

### Supplemental Note S3: Effects of altering the neutral substitution matrix for simulations

We considered the effect that altering the neutral substitution matrix would have on our simulations. In particular, we considered two possible limitations of our inferred neutral substitution matrix. First, our matrix was derived from flanking X chromosome sequence with a median GC content around 40%, while palindrome arms have a median GC content around 46%. Previous work has shown that substitution patterns can differ based on regional GC content, with regions with high GC content showing a lower rate of strong (GC) to weak (AT) mutations (Duret and Arndt 2008). Using a matrix derived from sequence with a lower GC content could in theory lead to over-estimation of AT mutation bias, and subsequent over-estimation of the GC conversion bias required to balance it. We therefore re-calculated our neutral substitution matrix using a subset of flanking sequence (1.3 Mb) with a total GC content of 45% and repeated our simulations (figure below, top panel). Our inference of GC bias remained unchanged: Using 20 simulations of 12 palindromes each, our observed results were still most consistent with a GC bias magnitude of 0.70 (difference between observed and simulated results not significant,  $p > 0.05$ ).

The second limitation we considered was our use of parsimony to infer substitution events for our neutral substitution matrix. While this is appropriate for most types of substitutions on the time scale of human-chimpanzee evolution, it has been shown to under-estimate the frequency of CpG substitutions, which occur more frequently than other substitutions and can thus occur twice at the same site (Duret 2006). Under-estimation of CpG substitutions could lead us to under-estimate the AT mutation bias, and therefore under-estimate the magnitude of GC conversion bias. To determine the impact on our simulations, we re-estimated our rate of CpG substitutions to align with the values found by Duret and Arndt 2008, who found that CpG substitutions (CG  $\rightarrow$  AT at CpG sites) are 14 times more common than the same substitutions at non-CpG sites using a maximum-likelihood method. While adjusting the frequency of CpG adjustments shifted our simulated differences in GC content between derived and ancestral bases slightly downwards, our results were still consistent with a magnitude of GC bias if 0.70 (difference between observed and simulated results not significant,  $p > 0.05$ ) We conclude that our results are robust to reasonable shifts in the neutral substitution matrix.

