Convergence and conflict among telomere specialized transposons across 60 million years of Drosophilid evolution

Jae Hak Son, Matthew A Lawlor, Mahek Virani, et al.

Genome Res. published online July 1, 2026
Access the most recent version at doi:10.1101/gr.281112.125

Convergence and conflict among telomere specialized transposons across 60 million years of Drosophilid evolution

Jae Hak Son[1,2], Matthew A. Lawlor[1,2], Mahek Virani[1], Weihuan Cao[1,2], Mia T. Levine[3,4], Christopher E. Ellison[1,2,5]

[1]Department of Genetics, Rutgers, the State University of New Jersey, Piscataway NJ, 08854, USA

[2]Human Genetics Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway NJ, 08854, USA

[3]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

[4]Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA

[5]To whom correspondence may be addressed. Email: chris.ellison@rutgers.edu

Abstract

The *Drosophila* telomere is one of the best-studied examples of active transposable elements (TEs) benefitting, rather than harming, a host genome. All *Drosophila* species lack telomerase and most species instead have telomeres composed of head-to-tail arrays of specialized retrotransposons. These TEs ostensibly act as mutualists by elongating chromosome ends, but evidence from species closely related to *Drosophila melanogaster* suggests that telomeric transposons may also antagonize their host genome. Importantly, the limited number of *Drosophila* species characterized thus far has precluded our ability to delineate idiosyncrasies from universal evolutionary forces and genetic mechanisms that shape the history of these TEs. Here, we have surveyed long-read genome assemblies of over 100 species of *Drosophila*, identifying a total of 396 telomeric TE families. Our findings show that these telomere-specialized elements evolve dynamically and also undergo striking convergent evolution: the complete loss of telomeric TEs has occurred repeatedly across the genus while individual telomeric TE lineages have repeatedly lost one of their two protein-coding genes. These elements have also repeatedly undergone horizontal transfer between distantly related *Drosophila* lineages and have repeatedly captured host gene fragments that promote their selfish suppression of host TE-silencing systems. Furthermore, telomere specialization itself appears to have evolved convergently, as some non-telomeric families have gained the ability to target their insertions to telomeres. These results provide unprecedented resolution into the evolution of these unusual TEs and highlight several novel mechanisms by which they evolve in conflict both with each other and their host genome despite the essential telomere function they provide.

Introduction

Transposable elements (TEs) selfishly replicate in the genomes of most species, often at the expense of their host. Although most new TE insertions are either neutral or deleterious, TE copies can also be co-opted by their host genome to serve a beneficial purpose, usually as a source of new genes or novel gene regulatory elements (Sundaram and Wysocka 2020; Fueyo et al. 2022; Almeida et al. 2022; Feschotte and Pritham 2007; Cosby et al. 2019). A large number of eukaryotic genes and regulatory sequences have been acquired from TEs. For the vast majority of these cases, the co-opted TE no longer actively mobilizes in the host genome and may evolve beyond recognition.

However, not all adaptive TEs are immobile. Work from a variety of species raises the possibility that active TEs can perform essential host functions, a process recently termed “TE addiction” (Shimada et al. 2024; Chang et al. 2025). This phenomenon refers to a period of host/TE cooperation that occurs during the transition of the TE from a selfish parasite to a fully co-opted, immobile element (Chang et al. 2025). Arguably the most well-characterized example of an active TE providing a benefit to its host genome involves the telomeric transposons of the *Drosophila* genus. All “true fly” Dipteran species lack telomerase activity (Mason et al. 2016) and the vast majority of *Drosophila* species have telomeres composed of head-to-tail arrays of specialized retrotransposons (Levis et al. 1993; Abad et al. 2004; Villasante et al. 2007). These TEs are derived from the Jockey clade of LINE retrotransposons and were first characterized in *D. melanogaster*, where three telomeric TE families have been identified: *HeT-A*, *TAHRE*, and *TART*, collectively known as *HTT* elements (Mason and Biessmann 1995; Frydrychova et al. 2008; Mason et al. 2008; Abad et al. 2004). The Jockey clade of LINEs generally contains two open reading frames (ORFs): a gag-like ORF1, characterized by major homology region (MHR) and zinc knuckle (CCHC) motifs, which play roles in nuclear localization and multimerization

(Fuller et al. 2010), and ORF2, which contains both endonuclease and reverse transcriptase domains (Casacuberta 2017).

These telomeric elements provide a clear benefit to their host: they prevent chromosome shortening due to the incomplete replication of chromosomal DNA during cell division (i.e. the end-replication problem (Olovnikov 1973)). Furthermore, the transcriptional silencing of these elements by the host genome may promote the assembly of a multi-protein complex, known as Terminin (analogous to Shelterin), that protects the chromosome ends from inappropriate double-strand break repair (Raffa et al. 2009; Cacchione et al. 2020). However, there is substantial evidence that these telomeric TEs are evolving in conflict with their host genome rather than existing exclusively as mutualists (Markova et al. 2020; Saint-Leandre and Levine 2020; Ellison et al. 2020). For example, genes required for telomere integrity tend to evolve rapidly under positive selection (Lee et al. 2017), a signature of genetic conflict. Furthermore, telomeric TEs themselves show dynamic patterns of evolution. Recent work found multiple cases of replacement of telomeric TE lineages among closely related *Drosophila* species. Notably, the same study found that telomeric TEs were completely lost in *D. biarmipes* (Saint-Leandre et al. 2019), suggesting that this ostensibly essential, mutualistic relationship between TE and host is not universally required for telomere maintenance. Furthermore, swapping into *D. melanogaster* an adaptively diverged version of a Terminin protein from a close relative resulted in a burst of transposition to telomere ends, suggesting that telomere binding proteins in *Drosophila* play a role in constraining telomeric TE activity, in addition to their end-capping function (Saint-Leandre et al. 2020; Perrini et al. 2004; Cui et al. 2021).

The piRNA pathway also constrains the activity of telomeric TEs (Khurana et al. 2010; Savitsky et al. 2006; Shpiz et al. 2007, 2011). Many TE transcripts, including those from telomeric TEs, are cleaved into sense and antisense piRNAs via the Argonaute proteins Ago3 and Aub

(Brennecke et al. 2007). These piRNAs then direct the formation of heterochromatin at TE loci in the genome via another Argonaute protein, Piwi, and a variety of accessory proteins including Panoramix and Nxf2 (Fabry et al. 2021; Batki et al. 2019; Murano et al. 2019; Zhao et al. 2019). Mutations in the piRNA pathway result in increased retrotransposition of telomeric transposons to chromosome ends (Savitsky et al. 2006). Conversely, increasing the efficiency of the piRNA pathway results in telomere shortening (Ryazansky et al. 2017). Thus, there is a tradeoff between genome defense and telomere elongation (Kalmykova and Sokolova 2023). However, the potential for conflict between these TEs and their host genome remains: unless they are actively constrained by the host genome, telomeric TEs hyperproliferate at chromosome ends, which has previously been shown to negatively affect host fitness (Walter et al. 2007).

Additional evidence of ongoing host-TE conflict comes from the *D. melanogaster TART* telomeric TE, which produces abundant sense and antisense piRNAs across its entire length (Ellison et al. 2020). The imprecise replication of a variety of TE families, including LINE elements, has been shown to result in the acquisition of host DNA sequence by the TE, a phenomenon known as transduction or gene capture (Catoni et al. 2019; Pickeral et al. 2000; Grabundzija et al. 2016). Through this process, *TART* has captured a portion of the host piRNA pathway gene *nxf2* (Ellison et al. 2020). A subset of the abundant antisense piRNAs produced from *TART* are able to target *nxf2* for suppression, which represents a form of host anti-silencing (Cosby et al. 2019). The host gene *nxf2* is evolving rapidly specifically in the region that was captured by *TART*, likely due to selection to escape targeting by *TART*-derived piRNAs (Ellison et al. 2020). While these results support an antagonistic relationship between telomeric TEs and their host genomes, other work raises the possibility that the rapid evolution of telomeric TEs is instead a consequence of their adaptation to the telomeric niche, which itself is inherently unstable (McGurk et al. 2021; Cacchione et al. 2025).

Are telomeric TEs in the process of being tamed by their host genome, as predicted by the "TE addiction" theory? Or, do they remain fundamentally selfish, despite the beneficial function they provide? If these elements are true mutualists, we would expect their evolutionary diversification to be tightly coupled to that of their host, resulting in a pattern of co-speciation, as has been observed for obligate endosymbionts of various insect species (Moran and Baumann 1994). Under this scenario, the previously described examples of gene capture and loss of telomeric TEs would represent rare edge cases where mutualism has reverted back to parasitism. On the other hand, a lack of co-speciation, repeated gene capture and loss of telomeric TEs would be more consistent with a parasitic relationship. These questions remain unresolved, in part because telomeric TEs have been described in only a tiny fraction of *Drosophila* species (16 (Saint-Leandre et al. 2019; Villasante et al. 2007) out of more than 1,600 species in the genus (O'Grady and DeSalle 2018)). Indeed, this limited picture of telomeric TE diversity precludes our ability to delineate universal from idiosyncratic evolutionary forces and genetic mechanisms that shape the history of these seemingly beneficial but active transposable elements. Here, we manually curate 396 telomere-specialized TE families from over 100 species of *Drosophila* and perform phylogenetic and sequence analysis of these transposons to determine whether telomere-specialized TEs evolve antagonistically with their hosts, despite the essential function they serve.

Results

Identification of telomere-specialized retrotransposons

We searched for non-LTR retrotransposons from the Jockey superfamily in long-read genome assemblies from a total of 106 *Drosophila* species and three outgroup species of Drosophilidae (**Table S1**). We used an automated pipeline to detect ORFs with homology to known *Jockey*

superfamily ORF1 and ORF2 peptides in each assembly, followed by a phylogenetic approach to assign ORFs to telomeric versus non-telomeric clades (see Methods, **Figure S1**). We then used manual curation to generate full-length consensus sequences for each TE family and to confirm that candidate telomeric TEs form head-to-tail arrays at contig ends, consistent with telomere-specialization (see Methods, **Figure S1**, **Figure S2**).

We identified a total of 396 putatively telomere-specialized TE families across 109 species (**Table S2**). Of the 396 telomeric TE families, we detected 188 families in head-to-tail arrays at the extreme ends/termini of gene-rich, megabase-long contigs in their host species' genome assembly. 189 TE families were found in multiple head-to-tail copies but only on short contigs lacking genes that would allow their assignment to chromosome arms. The remaining 19 TE families were present in a single copy in their host species genome assembly but were supported as telomeric based only on their position in our ORF1 and/or ORF2 gene trees (**Figure S3**). We were unable to identify Jockey telomeric TEs in a total of 15 species, including the previously reported case in *D. biarmipes* (Saint-Leandre et al. 2019). In three of these species (*D. kurseongensis*, *D. orena*, and an undescribed species from the *funebris* group, labeled as *D. sp.St01m* in (Kim et al. 2021)), we identified fragments of Jockey telomeric TEs lacking intact ORFs, but no full-length TEs (**Table S3**). To confirm the absence of active Jockey telomeric TEs in all 15 species, we analyzed the raw genome sequencing reads and recovered only the same TE fragments present in the genome assemblies of the three species above, confirming the absence of Jockey telomeric TEs in these species. To determine whether other non-Jockey TEs could have become telomere specialized in these species, we manually searched telomeric (based on synteny) scaffold ends in these genome assemblies but were unable to find any TE arrays. Thus, in total, we infer that telomere-specialized TEs were independently lost at least 10 times across the genus (**Figure 1**).

In the remaining species, we identified ~4 telomere-specialized families per species, on average (**Figure S4**). On average, full-length elements were 8.00 kb in size and present in approximately 11 copies per genome, based on read depth (see Methods)(**Figure S4**). The vast majority of species with telomeric TEs have at least one telomeric TE family that carries both ORF1 and ORF2; however, we identified full-length TEs carrying only ORF1 in the genomes of *Z. ghesquierei* and *D. rufa* but only fragments of telomeric TEs with ORF2. Overall, ~60% of telomeric TE families contained both ORF1 and ORF2, while ~40% contained ORF1 only, having lost ORF2 (**Figure S4**). We also found 18 TE families where ORF2 was present but ORF1 was missing, a structure that has not been previously reported. In five of these families, ORF1 was interrupted by a premature stop codon while the remainder completely lacked ORF1 sequence (**Table S2**). These TEs have significantly reduced copy numbers compared to TE families with either ORF1 only or both ORF1 and ORF2 (Wilcoxon test $P = 0.0004$ and $P = 0.0022$, respectively)(**Figure S4**). We therefore conclude that these likely represent older fragments of previously active TEs.

To assess the accuracy of our classification of Jockey elements as telomeric versus non-telomeric, we focused on a subset of 30 *Drosophila* species with chromosome-level genome assemblies. We identified 363 insertions of telomeric clade Jockey TEs across these species, 337 (93%) of which were located within 500 kb of the telomeric end of a chromosome-length scaffold (see Supplemental Materials for discussion of telomeric Jockey elements with locations outside the telomere). Insertions of TE families belonging to the non-telomeric clade of Jockey elements showed the opposite pattern: Of the 57,760 insertions of non-telomeric Jockey elements across all 30 species, only 321 (0.5%) were located within 500 kb of the telomeric scaffold end (**Figure S5**). The non-telomeric Jockey families are instead highly enriched in pericentromeric heterochromatin (**Figure S5**). These results strongly suggest that our phylogenetic classification of telomeric versus non-telomeric Jockey elements is accurate.

Taken together, our results show that the genomes of the vast majority of *Drosophila* species contain multiple families of telomere-specialized retrotransposons and that the dependence of ORF1-only telomeric TEs on ORF2 proteins encoded by other telomeric TE families is widespread across the genus. The two species for which we were unable to identify any intact telomeric TEs with ORF2 (in either the genome assembly or the raw reads) raise the possibility that some telomeric TEs could be using ORF2 from other non-telomeric Jockey clade TEs for replication.

Evolutionary Diversification of telomere-specialized retrotransposons

We used ORF1 and ORF2 peptide sequences to create phylogenetic trees showing the evolutionary relationships among all 396 telomeric TE families. We identify 6 major clades of telomeric retrotransposons across 109 *Drosophila* species. These clades are present in both our ORF1 and ORF2 trees and include all original telomeric TE clades described by Villasante *et al* : *TR1*, *TR2*, *TR3*, *TR4*, and *HTT*, though we note that our more comprehensive trees suggest the *TR4* clade is actually part of the *HTT* clade (**Figure 2A, Figure S6**) (Villasante et al. 2007). We therefore use *HTT/TR4* here to encompass both groups (**Figure 2A, Figure S6,** see Methods). Our results confirm the monophyly of the Villasante *et al* clades and also provide evidence of two novel clades, which we name *TR5* and *TR6*. Many species contain diverse assemblages of telomeric TEs: 94 species have telomeric TEs from more than one *TR* clade, with 18 species harboring three or more *TR* clades (**Figure 2A, Table S2**).

We used our ORF1 tree to investigate the evolutionary relationships among the abundant ORF1-only TEs found across the *Drosophila* genus. The location of these TEs within our ORF1 tree suggests that ORF2 has been repeatedly lost across dozens of independent TE lineages (**Figure S7**). However, we also find monophyletic clades containing between two and six ORF1-

only TE families, which suggests that not all ORF1-only TEs arise from independent loss of ORF2 (**Figure S7**). Instead, ancestral ORF1-only TE lineages can birth new ORF1-only TE lineages. The repeated loss of ORF2 from telomeric TEs is particularly striking given that this phenomenon does not seem to occur in non-telomeric Jockey elements (Arkhipova 2012).

We next sought to investigate the evolutionary history of these telomeric TEs across the *Drosophila* genus. In a mutualistic relationship, where the symbiont is transmitted vertically from parent to offspring, its evolution will be tightly coupled to that of its host, resulting in a pattern of co-speciation. Such a pattern has been observed in a variety of obligate endosymbionts (Moran and Baumann 1994). Thus, if telomeric TEs are acting as mutualists, the major telomeric TE clades should mirror the major clades found in the *Drosophila* species tree. We would also expect that the most ancient TE clade should be found broadly, across all extant *Drosophila* species, since its origin would coincide with or predate the most recent common ancestor of the genus. Instead, we find that many distantly related species harbor closely related telomeric TEs. For example, *TR1*, *TR2*, and *TR3* clade elements are found in species from both the *Sophophora* and *Drosophila* subgenera. Furthermore, the early diverging TE clade, *TR1*, is found in only 6 of 10 major species clades, while *TR2*, one of the more derived clades, is found across all 10 species clades (**Figure 2**).

There are two evolutionary scenarios that could explain these observations: (1) the divergence of the majority of telomeric TE clades predated the common ancestor of the *Drosophila* genus and different TE clades were subsequently lost from different *Drosophila* lineages or (2) divergence of the TE clades accompanied the divergence of the genus, followed by frequent horizontal transfer of TE clades among *Drosophila* species groups. To assess the evidence for these two scenarios, we performed gene tree/species tree reconciliation (see Methods) to infer duplication (in the case of TEs, this would be divergence of an ancestral TE family into two

related families within a single host lineage), transfer, and loss events for the telomeric TEs. To more easily visualize the reconciliation results, we focused on 9 major species clades (Suvorov et al. 2022) and pruned the species and TE trees (see **Figure S6** for pruned TE tree) to include the minimum number of species necessary to account for the presence of each *TR* clade within each species clade (**Figure 2B**). We found that the inferred rate of horizontal transfer (HT) exceeded the inferred rates for both duplications and losses (HT: 22, duplications: 6, losses: 13). We confirmed that these putatively horizontally transferred TEs showed patterns of sequence similarity consistent with horizontal transfer (**Figure S8, Figure S9**) and repeated this analysis on the unpruned species and gene trees, which infers a total of 82 HT events, 12 duplication events, and 70 losses. For comparison, we ran the same analysis pipeline on a set of 465 non-telomeric Jockey TEs recently identified in *Drosophila* (Tambones et al. 2019). The rate of horizontal transfer for telomeric TEs is approximately 63% higher than that of non-telomeric TEs, a difference that is highly significant (Wilcoxon test $P < 2.2 \times 10^{-16}$, **Figure 2C**). Thus, our reconciliation analysis suggests that the six major telomeric TE clades originated within different *Drosophila* lineages at different timepoints during the diversification of the genus and subsequently expanded their host range via horizontal transfer.

Convergent evolution of telomere-specialization

Telomeres can act as a “safe harbor” where TEs can insert without deleterious consequences. Indeed, transposon insertions have been identified within the canonical telomere repeats in species of fungi, silk moths, and plants (Fujiwara et al. 2005; Higashiyama et al. 1997; Rahnama et al. 2020; Gladyshev and Arkhipova 2007). One might therefore expect other transposons to target their insertions to *Drosophila* telomeres. To assess this possibility, we asked whether there were Jockey family TEs from the non-telomeric clade that form head to tail arrays at the telomeric regions of any species included in our analysis. We identified Jockey retrotransposons from the non-telomeric clade at the telomeres of six species: *D. cardini*, *D.*

funebris, *D. littoralis*, *D. virilis*, *Scaptomyza hsui*, and *Scaptomyza pallida* (**Figure 3**). These TEs form a monophyletic clade within the larger clade of non-telomeric Jockey elements, despite their appearance in species that are only distantly related. For example, *D. cardini*, *D. funebris*, and *S. hsui/pallida* diverged from each other more than 20 mya (Kumar et al. 2022). The presence of these closely related TEs in distantly related species is indicative of horizontal transfer (**Figure 3, Figure S10**). Hereafter, we refer to this as the NTT clade (Non-telomeric clade TEs at Telomeres).

These elements are not found elsewhere in the genomes of any of the six species listed above, consistent with telomere specialization; however, in all six species, their arrays are found in between and/or adjacent to arrays of telomeric-clade *Jockey* TEs (**Figure 3C**, **Figure S10**). Their genomic location suggests that these elements have acquired the ability to target their insertions to the telomeric regions of the genome. Nevertheless, it remains unclear whether they are able to act as *bona fide* telomeres (i.e. by mobilizing directly to chromosome ends) or if they are instead parasitizing the telomeric niche by homing in on pre-existing telomeric TE arrays.

It is also conceivable that other non-Jockey TEs have evolved telomere specialization. To assess this possibility, we calculated the median distance to the telomere for all TE families from 30 *Drosophila* species with chromosome-level genome assemblies. With the exception of the above NTT elements and telomeric Jockey TEs, no other TE families showed a telomere-biased distribution in their genomic locations, suggesting that NTT and Jockey elements are the only telomere specialized TEs across these 30 species (**Figure S11**).

Potential neofunctionalization of ORF2

The ORF2 of all LINE elements, including mammalian LINE-1, contains, at a minimum, an endonuclease and a reverse transcriptase (RT) domain (Eickbush and Malik 2002). The domain

organization of ORF2 within the Jockey clade of LINEs, which includes all *Drosophila* telomeric TEs, is similar to that of LINE-1 elements, with the endonuclease domain preceding the RT domain (Eickbush and Malik 2002; Wicker et al. 2007). The presence of a conserved endonuclease domain within the ORF2 carried by *Drosophila* telomeric TEs is surprising, considering that transposition to chromosome ends should not require DNA nicking (Casacuberta 2017). In order to better characterize the extent of ORF2 domain conservation across *Drosophila* telomeric TEs, we used InterPro to identify conserved domains in the ORF2 sequences from our library of telomeric TEs. Seven and nine ORF2 proteins were found to be missing the RT and endonuclease domains, respectively (**Table S4**). Two of these cases were due to a premature stop codon truncating the ORF2 reading frame whereas the seven remaining ORF2 proteins showed no signs of truncation but were missing both RT and endonuclease domains. All seven ORF2 proteins are from species in the *virilis* group, and, according to InterPro, these unusual ORFs contain no conserved domains at all, aside from a single disordered region. The first ORF in these TEs is clearly homologous to the ORF1 protein found in other *Drosophilia* telomeric TEs, which is how these elements were classified as telomeric TEs in the first place. However, a peptide BLAST search was unable to detect any significant amino acid similarity between these unusual ORF2 peptides and the canonical ORF2 peptides from other telomeric TEs (**Figure 3D**).

In total, we identified seven telomeric TE families carrying this unusual ORF from the following species: *D. virilis*, *D. americana*, *D. novamexicana*, and *D. littoralis*. The peptides encoded by these ORFs range in size from 793 to 1219 amino acids and they range from 30% to 96% identical among the seven TE families (**Figure 3D, Figure S12**). We were unable to identify homologs of these peptides outside of the above species, despite employing sensitive HMM-based search strategies (see Methods). We calculated the ratio of non-synonymous (*dN)* to synonymous (*dS)* substitutions for all pairs of sequences and found that, in 20 out of 21 pairwise

comparisons, there is statistically significant evidence of purifying selection, which suggests that the ORFs encode functional proteins (**Table S5**). The pair that did not reach significance are highly similar to one another resulting in a lack of power to detect purifying selection (**Table S5**).

In addition to the conserved endonuclease and RT domains, the ORF2 proteins of some *Drosophila* telomeric TEs carry an additional ~400 amino acid, glutamine rich region at the C terminus, referred to as the "X domain" (Casacuberta and Pardue 2003). We hypothesized that perhaps the unusual ORF that we discovered was derived from this X domain, having lost both the endonuclease and RT domains. To test this hypothesis, we used a Hidden Markov Model (HMM) approach (see Methods) to compare the 7 unusual ORF2 peptides identified here to all telomeric TE ORF2 peptides containing the endonuclease, RT, and X domains (hereafter referred to as EN/RT/X ORF2 peptides). These searches revealed several regions of homology between a *D. littoralis* EN/RT/X ORF2 peptide and our HMM (**Figure 3E**). Small portions of both the endonuclease and RT domains in the *D. littoralis* EN/RT/X peptide show weak evidence for homology with our HMM (HMMER E-value = 1.8×10^{-4} and 0.2, respectively). There is much stronger evidence for homology between the C terminus of the HMM and the X domain of this peptide (HMMER E-value = 2.2×10^{-14}). These results suggest that the unusual ORF2 peptides we discovered are in fact derived from the canonical ORF2. However, the endonuclease and RT domain regions have evolved so rapidly that they are no longer recognizable as conserved domains and may no longer retain their ancestral function, raising the possibility that this ORF has acquired a novel function in these species. This represents, to our knowledge, the only known example of a LINE ORF2 peptide that lacks both RT and EN domains.

Gene capture by telomeric transposons

In plants, gene capture may allow TEs to evade host silencing by forcing the host to reduce the efficiency of TE suppression mechanisms in order to avoid self-silencing (Lisch 2009; Muyle et

al. 2021). Similarly, our prior work showed that the *D. melanogaster* telomeric *TART* element has captured a fragment of the host piRNA pathway gene *nxf2* and that *TART*-derived piRNAs likely target *nxf2* for suppression (Ellison et al. 2020). This anti-silencing strategy imposes a cost to the host genome while benefitting the TE, consistent with *TART* acting as a parasite rather than a mutualist. We therefore sought to determine if gene capture occurred in the telomeric transposons of other *Drosophila* species by searching the full length consensus of all *Drosophila* telomeric TEs identified here for homology to *D. melanogaster* peptides. In *D. tristis*, we identified multiple fragments of the host gene *pangolin* embedded within the telomeric TE array of the dot chromosome, however upon further inspection, we concluded that the presence of *pangolin* coding sequence within the telomere of this species is likely due to unequal crossing over rather than gene capture by telomeric TEs (**Figure S13**). After manual curation of all remaining BLAST hits (see Methods), we identified 20 telomeric TE families that show evidence of gene capture (**Table S6**), which represents at least 9 independent gene capture events across the genus, including the previously described capture of *nxf2* (**Figure 4A**). Eight of the nine events involve capture of known piRNA pathway genes: the capture of *nxf2* by *D. melanogaster* plus seven additional capture events involving either *piwi* or *aubergine* (*aub*), both of which are members of the Piwi subfamily of Argonaute proteins that bind piRNAs and are required for transposon silencing (Vagin et al. 2006; Gunawardane et al. 2007; Saito et al. 2006; Brennecke et al. 2007).

We find that a fragment of *piwi* was captured by a telomeric TE family present in both *D. auraria* and *D. triauraria* and fragments of *aub* were captured by a telomeric TE family in *D. sucinea* and nine different species within the *vittiger* subgroup of the *Zaprionus* clade. The other novel capture event involves another telomeric TE family from *D. auraria* and *D. triauraria,* which carries fragments from a homolog of the F-box domain containing *D. melanogaster* gene *CG12520*. F-box proteins are known to provide substrate specificity to the SCF ubiquitin ligase

364 complex (Kipreos and Pagano 2000). *CG12520* is expressed in the female germline (Tiwari et
365 al. 2019) and knockdown of *CG12520* causes a very modest upregulation of two telomeric TEs
366 (Czech et al. 2013). This gene is largely uncharacterized; for example, its binding substrate
367 remains unknown (Dui et al. 2012). Across all gene capture TE families, the length of the
368 captured gene fragments ranges from 140bp to 235bp and their similarity to their host gene
369 ranges from a percent identity of 76.4% to 94.82% (**Figure 4B**, **Table S6**).
370
371 Five of the nine independent gene capture events occurred in the *Zaprionus* clade. Out of 17
372 *Zaprionus* species with sequenced genomes, we find ten species harboring one or more
373 telomeric TEs that contain *aub* gene fragments (**Figure 5A**). In total, four different regions of the
374 *aub* gene were captured by different TE lineages within this group (**Figure 5A**). Some of these
375 telomeric TEs, such as those from *Z. africanus*, *Z. gabonicus*, and *Z. indianus*, captured two
376 different regions of the *aub* gene (**Figure 5A**). The captured *aub* fragments were subsequently
377 amplified within these TEs, up to 7 copies in the *Z. africanus* TE and 8 copies in *Z. gabonicus*
378 and *Z. indianus* TEs (**Figure 5A**).
379
380 It is possible that these seemingly independent gene capture events actually originated via a
381 single capture event that included a large portion of the *aub* gene, followed by independent
382 deletions of *aub* sequence in different TE lineages. Indeed, the *Zaprionus* TE families that carry
383 *aub* sequence form a monophyletic clade in the ORF2 TE tree (**Figure S14**). We tested this
384 prediction by reconstructing the evolutionary history of each capture event (**Figure S14**). We
385 created gene trees using alignments containing the captured *aub*-like TE sequences along with
386 their homologous sequences from the *aub* gene of each *Zaprionus* species. These trees
387 confirmed our initial inference that five independent capture events occurred within *Zaprionus*
388 and allowed us to assign each event to a node in the *Zaprionus* species tree (**Figure 5A**).
389 Unexpectedly, we find that region B of *aub* was captured in the ancestor of *Z. vittiger* and *Z.*

capensis and subsequently lost in the *Z. davidi* / *Z.taronus* / *Z.capensis* clade, whose TEs now carry region C of *aub* (**Figure 5A, Figure S14**). We also note that region B of *aub* was captured independently by two different *Zaprionus* clades; however, the captured regions overlap but do not share the same boundaries.

The repeated capture of piRNA pathway genes by *Drosophila* telomeric TEs suggests that acquisition of these gene fragments may increase TE fitness. We previously found evidence that antisense piRNAs produced from the *nxf2*-like region of *TART* are capable of targeting the *nxf2* host gene for suppression, consistent with an anti-silencing strategy to escape host suppression (Ellison et al. 2020). To determine if a similar scenario is occurring in these other species, we performed small RNA sequencing from ovaries for five species where telomeric TEs had captured either *aub* or *piwi* as well as 4 related species where gene capture had not occurred to use as controls**.** The median identity between captured sequence and host gene is ~90% across all gene capture TEs (**Table S6**). This identity is noticeably smaller for two gene capture cases: *CG12520* in *D. auraria*/*triauraria* (~79% identity) and *aub* in *Z. sp. sepsoides* (~82% identity)(**Table S6**). We find almost a complete lack of piRNAs mapping to the host genes in these latter two cases of elevated divergence (**Figure S15**), suggesting that the similarity between captured sequence and host gene is too low to direct host gene targeting by TE-derived piRNAs.

In the species with high sequence identity between the captured sequence and the host gene, we found small RNAs from both the gene capture telomeric TEs and the *aub* (or *piwi*) host genes whose lengths and 5’ U bias are consistent with piRNAs (**Figures S16, S17, S18**). Furthermore, in species where *aub* (but not *piwi*) was captured, the abundance of piRNAs from *aub* is significantly larger than from *piwi* while the reverse is true for species where *piwi* (but not *aub*) was captured (**Figure 5B**). Overall, piRNA abundance from the captured host gene is

significantly larger in the species harboring gene capture TEs compared to species whose TEs have not captured any host gene (Wilcoxon test P = 0.0036, **Figure S19**), consistent with the host gene being targeted by piRNAs derived from the gene capture region of the telomeric TE. Across all species harboring gene capture TEs, the telomeric TE families that have captured a host gene fragment have significantly higher copy numbers compared to the other non-capture telomeric TEs in the same genome (paired Wilcoxon test, one-sided *P* = 0.02, **Figure S20**), suggesting that gene capture increases TE fitness beyond its role in suppressing the host piRNA pathway (see Discussion).

There are two ways that piRNA biogenesis from the *piwi*/*aub* host genes could be initiated: (1) via ping-pong amplification or (2) via a "trigger" piRNA, as has been described for the *D. melanogaster row* gene, where phased piRNA production is triggered by a piRNA from the *1360* TE that shares sequence similarity with *row* (Mohn et al. 2015). PiRNA biogenesis from *D. melanogaster nxf2* mRNA is more consistent with the latter phenomenon where a trigger piRNA from the *TART* telomeric TE induces phased piRNA biogenesis downstream of the region of *nxf2* that was captured by *TART* (Ellison et al. 2020).

To better assess ping-ping and phasing signatures in *piwi* and *aub* derived piRNAs, we performed additional deep sequencing of piRNAs (see Methods) from four species: *D. auraria*, *D. triauraria*, *Z. gabonicus*, and *Z. indianus*. For both *piwi* and *aub*, we observe sense-strand piRNA production both upstream and downstream of the region that was captured by a telomeric TE, however only the downstream piRNAs show a strong signature of phasing (**Figure S18**). In all four species, antisense piRNAs derived from telomeric TEs align to the captured region of the host gene and overlap with host gene derived sense piRNAs (**Figure S17**). In three species (*D. auraria*, *D. triauraria* and *Z. gabonicus*), the overlap between these sense and antisense piRNAs show the signature of ping-pong amplification (**Figure S17**).

These results are consistent with ping-pong amplification and/or trigger piRNAs resulting in the biogenesis of phased piRNAs downstream from the captured region of the host genes *aub* and *piwi*.

If *aub* is being targeted for suppression by TE-derived piRNAs, we expect to see accelerated evolution in the corresponding region of *aub*, which would reduce piRNA targeting by decreasing the similarity between *aub* and the *aub*-like sequences within the telomeric TE. We focused on the clade of *aub* with sequences from the following species: *Z. nigranus*, *taronus*, *davidi*, and *capensis*, since the telomeric TEs in all of these species have captured the same single region of *aub*, and the captured sequences have not been amplified within the TE (**Figure 5A**). We find that the region of *aub* that was captured by TEs is evolving approximately twice as fast as the uncaptured portion of the gene, based on a comparison of trees constructed from the captured region of *aub* versus *aub* sequence that was not captured by any *Zaprionus* TEs (Fisher's Exact Test $P = 2.7 \times 10^{-5}$, see Methods). As a control, we repeated this analysis on the *aub* clade from *Z. camerounensis*, *lachaisei*, and *vittiger*, whose TEs have not captured this region of *aub*. In this case, there was no difference in the rate of evolution (Fisher's Exact Test $P = 0.64$, **Figure 5D**).

We also find that *aub* has undergone two independent duplication events in *Zaprionus*, once in the *Z. capensis* lineage and once in *Z. africanus* (**Figure 5C**). This finding alone is remarkable given that a previous survey of 39 Dipteran species (including 22 species of *Drosophila*) identified only a single duplication event involving *aub* since its origin over 150 million years ago, which occurred in stalk-eyed flies (Lewis et al. 2016). Notably, both *Zaprionus aub* duplications occurred in species with telomeric TEs that have captured an *aub* fragment (**Figure 5A**), raising the possibility that the duplications were selected for as a means to increase *aub* dosage in response to it being targeted by TE-derived piRNAs.

In addition to duplications, we also find strong support for introgression of *aub* from *Z. taronus* to *Z. nigranus*, likely due to recent gene flow between these two species (**Figure 5C, Figure S21, Figure S22**). This transfer of *aub* from *taronus* to *nigranus* potentially set the stage for a subsequent invasion of the *nigranus* genome by *taronus* telomeric TEs, whose *aub*-like sequences are 92.26% identical to those of the *taronus aub* gene. Indeed, two telomeric TEs, both of which carry an *aub* gene fragment, show evidence of horizontal transfer from *Z. taronus* to *Z. nigranus* (**Figure S21)**. Based on sequence similarity, we infer that the two telomeric TEs initially diverged in *Z. taronus*, with one family subsequently losing the ORF2 gene (**Figure S21)**. Synonymous divergence between *taronus* and *nigranus* copies of the *aub* gene (K_S=0.034) is larger than that of the ORF1 genes for the two TE families (TE-1 K_S=0, TE-2 K_S=0.023), suggesting that the transfer of *aub* from *taronus* to *nigranus* occurred first, followed by the transfer of TE-2 and, most recently, the transfer of TE-1 (**Figure S21**). Notably, these recently transferred TEs (which captured region C of *aub*, **Figure 5A**) have replaced the ancestral telomeric TEs in *Z. nigranus*, which carried region B of *aub*. Thus, different lineages of gene capture TEs may compete with one another to maintain their residence in a given host genome.

Together, these results suggest that capture of piRNA pathway genes by telomeric TEs has occurred repeatedly across the *Drosophila* genus, likely as a form of counter-defense, where TEs selfishly target host suppression systems for silencing to increase their own fitness. The accelerated evolution of the region of the host gene that was captured, which was previously described for *nxf2* (Ellison et al. 2020) and shown here for *aub*, suggests there is selection to avoid this targeting by “erasing” the similarity between the captured sequence carried by the TE and the host gene.

Discussion

Recent work in both *Arabidopsis* and Zebrafish has introduced the concept of transposon "addiction" where an active TE family serves an essential function for its host genome despite causing mutational damage via insertion of new TE copies (Shimada et al. 2024; Chang et al. 2025). The telomeric TEs of *Drosophila* fit nicely into this framework: they play a critical role in protecting and extending chromosome ends and, importantly, the benefit they provide to their host *depends* upon their transposition activity. However, the host genome also constrains the activity of these elements, both to ensure that transposition occurs only on chromosome ends (Cui et al. 2021) and to control telomere length as ultralong telomeres have been shown to reduce fertility (Walter et al. 2007). The TE addiction model predicts that telomeric TEs should eventually be fully co-opted by their host genome; however, our results suggest that the tension between selfish mobilization and host control has led to a prolonged period of evolution between these TEs and their host genome that is dominated both by instability and genetic conflict.

Rather than the co-speciation that one might expect from a true genetic mutualist, the telomeric TE phylogeny is dominated by horizontal transfer and lineage-specific extinction events, consistent with a parasitic relationship between these elements and their hosts. This relationship is similar to what has been described for facultative bacterial endosymbionts like *Wolbachia*, *Rickettsiella*, and *Spiroplasma*, all of which have been shown to undergo frequent horizontal transfer among host species while also selfishly enhancing their own fitness at the expense of their host through strategies such as cytoplasmic incompatibility and male killing (Jonathan et al. 2024; Gu et al. 2023; Hoffmann and Cooper 2024; Floriano et al. 2025). These facultative endosymbionts are distinct from obligate mutualists, like *Buchnera*, which are vertically transmitted and provide essential nutrients to their aphid hosts, with whom they have codiverged for over 200 million years (Moran et al. 1993; Moran and Baumann 1994).

We further find that the relationship between telomeric TEs and their host is unstable: it has completely unraveled at least 10 separate times across the genus, with a total of 13 *Drosophila* species (and two outgroup species) having lost telomeric TEs entirely, and thus presumably relying on an alternative, recombination-based mechanism of chromosome elongation (*i.e.* Alternative Lengthening of Telomeres (ALT) (Bryan et al. 1997; Biessmann et al. 2000)). This instability may be due in part to the dynamic nature of the telomeric niche (McGurk et al. 2021); however, we also find evidence of competition among telomeric TE clades. Both the TR1 and TR6 clades predate the origin of the TR2 clade, but TR2 elements are found in more than half of all species studied here, which is ~2-fold as many species compared to TR1 and ~10-fold the number of species carrying TR6 clade TEs. TR2 has therefore spread via horizontal transfer into lineages whose ancestral telomeres were occupied by TR1 and/or TR6 clade TEs but have since been replaced by TR2 elements.

We also find evidence of potential competition among TEs from non-telomeric *Jockey* clades for the telomeric niche. The non-telomeric clade TEs at telomeres (NTTs) we identify here are well supported as members of the non-telomeric *Jockey* clade, yet they appear to be present in head-to-tail arrays exclusively at the telomeres of six *Drosophila* species, always either adjacent to, or in between, arrays of telomeric *Jockey* clade TEs. This localization pattern suggests that these elements are able to target their insertions to the telomeric regions of chromosomes. Future work will determine whether they are able to directly mobilize to chromosome ends or if they instead only insert into pre-existing arrays of telomeric clade TEs.

In contrast to what would be expected from a genetic mutualist, we find evidence of extremely rapid evolution and potential functional innovation within a subset of telomeric TE families from several *virilis* group species. The ORF2 amino acid sequence of these TE families has evolved

so rapidly that it is almost unrecognizable as an ORF2 homolog, which is especially notable given that ORF2 (as opposed to ORF1) is preferentially used for phylogenetic classification of LINEs due to its slower rate of evolution. The endonuclease and reverse transcriptase conserved domains are hallmarks of ORF2 encoded peptides in LINE non-LTR retrotransposons from humans to nematodes (Eickbush and Malik 2002). Neither of these domains are detectable by InterPro (Paysan-Lafosse et al. 2023) in the amino acid sequences of these unusual ORF2s. It is possible that the loss of these conserved domains is caused by degeneration or pseudogenization. However, our d_N/d_S analysis shows that these unusual ORF2 sequences are evolving under purifying selection. Thus, these peptides may have evolved a novel function that is different from the host DNA cleavage and reverse transcription functions provided by canonical ORF2s, which is now being maintained by purifying selection. How could these TEs continue to mobilize after loss of these two critical ORF2 functions? It is notable that the entire ORF2 is dispensable from a given TE as long as it is encoded by other related TE families, as evidenced by the repeated loss of ORF2 from many telomeric TEs. Such dispensability may have provided this unusual ORF2 the flexibility to evolve a novel function similar to the phenomenon of neofunctionalization after gene duplication (Ohno 1970).

Frequent horizontal transfer, competition among telomeric TE lineages, and functional innovation within certain lineages are all consistent with antagonistic evolution between telomeric TEs and their hosts. The possibility of antagonistic coevolution is further supported by our evidence that telomeric TEs have repeatedly and specifically captured fragments of piRNA pathway genes, including *aubergine* and *piwi*, in addition to the previously described capture of *nxf2* by the *D. melanogaster* telomeric transposon *TART* (Ellison et al. 2020). PiRNA abundance from a subset of these species is consistent with our previous work, suggesting that piRNAs derived from the captured gene fragment(s) embedded within the telomeric TEs are capable of targeting the host gene for suppression (Ellison et al. 2020). This phenomenon thus

likely represents a counter-defense or anti-silencing strategy deployed by these TEs (Cosby et al. 2019; Sasaki et al. 2022; Hosaka et al. 2017; Sasaki et al. 2023; Fu et al. 2013; Lawlor and Ellison 2023). It may seem self-destructive for a single TE family to impair the piRNA pathway, thus potentially leading to the upregulation of the majority of active TEs in the genome. However, the telomeric transposons of *D. melanogaster* are exquisitely sensitive to disruption of the piRNA pathway (Czech et al. 2013; Kalmykova 2023; Morgunova et al. 2021). Assuming telomeric TEs in other species are similarly sensitive, they may be able to suppress piRNA activity just enough to benefit themselves without causing the upregulation of other TE families, who are less sensitive to piRNA pathway disruption.

Similar to what we observed for *nxf2* in *D. melanogaster*, we find accelerated evolution of the *aub* gene in *Zaprionus*, specifically in species harboring *aub*-capture TEs and specifically in the region of *aub* that was captured. This process is consistent with antagonistic coevolution: the telomeric TE selfishly captures a fragment of the host gene, which likely decreases expression of the host gene via piRNA-mediated silencing (though we note that this targeting has not been experimentally verified). Host gene mutations that disrupt piRNA targeting would then be favored by natural selection, leading to accelerated evolution of the host gene in the region that was captured. The TE may respond by capturing a new region of the same gene, as we see in *Zaprionus*. Amplification of the captured gene fragment (as seen in *Z. indianus*, *Z. africanus*, and *Z. gabonicus*) may also be beneficial by allowing the TE to more efficiently acquire mutations that increase or preserve the similarity between the captured fragment and host gene via non-allelic gene conversion among the duplicated fragments (Mano and Innan 2008; Ellison and Bachtrog 2015).

We previously observed that all copies of *TART* carry the captured fragment of *nxf2*, both in the *D. melanogaster* reference genome, as well as in other wild strains (Ellison et al. 2020).

However, *nxf2*-positive *TART* copies impose a fitness cost on their host genome, which in turn reduces their own fitness because, unlike viruses, their survival depends on vertical transmission from parent to offspring (Cosby et al., 2019). If the sole benefit of the captured *nxf2* fragment is derived from its ability to suppress the piRNA pathway, we would instead expect this fragment to be polymorphic among *TART* copies because *nxf2*-negative *TART* copies would benefit from the presence of other *nxf2*-positive elements without the associated fitness cost. We therefore previously proposed that capturing host gene fragments may increase TE fitness in other ways, potentially by enhancing transposition (Ellison et al. 2020). Our results here support this prediction: we find that telomeric TEs that have captured host gene fragments have significantly higher copy numbers compared to other non-capture telomeric TEs in the same genome (**Figure S20**) which suggests there is another benefit provided by the captured gene fragment in addition to anti-silencing.

The well-ordered arrays of telomeric TEs found at *Drosophila* chromosome ends are the result of a complex interplay between TE families that both cooperate and compete with each other as well as the host piRNA (Khurana et al. 2010; Savitsky et al. 2006; Shpiz et al. 2007, 2011), DNA repair (Cenci et al. 2005; Melnikova et al. 2005), chromosome end-capping (Raffa et al. 2011), and heterochromatin maintenance (Savitsky et al. 2002; Perrini et al. 2004) pathways, each of which is necessary for proper telomere function. Telomeric TE lineages are likely to have co-evolved to at least some degree with the host proteins involved in these pathways to ensure the correct formation and maintenance of telomeric arrays. It is therefore remarkable that host proteins from all three of these pathways are known to evolve rapidly (Lee et al. 2017; Parhad and Theurkauf 2019; Lin et al. 2024). Previous work has proposed various types of genetic conflict that could explain this observation (Lee et al. 2017). Our work here adds two new possibilities: (1) the frequent horizontal transfer of telomeric TEs among species, as well as the invasion of the telomeric niche by NTTs, could require rapid host protein evolution to maintain

coordination between TE mobility and telomere maintenance and (2) repeated gene capture by telomeric TE families could antagonize the piRNA pathway, leading to rapid evolution of pathway members.

Another striking outcome of our study is the degree to which convergent evolution occurs among *Drosophila* species and among telomeric TE lineages. To the extent that this convergence is a result of adaptive evolution, these events provide insight into both host and TE biology. The repeated loss of telomeric TEs from various *Drosophila* species raises the possibility that these elements impose a fitness cost to their host and that it is advantageous to be rid of them when an alternative mode of chromosome elongation is available. In terms of the TEs themselves, their frequent loss of ORF2 raises the possibility that TEs lacking this ORF are more fit than their progenitors (as long as they can utilize ORF2 peptides encoded other TE families). Similarly, the repeated capture of piRNA pathway genes by telomeric TEs suggests that gene capture increases TE fitness (see above). Horizontal transfer is emerging as a key mechanism by which TEs avoid extinction (Schaack et al. 2010; Venner et al. 2017) and our observation of repeated horizontal transfer of telomeric TEs supports such findings. Finally, our finding of convergent evolution of telomere localization is consistent with the telomeric niche acting as a “safe harbor” for TE insertions and thus, TEs with telomeric insertion biases may have increased fitness relative to other TEs.

These findings are not without limitations. Although it is a common approach in the field, and often necessary for analysis feasibility, our usage of consensus sequences to represent telomeric TE families masks sequence variation that may be present among individual TE copies. This sequence variation, which remains unaccounted for, could affect the topology of our TE trees which, in turn, could alter our estimates of horizontal transfer. Though we attempt to control for this issue by comparing the horizontal transfer rates of telomeric Jockey TE

consensi to those of non-telomeric Jockey TE consensi, this comparison could be confounded by differences in the biology and/or evolutionary constraints between these two groups.

The “transposon addiction” model envisions an intermediate period of cooperation between TE and host in the continuum between selfish TE mobilization and host co-option (or TE extinction) (Chang et al. 2025). We show that this cooperation period can persist for millions of years: Telomeric TEs originated over 60 million years ago, before the origin of the *Drosophila* genus, and remain present in the vast majority of *Drosophila* species that we surveyed in our study. However, cooperation and conflict are not mutually exclusive. Throughout the 60+ million-year period of apparent cooperation, we find numerous signs of conflict both between different telomeric TE lineages and between these TEs and their host genome. It is striking that these signatures of antagonistic evolution recurrently appear across the *Drosophila* phylogeny, showcasing the major role of convergence in the evolution of these TEs. This convergence extends beyond *Drosophila*, as well. Telomerase, the holoenzyme responsible for telomere elongation across most eukaryotes, utilizes a reverse transcriptase that likely originated from an ancient non-LTR retrotransposon (Eickbush 1997; Nakamura et al. 1997; Lingner et al. 1997). Thus, the same phenomenon of TE addiction that is currently playing out in *Drosophila* may have also occurred over a billion years ago, ultimately leading to the co-option of a retrotransposon to form the reverse transcriptase component of the holoenzyme that we now know as telomerase (Eickbush 1997). There are echoes of convergence in other TE addiction systems as well. For example, telomeric TEs are not the only “beneficial” TEs that are able to target their insertions to specific locations in the genome. The *G2/Jockey-3* and *ATHILA* retrotransposons that likely aid in centromere formation in *Drosophila* and *Arabidopsis*, respectively, have evolved mechanisms to target their insertions to centromeric chromatin (Shimada et al. 2024; Naish et al. 2021; Chabot et al. 2024). Similarly, the *Drosophila R2* retrotransposon, which plays a role in the maintenance of ribosomal DNA repeats, is able to

target its insertions to rDNA (Yang et al. 1999). In summary, we find that what appears to be a long-standing period of cooperation between telomeric TEs and their host genome is also rife with repeated occurrences of genetic conflict and dynamic evolution. Discovery and characterization of additional TE addiction systems will help to determine the generality of the TE addiction process observed at the *Drosophila* telomere.

Methods

Species

Drosophila species used in this study were provided by Matute lab (*Z. indianus* [RCR 17.04], *Z. gabonicus* [JD'18], *Z. davidi* [JD'18], *Z. nigranus* [18 CAR07 Z03]) and ordered from the National Drosophila Species Stock Center at Cornell (*D. auraria* [strain: 14028-0471.00] and *D. triauraria* [strain: 14028-0651.00], *Z. ghesquierei* [strain: 50000-2743.00], *Z. kolodkinae* [strain: 50000-2748.00]). *Z. taronus* [strain: 50001-1020.00] and *Z. camerounensis* [strain: 50001-1010.01] were ordered from the National Drosophila Species Stock Center at Cornell, but our sequencing data from these species suggested that they were misidentified and that the two stocks are likely to be from unsequenced species related to *Z. megalorchis* and *Z. sepsoides,* respectively. We therefore refer to these stocks as *Z. sp. megalorchis* and *Z. sp. sepsoides.*

Data acquisition

Out of the 109 long-read genome assemblies used in this study, 106 assemblies were acquired from NCBI: *D. albomicans* (NCBI BioProject PRJNA630751) (Mai et al. 2020), *D. bifasciata* (PRJNA565796) (Bracewell et al. 2020), *D. innubila* (PRJNA524688) (Hill et al. 2019), *D. miranda* (PRJNA474939) (Mahajan et al. 2018), *D. pseudoobscura* (PRJNA596268) (Liao et al. 2021), *D. serrata* (PRJNA355616) (Allen et al. 2017), *D. suzukii* (PRJNA594550) (Paris et al. 2020), *D. triauraria* (PRJNA627893) (Torosin et al. 2020), *D. mauritiana*, *D. sechellia*, and *D.*

simulans (PRJNA383250) (Chakraborty et al. 2021), *D. athabasca*, *D. lowei*, and *D. subobscura* (PRJNA545704) (Bracewell et al. 2019), and *D.ananassae*, *D. azteca*, *D. erecta*, *D. hydei*, *D. novamexicana*, *D. orena*, *D. persimilis*, *D. virilis*, and *Scaptodrosophila lebanonensis* (PRJNA475270). The remaining 83 NCBI assemblies were obtained from (PRJNA675888) (Kim et al. 2021). Three new assemblies were generated in our lab: *D. auraria*, *Z. sp. megalorchis* and *Z. sp. sepsoides* (see below).

Sequencing and Genome Assemblies

Using Monarch® HMW DNA Extraction Kit for Tissue, we extracted DNA from ~20 females of *D. auraria (*NEB T3010*)*, ~30 males of *Z. sp. megalorchis*, and *Z. sp. sepsoides* (NEB T3060S). We used Oxford Nanopore Technologies (ONT) SQK-LSK109 library preparation kit for *D. auraria* and the ONT SQK-LSK114 library preparation kit for the two *Zaprionus* species to construct libraries following the ONT Ligation Sequencing Kit protocol. The library of *D. auraria* was sequenced on a MinION R9.4 flow cell. Each library of two *Zaprionus* species was sequenced on a MinION R10.4.1 flow cell. Raw signal data were basecalled using the ONT Guppy software package version 4.0.15 for *D. auraria* and version 6.4.8 for the two *Zaprionus* species with default parameters.

We used Flye (version 2.8.1) (Kolmogorov et al. 2019) to assemble the *D. auraria* genome, followed by polishing with Medaka (version 1.1.3) (https://github.com/nanoporetech/medaka) and Nextpolish (version 1.3.1) (Hu et al. 2020). We removed allelic contigs using Purge Haplotigs (version 1.1.2) (Roach et al. 2018). We then performed Hi-C scaffolding of the remaining contigs using the 3D de novo assembly (3D-DNA) pipeline (version 201008). Hi-C data were generated as described in (Torosin et al. 2020). The two *Zaprionus* species genomes were assembled using Flye (version 2.9.2) with the option *--no-alt-contigs*, and polished using Medaka (version 1.6.0).

Identification of telomeric retrotransposons

We used telomeric retrotransposons of the genus *Drosophila* available on the Repbase database (TART_DVIR, HeT-A_DYAK, TART-A_DMEL, TART-B_DMEL, TART-C_DMEL, TAHRE_DMEL, and HeT-A_DMEL) as the query to run *RepeatProteinMask* (RepeatMasker version 4.1.2) (http://www.repeatmasker.org) on each species genome with the parameters *-noLowSimple -engine ncbi*. We clustered similar DNA sequences within species for ORF1 and ORF2 genes separately using CD-HIT-EST (version 4.8.1) (Li and Godzik 2006) with the parameters *-d 0 -r 0 -c 0.9 -n 8 -g 1 -T 4 -M 32000*. We aligned the DNA sequences of each cluster using MUSCLE (version 3.8.31) (Edgar 2004) with the parameters *-maxiters 1 -diags* and then ran PILER (version 1.0) (Edgar and Myers 2005) to construct consensus sequences for each cluster, removing consensi less than 600 bp in length. We used the MACSE pipeline (OMM_MACSE version 10.01) (Ranwez et al. 2018) with parameters *--no_prefiltering --no_postfiltering*) to generate frameshift-aware amino acid translations for ORF1 and ORF2 from each consensus sequence.

We concatenated the ORF1 amino acid sequences generated by the MACSE pipeline from all species, additionally adding the following known telomeric ORF1 peptides: TART_DVIR, HeT-A_DYAK, TART-A_DMEL, TART-B_DMEL, TART-C_DMEL, TAHRE_DMEL, and HeT-A_DMEL and the following known non-telomeric ORF1 sequences: Juan_DMEL, Jockey_DMEL, Doc_DMEL, and F-element_DMEL. We did the same thing for ORF2 (minus the HeT-A peptides because this TE family lacks ORF2) and ran MAFFT (version 7.471) (Katoh and Standley 2013) with *--auto* to generate ORF1 and ORF2 amino acid multiple sequence alignments. We trimmed the aligned sequences using ClipKIT (version 1.1.5) (Steenwyk et al. 2020) with parameters *-m kpic-gappy -g 0.5* for ORF1 and *-m kpic-gappy -g 0.6* for ORF2. We removed the trimmed sequences with more than 20% and 10% gaps, for ORF1 and ORF2, respectively. We

constructed ORF1 and ORF2 gene trees using IQ-TREE (version 1.6.12) (Nguyen et al. 2015) with parameters *-m TEST -abayes -bb 1000*.

We used an unbiased approach to define the telomeric clade for the ORF1 and ORF2 gene trees by calculating branch lengths between each novel ORF1/ORF2 gene and the known telomeric and non-telomeric ORF1/ORF2 genes included in our unrooted trees (see GitHub for code). The novel ORF1 and ORF2 genes that showed shorter branch lengths to their known telomeric homologs (compared to their non-telomeric homologs) were assigned as putative telomeric clade peptides. This approach resulted in a single monophyletic clade of putative telomeric ORF1 and ORF2 genes (**Figure S3**).

We next sought to identify the genomic locations of each candidate telomeric ORF1 and ORF2 gene. The telomeric clade ORF1 and ORF2 consensus DNA sequences from each species were used as a custom library when running RepeatMasker (version 4.1.2) with parameters *-no_is -norna -nolow -engine ncbi*, retaining matches with <= 3% divergence from their consensus. We used the RepeatMasker hits as starting points to generate manually curated, full-length telomeric TE consensi. We identified arrays of RepeatMasker hits at contig ends (or within short unplaced contigs) and generated a dot plot of the array sequence using YASS, a genomic similarity search web tool (Noé and Kucherov 2005). We manually examined the dotplots to identify the boundaries of the monomers that compose the head-to-tail TE arrays. We then compared the monomers from the same TE family to identify the longest, unfragmented, representative sequence for each TE family. For species with Oxford Nanopore data, we aligned the nanopore genomic sequencing data to the representative TE sequence and performed one round of polishing using medaka (version 1.6.0). All consensi (except those from species lacking Illumina data, see **Table S7**) were then further iteratively polished using Pilon (version 1.24) (Walker et al. 2014) with Illumina data until no new changes were made to

776 the consensus. We searched each consensus for intact ORF1 and ORF2 open reading frames
777 and performed an additional round of polishing using Pilon with Illumina data (where available,
778 see **Table S7**) when the consensus sequence contained fragmented ORFs.
779
780 All manually curated telomeric TE consensi include both 5' and 3' UTRs. Recently derived
781 ORF2-defective, nonautonomous elements share the same 5' and 3' UTRs with their
782 autonomous progenitors. Our approach classifies the autonomous and non-autonomous
783 versions of these elements as different TE families (**Table S2**). In the cases where there was no
784 autonomous progenitor present in the genome, we defined the non-autonomous TE boundaries
785 by aligning multiple copies of the same TE, requiring that multiple copies support the same 5'
786 and 3' boundaries.
787
788 To estimate the genomic copy number of each telomeric TE family, we aligned the long-read
789 genomic sequencing data from the host species to the TE consensi. To avoid underestimation
790 of copy number due to 5' end truncation of telomeric TEs, we calculated the median per-base
791 read depth across a 1 kb interval of the TE consensus beginning 500 bp from the 3' end of the
792 sequence, similar to the approach described in (McGurk et al. 2021). We then estimated copy
793 number by dividing this value by the median whole-genome read depth (averaged across 100bp
794 windows).
795
796 **Telomeric Retrotransposon (TR) clade definitions**
797 We defined TR clades so that they would be as consistent as possible with the original
798 Villasante designations while also being monophyletic (**Figure S5**). *TR1*: Villasante et al (2007)
799 identified *TR1* elements in species from the *obscura*, *virilis*, and *repleta* groups. In our ORF2
800 tree, these sequences form a clade that is closely related to two other clades: one composed of
801 TRs from species in the *Zaprionus*, *Scaptomyza*, and *cardini* groups, and the other containing

802 species from the *obscura* group. The *virilis*, *repleta*, *Zaprionus*, *Scaptomyza*, and *cardini* group
803 species are all members of the *Drosophila* subgenus, thus we expanded the *TR1* clade to
804 encompass the related TRs from these groups. *TR2*: Villasante et al (2007) identified *TR2*
805 elements in species from the *ananassae*, *obscura*, *virilis* and Hawaiian *Drosophila* groups. Our
806 *TR2* clade represents the smallest possible monophyletic clade that includes these elements.
807 *TR3*: Villasante et al (2007) identified *TR3* elements in species from the *obscura* and *repleta*
808 species groups. Our *TR3* clade represents the smallest possible monophyletic clade that
809 includes these elements. *TR4* and *HTT*: Villasante et al (2007) identified *TR4* elements in
810 species from the *obscura* group. These elements are closely related to TRs that we identified in
811 species from the *montium* group, however expanding the *TR4* clade to include the *montium*
812 group elements would result in a clade that is not monophyletic. We therefore combined both
813 *HTT* and *TR4* elements into a single monophyletic clade, which we refer to as "*HTT/TR4*". *TR5*
814 and *TR6*: these are novel monophyletic clades that were not previously identified by Villasante
815 et al.
816
817 **Identification of the telomeric retrotransposons with host gene capture.**
818 We used *D. melanogaster* amino acid sequences (longest isoform per gene, r6.42 annotation)
819 as queries and the DNA consensus sequences of all species' telomeric transposons as the
820 database to run TBLASTN (version 2.10.1) (Camacho et al. 2009) with parameters *-outfmt 6 -*
821 *evalue 1e-5*. For each TBLASTN hit, we determined whether there was significant similarity
822 between the telomeric TE and its host species gene at the DNA level by running BLASTN
823 (version 2.10.1) with parameters *-outfmt 6 -evalue 1e-3*. To search for older gene capture
824 events we also ran BLASTN with parameters *-outfmt 6 -word_size 4* to identify host gene/TE
825 sequences with higher levels of sequence divergence.
826

Phylogenetic reconstruction (species, telomeric retrotransposons, and their reconciliation)

We used full-length ORF1 and ORF2 sequences from our curated telomeric TE consensi to create telomeric TE gene trees. We additionally added the following known non-telomeric Jockey clade ORFs from Repbase TE families to use as outgroup sequences: Jockey_DMEL, Jockey-1_DEr, Jockey-1_DEu, Jockey-1_DF, Jockey-1_DGri, Jockey-1_DK, Jockey-1_DRh, Jockey-1_DT, Jockey-1_DVi, Jockey-1_DWi, Jockey-1_DYa, Jockey-10_DAn, Jockey-10_DRh, Jockey-11_DAn, Jockey-11B_DBp, Jockey-14_DBp, Jockey-2_DEl, Jockey-2_DEu, Jockey-2_DRh, Jockey-2_DT, Jockey-3_DEu, Jockey-3_DF, Jockey-3_DK, Jockey-3_DRh, Jockey-3_DVi, Jockey-4_DEl, Jockey-4_DEu, Jockey-4_DF, Jockey-4_DPer, Jockey-4_DVi, Jockey-5_DAn, Jockey-5_DBp, Jockey-5_DTa, Jockey-6_DAn, Jockey-6_DEl, Jockey-6_DEu, Jockey-6_DF, Jockey-7_DAn, Jockey-7_DF, Jockey-8_DAn. We aligned the ORF peptides using MAFFT (version 7.471) and trimmed the alignments using ClipKIT (version 1.1.5) with parameters *-m kpic-gappy -g 0.1*. We removed sequences with > 20% gaps from each ORF alignment and then created the telomeric ORF1 and ORF2 gene trees using IQ-TREE (version 1.6.12) with parameters *-m TEST -abayes -bb 1000* and rooted the trees using the non-telomeric outgroup sequences. All tree figures were created using the iTOL tool (Letunic and Bork 2024).

To reconcile the above TE ORF1 and ORF2 trees with the *Drosophila* species tree, we first pruned the dated species tree from (Suvorov et al. 2022), keeping only the species studied here. We then ran Ranger-DTL-Dated (version 2.0) (Bansal et al. 2018) with default parameters to perform the gene tree/species tree reconciliation. We calculated support values for each reconciliation event using the AggregateRanger program on the output of 1000 Ranger-DTL-Dated reconciliation runs (**Table S8**). We used a single consensus sequence for each TE family as input for reconciliation. Note that Ranger-DTL identifies TE “duplications” as cases where an

ancestral TE lineage gives rise to two novel TE lineages, which is akin to “speciation” from the point of view of the TE. Ranger-DTL compares the TE tree to the species tree and identifies points of incongruence between the two. It then finds the most parsimonious combination of duplication (i.e. TE “speciation”), horizontal transfer, and loss (i.e. TE extinction) events to reconcile these points of incongruence using a reconciliation cost for each of these events. Horizontal transfer is assigned the highest cost: Transfer events are assigned a cost that is 1.5-fold higher than the duplication cost and three-fold higher than the loss cost, which makes our approach conservative in terms of identifying horizontal transfer events. To assess the accuracy of horizontal transfer events identified by Ranger-DTL, we used BLAST to create local alignments between pairs of TE sequences and recorded the alignment bit score, which is a normalized measure of both the percent identity and alignment length of a given BLAST match. For all Ranger-DTL events involving horizontal transfers between two tips of the species tree, we recorded the best alignment bit score for the BLAST search between the donor telomeric TE lineage and the recipient lineage. As a control, we used bit scores from related, vertically inherited, non-telomeric Jockey elements between the same species pair.

We used the same Ranger-DTL approach to infer horizontal transfer events from non-telomeric Jockey sequences obtained from (Tambones et al. 2019). The total number of transfer events for each group (i.e. telomeric versus non-telomeric) was then normalized by the total number of nodes in the corresponding tree. The boxplots in Figure 2C show the distribution of these estimates across 1000 runs of Ranger-DTL.

We generated two novel *de novo* genome assemblies for this study: *Z. sp. megalorchis* and *Z. sp. sepsoides*. We placed these species within the *Zaprionus* species tree using the same methods as described in (Suvorov et al. 2022). Briefly, we identified single-copy orthologs from each *Zaprionus* species genome assembly by running BUSCO (version 5.4.7) (Manni et al.

2021) with the Diptera data set of the OrthoDB v10 release (diptera_odb10) (Kriventseva et al. 2019) using default parameters. We used IQ-TREE (version 1.6.12) to obtain the gene trees for each single-copy ortholog, as described in (Suvorov et al. 2022). The gene trees were used to infer the *Zaprionus* species tree using ASTRAL (version 5.7.8) (Zhang et al. 2018) with default parameters.

To make the *aub* gene tree for *Zaprionus* species, we first built *aub* gene models in each species using genBlastG (She et al. 2011) with parameters *-e 1e-5 -g T -r 3 -c 0.6* and the *D. melanogaster* Aub peptide sequence as a query. We then used MAFFT (version 7.5.20) with *–auto* to align the *aub* coding sequences and IQ-TREE (version 1.6.12) with parameters *-m TEST -bb 1000* to infer the gene trees.

To test for accelerated evolution of the *aub* gene in *Zaprionus* we first identified all regions of *aub* coding sequence that were not captured by any *Zaprionus* species (i.e. uncaptured sequence) and concatenated these regions together for each *aub* gene in each species. We aligned these sequences using MAFFT (version 7.520) with *–auto*. We then extracted two sequence subsets from this alignment: those from *Z. nigranus*, *taronus*, *davidi*, and *capensis* (capture clade), and those from *Z. camerounensis*, *lachaisei*, and *vittiger* (control clade). We added *Z. inermis aub* sequence to each sequence subset as an outgroup and used ModelFinder from IQ-TREE which identified “TPM2+G4” as the best-fitting DNA substitution model for both sequence subsets. We then inferred gene trees for both the capture and control sequence subsets from the uncaptured *aub* sequence using IQ-TREE, specifying *Z. inermis* as the outgroup and “TPM2+G4” as the substitution model and summed the branch lengths of the resulting trees.

We then repeated the above procedure using Region C of *aub*, which was captured by the *Z. nigranus*, *taronus*, *davidi*, and *capensis* telomeric TEs, providing the corresponding uncaptured tree file as a constraint tree topology. We then compared the summed branch lengths between the captured and uncaptured *aub* regions, for both the capture clade and the control clade. We used Fisher's Exact test to compare the proportion of substitutions among all sites between captured and uncaptured sequence subsets.

To test for introgression between *Z. taronus* and *Z. nigranus*, we downloaded gene trees from BUSCO single-copy orthologs generated in a previous study (Suvorov et al. 2022). Across all ~2500 trees, we counted the number of times we observed the following species pairs as sister taxa, which all share the same most recent common ancestor, according to the species tree in (Suvorov et al. 2022): *Z. taronus* vs *Z. nigranus* (n=71), *Z. davidi* vs *Z. nigranus* (n=9), *Z. capensis* vs *Z. nigranus* (n=12), *Z. taronus* vs *Z. vittiger* (n=11), *Z. taronus* vs *Z. camerounensis* (n=6), *Z. taronus* vs *Z. lachaisei* (n=8). Overall, 2.77% of gene trees show *Z. taronus* and *Z. nigranus* as sister taxa compared to an average of 0.35% support for the other pairs shown above.

Small RNA sequencing

For 7 *Zaprionus* species plus *D. auraria* and *D. triauraria*, we extracted small RNAs from 10-24 pairs of ovaries using the TraPR Small RNA Isolation and Library Prep Kit (Lexogen Inc, 135.08) (Grentzinger et al. 2020), following the kit protocol. The small RNA libraries were sent to Novogene Corporation Inc. for NovaSeq SE50 sequencing.

We subsequently performed deep sequencing of sodium periodate treated small RNAs, with 2S rRNA blocking using a terminator oligo (Wickersheim and Blumenstiel 2013), from four species whose telomeric TEs had captured the host genes *aub* or *piwi*: *D. auraria*, *D. triauraria*, *Z.*

gabonicus, and *Z. indianus*. We unfortunately lost our *Z. davidii* stock after the initial small RNA library was generated and were unable to obtain a replacement. See Supplemental Methods for protocol.

We used Trim Galore! (version 0.6.10) (https://github.com/FelixKrueger/TrimGalore) to trim adapter sequences from raw small RNA reads and SortMeRNA (version 4.3.6) (Kopylova et al. 2012) to remove reads arising from tRNAs and rRNAs. TRNA genes were predicted by tRNAscan-SE (version 2.0.11) (Chan et al. 2021) and rRNA databases were generated by BLASTN searches with *D. melanogaster* rRNA queries. The filtered small RNA reads were aligned using ShortStack (version 3.8.5) (Axtell 2013) with the parameters *--dicermin 20 --dicermax 35 --mismatches 2*.

Detection of piRNA reads mapped to telomeric TEs and host genes, *piwi* and *aub*

We sought to quantify piRNA abundance from each telomeric TE family, as well as the host genes that were captured by telomeric TEs, excluding the captured region of the host gene to avoid cross-mapping of TE-derived piRNAs to the host gene. To do this, we masked the telomeric TEs and host genes (i.e. *piwi*, *aub*, and *CG12520*) in the appropriate species genome using RepeatMasker (version 4.1.2) with parameters *-no_is -norna -nolow -pa 4 -engine ncbi*. We then created a custom reference genome by appending consensus sequences of each telomeric TE family as well as mRNA sequences of the longest isoform of the captured host gene(s), where the captured region was masked using BEDTools, version 2.25.0 (Quinlan and Hall 2010), to the masked genome assembly. Next, we used ShortStack (version 3.8.5) with the parameters *--dicermin 20 --dicermax 35 --mismatches 2* to align the small RNA reads to the custom reference genome. We retained reads with lengths between 23 bp and 30 bp and used BEDTools (version 2.25.0) to calculate coverage of sense and anti-sense alignments on the telomeric TEs and the relevant host transcript(s) (i.e. *piwi*, *aub*, and/or *CG12520*).

956

957 **Phasing and ping-pong signatures**

958 To assess signatures of phasing, we used sense strand piRNA alignments from three regions of
959 *piwi*/*aub*: upstream from the gene capture region, downstream, and the full gene, excluded the
960 capture region. We collapsed exact duplicates and, for each 3' end, we calculated the
961 downstream distance to the nearest piRNA 5' end. For each distance between 0-20 bp, we
962 calculated the fraction of 3' – 5' piRNAs pairs whose ends were separated by that distance.

963

964 To assess ping-pong signatures, we used SAMtools (Danecek et al. 2021) to extract TE-derived
965 antisense piRNAs and *piwi*/*aub* derived sense piRNAs, from alignments to the unmasked
966 captured host gene and unmasked gene capture TE generated using ShortStack (version 3.8.5;
967 *--dicermin 20 --dicermax 35 --mismatches 2*). We then aligned both TE-derived antisense
968 piRNAs and *piwi*/*aub* derived sense piRNAs to the unmasked *piwi*/*aub* host gene and used
969 PingPongPro (Uhrig and Klein 2019) to identify ping-pong signatures.

970

971 **d_N/d_S (K_a/K_s) analysis**

972 We used MEGA X (Kumar et al. 2018) to align coding sequences and used *KaKs_Calculator*
973 *2.0* (Wang et al. 2010) to calculate the ratio of the number of nonsynonymous substitutions per
974 nonsynonymous site to the number of synonymous substitutions per synonymous site. Fisher's
975 Exact Test was used (within *KaKs_Calculator*) to test the null hypothesis of neutral evolution
976 (i.e. equal rates of synonymous versus non-synonymous substitutions).

977

978 **HMM searches**

979 We initially used an alignment of all seven unusual ORF2 sequences as a query to perform
980 sensitive HMM-HMM sequence searches against the *UniRef30_2023_02* database on the MPI
981 Bioinformatics Toolkit webserver (Zimmermann et al. 2018) using HHblits (Remmert et al. 2011)

but found no significant hits outside of our query species. We then aligned the seven unusual ORF2 amino acid sequences using MAFFT (version 7.520) and created an HMM profile from the alignment using hmmbuild from HMMER3 (version 3.4)(Eddy 2011). We compiled a database of all ORF2 peptides in *D. virilis*, *D. americana*, *D. novamexicana*, and *D. littoralis* telomeric TEs that contained the endonuclease, RT, and X domains and searched this database with our HMM profile as the query using hmmsearch from HMMER3.

Data access

The sequencing and genome assembly data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA1183390. Manually curated telomeric TE consensus sequences have been submitted to GitHub (https://github.com/jaehakson/DrosophilaTelomericRetrotransposons). All code used for this study has been submitted to GitHub (https://github.com/jaehakson/DrosophilaTelomericRetrotransposons) and is available as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgements

The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here. We also acknowledge the Cornell National Drosophila Species Stock Center and the laboratory of Daniel

Matute for providing fly stocks used in this study. This work was supported by the National Institutes of Health grants R01GM130698 and R35GM152168 to C.E.E. and R35GM124684 M.T.L.

Author contributions: C.E.E., M.T.L., and J.H.S. contributed to project design. W.C. performed all experiments. J.H.S., M.A.L., C.E.E. and M.V performed all computational analyses. J.H.S., M.T.L., and C.E.E. interpreted the data and wrote the manuscript. C.E.E. supervised the project.

Figure Legends

Figure 1. **Evolutionary relationships of the species included in this study.** Colors correspond to major *Drosophila* clades, as delineated in (Suvorov et al. 2022). The dashed grey branches indicate the 15 species for which no telomeric TEs were identified. The barplot shows the number of telomeric TE families identified in each species.

Figure 2. Evolutionary history of telomeric TE families. (A) Phylogeny of telomeric TE families inferred from ORF2 peptide sequences. Six major telomeric retrotransposon (*TR*) clades were identified, each with 100% bootstrap support (see **Figure S6**). Naming scheme is based on *TR* clades originally described in (Villasante et al. 2007), with the addition of the novel *TR5* and *TR6* clades identified here. The tree branches are colored by *TR* clade, while the outer circle colors correspond to the host species clades shown in Panel B. The tree is rooted using ORF2 sequences from non-telomeric jockey clade TEs (not shown, see **Figure S6**). (B) Dated species tree showing major events in the diversification of *Drosophila* telomeric TEs based on species tree/TE tree reconciliation (see Methods). Rectangles are colored based on the *TR* clades shown in Panel A. *TR* clade origins are indicated by filled rectangles while losses are indicated by an “X”. Horizontal transfer events are summarized by circular (transfers within subgenera) and vertical (transfers between subgenera) arrows. Both types of arrows are colored based on TR clade. Two additional transfer events are shown by curved arrows: one giving rise to the origin of the *TR2* clade (black arrow) and the other showing the transfer of a *TR2* clade TE to the outgroup species *Leucophenga varia* (blue arrow). See Supplemental Text in Supplemental Materials for additional descriptions of TR clade evolution. Parentheses next to the major *Drosophila* species clades indicate the number of species harboring the corresponding *TR* clade over the

1030 total number of species within each species clade. (C) Rates of horizontal transfer for non-telomeric versus telomeric
1031 Jockey clade TEs.
1032
1033 **Figure 3**. **Convergence and neofunctionalization in telomeric TE families.** (A) Species tree showing the six
1034 species harboring non-telomeric clade TEs at their telomeres (NTTs, red branches) and the four species carrying
1035 potentially neo-functionalized ORF2 (blue arrows). The outer circle colors delineate major species clades. (B) Pruned
1036 TE tree showing that the NTT families (red branches) form a monophyletic subclade within the larger clade of non-
1037 telomeric Jockey elements, inferred from ORF2 peptide sequences (C) Schematic examples of two species (*D. virilis*
1038 and *S. hsui*) showing telomeric regions composed of head-to-tail arrays of telomeric TEs and NTTs. In *S. hsui*, we
1039 found a fragmented TE sequence (red arrow) at the telomeric end of the contig whose partial ORF2 sequence placed
1040 it within the telomeric clade. (D) Seven telomeric TEs from members of the *virilis* species group (see **Figure 3A**)
1041 contain an ORF2 (brown box) that shows no significant homology to the ORF2 (purple box) of other telomeric TE
1042 families from this same group via peptide BLAST. Lengths correspond to the full length TE in basepairs (bp) and the
1043 ORF2-encoded peptide in amino acids (aa). (E) A profile HMM created from the seven unusual ORF2 sequences
1044 shows homology to several regions of the ORF2 from a *D. littoralis* telomeric TE that also contains the canonical
1045 endonuclease and RT domains usually found in Jockey clade ORF2 peptides, plus a domain of unknown function
1046 termed X. The unusual ORF2 sequences in Panel E may represent a case of neofunctionalization, where the
1047 ancestral function provided by the endonuclease and RT domains has been lost and replaced by a currently unknown
1048 function.
1049
1050 **Figure 4 Repeated capture of piRNA pathway genes by telomeric TEs**. (A) Species tree showing independent
1051 gene capture events (red thick lines) at the nodes. Labels indicate the gene that was captured and the outer circle is
1052 colored by major species clade. (B) Diagrams showing the location of the captured sequence within the full length TE
1053 consensus sequence and within the host gene. Green and purple boxes indicate the locations of ORF1 and ORF2,
1054 respectively. TEs lacking a purple box have lost ORF2. Note that the 5' UTR of *TART-A* is copied from a portion of
1055 the 3' UTR during replication, thus the 3'UTR of *TART-A* also carries *nxf2* sequences (Ellison et al. 2020), but they
1056 are not shown here for clarity.
1057
1058 **Figure 5 Multiple captures of *aubergine* by telomeric TEs in *Zaprionus*. (**A) *Zaprionus* species tree showing
1059 independent gene capture events. The rectangles are colored based on the *aub* region that was captured. Diagrams

show the location of the captured region within the full length TE. Note: region B of *aub* was captured in the ancestor of *Z. vittiger* and *Z. capensis* (solid blue rectangle) and subsequently lost in the *Z. davidi* / *Z.taronus* / *Z.capensis* clade (open blue rectangle), whose TEs now carry region C of *aub*. (B) Normalized piRNA abundance from the captured host gene (red = *aubergine*, green = *piwi*) calculated in 80 bp sliding windows across the gene transcript. *Aub* is used as a control (blue) in species where *piwi* was captured and *piwi* is used as a control in species where *aub* was captured. The control gene produces significantly fewer piRNAs in all within-species comparisons (Wilcoxon test $P < 0.05$ in all cases, see individual p-values below). (C) Gene tree inferred using *aub* coding sequence. The clades analyzed in panel D are indicated by vertical bars. (D) Comparison of rate of evolution of the captured (red) portion of *aubergine* versus the uncaptured portion. In the species where telomeric TEs have captured part of *aub* (capture clade), the homologous sequence in the *aub* gene is evolving approximately twice as fast as the uncaptured portion of the gene (Fisher's Exact Test $P = 2.7 \times 10^{-5}$). In species where this region has not been captured, it evolves at a similar rate as the uncaptured portion of the gene (Fisher's Exact Test $P = 0.64$). P-values for Panel B Wilcoxon Tests: *D. auraria* $P = 0.026$, *D. triauraria* $P = 1.3 \times 10^{-4}$, *Z. davidi* $P = 3.0 \times 10^{-4}$, *Z. gabonicus* $P = 6.3 \times 10^{-4}$, *Z. indianus* $P = 5.2 \times 10^{-8}$.

References

Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martín-Gallardo A, Villasante A. 2004. TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. *Mol Biol Evol* **21**: 1620–1624.

Allen SL, Delaney EK, Kopp A, Chenoweth SF. 2017. Single-molecule sequencing of the Drosophila serrata genome. *G3 (Bethesda)* **7**: 781–788.

Almeida MV, Vernaz G, Putman ALK, Miska EA. 2022. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet* **38**: 529–553.

Arkhipova IR. 2012. Telomerase, Retrotransposons, and Evolution. In *Telomerases*, pp. 265–299, John Wiley & Sons, Inc., Hoboken, NJ, USA.

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19**: 740–751.

1086 Bansal MS, Kellis M, Kordi M, Kundu S. 2018. RANGER-DTL 2.0: rigorous reconstruction of gene-family
1087 evolution by duplication, transfer and loss. *Bioinformatics* **34**: 3214–3216.

1088 Batki J, Schnabl J, Wang J, Handler D, Andreev VI, Stieger CE, Novatchkova M, Lampersberger L,
1089 Kauneckaite K, Xie W, et al. 2019. The nascent RNA binding complex SFiNX licenses piRNA-
1090 guided heterochromatin formation. *Nat Struct Mol Biol* **26**: 720–731.

1091 Biessmann H, Zurovcova M, Yao JG, Lozovskaya E, Walter MF. 2000. A telomeric satellite in Drosophila
1092 virilis and its sibling species. *Chromosoma* **109**: 372–380.

1093 Bracewell R, Chatla K, Nalley MJ, Bachtrog D. 2019. Dynamic turnover of centromeres drives karyotype
1094 evolution in Drosophila. *Elife* **8**. http://dx.doi.org/10.7554/eLife.49002.

1095 Bracewell R, Tran A, Chatla K, Bachtrog D. 2020. Chromosome-level assembly of Drosophila bifasciata
1096 reveals important karyotypic transition of the X chromosome. *G3 (Bethesda)* **10**: 891–897.

1097 Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small
1098 RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089–
1099 1103.

1100 Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA, Reddel RR. 1997. Evidence for an alternative
1101 mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat*
1102 *Med* **3**: 1271–1274.

1103 Cacchione S, Cenci G, Dion-Côté A-M, Barbash DA, Raffa GD. 2025. Maintaining Telomeres without
1104 Telomerase in Drosophila: Novel Mechanisms and Rapid Evolution to Save a Genus. **17**:
1105 a041708.

1106 Cacchione S, Cenci G, Raffa GD. 2020. Silence at the End: How Drosophila Regulates Expression and
1107 Transposition of Telomeric Retroelements. *J Mol Biol* **432**: 4305–4321.

1108 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:

architecture and applications. *BMC Bioinformatics* **10**: 421.

Casacuberta E. 2017. Drosophila: Retrotransposons making up telomeres. *Viruses* **9**. http://dx.doi.org/10.3390/v9070192.

Casacuberta E, Pardue M-L. 2003. Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilis group. *Proc Natl Acad Sci U S A* **100**: 3363–3368.

Catoni M, Jonesman T, Cerruti E, Paszkowski J. 2019. Mobilization of Pack-CACTA transposons in Arabidopsis suggests the mechanism of gene shuffling. *Nucleic Acids Res* **47**: 1311–1320.

Cenci G, Ciapponi L, Gatti M. 2005. The mechanism of telomere protection: a comparison between Drosophila and humans. *Chromosoma* **114**: 135–145.

Chabot BJ, Sun R, Amjad A, Hoyt SJ, Ouyang L, Courret C, Drennan R, Leo L, Larracuente AM, Core LJ, et al. 2024. Transcription of a centromere-enriched retroelement and local retention of its RNA are significant features of the CENP-A chromatin landscape. *Genome Biol* **25**: 295.

Chakraborty M, Chang C-H, Khost DE, Vedanayagam J, Adrion JR, Liao Y, Montooth KL, Meiklejohn CD, Larracuente AM, Emerson JJ. 2021. Evolution of genome structure in the Drosophila simulans species complex. *Genome Res* **31**: 380–396.

Chan PP, Lin BY, Mak AJ, Lowe TM. 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res* **49**: 9077–9096.

Chang N-C, Wells JN, Wang AY, Schofield P, Huang Y-C, Truong VH, Simoes-Costa M, Feschotte C. 2025. Gag proteins encoded by endogenous retroviruses are required for zebrafish development. *Proc Natl Acad Sci U S A* **122**: e2411446122.

Cosby RL, Chang N-C, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev* **33**: 1098–1116.

Cui M, Bai Y, Li K, Rong YS. 2021. Taming active transposons at Drosophila telomeres: The

interconnection between HipHop's roles in capping and transcriptional silencing. *PLoS Genet* **17**: e1009925.

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. *Mol Cell* **50**: 749–761.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**: giab008.

Dui W, Lu W, Ma J, Jiao R. 2012. A systematic phenotypic screen of F-box genes through a tissue-specific RNAi-based approach in Drosophila. *J Genet Genomics* **39**: 397–413.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* **7**: e1002195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Edgar RC, Myers EW. 2005. PILER: identification and classification of genomic repeats. *Bioinformatics* **21 Suppl 1**: i152-8.

Eickbush TH. 1997. Telomerase and retrotransposons: which came first? *Science* **277**: 911–912.

Eickbush TH, Malik HS. 2002. Origins and evolution of retrotransposons. In *Mobile DNA II*, pp. 1111–1144, American Society of Microbiology.

Ellison CE, Bachtrog D. 2015. Non-allelic gene conversion enables rapid evolutionary change at multiple regulatory sites encoded by transposable elements. *Elife* **4**. https://elifesciences.org/articles/05899 (Accessed October 14, 2024).

Ellison CE, Kagda MS, Cao W. 2020. Telomeric TART elements target the piRNA machinery in Drosophila. *PLoS Biol* **18**: e3000689.

Fabry MH, Falconio FA, Joud F, Lythgoe EK, Czech B, Hannon GJ. 2021. Maternally inherited piRNAs

direct transient heterochromatin formation at active transposons during early Drosophila embryogenesis. *Elife* **10**. http://dx.doi.org/10.7554/eLife.68573.

Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu Rev Genet* **41**: 331–368.

Floriano AM, El-Filali A, Amoros J, Buysse M, Jourdan-Pineau H, Sprong H, Kohl R, Dirks RP, Schaap P, Koehorst J, et al. 2025. Comparative genomics of *Rickettsiella* bacteria reveal variable metabolic pathways potentially involved in symbiotic interactions with arthropods. *Peer Community J* **5**. http://dx.doi.org/10.24072/pcjournal.633 (Accessed April 10, 2026).

Frydrychova RC, Mason JM, Archer TK. 2008. HP1 is distributed within distinct chromatin domains at Drosophila telomeres. *Genetics* **180**: 121–131.

Fu Y, Kawabe A, Etcheverry M, Ito T, Toyoda A, Fujiyama A, Colot V, Tarutani Y, Kakutani T. 2013. Mobilization of a plant transposon by expression of the transposon-encoded anti-silencing factor. *EMBO J* **32**: 2407–2417.

Fueyo R, Judd J, Feschotte C, Wysocka J. 2022. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* **23**: 481–497.

Fujiwara H, Osanai M, Matsumoto T, Kojima KK. 2005. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, Bombyx mori. *Chromosome Res* **13**: 455–467.

Fuller AM, Cook EG, Kelley KJ, Pardue M-L. 2010. Gag proteins of Drosophila telomeric retrotransposons: collaborative targeting to chromosome ends. *Genetics* **184**: 629–636.

Gladyshev EA, Arkhipova IR. 2007. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc Natl Acad Sci U S A* **104**: 9352–9357.

Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, et al. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism

of genome shuffling in eukaryotes. *Nat Commun* **7**: 10716.

Grentzinger T, Oberlin S, Schott G, Handler D, Svozil J, Barragan-Borrero V, Humbert A, Duharcourt S, Brennecke J, Voinnet O. 2020. A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic Acids Res* **48**: e79.

Gu X, Ross PA, Gill A, Yang Q, Ansermin E, Sharma S, Soleimannejad S, Sharma K, Callahan A, Brown C, et al. 2023. A rapidly spreading deleterious aphid endosymbiont that uses horizontal as well as vertical transmission. *Proc Natl Acad Sci U S A* **120**: e2217278120.

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**: 1587–1590.

Higashiyama T, Noutoshi Y, Fujie M, Yamada T. 1997. Zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region. *EMBO J* **16**: 3715–3723.

Hill T, Koseva BS, Unckless RL. 2019. The genome of Drosophila innubila reveals lineage-specific patterns of selection in immune genes. *Mol Biol Evol* **36**: 1405–1417.

Hoffmann AA, Cooper BS. 2024. Describing endosymbiont-host interactions within the parasitism-mutualism continuum. *Ecol Evol* **14**: e11705.

Hosaka A, Saito R, Takashima K, Sasaki T, Fu Y, Kawabe A, Ito T, Toyoda A, Fujiyama A, Tarutani Y, et al. 2017. Evolution of sequence-specific anti-silencing systems in Arabidopsis. *Nat Commun* **8**: 2161.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**: 2253–2255.

Jonathan F, Lopez-Villavicencio M, Vincent D, Rignault G, Rachel F, Camilo S, Karina LS-B, Patrick B, Andre VLF, Carolina P-D, et al. 2024. Genome evolution and between-host transmission of

Spiroplasma endosymbiont in wild communities of Morpho butterflies. *bioRxiv* 2024.02.22.581604. https://www.biorxiv.org/content/10.1101/2024.02.22.581604v2.abstract (Accessed April 17, 2025).

Kalmykova A. 2023. Telomere checkpoint in development and aging. *Int J Mol Sci* **24**: 15979.

Kalmykova AI, Sokolova OA. 2023. Retrotransposons and telomeres. *Biochemistry (Mosc)* **88**: 1739–1753.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Khurana JS, Xu J, Weng Z, Theurkauf WE. 2010. Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. *PLoS Genet* **6**: e1001246.

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ERR, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid genomes. *Elife* **10**. http://dx.doi.org/10.7554/eLife.66405 (Accessed June 5, 2025).

Kipreos ET, Pagano M. 2000. The F-box protein family. *Genome Biol* **1**: REVIEWS3002.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**: 540–546.

Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res* **47**: D807–D811.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549.

1223 Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022.
1224 TimeTree 5: An expanded resource for species divergence times. *Mol Biol Evol* **39**.
1225 http://dx.doi.org/10.1093/molbev/msac174.

1226 Lawlor MA, Ellison CE. 2023. Evolutionary dynamics between transposable elements and their host
1227 genomes: mechanisms of suppression and escape. *Curr Opin Genet Dev* **82**: 102092.

1228 Lee YCG, Leek C, Levine MT. 2017. Recurrent innovation at genes required for telomere integrity in
1229 Drosophila. *Mol Biol Evol* **34**: 467–482.

1230 Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display
1231 and annotation tool. *Nucleic Acids Res* **52**: W78–W82.

1232 Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM. 1993. Transposons in place of telomeric
1233 repeats at a Drosophila telomere. *Cell* **75**: 1083–1093.

1234 Lewis SH, Salmela H, Obbard DJ. 2016. Duplication and diversification of Dipteran Argonaute genes, and
1235 the evolutionary divergence of Piwi and aubergine. *Genome Biol Evol* **8**: 507–518.

1236 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or
1237 nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

1238 Liao Y, Zhang X, Chakraborty M, Emerson JJ. 2021. Topologically associating domains and their role in
1239 the evolution of genome structure and function in Drosophila. *Genome Res* **31**: 397–410.

1240 Lin L, Huang Y, McIntyre J, Chang C-H, Colmenares S, Lee YCG. 2024. Prevalent fast evolution of genes
1241 involved in heterochromatin functions. *Mol Biol Evol* **41**.
1242 http://dx.doi.org/10.1093/molbev/msae181.

1243 Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR. 1997. Reverse transcriptase
1244 motifs in the catalytic subunit of telomerase. *Science* **276**: 561–567.

1245 Lisch D. 2009. Epigenetic regulation of transposable elements in plants. *Annu Rev Plant Biol* **60**: 43–66.

Mahajan S, Wei KH-C, Nalley MJ, Gibilisco L, Bachtrog D. 2018. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biol* **16**: e2006348.

Mai D, Nalley MJ, Bachtrog D. 2020. Patterns of Genomic Differentiation in the Drosophila nasuta Species Complex. *Mol Biol Evol* **37**: 208–220.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **38**: 4647–4654.

Mano S, Innan H. 2008. The evolutionary rate of duplicated genes under concerted evolution. *Genetics* **180**: 493–505.

Markova DN, Christensen SM, Betrán E. 2020. Telomere-specialized retroelements in Drosophila: Adaptive symbionts of the genome, neutral, or in conflict? *Bioessays* **42**: e1900154.

Mason JM, Biessmann H. 1995. The unusual telomeres of Drosophila. *Trends Genet* **11**: 58–62.

Mason JM, Frydrychova RC, Biessmann H. 2008. Drosophila telomeres: an exception providing new insights. *Bioessays* **30**: 25–37.

Mason JM, Randall TA, Capkova Frydrychova R. 2016. Telomerase lost? *Chromosoma* **125**: 65–73.

McGurk MP, Dion-Côté A-M, Barbash DA. 2021. Rapid evolution at the Drosophila telomere: transposable element dynamics at an intrinsically unstable locus. *Genetics* **217**. http://dx.doi.org/10.1093/genetics/iyaa027.

Melnikova L, Biessmann H, Georgiev P. 2005. The Ku protein complex is involved in length regulation of Drosophila telomeres. *Genetics* **170**: 221–235.

Mohn F, Handler D, Brennecke J. 2015. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348**: 812–817.

1269 Moran N, Baumann P. 1994. Phylogenetics of cytoplasmically inherited microorganisms of arthropods.
1270 *Trends Ecol Evol* **9**: 15–20.

1271 Moran NA, Munson MA, Baumann P, Ishikawa H. 1993. A molecular clock in endosymbiotic bacteria is
1272 calibrated using the insect hosts. *Proc Biol Sci* **253**: 167–171.

1273 Morgunova V, Kordyukova M, Mikhaleva EA, Butenko I, Pobeguts OV, Kalmykova A. 2021. Loss of
1274 telomere silencing is accompanied by dysfunction of Polo kinase and centrosomes during
1275 Drosophila oogenesis and early development. *PLoS One* **16**: e0258156.

1276 Murano K, Iwasaki YW, Ishizu H, Mashiko A, Shibuya A, Kondo S, Adachi S, Suzuki S, Saito K, Natsume
1277 T, et al. 2019. Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional
1278 silencing. *EMBO J* **38**: e102870.

1279 Muyle A, Seymour D, Darzentas N, Primetis E, Gaut BS, Bousios A. 2021. Gene capture by transposable
1280 elements leads to epigenetic conflict in maize. *Mol Plant* **14**: 237–252.

1281 Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, Mandáková T, Jamge B,
1282 Lambing C, Kuo P, et al. 2021. The genetic and epigenetic landscape of the Arabidopsis
1283 centromeres. *Science* **374**: eabi7489.

1284 Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR.
1285 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**: 955–
1286 959.

1287 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic
1288 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274.

1289 Noé L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res*
1290 **33**: W540-3.

1291 O'Grady PM, DeSalle R. 2018. Phylogeny of the genus Drosophila. *Genetics* **209**: 1–25.

1292 Ohno S. 1970. *Evolution by gene duplication*. Springer, Berlin, Germany
1293 https://link.springer.com/book/10.1007/978-3-642-86659-3 (Accessed October 11, 2024).

1294 Olovnikov AM. 1973. A theory of marginotomy. The incomplete copying of template margin in enzymic
1295 synthesis of polynucleotides and biological significance of the phenomenon. *J Theor Biol* **41**:
1296 181–190.

1297 Parhad SS, Theurkauf WE. 2019. Rapid evolution and conserved function of the piRNA pathway. *Open*
1298 *Biol* **9**: 180181.

1299 Paris M, Boyer R, Jaenichen R, Wolf J, Karageorgi M, Green J, Cagnon M, Parinello H, Estoup A, Gautier
1300 M, et al. 2020. Near-chromosome level genome assembly of the fruit pest *Drosophila suzukii*
1301 using long-read sequencing. *Sci Rep* **10**: 11227.

1302 Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge
1303 A, Colwell L, et al. 2023. InterPro in 2022. *Nucleic Acids Res* **51**: D418–D427.

1304 Perrini B, Piacentini L, Fanti L, Altieri F, Chichiarelli S, Berloco M, Turano C, Ferraro A, Pimpinelli S.
1305 2004. HP1 controls telomere capping, telomere elongation, and telomere silencing by two
1306 different mechanisms in Drosophila. *Mol Cell* **15**: 467–476.

1307 Pickeral OK, Makałowski W, Boguski MS, Boeke JD. 2000. Frequent human genomic DNA transduction
1308 driven by LINE-1 retrotransposition. *Genome Res* **10**: 411–415.

1309 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
1310 *Bioinformatics* **26**: 841–842.

1311 Raffa GD, Ciapponi L, Cenci G, Gatti M. 2011. Terminin: a protein complex that mediates epigenetic
1312 maintenance of Drosophila telomeres. *Nucleus* **2**: 383–391.

1313 Raffa GD, Siriaco G, Cugusi S, Ciapponi L, Cenci G, Wojcik E, Gatti M. 2009. The Drosophila modigliani
1314 (moi) gene encodes a HOAP-interacting protein required for telomere protection. *Proc Natl Acad*

Sci U S A **106**: 2271–2276.

Rahnama M, Novikova O, Starnes JH, Zhang S, Chen L, Farman ML. 2020. Transposon-mediated telomere destabilization: a driver of genome evolution in the blast fungus. *Nucleic Acids Res* **48**: 7197–7217.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the alignment of Coding Sequences accounting for frameshifts and stop codons. *Mol Biol Evol* **35**: 2582–2584.

Remmert M, Biegert A, Hauser A, Söding J. 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**: 173–175.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**: 460.

Ryazansky S, Radion E, Mironova A, Akulenko N, Abramov Y, Morgunova V, Kordyukova MY, Olovnikov I, Kalmykova A. 2017. Natural variation of piRNA expression affects immunity to transposable elements. *PLoS Genet* **13**: e1006731.

Saint-Leandre B, Christopher C, Levine MT. 2020. Adaptive evolution of an essential telomere protein restricts telomeric retrotransposons. *Elife* **9**. http://dx.doi.org/10.7554/eLife.60987.

Saint-Leandre B, Levine MT. 2020. The telomere paradox: Stable genome preservation with rapidly evolving proteins. *Trends Genet* **36**: 232–242.

Saint-Leandre B, Nguyen SC, Levine MT. 2019. Diversification and collapse of a telomere elongation mechanism. *Genome Res* **29**: 920–931.

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**: 2214–2222.

Sasaki T, Kato K, Hosaka A, Fu Y, Toyoda A, Fujiyama A, Tarutani Y, Kakutani T. 2023. Arms race

between anti-silencing and RdDM in noncoding regions of transposable elements. *EMBO Rep* e56678.

Sasaki T, Ro K, Caillieux E, Manabe R, Bohl-Viallefond G, Baduel P, Colot V, Kakutani T, Quadrana L. 2022. Fast co-evolution of anti-silencing systems shapes the invasiveness of Mu-like DNA transposons in eudicots. *EMBO J* **41**: e110070.

Savitsky M, Kravchuk O, Melnikova L, Georgiev P. 2002. Heterochromatin protein 1 is involved in control of telomere elongation in *Drosophila melanogaster*. *Mol Cell Biol* **22**: 3204–3218.

Savitsky M, Kwon D, Georgiev P, Kalmykova A, Gvozdev V. 2006. Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. *Genes Dev* **20**: 345–354.

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

She R, Chu JS-C, Uyar B, Wang J, Wang K, Chen N. 2011. genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics* **27**: 2141–2143.

Shimada A, Cahn J, Ernst E, Lynn J, Grimanelli D, Henderson I, Kakutani T, Martienssen RA. 2024. Retrotransposon addiction promotes centromere function via epigenetically activated small RNAs. *Nat Plants* **10**: 1304–1316.

Shpiz S, Kwon D, Uneva A, Kim M, Klenov M, Rozovsky Y, Georgiev P, Savitsky M, Kalmykova A. 2007. Characterization of Drosophila telomeric retroelement TAHRE: transcription, transpositions, and RNAi-based regulation of expression. *Mol Biol Evol* **24**: 2535–2545.

Shpiz S, Olovnikov I, Sergeeva A, Lavrov S, Abramov Y, Savitsky M, Kalmykova A. 2011. Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. *Nucleic Acids Res* **39**: 8703–8711.

Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A. 2020. ClipKIT: A multiple sequence alignment

trimming software for accurate phylogenomic inference. *PLoS Biol* **18**: e3001007.

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* **375**: 20190347.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 Drosophila genomes. *Curr Biol* **32**: 111-123.e5.

Tambones IL, Haudry A, Simão MC, Carareto CMA. 2019. High frequency of horizontal transfer in Jockey families (LINE order) of drosophilids. *Mob DNA* **10**: 43.

Tiwari MD, Zeitler DM, Meister G, Wodarz A. 2019. Molecular profiling of stem cell-like female germ line cells in Drosophila delineates networks important for stemness and differentiation. *Biol Open* **8**: bio046789.

Torosin NS, Anand A, Golla TR, Cao W, Ellison CE. 2020. 3D genome evolution and reorganization in the *Drosophila melanogaster* species group. *PLoS Genet* **16**: e1009229.

Uhrig S, Klein H. 2019. PingPongPro: a tool for the detection of piRNA-mediated transposon-silencing in small RNA-Seq data. *Bioinformatics* **35**: 335–336.

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.

Venner S, Miele V, Terzian C, Biémont C, Daubin V, Feschotte C, Pontier D. 2017. Ecological networks to unravel the routes to horizontal transposon transfers. *PLoS Biol* **15**: e2001536.

Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE, de Pablos B. 2007. Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res* **17**: 1909–1918.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J,

Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.

Walter MF, Biessmann MR, Benitez C, Török T, Mason JM, Biessmann H. 2007. Effects of telomere length in Drosophila melanogaster on life span, fecundity, and fertility. *Chromosoma* **116**: 41–51.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**: 77–80.

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

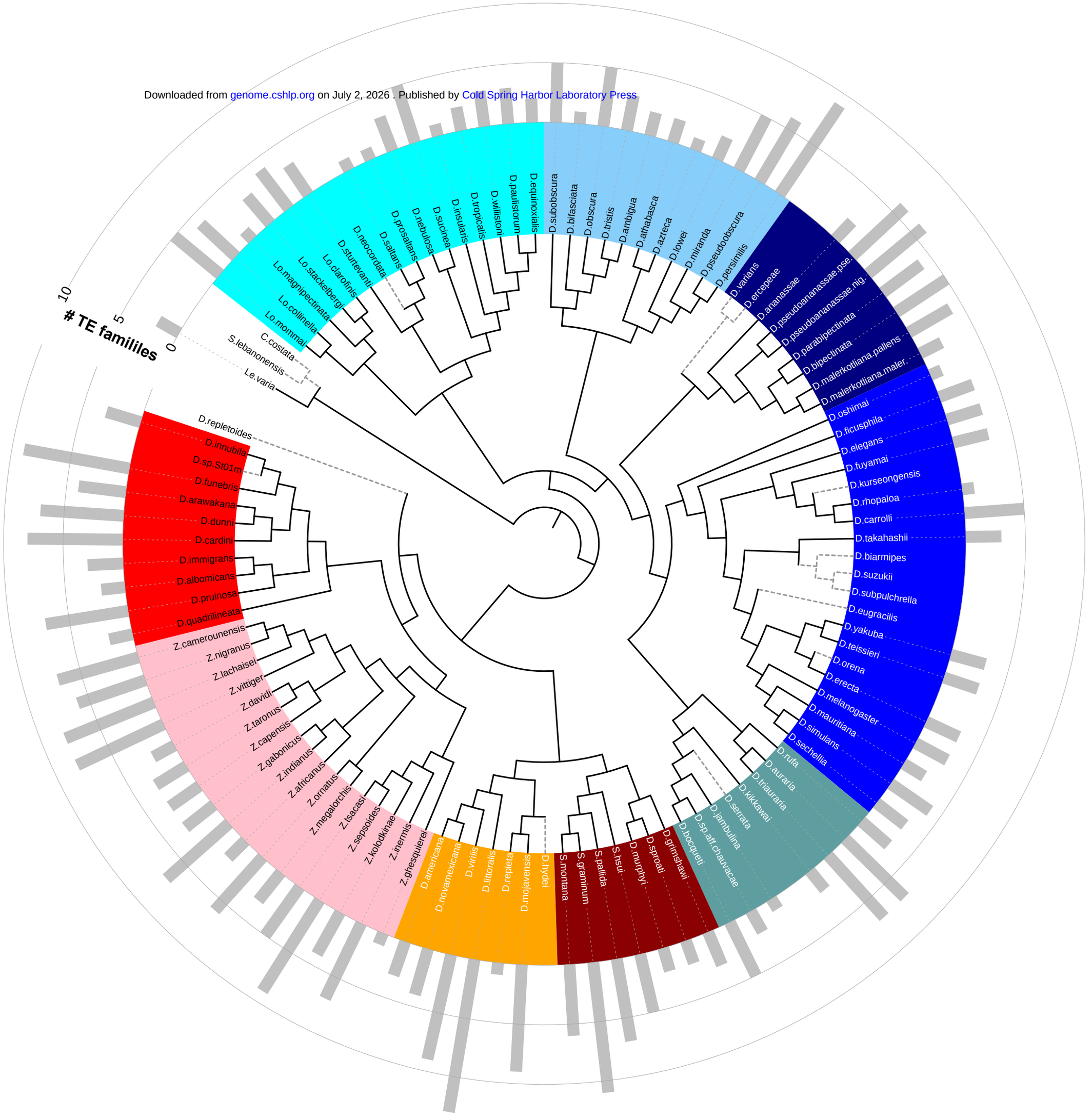
Wickersheim ML, Blumenstiel JP. 2013. Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. *Biotechniques* **55**: 269–272.

Yang J, Malik HS, Eickbush TH. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc Natl Acad Sci U S A* **96**: 7847–7852.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

Zhao K, Cheng S, Miao N, Xu P, Lu X, Zhang Y, Wang M, Ouyang X, Yuan X, Liu W, et al. 2019. A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. *Nat Cell Biol* **21**: 1261–1272.

Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. *J Mol Biol* **430**: 2237–2243.



clade1
Lordiphosa, saltans, willistoni

clade2
obscura

clade3
ananassae

clade4
melanogaster

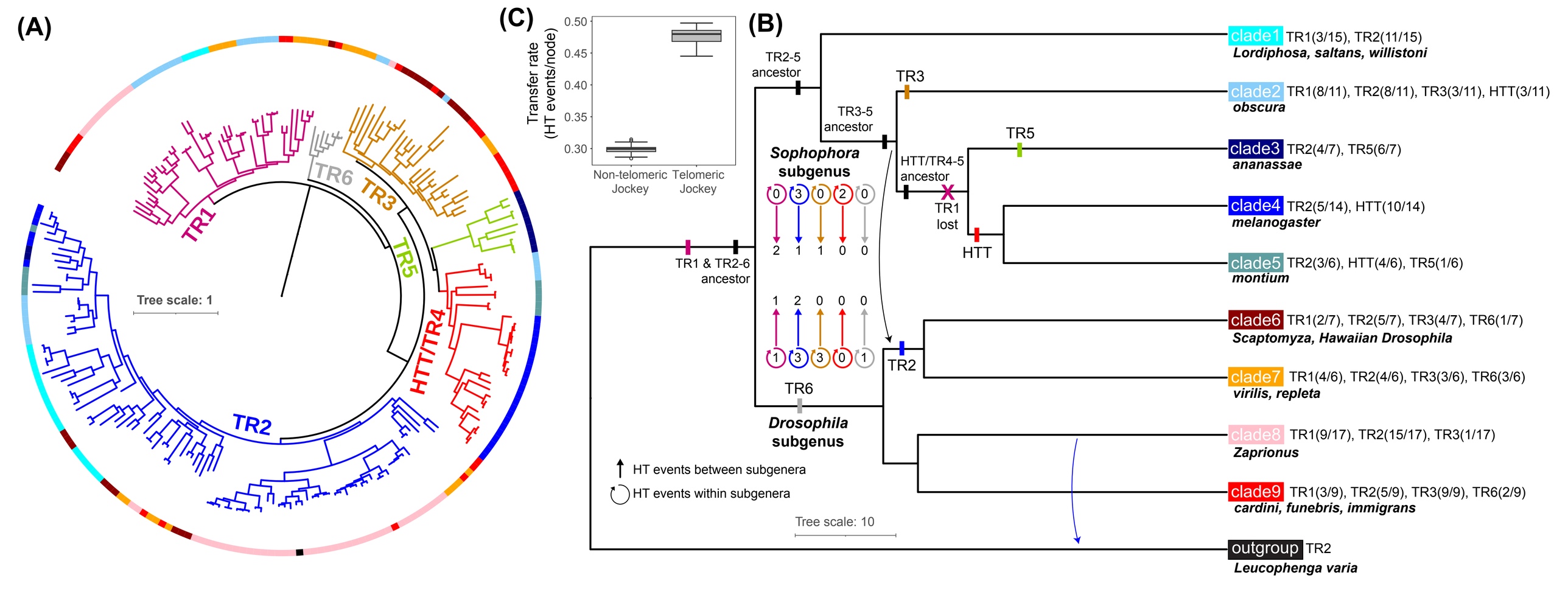
clade5
montium

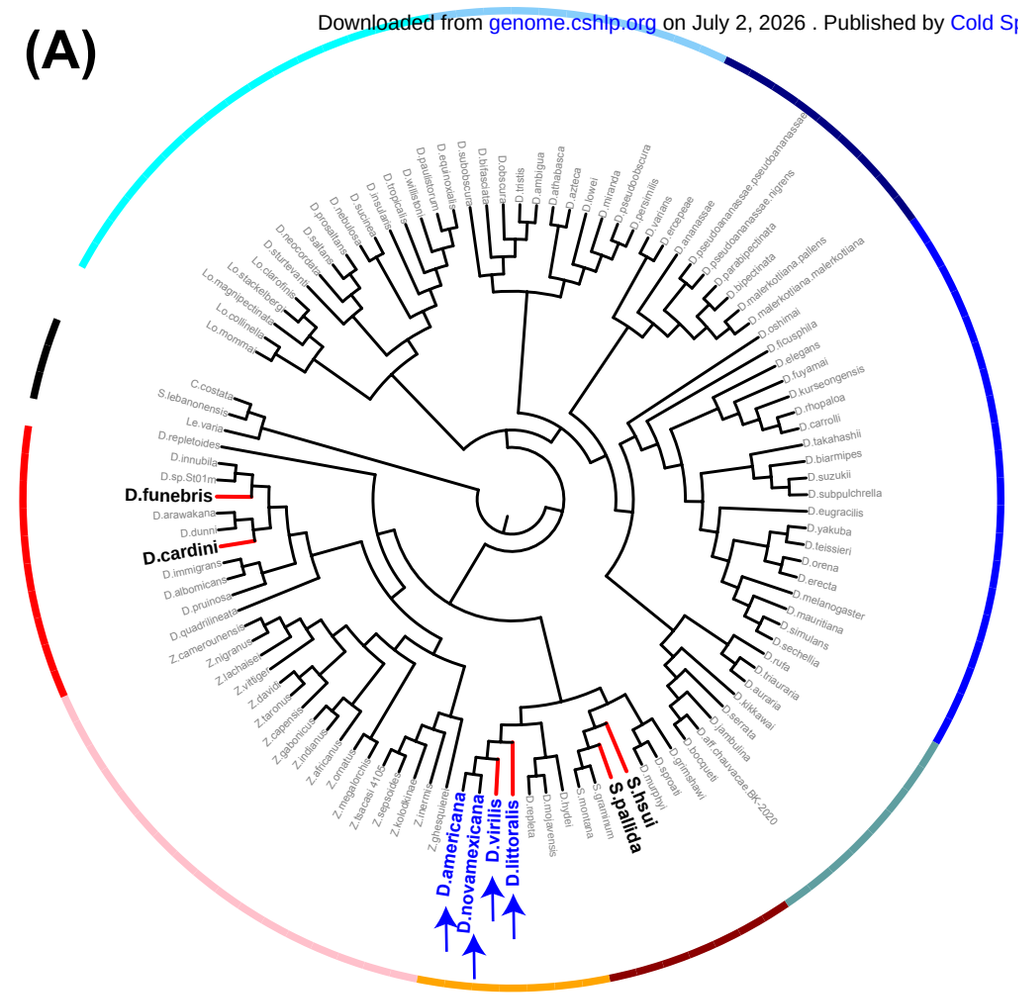
clade6
Scaptomyza, Hawaiian Drosophila

clade7
virilis, repleta

clade8
Zaprionus

clade9
cardini, funebris, immigrans


(A)
TR1
TR6
TR3
TR5
HTT/TR4
TR2
Tree scale: 1
(C)
Transfer rate
(HT events/node)
0.50
0.45
0.40
0.35
0.30
Non-telomeric Jockey
Telomeric Jockey
(B)
TR2-5 ancestor
TR3
TR3-5 ancestor
TR5
HTT/TR4-5 ancestor
TR1 lost
HTT
Sophophora subgenus
0 3 0 2 0
2 1 1 0 0
TR1 & TR2-6 ancestor
1 2 0 0 0
1 3 3 0 1
TR2
TR6
Drosophila subgenus
HT events between subgenera
HT events within subgenera
Tree scale: 10
clade1 TR1(3/15), TR2(11/15)
Lordiphosa, saltans, willistoni
clade2 TR1(8/11), TR2(8/11), TR3(3/11), HTT(3/11)
obscura
clade3 TR2(4/7), TR5(6/7)
ananassae
clade4 TR2(5/14), HTT(10/14)
melanogaster
clade5 TR2(3/6), HTT(4/6), TR5(1/6)
montium
clade6 TR1(2/7), TR2(5/7), TR3(4/7), TR6(1/7)
Scaptomyza, Hawaiian Drosophila
clade7 TR1(4/6), TR2(4/6), TR3(3/6), TR6(3/6)
virilis, repleta
clade8 TR1(9/17), TR2(15/17), TR3(1/17)
Zaprionus
clade9 TR1(3/9), TR2(5/9), TR3(9/9), TR6(2/9)
cardini, funebris, immigrans
outgroup TR2
Leucophenga varia



clade1
Lordiphosa, saltans, willistoni

clade2
obscura

clade3
ananassae

clade4
melanogaster

clade5
montium

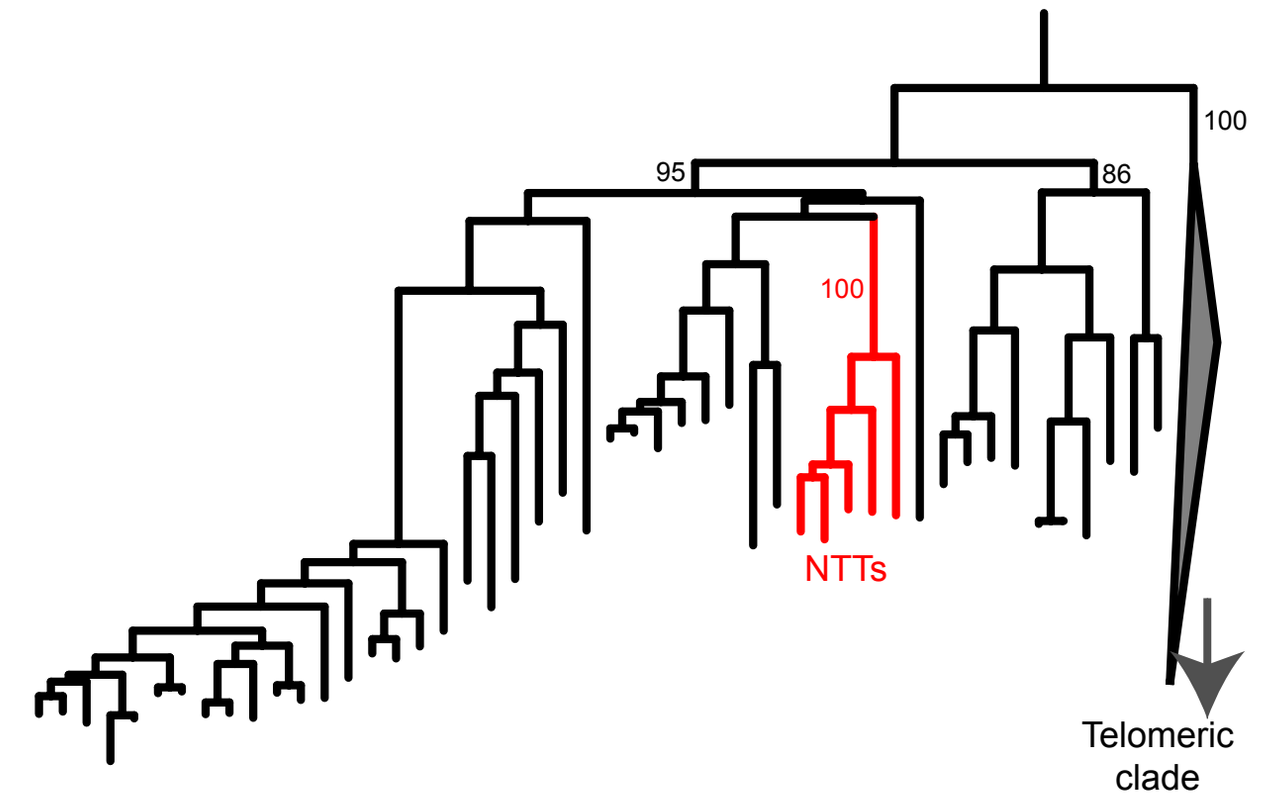
clade6
Scaptomyza, Hawaiian Drosophila

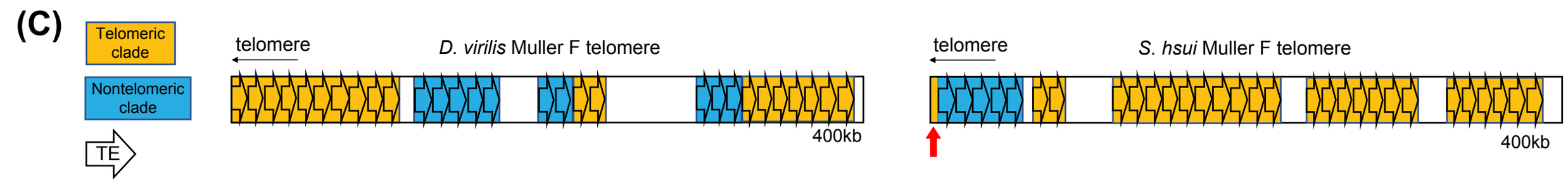
clade7
virilis, repleta

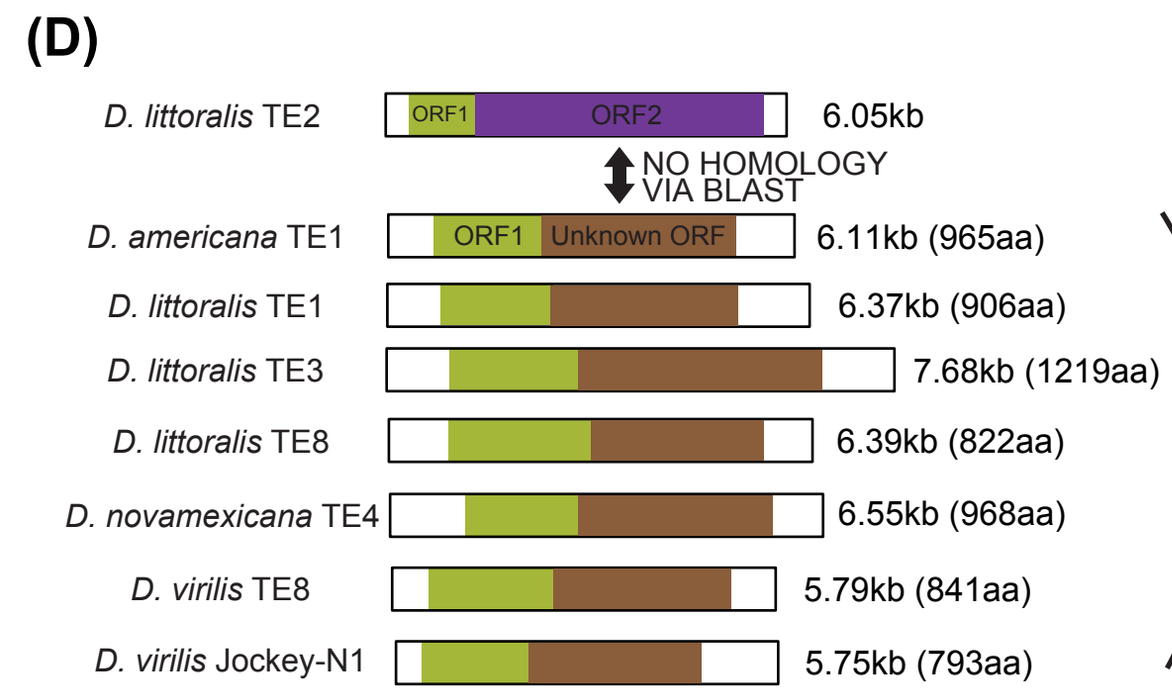
clade8
Zaprionus

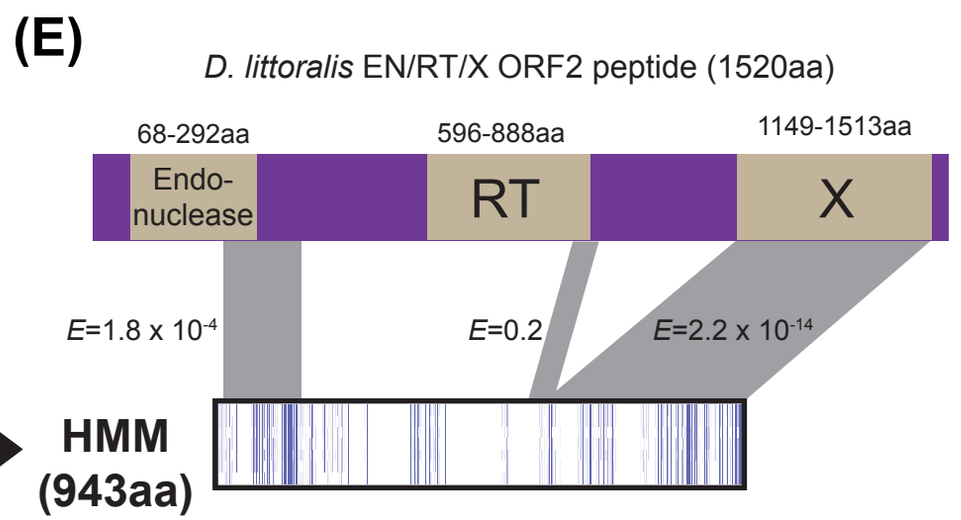
clade9
cardini, funebris, immigrans

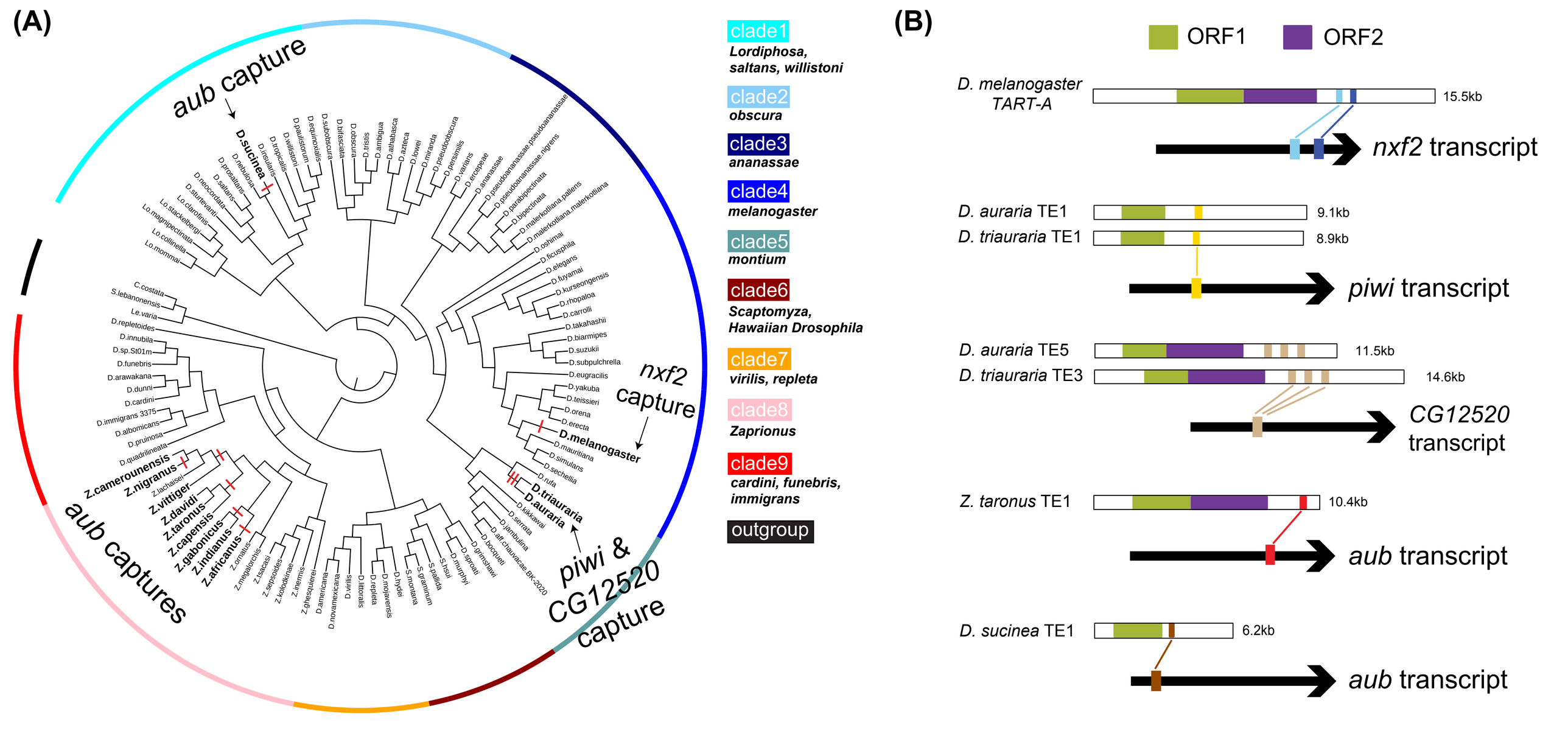
outgroup










(A)
aub capture
nxf2 capture
piwi & CG12520 capture
aub captures
clade1
Lordiphosa, saltans, willistoni
clade2
obscura
clade3
ananassae
clade4
melanogaster
clade5
montium
clade6
Scaptomyza, Hawaiian Drosophila
clade7
virilis, repleta
clade8
Zaprionus
clade9
cardini, funebris, immigrans
outgroup
(B)
ORF1
ORF2
D. melanogaster TART-A
15.5kb
nxf2 transcript
D. auraria TE1
9.1kb
D. triauraria TE1
8.9kb
piwi transcript
D. auraria TE5
11.5kb
D. triauraria TE3
14.6kb
CG12520 transcript
Z. taronus TE1
10.4kb
aub transcript
D. sucinea TE1
6.2kb
aub transcript