Package: RevEcoR (via r-universe)

August 28, 2024

Type Package

Title Reverse Ecology Analysis on Microbiome

Version 0.99.3

Date 2016-3-28

Description An implementation of the reverse ecology framework. Reverse ecology refers to the use of genomics to study ecology with no a priori assumptions about the organism(s) under consideration, linking organisms to their environment. It allows researchers to reconstruct the metabolic networks and study the ecology of poorly characterized microbial species from their genomic information, and has substantial potentials for microbial community ecological analysis.

Depends R (>= 2.14)

Imports Matrix, igraph, XML, stringr, magrittr, gtools, plyr, purrr, methods

Suggests knitr

VignetteBuilder knitr

License GPL (>= 2)

LazyData yes

RoxygenNote 5.0.1

Repository https://yiluheihei.r-universe.dev

RemoteUrl https://github.com/yiluheihei/revecor

RemoteRef HEAD

RemoteSha cedbfeb1b4fa3fcab04b53e5b37f96b8b2a46175

Contents

anno.species 2
calculateCooperationIndex 3
compose 4
confidencescore-methods 5

getGsMN-methods 5
getOrgMetabolicData 6
getSeedSets 7
gut_microbiome 8
Interactions 8
kegg_buc 10
kegg_ptr 10
KosarajuSCC 11
len-methods 12
nonseed-methods 12
reconstructGsMN 13
RefDbcache 14
RevEcoR 15
seedset-class 15
seedSize-methods 16
show,seedset-method 17

Index **18**

`anno.species` *Annotation profiles of seven well-studied oral species*

Description

A dataset containing the the KEGG orthology annotation profiles of seven oral species which was downloaded from the Integrated Microbial Genomes (IMG).

Format

A list with seven elements and each elements represents the annotation profile of the species

Details

This datasets constains the KEGG orthology annotation information of seven oral species whose interactions were carefully and well characterized. The human oral microbiota is relatively #' well described. The name of these seven species is: *Aggregatibacter*, *actinomycetemcomitans D7S-1*, *Fusobacterium nucleatum polymorphum ATCC 10953*, *Porphyromonas gingivalis ATCC 33277*, *Streptococcus gordonii str. Challis substr. CH1*, *Streptococcus oralis SK23, ATCC 35037*, *Veillonella atypica ACS-134-V-Col7a*. For more annotation information on these species, see `img.jgi.doe.gov/`.

- Aa, Aggregatibacter actinomycetemcomitans D7S-1
- Ao, Actinomyces oris K20
- Fn, Fusobacterium nucleatum polymorphum ATCC 10953
- Pg, Porphyromonas gingivalis ATCC 33277
- Sg, Streptococcus gordonii str. Challis substr. CH1
- So, Streptococcus oralis SK23, ATCC 35037
- Va, Veillonella atypica ACS-134-V-Col7a

Source

img.jgi.doe.gov/

Examples

```
data(anno.species)
```

`calculateCooperationIndex` *Calculating the metabolic competition and complementarity index*

Description

Calculating the metabolic competition complementarity index among all metabolic networks

Usage

```
calculateCooperationIndex(g, ..., threshold = 0, p = FALSE, nperm = 1000)
```

Arguments

`g`	igraph that represents a metabolic network, see `reconstructGsMN`
`...`	a list of metabolic networks or a network append to g
`threshold`	threshold, the cutoff of confidence score to be serve as a seed set, default is 0.2
`p`	a logical value which determins whether the calculated index is statistical or biological significant. default is FALSE
`nperm`	the number of permuations of metabolic network node labes, which is used for p value calculation, default is 1000.

Details

Metabolic competition index is defined as the fraction of compounds in a species seed set of metabolic network that are alse included in its partner; However, metabolic complementarity index is the fraction of compounds in one species seed set of metabolic network appearing in the metabolic network but not in the seed set of its partner; The biosynthetic support score represents the extent to which the metabolic requirements of a potential parasitic organism can be supported by the biosynthetic capacity of a potential host. It is measured by calculating the fraction of the source components of a, in which at least one of the compounds can be found in the network of b. However, seed compounds are associated with a confidence score (1/size of SCC), so this fraction is calculated as a mormalized weighted sum.

The ith row and jth col elements of the returnd matrix represents the metabolic competition index or complementarity index of the ith network on the jth metabolic network.

Value

a cooperation index matrix whose nrow and ncol is equal to the number of species to be compared, for more see details.

See Also

complementarityIndex, competitionIndex

Examples

```
## Not run:
## metabolic network reconstruction and seed set identity of sample data anno.species
net <- lapply(anno.species,reconstructGsMN)
interactions <- calculateCooperationIndex(net)

## End(Not run)
```

---

compose *Compose multiple functions*

Description

In infix and prefix forms.

Usage

```
compose(...)

f %.% g
```

Arguments

...	n functions to apply in order from right to left
f, g	two functions to compose for the infix form

Details

This function was from hadley wickham's package pryr, for more details see https://github.com/hadley/pryr

Author(s)

Hadley wickham

Examples

```
not_null <- `!` %.% is.null
not_null(4)
not_null(NULL)

add1 <- function(x) x + 1
compose(add1,add1)(8)
```

confidencescore-methods *Conficence score*

Description

Caculate confidence score of seed set

Usage

```
confidencescore(object)

## S4 method for signature 'seedset'
confidencescore(object)
```

Arguments

object	seedset class

Value

a list

See Also

seedset-class

Examples

```
## Not run:
confidencescore(seed.set)

## End(Not run)
```

getGsMN-methods *The genome scale metabolic network*

Description

T he genome scale metabolic network (GsMN) whose seed set is caculated.

Usage

```
getGsMN(object)

## S4 method for signature 'seedset'
getGsMN(object)
```

Arguments

object	seedset class

Value

a igraph

See Also

seedset-class

Examples

```
## Not run:
getGsMN(seed.set)

## End(Not run)
```

---

getOrgMetabolicData *Get organism metabolic data from KEGG database*

Description

This function helps us to obtain the specific-organism pathway map, prasing this maps to get metabolic data contains reaction, substrate and product.

Usage

```
getOrgMetabolicData(org)
```

Arguments

org	characters, the KEGG organism code, e.g. "buc".

Details

Function getOrgMetabolicData helps us to download metabolic data of a given organism from KEGG database with REST-style KEGG API. Enzyme reactions take place in this organism (org) and its metabolites (substrates and products), that will be used for organism-specific genome scale metabolic network reconstruction, can be obtained with this function.

Value

a three length df, consists of enzyme reaction names, substrates and products

See Also

getSeedSets

Examples

```
## Not run:
metabolic.data <- getOrgMetabolicData("buc")

## End(Not run)
```

---

getSeedSets *Identify seed compounds of each organism*

Description

Detect a given metabolic network and idendity the seed compounds of each organism

Usage

```
getSeedSets(g, threshold = 0)
```

Arguments

g	an igraph object which represents a given organism-specific metaboliic network
threshold	numeric constant ranges from 0 to 1, default is 0.

Details

All the compound in the same source SCC all equally to be included in the seed set, each of these compounds was assigned a confidence level, C=1/(size of souce SCC), denoting the compounds probability of being a seed. This threshold was used to determin whether a compound should be a seed.

Value

a two-length list which consists of network and the seed set compounds of the given organism-specific metabolic network, ...

See Also

KosarajuSCC,seedset-class

Examples

```
## Not run:
## get metabolic annotated data of a specific species
metabolic.data <- getOrgMetabolicData("buc")
## metabolic network reconstruction
net <- reconstructGsMN(metabolic.data)

## End(Not run)
```

gut_microbiome *Annotation profiles of 116 gut prevalent species*

Description

A dataset containing the the KEGG orthology annotation profiles of 116 gut prevlent species which was downloaded from the Integrated Microbial Genomes (IMG).

Format

A list with 116 elements and each elements represents the annotation profile of the species

Details

This dataset focused on a list of 116 prevalent gut species, whose genome sequence is available in IMG database and sequence coverage is more than 1 annotation profiles of this 116 species was collected from IMG database.

With a in-house R script, we obtained genomic data for all organisms from the Department of Integrated Microbial Genomes project (IMG). For each species, the list of genes mapped to the Kyoto Encyclopedia of Genes and Genomes orthologous groups (KEGG KOs) was downloaded. For more annotation information on these species, see img.jgi.doe.gov/.

Source

img.jgi.doe.gov/

Interactions *Calculating the species interactions*

Description

Calculating the metabolic complementarity index and complementarity index of based on species metabolic network.

Usage

```
complementarityIndex(g1, g2, seed.set1, seed.set2, threshold = 0, p = FALSE,
  nperm = 1000)

competitionIndex(g1, g2, seed.set1, seed.set2, threshold = 0, p = FALSE,
  nperm = 1000)
```

Arguments

g1	igraph object, a species-specific metabolic network.
g2	igraph object, a species-specific metabolic network, the complementary network of g1
seed.set1	seeds slot of a seed-set object, seeds of the metabolic network g1, more details see seedset-class.
seed.set2	seeds slot of a seed-set object, seeds of the metabolic network g2, more details see seedset-class.
threshold	the cutoff of confidence score to be serve as a seed set, default is 0.
p	a logical value which determins whether the calculated index is statistical or biological significant. default is FALSE.
nperm	the number of permuations of metabolic network node labes, which is used for complementarity index's P value calculating, default is 1000.

Details

Metabolic competition index is defined as the fraction of compounds in a species seed set of metabolic network that are also included in its partner; However, metabolic complementarity index is the fraction of compounds in one species seed set of metabolic network appearing in the metabolic network but not in the seed set of its partner. However, seed compounds are associated with a confidence score (1/size of SCC), so this fraction is calculated as a normalized weighted sum.

Based on the metabolic network and seed sets of species, this functions help us to predict the species interactions of species1 on the presence of species2.

Value

a two length list: complementarity index or competition index: range from 0 to 1, p value of complementarity index. Or a single value of complementarity or competition index while p is FALSE.

See Also

getSeedSets, calculateCooperationIndex

Examples

```
## Not run:
## metabolic network reconstruction and seed set identity of sample data anno.species
net <- lapply(anno.species,reconstructGsMN)
seed.sets <- lapply(net, getSeedSets)
seed.sets <- lapply(seed.sets, function(x)x@seeds)

## calculate the complementarity index of the first species
complementarity.index <- complementarityIndex(net[[1]],net[[2]],
 seed.sets[[1]], seed.sets[[2]])
competition.index <- competitionIndex(net[[1]],net[[2]],
 seed.sets[[1]], seed.sets[[2]])

## End(Not run)
```

kegg_buc *Metabolic profiles of KEGG organism Buchnera aphidicola APS (Acyrthosiphon pisum) (KEGG organism code: buc)*

Description

kegg organism buc metabolic information, which consists of enzymatic reactions and metabolites.

Format

A data frame with 418 observations on three variables.

[,1] .attrs.name, character (reaction: R)

[,2] substrate.name, list (substrates: cpd)

[,3] product.name, list (products: cpd)

Details

buc metatolic information:

- .attrs.name: Enzymatic reactions that organism involved
- substrate.name: Substrates of the corresponding reaction.
- product.name: Products of the corresponding reaction.

kegg_ptr *Metabolic profiles of KEGG organism Pan troglodytes (chimpanzee) (KEGG organism code: ptr)*

Description

kegg organism ptr metabolic information, which consists of enzymatic reactions and metabolites.

Format

A data frame with 1858 observations on three variables.

[,1] .attrs.name, character (reaction: R)

[,2] substrate.name, list (substrates: cpd)

[,3] product.name, list (products: cpd)

Details

ptr metatolic information:

- .attrs.name: Enzymatic reactions that organism involved
- substrate.name: Substrates of the corresponding reaction.
- product.name: Products of the corresponding reaction.

KosarajuSCC *Caculating the strong connected components (SCC) of a network*

Description

This function utilizes Kosaraju's algorithm to caculate the strong connetected components descomposition of a given network

Usage

```
KosarajuSCC(g)
```

Arguments

g	a igraph object to be caculated

Value

a list which length is equal to the number of SCCs, each element represents a Scc

References

AV Aho, JE Hopcroft, JD Ullman: The design and analysis of computer algorithms, 1974

See Also

getSeedSets

Examples

```
## Not run:
metabolic.data <- getOrgMetabolicData("buc")
## metabolic network reconstruction
net <- reconstructGsMN(metabolic.data)
scc <- KosarajuSCC(net)

## End(Not run)
```

len-methods *the length of the seed set*

Description

Caculate the number of the seed source components.

Usage

```
len(object)

## S4 method for signature 'seedset'
len(object)
```

Arguments

object	seed-set class

Value

an interger

See Also

seedset-class

Examples

```
## Not run:
len(seed.set)

## End(Not run)
```

nonseed-methods *Non seed of the network*

Description

Non seed of the network.

Usage

```
nonseed(object)

## S4 method for signature 'seedset'
nonseed(object)
```

Arguments

object	seedset class

Value

a vector

See Also

seedset-class

Examples

```
## Not run:
nonseed(seed.set)

## End(Not run)
```

---

reconstructGsMN *Reconstuction of the specific-organism genome-scale metabolic network*

Description

Reconstruction of genome-scale metabolic network (GsMN) whose nodes represents compounds and whose edges represents reactions.

Usage

```
reconstructGsMN(metabolic.data, RefData = RefDbcache, threshold = 10,
  is.gaint = TRUE)
```

Arguments

metabolic.data	df or a character vector. More details see function getOrgMetabolicData and details
RefData	The reference metabolic data. It does not need reference data While organism metabolic data was collected from KEGG database, and RefData is set to NULL. Otherwise, RefDbCache, an internal dataset in this package, was taken as the Reference metabolic data for Genome scale metabolic reconstruction.
threshold	numeric, Nodes belonging to components with fewer than the value of threshold nodes will be ignored. This is a good option for networks that contain many small and trivial components. Default is 10.
is.gaint	logical, Ignore all nodes except those in the giant component: selecting the only main largest component (connected set of nodes) of the network. All smaller components will be ignored. This is a good option for networks with a dominant component. Default is TRUE.

Details

The input of this function can be of two forms. If organims is collected in KEGG database, it can be obtained with `getOrgMetabolicData` which is a data frame. Otherwise, `metabolic.data` could be a character vecotr which contains the KEGG Orthology annotated information on this organism, e.g. we can download this KO annotation profile in the https://img.jgi.doe.gov website for species detected in a human microbime which not contained in KEGG organism database. Several functions, such as `link{read.table}` and `read.delim` could help us to read KO annotation profile.

Value

igraph object

See Also

`getOrgMetabolicData`

Examples

```
## not run (organism in KEGG)
## metabolic.data <- getOrgMetabolicData("buc")
## g <- reconstructGsMN(metabolic.data)

## species detected in a human microbiome
annodir <- system.file("extdata","koanno.tab",package = "RevEcoR")
metabolic.data <- read.delim2(file=annodir,stringsAsFactors=FALSE)
##load the reference metabolic data
data(RefDbcache)
g2 <- reconstructGsMN(metabolic.data, RefData = RefDbcache)
```

RefDbcache *Reference data for global metabolic construction The reference metabolic pathway data contains KOs, substrates and products, as well as a constructed reference global network, which used for metabolic network reconstruction*

Description

Reference data for global metabolic construction

The reference metabolic pathway data contains KOs, substrates and products, as well as a constructed reference global network, which used for metabolic network reconstruction

Format

The format is: List of 7 KO, substrate, product, user, date, version, reference network

Details

Information this dataset is involved:

- KO, all KEGG orthlogy enties in KEGG metabolic pathways.
- substrate, substrate of enzymatic reactions in all KEGG metabolic pathways.
- product, product of enzymatic reactions in all KEGG metabolic pathways.
- user who download this data.
- date, the date this data is downloaded.
- version, R version used to obtained it.
- network, the global network which is reconstructed based on all the metabolites.

References

https://www.bioconductor.org/packages/release/bioc/html/mmnet.html

---

`RevEcoR` *The RevEcoR package*

Description

This package implementation the applications of reverse ecology. Reverse ecology refers to the use of genomics to study ecology with no a priori assumptions about the organism(s) under consideration, linking the organism and their environment. Prediction the cooperation among species and hosts.

`seedset-class` `seedset-class`

Description

Object representing the seed sets of a given metabolic network

Slots

`GsMN`, a igraph network

`seeds`, a character list represents seeds of a given metabolic network which is composed of the KEGG compound index.

method

- getGsMN, `signature(object = "seedset")`: get the genome scale metabolic network whose seed set is caculated
- len, `signature(object = "seedset")`: return the number of source SCC
- seedSize, `signature(object = "seedset")`: returns the sizes of each source SCCs
- nonseed, `signature(object = "seedset")`: the non seeds of the GsMN
- show, `signature(object = "seedset")`: show the short summary of a seedset class
- confidencescore, `signature(object = "seedset")`: confidence score of the seed set

See Also

getSeedSets,getGsMN,len, nonseed,seedSize,confidencescore

Examples

```
## Not run:
#' ## generate a metabolic network in igraph class and a seed set of this graph
annodir <- system.file("extdata","koanno.tab",package = "RevEcoR")
metabolic.data <- read.delim2(file=annodir,stringsAsFactors=FALSE)
g <- reconstructGsMN(metabolic.data)
seeds <- getSeedSets(g)@seeds
seed.set <- new("seedset",GsMN = g, seeds = seeds)

## End(Not run)
```

---

seedSize-methods *Size of the each seed source component*

Description

Caculate the size of each seed source component.

Usage

```
seedSize(object)

## S4 method for signature 'seedset'
seedSize(object)
```

Arguments

object	seedset class

Value

a vector represents size of each source seed componet of network

See Also

seedset-class

Examples

```
## Not run:
seedSize(seed.set)

## End(Not run)
```

---

show,seedset-method *The show generic function*

Description

Show a short summary of seedset object

Usage

```
## S4 method for signature 'seedset'
show(object)
```

Arguments

object	seed-set class

See Also

seedset-class

Examples

```
## Not run:
show(seed.set)

## End(Not run)
```

Index

%.% (compose), 4

anno.species, 2

calculateCooperationIndex, 3, *9*
competitionIndex, *4*
competitionIndex (Interactions), 8
complementarityIndex, *4*
complementarityIndex (Interactions), 8
compose, 4
confidencescore, *16*
confidencescore
        (confidencescore-methods), 5
confidencescore,seedset-method
        (confidencescore-methods), 5
confidencescore-methods, 5

getGsMN, *16*
getGsMN (getGsMN-methods), 5
getGsMN,seedset-method
        (getGsMN-methods), 5
getGsMN-methods, 5
getOrgMetabolicData, 6, *14*
getSeedSets, *6*, 7, *9*, *11*, *16*
gut_microbiome, 8

Interactions, 8

kegg_buc, 10
kegg_ptr, 10
KosarajuSCC, *7*, 11

len, *16*
len (len-methods), 12
len,seedset-method (len-methods), 12
len-methods, 12

nonseed, *16*
nonseed (nonseed-methods), 12
nonseed,seedset-method
        (nonseed-methods), 12
nonseed-methods, 12

read.delim, *14*
reconstructGsMN, *3*, 13
RefDbcache, 14
RevEcoR, 15
RevEcoR-package (RevEcoR), 15

seedset-class, 15
seedSize, *16*
seedSize (seedSize-methods), 16
seedSize,seedset-method
        (seedSize-methods), 16
seedSize-methods, 16
show (show,seedset-method), 17
show,seedset-method, 17
show-methods (show,seedset-method), 17