INSTALLATION

The script will run on Linux operation system. To install, you must have the following softwares in place, please contact system admin if you do not know how to install,

   a. biopython (http://www.biopython.org)
   b. CLUSTALW (ftp://ftp.ebi.ac.uk/pub/software/clustalw2/)
   c. PAL2NAL (http://coot.embl.de/pal2nal/)
   d. PAML (http://abacus.gene.ucl.ac.uk/software/paml.html)

Please remember the installation path for CLUSTALW, PAL2NAL and PAML. You will then need to modify the script 'synonymous_calc.py' line 22-24 to change the binary path.

DATA PREPARATION

You need to prepare two FASTA files, one file with protein seqs, one file with gene seqs, they contain the gene pairs, e.g.

>gene1
ATAGATATATATA
>gene2
ATATAGAGAGAGA
>gene3
AGAGAGAGAGAGA
>gene4
ATAGAGAGAGAGA

This will calculate two pairs: gene1-gene2 Ks value and gene3-gene4 Ks value.
Make sure that your protein seq file corresponds to your gene seq file, in the same order.

Usage

Finally, run the command like this,
$ python synonymous_calc.py test.pep test.cds test.ks

where test.pep is your protein file, test.cds is your CDS file and your result is in test.ks.

The result is a comma-delimited file, you can open it in EXCEL, columns correspond to
Pair_ID; Yang-Nielson method Ks, Yang-Nielson method Ka, Nei-Gojobori method Ks, Nei-Gojobori method Ka