PAML description (note $\omega = d_N/d_S$ rate ratio)

1. Format sequences into a modified phylip format (see below). You can use Se-Al to edit the multiple alignment, and then export sequences into phylip format. The file starts with two numbers: # sequences # base pairs. The name of the sequence should be 10 characters in length including spaces; the last two characters must be spaces. Since we are analyzing codon data, the sequence length must be a multiple of three and no stop codons allowed. Save file with Unix line breaks using text edit such as bbedit. Place file in PAML folder. Example:

```
2   9
species1
atgagtcga
species2
atgcgagag
```

2. Generate a tree using Phylip, PAUP*, MrBayes or other program. Save file with Unix line breaks using text editor such as bbedit. Place file in PAML folder. Example tree: ((Species1,Species2),Species3)

3. Modify the codonml.ctl file to perform the analyses you want, save it with unix line breaks in PAML folder. Not all options in the command file are used. Typically modified options are:

Seqfile= any name you want for the phylip formatted aligned sequences

Treefile= any name you want for the treefile. No treefile needed for pairwise estimates

Runmode= use "0" for site and lineage variation analyses, use "-2" for pairwise

Model=0 if performing variation in the ω ratio between sites. 0 means one ω ratio for all lineages. For different ω ratios for each lineage ("free ratio" model) use 1.

NSsites= This parameter sets variation between sites. Use 0 for one ω ratio. The nested models typically used for variation between sites are:

Model 1 (neutral) versus Model 2 (selection)
Model 7 (beta, ω ratio between 0,1) versus Model 8 (beta+ω, additional ω)
You can specify multiple models to be run, for example: NSsites= 1 2 7 8

Fix_omega=0 for estimating the ω ratio, 1 to fix ω to test for significance

Omega=1. Start or fixed ω. To check for convergence, you should run the analysis a minimum of twice starting from different ω (< and >1, for example 0.2 and 2). If using Fix_omega, this is the value of ω.

ncatG= 1. The number relates to NSsites parameter, but is set automatically when multiple models are tested. Set to 1 for pairwise estimates and lineage variation

Common parameter setting for other options:

Seqtype=1 for all codon comparisons

CodonFreq=3 calculated codon table frequencies

4. The run the program, open the terminal window. Change in the directory that contains the PAML program, your sequence file and treefile. The > prompt, type: ./codeml codonml.ctl

5. View output in the outfile name you specified and the rst file automatically generated. Both will be located in the folder from which you executed the program.

Typical analyses performed in PAML:

1. Pairwise estimates. This averages the dn/ds ratio across all sites. You usually want to determine if the ω ratio is significantly different from 1. Run the program twice with different control files, such as:

To get the maximum likelihood estimates of ω:

Seqfile= whatever
Outfile= whatever you want
runmode = -2
model=0
NSsites=0
Fix_omega=0

To get the likelihood with dn/ds = 1:

Seqfile= whatever
Outfile= whatever, but different from previous
runmode = -2
model=0
NSsites=0
Fix_omega=1
Omega=1

The analysis is to compare the log likelihood where you estimate the ω ratio to the log likelihood where ω was fixed. The negative of twice of difference is compared to χ^2 with 1 degree of freedom (1 in this case)

2. Variation in the ω ratio between sites. You want to compare the likelihood of a model where the ω ratio is constrained to be equal or less than 1 to the likelihood of a model where a class of sites is allowed to have a ω ratio greater than 1. You can run all models by executing one command file with the following:

Seqfile= whatever you want to name the sequence file
Outfile= whatever you want outfile
runmode = 0
model=0
NSsites= 1 2 7 8
Fix_omega=0
Omega= try different values such as 0.3 and 3 to check for convergence.

The analysis is to compare the likelihood of the neutral versus the selection model. You compare the twice the difference between the log likelihood (lnl) of model 1 to model 2, and model 7 to model 8 with χ^2 distribution with degrees of freedom equal to the difference in the number of parameters. This should be run with different stating ω values (< and > 1) to check for convergence.

3. Variation in the ω ratio between lineages. You want to compare the likelihood where every lineage has the same ω ratio to the likelihood where each lineage (or a specific lineage of interest) has a different ω ratio. You compare twice the difference between the log likelihood (lnl) of the models with the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters. The command files should have the following:

One ratio:

runmode = 0
model=0
NSsites= 0

Free ratio:

runmode = 0
model=1
NSsites= 0