

Package ‘geneplast’

April 14, 2017

Type Package

Title Evolutionary and plasticity analysis based on orthologous groups distribution

Version 1.0.0

Date 2016-06-13

Author Rodrigo Dalmolin, Mauro Castro

Maintainer Mauro Castro <mauro.a.castro@gmail.com>

Depends R (>= 3.3), methods

Imports snow,ape,grDevices,graphics,stats,utils

Suggests BiocStyle, RUnit, BiocGenerics

Description Geneplast is designed for evolutionary and plasticity analysis based on orthologous groups distribution in a given species tree. It uses Shannon information theory and orthologs abundance to estimate the Evolutionary Plasticity Index. Additionally, it implements the Bridge algorithm to determine the evolutionary root of a given gene based on its orthologs distribution.

biocViews Genetics, GeneRegulation, SystemsBiology

Collate AllChecks.R AllClasses.R AllGenerics.R AllMethods.R AllConstructors.R AllSupplements.R AllPlots.R

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

R topics documented:

geneplast-package	2
gpdata	2
gplast	3
gplast.get	4
gplast.preprocess	5
groot	6
groot.get	8
groot.plot	9
groot.preprocess	10
OGP-class	11
OGR-class	12

Index**13**

geneplast-package	<i>Geneplast: an R package for evolutionary rooting and plasticity inference based on orthologous groups distribution.</i>
-------------------	--

Description

Geneplast is designed for evolutionary and plasticity analysis based on orthologous groups distribution in a given species tree. It uses Shannon information theory and orthologs abundance to estimate the Evolutionary Plasticity Index. Additionally, it implements the Bridge algorithm to determine the evolutionary root of a given gene based on its orthologs distribution.

Details

Package: geneplast
 Type: Package
 Version: 1.0
 Date: 2016-06-10
 License: GPL (>= 2)

R package for gene plasticity inference based on orthologous groups distribution.

Author(s)

Rodrigo JS Dalmolin, Mauro AA Castro.

References

Dalmolin RJS et al. Geneplast: an R package for gene plasticity inference based on orthologous groups distribution. Journal paper (in preparation).

gpdata	<i>A pre-processed dataset for the geneplast package.</i>
--------	---

Description

A dataset used to demonstrate geneplast main functions.

Usage

```
data(gpdata.gs)
```

Format

gpdata.gs data frame containing information about clusters of orthologous groups.

Details

The dataset consists of 4 R objects used in the geneplast vignettes.

cogdata A data frame with three columns containing the Clusters of Orthologous Groups information retrieved from the STRING database (<http://string-db.org/>), release 9.1. Column 1 = Ensembl protein ID; column 2 = NCBI species ID; column 3 = OG ID.

sspids A data frame containing the species listed in STRING database (<http://string-db.org/>), release 9.1. Column 1 = NCBI species ID; column 2 = NCBI species name; column 3 = species domain (eukaryotes).

cogids A one-column data.frame with orthologous groups IDs available in the 'cogdata' object.

phyloTree An object of class "phylo" listing the eukaryotes in the STRING database, release 9.1.

Value

a dataset.

References

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D808-15. doi: 10.1093/nar/gks1094. Epub 2012 Nov 29.

Examples

```
data(gpdata.gs)
```

gplast	<i>Evolutionary plasticity inference.</i>
--------	---

Description

Function to calculate abundance, diversity, and evolutionary plasticity of an orthologous group (OG).

Usage

```
gplast(object, verbose=TRUE)
```

Arguments

object	this argument is an object of class 'OGP' (OGP-class).
verbose	a single logical value specifying to display detailed messages (when verbose=TRUE) or not (when verbose=FALSE).

Details

This method computes the abundance and diversity of an OG, and derives the evolutionary plasticity as described in Castro et al. (2008) and Dalmolin et al. (2011). The OG diversity corresponds to the normalized Shannon's diversity index and estimates the distribution of orthologous proteins across the species listed in the input dataset. The OG abundance is simply the ratio between the number of orthologs of a given OG and the number of organisms listed in the group. Evolutionary Plasticity Index is calculated as described in Dalmolin et al (2011).

Value

A processed object of class 'OGP', including COG's abundance, diversity, and plasticity results.

Author(s)

Rodrigo Dalmolin, Mauro Castro

References

Dalmolin, RJ, Castro, MA, Rybarczyk-Filho JL, Souza LH, de Almeida RM, Moreira JC. Evolutionary plasticity determination by orthologous groups distribution. Biol Direct. 2011 May 17;6:22. DOI: 10.1186/1745-6150-6-22.

Castro MA, Dalmolin RJ, Moreira JC, Mombach JC, de Almeida RM. Evolutionary origins of human apoptosis and genome-stability gene networks. Nucleic Acids Res. 2008 Nov;36(19):6269-83. DOI: 10.1093/nar/gkn636.

See Also

[OGP-class](#)

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGP'
ogp <- gplast.preprocess(cogdata=cogdata, sspids=sspids, cogids=cogids)

## run the gplast function
## this example uses the especies/COGs listed in the gpdata object
ogp <- gplast(ogp)
res <- gplast.get(ogp,what="results")
```

gplast.get

Get information from individual slots in an OGP object.

Description

Get information from individual slots in an OGP object and any available results from previous analysis.

Usage

```
gplast.get(object, what="status")
```

Arguments

object an object of class 'OGP' [OGP-class](#).

what a single character value specifying which information should be retrieved from the slots. Options: "cogids", "sspids", "orthodist", "results", and "status".

Value

slot content from an object of class 'OGP' [OGP-class](#).

Author(s)

Rodrigo Dalmolin, Mauro Castro

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGP'
ogp <- gplast.preprocess(cogdata=cogdata, sspids=sspids, cogids=cogids)

## run the gplast function
## this example uses the especies/COGs listed in the gpdata object
ogp <- gplast(ogp)
res <- gplast.get(ogp,what="results")
```

`gplast.preprocess` *Evolutionary plasticity inference.*

Description

Constructor for the 'OGP-class' object.

Usage

```
gplast.preprocess(cogdata, sspids=NULL, cogids=NULL, verbose=TRUE)
```

Arguments

<code>cogdata</code>	a data frame or matrix with COG's data.
<code>sspids</code>	an optional data frame with species annotation. Alternatively, it can be a character vector with species IDs.
<code>cogids</code>	an optional data frame with COG's annotation. Alternatively, it can be a character vector with COG IDs.
<code>verbose</code>	a single logical value specifying to display detailed messages (when <code>verbose=TRUE</code>) or not (when <code>verbose=FALSE</code>).

Details

This function creates an [OGP-class](#) object and checks the consistency of the input data for the evolutionary plasticity pipeline. Internally, the function counts the number of orthologs for each species in a given OG and computes the orthodist matrix.

Value

A preprocessed object of class 'OGP'.

Author(s)

Rodrigo Dalmolin, Mauro Castro

References

Dalmolin, RJ, Castro, MA, Rybarczyk-Filho JL, Souza LH, de Almeida RM, Moreira JC. Evolutionary plasticity determination by orthologous groups distribution. *Biol Direct*. 2011 May 17;6:22. DOI: 10.1186/1745-6150-6-22.

Castro MA, Dalmolin RJ, Moreira JC, Mombach JC, de Almeida RM. Evolutionary origins of human apoptosis and genome-stability gene networks. *Nucleic Acids Res*. 2008 Nov;36(19):6269-83. DOI: 10.1093/nar/gkn636.

See Also

[OGP-class](#)

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGP'
ogp <- gplast.preprocess(cogdata=cogdata, sspids=sspids, cogids=cogids)
```

groot

Evolutionary rooting inference.

Description

Function to determine the evolutionary root of a gene based on its orthologs distribution.

Usage

```
groot(object, method="BR", penalty=2, cutoff=0.3, nPermutations=1000,
pAdjustMethod="bonferroni", verbose=TRUE)
```

Arguments

object	this argument is an object of class 'OGR' (OGR-class).
method	a single character value specifying the rooting algorithm. Options: "BR" and "KS" (see details).
penalty	a single numeric value specifying the penalty used in the rooting algorithm (see details).
cutoff	a single numeric value in [0,1] specifying the cutoff used in the BR statistics (see details).
nPermutations	a single integer value specifying the number of permutations used to compute a null distribution for the inferred roots in the species tree.

- `pAdjustMethod` a single character value specifying the p-value adjustment method to be used (see 'p.adjust' for details).
- `verbose` a single logical value specifying to display detailed messages (when `verbose=TRUE`) or not (when `verbose=FALSE`).

Details

This function addresses the problem of finding the evolutionary root of a feature in an evolutionary tree. The method infers the probability that such feature had been present in each last common ancestor (LCA) of a given existent species in the tree. Regarding genes, this is a delicate task, especially due to mechanisms like horizontal gene transfer, gene deletion, de novo gene formation, among others, which add noise to the vertical heritage pattern. The `groot` function estimates this information by accessing the presence and absence of orthologs in the tree branches and summarizes the available evidences supporting the ancestor ortholog in the LCA. A probability distribution is derived from this process, which sets the searching space used in the rooting algorithm described next. The penalty argument allows weighing gene gain and loss during speciation; `penalty=1` indicates same probability; `penalty > 1` indicates higher probability of gene loss while `penalty < 1` indicates higher probability of gene gain. Penalty value should be greater than zero (default `penalty=2`). The rooting algorithm is set by the argument `method`. The current version introduces the Bridge (BR) algorithm, which search the previously computed probability distribution to find the most reliable evolutionary root. The algorithm search for the optimum point that splits the probability distribution in two components: one enriched with the queried feature and another with low evidence in favor of the feature's presence. The `cutoff` sets the tolerance for the discrimination (default `cutoff=0.3`). The `groot` function also computes the inconsistency score `Dscore` in order to access the significance of the observed root, which is tested against a null distribution derived by permutation analysis.

Value

An processed object of class 'OGR', including results from the rooting algorithm.

Author(s)

Rodrigo Dalmolin, Mauro Castro

References

Dalmolin RJ and Castro, MA. Geneplast: Evolutionary rooting using orthologous groups distribution. Journal Paper (in preparation), 2016.

See Also

[OGR-class](#)

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGR' for H. sapiens
ogr <- groot.preprocess(cogdata=cogdata, phyloTree=phyloTree, spid="9606", cogids=cogids)

## run the groot function
## this example uses the orthologous groups listed in the gpdata object
```

```

ogr <- groot(ogr, nPermutations=100)
res <- groot.get(ogr, what="results")

## Not run:
# Option: parallel version with SNOW package!
library(snow)
options(cluster=makeCluster(2, "SOCK"))
ogr <- groot(ogr, nPermutations=100)
stopCluster(getOption("cluster"))

## End(Not run)

```

groot.get

Get information from individual slots in an OGR object.

Description

Get information from individual slots in an OGR object and any available results from a previous analysis.

Usage

```
groot.get(object, what="status")
```

Arguments

object	an object of class 'OGR' OGR-class .
what	a single character value specifying which information should be retrieved from the slots. Options: "cogids", "spbranches", "orthoroot", "results" and "status".

Value

slot content from an object of class 'OGR' [OGR-class](#).

Author(s)

Rodrigo Dalmolin, Mauro Castro

Examples

```

#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGR' for H. sapiens
ogr <- groot.preprocess(cogdata=cogdata, phyloTree=phyloTree, spid="9606", cogids=cogids)

## run the groot function
## this example uses the orthologous groups listed in the gpdata object
ogr <- groot(ogr, nPermutations=100)
res <- groot.get(ogr, what="results")

```

groot.plot	<i>Plot the inferred evolutionary root of a given OG or the map of LCAs of a given species.</i>
------------	---

Description

Plot the inferred evolutionary root of a given OG onto the species tree or the map of LCAs of a given species.

Usage

```
groot.plot(ogr, whichOG, fname="gproot", width=4.5, height=6.5, cex.lab=0.3,
  cex.nodes=0.6, adj.tips=c(1, 0.5), lab.offset=1.5, col.tips=c("green2", "grey"),
  col.edges=c("black", "grey"), col.root="red", plot.sspnames=TRUE,
  plot.subtree=FALSE, plot.lcas=FALSE)
```

Arguments

ogr	this argument is an object of class 'OGR' evaluated by the <code>groot</code> groot method.
whichOG	a single character value indicating the OG to be plotted.
fname	a character string naming a file.
width	a single numeric value specifying the width of the graphics region in inches.
height	a single numeric value specifying the height of the graphics region in inches.
cex.lab	numeric character expansion factor for tip labels.
cex.nodes	numeric expansion factor for node symbols.
adj.tips	two numeric values specifying the adjustment of the labels.
lab.offset	a single numeric value specifying the offset of the labels.
col.tips	a character vector of length=2 specifying the colors of the tips.
col.edges	a character vector of length=2 specifying the colors of the edges.
col.root	a character value specifying the color of the inferred root.
plot.sspnames	a single logical value specifying whether ssp names should be used to generate the plot.
plot.subtree	a single logical value specifying whether a sub-species tree should be used to generate the plot.
plot.lcas	a single logical value specifying whether a species tree should be generated mapping the positions of all possible roots.

Value

a pdf file.

Author(s)

Rodrigo Dalmolin, Mauro Castro

References

Dalmolin RJ and Castro, MA. Geneplast: Evolutionary rooting using orthologous groups distribution. Journal Paper (in preparation), 2016.

See Also

[groot](#)

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGR' for H. sapiens
ogr <- groot.preprocess(cogdata=cogdata, phyloTree=phyloTree, spid="9606", cogids=cogids)

## run the groot function
ogr <- groot(ogr, nPermutations=100)

## this example plots NOG40170 in the phyloTree
groot.plot(ogr,whichOG="NOG40170")
```

`groot.preprocess` *Evolutionary rooting inference.*

Description

Constructor for the 'OGR-class' object.

Usage

```
groot.preprocess(cogdata, phyloTree, spid, cogids=NULL, verbose=TRUE)
```

Arguments

<code>cogdata</code>	a data frame with COG's data.
<code>phyloTree</code>	an object of class "phylo".
<code>spid</code>	a single character or integer value specifying the reference species to be used in the rooting algorithm. This species should be listed in the 'phyloTree'.
<code>cogids</code>	an optional data frame with COG's annotation. Alternatively, it can be a character vector with COG IDs.
<code>verbose</code>	a single logical value specifying to display detailed messages (when <code>verbose=TRUE</code>) or not (when <code>verbose=FALSE</code>).

Details

This function creates an [OGR-class](#) object and checks the consistency of the input data for the evolutionary root pipeline. Internally, the function access the presence and absence of orthologs for each species in a given OG and computes the orthoct data.frame.

Value

A preprocessed object of class 'OGR'.

Author(s)

Rodrigo Dalmolin, Mauro Castro

References

Dalmolin RJ and Castro, MA. Geneplast: Evolutionary rooting using orthologous groups distribution. Journal Paper (in preparation), 2016.

See Also

[OGR-class](#)

Examples

```
#load datasets used for demonstration
data(gpdata.gs)

#create and object of class 'OGR' for H. sapiens
ogr <- groot.preprocess(cogdata=cogdata, phyloTree=phyloTree, spid="9606", cogids=cogids)
```

OGP-class

Class "OGP": an S4 class for genetic plasticity analysis.

Description

This S4 class includes methods to access the genetic plasticity of orthologous groups.

Objects from the Class

Objects can be created by calls to the "gplast.preprocess" constructor.

Slots

sspids: Object of class "data.frame", a data frame with species annotation.
cogids: Object of class "data.frame", a data frame with COG's data.
orthodist: Object of class "matrix", a matrix with COG's information (see return values in the OGP methods).
abundance: Object of class "numeric", a numeric vector with results from the gplast function (see return values in the OGP methods).
diversity: Object of class "numeric", a numeric vector with results from the gplast function (see return values in the OGP methods).
plasticity: Object of class "numeric", a numeric vector with results from the gplast function (see return values in the OGP methods).
status: Object of class "character", a character value specifying the status of the OGP object based on the available methods.

Methods

gplast signature(object = "OGP"): see [gplast](#)

gplast.get signature(object = "OGP"): see [gplast.get](#)

Author(s)

Rodrigo Dalmolin, Mauro Castro

See Also

[gplast.preprocess](#)

OGR-class

Class "OGR": an S4 class for rooting analysis.

Description

This S4 class includes methods to do inferential analysis of evolutionary roots in a given species tree.

Objects from the Class

Objects can be created by calls to the "groot.preprocess" constructor.

Slots

cogids: Object of class "data.frame", a data frame with COG's data.

tree: Object of class "phylo", a given species tree.

sbranches: Object of class "data.frame", a data frame listing branches of a given species tree.

orthoroot: Object of class "data.frame", a data.frame with results from the 'groot' function (see return values in the OGR methods).

orthoct: Object of class "data.frame", a data.frame with results from the 'groot.preprocess' function (see return values in the OGR methods).

status: Object of class "character", a character value specifying the status of the OGR object based on the available methods.

Methods

groot signature(object = "OGR"): see [groot](#)

groot.get signature(object = "OGR"): see [groot.get](#)

Author(s)

Rodrigo Dalmolin, Mauro Castro

See Also

[groot.preprocess](#)

Index

*Topic **classes**

OGP-class, [11](#)

OGR-class, [12](#)

*Topic **dataset**

gpdata, [2](#)

*Topic **methods**

gplast, [3](#)

gplast.get, [4](#)

gplast.preprocess, [5](#)

groot, [6](#)

groot.get, [8](#)

groot.plot, [9](#)

groot.preprocess, [10](#)

*Topic **package**

geneplast-package, [2](#)

geneplast (geneplast-package), [2](#)

geneplast-package, [2](#)

gpdata, [2](#)

gplast, [3](#), [12](#)

gplast, OGP-method (OGP-class), [11](#)

gplast.get, [4](#), [12](#)

gplast.get, OGP-method (OGP-class), [11](#)

gplast.preprocess, [5](#), [12](#)

groot, [6](#), [9](#), [10](#), [12](#)

groot, OGR-method (OGR-class), [12](#)

groot.get, [8](#), [12](#)

groot.get, OGR-method (OGR-class), [12](#)

groot.plot, [9](#)

groot.preprocess, [10](#), [12](#)

OGP-class, [11](#)

OGR-class, [12](#)