**File Preparation:**

We need four files to run CodeML:

1) The multiple nucleotide (CDS) alignment, in PHYLIP format. CodeML will strictly remove any position that contains at least one gap or an unknown "N" nucleotide: TF105351.Eut.3.phy

2) The phylogenetic tree in newick format, with the branch of interest specified by "#1"(You can view it with NJplot or FigTree): TF105351.Eut.3.53876.tree



3) A command file where all parameters to run CodeML under the alternative model are specified: TF105351.Eut.3.53876.ctl

4) A command file where all parameters to run CodeML under the null model are specified: TF105351.Eut.3.53876.fixed.ctl

**Execute CodeML**

Run command file (alternative model):

We estimate the Ts/Tv ratio (fix\_kappa = 0) and the dN/dS (fix\_omega = 0). The branch-site model is specified by setting the model parameter to 2 (different dN/dS for branches) and the NSsites value to 2 (which allows 3 categories for sites: purifying, neutral and positive selection).

Run command file (null model):

The command file for the null model is the same as for the alternative model, except for two parameters (in red):

1) The name of the output file (outfile) is different.

2) The dN/dS ratio is fixed to 1 (fix\_omega = 1).

**Launch CodeML:**

In Unix (Linux, MacOSX), this will look like:

codeml ./TF105351.Eut.3.53876.ctl

codeml ./TF105351.Eut.3.53876.fixed.ctl

**Analyse results:**

1. Find the lnL in both the output files.
2. Calculate the Likelihood Ratio Test: 2\*ABS(lnL1 – lnL2)
3. Compare the Likelihood Ratio to the critical values of a [chi-squared distribution](https://people.richland.edu/james/lecture/m170/tbl-chi.html) with 1 degree of freedom.
4. If significant (p<0.05), which sites have evolved under positive selection?