



S3 Fig. Selection inference based on amino acid distances. (a) The cumulative distribution of mean amino acid distances d between reassortant and parent strains for the HA-NA reassortments in influenza A/H3N2 (red line) is compared to the corresponding distribution of distances for co-circulating strains in the same influenza season (solid blue line) and from the New York area only (dashed blue line). (b) The ratio of reassortment counts to background counts in the interval $d \geq d_{\min}$ (red circles) decreases with increasing lower threshold d_{\min} and drops significantly below 1 (blue line). The suppression of reassortment at larger values of d signals distance-dependent negative selection. Bars show statistical errors due to the finite number of inferred reassortments. See Fig. 5 for the same analysis using nucleotide distances. (c) The average number of strains in the reassortant clades with amino acid distance $\leq \tau_A$ from the focal node, $\langle N_r \rangle(\tau_A)$ (red line) is compared to the corresponding average number of strains in the parent clades, $\langle N_0 \rangle(\tau_A)$. For $\tau_A \lesssim 4$, both functions increase with τ_A in an approximately exponential way; we estimate growth rates $f_r(A) \approx 0.2$ and $f_0(A) \approx 0.7$, respectively (dashed lines; cf. equation 1). The growth rate difference $\bar{s}_A \equiv f_0(A) - f_r(A) \approx 0.5$ inferred from distances in amino acid units is similar to $\bar{s} \approx 0.4$ for nucleotide distances; cf. Fig. 5c.