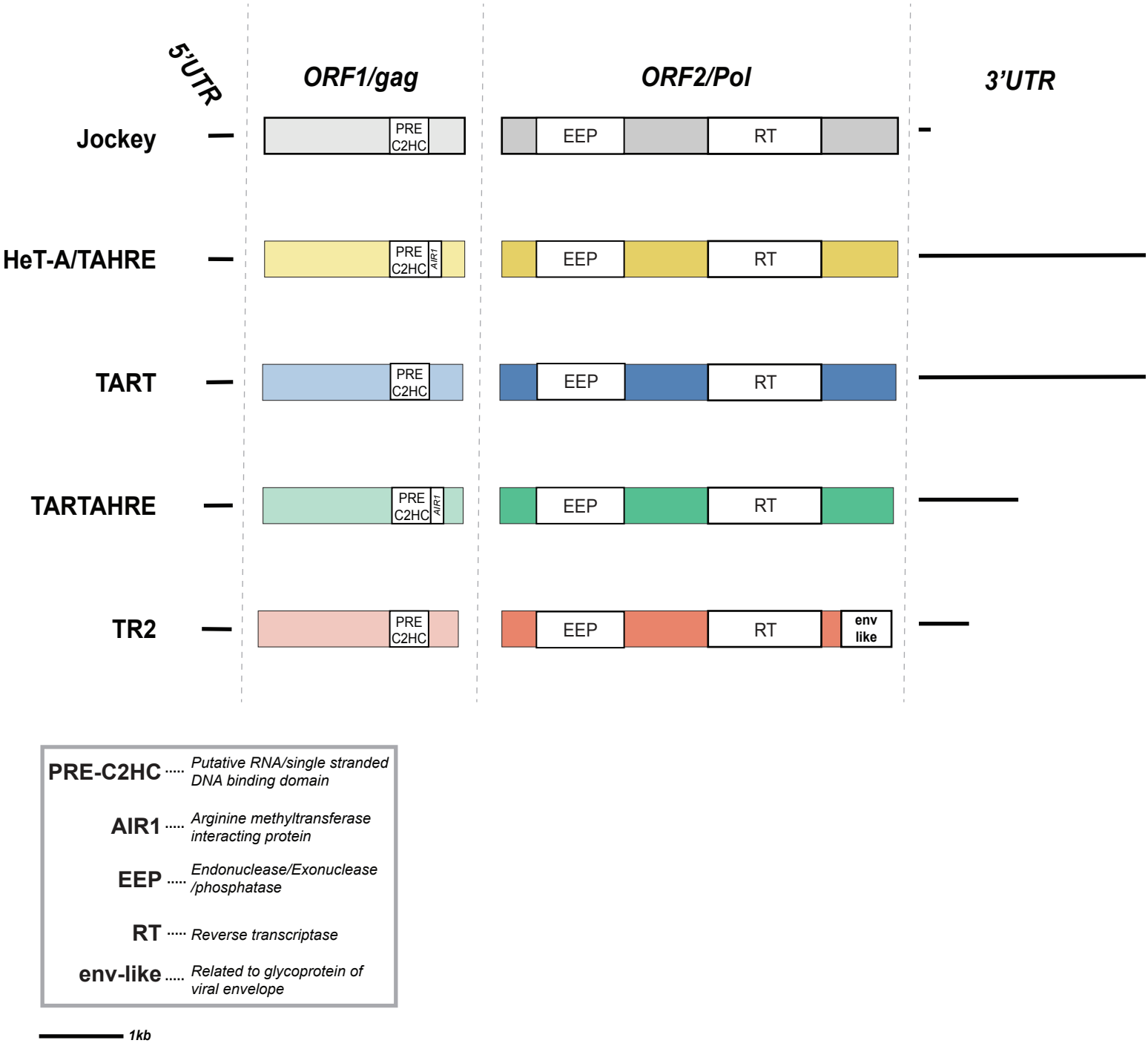

Figure S1. Canonical domain structure of jockey family elements. Output from screening our final consensus sequences for protein domains using the Conserved Domain Database (Marchler-Bauer et al. 2017). Significant protein domains are designated as white boxes along ORF1 (lighter shade) and ORF2 (darker shade).

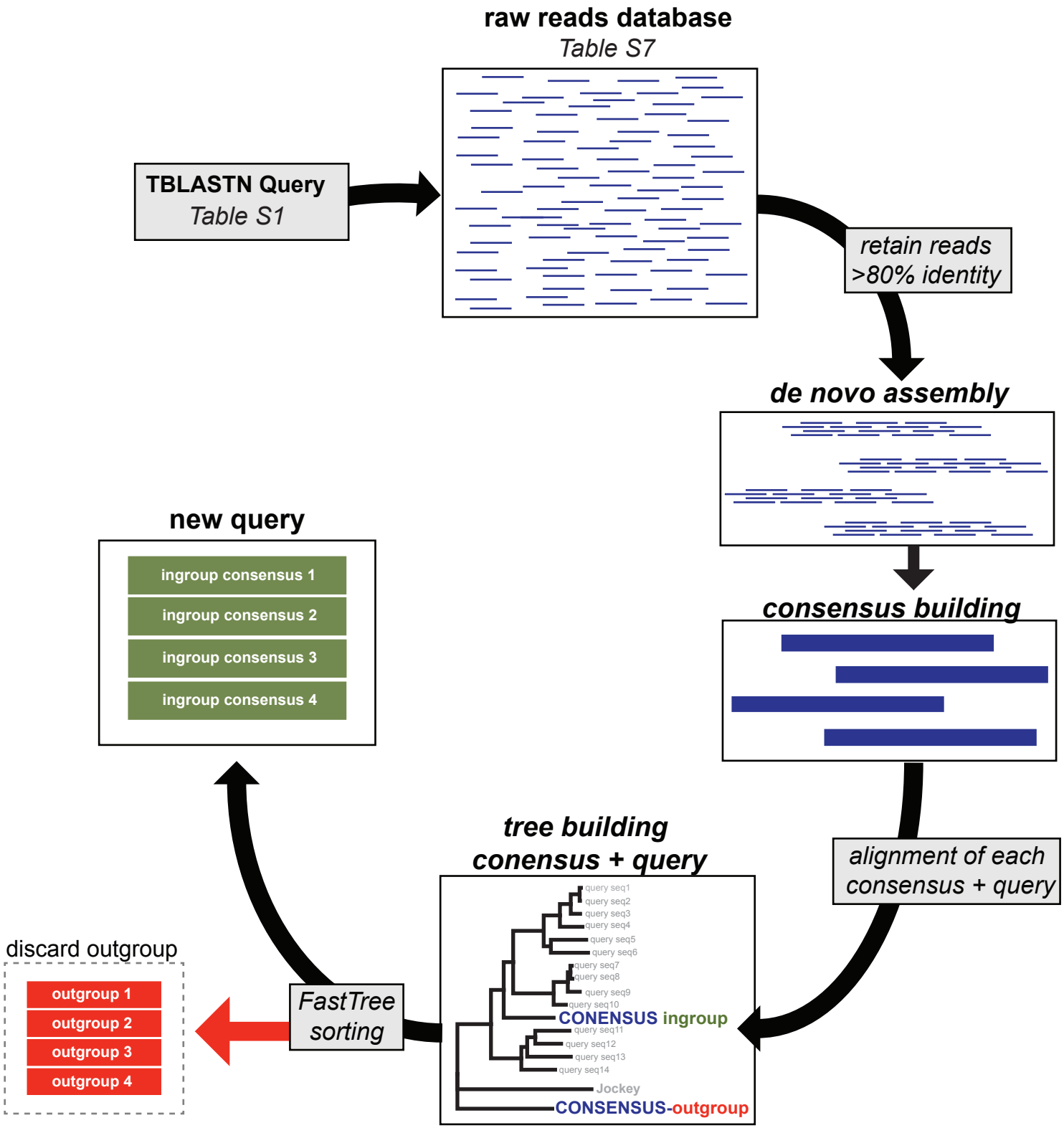

Figure S2. Visual summary of custom pipeline for defining jockey subclade elements. The first round starts with a query compiled from the literature (Table S1) used to search the raw reads of 10 Drosophila genomes. The final step generates “in group” consensus sequences from each species (green), which serve as the new query in the second round.

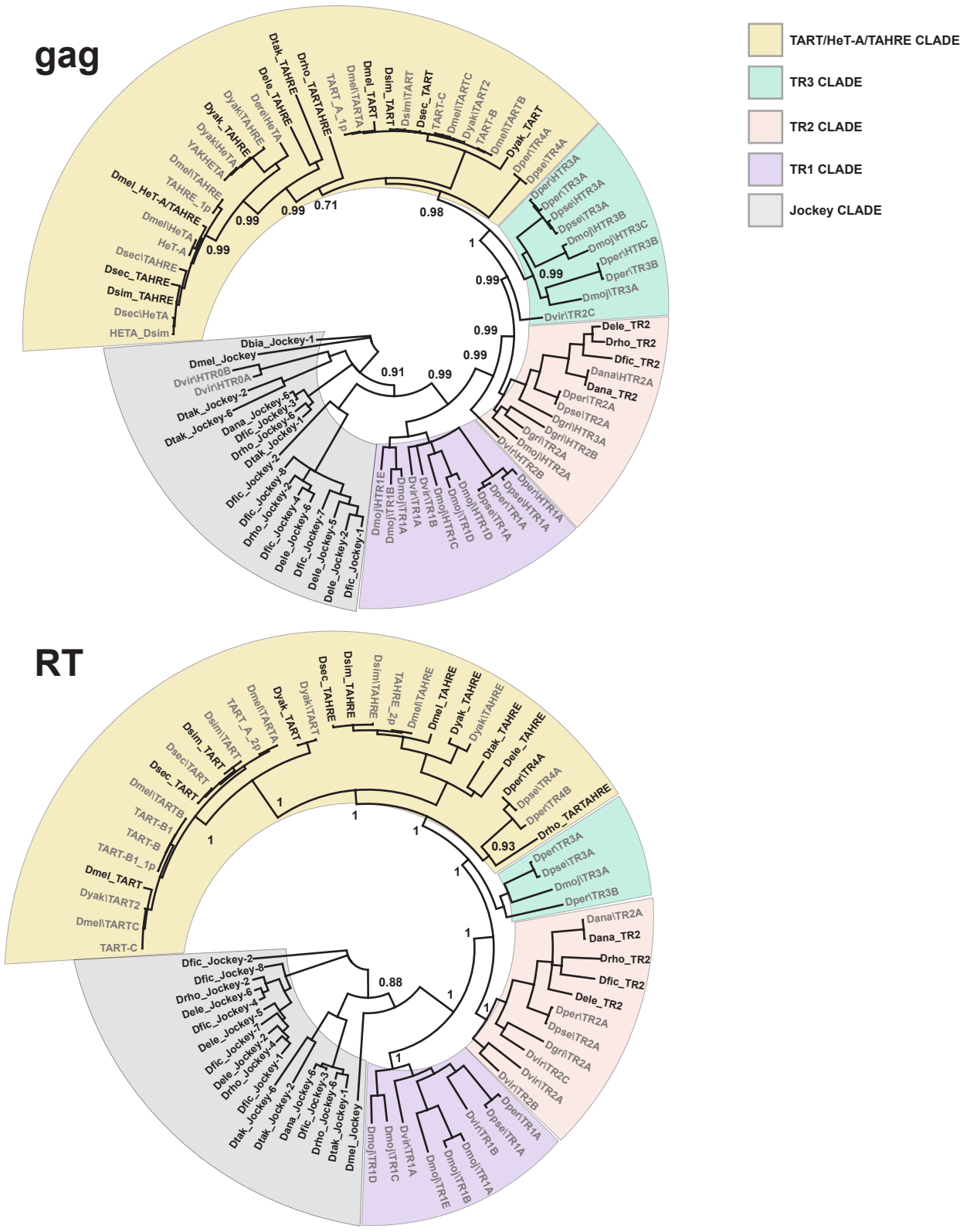

Figure S3. Phylogenetic tree of jockey family consensus sequences reported in Fig. 1 as well as elements described in previous publications. Black font names correspond to consensus sequences generated by our pipeline. Gray font names correspond to elements/consensus sequences previously defined in the literature (Table S1). Node support values are posterior probabilities generated by MrBayes. The gray cloud designates jockey elements distantly related to the telomere-specialized clade. Various colors delineate candidate telomere-specialized elements along with candidate telomeric elements that form monophyletic clades. Of note are the HeT-A and TAHRE consensus gag domains, which are phylogenetically intermingled. We thus refer to the Het-A gag as a HeT-A/TAHRE gag in the text. Moreover, the TART clade is most closely related to TR3, an element encoded by none of our focal species.

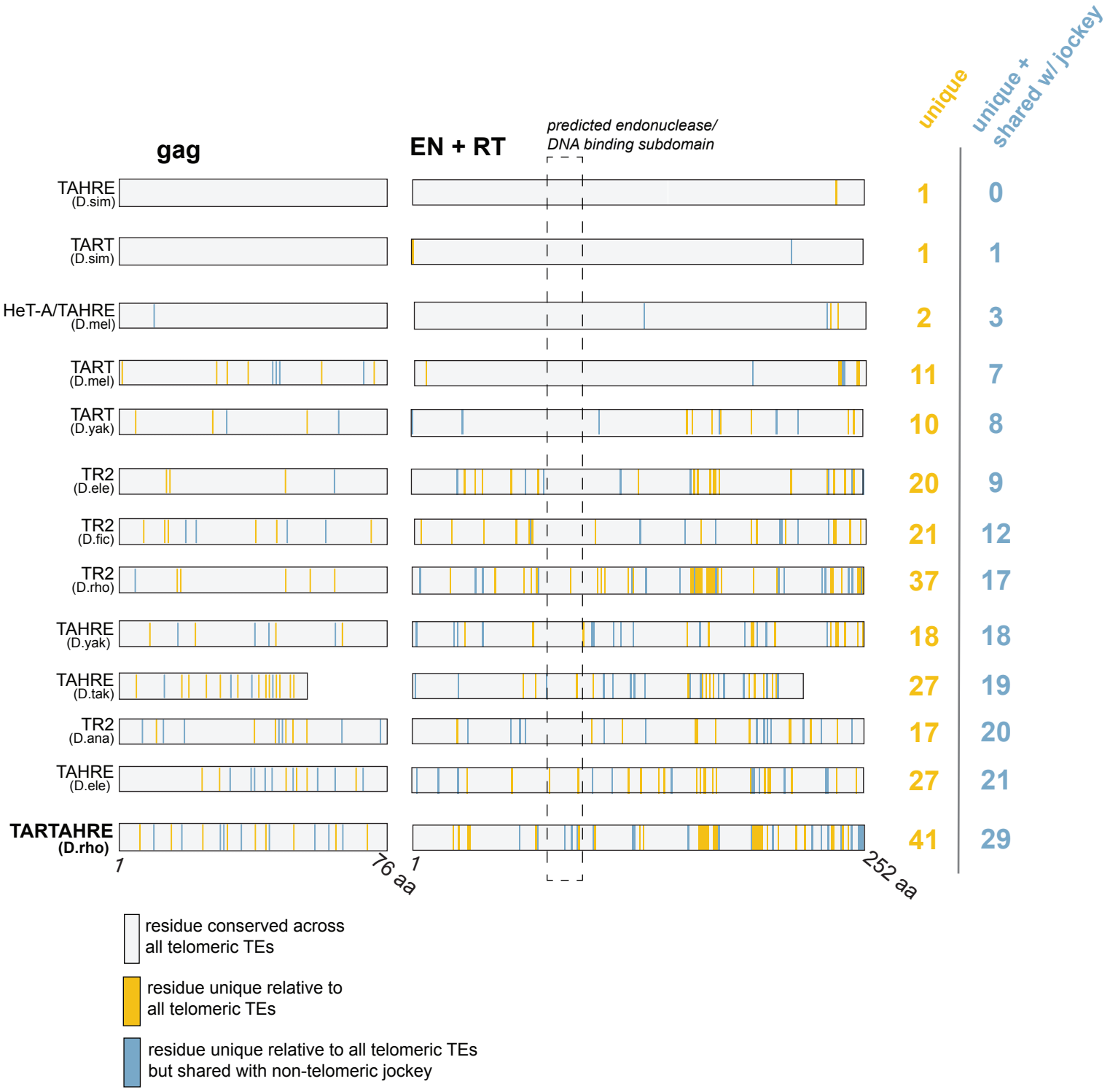

Figure S4.TARTAHRE residues are unusual compared with other telomeric retrotransposons. Graphical representation of residues in the gag, RT, and EN (endonuclease) domains unique to a particular telomeric retrotransposon compared with all other telomeric retrotransposon consensus sequences (yellow) or instead unique residues compared with all other telomeric retrotransposons and shared with a non-telomeric jockey element (blue). Dashed line box delineates to the endonuclease domain predicted to recognize and nick internal sequences (Marchler-Bauer et al. 2017).

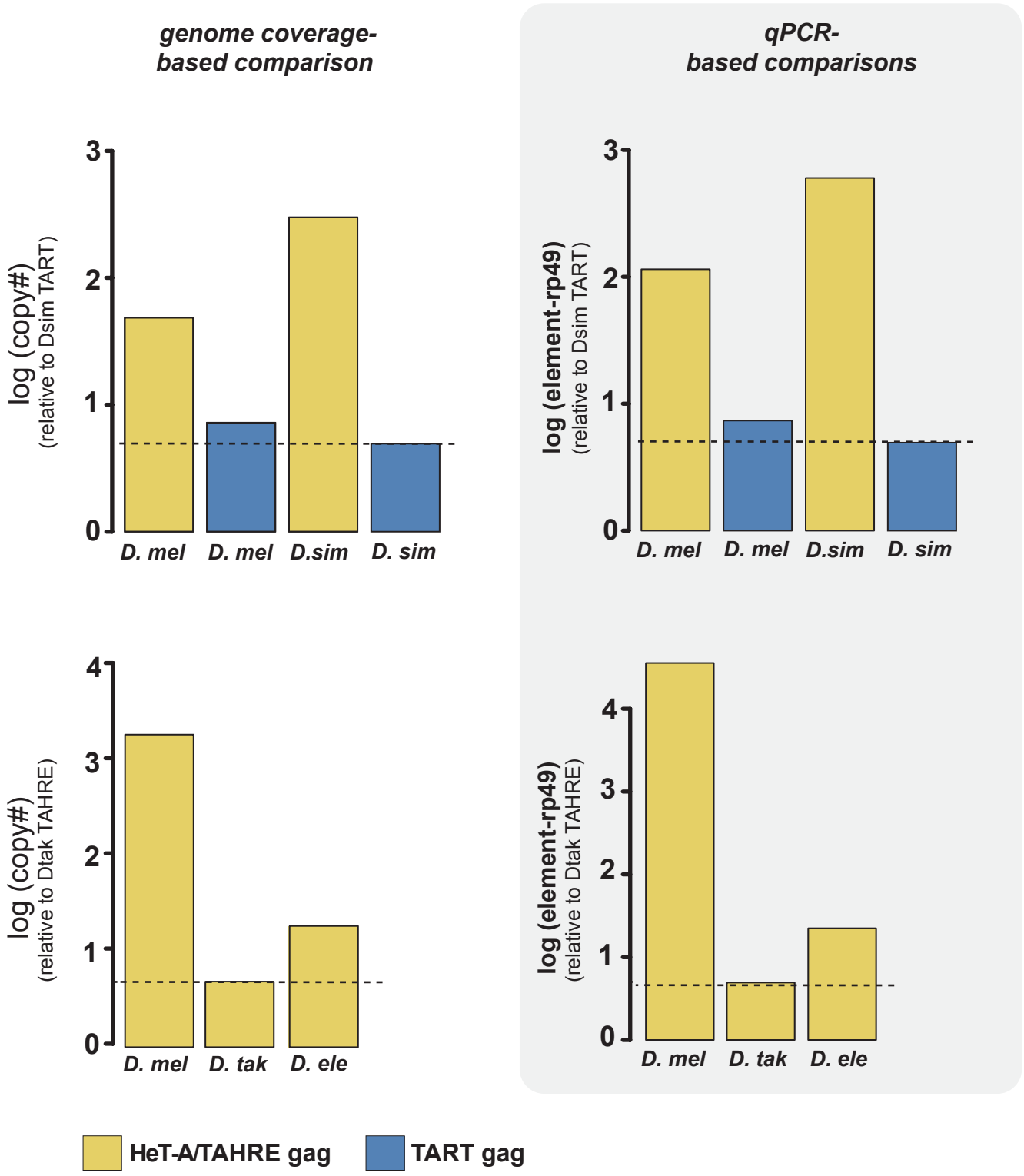

Figure S5. qPCR-based validation of copy number differences inferred from genome data. Inferred copy numbers differences of HeT-A/TAHRE and TART from *D. melanogaster (D. mel), D. simulans (D.sim), D. takahashii (D. tak),* and *D. elegans (D. ele)* species. The left panel illustrates copy number estimated after normalization of genomic reads mapping to the consensus divided by the mean genome coverage. The right panel shows copy number quantified by qPCR using primers that amplify the element normalized by a single copy gene (rp49). Estimates are expressed relative to the lowest copy element (dotted line). Genomic DNA was prepared from females only.

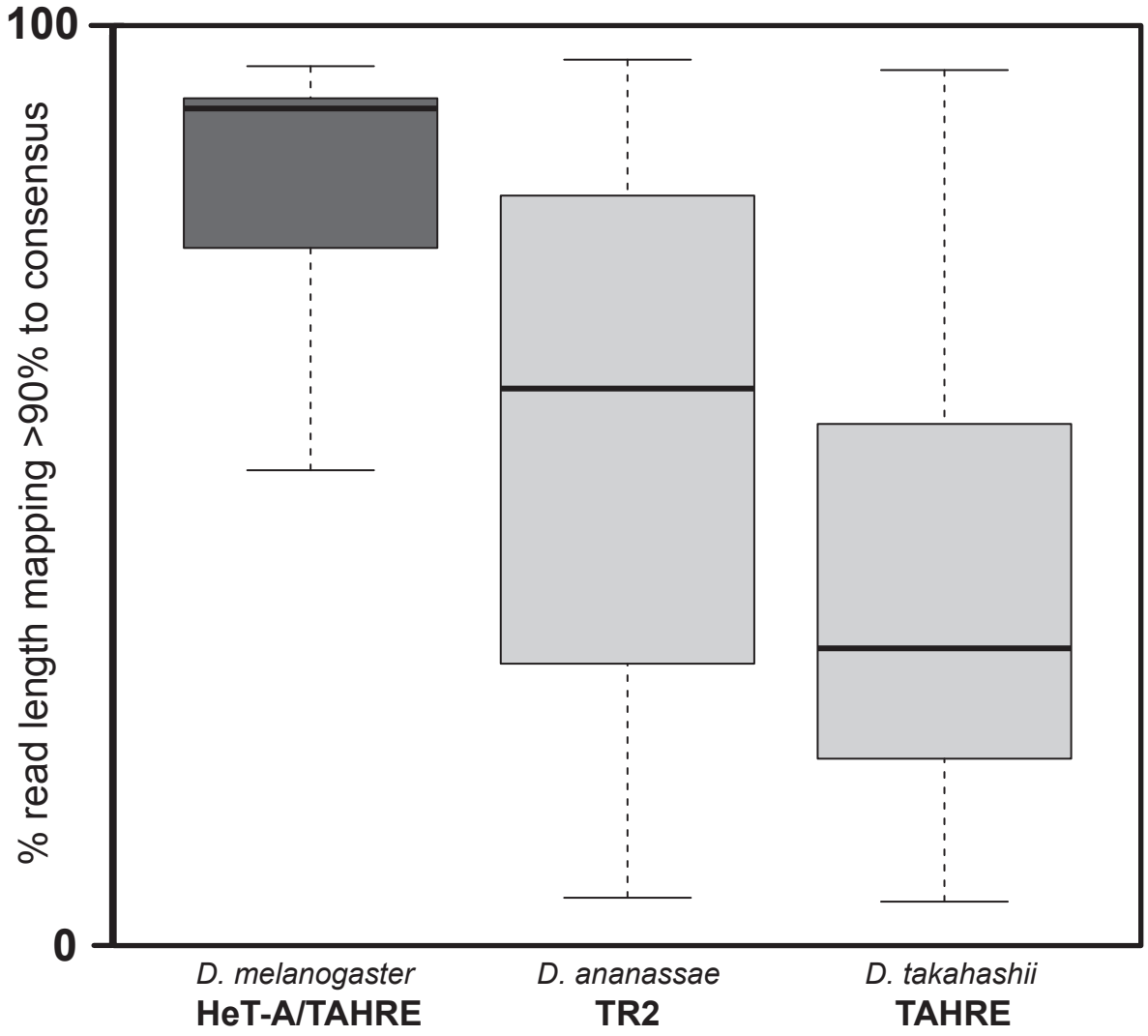

Figure S6. Sequence degradation estimates across an actively transposing elements and those hypothesized to be mostly immobile. Box plots of the percent of total read length captured by the span of the significant BLAST hit to the consensus reveals that *D. takahashii*'s and *D. ananassae*'s telomeric elements (light gray) are frequently juxtaposed with sequence diverged from the consensus. This contrasts with *D. melanogaster* HeT-A/TAHRE, whose reads span most of the significant BLAST hit.

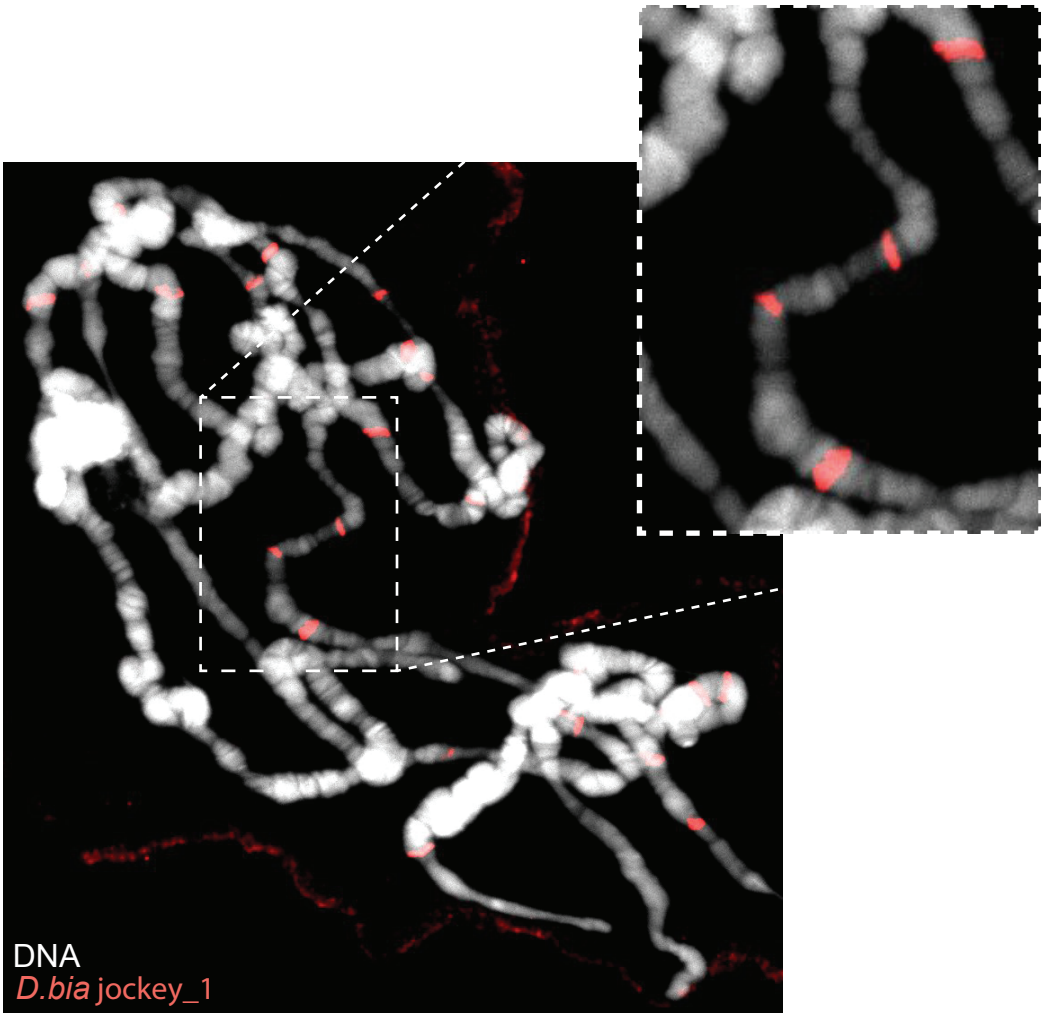

Figure S7. The jockey element from *D. biarmipes* localizes outside the telomere. DNA-FISH with a probe cognate to the 'jockey_1' element (arrow, Fig. 1A) hybridized to polytene chromosomes from *D. biarmipes*. Inset shows banding along a chromosome arm.

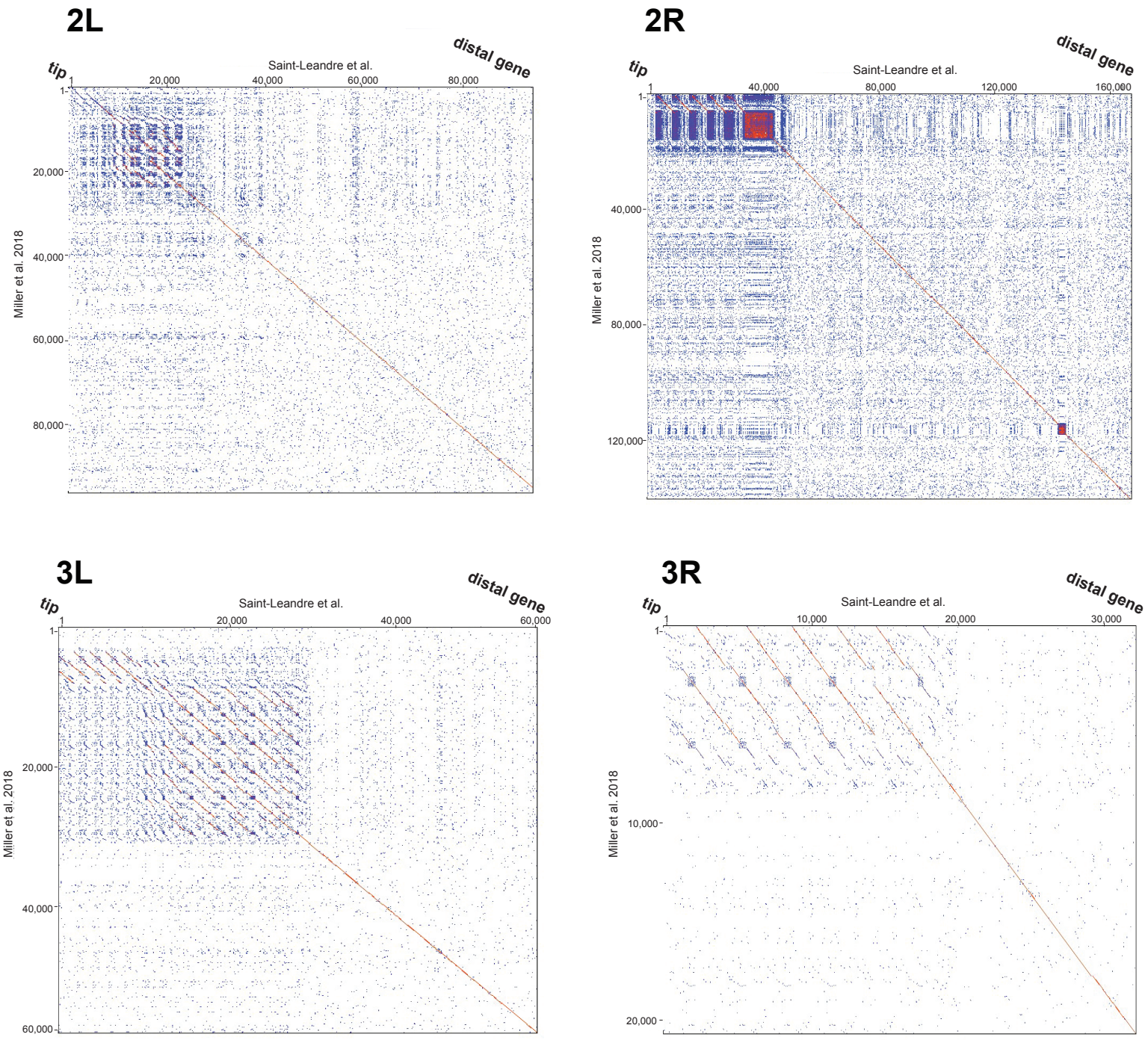

Figure S8. Comparison of *D. biarmipes* telomeric contigs from two independent, long-read-based genome assemblies. Dot matrix of the terminal sequences from two independent genome assemblies of the *D. biarmipes* genome. The Y-axis displays chromosome ends assembled from Oxford Nanopore and Illumina reads reported in (Miller et al. 2018). The X-axis displays chromosome ends from the assembly of PacBio long reads reported in the present study. Red pixels indicate 1:1 alignment over more than 100bp while blue indicates 1:1 alignment <100bp. White indicates no alignment.

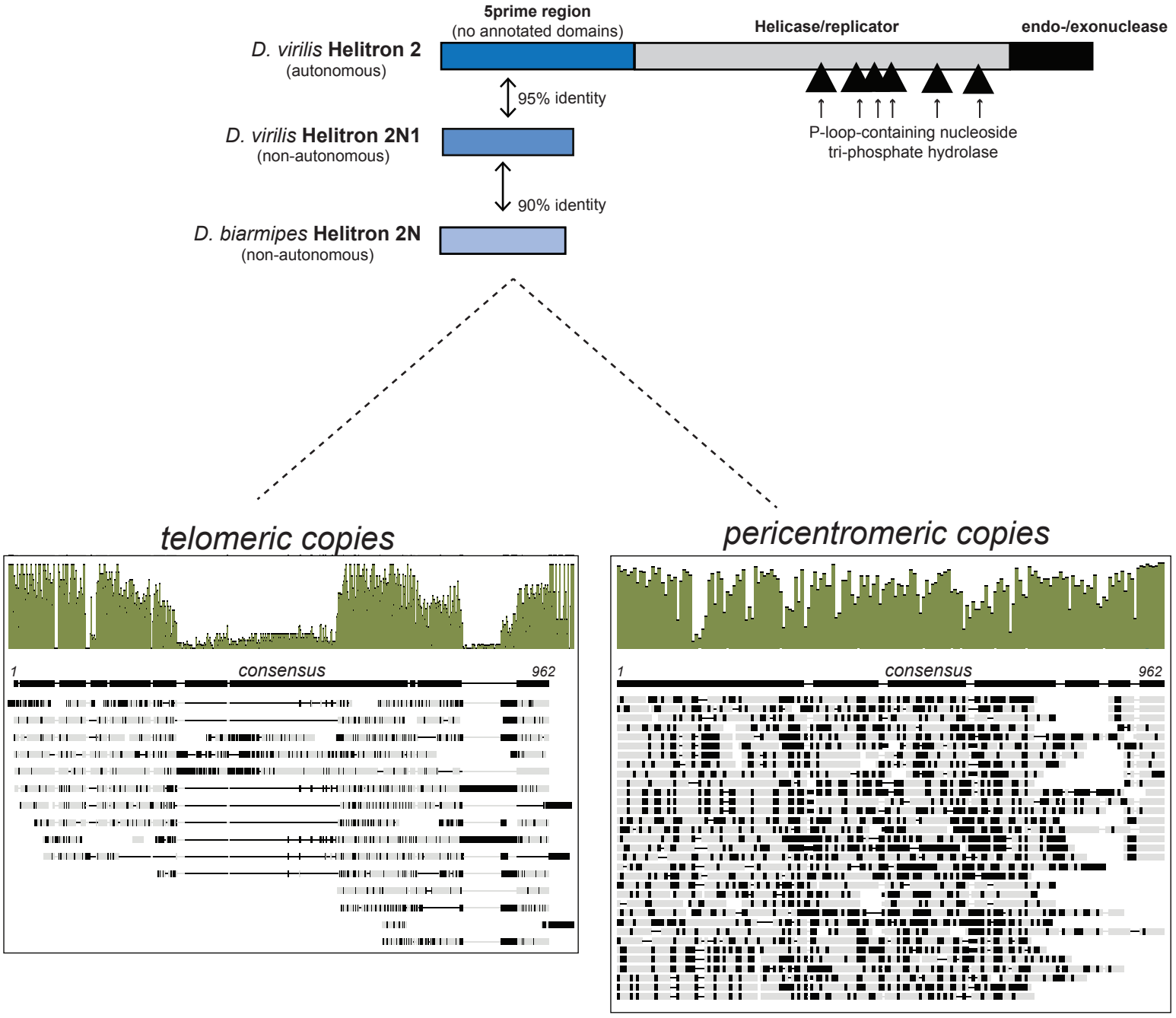

Figure S9. Alignments of Helitron copies from the telomeric and pericentromeric copies in *D. biarmipes*. Schematic representation of canonical full-length Helitron ('Helitron 2') and its non-autonomous derivatives (*D. virilis* 'Helitron 2N1' and *D. biarmipes* 'Helitron 2N'). The autonomous Helitron 2 harbors two conserved domains (gray, black). P-loop containing nucleoside triphosphate hydrolases bind DNA (triangles). No recognizable domains appear in the 5' region homologous to the non-autonomous versions found in *D. virilis* and *D. biarmipes*. The green histograms show the percent identity per nucleotide position along the alignment of Helitron copies to the Helitron 2N reference. Black lines represent divergent nucleotides and gray lines represent identical nucleotides.

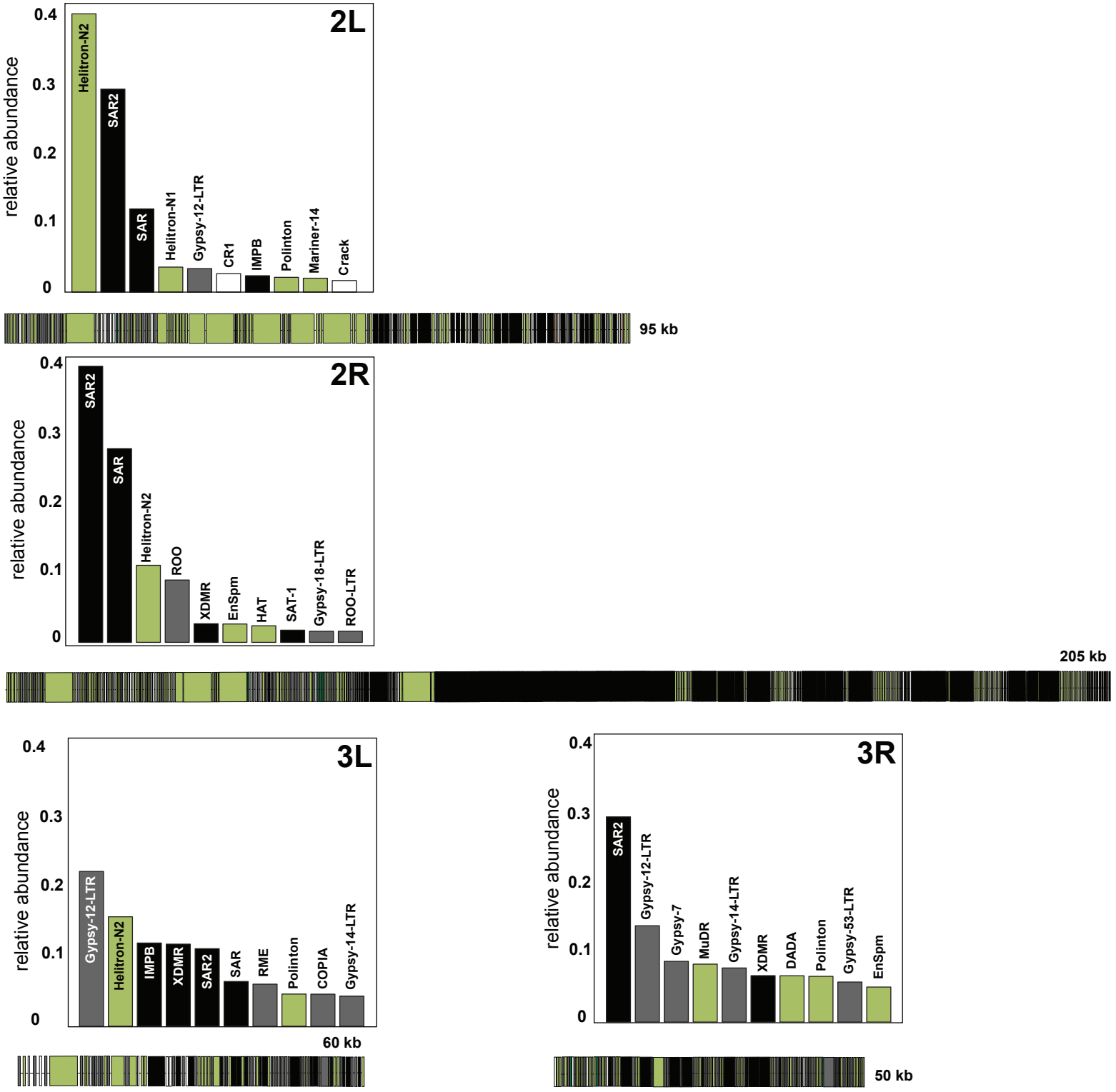

Figure S10. Repetitive DNA composition and physical distribution along *D. biarmipes* telomeres. Relative abundance of the 10 most abundant repeats along *D. biarmipes* telomeres (chromosomes 2R, 2L, 3R and 3L). Rank order of DNA transposons (green), satellite repeats (black), LTR retrotransposons (gray) and non-LTR retrotransposons (white) vary across chromosome arms. Schematic representation of the physical distribution of these repeats by class appears below the bar chart.

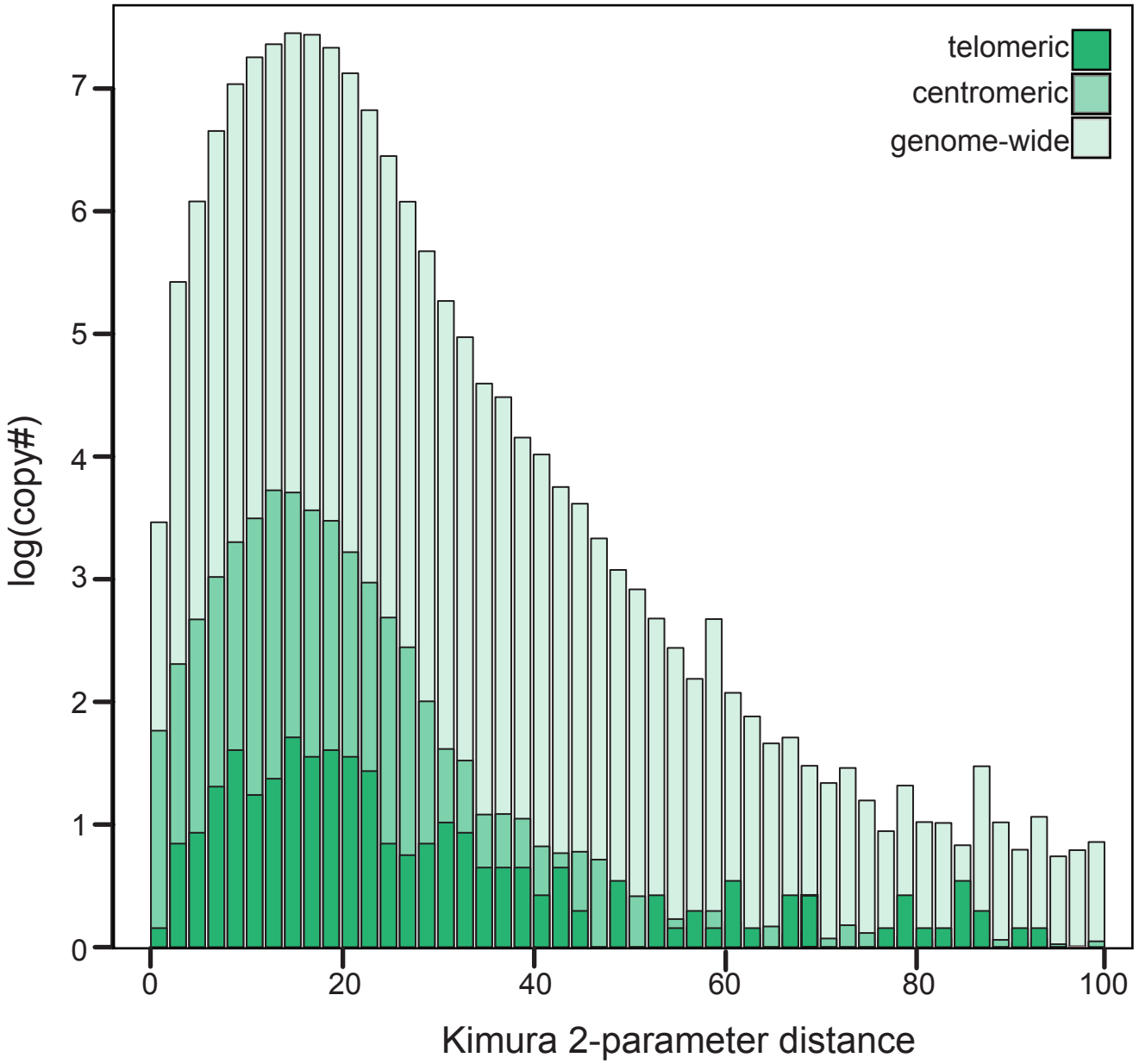

Figure S11. Helitron 2N divergence by chromosome region. Genome abundance (copy number, Y-axis) of pairwise divergence (X-axis, Kimura 2-parameter distance) of *D. biarmipes*' Helitron 2N. The bins closest to one along the X-axis share the highest identity. The pairwise divergence peaks at 20% independent of chromosomal region, consistent with a single ancient transposition burst followed by mutation accumulation.

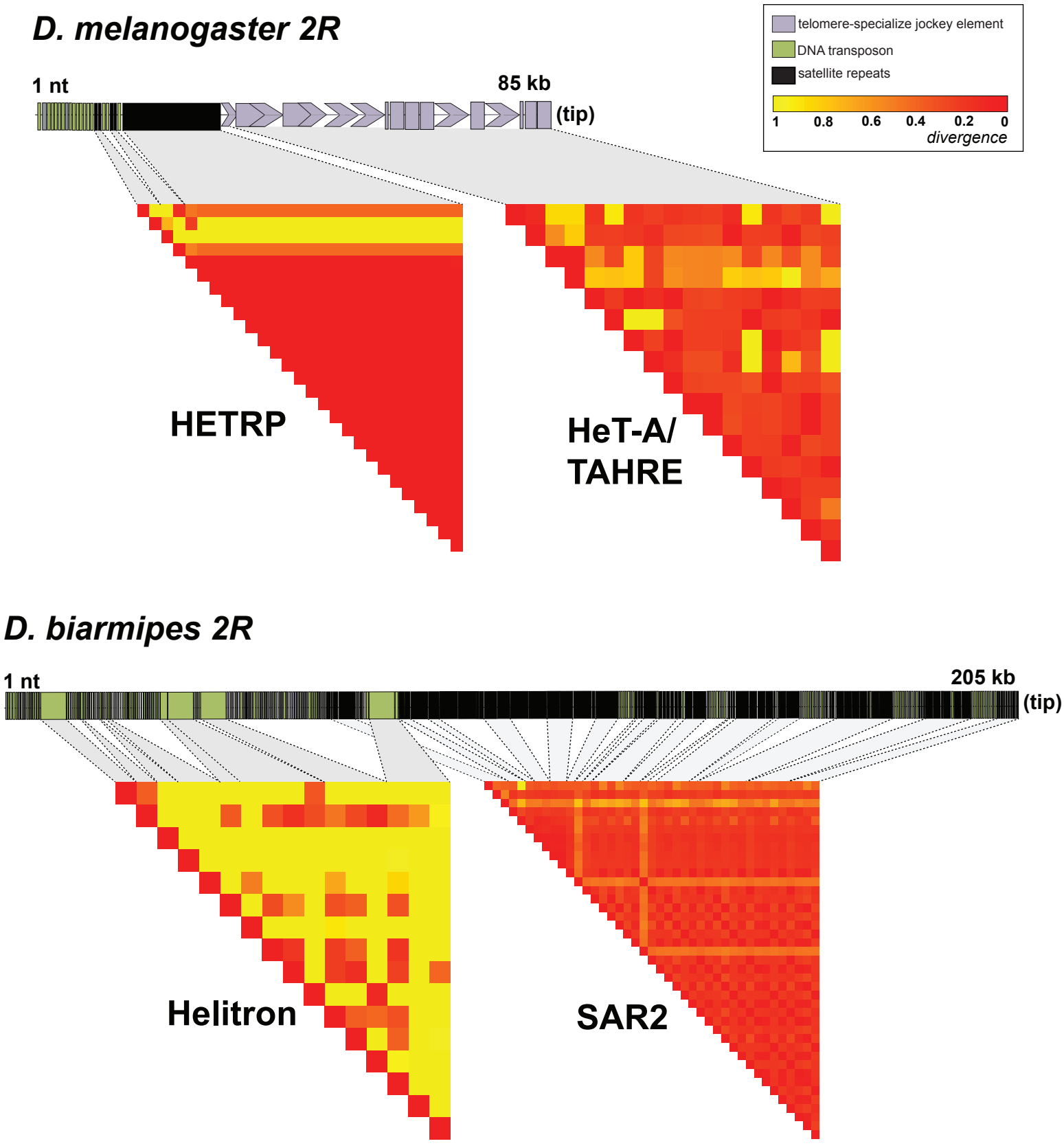

Figure S12. Correlation plots for satellites and mobile elements along telomeres of *D. melanogaster* and *D. biarmipes*. Chromosome 2R satellite blocks (HETRP or SAR2) or concatenated mobile elements (HeT-A/TAHRE gag domain or Helitron) plotted against themselves. Only the Helitron sequence is highly diverged (yellow). Too few TART insertions precluded inclusion of TART in the analysis.

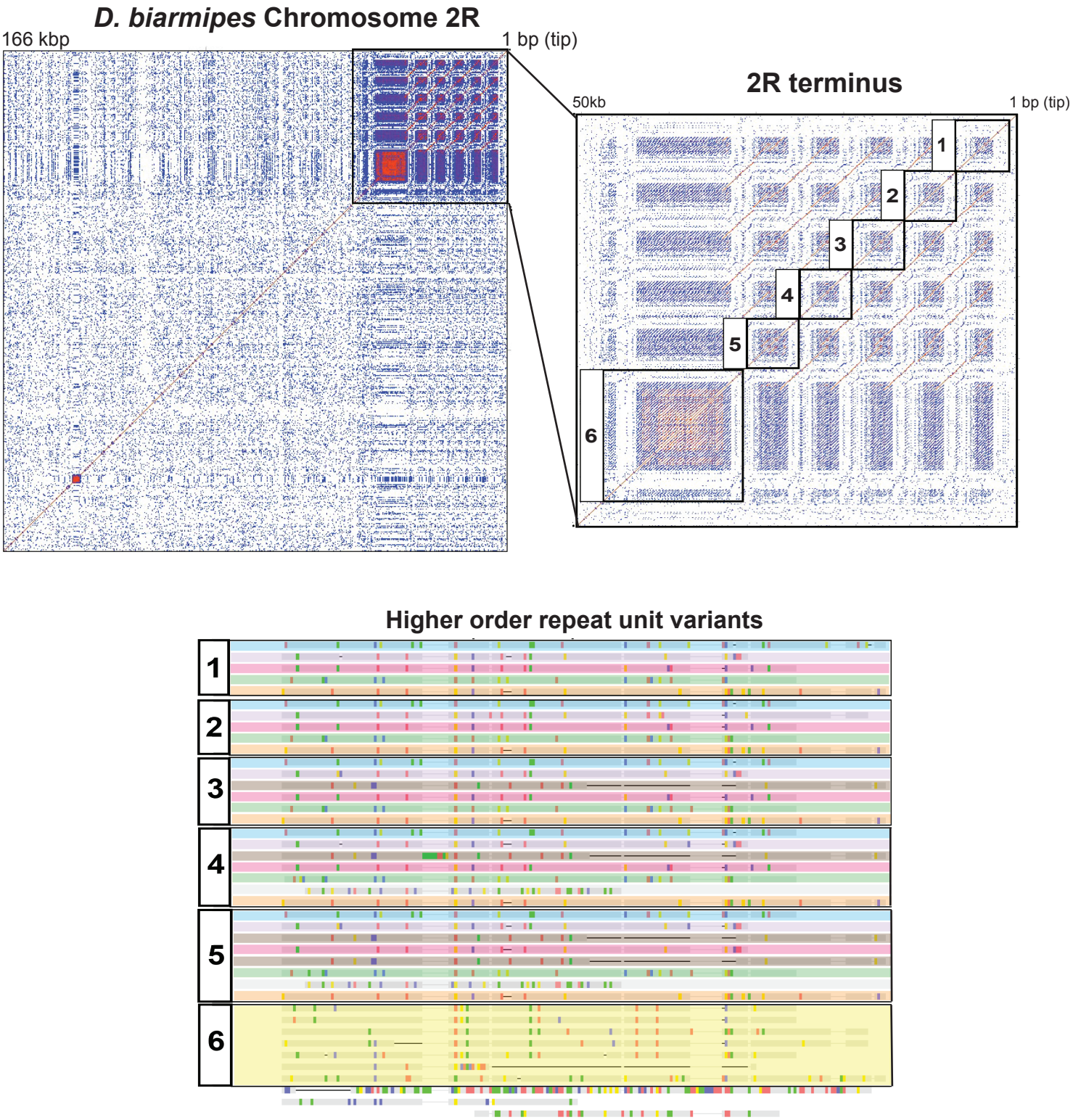

Figure S13. Higher order repeat structure of SAR2 in *D. biarmipes*. Dot matrix of the *D. biarmipes* 2R telomere plotted against itself (left). Blue lines off the main diagonal represent repeat structure along the telomere. The zoom window of the final 50kb reveals the complex structure above the level of single repeat units (right). We delineated six distinct blocks of five SAR2 repeat units sharing high levels of similarity. The five SAR2 variants that make up a higher order unit appear in the five different colors repeated, where the number refers to the most distal (1) to most proximal (5) higher order unit. This pattern breaks down at unit “6” (yellow).

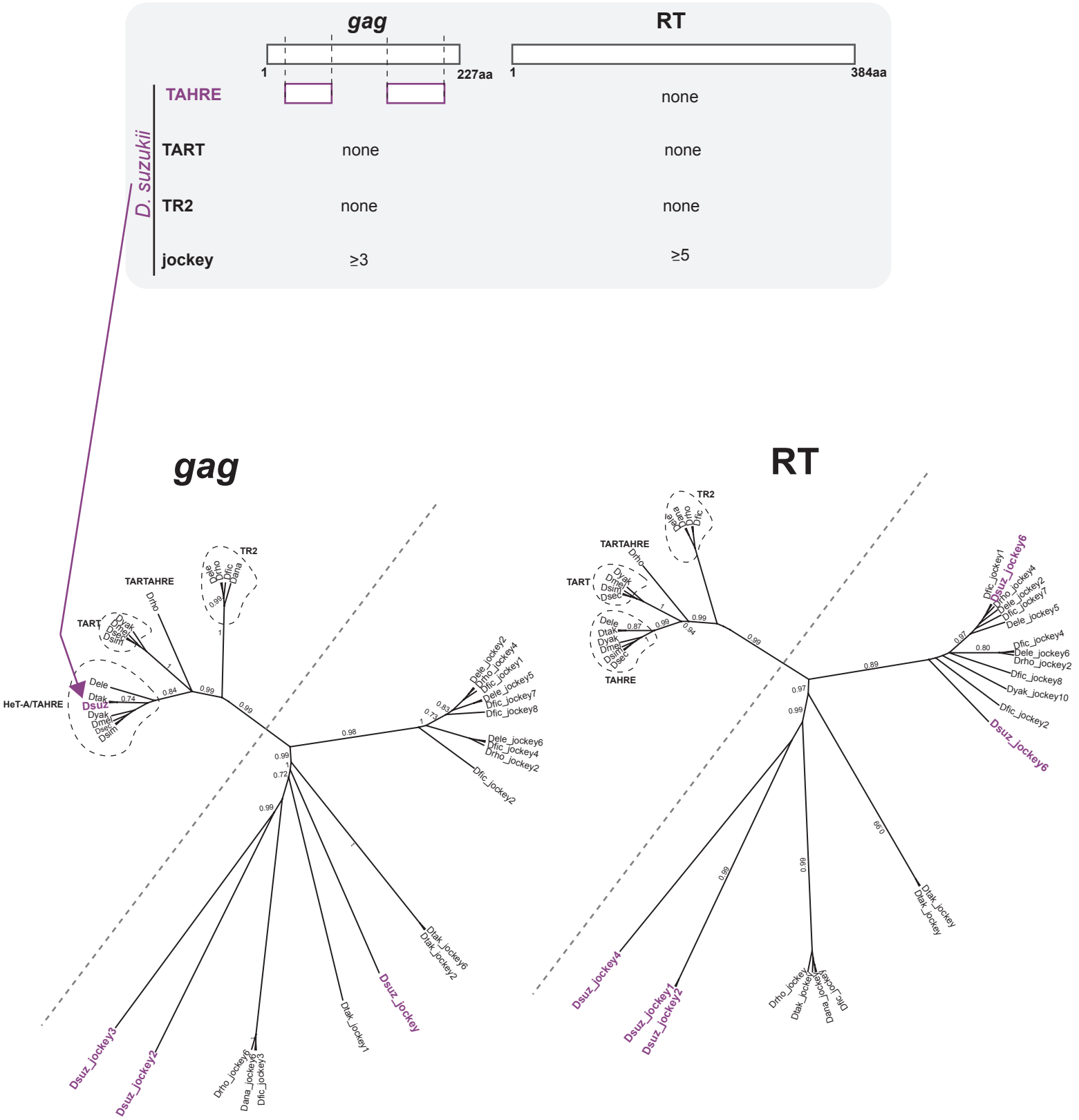

Figure S14. Full length, telomere-specialized jockey subclade retrotransposons were undetectable in the *D. biarmipes* close relative, *D. suzukii*. After discovering an absence of jockey subclade retrotransposons in *D. biarmipes,* we investigated the *D. suzukii* genome* using our custom pipeline. *D. suzukii* retains only remnants of the HeT-A/TAHRE gag lineage (purple boxes) but no RT domain from the telomere specialized jockey subclade. *databases: NCBI SRA: SRR1002946, SRR942799, SRR942803, SRR942804, SRR942797, SRR942800, SRR942802, SRR942805, SRR942798, SRR942803.

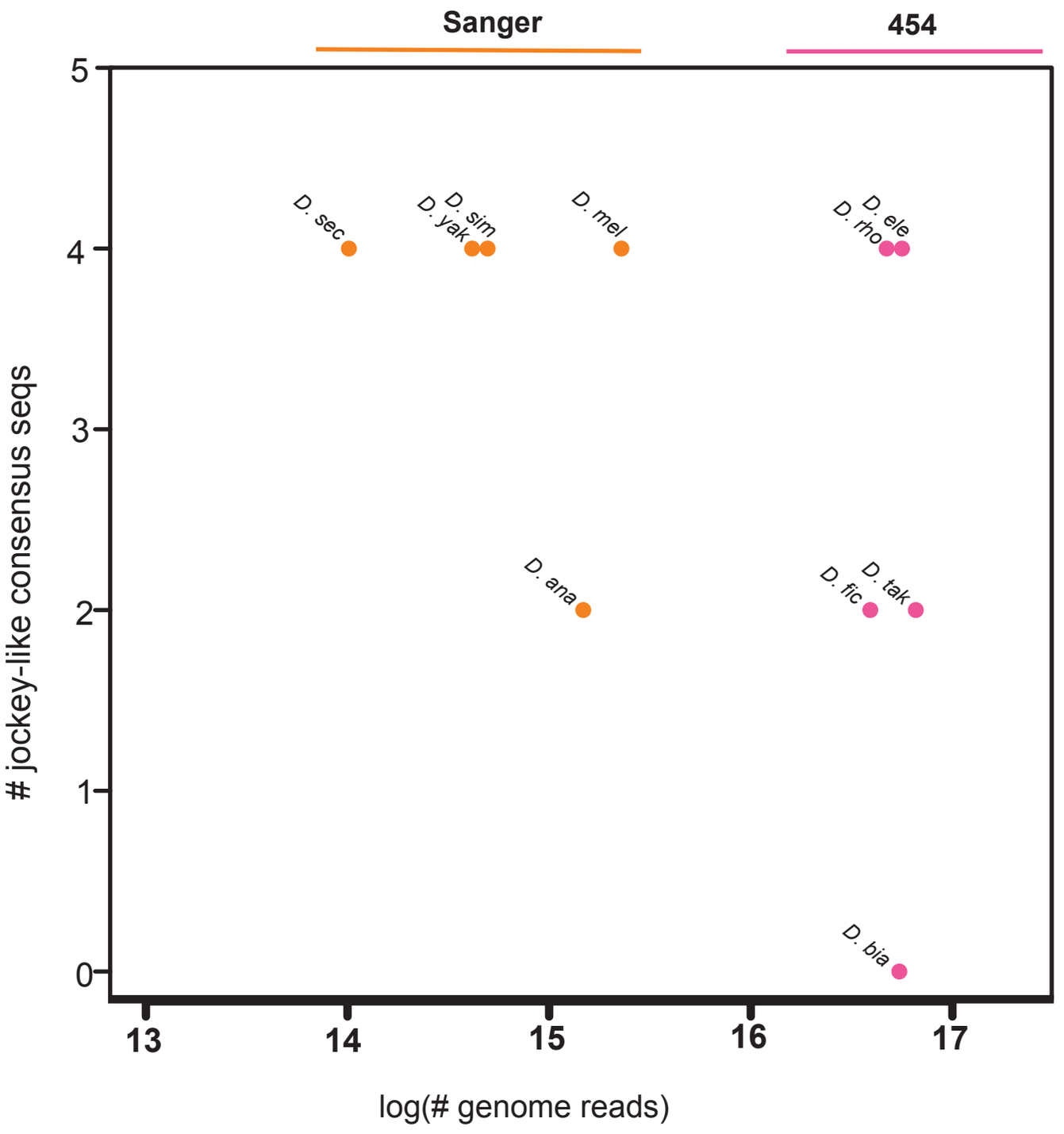

Figure S15. Number of major retrotransposon lineage defined do not correlate with short-read sequencing depth or with sequencing platform. Sequencing projects for each species can be found in Table S7.

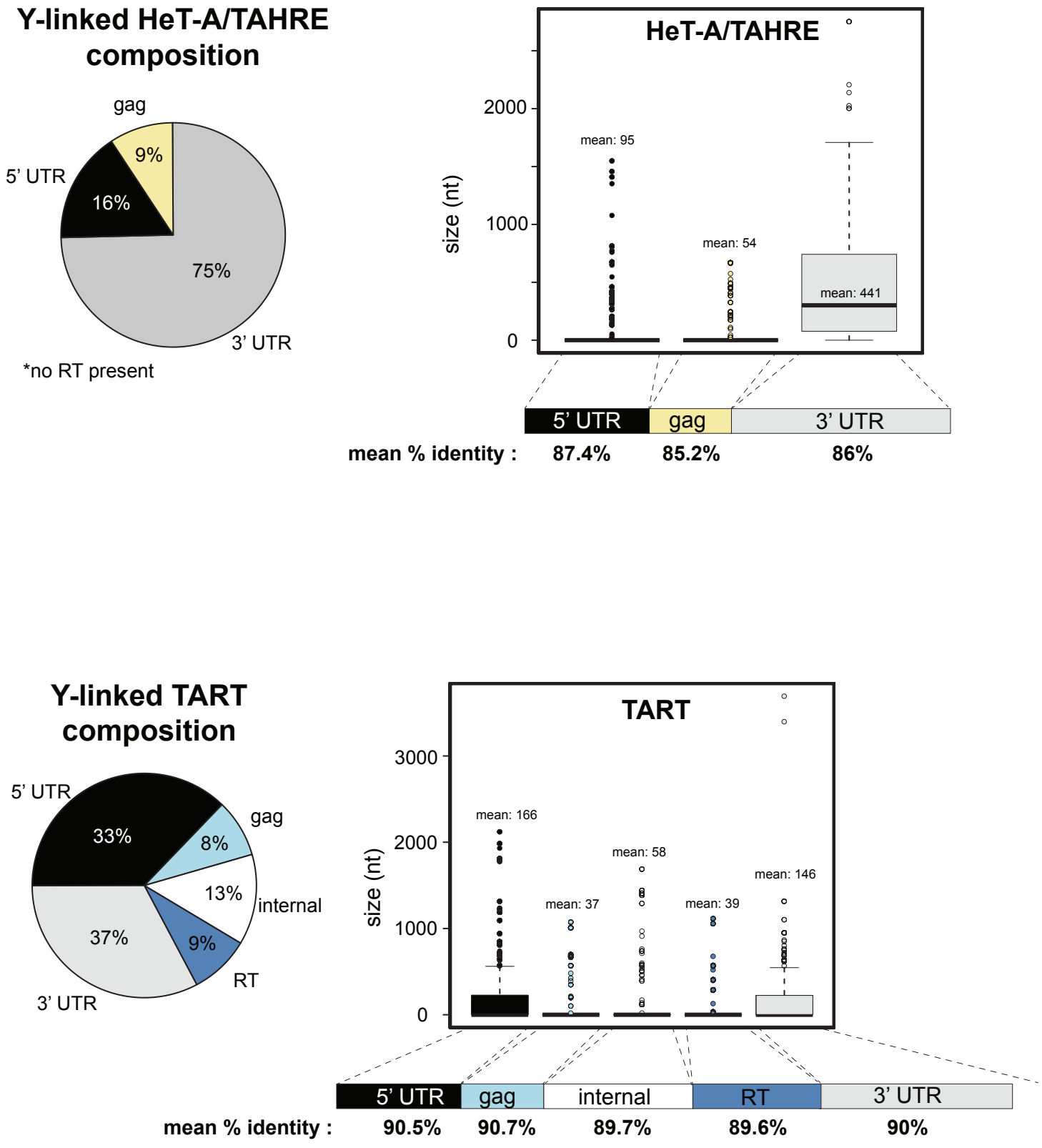

Figure S16. Internal, Y-linked HeT-A/TAHRE do not contribute to our genome-wide copy number estimates. Mapping of *D. melanogaster* raw reads to a recently assembled Y chromosome (Chang and Larracuente 2019) reveal the dominance of 3' HeT-A/TAHRE inserts and 5' and 3' UTR TART sequences instead of gag and RT domains, from which we estimated copy number.