NCBI Resources: from Sequence to Function

Medha Bhagwat, NCBI

Current Topics in Genome Analysis
January 18, 2005

NCBI

Outline

About NCBI
NCBI databases and tools
The Entrez- search and retrieval system
Training at NCBI

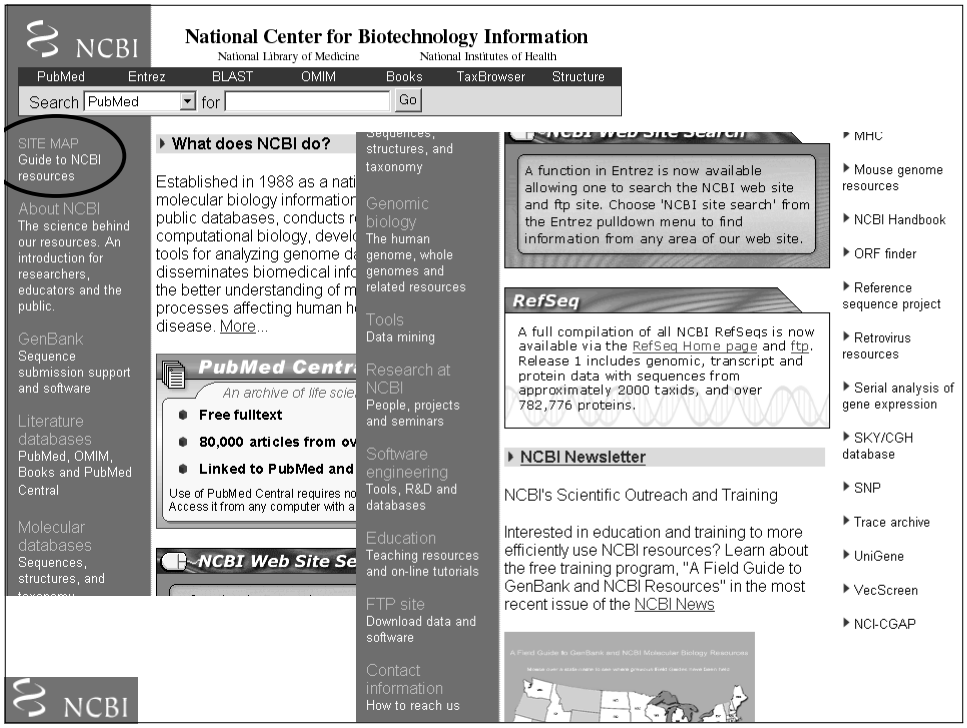
NCBI

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



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 & CGH 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.*

NCBI

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 Go

Entrez
NCBI Web Site
PubMed
Protein
Nucleotide
Structure
Genome
Books
CancerChromosomes
Conserved Domains

SITE MA
Alphabe
Resourc
About N
An intro
for rese
educato
the public
GenBank
Sequence

es NCBI do?

d in 1988 as a national resource for
biology information, NCBI creates
bases, conducts research in
nal biology, develops software
alyzing genome data, and
es 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

NCBI Databases and Sequence Analysis Tools

Entrez: Search and Retrieval System

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

NCBI

Databases

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

NCBI

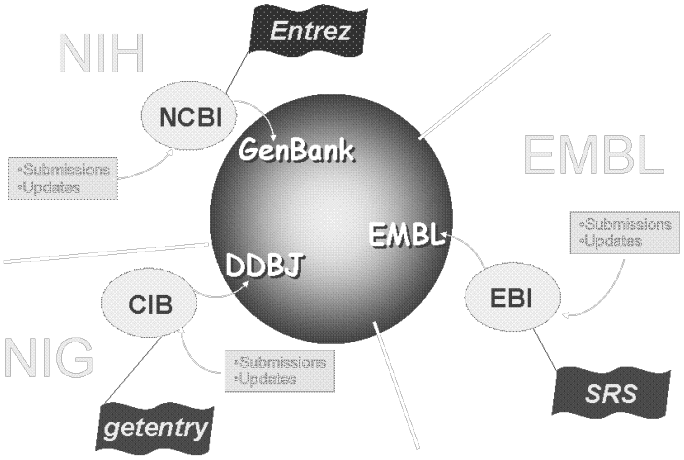
NCBI Databases

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

NCBI Databases

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

NCBI

RefSeq genomes·transcripts·proteins

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

NCBI

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
KEYWORDS
SOURCE
  ORGANIS

REFERENCE
  AUTHORS
  TITLE

  JOURNAL
   PUBMED
  REMARK

REFERENCE
  AUTHORS

  TITLE

  JOURNAL
   PUBMED
  REMARK

FEATURES             Location/Qualifiers
     source          1..2376
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q14"
     gene            1..2376
                     /gene="BIN1"
                     /note="synonyms: AMPH2, AMPHL, SH3P9, MGC10367,
                     DKFZp547F068"
                     /db_xref="GeneID:274"
                     /db_xref="LocusID:274"
                     /db_xref="MIM:601248"
     misc_feature    1
                     /gene="BIN1"
                     /note="5'-most transcription initiation site is
                     undetermined"
     misc_feature    189
                     /gene="BIN1"
                     /note="alternate transcription initiation site"
     CDS             346..1866
                     /gene="BIN1"
                     /note="isoform 3 is encoded by transcript variant 3;
                     amphiphysin-like; amphiphysin II; box dependant MYC
                     interacting protein 1;
                     go_component: nucleus [goid 0005634] [evidence IEA];
                     go_component: cytoplasm [goid 0005737] [evidence IEA];
                     go_component: actin cytoskeleton [goid 0015629] [evidence
                     TAS] [pmid 9182687];
                     go_function: protein binding [goid 0005515] [evidence
                     IEA];
```

NCBI Databases

Protein

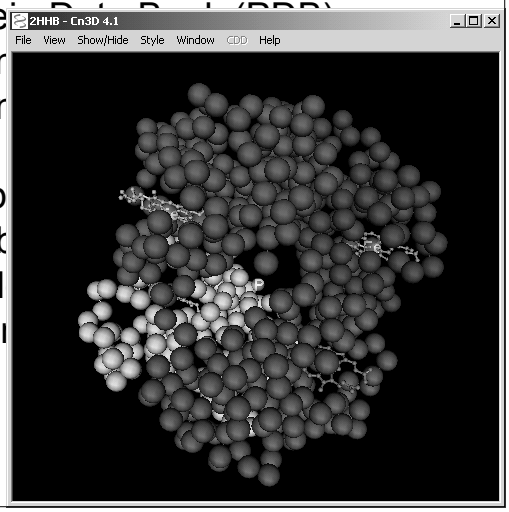
- Conceptual translations of GenBank and RefSeq records
- SwissProt, PIR, PRF, PDB

Molecular Modeling DataBase (MMDB)

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

- obtained from the Protei
- experimentally determin
- can be viewed using Cr

- sequences also availab
  Entrez protein datab
- useful for finding homol
  structures for a protei



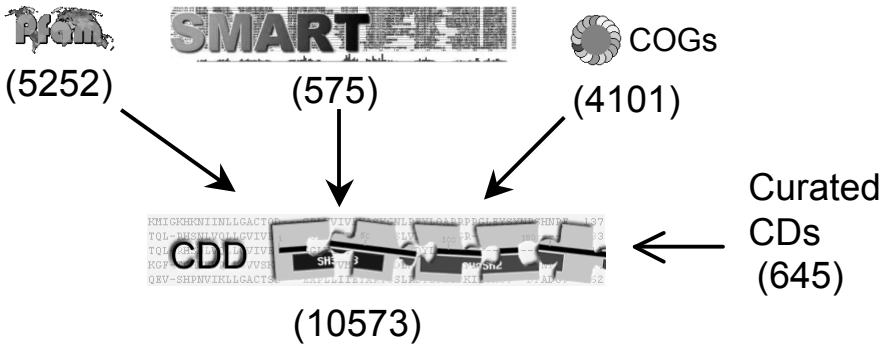
http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

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

NCBI

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

NCBI

Conserved Domain in Beta Globin

cd01040.1 — **globin**

Links:

Source: CDD
Taxonomy: cellular organisms
Pubmed: 5 links
Book: 3 book links
Proteins: cd01040 related
Related CD: 4 links

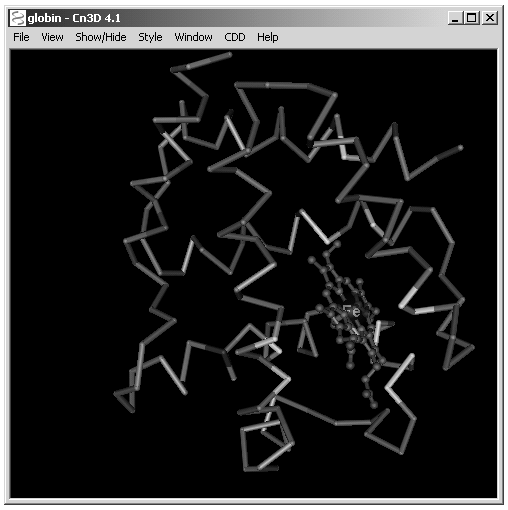
Globins are heme proteins, which bind and transport oxygen. This family summarizes a diverse set of homologous protein domains, including: (1) tetrameric vertebrate hemoglobins, which are the major protein component of erythrocytes and transport oxygen in the bloodstream, (2) microorganismal flavohemoglobins, which are linked to C-terminal FAD-dependend reductase domains, (3) homodimeric bacterial hemoglobins, such as from Vitreoscilla, (4) plant leghemoglobins (symbiotic hemoglobins, involved in nitrogen metabolism in plant rhizomes), (5) plant non-symbiotic hexacoordinate globins and hexacoordinate globins from bacteria and animals, such as neuroglobin, (6) invertebrate hemoglobins, which may occur in tandem-repeat arrangements, and (7) monomeric myoglobins found in animal muscle tissue.

Feature 1: heme-binding site
Evidence: Structure: Ascaris hemoglobin with bound heme and oxygen molecule - View structure with Cn3D 4.1
Comment: Ascaris hemoglobin exhibits strong affinity to oxygen
Citation: PMID 7753786
Structure: Bovine deoxy-hemoglobin A with bound heme - View structure with Cn3D 4.1
Citation: PMID 8411160

Show Alignment | Format: Hypertext | Row Display: up to 5 | Color Bits: 2.0 bits | Type Selection: the most similar members | Feature Display: heme-binding site

```
                    10        20        30        40        50        60        70        80
                ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
Feature 1                                       #         #  ##                  #  #
consensus     1 SAEEKKLVKASWAKLk----aDREEIGLEFYERLFKAHPETRALFSRFGGLSA--ALKGSPKFKAHGKRVLNALDEAIKN 74
query         5 TPEEKSAVTALWGKV------NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTpdAVMGNPKVKAHGKKVLGAFSDGLAH 78
1ASH          1 ANKTRELCMKSLehakvdtsnearQDGIDLYKHMFENYppLRKYFksreeyta-edvqndpfFAKQGQKILLACHVLCAt 79
1FDH_G        5 TEEDKATITSLWgkv------nveDAGGETLGRLLVVYpwTQRFFdsfgnlssasaimgnpkVKAHGKKVLTSLGDAIKh 78
1PBX_B        4 TDKERSIISDIFshm------dydDIGPKALSRCLIVYpwTQRHFsgfgnlynaeaiignanVAAHGIKVLHGLDRGVKn 77
gi 122300     4 SAEEKALVVGLCgkis----ghcdALGGEALDRLFASFgqTRTYFshfdls------pgsadVKRHGGKVLSAIGEAAKh 73
gi 122536     5 TAEEKAAITSVWgkv------nveHDGHDALGRLLIVYpwTQRYFsnfgnlsnsaavagnakVQAHGKKVLSAVGNAISh 78
gi 122542     1 GGSDVSAFLAKVdk---------rAVGGEALARLLIVYpwTQRYFstfgnlgsadaishnskVLAHGQRVLDSIEEGLKh 71
gi 122690     4 TGEEKALVNAVWtkt------dhqAVVAKALERLFVVYpwTKTYFvkfngkf----hasdstVQTHAGKVVSALTVAYNh 73
gi 229556     3 SIADKTSLKNAWgkis----tdttEIGTEALERLHLSFp-TQKFLshg-----------lahVKAHGSKVAGALTSILGp 66
                        90        100       110       120       130       140
```

Conserved Domain in Beta Globin



NCBI Databases

Organisms

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

incorporates phylogenetic and taxonomic knowledge from a variety of sources

NCBI Taxonomy Browser

PubMed Entrez BLAST OMIM

Search for **chiken** As complete name

complete name
wild card
token set
phonetic name
taxonomy id

Taxonomy browser
Archaea
Bacteria

The NCBI Taxonomy

Taxonomy Tip of the Day

Search for chiken as phonetic name

Display 3 levels using filter: none

Search results for phonetic name

chiken

- **Gallus gallus [genbank common name: chicken]**
- **Siagona**

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	**2,491,595**

NCBI

NCBI Home ▸ Genomic Biology ▸ Chicken Genome Resources

Search Map Viewer chicken or (Gallus gallus) Go

Clear

Chicken Genome Resources

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.

Jump to the Genome!
Chromosome: 1
GO

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 Genome

Limits

About Entrez

Entrez Genome
Help

Submiting
Genome Project
Genome sequence

Microbial
Genome Projects
PDB neighbors

Genomic BLAST
Microbial
Eukaryotic

FUNGI Genome projects

WGS projects

New! Asser
NCBI links t
sequence p
about a ger
Archive's f
related str
assemblies
See more ab

Bacillus li

Release Dat
Reference: B
Complete ge
bacterium Ba
comparisons
species
(er) Genome
Lineage: *Ba*
Bacillaceae;
Organism:
Bacillus liche
Genome seq

projects

WGS projects

Archaea
Chromosome
Plasmid
DarftAssembly

Bacteria
Chromosome
Plasmid
DarftAssembly

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: Brettin,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

Submiting
Genome Project
Genome sequence

Microbial
Genome Projects
PDB neighbors

Genomic BLAST
Microbial
Eukaryotic

FUNGI Genome projects

WGS projects

Archaea
Chromosome
Plasmid
DraftAssembly

Bacteria
Chromosome
Plasmid
DraftAssembly

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 VF5		NC_000918	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_003995	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	4303871 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_005957	5237682 bp	Jun 30 2004
Bacteroides fragilis YCH46		NC_006347	5277274 bp	Oct 1 2004

NCBI

NCBI Taxonomy Browser

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books

Search for as complete name lock Go Clear

Display 3 levels using filter: none

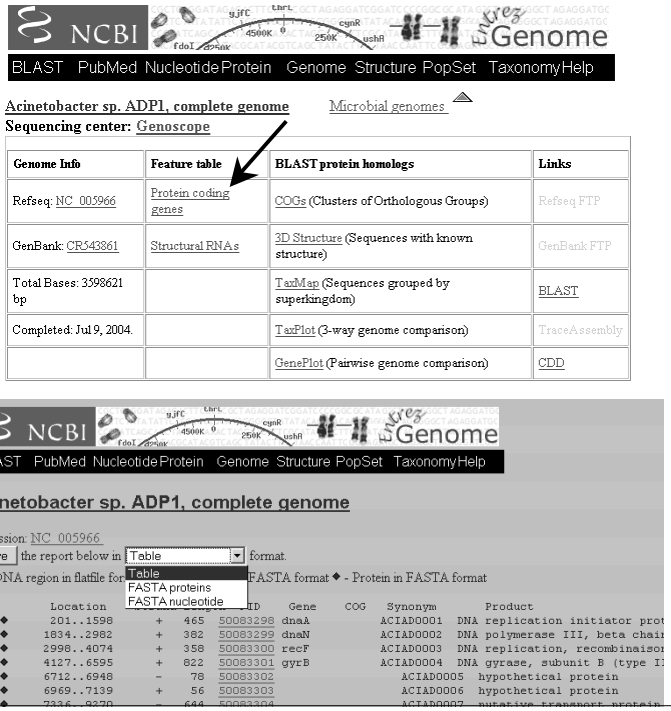
Acinetobacter sp. ADP1

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	2
Genome	1
3D Domains	33
PubMed Central	4
Gene	3,425
Taxonomy	1

NCBI



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

Entrez Gene

NCBI

Entrez Gene

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books | OMIM

Search Gene for Go Clear ☑ current records only

Limits Preview/Index History Clipboard Details

Display Summary Send to Text

Summary
Brief
ASN.1
XML
Graphics
Gene Table
UI List
LinkOut
Books Links
Conserved Domain Links
GENSAT Links

Links
PROTEIN LINKS
FASTA
GENPEPT
BLink
Conserved Domains

☐ 1: HB *sapiens*] MGC cDNA clone, Links
GeneID: MIM: 141900 updated 04-Jan-2005
Transcri verse complement genome) RefSeq below

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; 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 PubMed

Entrez
SITE MAP
Entrez Help
Gene
Search
Gene Help
FAQ
FTP site
Related sites
Entrez Genome
Genomic Biology
HomoloGene
LocusLink
Map Viewer
OMIM
RefSeq
UniGene
Feedback
Help Desk
Corrections
About GeneRIFs
Subscriptions
RefSeq
Gene
Map Viewer

NCBI

Entrez Gene

NCBI

Entrez Gene

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books | OMIM

Search Gene for Go Clear ☑ current records only

Limits Preview/Index History Clipboard Details

Display Gene Table Show: 5 Send to Text

☐ **1: HBB hemoglobin, beta** [*Homo sapiens*] MGC cDNA clone, Links
GeneID: 3043 Locus tag: HGNC:4827; MIM: 141900 updated 04-Jan-2005
total gene size: 1606 bp
Transcripts and products: (shown on reverse complement genome)

NC_000011

5'[5294077] [5293272]3'
NM_000518 NP_000509
■ - coding region ■ - untranslated region

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

Display Gene Table Show: 5 Send to Text

Entrez
SITE MAP
Entrez Help
Gene
Search
Gene Help
FAQ
FTP site
Related sites
Entrez Genome
Genomic Biology
HomoloGene
LocusLink
Map Viewer
OMIM
RefSeq
UniGene
Feedback
Help Desk
Corrections
About GeneRIFs
Subscriptions
RefSeq
Gene

NCBI

Entrez Gene

General gene information

GeneOntology

Provided by GOA

Function	**Evidence**	
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
HPFH, 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

Entrez Gene

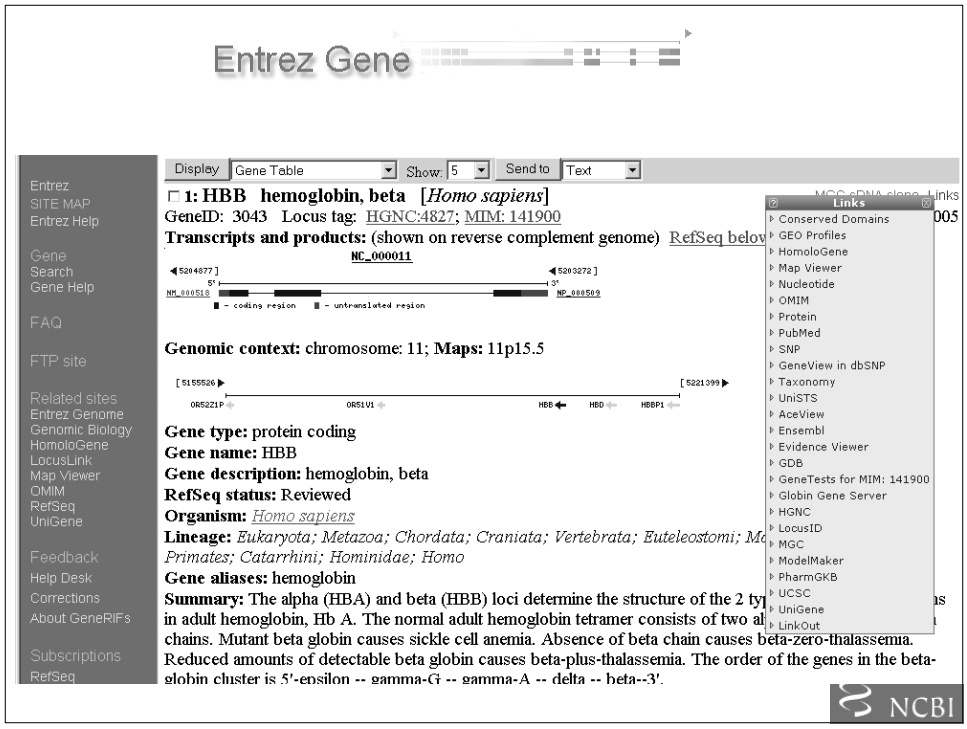
NCBI Reference Sequences (RefSeq)

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

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	P
NM_000518.4	Homo sapiens hemoglobin, beta (HBB), mRNA	P
AY509193.1	Homo sapiens hemoglobin beta mRNA, complete cds	P
CR536530.1	Homo sapiens full open reading frame cDNA clone	P

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

EST Sequences (10 of 2

BI518741.1 cDNA clone IMAGE:5211
BQ890006.1 cDNA clone IMAGE:6295
BQ898811.1 cDNA clone IMAGE:6302

Unix Download se
Unix
PC
Mac
Switch 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 / 28932
Testis	29		4 / 136540
Thymus	146		1 / 6848
Uterus	115		21 / 181622
Vascular	77		2 / 25883

Breakdown by Developmental Stage

Hs.523443

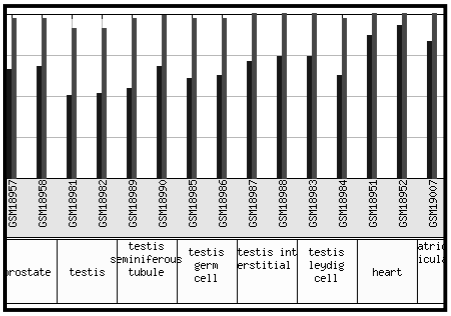
Embryo	1054		536 / 508346
Juvenile	33		2 / 59542
Adult	513		500 / 974089

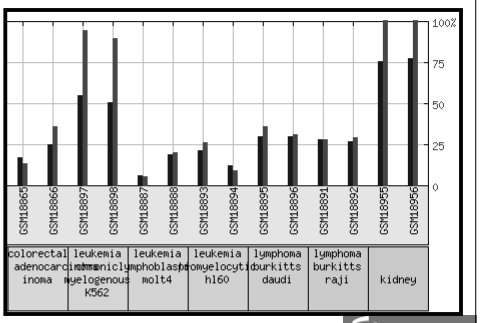
GEO Gene Expression Omnibus

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)





NCBI

Homologene
Discover Homologs

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

NCBI

Homologene for Human HBB
Discover Homologs

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

NCBI

OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University

- 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 mature protein
(selected examples)

Glu7Val in the precursor
Hemoglobin S
sickle cell anemia

- 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, GLU91GLN]
- 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]

NCBI

Entrez Gene

PubMed | 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
[5204877] [5203272]
5' 3'
NM_000518 NP_000509
■ - coding region ■ - untranslated region

Genomic context: chromosome: 11; **Maps:** 11p15.5
[5155526] [5221399]
OR5221P OR51V1 HBB HBD 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 chains in adult hemoglobin, Hb A. The normal adult hemoglobin tetramer consists of two alpha 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

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

Contig position | dbSNP rs# 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
		N.D.		Yes		contig reference	G	Leu [L]	3	89
4035112	rs1803195	N.D.		Yes		nonsynonymous	T	Val [V]	2	84
		N.D.		Yes		contig reference	G	Gly [G]	2	84
4035119	rs11549406	N.D.		Yes		nonsynonymous	G	Val [V]	1	82
		N.D.		Yes		contig reference	C	Leu [L]	1	82
4035245	rs11549407	N.D.		Yes		nonsynonymous	A	Lys [K]	1	40
		N.D.		Yes		contig reference	C	Gln [Q]	1	40
4035473	rs334	N.D.		Yes		nonsynonymous	T	Val [V]	2	7
		N.D.		Yes		contig reference	A	Glu [E]	2	7
4035484	rs713040	0.849		Yes		synonymous	C	His [H]	3	3
		0.849		Yes		contig reference	T	His [H]	3	3

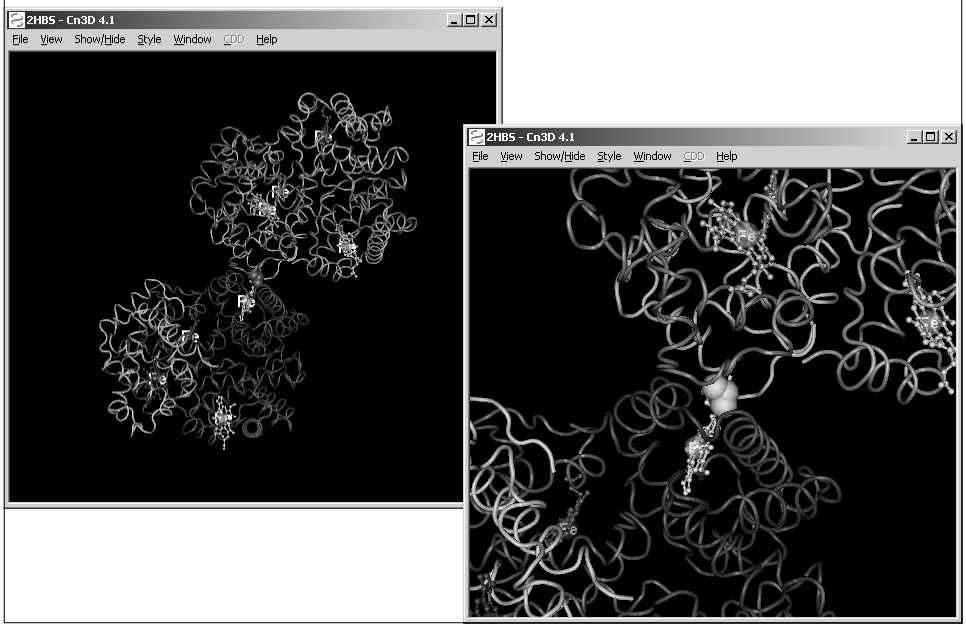
Hemoglobin S

⇧

⇨ Glu6Val in the mature protein

NCBI

Structure of Deoxyhemoglobin S



NCBI

Literature Databases

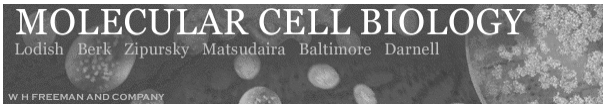
Pubmed	Biomedical literature
PubMed Central	Free online journals

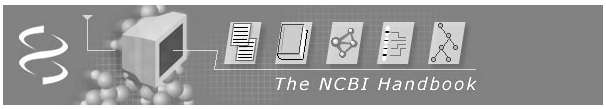
http://www.pubmedcentral.gov

Books	Free online textbooks

Online Books

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





NCBI

Other Databases in Entrez

Cancer Choromosomes 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

NCBI Databases and Sequence Analysis Tools

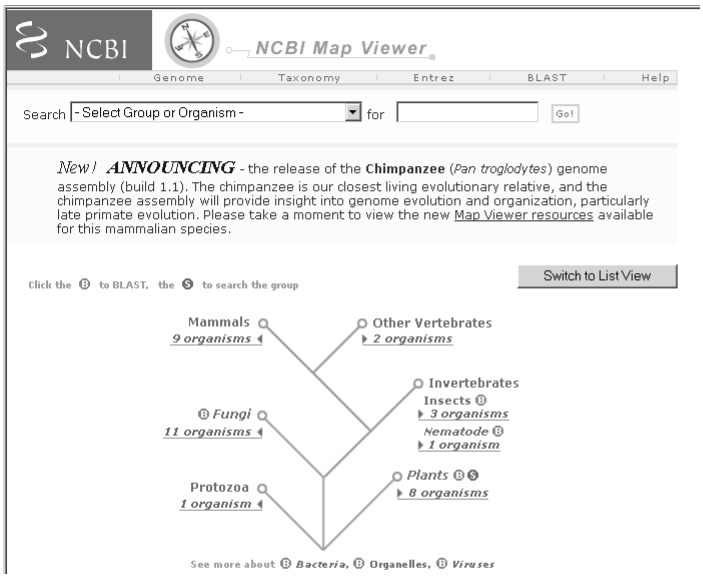
NCBI

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

HBB Gene in the Human Map Viewer

NCBI

NCBI Map Viewer

PubMed | Entrez | BLAST | OMIM | 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

Node1 Genes_seq.

NCBI Map Viewer

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

NCBI

BLAST

NCBI

PubMed | Entrez | BLAST | OMIM | Taxonomy | Structure

About BLAST
- 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

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 (tblastn)
- Translated query vs. translated database (tblastx)

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)
- Immunoglobin BLAST (IgBlast)
- Human SNP BLAST NEW

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

NCBI

Spidey

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

Genomic sequence (FASTA or GI/Accession):

Upload file: Browse...

AC002390

From: 0 To: 0

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

Upload file: Browse...

NM_014164
AF177940

Align
Clear

divergent sequences
Use large intron sizes

Minimum mRNA-genomic identity 0 %
Minimum length of mRNA covered 0 %

Genomic sequence is:

- Vertebrate
- Drosophila
- C. elegans
- Plant

Output options:

- Text/summary
- Summary only
- ASN.1
- Print multiple alignment

NCBI

Spidey

Genomic: gi|2282011|gb|AC002390.1|AC002390 Human DNA from overlapping chromosome 19-specific cosmids R30072 and R28588, genomic sequence

mRNA: gi|21618360|ref|NM_014164.3| Homo sapiens FXYD domain containing ion transport regulator 5 (FXYD5), mRNA

Alignment is on plus strand of genomic sequence and on plus strand of mRNA sequence
mRNA coverage: 100%
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

NCBI

VecScreen

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

Distribution of Vector Matches on the Query Sequence

1 264 528 792 1057

Match to Vector: **Strong** **Moderate** **Weak**

Segment of suspect origin:

Segments matching vector:
Strong match: 1009-1057
Moderate match: 12-33
Weak match: 19-34
Suspect origin: 1-11

```
gnl|uv|X52328.1:714-808 pBluescript II SK(+) vector DNA, phagemid
          excised from lambda ZAPII
          Length = 95

 Score = 84.4 bits (42), Expect = 7e-14
 Identities = 48/49 (97%), Gaps = 1/49 (2%)
 Strand = Plus / Plus

Query: 1009 gccgccaccgcggtggagctcccagcttttgttccctttagtgagggtt 1057
           |||||||||||||||||||| |||||||||||||||||||||||||||
Sbjct: 27  gccgccaccgcggtggagctcc-agcttttgttccctttagtgagggtt 74
```

NCBI

Outline

About NCBI
NCBI databases and tools
The Entrez- search and retrieval system
Training at NCBI

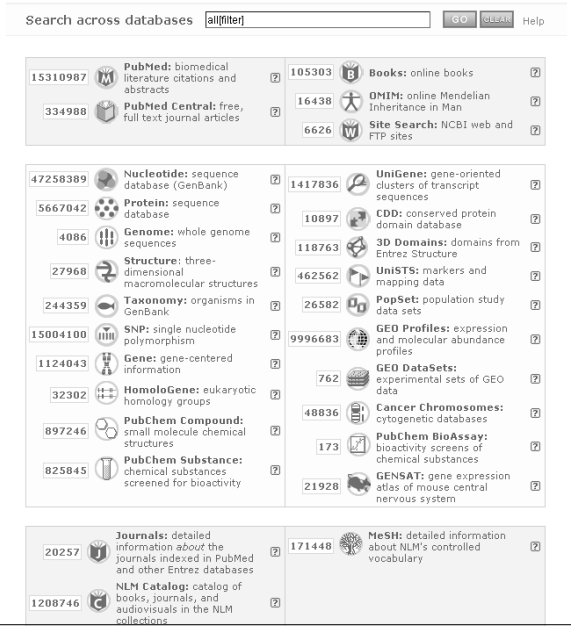
NCBI

Entrez, The Life Sciences Search Engine

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

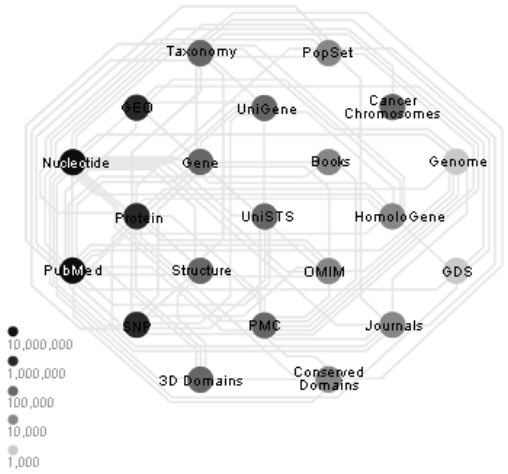
NCBI

Entrez: Search and Retrieval System



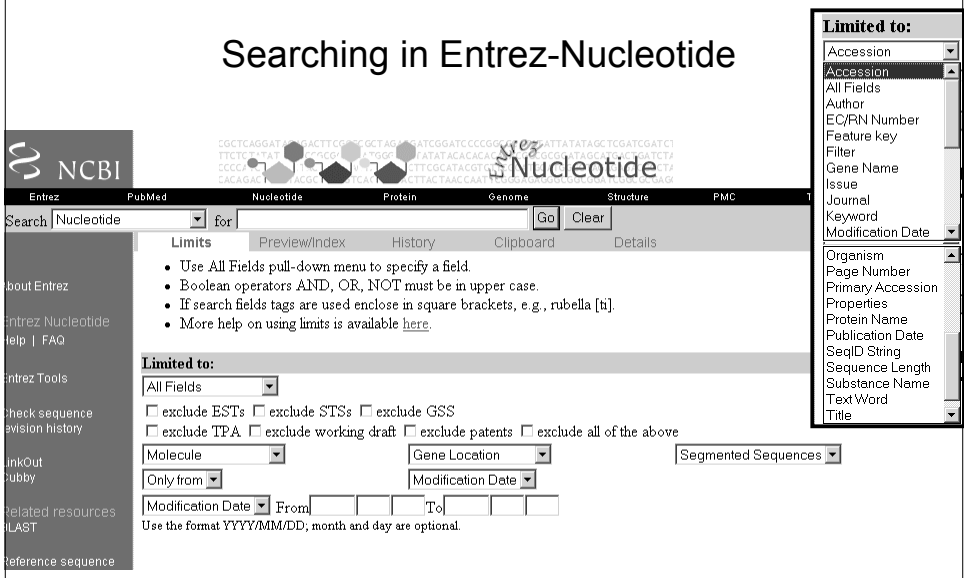
NCBI

Linking within Databases in Entrez



NCBI

Searching in Entrez-Nucleotide



NCBI

Searching for Virus Sequences excluding HIV 1

NCBI Entrez Nucleotide

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy

Search Nucleotide for [ ] Preview Go Clear

Limits | Preview/Index | History | Clipboard | Details

About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut Cubby

- 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.

No history available

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.

Organism | virus | Preview | Index

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

Accession
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

LinkOut Cubby
Related resources BLAST
Reference sequence project
Search for Genes
Submit to GenBank

- Multiple terms selected from Index will be ORed; click AND to add to search.

Organism | virus | Preview | Index

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

virus like particle cak1(1)
virus of serpulina hyodysenteriae 1(1)
virus phich1(8)
viruses(293378)
viscaceae(474)
viscacha rat(5)
viscainoa(3)
viscainoa geniculata(3)
viscaria(1125)
viscaria alpina(1)

Up | Down

NCBI

Searching for Virus Sequences excluding HIV 1

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search Nucleotide for "viruses"[Organism] Preview Go Clear

Limits | Preview/Index | History | Clipboard | Details

About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut Cubby
Related resources BLAST
Reference sequence

- 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
#2	Search **"viruses"[Organism]**	18:06:27	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.

Organism | Preview | Index

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

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

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

Limits | Preview/Index | History | Clipboard | Details

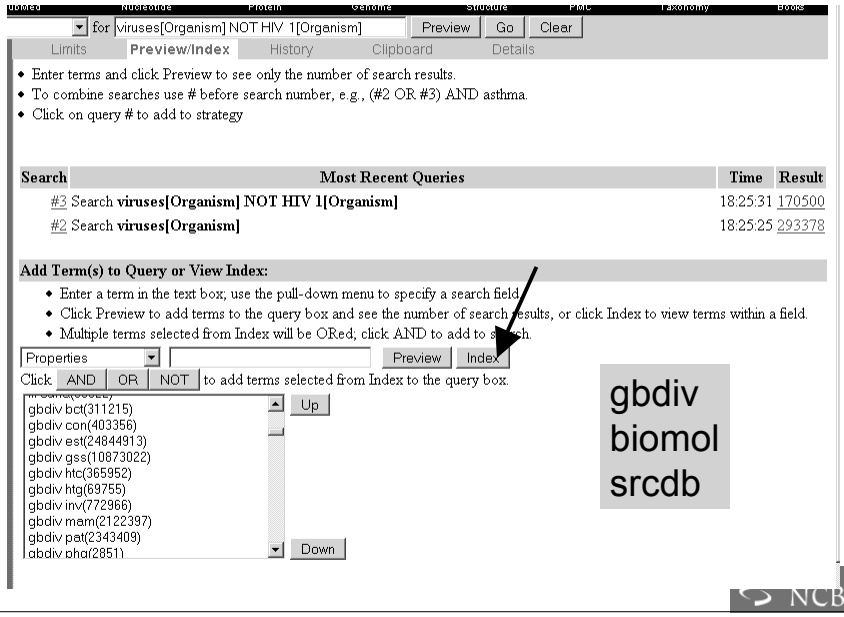
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history

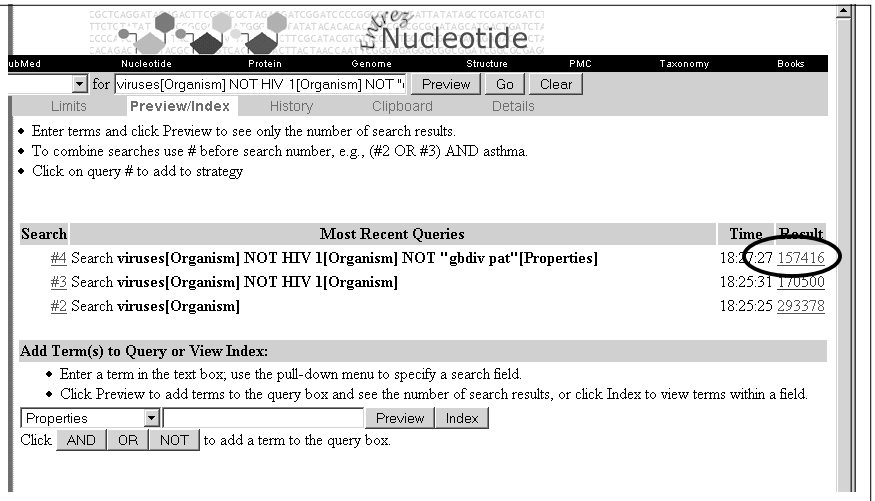
- Enter terms and click Preview to see only the number of search results.
- To combine searches use # before search number, e.g., (#2 OR #3) AND asthma.
- Click on query # to add to strategy

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

NCBI

Searching in Entrez Nucleotide Properties Field



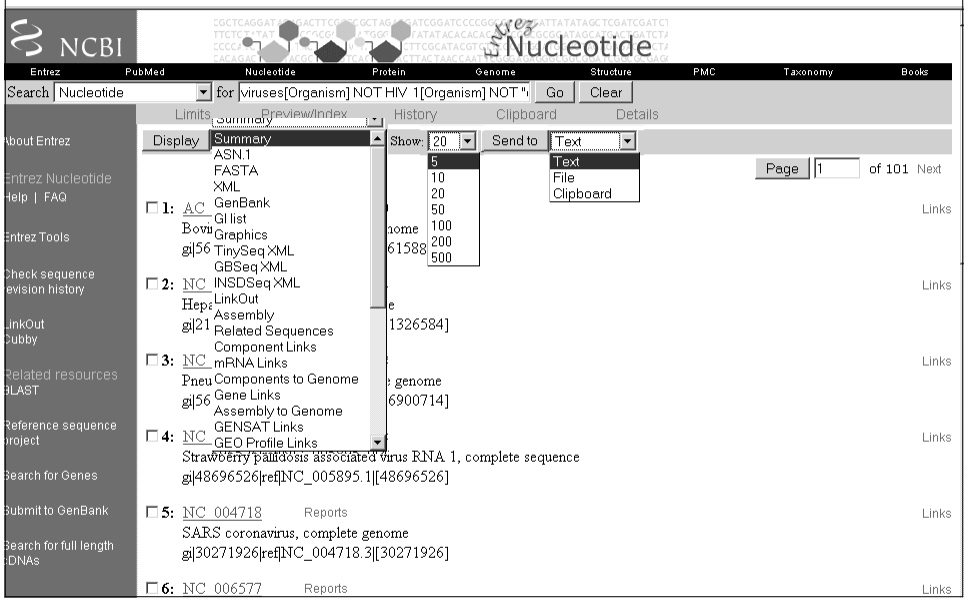


viruses[Organism] NOT HIV 1[Organism] NOT "gbdiv pat"[Properties]

"biomol genomic"[Properties] AND "srcdb refseq"[Properties]

NCBI

Displaying and Saving Sequences in Entrez Nucleotide



Searching in Entrez Nucleotide

NCBI Entrez Nucleotide

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

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

Limits | Preview/Index | History | Clipboard | Details

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

Display Summary Show: 20 Send to Text

Items 1 - 20 of 2010 Page 1 of 101 Next

1: AC_000001 Reports Links
Bovine adenovirus 2, com
gi|56158826|ref|AC_0000

2: NC_003977 Reports Links
Hepatitis B virus, complete
gi|21326584|ref|NC_0039

3: NC_006579 Reports Links
Pneumonia virus of mice, c
gi|56900714|ref|NC_0065

4: NC_005895 Reports Links
Strawberry pallidosis associated virus RNA 1, complete sequence
gi|48696526|ref|NC_005895.1|[48696526]

5: NC_004718 Reports Links
SARS coronavirus, complete genome
gi|30271926|ref|NC_004718.3|[30271926]

6: NC_006577 Reports Links

Reports: ASN.1, XML, Summary, FASTA, TinySeq XML, GenBank, GBSeq XML, INSDSeq XML, GenBank(Full), GI list, Graphic, Revision History

Links: Gene, Protein, PubMed, Taxonomy

Sequence Revision History

Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | OMIM | Books

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 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

PubMed Nucleotide Protein Genome Structure PMC

Search Nucleotide for Go Clear

Limits Preview/Index History Clipboard

Display GenBank Send all to file

Range: from begin to end Reverse complemented strand Features: SNP CDD MGC

1: NC_004718. Reports SARS coronavirus,... [gi:30271926]

```
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
            Viruses; ssRNA positive-stran
            Coronaviridae; Coronavirus.
REFERENCE   1  (bases 1 to 29751)
  AUTHORS   He,R., Dobie,F., Ballantine,
            Cutts,T., Andonov,A., Cao,J.
            Baker,L. and Li,X.
  CONSRTM   BCCA Genome Sciences Centre,
            Control and National Microbio
  TITLE     Analysis of multimerization
            protein
  JOURNAL   Biochem. Biophys. Res. Commun
   PUBMED   15020242
```

PubMed National Library of Medicine NLM

Nucleotide Protein Genome Structure OMIM PMC Journals Books

Limits Preview/Index History Clipboard Details

Display Books Show: 20 Sort Send to Text

1: Biochem Biophys Res Commun. 2004 Apr 2;316(2):476-83. Related Articles, Links

ELSEVIER SCIENCE DIRECT® FULL TEXT ARTICLE

Analysis of multimerization of the SARS coronavirus nucleocapsid protein.

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.

National Microbiology Laboratory, Health Canada, 1015 Arlington St., Winnipeg, MB, Canada R3E 3R2. Runato_He@hc-sc.gc.ca

Severe Acute Respiratory Syndrome (SARS), an emerging disease characterized by atypical pneumonia, has recently been attributed to a novel coronavirus. The genome of SARS Coronavirus (SARS-CoV) has recently been sequenced, and a number of genes identified, including that of the nucleocapsid protein (N). It is noted, however, that the N protein of SARS-CoV (SARS-CoV N) shares little homology with nucleocapsid proteins of other members of the coronavirus family [Science 300 (2003) 1399; Science 300 (2003) 1394]. N proteins of other coronavirus have been reported to be involved in forming the viral core and also in the packaging and transcription of the viral RNA. As data generated from some viral systems other than coronaviruses suggested that viral N-N self-interactions may be necessary for subsequent formation of the nucleocapsid and assembly of the viral particles, we decided to investigate SARS-CoV N-N interaction. By using mammalian two-hybrid system and sucrose gradient fractionations, a homotypic

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 pccompound"[Filter]

Homologene:

Entries for human disease genes

"link phenotype omim"[Properties]

NCBI

NCBI

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | OMIM | PMC | Journals

Advanced Entrez 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.

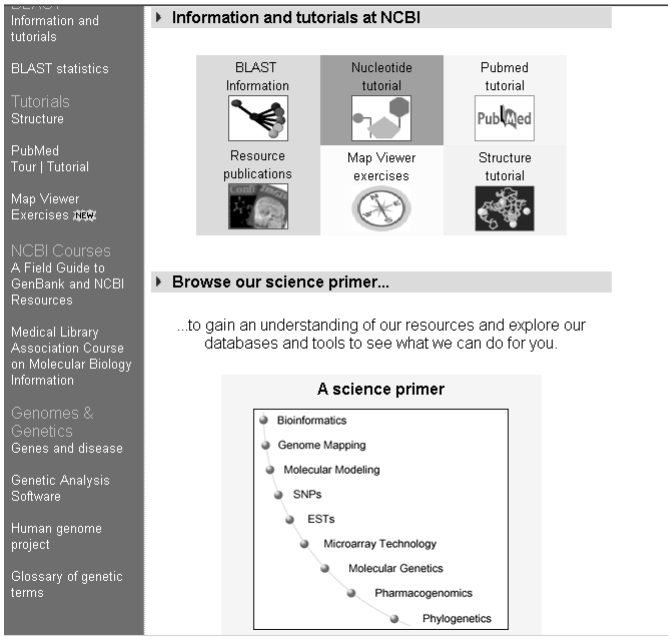
http://www.ncbi.nlm.nih.gov/entrez/query/static/advancedentrez.html

NCBI

Outline

About NCBI
NCBI databases and tools
The Entrez- search and retrieval system
Training at NCBI

NCBI



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

NCBI

NCBI Training

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

NCBI Field Guide

A Field Guide to GenBank and NCBI Molecular Biology Resources

3 hour lecture and 2 hour hands-on

Mini Courses — NCBI Resources one course at a time

on specific topics 2 hour lecture and hands-on

POWER NCBI ENTREZ E-utilities course tools

Three day workshops at NCBI

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

NCBI

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

NCBI

Outline

About NCBI
NCBI databases and tools
The Entrez- search and retrieval system
Training at NCBI

NCBI