Recent horizontal transfer of transposable elements in *Drosophila*

Shashank Pritam[1,] and Sarah Signor[1,†]

[1]Biological Sciences, North Dakota State University, USA
[†]Corresponding author email: sarah.signor@ndsu.edu

Abstract

Transposable elements are genetic elements also known as "jumping genes" that increase their copy number within a host through various mechanisms of transposition. TEs can also move between species through unknown intermediaries in a process known as horizontal transfer, infecting novel genomes and increasing in copy number. While many individual invasions have been documented, a large dataset of recent horizontal transfer events that will allow us to understand larger more general HT patterns has not been assembled. In this manuscript we used almost 400 dipteran genomes to uncover 637 recent transposon invasions, mostly in *Drosophila*. The majority of transfers occurred between closely related species, with the cosmopolitan *melanogaster* group showing the highest recent transfer activity. We even documented a single transposon with 16 recent transfers, many between different *Drosophila* groups. Using species distance on a phylogenetic tree to measure the distance that transposons travel, we found that DNA transposons transfer between distantly related species much more frequently compared to retrotransposons. This potentially represents a different evolutionary strategy for exploiting naive genomes. Our phylogenetic framework advances the understanding of horizontal transfer dynamics at the species level within *Drosophila*.

Introduction

Transposable elements (TEs) are mobile DNA parasites which increase in copy number within their hosts' genome. They frequently interrupt functional sequences, cause double stranded breaks, and induce ectopic recombination. Due to this and other negative side effects, host species have specialized systems for repressing the movement of transposons. In *Drosophila*, this relies on specialized small RNA from the PIWI-interacting RNA (piRNA) pathway [Senti and Brennecke, 2010, Brennecke et al., 2007, Gunawardane et al., 2007]. These piRNA are cognate to the TE, and can then be used to target TE transcripts for degradation, or TE sequences for silencing. piRNA are typically generated from TE copies inserted within specialized genomic regions, known as piRNA clusters [Brennecke et al., 2007, Malone et al., 2009]. This system adapts to the invasion of a new TE through an as-yet not fully understood mechanism. In the germline, the TE will either insert into an established piRNA cluster such as *42AB* in *Drosophila*, or will begin producing piRNA *de novo* [Srivastav et al., 2023]. While TE silencing is not absolute, there is evidence that over time selection is lessened on the TE and the TE sequence will begin to degrade. One way in which TE can escape this fate is by transferring to naive genomes of other species, or horizontal transfer (HT) [Kofler et al., 2015a, Scarpa et al., 2025, Pianezza et al., 2023].

HT of TEs has been documented extensively for a limited number of TEs. For example, the transfer of the *P*-element from *D. willistoni* to *D. melanogaster* and from *D. melanogaster* to *D. simulans* was tracked both through hybrid dysgenesis between populations and fortuitous short read sequencing [Kofler et al., 2015a, Daniels et al., 1990]. *D. willistoni* is native to South and Central America, thus this transfer was only able to occur when *D. melanogaster* extended its range into the Americas about 200 years ago [Engels, 1983, Keller, 2007]. Following documentation of the *P*-element invasions, it was later shown that the *I*-element,

hobo, and *Tirant* had also invaded *D. melanogaster* in the last century (1930-1960) [Bucheton et al., 1992, Bonnivard et al., 2000, Schwarz et al., 2021]. More recently, the invasion of *D. melanogaster* by three more transposable elements in the last 200 years was documented through DNA sequencing of museum specimens [Scarpa et al., 2023]. These three TEs were posited to have been transferred between *D. melanogaster* and *D. simulans* [Sánchez-Gracia et al., 2005, Bartolomé et al., 2009, Modolo et al., 2014]. Lastly, four TEs invaded much more recently in *D. melanogaster* (1980s-2010s), bringing the total to 11 in the last 200 years [Pianezza et al., 2025, 2023]. This was surprising in part due to the time scale of the invasions, 11 invasions in 200 years are much more than would be expected based on the known number of TE families in *Drosophila* [Pianezza et al., 2025]. However, these TEs also spread much faster than expected, some in as little as 2 years [Pianezza et al., 2025].

While the phenomenon of horizontal transfer has been clearly documented, its mechanism is not. One of the more widespread hypotheses is that viruses facilitate their transfer [Miller and Miller, 1982, Fraser et al., 1985, Jehle et al., 1995], but parasitoid wasps and parasitic mites have also been implicated in HT [Yoshiyama et al., 2001, Houck et al., 1991]. *Wolbachia*, an intracellular parasite that is widespread in insects, has also been considered a possible vector [Kondo et al., 2002, Heath et al., 1999]. None of these vectors has been demonstrated to mediate the transfer of a transposon in laboratory or natural settings. One class of TEs (*gypsy/gypsy* contains an *envelope* protein which allows the TE to act like an endogenous retrovirus and gives them infectious properties (*ZAM, springer, Tirant, Quasimodo*) [Song et al., 1994, Senti et al., 2025]. However, in *Drosophila*, it is only one clade of TEs that possess *envelope* proteins, and HT is observed for all classes of TEs regardless of the presence of this protein. Some TEs do appear to engage in HT more frequently than others - for example *Mariner* elements have been documented as crossing large phylogenetic distances in *Drosophila* [Wallau et al., 2014, Maruyama and Hartl, 1991a, Robertson, 1993]. However, depending upon the study, it is not always clear if it is the same TE or a member of the same class of TEs.

In order for a TE to spread between species it must either come into contact with a host species or with a vector from the original host species. This generally requires ecological proximity [Venner et al., 2017]. TEs are also much more readily transferred between more closely related species [Feschotte and Pritham, 2007, Peccoud et al., 2017]. Thus, for example, many of the recently acquired TEs in *D. melanogaster* are transferred from *D. simulans*, a close relative that is also a cosmopolitan human commensal which occupies many of the same ecological niches as *D. melanogaster* [Kofler et al., 2015a, Scarpa et al., 2023, Pianezza et al., 2023, 2025, Scarpa et al., 2025]. As humans move organisms and goods around the globe, the rate at which organisms experience novel congeners will also increase. A not-as-yet appreciated byproduct of this contact between congeners will be the exchange of novel TEs [Scarpa et al., 2025]. The expectation is that the transfer of TEs may have accelerated in the last 200 years. What the implications of that may be for genomic stability are unknown. For example, in *D. melanogaster* the last 11 invasions have increased its genome size by 1 MB [Pianezza et al., 2025].

In this manuscript we aim to investigate HT more broadly in *Drosophila*, limiting ourselves to recent invasions. We were interested to see the rate of horizontal transfer of TEs, the species between which transfer was occurring, and the phylogenetic distance between donor species. We found evidence of more than 600 horizontal transfers within *Drosophila* and other closely related dipterans such as *Zaprionus* and *Lordiphosa*. We recapitulated previous results, such as the invasion of the *P*-element and *Spoink*, as well as many novel transfers. Some of our results were expected, for example, the majority of HT events was found in LTRs, which are known to be the most active TEs in *Drosophila* [Kofler et al., 2015b, Mérel et al., 2020]. Other results were less expected, such as the observation that DNA transposons transfer less but cover larger phylogenetic distances than LTR transposons. This could represent separate evolutionary strategies for the two types of transposon.

Methods

Genetic resources

All of the *Drosophila* genomes used in this manuscript were obtained from NCBI and previously published [Kim et al., 2023, 2021]. The TE library was obtained from Chakraborty (2021) [Chakraborty et al., 2021]. This TE library contains $2,332$ repetitive elements from a variety of species groups, including *grimshawii*, *ananassae*, *willistoni*, and *ficusphila*. We excluded satellite repeats and *R* elements (which are vertically inherited)[Eickbush and Eickbush, 1995] as well as redundant TEs. The TE library is biased towards *D. melanogaster* and its relatives. However, many TEs are assigned to specific groups but are active in different groups, i.e. *Mariner-1 D. elegans* (see below). Therefore, the group label for the TE is not a good guide to distribution. Nonetheless, the results will still be biased towards the groups were TEs were most heavily sampled.

Identifying horizontal transfer

TE insertions were initially located using the entire library of TEs and genomes with blastn [Camacho et al., 2009]. These files were filtered for alignments of at least 80 percent similarity to the reference and 1000 bp in length. To obtain recently active TEs, they were then filtered for hits that were at least 80 percent of the reference length for the TE internal sequence. This removed short elements such as DNAREP from the dataset. To obtain a consensus sequence of each TE from each species, the sequence of each blast hit was extracted and aligned with mafft [Katoh et al., 2002, Katoh and Toh, 2008, Katoh et al., 2009]. For each species a consensus sequence was created with EMBOSS consensus [Rice et al., 2000]. Following this, the sequences from each species were aligned with mafft [Katoh et al., 2002, Katoh and Toh, 2008, Katoh et al., 2009]. To verify the results of the consensus sequence approach, we also created a dataset of aligned individual insertions using the same basic pipeline but without the creation of a consensus sequence (Supplemental File 1).

Alignments were hand-checked for errors. TEs present in three species or less were excluded from the dataset. Alignments from at least four species were used as input to MrBayes [Ronquist et al., 2012, Huelsenbeck and Ronquist, 2001]. The phylogenetic trees were built using a GTR substitution model and gamma distributed rate variation across sites. The Markov chain Monte Carlo chains were run until the standard deviation of split frequencies was below 0.01. The consensus trees were generated using sumt conformat = simple. While constructing branch lengths, MrBayes uses evolutionary models for nucleotide and branch lengths that are supposed to represent the expected number of substitutions per site [Nascimento et al., 2017]. A tree for each TE was constructed in R using the package ape [Paradis et al., 2004, Paradis and Schliep, 2019].

Data analysis was performed using Python libraries including pandas, NumPy, SciPy, and DuckDB for data processing [The pandas development team, 2025, Harris et al., 2020, Virtanen et al., 2020, Raasveldt and Mühleisen, 2019]. Visualizations were generated using matplotlib, seaborn, ggplot2 and ggtree [Hunter, 2007, Waskom, 2021, Wickham, 2016, Yu et al., 2017]. Phylogenetic analyses utilized DendroPy [Moreno et al., 2024]. R data manipulation was handled with dplyr and readr, while bioinformatics operations used Biostrings [Wickham et al., 2025, 2024, Pagès et al., 2025]. Additional R utilities included ctv, and machine learning analyses employed scikit-learn [Zeileis et al., 2023, Pedregosa et al., 2011].

TE and species similarity

CLUSTAL was used to generate distance matrices from each of the aligned TE sequences [Sievers and Higgins, 2018]. We selected TE alignments that were $> 99\%$ similar in different species as recent invaders. This is an arbitrary threshold, however, it does guarantee that any TE in the dataset was very recently active. To be sure this represents an HT event and not interbreeding or recent speciation, we excluded species that were not at least 98% diverged. We used sequences from *Adh*, *Gpdh*, and *Ddc* to gather the percent divergence for species pairs in the dataset, see (Supplemental File 4). This excluded transfers between species such as *D.*

saltans and *D. austrosaltans*, or *D. rhopaloa* and *D. carrolli*. At times these species are treated separately - for example if a TE is present only in *D. carrolli* at > 99% but not in *D. rhopaloa*. Otherwise, they were treated as a single unit. Note that if a TE is active in two different groups, i.e. *Zaprionus* and *melanogaster*, it is not necessarily > 99% similar between the two groups. However, to ensure that we are annotating the same TE, we followed the 80-80-80 rule, such that they are 80% similar for 80% of their length (Supplemental File 6). Whether two TEs that are 80% similar and active in *Zaprionus* and *melanogaster* are the same TE is a philosophical matter for another manuscript. However, please see Supplemental File 6 for a discussion of the impact of this rule upon the dataset.

Manual curation of the HT dataset

Because of the difficulty in making phylogenetic inferences about HT, we manually curated a dataset using the HT events described above as our starting point. It is quite difficult to infer the direction of transfer of a TE when it is > 99% similar. In terms of which TE transferred to who our dataset is direction agnostic. When transfer between more than two species that are all > 99% similar was noted, we used the branch length and phylogenetic position assigned by MrBayes to assign transfer partners. This may be somewhat arbitrary, but the majority of our analysis is not based on the exact donor species within a group. Note that if the route by which a TE transferred was unclear - i.e. equal similarity between *D. bipectinata*, *D. pseudoananassae*, *Z. inermis*, and *Z. nigranus*, we took the most parsimonious route. In this example, a transfer would occur between *D. bipectinata* and *D. pseudoananassae*, followed by a single transfer to either of these species to the *Zaprionus* species, and a transfer between the two *Zaprionus* species. Essentially, intra-group transfers were favored over inter-group transfers, and inter-group transfers were favored over inter-genera transfers.

Finally, we split our dataset into two groups - one group contains all of the transfers that were manually annotated, and the other includes only those for which we also have information on the species tree distance [Finet et al., 2021]. Everywhere that we do not mention tree distance we are using the former dataset, as it contains more transfers. The creation of the dataset which includes tree distances is described in the following section.

Curation of HT events with tree distance

Not all species that participated in HT were included in the larger phylogenetic tree that we used to determine species distance [Finet et al., 2021]. To make conclusions based on the distance between the species transferring TEs, we curated a second dataset including only those present in the tree. In addition, we excluded species that were very closely related, as aforementioned. This analysis also provides validation for our data through statistical means.

Results

TE Dataset

TEs less than 1000 bp were removed from the dataset, as were *R* elements and satellite repeats. The final dataset included 1410 TEs. After blasting to the *Drosophila* assemblies we removed TEs found in three species or less. We identified 637 invasions with > 99% similarity between species (Supplemental File 3). The majority of these invasions occurred within the *melanogaster*, *ananassae*, *willistoni*, and *zaprionus* species groups (fig. 1). TEs are frequently active in different groups and therefore at some time a transfer between groups was necessary, but it is only included in the dataset if the TEs are > 99% similar between species. In addition, it is possible that the TE is vertically inherited between species and we make no claim in that regard if it is not included in our dataset. In general, we assumed the most parsimonious explanation for TE transfer if the exact transfer partner was not clear - we favored transfers within groups and within genera. For example, if a TE transferred from *Drosophila* to multiple species of *Zaprionus*, we assumed a single transfer from *Drosophila* followed by within-group transfers rather than multiple transfers from *Drosophila* (fig. 1). The most common TE type to be involved in an HT was *BEL*, *Mariner*, and *Gypsy*. The *Mariner*

has previously been recognized to be involved in frequent HT between species [Maruyama and Hartl, 1991b,a, Wallau et al., 2014, Robertson, 1993]. For a discussion of the effect of using a consensus sequences versus an individual copy of a TE, please see (Supplemental File 1).

TE dataset including species distance

We produced a second dataset which is a subset of the aforementioned dataset (Supplemental File 2, 5). In this second dataset we began with the 637 HT events documented in the previous section, then removed transfers that involved species not included in [Finet et al., 2021]. From the Drosophyla dataset we were able to associate a tree distance with each species, so that we had a numerical record of the distance within the tree that a TE had traveled. Then, from the trees we constructed in MrBayes, we could also include a TE distance which represents the measure of evolutionary similarity of TE sequences between species pairs. An overview of this data is shown in (Supplemental File 1, Table 1).

Horizontal transfer of transposable elements

The total number of TEs involved in HT is 272 (for 637 invasions), 47% of which are documented in a single transfer. Of the 272 TEs included in the HT dataset, 57% of these TEs are LTRs, 21% are DNA elements, and 21% are Non-LTRs. Given that the dataset consists of TEs curated from *Drosophila*, the vast majority of transfers were recorded between species of *Drosophila* (560). However, the second most common genus in which we detected transfers was between species of the *Zaprionus* genus. The most common type of transfer was LTR elements in *Drosophila*, followed by the second and third most common types of TE to transfer were DNA elements and Non-LTRs in *Drosophila*, and LTRs in *Zaprionus*.

The vast majority of HT events occurred between closely related species within the same subgroup. However, 149 of the transfers occur between different subgroups (i.e. the *melanogaster* or *ananassae* species subgroup). 46 transfers occurred between different genera, such as between *zaprionus* and *Drosophila*. Furthermore, 111 TEs are active in more than one group, even if a transfer between the groups is not documented. This portion of the dataset is dependent upon the 80-80-80 rule, where the number of TEs included in this group would be lower if we required higher similarity between all TEs. If we raised the similarity requirement to 90%, 18 TEs would be split into two TEs and reduce this number. This list overlaps with the number where a transfer is recorded, but is largely distinct. For example, *shellder* is active in the *cardini*, *mulleri*, *willistoni*, and *melanogaster* groups. However, only transfers between *cardini*, *mulleri*, and *melanogaster* are recorded. This dataset is already very rich and complicated, and we did not wish to make it more so by including older transfers, but this is worth noting. In other cases, such as *DOC*, all of the transfers between groups are included. For example, *DOC* is active in the *zaprionus*, *leucophenga*, and *melanogaster* groups. The dataset includes transfers between *D. yakuba* (*melanogaster*) and *Leucophenga varia* as well as *Zaprionus africanus* and *Leucophenga varia*.

TEs with the most horizontal transfer events

The top five TEs involved in HT are all *Mariner* elements. 7 of the top 15 HT TEs are *gypsy*-class elements, while *Doc* and *I-2* are Non-LTRs. *Transib-5 D. bipectinata* is a DNA element like *Mariner* but of a different class. The most active element is *Mariner-1 D. elegans*, with 16 transfers (fig. 3). This TE has no transfers recorded from *D. elegans*, but is active in the *ananassae*, *ficusphila*, *chymomyza*, *quinaria*, *fenestrarum* (*Lordophisa*), *leucophenga*, *miki* (*Lordiphosa*), *montium*, *obscura*, *testacea* , and *rhopaloa* subgroups. We do not believe the HT events in this transposon have been previously documented, though it is difficult to tell given the vagaries of TE nomenclature and phylogeny. We propose to name this TE *Minona*. Between groups this TE is $> 96\%$ similar, thus the number of transfers would not be affected by changes to the 80-80-80 rule (fig. 3).

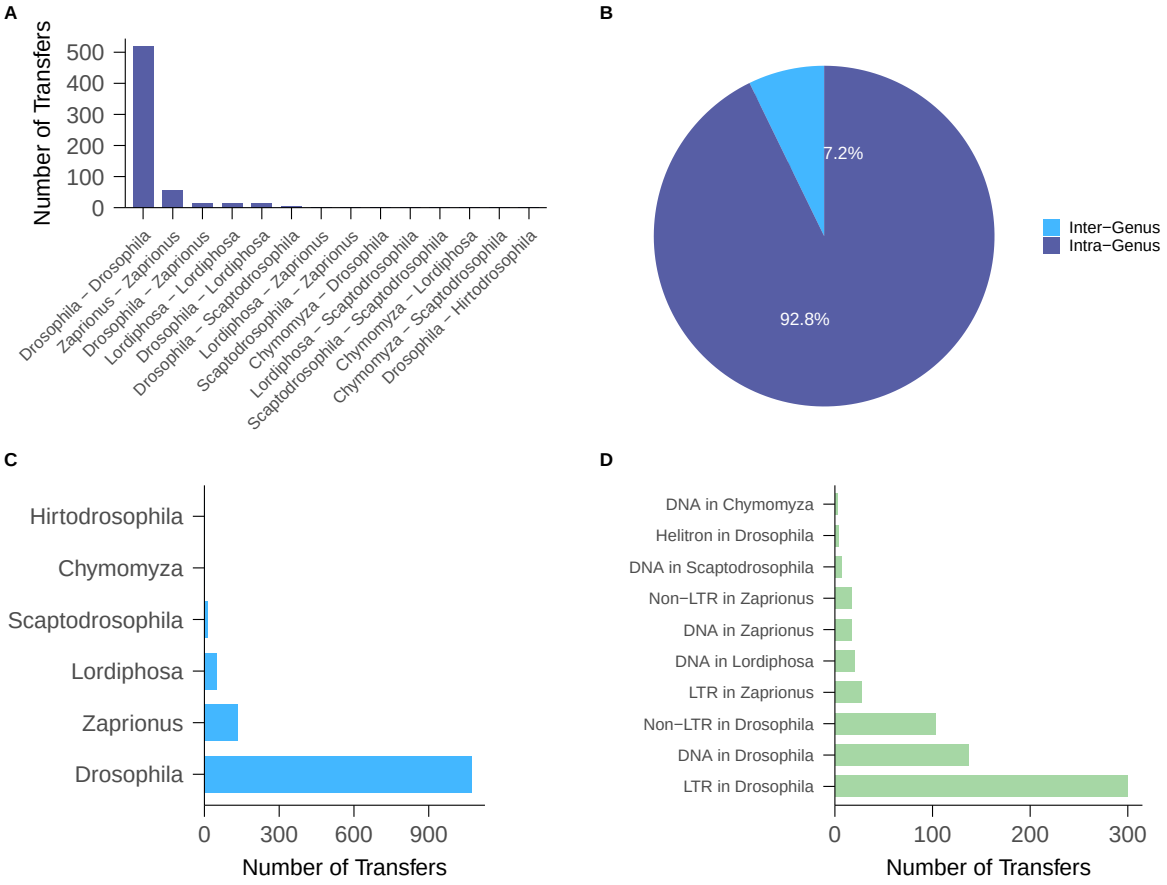


Figure 1: HT at the level of genus, i.e. *Drosophila* or *Zaprionus*. **A**. The number of TE transfers between genera for each pair in the dataset. This is agnostic to the direction of transfer. **B**. Summary of the number of HT evens that happen between different genera versus within a genus. **C**. The number of different transfers recorded for each genus included in the dataset. Note that both partners are counted for this dataset so it totals twice the other datasets. **D**. Classes of TE and genera of the top 10 types of HT observed in the dataset.

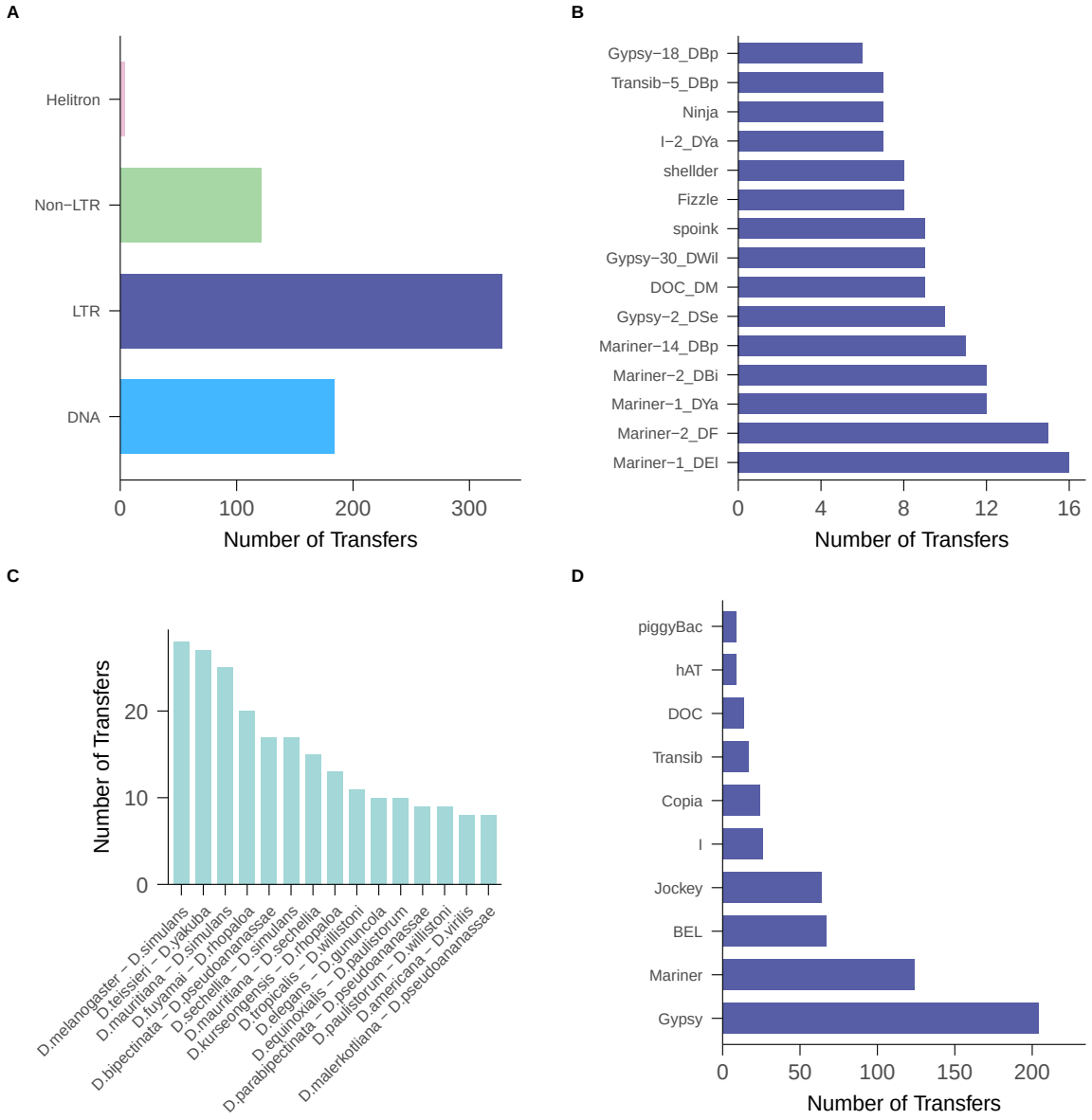


Figure 2: HT patterns among different TEs. A. The frequency of HT observed for each type of TE. B. The TEs with the 15 top recorded transfer events. C. The species pairs with the largest number of recorded transfer events. Note that the exact species pair that a TE transferred between is often ambiguous, and we favored species pairs over more distant transfers. However, this graph is largely reflective of overall patterns in the data, i.e. the high rate of transfer in the *rhopaloa* subgroup, *bipectinata* complex, and *melanogaster* subgroup. Species pairs that were not included in these groups include *D. elegans* - *D. gununcola* and *D. ercepeae* - *D. merina*. The latter is interesting in particular as neither species has been the subject of TE annotation (unlike *D. elegans*). D. TE classes with top 10 recorded transfer events.

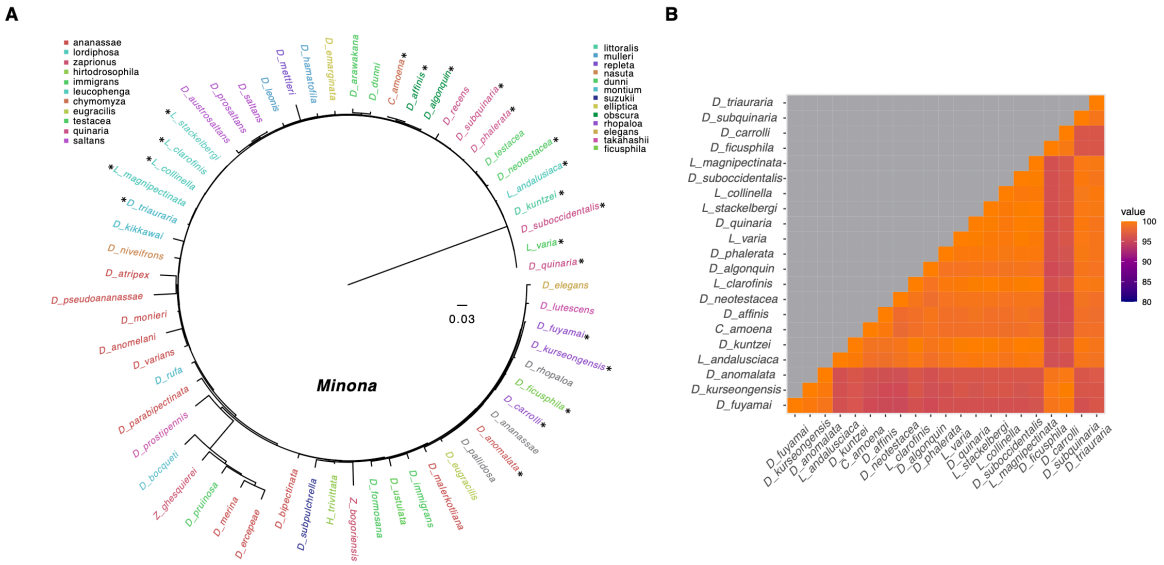


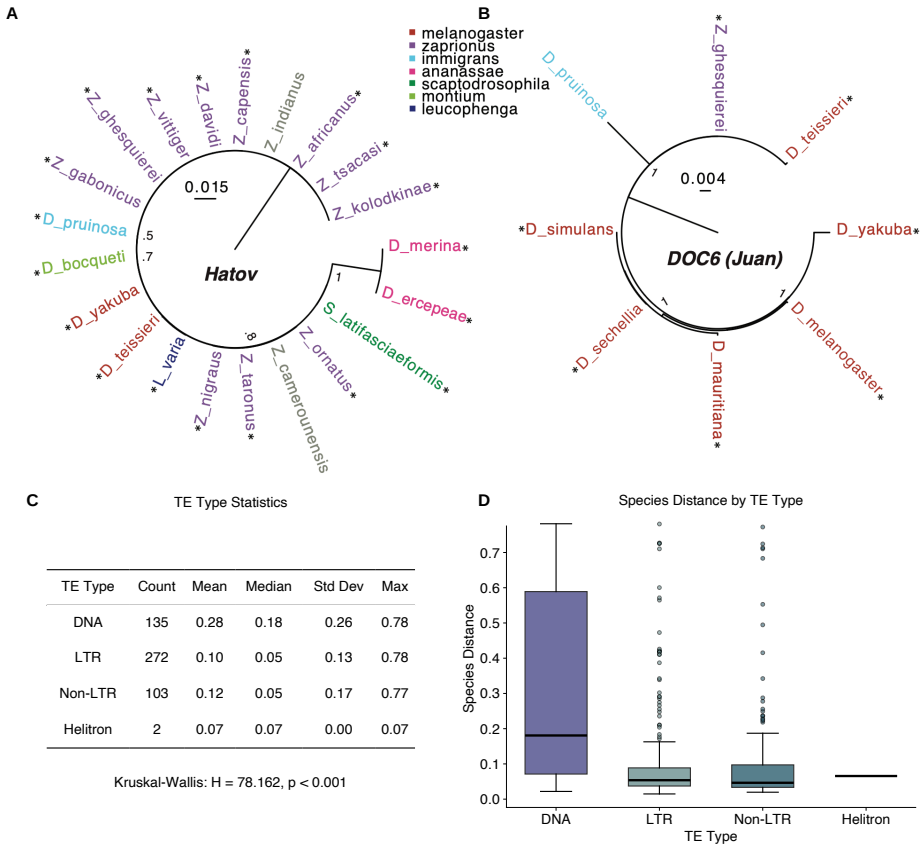
Figure 3: Phylogenetic tree of *Minona*. Each species name is colored according to its subgroup. *Zaprionus* species were not separated by subgroup in this figure but rather by genus. Gray species indicate that a transfer was documented in a relative too close to consider it an independent transfer event. For example, *D. ananassae*, *D. anomalata*, and *D. pallidosa* are all too similar to rule out inheritance by common descent, but a transfer into *D. pallidosa* is included. A star following the species name indicates a horizontal transfer events. B. A heatmap of similarity between TE sequences from species with a horizontal transfer event. Across all of the species with an active *Minona* element, the TE sequences are > 96% similar. In addition, similarity is not predicted by phylogenetic relationship - TE sequences from *Lordiphosa* species are more closely related to *D. kuntzei* and *D. triauraria* than *D. fuyamai* is to either.

The second most active TE is *Mariner-2 D. ficusphila*, which is active in the *chymomyza*, *grimshawii*, *planitibia*, *crucigera*, *leucophenga*, *fenestrarum* (*Lordiphosa*), *orphnopeza*, *picticornis*, and *split-tarsi* subgroups. The Hawaiian *Drosophila* contains many members that are closely related, thus these datasets were heavily pruned to remove any species that could have shared TEs through shared ancestry. These are all Hawaiian groups other than *chymomyza*, *lecophenga* and *fenestrarum* (*Lordiphosa*). If the threshold for inclusion is changed to 98.5%, then the number of transfers increases by two to include *D. ficusphila* and *D. rufa*. There is more than > 10% divergence between the Hawaiian and non-Hawaiian TE populations, thus they could be considered separate invasions depending upon the philosophy of TE divergence. We propose to name this TE *Veena*.

Gypsy-30 D. willistoni is among the most active TEs, though only two transfers occurred in the *willistoni* group. Rather, this TE is active in the *obscura*, *melanica*, *montana*, *robusta*, *saltans*, *willistoni*, and *zaprionus* groups. While the transfer to *zaprionus* is not included in the dataset, it most likely came from the *melanogaster* group. Increasing the threshold to 98.5% increases the number of transfers by five, to include two more members of the *willistoni* group (*D. paulistorum* and *D. equinoxialis*, as well as *D. tripunctata, D. putrida*, and *D. americana*. This would increase the number of subgroups in which the TE is active as well, to include *tripunctata* and *testacea*.This helps demonstrate the difficulty with naming TEs after a number and species, as they are typically active in groups outside of that species name, if they are active in the named species at all. Copies of this TE are more than 90% similar between groups. Typically copies are more than 97% similiar, with those between *Zaprionus* and *Drosophila* groups between 92%-94% similar. We propose to name this TE *Finnegas*.

Another TE with extensive HT is *Mariner-1 D. yakuba*, which is not active within most members in the *melanogaster* group, but is present in *D. yakuba* and *D. teissieri* (fig. 4). It has also transferred between

10 species of the *zaprionus* group, *Leucophenga varia*, *Scaptodrosophila latifasciaeformis*, and two members of the *ananassae* subgroup *D.* merina, D. ercepeae. Increasing the threshold to 98.5% does not change the results. We propose to name this TE *Hatov*. The transfer between *D. merina* and *D. erecepeae* would be attributed to a different TE if we were to raise the required similarity between all TEs to 90%. We have also named three additional TEs in this dataset - *Sojourner*, Judii, and *Fizzle* - which corresponded to more than one other generically named TE and were renamed to reduce confusion (i.e. *Fizzle* corresponds to three other TEs in Repbase with different names consisting of *gypsy* and a number).



TE Type	Count	Mean	Median	Std Dev	Max
DNA	135	0.28	0.18	0.26	0.78
LTR	272	0.10	0.05	0.13	0.78
Non-LTR	103	0.12	0.05	0.17	0.77
Helitron	2	0.07	0.07	0.00	0.07

Figure 4: A-B. Phylogenetic trees showing the relationship between TEs in different species of dipterans. A. *Hatov*, a TE that has been spreading primarily in African dipterans. Species from different groups are colored according to subgroup. The key indicates colors for both panels. B. The second TE is *DOC6* (*Juan*) which has been documented spreading in the *melanogaster* group. C-D. TE type distribution and species distance analysis across 512 mapped transfers. C. Table shows the distribution statistics of the species pair distance for each transfer for all four TE types: DNA, LTR, Non-LTR and Helitron. The mean and median corresponding the DNA TEs are significantly greater than other TEs. D. A visualization of the same data.

Phylogenetic distance of horizontal transfer

We also wished to determine the species distance that different types of TEs are traveling. For example, do LTRs move between different genera more commonly than non-LTRs? Using the species distance from Drosophyla [Finet et al., 2021], we found that DNA transposons move larger distances by far compared to LTRs, Non-LTRs, and Helitrons ($p < 0.001$, fig. 4). While LTRs make up the vast majority of transfers, DNA

transposons on average move larger phylogenetic distances. Note that this result is not affected by the 80-80-80 rule, which only applies to the number of TEs given the same name - distance of transfer was measured using only TEs that were > 99% similar. This could represent a separate evolutionary strategy, in which DNA transposons move less frequently to more naive genomes, potentially facilitating their establishment (fig. 4). This pattern is largely driven by *Mariner* elements, which make up 67% of transfers due to DNA transposons in the dataset. It is also possible that *Mariner* elements evolve more slowly, but then the expectation would be to see more intermediate transfers, because it would need to make a series of smaller steps to spread into more distant species.

The TE with some of the largest transfer distances is *POGO* (Figure 5). The *POGO* sequences present in different species are more than 98% similar across all genomes, thus this is not affected by the 80-80-80 rule. However, because the TE sequences are so closely related, the exact order of transfer is unclear, i.e. did *L. varia* donate to *D. melanogaster* or did *Zaprionus*? *POGO* has never been detected outside the *D. melanogaster* (to our knowledge) [Tudor et al., 1992]. Its presence in multiple species of *Zaprionus* suggests a likely origin in that genus, but more species of *Scaptodrosophila* and *Leucophenga* would need to be sequenced.

HT in D. melanogaster

50 of the recent TE transfers involve *D. melanogaster* and have largely been documented elsewhere, for example *opus*, *Burdock*, and *Max*-element [Vieira et al., 1999, Pianezza and Kofler, 2025, Scarpa et al., 2023, Signor et al., 2023]. Indeed, the 11 transfers documented in *D. melanogaster* in the last 200 years were recapitulated in this dataset, apart from the three most recent which had not yet invaded the reference genome (*Souslik*, *MCLE*, and *Transib1*) [Scarpa et al., 2025, 2023, Pianezza et al., 2023]. *Opus* was fragmented in *D. melanogaster* and was not included in this dataset. Less well documented transfers include *Chimpo*, *BEL*, and *Blastopia* [Bargues and Lerat, 2017]. For example, in *Blastopia* transfers between *D. melanogaster*, *D. sechellia* and *D. simulans* were documented, but we were able to add a transfer to *D. merina* and *D. mauritiana*. *D. melanogaster* also either donated or acquired the *POGO* element from *Z. capensis* and *S.latifasciaformis*, which was not previously documented (Figure 5). Aside from *POGO* (and the well documented *P*-element) species outside the *melanogaster* complex were not frequent donors and/or receivers with *D. melanogaster*, however *D. merina* likely donated or received *Transpac* and *ACCORD*, while *D. vallismaia* was likely involved in HT of *Bari1*. *D. vallismaia* and *D. merina* are both members of the *ananassae* group. *D. merina* is native to Madagascar, while *D. vallismaia* is native to the Seychelles. The relative importance of *D. merina* and other members of that subgroup such as *D. ercepeae* in HT with the *melanogaster group* is surprising, and has been noted elsewhere [Pianezza and Kofler, 2025].

In general, our results were consistent with those previously produced when the datasets overlapped [Pianezza and Kofler, 2025]. For example, we document the transfer of *412* likely from *D. simulans* to *D. melanogaster*. *DOC6* shows the same pattern of spread in our dataset as previously documented - between members of the *melanogaster* species group and *Z. ghesquierei* - with the addition of its presence in *D. yakuba* (Figure 4). For *diver* we document the donation from *D. simulans* to *D. melanogaster* [Pianezza and Kofler, 2025], as well as transfers to *D. sechellia*, *D. mauritiana*, between *D. teissieri* and *D. yakuba*, and between *Z. tsacasi* and *Z. kolodkinae*. Differences in TEs and species used as well as the timescale over which HT was considered contributed to differences in the datasets. For example, previously the transfer of *Stalker4* was documented between *D. mauritiana* and *D. melanogaster*, which our broader dataset also confirms, but is less than 99% similar. The spread of *Spoink* and *Shellder* was recently documented and is also recapitulated in our dataset [Pianezza et al., 2023, Scarpa et al., 2025].

Abundant HT in other species

At the level of subgroup, as expected, *melanogaster* has the most transfers at 159. The *D. melanogaster* genome has been extensively analyzed for TE content, and all of those TEs are in this dataset, so the dataset is inherently biased towards detecting transfers in the *melanogaster* group. However, the *rhopaloa* subgroup has 56 annotated HT events. There are also 46 within the *bipectinata* species complex, a subset of the *ananassae* subgroup. While the abundance of HT in the *melanogaster* subgroup is undoubtedly due

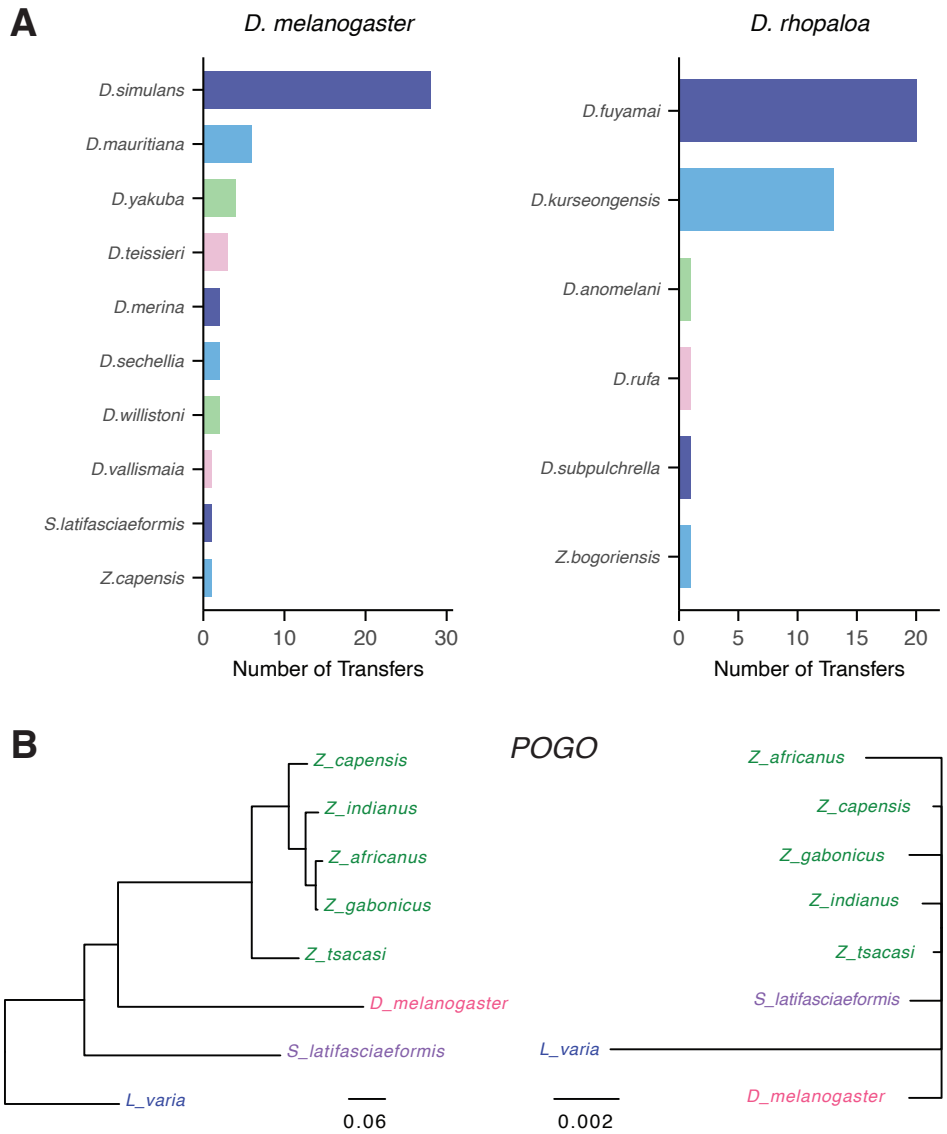


Figure 5: *D. melanogaster* and *D. rhopaloa* were involved in a large number of HT events, shown here split by species partner. This figure does not imply directionality, i.e. that *D. melanogaster* is the donor or acceptor of the TEs. *D. melanogaster* is more concentrated with its sister species *D. simulans* than *D. rhopaloa*, but both species are involved in a number of transfer events. B. Phylogenetic tree of the species whose genomes harbor *POGO* using distances from Drosophyla, and a tree of *POGO* itself (right hand side). Note the large difference in scale between the two trees. The *POGO* phylogeny is essentially a star, with no differentiation between the majority of the species.

somewhat to the level of scrutiny the group has received, the *bipectinata* and *rhopaloa* groups have not been extensively annotated.

D. yakuba has been involved in 39 HT events that happened recently enough to be 99% similar. The majority of the transfers occur between *D. yakuba* and *D. teissieri*, totaling 27 and including TEs such as *MDG3, Minos, Burdock, Chimpo*, and *Tabor*. This also includes instances between other genera such as *DOC* which likely moved between *D. yakuba* and *Leucophenga varia*. *Mariner-1 D. yakuba* also likely moved from *D. yakuba* to *Z. inermis*.

D. rhopaloa is involved in 37 HT events. 33 of these occur within the *rhopaloa* subgroup and include TEs such as *CR1A*, *Outcast-1 D. eugracilis*, *Copia-3 D. bipectinata*, and *I-2 D. elegans*. *D. rhopaloa* was also involved in HT with *D. rufa* (*BEL-1 D.bipectinata*), *Z. bogoriensis* and *D. subpulchrella* (*Mariner-2 D. bipectinata*), and *D. anomelani* (*Mariner-1 D. biarmipes*). We do not infer directionality here, but this does demonstrate that HT is occurring at high levels outside of *D. melanogaster*, even in species that are not considered human commensals. *dmel* does appear to most frequently receive TEs rather than donate them, and these species may not be biased in such a manner [Pianezza and Kofler, 2025].

The *D. bipectinata* subgroup is of particular interest here, as we did not allow transfers between members of the *bipectinata* complex: *D. bipectinata, D. malerkotliana*, or *D. parabipectinata* because they are suspected to still be interbreeding [Kopp and Barmina, 2005]. We did allow transfers from these species to *D. pseudoananassae* or any other species. Even restricting transfers to outside the complex, we still recorded 46 HT events involving this subgroup. The vast majority of these are from the three species in the *bipectinata* complex to *D. pseudoananassae* (34), which do not inter-breed. Other species involved in HT included *D. atripex, D. anomelani, D. eugracilis, D. birchii, D. biarmipes, D. ananassae, D. rufa*, and *D. varians*. Indeed, *ninja* (also known as *Juan*) was documented as moving between *D. birchii* and *D. pseudoananassae*. *Ninja* was only previously documented within the *melanogaster* complex, however we documented it within the *ananassae, rhopaloa*, and *Zaprionus* groups. Note that *Ninja* is more than 10% diverged outside of the *zaprionus* and *melanogaster* groups, so these conclusions would be altered by changing the 80-80-80 rule. It is unclear why the *bipectinata* complex would be such a hotspot for horizontal transfer - its TEs are no more well annotated than any other non-*melanogaster* group in the dataset.

Discussion

In this paper we set out to track recent invasions of TEs across almost 400 *Drosophila* genomes (as well as close relatives). We detected 637 recent invasions of 272 different TEs, filtering for TE sequences that were $> 99\%$ similar. LTRs were responsible for almost 50% of TE invasions, however these invasions typically occurred over shorter evolutionary distances than invasions by DNA elements. This suggests that the two types of TEs may have different invasion strategies. For example, LTRs may move more quickly but be restricted to more similar genomes, while DNA transposons are more flexible with regard to the genome but move larger phylogenetic distances less frequently. Larger phylogenetic distances may ensure that the genome is naive with regard to the TE, compared to a close relative which may have already encountered a similar TE. Invasion of a similar TE would trigger the piRNA pathway to suppress the TE as it attempted to invade [Gainetdinov et al., 2023]. It is also possible that they use different vectors to spread, some of which are less specialized to a particular genus of drosophilid [Silva et al., 2004, Loreto et al., 2008, Silva and Kidwell, 2000]. If *Mariner* elements were slower to evolve, the expectation would be to find more intermediate species rather than large jumps. *Mariner* elements are known to be phylogenetically widespread and active in many different species [Capy et al., 1990, Lohe et al., 1995, Robertson, 1993, Wallau et al., 2014]. An interesting future direction would be to compare copy number between the different types of TEs, as this could also reflect different evolutionary strategies between TE types.

Many factors are going to bias the distribution of HT that we uncover - for example the number of species within a group that is sequenced or the number of annotated TEs that originated from that group. For example, given the propensity for TEs to spread between close relatives it is more likely that you will detect an HT event if more species within a group have been sequenced. The relationship between TEs annotated in a group and the likelihood of finding HT is less clear, it was very common for a TE named after a species,

and for it to be active primarily in a completely different group. For example, *Mariner-1 D. elegans* (named *Minona* in this manuscript) is present in > 20 different subgroups. While present in *D. elegans*, it did not engage in HT recently from this species. This is an extreme example, however it goes to show that while bias in TE annotation (for example, the focus on *D. melanogaster*) will impact the distribution of HT we find, it is not as clear cut a relationship as one might expect.

Our summary of the number of HT events between subgroups does represent very different phylogenetic distances - for example a jump between the *montana* group and *littoralis* groups (both in the *virilis* group) is much different than a jump between *Leucophenga* and *melanogaster*. Regardless, it is a useful categorization of transfers between quite closely related species and those further away. The way in which we assigned transfer partners was also based on assumptions that are not always true - we prioritized within subgroup transfers over between subgroup transfers and both over between-genera transfers. It is entirely possible that the rate of transfer between different genera is higher, and has happened multiple times for a single TE (i.e. multiple transfers between *Drosophila* and *Zaprionus* rather than a single transfer followed by within-*Zaprionus* transfers). For the vast majority of the HT events categorized in this paper, there is no way to reliably distinguish transfer partners given the high sequence similarity between TEs.

We also find a high rate of recent HT in quite a number of subgroups in addition to *D. melanogaster*. This suggests that high rates of HT may not be unique to *D. melanogaster*. It has been suggested that HT has been accelerated by human movement, which is likely, but perhaps human movement of different drosophilids is extensive enough that it is not limited by the known distribution of the species i.e. the majority of species are essentially cosmopolitan. It is also possible that it is movement of the species pathogens that is impacting HT (if that is a vector for HT), as new pathogens might spread more readily than new flies. We do not yet know if this rate of invasion is historical for species not considered human commensals, but this is an interesting open question.

A thorough understanding of HT between species has been hindered by the lack of complete assemblies of repetitive regions, which was not possible until the advent of long-read sequencing [van Dijk et al., 2023]. The pace at which this data is now being produced is creating a unique opportunity to understand the dynamics of repetitive elements. However, it can still be difficult to place one's results in the context of existing literature. For example, one issue that has hindered our understanding of HT is confusion over the level at which HT is detected - for example if *Mariner* is detected across insect species, does this refer to the class, a group within *Mariner*, or an individual element. Another is naming, TEs are frequently given generic names such as *gypsy-5* through *50*, followed by a species name, and it can be difficult to differentiate different elements (and there are many repeats). They are also frequently not deposited in any central resource such as NCBI or Dfam. Improved central databases and naming conventions will help ameliorate these problems as more TEs are discovered. Approaches which emphasize reference-free TE discovery will also be of considerable help in this regard.

Many of the TEs in the original dataset were not found to have undergone recent HT. However, our cutoff of > 99% similarity was somewhat arbitrary and does not necessarily imply that they have not been active in the somewhat recent past. A subsequent study which focuses on estimating genome divergence relative to TE divergence will be a welcome contribution to our understanding of HT in the recent past. In addition, investment in museum sequencing will contribute to our understanding of the rate of TE invasion in these other species [Scarpa et al., 2023]. There are still many unanswered questions about HT in dipterans - for example how does a new TE invasion impact the population of flies? In one study, we found that the population genetics of *D. sechellia* was very different on islands that had been invaded by two TEs versus islands that had not [Scarpa et al., 2025]. Islands that had been invaded by the two TEs had flies that were all very closely related to each other, while islands without the two TEs had more population structure [Scarpa et al., 2025]. Can TE invasions cause population bottlenecks? Is there selection for TE insertions within TE silencing regions, or are insertions into these regions abundant as has been suggested for the *P*-element [Wang et al., 2023]? How do the invasion dynamics of TEs with envelope proteins differ from those that do not have envelope proteins [Senti et al., 2025]? While many questions remain about the dynamics and genomic consequences of TE invasions, it is clear that they are a powerful force shaping the evolution of species genomes.

Supplemental Data

Supplemental File 1 includes machine learning model analysis, methodological discussion of consensus sequence approaches, and supplementary figures referenced throughout this manuscript. We provide comprehensive datasets supporting this analysis: an extensive feature dataset (Supplemental File 2), a manually curated species invasion list (Supplemental File 3), a list of excluded species pairs (Supplemental File 4), and a manually curated species invasion list with features (Supplemental File 5). Supplemental File 6 includes a discussion of the application of the 80-80-80 rule to the dataset and a list of % identity between TEs spreading in different subgroups. These materials ensure reproducibility and support future comparative studies of horizontal transfer dynamics.

Acknowledgments

SS would like to thank F. Belt, S. & F. Emery, and P. Senn for inspiration during the production of this manuscript. SP would like to thank his late grandma Shanti Devi for helping him learn how to become a better human being.

Author contributions

S.S. conceived the project. S.P. performed computational analyses; S.S. performed manual curation. S.P. and S.S. co-wrote the manuscript.

Funding

This work was supported by the National Science Foundation Established Program to Stimulate Competitive Research grants NSF-EPSCoR-1826834 and NSF-EPSCoR-2032756 to SS. In addition this work was funding by the National Institutes of Health R35GM155272 to SS.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

Data Availability

The genomes used in this study are all publicly available from our cited sources, as are the TE sequences used [Chakraborty et al., 2021, Kim et al., 2021, 2023]. Newly named TEs such as *Minona* were deposited in Dfam and NCBI. The processed data discussed in this manuscript is all included in the supplemental files.

References

N. Bargues and E. Lerat. Evolutionary history of ltr-retrotransposons among 20 drosophila species. *Mobile DNA*, 8:1–15, 2017.

C. Bartolomé, X. Bello, and X. Maside. Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome biology*, 10(2):R22, 2009.

E. Bonnivard, C. Bazin, B. Denis, and D. Higuet. A scenario for the hobo transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem TPE repeats. *Genetical Research*, 75(1):13–23, 2000.

J. Brennecke, A. A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G. J. Hannon. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6):1089–1103, 2007.

A. Bucheton, C. Vaury, M. C. Chaboissier, P. Abad, A. Pélisson, and M. Simonelig. I elements and the *Drosophila* genome. *Genetica*, 86(1-3):175–90, 1992.

C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. Blast+: architecture and applications. *BMC bioinformatics*, 10(1):421, 2009.

P. Capy, F. Chakrani, F. Lemeunier, D. Hartl, and J. David. Active mariner transposable elements are widespread in natural populations of drosophila simulans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 242(1303):57–60, 1990.

M. Chakraborty, C. Chang, D. Khost, J. A. J Vedanayagam, Y. Liao, K. Montooth, C. Meiklejohn, A. Larracuente, and J. Emerson. Evolution of genome structure in the *Drosophila simulans* species complex. *Genome Research*, 31:380–396, 2021.

S. B. Daniels, K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell, and A. Chovnick. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*, 124(2):339–55, 1990.

D. G. Eickbush and T. H. Eickbush. Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics*, 139(2):671–684, 1995.

W. R. Engels. The P family of transposable elements in *Drosophila*. *Annual review of genetics*, 17:315–44, 1983. ISSN 0066-4197. URL `http://www.ncbi.nlm.nih.gov/pubmed/6320712`.

C. Feschotte and E. J. Pritham. DNA transposons and the evolution of eukaryotic genomes. *Annual review of genetics*, 41:331–68, 2007.

C. Finet, V. A. Kassner, A. B. Carvalho, H. Chung, J. P. Day, S. Day, E. K. Delaney, F. C. De Ré, H. D. Dufour, E. Dupim, et al. Drosophyla: resources for drosophilid phylogeny and systematics. *Genome biology and evolution*, 13(8):evab179, 2021.

M. Fraser, J. S. Brusca, G. E. Smith, and M. D. Summers. Transposon-mediated mutagenesis of a baculovirus. *Virology*, 145(2):356–361, 1985.

I. Gainetdinov, J. Vega-Badillo, K. Cecchini, A. Bagci, C. Colpan, D. De, S. Bailey, A. Arif, P.-H. Wu, I. J. MacRae, et al. Relaxed targeting rules help piwi proteins silence transposons. *Nature*, 619(7969):394–402, 2023.

L. S. Gunawardane, K. Saito, K. M. Nishida, K. Miyoshi, Y. Kawamura, T. Nagami, H. Siomi, and M. C. Siomi. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, 315(5818):1587–1590, 2007.

C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. Van Kerkwijk, M. Brett, A. Haldane, J. F. Del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2649-2. URL `https://www.nature.com/articles/s41586-020-2649-2`.

B. D. Heath, R. D. Butcher, W. G. Whitfield, and S. F. Hubbard. Horizontal transfer of wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. *Current Biology*, 9(6):313–316, 1999.

M. A. Houck, J. B. Clark, K. R. Peterson, and M. G. Kidwell. Possible horizontal transfer of drosophila genes by the mite proctolaelaps regalis. *Science*, 253(5024):1125–1128, 1991.

J. P. Huelsenbeck and F. Ronquist. Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17 (8):754–755, 2001.

J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55. URL `http://ieeexplore.ieee.org/document/4160265/`.

J. Jehle, E. Fritsch, A. Nickel, J. Huber, and H. Backhaus. Tcl4. 7: a novel lepidopteran transposon found in cydia pomonella granulosis virus. *Virology*, 207(2):369–379, 1995.

K. Katoh and H. Toh. Recent developments in the mafft multiple sequence alignment program. *Briefings in bioinformatics*, 9(4):286–298, 2008.

K. Katoh, K. Misawa, K.-i. Kuma, and T. Miyata. Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic acids research*, 30(14):3059–3066, 2002.

K. Katoh, G. Asimenos, and H. Toh. Multiple alignment of dna sequences with mafft. In *Bioinformatics for DNA sequence analysis*, pages 39–64. Springer, 2009.

A. Keller. Drosophila melanogaster's history as a human commensal. *Current biology*, 17(3):R77–R81, 2007.

B. Y. Kim, J. R. Wang, D. E. Miller, O. Barmina, E. Delaney, A. Thompson, A. A. Comeault, D. Peede, E. R. D'Agostino, J. Pelaez, J. M. Aguilar, D. Haji, T. Matsunaga, E. E. Armstrong, M. Zych, Y. Ogawa, M. Stamenković-Radak, M. Jelić, M. S. Veselinović, M. Tanasković, P. Erić, J.-J. Gao, T. K. Katoh, M. J. Toda, H. Watabe, M. Watada, J. S. Davis, L. C. Moyle, G. Manoli, E. Bertolini, V. Koštál, R. S. Hawley, A. Takahashi, C. D. Jones, D. K. Price, N. Whiteman, A. Kopp, D. R. Matute, and D. A. Petrov. Highly contiguous assemblies of 101 drosophilid genomes. *eLife*, 10:e66405, 2021.

B. Y. Kim, H. R. Gellert, S. H. Church, A. Suvorov, S. S. Anderson, O. Barmina, S. G. Beskid, A. A. Comeault, K. Nicole Crown, S. E. Diamond, S. Dorus, T. Fujichika, J. A. Hemker, J. Hrcek, M. Kankare, T. Katoh, K. N. Magnacca, R. A. Martin, T. Matsunaga, M. J. Medeiros, D. E. Miller, S. Pitnick, S. Simoni, T. E. Steenwinkel, M. Schiffer, Z. A. Syed, A. Takahashi, K. H.-C. Wei, T. Yokoyama, M. B. Eisen, A. Kopp, D. Matute, D. J. Obbard, P. M. O'Grady, D. K. Price, M. J. Toda, T. Werner, and D. A. Petrov. Single-fly assemblies fill major phylogenomic gaps across the drosophilidae tree of life. *bioRxiv*, Oct. 2023. doi: 10.1101/2023.10.02.560517. URL `http://dx.doi.org/10.1101/2023.10.02.560517`.

R. Kofler, T. Hill, V. Nolte, A. Betancourt, and C. Schlötterer. The recent invasion of natural *Drosophila simulans* populations by the P-element. *PNAS*, 112(21):6659–6663, 2015a.

R. Kofler, V. Nolte, and C. Schlötterer. Tempo and mode of transposable element activity in *Drosophila*. *PLoS Genetics*, 11(7):e1005406, 2015b.

N. Kondo, N. Nikoh, N. Ijichi, M. Shimada, and T. Fukatsu. Genome fragment of wolbachia endosymbiont transferred to x chromosome of host insect. *Proceedings of the National Academy of Sciences*, 99(22): 14280–14285, 2002.

A. Kopp and O. Barmina. Evolutionary history of the drosophila bipectinata species complex. *Genetics Research*, 85(1):23–46, 2005.

A. R. Lohe, E. N. Moriyama, D.-A. Lidholm, and D. L. Hartl. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Molecular biology and evolution*, 12(1):62–72, 1995.

E. L. S. Loreto, C. M. A. Carareto, and P. Capy. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity*, 100(6):545–54, 2008.

C. D. Malone, J. Brennecke, M. Dus, A. Stark, W. R. McCombie, R. Sachidanandam, and G. J. Hannon. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, 137(3): 522–535, 2009.

K. Maruyama and D. Hartl. Evolution of the transposable element mariner in drosophila species. *Genetics*, 128(2):319, 1991a.

K. Maruyama and D. L. Hartl. Evidence for interspecific transfer of the transposable element mariner between drosophila and zaprionus. *Journal of molecular evolution*, 33(6):514–524, 1991b.

V. Mérel, M. Boulesteix, M. Fablet, and C. Vieira. Transposable elements in drosophila. *Mobile DNA*, 11 (1):23, 2020.

D. W. Miller and L. K. Miller. A virus mutant with an insertion of a copia-like transposable element. *Nature*, 299(5883):562–564, 1982.

L. Modolo, F. Picard, and E. Lerat. A new genome-wide method to track horizontally transferred sequences: application to *Drosophila*. *Genome biology and evolution*, 6(2):416–432, 2014.

M. A. Moreno, M. T. Holder, and J. Sukumaran. DendroPy 5: a mature Python library for phylogenetic computing. *Journal of Open Source Software*, 9(101):6943, Sept. 2024. ISSN 2475-9066. doi: 10.21105/joss.06943. URL `https://joss.theoj.org/papers/10.21105/joss.06943`.

F. F. Nascimento, M. D. Reis, and Z. Yang. A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution*, 1(10):1446–1454, Sept. 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0280-x. URL `https://www.nature.com/articles/s41559-017-0280-x`.

H. Pagès, P. Aboyoun, R. Gentleman, and S. DebRoy. *Biostrings: Efficient manipulation of biological strings*, 2025. URL `https://bioconductor.org/packages/Biostrings`. R package version 2.76.0.

E. Paradis and K. Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. *Bioinformatics*, 35(3):526–528, 2019.

E. Paradis, J. Claude, and K. Strimmer. Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, 20(2):289–290, 2004.

J. Peccoud, V. Loiseau, Cordaux, and C. Gilbert. Massive horizontal transfer of transposable elements in insects. *Proc Natl Acad Sci U S A*, 114(18):4721–26, 2017.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

R. Pianezza and R. Kofler. Biogeography shapes the te landscape of drosophila melanogaster. *bioRxiv*, 2025. doi: 10.1101/2025.05.22.655554. URL `https://www.biorxiv.org/content/early/2025/05/27/2025.05.22.655554`.

R. Pianezza, A. Scarpa, P. Narayanan, S. Signor, and R. Kofler. Spoink, a ltr retrotransposon, invaded *D. melanogaster* populations in the 1990s. *bioRxiv*, 2023.

R. Pianezza, A. Scarpa, A. Haider, S. Signor, and R. Kofler. Spatio-temporal tracking of three novel transposable element invasions in drosophila melanogaster over the last 30 years. *Molecular Biology and Evolution*, page msaf143, 2025.

M. Raasveldt and H. Mühleisen. DuckDB: an Embeddable Analytical Database. In *Proceedings of the 2019 International Conference on Management of Data*, pages 1981–1984, Amsterdam Netherlands, June 2019. ACM. ISBN 9781450356435. doi: 10.1145/3299869.3320212. URL `https://dl.acm.org/doi/10.1145/3299869.3320212`.

P. Rice, I. Longden, and A. Bleasby. Emboss: the european molecular biology open software suite. *Trends in genetics*, 16(6):276–277, 2000.

H. M. Robertson. The mariner transposable element is widespread in insects. *Nature*, 362(6417):241–245, 1993.

F. Ronquist, M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3):539–542, 2012.

A. Sánchez-Gracia, X. Maside, and B. Charlesworth. High rate of horizontal transfer of transposable elements in *Drosophila*. *Trends in genetics : TIG*, 21(4):200–3, 2005.

A. Scarpa, R. Pianezza, F. Wierzbicki, and R. Kofler. Genomes of historical specimens reveal multiple invasions of ltr retrotransposons in *Drosophila melanogaster* populations during the 19th century. *PNAS*, 2023.

A. Scarpa, R. Pianezza, H. R. Gellert, A. Haider, B. Y. Kim, E. C. Lai, R. Kofler, and S. Signor. Double trouble: two retrotransposons triggered a cascade of invasions in drosophila species within the last 50 years. *Nature Communications*, 16(1):516, 2025.

F. Schwarz, F. Wierzbicki, K.-A. Senti, and R. Kofler. Tirant Stealthily Invaded Natural *Drosophila melanogaster* Populations during the Last Century . *Molecular Biology and Evolution*, 38(4):1482–1497, 2021.

K.-A. Senti and J. Brennecke. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in genetics*, 26(12):499–509, 2010.

K.-A. Senti, B. Rafanel, D. Handler, C. Kosiol, C. Schlötterer, and J. Brennecke. Co-evolving infectivity and expression patterns drive the diversification of endogenous retroviruses. *The EMBO Journal*, pages 1–20, 2025.

F. Sievers and D. G. Higgins. Clustal omega for making accurate alignments of many protein sequences. *Protein Science*, 27(1):135–145, 2018.

S. Signor, J. Vedanayagam, B. Kim, F. Wierzbicki, R. Kofler, and E. Lai. Rapid evolutionary diversification of the flamenco locus across simulans clade *Drosophila* species. *PLoS Genet*, 19, 2023.

J. C. Silva and M. G. Kidwell. Horizontal transfer and selection in the evolution of P elements. *Molecular biology and evolution*, 17(10):1542–57, 2000.

J. C. Silva, E. L. Loreto, and J. B. Clark. Factors that affect the horizontal transfer of transposable elements. *Current issues in molecular biology*, 6:57–71, 2004.

S. U. Song, T. Gerasimova, M. Kurkulos, J. D. Boeke, and V. G. Corces. An env-like protein encoded by a drosophila retroelement: evidence that gypsy is an infectious retrovirus. *Genes & development*, 8(17): 2046–2057, 1994.

S. Srivastav, C. Feschotte, and A. G. Clark. Rapid evolution of piRNA clusters in the *Drosophila melanogaster* ovary. *bioRxiv*, 2023.

The pandas development team. pandas-dev/pandas: Pandas, Aug. 2025. URL `https://zenodo.org/doi/10.5281/zenodo.3509134`.

M. Tudor, M. Lobocka, M. Goodell, J. Pettitt, and K. O'Hare. The pogo transposable element family of drosophila melanogaster. *Molecular and General Genetics MGG*, 232(1):126–134, 1992.

E. L. van Dijk, D. Naquin, K. Gorrichon, Y. Jaszczyszyn, R. Ouazahrou, C. Thermes, and C. Hernandez. Genomics in the long-read sequencing era. *Trends in Genetics*, 39(9):649–671, 2023.

S. Venner, V. Miele, C. Terzian, C. Biémont, V. Daubin, C. Feschotte, and D. Pontier. Ecological networks to unravel the routes to horizontal transposon transfers. *PLOS Biology*, 15(2):e2001536, 2017. ISSN 1545-7885. doi: 10.1371/journal.pbio.2001536.

C. Vieira, D. Lepetit, S. Dumont, and C. Biémont. Wake up of transposable elements following drosophila simulans worldwide colonization. *Molecular biology and evolution*, 16:1251–5, 10 1999. doi: 10.1093/oxfordjournals.molbev.a026215.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, Mar. 2020. ISSN 1548-7105. doi: 10.1038/s41592-019-0686-2. URL `https://www.nature.com/articles/s41592-019-0686-2`.

G. L. Wallau, P. Capy, E. Loreto, and A. Hua-Van. Genomic landscape and evolutionary dynamics of mariner transposable elements within the drosophila genus. *BMC genomics*, 15(1):727, 2014.

L. Wang, S. Zhang, S. Hadjipanteli, L. Saiz, L. Nguyen, E. Silva, and E. Kelleher. P-element invasion fuels molecular adaptation in laboratory populations of *Drosophila melanogaster*. *Evolution*, 77(4):980–994, 2023.

M. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, Apr. 2021. ISSN 2475-9066. doi: 10.21105/joss.03021. URL `https://joss.theoj.org/papers/10.21105/joss.03021`.

H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer Nature, Basel, Switzerland, 2016. ISBN 978-3-319-24277-4.

H. Wickham, J. Hester, and J. Bryan. *readr: Read Rectangular Text Data*, 2024. URL `https://CRAN.R-project.org/package=readr`. R package version 2.1.5.

H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan. *dplyr: A Grammar of Data Manipulation*, 2025. URL `https://dplyr.tidyverse.org`. R package version 1.1.4.

M. Yoshiyama, Z. Tu, Y. Kainoh, H. Honda, T. Shono, and K. Kimura. Possible horizontal transfer of a transposable element from host to parasitoid. *Molecular biology and evolution*, 18(10):1952–1958, 2001.

G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36, 2017. ISSN 2041-210X. doi: 10.1111/2041-210X.12628. URL `https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.12628`.

A. Zeileis, R. Bivand, D. Eddelbuettel, K. Hornik, and N. Vialaneix. CRAN Task Views: The Next Generation, 2023. URL `https://arxiv.org/abs/2305.17573`.