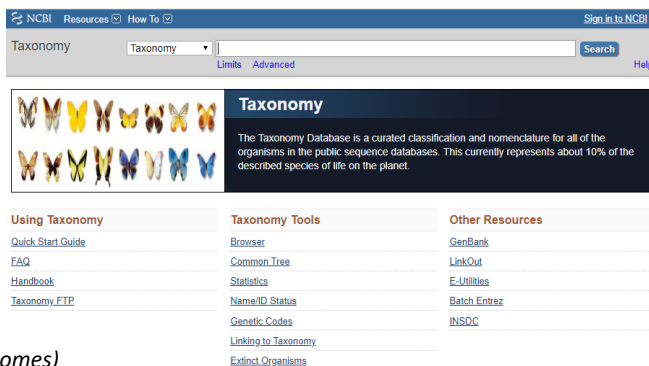


Exercise 3: Determine what types of data are available for my pathogen & find a cool experimental study dataset

- Find my organism in the taxonomic hierarchy
- Move through the hierarchy down to find specific strain information or up to find potentially helpful, related aggregate data
- Learn about my organism and find out what kind of related molecular and other data are available – including experimental studies with data!

NCBI Taxonomy Database <https://www.ncbi.nlm.nih.gov/taxonomy>



Note: Environmental samples (communities/metagenomes) are categorized as a discrete Taxonomic Group.

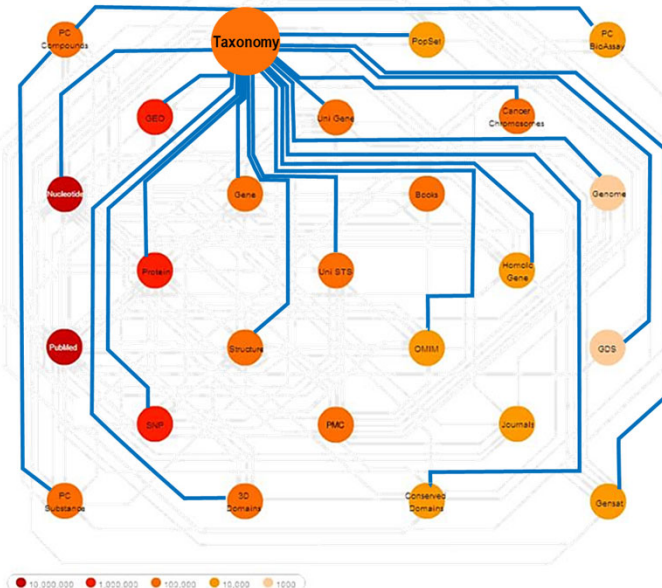
A very old image...

NCBI is known for linking related data within and across databases.

The Taxonomy database was originally developed to catalog GenBank sequences.

It was expanded to organize and connect data within and across all NCBI molecular databases.

And environmental samples (communities of organisms), for example: microbiomes.



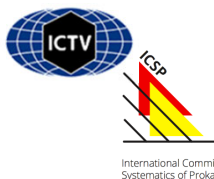
https://www.nlm.nih.gov/ncbi/workshops/2022-08_intro-to-pathogen-data/

Taxonomy – organization for NCBI’s molecular data

- Manually curated based on nucleotide information submitted to GenBank in consultation with Taxonomic Experts
- Continually reviewed and updated.

We work with taxonomic experts for organization and characterization of taxa including:

International Code of Virus Classification and Nomenclature
 International Code of Nomenclature of Prokaryotes
 International Code of Nomenclature for Algae, Fungi & Plants
 International Code of Zoological Nomenclature



NCBI Taxonomy Database

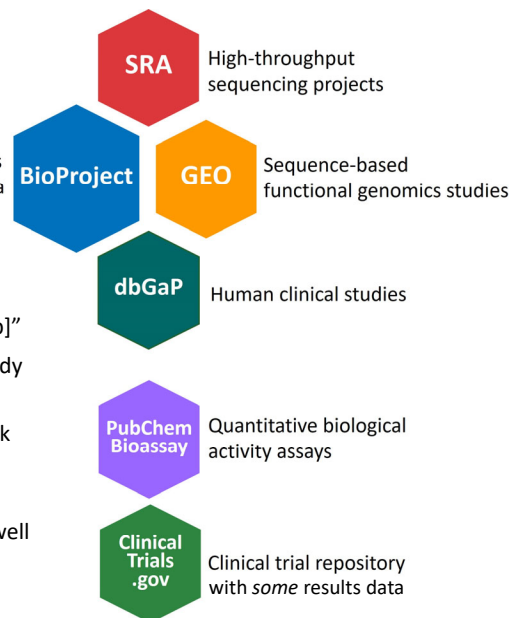
<https://www.ncbi.nlm.nih.gov/taxonomy>

Using Taxonomy	Taxonomy Tools	Other Resources
Quick Start Guide	Browser	GenBank
FAQ	Common Tree	LinkOut
Handbook	Statistics	E-Utilities
Taxonomy FTP	Name/ID Status	Batch Entrez
	Genetic Codes	INSDC
	Linking to Taxonomy	
	Extinct Organisms	

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7408187/>

Experimental Studies with Data at NCBI

A catalog of research projects & related data



Some Hints:

- **Start in Taxonomy** for organism-specific studies
 - For **BioProject** – add the search term “has data[prop]”
 - For **GEO** – set the filter for “Series” to see all the study records
 - For **SRA** – you may want to start in GEO and then link to SRA to get the data, unless you are looking for resequencing data.
 - For **PubChem BioAssay** – use the “Sort” feature as well as filters (on the right) to refine the list of studies