

DNA double-strand break repair and the evolution of intron density

Ashley Farlow, Eshwar Meduri and Christian Schlötterer

Institute of Population Genetics, Vetmeduni Vienna, Austria

Corresponding author: Schlötterer, C (christian.schloetterer@vetmeduni.ac.at)

Increased DNA flexibility in lost introns. Intronic sequences were taken from *D. melanogaster* for all losses occurring in *D. ananassae* and *D. pseudoobscura*. A small number of introns have undergone multiple losses, therefore the number of available sequence is slightly less than the number of reported loss events. Lost intronic sequence was concatenated and the total number of highly flexible regions was calculated using the TwistFlex program (<http://margalit.huji.ac.il/TwistFlex/>) using default parameters with window size = 50 & leap = 1. The count of highly flexible peaks was normalised to the number of lost introns (Figure S1). The same approach was applied to non-lost introns. The 95% confidence intervals were established by resampling (1,000 times) the same length of non-lost intronic sequence and calculating normalised flexibility as per lost introns (Table S1). p-values represent the number of bootstraps with a flexibility score more extreme than the observed value.

Table S1. Length statistics for lost introns. Resampling for flexibility scores was carried out such that the 1000 bootstraps had the same median length or the same mean length as the lost introns, p-values were significant with both approaches.

	n	Observed Length (bp)				Length in each bootstrap x n	
		min	max	median	mean	median	mean
non-lost introns	9222	44	19894	66	293		
Lost in ananassae	97	53	646	64	100	64	100
Lost in pseudoobscura	100	52	5476	62	171	62	171
All lost (present in mel)	624	49	4576	62	131		

n = number of introns in class

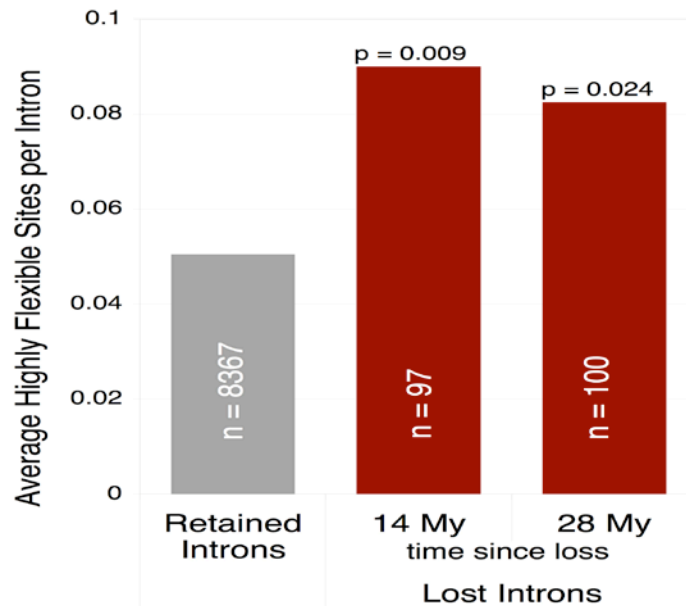


Figure S1. Recently lost introns in *Drosophila* contain more sequences with high flexibility than phylogenetically stable introns. Sequence regions that exhibit high flexibility are more susceptible independent recurrent DSBs [S1].

Table S2. Parameters used in Figure 2 of the text. We assume that the rate of intron gain or loss is roughly proportional to the level of activity of each pathway. While this need not be the case, it does not affect the quality of our findings. The values used to generate Figure 2 are arbitrary, but fall well within the reported range of *relative* activity for NHEJ and HR across different species (Table 1 of the text).

	Upper	Lower
NHEJ gain	1	1
NHEJ loss	1	1
HR loss	0.2	2

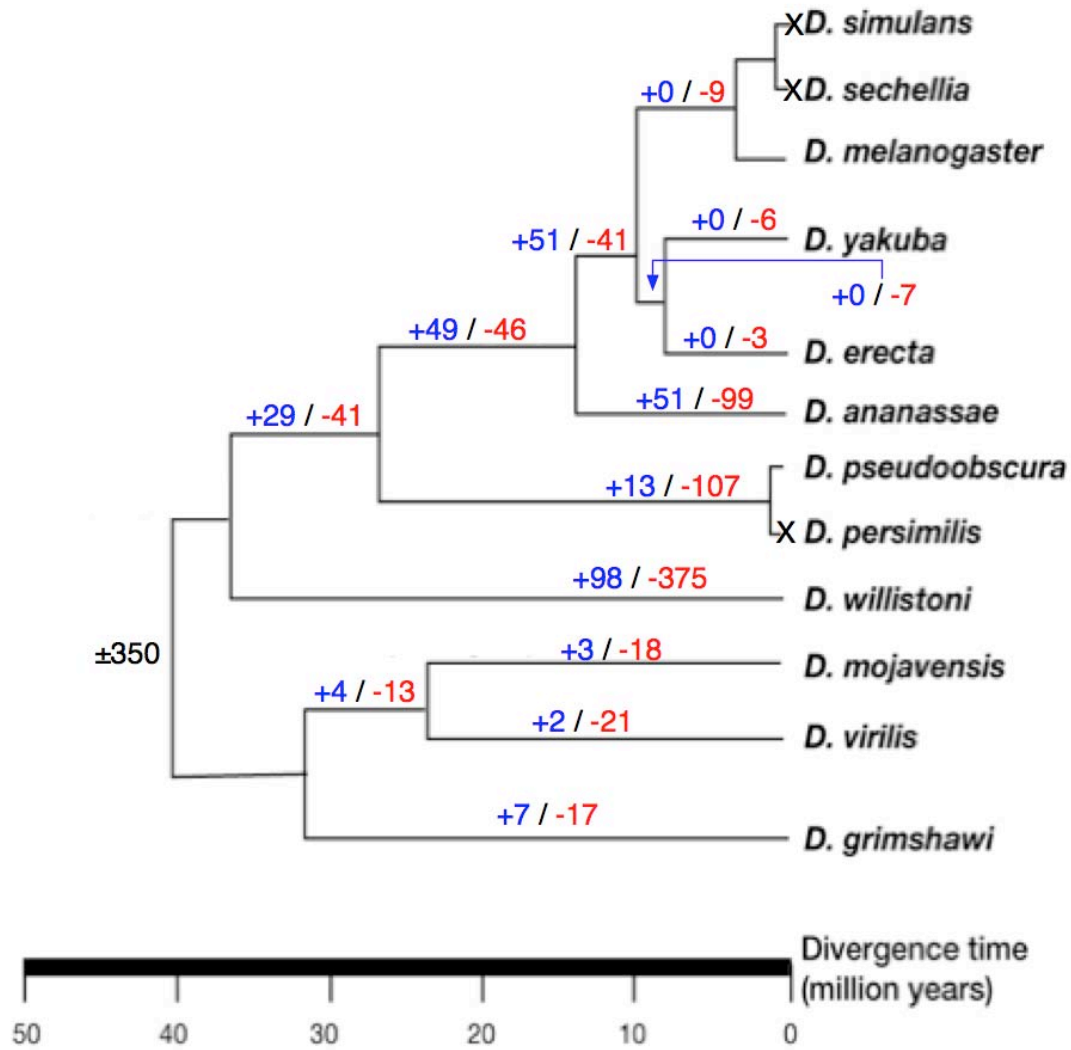


Figure S2. Intron gain and loss along each branch of the *Drosophila* species clade, related to Figure 1. We considered the large number of "recent" loss events leading to *D. ananassae* and *D. pseudoobscura* for our DNA flexibility analysis in Figure S1 due to their proximity to *D. melanogaster* (from which the "lost" sequence were taken). Much of eukaryotic evolution has occurred under a balanced rate of intron gain to loss [S2,S3], however, these rates are neither equal (intron loss predominates) nor constant (both rates vary over time). While speculative, it is possible that these differences may result from either genetic or environmental factors that cause a systematic change to the level of DSBs.

- S1. Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327: 302-305.
- S2. Carmel L, Wolf YI, Rogozin IB, Koonin EV (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Research* 17: 1034-1044.
- S3. Farlow A, Meduri E, Dolezal M, Hua L, Schlötterer C (2010) Nonsense-mediated decay enables intron gain in *Drosophila*. *PLoS Genet* 6: e1000819.