sol genomics network

Using R for GO Terms Analysis

sol genomics network

Boyce Thompson Institute for Plant Research
Tower Road
Ithaca, New York 14853-1801
U.S.A.

by

Aureliano Bombarely Gomez

Boyce Thompson Institute for Plant Research

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Enrinchment Analysis: TopGO
    - 4.3. GO profiles: GOProfiles
    - 4.4. Gene Similarities: GOSim
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Enrinchment Analysis: TopGO
    - 4.3. GO profiles: GOProfiles
    - 4.4. Gene Similarities: GOSim
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

1. What are GO Terms ?

Gene ontologies:

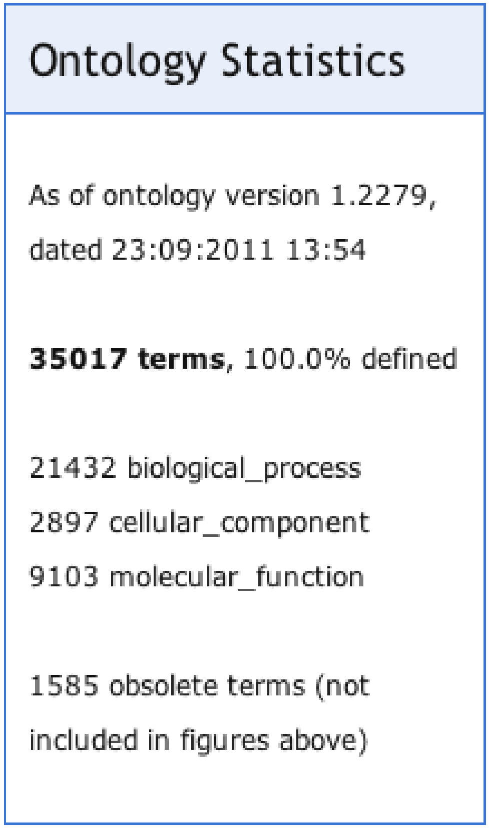
Structured controlled vocabularies (ontologies) that describe **gene products** in terms of their associated

biological processes,

cellular components and

molecular functions

in a species-independent manner



Ontology Statistics

As of ontology version 1.2279, dated 23:09:2011 13:54

35017 terms, 100.0% defined

21432 biological_process
2897 cellular_component
9103 molecular_function

1585 obsolete terms (not included in figures above)

http://www.geneontology.org/GO.doc.shtml

1. What are GO Terms ?

Biological processes,

Recognized series of events or molecular functions. A process is a collection of molecular events with a defined beginning and end.

Cellular components,

Describes locations, at the levels of subcellular structures and macromolecular complexes.

Molecular functions

Describes activities, such as catalytic or binding activities, that occur at the molecular level.

1. What are GO Terms ?

Gene ontologies:

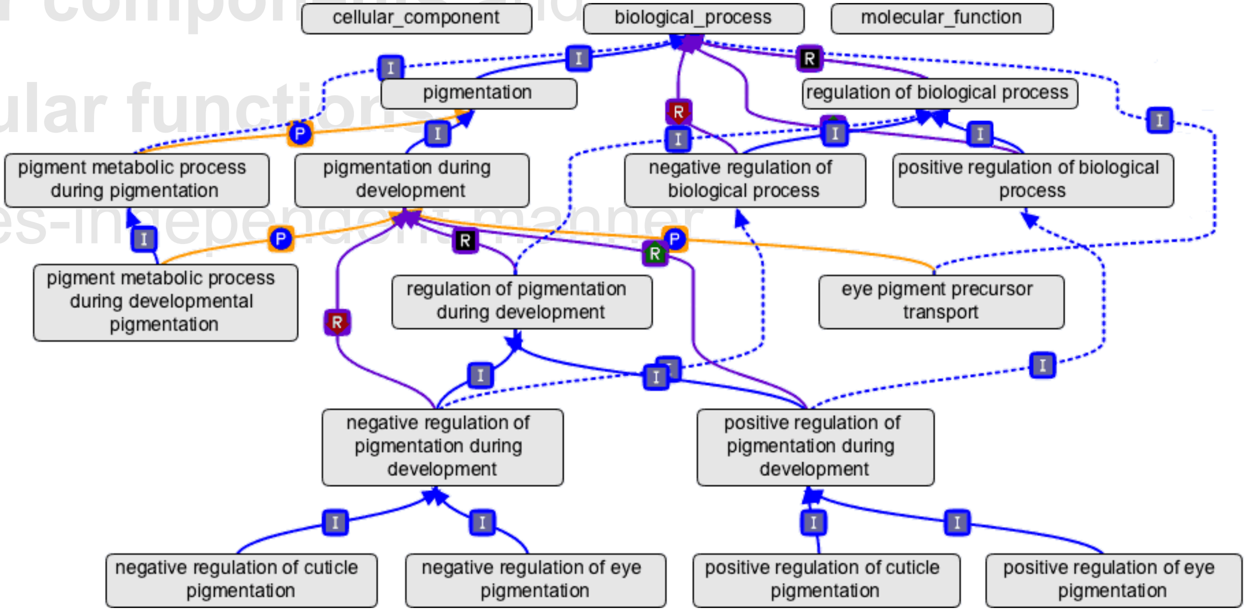
Structured controlled → **relationships between the terms**

(ontologies) that describe **gene products** in terms of their associated **biological processes**, **cellular components and molecular functions** in a species-independent manner

GO can be described in terms of a **acyclic graph**, where:

each **GO term** is a **node**, and the

relationships between the terms are **arcs** between the nodes



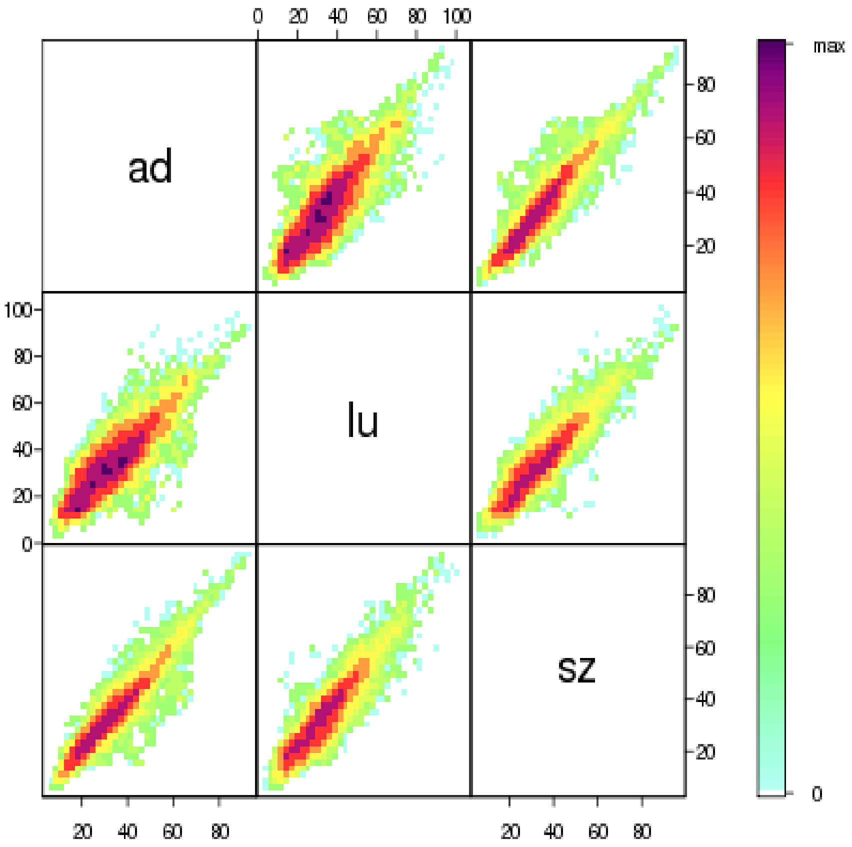
Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
   - 4.1. Basics: GO.db
   - 4.2. Gene Enrinchment Analysis: TopGO
   - 4.3. GO profiles: GOProfiles
   - 4.4. Gene Similarities: GOSim
5. Example.
   - 5.1. Comparing two arabidopsis chromosomes.

2. What is R ?

R is a **language** and **environment** for statistical computing and graphics..

```
## load required packages

library(IDPmisc)
library(SwissAir) # data for the example

## prepare the data

Ox <- AirQual[,c("ad.O3","lu.O3","sz.O3")] +
  AirQual[,c("ad.NOx","lu.NOx","sz.NOx")] -
  AirQual[,c("ad.NO","lu.NO","sz.NO")]
names(Ox) <- c("ad","lu","sz")

## draw graph

ipairs(Ox, ztransf = function(x){x[x<1] <- 1; log2(x)})
```



2. What is R ?

WEB:

OFICIAL WEB: http://www.r-project.org/index.html

QUICK-R: http://www.statmethods.net/index.html

BOOKS:

Introductory Statistics with R (Statistics and Computing), P. Dalgaard *[available as manual at R project web]*

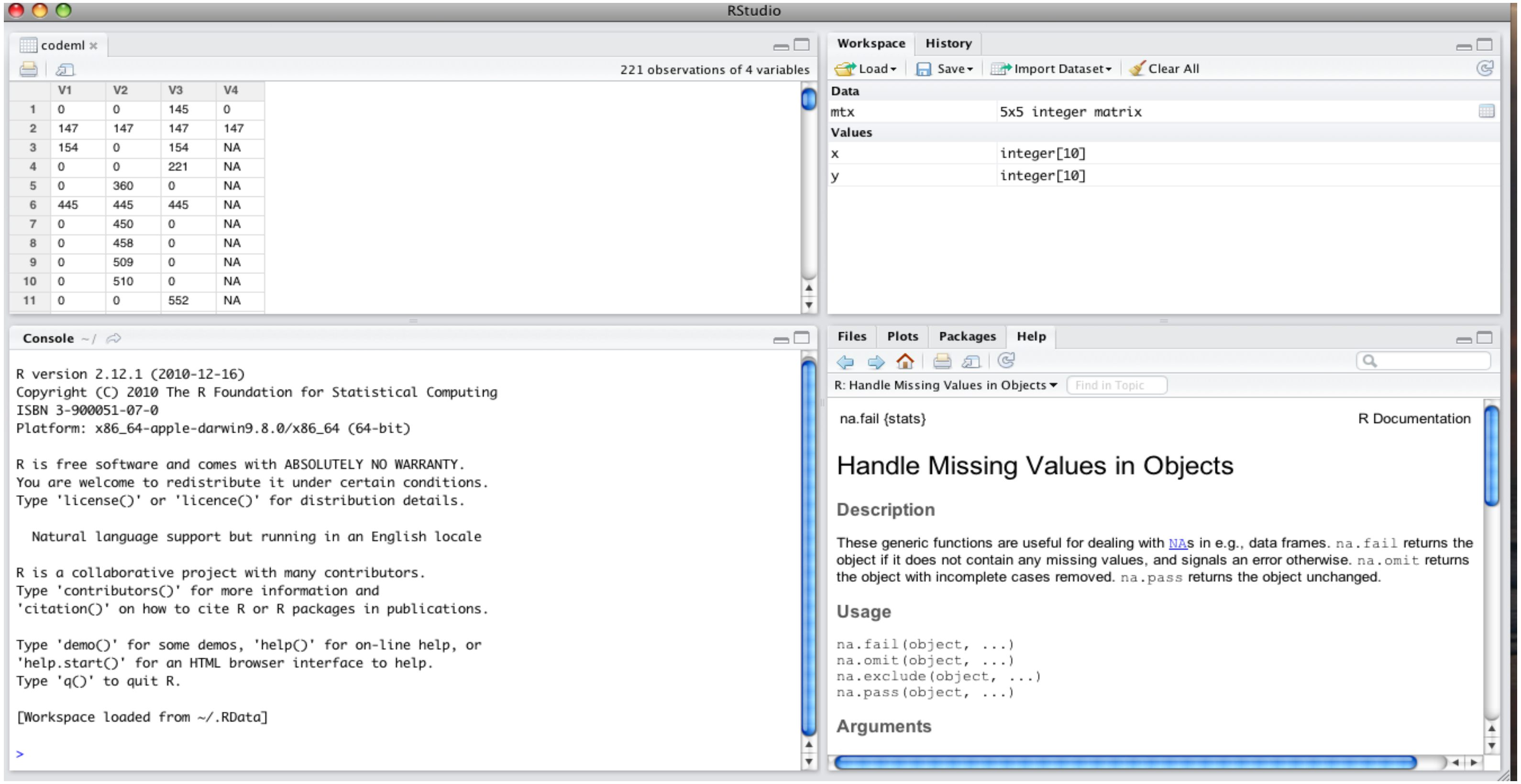
The R Book, MJ. Crawley

R itself: help() and example()

2. What is R ?

INTEGRATED DEVELOPMENT ENVIRONMENT (IDE):

R-Studio: http://rstudio.org/



221 observations of 4 variables

	V1	V2	V3	V4
1	0	0	145	0
2	147	147	147	147
3	154	0	154	NA
4	0	0	221	NA
5	0	360	0	NA
6	445	445	445	NA
7	0	450	0	NA
8	0	458	0	NA
9	0	509	0	NA
10	0	510	0	NA
11	0	0	552	NA

Data

mtx	5x5 integer matrix

Values

x	integer[10]
y	integer[10]

```
R version 2.12.1 (2010-12-16)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

>
```

na.fail {stats} R Documentation

Handle Missing Values in Objects

Description

These generic functions are useful for dealing with NAs in e.g., data frames. `na.fail` returns the object if it does not contain any missing values, and signals an error otherwise. `na.omit` returns the object with incomplete cases removed. `na.pass` returns the object unchanged.

Usage

```
na.fail(object, ...)
na.omit(object, ...)
na.exclude(object, ...)
na.pass(object, ...)
```

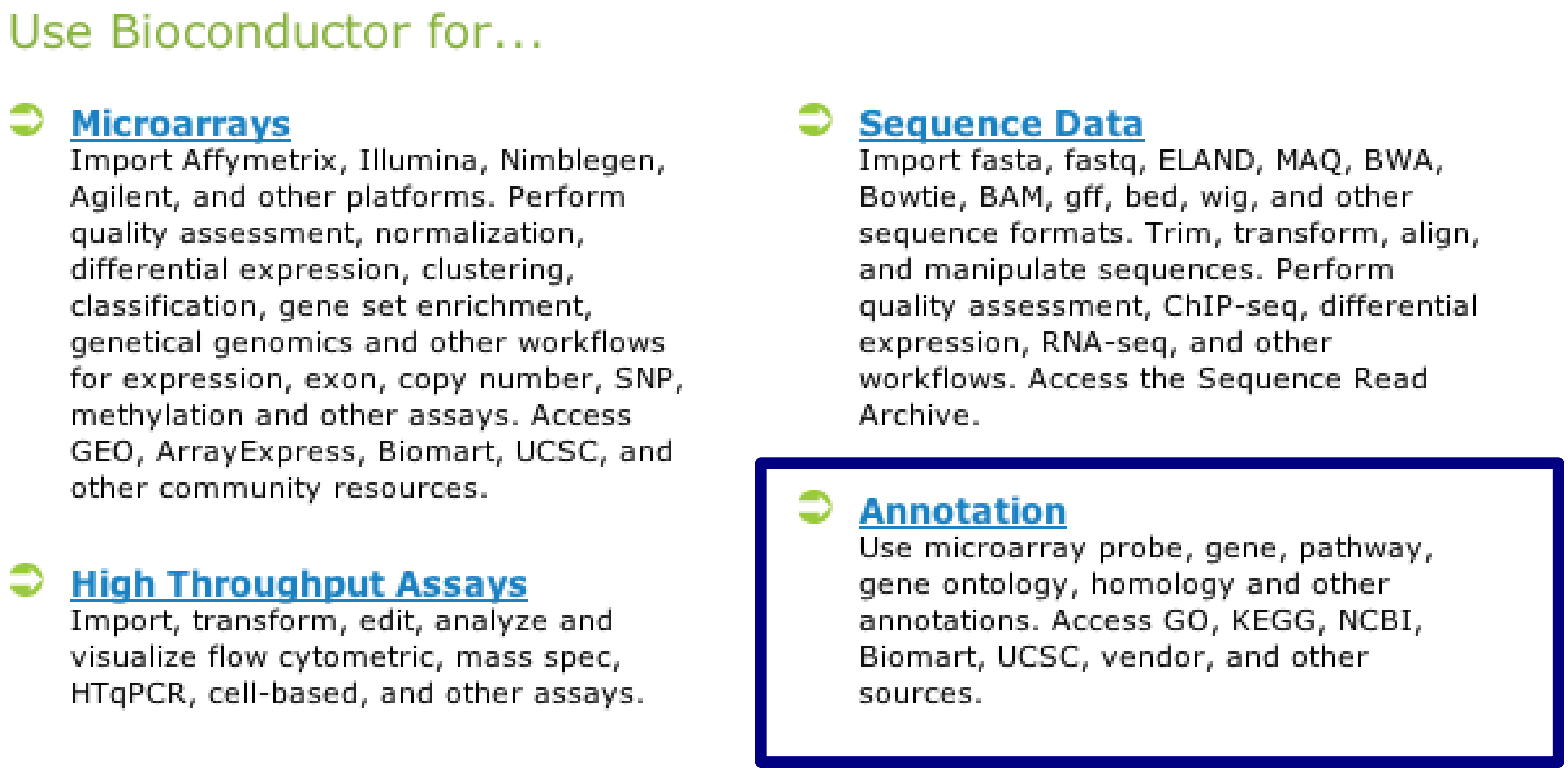
Arguments

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Enrinchment Analysis: TopGO
    - 4.3. GO profiles: GOProfiles
    - 4.4. Gene Similarities: GOSim
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

3. What is Bioconductor ?

Bioconductor is an open source, open development software project to provide **tools** for the analysis and comprehension of **high-throughput genomic data**. It is based primarily on the **R programming language**.



Use Bioconductor for…

Microarrays
Import Affymetrix, Illumina, Nimblegen, Agilent, and other platforms. Perform quality assessment, normalization, differential expression, clustering, classification, gene set enrichment, genetical genomics and other workflows for expression, exon, copy number, SNP, methylation and other assays. Access GEO, ArrayExpress, Biomart, UCSC, and other community resources.

High Throughput Assays
Import, transform, edit, analyze and visualize flow cytometric, mass spec, HTqPCR, cell-based, and other assays.

Sequence Data
Import fasta, fastq, ELAND, MAQ, BWA, Bowtie, BAM, gff, bed, wig, and other sequence formats. Trim, transform, align, and manipulate sequences. Perform quality assessment, ChIP-seq, differential expression, RNA-seq, and other workflows. Access the Sequence Read Archive.

Annotation
Use microarray probe, gene, pathway, gene ontology, homology and other annotations. Access GO, KEGG, NCBI, Biomart, UCSC, vendor, and other sources.

3. What is Bioconductor ?

WEB:

OFICIAL WEB: http://www.bioconductor.org/

BOOKS:

Bioinformatics and Computational Biology Solutions Using R and Bioconductor, 2005, Gentleman R. et al. (Springer)

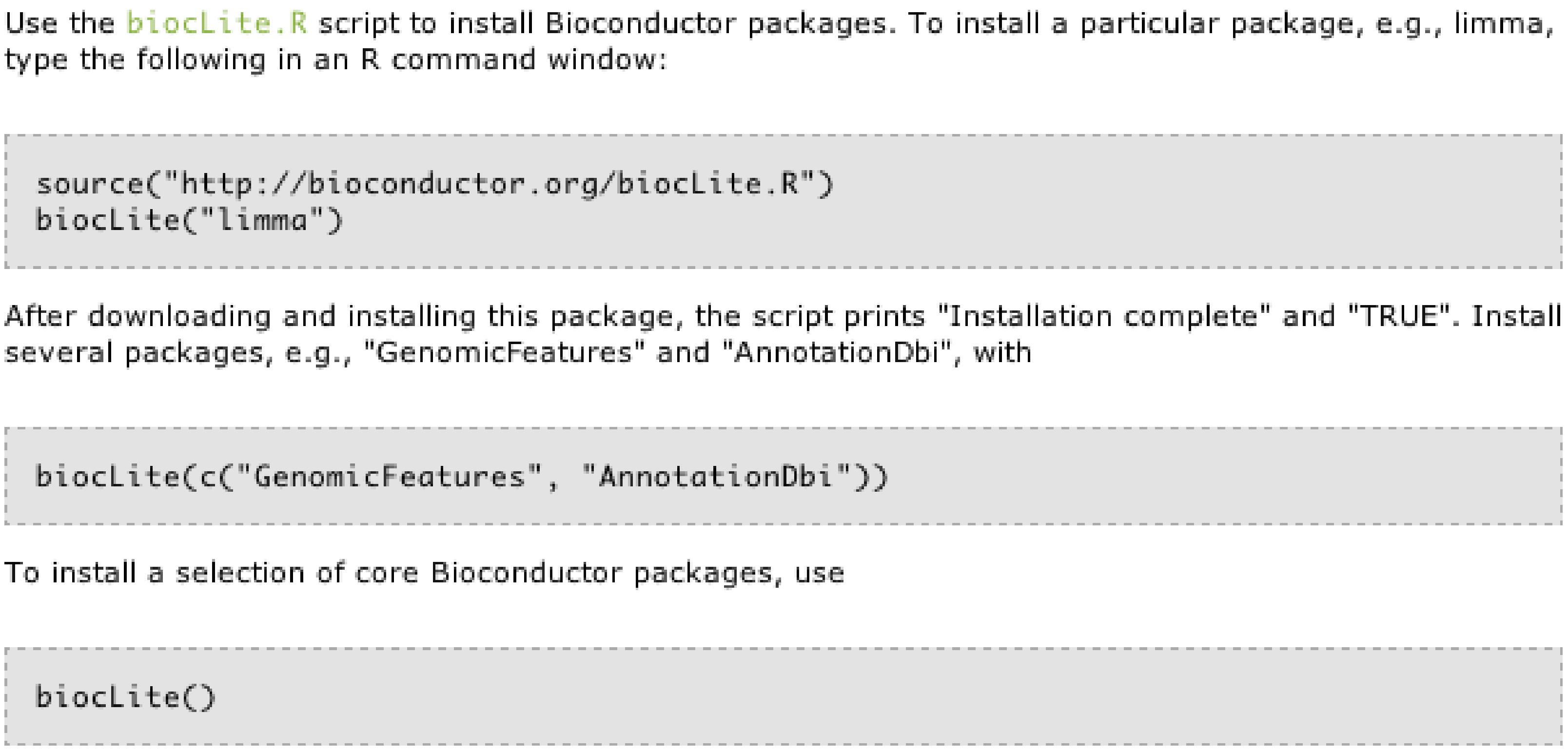
R Programming for Bioinformatics, 2008, Gentleman R. (CRC Press)

MAILING LIST:

http://www.bioconductor.org/help/mailing-list/

3. What is Bioconductor ?

Installation of Bioconductor:



Use the biocLite.R script to install Bioconductor packages. To install a particular package, e.g., limma, type the following in an R command window:

```
source("http://bioconductor.org/biocLite.R")
biocLite("limma")
```

After downloading and installing this package, the script prints "Installation complete" and "TRUE". Install several packages, e.g., "GenomicFeatures" and "AnnotationDbi", with

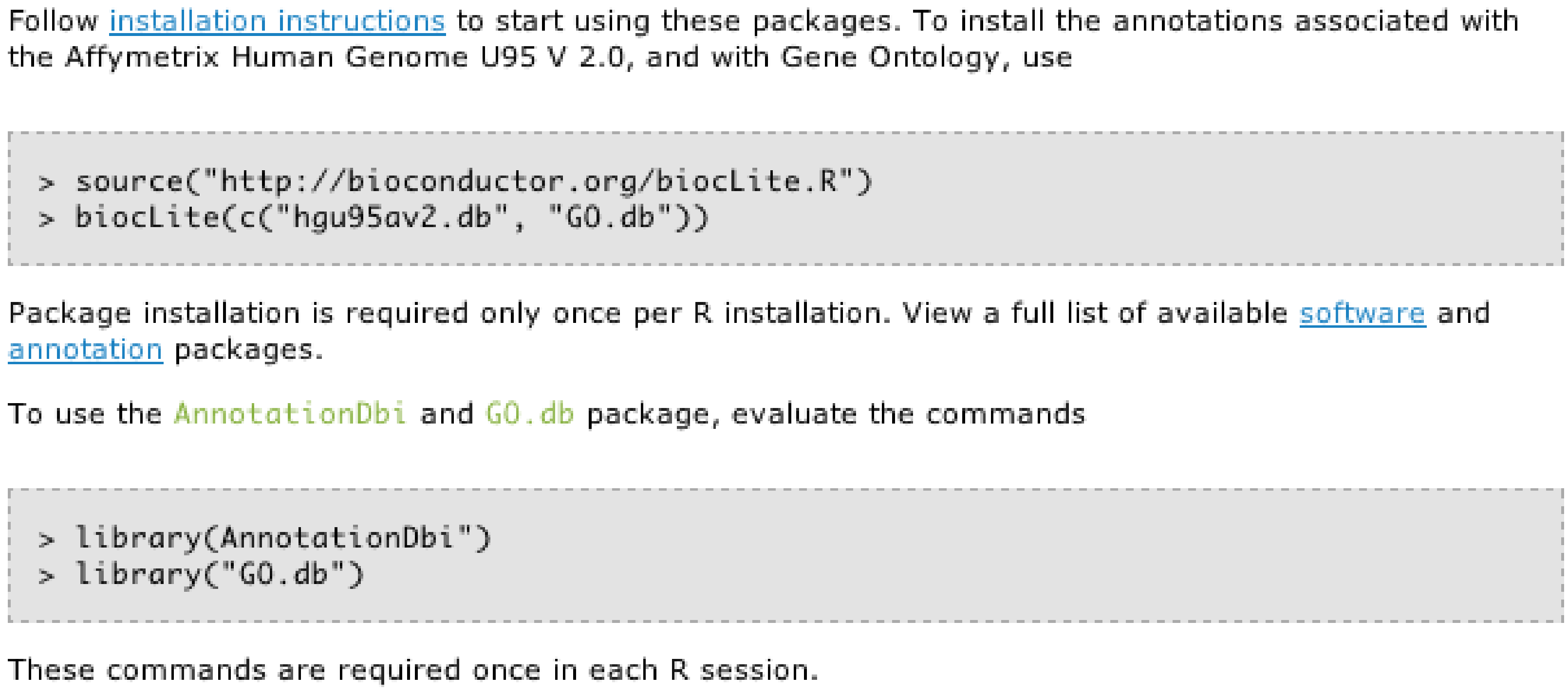
```
biocLite(c("GenomicFeatures", "AnnotationDbi"))
```

To install a selection of core Bioconductor packages, use

```
biocLite()
```

3. What is Bioconductor ?

Use of Bioconductor:



Follow installation instructions to start using these packages. To install the annotations associated with the Affymetrix Human Genome U95 V 2.0, and with Gene Ontology, use

```
> source("http://bioconductor.org/biocLite.R")
> biocLite(c("hgu95av2.db", "GO.db"))
```

Package installation is required only once per R installation. View a full list of available software and annotation packages.

To use the `AnnotationDbi` and `GO.db` package, evaluate the commands

```
> library(AnnotationDbi")
> library("GO.db")
```

These commands are required once in each R session.

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Similarities: GOSim
    - 4.3. GO profiles: GOProfiles
    - 4.4. Gene Enrinchment Analysis:TopGO
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

4. Bioconductor modules for GOterms

sol genomics network

Bioconductor Packages for GO Terms:

Bioconductor version 2.9 (Development)

- ▶ AnnotationData (604)
- ▶ ExperimentData (85)
- ▼ Software (499)
  - ▼ Annotation (66)
    - GO (15)
    - Pathways (23)
    - ProprietaryPlatforms (1)
    - ReportWriting (14)
  - ▶ AssayDomains (190)
  - ▶ AssayTechnologies (310)
  - ▶ Bioinformatics (290)
  - ▶ BiologicalDomains (50)
  - ▶ Infrastructure (201)

Packages

Software > Annotation > GO

- annaffy
- annotate
- Category
- clusterProfiler
- gage
- globaltest
- GOFunction
- goProfiles
- GOSemSim
- goseq
- GOstats
- goTools
- mgsa
- RamiGO
- RNAither

4. Bioconductor modules for GOterms

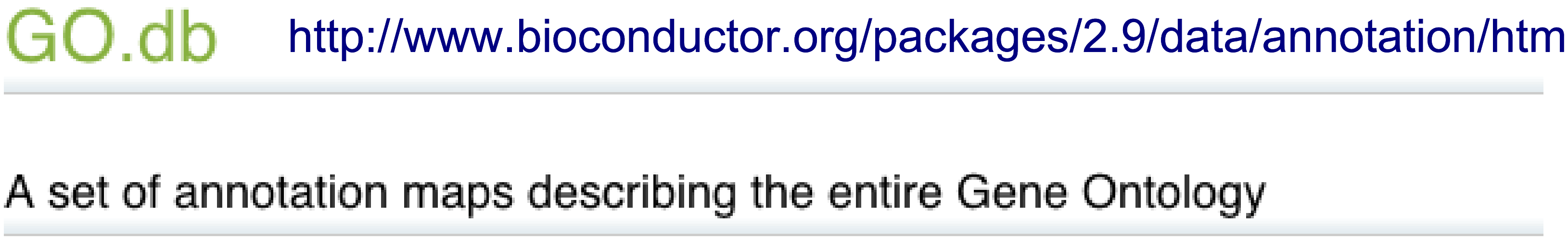
Bioconductor Packages for GO Terms:

GO.db	A set of annotation maps describing the entire Gene Ontology
Gostats	Tools for manipulating GO and microarrays
GOSim	functional similarities between GO terms and gene products
GOProfiles	Statistical analysis of functional profiles
TopGO	Enrichment analysis for Gene Ontology

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    4.1. Basics: GO.db
    4.2. Gene Enrinchment Analysis: TopGO
    4.3. GO profiles: GOProfiles
    4.4. Gene Similarities: GOSim
5. Example.
    5.1. Comparing two arabidopsis chromosomes.

4.1 GO.db



GO.db http://www.bioconductor.org/packages/2.9/data/annotation/html/GO.db.html

A set of annotation maps describing the entire Gene Ontology

1) GO terms stored in **objects**

GOTERM,

GOBPPARENTS, GOCCPARENTS, GOMFPARENTS

GOBPANCESTOR, GOCCANCESTOR, GOMFANCESTOR

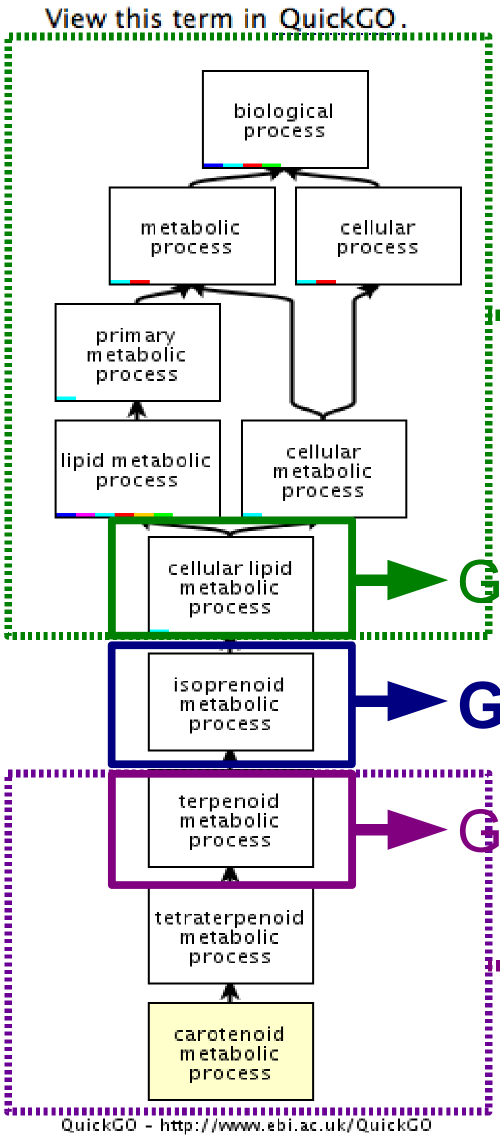
GOBPCHILDREN, GOCCCHILDREN, GOMFCHILDREN

GOBPOFFSPRING, GOCCOFFSPRING, GOMFOFFSPRING

More Information:

http://www.bioconductor.org/packages/2.6/bioc/vignettes/annotate/inst/doc/GOusage.pdf

4.1 GO.db



GOBPANCESTOR → GOBPANCESTOR$"GO:0006720”

GOBPPARENTS → GOBPPARENTS$"GO:0006720”

GOTERM → GOTERM$"GO:0006720”

GOBPCHILDREN → GOBPCHILDREN$"GO:0006720”

GOBPOFFSPRING → GOBPOFFSPRING$"GO:0006720”

4.1 GO.db

Exercise 1:

1.1 Which term correspond top the GOID: GO:0006720 ?

1.2 How many synonyms has this term ?

1.3 How many children ?

1.4 … and parents ?

1.5 … and offsprings ?

1.6 … and ancestors ?

4.1 GO.db

2) Mapping between gene and GO terms stored in **objects or dataframes.**

Bioconductor version 2.8 (Release)

- AnnotationData (593)
  - ChipManufacturer (356)
  - ChipName (190)
  - CustomArray (2)
  - CustomCDF (16)
  - CustomDBSchema (10)
  - FunctionalAnnotation (10)
  - Organism (424)
    - Anopheles_gambiae (4)
    - Apis_mellifera (2)
    - Arabidopsis_thaliana (11)

Packages

AnnotationData > Organism > Arabidopsis_thaliana

- ag.db ▪ agcdf ▪ agprobe ▪ arabidopsis.db0 ▪ ath1121501.db ▪ ath1121501cdf ▪ ath1121501probe
- BSgenome.Athaliana.TAIR.04232008 ▪ BSgenome.Athaliana.TAIR.TAIR9 ▪ hom.At.inp.db ▪ org.At.tair.db

org.At.tair.db

Genome wide annotation for Arabidopsis, primarily based on mapping using TAIR identifiers.

4.1 GO.db

2) Mapping between gene and GO terms stored in **objects (annotate package) or dataframes.**

```
> library("org.At.tair.db")
> org.At.tairGO[["AT5G58560"]]
```

Use a list:

```
> org.At.tairGO[["AT5G58560"]][[1]]$Ontology
```

Functions (“annotate” package):

+ getOntology(inlist, gocategorylist)

+ getEvidence(inlist)

```
> getOntology(org.At.tairGO[["AT5G58560"]])
> getEvidence(org.At.tairGO[["AT5G58560"]])
```

4.1 GO.db

Exercise 2:

2.1 How many terms map with the gene: AT5G58560 ?

2.2 What biological process terms map with this gene?

2.3 What evidences are associated with these terms ?

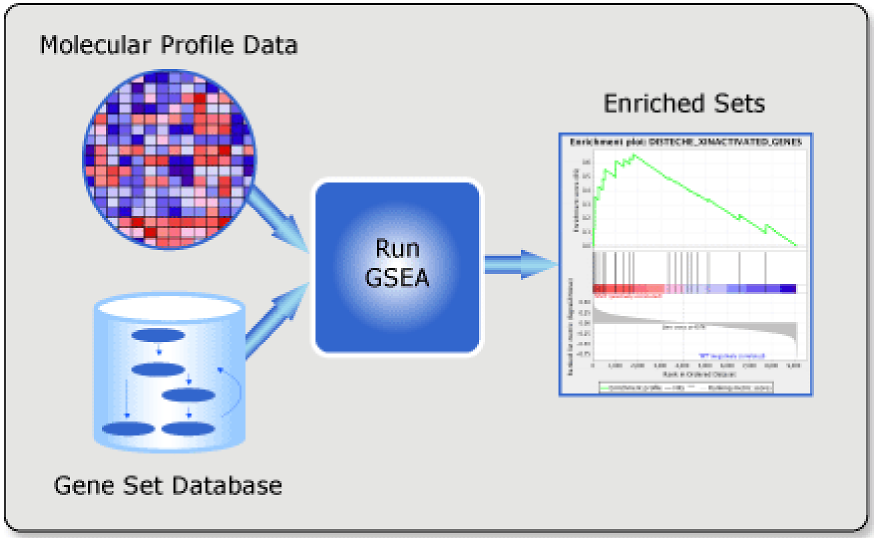
Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    4.1. Basics: GO.db
    4.2. Gene Enrinchment Analysis: TopGO
    4.3. GO profiles: GOProfiles
    4.4. Gene Similarities: GOSim
5. Example.
    5.1. Comparing two arabidopsis chromosomes.

4.2 TopGO

Gene Set Enrinchment Analysis (GSEA)

Computational method that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states (e.g. phenotypes).



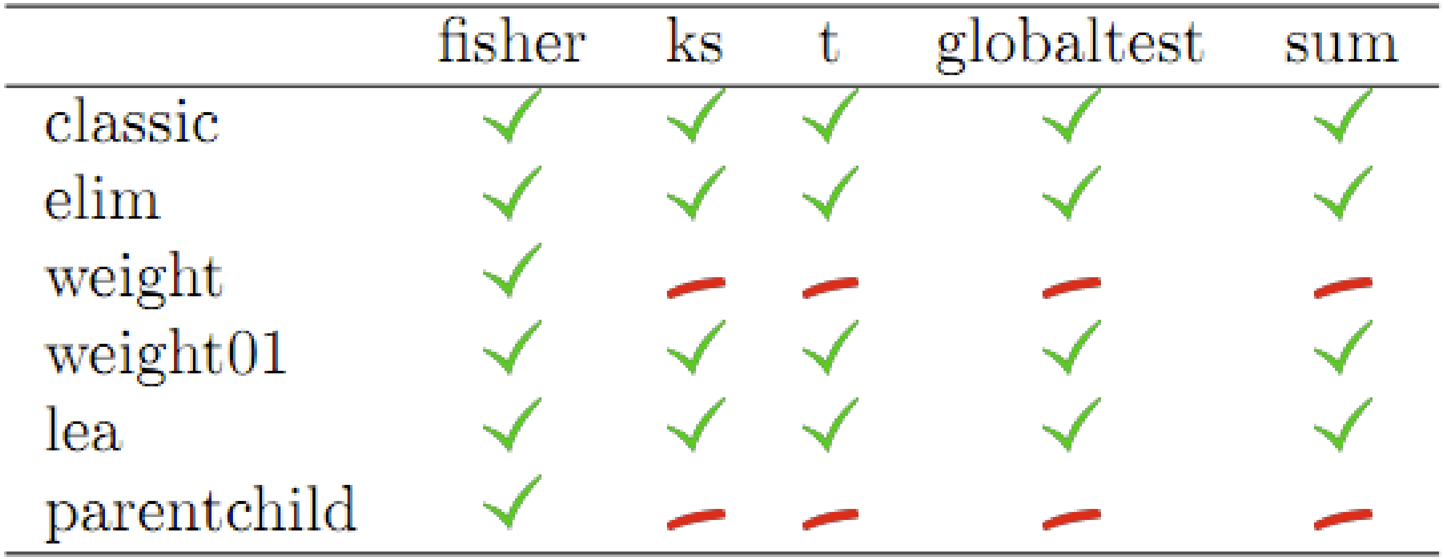
c1	**positional gene sets** for each human chromosome and each cytogenetic band.
c2	**curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.
c3	**motif gene sets** based on conserved *cis*-regulatory motifs from a comparative analysis of the human, mouse, rat and dog genomes.
c4	**computational gene sets** defined by expression neighborhoods centered on 380 cancer-associated genes.
c5	**GO gene sets** consist of genes annotated by the same GO terms.

Documentation: http://www.broadinstitute.org/gsea/index.jsp

4.2 TopGO

topGO http://www.bioconductor.org/packages/release/bioc/html/topGO.html

topGO: Enrichment analysis for Gene Ontology



	fisher	ks	t	globaltest	sum
classic	✓	✓	✓	✓	✓
elim	✓	✓	✓	✓	✓
weight	✓	–	–	–	–
weight01	✓	✓	✓	✓	✓
lea	✓	✓	✓	✓	✓
parentchild	✓	–	–	–	–

Table 1: *Algorithms currently supported by* **topGO**.

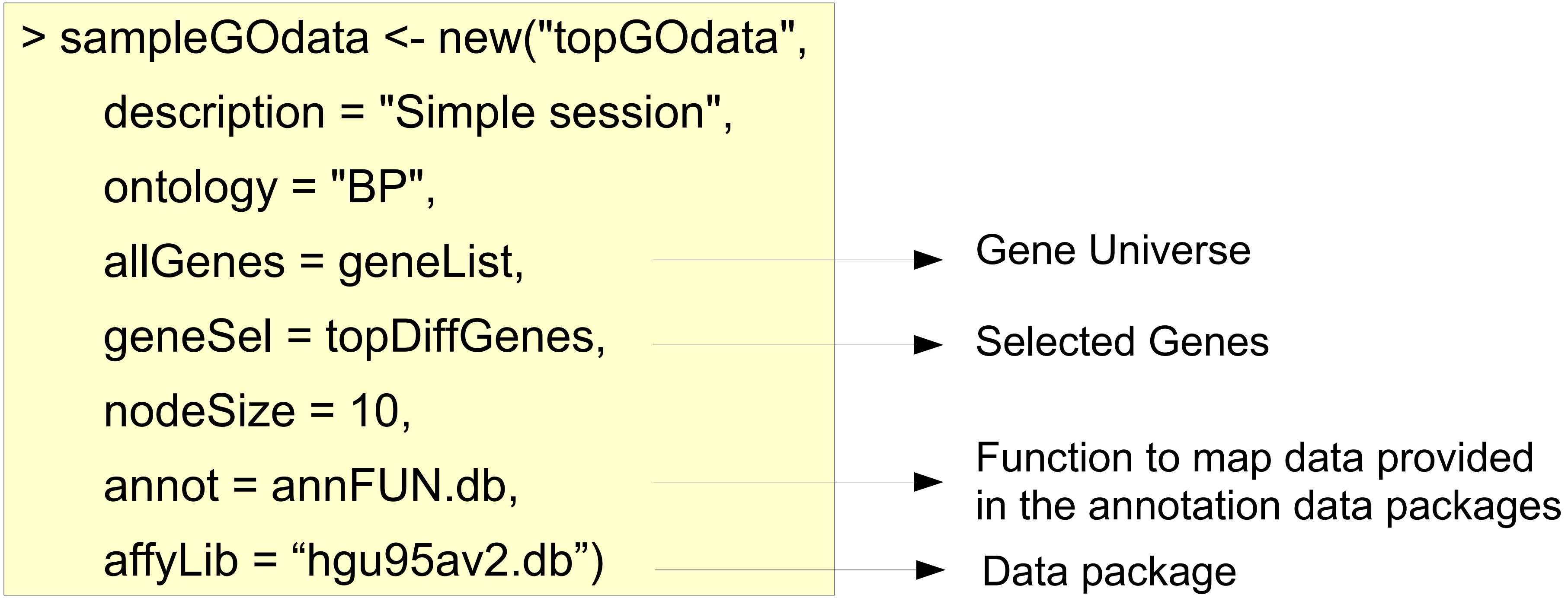
4.2 TopGO

Methodology:

1) Data preparation:
    - a) Gene Universe.
    - b) GO Annotation.
    - c) Criteria to select interesting genes.

OBJECT:

topGOdata



```
> sampleGOdata <- new("topGOdata",
      description = "Simple session",
      ontology = "BP",
      allGenes = geneList,
      geneSel = topDiffGenes,
      nodeSize = 10,
      annot = annFUN.db,
      affyLib = “hgu95av2.db”)
```

4.2 TopGO

sol genomics network

```
> sampleGOdata
------------------------- topGOdata object -------------------------
 Description:
   -  Simple session
 Ontology:
   -  BP
 323 available genes (all genes from the array):
   - symbol:  1095_s_at 1130_at 1196_at 1329_s_at 1340_s_at  ...
   - score :  1 1 0.62238 0.541224 1  ...
   - 50  significant genes.
 316 feasible genes (genes that can be used in the analysis):
   - symbol:  1095_s_at 1130_at 1196_at 1329_s_at 1340_s_at  ...
   - score :  1 1 0.62238 0.541224 1  ...
   - 50  significant genes.
 GO graph (nodes with at least  10  genes):
   - a graph with directed edges
   - number of nodes = 672
   - number of edges = 1358
------------------------- topGOdata object -------------------------
```

OBJECT:

topGOdata

4.2 TopGO

sol genomics network

Methodology:

2) Running the enrinchment test: (**runTest** function)

```
 > resultFisher <- runTest(sampleGOdata,
                          algorithm = "classic",
                          statistic = "fisher")

> resultFisher

Description: Simple session
Ontology: BP
'classic' algorithm with the 'fisher' test
672 GO terms scored: 66 terms with p < 0.01
Annotation data:
    Annotated genes: 316
    Significant genes: 50
    Min. no. of genes annotated to a GO: 10
    Nontrivial nodes: 607
```

`whichAlgorithms()`

`whichTest()`

"fisher"
"ks"
"t"
"globaltest"
"sum"
"ks.ties"

4.2 TopGO

Methodology:

2) Running the enrinchment test:

```
 > resultKS <- runTest(sampleGOdata,
                       algorithm = "classic",
                       statistic = "ks")

> resultKS

Description: Simple session
Ontology: BP
'classic' algorithm with the 'ks' test
672 GO terms scored: 57 terms with p < 0.01
Annotation data:
    Annotated genes: 316
    Significant genes: 50
    Min. no. of genes annotated to a GO: 10
    Nontrivial nodes: 672
```

whichAlgorithms()

whichTest()

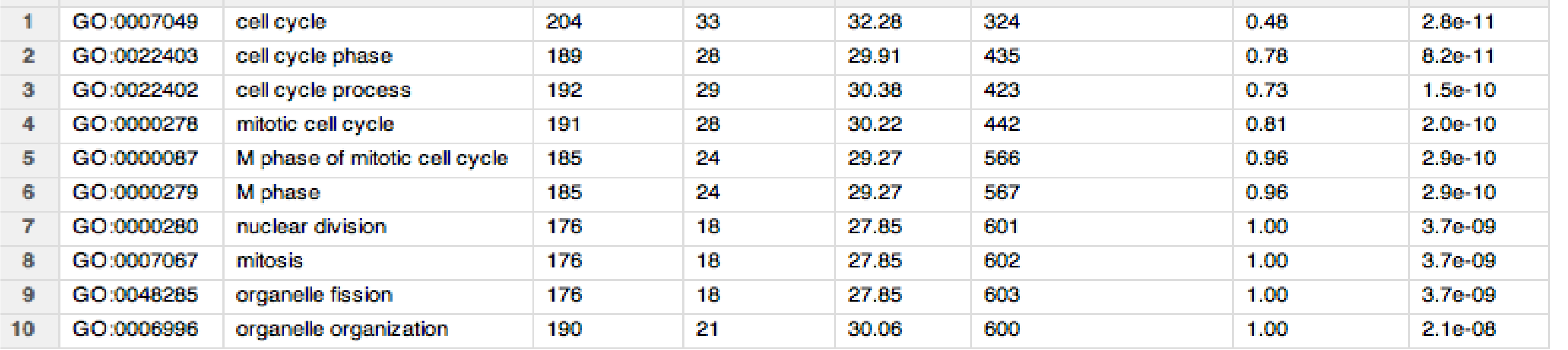
"fisher"
"ks"
"t"
"globaltest"
"sum"
"ks.ties"

4.2 TopGO

Methodology:

3) Analysis of the results: (**GenTable** function)

```
> allRes <- GenTable(sampleGOdata,
                                  classicFisher = resultFisher,
                                  classicKS = resultKS,
                                  orderBy = "classicKS",
                                  ranksOf = "classicFisher",
                                  topNodes = 10)

> allRes
```



	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	classicKS
1	GO:0007049	cell cycle	204	33	32.28	324	0.48	2.8e-11
2	GO:0022403	cell cycle phase	189	28	29.91	435	0.78	8.2e-11
3	GO:0022402	cell cycle process	192	29	30.38	423	0.73	1.5e-10
4	GO:0000278	mitotic cell cycle	191	28	30.22	442	0.81	2.0e-10
5	GO:0000087	M phase of mitotic cell cycle	185	24	29.27	566	0.96	2.9e-10
6	GO:0000279	M phase	185	24	29.27	567	0.96	2.9e-10
7	GO:0000280	nuclear division	176	18	27.85	601	1.00	3.7e-09
8	GO:0007067	mitosis	176	18	27.85	602	1.00	3.7e-09
9	GO:0048285	organelle fission	176	18	27.85	603	1.00	3.7e-09
10	GO:0006996	organelle organization	190	21	30.06	600	1.00	2.1e-08

Using R for GO Terms Analysis:

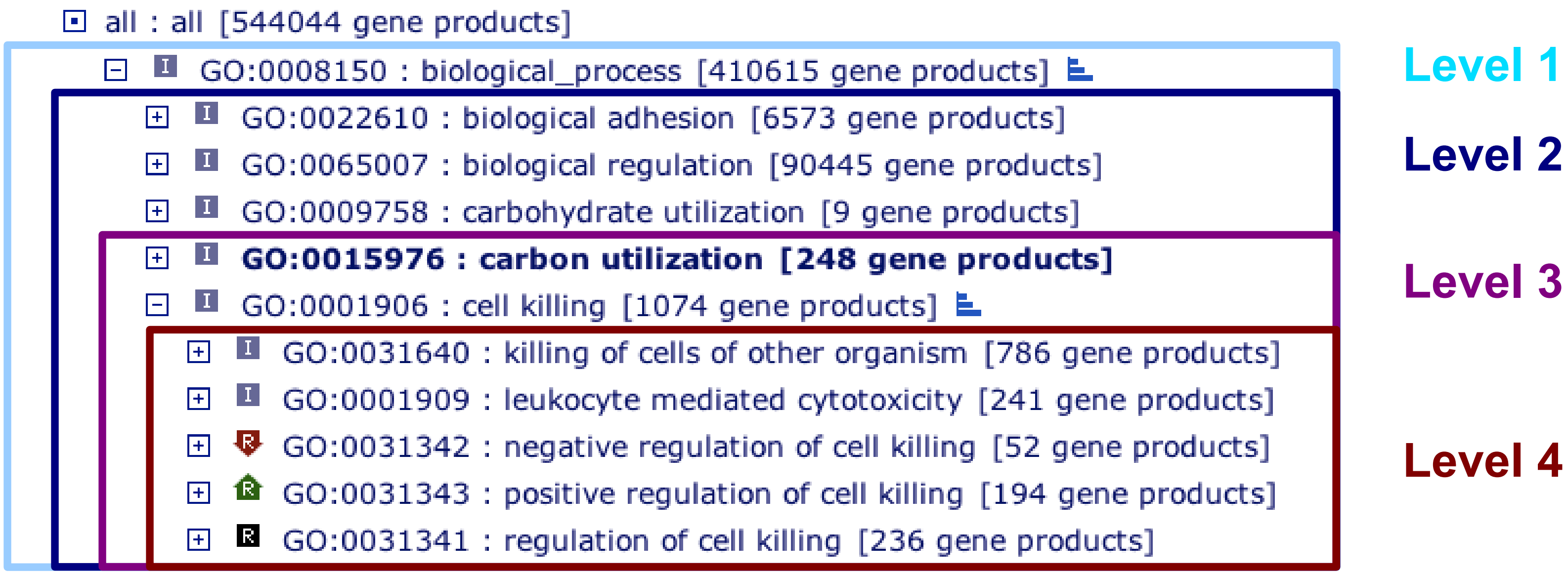
1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Enrinchment Analysis: TopGO
    - **4.3. GO profiles: GOProfiles**
    - 4.4. Gene Similarities: GOSim
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

4.3 goProfiles

goProfiles http://www.bioconductor.org/packages/2.8/bioc/html/goProfiles.html

goProfiles: an R package for the statistical analysis of functional profiles

1) Profiles are built by slicing the GO graph at a given level



⊡ all : all [544044 gene products]
⊟ GO:0008150 : biological_process [410615 gene products]
⊞ GO:0022610 : biological adhesion [6573 gene products]
⊞ GO:0065007 : biological regulation [90445 gene products]
⊞ GO:0009758 : carbohydrate utilization [9 gene products]
⊞ **GO:0015976 : carbon utilization [248 gene products]**
⊟ GO:0001906 : cell killing [1074 gene products]
⊞ GO:0031640 : killing of cells of other organism [786 gene products]
⊞ GO:0001909 : leukocyte mediated cytotoxicity [241 gene products]
⊞ GO:0031342 : negative regulation of cell killing [52 gene products]
⊞ GO:0031343 : positive regulation of cell killing [194 gene products]
⊞ GO:0031341 : regulation of cell killing [236 gene products]

Level 1

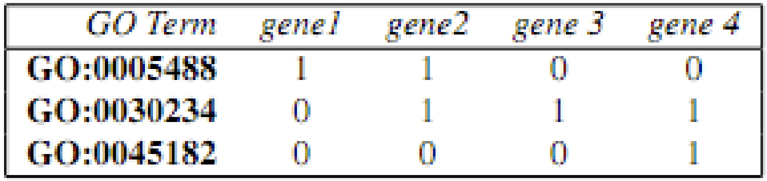
Level 2

Level 3

Level 4

4.3 goProfiles

2) Functional profile at a given GO level is obtained by counting the number of identifiers having a hit in each category of this level



GO Term	gene1	gene2	gene 3	gene 4
GO:0005488	1	1	0	0
GO:0030234	0	1	1	1
GO:0045182	0	0	0	1

Table 2.1: A simple list of 4 identifiers considered at level 2 of the MF ontology illustrates how some genes may have hits in several categories simultaneously.

3) Profiles comparissons:

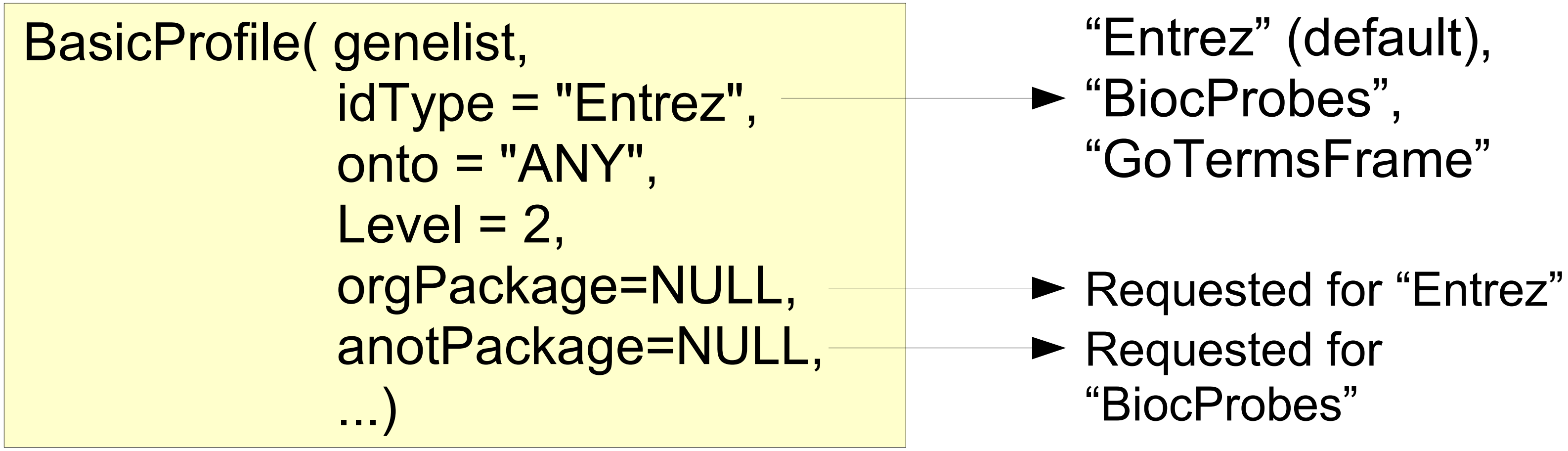
1. How different or representative of a given gene universe is a given set of genes?
   - Universe: All genes analyzed,
     Gene Set: Differentially expressed genes in a microarray experiment
   - Universe: All genes in a database,
     Gene Set: Arbitrarily selected set of genes
2. How biologically different are two given sets of genes?
   - Differentially expressed genes in two experiments
   - Arbitrarily chosen lists of genes

4.3 goProfiles

Methodology:

1) Data preparation:
    a) Gene GO term map (**object** or **dataframe**).
        Dataframe with 4 columns:
            + GeneID
            + Ontology
            + Evidence
            + GOID

2) Profile creation:
    a) **basicProfile** function



```
BasicProfile( genelist,
              idType = "Entrez",
              onto = "ANY",
              Level = 2,
              orgPackage=NULL,
              anotPackage=NULL,
              ...)
```

"Entrez" (default),
"BiocProbes",
"GoTermsFrame"

Requested for "Entrez"

Requested for
"BiocProbes"

4.3 goProfiles

Methodology:

2) Profile creation:

a) **basicProfile** function

```
> printProfiles(bpprofile)

Functional Profile
==================
[1] "BP ontology"
                    Description                  GOID              Frequency
25 cellular component biogen...         GO:0044085            5
14 cellular component organi...          GO:0016043            6
12              cellular process                GO:0009987          122
32 establishment of localiza...            GO:0051234            4
4      immune system process...          GO:0002376            1
31                   localization               GO:0051179            4
2              metabolic process              GO:0008152            3
33    multi-organism process...          GO:0051704            2
30         response to stimulus             GO:0050896            4
```

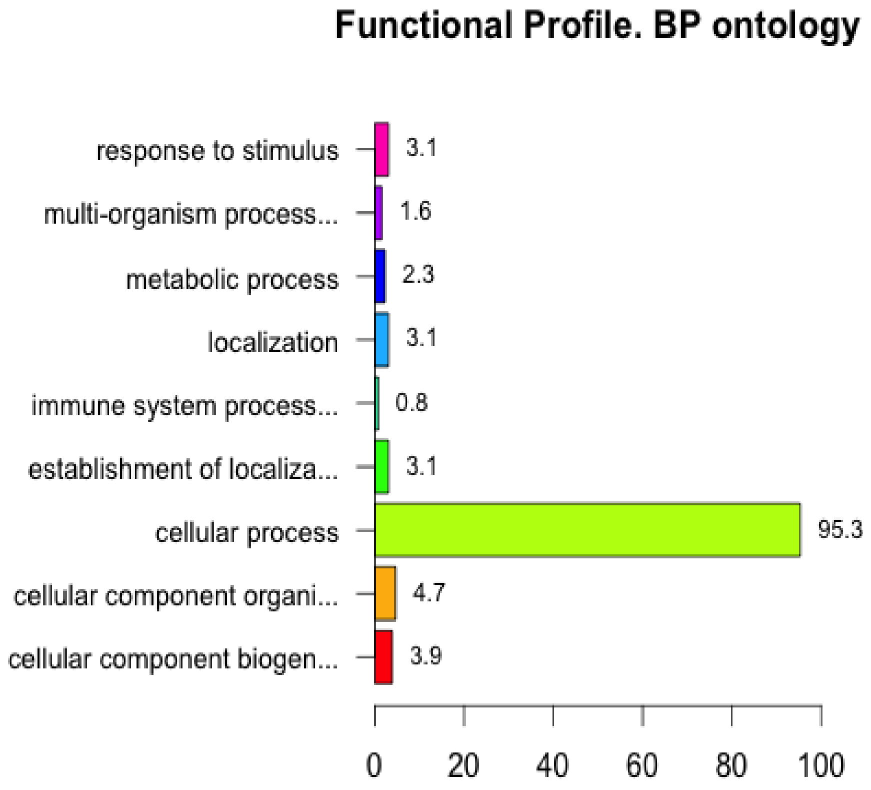
4.3 goProfiles

Methodology:

2) Profile creation:

a) **basicProfile** function

```
> plotProfiles(bpprofile)
```



4.3 goProfiles

Methodology:

2) Profile creation:

b) **expandedProfile** function

Used mainly for comparisons of profiles.

```
ExpandedProfile( genelist,
                 idType = "Entrez",
                 onto = "ANY",
                 Level = 2,
                 orgPackage=NULL,
                 anotPackage=NULL
                 ...)
```

3) Profile comparisons:

- Case I (INCLUSION): One list incluided in the other.
- Case || (DISJOINT): Non overlapping gene sets
- Case III (INTERSECTION): Overlapping genes

4.3 goProfiles

Methodology:

3) Profile comparisons:

- Case I (INCLUSION):	compareProfilesLists()
- Case || (DISJOINT):	compareGeneLists()
- Case III (INTERSECTION):	compareGeneLists()

```
> comp_ath_genes <- compareGeneLists(
          genelist1=ath_chl_list,
          genelist2=ath_mit_list,
          idType="Entrez", orgPackage="org.At.tair.db",
          onto="BP", level=2)

> print(compSummary(comp_ath_genes))

   Sqr.Euc.Dist  StdErr      pValue      0.95CI.low    0.95CI.up
   0.031401      0.024101  0.005000   -0.015837     0.078639
```

4.3 goProfiles

Methodology:

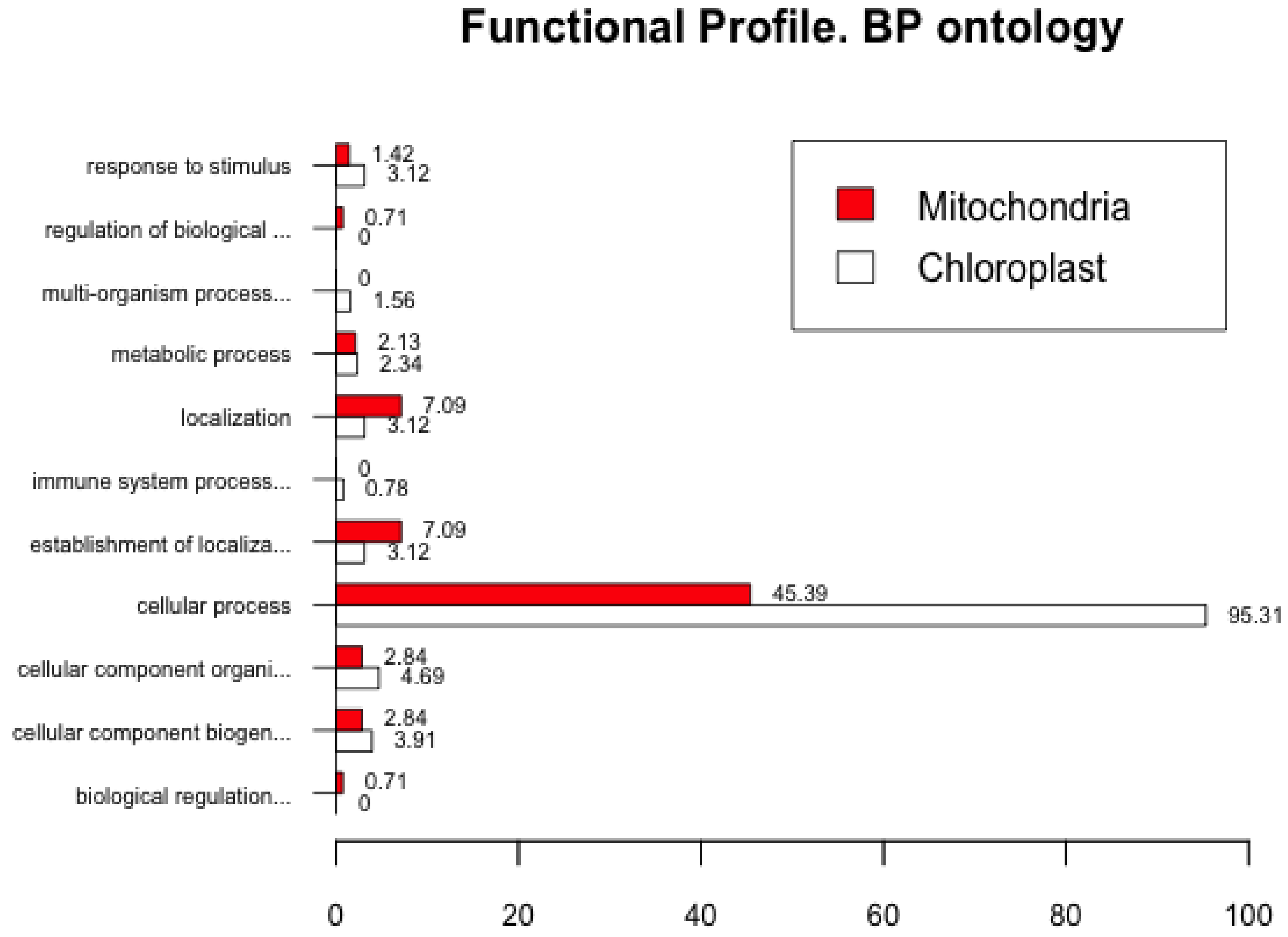
3) Profile comparisons:

```
 > basic_mitprof <- basicProfile(ath_mit_list, idType="Entrez",
             onto="BP", level=2, orgPackage="org.At.tair.db",
             empty.cats=TRUE)

> basic_chloprof <- basicProfile(ath_chl_list, idType="Entrez",
             onto="BP", level=2, orgPackage="org.At.tair.db",
             empty.cats=TRUE)

> merged_prof <- mergeProfilesLists(basic_chloprof, basic_mitprof,
             profNames=c("Chloroplast", "Mitochondria"))

> plotProfiles(merged_prof, percentage=TRUE, legend=TRUE)
```

4.3 goProfiles

Methodology:

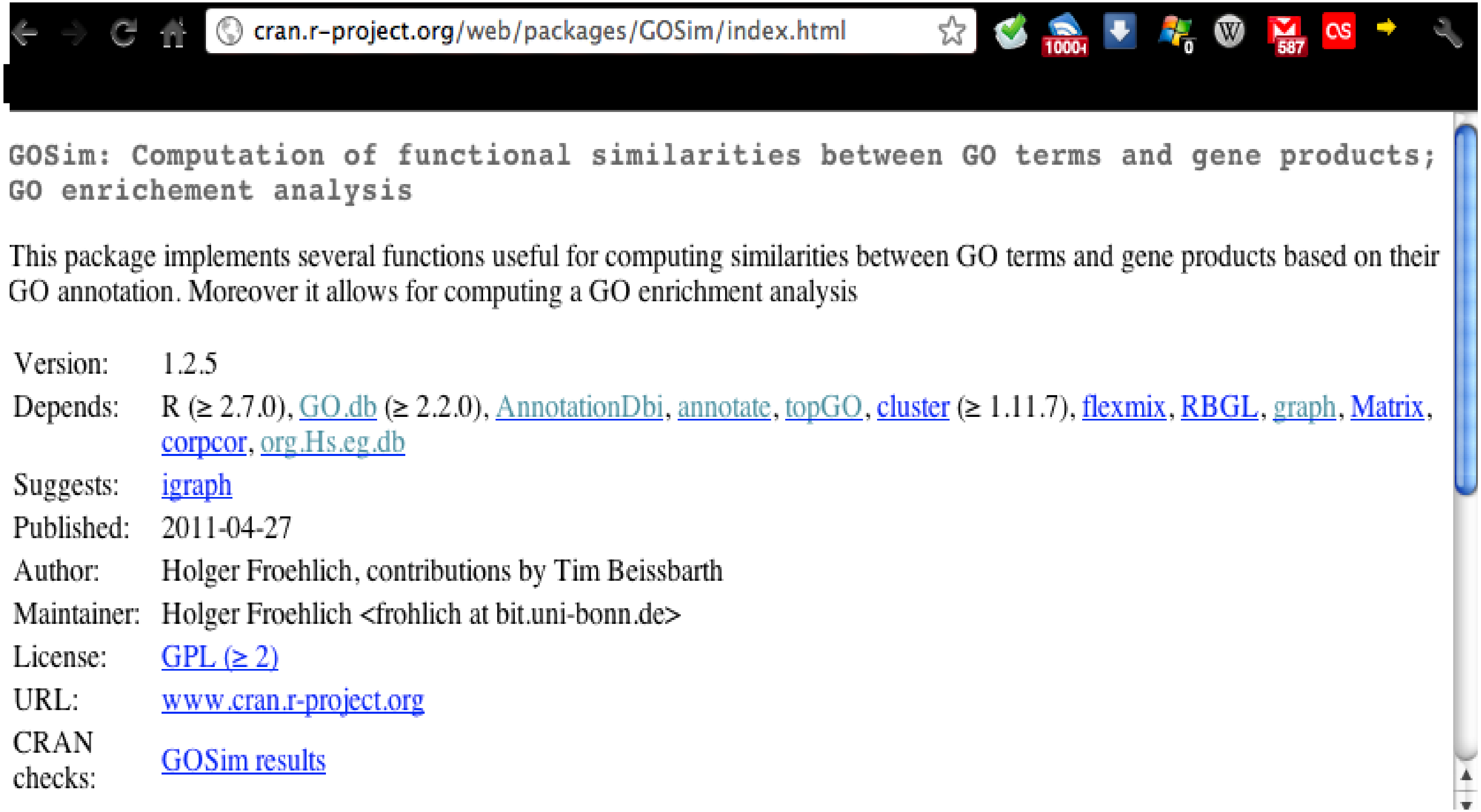
3) Profile comparisons:



Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    - 4.1. Basics: GO.db
    - 4.2. Gene Enrinchment Analysis: TopGO
    - 4.3. GO profiles: GOProfiles
    - 4.4. Gene Similarities: GOSim
5. Example.
    - 5.1. Comparing two arabidopsis chromosomes.

4.4 GOSim



GOSim: Computation of functional similarities between GO terms and gene products; GO enrichement analysis

This package implements several functions useful for computing similarities between GO terms and gene products based on their GO annotation. Moreover it allows for computing a GO enrichment analysis

Version:	1.2.5
Depends:	R (≥ 2.7.0), GO.db (≥ 2.2.0), AnnotationDbi, annotate, topGO, cluster (≥ 1.11.7), flexmix, RBGL, graph, Matrix, corpcor, org.Hs.eg.db
Suggests:	igraph
Published:	2011-04-27
Author:	Holger Froehlich, contributions by Tim Beissbarth
Maintainer:	Holger Froehlich <frohlich at bit.uni-bonn.de>
License:	GPL (≥ 2)
URL:	www.cran.r-project.org
CRAN checks:	GOSim results

Using R for GO Terms Analysis:

1. What are GO terms ?
2. What is R ?
3. What is Bioconductor ?
4. Bioconductor modules for GO terms
    4.1. Basics: GO.db
    4.2. Gene Enrinchment Analysis: TopGO
    4.3. GO profiles: GOProfiles
    4.4. Gene Similarities: GOSim
5. Example.
    5.1. Comparing two arabidopsis chromosomes.

5 Example

Exercise 3:

Compare the goProfiles for A. thaliana chromosomes 1 and 5.

Are they significantly differents ?

Source:

ftp://ftp.arabidopsis.org/home/tair/Genes/

TAIR10_genome_release/TAIR10_gene_lists/TAIR10_all_gene_models

R Package:

GoProfiles