

A Field Guide to GenBank and NCBI Molecular Biology Resources

slightly modified from

Peter Cooper

<ftp://ftp.ncbi.nih.gov/pub/cooper/FieldGuide/>

Eric Sayers

ftp://ftp.ncbi.nih.gov/pub/sayers/Field_Guide/U_Penn/

NCBI Resources

- About NCBI
- NCBI Sequence Databases
 - Primary Database - GenBank
 - Derivative Databases - RefSeq
- Entrez Databases and Text Searching
- BLAST Services
- Genomic Resources

The National Center for Biotechnology Information (NCBI)



Lister Hill Center



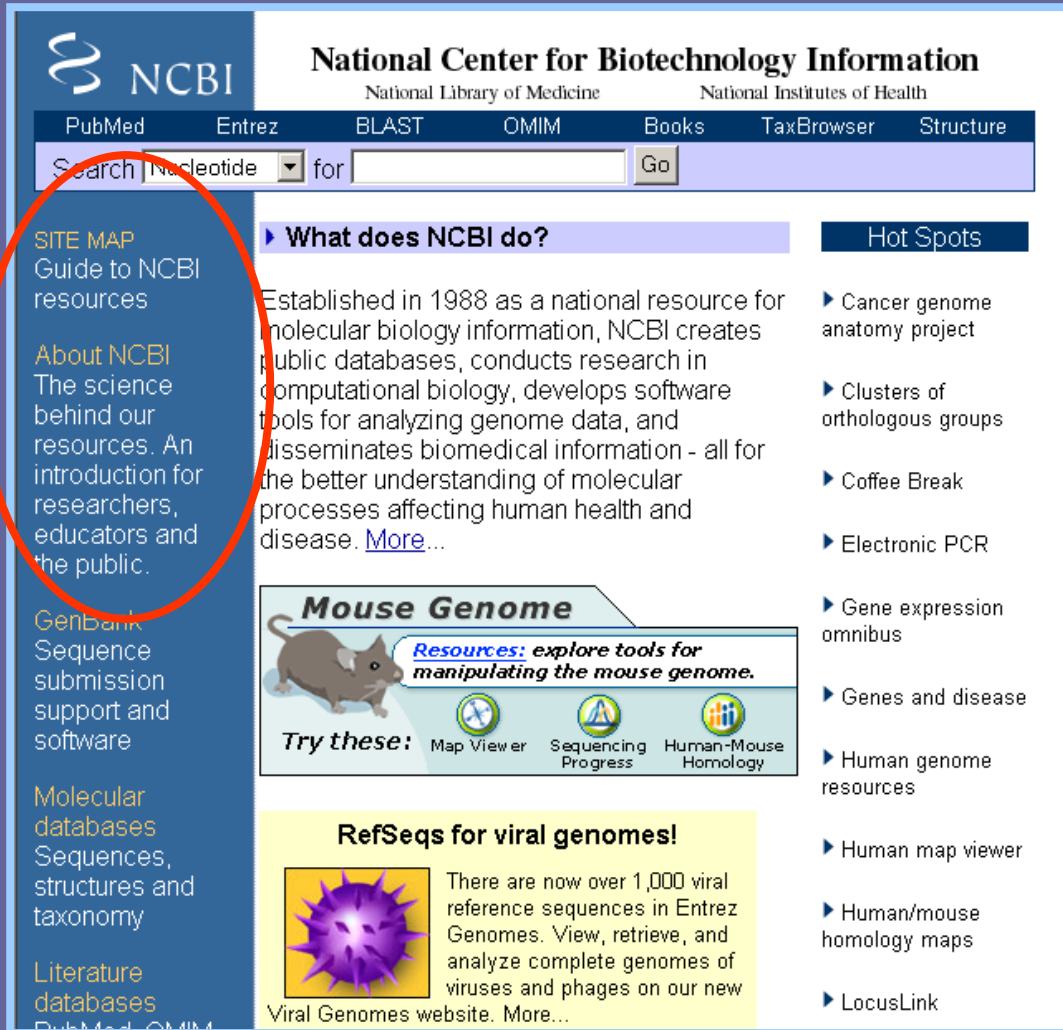
William H. Natcher Building

The National Center for Biotechnology Information (NCBI)

- **Created as a part of NLM in 1988**
 - Establish public databases
 - Perform research in computational biology
 - Develop software tools for sequence analysis
 - Disseminate biomedical information
- **Tools: BLAST(1990), Entrez (1992)**
- **GenBank (1992)**
- **Free MEDLINE (PubMed, 1997)**
- **Human genome (2001)**

NCBI Home Page

<http://www.ncbi.nlm.nih.gov>



NCBI
National Center for Biotechnology Information
National Library of Medicine National Institutes of Health

PubMed Entrez BLAST OMIM Books TaxBrowser Structure

Search Nucleotide for Go

SITE MAP
Guide to NCBI resources

About NCBI
The science behind our resources. An introduction for researchers, educators and the public.

Genbank
Sequence submission support and software

Molecular databases
Sequences, structures and taxonomy

Literature databases
PubMed, OMIM


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...](#)

Hot Spots

- ▶ Cancer genome anatomy project
- ▶ Clusters of orthologous groups
- ▶ Coffee Break
- ▶ Electronic PCR
- ▶ Gene expression omnibus
- ▶ Genes and disease
- ▶ Human genome resources
- ▶ Human map viewer
- ▶ Human/mouse homology maps
- ▶ LocusLink

Mouse Genome
Resources: explore tools for manipulating the mouse genome.

Try these: [Map Viewer](#) [Sequencing Progress](#) [Human-Mouse Homology](#)

RefSeqs for viral genomes!
 There are now over 1,000 viral reference sequences in Entrez Genomes. View, retrieve, and analyze complete genomes of viruses and phages on our new Viral Genomes website. [More...](#)

To learn more, visit the "Site Map" and "About NCBI" web pages

Site Map



Site Map

[PubMed](#)[Entrez](#)[BLAST](#)[OMIM](#)[Taxonomy](#)[Structure](#)

This site map is also a **guide to NCBI resources**. Each link leads to a **brief description of the resource** on this page, then to the resource itself. A **Quick Links** table is also available. It provides only an alphabetical list of the major resources with **direct links** to those resources, bypassing the descriptions.

RESOURCE CATEGORIES

About NCBI

[programs and services](#), [what's new](#), [NCBI News](#), [postdoctoral fellowships](#), [organizational structure](#), [contact information](#), [e-mail lists](#)

GenBank

[overview](#), [submit sequences](#), [submit genomes](#), [sample record](#), [GenBank divisions](#), [statistics](#), [release notes](#), [international collaboration](#), [FTP GenBank](#)

Molecular Databases

[nucleotides](#), [proteins](#), [structures](#), [taxonomy](#)

Literature Databases

[PubMed](#), [PubMedCentral](#), [OMIM](#), [Books](#), [Citation Matcher](#)

Genomes and Maps

[Entrez Genomes](#), [Map Viewer](#), [human](#), [mouse](#), [rat](#), [cow](#), [zebrafish](#), [Drosophila](#), [nematode](#), [plant genomes](#), [yeast](#), [malaria](#), [microbial genomes](#), [viruses](#), [viroids](#), [plasmids](#), [eukaryotic organelles](#)

ALPHABETICAL INDEX

WITH LINKS TO RESOURCE DESCRIPTIONS

(To bypass descriptions, use the [Quick Links](#) table.)

BankIt	GenBank sample record	Plant Genomes <small>NEW</small>
BLAST	GeneMap'99	Proteins Sequences
Books <small>NEW</small>	Genes and Disease	PROW
CCAP	Genomes and Maps	PubMed
CDD	GEO	PubMed Central
CGAP	HTGs	RefSeq
Clones	HomoloGene	Research at NCBI
Cn3D	Human Genome Resources	Retroviruses
Coffee Break	Human Genome Sequencing	SAGEmap
COGs	Human-Mouse Homology Maps	Seminars

About NCBI

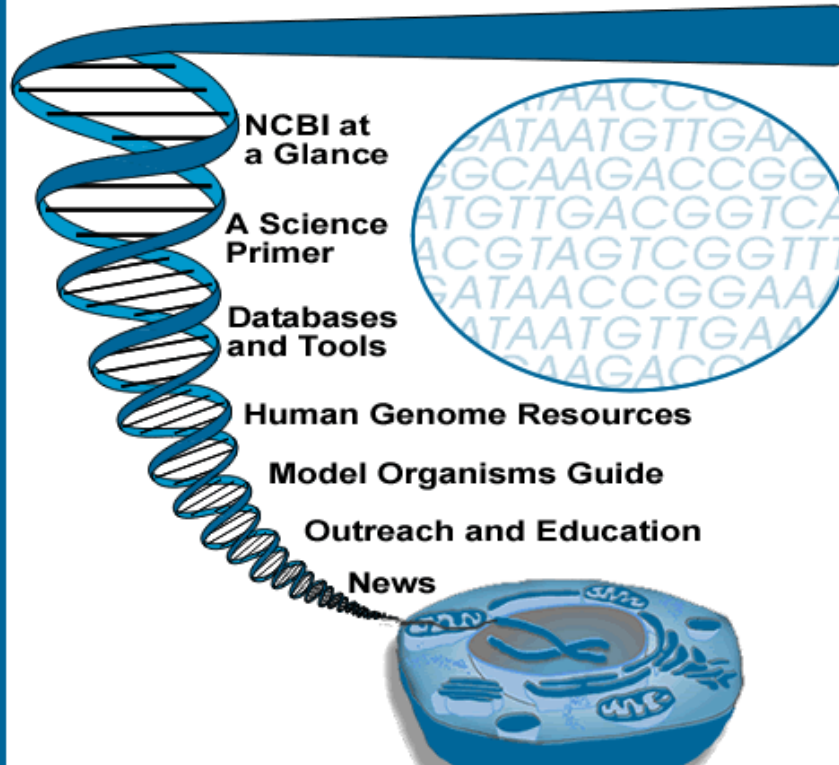


About NCBI

National Center for Biotechnology Information

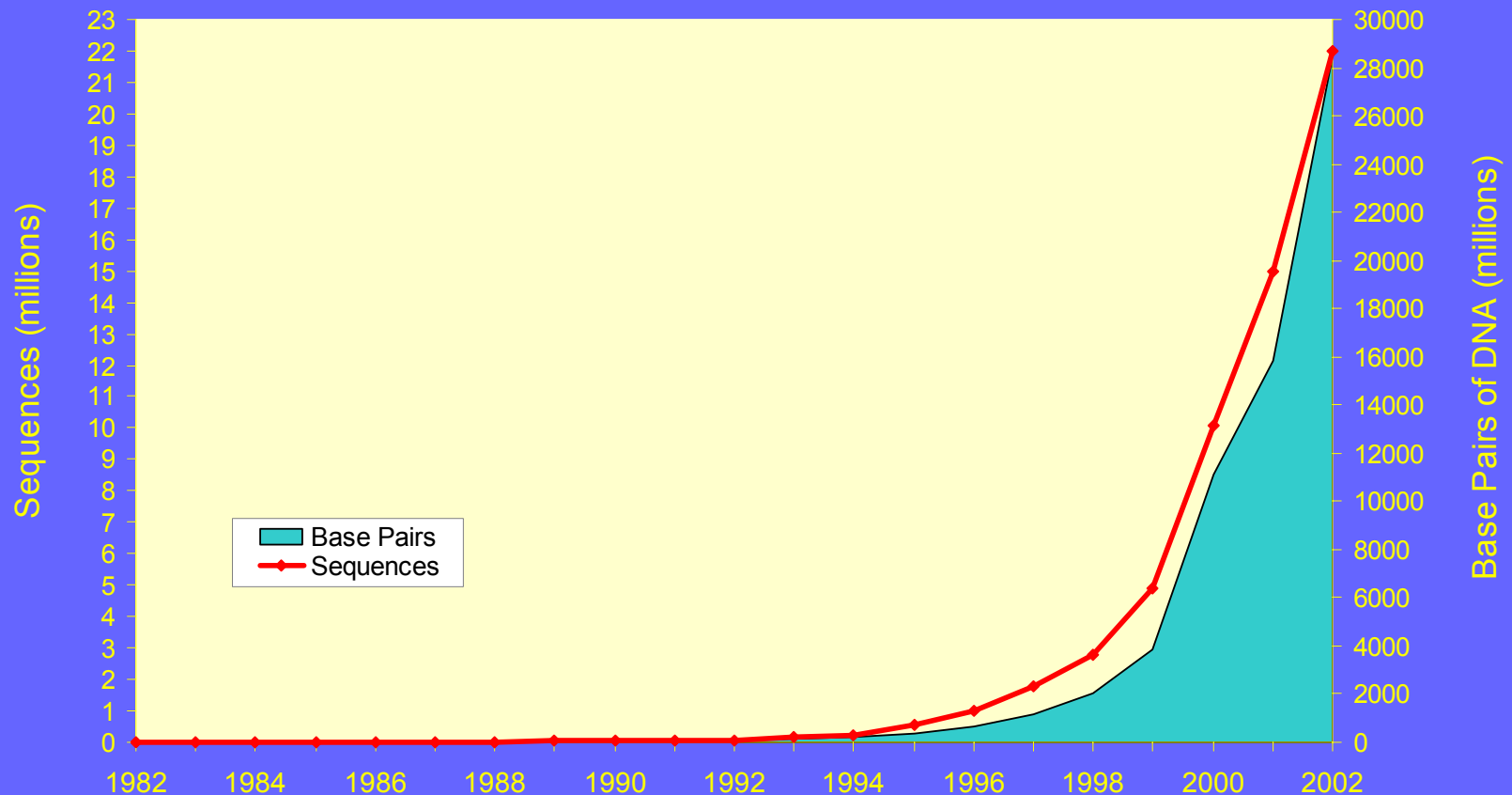
About NCBI
Site Map

About NCBI	NCBI at a Glance	A Science Primer	Databases and Tools
Human Genome Resources	Model Organisms Guide	Outreach and Education	News

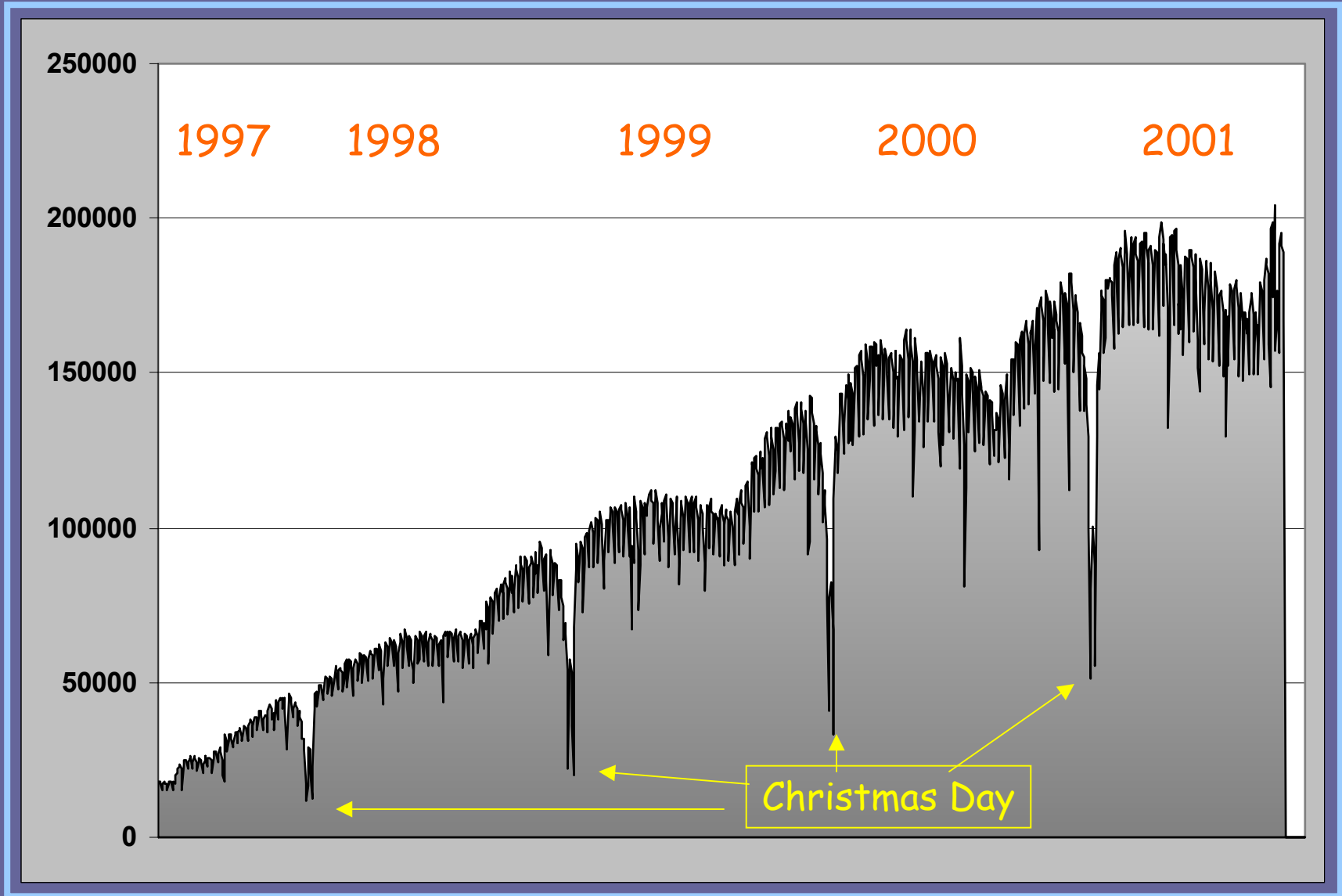


Some NCBI Statistics...

Growth of GenBank



Users per day



Molecular Databases

- Primary Databases

- Original submissions by experimentalists
- Database staff organize but don't add additional information

- **Example: GenBank**

- Derivative Databases

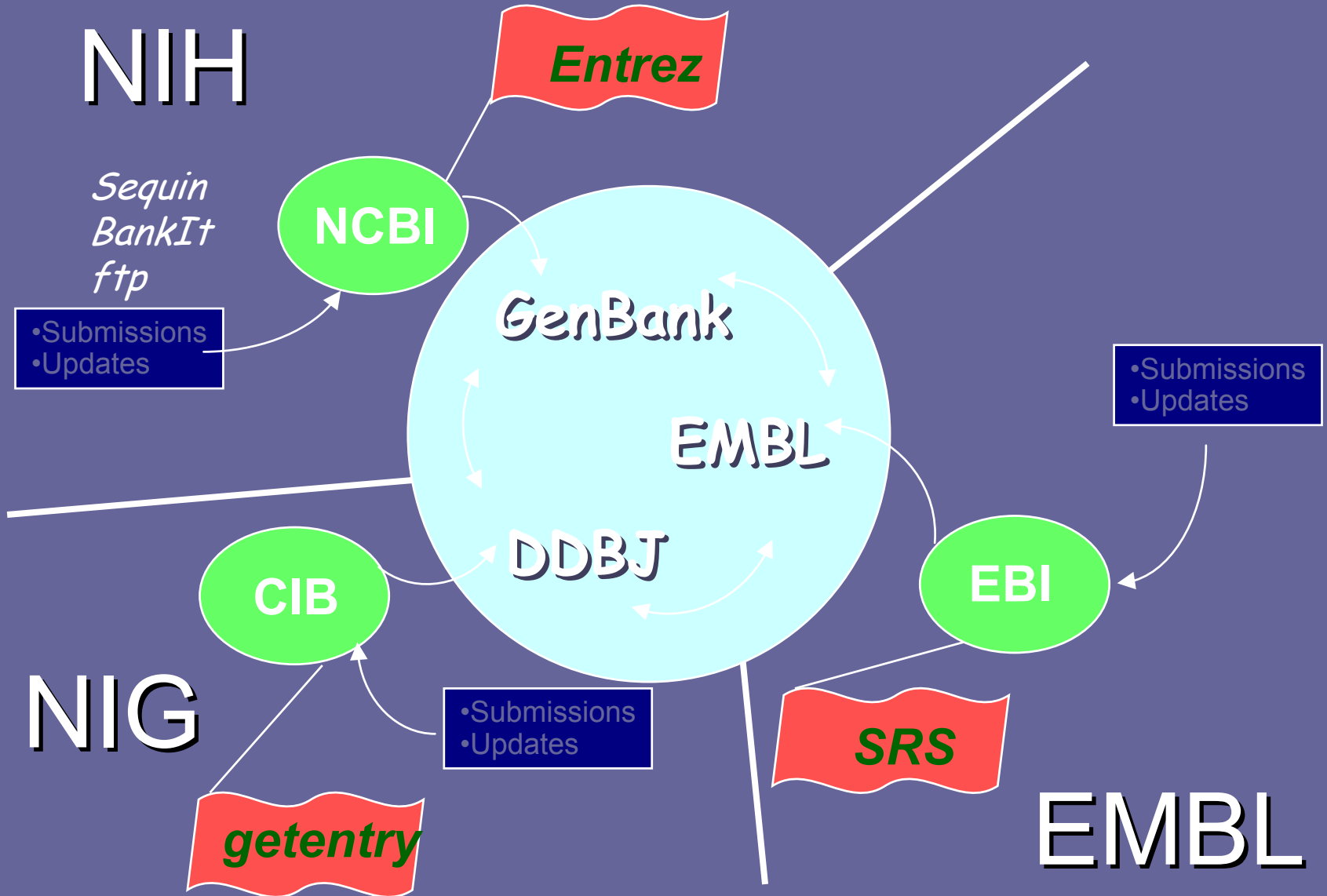
- Human curated
 - compilation and correction of data
 - **Example: SWISS-PROT, NCBI RefSeq mRNA**
- Computationally Derived
 - **Example: UniGene**
- Combinations
 - **Example: NCBI Genome Assembly**

What is GenBank?

NCBI's Primary Sequence Database

- **Nucleotide only sequence database**
- **GenBank Data**
 - Direct submissions individual records (BankIt, Sequin)
 - Batch submissions via email (EST, GSS, STS)
 - ftp accounts established for sequencing centers
- **Data shared amongst three collaborating databases:**
 - GenBank
 - DNA Database of Japan (DDBJ).
 - European Molecular Biology Laboratory Database (EMBL)

The International Nucleotide Sequence Database Collaboration



GenBank: NCBI's Primary Sequence Database

Release 133

December 2002

22,318,883

Records

28,507,990,166

Nucleotides

110,000 +

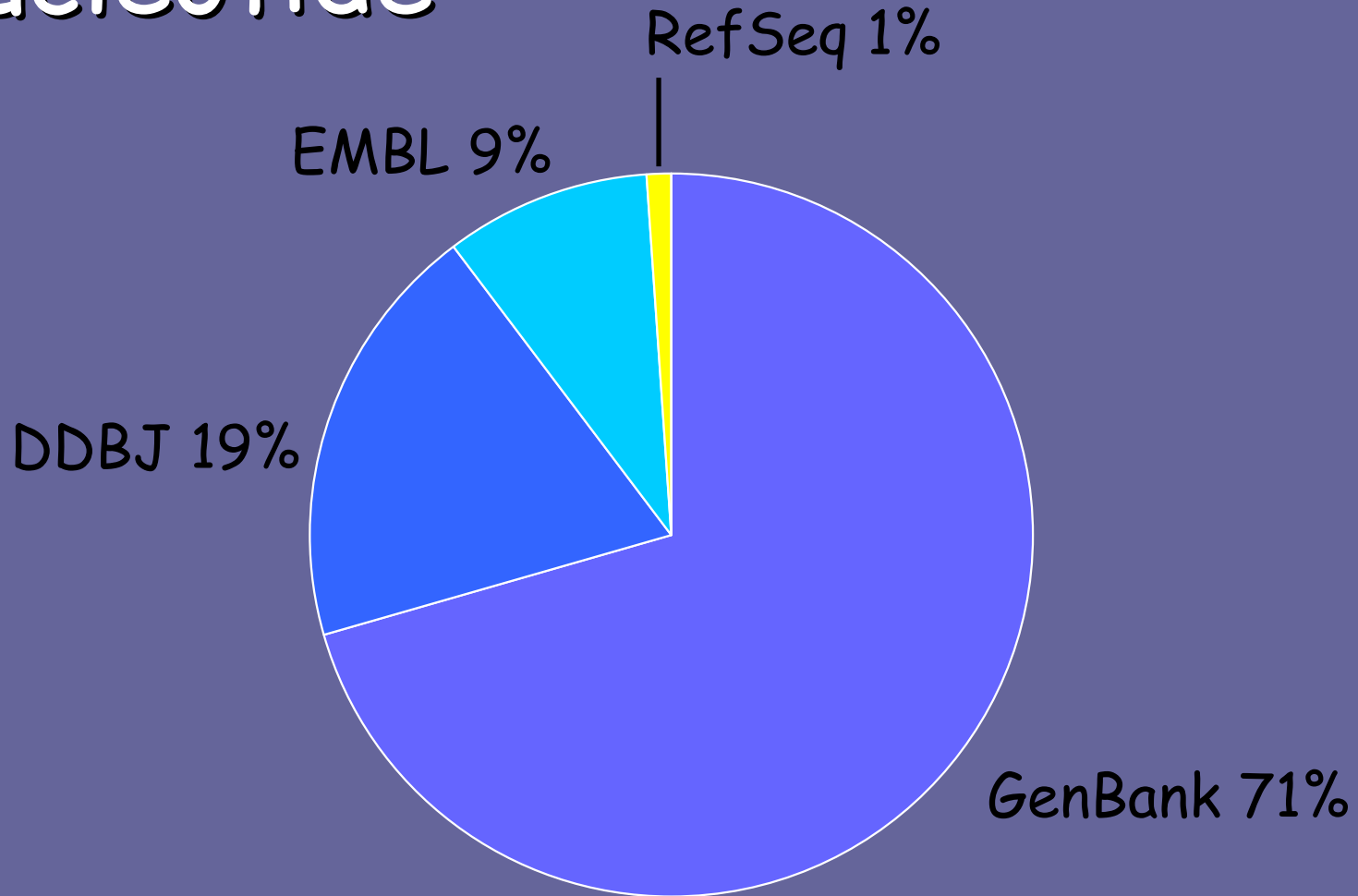
Species

- full release every two months
- incremental and cumulative updates daily
- available only through internet

<ftp://ftp.ncbi.nih.gov/genbank/>

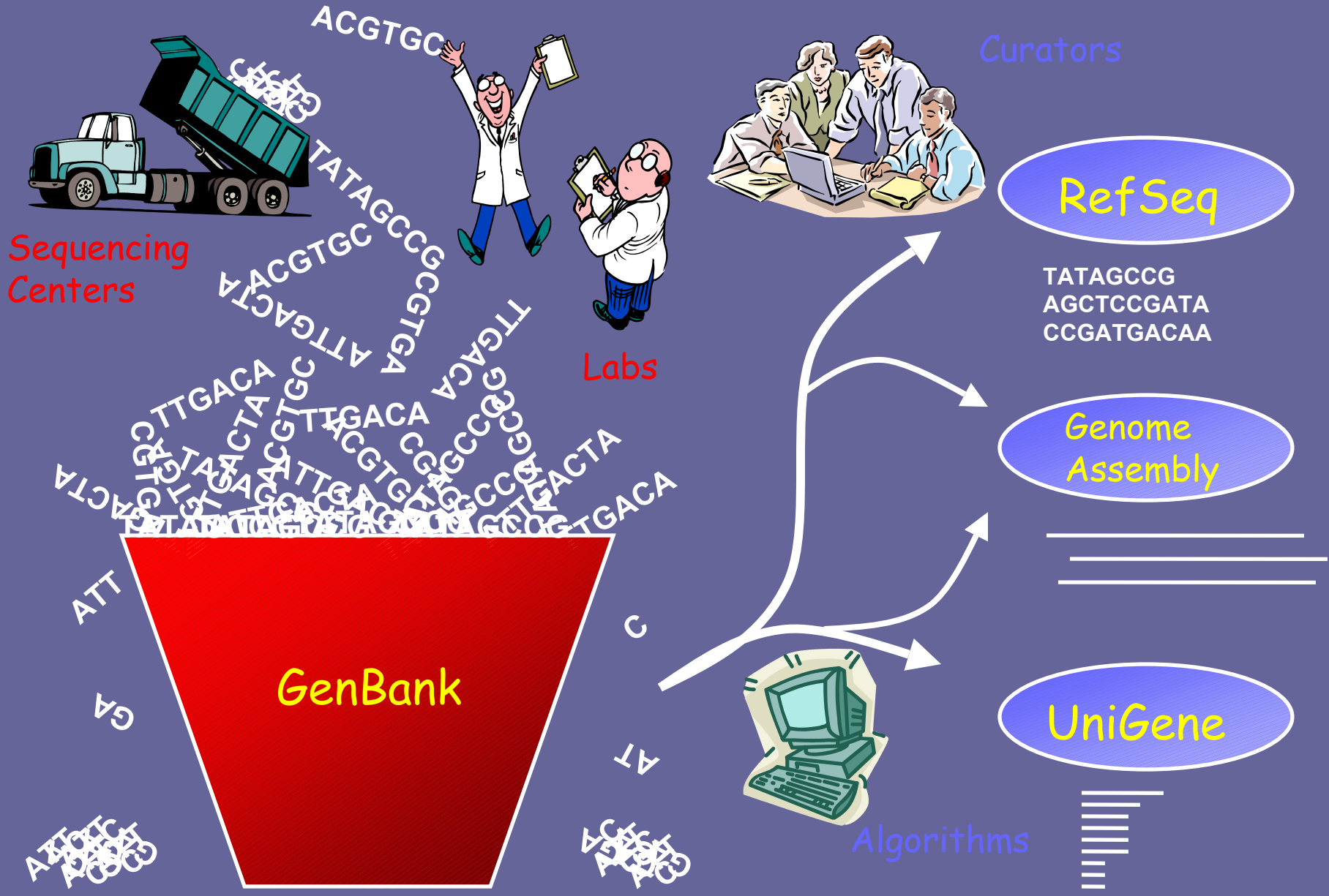
>90 Gigabytes of data

Entrez Nucleotide



23,464,770 records

Primary vs. Derivative Databases



Sequencing Centers

Labs

Curators

RefSeq

TATAGCCG
AGCTCCGATA
CCGATGACAA

Genome Assembly

=====
=====
=====

UniGene

Algorithms

GenBank

Traditional GenBank Divisions

- Direct Submissions (Sequin and BankIt)
- Accurate
- Well characterized

BCT	Bacterial and Archeal
INV	Invertebrate
MAM	Mammalian (ex. ROD and PRI)
PHG	Phage
PLN	Plant and Fungal
PRI	Primate
ROD	Rodent
SYN	Synthetic (cloning vectors)
VRL	Viral
VRT	Other Vertebrate

A Traditional GenBank Record

Locus Field

Molecule Type

Modification Date

Definition Line

GI (GenInfo)

Keywords

Taxonomy

```
LOCUS      AF437885                861 bp    mRNA     linear   INV 08-JUL-2002
DEFINITION Anopheles gambiae odorant binding protein (OBP-2) mRNA comple
           cds.
ACCESSION  AF437885
VERSION    AF437885.1  GI:1907
KEYWORDS   Anopheles gambiae; Odorant binding protein; OBP-2;
SOURCE     Anopheles gambiae (Malaria mosquito)
           yeota; Neozoa; Arthropoda; Hexapoda; Insecta; Pterygota;
           Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
           Anopheles.
REFERENCE  1 (bases 1 to 861)
AUTHORS    Biessmann, H., Walter, M.F., Dimitratos, S. and Woods, D.
TITLE      Isolation of cDNA clones encoding putative odourant binding
           proteins from the antennae of the malaria-transmitting mosquito,
           Anopheles gambiae
JOURNAL    Insect Mol. Biol. 11 (2), 123-132 (2002)
MEDLINE    21963933
PUBMED     11966877
REFERENCE  2 (bases 1 to 861)
AUTHORS    Biessmann, H., Walter, M.F., Dimitratos, S. and Woods, D.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2001) Dev. Biol. Center, Univ. of California,
           Irvine, CA 92697, USA
```

A Traditional GenBank Record

```
FEATURES             Location/Qualifiers
    source            1..861
                     /organism="Anopheles gambiae"
                     /db_xref="taxon:7165"
    gene               1..861
                     /gene="OBP-2"
    CDS              69..542
                     /gene="OBP-2"
                     /codon_start=1
                     /product="odorant binding protein"
                     /protein_id="AAL84180.1"
                     /db_xref="GI:19071274"
                     /translation="MLVQDSPLLLLVLLLVTQCLDGADCSTTTTQRPAPRRDGQYPPP
ETLAFLRPLGKLCLEETGVSPEAVKRFSDADPFDDNRALKCYMDCMERVTNVTDDRGE
LHMGKLLLEHVPTEFEDIALRMGVRCTRPKGKDV CERAFWFHKC WKTSDPVHYYLV"
```

BASE COUNT 223 a 201 c 242 g 195 t

ORIGIN

```
   1 aatataaaac cttccgcacg tgaggctggt cgtcgaccgt tttgtcggct cgcacagagg
  61 gcacagatat gttgggtcaa gattccccac tgttgctgct ggtgctactg ttagtgacgc
 121 agtgcttgga tggggcggat tgtagcacca ccaccactca acggccagcg ccgcgccggg
 181 acgggcaata tcctccgcca gagactcttg cctttctgcg accactgggc aagctgtgcc
 241 tggaggagac ggggtgtagc ccggaggcgg ttaagcgatt cagtgacgcc gatccgttcg
 301 acgataatcg tgcgctcaag tgctacatgg actgtatggt ccgagtaacg aacgtgaccg
 361 atgatcgcgg tgagctgcac atgggcaagc tgctcgaaca tgtgccgaca gagtttgaag
 421 acattgcact gcgaatgggc gtgaggtgta ctcggcccaa gggcaaggat gtgtgcgaac
 481 gggcgttctg gtttcacaag tgctggaaaa cgtcggacce tgtgcactac tatctggtgt
 541 gagaagcgac aaaagtcaac cgggtttggt gtggtgacce attaagtaag tatctagcta
 601 cattaagaca cacactgttg cctgctgaat gcctgcctgc tgaccattat acatcggcac
 661 aaaaaacaca agtgtgtgtg tgcgtgtctc tgtgatctga ggatggatgg tcacgaatgc
 721 gccacatcgt agagcaatca atcaatcaat gtacaacgtc ccgggagggt acgcagaatg
 781 gaagtacgag ttcaacgtga ccgtttttat gattgattct tgaataaaca gtaagcacga
 841 aaaaaaaaaa aaaaaaaaaa a
```

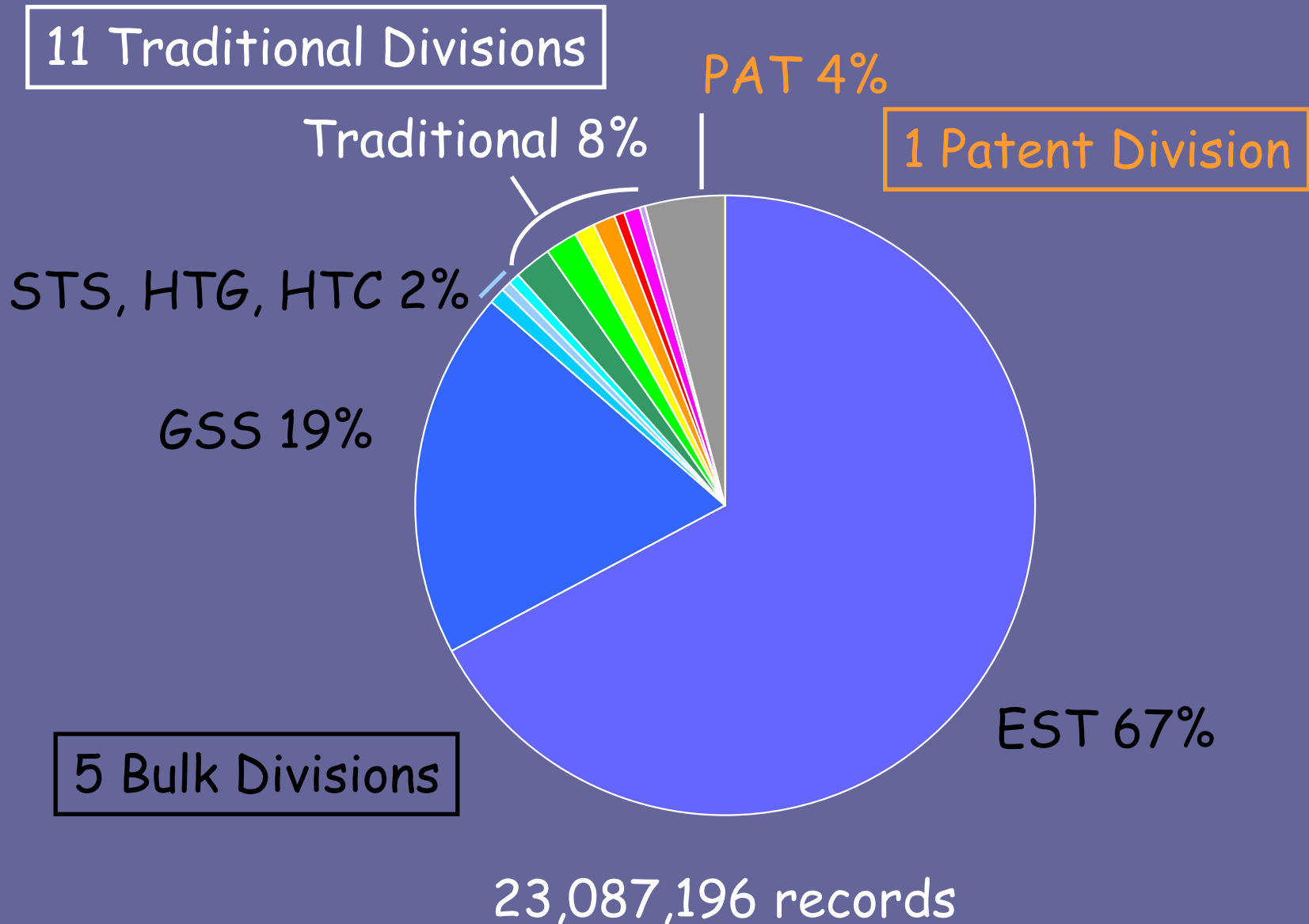
//

Bulk Sequence Divisions of GenBank

- Batch Submissions (email and ftp)
- Inaccurate
- Poorly Characterized

EST	Expressed Sequence Tag
STS	Sequence Tagged Site
GSS	Genome Survey Sequence
HTG	High Throughput Genomic
HTC	High Throughput cDNA

Organization of GenBank



EST Division: Expressed Sequence Tags

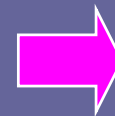
>IMAGE:275615 5' mRNA sequence

GACAGCATTCTGGGCCGAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGCGCTACTCTCTCTT
TCTGGCCTGGAGGTATCCAGCGTACTCCAAAGATTCAGGTTTACTCACGTCATCCAGCAGAG
AATGGAAAGTCAAATTTCTGAATTGCTATGTGTCTGGGTTTCATCCATCCGACATTGAAG
TTGACTTACTGAAGAATGGAGAGAGAATTGAAAAAGTGGAGCATTACAGACTTGTCTTTTCAG
CAAGGACTGGTCTTTCTATCTCTTGTACTACACTGAATTCACCCCCACTGAAAAAGATGAGT
ATGCCTGCCGTGTTGAACCATGTNGACTTTGTCACAGNCCCAAGTTNAGTTTAAGTGGGNA
TCGAGACATGTAAGGCAGGCATCATGGGAGGTTTTGAAGNATGCCGCNTTTTGGATTGGGA
TGAATTCCAAATTTCTGGTTTGCTTGNNTTTTTTAATATTGGATATGCTTTTG

>IMAGE:275615 3' mRNA sequence

NTCAAGTTTTATGATTTATTTAACTTGTGGAACAAAAATAAACCAGATTAACCACAACC
ATGCCTTACTTTATCAAATGTATAAGANGTAAATATGAATCTTATATGACAAAATGTTTC
ATTCATTATAACAAATTTCCAATAATCCTGTCAATNATATTTCTAAATTTTCCCCCAAATT
CTAAGCAGAGTATGTAAATTGGAAGTTAACTTATGCACGCTTAACTATCTTAACAAGCTTT
GAGTGCAAGAGATTGANGAGTTCAAATCTGACCAAGATGTTGATGTTGGATAAGAGAATT
CTCTGCTCCCACCTCTANGTTGCCAGCCCTC

make cDNA
library




80-100,000 unique
cDNA clones in library


What is UniGene?

A gene-oriented view of sequence entries

- MegaBlast-based automated sequence clustering
- Nonredundant set of gene-oriented clusters
- Each cluster represents a unique gene
- Provides information on tissue-specific expression and map locations
- Includes well-characterized genes and novel ESTs
- Useful for gene discovery and selection of mapping reagents

Organisms Represented in UniGene

	Chordata	
	Mammalia	
	<u>Bos taurus</u> (cow)	10,702 entries
	<u>Homo sapiens</u> (human)	121,062 entries
	<u>Mus musculus</u> (mouse)	88,647 entries
	<u>Rattus norvegicus</u> (rat)	62,990 entries
	<u>Sus scrofa</u> (pig)	14,561 entries
	Amphibia	
	<u>Xenopus laevis</u> (frog)	19,128 entries



	Chlorophyta	
	Chlorophyceae	
	<u>Chlamydomonas reinhardtii</u>	6,517 entries

Just In...

C.elegans

Ciona intestinalis

Gallus gallus

	Actinopterygii	
	<u>Danio rerio</u> (zebrafish)	15,963 entries
	Arthropoda	
	Insecta	
	<u>Anopheles gambiae</u> (malaria mosquito)	2,568 entries
	<u>Drosophila melanogaster</u> (fruit fly)	11,292 entries
	Embryophyta	
	Eudicotyledons	
	<u>Arabidopsis thaliana</u> (thale cress)	27,148 entries
	<u>Glycine max</u> (soybean)	8,729 entries
	<u>Lycopersicon esculentum</u> (tomato)	3,703 entries
	<u>Medicago truncatula</u> (barrel medic)	5,906 entries
	Liliopsida	
	<u>Hordeum vulgare</u> (barley)	7,683 entries
	<u>Oryza sativa</u> (rice)	16,718 entries
	<u>Triticum aestivum</u> (wheat)	13,284 entries
	<u>Zea mays</u> (maize)	12,072 entries

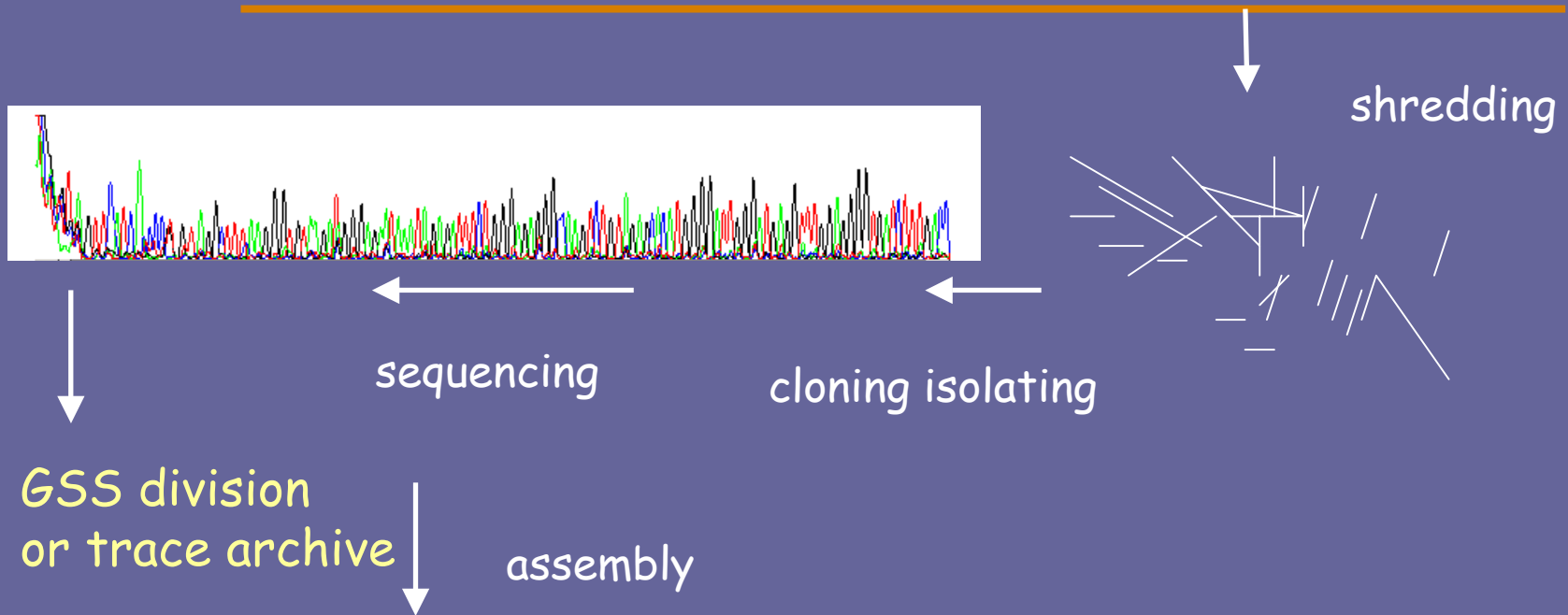
EST hits to *Homo sapiens* muscle creatine kinase mRNA

Query Sequence →



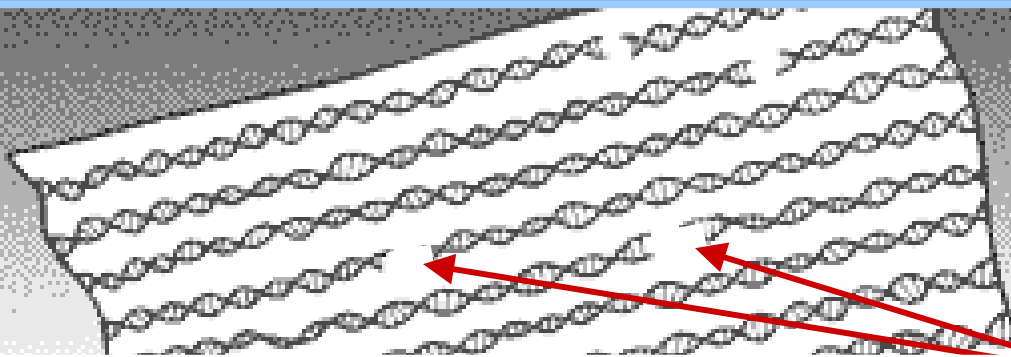
Genome Sequencing

Whole BAC insert (or genome)

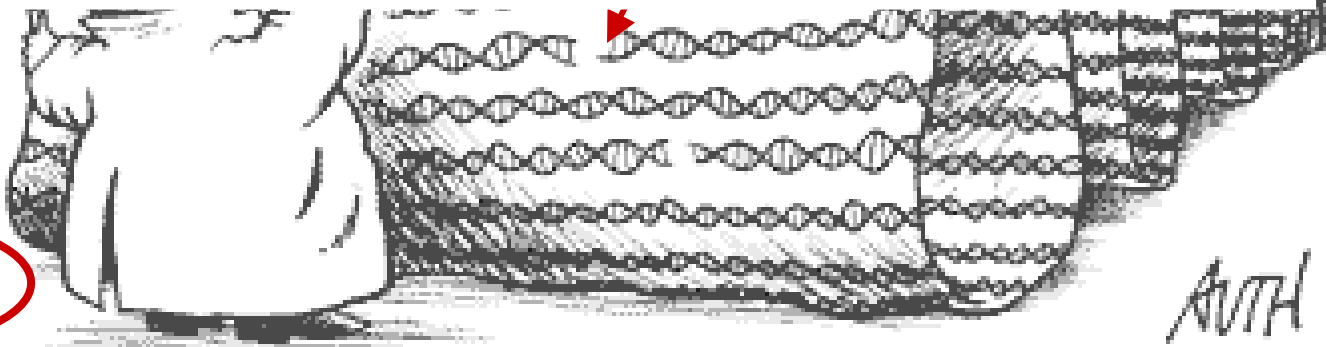


Draft Sequence (HTG division)

Working Draft Sequence



```
61681 gcecggtect etetegggett cetetgceat eggteectect tttatgggtcc agagctcett
61741 cegtgggata cgaggecaggt ggcacccga cttgtateca ctcacggcgt ggctcacc
61801 egtteccaag gcteteggcc acgccccctc gccacatacc cccatctccc atcacaggcc
61861 gggggcacc cgtggctctg catactccgg ggggggggggt ttggcgcacc agcaccggcg
61921 agacctatag agacgggccc agagaaaagg agannnnnnn nnnnnnnnnn nnnnnnnnnn
61981 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
62041 nnnnnnnnnn nnnngttcctc ggccatcgggt cctccttcta tgggtccagag ctccttcctt
62101 gggatccgag gcaggtaccgc cccgcacgtg tateactca cgccttggcc tcaccctggt
62161 ccenegggett tcggccaccgc cccctctcca cagggccata tgtagaatc tgccttgaca
62221 catgtgcgat gtcttaaat gtatggcaag tgtaaacaca cgtatcggat atgggtcaca
62281 attaaaagaa cttgtaaacc gacagataaa caaatctgat atacagcagg cagcaaatcg
```



HTG Division: High Throughput Genome

phase 1 → ← → → ← HTG
Acc = AC109609.1

phase 2 → → → HTG
Acc = AC109609.6

phase 3 → ROD
Acc = AC109609.10

40,000 to > 350,000 bp

HTG Division: High Throughput Genome

LOCUS	AC109609	154774 bp	DNA	linear	HTG 06-FEB-2002
DEFINITION	Mus musculus clone rp23-167h1, WORKING DRAFT SEQUENCE, 19 unordered pieces.				
ACCESSION	AC109609				
VERSION	AC109609.1 GI:18543009				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
LOCUS	AC109609	167074 bp	DNA	linear	HTG 07-MAY-2002
DEFINITION	Mus musculus clone rp23-167h1 strain C57BL/6J, WORKING DRAFT SEQUENCE, 3 ordered pieces.				
ACCESSION	AC109609				
VERSION	AC109609.6 GI:20087170				
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.				
LOCUS	AC109609	167617 bp	DNA	linear	ROD 10-AUG-2002
DEFINITION	Mus musculus chromosome 17 clone rp23-167h1 strain C57BL/6J, complete sequence.				
ACCESSION	AC109609				
VERSION	AC109609.10 GI:22203259				
KEYWORDS	HTG.				

NCBI's Third Party Annotation (TPA) Database



- NCBI now accepts the submission of new annotations of **existing** GenBank sequences;
- Facilitates the annotation of genomes by experts;

A Sample TPA record

LOCUS BK000147 3230 bp mRNA linear PRI 28-SEP-2002
DEFINITION TPA: Homo sapiens harmonin isoform b3 (USH1C) mRNA, complete cds;
alternatively spliced.
ACCESSION BK000147
VERSION BK000147.1 GI:23342606
KEYWORDS Third Party Annotation; TPA.
SOURCE human.
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3230)
AUTHORS Verpy,E., Leibovici,M., Zwaenepoel,I., Liu,X.Z., Gal,A., Salem,N.,
Mansour,A., Blanchard,S., Kobayashi,I., Keats,B.J., Slim,R. and
Petit,C.
TITLE A defect in harmonin, a PDZ domain-containing protein expressed in
the inner ear sensory hair cells, underlies Usher syndrome type 1C
JOURNAL Nat. Genet. 26 (1), 51-55 (2000)
MEDLINE [20428180](#)
PUBMED [10973247](#)
REFERENCE 2 (bases 1 to 3230)
AUTHORS Verpy,E., Leibovici,M., Zwaenepoel,I., Blanchard,S. and Petit,C.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Unite de Genetique des Deficits Sensoriels,
Institut Pasteur, 25 rue du Dr. Roux, Paris 75015, France
COMMENT THIRD PARTY ANNOTATION DATABASE: This TPA record uses data from
DDBJ/EMBL/GenBank entry AC005137
PRIMARY TPA-SPAN PRIMARY IDENTIFIER PRIMARY SPAN COMP

RefSeq: NCBI's Derivative Sequence Database

- **Curated transcripts and proteins**
 - reviewed
 - human, mouse, rat, fruit fly, zebrafish, arabidopsis
- **Human model transcripts and proteins**
- **Assembled Genomic Regions (contigs)**
 - draft human genome
 - mouse genome
- **Chromosome records**
 - Microbial
 - viral
 - organelle

The RefSeq Accession Numbers

mRNA

NM_1

NP_12

NR_12

XM_1

XP_12

XR_12

Gene

NG_1

Assem

NT_1

NW_1

NC_1

NR_1

NCBI Reference Sequences

The Reference Sequence (RefSeq) collection aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms.

RefSeq standards serve as the basis for medical, functional, and diversity studies; they provide a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analyses. RefSeqs are used as a reagent for the functional annotation of some genome sequencing projects, including those of human and mouse.

What's New: The RefSeq collection for *Rattus norvegicus* now includes genomic scaffolds. Rat genomic RefSeq records, with the accession prefix NW_, are based on the genomic sequence provided by the [Rat Genome Sequencing Consortium \(RGSC\)](#). The rat RefSeq scaffolds are

Hot Spots

Documentation:

[Accession Key](#)

[Entrez Queries](#)

[FAQ](#)

[NCBI Handbook](#)

[RefSeq Production](#)

[Status Key](#)

FTP:

[RefSeq](#)

[Genomes](#)

Statistics:

[LocusLink Pipeline](#)

Related Resources:

[COGs](#)

[LocusLink](#)

[Map Viewer](#)

[UniGene](#)

[Entrez Genomes](#)

human
mouse
rat
fruit fly
zebrafish
Arabidopsis

(se)

e (human)

Arabidopsis)
obial

Curated RefSeq Records: NM_, NP_

LOCUS Lpl 3969 bp mRNA linear ROD 07-JAN-2002
DEFINITION Mus musculus lipoprotein lipase (Lpl), mRNA.
ACCESSION NM_008509
VERSION NM_008509.1 GI:6678709
KEYWORDS .
SOURCE Mus musculus.
ORGANISM [Mus musculus](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3969)
AUTHORS Kirchgessner,T.G., Svenson,K.L., Lusic,A.J. and Schotz,M.C.
TITLE The sequence of cDNA encoding lipoprotein lipase. A member of a
lipase gene family
JOURNAL J. Biol. Chem. 262 (18), 8463-8466 (1987)
COMMENT PROVISIONAL [REFSEQ](#): This record has not yet been subject to final
NCBI review. The reference sequence was derived from [M60847.1](#).
REFERENCE 2 (bases 1 to 3969)
AUTHORS Zechner,B., Neuman,T.C., Steiner,E. and Breslow,J.L.
COMMENT REVIEWED [REFSEQ](#): This record has been curated by NCBI staff. The
reference sequence was derived from [M15856.1](#).
Summary: LPL encodes lipoprotein lipase, which is expressed in
heart, muscle, and adipose tissue. LPL functions as a homodimer,
and has the dual functions of triglyceride hydrolase and
ligand/bridging factor for receptor-mediated lipoprotein uptake.
Severe mutations that cause LPL deficiency result in type I
hyperlipoproteinemia, while less extreme mutations in LPL are
linked to many disorders of lipoprotein metabolism.
analysis
JOURNAL Gene 107 (2), 247-258 (1991)
MEDLINE [92084117](#)
PUBMED [1748295](#)
REFERENCE 4 (bases 1 to 3969)

Entrez:
Linking and Neighboring

The Entrez Databases

The screenshot shows the NCBI Entrez search interface. At the top left is the NCBI logo. The main header features the word "Entrez" in a large, stylized font, with "search and retrieval system" underneath. Below this is a navigation bar with tabs for "PubMed", "Nucleotide", "Protein", "Genome", "Structure", and "PMC". The "PubMed" tab is selected. A search bar contains the text "Search PubMed" and a dropdown menu. To the right of the search bar are "Go" and "Clear" buttons. Below the search bar are links for "Limits", "Preview/Index", "History", and "Clipboard". On the left side, there is a vertical menu with links for "About Entrez", "SITE MAP", "PubMed Help", "Entrez Help", "The Entrez Databases", "Network Entrez", and "Batch Entrez". The main content area contains the text: "Entrez is a retrieval system for searching several linked databases. It provides access to:" followed by a list of database descriptions with underlined links: PubMed, Nucleotide, Protein, Structure, Genome, PopSet, OMIM, Taxonomy, Books, ProbeSet, 3D Domains, UniSTS, SNP, CDD, Journals, UniGene, and PMC.

NCBI

Entrez
search and retrieval system

PubMed Nucleotide Protein Genome Structure PMC

Search PubMed for [] Go Clear

Limits Preview/Index History Clipboard

About Entrez
SITE MAP

PubMed Help
help
documentation for
PubMed

Entrez Help
help
documentation for
the rest of Entrez

The Entrez
Databases
cross-references
and information

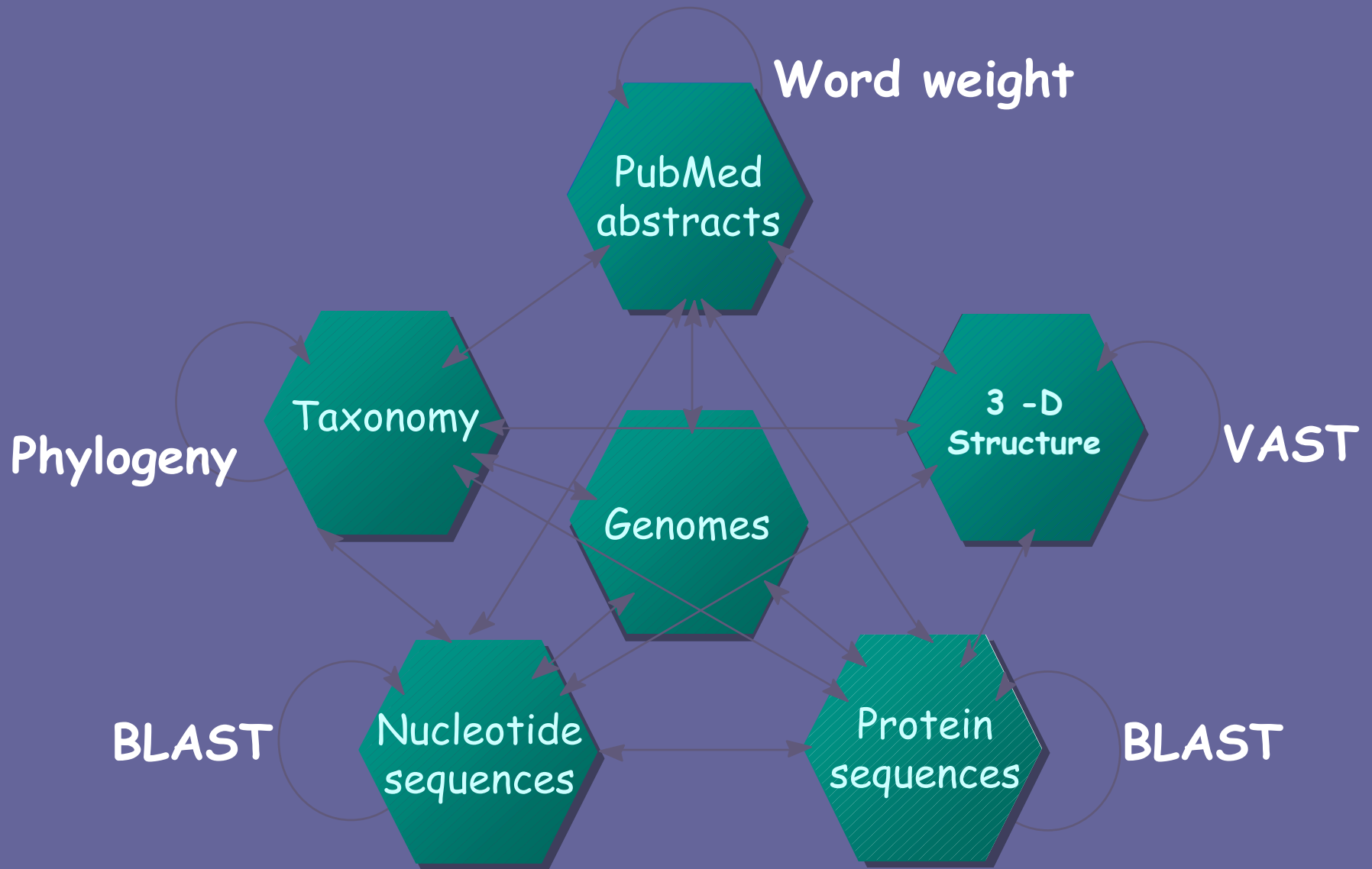
Network Entrez
TCP/IP-based
client-server
version of WWW
Entrez

Batch Entrez

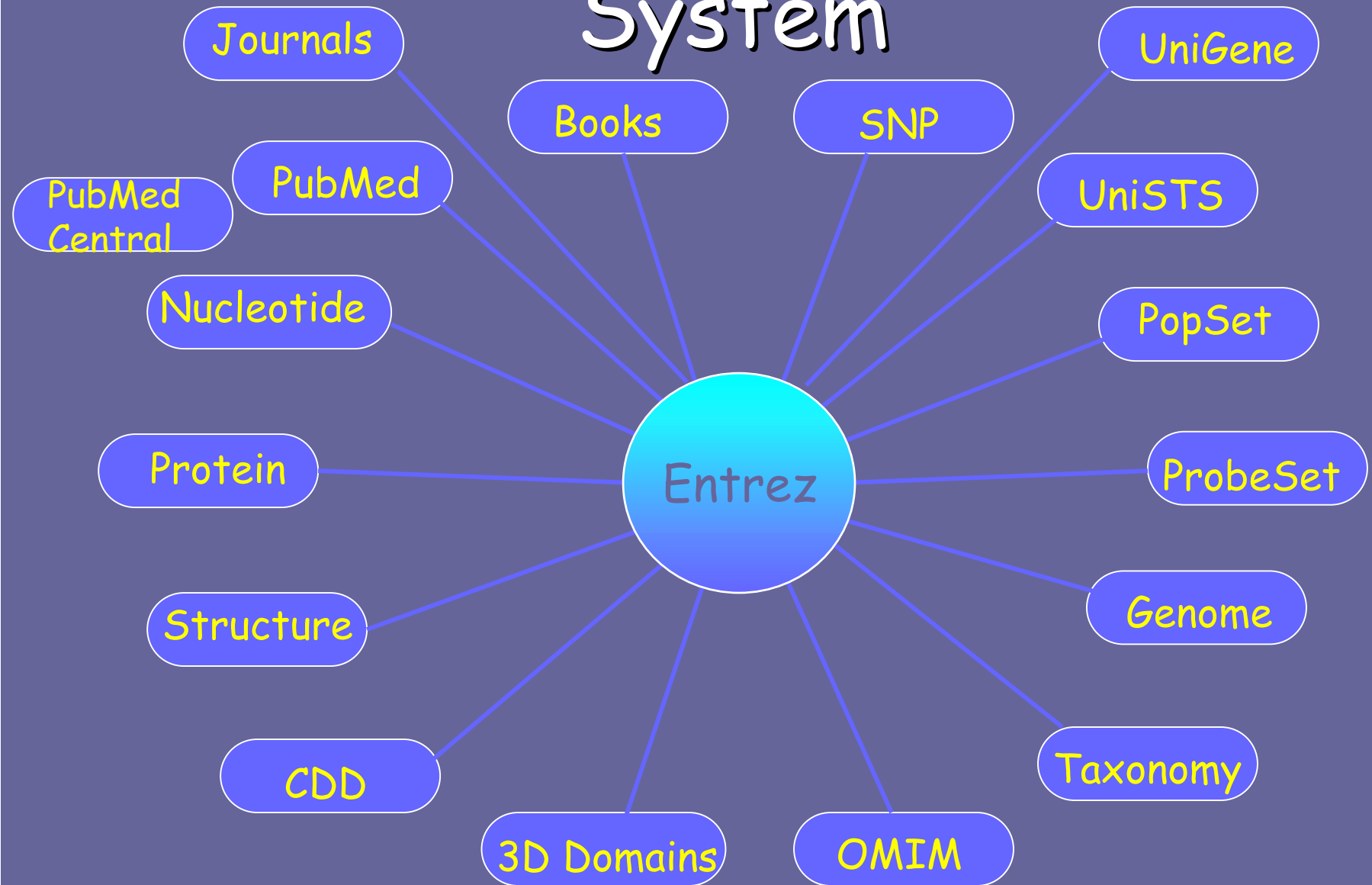
Entrez is a retrieval system for searching several linked databases. It provides access to:

- [PubMed](#): The biomedical literature (PubMed)
- [Nucleotide](#) sequence database (Genbank)
- [Protein](#) sequence database
- [Structure](#): three-dimensional macromolecular structures
- [Genome](#): complete genome assemblies
- [PopSet](#): population study data sets
- [OMIM](#): Online Mendelian Inheritance in Man
- [Taxonomy](#): organisms in GenBank
- [Books](#): online books
- [ProbeSet](#): gene expression and microarray datasets
- [3D Domains](#): domains from Entrez Structure
- [UniSTS](#): markers and mapping data
- [SNP](#): single nucleotide polymorphisms
- NEW** [CDD](#): conserved domains
- NEW** [Journals](#): journals in Entrez
- NEW** [UniGene](#): gene-oriented clusters of transcript sequences
- NEW** [PMC](#): full-text digital archive of life sciences journal literature

Entrez: Database Integration



The (ever) Expanding Entrez System



Entrez Nucleotides

The screenshot shows the NCBI Entrez Nucleotides search interface. At the top, the NCBI logo is on the left, and a decorative graphic with the text 'Entrez Nucleotide' is on the right. Below the logo is a navigation bar with tabs for PubMed, Nucleotide, Protein, Genome, Structure, PopSet, Taxonomy, OMIM, and Books. The search bar contains the text 'glucose 6 phosphate dehydrogenase' and has 'Go' and 'Clear' buttons. Below the search bar are links for Limits, Preview/Index, History, Clipboard, and Details. On the left side, there is a sidebar with links for 'About Entrez', 'Search for Genes LocusLink', 'Entrez Nucleotide Help | FAQ', 'Batch Entrez', 'Check sequence revision history', and 'How to create WWW links to Entrez'. The main content area features a yellow box with text about the database, a digital display showing '30061326645', a 'Draft Human Genome' section with a link to 'human genome resources', and a 'Building the human genome' section with a paragraph of text.

NCBI

Entrez Nucleotide

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books

Search Nucleotide for **glucose 6 phosphate dehydrogenase** Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

Search for Genes 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

The Entrez Nucleotides database is a collection of sequences from several sources, including GenBank, RefSeq, and PDB. The number of bases grows at an exponential rate. Today's total is:

30061326645

Draft Human Genome

Explore [human genome resources](#) or browse the human genome sequence using the [Map Viewer](#).

Building the human genome

While the draft sequence of the human genome was officially launched early in 2001, data are produced on a continuous basis. This allows the reference human genome sequence to be further refined and completed as an ongoing process. At NCBI, new versions of the assembled genomic sequence are released every few weeks. The current version is listed as a build number on the [Genome View](#) page. Each build has an

Document Summaries:

glucose 6 phosphate dehydrogenase[All Fields] = 748 hits

Show:

Items 1-20 of 748

Page 1 of 38

Select page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) »

1: [NT_025965](#)

[MapView](#), [Protein](#), [Taxonomy](#)

Homo sapiens chromosome X reference genomic contig
[gi|22059389|ref|NT_025965.8|HsX_26121\[22059389\]](#)

2: [AE008786](#)

[Related Sequences](#), [Protein](#), [PubMed](#), [Taxonomy](#)

Salmonella typhimurium LT2, section 90 of 220 of the complete genome
[gi|16420461|gb|AE008786.1|\[16420461\]](#)

3: [AE008784](#)

[Related Sequences](#), [Protein](#), [PubMed](#), [Taxonomy](#)

Salmonella typhimurium LT2, section 88 of 220 of the complete genome
[gi|16420418|gb|AE008784.1|\[16420418\]](#)

4: [AE013827](#)

[Related Sequences](#), [Protein](#), [Taxonomy](#)

Yersinia pestis KIM section 227 of 415 of the complete genome
[gi|21959085|gb|AE013827.1|\[21959085\]](#)

5: [NM_017006](#)

[Related Sequences](#), [Protein](#), [PubMed](#), [Taxonomy](#)

Rattus norvegicus Glucose-6-phosphate dehydrogenase (G6pd), mRNA
[gi|8393380|ref|NM_017006.1|\[8393380\]](#)

Entrez Nucleotides: Limits

NCBI
PubMed Nucleotide
Search Nucleotide for glucose 6
Limits
Structure PopSet Taxonomy OMIM Books
Clear
Details

Limited to:

Title

exclude ESTs exclude

exclude patents

mRNA

Only from

Modification Date

Use the format YYYY/

- Accession
- All Fields
- Author Name
- EC/RN Number
- Feature key
- Filter
- Gene Name
- Issue
- Journal Name
- Keyword
- Modification Date
- Organism
- Page Number
- Primary Accession
- Properties
- Protein Name
- Publication Date
- SeqID String
- Sequence Length
- Substance Name
- Text Word
- Title Word

include GSS exclude working draft
above
mRNA Segmented Sequences
To
are optional.

Entrez Nucleotides: Preview/Index

for


Limits **Preview/Index** History Clipboard Details

Field: Title, Limits: **Exclude all of the above, Genomic DNA/RNA, mRNA**

Show: Items 1-20 of 77 Page 1 of 4 Select page: [1](#) [2](#) [3](#) [4](#)

1: [NM_017006](#) Related Sequences, Protein, PubMed, Taxonomy
Rattus norvegicus Glucose-6-phosphate dehydrogenase (G6pd), mRNA
gi|8393380|ref|NM_017006.1|[8393380]

2: [NM_000402](#) MapView, Related Sequences, OMIM, Protein, PubMed, Taxonomy
Homo sapiens glucose-6-phosphate dehydrogenase (G6PD), nuclear gene encoding mitochondrial protein, mRNA
gi|21614519|ref|NM_000402.2|[21614519]



Adding Terms: Preview/Index

Accession
All Fields
Author Name

Add Term(s) to Query or View Index:

- Enter a term in the text box; use the pull-down menu to specify a search field.
- Click Preview to add terms to the query box and see the number of search results, or click Index to view terms within a field.
- Multiple terms selected from Index will be ORed; click AND to add to search.

Organism Preview Index

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

green plants(2774214) Up
green poison berry(1)
green porcelain crab(11)
green pygmy goose(1)
green rice leafhopper(1)
green ringtail possum(1)
green rock lobster(2)
green sea mat(3)
green sea turtle(105)
green striped grasshopper(5) Down



Plant G6PD mRNAs

Show:

Items 1-20 of 32

Page 1 of 2

Select page: [1](#) [2](#)

1: [AY086213](#)

[Related Sequences, Protein, Taxonomy](#)

Arabidopsis thaliana clone 22483 mRNA, complete sequence
gi|21404923|gb|AY086213.1|[21404923]

2: [AY117271](#)

[Protein, Taxonomy](#)

Arabidopsis thaliana putative glucose-6-phosphate dehydrogenase (At1g09420) mRNA, complete cds
gi|21436352|gb|AY117271.1|[21436352]

3: [AB085846](#)

[Protein, Taxonomy](#)

Chlorella vulgaris g6pdh mRNA for glucose-6-phosphate dehydrogenase 1, partial cds
gi|21262178|dbj|AB085846.1|[21262178]

4: [NM_123441](#)

[Protein, Taxonomy](#)

Arabidopsis thaliana glucose-6-phosphate dehydrogenase, predicted mRNA
gi|18422022|ref|NM_123441.1|[18422022]

5: [NM_122970](#)

[Related Sequences, Protein, Taxonomy](#)

Arabidopsis thaliana glucose-6-phosphate dehydrogenase, predicted mRNA
gi|18421498|ref|NM_122970.1|[18421498]

Display: Formats, Links, and Neighbors

The screenshot shows a web interface for displaying sequence data. At the top, there is a 'Display' dropdown menu set to 'Summary' and a 'Save' button. Below this is a 'Show: 20' dropdown and 'Items 1-3 of 3'. The main content area lists three items:

- 1: [AY086213](#)
Arabidopsis thaliana clone
gi|21404923|gb|AY086213
- 2: [AY117271](#)
Arabidopsis thaliana putative
mRNA, complete cds
gi|21436352|gb|AY117271.1
- 3: [AB085846](#)
Chlorella vulgaris g6pdh mRN
cds
gi|21262178|dbj|AB085846.

A yellow callout box points to the 'Display' dropdown menu and contains the following list of options:

- Summary
- Brief
- ASN.1
- FASTA
- XML
- GenBank
- GI list
- LinkOut
- Nucleotide Neighbors
- Genome Links
- ProbeSet Links
- OMIM Links
- PopSet Links
- Protein Links
- PubMed Links
- SNP Links
- Structure Links
- Taxonomy Links
- UniSTS Links

On the right side of the interface, there is a 'Select page: 1 2' dropdown and a list of links for each item, such as 'Sequences, Protein, Taxonomy' and 'Protein, Taxonomy'.


```
>gi|603218|gb|U18238.1|MSU18238 Medicago sativa glucose-6-phosphate dehyd  
CCACCAGATATAATTAAGTAGATCAGAGTAGAAGAAGATGGGAACAAATGAATGGCATGTAGAAAGAAGA  
GATACCATACCTACTCAATCTCCTCTACCAACAGACCTACTTCAAAGCTCCGACACTCTCTATTCTTCTCC
```

FASTA definition line

```
>gi|603218|gb|U18238.1|MSU18238
```

gi number

Locus name

Accession number

Database identifiers

gb	GenBank
emb	EMBL
dbj	DDBJ
sp	SWISS-PROT
pdb	Protein Databank
pir	PIR
prf	PRF
ref	RefSeq

```
CCTTCAGTGTTCATCCGTTTGCAATGATCAAACCTTGTTCGATGAATAAATCTGATCTTGGAT  
GTTGAGAAACCGGTAGGGATCTAGAATCTGAAGAAGACTCAGTACT  
TGAAGAACCAATCGTATTGATCACTATTTAGGAAGACTAGTG  
ATGTTAGTACTTCGTTTTGCAAAATTCTTGCCCTCTGTGGAACCACATTGACAATGTC  
AGATAGTATTTAGAGAGGTTTGGACCTGGAATTATCCG  
AGATATCATTCAAACCGTTGGZ  
CCTGAGCACATTCGAGACCCCT  
TTCTTGACAAATATGAAGGATTCAAACACCCCGACTTTTGC  
AACTACTATTCTGCGGATCATTGTGAAAGCAGGGAAGGCC  
CTAAATTCTAGGAAGGCATGGTGACATTTTCAGGAGTAAAA  
AGCAAGGGAGAAACGAGATTTTACATGAAGCTTACGGTCAA  
GCAACCTGGACTGGAAACATATGGGCAACGATATCAAGGG  
ATAACCATTCCAGAGGCATAGGTGATCAACAACATTTTGTTC  
GCAGAGACGAATTAAGCCACAAAATTGATAGAGGGGAGTT  
GAAGCCGTTCTTACATATGAGTTATTAGAAAAAGCTGGA  
TATGTTCAAACACCCGGTGGACCAATTCATAATAAAACA  
AGGATTAGGATTATCAGATGAAATTTTCGTTATAATCTCTC  
TCATTTTGGGGTGTATAATCAAGCATTTAAGCGCGTGTGACACAGTTGTGTAAATAGATTTGGCTCTGA  
ATGAAAATAAACGGGAATTGTTTCTTTTTGTTTTA
```


Entrez Genome



NCBI Map Viewer

Genome

Taxonomy

Entrez

BLAST

Help

Search

Select Organism

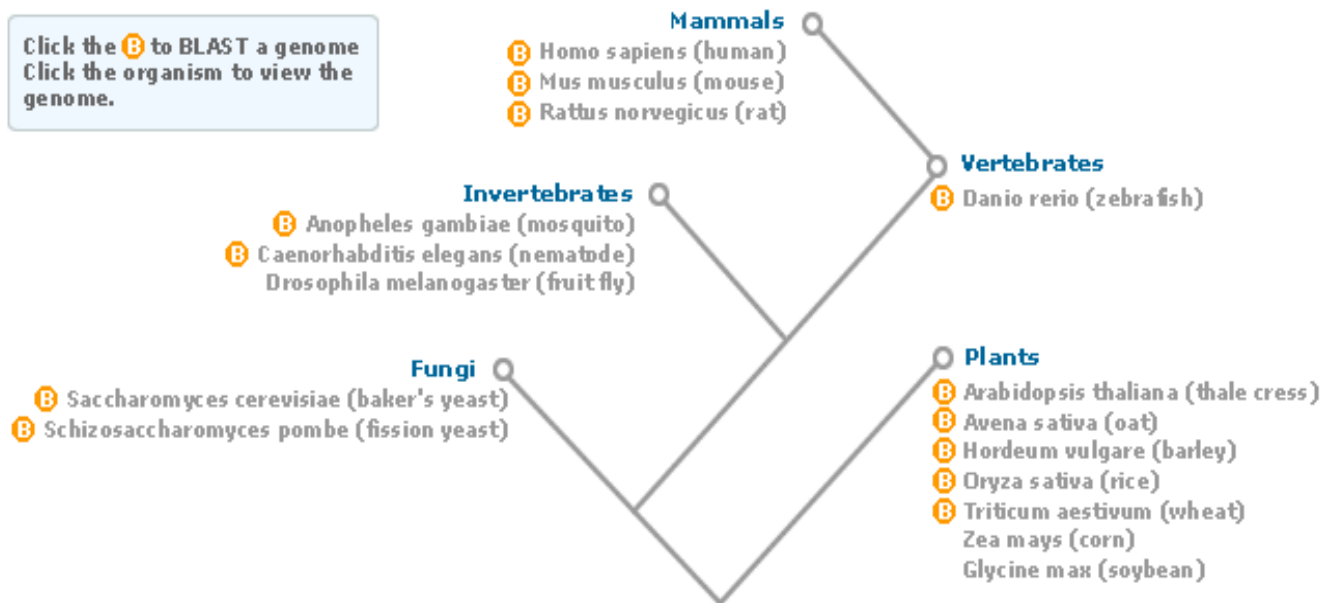
for

Go!

See new genome in MapViewer: [Plasmodium falciparum](#)

Almost complete DNA sequence of *Plasmodium falciparum* -- one of the parasites that causes malaria -- appears in [Science](#), 2002 Oct 4;298(5591):129 and [Nature](#), 2002 Oct 4

Click the **B** to BLAST a genome
Click the organism to view the genome.



See more about [Bacteria](#), [Organelles](#), [Viruses](#)

Web Resources

BLAST. Compare sequence to the genome or its gene products.

Cytogenetics. A cytogenetic resolution of FISH-mapped, sequence-tagged clones.

dbSNP. Database of SNPs and other genetic variations

e-PCR. Check your sequence for STSs and view in genomic context.

GEO. Gene Expression Omnibus, a public repository of expression data.

HomoloGene. Putative homologies among human, mouse, rat, and zebrafish.

Homology Map. Blocks of conserved synteny between mouse and human

Mouse Genome Resources

NCBI Web Resources:

BLAST:

Compare your sequence to different mouse specific sequences.

Clone Registry: Find information about specific BAC clones, including sequencing status, end sequence information and Fingerprint information.

dbSNP:

Database of SNPs and other genetic variation.

e-PCR: Check your sequence for STSs and view in genomic context.

GEO:

Gene Expression Omnibus, a public repository for expression data.

HomoloGene:

Putative homologies among human, mouse, rat, and zebrafish.

Homology Map:

Blocks of conserved synteny between mouse and human.

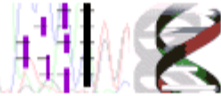
LocusLink:

Focal point for genes and associated information.

Genomic Biology
The value of
Pr
Ph
M
fr
mc
I
S
/



Zebrafish Genome Resources



PubMed Entrez BLAST OMIM Taxonomy Structure

Search for

Genomic Biology

[The Danio rerio Sequencing Project](#)

[SSAHA](#)

[What is SSAHA Search the Trace Repository with SSAHA](#)

[Search Tool for Zebrafish, Fugu and Human Orthologs](#)

Zebrafish Genome Resources

Welcome to the zebrafish genome resources page. This homepage will bring together information on diverse zebrafish-related resources from multiple centers: sequence, mapping and clone information as well as pointers to strain and mutant resources. We encourage your suggestions. Stay tuned for more information.

This month in...

[PubMed](#)

[Genbank](#)

Meetings and Courses News

NEW [Sanger Releases Preliminary Zebrafish Genome Assembly](#) [More...](#)

[International Conference on Zebrafish Development & Genetics University of Wisconsin, Madison June 12-16, 2002](#) [More...](#)

1 March 2002 [Zebrafish Reference Sequences Released](#) [More...](#)

[Trans-NIH ZebraFish Initiative](#)

[Zebrafish Information](#)

[ZFIN](#)

[RHdb](#)

[WashU-Zebrafish Genome Resources Project](#)

[Radiation Hybrid Maps](#)

[T51](#)

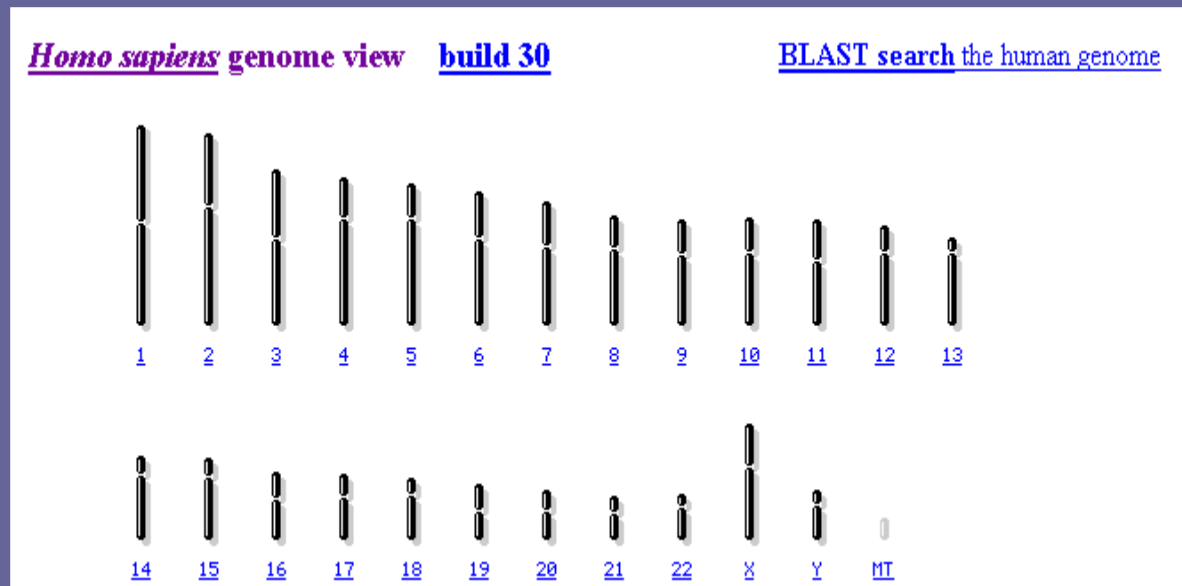
[Map-Tübingen](#)

[T51-Zon](#)

[RHMMapper](#)

[LN54](#)

The Map Viewer: a common platform for integrated display



The Map Viewer

NCBI Entrez BLAST OMIM Taxonomy Structure

Search Find in This View Find Advanced Search

[MapViewer Home](#)
[Map Viewer Help](#)
[Human Maps Help](#)
[FTP](#)
[Chr. Y Resource](#)

Data As Table View
Maps&Options
 Compress Map

Region Shown:

 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](#)

Master Map: [Genes On Sequence](#) **Maps & Options**

Total Genes On Chromosome: **161**

Region Displayed: **0-50M bp** [Download/View Sequence/Evidence](#)

Genes Labeled: **20** Total Genes in Region: **161**

Contig	Uni...	Genes_seq	symbol	orient.	links	evidence	cyto.	full name
Yp11.32			SRY	↑	sv ev hm seq mm	C	Yp11.3	sex determining region Y
Yp11.31			TGIF2LY	↓	sv ev - seq mm	C	Yp11.2	TGFB-induced factor 2-like, Y-linked
			TBL1Y	↓	sv ev - seq mm	C	Yp11.2	transducin (beta)-like 1Y-linked
Yp11.2			TTTY12	↓	sv ev - seq mm	C	Yp11.2	testis-specific transcript, Y-linked 12
			TSPYQ1	↓	sv ev - seq mm	C	Yp11.2	TSPYq1
Yp11.1			TTTY2	↓	sv ev - seq mm	C	Yp11.1	testis-specific transcript, Y-linked 2
Yq11.1			RPS24P1	↓	- - - seq -	C	Y	ribosomal protein S24 pseudogene 1
Yq11.21			USP9Y	↓	sv ev - seq mm	C	Yq11.2	ubiquitin specific protease 9, Y chromosome (f
Yq11.221			DBY	↓	sv ev hm seq mm	C	Yq11	DEAD/H (Asp-Glu-Ala-Asp/His) box polypept
Yq11.222			TMSB4Y	↓	sv ev - seq mm	C	Yq11.1	thymosin, beta 4, Y chromosome
Yq11.223			XKRY	↑	sv ev - seq mm	C	Yq11.221	X.Kell blood group precursor-related, Y-linked
Yq11.23			ACTGP2	↓	- - - seq -	C	Yq11	actin, gamma pseudogene 2
			CYorf14	↑	sv ev - seq mm	C	Yq11.221	chromosome Y open reading frame 14
			CYorf15B	↓	sv ev - seq mm	C	Yq11.221	chromosome Y open reading frame 15B
			EIF1AY	↑	sv ev hm seq mm	C	Yq11.221	eukaryotic translation initiation factor 1A, Y ch
			TTTY13	↑	sv ev - seq mm	C	Yq11.222	testis-specific transcript, Y-linked 13
			TTY7	↑	sv ev - seq mm	C	Yq11.222	testis transcript Y7
			VCY2	↓	sv ev - seq mm	C	Yq11	variable charge, Y chromosome, 2
Yq12			DAZ2	↓	sv ev - seq mm	C	Yq11.223	deleted in azoospermia 2

Entrez PubMed

□ 1: Plant Mol Biol 1995 Aug;28(5):885-900

[Related Articles](#), [Nucleotide](#), [Protein](#), [NEW Books](#), [LinkOut](#)

Stress responses in alfalfa (*Medicago sativa* L.) XIX. Transcriptional activation of

Stress responses in alfalfa (*Medicago sativa* L.) XIX. [Transcriptional activation](#) of oxidative [pentose](#) phosphate pathway genes at the onset of the isoflavonoid phytoalexin response.

Fahrendorf T, Ni W, Shorrosh BS, Dixon RA.

Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, OK 73401, USA.

We have isolated [cDNA clones](#) encoding the [pentose](#) phosphate [pathway enzymes](#) 6-phosphogluconate [dehydrogenase](#) (6PGDH, EC 1.1.1.44) and [glucose 6-phosphate dehydrogenase](#) (G6PDH, EC 1.1.1.49) from alfalfa (*Medicago sativa* L.). These exhibit extensive nucleotide and [amino acid sequence](#) similarity to the corresponding genes from bacteria, [Drosophila](#) and [mammals](#). Transcripts encoding both enzymes are expressed at high levels in roots and nodules. Exposure of alfalfa suspension cells to an elicitor from [yeast cell](#) walls results in coordinated increases in [transcription rates](#) for both genes, followed by increased steady state transcript levels but only slightly increased extractable enzyme activities, at the onset of accumulation of isoflavonoid phytoalexins. Levels of [NADPH](#) and [NADP](#) remain relatively constant in alfalfa cells following elicitation. The rapid [transcriptional activation](#) of 6PGDH and G6PDH does not therefore appear to be a response to altered pyridine nucleotide [redox state](#). These genes appear to respond to early events in elicitor-mediated signalling rather than to subsequent elicitor-induced changes in secondary metabolism. [Hydrogen peroxide](#), a potential signal for elicitation of anti-oxidative genes in biologically stressed [plant cells](#) did not induce 6PGDH or G6PDH transcripts or enzymatic activity.

Online Books

NCBI **Books**

PubMed Nucleotide Protein G

Search Books for cDNA clones

Limits Preview/Index

Display Books Save

20 items in Molecular Biology of the Cell, 3rd ed. Lodish, Harvey; Berk, Arnold J.; Martin, James E. New York: W H Freeman & Co. 1996.

11 items in C. elegans: A Practical Approach, 2nd ed. Riddle, Donald L.; Blumenthal, David. Plainview (NY): Cold Spring Harbor Laboratory Press. 1997.

9 items in Molecular Biology of the Cell, 4th ed. Alberts, Bruce; Bray, Dennis; Johnson, Alexander D.; Lewis, Julian; Raff, Martin; Roberts, Keith; Waterman, Peter. New York and London: Garland Science. 2002.

6 items in Introduction to Molecular Biology and Biotechnology, 2nd ed. Griffiths, Anthony J.F.; Miller, John H. New York: W H Freeman & Co. 1999.

6 items in Cancer Medicine, 5th ed. Bast, Robert C.; Kufe, Donald W.; Frei, Emil, editors. Canada: BC Decker Inc. 2004.

3 items in Retroviruses: A Practical Approach, 2nd ed. Coffin, John M.; Hughes, Michael. Plainview (NY): Cold Spring Harbor Laboratory Press. 1997.

3 items in Modern Molecular Biology, 2nd ed. Griffiths, Anthony J.F.; Miller, John H. New York: W H Freeman & Co. 1999.

About Entrez

Books

Overview

Using the books

Information for authors and publishers

Contact us

Project background

FAQ

Cubby

Privacy Policy

GARLAND PUBLISHING
Taylor & Francis Group

MOLECULAR BIOLOGY OF THE CELL

NCBI

PUBMED NCBI HOMEPAGE MOLECULAR BIOLOGY OF THE CELL HOMEPAGE

Short Contents | Full Contents | Other books @ NCBI

Molecular Biology of the Cell, 3rd edn. → Part II. Molecular Genetics → **Chapter 7. Recombinant DNA Technology** → DNA Cloning 15

Navigation

7. Recombinant DNA Technology

[Introduction](#)

[The Fragmentation, Separation, and Sequencing of DNA Molecules](#)

[Nucleic Acid Hybridization](#)

➔ [DNA Cloning](#)

[DNA Engineering](#)

[References](#)

Search

Go

This book All books

Entrez Specialized Databases

Taxonomy

Searchable taxonomic tree having nodes for all species with records in an Entrez database

OMIM

Online Mendelian Inheritance in Man: A database of genetically linked human diseases

ProbeSet

Expression data (GEO) and microarray datasets

Entrez Taxonomy

NCBI Taxonomy Browser

PubMed Entrez BLAST OMIM

Search for As complete name lock

Display levels nucleotides proteins structures genome records

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

- [Primates](#) [[6267223](#) [184321](#) [3684](#) [14](#)] Click on name to get more information.
 - [Catarrhini](#) [[6262446](#) [180921](#) [3684](#) [10](#)]
 - [Cercopithecoidea](#) (Old World monkeys) [[8304](#) [5400](#) [4](#) [3](#)]
 - [Cercopithecoidea](#) [[7981](#) [5045](#) [3](#) [3](#)]
 - [Allenopithecus](#) [[9](#) [5](#)]
 - [Cercocebus](#) [[81](#) [96](#)]
 - [Cercopithecus](#) [[492](#) [387](#)]
 - [Erythrocebus](#) [[40](#) [40](#)]
 - [Lophocebus](#) [[17](#) [6](#)]
 - [Macaca](#) (macaques) [[6291](#) [3897](#) [1](#) [1](#)]
 - [Mandrillus](#) (forest baboons) [[41](#) [42](#)]
 - [Miopithecus](#) [[30](#) [15](#)]
 - [Papio](#) (baboons) [[956](#) [543](#) [2](#) [2](#)]
 - [Theropithecus](#) [[24](#) [30](#)]
 - [Colobinae](#) [[312](#) [351](#)]
 - [Colobus](#) (black-and-white colobus monkeys) [[63](#) [58](#)]
 - [Nasalis](#) [[23](#) [12](#)]
 - [Presbytis](#) [[60](#) [64](#)]
 - [Procolobus](#) [[2](#) [2](#)]
 - [Pygathrix](#) [[69](#) [100](#)]
 - [Semnopithecus](#) [[18](#) [16](#)]
 - [Trachypithecus](#) (leaf monkeys) [[77](#) [100](#)]
 - [Cercopithecoidea gen. sp.](#) [[11](#) [5](#) [1](#)]
 - [Hominoidea](#) [[6253631](#) [175150](#) [3680](#) [6](#)]
 - [Homo/Pan/Gorilla group](#) [[6252769](#) [174616](#) [3680](#) [4](#)]
 - [Gorilla](#) [[878](#) [512](#) [1](#)]
 - [Homo](#) [[6090425](#) [172807](#) [3680](#) [1](#)]
 - [Pan](#) (chimpanzees) [[161466](#) [1310](#) [2](#)]

Entrez OMIM



OMIM

Online Mendelian Inheritance in Man



Johns
Hopkins
University

PubMed

Nucleotide

Protein

Genome

Structure

PMC

Taxonomy

Search OMIM for

Limits

Preview/Index

History

Clipboard

Details

Display Detailed Show: 20 Send to Text

***261600**

PHENYLKETONURIA

Alternative titles; symbols

PKU

PHENYLALANINE HYDROXYLASE DEFICIENCY

PAH DEFICIENCY

OLIGOPHRENIA PHENYLPYRUVICA

FOLLING DISEASE

PHENYLALANINE HYDROXYLASE, INCLUDED; PAH, INCLUDED

PKU1, INCLUDED

HYPERPHENYLALANINEMIA, MILD, INCLUDED; HPA, INCLUDED

PHENYLALANINEMIA, INCLUDED

Gene map locus [12q24.1](#)

TEXT

DESCRIPTION

MIM *261600

Description

Clinical Features

Biochemical Features

Inheritance

Mapping

Molecular Genetics

Pathogenesis

Diagnosis

Clinical Management

Population Genetics

Evolution

Animal Model

History

Allelic Variants

• View List

See Also

References

Contributors

Creation Date

Edit History

• MINI-MIM

• Clinical Synopsis

• Gene map

• Pathway

LocusLink

N Nomenclature

R RefSeq

G GenBank

P Protein

Entrez ProbeSet



□1: [GSM1130](#): ZR75-1 tamoxifen

SAGE NlaIII

Organism: Homo sapiens

Source: ZR75-1 cell line, tamoxifen treated

Sample synopsis: ZR75-1 estrogen receptor positive, estrogen dependent human breast cancer cell line, estrogen deprived for a week, cells were collected 16 hours following addition of fresh medium containing 10 uM 4-OH-tamoxifen

Published: 2002/02/22

□2: [GSM1129](#): ZR75-1 estrogen treated

SAGE NlaIII

Organism: Homo sapiens

Source: ZR75-1 human breast cancer cell line

Sample synopsis: ZR75-1 estrogen receptor positive, estrogen dependent human breast cancer cell line, estrogen deprived for a week, cells were collected 16 hours following addition of fresh medium containing 10 nM estrogen

Published: 2002/02/22

□3: [GSM1128](#): ZR75-1 untreated

SAGE NlaIII


Organism: Homo sapiens

Source: ZR75-1 human breast cancer cell line

Sample synopsis: ZR75-1 human estrogen receptor positive, estrogen dependent breast cancer cell line, estrogen deprived for 1 week, cells collected 16 hours following change to fresh medium (no estrogen added)

Published: 2002/02/22

Entrez Structure

NCBI  Entrez Structure

PubMed Nucleotide Protein Genome Structure PopSet

Search for

Go Clear

Limits Preview/Index History Clipboard

NCBI Structure

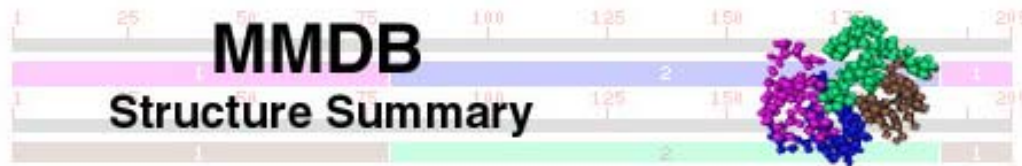
PubMed BLAST OMIM Taxonomy

Search Entrez for

[Structure summary](#)
via
PDB/MMDB-Id: Go

Go

Structure Summary



[PubMed](#) [BLAST](#) [Structure](#) [Taxonomy](#) [OMIM](#) [Help?](#) [Cn3d](#)

Description: Chloroquine Binds In The Cofactor Binding Site Of Plasmodium Falciparum Lactate Dehydrogenase..

Deposition: J.A.Read, K.W.Wilkinson, R.Tranter, R.B.Sessions & R.L.Brady, 10-Mar-99

Taxonomy: [Plasmodium falciparum](#)

Reference: [PubMed](#) **MMDB:** [9873](#) **PDB:** [1CET](#)

Cn3D viewer

[View 3D Structure](#)

Related Structures

Display

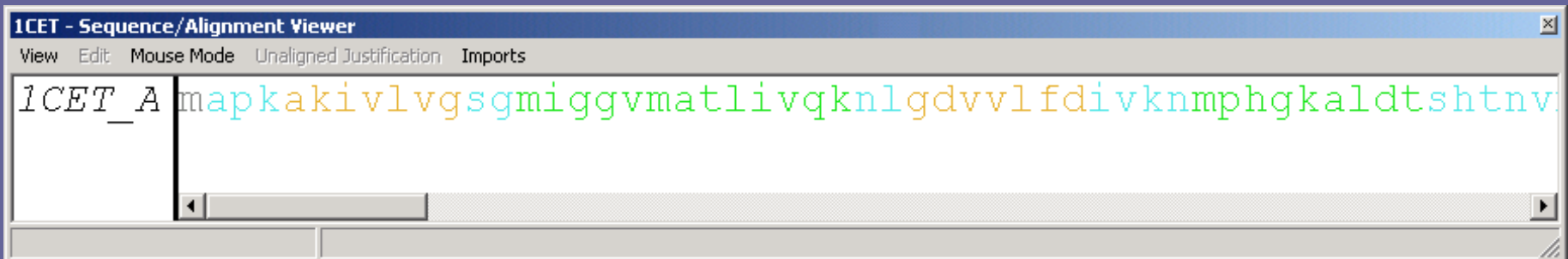
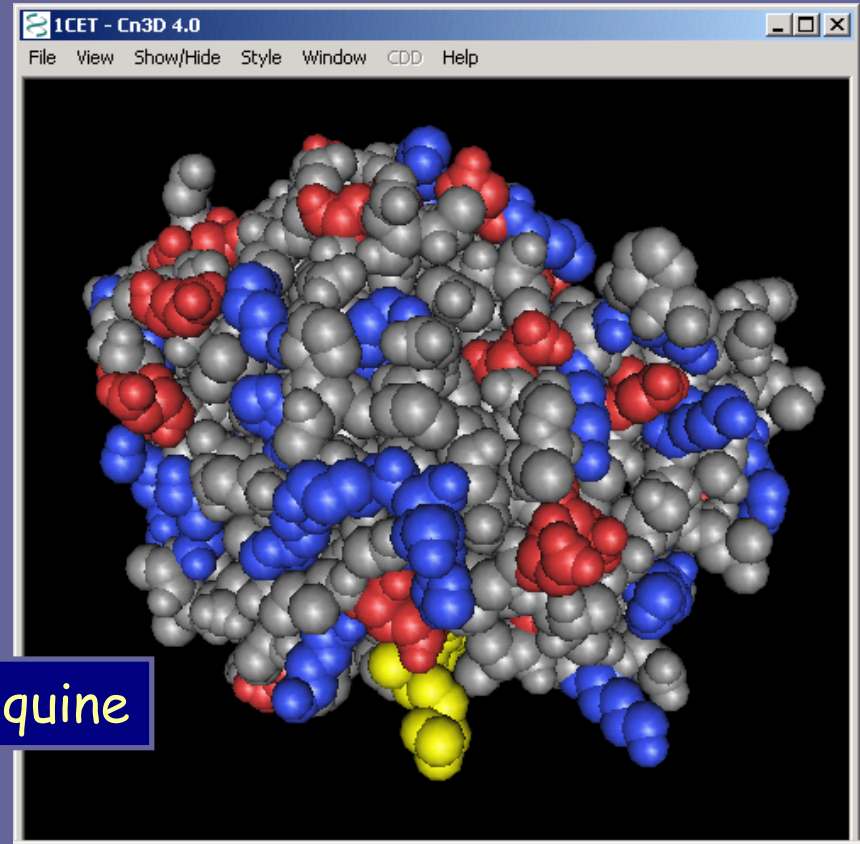
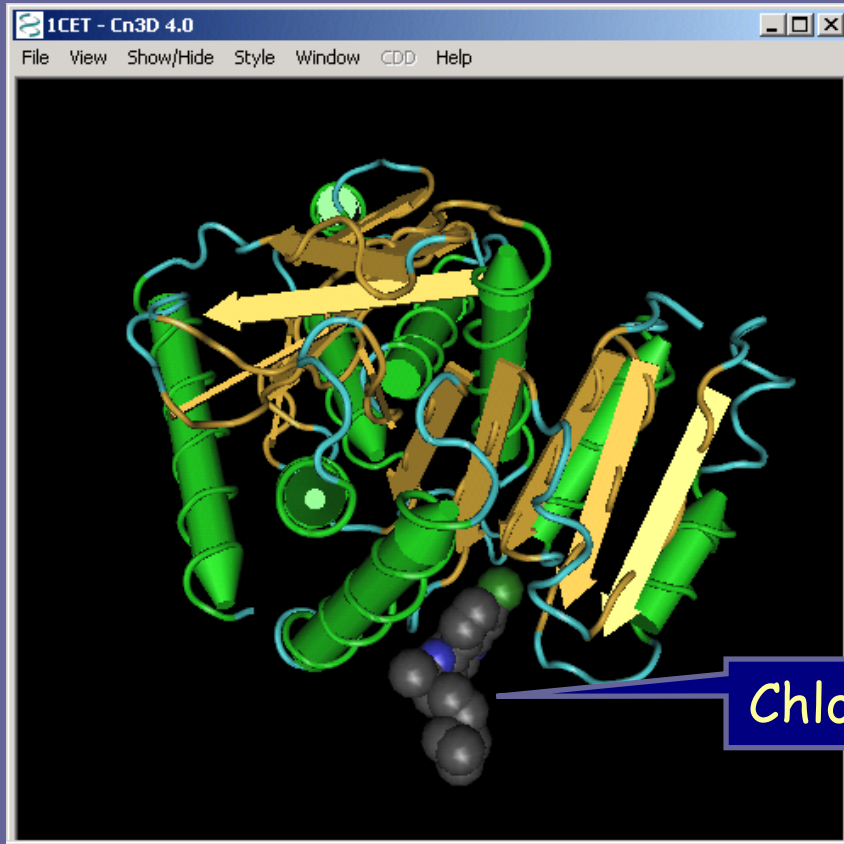
NEW

[Get Cn3D 4.0!](#)



Conserved Domains

Cn3D: Displaying Structures

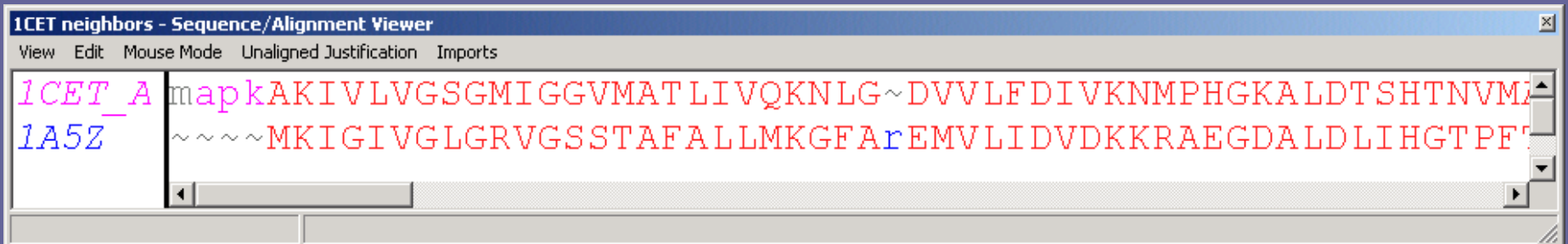
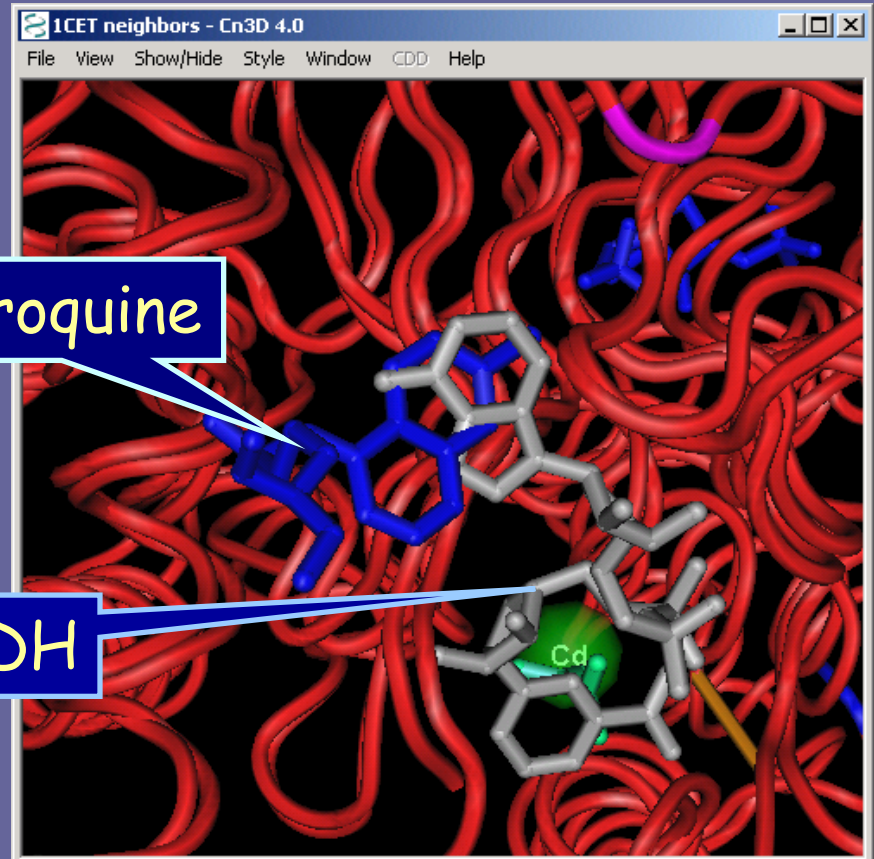
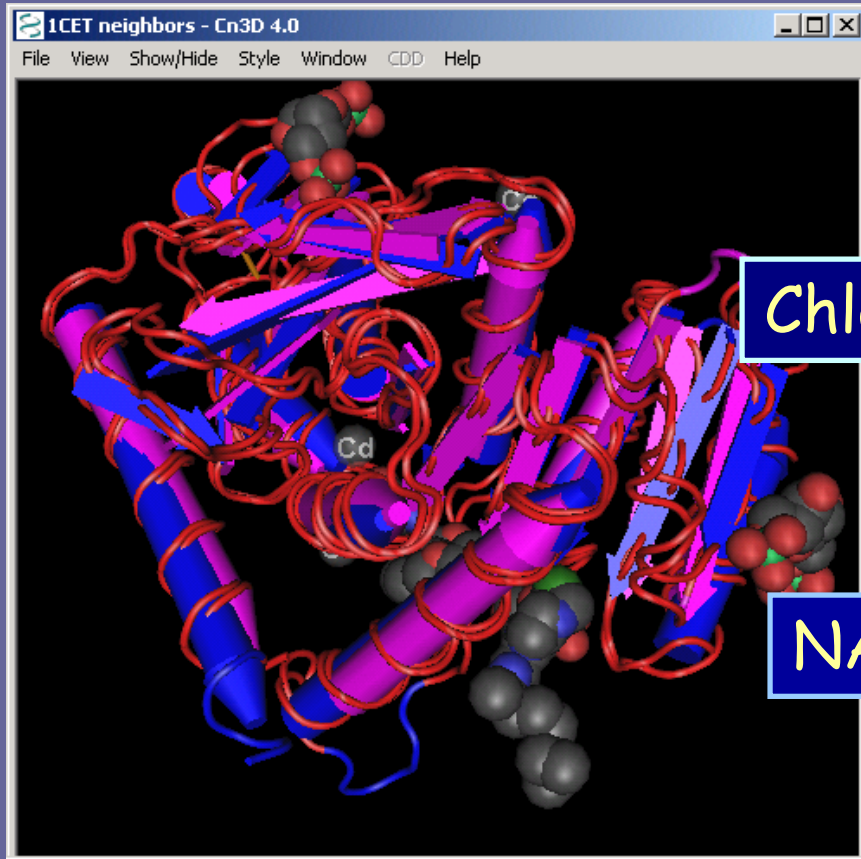


Structure Neighbors

Structure neighbors 1-20 out of 208 displayed. Page 1 of 11.

	<u>PDB</u>	<u>C</u>	<u>D</u>	<u>RMSD</u>	<u>NRES</u>	<u>%Id</u>	<u>Description</u>
<input type="checkbox"/>	1LLD	A		1.8	294	28.9	L-Lactate Dehydrogenase (E.C.1.1.1.27) (T-State) Mutant With Cys 199 Replaced By Ser (C199s) Complex With Nadh
<input type="checkbox"/>	1LDG			0.6	300	100.0	Plasmodium Falciparum L-Lactate Dehydrogenase Complexed With Nadh And Oxamate
<input type="checkbox"/>	1CEQ	A		0.2	304	99.7	Chloroquine Binds In The Cofactor Binding Site Of Plasmodium Falciparum Lactate Dehydrogenase.
<input type="checkbox"/>	6LDH			1.7	298	27.8	M4 Apo-Lactate Dehydrogenase (E.C.1.1.1.27)
<input checked="" type="checkbox"/>	1A5Z			1.6	300	36.3	Lactate Dehydrogenase From Thermotoga Maritima (Tmldh)
<input type="checkbox"/>	1LDN	A		1.6	300	28.7	L-Lactate Dehydrogenase (E.C.1.1.1.27) Complexed With Nadh, Oxamate, And Fructose-1,6-Bisphosphate

Structural Alignment



MMDB: Molecular Modeling Data Base

- Derived from experimentally determined PDB records
- Value added to PDB records including:
 - Addition of explicit chemical graph information
 - Validation
 - Inclusion of Taxonomy, Citation, and other information
 - Conversion to ASN.1 data description language
- Structure neighbors determined by Vector Alignment Search Tool (VAST)