and Kidwell 2000; Lerat et al. 2000). Nevertheless, the identification of HTT events can still be difficult, even when combinations of several methodologies are used, because these methods can both overestimate and underestimate the occurrence of HTT events depending on when and in which species the HTT occurred. The astonishing number of new genome projects, in all branches of life, presents an unprecedented challenge to the field of comparative genomics. The amazing quantity of genomic data that is now available for many taxa urgently calls for the development and application of standardized methodologies that will produce widely comparable results. To date, there is no gold-standard approach to clearly discern between alternative explanations and HTT events.

The main purpose of this article is to propose a standard hypothesis test for the evaluation of HTT events. We discuss the biological bias found in the distribution of the HTT events described in the literature and caution against methodological biases in regards to inferring the number of HTT events.

HTT Detection

Currently, the most robust approach for evaluating potential HTT events is a combination of evidence supported by statistical tests (Loreto et al. 2008). However, in some cases, only one type of evidence, such as HS, PD, or PI, is necessary to support HTT. For example, the classical and unequivocal uptake of *P*-elements by *Drosophila melanogaster* from *Drosophila willistoni* is supported by the PD of this element in the *melanogaster* species group, where it is present despite being absent from the genomes of related species (Daniels et al. 1990).

One of the most promising methodologies for the detection of HTT is based on a between-species comparison of the neutral rate of evolution (assessed by synonymous substitution divergence) for both the TEs and the HGs. This approach assumes that, if TEs have been vertically transmitted and maintained by neutral evolutionary processes in the genomes of two different species since their last common ancestor, the number of synonymous substitutions per synonymous site (dS) of the TEs should be equal to or greater than that of the vertically transmitted HGs. However, if the dS obtained for the TE is significantly lower than the dS for the vertically transferred HG, the most probable explanation is that these elements were exchanged by HT between the species after their reproductive isolation. This pattern can be observed because a horizontally transferred TE has spent less time in the new host genome than the original HGs. These HGs have been in the genome since the last common ancestor of the species involved in the HT. Therefore, these TEs have had less time to accumulate synonymous substitutions than the HGs. It is noteworthy to state that even if a TE shows a dS value equal to or greater than the HG dS, it does not necessarily imply vertical transmission (VT). This pattern can also be generated by an HTT event occurring just after the split of the involved species. For these comparisons, it is necessary to choose HGs with similar codon usages to those of the TEs (Silva and Kidwell 2000; Ludwig et al. 2008). If an HG with a higher codon usage bias is chosen, it can present low dS values and results in the underestimation of the number of HTT events (Silva and Kidwell 2000; Vidal et al. 2009).

Another interesting method for evaluating HTT involves the use of the unique codon usage bias of each genome (Lerat et al. 2002; Jia and Xue 2009; Plotkin and Kudla 2011). Differences in the codon usage bias are expected to be higher among genomes from different species than among the genes within the genome of the same species. According to this premise, it should be possible to detect the recent invasion of a genome by TEs from the patterns of codon usage bias because the TE's codon usage should be more homologous to that of the donor species than that of the receptor species. Recently, Rodelsperger and Sommer (2011) showed the utility of this methodology for detecting HTT events between a beetle species and its associated nematode. It is noteworthy that the species–specific codon usage bias becomes less evident when more closely related species are considered because of their phylogenetic similarity (Sharp et al. 1995). Thus, although this methodology can be very useful in detecting HTT between distantly related species, there are limitations to its application in related species.

Multiple hypothesis testing using several methodologies could be an efficient approach for discriminating between HTT and alternative hypotheses. On the basis of recent reports

on TE characteristics and HTT events, we propose a straightforward hypothesis test to evaluate potential HTT events (fig. 1).

Hypotheses Test

Normally, the first sign of evidence to suggest HTT comes from PI, a PD, or a HS between the TEs from distantly related host species. PI is inferred if a phylogeny of TE does not match the host phylogeny (fig. 1). PD is detected when a specific TE shows a random distribution, characterized by the presence TE in one or a few species from a phylogenetic branch that otherwise lacks the TE. However, although these patterns can be generated by HTT events, they can also be the result of the natural degeneration of TEs inside the host genomes, when combined with ancestral polymorphism and stochastic loss.

The first step of the hypothesis test is the implementation of two different tests (T1 in fig. 1): 1) a comparison of the dS between the TEs and HGs and 2) a comparison of the codon usage bias between the TEs and the host genome. These two tests can be complementary in HTT detection. The codon usage bias comparison can be used to evaluate HTT in distantly related species; however, the difference in the codon usage bias among closely related species is normally low, which does not allow the donor (and TE) and the recipient species codon usages to be distinguished. TE and HG dS comparisons can be used to evaluate HTT in closely related species and in distantly related species alike. If we find that the codon usage bias is similar between the TEs and the host genome and that the TE dS values are equal to or greater compared with the HGs, then it is likely that the TEs are being inherited by vertical transfer (fig. 1*B*). Otherwise, if the TE's codon usage bias is different from that of the host genome or if the dS is significantly lower for the TEs than for the nuclear HGs, then the TEs were most likely exchanged among the species by HT (fig. 1*A*). It is necessary to perform dS and codon usage bias comparisons even if PI or PD were not detected because the absence of these evidences does not guarantee that an HTT event has not occurred.

Nevertheless, alternative hypotheses attempting to explain the observed differences in the dS values between TEs and HGs have also been suggested. For example, selective constraint can act at the RNA/DNA level as a pressure established on the mRNA structural stability or on splicing sites or if a TE is integral in the siRNA regulatory machinery (Rubinstein et al. 2011; Plotkin and Kudla 2011). However, as these constraints are expected to act on specific sites and not on the sequence as a whole, the magnitude of these constraints should be small. Therefore, these factors cannot explain the dS differences observed between HGs and TEs when the TE is conserved across the entire sequence. In fact, sometimes the dS values between TEs from different species are very low, the magnitude of which could not be easily explained by the previously described constraints. Therefore, a very low dS measurement is better explained by the occurrence of an HTT.

When HTT events among distantly related species are considered, only the T1 stage of the hypothesis testing is necessary for validation. However, HTT events can occur among

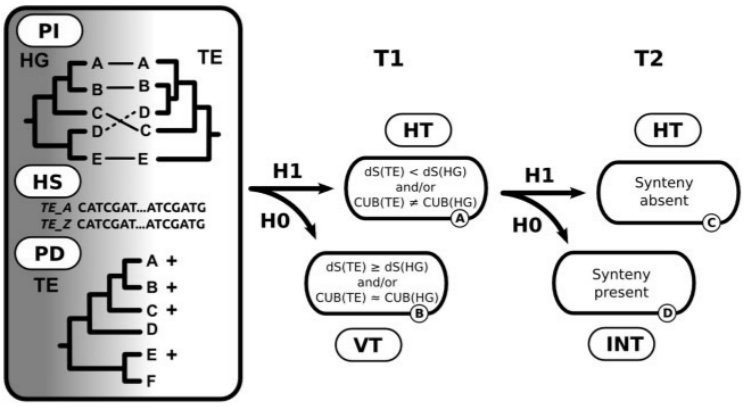

FIG. 1.—A schematic representation of a hypothesis test for discerning between HT and the natural stages of the TE life cycle. BOX: The first line of evidence for HTT: Phylogenetic incongruence (PI) between the host and TE phylogenies. Patchy distribution (PD) of a given TE within a group of species and high similarity (HS) between the TEs from different species. T1—The first test to distinguish between HTT and vertical transmission (VT)—comparing the dS between the TE and host genes (HGs) and species-specific codon usage bias (CUB). H0—vertical transfer is more probable if the dS values for the TEs are greater than or equal to the dS values of the vertically transmitted host genes and if the TE codon usage bias is similar to the codon usage bias in the host species. H1—HT will be selected if the TE's dS value is significantly lower than the dS values of the vertically transmitted host genes or if the TE codon usage bias is different from the host species codon usage bias. T2—A second step can be used to evaluate HTT between closely related species. H0—If there is synteny beyond the border of the TE copies, it is more probable that these copies were shared by hybridization among the host species (an introgression [INT] occurred). H1—If there is no synteny, it is more probable that these copies were shared by an HTT event between host species.

individuals encompassing any taxonomic level, from different phyla to closely related species (Bartolomé et al. 2009). It is very difficult to prove HTT among closely related species, and in this case, the sharing of TEs between species can be the result of the occasional cross-fertilization between species. Introgression events between closely related species can generate significantly lower dS values for the TEs compared with the nonintrogressed HGs. The analysis of synteny beyond the border of the TE copies, that is, analysis all the TE copies present in one species, and an evaluation of whether they are found at the same locus in another species is one method that has been suggested to discern between introgression and HTT (Fortune et al. 2008) (T2 in fig. 1). Introgression events normally maintain synteny among the species involved in the hybridization; in other words, homology and high identity are encountered not only in the TE sequences but also in the neighboring DNA regions (fig. 1*D*). However, when HTT events occur, only variability with the absence of synteny is typically encountered in the TE-neighboring regions (Fortune et al. 2008) (fig. 1*C*). Nevertheless, despite the fact that this methodology is consistent and straightforward, it has yet to be tested, and it could proved to be particularly difficult to evaluate the synteny of TEs because of their inherent mobility. It is likely that this methodology will be restricted to the analysis of nonautonomous TEs, but even nonautonomous elements can be mobilized by other TEs in trans, a factor that would complicate the analysis. Regardless, if synteny is found, it is taken as evidence that hybridization occurred; therefore, in the absence of synteny, the probability that the sharing of TEs between species as the result of hybridization decreases, whereas the probability of an HTT event increases.

High similarities between the TE sequences in different species can also be the result of TE domestication, where a TE region is co-opted to perform a new, useful function in the genome of the host (Gould and Vrba 1982; Huda et al. 2010). Domestication can be detected using features such as copy number, orthologous position, and evaluating the selective constraint (dN:dS ratio) acting between the TEs that are incongruent with the host species' phylogeny and comparing this constraint with the selective constraint on the HGs. Thereby, we can discern whether the HS found between the TEs from different host species is due to domestication events, different evolutionary rates, or ancestral polymorphism. Other analyses can also reveal clues as to whether a TE is domesticated, such as the presence of only one TE copy in the genome or the observation that the TEs occur at orthologous positions in different species (Sinzelle et al. 2009). Another approach that can be used to gather clues about TE domestication is the analysis of full-length TE copies (including inverted terminal repeats, long terminal repeats, and coding and noncoding regions). If there are high similarities along all the TE sequences, the best explanation for the sequence conservation is the occurrence of an HTT event. This is because TE domestication only imposes strong selective constraints on one region of a TE and not in the full-length copies (Feschotte 2008; Sinzelle et al. 2009). Even if a domestication occurred, a dS TE smaller than dS HG is unlikely to be observed, because negative selection acts only in nonsynonymous substitutions and not over neutral synonymous substitutions. Therefore, we also can evaluate if occurred HT events before the domestication event using the dS analysis (T1 in fig. 1).

Another analysis that can be useful for understanding HTT is the dating of these events along the molecular clock. One way to perform this analysis is by evaluating the molecular evolution rate of the nuclear genes with a codon usage bias similar to the TE to estimate the time of divergence between horizontally transferred TEs copies (Ludwig et al. 2008). A second type of analysis can be performed when the entire host genome is available. In this case, an ancestral sequence can be inferred when evaluating many copies of one horizontally transferred element. This analysis is based on the premise that these elements have been evolving neutrally since the HT event; therefore, we can estimate the time of the first insertion event and the subsequent amplification inside of the host genome using a neutral substitution rate (Mouse Genome Sequencing Consortium 2002; Yang et al. 2004; Khan et al. 2006; Pace and Feschotte 2007). This neutral substitution rate can be estimated from an ancestral TE present in an orthologous position (inherited vertically) in genomes where we have an estimate of the host species' divergence time (Pace et al. 2008). Therefore, with these type of data, we can evaluate whether a TE is more recent than expected for vertical transfer, and by comparing this activity estimative among different species, we also can reveal relationships between the donor and the receptor species.

HTT Distribution and Frequency

HTT Rates

Here, we analyze the HTT events previously collected from the literature by Schaack et al. (2010) along with new events to compile all the HTT events described to date (supplementary table 1, Supplementary Material online). HTT events have already been detected in three eukaryote kingdoms: Animalia, Fungi, and Plantae (fig. 2). The majority (94.37%) of the HTT events were detected in Animalia, followed by Plantae (4.30%) and Fungi (1.32%). The differences in the HTT frequencies among kingdoms may be explained by differential susceptibilities of taxas to experiencing HTT. However, these differences could also be due to a historical bias for the use of animal model organisms in TE research or the differential abilities of the studied TEs to undergo HT (Pritham 2009; Schaack et al. 2010). To date, 178 of the 330 HTT cases described in the literature were detected among *Drosophila* species (54%). This disproportionate number of HTTs in *Drosophila* could be biased because some of the pioneering studies in TEs,

including the first well-documented case of HTT (Daniels et al. 1984), were performed in these model organisms. Thus, these studies opened the door for TE research using the *Drosophila* genus. Several recent publications have shown evidence of HTT events in other Animalia taxa, such as crustaceans and mammals (Casse et al. 2006; Gilbert et al. 2010; Novick et al. 2010), further suggesting that the elevated number of HTT events described in *Drosophila* may show a historical bias.

Genome Projects, TE, and HTT Bias

Although exponentially growing, global species biodiversity is still poorly represented in current genome projects. In Eukarya, only the Animalia (270 projects), Fungi (234 projects), and Plantae (101 projects) kingdoms have a large number of genome projects (http://www.ncbi.nlm.nih.gov [cited 2011 October 12]). Many of these genomes are still undergoing sequencing or are in other steps of analysis; thus, we have differing knowledge about the TE content in these genomes (fig. 2). Moreover, many studies remove these elements to facilitate genome assembly or analysis (Bergman and Quesneville 2007; Treangen and Salzberg 2011). The lack of knowledge about the TE content in some taxa could strongly bias the descriptions of HTT distribution and frequency.

To evaluate how genomic analysis can influence the TE and HTT descriptions, we collected, for each of the aforementioned kingdoms, the number of genome projects in NCBI and the TE entries from the Repbase site (http://www.girinst.org [cited 2011 October 12]; Kohany et al. 2006) (fig. 2*A*). For this evaluation, two points should be noted: 1) the genome projects are in different stages and many have not yet analyzed the TE content and 2) the entries in Repbase are not limited to the TEs from genome projects.

Most of the HTT events described in the literature were from Animalia (fig. 2*C* and *D*). This finding likely reflects the larger number of genome projects for animals. Moreover, on the basis of TE entries available in Repbase for different taxa, we noted that animal species have been analyzed more deeply in regards to their TE content compared with the other phyla (12,565 TE entries) (fig. 2*A*).

The Plantae kingdom is an intriguing case; some species have high TE content (more than 60% in maize; Biémont and Vieira 2006), and a large number of elements have been characterized (4,638 TEs entries Repbase); however, only 13 HTT events have been detected in this kingdom (fig. 2*C*). This discrepancy could be explained by the following: 1) the smaller number of genome projects in Plantae compared with the Animalia and Fungi kingdoms; 2) some unknown, specific features of these organisms; or 3) historical bias in the HTT analysis, despite TE characterization.

In fungi, there is no apparent bias due to the number of genomes available as there are a similar number of projects when compared with Animalia (fig. 2*A*); however, to date, only four HTT events have been described for fungi (fig. 2*C*). One possible explanation for this fact could be related to the N_e (effective population size) of these organisms because they have among the largest eukaryotic N_e (Lynch and Conery 2003). It has been shown that there is a negative correlation between the N_e and TE maintenance in host genomes (Lynch and Conery 2003). Moreover, fungi present a low, and most likely poorly studied, TE content (1,603 TEs in Repbase) compared with animals (12,565) or plants (4,638) (fig. 2*A*). It is important to note that the existence of only a few described HTT events in fungi does not mean that HTT does not occur; it more likely indicates that HTT occurs but cannot be detected due to the high turnover of TEs in species with large N_e values and small genomes. However, this is not always the case. *D. melanogaster*, for example, has a small N_e compared with most fungi species but has a high turnover for retrotransposons and a high rate of HTT (Lerat et al. 2003).

Excavates, Chromalveolates, and Rhizaria are the least represented of the kingdoms in the NCBI genome projects database, and they also have fewer entries in the Repbase repository (fig. 2*A* and *C*). The lack of knowledge about the TE content in these groups, along with the high turnover of TEs in taxa with large N_e values, may explain why there have been no HTT cases reported for these groups thus far.

TE Features Influencing HTT Frequency

Despite historical bias in the evaluation of HTT among taxa, we can observe patterns in HTT distribution and frequency that are associated with different TE features. Silva et al. (2004) suggested that an effective HTT event may be related to the presence of a stable intermediate during the transposition process. Moreover, TE self-regulatory mechanisms can also influence the success of certain HTT events. HTT events appear to be more frequent for LTR retrotransposons and DNA transposons when compared with non-LTR retrotransposons (Silva et al. 2004; Loreto et al. 2008; Schaack et al. 2010).

The evolutionary relationship between LTR retrotransposons and retroviruses is well established (Xiong and Eickbush 1988, 1990; Poch et al. 1989). This evolutionary link suggests that some LTR retrotransposons can undergo HTT by themselves if they are capable of producing viral capsids and envelopes (*env* gene), hence promoting a viral-like infection and thereby eliminating the requirement for a vector. It has been shown that *gypsy* elements are capable of producing viral capsids and infecting *gypsy*-free *D. melanogaster* strains (Kim et al. 1994; Song et al. 1994). Even LTR retrotransposons that lack the *env* gene and the gene responsible for producing viral capsids can use the viral capsids from other LTR retrotransposons in trans, allowing a "helped" infection (Coffin et al. 1997). Recently, Routh et al. (2012) showed that at least 5.3% of the RNAs packaged inside of viral-like particles contain sequences derived from TEs, including DNA transposons, LTR and non-LTR retrotransposons. However, the capacity of *gypsy* viral capsids to infect other *Drosophila* species still

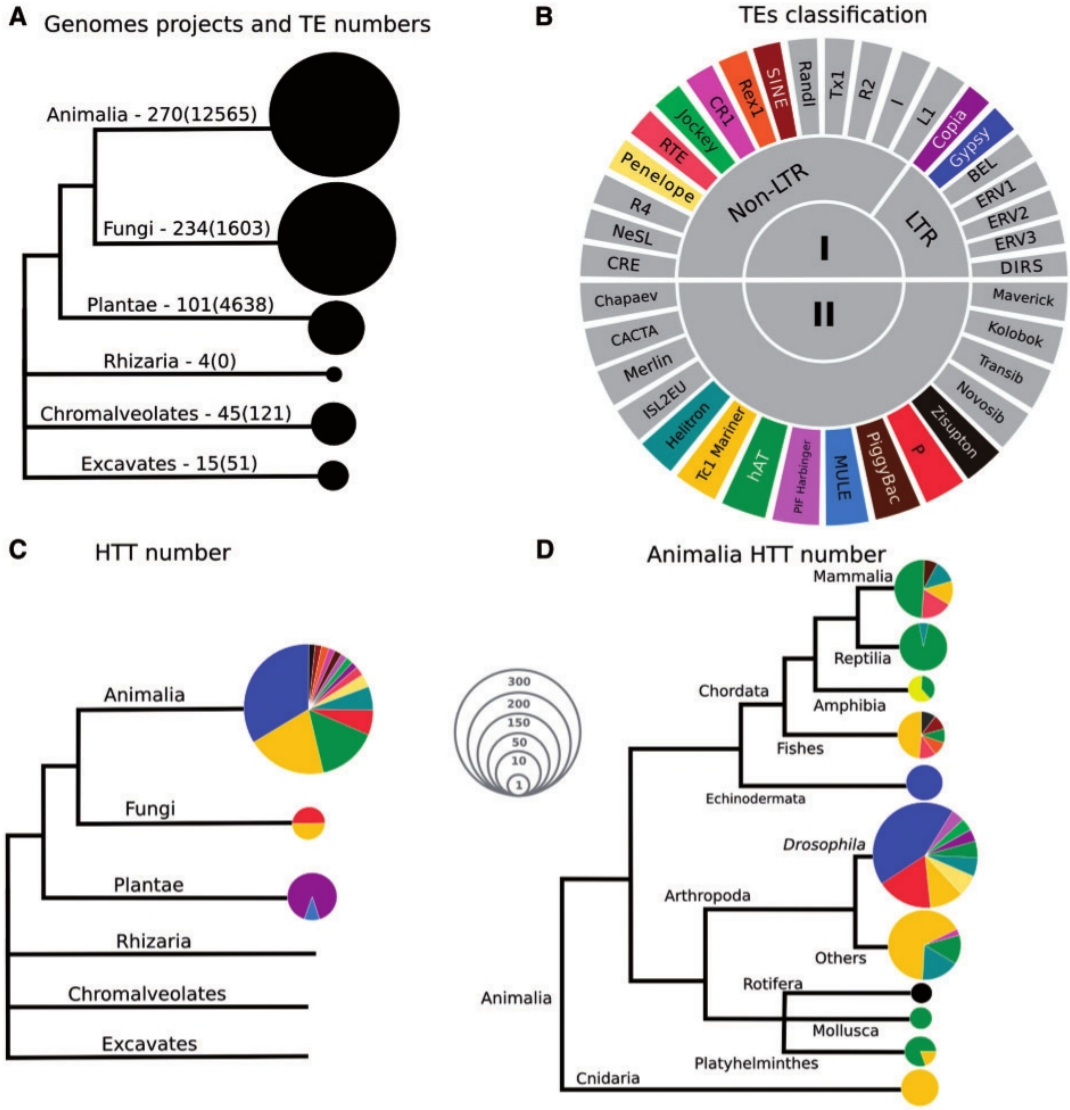

Fig. 2.—A representation of the genome projects, TEs, and number of HTT events in each major eukaryotic taxon. (*A*) The number of genome projects from the NCBI database (corresponding to cycle size and the number after the branch name, respectively) and TE RepBase entries (indicated by the number within the parentheses) in each major branch of the tree of life. (*B*) TE superfamily classifications based on RepBase. (*C*) The distribution of HTT events in each major eukaryotic taxon. (*D*) Distribution of HTT events within Animalia. The colors represent the TE superfamilies described in (*B*), and the cycle size represents the number of HTT events for each host taxon.

remains unclear and requires further elucidation. The same holds true for the in trans infection hypothesis.

If we suppose that the infective ability of LTR retrotransposons plays a significant role in promoting HTT events among species, we should expect that LTR retrotransposons would be preferentially transferred among species with cell structures that are similarly recognized by the LTR retrotransposon's capsids. This assumption is based on the premise that a retrotransposon's recognition machinery is analogous to that of a virus, which recognizes a restricted set of cellular receptors from a particular group of species. Furthermore, this analogy allows the extrapolation that HTT events should occur in waves, similar to those in a viral infection. When we look at all the previously described examples of HTT involving LTR retrotransposons, we note that 88.88% (104 of 117) of the events are among species from the same genus, 3.41% (4 of 117)

occur among species from different genera, 6.83% (8 of 117) occur among species from different orders, and only one event was observed among species from different phyla. However, the tendency for a higher frequency of LTR retrotransposon HTT events among closely related species could represent a strong taxonomic bias because 74 of the 86 described HTT events involving retrotransposons were described in the *Drosophila*. Regarding the HTT waves, one study reported that retrotransposon HTT waves occurred among *Drosophila* species (de Setta et al. 2009). Because of this, future studies are required to evaluate whether the LTR retrotransposon HTT events also occur more frequently among closely related species in other taxa.

The most widely distributed DNA transposon elements, from the *Tc1-mariner* superfamily, are simple in structure (presenting one or only a few ORFs and a primary structure rarely longer than 4 kb) (Wicker et al. 2007) and possess self-regulatory mechanisms. This structural simplicity can increase the likelihood of stable vector transportation during an HT event and is thought to represent an adaptation for HTTs (Schaack et al. 2010). O'Brochta et al. (2009) observed that *hobo/Hermes hAT* elements commonly produce stable and recombinogenic episomes; the circular extrachromosomal DNA of these transposons is a stable excision product that can reintegrate at a new site. Therefore, it is possible that these episomes could maintain TE recombinogenic properties following the transport by a vector into another species.

In addition to being carried by vectors, we cannot rule out the possibility that some DNA transposons may be self-transmissible. It is important to note that some complex DNA transposons may have originated from virophages (Fischer and Suttle 2011) and single-stranded DNA viruses (Liu et al. 2011).

As mentioned previously, some LTR retrotransposons are able to produce virus-like particles, and this has been suggested as a mechanism for HTT. Based on the common features shared between some TEs and viruses, one would assume that if the infective capacity is an important step in the HTT of LTR retrotransposons, then jumps by viral species between host species are expected to be common. In fact, there are many works reporting the jumping of viral species between host species. These events are commonly called viral-host switches (Gibbs and Weiller 1999; Nemirov et al. 2002; Vijaykrishna et al. 2007; Kang et al. 2010; Liu et al. 2010, 2011; Longdon et al. 2011) or species jumps. Viral-host switches have been primarily described in vertebrate species, the majority of which are related to human infectious diseases such as HIV, SARS, and H5N1 (Woolhouse et al. 2005; Parrish et al. 2008). Thus far, little importance has been given to these events in other taxa; however, some examples of viral species jumps have recently been described in *Drosophila* species (Liu et al. 2010). Altogether, these viral and TE data suggest that the infective capacity of some TEs is likely a key step that allows their horizontal transmission across species.

Once a host-switch event occurs, some virus integrates into the germ cells of the host genome as a provirus by a process known as endogenizaton [retroviruses—see Patel et al. (2011) and Feschotte and Gilbert (2012)]. These proviruses can be maintained from parent to offspring by VT. Even viruses that do not have a natural or obligate integration step into the host genome can also be endogenized using the LTR retrotransposon and endogenous retroviruses machinery (Holmes 2011; Patel et al. 2011). Currently, the studies have shown that the majority of eukaryotic viruses can be integrated in the host chromosomes via different pathways. Therefore, once a virus is endogenized, the host switch can be detected by analyzing the same evidence used to detect HTT, such as our hypothesis test.

The discussion of self-regulatory mechanisms yields support for the high efficiency of *mariner* elements to perform an efficient invasion strategy in a new genome (Lohe et al. 1995). Under excessive transposase production, *mariner* transposases aggregate together, causing a decrease in the transposition rate (Hartl et al. 1997; Lohe et al. 1997). When an organism acquires a new active transposon by HTT, a burst of transposition events typically follows, until all copies are mutationally inactivated or regulated by the inner host regulation mechanisms (Boer et al. 2007). High TE activity may cause detrimental changes in the host genome with TE insertions in coding or regulatory gene sequences. Self-regulatory mechanisms can be advantageous to TEs because these mechanisms can decrease the probability that a detrimental mutation will be introduced into the host genome, thereby increasing the TE's odds for inheritance by the host's descendants. Thus, TEs with self-regulatory mechanisms appear to have evolved a more effective strategy for an efficient invasion and for being maintained in host descendants reducing their harmful effects for the new host's genome.

Several characteristics of DNA transposons, such as autonomous transposition capacity (independent of the host's proteins), short length, and the presence of self-regulatory mechanisms, can enhance the probability that these elements will undergo HT. To date, only one event has been reported to have occurred among domains; 25.49% (39 of 153) of the HTT events occurred among phyla, 3.26% (5 of 153) occurred among classes, 17.54% (27 of 153) occurred among orders, 11.76% (18 of 153) occurred among families, 3.26% (5 of 153) occurred among genera, and 37.90% (58 of 153) occurred among species of the same genus. These data suggest that HTT involving DNA transposons can occur in all taxonomic levels, including the more distantly related levels.

Host Features Influencing HTT Frequency

Intrinsic host features can also influence HTT rates. The frequency of HTT events can be influenced by factors such as the

natural history or life cycle of the host species. For example, if two species have a close ecological relationship, such as a predator–prey relationship, symbiotic contact, the sharing of parasites, or even the use of the same natural resources, the chances that an HTT event will occur between these species increases (Houck et al. 1991; Yoshiyama et al. 2001; Loreto et al. 2008; Gilbert et al. 2010; Schaack et al. 2010). This scenario has been used to explain cases of HTT among sympatric crustaceans (Casse et al. 2006) and in *Drosophila* species (Mota et al. 2010; Carareto et al. 2011).

In the majority of multicellular eukaryotes, the reproductive and somatic cells are differentiated. Therefore, the TEs must be transmitted to the reproductive cells to be inherited by the descendants of a new host, that is, to gain entry into a new host genome via HTT. Thus, we might expect that HTTs should be more prevalent in unicellular eukaryotes and multicellular eukaryotes with undifferentiated reproductive and somatic cells because any cell in the body that has acquired a new TE can transmit it to future generations. Along these lines, Pritham (2009) suggested that unicellular eukaryotes should be particularly susceptible to HT due to the lack of a protected germline. Supporting these ideas, Robertson (1997) identified seven putative HTT events between insects and *Hydra* and one HTT case between the planarian *Dugesia tigrina* and the ant *Crematogaster cerasi*. In line with these findings, Chapman et al. (2010) more recently identified at least 90 potential HTT events in the *Hydra magnipapillata* genome. Hydras and planarians are animals without germ and somatic cell differentiation. Nevertheless, despite the difficulty imposed by cellular segregation, almost all HTT cases have been described in multicellular eukaryotes that have reproductive and somatic cell differentiation.

The mutation rate ($N_e\mu$) and the N_e of a receptor–host species can also influence the probability of a successful TE invasion by HT. For example, successful TE invasions by HT are less likely in species with a higher $N_e\mu$ and shorter generation times because there is an increased probability of the TE being inactivated. The influence of the N_e is a result of the balance between natural selection and genetic drift (Lynch and Conery 2003). The genomes of host species with large population sizes, as many unicellular organisms possess, are also subject to a strong purifying selection (Lynch 2007). Thus, if an HTT event occurs in these species, the probability that the TEs will be quickly eliminated by natural selection is high. On the other hand, in species with small population sizes, such as many tetrapods, genetic drift increases the probability that a new TE will be maintained in the host genome following an HTT event. Lynch and Coney (2003) reported that host species should have an N_e less than approximately equal to 7×10^7 to allow retrotransposon proliferation and an N_e less than approximately equal to 2×10^7 to allow the proliferation of DNA transposons.

As expected, each species set has unique ecological interactions (among species and among their parasites) leading to differential probabilities of HTT. However, it seems likely that there are some patterns that will be useful for predicting HTT in a broad range of species due to host reproduction and population features.

HTT Underestimation and Overestimation

Even when all the available approaches for detecting HTT are used, it is likely that many events will remain undetected. The inability to detect HTT results from the high turnover of TEs in host genomes (Lerat et al. 2003). When a TE arrives in a new genome, it usually occurs through a transposition burst that can be detrimental to the new host. The individuals bearing these detrimental changes can then be eliminated by natural selection, hence abolishing the signal of the primary invasion. When TEs successfully invade and are maintained in a new genome, the TE copies will evolve under neutral or weak natural selection (Silva and Kidwell 2000). Both low-dS measures obtained from the TEs compared with the HGs and species-specific codon usage biases from donor species tend to degenerate over the course of time. Thus, the more ancient an HTT event, the more difficult it will be to detect (fig. 3*A*). This promotes a weak signal of HTT events, leading to underestimation (fig. 3*B*).

Alternatively, HTT can also be overestimated. The overestimation of HTT events is directly related to the number of species in both the donor and the receptor clades. For example, if HTT occurs in the ancestor of two clades (fig. 3*A*–*C*), comparisons of the dS TE/dS HG and the codon usage bias could be significant for all pairwise species comparisons, suggesting many HTT events, when in reality only one has occurred. The maximum number of overestimated HTT events will be the number of analyzed species derived from the donor clade since the last common ancestor, multiplied by the number of analyzed species derived from the receptor clade since the HTT event.

A more complex scenario can also be considered. For example, members of a family of related TEs could have undergone HTT at different evolutionary times (fig. 3*D*). In this situation, overestimation will occur if we count each pairwise comparison resulting in a dS TE $<$ dS HG as one event, as mentioned earlier. However, in the scenario depicted in figure 3*D*, for example, if we consider the observed cases as a unique, ancient HTT events, we will obtain an underestimate because three independent HTT events have occurred. In some specific cases, however, dS values can be used to date these HTT events (fig. 3*E*). For example, HTT events may be dated when the time since the occurrence of the HTT is long enough to result in differentiated in dS values, when the studied species have a well-resolved phylogeny and when a calibrated molecular clock is available. The number of dates obtained in these analyses may then be used to parsimoniously estimate the number of HTT cases.

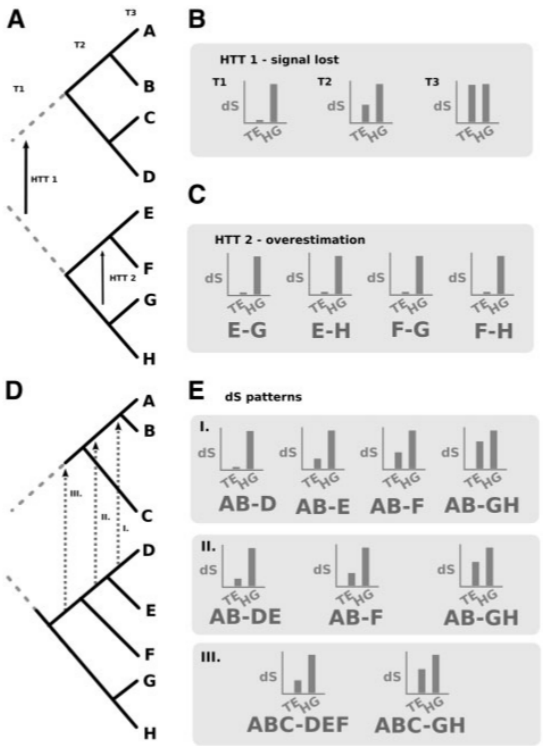

Fig. 3.—The underestimation and overestimation of HTT events. (*A*) The host species' phylogenies that represent HTTs at different evolutionary times (HTT1 and HTT2). (*B*) In HTT1, as the TEs evolve under neutral- or weak-natural selection, the dS value will increase over time (T1–T2–T3), and the species-specific codon usage bias from the donor species will be lost (resulting in the underestimation of old events). (*C*) In HTT2, all the TE-dS comparisons among species E, F, G, and H will be significantly lower than the HG's dS due to the maintenance of only one HT signal. (*D*) The host species' phylogenies that represent a more complex scenario with three HTT events (I, II, and III). (*E*) The TE dS patterns resulting from the HTT events in (*D*).

These theoretical models are simplistic compared with the complex evolution of TEs, where HTT is common. However, the use of these models can allow us to describe the degree of overestimation for a given situation. Moreover, because we observe a number of significantly lower dS values in the potential HTT events among current species, we may propose, by parsimony, the probable donor and receptor species by identifying the lower dS value.

Conclusions

Currently, there are no doubts as to the impact of TEs on eukaryotic genome evolution. There is a growing amount of data showing that HTT is a common and widespread phenomenon in eukaryote evolution. In light of the currently astonishing number of new eukaryotic genomes, it has become necessary to use a standardized methodology for the detection of HTT if these analyses are to be comparable across a wide range of eukaryotic taxa. Currently, different software is available to perform the analyses proposed in the hypothesis test (fig. 1), although one major challenge is to automate the data mining in the genomes to perform the analyses and organize the programs in a pipeline. This process can then facilitate and increase the discovery of HTT cases.

A strong HTT bias can be observed among eukaryotic taxa, primarily resulting from a historical bias for TE research in the *Drosophila* genus. However, even with this bias, we can observe trends that might be explained by the biological features of TEs and their hosts. HTT detection is a difficult task because of the high turnover of TEs inside host genomes and the number of species analyzed. These issues can lead to the underestimation or overestimation of HTT events between ancestral and current eukaryotic species; therefore, careful evaluation is warranted.

Supplementary Material

Supplementary table 1 is available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Dr Lizandra J. Robe, Dr Daniel Ângelo Sganzerla Graichen, and the two anonymous reviewers for valuable suggestions. This work was supported by Conselho Nacional de Desenvolvimento Científica e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Literature Cited

Anderson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62:1182–1197.

Bartolomé C, Bello X, Maside X. 2009. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. Genome Biol. 10:R22.

Bergman CM, Quesneville H. 2007. Discovering and detecting transposable elements in genome sequences. Brief Bioinform. 8:382–392.

Biémont C, Vieira C. 2006. Junk DNA as an evolutionary force. Nature 443: 521–524.

Boer J de, Yazawa R, Davidson W, Koop B. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics. 8:422.

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

Capy P, Bazin C, Higuet D, Langin T. 1998. Dynamics and evolution of transposable elements. Austin (TX): Landes Bioscience.

Carareto C. 2011. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. Mob Genet Elem. 1:179–186.

Casse N, et al. 2006. Species sympatry and horizontal transfers of *Mariner* transposons in marine crustacean genomes. Mol Phylogenet Evol. 40: 609–619.

Chapman JA, et al. 2010. The dynamic genome of *Hydra*. Nature 464: 592–596.
Coffin JM, Hughes SH, Varmus HE. 1997. Retroviruses. New York: Cold Spring Harbor Laboratory Press. p. 843.
Cummings MP. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol Evol. 9:141–145.
Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124:339–355.
Daniels SB, Strausbaugh LD, Ehrman L, Armstrong R. 1984. Sequences homologous to *P* elements occur in *Drosophila paulistorum*. Proc Natl Acad Sci U S A. 81:6794–6797.
de Setta N, Van Sluys MA, Capy P, Carareto CMA. 2009. Multiple invasions of *Gypsy* and *Micropia* retroelements in genus *Zaprionus* and *melanogaster* subgroup of the genus *Drosophila*. BMC Evol Biol. 9: 279.
Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 9:397–405.
Feschotte C, Gilbert C. 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 13: 283–296.
Fischer MG, Suttle CA. 2011. A virophage at the origin of large DNA transposons. Science 332:231–234.
Fortune PM, Roulin A, Panaud O. 2008. Horizontal transfer of transposable elements in plants. Commun Integr Biol. 1:74–77.
Gibbs MJ, Weiller GF. 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. Proc Natl Acad Sci U S A. 96:8022–8027.
Gilbert C, Schaack S, Pace JK II, Brindley PJ, Feschotte C. 2010. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature 464:1347–1350.
Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. Nat Rev Microbiol. 3:679–687.
Gould SJ, Vrba ES. 1982. Exaptation—a missing term in the science of form. Paleobiology 8:4–15.
Hartl D, Lohe A, Lozovskaya ER. 1997. Regulation of the transposable element mariner. Genetica 100:177–184.
Holmes EC. 2011. The evolution of endogenous viral elements. Cell Host Microbe. 10:368–377.
Houck MA, Clark JB, Peterson KR, Kidwell MG. 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. Science 253:1125–1129.
Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. 2011. The struggle for life of the genome's selfish architects. Biol Direct. 6:19.
Huda A, Mariño-Ramírez L, Jordan IK. 2010. Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. Mobile DNA. 1:2.
Jia J, Xue Q. 2009. Codon usage biases of transposable elements and host nuclear genes in *Arabidopsis thaliana* and *Oryza sativa*. Genomics Proetomics Bioinforma. 7:175–184.
Kang HJ, et al. 2010. Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (*Neurotrichus gibbsii*). Virology 388:8–14.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9:605–618.
Khan H, Smit A, Boissinot S. 2006. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 16:78–87.
Kidwell MG. 1993. Lateral transfer in natural populations of eukaryotes. Annu Rev Genet. 27:235–256.
Kim A, et al. 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 91:1285–1289.
Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 7:474.
Le Rouzic A, Capy P. 2006. Population genetics models of competition between transposable element subfamilies. Genetics 174:785–793.
Lerat E, Biemont C, Capy P. 2000. Codon usage and the origin of *P* elements. Mol Biol Evol. 17:467–468.
Lerat E, Capy P, Biémont C. 2002. Codon usage by transposable elements and their host genes among five species. J Mol Evol. 54:625–637.
Lerat E, Rizzon C, Biémont C. 2003. Sequence divergence within transposable element families in the *Drosophila melanogaster* genome. Genome Res. 13:1889–1896.
Liu H, et al. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J Virol. 84:11876–1187.
Liu H, et al. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol Biol. 11:276.
Lohe A, Hartl DL, Lozovskaya ER. 1997. Modern thoughts on an ancient marinere: function, evolution, regulation. Annu Rev Genet. 31: 337–358.
Lohe A, Moriyama E, Lidholm D, Hartl D. 1995. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. Mol Biol Evol. 12:62–72.
Longdon B, et al. 2011. Host-switching by a vertically transmitted rhabdovirus in *Drosophila*. Biol Lett. 7:747–750.
Loreto ELS, Carareto C, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity 100:545–554.
Ludwig A, Valente VLS, Loreto ELS. 2008. Multiple invasions of *Errantivirus* in the genus *Drosophila*. Insect Mol Biol. 17:113–124.
Lynch M. 2007. The origin of genome architecture. Sunderland (MA): Sinauer Associates, p. 494.
Lynch M, Conery JS. 2003. The origins of genome complexity. Science 302:1401–1404.
McVean G. 2010. What drives recombination hotspots to repeat DNA in humans? Philos Trans R Soc Lond B Biol Sci. 365:1213–1218.
Mota NR, Ludwig A, Silva Valente VL, Loreto ELS. 2010. Harrow: new *Drosophila hAT* transposons involved in horizontal transfer. Insect Mol Biol. 19:217–228.
Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520–562.
Nemirov K, Henttonen H, Vaheri A, Plyusnin A. 2002. Phylogenetic evidence for host switching in the evolution of hantaviruses carried by Apodemus mice. Virus Res. 90:207–215.
Novick P, Smith J, Ray D, Boissinot S. 2010. Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. Gene 449:85–94.
O'Brochta D, et al. 2009. Transpositionally active episomal *hAT* elements. BMC Mol Biol. 10:108.
Pace JK, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 17:422–432.
Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci U S A. 105:17023–17028.
Parrish CR, et al. 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol R. 72:457–470.
Patel MR, Emerman M, Malik HS. 2011. Paleovirology—ghosts and gifts of viruses past. Curr Opin Virol. 1:304–309.
Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat Rev Genet. 12:32–42.
Poch O, Sauvaget I, Delarue M, Tordo N. 1989. Identification of four conserved motifs among the RNA dependent polymerase encoding elements. EMBO J. 8:3867–3874.

Pritham EJ. 2009. Transposable elements and factors influencing their success in eukaryotes. J Hered. 100:648–655.
Robertson HM. 1997. Multiple mariner transposons in flatworms and hydras are related to those of insects. J Hered. 88:195–201.
Rodelsperger C, Sommer RJ. 2011. Computational archaeology of the *Pristionchus pacificus* genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol. 11:239.
Routh A, Domitrovic T, Johnson JE. 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U S A. 109:1907–1912.
Rubinstein ND, Doron-Faigenboim A, Mayrose I, Pupko T. 2011. Evolutionary models accounting for layers of selection in protein-coding genes and their impact on the inference of positive selection. Mol Biol Evol. 28:3297–3308.
Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 25:537–546.
Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA sequence evolution: the sounds of silence. Philos Trans R Soc Lond B. 349: 241–247.
Skippington E, Ragan MA. 2011. Lateral genetic transfer and the construction of genetic exchange communities. FEMS Microbiol Rev. 35: 707–735.
Silva JC, Kidwell MG. 2000. Horizontal transfer and selection in the evolution of P elements. Mol Biol Evol. 17:1542–1557.
Silva JC, Loreto ELS, Clark JB. 2004. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 6:57–71.
Sinzelle L, Izsvák Z, Ivics Z. 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci. 66:1073–1093.
Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. 1994. An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. Gene Dev. 8:2046–2057.
Thomas J, Schaack S, Pritham EJ. 2010. Pervasive horizontal transfer of rolling-circle transposons among animals. Genome Biol Evol. 2: 656–664.
Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 13: 36–46.
Venner S, Feschotte C, Biémont C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. Trends Genet. 25:317–323.
Vidal NMM, Ludwig A, Loreto ELS. 2009. Evolution of Tom, 297, 17.6 and rover retrotransposons in Drosophilidae species. Mol Genet Genomics. 282:351–362.
Vijaykrishna D, et al. 2007. Evolutionary insights into the ecology of coronaviruses. J Virol. 81:4012–4020.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8:973–982.
Woolhouse MEJ, Haydon DT, Antia R. 2005. Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol. 20: 238–244.
Xiong Y, Eickbush TH. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable element, and mitochondrial introns. Mol Biol Evol. 5:675–690.
Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9: 3353–3362.
Yang S, et al. 2004. Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. Genome Res. 14: 517–527.
Yoshiyama M, et al. 2001. Possible horizontal transfer of a transposable element from host to parasitoid. Mol Biol Evol. 18: 1952–1958.

Associate editor: Ken Wolfe