

Package ‘OrganismDbi’

April 10, 2015

Title Software to enable the smooth interfacing of different database packages.

Description The package enables a simple unified interface to several annotation packages each of which has its own schema by taking advantage of the fact that each of these packages implements a select methods.

Version 1.8.1

Author Marc Carlson, Herve Pages, Martin Morgan, Valerie Obenchain

Maintainer Biocore Data Team <maintainer@bioconductor.org>

Depends R (>= 2.14.0), methods, AnnotationDbi (>= 1.16.10), GenomicFeatures (>= 1.17.13)

Imports BiocGenerics, graph, RBGL, AnnotationDbi, stats

Suggests Homo.sapiens, Rattus.norvegicus, RUnit

Collate AllGenerics.R AllClasses.R methods-select.R
methods-transcripts.R createOrganismPackage.R seqinfo.R
test_OrganismDbi_package.R

License Artistic-2.0

biocViews Annotation, Infrastructure

R topics documented:

makeOrganismPackage	2
OrganismDb-class	3
rangeBasedAccessors	5

Index	7
--------------	----------

makeOrganismPackage *Making OrganismDb packages from annotation packages.*

Description

makeOrganismPackage is a method that generates a package that will load an appropriate annotationOrganismDb object that will in turn point to existing annotation packages.

Usage

```
makeOrganismPackage (pkgname,
                    graphData,
                    organism,
                    version,
                    maintainer,
                    author,
                    destDir,
                    license="Artistic-2.0")
```

Arguments

pkgname	What is the desired package name. Traditionally, this should be the genus and species separated by a ".". So as an example, "Homo.sapiens" would be the package name for human
graphData	A list of short character vectors. Each character vector in the list is exactly two elements long and represents a join relationship between two packages. The names of these character vectors are the package names and the values are the foreign keys that should be used to connect each package. All foreign keys must be values that can be returned by the columns method for each package in question, and obviously they also must be the same kind of identifier as well.
organism	The name of the organism this package represents
version	What is the version number for this package?
maintainer	Who is the package maintainer? (must include email to be valid)
author	Who is the creator of this package?
destDir	A path where the package source should be assembled.
license	What is the license (and it's version)

Details

The purpose of this method is to create a special package that will depend on existing annotation packages and which will load a special annotationOrganismDb object that will allow proper dispatch of special select methods. These methods will allow the user to easily query across multiple annotation resources via information contained by the annotationOrganismDb object. Because the end result will be a package that treats all the data mapped together as a single source, the user is encouraged to take extra care to ensure that the different packages used are from the same build etc.

Value

A special package to load an [OrganismDb](#) object.

Author(s)

M. Carlson

See Also

[OrganismDb](#)

Examples

```
## set up the list with the relevant relationships:
gd <- list(join1 = c(GO.db="GOID", org.Hs.eg.db="GO"),
           join2 = c(org.Hs.eg.db="ENTREZID",
                    TxDb.Hsapiens.UCSC.hg19.knownGene="GENEID"))

## sets up a temporary directory for this example
## (users wont need to do this step)
destination <- tempfile()
dir.create(destination)

## makes an Organism package for human called Homo.sapiens
makeOrganismPackage(pkgname = "Homo.sapiens",
                    graphData = gd,
                    organism = "Homo sapiens",
                    version = "1.0.0",
                    maintainer = "Bioconductor Package Maintainer <maintainer@bioconductor.org>",
                    author = "Bioconductor Core Team",
                    destDir = destination,
                    license = "Artistic-2.0")
```

OrganismDb-class

OrganismDb objects

Description

The `OrganismDb` class is a container for storing knowledge about existing Annotation packages and the relationships between these resources. The purpose of this object and its associated methods is to provide a means by which users can conveniently query for data from several different annotation resources at the same time using a familiar interface.

The supporting methods `select`, `columns` and `keys` are used together to extract data from an `OrganismDb` object in a manner that should be consistent with how these are used on the supporting annotation resources.

The family of `seqinfo` style getters (`seqinfo`, `seqlevels`, `seqlengths`, `isCircular`, `genome`, and `seqnameStyle`) is also supported for `OrganismDb` objects provided that the object in question has an embedded `TxDb` object.

Methods

In the code snippets below, `x` is a `OrganismDb` object. For the metadata and show methods, there is also support for `FeatureDb` objects.

`keytypes(x)`: allows the user to discover which keytypes can be passed in to `select` or `keys` and the `keytype` argument.

`keys(x, keytype, pattern, column, fuzzy)`: Return keys for the database contained in the `TxDB` object .

The `keytype` argument specifies the kind of keys that will be returned and is always required.

If `keys` is used with `pattern`, it will pattern match on the `keytype`.

But if the `column` argument is also provided along with the `pattern` argument, then `pattern` will be matched against the values in `column` instead.

If `keys` is called with `column` and no `pattern` argument, then it will return all keys that have corresponding values in the `column` argument.

Thus, the behavior of `keys` all depends on how many arguments are specified.

Use of the `fuzzy` argument will toggle fuzzy matching to `TRUE` or `FALSE`. If `pattern` is not used, `fuzzy` is ignored.

`columns(x)`: shows which kinds of data can be returned for the `OrganismDb` object.

`select(x, keys, columns, keytype)`: When all the appropriate arguments are specified `select` will retrieve the matching data as a `data.frame` based on parameters for selected keys and columns and `keytype` arguments.

Author(s)

Marc Carlson

See Also

- [AnnotationDb-class](#) for more description of methods `select`, `keytypes`, `keys` and `columns`.
- [makeOrganismPackage](#) for functions used to generate an `OrganismDb` based package.
- [rangeBasedAccessors](#) for the range based methods used in extracting data from a `OrganismDb` object.
- [seqlevels](#) .
- [seqlengths](#) .
- [isCircular](#) .
- [genome](#) .

Examples

```
## load a package that creates an OrganismDb
library(Homo.sapiens)
ls(2)
## then the methods can be used on this object.
columns <- columns(Homo.sapiens)[c(7,10,11,12)]
keys <- head(keys(org.Hs.eg.db, "ENTREZID"))
keytype <- "ENTREZID"
```

```
res <- select(Homo.sapiens, keys, columns, keytype)
head(res)
```

rangeBasedAccessors *Extract genomic features from an object*

Description

Generic functions to extract genomic features from an object. This page documents the methods for [OrganismDb](#) objects only.

Usage

```
## S4 method for signature OrganismDb
transcripts(x, vals=NULL, columns=c("TXID", "TXNAME"))
```

```
## S4 method for signature OrganismDb
exons(x, vals=NULL, columns="EXONID")
```

```
## S4 method for signature OrganismDb
cds(x, vals=NULL, columns="CDSID")
```

```
## S4 method for signature OrganismDb
genes(x, vals=NULL, columns="GENEID")
```

```
## S4 method for signature OrganismDb
transcriptsBy(x, by, columns)
```

```
## S4 method for signature OrganismDb
exonsBy(x, by, columns)
```

```
## S4 method for signature OrganismDb
cdsBy(x, by, columns)
```

Arguments

x	A TxDb object.
...	Arguments to be passed to or from methods.
by	One of "gene", "exon", "cds" or "tx". Determines the grouping.
columns	The columns or kinds of metadata that can be retrieved from the database. All possible columns are returned by using the columns method.
vals	Either NULL or a named list of vectors to be used to restrict the output. Valid names for this list are: "gene_id", "tx_id", "tx_name", "tx_chrom", "tx_strand", "exon_id", "exon_name", "exon_chrom", "exon_strand", "cds_id", "cds_name", "cds_chrom", "cds_strand" and "exon_rank".

Details

These are the range based functions for extracting transcript information from a [OrganismDb](#) object.

Value

a GRanges or GRangesList object

Author(s)

M. Carlson

See Also

- [OrganismDb-class](#) for how to use the simple "select" interface to extract information from a OrganismDb object.
- [transcripts](#) for the original transcripts method and related methods.
- [transcriptsBy](#) for the original transcriptsBy method and related methods.

Examples

```
## extracting all transcripts from Homo.sapiens with some extra metadata
library(Homo.sapiens)
cols = c("TXNAME", "SYMBOL")
res <- transcripts(Homo.sapiens, columns=cols)

## extracting all transcripts from Homo.sapiens, grouped by gene and
## with extra metadata
res <- transcriptsBy(Homo.sapiens, by="gene", columns=cols)

## list possible values for columns argument:
columns(Homo.sapiens)
```

Index

*Topic **methods**

- rangeBasedAccessors, 5
- AnnotationDb-class, 4
- cds, OrganismDb-method
 - (rangeBasedAccessors), 5
- cdsBy, OrganismDb-method
 - (rangeBasedAccessors), 5
- class:OrganismDb (OrganismDb-class), 3
- columns, OrganismDb-method
 - (OrganismDb-class), 3
- exons, OrganismDb-method
 - (rangeBasedAccessors), 5
- exonsBy, OrganismDb-method
 - (rangeBasedAccessors), 5
- genes, OrganismDb-method
 - (rangeBasedAccessors), 5
- genome, 4
- isCircular, 4
- keys, OrganismDb-method
 - (OrganismDb-class), 3
- keytypes, OrganismDb-method
 - (OrganismDb-class), 3
- makeOrganismPackage, 2, 4
- metadata, OrganismDb-method
 - (OrganismDb-class), 3
- OrganismDb, 3, 5, 6
- OrganismDb (OrganismDb-class), 3
- OrganismDb-class, 3, 6
- rangeBasedAccessors, 4, 5
- select, OrganismDb-method
 - (OrganismDb-class), 3
- seqinfo, OrganismDb-method
 - (OrganismDb-class), 3
- seqlengths, 4
- seqlevels, 4
- transcripts, 6
- transcripts, OrganismDb-method
 - (rangeBasedAccessors), 5
- transcriptsBy, 6
- transcriptsBy, OrganismDb-method
 - (rangeBasedAccessors), 5
- TxDb, 4, 5