

Package ‘TFBSTools’

April 5, 2014

Version 1.0.0

Date 2013-10-07

Title Software package for TFBS

Description Software package for TFBS.

Author Ge Tan <ge.tan09@imperial.ac.uk>

Maintainer Ge Tan <ge.tan09@imperial.ac.uk>

Imports Biostrings(>= 2.29.19), RSQLite(>= 0.11.4), seqLogo

Depends R (>= 3.0.1), methods, IRanges(>= 1.19.38)

License GPL-2

URL <http://jaspardev.genereg.net/>

Type Package

biocViews MotifAnnotation, GeneRegulation, MotifDiscovery

NeedsCompilation no

LazyData yes

Collate AllGenerics.r AllClasses.r show-methods.r util-methods.r
XMatrix-methods.r XMatrixList-methods.r Site-methods.r
coercion-methods.r DB-methods.r PairwiseAlignment-methods.r
ICM-methods.r PWM-methods.r

R topics documented:

TFBSTools-package	2
deleteMatrixHavingID	2
getMatrixByID	3
getMatrixSet	4
XMatrix	6
XMatrixList	8

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

TFBSTools-package *TFBS software package.*

Description

TFBS includes a set of tools for transcription factor binding site detection and analysis as well as database interface functions for JASPAR, etc.

Author(s)

Ge Tan

deleteMatrixHavingID *JASPAR database operations*

Description

The functions to initialize, store matrix or delete matrix in JASPAR database.

Usage

```

## S4 method for signature character
deleteMatrixHavingID(x, IDs)
## S4 method for signature SQLiteConnection
deleteMatrixHavingID(x, IDs)
## S4 method for signature JASPAR2014
deleteMatrixHavingID(x, IDs)
## S4 method for signature character,PFMatrixList
storeMatrix(x, pfmList)
## S4 method for signature SQLiteConnection,PFMatrixList
storeMatrix(x, pfmList)
## S4 method for signature JASPAR2014,PFMatrixList
storeMatrix(x, pfmList)
## S4 method for signature character,PFMatrix
storeMatrix(x, pfmList)
## S4 method for signature SQLiteConnection,PFMatrix
storeMatrix(x, pfmList)
## S4 method for signature JASPAR2014,PFMatrix
storeMatrix(x, pfmList)
## S4 method for signature SQLiteConnection
initializeJASPARDB(x)
## S4 method for signature character
initializeJASPARDB(x)
## S4 method for signature JASPAR2014
initializeJASPARDB(x)

```

Arguments

x	a character vector of length 1 for the path of JASPAR SQLite file, or a SQLiteConnection object.
IDs	JASPAR stable IDs.
pfmList	The PFMatrixList object, or pfm object.

Examples

```
## Not run:
library(JASPAR2014)
deleteMatrixHavingID(JASPAR2014, "MA0003")
storeMatrix(JASPAR2014, pfm)
initializeJASPARDB("jaspar.sqlite")

## End(Not run)
```

getMatrixByID	<i>Basic JASPAR database search functionis</i>	getMatrixByID, getMatrixByName
---------------	--	-----------------------------------

Description

This method fetches matrix data under the given ID or name from the database and returns a XMatrix object.

Usage

```
## S4 method for signature character
getMatrixByID(x, ID)
## S4 method for signature SQLiteConnection
getMatrixByID(x, ID)
## S4 method for signature JASPAR2014
getMatrixByID(x, ID)
## S4 method for signature character
getMatrixByName(x, name)
## S4 method for signature SQLiteConnection
getMatrixByName(x, name)
## S4 method for signature JASPAR2014
getMatrixByName(x, name)
```

Arguments

x	a character vector of length 1 for the path of JASPAR SQLite file, a SQLiteConnection object, or a JASPAR2014 object.
ID	a character vector of length 1 for the JASPAR stable ID. See more details below.
name	a character vector of length 1 for the JASPAR profile name.

Details

For getMatrixByID, ID is a string which refers to the stable JASPAR ID (usually something like "MA0001") with or without version numbers. "MA0001" will give the latest version on MA0001, while "MA0001.2" will give the second version, if existing.

For getMatrixByName, according to the current JASPAR data model, name is not necessarily a unique identifier. Also, names change over time. In the case where there are several matrices with the same name in the database, the function fetches the first one and prints a warning. You've been warned. Some matrices have multiple versions. The function will return the latest version. For specific versions, use getMatrixByID(ID.version)

Value

A XMatrix object is returned. The exact type of the object depending on the type argument. NA is returned if matrix with the given ID/name is not found.

Author(s)

Ge Tan

See Also

[getMatrixSet](#)

Examples

```
## Not run:
db = "/Users/gtan/CSC/JASPAR/JASPAR2014/inst/extdata/JASPAR2014.sqlite"
pfm = getMatrixByID(db, ID="MA0003")
pfm2 = getMatrixByName(db, name="TFAP2A")

library(JASPAR2014)
pfm3 = getMatrixByID(JASPAR2014, ID="MA0003")

## End(Not run)
```

getMatrixSet

Advanced JASPAR database search functions get_MatrixSet

Description

This function fetches matrix data for all matrices in the database matching criteria defined by the named arguments and returns a PFMATRIX object

Usage

```
## S4 method for signature character
getMatrixSet(x, opts)
## S4 method for signature SQLiteConnection
getMatrixSet(x, opts)
## S4 method for signature JASPAR2014
getMatrixSet(x, opts)
```

Arguments

x a character vector of length 1 for the path of JASPAR SQLite file, a SQLiteConnection object, or a JASPAR2014 object.

opts a search options list. See more details below.

Details

The search options include three categories:

(1) Database basic criterias:

all=c(TRUE, FALSE)

ID, name

collection=c("CORE", "CNE", "PHYLOFACTS", ...)

all_versions=c(TRUE, FALSE)

species, matrixtype=c("PFM", "PWM", "ICM")

(2) Tags based criterias:

class, type, family

medline, tax_group

(3) Further criterias:

min_ic (minimum total information content of the matrix)

length (minimum sites length)

sites (minimum average sites number per base)

When all is TRUE, it will get all the matrices and has higher priority over other options. Then ID has the second highest priority, and will ignore all the following options. The rest options are combined in search with AND, while multiple elements under one options have the logical operator OR.

Author(s)

Ge Tan

See Also

[getMatrixByID](#), [getMatrixByName](#)

Examples

```

## Not run:
db = "/Users/gtan/CSC/JASPAR/JASPAR2014/inst/extdata/JASPAR2014.sqlite"
opts = list()
opts[["species"]] = 9606
opts[["name"]] = "RUNX1"
#opts[["class"]] = "Ig-fold"
opts[["type"]] = "SELEX"
opts[["all_versions"]] = TRUE
siteList = getMatrixSet(db, opts)

library(JASPAR2014)
siteList2 = getMatrixSet(JASPAR2014, opts)

## End(Not run)

```

XMatrix	"XMatrix" <i>objects</i>
---------	--------------------------

Description

XMatrix is a virtual class. No objects can be created from it directly. Three classes are derived from this class: PFMatrix, PWMatrix and ICMatrix.

PFMatrix is a class whose instances are objects representing raw position frequency matrices (PFMs).

Usage

```

## Constructors:
PFMatrix(ID="Unknown", name="Unknown", matrixClass="Unknown",
         strand="*", bg=c(A=0.25, C=0.25, G=0.25, T=0.25),
         tags=list(), matrix=matrix())
## Accessor-like methods:
## S4 method for signature XMatrix
ID(x)
## S4 method for signature XMatrix
bg(x)

## ... and more (see Methods)

```

Arguments

ID	Object of class "character": a unique identifier for each matrix.
name	Object of class "character": The name of the transcription factor. In JASPAR, as far as possible, the name is based on the standardized Entrez gene symbols. In the case the model describes a transcription factor hetero-dimer, two names are concatenated, such as RXR-VDR. In a few cases, different splice forms of

the same gene have different binding specificity: in this case the splice form information is added to the name, based on the relevant literature.

matrixClass	Object of class "character": Structural class of the transcription factor, based on the TFCaT system
strand	Object of class "character": which strand is the binding sites sequences from.
bg	Object of class "numeric": Background frequencies of the four bases. By default, it is equally 0.25.
tags	Object of class "list": Some tags information about this model. Tags include: <ol style="list-style-type: none"> (1) "family": Structural sub-class of the transcription factor, based on the TFCaT system. (2) "species": The species source for the sequences, in NCBI tax IDs. (3) "tax_group": Group of species, currently consisting of 4 larger groups: vertebrate, insect, plant, chordate. (4) "medline": a ID to the relevant publication reporting the sites used in the mode building. (5) "type": Methodology used for matrix construction. (6) "ACC": A representative protein accession number in Genbank for the transcription factor. Human takes precedence if several exists. (6) "pazar_tf_id": a ID to PAZAR database. (7) "TFBSshape_ID": a ID to TFBSshape database. (8) "TFencyclopedia_ID": a ID to the Transcription Factor Encyclopedia. (9) "comment": For some matrices, a curator comment is added.
matrix	Object of class "matrix": This is the matrix information.
x	Object of class XMatrix.

Methods

ID signature(x = "XMatrix"): Gets the ID slot.

length signature(x = "XMatrix"): Gets the pattern length in nucleotides (i.e. number of columns in the matrix).

reverseComplement signature(x = "XMatrix"): Generates the reverse complement matrix object. Note than the strand is XMatrix will also be changed to the oposite strand.

as.matrix signature(x = "XMatrix"): Returns the matrix in the XMatrix class.

Author(s)

Ge Tan

Examples

```
## -----
## Constructor
## Note that there is no XMatrix() constructor,
## but an XMatrix family of constructors: PFMatrix(), PWMMatrix(), ICMatrix()
pfm = PFMatrix(ID="MA0004.1", name="Arnt", matrixClass="Zipper-Type", strand="+",
```

```

bg=c(A=0.25, C=0.25, G=0.25, T=0.25),
tags=list(family="Helix-Loop-Helix", species="10090", tax_group="vertebrates",
medline="7592839", type="SELEX", ACC="P53762", pazar_tf_id="TF0000003",
TFBSShape_ID="11", TFencyclopedia_ID="580"),
matrix=matrix(c(4L, 19L, 0L, 0L, 0L, 0L,
16L, 0L, 20L, 0L, 0L, 0L,
0L, 1L, 0L, 20L, 0L, 20L,
0L, 0L, 0L, 0L, 20L, 0L),
byrow=TRUE, nrow=4, dimnames=list(c("A", "C", "G", "T")))
)

## Coersion
as.matrix(pfm)
as(pfm, "matrix")

## Methods
reverseComplement(pfm)
length(pfm)

```

XMatrixList

Class "XMatrixList"

Description

The XMatrixList virtual class is a container for storing a collection of XMatrix objects. No object can be constructed directly from this virtual and it has three subclasses: PFMatrixList, PWMatrixList and ICMatrixList. Basically it is a SimpleList and is designed for manipulating the set of XMatrix objects as a whole.

Usage

```

## Constructors:
PFMatrixList(..., use.names=TRUE)

## Accessor-like methods:
## S4 method for signature XMatrixList
ID(x)
## S4 method for signature XMatrixList
name(x)

```

Arguments

...	The XMatrix objects are supplied in
use.names	A logical value. When TRUE, the names of the XMatrix will be kept.
x	A XMatrixList object.

Author(s)

Ge Tan

See Also

[XMatrix](#),

Examples

```
## Not run:  
  PFMatrixList(pfm1=pfm, pfm2=pfm, use.names=TRUE)  
  
## End(Not run)
```

Index

- *Topic **\textasciitilde\textasciitilde**
other possible keyword(s)
\textasciitilde\textasciitilde
 - deleteMatrixHavingID, 2
 - getMatrixByID, 3
 - getMatrixSet, 4
- *Topic **classes**
 - XMatrix, 6
 - XMatrixList, 8
- *Topic **methods**
 - deleteMatrixHavingID, 2
 - getMatrixByID, 3
 - getMatrixSet, 4
- *Topic **package**
 - TFBSTools-package, 2
 - [, Site-method (XMatrix), 6
 - [-methods (XMatrix), 6
- bg (XMatrixList), 8
- bg, XMatrix-method (XMatrix), 6
- bg, XMatrixList-method (XMatrixList), 8
- bg<- , XMatrix-method (XMatrix), 6
- deleteMatrixHavingID, 2
- deleteMatrixHavingID, character-method
(deleteMatrixHavingID), 2
- deleteMatrixHavingID, JASPAR2014-method
(deleteMatrixHavingID), 2
- deleteMatrixHavingID, SQLiteConnection-method
(deleteMatrixHavingID), 2
- getMatrixByID, 3, 5
- getMatrixByID, character-method
(getMatrixByID), 3
- getMatrixByID, JASPAR2014-method
(getMatrixByID), 3
- getMatrixByID, SQLiteConnection-method
(getMatrixByID), 3
- getMatrixByName, 5
- getMatrixByName (getMatrixByID), 3
- getMatrixByName, character-method
(getMatrixByID), 3
- getMatrixByName, JASPAR2014-method
(getMatrixByID), 3
- getMatrixByName, SQLiteConnection-method
(getMatrixByID), 3
- getMatrixSet, 4, 4
- getMatrixSet, character-method
(getMatrixSet), 4
- getMatrixSet, JASPAR2014-method
(getMatrixSet), 4
- getMatrixSet, SQLiteConnection-method
(getMatrixSet), 4
- ID (XMatrixList), 8
- ID, XMatrix-method (XMatrix), 6
- ID, XMatrixList-method (XMatrixList), 8
- ID<- , XMatrix-method (XMatrix), 6
- initializeJASPARDB
(deleteMatrixHavingID), 2
- initializeJASPARDB, character-method
(deleteMatrixHavingID), 2
- initializeJASPARDB, JASPAR2014-method
(deleteMatrixHavingID), 2
- initializeJASPARDB, SQLiteConnection-method
(deleteMatrixHavingID), 2
- length, Site-method (XMatrix), 6
- length, SitePair-method (XMatrix), 6
- length, XMatrix-method (XMatrix), 6
- length-methods (XMatrix), 6
- Matrix (XMatrixList), 8
- Matrix, XMatrix-method (XMatrix), 6
- Matrix, XMatrixList-method
(XMatrixList), 8
- Matrix<- , XMatrix-method (XMatrix), 6
- matrixClass (XMatrixList), 8
- matrixClass, XMatrix-method (XMatrix), 6

- matrixClass, XMatrixList-method (XMatrixList), 8
- matrixClass<-, XMatrix-method (XMatrix), 6
- matrixType (XMatrixList), 8
- matrixType, ICMatrix-method (XMatrix), 6
- matrixType, PFMatrix-method (XMatrix), 6
- matrixType, PWMMatrix-method (XMatrix), 6
- matrixType, XMatrixList-method (XMatrixList), 8
- matrixType-methods (XMatrix), 6

- name (XMatrixList), 8
- name, XMatrix-method (XMatrix), 6
- name, XMatrixList-method (XMatrixList), 8
- name<-, XMatrix-method (XMatrix), 6

- PFMatrix (XMatrix), 6
- PFMatrix-class (XMatrix), 6
- PFMatrixList (XMatrixList), 8
- PFMatrixList-class (XMatrixList), 8

- reverseComplement, XMatrix-method (XMatrix), 6

- storeMatrix (deleteMatrixHavingID), 2
- storeMatrix, character, PFMatrix-method (deleteMatrixHavingID), 2
- storeMatrix, character, PFMatrixList-method (deleteMatrixHavingID), 2
- storeMatrix, JASPAR2014, PFMatrix-method (deleteMatrixHavingID), 2
- storeMatrix, JASPAR2014, PFMatrixList-method (deleteMatrixHavingID), 2
- storeMatrix, SQLiteConnection, PFMatrix-method (deleteMatrixHavingID), 2
- storeMatrix, SQLiteConnection, PFMatrixList-method (deleteMatrixHavingID), 2
- storeMatrix-methods (deleteMatrixHavingID), 2
- strand (XMatrixList), 8
- strand, XMatrix-method (XMatrix), 6
- strand, XMatrixList-method (XMatrixList), 8
- strand<-, XMatrix-method (XMatrix), 6

- tags (XMatrix), 6
- tags, XMatrix-method (XMatrix), 6
- tags, XMatrixList-method (XMatrixList), 8

- TFBSTools (TFBSTools-package), 2
- TFBSTools-package, 2

- XMatrix, 6, 9
- XMatrix-class (XMatrix), 6
- XMatrixList, 8
- XMatrixList, list-method (XMatrixList), 8