

DMS report for Juhye

Nov 13, 2017, 2:37 PM

#collaboration/bloom

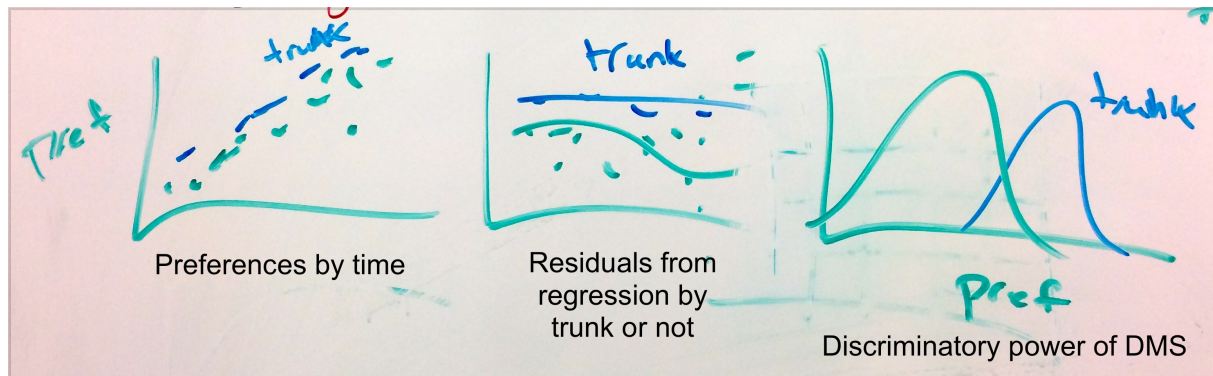


Figure 0. Expectations for a) linear regression between sequence preference and time, b) residuals from the linear regression, and c) distribution of preferences by trunk status. Ideally, the latter distribution indicates discriminatory power of sequence preference between trunk and side branch status which would then be confirmed by the linear regression and residuals.

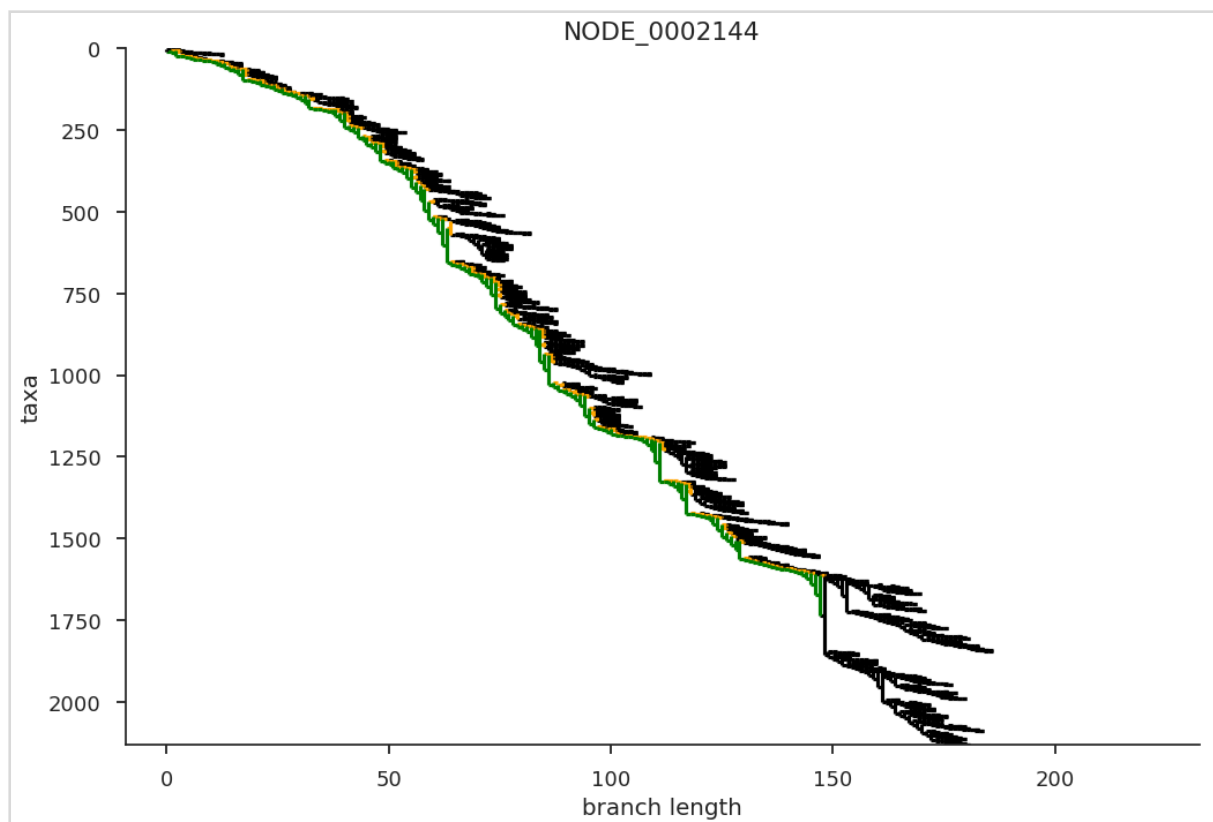


Figure 1. A/H3N2 tree built with ~3,000 sequences from 1968-2017. Branches on the trunk are shown in green while ancestors of side branch clades are shown in orange.

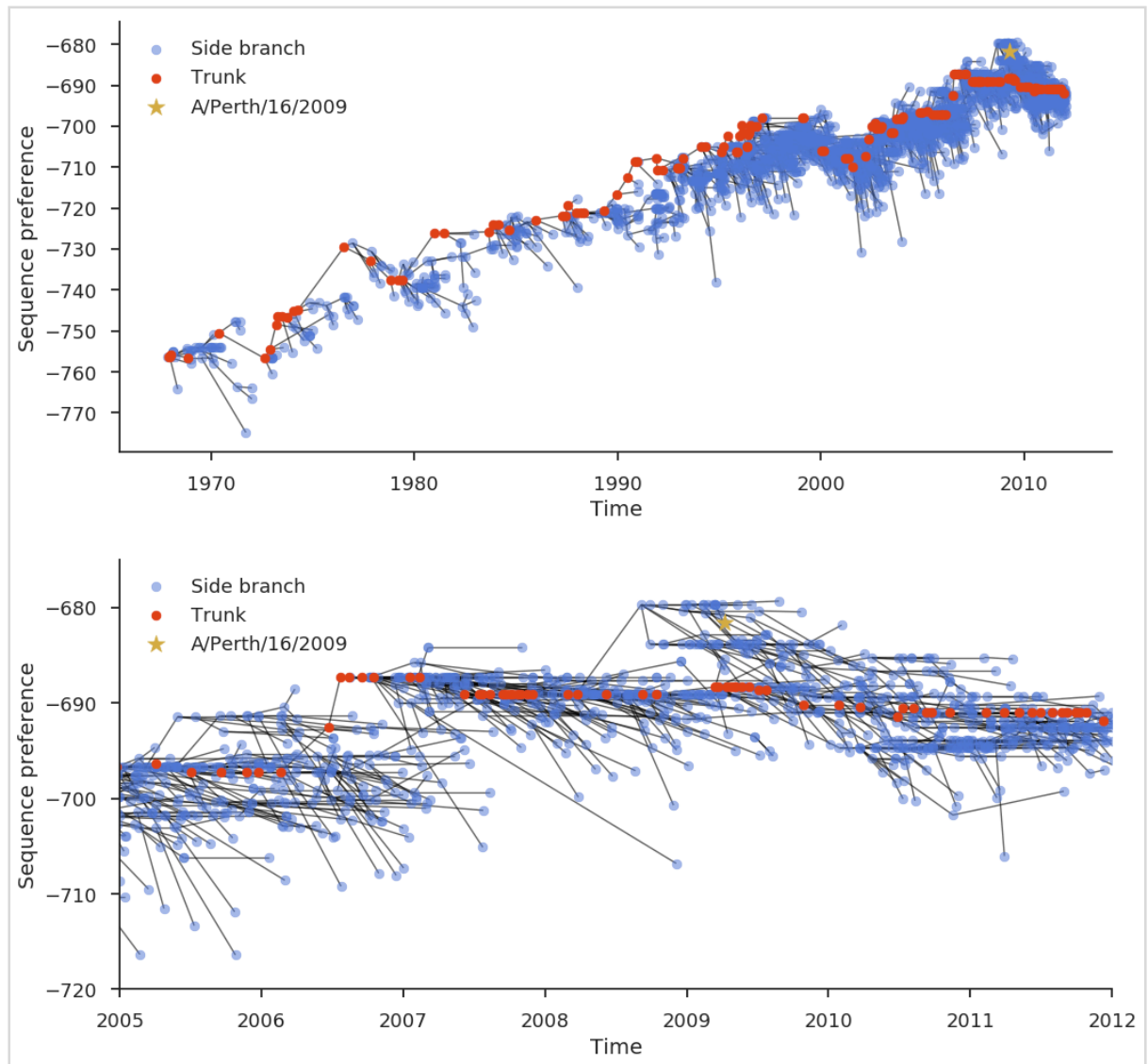


Figure 2. Sequence preference by time for a) 1968-2012 and b) 2005-2012. Trunk nodes are shown in red, side branch nodes are in blue, and the DMS strain is a gold star. Each node is connected to its direct ancestor by a line.

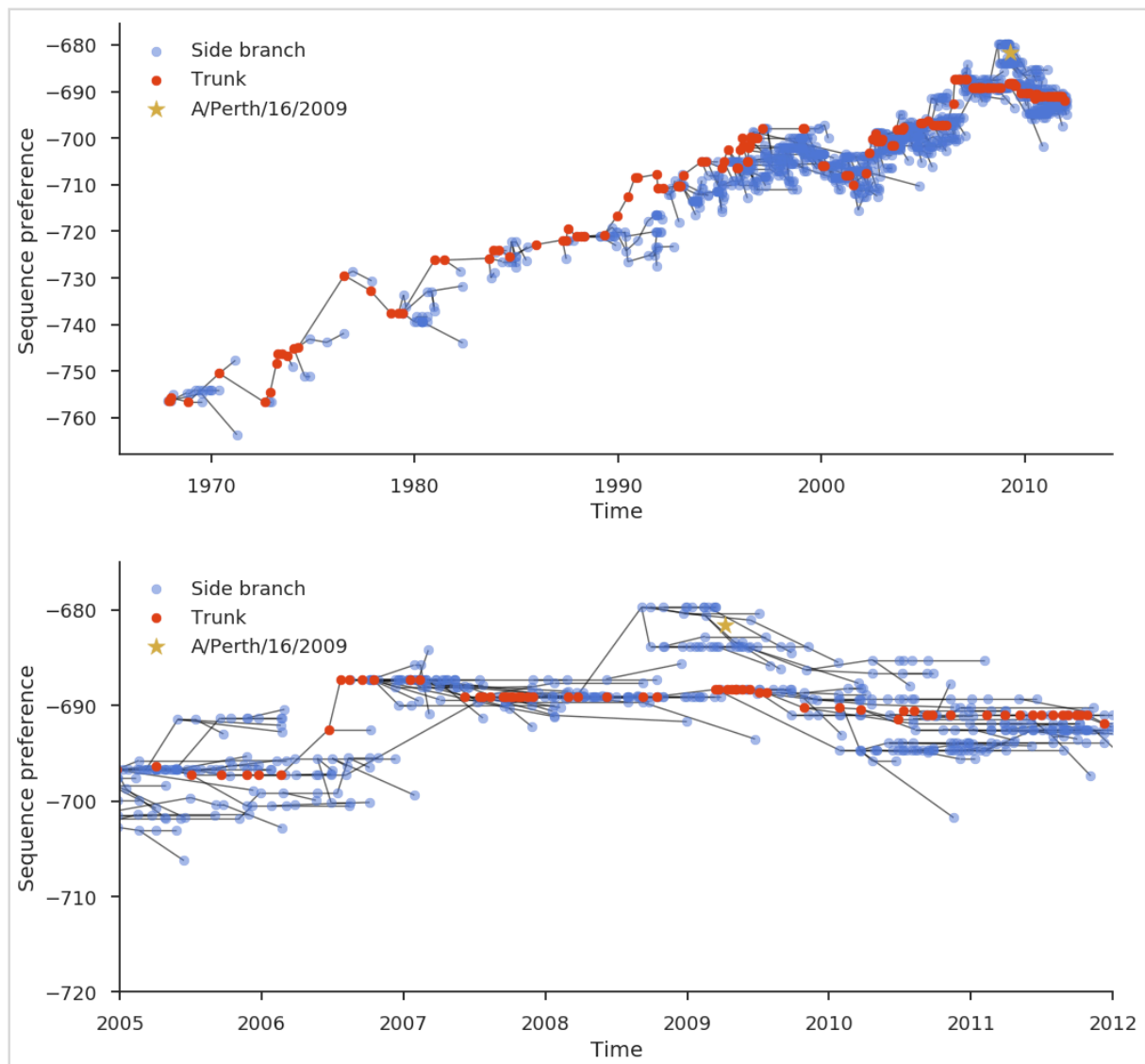


Figure 3. Sequence preference by time as in Figure 2 except with all terminal nodes eliminated. The terminal node for the DMS strain is shown by a gold star for reference.

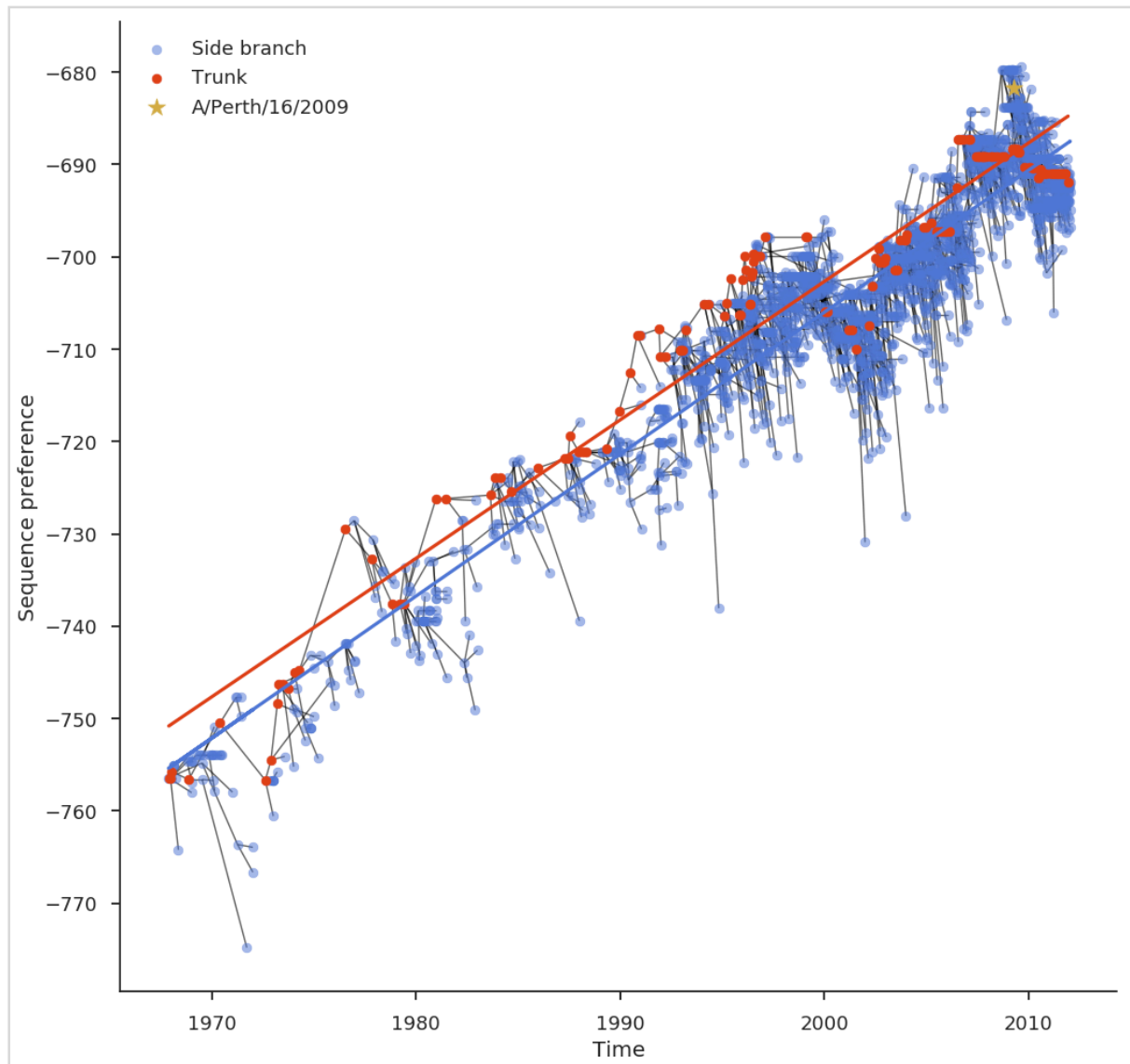


Figure 4. Sequence preference by time for sequences from 1968-2012. Least squares regression lines are shown for trunk (red) and side branch (blue) nodes. The trunk line corresponds to the equation $y = 1.54x - 3779.83$ with an adjusted R-squared value of 0.87 ($p < 0.01$). The side branch line corresponds to the equation $y = 1.50x - 3700$ with an adjusted R-squared value of 0.88 ($p < 0.01$). Note: the differences between these two equations is effectively eliminated when the same analysis is performed without terminal nodes.

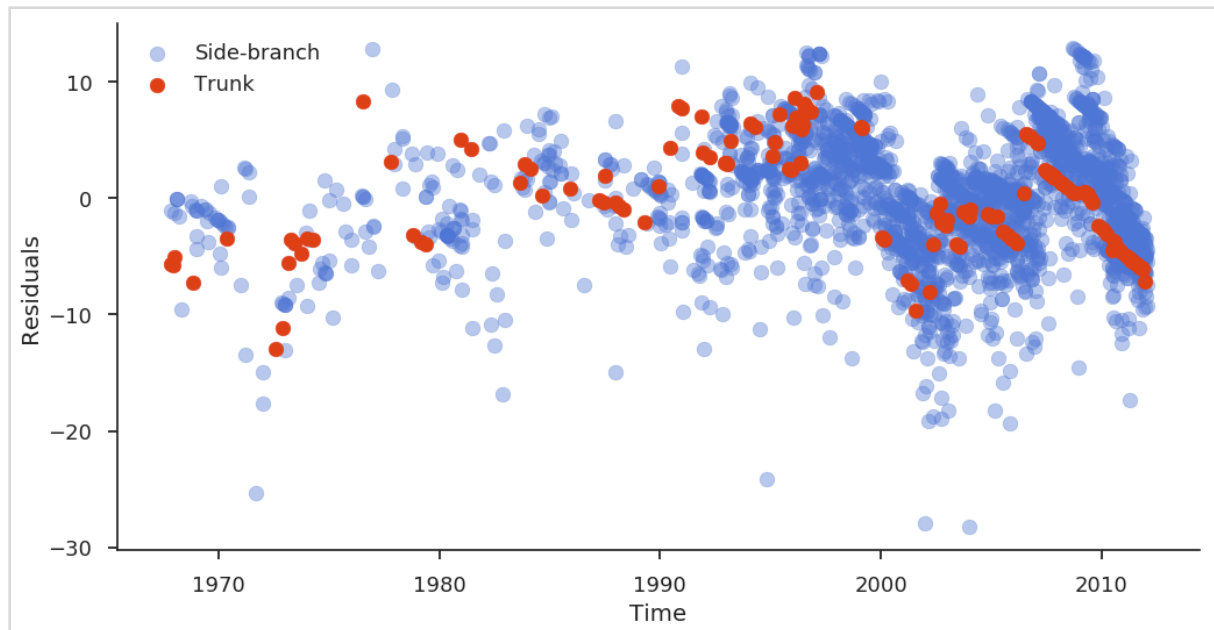


Figure 5. Residuals by time based on linear regression equations in Figure 4. The null expectation for the residuals is a horizontal line for both trunk and side-branch nodes with a trunk line slightly above the side-branch line.

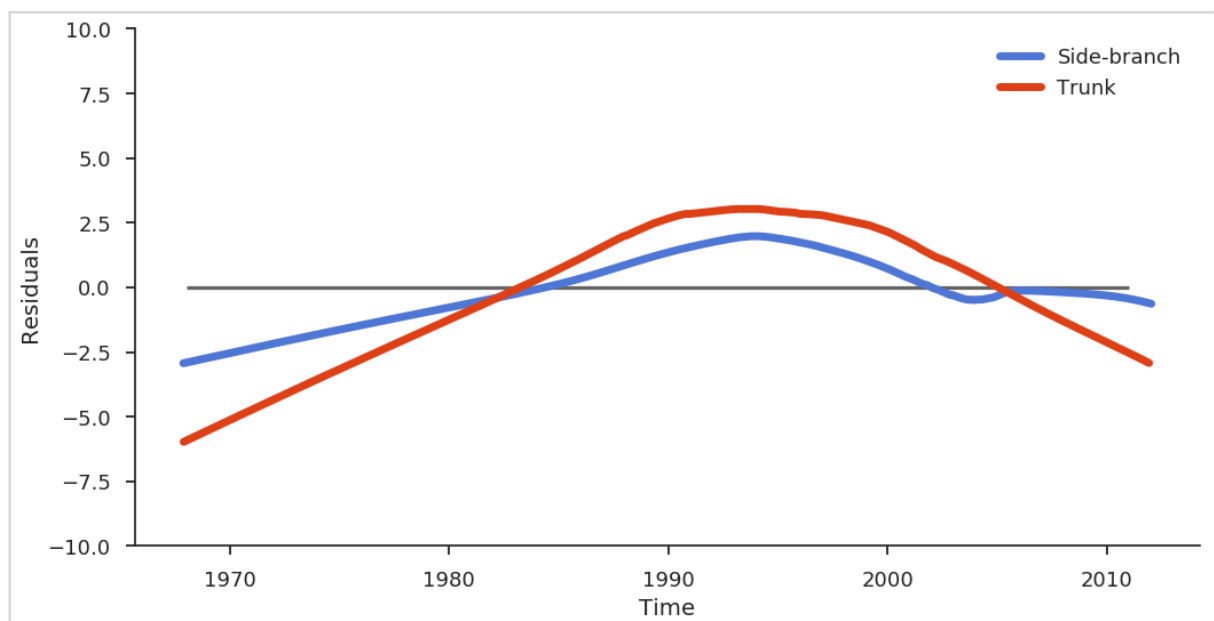


Figure 6. Local weighted regression (LOWESS) lines for trunk and side-branch residuals from linear regression equations in Figure 4.

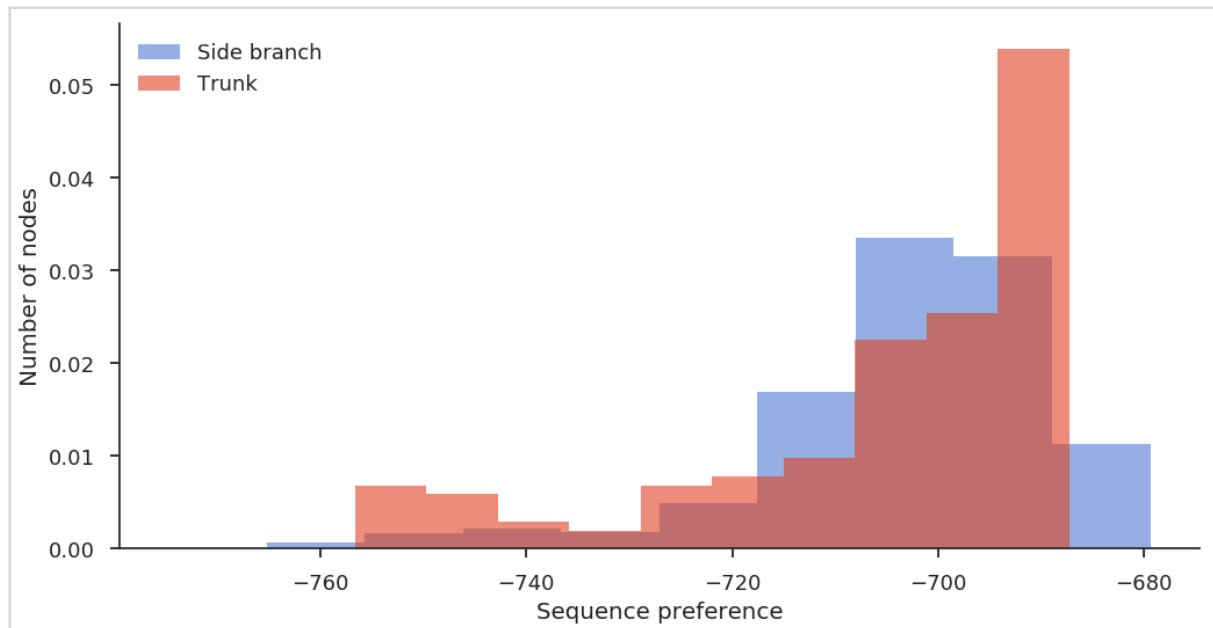


Figure 7. Distribution of all sequence preference values between 1968-2012 by trunk and side-branch status. Overlapping distributions indicate the difficulty of discriminating between trunk and side-branch status by sequence preference alone across the entire time period.

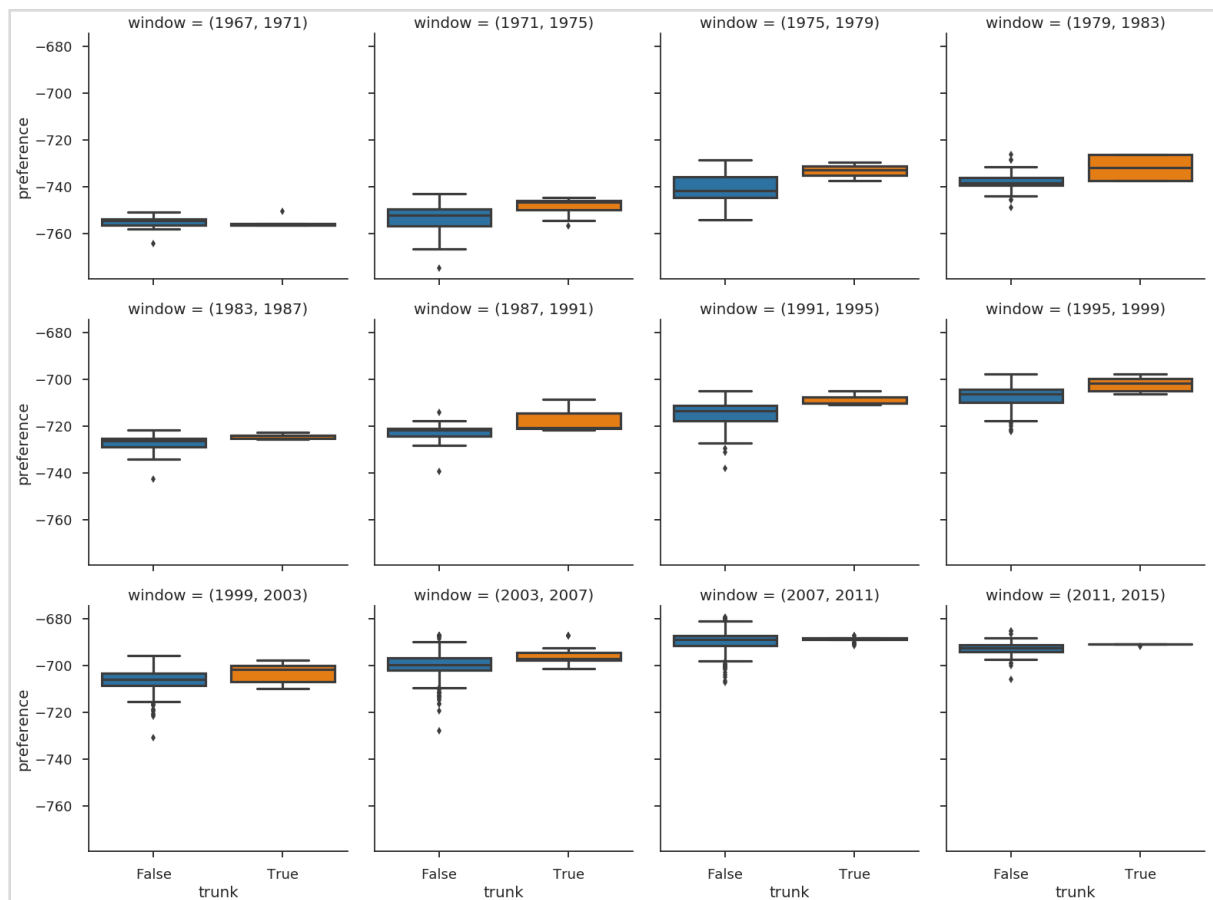


Figure 8. Distributions of sequence preferences by trunk status over four-year windows. Trunk and side-branch nodes appear to be distinguished by sequence preference between 1971 and 2007 after which there is little distinction between the trunk status.

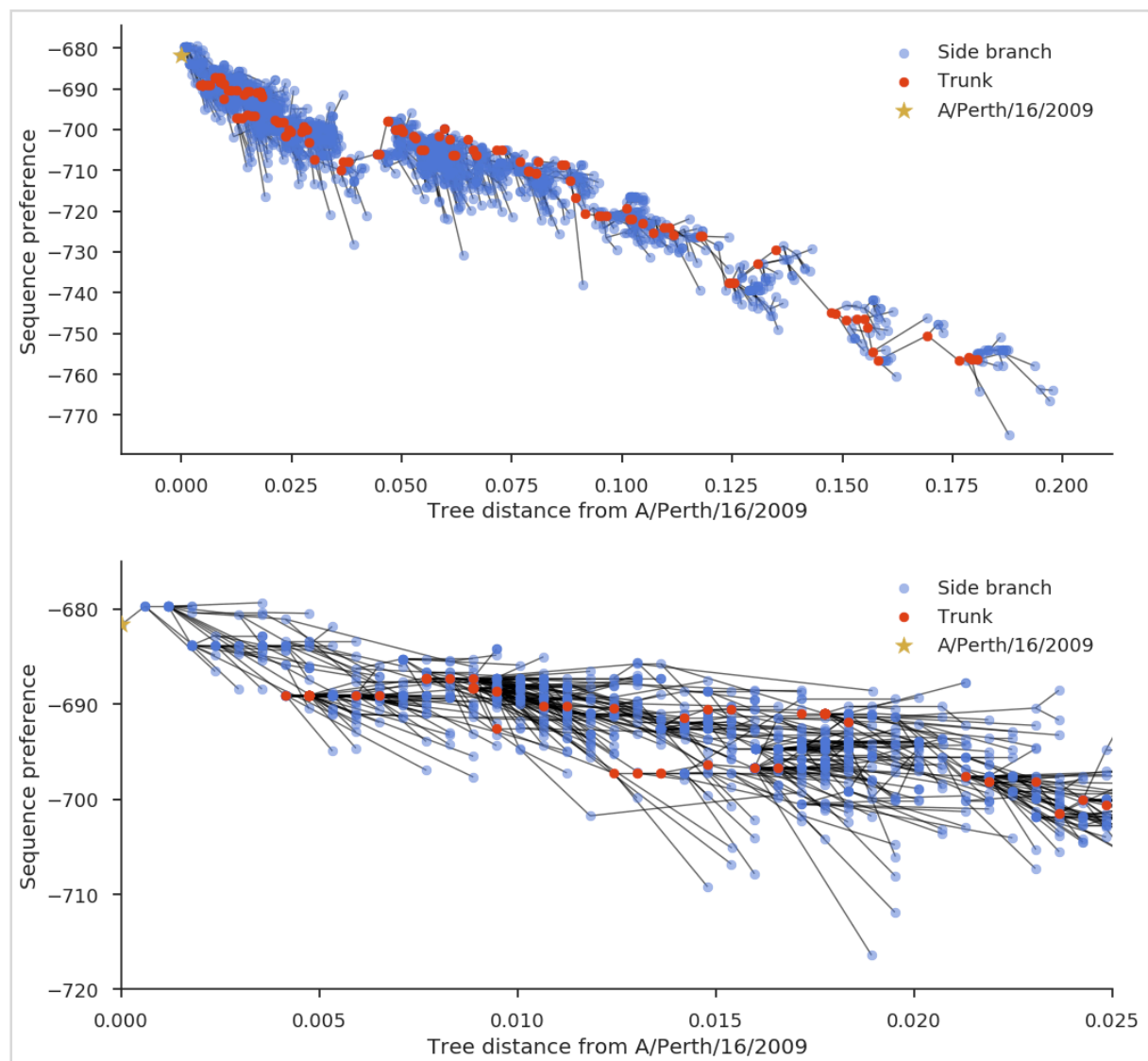


Figure 9. Sequence preference by tree distance from DMS strain for sequences between a) 1968-2012 and b) 2005-2012. Trunk nodes are shown in red and side-branch nodes are shown in blue. The DMS strain is shown by a gold star.