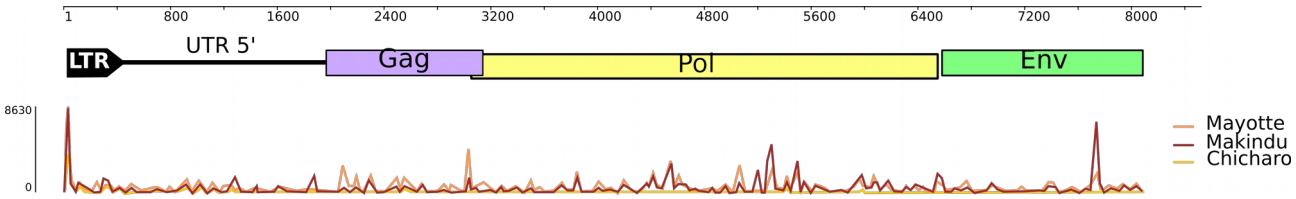

Figure S1 Mapping of piRNA reads against *tirant* reference sequence.

The upper part represents *tirant* structure, to scale (in bp). The lower part depicts read coverage along *tirant* sequence (subfamily C), obtained from samples of 40,000,000 small RNA reads, which were further size-selected to be 23-30 nt in length. We removed LTR 3' to get rid of multi-mapping issues.

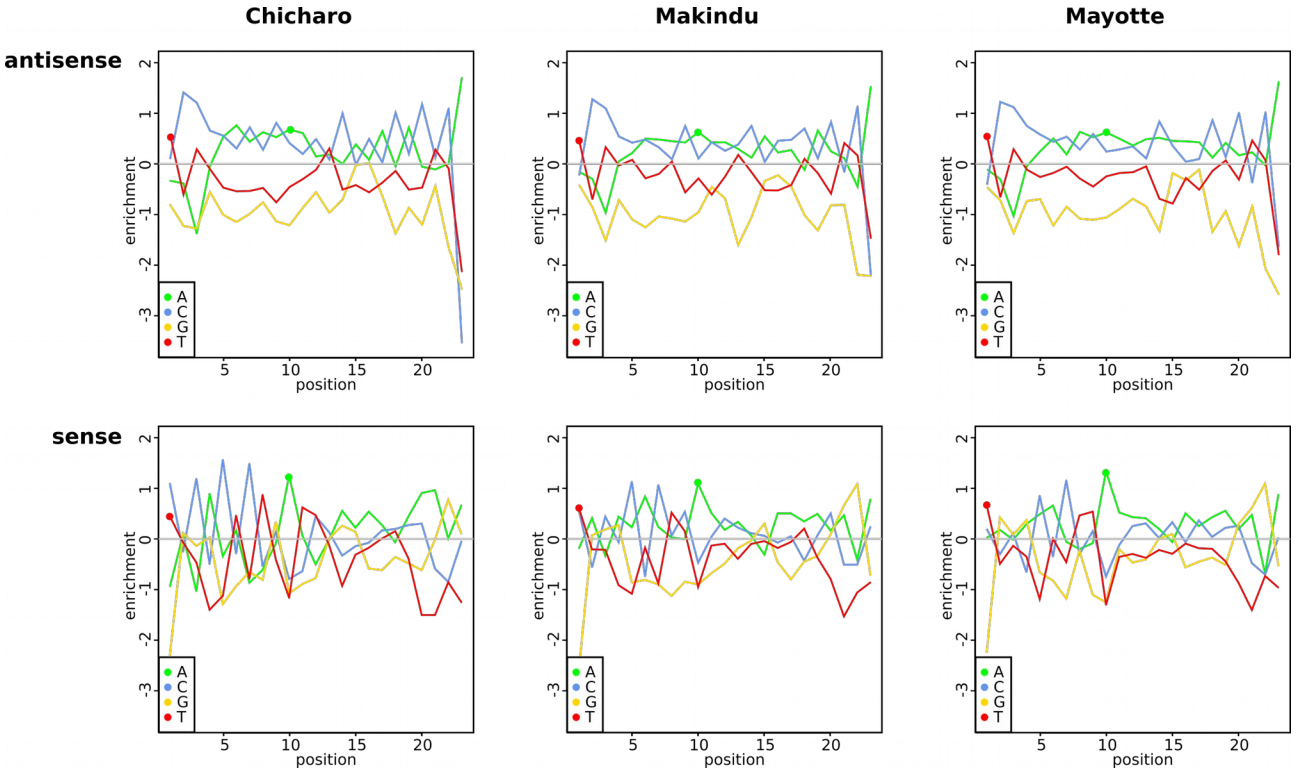

Figure S2 Per position base composition for piRNAs aligned against *tirant*.

Shown are log2 odds ratios comparing emission probabilities in match states to background nucleotide probabilities, as provided by SAMStat. Values above 0 indicate positional enrichment of a particular nucleotide. 1U and 10A are highlighted by colored dots.

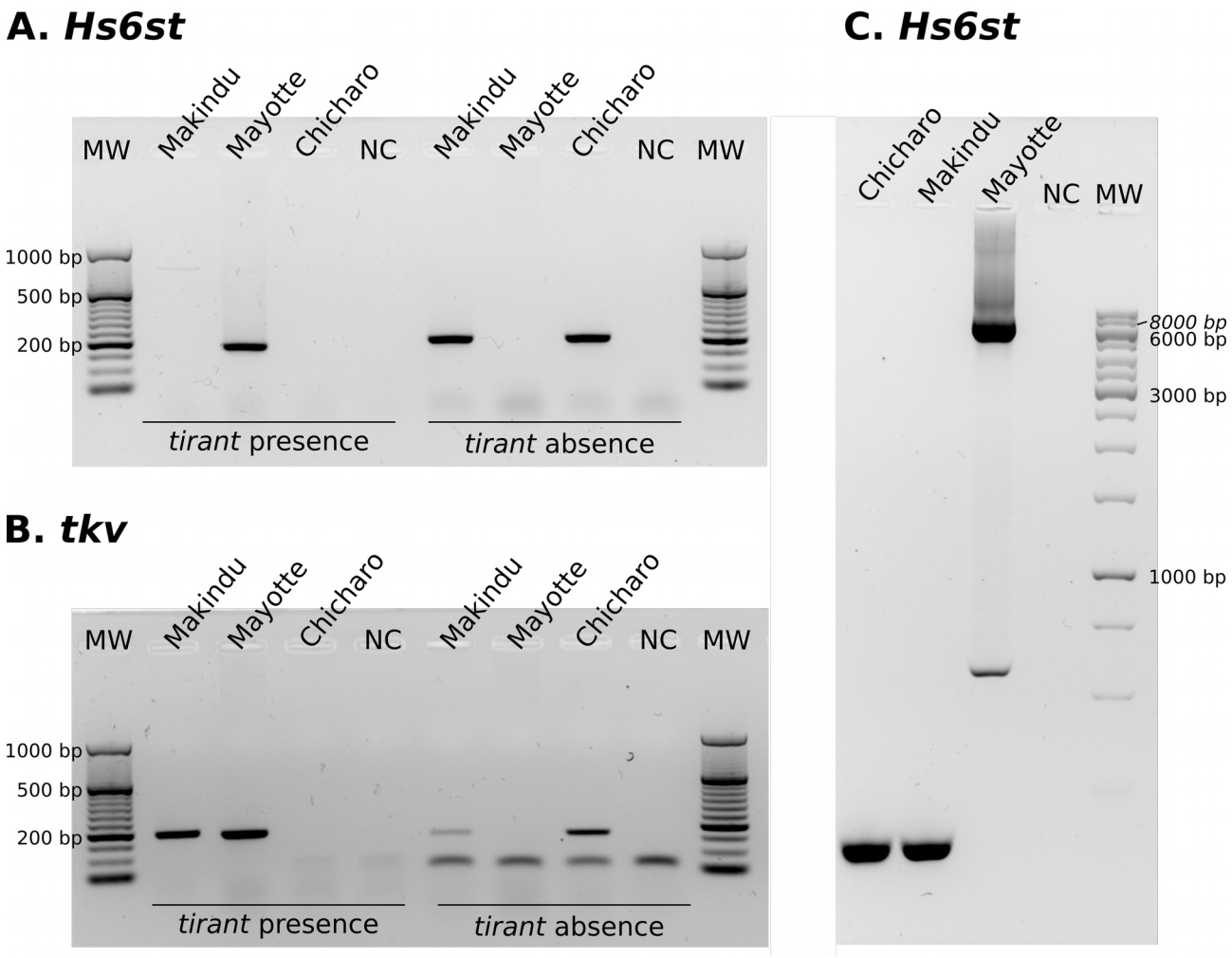

Figure S3 PCR validations for *tirant* insertions into *Hs6st* and *tkv* genes.

MW: Molecular Weight, NC: Negative Control. **A** and **B.** *Tirant* presence/absence PCR tests for *Hs6st* and *tkv* genes, respectively. "*tirant* presence": the PCR reaction is expected to result in a band only if *tirant* is present (the primers used correspond to insertion site with *tirant* in Suppl Mat S1). "*tirant* absence": the PCR reaction is expected to result in a band only if *tirant* is absent (the primers used correspond to insertion site without *tirant* in Suppl Mat S1). **C.** Long PCR amplifying full *tirant* insertion within the *Hs6st* gene.

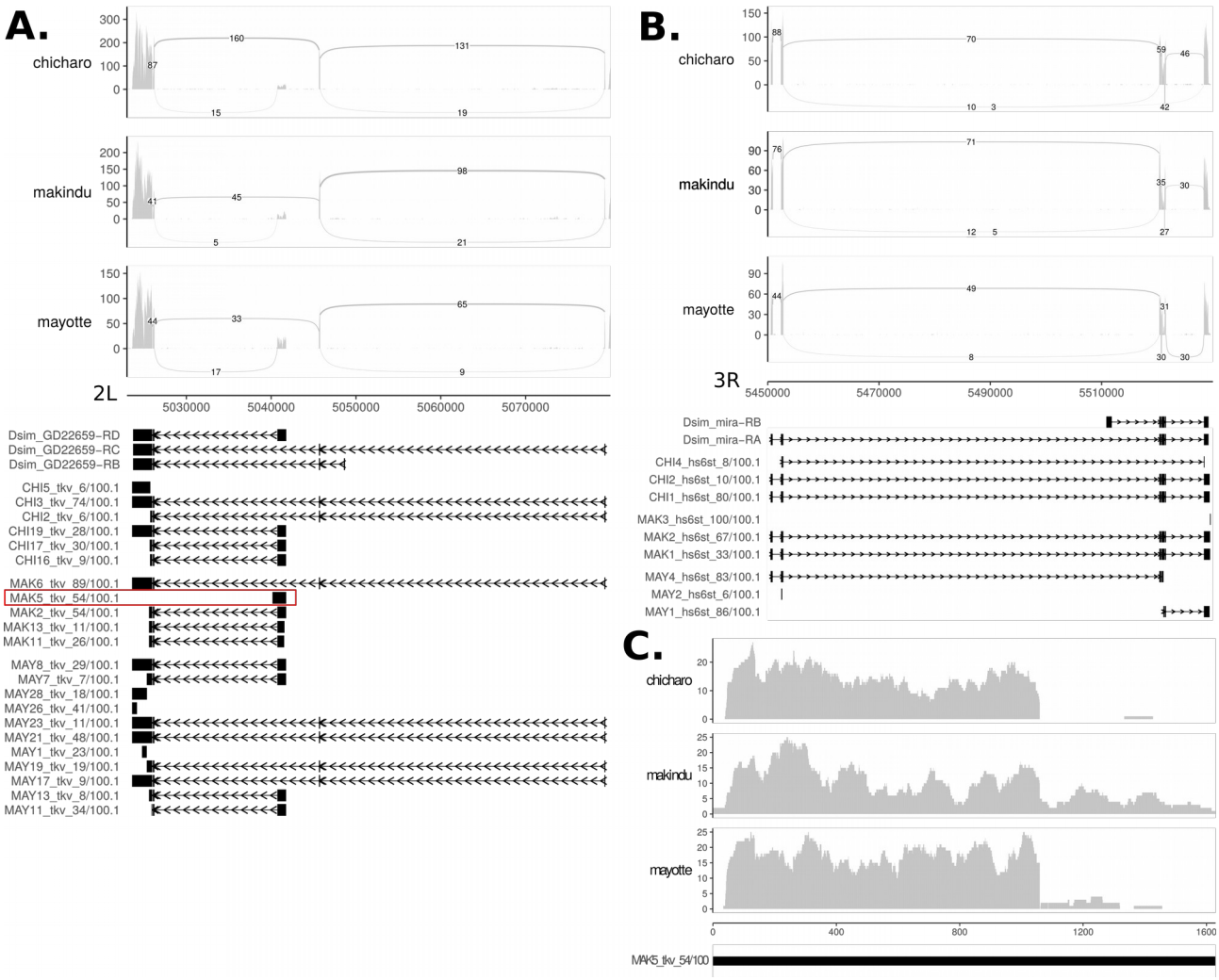

Figure S4 Sashimi plots for *Hs6st* and *tkv* splicing.

Sashimi plots of regions **A.** 2L:5023000-5080000 (*tkv*) and **B.** 3R:5450000-5530000 (*Hs6st*) obtained by ggsashimi. For each panel, the top subplot is the read coverage related to the position on the reference genome. Lines highlight supported junctions and numbers are the numbers of reads supporting the junction. The bottom panel is the gene annotation of each corresponding genomic region. Transcripts labeled “Dsim” come from the reference genomic annotation (ASM75419v3.41), and transcripts labeled “CHI”, “MAK”, and “MAY” are transcripts reconstructed by apytram for strains Chicharo, Makindu, and Mayotte, respectively. As the read coverage is low for these genes, there is stochasticity in the outputs inherent to BLAST and Trinity, the major components of apytram. To consider this issue, we run apytram 100 times and we only reported transcripts observed more than five times (the number of times each transcript was observed among the 100 iterations is mentioned at the end of its name).

MAK5_tkv is a Makindu-specific transcript, which is not described in the reference genome annotation. It does not result from a splicing event, therefore it does not appear in the junctions shown in B. However, it is responsible for the Makindu-specific profile in **C.** We checked that the unshared part of this transcript was not found in Chicharo and Mayotte transcriptomes. In addition, the reads supporting the unshared part do not map anywhere else in the reference genome.