Ontologies

- What is an ontology?
  - What is a biomedical ontology?
    - How is it generated?
    - How is it used?
    - .....

Ontology

onto-, of being or existence;
-logy, study.
Greek origin; Latin, *ontologia*, 1606

- In ***philosophy***, it seeks to describe basic categories and relationships of being or existence to define entities and types of entities within its framework:
    - What do you know? How do you know it?
    - What is existence? What is a physical object?
    - What constitutes the identity of an object? .....
- Central goal is to have a definitive and exhaustive classification of all entities.

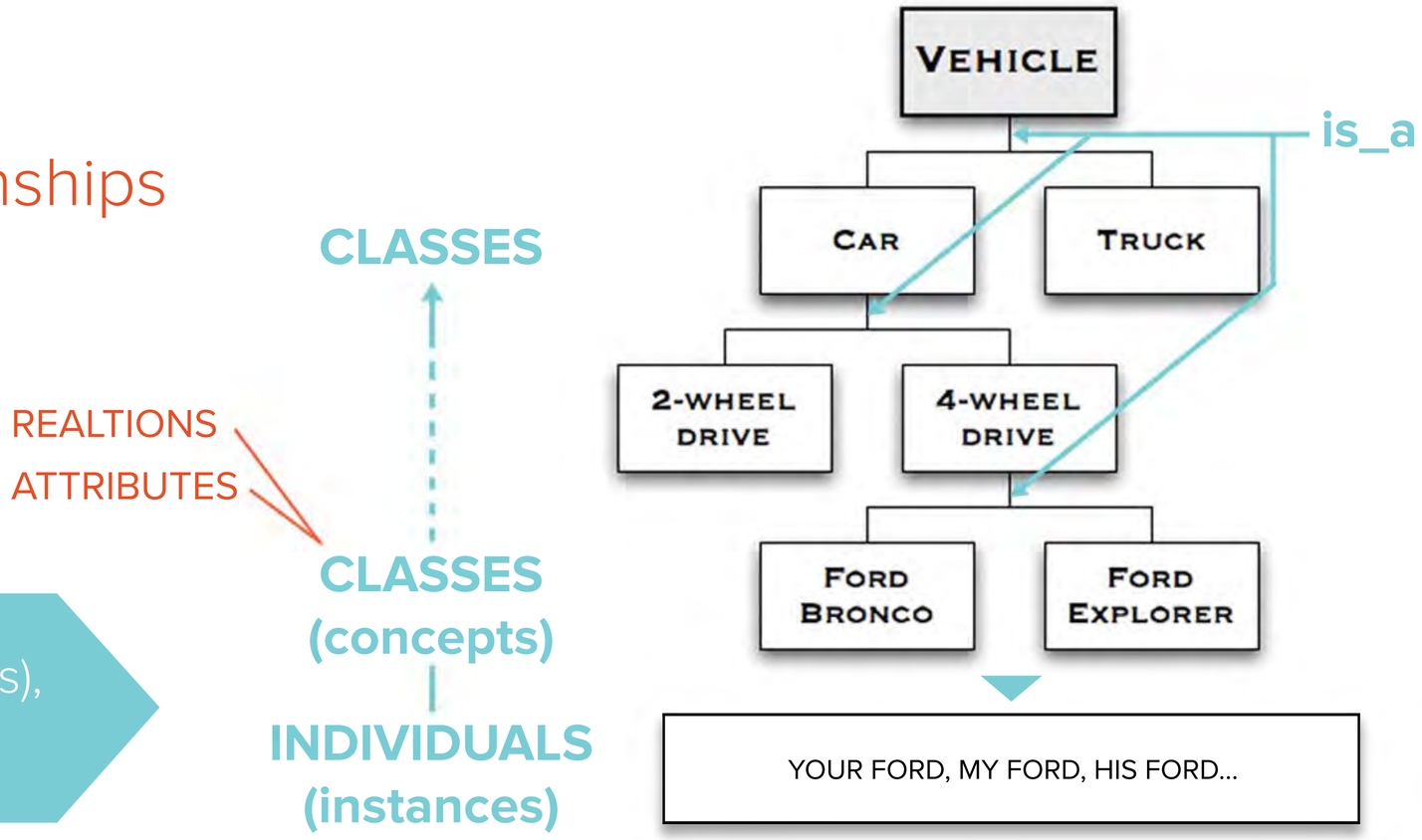
"The science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality"

- Barry Smith, U Buffalo

In Computer and Information Science

Ontology is a data model that represents a set of concepts within a domain and the relationships between those concepts.
It is used to reason about objects within the domain.

Most ontologies describe individuals (instances), classes (concepts), attributes, and relations



What are Ontologies Useful for?

- Ontology is a form of knowledge representation about the world or some part of it. They can help with:
    - Terminology management
    - Integration, interoperability, and sharing of data
        - promote precise communication between scientists
        - enable information retrieval across multiple resources
    - Knowledge reuse and decision support
        - extend the power of computational approaches to perform data exploration, inference, and data mining

Biomedical *Terminology* vs. *Ontology*

Some common Biomedical Terminologies:

- MeSH (medical subject heading)
- NCI Thesaurus
- SNOMED / SNODENT
- UMLS (unified medical language system)

Terminology Collections of concepts and terms in a certain language in a specific field. May or may not be structured.

Ontology A formal, explicit (conceptual) model of objects in a structured computational representation.

Both are knowledge organization systems: structured terminology systems can be “ontologized“ and ontology terms are routinely included in terminologies.

Ontology Enables Large-Scale Biomedical Science

The center of two major activities currently in biomedical research:

Structured representation of biomedicine

For different types of entities and relations to describe biomedicine (ontology content curation).

Annotation

Using ontologies to summarize and describe biomedical experimental results to enable:

- Integration of their data with other researchers' results
- Cross-species analyses

Enrichment analysis

Your gene IDs here...

biological process

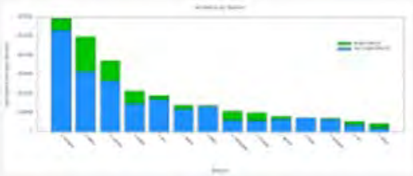
Homo sapiens

Submit

Advanced options / Help

Powered by PANTHER

Statistics



Other GOC tools

Explore other GOC tools in the AmiGO software suite.

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

Filter classes

Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products
cellular component
where gene products are active
biological process
pathways and larger processes made up of the activities of multiple gene products.
more

Annotations

Download annotations (standard files)

Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. more

What is the Gene Ontology?

GO provides controlled vocabulary to describe gene and gene product attributes and relationships in any organism.

Three separate ontologies:

Biological process

Series of events accomplished by one or more ordered assemblies of molecular functions, e.g. signal transduction, or pyrimidine metabolism.

What is the Gene Ontology?

GO provides controlled vocabulary to describe gene and gene product attributes and relationships in any organism.

Three separate ontologies:

Biological process

Molecular function

Describes activities, such as catalytic or binding activities, that occur at the molecular level. Activities that can be performed by individual gene products, or by assembled complexes of gene products; e.g. catalytic activity, transporter activity.

What is the Gene Ontology?

GO provides controlled vocabulary to describe gene and gene product attributes and relationships in any organism.

Three separate ontologies:

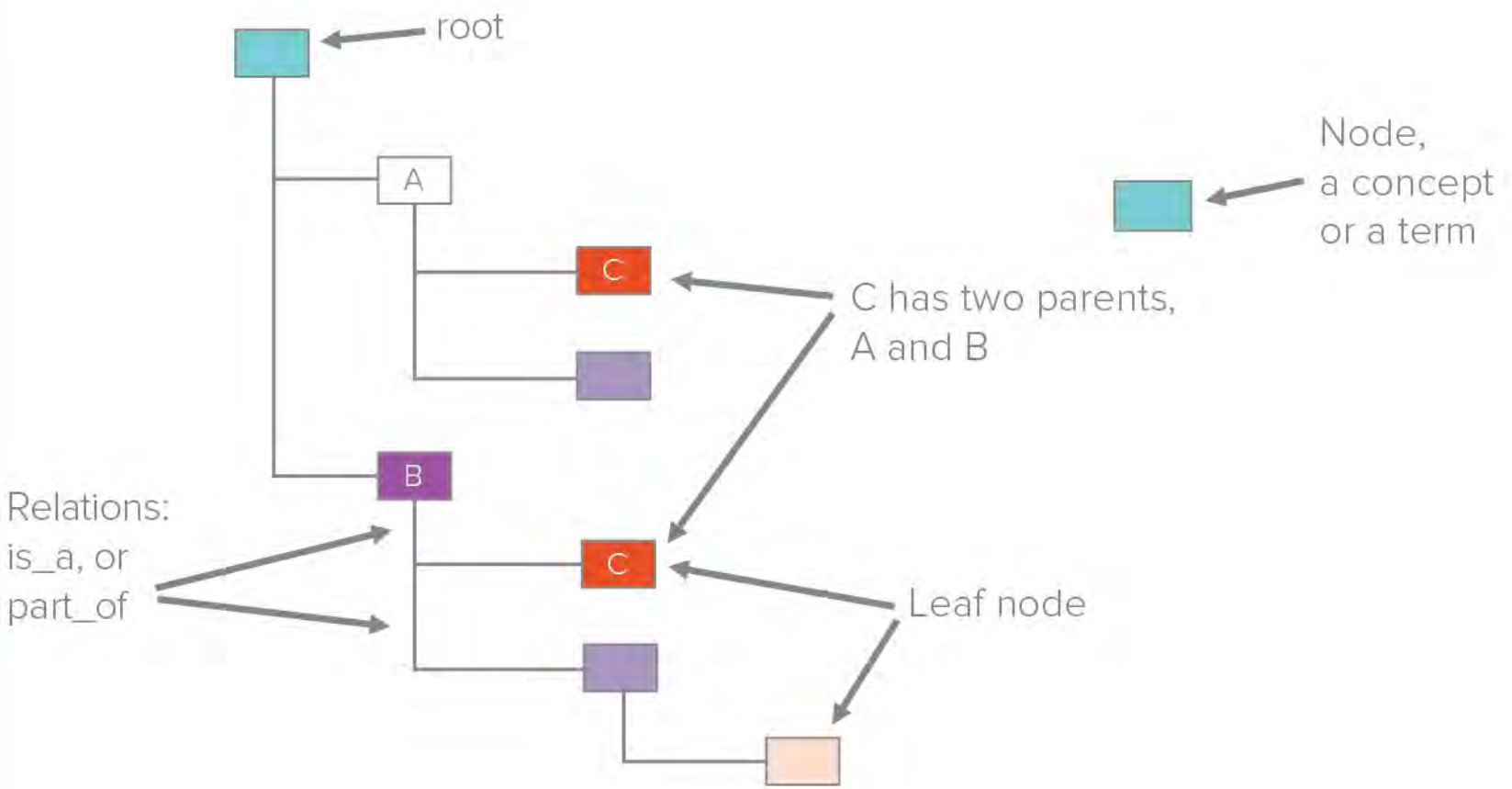
Biological process

Molecular function

Cellular component

A component of a cell that it is part of some larger object, maybe an anatomical structure (e.g. ER or nucleus) or a gene product group.

GO is a Network Structure



Protein **Cytochrome c**

Gene **CYCS**

Organism *Homo sapiens (Human)*

Status Reviewed - Annotation score: ◉◉◉◉◉ - Experimental evidence at protein level[i]

Function[i]

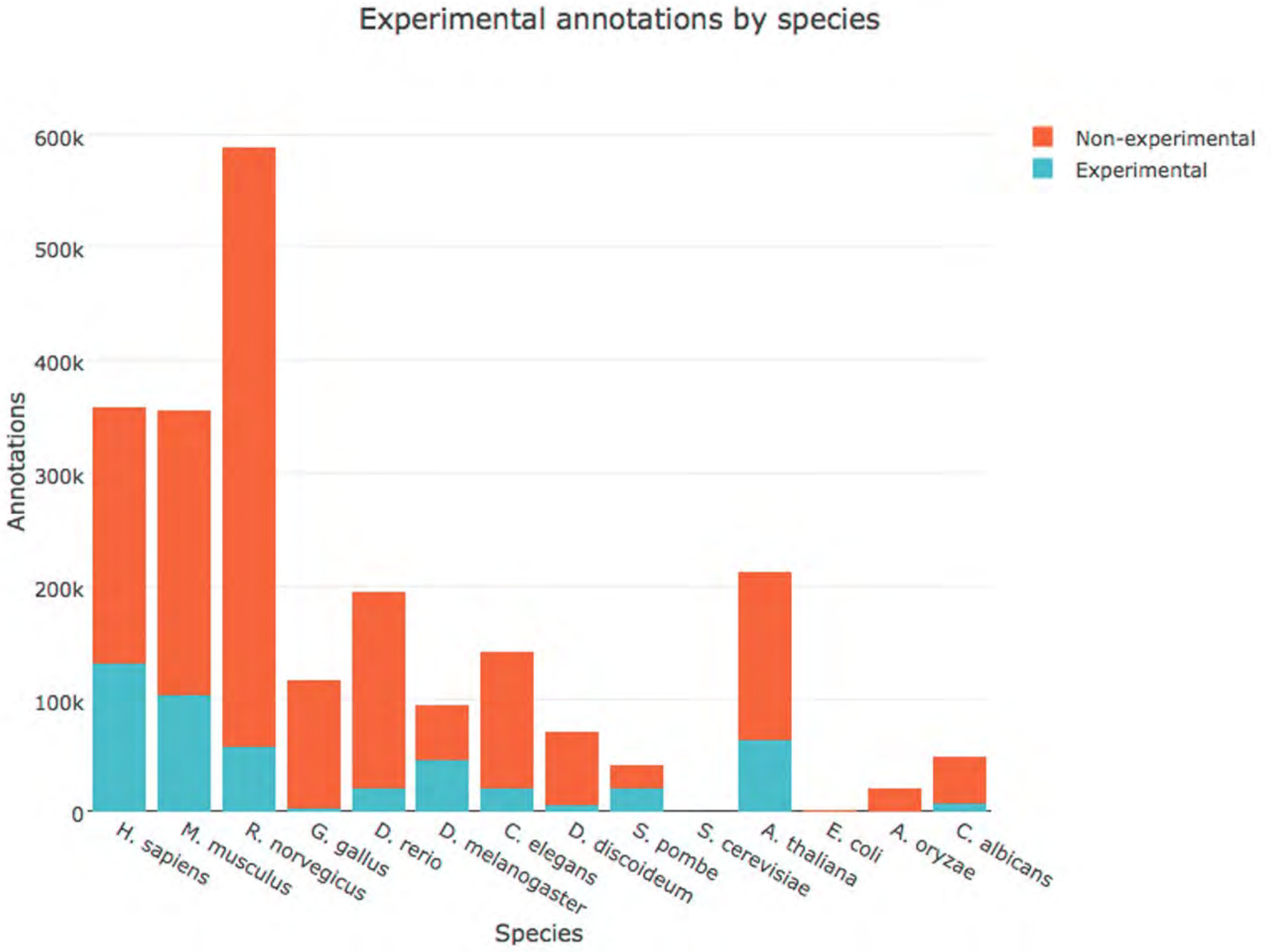
Electron carrier protein. The oxidized form of the cytochrome c heme group can accept an electron from the heme group of the cytochrome c1 subunit of cytochrome reductase. Cytochrome c then transfers this electron to the cytochrome oxidase complex, the final protein carrier in the mitochondrial electron-transport chain.
Plays a role in apoptosis. Suppression of the anti-apoptotic members or activation of the pro-apoptotic members of the Bcl-2 family leads to altered mitochondrial membrane permeability resulting in release of cytochrome c into the cytosol. Binding of cytochrome c to Apaf-1 triggers the activation of caspase-9, which then accelerates apoptosis by activating other caspases.

GO - Molecular function[i]

- electron transporter, transferring electrons from CoQH2-cytochrome c reductase complex and cytochrome c oxidase complex activity Source: UniProtKB
- heme binding Source: UniProtKB
- metal ion binding Source: UniProtKB-KW

GO - Biological process[i]

- activation of cysteine-type endopeptidase activity involved in apoptotic process by cytochrome c Source: UniProtKB
- cellular respiration Source: UniProtKB
- intrinsic apoptotic signaling pathway Source: Reactome
- mitochondrial electron transport, cytochrome c to oxygen Source: GO_Central
- mitochondrial electron transport, ubiquinol to cytochrome c Source: GO_Central
- mitochondrion organization Source: Reactome
- protein dephosphorylation Source: GOC
- response to reactive oxygen species Source: Reactome


Experimental annotations by species
Non-experimental
Experimental
600k
500k
400k
300k
200k
100k
0
Annotations
H. sapiens
M. musculus
R. norvegicus
G. gallus
D. rerio
D. melanogaster
C. elegans
D. discoideum
S. pombe
S. cerevisiae
A. thaliana
E. coli
A. oryzae
C. albicans
Species

What GO is NOT.....

- GO is not an ontology of genes or gene products (so it is somewhat a misnomer): e.g. cytochrome c is not in GO, but attributes of cytochrome c are, e.g. **oxidoreductase activity**.
- Processes, functions and components unique to variants or diseases: e.g. **oncogenesis** is not a valid GO term.
- Protein domains or structural features.
- Protein-protein interactions.
- Environment, evolution and expression.
- Anatomical or histological features above the level of cellular components, including cell types.

More Ontologies

- Sequence Ontology (SO) www.sequenceontology.org
  - Sequence features used in biological sequence annotation
- Protein Ontology (PRO) pir.georgetown.edu/pro/
  - Representation of protein-related entities
- And many more on Phenotypes, Anatomies, Cell structures .....

PRO Hierarchy (Note that the implicit relationship is *is_a*, whereas d indicates *derives_from* relationship.)

8 shown of 208460 records

PMID | Taxon | PANTHER | EcoCyc | Definition
Synonym | Gene | MGI | HGNC | Pfam | PIRSF | Reactome | UniProtKB

expand	sort (ID)	sort (STR)	find	Category
PR:000018263 ***amino acid chain***				*external*
PR:000000001 ***protein***				
PR:000000027 *smad protein*				*family*
PR:000000066 *TGF-beta receptor-regulated smad protein*				*family*
PR:000000364 *smad2*				*gene*
PR:000000468 *smad2 isoform 1*				*sequence*
PR:Q15796-1 *mothers against decapentaplegic homolog 2 isoform Long (human)*				*organism-sequence*
PR:000045371 *smad2 isoform 1 unphosphorylated 1 (human)*				*organism-modification*

The OBO Foundry http://www.obofoundry.org

Download table as: [YAML | JSON-LD | RDF/Turtle]

chebi	Chemical Entities of Biological Interest	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds. Detail
doid	Human Disease Ontology (cc) BY	An ontology for describing the classification of human diseases organized by etiology. Detail
go	Gene Ontology (cc) BY	An ontology for describing the function of genes and gene products Detail
obi	Ontology for Biomedical Investigations (cc) BY	An integrated ontology for the descript and clinical investigations Detail
pato	Phenotypic quality (cc) BY	An ontology of phenotypic qualities (pr characteristics) Detail
po	Plant Ontology (cc) BY	The Plant Ontology is a structured voca resource that links plant anatomy, morp and development to plant genomics da

ID Space	doid
PURL	http://purl.obolibrary.org/obo/doid.owl
License	CC-BY
Homepage	http://www.disease-ontology.org
Contact	Lynn Schriml
Trackers	https://github.com/DiseaseOntology/HumanDiseaseOntology/issues
Domain	disease
Taxon	Homo sapiens
Cite	Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data

Ontology and Terminology Servers

- BioPortal bioportal.bioontology.org
- NCI Metathesaurus ncim.nci.nih.gov/ncimbrowser/
- EMBL-EBI OLS www.ebi.ac.uk/ols/index

Summary

- BioMedical big data needs to be shared
- Sharing data requires high quality:
  - Data Standards
  - Metadata
  - Terminologies and Ontologies
  - *Better tools to apply and work with them*