

```

library(RERconverge)

## Loading required package: RColorBrewer
## Loading required package: gplots
##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##      lowess
## Loading required package: ape
## Loading required package: phytools
## Loading required package: maps
## Loading required package: geiger

```

Read in some the tree data. This function puts trees with different number of present species into a unified format and therefore takes some time.

```

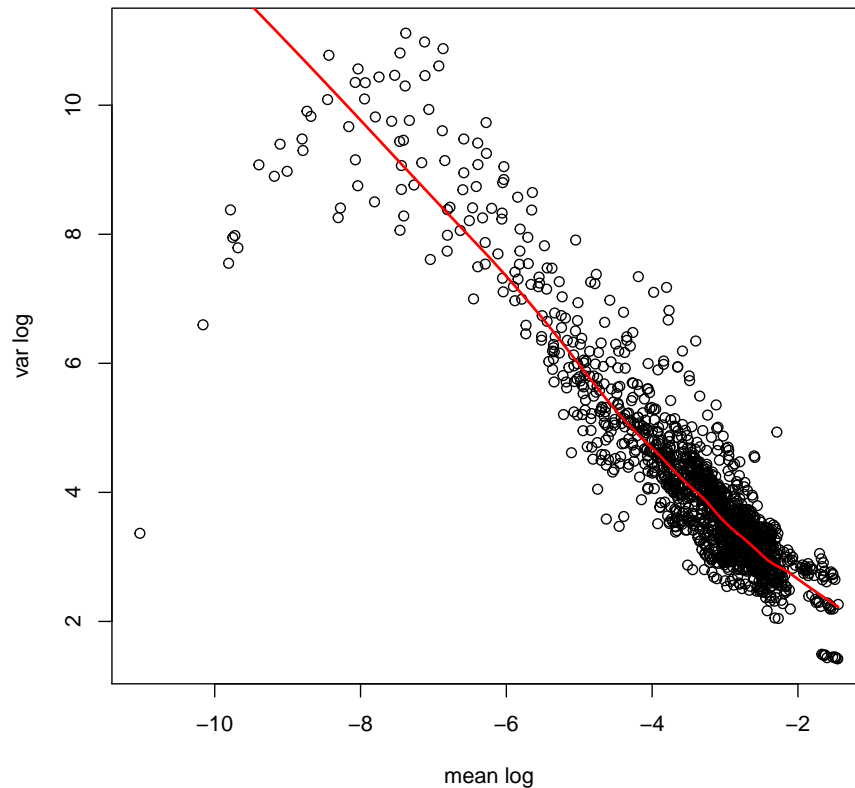
treefile=system.file("extdata", "mammal_62_aa_sub.tre", package = "RERconverge")
show(treefile)

## [1] "/home/mchikina/R/x86_64-pc-linux-gnu-library/3.4/RERconverge/extdata/mammal_62_aa_sub.tre"

mamTrees=readTrees(treefile,max.read = 200)

## max is 62
## estimating master tree branch lengths from 154 genes

```



This generates the mean variance plot for log transformed measurements and the corresponding weights fit. The plot can also be used to define a reasonable cutoff for the residual calculation based on the region of low values where the variance starts to decrease. Here around $\exp(-8)$. We set the cutoff to be $\exp(-7)$ (approximately 0.001) to be slightly more conservative.

Next we calculate the relative evolutionary rate (RER).

This is the basic method that performs a simple regression on the original CODEML output values. This takes some time as separate the expected rates are calculated for every subset of species that is represented by a single tree.

```
mamRER=getAllResiduals(mamTrees,transform = "none", cutoff=0.001)

## i= 1
## i= 2
## i= 9
## i= 11
## i= 15
```

```
## i= 24
## i= 28
## i= 35
## i= 50
## i= 51
## i= 68
## i= 69
## i= 73
## i= 78
## i= 79
## i= 81
## i= 91
## i= 98
## i= 104
## i= 107
## i= 108
## i= 109
## i= 110
## i= 112
## i= 114
## i= 120
## i= 123
## i= 129
## i= 143
## i= 144
## i= 148
## i= 150
## i= 151
## i= 153
## i= 173
## i= 176
## i= 178
## i= 183
## i= 184
## i= 187
## i= 192
## i= 193
```

We have found that scaling each branch generally improves results so we can generate a scaled version of the same RER data.

```
mamRERs=scale(mamRER)
```

This method performs a weighted regression in log space and has been found to perform better in benchmarks

```

mamRERlogW=getAllResiduals(mamTrees, transform = "log",weighted = T, cutoff=0.001)

## i= 1
## i= 2
## i= 9
## i= 11
## i= 15
## i= 24
## i= 28
## i= 35
## i= 50
## i= 51
## i= 68
## i= 69
## i= 73
## i= 78
## i= 79
## i= 81
## i= 91
## i= 98
## i= 104
## i= 107
## i= 108
## i= 109
## i= 110
## i= 112
## i= 114
## i= 120
## i= 123
## i= 129
## i= 143
## i= 144
## i= 148
## i= 150
## i= 151
## i= 153
## i= 173
## i= 176
## i= 178
## i= 183
## i= 184
## i= 187
## i= 192
## i= 193

```

These results can also be scaled, almost always improves results in our ex-

prience

```
mamRERlogWs=scale(mamRERlogW)
```