

Gene Ontology

Till Helge Helwig

Praktikum "Bioinformatics Software Tools"

March 30 - April 10, 2009

1 Introduction

- Problems with biological databases
- The answer to the problems

2 Database structure

- General decisions and techniques
- The three domains

3 Tools

- General information
- A small selection of tools

Why Gene Ontology?

Example (Development of new antibiotic)

“Let’s find all gene products involved in bacterial protein synthesis, that differ from their equivalents in human cells.”

Why Gene Ontology?

Example (Development of new antibiotic)

“Let’s find all gene products involved in bacterial protein synthesis, that differ from their equivalents in human cells.”

Problems:

- Many relevant entries in many different databases
- Each database has its **own terminology** for “bacterial protein synthesis”

Why Gene Ontology?

Example (Development of new antibiotic)

“Let’s find all gene products involved in bacterial protein synthesis, that differ from their equivalents in human cells.”

Problems:

- Many relevant entries in many different databases
 - Each database has its **own terminology** for “bacterial protein synthesis”
- ⇒ No way to perform an **automated search**

Why Gene Ontology?

Example (Development of new antibiotic)

“Let’s find all gene products involved in bacterial protein synthesis, that differ from their equivalents in human cells.”

Problems:

- Many relevant entries in many different databases
 - Each database has its **own terminology** for “bacterial protein synthesis”
- ⇒ No way to perform an **automated search**

⇒ **Unification of terms** needed to allow simultaneous searches on many databases

⇒ **Ontologies**

Why Gene Ontology? (2)

- Genetical knowledge **spread** over lots of different projects

Why Gene Ontology? (2)

- Genetical knowledge **spread** over lots of different projects
- There are genes and proteins **shared** by most or even all living cells
- Knowing a protein's function in one organism **helps to understand** its function in other organisms

Why Gene Ontology? (2)

- Genetical knowledge **spread** over lots of different projects
- There are genes and proteins **shared** by most or even all living cells
- Knowing a protein's function in one organism **helps to understand** its function in other organisms
- **Interoperability of databases** should be ensured to enable automatic **transfer of annotations**
- Search queries covering **several** databases while applying **filters** concerning i.e. a biological function needed

What is Gene Ontology?

- Founded in **1998** as collaboration of:
 - Flybase
 - Saccharomyces Genome Database
 - Mouse Genome Database
- Tries to build a **set of ontologies** providing unified terminologies for a set of domains

What is Gene Ontology?

- Founded in **1998** as collaboration of:
 - Flybase
 - Saccharomyces Genome Database
 - Mouse Genome Database
- Tries to build a **set of ontologies** providing unified terminologies for a set of domains
- Since its foundation **13 projects** joined the consortium and another **4 are contributing associates**

What is Gene Ontology?

- Founded in **1998** as collaboration of:
 - Flybase
 - Saccharomyces Genome Database
 - Mouse Genome Database
- Tries to build a **set of ontologies** providing unified terminologies for a set of domains
- Since its foundation **13 projects** joined the consortium and another **4 are contributing associates**
- On March 25th, 2009 GO contained **27009 terms**

Three domains

- GO covers **three different domains** to describe gene products:
 - 1 Cellular component
 - 2 Molecular function
 - 3 Biological process

Three domains

- GO covers **three different domains** to describe gene products:
 - ① Cellular component
 - ② Molecular function
 - ③ Biological process
- Every gene product can be **annotated** with one or more terms from each of these ontologies
- **Not** meant to provide a complete cover of biology, but **consensus terminologies** used by all participating projects

How does GO work?

- Ontologies built as **directed acyclic graphs**
- Terms can have **more than one parent**

How does GO work?

- Ontologies built as **directed acyclic graphs**
- Terms can have **more than one parent**
- Between terms **relationships** are established, i.e.:
 - is a: class-subclass relation
(*"nucleus" is a "cell part"*)

How does GO work?

- Ontologies built as **directed acyclic graphs**
- Terms can have **more than one parent**
- Between terms **relationships** are established, i.e.:
 - is a:** class-subclass relation
(*"nucleus" is a "cell part"*)
 - part of:** whenever the former is present, the latter is too, but not necessarily in reverse
(*"cell part" part of "cell"*)

How does GO work?

- Ontologies built as **directed acyclic graphs**
- Terms can have **more than one parent**
- Between terms **relationships** are established, i.e.:
 - is a**: class-subclass relation
(*"nucleus" is a "cell part"*)
 - part of**: whenever the former is present, the latter is too, but not necessarily in reverse
(*"cell part" part of "cell"*)
 - regulates**: in a biological process the former has an influence on the latter
(*"translational frameshifting" regulates "translation"*)

How does GO work?

- Ontologies built as **directed acyclic graphs**
- Terms can have **more than one parent**
- Between terms **relationships** are established, i.e.:
 - is a**: class-subclass relation
(*"nucleus" is a "cell part"*)
 - part of**: whenever the former is present, the latter is too, but not necessarily in reverse
(*"cell part" part of "cell"*)
 - regulates**: in a biological process the former has an influence on the latter
(*"translational frameshifting" regulates "translation"*)
- There are **transitive relations** between these relationships

translation

[Term information](#) [Term lineage](#) [External references](#) [6399 gene product associations](#)

Term Information

Accession GO:0006412

Ontology biological process

Synonyms
exact: protein anabolism
exact: protein biosynthesis
exact: protein formation
exact: protein synthesis
exact: protein translation
alt_id: GO:0006416
alt_id: GO:0006453
alt_id: GO:0043037

Definition The chemical reactions and pathways resulting in the formation of a protein. This is a ribosome-mediated process in which the information in messenger RNA (mRNA) is used to specify the sequence of amino acids in the protein. [source: GOC:go_curators]

Comment None

Subset
[goslim_candida](#)
[goslim_generic](#)
[goslim_pir](#)
[goslim_plant](#)
[goslim_yeast](#)

[Back to top](#)

Screenshot: Term "translation"

Term Lineage

[Switch to viewing term ancestors](#)

Filter tree view

Filter Gene Product Counts

Data source

All
CGD
dictyBase
EcoCyc

Species
















All
Anaplasma phagocy...
Arabidopsis thaliana
Bacillus anthraci...

View Options

Tree view Full Compact






Set filters

Remove all filters

- GO:0034961 : cellular biopolymer biosynthetic process [19483 gene products]
 -  GO:0033692 : cellular polysaccharide biosynthetic process [779 gene products]
 -  GO:0006260 : DNA replication [1817 gene products]
 -  GO:0070501 : poly-gamma-glutamate biosynthetic process [0 gene products]
 -  GO:0006350 : transcription [10789 gene products]
 -  **GO:0006412 : translation [6399 gene products]**
 -  GO:0032543 : mitochondrial translation [120 gene products]
 -  GO:0017148 : negative regulation of translation [137 gene products]
 -  GO:0032544 : plastid translation [1 gene product]
 -  GO:0045727 : positive regulation of translation [70 gene products]
 -  GO:0006417 : regulation of translation [848 gene products]
 -  GO:0006414 : translational elongation [1654 gene products]
 -  GO:0006413 : translational initiation [455 gene products]
 -  GO:0006415 : translational termination [93 gene products]
 -  GO:0006418 : tRNA aminoacylation for protein translation [810 gene products]
 -  GO:0006414 : translational elongation [1654 gene products]
- GO:0044267 : cellular protein metabolic process [17264 gene products]

Actions...

Last action: Reset
the tree
Graphical View
View in tree browser

 negatively regulates
 is a
 positively regulates
 regulates
 part of

GO: Cellular component

- Describes **locations** on the level of subcellular structures and macromolecular complexes
- Multicellular (or anatomical) locations and individual proteins are **not included**
- Root node: **cell (GO:0005623)**
- Locations are clearly **seperated into classes** like “Intracellular” or “Plasma membrane”

GO: Cellular component

- Describes **locations** on the level of subcellular structures and macromolecular complexes
- Multicellular (or anatomical) locations and individual proteins are **not included**
- Root node: **cell (GO:0005623)**
- Locations are clearly **seperated into classes** like “Intracellular” or “Plasma membrane”

Example

```
sarcoplasmic reticulum
[p] sarcoplasmic reticulum membrane
---[i] free sarcoplasmic reticulum membrane
---[i] junctional sarcoplasmic reticulum membrane
[p] sarcoplasmic reticulum lumen
```

GO: Molecular function

- Covers **abilities** a gene product has, i.e. transport or binding
- Are often named “**activity**” to clearly separate them from the gene product itself
- There are **no terms** describing more than one function at the same time

GO: Molecular function

- Covers **abilities** a gene product has, i.e. transport or binding
- Are often named “**activity**” to clearly separate them from the gene product itself
- There are **no terms** describing more than one function at the same time

Example

regulation of transcription

[i] regulation of transcription factor activity

---[i] positive reg. of transcription factor activity

---[i] negative reg. of transcription factor activity

[i] regulation of transcription, start site selection

GO: Biological process

- Contains terms that describe **series of events** or molecular functions
- Terms describe processes or represent a **collection** of processes
- Processes might be summarized in a **single “development node”** to prevent a proliferation
- Some processes (i.e. the cell cycle) are **split** into molecular events and temporal stages

GO: Biological process

- Contains terms that describe **series of events** or molecular functions
- Terms describe processes or represent a **collection** of processes
- Processes might be summarized in a **single “development node”** to prevent a proliferation
- Some processes (i.e. the cell cycle) are **split** into molecular events and temporal stages

Example

```
cell cycle
[p] cell cycle process
---[i] cell cycle phase
-----[i] anaphase
---[i] centromere separation
```

Standard definitions

- For all three ontologies there are **defined standards**
- They set rules for:
 - the choice of **term names**
 - the creation of **new subclasses**
 - **tests** to ensure consistency and correctness
 - **ontology specific things** like terms for start and end of a process
- These standards are especially relevant if you want to **contribute** to GO

Working with GO

- There is a **huge amount of tools** (over 100) that either use GO or can be used to work with GO
- Only **two official tools** are maintained by the GO consortium
- The remaining ones were created by the **community** for many different tasks

Working with GO

- There is a **huge amount of tools** (over 100) that either use GO or can be used to work with GO
 - Only **two official tools** are maintained by the GO consortium
 - The remaining ones were created by the **community** for many different tasks
 - On the **GO homepage** there is a complete list available:
<http://www.geneontology.org/GO.tools.shtml>
- ⇒ I will only provide a **very short overview**

Official tools

● AmiGO

- Web-based **browsing and searching** tool to access the GO database
- Searches GO for terms or gene products
- Shows the **position** of terms in the ontology (parents, siblings)
- Gene products are shown with their GO term **annotations**

Official tools

● AmiGO

- Web-based **browsing and searching** tool to access the GO database
- Searches GO for terms or gene products
- Shows the **position** of terms in the ontology (parents, siblings)
- Gene products are shown with their GO term **annotations**

● OBO-Edit

- Graph-based tool to **view and edit** OBO format ontologies
- **Easy user interface** focussing on the display of the ontology structure

Unofficial tools

Tasks performed by these tools:

- **Searching and browsing** GO
e.g. DynGO, GoFish, QuickGO, GenNav, PANDORA, ...

Unofficial tools

Tasks performed by these tools:

- **Searching and browsing** GO
e.g. DynGO, GoFish, QuickGO, GenNav, PANDORA, ...
- **Annotation** of data sets
e.g. Blast2GO, GOanna, GOtcha, Manatee, ...

Unofficial tools

Tasks performed by these tools:

- **Searching and browsing** GO
e.g. DynGO, GoFish, QuickGO, GenNav, PANDORA, ...
- **Annotation** of data sets
e.g. Blast2GO, GOanna, GOtcha, Manatee, ...
- **Analyzation** of data sets
e.g. BiNGO, EasyGO, FatiGO, GOArray, MultiGO, ...

Unofficial tools

Tasks performed by these tools:

- **Searching and browsing** GO
e.g. DynGO, GoFish, QuickGO, GenNav, PANDORA, ...
- **Annotation** of data sets
e.g. Blast2GO, GOanna, GOtcha, Manatee, ...
- **Analyzation** of data sets
e.g. BiNGO, EasyGO, FatiGO, GOArray, MultiGO, ...
- Analysis of the **GO database** itself and other things
e.g. CateGORizer, FunSpec, GOChase, WEGO, Whatizit, ...

The last slide

Thank you for listening.

References



M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, et al.
Gene Ontology: tool for the unification of biology.
Nature Genetics, 25:25–29, 2000.



Gene Ontology Consortium.
Homepage: www.genetontology.org.