

生物信息学：导论与方法

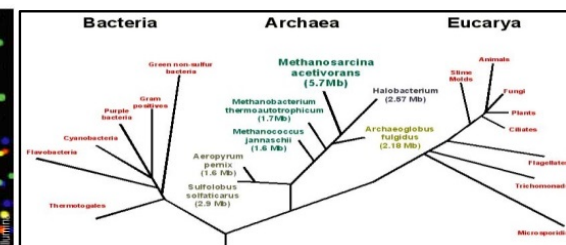
Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍

Center for Bioinformatics, Peking University



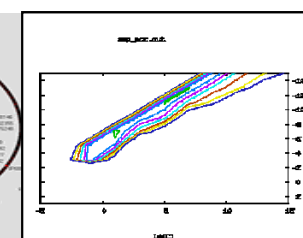
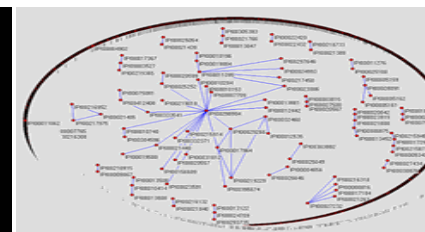
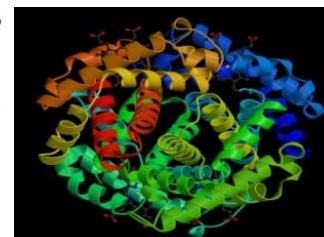
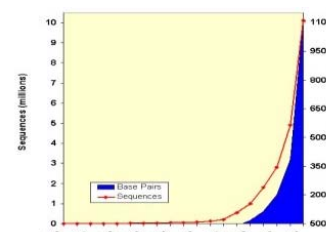
<https://www.coursera.org/course/pkubioinfo>



北京大学生物信息学中心 魏丽萍

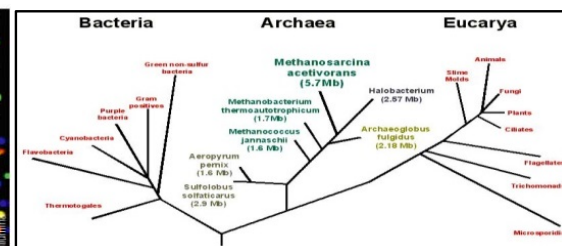
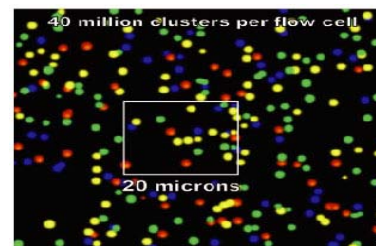
Liping Wei, Ph.D.

Center for Bioinformatics, Peking University





TAACCCTAACCCTAACCCTAACCCTAACCCTA
CCTAACCCTAACCCTAACCCTAACCCTAACC
CCCTAACCCTAACCCTAACCCTAACCCTAAC
AACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCTAACCCTAAC
CTACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCTAACCCTAAC

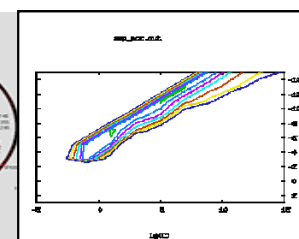
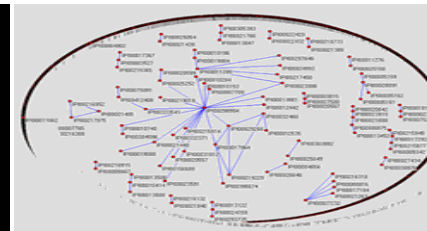
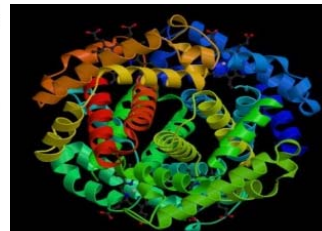
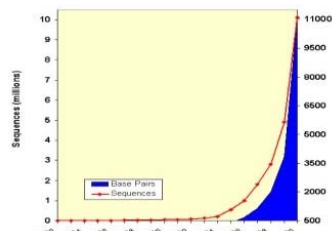
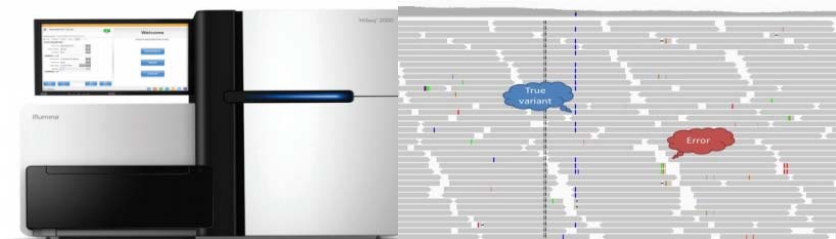


Unit 1: Ontology and Gene Ontology

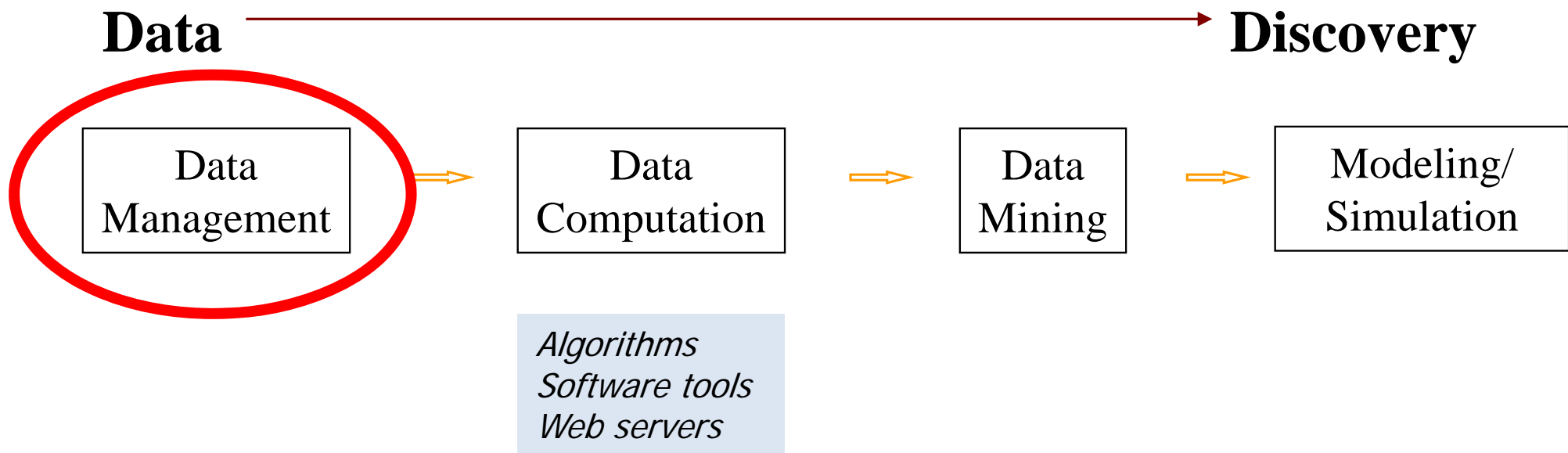
北京大学生物信息学中心 魏丽萍

Liping Wei, Ph.D.

Center for Bioinformatics, Peking University



The –informatics in Bioinformatics



How can a computer know?!

wnt1

wnt-1

int1

WINGLESS-TYPE MMTV INTEGRATION SITE FAMILY, MEMBER 1

WINGLESS-TYPE MMTV INTEGRATION SITE FAMILY

wingless

wg



http://commons.wikimedia.org/wiki/File:Wingless_Drosophila_hydei.jpg







Copyright © Peking University

Another example: delta DNA polymerase

Saccharomyces cerevisiae:
cdc2 (now *pol3*)

Drosophila melanogaster:
DNApol-delta

Mus musculus: *pold1*

 POLD1, <i>H.sapiens</i> polymerase (DNA directed), delta 1, catalytic subunit	 NP_002682.2 1107 aa
 POLD1, <i>P.troglodytes</i> polymerase (DNA directed), delta 1, catalytic subunit	 XP_003316601.2 648 aa
 POLD1, <i>M.mulatta</i> polymerase (DNA directed), delta 1, catalytic subunit	 XP_001116065.1 1327 aa
 POLD1, <i>C.lupus</i> polymerase (DNA directed), delta 1, catalytic subunit	 XP_851285.1 1107 aa
 POLD1, <i>B.taurus</i> polymerase (DNA directed), delta 1, catalytic subunit	 NP_776852.1 1106 aa
 Pold1, <i>M.musculus</i> polymerase (DNA directed), delta 1, catalytic subunit	 NP_035261.3 1105 aa
 Pold1, <i>R.norvegicus</i> polymerase (DNA directed), delta 1, catalytic subunit	 NP_067694.1 1103 aa
 pold1, <i>D.rerio</i> polymerase (DNA directed), delta 1, catalytic subunit	 NP_001034899.1 1105 aa
 DNApol-delta, <i>D.melanogaster</i> DNA-polymerase-delta	 NP_524099.2 1092 aa
 F10C2.4, <i>C.elegans</i> Protein F10C2.4	 NP_506017.1 1081 aa
 POL3, <i>S.cerevisiae</i> Pol3p	 NP_010181.2 1097 aa
 KLLA0E01607g, <i>K.lactis</i> hypothetical protein	 XP_454020.1 1101 aa
 AGOS_AFL189W, <i>E.gossypii</i> AFL189Wp	 NP_985361.1 1092 aa
 cdc6, <i>S.pombe</i> DNA polymerase delta catalytic subunit Cdc6	 NP_596124.1 1086 aa
 MGG_08071, <i>M.musculus</i>	 XP_362488.1

We need to define it for the computer!

- Entity: name, synonyms, acronyms, properties
- Relationship: e.g., is-a

hierarchical, common, controlled vocabulary

the use of predefined, authorized terms that have been preselected by the designer of the vocabulary, in contrast to free natural language vocabularies.

Ontology

A specification of a conceptualization

<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

A set of concepts within a domain, defined by a shared vocabulary to denote the types and properties of the concepts as well as the relationships between the concepts

In philosophy: the study of the nature of being, becoming, existence, or reality, as well as the basic categories of being and their relations

What does an ontology enable?

Communication: to be able to communicate unambiguously

e.g., to understand across different groups' annotations of various genomes

Computation: to be able to represent knowledge in a computable form

e.g., to represent literature and data in a structured form to enable automated analyses by computer programs

Discovery of Patterns: to be able to traverse different hierarchies

e.g., to go above a set of individual genes to find the larger functional categories or pathways involved, with a bird's-eye view.

Open Biomedical Ontologies (OBO)

Gene Ontology

Anatomical Entity Ontology

Disease Ontology

Sequence Ontology

System Biology Ontology

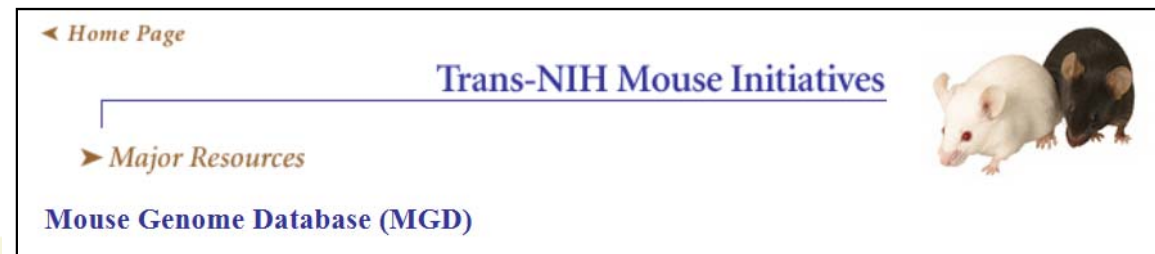
.....

<http://www.obofoundry.org/>

Gene Ontology (GO)

Ashburner *et al.*, *Nat. Genet.*, 2000

- The project began as a collaboration between three model organism genome databases: FlyBase, *Saccharomyces* Genome Database (SGD), and Mouse Genome Database (MGD) in 1998.
- Defines a structured, common, controlled vocabulary to describe attributes of genes and gene products across organisms.



GO consortium

<http://www.geneontology.org/>

Berkeley Bioinformatics Open-source Project (BBOP)
British Heart Foundation - University College London
dictyBase
EcoliWiki
FlyBase
GeneDB
GO Editorial Office at the EBI
Gramene
Institute of Genome Sciences, Univ. of Maryland
InterPro

J Craig Venter Institute
Mouse Genome Informatics (MGI)
Pombase
Rat Genome Database (RGD)
Reactome
Saccharomyces Genome Database (SGD)
The Arabidopsis Information Resource (TAIR)
UniProtKB-Gene Ontology Annotation
(UniProtKB-GOA)
WormBase
The Zebrafish Information Network (ZFIN)

Three categories

Molecular Function = elemental activity/task

the tasks performed by individual gene products; examples are carbohydrate binding and ATPase activity

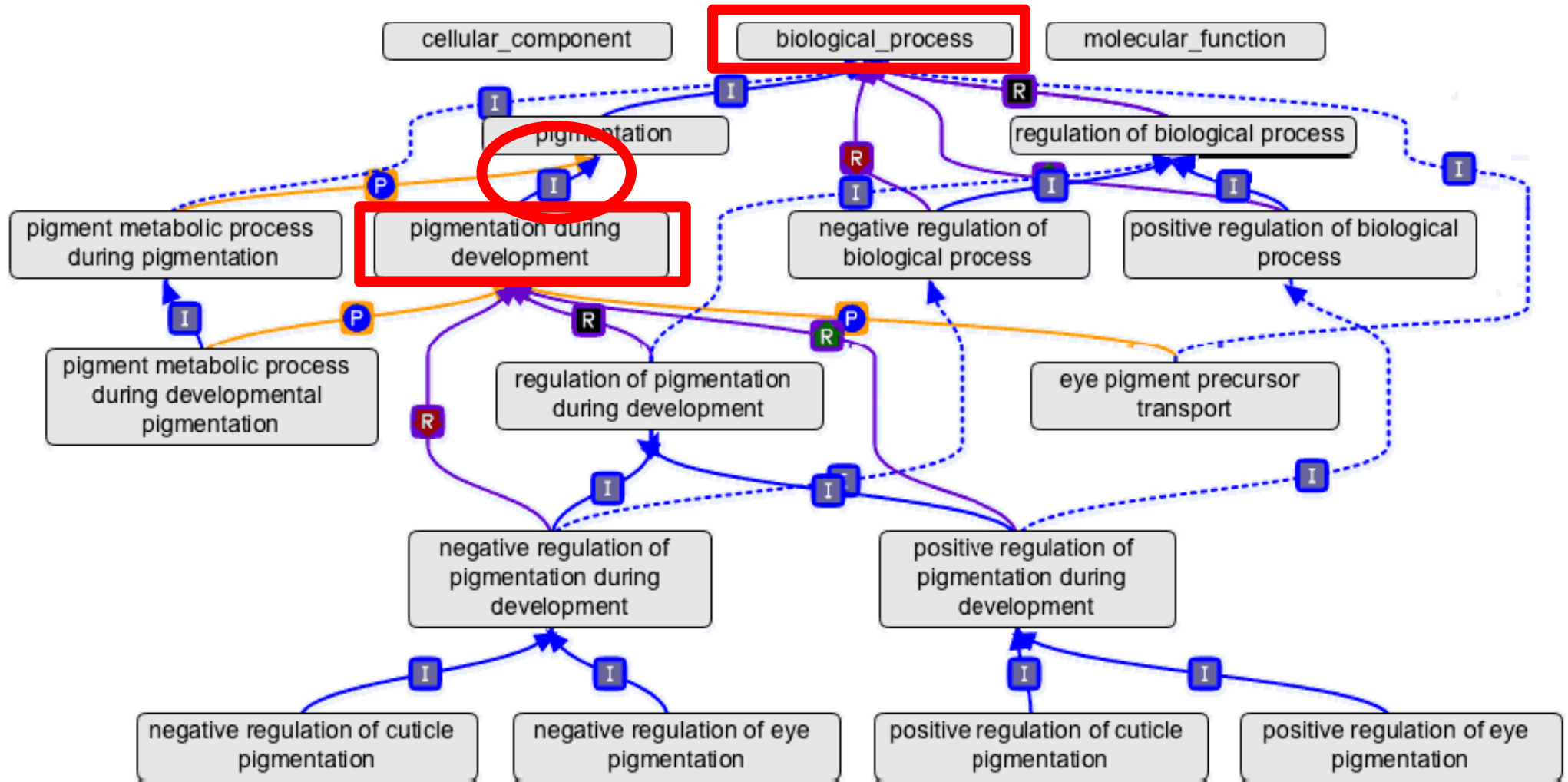
Biological Process = biological goal or objective

broad biological goals, such as mitosis or purine metabolism, that are accomplished by ordered assemblies of molecular functions

Cellular Component = location or complex

Subcellular structures, locations, and macromolecular complexes; examples include nucleus, telomere, and RNA polymerase II holoenzyme

GO structure: Directed Acyclic Graph (DAG)



<http://www.geneontology.org/images/diag-ontology-graph.gif>

Copyright © Peking University

How to store this graph in a computer?

OBO File Format

[Term]

id

name

namespace

def

synonym

is_a

OBO format example

```
[Term]
id: GO:0000001
name: mitochondrion inheritance
namespace: biological_process
def: "The distribution of mitochondria, including the mitochondrial genome, into daughter cells after mitosis mediated by interactions between mitochondria and the cytoskeleton." [GOC:mcc, PMID:10873824, PMID:11389764]
synonym: "mitochondrial inheritance" EXACT []
is_a: GO:0048308 ! organelle inheritance
is_a: GO:0048311 ! mitochondrion distribution

[Term]
id: GO:0000002
name: mitochondrial genome maintenance
namespace: biological_process
def: "The maintenance of the structure and integrity of the mitochondrial genome; includes replication and segregation of the mitochondrial chromosome." [GOC:ai, GOC:vw]
is_a: GO:0007005 ! mitochondrion organization

[Term]
id: GO:0000003
name: reproduction
namespace: biological_process
```

XML format

go:term

go:accession

go:name

go:synonym

go:definition

go:isa

go:dbxref

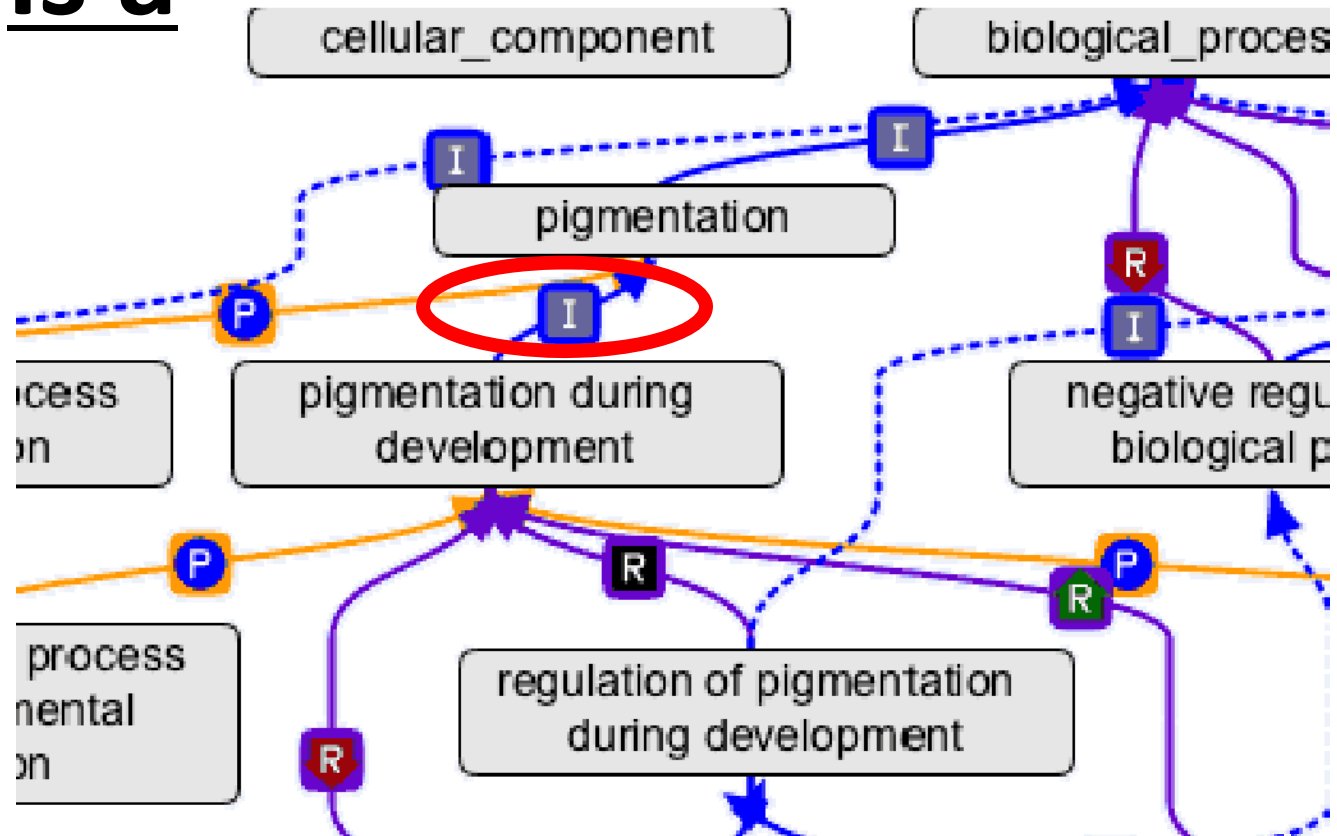
RDF-XML format example

```
<go:term rdf:about="http://www.geneontology.org/go#GO:0000001">
  <go:accession>GO:0000001</go:accession>
  <go:name>mitochondrion inheritance</go:name>
  <go:synonym>mitochondrial inheritance</go:synonym>
  <go:definition>The distribution of mitochondria, including the mitochondrial genome, into daughter cells after mitosis or meiosis, mediated by interactions between mitochondria and the cytoskeleton.</go:definition>
  <go:is_a rdf:resource="http://www.geneontology.org/go#GO:0048308" />
  <go:is_a rdf:resource="http://www.geneontology.org/go#GO:0048311" />
</go:term>

<go:term rdf:about="http://www.geneontology.org/go#GO:0000002">
  <go:accession>GO:0000002</go:accession>
  <go:name>mitochondrial genome maintenance</go:name>
  <go:definition>The maintenance of the structure and integrity of the mitochondrial genome; includes replication and segregation of the mitochondrial chromosome.</go:definition>
  <go:is_a rdf:resource="http://www.geneontology.org/go#GO:0007005" />
  <go:dbxref rdf:parseType="Resource">
    <go:database_symbol>InterPro</go:database_symbol>
    <go:reference>IPR009446</go:reference>
  </go:dbxref>
  <go:dbxref rdf:parseType="Resource">
    <go:database_symbol>InterPro</go:database_symbol>
    <go:reference>IPR016610</go:reference>
  </go:dbxref>
```

GO relationship: is a

“B is a A”: B is a subtype of A



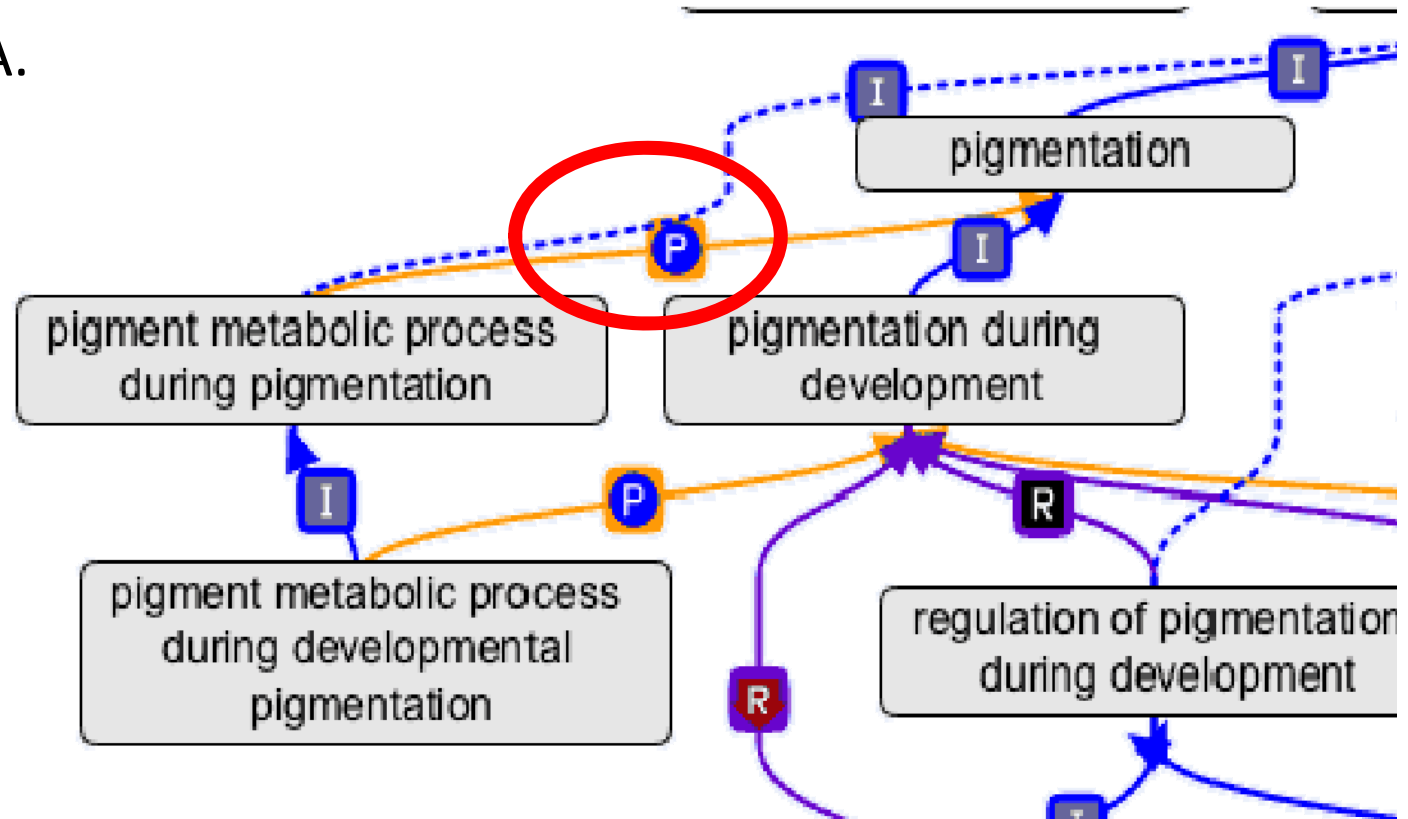
Examples:

mitochondrion inheritance is an organelle inheritance

pigmentation during development is a pigmentation

GO relationship: part of

B **part of** A: B is a part of A.



Examples:

ribosomal large subunit assembly is **part of** ribosome assembly

pigment metabolic process during pigmentation is **part of** pigmentation

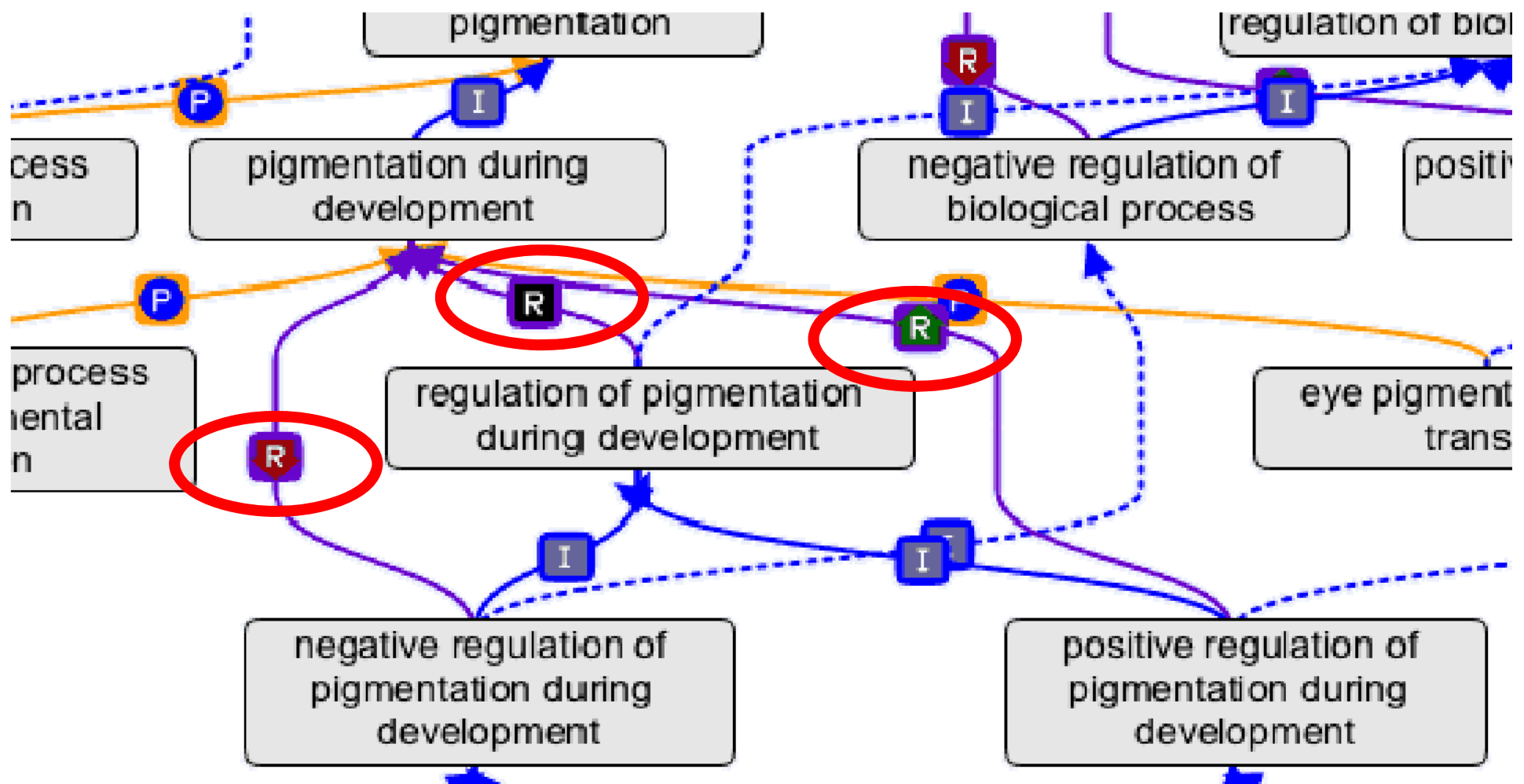
GO relationship: regulates

B regulates A

Sub-relationships:

positively regulates

negatively regulates



Reasoning over the relationships

A is a B ◦ B is a C → A is a C

A is a B ◦ B part of C → A part of C

A part of B ◦ B is a C → A part of C

A part of B ◦ B part of C → A part of C

A is a B ◦ B regulates C → A regulates C

A regulates B ◦ B is a C → A regulates C

A regulates B ◦ B part of C → A regulates C

Reasoning over the relationships

A <u>is a</u> B	◦	B <u>positively regulates</u> C	→	A <u>positively regulates</u> C
A <u>positively regulates</u> B	◦	B <u>is a</u> C	→	A <u>positively regulates</u> C
A <u>positively regulates</u> B	◦	B <u>part of</u> C	→	A <u>positively regulates</u> C
A <u>is a</u> B	◦	B <u>negatively regulates</u> C	→	A <u>negatively regulates</u> C
A <u>negatively regulates</u> B	◦	B <u>is a</u> C	→	A <u>negatively regulates</u> C
A <u>negatively regulates</u> B	◦	B <u>part of</u> C	→	A <u>negatively regulates</u> C

Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides [a controlled vocabulary of terms](#) for describing gene product characteristics and [gene product annotation data](#) from GO Consortium members, as well as [tools to access and process this data](#). [Read more about the Gene Ontology...](#)

Search the Gene Ontology Database

Search for genes, proteins or GO terms using [AmiGO](#):

☒ gene or protein name ☐ GO term or ID

[AmiGO](#) is the official GO browser and search engine. [Browse the Gene Ontology with AmiGO](#).

Quick Links

Tools

[AmiGO browser](#)

[Submit GO Annotations](#)

[OBO-Edit ontology editor](#)

[Ontology downloads](#)

[Annotation downloads](#)

[Database downloads](#)

[Documentation](#)

[GO FAQ](#)

[GO on SourceForge](#)

[Contact GO](#)

News

[GO on Twitter](#)

Finding updates...

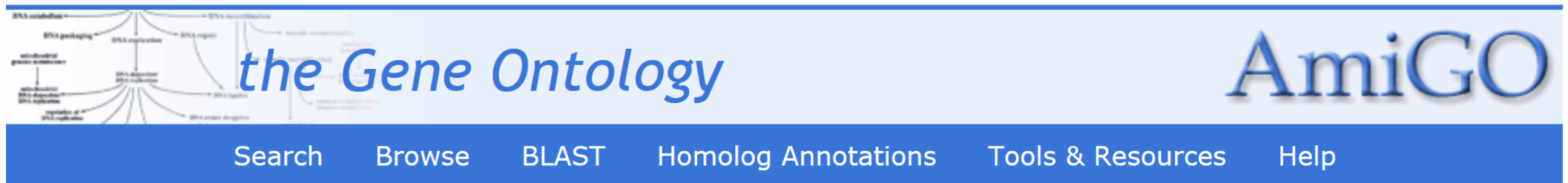
<http://www.geneontology.org/>

Copyright © Peking University

Current statistics of GO

Number of terms	39972
Number of species (filtered)	2838
Number of genes annotated (filtered)	573380
Number of relations	75856
is a	62870
part of	6289
regulates	2490
positively regulates	2089
negatively regulates	2118
Deepest level	12

AmiGO: GO browser and search engine



Search the Gene Ontology database

☒ GO terms ☐ genes or proteins ☐ exact match

Submit Query

Beta
AmiGO 2

AmiGO version: 1.8


Try AmiGO Labs

GO database release 2013-10-26

Cite this data • Terms of use • GO helpdesk

Copyright © 1999-2010 the Gene Ontology

Browse GO

 *the Gene Ontology*

AmiGO

Search **Browse** BLAST Homolog Annotations Tools & Resources Help

Search GO ☒ terms ☐ genes or proteins ☐ exact match

Tree Browser

▼ **Filter tree view** ?

Filter by ontology

Ontology

All
biological process
cellular component
molecular function

Filter Gene Product Counts

Data source

All
ASAP
AspGD
CGD

Species

All
Arabidopsis thaliana
Aspergillus fumig...
Aspergillus fumig...

View Options

Tree view ☒ Full ☐ Compact

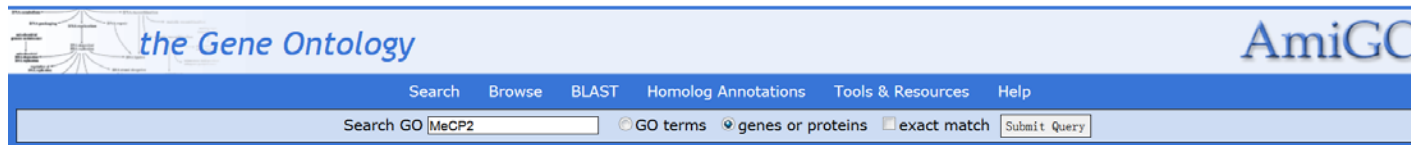
- all : all [699004 gene products]
- GO:0008150 : biological_process [543249 gene products]
- GO:0022610 : biological_adhesion [8711 gene products]
- GO:0051825 : adhesion to other organism involved in symbiotic interaction [970 gene products]
- GO:0044406 : adhesion to host [954 gene products]
- GO:0075001 : adhesion of symbiont infection structure to host [1 gene product]**
 - GO:0075003 : adhesion of symbiont appressorium to host [0 gene products]
 - GO:0075002 : adhesion of symbiont germination tube to host [0 gene products]
 - GO:0075196 : adhesion of symbiont haustorium mother cell to host [0 gene products]
 - GO:0075070 : adhesion of symbiont hyphopodium to host [0 gene products]
 - GO:0075069 : adhesion of symbiont infection cushion to host [0 gene products]
 - GO:0075004 : adhesion of symbiont spore to host [0 gene products]

Actions...

Last action: Opened
GO:0075001

[Graphical View](#)
[Permalink](#)
[Download...](#)
[OBO](#)
[RDF/XML](#)
[GraphViz dot](#)

Search for GO terms or genes



Gene Product

11 results for **MeCP2** in genes or proteins fields **symbol, full name(s) and synonym**

Filter search results

Filter Gene Products

Gene Product Type	Data source	Species
All	All	All
complex	ASAP	Arabidopsis thaliana
gene	AspGD	Aspergillus fumig...
gene product	CCD	Aspergillus fumig...

Filter Gene Products by Ontology

Ontology
All
biological process
cellular component
molecular function

Results are sorted by **relevance**. To change the sort order, click on the column header: **Symbol, full name**

★ indicates that the gene product is a member of a [homolog set](#). Click on the gene product to view details.

Select all Clear all Perform an action with this page's selected gene products... Go!

rel ↓	Symbol, full name	associations
<input type="checkbox"/>	MeCP2 ★ methyl CpG binding protein 2	78 associations BLAST
<input type="checkbox"/>	MeCP2 ★ methyl CpG binding protein 2	78 associations BLAST
<input type="checkbox"/>	mecp2 ★ methyl CpG binding protein 2	3 associations BLAST

mitochondrion inheritance

Term information • Term neighborhood • External references • 137 gene product associations •

Term Information

Accession GO:0000001

Ontology Biological Process

Synonyms exact: mitochondrial inheritance

Definition The distribution of mitochondria, including the mitochondrial genome, into daughter cells after mitosis or meiosis, mediated by interactions between mitochondria and the cytoskeleton.
Source: GOC:mcc, PMID:10873924, PMID:11389764

Comment None

Subset None

Community [Add](#) usage comments for this term on the GORU's wiki.

Term Neighborhood for mitochondrion inheritance (GO:0000001)

Filter lineage gene product counts

Data source	Species
All	All
ASAP	A. fumigatus
AspGD	A. fumigatus AF293
CCD	A. nidulans PGSC A4

Ancestors and Children Inferred Tree View Graph View Other Views Downloads Mappings

Ancestors of mitochondrion inheritance (GO:0000001)

subject	relation	object	annotations
mitochondrion inheritance	is_a (inferred)	biological_process (GO:0008150)	543249
mitochondrion inheritance	is_a (inferred)	localization (GO:0051179)	73695
mitochondrion inheritance	is_a (inferred)	cellular_component_organization_or_biogenesis (GO:0071840)	65628
mitochondrion inheritance	is_a (inferred)	cellular_localization (GO:0051641)	26549
mitochondrion inheritance	is_a (inferred)	cellular_process (GO:0009987)	302930

Summary Questions

What is ontology and can you name a few important Open Biomedical Ontologies?

What is Gene Ontology and what are the three categories of GO?

What is the structure of GO and can you name any GO relations?

Can you browse GO or search GO terms and genes online? Where?

生物信息学：导论与方法

Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍

Center for Bioinformatics, Peking University



<https://www.coursera.org/course/pkubioinfo>