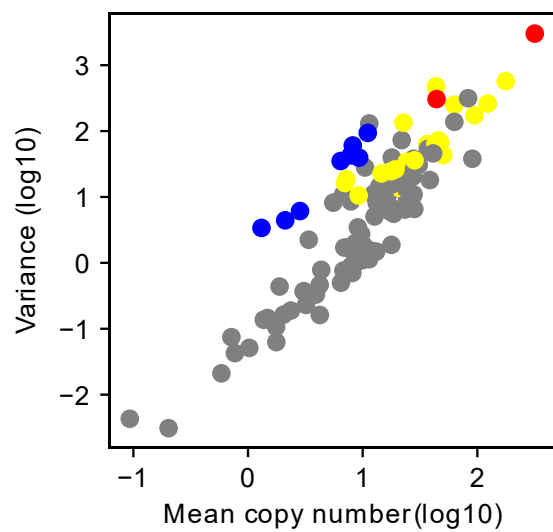
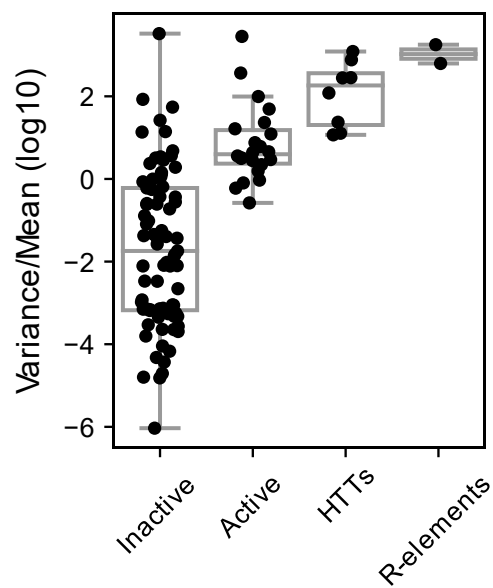


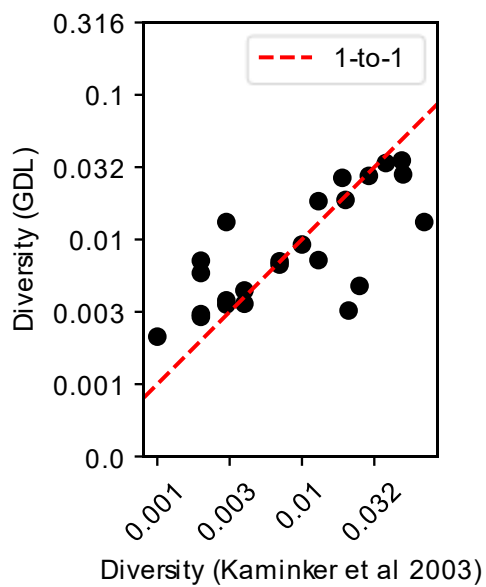
A



B



C



D

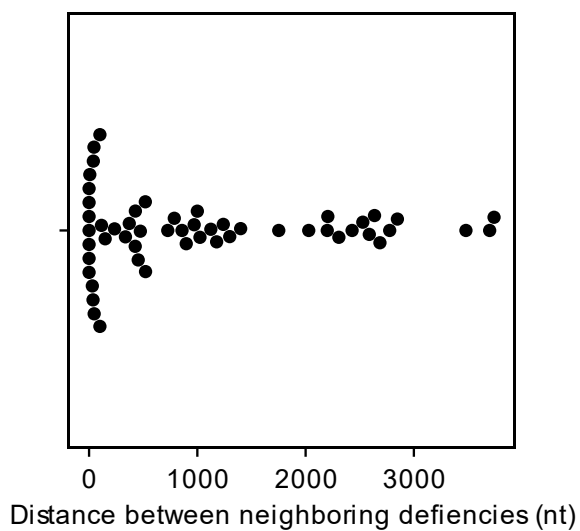


Figure S2. Alternative and supporting visualizations of data. (A) Scatter plot showing the relationship between mean copy number (log scale) and variance (log scale) of TE families across the GDL as estimated from the read depth at the 3'-end of TE consensus sequences rather than junctions. Designations of active and inactive TEs are from prior estimates of sequence divergence and population frequency as described in the Materials and Methods. Solid line represents the expected relationship under the assumption of little variation in population frequency and low linkage disequilibrium among insertions. (B) Boxplots depicting distributions of variance-to-mean ratios (\log_{10}) for each of the four categories of the TE families based on these copy number estimates. (C) A scatter plot comparing our estimates of sequence diversity for the TE families in Figure 4B with those reported in Kaminker et al (2003). This is restricted to TE families included in both analyses. (D) A swarm plot depicting the distribution of distances between terminal deficiency breakpoints, where for each breakpoint the distance is plotted relative to the nearest other breakpoint we discovered in a different sample.