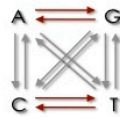


## Exercise 2

in this exercise you will investigate the sensitivity of  $\omega$  to the transition/transversion ratio (*kappa*), and to the assumed codon frequencies (*rt*s). After you collect the required data you will determine which assumptions yield the largest and smallest values of S, and what effect it has on the value of  $\omega$ .

transitions vs. transversions:



$Kappa (ts/tv)$   
= 2.71

preferred vs. un-preferred codons:

partial codon usage table for the *GstD* gene of *Drosophila*

Phe F TTT	0	Ser S TCT	0	Tyr Y TAT	1	Cys C TGT	0
<b>TTC</b>	<b>27</b>	<b>TCC</b>	<b>15</b>	<b>TAC</b>	<b>22</b>	<b>TGC</b>	<b>6</b>
Leu L TTA	0	TCA	0	*** * TAA	0	*** * TGA	0
TTG	1	TCG	1	TAG	0	Trp W <b>TGG</b>	<b>8</b>
Leu L CTT	2	Pro P CCT	1	His H CAT	0	Arg R CGT	1
CTC	2	<b>CCC</b>	<b>15</b>	CAC	4	<b>CGC</b>	<b>7</b>
CTA	0	CCA	3	Gln Q CAA	0	CGA	0
<b>CTG</b>	<b>29</b>	CCG	1	<b>CAG</b>	<b>14</b>	CGG	0

- Find the files for Exercise 2 (**ex2\_codeml.ctl**, **seqfile.txt**) and familiarize yourself with them. It would be best to create a new directory for Exercise 2. When you are ready to run CODEML, delete the **ex2\_** prefix (the control file must be called **codeml.ctl**).
- Run CODEML using the settings in the control file for Exercise 2. Familiarize yourself with the results ([ex2 HelpFile.pdf](#)). In addition to the likelihood score you must be able to identify the part of the result file that provides estimates of the following:
  - Number of synonymous or nonsynonymous sites (S and N))
  - Synonymous and nonsynonymous rates (dS and dN)
- As in Exercise 1, you will need to *change and save* the control files and re-run CODEML. The control file "quick guide" might be helpful here ([quick guide](#)). The objective is to compute the likelihood of the example dataset under different model assumptions. To do this you must:

- *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 codon frequencies (via `CodonFreq=`) and kappa (via `kappa=` and `fix_kappa=`).
4. Repeat step 3 for each set of assumptions about codon frequencies and kappa given as comments at the bottom of the example control file, and shown in the figure below.

Further details for about the assumptions tested in Exercise 2

<b>Assumption set 1:</b> Control file..	<b>Codon bias = none;</b> CodonFreq=0;	<b>Ts/Tv bias = none</b> kappa=1; fix_kappa=1
<b>Assumption set 2:</b> Control file..	<b>Codon bias = none;</b> CodonFreq=0;	<b>Ts/Tv bias = Yes</b> kappa=1; fix_kappa=0
<b>Assumption set 3:</b> Control file..	<b>Codon bias = yes [F3x4];</b> CodonFreq=2;	<b>Ts/Tv bias = none</b> kappa=1; fix_kappa=1
<b>Assumption set 4:</b> Control file..	<b>Codon bias = yes [F3x4];</b> CodonFreq=2;	<b>Ts/Tv bias = Yes</b> kappa=1; fix_kappa=0
<b>Assumption set 5:</b> Control file..	<b>Codon bias = yes [F61];</b> CodonFreq=3;	<b>Ts/Tv bias = none</b> kappa=1; fix_kappa=1
<b>Assumption set 6:</b> Control file..	<b>Codon bias = yes [F61];</b> CodonFreq=3;	<b>Ts/Tv bias = Yes</b> kappa=1; fix_kappa=0

5. In your favorite spreadsheet program create a table like **Table E2** in the slides (see [ex2 table template.pdf](#)) and fill it in with your results.
6. Use your table to determine:
- *Which assumptions yield the largest and smallest values of S?*
  - *What is the effect on  $\omega$ ?*
  - *What model would you choose?*