

Package ‘helminthR’

August 16, 2021

Type Package

Title Access London Natural History Museum Host-Helminth Record Database

Version 1.0.9

Author Tad Dallas

Imports xml2, rvest, httr, magrittr, plyr, utils, taxize

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

Maintainer Tad Dallas <tad.a.dallas@gmail.com>

URL <https://docs.ropensci.org/helminthR/>,
<https://github.com/rOpenSci/helminthR/>

BugReports <https://github.com/rOpenSci/helminthR/issues/>

Description Access to large host-parasite data is often hampered by the availability of data and difficulty in obtaining it in a programmatic way to encourage analyses. 'helminthR' provides a programmatic interface to the London Natural History Museum's host-parasite database, one of the largest host-parasite databases existing currently <<https://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/>>. The package allows the user to query by host species, parasite species, and geographic location.

License GPL-3

LazyData yes

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-08-16 21:50:02 UTC

R topics documented:

helminthR-package	2
cleanData	2
findHost	3
findLocation	5
findParasite	7
locations	9

Index	10
--------------	-----------

helminthR-package	<i>Access London Natural History Museum host-helminth record database</i>
-------------------	---

Description

'helminthR': A programmatic interface to the London Natural History Museum's host-parasite database.

The package currently allows you to query by host species, parasite species, and geographic location. No information is provided on parasite prevalence or intensity.

Author(s)

Tad Dallas <tad.a.dallas@gmail.com>

References

Gibson, D. I., Bray, R. A., & Harris, E. A. (Compilers) (2005). Host-Parasite Database of the Natural History Museum, London. <<http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/>>

cleanData	<i>Clean helminth parasite occurrence data</i>
-----------	--

Description

Given a host-parasite edgelist, this function can validate species names, provide further taxonomic information (thanks to taxize), and remove records only to genus level.

Usage

```
cleanData(edge, speciesOnly = FALSE, validateHosts = FALSE)
```

Arguments

edge	Host-parasite edgelist obtained from findLocation , findHost , or findParasite
speciesOnly	boolean flag to remove host and parasite species where data are only available at genus level (default = FALSE)
validateHosts	boolean flag to check host species names against Catalogue of Life information and output taxonomic information (default = FALSE)

Details

Use `data(locations)` for a list of possible locations.

Value

`cleanEdge` Host-parasite edgelist, but cleaned

Author(s)

Tad Dallas

findHost	<i>Find parasite occurrence data for given host.</i>
----------	--

Description

Given a host genus, species, and/or location, returns a list of parasite occurrences on that host or for that location. Use `data(locations)` for a list of possible locations.

Usage

```
findHost(  
  genus = NULL,  
  species = NULL,  
  location = NULL,  
  citation = FALSE,  
  hostState = NULL,  
  speciesOnly = FALSE,  
  validateHosts = FALSE,  
  parGroup = NULL,  
  removeDuplicates = FALSE  
)
```

Arguments

genus	Host genus
species	Host species
location	Geographic location.
citation	Boolean. Should the output include the citation link and the number of supporting citations? default is FALSE
hostState	number corresponding to one of six different host states. The default value is NULL and includes all host states
speciesOnly	boolean flag to remove host and parasite species where data are only available at genus level (default = FALSE)
validateHosts	boolean flag to check host species names against Catalogue of Life information and output taxonomic information (default = FALSE)
parGroup	name of parasite group to query (default queries all groups)
removeDuplicates	(boolean) should duplicate host-parasite combinations be removed? (default is FALSE)

Details

hostState can take values 1-6 corresponding to if the recorded host was found

- (1) "In the wild"
- (2) "Zoo captivity"
- (3) "Domesticated"
- (4) "Experimental"
- (5) "Commercial source"
- (6) "Accidental infestation"

A value of NULL should be entered if you would like to include all hostStates.

parGroup can be specified as "Acanthocephalans", "Cestodes", "Monogean", "Nematodes", "Trematodes", or "Turbs" (Turbellarians etc.). The default is to query all helminth parasite taxa.

Value

Three (or five) column data.frame containing host species, parasite species (shortened name and full name), and citation link and number of citations (if 'citation'=TRUE), with each row corresponding to an occurrence of a parasite species on a host species.

Author(s)

Tad Dallas

References

Gibson, D. I., Bray, R. A., & Harris, E. A. (Compilers) (2005). Host-Parasite Database of the Natural History Museum, London. <<http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/>>

See Also[findParasite](#)**Examples**

```
gorillaParasites <- helminthR::findHost("Gorilla", "gorilla")

# An example of how to query multiple hosts when you have a
# vector of host species names

hosts <- c("Gorilla gorilla", "Peromyscus leucopus")
plyr::ldply(hosts, function(x)
  {helminthR::findHost(unlist(strsplit(x, " ")[1], unlist(strsplit(x, " ")[2]))})
```

findLocation

*Find host-parasite interactions for a given location***Description**

Given a location (available from `data{locations}`) this function returns all host-parasite associations in that location.

Usage

```
findLocation(
  location = NULL,
  group = NULL,
  citation = FALSE,
  hostState = NULL,
  speciesOnly = FALSE,
  validateHosts = FALSE,
  removeDuplicates = FALSE
)
```

Arguments

location	Location of host-parasite interaction.
group	Parasite group - Cestodes, Acanthocephalans, Monogeneans, Nematodes, Trematodes, or Turbellarian etc. (Turb)
citation	Boolean. Should the output include the citation link and the number of supporting citations? default is FALSE
hostState	number corresponding to one of six different host states. The default value is NULL and includes all host states.
speciesOnly	boolean flag to remove host and parasite species where data are only available at genus level (default = FALSE)

`validateHosts` boolean flag to check host species names against Catalogue of Life information and output taxonomic information (default = FALSE)

`removeDuplicates` (boolean) should duplicate host-parasite combinations be removed? (default is FALSE)

Details

`hostState` can take values 1-6 corresponding to if the recorded host was found

- (1) "In the wild"
- (2) "Zoo captivity"
- (3) "Domesticated"
- (4) "Experimental"
- (5) "Commercial source"
- (6) "Accidental infestation"

Value

Three (or five) column data.frame containing host species, parasite species (shortened name and full name), and citation link and number of citations (if `citation = TRUE`), with each row corresponding to an occurrence of a parasite species on a host species.

Author(s)

Tad Dallas

References

Gibson, D. I., Bray, R. A., & Harris, E. A. (Compilers) (2005). Host-Parasite Database of the Natural History Museum, London. <<http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/>>

See Also

[findHost](#)

Examples

```
FrenchHostPars <- helminthR::findLocation(location="France")
```

findParasite	<i>Find host-parasite interactions for a given parasite species.</i>
--------------	--

Description

Given a host genus and/or species, this function returns a matrix containing host-parasite interaction data. Search available locations using `data(locations)`.

Usage

```
findParasite(
  genus = NULL,
  species = NULL,
  group = NULL,
  subgroup = NULL,
  location = NULL,
  citation = FALSE,
  hostState = NULL,
  speciesOnly = FALSE,
  validateHosts = FALSE,
  removeDuplicates = FALSE
)
```

Arguments

<code>genus</code>	Parasite genus
<code>species</code>	Parasite species
<code>group</code>	Parasite group - Cestodes, Acanthocephalans, Monogeneans, Nematodes, Trematodes, or Turbellarian etc. (Turb)
<code>subgroup</code>	Parasite subgroup (family names largely)
<code>location</code>	Location of host-parasite interaction.
<code>citation</code>	Boolean. Should the output include the citation link and the number of supporting citations? default is FALSE
<code>hostState</code>	number corresponding to one of six different host states. The default value is NULL includes all host states
<code>speciesOnly</code>	boolean flag to remove host and parasite species where data are only available at genus level (default = FALSE)
<code>validateHosts</code>	boolean flag to check host species names against Catalogue of Life information and output taxonomic information (default = FALSE)
<code>removeDuplicates</code>	(boolean) should duplicate host-parasite combinations be removed? (default is FALSE)

Details

hostState can take values 1-6 corresponding to if the recorded host was found

- (1) "In the wild"
- (2) "Zoo captivity"
- (3) "Domesticated"
- (4) "Experimental"
- (5) "Commercial source"
- (6) "Accidental infestation"

Value

Three (or five) column data.frame containing host species, parasite species (shortened name and full name), and citation link and number of citations (if `citation = TRUE`), with each row corresponding to an occurrence of a parasite species on a host species.

Author(s)

Tad Dallas

References

Gibson, D. I., Bray, R. A., & Harris, E. A. (Compilers) (2005). Host-Parasite Database of the Natural History Museum, London. <<http://www.nhm.ac.uk/research-curation/scientific-resources/taxonomy-systematics/host-parasites/>>

See Also

[findHost](#)

Examples

```
strongHosts <- helminthR::findParasite(genus = "Strongyloides")

# An example of how to query multiple parasite species when
# you have a vector of parasite species names

parasites <- c("Ascaris aculeati", "Oxyuris flagellum")

plyr::ldply(parasites,
  function(x){
    helminthR::findParasite(unlist(strsplit(x, " "))[1],
      unlist(strsplit(x, " "))[2])
  }
)
```

locations

Table of geographic location names, and associated coordinates

Description

Lists geographic locations that can be input to `findHost` or `findParasite` and the corresponding latitude and longitude coordinates of the country's centroid. The georeferencing was performed dynamically using the Google Maps API, but they have since restricted access. The data on locations is now provided in this data file called `locations` – `data(locations)` – and is based on an earlier usage of `ggmap`. The geographic coordinates may not be accurate, and users should check for accuracy (and feel free to file an issue or PR on Github with corrections).

Usage

```
data(locations)
```

Format

Location Name of geographic location

Latitude Latitude of location centroid

Longitude Longitude of location centroid

References

Gibson, D. I., Bray, R. A., & Harris, E. A. (Compilers) (2005). Host-Parasite Database of the Natural History Museum, London.

Index

* **data**

locations, [9](#)

cleanData, [2](#)

findHost, [3](#), [3](#), [6](#), [8](#), [9](#)

findLocation, [3](#), [5](#)

findParasite, [3](#), [5](#), [7](#), [9](#)

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

helminthR-package, [2](#)

locations, [9](#)