GenBank

GenBank® is the NIH nucleotide sequence database. It is part of the International Nucleotide Sequence Database Collaboration, with the DNA DataBank of Japan (DDBJ) and the European Molecular Biology Laboratory (EMBL). These three organizations exchange data on a daily basis that now includes sequences from about 407,000 organisms.

The exercises cover following topics:

* History and growth of GenBank
* Different types of entries and how to identify them from the accession numbers
* The format of a typical GenBank entry and understanding the annotated features
* The submission tools, BankIt and Sequin, and choosing which one to use
* Overview of the processing of GenBank entries
* Overview of the GenBank FTP directory: how to download sequences in batch
* BLAST databases and their equivalents in GenBank
* Some tips on searching GenBank efficiently

The following handout includes the screen shots of the first exercise as well as background slides**.**

Exercise 1

1. Retrieve entries containing the word "pannexin" from the nucleotide database. The Entrez Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, and the Protein Data Bank. The summary of the top 20 entries is displayed by default. Currently, how many entries are there in the CoreNucleotide, EST and GSS databases associated with pannexin?

Change the pull down menu "Show" to a number large enough to view the summaries of all entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the Accession Number Prefixes page (http://www.ncbi.nlm.nih.gov/Sequin/acc.html), if necessary.

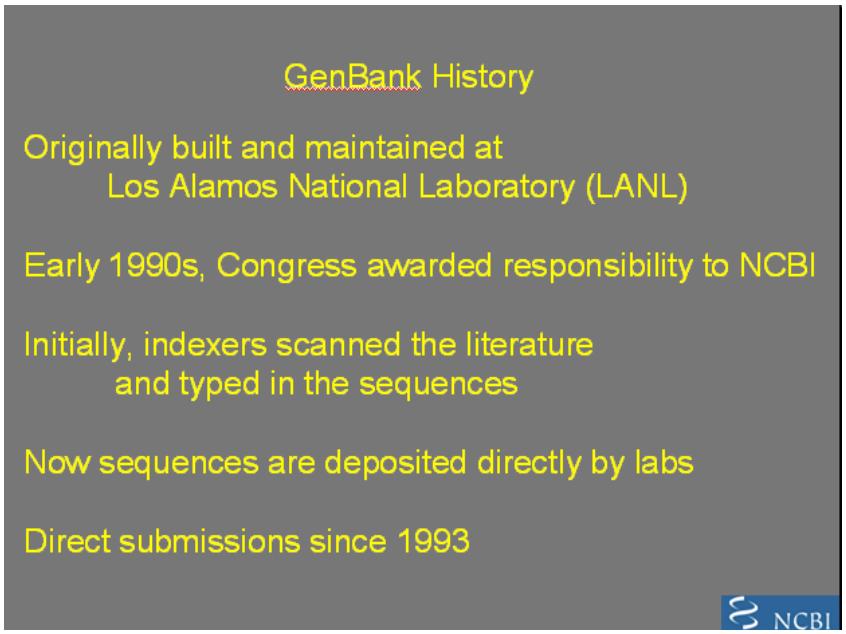
Access entries from the Nucleotide database. Keep the genomic entries from GenBank by using the Preview/Index page (biomol genomic [properties] AND srcdb genbank [properties]). Is GenBank a redundant database? Note multiple pannexin 1 segmented sets and individual exon sequences.

Let us use the entry AY048509 as a sample GenBank entry with feature annotations. When the entry was last updated? Was its sequence updated after

release? Does it contain a complete coding region? What are the locations of the 5'UTR? What is the accession number for the protein entry?

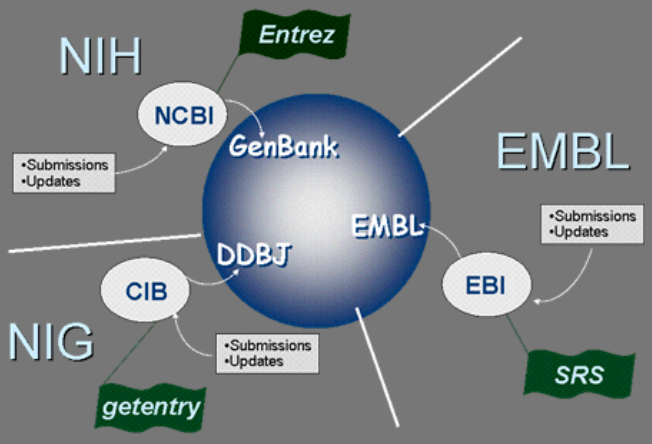
Go back to the Entrez report. Change the source database to RefSeq, molecule type to mRNA and human as organism by using the Preview/index page. Now, RefSeq entries for human pannexin 1, pannexin 2 and pannexin 3 mRNAs are displayed. Save the sequences to a file by changing the "Display" to FASTA and "Send to" file.

Check the sequence revision history of NM_015368. What is the most recent change? How many times has the sequence of the entry been updated? When did these updates occur?



International Nucleotide Sequence Database Collaboration

http://www.ncbi.nlm.nih.gov/Genbank/index.html



NCBI

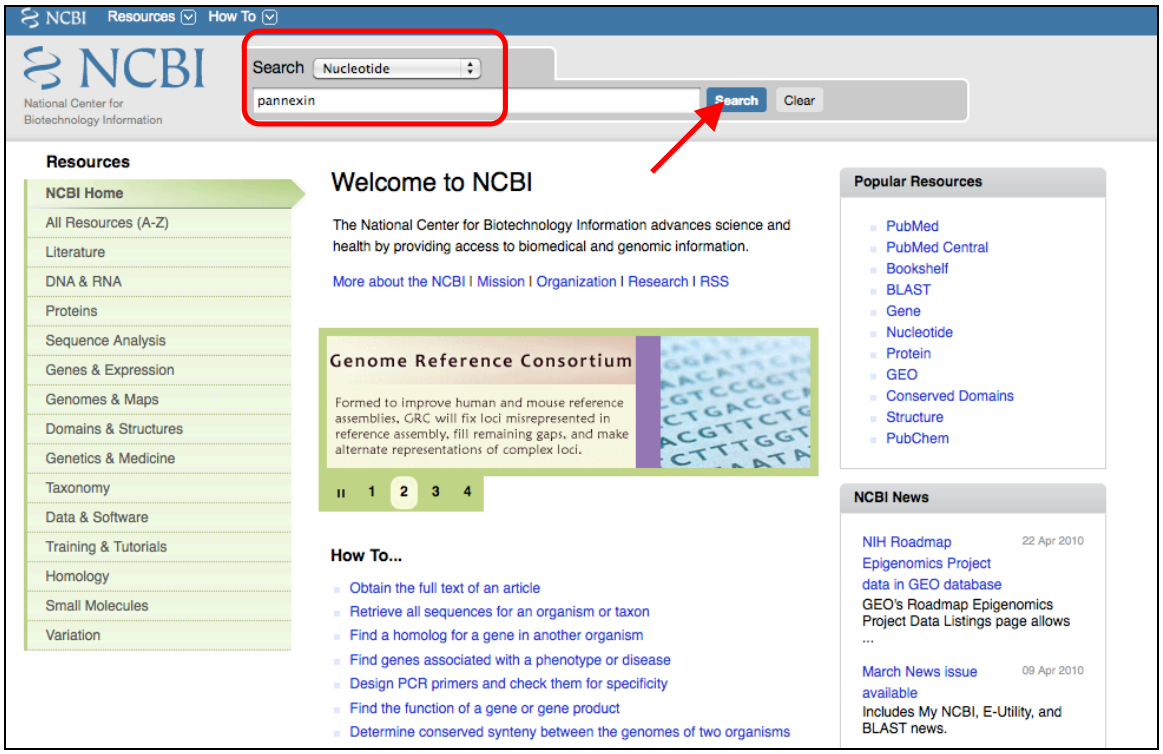
GenBank	RefSeq
Archival/repository	Curated
Redundant	Non-redundant
Submitter owner	NCBI owner
Sequenced	Combined/edited

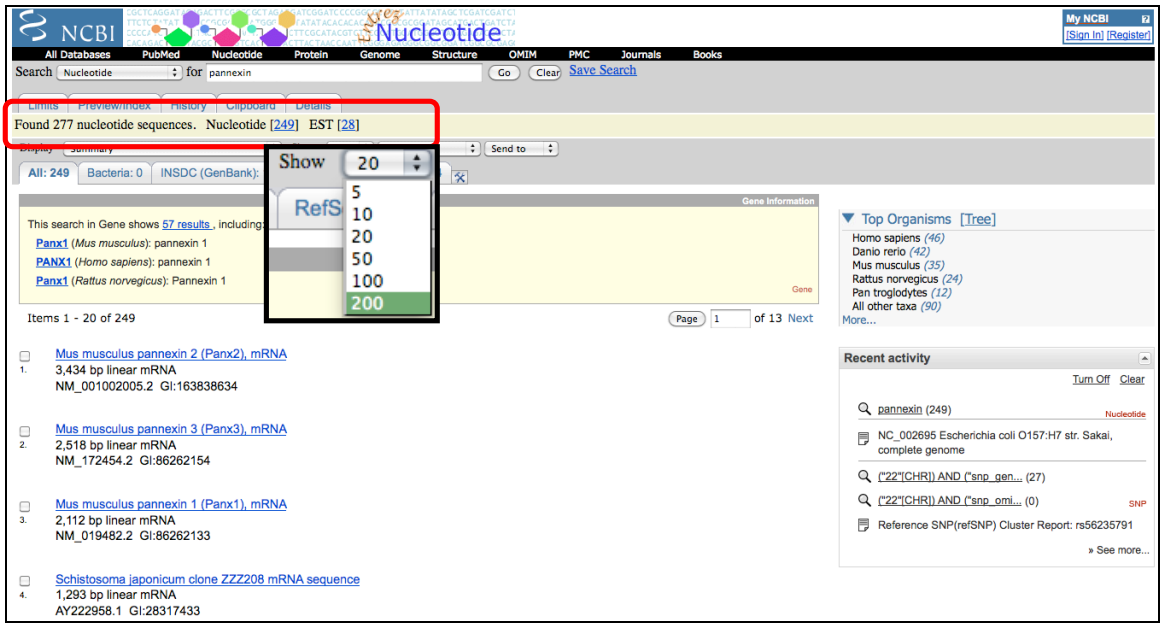
NCBI

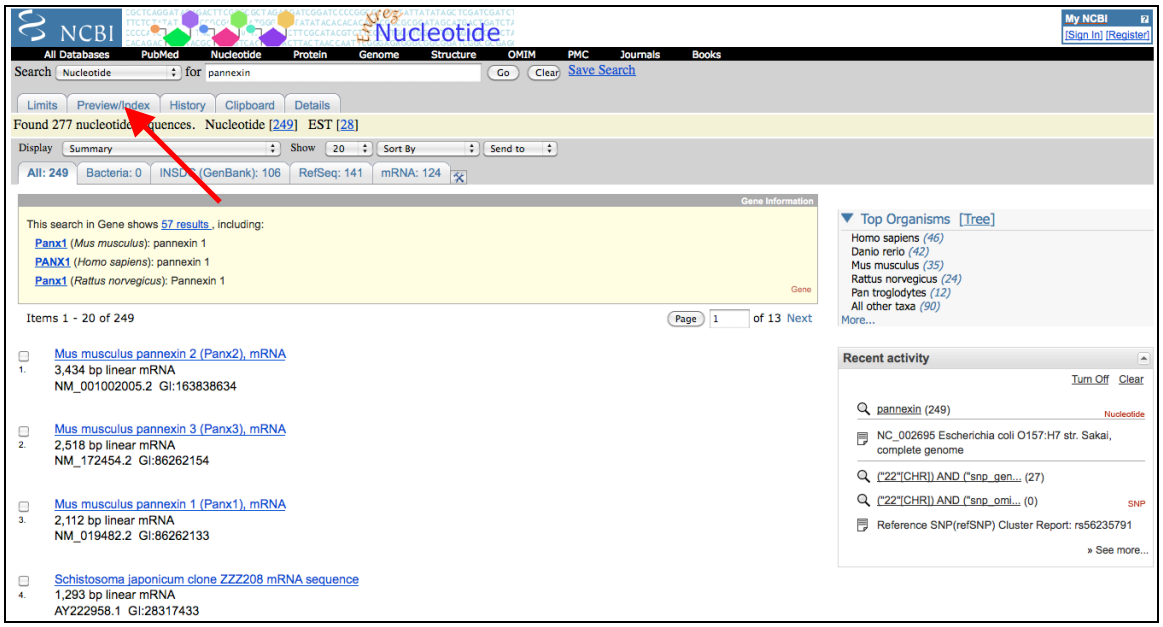
Types of Entries

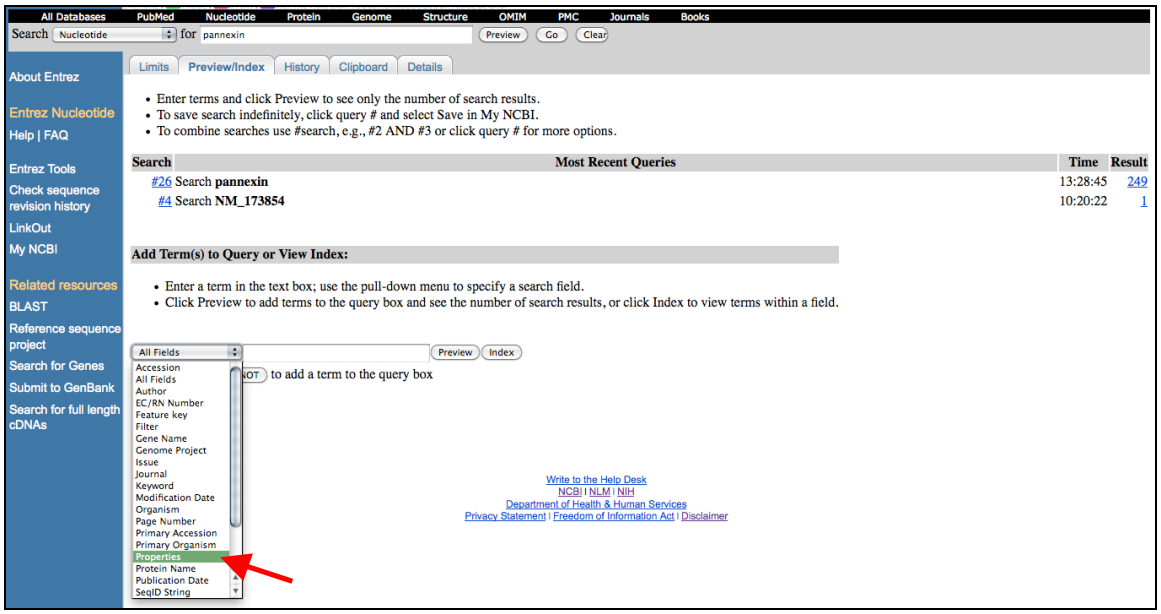
1. Individual mRNA/Genomic
2. Sets such as Pop, Phy, Mut and environmental
3. Segmented sets
4. Expressed Sequence Tags (EST)
5. Genome Survey Sequence (GSS)
6. Sequence Tagged Site (STS)
7. Whole Genome Shotgun (WGS)
8. High Throughput Genomic (HTG)
9. High Throughput cDNA (HTC)
10. Full-Length Insert cDNA (FLIC)
11. Complete genomes
12. Third Party Annotation (TPA)

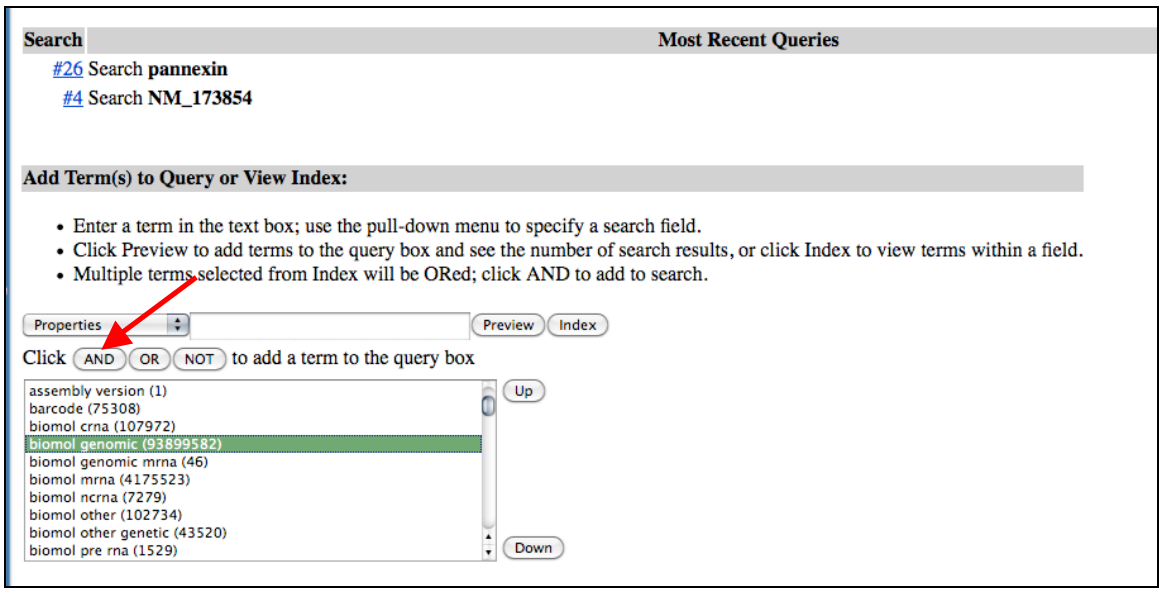
NCBI



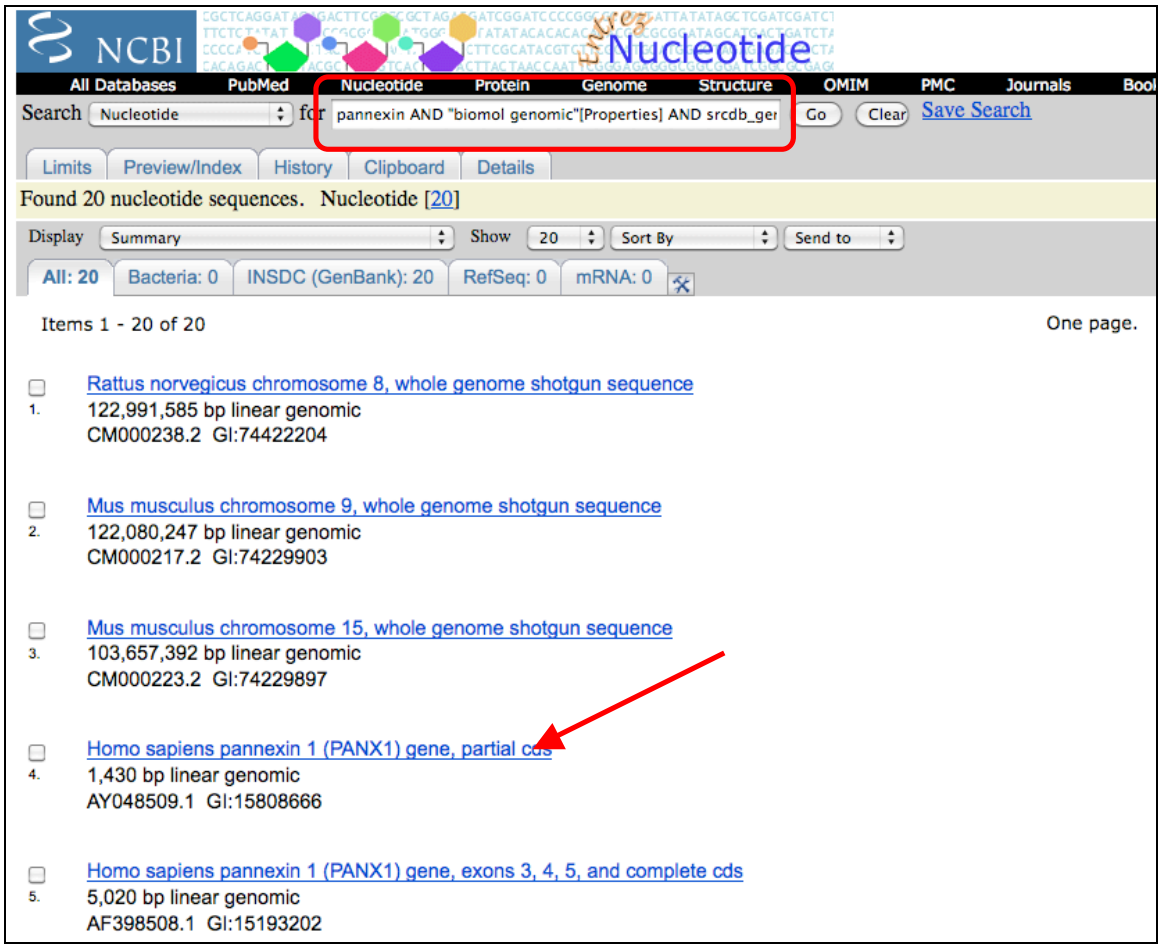

NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for pannexin Go Clear Save Search
Limits Preview/Index History Clipboard Details
Found 277 nucleotide sequences. Nucleotide [249] EST [28]
Show 20
5
10
20
50
100
200
Send to
All: 249 Bacteria: 0 INSDC (GenBank):
Gene Information
This search in Gene shows 57 results , including:
Panx1 (Mus musculus): pannexin 1
PANX1 (Homo sapiens): pannexin 1
Panx1 (Rattus norvegicus): Pannexin 1
Gene
Items 1 - 20 of 249
Page 1 of 13 Next
Top Organisms [Tree]
Homo sapiens (46)
Danio rerio (42)
Mus musculus (35)
Rattus norvegicus (24)
Pan troglodytes (12)
All other taxa (90)
More...
1. Mus musculus pannexin 2 (Panx2), mRNA
3,434 bp linear mRNA
NM_001002005.2 GI:163838634
2. Mus musculus pannexin 3 (Panx3), mRNA
2,518 bp linear mRNA
NM_172454.2 GI:86262154
3. Mus musculus pannexin 1 (Panx1), mRNA
2,112 bp linear mRNA
NM_019482.2 GI:86262133
4. Schistosoma japonicum clone ZZZ208 mRNA sequence
1,293 bp linear mRNA
AY222958.1 GI:28317433
Recent activity
Turn Off Clear
pannexin (249) Nucleotide
NC_002695 Escherichia coli O157:H7 str. Sakai, complete genome
("22"[CHR]) AND ("snp_gen... (27)
("22"[CHR]) AND ("snp_omi... (0) SNP
Reference SNP(refSNP) Cluster Report: rs56235791
» See more...

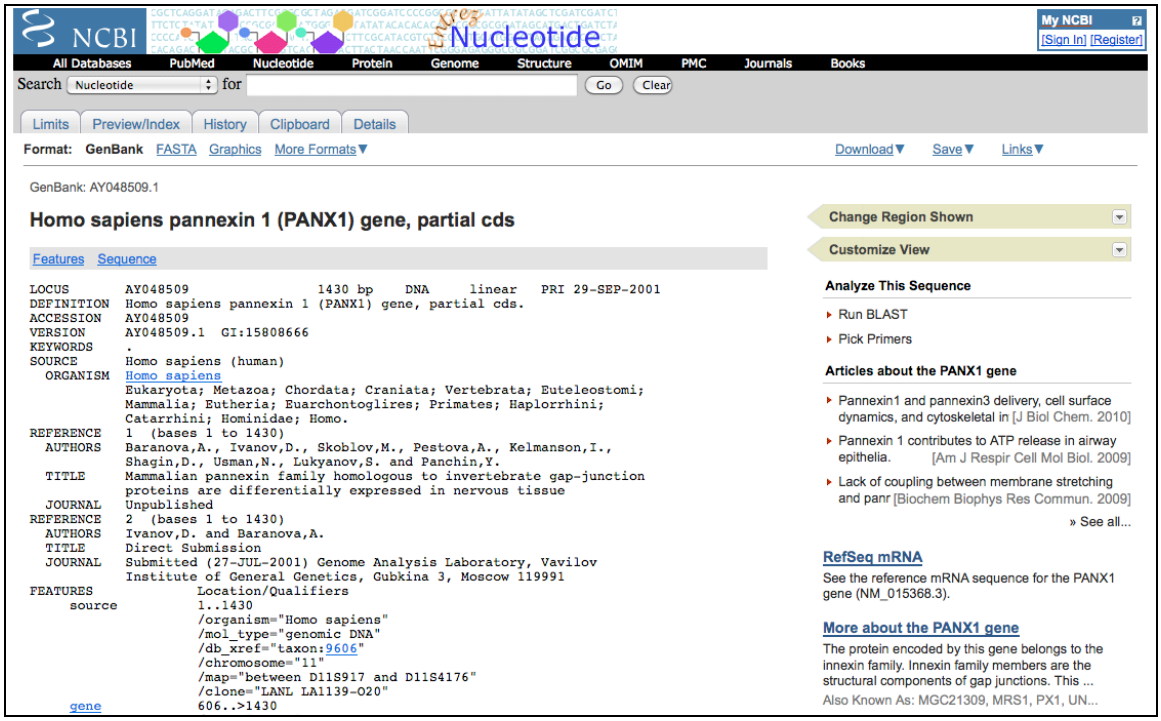

NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for pannexin Go Clear Save Search
Limits Preview/Index History Clipboard Details
Found 277 nucleotide sequences. Nucleotide [249] EST [28]
Display Summary Show 20 Sort By Send to
All: 249 Bacteria: 0 INSDC (GenBank): 106 RefSeq: 141 mRNA: 124
Gene Information
This search in Gene shows 57 results , including:
Panx1 (Mus musculus): pannexin 1
PANX1 (Homo sapiens): pannexin 1
Panx1 (Rattus norvegicus): Pannexin 1
Gene
Items 1 - 20 of 249
Page 1 of 13 Next
Top Organisms [Tree]
Homo sapiens (46)
Danio rerio (42)
Mus musculus (35)
Rattus norvegicus (24)
Pan troglodytes (12)
All other taxa (90)
More...
1. Mus musculus pannexin 2 (Panx2), mRNA
3,434 bp linear mRNA
NM_001002005.2 GI:163838634
2. Mus musculus pannexin 3 (Panx3), mRNA
2,518 bp linear mRNA
NM_172454.2 GI:86262154
3. Mus musculus pannexin 1 (Panx1), mRNA
2,112 bp linear mRNA
NM_019482.2 GI:86262133
4. Schistosoma japonicum clone ZZZ208 mRNA sequence
1,293 bp linear mRNA
AY222958.1 GI:28317433
Recent activity
Turn Off Clear
pannexin (249) Nucleotide
NC_002695 Escherichia coli O157:H7 str. Sakai, complete genome
("22"[CHR]) AND ("snp_gen... (27)
("22"[CHR]) AND ("snp_omi... (0) SNP
Reference SNP(refSNP) Cluster Report: rs56235791
» See more...

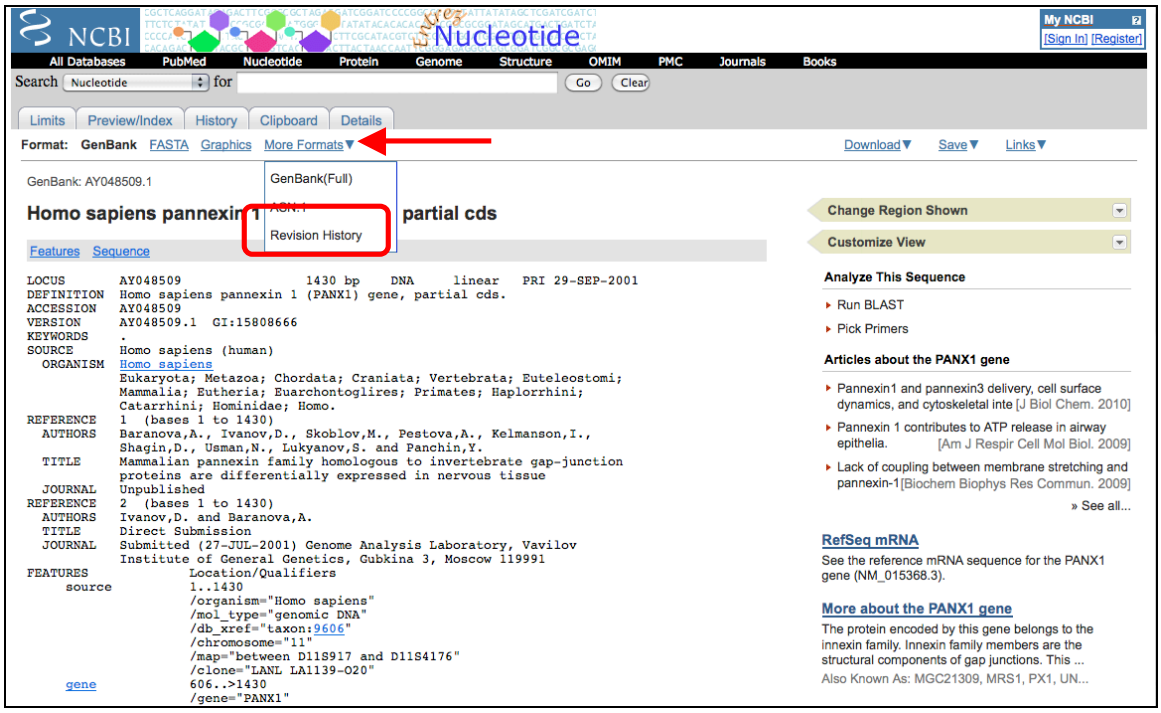

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for pannexin
Preview Go Clear
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Limits Preview/Index History Clipboard Details
• Enter terms and click Preview to see only the number of search results.
• To save search indefinitely, click query # and select Save in My NCBI.
• To combine searches use #search, e.g., #2 AND #3 or click query # for more options.
Search Most Recent Queries Time Result
#26 Search pannexin 13:28:45 249
#4 Search NM_173854 10:20:22 1
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.
All Fields Preview Index
Accession
All Fields
Author
EC/RN Number
Feature key
Filter
Gene Name
Genome Project
Issue
Journal
Keyword
Modification Date
Organism
Page Number
Primary Accession
Primary Organism
Properties
Protein Name
Publication Date
SeqID String
to add a term to the query box
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer

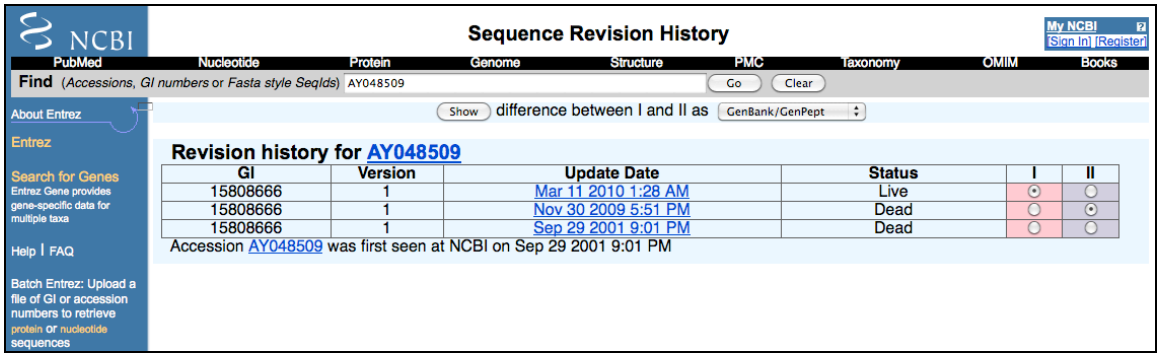

Search Most Recent Queries
#26 Search pannexin
#4 Search NM_173854
Add Term(s) to Query or View Index:
• Enter a term in the text box; use the pull-down menu to specify a search field.
• Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
• Multiple terms selected from Index will be ORed; click AND to add to search.
Properties Preview Index
Click AND OR NOT to add a term to the query box
assembly version (1)
barcode (75308)
biomol crna (107972)
biomol genomic (93899582)
biomol genomic mrna (46)
biomol mrna (4175523)
biomol ncrna (7279)
biomol other (102734)
biomol other genetic (43520)
biomol pre rna (1529)
Up
Down


NCBI
Entrez Nucleotide
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals
Search Nucleotide for pannexin AND "biomol genomic"[Properties] Go Clear Save Search
Limits Preview/Index History Clipboard Details
Found 119 nucleotide sequences. Nucleotide [119]
Display Summary Show 20 Sort By Send to
All: 119 Bacteria: 0 INSDC (GenBank): 38 RefSeq: 81 mRNA: 0
Items 1 - 20 of 119
Page 1 of 6 Next
1. Canis familiaris chromosome 21, whole genome shotgun sequence
54,024,781 bp linear genomic
NC_006603.2 GI:74033520
2. Canis familiaris chromosome 10, whole genome shotgun sequence
72,488,556 bp linear genomic
NC_006592.2 GI:74029679
3. Canis familiaris chromosome 5, whole genome shotgun sequence
91,976,430 bp linear genomic
NC_006587.2 GI:74027287
4. Taeniopygia guttata chromosome 1A, reference assembly (based on Taeniopygia_guttata-3.2.4), whole genome shotgun sequence
73,657,157 bp linear genomic
NC_011463.1 GI:224381677

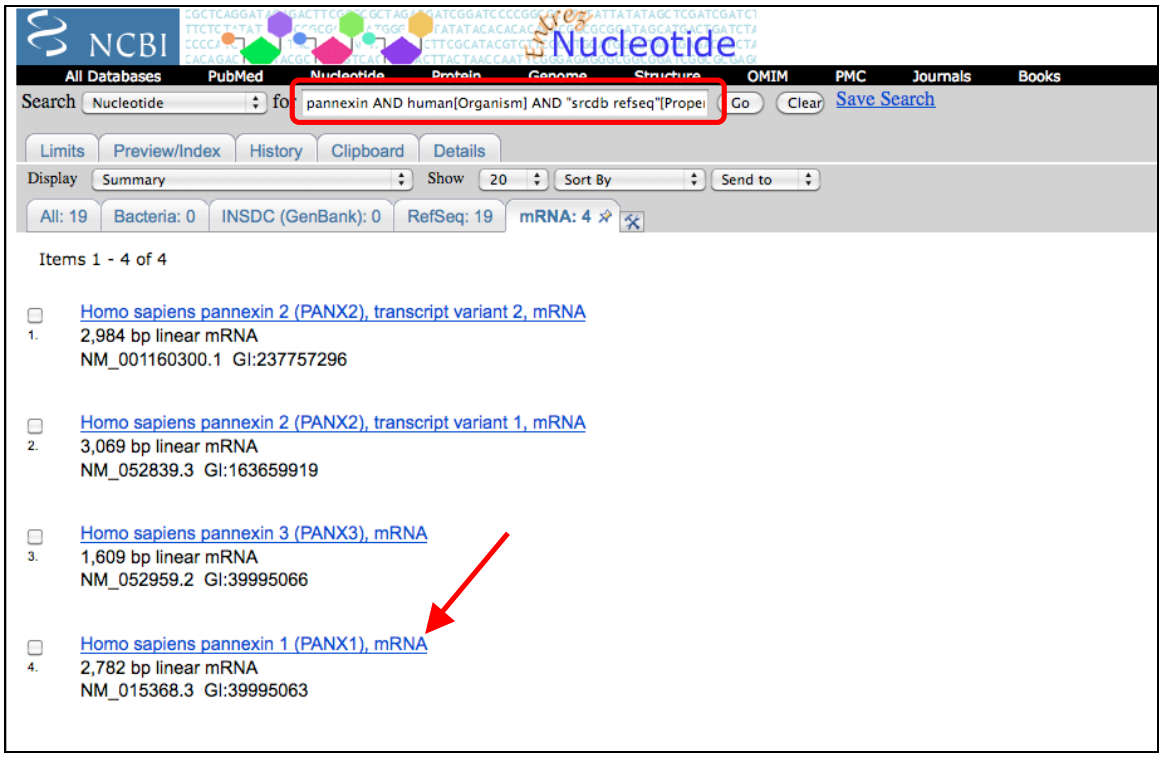

NCBI
Nucleotide
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals
Search Nucleotide for pannexin AND "biomol genomic"[Properties] AND srcdb_ge
Go Clear Save Search
Limits Preview/Index History Clipboard Details
Found 20 nucleotide sequences. Nucleotide [20]
Display Summary Show 20 Sort By Send to
All: 20 Bacteria: 0 INSDC (GenBank): 20 RefSeq: 0 mRNA: 0
Items 1 - 20 of 20
One page.
1. Rattus norvegicus chromosome 8, whole genome shotgun sequence
122,991,585 bp linear genomic
CM000238.2 GI:74422204
2. Mus musculus chromosome 9, whole genome shotgun sequence
122,080,247 bp linear genomic
CM000217.2 GI:74229903
3. Mus musculus chromosome 15, whole genome shotgun sequence
103,657,392 bp linear genomic
CM000223.2 GI:74229897
4. Homo sapiens pannexin 1 (PANX1) gene, partial cds
1,430 bp linear genomic
AY048509.1 GI:15808666
5. Homo sapiens pannexin 1 (PANX1) gene, exons 3, 4, 5, and complete cds
5,020 bp linear genomic
AF398508.1 GI:15193202

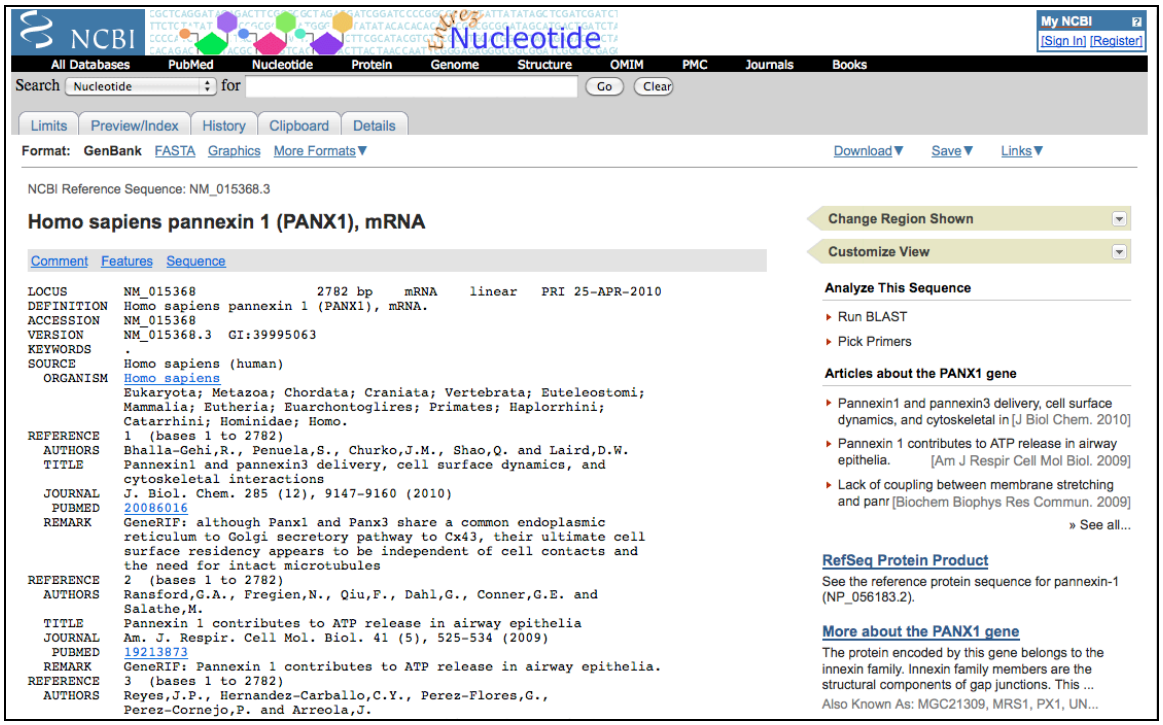

NCBI
Nucleotide
My NCBI [Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Format: GenBank FASTA Graphics More Formats
Download Save Links
GenBank: AY048509.1
Homo sapiens pannexin 1 (PANX1) gene, partial cds
Features Sequence
LOCUS AY048509 1430 bp DNA linear PRI 29-SEP-2001
DEFINITION Homo sapiens pannexin 1 (PANX1) gene, partial cds.
ACCESSION AY048509
VERSION AY048509.1 GI:15808666
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1430)
AUTHORS Baranova,A., Ivanov,D., Skoblov,M., Pestova,A., Kelmanson,I.,
Shagin,D., Usman,N., Lukyanov,S. and Panchin,Y.
TITLE Mammalian pannexin family homologous to invertebrate gap-junction
proteins are differentially expressed in nervous tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1430)
AUTHORS Ivanov,D. and Baranova,A.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov
Institute of General Genetics, Gubkina 3, Moscow 119991
FEATURES Location/Qualifiers
source 1..1430
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="between D11S917 and D11S4176"
/clone="LANL LA1139-O20"
gene 606..>1430
Change Region Shown
Customize View
Analyze This Sequence
Run BLAST
Pick Primers
Articles about the PANX1 gene
Pannexin1 and pannexin3 delivery, cell surface dynamics, and cytoskeletal in [J Biol Chem. 2010]
Pannexin 1 contributes to ATP release in airway epithelia. [Am J Respir Cell Mol Biol. 2009]
Lack of coupling between membrane stretching and panr [Biochem Biophys Res Commun. 2009]
» See all...
RefSeq mRNA
See the reference mRNA sequence for the PANX1 gene (NM_015368.3).
More about the PANX1 gene
The protein encoded by this gene belongs to the innexin family. Innexin family members are the structural components of gap junctions. This ...
Also Known As: MGC21309, MRS1, PX1, UN...

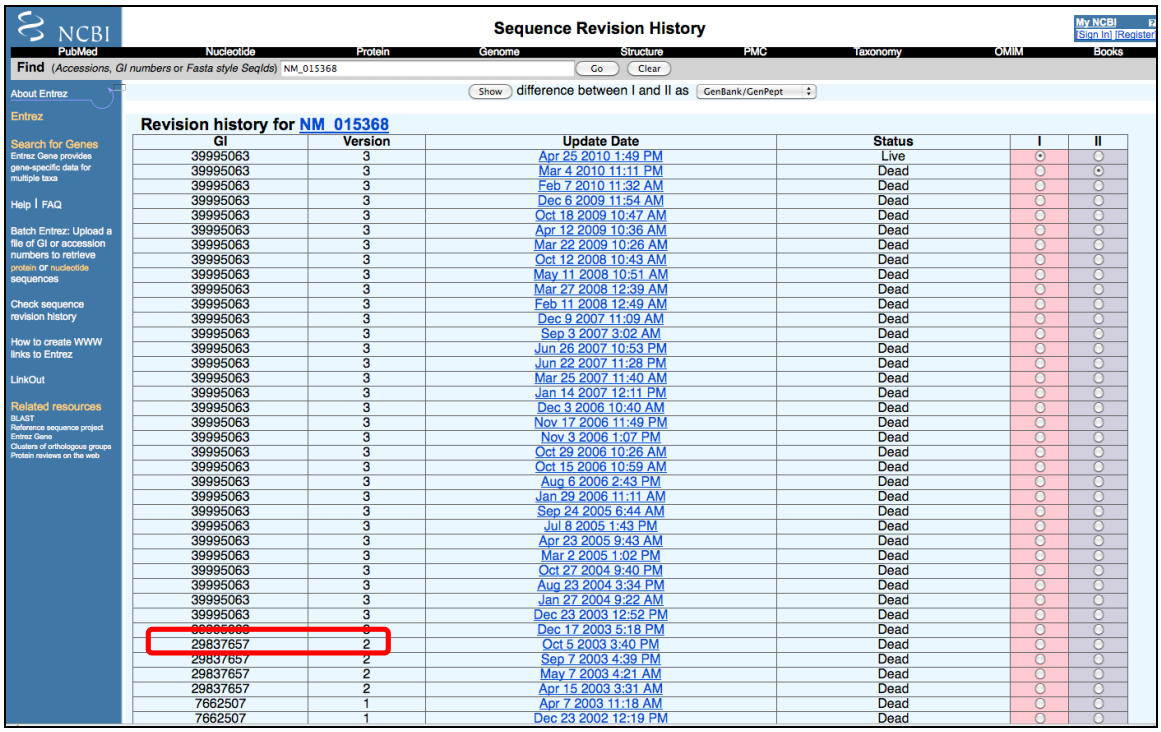

NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Format: GenBank FASTA Graphics More Formats▼
Download▼ Save▼ Links▼
GenBank: AY048509.1
GenBank(Full)
ASN.1
Revision History
Homo sapiens pannexin 1 partial cds
Features Sequence
LOCUS AY048509 1430 bp DNA linear PRI 29-SEP-2001
DEFINITION Homo sapiens pannexin 1 (PANX1) gene, partial cds.
ACCESSION AY048509
VERSION AY048509.1 GI:15808666
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1430)
AUTHORS Baranova,A., Ivanov,D., Skoblov,M., Pestova,A., Kelmanson,I.,
Shagin,D., Usman,N., Lukyanov,S. and Panchin,Y.
TITLE Mammalian pannexin family homologous to invertebrate gap-junction
proteins are differentially expressed in nervous tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1430)
AUTHORS Ivanov,D. and Baranova,A.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov
Institute of General Genetics, Gubkina 3, Moscow 119991
FEATURES Location/Qualifiers
source 1..1430
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="between D11S917 and D11S4176"
/clone="LANL LA1139-O20"
gene 606..>1430
/gene="PANX1"
Change Region Shown
Customize View
Analyze This Sequence
Run BLAST
Pick Primers
Articles about the PANX1 gene
Pannexin1 and pannexin3 delivery, cell surface dynamics, and cytoskeletal inte [J Biol Chem. 2010]
Pannexin 1 contributes to ATP release in airway epithelia. [Am J Respir Cell Mol Biol. 2009]
Lack of coupling between membrane stretching and pannexin-1[Biochem Biophys Res Commun. 2009]
» See all...
RefSeq mRNA
See the reference mRNA sequence for the PANX1 gene (NM_015368.3).
More about the PANX1 gene
The protein encoded by this gene belongs to the innexin family. Innexin family members are the structural components of gap junctions. This ...
Also Known As: MGC21309, MRS1, PX1, UN...

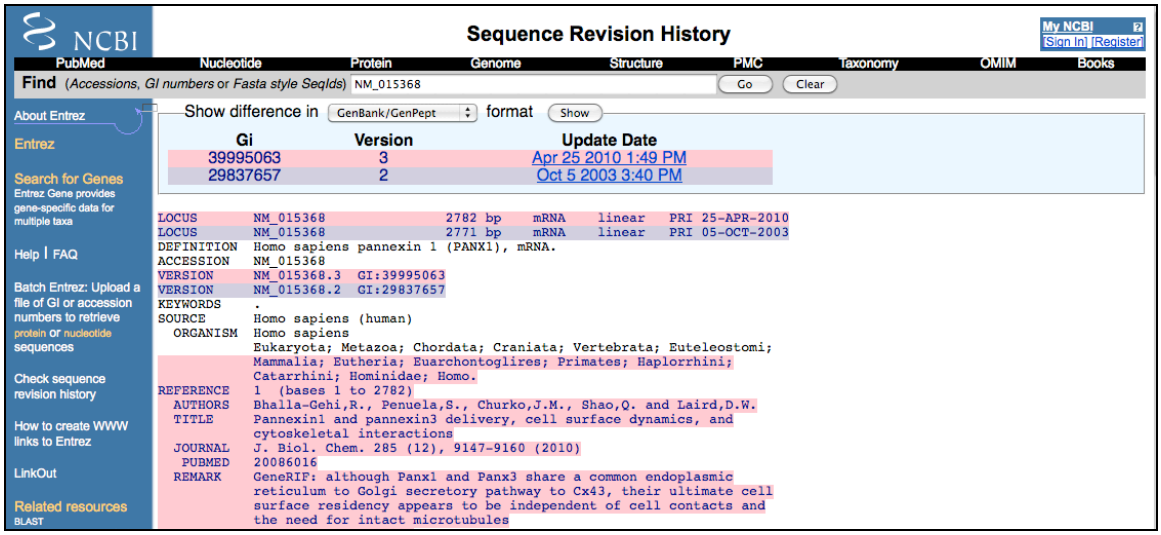

NCBI
Sequence Revision History
My NCBI
[Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Find (Accessions, GI numbers or Fasta style SeqIds) AY048509 Go Clear
About Entrez
Entrez
Search for Genes
Entrez Gene provides gene-specific data for multiple taxa
Help | FAQ
Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences
Show difference between I and II as GenBank/GenPept
Revision history for AY048509
GI Version Update Date Status I II
15808666 1 Mar 11 2010 1:28 AM Live
15808666 1 Nov 30 2009 5:51 PM Dead
15808666 1 Sep 29 2001 9:01 PM Dead
Accession AY048509 was first seen at NCBI on Sep 29 2001 9:01 PM

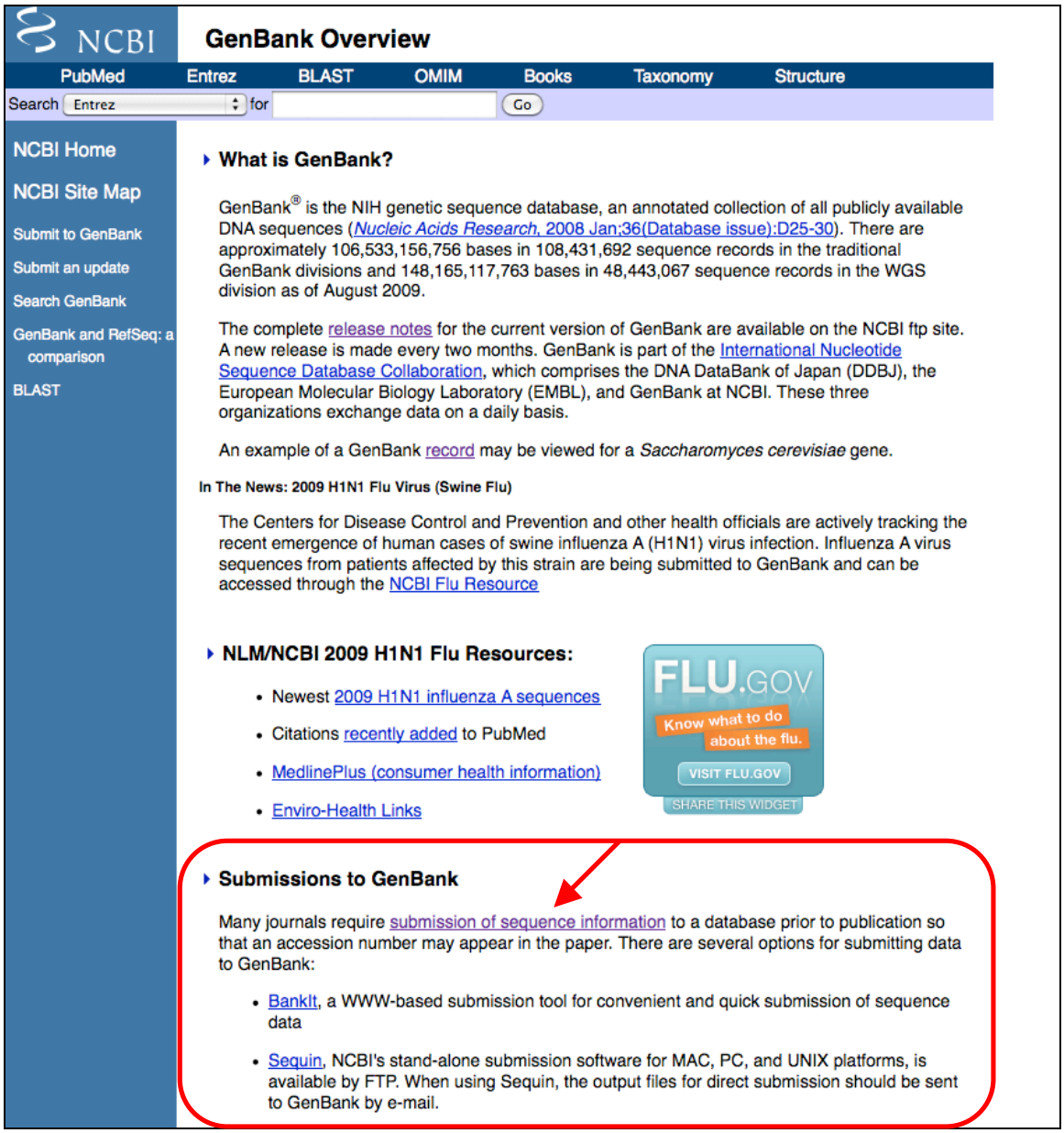
```
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2001) Genome Analysis Laboratory, Vavilov
            Institute of General Genetics, Gubkina 3, Moscow 119991
FEATURES             Location/Qualifiers
     source          1..1430
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="between D11S917 and D11S4176"
                     /clone="LANL LA1139-020"
     gene            606..>1430
                     /gene="PANX1"
     mRNA            606..>1170
                     /gene="PANX1"
                     /product="pannexin 1"
     exon            606..1170
                     /gene="PANX1"
                     /number=1
     CDS             990..>1170
                     /gene="PANX1"
                     /note="putative gap junction protein; contains four
                     putative transmembrane domains; similar to innexin gap
                     junction gene family"
                     /codon_start=1
                     /product="pannexin 1"
                     /protein_id="AAL06604.1"
                     /db_xref="GI:15808667"
                     /translation="MAIAQLATEYVFSDFLLKEPTEPKFKGLRLELAVDKMVTCIAVG
                     LPLLLISLAFAQEISI"
ORIGIN
        1 acctgcaggc ccccctctct ctgctggctc ccggaacaag gctctgattg ggatggcaga
       61 ggaaagagaa acgggaccag cagcgctact caggcctcga aactccacac tcactaccgt
```

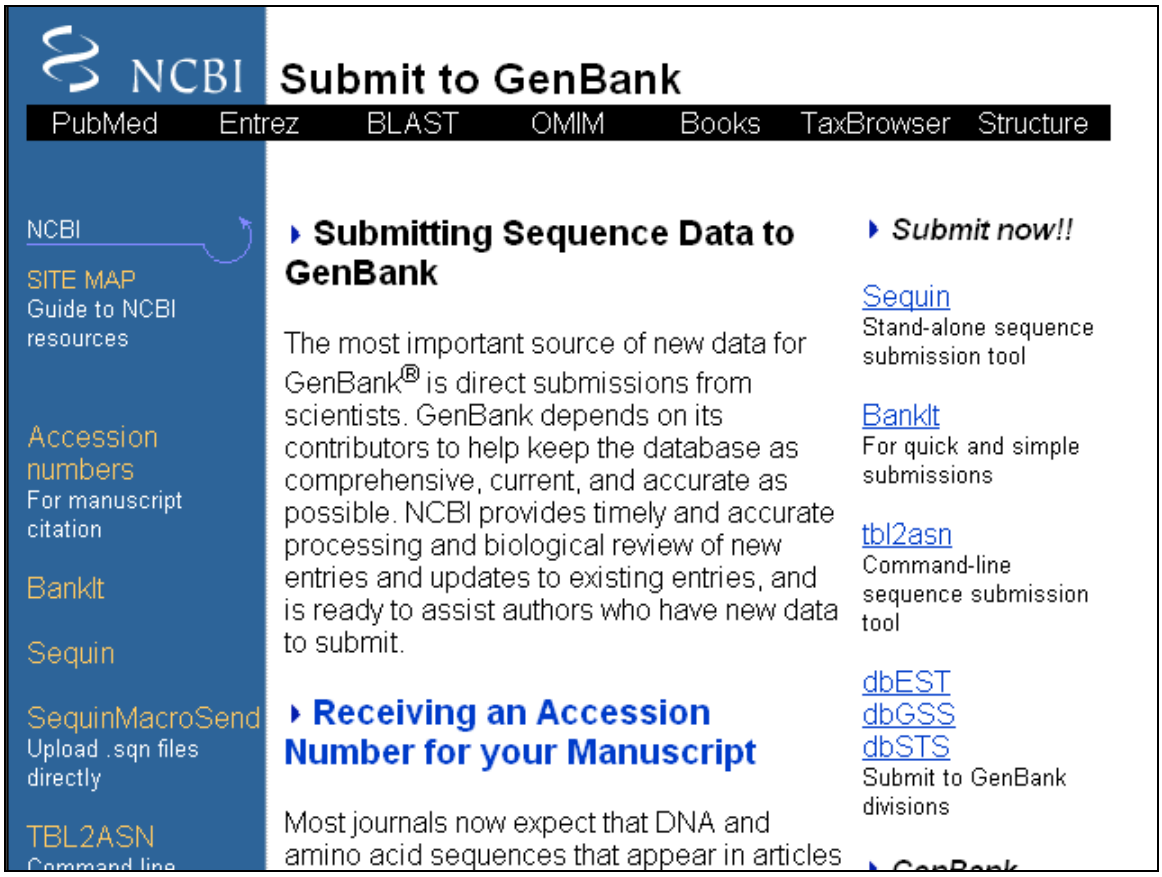

NCBI
Nucleotide
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for pannexin AND human[Organism] AND "srcdb refseq"[Prope Go Clear Save Search
Limits Preview/Index History Clipboard Details
Display Summary Show 20 Sort By Send to
All: 19 Bacteria: 0 INSDC (GenBank): 0 RefSeq: 19 mRNA: 4
Items 1 - 4 of 4
1. Homo sapiens pannexin 2 (PANX2), transcript variant 2, mRNA
2,984 bp linear mRNA
NM_001160300.1 GI:237757296
2. Homo sapiens pannexin 2 (PANX2), transcript variant 1, mRNA
3,069 bp linear mRNA
NM_052839.3 GI:163659919
3. Homo sapiens pannexin 3 (PANX3), mRNA
1,609 bp linear mRNA
NM_052959.2 GI:39995066
4. Homo sapiens pannexin 1 (PANX1), mRNA
2,782 bp linear mRNA
NM_015368.3 GI:39995063


NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Format: GenBank FASTA Graphics More Formats Download Save Links
NCBI Reference Sequence: NM_015368.3
Homo sapiens pannexin 1 (PANX1), mRNA
Comment Features Sequence
LOCUS NM_015368 2782 bp mRNA linear PRI 25-APR-2010
DEFINITION Homo sapiens pannexin 1 (PANX1), mRNA.
ACCESSION NM_015368
VERSION NM_015368.3 GI:39995063
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2782)
AUTHORS Bhalla-Gehi,R., Penuela,S., Churko,J.M., Shao,Q. and Laird,D.W.
TITLE Pannexin1 and pannexin3 delivery, cell surface dynamics, and cytoskeletal interactions
JOURNAL J. Biol. Chem. 285 (12), 9147-9160 (2010)
PUBMED 20086016
REMARK GeneRIF: although Panx1 and Panx3 share a common endoplasmic reticulum to Golgi secretory pathway to Cx43, their ultimate cell surface residency appears to be independent of cell contacts and the need for intact microtubules
REFERENCE 2 (bases 1 to 2782)
AUTHORS Ransford,G.A., Fregien,N., Qiu,F., Dahl,G., Conner,G.E. and Salathe,M.
TITLE Pannexin 1 contributes to ATP release in airway epithelia
JOURNAL Am. J. Respir. Cell Mol. Biol. 41 (5), 525-534 (2009)
PUBMED 19213873
REMARK GeneRIF: Pannexin 1 contributes to ATP release in airway epithelia.
REFERENCE 3 (bases 1 to 2782)
AUTHORS Reyes,J.P., Hernandez-Carballo,C.Y., Perez-Flores,G., Perez-Cornejo,P. and Arreola,J.
Change Region Shown
Customize View
Analyze This Sequence
Run BLAST
Pick Primers
Articles about the PANX1 gene
Pannexin1 and pannexin3 delivery, cell surface dynamics, and cytoskeletal in [J Biol Chem. 2010]
Pannexin 1 contributes to ATP release in airway epithelia. [Am J Respir Cell Mol Biol. 2009]
Lack of coupling between membrane stretching and panr [Biochem Biophys Res Commun. 2009]
» See all...
RefSeq Protein Product
See the reference protein sequence for pannexin-1 (NP_056183.2).
More about the PANX1 gene
The protein encoded by this gene belongs to the innexin family. Innexin family members are the structural components of gap junctions. This ...
Also Known As: MGC21309, MRS1, PX1, UN...


NCBI
Sequence Revision History
My NCBI [Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Find (Accessions, GI numbers or Fasta style SeqIds) NM_015368 Go Clear
About Entrez
Entrez
Search for Genes
Entrez Gene provides gene-specific data for multiple taxa
Help | FAQ
Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences
Check sequence revision history
How to create WWW links to Entrez
LinkOut
Related resources
BLAST
Reference sequence project
Entrez Gene
Clusters of orthologous groups
Protein reviews on the web
Show difference between I and II as GenBank/GenPept
Revision history for NM_015368
GI Version Update Date Status I II
39995063 3 Apr 25 2010 1:49 PM Live
39995063 3 Mar 4 2010 11:11 PM Dead
39995063 3 Feb 7 2010 11:32 AM Dead
39995063 3 Dec 6 2009 11:54 AM Dead
39995063 3 Oct 18 2009 10:47 AM Dead
39995063 3 Apr 12 2009 10:36 AM Dead
39995063 3 Mar 22 2009 10:26 AM Dead
39995063 3 Oct 12 2008 10:43 AM Dead
39995063 3 May 11 2008 10:51 AM Dead
39995063 3 Mar 27 2008 12:39 AM Dead
39995063 3 Feb 11 2008 12:49 AM Dead
39995063 3 Dec 9 2007 11:09 AM Dead
39995063 3 Sep 3 2007 3:02 AM Dead
39995063 3 Jun 26 2007 10:53 PM Dead
39995063 3 Jun 22 2007 11:28 PM Dead
39995063 3 Mar 25 2007 11:40 AM Dead
39995063 3 Jan 14 2007 12:11 PM Dead
39995063 3 Dec 3 2006 10:40 AM Dead
39995063 3 Nov 17 2006 11:49 PM Dead
39995063 3 Nov 3 2006 1:07 PM Dead
39995063 3 Oct 29 2006 10:26 AM Dead
39995063 3 Oct 15 2006 10:59 AM Dead
39995063 3 Aug 6 2006 2:43 PM Dead
39995063 3 Jan 29 2006 11:11 AM Dead
39995063 3 Sep 24 2005 6:44 AM Dead
39995063 3 Jul 8 2005 1:43 PM Dead
39995063 3 Apr 23 2005 9:43 AM Dead
39995063 3 Mar 2 2005 1:02 PM Dead
39995063 3 Oct 27 2004 9:40 PM Dead
39995063 3 Aug 23 2004 3:34 PM Dead
39995063 3 Jan 27 2004 9:22 AM Dead
39995063 3 Dec 23 2003 12:52 PM Dead
39995063 3 Dec 17 2003 5:18 PM Dead
29837657 2 Oct 5 2003 3:40 PM Dead
29837657 2 Sep 7 2003 4:39 PM Dead
29837657 2 May 7 2003 4:21 AM Dead
29837657 2 Apr 15 2003 3:31 AM Dead
7662507 1 Apr 7 2003 11:18 AM Dead
7662507 1 Dec 23 2002 12:19 PM Dead


NCBI
Sequence Revision History
My NCBI [Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Find (Accessions, GI numbers or Fasta style SeqIds) NM_015368 Go Clear
About Entrez
Entrez
Search for Genes
Entrez Gene provides gene-specific data for multiple taxa
Help | FAQ
Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences
Check sequence revision history
How to create WWW links to Entrez
LinkOut
Related resources
BLAST
Show difference in GenBank/GenPept format Show
GI Version Update Date
39995063 3 Apr 25 2010 1:49 PM
29837657 2 Oct 5 2003 3:40 PM
LOCUS NM_015368 2782 bp mRNA linear PRI 25-APR-2010
LOCUS NM_015368 2771 bp mRNA linear PRI 05-OCT-2003
DEFINITION Homo sapiens pannexin 1 (PANX1), mRNA.
ACCESSION NM_015368
VERSION NM_015368.3 GI:39995063
VERSION NM_015368.2 GI:29837657
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2782)
AUTHORS Bhalla-Gehi,R., Penuela,S., Churko,J.M., Shao,Q. and Laird,D.W.
TITLE Pannexin1 and pannexin3 delivery, cell surface dynamics, and
cytoskeletal interactions
JOURNAL J. Biol. Chem. 285 (12), 9147-9160 (2010)
PUBMED 20086016
REMARK GeneRIF: although Panx1 and Panx3 share a common endoplasmic
reticulum to Golgi secretory pathway to Cx43, their ultimate cell
surface residency appears to be independent of cell contacts and
the need for intact microtubules


NCBI
GenBank Overview
PubMed
Entrez
BLAST
OMIM
Books
Taxonomy
Structure
Search
Entrez
for
Go
NCBI Home
NCBI Site Map
Submit to GenBank
Submit an update
Search GenBank
GenBank and RefSeq: a comparison
BLAST
What is GenBank?
GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (Nucleic Acids Research, 2008 Jan;36(Database issue):D25-30). There are approximately 106,533,156,756 bases in 108,431,692 sequence records in the traditional GenBank divisions and 148,165,117,763 bases in 48,443,067 sequence records in the WGS division as of August 2009.
The complete release notes for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.
An example of a GenBank record may be viewed for a Saccharomyces cerevisiae gene.
In The News: 2009 H1N1 Flu Virus (Swine Flu)
The Centers for Disease Control and Prevention and other health officials are actively tracking the recent emergence of human cases of swine influenza A (H1N1) virus infection. Influenza A virus sequences from patients affected by this strain are being submitted to GenBank and can be accessed through the NCBI Flu Resource
NLM/NCBI 2009 H1N1 Flu Resources:
Newest 2009 H1N1 influenza A sequences
Citations recently added to PubMed
MedlinePlus (consumer health information)
Enviro-Health Links
FLU.GOV
Know what to do about the flu.
VISIT FLU.GOV
SHARE THIS WIDGET
Submissions to GenBank
Many journals require submission of sequence information to a database prior to publication so that an accession number may appear in the paper. There are several options for submitting data to GenBank:
BankIt, a WWW-based submission tool for convenient and quick submission of sequence data
Sequin, NCBI's stand-alone submission software for MAC, PC, and UNIX platforms, is available by FTP. When using Sequin, the output files for direct submission should be sent to GenBank by e-mail.



Submission Procedure

Individual mRNA/Genomic	BankIt
Sets	Sequin
EST	batch-sub
GSS	batch-sub
STS	batch-sub
WGS	tbl2asn/ftp
HTG	fa2htgs/ftp
HTC	tbl2asn
FLIC	tbl2asn/ftp
Complete genomes	tbl2asn
TPA	BankIt/Sequin

Accession Numbers

GenBank/DDBJ/EMBL

Nucleotide:1 letter + 5 numbers
OR 2 letters + 6 numbers (AF123456)
Protein:3 letters + 5 numerals
WGS:4 letters + 2 numbers

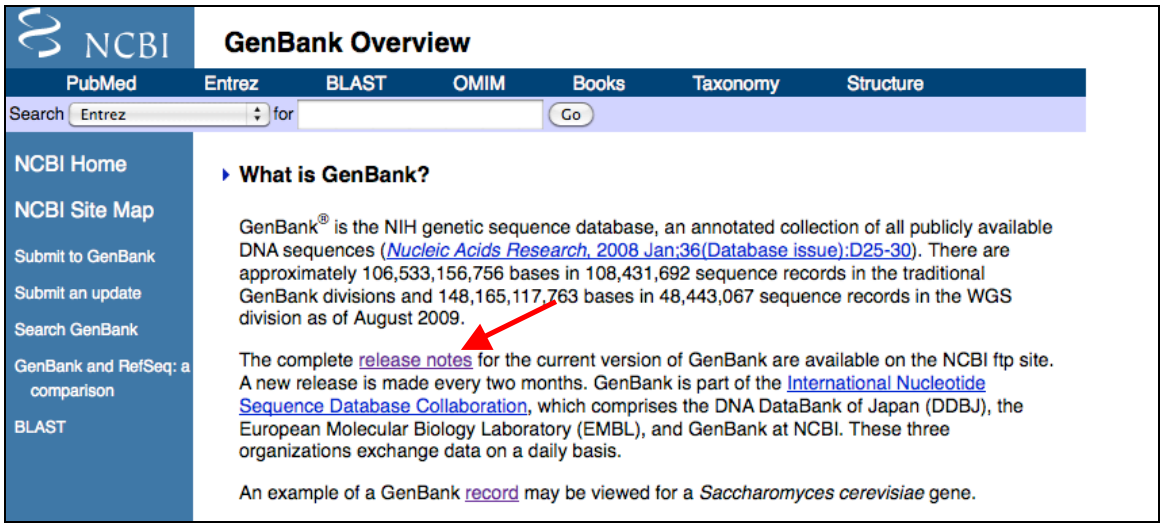
RefSeq

2 letters + underscore + 6 numbers (NM_123456)

Processing of Direct Submissions

Entries are checked for:
- Biological validity
- Vector contamination
- Publication Status
- Formatting/spelling

Completed sequences sent to submitters for review
5 days to review/update
Can hold until published
Released records available for searching in Entrez and BLAST within few days
Release to ftp every 2 months and daily updates




```
GBREL.TXT          Genetic Sequence Data Bank
                         April 15 2010

               NCBI-GenBank Flat File Release 177.0

                    Distribution Release Notes

 119112251 loci, 114348888771 bases, from 119112251 reported sequences

 This document describes the format and content of the flat files that
comprise releases of the GenBank nucleotide sequence database. If you
have any questions or comments about GenBank or this document, please
contact NCBI via email at info@ncbi.nlm.nih.gov or:

   GenBank
   National Center for Biotechnology Information
   National Library of Medicine, 38A, 8N805
   8600 Rockville Pike
   Bethesda, MD  20894
   USA
   Phone:  (301) 496-2475
   Fax:    (301) 480-9241

 GenBank releases do not include sequence records that originate from
third-parties (TPA) or from NCBI's Reference Sequence (RefSeq) project.
Rather, GenBank is the archival/primary resource which those other
efforts draw upon. For information about TPA and RefSeq, please visit:

        http://www.ncbi.nih.gov/Genbank/TPA.html
        http://www.ncbi.nlm.nih.gov/RefSeq

================================================================================
TABLE OF CONTENTS
================================================================================

1. INTRODUCTION

1.1 Release 177.0
1.2 Cutoff Date
1.3 Important Changes in Release 177.0
1.4 Upcoming Changes
1.5 Request for Direct Submission of Sequence Data
1.6 Organization of This Document
```

GenBank Database Divisions

BCT	Bacterial sequences
PRI	Primate sequences
ROD	Rodent sequences
MAM	Other mammalian sequences
VRT	Other vertebrate sequences
INV	Invertebrate sequences
PLN	Plant and Fungal sequences
VRL	Viral sequences
PHG	Phage sequences
SYN	Synthetic sequences
UNA	Unannotated sequences

BLAST database: nr, month except for environmental samples

GenBank Database Divisions

Division		BLAST database
EST	Expressed Sequence Tags	dbest, month
STS	Sequence Tagged Sites	dbsts, month
GSS	Genome Survey Sequences	dbgss, month
HTG	High Throughput Genomic sequences Phase 0, 1 and 2	htgs, month
PAT	Patent sequences	patent, month
HTC	High Throughput cDNA	nr, month
ENV	Environmental sampling sequences	env, month

Problem 2: Retrieve entries containing the word "FOXP2" from Entrez-Nucleotide. The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The summary of top 20 entries is displayed by default. Currently, how many entries are there in the Entrez-nucleotides CoreNucleotide, EST and GSS databases associated with FOXP2?

Change the pull down menu "Show" to a number large enough to view the summaries of all entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the Accession Number Prefixes page, if necessary.

Access entries from the CoreNucleotide database. Keep the genomic entries from GenBank by using the Preview/Index page (biomol genomic [properties] AND srcdb genbank [properties]). Is GenBank a redundant database? Note the multiple FOXP2 exon sequence entries. Are they duplicating entries?

Let us use the entry AF515032 as a sample GenBank entry with feature annotations. When the entry was last updated? Was its sequence updated after release? Does it contain a complete coding region? What is the accession number for the protein entry?

Go back to the Entrez report. Change the source database to RefSeq, molecule type to mRNA and human as organism by using the Preview/index page. Note that some of the entries are not for the FOXP2 gene. Restrict to the FOXP2 gene using the Preview/index page and the gene name field from the "All Fields" pull down menu. How many human FOXP2 transcript variants have been annotated? Save the sequences to a file by changing the "Display" to FASTA and "Send to" file.

Check the sequence revision history of NM_148900. What is the most recent change? How many times has the sequence of the entry been updated? When did these updates occur?

Problem 3: Retrieve entries containing the word "bacteriophage" from Entrez-Nucleotide. The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. Currently, how many entries are there in the Entrez-nucleotides CoreNucleotide, EST and GSS databases associated with bacteriophage?

Click on the CoreNucleotide link. The summary of top 20 entries is displayed by default. Select '500' from the Show pull-down menu to see a variety of different Genbank entries. Determine the source databases and types of entries such as mRNA or genomic. Consult the Accession Number Prefixes page, if necessary.

Next, keep the GenBank entries containing the terms bacteriophage and complete sequence in the title of the record by using the Preview/Index page (bacteriophage[Title] AND "complete sequence"[Title] AND "srcdb Genbank"[Properties]). Are all of these complete bacteriophage genome sequences?

Let us use the entry U32222 as a sample GenBank entry with feature annotations. Which bacteriophage sequence does it represent? When was the entry last updated? Was its sequence updated after the first release? What are the different ways to determine this? Access the Revision History page. Notice the 3 other accession numbers on the upper right side of the page. What do they represent? Return to the GenBank entry.

Count the number of entries under the Reference field. How many of these are actual PubMed entries? What are the other types of References? Now scroll to the Features table. This GenBank record is a well-annotated phage genome entry. Note the different types of features are annotated here. Navigate to the variation feature at position 3980. What is the nucleotide sequence in this entry at position 3980? What is it replaced with in the Vam38 mutant? What is the amino acid change caused due to this nucleotide change? What is the translation table (genetic code) used for the translation of the bacteriophage sequence? Access that translation table. What are the codons for glutamine using this translation table? Find the codon that is present in the DNA sequence of this bacteriophage (displayed at the bottom of the record) for the wild type amino acid (glutamine). Determine the variant codon in the Vam38 mutant. Which gene includes this variation? Click on the corresponding CDS feature. What sequence is now displayed? Navigate back to the full GenBank record for U32222. Display the entry in the "Graph" format to display the graphical representation of the annotation of various features on the entry. Finally, save the sequence to a file by changing the "Display" to FASTA and "Send to" file.