Katie Wendell
November 2020

How to use **PAML** on *BisonNet*

Module: molecular_evolution

Version: 4.9h

Description on BisonNet: Testing genes for selection

What it really does:

PAML is a group of programs that allows the phylogenetic analysis of DNA or protein sequences through use of maximum likelihood. **PAML** is an acronym which stands for *phylogenetic analysis by maximum likelihood*, the description is in the name! It is mainly a collection of the following programs when instructed to do so: *baseml, basemlg, codeml, evolver, pamp, yn00, mcmctree, and chi2*. These programs offer a wide range of analyses from comparing and testing phylogenetic trees to estimating synonymous and nonsynonymous substitution rates. The full list of **PAML** capabilities is listed below:

- Comparison and tests of phylogenetic trees (baseml and codeml);
- Estimation of parameters in sophisticated substitution models, including models of variable rates among sites and models for combined analysis of multiple genes or site partitions (baseml and codeml);
- Likelihood ratio tests of hypotheses through comparison of implemented models (baseml, codeml, chi2);
- Estimation of divergence times under global and local clock models (baseml and codeml);
- Likelihood (Empirical Bayes) reconstruction of ancestral sequences using nucleotide, amino
  acid and codon models (baseml and codeml);
- Generation of datasets of nucleotide, codon, and amino acid sequence by Monte Carlo simulation (evolver);
- Estimation of synonymous and nonsynonymous substitution rates and detection of positive selection in protein-coding DNA sequences (yn00 and codeml).
- Bayesian estimation of species divergence times incorporating uncertainties in fossil calibrations (mcmctree).

Katie Wendell
November 2020

How to Use **PAML** on BisonNet:

In preparation for using **PAML**, you need to make sure your sequence is in one of the following formats:

Sequence data file ("PHYLIP" format)-

```
   4  60
sequence 1
AAGCTTCACCGGCGCAGTCATTCTCATAAT
CGCCCACGGACTTACATCCTCATTACTATT
sequence 2
AAGCTTCACCGGCGCAATTATCCTCATAAT
CGCCCACGGACTTACATCCTCATTATTATT
sequence 3
AAGCTTCACCGGCGCAGTTGTTCTTATAAT
TGCCCACGGACTTACATCATCATTATTATT
sequence 4
AAGCTTCACCGGCGCAACCACCCTCATGAT
TGCCCATGGACTCACATCCTCCCTACTGTT
```

To Note:

The "4" represents the # of species.
The "60" represents the # of nucleotides.
- divide by 3 to get # of codons

The sequence name should be limited to 10 characters.
Name and sequence should be separated by at least two spaces

Tree file-

```
((1,2),3,4);    OR
((human:.1,chimpanzee:.2):.05,gorilla:.3,orangutan:.5);
```

Control file (____.ctl)

```
      seqfile = seqfile.txt         * sequence data filename
      outfile = results_0.txt   * main result file name

        noisy = 9      * 0,1,2,3,9: how much rubbish on the screen
      verbose = 1      * 1:detailed output
      runmode = -2     * -2:pairwise

      seqtype = 1      * 1:codons
    CodonFreq = 3      * 0:equal, 1:F1X4, 2:F3X4, 3:F61
        model = 0      *
     NSsites = 0       *
        icode = 0      * 0:universal code

    fix_kappa = 0      * 1:kappa fixed, 0:kappa to be estimated
        kappa = 2      * initial or fixed kappa

    fix_omega = 0     * 1:omega fixed, 0:omega to be estimated
        omega = 0.09    * 1st fixed omega value [change this]

     * EXCERCISE 1
     *alternate fixed omega values
     *omega = 0.005  * 2nd fixed value
     *omega = 0.01   * 3rd fixed value
     *omega = 0.05   * 4th fixed value
     *omega = 0.10   * 5th fixed value
     *omega = 0.20   * 6th fixed value
     *omega = 0.40   * 7th fixed value
     *omega = 0.80   * 8th fixed value
     *omega = 1.60   * 9th fixed value
     *omega = 2.00   * 10th fixed value
```

To Note:

- Make sure code name is codeml.ctl

- change omega values for desired dN/dS

Once the data is formatted, this code can be followed in order to run **PAML**.

Load the module

Katie Wendell
November 2020

```
module load molecular_evolution
```

Move the desired .ctl file file into your home directory and rename it codeml.ctl.

```
cp -rp /data/courses/biol325_evolgen/PAML_activity/ex1_codeml.ctl .
```

```
mv ex1_codeml.ctl codeml.ctl
```

Create directories for each omega value results. Then move your desired sequence and the codeml.ctl file into the first directory.

```
mkdir results_for_omega_equals_0.001
```

```
mv codeml.ctl results_for_omega_equals_0.001
```

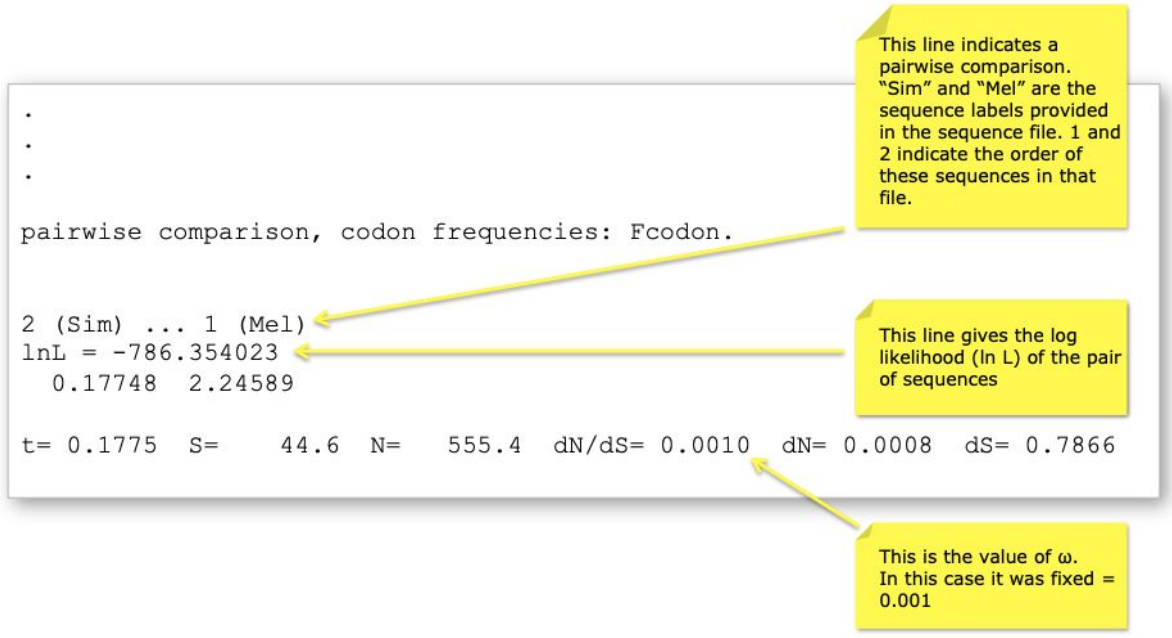
```
mv seqfile.txt results_for_omega_equals_0.001
```

Run CODEML using the following command, changing the .ctl file for the corresponding omega values. Repeat for each value.

```
nano codeml.ctl
```

```
/software/apps/paml/current/bin/codeml codeml.ctl
```

Finally, observe and record the results utilizing this graphic.



Katie Wendell
November 2020