Package 'OptM'

June 17, 2024

Title Estimating the Optimal Number of Migration Edges from 'Treemix'

Version 0.1.8

Description The popular population genetic software 'Treemix' by
'Pickrell and Pritchard' (2012) <DOI:10.1371/journal.pgen.1002967>
estimates the number of migration edges on a population tree.
However, it can be difficult to determine the number of migration
edges to include. Previously, it was customary to stop adding migration
edges when 99.8% of variation in the data was explained, but 'OptM'
automates this process using an ad hoc statistic based on the
second-order rate of change in the log likelihood. 'OptM' also
has added functionality for various threshold modeling
to compare with the ad hoc statistic.

Maintainer Robert Fitak <rfitak9@gmail.com>

Author Robert Fitak [aut, cre]

Depends R (>= 3.2.2)

License GPL (>= 2)

Encoding UTF-8

NeedsCompilation no

Imports SiZer (>= 0.1-4), stats, grDevices

Date 2024-6-16

Repository CRAN

RoxygenNote 7.3.1

Date/Publication 2024-06-17 03:50:02 UTC

Contents

---

optM   *optM function*

Description

Load a folder of .llik files from the program Treemix and determine the optimal number of migration edges to include

Usage

```
optM(
  folder,
  orientagraph = F,
  tsv = NULL,
  method = "Evanno",
  ignore = NULL,
  thresh = 0.05,
  ...
)
```

Arguments

folder	A character string of the path to a directory containing .llik, .cov.gz and .model-cov.gz files produced by Treemix
orientagraph	A logical indicating whether the files were produced from Treemix (FALSE) or OrientAGraph (TRUE). Default = F
tsv	a string defining the name of the tab-delimited output file. If NULL (default), then no data file is produced.
method	a string containing the method to use, either "Evanno", "linear", or "SiZer". Default is "Evanno".
ignore	a numeric vector of whole numbers indicating migration edges to ignore. Useful when running Treemix on a prebuilt tree (ignore = 0). Default is NULL.
thresh	a numeric value between 0 and 1 for the threshold to use for the proportion of increase in likelihood that defines when a plateau is reached. Default is 0.05 (5%), only applicable for method = "linear".
...	other options sent to the function "SiZer" - see the R package 'SiZer'

Value

If method = "Evanno": A data frame with 17 columns summarizing the results for each migration edge (rows).

The columns are: "m" - number of migration edges from the model; "runs" = number of iterations for "m"; "mean(Lm)" - mean log likelihood across runs; "sd(Lm)" - standard deviation of log likelihood across runs; "min(Lm)" - minimum log likelihood across runs; "max(Lm)" - maximum log likelihood across runs; "L'(m)" - first-order rate of change in log likelihood; "sdL'(m)"

- standard deviation of first-order rate of change in log likelihood; "minL'(m)" - minimum first-order rate of change in log likelihood; "maxL'(m)" - maximum first-order rate of change in log likelihood; "L''(m)" - second-order rate of change in log likelihood; "sdL''(m)" - standard deviation of the second-order rate of change in log likelihood; "minL''(m)" - minimum second-order rate of change in log likelihood; "maxL''(m)" - maximum second-order rate of change in log likelihood; "Deltam" - the ad hoc deltaM statistic (secord order rate of change in log likelihood); "mean(f)" - mean proportion of variation explained by the models; "sd(f)" - standard deviation of the proportion of variation explained by the models

If method = "linear": A list containing 5 elements:

$out - a data frame with the name of each model, the degrees of freedom (df), the Akaike information criterion (AIC), the deltaAIC, and the optimal estimate for m based on the model.

$PiecewiseLinear - the piecewise linear model object

$BentCable - the bent cable model object

$SimpleExponential - the simple exponential model object

$NonLinearLeastSquares - the NLS model object

If method = "SiZer": an object of class "SiZer" (see the R package 'SiZer' for more information)

Examples

```
# Load a folder of simulated test data for m = 3
folder <- system.file("extdata", package = "OptM")
test.optM = optM(folder)

# To view the various linear modeling estimates:
  # test.linear = optM(folder, method = "linear")

# To view the results from the SiZer package:
  # test.sizer = optM(folder, method = "SiZer")
```

---

plot_optM   *plot_optM function*

Description

Plotting the optM results. This function visualizes the output of optM, including the amount of total variation explained across each value of the migration rate

Usage

```
plot_optM(input, method = "Evanno", plot = TRUE, pdf = NULL)
```

Arguments

`input`	an object produced by the fucntion 'optM'
`method`	a string containing the method to use, either "Evanno", "linear", or "SiZer". Default is "Evanno", but needs to match that used in 'optM'
`plot`	logical of whether or not to display the plot
`pdf`	string of the file name to save the resulting pdf plot. If NULL, no file is saved. Default is NULL

Value

a plot or pdf of a plot

Examples

```
# Load a folder of simulated test data for m = 3
folder <- system.file("extdata", package = "OptM")
# Run the Evanno method and plot the results
test.optM = optM(folder)
plot_optM(test.optM, method = "Evanno")

# To view the various linear modeling estimates and plot:
   # test.linear = optM(folder, method = "linear")
   # plot_optM(test.linear, method = "linear")

# To view the results from the SiZer package:
   # test.sizer = optM(folder, method = "SiZer")
   # plot_optM(test.sizer, method = "SiZer")
```

Index