org.Cf.eg.db
April 10, 2024

org.Cf.egACCNUM Map Entrez Gene identifiers to GenBank Accession Numbers

Description
org.Cf.egACCNUM is an R object that contains mappings between Entrez Gene identifiers and GenBank accession numbers.

Details

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

See Also
• AnnotationDb-class for use of the select() interface.

Examples
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egACCNUM
# Get the entrez gene identifiers that are mapped to an ACCNUM
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the ACCNUM for the first five genes
  xx[1:5]
  # Get the first one
## org.Cf.egALIAS2EG

### Map between Common Gene Symbol Identifiers and Entrez Gene

#### Description

org.Cf.egALIAS is an R object that provides mappings between common gene symbol identifiers and entrez gene identifiers.

#### Details

Each gene symbol maps to a named vector containing the corresponding entrez gene identifier. The name of the vector corresponds to the gene symbol. Since gene symbols are sometimes redundantly assigned in the literature, users are cautioned that this map may produce multiple matching results for a single gene symbol. Users should map back from the entrez gene IDs produced to determine which result is the one they want when this happens.

Because of this problem with redundant assignment of gene symbols, it is never advisable to use gene symbols as primary identifiers.

This mapping includes ALL gene symbols including those which are already listed in the SYMBOL map. The SYMBOL map is meant to only list official gene symbols, while the ALIAS maps are meant to store all used symbols.

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

#### References


#### See Also

- [AnnotationDb-class](#) for use of the select() interface.
Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
# Convert the object to a list
xx <- as.list(org.Cf.egALIAS2EG)
# Remove pathway identifiers that do not map to any entrez gene id
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # The entrez gene identifiers for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}

---

org.Cf.eg.db  
Bioconductor annotation data package

Description

Welcome to the org.Cf.eg.db annotation Package. This is an organism specific package. The purpose is to provide detailed information about the species abbreviated in the second part of the package name org.Cf.eg.db. This package is updated biannually.

Objects in this package are accessed using the select() interface. See ?select in the AnnotationDbi package for details.

See Also

- AnnotationDb-class for use of keys(), columns() and select().

Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.
columns(org.Cf.eg.db)

## Bimap interface:
## The 'old style' of interacting with these objects is manipulation as
## bimaps. While this approach is still available we strongly encourage the
## use of select().
ls("package:org.Cf.eg.db")
org.Cf.egCHR

Map Entrez Gene IDs to Chromosomes

Description

org.Cf.egCHR is an R object that provides mappings between entrez gene identifiers and the chromosome that contains the gene of interest.

Details

Each entrez gene identifier maps to a vector of a chromosome.

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

See Also

• AnnotationDb-class for use of the select() interface.

Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egCHR
# Get the entrez gene identifiers that are mapped to a chromosome
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
    # Get the CHR for the first five genes
    xx[1:5]
    # Get the first one
    xx[1]
}

org.Cf.egCHRLENGTHS

A named vector for the length of each of the chromosomes

Description

org.Cf.egCHRLENGTHS provides the length measured in base pairs for each of the chromosomes.
**Details**

This is a named vector with chromosome numbers as the names and the corresponding lengths for chromosomes as the values.

Total lengths of chromosomes were derived by calculating the number of base pairs on the sequence string for each chromosome.

**See Also**

- *AnnotationDb-class* for use of the select() interface.

**Examples**

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
tt <- org.Cf.egCHRLENGTHS
# Length of chromosome 1
tt["1"]
```

---

**Description**

org.Cf.egCHRLOC is an R object that maps entrez gene identifiers to the starting position of the gene. The position of a gene is measured as the number of base pairs.

The CHRLOCEND mapping is the same as the CHRLOC mapping except that it specifies the ending base of a gene instead of the start.

**Details**

Each entrez gene identifier maps to a named vector of chromosomal locations, where the name indicates the chromosome.

Chromosomal locations on both the sense and antisense strands are measured as the number of base pairs from the p (5' end of the sense strand) to q (3' end of the sense strand) arms. Chromosomal locations on the antisense strand have a leading "-" sign (e.g. -1234567).

Since some genes have multiple start sites, this field can map to multiple locations.

Mappings were based on data provided by: UCSC Genome Bioinformatics (Canis familiaris)

With a date stamp from the source of: 2021-Nov23

**See Also**

- *AnnotationDb-class* for use of the select() interface.
Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egCHRLOC
# Get the entrez gene identifiers that are mapped to chromosome locations
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
    # Get the CHRLOC for the first five genes
    xx[1:5]
    # Get the first one
    xx[[1]]
}

---

org.Cf.egENSEMBL  Map Ensembl gene accession numbers with Entrez Gene identifiers

Description

org.Cf.egENSEMBL is an R object that contains mappings between Entrez Gene identifiers and Ensembl gene accession numbers.

Details


Mappings were based on data provided by BOTH of these sources: http://www.ensembl.org/biomart/martview/ ftp://ftp.ncbi.nlm.nih.gov/gene/DATA

For most species, this mapping is a combination of NCBI to ensembl IDs from BOTH NCBI and ensembl. Users who wish to only use mappings from NCBI are encouraged to see the ncbi2ensembl table in the appropriate organism package. Users who wish to only use mappings from ensembl are encouraged to see the ensembl2ncbi table which is also found in the appropriate organism packages. These mappings are based upon the ensembl table which is contains data from BOTH of these sources in an effort to maximize the chances that you will find a match.

For worms and flies however, this mapping is based only on sources from ensembl, as these organisms do not have ensembl to entrez gene mapping data at NCBI.

See Also

- AnnotationDb-class for use of the select() interface.
Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egENSEMBL
# Get the entrez gene IDs that are mapped to an Ensembl ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
    # Get the Ensembl gene IDs for the first five genes
    xx[1:5]
    # Get the first one
    xx[[1]]
}

# For the reverse map ENSEMBL2EG:
# Convert to a list
xx <- as.list(org.Cf.egENSEMBL2EG)
if(length(xx) > 0){
    # Gets the entrez gene IDs for the first five Ensembl IDs
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

Description

org.Cf.egENSEMBL is an R object that contains mappings between Entrez Gene identifiers and Ensembl protein accession numbers.

Details


See Also

- AnnotationDb-class for use of the select() interface.
Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:

x <- org.Cf.egENSEMBLPROT
# Get the entrez gene IDs that are mapped to an Ensembl ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
    # Get the Ensembl gene IDs for the first five proteins
    xx[1:5]
    # Get the first one
    xx[[1]]
}

# For the reverse map ENSEMBLPROT2EG:

# Convert to a list
xx <- as.list(org.Cf.egENSEMBLPROT2EG)
if(length(xx) > 0) {
    # Gets the entrez gene IDs for the first five Ensembl IDs
    xx[1:5]
    # Get the first one
    xx[[1]]
}

Description

org.Cf.egENSEMBL is an R object that contains mappings between Entrez Gene identifiers and Ensembl transcript accession numbers.

Details


See Also

- *AnnotationDb-class* for use of the select() interface.
Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egENSEMBLTRANS
# Get the entrez gene IDs that are mapped to an Ensembl ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the Ensembl gene IDs for the first five proteins
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

```r
# For the reverse map ENSEMBLTRANS2EG:
# Convert to a list
xx <- as.list(org.Cf.egENSEMBLTRANS2EG)
if(length(xx) > 0) {
  # Gets the entrez gene IDs for the first five Ensembl IDs
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

---

org.Cf.egENZYME  
Map between Entrez Gene IDs and Enzyme Commission (EC) Numbers

**Description**

org.Cf.egENZYME is an R object that provides mappings between entrez gene identifiers and EC numbers.

**Details**

Each entrez gene identifier maps to a named vector containing the EC number that corresponds to the enzyme produced by that gene. The name corresponds to the entrez gene identifier. If this information is unknown, the vector will contain an NA.

Enzyme Commission numbers are assigned by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology [http://www.chem.qmw.ac.uk/iubmb/enzyme/](http://www.chem.qmw.ac.uk/iubmb/enzyme/) to allow enzymes to be identified.

An Enzyme Commission number is of the format EC x.y.z.w, where x, y, z, and w are numeric numbers. In org.Cf.egENZYME2EG, EC is dropped from the Enzyme Commission numbers.
Enzyme Commission numbers have corresponding names that describe the functions of enzymes in such a way that EC x is a more general description than EC x.y that in turn is a more general description than EC x.y.z. The top level EC numbers and names are listed below:

EC 1 oxidoreductases
EC 2 transferases
EC 3 hydrolases
EC 4 lyases
EC 5 isomerases
EC 6 ligases

The EC name for a given EC number can be viewed at http://www.chem.qmul.ac.uk/iupac/jcbn/index.html#6

Mappings between entrez gene identifiers and enzyme identifiers were obtained using files provided by: KEGG GENOME ftp://ftp.genome.jp/pub/kegg/genomes With a date stamp from the source of: 2011-Mar15

For the reverse map, each EC number maps to a named vector containing the entrez gene identifier that corresponds to the gene that produces that enzyme. The name of the vector corresponds to the EC number.

References

See Also
- AnnotationDb-class for use of the select() interface.

Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egENZYME
# Get the entrez gene identifiers that are mapped to an EC number
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the ENZYME for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
# For the reverse map:
# Convert to a list
xx <- as.list(org.Cf.egENZYME2EG)
if(length(xx) > 0){
```
# Get the entrez gene identifiers for the first five enzyme commission numbers
xx[1:5]
# Get the first one
xx[[1]]

Description

org.Cf.egGENENAME is an R object that maps entrez gene identifiers to the corresponding gene name.

Details

Each entrez gene identifier maps to a named vector containing the gene name. The vector name corresponds to the entrez gene identifier. If the gene name is unknown, the vector will contain an NA.

Gene names currently include both the official (validated by a nomenclature committee) and preferred names (interim selected for display) for genes. Efforts are being made to differentiate the two by adding a name to the vector.

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

See Also

- AnnotationDb-class for use of the select() interface.

Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egGENENAME
# Get the gene names that are mapped to an entrez gene identifier
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
    # Get the GENE NAME for the first five genes
    xx[1:5]
    # Get the first one
    xx[[1]]
}
**Map between Entrez Gene Identifiers and Gene Type**

**Description**

`org.Cf.egGENETYPE` is an R object that provides mappings between entrez gene identifiers and gene types.

**Details**

Each entrez gene identifier is mapped to the reported type of the corresponding gene. An NA is reported if there is no known type for a given gene.

**References**


**See Also**

- `AnnotationDb-class` for use of the `select()` interface.

**Examples**

```
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egGENETYPE
# Get the gene symbol that are mapped to an entrez gene identifiers
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the SYMBOL for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
# Get the entrez gene identifiers that are mapped to a gene type
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the entrez gene ID for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```
org.Cf.egGO is an R object that provides mappings between entrez gene identifiers and the GO identifiers that they are directly associated with. This mapping and its reverse mapping do NOT associate the child terms from the GO ontology with the gene. Only the directly evidenced terms are represented here.

org.Cf.egGO2ALLEGS is an R object that provides mappings between a given GO identifier and all of the Entrez Gene identifiers annotated at that GO term OR TO ONE OF IT’S CHILD NODES in the GO ontology. Thus, this mapping is much larger and more inclusive than org.Cf.egGO2EG.

Details

If org.Cf.egGO is cast as a list, each Entrez Gene identifier is mapped to a list of lists. The names on the outer list are GO identifiers. Each inner list consists of three named elements: GOID, Ontology, and Evidence.

The GOID element matches the GO identifier named in the outer list and is included for convenience when processing the data using ‘lapply’.

The Ontology element indicates which of the three Gene Ontology categories this identifier belongs to. The categories are biological process (BP), cellular component (CC), and molecular function (MF).

The Evidence element contains a code indicating what kind of evidence supports the association of the GO identifier to the Entrez Gene id. Some of the evidence codes in use include:

- IMP: inferred from mutant phenotype
- IGI: inferred from genetic interaction
- IPI: inferred from physical interaction
- ISS: inferred from sequence similarity
- IDA: inferred from direct assay
- IEP: inferred from expression pattern
- IEA: inferred from electronic annotation
- TAS: traceable author statement
- NAS: non-traceable author statement
- ND: no biological data available
- IC: inferred by curator

A more complete listing of evidence codes can be found at:


If org.Cf.egGO2ALLEGS or org.Cf.egGO2EG is cast as a list, each GO term maps to a named vector of entrez gene identifiers and evidence codes. A GO identifier may be mapped to the same entrez
gene identifier more than once but the evidence code can be different. Mappings between Gene Ontology identifiers and Gene Ontology terms and other information are available in a separate data package named GO.

Whenever any of these mappings are cast as a data.frame, all the results will be output in an appropriate tabular form.

Mappings between entrez gene identifiers and GO information were obtained through their mappings to Entrez Gene identifiers. NAs are assigned to entrez gene identifiers that can not be mapped to any Gene Ontology information. Mappings between Gene Ontology identifiers an Gene Ontology terms and other information are available in a separate data package named GO.

All mappings were based on data provided by: Gene Ontology http://current.geneontology.org/ontology/go-basic.obo With a date stamp from the source of: 2023-07-27

For GO2ALL style mappings, the intention is to return all the genes that are the same kind of term as the parent term (based on the idea that they are more specific descriptions of the general term). However because of this intent, not all relationship types will be counted as offspring for this mapping. Only “is a” and “has a” style relationships indicate that the genes from the child terms would be the same kind of thing.

References


See Also

- org.Cf.egGO2ALLEGS
- AnnotationDb-class for use of the select() interface.

Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egGO
# Get the entrez gene identifiers that are mapped to a GO ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Try the first one
got <- xx[[1]]
got[[1]]["GOID"]
got[[1]]["Ontology"]
got[[1]]["Evidence"]
}
# For the reverse map:
# Convert to a list
xx <- as.list(org.Cf.egGO2EG)
if(length(xx) > 0){
```
# Gets the entrez gene ids for the top 2nd and 3nd GO identifiers
goids <- xx[2:3]
# Gets the entrez gene ids for the first element of goids
goids[[1]]
# Evidence code for the mappings
names(goids[[1]])
}

# For org.Cf.egGO2ALLEGS
xx <- as.list(org.Cf.egGO2ALLEGS)
if(length(xx) > 0){
  # Gets the Entrez Gene identifiers for the top 2nd and 3nd GO identifiers
goids <- xx[2:3]
  # Gets all the Entrez Gene identifiers for the first element of goids
goids[[1]]
  # Evidence code for the mappings
  names(goids[[1]])
}

org.Cf.egMAPCOUNTS

Number of mapped keys for the maps in package org.Cf.eg.db

Description

org.Cf.egMAPCOUNTS provides the "map count" (i.e. the count of mapped keys) for each map in package org.Cf.eg.db.

Details

This "map count" information is precalculated and stored in the package annotation DB. This allows some quality control and is used by the checkMAPCOUNTS function defined in AnnotationDbi to compare and validate different methods (like count.mappedkeys(x) or sum(!is.na(as.list(x)))) for getting the "map count" of a given map.

See Also

- mappedkeys
- count.mappedkeys
- checkMAPCOUNTS
- AnnotationDb-class for use of the select() interface.

Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
org.Cf.egMAPCOUNTS
mapnames <- names(org.Cf.egMAPCOUNTS)
org.Cf.egMAPCOUNTS[mapnames[1]]
x <- get(mapnames[1])
sum(!is.na(as.list(x)))
count.mappedkeys(x)  # much faster!

## Check the "map count" of all the maps in package org.Cf.eg.db
checkMAPCOUNTS("org.Cf.eg.db")

---

org.Cf.egORGANISM  
**The Organism for org.Cf.eg**

**Description**

org.Cf.egORGANISM is an R object that contains a single item: a character string that names the organism for which org.Cf.eg was built.

**Details**

Although the package name is suggestive of the organism for which it was built, org.Cf.egORGANISM provides a simple way to programmatically extract the organism name.

**See Also**

- [AnnotationDb-class](#) for use of the `select()` interface.

**Examples**

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
org.Cf.egORGANISM
```

---

org.Cf.egPATH  
**Mappings between Entrez Gene identifiers and KEGG pathway identifiers**

**Description**

KEGG (Kyoto Encyclopedia of Genes and Genomes) maintains pathway data for various organisms. org.Cf.egPATH maps entrez gene identifiers to the identifiers used by KEGG for pathways
Details

Each KEGG pathway has a name and identifier. Pathway name for a given pathway identifier can be obtained using the KEGG data package that can either be built using AnnBuilder or downloaded from Bioconductor [http://www.bioconductor.org](http://www.bioconductor.org).


Mappings were based on data provided by: KEGG GENOME ftp://ftp.genome.jp/pub/kegg/genomes With a date stamp from the source of: 2011-Mar15

References


See Also

- [AnnotationDb-class](http://www.genome.ad.jp/kegg/) for use of the select() interface.

Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egPATH
# Get the entrez gene identifiers that are mapped to a KEGG pathway ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the PATH for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}

# For the reverse map:
# Convert the object to a list
xx <- as.list(org.Cf.egPATH2EG)
# Remove pathway identifiers that do not map to any entrez gene id
xx <- xx[!is.na(xx)]
if(length(xx) > 0) {
  # The entrez gene identifiers for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```
Description

org.Cf.egPMID is an R object that provides mappings between entrez gene identifiers and PubMed identifiers.

Details

Each entrez gene identifier is mapped to a named vector of PubMed identifiers. The name associated with each vector corresponds to the entrez gene identifier. The length of the vector may be one or greater, depending on how many PubMed identifiers a given entrez gene identifier is mapped to. An NA is reported for any entrez gene identifier that cannot be mapped to a PubMed identifier.

Titles, abstracts, and possibly full texts of articles can be obtained from PubMed by providing a valid PubMed identifier. The pubmed function of annotate can also be used for the same purpose.

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

References


See Also

• AnnotationDb-class for use of the select() interface.

Examples

## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egPMID
# Get the entrez gene identifiers that are mapped to any PubMed ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0){
    # The entrez gene identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
    if(interactive() & & !is.null(xx[[1]]) & & !is.na(xx[[1]]) & & require(annotate)){
        # Gets article information as XML files
        xmls <- pubmed(xx[[1]], disp = "data")
    }
# Views article information using a browser
pubmed(xx[[1]], disp = "browser")
}
#
# For the reverse map:
# Convert the object to a list
xx <- as.list(org.Cf.egPMID2EG)
if(length(xx) > 0){
    # The entrez gene identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
    if(interactive() && require(annotate)){
        # Gets article information as XML files for a PubMed id
        xmls <- pubmed(names(xx)[1], disp = "data")
        # Views article information using a browser
        pubmed(names(xx)[1], disp = "browser")
    }
}

---

**org.Cf.egREFSEQ**

*Map between Entrez Gene Identifiers and RefSeq Identifiers*

**Description**

org.Cf.egREFSEQ is an R object that provides mappings between entrez gene identifiers and RefSeq identifiers.

**Details**

Each entrez gene identifier is mapped to a named vector of RefSeq identifiers. The name represents the entrez gene identifier and the vector contains all RefSeq identifiers that can be mapped to that entrez gene identifier. The length of the vector may be one or greater, depending on how many RefSeq identifiers a given entrez gene identifier can be mapped to. An NA is reported for any entrez gene identifier that cannot be mapped to a RefSeq identifier at this time.

RefSeq identifiers differ in format according to the type of record the identifiers are for as shown below:

NG\_XXXXX: RefSeq accessions for genomic region (nucleotide) records  
NM\_XXXXX: RefSeq accessions for mRNA records  
NC\_XXXXX: RefSeq accessions for chromosome records  
NP\_XXXXX: RefSeq accessions for protein records  
XR\_XXXXX: RefSeq accessions for model RNAs that are not associated with protein products  
XM\_XXXXX: RefSeq accessions for model mRNA records  
XP\_XXXXX: RefSeq accessions for model protein records

Where XXXXX is a sequence of integers.
NCBI [https://www.ncbi.nlm.nih.gov/RefSeq/] allows users to query the RefSeq database using RefSeq identifiers.


References


See Also

- [AnnotationDbi-class](https://www.bioconductor.org/packages/vignettes/statistics/data/AnnotationDbi.pdf) for use of the `select()` interface.

Examples

```r
## select() interface:
## Objects in this package can be accessed using the `select()` interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egREFSEQ
# Get the entrez gene identifiers that are mapped to any RefSeq ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the REFSEQ for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}

# For the reverse map:
x <- org.Cf.egREFSEQ2EG
# Get the RefSeq identifier that are mapped to an entrez gene ID
mapped_seqs <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_seqs])
if(length(xx) > 0) {
  # Get the entrez gene for the first five Refseqs
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

---

**org.Cf.egSYMBOL**  
Map between Entrez Gene Identifiers and Gene Symbols
**Description**

org.Cf.egSYMBOL is an R object that provides mappings between entrez gene identifiers and gene abbreviations.

**Details**

Each entrez gene identifier is mapped to the a common abbreviation for the corresponding gene. An NA is reported if there is no known abbreviation for a given gene.

Symbols typically consist of 3 letters that define either a single gene (ABC) or multiple genes (ABC1, ABC2, ABC3). Gene symbols can be used as key words to query public databases such as Entrez Gene.

Mappings were based on data provided by: Entrez Gene ftp://ftp.ncbi.nlm.nih.gov/gene/DATA With a date stamp from the source of: 2023-Sep11

**References**


**See Also**

- AnnotationDb-class for use of the select() interface.

**Examples**

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egSYMBOL
# Get the gene symbol that are mapped to an entrez gene identifiers
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the SYMBOL for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
# For the reverse map:
x <- org.Cf.egSYMBOL2EG
# Get the entrez gene identifiers that are mapped to a gene symbol
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the entrez gene ID for the first five genes
  xx[1:5]
  # Get the first one
}
Description

org.Cf.egUNIPROT is an R object that contains mappings between Entrez Gene identifiers and Uniprot accession numbers.

Details


Mappings were based on data provided by NCBI (link above) with an exception for fly, which required retrieving the data from ensembl [http://www.ensembl.org/biomart/martview/](http://www.ensembl.org/biomart/martview/)

See Also

- [AnnotationDb-class](#) for use of the select() interface.

Examples

```r
## select() interface:
## Objects in this package can be accessed using the select() interface
## from the AnnotationDbi package. See ?select for details.

## Bimap interface:
x <- org.Cf.egUNIPROT
# Get the entrez gene IDs that are mapped to a Uniprot ID
mapped_genes <- mappedkeys(x)
# Convert to a list
xx <- as.list(x[mapped_genes])
if(length(xx) > 0) {
  # Get the Uniprot gene IDs for the first five genes
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```
Collect information about the package annotation DB

Description

Some convenience functions for getting a connection object to (or collecting information about) the package annotation DB.

Usage

org.Cf.eg_dbconn()
org.Cf.eg_dbfile()
org.Cf.eg_dbschema(file="", show.indices=FALSE)
org.Cf.eg_dbInfo()

Arguments

file A connection, or a character string naming the file to print to (see the file argument of the cat function for the details).
show.indices The CREATE INDEX statements are not shown by default. Use show.indices=TRUE to get them.

Details

org.Cf.eg_dbconn returns a connection object to the package annotation DB. IMPORTANT: Don’t call dbDisconnect on the connection object returned by org.Cf.eg_dbconn or you will break all the AnnDbObj objects defined in this package!
org.Cf.eg_dbfile returns the path (character string) to the package annotation DB (this is an SQLite file).
org.Cf.eg_dbschema prints the schema definition of the package annotation DB.
org.Cf.eg_dbInfo prints other information about the package annotation DB.

Value

org.Cf.eg_dbconn: a DBIConnection object representing an open connection to the package annotation DB.
org.Cf.eg_dbfile: a character string with the path to the package annotation DB.
org.Cf.eg_dbschema: none (invisible NULL).
org.Cf.eg_dbInfo: none (invisible NULL).

See Also
dbGetQuery, dbConnect, dbconn, dbfile, dbschema, dbInfo
Examples

library(DBI)
## Count the number of rows in the "genes" table:
dbGetQuery(org.Cf.eg_dbconn(), "SELECT COUNT(*) FROM genes")

org.Cf.eg_dbschema()
org.Cf.eg_dbInfo()
Index

* datasets
  org.Cf.eg.db, 3
  org.Cf.eg_dbconn, 23
  org.Cf.egACCCOUNT, 1
  org.Cf.egALIAS2EG, 2
  org.Cf.egCHR, 4
  org.Cf.egCHRLENGTHS, 4
  org.Cf.egCHRLOC, 5
  org.Cf.egENSEMBL, 6
  org.Cf.egENSEMBLPROT, 7
  org.Cf.egENSEMBLELED, 7
  org.Cf.egENSEMBLTRANS, 8
  org.Cf.egENSEMBL2EG, 8
  org.Cf.egENZYME, 9
  org.Cf.egGENENAME, 11
  org.Cf.egGENETYPE, 12
  org.Cf.egGO, 13
  org.Cf.egMAPCOUNTS, 15
  org.Cf.egORGANISM, 16
  org.Cf.egPATH, 16
  org.Cf.egPMID, 18
  org.Cf.egREFSEQ, 19
  org.Cf.egSYMBOL, 20
  org.Cf.egUNIPROTEIN, 22

* utilities
  org.Cf.eg_dbconn, 23

AnnDbObj, 23

cat, 23
checkMAPCOUNTS, 15
count.mappedkeys, 15

dbconn, 23
dbConnect, 23
dbDisconnect, 23
dbfile, 23
dbGetQuery, 23
dbInfo, 23
dbschema, 23

mappedkeys, 15
org.Cf.egSYMBOL, 20
org.Cf.egSYMBOL2EG (org.Cf.egSYMBOL), 20
org.Cf.egUNIPROT, 22