File S4: The likelihood that mutation creates or destroys a binding site differs substantially between biological binding sites and random binding sites.

PBM data can not only help estimate p_B, but also the likelihood that a mutation which alters a single nucleotide in a DNA sequence of a given length creates or destroys a TF binding site. This likelihood may depend on whether TF binding sites are uniformly distributed in sequence space. To find out, I numerically determined for multiple values of p_B and for $L = 8$ the probabilities μ^+ and μ^- (i) for mouse binding sites obtained from PBM data, and (ii) for random DNA sequences of length $L = 8$ that are uniformly distributed in sequence space, and that fill the same fraction p_B of sequence space as the biologicl sequences (Methods). Both the probabilities of creating and destroying binding sites (μ^+ and μ^-) are substantially smaller for the biological binding sites than for random and uniformly distributed binding sites (Figure S2A and S2B). The likely reason is that binding sites for any one TF are more aggregated or localized in sequence space than expected by chance alone (Aguilar-Rodríguez et al., 2018). This analysis also shows that even for p_B as low as 0.05 almost one half of all sequences in sequence space are no more than one mutation away from a TF binding site (Figure S2C).

In sum, biological binding sites do not behave like random sequences under mutation pressure, I will thus use estimates of mutation probabilities from biological binding sites in my population genetic simulations. I note that the likelihoods shown in Figure S2A and S2B are based on one nucleotide mutation event per binding site, and still need to be multiplied with the nucleotide mutation rate per binding site $L\mu$ to obtain an overall mutation rate for the destruction or creation of a binding site.

References

Aguilar-Rodríguez, J., L. Peel, M. Stella, A. Wagner, and J. L. Payne (2018). The architecture of an empirical genotype-phenotype map. *Evolution 72*(6), 1242–1260.