Helena and *BS*: Two Travellers between the Genera *Drosophila* and *Zaprionus*

Maryanna C. Simão[1], Annabelle Haudry[2], Adriana Granzotto[1], Nathalia de Setta[3], and Claudia M.A. Carareto[1,*]

[1]Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas (Ibilce), Câmpus São José do Rio Preto, SP, Brazil

[2]Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France

[3]Universidade Federal do ABC (UFABC), Centro de Ciências Naturais e Humanas (CCNH), São Bernardo do Campo, SP, Brazil

*Corresponding author: E-mail: claudia.carareto@unesp.br.

Accepted: August 22, 2018

Data deposition: The sequences of reverse transcriptase gene of the Helena and BS non-LTR retrotransposons generated in the study have been deposited in GenBank under accession numbers MH047863 to MH047945.

Abstract

The frequency of horizontal transfers of transposable elements (HTTs) varies among the types of elements according to the transposition mode and the geographical and temporal overlap of the species involved in the transfer. The drosophilid species of the genus *Zaprionus* and those of the *melanogaster*, *obscura*, *repleta*, and *virilis* groups of the genus *Drosophila* investigated in this study shared space and time at some point in their evolutionary history. This is particularly true of the subgenus *Zaprionus* and the *melanogaster* subgroup, which overlapped both geographically and temporally in Tropical Africa during their period of origin and diversification. Here, we tested the hypothesis that this overlap may have facilitated the transfer of retrotransposons without long terminal repeats (non-LTRs) between these species. We estimated the HTT frequency of the non-LTRs *BS* and *Helena* at the genome-wide scale by using a phylogenetic framework and a vertical and horizontal inheritance consistence analysis (VHICA). An excessively low synonymous divergence among distantly related species and incongruities between the transposable element and species phylogenies allowed us to propose at least four relatively recent HTT events of *Helena* and *BS* involving ancestors of the subgroup *melanogaster* and ancestors of the subgenus *Zaprionus* during their concomitant diversification in Tropical Africa, along with older possible events between species of the subgenera *Drosophila* and *Sophophora*. This study provides the first evidence for HTT of non-LTRs retrotransposons between *Drosophila* and *Zaprionus*, including an in-depth reconstruction of the time frame and geography of these events.

Key words: transposable elements, horizontal transfer, non-LTR retrotransposons, drosophilids.

Introduction

The connections among all living entities are determined by the flow of genetic information through the generations, which occurs predominantly through the reproductive process and is termed vertical transfer (VT). However, other nonvertical forms of inheritance have been demonstrated that are generically termed horizontal transfer (HT), by which genes and other genetic sequences are transferred between species through nonsexual means (Schaack et al. 2010). Horizontal transfer of transposable elements (HTT) was long believed to be restricted to prokaryotes (Andersson 2005). Until the 1990s, only a small number of HTTs among eukaryotes had been reported (e.g., Daniels et al. 1984; Anxolabéhère et al. 1988; Daniels et al. 1990; Clark et al. 1994; Robertson and Zumpano 1997; Kordis and Gubensek 1998), but in the last two decades, the number of HTTs described in invertebrates, vertebrates and plants, and between species from different higher taxa or realms has increased considerably (e.g., Silva and Kidwell 2000; De Almeida and Carareto 2006; Diao et al. 2006; Ludwig et al. 2008; Pace et al. 2008; Ray et al. 2008; de Setta et al. 2009; Dias and Carareto 2012; Thomas et al. 2010; Gilbert et al. 2010; de Setta et al. 2011; El Baidouri et al. 2014; Dias et al. 2015; Kofler et al. 2015; Suh et al. 2016; Peccoud et al. 2017; Gao et al. 2018).

Among the genetic elements involved in HTs in eukaryotes, transposable elements (TEs) are the most frequently reported

(Modolo et al. 2014). TEs are repetitive sequences of DNA that have the ability to mobilize in the genome, and due to this mobility, they stand out in HTT studies. The number of identified HTT events increased markedly in the last decade, from 156 cases described prior to 2007 to 2,855 at the time of this query (December 2, 2018), as shown in the database HTT-DB (Horizontally Transferred Transposable Elements DataBase, http://lpa.saogabriel.unipampa.edu.br:8080/HTTdatabase/; last accessed December 2016). This database, as well as the results of previous studies (reviewed in Schaack et al. 2010; Carareto 2011; Peccoud et al. 2017), showed that the frequency of HTT is not constant among the different types of elements. A greater number of HTTs involve DNA transposons (2,178) rather than long terminal repeat (LTR) retrotransposons (353) and non-LTR retrotransposons (324). The difference in the frequency of HTTs can be explained by both the structural characteristics and the mechanisms of transposition of the different types of TEs, along with the functionality of the enzymatic machinery necessary to perform the initial transposition events. The presence of a TE as a free DNA molecule in the cell facilitates the transfer out of the cell and the insertion into another host genome, particularly for DNA transposons. The simplicity of the structure of DNA transposons, such as those of the *Tc1-Mariner* superfamily, could facilitate their transport in a vector and increase the probability of HTT (Schaack et al. 2010). Moreover, the success of an HTT event may be related to the ability of the promoter to drive gene expression in a new genetic and genomic environment, as demonstrated by Palazzo et al. (2017) using the *Bari* family of DNA transposons. HTT may also be facilitated for retrotransposons such as the *Gypsy* superfamily LTRs. Some *Gypsy* elements, such as those belonging to the clades *Gypsy* and *17.6* in *Drosophila*, and in the lepidopteran *Trichoplusia ni*, have an additional open reading frame (ORF) that encodes a retroviral envelope protein (http://www.gydb.org/index.php/Ty3/Gypsy; last accessed September 2017; Friesen and Nissen 1990; Pélisson et al. 1994; Song et al. 1994). The *gypsyDM* of *D. melanogaster*, for example, produces a fully functional protein that provides it with both infectious ability and the possibility to be horizontally transferred by contact or feeding when individuals of an "empty" stock are raised on medium containing ground pupae of the stock possessing transposable elements (Kim et al. 1994). These results raise the possibility of *Gypsy* being both a retrotransposon and a facultative retrovirus, expressing an infectious form only under special circumstances (Song et al. 1994).

The frequency of HTT also differs among groups of organisms. Thus far, there has been an observable bias towards *Drosophila*/Insecta in the literature for documented HTTs. For example, of the 330 HTT cases published by 2012, 178 were described among *Drosophila* species. This disproportionate number of HTTs in *Drosophila* may be a historical deviation, as the pioneer studies of HTT in TEs were performed with this model organism (Carareto 2011; Wallau et al. 2012). Among the events described as HTTs in *Drosophila*, we previously proposed the transfer of LTR retrotransposons between two groups of species: the *melanogaster* subgroup of the *melanogaster* group (subgenus *Sophophora*, genus *Drosophila*, Drosophilidae) and the subgenus *Zaprionus* (genus *Zaprionus*, Drosophilidae) (De Setta et al. 2009, 2011). The genus *Zaprionus*, which is divided into two subgenera—*Anaprionus*, of Oriental region distribution, and *Zaprionus*, of Tropical Africa distribution from 7 Mya (Okada and Carson 1983; Yassin et al. 2008)—originated in the middle and late Miocene in the Eastern biogeographic region (Yassin et al. 2008). The *melanogaster* subgroup, one of the seven subgroups that form the *melanogaster* group of the genus *Drosophila*, also originated in Tropical Africa from a founding lineage that reached the African continent between 17 and 20 Mya from the Eastern region when the faunal interchange between Africa and Eurasia first became possible (Jeffs et al. 1994; Lachaise and Silvain 2004). It has been proposed that three speciation centers in the African continent were responsible for the formation of three species complexes of the *melanogaster* subgroup: the *erecta*, the *yakuba*, and the *melanogaster* complexes. The *erecta* (*D. erecta* and *D. orena*) and *yakuba* (*D. yakuba*, *D. teissieri*, and *D. santomea*) complexes evolved in the western Africa from the common ancestral *melanogaster* lineage, with the first complex between 13 and 15 Mya and the second between 8 and 15 Mya (Lachaise and Silvain 2004) or even later, approximately 6 Mya (Russo et al. 1995). In the *melanogaster* complex (*D. melanogaster*, *D. simulans*, *D. sechellia*, and *D. mauritiana*), the dating of the divergence between *D. melanogaster* and *D. simulans* from an ancestral lineage in Tropical Africa remains controversial but may have occurred between 3 and 4 Mya (Tamura et al. 2004a, 2004b; Cutter 2008). Later, the ancestor of the subcomplex *simulans* (*D. simulans*, *D. sechellia*, and *D. mauritiana*) arose during the colonization of Madagascar and the islands of the Indian Ocean, approximately 0.4 Mya (Lachaise and Silvain 2004). The above-proposed evolutionary scenario is shown in figure 1.

The high number of HTTs among species of the subgroup *melanogaster* and the subgenus *Zaprionus* could be explained by the presence of shared putative vectors such as viruses, bacteria and microparasitoids that would have facilitated transfers of genetic material between those species that overlapped geographically and temporally during their origin and diversification (De Setta et al. 2009, 2011). A small effective population size (*Ne*) during the species origin could be another factor facilitating HTT, since permissiveness to the fixation of TEs in the genome (in general, including those introduced by HTTs) can be increased in newly formed species due to small *Ne*, which reduces the efficacy of natural selection against invasive DNA (reviewed in Carareto 2011). However, how many and which of these factors contributed to the high exchange rate of genetic material in TEs between these two species groups remains unknown.

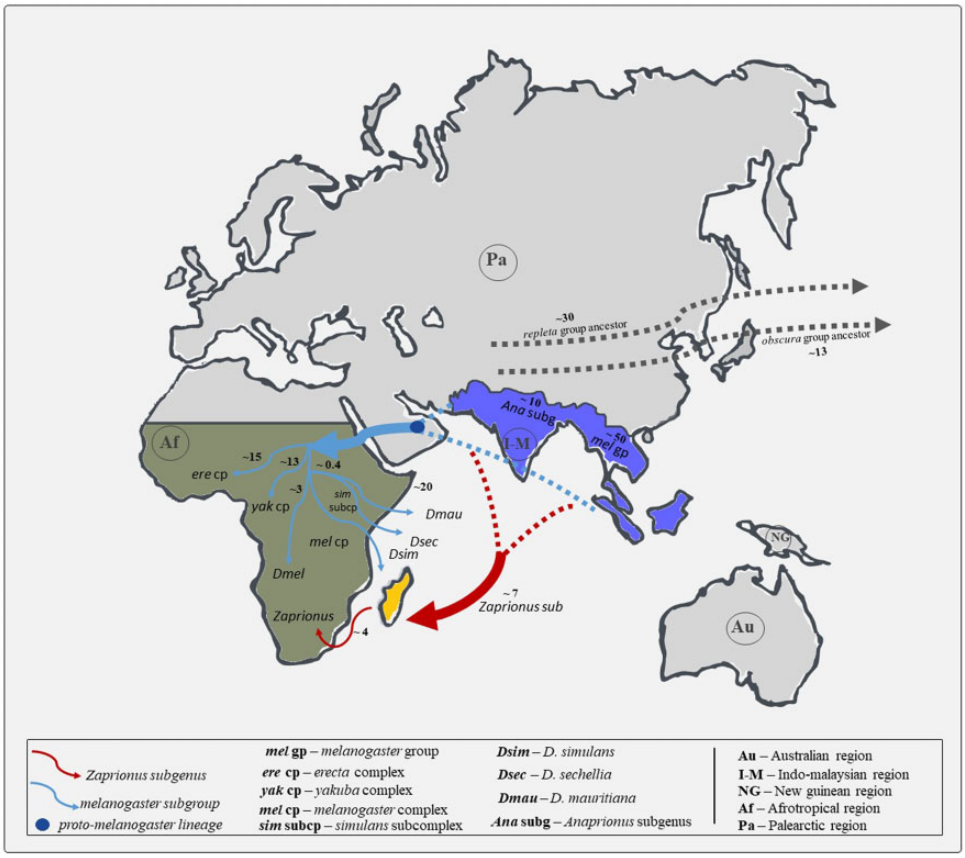

Fig. 1.—Evolutionary scenario and historical biogeography of drosophilid in the Old World with emphasis on the subgroup *melanogaster* (*melanogaster* group, *Sophophora* subgenus, genus *Drosophila*) and the subgenus *Zaprionus* (*Zaprionus* genus). The ages (numbers in My) of the African continent colonization of the *melanogaster* subgroup and *Zaprionus* subgenus, migrations (arrows) and lineages diversification in each region are indicated (Okada and Carson 1983, Jeffs et al. 1994, Russo et al. 1995, Yassin et al. 2008, Lachaise and Silvain 2004).

This study is the first to investigate the occurrence and evolutionary relationships of two families of non-LTR retrotransposons *Helena* (Petrov et al. 1995) and *BS* (Udomkit et al. 1995) in species of the genus *Zaprionus* and to perform comparative analyses with the sequences of both families in species of the genus *Drosophila*. *Helena* and *BS* belong to the *Jockey* superfamily of the LINE order (Wicker et al. 2007). *Helena* has a 25 bp poly-A tail and two overlapping ORFs (ORF1 and ORF2). The first ORF is 1,737 bp long, encodes a 579 aa protein that has high similarity to the gag protein of other LINE-like elements and contains a domain PRE_C2H2 (associated with zinc fingers). The second ORF, which starts on the last base of ORF1, is 2,721 bp, encodes a 907 aa protein corresponding to the pol gene and contains the apyrimidic endonuclease and exonuclease (ENDO_EXO) domains and the reverse transcriptase (RTASE) domain (Rebollo et al. 2008). The *BS* element has a structure similar to the *Helena* element, with two ORFs and a 13 bp poly-A tail at the 3′ end. The ORF1 is 2,580 bp long and encodes a 860 aa protein with the PRE_C2H2 domain; the ORF2 is 2,911 bp long and encodes a 970 aa protein with the ENDO_EXO and RTASE domains.

Both elements show discontinuous distribution within the genus *Drosophila*, varying in terms of structure, sequence conservation, number of copies and transcriptional activity (Petrov et al. 1998; Rebollo et al. 2008; Granzotto et al. 2009, 2011). The element *Helena* has been identified in six species of the *melanogaster* subgroup of the *melanogaster* group (*D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta*, and *D. ananassae*), three species of the *repleta* group (*D. mojavensis, D. koepferae*, and *D. buzzatii*), and one species each of the *obscura* (*D. pseudoobscura*) and the *virilis* (*D. virilis*) groups of the genus *Drosophila* (Petrov et al. 1995; Granzotto et al. 2009; Romero-Soriano and Guerreiro 2016). *Helena* shows mostly truncated copies, except for the *D. mojavensis* elements of the *repleta* group, which are numerous and have high levels of transcriptional activity (Granzotto et al. 2009). In contrast, the element *BS* has also been reported in six species of the *melanogaster* group (*D. yakuba, D. erecta, D. simulans, D. melanogaster*, *D. sechellia*, and *D. ananassae*), two species of the *obscura* group (*D. pseudoobscura* and *D. persimilis*) and *D. mojavensis* (Granzotto et al. 2011). *BS* was characterized recently in a

larger number of species—*D. melanogaster*, *D. erecta*, and *D. mojavensis*—with complete ORFs and transcriptional activity (Granzotto et al. 2011).

Our study shows evidence of HTTs of the *Helena* and the *BS* elements that could have occurred concomitantly with the HTTs of LTR retrotransposons (De Setta et al. 2009, 2011). We were able to detect at least four HTT events that took place in Tropical Africa between the ancestral species of the *melanogaster* subgroup and the ancestral *Zaprionus* species, along with other putative transfers of both elements between species of the subgenera *Drosophila* and *Sophophora* with an Oriental origin.

Material and Methods

Biological Material and Sequencing of Elements

The occurrence of the *Helena* and *BS* elements were investigated through PCR and Sanger sequencing in 11 species of the genus *Zaprionus* (supplementary table S1, Supplementary Material online): ten species of the subgenus *Zaprionus* (*Z. indianus, Z. gabonicus, Z. africanus, Z. ornatus, Z. camerounensis, Z. davidi, Z. tuberculatus, Z. inermis, Z. nigranus* and *Z. sepsoides*) and one species of the subgenus *Anaprionus* (*Z. bogoriensis*). Genomic DNA was extracted from the ovaries of 20 individuals of each species using the phenol-chloroform method (Jowett 1986).

For amplification of the *Helena* element, a pair of primers (DsechF: 5′ AGG ATT TGT CAT GCC ACG CT 3′ and DsechR: 5′ TGT TTG GTG CTG CCA TGT GT 3′) that amplifies a 640 bp sequence of the gene of the complete element of *D. sechellia* were used using the following PCR conditions: 200 ng of genomic DNA, each dNTP 0.5 µl at 0.625 mM, 0.75 µl at 0.8 mM $MgCl_2$, each primer at 0.5 µl at 10 mM, and 1 U of Platinum Taq (Invitrogen) in 1× buffer, in a final volume of 25 µl. The amplification reaction was performed with the following parameters: 95 °C for 5 min, 24 cycles of 95 °C for 1 min, 62 °C for 1 min, 72 °C for 1 min, and a final extension at 72 °C for 10 min. *BS* element amplifications were performed with a pair of primers (DmelF: 5′ TGA AGA GAG CCC TGA ATC GT 3′ and DmelR: 5′ GTG AAG CAG GGA TTG ATG GT 3′) that amplifies a 774 bp sequence of the gene of the complete element of *D. melanogaster*. The amplification reaction was performed under the same conditions described above, with cycling as follows: 95 °C for 5 min, 35 cycles of 95 °C for 2 min, 62 °C for 1 min, 72 °C for 1 min, and a final extension at 72 °C for 5 min. PCR fragments were purified with an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) and cloned with a TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instructions. Three to ten clones were randomly selected and sequenced in an ABI 3730 xl DNA Analyzer (Applied Biosystems), at the Center for Biological Resources and Genomic Biology (CREBIO, UNESP, Jaboticabal, Brazil), using the M13 primers. The RT sequences generated have been deposited in GenBank under accession numbers MH047863 to MH047945.

Identification of Elements in silico

Sequences of the *Helena* and *BS* retrotransposons were also obtained both from public databases and from 26 genomes of *Drosophila* (Sessegolo et al. 2016) and three genomes of *Zaprionus* made available by A. Haudry (supplementary tables S2, S3 and S4, Supplementary Material online). Raw reads were automatically filtered for quality using UrQt (Modolo and Lerat 2015). TEs were identified by de novo assembly and annotation in dnaPipeTE (Goubert et al. 2015) from random samples corresponding to 0.25× coverage.

To identify the TEs in the publicly available genomes, we searched for sequence similarity with the reference sequences of the *Helena* element of *D. simulans* (4,912 bp; Rebollo et al. 2008) and the *BS* element of *D. melanogaster* (5,124 bp; Udomkit et al. 1995) using BLASTn (Altschul et al. 1990) with a cut-off value $\leq 1e^{-10}$. The three best hits were extended by 3 kb in the 5′ and 3′ flanking regions to allow identification of the ends of the elements. The most conserved and most complete sequence obtained by BLASTn was considered as the reference sequence for each element in each genome. With the reference sequence of each species, the BLASTn search was performed a second time, using as parameters length ≥300 bp and identity ≥80%. Finally, each putative TE sequence was analyzed for the presence and integrity of the PRE_C2H2 (ORF1), ENDO_EXO and RTASE (ORF2) of the *Helena* and *BS* coding sequences using a BLAST CD-search against the Conserved Domain Database (CDD) (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; last accessed July 2016).

Phylogenetic Inferences

From all *Helena* and *BS* sequences obtained, the RT sequences were extracted and aligned with the RT-amplified sequences of *Zaprionus* species using MAFFT software (Katoh et al. 2017). The phylogenies were reconstructed using the maximum likelihood (ML) method with 1,000 bootstrap replicates to validate the robustness of the phylogeny (Felsenstein 1985). Additionally, Bayesian inferences of phylogeny (BI) were performed with the program BEAST v16.1 (Drummond et al. 2012) with an *a posteriori* phylogenetic support test, which involved the sampling of 100,000 trees with 10% burn-in. The evolutionary model of substitution that best fit the data was determined by the Find Best DNA Model (MEGA7, Kumar et al. 2016). The sequences of elements belonging to the *Jockey* clade to which *Helena* and *BS* belong (Metcalfe and Casane 2014), *Jockey1* (zind_LINE_comp657, zafr_LINE_comp2868, divir_LINE_comp1826) and *Jockey2* (dan_LINE_comp1636, dbip_LINE_comp1936) obtained from the dnaPipeTE data set, were used as an outgroup. Sequences of the elements *Doc*

(X17551.1) and TART (U02279.1), which also belong to the *Jockey* clade, were obtained from GenBank and used as an outgroup. The nucleotide divergence (*p*-distance) was estimated using MEGA 7.

The evolutionary relationships between the sequences of *Helena* and *BS* were also inferred using the median-joining algorithm implemented by the Network 5.0.0.1 program (Bandelt et al. 1999), with the default parameters. Amino acid sequences were used due to the high nucleotide variation between sequences of distantly related species. The networks show a feature not found in phylogenetic analyses: the proposition of median vectors that represent sequences inferred either because they were not sampled or because they were lost during evolution, a likely pattern in the evolutionary cycle of TEs. These vectors can also represent the ancestral state of a sequence, a plausible scenario, because once the element transposes, it begins to evolve independently in the genome (Cordaux et al. 2004).

The divergence times of the *BS* and the *Helena* sequences from the most recent ancestral sequence shared by the species were estimated using the Bayesian approach and BEAST v16.1 (Drummond et al. 2012), with a neutral nucleotide substitution rate of $r = 0.016$/site/My (Sharp and Li 1989) for the calibration of the phylogenetic tree.

The phylogeny of the 26 drosophilid species involved in this study was reconstructed based on the nuclear gene *Amyrel*, which was obtained from GenBank for most of the species (supplementary table S5, Supplementary Material online). For *D. mojavensis* and *D. ananassae*, the sequences were obtained using *Genome Browser* tools (https://genome.ucsc.edu/cgi-bin/hgBlat; last accessed December 2016). For *Z. gabonicus* and *Z. africanus*, the sequences were retrieved from the genomes using the *Z. indianus* sequence (EF458322.1) as the query. The phylogenetic inference was performed with MEGA 7 using the Tamura-3-parameter as the substitution model.

Vertical and Horizontal Inheritance Consistence Analysis (VHICA)

The VHICA approach (Wallau et al. 2016) was used to corroborate the phylogenetic inferences of vertical and horizontal transfer by providing statistical support. The method is based on discrepancies between the rate of evolution in synonymous sites (dS) and the preferential use of codons (ENC) between pairs of TE sequences and vertically transferred orthologous genes. Statistical support for the HTT inferences is given by a linear regression between the distribution of ENC and dS values (with the Bonferroni correction, $P < 0.01$). For each pair of species, the correlation between ENC and dS is calculated, and the residuals of a linear regression $ENC = a\,dS + b$ among reference genes, assumed to be vertically transmitted, and those of the TEs are plotted in a graph. Statistically significant deviation is interpreted as indicative of HTT. Thirty orthologous genes from 20 species were used. The gene sequences of *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. biarmipes*, *D. bipectianta*, *D. elegans*, *D. ficusphila*, *D. takahashii*, *D. kikkawai*, *D. grimshawi*, *D. persimilis*, *D. pseudoobscura*, *D. mojavensis*, and *D. virilis* were obtained from the HTTDB (http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase; last accessed December 2016) whereas those of the species *Z. africanus*, *Z. gabonicus* and *Z. indianus* were obtained through searching directly the genomes. The orthology of these genes was verified using OrthoDB—The Hierarchical Catalog of Orthologs v9.1 (http://www.orthodb.org/; last accessed February 2017). All genes were single-copy, except for the CG4386 gene of *D. melanogaster* (a peptidase S1 gene family), which presents two paralogs in eight species (*D. melanogaster, D. ananassae, D. erecta, D. simulans, D. yakuba, D. persimilis, D. pseudoobscura*, and *D. grimshawii*). In this case, phylogenetic analysis identified the orthologous sequences, and only one of the two sets of orthologues were used in the VHICA analysis. The same 30 genes were used for reconstructing a phylogeny required for interpreting the HTT signals (supplementary table S6, Supplementary Material online). For both TEs, a consensus was used when the sequences from each genome had <10% nucleotide divergence, as recommended in the VHICA tool. In this analysis, it was possible to use only the sequence of the *BS* element found in *D. yakuba* genome, which was not used in the phylogenetic analysis because of a deletion comprising a large part of the reverse transcriptase gene.

Results

Distribution of the Elements *Helena* and *BS* in the Genomes of *Zaprionus* and *Drosophila*

In addition to the 11 species of *Drosophila* in which the element *Helena* had been previously identified (Petrov et al. 1995; Granzotto et al. 2009), we showed its presence in four other species of the *melanogaster* group, three species of the *melanogaster* subgroup (*D. teissieri, D. mauritiana*, and *D. orena*) and one species of the *ananassae* subgroup (*D. bipectinata*), all sequences found by genome sequence search. However, *Helena* was not found in the genomes of the nine Oriental species of the *melanogaster* group (*D. malerkotliana, D. suzukii*, *D. biarmipes, D. elegans*, *D. eugracilis, D. ficusphila, D. kikkawai, D. rhopaloa*, and *D. takahashii*) among the 11 Oriental species whose genomes were investigated (supplementary table S2, Supplementary Material online). The element *Helena* was found in 11 of the *Zaprionus* species studied, except for the Oriental *Z. bogoriensis*, which was the only species of the subgenus *Anaprionus* studied (supplementary tables S1 and S3, Supplementary Material online). Analysis of the occurrence and integrity of the domains PRE_C2H2 (ORF1), ENDO_EXO and RTASE (ORF2) in all insertions of the *Z. africanus, Z. gabonicus*, and *Z. indianus* genomes showed that a single copy of *Helena* from *Z.*

indianus had the three domains, but two of them—ENDO_EXO and RTASE—were incomplete. In *Z. gabonicus*, only two domains were predicted, but both were incomplete (supplementary fig. S1, Supplementary Material online).

The *BS* element had been previously identified in nine species of *Drosophila* (Udomkit et al. 1995; Granzotto et al. 2011). We were able to show its presence in only two other species of the *melanogaster* group (*D. ficusphila* and *D. bipectinata*), which were found by genome sequence searching, both of which are of Oriental origin (supplementary table S4, Supplementary Material online). In *Zaprionus, BS* was found in six of the 11 species investigated (supplementary tables S1 and S3, Supplementary Material online): five species of the *vittiger* group (*Z. indianus, Z. africanus, Z. gabonicus, Z. davidi*, and *Z. ornatus*) and one species of the *inermis* group (*Z. sepsoides*). Analyses of the genomes of *Z. africanus, Z. gabonicus*, and *Z. indianus* showed that at least three copies of *Z. indianus* and one copy of *Z. gabonicus* present the domains PRE_C2H2, ENDO_EXO and RTASE, indicating they are putative full-length insertions (supplementary fig. S2, Supplementary Material online).

Phylogenetic Inferences

The evolutionary relationships of the elements *Helena* and *BS* were evaluated by the maximum likelihood (ML) and Bayesian inference (BI) methods and produced similar results. The main differences between the ML and the BI trees were in the robustness values of the clades (ML trees not shown). Sixty-six *Helena* sequences were used: 39 from the subgenus *Zaprionus* and 28 from the genus *Drosophila* (fig. 2). All the sequences identified as *Helena* form a monophyletic group, corroborating the identification method used. Several phylogenetic incongruities can be identified among the sequences of the *Helena* elements of the *Drosophila* species from the *melanogaster*, *repleta*, and *virilis* groups and those of the subgenus *Zaprionus* species. The two clades that diverged approximately 9.5 Mya are incongruent with the species phylogeny (mirroring the *Helena* tree in fig. 2) despite strong posterior probability support (PP = 1). The first clade contains sequences from Oriental species of the group *melanogaster* (*D. bipectinata* and *D. ananassae*) and from Nearctic and Neotropical species of the *repleta* (*D. mojavensis, D. koepferae*, and *D. buzzatii*) and *virilis* (*D. virilis*) groups, sharing a common ancestral sequence dated at 7.4 Mya. The second clade contains *Helena* sequences of Tropical African species belonging to the subgroup *melanogaster* and the subgenus *Zaprionus*, which share a common ancestral sequence dated at 5.5 Mya. In this clade, the sequences of the *melanogaster* complex (*D. melanogaster*, *D. simulans*, *D. sechellia*, and *D. mauritiana*) are more closely related to those of the subgenus *Zaprionus* (node PP = 0.91) than to the other species of its subgroup (*erecta* complex: *D. erecta*, *D. orena; yakuba* complex: *D. yakuba* and *D. teissieri*), and they share a common ancestral sequence dated at 4.1 Mya. Finally, the *Z. ornatus* sequences are grouped with *D. sechellia* sequences (node PP: 0.95), with the time of divergence from an ancestral sequence dated at 0.2 Mya. Apart from *Z. ornatus*, the sequences of the species of each complex of the subgenus *Zaprionus* grouped as expected according to the phylogeny of the species proposed by Yassin et al. (2008), as expected by vertical inheritance.

We observed a smaller divergence between *Helena* sequences of distantly related species than between closely related species. For example, the mean divergence between the *Helena* sequences of the *melanogaster* complex and the *Zaprionus* genus is lower (6.7%) than that between the *Helena* sequences of this complex and the *yakuba* and the *erecta* complexes (10.5% and 10.2%, respectively), which form the *melanogaster* subgroup (supplementary table S7, Supplementary Material online). More interestingly, *Helena* sequences of *Z. ornatus* are more similar to the sequences of *D. sechellia* (2%) than to the *Helena* sequences of the other *Zaprionus* species (7%). However, although they are grouped in the phylogeny, the sequences of the Oriental *melanogaster* group (*D. bipectinata* and *D. ananassae*) are as divergent from those of the *repleta* and *virilis* groups (25%) as they are from those of the *melanogaster* complex (25%) and the *Zaprionus* subgenus (28%). These degrees of divergence are incongruent with the phylogenetic relationships between these species.

The phylogenetic analyses for the element *BS* corroborated our method of identification for this element because the *BS* sequences are grouped in a clade with high posterior probability (PP = 1) (fig. 3). Such as the *Helena* element, phylogenetic incongruities involving the *BS* sequences were observed when compared with the species phylogeny (mirrored the *BS* tree). *BS* sequences of the Oriental species of the group *melanogaster, D. ficusphila*, and *D. bipectinata*, are positioned basally in the *BS* clade. The other clade contains all the other sequences, and the Nearctic species of the group *repleta* of the genus *Drosophila* (*D. mojavensis*) and the *obscura* group of the subgenus *Sophophora* (*D. persimilis* and *D. pseudoobscura*) occupy a basal position. Interestingly, as with the element *Helena*, the sequences of the *melanogaster* species complex (*D. melanogaster*, *D. simulans*, and *D. sechellia*) are more closely related to the sequences of the *Zaprionus* species than to the sequences of *D. erecta* (*melanogaster* subgroup), with high posterior probability (PP = 1). The origin of the clade of sequences from species in the *melanogaster* complex and the *Zaprionus* subgenus is dated at 2.7 Mya, whereas that of the clade that includes the sequences of *erecta* and the *melanogaster* complexes and *Zaprionus* sequences (PP = 1) is dated at 7.3 Mya. The divergence of this clade from the one possessing sequences of the species *D. mojavensis*, *D. persimilis*, and *D. pseudoobscura* is dated at 12.2 Mya (PP = 1). Finally, the unexpected clustering of a sequence of *Z. africanus* (Z. afr_9) with the clade of *Z. sepsoides* (PP = 1) is incongruent with the phylogenetic relationships between the two species.

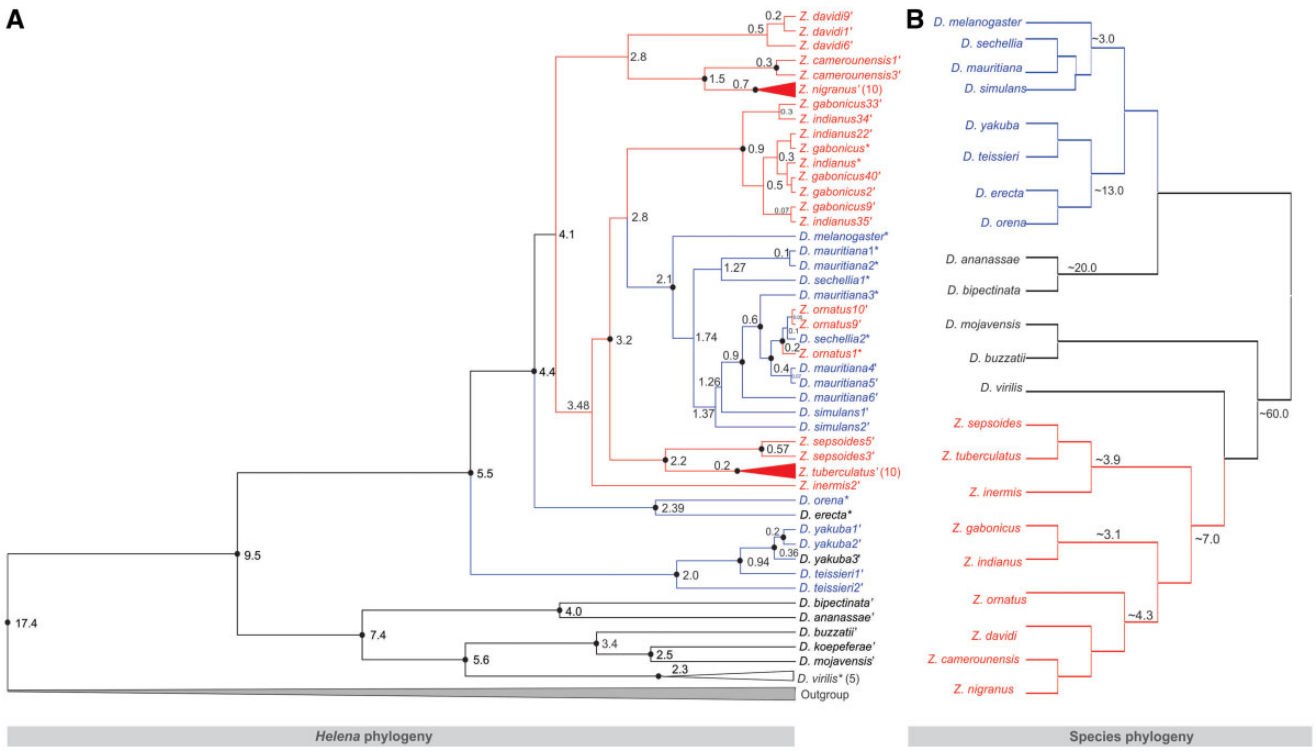

Fig. 2.—Calibrated tree of *Helena* sequences mirrored by a phylogenetic species tree reconstructed with sequences of the gene *Amyrel*. (*A*) The *Helena* tree was reconstructed with partial sequences of the reverse transcriptase gene using the Bayesian phylogenetic inference method and the Tamura-Nei nucleotide substitution model (Tamura and Nei 1993). The analysis involved 78 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were 397 positions in the final alignment. Branch support values >0.7 are indicated by black circles at the root of each clade, with the age estimated for each branching. Asterisks indicate sequences retrieved from sequenced genomes, and apostrophes indicate the PCR amplified sequences. The *Helena* sequences of the subgenus *Zaprionus* and the subgroup *melanogaster* are shown in red and blue, respectively. The branch with 11 sequences of the non-LTR elements *Doc* (1), *Jockey* (9), and *TART* (1), used as outgroup, was collapsed. The evolutionary analyses were conducted in BEAST v16.1 (Drummond et al. 2012). (*B*) The species tree was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter substitution model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 21 nucleotide sequences with 1,489 positions in the final alignment. Branch support values >0.7 are indicated by black circles. The ages of divergence of the *melanogaster* group and complexes indicated in the branches follow Lachaise and Silvain (2004), and those of the *Zaprionus* subgroups follow Yassin et al. (2008). Evolutionary analyses were conducted in MEGA 7 (Kumar et al. 2016).

The divergence estimated between the *BS* sequences of the different clades highlights incongruences with the species phylogeny (supplementary table S8, Supplementary Material online). Mean divergence between the *BS* sequences of the *melanogaster* complex and the subgenus *Zaprionus* (8%) is lower than the divergence between the sequences of this complex and that of *D. erecta* (18.5%). Additionally, the *BS* sequences of the species of the Oriental *melanogaster* group (*D. bipectinata* and *D. ficusphila*) are less divergent from those included in the clade that groups the sequences of the *obscura* and *repleta* groups (21%) than from those of the *melanogaster* (34%) and *Zaprionus* (33%) complexes.

Network Analysis

The network analyses corroborated the results of the phylogenies showing that the *BS* and *Helena* sequences from species belonging to the *melanogaster* complex are more closely related to the sequences from species of the subgenus *Zaprionus* (figs. 4 and 5).

The *Helena* sequences of the *Zaprionus* subgenus and the *melanogaster* subgroup are very similar, separated by short branches that correspond to substitutions of fewer than ten amino acids. In contrast, the sequences of *Zaprionus* and the Tropical African species of the *melanogaster* group are separated from the sequences of the Oriental species of the *melanogaster* group by long branches corresponding to substitutions of at least 30 amino acids. The center of the network is occupied by many median vectors, without establishing direct relationships among the sequences of the species of the *Zaprionus* subgenus and the *melanogaster* subgroup. The sequences of the *melanogaster* species group from the Oriental region are also bound by median vectors to the sequences of species of the subgenus *Drosophila* (*virilis* and *repleta* groups), which might represent either unsampled sequences or ancestral states of the sequences that were

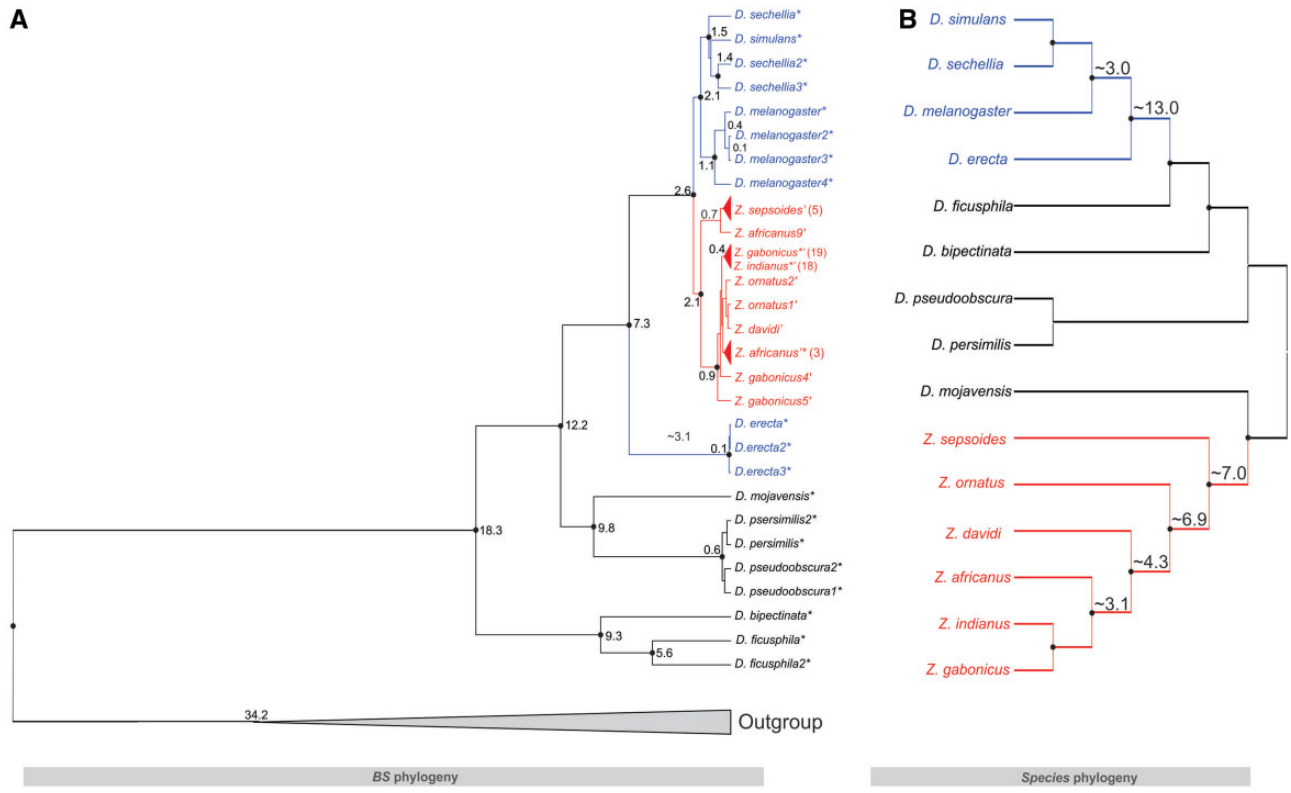

Fig. 3.—Calibrated tree of *BS* sequences mirrored by a phylogenetic species tree reconstructed with sequences of the gene *Amyrel*. (*A*) The *BS* tree was reconstructed with partial sequences of the reverse transcriptase gene using the Bayesian phylogenetic inference method and the Kimura 2-parameter substitution model (Kimura 1980). The analysis involved 86 nucleotide sequences. All ambiguous positions were removed for each sequence pair, leaving 762 positions in the final alignment. Branch support values >0.7 are indicated by black circles at the root of each clade, with the age estimated for each branching. Asterisks indicate *BS* sequences obtained from sequenced genomes, and apostrophes indicate the PCR amplified sequences The *BS* sequences of the subgenus *Zaprionus* and subgroup *melanogaster* are represented in red and blue, respectively. The branch with seven sequences of the non-LTR elements *Doc* (1), *Jockey* (5), and *TART* (1), used as outgroup, was collapsed. Evolutionary analyses were conducted in BEAST v16.1 (Drummond et al. 2012). (*B*) The species tree is inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences with 1,489 positions in the final alignment. Branch support values >0.7 are indicated by black circles. The ages of divergence of the *melanogaster* group and complexes indicated in the branches follow Lachaise and Silvain (2004), and those of the *Zaprionus* subgroups follow Yassin et al. (2008). Evolutionary analyses were conducted in MEGA 7 (Kumar et al. 2016).

sampled. One exception is the direct relationship between the sequences of *D. sechellia* and the sequences of *Z. ornatus*. This suggestion of direct ancestry is concordant with the phylogenetic reconstruction (fig. 4).

Similar to *Helena*, the *BS* sequences are closely related to those of the *melanogaster* complex (fig. 5). Their sequences are separated by branches corresponding to substitutions of an average of 13 amino acids, fewer than those of the *melanogaster* group species that diverged in the Oriental region, which showed substitutions of at least 146 amino acids, or 59 amino acids substitutions relative to the sequences of species of the groups *obscura* (*D. pseudoobscura*, *D. persimilis*) or *repleta* (*D. mojavensis*). For the *Helena* sequences, the central median vectors may represent ancestral sequences not sampled, which preclude direct inference of the relationship between the sequences of the species of *Zaprionus* and those of the subgroup *melanogaster*. However, these vectors reinforce the separated clustering of the *BS* sequences from the groups *vittiger* and *inermis*, along with the clustering of Z_afr9 with the sequences of *Z. sepsoides*.

Identification of Vertical and Horizontal Transfers

To evaluate whether the phylogenetic incongruities observed for the *Helena* and *BS* elements can be explained by HTT events, we used the VHICA method, a recently proposed strategy, to differentiate HTT and VT events involving TEs among related species (Wallau et al. 2016). In this study, VHICA was used only for comparisons with the species of the subgenus *Zaprionus* with sequenced genomes—*Z. indianus*, *Z. gabonicus* and *Z. africanus*—which allowed access to sequences of genes orthologous to those of *Drosophila*

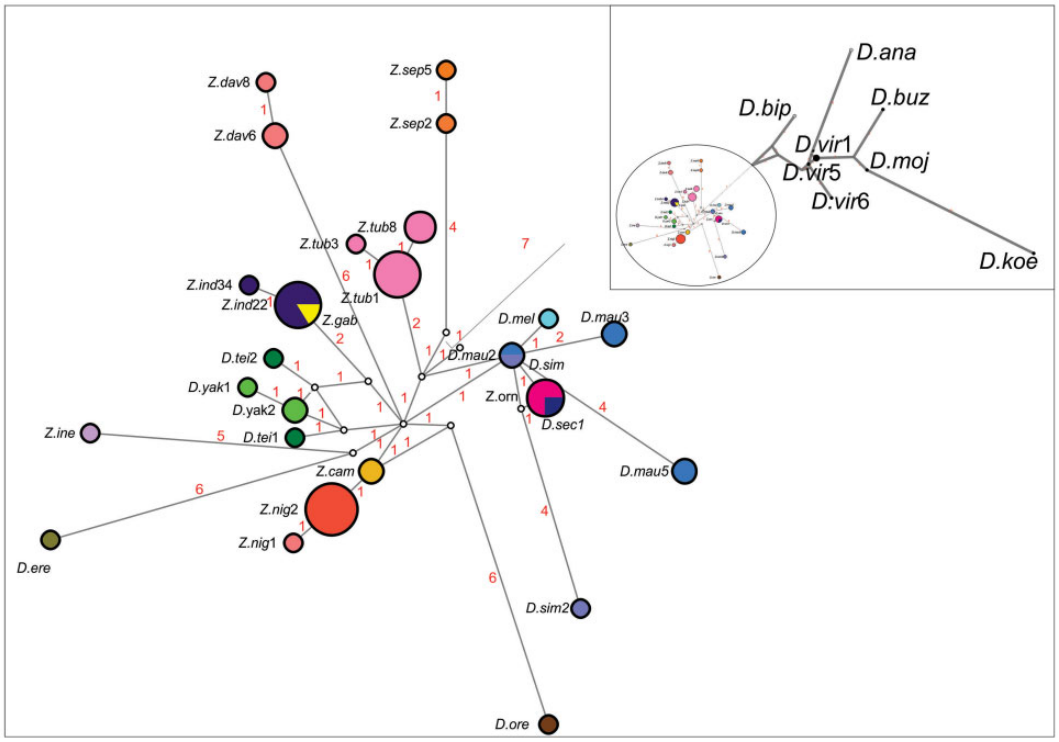

Fig. 4.—Phylogenetic network reconstruction for the non-LTR retrotransposon *Helena* of *Zaprionus* and *Drosophila* species with emphasis on the relationships between the sequences of *Zaprionus* and the species of the *melanogaster* complex. The network was constructed with partial amino acid sequences of the reverse transcriptase gene using the median-joining algorithm implemented in Network 5.0.0.1 (Bandelt et al. 1999). The size of each circle denotes the number of sequences grouped together, and the branch lengths are proportional to the number of substitutions between two nodes. Small empty circles represent the ancestor vectors, black circles represent sequences that do not belong to the *melanogaster* group, and grey circles represent sequences that belong to the *melanogaster* group from the Oriental region. The sequences from the *melanogaster* subgroup are represented by cool colors (blue to green), and *Zaprionus* sequences are represented by warm colors (red to purple). *Z. cam*: *Z. camerounensis*, *Z. sep*: *Z. sepsoides*, *Z. ind*: *Z. indianus*, *Z. gab*: *Z. gabonicus*, *Z. ine*: *Z. inermis*, *Z. orn*: *Z. ornatus*, *Z. dav*: *Z. davidi*, *Z. tub*: *Z. tuberculatus*, *Z. nig*: *Z. nigranus*, *D. mau*: *D. mauritiana*, *D. sec*: *D. sechellia*, *D. sim*: *D. simulans*, *D. mel*: *D. melanogaster*, *D. yak*: *D. yakuba*, *D. tei*: *D. teissieri*, *D. ore*: *D. orena*, *D. ere*: *D. erecta*, *D. bip*: *D. bipectinata*, *D. ana*: *D. ananassae*, *D. vir*: *D. virilis*, *D._moj*: *D. mojavensis*, *D. buz*: *D. buzzatii*, *D. koe*: *D. koepferae*. In this analysis, 86 codons and 65 sequences were used.

species (see supplementary table S6, Supplementary Material online). For the *Drosophila* species, this analysis only used sequences from species belonging to the *melanogaster* subgroup because the use of VHICA is not recommended when the sequences exhibit divergence (pairwise distance) higher than 30%.

The linear regression analyses performed by VHICA showed a scenario in which the sequences of the *Helena* and the *BS* elements are outside the limit of variance for VT in several comparisons between species (fig. 6 and supplementary figs. S3–S6, Supplementary Material online). For example, in comparisons of *D. simulans* versus *Z. indianus* and *D. sechellia* versus *Z. gabonicus*, the sequences have a low rate of evolution in synonymous sites (dS) and a high effective number of codons (ENC), values that are significantly different from the estimated variance limit for the orthologous genes in the linear regression, thus supporting the inference of HTT. However, there is no evidence of HTT events among species of the *melanogaster* subgroup, as exemplified by the comparisons between *D. yakuba* and *D. simulans*, in which the dS and ENC values of the two elements are within the range of estimated values for the orthologous genes. In the heatmaps, which present the results of the linear regressions, we can observe strong signals of HTT between the sequences of the *Helena* and the *BS* elements of the *Drosophila* species of the subgroup *melanogaster* (*D. mauritiana*, *D. sechellia*, *D. simulans*, *D. melanogaster*, *D. erecta*, and *D. yakuba)* and the *Zaprionus* species subgenus (*Z. africanus, Z. gabonicus, Z. indianus* for *BS*, and *Z. gabonicus, Z. indianus* for *Helena*). However, there is no evidence of HTT events either between the species of the *melanogaster* subgroup or between the species of the subgenus *Zaprionus* (supplementary fig. S7, Supplementary Material online).

Discussion

Three requirements must be considered when proposing HTTs: 1) the high similarity between sequences of TEs from

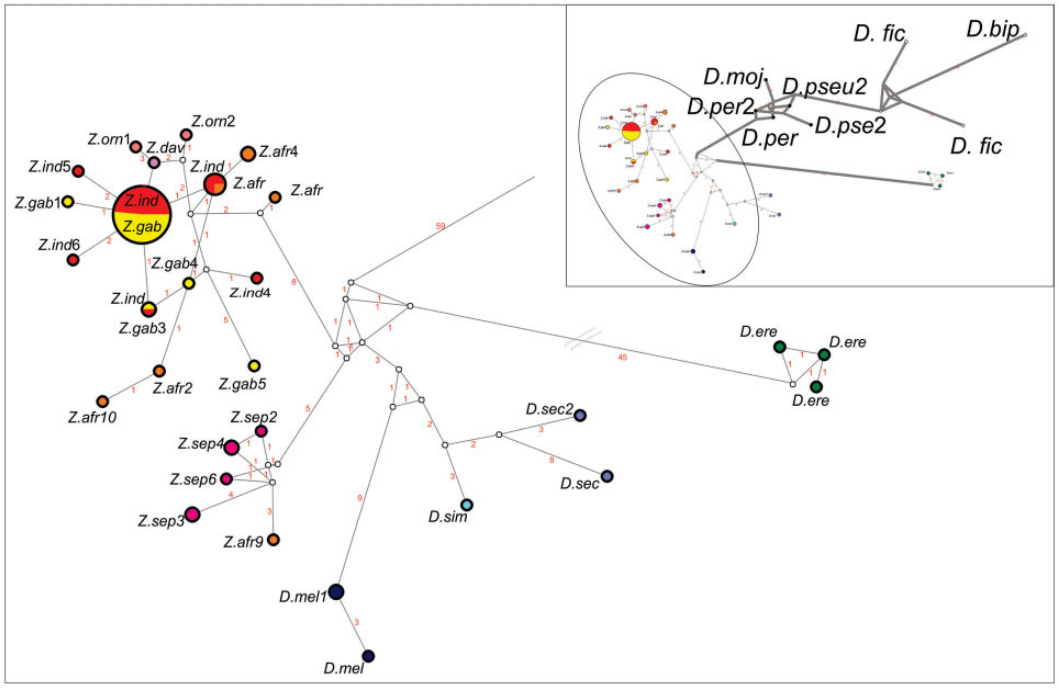

Fig. 5.—Phylogenetic network reconstruction for the non-LTR retrotransposon *BS* of *Zaprionus* and *Drosophila* species with emphasis on the relationships between the sequences of *Zaprionus* and the species of the *melanogaster* complex. The network was constructed with amino acid sequences of the reverse transcriptase gene using the median-joining algorithm implemented in Network 5.0.0.1 (Bandelt et al. 1999). The size of each circle denotes the number of sequences grouped together, and the branch lengths are proportional to the number of substitutions between two nodes. Small empty circles represent the ancestor vectors, black circles represent sequences that do not belong to the *melanogaster* group, and grey circles represent sequences that belong to the *melanogaster* group from the Oriental region. The sequences from the *melanogaster* subgroup are represented by cool colors (blue to green), and *Zaprionus* sequences are represented by warm colors (red to purple). *Z. ind*: *Z. indianus*, *Z. gab*: *Z. gabonicus*, *Z. afr*: *Z. africanus*, *Z. sep*: *Z. sepsoides*, *Z. dav*: *Z. davidi*, *Z. orn*: *Z. ornatus*, *D. mel*: *D. melanogaster*, *D. sim*: *D. simulans*, *D. sec*: *D. sechellia*, *D. ere*: *D. erecta*, *D. bip*: *D. bipectinata*, *D. fic*: *D. ficusphila*, *D. per*: *D. persimilis*, *D. pse*: *D. pseudoobscura*, and *D. moj*: *D. mojavensis.* This analysis used 227 codons and 67 sequences.

distantly related species, 2) inconsistencies between species and TEs phylogenies, and 3) the discontinuous distribution of TEs in a group of species (reviewed in Loreto et al. 2008; Cararetto 2011; Wallau et al. 2012). It is also important to consider whether the elements are amenable to mobilization due to the conservation of their structure and whether the lifestyle and geographical distribution of host species corroborate the molecular results (Loreto et al. 2008). In addition to those factors, the mechanism of transposition has been used to justify the differential frequency of HT among types of TEs. Peccoud et al. (2017) used bioinformatics analyses to identify 206 cases of HTT of non-LTR retrotransposons in insects. Combined with the HTT events published before 2018, 302 HTT cases are reported at http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase/resultado/resultado.jsp?organism; last accessed December 2016, but in most cases, the TEs involved are identified only at the superfamily level. Among them, 18 HTT cases deserve special mention because they exemplify the ability of specific LINE families to travel across genomes from species belonging to the same genus to higher taxa. Of these cases, 16 events involved the transfers of five LINE families in *Drosophila:* two events involved *Jockey* (Mizrokhi and Mazo 1990; Sánchez-Gracia et al. 2005), one event involved the *F* element, one event involved *Doc* (Sánchez-Gracia et al. 2005), one event involved the *I* element (Kidwell 1983; Brégliano and Kidwell 1983; Bucheton et al. 1984) and 11 events involved *Penelope* (Evgen'ev et al. 2000; Lyozin et al. 2001; Morales-Hojas et al. 2006). Additionally, the transfer of two RTE elements between species of different classes of vertebrata were reported, as Bov-B was transferred between snakes and ruminants probably by reptile ticks (Kordis and Gubensek 1998; Walsh et al. 2013) and two waves of transfers of AviRTE occurred between birds and parasitic nematodes (Suh et al. 2016). The results presented herein with the *Helena* and *BS* elements contribute to the enrichment of our knowledge of the HTT of non-LTR retrotransposons by increasing the number of events, TEs and species evaluated.

The three main requirements above-mentioned to propose HTTs are met in our study: high similarity between sequences of the *Helena* and *BS* from distantly related species, inconsistencies between the species and the element phylogenies, and their discontinuous distribution (absence

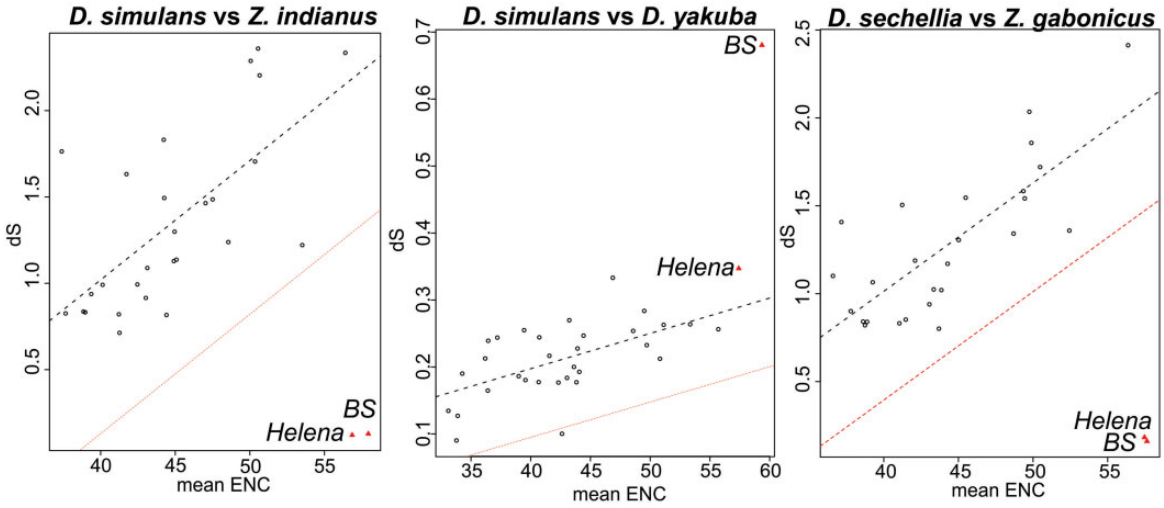

Fig. 6.—ENC–dS correlation graph obtained from the comparison between *D. simulans* versus *Z. indianus*, *D. simulans* versus *D. yakuba*, and *D. sechellia* versus *Z. gabonicus* representing inferences of HT and VT. Black empty circles represent the 30 host genes used as controls for vertically transmitted genetic information, red solid triangles represent the *Helena* and *BS* ENC–dS plotted against the vertically inherited host genes, the dotted black lines represent the predicted distribution of the ENC–dS correlation between host genes derived from the observed data, and the dotted red lines represent the variance of the observed measurements. If the TE ENC–dS red triangle is plotted within the variance of the host data, then it is not significantly different from the host genes and is considered vertically transmitted. In contrast, if it is plotted far from the dotted red line, then it is significantly different from the host genes, and therefore, will be considered horizontally transferred between the two species. dS is the number of synonymous substitutions per synonymous site, and ENC is the effective number of codons (according to Wallau et al. 2016).

of *Helena* and *BS* in *Z. bogoriensis*, which belongs to the *Anaprionus* subgenus, and of *BS* in four species of the *Zaprionus* subgenus). The discontinuous distribution scenario could be explained by two hypotheses. The first hypothesis assumes the presence of *Helena* and *BS* in the ancestor of the two subgenera of the genus *Zaprionus* and later losses of both elements in the subgenus *Anaprionus*, at least in *Z. bogoriensis*, and of *BS* also in some species of the subgenus *Zaprionus*. These losses could have occurred gradually due to either the accumulation of mutations, which would indicate a process of extinction of the TE, or genetic drift. Because the only member of the subgenus *Anaprionus* available for analysis was *Z. bogoriensis*, we could not test the first hypothesis. The second hypothesis assumes the complete absence of *Helena* and *BS* in the ancestral of the genus *Zaprionus* and recent introduction in the subgenus *Zaprionus* by HTT. In the case of HTT into the subgenus *Zaprionus*, the introduction of both elements would have occurred concomitantly with the divergence of the species belonging to this subgenus in Africa. However, we did not discard the possibility that the Asian species carry very divergent *Helena* and *BS* sequences, but unfortunately, except for *Z. bogoriensis* the Asian species were not available for analysis. Even though, this does not invalidate our second hypothesis, because it was formulated for the sharing of similar *Helena* and *BS* sequences between species of *Zaprionus* and those of the subgroup *melanogaster*. We tested this hypothesis using phylogenetic reconstruction, networks and the VHICA method and investigated whether both non-LTR retrotransposons were inserted into the ancestor of the subgenus *Zaprionus* by one or more independent HTT events.

The *Helena* Element

The *Helena* element was identified in all species of the African *Zaprionus* subgenus; however, it was not identified in *Z. bogoriensis*, the only species analyzed from the Oriental *Anaprionus* subgenus. As shown in the phylogeny, the clustering of all the *Helena* sequences sampled in species of the subgenus *Zaprionus* within the clade of the African species of the *melanogaster* group, on one hand, and that of *Helena* sequences of *Z. ornatus* together with those of *D. sechellia*, on the other hand, indicate that more than one HTT event occurred among the Tropical African drosophilids.

The divergence of the *Drosophila* and *Zaprionus* genera is estimated to have occurred between 40 and 60 Mya (Russo et al. 1995; Yassin et al. 2008). As illustrated in figure 1, the genus *Zaprionus* diversified on the Oriental region during the Quaternary period at approximately 7 Mya and from there, a lineage would have migrated to the islands of the Indian Ocean and to the African continent, where the diversification of the *Zaprionus* subgenus complexes is proposed to have occurred at approximately 4 Mya (Yassin et al. 2008). The clade that aggregates all the sequences of *Helena* from *Zaprionus* was estimated to be 4.1 My old (fig. 2), values consistent with the diversification of the subgenus in tropical Africa. Estimates of the origin of the element *Helena* are

therefore consistent with the period of diversification in the African continent of the subgenus *Zaprionus*, along with the divergence of the *melanogaster* complex (Lachaise and Silvain 2004; Tamura et al. 2004a, 2004b; Cutter 2008). The clustering of *Z. ornatus* with *D. mauritiana* and *D. sechellia* is also incongruent with the divergence time of the species included in this clade, whose *Helena* ancestral sequence was estimated to originate 0.4 Mya. The dating of this clade is similar to the divergence time of *D. mauritiana* and *D. sechellia* (0.4 Mya, Lachaise and Silvain 2004) and is much more recent than the diversification of the *Z. ornatus* species complex (4.4 Mya, Yassin et al. 2008).

The above-mentioned incongruities are also observable in the network in which the *Helena* sequences of the subgroup *melanogaster* and *Zaprionus* form a separate group from the other sequences of species that did not share an evolutionary period in Africa. However, due to the large number of median vectors between the sequences, the direction of the HTT cannot be clearly inferred, except for the *Helena* sequences of *D. sechellia* and *Z. ornatus*, for which a direct relationship of ancestry-descent exists (*D. sechellia* to *Z. ornatus*). In summary, our results suggest the occurrence of two HTT events of *Helena* among the species of the subgenus *Zaprionus* and the subgroup *melanogaster*. The first event would have occurred at approximately 4 Mya between the ancestor of the *melanogaster* complex and the ancestor of the subgenus *Zaprionus*, and the second would have occurred at <0.5 Mya and involved the transfer of *Helena* from *D. sechellia* to *Z. ornatus*.

The sequences of *Helena* of the Oriental species of the *melanogaster* group, *D. bipectinata* and *D. ananassae*, also suggest a relationship inconsistent with the species phylogeny of the *melanogaster* subgroup, as they clustered with sequences of the species belonging to the *repleta* and *virilis* groups. Although the species of the *repleta* group in which *Helena* was sampled are currently native to the Nearctic region, the ancestor of the group evolved in Asia, from whence it migrated to the Americas at approximately 30 Mya (Throckmorton 1982), as shown in figure 1. This grouping suggests another HTT event between ancestors of subgroups of the *melanogaster*, *virilis* and *repleta* groups.

The *BS* Element

Similar to *Helena*, *BS* presented a discontinuous distribution pattern: it was found in six of the 11 species of *Zaprionus* (belonging to the complexes *sepsoides*, *davidi* and *ornatus*) and in 14 of the 21 *Drosophila* species tested, including only two of the 11 Oriental species of the *melanogaster* group. Two basal incongruities call attention to the phylogenetic tree of *BS*. The first is due to the *BS* sequences of the *melanogaster* complex clustering more closely with the sequences of the *Zaprionus* subgenus than with those of *D. erecta*, which belongs to the same subgroup (fig. 3). The clade that combines these sequences is dated at 2.6 Mya, which is near the period of diversification of the species of the complex *melanogaster* (Lachaise and Silvain 2004; Tamura et al. 2004a, 2004b; Cutter 2008). This phylogenetic incongruence, the distance values, the network and the statistical results provided by the VHICA analysis support our hypothesis of the occurrence of one or more HTT events of the element *BS* between the ancestor of the *melanogaster* complex and ancestors of the complexes belonging to the *inermis* and *vittiger* groups, which evolved from 3.9 Mya in the Indian Ocean Islands and from 4.4 Mya in Central Africa, respectively (Yassin et al. 2008; Yassin and David 2010). In addition, the polyphyly inside the clade *Zaprionus* and the relatively short time of species divergence makes incomplete lineage sorting an equally probable explanation for the phylogenetic incongruities within this clade.

Two additional *BS* HTT events were detected in our phylogenetic analyses. The sequences of the species *D. persimilis* and *D. pseudoobscura* of the group *obscura* (subgenus *Sophophora*) are grouped more closely with those of *D. mojavensis* (subgenus *Drosophila*). These sequences form a clade with high support that are grouped with the sequences of the *melanogaster-Zaprionus* subgroup dated at 12.5 Mya. The sequences of the Oriental species *D. bipectinata* and *D. ficusphila* of the *melanogaster* group are grouped basally to this clade. These species belong to three subgenera and five different species groups (subgenus *Sophophora*: *melanogaster* and *obscura* groups; subgenus *Drosophila*: *repleta* group; and subgenus *Zaprionus*: *vittiger* and *inermis* groups) that share a recent common ancestral *BS* sequence dated at 18.4 Mya. This date coincides with that of the migration of the proto-*melanogaster* founder population to Africa (between 17 and 20 Mya) from the Eastern region (Lachaise and Silvain 2004). Although *Drosophila* species of the *pseudoobscura* subgroup and the *repleta* group are native to the New World, their ancestors are of Asian origin. The *obscura* group would have been subdivided into *obscura* and *pseudoobscura* subgroups while still in Asia, and the latter would have been introduced in the Late Miocene in the Americas, at approximately 13 Mya (Russo et al. 1995). The HTTs of *BS* among the ancestral strains of one or more subgroups of the *melanogaster* group to the ancestor of the *pseudoobscura* subgroup could have occurred, and from this, HTT could have subsequently occurred to species of the *repleta* group (here represented by *D. mojavensis*) in the Americas (fig. 1).

Genomic Promiscuity between the *melanogaster* Complex and the Subgenus *Zaprionus*

We present robust evidence supporting the hypothesis of the occurrence of several HTT events involving the non-LTR retrotransposons *Helena* and *BS* between species of the *melanogaster* complex and species of the subgenus *Zaprionus*. *Helena* and *BS* were present in the *melanogaster* group

before the divergence of the African subgroup. Because these species have evolved in Tropical Africa over the past 10 My, several opportunities for HTT must have occurred, allowing the transfer of TEs with a low probability of invasion of new genomes.

Importantly, the non-LTR retrotransposons examined in this study were not the only participants in HTTs between these species in this period. Some studies previously showed evidence of HTTs between species of the *melanogaster* and *Zaprionus* involving the LTR retrotransposons *Tom*, *297*, *17.6, rover* (Vidal et al. 2009), *Gypsy* (Herédia et al. 2004; De Setta et al. 2009), *Micropia* (De Setta et al. 2009), *Copia* (De Setta et al. 2011), and DNA transposons *Mariner* (Maruyama and Hartl 1991), Mos1-Like (Brunet et al. 1999) and hAT (Deprá et al. 2010). In most of these studies, only one species of *Zaprionus* (*Z. indianus*) and few TEs were studied, and the direction of the HTT was unclear. We previously proposed that three *Gypsy* variants and one *Micropia* variant, both LTR retrotransposons, invaded the genomes of species of the subgenus *Zaprionus* in waves of HTT from the ancestor of the *melanogaster* group between 0.6 and 10 Mya (De Setta et al. 2009). Earlier still, at approximately 11 Mya, an invasion of the LTR retrotransposon *Copia* occurred in the genus *Zaprionus* from an ancestor of the group *melanogaster* (De Setta et al. 2011). This element is even present in *Z. bogoriensis* (subgenus *Anaprionus*), which does not have *Helena*, *BS*, *Gypsy*, or *Micropia*, reinforcing HTT at an earlier point in the evolution of these drosophilids.

The results of the above-cited studies showed that after invasion, the LTR retrotransposons maintained their transposition activity and participated in some HTT events within the subgenus *Zaprionus*. In this study, we also identified putatively complete copies of at least *BS* in the subgenus *Zaprionus* through analysis of the integrity of the protein domains. Because the number of sequenced *Zaprionus* genomes remains very low, we cannot exclude the possibility that *BS* and *Helena* elements are still transpositionally active in this subgenus or that they had been active until recently. Together, the data presented herein reveal a panorama of extensive genetic material exchange between species of the subgroup *melanogaster* and those of the subgenus *Zaprionus* during the origin and diversification of the species in Tropical Africa between 4 and 1 Mya. This findings also support the hypothesis of permissiveness to the fixation of transferred TEs in the genomes of newly diversified species, mainly associated with small effective population sizes, which could reduce the efficacy of natural selection against invasive DNA (reviewed in Carareto 2011).

Conclusions and Perspectives

The importance of this study goes beyond stating whether the HT of TEs occurred between species of the subgenera *Sophophora* and *Zaprionus*. The in-depth reconstruction of the timeframe and geography of these events allowed us to reaffirm a historical scenario for the evolution of the genomes of these groups of species. For the first time, we identified the HTT of non-LTR retrotransposon in *Zaprionus*. The growth of genomic data for nonmodel species and high-throughput genomic analysis, which were combined in this study, corroborate recent data indicating that the HTT of non-LTR retrotransposons is not as rare as previously thought. Knowledge of the period of species diversification, the life history and the environment in which they evolved facilitates an understanding of not only population history but also the processes of diversification for one of the most important known sources of genetic diversity: transposable elements.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This research was supported by a grant from the Brazilian agency FAPESP—Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant number 2016/19271-2 awarded to CMAC), by APEGE CNRS DynET and BQR Université Lyon1 EC Grants awarded to AH, and by CNPq-Conselho Nacional de Desenvolvimento Científico e Tecnológico (Grant number 303455/2017-9) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) fellowships awarded to CMAC and MCS, respectively. We thank J.R. David for kindly providing the *Zaprionus* lines, G.L. Wallau for supporting the VHICA analyses, and I.L. Tambones for helping with the design of the figures. We also thank two anonymous reviewers for their valuable contributions.

Literature Cited

De Almeida LM, Carareto CM. 2006. Sequence heterogeneity and phylogenetic relationships between the copia retrotransposon in *Drosophila* species of the *repleta* and *melanogaster* groups. Genet Sel Evol. 38(5):535–550.

Altschul SF, et al. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 62(11):1182–1186.

Anxolabéhère D, Kidwell MG, Periquet G. 1988. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile *P* elements. Mol Biol Evol. 5(3):252–269.

Baidouri ME, et al. 2014. Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res. 24(5):831–838.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16(1):37–48.

Brégliano JC, Kidwell MG. 1983. Hybrid dysgenesis determinants. In: Shapiro J, editor. Mobile genetic elements. New York: Academic Press. p. 363–410.

Brunet F, Godin F, Bazin C, Capy P. 1999. Phylogenetic analysis of Mos1-like transposable elements in the Drosophilidae. J Mol Evol. 49(6):760–768.
Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan D. 1984. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the *I* factor. Cell 38(1): 153–163.
Carareto CM. 2011. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. Mob Genet Elements 1(3):179–186.
Clark JB, Maddison WP, Kidwell MG. 1994. Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol Biol Evol. 11(1):40–50.
Cordaux R, Hedges DJ, Batzer MA. 2004. Retrotransposition of Alu elements: how many sources? Trends Genet. 20(10):464–467.
Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. Mol Biol Evol. 25(4):778–786.
Daniels SB, Peterson KR, Strausbaugh KMG, Chovnicl 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 USA. 81(21):6794–6797.
Deprá M, Panzera Y, Ludwig A, Valente VLS, Loreto ELS. 2010. *hosimary*: a new *hAT* transposon group involved in horizontal transfer. Mol Genet Genomics 283(5):451–459.
Diao X, Freeling M, Lisch D. 2006. Horizontal transfer of a plant transposon. PLoS Biol. 4(1):e5.
Dias ES, Carareto CMA. 2012. Ancestral polymorphism and recent invasion of transposable elements in *Drosophila* species. BMC Evol Biol. 12(1):119.
Dias ES, et al. 2015. Large distribution and high sequence identity of a *Copia-type* retrotransposon in angiosperm families. Plant Mol Biol. 89(1–2):83–97.
Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29(8):1969–1973.
Evgen'ev M, Zelentsova H, Mnjoian L, Poluectova H, Kidwell MG. 2000. Invasion of *Drosophila virilis* by the *Penelope* transposable element. Chromosoma 109(5):350–357.
Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat. 125(1):1–15.
Friesen PD, Nissen MS. 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the *baculovirus* genome. Mol Cell Biol. 10(6):3067–3077.
Gao D, et al. 2018. Horizontal transfer of non-LTR retrotransposons from arthropods to flowering plants. Mol Biol Evol. 35(2):354–364.
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(7293):1347–1350.
Goubert C, et al. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Genome Biol Evol. 7(4):1192–1205.
Granzotto A, Lopes FR, Lerat E, Vieira C, Carareto CMA. 2009. The evolutionary dynamics of the *Helena* retrotransposon revealed by sequenced *Drosophila* genomes. BMC Evol Biol. 9(1):174.
Granzotto A, Lopes FR, Vieira C, Carareto CMA. 2011. Vertical inheritance and bursts of transposition have shaped the evolution of the *BS* non-LTR retrotransposon in *Drosophila*. Mol Genet Genomics 286(1):57–66.
Herédia F, Loreto ELS, Valente VLS. 2004. Complex evolution of *gypsy* in drosophilid species. Mol Biol Evol. 21(10):1831–1842.
Jeffs PS, Holmes EC, Ashburner M. 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol Biol Evol. 11(2):287–304.
Jowett T. 1986. Preparation of nucleic acids. In: Roberts DB, editor. *Drosophila*: a practical approach, vol. 335. Washington (DC): Oxford. p. 275–277.
Katoh K, Rozewicki J, Yamada KD. 2017. Article Navigation MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. bbx108.
Kidwell MG. 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 80(6): 1655–1659.
Kim AI, et al. 1994. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc Natl Acad Sci USA. 91(4):1285–1289.
Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 16(2):111–120.
Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. 2015. The recent invasion of natural *Drosophila simulans* populations by the *P*-element. Proc Natl Acad Sci USA. 112(21):6659–6663.
Kordis D, Gubensek F. 1998. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes. Proc Natl Acad Sci USA. 95(18):10704–10709.
Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870–1874.
Lachaise D, Silvain JF. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: *the Drosophila melanogaster–D. simulans* palaeogeographic riddle. Genetica 120(1–3):17–39.
Loreto ELS, Carareto CMA, Capy P. 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. Heredity (Edinb) 100(6):545–554.
Ludwig A, Valente VL, Loreto EL. 2008. Multiple invasions of Errantivirus in the genus *Drosophila*. Insect Mol Biol. 17(2):113–124.
Lyozin GT, et al. 2001. The structure and evolution of *Penelope* in the *virilis* species group of *Drosophila*: an ancient lineage of retroelements. J Mol Evol. 52(5):445–456.
Maruyama K, Hartl DL. 1991. Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus*. J Mol Evol. 33(6):514–524.
Metcalfe CJ, Casane D. 2014. Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements. 5:19.
Mizrokhi LJ, Mazo AM. 1990. Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. Proc Natl Acad Sci USA. 87(23):9216–9220.
Modolo L, Lerat E. 2015. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. BMC Bioinformatics 16:137.
Modolo L, Picard F, Lerat E. 2014. A new genome-wide method to track horizontally transferred sequences: application to *Drosophila*. Genome Biol Evol. 6(2):416–432.
Morales-Hojas R, Vieira CP, Vieira J. 2006. The evolutionary history of the transposable element *Penelope* in the *Drosophila virilis* group of species. J Mol Evol. 63(2):262–273.
Okada T, Carson HL. 1983. The genera *Phorticella* DUDA and *Zaprionus* COQUILLETT (Diptera, Drosophilidae) of the Oriental region and New Guinea. Jpn J Entomol 51:539–553.
Pace JK, II, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. Proc Natl Acad Sci USA. 105(44):17023–17028.
Palazzo A, Caizzi R, Viggiano L, Marsano RM. 2017. Does the promoter constitute a barrier in the horizontal transposon transfer

process? Insight from *Bari* Transposons. Genome Biol. Evol. 9(6):1637–1645.
Peccoud J, Loiseau V, Cordaux R, Gilbert C. 2017. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci USA. 114(18):4721–4726.
Pélisson A, et al. 1994. *Gyspy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. EMBO J. 13:4401–4411.
Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER. 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proc Natl Acad Sci USA. 92(17):8050–8054.
Petrov DA, Chao Y-C, Stephenson EC, Hartl DL. 1998. Pseudogene evolution in *Drosophila* suggests a high rate of DNA loss. Mol Biol Evol. 15(11):1562–1567.
Ray DA, et al. 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. Genome Res. 18(5):717–728.
Rebollo R, Lerat E, Kleine LL, Biémont C, Vieira C. 2008. Losing *Helena*: the extinction of a *Drosophila* line-like element. BMC Genomics 9:149–160.
Robertson HM, Zumpano KL. 1997. Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. Gene 205(1–2):203–217.
Romero-Soriano V, Guerreiro MPG. 2016. Expression of the retrotransposon *Helena* reveals a complex pattern of TE deregulation in *Drosophila* hybrids. PLoS One 11(1):e0147903.
Russo CAM, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol. 12(3):391–404.
Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in *Drosophila*. Trends Genet. 21(4):200–203.
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(9):537–546.
De Setta N, Van Sluys MA, Capy P, Carareto CM. 2009. Multiple invasions of *Gypsy* and *Micropia* retroelements in genus *Zaprionus* and *melanogaster* subgroup of the genus *Drosophila*. BMC Evol Biol. 9:279–297.
De Setta N, Van Sluys MA, Capy P, Carareto CMA. 2011. *Copia* retrotransposon in the *Zaprionus* genus: another case of transposable element sharing with the *Drosophila melanogaster* subgroup. J Mol Evol. 72(3):326–338.
Sharp PM, Li WH. 1989. On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol. 28(5):398–402.
Sessegolo C, Burlet N, Haudry A. 2016. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. Biol Lett. 12(8):20160407.
Silva JC, Kidwell MG. 2000. Horizontal transfer and selection in the evolution of *P* elements. Mol Biol Evol. 17(10):1542–1557.
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. Genes Dev. 8(17):2046–2057.
Suh A, et al. 2016. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. Nat Commun. 7:11396.
Tamura K, Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10(3):512–526.
Tamura K, Nei M, Kumar S. 2004a. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA. 101(30):11030–11035.
Tamura K, Subramanian S, Kumar S. 2004b. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol Biol Evol. 21(1):36–44.
Thomas T, et al. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J. 4(12):1557–1567.
Throckmorton LH. 1982. Pathways of evolution in the genus *Drosophila* and the founding of the *Repleta* group. In: Barker JSF and Starmer WT (eds). Ecological Genetics and Evolution – The Cactus-Yeast *Drosophila* Model System. Academic Press, Sydney, p. 33–47.
Udomkit A, Forbes S, Daigleish G, Finnegan DJ. 1995. *BS* a novel LINE-like element in *Drosophila melanogaster*. Nucleic Acids Res. 23(8):1354–1358.
Vidal NM, Ludwig A, Loreto ELS. 2009. Evolution of *Tom*, *297*, *17.6* and *rover* retrotransposons in Drosophilidae species. Mol Genet Genomics 282(4):351–362.
Wallau GL, Capy P, Loreto E, Rouzic AL, Hua-Van A. 2016. VHICA, a New method to discriminate between vertical and horizontal transposon transfer: application to the *mariner* family with in *Drosophila*. Mol Biol Evol. 33(4):1094–1109.
Wallau GL, Ortiz MF, Loreto ELS. 2012. Horizontal transposon transfer in eukarya: detection, bias, and perspectives. Genome Biol Evol. 4(8):689–699.
Walsh AM, Kortschak RD, Gardner MG, Bertozzi T, Adelson DL. 2013. Widespread horizontal transfer of retrotransposons. Proc Natl Acad Sci USA. 110(3):1012–1016.
Wicker T, et al. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 8(12):973–982.
Yassin A, et al. 2008. Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Diptera: Drosophilidae). Mol Phylogenet Evol. 47(3):903–915.
Yassin A, David JR. 2010. Revision of the Afrotropical species of *Zaprionus* (Diptera, Drosophilidae), with descriptions of two new species and notes on internal reproductive structures and immature stages. Zookeys 51:33–72.

Associate editor: Ellen Pritham