

An rbrothers tutorial

Jan Irvahn and Vladimir N. Minin

Department of Statistics, University of Washington Seattle, WA, 98195, USA

This is a tutorial demonstrating the usage of the java program, DualBrothers, in R. To get started, install the rbrothers R package from R-forge.

```
> install.packages("rbrothers", repos="http://R-Forge.R-project.org")
```

The rJava package and a couple of other packages will need to be installed.

Start R and load the rbrothers library.

```
> library(rbrothers)
```

Copy the KAL153.phy file from the rbrothers package to your current working directory.

```
> my.align = read.dna(file=system.file("extdata/KAL153/KAL153.phy",  
                                     package="rbrothers"))  
> write.dna(my.align, "KAL153.phy")
```

Now you can run DualBrothers with a single command.

```
> db<-dualbrothers(123, "KAL153", format="interleaved")
```

If you already have the output files of a DualBrothers run you can read the information in directly.

```
> db<-readdb("KAL153")
```

Plots can be created with the following commands.

```
> plot(db, threshold=.5)
```

```
> plottree.db(db, type="phylogram", threshold=.5)
```

The average number of breakpoints, the posterior probability of at least one breakpoint and DualBrothers parameters associated with the run can be easily accessed.

```
> summary(db)
```

DualBrothers output for KAL153

4 sequences of length 8588

MCMC settings:

length of the MCMC chain: 2100000

burn-in length: 100000

subsample frequency: 200

Prior parameters:

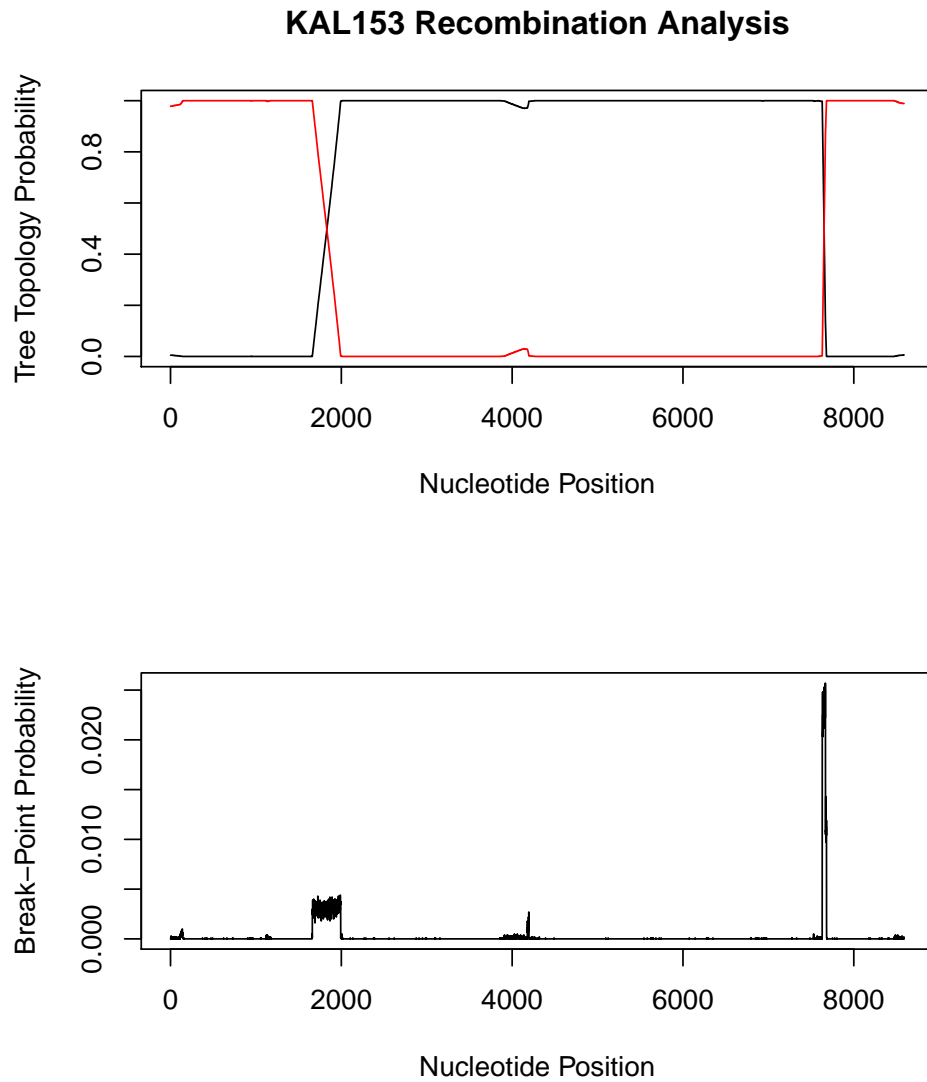


Figure 1: The top plot shows site specific posterior probabilities for the two most probable phylogenetic tree topologies. The bottom plot shows the site specific posterior probability of a breakpoint.

```
prior mean number of substitution process change points: 5
prior mean number of topology change points: 0.693
```

```
average number of breakpoints in the posterior: 2.1
```

```
The posterior probability of at least one breakpoint was 1 (> 0.999 required for a Bay
3 trees considered
```

You can calculate a 95% Bayesian credible interval for the first breakpoint.

```
> breakpointCI(db,1000,3000)
```

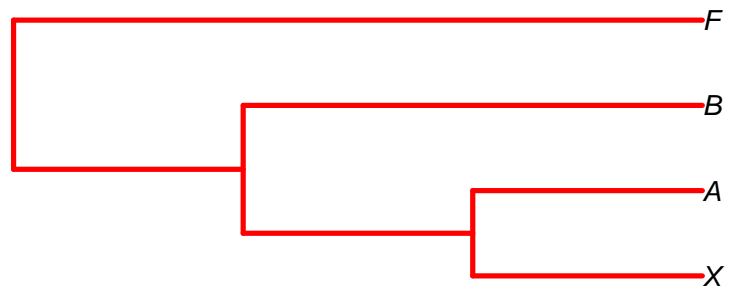
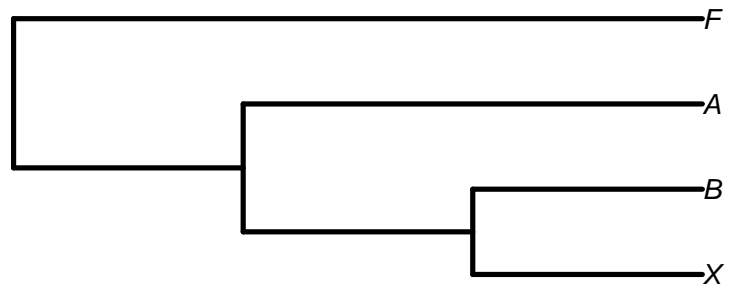


Figure 2: These are the two most probable phylogenetic tree topologies.

The 95% credible interval for a single break point between nucleotide number 1000 and nu
 2.5% 97.5%
 1666 1985