Ks calculation script

Author: Haibao Tang (tanghaibao), Brad Chapman
Email: tanghaibao@gmail.com

Installation

The python 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,

- biopython
- CLUSTALW
- PAL2NAL
- PAML

Please remember the installation path for CLUSTALW, PAL2NAL and PAML. You will then need to modify the script `synonymous_calc.py` line `~20` to change to the absolute path.

Preparing data

The input file is Fasta formatted, at least the CDS (coding DNA sequence) file needs to be given. Another protein sequence file is optional (if not given, the script will first try to translate the CDS file). The names for consecutive records are the gene pairs in which you wish to calculate Ks and Ka, like the following:

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

This will calculate two pairs: gene1-gene2 and gene3-gene4. Make sure that your protein seq file corresponds to your gene seq file, in exactly 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`.

You can in fact skip the protein file (`test.pep`), and only provide the CDS file (`test.cds`). The program will just assume the CDS file contains the frame-0 sequence and generate a file with translated sequence and continue. So the following command will also work:

```
$ python synonymous_calc.py test.cds >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
```

The script `report_ks.py` will generate a nice text-based report on the Ks output:

```
$ python report_ks.py test.ks
```

will generate the following report:

```
File `data/test.ks` contains a total of 166 gene pairs
------------------------------------------------------------
Yang-Nielson method of Ks estimate : 0.54285
  0.0|===============
  0.2|===
  0.4|=============
  0.6|======
  0.8|====
  1.0|===
  1.2|=
  1.4|
  1.6|
  1.8|
Yang-Nielson method of Ka estimate : 0.0697
  0.0|==============================
  0.1|===========
  0.2|===
  0.3|==
  0.4|
Nei-Gojobori method of Ks estimate : 0.45565
  0.0|=================
  0.2|====
  0.4|===============
  0.6|========
  0.8|=
  1.0|=
  1.2|
  1.4|
  1.6|
  1.8|
Nei-Gojobori method of Ka estimate : 0.067
  0.0|==============================
  0.1|===========
  0.2|====
  0.3|===
  0.4|
```

I personally recommend Nei-Gojobori method, from past experience. Nei-Gojobori method is a simple correction for multiple substitutions.