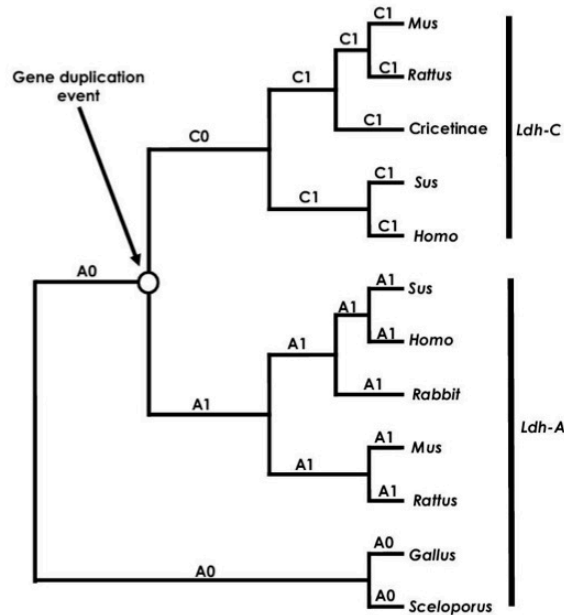


Exercise 3

The objective of this exercise is to use three LRTs to evaluate the following possibilities: (1) the mutation rate of Ldh-C has increased relative to Ldh-A, (2) a burst of positive selection for functional divergence occurred following the duplication event that gave rise to Ldh-C, and (3) there was a long term shift in selective constraints following the duplication event that gave rise to Ldh-C.

1. Obtain the files for Exercise 3 from the course web-site (**ex3_codeml.ctl**, **seqfile.txt**, **treeH0.txt**, **treeH1.txt**, **treeH2.txt**, **treeH3.txt**). The tree files represent different hypotheses denoted H0, H1, H2 & H3 (see diagram below, or download this file [genetree](#)). These evolutionary concepts described above are covered by these four precise hypotheses:
 - H0: Homogeneous selection pressure over the tree (concept = any rate differences are due to changes in mutation rate).
 - H1: Episodic change in selection pressure in Ldh-C (concept = only in the branch that immediately follows the gene duplication event).
 - H2: A long term shift in selection pressure in Ldh-C only (concept = Ldh-C has a novel selection pressure, as compared to its ancestors, whereas Ldh-A remains subject to the ancestral level of selection pressure).
 - H3: A long term shift in selection on both Ldh-C and Ldh-A (concept = both paralogous lineages experience novel selection pressures; i.e., different from each other, and from the ancestor).



- H₀:** $\omega_{A0} = \omega_{A1} = \omega_{C1} = \omega_{C0}$
H₁: $\omega_{A0} = \omega_{A1} = \omega_{C1} \neq \omega_{C0}$
H₂: $\omega_{A0} = \omega_{A1} \neq \omega_{C1} = \omega_{C0}$
H₃: $\omega_{A0} \neq \omega_{A1} \neq \omega_{C1} = \omega_{C0}$

2. Run CODEML using the settings in the control file for Exercise 3. Familiarize yourself with the results ([ex3_HelpFile.pdf](#)). In addition to the likelihood score you must be able to identify the branch-specific estimates of the ω parameter. (In the first run, the branch specific values for ω will all be the same. In later runs there will be differences among some branches).
3. As in the previous exercises, you will need to *change the control* file and re-run CODEML. The control file "quick guide" might be helpful here ([quick guide](#)). The objective is to compute the likelihood, and estimate ω parameters, under different models of how selection pressure changes in different parts of the tree. Because the relevant model information is contained in the tree file, you will need specify one of several tree files in each analysis, and change the control file so that it reads the required tree file.
 - As always, you should change the name of the main result file (via `outfile=` in the control file) or you will overwrite your previous results.
 - Change the model assumptions about branch specific ω values by changing the tree files (via `treefile=` and `model=`) set within the control file.

4. Repeat step 3 for each of the four tree files that have been provided to you. Again, keep track of your results by using a table like **Table E3** shown in the slides (see [ex3 table template.pdf](#)). In addition, carry out likelihood ratio tests (LRT) of the hypotheses below. See the lecture notes for additional details about LRTs. Use 1 degree of freedom to obtain the *P*-value for each LRT. (Helpful for computing the *P*-value: [Chi-Square calculator](#))

- H0 vs. H1
- H0 vs. H2
- H2 vs. H3

5. Use your table of results to determine:

- *Which model(s) are supported by the data?*
- *What evolutionary scenario is the best explanation of Ldh gene-family evolution?*
- *Is there evidence of positive selection during the history of Ldh evolution?*
- *Are there any scenarios in which Ldh could have evolved by positive selection that would be undetectable by these LRTs?*