Table S1 Illumina sequencing error frequency distribution empirically estimated based on D. melanogaster MA line whole-genome sequence data.

		To			
		A	T	G	C
From	A	-	0.222	0.257	0.521
	T	0.222	-	0.519	0.260
	G	0.199	0.696	-	0.106
	C	0.695	0.199	0.105	-

We empirically estimated a matrix of nucleotide sequencing error frequencies from Illumina whole-genome sequences of three Drosophila melanogaster mutation accumulation (MA) lines aligned to the reference D. melanogaster genome by the MAQ aligner (Li, H.et al. 2008. Genome Res. 18: 1851-1858), which have been previously published (Keightley et al. 2009). The three initially isogenic lines were maintained for 262 generations by full-sib mating (Fernandez, J. and López-Fanjul, C. 1996. Genetics 143: 829 837), and are therefore close to 100% inbred. Almost all sites sequenced are therefore expected to be homozygous, and any base reads that are different from a consensus base are likely to be errors. We counted the frequency of putative errors at sites that had a depth of coverage of at least 10, and classified them according to the kind of nucleotide change involved. There were 1,788,344 putative errors in 1,717,965,680 base reads, giving an estimated mean error rate of 0.00104 per base read. The data indicate a substantial departure from uniform errors, with errors of type A/T → C/G and C/G → A/T predominating.