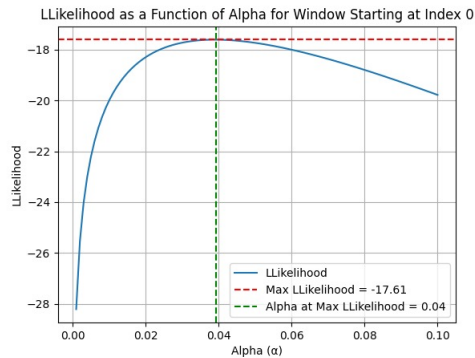


# Exercise 3 - Continuous-time Markov Model

Adi Rabinovitz, Avi Kupinsky

- For the following graph, The window we chose is of size 11, with the example of the phylogenetic tree and the sequences as we received with the file defining the assignment:



- The graph above shows how the log likelihood (of base  $e$ ) of observing the sequences changes between child sequence and parent sequence as the conservation coefficient,  $\alpha$ , varies
- The algorithm iterates through  $\alpha$  between the values 0 and 0.1 in increments of 0.001, and calculate the log likelihood for each one.
- The variable  $\alpha$  denotes the conservation coefficient used to adjust the size of all edges of the tree, either “inflating” or “shrinking” them.
- In the graph above, the curve begins at a low point, rises to a peak, and then declines, demonstrating the varying log likelihoods of different  $\alpha$  values producing on the sequences.
- The maximum in the graph above (which is at  $\alpha=0.04$ ) represents the conservation coefficient,  $\alpha$ , that optimizes the log likelihood of observing the sequences under the given phylogenetic tree. This  $\alpha$  value represents the level of conservation that best fits the observed sequence data according to the given tree.
- To determine the optimal  $\alpha$  value for each window, we used the **minimize\_scalar** function, specifically designed for minimizing scalar functions. Given our focus on maximizing the log likelihood, we made an **objective** function that returns the negative of the log likelihood.
- This approach is computationally efficient as it confines the search within a predetermined range and employs an optimization algorithm to identify the value that maximizes the log likelihood, eliminating the need to iterate over every possible  $\alpha$  value.