SUPPLEMENTARY MATERIALS

Sampling

Codon bias is determined by a number of factors, such as expression level and protein length and to quantify the effects of simply X-linkage on codon bias, we conducted a series of sampling experiments (see Methods). Controlling for the individual effects of protein length, expression level, and neighboring GC content suggests that the effect of X-linkage on codon bias is not systematically enhanced or repressed by these genic characteristics (Figure S1). In *D. melanogaster* (Figure S1a), the initial 0.026 difference in codon bias between X-linked and autosomal genes increases to 0.027 when the effects of length are taken into consideration, and increases to 0.030 when GC content is taken into account, though neither of these increases s statistically significant. When genes are paired according to all three criteria or by expression level alone, the difference in codon bias between X-linked and autosomal genes decreases, though not significantly so. In *C. elegans*, pairing genes along individual characteristics or along all three criteria decreases the magnitude of the difference in codon bias between X-linked and autosomal genes, though not significantly (Figure S1b). As a result, while there are differences in GC content of noncoding sequences, protein length and expression level of X-linked versus autosomal genes, these differences only nominally affect the measured difference in levels of codon bias between autosomal and X-linked genes.

Population Genetic Model

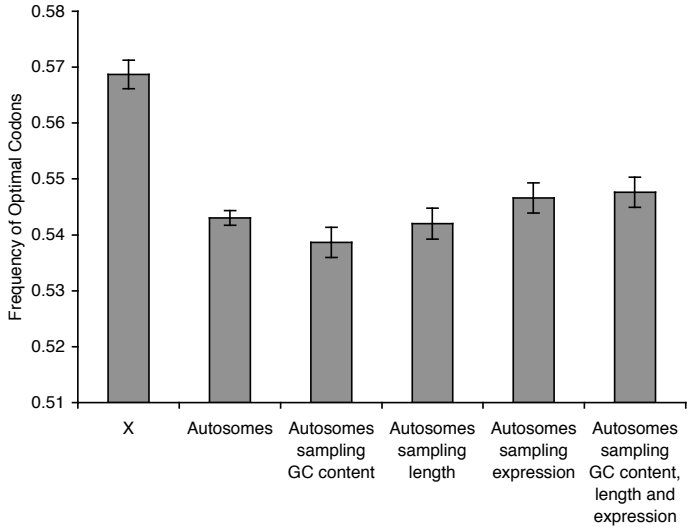
We have seen that for some combinations of c (N_M/N_F) and F (s_1/s_2), $N_e s_X > N_e s_A$, and as a result codon bias of X-linked genes could be increased relative to codon bias of autosomal genes. Using the same equations we considered what values of c and F account for the observed difference in codon bias in Drosophila under three simplifying assumptions. We assume that $r = \mu_p/\mu_u$, which is the probability of mutation from a preferred codon over the probability of a mutation from an unpreferred codon is constant for all codons. Under this assumption, $r = 3$ may be appropriate for the 4-fold degenerate amino acids with a single major codon where all mutation rates are equal. We also assume that all sites are under selection to the same extent and that all sites are unlinked. These assumptions lead to the following two equations for explaining the levels of codon bias that we observe:

$$\mathrm{FOP_A} = \frac{1}{1 + (\mu_p/\mu_u)e^{-4N_{eA}s_A}} \text{ and } \mathrm{FOP_X} = \frac{1}{1 + (\mu_p/\mu_u)e^{-4N_{eX}s_X}}$$

We can then investigate the dependence of F on c given our values of codon bias on the X chromosome and the autosomes for different values of r (Figure S2).

Figure S1: Optimal codon frequencies for sampling experiments in a) *D. melanogaster* and b) *C. elegans*. "X" denotes FOP of X-linked genes, and "autosomes" depicts FOP of autosomal genes. Mean optimal codon frequencies of autosomal genes after sampling with respect to one of more variables (see Methods) are also depicted. Autosomal genes were sampled based on neighboring GC content, protein length, expression level, and all three properties combined.

a)



b)

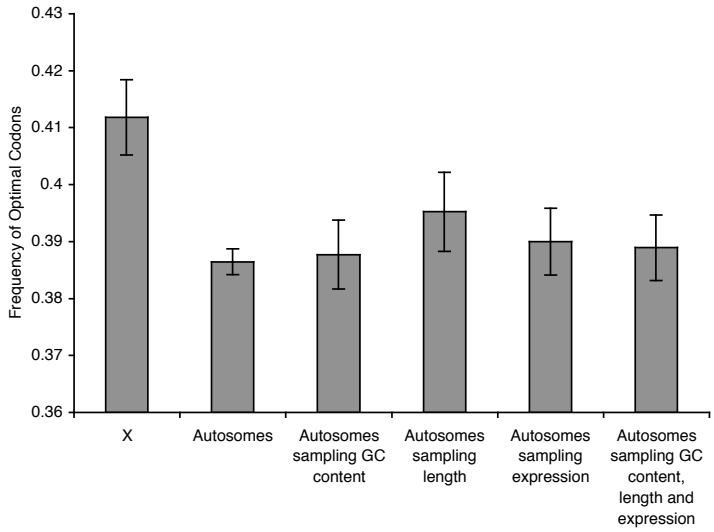


Figure S2: Pairs of c (N_M/N_F) and F (s_1/s_2) that provide solutions to our measured difference in codon bias between X-linked and autosomal genes in Drosophila for several values of r, where r is the ratio of probabilities of mutation away from and to the preferred codon, respectively.