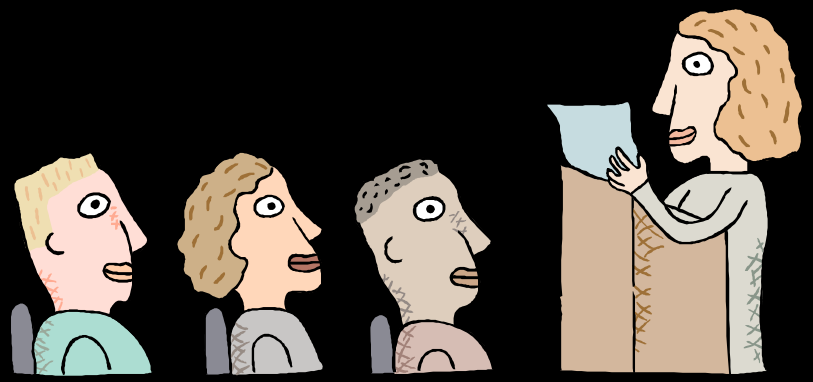


BMG 744
Bioinformatics

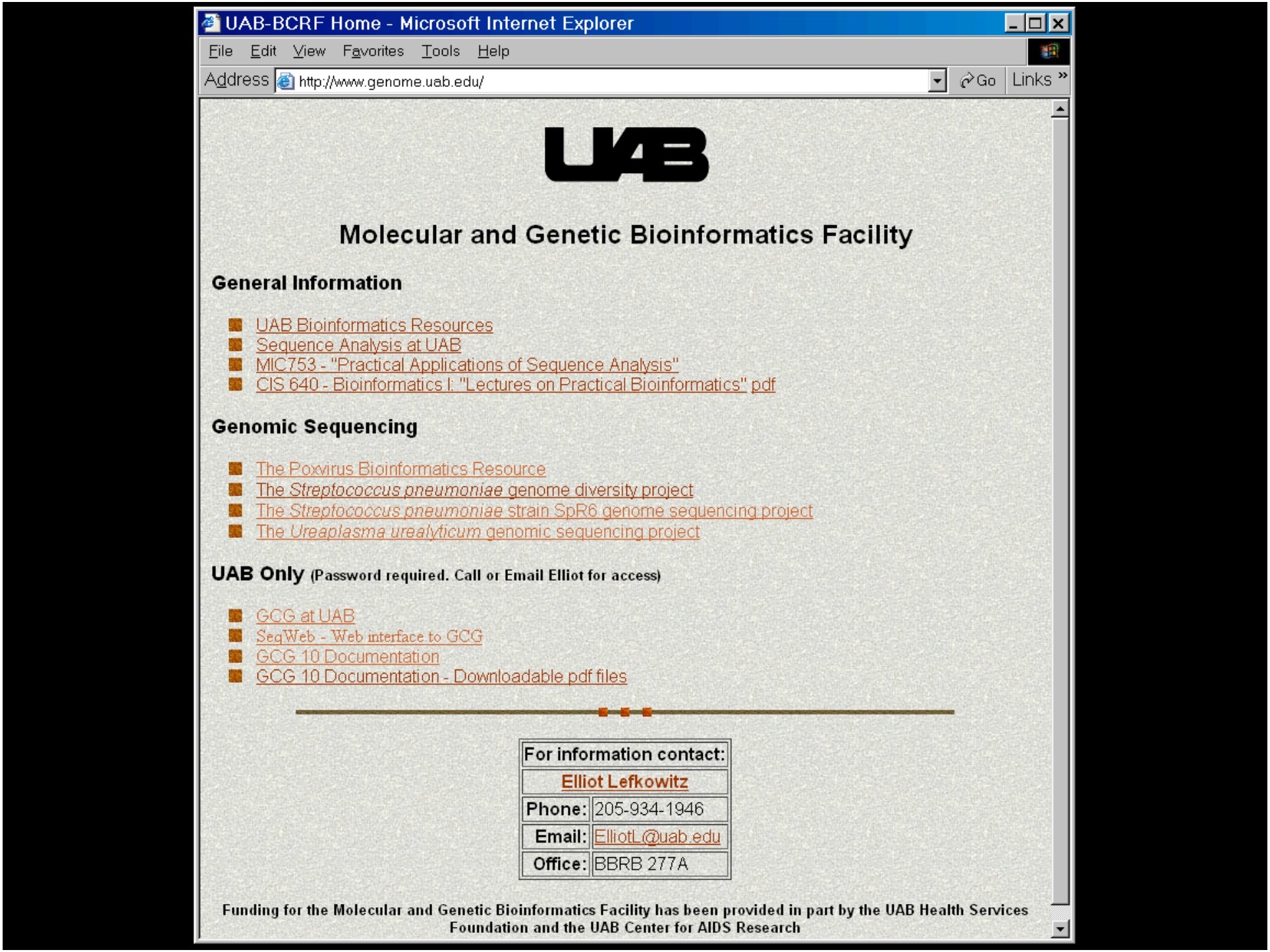
February, 2004

Bioinformatics

Management and Analysis of
Biological Data

Contact Information
Elliot Lefkowitz

- Email
  - ElliotL@uab.edu
- Web Site
  - http://www.genome.uab.edu
- Office
  - BBRB 277A
- Phone
  - 934-1946


UAB-BCRF Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.genome.uab.edu/
Go
Links
UAB
Molecular and Genetic Bioinformatics Facility
General Information
UAB Bioinformatics Resources
Sequence Analysis at UAB
MIC753 - "Practical Applications of Sequence Analysis"
CIS 640 - Bioinformatics I: "Lectures on Practical Bioinformatics" pdf
Genomic Sequencing
The Poxvirus Bioinformatics Resource
The Streptococcus pneumoniae genome diversity project
The Streptococcus pneumoniae strain SpR6 genome sequencing project
The Ureaplasma urealyticum genomic sequencing project
UAB Only (Password required. Call or Email Elliot for access)
GCG at UAB
SeqWeb - Web interface to GCG
GCG 10 Documentation
GCG 10 Documentation - Downloadable pdf files
For information contact:
Elliot Lefkowitz
Phone: 205-934-1946
Email: ElliotL@uab.edu
Office: BBRB 277A
Funding for the Molecular and Genetic Bioinformatics Facility has been provided in part by the UAB Health Services Foundation and the UAB Center for AIDS Research

Lecture Goals

- Defining and understanding the role of Bioinformatics in modern biological sciences
- Becoming familiar with the basic bioinformatic vocabulary
- Providing an overview of biomedical data and databases
- Providing an overview of biomedical analytical tools
- Learning how to discover, access, and utilize information resources

Bioinformatics

- Managing Complexity
  - Technology development
- Enhancing Understanding
  - Research

Managing Complexity

- Data
  - Acquisition
  - Storage
  - Manipulation
  - Retrieval

Managing Complexity…

- Data Analysis
  - Development and Utilization of
    - Analytical tools
    - Visualization tools
- Analyses provides the interpretations necessary for…

Enhancing Understanding

What distinguishes one organism from another?

- Sequence
- Molecular Biology
- Physiology
- Pathogenesis
- Epidemiology
- Evolution

Will the genomic sequence provide an explanation for the differences?

Caveat

- In the end, bioinformatics (a.k.a. computers) can only help in making inferences concerning biological processes
- These inferences (or hypotheses) have to be tested in the laboratory

Genomics

The Human Genome Project

- Mapping and Sequencing the Genomes of Model Organisms
- Data Collection and Distribution
- Ethical, Legal, and Social Considerations
- Research Training
- Technology Development
- Technology Transfer

Genomes of Humans and their "cousins"

- Eukaryotic
- Prokaryotic
- Archaea
- Viruses

April 14, 2003

- The Human Genome Project announces completion of the DNA reference sequence of Homo sapiens.

The Human Genome Sequence

Reference Sequence Properties

Chrom. number	Reference accession	Sequence length	Determined bases*
1	NC_000001.4	245,203,898	218,712,898
2	NC_000002.5	243,315,028	237,043,673
3	NC_000003.5	199,411,731	193,607,218
4	NC_000004.5	191,610,523	186,580,523
5	NC_000005.4	180,967,295	177,524,972
6	NC_000006.5	170,740,541	166,880,540
7	NC_000007.7	158,431,299	154,546,299
8	NC_000008.5	145,908,738	141,694,337
9	NC_000009.5	134,505,819	115,187,714
10	NC_000010.4	135,480,874	130,710,865
11	NC_000011.4	134,978,784	130,709,420
12	NC_000012.5	133,464,434	129,328,332
13	NC_000013.5	114,151,656	95,511,656
14	NC_000014.4	105,311,216	87,191,216
15	NC_000015.4	100,114,055	81,117,055
16	NC_000016.4	89,995,999	79,890,791
17	NC_000017.5	81,691,216	77,480,855
18	NC_000018.4	77,753,510	74,534,531
19	NC_000019.5	63,790,860	55,780,860
20	NC_000020.5	63,644,868	59,424,990
21	NC_000021.3	46,976,537	33,924,742
22	NC_000022.4	49,476,972	34,352,051
X	NC_000023.4	152,634,166	147,686,664
Y	NC_000024.3	50,961,097	22,761,097
unplaced	various	25,263,157	25,062,835

* HGP goals called for determination of only the euchromatic portion of the genome. Telomeres, centromeres, and other heterochromatic regions have been left undetermined, as have a small number of unclonable gaps.

Genome Project Organization

- Cloning
- Mapping
- Sequencing
  - Sequence Assembly
- Annotation
  - Feature identification and prediction
    - Genes, Regulatory regions…
- Analysis

Bioinformatic Information Flow

- "Raw" data generation
  - Sequence generation and assembly
- Analytical tools
  - Pattern matching
- Database generation
  - Construction and data import
- Visualization (publication) of results
  - Static: Table or graph
  - Dynamic: Web page/Java applet

Annotation and Analysis

- Gene prediction
  - Identify patterns characteristic of ORFs
- Functional assignment
  - Similarity searching
- Metabolic pathway modeling
- Comparative analysis
  - Identification and comparison with related genes

What is a gene?

- Does it looks like a gene?
  - Open Reading Frame
  - Base composition
  - Codon usage
- Is it expressed?
  - Regulatory signals
  - Transcription
  - Translation
- When is it expressed?

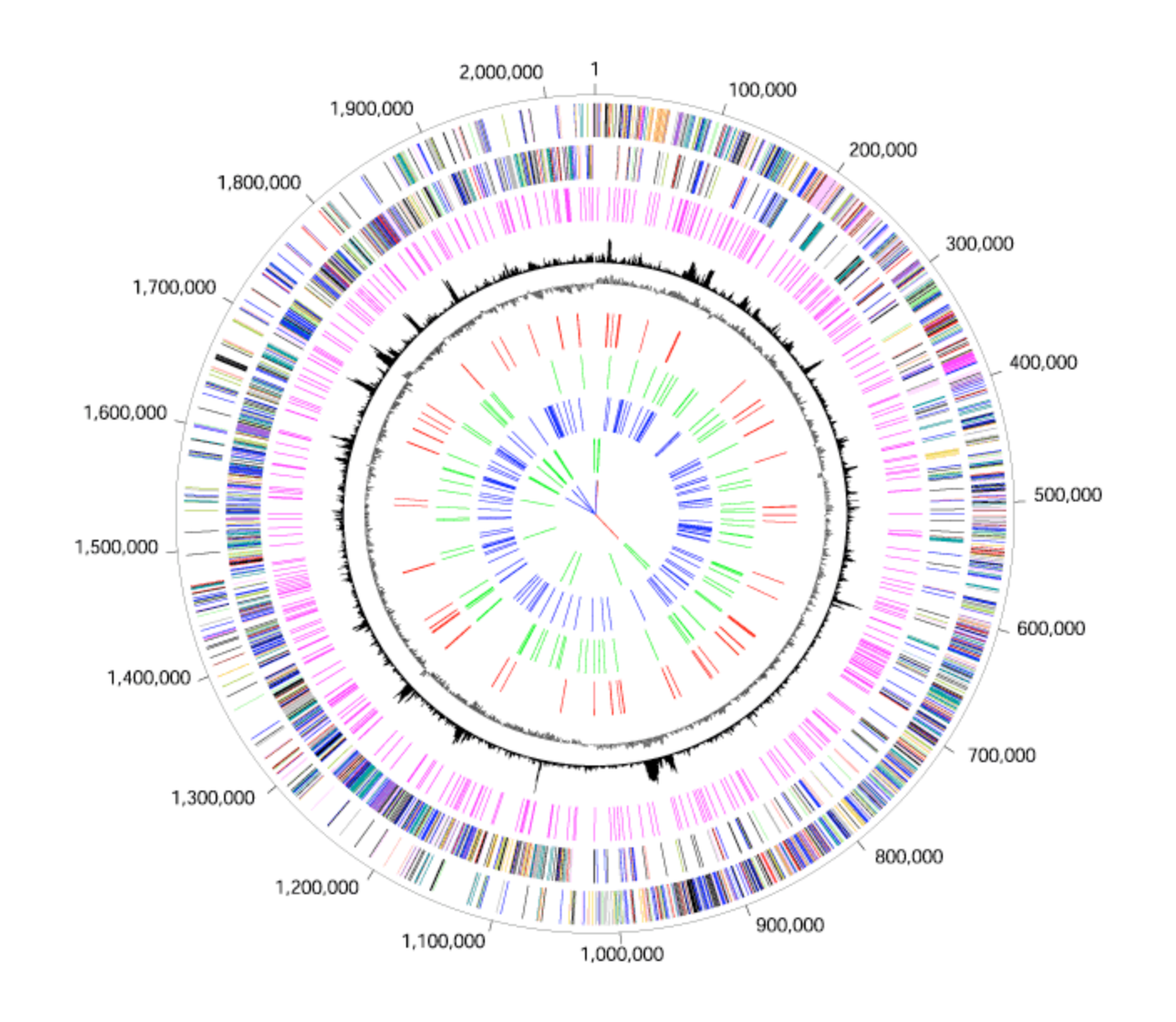
Prokaryotic vs. Eukaryotic Genes

- Prokaryotes
  - small genomes
  - high gene density
  - no introns (or splicing)
  - no RNA processing
  - “simple” promoters
  - terminators important
  - Few overlapping genes
- Eukaryotes
  - large genomes
  - low gene density
  - introns (splicing)
  - RNA processing
  - Complex gene regulation
  - terminators not important
  - polyadenylation

Intrinsic & extrinsic information about gene locations

- Intrinsic information
  - Buried in the primary DNA sequence
    - Open Reading Frame
    - Base composition
- Extrinsic information
  - Evidence inferred from database searching and genomic comparison.
    - BLAST searches
- Laboratory data
  - Expression arrays
  - mRNAs, ESTs

Streptococcus pneumoniae R6 genome



Metabolic Pathway Modeling

- Role assignment
- Metabolic Pathway Reconstruction
  - BioCyc Knowledge Library, Peter Karp, SRI
  - http://biocyc.org/
  - EcoCyc
- Navigation and analysis
- Pathway editing

Pangea Systems EcoCyc version 4.5, Pathway Tools version 4.3

E. coli

Summary of Organisms

Compound Mode
Reaction Mode
Protein Mode
Gene Map Mode
Gene Mode
Pathway Mode
Overview Mode

Backward in History
Forward in History
Select from History

Select Answer
Next Answer

Clone Window
Fix Window
Unfix Window

Preferences
Help
Print to File
Exit

E. coli

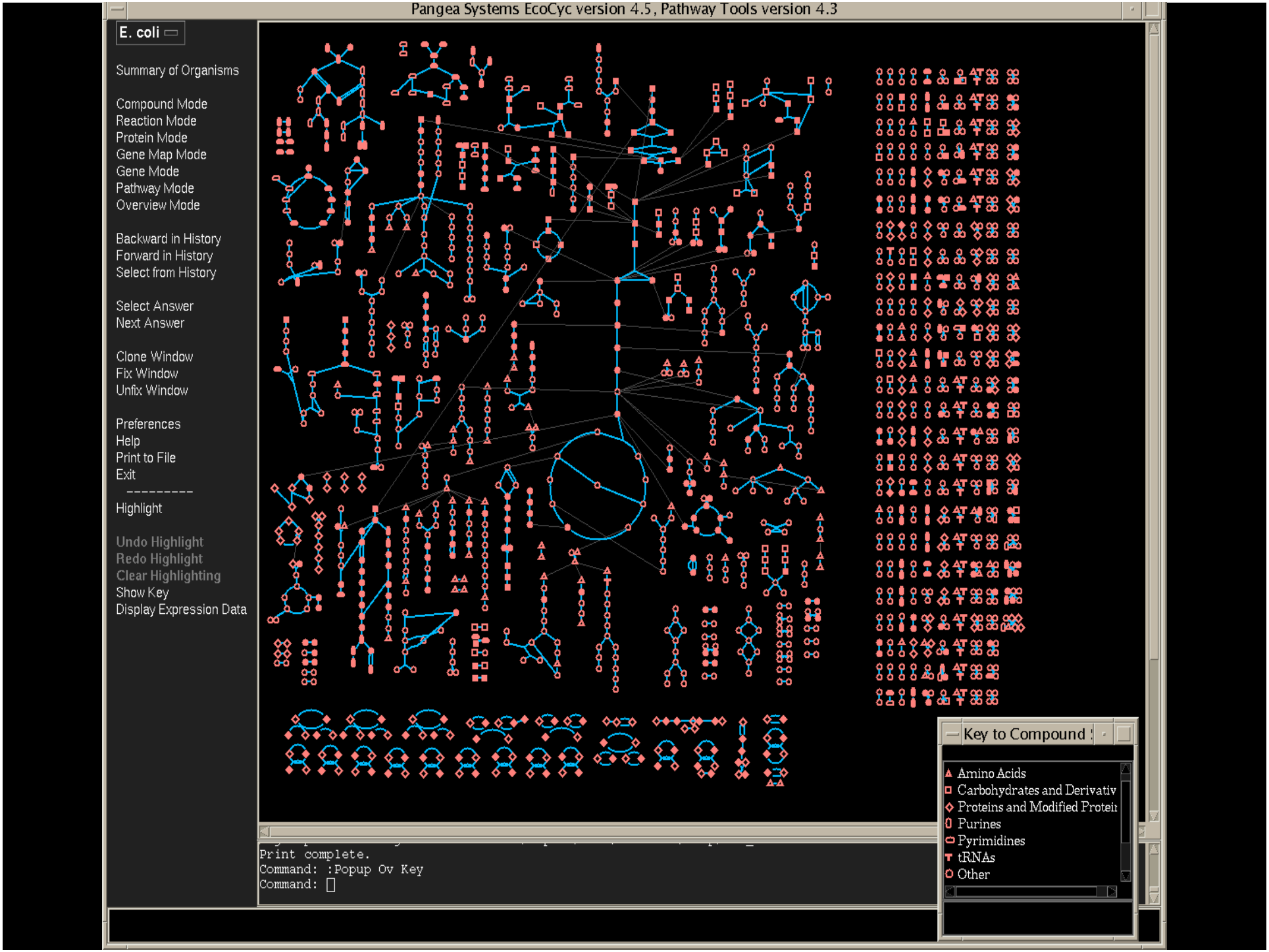
Strain: K-12

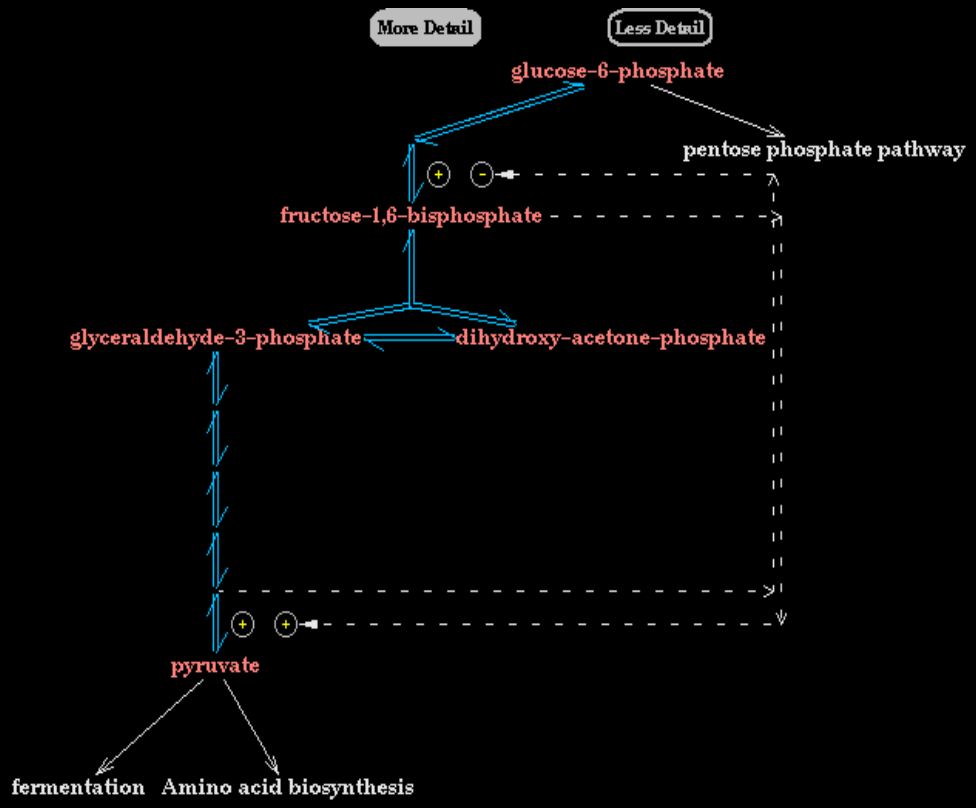
Genetic Element	Genes				Size (bp)	GC %
	Mapped	Protein	RNA	Unidentified ORFs		
K-12 Chromosome	4384	4281	103	1398	4,639,221	50.8

Genes without a physical map position: 286

Pathways:	150
Enzymatic Reactions:	905
Transport Reactions:	0
Polypeptides:	894
Protein Complexes:	498
Enzymes:	655
Transporters:	0
Compounds:	1308
Operons:	0
tRNAs:	79

```
Current organism for command modes is now E. coli
Command:
```


Pangea Systems EcoCyc version 4.5, Pathway Tools version 4.3
E. coli
Summary of Organisms
Compound Mode
Reaction Mode
Protein Mode
Gene Map Mode
Gene Mode
Pathway Mode
Overview Mode
Backward in History
Forward in History
Select from History
Select Answer
Next Answer
Clone Window
Fix Window
Unfix Window
Preferences
Help
Print to File
Exit
Highlight
Undo Highlight
Redo Highlight
Clear Highlighting
Show Key
Display Expression Data
Print complete.
Command: :Popup Ov Key
Command:
Key to Compound
Amino Acids
Carbohydrates and Derivativ
Proteins and Modified Protei
Purines
Pyrimidines
tRNAs
Other



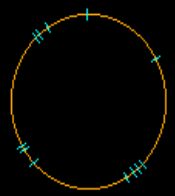
Synonyms: Embden–Meyerhof pathway

Superclasses: **Energy metabolism**

Net reaction equation: Glucose + 2 Pi + 2 ADP + 2 NAD = 2 pyruvate + 2 ATP + 2 NADH + 2 H + 2 H(2)O

Superpathways: **glycolysis+Entner–Doudoroff**, **glycolysis+TCA+glyoxylate bypass**

Locations of Mapped Genes:



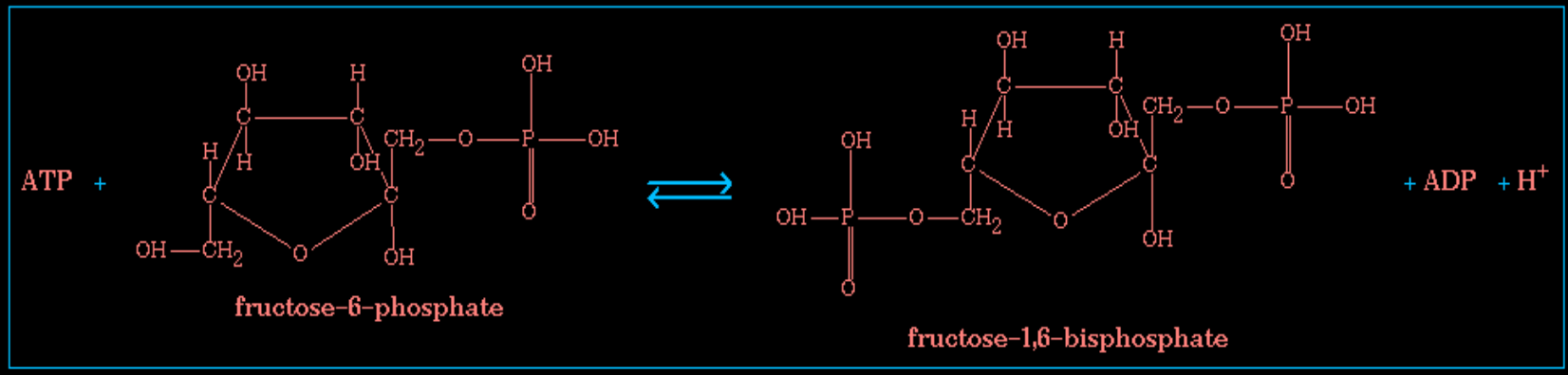
E. coli Reaction: 2.7.1.11

Superclasses: **2.7.1 -- PHOSPHOTRANSFERASES WITH AN ALCOHOL GROUP AS ACCEPTOR**

Enzymes and Genes:

6-phosphofructokinase-1: **pfkA**,
6-phosphofructokinase-2: **pfkB**

In pathway: **mannitol degradation**, **sorbitol degradation**, **glycolysis**



$\Delta G^{0'}$ (kcal/mol): -3.4 [1]

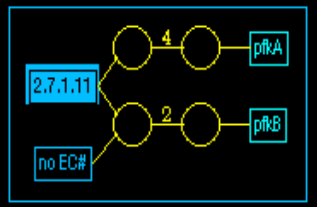
Comment: This is a key control step in glycolysis [2]

This reaction occurs in E. coli.

Citations: [2,1]

Unification Links: **ENZYME.2.7.1.11**

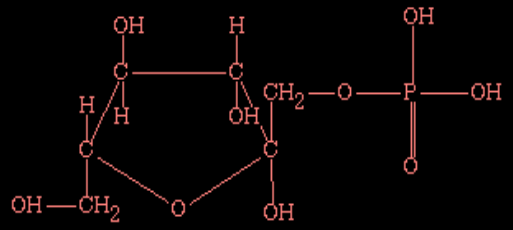
Gene-Reaction Schematic:



Superclasses: Carbohydrate-Derivatives

Empirical formula: $C_6H_{13}O_9P$

Molecular weight: 260.14



In Reactions:

UDP-N-acetylglucosamine biosynthesis:

fructose-6-phosphate + L-glutamine = D-glucosamine-6-phosphate + L-glutamate

colanic acid biosynthesis:

mannose-6-phosphate = fructose-6-phosphate

gluconeogenesis:

fructose-1,6-bisphosphate + H_2O = fructose-6-phosphate + phosphate,

glucose-6-phosphate = fructose-6-phosphate

glucosamine catabolism:

H_2O + D-glucosamine-6-phosphate = NH_3 + fructose-6-phosphate

glycolysis:

glucose-6-phosphate = fructose-6-phosphate ,

ATP + fructose-6-phosphate = fructose-1,6-bisphosphate + ADP + H^+

mannitol degradation:

mannitol-1-phosphate + NAD = fructose-6-phosphate + NADH,

ATP + fructose-6-phosphate = fructose-1,6-bisphosphate + ADP + H^+

mannose and GDP-mannose metabolism:

mannose-6-phosphate = fructose-6-phosphate

mannose catabolism:

mannose-6-phosphate = fructose-6-phosphate

non-oxidative branch of the pentose phosphate pathway:

erythrose-4-phosphate + xylulose-5-phosphate = fructose-6-phosphate + glyceraldehyde-3-phosphate,

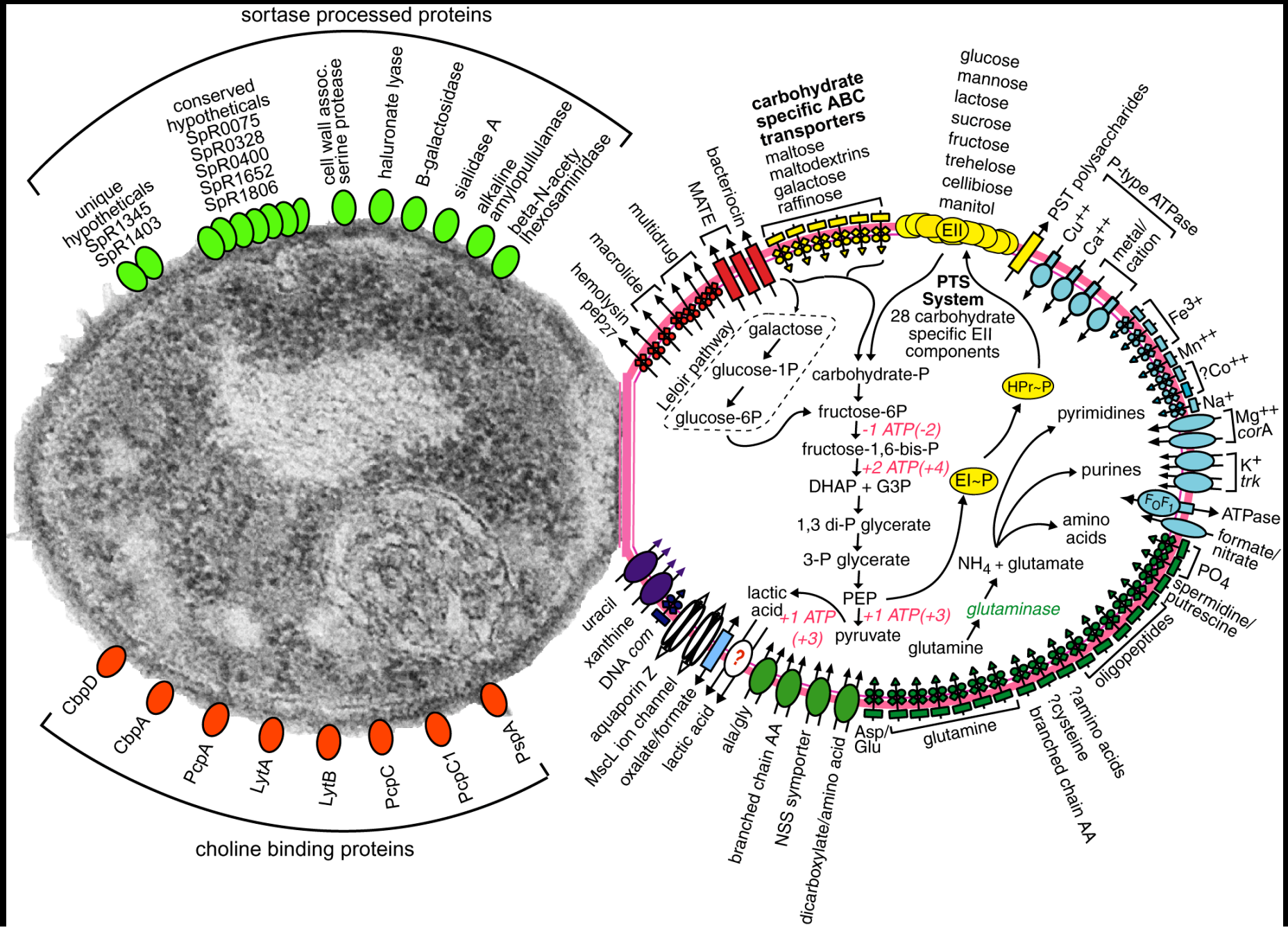
glyceraldehyde-3-phosphate + sedoheptulose-7-phosphate = fructose-6-phosphate + erythrose-4-phosphate

sorbitol degradation:

D-sorbitol 6-phosphate + NAD = fructose-6-phosphate + NADH,

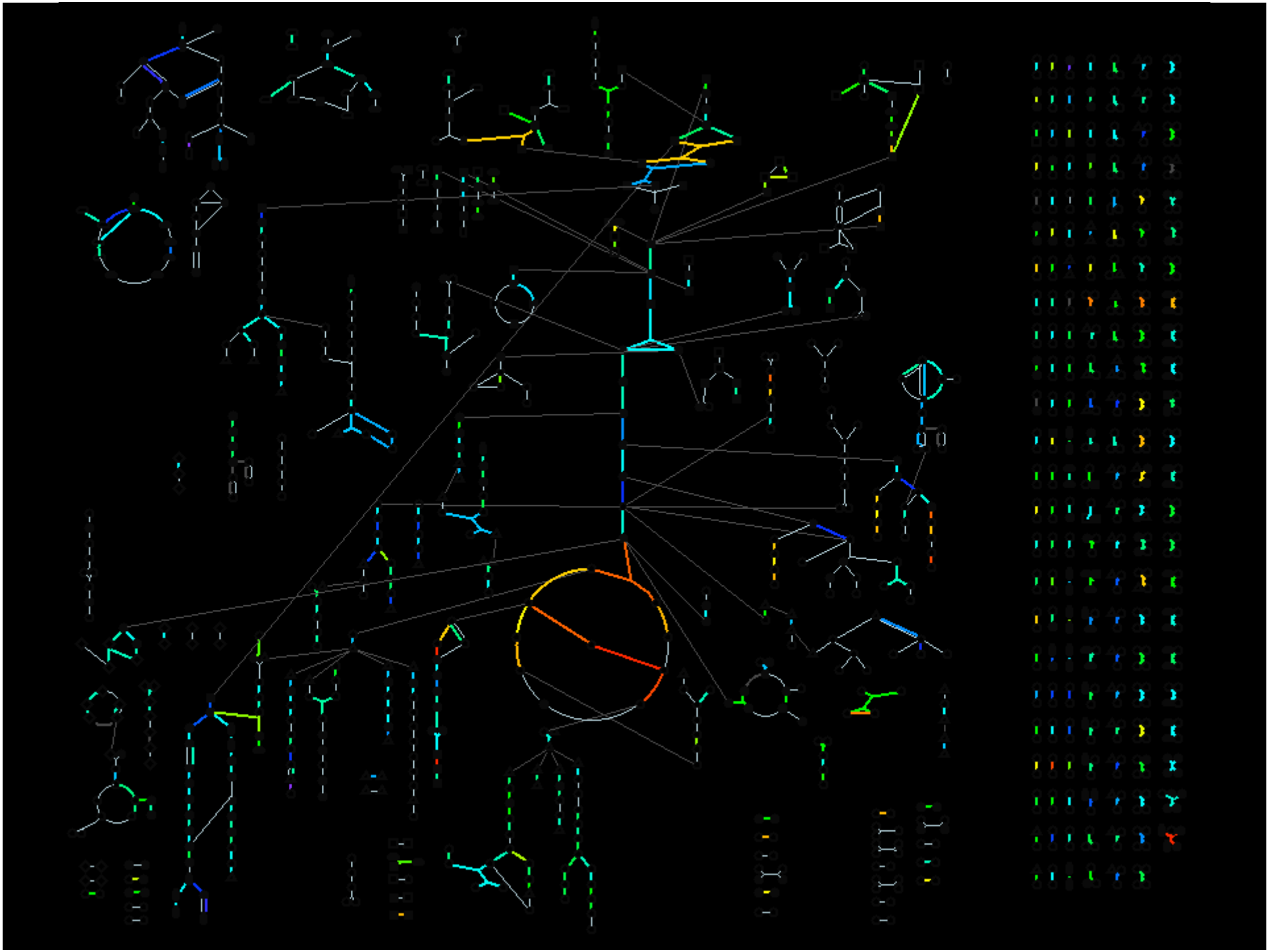
ATP + fructose-6-phosphate = fructose-1,6-bisphosphate + ADP + H^+

Streptococcus pneumoniae R6 metabolism



Expression Arrays

- Cell growth in different environments
- Isolate cDNAs
- Measure expression using array technology
- Create database of expression information
- Display information in an easy-to-use format
  - Show ratio of expression under different conditions



Comparative Genomics

- Identification of similarities
  - Primary sequence
  - Structure
  - Function
- Identification of differences
  - Gene complement
  - Genotypic differences resulting in phenotypic changes
- Phylogenetic inference
  - Predicting evolutionary history

Comparative Genomics

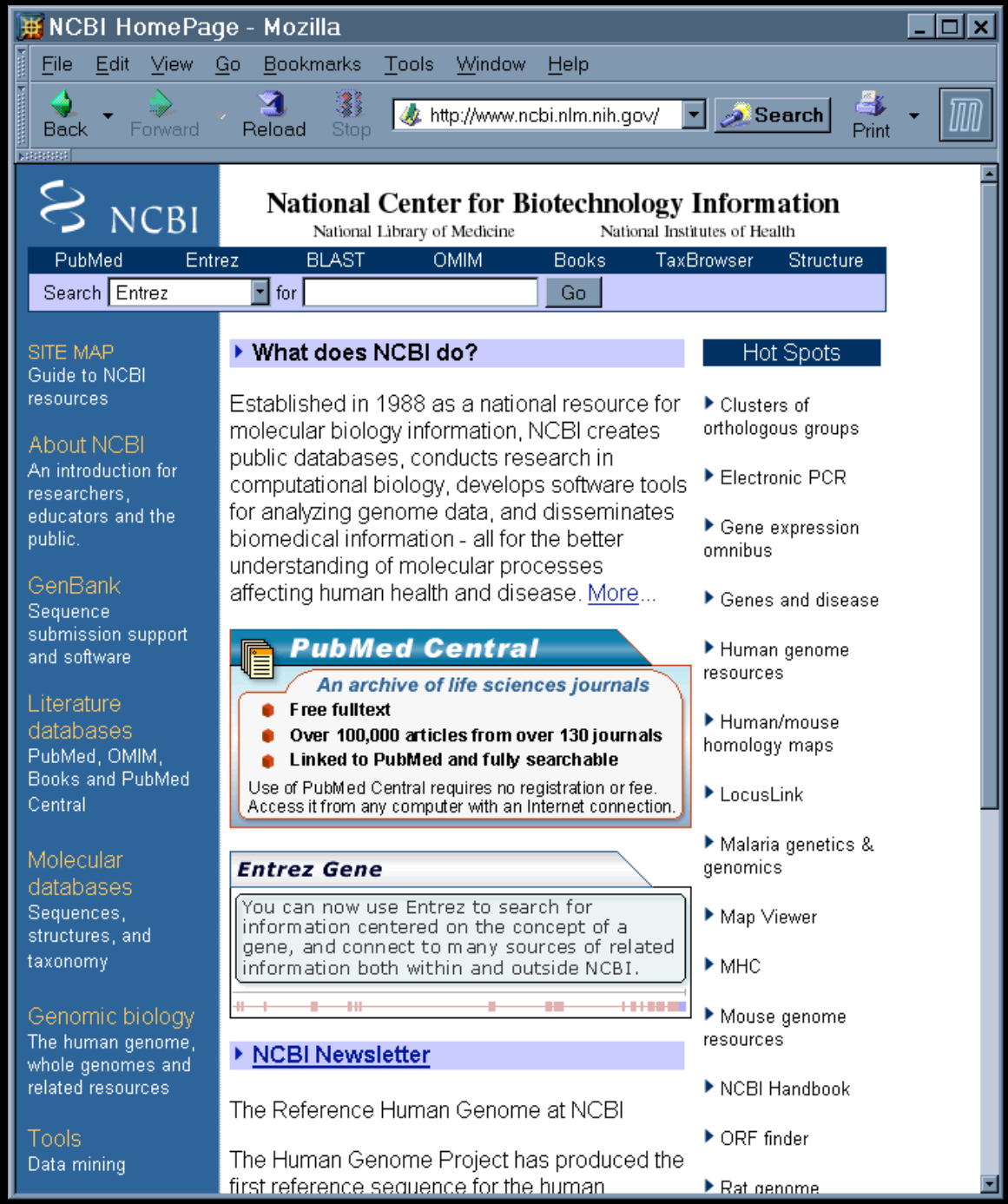
- "Similar" sequences
  - Sequences related by primary sequence similarity
- Homologs
  - Sequences related by evolution
  - Orthologs
    - Related due to speciation
  - Paralogs
    - Related due to gene duplication

Biological Information

Access and Analysis

Information Resources

- NCBI – Databases, tools, links
  - National Center for Biotechnology Information
  - http://www.ncbi.nih.gov/
- General Protein Analysis Tools
  - http://us.expasy.org/


NCBI HomePage - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://www.ncbi.nlm.nih.gov/
Search
Print
NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed Entrez BLAST OMIM Books TaxBrowser Structure
Search Entrez for Go
SITE MAP
Guide to NCBI resources
About NCBI
An introduction for researchers, educators and the public.
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books and PubMed Central
Molecular databases
Sequences, structures, and taxonomy
Genomic biology
The human genome, whole genomes and related resources
Tools
Data mining
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
PubMed Central
An archive of life sciences journals
Free fulltext
Over 100,000 articles from over 130 journals
Linked to PubMed and fully searchable
Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.
Entrez Gene
You can now use Entrez to search for information centered on the concept of a gene, and connect to many sources of related information both within and outside NCBI.
NCBI Newsletter
The Reference Human Genome at NCBI
The Human Genome Project has produced the first reference sequence for the human
Hot Spots
Clusters of orthologous groups
Electronic PCR
Gene expression omnibus
Genes and disease
Human genome resources
Human/mouse homology maps
LocusLink
Malaria genetics & genomics
Map Viewer
MHC
Mouse genome resources
NCBI Handbook
ORF finder
Rat genome

ExPASy Molecular Biology Server - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://us.expasy.org/ Search Print

Site Map | Search ExPASy | Contact us

Hosted by NCSC US | Mirror sites: Bolivia Canada China Korea Switzerland Taiwan

Search Swiss-Prot/TrEMBL for Go Clear

ExPASy Molecular Biology Server

The ExPASy (**Expert Protein Analysis System**) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).

ExPASy celebrates 10 years of continuing service!
What do you like best on ExPASy, what do you like least?

[Announcements] [Job opening] [Mirror Sites]

Databases

- Swiss-Prot and TrEMBL - Protein knowledgebase
- PROSITE - Protein families and domains
- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- ENZYME - Enzyme nomenclature
- SWISS-3DIMAGE - 3D images of proteins and other biological macromolecules
- SWISS-MODEL Repository - Automatically generated protein models
- CD40Lbase - CD40 ligand defects
- SeqAnalRef - Sequence analysis bibliographic references

- Links to many other molecular biology databases

Tools and software packages

- Proteomics and sequence analysis tools
  - Proteomics [PeptIdent, PeptideMass, ...]
  - DNA -> Protein [Translate]
  - Similarity searches [BLAST]
  - Pattern and profile searches [ScanProsite]
  - Post-translational modification and topology prediction
  - Primary structure analysis [ProtParam, pI/MW, ProtScale]
  - Secondary and tertiary structure prediction [SWISS-MODEL, Swiss-PdbViewer]
  - Alignment [T-COFFEE, SIM]
  - Biological text analysis
- Melanie 4 - Software for 2-D PAGE analysis
- Roche Applied Science's Biochemical Pathways

Education and services

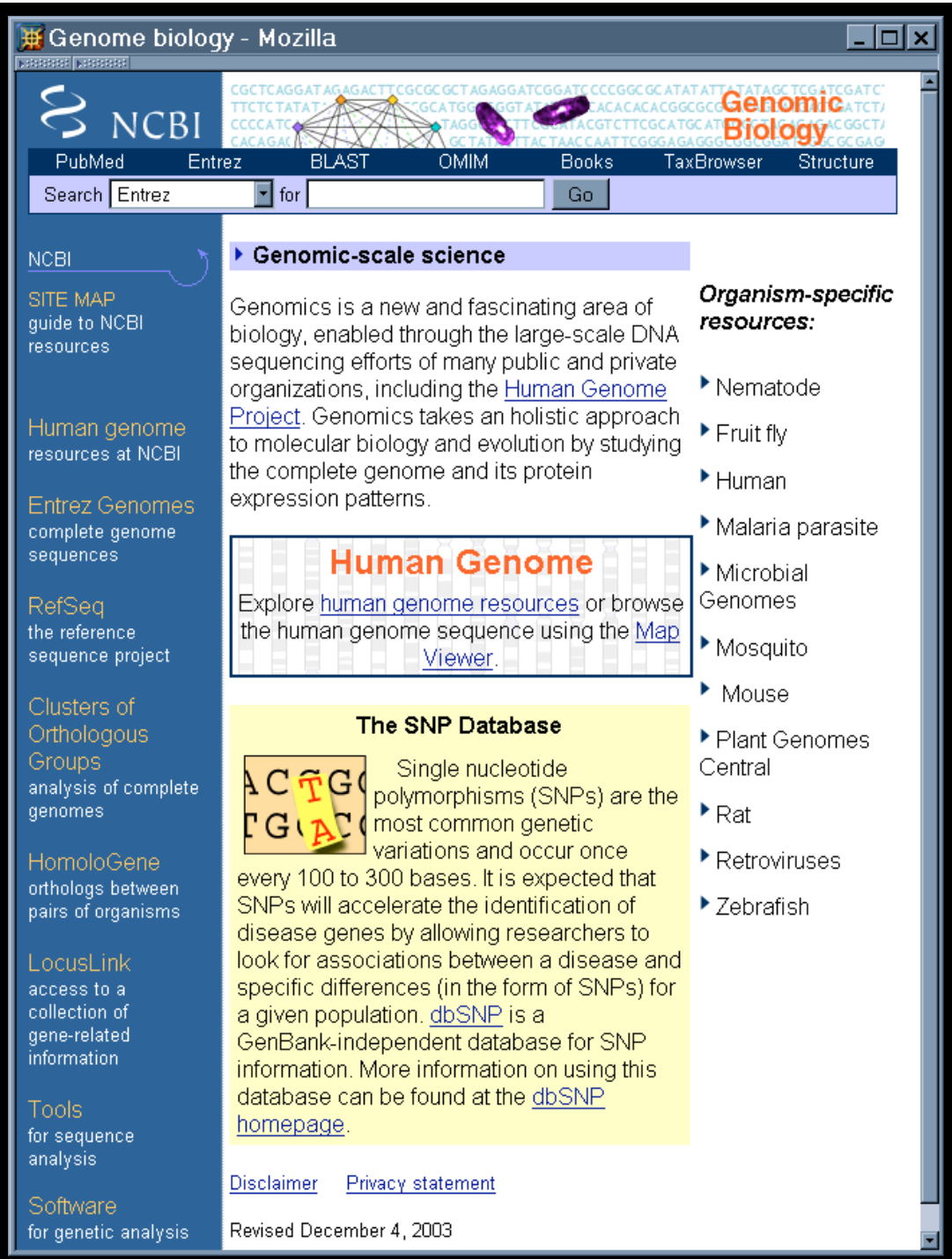
- The ExPASy FTP server

Documentation

- What's New on ExPASy

Genomic Resources

- NCBI Genome Resources
  - http://www.ncbi.nih.gov/Genomes/
- Ensembl Genome Server
  - www.ensembl.org
- UCSC Genome Browser
  - genome.ucsc.edu


Genome biology - Mozilla
NCBI
Genomic Biology
PubMed
Entrez
BLAST
OMIM
Books
TaxBrowser
Structure
Search
Entrez
for
Go
NCBI
SITE MAP
guide to NCBI resources
Human genome
resources at NCBI
Entrez Genomes
complete genome sequences
RefSeq
the reference sequence project
Clusters of Orthologous Groups
analysis of complete genomes
HomoloGene
orthologs between pairs of organisms
LocusLink
access to a collection of gene-related information
Tools
for sequence analysis
Software
for genetic analysis
Genomic-scale science
Genomics is a new and fascinating area of biology, enabled through the large-scale DNA sequencing efforts of many public and private organizations, including the Human Genome Project. Genomics takes an holistic approach to molecular biology and evolution by studying the complete genome and its protein expression patterns.
Human Genome
Explore human genome resources or browse the human genome sequence using the Map Viewer.
The SNP Database
Single nucleotide polymorphisms (SNPs) are the most common genetic variations and occur once every 100 to 300 bases. It is expected that SNPs will accelerate the identification of disease genes by allowing researchers to look for associations between a disease and specific differences (in the form of SNPs) for a given population. dbSNP is a GenBank-independent database for SNP information. More information on using this database can be found at the dbSNP homepage.
Disclaimer
Privacy statement
Revised December 4, 2003
Organism-specific resources:
Nematode
Fruit fly
Human
Malaria parasite
Microbial Genomes
Mosquito
Mouse
Plant Genomes Central
Rat
Retroviruses
Zebrafish

Ensembl Genome Browser - Mozilla

project Ensembl

The Wellcome Trust Sanger Institute

EBI

Ensembl Genome Browser

Search Ensembl

Search all species for Anything with Lookup

About Ensembl

Ensembl is a joint project between EMBL - EBI and the Sanger Institute to develop a software system which produces and maintains automatic annotation on eukaryotic genomes. Ensembl is primarily funded by the Wellcome Trust.

Access to all the data produced by the project, and to the software used to analyse and present it, is provided free and without constraints. Some data and software may be subject to third-party constraints [details].

Ensembl presents up-to-date sequence data and the best possible annotation for metazoan genomes. Available now are human, mouse, rat, fugu, zebrafish, mosquito, Drosophila, C. elegans, and C. briggsae, Others will be added soon.

For an introduction to the Ensembl project, take the Ensembl tour, and then go through a step-by-step worked example which introduces Ensembl's main functions. For more information read these short papers (Jan 2002, Jan 2003), in Nucleic Acids Research.

For all enquiries, please contact the Ensembl *HelpDesk* (helpdesk@ensembl.org).

Ensembl provides ....

- Easy access to sequence data
- For known genes, predicted structure and location in the genome sequence
- Prediction of novel genes, all with supporting evidence
- Annotation of other features of the genome
- Targeted connections to other genome resources worldwide

Easy access to the data via ....

- A web-based genome browser (which can be customized as required)
- A web-based system for data export and data mining
- 'Dumps' of sequence and other data sets for you to download
- Direct access to the databases
- A Perl-based object layer

Ensembl Species

Human	v. 18.34.1	4 Nov 2003
Mouse	v. 18.30.1	6 May 2003
Rat	v. 18.3.1	4 Nov 2003
Zebrafish	v. 18.2.1	2 Jul 2003
Fugu	v. 18.2.1	3 Mar 2003
Mosquito	v. 18.2a.1	1 Oct 2003
Fruitfly	v. 18.3a.1	2 Jul 2003
C. elegans	v. 18.102.1	2 Jul 2003
C. briggsae	v. 18.25.1	3 Mar 2003

Similarity searches (multi-species) BLAST/SSAHA

Fast data/sequence retrieval (multi-species) EnsMart

Access to whole genome shotgun data (includes additional species) Trace Server

Help and documentation

- Species-specific documentation is available via the species home pages above.
- Take the Ensembl tour, go through a step-by-step worked example, or read this short paper in Nucleic Acids Research.
- For context-sensitve help on any web page click: Help
- There is also an index of context-sensitive help pages, and a set of guided How do I....? trails.

Recent Ensembl news News

Display your own data in Ensembl DAS

Apollo genome browser Apollo

Questions or suggestions? Try the Help Desk

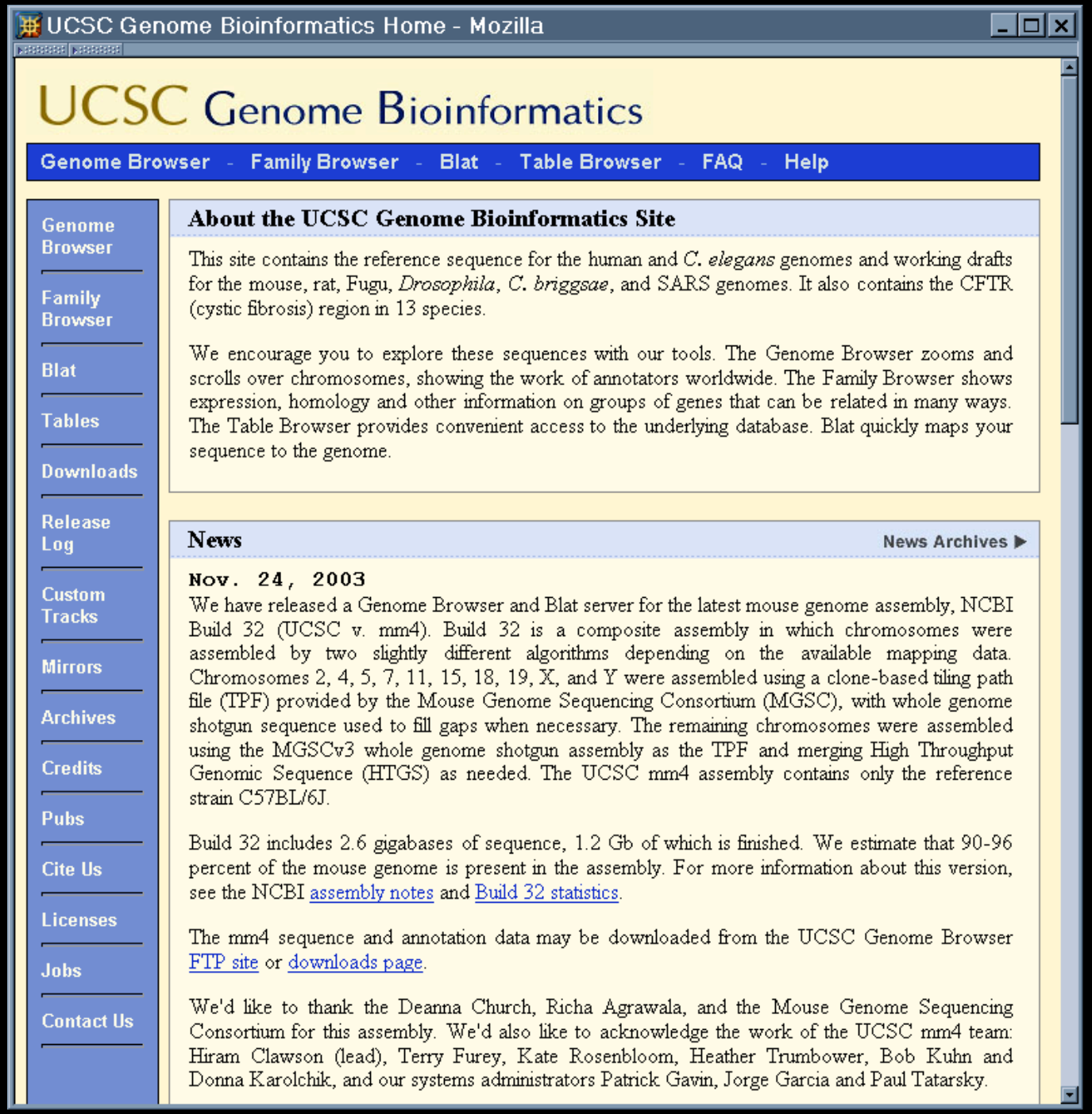
Documentation (includes tutorial on direct data access & instructions for installing Ensembl on your own site) Documentation

Have you tried ....?

Fly

Fly*(Drosophila melanogaster)* with data imported from FlyBase is now available at Ensembl

Click for more information


UCSC Genome Bioinformatics Home - Mozilla
UCSC Genome Bioinformatics
Genome Browser - Family Browser - Blat - Table Browser - FAQ - Help
Genome Browser
Family Browser
Blat
Tables
Downloads
Release Log
Custom Tracks
Mirrors
Archives
Credits
Pubs
Cite Us
Licenses
Jobs
Contact Us
About the UCSC Genome Bioinformatics Site
This site contains the reference sequence for the human and C. elegans genomes and working drafts for the mouse, rat, Fugu, Drosophila, C. briggsae, and SARS genomes. It also contains the CFTR (cystic fibrosis) region in 13 species.
We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Family Browser shows expression, homology and other information on groups of genes that can be related in many ways. The Table Browser provides convenient access to the underlying database. Blat quickly maps your sequence to the genome.
News
News Archives ▶
Nov. 24, 2003
We have released a Genome Browser and Blat server for the latest mouse genome assembly, NCBI Build 32 (UCSC v. mm4). Build 32 is a composite assembly in which chromosomes were assembled by two slightly different algorithms depending on the available mapping data. Chromosomes 2, 4, 5, 7, 11, 15, 18, 19, X, and Y were assembled using a clone-based tiling path file (TPF) provided by the Mouse Genome Sequencing Consortium (MGSC), with whole genome shotgun sequence used to fill gaps when necessary. The remaining chromosomes were assembled using the MGSCv3 whole genome shotgun assembly as the TPF and merging High Throughput Genomic Sequence (HTGS) as needed. The UCSC mm4 assembly contains only the reference strain C57BL/6J.
Build 32 includes 2.6 gigabases of sequence, 1.2 Gb of which is finished. We estimate that 90-96 percent of the mouse genome is present in the assembly. For more information about this version, see the NCBI assembly notes and Build 32 statistics.
The mm4 sequence and annotation data may be downloaded from the UCSC Genome Browser FTP site or downloads page.
We'd like to thank the Deanna Church, Richa Agrawala, and the Mouse Genome Sequencing Consortium for this assembly. We'd also like to acknowledge the work of the UCSC mm4 team: Hiram Clawson (lead), Terry Furey, Kate Rosenbloom, Heather Trumbower, Bob Kuhn and Donna Karolchik, and our systems administrators Patrick Gavin, Jorge Garcia and Paul Tatarsky.

Bioinformatic Databases

Something to compare against

Major Sequence Databases

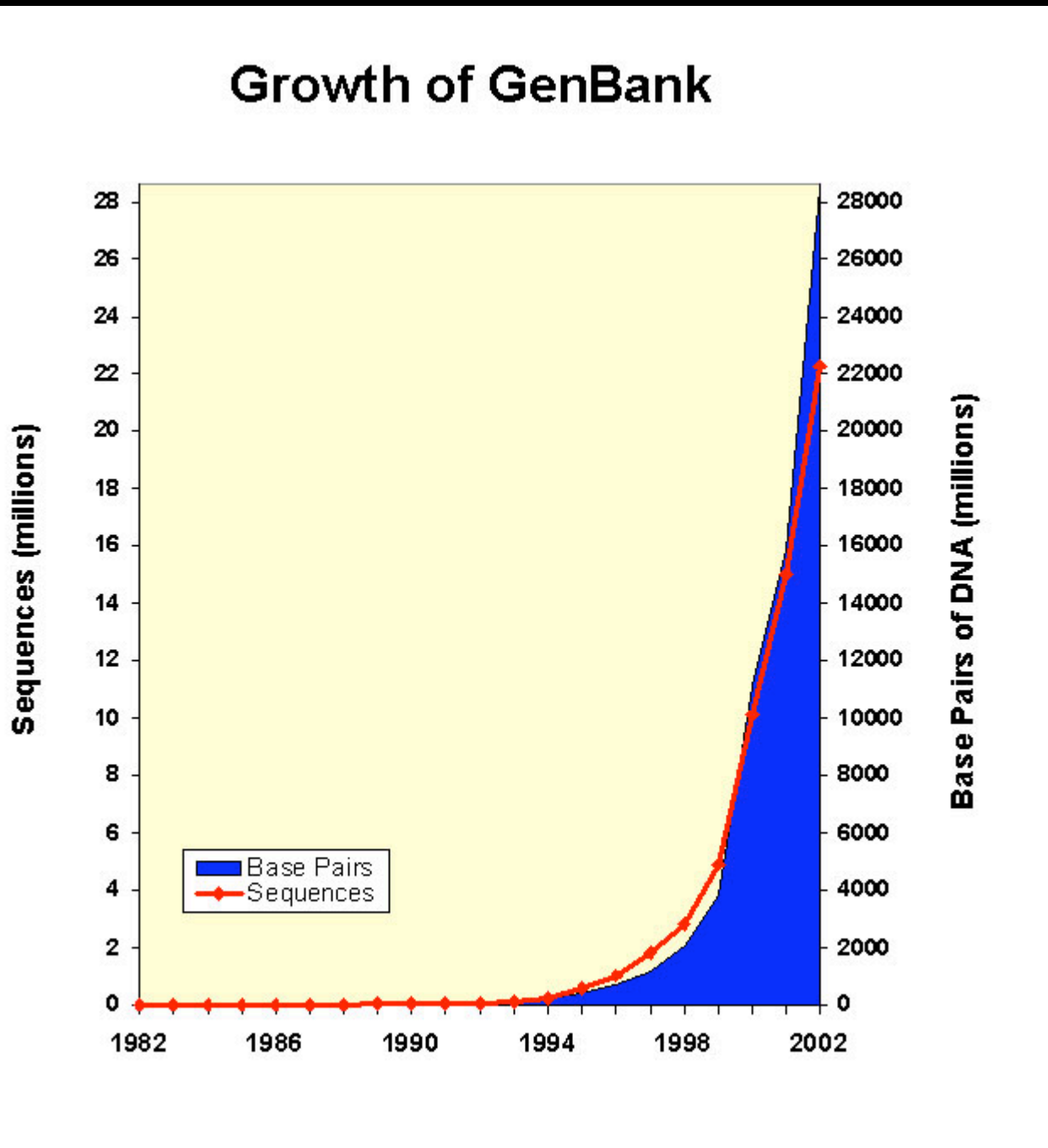
- DNA
  - Genbank (NCBI)
  - EMBL
  - DDBJ
- Protein
  - PIR
  - Swiss-Prot
  - Swiss-Prot TrEMBL
  - UniProt

Other Databases

- Structural
  - Protein Data Bank (PDB): http://www.rcsb.org/pdb/
- Expression
  - Microarray Gene Expression Data Society (MGED): http://www.mged.org/
  - Gene Expression Omnibus (GEO – NCBI)
- Proteomic
  - Mascot: http://www.matrixscience.com/
- Metabolism
  - BioCyc: http://biocyc.org/
- Ontology
  - Gene Ontology (GO) Consortium: http://www.geneontology.org/
  - Controlled vocabulary for the description of biological processes

Genbank

- Primary nucleic acid sequence database
- Maintained by NCBI
  - National Center for Biotechnology Information
  - http://www.ncbi.nlm.nih.gov
- December 15, 2003; Release 139
  - 36,553,368,485 bases
  - 30,968,418 sequences


Growth of GenBank
Sequences (millions)
Base Pairs of DNA (millions)
Base Pairs
Sequences
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
22000
24000
26000
28000
1982
1986
1990
1994
1998
2002

Some GenBank Divisions

- EST: Expressed Sequence Tags
  - “Single-pass" cDNA sequences
  - Generally representative of the 3’ ends of cDNAs
  - More “full-length” ESTs now available
- STS: Sequence Tagged Sites
  - Sequence and mapping data
  - Short genomic landmark sequences
- HTGS: High Throughput Genomic Sequences
  - “Unfinished” DNA sequences generated by the high-throughput sequencing centers

Other NCBI Databases

- RefSeq
- Unigene
- HomoloGene
- Genomic
- SNPs

RefSeq

- NCBI Reference Sequence project
- Provides reference sequence standards for the naturally occurring molecules from chromosomes to mRNAs to proteins
- Stable reference point for:
  - mutation analysis
  - gene expression studies
  - polymorphism discovery
- Accession numbers have two letters, an underscore, and six numbers
  - NM_123456

Unigene

- GenBank sequences partitioned into a non-redundant set of gene-oriented clusters
  - Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.
- Includes EST and complete cDNA sequences
- Provides information on differentially-spliced transcripts

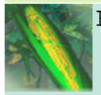
Unigene Organisms

Vertebrata	
Mammalia	
Bos taurus (cow)	12,808 entries
Homo sapiens (human)	128,826 entries
Mus musculus (mouse)	90,444 entries
Rattus norvegicus (rat)	63,253 entries
Sus scrofa (pig)	14,344 entries
Aves	
Gallus gallus (chicken)	5,068 entries
Amphibia	
Xenopus laevis (frog)	19,512 entries
Actinopterygii	
Danio rerio (zebrafish)	16,355 entries

Urochordata	
Ascidiacea	
Ciona intestinalis (sea squirt)	13,674 entries

Arthropoda	
Insecta	
Anopheles gambiae (malaria mosquito)	3,270 entries
Drosophila melanogaster (fruit fly)	14,779 entries

Nematoda	
Chromadorea	
Caenorhabditis elegans	20,137 entries



Embryophyta	
Eudicotyledons	
Arabidopsis thaliana (thale cress)	27,141 entries
Glycine max (soybean)	8,987 entries
Lycopersicon esculentum (tomato)	3,740 entries
Medicago truncatula (barrel medic)	5,729 entries
Liliopsida	
Hordeum vulgare (barley)	7,944 entries
Oryza sativa (rice)	19,223 entries
Triticum aestivum (wheat)	20,454 entries
Zea mays (maize)	13,512 entries

Chlorophyta	
Chlorophycaea	
Chlamydomonas reinhardtii	6,448 entries

HomoloGene

- Curated and calculated orthologs and homologs for genes represented in UniGene and LocusLink

Genomic DBs

- Human
- Mouse
- Rat
- Zebrafish
- Drosophila
- Nematode
- Plant genomes
- Yeast
- Malaria
- Microbial genomes
- Viruses
- Viroids
- Plasmids
- Eukaryotic organelles

dbSNP

- Single Nucleotide Polymorphisms
  - Single base changes
  - Small-scale insertions/deletions
  - Polymorphic repetitive elements
  - Microsatellite variation

LocusLink

- Provides a single query interface to curated sequence and descriptive information about genetic loci
  - Nomenclature
  - Aliases
  - Sequence accessions
  - Phenotypes
  - EC numbers
  - OMIM numbers
  - UniGene clusters
  - Homology
  - Map locations
  - Web sites

OMIM

- Online Mendelian Inheritance in Man
- Database of gene-linked genetic disorders
- Maintained at Johns Hopkins University
  - Dr. Victor A. McKusick
- Provides link to GeneTests
  - Laboratories that provide testing for specific genetic disorders

Sample OMIM Queries
(From the OMIM Help Docs)

- What human genes are related to hypertension? Which of those genes are on chromosome 17?
- List the OMIM entries that describe genes on chromosome 10.
- List the OMIM entries that contain information about allelic variants.
- Retrieve the OMIM record for the cystic fibrosis transmembrane conductance regulator (CFTR), and link to related protein sequence records via Entrez.
- Find the OMIM record for the p53 tumor protein, and link out to related information in LocusLink and the p53 Mutation Database.

EMBL and DDBJ

- European Molecular Biology Laboratory
  - Hinxton, UK
  - http://www.ebi.ac.uk/
- DNA Data Bank of Japan
  - Mishima, Japan
  - http://www.ddbj.nig.ac.jp/

Coordination with Genbank

- Prevents duplication
- Genbank enters sequences from U.S. journals and researchers
- EMBL handles European data
- DDBJ handles Asian data
- Data exchanged daily

The Sequence Record

- Different for each database
- Locus (Name)
- Accession Number
- Keywords
- Description
- Properties
- References
- The Sequence

GenBank Sample Record

- http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html

```
analyze% typedata ge:humcftrm
!!NA_SEQUENCE 1.0
LOCUS       HUMCFTRM     6129 bp    mRNA            PRI       15-DEC-1989
DEFINITION  Human cystic fibrosis mRNA, encoding a presumed transmembrane
            conductance regulator (CFTR).
ACCESSION   M28668
NID         g180331
KEYWORDS    cystic fibrosis; transmembrane conductance regulator.
SOURCE      Human, cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6129)
  AUTHORS   Riordan,J.R., Rommens,J.M., Kerem,B., Alon,N., Rozmahel,R.,
            Grzelczak,Z., Zielenski,J., Lok,S., Plavsic,N., Chou,J.-L.,
            Drumm,M.L., Iannuzzi,M.C., Collins,F.S. and Tsui,L.-C.
  TITLE     Identification of the cystic fibrosis gene: Cloning and
            characterization of complementary DNA
  JOURNAL   Science 245, 1066-1073 (1989)
  MEDLINE   89368940
```

Accession Numbers

- Each sequence submitted to a database is assigned a unique primary accession number
- Accession numbers do not change
- If a sequence is merged with another, a new accession number is assigned, and the original number becomes a secondary accession number
- Accession numbers may include version numbers
  - AO2428.2

```
COMMENT     A three base-pair deletion spanning positions 1654-1656 is observed
            in cDNAs from cystic fibrosis patients.
FEATURES             Location/Qualifiers
     source          1. .6129
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             133. .4575
                     /note="cystic fibrosis transmembrane conductance
                     regulator"
                     /codon_start=1
                     /db_xref="PID:g180332"
                     /translation="MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVD
                     SADNLSEKLEREWDRELASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLL
                     LNRFSKDIAILDDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLR
                     AYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLWTLRAFGRQPYFETLFHKALNLHTAN
                     WFLYLSTLRWFQMRIEMIFVIFFIAVTFISILTTGEGEGRVGIILTLAMNIMSTLQWA
                     VNSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKVMIIENSHVKKDDIW
                     PSGGQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRLLN
                     TEGEIQIDGVSWDSITLQQWRKAFGVIPQKVFIFSGTFRKNLDPYEQWSDQEIWKVAD
                     EVGLRSVIEQFPGKLDFVLVDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDP
                     VTYQIIRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSL
                     FRQAISPSDRVKLFPHRNSSKCKSKPQIAALKEETEEEVQDTRL"
BASE COUNT     1886 a   1181 c   1330 g   1732 t
ORIGIN
```

```
HUMCFTRM  Length: 6129  April 13, 1998 13:00  Type: N  Check: 6781  ..

      1  AATTGGAAGC AAATGACATC ACAGCAGGTC AGAGAAAAAG GGTTGAGCGG

     51  CAGGCACCCA GAGTAGTAGG TCTTTGGCAT TAGGAGCTTG AGCCCAGACG

    101  GCCCTAGCAG GGACCCCAGC GCCCGAGAGA CCATGCAGAG GTCGCCTCTG

    151  GAAAAGGCCA GCGTTGTCTC CAAACTTTTT TTCAGCTGGA CCAGACCAAT

    201  TTTGAGGAAA GGATACAGAC AGCGCCTGGA ATTGTCAGAC ATATACCAAA

    251  TCCCTTCTGT TGATTCTGCT GACAATCTAT CTGAAAAATT GGAAAGAGAA

    301  TGGGATAGAG AGCTGGCTTC AAAGAAAAAT CCTAAACTCA TTAATGCCCT

    351  TCGGCGATGT TTTTTCTGGA GATTTATGTT CTATGGAATC TTTTTATATT

    401  TAGGGGAAGT CACCAAAGCA GTACAGCCTC TCTTACTGGG AAGAATCATA

    451  GCTTCCTATG ACCCGGATAA CAAGGAGGAA CGCTCTATCG CGATTTATCT
```

Swiss-Prot

- http://www.expasy.ch/sprot/
- Protein Database
- University of Geneva
- Arranged by protein function
- Release 42.9
- February 2, 2004
- 53,044,352 amino acids 143,790 entries
- Provides annotated protein records

Swiss-Prot TrEMBL

- Translation of all EMBL Nucleic Acid coding sequences not yet present in Swiss-Prot
- Allows rapid availability without immediate annotation
- Release 25.9
- February 2, 2004
- 1,075,779 entries

PIR

- http://pir.georgetown.edu/
- Protein Identification Resource
  - PIR-International Protein Sequence Database (PSD)
- National Biomedical Research Foundation
- Georgetown University
- Release 78.03, November 24, 2003
- 283,366 Entries

PIR-NREF

- Non-redundant REFerence protein database
- Current Release 1.4
- February 2, 2004
- 1,485,025 Entries

iProClass Database - PIR

http://pir.georgetown.edu/iproclass/

- Comprehensive family relationships and structural/functional classifications and features of proteins
  - Superfamilies
  - Families
  - Domains

UniProt (United Protein Databases)

- Unified, coordinated database of protein information
- Integration of SwissProt, TrEMBL, and PIR
- http://www.uniprot.org/

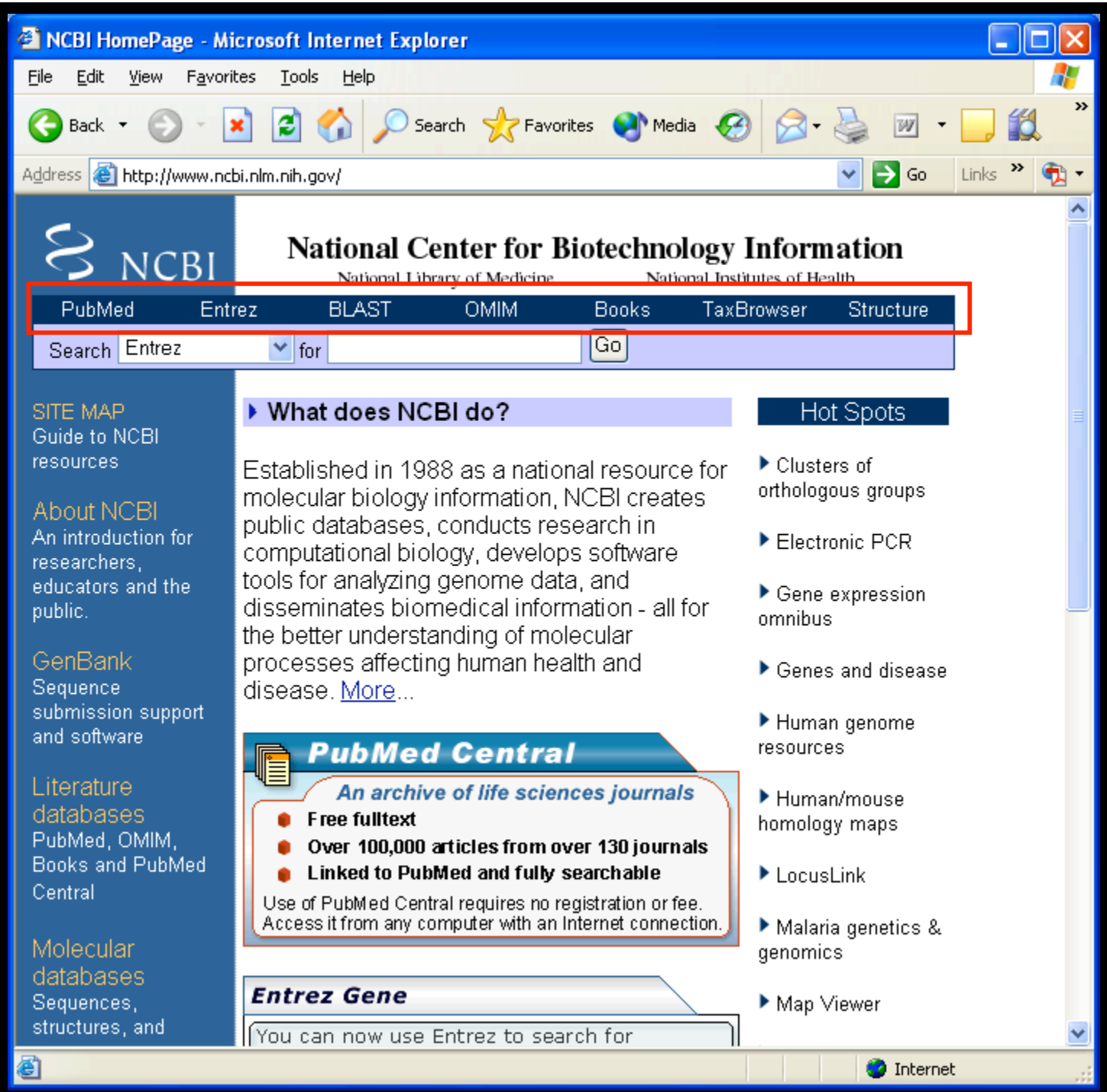
UniProt databases

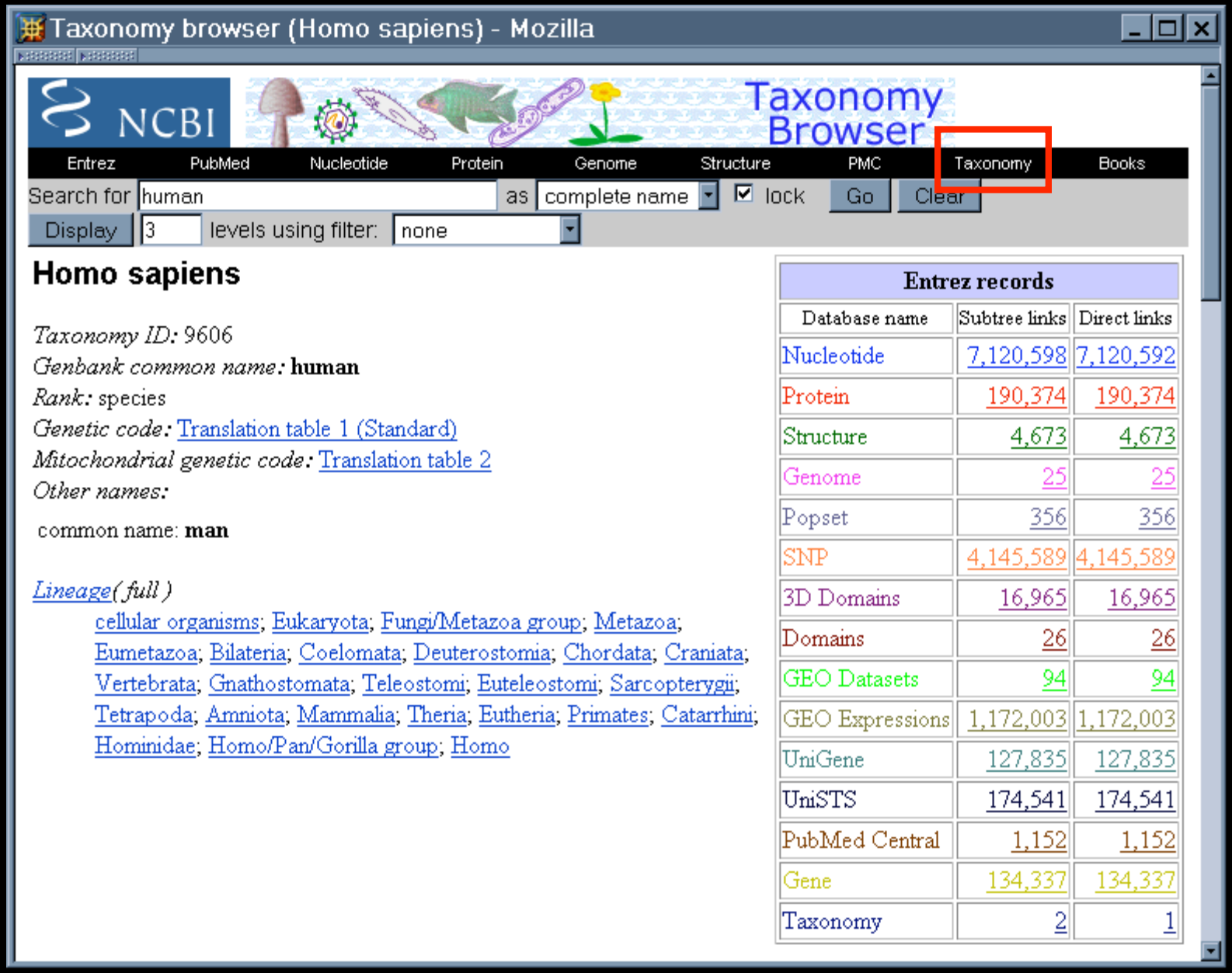
- The UniProt Archive (UniParc) provides a stable, comprehensive sequence collection without redundant sequences
- The UniProt Knowledgebase (UniProt) provides the central database of protein sequences with accurate, consistent, rich sequence and functional annotation.
- The UniProt Non-redundant Reference (UniRef) databases provide condensed data collections based on the UniProt knowledgebase in order to obtain complete coverage of sequence space at several resolutions.

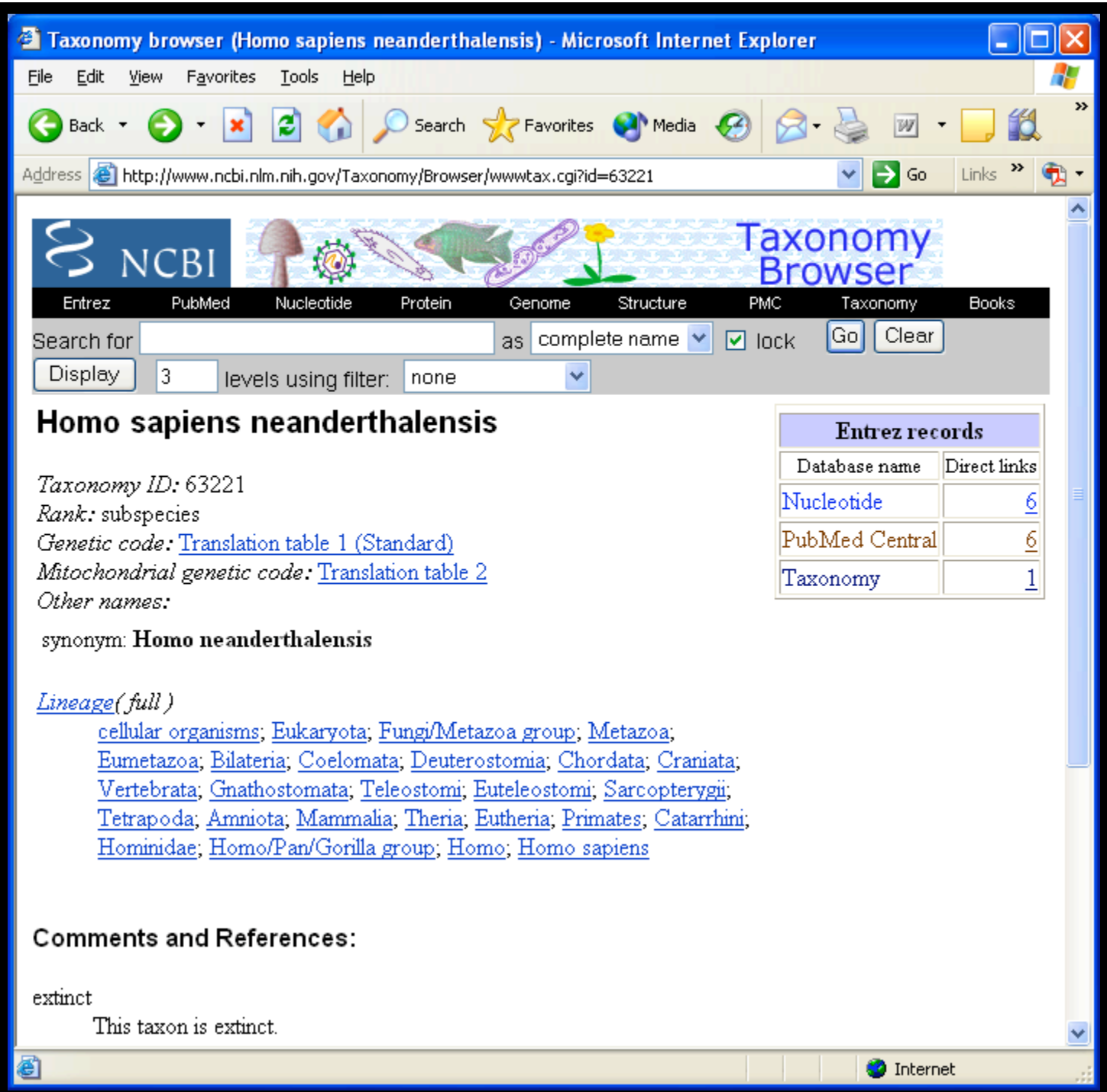
Searching for Information

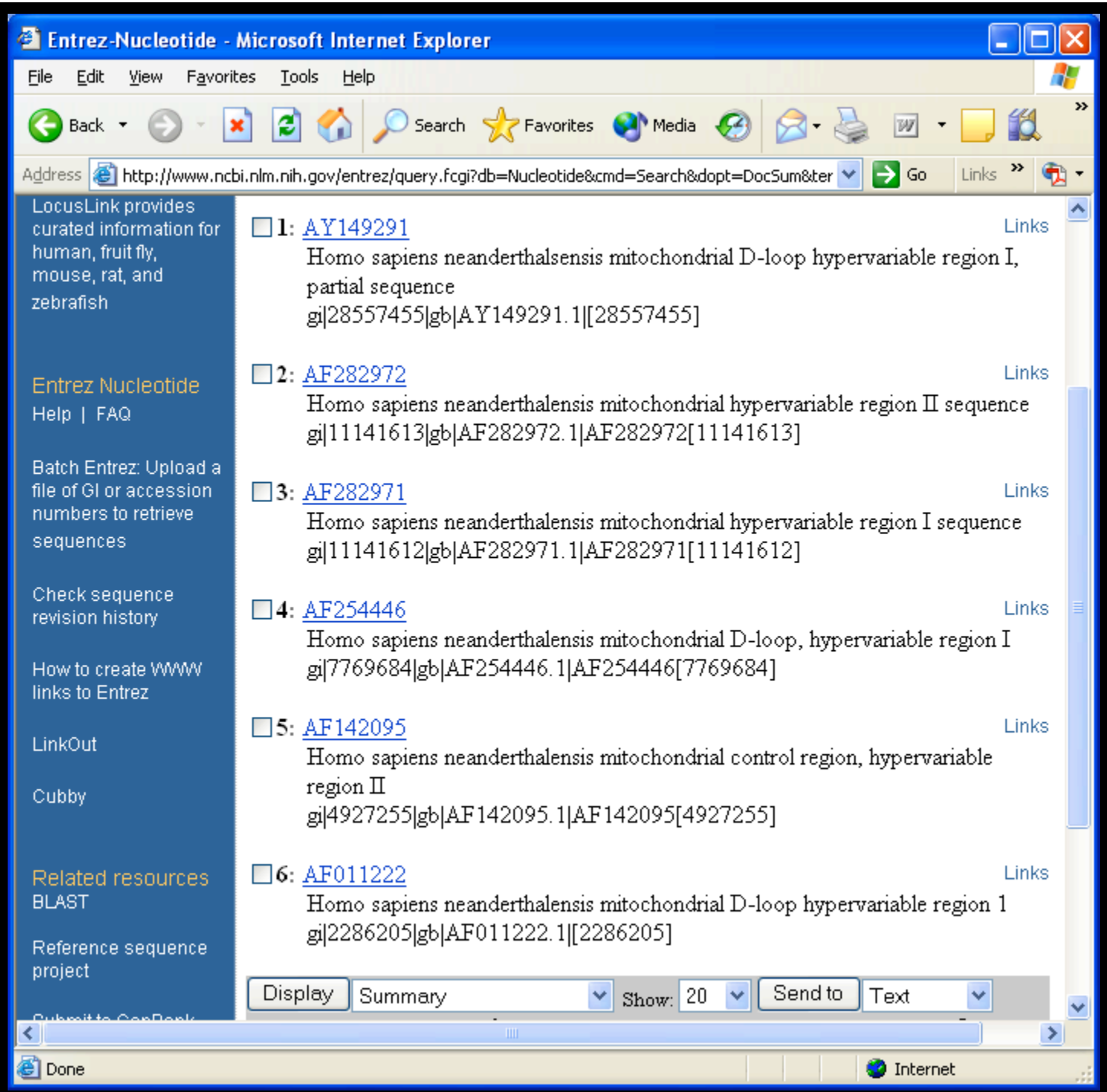
Information Searching at NCBI

- Publications
  - PubMed
- Sequences
  - Entrez
- Structures
  - PDB
- Taxonomy
- …


NCBI HomePage - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/
Go Links
NCBI
National Center for Biotechnology Information
National Library of Medicine National Institutes of Health
PubMed Entrez BLAST OMIM Books TaxBrowser Structure
Search Entrez for Go
SITE MAP
Guide to NCBI resources
About NCBI
An introduction for researchers, educators and the public.
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books and PubMed Central
Molecular databases
Sequences, structures, and
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
PubMed Central
An archive of life sciences journals
Free fulltext
Over 100,000 articles from over 130 journals
Linked to PubMed and fully searchable
Use of PubMed Central requires no registration or fee. Access it from any computer with an Internet connection.
Entrez Gene
You can now use Entrez to search for
Hot Spots
Clusters of orthologous groups
Electronic PCR
Gene expression omnibus
Genes and disease
Human genome resources
Human/mouse homology maps
LocusLink
Malaria genetics & genomics
Map Viewer
Internet


Taxonomy browser (Homo sapiens) - Mozilla
NCBI
Taxonomy Browser
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search for human as complete name lock Go Clear
Display 3 levels using filter: none
Homo sapiens
Taxonomy ID: 9606
Genbank common name: human
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2
Other names:
common name: man
Lineage (full)
cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Primates; Catarrhini; Hominidae; Homo/Pan/Gorilla group; Homo
Entrez records
Database name | Subtree links | Direct links
Nucleotide | 7,120,598 | 7,120,592
Protein | 190,374 | 190,374
Structure | 4,673 | 4,673
Genome | 25 | 25
Popset | 356 | 356
SNP | 4,145,589 | 4,145,589
3D Domains | 16,965 | 16,965
Domains | 26 | 26
GEO Datasets | 94 | 94
GEO Expressions | 1,172,003 | 1,172,003
UniGene | 127,835 | 127,835
UniSTS | 174,541 | 174,541
PubMed Central | 1,152 | 1,152
Gene | 134,337 | 134,337
Taxonomy | 2 | 1

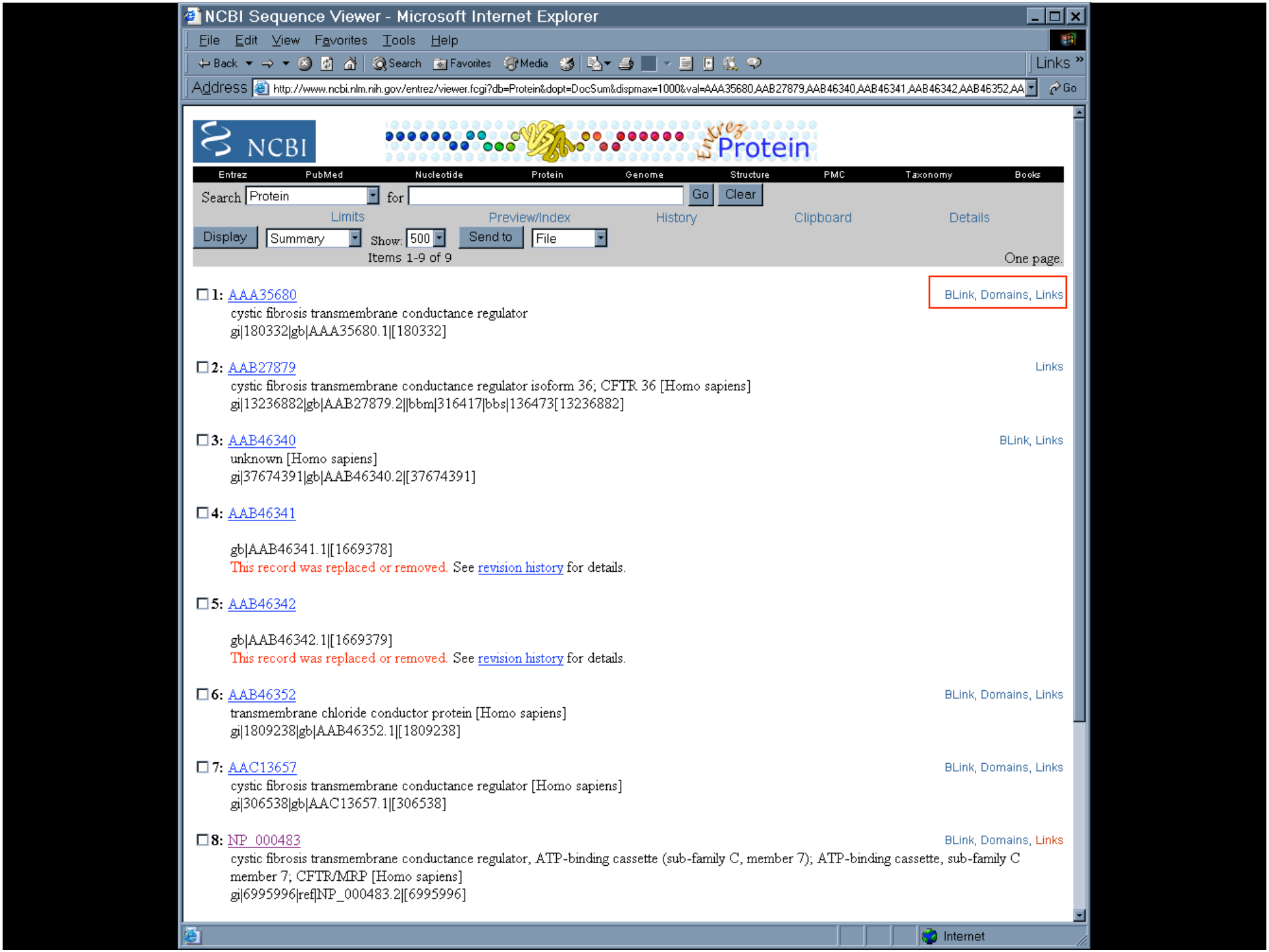

Taxonomy browser (Homo sapiens neanderthalensis) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=63221
Go Links
NCBI
Taxonomy Browser
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search for as complete name lock Go Clear
Display 3 levels using filter: none
Homo sapiens neanderthalensis
Taxonomy ID: 63221
Rank: subspecies
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2
Other names:
synonym: Homo neanderthalensis
Lineage (full)
cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Primates; Catarrhini; Hominidae; Homo/Pan/Gorilla group; Homo; Homo sapiens
Entrez records
Database name
Direct links
Nucleotide 6
PubMed Central 6
Taxonomy 1
Comments and References:
extinct
This taxon is extinct.
Internet


Entrez-Nucleotide - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide&cmd=Search&dopt=DocSum&ter
Go Links
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish
Entrez Nucleotide
Help | FAQ
Batch Entrez: Upload a file of GI or accession numbers to retrieve sequences
Check sequence revision history
How to create WWW links to Entrez
LinkOut
Cubby
Related resources
BLAST
Reference sequence project
1: AY149291 Links
Homo sapiens neanderthalsensis mitochondrial D-loop hypervariable region I, partial sequence
gi|28557455|gb|AY149291.1|[28557455]
2: AF282972 Links
Homo sapiens neanderthalensis mitochondrial hypervariable region II sequence
gi|11141613|gb|AF282972.1|AF282972[11141613]
3: AF282971 Links
Homo sapiens neanderthalensis mitochondrial hypervariable region I sequence
gi|11141612|gb|AF282971.1|AF282971[11141612]
4: AF254446 Links
Homo sapiens neanderthalensis mitochondrial D-loop, hypervariable region I
gi|7769684|gb|AF254446.1|AF254446[7769684]
5: AF142095 Links
Homo sapiens neanderthalensis mitochondrial control region, hypervariable region II
gi|4927255|gb|AF142095.1|AF142095[4927255]
6: AF011222 Links
Homo sapiens neanderthalensis mitochondrial D-loop hypervariable region 1
gi|2286205|gb|AF011222.1|[2286205]
Display Summary Show: 20 Send to Text
Done
Internet

Entrez Searching

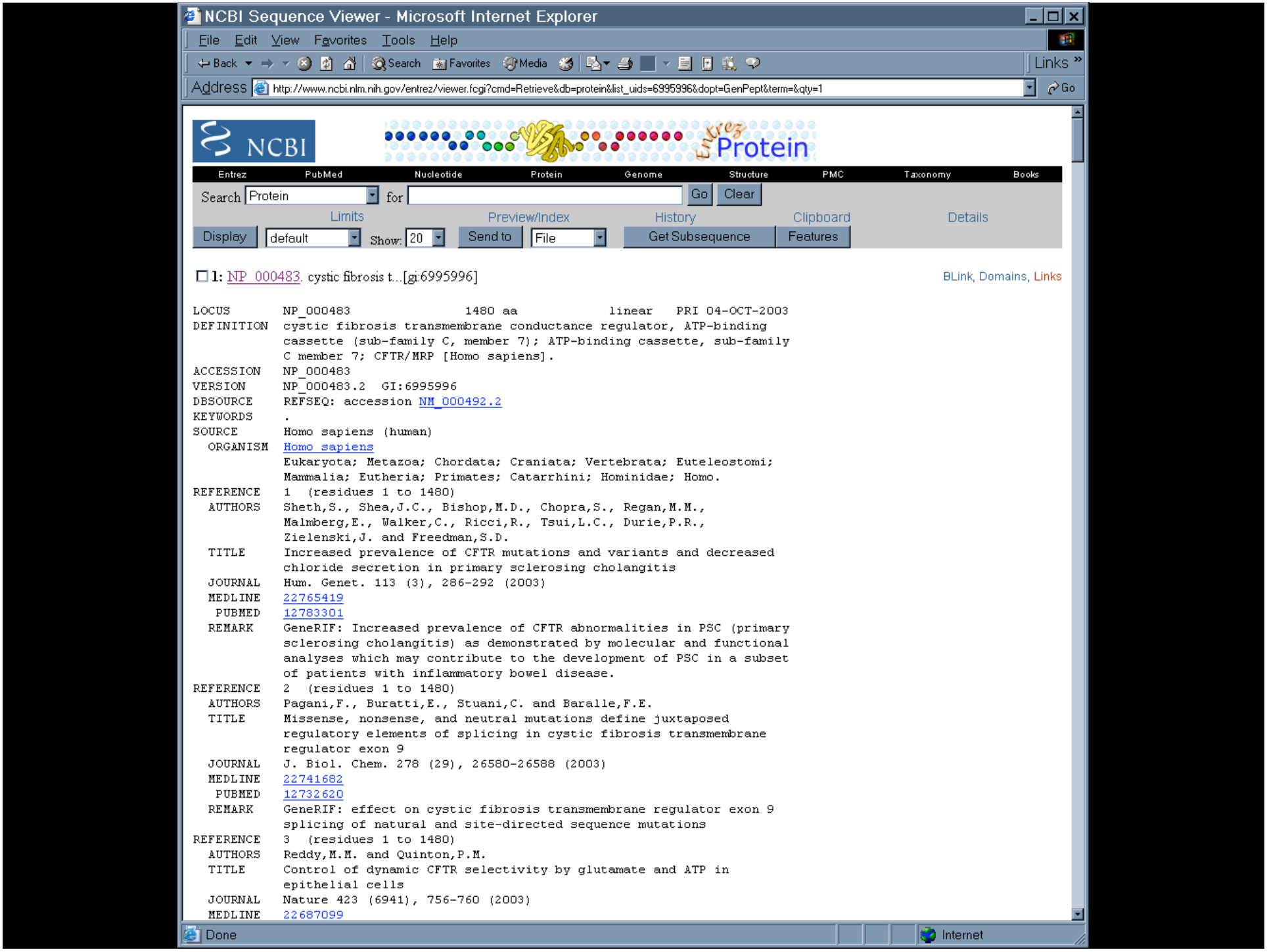
- http://www.ncbi.nlm.nih.gov/entrez/
- Search via text patterns
- Cross-database search interface
  - Sequence
  - PubMed
  - OMIM
  - Linkage information
  - …

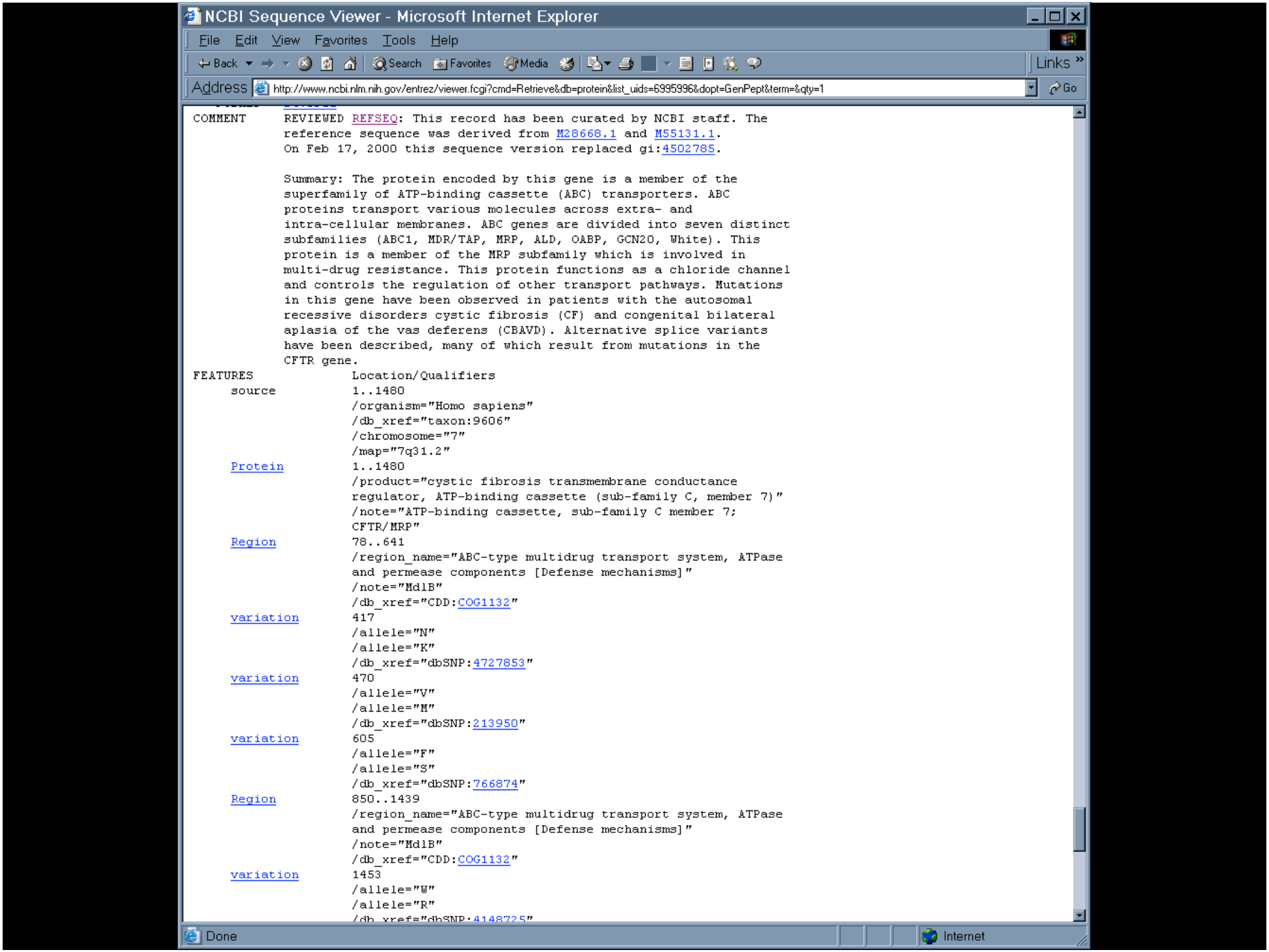

Entrez cross-database search - Mozilla
NCBI
Entrez, The Life Sciences Search Engine
HOME
SEARCH
SITE MAP
PubMed
Entrez
Human Genome
GenBank
Map Viewer
BLAST
Search across databases
cystic fibrosis
GO
CLEAR
Help
22766 PubMed: biomedical literature citations and abstracts
1331 PubMed Central: free, full text journal articles
none Journals: detailed information about journals in Entrez
3 MeSH: detailed information about NLM's controlled vocabulary
88 Books: online books
90 OMIM: online Mendelian Inheritance in Man
35 Site Search: NCBI web and FTP sites
11918 Nucleotide: sequence database (GenBank)
573 Protein: sequence database
3 Genome: whole genome sequences
8 Structure: three-dimensional macromolecular structures
none Taxonomy: organisms in GenBank
702 SNP: single nucleotide polymorphism
10 UniGene: gene-oriented clusters of transcript sequences
2 CDD: conserved protein domain database
11 3D Domains: domains from Entrez Structure
78 UniSTS: markers and mapping data
4 PopSet: population study data sets
135 GEO: expression and molecular abundance profiles
none GEO DataSets: experimental sets of GEO data
- Result counts displayed in gray indicate one or more terms not found
Disclaimer
Privacy statement
Accessibility

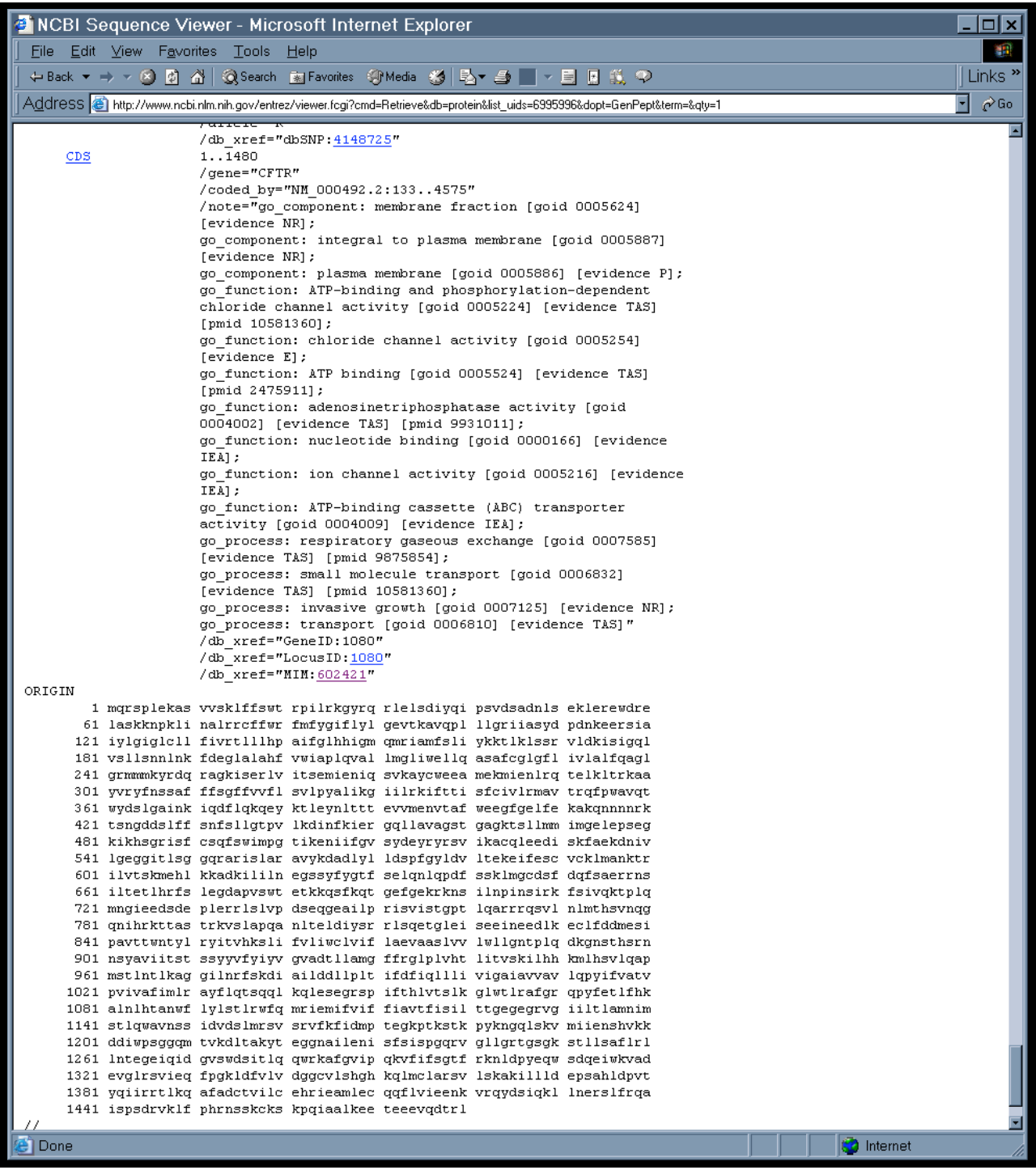

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=Protein&dopt=DocSum&dispmax=1000&val=AAA35680,AAB27879,AAB46340,AAB46341,AAB46342,AAB46352,AA
Go
NCBI
Entrez Protein
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display Summary Show: 500 Send to File
Items 1-9 of 9
One page.
1: AAA35680
cystic fibrosis transmembrane conductance regulator
gi|180332|gb|AAA35680.1|[180332]
BLink, Domains, Links
2: AAB27879
cystic fibrosis transmembrane conductance regulator isoform 36; CFTR 36 [Homo sapiens]
gi|13236882|gb|AAB27879.2||bbm|316417|bbs|136473[13236882]
Links
3: AAB46340
unknown [Homo sapiens]
gi|37674391|gb|AAB46340.2|[37674391]
BLink, Links
4: AAB46341
gb|AAB46341.1|[1669378]
This record was replaced or removed. See revision history for details.
5: AAB46342
gb|AAB46342.1|[1669379]
This record was replaced or removed. See revision history for details.
6: AAB46352
transmembrane chloride conductor protein [Homo sapiens]
gi|1809238|gb|AAB46352.1|[1809238]
BLink, Domains, Links
7: AAC13657
cystic fibrosis transmembrane conductance regulator [Homo sapiens]
gi|306538|gb|AAC13657.1|[306538]
BLink, Domains, Links
8: NP_000483
cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7); ATP-binding cassette, sub-family C member 7; CFTR/MRP [Homo sapiens]
gi|6995996|ref|NP_000483.2|[6995996]
BLink, Domains, Links
Internet

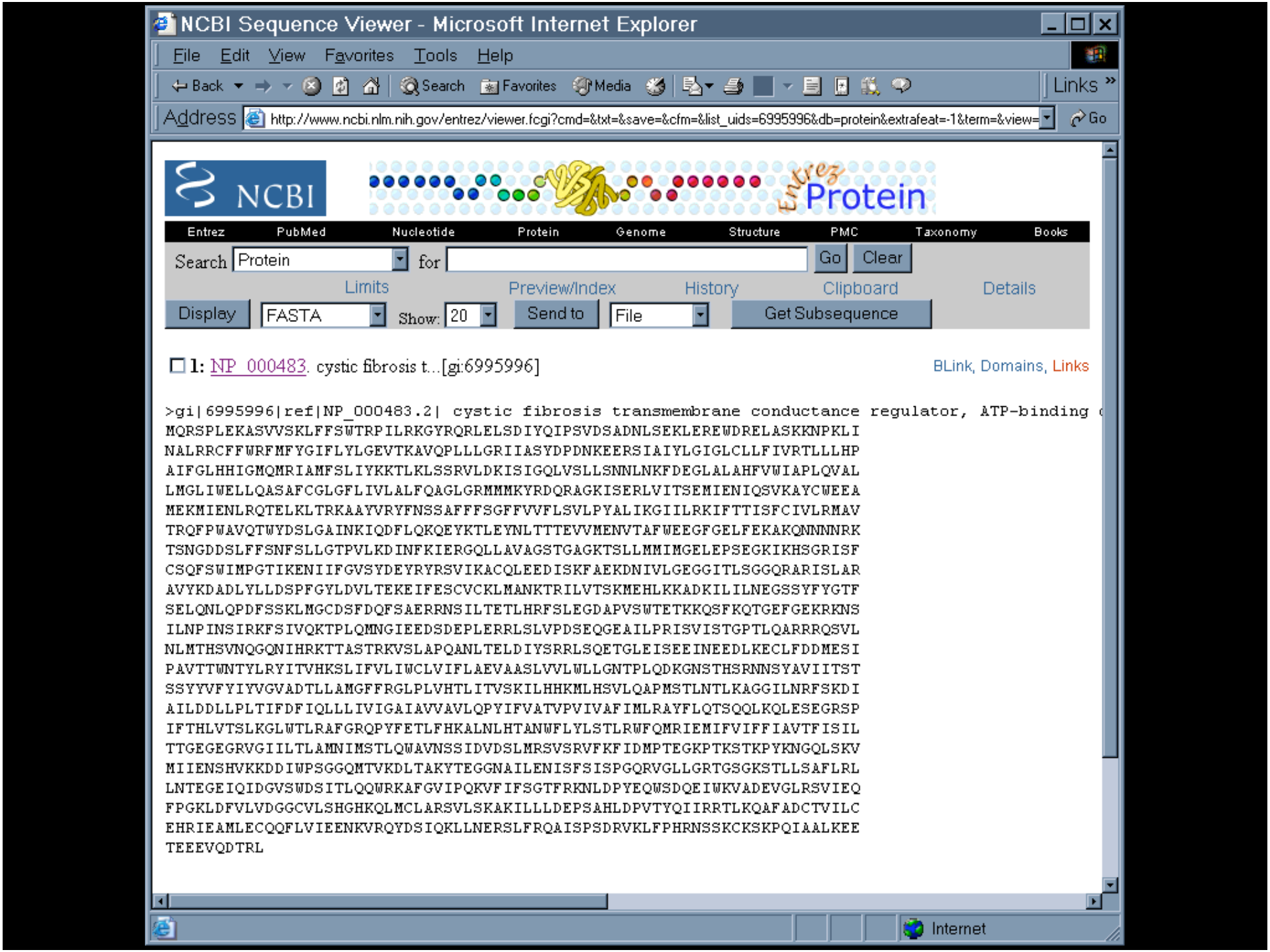
Gene Information

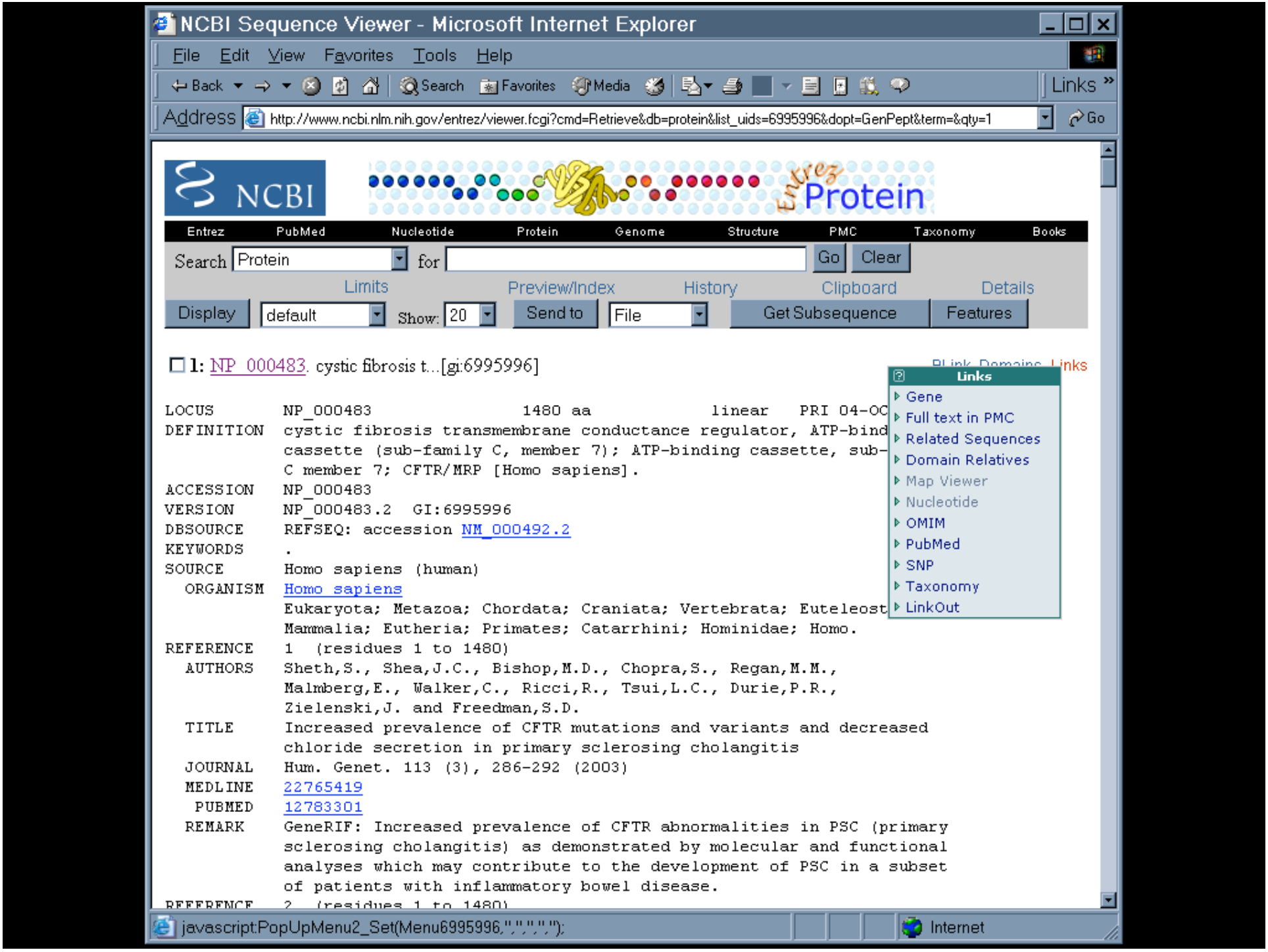
- BLink
  - BLAST Hits
- Domains
  - Protein domains
- Links
  - Varies with available information
- LinkOut
  - “Custom” links to other relevant databases

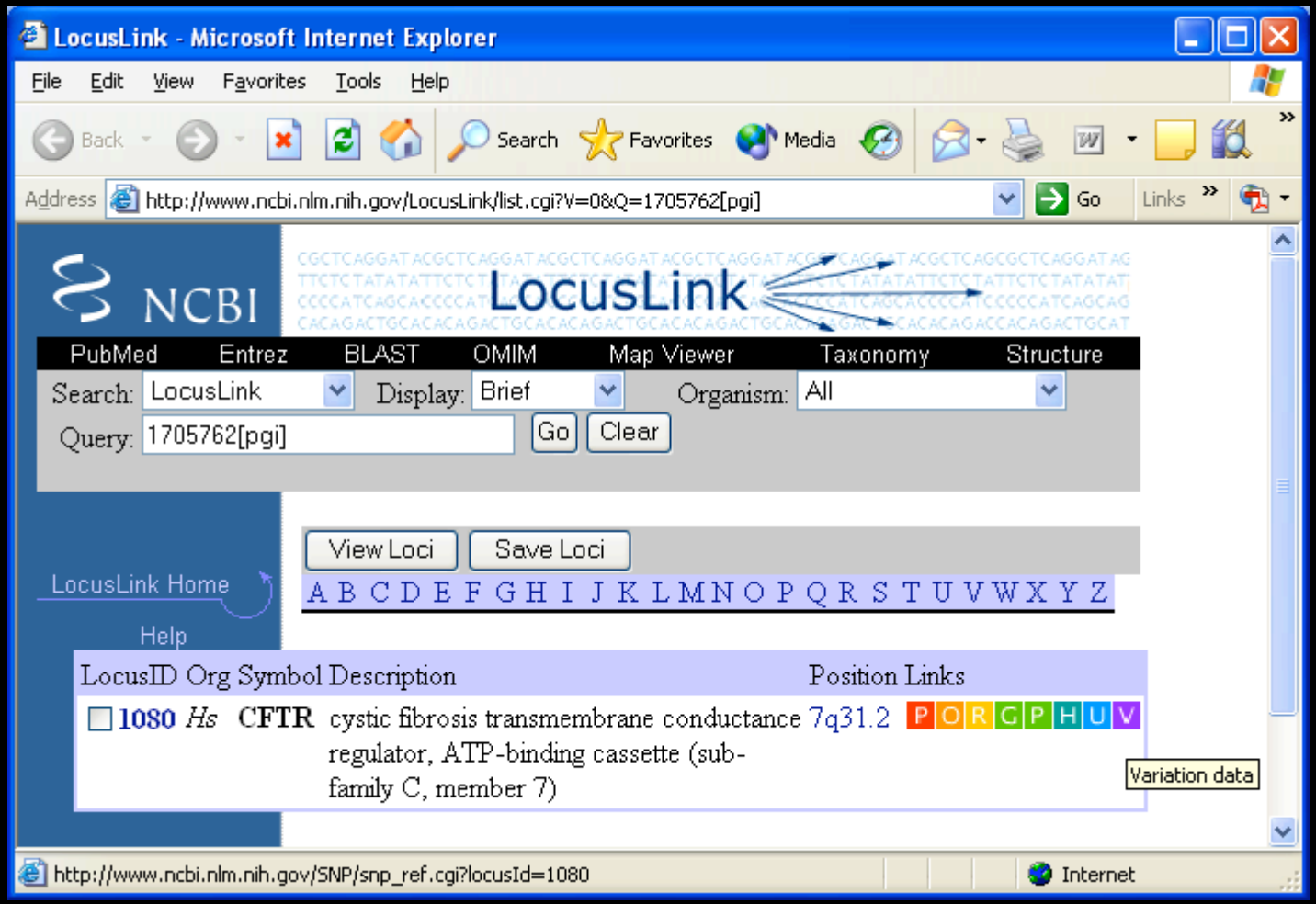

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=Retrieve&db=protein&list_uids=6995996&dopt=GenPept&term=&qty=1
Go
NCBI
Entrez Protein
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display default Show: 20 Send to File Get Subsequence Features
1: NP_000483. cystic fibrosis t...[gi:6995996]
BLink, Domains, Links
LOCUS NP_000483 1480 aa linear PRI 04-OCT-2003
DEFINITION cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7); ATP-binding cassette, sub-family C member 7; CFTR/MRP [Homo sapiens].
ACCESSION NP_000483
VERSION NP_000483.2 GI:6995996
DBSOURCE REFSEQ: accession NM_000492.2
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 1480)
AUTHORS Sheth,S., Shea,J.C., Bishop,M.D., Chopra,S., Regan,M.M., Malmberg,E., Walker,C., Ricci,R., Tsui,L.C., Durie,P.R., Zielenski,J. and Freedman,S.D.
TITLE Increased prevalence of CFTR mutations and variants and decreased chloride secretion in primary sclerosing cholangitis
JOURNAL Hum. Genet. 113 (3), 286-292 (2003)
MEDLINE 22765419
PUBMED 12783301
REMARK GeneRIF: Increased prevalence of CFTR abnormalities in PSC (primary sclerosing cholangitis) as demonstrated by molecular and functional analyses which may contribute to the development of PSC in a subset of patients with inflammatory bowel disease.
REFERENCE 2 (residues 1 to 1480)
AUTHORS Pagani,F., Buratti,E., Stuani,C. and Baralle,F.E.
TITLE Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9
JOURNAL J. Biol. Chem. 278 (29), 26580-26588 (2003)
MEDLINE 22741682
PUBMED 12732620
REMARK GeneRIF: effect on cystic fibrosis transmembrane regulator exon 9 splicing of natural and site-directed sequence mutations
REFERENCE 3 (residues 1 to 1480)
AUTHORS Reddy,M.M. and Quinton,P.M.
TITLE Control of dynamic CFTR selectivity by glutamate and ATP in epithelial cells
JOURNAL Nature 423 (6941), 756-760 (2003)
MEDLINE 22687099
Done
Internet


NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=Retrieve&db=protein&list_uids=6995996&dopt=GenPept&term=&qty=1
Go
COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from M28668.1 and M55131.1.
On Feb 17, 2000 this sequence version replaced gi:4502785.
Summary: The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABC proteins transport various molecules across extra- and intra-cellular membranes. ABC genes are divided into seven distinct subfamilies (ABC1, MDR/TAP, MRP, ALD, OABP, GCN20, White). This protein is a member of the MRP subfamily which is involved in multi-drug resistance. This protein functions as a chloride channel and controls the regulation of other transport pathways. Mutations in this gene have been observed in patients with the autosomal recessive disorders cystic fibrosis (CF) and congenital bilateral aplasia of the vas deferens (CBAVD). Alternative splice variants have been described, many of which result from mutations in the CFTR gene.
FEATURES Location/Qualifiers
source 1..1480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31.2"
Protein 1..1480
/product="cystic fibrosis transmembrane conductance regulator, ATP-binding cassette (sub-family C, member 7)"
/note="ATP-binding cassette, sub-family C member 7; CFTR/MRP"
Region 78..641
/region_name="ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]"
/note="MdlB"
/db_xref="CDD:COG1132"
variation 417
/allele="N"
/allele="K"
/db_xref="dbSNP:4727853"
variation 470
/allele="V"
/allele="M"
/db_xref="dbSNP:213950"
variation 605
/allele="F"
/allele="S"
/db_xref="dbSNP:766874"
Region 850..1439
/region_name="ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]"
/note="MdlB"
/db_xref="CDD:COG1132"
variation 1453
/allele="W"
/allele="R"
Done
Internet

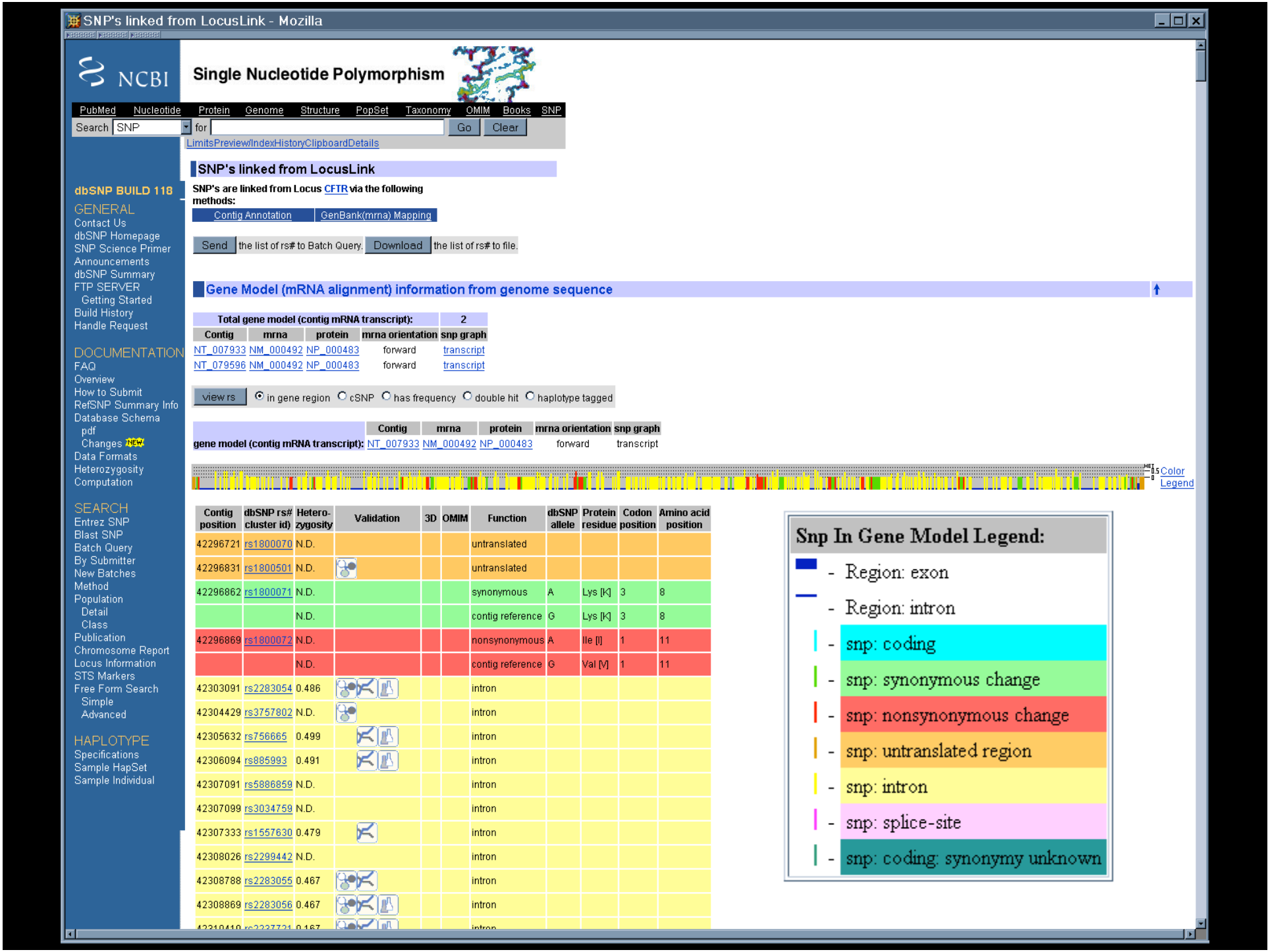

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=Retrieve&db=protein&list_uids=6995996&dopt=GenPept&term=&qty=1
Go
/db_xref="dbSNP:4148725"
CDS 1..1480
/gene="CFTR"
/coded_by="NM_000492.2:133..4575"
/note="go_component: membrane fraction [goid 0005624]
[evidence NR];
go_component: integral to plasma membrane [goid 0005887]
[evidence NR];
go_component: plasma membrane [goid 0005886] [evidence P];
go_function: ATP-binding and phosphorylation-dependent
chloride channel activity [goid 0005224] [evidence TAS]
[pmid 10581360];
go_function: chloride channel activity [goid 0005254]
[evidence E];
go_function: ATP binding [goid 0005524] [evidence TAS]
[pmid 2475911];
go_function: adenosinetriphosphatase activity [goid
0004002] [evidence TAS] [pmid 9931011];
go_function: nucleotide binding [goid 0000166] [evidence
IEA];
go_function: ion channel activity [goid 0005216] [evidence
IEA];
go_function: ATP-binding cassette (ABC) transporter
activity [goid 0004009] [evidence IEA];
go_process: respiratory gaseous exchange [goid 0007585]
[evidence TAS] [pmid 9875854];
go_process: small molecule transport [goid 0006832]
[evidence TAS] [pmid 10581360];
go_process: invasive growth [goid 0007125] [evidence NR];
go_process: transport [goid 0006810] [evidence TAS]"
/db_xref="GeneID:1080"
/db_xref="LocusID:1080"
/db_xref="MIM:602421"
ORIGIN
1 mqrsplekas vvsklffswt rpilrkgyrq rlelsdiyqi psvdsadnls eklerewdre
61 laskknpkli nalrrcffwr fmfygiflyl gevtkavqpl llgriiasyd pdnkeersia
121 iylgiglcll fivrtlllhp aifglhhigm qmriamfsli ykktlklssr vldkisigql
181 vsllsnnlnk fdeglalahf vwiaplqval lmgliwellq asafcglgfl ivlalfqagl
241 grmmmkyrdq ragkiserlv itsemieniq svkaycweea mekmienlrq telkltrkaa
301 yvryfnssaf ffsgffvvfl svlpyalikg iilrkiftti sfcivlrmav trqfpwavqt
361 wydslgaink iqdflqkqey ktleynlttt evvmenvtaf weegfgelfe kakqnnnnrk
421 tsngddslff snfsllgtpv lkdinfkier gqllavagst gagktsllmm imgelepseg
481 kikhsgrisf csqfswimpg tikeniifgv sydeyryrsv ikacqleedi skfaekdniv
541 lgeggitlsg gqrarislar avykdadlyl ldspfgyldv ltekeifesc vcklmanktr
601 ilvtskmehl kkadkililn egssyfygtf selqnlqpdf ssklmgcdsf dqfsaerrns
661 iltetlhrfs legdapvswt etkkqsfkqt gefgekrkns ilnpinsirk fsivqktplq
721 mngieedsde plerrlslvp dseqgeailp risvistgpt lqarrrqsvl nlmthsvnqg
781 qnihrkttas trkvslapqa nlteldiysr rlsqetglei seeineedlk eclfddmesi
841 pavttwntyl ryitvhksli fvliwclvif laevaaslvv lwllgntplq dkgnsthsrn
901 nsyaviitst ssyyvfyiyv gvadtllamg ffrglplvht litvskilhh kmlhsvlqap
961 mstlntlkag gilnrfskdi ailddllplt ifdfiqllli vigaiavvav lqpyifvatv
1021 pvivafimlr ayflqtsqql kqlesegrsp ifthlvtslk glwtlrafgr qpyfetlfhk
1081 alnlhtanwf lylstlrwfq mriemifvif fiavtfisil ttgegegrvg iiltlamnim
1141 stlqwavnss idvdslmrsv srvfkfidmp tegkptkstk pykngqlskv miienshvkk
1201 ddiwpsggqm tvkdltakyt eggnaileni sfsispgqrv gllgrtgsgk stllsaflrl
1261 lntegeiqid gvswdsitlq qwrkafgvip qkvfifsgtf rknldpyeqw sdqeiwkvad
1321 evglrsvieq fpgkldfvlv dggcvlshgh kqlmclarsv lskakillld epsahldpvt
1381 yqiirrtlkq afadctvilc ehrieamlec qqflvieenk vrqydsiqkl lnerslfrqa
1441 ispsdrvklf phrnsskcks kpqiaalkee teeevqdtrl
//
Done
Internet

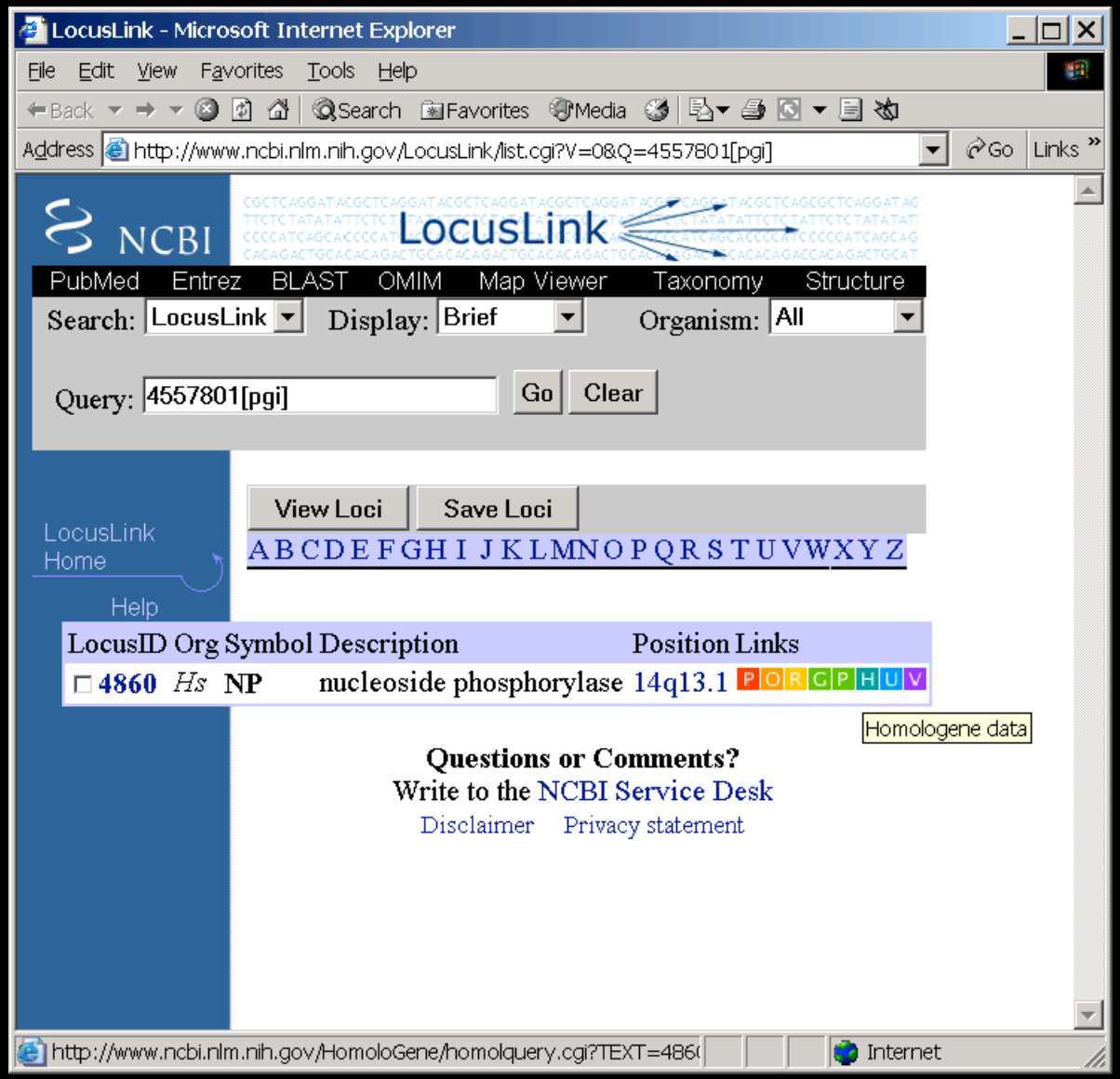

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=&txt=&save=&cfm=&list_uids=6995996&db=protein&extrafeat=-1&term=&view=
Go
NCBI
Entrez Protein
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display FASTA Show: 20 Send to File Get Subsequence
1: NP_000483. cystic fibrosis t...[gi:6995996]
BLink, Domains, Links
>gi|6995996|ref|NP_000483.2| cystic fibrosis transmembrane conductance regulator, ATP-binding
MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRELASKKNPKLI
NALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIAIYLGIGLCLLFIVRTLLLHP
AIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQLVSLLSNNLNKFDEGLALAHFVWIAPLQVAL
LMGLIWELLQASAFCGLGFLIVLALFQAGLGRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEA
MEKMIENLRQTELKLTRKAAYVRYFNSSAFFFSGFFVVFLSVLPYALIKGIILRKIFTTISFCIVLRMAV
TRQFPWAVQTWYDSLGAINKIQDFLQKQEYKTLEYNLTTTEVVMENVTAFWEEGFGELFEKAKQNNNNRK
TSNGDDSLFFSNFSLLGTPVLKDINFKIERGQLLAVAGSTGAGKTSLLMMIMGELEPSEGKIKHSGRISF
CSQFSWIMPGTIKENIIFGVSYDEYRYRSVIKACQLEEDISKFAEKDNIVLGEGGITLSGGQRARISLAR
AVYKDADLYLLDSPFGYLDVLTEKEIFESCVCKLMANKTRILVTSKMEHLKKADKILILNEGSSYFYGTF
SELQNLQPDFSSKLMGCDSFDQFSAERRNSILTETLHRFSLEGDAPVSWTETKKQSFKQTGEFGEKRKNS
ILNPINSIRKFSIVQKTPLQMNGIEEDSDEPLERRLSLVPDSEQGEAILPRISVISTGPTLQARRRQSVL
NLMTHSVNQGQNIHRKTTASTRKVSLAPQANLTELDIYSRRLSQETGLEISEEINEEDLKECLFDDMESI
PAVTTWNTYLRYITVHKSLIFVLIWCLVIFLAEVAASLVVLWLLGNTPLQDKGNSTHSRNNSYAVIITST
SSYYVFYIYVGVADTLLAMGFFRGLPLVHTLITVSKILHHKMLHSVLQAPMSTLNTLKAGGILNRFSKDI
AILDDLLPLTIFDFIQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLRAYFLQTSQQLKQLESEGRSP
IFTHLVTSLKGLWTLRAFGRQPYFETLFHKALNLHTANWFLYLSTLRWFQMRIEMIFVIFFIAVTFISIL
TTGEGEGRVGIILTLAMNIMSTLQWAVNSSIDVDSLMRSVSRVFKFIDMPTEGKPTKSTKPYKNGQLSKV
MIIENSHVKKDDIWPSGGQMTVKDLTAKYTEGGNAILENISFSISPGQRVGLLGRTGSGKSTLLSAFLRL
LNTEGEIQIDGVSWDSITLQQWRKAFGVIPQKVFIFSGTFRKNLDPYEQWSDQEIWKVADEVGLRSVIEQ
FPGKLDFVLVDGGCVLSHGHKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQIIRRTLKQAFADCTVILC
EHRIEAMLECQQFLVIEENKVRQYDSIQKLLNERSLFRQAISPSDRVKLFPHRNSSKCKSKPQIAALKEE
TEEEVQDTRL
Internet

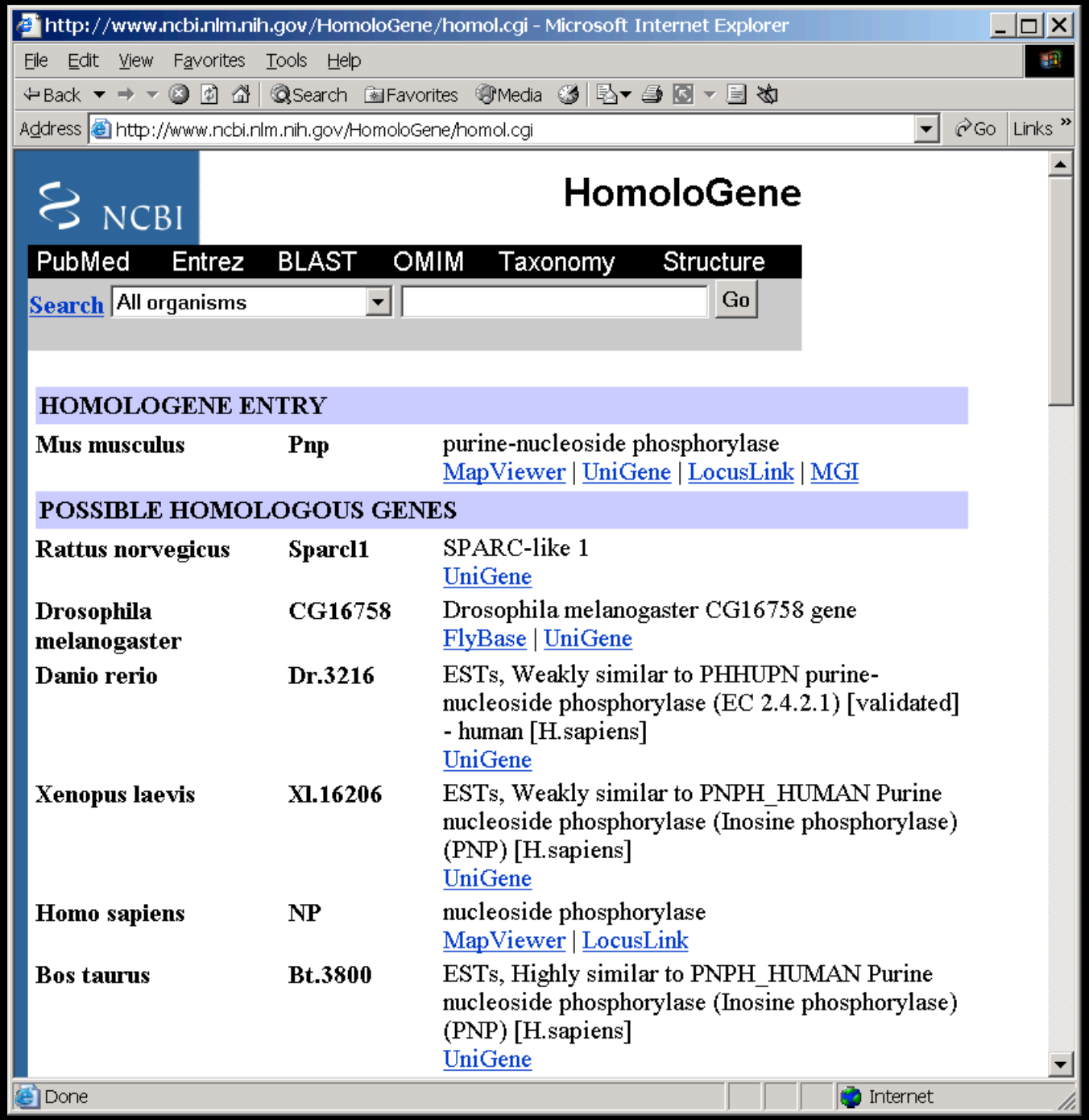

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media Links
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?cmd=Retrieve&db=protein&list_uids=6995996&dopt=GenPept&term=&qty=1
Go
NCBI
Entrez Protein
Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display default Show: 20 Send to File Get Subsequence Features
1: NP_000483. cystic fibrosis t...[gi:6995996]
Links
Gene
Full text in PMC
Related Sequences
Domain Relatives
Map Viewer
Nucleotide
OMIM
PubMed
SNP
Taxonomy
LinkOut
LOCUS NP_000483 1480 aa linear PRI 04-OC
DEFINITION cystic fibrosis transmembrane conductance regulator, ATP-bind
cassette (sub-family C, member 7); ATP-binding cassette, sub-
C member 7; CFTR/MRP [Homo sapiens].
ACCESSION NP_000483
VERSION NP_000483.2 GI:6995996
DBSOURCE REFSEQ: accession NM_000492.2
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleost
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 1480)
AUTHORS Sheth,S., Shea,J.C., Bishop,M.D., Chopra,S., Regan,M.M.,
Malmberg,E., Walker,C., Ricci,R., Tsui,L.C., Durie,P.R.,
Zielenski,J. and Freedman,S.D.
TITLE Increased prevalence of CFTR mutations and variants and decreased
chloride secretion in primary sclerosing cholangitis
JOURNAL Hum. Genet. 113 (3), 286-292 (2003)
MEDLINE 22765419
PUBMED 12783301
REMARK GeneRIF: Increased prevalence of CFTR abnormalities in PSC (primary
sclerosing cholangitis) as demonstrated by molecular and functional
analyses which may contribute to the development of PSC in a subset
of patients with inflammatory bowel disease.
javascript:PopUpMenu2_Set(Menu6995996,",",",",",");
Internet

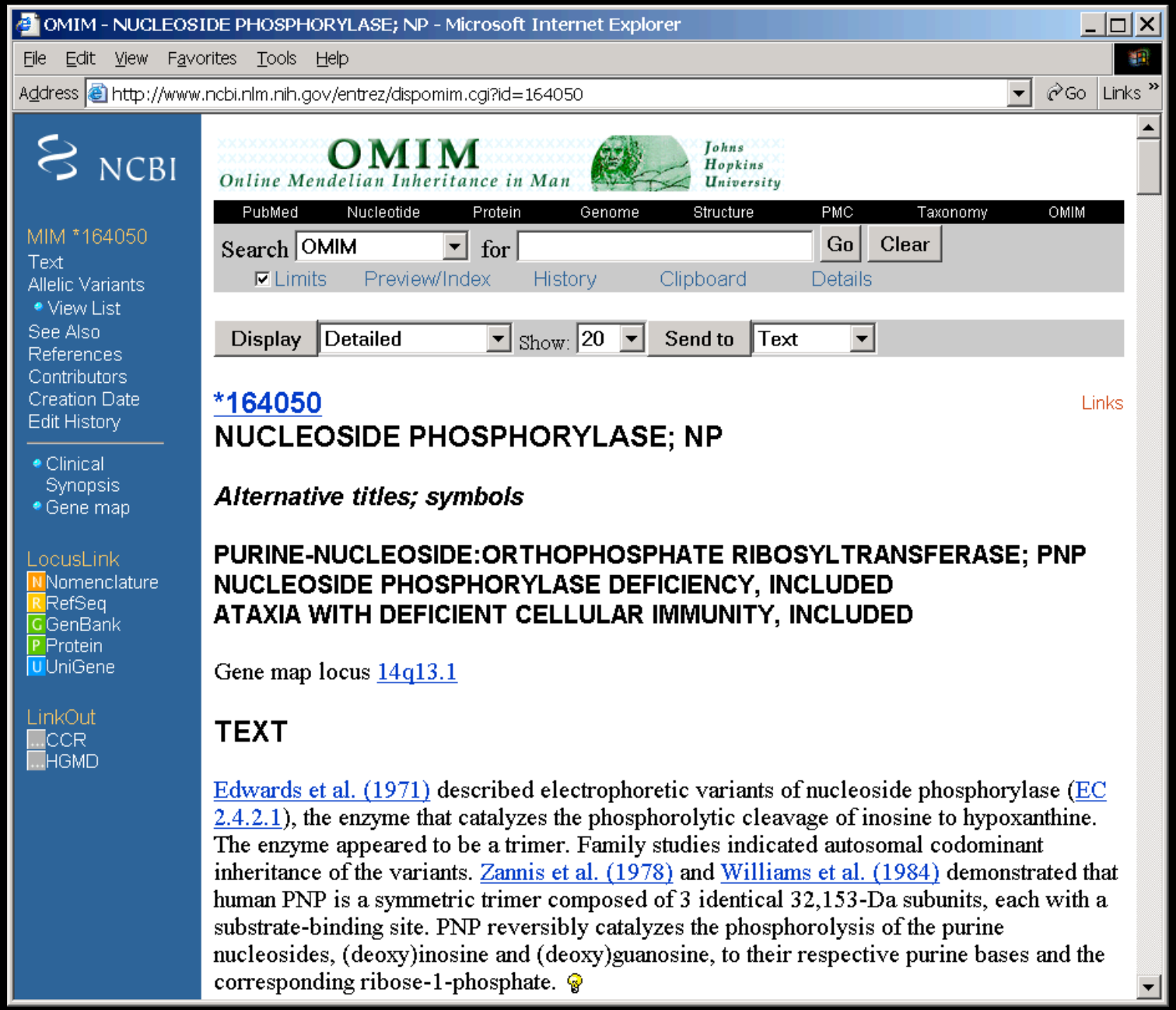


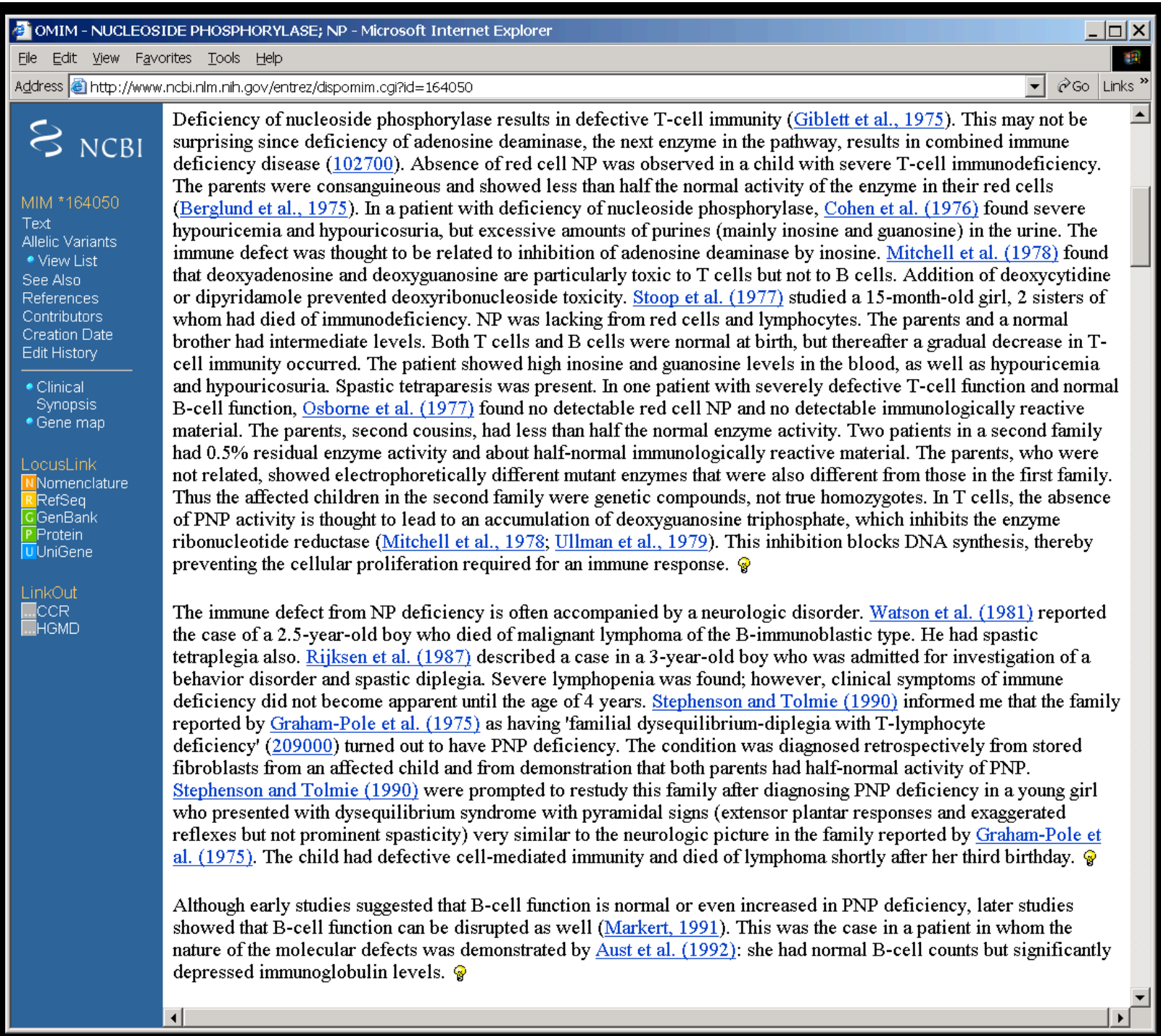
P: PubMed
O: OMIM
R: RefSeq
G: GenBank
P: Protein
H: Homologene
U: Unigene
V: Variation


SNP's linked from LocusLink - Mozilla
NCBI
Single Nucleotide Polymorphism
PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP
Search SNP for Go Clear
Limits Preview/Index History Clipboard Details
dbSNP BUILD 118
GENERAL
Contact Us
dbSNP Homepage
SNP Science Primer
Announcements
dbSNP Summary
FTP SERVER
Getting Started
Build History
Handle Request
DOCUMENTATION
FAQ
Overview
How to Submit
RefSNP Summary Info
Database Schema
pdf
Changes NEW
Data Formats
Heterozygosity
Computation
SEARCH
Entrez SNP
Blast SNP
Batch Query
By Submitter
New Batches
Method
Population
Detail
Class
Publication
Chromosome Report
Locus Information
STS Markers
Free Form Search
Simple
Advanced
HAPLOTYPE
Specifications
Sample HapSet
Sample Individual
SNP's linked from LocusLink
SNP's are linked from Locus CFTR via the following methods:
Contig Annotation
GenBank(mrna) Mapping
Send the list of rs# to Batch Query.
Download the list of rs# to file.
Gene Model (mRNA alignment) information from genome sequence
Total gene model (contig mRNA transcript): 2
Contig mrna protein mrna orientation snp graph
NT_007933 NM_000492 NP_000483 forward transcript
NT_079596 NM_000492 NP_000483 forward transcript
view rs
in gene region
cSNP
has frequency
double hit
haplotype tagged
gene model (contig mRNA transcript): NT_007933 NM_000492 NP_000483 forward transcript
Color Legend
Contig position
dbSNP rs# cluster id
Hetero-zygosity
Validation
3D
OMIM
Function
dbSNP allele
Protein residue
Codon position
Amino acid position
42296721 rs1800070 N.D. untranslated
42296831 rs1800501 N.D. untranslated
42296862 rs1800071 N.D. synonymous A Lys [K] 3 8
N.D. contig reference G Lys [K] 3 8
42296869 rs1800072 N.D. nonsynonymous A Ile [I] 1 11
N.D. contig reference G Val [V] 1 11
42303091 rs2283054 0.486 intron
42304429 rs3757802 N.D. intron
42305632 rs756665 0.499 intron
42306094 rs885993 0.491 intron
42307091 rs5886859 N.D. intron
42307099 rs3034759 N.D. intron
42307333 rs1557630 0.479 intron
42308026 rs2299442 N.D. intron
42308788 rs2283055 0.467 intron
42308869 rs2283056 0.467 intron
Snp In Gene Model Legend:
- Region: exon
- Region: intron
- snp: coding
- snp: synonymous change
- snp: nonsynonymous change
- snp: untranslated region
- snp: intron
- snp: splice-site
- snp: coding: synonymy unknown


LocusLink - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi?V=0&Q=4557801[pgi]
Go Links
NCBI
LocusLink
PubMed Entrez BLAST OMIM Map Viewer Taxonomy Structure
Search: LocusLink Display: Brief Organism: All
Query: 4557801[pgi]
Go Clear
View Loci Save Loci
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
LocusLink Home
Help
LocusID Org Symbol Description Position Links
4860 Hs NP nucleoside phosphorylase 14q13.1 P O R G P H U V
Homologene data
Questions or Comments?
Write to the NCBI Service Desk
Disclaimer Privacy statement
http://www.ncbi.nlm.nih.gov/HomoloGene/homolquery.cgi?TEXT=486
Internet


http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/HomoloGene/homol.cgi
Go Links
NCBI
HomoloGene
PubMed Entrez BLAST OMIM Taxonomy Structure
Search All organisms Go
HOMOLOGENE ENTRY
Mus musculus Pnp purine-nucleoside phosphorylase
MapViewer | UniGene | LocusLink | MGI
POSSIBLE HOMOLOGOUS GENES
Rattus norvegicus Sparcl1 SPARC-like 1
UniGene
Drosophila melanogaster CG16758 Drosophila melanogaster CG16758 gene
FlyBase | UniGene
Danio rerio Dr.3216 ESTs, Weakly similar to PHHUPN purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human [H.sapiens]
UniGene
Xenopus laevis Xl.16206 ESTs, Weakly similar to PNPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens]
UniGene
Homo sapiens NP nucleoside phosphorylase
MapViewer | LocusLink
Bos taurus Bt.3800 ESTs, Highly similar to PNPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens]
UniGene
Done
Internet

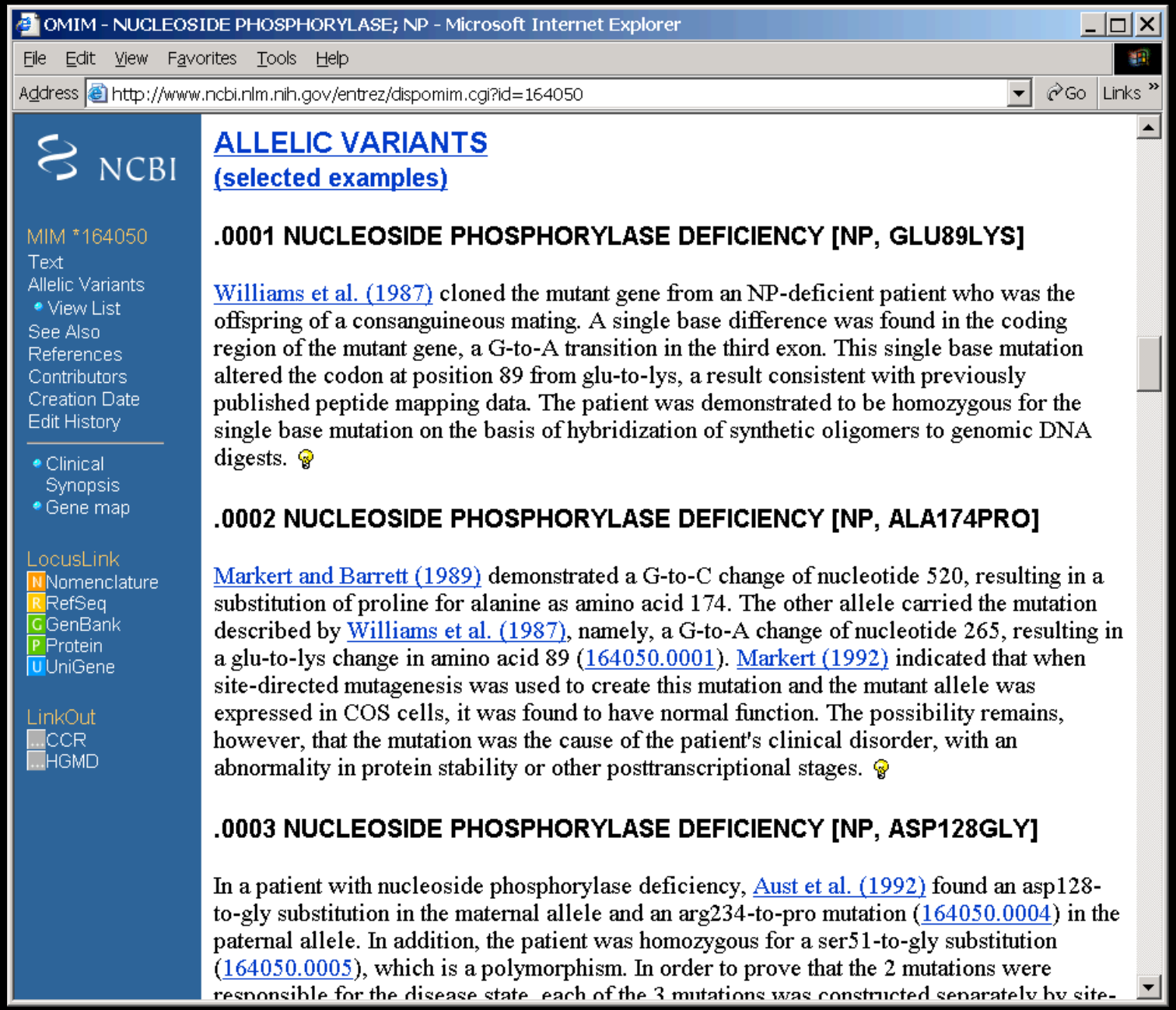

OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050
Go Links
NCBI
OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM
Search OMIM for Go Clear
Limits Preview/Index History Clipboard Details
Display Detailed Show: 20 Send to Text
MIM *164050
Text
Allelic Variants
View List
See Also
References
Contributors
Creation Date
Edit History
Clinical Synopsis
Gene map
LocusLink
Nomenclature
RefSeq
GenBank
Protein
UniGene
LinkOut
CCR
HGMD
*164050
Links
NUCLEOSIDE PHOSPHORYLASE; NP
Alternative titles; symbols
PURINE-NUCLEOSIDE:ORTHOPHOSPHATE RIBOSYLTRANSFERASE; PNP
NUCLEOSIDE PHOSPHORYLASE DEFICIENCY, INCLUDED
ATAXIA WITH DEFICIENT CELLULAR IMMUNITY, INCLUDED
Gene map locus 14q13.1
TEXT
Edwards et al. (1971) described electrophoretic variants of nucleoside phosphorylase (EC 2.4.2.1), the enzyme that catalyzes the phosphorolytic cleavage of inosine to hypoxanthine. The enzyme appeared to be a trimer. Family studies indicated autosomal codominant inheritance of the variants. Zannis et al. (1978) and Williams et al. (1984) demonstrated that human PNP is a symmetric trimer composed of 3 identical 32,153-Da subunits, each with a substrate-binding site. PNP reversibly catalyzes the phosphorolysis of the purine nucleosides, (deoxy)inosine and (deoxy)guanosine, to their respective purine bases and the corresponding ribose-1-phosphate.

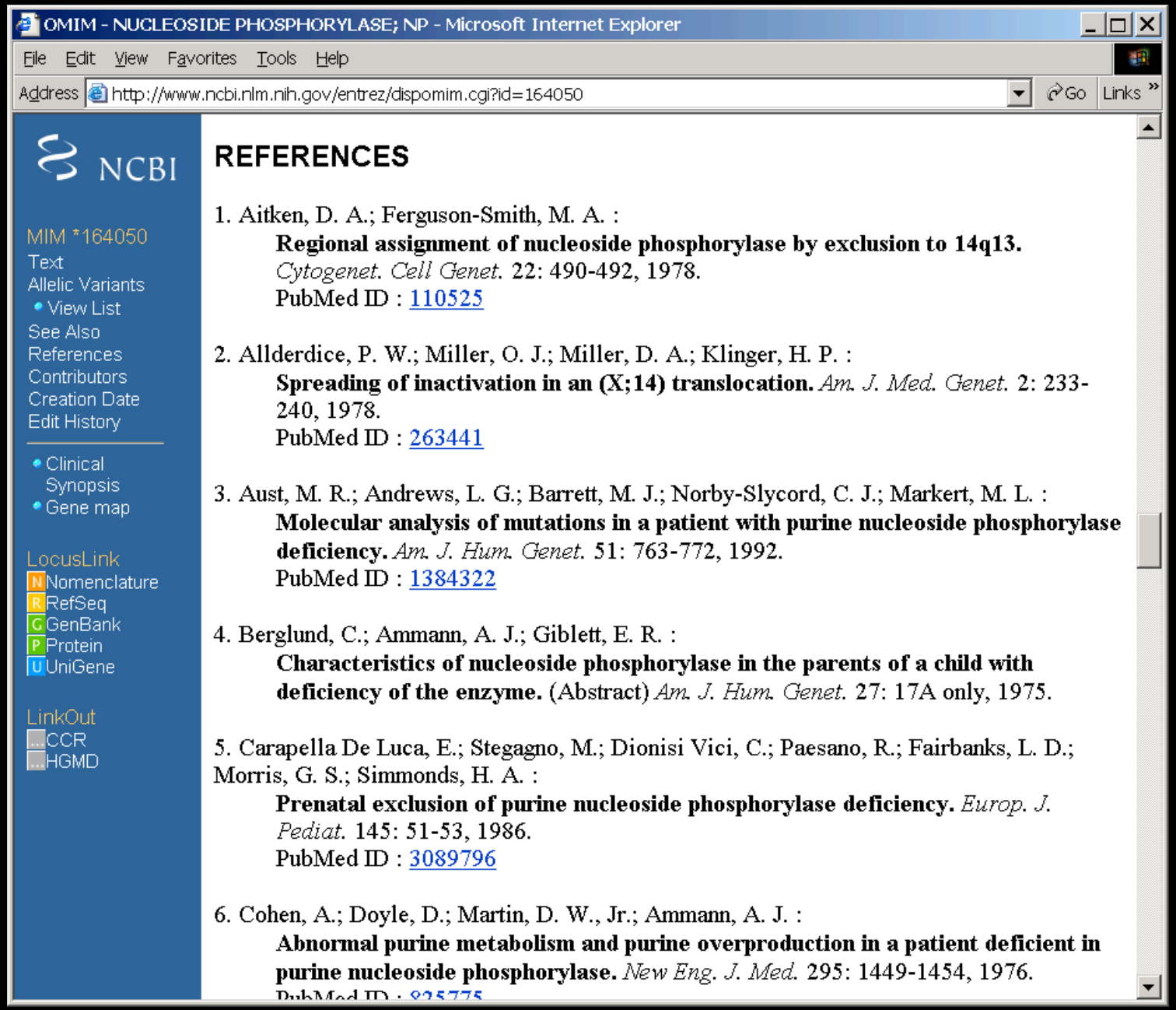


Deficiency of nucleoside phosphorylase results in defective T-cell immunity (Giblett et al., 1975). This may not be surprising since deficiency of adenosine deaminase, the next enzyme in the pathway, results in combined immune deficiency disease (102700). Absence of red cell NP was observed in a child with severe T-cell immunodeficiency. The parents were consanguineous and showed less than half the normal activity of the enzyme in their red cells (Berglund et al., 1975). In a patient with deficiency of nucleoside phosphorylase, Cohen et al. (1976) found severe hypouricemia and hypouricosuria, but excessive amounts of purines (mainly inosine and guanosine) in the urine. The immune defect was thought to be related to inhibition of adenosine deaminase by inosine. Mitchell et al. (1978) found that deoxyadenosine and deoxyguanosine are particularly toxic to T cells but not to B cells. Addition of deoxycytidine or dipyridamole prevented deoxyribonucleoside toxicity. Stoop et al. (1977) studied a 15-month-old girl, 2 sisters of whom had died of immunodeficiency. NP was lacking from red cells and lymphocytes. The parents and a normal brother had intermediate levels. Both T cells and B cells were normal at birth, but thereafter a gradual decrease in T-cell immunity occurred. The patient showed high inosine and guanosine levels in the blood, as well as hypouricemia and hypouricosuria. Spastic tetraparesis was present. In one patient with severely defective T-cell function and normal B-cell function, Osborne et al. (1977) found no detectable red cell NP and no detectable immunologically reactive material. The parents, second cousins, had less than half the normal enzyme activity. Two patients in a second family had 0.5% residual enzyme activity and about half-normal immunologically reactive material. The parents, who were not related, showed electrophoretically different mutant enzymes that were also different from those in the first family. Thus the affected children in the second family were genetic compounds, not true homozygotes. In T cells, the absence of PNP activity is thought to lead to an accumulation of deoxyguanosine triphosphate, which inhibits the enzyme ribonucleotide reductase (Mitchell et al., 1978; Ullman et al., 1979). This inhibition blocks DNA synthesis, thereby preventing the cellular proliferation required for an immune response.

The immune defect from NP deficiency is often accompanied by a neurologic disorder. Watson et al. (1981) reported the case of a 2.5-year-old boy who died of malignant lymphoma of the B-immunoblastic type. He had spastic tetraplegia also. Rijksen et al. (1987) described a case in a 3-year-old boy who was admitted for investigation of a behavior disorder and spastic diplegia. Severe lymphopenia was found; however, clinical symptoms of immune deficiency did not become apparent until the age of 4 years. Stephenson and Tolmie (1990) informed me that the family reported by Graham-Pole et al. (1975) as having 'familial dysequilibrium-diplegia with T-lymphocyte deficiency' (209000) turned out to have PNP deficiency. The condition was diagnosed retrospectively from stored fibroblasts from an affected child and from demonstration that both parents had half-normal activity of PNP. Stephenson and Tolmie (1990) were prompted to restudy this family after diagnosing PNP deficiency in a young girl who presented with dysequilibrium syndrome with pyramidal signs (extensor plantar responses and exaggerated reflexes but not prominent spasticity) very similar to the neurologic picture in the family reported by Graham-Pole et al. (1975). The child had defective cell-mediated immunity and died of lymphoma shortly after her third birthday.

Although early studies suggested that B-cell function is normal or even increased in PNP deficiency, later studies showed that B-cell function can be disrupted as well (Markert, 1991). This was the case in a patient in whom the nature of the molecular defects was demonstrated by Aust et al. (1992): she had normal B-cell counts but significantly depressed immunoglobulin levels.


OMIM - NUCLEOSIDE PHOSPHORYLASE; NP - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=164050
Go Links
NCBI
MIM *164050
Text
Allelic Variants
View List
See Also
References
Contributors
Creation Date
Edit History
Clinical Synopsis
Gene map
LocusLink
Nomenclature
RefSeq
GenBank
Protein
UniGene
LinkOut
CCR
HGMD
ALLELIC VARIANTS
(selected examples)
.0001 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, GLU89LYS]
Williams et al. (1987) cloned the mutant gene from an NP-deficient patient who was the offspring of a consanguineous mating. A single base difference was found in the coding region of the mutant gene, a G-to-A transition in the third exon. This single base mutation altered the codon at position 89 from glu-to-lys, a result consistent with previously published peptide mapping data. The patient was demonstrated to be homozygous for the single base mutation on the basis of hybridization of synthetic oligomers to genomic DNA digests.
.0002 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ALA174PRO]
Markert and Barrett (1989) demonstrated a G-to-C change of nucleotide 520, resulting in a substitution of proline for alanine as amino acid 174. The other allele carried the mutation described by Williams et al. (1987), namely, a G-to-A change of nucleotide 265, resulting in a glu-to-lys change in amino acid 89 (164050.0001). Markert (1992) indicated that when site-directed mutagenesis was used to create this mutation and the mutant allele was expressed in COS cells, it was found to have normal function. The possibility remains, however, that the mutation was the cause of the patient's clinical disorder, with an abnormality in protein stability or other posttranscriptional stages.
.0003 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ASP128GLY]
In a patient with nucleoside phosphorylase deficiency, Aust et al. (1992) found an asp128-to-gly substitution in the maternal allele and an arg234-to-pro mutation (164050.0004) in the paternal allele. In addition, the patient was homozygous for a ser51-to-gly substitution (164050.0005), which is a polymorphism. In order to prove that the 2 mutations were



REFERENCES

1. Aitken, D. A.; Ferguson-Smith, M. A. :
Regional assignment of nucleoside phosphorylase by exclusion to 14q13.
Cytogenet. Cell Genet. 22: 490-492, 1978.
PubMed ID : 110525

2. Allderdice, P. W.; Miller, O. J.; Miller, D. A.; Klinger, H. P. :
Spreading of inactivation in an (X;14) translocation. *Am. J. Med. Genet.* 2: 233-240, 1978.
PubMed ID : 263441

3. Aust, M. R.; Andrews, L. G.; Barrett, M. J.; Norby-Slycord, C. J.; Markert, M. L. :
Molecular analysis of mutations in a patient with purine nucleoside phosphorylase deficiency. *Am. J. Hum. Genet.* 51: 763-772, 1992.
PubMed ID : 1384322

4. Berglund, C.; Ammann, A. J.; Giblett, E. R. :
Characteristics of nucleoside phosphorylase in the parents of a child with deficiency of the enzyme. (Abstract) *Am. J. Hum. Genet.* 27: 17A only, 1975.

5. Carapella De Luca, E.; Stegagno, M.; Dionisi Vici, C.; Paesano, R.; Fairbanks, L. D.; Morris, G. S.; Simmonds, H. A. :
Prenatal exclusion of purine nucleoside phosphorylase deficiency. *Europ. J. Pediat.* 145: 51-53, 1986.
PubMed ID : 3089796

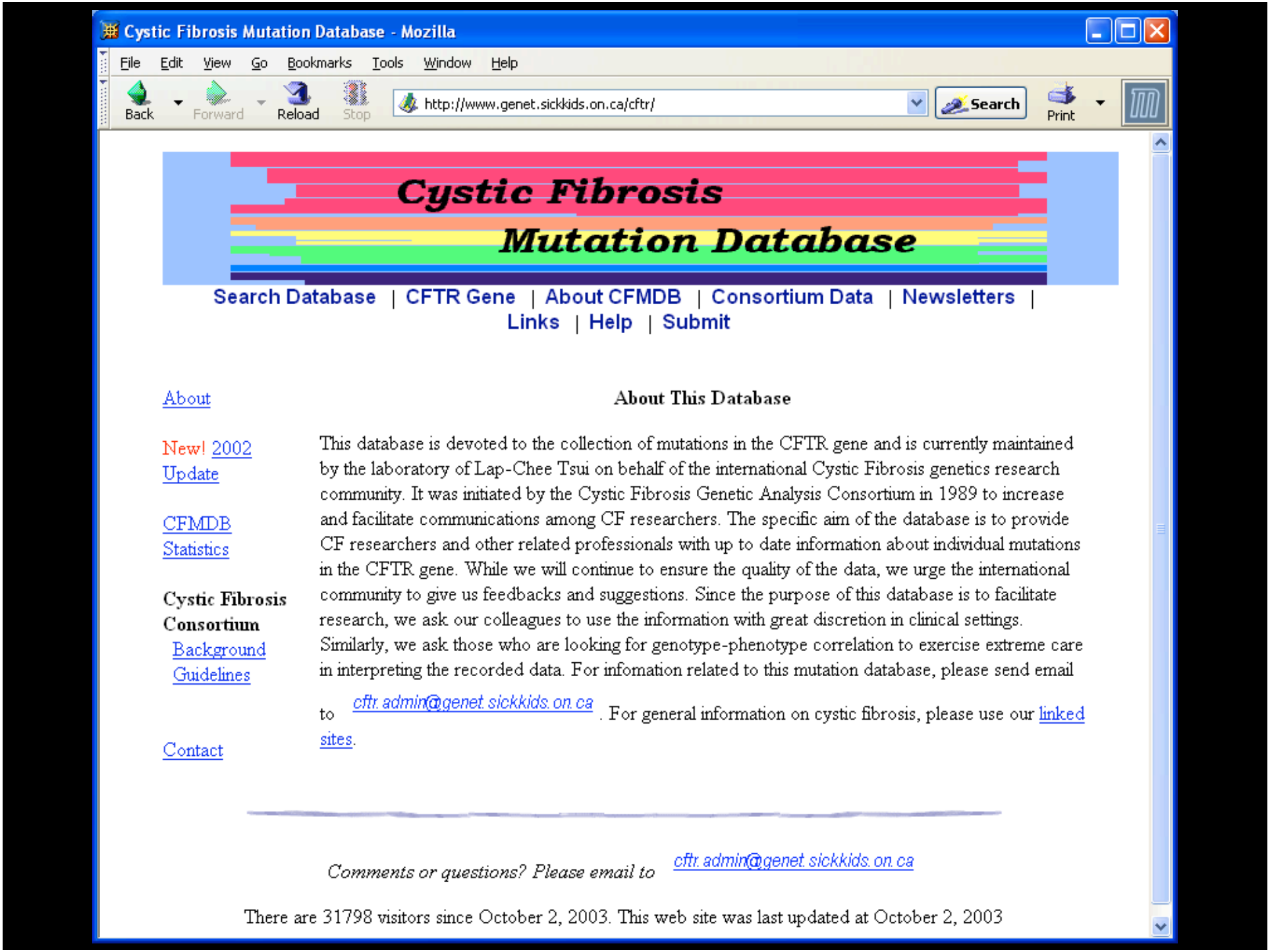
6. Cohen, A.; Doyle, D.; Martin, D. W., Jr.; Ammann, A. J. :
Abnormal purine metabolism and purine overproduction in a patient deficient in purine nucleoside phosphorylase. *New Eng. J. Med.* 295: 1449-1454, 1976.

OMIM Allied Resources

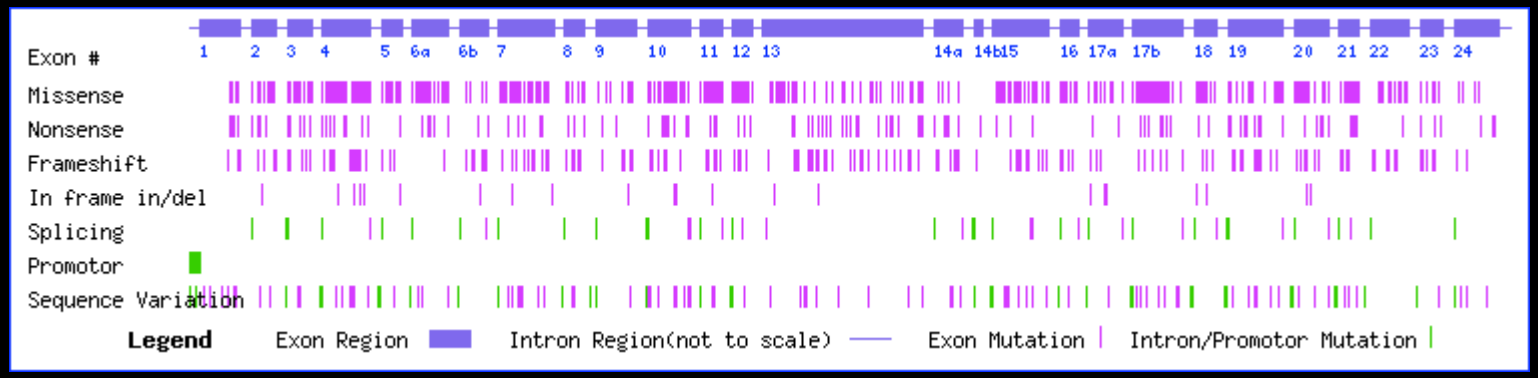
- Locus Specific Mutation Databases
- Model Organisms
- Phenotypes and Clinical Resources
- Additional Resources
- Mapping Resources
- Organizations and Research Programs
- Listservs

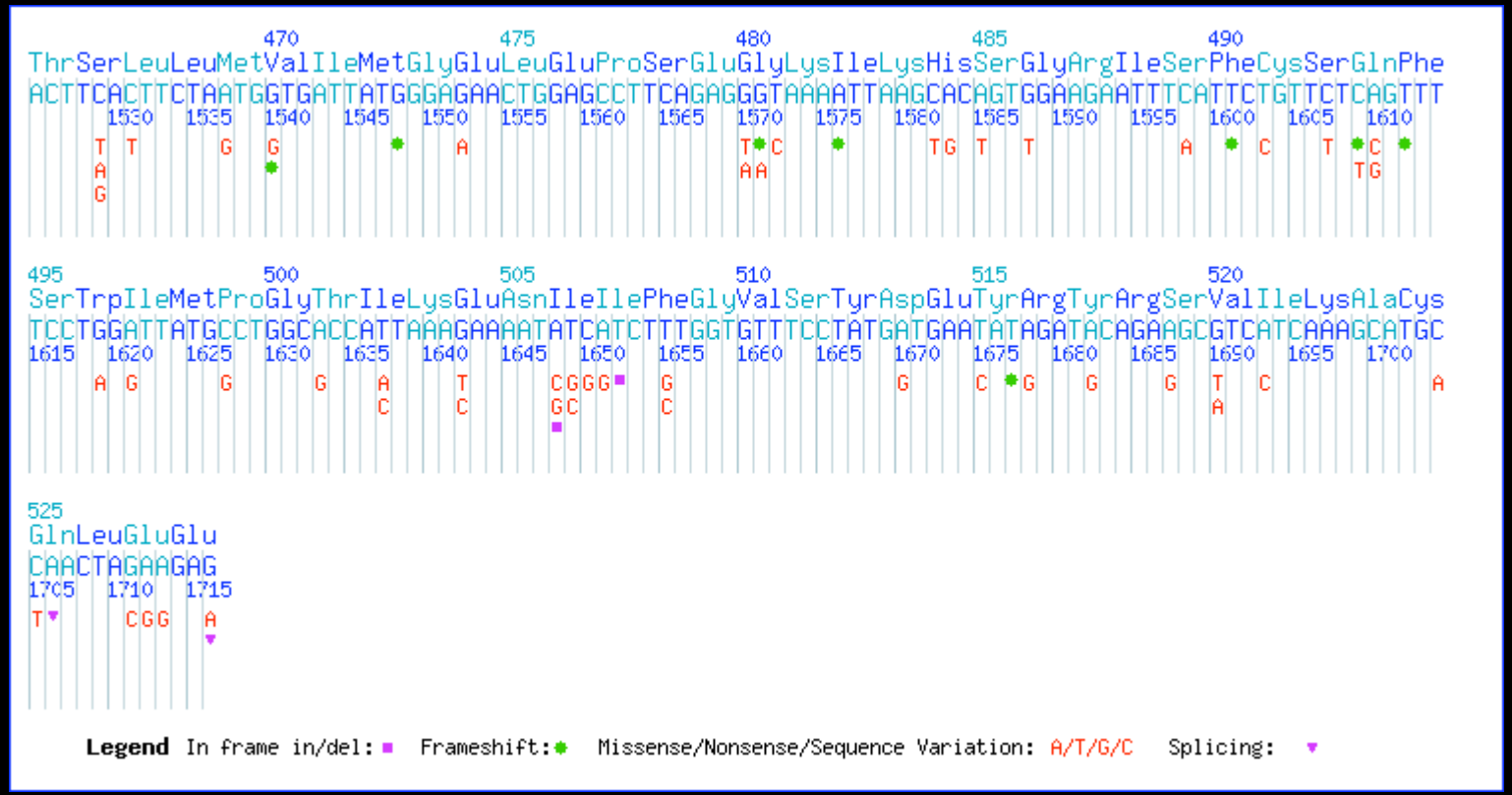
Locus Specific Mutation Databases

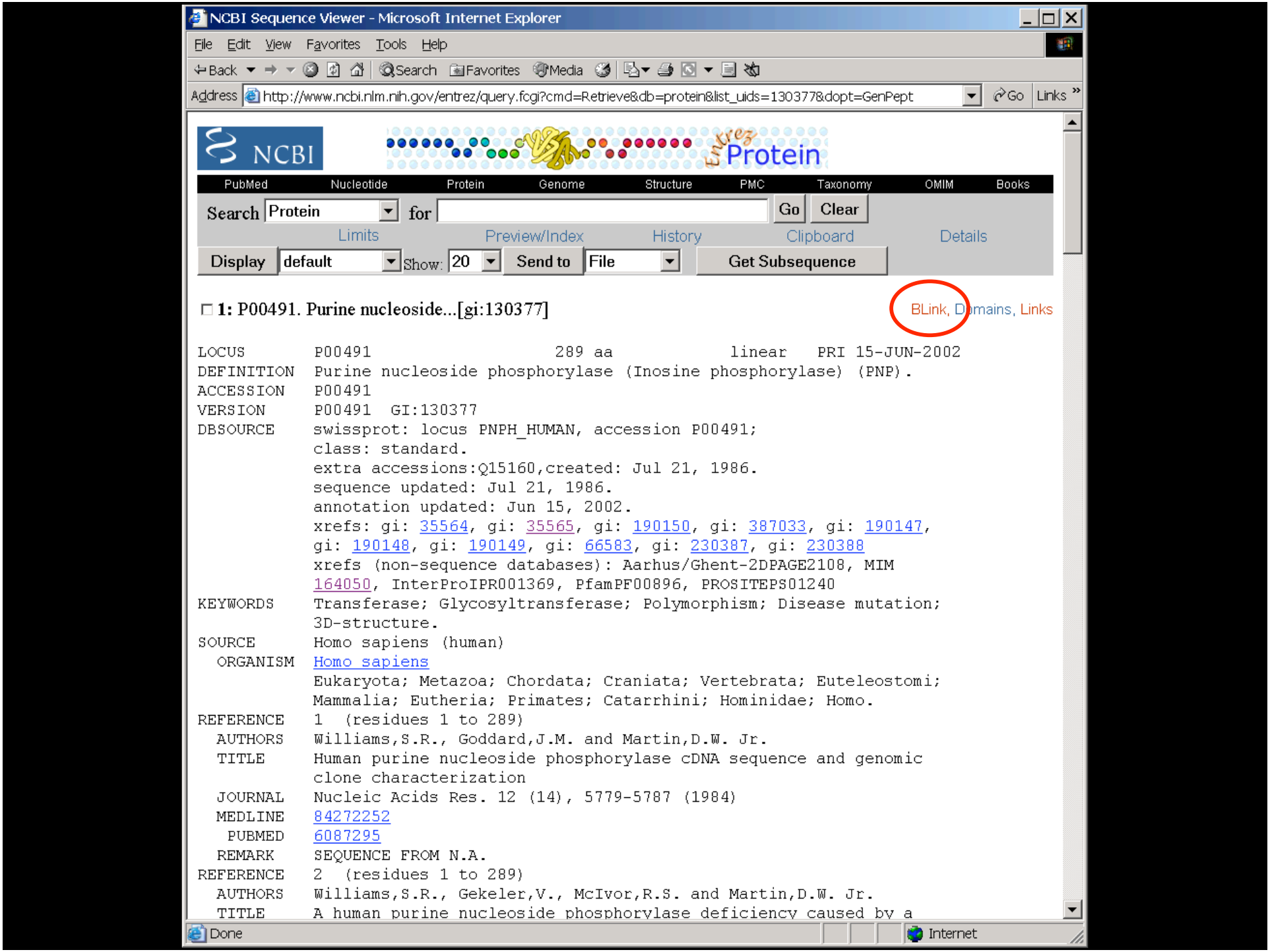
- Androgen Receptor Mutations Database
- Ataxia-Telangiectasia Database
- BIOMDB database of mutations causing tetrahydrobiopterin deficiencies
- BTKBase mutation registry for X-Linked agammaglobulinemia
- Cystic Fibrosis Mutation Database
- Emery-Dreifuss Muscular Dystrophy Mutation Database
- Emory University MitoMap mitochondrion genome database
- Factor VII Mutation Database
- Fanconi Anemia Mutation Database
- Favism Database of Glucose-6-Phosphate Mutations
- Glycogen Storage Disease Type II (Pompe Disease) Mutation Database
- Hemophilia A Mutation Database
- Hereditary Non-Polyposis Colorectal Cancer Database
- Hexosaminidase A Locus Database
- Human Type I and Type III Collagen Mutation Database
- IARC TP53 Mutation Database
- IL2RG mutation database for X-linked SCID
- Iowa Compendium of Rhodopsin and RDS Mutations
- L1CAM Mutation Database
- LDL Receptor Mutation Database
- Ornithine Transcarbamylase Structure and Mutation Database
- PAX6 Mutation Database
- Phenylalanine Hydroxylase Locus Database
- RB1 Gene Mutation Database
- Tuberous Sclerosis Mutation Database
- von Willebrand Factor Database
- Werner Syndrome Mutation Database

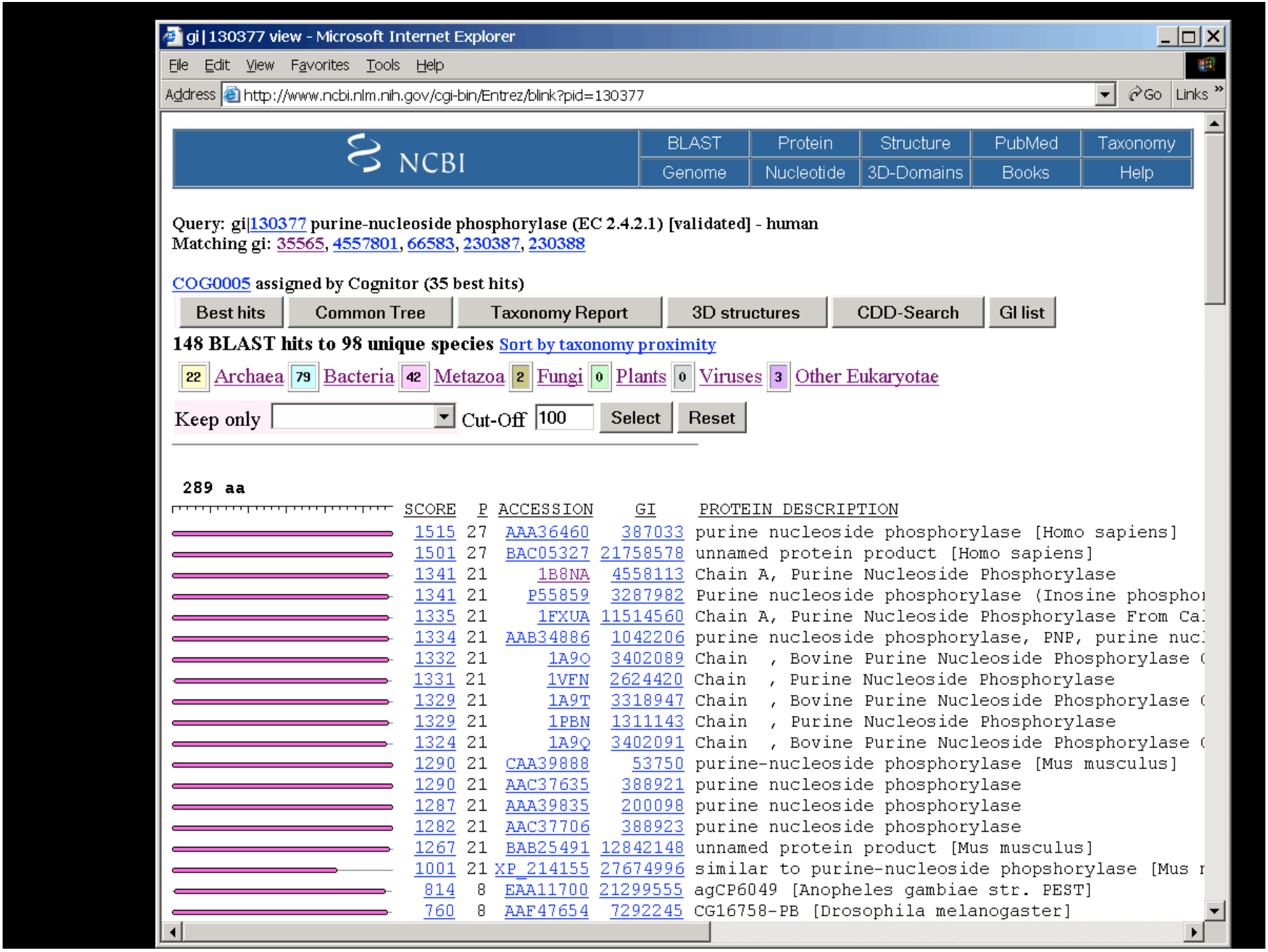

Cystic Fibrosis Mutation Database - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://www.genet.sickkids.on.ca/cftr/
Search
Print
Cystic Fibrosis
Mutation Database
Search Database | CFTR Gene | About CFMDB | Consortium Data | Newsletters |
Links | Help | Submit
About
New! 2002 Update
CFMDB Statistics
Cystic Fibrosis Consortium
Background
Guidelines
Contact
About This Database
This database is devoted to the collection of mutations in the CFTR gene and is currently maintained by the laboratory of Lap-Chee Tsui on behalf of the international Cystic Fibrosis genetics research community. It was initiated by the Cystic Fibrosis Genetic Analysis Consortium in 1989 to increase and facilitate communications among CF researchers. The specific aim of the database is to provide CF researchers and other related professionals with up to date information about individual mutations in the CFTR gene. While we will continue to ensure the quality of the data, we urge the international community to give us feedbacks and suggestions. Since the purpose of this database is to facilitate research, we ask our colleagues to use the information with great discretion in clinical settings. Similarly, we ask those who are looking for genotype-phenotype correlation to exercise extreme care in interpreting the recorded data. For infomation related to this mutation database, please send email to cftr.admin@genet.sickkids.on.ca. For general information on cystic fibrosis, please use our linked sites.
Comments or questions? Please email to cftr.admin@genet.sickkids.on.ca
There are 31798 visitors since October 2, 2003. This web site was last updated at October 2, 2003

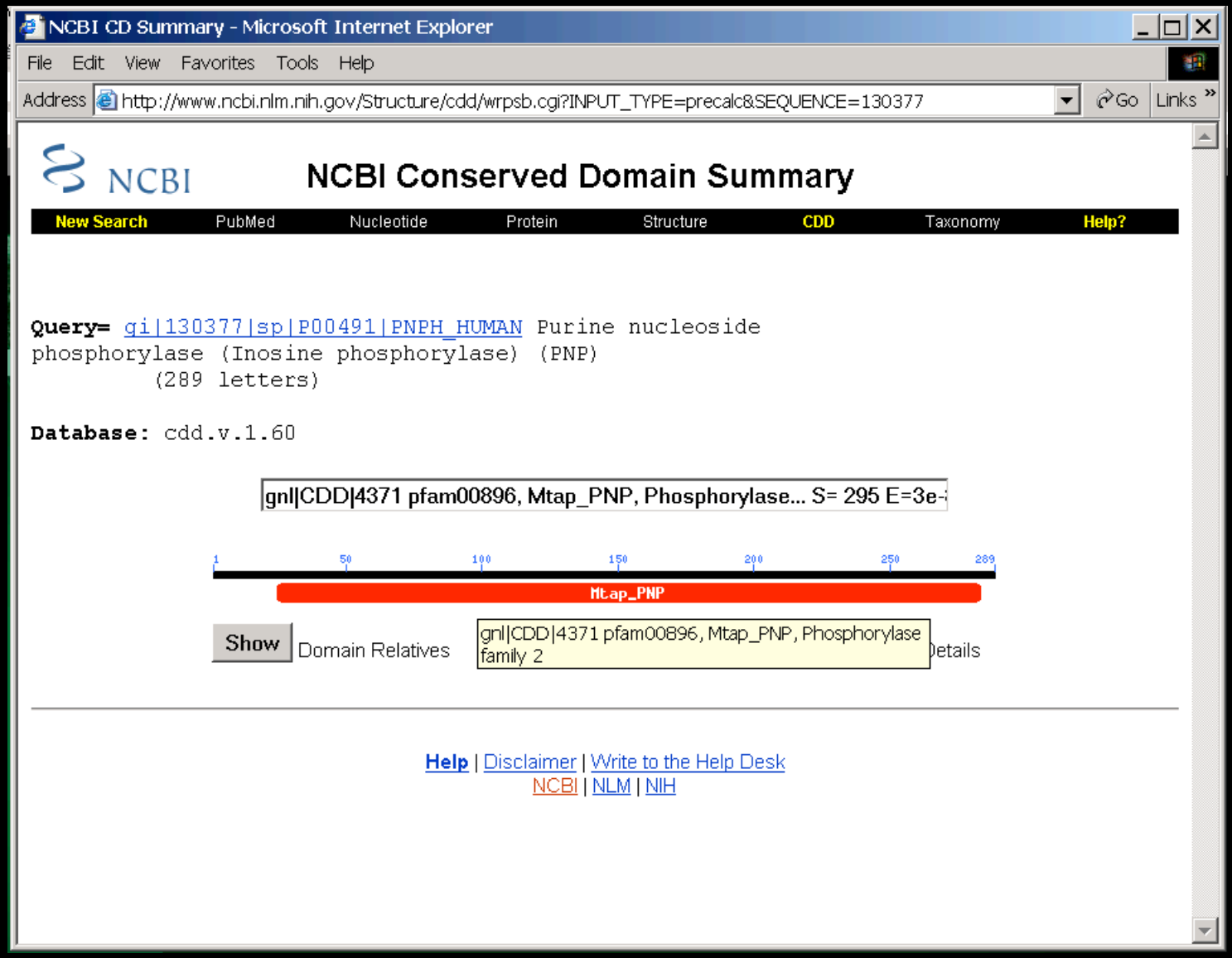
CFTR Search – Exon 10

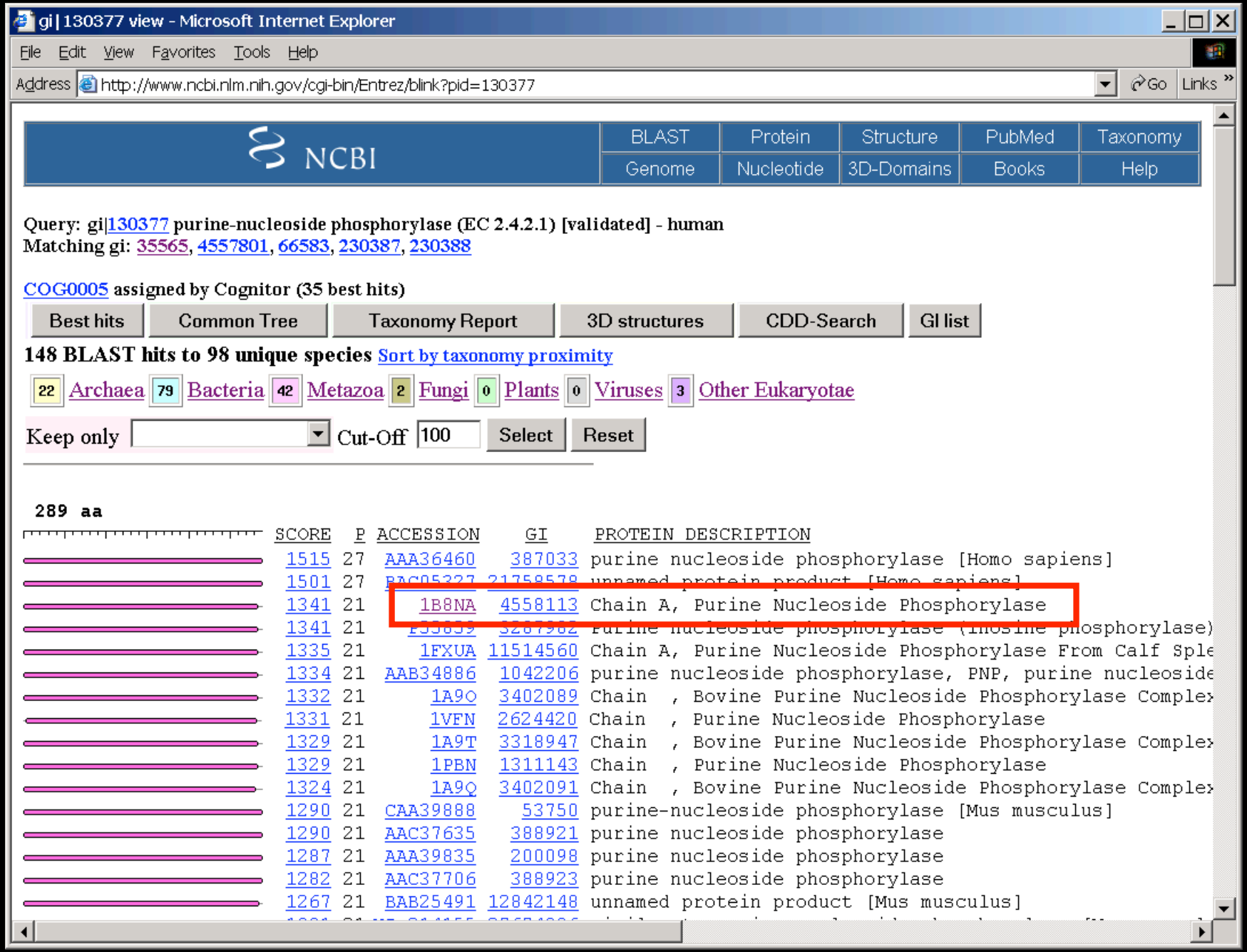


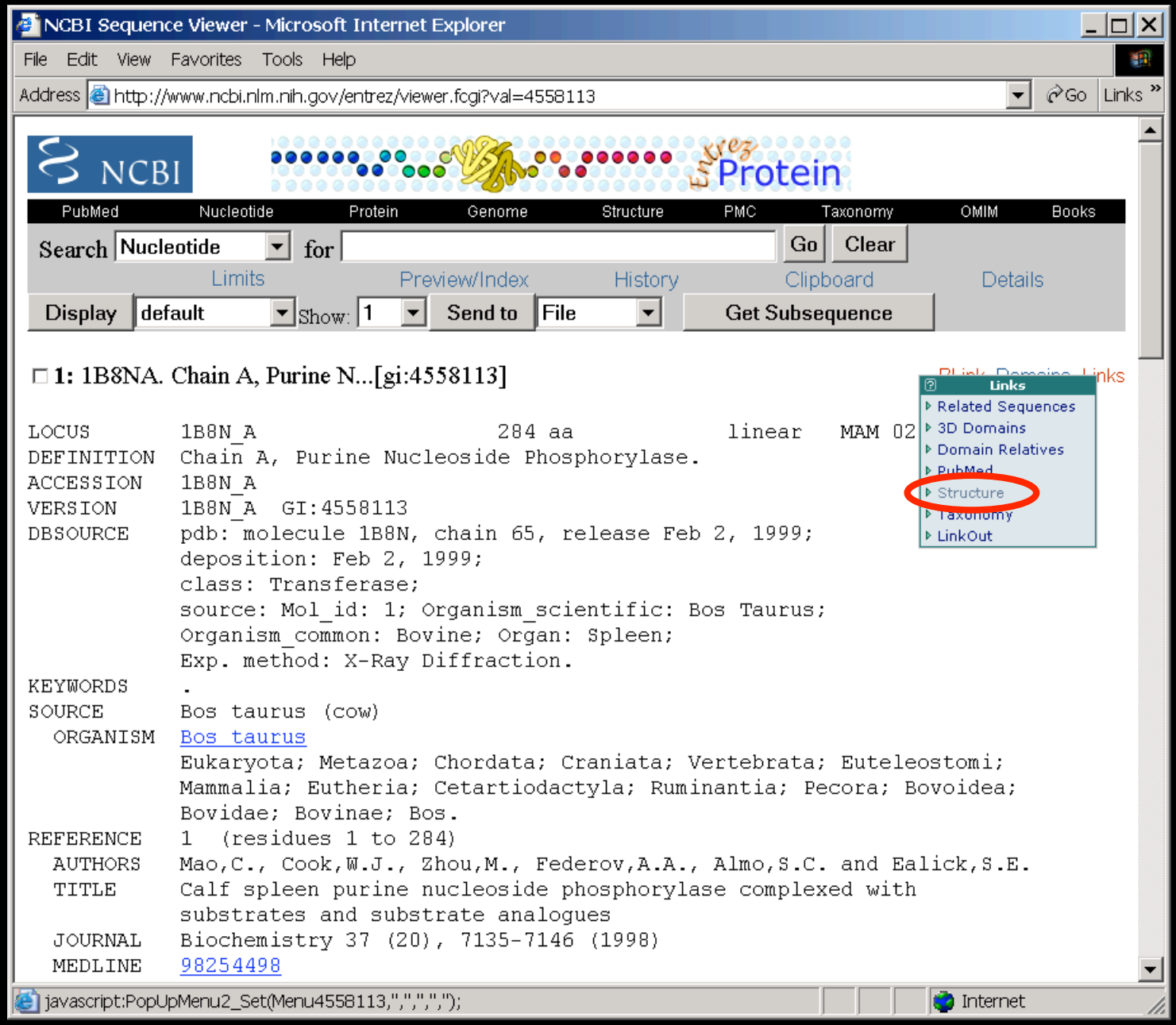


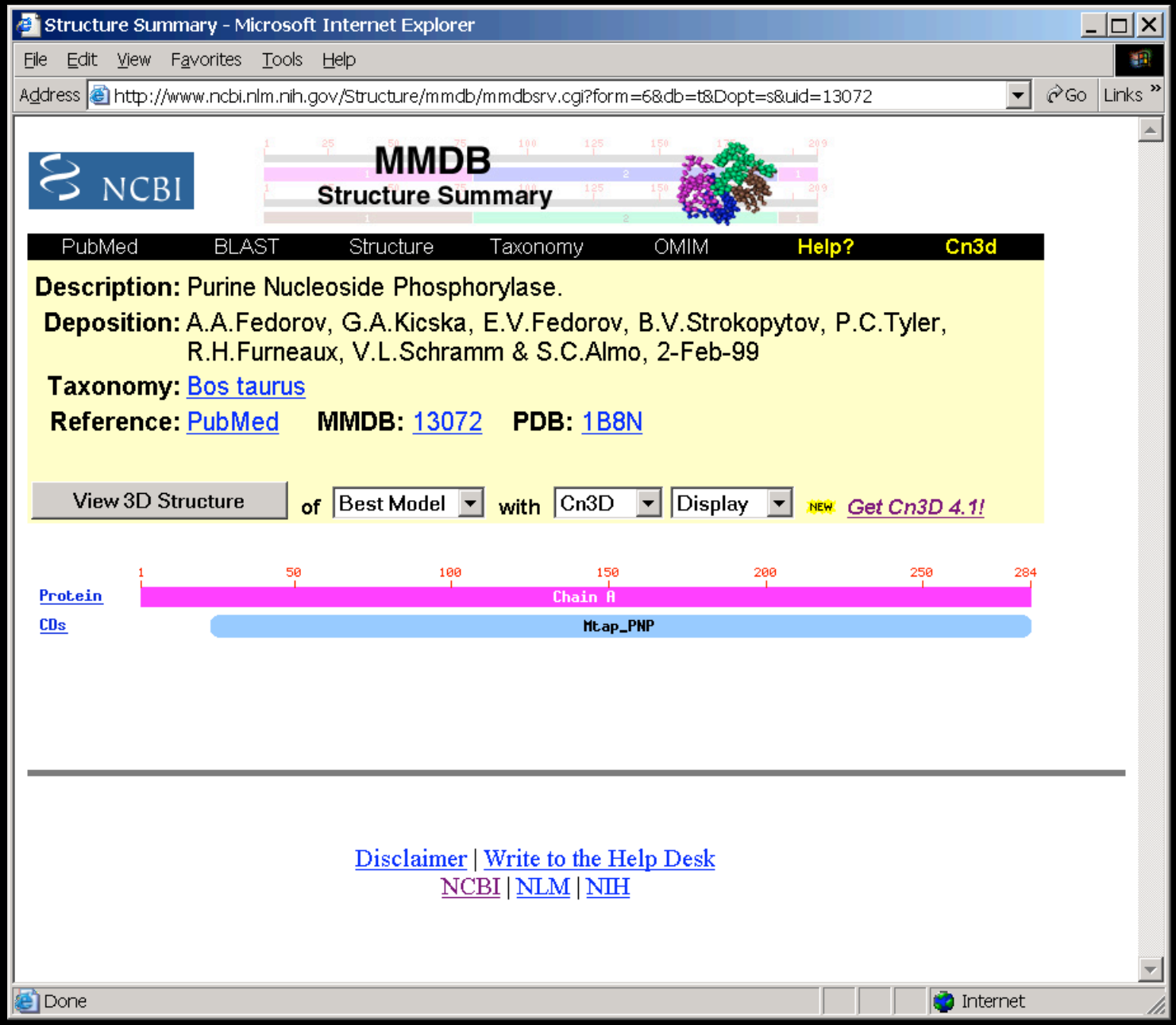

NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=130377&dopt=GenPept
Go Links
NCBI
Entrez Protein
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Protein for Go Clear
Limits Preview/Index History Clipboard Details
Display default Show: 20 Send to File Get Subsequence
1: P00491. Purine nucleoside...[gi:130377]
BLink, Domains, Links
LOCUS P00491 289 aa linear PRI 15-JUN-2002
DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP).
ACCESSION P00491
VERSION P00491 GI:130377
DBSOURCE swissprot: locus PNPH_HUMAN, accession P00491;
class: standard.
extra accessions:Q15160,created: Jul 21, 1986.
sequence updated: Jul 21, 1986.
annotation updated: Jun 15, 2002.
xrefs: gi: 35564, gi: 35565, gi: 190150, gi: 387033, gi: 190147,
gi: 190148, gi: 190149, gi: 66583, gi: 230387, gi: 230388
xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
164050, InterProIPR001369, PfamPF00896, PROSITEPS01240
KEYWORDS Transferase; Glycosyltransferase; Polymorphism; Disease mutation;
3D-structure.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 289)
AUTHORS Williams,S.R., Goddard,J.M. and Martin,D.W. Jr.
TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
clone characterization
JOURNAL Nucleic Acids Res. 12 (14), 5779-5787 (1984)
MEDLINE 84272252
PUBMED 6087295
REMARK SEQUENCE FROM N.A.
REFERENCE 2 (residues 1 to 289)
AUTHORS Williams,S.R., Gekeler,V., McIvor,R.S. and Martin,D.W. Jr.
TITLE A human purine nucleoside phosphorylase deficiency caused by a
Done
Internet


gi|130377 view - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/blink?pid=130377
Go Links
NCBI
BLAST Protein Structure PubMed Taxonomy
Genome Nucleotide 3D-Domains Books Help
Query: gi|130377 purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
Matching gi: 35565, 4557801, 66583, 230387, 230388
COG0005 assigned by Cognitor (35 best hits)
Best hits Common Tree Taxonomy Report 3D structures CDD-Search GI list
148 BLAST hits to 98 unique species Sort by taxonomy proximity
22 Archaea 79 Bacteria 42 Metazoa 2 Fungi 0 Plants 0 Viruses 3 Other Eukaryotae
Keep only Cut-Off 100 Select Reset
289 aa
SCORE P ACCESSION GI PROTEIN DESCRIPTION
1515 27 AAA36460 387033 purine nucleoside phosphorylase [Homo sapiens]
1501 27 BAC05327 21758578 unnamed protein product [Homo sapiens]
1341 21 1B8NA 4558113 Chain A, Purine Nucleoside Phosphorylase
1341 21 P55859 3287982 Purine nucleoside phosphorylase (Inosine phospho
1335 21 1FXUA 11514560 Chain A, Purine Nucleoside Phosphorylase From Ca
1334 21 AAB34886 1042206 purine nucleoside phosphorylase, PNP, purine nuc
1332 21 1A9O 3402089 Chain , Bovine Purine Nucleoside Phosphorylase (
1331 21 1VFN 2624420 Chain , Purine Nucleoside Phosphorylase
1329 21 1A9T 3318947 Chain , Bovine Purine Nucleoside Phosphorylase (
1329 21 1PBN 1311143 Chain , Purine Nucleoside Phosphorylase
1324 21 1A9Q 3402091 Chain , Bovine Purine Nucleoside Phosphorylase (
1290 21 CAA39888 53750 purine-nucleoside phosphorylase [Mus musculus]
1290 21 AAC37635 388921 purine nucleoside phosphorylase
1287 21 AAA39835 200098 purine nucleoside phosphorylase
1282 21 AAC37706 388923 purine nucleoside phosphorylase
1267 21 BAB25491 12842148 unnamed protein product [Mus musculus]
1001 21 XP_214155 27674996 similar to purine-nucleoside phopshorylase [Mus
814 8 EAA11700 21299555 agCP6049 [Anopheles gambiae str. PEST]
760 8 AAF47654 7292245 CG16758-PB [Drosophila melanogaster]


NCBI CD Summary - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=precalc&SEQUENCE=130377
Go
Links
NCBI
NCBI Conserved Domain Summary
New Search
PubMed
Nucleotide
Protein
Structure
CDD
Taxonomy
Help?
Query= gi|130377|sp|P00491|PNPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)
(289 letters)
Database: cdd.v.1.60
gnl|CDD|4371 pfam00896, Mtap_PNP, Phosphorylase... S= 295 E=3e-
1
50
100
150
200
250
289
Mtap_PNP
Show
Domain Relatives
gnl|CDD|4371 pfam00896, Mtap_PNP, Phosphorylase family 2
Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH


gi|130377 view - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/blink?pid=130377
Go Links
NCBI
BLAST Protein Structure PubMed Taxonomy
Genome Nucleotide 3D-Domains Books Help
Query: gi|130377 purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
Matching gi: 35565, 4557801, 66583, 230387, 230388
COG0005 assigned by Cognitor (35 best hits)
Best hits Common Tree Taxonomy Report 3D structures CDD-Search GI list
148 BLAST hits to 98 unique species Sort by taxonomy proximity
22 Archaea 79 Bacteria 42 Metazoa 2 Fungi 0 Plants 0 Viruses 3 Other Eukaryotae
Keep only Cut-Off 100 Select Reset
289 aa
SCORE P ACCESSION GI PROTEIN DESCRIPTION
1515 27 AAA36460 387033 purine nucleoside phosphorylase [Homo sapiens]
1501 27 BAC05327 21758578 unnamed protein product [Homo sapiens]
1341 21 1B8NA 4558113 Chain A, Purine Nucleoside Phosphorylase
1341 21 P33859 3287982 Purine nucleoside phosphorylase (Inosine phosphorylase)
1335 21 1FXUA 11514560 Chain A, Purine Nucleoside Phosphorylase From Calf Sple
1334 21 AAB34886 1042206 purine nucleoside phosphorylase, PNP, purine nucleoside
1332 21 1A9O 3402089 Chain , Bovine Purine Nucleoside Phosphorylase Complex
1331 21 1VFN 2624420 Chain , Purine Nucleoside Phosphorylase
1329 21 1A9T 3318947 Chain , Bovine Purine Nucleoside Phosphorylase Complex
1329 21 1PBN 1311143 Chain , Purine Nucleoside Phosphorylase
1324 21 1A9Q 3402091 Chain , Bovine Purine Nucleoside Phosphorylase Complex
1290 21 CAA39888 53750 purine-nucleoside phosphorylase [Mus musculus]
1290 21 AAC37635 388921 purine nucleoside phosphorylase
1287 21 AAA39835 200098 purine nucleoside phosphorylase
1282 21 AAC37706 388923 purine nucleoside phosphorylase
1267 21 BAB25491 12842148 unnamed protein product [Mus musculus]


NCBI Sequence Viewer - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=4558113
Go Links
NCBI
Entrez Protein
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Display default Show: 1 Send to File Get Subsequence
1: 1B8NA. Chain A, Purine N...[gi:4558113]
Links
Related Sequences
3D Domains
Domain Relatives
PubMed
Structure
Taxonomy
LinkOut
LOCUS 1B8N_A 284 aa linear MAM 02
DEFINITION Chain A, Purine Nucleoside Phosphorylase.
ACCESSION 1B8N_A
VERSION 1B8N_A GI:4558113
DBSOURCE pdb: molecule 1B8N, chain 65, release Feb 2, 1999;
deposition: Feb 2, 1999;
class: Transferase;
source: Mol_id: 1; Organism_scientific: Bos Taurus;
Organism_common: Bovine; Organ: Spleen;
Exp. method: X-Ray Diffraction.
KEYWORDS .
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
REFERENCE 1 (residues 1 to 284)
AUTHORS Mao,C., Cook,W.J., Zhou,M., Federov,A.A., Almo,S.C. and Ealick,S.E.
TITLE Calf spleen purine nucleoside phosphorylase complexed with
substrates and substrate analogues
JOURNAL Biochemistry 37 (20), 7135-7146 (1998)
MEDLINE 98254498
javascript:PopUpMenu2_Set(Menu4558113,",",",",",");
Internet


Structure Summary - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?form=6&db=t&Dopt=s&uid=13072
Go Links
NCBI
MMDB
Structure Summary
PubMed BLAST Structure Taxonomy OMIM Help? Cn3d
Description: Purine Nucleoside Phosphorylase.
Deposition: A.A.Fedorov, G.A.Kicska, E.V.Fedorov, B.V.Strokopytov, P.C.Tyler, R.H.Furneaux, V.L.Schramm & S.C.Almo, 2-Feb-99
Taxonomy: Bos taurus
Reference: PubMed MMDB: 13072 PDB: 1B8N
View 3D Structure of Best Model with Cn3D Display NEW Get Cn3D 4.1!
1 50 100 150 200 250 284
Protein
Chain A
CDs
Mtap_PNP
Disclaimer | Write to the Help Desk
NCBI | NLM | NIH
Done
Internet


Ma

Sequence Similarity Searching

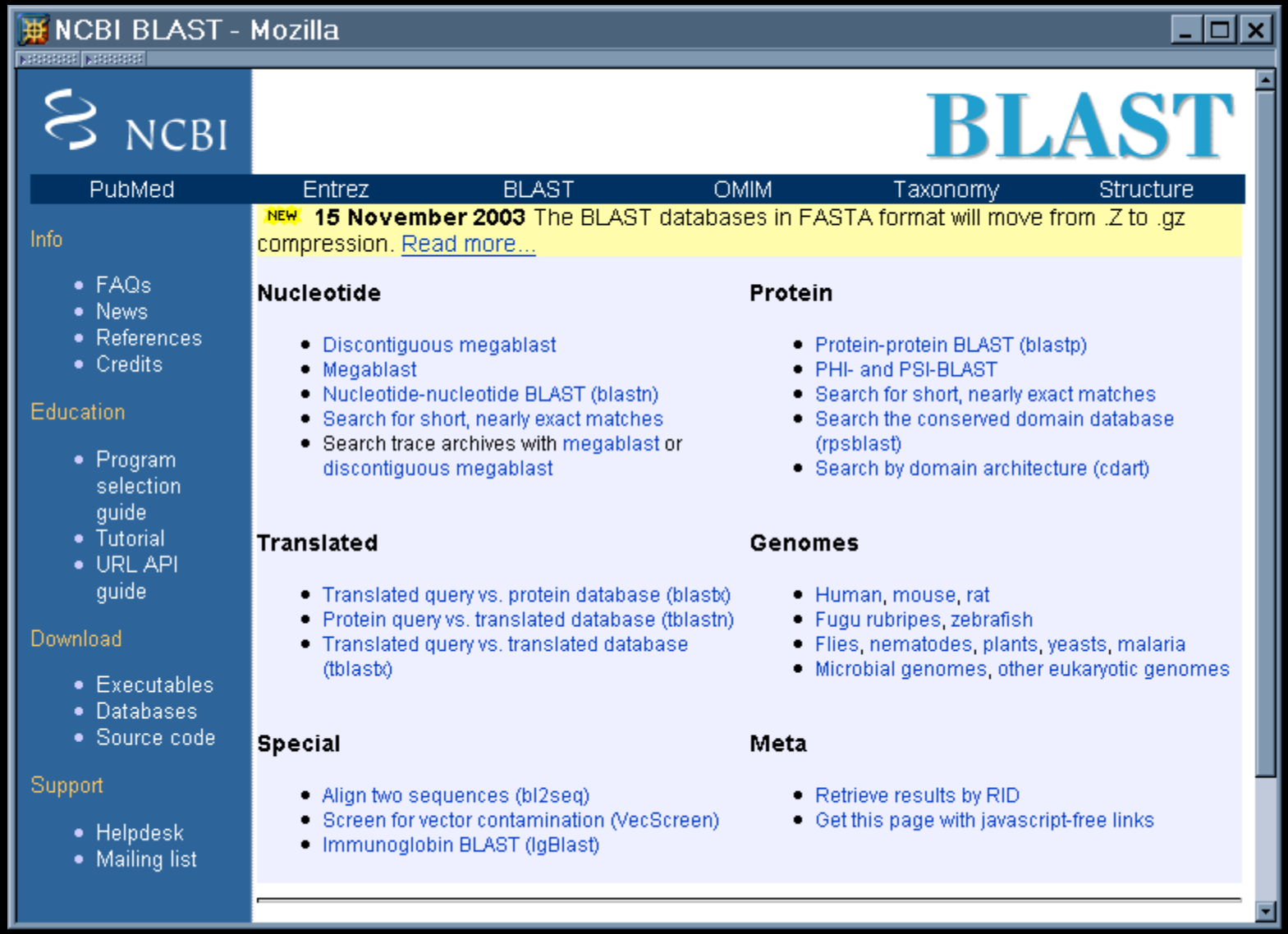
What other sequences have some primary sequence similarity to my query sequence?

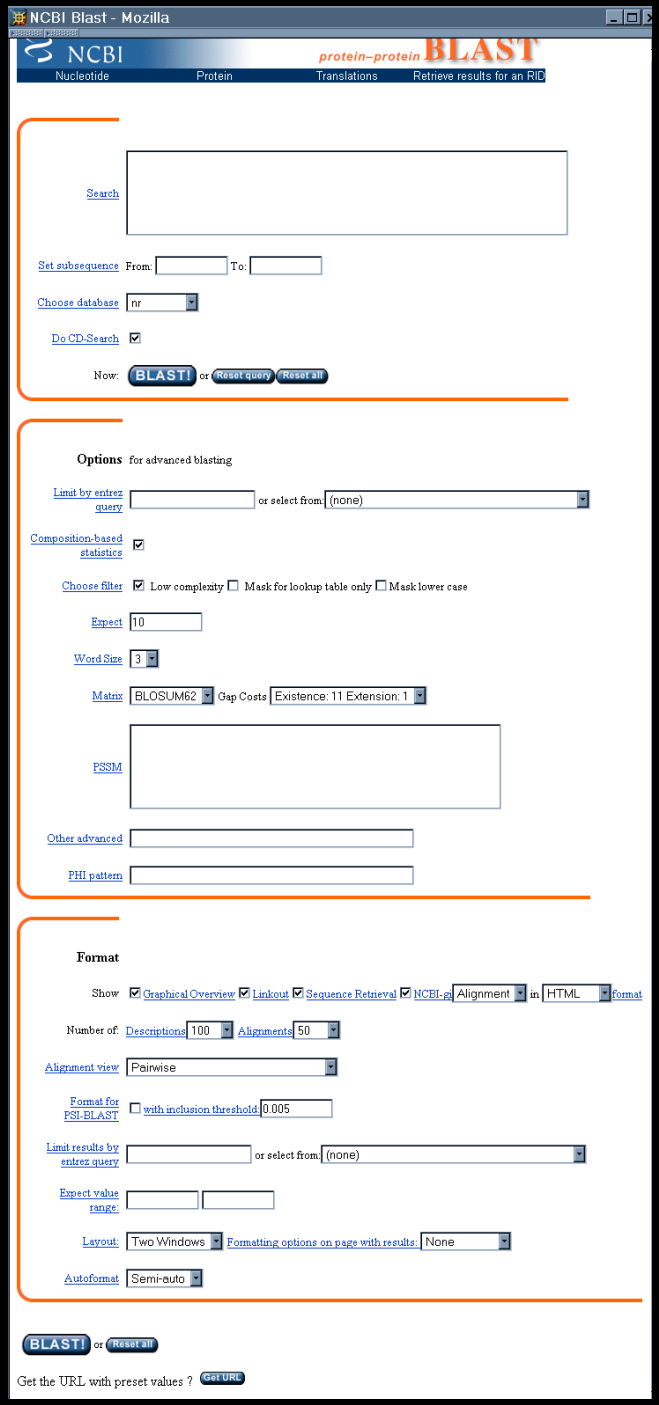
BLAST

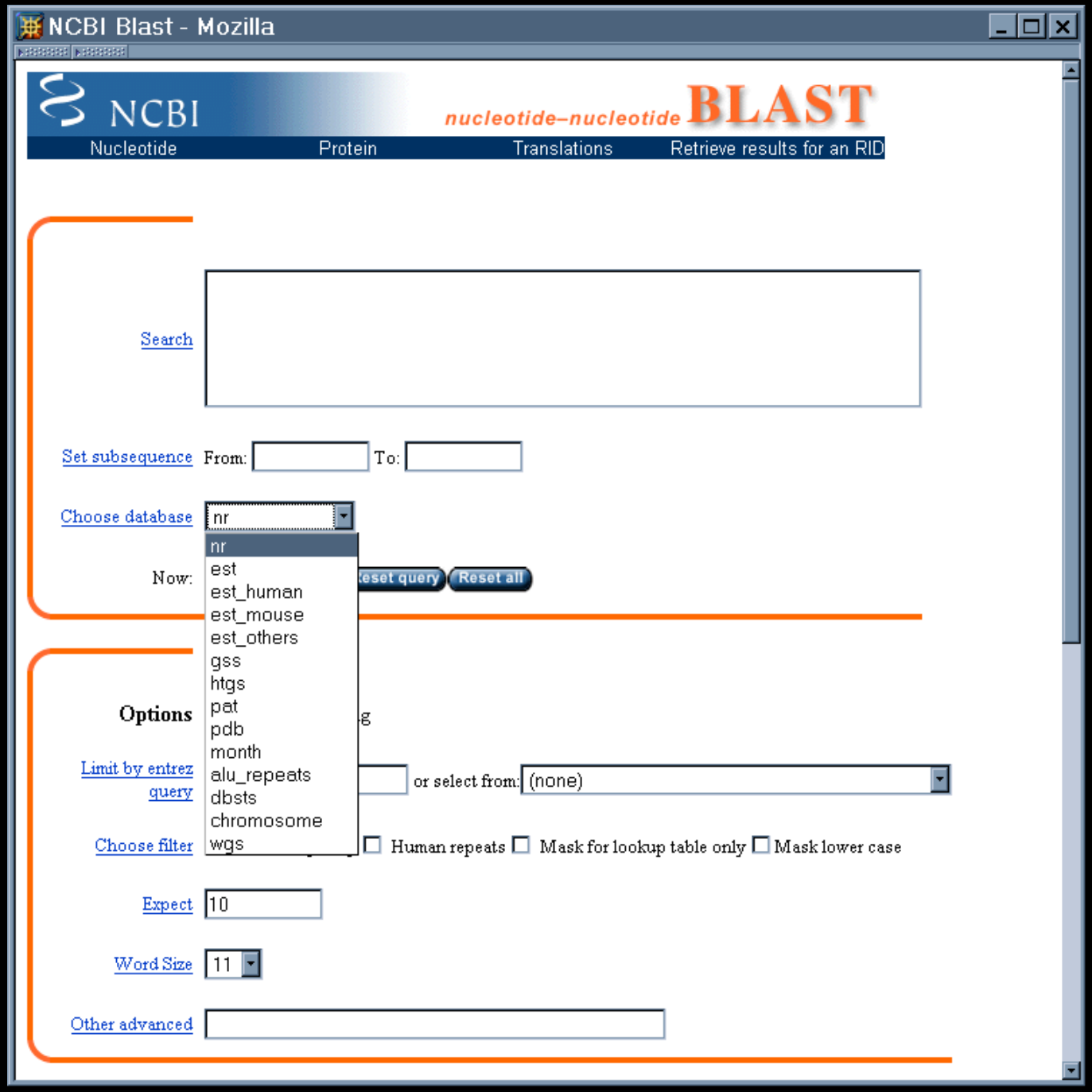
- Basic Local Alignment Search Tool
- Search a sequence database for primary sequence similarities to some query sequence
- Provides a measure of the significance of the similarity
- Does not necessarily imply common evolutionary origin

BLAST

- All search combinations possible
- nt vs. nt database
  - blastn
- protein vs. protein database
  - blastp
- translated nt vs. protein database
  - blastx
- protein vs. translated nt database
  - tblastn
- translated nt vs. translated nt database
  - tblastx


NCBI BLAST - Mozilla
NCBI
BLAST
PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure
NEW 15 November 2003 The BLAST databases in FASTA format will move from .Z to .gz compression. Read more...
Info
FAQs
News
References
Credits
Education
Program selection guide
Tutorial
URL API guide
Download
Executables
Databases
Source code
Support
Helpdesk
Mailing list
Nucleotide
Discontiguous megablast
Megablast
Nucleotide-nucleotide BLAST (blastn)
Search for short, nearly exact matches
Search trace archives with megablast or discontiguous megablast
Protein
Protein-protein BLAST (blastp)
PHI- and PSI-BLAST
Search for short, nearly exact matches
Search the conserved domain database (rpsblast)
Search by domain architecture (cdart)
Translated
Translated query vs. protein database (blastx)
Protein query vs. translated database (tblastn)
Translated query vs. translated database (tblastx)
Genomes
Human, mouse, rat
Fugu rubripes, zebrafish
Flies, nematodes, plants, yeasts, malaria
Microbial genomes, other eukaryotic genomes
Special
Align two sequences (bl2seq)
Screen for vector contamination (VecScreen)
Immunoglobin BLAST (IgBlast)
Meta
Retrieve results by RID
Get this page with javascript-free links


NCBI Blast - Mozilla
NCBI
protein–protein BLAST
Nucleotide
Protein
Translations
Retrieve results for an RID
Search
Set subsequence
From:
To:
Choose database
nr
Do CD-Search
Now:
BLAST!
or
Reset query
Reset all
Options
for advanced blasting
Limit by entrez query
or select from:
(none)
Composition-based statistics
Choose filter
Low complexity
Mask for lookup table only
Mask lower case
Expect
10
Word Size
3
Matrix
BLOSUM62
Gap Costs
Existence: 11 Extension: 1
PSSM
Other advanced
PHI pattern
Format
Show
Graphical Overview
Linkout
Sequence Retrieval
NCBI-gi
Alignment
in
HTML
format
Number of:
Descriptions
100
Alignments
50
Alignment view
Pairwise
Format for PSI-BLAST
with inclusion threshold:
0.005
Limit results by entrez query
or select from:
(none)
Expect value range:
Layout:
Two Windows
Formatting options on page with results:
None
Autoformat
Semi-auto
BLAST!
or
Reset all
Get the URL with preset values ?
Get URL


NCBI Blast - Mozilla
NCBI
nucleotide–nucleotide BLAST
Nucleotide
Protein
Translations
Retrieve results for an RID
Search
Set subsequence
From:
To:
Choose database
nr
Now:
Reset query
Reset all
nr
est
est_human
est_mouse
est_others
gss
htgs
pat
pdb
month
alu_repeats
dbsts
chromosome
wgs
Options
Limit by entrez query
or select from:
(none)
Choose filter
Human repeats
Mask for lookup table only
Mask lower case
Expect
10
Word Size
11
Other advanced

RID=1070653888-8986-72268071872.BLASTQ3, cftr_human - Mozilla

NCBI *results of* **BLAST**

BLASTP 2.2.6 [Apr-09-2003]

Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

RID: 1070653888-8986-72268071872.BLASTQ3

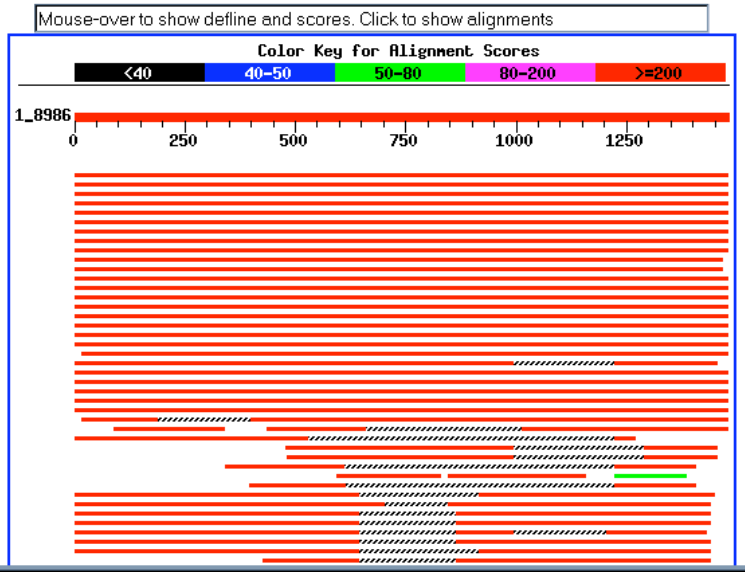
Query= cftr_human
(1480 letters)

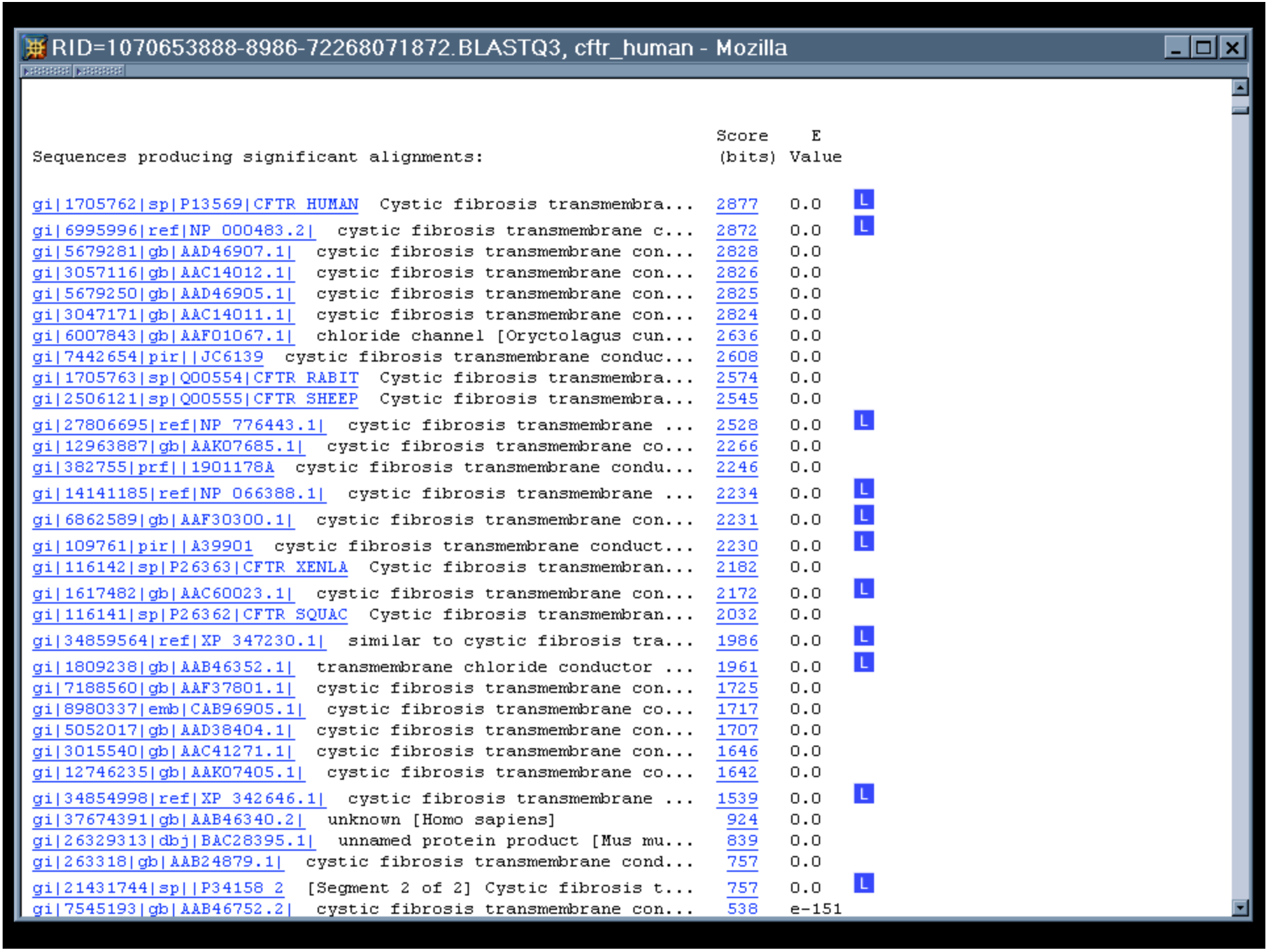
Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,550,899 sequences; 506,521,537 total letters

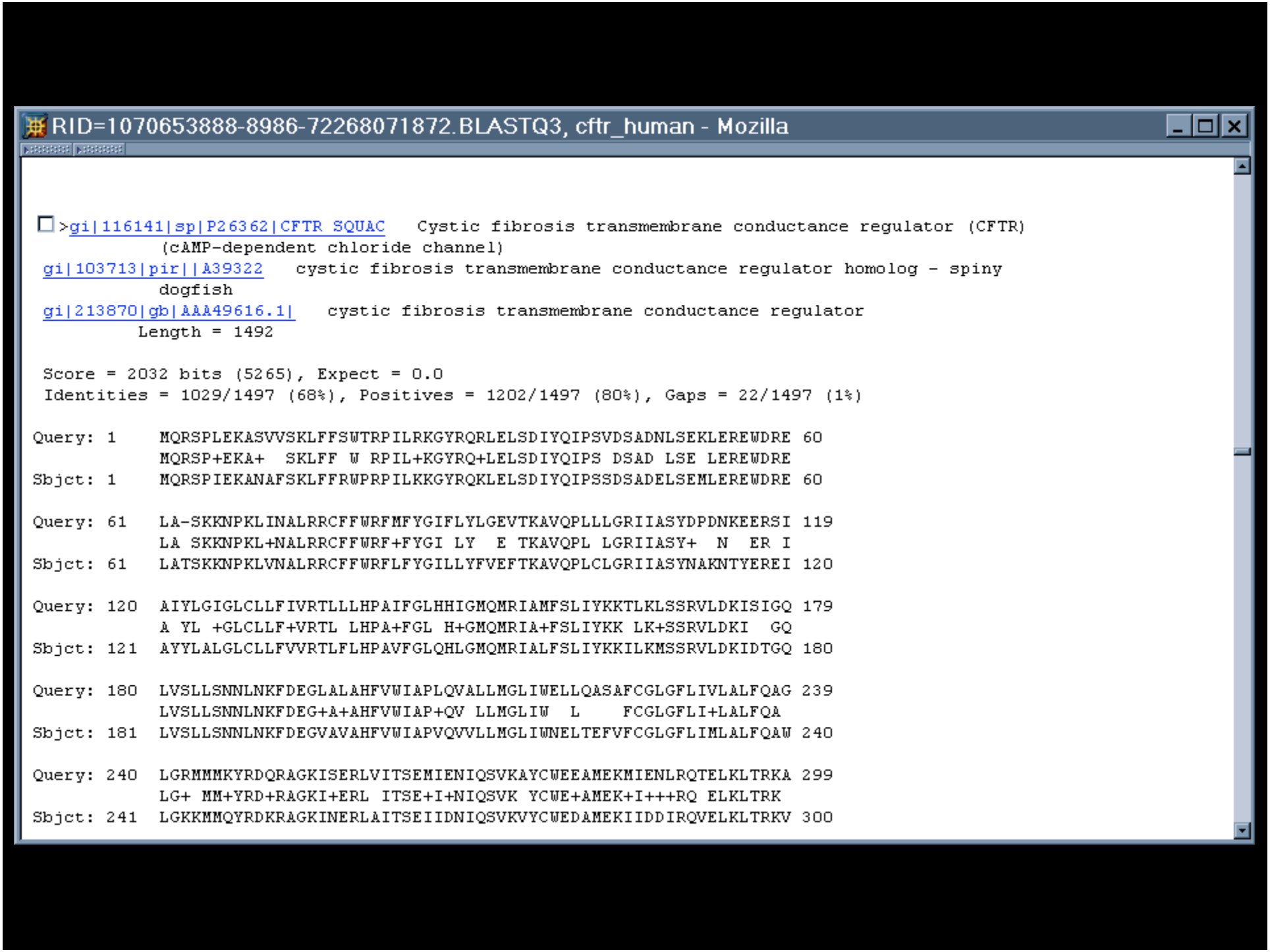
If you have any problems or questions with the results of this search
please refer to the **BLAST FAQs**

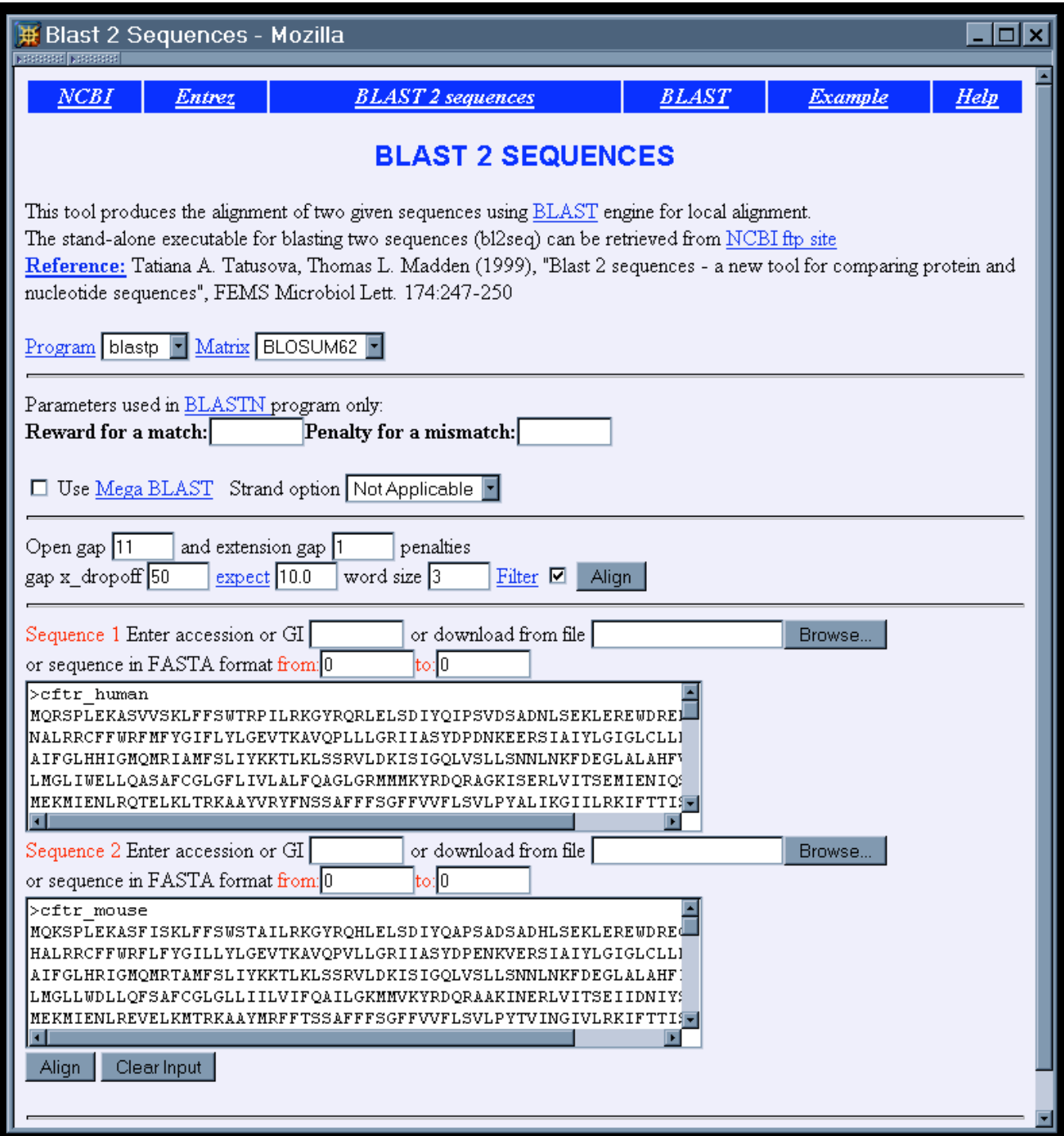
Taxonomy reports

Distribution of 100 Blast Hits on the Query Sequence




RID=1070653888-8986-72268071872.BLASTQ3, cftr_human - Mozilla
Sequences producing significant alignments:
Score (bits)
E Value
gi|1705762|sp|P13569|CFTR_HUMAN Cystic fibrosis transmembra... 2877 0.0 L
gi|6995996|ref|NP_000483.2| cystic fibrosis transmembrane c... 2872 0.0 L
gi|5679281|gb|AAD46907.1| cystic fibrosis transmembrane con... 2828 0.0
gi|3057116|gb|AAC14012.1| cystic fibrosis transmembrane con... 2826 0.0
gi|5679250|gb|AAD46905.1| cystic fibrosis transmembrane con... 2825 0.0
gi|3047171|gb|AAC14011.1| cystic fibrosis transmembrane con... 2824 0.0
gi|6007843|gb|AAF01067.1| chloride channel [Oryctolagus cun... 2636 0.0
gi|7442654|pir||JC6139 cystic fibrosis transmembrane conduc... 2608 0.0
gi|1705763|sp|Q00554|CFTR_RABIT Cystic fibrosis transmembra... 2574 0.0
gi|2506121|sp|Q00555|CFTR_SHEEP Cystic fibrosis transmembra... 2545 0.0
gi|27806695|ref|NP_776443.1| cystic fibrosis transmembrane ... 2528 0.0 L
gi|12963887|gb|AAK07685.1| cystic fibrosis transmembrane co... 2266 0.0
gi|382755|prf||1901178A cystic fibrosis transmembrane condu... 2246 0.0
gi|14141185|ref|NP_066388.1| cystic fibrosis transmembrane ... 2234 0.0 L
gi|6862589|gb|AAF30300.1| cystic fibrosis transmembrane con... 2231 0.0 L
gi|109761|pir||A39901 cystic fibrosis transmembrane conduct... 2230 0.0 L
gi|116142|sp|P26363|CFTR_XENLA Cystic fibrosis transmembran... 2182 0.0
gi|1617482|gb|AAC60023.1| cystic fibrosis transmembrane con... 2172 0.0 L
gi|116141|sp|P26362|CFTR_SQUAC Cystic fibrosis transmembran... 2032 0.0
gi|34859564|ref|XP_347230.1| similar to cystic fibrosis tra... 1986 0.0 L
gi|1809238|gb|AAB46352.1| transmembrane chloride conductor ... 1961 0.0 L
gi|7188560|gb|AAF37801.1| cystic fibrosis transmembrane con... 1725 0.0
gi|8980337|emb|CAB96905.1| cystic fibrosis transmembrane co... 1717 0.0
gi|5052017|gb|AAD38404.1| cystic fibrosis transmembrane con... 1707 0.0
gi|3015540|gb|AAC41271.1| cystic fibrosis transmembrane con... 1646 0.0
gi|12746235|gb|AAK07405.1| cystic fibrosis transmembrane co... 1642 0.0
gi|34854998|ref|XP_342646.1| cystic fibrosis transmembrane ... 1539 0.0 L
gi|37674391|gb|AAB46340.2| unknown [Homo sapiens] 924 0.0
gi|26329313|dbj|BAC28395.1| unnamed protein product [Mus mu... 839 0.0
gi|263318|gb|AAB24879.1| cystic fibrosis transmembrane cond... 757 0.0
gi|21431744|sp||P34158_2 [Segment 2 of 2] Cystic fibrosis t... 757 0.0 L
gi|7545193|gb|AAB46752.2| cystic fibrosis transmembrane con... 538 e-151


RID=1070653888-8986-72268071872.BLASTQ3, cftr_human - Mozilla
>gi|116141|sp|P26362|CFTR SQUAC Cystic fibrosis transmembrane conductance regulator (CFTR)
(cAMP-dependent chloride channel)
gi|103713|pir||A39322 cystic fibrosis transmembrane conductance regulator homolog - spiny
dogfish
gi|213870|gb|AAA49616.1| cystic fibrosis transmembrane conductance regulator
Length = 1492
Score = 2032 bits (5265), Expect = 0.0
Identities = 1029/1497 (68%), Positives = 1202/1497 (80%), Gaps = 22/1497 (1%)
Query: 1 MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRE 60
MQRSP+EKA+ SKLFF W RPIL+KGYRQ+LELSDIYQIPS DSAD LSE LEREWDRE
Sbjct: 1 MQRSPIEKANAFSKLFFRWPRPILKKGYRQKLELSDIYQIPSSDSADELSEMLEREWDRE 60
Query: 61 LA-SKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSI 119
LA SKKNPKL+NALRRCFFWRF+FYGI LY E TKAVQPL LGRIIASY+ N ER I
Sbjct: 61 LATSKKNPKLVNALRRCFFWRFLFYGILLYFVEFTKAVQPLCLGRIIASYNAKNTYEREI 120
Query: 120 AIYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQ 179
A YL +GLCLLF+VRTL LHPA+FGL H+GMQMRIA+FSLIYKK LK+SSRVLDKI GQ
Sbjct: 121 AYYLALGLCLLFVVRTLFLHPAVFGLQHLGMQMRIALFSLIYKKILKMSSRVLDKIDTGQ 180
Query: 180 LVSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAG 239
LVSLLSNNLNKFDEG+A+AHFVWIAP+QV LLMGLIW L FCGLGFLI+LALFQA
Sbjct: 181 LVSLLSNNLNKFDEGVAVAHFVWIAPVQVVLLMGLIWNELTEFVFCGLGFLIMLALFQAW 240
Query: 240 LGRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKA 299
LG+ MM+YRD+RAGKI+ERL ITSE+I+NIQSVK YCWE+AMEK+I+++RQ ELKLTRK
Sbjct: 241 LGKKMMQYRDKRAGKINERLAITSEIIDNIQSVKVYCWEDAMEKIIDDIRQVELKLTRKV 300


Blast 2 Sequences - Mozilla
NCBI
Entrez
BLAST 2 sequences
BLAST
Example
Help
BLAST 2 SEQUENCES
This tool produces the alignment of two given sequences using BLAST engine for local alignment.
The stand-alone executable for blasting two sequences (bl2seq) can be retrieved from NCBI ftp site
Reference: Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250
Program blastp Matrix BLOSUM62
Parameters used in BLASTN program only:
Reward for a match: Penalty for a mismatch:
Use Mega BLAST Strand option Not Applicable
Open gap 11 and extension gap 1 penalties
gap x_dropoff 50 expect 10.0 word size 3 Filter Align
Sequence 1 Enter accession or GI or download from file Browse...
or sequence in FASTA format from: 0 to: 0
>cftr_human
MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRE
NALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIAIYLGIGLCLL
AIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQLVSLLSNNLNKFDEGLALAHF
LMGLIWELLQASAFCGLGFLIVLALFQAGLGRMMMKYRDQRAGKISERLVITSEMIENIQ
MEKMIENLRQTELKLTRKAAYVRYFNSSAFFFSGFFVVFLSVLPYALIKGIILRKIFTTI
Sequence 2 Enter accession or GI or download from file Browse...
or sequence in FASTA format from: 0 to: 0
>cftr_mouse
MQKSPLEKASFISKLFFSWSTAILRKGYRQHLELSDIYQAPSADSADHLSEKLEREWDRE
HALRRCFFWRFLFYGILLYLGEVTKAVQPVLLGRIIASYDPENKVERSIAIYLGIGLCLL
AIFGLHRIGMQMRTAMFSLIYKKTLKLSSRVLDKISIGQLVSLLSNNLNKFDEGLALAHF
LMGLLWDLLQFSAFCGLGLLIILVIFQAILGKMMVKYRDQRAAKINERLVITSEIIDNIY
MEKMIENLREVELKMTRKAAYMRFFTSSAFFFSGFFVVFLSVLPYTVINGIVLRKIFTTI
Align Clear Input

Blast Result - Mozilla

NCBI **Blast 2 Sequences results**

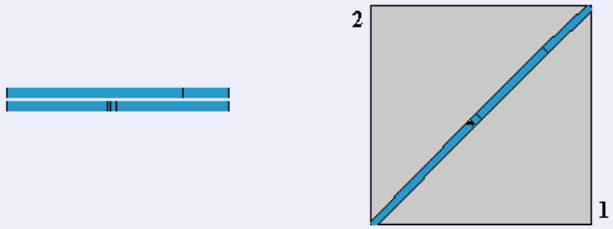
PubMed Entrez BLAST OMIM Taxonomy Structure

BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.6 [Apr-09-2003]

Matrix: BLOSUM62 gap open: 11 gap extension: 1
x_dropoff: 50 expect: 10.000 wordsize: 3 Filter ☑ Align

Sequence 1 lcl|1_cftr_human **Length** 1480 (1 .. 1480)

Sequence 2 lcl|2_cftr_mouse **Length** 1476 (1 .. 1476)



NOTE:The statistics (bitscore and expect value) is calculated based on the size of nr database

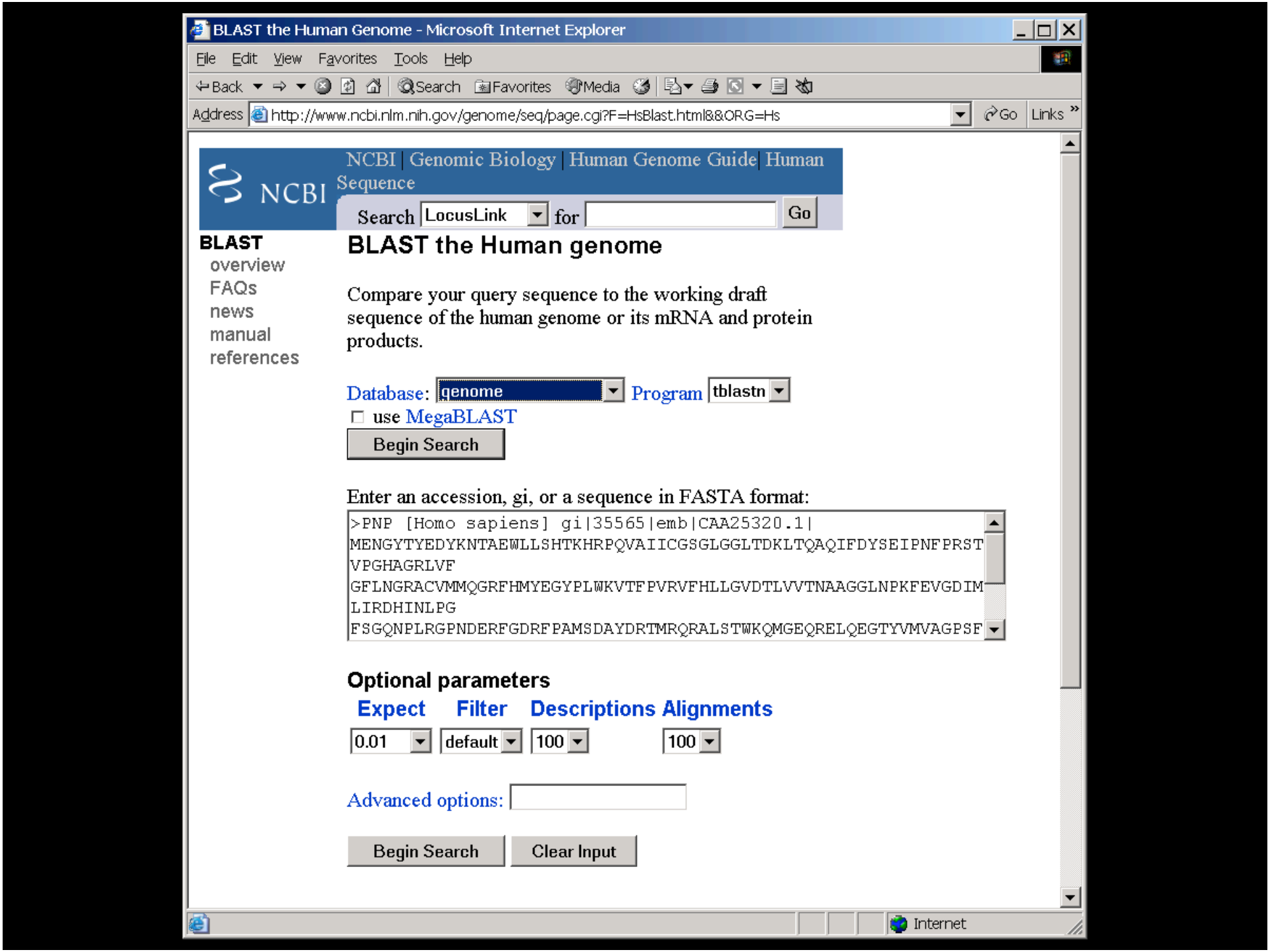
```
Score = 2228 bits (5774), Expect = 0.0
Identities = 1119/1481 (75%), Positives = 1270/1481 (85%), Gaps = 6/1481 (0%)
```

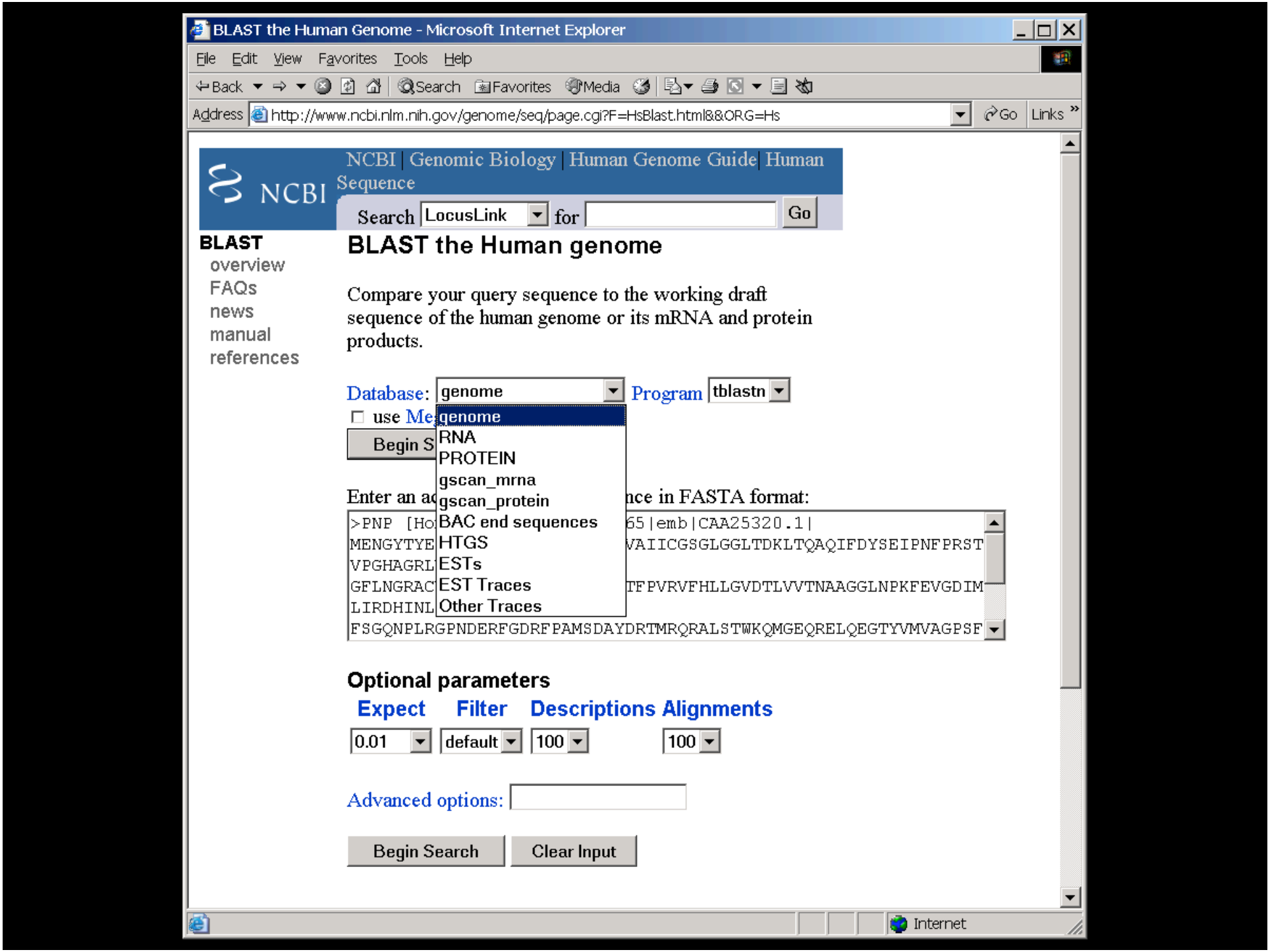
```
Query: 1    MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLELSDIYQIPSVDSADNLSEKLEREWDRE 60
            MQ+SPLEKAS +SKLFFSW+  ILRKGYRQ LELSDIYQ PS DSAD+LSEKLEREWDRE
Sbjct: 1    MQKSPLEKASFISKLFFSWSTAILRKGYRQHLELSDIYQAPSADSADHLSEKLEREWDRE 60

Query: 61   LASKKNPKLINALRRCFFWRFMFYGIFLYLGEVTKAVQPLLLGRIIASYDPDNKEERSIA 120
             ASKKNP+LI+ALRRCFFWRF+FYGI LYLGEVTKAVQP+LLGRIIASYDP+NK ERSIA
Sbjct: 61   QASKKNPQLIHALRRCFFWRFLFYGILLYLGEVTKAVQPVLLGRIIASYDPENKVERSIA 120

Query: 121  IYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSSRVLDKISIGQL 180
            IYLGIGLCLLFIVRTLLLHPAIFGLH IGMQMR AMFSLIYKKTLKLSSRVLDKISIGQL
Sbjct: 121  IYLGIGLCLLFIVRTLLLHPAIFGLHRIGMQMRTAMFSLIYKKTLKLSSRVLDKISIGQL 180

Query: 181  VSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWELLQASAFCGLGFLIVLALFQAGL 240
            VSLLSNNLNKFDEGLALAHF+WIAPLQV LLMGL+W+LLQ SAFCGLG LI+L +FQA L
Sbjct: 181  VSLLSNNLNKFDEGLALAHFIWIAPLQVTLLMGLLWDLLQFSAFCGLGLLIILVIFQAIL 240

Query: 241  GRMMMKYRDQRAGKISERLVITSEMIENIQSVKAYCWEEAMEKMIENLRQTELKLTRKAA 300
            G+MM+KYRDQRA KI+ERLVITSE+I+NI SVKAYCWE AMEKMIENLR+ ELK+TRKAA
Sbjct: 241  GKMMVKYRDQRAAKINERLVITSEIIDNIYSVKAYCWESAMEKMIENLREVELKMTRKAA 300
```

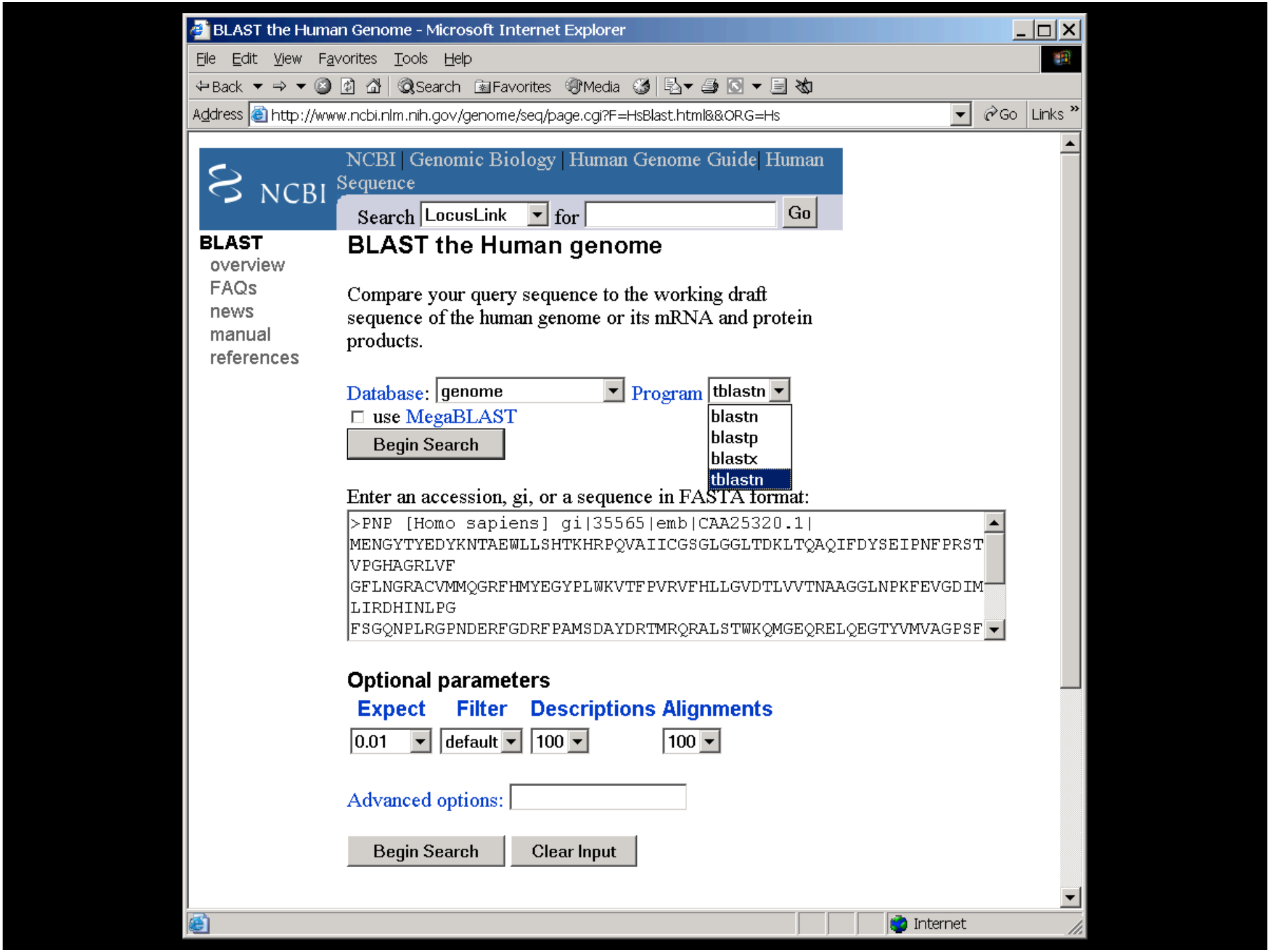
Genomic Biology

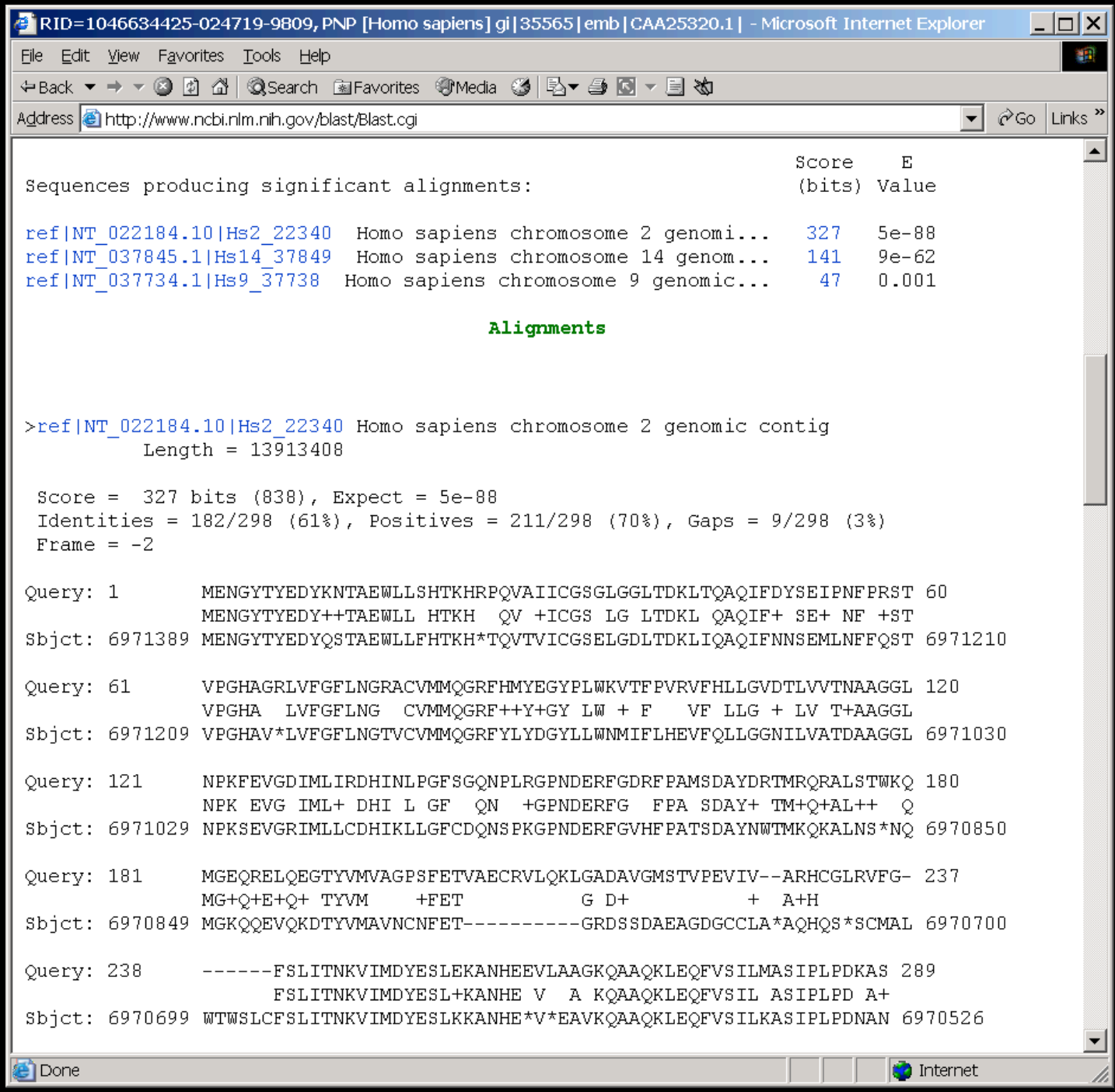
Searching Genomic Sequences

- Where is my sequence located in the human genome?
  - Chromosome; band; mapping data
  - Genetic linkage relationships
- What is the genomic context of my sequence?
  - Alternative splicing
  - Regulation
- Are there any paralogs?
- Are there any pseudogenes?
- Comparative analysis with the same gene in other genomes


BLAST the Human Genome - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs
Go Links
NCBI
NCBI | Genomic Biology | Human Genome Guide | Human Sequence
Search LocusLink for Go
BLAST
overview
FAQs
news
manual
references
BLAST the Human genome
Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.
Database: genome Program tblastn
use MegaBLAST
Begin Search
Enter an accession, gi, or a sequence in FASTA format:
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST
VPGHAGRLVF
GFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM
LIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQMGEQRELQEGTYVMVAGPSF
Optional parameters
Expect Filter Descriptions Alignments
0.01 default 100 100
Advanced options:
Begin Search Clear Input
Internet


BLAST the Human Genome - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs
Go Links
NCBI
NCBI | Genomic Biology | Human Genome Guide | Human Sequence
Search LocusLink for Go
BLAST
overview
FAQs
news
manual
references
BLAST the Human genome
Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.
Database: genome
Program tblastn
use Me
Begin S
genome
RNA
PROTEIN
gscan_mrna
gscan_protein
BAC end sequences
HTGS
ESTs
EST Traces
Other Traces
Enter an ac nce in FASTA format:
>PNP [Ho 65|emb|CAA25320.1|
MENGYTYE VAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST
VPGHAGRL
GFLNGRAC TFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM
LIRDHINL
FSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQMGEQRELQEGTYVMVAGPSF
Optional parameters
Expect Filter Descriptions Alignments
0.01 default 100 100
Advanced options:
Begin Search Clear Input
Internet


BLAST the Human Genome - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs
Go Links
NCBI
NCBI | Genomic Biology | Human Genome Guide | Human Sequence
Search LocusLink for Go
BLAST
overview
FAQs
news
manual
references
BLAST the Human genome
Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.
Database: genome
Program tblastn
blastn
blastp
blastx
tblastn
use MegaBLAST
Begin Search
Enter an accession, gi, or a sequence in FASTA format:
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST
VPGHAGRLVF
GFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIM
LIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQMGEQRELQEGTYVMVAGPSF
Optional parameters
Expect Filter Descriptions Alignments
0.01 default 100 100
Advanced options:
Begin Search Clear Input
Internet


RID=1046634425-024719-9809, PNP [Homo sapiens] gi|35565|emb|CAA25320.1| - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
Go Links
Sequences producing significant alignments: Score (bits) E Value
ref|NT_022184.10|Hs2_22340 Homo sapiens chromosome 2 genomi... 327 5e-88
ref|NT_037845.1|Hs14_37849 Homo sapiens chromosome 14 genom... 141 9e-62
ref|NT_037734.1|Hs9_37738 Homo sapiens chromosome 9 genomic... 47 0.001
Alignments
>ref|NT_022184.10|Hs2_22340 Homo sapiens chromosome 2 genomic contig
Length = 13913408
Score = 327 bits (838), Expect = 5e-88
Identities = 182/298 (61%), Positives = 211/298 (70%), Gaps = 9/298 (3%)
Frame = -2
Query: 1 MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST 60
MENGYTYEDY++TAEWLL HTKH QV +ICGS LG LTDKL QAQIF+ SE+ NF +ST
Sbjct: 6971389 MENGYTYEDYQSTAEWLLFHTKH*TQVTVICGSELGDLTDKLIQAQIFNNSEMLNFFQST 6971210
Query: 61 VPGHAGRLVFGFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGL 120
VPGHA LVFGFLNG CVMMQGRF++Y+GY LW + F VF LLG + LV T+AAGGL
Sbjct: 6971209 VPGHAV*LVFGFLNGTVCVMMQGRFYLYDGYLLWNMIFLHEVFQLLGGNILVATDAAGGL 6971030
Query: 121 NPKFEVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQ 180
NPK EVG IML+ DHI L GF QN +GPNDERFG FPA SDAY+ TM+Q+AL++ Q
Sbjct: 6971029 NPKSEVGRIMLLCDHIKLLGFCDQNSPKGPNDERFGVHFPATSDAYNWTMKQKALNS*NQ 6970850
Query: 181 MGEQRELQEGTYVMVAGPSFETVAECRVLQKLGADAVGMSTVPEVIV--ARHCGLRVFG- 237
MG+Q+E+Q+ TYVM +FET G D+ + A+H
Sbjct: 6970849 MGKQQEVQKDTYVMAVNCNFET----------GRDSSDAEAGDGCCLA*AQHQS*SCMAL 6970700
Query: 238 ------FSLITNKVIMDYESLEKANHEEVLAAGKQAAQKLEQFVSILMASIPLPDKAS 289
FSLITNKVIMDYESL+KANHE V A KQAAQKLEQFVSIL ASIPLPD A+
Sbjct: 6970699 WTWSLCFSLITNKVIMDYESLKKANHE*V*EAVKQAAQKLEQFVSILKASIPLPDNAN 6970526
Done
Internet

Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.ensembl.org/Homo_sapiens/blastview?id=hs_s3a69Ev66Yg3&format=karyo_format Go Links

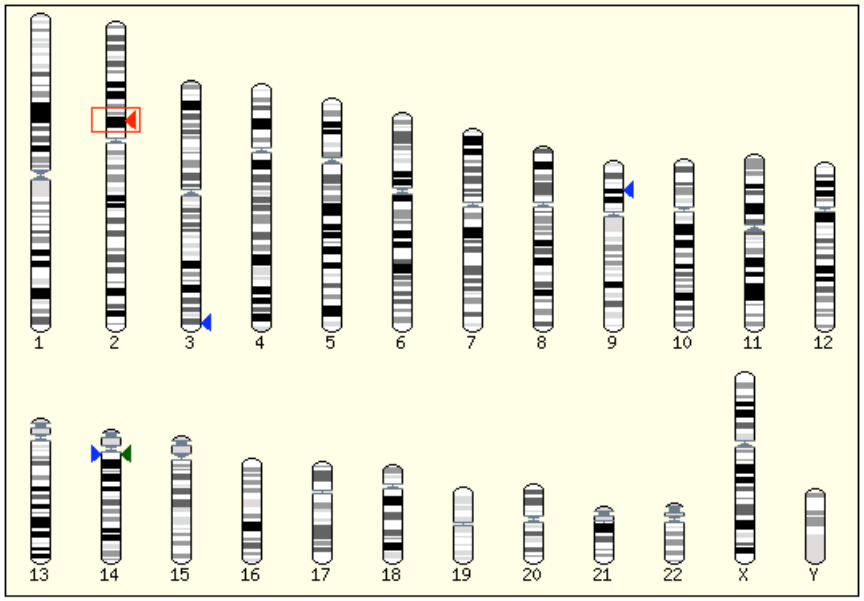
Google Search Web Search Site News PageRank Page Info Up Highlight

Ensembl Human BLASTView

The Wellcome Trust Sanger Institute EBI

Home ▸ Human ▲ What's New ▲ BLAST ▲ SSAHA ▲ EnsMart ▲ Export Data ▲ Download ▲ Disease Browser ▲ Docs ▲

Find All Lookup [e.g. AP000462, RH9632, cancer] Help

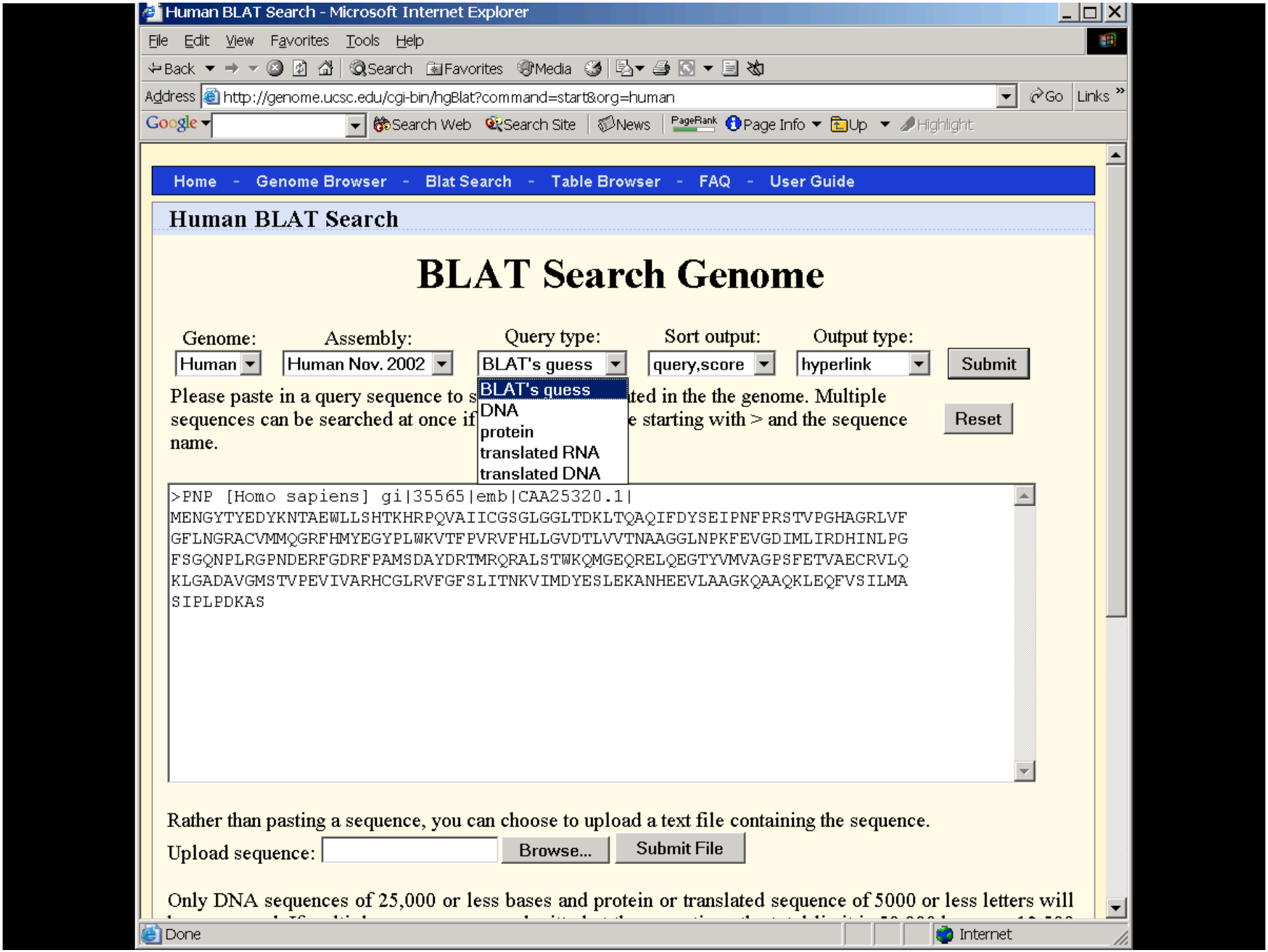


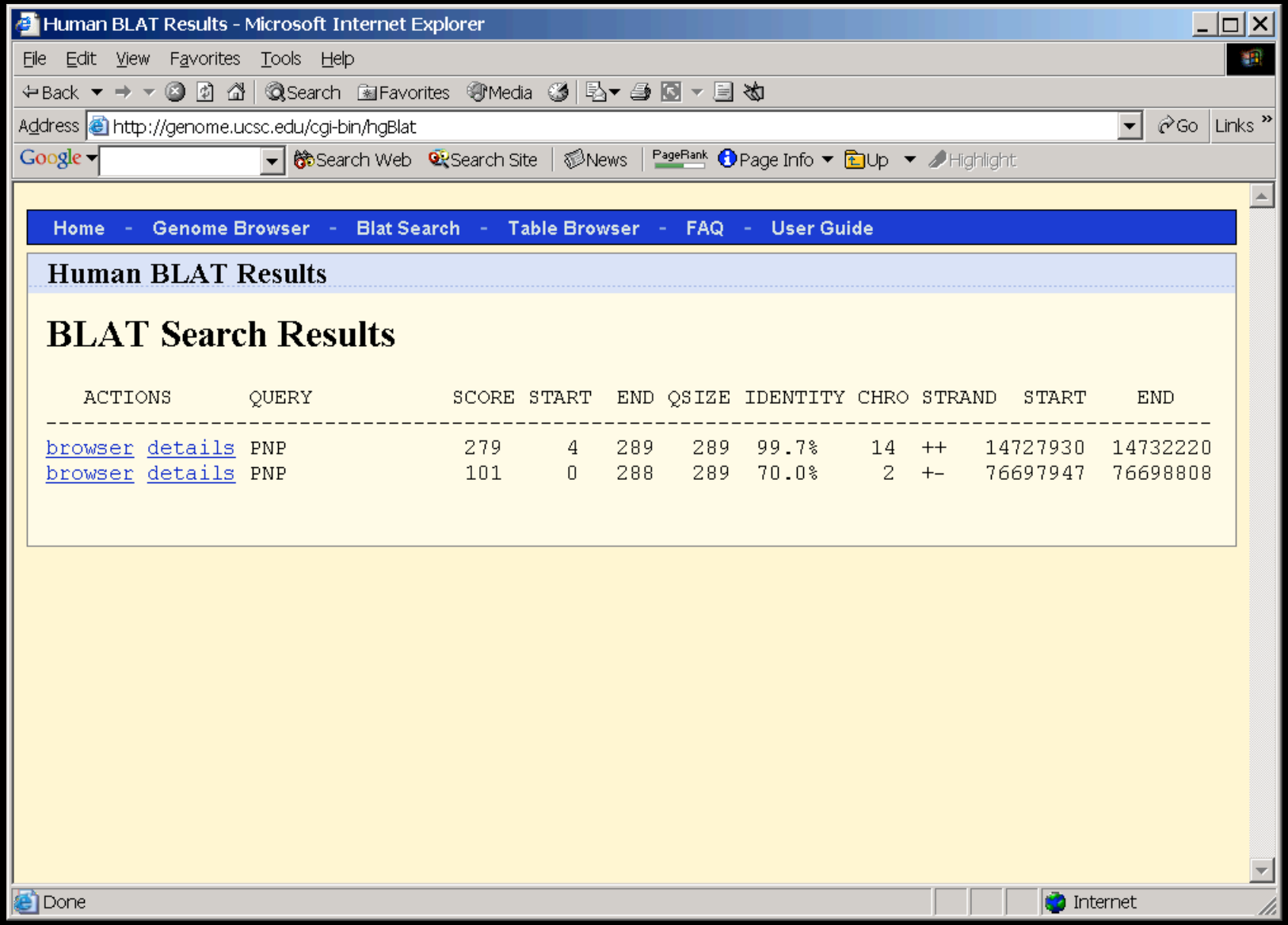
Blast score ranges for this search: [The highest scoring hit(s) are boxed]

42 - 308 | 309 - 575 | 576 - 842

Location of Blast hits

Internet


Human BLAT Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=human
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Home - Genome Browser - Blat Search - Table Browser - FAQ - User Guide
Human BLAT Search
BLAT Search Genome
Genome:
Human
Assembly:
Human Nov. 2002
Query type:
BLAT's guess
Sort output:
query,score
Output type:
hyperlink
Submit
BLAT's guess
DNA
protein
translated RNA
translated DNA
Please paste in a query sequence to s ted in the the genome. Multiple sequences can be searched at once if e starting with > and the sequence name.
Reset
>PNP [Homo sapiens] gi|35565|emb|CAA25320.1|
MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRSTVPGHAGRLVF
GFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGLNPKFEVGDIMLIRDHINLPG
FSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQMGEQRELQEGTYVMVAGPSFETVAECRVLQ
KLGADAVGMSTVPEVIVARHCGLRVFGFSLITNKVIMDYESLEKANHEEVLAAGKQAAQKLEQFVSILMA
SIPLPDKAS
Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence:
Browse...
Submit File
Only DNA sequences of 25,000 or less bases and protein or translated sequence of 5000 or less letters will
Done
Internet


Human BLAT Results - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://genome.ucsc.edu/cgi-bin/hgBlat
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Home - Genome Browser - Blat Search - Table Browser - FAQ - User Guide
Human BLAT Results
BLAT Search Results
ACTIONS QUERY SCORE START END QSIZE IDENTITY CHRO STRAND START END
browser details PNP 279 4 289 289 99.7% 14 ++ 14727930 14732220
browser details PNP 101 0 288 289 70.0% 2 +- 76697947 76698808
Done
Internet

Entrez Map View - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?org=hum&MAPS=loc%2Ccntg-r&db=genome%2C Go Links

MapViewer Home

Map Viewer Help
Human Maps Help
FTP

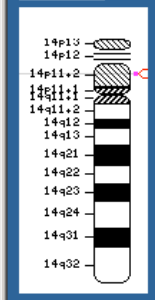
Data As Table View

Maps&Options

Compress Map

Region Shown:
14727503
14732649 Go

out zoom in



ideogram
master

Homo sapiens Map View build 31 BLAST the Human Genome

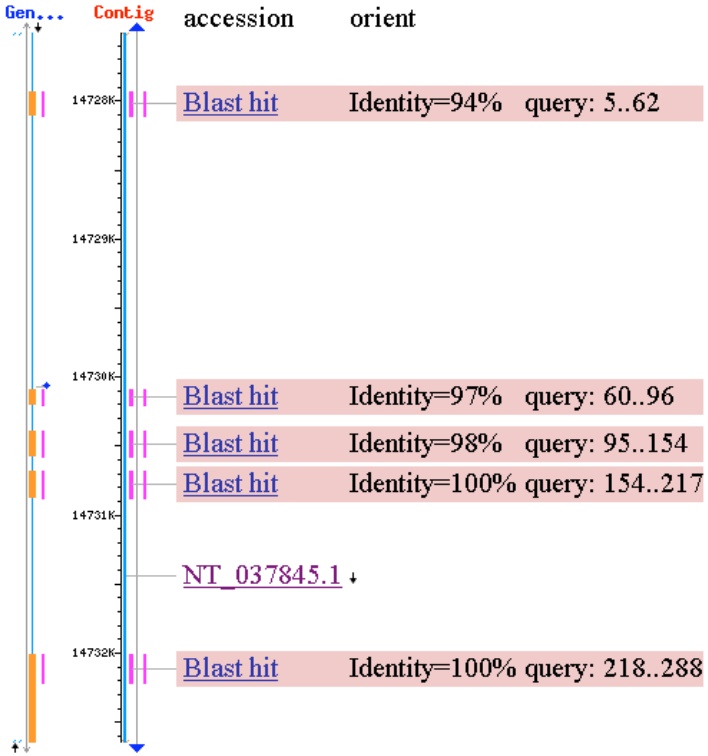
Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 [**14**] 15 16 17 18 19 20 21 22 X Y

Query: BLAST: PNP [Homo sapiens]
gi|35565|emb|CAA25320.1| [clear]

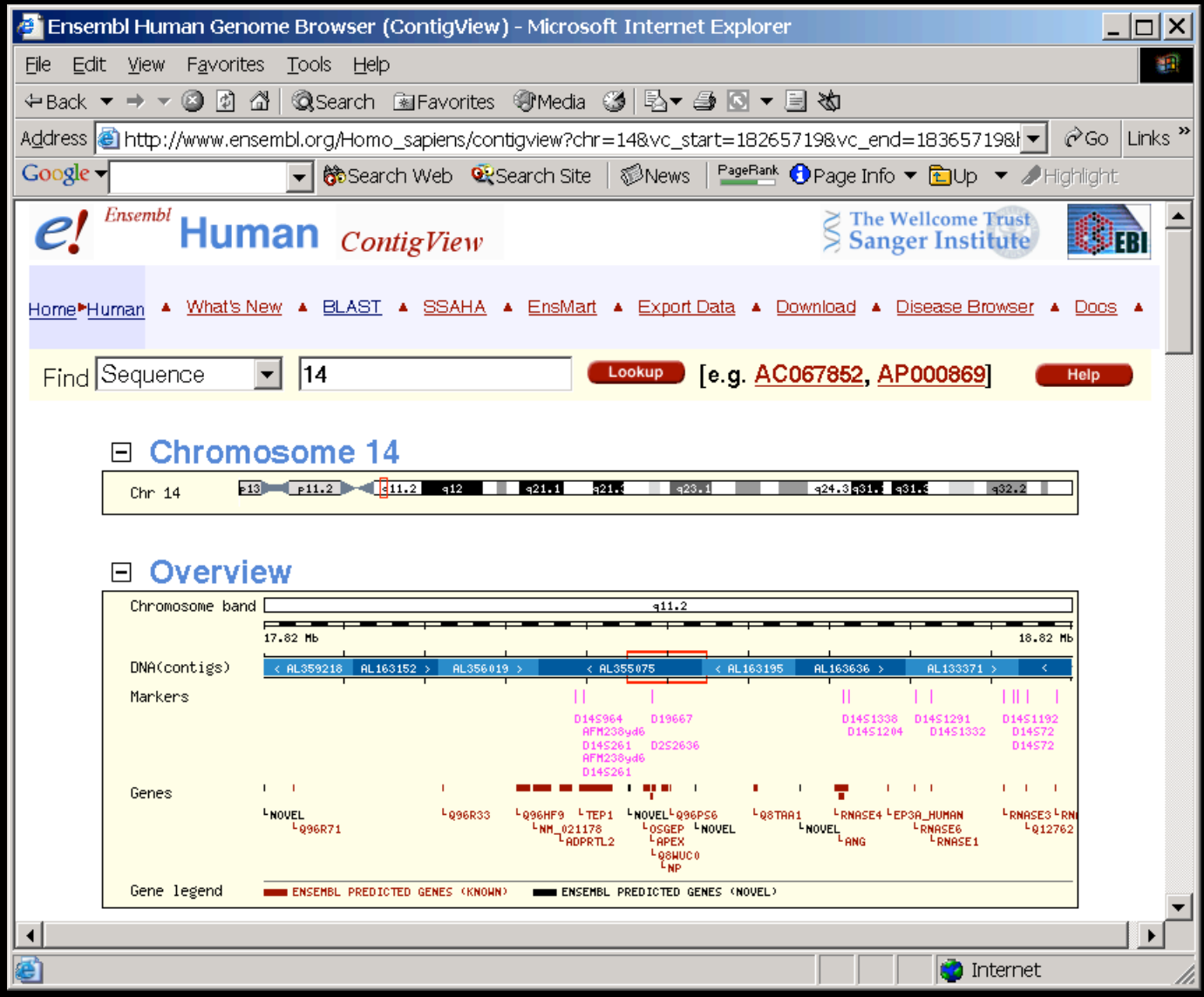
Color Key for Alignment Scores: <40 | 40-50 | 50-80 | 80-200 | >=200

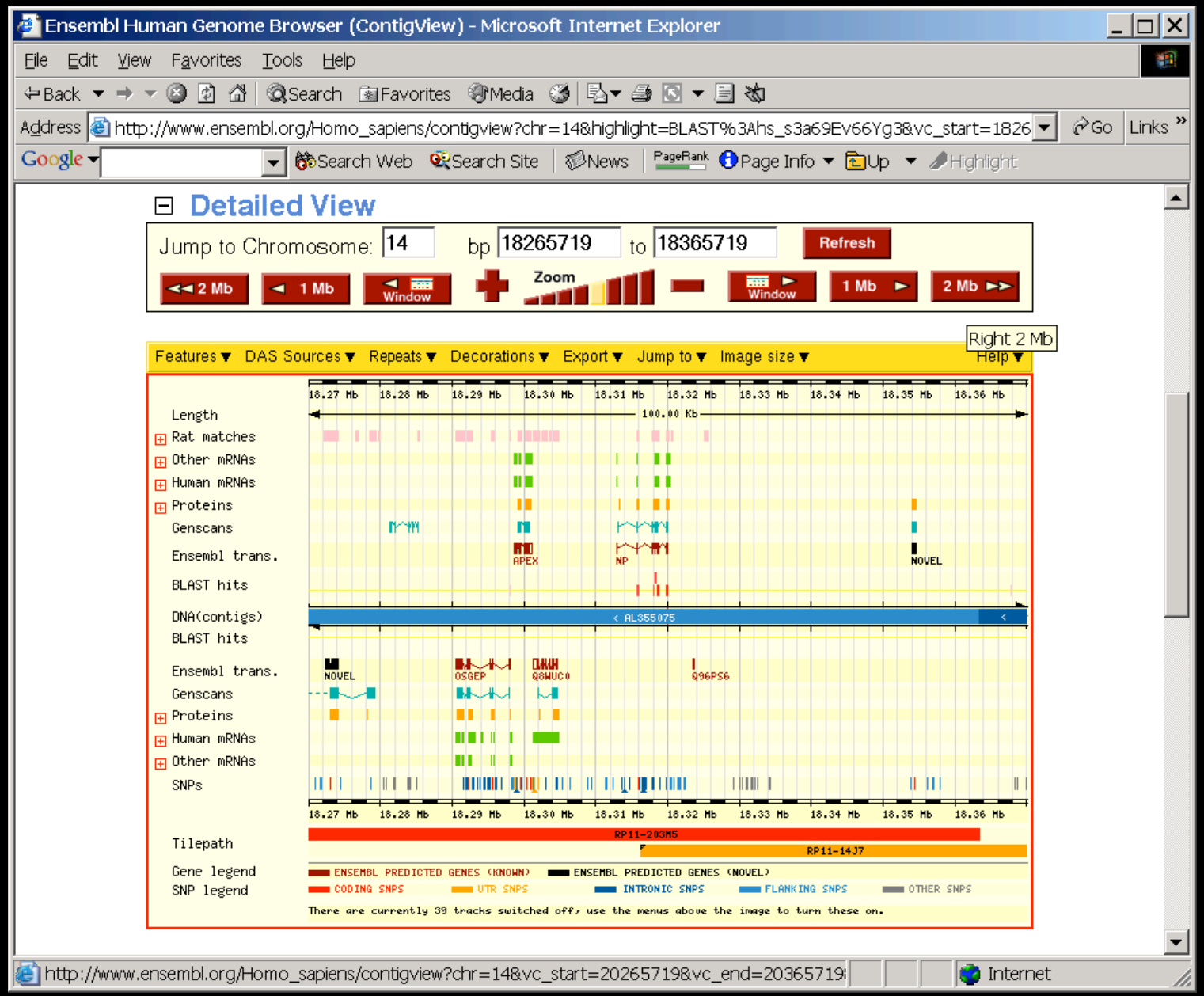
Master Map: Contig — Maps & Options

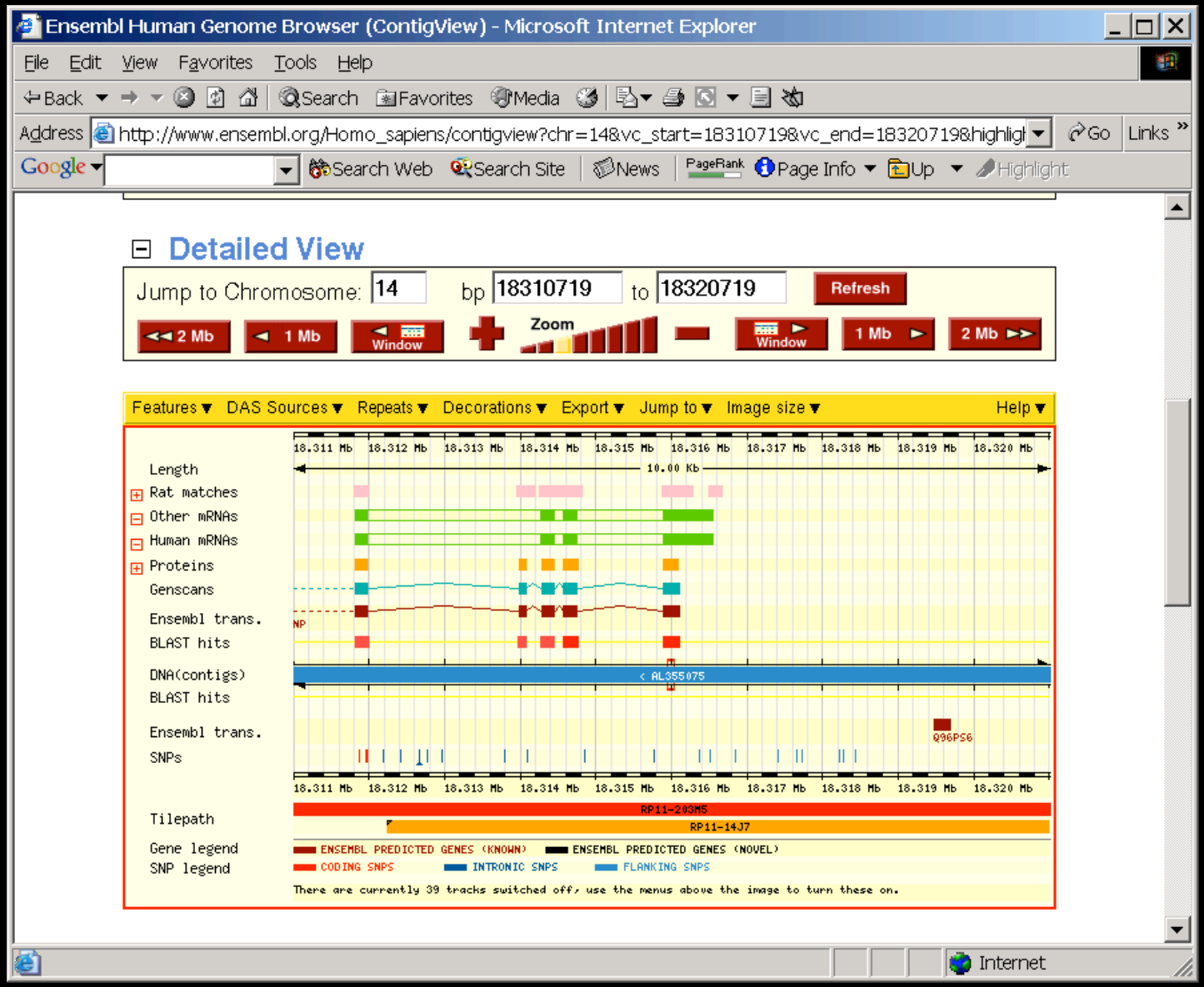
Total Contigs On Chromosome: 7 [1 not localized]
Region Displayed: **14,727K-14,732K**
bp Download/View Sequence/Evidence
Contigs Labeled: **6** Total Contigs in Region: **6**



Error on page. Internet


Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=18265719&vc_end=18365719&
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Ensembl Human ContigView
The Wellcome Trust Sanger Institute
EBI
Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs
Find Sequence 14 Lookup [e.g. AC067852, AP000869] Help
Chromosome 14
Chr 14
Overview
Chromosome band
q11.2
17.82 Mb
18.82 Mb
DNA(contigs)
AL359218 AL163152 AL356019 AL355075 AL163195 AL163636 AL133371
Markers
D14S964 D19667 AFM238yd6 D14S261 D2S2636 AFM238yd6 D14S261
D14S1338 D14S1204 D14S1291 D14S1332 D14S1192 D14S72 D14S72
Genes
NOVEL Q96R71 Q96R33 Q96HF9 NM_021178 ADPRTL2 TEP1 NOVEL OSGEP APEX Q8WUC0 NP Q96PS6 NOVEL Q8TAA1 NOVEL RNASE4 ANG EP3A_HUMAN RNASE6 RNASE1 RNASE3 Q12762
Gene legend
ENSEMBL PREDICTED GENES (KNOWN)
ENSEMBL PREDICTED GENES (NOVEL)
Internet


Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&highlight=BLAST%3Ahs_s3a69Ev66Yg3&vc_start=1826
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Detailed View
Jump to Chromosome: 14 bp 18265719 to 18365719 Refresh
2 Mb 1 Mb Window Zoom Window 1 Mb 2 Mb
Right 2 Mb
Features DAS Sources Repeats Decorations Export Jump to Image size Help
18.27 Mb 18.28 Mb 18.29 Mb 18.30 Mb 18.31 Mb 18.32 Mb 18.33 Mb 18.34 Mb 18.35 Mb 18.36 Mb
Length
100.00 Kb
Rat matches
Other mRNAs
Human mRNAs
Proteins
Genscans
Ensembl trans.
APEX
NP
NOVEL
BLAST hits
DNA(contigs)
AL355075
BLAST hits
Ensembl trans.
NOVEL
OSGEP
Q8WUC0
Q96PS6
Genscans
Proteins
Human mRNAs
Other mRNAs
SNPs
18.27 Mb 18.28 Mb 18.29 Mb 18.30 Mb 18.31 Mb 18.32 Mb 18.33 Mb 18.34 Mb 18.35 Mb 18.36 Mb
Tilepath
RP11-203M5
RP11-14J7
Gene legend
ENSEMBL PREDICTED GENES (KNOWN)
ENSEMBL PREDICTED GENES (NOVEL)
SNP legend
CODING SNPS
UTR SNPS
INTRONIC SNPS
FLANKING SNPS
OTHER SNPS
There are currently 39 tracks switched off; use the menus above the image to turn these on.
http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=20265719&vc_end=20365719
Internet


Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/contigview?chr=14&vc_start=18310719&vc_end=18320719&highligh
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Detailed View
Jump to Chromosome: 14 bp 18310719 to 18320719 Refresh
2 Mb 1 Mb Window Zoom Window 1 Mb 2 Mb
Features DAS Sources Repeats Decorations Export Jump to Image size Help
18.311 Mb 18.312 Mb 18.313 Mb 18.314 Mb 18.315 Mb 18.316 Mb 18.317 Mb 18.318 Mb 18.319 Mb 18.320 Mb
Length
10.00 Kb
Rat matches
Other mRNAs
Human mRNAs
Proteins
Genscans
Ensembl trans.
BLAST hits
DNA(contigs)
AL355075
BLAST hits
Ensembl trans.
SNPs
Q96PS6
Tilepath
RP11-203M5
RP11-14J7
Gene legend
ENSEMBL PREDICTED GENES (KNOWN)
ENSEMBL PREDICTED GENES (NOVEL)
SNP legend
CODING SNPS
INTRONIC SNPS
FLANKING SNPS
There are currently 39 tracks switched off, use the menus above the image to turn these on.
Internet

Ensembl Human Genome Browser (SyntenyView) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://www.ensembl.org/Homo_sapiens/syntenyview?species=Mus_musculus&chr=14&loc=18290719 Go Links

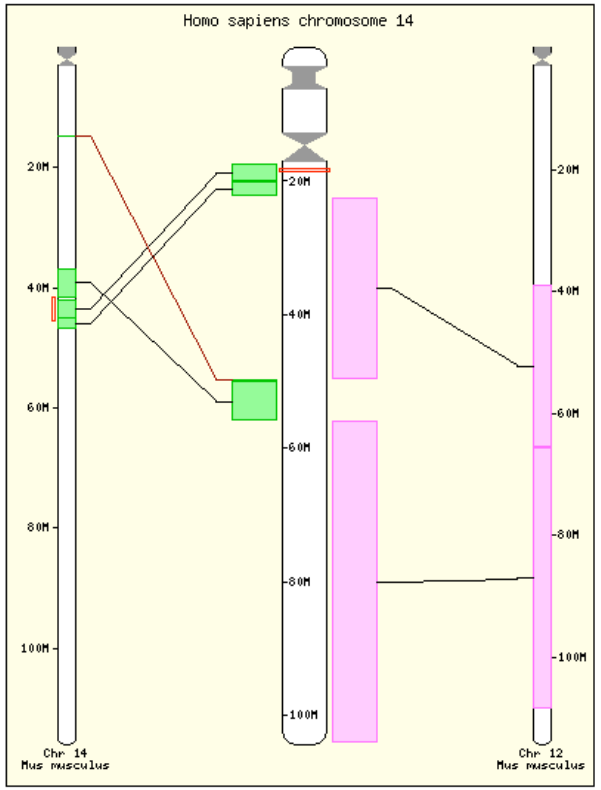
Google Search Web Search Site News PageRank Page Info Up Highlight

Ensembl Human SyntenyView

The Wellcome Trust Sanger Institute EBI

Home ▸ Human ▲ What's New ▲ BLAST ▲ SSAHA ▲ EnsMart ▲ Export Data ▲ Download ▲ Disease Browser ▲ Docs ▲

Find All Lookup [e.g. AP000462, RH9632, cancer] Help



Human Chromosome 14

Jump to chromosome 14 Lookup

Jump to mapview for chromosome statistics.

Homology Matches

Homo_sapiens Genes		*Mus_musculus* Homologues
OSGEP (18.29 Mb)	->	Q99LN8 (chr 14 : 42.78 Mb)
APEX (18.29 Mb)	->	Apex1 (chr 14 : 42.79 Mb)
Q8WUC0 (18.30 Mb)	->	ENSMUSG00000035953 (chr 14 : 42.79 Mb)
NP (18.31 Mb)	->	Pnp (chr 14 : 42.81 Mb)
Q96PS6 (18.32 Mb)		
ENSG00000165787 (18.35 Mb)	->	4930474F22Rik (chr 14 : 42.87 Mb)
Q8TAA1 (18.42 Mb)		
ENSG00000169431 (18.48 Mb)	->	ENSMUSG00000035932 (chr 14 : 42.93 Mb)
RNASE4 (18.52 Mb)	->	Rnase4 (chr 14 : 42.96 Mb)
ANG (18.53 Mb)	->	Angrp (chr 14 : 43.06 Mb) Ang (chr 14 : 42.97 Mb)
EP3A_HUMAN (18.59 Mb)	->	ENSMUSG00000021878 (chr 14 : 42.98 Mb)

Internet

UCSC Genome Browser v17 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr14:14727931-14732220&db=hg13&ss=../trash/hgSs_genome.

Go Links

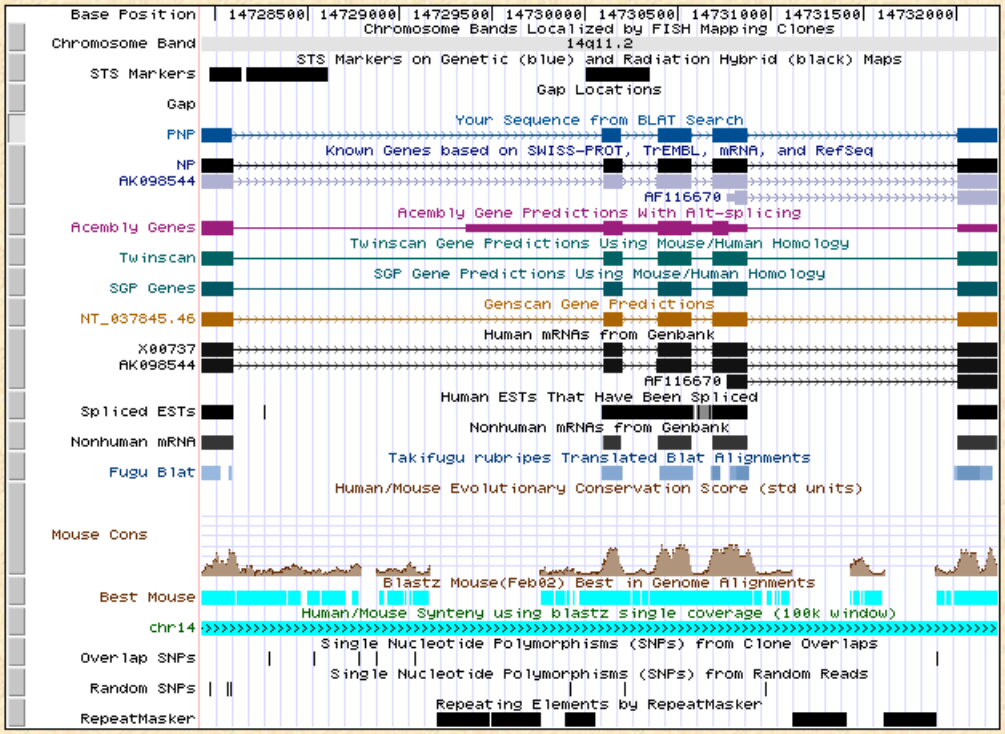
Google Search Web Search Site News PageRank Page Info Up Highlight

Home BLAT DNA Tables Convert Map View PDF/PS Guide

UCSC Genome Browser on Human Nov. 2002 Freeze

move <<< << < > >> >>> zoom in 1.5x 3x 10x zoom out 1.5x 3x 10x

position chr14:14727931-14732220 size 4,290 image width 620 jump

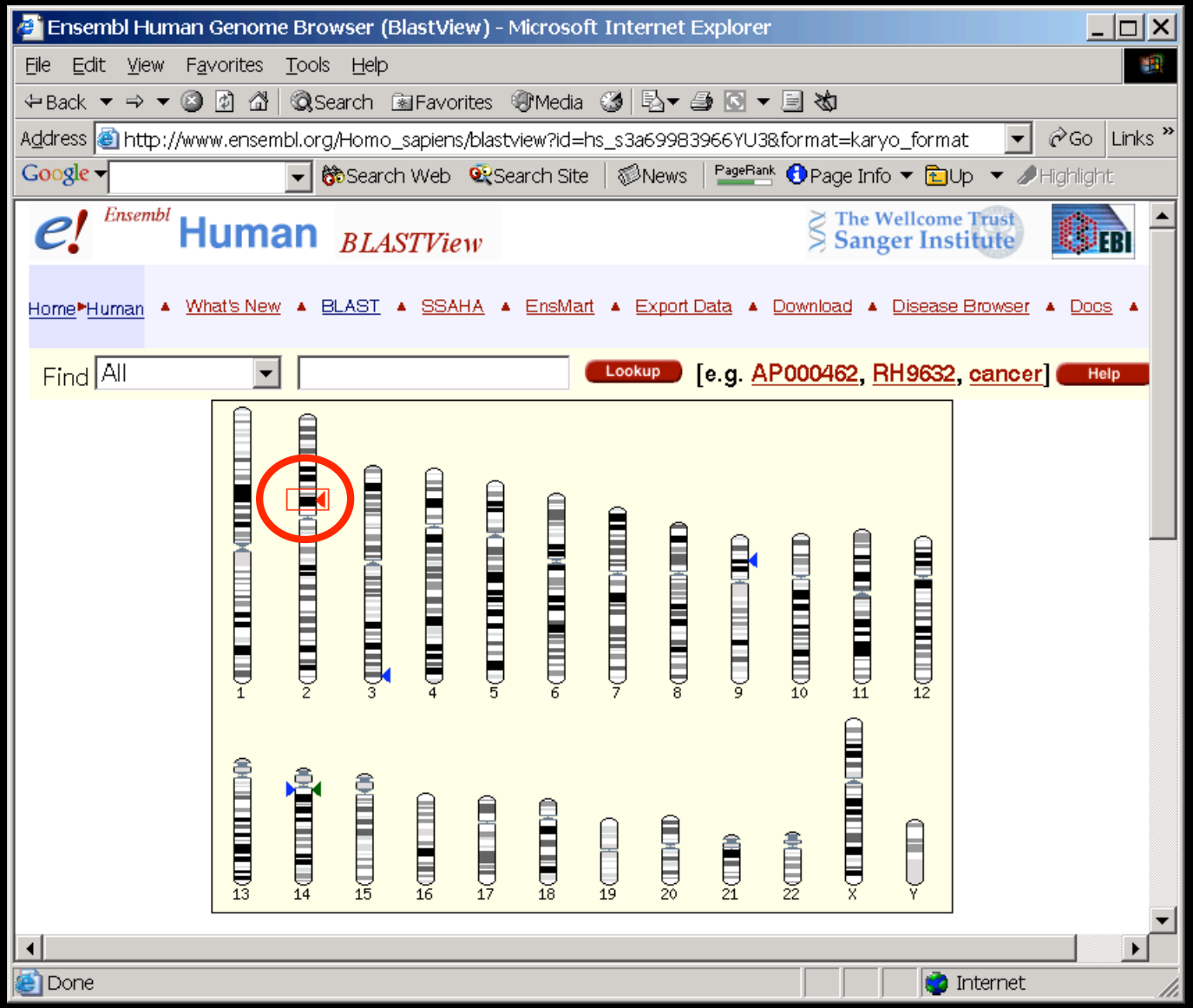


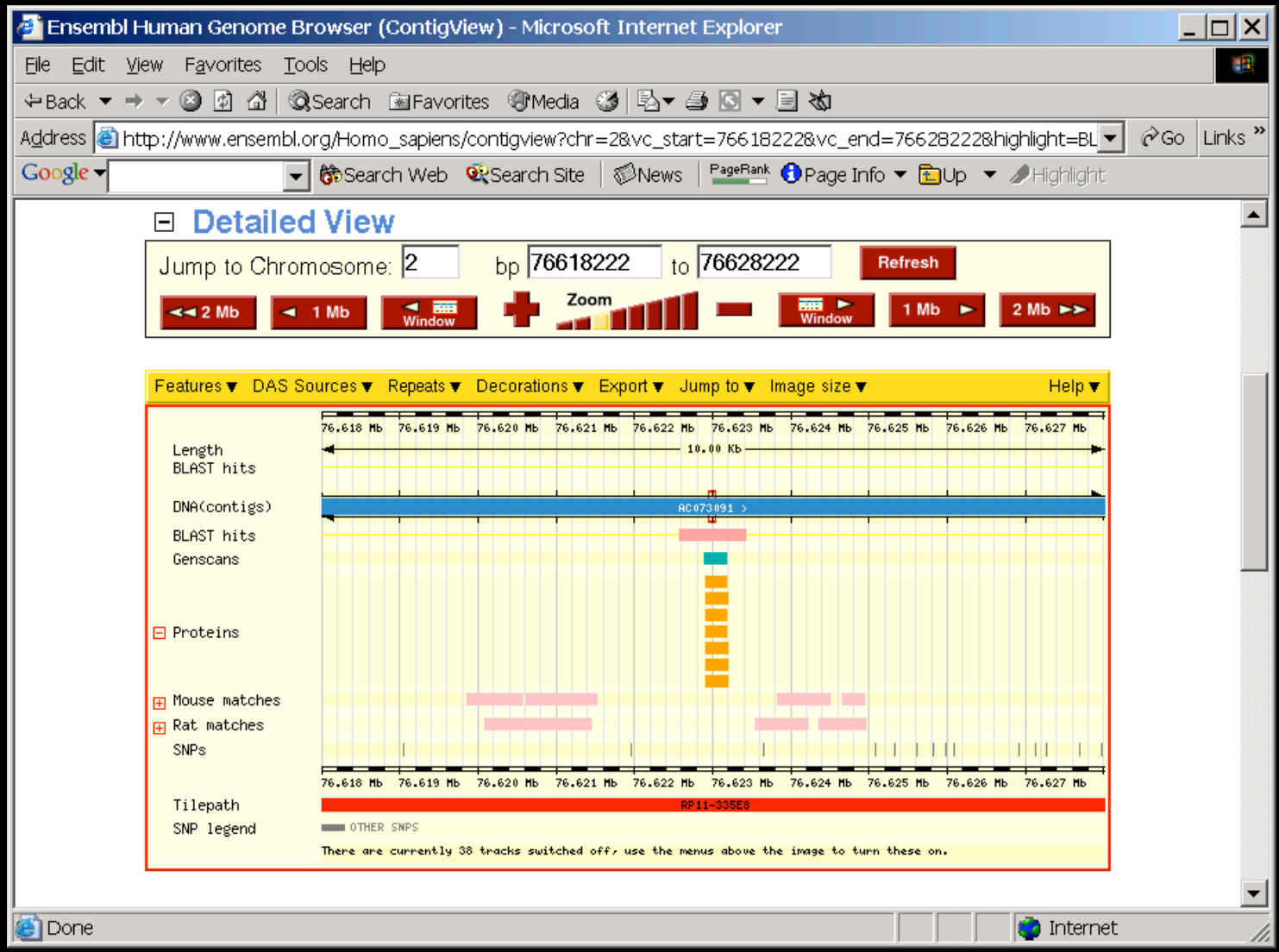
move start < 2.0 >

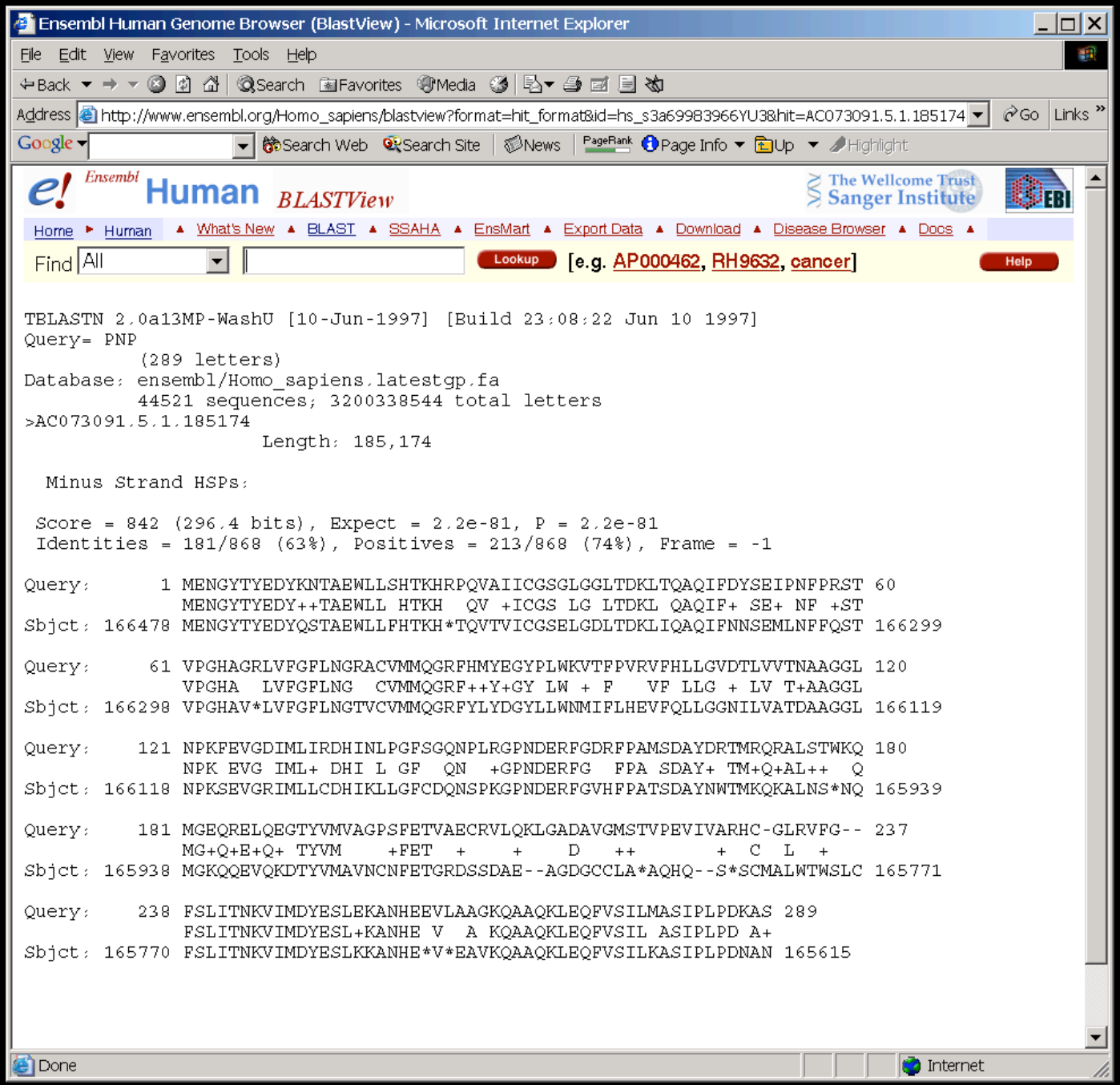
Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options

move end < 2.0 >

Internet

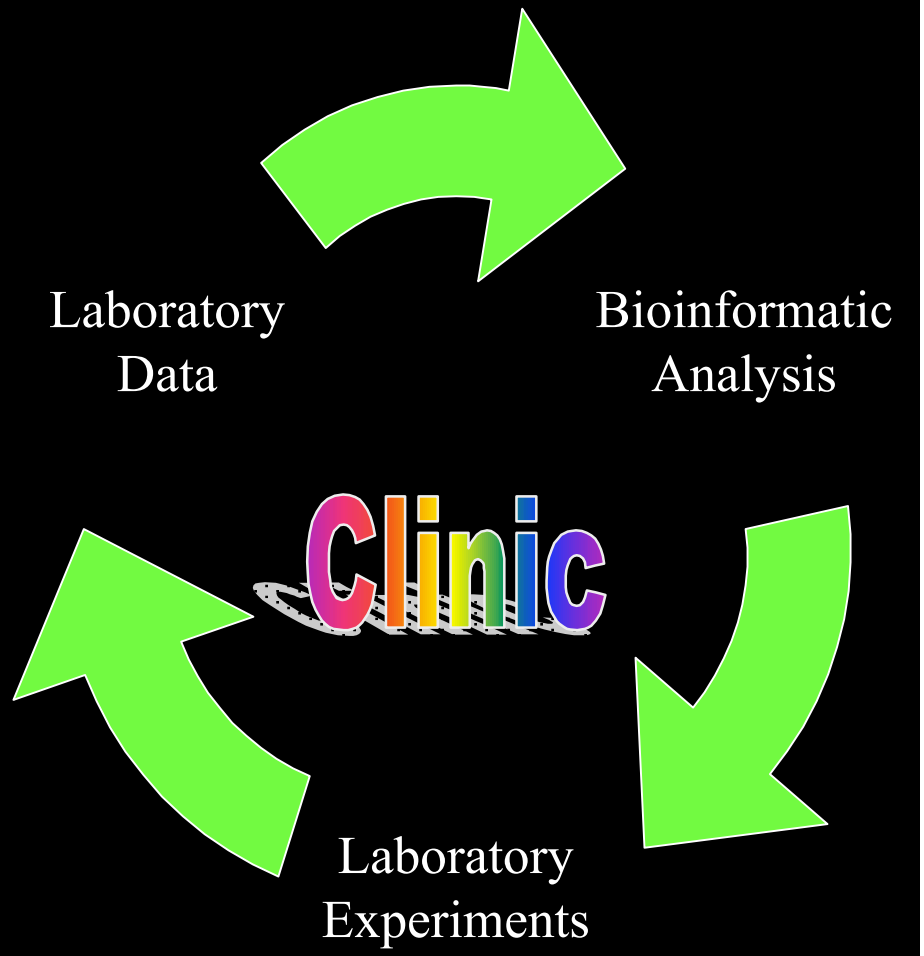

Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/blastview?id=hs_s3a69983966YU3&format=karyo_format
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Ensembl Human BLASTView
The Wellcome Trust Sanger Institute
EBI
Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs
Find All Lookup [e.g. AP000462, RH9632, cancer] Help
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 X Y
Done
Internet


Ensembl Human Genome Browser (ContigView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/contigview?chr=2&vc_start=76618222&vc_end=76628222&highlight=BL
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Detailed View
Jump to Chromosome: 2 bp 76618222 to 76628222 Refresh
<< 2 Mb < 1 Mb Window Zoom Window > 1 Mb 2 Mb >>
Features DAS Sources Repeats Decorations Export Jump to Image size Help
76.618 Mb 76.619 Mb 76.620 Mb 76.621 Mb 76.622 Mb 76.623 Mb 76.624 Mb 76.625 Mb 76.626 Mb 76.627 Mb
Length
10.00 Kb
BLAST hits
DNA(contigs)
AC073091 >
BLAST hits
Genscans
Proteins
Mouse matches
Rat matches
SNPs
Tilepath
RP11-335E8
SNP legend
OTHER SNPS
There are currently 38 tracks switched off; use the menus above the image to turn these on.
Done
Internet


Ensembl Human Genome Browser (BlastView) - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.ensembl.org/Homo_sapiens/blastview?format=hit_format&id=hs_s3a69983966YU3&hit=AC073091.5.1.185174
Go Links
Google Search Web Search Site News PageRank Page Info Up Highlight
Ensembl Human BLASTView
The Wellcome Trust Sanger Institute
EBI
Home Human What's New BLAST SSAHA EnsMart Export Data Download Disease Browser Docs
Find All Lookup [e.g. AP000462, RH9632, cancer] Help
TBLASTN 2.0a13MP-WashU [10-Jun-1997] [Build 23:08:22 Jun 10 1997]
Query= PNP
(289 letters)
Database: ensembl/Homo_sapiens.latestgp.fa
44521 sequences; 3200338544 total letters
>AC073091.5.1.185174
Length: 185,174
Minus Strand HSPs:
Score = 842 (296.4 bits), Expect = 2.2e-81, P = 2.2e-81
Identities = 181/868 (63%), Positives = 213/868 (74%), Frame = -1
Query: 1 MENGYTYEDYKNTAEWLLSHTKHRPQVAIICGSGLGGLTDKLTQAQIFDYSEIPNFPRST 60
MENGYTYEDY++TAEWLL HTKH QV +ICGS LG LTDKL QAQIF+ SE+ NF +ST
Sbjct: 166478 MENGYTYEDYQSTAEWLLFHTKH*TQVTVICGSELGDLTDKLIQAQIFNNSEMLNFFQST 166299
Query: 61 VPGHAGRLVFGFLNGRACVMMQGRFHMYEGYPLWKVTFPVRVFHLLGVDTLVVTNAAGGL 120
VPGHA LVFGFLNG CVMMQGRF++Y+GY LW + F VF LLG + LV T+AAGGL
Sbjct: 166298 VPGHAV*LVFGFLNGTVCVMMQGRFYLYDGYLLWNMIFLHEVFQLLGGNILVATDAAGGL 166119
Query: 121 NPKFEVGDIMLIRDHINLPGFSGQNPLRGPNDERFGDRFPAMSDAYDRTMRQRALSTWKQ 180
NPK EVG IML+ DHI L GF QN +GPNDERFG FPA SDAY+ TM+Q+AL++ Q
Sbjct: 166118 NPKSEVGRIMLLCDHIKLLGFCDQNSPKGPNDERFGVHFPATSDAYNWTMKQKALNS*NQ 165939
Query: 181 MGEQRELQEGTYVMVAGPSFETVAECRVLQKLGADAVGMSTVPEVIVARHC-GLRVFG-- 237
MG+Q+E+Q+ TYVM +FET + + D ++ + C L +
Sbjct: 165938 MGKQQEVQKDTYVMAVNCNFETGRDSSDAE--AGDGCCLA*AQHQ--S*SCMALWTWSLC 165771
Query: 238 FSLITNKVIMDYESLEKANHEEVLAAGKQAAQKLEQFVSILMASIPLPDKAS 289
FSLITNKVIMDYESL+KANHE V A KQAAQKLEQFVSIL ASIPLPD A+
Sbjct: 165770 FSLITNKVIMDYESLKKANHE*V*EAVKQAAQKLEQFVSILKASIPLPDNAN 165615
Done
Internet

Biological Information Flow

- Laboratory
  - Data generation
    - Sequence, expression, proteomic…
- Bioinformatics
  - Data management
  - Data analysis
  - Biological inferences
- Laboratory
  - Hypothesis testing
- Clinical applications
  - Diagnostics
  - Prophylaxis
  - Targeted therapeutics



One Final Word of Wisdom...

- "...although the computer is a wonderful helpmate for the sequence searcher and comparer, biochemists and molecular biologists must guard against the blind acceptance of any algorithmic output; given the choice, think like a biologist and not a statistician."
  - - Russell F. Doolittle, 1990


Farewell!