

Organism- & Genome-centered Information Resources

You will be introduced to the main information for molecular (and other) data of prokaryotic and eukaryotic organisms. The tour includes NCBI Taxonomy Browser, Genome Biology Hubs and others.

(1) Genomic Biology (Link from the NCBI homepage – left blue column)

Scope: This page is an information hub for genomic biology resources that are organized in organism-specific categories. Access to NCBI databases and tools that are useful to search and analyse genomic data is also provided from this page.

- Genome Resources
- Organism-Specific Resources

(2) Taxonomy Database/Browser (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>)

Definition: Taxonomy

Taxonomy is the science of classification, identification, and nomenclature. For classification purposes, organisms are usually organized into subspecies, species, genera, families, and higher orders. For eukaryotes, the definition of the species usually stresses the ability of similar organisms to reproduce sexually with the formation of a zygote and to produce fertile offspring. ... Other criteria such as the production and utilization of energy are used for classification of prokaryotes. (For Definitions on *Classification, Identification, Nomenclature* or *Species* please go to Medical Microbiology, Baron S, 1996, 4th Ed. (ISBN 0-9631172-1-1))

- Life on Earth from the “Tree of Life” Web Project (<http://tolweb.org/>, *Root of the tree*)

Scope: “The NCBI Taxonomy Database contains the names and lineages of ~247,000 organisms, both living and extinct, that are represented in the genetic databases with at least one nucleotide or protein sequence”.

Search examples: “dog”, HIV

Practice question: Substantial data are available for two species of filarial nematodes that are human parasites. Use the Taxonomy Browser to examine the number of nucleotide sequences for the superfamily Filarioidea and determine which species these are. How many nucleotide and protein sequences are there for each of these two species? Display nucleotide records for each of these. What kinds of sequences are most of these? (@

http://www.ncbi.nlm.nih.gov/Class/FieldGuide/problem_set.html)

(3) Entrez-Genome Project (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj>)

Scope: “The NCBI Entrez Genome Project database is intended to be a searchable collection of complete and incomplete (in-progress) large-scale sequencing, assembly, annotation, and mapping projects for cellular organisms. The database is organized into organism-specific overviews that function as portals from which all projects in the database pertaining to that organism can be browsed and retrieved.”

Examples: Prokaryotes/Eukaryotes

(4) **Entrez-Genome** (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>)

Scope: “Sequence and map data from the whole genomes of over 1000 organisms are provided. The genomes represent both completely sequenced organisms and those for which sequencing is in progress. All three main domains of life - bacteria, archaea, and eukaryotes - are represented, as well as many viruses, phages, viroids, plasmids, and organelles. *Entrez-Genome* provides graphical overviews of complete genomes/chromosomes, and the ability to explore regions of interest in progressively greater detail”.

Searching *Entrez-Genome* with Standard *Entrez-Genome* functions:

- Limits - limit search using specific criteria
- Preview/Index - view search counts and terms
- History - use your search history
- Clipboard - temporarily store a list of items
- Details - explain search strategy

Search examples: dog (1. without field qualification, 2. with field qualifier “Organism”), compare the retrieval for “human” in *Entrez-Genome* to the search in *Entrez-Nucleotide*.

Comparison of Entrez Databases.

(http://www.ncbi.nlm.nih.gov/genomes/static/gprj_help.html#introduction)

Entrez Database	Organism-specific sequences	Project-specific sequences	Submitter-specific sequences	Complete and in progress	GenBank and RefSeq sequences
Genome	Yes	No	Yes	No	Separated
Taxonomy	Yes	No	No	No	Together
Genome Project	Yes	Yes	Yes	Yes	Separated

(5) **RefSeq** (<http://www.ncbi.nlm.nih.gov/RefSeq/>)

Scope: The Reference Sequence (RefSeq) collection aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms.

- Explanations & Accessions (<http://www.ncbi.nlm.nih.gov/RefSeq/key.html#accessions>)

(6) **MapView** (Link from NCBI homepage and other pages)

Scope: The MapViewer is a software component of *Entrez-Genome* that provides special browsing capabilities for a subset of organisms. It allows you to view and search an organism's complete genome, display chromosome maps, and zoom into progressively greater levels of detail, down to the sequence data for a region of interest.

Example: Browse Chromosomes and Maps, Look at the MapViewer Help pages.

(7) **ENSEMBL Genome Browser** (<http://www.ensembl.org/index.html>)

Scope: This site provides free access to all the data and software from the Ensembl project, a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on selected eukaryotic genomes. Click on a species name to browse the data.

(8) **UCSC Genome Bioinformatics** (<http://genome.ucsc.edu/>)

Scope: This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

(9) **Organism-specific databases & Web Server @ Nucleic Acids Research**

- Molecular Biology Database Collection
(http://nar.oxfordjournals.org/content/vol36/suppl_1/index.dtl)
- Web Server issue (http://nar.oxfordjournals.org/content/vol35/suppl_2/)

MPG Bioinformatics Support Service

(<http://wwwex.biochem.mpg.de/iv/bioinfo.html>)

- Internal Resources, Tools & Groups
(<http://wwwex.biochem.mpg.de/iv/mpsinternal.html>)
- Selected External Resources
(<http://wwwex.biochem.mpg.de/iv/external.html>)