Mini Courses NCBI Resources one course at a time

NCBI's GenBank Quick Start
An NCBI Mini-Course

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 160,000 organisms.

The mini-course covers 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 features annotated on it
* 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 exercise demonstrated in the workshop**.**

URL: http://www.ncbi.nlm.nih.gov/Class/minicourses/quickgenbank.html

Course developed by Medha Bhagwat (bhagwat@ncbi.nlm.nih.gov)

Problems:

1. Retrieve entries containing the word "pannexin" from Entrez-Nucleotide. The Entrez Nucleotides 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 Entrez-nucleotides 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, 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 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?

GenBank History

Originally built and maintained at
Los Alamos National Laboratory (LANL)

Early 1990s, Congress awarded responsibility to NCBI

Initially, indexers scanned the literature
and typed in the sequences

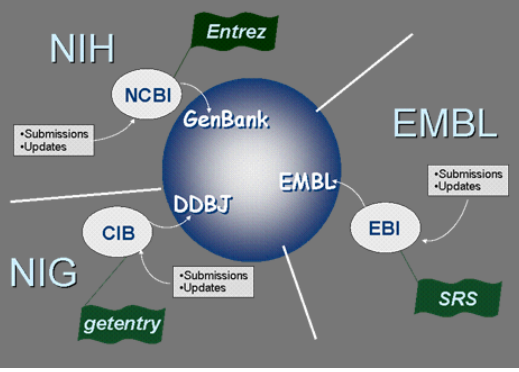
Now sequences are deposited directly by labs

Direct submissions since 1993

NCBI

International Nucleotide Sequence Database Collaboration

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



NCBI

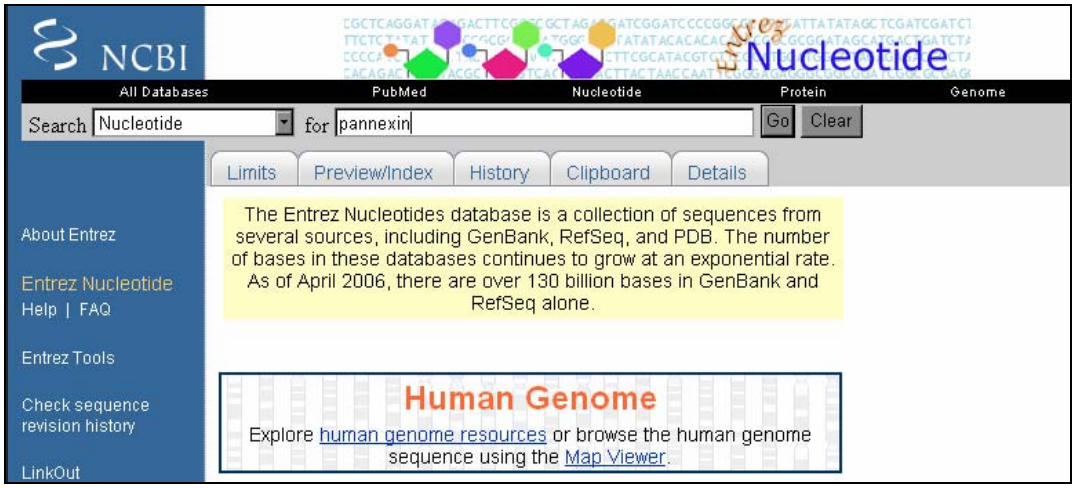
GenBank	RefSeq
Archival/repository	Curated
Redundant	Non-redundant
Submitter owner	NCBI owner
Sequenced	Combined/edited

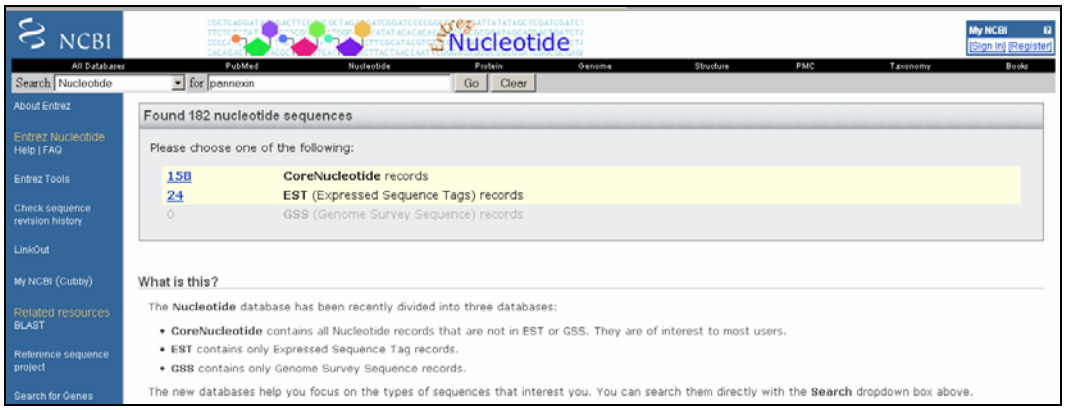
NCBI

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)

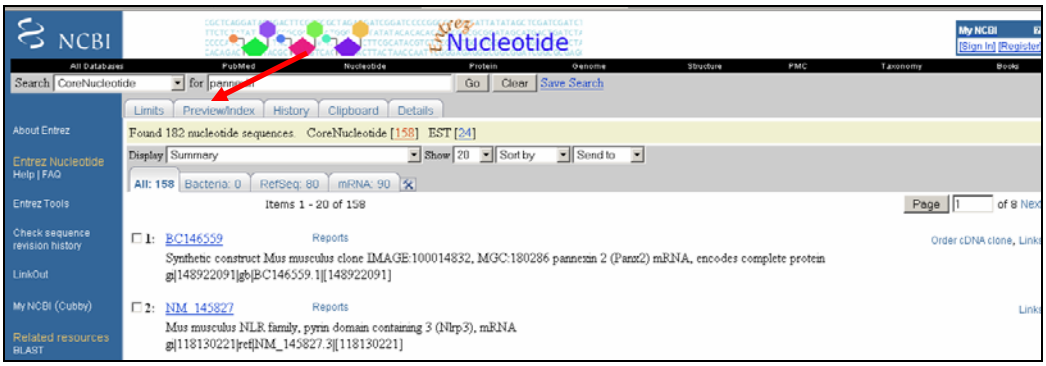
NCBI

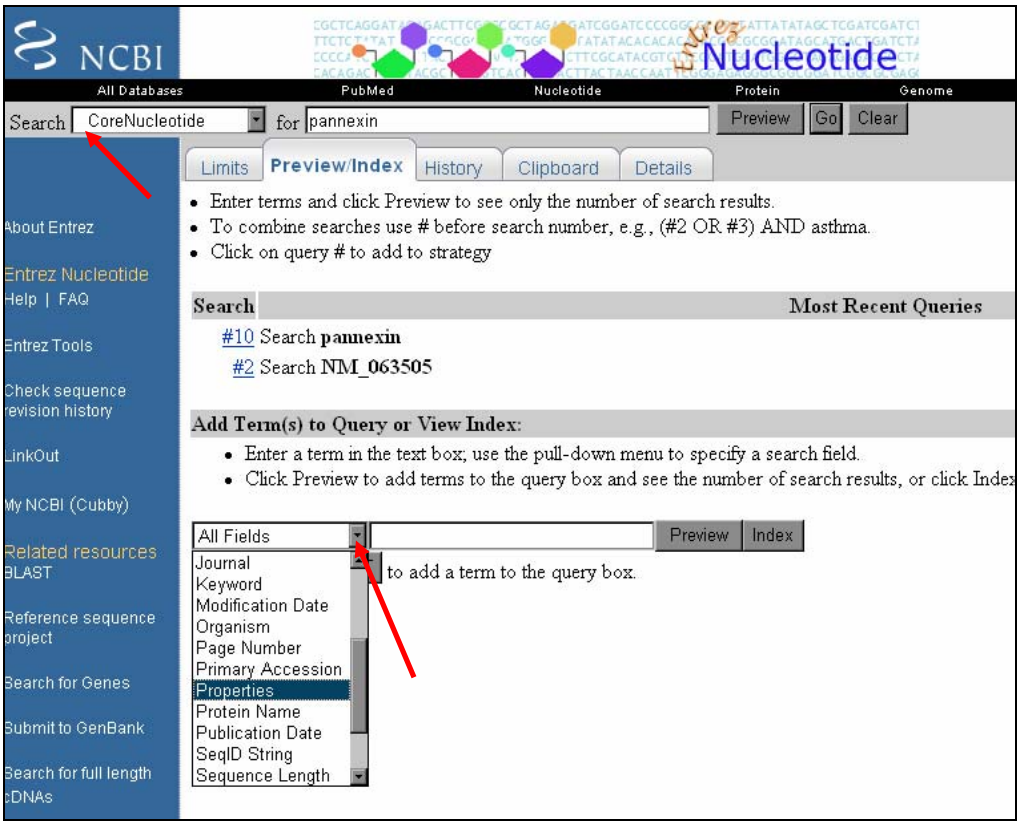




NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for pannexin Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Found 182 nucleotide sequences. CoreNucleotide [158] EST [24]
Display Summary Show 20 Sort by Send to
All: 158 Bacteria: 0 RefSeq: 80 mRNA: 90
Page 1 of 8 Next
1: BC146559 Reports Order cDNA clone, Links
Synthetic construct Mus musculus clone IMAGE:100014832, MGC:180286 pannexin 2 (Panx2) mRNA, encodes complete protein
gi|148922091|gb|BC146559.1|[148922091]
2: NM_145827 Reports Links
Mus musculus NLR family, pyrin domain containing 3 (Nlrp3), mRNA
gi|118130221|ref|NM_145827.3|[118130221]
3: BC141516 Reports Order cDNA clone, Links
Synthetic construct Mus musculus clone IMAGE:100014580, MGC:175435 pannexin 2 (Panx2) mRNA, encodes complete protein
gi|146327291|gb|BC141516.1|[146327291]
4: NW_001800828 Reports Links
Equus caballus genomic contig, reference assembly (based on EquCab1 scaffold_184), whole genome shotgun sequence
gi|149759322|ref|NW_001800828.1|EcaUn_WGA185_1|[149759322]
5: XM_001489989 Reports Links
PREDICTED: Equus caballus similar to Pannexin 2 (LOC100056108), mRNA
gi|149759299|ref|XM_001489989.1|[149759299]
6: NW_001799731 Reports Links
Equus caballus chromosome 7 genomic contig, reference assembly (based on EquCab1 scaffold_9), whole genome shotgun sequence
gi|149720118|ref|NW_001799731.1|Eca7_WGA30_1|[149720118]

NCBI
Nucleotide
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search CoreNucleotide for pannexin Go Clear Save Search
Limits Preview/Index History Clipboard Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
Found 182 nucleotide sequences. CoreNucleotide [158] EST [24]
Display Summary Show 20 Sort by Send to
5
10
20
50
100
200
500
All: 158 Bacteria: 0 RefSeq: 80 mRNA: 90
Items 1 - 20 of 158
Page 1 of 8 Next
1: BC146559 Reports Order cDNA clone, Links
2: NM_145827 Reports Links
3: BC141516 Reports Order cDNA clone, Links
4: NW_001800828 Reports Links
5: XM_001489989 Reports Links
6: NW_001799731 Reports Links
7: XM_001491287 Reports Links


NCBI
Nucleotide
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
Books
Search CoreNucleotide for pannexin
Go
Clear
Save Search
Limits
Preview/Index
History
Clipboard
Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Found 182 nucleotide sequences. CoreNucleotide [158] EST [24]
Display Summary
Show 20
Sort by
Send to
All: 158
Bacteria: 0
RefSeq: 80
mRNA: 90
Items 1 - 20 of 158
Page 1 of 8 Next
1: BC146559
Reports
Synthetic construct Mus musculus clone IMAGE:100014832, MGC:180286 pannexin 2 (Panx2) mRNA, encodes complete protein
gi|148922091|gb|BC146559.1|[148922091]
2: NM_145827
Reports
Mus musculus NLR family, pyrin domain containing 3 (Nlrp3), mRNA
gi|118130221|ref|NM_145827.3|[118130221]


NCBI
Nucleotide
All Databases
PubMed
Nucleotide
Protein
Genome
Search CoreNucleotide for pannexin
Preview
Go
Clear
Limits
Preview/Index
History
Clipboard
Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
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
#10 Search pannexin
#2 Search NM_063505
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
All Fields
Preview
Index
to add a term to the query box.
Journal
Keyword
Modification Date
Organism
Page Number
Primary Accession
Properties
Protein Name
Publication Date
SeqID String
Sequence Length

All Databases PubMed Nucleotide Protein

Search CoreNucleotide for pannexin Preview Go Cl

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Nucleotide
Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources
BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

- 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 Re**

#10 Search **pannexin**

#2 Search **NM_063505**

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

Properties biomol Preview Index

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

biomol crna(1511)
biomol genomic(25885125)
biomol genomic mrna(42)
biomol mrna(1843056)
biomol other(15549)
biomol other genetic(2214)
biomol pre rna(976)
biomol rna(2029649)
biomol rrna(7399)
biomol scrna(88)

Up Down

All Databases PubMed Nucleotide Protein Genome

Search CoreNucleotide for pannexin AND "biomol genomic"[Properties] Preview Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

Entrez Nucleotide
Help | FAQ

Entrez Tools

Check sequence revision history

LinkOut

My NCBI (Cubby)

Related resources
BLAST

Reference sequence project

Search for Genes

Submit to GenBank

Search for full length cDNAs

- 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**

#10 Search **pannexin**

#2 Search **NM_063505**

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

Properties srcdb Preview Index

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

srcdb ddbj(1742576)
srcdb ddbj/embl/genbank(26565711)
srcdb embl(2703718)
srcdb genbank(22119417)
srcdb pdb(6845)
srcdb refseq(1440162)
srcdb refseq inferred(52)
srcdb refseq known(500945)
srcdb refseq model(329471)
srcdb refseq predicted(11566)

Up Down

NCBI Nucleotide

My NCBI [Sign In] [Register]

All Databases | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search CoreNucleotide for pannexin AND "biomol genomic"[Properties] AN[Go Clear Save Search

Limits | Preview/Index | History | Clipboard | Details

About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes

Found 15 nucleotide sequences. CoreNucleotide [15]

Display Summary Show 20 Sort by Send to

All: 15 | Bacteria: 0 | RefSeq: 0 | mRNA: 0

Items 1 - 15 of 15 One page.

1: CH473993 Reports Links
Rattus norvegicus 181000120144526 genomic scaffold, whole genome shotgun sequence
gi|74325685|gb|CH473993.2|[74325685]

2: CH474096 Reports Links
Rattus norvegicus 342000077087993 genomic scaffold, whole genome shotgun sequence
gi|74325592|gb|CH474096.2|[74325592]

3: CH466550 Reports Links
Mus musculus 232000009743879 genomic scaffold, whole genome shotgun sequence
gi|74140747|gb|CH466550.2|[74140747]

6: CH471138 Reports Links
Homo sapiens 211000035835149 genomic scaffold, whole genome shotgun sequence
gi|71514907|gb|CH471138.1|[71514907]

7: AF398508 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exons 3, 4, 5, and complete cds
gi|15193202|gb|AF398508.1|AF398506S3[15193202]

8: AF398507 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exon 2
gi|15193201|gb|AF398507.1|AF398506S2[15193201]

9: AF398506 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exon 1
gi|15193200|gb|AF398506.1|AF398506S1[15193200]

10: AH010983 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, complete cds
gi|15193199|gb|AH010983.1|SEG_AF398506S[15193199]

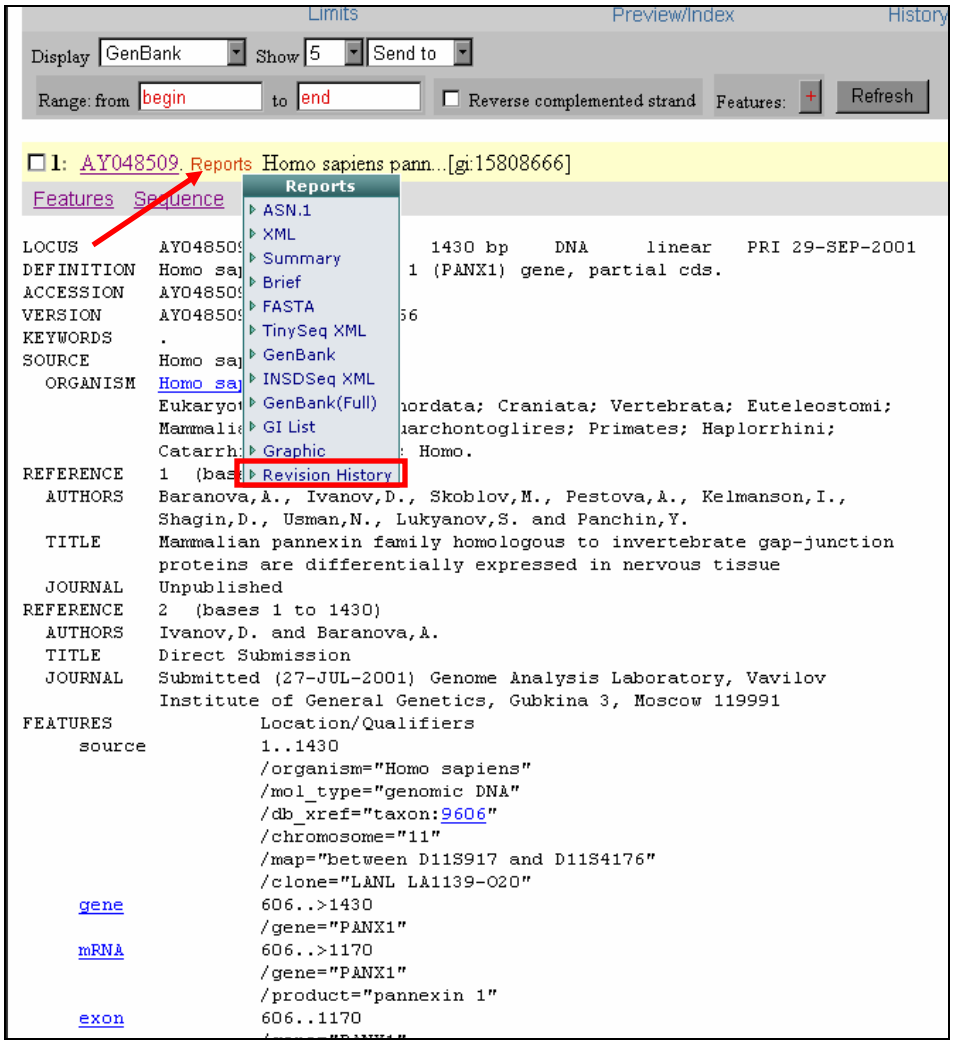
11: AY048509 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, partial cds
gi|15808666|gb|AY048509.1|[15808666]

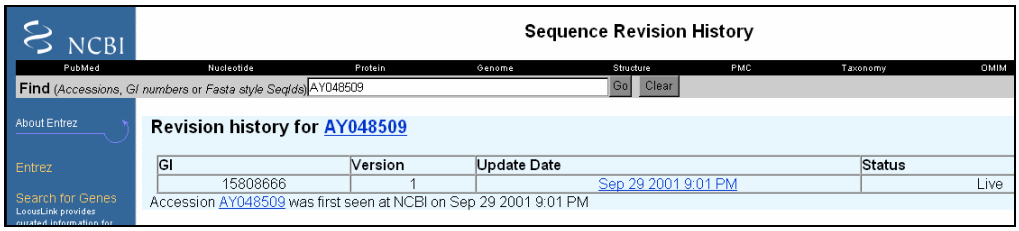
12: AF283663 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exons, partial cds
gi|14794510|gb|AF283663.1|AF283661S3[14794510]

13: AF283662 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exon
gi|14794509|gb|AF283662.1|AF283661S2[14794509]

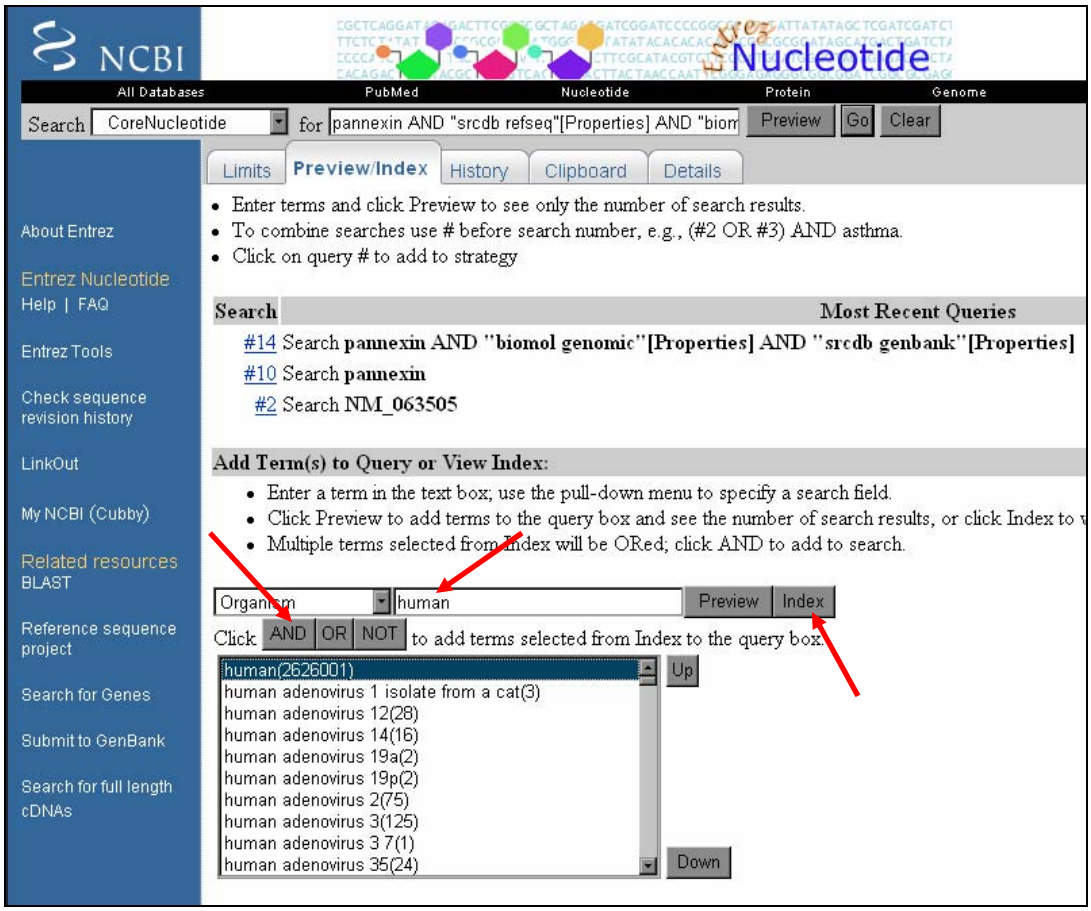
14: AF283661 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, exon
gi|14794508|gb|AF283661.1|AF283661S1[14794508]

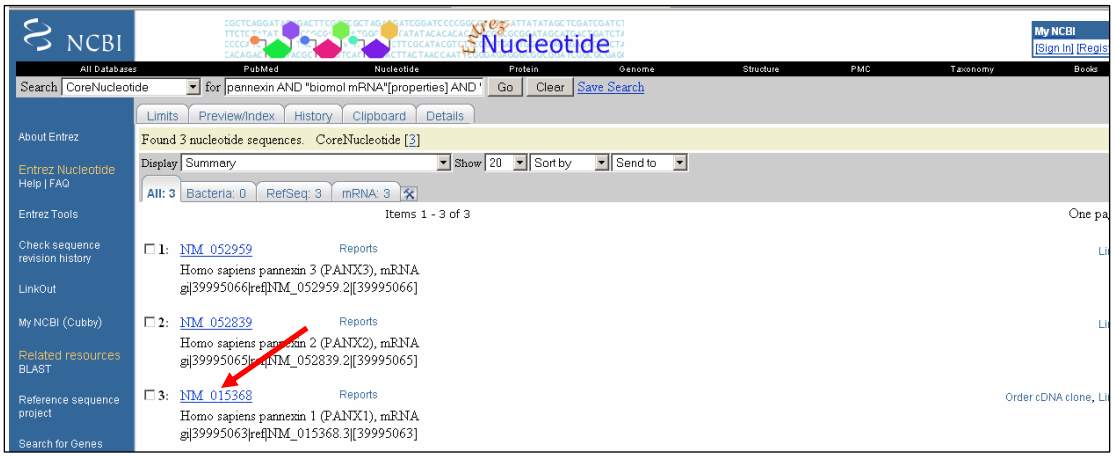
15: AH010945 Reports Links
Homo sapiens pannexin 1 (PANX1) gene, partial cds
gi|14794507|gb|AH010945.1|SEG_AF283661S[14794507]

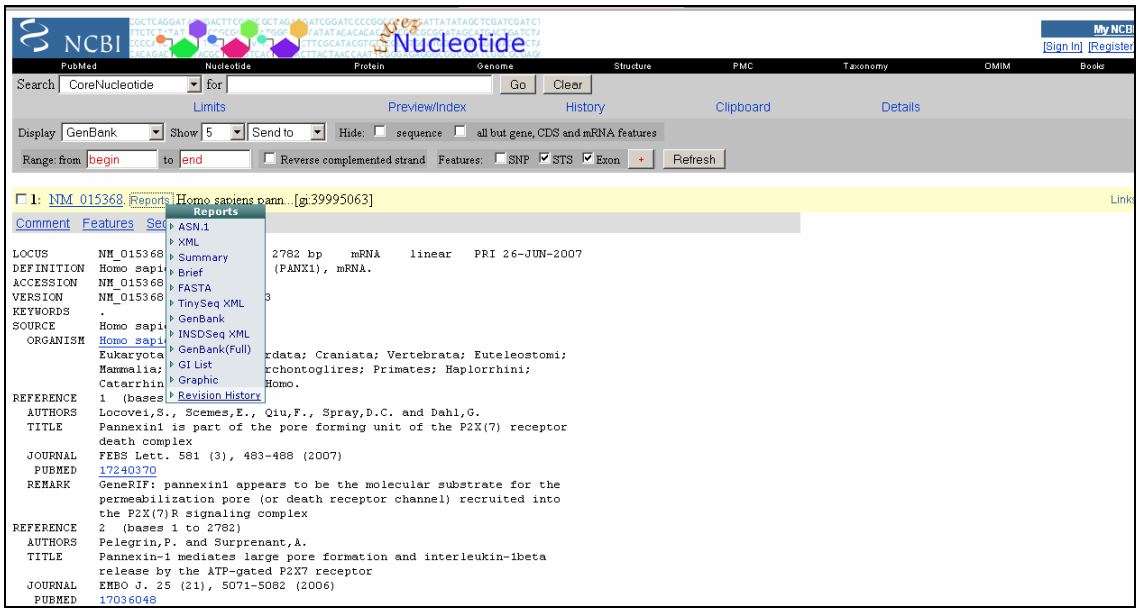

Limits
Preview/Index
History
Display GenBank
Show 5
Send to
Range: from begin to end
Reverse complemented strand
Features:
Refresh
1: AY048509. Reports Homo sapiens pann...[gi:15808666]
Features
Sequence
Reports
ASN.1
XML
Summary
Brief
FASTA
TinySeq XML
GenBank
INSDSeq XML
GenBank(Full)
GI List
Graphic
Revision History
LOCUS 1430 bp DNA linear PRI 29-SEP-2001
DEFINITION 1 (PANX1) gene, partial cds.
ACCESSION
VERSION
KEYWORDS .
SOURCE
ORGANISM
REFERENCE 1 (bas
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 LA1139-020"
gene 606..>1430
/gene="PANX1"
mRNA 606..>1170
/gene="PANX1"
/product="pannexin 1"
exon 606..1170

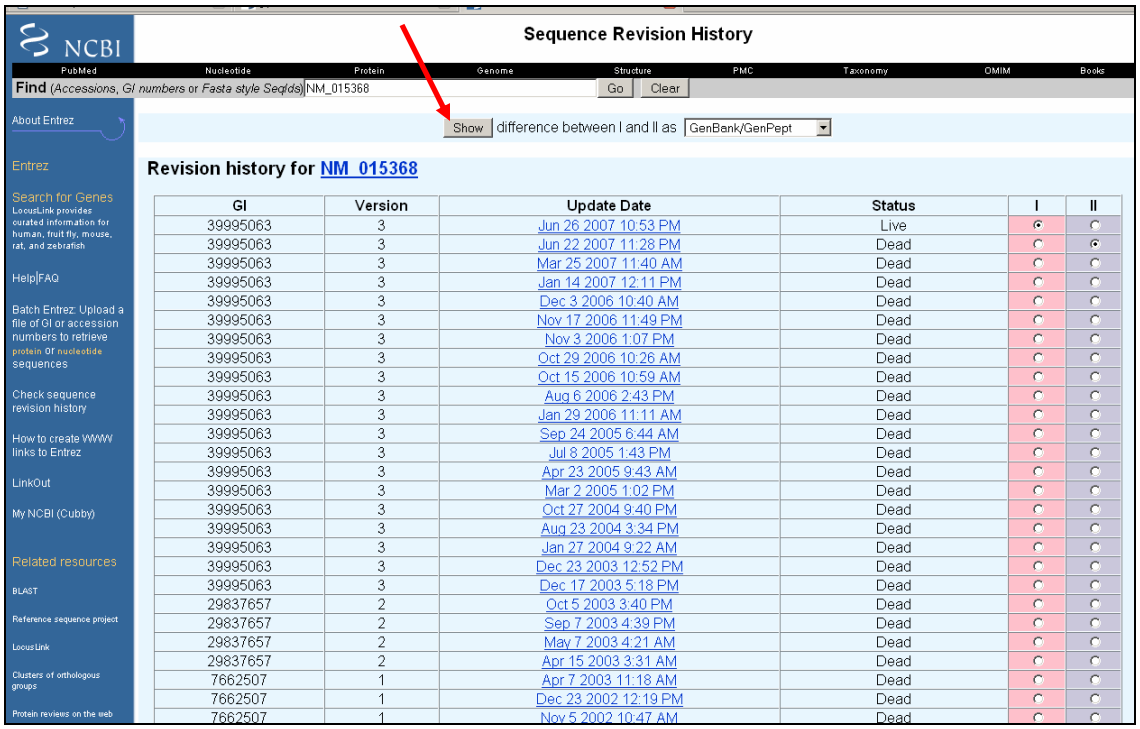

NCBI
Sequence Revision History
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM
Find (Accessions, GI numbers or Fasta style SeqIds) AY048509 Go Clear
About Entrez
Entrez
Search for Genes
Revision history for AY048509
GI Version Update Date Status
15808666 1 Sep 29 2001 9:01 PM Live
Accession 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-O20"
     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
Search CoreNucleotide for pannexin AND "srcdb refseq"[Properties] AND "biom
Preview
Go
Clear
Limits
Preview/Index
History
Clipboard
Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Submit to GenBank
Search for full length cDNAs
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
#14 Search pannexin AND "biomol genomic"[Properties] AND "srcdb genbank"[Properties]
#10 Search pannexin
#2 Search NM_063505
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
Multiple terms selected from Index will be ORed; click AND to add to search.
Organism
human
Preview
Index
Click AND OR NOT to add terms selected from Index to the query box.
human(2626001)
human adenovirus 1 isolate from a cat(3)
human adenovirus 12(28)
human adenovirus 14(16)
human adenovirus 19a(2)
human adenovirus 19p(2)
human adenovirus 2(75)
human adenovirus 3(125)
human adenovirus 3 7(1)
human adenovirus 35(24)
Up
Down


NCBI
Nucleotide
My NCBI
[Sign In] [Regis
All Databases
PubMed
Nucleotide
Protein
Genome
Structure
PMC
Taxonomy
Books
Search CoreNucleotide for pannexin AND "biomol mRNA"[properties] AND '
Go
Clear
Save Search
Limits
Preview/Index
History
Clipboard
Details
About Entrez
Entrez Nucleotide
Help | FAQ
Entrez Tools
Check sequence revision history
LinkOut
My NCBI (Cubby)
Related resources
BLAST
Reference sequence project
Search for Genes
Found 3 nucleotide sequences. CoreNucleotide [3]
Display Summary
Show 20
Sort by
Send to
All: 3
Bacteria: 0
RefSeq: 3
mRNA: 3
Items 1 - 3 of 3
1: NM_052959
Reports
Homo sapiens pannexin 3 (PANX3), mRNA
gi|39995066|ref|NM_052959.2|[39995066]
2: NM_052839
Reports
Homo sapiens pannexin 2 (PANX2), mRNA
gi|39995065|ref|NM_052839.2|[39995065]
3: NM_015368
Reports
Homo sapiens pannexin 1 (PANX1), mRNA
gi|39995063|ref|NM_015368.3|[39995063]
Order cDNA clone





GI	Version	Update Date	Status	I	II
39995063	3	Jun 26 2007 10:53 PM	Live	●	○
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	○	○
7662507	1	Nov 5 2002 10:47 AM	Dead	○	○

NCBI

Sequence Revision History

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
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

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

My NCBI (Cubby)

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Show difference in GenBank/GenPept format Show

Gi	Version	Update Date
39995063	3	Jun 26 2007 10:53 PM
39995063	3	Jun 22 2007 11:28 PM

```
LOCUS       NM_015368                2782 bp    mRNA    linear   PRI 26-JUN-2007
LOCUS       NM_015368                2782 bp    mRNA    linear   PRI 22-JUN-2007
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   Locovei,S., Scemes,E., Qiu,F., Spray,D.C. and Dahl,G.
  AUTHORS   Locovei S, Scemes E, Qiu F, Spray DC and Dahl G.
  TITLE     Pannexin1 is part of the pore forming unit of the P2X(7) receptor
            death complex
  JOURNAL   FEBS Lett. 581 (3), 483-488 (2007)
   PUBMED   17240370
  REMARK    GeneRIF: pannexin1 appears to be the molecular substrate for the
            permeabilization pore (or death receptor channel) recruited into
            the P2X(7)R signaling complex
REFERENCE   2  (bases 1 to 2782)
  AUTHORS   Pelegrin,P. and Surprenant,A.
  AUTHORS   Pelegrin P and Surprenant A.
  TITLE     Pannexin-1 mediates large pore formation and interleukin-1beta
            release by the ATP-gated P2X7 receptor
  JOURNAL   EMBO J. 25 (21), 5071-5082 (2006)
   PUBMED   17036048
  REMARK    GeneRIF: Panx1 in the plasma membrane of the macrophage couples to
  REMARK    GeneRIF: pannexin-1 is required for processing of caspase-1 and
            release of mature IL-1beta induced by P2X(7) receptor activation.
            GeneRIF: Panx1 in the plasma membrane of the macrophage couples to
            the purinergic P2X7 receptor and permeabilize the macrophage
            membrane, this signaling is required for the processing and release
```

NCBI

Submit to GenBank

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

NCBI

SITE MAP
Guide to NCBI resources

Accession numbers
For manuscript citation

BankIt

Sequin

SequinMacroSend
Upload .sqn files directly

TBL2ASN
Command line

Submitting Sequence Data to GenBank

The most important source of new data for GenBank® is direct submissions from scientists. GenBank depends on its contributors to help keep the database as comprehensive, current, and accurate as possible. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit.

Receiving an Accession Number for your Manuscript

Most journals now expect that DNA and amino acid sequences that appear in articles

Submit now!!

Sequin
Stand-alone sequence submission tool

BankIt
For quick and simple submissions

tbl2asn
Command-line sequence submission tool

dbEST
dbGSS
dbSTS
Submit to GenBank divisions

GenBank

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

NCBI **GenBank Overview**

PubMed Entrez BLAST OMIM Books Taxonomy Structure

Search Entrez for Go

NCBI

SITE MAP

Submit to GenBank

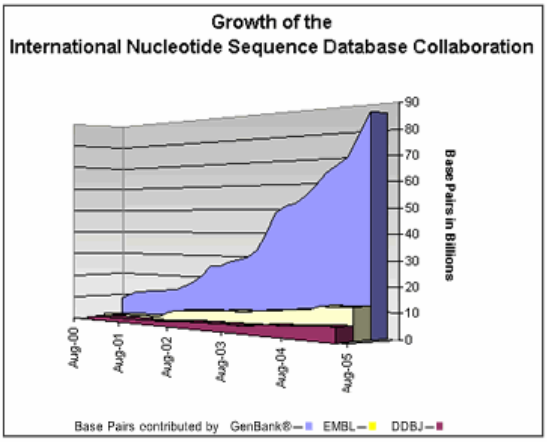
Updates

Search GenBank
Entrez Nucleotide

BLAST

International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related press release.



What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research* 2007 Jan ;35(Database issue):D21-5). There are approximately 65,369,091,950 bases in 61,132,599 sequence records in the traditional GenBank divisions and 80,369,977,826 bases in 17,960,667 sequence records in the WGS division as of August 2006.

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.

```
GBREL.TXT          Genetic Sequence Data Bank
                         June 15 2007

               NCBI-GenBank Flat File Release 160.0

                    Distribution Release Notes

 73078143 loci, 77248690945 bases, from 73078143 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 160.0
1.2 Cutoff Date
1.3 Important Changes in Release 160.0
1.4 Upcoming Changes
1.5 Request for Direct Submission of Sequence Data
1.6 Organization of This Document

2. ORGANIZATION OF DATA FILES
```

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

NCB

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 duplicate 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?