

NCBI Resources: from Sequence to Function

Medha Bhagwat, NCBI

Current Topics in Genome Analysis
January 18, 2005



Outline

About NCBI

NCBI databases and tools

The Entrez- search and retrieval system

Training at NCBI



National Center for Biotechnology Information

<http://www.ncbi.nlm.nih.gov/>

Created as a part of NLM in 1988

- To establish public databases
GenBank and others
- To perform research in computational biology
- To develop software tools for sequence analysis
- To disseminate biomedical information



The screenshot shows the NCBI homepage with the following elements:

- Header:** NCBI logo, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health.
- Navigation:** PubMed, Entrez, BLAST, OMIM, Books, TaxBrowser, Structure.
- Search:** Search PubMed for [] Go
- Left Sidebar:**
 - SITE MAP:** Guide to NCBI resources (circled in pink)
 - About NCBI:** The science behind our resources. An introduction for researchers, educators and the public.
 - GenBank:** Sequence submission support and software
 - Literature databases:** PubMed, OMIM, Books and PubMed Central
 - Molecular databases:** Sequences, structures, and taxonomy
- Main Content:**
 - What does NCBI do?** Established in 1988 as a national molecular biology information public databases, conducts research in computational biology, develops tools for analyzing genome data, disseminates biomedical information, and provides the better understanding of molecular processes affecting human health disease. [More...](#)
 - Sequences, structures, and taxonomy**
 - Genomic biology:** The human genome, whole genomes and related resources
 - Tools:** Data mining
 - Research at NCBI:** People, projects and seminars
 - Software engineering:** Tools, R&D and databases
 - Education:** Teaching resources and on-line tutorials
 - FTP site:** Download data and software
 - Contact information:** How to reach us
- Right Sidebar:**
 - NCBI Web Site Search:** A function in Entrez is now available allowing one to search the NCBI web site and ftp site. Choose 'NCBI site search' from the Entrez pulldown menu to find information from any area of our web site.
 - RefSeq:** A full compilation of all NCBI RefSeqs is now available via the [RefSeq Home page](#) and [ftp](#). Release 1 includes genomic, transcript and protein data with sequences from approximately 2000 taxids, and over 762,776 proteins.
 - NCBI Newsletter:** NCBI's Scientific Outreach and Training
 - Interested in education and training to more efficiently use NCBI resources? Learn about the free training program, "A Field Guide to GenBank and NCBI Resources" in the most recent issue of the [NCBI News](#)
 - Navigation List:**
 - MHC
 - Mouse genome resources
 - NCBI Handbook
 - ORF finder
 - Reference sequence project
 - Retrovirus resources
 - Serial analysis of gene expression
 - SKY/CGH database
 - SNP
 - Trace archive
 - UniGene
 - VecScreen
 - NCI-CGAP

NCBI **Alphabetical Quicklinks Table**

PubMed Entrez BLAST OMIM Taxonomy Structure

ALPHABETICAL QUICKLINKS TABLE
 (To view resource descriptions and a complete list of services, see the [NCBI Resource Guide](#)
 To view resources by category, see the graphical [Site Map](#).)

About NCBI	Education	LocusLink	SAGEmap
Announcements	e-PCR	Malaria	Science Primer
ASN.1	Entrez	Map Viewer	Seminars
BankIt	Entrez Utilities	MeSH	Sequin
BLAST	Expression	MGC	Site Search
BLink	FTP	Microbial Genomes	SKY/M-FISH & FISH Database
Books	GenBank	MMDB	Software Engineering
Cancer Chromosomes	GenBank sample record	Model Maker	Spidey
CDART	Genes NEW	Mutation Databases (external)	Statistics
CDD	Genes and Disease	NCBI Home	Structures
CGAP	Genomes	NCBI News	Submit Data
Clones	GENSAT NEW	Nucleotide Sequences (Entrez)	Taxonomy
Cn3D	GEO (Expression)	OMIM	Tools
Coffee Break	Glossary	ORF Finder	TPA
COGs	Handbook	Plant Genomes	Trace Archive
Computational Biology Branch	HIV Interactions NEW	Protein Sequences (Entrez)	UniGene
Data Submissions	HTGs	PubChem NEW	UniSTS
dbEST	HomoloGene	PubMed	VAST
dbGSS	Human Genome Resources	PubMed Central	VecScreen
dbMHC	Human-Mouse Homology Maps	RefSeq	Viruses
dbSNP	Journals	Research at NCBI	WGS
dbSTS	LinkOut	Retroviruses	What's New NEW

NEW indicates a resource which has become available in the last 12 months.

Web Site Search

NCBI **National Center for Biotechnology Information**
 National Library of Medicine National Institutes of Health

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Entrez for

SITE MAP

- Alphabetical Quicklinks Table
- Resources
- Protein
- Nucleotide
- Structure
- Genome
- An introduction to NCBI
- for researchers
- educators
- the public
- GenBank
- Sequence

What does NCBI do?

Established in 1988 as a national resource for biological information, NCBI creates databases, conducts research in computational biology, develops software for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. [More...](#)

Hot Spots

- Assembly Archive
- Clusters of orthologous groups
- Coffee Break, Genes & Disease, NCBI Handbook

NCBI Databases and Sequence Analysis Tools



Entrez: Search and Retrieval System

<http://www.ncbi.nlm.nih.gov/Entrez/>

NCBI Entrez, The Life Sciences Search Engine

HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

Search across databases Help

Welcome to the new Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
Nucleotide: sequence database (GenBank)	Site Search: NCBI web and FTP sites
Protein: sequence database	UniGene: gene-oriented clusters of transcript sequences
Genome: whole genome sequences	CDD: conserved protein domain database
Structure: three-dimensional macromolecular structures	3D Domains: domains from Entrez Structure
Taxonomy: organisms in GenBank	UniSTS: markers and mapping data
SNP: single nucleotide polymorphism	PopSet: population study data sets
Gene: gene-centered information	GEO Profiles: expression and molecular abundance profiles
HomoloGene: eukaryotic homology groups	GEO DataSets: experimental sets of GEO data
PubChem Compound: small molecule chemical structures	Cancer Chromosomes: cytogenetic databases
PubChem Substance: chemical substances screened for bioactivity	PubChem BioAssay: bioactivity screens of chemical substances
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	GENSAT: gene expression atlas of mouse central nervous system
NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections	MeSH: detailed information about NLM's controlled vocabulary

NCBI



Nucleotide sequences
Protein sequences
Structures
Taxonomy
Genomes
Expression
Chemical
Literature

An Array of Sequence Analysis Tools

<http://www.ncbi.nlm.nih.gov/Tools/>

Nucleotide sequence analysis
Protein sequence analysis
Genome analysis
Structure
Gene expression



Nucleotide Databases

GenBank

Individual submissions

Bulk submissions

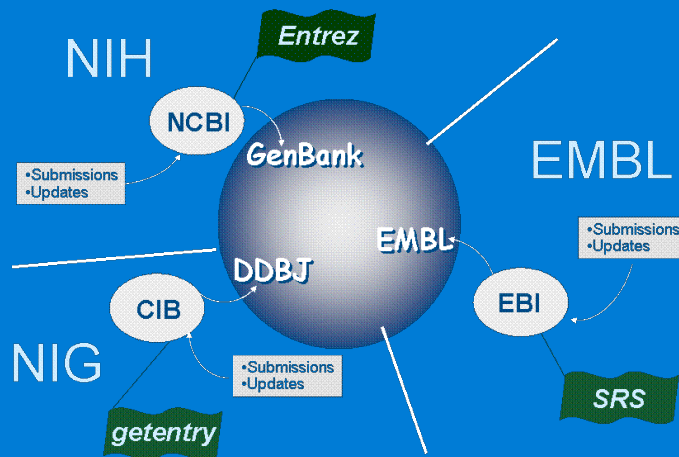
EST, GSS, HTGS, WGS

Derived database

RefSeq

International Nucleotide Sequence Database Collaboration

<http://www.ncbi.nlm.nih.gov/Genbank/>



NCBI Databases

Primary	Derived
Redundant	Non-redundant
Archival/repository	Curated
Submitter owner	NCBI owner
Sequenced	Combined/edited
Ex: GenBank	Ex: RefSeq



<http://www.ncbi.nlm.nih.gov/RefSeq/>

- best, comprehensive, non-redundant set of sequences
- for genomic DNA, transcript (RNA), and protein
- for major research organisms
2645 organisms
- based on GenBank derived sequences
- ongoing curation by NCBI staff and collaborators, with review status indicated on each record
- updates to reflect current knowledge of sequence data and biology





Partial Accession Number List

NM_123456	mRNA	
NP_123456	Protein	
NR_123456	RNA	Non-coding transcripts
NG_123456	Genomic	Incomplete genomic region
NT_123456	Genomic	BAC sequence assemblies
NW_123456	Genomic	WGS sequence assemblies
NC_123456	Genomic	Complete genomic molecules
XM_123456	mRNA	Genome Annotation
XR_123456	RNA	Genome Annotation
XP_123456	Protein	Genome Annotation



A RefSeq Record

LOCUS	NM_139344	2508 bp	mRNA	linear	PRI 27-OCT-2004
DEFINITION	Homo sapiens bridging integrator 1 (BIN1), transcript variant 2, mRNA.				
ACCESSION	NM_139344				
VERSION	1..2376				
KEYWORDS	source	Location/Qualifiers			
SOURCE		/organism="Homo sapiens"			
ORGANISM		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/chromosome="2"			
		/map="2q14"			
REFERENCE	gene	1..2376			
AUTHOR		/gene="BIN1"			
TITLE		/note="synonyms: AMPH2, AMPHL, SH3P9, MGC10367, DKFZp547F068"			
JOURNAL		/db_xref="GeneID:274"			
PUBMED		/db_xref="LocusID:274"			
REMARK	misc_feature	/db_xref="MIM:601248"			
		1			
REFERENCE	misc_feature	/gene="BIN1"			
AUTHOR		/note="5'-most transcription initiation site is undetermined"			
		189			
TITLE	misc_feature	/gene="BIN1"			
		/note="alternate transcription initiation site"			
		346..1866			
JOURNAL	CDS	/gene="BIN1"			
PUBMED		/note="isoform 3 is encoded by transcript variant 3; amphiphysin-like; amphiphysin II; box dependant MYC interacting protein 1;			
REMARK		go_component: nucleus [goid 0005634] [evidence IEA];			
		go_component: cytoplasm [goid 0005737] [evidence IEA];			
		go_component: actin cytoskeleton [goid 0015629] [evidence TAS] [pmid 9182667];			
		go_function: protein binding [goid 0005515] [evidence IEA];			

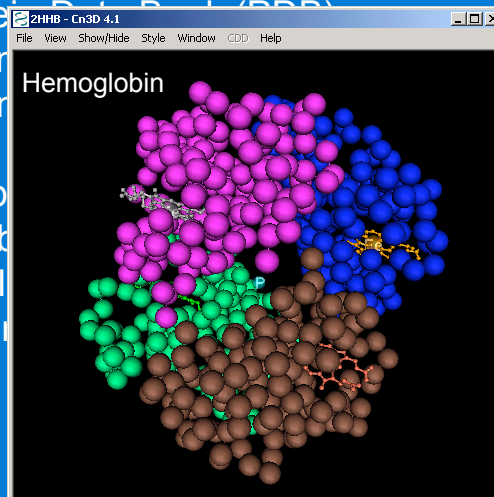
Protein

- Conceptual translations of GenBank and RefSeq records
- SwissProt, PIR, PRF, PDB

Molecular Modeling DataBase (MMDB)

<http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>

- obtained from the Protein Data Bank (PDB)
- experimentally determined
- can be viewed using Crystallographic Object Manipulation (COM) software
- sequences also available in Entrez protein database
- useful for finding homologous structures for a protein





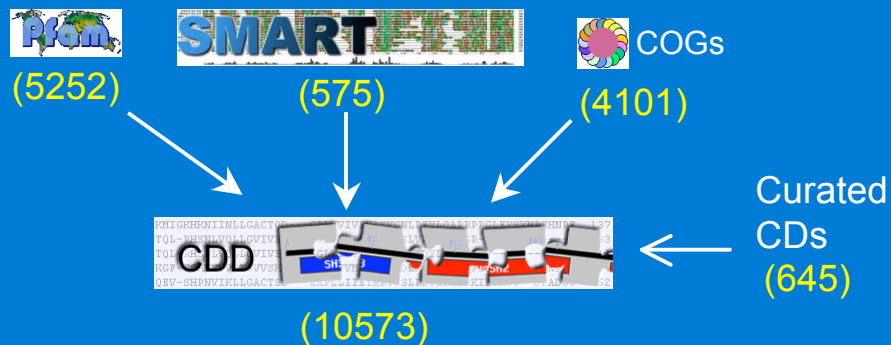
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>

Conserved Domain

- recurring unit in molecular evolution, whose extents can be determined by sequence and structure analysis
- performs a particular function
- represented as a multiple local sequence alignment of proteins containing the domain



Conserved Domain Database



- A position-specific scoring matrix (PSSM) is calculated
- CD-Search can be used to search against the PSSMs
- Manual curation of CDs has begun





Organisms

http://www.ncbi.nlm.nih.gov/Taxonomy

incorporates phylogenetic and taxonomic knowledge from a variety of sources

The screenshot shows the NCBI Taxonomy Browser interface. The search bar contains 'chicken' and the search type is set to 'phonetic name'. The results show 'chicken' with two entries: 'Gallus gallus [genbank common name: chicken]' and 'Siagona'. A 'Taxonomy Tip of the Day' is also visible.

Taxonomy Browser

Gallus gallus

Taxonomy ID: 9031
 Genbank common name: **chicken**
 Rank: **species**
 Genetic code: [Translation table 1 \(Standard\)](#)
 Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)
 Other names:

common name: **chickens**
 includes: **dwarf Leghorn chickens**
 includes: **red junglefowl**
 misnomer: **Gallus domesticus**
 misnomer: **Gallus gallus domesticus**

[Lineage \(full\)](#)
 cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Sauropsida; Sauria; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	871,867	871,825
Protein	29,194	29,192
Structure	453	453
Genome	31	31
Popset	31	31
3D Domains	1,967	1,967
Domains	1	1
UniGene	21,447	21,447
UniSTS	1,958	1,958
PubMed Central	243	242
Gene	18,505	18,505
HomoloGene	9,700	9,700
Taxonomy	3	1

Genome Information

[See the NCBI Genome homepage](#)

Trace records (raw single-pass reads of DNA sequence)				
Center name	Record counts per type			
	FINISHING	SHOTGUN	WGS	ALL
JGI - Joint Genome Institute, U.S. Department of Energy	0	0	2,477,710	2,477,710
UOKNOR - University of Oklahoma Norman Campus, Advanced Center for Genome Technology	223	13,662	0	13,885
Total	223	13,662	2,477,710	2,491,598



NCBI Home > Genomic Biology > Chicken Genome Resources

Search chicken or (Gallus gallus)


Clear

Chicken Genome Resources

Jump to the Genome!
Chromosome:

NCBI Web Resources:
Global Query. Query all NCBI Entrez databases in one step.
BLAST. Compare your sequence to different organism-specific sequences.
Clone Registry. Find information about specific BAC clones, including sequencing status and end sequence information.
dbSNP. Database of SNPs and other genetic variation.
Entrez Gene. Focal point for genes and associated information.
e-PCR. Check your sequence for STSs and view in genomic context.
HomoloGene. Putative homologies among human, mouse, rat, and zebrafish.
Map Viewer. Interactive viewer for genome maps, sequence, and genes.
PopSet. Population study data sets.
PubMed Central. Digital archive of full text and content from life

Welcome to the *Gallus gallus* Genome Resources page. This homepage provides information on chicken- and avian-related resources from NCBI and the chicken research community. [We encourage your suggestions.](#)



The chicken (*Gallus gallus*) is an important model organism for biomedical research, development, and aging, in addition to being important agriculturally. This handsome rooster is a Red Jungle Fowl, considered the ancestor of all breeds of domestic chickens.

Additional Resources
New This Month In:
 • PubMed
 • PubMed Central
 • GenBank

ANNOUNCING the release of the chicken genome assembly (build 1.1) in Map Viewer. Take a moment to BLAST your favorite gene sequence against the genome and explore the maps available for viewing. [Learn more about the Gnomon gene prediction program and the resulting models available in Map Viewer.](#)

Documentation:
 • [Chicken Map Viewer Help](#)
 • [Chicken Sequencing White Paper](#)
 Maps and Sequence:
 • [Chicken Map Viewer](#)
 • [Gallus gallus Traces](#)
 Annotation Projects:
 • [Ensembl Annotation](#)
 • [UCSC Annotation](#)

NCBI

Genomes

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>

NCBI

Entrez PubMed

Search

Limits

About Entrez

Entrez Genome Help

Submitting

Genome Project

Genome sequence

Microbial

Genome Projects

PDB neighbors

Genomic BLAST

Microbial

Eukaryotic

FUNGI Genome projects

WGS projects

projects

WGS projects

Archaea

Chromosome

Plasmid

DarttAssembly

Bacteria

Chromosome

Plasmid

DarttAssembly

Eukaryota

Chromosome

Plasmid

Organelles

Viruses

Phages

Viroids

All Plasmids

Environmental

Sample

Organism:
[Bacillus licheniformis ATCC 14580](#)
Genome sequence information
 chromosome - [CP000002](#) - [NC_006270](#)
 Size: 4,222,336 bp Proteins: 4161
 Sequence data files submitted to GenBank/EMBL/DDBJ can be found at NCBI FTP:
[GenBank](#); or [RefSeq Genomes](#)

Bacillus cereus ZK
 Release Date: September 15, 2004
 Reference: Brettn, T.S., et al.
 Complete genome sequence of Bacillus cereus ZK
 Unpublished
 Lineage: Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
 Organism:
[Bacillus cereus ZK](#)
Genome sequence information
 chromosome - [CP000001](#) - [NC_006274](#)
 Size: : 5,300,915 bp Proteins: 5134
 Sequence data files submitted to GenBank/EMBL/DDBJ can be found at NCBI FTP:
[GenBank](#) or [RefSeq Genomes](#)

Influenza Virus Resource
 sequence database and analyses
[WGS Projects](#)
 Whole Genome
 Shotgun sequencing
 ▶ **Tools and Analysis**
[Map Viewer](#)
 genome browser for eukaryotic genomes
[TaxPlot](#)
 3-Way View of Genome Similarities
[COGs](#)
 clusters of orthologous groups
[BLAST](#)
 with completed and unfinished genomes
 ▶ [Major Sequencing Centers](#)

Genomes

Entrez Genome Help

Submitting
Genome Project
Genome sequence

Microbial
Genome Projects
PDB neighbors

Genomic BLAST
Microbial
Eukaryotic

FUNGI Genome projects

WGS projects

Archaea
Chromosome
Plasmid
DartAssembly

Bacteria
Chromosome
Plasmid
DartAssembly

Bacteria Complete Chromosome	Taxonomy	/List	206
Acinetobacter sp. ADP1	NC 005966	3598621 bp	Jul 9 2004
Agrobacterium tumefaciens str. C58	circular NC 003062	2841581 bp	Oct 3 2001
Agrobacterium tumefaciens str. C58	linear NC 003063	2074782 bp	Oct 3 2001
Agrobacterium tumefaciens str. C58	circular NC 003304	2841490 bp	Dec 14 2001
Agrobacterium tumefaciens str. C58	linear NC 003305	2075560 bp	Dec 14 2001
Anaplasma marginale str. St. Maries	NC 004842	1197687 bp	Dec 8 2004
Aquifex aeolicus VF3	NC 000913	1551335 bp	Sep 7 2001
Azoarcus sp. EbN1	NC 006513	4296230 bp	Dec 9 2004
Bacillus anthracis str. Ames Ancestor	NC 007530	5227419 bp	May 20 2004
Bacillus anthracis str. A2012	NC 003925	5093554 bp	Jun 13 2002
Bacillus anthracis str. Ames	NC 003997	5227293 bp	Apr 30 2003
Bacillus anthracis str. Sterne	NC 005945	5228663 bp	Jun 24 2004
Bacillus cereus ATCC 10987	NC 003909	5224283 bp	Feb 24 2004
Bacillus cereus ATCC 14579	NC 004722	5411809 bp	Apr 17 2003
Bacillus cereus ZK	NC 006274	5300915 bp	Sep 16 2004
Bacillus clausii KSM-K16	NC 006582	4803871 bp	Jan 3 2005
Bacillus halodurans C-125	NC 002570	4202353 bp	Sep 10 2001
Bacillus licheniformis ATCC 14580	NC 006270	4222334 bp	Sep 15 2004
Bacillus licheniformis ATCC 14580	NC 006322	4222645 bp	Sep 28 2004
Bacillus subtilis subsp. subtilis str. 168	NC 000964	4214630 bp	Nov 20 1997
Bacillus thuringiensis serovar konkukian str. 97-27	NC 005927	5237682 bp	Jun 30 2004
Bacteroides fragilis YCH46	NC 006347	5277274 bp	Oct 1 2004
Bacteroides thetaiotaomicron VPI 5402	NC 004662	6760361 bp	Mar 18 2004



Taxonomy Browser

Entrez
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
Books

Search for as complete name lock

Display levels using filter: none

Acinetobacter sp. ADP1

Taxonomy ID: 62977

Rank: species

Genetic code: [Translation table 11 \(Bacterial and Plant Plastid\)](#)

Other names:

synonym: **Acinetobacter calcoaceticus ADP1**

Lineage (full)

[cellular organisms](#); [Bacteria](#); [Proteobacteria](#); [Gammaproteobacteria](#); [Pseudomonadales](#); [Moraxellaceae](#); [Acinetobacter](#)

Entrez records	
Database name	Direct links
Nucleotide	20
Protein	6,891
Structure	9
Genome	1
2D Domains	33
PubMed Central	4
Gene	3,425
Taxonomy	1



NCBI Genomes

BLAST PubMed Nucleotide Protein Genome Structure PopSet Taxonomy/Help

Acinetobacter sp. ADP1, complete genome [Microbial genomes](#)

Sequencing center: [Genoscope](#)

Genome Info	Feature table	BLAST protein homologs	Links
Refseq: NC_005966	Protein coding genes	COGs (Clusters of Orthologous Groups)	Refseq FTP
GenBank: CR543861	Structural RNAs	3D Structure (Sequences with known structure)	GenBank FTP
Total Bases: 3598621 bp		TaxMap (Sequences grouped by superkingdom)	BLAST
Completed: Jul 9, 2004		TaxPlot (3-way genome comparison)	TraceAssembly
		GenePlot (Pairwise genome comparison)	CDD

Accession [NC_005966](#)

Save the report below in [Table](#) format.

FASTA format - Protein in FASTA format

Location	FASTA proteins	FASTA nucleotide	ID	Gene	COG	Synonym	Product
201..1598	+	465	S0083298	dnaA		ACIAD0001	DNA replication initiator protein
1834..2982	+	382	S0083299	dnaN		ACIAD0002	DNA polymerase III, beta chain
2998..4074	+	358	S0083300	recF		ACIAD0003	DNA replication, recombination
4127..6595	+	822	S0083301	gyrB		ACIAD0004	DNA gyrase, subunit B (type II)
6712..6948	-	78	S0083302			ACIAD0005	hypothetical protein
6969..7139	+	56	S0083303			ACIAD0006	hypothetical protein
7336..8270	+	684	S0083304			ACIAD0007	putative transmembrane protein

Entrez Gene

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>

- A record represents a single gene from an organism
- A gene-specific information such as map, sequence, expression, structure, function, homology and publications
- Includes data for all organisms that have RefSeq genome records
- Successor to LocusLink
 - more organisms
 - efficient searching options

NCBI



NCBI Entrez Gene


Search Gene for [] Go Clear current records only

Display Summary Send to Text

Gene Table

Gene name: HBB
 Gene description: hemoglobin, beta
 RefSeq status: Reviewed
 Organism: *Homo sapiens*
 Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo
 Gene aliases: hemoglobin
 Summary: The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of polypeptide chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta-3'.
 Bibliography: Gene References into Function (GeneRIF): [Submit](#) [help](#)
 PubMed links
 GeneRIFs:
 1. co-transcriptional cleavage process in the human beta-globin gene involves an RNA self-cleaving activity.

NCBI



NCBI Entrez Gene

Search Gene for [] Go Clear current records only

Display Gene Table Show: 5 Send to Text

Gene Table

Gene name: HBB hemoglobin, beta [*Homo sapiens*]
 GeneID: 3043 Locus tag: [HGNC:4827](#); [MIM: 141900](#)
 total gene size: 1606 bp
 Transcripts and products: (shown on reverse complement genome)
 NC_000011
 mRNA bp exons Protein aa exons
 NM_000518 626 3 NP_000509 148 3
 Exon information:
 NM_000518 length: 626 bp, number of exons: 3
 NP_000509 length: 148 aa, number of exons: 3

EXON	coords	length	Coding EXON	coords	length	INTRON	coords	length
1	142	142 bp	51 - 142	92 bp	143 - 272	130 bp		
2	273 - 495	223 bp	273 - 495	223 bp	496 - 1345	850 bp		
3	1346 - 1606	261 bp	1346 - 1474	129 bp				

Display Gene Table Show: 5 Send to Text

NCBI

► General gene information

GeneOntology

Provided by [GOA](#)

Function

[oxygen transporter activity](#) IEA [PubMed](#)

[oxygen transporter activity](#) NAS [PubMed](#)

Process

[oxygen transport](#) IEA

[oxygen transport](#) NAS

[transport](#) IEA

Component

[hemoglobin complex](#) NAS

Phenotypes

Erythremias, beta- [MIM: 141900](#)

Heinz body anemias, beta- [MIM: 141900](#)

HPPFH, deletion type [MIM: 141900](#)

Methemoglobinemias, beta- [MIM: 141900](#)

Sickle cell anemia [MIM: 141900](#)

Thalassemia-beta, dominant inclusion-body [MIM: 603902](#)

Thalassemias, beta- [MIM: 141900](#)

Markers (Sequence Tagged Sites/STS)

[STS-L48931](#) (e-PCR)

Alternate name RH39984

Alternate name sts-L48931

[RH41842](#) (e-PCR)

Alternate name STS-F17257



► NCBI Reference Sequences (RefSeq)

Reference [NG_000007](#)

mRNA Sequence [NM_000518](#)

Source Sequence [L48217](#)

Product [NP_000509](#) beta globin

Conserved Domains (1) [summary](#)

[cd01040: globin](#); [Globins are heme proteins, which bind and transport oxygen](#)

Location: 5 - 142 Blast Score: 278

► Related Sequences

Nucleotide	Protein
Genomic A01592	CAA00182
Genomic AF007546	AAB62944
Genomic AF104901	AAC97372
Genomic AF105973	AAC97959
Genomic AF186606	AAF08258
Genomic AF186607	AAF08259
Genomic AF186608	AAF08260
Genomic AF186609	AAF08261
Genomic AF186610	AAF08262
Genomic AF186611	AAF08263
Genomic AF186612	AAF08264
Genomic AF186613	AAF08265
Genomic AF186614	AAF08266
Genomic AF186615	AAF08267



Entrez Gene

Display: Gene Table Show: 5 Send to: Text

1: HBB hemoglobin, beta [*Homo sapiens*]
 GeneID: 3043 Locus tag: [HGNC:4827](#); [MIM: 141900](#)

Transcripts and products: (shown on reverse complement genome) [RefSeq below](#)

NC_000011

Genomic context: chromosome: 11; Maps: 11p15.5

Gene type: protein coding
Gene name: HBB
Gene description: hemoglobin, beta
RefSeq status: Reviewed
Organism: [Homo sapiens](#)
Lineage: *Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Catarrhini; Hominidae; Homo*
Gene aliases: hemoglobin
Summary: The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of globin chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia. Reduced amounts of detectable beta globin causes beta-plus-thalassemia. The order of the genes in the beta-globin cluster is 5'-epsilon -- gamma-G -- gamma-A -- delta -- beta-3'.

Links

- Conserved Domains
- GEO Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- Protein
- PubMed
- SNP
- GeneView in dbSNP
- Taxonomy
- UniSTS
- AceView
- Ensembl
- Evidence Viewer
- GDB
- GeneTests for MIM: 141900
- Globin Gene Server
- HGNC
- LocusID
- MGC
- ModelMaker
- PharmGKB
- UCSC
- UniGene
- LinkOut

NCBI

UniGene

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>

- An evolving system
- Automatically partitioning expressed sequences
- Non-redundant set of gene-oriented clusters

NCBI

UniGene Cluster for Human HBB

SEQUENCES
Sequences representing this gene; mRNAs, ESTs, and gene predictions supported by transcribed sequences.

mRNA sequences (20)

[AF117710.1](#) Homo sapiens hemoglobin beta chain (HBB) mRNA, complete cds

[NM_000518.4](#) Homo sapiens hemoglobin, beta (HBB), mRNA

[AY509193.1](#) Homo sapiens hemoglobin beta mRNA, complete cds

[CR536530.1](#) Homo sapiens full open reading frame cDNA clone

EST Sequences (10 of 20)

[B1518741.1](#) cDNA clone IMAGE:5211

[BQ890006.1](#) cDNA clone IMAGE:6295

[BQ898811.1](#) cDNA clone IMAGE:6302

Download sequences

Unix
 PC
 Mac

[Convert to text mode](#)

Breakdown by Tissue
Hs. 523443

Bladder	0	0/21715
Blood	6603	517/78292
Bone	53	3/55730
Bone Marrow	3065	112/36541
Brain	123	57/462100
Cervix	0	0/41264
Colon	55	10/179987
Eye	130	22/168244
Heart	220	13/58912
Kidney	95	13/135458
Larynx	0	0/27551
Liver	167	22/131463
Lung	328	95/288794
Lymph Node	23	3/128142
Mammary Gland	124	16/128200
Muscle	2428	265/109115
Ovary	83	8/95612
Pancreas	555	47/84639
Peripheral ...	80	2/24996
Placenta	946	225/237797
Prostate	149	20/133636
Skin	66	11/165608
Small Intes...	0	0/14090
Soft Tissue	715	17/23760
Spleen	3473	67/19290
Stomach	18	2/108238
Tongue	0	0/23932
Testis	29	4/136540
Thymus	146	1/6848
Uterus	115	21/181622
Vascular	77	2/25883

Breakdown by Developmental Stage
Hs. 523443

Embryo	1054	536/508346
Juvenile	33	2/59542
Adult	513	500/974089

GENE EXPRESSION
Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

cDNA sources: Blood, Bone, Bone Marrow, Brain, Colon, Eye, Heart, Kidney, Liver, Lung, Lymph Node, Mammary Gland, Muscle, Ovary, Pancreas, Peripheral Nervous System, Placenta, Prostate, Skin, Soft Tissue, Spleen, Stomach, Testis, Thymus, Uterus, Vascular, Embryo, Juvenile, Adult

Restricted Expression: Embryo [[Show more like this](#)]

[Expression Profile:](#) View expression levels using UniGene's EST ProfileViewer

Note: Highly represented in many libraries

Gene Expression Omnibus

<http://www.ncbi.nlm.nih.gov/geo/>

- First fully public high-throughput gene expression data repository
- Curated, online resource for gene expression data browsing, query and retrieval

GDS596: Large-scale analysis of the human transcriptome (HG-U133A)



<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene>

- An automated system
- Detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes



Genes

Genes identified as putative homologs of one another during the construction of HomoloGene

- [H.sapiens HBB](#)
hemoglobin, beta.
- [P.troglodytes LOC450978](#)
similar to beta globin; hemoglobin beta chain; beta globin mutant, beta globin chain.
- [M.musculus LOC436003](#)
similar to hemoglobin beta chains - white rhinoceros.
- [R.norvegicus Hbb](#)
hemoglobin beta chain complex.

Proteins

Proteins used in sequence comparisons and the conserved domain architectures.

- [NP_000509.1](#)
147 aa
- [XP_508242.1](#)
147 aa
- [XP_488069.1](#)
147 aa
- [NP_150237.1](#)
147 aa
- [XP_215033.1](#)
147 aa

Species	Gene	aa%ID	nt%ID	D	Ka/Ks	Knr/Knc	
H.sapiens	HBB						
vs. M.musculus	LOC436003	57.1	71.7	0.356	0.555	0.820	Blast
vs. R.norvegicus	Hbb	81.6	82.8	0.196	0.263	0.480	Blast
vs. R.norvegicus	LOC293265	78.9	81.0	0.220	0.271	0.694	Blast
vs. P.troglodytes	LOC450978	100.0	99.8	0.002	0	0	Blast

D evolutionary distance
 Ka/Ks non-synonymous/synonymous changes
 Knr/Knc conserved/non-conserved changes



- A catalog of human genes and genetic disorders at John Hopkins
- Developed for the World Wide Web by NCBI

+141900

HEMOGLOBIN--BETA LOCUS; HBB

ALLELIC VARIANTS (selected examples) **mature protein**

- 0001 HEMOGLOBIN AALBORG [HBB, GLY74ARG]
- 0002 HEMOGLOBIN ABRUZZO [HBB, HIS143ARG]
- 0003 HEMOGLOBIN AGENOGI [HBB, GLU90LYS]
- 0004 HEMOGLOBIN ALABAMA [HBB, GLN39LYS]
- 0005 HEMOGLOBIN ALAMO [HBB, ASN19ASP]

- 0242 HEMOGLOBIN RUSH [HBB, GLU101GLN]
- 0243 HEMOGLOBIN S [HBB, GLU6VAL]
- 0244 HEMOGLOBIN S (ANTILLES) [HBB, GLU6VAL AND VAL23ILE]
- 0245 HEMOGLOBIN S (OMAN) [HBB, GLU6VAL AND GLU121LYS]
- 0246 HEMOGLOBIN S (PROVIDENCE) [HBB, GLU6VAL AND LYS82]
- 0247 HEMOGLOBIN S (TRAVIS) [HBB, GLU6VAL AND ALA142VAL]
- 0248 HEMOGLOBIN SABINE [HBB, LEU91PRO]
- 0249 HEMOGLOBIN SAINT JACQUES [HBB, ALA140THR]
- 0250 HEMOGLOBIN SAFFAMA [HBB, HIS117PHE]

Glu7Val in the precursor Hemoglobin S sickle cell anemia

Entrez Gene

Med Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

for [] Go Clear current records only

Limits Preview/Index History Clipboard Details

Display Graphics Show: 5 Send to: Text

1: HBB hemoglobin, beta [*Homo sapiens*]
GeneID: 3043 Locus tag: HGNC:4827; MIM: 141900

Transcripts and products: (shown on reverse complement genome) RefSeq below

NC_000011

5284877 5283272
5' 3'
NM_000516 NP_000509
- coding region - untranslated region

Genomic context: chromosome: 11; Maps: 11p15.5

[5155526 5221399]
OR5221P OR51V1 HBB HBBP1

Gene type: protein coding
Gene name: HBB
Gene description: hemoglobin, beta
RefSeq status: Reviewed
Organism: *Homo sapiens*
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Primates; Catarrhini; Hominidae; Homo
Gene aliases: hemoglobin

Summary: The alpha (HBA) and beta (HBB) loci determine the structure of the 2 types of globin chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha chains and two beta chains. Mutant beta globin causes sickle cell anemia. Absence of beta chain causes beta-zero-thalassemia.

Links
Conserved Domains
GEO Profiles
HomoloGene
Map Viewer
Nucleotide
OMIM
Protein
PubMed
SNP
GeneView in dbSNP
Taxonomy
UniSTS
AceView
Ensembl
Evidence Viewer
GDB
GeneTests for MIM: 141900
Globin Gene Server
HGNC
LocusID
MGC
ModelMaker
PharmGKB
UCSC
UniGene
LinkOut

NCBI

SNPs in the HBB Gene

Contig	mrna	protein	mrna orientation	transcript	snp list
NT_009237	NM_000518	NP_000509	reverse	minus strand	currently shown

view rs In gene region cSNP has frequency double hit haplotype tagged

Contig position	dbSNP rs#	dbSNP cluster id	Hetero-zygosity	Validation	3D	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid
4035096	rs11549405	N.D.		Yes			synonymous	C	Leu [L]	3	89
4035112	rs1803195	N.D.		Yes			nonsynonymous	T	Val [M]	2	84
4035119	rs11549406	N.D.		Yes			nonsynonymous	G	Val [M]	1	82
4035245	rs11549407	N.D.		Yes			nonsynonymous	A	Lys [K]	1	40
4035473	rs334	N.D.		Yes			nonsynonymous	T	Val [M]	2	7
4035484	rs713040	N.D.		Yes			synonymous	C	His [H]	3	3
		0.849		Yes			contig reference	T	His [H]	3	3

Snp In Gene Model Legend:

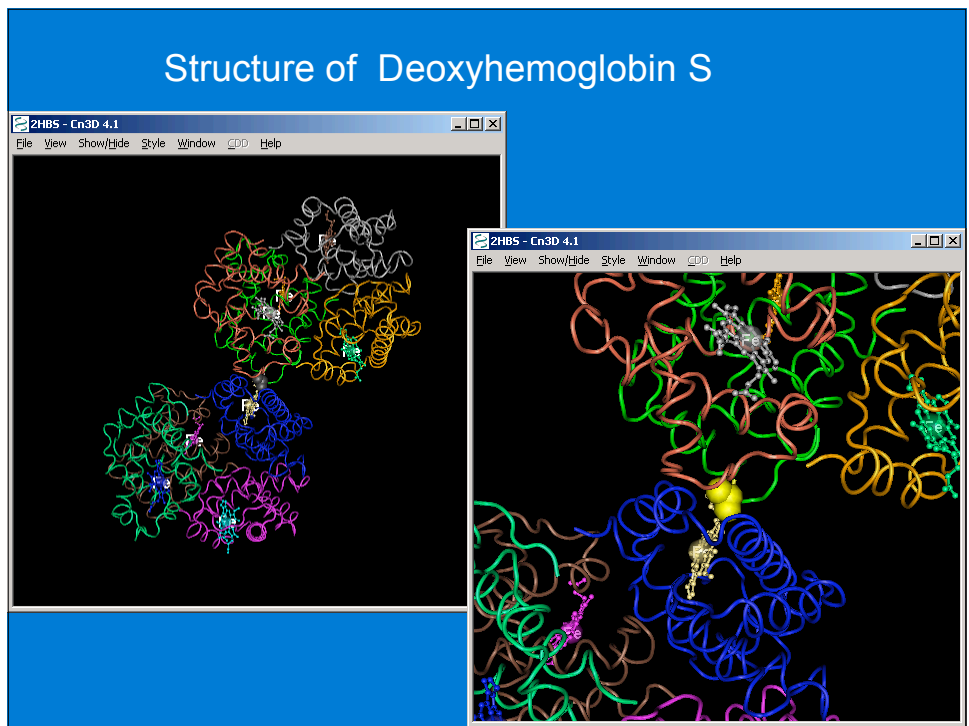
- Region: exon
- Region: intron
- snp: coding
- snp: synonymous
- snp: nonsynonymous
- snp: untranslated
- snp: intron
- snp: splice-site
- snp: coding: synonymy unknown

Hemoglobin S

↑

Glu6Val in the mature protein

Structure of Deoxyhemoglobin S





Literature Databases



Pubmed

Biomedical literature

PubMed Central

Free online journals

<http://www.pubmedcentral.gov>

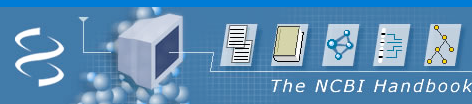


Books

Free online textbooks

Online Books

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=books>



Other Databases in Entrez

Cancer Chromosomes chromosomal aberrations

NCI/NCBI SKY/M-FISH & CGH Database
NCI Mitelman Database of Chromosome Aberrations in Cancer
NCI Recurrent Aberrations in Cancer

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=CancerChromosomes>

PubChem catalog of small organic molecules

To support the Molecular Libraries and Imaging component of the NIH
Roadmap Initiative

- chemical structures
- information on their biological activities

<http://pubchem.ncbi.nlm.nih.gov/>



NCBI Databases and Sequence Analysis Tools



An Array of Sequence Analysis Tools

<http://www.ncbi.nlm.nih.gov/Tools/index.html>

- Nucleotide sequence analysis
- Protein sequence analysis
- Genome analysis
- Structure
- Gene expression



NCBI **NCBI Map Viewer**

Genome Taxonomy Entrez BLAST Help

Search for

New! ANNOUNCING - the release of the **Chimpanzee** (*Pan troglodytes*) genome assembly (build 1.1). The chimpanzee is our closest living evolutionary relative, and the chimpanzee assembly will provide insight into genome evolution and organization, particularly late primate evolution. Please take a moment to view the new [Map Viewer resources](#) available for this mammalian species.

Click the to BLAST, the to search the group

- Mammals 9 organisms
- Other Vertebrates 2 organisms
- Invertebrates
 - Insects 3 organisms
 - Nematode 1 organism
- Fungi 11 organisms
- Protozoa 1 organism
- Plants 8 organisms

See more about Bacteria, Organelles, Viruses



HBB Gene in the Human Map Viewer

NCBI NCBI Map Viewer

PubMed Entrez BLAST OMM Taxonomy Structure

Search Find Find in This View Advanced Search

Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View

Maps & Options
Compress Map
Region Shown:
5,203K
5,205K Go
out
zoom
in
You are here:
Ideogram
default
master

Homo sapiens build 35.1 BLAST The Human Genome

Chromosome: 1 2 3 4 5 6 7 8 9 10 [11] 12 13 14 15 16 17 18 19 20 21 22 X Y MT

Query: 3043[gene_id] [clear]

Master Map: Genes On Sequence Summary of Maps Maps & Options

Region Displayed: 5,203K-5,205K bp Download/View Sequence/Evidence

HBB + OMIM sv pr dl ev nm hm C 11p15.5 hemoglobin, beta

NCBI

NCBI BLAST

PubMed Entrez BLAST OMM Taxonomy Structure

About BLAST **NEW 15 Nov 2004** Download the [BLAST poster](#) from SC2004!

Nucleotide

- Quickly search for highly similar sequences (megablast)
- Quickly search for divergent sequences (discontiguous megablast)
- Nucleotide-nucleotide BLAST (blastn)
- Search for short, nearly exact matches
- Search trace archives with megablast or discontiguous megablast

Protein

- Protein-protein BLAST (blastp)
- PHI- and PSI-BLAST
- Search for short, nearly exact matches
- Search the conserved domain database (rpsblast)
- Search by domain architecture (cdart)

Translated

- Translated query vs. protein database (blastx)
- Protein query vs. translated database (tblastx)
- Translated query vs. translated database (tblastb)

Genomes

- Chicken, cow, pig, dog, sheep, cat
- Environmental samples
- Human, mouse, rat
- Fugu rubripes, zebrafish
- Insects, nematodes, plants, fungi, malaria
- Microbial genomes, other eukaryotic genomes

Special

- Search for gene expression data (GEO BLAST)
- Align two sequences (bl2seq)
- Screen for vector contamination (VecScreen)
- Immunoglobulin BLAST (IgbLst)
- Human SNP BLAST **NEW**

Meta

- Retrieve results by RID
- Get this page with javascript-free links

News
Mailing list
References
NCBI Contributors

BLAST Services

- FAQs
- Program selection guide
- Web service interface

BLAST Software

- Databases
- Documentation
- Errata
- Executables
- Source code

Support

- Contact us

NCBI

Spidey



http://www.ncbi.nlm.nih.gov/spidey

Genomic sequence (FASTA or GI/Accession):

Upload file:

AC002390

From: To:

mRNA sequence(s) (One or more FASTA or GI/Accession):

Upload file:

NM_014164
AF177940

- divergent sequences
 Use large intron sizes

Minimum mRNA-genomic identity %
Minimum length of mRNA covered %

Genomic sequence is:

- Vertebrate
 Drosophila
 C. elegans
 Plant

Output options:

- Text/summary
 Summary only
 ASN 1
 Print multiple alignment



Spidey



Genomic: [gi2282011|gb|AC002390.1|AC002390](#) Human DNA from overlapping chromosome 19-specific cosmids R30072 and R28588, genomic sequence

mRNA: [gi21618360|ref|NM_014164.3](#) Homo sapiens FXYP domain containing ion transport regulator 5 (FXYP5), mRNA

Alignment is on plus strand of genomic sequence and on plus strand of mRNA sequence
mRNA coverage: 100.0%
Overall percent identity: 100.0%

516 15730

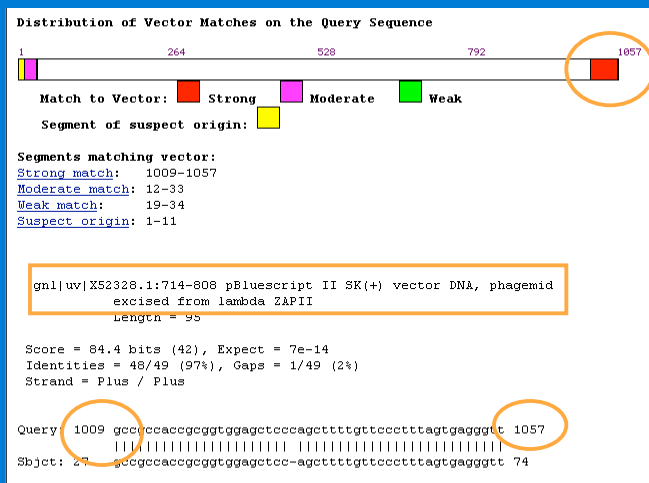
	Genomic coordinates	mRNA coordinates	length	identity	mismatches	gaps	Donor site	Acc. site
Exon 1	516-657	1-142	142	100.0%	0	0	d	
Exon 2	1399-1459	143-203	61	100.0%	0	0	d	a
Exon 3	3269-3349	204-284	81	100.0%	0	0	d	a
Exon 4	4192-4248	285-341	57	100.0%	0	0	d	a
Exon 5	6557-6649	342-434	93	100.0%	0	0	d	a
Exon 6	10004-10093	435-524	90	100.0%	0	0	d	a





VecScreen

<http://www.ncbi.nlm.nih.gov/VecScreen/>



Outline

About NCBI

NCBI databases and tools

The Entrez- search and retrieval system

Training at NCBI





http://www.ncbi.nlm.nih.gov/Entrez/

HOME SEARCH SITE MAP PubMed Entrez Human Genome GenBank Map Viewer BLAST

Search across databases GO CLEAR Help

Welcome to the new Entrez cross-database search page

PubMed: biomedical literature citations and abstracts	Books: online books
PubMed Central: free, full text journal articles	OMIM: online Mendelian Inheritance in Man
Nucleotide: sequence database (GenBank)	Site Search: NCBI web and FTP sites
Protein: sequence database	UniGene: gene-oriented clusters of transcript sequences
Genome: whole genome sequences	CDD: conserved protein domain database
Structure: three-dimensional macromolecular structures	3D Domains: domains from Entrez Structure
Taxonomy: organisms in GenBank	UniSTS: markers and mapping data
SNP: single nucleotide polymorphism	PopSet: population study data sets
Gene: gene-centered information	GEO Profiles: expression and molecular abundance profiles
HomoloGene: eukaryotic homology groups	GEO DataSets: experimental sets of GEO data
PubChem Compound: small molecule chemical structures	Cancer Chromosomes: cytogenetic databases
PubChem Substance: chemical substances screened for bioactivity	PubChem BioAssay: bioactivity screens of chemical substances
Journals: detailed information about the journals indexed in PubMed and other Entrez databases	GENSAT: gene expression atlas of mouse central nervous system
NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections	MeSH: detailed information about NLM's controlled vocabulary



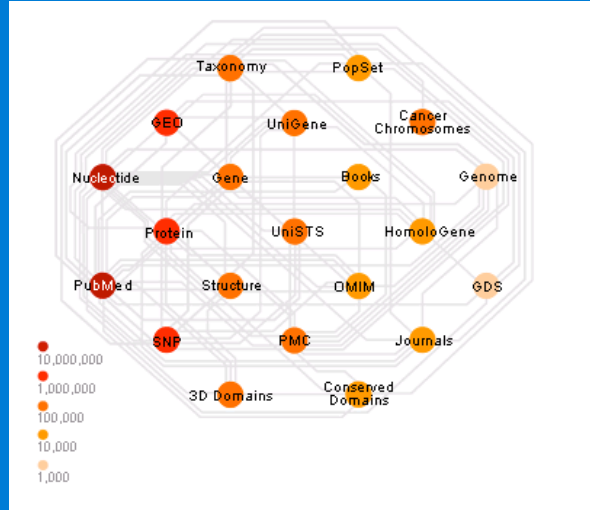
Entrez: Search and Retrieval System

Search across databases GO CLEAR Help

15310987 PubMed: biomedical literature citations and abstracts	105303 Books: online books
334988 PubMed Central: free, full text journal articles	16438 OMIM: online Mendelian Inheritance in Man
47258389 Nucleotide: sequence database (GenBank)	6626 Site Search: NCBI web and FTP sites
5667042 Protein: sequence database	1417836 UniGene: gene-oriented clusters of transcript sequences
4086 Genome: whole genome sequences	10897 CDD: conserved protein domain database
27968 Structure: three-dimensional macromolecular structures	118763 3D Domains: domains from Entrez Structure
244359 Taxonomy: organisms in GenBank	462562 UniSTS: markers and mapping data
15004100 SNP: single nucleotide polymorphism	26582 PopSet: population study data sets
1124043 Gene: gene-centered information	9996683 GEO Profiles: expression and molecular abundance profiles
32302 HomoloGene: eukaryotic homology groups	762 GEO DataSets: experimental sets of GEO data
897246 PubChem Compound: small molecule chemical structures	48836 Cancer Chromosomes: cytogenetic databases
825845 PubChem Substance: chemical substances screened for bioactivity	173 PubChem BioAssay: bioactivity screens of chemical substances
20257 Journals: detailed information about the journals indexed in PubMed and other Entrez databases	21928 GENSAT: gene expression atlas of mouse central nervous system
1208746 NLM Catalog: catalog of books, journals, and audiovisuals in the NLM collections	171448 MeSH: detailed information about NLM's controlled vocabulary



Linking within Databases in Entrez



Searching in Entrez-Nucleotide



Entrez PubMed Nucleotide Protein Genome Structure PMC

Search Nucleotide for [] Go Clear

Limits Preview/Index History Clipboard Details

- Use All Fields pull-down menu to specify a field.
- Boolean operators AND, OR, NOT must be in upper case.
- If search fields tags are used enclose in square brackets, e.g., rubella [ti].
- More help on using limits is available [here](#).

Limited to:

All Fields

exclude ESTs exclude STSs exclude GSS

exclude TPA exclude working draft exclude patents exclude all of the above

Molecule Gene Location Segmented Sequences

Only from Modification Date

Modification Date From To

Use the format YYYY/MM/DD; month and day are optional.

Limited to:

- Accession
- All Fields
- Author
- EC/RN Number
- Feature key
- Filter
- Gene Name
- Issue
- Journal
- Keyword
- Modification Date
- Organism
- Page Number
- Primary Accession
- Properties
- Protein Name
- Publication Date
- SeqID String
- Sequence Length
- Substance Name
- Text Word
- Title



Searching for Virus Sequences excluding HIV 1

NCBI Entrez Nucleotide search interface. Search term: virus. The search results list includes:

- virus like particle cak1(1)
- virus of serpulina hydodysenteriae 1(1)
- virus phich1(8)
- viruses(293378)**
- viscaceae(474)
- viscacha rat(5)
- viscainoa(3)
- viscainoa geniculata(3)
- viscaria(125)
- viscaria alpina(1)

Searching for Virus Sequences excluding HIV 1

NCBI Entrez Nucleotide search interface. Search term: viruses[Organism]. The search results list includes:

- virus like particle cak1(1)
- virus of serpulina hydodysenteriae 1(1)
- virus phich1(8)
- viruses(293378)**
- viscaceae(474)
- viscacha rat(5)
- viscainoa(3)
- viscainoa geniculata(3)
- viscaria(125)
- viscaria alpina(1)

NCBI Entrez Nucleotide search interface. Search term: viruses[Organism] NOT HIV 1[Organism]. The search results list includes:

- virus like particle cak1(1)
- virus of serpulina hydodysenteriae 1(1)
- virus phich1(8)
- viruses(293378)**
- viscaceae(474)
- viscacha rat(5)
- viscainoa(3)
- viscainoa geniculata(3)
- viscaria(125)
- viscaria alpina(1)

Most Recent Queries table:

Search	Time	Result
#2 Search "viruses"[Organism]	18:06:27	293378
#3 Search "viruses"[Organism] NOT HIV 1[Organism]	18:08:18	170500
#2 Search "viruses"[Organism]	18:06:27	293378

NCBI Nucleotide

Search: Nucleotide for "viruses[Organism] NOT HIV 1[Organism]"

Display: Summary Show: 20 Send to: Text

Items 1 - 20 of 170500 Page 1 of 8525

- 1: [AH004344](#) Reports Links
VP1/2A (5' region, capsid/protease junction) [poliovirus type 3 P3, southern Alberta isolate, Genomic RNA, 230 nt 2 segments]
gi|57165425|gb|AH004344.2|bbm|322237[57165425]
- 2: [CQ972063](#) Reports Links
Sequence 8 from Patent WO2004108922
gi|57163376|emb|CQ972063.1|pat|WO|2004108922|8[57163376]
- 3: [CQ972062](#) Reports Links
Sequence 7 from Patent WO2004108922
gi|57163375|emb|CQ972062.1|pat|WO|2004108922|7[57163375]
- 4: [CQ972016](#) Reports Links
Sequence 3 from Patent WO2004108159
gi|57163356|emb|CQ972016.1|pat|WO|2004108159|3[57163356]
- 5: [CQ972014](#) Reports Links
Sequence 1 from Patent WO2004108159
gi|57163354|emb|CQ972014.1|pat|WO|2004108159|1[57163354]
- 6: [CQ971747](#) Reports Links
Sequence 13 from Patent WO2004108754
gi|57163188|emb|CQ971747.1|pat|WO|2004108754|13[57163188]
- 7: [CQ971743](#) Reports Links

Searching in Entrez Nucleotide Properties Field

NCBI Nucleotide

Search: Nucleotide for "viruses[Organism] NOT HIV 1[Organism]"

- Enter terms and click Preview to see only the number of search results.
- To combine searches use # before search number, e.g., (#2 OR #3) AND asthma.
- Click on query # to add to strategy

Search	Most Recent Queries	Time	Result
#3	Search viruses[Organism] NOT HIV 1[Organism]	18:25:31	170500
#2	Search viruses[Organism]	18:25:25	293378

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.

Properties: Preview Index

Click **AND** **OR** **NOT** to add terms selected from Index to the query box.

- gbdiv bcl(311215)
- gbdiv con(403356)
- gbdiv est(24844913)
- gbdiv gss(10873022)
- gbdiv htc(365952)
- gbdiv htg(69755)
- gbdiv imv(772966)
- gbdiv mem(2122397)
- gbdiv pat(2343409)
- gbdiv rho(28511)

gbdiv
biomol
srcdb

CGCTCAGGAT... GACTTCGG... GCTAG... ATCGGATCCCGG... ATTATATAGCTGGATGATC1
 TTCTCTAT... TGG... FATATACACACA... GCGATAGCTGATCTGATCTA
 CCCC... CTTCGCATACGTC...
 CACAG... GGG... ATTATATAGCTGGATGATC1

Nucleotide

PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for viruses[Organism] NOT HIV 1[Organism] NOT "gbdiv pat"[Properties] Preview Go Clear

Limits **Preview/Index** History Clipboard Details

- Enter terms and click Preview to see only the number of search results.
- To combine searches use # before search number, e.g., (#2 OR #3) AND asthima.
- Click on query # to add to strategy

Search	Most Recent Queries	Time	Result
#4	Search viruses[Organism] NOT HIV 1[Organism] NOT "gbdiv pat"[Properties]	18:27:27	157416
#3	Search viruses[Organism] NOT HIV 1[Organism]	18:25:31	170500
#2	Search viruses[Organism]	18:25:25	293378

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.

Properties [dropdown] Preview Index

Click AND OR NOT to add a term to the query box.

viruses[Organism] NOT HIV 1[Organism] NOT "gbdiv pat"[Properties]
 "biomol genomic"[Properties] AND "srcdb refseq"[Properties]

NCBI

Displaying and Saving Sequences in Entrez Nucleotide

CGCTCAGGAT... GACTTCGG... GCTAG... ATCGGATCCCGG... ATTATATAGCTGGATGATC1
 TTCTCTAT... TGG... FATATACACACA... GCGATAGCTGATCTGATCTA
 CCCC... CTTCGCATACGTC...
 CACAG... GGG... ATTATATAGCTGGATGATC1

Nucleotide

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for viruses[Organism] NOT HIV 1[Organism] NOT "gbdiv pat"[Properties] Go Clear

Limits Summary **Preview/Index** History Clipboard Details

about Entrez Entrez Nucleotide Help | FAQ Entrez Tools Check sequence revision history LinkOut Pubby Related resources BLAST Reference sequence project Search for Genes Submit to GenBank Search for full length DNAs

Display	Summary	Show	Send to
<input type="checkbox"/>	ASN1	5	Text
<input type="checkbox"/>	FASTA	10	File
<input type="checkbox"/>	XML	20	Clipboard
<input type="checkbox"/>	GenBank	50	
<input type="checkbox"/>	GI list	100	
<input type="checkbox"/>	Graphics	200	
<input type="checkbox"/>	TinySeq XML	61588	
<input type="checkbox"/>	GBSeq XML		
<input type="checkbox"/>	INSDSeq XML		
<input type="checkbox"/>	LinkOut		
<input type="checkbox"/>	Assembly	1326584	
<input type="checkbox"/>	Related Sequences		
<input type="checkbox"/>	Component Links		
<input type="checkbox"/>	mRNA Links		
<input type="checkbox"/>	Components to Genome	genome	
<input type="checkbox"/>	Gene Links	6900714	
<input type="checkbox"/>	Assembly to Genome		
<input type="checkbox"/>	GENSAT Links		
<input type="checkbox"/>	GEO Profile Links		
<input type="checkbox"/>	Strawberry palmopsis associated virus RNA 1, complete sequence		
<input type="checkbox"/>	NC_005895.1[48696526]		
<input type="checkbox"/>	NC_004718 Reports		
<input type="checkbox"/>	SARS coronavirus, complete genome		
<input type="checkbox"/>	NC_004718.3[30271926]		
<input type="checkbox"/>	NC_006577 Reports		

Page 1 of 101 Next

Searching in Entrez Nucleotide

NCBI Nucleotide

Search Nucleotide for viruses[Organism] NOT HIV 1[Organism] NOT % [Go] [Clear]

Display Summary Show: 20 Send to Text

Items 1 - 20 of 2010 Page 1 of 101 Next

- 1: [AC_000001](#) Reports
 - Bovine adenovirus 2, complete genome sequence [56158826]ref[AC_000001]
 - Reports: ASN.1, XML, Summary, FASTA
 - Links: Gene, Protein, PubMed, Taxonomy
- 2: [NC_003977](#) Reports
 - Hepatitis B virus, complete genome sequence [21326584]ref[NC_003977]
 - Reports: TinySeq XML, GenBank, GBSeq XML, INSDSeq XML
 - Links: GenBank(Full), GI list, Graphic, Revision History
- 3: [NC_006579](#) Reports
 - Pneumonia virus of mice, complete genome sequence [56900714]ref[NC_006579]
 - Reports: GenBank(Full), GI list, Graphic, Revision History
 - Links: GenBank(Full), GI list, Graphic, Revision History
- 4: [NC_005895](#) Reports
 - Strawberry pallidosis associated virus RNA 1, complete sequence [48696526]ref[NC_005895.1][48696526]
 - Reports: GenBank(Full), GI list, Graphic, Revision History
 - Links: GenBank(Full), GI list, Graphic, Revision History
- 5: [NC_004718](#) Reports
 - SARS coronavirus, complete genome sequence [30271926]ref[NC_004718.3][30271926]
 - Reports: GenBank(Full), GI list, Graphic, Revision History
 - Links: GenBank(Full), GI list, Graphic, Revision History
- 6: [NC_006577](#) Reports

Sequence Revision History

Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for [GI numbers or Fasta style Seqids] NC_004718 [Go] [Clear]

Show difference between I and II as GenBank/GenPept

Revision history for NC_004718

GI	Version	Update Date	Status	I	II
30271926	3	Jan 4 2005 1:33 AM	Live	☑	☑
30271926	3	Sep 30 2004 1:34 AM	Dead	☑	☑
30271926	3	Aug 31 2004 2:06 AM	Dead	☑	☑
30271926	3	Aug 3 2004 11:29 AM	Dead	☑	☑
30271926	3	Apr 9 2004 3:47 PM	Dead	☑	☑
30271926	3	Apr 1 2004 1:33 AM	Dead	☑	☑
30271926	3	May 2 2003 12:19 PM	Dead	☑	☑
30271926	3	May 1 2003 4:13 PM	Dead	☑	☑
30124072	2	Apr 25 2003 5:30 PM	Dead	☑	☑
30124072	2	Apr 25 2003 5:15 PM	Dead	☑	☑
29826277	1	Apr 25 2003 1:15 AM	Dead	☑	☑
29826277	1	Apr 21 2003 2:58 PM	Dead	☑	☑
29826277	1	Apr 18 2003 1:09 AM	Dead	☑	☑
29826277	1	Apr 15 2003 4:05 PM	Dead	☑	☑
29826277	1	Apr 15 2003 9:15 AM	Dead	☑	☑
29826277	1	Apr 15 2003 8:11 AM	Dead	☑	☑
29826277	1	Apr 15 2003 1:33 AM	Dead	☑	☑
29826277	1	Apr 15 2003 1:33 AM	Dead	☑	☑
29826277	1	Apr 14 2003 10:01 PM	Dead	☑	☑
29826277	1	Apr 14 2003 12:19 PM	Dead	☑	☑

Accession NC_004718.3 was first seen at NCBI on Apr 14 2003 12:19 PM

NCBI

Accessing the Sequence and Annotation Information

The screenshot displays the NCBI Entrez interface for the sequence NC_004718.3. The top navigation bar includes tabs for Nucleotide, Protein, Genome, Structure, and PDB. The search bar shows 'Nucleotide' selected. Below the search bar, there are options for 'Display' (GenBank), 'Send' (all to file), and 'Range: from begin to end'. The sequence information is as follows:

```

LOCUS       NC_004718             29751 bp ss-RNA           linear   VRL 04-JAN-2005
DEFINITION  SARS coronavirus, complete genome.
ACCESSION  NC_004718
VERSION    NC_004718.3  GI:30271926
KEYWORDS   .
SOURCE     SARS coronavirus
ORGANISM   SARS coronavirus
REFERENCE  1 (bases 1 to 29751)
AUTHORS   He,R., Dobie,F., Ballantine,M, Cutts,T., Andonov,A., Cao,J., Baker,L. and Li,X.
CONSTRM   ECCA Genome Sciences Centre, Control and National Microbiology Institute
TITLE     Analysis of multimerization of the SARS coronavirus nucleocapsid protein
JOURNAL   Biochem. Biophys. Res. Commun. 2004 Apr 2;316(2):476-83.
PUBMED   15020242
    
```

The detailed view shows the title 'Analysis of multimerization of the SARS coronavirus nucleocapsid protein.' and the authors 'He R, Dobie F, Ballantine M, Leeson A, Li Y, Bastien N, Cutts T, Andonov A, Cao J, Booth TF, Plummer FA, Tyler S, Baker L, Li X.' The abstract text is partially visible, discussing the SARS-CoV genome and the nucleocapsid protein (N).

Examples of Searching in Entrez

Nucleotide:

Mouse EST sequences
 mouse[Organism] AND "gbdiv est"[Properties]
 DNA barcode sequences
 "barcode"[Properties]

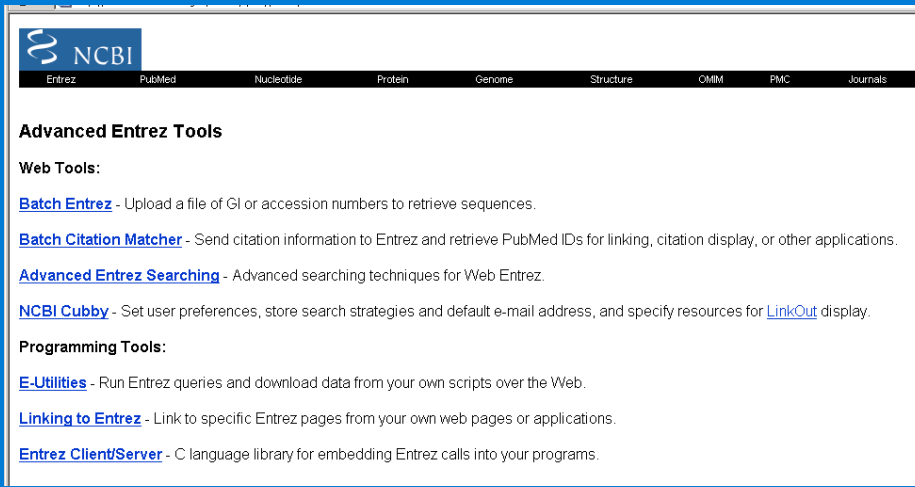
Protein:

Peptide sequences of length between 40 and 50
 40:50[Sequence Length]
 Proteins with links to PubChem Compound
 "protein pcompound"[Filter]

Homologene:

Entries for human disease genes
 "link phenotype omim"[Properties]





The screenshot shows the NCBI website's "Advanced Entrez Tools" page. At the top, there is a navigation bar with the NCBI logo and links for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, OMIM, PMC, and Journals. Below the navigation bar, the page is titled "Advanced Entrez Tools" and is divided into two sections: "Web Tools" and "Programming Tools".

Web Tools:

- [Batch Entrez](#) - Upload a file of GI or accession numbers to retrieve sequences.
- [Batch Citation Matcher](#) - Send citation information to Entrez and retrieve PubMed IDs for linking, citation display, or other applications.
- [Advanced Entrez Searching](#) - Advanced searching techniques for Web Entrez.
- [NCBI Cubby](#) - Set user preferences, store search strategies and default e-mail address, and specify resources for [LinkOut](#) display.

Programming Tools:

- [E-Utilities](#) - Run Entrez queries and download data from your own scripts over the Web.
- [Linking to Entrez](#) - Link to specific Entrez pages from your own web pages or applications.
- [Entrez Client/Server](#) - C language library for embedding Entrez calls into your programs.

Below the tools list, the URL <http://www.ncbi.nlm.nih.gov/entrez/query/static/advancedentrez.html> is displayed. The NCBI logo is visible in the bottom right corner of the screenshot.

Outline

About NCBI

NCBI databases and tools

The Entrez- search and retrieval system

Training at NCBI

[Information and tutorials](#)
[BLAST statistics](#)
[Tutorials](#)
[Structure](#)
[PubMed](#)
[Tour | Tutorial](#)
[Map Viewer](#)
[Exercises **NEW**](#)
NCBI Courses
[A Field Guide to GenBank and NCBI Resources](#)
[Medical Library Association Course on Molecular Biology Information](#)
[Genomes & Genetics](#)
[Genes and disease](#)
[Genetic Analysis Software](#)
[Human genome project](#)
[Glossary of genetic terms](#)

Information and tutorials at NCBI

BLAST Information 	Nucleotide tutorial 	Pubmed tutorial
Resource publications 	Map Viewer exercises 	Structure tutorial

Browse our science primer...
 ...to gain an understanding of our resources and explore our databases and tools to see what we can do for you.

A science primer

<http://www.ncbi.nlm.nih.gov/Education/>

NCBI Training

<http://www.ncbi.nlm.nih.gov/Education>

A Field Guide to GenBank and NCBI Molecular Biology Resources

3 hour lecture and 2 hour hands-on

on specific topics 2 hour lecture and hands-on

Three day workshops at NCBI

NCBI Core Bioinformatics Facility

- Supports a network of bioinformatics specialists serving individual institutes at NIH
- Trains Core Members in the use of NCBI tools
- The Core Members, in turn, support the use of NCBI's tools and databases by researchers in their institutes
- Currently 18 Members from 14 institutes

Refer to the handout for the Core Member from your institute



Access More Information at

1.



<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books>

2. Database Resources of the National Center for Biotechnology Information
Nucleic Acids Res. 2005 Jan 1;33 Database Issue:D39-45
3. GenBank
Nucleic Acids Res. 2005 Jan 1;33 Database Issue:D34-8



Outline

About NCBI

NCBI databases and tools

The Entrez- search and retrieval system

Training at NCBI

