



NCBI FieldGuide Overview

National Center for Biotechnology Information ■ National Library of Medicine ■ National Institutes of Health ■ Department of Health and Human Services

SELECTED DATABASES

Nucleic Acids Research Database Issue:
nar.oxfordjournals.org/cgi/content/full/34/suppl_1/D173

Literature	 PubMed	Biomedical literature citations and abstracts	
	 PubMed Central	Free, full text life sciences articles	
	 OMIM	“Online Mendelian Inheritance in Man” articles on human genes and phenotypes	
	 Books	Online biomedical and scientific texts	
Organism	 Taxonomy	Organisms represented in the NCBI molecular databases	
Sequence	Nucleotide: Three subset databases containing GenBank, EMBL, DDBJ, PDB, RefSeq, TPA and patented nucleotide sequences		
	 CoreNucleotide	A subset of nucleotide sequence records, including the RefSeq collection, that are not included in the EST or GSS databases	
	 EST	High-throughput Expressed Sequence Tag mRNA records	
	 GSS	High-throughput Genome Survey Sequence DNA records	
	 Protein	Translations of GenBank, EMBL, and DDBJ records, as well as PDB, PIR, PRF, SWISS-PROT and patented protein sequences	
	 Genome	Complete genome records integrated with sequence, genetic and physical maps	
	 PopSet	Sets of sequences from population, phylogenetic and mutation studies	
	 Trace Archives	Raw sequence traces generated by large sequencing projects	
	Genome	 Genome Project	Genome project information
		 dbSNP	Single nucleotide polymorphism as well as insertion, deletion and microsatellite data
 UniSTS		Sequence Tagged Site marker and mapping data	
 Gene		Summary pages of Gene-centered information	
 HomoloGene		Computationally-determined homologs among selected eukaryotic organisms with complete genomes	
 Protein Clusters		A collection of related protein sequences (clusters) encoded by complete prokaryotic and chloroplast plasmids and genomes	
Expression	 GEO: Profiles & DataSets	“Gene Expression Omnibus”, a repository for gene expression and hybridization array data	
	 UniGene	Gene-based clusters of EST and mRNA sequences	
	 GenSat	Gene expression atlas of mouse central nervous system	
	 Probe	A public registry of nucleic acid reagents listing effectiveness and distributors	
Structure	 CDD	“Conserved Domain Database” of identified protein domains	
	 CDART	Modular domain architecture of proteins	
	 MMDB/Structure	Experimental 3D structure data from PDB	
	 PubChem: Compound, Substance & BioAssay	Information on chemical compounds and substances as well as biological assay data	
Clinical	 dbGaP	Large-scale patient genotype and phenotype association studies	

SEARCH METHODS

ENTREZ	The Entrez database search protocol allows for keyword searching and integrates databases encompassing scientific literature, taxonomy, sequence, genome, expression and structure data into a tightly interlinked system. (www.ncbi.nlm.nih.gov/sites/gquery)
BLAST	Basic Local Alignment Search Tool (BLAST) is a sequence similarity search program that can be used via a web interface or as a stand-alone tool. BLAST compares combinations of nucleotide or protein queries to those in sequence databases and provides statistical information to help decipher the biological significance of the alignment. (www.ncbi.nlm.nih.gov/BLAST)
VAST	Vector Alignment Search Tool (VAST) is a structure similarity search service. VAST compares the 3D organization of protein secondary structure elements in the MMDB/Structure database and allows for viewing superpositions and alignments using Cn3D. (www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml)
PubChem Structure Search	PubChem Structure Search queries the PubChem Compound Database with a chemical structure based on Tanimoto scoring of substructure binary fingerprints. (pubchem.ncbi.nlm.nih.gov/search/search.cgi)

SELECTED TOOLS

Sequin & BankIt	Stand-alone software and web form tools, respectively, developed by the NCBI for submitting and updating entries to the GenBank, EMBL, or DDBJ sequence databases (www.ncbi.nlm.nih.gov/Sequin and www.ncbi.nlm.nih.gov/BankIt)
Batch Entrez	An interface for uploading a file of UIDs or accession numbers to retrieve records (www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi)
Genome Workbench	A stand-alone application integrating the ability to view and analyze sequence data in many ways, including graphical sequence, varied alignment, phylogenetic tree and tabular views of data (www.ncbi.nlm.nih.gov/projects/gbench).
MapView	A genome browser displaying integrated views of sequence, genetic and physical chromosome maps for eukaryotic organisms (www.ncbi.nlm.nih.gov/mapview)
Model Maker	A tool for viewing evidence for gene annotations on assembled genomic sequence, and create your own model transcripts (www.ncbi.nlm.nih.gov/mapview/static/ModelMakerHelp.html)
Splign	A stand-alone or web application that aligns cDNA with genomic sequences (www.ncbi.nlm.nih.gov/sutils/splign)
ORFinder	A tool that finds potential ORFs in an RNA sequence (www.ncbi.nlm.nih.gov/gorf/gorf.html)
OMSSA	A stand-alone or web application that identifies MS/MS peptide spectra by searching libraries of known protein sequences (pubchem.ncbi.nlm.nih.gov/omssa).
Cn3D	A helper application for web browsers that allow for simultaneous viewing of 3D structures with sequence alignments. Also has powerful annotation and alignment editing features. (www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml)
The BLAST Suite	The NCBI BLAST ftp site provides standalone blast, client server blast, and wwwblast packages for different platforms. It also provides commonly used blast databases in preformatted as well as FASTA format. (ftp.ncbi.nih.gov/blast)

PROGRAMMING TOOLS & INFORMATION

E-Utilities	The Entrez Programming Utilities are a set of server-side programs that provide a stable interface to access NCBI's databases by scripted "URL calls". (eutils.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html)
NCBI Toolkit	The NCBI Software Development Toolkit was developed for the production and distribution of GenBank, Entrez, BLAST and other NCBI services. Source code is provided as well as information on NCBI's data model, data encoding and programming libraries. (www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=toolkit.TOC)
Creating Weblinks	General Entrez Links www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helplinks.chapter.linkshelp OMIM page Links www.ncbi.nlm.nih.gov/Omim/omimhelp.html#HowToLink UniGene page Links www.ncbi.nlm.nih.gov/UniGene/FAQ.shtml Structure page Links www.ncbi.nlm.nih.gov/Structure/MISC/linking.html#mmdbsrv Taxonomy page Links www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=howlink

MyNCBI

MyNCBI is a central place to customize NCBI Web services. You can use My NCBI to: Save searches, Set up e-mail alerts for new content, Display links to Web resources (LinkOut), Choose filters that group search results. (<http://www.ncbi.nlm.nih.gov/entrez/login.fcgi?>)

NCBI Handbook

Individual chapters describing databases, major search algorithms and data processing pipelines; each describes the database design or how the resource works and how the different databases and resources relate to each other, along with a brief overview on using the resource. (www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook)

NCBI Training

Education www.ncbi.nlm.nih.gov/Education
FieldGuide & FieldGuide Plus www.ncbi.nlm.nih.gov/Class/FieldGuide & [/FGPlus](http://www.ncbi.nlm.nih.gov/Class/FieldGuidePlus)
MiniCourses www.ncbi.nlm.nih.gov/Class/minicourses
Structure Course www.ncbi.nlm.nih.gov/Class/Structure/course.html
PubChem Course www.ncbi.nlm.nih.gov/Class/PubChem/course.html
PowerTools Workshop www.ncbi.nlm.nih.gov/Class/PowerTools

HELP!

General Help info@ncbi.nlm.nih.gov
BLAST-specific Help blast-help@ncbi.nlm.nih.gov
GenBank Submissions/Updates Help gb-admin@ncbi.nlm.nih.gov
Telephone (8³⁰ am to 5pm Eastern time) (301) 496-2475