Package 'PSICQUIC'

October 8, 2014

Type Package

Title Protemics Standard Initiative Common QUery InterfaCe

Version 1.2.1

Date 2014-07-10

Author Paul Shannon

Maintainer Paul Shannon<pshannon@fhcrc.org>

Depends R (>= 2.15.0), methods, IRanges, biomaRt, BiocGenerics

Suggests org.Hs.eg.db

Imports RCurl

Description PSICQUIC is a project within the HUPO Proteomics Standard
Initiative (HUPO-PSI). It standardises programmatic access to molecular interaction databases.

License Apache License 2.0

biocViews DataImport, GraphAndNetwork, ThirdPartyClient

R topics documented:

PSICQUIC-package   *Protemics Standard Initiative Common QUery InterfaCe*

Description

This package provides a simple functional interface to the rich resources of the *PSICQUIC*, the HUPO Proteomics Standard Initiative (HUPO-PSI) project to offer programmatic access to molecular interaction databases.

<http://code.google.com/p/psicquic/>

There are two classes in the package: PSICQUIC provides the query interface via the *interactions* method, whose results are returned in a data.frame; and IDMapper adds gene symbol and entrez geneIDs to the PSICQUIC results data.frame, for easier interpretation and subsequent use.

Author(s)

Paul Shannon<pshannon@fhcrc.org>

References

http://code.google.com/p/psicquic/ Aranda, Bruno, et al. "PSICQUIC and PSISCORE: accessing and scoring molecular interactions." Nature methods 8.7 (2011): 528-529.

See Also

providers, interactions, rawQuery, addGeneInfo, IDMapper, interactionTypes, detectionMethods, speciesIDs

---

IDMapper-class   *IDMapper*

Description

PSICQUIC-compliant services typically return verbose protein identifiers for the interactions they report. However, we often wish to know the entrez geneID or HUGO symbol of the gene associated with the interacting protein. The IDMapper class, using biomaRt, provides that mapping. The addGeneInfo method adds four columns to the data.frame it is passed: gene symbol and geneID for both interacting proteins.

Usage

```
IDMapper(species)
## S4 method for signature IDMapper
addGeneInfo(object, tbl.interactions)
```

Arguments

species
:   a character string, as specified by the NCBI. Only "9606" (Homo sapiens) is currently supported.

object
:   a PSICQUIC object.

tbl.interactions

>   A data.frame such as one returned by the interactions method. Columns named A and B are required, and are presumed to contain PSICQUIC provider identifers in a few formats. Most, but not all identifiers returned by all PSIC-QUIC providers are recognized.

Value

A copy of the data.frame passed in, with 4 columns added: A.sym, B.sym, A.geneID, B.geneID.

Constructor

IDMapper(species): only human ("9606") is currently supported.

Methods

addGeneInfo(x,tbl): add geneID and symbol for the A and B columns in the data.frame "tbl", which is expected to be a result returned by the interactions method.

Author(s)

Paul Shannon

See Also

providers, interactions, rawQuery, addGeneInfo, IDMapper, interactionTypes, detectionMethods, speciesIDs

Examples

```
mapper <- IDMapper("9606")
tbl <- data.frame(A="entrez gene/locuslink:238|BIOGRID:106739",
                  B="entrez gene/locuslink:3718|BIOGRID:109921",
                  stringsAsFactors=FALSE)
tbl.withGeneInfo <- addGeneInfo(mapper, tbl)
```

---

interactions   *interactions*

Description

This is the primary interface to the PSICQUIC services, in which approximately thirty interaction providers return annotated molecular (often protein-protein) interactions in response to a standard query language ("MIQL" – the molecular interaction query language). This method is thus an R interface to MIQL, with several conveniences added.

In the most typical and basic use, one specifies a species and a gene symbol, and PSICQUIC returns a data.frame listing all of the interactions for that gene (or the protein it encodes) from all PSICQUIC providers, of all interaction types and all detection methods.

More focused queries are easily accomplished: all arguments (execpt for the PSICQUIC object the method dispatches on) are optional, and each specified argument limits the search space: to one or more genes, a publication identifier (typically a pubmed id), interaction types, a detection methods, and/or one or more species.

In principle you could obtain all interactions, of all types, from all providers, by using default values for all of the search-constraining aruguments. But this would be very unwise! Highly specific or exploratory queries of modest scope are encouraged, as are bulk downloads using *ftp* directed at the providers' web sites for acquiring large interaction datasets.

Usage

```
## S4 method for signature PSICQUIC
interactions(object,
                          id=NA,
                          species=NA,
                          speciesExclusive=TRUE,
                          type=NA,
                          provider=NA,
                          detectionMethod=NA,
                          publicationID=NA,
                          quiet=TRUE)
```

Arguments

object   a PSICQUIC object.

id   one or more character strings. These are most often gene symbols (e.g., "RAD17"), to which most of the PSICQUIC providers respond, translating these into the native identifiers of their data store, which is typically a protein identifier of one sort or another. You may also provide such a protein identifier directly (e.g., uniprot "O75943", or refseq "NP_002684"). Protein identifiers typically return fewer interactions, presumably because providers do not translate them into their native identifiers, and only a subset of the providers find matches.

If multiple ids are supplied, then only interactions which occur between any two of the ids is returned. If only one id is supplied, all interactions are reported in which that id participates.

species   One or more `character` strings. Use an NCBI taxon code, for which the `speciesIds` method is useful. Note the next argument.

speciesExclusive

     a `logical` variable, default TRUE. If set to FALSE than interactions will be included with proteins from species other than those explicitly named. Infection and transgenic experiments are examples of cross-species protein-protein interactions. We anticipate that the most common query seeks interactions among proteins in the same cell of the same organism. For this reason, this argument defaults to TRUE.

type   a `list` of `character` strings. See the method `interactionTypes`.

provider   a `list` of `character` strings. See the method `providers` for the currently available set. This argument defaults to NA, and all currently live providers are queried.

detectionMethod

     a `list` of `character` strings. See the method `detectionMethods`

publicationID   a `list` of `character` strings. Usually a pubmedID, but sometimes an OMIM reference number.

quiet   a `logical` of `character` strings, default TRUE. Set to FALSE if you wish to trace the exection of your query against all providers, and explore the details of the REST API.

Value

A data.frame with 16 columns, and one row for every interaction, described as an annotated A/B relationship, with self-describing column names.

Author(s)

Paul Shannon

See Also

providers, rawQuery, addGeneInfo IDMapper, interactionTypes, detectionMethods, speciesIDs

Examples

```
psicquic <- PSICQUIC()
providers(psicquic)
    # query all providers for all known Myc interactions
tbl.1 <- interactions(psicquic, id="Myc", species="9606",
                    provider=providers(psicquic)[1:3])

    # all Myc/EP400 interactions known to BioGrid
    # make sure that BioGrid is currently available

if("BioGrid" %in% providers(psicquic))
   tbl.2 <- interactions(psicquic, c("Myc", "EP400"), species="9606",
                        provider="BioGrid")
```

```
      # or those between Myc and any other molecule, detected by
      # the "pull down" proteomics technique and judged to be
      # a "direct interaction"

    tbl.3 <- interactions(psicquic, "Myc", species="9606",
                          detectionMethod="pull down",
                          type="direct interaction")
```

---

providers   *providers*

Description

The central PSICQUIC server has only one purpose: to return a list of, and metadata about, the currently "live" PSICQUIC-compliant MI-QL websites. A list of these providers by the PSICQUIC object you create, and is returned by this method.

Usage

```
## S4 method for signature PSICQUIC
providers(object)
```

Arguments

object   a PSICQUIC object.

Value

A list of character strings, short readable names for the providers, e.g., "BioGrid", "IntAct".

Author(s)

Paul Shannon

See Also

interactions, rawQuery, addGeneInfo, IDMapper, interactionTypes, detectionMethods, speciesIDs

Examples

```
psicquic <- PSICQUIC()
providers(psicquic)
```

PSICQUIC-class   *PSICQUIC*

Description

PSICQUIC is an effort from the HUPO Proteomics Standard Initiative (HUPO-PSI) to standardise programmatic access to molecular interaction databases. The Bioconductor PSICQUIC package provides a traditional R interface layered on top of the PSICQUIC REST interface. Gene symbols are most commonly used in queries; interactions are returned in a data.frame, characterized by interaction type, detection method, and publication references. Confidence scores are sometimes avaialable. Queries may be constrained by many of these same attributes, i.e., interaction type, detection method, species, publication identifier, and source database.

Details

There are two operational differences between the native PSIQUIC REST interface and that offered here via the interactions method:

species exclusivity: The REST interface requires only that one participant in an interaction be from the named species. By default, we require that both participants are from the named species. This can be controlled by the speciesExclusive logical argument to the interactions method.

number of molecules (or identifiers): the REST interface permits only zero, one or two gene or protein identifiers per query. We allow any number, zero or more and, when the number is greater than or equal to two, the interactions returned are only those in which any two of those identifiers participate. If you want all interactions which include any of a list of identifiers, and you don't care to control for their partners, you can accomplish this by issuing successive single-identifer interaction queries.

Constructor

PSICQUIC: contacts the central PSICQUIC web server, discovers currently functioning servers, returns an object used in the methods below.

Methods

providers(x): lists the short names of the data providers

interactions(x,id, species, speciesExclusive, type, provider,detectionMethod,   publicationID, qu
retrieves all interactions matching the specified pattern.

rawQuery(x, provider, rawArgs): query terms in native PSICQUIC REST style

show(x): displays current providers and related data

Functions

detectionMethods(): your web browser will display the PSI-MI ontology for detection methods

interactionTypes(): your web browser will display the PSI-MI ontology for molecular interaction types

Author(s)

Paul Shannon

See Also

providers, interactions, rawQuery, addGeneInfo, IDMapper, interactionTypes, detectionMethods,
speciesIDs

Examples

```
psicquic <- PSICQUIC()
  # obtain the list of two dozen (or so) currently live
  # PSICQUIC-compliant data providers
providers(psicquic)
  # a minimal call: get all interactions with MAP3K3, of all types,
  # from all providers.  a data.frame is returned
tbl.0 <- interactions(psicquic, "MAP3K3", species="9606")

  # build a contingency table, sort it, and see
  # what kinds of interactions were returned, obtained
  # by what detection methods.
  # "-" is used when the provider does not specify a value.
  # you will see a wide range of specificity, in detection method,
  # interaction type, and number of interactions found.

xtab <- with(tbl.0, as.data.frame(
              table(type, detectionMethod,  provider)))
xtab <- subset(xtab, Freq > 0) # [order(xtab$Freq, decreasing=TRUE),]
xtab <- xtab[order(xtab$Freq, decreasing=TRUE),]

  # what interactors were returned?  the IDMapper class in this
  # package converts many PSICQUIC providers protein identifiers to
  # entrez geneIDs and HUGO gene symbols, via remote calls to
  # biomaRt:

idMapper <- IDMapper("9606")
tbl.0g <- addGeneInfo(idMapper, tbl.0)
with(tbl.0g, head(unique(c(A.sym, B.sym))))

   # we see that MAP2K5 is the most frequently mentioned interacator:

xtab.sym <- with(tbl.0g, table(c(A.sym, B.sym)))
head(sort(xtab.sym, decreasing=TRUE))

  # PSIQUIC uses well-devloped ontologies -- controlled vocabularies --
  # which are currently best viewed in a web browser.
  # we provide two convenience functions which will display these
  # hierarchically defined vocabularies:

# interactionTypes()
# detectionMethods()
```

```
        # NCBI curates taxonomy codes, such as "9606" for Homo sapiens.
        # you can find these codes by using this method, which will
        # drive your browser to the appropriate NCBI web page.

    # speciesIds()

        # use terms from these vocabularies to retrieve interaction
        # information for these two proteins.  note that both of
        # these terms are mid-level in their respective hierarchies
        # and will likely retrieve more specific nested terms

    tbl.2 <- interactions(psicquic, id=c("MAP3K3", "MAP2K5"),
                         species="9606",
                         type="physical association",
                         detectionMethod="affinity chromatography technology")

        # add gene IDs and symbols

    tbl.2g <- addGeneInfo(idMapper, tbl.2)

        # how many publications lie behind these interactions?
    tbl.2g[, c("A.sym", "B.sym", "detectionMethod", "firstAuthor")]

        # the package also provides a convenience method for submitting
        # queries in native MIQL (Molecular Interaction Query Language).
        # the language is defined here:
        # http://code.google.com/p/psicquic/wiki/MiqlReference27

    if("BioGrid" %in% providers(psicquic)){
        tbl.3 <- rawQuery(psicquic, "BioGrid", "identifier:ALK AND species:9606")
            # what publications?
        table(tbl.3$V8)
        }
```

---

rawQuery   *rawQuery*

Description

For exploratory situations, or by personal preference, one can bypass the interactions API (which we expect will be used by most people most of the time). The rawQuery method allows you to craft your own MIQL (molecular interaction query language) query but without needing to obtain per-provider URLs. The sixteen columns of data are returned in a data.frame without column titles.

Usage

```
## S4 method for signature PSICQUIC
rawQuery(object, provider, rawArgs)
```

Arguments

`object`	a PSICQUIC object.
`provider`	a character object, one of the approximately two dozen PSCIQUIC-compliant services.
`rawArgs`	A character string which follows MIQL syntax, which is defined here:
	http://code.google.com/p/psicquic/wiki/MiqlReference27
	The encoding of special characters (spaces and parentheses in particular) is handled by this method, so there is no need to replace, for instance, spaces or parens with their url-safe hex codes.

Value

A data.frame with 15 columns, and one row for every interaction, described as an annotated A/B relationship, with self-describing column names.

Author(s)

Paul Shannon

See Also

providers, interactions, addGeneInfo, IDMapper, interactionTypes, detectionMethods, speciesIDs

Examples

```
psicquic <- PSICQUIC()
providers <- providers(psicquic)

   # query BioGrid for all known ALK interactions

if("BioGrid" %in% providers){
    tbl.1 <- rawQuery(psicquic, providers[1], "identifier:ALK AND species:9606")
      # what publications?
    table(tbl.1$V8)
    }

  # query the iRefIndex provider for interactions between
  # two specified ids.  then further contrain by publication.

if("iRefIndex" %in% providers){
   rawArgs.1 <- "identifier:(ALK AND MAP3K3) AND species:9606"
   tbl.2 <- rawQuery(psicquic, "iRefIndex", rawArgs.1)
   rawArgs.2 <- paste(rawArgs.1, " AND pubid:(15657099 OR 14743216)", sep="")
   tbl.3 <- rawQuery(psicquic, "iRefIndex", rawArgs.2)
   }
```

speciesIDs   *Web Browser Functions*

Description

PSICQUIC uses standard codes to identify species, detection methods, and interaction types. In the absence of easy-to-use data files which we could store with, and provide from this package, we instead offer methods which drive your web browser to pages which offer the relevant information.

Usage

```
interactionTypes()
detectionMethods()
speciesIDs()
```

Details

interactionTypes will drive your web browser to a EMBL-EBI Ontology Lookup Service devoted to protein-protein interaction types.

detectionMethods will drive your web browser to a EMBL-EBI Ontology Lookup Service devoted to protein interaction detection methods.

speciesIDs helps in finding the NCBI taxon ids which are used to identify species (e.g., "9606" for Homo sapiens). This method drives your web browser to the NCBI Taxonomy Browser (the NCBI Taxonomy Homepage), where you can use the search facilities to obtain the taxon id for your organism/s of interest.

Value

None

Author(s)

Paul Shannon

See Also

providers, interactions, rawQuery, addGeneInfo, IDMapper,

Examples

```
## Not run:
  psicquic <- PSICQUIC()
  speciesIDs()
  interactionTypes()
  detectionMethods()

## End(Not run)
```

Index