

Chapter 5-7. Ontology (GO, HPO and PTO)

Jingbo Xia

Huazhong Agricultural University

xiajingbo.math@gmail.com

2022-04

What is Ontology?

什么是本体 (Ontologies) ^a ?

- 1 are a set of concepts in a subject area or domain that shows relations between concepts represented by properties.
- 2 go beyond lists, thesauri, and taxonomies to provide a formal description of definitions of conceptual classes and their relations.
- 3 Ontologies are one of the mechanisms to encode the semantics for an area of human knowledge in a machine-readable manner.

^aHarrow, Ian, Rama Balakrishnan, Ernesto Jimenez-Ruiz, Simon Jupp, Jane Lomax, Jane Reed, Martin Romacker et al. "Ontology mapping for semantically enabled applications." Drug discovery today (2019).

<https://www.sciencedirect.com/science/article/pii/S1359644618304215>



[Jingbo] 由于历史的原因, 这个 Slides 一开始就计划用于双语教学, 故没有采用过多的中文翻译。如遇到不熟悉的专有名词, 请勤记、辨识。

Famous ontology I

UMLS ^a

Statistics - 2021AB Release

Concepts: 4,536,653

Number of concept names (AUIs): 16,543,671

Number of distinct concept names (SUIs): 14,005,865

Number of distinct normalized concept names (LUIs): 12,849,628

Number of sources (distinct source families by language): 158

Number of sources contributing concept names: 220

Number of languages contributing concept names: 25

^ahttps://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html.

Famous ontology II

ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO).

It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. Work on ICD-10 began in 1983, became endorsed by the Forty-third World Health Assembly in 1990, and was first used by member states in 1994.

It was replaced by ICD-11 on January 1, 2022. ^a

In the base classification, the code set allows for more than 14,000 different codes. The adapted versions may differ in a number of ways, and some national editions have expanded the code set even further; with some going so far as to add procedure codes. ICD-10-CM, for example, has over 70,000 codes.

^a<https://icd.who.int/en>.

Famous ontology III

Gene ontology. ^a

Valid terms 43699 ($\Delta = -5$)

Obsoleted terms 3574 ($\Delta = 2$)

Merged terms 2369 ($\Delta = 11$)

Biological process terms 28329

Molecular function terms 11187

Cellular component terms 4183

^a<http://geneontology.org/stats.html>.



[Jingbo] 请点击链接了解当前更新。

Famous ontology IV

Property Value	Number of annotations	7,838,790
Annotations for biological process		2,952,731
Annotations for molecular function		2,511,800
Annotations for cellular component		2,374,259
Annotations for evidence PHYLO		3,871,123
Annotations for evidence IEA		1,891,133
Annotations for evidence OTHER		853,417
Annotations for evidence EXP		901,133
Annotations for evidence ND		264,272
Annotations for evidence HTP		57,712
Number of annotated scientific publications		169,555

Famous ontology V

Human phenotype ontology. ^a

The Human Phenotype Ontology (HPO) project provides an ontology of medically relevant phenotypes, disease-phenotype annotations, and the algorithms that operate on these.

The HPO is being used for computational deep phenotyping and precision medicine as well as integration of clinical data into translational research.

The HPO is being increasingly adopted as a standard for phenotypic abnormalities by diverse groups such as international rare disease organizations, registries, clinical labs, biomedical resources, and clinical software tools and will thereby contribute toward nascent efforts at global data exchange for identifying disease etiologies (Köhler et al, 2017).

The HPO currently contains over 13,000 terms arranged in a directed acyclic graph and are connected by is-a (subclass-of) edges.

^a<https://hpo.jax.org/app/help/introduction>.

Table of contents I

1	Gene ontology	9
	• Why gene ontology?	10
	• Basic knowledge of GO	12
	• Case study. GO enrichment by using R codes (代码)	15
	• *Assignment this week (GO enrichment analysis by using R packages)	23
2	Mathematical Foundation of Enrichment Analysis	24
	• Hyper geometric test (数学原理部分)	25
	• Application of enrichment analysis on pathway or GO terms	30
	• Development of novel metrics for gene set enrichment (已发表工作)	32
3	Human Phenotype Ontology	34
	• Basic knowledge of HPO	35
	• HPO-Shuffle on HPO-aided gene prioritization (已发表工作)	38
	• *Assignment this week (HPO enrichment)	46
4	Plant Trait Ontology	47
	• Basic knowledge of HPO	48
	• PTO researches on plant breeding	53
	• Pilot study: plant trait ontology mapping of 2.1k literatures (讨论部分)	62
	• *Assignment this week (PTO mapping)	66

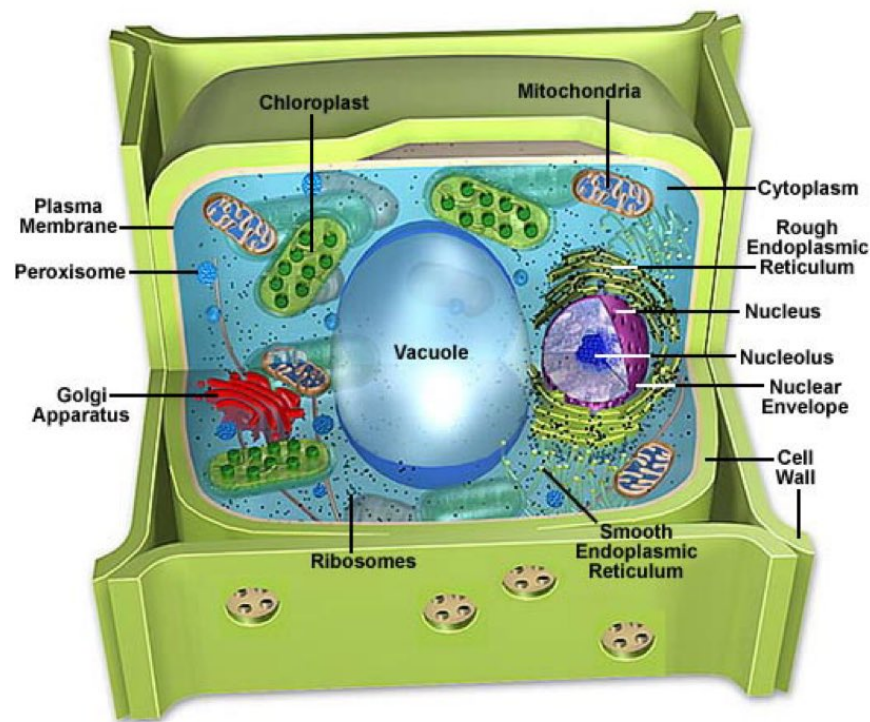
Outline

- 1 Gene ontology 9
 - Why gene ontology? 10
 - Basic knowledge of GO 12
 - Case study. GO enrichment by using R codes (代码) 15
 - *Assignment this week (GO enrichment analysis by using R packages) 23
- 2 Mathematical Foundation of Enrichment Analysis 24
 - Hyper geometric test (数学原理部分) 25
 - Application of enrichment analysis on pathway or GO terms 30
 - Development of novel metrics for gene set enrichment (已发表工作) 32
- 3 Human Phenotype Ontology 34
 - Basic knowledge of HPO 35
 - HPO-Shuffle on HPO-aided gene prioritization (已发表工作) 38
 - *Assignment this week (HPO enrichment) 46
- 4 Plant Trait Ontology 47
 - Basic knowledge of HPO 48
 - PTO researches on plant breeding 53
 - Pilot study: plant trait ontology mapping of 2.1k literatures (讨论部分) 62
 - *Assignment this week (PTO mapping) 66

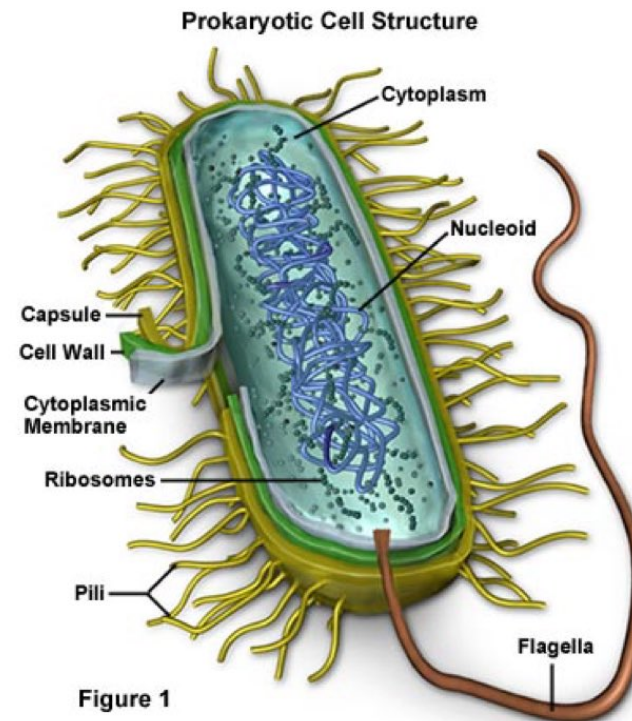
1	Gene ontology	9
	• Why gene ontology?	10
	• Basic knowledge of GO	12
	• Case study. GO enrichment by using R codes (代码)	15
	• *Assignment this week (GO enrichment analysis by using R packages)	23
2	Mathematical Foundation of Enrichment Analysis	24
3	Human Phenotype Ontology	34
4	Plant Trait Ontology	47

Gene Ontology I

Why gene ontology?



Eukaryotic cell;



Prokaryotic cell

图 1: 生物概念具有的复杂性