Fabric Model Post Processor

This post-processor implements the selection criteria that we apply to the stationary Markov chain to identify ‘significant’ directional and evolvability parameters in applications of the Fabric model (see [1]). The post processor converts the results of a BayesTraits run which includes evolvability scalars and directional change (betas) estimated using RJ MCMC into a table format, for further analysis. The script takes the tree file used in the BayesTraits analysis and the “.VarRates.txt” and “.Log.txt” file produced by the analysis and outputs a tab delimited table summarising the results on a per node bases. Please note this post processor is fundamentally different from the Variable Rates Post Processor (http://www.evolution.reading.ac.uk/VarRatesWebPP).

The example included here is not designed to make a biological statement but rather is used to illustrate how to run the program. The program may take a while to run depending on the size of the tree and the number of samples in the posterior.

Running the post processor

The easiest way to run the post processor is to use the provided binaries; download and unpack the file appropriate for your operating system. The program is run from the command line / terminal and takes a tree file, used in the original analysis a Variable rates file (VarRates.txt) and the log file (.Log.txt) produced by BayesTraits.

Linux / OS X

```
./FabricPostProcessor TreeFile.tree VarRates.txt LogFile.Log.txt
```

For example

```
./FabricPostProcessor Cat.trees Cat.VarRates.txt Cat.Log.txt
```

Windows

```
FabricPostProcessor.exe TreeFile.tree VarRates.txt LogFile.Log.txt
```

For example

```
FabricPostProcessor.exe Cat.trees Cat.VarRates.txt Cat.Log.txt
```

Running the python script

The post processor can be run using the python script directly, it requires a new (ish) version of python, and has been tested with version 3.6, 3.8 and 3.9. It requires three libraries numpy, scipy, ply. Its been tested with ply 3.11, scipy 1.3.1 and numpy 1.16.2 but newer (and possibly older) versions may work. The packages can be installed with pip3 using the command

```
pip3 install pandas numpy scipy ply
```

the program can be run.

```
python3 FabricPostProcessor.py TreeFile.tree VarRates.txt Log.txt
```

for the included example

```
python3 FabricPostProcessor.py Cat.trees Cat.VarRates.txt Cat.Log.txt
```

Output

The output is a csv (Comma-separated values) file, with the same name as the VarRates.txt file with csv appended, it can be loaded into programs like Excel and JMP. The headings are defined below.

With each row of the output relating to a node on the tree.

Header	Meaning
ID	ID for each node
Branch Length	Branch length from phylogenetic tree
Height	Height of the node, with the root having the largest height
Height Mid-Point	Height of the node at its mid-point
Sum BL	Sum of all branches that contained within the node
No Descendants	Number of descendant nodes, 0 for a tip, 2 for binary node
No Concurrent Linages Start	Number of concurrent lineages at the start of the node.
No Concurrent Linages Mid-point	Number of concurrent lineages at the mid-point of the node.
No Concurrent Linages End	Number of concurrent lineages at the end of the node.
Mean (Beta * BL)	Mean beta
SD (Beta * BL)	Beta standard deviation
Mean (Beta * BL) NZ	Mean beta excluding samples when the beta is set to zero. Note: we recommend that this should be used for estimating the size of the directional effect.
SD (Beta * BL) NZ	Standard deviation excluding samples when the beta is set to zero

Sample Size	Number of samples in the MCMC posterior
P < 0	Probability beta is less than 0
P == 0	Probability beta is 0
P > 0	Probability beta is greater than 0
P (Beta * BL)	The probability of Beta * BL, given the prior and threshold value
Observed P (Beta * BL)	The observed probability of Beta * BL in the Markov chain
SD Observed P(Beta * BL)	The standard deviation of observed P (Beta * BL)
Z (Beta * BL)	The Z score of Beta * BL
Sig (Beta * BL)	1 if the Z scores is significant (Z >= 2), 0 otherwise
Mean Evolvability Scalar	The mean scalar
SD Scalar	The standard deviation of the scalar
Mean Non 1 Scalar	The mean scalar excluding samples when it is 1. Note: we recommend that this should be used for estimating the size of the evolvability effect.
Median Non 1	The median scalar excluding samples when it is 1. Worth checking to see if the mean and median agree. Lack of agreement indicates a skewed posterior.
SD Non 1 Scalar	The standard deviation of the scalar excluding samples when it is 1
P < 1	The probability the scalar is less than one
P == 1	The probability the scalar is one
P > 1	The probability the scalar is greater than one
P Scalar	The probability of an evolvability scalar, given the prior and threshold value
Observed P Scalar	The observed probability of the evolvability scalar
SD Observed P Scalar	The standard deviation of observed P evolvability scalar
Z Scalar	The Z score of the evolvability scalar
Sig Scalar	1 if the Z scores is significant (Z >= 2), 0 otherwise
Prop beta followed by a variance scalar	The probability of a beta followed by an evolvability scalar
No Beta or Node	Number of times the sample has a beta or evolvability scalar
Prop Variance scalar followed by a beta	The probability of an evolvability scalar followed by a beta
Prop Beta on Dec Nodes	The probability of a beta on a descendant node
MD5 Sum	An MD5 check sum based on the taxa names that define the node, sorted into alphabetical order.
No Taxa	Number of taxa which define the node
Taxa List	A list of taxa names which define the node

Combining runs

The nature of MCMC analysis requires multiple independent runs to establish stationarity. The program CombineFabric (replaced MergeFabric) takes the results from the fabric post processor and combines them, identifying how often a directional change or node scalar was significant. Please see [1]

The program takes a number of tables created by the fabric post processor (above). The example below assumes that 6 files from the fabric post processor have been created, called T1.txt … T6.txt

Linux / OS X

```
./CombineFabric T1.csv T2.csv T3.csv T4.csv T5.csv T6.csv
```

Windows

```
CombineFabric.exe T1.csv T2.csv T3.csv T4.csv T5.csv T6.csv
```

The program produces a csv file called master.csv (which can be read into excel) that shows how often directional change or evolvability scalars were significant across the analysis provided. Where these fall on the tree can be identified via there MD5 check sum.

Header	Meaning
ID	A unique number for each node
No Sig Beta	The number of times the Z score of Beta * BL was >= 2.0
No Sig Nodes	The number of times Z score of the evolvability scalar was >= 2.0
Mean (Beta * BL) NZ -	The mean Beta * BL, when nonzero, for each of the input files
Mean Non 1 Scalar -	The mean node scalar value, when non one, for each of the input files
Z (Beta * BL)	The Z score of each Beta * BL, for each of the input files
Z Scalar	The Z score of each evolvability scalar, for each of the input files
All other columns	The remaining columns are taken from the first input file

Example

Six independent analyses of the fabric model were run on Marsupial body data and tree using the commands below. The data files were renamed Marsupials-001.txt … Marsupials-006.txt to prevent the analysis producing file names that clashed. The runs will all differ as the random seed as not been set.

```
7

2

Fabric

Burnin 5000000

Iterations 105000000

Sample 100000

Run
```

The results for each analysis were converted into csv files using the commands below.

Linux / OS X

```
./FabricPostProcessor Marsupials.trees Marsupials-001.txt.VarRates.txt Marsupials-001.txt.Log.txt

./FabricPostProcessor Marsupials.trees Marsupials-002.txt.VarRates.txt Marsupials-002.txt.Log.txt

./FabricPostProcessor Marsupials.trees Marsupials-003.txt.VarRates.txt Marsupials-003.txt.Log.txt

./FabricPostProcessor Marsupials.trees Marsupials-004.txt.VarRates.txt Marsupials-004.txt.Log.txt

./FabricPostProcessor Marsupials.trees Marsupials-005.txt.VarRates.txt Marsupials-005.txt.Log.txt

./FabricPostProcessor Marsupials.trees Marsupials-006.txt.VarRates.txt Marsupials-006.txt.Log.txt
```

Windows

```
FabricPostProcessor.exe Marsupials.trees Marsupials-001.txt.VarRates.txt Marsupials-001.txt.Log.txt

FabricPostProcessor.exe Marsupials.trees Marsupials-002.txt.VarRates.txt Marsupials-002.txt.Log.txt

FabricPostProcessor.exe Marsupials.trees Marsupials-003.txt.VarRates.txt Marsupials-003.txt.Log.txt
FabricPostProcessor.exe Marsupials.trees Marsupials-004.txt.VarRates.txt Marsupials-
004.txt.Log.txt
```

```
FabricPostProcessor.exe Marsupials.trees Marsupials-005.txt.VarRates.txt Marsupials-
005.txt.Log.txt
```

```
FabricPostProcessor.exe Marsupials.trees Marsupials-006.txt.VarRates.txt Marsupials
006.txt.Log.txt
```

This produces six csv files (Marsupials-001.txt.VarRates.txt.csv … Marsupials 006.txt.VarRates.txt.csv). These files can then be combined using the CombineFabric program to identify how often each of the independent analyses agree.

Linux / OS X

```
./CombineFabric Marsupials-001.txt.VarRates.txt.csv Marsupials-002.txt.VarRates.txt.csv
Marsupials-003.txt.VarRates.txt.csv Marsupials-004.txt.VarRates.txt.csv Marsupials-
005.txt.VarRates.txt.csv Marsupials-006.txt.VarRates.txt.csv
```

Windows

```
CombineFabric.exe Marsupials-001.txt.VarRates.txt.csv Marsupials-002.txt.VarRates.txt.csv
Marsupials-003.txt.VarRates.txt.csv Marsupials-004.txt.VarRates.txt.csv Marsupials-
005.txt.VarRates.txt.csv Marsupials-006.txt.VarRates.txt.csv
```

This will create a csv file (master.csv) summarising the six runs, from the file it will be possible to determine the location of the directional and evolvability effects within the tree. For example the branch that corresponds to the checksum de766dd96a5bd3ee3d43aed8dafa78f2 has a significant directional effect in all six runs, with an average z score of 117. The node f8679af46437940427a61adc8a584a85 has a significant variance scalar in all six runs with an average z score of 17.

1. Pagel, M., C. O'Donovan, and A. Meade, *General statistical model shows that macroevolutionary patterns and processes are consistent with Darwinian gradualism.* Nature Communications, 2022. **13**(1): p. 1113.