

List of figures and explanations for them, including the captions for the figures, and how to reproduce them.

To reproduce Figure 1, use the function *simulation* to do the simulation and then plot the path. The number of amino acids in the simulated protein can be longer than 1.

Plot function for this Figure 2 is *matrix.bubble*.

Like Figure 4, plots of all substitution rate matrices for each amino acid being optimal are in **submit_all_opaa.pdf**.

All the network figures like Figure 5 are in **network_plots.pdf** and **network_plots_nolevel.pdf**. They are produced by the functions *fgraph* and *fgraph1*.

Figure 6 is plotted by **gatherdata.R**.

Figure 7 is plotted by **test.R**, and plots for 16 and 32 taxa are in **opaa_estimation_combined.pdf**.

All figures like Figure 8 are in this directory:

~/BackupProEvo/Newton/yeast/adequacy

All figures like Figure 9 and Figure 10 are in this directory:

~/BackupProEvo/Newton/yeast/rootEqm.

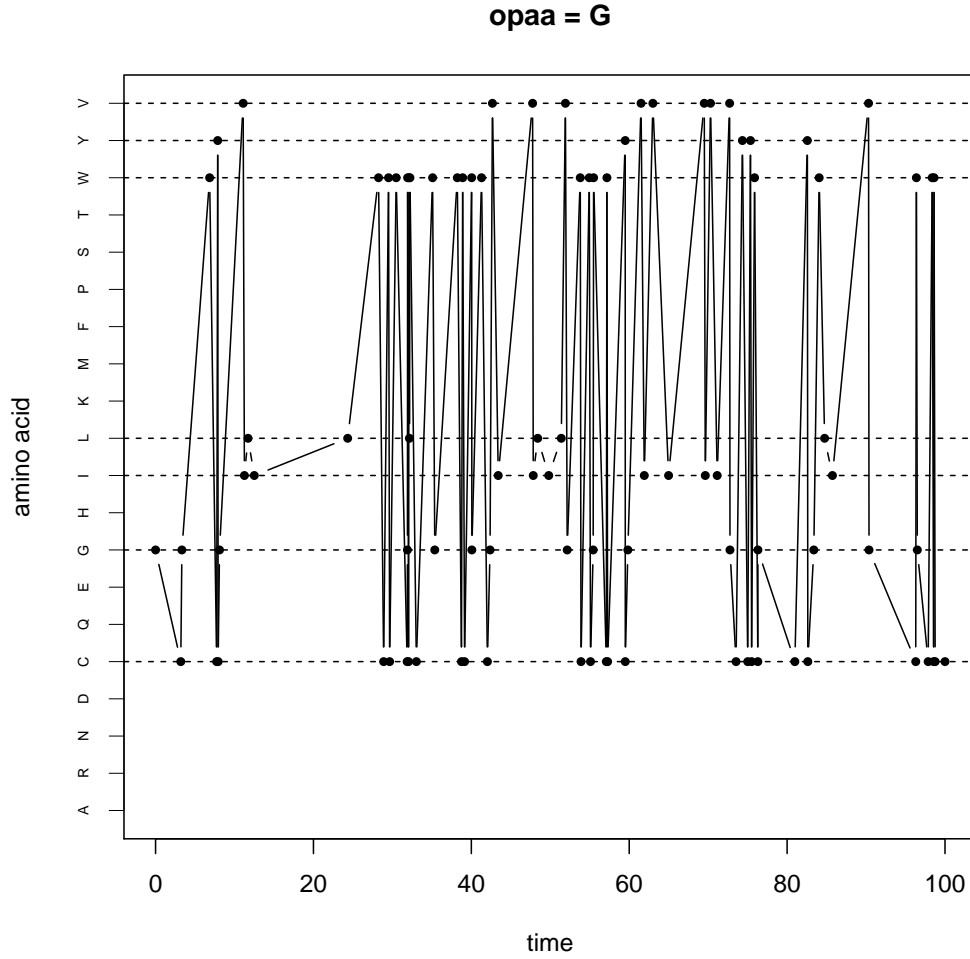


Figure 1: A long evolutionary path when the optimal amino acid is (8th) G. With 84 substitutions in the path, only 7 states are visited. Parameter values are: $g\phi = 2.005282$, $\beta = 0.1938331$, $\gamma = 0.0003889893$, $Q = 1.129544, 3.032109, 1.381705, 1.751379, 3.799967, 1.000000$

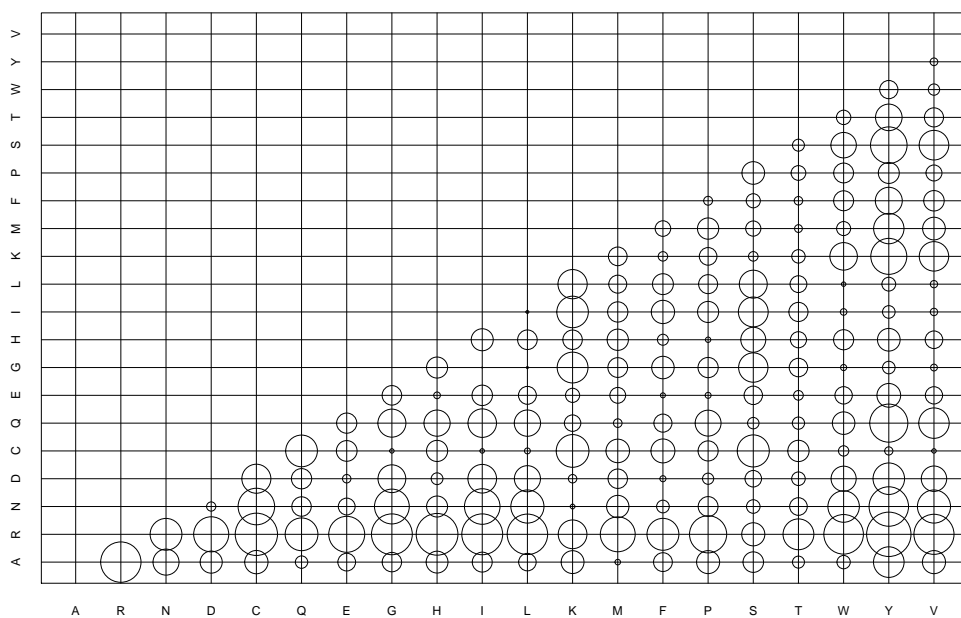


Figure 2: Bubble plot of the Grantham Distance matrix with Grantham's weights. Only half of the matrix is plotted since it's symmetric.

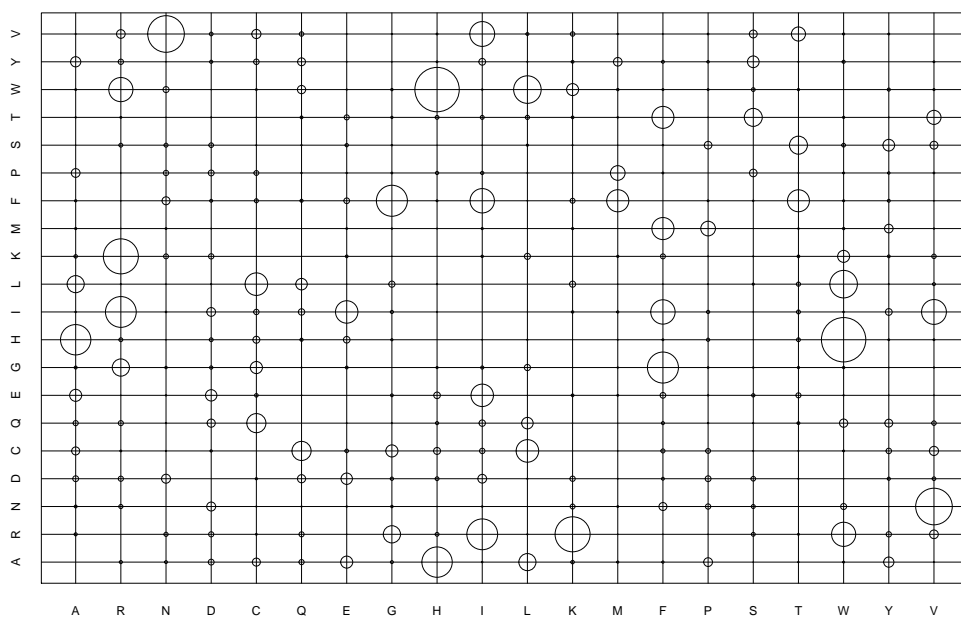


Figure 3: Bubble plot of the substitution rates matrix under WAG model.
Only half of the matrix is plotted since it's symmetric.

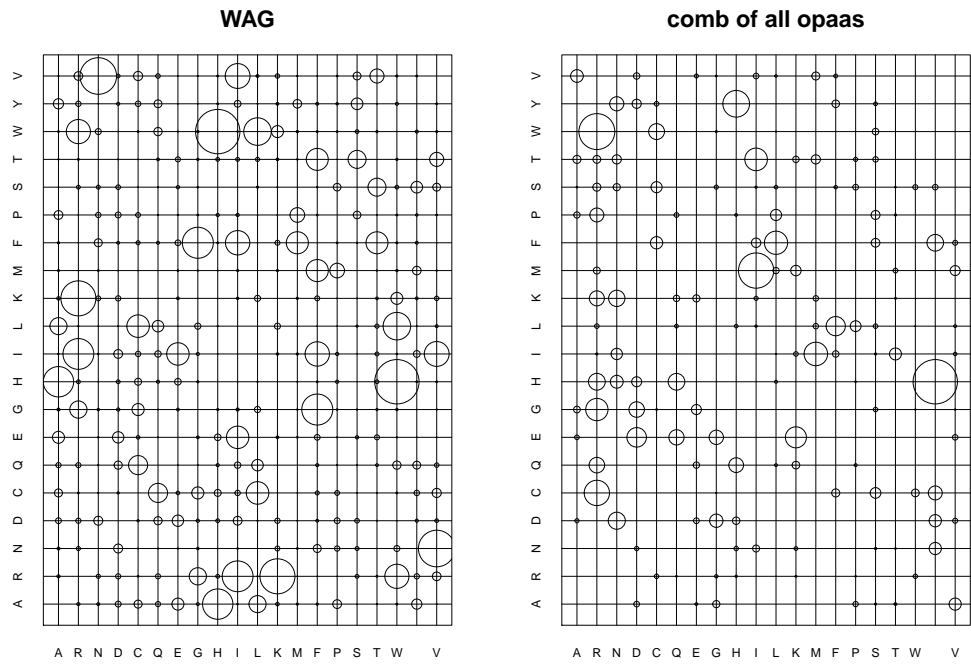


Figure 4: Substitution rate matrix under WAG model and a mash of substitution rate matrices for all amino acids being optimal.

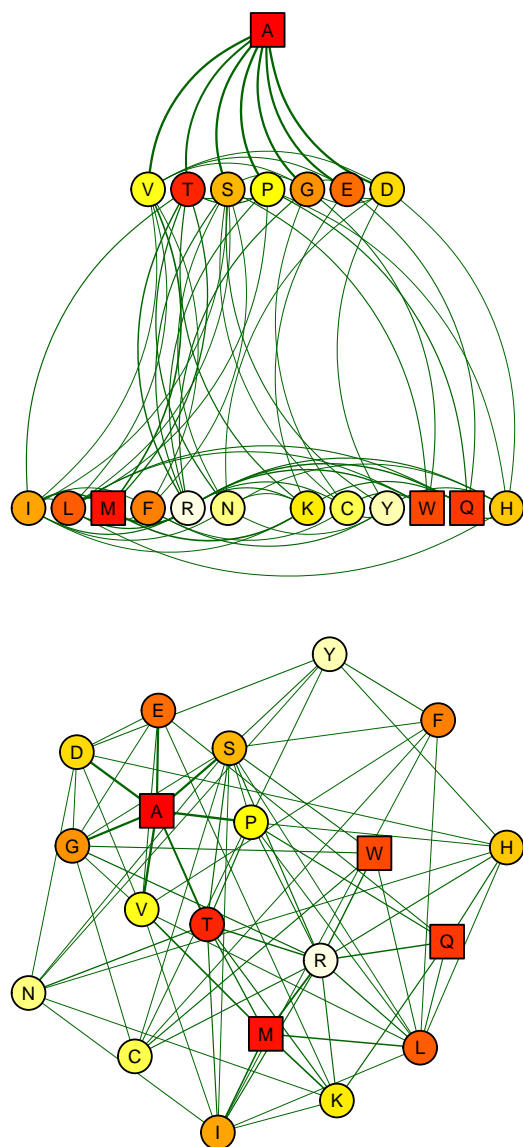


Figure 5: Another visualization of the substitution matrix. Squared nodes are local optima, the top one shows the neighboring amino acids more clearly.

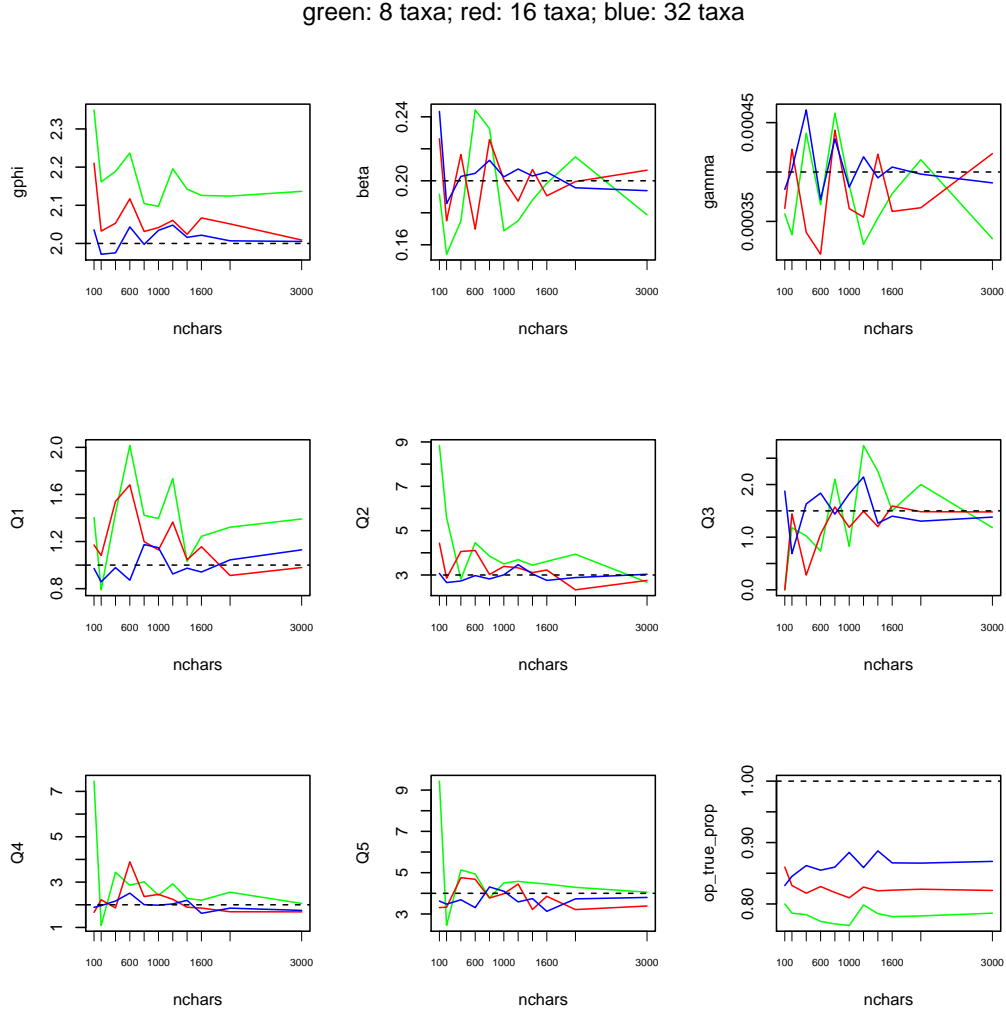


Figure 6: Parameter estimation from simulated data. The last panel is the proportion of the correctly inferred optimal amino acids of all sites. This shows that as the number of taxa or characters increases the parameter estimates are more accurate.

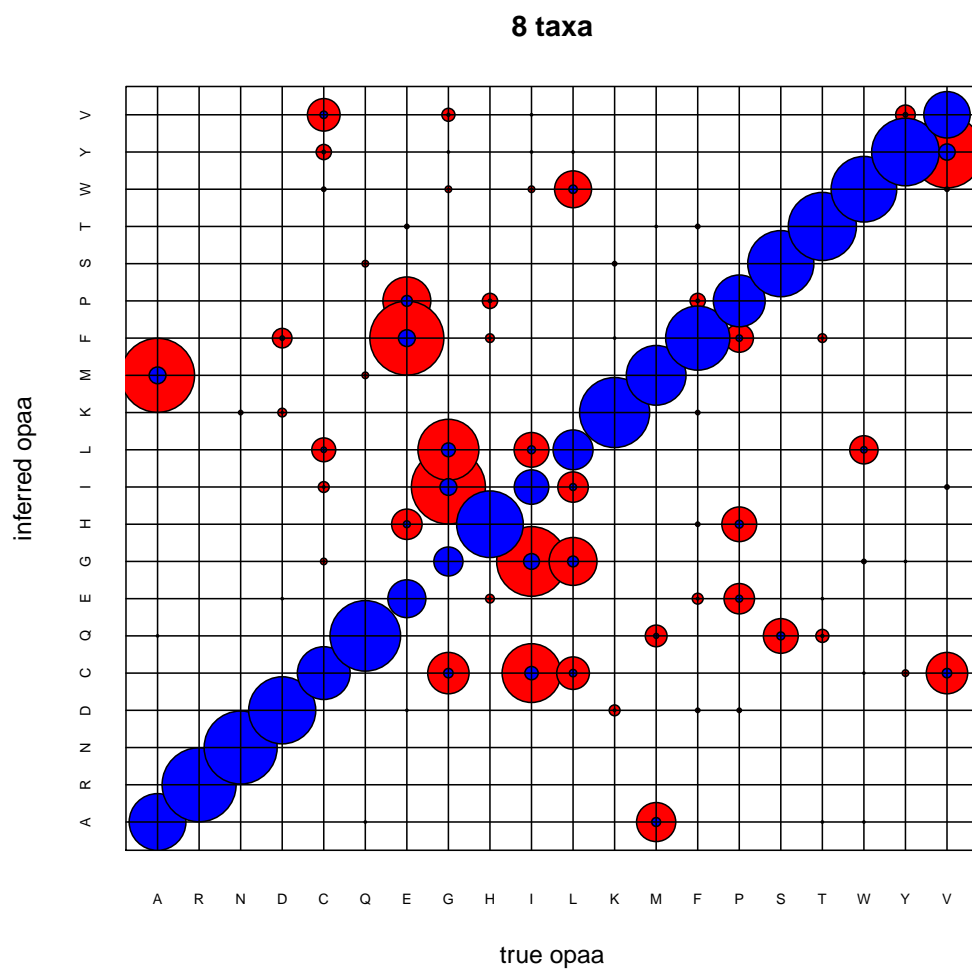


Figure 7: Inference of optimal amino acids from simulated data. Red ones are the magnified version of the incorrectly inferred optimal amino acids.

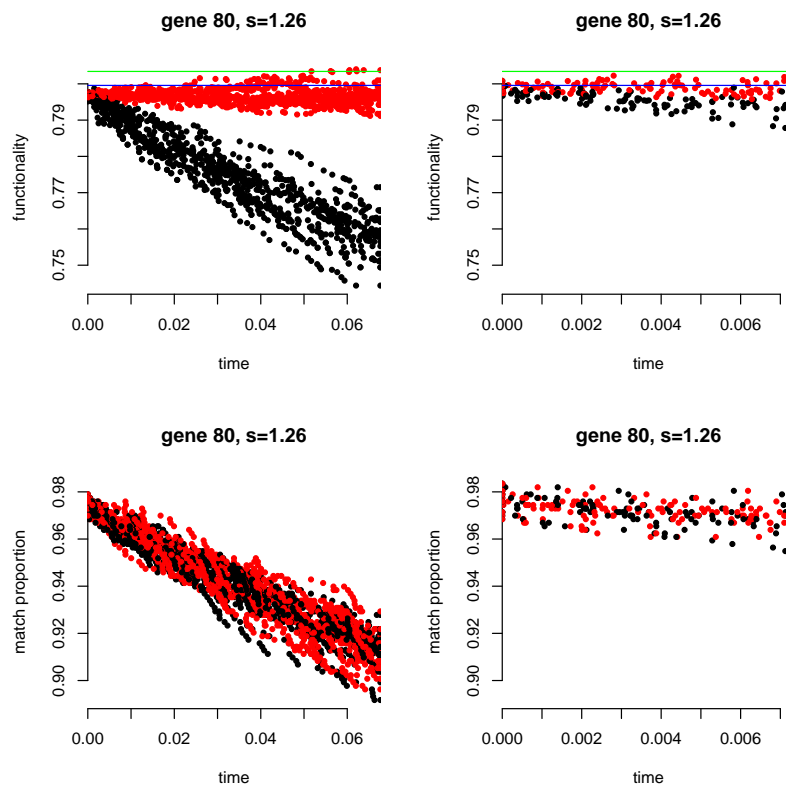


Figure 8: Model adequacy illustration. One external branch is broken off the tree for yeast data, and the grafted back, under both best empirical model for the gene and the new model. Top panels are the functionalities of protein sequence along the evolutionary path, and the bottom panels are the matching proportion to the observed data. Red is simulation under new model, black is simulation under empirical model. Blue line the observed functionality, green line is the equilibrium functionality.

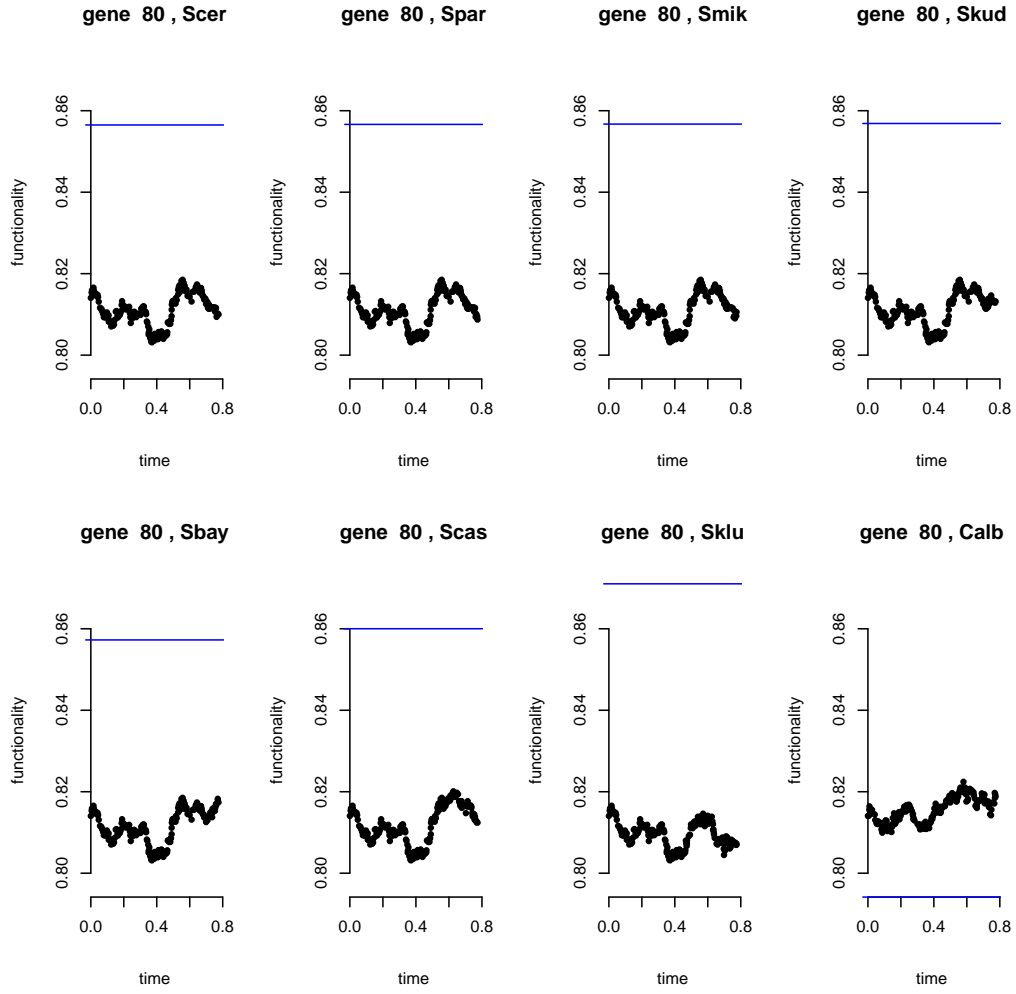


Figure 9: Same gene (80th), functionality of simulated sequence under the estimated parameter values, comparing to the observed sequences.

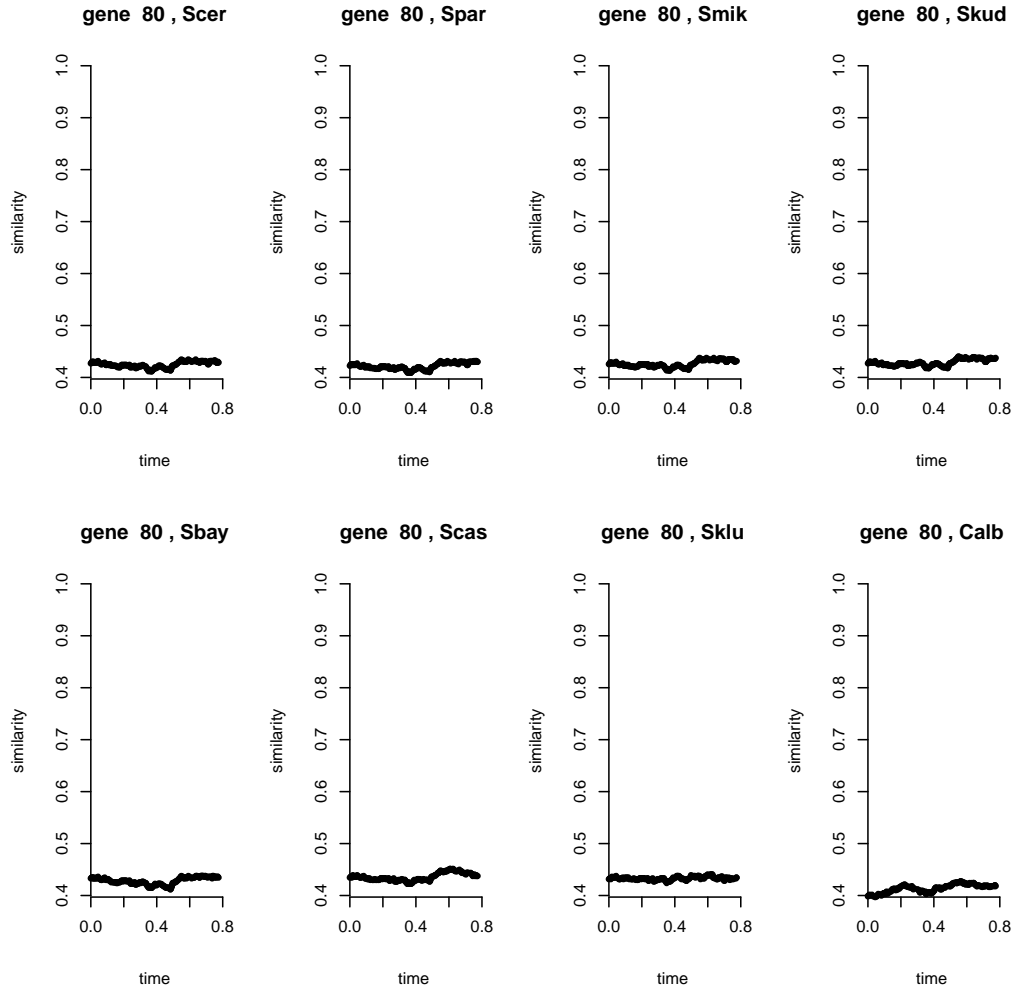


Figure 10: Same gene (80th), similarity of simulated sequence under the estimated parameter values, to the observed sequences.