Package 'BOLD.R'

May 1, 2021

Type Package

Title A package to interface with the Barcode of Life Database through R

Version 0.4.0

Author Nishan Mudalige

Maintainer Nishan Mudalige <mudalign@uoguelph.ca>

Description

BOLDR is a package which allows users to access data directly from BOLD into R via current APIs maintained by the BOLD system. Users can access private data by logging in to BOLD using BOLDR, or they can access public data without the need to login. All data accessed using BOLDR is stored in R with a consistent internal structure and this allows the user to employ the suite of functions provided by BOLDR along with existing packages to perform statistical analysis on DNA sequences.
BOLDR is a powerful, flexible and convenient platform which provides the user with the capacity to access and analyze large amounts of private and public data on BOLD directly through R.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 4.0.0)

Imports jsonlite, gtools, RCurl, ape, plyr, seqinr, getPass

RoxygenNote 7.1.1

NeedsCompilation no

R topics documented:

1

---

all.primers	*A function gives informaiton on the primers in a data set.*

Description

A function which returns a vector with the names of all of the primers in a dataframe.

Usage

```
all.primers(df)
```

Arguments

df   A dataframe containing data obtained from BOLD.

Details

The dataframe entered into this function can be a dataframe that was created after merging several dataframes with data obtained from BOLD.

Value

A vector of primers in the dataframe.

Author(s)

Nishan Mudalige

Examples

```
example.df <- get.public(container="PRT")
all.primers(example.df)

example.df <- get.public(container=c("PRT","ACAGA"))
all.primers(example.df)
```

BOLD.R	*BOLD.R: A package to interface with the Barcode of Life Database through R*

Description

BOLD.R is a package which allows users to access data directly from BOLD into R via current APIs maintained by the BOLD system. Users can access private data by logging in to BOLD using BOLD.R, or they can access public data without the need to login. All data accessed using BOLD.R is stored in R with a consistent internal structure and this allows the user to employ the suite of functions provided by BOLD.R along with existing packages to perform statistical analysis on DNA sequences.

BOLD.R is a powerful, flexible and convenient platform which provides the user with the capacity to access and analyze large amounts of private and public data on BOLD directly through R.

Author(s)

Nishan Mudalige, Joris D'hondt

---

fasta.out   *A function to write a fasta file*

Description

Write a fasta file using datafram with data obtained from BOLD. The file is written by marker code.

Usage

```
fasta.out(x.df, marker.name, output.name = "output", remove.empty = FALSE, nbchar = 60)
```

Arguments

x.df	A dataframe containing data obtained from BOLD.
marker.name	A DNA marker.
output.name	The name of the output file. The extension ".fasta" is added by default.
remove.empty	Remove rows which do not have an entry for the chosen marker type.
nbchar	The number of characters per line.

Value

none.

Author(s)

Nishan Mudalige

Examples

```
example.df <- get.public(container="PRT")

fasta.out(example.df, "COI-5P_nucraw", output.name = "testfile1")

fasta.out(example.df, "COI-5P_nucraw", output.name = "testfile2", remove.empty=TRUE)
```

---

gen.DNAbin   *A function to generate a DNAbin object*

Description

A function that uses information in a dataframe with data obtained from BOLD to genereate a
DNAbin object. The DNAbin object returned can be used with the 'ape' package.

Usage

```
gen.DNAbin(x.df = NULL, alignment = NULL, labels.headers = c(...))
```

Arguments

x.df	A dataframe with data obtained from BOLD.
alignment	Specify the alignment type.
labels	The column names of the data frame to use to create the label.

Value

a DNAbin object.

Author(s)

Nishan Mudalige

Examples

```
example.df <- get.public(container="PRT")

example.DNAbin <- gen.DNAbin(example.df, alignment="COI-5P_nucraw", labels=c("processid","taxon"))
```

---

get.fasta   *A function to retrieve DNA sequences in fasta format from BOLD*

Description

A function to retrieve DNA sequences in fasta format from BOLD using the public API. The user
can access public data bsed on taxon, id, bin, container, institution, researchers, geographic location
or marker.

Usage

```
get.fasta(taxon = NULL, ids = NULL, bin = NULL, container = NULL, institutions = NULL, researchers = N
```

Arguments

taxon	Returns all records containing matching taxa, defined in a pipe or comma delimited list. (spaces need to be encoded).Taxa includes scientific names at phylum, class, order, family, subfamily, genus, and species levels.
ids	Returns all records containing matching IDs, defined in a pipe or comma delimited list. (spaces need to be encoded). IDs include Sample IDs, Process IDs, Museum IDs and Field IDs.
bin	Returns all records contained in matching BINs, defined in a pipe or comma delimited list. A BIN is defined by a Barcode Index Number URI.
container	Returns all records contained in matching projects or datasets, defined in a pipe or comma delimited list. Containers include project codes and dataset codes.
institutions	Returns all records stored in matching institutions, defined in a pipe or comma delimited list. Institutions are the Specimen Storing Site (spaces need to be encoded).
researchers	Returns all records containing matching researcher names, defined in a pipe or comma delimited list. Researchers include collectors and specimen identifiers (spaces need to be encoded).
geo	Returns all records collected in matching geographic sites, defined in a pipe or comma delimited list. Geographic sites includes countries and province/states (spaces need to be encoded).
marker	Returns all specimen records containing matching marker codes defined in a pipe or comma delimited list. All markers for a specimen matching the search string will be returned. ie. A record with COI-5P and ITS will return sequence data for both markers even if only COI-5P was specified.

Value

A dataframe which contains the name and the corresponding sequences that were retrieved using the recordset codes entered.

Author(s)

Nishan Mudalige

Examples

```
## Returns all records matching a project id
## example.df <- get.fasta(container="SSBAA")
## example.df


## Returns all records matching multiple project ids
## example.df <- get.fasta(container=c("SSBAA","SSBAB"))
## example.df


## Returns all records matching multiple project ids
## example.df <- get.fasta(container="SSBAA,SSBAB")
## example.df


## Returns all records matching multiple project ids
```

```
## example.df <- get.fasta(container="SSBAA|SSBAB")
## example.df


## Returns all records matching the taxon Bos taurus.
## example.df <- get.fasta(taxon="Bos taurus")
## example.df


## Returns all records matching the taxa Aves or Reptilia
## example.df <- get.fasta(taxon="Aves|Reptilia")
## example.df


## Returns records matching these Process IDs.
## example.df <- get.fasta(ids="ACRJP618-11|ACRJP619-11")
## example.df

## example.df <- get.fasta(ids="Example 10|Example 11|Example 12")
## example.df


## Returns records matching these BIN URIs.
## example.df <- get.fasta(bin="BOLD:AAA5125|BOLD:AAA5126")
## example.df


## Returns records for specimens stored within matching institutions.
## example.df <- get.fasta(institutions="Biodiversity Institute of Ontario|York University")
## example.df


## Returns records for specimens collected or identified by the matching researchers.
## example.df <- get.fasta(researchers="Thibaud Decaens|Rodolphe Rougerie")
## example.df


## Returns records for specimens collected in the matching geographic sites.
## example.df <- get.fasta(geo="Canada|Alaska")
## example.df


## Returns records for specimens collected with the matching marker.
## example.df <- get.fasta(marker="COI-5P")
## example.df

## example.df <- get.fasta(marker="matK|rbcL")
## example.df
```

get.private   *A function to retrieve data from BOLD using the Private API*

Description

This function retrieves data from BOLD using the private API. The user can access both public data and their own private data using this function.

Usage

```
get.private(record.codes, token)
```

Arguments

record.codes	A vector of dataset codes or project codes
token	A valid token obtained from loging in to BOLD.

Value

A dataframe of records obtaiend through the private API corresponding to the record.codes parameter entered into the function.

Author(s)

Nishan Mudalige

Examples

```
## token <- login.bold("username")

## example.df <- get.private("ACAGA", token)
## example.df

## example.df <- get.private(c("ACAGA","CCANN","PRT"), token)
## example.df
```

---

get.public   *A function to retrieve public data using the public API.*

Description

A function to retrieve public data using the public API. The user can access public data based on taxon, id, bin, container, institution, researchers, geographic location or marker.

Usage

```
get.public(taxon = NULL, ids = NULL, bin = NULL, container = NULL, institutions = NULL, researchers =
```

Arguments

taxon	Returns all records containing matching taxa, defined in a pipe or comma delimited list. (spaces need to be encoded).Taxa includes scientific names at phylum, class, order, family, subfamily, genus, and species levels.
ids	Returns all records containing matching IDs, defined in a pipe or comma delimited list. (spaces need to be encoded). IDs include Sample IDs, Process IDs, Museum IDs and Field IDs.
bin	Returns all records contained in matching BINs, defined in a pipe or comma delimited list. A BIN is defined by a Barcode Index Number URI.
container	Returns all records contained in matching projects or datasets, defined in a pipe or comma delimited list. Containers include project codes and dataset codes.

institutions	Returns all records stored in matching institutions, defined in a pipe or comma delimited list. Institutions are the Specimen Storing Site (spaces need to be encoded).
researchers	Returns all records containing matching researcher names, defined in a pipe or comma delimited list. Researchers include collectors and specimen identifiers (spaces need to be encoded).
geo	Returns all records collected in matching geographic sites, defined in a pipe or comma delimited list. Geographic sites includes countries and province/states (spaces need to be encoded).
marker	Returns all specimen records containing matching marker codes defined in a pipe or comma delimited list. All markers for a specimen matching the search string will be returned. ie. A record with COI-5P and ITS will return sequence data for both markers even if only COI-5P was specified.

Value

A dataframe which contains information corresponding to the recordset codes entered.

Author(s)

Nishan Mudalige

Examples

```
## Returns all records matching a project id
## example.df <- get.public(container="CCANN")
## example.df


## Returns all records matching multiple project ids
## example.df <- get.public(container=c("ACAGA","CCANN","AICC","PRT"))
## example.df


## Returns all records matching multiple project ids
## example.df <- get.public(container=c("ACAGA,CCANN,AICC,PRT"))
## example.df


## Returns all records matching multiple project ids
## example.df <- get.public(container=c("ACAGA|CCANN|AICC|PRT"))
## example.df


## Returns all records matching multiple project ids
## example.df <- get.public(container=c("ACAGA,CCANN|AICC,PRT"))
## example.df


## Returns all records matching the taxon Bos taurus.
## example.df <- get.public(taxon="Bos taurus")
## example.df


## Returns all records matching the taxa Aves or Reptilia
```

```
## example.df <- get.public(taxon="Aves|Reptilia")
## example.df


## Returns records matching these Process IDs.
## example.df <- get.public(ids="ACRJP618-11|ACRJP619-11")
## example.df


## Returns records matching these BIN URIs.
## example.df <- get.public(bin="BOLD:AAA5125|BOLD:AAA5126")
## example.df


## Returns records for specimens stored within matching institutions.
## example.df <- get.public(institutions="Biodiversity Institute of Ontario|York University")
## example.df


## Returns records for specimens collected or identified by the matching researchers.
## example.df <- get.public(researchers="Thibaud Decaens|Rodolphe Rougerie")
## example.df


## Returns records for specimens collected in the matching geographic sites.
## example.df <- get.public(geo="Canada|Alaska")
## example.df


## Returns records for specimens collected with the matching marker.
## example.df <- get.public(marker="COI-5P")
## example.df

## example.df <- get.public(marker="matK|rbcL")
## example.df
```

---

login.bold   *A function to login to BOLD.*

Description

A function to login to BOLD using a valid username and password. Returns a token that identifies the user to the BOLD system.

Usage

```
login.bold(username)
```

Arguments

username   A valid username to login to BOLD

Value

Returns a token which can be used to retrieve private data from BOLD.

Author(s)

Nishan Mudalige

Examples

```
## token <- login.bold("username")
## token
```

---

merge.bold   *A function to merge two dataframes with data obtained from BOLD.*

Description

A function which merges two data frames which contain data obtained from BOLD. Merging data follows a hierarchy for overwriting data. Data obtained from the private API has higher priority than the same data obtained from the public API and newer data has higher priority over older data.

Usage

```
merge.bold(df.x, df.y)
```

Arguments

df.x   A dataframe with data obtaiend from BOLD.

df.y   A dataframe with data obtaiend from BOLD.

Value

A dataframe with merged data.

Author(s)

Nishan Mudalige

Examples

```
## token <- login.bold("username")

## example.df1 <- get.public(container="CCANN")
## example.df2 <- get.private("CCANN", token)
## example.df3 <- get.public(container="ACAGA")
## example.df4 <- get.private("ACAGA", token)
## example.df5 <- get.public(container="DS-ACGAG1")
## example.df6 <- get.private("DS-ACGAG1", token)
## example.df7 <- get.public(taxon="Reptilia")

## example.df.A <- merge.bold(## example.df1, ## example.df1)
## example.df.A

## example.df.B <- merge.bold(## example.df1, ## example.df2)
## example.df.B

## example.df.C <- merge.bold(## example.df1, ## example.df3)
```

```
## example.df.C

## example.df.D <- merge.bold(## example.df4, ## example.df5)
## example.df.D

## example.df.E <- merge.bold(## example.df6, ## example.df7)
## example.df.E
```

---

nucleotides	*A function to view the different nucleotides in a dataframe*

Description

The user should input a datafarme with data obtained from BOLD. This function will then return a vector of all of the different nucleotide markers in the dataframe.

Usage

```
nucleotides(df)
```

Arguments

df   A dataframe with data obtained form BOLD.

Value

A vector of all of the different nucleotide markers in the dataframe.

Author(s)

Nishan mudalige

Examples

```
example.df <- get.public(container=c("AICC, PRT"))
nucleotides(example.df)
```

---

summary.bold	*A function which returns useful summary informaiton.*

Description

The user should enter a dataframe with data obtained from BOLD. The function will provide summary information on: * The Number of records in the dataframe * The Number of unique record codes in the dataframe * The Number of duplicate records in the dataframe * The Number of different primers in the dataframe

Usage

```
summary.bold(df)
```

Arguments

df   A dataframe with data obtained form BOLD.

Value

A summary of the items mentioned in the description.

Author(s)

Nishan Mudalige

Examples

```
example.df <- get.public(container=c("AICC, PRT"))
summary(example.df)
```

Index