Diversification and collapse of a telomere elongation mechanism

Bastien Saint-Leandre[1,2], Son C. Nguyen[2,3], and Mia T. Levine[1,2*]

[1]Department of Biology, [2]Epigenetics Institute, University of Pennsylvania, Philadelphia, PA
[3]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104.

Running head: diversification of Drosophila telomeres

Keywords: telomere, retrotransposon, intra-genomic conflict, Drosophila

[*]Corresponding Author:
Mia Levine, Department of Biology, University of Pennsylvania, 433 S. University Avenue, Philadelphia, PA 19104. p: 215-573-9709 e: m.levine@sas.upenn.edu.

ABSTRACT

In virtually all eukaryotes, telomerase counteracts chromosome erosion by adding repetitive sequence to terminal ends. *Drosophila melanogaster* instead relies on specialized retrotransposons that insert preferentially at telomeres. This exchange of goods between host and mobile element—wherein the mobile element provides an essential genome service and the host provides a hospitable niche for mobile element propagation—has been called a 'genomic symbiosis'. However, these telomere-specialized, 'jockey' family elements may actually evolve to selfishly over-replicate in the genomes that they ostensibly serve. Under this intra-genomic conflict model, we expect rapid diversification of telomere-specialized retrotransposon lineages and possibly, the breakdown of this tenuous relationship. Here we report data consistent with both predictions. Searching the raw reads of the 15-million-year-old 'melanogaster species group', we generated *de novo* jockey retrotransposon consensus sequences and used phylogenetic tree-building to delineate four distinct telomere-associated lineages. Recurrent gains, losses, and replacements account for this striking retrotransposon lineage diversity. Moreover, an ancestrally telomere-specialized element has 'escaped,' residing now throughout the genome of *D. rhopaloa*. In *D. biarmipes*, telomere-specialized elements have disappeared completely. *De novo* assembly of long-reads and cytogenetics confirmed this species-specific collapse of retrotransposon-dependent telomere elongation. Instead, telomere-restricted satellite DNA and DNA transposon fragments occupy its terminal ends. We infer that *D. biarmipes* relies instead on a recombination-based mechanism conserved from yeast to flies to humans. Combined with previous reports of adaptive evolution at host proteins that regulate telomere length, telomere-associated retrotransposon diversification and disappearance offer compelling evidence that intra-genomic conflict shapes Drosophila telomere evolution.

INTRODUCTION

Transposable elements infest eukaryotic genomes, ever-evolving to increase in copy number over time (Feschotte and Pritham 2007; Beauregard et al. 2008). These 'selfish elements' compromise host fitness by disrupting host genes (Hancks and Kazazian 2012), nucleating local epigenetic silencing (Slotkin and Martienssen 2007; Lee and Karpen 2017), and triggering catastrophic recombination between non-homologous genomic regions (Langley et al. 1988; Beck et al. 2011). Transposable elements ('TEs') also provide raw material for genome adaptation (Jangam et al. 2017): across eukaryotes, host genomes re-purpose TE-derived sequence for basic cellular and developmental processes, from immune response (van de Lagemaat et al. 2003) to placental development (Lynch et al. 2015) to programmed genome rearrangements (Cheng et al. 2010; Cheng et al. 2016). These diverse 'molecular domestication' events share a common feature—the degeneration or deletion of the TE's capacity to propagate. Consequently, the TE-derived sequence resides permanently at a single genome location, just like any other host gene sequence. Inability to selfishly over-replicate resolves prior conflict of interest between the host and the TE (Jangam et al. 2017). However, not all adaptive molecular domestication events necessitate TE immobilization; in rare cases, essential host functions rely on retention of the mobilization machinery that promotes recurrent TE insertions into host DNA. The non-canonical telomere elongation mechanism of Drosophila is exemplary (Pardue and DeBaryshe 2003; Casacuberta 2017).

In virtually all eukaryotes beyond Drosophila, telomerase-added DNA repeats (Greider and Blackburn 1989; Zakian 1989; Blackburn 1991; Zakian 1996) counteract the 'end-replication problem' that otherwise erodes unique DNA sequence at chromosome termini (Watson 1972). However, telomeres of select fungi (Starnes et al. 2012), algae (Higashiyama et al. 1997), moths (Osanai-Futahashi and Fujiwara 2011), crustaceans (Gladyshev and Arkhipova 2007),

86 and DNA repair-deficient mammalian cells (Morrish et al. 2007) encode not only telomerase-
87 added repeat elements but also TEs that insert preferentially at chromosome termini. These
88 telomeric mobile elements are typically derived from a single class of TE—the non-long terminal
89 repeat ("non-LTR") retrotransposons (Beck et al. 2011), which mobilize via reverse transcription
90 and insertion into new genomic locations. The most extreme example of this co-option is found
91 in *Drosophila melanogaster*, whose telomeres harbor no telomerase-added repeats. In fact, the
92 220 million-year-old 'true fly' insect Order, Diptera, completely lacks the genes encoding the
93 telomerase holoenzyme (Pardue and DeBaryshe 2003; Casacuberta 2017). Instead, *D.*
94 *melanogaster* harbors three highly specialized retrotransposons–HeT-A, TART, and TAHRE—
95 charged with the important job of preserving distal, unique sequence (Pardue and DeBaryshe
96 2011). These telomere-specialized retrotransposons represent a monophyletic clade within the
97 larger 'jockey' element family (Villasante et al. 2007), whose members are more typically found
98 along chromosome arms (Xie et al. 2013). The telomere-specialized jockey subclade, in
99 contrast, rarely inserts outside the telomere (Pardue and DeBaryshe 2003; Berloco et al. 2005;
100 Pardue and DeBaryshe 2011).
101
102 This molecular domestication of still-mobile retrotransposons into an essential genome function
103 is often referred to as a genomic 'symbiosis' (Pardue and DeBaryshe 2008). Evidence for such
104 a mutualism is compelling. The elements that maintain telomere ends in *D. melanogaster*
105 comprise a monophyletic clade ostensibly specialized to replicate only at chromosome ends
106 (Pardue and DeBaryshe 2008). Elements from this jockey clade also appear at terminal ends in
107 distant Drosophila species, consistent with a single domestication event >40 million years ago
108 followed by faithful vertical transmission (Casacuberta and Pardue 2003; Villasante et al. 2007).
109 Moreover, no other mobile element family appears at terminal ends, consistent with host-
110 dependent molecular mechanisms that block 'generalist' TEs from invading this vital genome
111 region (Mason and Biessmann 1995; Mason et al. 2016). These data suggest that

112 retrotransposon-mediated chromosome elongation represents a long-term, conserved
113 relationship between a host genome and its domesticated, but still-mobile, retrotransposons.
114
115 This picture of cooperativity was complicated by the discovery that the telomere-associated
116 subclade of jockey elements evolves rapidly across Drosophila. Leveraging 12 Drosophila
117 genomes that span 40 million years of evolution, Villasante, Abad, and colleagues detected at
118 least one jockey-like, candidate telomeric retrotransposon in all 12 species (Villasante et al.
119 2007). These data implicated a single evolutionary event in a common ancestral sequence that
120 conferred telomere-specificity. However, these elements represent distinct phylogenetic
121 lineages rather than species-specific versions of the HeT-A, TART, and TAHRE elements well-
122 studied in *D. melanogaster.* For example, *D. melanogaster* and its close relatives encode HeT-
123 A, TART, and TAHRE, while the 20 million year-diverged *D. ananassae* encodes a single,
124 phylogenetically distinct candidate retrotransposon lineage, 'TR2'. The 30 million-year diverged
125 *D. pseudoobscura* species encodes yet other phylogenetically distinct lineages within this
126 jockey subclade. This expansive evolutionary lens revealed a previously unappreciated,
127 dynamic evolutionary history of these jockey subclade retrotransposons. However, the large
128 evolutionary distances between species left fine scale dynamics unknown and telomere-specific
129 localization was not explored (Villasante et al. 2007). The evolutionary origin(s), ages, between-
130 species differences, and genome locations of these candidate length regulators remain obscure.
131 Elucidating the evolutionary history of these elements is essential to address the possibility that
132 this molecular domestication event is less a stable, long-term genomic symbiosis and instead a
133 tenuous relationship between the domesticator and the domesticated.
134
135 Our recent report that pervasive positive selection shapes Drosophila telomere proteins—
136 especially genes that regulate chromosome length (Lee et al. 2017)—raises the intriguing
137 possibility that telomeric retrotransposons recurrently evolve to escape host control. Under this

138 model, domesticated retrotransposons evolve to increase copy number either by excessively
139 lengthening chromosome ends or by invading non-telomeric locations. Host dependency on still-
140 mobile TEs to perform an essential genome function may, in fact, be an unstable telomere
141 elongation strategy. To evaluate this possibility, we set out to resolve the finer scale
142 evolutionary history of telomere-specialized elements. Specifically, we investigated the
143 'melanogaster species group', which captures an informative three-15 million years of evolution
144 (Chen et al. 2014). This clade spans the distance between HeT-A-, TART-, and TARHE-
145 encoding *D. melanogaster* and its close relatives and *D. ananassae,* which encodes the distant
146 TR2 lineage not yet localized to telomeres (Danilevskaya et al. 1998; Casacuberta and Pardue
147 2002; Villasante et al. 2007). We probed the raw reads of melanogaster species group genomes
148 and used *de novo* consensus building and experimental validation to define young and ancient
149 telomere-associated retrotransposons from this jockey subclade. We uncovered diversification
150 via gain and loss, both with and without replacement, of telomere-associated, retrotransposon
151 lineages across three to 15 million years of evolution. We also observe dramatic differences in
152 telomere-specialized retrotransposon copy number across these closely related species,
153 consistent with chromosome length evolving under minimal host constraint. In the most extreme
154 case, we identified a species that has undergone wholesale loss of active, telomere-specialized
155 retrotransposons. Collapse of this telomere elongation mechanism highlights the potential
156 hazards of depending on still-mobile elements to perform an essential genome service.
157
158 **RESULTS**
159
160 *Candidate telomere-specialized retrotransposons identified in the melanogaster species group*
161
162 *D. melanogaster* encodes telomere-specialized, non-LTR retrotransposons that increase in copy
163 number by a "copy and paste" mechanism (Pardue and DeBaryshe 2011). Transcripts encoded

164 by these elements are localized, reverse transcribed, and integrated at the terminal nucleotides
165 of chromosome ends, resulting in stereotypical head-to-tail arrays (Pardue and DeBaryshe
166 2003). Full length, autonomous retrotransposons typically contain two open reading frames
167 (ORFs) between the variable 5' and 3' UTRs. ORF1 encodes an RNA binding domain ('Gag')
168 and ORF2 encodes a reverse transcriptase domain ("RT") and an endonuclease domain ("EN",
169 Fig. S1). The telomere-specialized HeT-A (ORF1 only), TART, and TAHRE elements form a
170 monophyletic clade within the jockey family of non-LTR retrotransposons (Villasante et al.
171 2007).
172
173 To elucidate the fine scale evolutionary history of telomere-specialized elements in Drosophila,
174 we searched for jockey-like, telomere-specialized retrotransposons across lineages that span
175 three-15 million years of evolution (Drosophila 12 Genomes et al. 2007; Chen et al. 2014). *D.*
176 *melanogaster* and its close relatives (the 'melanogaster subgroup,' including *D. melanogaster,*
177 *D. simulans, D. sechellia, D. yakuba*) share the well-studied retrotransposon lineages HeT-A,
178 TART, and TAHRE (Danilevskaya et al. 1998; Casacuberta and Pardue 2002; Berloco et al.
179 2005; Villasante et al. 2007) and so offer limited traction for capturing evolutionary turnover
180 events of major telomeric retrotransposon lineages. Beyond these close relatives, previous work
181 defined a candidate telomeric retrotransposon from the jockey family in the 15-million year-
182 diverged *D. ananassae* (Villasante et al. 2007). This element, however, represents a lineage
183 distinct from HeT-A, TART, and TAHRE and has yet to be validated cytogenetically as telomere-
184 specialized. To gain both sufficient divergence to reveal telomeric retrotransposon turnover and
185 sufficient resolution to estimate retrotransposon lineage birth and death, we focused on an
186 intermediate evolutionary distance from *D. melanogaster* by including other members of the
187 'melanogaster species group'. Five focal species—*D. rhopaloa, D. biarmipes, D. takahsashii, D.*
188 *elegans* and *D. ficusphila* — span 10 million years of divergence and share a common ancestor
189 with *D. melanogaster* <15 million years ago (Chen et al. 2014). The genomes of *D.*

190 *melanogaster* and its three close relatives served as positive controls for our *de novo*
191 identification of retrotransposons phylogenetically related to the jockey subclade specialized at
192 telomeres. *D. ananassae* served as an outgroup.
193
194 We developed a custom pipeline to discover jockey family elements ancestrally related to
195 previously defined lineages that maintain chromosome ends (Fig. S2). Briefly, we conducted
196 tBLASTn searches against raw reads from each of the ten species using a query that included
197 both fully and minimally validated telomere-specialized retrotransposons from across Drosophila
198 (Table S1). We also included the reference non-telomeric jockey element (from *D.*
199 *melanogaster*) defined in Repbase (www.girinst.org/Repbase/, Table S1). To facilitate alignment
200 across distant lineages, we restricted our search to the Gag domain of ORF1 and the RT of
201 ORF2 (Fig. S1). These 63 Gag and 48 RT domains span 40 million years of Drosophila
202 evolution (Table S1). A detailed description of our pipeline of iterative BLAST searches to raw
203 reads, *de novo* consensus building, and phylogenetic tree-assisted sorting can be found in the
204 Materials and Methods. This pipeline (Fig. S2) generated a refined list of consensus sequences
205 that included previously described telomere-specialized elements from *D. melanogaster* and its
206 close relatives (our "positive controls") as well as jockey family elements outside the specialized
207 telomeric subclade (Table S2). These latter elements indicated that our search was
208 exhaustive—we effectively overshot the jockey subclade associated with telomeres for all
209 species. We built Bayesian phylogenetic trees based on the Gag and RT domains of all final
210 consensus sequences (Fig. 1 and Fig. S3). Our Gag-based trees revealed that the 15
211 consensus sequences form a well-supported, monophyletic subclade within the jockey family
212 but distinct from 18 generalist jockey element consensus sequences that too form a distinct,
213 well-supported monophyletic clade (gray, Fig 1A). The candidate telomeric Gag consensus
214 sequences form four distinct lineages—TAHRE, TART, TR2, and a previously undefined lineage
215 that we named 'TARTAHRE' for its labile phylogenetic position between TART and TAHRE. The

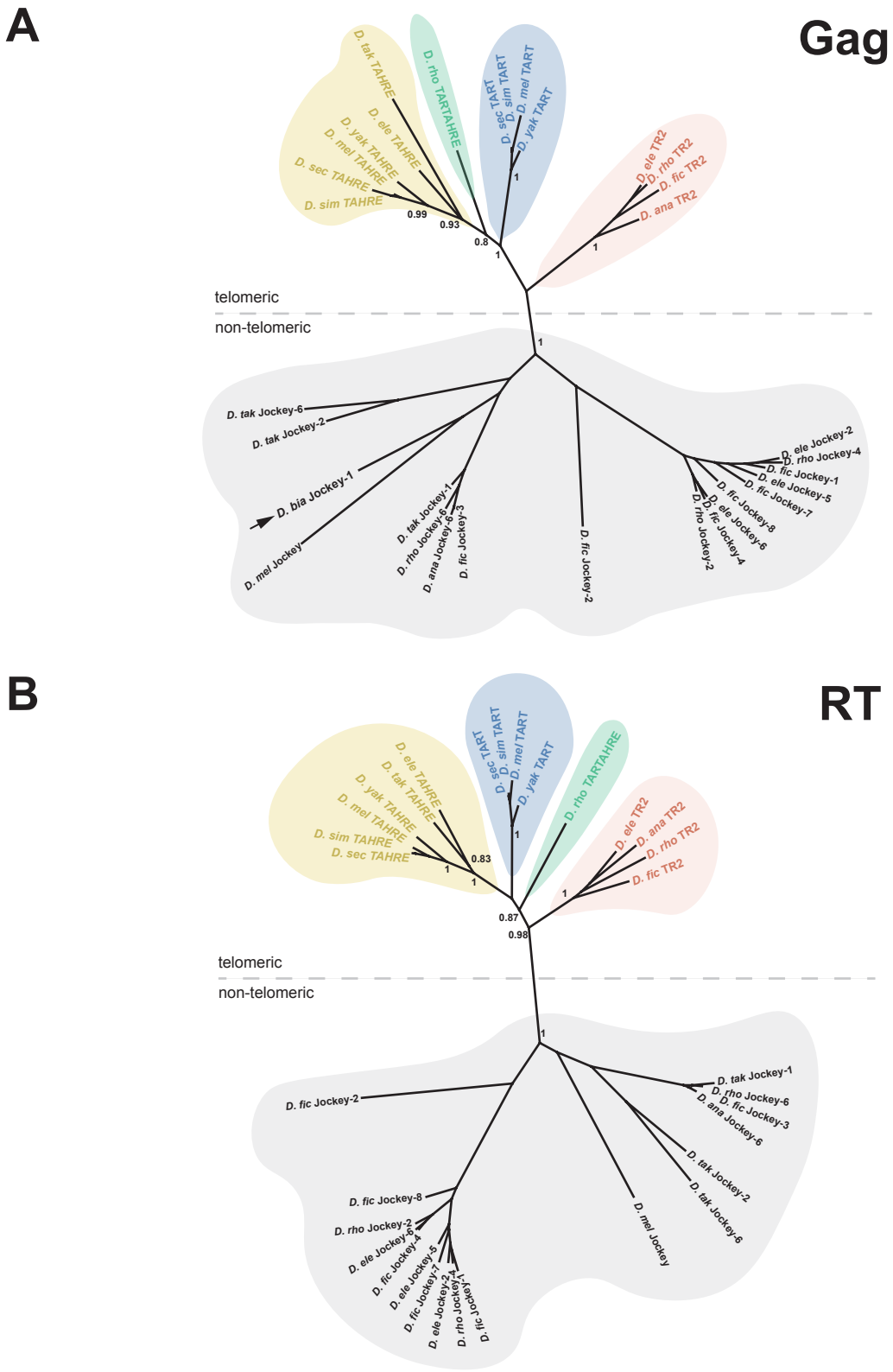

Figure 1. Phylogenetic relationships among previously and newly defined jockey subclade elements. Unrooted phylogenetic trees built from Gag domain **(A)** and RT domain **(B)** consensus sequences. Node support values are posterior probabilities generated by MrBayes. Gray designates jockey elements that passed the final pipeline filter but are distantly related to the telomere-specialized subclade. Various colors delineate candidate telomere-specialized elements along with previously characterized elements that form monophyletic clades. Only the D. rhopaloa-restricted element, TARTARHE (green), 'migrates' between TAHRE and TART clades across the two trees and may represent a chimera of the two lineages. The black arrow corresponds to the closest D. biarmipes jockey family element to the telomere-specialized subclade.

217 TR2 lineage shares a more recent common ancestor with jockey elements compared to the
218 TART/TAHRE/TARTHARE clade. Moreover, our trees support previous inferences that HeT-A
219 Gags are phylogenetically indistinguishable from TAHRE Gags ((Villasante et al. 2007),Fig.
220 S3)—HeT-A encodes its own 5'UTR, Gag, and 3'UTR (Pardue and DeBaryshe 2003), yet its
221 phylogenetic position reveals that this single-domain element is effectively a TAHRE element
222 missing an RT domain. Henceforth, we refer to this single lineage as TAHRE. The RT domain-
223 based tree (Fig. 1B) revealed strikingly similar topology. We were surprised to recover complete
224 consensus sequences of candidate telomere-specialized elements from only seven species of
225 the 10 species. Our pipeline detected only a partial TAHRE Gag in *D. takahashii* (Table S2) and
226 no evidence at all of the telomere-associated, jockey subclade elements in its close relative, *D.*
227 *biarmipes* (see below). We explore below the possibility that *D. biarmipes* telomeres are
228 functionally distinct from other sampled species.
229
230 *PCR and cytogenetic validation of computationally predicted telomeric retrotransposons*
231
232 Despite our search being inherently conservative—our consensus-building approach may
233 average out sub-lineages within major named retrotransposon lineages—we uncovered striking
234 diversity across only 15 million years of Drosophila evolution. By virtue of phylogenetic
235 relatedness, we predict that the jockey subclade retrotransposons uncovered by our pipeline
236 specialize at chromosome ends in their respective host species. Testing this prediction first
237 requires molecular biology to confirm that 1) these *in silico*-generated consensus sequences
238 represent actual elements in the targeted genomes and 2) the elements are arrayed in
239 characteristic head-to-tail orientation that arises from exclusive end-integration of the poly-
240 adenylated 3' end (Pardue and DeBaryshe 2008; Pardue and DeBaryshe 2011). Using Sanger
241 sequencing of PCR products amplified from genomic DNA, we discovered that the *in silico*
242 sequences represent true DNA elements in their host genomes (Fig. 2A).

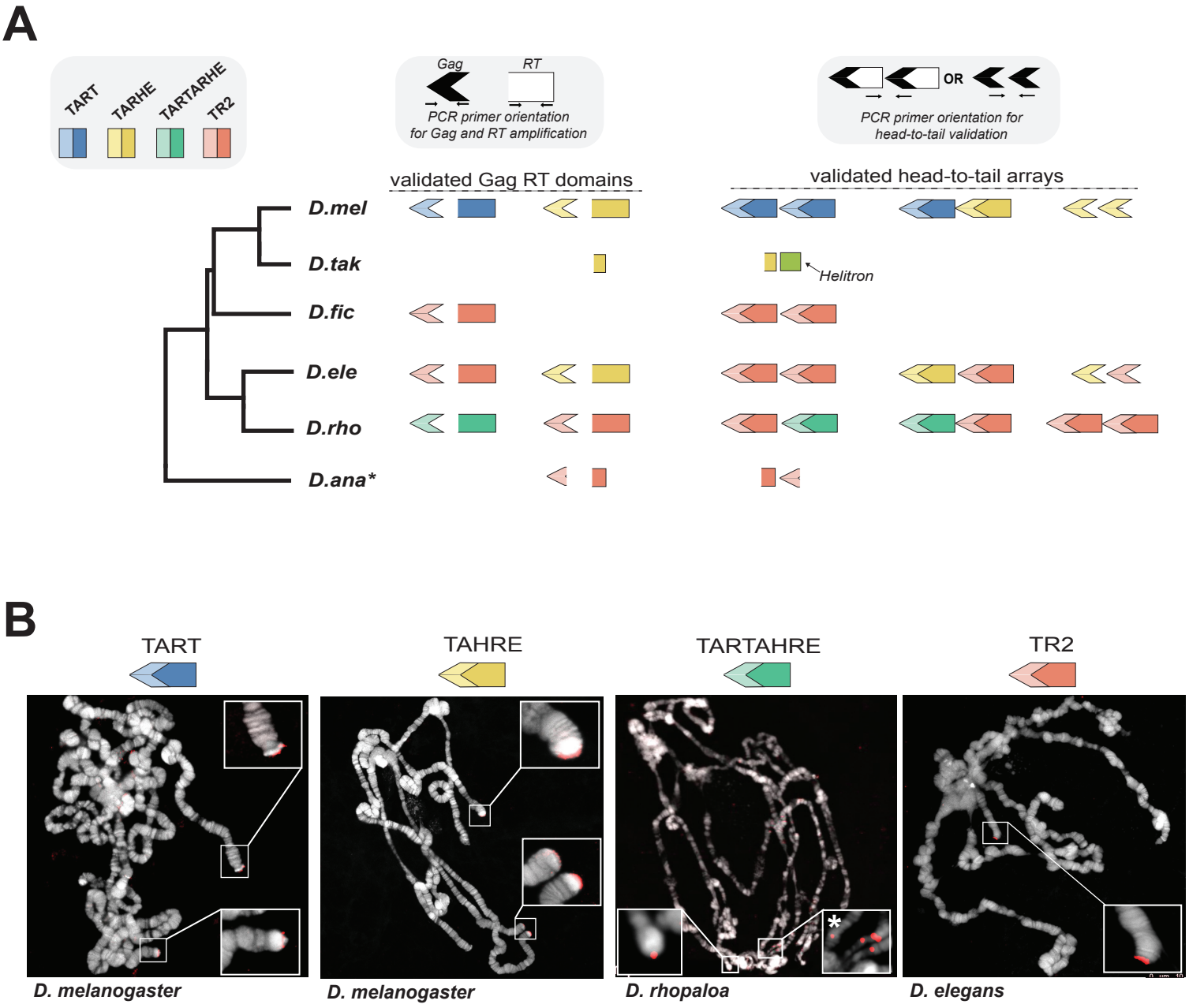

Figure 2. PCR- and cytology-based validation of in silico predicted, telomere-specialized elements. **(A)** Cartoon representation of our PCR-based validation of in silico – predicted Gag and RT domains and head–to–tail orientation of candidate telomeric retrotransposons (and the previously validated TAHRE and TART). Primer orientation to test head-to–tail positioning is represented above as cartoons in white and black. Gag (arrowhead) and RT (rectangle) domains are represented by lighter or darker shades, respectively. PCR-validated partial Gag and partial RT domains are represented as truncated symbols in *D. takahashii* and *D. ananassae*. **(B)** DNA-FISH or oligopainting with probes/paints cognate to TAHRE, TART, TARTAHRE and TR2 on polytene chromosomes from representative species. TART and TAHRE from D. melanogaster serve as positive controls. All insets show telomere hybridization exclusively except TARTARHE, which hybridized to both telomeric and non-telomeric locations (inset designated with a '*').

244

245 The PCR-amplified/Sanger-sequenced domains typically share ~97.5% sequence identity to a
246 given consensus (Table S3). For *D. takahashii* and *D. ananassae,* we successfully amplified the
247 predicted partial TAHRE and TR2 RT domains, respectively, as well as *D. ananassae's* partial
248 TR2 Gag domain. We also confirmed head-to-tail orientation using primers that annealed to the
249 3' 'tail' of one copy and the 5' 'head' of another copy predicted to reside at the same telomere
250 (Fig. 2A, Table S3). In *D. takahashii*, we successfully PCR-amplified an *in silico*-predicted
251 junction between its partial TAHRE and a Helitron. Helitrons are DNA transposons that we
252 discovered in subsequent experiments have colonized the telomeres of *D. takahashii*'s closest
253 relative in our sampled species—*D. biarmipes* (see below). PCR- and Sanger sequencing-
254 based validation suggests that virtually all consensus sequences represented in Fig. 1
255 correspond to actual jockey subclade elements found in their respective genomes and in an
256 orientation stereotypical of telomere-specialized retrotransposons. Restriction to telomere ends,
257 however, has only been demonstrated previously for the TAHRE and TART lineages.
258

259 To investigate chromosome localization of the newly defined elements, TR2 and TARTAHRE,
260 we conducted DNA FISH or 'oligopainting' on polytene chromosomes in representative species.
261 TAHRE/HeT-A and TART probes in *D. melanogaster* served as a positive control. Like the
262 striking telomere-specialization of TAHRE/HeT-A and TART in *D. melanogaster*, the TR2 probe
263 hybridized exclusively to telomere ends (Fig. 2B). However, TARTAHRE probe hybridization
264 revealed *both* telomere localization and pericentromeric localization, as well as localization at
265 other euchromatic sites (Fig. 2B). The promiscuous localization of TARTAHRE implicates either
266 incomplete domestication or escape from domestication. Its nested phylogenetic position within
267 a clade of telomere-specialized elements favors an innovation event leading to "escape" from
268 the telomere (see Discussion).
269

270

271 *Recurrent turnover of telomere-specialized retrotransposons in the melanogaster species group*

272

273 To infer telomeric retrotransposon lineage turnover across the melanogaster species group, we
274 summarized presence/absence of these validated elements across the species tree (Fig. 3A,
275 (Chen et al. 2014)). The TART lineage, well-known from *D. melanogaster*, is relatively young,
276 emerging in the ancestor of the 'subgroup' between 10 and 15 million years ago. The TAHRE
277 lineage is more ancient, emerging either along the lineage leading to the melanogaster species
278 group or instead before the common ancestor of the melanogaster species group and *D.*
279 *ananassae* (and subsequently lost along the *D. ananassae* lineage). TAHRE has been lost at
280 least three times, along lineages leading to *D. rhopaloa, D. ficusphila* and *D. biarmipes. D.*
281 *biarmipes'* closest relative on the tree, *D. takahashii,* encodes only a truncated TAHRE. These
282 two species together suggest that TAHRE loss began in their common ancestor, though only in
283 *D. biarmipes* is the loss event complete. The 'generalist' TARTARHE lineage appears in *D.*
284 *rhopaloa* only, making it the only jockey subclade element restricted to a single species on the
285 densely sampled, melanogaster species group tree. TR2 absence from *D. melanogaster* and its
286 close relatives suggests the possibility that this lineage was functionally replaced by TAHRE
287 and/or TART. Finally, we infer that TR2 was lost at least once along the lineage leading to the
288 *D. takahashii/D. biarmipes* clade.

289

290 While some lineages appear to rely on only one element, others harbor multiple retrotransposon
291 lineages. The observation that *D. elegans*, for example, encodes both TAHRE and TR2, *D.*
292 *rhopaloa* both TARTAHRE and TR2, and *D. melanogaster* both TART and TAHRE, rejects the
293 null expectation that the retrotransposon lineage tree recapitulates the species tree. Instead,
294 retrotransposon lineages are selectively retained and selectively lost across the species
295 phylogeny, leading to retrotransposon tree-species tree discordance (Fig. 1 versus 3).

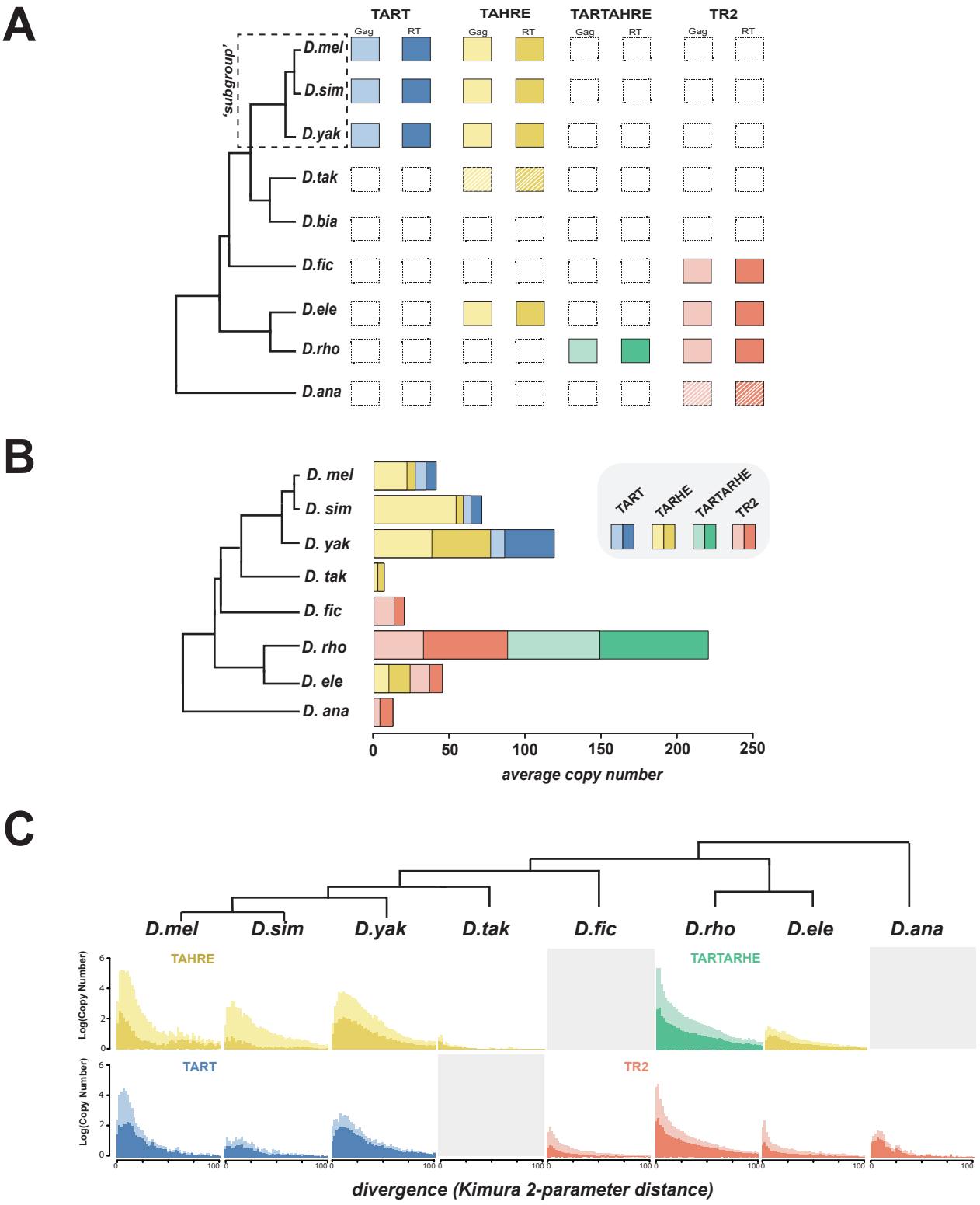

Figure 3. Element identity, copy number, and history across the melanogaster species group. **(A)** Presence/absence of telomere-localized elements across the melanogaster species group and *D. ananassae.* Each column represents a phylogenetically distinct lineage defined and validated in Fig. 1 and Fig. 2, respectively. Hatched lines delineate elements for which only a degraded version was recovered. Gag domains are represented by lighter shaded arrows and RT domains are shown as the darker shaded rectangles. **(B)** Estimated Gag (light) and RT (dark) copy number per species, calculated from the average read depth of a consensus sequence relative to the genome-wide estimate. **(C)** Repeat landscapes of telomere-specialized retrotransposons captured by Kimura 2-p distance between genomic reads with significant BLAST hits to the consensus (>90% identity). Copy number (consensus read # / genome-wide average read #) appears on the Y-axis and binned divergence classes appears on the X-axis. The bins closest to zero putatively represent the youngest classes. Gag (lighter shade), RT (darker shade).

297 Overall, these data are consistent with diversification via gain and loss, both with and without
298 replacement, of major retrotransposon lineages across three to 15 million years of evolution.
299 The melanogaster species group is especially dynamic. We observe pervasive lineage-specific
300 presence/absence of TR2, TARTAHRE, and TAHRE and even extreme cases of wholesale loss
301 of jockey subclade, telomere-specialized elements.
302
303 *Expansions and contractions of DNA content derived from telomere-specialized elements*
304
305 Retrotransposon expansions and contractions over time result in contemporary species-
306 restriction of specific retrotransposon lineages; that is, progressively diminishing copy number
307 precedes the wholesale loss of a retrotransposon lineage. To evaluate such bulk sequence
308 changes across species, we mapped raw reads to a given domain consensus sequence and
309 quantified read depth (Fig. 3B, Table S4). This copy number estimate serves also as a proxy for
310 relative telomere length for all elements except TARTARHE, which localizes to both telomeric
311 and non-telomeric sites. We note that the highly variable telomeric retrotransposon abundance
312 within *D. melanogaster* (Wei et al. 2017) suggests that our single-genome estimates offer only a
313 partial picture of divergence in bulk content.
314
315 We observe dramatic between-species differences in copy number, even for species that share
316 common retrotransposon lineages. The degree and direction of these between-species
317 differences were robust to multiple percent similarity thresholds to the consensus, minimizing
318 the likelihood that we inadvertently underestimated copy number due to highly diverged variants
319 (Table S4). qPCR on genomic DNA in cases of extreme copy number differences across
320 species validated these computationally-generated estimates (Fig. S4). Across *melanogaster*
321 subgroup species, telomeric retrotransposon content is 2-fold larger in *D. yakuba* compared to
322 *D. melanogaster*, at least in the sequenced strains. Moreover, *D. melanogaster* and *D. simulans*

323 telomeres are predominantly comprised of the TAHRE Gag (HeT-A), while *D. yakuba* encodes
324 relatively more TART RT (Fig. 3B). TARTARHE in *D. rhopaloa* is especially abundant, an
325 unsurprising discovery given its localization not only to chromosome ends but also to many
326 other chromosomal locations (Fig. 2B). The telomere-restricted element, TR2, is also highly
327 abundant in *D. rhopaloa* but depauperate in *D. ananassae. D. takahashii* too harbors low copy
328 number of TAHRE Gag and RT domains. Notably, we were also unable to recover using *in silico*
329 or PCR-based methods a single complete TAHRE domain in *D. takahashii* (we recovered only
330 truncated versions of both domains), suggestive of ongoing loss of TAHRE-dependent telomere
331 elongation. These data highlight the divergent jockey subclade content even across species that
332 share a common retrotransposon lineage.
333
334 To further refine our snapshot of retrotransposon invasion and degeneration history, we
335 generated frequency distributions of pairwise read divergence for each element in each species.
336 If transposition events were recent, we expect the highest frequency classes to exhibit lowest
337 divergence. Alternatively, an element that has undergone only an ancient episode of expansion
338 exhibits an excess of high frequency reads with similar but elevated divergence from the
339 consensus. Since fully functional Drosophila telomeres must constantly renew their telomere-
340 specialized retrotransposon template, we expect actively transposing telomeric elements to
341 have high similarity between consensus-mapping reads. Profiles of sequence divergence
342 (estimated by Kimura 2-p distance) support this mechanism—virtually all distributions reveal an
343 enrichment of reads diverging minimally or not at all (Fig. 3C). Indeed, most distributions are
344 typical of recent transposition bursts characterized by a preponderance of highly similar
345 sequences. We cannot formally rule out the homogenizing force of gene conversion to the
346 excess of highly similar reads; however, the empirically verified rates of transposition in *D.*
347 *melanogaster* (Kahn et al. 2000) suggests that gene conversion may contribute only moderately
348 to the observed patterns. The distributions of *D. ananassae* TR2, *D. simulans* TART, and *D.*

349 *elegans* TAHRE instead appear more uniform, consistent with ancient activity followed by
350 mutation accumulation.
351
352 *Telomere-specialized retrotransposons are absent in D. biarmipes*
353
354 Our data suggest that jockey subclade, telomere-specialized retrotransposons frequently
355 occupy the melanogaster species group's telomeres. However, these elements appear as full
356 length in only a subset of the 10 species. To evaluate the possibility that *active* insertions by
357 telomere-specialized retrotransposons can be lost completely, we exploited the most extreme
358 case of telomeric retrotransposon degeneration uncovered by our pipeline. We detected no *D.*
359 *biarmipes* telomere-specialized elements branching inside the monophyletic jockey subclade
360 (Fig. 1). From these short-read data, we recovered instead a jockey family Gag (arrow, Fig. 1A).
361 Hybridizing a FISH probe cognate to 'jockey_1' confirmed our phylogeny-based inference.
362 Specifically, this computationally predicted jockey element localized across *D. biarmipes*
363 chromosome arms but not to telomeres (Fig. S5), a typical, jockey family chromosomal
364 distribution (Kaminker et al. 2002; Xie et al. 2013). To test the hypothesis that this exceptional
365 species has lost jockey family retrotransposons at its chromosome ends, we performed long-
366 read sequencing using the Pacific Biosciences (PacBio) SMRT sequencing platform. We
367 predicted that upon assembling whole telomeres from *D. biarmipes*, we would discover a newly
368 domesticated, telomere-specialized mobile element lineage unrelated to the jockey family, the
369 wholesale loss of telomere-specialized mobile elements, or both.
370
371 Our long read-based assembly of the *D. biarmipes* genome revealed dramatically different
372 chromosome end composition from the well-studied telomeres of *D. melanogaster* (Fig. 4A).
373 Using the packages PBcR (Berlin et al. 2015) or DBG2OLC (Ye et al. 2016), we assembled *D.*
374 *biarmipes* telomeres corresponding to *D. melanogaster*'s 2L, 2R, 3L, and 3R chromosomes. To

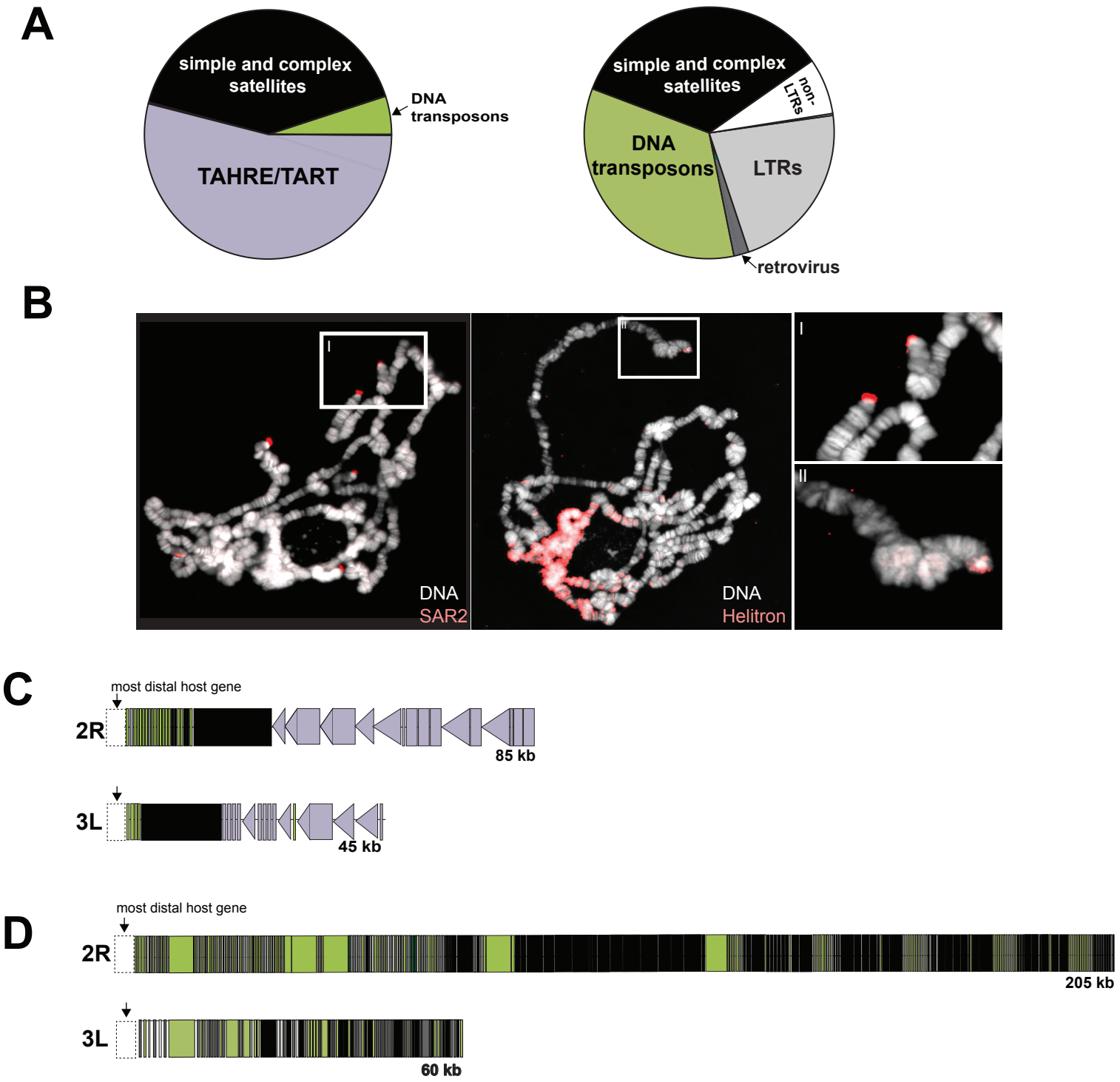

Figure 4. **Collapse of the retrotransposon-based telomere elongation mechanism in *D. biarmipes.*** **(A)** Composition of *D. melanogaster* and *D. biarmipes* chromosomes from the most distal, protein-coding gene and the terminal nucleotide (assembled from PacBio-generated long reads for both species) for Muller elements corresponding to 2L, 2R, 3L, and 3R. Purple corresponds to telomere-specialized, jockey family retrotransposons found in *D. melanogaster* (but not in *D. biarmipes*). **(B)** Fluorescent *in situ* hybridization of a SAR2 and Helitron probes to polytene chromosomes from *D. biarmipes.* Inset "I" and "II" shows telomere localization of SAR2 and Helitrons, respectively, on *D. biarmipes* polytene chromosomes. **(C)** Schematic representation of the long read-based assembly of *D. melanogaster* 2R and 3L telomeric DNA. Telomere-specialized, jockey-like elements (purple) are distal to a block of simple and complex satellites (black), consistent with previous reports. Triangles represent full-length elements and rectangles represent partially degenerated elements. **(D)** Schematic representation the of the long read-based assembly of 2R and 3L telomeres from *D. biarmipes* where simple and complex satellite DNA (black) is juxtaposed with primarily Helitron DNA transposons (green).

376 determine composition in both species, we delineated *D. melanogaster* and *D. biarmipes*
377 telomeric DNA as distal to the most terminally annotated, orthologous genes from *D.*
378 *melanogaster* (the two species share a common karyotype (Deng et al. 2007)). Consistent with
379 previous literature (Pardue and DeBaryshe 2011; Mason et al. 2016), *D. melanogaster*
380 telomeres are dominated by telomere-specialized elements (HeT-A/TAHRE and TART) and
381 satellite repeats (primarily "HETRP_DM', Fig. 4A). In sharp contrast, the 50-200kb *D. biarmipes*
382 telomeres harbor no jockey subclade retrotransposons, consistent with our short read-based
383 pipeline results above. We uncovered instead DNA transposons called Helitrons, and to a lesser
384 extent, Gypsy family LTR elements (Fig. 4A). We validated these *D. biarmipes* telomere
385 assemblies in two ways. First, we observed 99% correspondence and no structural
386 discrepancies between our assembled *D. biarmipes* telomeres and those of an independent,
387 recently published hybrid assembly based on 454 and Nanopore reads (Fig. S6, (Miller et al.
388 2018)). Second, hybridization of FISH probes cognate to the telomeric Helitron and telomeric
389 satellite sequence confirmed that we successfully assembled the termini of *D. biarmipes*
390 chromosomes (Fig. 4B, insets).
391
392 The rough equivalency of *D. melanogaster* and *D. biarmipes* telomeric DNA attributed to mobile
393 elements and to satellite sequence suggests functional replacement of jockey subclade,
394 telomere-specialized retrotransposons with other mobile element families. However, the integrity
395 of these elements and their physical distribution challenges this inference. Starting at the most
396 distally annotated host gene, *D. melanogaster* encodes long, uniform tracts of satellite DNA
397 (primarily "HETRP-DM", (Karpen and Spradling 1992)) followed by equivalently uniform tracts of
398 the end-integrating TART and TAHRE (Fig. 4C, Table S5). Many of these elements are full-
399 length and so presumably still active. Pervasive degradation instead dominates the Helitrons of
400 *D. biarmipes* (Fig. S7). Moreover, these fragments are interspersed amidst satellite sequence
401 ('SAR' and 'SAR2', Fig. 4D) and orientated randomly (Table S6), unlike the head-to-tail arrays of

402 jockey subclade retrotransposons in *D. melanogaster*. All assembled *D. biarmipes* telomeres
403 exhibit this pattern, despite varying in length from 50-200kb and varying in enrichment for
404 transposable element family (Fig. 4D, Fig. S8, Table S6). Finally, the abundant Helitron signal at
405 the chromocenter (Fig. 4B) and the patterns of divergence between copies (Fig. S9) highlights
406 the similarity between *D. biarmipes* telomeres and its pericentric heterochromatin, consistent
407 with 'generalist' mobile elements occupying a niche once-restricted to telomere-specialized
408 elements.
409
410 Patterns of divergence between Helitron insertions and repeat structure of satellite sequence
411 implicates an alternative lengthening mechanism in *D. biarmipes.* Using the uniquely long
412 chromosome 2R as model, we observed that Helitron divergence is uniformly elevated between
413 both nearby and distant insertions, whereas TAHRE Gag divergence is uniformly low (Fig. S10).
414 We attribute this pattern of divergence to an ancient Helitron invasion followed by mutation
415 accumulation. This pattern contrasts sharply with uniformly low divergence of TAHRE Gag from
416 *D. melanogaster* **(**Fig. S10), likely driven by constant "refreshment" of end-inserting TAHRE Gag
417 ("HeT-A") elements. The absence of full-length Helitrons anywhere in our long-read *D.*
418 *biarmipes* assembly (Fig.S7) suggests that these *D. biarmipes* telomeric Helitron fragments do
419 not jump via Helitron mobilization machinery in *trans*. The *D. biarmipes* telomere tracts of
420 repeats are instead reminiscent of Drosophila pericentromeric heterochromatin, which encodes
421 long tracts of simple and complex satellites interspersed with dead TEs (Hoskins et al. 2007). *D.*
422 *biarmipes* telomeres reveal the dramatic loss of a greater than 40 million-year old telomere
423 elongation mechanism.
424
425
426
427

428 **DISCUSSION**

429

430 Virtually all eukaryotes rely on end-targeting and reverse transcription by telomerase to add
431 DNA repeats to chromosome termini (Greider and Blackburn 1989; Zakian 1989; Blackburn
432 1991; Zakian 1996). The chromosome ends of *D. melanogaster* instead rely on telomere-
433 specialized retrotransposons (Pardue and DeBaryshe 2011). These 'domesticated' mobile
434 elements, too, end-target and reverse transcribe RNA into repetitive DNA at the most terminal
435 base pairs, preserving unique genes several thousand nucleotides away. These alternative
436 mechanisms differ not only in the molecular players that preserve chromosome ends but also in
437 the rates of molecular evolution of telomeric DNA sequence. The telomerase-associated,
438 nucleotide repeat unit composition changes slowly across eukaryotes (Meyne et al. 1989;
439 Mason et al. 2016; Podlevsky and Chen 2016). Our investigation of the melanogaster species
440 group reveals instead rapid diversification of the repeats charged with telomere length
441 maintenance in Drosophila.

442

443 Building on previous discoveries of phylogenetically distinct, candidate telomeric jockey
444 elements (Villasante et al. 2007), we report rapid turnover of major retrotransposon lineages
445 across the melanogaster species group. Our comprehensive computational search, supported
446 by experimental validation, revealed that an ancient TR2 lineage is retained in some species but
447 sporadically lost and replaced by TAHRE, and/or TART in others. Notably, some species harbor
448 a single retrotransposon lineage, like *D. ficusphila*, while others maintain multiple
449 retrotransposon lineages, like *D. melanogaster* and *D. elegans*. Still others, like *D. biarmipes,*
450 appear to lack telomere-restricted retrotransposons altogether. Consistent with a dynamic
451 history of turnover, telomere-associated retrotransposon copy number also varies broadly. *D.*
452 *rhopaloa*, for example, encodes 10x the TR2-derived sequence of *D. ananassae*. Appreciating
453 that copy number also varies dramatically within species (Wei et al. 2017), we presume that

454 between-species divergence documented here captures only a partial picture of copy number
455 divergence. Together, these data highlight the remarkable diversity of telomere-specialized
456 sequence—both identity and amount—across only 15 million years of Drosophila evolution.
457
458 Partially degenerated elements, rather than full length elements, dominate chromosome ends of
459 several focal species, e.g., *D. takahashii* and *D. ananassae.* This diversity of telomere
460 composition challenges the widely cited idea that active transposition is the primary mechanism
461 of length regulation genus-wide (Casacuberta and Pardue 2003; Villasante et al. 2007). Active
462 retrotransposons indeed populate the well-studied telomeres of *D. melanogaster*—the long
463 read-based assembly of *D. melanogaster*'s telomeres ((Kim et al. 2014), Fig. 4C) supports
464 decades of work defining this system (Biessmann et al. 1990; Mason and Biessmann 1995;
465 Pardue and DeBaryshe 2003; Pardue and DeBaryshe 2011; Mason et al. 2016). Moreover, it is
466 well-established that these active elements insert even at terminal deficiencies proximal to *D.*
467 *melanogaster's* telomeric retrotransposon array (Biessmann et al. 1990; Kahn et al. 2000).
468 Finally, all Drosophila species investigated over the past several decades harbor telomere-
469 specialized jockey subclade elements (Casacuberta and Pardue 2003; Berloco et al. 2005;
470 Villasante et al. 2007). However, in contrast to *D. melanogaster*, species like *D. takahashii*
471 appear to encode only partial copies of its telomeric retrotransposons (Fig. 2A, Fig. S11).
472
473 The predominance of only degenerated versions in this species raises the possibility that
474 chromosome length maintenance depends on an alternative class of telomere-specialized
475 mobile elements or instead, a non-mobile element-based mechanism. If newly domesticated
476 mobile elements readily replace ancestral, telomere-specialized lineages, we would expect the
477 species ostensibly lacking active jockey subclade elements—*D. biarmipes*—to encode such
478 new recruits. The extreme terminal ends of *D. biarmipes* instead harbor AT-rich satellite "SAR"
479 DNA (Mirkovitch et al. 1984; Kas and Laemmli 1992) at all telomeres as well as Helitron

480 fragments (Kapitonov and Jurka 2007) and Gypsy-derived LTR fragments (Nefedova and Kim
481 2017) at variable frequencies across different telomeres. These non-autonomous elements are
482 found both inside and outside the *D. biarmipes* telomeres (Fig. 4D) and found more typically
483 outside the telomere in other Drosophila species (Kas and Laemmli 1992; Kapitonov and Jurka
484 2007; Nefedova and Kim 2017). Moreover, we found not one full-length Helitron in the *D.*
485 *biarmipes* genome, rejecting the possibility of contemporary propagation using mobilization
486 proteins in *trans* (Fig. S7). These data suggest that active transposition is not responsible for
487 telomere length regulation, at least recently, in *D. biarmipes*. Intriguingly, Dias *et al.* (2015)
488 previously reported dramatic expansion of a Helitron tandem repeat, DINE-TRI, along only the
489 lineage leading to *D. biarmipes* and the lineage leading to *D. virilis/D. americana* relative to 25
490 other Drosophila genomes (Dias et al. 2015). Hybridization of DINE-TRI probes to *D. virilis*
491 polytene chromosomes revealed striking localization to multiple telomeres, in addition to
492 pericentromeric heterochromatin, reminiscent of the localization we detected in *D. biarmipes*.
493 Like *D. biarmipes* telomeres, telomeric Helitrons in *D. virilis* represent only partial fragments
494 rather than full-length, actively transposing copies (Dias et al. 2015). However, unlike *D.*
495 *biarmipes*, *D. virilis* telomeres also encode jockey subclade retrotransposons predicted to
496 maintain its chromosome length (Casacuberta and Pardue 2003).
497
498 In the absence of active telomeric transposition, how might *D. biarmipes* (and possibly *D.*
499 *takahashii*) telomeres be maintained? *D. melanogaster* telomeres elongate not only by
500 transposition but also by a recombination-based mechanism called ‘terminal gene conversion’
501 (Kahn et al. 2000; Melnikova and Georgiev 2002). This alternative lengthening pathway, utilized
502 across eukaryotes alongside or instead of telomerase (McEachern and Haber 2006; Sakofsky
503 and Malkova 2017; Sobinoff and Pickett 2017), relies on the shorter “broken” telomere resecting
504 and invading a longer telomere. Second strand synthesis extends the chromosome end. This
505 newly synthesized strand serves as the template for synthesis of its sister. In addition to

506 terminal gene conversion, nonallelic recombination between two short telomeres could also
507 lengthen deficient chromosomes, where the shorter product is ultimately removed by natural
508 selection.
509
510 We predict that *D. biarmipes* depends on this alternative pathway. All studied, non-Drosophila
511 species from the insect Order, Diptera, lack both telomerase-added repeats *and* telomere-
512 specialized TEs and so are presumed to depend on terminal gene conversion (Mason et al.
513 2016). Specifically, Dipterans such as *Anopheles gambiae* (Biessmann et al. 1998)*,*
514 *Chironomous sp.* (Lopez et al. 1996; Rosen and Edstrom 2000), and *Rhynchosciara americana*
515 (Lopez et al. 1996) encode only simple and/or complex satellites at chromosome ends. We,
516 therefore, predict that Drosophila species lacking full length, telomere-specialized
517 retrotransposons rely exclusively on this alternative elongation mechanism. Consistent with
518 recombination shaping *D. biarmipes* telomere sequence evolution, we observe higher order
519 repeat unit structure of telomeric SAR (Fig. S12). The concomitant proliferation of non-
520 autonomous Helitrons at *D. biarmipes* telomeres may serve to replenish the repetitive sequence
521 that mediates resected end homology searches and ultimately maintains the repair template. *D.*
522 *biarmipes'* close relative*, D. takahashii,* likely represents a transitional state. In *D. takahashii,* we
523 detected no evidence of a full-length TR2. Instead we recovered *in silico,* and then confirmed
524 via PCR, at least one chimeric TR2-Helitron instance. The variable retention of a
525 retrotransposon-based telomere elongation mechanism, together with the inferred ancestral
526 Dipteran telomere state, helps us to contextualize what seemed initially like an aberration.
527 Specifically, *D. biarmipes'* jockey element-poor, Helitron-rich telomeres may, in fact, represent a
528 reversion to an ancestral, Dipteran-like state rather than a previously unseen innovation (Fig. 5).
529
530 The discovery of "undomesticated" Helitrons at *D. takahashii* and *D. biarmipes* telomeres raises
531 the possibility that some telomeric mobile elements act akin to ecological 'facilitators' rather than

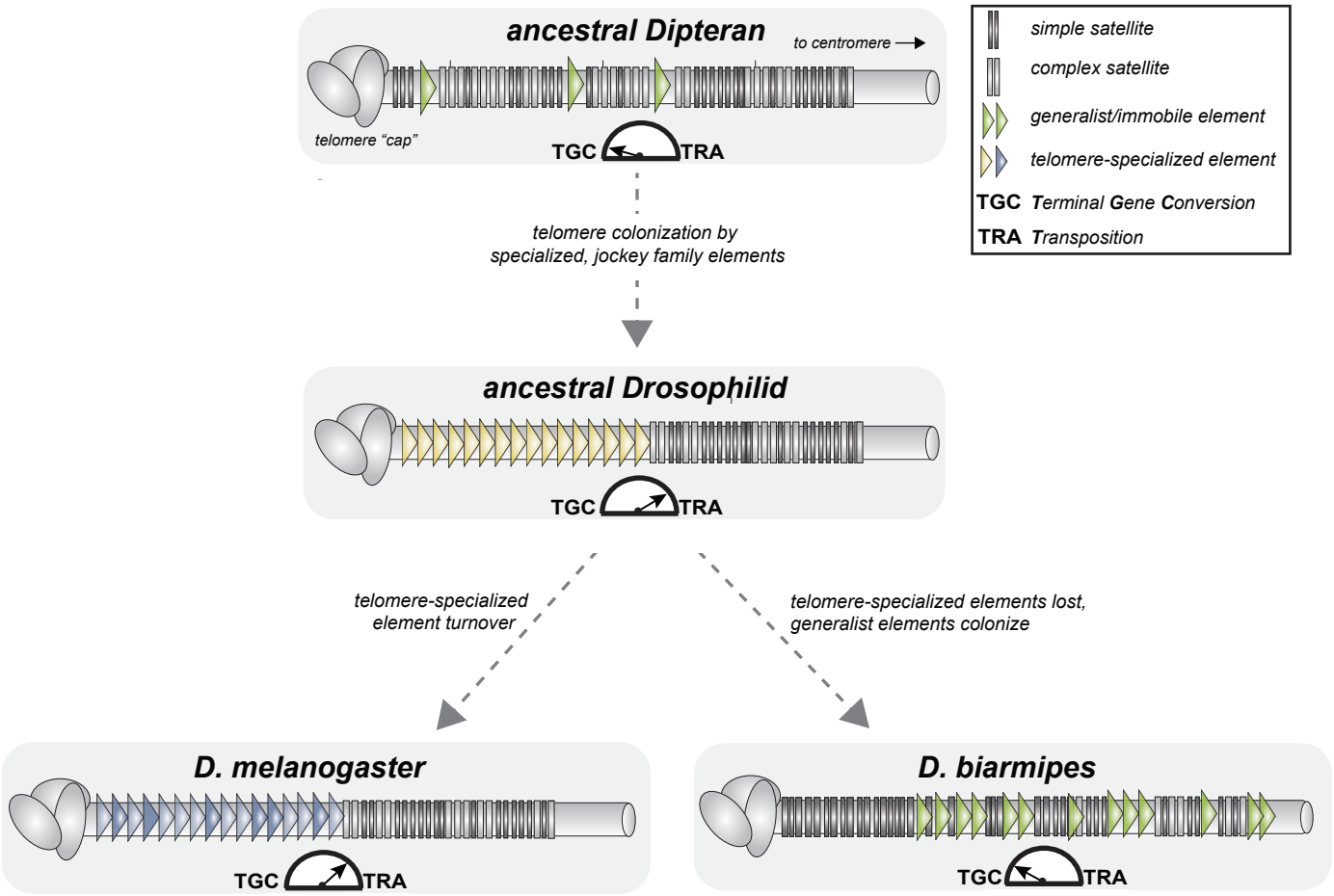

Figure 5. Model of telomere elongation mechanism evolution pre- and post- birth of Drosophilids. The Dipteran ancestor of Drosophila encodes neither telomerase nor telomere-specialized mobile elements. Instead, a recombination-based mechanism, possibly 'terminal gene conversion', lengthens the repetitive DNA (e.g., in mosquitos and midges). Exclusive chromosome-end insertions by a jockey family element becomes the primary, Drosophila-wide telomere elongation mechanism. Major jockey family lineages turn over across Drosophila species that retain this lengthening mechanism (bottom left). In species like D. biarmipes, the subsequent loss of telomere-specialized elements but presence of 'generalist' mobile elements illustrates how some Drosophila species may revert to the ancestral, predominantly recombination-based telomere lengthening mechanism (bottom right).

533 ‘mutualists’. Like a canopy tree inadvertently cultivating a favorable environment for shade-
534 tolerant herbs, these immobile elements, by virtue of where they inserted during an ancient
535 transposition burst, serve inadvertently the genome’s recombination-based telomere
536 lengthening mechanism. Moreover, we cannot rule out the possibility that reverse transcription
537 of satellite or Helitron RNA may also extend these telomeres (and other species’) by utilizing a
538 non-telomeric retrotransposon’s reverse transcriptase in *trans* (Gorab 2003). We speculate that
539 these ‘back up’ mechanisms reduce constraint on telomere-maintenance by active
540 transposition, setting the stage for intra-genomic conflict to shape mobile element diversity
541 across the melanogaster species group.
542
543 If intra-genomic conflict lies beneath this genomic symbiosis, who are the dueling parties?
544 Compromised host fitness may ultimately select for expulsion of ‘disobedient’ retrotransposons
545 that excessively lengthen telomeres or disrupt the euchromatic chromosome arms upon
546 ‘escape’ from this chromosomal niche, as TARTAHRE appears to have done (host-
547 retrotransposon conflict). Indeed, evolutionary pressure to police these telomeric
548 retrotransposons may shape the adaptive evolution detected at telomere length regulator
549 proteins encoded by the host (Lee et al. 2017). Alternatively, transposable elements may
550 compete among themselves for dominance within this genomic habitat (TE-TE conflict). More
551 likely, both host genome and mobile element fitness together shape telomeric DNA diversity.
552 Future work will define the genetic determinants and the molecular mechanisms shaping this
553 ostensibly tenuous relationship between the host genome and its ever-evolving telomeric
554 retrotransposons.
555
556
557
558

559 **MATERIALS AND METHODS**

560

561 **Species selection**

562

563 To elucidate lineage turnover of Drosophila telomeric retrotransposons, we harnessed the
564 'melanogaster species group', which captures 15 million years of evolution (Chen et al. 2014).
565 This slice of evolutionary time offers a previously unexplored scale of resolution: published
566 investigations have targeted either short evolutionary distances from *D. melanogaster*—the
567 melanogaster subgroup, where major telomeric retrotransposon lineages are conserved
568 (Danilevskaya et al. 1998; Casacuberta and Pardue 2002; Berloco et al. 2005; Villasante et al.
569 2007)—or instead very long evolutionary distances from *D. melanogaster* (e.g., *D. ananassae,*
570 *D. pseudoobscura*), where telomeric retrotransposon lineages are wildly diverged and never
571 localized to chromosome ends (Villasante et al. 2007). The melanogaster species group spans
572 an intermediate distance poised to capture the degradation of ancestral retrotransposons that
573 maintain telomeres and the birth of new ones. *D. takahashii, D. ficusphila, D. biarmipes, D.*
574 *elegans* and *D. rhopaloa* broadly sample this unexplored part of the Drosophila tree. To these
575 species, we compare the melanogaster subgroup, including *D. simulans, D. sechellia, D.*
576 *yakuba. D. ananassae* serves as a distant outgroup previously investigated by (Villasante et al.
577 2007).

578

579 **Drosophila genome sequence data used to define telomeric retrotransposons**

580

581 Reference genomes assembled from short read sequences rarely include the repetitive DNA
582 elements that accumulate at chromosome ends. We searched instead the publicly available raw
583 sequence reads derived from 10 species (Table S7) sequenced with either Sanger (ABI 3700)
584 and/or 454 (Genome Sequencer FLX). Although sequencing depth varies across these 10

585 species, we discovered jockey subclade, candidate telomere-specialized elements in even the
586 lowest-coverage genome. The only exception was *D. biarmipes,* a species to which we
587 ultimately subjected long-read sequencing (see below). Our single molecule-based genome
588 assembly confirmed the absence of jockey subclade elements. For all other genomes, we
589 detected no biased enrichment of jockey family lineages in genomes sequenced using one
590 platform or another (Fig. S13). We also used publicly available Illumina-based short read data
591 for consensus sequence correction (Table S7, see below).
592
593 **Identification of candidate telomeric retrotransposons from raw reads**
594
595 We developed a custom method (Fig. S2) to search raw reads from 10 Drosophila species for
596 the non-LTR retrotransposons related to jockey subclade of telomere-specialized elements. We
597 designed this two-step method to identify phylogenetically distinct elements that maintain
598 chromosome ends—both active and degraded copies. Our initial query included all previously
599 characterized, telomeric-specialized retrotransposons from the melanogaster subgroup (*D.*
600 *melanogaster, D. simulans, D. sechellia, D. yakuba*) as well as those uncharacterized elements
601 only predicted to maintain telomere ends in more distant species (Table S1). To ensure that our
602 search was exhaustive, we included even retrotransposons for which telomeric localization or
603 head-to-tail orientations had not been confirmed (Table S1, references therein). We also
604 included the reference jockey element from *D. melanogaster* (Repbase,
605 www.girinst.org/Repbase/). These input sequences served as the query for a tBLASTn search
606 of a given species' raw read database (Table S7). We retained any read that shared 80%
607 sequence identity with any one of the query elements. We then *de novo* assembled this subset
608 of raw reads into consensus sequences (i.e., two or more reads assembled by the Geneious
609 assembler: Medium Sensitivity/Fast parameters). We aligned each consensus sequence to the
610 query sequences plus the reference jockey element from *D. melanogaster* using MAFFT (k =2,

611 Gap penalty=1.53, Offset=0.123). For each alignment of either the Gag or RT domain, we built a
612 phylogenetic tree using FastTree (GTR+CAT, (Price et al. 2009)) and determined if the focal
613 consensus sequence branched as an ingroup or an outgroup (outside the jockey element). We
614 retained only ingroup consensus sequences for subsequent analysis (482 out of 4,583
615 consensus sequences).
616
617 We repeated this pipeline by inputting the 482 consensus sequences from round one (length
618 mean= 4143, standard deviation= 1656) as a new query in a BLASTn search of the raw reads
619 from each species. This second iteration enriched our pool substantially: we obtained ~2 million
620 total reads from all 10 species in this round compared to 60,000 from the first round (Table S8).
621 We filtered uninformative reads by mapping the two million reads to the 482 consensus
622 sequences sited above (Geneious mapper: Medium Sensitivity/Medium; Iterative Fine Tuning =
623 5). We refined consensus accuracy from the pool of 754,230 reads retained with filtering. Again,
624 we *de novo* assembled the pool of retained reads, which generated 62,383 new consensus
625 sequences. We repeated the FastTree sorting method and retained only consensus sequences
626 that branched internally to the jockey clade, leaving us with 3,112 consensus sequences (length
627 mean= 1531, standard deviation = 732).
628
629 We next generated full-length Gag and RT domains from our consensus sequences. We first
630 translated the nucleotide sequences using “six-pack” (Rice et al. 2000) and aligned using
631 MAFFT to known telomere-specialized retrotransposon domains (parameters: -k =2, Gap
632 penalty=1.53, Offset=0.123, Table S1, (Katoh and Standley 2013)). We then parsed the Gag
633 and RT consensus sequences based on its branch position (*i.e.* TAHRE-like, TART-like, TR2-
634 like, jockey-like) from the previous FastTree sorting step. Guided by the alignments, we
635 removed frameshifts and unalignable sequence. We discarded entire consensus sequences
636 harboring less than 80% identity over 80% of its alignment length to any other consensus within

637 the subset (Wicker et al. 2007). For each subset of consensus sequences per species, we
638 generated final consensus sequences using majority rule. We also used Repbase to infer
639 consensus identity for all discarded sequences described above. These sequences represented
640 mostly jockey elements (Table S2) and also several highly degraded, telomere-associated
641 elements juxtaposed with DNA unrelated to the jockey family. We conducted a final refinement
642 step for each complete Gag and RT domain by mapping raw reads to our set of telomeric
643 retrotransposon candidates. These raw reads came from Illumina-based short read sequence
644 databases generated independently of the Sanger and 454 reads used to build the consensus
645 sequences (Table S7). This step successfully called many sites previously designated
646 ambiguous based on majority rule. In addition, we elongated the domain boundaries using
647 reads that spanned the consensus and sequence outside the domain. Again, we used majority
648 rule to infer the final sequence. We repeated cycles of mapping and extension up to 10 times
649 until we detected a full-length open reading frame and, in most cases, the sequence spanning
650 Gag and RT for a given element. All scripts are available upon request.

651 **Phylogenetic tree building of final consensus sequences**

652 We built Gag- and RT- based phylogenetic trees of the retained consensus non-LTR
653 retrotransposons using MrBayes 3.2.6 (Ronquist et al. 2012), implemented in Geneious (Kearse
654 et al. 2012). We determined the amino acid substitution model using ProTest 3.4.2 (Darriba et
655 al. 2011). We performed MCMC analysis over 1,100,000 generations. We estimated
656 convergence of the joint posterior probability density each 200 generations with a 25% burn-in
657 of the sampled chains. We stopped the analysis once the maximum standard deviation of split
658 frequencies was < 0.001. The *D. melanogaster* reference jockey element served as the
659 outgroup. We built phylogenies of Gag and RT domains defined by our pipeline using the
660 majority-rule consensus sequences only (Fig. 1). We also built Gag and RT trees that included
661 all previously described telomeric retrotransposon consensus within the melanogaster subgroup

662 and all previously predicted candidate telomeric elements outside of the melanogaster species
663 group to contextualize our results in the previous literature (Table S1, Fig. S3).

664 **Validation of computationally-defined candidate telomeric retrotransposons**

665 Our custom pipeline detected many species-restricted, non-LTR retrotransposons that form a
666 monophyletic clade with previously described telomere-specialized elements. These
667 phylogenetically distinct, intact open reading frames are supported by overlapping reads along
668 each sequence and high coverage (>27X on average). However, assemblies of short reads
669 mapping to repetitive sequence and majority rule-based consensus sequences may include
670 computational artifacts that do not represent sequences that truly exist in a given genome.

671 *PCR validation*

672 We used PCR to validate 1) domain consensus sequences and 2) head-to-tail tandem array
673 orientation stereotypical of telomere-lengthening retrotransposons. Specifically, we designed
674 primers (Table S9) from our consensus sequences and performed PCR on genomic DNA
675 (Table S10) prepared from 10 females from each species (ArchivePure™, 5 PRIME). We gel-
676 extracted and Sanger-sequenced (Genewiz, South Plainfield, NJ) the PCR products of the
677 expected size (QIAquick® Gel Extraction kit**,** Table S3).

678 *Cytogenetic validation*

679 PCR-validated elements may be phylogenetically related to telomere-specialized elements but
680 not necessarily telomere-restricted. We used fluorescent *in situ* hybridization (FISH) probes
681 and/or 'Oligopaints' (Beliveau et al. 2015) to determine whether the computationally predicted,
682 PCR-validated retrotransposons localize to chromosome ends. Specifically, we hybridized to
683 polytene chromosomes FISH probes cognate to HeT-A, TART (both *D. melanogaster* controls),
684 and TARTAHRE (*D. rhopaloa)* as well as Oligopaints cognate to TR2 (*D. elegans,* (Table S9)).

685 We also evaluated whether a jockey family element from *D. biarmipes*, a species that lacks
686 HeT-A, TART, TARTAHRE, and TR2, localized to telomeres using FISH (previously referred to
687 HTR0 (Villasante et al. 2007)). Finally, we hybridized Helitron and SAR2 sequence-derived
688 FISH probes to *D. biarmipes* polytene chromosomes to validate our long read-based assembly
689 of its chromosome ends (Table S9).
690
691 We first dissected salivary glands from third-instar larvae in a PBS-Tween 0.1% solution. We
692 then transferred salivary glands into fresh 45% acetic acid and 2% paraformaldehyde for one
693 minute on a coverslip treated with Sigmacote® (Sigma-Aldrich). We lowered Poly-L-lysine
694 (Sigma-Aldrich) treated slides onto the coverslip and spread the polytene chromosomes by
695 tapping the slide with a rubber hammer. We flash-froze the slides in liquid nitrogen and
696 transferred into 100% ethanol (-20°C) for 10 minutes. To prepare chromosomes for probe
697 hybridization, we washed twice in PBS-Tween 0.1% for 5 minutes (min) and then transferred
698 slides in successive pre-hybridization solutions (5 minutes in 2X SSC plus 0.1% Tween, 5
699 minutes in 2X SSC plus 0.1% Tween and 50% formamide, 2.5 min at 92 °C in 2X SSC plus
700 0.1% Tween and 50% formamide, and then 20 min at 60 °C in 2X SSC plus 0.1% Tween and
701 50% formamide).
702
703 We labeled DNA FISH probes either with DIG dUTP (Sigma-Aldrich, HeT-A, TART,
704 TARTAHRE) or used Oligopaints (TR2, Table S9). We diluted probes (15μg/ml for PCR DIG
705 probes and 2μg/ml for Oligopaints) in a hybridization buffer (50% formamide, 25% Dextran
706 Sulfate, 10μg of RNase A and 12.5% of ddH2O), denatured probes for 2.5 min at 95 °C and
707 then incubated on ice for 2 min. We performed hybridization overnight at 50°C for PCR-DIG
708 probes and 37°C for Oligopaints. We removed unbound probes with two 10 min washes in 2X
709 SSC at 60 °C and three 10 min washes in PBS at room temperature. We stained PCR-DIG

710 probes with anti-Dioxigenin-AP Fab fragments (Sigma-Aldrich), diluted 1:250 in PBST+BSA 3%,
711 1-hour incubation) and then anti-Sheep IgG (H+L) Alexa Fluor 555 conjugated secondary
712 antibody (Invitrogen^{TM}, diluted 1:500 in PBST+BSA 3%, 1-hour incubation). For Oligopaints, we
713 followed the (Beliveau et al. 2015) protocol for secondary incubation. We visualized
714 hybridization experiments on a Leica SP8 confocal microscope.

715 **Relative telomeric retrotransposon copy number and estimates of within-genome**
716 **diversity**

717 To estimate the relative abundance of our validated telomeric retrotransposons in a given
718 species genome, we mapped raw reads (Sanger or 454) to both our consensus sequences and
719 the most recent genome assembly (Table S7) using SMALT
720 (http://sourceforge.net/projects/smalt/) with varying parameters (k: length of the hashed word
721 index, s: the sampling distance between successive words, and y: the percent identity allowing
722 a word to match). We selected the parameters that maximized coverage for both genome and
723 TE consensus (index: -k 20, -s 13; map: -y 0.9). To infer retrotransposon copy number, we
724 divided a given retrotransposon average coverage by the genome-wide average coverage.
725 Previous reports suggested that a non-trivial fraction of telomere-specialized retrotransposons
726 are partially degenerated (Mason and Biessmann 1995). We indeed detected widespread
727 degradation in all sampled genomes. To account for retrotransposon sequence represented by
728 these reads encoding partially degraded sequence, we also conducted a BLAST-guided
729 approach to capture all reads harboring >90% identity to the consensus sequence. We trimmed
730 these reads from their non-telomeric specific sequence and re-calculated normalized coverage
731 with these trimmed reads. We report the latter analysis. In addition, we estimated divergence
732 between these filtered reads and retrotransposon consensus in order to infer retrotransposon
733 invasion history. We aligned in PHYLIP a given consensus sequence and all significant, >90%

734 identity reads. After visual inspection, we estimated the rates of transitions and transversions
735 from alignment files and calculated Kimura 2-parameter distances (Kimura 1980).

736 *Validation of copy number estimates*

737 We evaluated our computational predictions of copy number variation using qPCR on genomic
738 DNA prepared from 10 females per species (ArchivePure^{TM}, 5 PRIME). We designed primers to
739 amplify the region of highest mean coverage (estimated from the BLAST analysis above) and
740 used the ΔΔCq method to quantitate abundance relative to a single copy gene (rp49).

741 **Single molecule sequencing, assembly, and validation of *D. biarmipes***

742 *DNA preparation for single molecule sequencing*

743 We detected no evidence of retrotransposons related to TAHRE, TART, TARTARHE, or TR2 in
744 the (short) raw reads of *D. biarmipes.* To define the set of telomeric genetic elements found in
745 this unusual species, we conducted single-molecule sequencing using the Pacific Biosciences
746 platform. We prepared high quality DNA using Qiagen Blood and Cell culture DNA MIDI Kit
747 according to (Chakraborty et al. 2018) with some modifications (Qiagen). Specifically, we
748 collected 150mg of tissue from 200 zero to two-day old, whole females. We held the females in
749 empty vials for two hours prior flash-freezing in liquid nitrogen and grinding to a fine powder. We
750 transferred the sample into 9.5ml of buffer G2 with addition of 38μl of RNase-A (100mg/ml) and
751 250μl of protease (0.75AU). We homogenized the sample and then incubated at 50°C overnight
752 with gentle shaking. The next day, we centrifuged the sample at 5000 x g for 10 min at 4°C. We
753 decanted the supernatant into a fresh 15ml tube, vortexed, and applied it to the anion exchange
754 column. We washed the column and precipitated genomic DNA with 0.7 volume of isopropanol
755 resuspended in Tris buffer (pH 8) and stored at 4°C overnight. Genewiz (South Plainfield, NJ)
756 prepared a 20 kb PacBio SMRTbell library with Blue Pippin size selection per manufacturer's
757 instruction.

758 *Assembling D. biarmipes genome with long reads*

759 We built both PacBio-only assemblies and PacBio/Illumina hybrid assemblies. We used the
760 PBcR pipeline (Celera 8.3) to assemble PacBio reads only. First, we executed a self-correction
761 step ('PBcR-MHAP', k = 14, merSize=16, sketches = 1024, coverage = 25), which generated a
762 pre-assembly (N50= 480 kb; longest contig = 4.3 Mb). We passed the Celera assembler (CA
763 8.3) the longest 25X coverage corrected contigs, using parameters optimized for genomes
764 >100Mb (Berlin et al. 2015). The N50 and longest contig both increased to a final 718 kb and
765 9.1 Mb, respectively. We generated a hybrid assembly of our PacBio reads and the publicly
766 available Illumina short reads from *D. biarmipes* (Table S7) using the software package
767 DGB2OLC (Ye et al. 2016), following parameters described in (Chakraborty et al. 2018). This
768 hybrid assembly increased the N50 and longest contig to 2.5Mb and 9.2Mb, respectively. Both
769 assemblies and raw reads have been deposited at NCBI under XXXXXX.

770 *Defining telomeric DNA from D. biarmipes*

771

772 To extract telomeric DNA sequence from the *D. biarmipes* PBcR assembly, we relied on the
773 well-annotated, full-length chromosome arms of *D. melanogaster* (Release r6.17). Starting with
774 a query of terminal *D. melanogaster* genes (~20 genes per chromosome arm), we used
775 BLASTn to identify *D. biarmipes* contigs encoding terminal sequence (*D. biarmipes* and *D.*
776 *melanogaster* share a common karyotype (Deng et al. 2007)). We identified exactly four *D.*
777 *biarmipes* contigs with homology to *D. melanogaster* 2L, 2R, 3L and 3R chromosome ends. We
778 next used these four contigs as queries in a BLASTn search of the *D. biarmipes* assembly to
779 identify any additional contigs with homology to the terminal sequence. We identified only a
780 single, 29kb contig corresponding to 3R, which overlapped the original 3R contig over 15kb
781 (>99% identity). This new contig augmented the 3R telomeric sequence by an additional 14kb of
782 distal sequence. To further validate our assembly of four *D. biarmipes* telomeres, we compared

783 them to a second hybrid assembly generated independently from Nanopore and Illumina reads
784 (Miller et al. 2018). In no case did we observe in Miller *et al.* contigs that extend beyond the
785 most distal sequence from our assembly (see Fig. S6).
786
787 *Analysis of telomeric DNA sequence in D. biarmipes and D. melanogaster*
788
789 For both our *D. biarmipes* telomeric contigs and those defined using the same methods for *D.*
790 *melanogaster* (Kim et al. 2014), we annotated the sequence distal to the most terminal gene
791 using RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*). From this
792 annotation, we plotted TE family abundance (Fig. S8) and then focused our analyses on the two
793 highest repeat classes, the Helitron 2N DNA transposon and the SAR2 complex satellite repeat.
794 We extracted the repeats from annotated contigs and aligned them to their respective
795 consensus. From these PHYLIP alignments, we calculated Kimura 2-parameter distance. We
796 summarized distance estimates on heatmaps according to their position along telomere terminal
797 sequences (Fig S10). We also annotated higher order repeat structure of individual SAR2
798 variants from PHYLIP alignments by ordering SAR2 repeat according to their position along
799 telomeres (Fig S12). Finally, we compared telomeric Helitrons to peri-centromeric (those
800 sharing a contig with most proximal genes) and euchromatic Helitrons by parsing copies from
801 our PacBio assembly using the Helitron 2N consensus as a BLASTn query (>80% identity
802 threshold). Using the methods described above, we calculated Kimura 2-p distance of copies
803 derived from telomeric, pericentromeric, and euchromatic regions as above (Fig S9).
804
805
806
807
808

809 **DATA ACCESS**

810 All consensus sequences and PCR-amplified elements can be found in Table S2 and S3,
811 respectively. Our PacBio-generated long reads for *D. biarmipes* were deposited at NCBI under
812 SRA XXXXX. Genome assemblies will be deposited at www.flybase.org.

813

814

815

816 **AUTHOR CONTRIBUTIONS**

817 M.T.L and B.S. designed the experiments. B.S. performed the bioinformatics, the experiments,
818 and the analyses. S.C.N. designed the Oligopaints. M.T.L and B.S. wrote the manuscript.

819

820

821 **ACKNOWLEDGEMENTS**

822 The authors thank C. Leek for assistance with PCR-based validation experiments and G. Lee and
823 S. Zanders for their comments on earlier versions of the manuscript. This work was supported
824 by an NIH NIGMS R00GM107351 and an NIH NIGMS R35GM124684 to MTL.

825

826

827

828

829

830

831

832

833

834

835 **LITERATURE CITED**

836 Beauregard A, Curcio MJ, Belfort M. 2008. The take and give between retrotransposable
837 elements and their hosts. *Annu Rev Genet* **42**: 587-617.

838 Beck CR, Garcia-Perez JL, Badge RM, Moran JV. 2011. LINE-1 elements in structural variation
839 and disease. *Annu Rev Genomics Hum Genet* **12**: 187-215.

840 Beliveau BJ, Boettiger AN, Avendano MS, Jungmann R, McCole RB, Joyce EF, Kim-Kiselak C,
841 Bantignies F, Fonseka CY, Erceg J et al. 2015. Single-molecule super-resolution
842 imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH
843 probes. *Nat Commun* **6**.

844 Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large
845 genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol*
846 **33**: 623-630.

847 Berloco M, Fanti L, Sheen F, Levis RW, Pimpinelli S. 2005. Heterochromatic distribution of HeT-
848 A- and TART-like sequences in several Drosophila species. *Cytogenet Genome Res*
849 **110**: 124-133.

850 Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW. 1998. DNA organization and length
851 polymorphism at the 2L telomeric region of Anopheles gambiae. *Insect Molecular*
852 *Biology* **7**: 83-93.

853 Biessmann H, Mason JM, Ferry K, d'Hulst M, Valgeirsdottir K, Traverse KL, Pardue ML. 1990.
854 Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends
855 in Drosophila. *Cell* **61**: 663-673.

856 Blackburn EH. 1991. Structure and function of telomeres. *Nature* **350**: 569-573.

857 Casacuberta E. 2017. Drosophila: Retrotransposons Making up Telomeres. *Viruses* **9**.

858 Casacuberta E, Pardue ML. 2002. Coevolution of the telomeric retrotransposons across
859 Drosophila species. *Genetics* **161**: 1113-1124.

860 Casacuberta E, Pardue ML. 2003. Transposon telomeres are widely distributed in the
861 Drosophila genus: TART elements in the virilis group. *Proc Natl Acad Sci U S A* **100**:
862 3363-3368.
863 Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2018. Hidden genetic
864 variation shapes the structure of functional elements in Drosophila. *Nat Genet* **50**: 20-25.
865 Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC,
866 FitzGerald PC et al. 2014. Comparative validation of the D. melanogaster modENCODE
867 transcriptome annotation. *Genome Res* **24**: 1209-1223.
868 Cheng CY, Vogt A, Mochizuki K, Yao MC. 2010. A domesticated piggyBac transposase plays
869 key roles in heterochromatin dynamics and DNA cleavage during programmed DNA
870 deletion in Tetrahymena thermophila. *Mol Biol Cell* **21**: 1753-1762.
871 Cheng CY, Young JM, Lin CG, Chao JL, Malik HS, Yao MC. 2016. The piggyBac transposon-
872 derived genes TPB1 and TPB6 mediate essential transposon-like excision during the
873 developmental rearrangement of key genes in Tetrahymena thermophila. *Genes Dev* **30**:
874 2724-2736.
875 Danilevskaya ON, Tan C, Wong J, Alibhai M, Pardue ML. 1998. Unusual features of the
876 Drosophila melanogaster telomere transposable element HeT-A are conserved in
877 Drosophila yakuba telomere elements. *Proc Natl Acad Sci U S A* **95**: 3770-3775.
878 Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models
879 of protein evolution. *Bioinformatics* **27**: 1164-1165.
880 Deng Q, Zeng Q, Qian Y, Li C, Yang Y. 2007. Research on the karyotype and evolution of
881 Drosophila melanogaster species group. *J Genet Genomics* **34**: 196-213.
882 Dias GB, Heringer P, Svartman M, Kuhn GC. 2015. Helitrons shaping the genomic architecture
883 of Drosophila: enrichment of DINE-TR1 in alpha- and beta-heterochromatin, satellite
884 DNA emergence, and piRNA expression. *Chromosome Res* **23**: 597-613.

885 Drosophila 12 Genomes C Clark AG Eisen MB Smith DR Bergman CM Oliver B Markow TA
886 Kaufman TC Kellis M Gelbart W et al. 2007. Evolution of genes and genomes on the
887 Drosophila phylogeny. *Nature* **450**: 203-218.
888 Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes.
889 *Annu Rev Genet* **41**: 331-368.
890 Gladyshev EA, Arkhipova IR. 2007. Telomere-associated endonuclease-deficient Penelope-like
891 retroelements in diverse eukaryotes. *Proc Natl Acad Sci U S A* **104**: 9352-9357.
892 Gorab E. 2003. Reverse transcriptase-related proteins in telomeres and in certain chromosomal
893 loci of Rhynchosciara (Diptera: Sciaridae). *Chromosoma* **111**: 445-454.
894 Greider CW, Blackburn EH. 1989. A telomeric sequence in the RNA of Tetrahymena telomerase
895 required for telomere repeat synthesis. *Nature* **337**: 331-337.
896 Hancks DC, Kazazian HH, Jr. 2012. Active human retrotransposons: variation and disease. *Curr*
897 *Opin Genet Dev* **22**: 191-203.
898 Higashiyama T, Noutoshi Y, Fujie M, Yamada T. 1997. Zepp, a LINE-like retrotransposon
899 accumulated in the Chlorella telomeric region. *Embo J* **16**: 3715-3723.
900 Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S,
901 Mendez-Lago M, Rossi F et al. 2007. Sequence finishing and mapping of Drosophila
902 melanogaster heterochromatin. *Science* **316**: 1625-1628.
903 Jangam D, Feschotte C, Betran E. 2017. Transposable Element Domestication As an
904 Adaptation to Evolutionary Conflicts. *Trends Genet* **33**: 817-831.
905 Kahn T, Savitsky M, Georgiev P. 2000. Attachment of HeT-A sequences to chromosomal
906 termini in Drosophila melanogaster may occur by different mechanisms. *Mol Cell Biol* **20**:
907 7634-7642.
908 Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA,
909 Lewis SE, Rubin GM et al. 2002. The transposable elements of the Drosophila
910 melanogaster euchromatin: a genomics perspective. *Genome Biol* **3**: RESEARCH0084.

911 Kapitonov VV, Jurka J. 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends*
912 *Genet* **23**: 521-529.
913 Karpen GH, Spradling AC. 1992. Analysis of subtelomeric heterochromatin in the Drosophila
914 minichromosome Dp1187 by single P element insertional mutagenesis. *Genetics* **132**:
915 737-753.
916 Kas E, Laemmli UK. 1992. In vivo topoisomerase II cleavage of the Drosophila histone and
917 satellite III repeats: DNA sequence and structural characteristics. *Embo J* **11**: 705-716.
918 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
919 improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.
920 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
921 Markowitz S, Duran C et al. 2012. Geneious Basic: an integrated and extendable
922 desktop software platform for the organization and analysis of sequence data.
923 *Bioinformatics* **28**: 1647-1649.
924 Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, Fisher WW, Chin CS, Rapicavoli NA, Rank
925 DR, Li J et al. 2014. Long-read, whole-genome shotgun sequence data for five model
926 organisms. *Sci Data* **1**: 140045.
927 Kimura M. 1980. A Simple Method for Estimating Evolutionary Rates of Base Substitutions
928 through Comparative Studies of Nucleotide-Sequences. *J Mol Evol* **16**: 111-120.
929 Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of
930 unequal exchange in the containment of transposable element copy number. *Genet Res*
931 **52**: 223-235.
932 Lee YC, Leek C, Levine MT. 2017. Recurrent Innovation at Genes Required for Telomere
933 Integrity in Drosophila. *Mol Biol Evol* **34**: 467-482.
934 Lee YCG, Karpen GH. 2017. Pervasive epigenetic effects of Drosophila euchromatic
935 transposable elements impact their evolution. *Elife* **6**.

936 Lopez CC, Nielsen L, Edstrom JE. 1996. Terminal long tandem repeats in chromosomes from
937 Chironomus pallidivittatus. *Mol Cell Biol* **16**: 3285-3290.
938 Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ,
939 Grutzner F, Bauersachs S et al. 2015. Ancient transposable elements transformed the
940 uterine regulatory landscape and transcriptome during the evolution of mammalian
941 pregnancy. *Cell Rep* **10**: 551-561.
942 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC,
943 Gonzales NR et al. 2017. CDD/SPARCLE: functional classification of proteins via
944 subfamily domain architectures. *Nucleic Acids Res* **45**: D200-D203.
945 Mason JM, Biessmann H. 1995. The Unusual Telomeres of Drosophila. *Trends in Genetics* **11**:
946 58-62.
947 Mason JM, Randall TA, Capkova Frydrychova R. 2016. Telomerase lost? *Chromosoma* **125**:
948 65-73.
949 McEachern MJ, Haber JE. 2006. Break-induced replication and recombinational telomere
950 elongation in yeast. *Annu Rev Biochem* **75**: 111-135.
951 Melnikova L, Georgiev P. 2002. Enhancer of terminal gene conversion, a new mutation in
952 Drosophila melanogaster that induces telomere elongation by gene conversion.
953 *Genetics* **162**: 1301-1312.
954 Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the Human Telomere Sequence
955 (Ttaggg)N among Vertebrates. *P Natl Acad Sci USA* **86**: 7049-7053.
956 Miller DE, Staber C, Zeitlinger J, Hawley RS. 2018. GENOME REPORT: Highly Contiguous
957 Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing.
958 *G3 (Bethesda)* doi:10.1534/g3.118.200160.
959 Mirkovitch J, Mirault ME, Laemmli UK. 1984. Organization of the higher-order chromatin loop:
960 specific DNA attachment sites on nuclear scaffold. *Cell* **39**: 223-232.

961 Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV. 2007.
962 Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature*
963 **446**: 208-212.
964 Nefedova L, Kim A. 2017. Mechanisms of LTR-Retroelement Transposition: Lessons from
965 Drosophila melanogaster. *Viruses* **9**.
966 Osanai-Futahashi M, Fujiwara H. 2011. Coevolution of telomeric repeats and telomeric repeat-
967 specific non-LTR retrotransposons in insects. *Mol Biol Evol* **28**: 2983-2986.
968 Pardue ML, DeBaryshe PG. 2003. Retrotransposons provide an evolutionarily robust non-
969 telomerase mechanism to maintain telomeres. *Annu Rev Genet* **37**: 485-511.
970 Pardue ML, DeBaryshe PG. 2008. Drosophila telomeres: A variation on the telomerase theme.
971 *Fly (Austin)* **2**: 101-110.
972 Pardue ML, DeBaryshe PG. 2011. Retrotransposons that maintain chromosome ends. *Proc Natl*
973 *Acad Sci U S A* **108**: 20317-20324.
974 Podlevsky JD, Chen JJL. 2016. Evolutionary perspectives of telomerase RNA structure and
975 function. *Rna Biol* **13**: 720-732.
976 Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with
977 profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641-1650.
978 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software
979 Suite. *Trends Genet* **16**: 276-277.
980 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L,
981 Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic
982 inference and model choice across a large model space. *Syst Biol* **61**: 539-542.
983 Rosen M, Edstrom J. 2000. DNA structures common for chironomid telomeres terminating with
984 complex repeats. *Insect Mol Biol* **9**: 341-347.
985 Sakofsky CJ, Malkova A. 2017. Break induced replication in eukaryotes: mechanisms,
986 functions, and consequences. *Crit Rev Biochem Mol* **52**: 395-413.

987 Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the
988 genome. *Nat Rev Genet* **8**: 272-285.
989 Sobinoff AP, Pickett HA. 2017. Alternative Lengthening of Telomeres: DNA Repair Pathways
990 Converge. *Trends Genet* **33**: 921-932.
991 Starnes JH, Thornbury DW, Novikova OS, Rehmeyer CJ, Farman ML. 2012. Telomere-targeted
992 retrotransposons in the rice blast fungus Magnaporthe oryzae: agents of telomere
993 instability. *Genetics* **191**: 389-406.
994 van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. 2003. Transposable elements in
995 mammals promote regulatory variation and diversification of genes with specialized
996 functions. *Trends Genet* **19**: 530-536.
997 Villasante A, Abad JP, Planello R, Mendez-Lago M, Celniker SE, de Pablos B. 2007. Drosophila
998 telomeric retrotransposons derived from an ancestral element that was recruited to
999 replace telomerase. *Genome Res* **17**: 1909-1918.
1000 Watson JD. 1972. Origin of concatemeric T7 DNA. *Nat New Biol* **239**: 197-201.
1001 Wei KHC, Reddy HM, Rathnam C, Lee J, Lin DAN, Ji SQ, Mason JM, Clark AG, Barbash DA.
1002 2017. A Pooled Sequencing Approach Identifies a Candidate Meiotic Driver in
1003 Drosophila. *Genetics* **206**: 451-465.
1004 Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P,
1005 Morgante M, Panaud O et al. 2007. A unified classification system for eukaryotic
1006 transposable elements. *Nat Rev Genet* **8**: 973-982.
1007 Xie W, Donohue RC, Birchler JA. 2013. Quantitatively increased somatic transposition of
1008 transposable elements in Drosophila strains compromised for RNAi. *PLoS One* **8**:
1009 e72163.
1010 Ye C, Hill CM, Wu S, Ruan J, Ma ZS. 2016. DBG2OLC: Efficient Assembly of Large Genomes
1011 Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Sci*
1012 *Rep* **6**: 31900.

1013 Zakian VA. 1989. Structure and function of telomeres. *Annu Rev Genet* **23**: 579-604.

1014 Zakian VA. 1996. Structure, function, and replication of Saccharomyces cerevisiae telomeres.

1015 *Annu Rev Genet* **30**: 141-172.

1016

1017