Figure S1. Flowchart for the identification and calling presence/absence of TE insertions. We employed a two-step approach, which first calls the presence of TE insertions using TIDAL, followed by determining the presence/absence status of that TE insertion in all genomes using methods developed in Cridland et al. 2013 (see Materials and Methods). When contigs are aligned across TE breakpoints, a TE is confirmed absent in a strain (green box). When there is a contig mapped to TE sequences and a contig mapped to the flanking of TE breakpoint but not across it, a TE is considered potentially present (blue box). All other cases are considered missing data (the orange box on the right). We also determined whether the identified TE family matches that identified by TIDAL. Discrepancies mainly arose for insertions that are not identified in the strain from which TIDAL first called the TE insertion (e.g., TE2 in genome 3 in the cartoon example), and such TE insertions are excluded from the analysis. In the cartoons, different colors of triangles represent different called TE families.

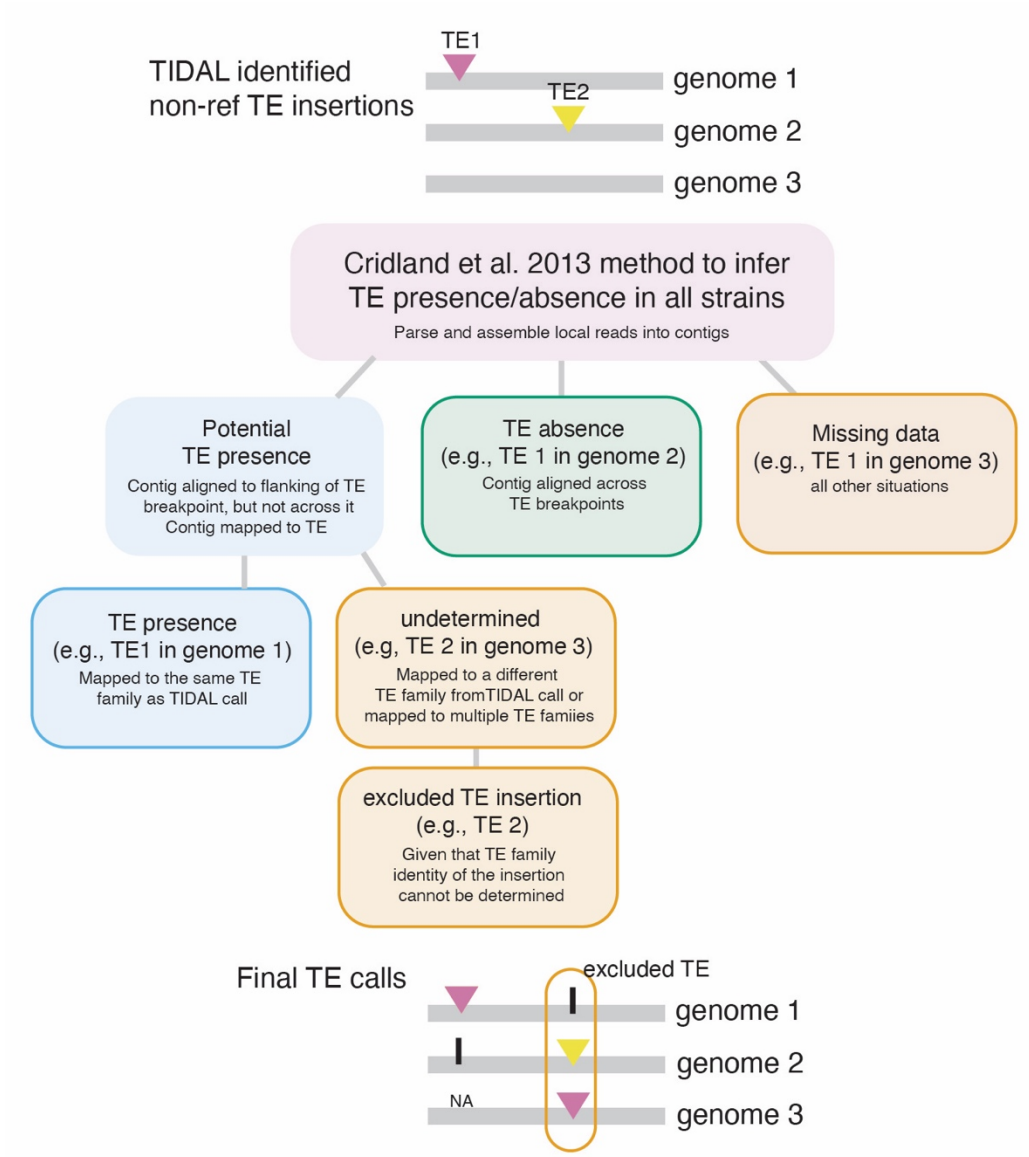

Figure S2. X-Y plot for average sequencing depth and TE burden per genome. There is no significant correlation between average sequencing depth and TE burden across genomes included in our analysis (*Spearman Rank test, ρ = 0.0351, p = 0.6465*). Sequencing depth estimates are from Lack et al. 2015.

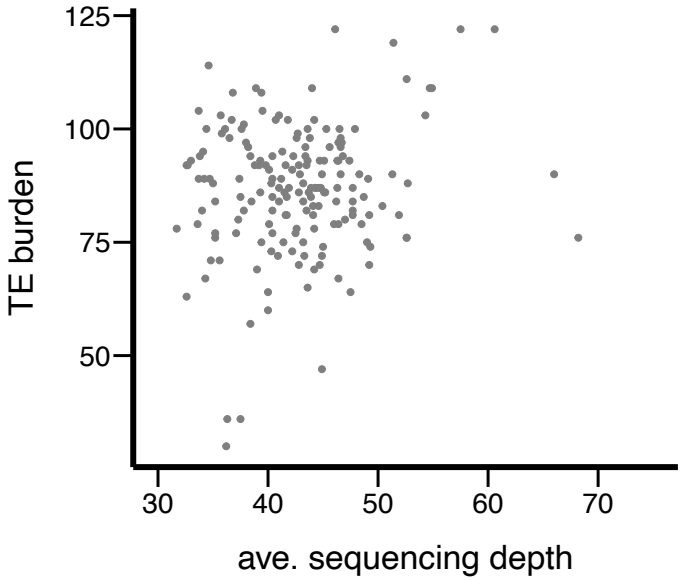

Figure S3. The distribution of missing data per genome.

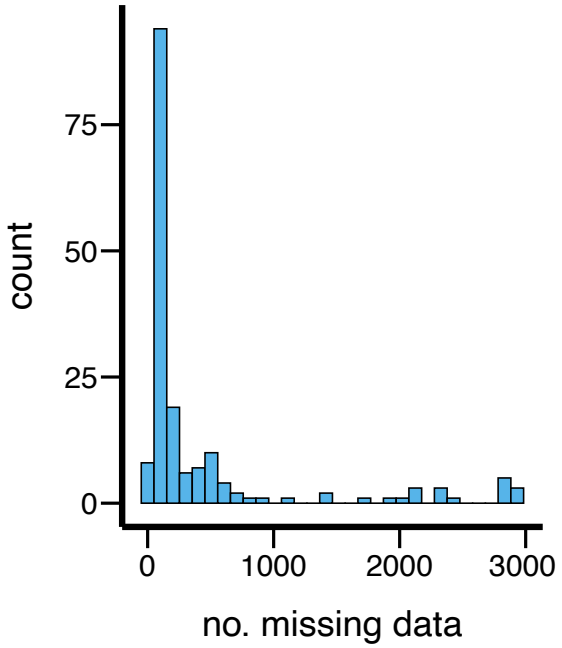

Figure S4. Negative correlations between the amount of missing data and TE burden across genomes.

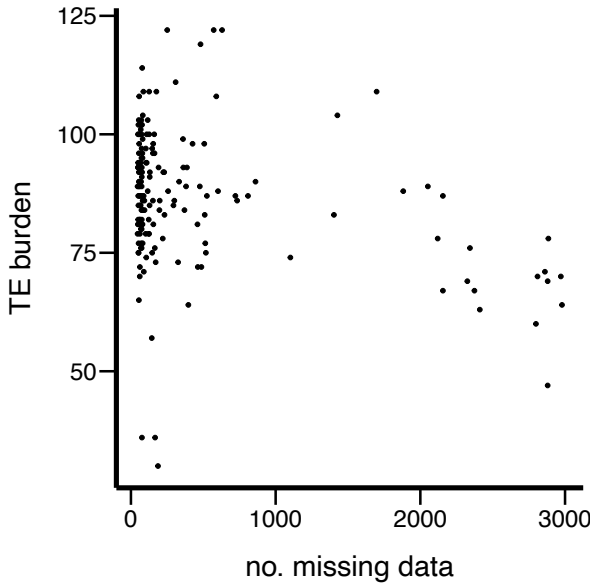

Figure S5. Distribution of mean LD per TE family with respect to different TE family attributes (TE copy number, mean length, and pairwise differences).

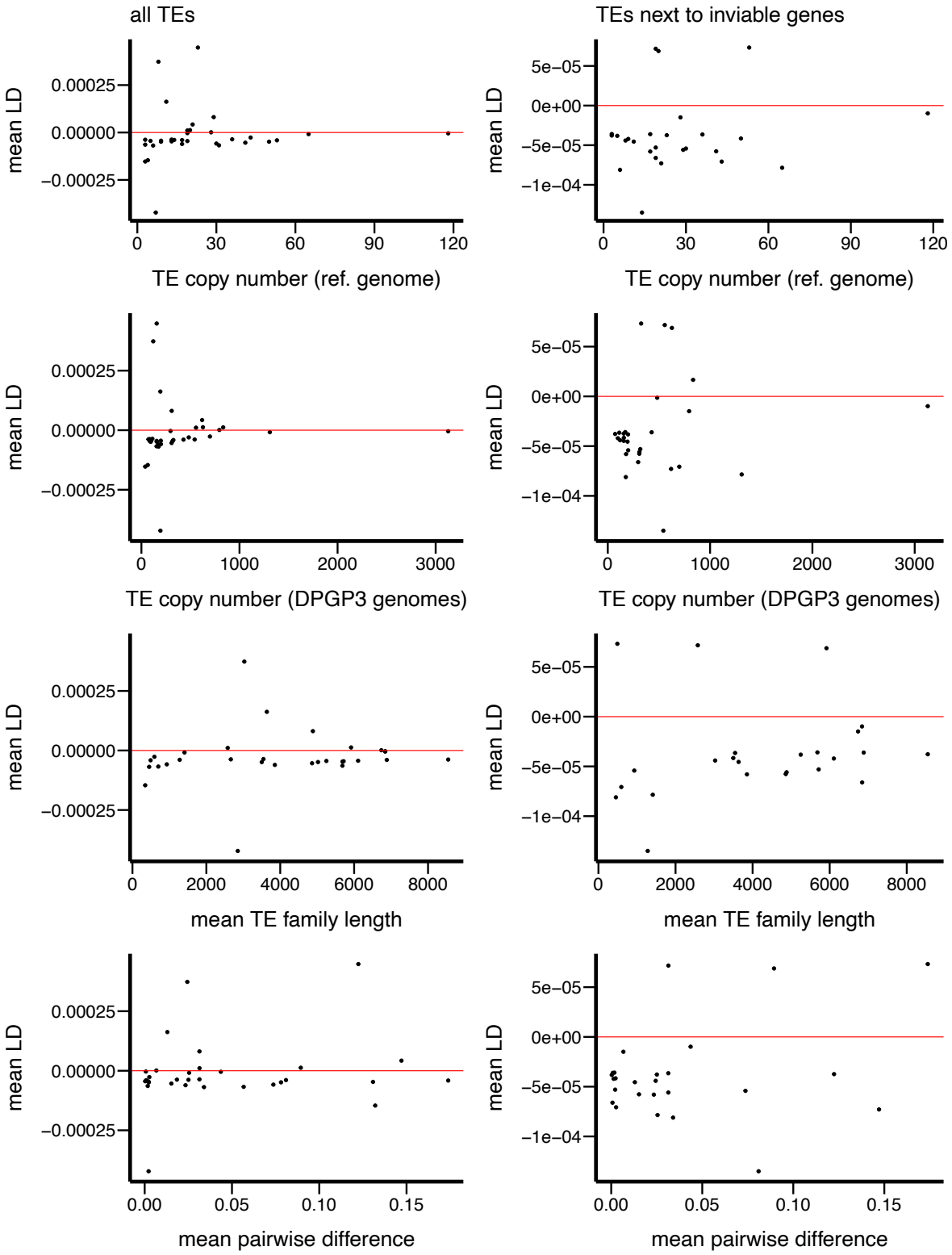

Figure S6. Distribution of mean LD per TE family with respect to the epigenetic effects of TE families.

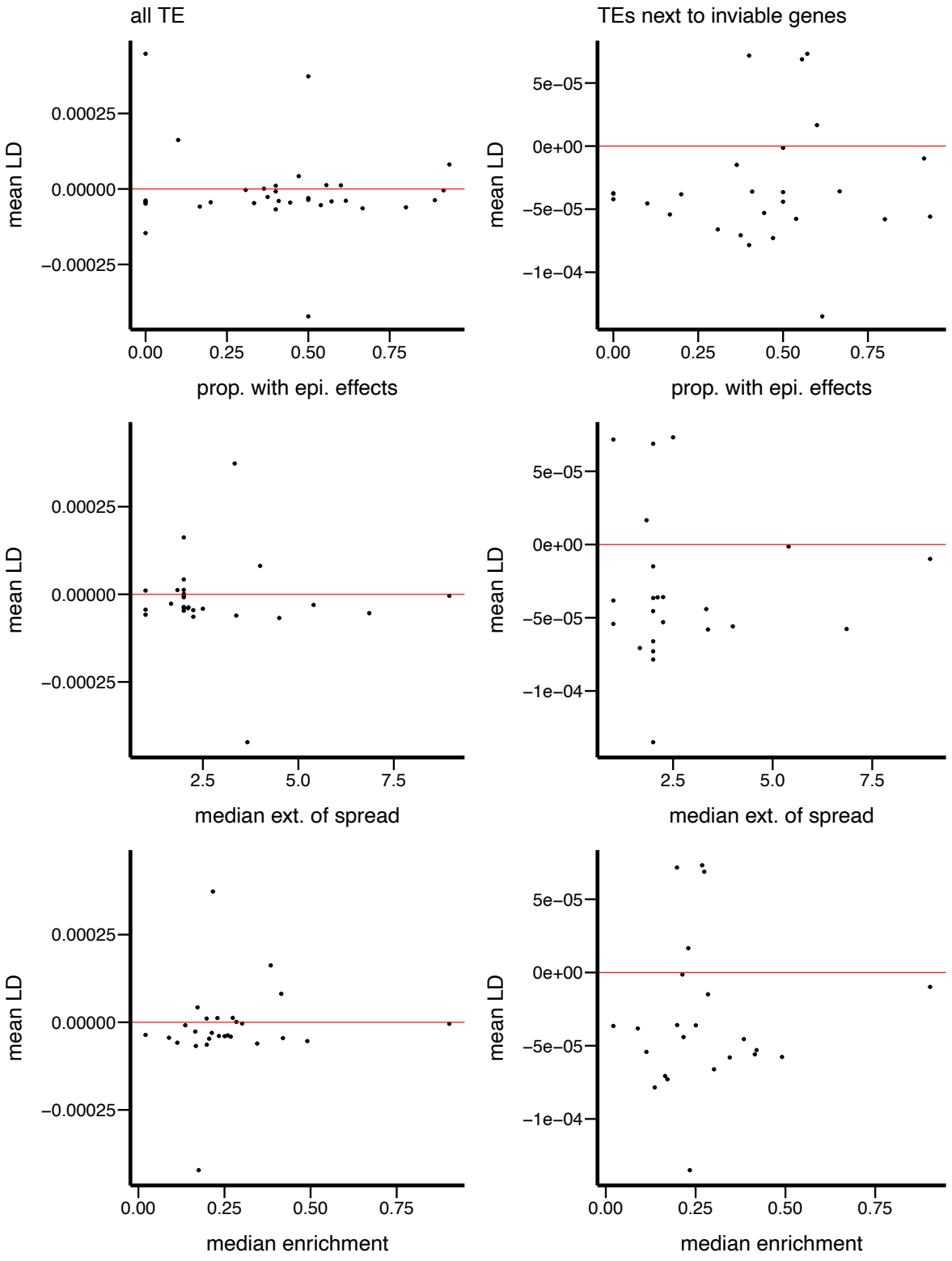

Figure S7. Distribution of mean LD per TE family with respect to different piRNA indexes.