

RESEARCH

Open Access

# High frequency of horizontal transfer in Jockey families (LINE order) of drosophilids



Izabella L. Tambones<sup>1</sup>, Annabelle Haudry<sup>2</sup>, Maryanna C. Simão<sup>1</sup> and Claudia M. A. Carareto<sup>1\*</sup> 

## Abstract

**Background:** The use of large-scale genomic analyses has resulted in an improvement of transposable element sampling and a significant increase in the number of reported HTT (horizontal transfer of transposable elements) events by expanding the sampling of transposable element sequences in general and of specific families of these elements in particular, which were previously poorly sampled. In this study, we investigated the occurrence of HTT events in a group of elements that, until recently, were uncommon among the HTT records in *Drosophila* – the Jockey elements, members of the LINE (long interspersed nuclear element) order of non-LTR (long terminal repeat) retrotransposons. The sequences of 111 Jockey families deposited in Repbase that met the criteria of the analysis were used to identify Jockey sequences in 48 genomes of Drosophilidae (genus *Drosophila*, subgenus *Sophophora*: melanogaster, obscura and willistoni groups; subgenus *Drosophila*: immigrans, melanica, repleta, robusta, virilis and grimshawi groups; subgenus *Dorsilopa*: busckii group; genus/subgenus *Zaprionus* and genus *Scaptodrosophila*).

**Results:** Phylogenetic analyses revealed 72 Jockey families in 41 genomes. Combined analyses revealed 15 potential HTT events between species belonging to different genera and species groups of Drosophilidae, providing evidence for the flow of genetic material favoured by the spatio-temporal sharing of these species present in the Palaearctic or Afrotropical region.

**Conclusions:** Our results provide phylogenetic, biogeographic and temporal evidence of horizontal transfers of the Jockey elements, increase the number of rare records of HTT in specific families of LINE elements, increase the number of known occurrences of these events, and enable a broad understanding of the evolutionary dynamics of these elements and the host species.

**Keywords:** transposable elements, HTT, *Drosophila*, *Zaprionus*, *Scaptodrosophila*, VHICA

## Background

Transposable elements (TEs) are abundant mobile repetitive sequences in eukaryotes that are capable of moving from one place to another in the genome. The introduction of a TE into a *naïve* genome may occur by three mechanisms: (i) *de novo*, by means of recombination between pre-existing and degraded TE sequences that restores their coding capacity and transpositional activity; (ii) introgression of sequences consequent to reproduction between closely related species; and (iii) horizontal transfer (HT) of TE sequences between different species via a non-sexual route. After such introduction, the TEs can

proliferate both throughout the host genome and throughout the population by vertical transfer (VT) from parents to offspring. During this phase, it is possible for the TE to be transferred horizontally to another genome or lost as a result of random processes or of purifying selection due to deleterious effects of the insertion (for a review, see [1–3]). TEs may escape extinction during vertical transmission across generations if HT occurs [4].

Horizontal transfer of TEs (HTT), a quite frequent phenomenon in prokaryotes, has been increasingly documented in eukaryotes. Among the events recorded in metazoans, most are reported in Drosophilidae, a Diptera family widely used to study TEs since their discovery as important sources of genetic variability and genomic evolution. Since the first report of HTT in the 1980s, i.e., the transfer of the *P* element from *Drosophila willistoni* to *D. melanogaster* [5, 6], the number of such

\* Correspondence: [claudia.carareto@unesp.br](mailto:claudia.carareto@unesp.br)

<sup>1</sup>Department of Biology, Institute of Biosciences, Humanities and Exact Sciences (IBILCE), UNESP – São Paulo State University, Campus São José do Rio Preto, São Paulo, SP 15054-000, Brazil

Full list of author information is available at the end of the article



studies in drosophilids has increased, doubling from 101 in 2008 (reviewed in [1, 2]) to 218 in 2010 [3], and 243 reports are currently documented [7, 8]. This increasing frequency illustrates how advances in comparative analyses of complete genomes improved the probability of HTT identification and revealed that the rate of HTT is variable between the different types of TEs. Of the 243 cases reported, 126 (51.9%) are of DNA transposons, 103 (42.4%) are of LTR (long terminal repeat) retrotransposons, and only 14 (5.8%) involve non-LTR retrotransposons. For non-LTR retrotransposons in particular, only six families of the LINE order are involved, such as *Jockey* (*D. melanogaster*/*D. funebris*), *F* (*D. melanogaster*/*D. yakuba*), *Doc* (*D. melanogaster*/*D. yakuba*), *I* (*D. simulans*/*D. melanogaster*) (reviewed in [1, 2, 8]; HTT-DB platform), and *Helena* and *BS* (*melanogaster* complex/subgenus *Zaprionus*) [9]. The same marked difference in the HTT rate of different types of TEs was also recently detected by Reiss et al. [10], in a study of 460 species covering 19 orders of arthropods that focused on one LTR retrotransposon (*Copia*), one non-LTR retrotransposon (*Jockey*) and one DNA transposon (*Mariner*). The authors found that the HTT rate of three widely distributed TE superfamilies in Arthropoda, namely, *Mariner* (DNA transposons), *Copia* (LTR retrotransposons) and *Jockey* (non-LTR retrotransposons), which differ in their transposition mode, was 52, 37 and 10%, respectively. Interestingly, in contrast to having a low HTT rate, the *Jockey* superfamily is the most diverse in terms of the number of families in each genome (~15 families on average per species versus 4 and 3.5 for *Copia* and *Mariner*, respectively) and species distribution (252 species versus 174 and 129 for *Copia* and *Mariner*, respectively). Also interesting is that although arthropods generally present a low rate of HTT, butterflies and moths show a large excess of HTT: 56 events in Lepidoptera compared with the expected average of 13.7. In contrast, only two events were registered in Drosophilidae, one involving *D. melanogaster* and the other in *D. kikkawai*, two species belonging to the *melanogaster* group of the genus *Drosophila*.

*Jockey* is a superfamily of non-LTR retrotransposons found only in Arthropoda. The full-length element is ~5 kb in size [11, 12] and is formed by two ORFs (open reading frames), the first with 568 aa residues and the second with 916 aa residues. In contrast to ORF1, ORF2, which encodes an apurinic endonuclease (APE) and a reverse transcriptase (RT), is well conserved, shows tree-like evolution [13] and is very appropriate for phylogenetic analyses. The low rate of HTT reported for this superfamily has been associated with the target-primed reverse transcription mode of transposition of the non-LTR elements, in which the cDNA strand is reverse-transcribed from an RNA template directly onto a

chromosomal target site [14], thus not producing extra-chromosomal copies in the stable form of DNA.

The current study aimed to further investigate the rate of HTT in the *Jockey* superfamily, focusing on the occurrence of its families in drosophilids. For a definition of family, we followed Wicker et al. [15]: a family is a group of highly conserved TEs with a sequence similarity of  $\geq 80\%$  in at least 80% of their coding region in an aligned region of at least 80 nucleotides. The term clade (instead of superfamily) is used by the Repbase database [16], the most commonly used database for TE classification. Repbase also uses a somewhat different criterion from the TE hierarchical classification of Wicker et al. [15], based on studies of enzymology, structural similarities and relationships between sequences [17]. As an example, the 123 *Jockey* sequences within the *Jockey* clade are distributed in up to 15 families in each *Drosophila* species (named *Jockey-1* to *Jockey-15*) in Repbase, and four *Jockey* families were recorded in *D. simulans* (*Jockey-1\_DSim*, *Jockey-2\_DSim*, *Jockey-3\_DSim*, and *Jockey-4\_DSim*). Because of the nomenclature criterion used by Repbase, the family *Jockey-1\_DSim* has the same name (*Jockey-1*) but may not be closely related to the *Jockey-1* family of other *Drosophila* species. Therefore, the criteria for family denomination used by Repbase does not allow identifying the sharing of families between species, either by HT or by VT.

It is important to highlight, as considered by Wicker et al. [15], that “the precise definition of a family is problematic because groups of TEs with similar features sometimes form a continuum of sequence homology; [...] it seems that evolutionary lineages are sufficiently distinct to allow the borders of such a continuum to distinguish a family”. Because a family of TEs is a member of an evolutionary lineage, it is essential to consider that its sequences may not be restricted to a single species but shared by a group of species through VT or HTT and to use phylogenetic approaches to identify families. Given the paucity of HTT records for *Jockey* families in drosophilids (only three cases has been reported so far), one may wonder if the elements of this superfamily are less prone to HTT than elements from other families for which more HTT events have been documented [8]. This study aimed to broaden our understanding of the evolution of *Jockey* families in drosophilids and to provide an estimate of the rate of HTT involving these families by using phylogenetic approaches. To do so, we sampled 48 genomes of Drosophilidae for the occurrence of *Jockey* sequences by using the 111 families deposited in Repbase as a query.

## Results

A broad homology-based search using 111 downloaded consensus sequences of *Jockey* families deposited in the

Rebase database allowed us to identify Jockey sequences in 44 genomes of *Drosophilidae* species (Additional file 1: Table S1) belonging to the genera *Drosophila* (subgenera *Sophophora*, *Drosophila*, and *Dorsilopha*), *Zaprionus* (subgenus *Zaprionus*) and *Scaptodrosophila* (Additional file 1: Table S2). Sequences not yet catalogued in Rebase (Additional file 1: Table S3) were annotated in 28 species. In the subgenus *Sophophora*, new Jockey sequences were annotated in 12 genomes: five in species in the melanogaster group (*D. biarmipes*, *D. mauritiana*, *D. malerkotliana*, *D. serrata*, and *D. suzukii*) and seven in the *obscura* group (*D. athabasca*, *D. guanche*, *D. lowei*, *D. miranda*, *D. obscura*, *D. persimilis*, and *D. pseudoobscura*). In the subgenus *Drosophila*, 11 genomes had new Jockey sequences: two in species of the *immigrans* group (*D. albomicans* and *D. nasuta*), three in the *melanica* group (*D. melanica*, *D. micromelanica*, and *D. nigromelanica*), one in the *repleta* group (*D. mojavenensis*), three in the *virilis* group (*D. americana*, *D. montana*, and *D. novamexicana*) and two in the *robusta* group (*D. lacertosa* and *D. robusta*). In the *Dorsilopha* subgenus, only *D. busckii* presented Jockey sequences. In addition, three species of the *Zaprionus* genus (*Z. africanus*, *Z. gabonicus*, and *Z. indianus*) and one of the genus *Scaptodrosophila* (*S. lebanonensis*) presented Jockey sequences. Among these, the sequences of *D. malerkotliana*, *D. nasuta* and *D. robusta* were not used in the analyses due to their small sizes.

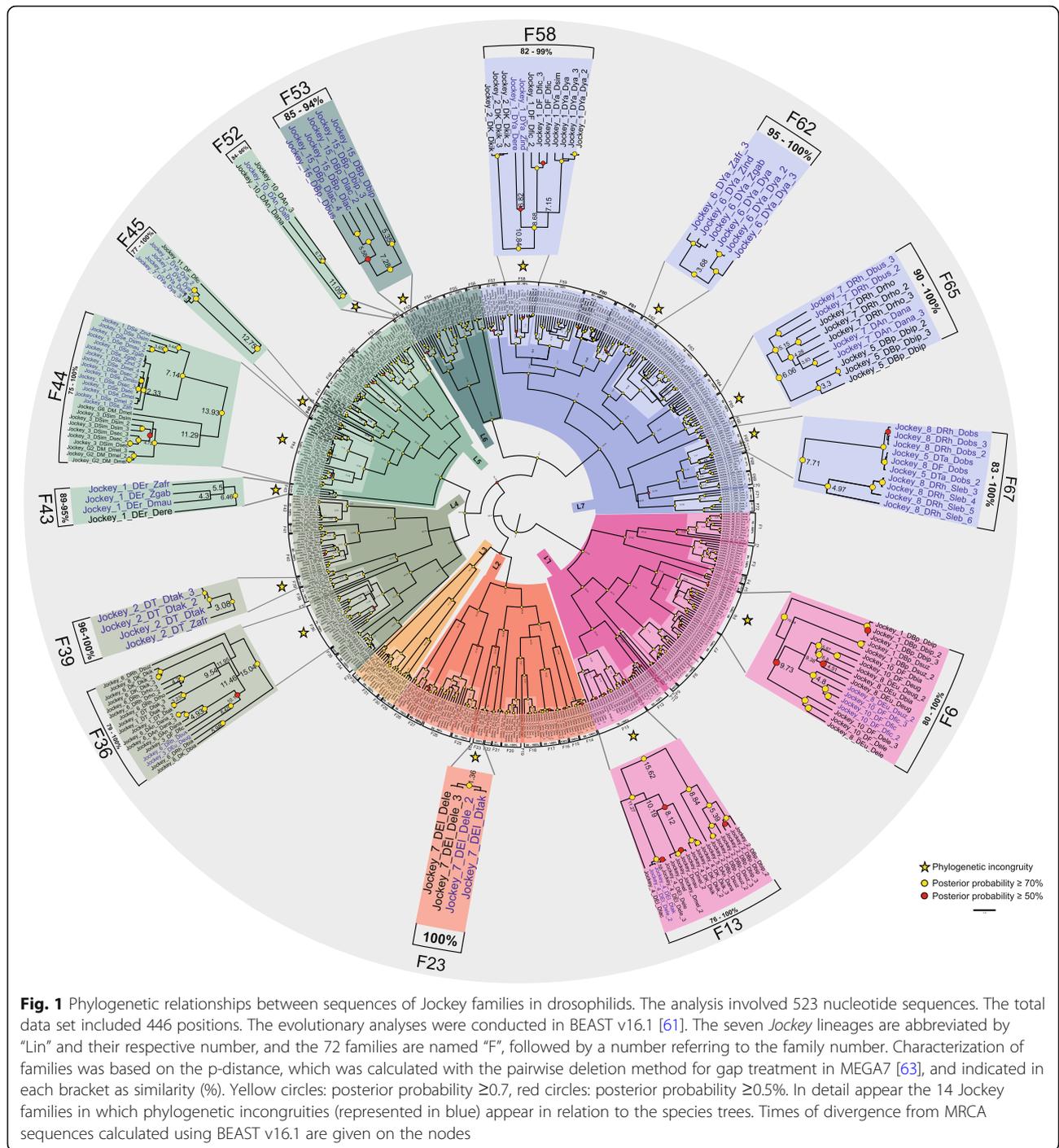
### Phylogenetic analysis of the jockey families

A phylogenetic tree was reconstructed using a multiple alignment of 523 RT sequences of Jockey families in 41 drosophilids that met the search criteria, in which the shortest sequence had 221 nt and the largest was 446 nt in length. To analyse the phylogeny, we considered the existence of Jockey evolutionary lineages and families. An evolutionary lineage is a large and robust monophyletic clade (posterior probability  $\geq 70\%$ ) formed by several families. A family is constituted by Jockey sequences of the same or different species that share nucleotide similarity  $\geq 80\%$  and are grouped with  $\geq 70\%$  support (Fig. 1). In practical terms, two clades of the same lineage constitute different families when the minimum similarity between them is 80%. Exceptions to this criterion were allowed in three cases (families F13, F36 and F44), where a sequence was grouped with  $\geq 75\%$  similarity to a sequence group that constituted a family of the same evolutionary lineage if the average similarity of this family was  $\geq 80$ . These criteria allowed us to classify all 523 *Drosophila* Jockey sequences into 72 families (F1 to F72), which were grouped into seven evolutionary lineages (Lin1 to Lin7). The number of families in each species varied from one (*D. athabasca*, *D. guanche*, *D. lowei*, *D. melanica*, *D. micromelanica*, *D. mojavenensis*, *D. nigromelanica*, and *D. willistoni*) to 14 (*D. ananassae*)

(Additional file 1: Table S3), indicating that the Jockey clade is very diverse in *Drosophila* genomes, as found in Arthropoda [10]. The divergence (p-distance) within (Additional file 2: Table S1) and between (Additional file 2: Table S2) the 72 Jockey families is given.

The most basal Jockey clade forms Lineage 1 (Fig. 1). This lineage is formed by 13 families, among which six families (F2, F4, F5, F9, F10 and F11) are formed by just one to three sequences of the same genome, which are divergent from the neighbouring clade (mean similarity: 62.5 to 75%). The other families include sequences of just one group or subgroup of species, such as the *virilis* group of the *Drosophila* subgenus (F1, F3, and F12) and the melanogaster subgroup of the *Sophophora* subgenus (F7). In addition, there are families formed by sequences of species belonging to different subgenera, such as the *robusta* group of the *Drosophila* subgenus and the *obscura* group of the *Sophophora* subgenus (F8), or to subgroups of the same species group (F6 and F13: subgroups of the melanogaster group). There is relative congruence between the phylogenetic relationships of the sequences within families F1, F3, F7, F8 and F12 and the species phylogeny (Fig. 1). However, F6 and F13 reveal incongruence with the species phylogeny. Family F6 includes two sequences of *D. suzukii*, one basal to the clade formed by *D. biarmipes*, *D. eugracilis*, *D. ficusphila* and *D. elegans* and the other closely related to sequences of *D. ficusphila* (average similarity = 96%), but not with the species of its own subgroup (*D. biarmipes*), representing the first phylogenetic incongruity in the tree. In addition, sequences of *D. elegans* and *D. takahashii* (F13) that belong to subgroups not closely related are clustered with a high degree of similarity (100%), representing the second incongruity.

Lineage 2 is also formed by 15 families (F14-F28), mostly represented by sequences of the *obscura* (F14, F15, F17, F20, F26 and F27) and *virilis* (F18, F25 and F28) groups, which follow the species phylogeny (Fig. 1). In this lineage, one sequence (Jockey\_3\_Dan\_Dana) forms a single divergent family (F24), which clusters with F23 with support below 70%. Among the families formed by sequences from the melanogaster group (F21, F22, F23 and F24), the Oriental subgroups show the third phylogenetic incongruence. As also seen in F13, in F23 the sequences of *D. elegans* and *D. takahashii* are clustered together with high similarity (100%), which is not consistent with the phylogenetic relationship between the two subgroups. In turn, Lineage 3 is formed by only four families (F29-F32) that include only sequences of the melanogaster group (African and Oriental subgroups), three of them formed by sequences of a single genome (F30-F32).



Next, in the phylogenetic tree, 10 families (F33-F42) form Lineage 4 (Fig. 1), and four of them are constituted by sequences of a single genome (F33, F35, F37 and F41). The other families include sequences of the subgenus *Sophophora* (F34: *obscura* and F38, F40: Oriental melanogaster subgroups) and the subgenus *Drosophila* (F42: *virilis* group), which generally follows the species phylogeny. On the other hand, two incongruities are seen in this

lineage. The clustering of sequences of *D. busckii* (*busckii* group) and *D. eugracilis* (*melanogaster* group) with high similarity (81%) in F36 reveals the fourth incongruity, considering that these sequences belong to species of two subgenera of the *Drosophila* genus, namely, *Dorsilopha* and *Sophophora*, respectively. The fifth incongruity can be seen in F39 because a sequence of *Z. indianus* (genus *Zapriionus*) clustered with sequences of *D. takahashii*

(genus *Drosophila*), with average similarity incompatible with the phylogenetic relationships between the two groups of species (98%).

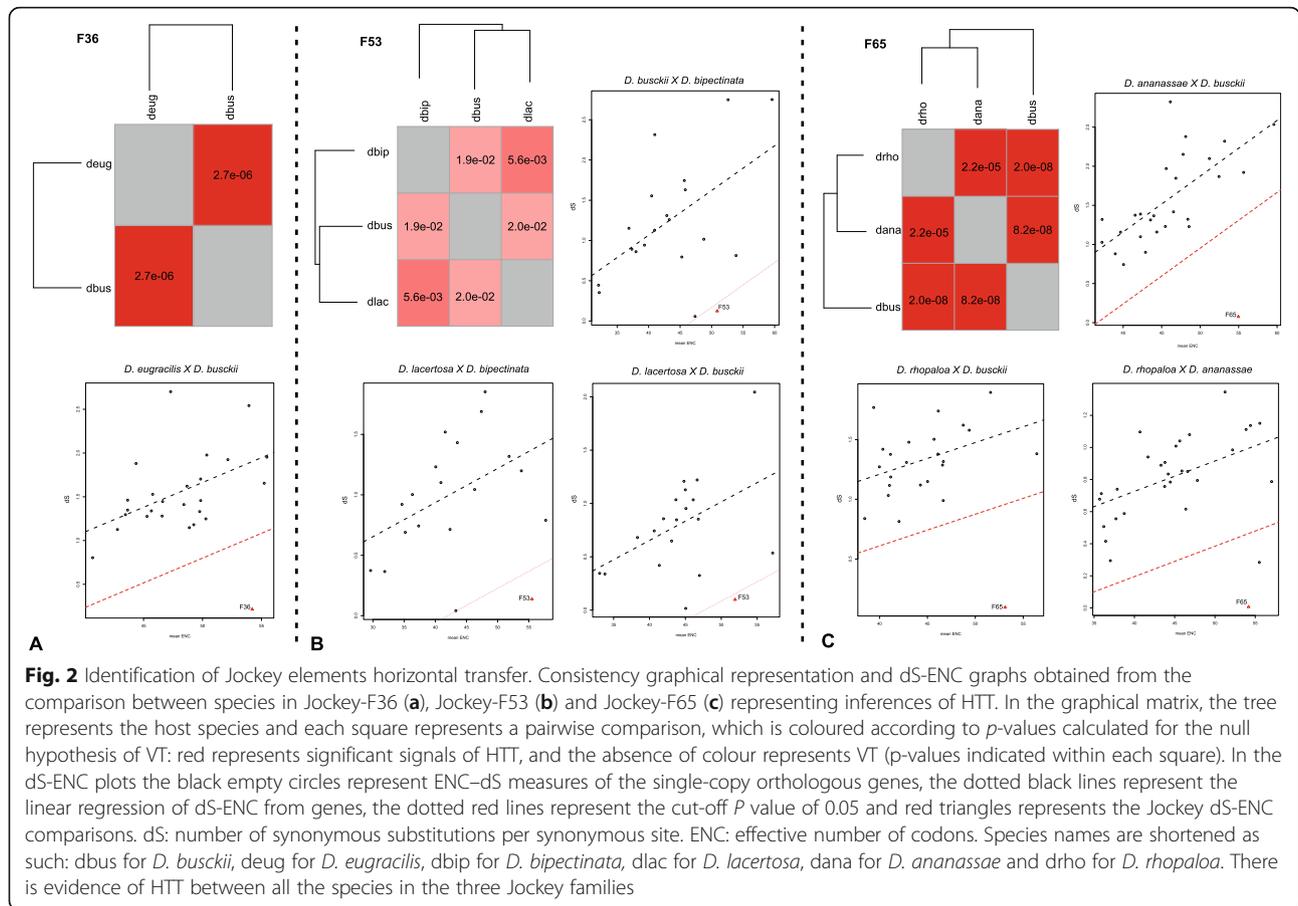
Ten families also compose Lineage 5 (F43-F52), of which four include the sequences of the melanogaster complex, all showing phylogenetic incongruities (Fig. 1). The sixth and seventh incongruities refer to *Zaprionus* species sequences: in F43, sequences of *Z. africanus* and *Z. gabonicus* cluster with a sequence of *D. mauritiana* (average similarity = 94%), and in F44, sequences of *Z. africanus*, *Z. gabonicus*, and *Z. indianus* cluster inside a clade with sequences of the melanogaster complex (average similarity = 95%). The eighth and ninth incongruities occur in F45 and F52, respectively, the first due to the clustering of a Jockey sequence of *D. simulans* inside a cluster of *D. yakuba* (average similarity = 98.4%) and the second involving two sequences of *D. ananassae* and a sequence of *D. albomicans* (immigrans group, *Drosophila* subgenus), which present a similarity equal to 86%. On the other hand, Lineage 6, which is composed of only five families (F53-F56), also presents an incongruity between species belonging to different subgenera in F53, the tenth incongruity of the phylogeny (Fig. 1), because the sequence of *D. busckii* is closely related (similarity = 85%) to sequences of *D. lacertosa* (robusta group, *Drosophila* subgenus).

The last lineage (Lin 7) is the largest (Fig. 1), composed of 15 Jockey families (F57-F72). The majority of Lin7 includes sequences of the melanogaster group (F57-F60, F63-F64, F68-F69 and F71-F72) or sequences clustered with those of other groups (F58, F62, F63 and F65), and sequences of the obscura group (F67) clustered with a sequence of *S. lebanonensis* and of *D. willistoni* (F70). Four incongruities can be observed in families that combine sequences of different species groups, subgenera or genera (F58, F62, F65 and F67). In the other families, the phylogenies of the Jockey sequences correspond relatively well to those of the species. The eleventh and twelfth incongruities involve sequences of *Zaprionus* clustered with sequences of species belonging to the melanogaster subgroup, the first with *D. erecta* (F58, similarity = 83%) and the second with *D. yakuba* (F62, similarity = 83%). The thirteenth and fourteenth incongruities occur in F65. The first is due to the clustering of *D. busckii* sequences with species of the Oriental melanogaster group with a similarity higher than expected (93%) between sequences belonging to species of two subgenera. The second is due to the clustering of *D. rhopaloea* and *D. ananassae* sequences, which belong to different subgroups (95%). Finally, in the fifteenth incongruity (F67), four sequences of *S. lebanonensis* cluster with sequences of *D. obscura* (similarity = 92%), thus involving two subgenera.

#### Identification of jockey elements horizontal transfer

Candidate cases of HTT were identified based on the 15 observed phylogenetic incongruities of Jockey elements (represented by stars in Fig. 1) from 14 families, hereafter named Jockey-F (1–72), belonging to six evolutionary strains. We used the VHICA (Vertical and Horizontal Inheritance Consistence Analysis) method based on comparison of dS rates of evolution and effective use of codons (ENC) to test for statistical support [18] of each inferred HTT event. Only the incongruities with a  $p$ -value < 0.01 were accepted as evidence for HTT: briefly, 15 phylogenetic incongruities were confirmed as HTTs. Among these, Fig. 2 shows the results of Jockey-F36, Jockey-F53 and Jockey-F65 obtained from dS-ENC pairwise comparisons between sequences of *D. busckii* and species of the Oriental melanogaster group with *D. eugracilis* (Jockey-F36), with *D. lacertosa*/*D. bipectinata* (Jockey-F53), and with *D. rhopaloea*/*D. ananassae* (Jockey-F65), respectively. The linear regression between the *D. busckii* dS and ENC of Jockey sequences and the host genes is significant in F36 (Fig. 2a), as well as in F53 (Fig. 2b) and F65 (Fig. 2c), confirming the hypotheses of HTT based on phylogenetic incongruities. However, in Jockey-F53, all dS and ENC comparisons revealed significant differences, suggesting the occurrence of HTT involving the three species; hence, instead of one HTT, we should compute two, the first possibly between *D. bipectinata* and *D. lacertosa* and the second between this species and *D. busckii*. In addition, the comparison between the sequences of *D. rhopaloea* and *D. ananassae* in F65 also shows a significant signal of HTT. The positions of the points in the plots are beyond the limit of variance in the genes, supporting the hypothesis of five HTT events in these families. The consistency graph reflects these results, in which each square that represents a pairwise comparison is coloured according to the coloured bar of  $p$ -values calculated for the null hypothesis of VT.

The results for the Jockey-F43-F44, F58 and F62 sequences were obtained from pairwise comparisons between species of the melanogaster subgroup and of the genus *Zaprionus*. The distribution of linear regressions between the dS and ENC values of Jockey-F44 sequences and of host genes of the species of the melanogaster complex (*D. melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana*) and of the genus *Zaprionus* (*Z. africanus*, *Z. indianus* and *Z. gabonicus*) is beyond the limit of variance in the genes, supporting the hypothesis of HTT between these species (Fig. 3). In contrast, between the *Zaprionus* species and between the species of the melanogaster complex, there is no signal of HTT, with a sole exception: the position of Jockey-F44 in the graph is beyond the limit of variance in the comparison between *D. melanogaster* and *D. sechellia*. This result suggests the

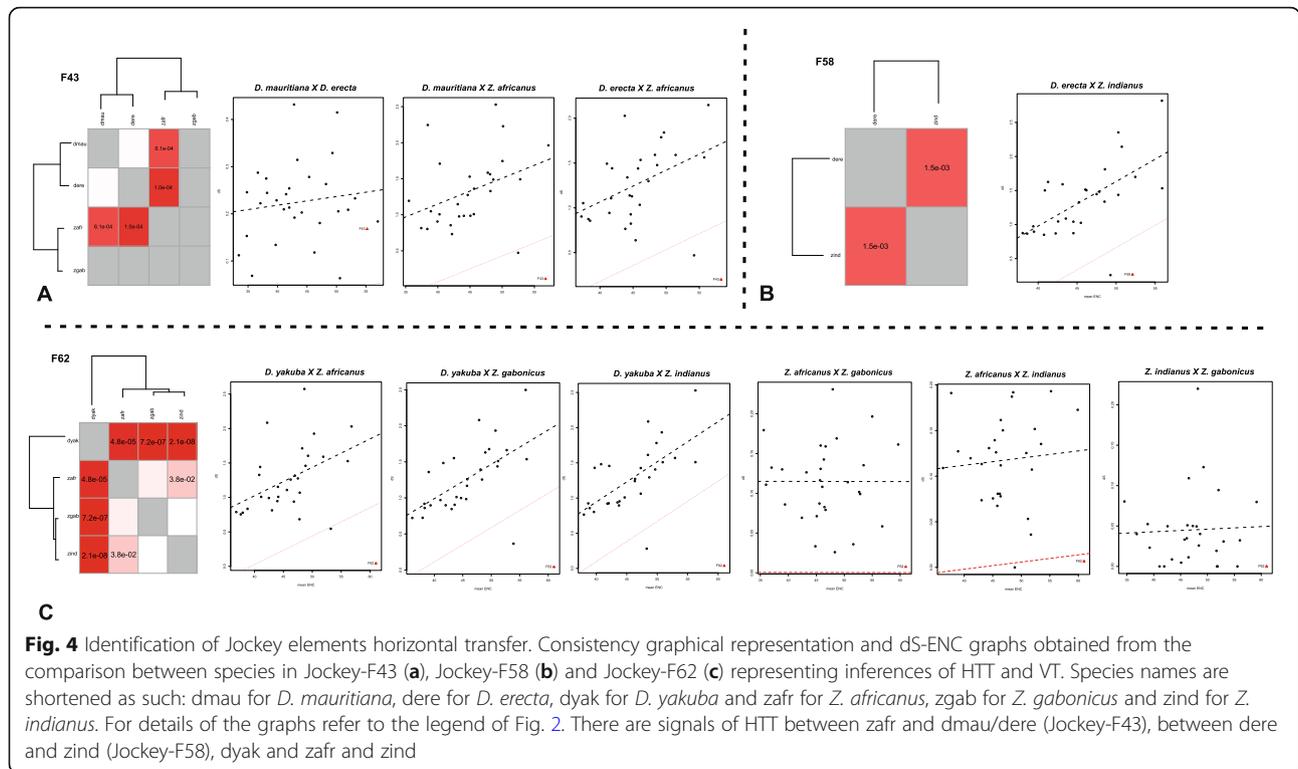


loss of Jockey-F44 ancestral copies in *D. sechellia* and re-introduction from *D. melanogaster*. Stochastic loss of mariner transposon has also been described in *D. sechellia*, as well as in other species of the melanogaster subgroup [19]. Similarly, the positions of the Jockey-F43 sequences of *Z. africanus* and both *D. erecta* and *D. mauritiana*, and *D. erecta* and *Z. indianus* (Fig. 4a), or also the *D. erecta* and *Z. indianus* sequences in F58 (Fig. 4b) are beyond the limit of variance in the genes, supporting the hypothesis of HTT between these species. A comparison between the Jockey-F43 sequences of *Z. gabonicus* with the others was not possible due to the length of the sequence, 189 bp, which resulted in only 63 codons. Additionally, significant signals of HTT are seen between the Jockey-F62 sequences of *D. yakuba* and the three *Zaprionus* species and between *Z. africanus* and *Z. indianus*, two sibling species [20] with very recent divergence (Fig. 4c). This signal of HTT is not supported by the significant difference criterion adopted here as well as could be a false-positive result because VHICA is not suitable for very closely related species [18], as well as due to incomplete lineage sorting; therefore, it was not considered HTT. The HTT signals between the three species of the genus *Zaprionus* and the species of the

melanogaster subgroup, but not between the species within these groups, with the two exceptions aforementioned, suggest the occurrence of a single transfer involving ancestors of these species groups in the four Jockey families, and one transfer between *D. melanogaster* and *D. sechellia*, for a total of five HTTs.

Four incongruities occur between species of the melanogaster group, one among species belonging to the African subgroup, *D. simulans* and *D. yakuba* (Jockey-F45, Additional file 3: Figure S1), and three between species belonging to the Oriental subgroups (Jockey-F6, Jockey-F13 and Jockey-F23, Additional file 3: Figure S2). The linear regression between the *D. simulans*-*D. yakuba* dS and ENC is highly significant, supporting the occurrence of HTT. Although the phylogenetic relationships between the Oriental subgroups of the melanogaster group are to some extent uncertain [21], two incongruities reported above (Jockey-F13 and Jockey-F23) were validated as HTTs (Additional file 3: Figure S2 B and C). From the last three incongruities (Jockey-F39, Jockey-F52 and Jockey-F67), only two, namely, those between *D. takahashii* and *Z. africanus* (Additional file 3: Figure S3 A) and between *D. obscura* and *S. lebanonensis* (Additional file 3: Figure S3 B),





were validated as HTTs, excluding that between *D. albicans* and *D. ananassae* (Additional file 3: Figure S3C). The last two results should be interpreted with caution because only 12 (Jockey-F67) and 16 (Jockey-F52) host genes were obtained due to the poor genome quality of *S. lebanonensis* and *D. albicans*. As pointed out by Wallau et al. [18], the use of few genes can lead to false positives and false negatives. In summary, in view of the premises established to propose a case of HTT, phylogenetic incongruence validated by significant discrepancies in the dS evolutionary rate between the Jockey elements and a set of vertically transferred genes of the host species, 15 phylogenetic incongruities were validated as potential cases of HTT. These transfers are depicted in Fig. 5.

**Discussion**

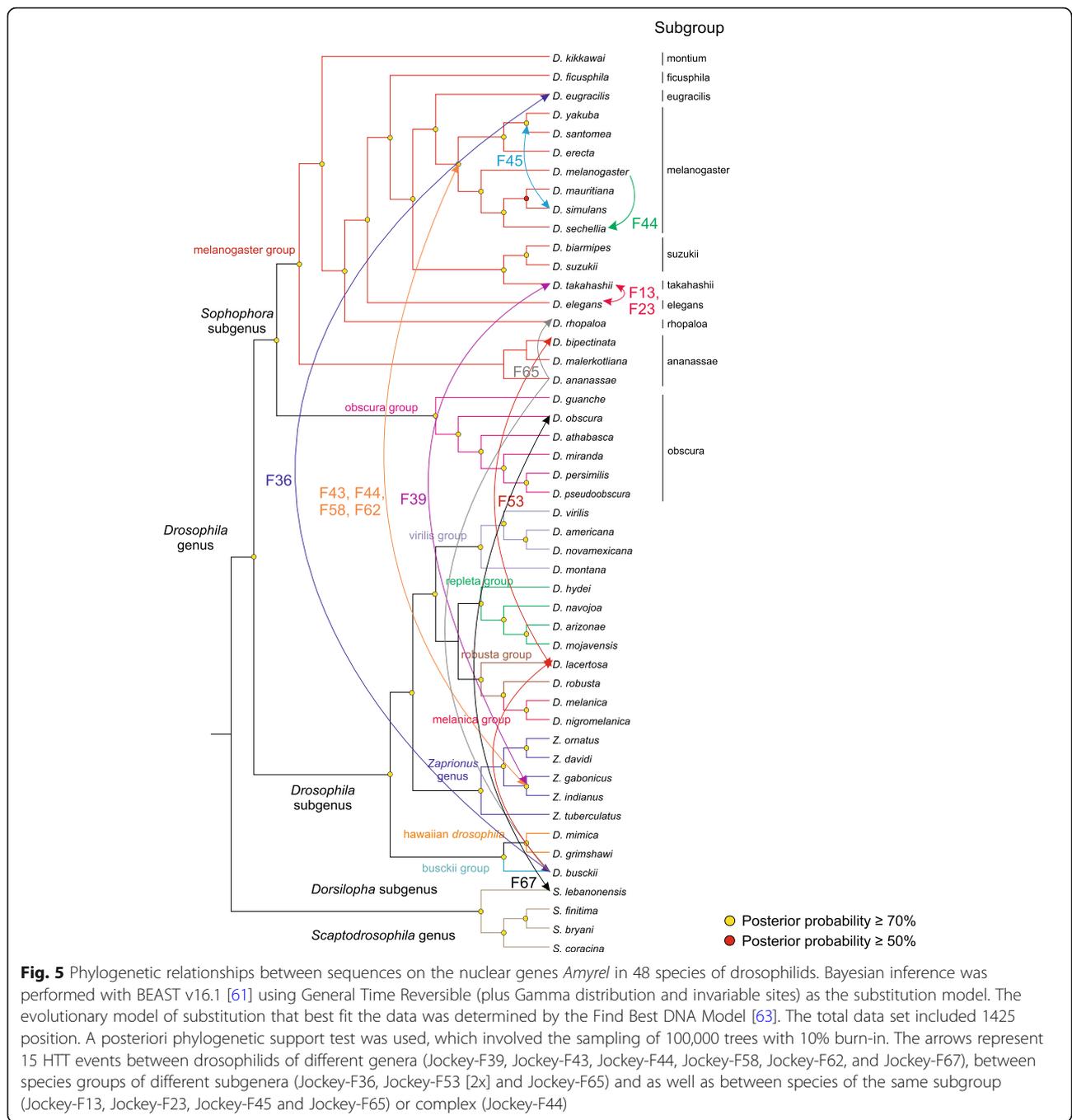
The 41 genomes analysed here provided a broad sampling of Drosophilidae diversity, including three subgenera of the genus *Drosophila* (*Sophophora*, *Drosophila* and *Dorsilopha*) and two other genera, namely, *Zaprionus* and *Scaptodrosophila*. The genus *Scaptodrosophila* occupies a basal position within the Drosophilinae [22], and the placement of the genus *Zaprionus* within the subfamily stands for now, but there is a current consensus that it is imbedded within the genus *Drosophila*. The locations of origin of the species sampled are in all biogeographical

regions where drosophilids occur, and several of the species are currently distributed worldwide.

To infer the occurrence of HTT, three main criteria should be considered: phylogenetic incongruence between the TE tree and the host tree; an irregular distribution of the element in a group of species; and high similarity between sequences of TEs from distantly related species [2, 23]. TEs are sets of copies dispersed throughout the genome of the host species, occurring in tens, hundreds or even thousands of families, which may have different levels of similarity between them and may or may not reflect common ancestry. Therefore, inferences of HTT are reliable if the TE phylogenetic relationships and classification are characterized with robustness; in such cases, we can be confident that the elements compared between two species belong to the same evolutionary unit (family).

**Identification and characterization of jockey families in Drosophilidae**

The LINE elements of the Jockey superfamily are characterized by being exclusively present in arthropods, mainly insects [24]. The term family adopted here differs from that used by Repbase in that we consider a family of TEs to represent an evolutionary lineage that is often shared by several species. Accordingly, we propose that to infer events of HTT, sequences belonging to a family should be characterized following similarity ( $\geq 80\%$ ) and



phylogenetic support (posterior probability  $\geq 70\%$ ) criteria. Thus, the identification of sequences of a TE family in a species reflects the evolutionary relationships of each family, which may (VT) or may not (HTT) be congruent with the phylogenetic relationships of the host species. Following these criteria, we were able to characterize 72 Jockey families in 41 Drosophilidae species. It is important to emphasize that the robustness of HTT hypotheses in our study was initially based on the classical criteria for these kinds of inferences – a high

sequence similarity and phylogenetic incongruities - associated with statistical support provided by the comparison of rates of evolution and codon usage among sequences [18]. However, HTT inferences become more reliable if they are based not only on the phylogenetic aspects but also on the biogeographic-evolutionary history of the species involved. The 15 candidate cases of HTT (20.8% of total families) suggested in this study are discussed in terms of this latter aspect plus the divergence times of the taxa (Fig. 1).

### The Orient: the point of origin and diversification of the melanogaster group

The melanogaster group, which originated in Asia approximately 50 Mya (million years ago), is one of the main groups of the genus *Drosophila* that inhabits the Old Continent [25]. All melanogaster subgroups diversified in the Oriental biogeographic region, except the melanogaster subgroup, which originated and diversified in Africa from a proto-melanogaster founder lineage that migrated to the African continent ~ 17 to 20 Mya [26–28]. In contrast to the melanogaster subgroup, the phylogenetic relationships and dating of divergence between the eastern subgroups are not fully elucidated [29–31], which causes difficulties in interpreting the phylogenetic incongruities between the Jockey sequences. In these cases, and in general inferences of HTT, the spatio-temporal overlap of species, which facilitated TE exchanges at some point in their evolutionary history, is an important criterion with which to add robustness to the inferences. Moreover, it is important to stress that unsampled species could also have been involved since species may share niches and direct or indirect interspecific ecological relations [32].

In the present study, we propose that due to sympatry, the possibility of sharing niches associated with the high movement of species during late Miocene may have contributed to the establishment of a scenario permissive to the occurrence of HTT involving four Jockey families in species of the Oriental melanogaster group. However, as the phylogenetic relationships between some Oriental melanogaster groups and their times of divergence are not completely established, the HTT inferences in these cases should be viewed with caution, even though they are supported by phylogenetic incongruities, later coalescence times of the sequence shared by two species than the species divergence times, and the significance of evolutionary rates and codon usage. Two of these proposed HTTs involve families that occur in species of the subgroups takahashii and elegans (Jockey-F13 and Jockey-F23). The Jockey most recent common ancestor (MRCA) sequence of the clade *D. takahashii*/*D. elegans* dates to only 1.58 Mya (F13) and 1.36 Mya (F23), times of divergence that are very recent for species belonging to different subgroups.

The Eastern and African groups of drosophilids, hosts of the Jockey families sampled here, may have participated in a large wildlife exchange between Eurasia and Africa, when an intercontinental route became possible for the first time between 17 and 20 Mya [26–28], or even more recently, between 6 and 13 Mya [33]. This high species dispersal may have facilitated genetic exchange among drosophilids, resulting in the high amount of HTT between several subgroups of the melanogaster group reported here for Jockey elements as well as elsewhere for other TEs (reviewed in [1, 2]). However,

it is necessary to emphasize that not all phylogenetic incongruities are necessarily due to HTT because different evolutionary rates, ancestral polymorphism, stochastic loss, and introgressive hybridization also result in phylogenetic incongruities [2]. Moreover, incomplete lineage sorting (ILS) due to a short period of species divergence may also cause phylogenetic incongruities [34–36]. Phylogenomic studies of the 12 *Drosophila* genomes provide evidence of some ILS among *Drosophila* (30–40% of the loci) phylogenies, showing discrepancies with the species phylogeny [37]. The significant difference between dS and ENC of Jockey-F62 *Zaprionus* sequences and of host genes may be one of these cases. Another difficulty, in addition to the uncertainty about relationships between some groups of species, regards dating key nodes in the *Drosophila* phylogeny, which can vary widely, regardless of whether the calibration points used are from amber fossils [38], Hawaiian phylogeography [33, 39, 40] and island formation dates [41] or mutation rates [42]. In this study, we considered only HTT events as phylogenetic incongruities whose MRCA sequence dating was more recent than all the possible dates produced using these different calibration points.

### Tropical Africa: a permissive environment for HTT events

Tropical Africa was the location for the radiation of different groups of organisms, including drosophilids, mainly those belonging to the melanogaster subgroup and the *Zaprionus* subgenus. There is an important chapter in the history of the melanogaster group on the African continent, where the melanogaster subgroup originated from the proto-melanogaster lineage and gave rise to three speciation centres: the first, in the western region (15 to 13 Mya), produced the erecta complex (*D. erecta* and *D. oreana*); the second, also originating in the western region (15 to 8 Mya), resulted in the yakuba complex (*D. yakuba*, *D. teissieri* and *D. santomea*); and the third gave rise to the melanogaster complex (*D. melanogaster* and the *simulans* subcomplex) approximately 3 to 2 Mya in central Africa. Later, an ancestral lineage of the simulans subcomplex (*D. simulans*, *D. sechellia* and *D. mauritiana*) dispersed to the islands of the Indian Ocean and Madagascar, approximately 400 thousand years ago (reviewed in 1). These periods are significantly more recent, based on dating using mutation rates as calibration points [41]. On the other hand, a lineage of the genus *Zaprionus* (Drosophilidae) also dispersed from Asia, reaching the African continent approximately 7 Mya [20] or even in the early Oligocene (~ 29.4 Mya, 29), giving rise to the African subgenus *Zaprionus*. The biogeographic history of this genus reveals possible geographic and temporal overlap with the species of the melanogaster subgroup. Thus, the species of the subgenus *Zaprionus*, have an evolutionary history

distinct from that of the melanogaster group, and their biogeographic history reveals possible geographic and temporal overlap with the species of the melanogaster subgroup in the last 7 My.

The incongruities involving Jockey sequences of four families of two evolutionary lineages (Lin5: Jockey-F43 and Jockey-F44 and Lin7: Jockey-F58 and Jockey-F62), the high similarity, which reflects the recent times of divergence from the MRCA sequences, and the VHICA results strongly support the hypothesis of multiple HTT occurrences involving these families between species of the *Zaprionus* subgenus and of the melanogaster subgroup. The MRCAs of Jockey-F43 sequences that *D. erecta* and *D. mauritiana* shares with *Z. gabonicus* and *Z. africanus* are dated to only 6.5 Mya or less (Fig. 1). Jockey-F44 sequences of *Z. indianus*, *Z. gabonicus* and *Z. africanus* are imbedded within the clade containing sequences of the melanogaster complex, with the MRCA dating to 7.14 Mya or less. On the other hand, the MRCAs of the Jockey-F58 shared by *Z. indianus* and *D. erecta* and of Jockey-F62 shared by *Z. africanus*, *Z. indianus*, *Z. gabonicus* and *D. yakuba* are dated to 10.84 and 3.68 Mya, respectively. As we proposed earlier, the overlap in space and time during the radiation of the *Zaprionus* subgenus [20] and of the *melanogaster* subgroup in central Africa [26, 27] could have facilitated the exchange of Jockey sequences between ancestors of these two groups (Fig. 5).

Our research group has been dedicated to the study of HTT between *Zaprionus* and the species of the subgroup *melanogaster* of *Drosophila* for a decade. The results reported in this study for the Jockey families agree with the hypothesis of recent transfer of retrotransposons with LTR (Gypsy, Micropia and Copia) [42, 43] and the non-LTR retrotransposons *Helena* and *BS* [9] between genomes of species of the subgenus *Zaprionus* and the subgroup *melanogaster*. Thus, the evidence we add here provides a consolidated picture of the extensive exchange of TEs between the species of these two groups in central Africa during the late Miocene. In addition, there is evidence of HTT between *Z. africanus* and *D. takahashii* (Jockey-F39, Lin4). Since the MRCA of these sequences is dated to only 3.08 Mya, the hypothesis of HTT due to spatio-temporal overlap between the two groups of species cannot be applied if we assume the period of *Zaprionus* subgenus radiation in Africa (~7 Mya) given [20]. A possible explanation for this time incongruence is that a third species was involved in this transfer.

#### Widespread species distribution and HTT

The rate of HTT depends on both the biology of TEs and various aspects of the species involved, such as their ecology, relatedness, and biology. From the TE point of view, studies indicate that DNA transposons move more

often than LTR retrotransposons between species and more often than non-LTR retrotransposons due to the mode of transposition, which explains the low rate of HTT reported for Jockey (reviewed in [1–3, 18, 44]). Studies have also shown that HTT occurs more frequently between closely related species than between more distantly related ones [45–47]. Regarding the ecology and biology of the species involved in HTT, it has been shown that species with overlapping habitats share more TEs by HTT than species that are distributed in different environments [9, 42], as well as share more viruses that can be vectors of TEs [48]. Although susceptibility to viruses shows a strong phylogenetic correlation, susceptible hosts can occasionally be grouped into phylogenetically distant groups, allowing parasites to jump great phylogenetic distances [49].

Our data possibly show a further trend that can be added to the aforementioned ones - colonizing species that disperse across different geographic regions have increased HTT rates. It is noteworthy that of the 12 Jockey families in which HTT events were validated, in five (Jockey-F39, Jockey-F43, Jockey-F44, Jockey-F58, and Jockey-F62), the transfers occurred between two distantly related groups of species (the melanogaster group and *Zaprionus*) that migrated from the Eastern region to Africa. In addition, in three other families (Jockey-F36, Jockey-F53, and Jockey-F65), the transfer involved *D. busckii*, which is a cosmopolitan species [50] as well as *D. bipunctata*, which has wide distribution and spread across the Oriental-Australian biogeographic zone [51]. Additionally, in Jockey-F67, the HTT occurred between *D. obscura* and *S. lebanonensis*, which are widespread in the Palaearctic region. Moreover, although not validated by the VHICA and thus not considered an HTT event here, two extra incongruities involved widespread species, namely, *D. suzukii* [52] and *D. albomicans* [53], respectively in Jockey-F6 and Jockey-F52. The reason for the relationship between a widespread distribution and HTT needs to be investigated; however, what seems clear is that the ability of these species to disperse increases the chance of phylogenetic jumps of viruses, potential candidates to move TEs from one genome into another [54].

#### Conclusion and perspectives

Recently, there has been an increase in the number of reported cases of HTT of non-LTR elements in *Drosophila*. Although only five cases were previously reported in drosophilids, involving only four LINE elements - Jockey, Doc, F and I [2], 20 cases were reported in Drosophilidae and 303 cases were reported in Insecta in 2017 [8] due to the use of large-scale genomic analyses [47, 55]. Despite this broad sampling, the identification of HTT events for specific families in the LINE

order is lacking, and little is known about the species involved in these events. In the present study, we were able to characterize 15 new cases of HTT for Jockey families, which represents 20.8% of the families sampled in 41 species of *Drosophila*. This value is twice that obtained in Arthropoda, and this difference is due to poor sampling of drosophilids, with only nine species [10]. Equally importantly, we identified the species involved in these exchanges and formulated hypotheses about the spatio-temporal relationships between species that made these exchanges possible. While the HTT events between species of different subgenera were strongly corroborated, the inferences related to the eastern species of the *melanogaster* group, despite meeting all the criteria for HTT used in the literature, need greater support, which could be obtained by increasing the sampling of these elements in *Drosophila* species as well as producing more robust phylogenies of these species. Moreover, more complex phylogenetic approaches/models integrating ILS, duplication/loss and introgression should be developed.

## Material and methods

### Extraction of jockey sequences from genome assemblies

Genome assemblies for 48 species were downloaded from public databases (Additional file 1: Table S1): The European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena>) for 31 of the 48, Ensembl Metazoa for the 12 first sequenced genomes [56], <http://popoolation.at/> for *D. lowei* and *D. mauritiana* and [https://www.diark.org/diark/species\\_list/Drosophila](https://www.diark.org/diark/species_list/Drosophila) for *D. malerkotiana palensis*. We also used draft genomes of the three *Zaprionus* species (*Z. africanus*, *Z. indianus*, and *Z. gabonicus*), for which we provide RepeatMasker outputs, but the genome assemblies are not yet available (Haudry, [pers.com](mailto:pers.com)). On each assembly, RepeatMasker was run using the complete Repbase library to detect all elements in each genome and their closest match, using the *Drosophila* model. We then used the program OneCodeToFindThemAll [57] to parse RepeatMasker outputs, identify copies with respect to the 80–80–80 rule (using the “strict” option) and export sequences in FASTA format. From each set containing all 80–80–80 copies of the genome, a manually curated Repbase library, 111 Jockey families of *Drosophila* containing sequences of transcriptase reverses [Additional file 4: Data S1] was blasted to select Jockey copies with RT, with the option of *eval* <1e-5 and a percentage of identity > 80%. Finally, a custom script was run to select the sequences corresponding to the three best hits (based on the bit-score criteria) per Jockey family in FASTA format.

All Jockey sequences extracted from the genomes were submitted to the NCBI’s Conserved Domains Database [58] for identification of the RT domain region, and

sequences that presented a minimum extension of 300 nt for this domain were selected for the analyses. A list of the species in whose genomes Jockey sequences that met the search criteria were found is given [Additional file 1: Table S3].

### Phylogenetic analyses

Phylogenetic analyses were performed using the RT gene sequences, one of the most conserved regions of TEs that is, as a rule, used for phylogenetic analyses. The alignment was performed with MAFFT [59] and trimmed visually. Poorly aligned regions were removed using trimALL version 1.3 [60]. Then, sequences that aligned poorly or  $\leq$  200 bp of the filtered alignment were also removed. For the phylogenetic reconstructions, Bayesian inference (BI) was performed using BEAST v16.1 [61] with an a posteriori phylogenetic support test, using sampling of 100,000 trees and a burn-in of 10%. The times of divergence of the Jockey sequences belonging to the same family from the most recent common ancestor (MRCA) sequence were estimated using the Bayesian approach and BEAST v16.1, with a neutral nucleotide substitution rate of  $r = 0.016/\text{site}/\text{My}$  [62] for the calibration of the phylogenetic tree. Moreover, pairwise (p) distances between the Jockey sequences were estimated to classify sets of sequences belonging to the same family following the 80–80–80 criterion of Wicker et al. [15] using MEGA7 [63]. An evolutionary lineage was defined as a group that branched off early in the phylogeny and had an a posteriori probability  $\geq 0.7$ . Within each lineage, the later clades, which include sequences with a divergence (p-distance) < 20% (similarity of  $\geq 80\%$ ) and support of  $\geq 0.7$  or single sequences, constitute the families. Two clades of the same lineage constitute different families when the mean divergence between them is > 20%, which corresponds to a similarity < 80%.

The phylogeny of the drosophilid species involved in this study was reconstructed based on 48 sequences of the nuclear gene *Amyrel*, which was obtained from GenBank for most of the species (Additional file 1: Table S4). For *Z. gabonicus* and *Z. africanus*, the sequences were retrieved from the genomes using the *Z. indianus* sequence (EF458322.1) as a query. Bayesian inference was performed with BEAST v16.1 [61] using General Time Reversible (plus Gamma distribution and invariable sites) as the substitution model and an a posteriori phylogenetic support test, using sampling of 100,000 trees and a burn-in of 10%. The evolutionary model of substitution that best fit the data was determined by the Find Best DNA Model in MEGA7 [63].

### Inference of HTT

The inference of HTT was initially carried out by the classical criteria: sequence similarity and phylogenetic incongruence. When the necessary data were available,

these inferences were validated by the application of VHICA (vertical and horizontal inheritance consistence analysis) methodology, which provides statistical support for corroboration of the inferences of HTT [18]. The method is based on discrepancies between the rate of evolution at synonymous sites (dS) and the ENC between pairs of TE sequences and vertically transferred orthologous genes. ENC is used to estimate selection due to codon usage bias (CUB) at synonymous sites. The use of dS and CUB together provides robustness in the HTT inferences because being dS and CUB correlated, a low dS does not necessarily indicates HTT if associated with a high CUB, but low dS associated with low CUB is inconsistent with VT [18]. Statistical support for the HTT inferences is given by a linear regression between the distribution of ENC and dS values (with Bonferroni correction,  $P < 0.01$ ). For VHICA application, up to 29 single copy orthologous genes of each analysed species were used (Additional file 1: Table S5), with the exception for species involved in HTT with *D. lacertosa*, *D. albomicans* and *S. lebanonensis*, in which it was possible to extract only 19, 16 and 12 genes with good sequence quality, respectively, due to poor genome quality. The sequences of these genes were aligned using MAFFT [59] and later concatenated to reconstruct a phylogenetic tree of the host species. Additionally, the tree was inferred by maximum likelihood (ML), as described in Simão et al. [9].

### Gene and TE sequences used in the VHICA analysis

For the candidates of HTT among species (Fig. 1), we performed de novo gene prediction using Augustus [64], with “fly” as the training species, except for *D. erecta*, for which annotated CDSs were retrieved from the file “*Drosophila erecta.dere\_caf1.cds.all.fa*” available from the NCBI (<https://www.ncbi.nlm.nih.gov/>). A custom nucleotide database was created for each species that contained all the predicted/annotated (for *D. erecta*) coding sequences using makeblastdb and then used for a blastn search against the sequences of 50 single copy *D. melanogaster* genes from a set of 50 *Drosophila* single copy orthologous genes [18], using the following command: `blastn -db $species.codingseq -query dmel_vhica_genes.fas -outfmt 6 | sort -k1,1 -k12,12nr -k11,11n | sort -u -k1,1 --merge >best_single_hits.blastn`. Probably due to variations in sequencing quality and genomic assemblies, as well to sequence divergence, the maximum number of genes found in all genomes, with sufficient quality and length to be used in the analysis was 29. For the TE dS estimation we selected the best copy of each TE family extracted from the genomes, regarding its length and sequence integrity.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13100-019-0184-1>.

**Additional file 1: Table S1.** Genome accessions for 48 species analysed. **Table S2.** Species of the genus *Drosophila* and *Zaprionus* analysed in this study, its hierarchical divisions and geographical origin. **Table S3.** Number of Jockey families and classification after phylogenetic analyses in 41 species of Drosophilidae. **Table S4.** Orthologous single copy genes vertically-transmitted genes used for each species used for the VHICA analyses. **Table S5.** Accession of Amyrel sequences used for reconstructing the species phylogeny.

**Additional file 2: Table S1.** Pairwise distances (p-distances) between 72 Jockey families. **Table S2.** Pairwise distances (p-distances) and similarity (1-p\*100) within 72 Jockey families.

**Additional file 3: Figure S1.** Identification of Jockey elements horizontal transfer. Consistency graphical representation and dS-ENC graphs obtained from the comparison between species in Jockey-F45 representing inferences of HTT between *D. yakuba* and *D. simulans*. For details of the graphs refer to the legend of Fig. 2. **Figure S2.** Identification of Jockey elements horizontal transfer. Consistency graphical representation and dS-ENC graphs obtained from the comparison between species in Jockey-F6 (a), Jockey-F13 (b) and Jockey-F23 (c) representing inferences of HTT and VT. For details of the graphs refer to the legend of Fig. 2. There are signals of HTT between *D. takahashii* and *D. elegans* in both Jockey-F13 and Jockey-F23. **Figure S3.** Identification of Jockey elements horizontal transfer. Consistency graphical representation and dS-ENC graphs obtained from the comparison between species in Jockey-F39 (a), Jockey-F62 (b) and Jockey-F67 (c) representing inferences of HTT and VT. For details of the graphs refer to the legend of Fig. 2. There are signals of HTT between *D. takahashii* and *Z. africanus* (Jockey-F39) and *S. lebanonensis* and *D. obscura* (Jockey-F62).

**Additional file 4: Data S1.** Sequences of the transcriptase reverse gene from 111 families of Jockey clade downloaded from Repbase database [16] that occurs in species of the genus *Drosophila*.

### Abbreviations

CUB: Codon Usage Bias; dS: Synonymous sites; ENC: Preferential use of codons; HT: Horizontal transfer; HTT: Horizontal transfer of transposable elements; LINE: Long interspersed nuclear element; MRCA: Most Recent Common Ancestral; Mya: millions of years ago; TEs: Transposable elements; VHICA: Vertical and Horizontal Inheritance Consistence Analysis; VT: Vertical transfer

### Acknowledgements

Genomic analyses by AH were performed using the computing facilities of the CC LBBE/PRABI.

### Authors' contributions

ILT performed all the analyses, interpreted the data, drafted figures and wrote the manuscript. AH performed extraction of Jockey and the host gene sequences from genome assemblies and revised the manuscript. MCS performed the phylogenetic analyses, interpreted the data, and drafted figures. CMAC designed the study, interpreted the data, and wrote the manuscript. All the authors read and approved the final manuscript.

### Funding

This research was funded by the Brazilian agency FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo to CMAC (grant number 2016/19271-2). APEGE CNRS DynET and BQR Université Lyon1 EC grants awarded to AH. CNPq-Conselho Nacional de Desenvolvimento Científico e Tecnológico awarded fellowship to CMAC (grant number 303455/2017-9) and to ILT (grant number: 158389/2017-4).

### Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. Jockey sequences dataset is available from the corresponding author on reasonable request.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing of interests.

**Author details**

<sup>1</sup>Department of Biology, Institute of Biosciences, Humanities and Exact Sciences (IBILCE), UNESP – São Paulo State University, Campus São José do Rio Preto, São Paulo, SP 15054-000, Brazil. <sup>2</sup>Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, Université Lyon 1, CNRS, UMR 5558, F-69622 Villeurbanne, France.

Received: 22 March 2019 Accepted: 9 October 2019

Published online: 04 November 2019

**References**

- Carareto CMA. Tropical Africa as a cradle for horizontal transfers of transposable elements between species of the genera *Drosophila* and *Zaprionus*. *Mob Genet Elements*. 2011;1:179–86 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22312591>. Cited 2017 Apr 12.
- Loreto ELS, Carareto CMA, Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* (Edinb). 2008;100:545–54 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18431403>. Cited 2017 Apr 12.
- Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol*. 2010;25:537–46 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20591532>. Cited 2017 Apr 12.
- Hartl DL, Lohe AR, Lozovskaya ER. Modern thoughts on an ancient mariner: function, evolution, regulation. *Annu Rev Genet*. 1997; 31:337–358. Available from: <https://doi.org/10.1146/annurev.genet.31.1.337>. Cited 2017 Apr 3
- Daniels SB, Strausbaugh LD, Ehrman L, Armstrong R. Sequences homologous to P elements occur in *Drosophila paulistorum*. *Proc Natl Acad Sci U S A*. 1984;81:6794–7 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6093120>. Cited 2017 Jul 10.
- Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*. 1990;124:339–55 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2155157>. Genetics Society of America. Cited 2018 Dec 21.
- Dotto BR, Carvalho EL, Silva AF, Duarte Silva LF, Pinto PM, Ortiz MF, et al. HTT-DB: Horizontally transferred transposable elements database: Fig. 1. *Bioinformatics*. 2015;31:2915–7 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25940562>. Cited 2018 Dec 21.
- HTTDB - Horizontally transferred transposable elements database. Available from: <http://pa.saogabriel.unipampa.edu.br:8080/httdatabase/resultado/resultado.jsp?organism>
- Simão MC, Haudry A, Granzotto A, Setta N, CMA C. Helena and BS: two travellers between the genera *Drosophila* and *Zaprionus*. *Genome Biol Evol*. 2018;10:2671–85 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30165545>. Cited 2019 Sep 18.
- Reiss D, Mialdea G, Miele V, de Vienne DM, Peccoud J, Gilbert C, et al. Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. *Buerkle A, editor. PLOS Genet*. 2019;15:e1007965. Cited 2019 Feb 20. Public Library of Science. Available from: <https://doi.org/10.1371/journal.pgen.1007965>
- Priimägi AF, Mizrokhi LJ, Ilyin YV. The *Drosophila* mobile element jockey belongs to LINES and contains coding sequences homologous to some retroviral proteins. *Gene*. 1988;70:253–62 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2463954>. Cited 2019 Sep 18.
- Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, et al. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol*. 2002;3:research0084.1 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12537573>. Cited 2019 Sep 18.
- Metcalfe CJ, Casane D. Modular organization and reticulate evolution of the ORF1 of Jockey superfamily transposable elements. *Mob DNA*. 2014;5:19 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25093042>. BioMed Central. cited 2019 Sep 18.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*. 1993;72:595–605 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7679954>. Cited 2019 Sep 18.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8:973–82 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17984973>. Cited 2017 Apr 3.
- Repbase - GIRI. [cited 2019 Mar 20]. Available from: <https://www.girinst.org/repbase/>. Accessed 20 Mar 2019.
- Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26045719>. Cited 2019 Sep 18.
- Wallau GL, Capy P, Loreto E, Le Rouzic A, Hua-Van A. VHICA, a New Method to Discriminate between Vertical and Horizontal Transposon Transfer: Application to the Mariner Family within *Drosophila*. *Mol Biol Evol*. 2016;33:1094–109 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26685176>. Cited 2017 Apr 12. Oxford University Press.
- Lohe AR, Moriyama EN, Lidholm DA, Hartl DL. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol Biol Evol*. 1995;12:62–72 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7877497>. Cited 2019 Jul 23.
- Yassin A, Araripe LO, Capy P, Da Lage J-L, Klaczko LB, Maisonhaute C, et al. Grafting the molecular phylogenetic tree with morphological branches to reconstruct the evolutionary history of the genus *Zaprionus* (Diptera: Drosophilidae). *Mol Phylogenet Evol*. 2008;47:903–15 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S105579030800047X>. Cited 2019 Sep 18.
- Markow TA, O'Grady PM. A guide to species identification and use. *Entomol Ber. Elsevier*; 2006;66:157–58.
- Grimaldi D. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). *Bull. Am. Museum Nat. Hist*. 1990;197:139.
- Wallau GL, Vieira C, ÉLS L. Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. *Mob DNA*. 2018;9:6. BioMed Central. Cited 2018 Dec 21. Available from: <https://doi.org/10.1186/s13100-018-0112-9>.
- Berezikov E, Bucheton A, Busseau I. A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of *Drosophila melanogaster*. *Genome Biol*. 2000;1:RESEARCH0012 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11178266>. BioMed Central. Cited 2017 Jul 4.
- Kopp A, True JR. Phylogeny of the Oriental *Drosophila melanogaster* Species Group: A Multilocus Reconstruction. *Whitfield J, editor. Syst Biol*. 2002;51:786–805. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12396591>. Cited 2019 Sep 18.
- Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. Historical Biogeography of the *Drosophila melanogaster* Species Subgroup. *Evol Biol*. 1988;159–225. Available from: [https://doi.org/10.1007/978-1-4613-0931-4\\_4](https://doi.org/10.1007/978-1-4613-0931-4_4)
- Lachaise D, Silvain J-F. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*-*D. simulans* palaeogeographic riddle. *Genetica*. 2004;120:17–39 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15088644>. Cited 2017 May 3.
- Jeffs PS, Holmes EC, Ashburner M. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. *Mol Biol Evol*. 1994;11:287–304 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/8170369>. Cited 2019 Sep 18.
- Ashburner M, Bodmer M, Lemeunier F. On the Evolutionary relationships of *Drosophila-melanogaster*. *Dev Genet* 1984;4:295–312. Available from: <https://doi.org/10.1002/dvg.1020040407>
- Schwaroch V. Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). *Biol J Linn Soc*. 2002; 76:21–37 Available from: <https://academic.oup.com/biolinnean/article/76/1/21/2639664>. Oxford University Press.
- O'Grady PM, DeSalle R. Phylogeny of the Genus *Drosophila*. *Genetics*. 2018; 209:1–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29716983>. Cited 2019 Sep 18.
- Wang X, Liu X. Close ecological relationship among species facilitated horizontal transfer of retrotransposons. *BMC Evol Biol*. 2016;16:201. BioMed Central. Available from: <https://doi.org/10.1186/s12862-016-0767-0>.

33. Russo CAM, Takezaki N, Nei M. Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol.* 1995; Available from: <https://academic.oup.com/mbe/article/12/3/391/979928>.
34. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol.* 2009;24:332–40 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19307040>. Cited 2019 Sep 18.
35. Degnan JH, Salter LA. Gene tree distributions under the coalescent process. *Evolution.* 2005;59:24–37.
36. Hobolth A, Duthéil JY, Hawks J, Schierup MH, Mailund T. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* 2011;21:349–56 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21270173>. Cited 2019 Sep 18.
37. Pollard DA, Iyer VN, Moses AM, Eisen MB. Widespread Discordance of Gene Trees with Species Tree in *Drosophila*: Evidence for Incomplete Lineage Sorting. *PLoS Genet.* 2006;2:e173 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17132051>. Cited 2019 mar 19.
38. Grimaldi DA. Amber fossil *Drosophilidae* (Diptera), with particular reference to the Hispaniolan taxa. *Am Museum Novit.* 1987;1–23 Available from: <http://digitallibrary.amnh.org/handle/2246/5210>.
39. Tamura K, Subramanian S, Kumar S. Temporal Patterns of Fruit Fly (*Drosophila*) Evolution Revealed by Mutation Clocks. *Mol Biol Evol.* 2003;21:36–44 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12949132>. Cited 2019 Sep 18.
40. Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. Estimating Divergence Dates and Substitution Rates in the *Drosophila* Phylogeny. *Mol Biol Evol.* 2012;29:3459–73 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22683811>. Cited 2019 Sep 18.
41. Russo CAM, Mello B, Frazão A, Voloch CM. Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). *Zool J Linn Soc.* 2013; 169:765–775. Available from: <https://doi.org/10.1111/zoj.12062>. John Wiley & Sons, Ltd (10.1111). Cited 2019 Sep 18.
42. de Setta N, Van Sluys M-A, Capy P, Carareto CM. Multiple invasions of Gypsy and Microtia retroelements in genus *Zaprionus* and melanogaster subgroup of the genus *Drosophila*. *BMC Evol Biol.* 2009;9:279. *BioMed Central.* Cited 2019 Sep 18. Available from: <https://doi.org/10.1186/1471-2148-9-279>.
43. de Setta N, Van Sluys M-A, Capy P, Carareto CMA. Copia Retrotransposon in the *Zaprionus* Genus: Another Case of Transposable Element Sharing with the *Drosophila melanogaster* Subgroup. *J Mol Evol.* 2011;72:326–38 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21347850>. Cited 2019 Jul 25.
44. Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. *Curr Issues Mol Biol.* 2004;6:57–71 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14632259>. Cited 2017 Apr 3.
45. Bartolomé C, Bello X, Maside X. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol.* 2009; 10:R22 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19226459>. *BioMed Central.* Cited 2017 Jul 4.
46. El Baidouri M, Carpentier M-C, Cooke R, Gao D, Lasserre E, Llauro C, et al. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res.* 2014;24:831–8 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24518071>. Cited 2017 Apr 17.
47. Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci U S A.* 2017;114:4721–6 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28416702>. National Academy of Sciences. Cited 2017 Jul 4.
48. Gilbert C, Feschotte C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev.* 2018;49:15–24 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29505963>. Cited 2019 Sep 18.
49. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM. Host Phylogeny Determines Viral Persistence and Replication in Novel Hosts. *PLoS Pathog.* Schneider DS, editor. 2011;7:e1002260. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21966271>. Cited 2019 Sep 18.
50. Nunney L. The colonization of oranges by the cosmopolitan *Drosophila*. *Oecologia.* 1996;108:552–61 Available from: <https://www.jstor.org/stable/4221451>.
51. Banerjee P, Singh BN. The *Drosophila bipectinata* species complex: phylogenetic relationship among different members based on chromosomal variations. *J Genet.* 2017;96:97–107. Available from: <https://doi.org/10.1007/s12041-017-0746-2>
52. Cini A, Ioriatti C, Anfora G. A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. *Bull Insectology.* 2012;65:149–60 Available from: <https://www.researchgate.net/publication/229059768>.
53. Isobe K, Takahashi A, Tamura K. Cold tolerance and metabolic rate increased by cold acclimation in *Drosophila albomicans* from natural populations. *Genes Genet Syst.* 2013;88:289–300 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24694392>. Cited 2019 Sep 18.
54. Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature.* 2010;464:1347–50. Cited 2017 Apr 3 Available from: <https://doi.org/10.1038/nature08939>.
55. Peccoud J, Cordaux R, Gilbert C. Analyzing Horizontal Transfer of Transposable Elements on a Large Scale: Challenges and Prospects. *BioEssays.* 2017;1700177:1700177. Available from: <https://doi.org/10.1002/bies.201700177>.
56. Consortium D 12 G. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007;450:203–18 Available from: <http://www.nature.com/articles/nature06341>. Nature Publishing Group. Cited 2019 Jul 16.
57. Bailly-Bechet M, Haudry A, Lerat E. "One code to find them all": A perl tool to conveniently parse RepeatMasker output files. *Mob DNA.* 2014; 5. Available from: <https://doi.org/10.1186/1759-8753-5-13>
58. NCBI Conserved Domain Search. Available from: <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>. Accessed 20 Mar 2019.
59. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12136088>. Cited 2017 Apr 12.
60. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19505945>. Cited 2019 Sep 18.
61. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012;29:1969–73 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22367748>. Cited 2017 Apr 12.
62. Sharp PM, Li WH. On the rate of DNA sequence evolution in *Drosophila*. *J Mol Evol.* 1989;28:398–402 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2501501>. Cited 2017 Apr 12.
63. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33:1870–4 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27004904>.
64. Hoff KJ, Stanke M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinforma.* 2018; 65:e57. Available from: <https://doi.org/10.1002/cpbi.57>. John Wiley & Sons, Ltd. cited 2019 Jul 16.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

