Package 'TipDatingBeast'

October 12, 2022

Encoding UTF-8

Type Package

Title Using Tip Dates with Phylogenetic Trees in BEAST

Version 1.1-0

Date 2019-10-21

Author Adrien Rieux, Camilo Khatchikian

Maintainer Camilo Khatchikian <ckhatchikian@gmail.com>

Depends R (>= 3.1.2)

Imports mclust, TeachingDemos, DescTools, utils, stats

Description Assists performing tip-dating of phylogenetic trees with BEAST
BEAST is a popular software for phylogenetic analysis.
The package assists the implementation of various phylogenetic tip-
dating tests using BEAST. It contains two main functions.
The first one allows preparing date randomization analyses,
which assess the temporal signal of a data set.
The second function allows performing leave-one-out analyses,
which test for the consistency between independent calibration sequences
and allow pinpointing those leading to potential bias.
The included tutorial provides detailed step-by-step instructions.
An expanded description of the package can be found in article:
Rieux, A. and Khatchikian, C.E. (2017),
TIPDATINGBEAST: an R package to assist the implementation of phylogenetic
tip-dating tests using BEAST. Molecular Ecology Resources, 17: 608-613.
<onlinelibrary.wiley.com/doi/full/10.1111/1755-0998.12603>.

License GPL (>= 2)

URL https://www.r-project.org

Repository CRAN

Date/Publication 2019-10-22 08:20:05 UTC

Archs i386, x64

NeedsCompilation no

R topics documented:

TipDatingBeast-package

Tip dating of phylogenetic trees with BEAST

Description

Assist performing tip-dating of phylogenetic trees with BEAST. Main functions include randomization of dates among tips and producing new input files with such randomization and generating input files for leave-one-out analyses using BEAST software.

Details

Package:	TipDatingBeast
Type:	Package
Version:	1.1.0
Date:	2018-08-05
License:	GPL (>= 2)

The RandomDates function randomize tip dates. The RandomCluster function ramdomize tip dates between group of tips. The TaxaOut function generates input files for leave-one-analysis. The TaxonOut function generates a single input file by removing the date of a particular taxon. The ListTaxa function displays the names and order of the taxa in the xml file. The PlotDRT function plots BEAST output of ramdomize tip dates analysis. The PlotLOOCV function plots BEAST output of leave-one-analysis.

Author(s)

Rieux A & Khatchikian, C. Maintainer: C. Khatchikian <ckhatchikian@gmail.com>

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613.

Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology And Evolution 29: 1969-1973.

ListTaxa   *List taxa names present in the BEAST input file*

Description

This function list the taxa names present in an input xml file for the BEAST version 1 software. The function is intended to help using the function "TaxonOut" as it allows the identification of the order of each taxa present in the input file.

Usage

```
ListTaxa(name)
```

Arguments

name   Name of the input file should be a .xml file generated using BEAUTi. Quote the name ("example"). Do not included .xml.

Details

The function works only with a .xml file generated with BEAUti

Value

The function returns the names and order of the taxa present in an input xml file for the BEAST software.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
    # using the example files "Flu-BEAST-1.8.xml" found in example folder.
# example file can be found in the example folder
ListTaxa("Flu-BEAST-1.8")
# list all 21 taxa in the file in the console

## End(Not run)
```

PlotDRT   *List taxa names present in the BEAST input file*

Description

Once analyzed with BEAST, the PlotDRT function allows comparing the parameter estimates obtained with the original vs. the randomized datasets. This function requires uploading, reading and analyzing the "Log" files generated by BEAST. A Log file contains a row for each MCMC sampling and a column for each estimated parameter. When considered as frequency distributions, this file provides an estimate of the marginal posterior probability distribution for each parameter. The function "PlotDRT" enable a graphical comparison of the parameter estimates obtained with the original vs. the date-randomized datasets (whatever the shuffling procedure considered).

Usage

```
PlotDRT(name, reps, burnin = 0.1)
```

Arguments

name   The name of the original Log file (excluding the .log extension) to upload and
            compute the real parameter estimates on. The name of the Log files computed
            from the date-randomized datasets should look like "name.Rep[i].log"

reps   The number date-randomized log files

burnin   the fraction of the first MCMC sampling to exclude from the Log files when
            computing the parameter estimates distribution (default = 0.1, which means 10

Details

The function works only after all BEAST runs are completed

Value

The function produce a two plots; one in normal scale, one in log scale.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
# example create with example file
PlotDRT("Flu_BEAST_1.8", reps = 20, burnin = 0.1)
# produce DRT plot

## End(Not run)
```

PlotLOOCV	*List taxa names present in the BEAST input file*

Description

The LOOCV (Leave-one-out-cross-validation) is a statistical procedure aiming to detect whether some particular sequences, when used as calibration in a tip dating analyse, could lead to systematic bias. Once the XMLs produced with the "TaxaOut" function analyzed with BEAST, the PlotLOOCV function allows a graphical comparison of the estimated vs true sequence ages. Similarly to the PlotDRT function, PlotLOOCV requires uploading, reading and analysing the "Log" files generated by BEAST.

Usage

```
PlotLOOCV(name, burnin = 0.1)
```

Arguments

name   the name of the original XML file (excluding the .xml extension). In the same folder, should be present the i LogFiles generated from the i XMLs produced with the "TaxaOut" function (i being the number of Taxa in the dataset). The name of the Log files should look like "name.Taxon[i].log".

burnin   the fraction of the first MCMC sampling to exclude from the Log files when computing the parameter estimates distribution (default = 0.1, which means 10

Details

The function works only with a .xml file generated with BEAUti

Value

The function returns the names and order of the taxa present in an input xml file for the BEAST software.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
# example create with example file
PlotLOOCV("Flu_BEAST_1.8", burnin = 0.1)
# produce LOOCV plot

## End(Not run)
```

| RandomCluster | *randomize dates among tips in the BEAST input file* |

Description

This function is similar to "RandomDates" excepts that in "RandomCluster", samples are grouped into clusters and the shuffling procedure randomizes dates among the clusters but not within (see manual for more details on this procedure). There are two distinct ways to group the samples into clusters. The first one is through the upload of a csv.file containing the names of the samples (as given in the XML) and a cluster number. Any positive integer (i.e., positive number) can be used to identified cluster; if a "0" is given to any sample, it would be excluded from the procedure. The file containing the classification should be labeled: clusters."name".cvs. An example for such a file in the case of the Influenza dataset can be found distributed with the package.

In a second approach, a model-based clustering classification is automatically performed using the mclust library. In this case, the option loadCluster should be set to FALSE (loadCluster = F). If this option is chosen, the new classification is written in a csv file that is labeled: clusters."name".cvs.

Usage

```
RandomCluster(name, reps = 20, loadCluster = T, writeTrees = T)
```

Arguments

name	The name of the original XML-formatted input file on which to apply the date-randomization procedure. Quote the name ("example"). The .xml extension should not be included.
reps	The number of repetions required by the user. There will be as many date-randomized datasets produced as the value of reps (default = 20).
loadCluster	F or T (default T). If T, clusters are loaded from a cluster structure file. The file containing the cluster structure needs to follow the example provided. Any tip assigned to cluster "0" will not be included in any randomization. Tip dates will only be randomized between (and not within) clusters. The cluster file should be named "clusters.NAME.csv" where NAME is the XML file name. If F, clusters are calculated using the package "mclust" procedure and an output cvs file containing the cluster structure is produced.
writeTrees	This argument has no function in the current version (default = T).

Details

The function works only with a .xml file generated with BEAUti

Value

The function returns one or many files (the number is set by the "reps" argument; default is 20) In each new file, the date values are randomized among tips.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology And Evolution 29: 1969-1973. Fraley C & Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association 97: 611-631.

Examples

```
## Not run:
   # using the example files "Flu_BEAST_1.8.xml" and "clusters.Flu.csv" found in example folder
RandomCluster("Flu_BEAST_1.8", reps = 20, loadCluster = T)
# produce 20 replicate input files (.xml) in working directory

## End(Not run)
```

RandomDates   *randomize dates among tips in the BEAST input file*

Description

This function read an XML file (BEAST version 1), randomize dates among all tips and produce a new XML input file with such randomization. The process is repeated up to the number of replicates (default is 20).

Usage

```
RandomDates(name, reps = 20, writeTrees = T)
```

Arguments

name	The name of the original XML-formatted input file on which to apply the date-randomization procedure. Quote the name ("example"). The .xml extension should not be included.
reps	The number of repetions required by the user. There will be as many date-randomized datasets produced as the value of reps (default = 20).
writeTrees	This argument has no function in the current version (default = T).

Details

The function works only with a .xml file generated with BEAUti

Value

The function returns one or many files (the number is set by the "reps" argument; default is 20) In each new file, the date values are randomized among all tips.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
    # using the example files "Flu-BEAST-1.8.xml" found in example folder.
# example file can be found in the example folder
RandomDates("Flu-BEAST-1.8", reps = 20)
# produce 20 replicate input files (.xml) in working directory

## End(Not run)
```

TaxaOut   *generates input files for leave-one-out analyses in BEAST*

Description

This function produces input files to perform leave-one-out analyses using BEAST version 1 software. As many files as taxa present in the input file are produced; each one leaving each consecutive taxon out for analysis.

Usage

```
TaxaOut(name, lBound = 0, hBound = 1.0E100, writeTrees = T)
```

Arguments

name	The name of the original XML-formatted input file on which to apply the LOOCV procedure (the .xml extension should be excluded). This xml should be set up so that earlier dates have lower numerical values (i.e., set direction = "forwards" when setting up date in BEAUti). Place the name between quotes ("example").
lBound	The uniform prior lower bound for the age of the missing taxa (default = 0)
hBound	The uniform prior higher bound for the age of the missing taxa (default = 1.0E100)
writeTrees	This argument has no function in the current version (default = T).

Details

The function works only with a .xml file generated with BEAUti.

Value

The function returns as many files as taxa are present in the input file; each one leaving each consecutive taxon out for analysis.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
    # using the example files "Flu_BEAST_1.8.xml" found in example folder.
TaxaOut("Flu_BEAST_1.8")
# produce 21 input files, each one without the corresponding taxon

## End(Not run)
```

TaxonOut   *single Taxon out input file for leave-one-out analysis in BEAST*

Description

This function produces a single input file for leave-one-out analyses using BEAST version 1 software. In this analysis, the date of the chosen taxon is estimated using the remaining taxa dates. The function "ListTaxa" is intended to help identifying the order (parameter "takeOut") of the taxon desired for the leave-out-analysis.

Usage

```
TaxonOut(name, lBound = 0, hBound = 1.0E100, takeOut, writeTrees = T)
```

Arguments

name	Name of the input file should be a .xml file generated using BEAUTi. Quote the name ("example"). Do not included .xml.
lBound	The uniform prior lower bound for the age of the missing taxa (default = 0)
hBound	The uniform prior higher bound for the age of the missing taxa (default = 1.0E100)
takeOut	Taxon order for the take-one-out analysis.
writeTrees	This argument has no function in the current version (default = T).

Details

The function works only with a .xml file generated with BEAUti

Value

The function returns a single file to perform a leave-one-out analyses using BEAST software for the specific taxon.

References

Rieux, A. and Khatchikian, C.E., 2017. TipDatingBeast: An R package to assist the implementation of phylogenetic tip-dating tests using BEAST. Molecular ecology resources, 17(4), pp.608-613. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology And Evolution 29: 1969-1973.

Examples

```
## Not run:
    # using the example files "Flu_BEAST_1.8.xml" found in example folder.
# example using the 5th taxon ("CHICKEN_HONGKONG_915_1997")
TaxonOut("Flu_BEAST_1.8", takeOut = 5)
# produce a single input files without the corresponding taxon ("CHICKEN_HONGKONG_915_1997")

## End(Not run)
```

Index