

## S4: Assessment of flanking region alignments

In the aforementioned analyses, the shared-repeat graph was created by tracking the presence or absence of various repeats in multiple species. In order to decide if a repeat inserted in a given organism, flanking region alignments (described in Methods) were critical. To assess the quality of these alignments, we retrieved each upstream and downstream flanking sequence which resulted in a '1' or '0' (as defined in the Approach) and determined its percent identity to the reference sequence. Additionally, we randomly permuted each of these flanking sequences 10 times and determined the percent identity of the permuted sequence to the reference sequence. A histogram of these results, for our largest mammalian set, is shown in Figure 1. The original sequence overlaps with only the tail end of the permuted distribution. This supports the claim that the original flanking alignments represent significant orthology and are, therefore, useful in determining whether a repeat inserted prior to the divergence of two species.

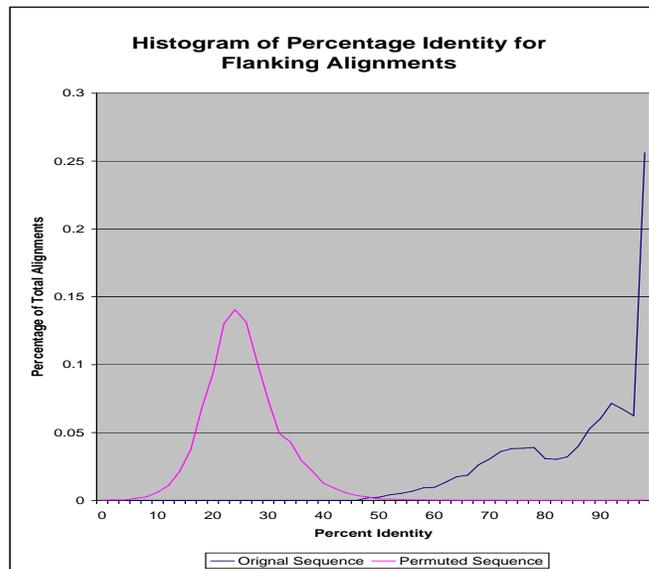


Figure 1: Histogram of percent identity for original and permuted sequences in flanking regions. The original sequence distribution appears skewed for two reasons. First, alignments cannot have percent identity greater than 100 percent (right side). Second, the cutoff for acceptable alignments required percent identity greater than 50 percent (left side).