



**S4 Fig. Background distribution and reassortment events as a function of the amino acid distances  $d_{HA}$  and  $d_{NA}$  between strains.** (a) The background distribution  $P_0^{aa}(d_{HA}, d_{NA})$  (contour plot) is compared to reassortment counts (red dots). (b) Conditional background distributions  $P_0^{aa}(d_{HA} - d_{NA} | d_{HA} + d_{NA})$  (whisker plots) are compared to reassortment counts (red dots). Whisker plots show the 0.25 quantile to the 0.75 quantile of the distribution (blue boxes); the white horizontal line represents the median, vertical bars span the dataset excluding outliers. The width of the bins is chosen to ensure a statistically relevant number of events for each bin. The reassortment data appear more spread in the coordinate  $d_{HA} - d_{NA}$  compared to the background (red points are mainly placed outside or at the border of the blue boxes).