

A Field Guide part 2

- Entrez searching
- Genome resources
- Sequence similarity

GenBank Records

The Flatfile Format

```
LOCUS       AAF02223             30 aa             linear   V01 02-OCT-2004
DEFINITION  BACH dehydrogenase subunit 2 [Crotalaria purcutata].
ACCESSION   AAF02223
VERSION     AAF02223.1  GI:19702641
REVISION   accession AF402997.1
KEYWORDS
SOURCE      Crotalaria purcutata (channel catfish)
ORGANISM   Crotalaria purcutata
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Psectroptera; Teleostei; Ostariophysi; Characiformes;
            Ictalurini; Ictalurus.
REFERENCE   1. (containing 1 to 30)
            Waldman, J.C., Wilson, A.L. and Rimmey, J.D.
            Complete sequence and characterization of the channel catfish
            mitochondrial genome.
            JOURNAL   DNA Seq. 14 (4), 265-277 (2003)
REFERENCE   2. (containing 1 to 30)
            Waldman, J.C.
            TITLE      Siren Subgenus
            JOURNAL   Submitted 03-OCT-1997 Catfish Genetics Research Unit, USDA -
            Agricultural Research Service, 141 Experiment Station Road,
            Stoneville, MS 38776, USA
            (containing 1 to 30)
REFERENCE   3. (containing 1 to 30)
            Waldman, J.C.
            TITLE      Siren Subgenus
            JOURNAL   Submitted 12-FEB-2002 Catfish Genetics Research Unit, USDA -
            Agricultural Research Service, 141 Experiment Station Road,
            Stoneville, MS 38776, USA
            Remark: conceptual translation supplied by author.
COMMENT     Location/Qualifiers
FEATURES             source
                    1..30
                    /organism="Crotalaria purcutata"
                    /organism_taxonid="30551"
                    /taxon="Bachi"
                    /db_xref="taxon:30551"
                    1..30
                    /product="BACH dehydrogenase subunit 2"
                    1..30
                    /gene="Bach"
                    /code_by="AF402997.1:4003..5940"
                    /transl="MGI:1929074"
ORIGIN          1  aagpvtlll pslglpqlat fawhllllw wplntllal plawdhhp swvwtatgt
                    41  lpsnallt lftalllnt tprmltlclt lpsallllt alawllwv mhdwprwq
                    81  qltllllla wvplwplq llwllwlll qllllllll pvlppwpl wplclllla
                    121  ywawllllw lrvlrvpplv wllvllplv wvavllllt lswvllllt wvavvplv
                    161  lswvllllt lrvplvplv wvllvllt wvllvllt wvllvllt wvllvllt
                    201  tlllplllw wvplwplq wplwllllt lalllllll pqlatllt
//
```

Header

Feature Table

Sequence

A Typical GenBank Record

```
LOCUS       NM_019570             4279 bp             mRNA             linear   ROD 28-OCT-2004
DEFINITION  Mus musculus Rev1-like (S. cerevisiae) (Rev1), mRNA.
ACCESSION   NM_019570
VERSION     NM_019570.3  GI:50811869
```

LOCUS NM_019570 4279 bp mRNA linear ROD 28-OCT-2004
DEFINITION Mus musculus Rev1-like (S. cerevisiae) (Rev1), mRNA
ACCESSION NM_019570
VERSION NM_019570.3 GI:50811869

= Title

Version, gi
only change when
sequence changes

```
KEYWORDS
.

FEATURES             source
                    1..4279
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /chromosome="1"
                    1..4279
                    /gene="Rev1"
                    /note="synonyms: Rev1, MGC66950, 1110027I23Rik"
                    /db_xref="GeneID:56210"
                    /db_xref="LocusID:56210"
                    /db_xref="MGI:1929074"
                    212..3961
                    /gene="Rev1"
                    /note="REV1 protein;
                    go_component: intracellular [goid 0005622] [evidence I28]
```

GenBank Record: Feature Table

FEATURES	Location/qualifiers
source	1..4279 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /chromosome="1"
gene	1..4279 /gene="Rev1" /note="synonyms: Rev1, MGC66950, 1110027I23Rik" /db_xref="GeneID:56210" /db_xref="LocusID:56210" /db_xref="MGI:1929074"
CDS	212..3961 /gene="Rev1" /note="REV1 protein;

GenBank Record: Feature Table, con't.

CDS

```

212..3961
/gene="Rev11"
/note="REV1 protein;
go_component: intracellular [goid 0005622] [evidence ISS]
[pmid 12466851];
go_function: transferase activity [goid 0016740] [evidence
IEA];
go_function: deoxycytidyl transferase activity [goid
0017125] [evidence IDA] [pmid 11711549];
go_process: DNA repair [goid 0006281] [evidence ISS] [pmid
12466851];
go_process: error-prone postreplication DNA repair [goid
0042276] [evidence IDA] [pmid 11711549]"
/codon_start=1
/product="REV1-like"
/protein_id="NP_062516.2"
/db_xref="GI:40254551"
/db_xref="GeneID:56210"
/db_xref="LocusID:56210"
/db_xref="MGI:1929074"
/translation="MRRGGWRKRTENDGWEKGGYMAAKVQKLEEQRTDAANQDGT
ASAI FSGVAIYVNGYTDPSAEELRNMLHGGQYHVYYSRSKTHI IATNLPNAKIKE
LKGKVI RPEWIVESIKAGRLLSAPYQLYTKPSAAQKSLNFNPKPEDPGPGPSNR
AKQLNNRVNHI IKKIESEVKNGLSWSNEDGVNDDFSEFLEHTFPKQNGVMHP
RDTAVIFNGHTSSNGALKTDQCLVPGNSVASRLSLDSTQEEKRAEKSNAFRDCTV
QHLQHSRSDALRSRHRTNSLSPSLHSNTKINGAHSIVQGPSSTKTSVLTLSKVA

```

GenBank Record: sequence

```

ORIGIN
1 ctgccgcgga ggcgcgaggc gggcgagcg actgagggag cgtgggccgc ggtgcggtgc
61 ggtgcgctg ggccgcgccc tgaccggtg ccttcccgc cgttcccctc agccgcgggg
121 aggcactgca acgacgcggc cgcccgagag gcccgaggaa cggcgtggcc ccggcatgg
181 tggcctgggg cggcgcgaaa gaagctccac catgagcga ggtgatggc ggaagagaac
241 ggagaatgat ggctgggaaa aatggggcgg gtacatgctt gccaaagtcc agaagtgga
301 ggagcagttc cgcacagatg ctgctaacca gaagatggg acgggctctg caatatttag
361 tggagttgct atctatgta atggatatac agatcctctc gggaggaaat tgagaatctc
421 aatgatgttg catggaggtc aatccatgtg gtattattct agatccaaaa caacacacat
481 tattgtctaca atcttctcta atgccaaaat taaagaatta aagggggaaa aagtaatacg
541 accagaatgg attgtggaaa gcatacaaac tgggagactc ttgtcctctg ctccctatca
601 gctctacacc aagccgctcg ctgcacagaa gactctcaac ttcaatcctg tgtgcaaac
661 tgaagatcct gggccaggtc caagcaatg agcacaacag ctcaacaaca ggtgaaatca
721 cataattaag aagattgaga cagaagaatg agtgaagaca aatggactga gcagctggaa
781 tgaagacggt gtaaatgatg attttagttt tgaggatctg gaacatacat ttccaggaag
841 gaaacagaat ggagtaatgc atcccagaga taccgctgtc attttaatg gacacacca
901 taattcaaat ggtgccttaa agacacagga ttgcttggtg cctgtgggca acagtgttc
961 tagcaggcta tctttagact ccaccaggga ggagaagagg gcagagaaga gcaacgggga
1021 cttcagagac tgacacagtc agcacttgca acacagacc agaagtgcag atgtcttgg
1081 gagtccacac agaactaatt ctctctcacc atctttgac agtaacacta aatcaatgg

skip
4201 tatggattt tacagtgaag tgtttacctg tgtttaataa agaactgtat gtatattttg
4261 taaaaaaaaa aaaaaaaaaa
//
NCBI FieldGuide

```

Indexing for Nucleotide UID 59958365

```

LOCUS       NM_001012399                1168 bp    mRNA    linear    MAM 19-FEB-2005
DEFINITION Bos taurus hemochromatosis (hfe), mRNA.
ACCESSION  NM_001012399
VERSION    NM_001012399.1  GI:59958365
KEYWORDS   .
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
            Pecora; Bovidae; Bovinae; Bos.

```

Field Indexed Terms

[primary accession]	NM_001012399 [accn]
[title]	Bos taurus hemochromatosis (hfe), mRNA.
[organism]	Bos taurus [orgn]
[sequence length]	1168 [slen] 3000:6000[slen]
[modification date]	2005/02/19 [mdat] 2006/01:2006/08[mdat]
[properties]	biomol mrna [prop]
	gbdiv mam
	srcdb refseq

Global Entrez Search: HFE

Search across databases **HFE**

137	Nucleotide: sequence database (GenBank)	3	UniGene: gene-oriented clusters of transcript sequences
1	Protein: sequence database	none	CDD: conserved protein domain database
1	Genome: whole genome sequences	32	3D Domains: domains from Entrez Structure
3	Structure: three-dimensional macromolecular structures	6	UniSTS: markers and mapping data
none	Taxonomy: organisms in GenBank	none	PopSet: population study data sets
127	SNP: single nucleotide polymorphism	1160	GEO Profiles: expression and molecular abundance profiles
22	Gene: gene-centered information	none	GEO DataSets: experimental sets of GEO data
16	HomoloGene: eukaryotic homology groups	none	Cancer Chromosomes: cytogenetic databases
2	PubChem Compound: small molecule chemical structures	none	PubChem BioAssay: bioactivity screens of chemical substances
2	PubChem Substance: chemical substances screened for bioactivity	none	GENSAT: gene expression atlas of mouse central nervous system
none	Genome Project: genome project information	none	

NCBI FieldGuide

137 records

- 1: [NM_001012399](#) Reports
Bos taurus hemochromatosis (hfe), mRNA
gi|59958365|ref|NM_001012399.1|[59958365]
- 2: [AY898807](#) Reports
Bos taurus HFE mRNA, complete cds
gi|58759917|gb|AY898807.1|[58759917]
- 3: [NM_002032](#) Reports
Homo sapiens ferritin, heavy polypeptide 1 (FTH1), mRNA
gi|56682958|ref|NM_002032.2|[56682958]
- 4: [NM_000146](#) Reports
Homo sapiens ferritin, light polypeptide (FTL), mRNA
gi|56682960|ref|NM_000146.3|[56682960]
- 5: [NM_139011](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 11, mRNA
gi|21040356|ref|NM_139011.1|[21040356]
- 6: [NM_139010](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 10, mRNA
gi|21040354|ref|NM_139010.1|[21040354]

hfe[title]

NCBI FieldGuide

Smarter Query

hfe[title] AND human[orgn]

42 records

- 1: [NM_139011](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 11, mRNA
gi|21040356|ref|NM_139011.1|[21040356]
- 2: [NM_139010](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 10, mRNA
gi|21040354|ref|NM_139010.1|[21040354]
- 3: [NM_139009](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 9, mRNA
gi|21040352|ref|NM_139009.1|[21040352]
- 4: [NM_139008](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 8, mRNA
gi|21040350|ref|NM_139008.1|[21040350]
- 5: [NM_139007](#) Reports
Homo sapiens hemochromatosis (HFE), transcript variant 7, mRNA
gi|21040348|ref|NM_139007.1|[21040348]

Curated HFE splice variants (11 total)

NCBI FieldGuide

hfe[title] AND human[orgn] (con't)

Primary data

- 15: [AF461190](#)
Homo sapiens hemochromatosis (HFE) gene, intron 3
gi|18483172|gb|AF461190.1|[18483172]
- 16: [AJ250635](#)
Homo sapiens mRNA for Hemochromatosis protein (HFE gene), DELEX2+3 splice form
gi|16751450|emb|AJ250635.1|HSA250635[16751450]
- 17: [AJ298840](#)
Homo sapiens partial mRNA for HFE gene, splice variant end form 2, 3' UTR
gi|15963339|emb|AJ298840.1|HSA298840[15963339]
- 18: [AJ298839](#)
Homo sapiens partial mRNA for HFE gene, splice variant end form 1, 3' UTR
gi|15963338|emb|AJ298839.1|HSA298839[15963338]
- 19: [AJ249338](#)
Homo sapiens mRNA for hemochromatosis protein (HFE gene) splice variant 4
gi|15485424|emb|AJ249338.1|HSA249338[15485424]

NCBI FieldGuide

Preview/Index

Gateway to Advanced Searches

The Entrez Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases in these databases continues to grow at an exponential rate. As of June 2005, there are over 89 billion bases in GenBank and RefSeq alone.

NCBI FieldGuide

Preview/Index

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

Preview to see only the number of search results.
Use # before search number, e.g., (#2 OR #3) AND asthma.
to strategy

Most Recent Queries	Time	Result
AND human[orgn]	11:00:40	42

or View Index:
Text box; use the pull-down menu to specify a search field.
Terms to the query box and see the number of search results,
Terms within a field.

Preview Index

to add a term to the query box.

NCBI FieldGuide

Preview/Index: Properties, srcdb

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 asthma.
- Click on query # to add to strategy

Search	Most Recent Queries	Time	Result
#1 Search hfe[title] AND human[orgn]		11:00:40	42

Add Term(s) to Query **srcdb = source database**

- 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 Preview Index

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

click to see Index

NCBI FieldGuide

Preview/Index: Properties, srcdb

Properties srcdb Preview Index

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

- srcdb ddbj(6971322)
- srcdb ddbj/embl/genbank(49484945)
- srcdb embl(5199812)
- srcdb genbank(37313811)
- srcdb pdb(5561)
- srcdb refseq(545430)
- srcdb refseq inferred(3494)
- srcdb refseq known(273967)
- srcdb refseq model(111494)
- srcdb refseq predicted(11603)

...AND srcdb refseq[Properties]

NCBI FieldGuide

Preview/Index: Properties, srcdb

Properties srcdb Preview Index

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

- srcdb ddbj(6971322)
- srcdb ddbj/embl/genbank(49484945)
- srcdb embl(5199812)
- srcdb genbank(37313811)
- srcdb pdb(5561)
- srcdb refseq(545430)
- srcdb refseq inferred(3494)
- srcdb refseq known(273967)
- srcdb refseq model(111494)
- srcdb refseq predicted(11603)

...AND srcdb ddbj/embl/genbank[Properties]

NCBI FieldGuide

'Properties' Search Field

#1	hfe	137
#2	hfe[title] AND human[orgn]	42
#3	#2 AND srcdb refseq[prop]	11
#4	#2 AND srcdb ddbj/embl/genbank[prop]	31
#5	#4 AND gbdiv pri[prop]	29
#4	#4 AND gbdiv est[prop]	2

Primate division gbdiv pri[prop]
EST division gbdiv est[prop]

NCBI FieldGuide

'Properties' Search Field: biomol

#1	hfe	116
#2	hfe[title] AND human[orgn]	42
#3	#2 AND biomol mrna[prop]	29
#4	#2 AND biomol genomic[prop]	13

Genomic DNA biomol genomic[prop]
cDNA biomol mrna[prop]

NCBI FieldGuide

More Queries...

Fields are database-specific

Entrez Nucleotide

Reviewed RefSeqs with transcript variants:

srcdb refseq reviewed[prop] AND transcript[title] AND variant[title]

Properties

Click to add terms selected from box.

- srcdb genbank(15715286)
- srcdb pdb(6371)
- srcdb refseq(1180672)
- srcdb refseq inferred(53)
- srcdb refseq known(508082)
- srcdb refseq model(254211)
- srcdb refseq predicted(13429)
- srcdb refseq provisional(416448)
- srcdb refseq reviewed(60744)
- srcdb refseq validated(6462)

NCBI Field

More Queries...

Fields are database-specific

Entrez Nucleotide

Reviewed RefSeqs with transcript variants:

srcdb refseq reviewed[prop] AND transcript[title] AND variant[title]

Entrez Gene

Topoisomerase genes from Archaea:

topoisomerase[gene name] AND archaea[organism]

Genes on human chromosome 2 with OMIM links

2[chromosome] AND human[organism] AND "gene omim"[filter]

Membrane proteins linked to cancer:

"integral to plasma membrane"[gene ontology] AND cancer[dis]

NCBI FieldGuide

Other Entrez Databases

UniGene: rat clusters that have at least one mRNA
 rat[organism] **NOT** 0[mrna count]

SNP: uniquely mapped microsatellites on human chr2

microsat[SNP Class] **AND** 1[Map Weight] **AND** 2[Chromosome] **AND** human[orgn]

UniSTS: markers on the Genethon map of human chromosome 12
 Genethon[Map Name] **AND** human[organism] **AND** 12[chromosome]

Structure: structures of bacterial kinases with resolutions below 2 Å
 bacteria[organism] **AND** kinase **AND** 000.00:002.00[resolution]

NCBI FieldGuide

Genome Resources

Genomic Biology



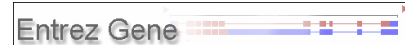
Homologene



Map Viewer



Entrez Gene



NCBI FieldGuide

Genomic biology
 The human genome, whole genomes, and related resources

NCBI FieldGuide

Organism-Specific

- Genome Resources
- BLAST
- Map Viewer
- Genome Project DB

Arabidopsis	Human
Aspergillus	Malaria
Bee	Mosquito
Cat	Mouse
Chicken	Nematode
Chimpanzee	Pig
Cow	Rat
Dictyostelium	Rhesus macaque
Dog	Sea Urchin
Frog	Sheep
Fruit Fly	Yeast (<i>S.cerevisiae</i>)
	Zebrafish

NCBI FieldGuide

Map Viewer – Genome Annotation Updates

Species	Build	Map Viewer Release
Bos taurus (cow)	3.1	January 3, 2007
Gallus gallus (chicken)	2.1	November 30, 2006
Drosophila mel...		November 21, 2006
Caenorhabditis (nematode)		November 21, 2006
Dictyostelium d...		November 2, 2006
Strongylocentrotus (urchin)		October 18, 2006
Pan troglodytes		October 5, 2006
Homo sapiens (human)	36.2	September 14, 2006
Apis mellifera (bee)	Amel_4.0	August 11, 2006
Rattus norvegicus (rat)	RGSC v3.4	July 6, 2006
Macaca mulatta (rhesus macaque)	1.1	June 23, 2006

Subscriptions

- ▶ [Gene Announce](#)
- ▶ [Genome Announce](#)
- ▶ [Map Viewer Announce](#)
- ▶ [NCBI Announce](#)
- ▶ [RefSeq Announce](#)

NCBI FieldGuide

Genome Resources

- ▶ [Entrez Genome](#)
- ▶ [Fungal Genomes Central](#) NEW
- ▶ [Genome Projects Database](#)
 - [Eukaryotic](#) ← ▶
 - [Fungi](#)
 - [Insects](#)
 - [Mammals](#)
 - [Microbial](#)
 - [Plants](#)
- ▶ [Map Viewer](#)
- ▶ [Organelles](#)
- ▶ [Plant Genomes Central](#)
- ▶ [Viral Resources](#)
 - [Influenza Virus Resource](#)
 - [Retroviruses](#)
 - [Viral Genomes](#)

NCBI FieldGuide

ENTREZ Genome Project

Organism Group: **All Eukaryotes**

Abbreviation: **-- All Animals --**

326 Eukaryotes

179 Org

- Mammals
- Birds
- Fishes
- Flatworms
- Insects
- Reptiles
- Roundworms
- Other Animals**
- All Fungi --
- Ascomycetes
- Basidiomycetes
- Other Fungi
- All Plants --
- Land Plants
- Green Algae

Status: [All] Sequencing Method: [All] [Go] [Reset]

PubMed: R - RefSeq Accessions, G - Entrez Gene, T - Trace Archive, B - BLAST, M - Viewer, F - FTP

Projects Selected: Complete - 23, Assembly - 124, In Progress

Sequence Information

Gene #	Chr	Status	Method	Depth	Release Date	Center/Consortium	CB	PM	G	T	B	Link
100	4	In Progress				National Human Genome Research Institute (NHGRI)						
800	3	Assembly	WGS	8X	02/11/2005	TIGR [more]	CB					
24	7	In Progress	WGS	2X		Washington University (WashU)						
24	7	In Progress	WGS	8X		Washington University (WashU)		PM				
28	7	Assembly	WGS	4X	09/21/2005	Broad Institute	CB					
28		In Progress	WGS	3X		Washington University (WashU)						
3		Assembly	WGS	10X	03/22/2002	The International Consortium for the Sequencing of Anopheles Genome	CB	PM	B	G	T	B

NCBI FieldGuide

ENTREZ Genome Project

13 Eukaryotic Genome Sequencing Projects Selected: Complete - 0, Assembly - 2, In Progress - 11

Organism	Group	Subgroup	TaxID	Genome Size (Mb)	# Chr	Status	Method
Aplysia californica	Animals	Other Animals	6500	1800	17	In Progress	
Biomphalaria glabrata	Animals	Other Animals	6526			In Progress	
Ciona intestinalis	Animals	Other Animals	7719	155	14	Assembly	WGS
Ciona savignyi	Animals	Other Animals	51511	180		Assembly	WGS
Daphnia pulex	Animals	Other Animals	6669			In Progress	WGS
Hydra magnipapillata	Animals	Other Animals	6085			In Progress	
Nematostella vectensis	Animals	Other Animals	45351			In Progress	WGS

NCBI FieldGuide

Genomic Biology

Organism-Specific

- Genome Resources
- BLAST
- Map Viewer
- Genome Project DB

- Aspergillus
- Bee
- Cat
- Chicken
- Chimpanzee
- Cow
- Dictyostelium
- Dog
- Frog
- Fruit Fly

NCBI FieldGuide

ENTREZ Genome Project

Genome Project > *Pan troglodytes* (chimpanzee)

The chimpanzee genome sequence is an important tool in understanding primate evolution

Lineage: *Eukaryota* ; *Melazoa* ; *Chordata* ; *Cranialia* ; *Vertebrata* ; *Eutheria* ; *Euarchontoglires* ; *Primates* ; *Catarrhini* ; *Hominidae* ;

Project data

- Genomic (25)
- Organelle (1)
- mRNA (18492)
- Protein (18504)
- WGS (2)
- EST (6930)
- Publication (6)

Photo: Gerald and Buff Corsi ? California Academy of Sciences

NCBI FieldGuide

ENTREZ Genome Project

Search Map Viewer for Find Advanced Search

Available maps: [View Available Assemblies](#)

Sequence Maps

- Ab initio
- Assembly
- Component
- Contig
- CpG Island
- estHs
- GenBank DNA
- Gene
- Reaseq Transcripts
- Repeats
- STS
- uqHs

NCBI FieldGuide

Genome Resources

Genomic Biology

Homologene **HomoloGene** Discover Homologs

Map Viewer **NCBI Map Viewer**

Entrez Gene **Entrez Gene**

NCBI FieldGuide

Homologene

- No longer UniGene based
- Protein similarities first
- Guided by taxonomic tree
- Includes orthologs and paralogs

orthologs

Species	Number of Genes	HomoloGene groups
H.sapiens	22,873	20,111
P.troglodytes	25,096	17,315
C.familiaris	19,766	16,638
M.musculus	24,175	20,528
R.novogicus	21,991	19,008
G.gallus	18,029	12,103
D.melanogaster	14,033	7,995
A.gambiae	13,909	8,394
C.elegans	20,956	5,187
S.pombe	5,043	3,203
S.cerevisiae	5,963	4,729
K.lactis	5,335	4,454
E.gossypii	4,726	3,943
M.grisea	11,109	6,302
N.craspa	10,079	5,915
A.thaliana	26,659	11,180
O.sativa	33,553	11,055
P.falciparum	5,222	975

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species.

NCBI FieldGuide

MLH1 Cluster

HomoloGene Discover Homologs

Display Summary Show 20 Send to

All: 11 Fungi: 0 Mammals: 0

Items 1 - 11 of 11 One page. Download, Links

1: HomoloGene:208. Gene conserved in Eukaryota

Species	Gene	Description
H.sapiens	MLH1	mutL homolog 1, colon cancer, nonpolyposis...
P.troglodytes	MLH1	MutL protein homolog 1
C.familiaris	LOC477019	similar to MutL protein homolog 1
M.musculus	Mlh1	mutL homolog 1 (E. coli)
R.novogicus	Mlh1	mutL homolog 1 (E. coli)
G.gallus	MLH1	mutL homolog 1, colon cancer, nonpolyposis...
D.melanogaster	Mlh1	Mlh1
A.gambiae	ENSANGG00...ENSANGP00000014016	
A.gambiae	ENSANGG00...ENSANGP00000013484	
S.pombe	SPBC1703.04	hypothetical protein
S.cerevisiae	MLH1	Protein required for mismatch repair in mi...

NCBI FieldGuide

Rice Homolog

Region Displayed: 42,260K-42,270,700 bp

Gene: *Oryza sativa* (rice) P0401G10.8

Links 1 MapViewer

Genes identified as putative homologs of Oryza sativa P0401G10.8 during the construction of HomoloGene:

- H.sapiens MLH1 mutL homolog 1, colon cancer, nonpolyposis...
- P.troglodytes MLH1 MutL protein homolog 1
- C.familiaris LOC477019 similar to MutL protein homolog 1
- M.musculus Mlh1 mutL homolog 1 (E. coli)
- R.novogicus Mlh1 mutL homolog 1 (E. coli)
- G.gallus MLH1 mutL homolog 1, colon cancer, nonpolyposis...
- D.melanogaster Mlh1 Mlh1
- A.gambiae ENSANGG00...ENSANGP00000014016
- A.gambiae ENSANGG00...ENSANGP00000013484
- S.pombe SPBC1703.04 hypothetical protein
- S.cerevisiae MLH1 Protein required for mismatch repair in mismatch repair...
- K.lactis KILAC00955 similar to MutL protein homolog 1
- E.gossypii AF15902 AF15902
- M.grisea GMS423710300 Neurexin protein NC002001.1
- N.craspa AT36211 AT36211
- O.sativa P0401G10.8 P0401G10.8

NCBI FieldGuide

Genome Resources

Genomic Biology

Homologene

Map Viewer

Entrez Gene

NCBI FieldGuide

Vertebrates

Mammals

- BLAST *Bos taurus* (cow) Build 3.1
- BLAST *Bos taurus* (cow) Build 2.1
- BLAST *Canis familiaris* (dog)
- BLAST *Felis catus* (cat)
- BLAST *Homo sapiens* (human) Build 36
- BLAST *Homo sapiens* (human) Build 35
- BLAST *Macaca mulatta* (Rhesus macaque)
- BLAST *Mus musculus* (mouse) Build 36
- BLAST *Mus musculus* (mouse) Build 35
- BLAST *Ovis aries* (sheep)
- BLAST *Pan troglodytes* (chimpanzee)
- BLAST *Rattus norvegicus* (rat)
- BLAST *Sus scrofa* (pig)

Other Vertebrates

- BLAST *Danio rerio* (zebrafish)
- BLAST *Gallus gallus* (chicken)

Invertebrates

Insects

- BLAST *Anopheles gambiae* (mosquito)
- BLAST *Apis mellifera* (honey bee) Amel_4.0
- BLAST *Apis mellifera* (honey bee) Amel_2.0
- BLAST *Drosophila melanogaster* (fruit fly)
- BLAST *Tribolium castaneum* (red flour beetle)

Nematode

- BLAST *Caenorhabditis elegans* (nematode)

Echinoderms

- BLAST *Strongylocentrotus purpuratus* (sea urchin) Build 2
- BLAST *Strongylocentrotus purpuratus* (sea urchin) Build 1

Protozoa

- BLAST *Cryptosporidium parvum*
- BLAST *Dicystidium discoidium*
- BLAST *Plasmodium falciparum*
- BLAST *Thelazia parva*
- BLAST *Trypanosoma brucei*

Plants

- BLAST *Agriopsis tauschii*
- BLAST *Agriopsis umbellata*
- BLAST *Allium cepa* (onion)
- BLAST *Arabidopsis thaliana* (thale cress)
- BLAST *Avena sativa* (oat)
- BLAST *Beta vulgaris* (beet)
- BLAST *Brassica juncea* (Indian mustard)
- BLAST *Brassica napus* (oilseed rape)
- BLAST *Brassica nigra* (black mustard)
- BLAST *Brassica oleracea*
- BLAST *Brassica rapa* (field mustard)
- BLAST *Capsicum annuum*
- BLAST *Eragrostis tef* (tef)
- BLAST *Cytisus medeolense*
- BLAST *Hordeum vulgare* (barley)
- BLAST *Lotus japonicus* (lotus)
- BLAST *Manihot esculenta* (cassava)
- BLAST *Medicago sativa* (alfalfa)
- BLAST *Oryza sativa* (rice)
- BLAST *Phaseolus vulgaris* (French bean)
- BLAST *Populus trichocarpa* (black cottonwood)
- BLAST *Prunus dulcis* (almond)
- BLAST *Sesuvium portulacastrum* (purslane)
- BLAST *Setaria italica* (foxtail millet)
- BLAST *Solanum lycopersicon* (tomato)
- BLAST *Solanum melongena* (eggplant)
- BLAST *Solanum peruvianum* (Peruvian tomato)
- BLAST *Sorghum bicolor* (broomcorn)
- BLAST *Theobroma cacao* (cacao)
- BLAST *Triticum aestivum* (wheat)
- BLAST *Triticum burgoldium* (English wheat)
- BLAST *Vigna radiata*
- BLAST *Zea mays* (corn)

Fungi

- BLAST *Aspergillus fumigatus*
- BLAST *Candida glabrata*
- BLAST *Cryptococcus neoformans*
- BLAST *Debaryomyces hansenii*
- BLAST *Encephalitozoon cuniculi*
- BLAST *Eremothecium gossypii*
- BLAST *Gibberella zeae*
- BLAST *Kluyveromyces fragilis*
- BLAST *Kluyveromyces fragilis*
- BLAST *Neurospora crassa*
- BLAST *Saccharomyces cerevisiae* (baker's yeast)
- BLAST *Schizosaccharomyces pombe* (fission yeast)
- BLAST *Ustilago maydis*
- BLAST *Yarrowia lipolytica*

Bacteria (BLAST), **Organelles** (BLAST), **Viruses** (BLAST)

NCBI Map Viewer

Search for **adar** on chromosome(s) assembly All

Show related entries Help FTP MapViewer home

Mus musculus (house mouse) genome view BLAST search the mouse genome

Build 36.1 statistics Switch to previous build

NCBI FieldGuide

Mus musculus (house mouse) genome view

Build 36.1 statistics Switch to previous build

Search results for query "adar": 28 hits

NCBI FieldGuide

MapViewer: Mouse ADAR, 28 Hits

Chr	Assembly	Match	Map element	Type	Maps
3	reference	all matches			
		Mus musculus RNA adenosine... Adar mRNA, complete cds...	AF291050.1	TRANSCRIPT	Mm RNA
		Mus musculus RNA adenosine... Adar mRNA, complete cds...	AF291875.1	TRANSCRIPT	Mm RNA
		Mus musculus RNA adenosine... Adar mRNA, complete cds...	AF291876.1	TRANSCRIPT	Mm RNA
		Mus musculus RNA adenosine... Adar mRNA, complete cds...	AF291877.1	TRANSCRIPT	Mm RNA
		Mus musculus full open reading... Adar , Adenosine deaminase...	CT010323.1	TRANSCRIPT	Mm RNA
		Mus musculus adenosine deaminase... Adar , transcript variant...	NM_001038587.1	TRANSCRIPT	Mm RNA RefSeq RNA
		Adar : double-stranded RNA-specific adenosine deaminase	Adar	GENE	Genes_seq
		Mus musculus adenosine deaminase... Adar , transcript variant...	NM_019655.2	TRANSCRIPT	Mm RNA RefSeq RNA

Quick Filter: Gene Transcript

Filter

NCBI FieldGuide

Mouse MapViewer: Gene Filter

Search results for query "adar AND (gene[obj_type])": 4 hits

Chr	Assembly	Match	Map element	Type	Map
3	reference	all matches			
		Adar : double-stranded RNA-specific adenosine deaminase	Adar	GENE	Genes_seq
		Adar : double-stranded RNA-specific adenosine deaminase	Adar	GENE	ensGenes
3	Celera	Adar : double-stranded RNA-specific adenosine deaminase	Adar	GENE	Genes_seq
3	MGSCv3	Adar : double-stranded RNA-specific adenosine deaminase	Adar	GENE	Genes_seq

NCBI FieldGuide

Mus musculus Build 36.1 (Current) BLAST The Mouse Genome

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

Query: adar [clear]

Region Displayed: 89,868K bp

Gene: [Adar](#)

Summary of Maps: Download/View Sequence/Evidence

Gene annotations: **exon**, **3' UTR**

Best RefSeq: **Adar** : double-stranded RNA-specific adenosine deaminase, RNA-specific

Maps & Options

2E [89868160.75]

Displayed (left to right):

- Sequence maps--
 - Ab initio
 - Assembly
 - BES_Clone
 - Component**
 - Contig
 - CpG island
 - Ensembl Genes
 - Ensembl Transcripts
 - GenBank DNA
 - Gene**
 - Gene Traps
 - MICER
 - Phenotype
 - RefSeq Transcripts
 - Repeats
 - rnaHS
 - rnaMm
 - rnaRn
 - STS
- Cytogenetic maps--
 - Ideogram
 - Genetic Maps--
 - MGI
 - WL_GEN
 - RH maps--
 - WI/MRC-RH
 - WI-YAC

ugHs = SNP

ugMm

ugRn

Variation

NCBI FieldGuide

Mus musculus Build 36.1 (Current) BLAST The Mouse Genome

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

Query: adar [clear]

Region Displayed: 89,800,500-89,850K bp

Gene: [Adar](#)

Summary of Maps: Download/View Sequence/Evidence

Gene annotations: **Gene annotations**

RefSeq RNA: **RefSeq RNA**

Tiling path: **Tiling path**

Variations: **Variations**

Maps & Options

Organism: Mus musculus [Help](#)
Chromosome: 3 **Region Shown:** 89772198.25 - 89868160.75

Available Maps: Org: mouse Assembly: reference Change Assembly
 ---Se human
 Ab in mouse
 Aese mouse
 BES rat
 Com human
 Contig
 CpG Island
 Ensembl Genes
 Ensembl Transcripts

Maps Displayed (left to right): reference, reference, Gene
 129_substrain
 A/J
 B6/CBAF1J
 Balb/c
 C3H
 Celera
 MGSCv3
 NOD
 unknown

More Options:
 Show Connections Verbose Mode
 Compress Map: off Auto Compress if > 350 px
 Page Length: 30
 Thumbnail View: default (ideogram) master

[Move UP] [Move DOWN] [Make Master/Move to Bottom] [Toggle Ruler] ([R] before map means 'ruler set')

[OK] [Apply] [Close]

NCBI FieldGuide

Synteny

Master Map: Genes On Sequence (M. musculus, chr.3)
 Region Displayed: 89,800,500-89,850K bp

NCBI FieldGuide

Genome Resources

Genomic Biology

Homologene

Map Viewer

Entrez Gene

NCBI FieldGuide

Entrez Gene

Human ADAR

All: 1 Current Only: 1 Genes Genomes: 1 SNP GeneView: 1

1: ADAR adenosine deaminase, RNA-specific [Homo sapiens] updated 24-Jan-2007
 GeneID: 103 Summary

Official Symbol ADAR provided by [HGNC](#)

Official Full Name adenosine deaminase, RNA-specific provided by [HGNC](#)

Primary source [HGNC:1225](#)

See related [HPRD:07528](#); [MIM:601059](#)

Gene type protein coding

RefSeq status Reviewed

Organism [Homo sapiens](#)

Lineage [Eukaryota](#); [Metazoa](#); [Chordata](#); [Craniata](#); [Vertebrata](#); [Euteleostomi](#); [Mammalia](#); [Eutheria](#);

Also known as DSH; G1P1; IFI4; p136; ADAR1; DRADA; DSRAD; IFI-4; K88dsRBP

This gene encodes the enzyme adenosine deaminase, RNA-specific (ADAR), which is involved in RNA editing. This enzyme destabilizes double stranded RNA through conversion of adenosine to inosine. Mutations in this gene have been associated with dyschromotaxis symmetrica hereditaria. Alternate transcriptional splice variants, encoding different isoforms, have been characterized.

NCBI FieldGuide

Entrez Gene

Human ADAR - Genomic regions, transcripts, and products

Genomic regions, transcripts, and products

(minus strand) Go to reference sequence details

Genomic context

chromosome: 1; Location: 1q21.1-q21.2

NCBI FieldGuide

Entrez Gene

Human ADAR - Interactions

Interactions

Description	Interactant	Other Gene	Complex	Source	Pubs
NP_001102.1	NP_003391.1	XPO1		HPRD	PubMed
in vitro; in vivo	BioGRID:113348	XPO1		BioGRID	PubMed

NCBI FieldGuide

Entrez Gene

BioGRID BETA

Search

home support contribute downloads mirrors about us

ADAR *Homo sapiens*

Aliases: ADAR1, DRADA, DSH, DSRAD, IFI4, K88dsRBP, p136, G1P1, HGNC:225

Description: adenosine deaminase, RNA-specific

Links: [Entrez-Gene] [OMIM] [HUGO] [HPRD]

Gene Ontology: 6 total GO mappings for this record. [click here for a full list]

Click here to download interactions associated with ADAR

ADAR was identified in association with 1 results:

Name	Aliases	Description	Evidence Code(s)	Author(s)
XPO1	<ul style="list-style-type: none"> CRU1 DRP2p558 HGNC:12825 	<ul style="list-style-type: none"> exportin 1 (CRM1 homolog, yeast) Gene Ontology: [View List] 	<ul style="list-style-type: none"> In vitro In vivo 	<ul style="list-style-type: none"> Poulsen H et al. Poulsen H et al.

NCBI FieldGuide

Entrez Gene

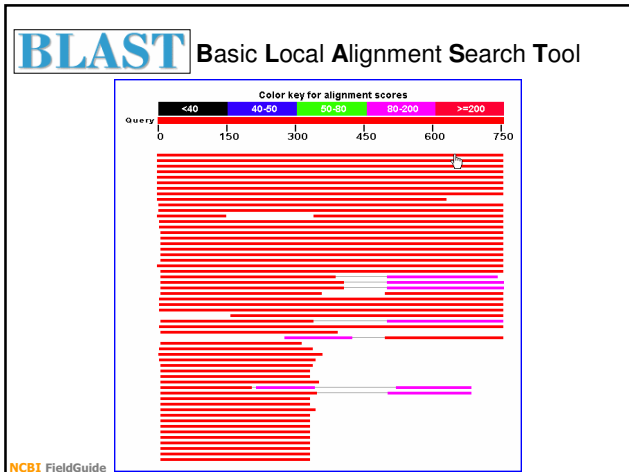
Links

▼ Links Explain

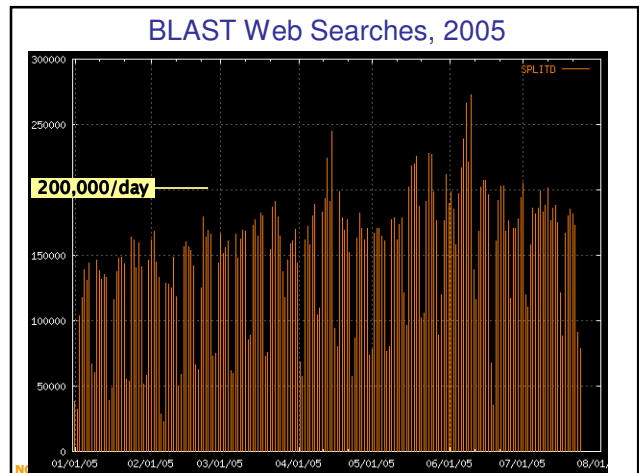
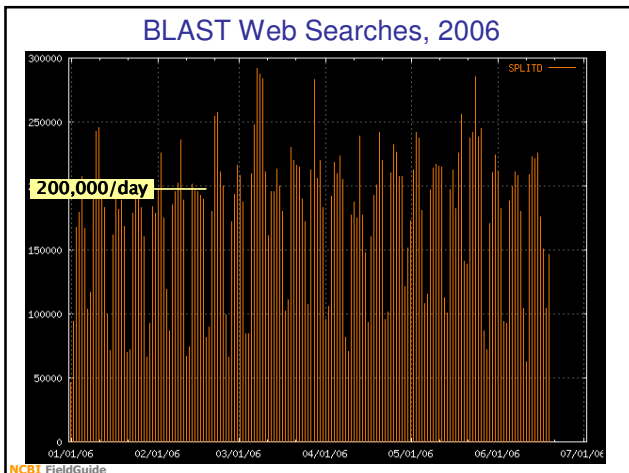
- Order cDNA clone
- Conserved Domains
- Genome
- GEO Profiles
- HomoloGene
- Map Viewer
- Nucleotide
- OMIM
- Full text in PMC
- Probe
- Protein
- PubMed
- PubMed (GeneRIF)
- SNP
- SNP: Genotype
- SNP: GeneView

- Taxonomy
- UniSTS
- AceView
- CCDS
- Evidence Viewer
- GDB
- HGNC
- HPRD
- KEGG
- MGC
- ModelMaker
- Reactome
- UniGene
- LinkOut

NCBI FieldGuide



- ### Web BLAST
- pre-computed results
 - how BLAST works
 - words; scoring matrices; statistics
 - specialized BLAST algorithms
 - what's new, or important
 - example oligo search
- NCBI FieldGuide



Precomputed BLAST Services

- Nucleotide or protein: **Related Sequences**
- BLAST link: **Blink**
- Transcript clusters **UniGene**
- Protein homologs **Homologene**

NCBI FieldGuide

Link to Related Sequences

1: [NM_010424](#) Mus musculus hemo...[gi31981696]

LOCUS NM_010424 1723 bp mRNA linear ROD 22
 DEFINITION Mus musculus hemochromatosis (Hfe), mRNA.
 ACCESSION NM_010424
 VERSION NM_010424.2 GI:31981696
 KEYWORDS .
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus_musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; Mus musculus
 REFERENCE
 AUTHORS Nicolas,G., Viatte,L., Lou,D.Q., Bennou,M., Beaumont,C., Andrews,N.C. and Vaulont,S.
 TITLE Constitutive hepcidin expression prevents iron overload in model of hemochromatosis

Links

- ▶ Gene
- ▶ HomoloGene
- ▶ Master
- ▶ Full text in PMC
- ▶ **Related Sequences**
- ▶ Map Viewer
- ▶ OMIM
- ▶ GEO Profiles
- ▶ Protein
- ▶ PubMed
- ▶ SNP
- ▶ Taxonomy
- ▶ UniGene
- ▶ UniSTS
- ▶ LinkOut

NCBI FieldGuide

Related Sequences

Display Summary Show: 20 Send to: Text One page

Items 1-11 of 11

1: [NM_010424](#)
 Mus musculus hemochromatosis (Hfe), mRNA
 gi31981696|ref|NM_010424.2|31981696| **Most similar** Links

2: [AK009581](#)
 Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone 231004 product:hemochromatosis, full insert sequence
 gi12844462|dbj|AK009581.1|12844462| Links

3: [AK088986](#)
 Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone E430034J19 product:hemochromatosis, full insert sequence
 gi26354115|dbj|AK088986.1|26354115| Links

4: [U66849](#)
 Mus musculus here:binary haemochromatosis homolog mRNA, complete cds
 gi1519484|gb|U66849.1|MMU66849|1519484| Links

5: [AJ306425](#)
 Mus musculus partial mRNA for hemochromatosis protein (Hfe gene), 3' UTR long form
 gi18873691|emb|AJ306425.1|MMU306425|18873691| Links

6: [AL590388](#)
 Mouse DNA sequence from clone RP23-480B19 on chromosome 13, complete sequence
 gi15858949|emb|AL590388.4|15858949| **Least similar** Links

BLink (BLAST Link)

1: [NP_034554](#) Reports
 hemochromatosis [Mus musculus]
 gi31981697|ref|NP_034554.2|31981697|

BLink

Links

- ▶ Gene
- ▶ Genome Project
- ▶ HomoloGene
- ▶ Full text in PMC
- ▶ Gene Genotype
- ▶ GeneView in dbSNP
- ▶ UniGene
- ▶ **Related Sequences**
- ▶ Domain Relatives
- ▶ Genome
- ▶ Map Viewer
- ▶ Nucleotide
- ▶ OMIM
- ▶ PubMed
- ▶ SNP
- ▶ Taxonomy

NCBI FieldGuide

BLink Output

Best hits
3D structures
CDD-Search

Query: gi|1981697|hemochromatosis [Mus musculus]
 Matching gi: 56237867, 1344463

Show identical | Best hits | Common Tree | Taxonomy Report | 3D structures | CDD-Search | GI list

200 BLAST hits to 22 unique species [Sort by taxonomy proximity](#)

Archaea
 Bacteria
 Metazoa
 Fungi
 Plants
 Viruses
 Other Eukaryote

Keep only: Cut-Off: [100]

359 aa	SCORE	E	ACCESSION	GI	PROTEIN DESCRIPTION
<i>Conserved Domain Database hits</i>					
1929	30		F70387	2219802	Hereditary hemochromatosis protein homolog
1914	30		BAC40698	26354116	unnamed protein product [Mus musculus]
1708	28		AAH70942	4747797	Hemochromatosis [Rattus norvegicus]
1522	30		CAA73197	2072657	HFE (HLA-H) [Mus musculus]
1337	21		Q9GL42	24418448	Hereditary hemochromatosis protein homolog
1334	21		Q9GL41	24418447	Hereditary hemochromatosis protein homolog
1333	21		NP_001...	5958366	hemochromatosis [Bos taurus]
1328	21		Q9GK20	24418446	Hereditary hemochromatosis protein homolog
1326	21		Q9GL43	24418449	Hereditary hemochromatosis protein homolog
1316	28		AAH86597	2624957	Hereditary hemochromatosis protein homolog
1303	30		CAT25847	56237864	HFE [Mus musculus]
1298	22		NP_001...	57114069	hemochromatosis protein [Pan troglodytes]
1269	21		AAG39940	11692703	HFE [Dicerops bicornis]

Basic Local Alignment Search Tool

- local, isolated, "surprising" regions of similarity
- breaks the query sequence into "words"
- word hits to database sequences become "seeds" for alignment extension

NCBI FieldGuide

Global vs Local Alignment

Seq 1

Seq 2

↓ Global alignment

Seq 1

Seq 2

↓ Local alignment

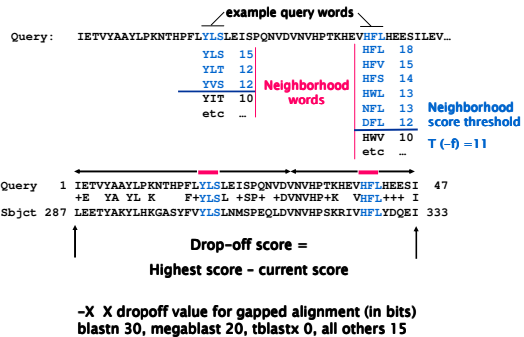
NCBI FieldGuide

How BLAST Works

1. Make lookup table of "words" for query
2. Scan database for hits
3. Extend alignment both directions
 - Ungapped extensions of hits (initial HSPs)
 - Gapped extensions (no traceback)
 - Gapped extensions (traceback - alignment details)

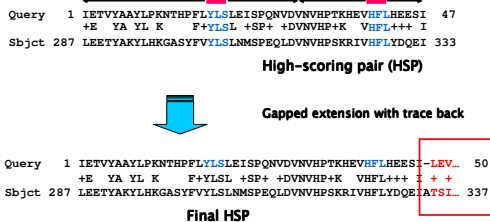
NCBI FieldGuide

BLASTP Summary



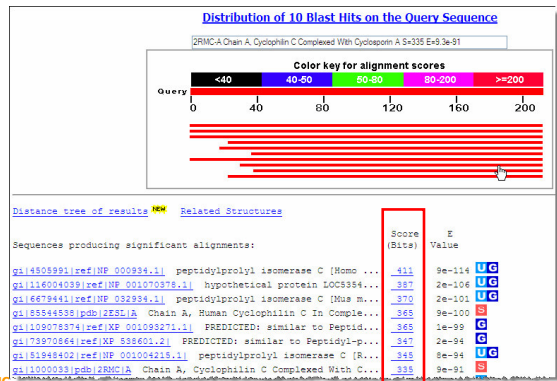
NCBI FieldGuide

BLASTP Summary (gapped extension)



NCBI FieldGuide

Scoring Systems



NCBI

Scoring Systems - Nucleotides

Identity matrix

	A	G	C	T
A	+1	-3	-3	-3
G	-3	+1	-3	-3
C	-3	-3	+1	-3
T	-3	-3	-3	+1

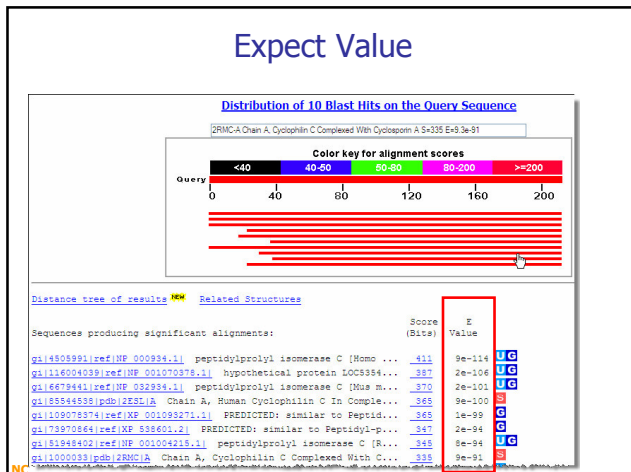
[-r 1 -q -3]

```
CAGGTAGCAAGCTTGCATGTCA
||| | | | | | | | | | | ||| | | raw score = 19 - (6*7) * = 6*
CACGTAGCAAGCTTGTGTCAC
```

* gap costs -G 5 -E 2

NCBI FieldGuide

Expect Value



Local Alignment Statistics

Expect Value

E = number of database hits you expect to find by chance, $\geq S$

$$E = Kmne^{-\lambda S} \quad \text{or} \quad E = mn2^{-S'}$$

K = scale for search space
 λ = scale for scoring system
 S' = bitscore = $(\lambda S - \ln K) / \ln 2$
 m = query length
 n = database length

E is dependent on m x n (search space)

More info: [The Statistics of Sequence Similarity Scores](#)

NCBI FieldGuide

E is dependent on m x n (search space)

Short query =

low score

high Expect

Score = 30.2 bits (15), Expect = 19
 Identities = 15/15 (100%), Gaps = 0/15 (0%)
 Strand=Plus/Plus

```
Query 4          GTAGCAAGCTTGCAT 18
                |||
Sbjct 40347983  GTAGCAAGCTTGCAT 40347997
```

NCBI FieldGuide

BLAST is a shortcut . . .

An alignment BLAST cannot make:

```
1 GAATATATGAAGACCAAGATTGCAGTCCTGCTGGCCTGAACCAACGCTATTCTTGTGTG
  |||
1 GAGTGACGATGAGCCCGAGTGTAGCAGTGAAGATCTGGACCAACGGTGTACTCGTTGTCC

61 GTTACGGAACCGAGAATGGTAAAGACTACTGGATCATTAAAGACTCCTGGGAGCCAGTT
  |||
61 GCTATGGTGTAAAGGTGGGAAGAAGTACTGGCTCGTCAAGAACAGCTGGGCTGAATCCT

121 GGGGTGAACAAGGTTATTTTCAGGCTTGCTCGTGGTAAAAAAC
  |||
121 GGGGAGACCAAGGCTACATCCTTATGTCCCGTGACAACAAC
```

Reason:

no contiguous exact match of 7 bp.

NCBI FieldGuide

An Alignment BLAST Can Make

Score = 290 bits (741), Expect = 7e-77
 Identities = 147/331 (44%), Positives = 206/331 (61%), Gaps = 8/331 (2%)
 Frame = +3

```

Query: 6  HVAIFFLLVVAISASISELGRFQAFKLERKTYLMAQAESKRFNIFTINVRAIEAH 185
      +++++ +++++ V++AS E L +++++R H KAY + EE EF IFTN I H
Sbjct: 2  LRLSVLCAIVAVTVAAASQELRFTQWEAFKTTHEKSYQSHMEELLRFKIFTENSLIAKH 61

Query: 186 NALYEQGVYKKGINKFTDMSQEEFKTHLTLASRKPLETESSYVKTQVE---IPSSVD 356
      NA Y +G USFK G      +T V -+P VD
Sbjct: 62  NAKYAKGLVSYKLG      GSTLPPANVNDSESLPKVVD 121

Query: 357 WRKEGRVTVKDGQ      SLSEQQLDCCTD-TSAGCD 533
      WRK+G VT VRDQG      SLSEQ L+DC +GC+
Sbjct: 122 WRKGAIVTVKDGQ      SLSEQNLVDCSQSFGNGCE 191

Query: 534 GSSLDDNFKYV-MK      TRVSKYTSIPAEDEDALLEA 710
      GG ++D FRK+      + Y I A E L +A
Sbjct: 182 GGLMEDAFKTIKAM      ATDTGVVETRAQSEVDLKA 241

Query: 711 VATVGPVSVGMDAS      LAVGYGTENGKDYIILKNSW 884
      VATVGP+SV +DAS      L VGYG + GK YM++KNSW
Sbjct: 242 VATVGPISVAIDASHSSFLQYSEGVYDFEPCSELDLDRVLVVGTVGKGRKTLVKNW 301

Query: 885 GASWGEQGYFLAR-GNQCGISEDTVYPTI 974
      SWG+QY +R NCCGI+ YP +
Sbjct: 302 AESWGDQGYLMSRDNQCGIASQASTPLV 332
  
```

NCBI FieldGuide

Other BLAST Algorithms

- Megablast
- Discontiguous Megablast
- PSI-BLAST
- PHI-BLAST

NCBI FieldGuide

Megablast: NCBI's Genome Annotator

- Long alignments of similar DNA sequences
- Greedy algorithm
- Concatenation of query sequences
- Faster than blastn; less sensitive

WORD SIZE	default	minimum
blastn	11	7
megablast	28	8

NCBI FieldGuide

Discontiguous Megablast

- Uses *discontiguous* word matches
- Better for *cross-species* comparisons

NCBI FieldGuide

Templates for Discontiguous Words

W = 11, t = 16, coding: 1101101101101101
W = 11, t = 16, non-coding: 1110010110110111
W = 12, t = 16, coding: 1111101101101101
W = 12, t = 16, non-coding: 1110110110110111
W = 11, t = 18, coding: 101101100101101101
W = 11, t = 18, non-coding: 111010010110010111
W = 12, t = 18, coding: 101101101101101101
W = 12, t = 18, non-coding: 111010110010110111
W = 11, t = 21, coding: 100101100101100101101
W = 11, t = 21, non-coding: 111010010100010010111
W = 12, t = 21, coding: 100101101101100101101
W = 12, t = 21, non-coding: 111010010110010010111

W = word size; # matches in template
t = template length

Reference: Ma, B, Tromp, J, Li, M. PatternHunter: faster and more sensitive homology search. Bioinformatics March, 2002; 18(3):440-5


NCBI FieldGuide

BLAST

Nucleotide <ul style="list-style-type: none"> 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 <ul style="list-style-type: none"> Protein-protein BLAST (blastp) Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST) Search for short, nearly exact matches Search the conserved domain database (rpsblast) Protein homology by domain architecture (cdart)
Translated <ul style="list-style-type: none"> Translated query vs. protein database (tblastx) Protein query vs. translated database (tblastn) Translated query vs. translated database (tblastx) 	Genomes <ul style="list-style-type: none"> Human, mouse, rat, chimp ^{new}, cow, pig, dog, sheep, cat Chicken, puffer fish, zebrafish Environmental samples Malaria Insects, nematodes, plants, fungi, microbial genomes, other eukaryotic genomes
Special	Meta

NCBI

Discontiguous (Cross-species) MegaBLAST


megablast BLAST

[Nucleotide](#) [Protein](#) [Translations](#) [Retrieve results for an RCD](#)

[What is discontiguous Mega BLAST?](#)

Load query file from disk

Set subsequence From: To:

Choose database

Return alignment endpoints only

Discontiguous Word Options

Options for advanced blasting

Limit by or select from

[Choose filter](#) Low complexity Human repeats Mask for lookup table only Mask lower case

Expect

Word Size

Percent Identity, match, mismatch scores

Discontiguous Word options

Template length Template type Require 2 word hits for extension

Other advanced

NCBI FieldGuide

Megablast in New Interface

NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

[Learn more](#) about how to use the new BLAST design (beta)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#)

- Human
- Mouse
- Rat
- Arabidopsis thaliana*
- Oryza sativa*
- Bos taurus*
- Danio rerio*
- Drosophila melanogaster*
- Gallus gallus*
- Pan troglodytes*
- Microbes*
- Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

[nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query
Algorithms: [blastn](#), [megablast](#), [discontiguous megablast](#)

[protein blast](#) Search a **protein** database using a **protein** query
Algorithms: [blastp](#), [psi-blast](#), [phi-blast](#)

NCBI FieldGuide

Megablast in New Interface

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search database `Test/gpipe/9606/allcontig_and_rna` using Megablast
(Optimize for highly similar sequences)

Show results in a new window

Algorithm parameters

NCBI FieldGuide

Disco. Megablast Example . . .

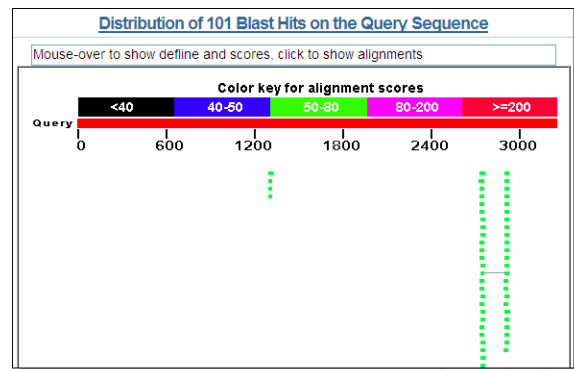
Query: NM_078651
Drosophila melanogaster CG18582-PA (mbt) mRNA, (3244 bp)
/note= mushroom bodies tiny; synonyms: Pak2, STE20, dPAK2

Database: nr (nt), Mammalia[orgn]

- > MegaBLAST = poor hits
- > Discontiguous megaBLAST = numerous hits . . .

NCBI FieldGuide

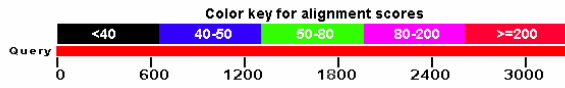
Example: MegaBLAST



Example: Discontiguous MegaBLAST

Distribution of 96 Blast Hits on the Query Sequence

NM_001014834 Homo sapiens p21(CDKN1A)-activated kinase 4 (PAK4), tra. S=365 E



Score = 365 bits (190), Expect = 2e-97
 Identities = 624/836 (74%), Gaps = 2/836 (0%)
 Strand=Plus/Plus

NCBI FieldGuide

PSI-BLAST

Position-specific Iterated BLAST

Example: N4/N6 Cytosine Methyltransferase

NCBI FieldGuide

NCBI/BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

[Learn more](#) about how to use the new BLAST design (beta)

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- Human
- Mouse
- Rat
- Arabidopsis thaliana*
- Oryza sativa*
- Bos taurus*
- Danio rerio*
- Drosophila melanogaster*
- Gallus gallus*
- Pan troglodytes*
- Microbes*
- Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontiguous megablast
- [protein blast](#) Search a **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nucleotide** query

NCBI FieldGuide

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Query subrange

Or, upload file

Job Title
Enter a descriptive title for your BLAST search

Choose Search Set

Database

Organism Any Human *A.thaliana* Mouse Custom...
Search only sequences from selected organism

Entrez Query
Optional
Enter an Entrez query to limit search

Program Selection

Algorithm blastp (protein-protein BLAST) **PSI-BLAST (Position-Specific Iterated BLAST)** PHI-BLAST (Pattern Hit Initiated BLAST)
Choose a BLAST algorithm

Search Parameter Settings:

Query Subsequence **P11409 70-280**

Database **swissprot selected**

PSI-BLAST **selected**

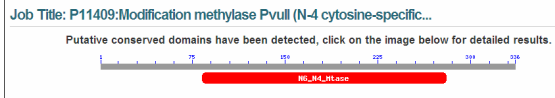
Max Hits **1000**

Expect **1**

Inclusion **0.005**

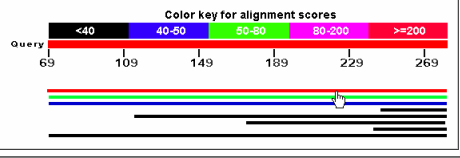
NCBI FieldGuide

First Iteration = BLASTP



Distribution of 8 Blast Hits on the Query Sequence

P11409 Modification methylase PvuII (N-4 cytosine-specific met. S=432 E=4.3e-121)



NCBI FieldGuide

First Iteration, con't

Sequences with E-value **BETTER** than threshold

Sequences producing significant alignments:	Score (Bits)	E Value
<input checked="" type="checkbox"/> sp P11409 MTF2_PROVU Modification methylase PvuII (N-4 cytosin...	432	4e-121
<input checked="" type="checkbox"/> sp P14751 MTB1_RHOSH Modification methylase RsrI (Adenine-spe...	30.6	3e-06
<input checked="" type="checkbox"/> sp Q45489 MTB2_BACSU Modification methylase BglII (N-4 cytosin...	42.0	0.001

Run PSI-Blast iteration 2

Sequences with E-value **WORSE** than threshold

<input type="checkbox"/> sp Q68556 MTB1_BACSU Modification methylase BglI (N-4 cytosin...	38.1	0.021
<input type="checkbox"/> sp Q4845 MTC1_CITFR Modification methylase CfrPI (N-4 cytosin...	37.4	0.043
<input type="checkbox"/> sp Q52692 MTS1_STRCS Modification methylase ScaI (N-4 cytosin...	33.3	0.44
<input type="checkbox"/> sp Q30569 MTS1_RHIME Modification methylase SmeIF (Adenine-sp...	33.2	0.43
<input type="checkbox"/> sp Q45971 MTC1_CAUCR Modification methylase CorMI (Adenine-sp...	32.7	0.88

Run PSI-Blast iteration 2

NCBI FieldGuide

Second Iteration

Sequences with E-value **BETTER** than threshold

Sequences producing significant alignments:	Score (Bits)	E Value
<input checked="" type="checkbox"/> sp P11409 MTF2_PROVU Modification methylase PvuII (N-4 cytosin...	358	9e-99
<input checked="" type="checkbox"/> sp Q45489 MTB2_BACSU Modification methylase BglII (N-4 cytosin...	275	1e-73
<input checked="" type="checkbox"/> sp P14751 MTB1_RHOSH Modification methylase RsrI (Adenine-spe...	273	2e-73
<input checked="" type="checkbox"/> sp Q52692 MTS1_STRCS Modification methylase ScaI (N-4 cytosin...	108	2e-23
<input checked="" type="checkbox"/> sp P29538 MTB1_HAEFA Modification methylase HpaI (Adenine-spe...	102	7e-22
<input checked="" type="checkbox"/> sp P23192 MTM2_MORBO Modification methylase MboII (Adenine-sp...	96.6	6e-20
<input checked="" type="checkbox"/> sp P28638 YHDJ_ECCLI Hypothetical adenine-specific methylase yhd	90.8	3e-18
<input checked="" type="checkbox"/> sp Q45971 MTC1_CAUCR Modification methylase CorMI (Adenine-sp...	83.5	5e-16
<input checked="" type="checkbox"/> sp Q30569 MTS1_RHIME Modification methylase SmeIF (Adenine-sp...	81.2	2e-15
<input checked="" type="checkbox"/> sp P20590 MTB1_HAEIN Modification methylase HinfI (Adenine-sp...	81.2	3e-15
<input checked="" type="checkbox"/> sp Q2YMK2 MTB1_BRUA2 Modification methylase BabI (Adenine-spe...	77.7	3e-14
<input checked="" type="checkbox"/> sp Q4845 MTC1_CITFR Modification methylase CfrBI (N-4 cytosin...	67.3	4e-11
<input checked="" type="checkbox"/> sp P30774 MTX1_XANCC Modification methylase XcyI (N-4 cytosin...	65.8	1e-10
<input checked="" type="checkbox"/> sp P14243 MTC9_CITFR Modification methylase CfrPI (N-4 cytosin...	64.3	3e-10
<input checked="" type="checkbox"/> sp P50178 MTL22_LACLC Modification methylase LlaDCRIB (Adeninn...	63.9	4e-10

Sequences with E-value **WORSE** than threshold

<input type="checkbox"/> sp P25238 MTK1_KLEPN Modification methylase RpnI (Adenine-spe...	36.9	0.055
<input type="checkbox"/> sp P29568 MTM2_METTF Modification methylase MthZII (N-4 cytosin...	33.1	0.78

Run PSI-Blast iteration 3

NCBI FieldGuide

PSI-BLAST: Comparison of 1st & 2nd Iterations

<input checked="" type="checkbox"/> New sp Q52692 MTS1_STRCS Modification methylase ScaI (N-4 cytosin...	108	2e-23 +
<input checked="" type="checkbox"/> New sp P29538 MTB1_HAEFA Modification methylase HpaI (Adenine-spe...	102	7e-22
<input checked="" type="checkbox"/> New sp P23192 MTM2_MORBO Modification methylase MboII (Adenine-sp...	96.6	6e-20 +
<input checked="" type="checkbox"/> New sp P28638 YHDJ_ECCLI Hypothetical adenine-specific methylase yhd	90.8	3e-18 +
<input checked="" type="checkbox"/> New sp Q45971 MTC1_CAUCR Modification methylase CorMI (Adenine-sp...	83.5	5e-16 +
<input checked="" type="checkbox"/> New sp Q30569 MTS1_RHIME Modification methylase SmeIF (Adenine-sp...	81.2	2e-15 +
<input checked="" type="checkbox"/> New sp P20590 MTB1_HAEIN Modification methylase HinfI (Adenine-sp...	81.2	3e-15 + Gene
<input checked="" type="checkbox"/> New sp Q2YMK2 MTB1_BRUA2 Modification methylase BabI (Adenine-spe...	77.7	3e-14 +
<input checked="" type="checkbox"/> New sp Q4845 MTC1_CITFR Modification methylase CfrBI (N-4 cytosin...	67.3	4e-11 +
<input checked="" type="checkbox"/> New sp P30774 MTX1_XANCC Modification methylase XcyI (N-4 cytosin...	65.8	1e-10 +
<input checked="" type="checkbox"/> New sp P14243 MTC9_CITFR Modification methylase CfrPI (N-4 cytosin...	64.3	3e-10 +
<input checked="" type="checkbox"/> New sp P50178 MTL22_LACLC Modification methylase LlaDCRIB (Adenin...	63.9	4e-10 +
<input checked="" type="checkbox"/> New sp P09358 MTD2_STRFM Modification methylase DpnIIB (Adenine-...	61.6	2e-09
<input checked="" type="checkbox"/> New sp P23941 MTB1_BACAM Modification methylase BamHI (N-4 cytosin...	60.0	5e-09
<input checked="" type="checkbox"/> New sp Q58893 MTM5_METJA Modification methylase MjaV (N-4 cytosin...	58.9	1e-08 +
<input checked="" type="checkbox"/> New sp P14230 MTSM_SERMA Modification methylase SmaI (N-4 cytosin...	58.1	2e-08
<input checked="" type="checkbox"/> New sp P34721 MTLB_MORBO Modifi...	44.3	1e-04
<input checked="" type="checkbox"/> New sp P43871 MTB3_HAEIN Modifi...		
<input checked="" type="checkbox"/> New sp P18051 MTB2_BACAM Modifi...		
<input checked="" type="checkbox"/> New sp Q58513 MTS1_STRPI Modification methylase SrrI (N-4 cytosin...	52.3	1e-05
<input checked="" type="checkbox"/> New sp Q58392 MTM1_METJA Modification methylase MjaI (N-4 cytosin...	50.8	4e-06 +
<input checked="" type="checkbox"/> New sp Q984X2 YUBD_ECCLI Putative methylase yubD	49.6	8e-06 +
<input checked="" type="checkbox"/> New sp Q68556 MTB1_BACSU Modification methylase BglI (N-4 cytosin...	46.9	5e-05 +
<input checked="" type="checkbox"/> New sp Q59647 MTM8_METRO Modification methylase MwoI (N-4 cytosin...	45.0	2e-04 +
<input checked="" type="checkbox"/> New sp P14827 MTFC_ENTCL Modification methylase EcaI (Adenine-spe...	43.5	6e-04
<input checked="" type="checkbox"/> New sp P71366 TSMH_HAEIN Putative type III restriction-modificati...	42.3	0.001
<input checked="" type="checkbox"/> New sp P40844 MTM10_METM10 Modifi...	41.3	0.001
<input checked="" type="checkbox"/> New sp P08		
<input checked="" type="checkbox"/> New sp Q91		
<input checked="" type="checkbox"/> sp P25076 CY11_SQLEU Cytochrome c1 heme protein, mitochondria...	31.2	2.8
<input checked="" type="checkbox"/> sp Q83DD0 SVV_COXBU Valyl-tRNA synthetase (Valine-tRNA ligase)	30.4	4.5

Red ones not found by regular blastp

These two found by regular blastp are not included

NCBI FieldGuide

Convergence

Results of PSI-Blast iteration 4

No new sequences were found above the 0.005 threshold!

Accession	Description	Score	E-value
gpi082692.MTHL_STBCE	Modification methylase Ssa1 (H-4 cytosin...	151	2e-06
gpi020590.MTHL_BAEH1	Modification methylase HinfI (Adenine-spc...	130	5e-30
gpi029318.MTHL_BAEH2	Modification methylase HpaI (Adenine-spc...	103	6e-22
gpi033130.MTHL_HOBB0	Modification methylase HsuRI (Adenine-spc...	102	1e-21
gpi050118.MTHL2_LAGL6	Modification methylase Ila4ORIS (Adenine...	94.3	2e-19
gpi003918.MTHL22_STBFB	Modification methylase DpnIIb (Adenine...	82.8	7e-19
gpi024483.MTHL22_SCOO1	Hypothetical adenine-specific methylase (M...	81.8	2e-17
gpi047474.MTHL1B_HOBB0	Modification methylase HpaII (Adenine-spc...	82.7	2e-17
gpi028883.MTHL5_METHA	Modification methylase MspV (H-4 cytosin...	75.1	1e-13
gpi030714.MTHL12_SAMP2	Modification methylase XpyI (H-4 cytosin...	73.1	6e-13
gpi048438.MTHL9_CITR1	Modification methylase CpRI (H-4 cytos...	72.7	6e-13
gpi015051.MTHL2_BACW0	Modification methylase BsaBI (H-4 cytos...	64.6	2e-10
gpi028936.MTHL1_BACQ2	Modification methylase BglI (H-4 cytosin...	63.1	6e-10
gpi044230.MTHL12_SFBW0	Modification methylase BsaI (H-4 cytosin...	62.1	7e-10
gpi023941.MTHL1_BACW0	Modification methylase BsaSI (H-4 cytos...	60.6	3e-09
gpi008641.MTHL0_METH0	Modification methylase HpaI (H-4 cytosin...	60.6	4e-09
gpi013973.MTHL1_BAEH1	Modification methylase HinfII (Adenine...	60.6	4e-09
gpi048938.MTHL12_METHA	Modification methylase HpaI (H-4 cytosin...	58.6	1e-07
gpi002613.MTHL12_SIFR1	Modification methylase SfiI (H-4 cytosin...	53.9	4e-07
gpi004342.MTHL12_BACQ2	Modification methylase BstI (H-4 cytosin...	53.2	3e-06
gpi008642.MTHL12_SCOO1	Purcative methylase yuaB	50.6	4e-04
gpi040811.MTHL0_SANTY	Type III restriction-modification system...	50.4	5e-04
gpi0087431.MTHL0_RPFL	Type III restriction-modification system...	50.0	5e-04
gpi029493.MTHL12_BAEH1	Modification methylase HsuII (H-4 cytos...	50.0	2e-05
gpi0124461.MTHL0_BACR1	Purcative type III restriction-modificati...	52.7	3e-05
gpi014847.MTHL2_METHC	Modification methylase EcoI (Adenine-spc...	51.2	3e-04

NCBI FieldGuide

Query Specific PSSM vs BLOSUM62

Last position-specific scoring matrix computed, weighted observed perc

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
265 P	-1	0	-1	-1	-5	-1	3	0	-3	-4	-4	2	-2	-3	5	0	-2	-5	-3	-3
266 D	-2	-3	2	2	0	-1	-1	5	-2	-3	-5	-2	-5	-4	-1	-1	-2	-5	-3	-2
267 D	-1	-1	-1	5	-1	0	2	-1	0	-4	-2	3	1	-5	-2	0	-1	-6	-5	-4
268 L	-3	4	-1	-1	-2	-1	-1	-5	1	0	2	-2	-3	-4	0	2	-2	-1	-1	-1
269 V	-1	-5	-6	-6	-1	-5	-5	-5	-6	3	1	-5	1	-1	-5	-4	-3	-5	-1	6
270 V	-1	-5	-6	-6	-2	-5	-6	-6	-5	2	-5	-5	0	0	-6	-4	-4	-2	-1	1
271 D	-5	-4	-1	8	-6	-3	3	-4	-6	-6	-3	-6	-6	-4	-3	-4	-7	-6	-3	1
272 I	0	-5	-6	-6	-1	-5	-5	-6	-5	2	-3	-5	1	0	2	-4	-3	-2	-1	4
273 F	-1	-5	-4	-5	2	-5	-5	-5	-2	-5	-4	-5	-1	-3	-6	-4	-3	-3	-3	-5
274 G	0	-5	-4	-5	9	-4	-5	-2	-5	-3	-3	-5	-1	-3	-1	0	2	-5	-5	-2
275 G	-2	-5	-3	-4	-5	-4	-4	7	-4	-6	-6	-3	-5	-5	-4	-2	-3	-4	-5	-5
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	6															
Q	-1	0	0	-3	3	6														
E	-1	0	0	2	-4	2	5	6												
G	-2	0	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
F	-1	-2	-1	-3	-1	-2	-2	-3	-3	-1	-2	-4	6							
S	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	6						
T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-2	-2	-1	6					
W	-3	-4	-4	-2	-2	-2	-2	-2	-3	-2	-3	-1	1	-1	-1	6				
Y	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-1	-1	-1	6			
V	0	-3	-3	-1	-2	-2	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	6		
X	0	-1	-1	-2	-3	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	6	

BLOSUM62

	270	280
query	268	LVVDIHGSGNNTTGLV
lg60_A	215	LVLDCHMGSGTTAIV
gi 3915983	217	IVLDPFAGSFTTGAIV
gi 1769991	212	IILDPMGSGTTGVA
gi 2313123	190	IVLDPFMGSGTTGLA

G274, G275

NCBI FieldGuide

BLAST

Basic Local Alignment and Search Tool

Save your searches

Choose a BLAST program to run:

- nucleotide_blast** Search a nucleotide database using a nucleotide query
- protein_blast** Search a protein database using a protein query
- tblastx** Search translated nucleotide database using a translated nucleotide query

What's New?

BLAST 2.14 now available

BLAST 2.14 now available

NCBI FieldGuide

Nucleotide BLAST Databases

- nr (nt)**
 - Traditional GenBank Divisions
 - NM_ and XM_ RefSeqs
- refseq_rna**
 - NM_, XM_, NR_
- refseq_genomic**
 - NC_, NT_, NG_
- est**
 - EST Division
- htgs**
 - HTG division
- dbsts**
 - STS Division
- chromosome**
 - NC genomic records
- gss**
 - GSS division
- wgs**
 - wgs entries from traditional divisions
- pdb**
 - Nucleotide sequences from structures
- env_nt**
 - environmental samples

combined, for human and mouse

NCBI FieldGuide

Protein BLAST Databases

Protein

- nr
traditional GenBank records
- refseq = NP_, XP_
- swissprot
- pdb
- pat
- env_nr

nr ≠ nr

NCBI

New Nucleotide Databases

NCBI *nucleotide-nucleotide* **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

nr
Genomic plus Transcript
Human genomic plus transcript
Mouse genomic plus transcript

Other Databases

- refseq_ma
- refseq_genomic
- est
- est_human
- est_mouse
- est_others
- gss
- htgs
- pat
- pdb
- month
- alu_repeats

Now: **BLAST!** or **FASTA** **FASTX** **FASTY**

NCBI FieldGuide

nr Two new **Human** and **Mouse** databases combine genomic plus transcript alignments in a single report. You can also choose from **Others** to use nr or an existing database.

New Formatter

Format

Show Graphical Overview Linkout Sequence Retrieval NCBI-qi Alignment in HTML format

Default(X for protein, n for nucleotide) Masking Black

Default(X for protein, n for nucleotide) Lower Case

Select lower case

Select red

Alignment view

Format for PSI-BLAST with inclusion threshold: 0.05

Limit results by: or select from: All organisms

Expect value range:

Layout: Two Windows Formatting options on page with results: None

Autoformat: Semi-auto

NCBI

BLAST Output: Alignments & Filter

```
>gi1466462.gb|AA17374.1| human homolog of E. coli mutI gene product, Swiss-Prot
Number: P23367
Length=756
Score = 219.935 bits (559), Expect = 8.71148e-57
Identities = 120/131 (91%), Positives = 125/131 (95%), Gaps = 2/131 (1%)
Query 1 IETVYAAYLEKNTHEFFLYLSLEISFQNVVDVNHPTKHEVHFLHEESILE-VQGHIESKLL 59
Sbjct 276 IETVYAAYLEKNTHEFFLYLSLEISFQNVVDVNHPTKHEVHFLHEESILE-VQGHIESKLL 335
Query 60 GSNSSRMYFTQTLPLPLAGSGEMVK-sttltssstgsgDKVYAHQMVRTDSREQK LDA 118
Sbjct 336 GSNSSRMYFTQTLPLPLAGSGEMVK +T+ +SS++ SDKVYAHQMVRTDSREQK LDA 395
Query 119 FLQPLSKELSS 129 low complexity sequence filtered
Sbjct 396 FLQPLSKELSS 406
```

NCBI FieldGuide

New Output View

Format

Show Graphical Overview Linkout Sequence Retrieval NCBI-g Alignment In HTML format

CDS feature

New View ←

Masking Character: Lower Case Masking Color: Red

Number of: Descriptions 100 Alignments 50 Graphic overview 100

Alignment view: Pairwise

Limit results by: AND All organisms

NCBI FieldGuide

New Output View

Legend filters: UniGene GEO Map

Results can be sorted

Transcript & genomic hits separated

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell divisio	2876	2876	95%	0.0	97%	U G M
XM_94493.2	PREDICTED: Homo sapiens chr	40.1	40.1	1%	8.2	100%	G M
XM_93301.2	PREDICTED: Homo sapiens chr	40.1	40.1	1%	8.2	100%	G M
Genomic sequences [show first]							
NT_023935.17	Homo sapiens chromosome 9 ge	2629	2673	94%	0.0	100%	
NW_024484.1	Homo sapiens chromosome 9 ge	2601	2645	94%	0.0	100%	
NT_032977.8	Homo sapiens chromosome 1 ge	428	3002	95%	9e-117	100%	
NW_021351.1	Homo sapiens chromosome 1 ge	428	3010	95%	9e-117	100%	
NT_009237.17	Homo sapiens chromosome 11 g	44.1	44.1	1%	0.53	100%	
Ad							

Pseudogene, chr 9

Functional gene, chr 1

NCBI FieldGuide

Sorting Results

Resorted by Total score

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell divisio	2876	2876	95%	0.0	97%	U G M
XM_933013.2	PREDICTED: Homo sapiens chr	40.1	40.1	1%	8.2	100%	G M
XM_944935.2	PREDICTED: Homo sapiens chr	40.1	40.1	1%	8.2	100%	G M
Genomic sequences [show first]							
NW_021351.1	Homo sapiens chromosome 1 ge	428	3010	95%	9e-117	100%	
NT_032977.8	Homo sapiens chromosome 1 ge	428	3002	95%	9e-117	100%	
NT_023935.17	Homo sapiens chromosome 9 ge	2629	2673	94%	0.0	100%	
NW_024484.1	Homo sapiens chromosome 9 ge	2601	2645	94%	0.0	100%	
NW_021807.1	Homo sapiens chromosome 3 ge	44.1	44.1	1%	0.53	100%	

Functional gene now first

NCBI FieldGuide

Sorting Hits: by Score

ref[NT_032977.8]Me1 33153 Homo sapiens chromosome 1 genomic contig, reference assembly
Length=7889525

Sort alignments for this subject sequence by:
[E value](#) [Score](#) [Percent identity](#)
[Query start position](#) [Subject start position](#)

Features in this part of subject sequence:
[cell division cycle 21](#)

Score = 428 bits (216), Expect = 9e-117
Identities = 231/236 (97%), Gaps = 0/236 (0%)
Strand=Plus/Plus

Query 965 CTCACAGTGGTTCAGGTTCTGAGCCACATCCACACCATGATGTGCGGGTACGAGAACCA 1024
II
Sbjct 13798816 CTCACAGTGGTTCAGGTTCTGAGCCACATCCACACCATGATGTGCGGGTACGAGAACCA 13798875

Query 1025 TGTGGCTACACTGAGTGGCCACAGCAGAAATGTGTGGCTGGCCCTGGGCCCCAGATGG 1084
II

Sort by Score: longest exon usually first

NCBI FieldGuide

Sorting Hits: by Query Start

ref[NT_032977.6]Hs1_33153 Homo sapiens chromosome 1 genomic contig, reference assembly
Length=73835825

Sort alignments for this subject sequence by:
E Value Score Percent Identity
Query start position Subject start position

Features flanking this part of subject sequence:
619 bp at 3' side: m1e1op1lifetative leukemia virus oncogene
223 bp at 3' side: cell division cycle 20

Score = 89.7 bits (45), Expect = 1e-14
Identities = 51/53 (96%), Gaps = 0/53 (0%)
Strand=Plus/Plus

Query 1
ACCGGAGAGTTTAAAGAGCGCTAAGCAGGCGTGTAAAGCCGCTCGAACTGC 53
|||||
Sbjct 13796530 ACCGGAGAGTTTAAAGAGCGCTAAGCAGGCGTGTAAAGCCGCTCGAACTGC 13796582

Sort by Query start:
Proper exon order

NCBI FieldGuide

Advanced Options

Options for advanced blasting

Limit by filter NOT ma or select from: All organisms
Example Entrez query: Limit to Organism

Composition-based statistics: all ch1 src

Choose filter: Low complexity Nucleotide OtherAdvanced

Expect: 10

Word Size: 3

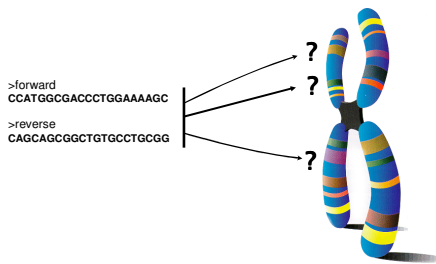
Matrix: BLOSUM62

PSSM:

Other advanced: descriptions alignments

NCBI FieldGuide

Example: Mapping Oligos Onto a Genome



Map Oligos Onto Genome

NCBI Home - Genomic Biology - Human Genome Guide

Search [LocusLink] for [] Go

BLAST

overview
FAQs
news
manual
references

Blast the Human Genome

Blast your sequence against Human specific sequences

Database: [genome] Program: [blastn]

Use MegaBLAST

Enter an accession, gi, or a sequence in FASTA format:

CCATGGCGACCCCTGGAAAAGC
forward primer reverse primer

Optional parameters

Expect: 1 Filter: default Descriptions: 100 Alignments: 100

0.0001
0.001
0.01
d options: [-W 7 -e 1000]

10 search Clear Input

NCBI FieldGuide

Genome BLAST Results

RID: 1076295772-31414-177358914251.BLASTQ3

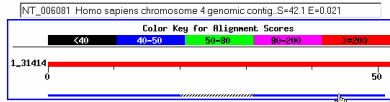
Query= CCATGGCGACCCCTGGAAAAGCANNNNNNNNNCAGCAGCGGCTGTGCCTGCGG
(52 letters)

Database: contig
498 sequences; 3,020,300,271 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Genome View](#) Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

Distribution of 2 Blast Hits on the Query Sequence



NCBI FieldGuide

Primer Alignments

>ref|NT_006081.16|Hs4_6238 Homo sapiens chromosome 4 genomic contig
Length = 1182262

Score = 42.1 bits (21), Expect = 0.021
Identities = 21/21 (100%)
Strand = Plus / Minus

Query: 32 cagcagcgctgtgctgctgcg 52
Sbjct: 463315 cagcagcgctgtgctgctgcg 463295

reverse primer

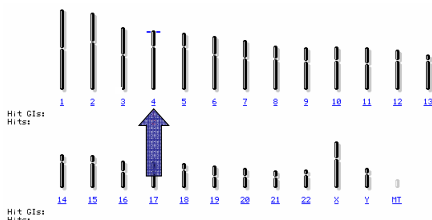
Score = 42.1 bits (21), Expect = 0.021
Identities = 21/21 (100%)
Strand = Plus / Plus

Query: 1 ccattggcgacctggaaaagc 21
Sbjct: 463128 ccattggcgacctggaaaagc 463148

forward primer

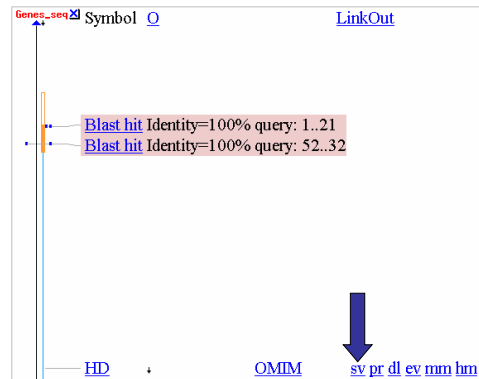
NCBI FieldGuide

MapViewer



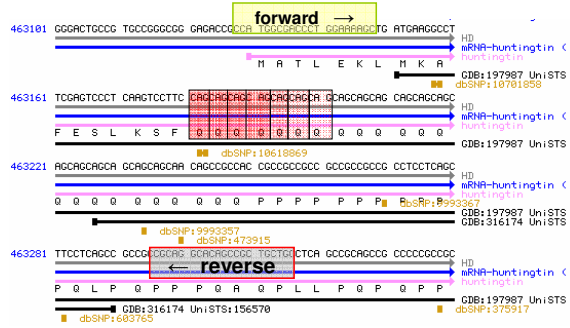
NCBI FieldGuide

MapViewer



NCBI FieldGuide

Sequence View (sv)



NCBI FieldGuide

Service Addresses

- **BLAST** blast-help@ncbi.nlm.nih.gov
- **General Help** info@ncbi.nlm.nih.gov
- **Wayne Matten** matten@ncbi.nlm.nih.gov

NCBI FieldGuide