keggorthology: the KEGG orthology as graph

VJ Carey

May 1, 2024

Contents

1 Introduction 1
2 KOgraph 1
3 Application to gene filtering 3
4 Infrastructure considerations 4
5 Session info 4

1 Introduction

KEGG is the Kyoto Encyclopedia of Genes and Genomes. An important product of the KEGG group is a catalog of pathways. The KEGG Orthology (KO) organizes the pathways into a conceptual hierarchy. This package encodes the hierarchy as a graph, and provides some support for deriving sets of array feature identifiers from the hierarchy.

2 KOgraph

> library(keggorthology)
> library(graph)
> data(KOgraph)
> KOgraph

A graphNEL graph with directed edges
Number of Nodes = 358
Number of Edges = 357

> nodes(KOgraph)[1:5]
Graph operations can be used to explore the orthology. For example, the context of the PPAR signaling pathway is found as follows:

```r
> library(RBGL)
> sp.between(KOgraph, nodes(KOgraph)[1], "PPAR signaling pathway")
```

```
KO.Feb10root:PPAR signaling pathway
KO.Feb10root:PPAR signaling pathway
KO.Feb10root:PPAR signaling pathway

KO.Feb10root->Organismal Systems
1
Organismal Systems->Endocrine System
1
Endocrine System->PPAR signaling pathway
1
```

Fixed-length identifiers are used to label pathways. These are available as the 'tag' nodeData attribute.

```r
> nodeData(KOgraph,"tag")[1:5]
```
3 Application to gene filtering

Several functions are available for retrieving relevant information from the orthology. If you know a substring of the pathway name of interest, you can obtain the numerical tag(s).

> getKOtags("insulin")

Insulin signaling pathway
   "04910"
We can get probe set identifiers corresponding to a term. The default chip annotation package used is hgu95av2.db.

```r
> library(hgu95av2.db)
> mp = getKOprobes("Methionine")
> library(ALL)
> data(ALL)
> ALL[mp,]
```

ExpressionSet (storageMode: lockedEnvironment)
assayData: 30 features, 128 samples
  element names: exprs
protocolData: none
phenoData
  sampleNames: 01005 01010 ... LAL4 (128 total)
  varLabels: cod diagnosis ... date last seen (21 total)
  varMetadata: labelDescription
featureData: none
experimentData: use 'experimentData(object)'
  pubMedIds: 14684422 16243790
Annotation: hgu95av2

4 Infrastructure considerations

Based on keggorthology read of KEGG orthology, March 2 2010. Specifically, we run wget on [ftp://ftp.genome.jp/pub/kegg/brite/ko/ko00001.keg](ftp://ftp.genome.jp/pub/kegg/brite/ko/ko00001.keg) and use parsing and modeling code given in inst/keggHTML to generate a data frame respecting the hierarchy, and then keggDF2graph function in keggorthology package to construct the graph.

5 Session info

```r
> sessionInfo()
```

R version 4.4.0 beta (2024-04-15 r86425)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 22.04.4 LTS

Matrix products: default
BLAS: /home/biocbuild/bbs-3.19-bioc/R/lib/libRblas.so
LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0

locale: