

BFCA tools v 1.01

BFCA tools are a set of Perl scripts to help in the generation of xml input files required for the multiple runs of BEAST and the posterior processing of log files.

Installation

The only requirement is to have a Perl 5.8.x (or higher) interpreter installed in your operating system. Most Linux distributions and Mac OS X have Perl installed by default. In Windows you can install any of the distributions recommended in <http://www.perl.org/get.html>.

The scripts are distributed in a compressed tar file (.tar.gz), so you will need to unpack and decompress them. On UNIX / UNIX-like environments execute:

```
tar xvfz bfca_tools_v1.tar.gz
```

In Windows you can use a tool like 7-zip to do it.

Usage

bfca_tool1_2by2.pl

This script is intended to generate xml files for BEAST with all pairwise combinations of calibration hypotheses given an xml template and a calibrations/constraints file. The syntax is:

```
bfca_tool1_2by2.pl -i <template xml> -c <calibrations file> -o <output directory>
```

When generating the BEAST xml template with BEAUTi you should follow these guidelines:

- You should leave the *Estimate* box/es in the *Clock Models* tab unchecked (uncheck all boxes in partitioned analyses) unless you want to set a specific prior distribution for the average evolutionary rate (parameter *clock.rate*, *ucl.d.mean* or *uced.mean*, depending on the clock model). By default (when left unchecked), the Perl script will change the xml code and a uniform distribution [0,100] will be used as prior in all cases. If you want to use a specific prior distribution, you should check (all) the *Estimate* box/es in the *Clock Models* tab and indicate the desired distribution/s in the *Priors* tab.
- If you are using a relaxed clock model with a parameter that sets the amount of variation in evolutionary rates among branches (e.g. *ucl.d.stdev* for the ULN clock model), we strongly recommend fixing it to a specific value or setting a specific prior distribution with (at least) an upper hard bound for that parameter. Such value or distribution should be estimated from an analysis without calibrations, reflecting this way the amount of rate variation estimated from the sequences alone.

The calibrations/constraints file must have the following structure (with columns separated by tabs):

Id	Taxa	Monophyly	Distribution	Parameter1	Parameter2	Constraint 1 / Offset	Constraint 2
group	taxa	yes/no	uniform	lower_bound	upper_bound		
group	taxa	yes/no	exponential	mean	offset		
group	taxa	yes/no	normal	mean	stdev		
group	taxa	yes/no	lognormal	mean	stdev	offset	
group	taxa	yes/no	gamma	shape	scale	offset	
group	taxa	yes/no	invgamma	mean	scale	offset	
group	taxa	yes/no	laplace	shape	scale	offset	
group	taxa	yes/no	truncated-normal	mean	stdev	lower_bound	upper_bound

An example of calibrations/constraints file with 5 calibrations and 2 topological constraints might be:

Id	Taxa	Monophyly	Distribution	Parameter1	Parameter2	Constraint1	Constraint2
Group1	A,B	no	lognormal 0.25	1	4.95		
Group2	C,D	no	exponential	0.5	10		
Group3	E,F	no	exponential	0.5	5		
Group4	A,B,C,D	no	lognormal 0.25	1	9.95		
Group5	E,F,G,H	no	truncated-normal	5	0.5	5	25
Ingroup	A,B,C,D,E,F,G,H	yes	NA				
Outgroup	I,J	yes	NA				

bfca_tool2_LH_merger.pl

This script is very helpful in cases where MCMC chains are very long. It simply extracts likelihood columns from the log files and put them together in a new file to avoid long reading times in R (and higher memory requirements). The syntax is:

```
bfca_tool2_LH_merger.pl -i <input directory> -o <output file>
```

The input directory should contain BEAST log files (with *.log* extension).

bfca_tool3_final.pl

This script generates a BEAST xml file given a template xml file and a calibrations file. Please note that although this script was written to generate the BEAST xml with the final congruent set of calibrations for the final analysis (avoiding the repetition of all the process with BEAUTi), it can also be helpful to incorporate a large set of calibration priors of topological constraints in a quicker (and perhaps even easier) way than using BEAUTi. The syntax is:

```
bfca_tool3_final.pl -i <template xml> -c <calibrations file> -o <output file>
```

The calibrations/constraints file must have the same structure described above (see *bfca_tool1_2by2.pl* section).

Citation

If you use the BFCA tools and/or the BFCA R package please cite:

Andújar, C., Soria-Carrasco, V., Serrano, J., and Gómez-Zurita, J. (2012) Congruence test of molecular clock calibration hypotheses based on Bayes factor comparisons