

GO.db

June 9, 2011

GO.db

Bioconductor annotation data package

Description

Welcome to the GO.db annotation Package. The purpose of this package is to provide detailed information about the latest version of the Gene Ontologies. This package is updated biannually.

You can learn what objects this package supports with the following command:

```
ls ("package:GO.db")
```

Each of these objects has their own manual page detailing where relevant data was obtained along with some examples of how to use it.

Examples

```
ls ("package:GO.db")
```

GOBPANCESTOR

Annotation of GO Identifiers to their Biological Process Ancestors

Description

This data set describes associations between GO Biological Process (BP) terms and their ancestor BP terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO BP terms to all ancestor terms, where an ancestor term is a more general GO term that precedes the given GO term in the DAG (in other words, the parents, and all their parents, etc.).

Details

Each GO BP term is mapped to a vector of ancestor GO BP terms.

Biological process is defined as the broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions as defined by Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOBPANCESTOR)
# Remove GO IDs that do not have any ancestor
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the ancestor GO IDs for the first two elements of xx
  goids <- xx[1:2]
}
```

GOBPCHILDREN

Annotation of GO Identifiers to their Biological Process Children

Description

This data set describes associations between GO molecular function (BP) terms and their direct children BP terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO BP terms to all direct children terms, where a direct child term is a more specific GO term that is immediately preceded by the given GO term in the DAG.

Details

Each GO BP term is mapped to a vector of children GO BP terms.

Biological process is defined as the broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions as defined by Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOBPCHILDREN)
# Remove GO IDs that do not have any children
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the parent GO IDs for the first elements of xx
  goids <- xx[[1]]
  # Find out the GO terms for the first parent goid
  GOID(GOTERM[[goids[1]]])
  Term(GOTERM[[goids[1]]])
}
```

```
Synonym(GOTERM[[goids[1]]])
Secondary(GOTERM[[goids[1]]])
Definition(GOTERM[[goids[1]]])
Ontology(GOTERM[[goids[1]]])
}
```

GOBPOFFSPRING

Annotation of GO Identifiers to their Biological Process Offspring

Description

This data set describes associations between GO molecular function (BP) terms and their offspring BP terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO BP terms to all offspring terms, where an ancestor term is a more specific GO term that is preceded by the given GO term in the DAG (in other words, the children and all their children, etc.).

Details

Each GO BP term is mapped to a vector of offspring GO BP terms.

Biological process is defined as the broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions as defined by Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOBPOFFSPRING)
# Remove GO IDs that do not have any offspring
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the offspring GO IDs for the first two elents of xx
  goids <- xx[1:2]
}
```

Description

This data set describes associations between GO molecular function (BP) terms and their direct parent BP terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO BP terms to all direct parent terms, where a direct parent term is a more general GO term that immediately precedes the given GO term in the DAG.

Details

Each GO BP term is mapped to a named vector of GO BP terms. The name associated with the parent term will be either *isa*, *hasa* or *partof*, where *isa* indicates that the child term is a more specific version of the parent, and *hasa* and *partof* indicate that the child term is a part of the parent. For example, a telomere is part of a chromosome.

Biological process is defined as the broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOBPPARENTS)
# Remove GO IDs that do not have any parent
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the children GO IDs for the first elents of xx
  goids <- xx[[1]]
  # Find out the GO terms for the first parent goid
  GOID(GOTERM[[goids[1]]])
  Term(GOTERM[[goids[1]]])
  Synonym(GOTERM[[goids[1]]])
  Secondary(GOTERM[[goids[1]]])
  Definition(GOTERM[[goids[1]]])
  Ontology(GOTERM[[goids[1]]])
}
```

Description

This data set describes associations between GO molecular function (CC) terms and their ancestor CC terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO CC terms to all ancestor terms, where an ancestor term is a more general GO term that precedes the given GO term in the DAG (in other words, the parents, and all their parents, etc.).

Details

Each GO CC term is mapped to a vector of ancestor GO C terms.

Cellular component is defined as the subcellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and origin recognition complex as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOCCANCESTOR)
# Remove GO IDs that do not have any ancestor
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the ancestor GO IDs for the first two elements of xx
  goids <- xx[1:2]
}
```

Description

This data set describes associations between GO molecular function (CC) terms and their direct children CC terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO CC terms to all direct children terms, where a direct child term is a more specific GO term that is immediately preceded by the given GO term in the DAG.

Details

Each GO CC term is mapped to a vector of children GO CC terms.

Cellular component is defined as the subcellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and origin recognition complex as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOCCCHILDREN)
# Remove GO IDs that do not have any children
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  goids <- xx[[1]]
  # Find out the GO terms for the first parent goid
  GOID(GOTERM[[goids[1]]])
  Term(GOTERM[[goids[1]]])
  Synonym(GOTERM[[goids[1]]])
  Secondary(GOTERM[[goids[1]]])
  Definition(GOTERM[[goids[1]]])
  Ontology(GOTERM[[goids[1]]])
}
```

GOCCOFFSPRING

Annotation of GO Identifiers to their Cellular Component Offspring

Description

This is an R object mapping GO identifiers to all the GO identifiers of their offspring based on the cellular component (CC) GO terms. Offspring are defined as the direct or indirect children/grand children nodes of a given node for a GO identifier based on the directed acyclic graph defined by Gene Ontology Consortium. GO terms for offspring nodes are more specific than that of the parent.

Details

GO identifiers are keys and the corresponding offspring GO identifiers are values. Values are vectors of GO identifiers for all the offspring of the key GO identifier. GO identifiers that do not have any offspring node are assigned NA as the value.

Cellular component is defined as the subcellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and origin recognition complex as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOCCOFFSPRING)
# Remove GO identifiers that do not have any offspring
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the offspring GO identifiers for the first two elents of xx
  goidentifiers <- xx[1:2]
}
```

GOCCPARENTS

Annotation of GO Identifiers to their Cellular Component Parents

Description

This data set describes associations between GO molecular function (CC) terms and their direct parent CC terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO CC terms to all direct parent terms, where a direct parent term is a more general GO term that immediately precedes the given GO term in the DAG.

Details

Each GO CC term is mapped to a named vector of GO CC terms. The name associated with the parent term will be either *isa*, *hasa* or *partof*, where *isa* indicates that the child term is a more specific version of the parent, and *hasa* and *partof* indicate that the child term is a part of the parent. For example, a telomere is part of a chromosome.

Cellular component is defined as the subcellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and origin recognition complex as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOCCPARENTS)
# Remove GO IDs that do not have any parent
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  goids <- xx[[1]]
  # Find out the GO terms for the first parent go ID
  GOID(GOTERM[[goids[1]]])
}
```

```

    Term(GOTERM[[goids[1]]])
    Synonym(GOTERM[[goids[1]]])
    Secondary(GOTERM[[goids[1]]])
    Definition(GOTERM[[goids[1]]])
    Ontology(GOTERM[[goids[1]]])
}

```

GO_dbconn

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

```

GO_dbconn()
GO_dbfile()
GO_dbschema(file="", show.indices=FALSE)
GO_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

`GO_dbconn` returns a connection object to the package annotation DB. **IMPORTANT:** Don't call `dbDisconnect` on the connection object returned by `GO_dbconn` or you will break all the `AnnDbObj` objects defined in this package!

`GO_dbfile` returns the path (character string) to the package annotation DB (this is an SQLite file).

`GO_dbschema` prints the schema definition of the package annotation DB.

`GO_dbInfo` prints other information about the package annotation DB.

Value

`GO_dbconn`: a `DBIConnection` object representing an open connection to the package annotation DB.

`GO_dbfile`: a character string with the path to the package annotation DB.

`GO_dbschema`: none (invisible NULL).

`GO_dbInfo`: none (invisible NULL).

See Also

[dbGetQuery](#), [dbConnect](#), [dbconn](#), [dbfile](#), [dbschema](#), [dbInfo](#)

Examples

```
## Count the number of rows in the "go_term" table:
dbGetQuery(GO_dbconn(), "SELECT COUNT(*) FROM go_term")

## The connection object returned by GO_dbconn() was
## created with:
dbConnect(SQLite(), dbname=GO_dbfile(), cache_size=64000,
synchronous=0)

GO_dbschema()

GO_dbInfo()
```

GOMAPCOUNTS

Number of mapped keys for the maps in package GO.db

Description

GOMAPCOUNTS provides the "map count" (i.e. the count of mapped keys) for each map in package GO.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](#)

Examples

```
GOMAPCOUNTS
mapnames <- names(GOMAPCOUNTS)
GOMAPCOUNTS[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 GO.db
checkMAPCOUNTS("GO.db")
```

Description

This data set describes associations between GO molecular function (MF) terms and their ancestor MF terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO MF terms to all ancestor terms, where an ancestor term is a more general GO term that precedes the given GO term in the DAG (in other words, the parents, and all their parents, etc.).

Details

Each GO MF term is mapped to a vector of ancestor GO MF terms.

Molecular function is defined as the tasks performed by individual gene products; examples are transcription factor and DNA helicase as defined by Gene Ontology Consortium.

Mappings were based on data provided: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOMFANCESTOR)
# Remove GO identifiers that do not have any ancestor
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the ancestor GO identifiers for the first two elents of xx
  goids <- xx[1:2]
}
```

Description

This data set describes associations between GO molecular function (MF) terms and their direct children MF terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO MF terms to all direct children terms, where a direct child term is a more specific GO term that is immediately preceded by the given GO term in the DAG.

Details

Each GO MF term is mapped to a vector of children GO MF terms.

Molecular function is defined as the tasks performed by individual gene products; examples are transcription factor and DNA helicase as defined by Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/ With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOMFCHILDREN)
# Remove GO identifiers that do not have any children
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the children GO identifiers for the first elents of xx
  goids <- xx[[1]]
  # Find out the GO terms for the first parent goid
  GOID(GOTERM[[goids[1]]])
  Term(GOTERM[[goids[1]]])
  Synonym(GOTERM[[goids[1]]])
  Secondary(GOTERM[[goids[1]]])
  Definition(GOTERM[[goids[1]]])
  Ontology(GOTERM[[goids[1]]])
}
```

Description

This data set describes associations between GO molecular function (MF) terms and their offspring MF terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO MF terms to all offspring terms, where an ancestor term is a more specific GO term that is preceded by the given GO term in the DAG (in other words, the children and all their children, etc.).

Details

Each GO MF term is mapped to a vector of offspring GO MF terms.

Molecular function is defined as the tasks performed by individual gene products; examples are transcription factor and DNA helicase as defined by Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/ With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOMFOFFSPRING)
# Remove GO identifiers that do not have any offspring
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the offspring GO identifiers for the first two elents of xx
  goids <- xx[1:2]
}
```

GOMFPARENTS

Annotation of GO Identifiers to their Molecular Function Parents

Description

This data set describes associations between GO molecular function (MF) terms and their direct parent MF terms, based on the directed acyclic graph (DAG) defined by the Gene Ontology Consortium. The format is an R object mapping the GO MF terms to all direct parent terms, where a direct parent term is a more general GO term that immediately precedes the given GO term in the DAG.

Details

Each GO MF term is mapped to a named vector of GO MF terms. The name associated with the parent term will be either *isa*, *hasa* or *partof*, where *isa* indicates that the child term is a more specific version of the parent, and *hasa* and *partof* indicate that the child term is a part of the parent. For example, a telomere is part of a chromosome.

Molecular function is defined as the tasks performed by individual gene products; examples are transcription factor and DNA helicase as defined by the Gene Ontology Consortium.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.geneontology.org/> and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOMFPARENTS)
# Remove GO identifiers that do not have any parent
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Get the parent GO identifiers for the first elents of xx
  goids <- xx[[1]]
  # Find out the GO terms for the first parent goid
}
```

```

GOID(GOTERM[[goids[1]]])
Term(GOTERM[[goids[1]]])
Synonym(GOTERM[[goids[1]]])
Secondary(GOTERM[[goids[1]]])
Definition(GOTERM[[goids[1]]])
Ontology(GOTERM[[goids[1]]])
}

```

GOOBSOLETE

*Annotation of GO identifiers by terms defined by Gene Ontology***Description**

This is an R object mapping GO identifiers to the specific terms in defined by Gene Ontology Consortium and their definition are obsolete

Details

All the obsolete GO terms that are collected in this index will no longer exist in other mapping objects.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```

# Convert the object to a list
xx <- as.list(GOTERM)
if(length(xx) > 0){
  # Get the TERMS for the first elent of xx
  GOID(xx[[1]])
  Ontology(xx[[1]])
}

```

GOSYNONYM

*Map from GO synonyms to GO terms***Description**

GOSYNONYM is an R object that provides mapping from GO synonyms to GO terms

Details

[TODO: Put some details here]

Examples

```
x <- GOSYNONYM
sample(x, 3)
# GO ID "GO:0009435" has a lot of synonyms
GOTERM[["GO:0009435"]]
# GO ID "GO:0006736" is a synonym of GO ID "GO:0009435"
GOID(GOSYNONYM[["GO:0006736"]])
```

GOTERM

*Annotation of GO Identifiers to GO Terms***Description**

This data set gives mappings between GO identifiers and their respective terms.

Details

Each GO identifier is mapped to a `GOTerms` object that has 6 slots: `GOID`: GO Identifier Term: The term for that GO id `Synonym`: Synonymous terms `Secondary`: Secondary terms that have been merged into this term `Definition`: Further definition of the GO term `Ontology`: One of MF - molecular function, BP - biological process, or CC - cellular component

All the obsolete GO terms are under the nodes "obsolete molecular function" (GO:0008369), "obsolete cellular component" (GO id GO:0008370), and "obsolete biological process" (GO:0008371). Each of these GO identifiers has a group of GO identifiers as their direct children with GO terms that were defined by GO but are deprecated in the current build. These deprecated GO terms were appended by "(obsolete)" when the data package was built.

Mappings were based on data provided by: Gene Ontology <ftp://ftp.geneontology.org/pub/go/godatabase/archive/latest-lite/> With a date stamp from the source of: 20110312

References

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

Examples

```
# Convert the object to a list
xx <- as.list(GOTERM)
if(length(xx) > 0){
  # Get the TERMS for the first element of xx
  GOID(xx[[1]])
  Term(xx[[1]])
  Synonym(xx[[1]])
  Secondary(xx[[1]])
  Definition(xx[[1]])
  Ontology(xx[[1]])
}
```

Index

*Topic **datasets**

GO.db, 1
GO_dbconn, 8
GOBPANCESTOR, 1
GOBPCHILDREN, 2
GOBPOFFSPRING, 3
GOBPPARENTS, 4
GOCCANCESTOR, 5
GOCCCHILDREN, 5
GOCCOFFSPRING, 6
GOCCPARENTS, 7
GOMAPCOUNTS, 9
GOMFANCESTOR, 10
GOMFCHILDREN, 10
GOMFOFFSPRING, 11
GOMFPARENTS, 12
GOBSOLETE, 13
GOSYNONYM, 13
GOTERM, 14

*Topic **utilities**

GO_dbconn, 8

AnnDbObj, 8

cat, 8

checkMAPCOUNTS, 9

count.mappedkeys, 9

dbconn, 8

dbConnect, 8

dbDisconnect, 8

dbfile, 8

dbGetQuery, 8

dbInfo, 8

dbschema, 8

GO (*GO.db*), 1

GO.db, 1

GO_dbconn, 8

GO_dbfile (*GO_dbconn*), 8

GO_dbInfo (*GO_dbconn*), 8

GO_dbschema (*GO_dbconn*), 8

GOBPANCESTOR, 1

GOBPCHILDREN, 2

GOBPOFFSPRING, 3

GOBPPARENTS, 4

GOCCANCESTOR, 5

GOCCCHILDREN, 5

GOCCOFFSPRING, 6

GOCCPARENTS, 7

GOMAPCOUNTS, 9

GOMFANCESTOR, 10

GOMFCHILDREN, 10

GOMFOFFSPRING, 11

GOMFPARENTS, 12

GOBSOLETE, 13

GOSYNONYM, 13

GOTERM, 14

mappedkeys, 9