

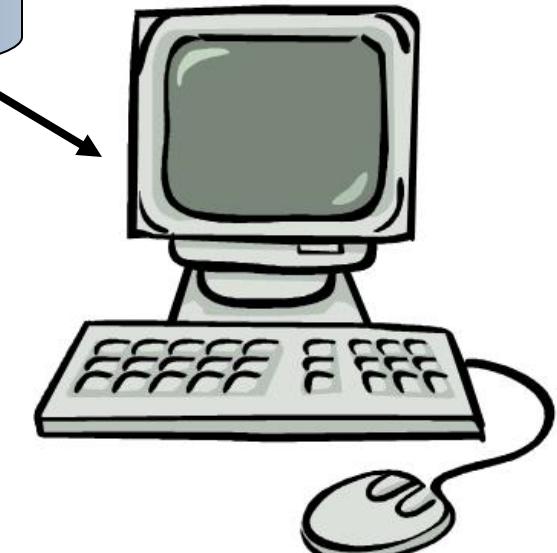
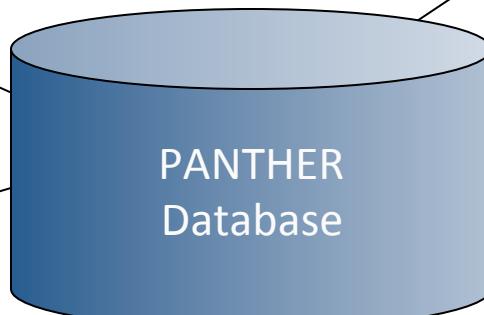
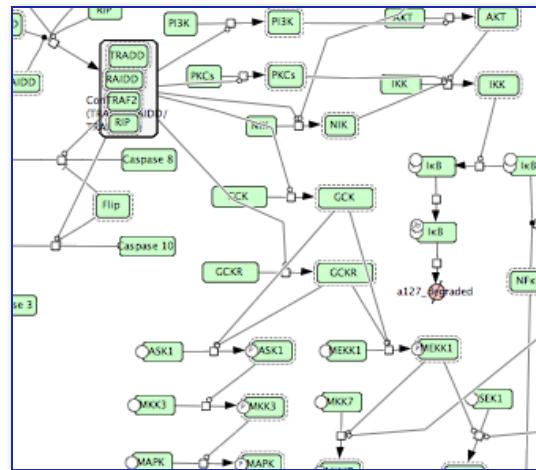


Updates on the PANTHER System

Huaiyu Mi

University of Southern California
Keck School of Medicine

COMBINE 11, Sept. 7, 2011, Heidelberg





Brief History



1998

Start of the project

Brief History



1998 → 2001

Start of the project

Participated in the human genome sequence project.



Brief History



Established methods for
large scale gene
expression analysis and
cSNP analysis
Clark AG. 2003, Science
Thomas PD, 2004, PNAS

1998 → 2001 → 2003

Start of the
project

Participated in the
human genome
sequence project.



Brief History



Established methods for
large scale gene
expression analysis and
cSNP analysis
Clark AG. 2003, Science
Thomas PD, 2004, PNAS

1998 → 2001 → 2003 → 2004

Start of the
project

Participated in the
human genome
sequence project.



Launch PANTHER
pathway system



GOOG ~\$100

GOOG > \$600

Brief History



Established methods for
large scale gene
expression analysis and
cSNP analysis
Clark AG. 2003, Science
Thomas PD, 2004, PNAS

1998

2001

2003

2004

2006

Start of the
project

Participated in the
human genome
sequence project.



Completed the
PANTHER tool
suite.
Thomas PD, 2006,
NAR.



GOOG ~\$100

Brief History



1998

Start of the project

2001

Participated in the human genome sequence project.

2003

Established methods for large scale gene expression analysis and cSNP analysis
Clark AG. 2003, Science
Thomas PD, 2004, PNAS



2004

Launch PANTHER pathway system



Completed the PANTHER tool suites.
Thomas PD, 2006, NAR.

2006



2008

Integrated the PANTHER into GO Reference Genome curation pipeline.
2009, PLoS Comput. Biology
Gaudet P, 2011, Brief Bioinformatics



Brief History



Established methods for large scale gene expression analysis and cSNP analysis
Clark AG. 2003, Science
Thomas PD, 2004, PNAS

1998

Start of the project

2001

Participated in the human genome sequence project.

2003



Launch PANTHER pathway system

2004



GOOG ~\$100



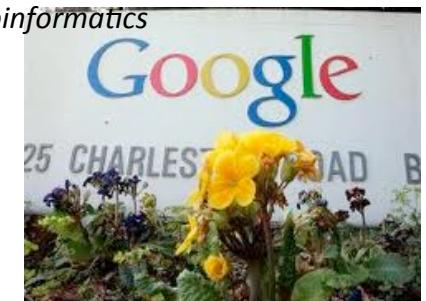
Completed the PANTHER tool suites.
Thomas PD, 2006, NAR.

2006

2008

Integrated the PANTHER into GO Reference Genome curation pipeline.
2009, PLoS Comput. Biology
Gaudet P, 2011, Brief Bioinformatics

2010



GOOG > \$600





Collaborators and Partners



CellDesigner.org

[YPED Repository](http://Yped.org)



[the Gene Ontology](http://Geneontology.org)

[NCBI](http://NCBI.nlm.nih.gov)



[UniProt](http://UniProt.org)
the universal protein resource

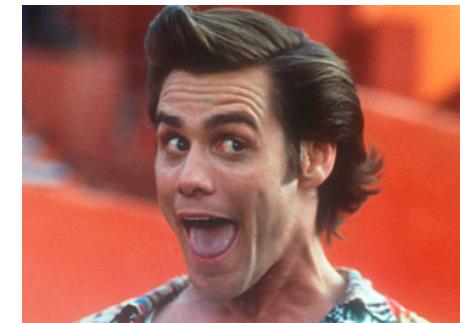
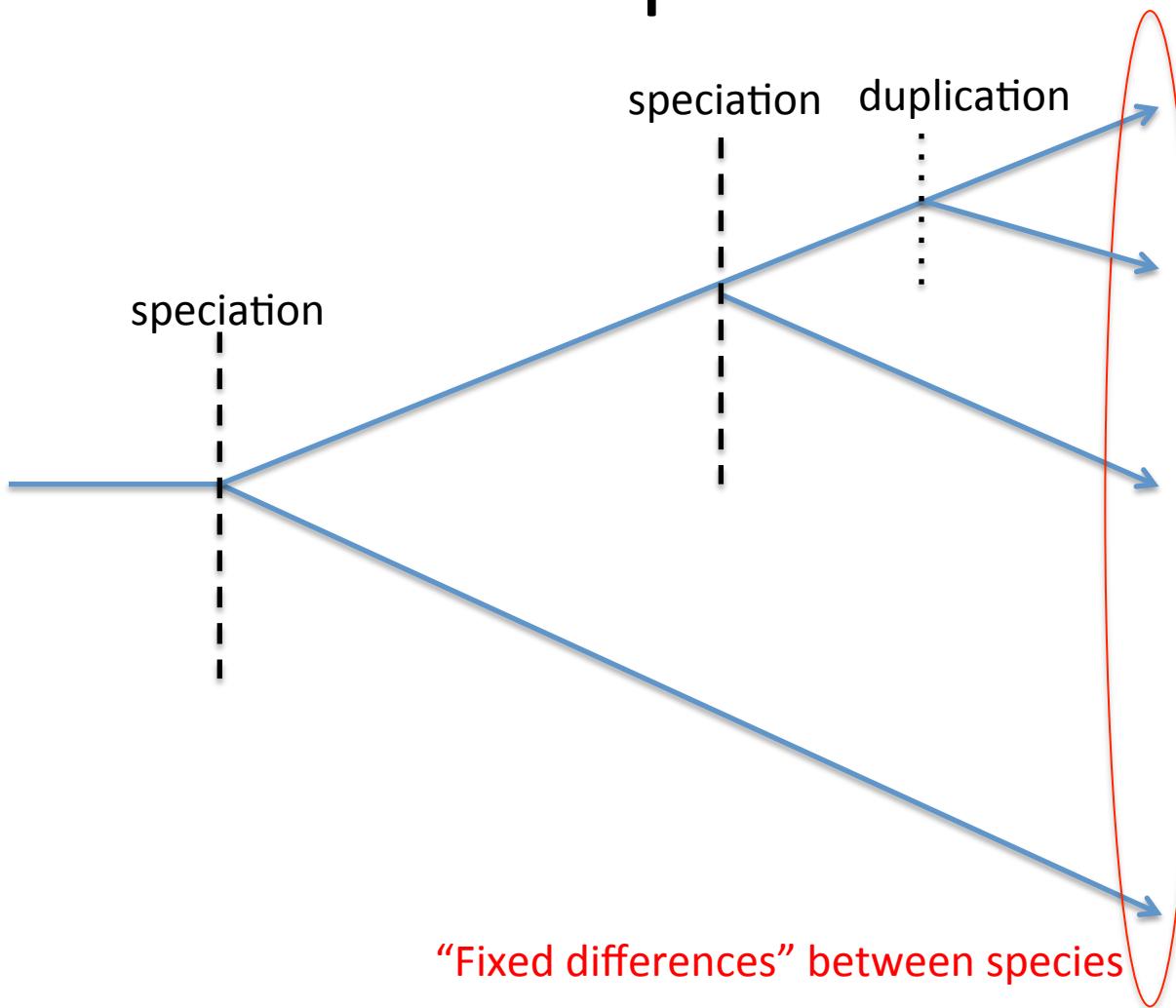


GIGA

- An algorithm that makes phylogenetic inferences under the constraint of the species tree.
- Use sequence-based distance from multiple sequence alignment at each step.
 - Speciation
 - Duplication
 - Ortholog group (subfamily)

Thomas, 2010 BMC Bioinformatics, 11:312

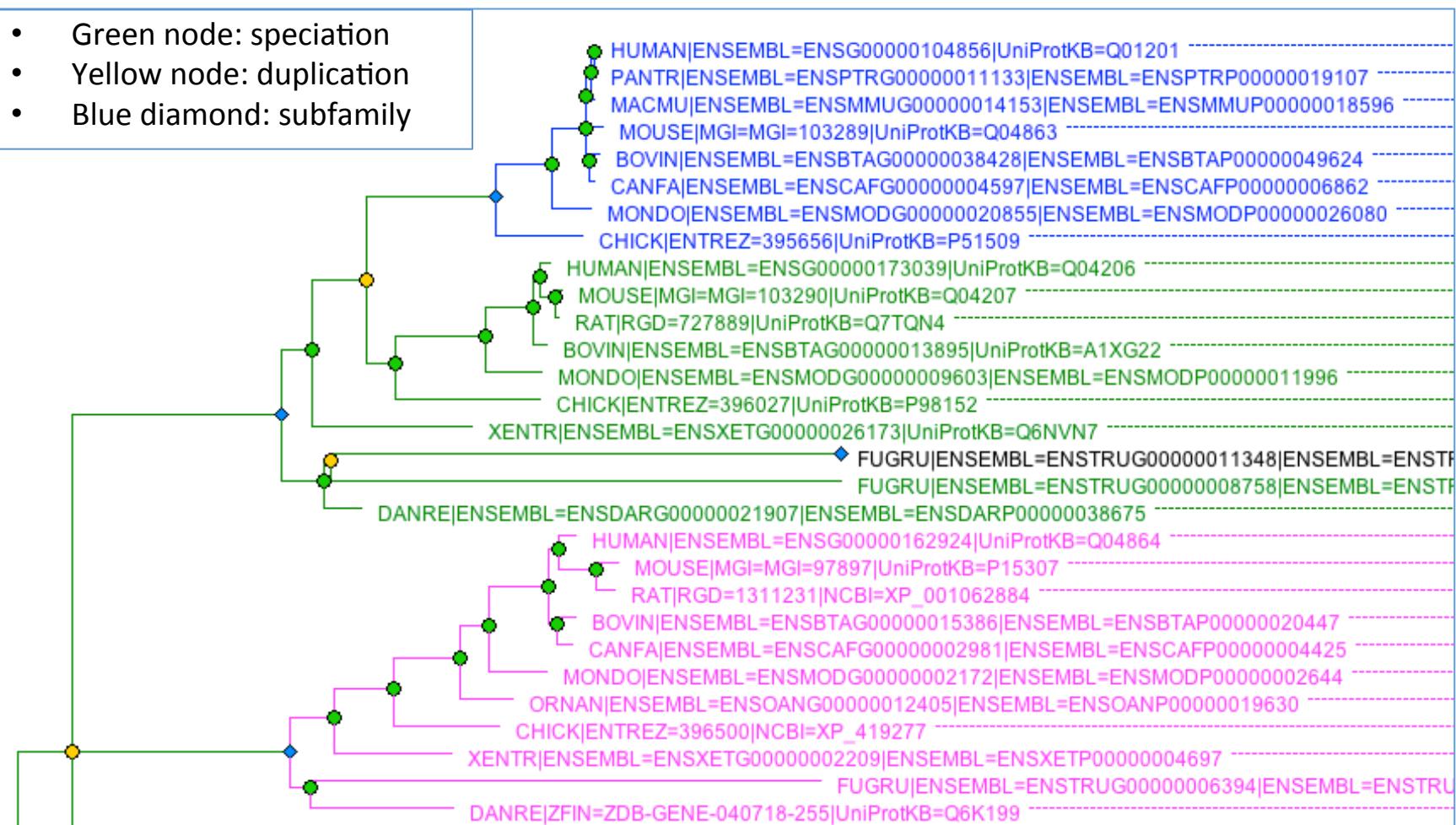
Phylogenetic inferences based on species tree



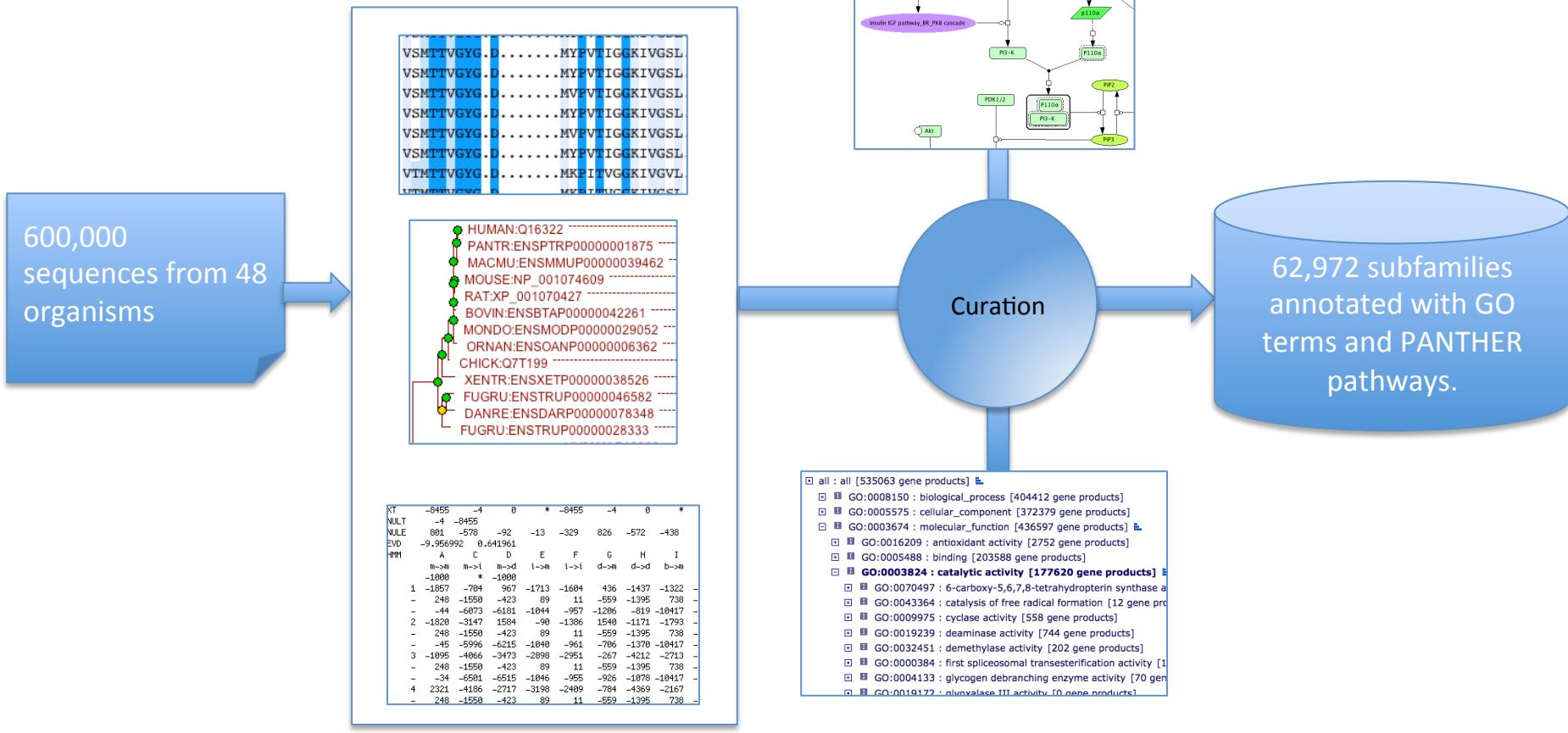
PANTHER Phylogenetic Tree

Tree from PTHR24169

- Green node: speciation
- Yellow node: duplication
- Blue diamond: subfamily

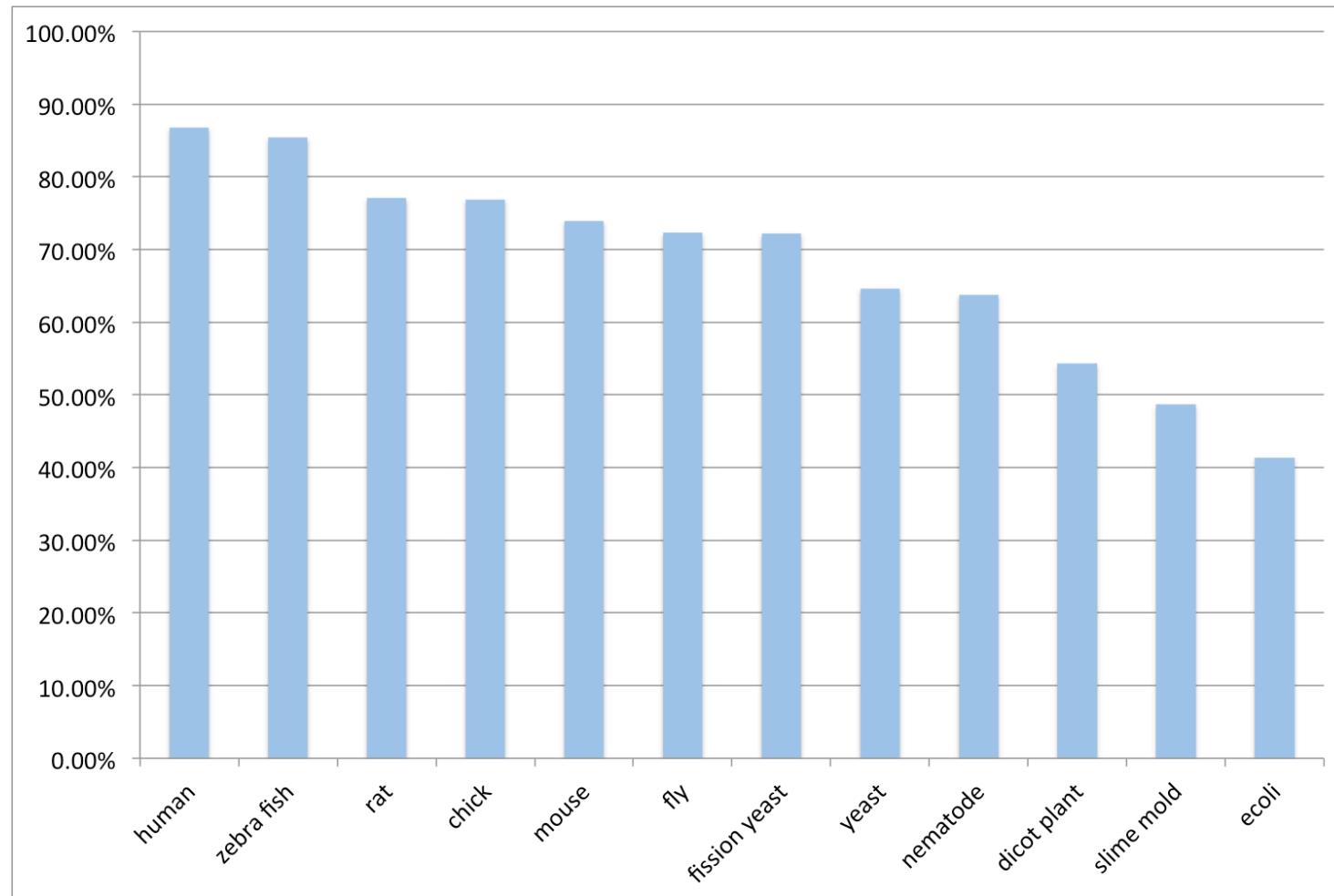


PANTHER Protein Library Building



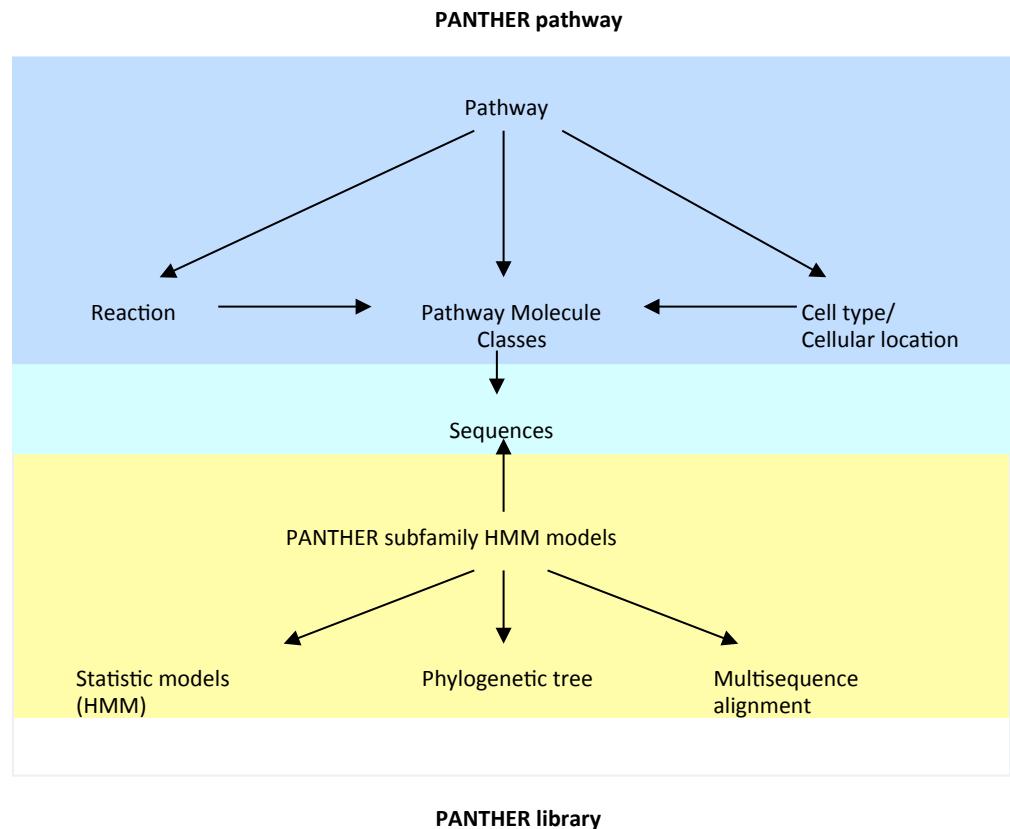
400,000 sequences
In 6594 family clusters

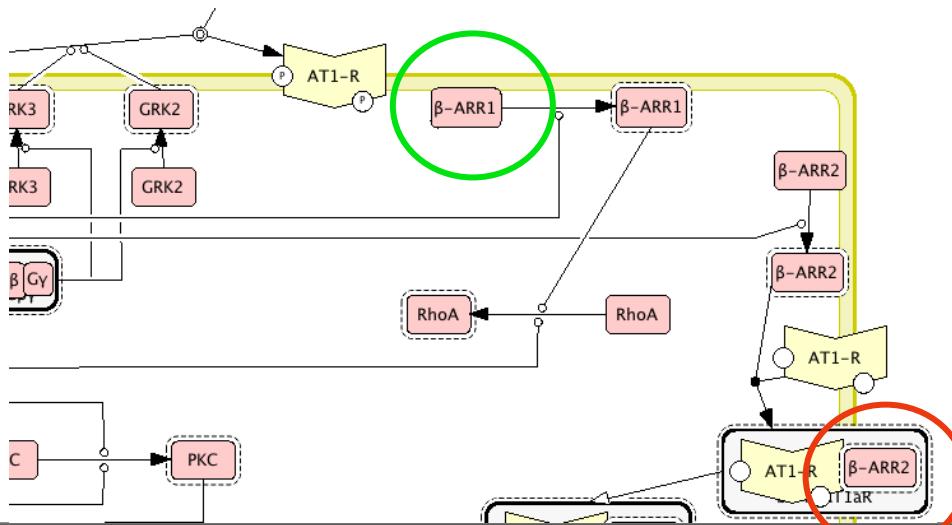
Genome Coverage in PANTHER



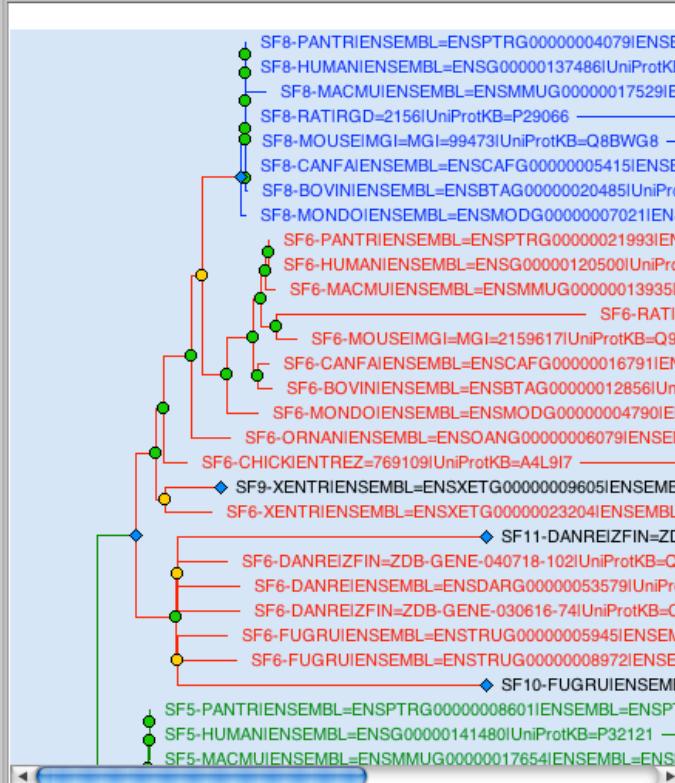
PANTHER Pathway Data Structure

- A pathway diagram
 - Curate the pathway
 - Display the pathway
- Unambiguous graphical representation of pathway data
- Structured data for pathway
- Link pathway classes to the sequence database





Tree



Grid

definition	organism
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Pan troglodytes
Beta-arrestin-1	Homo sapiens
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Macaca mulatta
Beta-arrestin-1	Rattus norvegicus
Beta-arrestin-1	Mus musculus
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Canis familiaris
Beta-arrestin-1	Bos taurus
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Monodelphis domestica
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Pan troglodytes
Arrestin-C	Homo sapiens
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Macaca mulatta
Arrestin-C (Fragment)	Rattus norvegicus
Arrestin-C	Mus musculus
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Canis familiaris
Arrestin-C	Bos taurus
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Monodelphis domestica
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Ornithorhynchus anatinus
Cone arrestin	Gallus gallus
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Xenopus tropicalis
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Xenopus tropicalis
ARRB1 protein (Fragment)	Danio rerio
Arrestin 3, retinal (X-arrestin), like	Danio rerio
Arrestin 3, retinal (X-arrestin) (Fragment)	Danio rerio
Arrestin 3, retinal (X-arrestin)	Danio rerio
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Fugu rubripes
Arrestin-C (Cone arrestin)(C-arrestin)(cArr)(Retinal cone arre...	Fugu rubripes
Beta-arrestin-1 (Arrestin beta-1) [Source:UniProtKB/Swiss-Pro...]	Fugu rubripes
Beta-arrestin-2 (Arrestin beta-2) [Source:UniProtKB/Swiss-Pro...	Pan troglodytes
Beta-arrestin-2	Homo sapiens
Beta-arrestin-2 (Arrestin beta-2) [Source:UniProtKB/Swiss-Pro...	Macaca mulatta

MSA

The MSA (Multiple Sequence Alignment) panel shows the sequence alignment for the highlighted proteins. The sequences are color-coded by species, matching the tree. Red arrows point to the 'Beta-arrestin-1' and 'Beta-arrestin-2' entries in the grid table above.

User Statistics

- 600 unique IPs every day.
- 5000 regular users (IP) each month.
- >90 countries around the world.
- 5000 page views per day.
- Cited in ~2500 publications.

PANTHER PATHWAY IN BIOPAX

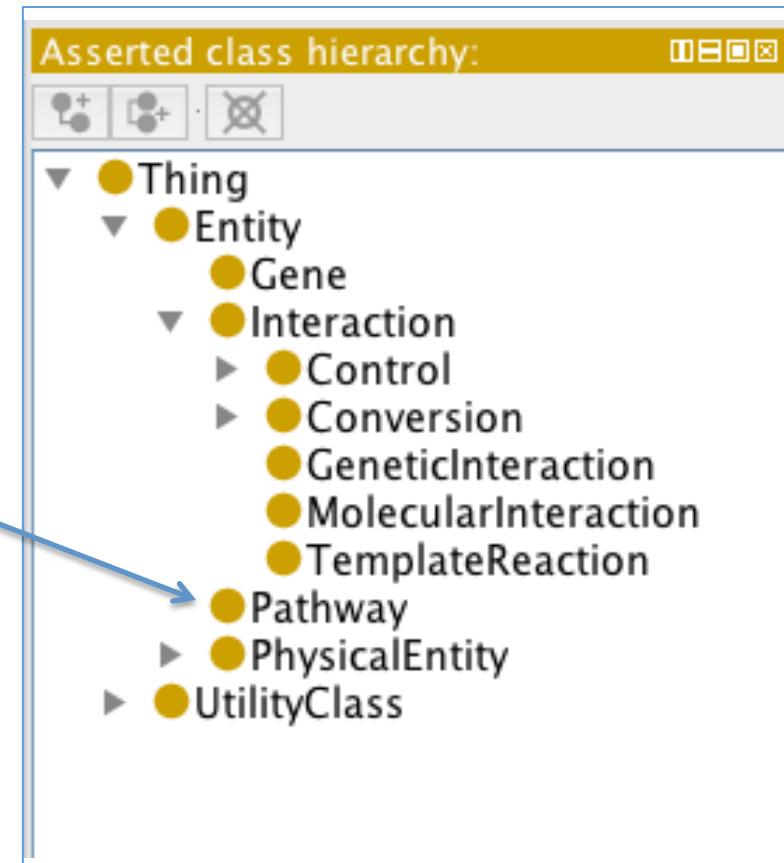
Motivation

- Most BioPAX files are generated by pathway database resources. The curation usually requires highly trained curators.
- To provide a graphical editing tool that can export data directly to BioPAX.
- PANTHER user community has requested to have PANTHER pathways available in BioPAX format.
 - Database: exchange
 - Data analysis: tools
 - Modeling
- PANTHER pathways are generated using CellDesigner, and are stored in PD format, which can be easily converted to BioPAX data structure.
- CellDesigner is widely used in the community for pathway curation.
- Approach: Build a CellDesigner to BioPAX translator.
- Long term goal is to build a seamless conversion between SBGN and BioPAX.

Mapping CellDesigner Notations to BioPAX Classes

CellDesigner notation

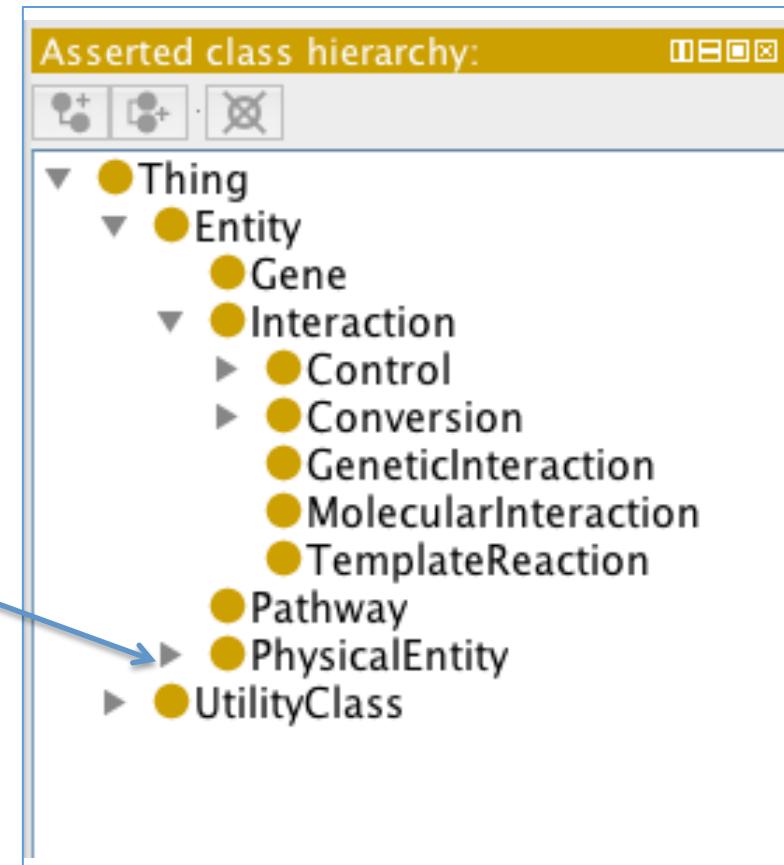
- Model
- Species
- Arcs
 - State transition arc
 - Modification arc



Mapping CellDesigner Notations to BioPAX Classes

CellDesigner notation

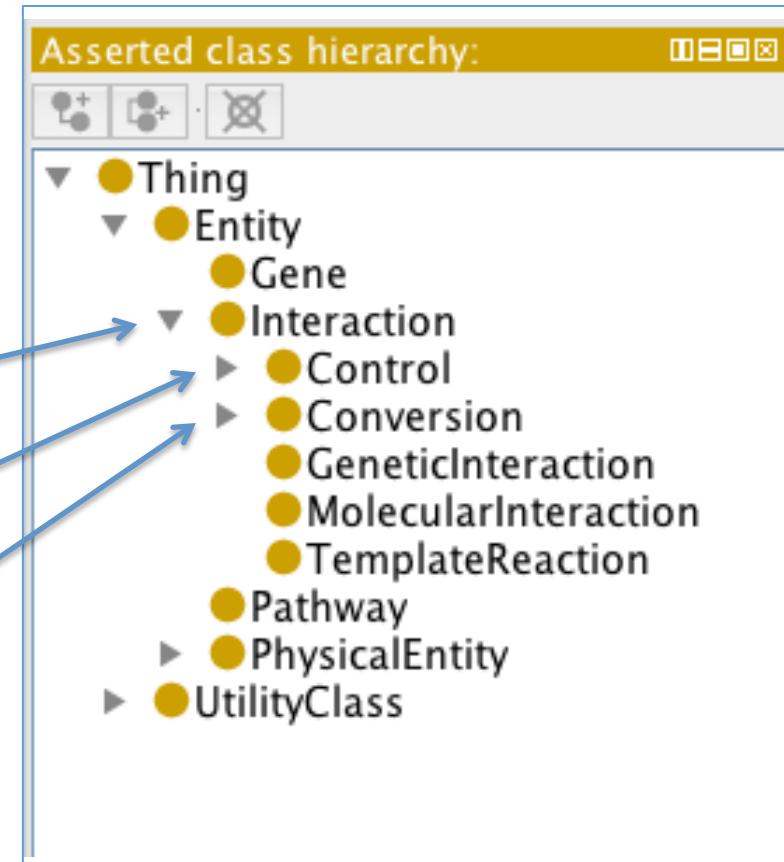
- Model
- Species
- Arcs
 - State transition arc
 - Modification arc



Mapping CellDesigner Notations to BioPAX Classes

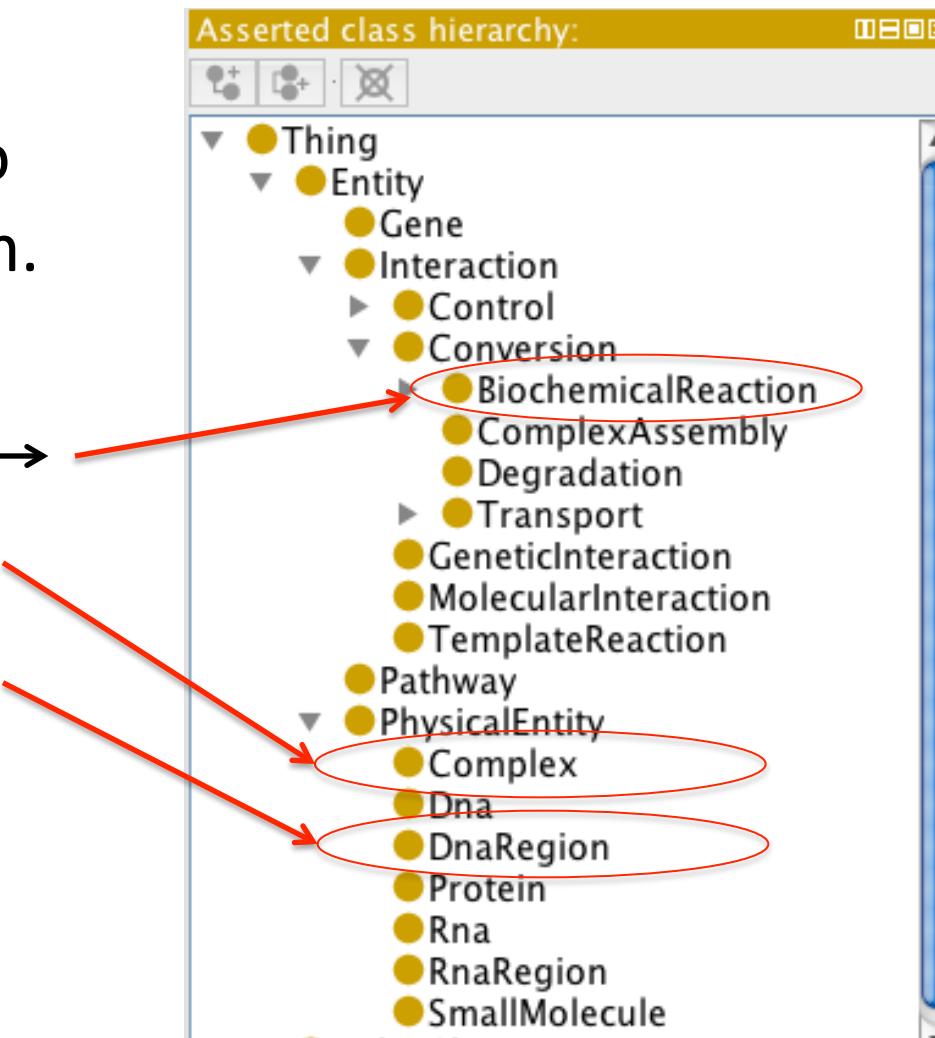
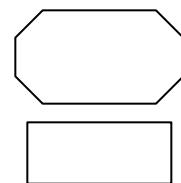
CellDesigner notation

- Model
- Species
- Arcs
 - Modification arc
 - Transition arc



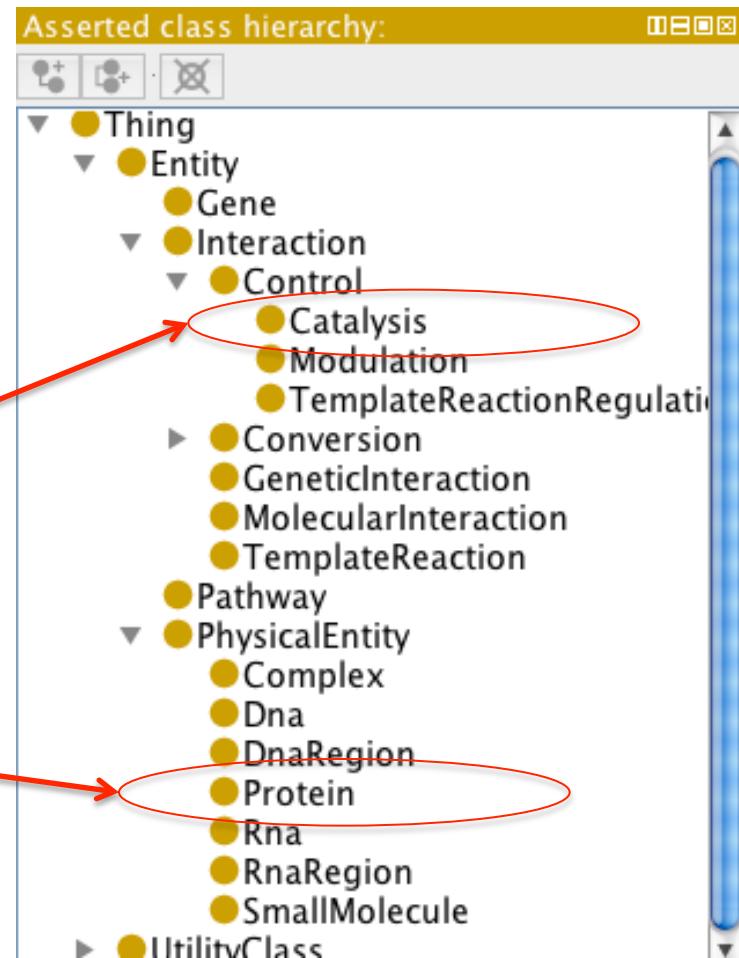
Exact map between CellDesigner notation and BioPAX term

- CellDesigner notations that have 1-1 match to a BioPAX subclass term.
- E.g.
 - State translation
 - Complex
 - Gene



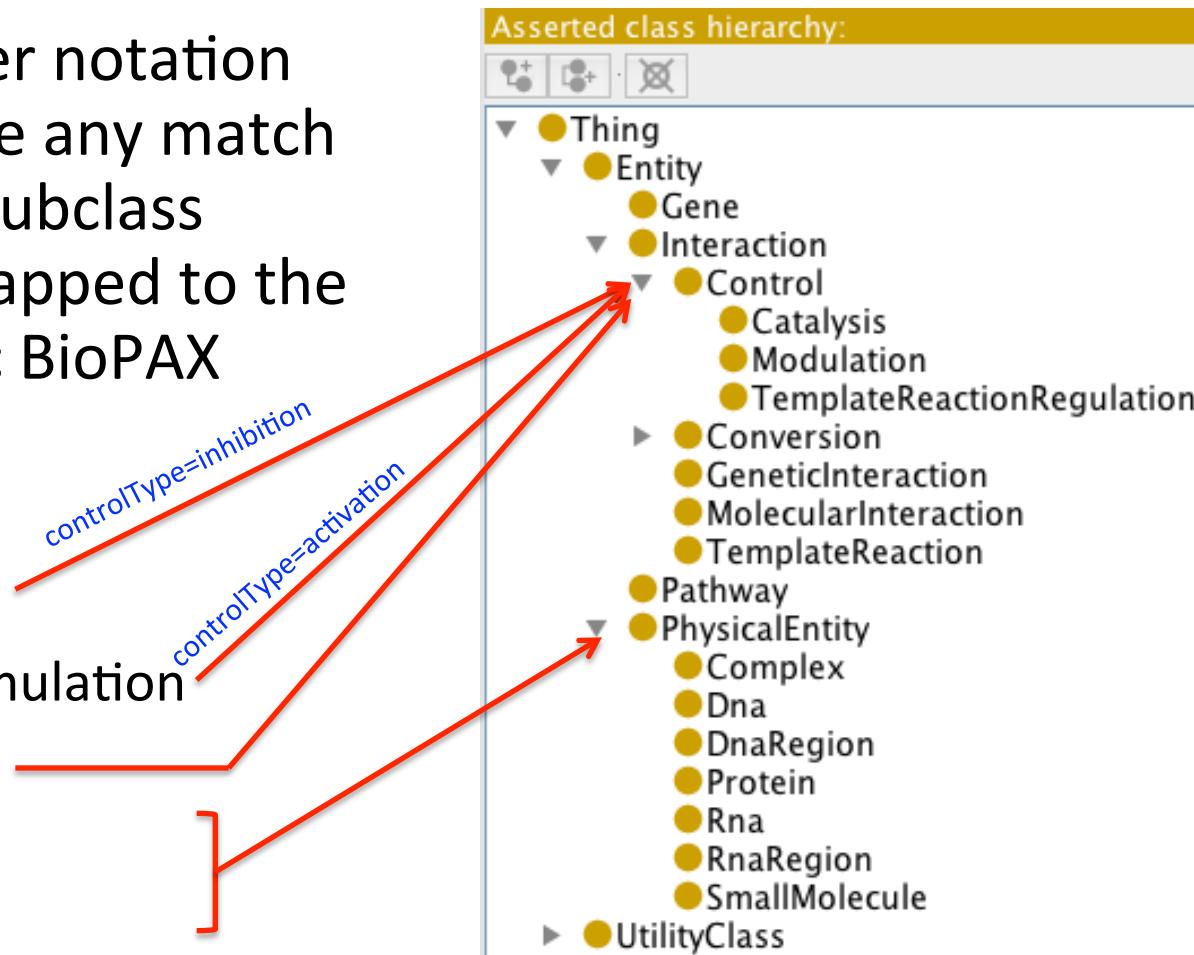
Multiple CellDesigner notations are mapped to a single BioPAX term

- Multiple CellDesigner notations mapped to a single BioPAX term
- E.g.
 - Catalysis
 - Unknown catalysis
 - Protein generic
 - Ion channel
 - receptor



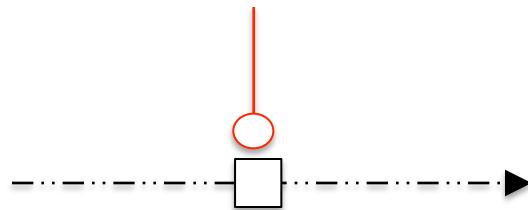
CellDesigner notations are mapped to the generic class term

- A CellDesigner notation does not have any match to a BioPAX subclass term. It is mapped to the more generic BioPAX term.
- E.g.,
 - Inhibition
 - Physical stimulation
 - Modulation
 - Drug
 - Unknown

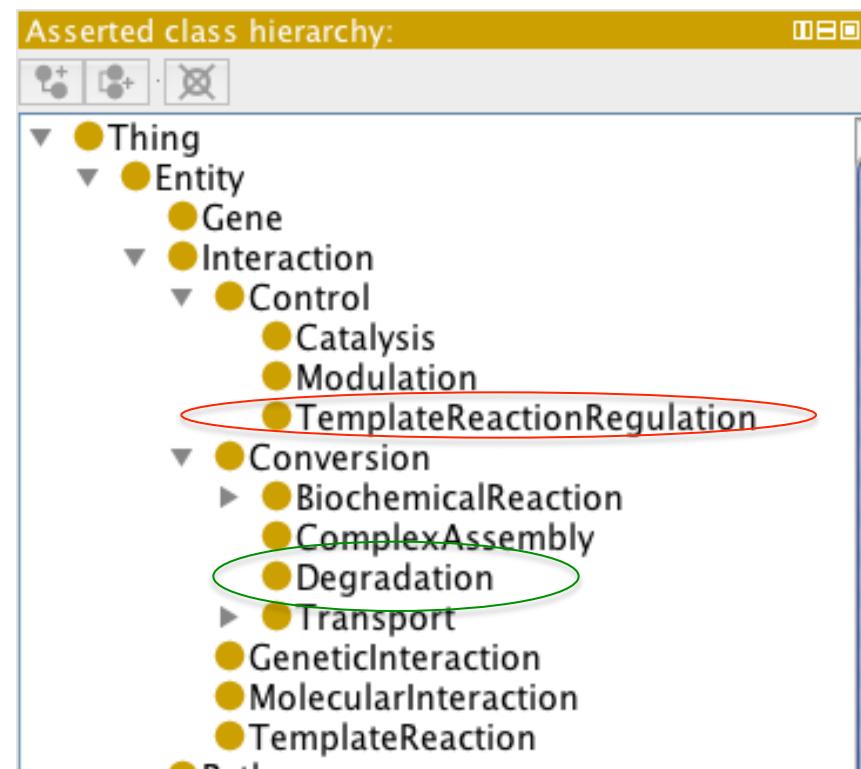
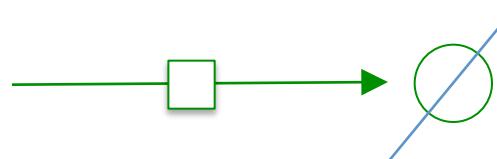


Special handling

- A modulation arc that controls transcription or translation



- State transition that leads to a sink

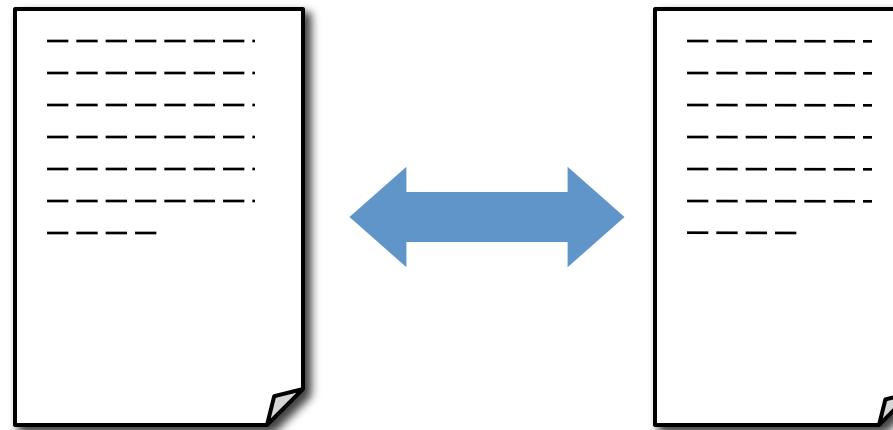


Controlled Vocabulary

- Use controlled vocabularies
 - Gene ontology (cellular component)
 - ChEBI
 - PSI-MI
 - Sequence ontology

CellDesigner BioPAX Converter

- Uses CellDesigner library to create input data types.
- Map to PAXTool library data type.
- Writes the PAXTools library to output an OWL file.



PAXTool

The BioPAX translator plugin

- Use the plugin infrastructure
- BioPAX validator is used throughout the development as a reference.
- Frequent updates based on the updates of CellDesigner, BioPAX, Paxtool and the validator.
- Plugin download:
 - <ftp://ftp.pantherdb.org/CellDesigner/plugins/BioPAX/>
- BioPAX file download:
 - <http://www.pantherdb.org/pathway/pathwayList.jsp>

Acknowledgements

- USC
 - Paul Thomas
 - Anushya Muruganujan
 - Stan Dong
- BioPAX community
 - Emek Demir
 - Igor Rodchenkov
- SBI
 - Hiroaki Kitano
 - Yukiko Matsuoka
 - Akira Funahashi
- GO community
 - Suzi Lewis
 - Pascal Gaudet