

WEBLEM: 1
National Centre for Biotechnology Information (NCBI)
(URL: <https://www.ncbi.nlm.nih.gov/>)

Introduction:

The National Center for Biotechnology Information (NCBI) was created in 1988 as a division of the National Library of Medicine (NLM) at the National Institutes of Health, to develop information systems in the field of molecular biology and bioinformatics. The data accessible from NCBI's home page (<https://www.ncbi.nlm.nih.gov/>) provides a range of information from short sequences representing parts of genes, such as sequence tags (ESTs), to complete genomic sequences, such as the 21 complete microbial genomes in GenBank today. In addition to maintaining the GenBank, NCBI also provides data analysis and retrieval and resources that operate on GenBank data and also on a variety of other biological data made available through NCBI.

Databases and tools associated with NCBI:

1. Entrez:

- It is an Integrated database retrieval system for access to publicly available GenBank and other sequence data mapping and complete genome data, 3D structures and the biomedical literature through PubMed.
- Entrez provides text searching of sequence or bibliographic records by simple Boolean queries. Plus extensive links to related information.

2. The Taxonomy Browser:

- It is a search tool for the NCBI taxonomy database, which indexes over 55,000 organisms represented in the sequence databases by at least one nucleotide or protein sequence.
- From the data displayed for a particular organism, one can retrieve and download the sequence data for that organism or protein 3D structure data if available.

3. UniGene:

- The UniGene system automatically partitions GenBank sequences, including EST's into a non-redundant set of gene-oriented clusters.
- This currently includes the total genome sequencing for human, mouse and rat in its database.

4. ORF Finder

- This is a tool used for performing a six-frame translation of a nucleotide query and returns a graphic that indicates the location of each ORF found.
- The restrictions on the size of the ORFs returned can be set by the user and the predicted protein sequences can be submitted directly for BLAST similarity searching.

5. Human Genome Resources:

- It provides a centralized access to a full range of human genome resources available from NCBI and elsewhere.
- It also includes the Gene and Disease site, which has general synopses of over 60 diseases of genetic origin.

6. GeneMap '99:

- An international effort was made in 1994 to construct a human gene map by determining the locations of ESTs relative to a framework of well-characterized genetic markers.
- The latest version of this map is called GeneMap '99 and features 30,261 unique gene loci representing approximately half of the 60k-80k genes that are thought to be a part of the human genome.

7. Entrez Genomes:

- Entrez Genomes Database provides access to genomic data contributed by the scientific community for over 600 species whose sequencing and mapping is either complete or in progress.
- In this each and every genome is represented as a clickable graphic for quick navigation to more detailed representations of smaller regions.

8. Retroviral Genotyping Tools:

- Genotyping of retrovirus sequences has a number of implications for attempts at characterization of viral genetic diversity, tracking of the epidemic and in case of HIV-1 the impact that it has on vaccine development.
- The tools helps detect recombinant genomes by performing local blastn comparisons over a sliding window of size and step values set by the user.

9. Cancer Genome Anatomy Project (CGAP):

- The CGAP service provides access to genetic data on normal precancerous and malignant cells generated by the CGAP initiative.
- The cDNA library information from CGAP can be retrieved by text words, or by using an interactive library by select a summary table that organizes data by gene, clone, tissue, type, method of sample preparation, stage of tumor development, etc

10. Molecular Modeling Database (MMDB):

- The MMDB is a structural database derived from the protein Data Bank and accessible via the Entrez system.

Conclusion:

In conclusion, NCBI is a great database with an amazing set of tools required by scientists and students alike to either organize or to collect data/information on a particular topic. The suite of tools that NCBI offers under one umbrella is expansive and ever mutating, receiving new features to handle new complications and edge cases in data management.

References:

1. Nucleic Acids Research, 2000, Vol. 28, No 1, Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L.
2. Science, 1998, Vol 282, Deloukas,P., Schuler,G.D., Gyapay,G., Beasley,E.M., Soderlund,C., Rodriguez-Tome,P., Hui,L., Matise,T.C., McKusick,K.B., Beckmann,S.
3. Nature Genet, 1998, Vol 20, Ermolaeva,O., Rastogi,M., Pruitt,K.D., Schuler,G.D., Bittner,M.L., Chen,Y., Simon,R., Meltzer,P., Trent,J.M. and Boguski,M.S.
4. Human genes and Genetic Disorders, 12th Edition, John Hopkins University Press, Baltimore.
5. Molecular biology, 1990, Vol 125, Ermolaeva,O., Rastogi,M., Pruitt,K.D., Schuler,G.D., Bittner,M.L., Chen,Y., Simon,R., Meltzer,P., Trent,J.M. and Boguski,M.S.