NCBI web resources I: databases and Entrez

Yanbin Yin

Fall 2015

Most materials are downloaded from ftp://ftp.ncbi.nih.gov/pub/education/

Homework assignment 1

- Two parts:
- Extract the gene IDs reported in table 1 of http://www.ncbi.nlm.nih.gov/pmc/articles/PMC523881/
  Using NCBI batch Entrez to download all refseq protein fasta sequences from Arabidopsis thaliana

- Search "oncogenes" using Entrez and report in human how many oncogenes correspond to how many total proteins and how many refseq proteins

- Write a report to explain all the operations and include screen shots

Due on 9/15 (send by email)

References

- NCBI mcbios workshop
  - ftp://ftp.ncbi.nih.gov/pub/education/mcbios2012/
- NCBI web resource tutorials
  - ftp://ftp.ncbi.nih.gov/pub/education/tutorials/
- NCBI discovery workshops
  - ftp://ftp.ncbi.nih.gov/pub/education/discovery_workshops/NLM/2012/Sept2012/
- NCBI Help Manual
  - http://www.ncbi.nlm.nih.gov/books/NBK3831/

Youtube

- http://www.youtube.com/ncbinlm

- Go to www.youtube.com
- Search “NCBI tutorial general”

Topics

- Intro. to NCBI
- Selected NCBI Databases
- The Entrez system
- Hands on practice

The National Center for Biotechnology Information

Created in 1988 as a part of the National Library of Medicine at NIH

- Establish public databases
- Research in computational biology
- Develop software tools for sequence analysis
- Disseminate biomedical information

GenBank history

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

Early 1990s, Congress awarded responsibility to NCBI

Initially, indexers scanned the literature
and typed in the sequences

Now sequences are deposited directly by labs

Direct submissions since 1993

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

Molecular Data

- Sequences
- Expression
- Genome Maps
- 3D Structures
- Protein Domains
- Homologous Genes, Proteins, Structures
- Pathways
- Genetic Variation

Selected NCBI Databases

- Biomedical literature
  - PubMed free Medline
  - PubMed Central full text online access
  - NCBI Bookshelf online biomedical textbooks
- Biomolecular Databases
  - Nucleotide
    - GenBank submitted sequence records
    - RefSeq curated NCBI reference sequences
  - Protein GenBank and RefSeq translations, outside protein
  - dbSNP small scale genetic variations
  - Structure biomolecular 3-D structures
    - MMDB NCBI's 3D structure database
  - GEO microarray expression data
  - SRA next-generation sequence data

Information Hubs: Aggregators

- Taxonomy access to NCBI data through source organism classification
- BioProjects molecular data and literature related to large scale molecular projects (genomes, transcriptomes, metagenomes)
- Genome specialized displays for complete genomes and access to microbial genome analysis tools
- Gene molecular data and literature related to genes
- BioSystems biochemical pathways and processes linked to NCBI genes, gene products, small molecules, and structures

Information Hubs: Analyses

- Analysis Results
  - HomoloGene homologous genes from selected eukaryotes
  - Protein Clusters homologs (proteins) from microbial genomes
  - UniGene sequence-based gene catalog (eukaryotes)
  - GEO Datasets microarray experiments and analyses

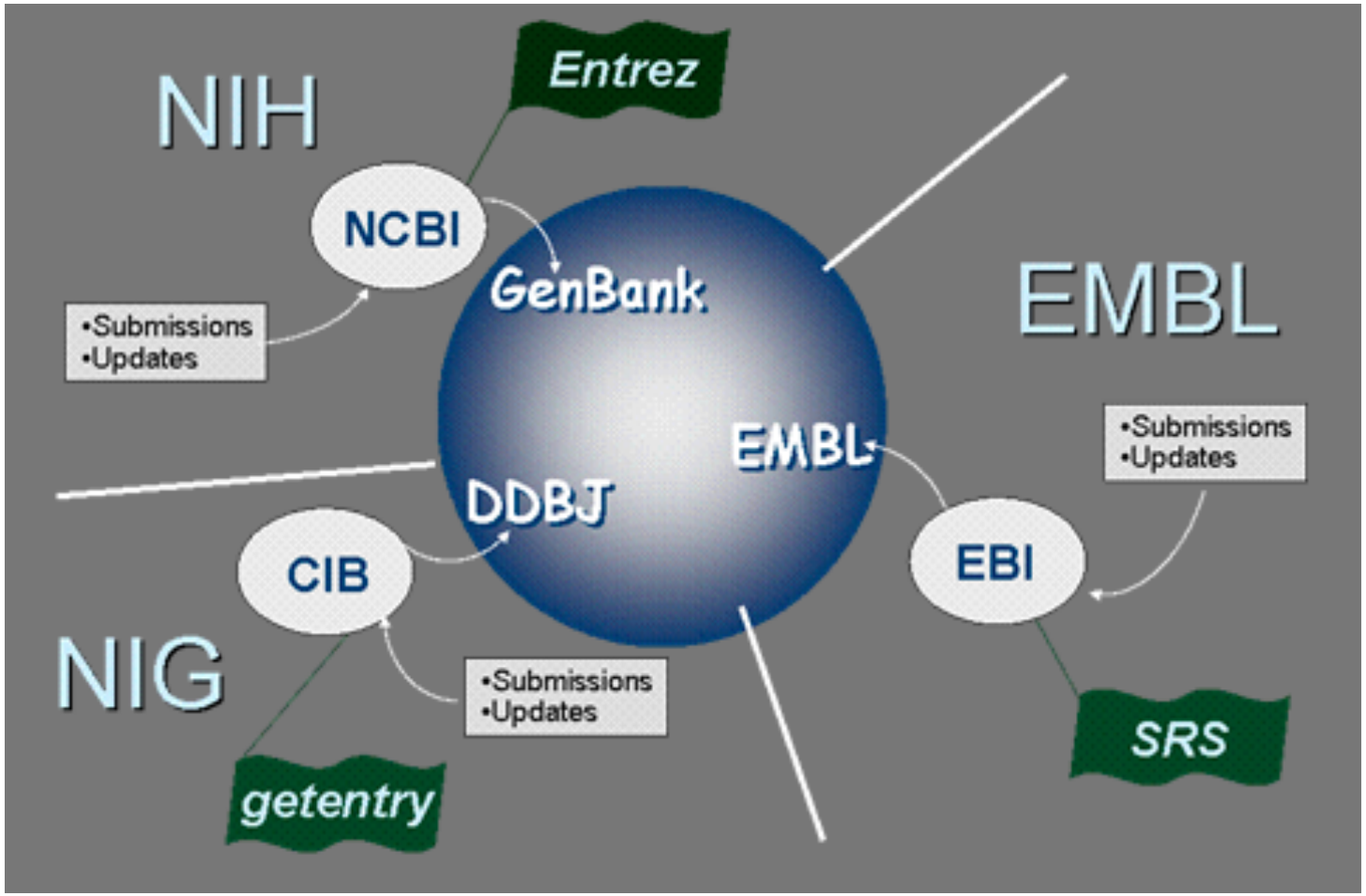
Sequence Databases at NCBI

- Primary
  - GenBank: NCBI's primary sequence database
  - Trace Archive: reads from capillary sequencers
  - Sequence Read Archive: next generation data
- Derivative
  - GenPept (GenBank translations)
  - Outside Protein (UniProt—Swiss-Prot, PDB)
  - NCBI Reference Sequences (RefSeq)

GenBank types of entries

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

Three international nucleotide sequence databases



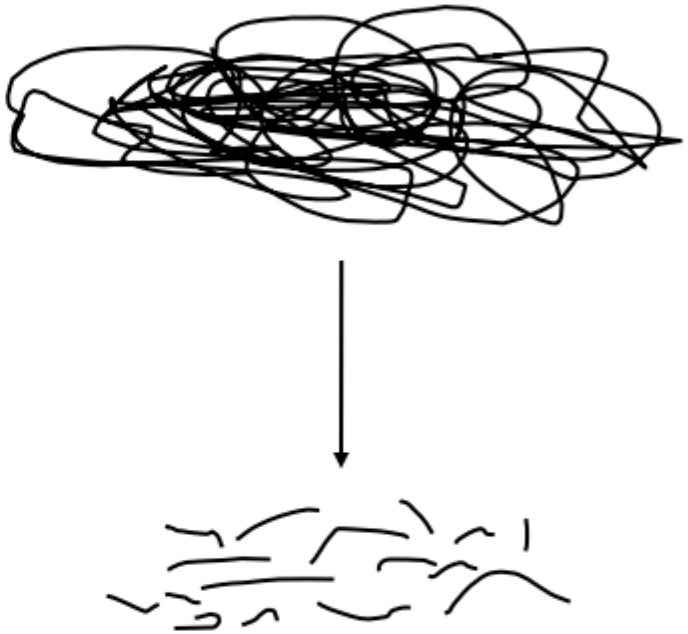
RefSeq: NCBI's Derivative Sequence Database

- **Experimentally verified / curated transcripts and proteins**
  - NM_, NP_ accession numbers
- **Model transcripts and proteins**
  - XM_, XP_ accession numbers
- **Assembled Genomic Regions (contigs)**
  - NT_, NW_ accession numbers
- **Chromosome records**
  - NC_, AC_ accession numbers
- **RefSeqGene Records**
  - NG_ accession numbers (NG_ also used pseudo genes and other fixed genomic sequences)
- **Draft whole genome shotgun assemblies (microbial)**
  - NZ_ accession numbers
- **Microbial proteins**
  - NP_, YP_, ZP_ accessions

ftp://ftp.ncbi.nih.gov/refseq/release/

Whole Genome Sequencing Approaches

Shotgun Approach



Genomic DNA

Shotgun Clones

http://www.bio.davidson.edu/genomics/method/shotgun.html

GCAATGAAATATGTTCTTGTAATTTAAGCTGACACTCCTAATTTAGCTCTTGTCCTCTACTGAGTCTACCTAATTATATGTATGGATTGACTTGG

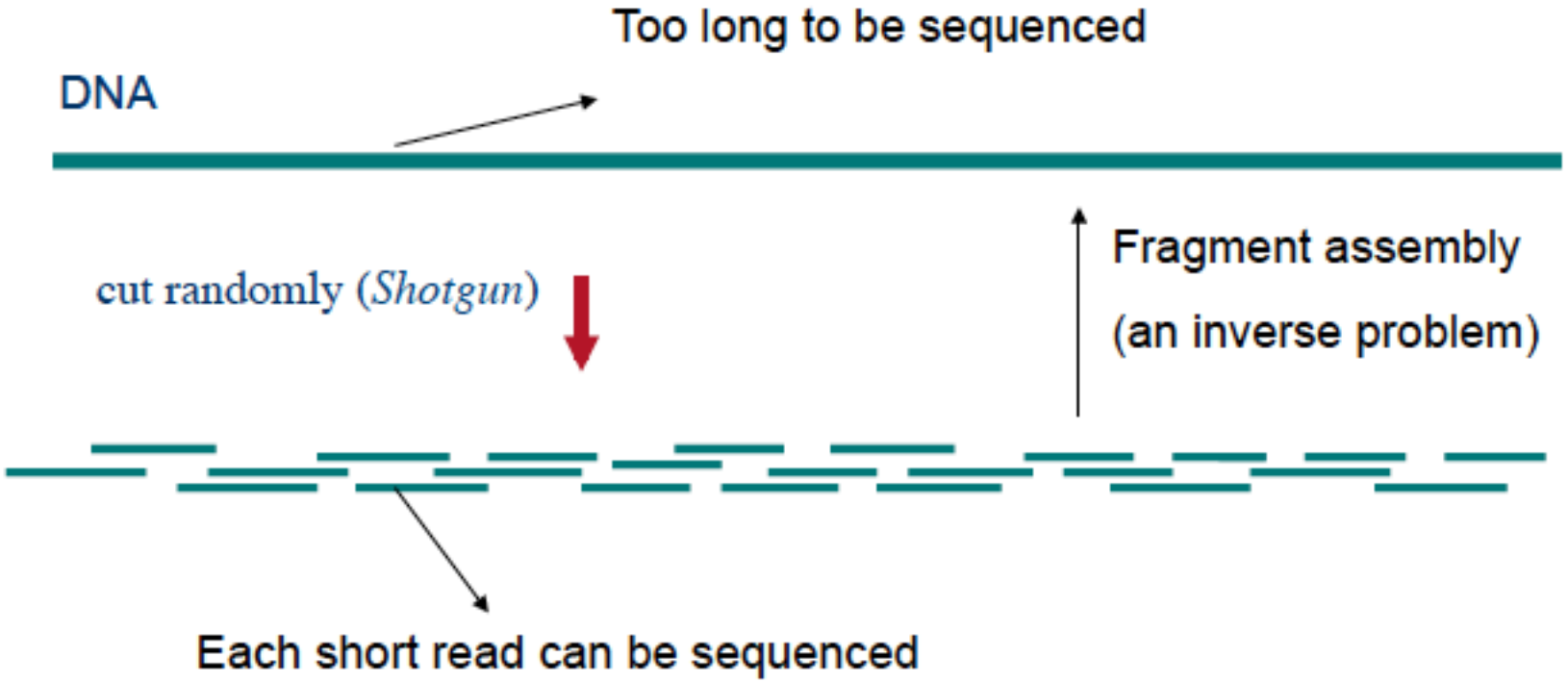
AGCTCTTGTCCTCTACTGAGTCTACCTAATTATATGTATGGATTGACTTGGTGTTTTCTCTTTTTCTTAAATAGTAATGCAGAAAGCCTGGAGAGAGAG

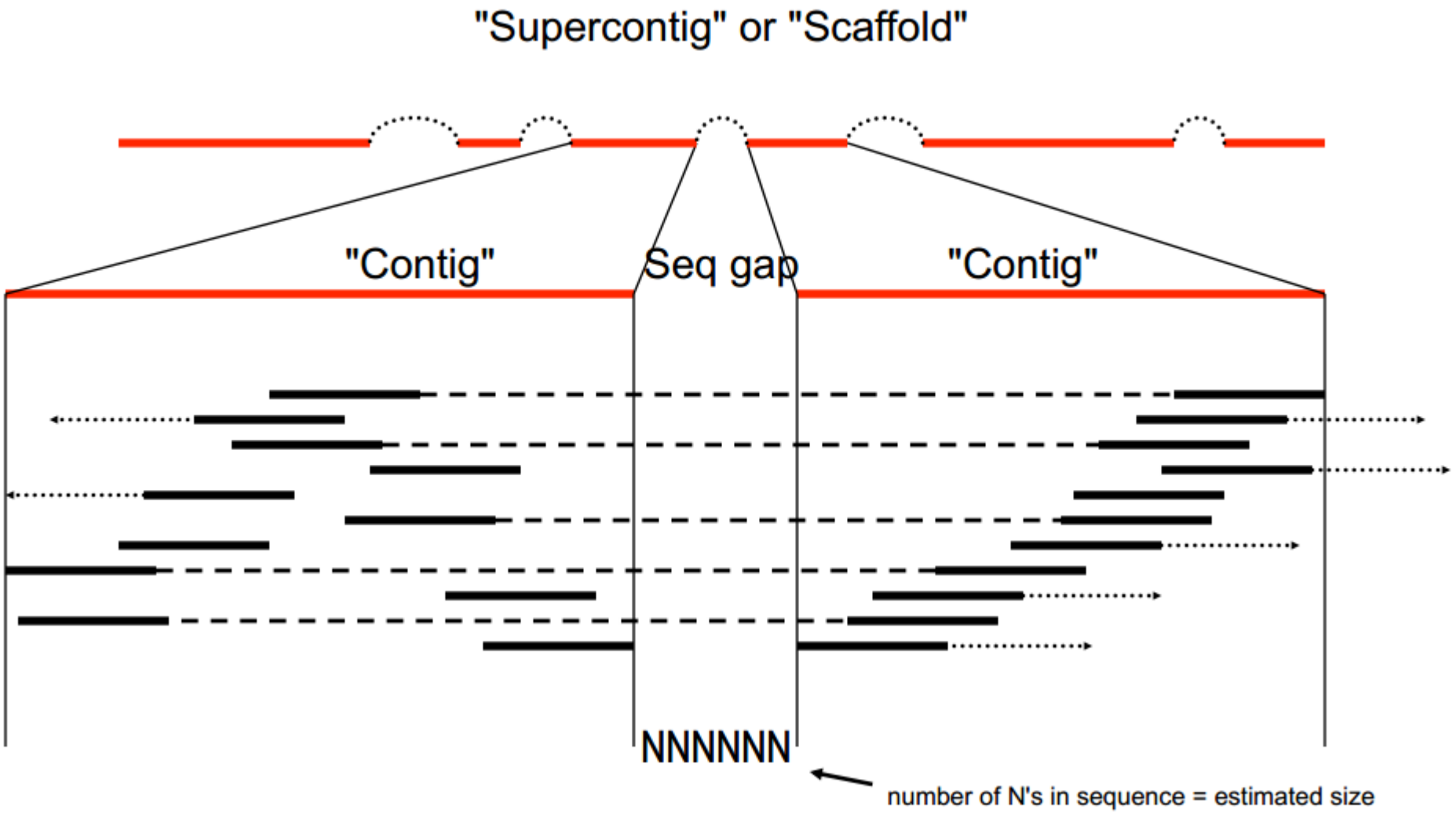
Reads

ATGTTCTTGTAATTTAAGCTGACACTCCTAATTTAGCTCTTGTCCTCTACTGAGTCTACCTAATTATATGTATGGATTGACTTGGTGTTTTCTCTTTTTCTTAAATAGTAATGCAGAAAGCCTGGAGAGAGAG

Assembly

Shotgun sequencing





Sequence assembly is the problem of merging and ordering shorter fragments, termed "reads," sampled from a set of larger sequences in order to reconstruct the larger sequences. The output of an assembly is typically a set of "contigs," which are contiguous sequence fragments, ordered and oriented into "scaffold" sequences, with gaps between contigs within scaffolds representing regions of uncertainty

Genome assemblies are composed of scaffolds and contigs.

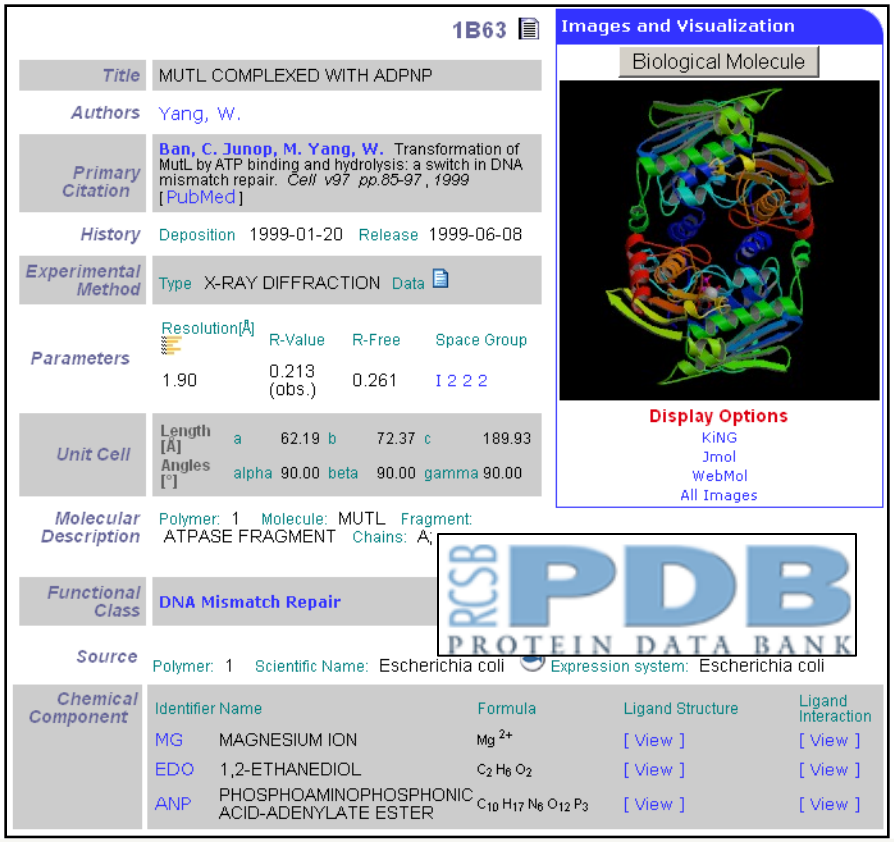
Contigs are contiguous consensus sequences that are derived from collections of overlapping reads (no gaps).

Scaffolds are ordered and orientated sets of contigs that are linked to one another by mate pairs of sequencing reads (have gaps).

GenBank & RefSeq

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

Protein Sequences from Structures



1B63

Title	MUTL COMPLEXED WITH ADPNP
Authors	Yang, W.
Primary Citation	Ban, C. Junop, M. Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell v97 pp.85-97 , 1999* [PubMed]
History	Deposition 1999-01-20 Release 1999-06-08
Experimental Method	Type X-RAY DIFFRACTION Data

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	1.90	0.213 (obs.)	0.261	I 2 2 2

Unit Cell						
Length [Å]	a	62.19	b	72.37	c	189.93
Angles [°]	alpha	90.00	beta	90.00	gamma	90.00

Molecular Description: Polymer: 1 Molecule: MUTL Fragment: ATPASE FRAGMENT Chains: A;

Functional Class: DNA Mismatch Repair

Source: Polymer: 1 Scientific Name: Escherichia coli Expression system: Escherichia coli

Chemical Component	Identifier	Name	Formula	Ligand Structure	Ligand Interaction
	MG	MAGNESIUM ION	Mg^{2+}	[View]	[View]
	EDO	1,2-ETHANEDIOL	$C_2 H_6 O_2$	[View]	[View]
	ANP	PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER	$C_{10} H_{17} N_6 O_{12} P_3$	[View]	[View]

Images and Visualization

Biological Molecule

Display Options

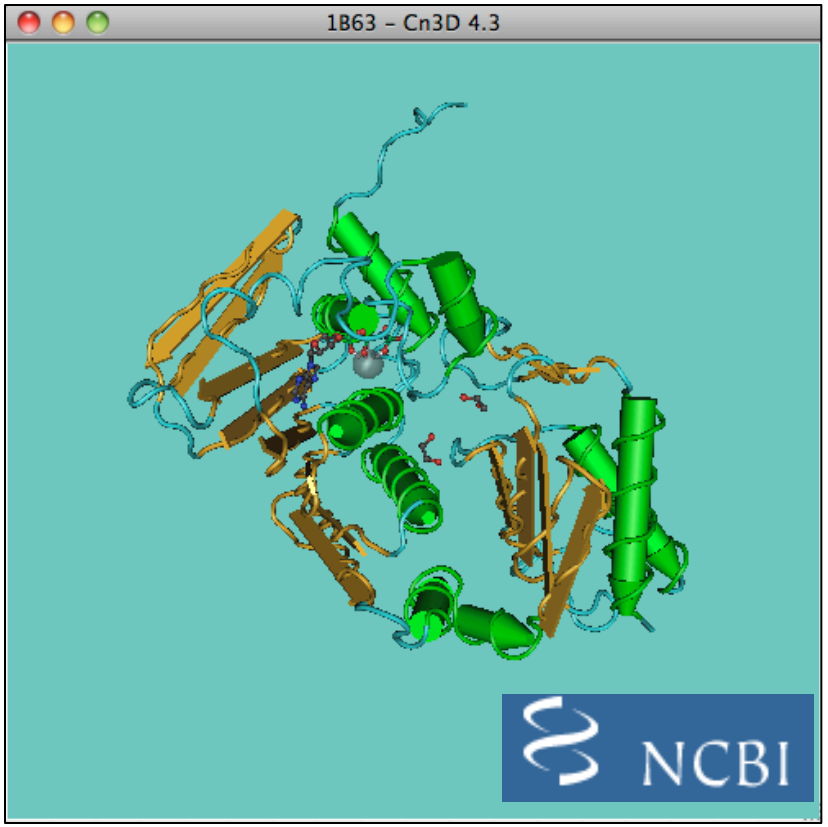
KiNG

Jmol

WebMol

All Images

RCSB PDB PROTEIN DATA BANK




```
>gi|5542073|pdb|1B63|A Chain A, Mutl Complexed With Adpnp
SHMPIQVLPPQLANQIAAGEVVERPASVVKELVENSLDAGATRIDIDIERGGAKLIRIRDNGCGIKKDEL
ALALARHATSKIASLDDLEAIISLGFRGEALASISSVSRLTLTSRTAEQQEAWQAYAEGRDMNVTVKPAA
HPVGTTLEVLDLFYNTPARRKFLRTEKTEFNHIDEIIRRIALARFDVTINLSHNGKIVRQYRAVPEGGQK
ERRLGAICGTAFLEQALAIEWQHGDLTLRGWVADPNHTTPALAEIQYCYVNGRMMRDRLINHAIRQACED
KLGADQQPAFVLYLEIDPHQVDVNVHPAKHEVRFHQSRLVHDFIYQGVLSVLQ
```

MMDB: Molecular Modeling Data Base

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

Protein Domains

- Structural Domain
  - Discrete independently folding unit of a protein
- Conserved Domain (sequence-based)
  - Protein region with recognizable position-specific pattern of sequence conservation
- Sequence-based domains often roughly correspond to structural domains
- Domains often have distinct, identifiable functions

NCBI's Conserved Domain Database

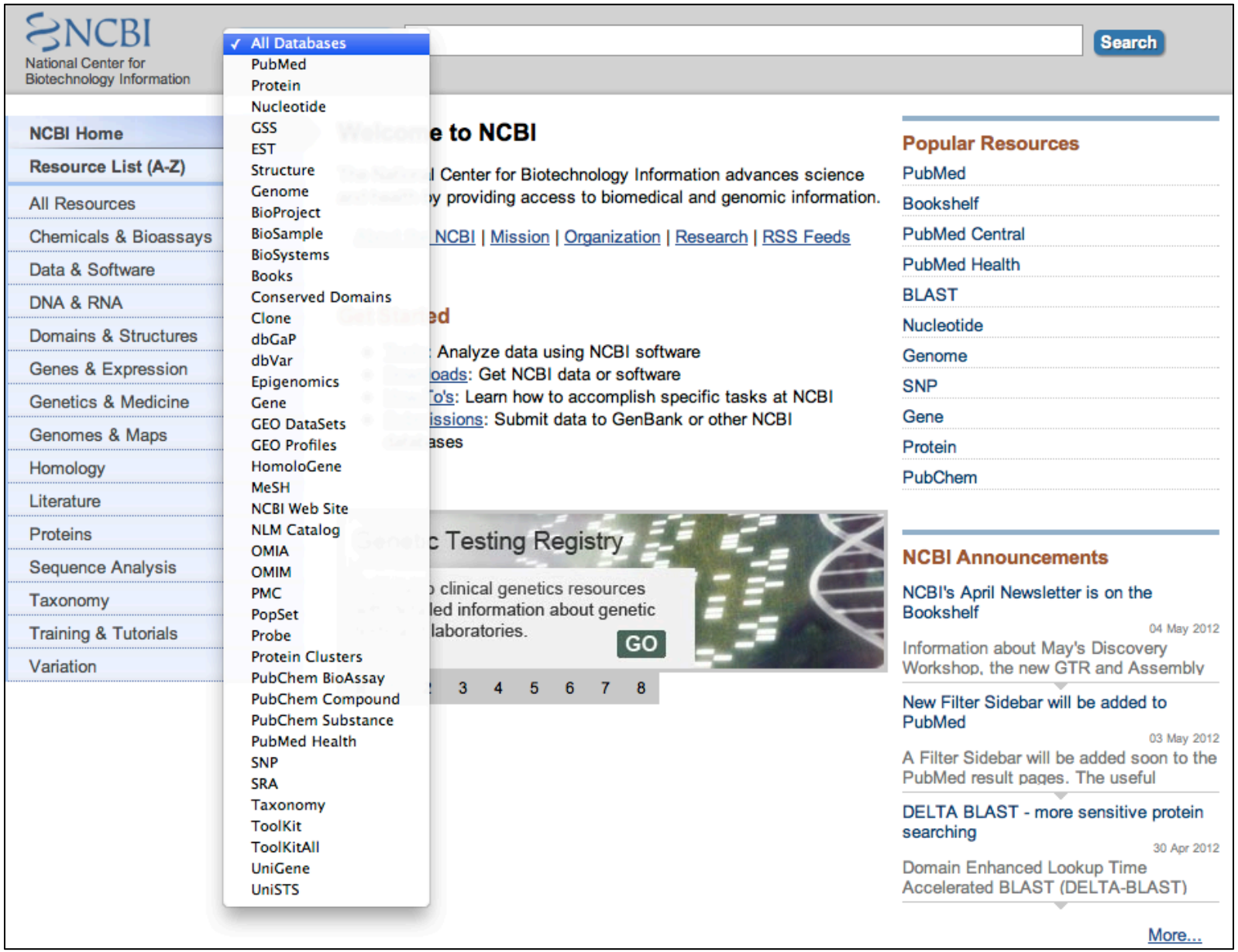
- Searchable with RPS-BLAST
- Sources
  - SMART
  - PFAM
  - COGs
  - NCBI curated domains
    - structure-informed alignments

NCBI Search Services and Tools

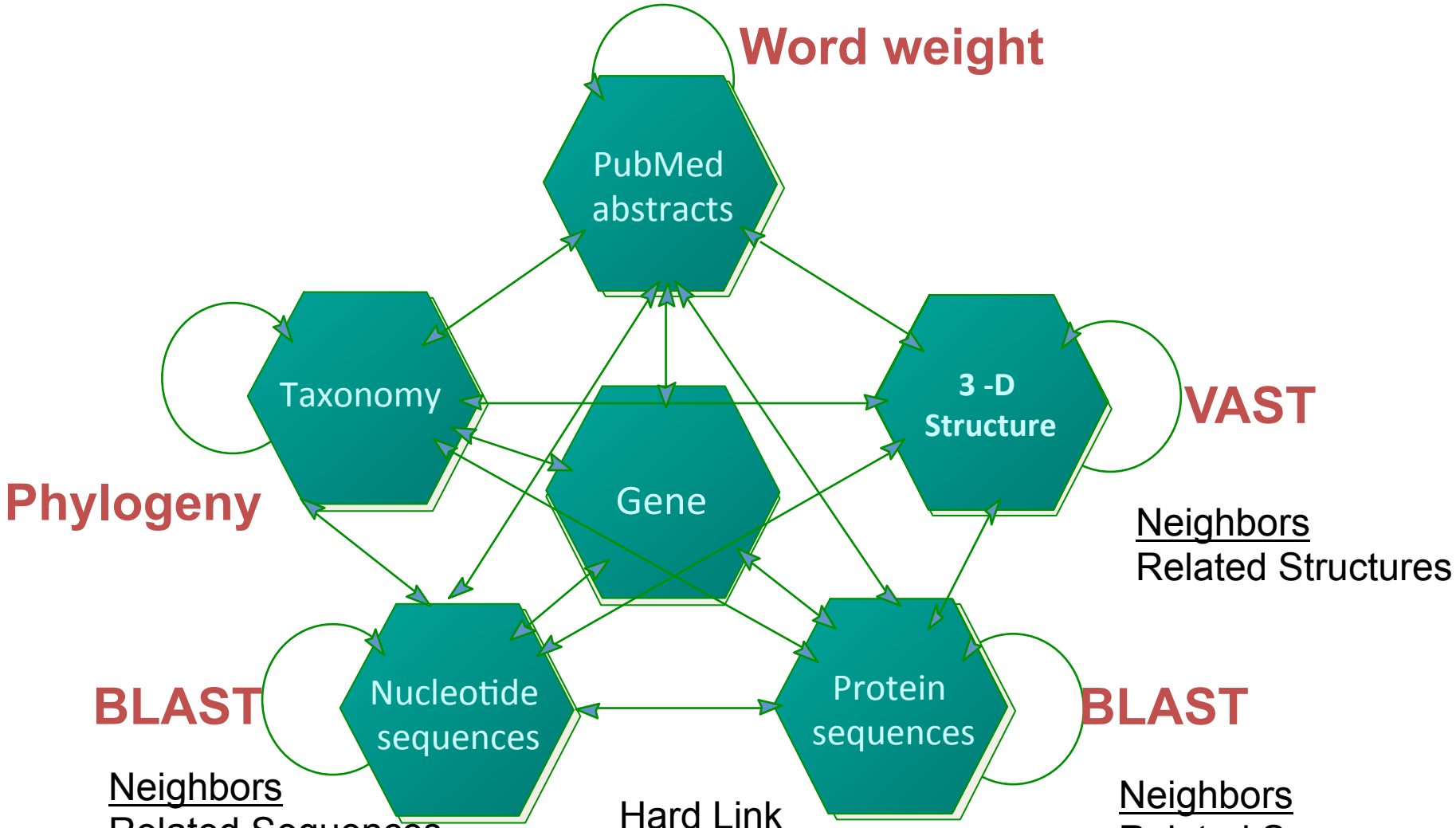
- Entrez integrated literature and molecular databases
  - BLink protein similarities
  - Graphical Sequence Viewer incipient genome browser
- BLAST highest volume sequence search service
- VAST structure similarity searches
- Map Viewer graphical genome map display (assembled eukaryotic genomes only)
- Cn3D 3D structure viewer
- Genome Workbench standalone sequence analysis annotation platform

Entrez: Integrated Molecular and Sequence Databases

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

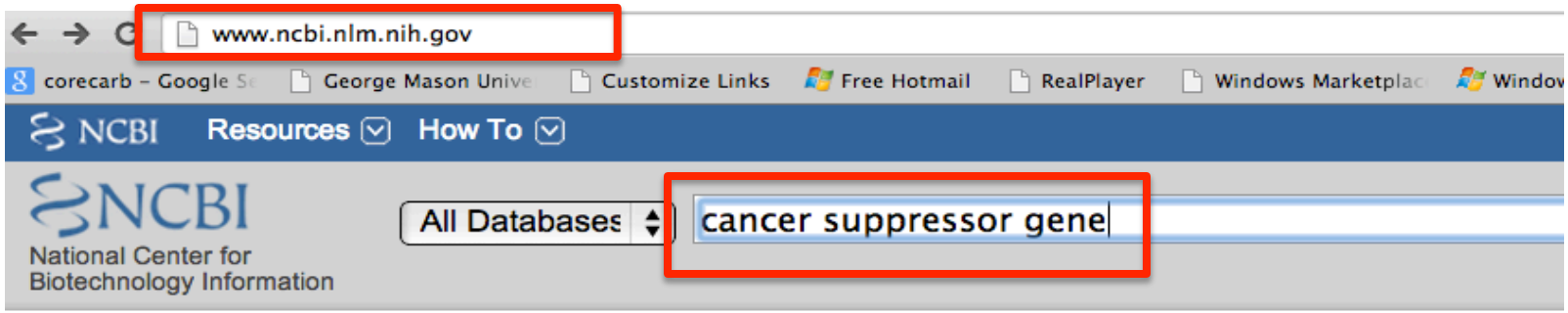


Entrez: A Discovery System



Hands-on exercise 1

Cancer related genes



NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

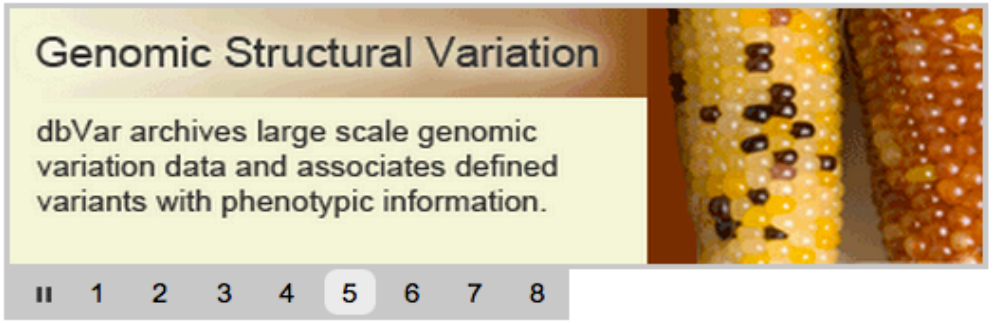
Welcome to NCBI

The National Center for Biotechnology Information advances science and h
biomedical and genomic information.

About the NCBI | Mission | Organization | Research | RSS Feeds

Get Started

- Tools: Analyze data using NCBI software
- Downloads: Get NCBI data or software
- How-To's: Learn how to accomplish specific tasks at NCBI
- Submissions: Submit data to GenBank or other NCBI databases



Search NCBI databases

cancer suppressor gene

Search

Results found in 22 databases for "cancer suppressor gene"

Literature

Books	1,400	books and reports
MeSH	1	ontology used for PubMed indexing
NLM Catalog	155	books, journals and more in the NLM Collections
PubMed	78,717	scientific & medical abstracts/citations
PubMed Central	126,287	full-text journal articles

Health

ClinVar	310	human variations of clinical significance
dbGaP	151	genotype/phenotype interaction studies
GTR	0	genetic testing registry
MedGen	0	medical genetics literature and links
OMIM	644	online mendelian inheritance in man
PubMed Health	62	clinical effectiveness, disease and drug reports

Genomes

Assembly	0	genome assembly information
BioProject	479	biological projects providing data to NCBI
BioSample	0	descriptions of biological source materials

Genes

EST	2	expressed sequence tag sequences
Gene	14	collected information about gene loci
GEO DataSets	6,697	functional genomics studies
GEO Profiles	0	gene expression and molecular abundance profiles
HomoloGene	0	homologous gene sets for selected organisms
PopSet	0	sequence sets from phylogenetic and population studies
UniGene	8	clusters of expressed transcripts

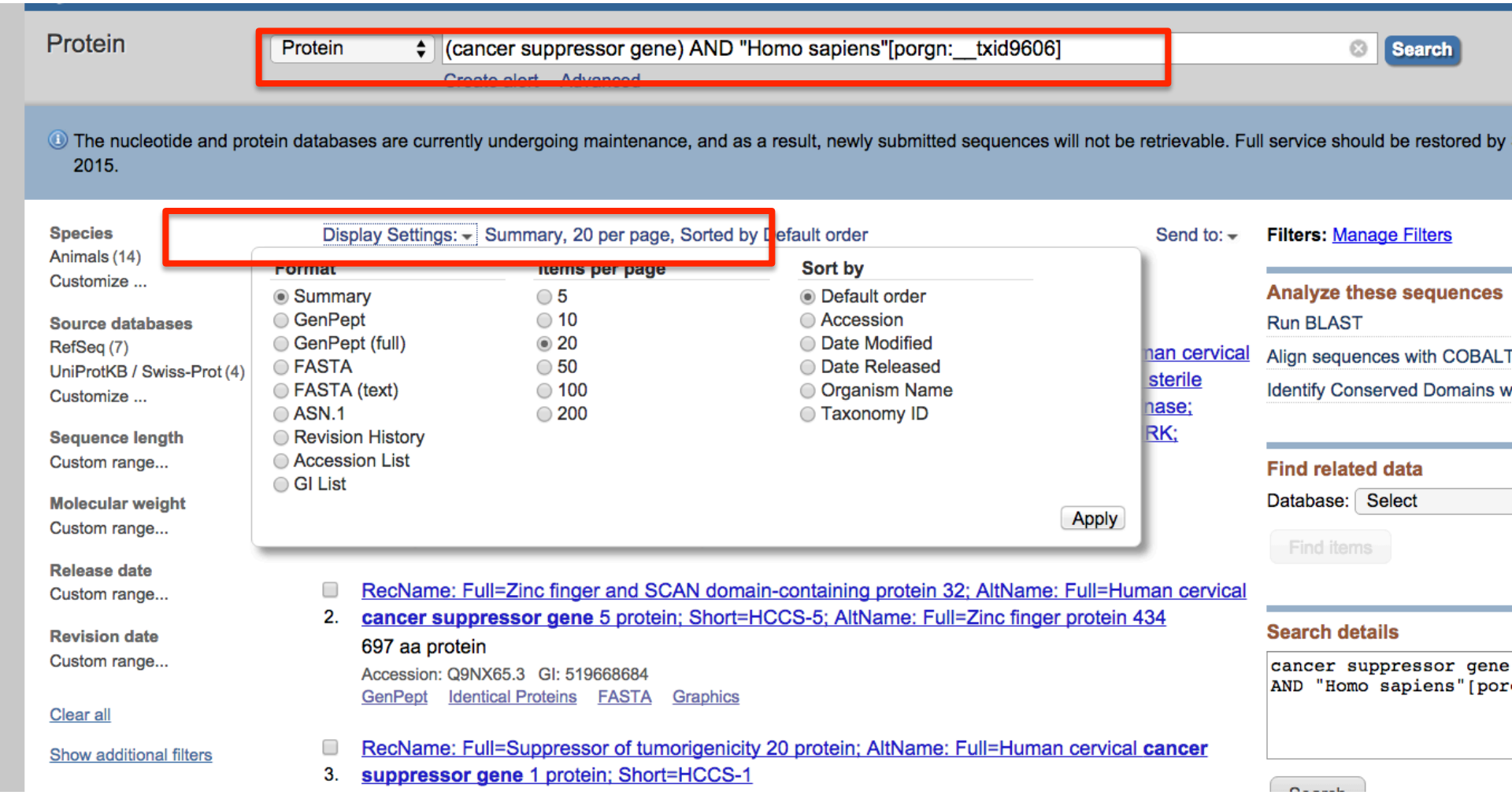
Proteins

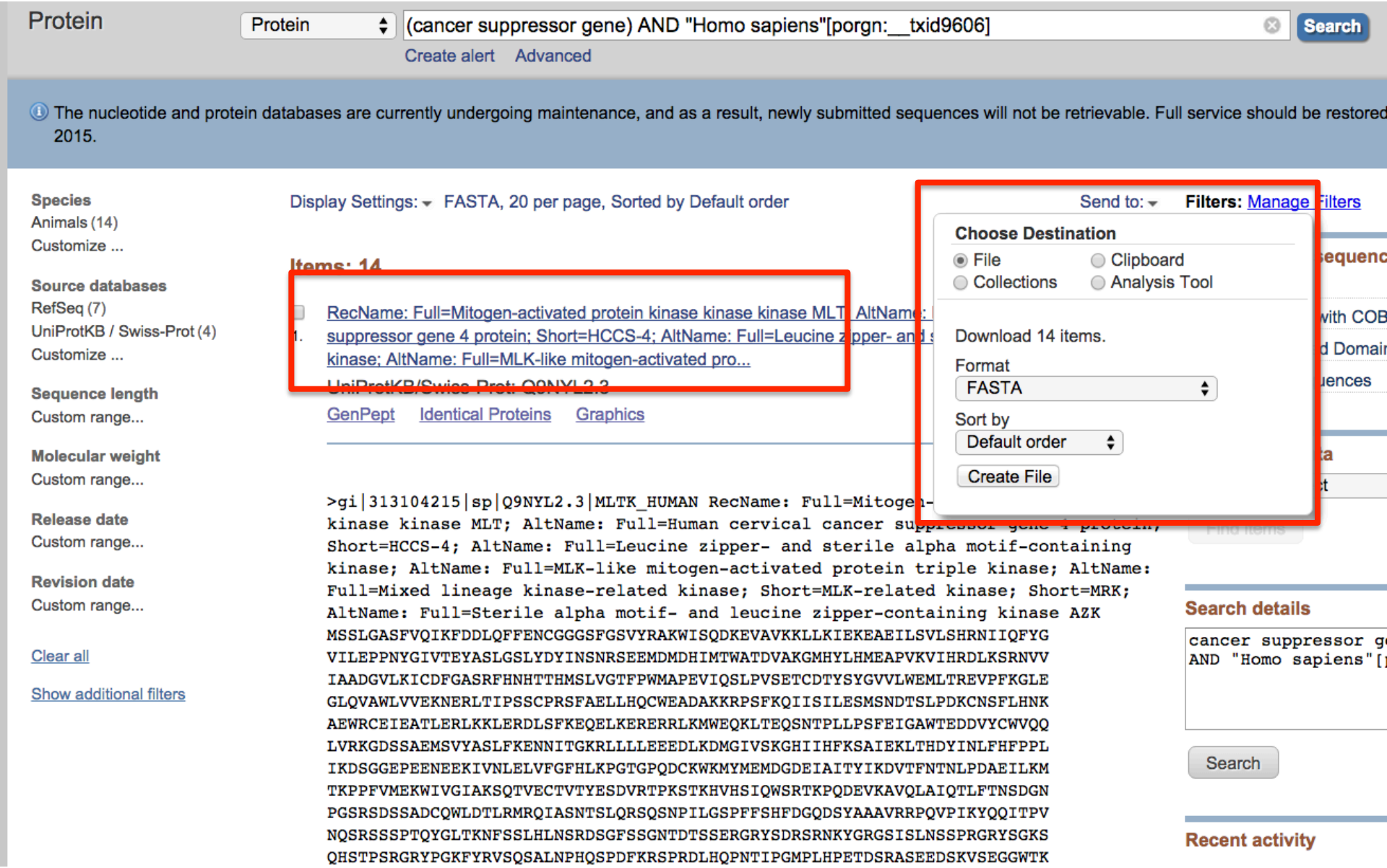
Conserved Domains	0	conserved protein domains
Protein	20	protein sequences
Protein Clusters	0	sequence similarity-based protein clusters
Structure	582	experimentally-determined biomolecular structures

Chemicals

BioSystems	273	molecular pathways with links to genes, proteins and chemicals


Species
Animals (20)
Customize ...
Source databases
RefSeq (7)
UniProtKB / Swiss-Prot (4)
Customize ...
Sequence length
Custom range...
Molecular weight
Custom range...
Release date
Custom range...
Revision date
Custom range...
Clear all
Show additional filters
Display Settings: Summary, 20 per page, Sorted by Default order
Send to:
Items: 20
1. RecName: Full=Mitogen-activated protein kinase kinase kinase MLT; AltName: Full=Human cervical cancer suppressor gene 4 protein; Short=HCCS-4; AltName: Full=Leucine zipper- and sterile alpha motif-containing kinase; AltName: Full=MLK-like mitogen-activated protein triple kinase; AltName: Full=Mixed lineage kinase-related kinase; Short=MLK-related kinase; Short=MRK; AltName: Full=Sterile alpha motif- and leucine zipper-containing kinase AZK
800 aa protein
Accession: Q9NYL2.3 GI: 313104215
GenPept Identical Proteins FASTA Graphics
2. RecName: Full=Zinc finger and SCAN domain-containing protein 32; AltName: Full=Human cervical cancer suppressor gene 5 protein; Short=HCCS-5; AltName: Full=Zinc finger protein 434
697 aa protein
Accession: Q9NX65.3 GI: 519668684
GenPept Identical Proteins FASTA Graphics
3. RecName: Full=Suppressor of tumorigenicity 20 protein; AltName: Full=Human cervical cancer suppressor gene 1 protein; Short=HCCS-1
79 aa protein
Accession: Q9HBF5.2 GI: 294862468
GenPept Identical Proteins FASTA Graphics
Filters: Manage Filters
Results by taxon
Top Organisms [Tree]
Homo sapiens (14)
Macaca mulatta (1)
Macaca fascicularis (1)
Phodopus roborovskii (1)
Phodopus sungorus (1)
All other taxa (2)
More...
Analyze these sequences
Run BLAST
Align sequences with COBALT
Identify Conserved Domains with CD-Search
Find related data
Database: Select
Find items


Protein
Protein
(cancer suppressor gene) AND "Homo sapiens"[porgn:__txid9606]
Search
Create alert
Advanced
The nucleotide and protein databases are currently undergoing maintenance, and as a result, newly submitted sequences will not be retrievable. Full service should be restored by
2015.
Species
Animals (14)
Customize ...
Source databases
RefSeq (7)
UniProtKB / Swiss-Prot (4)
Customize ...
Sequence length
Custom range...
Molecular weight
Custom range...
Release date
Custom range...
Revision date
Custom range...
Clear all
Show additional filters
Display Settings: Summary, 20 per page, Sorted by Default order
Send to:
Filters: Manage Filters
Format
Summary
GenPept
GenPept (full)
FASTA
FASTA (text)
ASN.1
Revision History
Accession List
GI List
Items per page
5
10
20
50
100
200
Sort by
Default order
Accession
Date Modified
Date Released
Organism Name
Taxonomy ID
Apply
RecName: Full=Zinc finger and SCAN domain-containing protein 32; AltName: Full=Human cervical
2. cancer suppressor gene 5 protein; Short=HCCS-5; AltName: Full=Zinc finger protein 434
697 aa protein
Accession: Q9NX65.3 GI: 519668684
GenPept Identical Proteins FASTA Graphics
RecName: Full=Suppressor of tumorigenicity 20 protein; AltName: Full=Human cervical cancer
3. suppressor gene 1 protein; Short=HCCS-1
Analyze these sequences
Run BLAST
Align sequences with COBALT
Identify Conserved Domains w
Find related data
Database: Select
Find items
Search details
cancer suppressor gene
AND "Homo sapiens"[por


Protein
Protein
(cancer suppressor gene) AND "Homo sapiens"[porgn:__txid9606]
Search
Create alert
Advanced
The nucleotide and protein databases are currently undergoing maintenance, and as a result, newly submitted sequences will not be retrievable. Full service should be restored
2015.
Species
Animals (14)
Customize ...
Source databases
RefSeq (7)
UniProtKB / Swiss-Prot (4)
Customize ...
Sequence length
Custom range...
Molecular weight
Custom range...
Release date
Custom range...
Revision date
Custom range...
Clear all
Show additional filters
Display Settings: FASTA, 20 per page, Sorted by Default order
Send to:
Filters: Manage Filters
Choose Destination
File
Clipboard
Collections
Analysis Tool
Download 14 items.
Format
FASTA
Sort by
Default order
Create File
Items: 14
RecName: Full=Mitogen-activated protein kinase kinase kinase MLT AltName:
1. suppressor gene 4 protein; Short=HCCS-4; AltName: Full=Leucine zipper- and
kinase; AltName: Full=MLK-like mitogen-activated pro...
GenPept
Identical Proteins
Graphics
>gi|313104215|sp|Q9NYL2.3|MLTK_HUMAN RecName: Full=Mitogen-
kinase kinase MLT; AltName: Full=Human cervical cancer
Short=HCCS-4; AltName: Full=Leucine zipper- and sterile alpha motif-containing
kinase; AltName: Full=MLK-like mitogen-activated protein triple kinase; AltName:
Full=Mixed lineage kinase-related kinase; Short=MLK-related kinase; Short=MRK;
AltName: Full=Sterile alpha motif- and leucine zipper-containing kinase AZK
MSSLGASFVQIKFDDLQFFENCGGGSFGSVYRAKWISQDKEVAVKKLLKIEKEAEILSVLSHRNIIQFYG
VILEPPNYGIVTEYASLGSLYDYINSNRSEEMDMDHIMTWATDVAKGMHYLHMEAPVKVIHRDLKSRNVV
IAADGVLKICDFGASRFHNHTTHMSLVGTFPWMAPEVIQSLPVSETCDTYSYGVVLWEMLTREVPFKGLE
GLQVAWLVVEKNERLTIPSSCPRSFAELLHQCWEADAKKRPSFKQIISILESMSNDTSLPDKCNSFLHNK
AEWRCEIEATLERLKKLERDLSFKEQELKERERRLKMWEQKLTEQSNTPLLPSFEIGAWTEDDVYCWVQQ
LVRKGDSSAEMSVYASLFKENNITGKRLLLLEEEDLKDMGIVSKGHIIHFKSAIEKLTHDYINLFHFPPL
IKDSGGEPEENEEKIVNLELVFGFHLKPGTGPQDCKWKMYMEMDGDEIAITYIKDVTFNTNLPDAEILKM
TKPPFVMEKWIVGIAKSQTVECTVTYESDVRTPKSTKHVHSIQWSRTKPQDEVKAVQLAIQTLFTNSDGN
PGSRSDSSADCQWLDTLRMRQIASNTSLQRSQSNPILGSPFFSHFDGQDSYAAAVRRPQVPIKYQQITPV
NQSRSSSPTQYGLTKNFSSLHLNSRDSGFSSGNTDTSSERGRYSDRSRNKYGRGSISLNSSPRGRYSGKS
QHSTPSRGRYPGKFYRVSQSALNPHQSPDFKRSPRDLHQPNTIPGMPLHPETDSRASEEDSKVSEGGWTK
Search details
cancer suppressor g
AND "Homo sapiens"[
Search
Recent activity

Display Settings: GenPept

Send to:

RecName: Full=Mitogen-activated protein kinase kinase kinase MLT; AltName: Full=Human cervical cancer suppressor gene 4 protein; Short=HCCS-4; AltName: Full=Leucine zipper- and sterile alpha motif-containing kinase; AltName: Full=MLK-like mitogen-activated pro...

UniProtKB/Swiss-Prot: Q9NYL2.3

FASTA Graphics

Go to:

```
LOCUS       MLTK_HUMAN               800 aa            linear   PRI 28-NOV-2012
DEFINITION  RecName: Full=Mitogen-activated protein kinase kinase kinase MLT;
            AltName: Full=Human cervical cancer suppressor gene 4 protein;
            Short=HCCS-4; AltName: Full=Leucine zipper- and sterile alpha
            motif-containing kinase; AltName: Full=MLK-like mitogen-activated
            protein triple kinase; AltName: Full=Mixed lineage kinase-related
            kinase; Short=MLK-related kinase; Short=MRK; AltName: Full=Sterile
            alpha motif- and leucine zipper-containing kinase AZK.
ACCESSION   Q9NYL2
VERSION     Q9NYL2.3  GI:313104215
DBSOURCE    UniProtKB: locus MLTK_HUMAN, accession Q9NYL2;
            class: standard.
            extra
            accessions:B3KPG2,Q53SX1,Q580W8,Q59GY5,Q86YW8,Q9HCC4,Q9HCC5,Q9HDD2,
            Q9NYE9
            created: Jul 5, 2005.
            sequence updated: Nov 30, 2010.
            annotation updated: Nov 28, 2012.
            xrefs: AF238255.1, AAF63490.1, AB049733.1, BAB16444.1, AB049734.1,
            BAB16445.1, AF325454.1, AAK11615.1, AF480461.1, AAL85891.1,
            AF480462.1, AAL85892.1, AB030034.1, BAB12040.1, AF251441.1,
            AAF65822.1, AF465843.1, AAO33376.1, AK056310.1, BAG51674.1,
            AB208974.1, BAD92211.1, AC092573.2, AAX82002.1, AC013461.9,
            EAX11164.1, BC001401.2,
```

ww.ncbi.nlm.nih.gov/protein/294862468?report=graph#

Change region shown

Customize view

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Articles about the ZAK gene

ZAK: a MAP3Kinase that transduces Shiga toxin- and ricin-induced pro [Cell Microbiol. 2008]

ZAK re-programs atrial natriuretic factor expressi [Biochem Biophys Res Commun. 2004]

A novel role for mixed-lineage kinase-like mitogen-activated protein trip [Cancer Res. 2004]

See all...

Identical proteins for Q9NYL2.3

mitogen-activated protein kinase kir [NP_057737]

plaucible mixed-lineage kinase prote [BAB12040]

See all...

Pathways for the ZAK gene

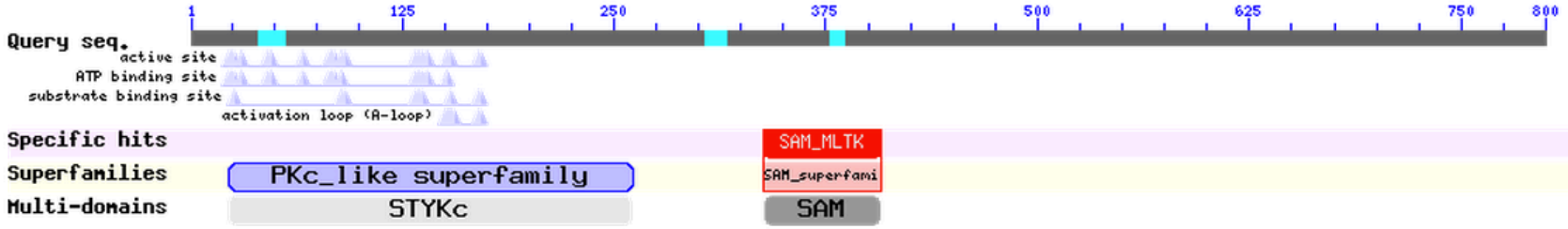
Conserved domains on [gi|313104215|sp|Q9NYL2|]

View full result

RecName: Full=Mitogen-activated protein kinase kinase kinase MLT; AltName: Full=Human cervical cancer suppressor gene 4 protein; Short=HCCS-4; AltName: Full=Leucine zipper- and sterile alpha motif-containing kinase; AltName: Full=MLK-like mitogen-activated protein triple kinase; AltName: Full=Mixed lineage kinase-related kinase; Short=MLK-related kinase; Short=MRK; AltName: Full=Sterile alpha motif- and leucine zipper-containing kinase AZK

Graphical summary

show options »



Search for similar domain architectures

Refine search

List of domain hits

Description	Pssmid	Multi-dom	E-value
[+]SAM_MLTK[cd09529], SAM domain of MLTK subfamily; SAM (sterile alpha motif) domain of MLTK subfamily is a protein-protein interaction ...	188928	no	1.48e-31
[+]PTKc[cd00192], Catalytic domain of Protein Tyrosine Kinases; Protein Tyrosine Kinase (PTK) family, catalytic domain. This PTKc family is part of a ...	173624	yes	3.00e-55
[+]STYKc[smart00221], Protein kinase; unclassified specificity.; Phosphotransferases. The specificity of this class of kinases can not be predicted. ...	197583	yes	4.51e-68
[+]SAM[smart00454], Sterile alpha motif.; Widespread domain in signalling and nuclear proteins. In EPH-related tyrosine kinases, ...	197735	yes	5.88e-05

References:

Pre-computed BLAST results for: gi|313104215|sp|Q9NYL2.3 RecName: Full=Mitogen-activated protein kinase kinase kinase MLT; AltName: Full=Human cervical cance

Matching gis: 82880648;9927293;

Total (score > 100) : 184000 hits in 181913 proteins in 7930 species

Selected: 184000 hits in 181913 proteins in 7930 species Filter: Min Score: 100 |

Other views (Reports): Taxonomy report | Multiple Alignment | Blast

Reset all filters

▶ Choose Display Options

152 Archaea | 27030 Bacteria | 67989 Metazoa | 20927 Fungi | 43342 Plants | 713 Viruses | 23847 The Others | reset selection

Results: 1 - 100 Next Page Last

% hits reset selection

800 aa

blink

SCORE	ACCESSION	Length	Protein Description
	Conserved Domain Database hits		
4234	NP_057737	800	mitogen-activated protein kinase kinase kinase MLT isoform 1 [Homo sapiens]
4234	BAB12040	800	plaucible mixed-lineage kinase protein [Homo sapiens]
4228	AAL85891	800	mixed lineage kinase-related kinase MRK-alpha [Homo sapiens]
4228	EAX11165	800	sterile alpha motif and leucine zipper containing kinase AZK, isoform CRA_b [Hom
4228	AAF65822	800	sterile-alpha motif and leucine zipper containing kinase AZK [Homo sapiens]
4228	BAD92211	845	Plaucible mixed-lineage kinase protein variant [Homo sapiens]
4228	BAG10659	800	mitogen-activated protein kinase kinase kinase MLT [synthetic construct]
4218	BAB16444	800	MLTK-alpha [Homo sapiens]
4217	AAF63490	800	mixed lineage kinase ZAK [Homo sapiens]
4209	XP_003824298	800	PREDICTED: mitogen-activated protein kinase kinase kinase MLT-like isoform 2 [Pa
4209	JAA10812	800	sterile alpha motif and leucine zipper containing kinase AZK [Pan troglodytes]
4209	JAA17990	800	sterile alpha motif and leucine zipper containing kinase AZK [Pan troglodytes]
4209	JAA30788	800	sterile alpha motif and leucine zipper containing kinase AZK [Pan troglodytes]
4209	JAA43585	800	sterile alpha motif and leucine zipper containing kinase AZK [Pan troglodytes]
4209	XP_003824297	800	PREDICTED: mitogen-activated protein kinase kinase kinase MLT-like isoform 1 [Pa
4208	XP_003253770	800	PREDICTED: mitogen-activated protein kinase kinase kinase MLT-like isoform 2 [No
4204	XP_003309476	800	PREDICTED: LOW QUALITY PROTEIN: mitogen-activated protein kinase kinase kinase M
4196	XP_003907669	800	PREDICTED: mitogen-activated protein kinase kinase kinase MLT-like isoform 2 [Pa

ZAK sterile alpha motif and leucine zipper containing kinase AZK [*Homo sapiens*]

Gene ID: 51776, updated on 6-Jan-2013

Summary

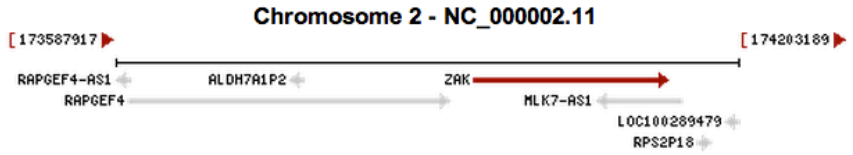
Gene symbol	ZAK
Gene description	sterile alpha motif and leucine zipper containing kinase AZK
Locus tag	HCCS4
See related	Ensembl:ENSG00000091436; HPRD:11791; MIM:609479; Vega:OTTHUMG00000132297
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	pk; AZK; MLT; MRK; MLK7; MLTK; mlklak
Summary	This gene is a member of the MAPKKK family of signal transduction molecules and encodes a protein with an N-terminal kinase catalytic domain, followed by a leucine zipper motif and a sterile-alpha motif (SAM). This magnesium-binding protein forms homodimers and is located in the cytoplasm. The protein mediates gamma radiation signaling leading to cell cycle arrest and activity of this protein plays a role in cell cycle checkpoint regulation in cells. The protein also has pro-apoptotic activity. Alternate transcriptional splice variants, encoding different isoforms, have been characterized. [provided by RefSeq, Jul 2008]

Genomic context

Location: 2q24.2

Sequence: Chromosome: 2; NC_000002.11 (173940565..174132737)

See ZAK in Epigenomics, MapViewer



Genomic regions, transcripts, and products

Genomic Sequence NC_000002 chromosome 2 reference GRCh37.p10 Primary Assembly

Go to reference sequence details

Related information

Hands-on exercise 2

Given a list of sequence IDs, get their sequences from NCBI

Suppose

- You read a paper which reported a list of genes (with a table e.g. to show all the IDs)

Or

- You have a collaborator sending you a file with all the IDs

You want to quickly get the sequences of these genes

Download the example id file at http://cys.bios.niu.edu/yyin/teach/PBB/gt8-id.txt

Plain text file!


www.ncbi.nlm.nih.gov
corecarb - Google Sc
George Mason Unive
Customize Links
Free Hotmail
RealPlayer
Windows Marketplac
Windows Media
Windows
Gmail - Inbox (77) -
travel expense
NCBI
Resources
How To
NCBI
National Center for Biotechnology Information
All Databases
Search
NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation
Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
About the NCBI | Mission | Organization | Research | RSS Feeds
Get Started
Tools: Analyze data using NCBI software
Downloads: Get NCBI data or software
How-To's: Learn how to accomplish specific tasks at NCBI
Submissions: Submit data to GenBank or other NCBI databases
NCBI Facebook page
Find out the latest news about NCBI resources and participate in community discussions.
GO
1 2 3 4 5 6 7 8
Popular Resources
PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem
NCBI Announcements
Come to the NCBI Disco
on February 4&5!
Spaces are still available
New version of Genome
available

Site Map

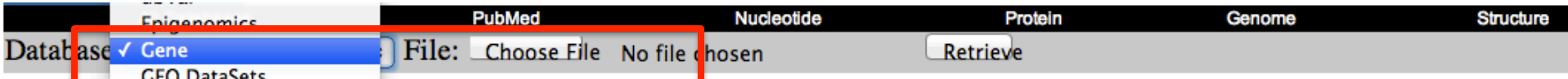
1 A B C D E F G H I J L M N O P R S T U V

Featured items are in bold.

1
1000 Genomes Browser

A
Amino Acid Explorer
ASN.1 Format Summary
Assembly Archive

B
Basic Local Alignment Search Tool (BLAST)
Batch Entrez
~~BioAssay Services~~
BioProject (formerly Genome Project)
BioProject Submission
BioSample
BioSystems
BLAST (Stand-alone)
BLAST Link (BLink)
BLAST Microbial Genomes
BLAST RefSeqGene
BLAST Tutorials and Guides
Bookshelf



Batch

Use Batch to upload a file of GIs or accession numbers from the Nucleotide or Protein database or a list of record identifiers from other Entrez databases.

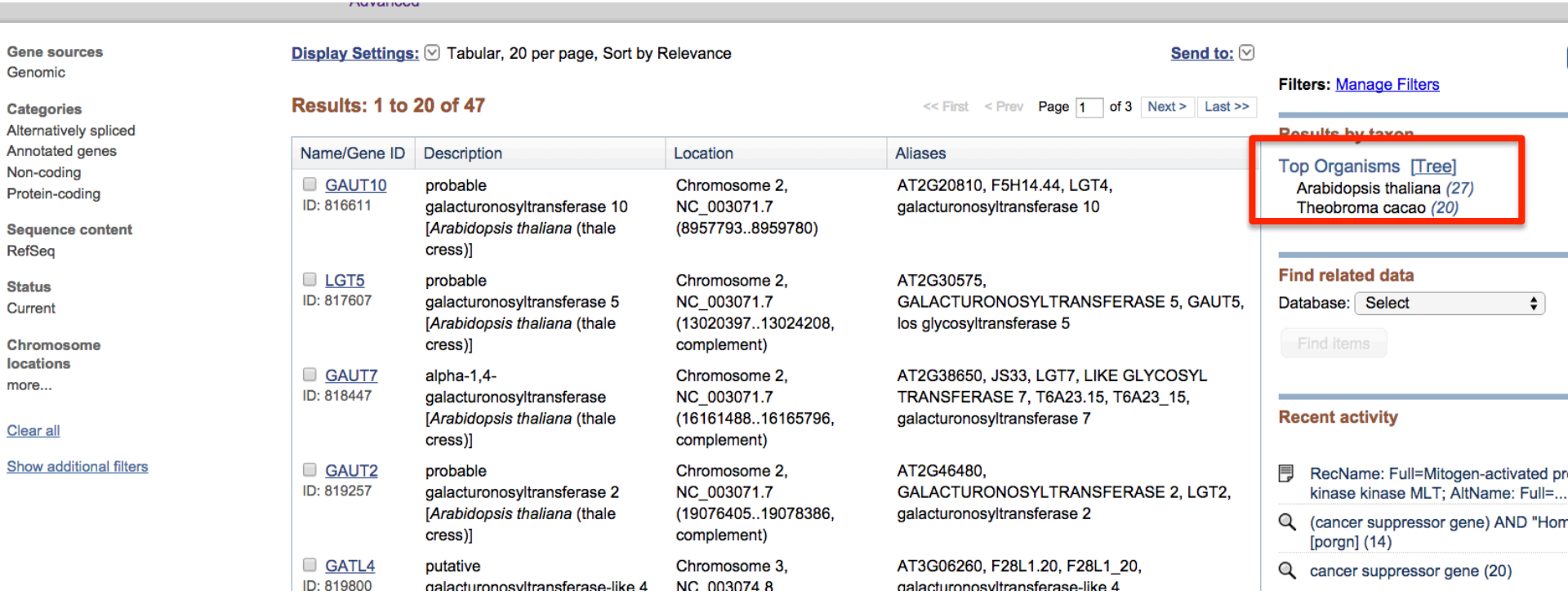
Tips : Sometimes lists of record identifiers can be tens of thousands of lines long and Batch Entrez may not retrieve all records from one list. Split the list of identifiers into smaller files using a file splitting software or a file split command at the command prompt in UNIX or LINUX systems. Please limit to a few thousand record identifiers per file, left-formatted, and one per line. This may be necessary if you are loading large numbers of genome records. You can check the NCBI FTP site to download entire genome records. Also, use GIs rather than 'accessions' when making lists for batch entrez to fetch.

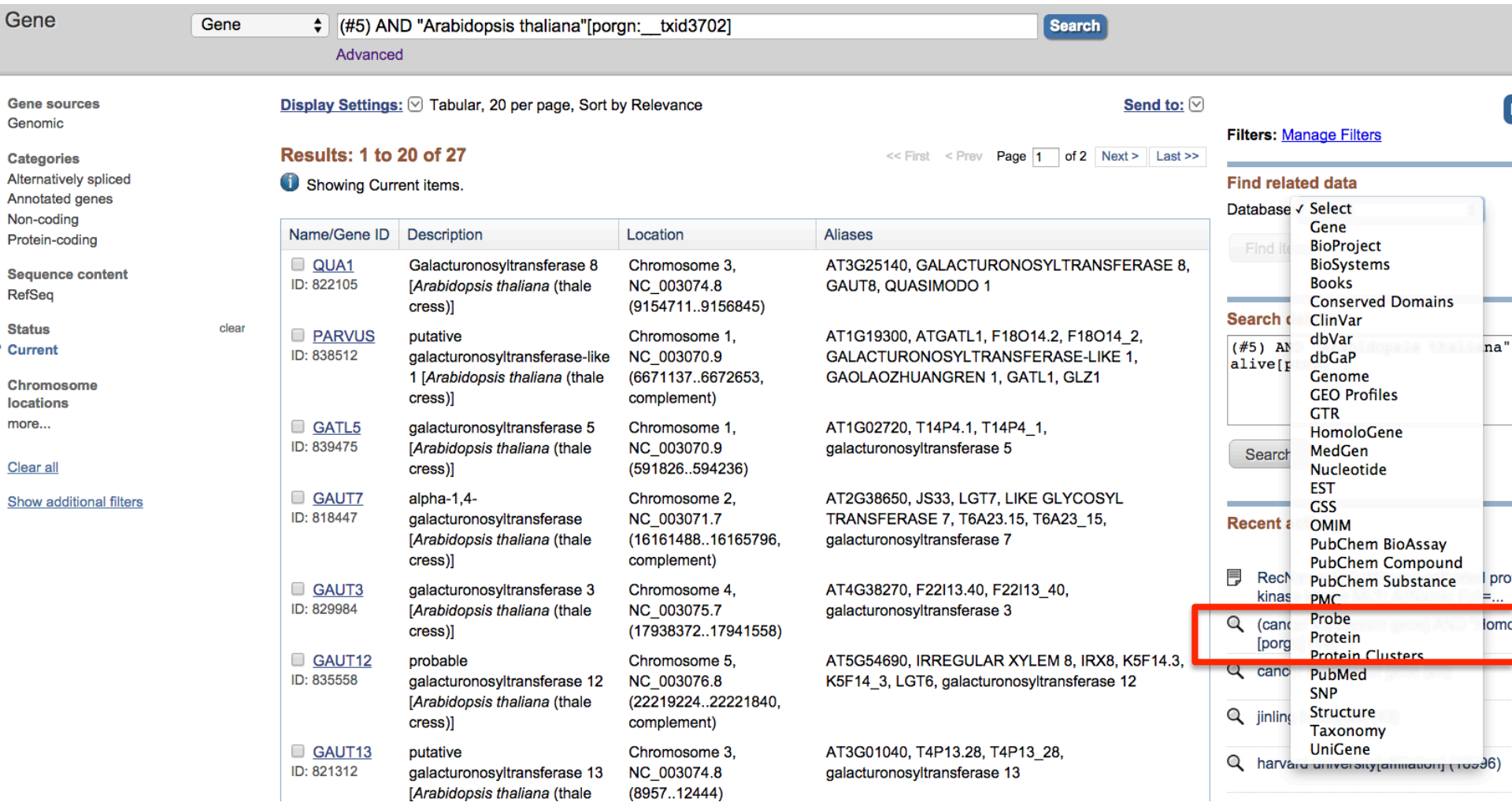
Please note that Batch Entrez will check for duplicate identifiers when reporting results from a list that you have imported.

When retrieving nucleotide accessions, you must select the specific component database from which the accessions or GIs were saved. For Nucleotide, choose either the CoreNucleotide, the EST or the GSS selection from the database menu. If you have a mixed list of nucleotide accessions or UIDs, you will need to run the Batch Entrez search three times. Select the database from the pull-down menu, CoreNucleotide, EST, and GSS separately.

In all cases, be certain to select the correct database for uploaded identifiers when using Batch Entrez, to ensure the expected records. For example, if you have saved a list of protein GIs, be sure to select the Protein database.

- Create a file with a list of GI or accession numbers and save it locally
- Select the database from which the list of accessions or UIDs originated
- Use the 'Browse' button to select the filename containing the list of UIDs from your system directory
- Press the Retrieve button and you will see a list of document summaries
- Select a format in which to display the data for viewing, and/or saving
- Select 'Send to file' to save the file.


Gene sources
Genomic
Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Sequence content
RefSeq
Status
Current
Chromosome locations
more...
Clear all
Show additional filters
Display Settings: Tabular, 20 per page, Sort by Relevance
Send to:
Results: 1 to 20 of 47
<< First < Prev Page 1 of 3 Next > Last >>
Name/Gene ID
Description
Location
Aliases
GAUT10
ID: 816611
probable galacturonosyltransferase 10 [Arabidopsis thaliana (thale cress)]
Chromosome 2, NC_003071.7 (8957793..8959780)
AT2G20810, F5H14.44, LGT4, galacturonosyltransferase 10
LGT5
ID: 817607
probable galacturonosyltransferase 5 [Arabidopsis thaliana (thale cress)]
Chromosome 2, NC_003071.7 (13020397..13024208, complement)
AT2G30575, GALACTURONOSYLTRANSFERASE 5, GAUT5, los glycosyltransferase 5
GAUT7
ID: 818447
alpha-1,4-galacturonosyltransferase [Arabidopsis thaliana (thale cress)]
Chromosome 2, NC_003071.7 (16161488..16165796, complement)
AT2G38650, JS33, LGT7, LIKE GLYCOSYL TRANSFERASE 7, T6A23.15, T6A23_15, galacturonosyltransferase 7
GAUT2
ID: 819257
probable galacturonosyltransferase 2 [Arabidopsis thaliana (thale cress)]
Chromosome 2, NC_003071.7 (19076405..19078386, complement)
AT2G46480, GALACTURONOSYLTRANSFERASE 2, LGT2, galacturonosyltransferase 2
GATL4
ID: 819800
putative
Chromosome 3,
AT3G06260, F28L1.20, F28L1_20,
Filters: Manage Filters
Top Organisms [Tree]
Arabidopsis thaliana (27)
Theobroma cacao (20)
Find related data
Database: Select
Find items
Recent activity
RecName: Full=Mitogen-activated pr
kinase kinase MLT; AltName: Full=...
(cancer suppressor gene) AND "Hom
[porgn] (14)
cancer suppressor gene (20)


Gene
Gene
(#5) AND "Arabidopsis thaliana"[porgn:__txid3702]
Search
Advanced
Gene sources
Genomic
Categories
Alternatively spliced
Annotated genes
Non-coding
Protein-coding
Sequence content
RefSeq
Status
clear
Current
Chromosome locations
more...
Clear all
Show additional filters
Display Settings: Tabular, 20 per page, Sort by Relevance
Send to:
Results: 1 to 20 of 27
<< First < Prev Page 1 of 2 Next > Last >>
Showing Current items.
Name/Gene ID
Description
Location
Aliases
QUA1
ID: 822105
Galacturonosyltransferase 8 [Arabidopsis thaliana (thale cress)]
Chromosome 3, NC_003074.8 (9154711..9156845)
AT3G25140, GALACTURONOSYLTRANSFERASE 8, GAUT8, QUASIMODO 1
PARVUS
ID: 838512
putative galacturonosyltransferase-like 1 [Arabidopsis thaliana (thale cress)]
Chromosome 1, NC_003070.9 (6671137..6672653, complement)
AT1G19300, ATGATL1, F18O14.2, F18O14_2, GALACTURONOSYLTRANSFERASE-LIKE 1, GAOLAOZHUANGREN 1, GATL1, GLZ1
GATL5
ID: 839475
galacturonosyltransferase 5 [Arabidopsis thaliana (thale cress)]
Chromosome 1, NC_003070.9 (591826..594236)
AT1G02720, T14P4.1, T14P4_1, galacturonosyltransferase 5
GAUT7
ID: 818447
alpha-1,4-galacturonosyltransferase [Arabidopsis thaliana (thale cress)]
Chromosome 2, NC_003071.7 (16161488..16165796, complement)
AT2G38650, JS33, LGT7, LIKE GLYCOSYL TRANSFERASE 7, T6A23.15, T6A23_15, galacturonosyltransferase 7
GAUT3
ID: 829984
galacturonosyltransferase 3 [Arabidopsis thaliana (thale cress)]
Chromosome 4, NC_003075.7 (17938372..17941558)
AT4G38270, F22I13.40, F22I13_40, galacturonosyltransferase 3
GAUT12
ID: 835558
probable galacturonosyltransferase 12 [Arabidopsis thaliana (thale cress)]
Chromosome 5, NC_003076.8 (22219224..22221840, complement)
AT5G54690, IRREGULAR XYLEM 8, IRX8, K5F14.3, K5F14_3, LGT6, galacturonosyltransferase 12
GAUT13
ID: 821312
putative galacturonosyltransferase 13 [Arabidopsis thaliana (thale
Chromosome 3, NC_003074.8 (8957..12444)
AT3G01040, T4P13.28, T4P13_28, galacturonosyltransferase 13
Filters: Manage Filters
Find related data
Database
Select
Gene
BioProject
BioSystems
Books
Conserved Domains
ClinVar
dbVar
dbGaP
Genome
GEO Profiles
GTR
HomoloGene
MedGen
Nucleotide
EST
GSS
OMIM
PubChem BioAssay
PubChem Compound
PubChem Substance
PMC
Probe
Protein
Protein Clusters
PubMed
SNP
Structure
Taxonomy
UniGene

Protein | Protein | Search | Limits | Advanced

Display Settings: Summary, 20 per page, Sorted by Default order | Send to:

Results: 1 to 20 of 213

<< First | < Prev | Page 1 of 11 | Next > | Last >>

1. RecName: Full=Probable galacturonosyltransferase 3
   680 aa protein
   Accession: Q0WQD2.2 GI: 357528801
   GenPept FASTA Graphics Related Sequences Identical Proteins

2. RecName: Full=Probable galacturonosyltransferase 6
   589 aa protein
   Accession: Q9M9Y5.1 GI: 75191689
   GenPept FASTA Graphics Related Sequences Identical Proteins

3. RecName: Full=Polygalacturonate 4-alpha-galacturonosyltransferase; AltName: Full=Alpha-1,4-galacturonosyltransferase 1; AltName: Full=Galacturonosyltransferase 1; AltName: Full=Like glycosyl transferase 1
   673 aa protein
   Accession: Q9LE59.1 GI: 75173891
   GenPept FASTA Graphics Related Sequences Identical Proteins

4. RecName: Full=Probable galacturonosyltransferase-like 5
   361 aa protein
   Accession: Q9FWY9.1 GI: 75172933
   GenPept FASTA Graphics Related Sequences Identical Proteins

5. RecName: Full=Probable galacturonosyltransferase 4; AltName: Full=Like glycosyl transferase 3
   616 aa protein
   Accession: Q93ZX7.1 GI: 75163841
   GenPept FASTA Graphics Related Sequences Identical Proteins

RecName: Full=Probable galacturonosyltransferase 7; AltName: Full=Like glycosyl transferase 7

Filter your results:

All (213)
Bacteria (0)
Related Structures (156)
RefSeq (35)

Mana

Find related data

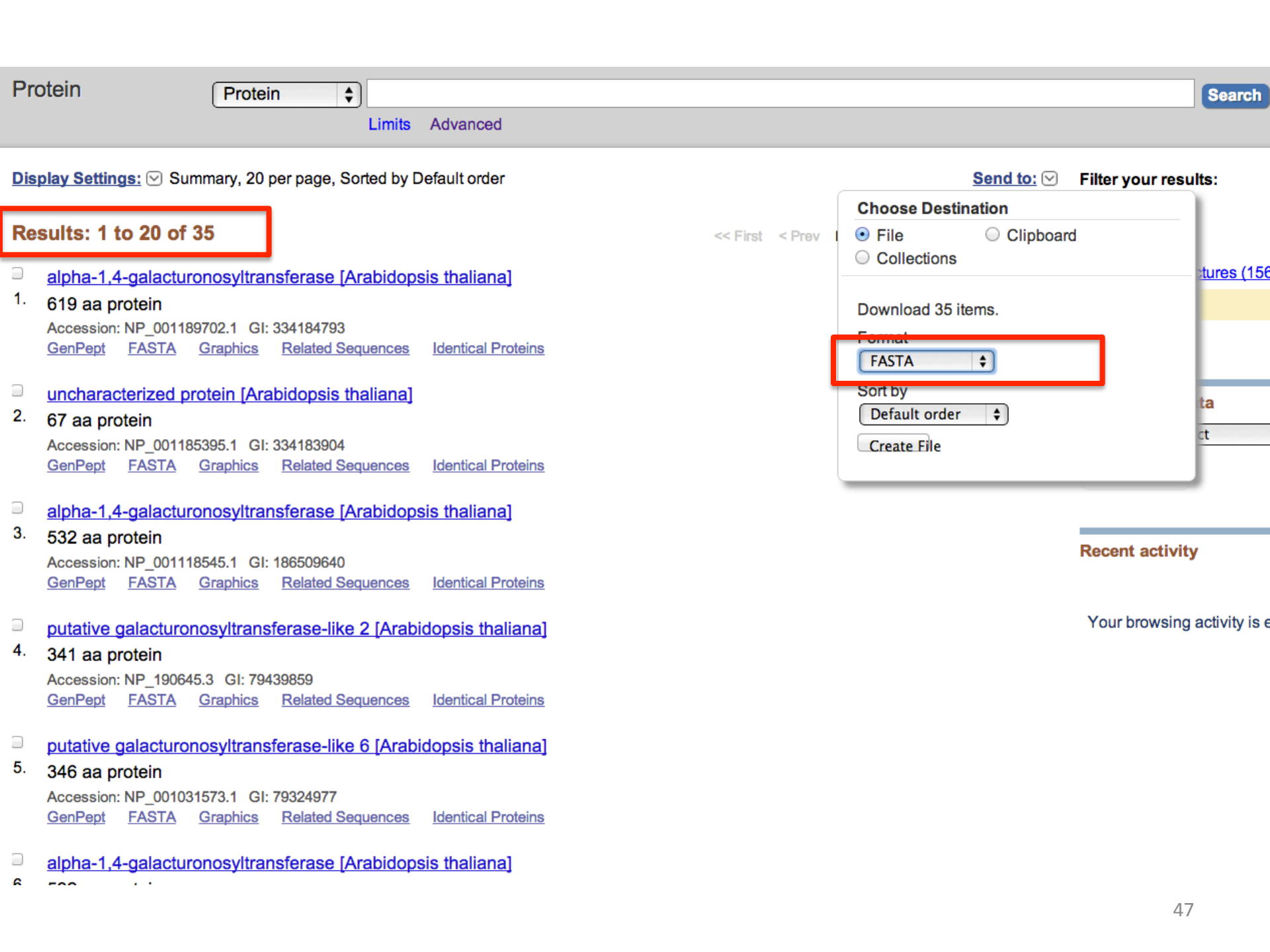
Database: Select

Find items

Recent activity

Turn C

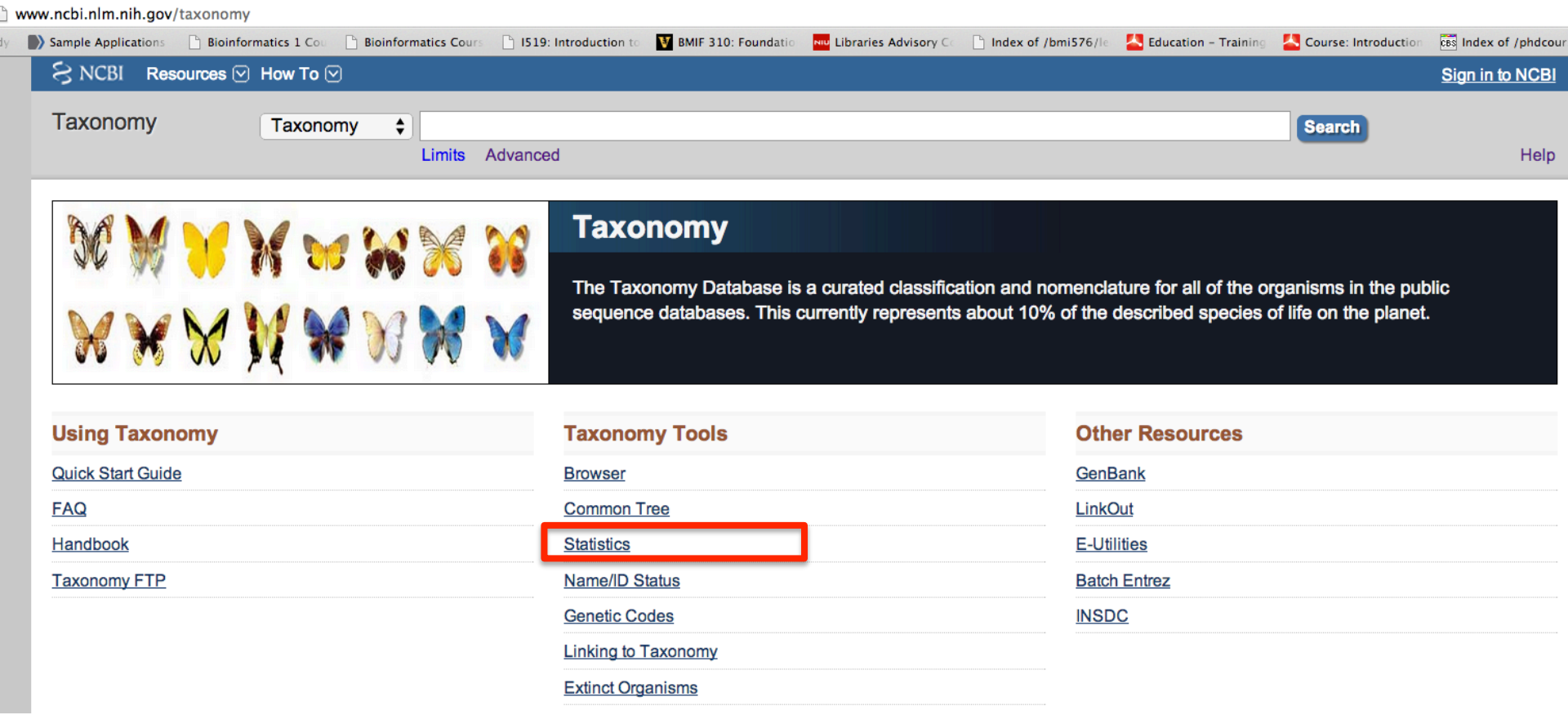
Your browsing activity is empty.


Protein
Protein
Search
Limits
Advanced
Display Settings: Summary, 20 per page, Sorted by Default order
Send to:
Filter your results:
Results: 1 to 20 of 35
<< First
< Prev
Choose Destination
File
Clipboard
Collections
Download 35 items.
FASTA
Sort by
Default order
Create File
alpha-1,4-galacturonosyltransferase [Arabidopsis thaliana]
1. 619 aa protein
Accession: NP_001189702.1 GI: 334184793
GenPept FASTA Graphics Related Sequences Identical Proteins
uncharacterized protein [Arabidopsis thaliana]
2. 67 aa protein
Accession: NP_001185395.1 GI: 334183904
GenPept FASTA Graphics Related Sequences Identical Proteins
alpha-1,4-galacturonosyltransferase [Arabidopsis thaliana]
3. 532 aa protein
Accession: NP_001118545.1 GI: 186509640
GenPept FASTA Graphics Related Sequences Identical Proteins
putative galacturonosyltransferase-like 2 [Arabidopsis thaliana]
4. 341 aa protein
Accession: NP_190645.3 GI: 79439859
GenPept FASTA Graphics Related Sequences Identical Proteins
putative galacturonosyltransferase-like 6 [Arabidopsis thaliana]
5. 346 aa protein
Accession: NP_001031573.1 GI: 79324977
GenPept FASTA Graphics Related Sequences Identical Proteins
alpha-1,4-galacturonosyltransferase [Arabidopsis thaliana]
Recent activity
Your browsing activity is e

Hands-on exercise 3

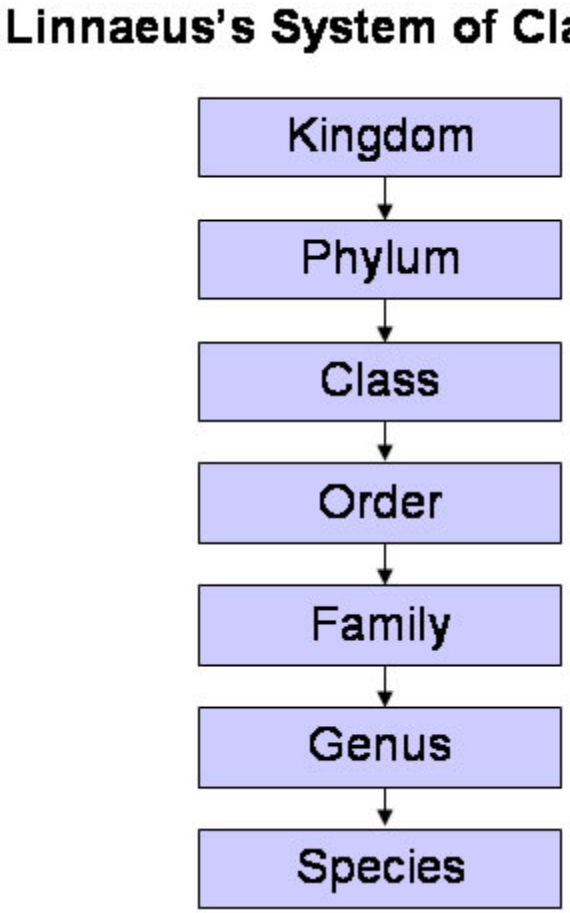
Find sequences through taxonomy database

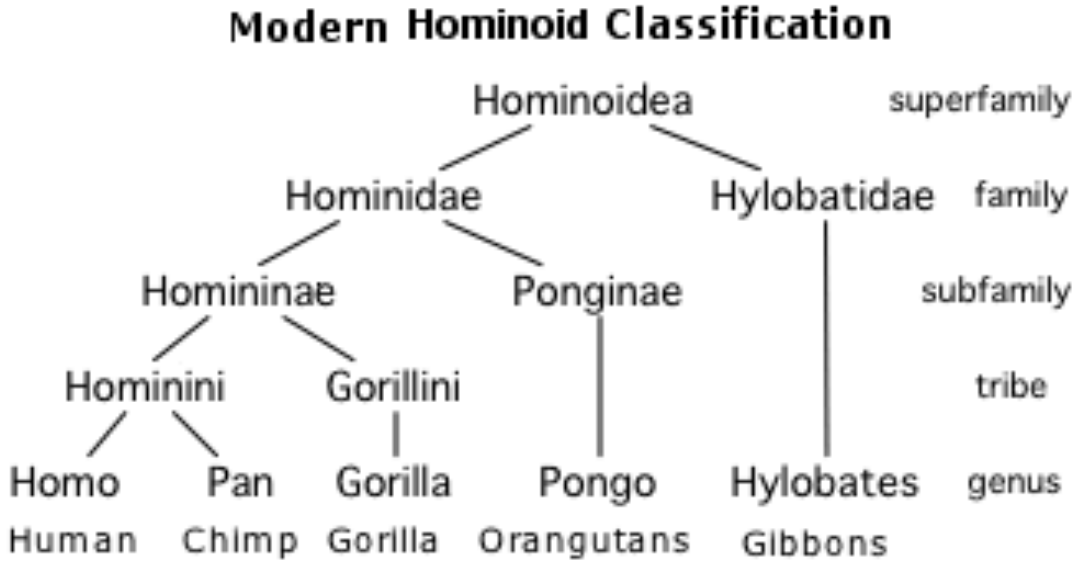
http://www.ncbi.nlm.nih.gov/taxonomy



http://nar.oxfordjournals.org/content/early/2011/12/01/nar.gkr1178.full-text-lowres.pdf

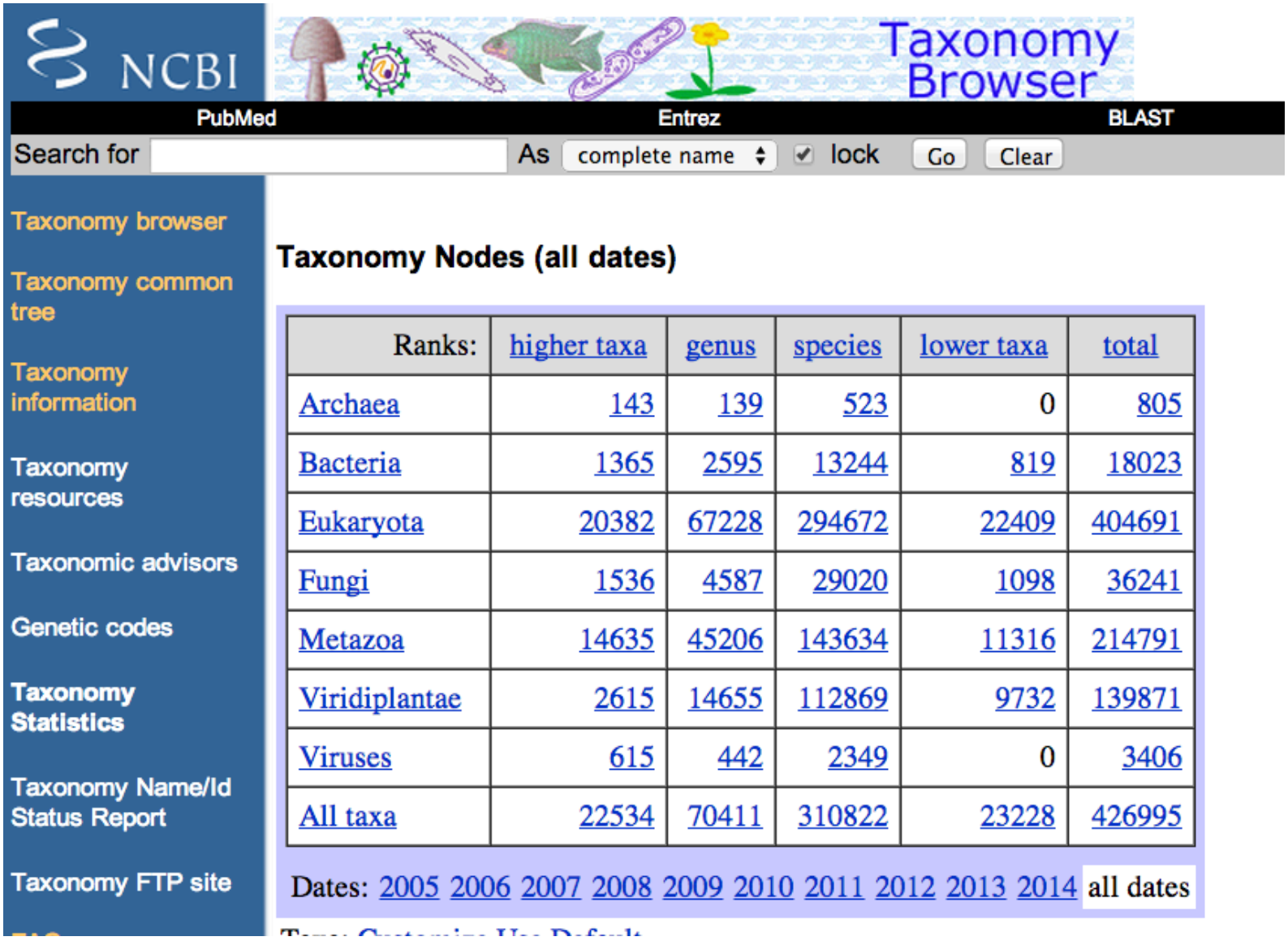
Taxonomy classification of species



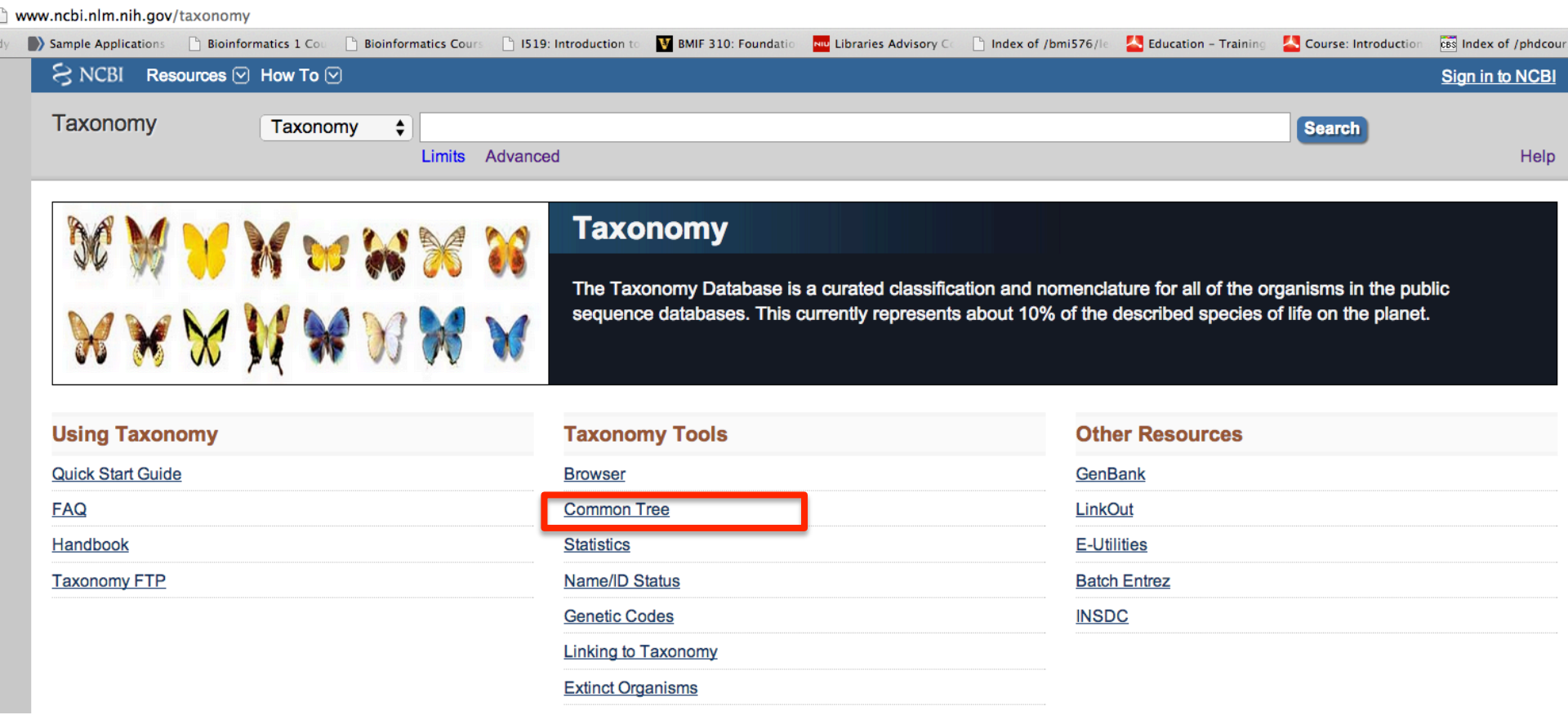


Every species has a unique taxonomy ID (e.g. human: 9606)

All species of GenBank have a taxonomy ID and lineage info


NCBI
Taxonomy Browser
PubMed
Entrez
BLAST
Search for
As
complete name
lock
Go
Clear
Taxonomy browser
Taxonomy common tree
Taxonomy information
Taxonomy resources
Taxonomic advisors
Genetic codes
Taxonomy Statistics
Taxonomy Name/Id Status Report
Taxonomy FTP site
Taxonomy Nodes (all dates)
Ranks: higher taxa genus species lower taxa total
Archaea 143 139 523 0 805
Bacteria 1365 2595 13244 819 18023
Eukaryota 20382 67228 294672 22409 404691
Fungi 1536 4587 29020 1098 36241
Metazoa 14635 45206 143634 11316 214791
Viridiplantae 2615 14655 112869 9732 139871
Viruses 615 442 2349 0 3406
All taxa 22534 70411 310822 23228 426995
Dates: 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 all dates

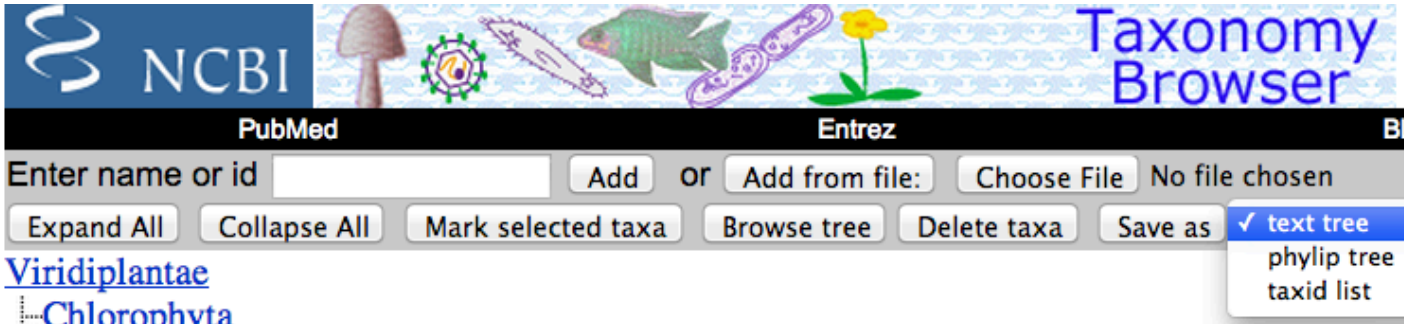
If you have a list of species names and you want to find out how they are related according to NCBI taxonomy database



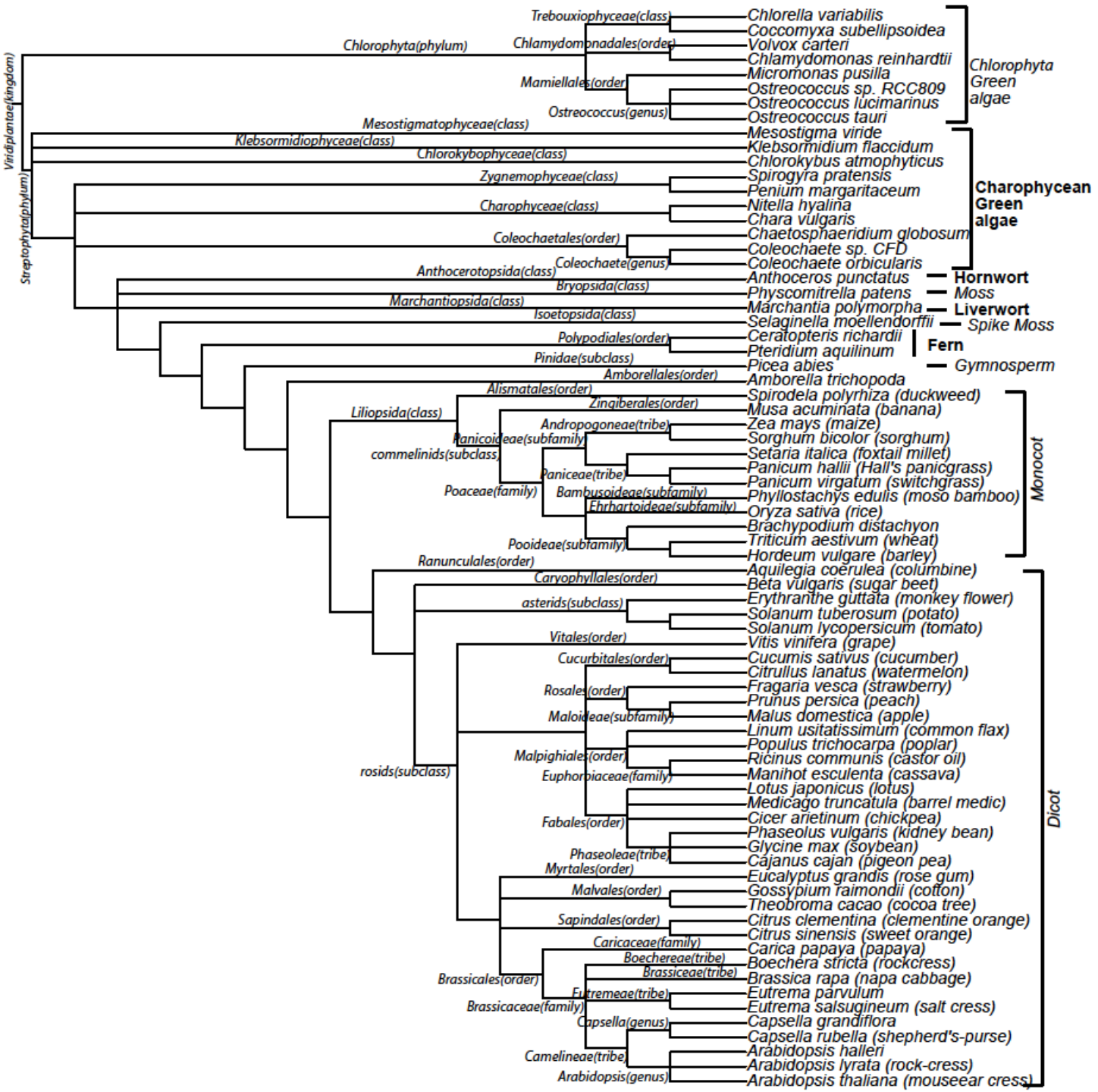
Download a list of plant species from:
http://cys.bios.niu.edu/yyin/teach/PBB/plant-genome.txt

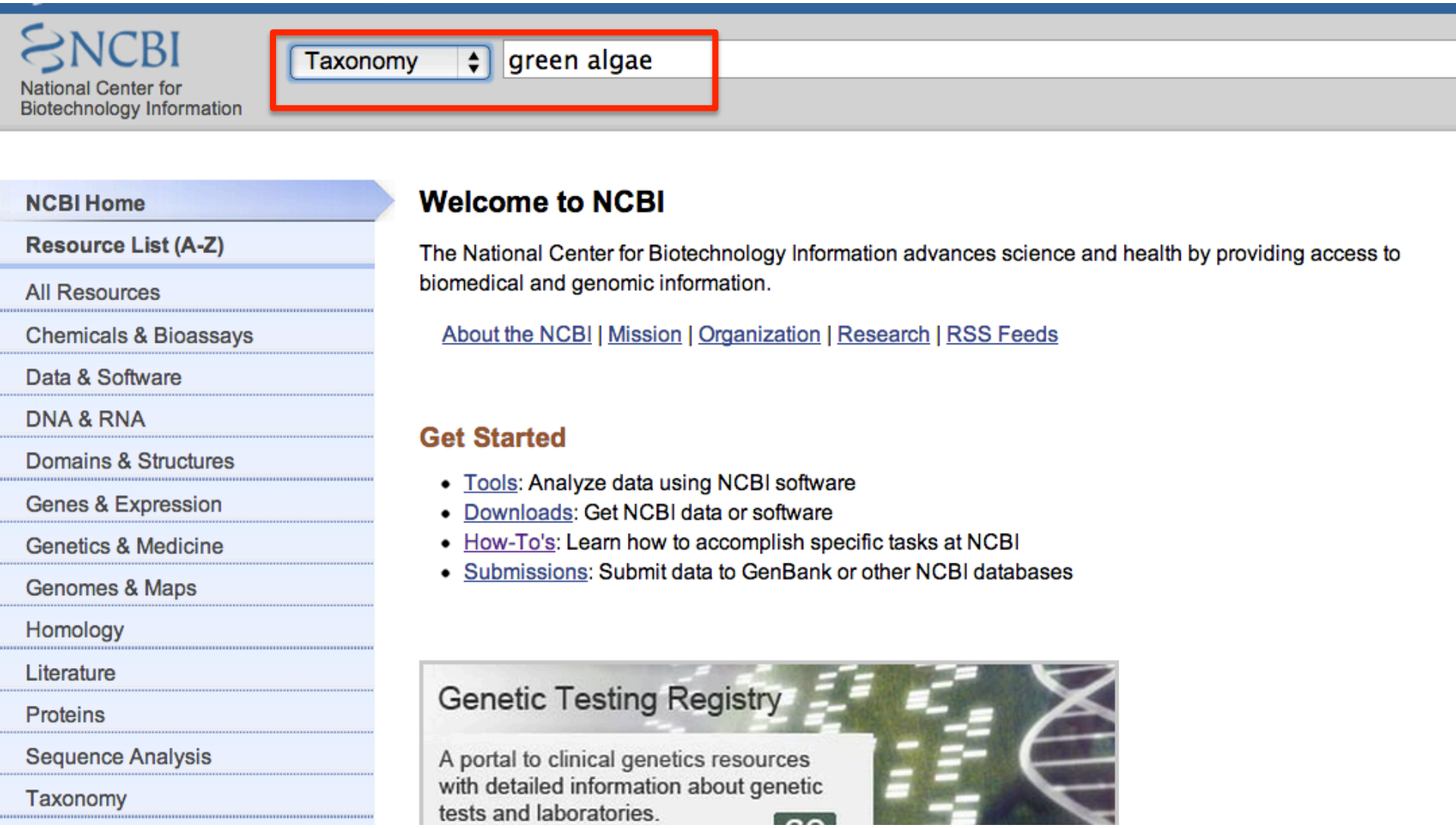
Choose file -> Add from file

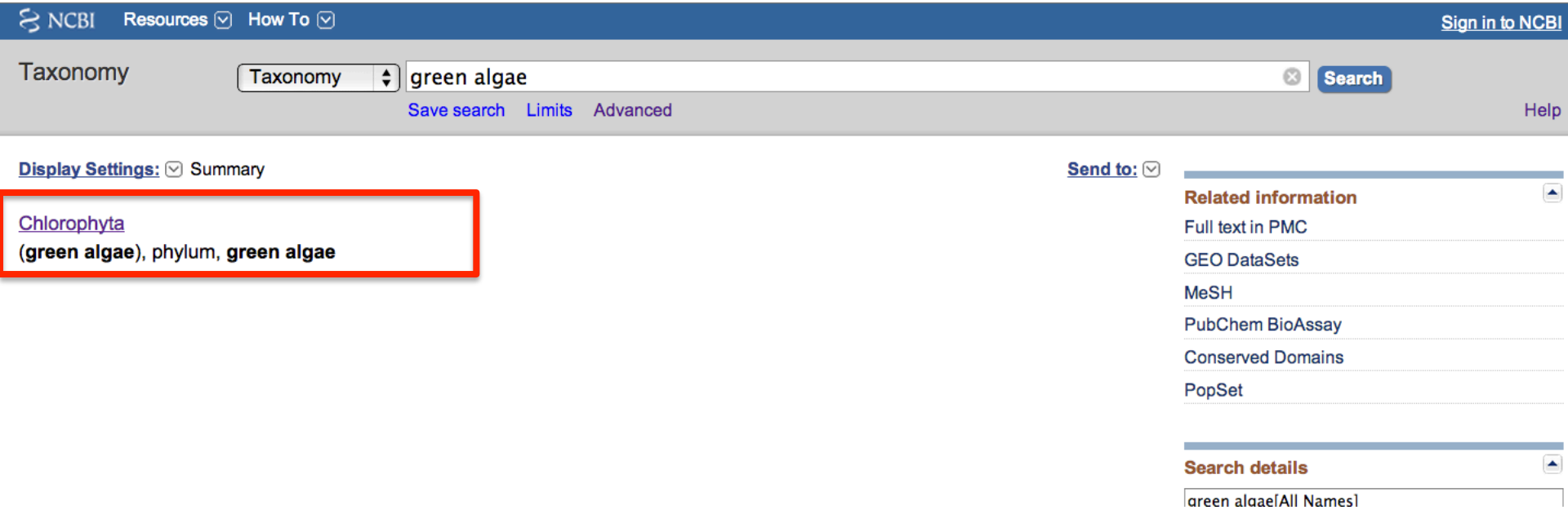
Here is how these plants are distributed in the taxonomic classification



You may save this as a phylip format file, which could be presented as a phylogeny graph using tree viewer softwares

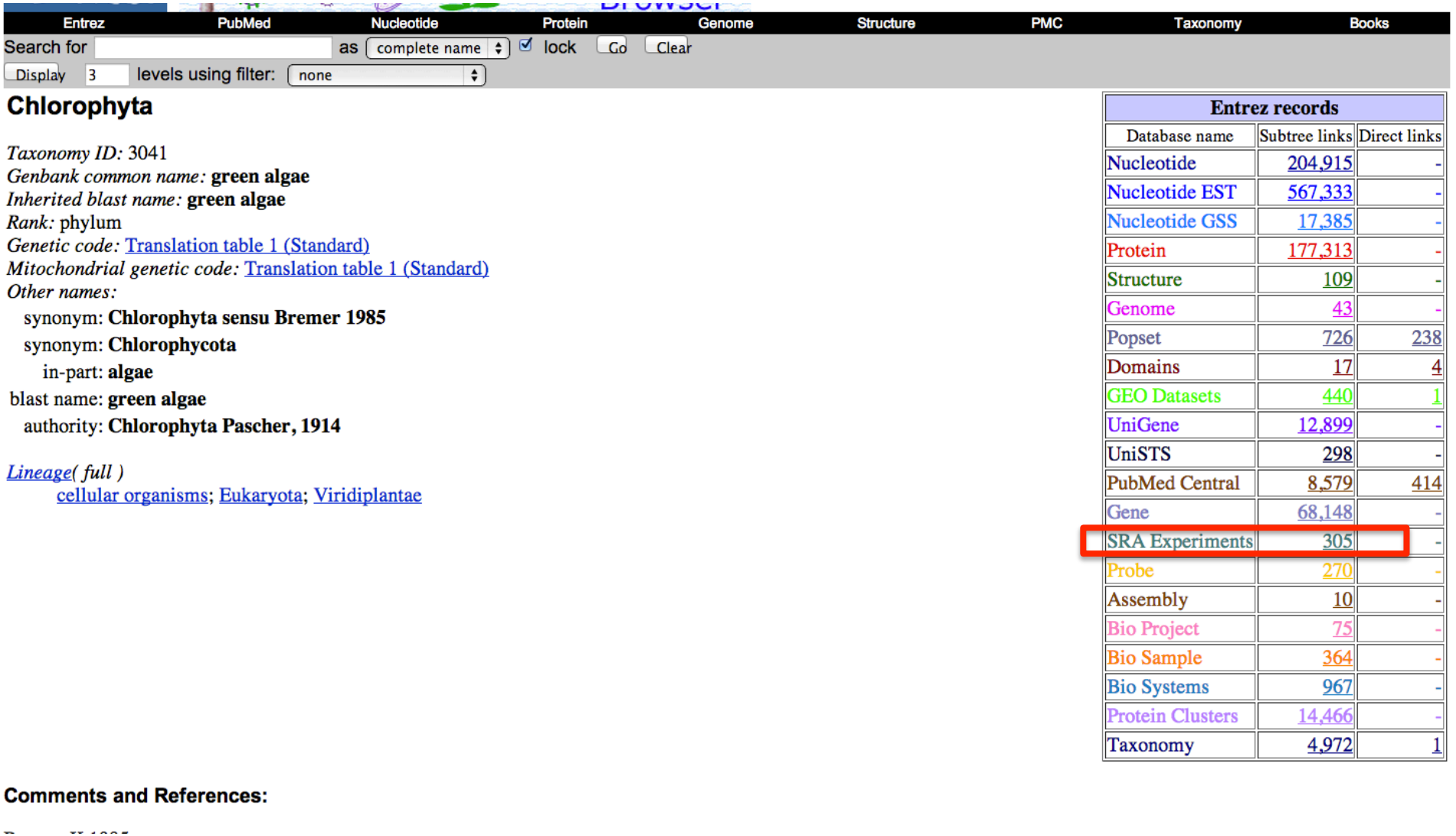

Viridiplantae(kingdom)
Streptophyta(phylum)
Chlorophyta(phylum)
Trebouxiophyceae(class)
Chlamydomonadales(order)
Mamiellales(order)
Ostreococcus(genus)
Chlorella variabilis
Coccomyxa subellipsoidea
Volvox carteri
Chlamydomonas reinhardtii
Micromonas pusilla
Ostreococcus sp. RCC809
Ostreococcus lucimarinus
Ostreococcus tauri
Chlorophyta Green algae
Mesostigmatophyceae(class)
Klebsormidiophyceae(class)
Chlorokybophyceae(class)
Zygnemophyceae(class)
Charophyceae(class)
Coleochaetales(order)
Coleochaete(genus)
Mesostigma viride
Klebsormidium flaccidum
Chlorokybus atmophyticus
Spirogyra pratensis
Penium margaritaceum
Nitella hyalina
Chara vulgaris
Chaetosphaeridium globosum
Coleochaete sp. CFD
Coleochaete orbicularis
Charophycean Green algae
Anthocerotopsida(class)
Bryopsida(class)
Marchantiopsida(class)
Isoetopsida(class)
Polypodiales(order)
Pinidae(subclass)
Amborellales(order)
Anthoceros punctatus
Physcomitrella patens
Marchantia polymorpha
Selaginella moellendorffii
Ceratopteris richardii
Pteridium aquilinum
Picea abies
Amborella trichopoda
Hornwort
Moss
Liverwort
Spike Moss
Fern
Gymnosperm
Alismatales(order)
Zingiberales(order)
Liliopsida(class)
Andropogoneae(tribe)
Panicoideae(subfamily)
commelinids(subclass)
Paniceae(tribe)
Poaceae(family)
Bambusoideae(subfamily)
Ehrhartoideae(subfamily)
Pooideae(subfamily)
Spirodela polyrhiza (duckweed)
Musa acuminata (banana)
Zea mays (maize)
Sorghum bicolor (sorghum)
Setaria italica (foxtail millet)
Panicum hallii (Hall's panicgrass)
Panicum virgatum (switchgrass)
Phyllostachys edulis (moso bamboo)
Oryza sativa (rice)
Brachypodium distachyon
Triticum aestivum (wheat)
Hordeum vulgare (barley)
Monocot
Ranunculales(order)
Caryophyllales(order)
asterids(subclass)
Vitales(order)
Cucurbitales(order)
Rosales(order)
Maloideae(subfamily)
Malpighiales(order)
Euphorbiaceae(family)
Fabales(order)
Phaseoleae(tribe)
Myrtales(order)
Malvales(order)
Sapindales(order)
Caricaceae(family)
Boechereae(tribe)
Brassicales(order)
Brassiceae(tribe)
Eutremeae(tribe)
Brassicaceae(family)
Capsella(genus)
Camelineae(tribe)
Arabidopsis(genus)
rosids(subclass)
Aquilegia coerulea (columbine)
Beta vulgaris (sugar beet)
Erythranthe guttata (monkey flower)
Solanum tuberosum (potato)
Solanum lycopersicum (tomato)
Vitis vinifera (grape)
Cucumis sativus (cucumber)
Citrullus lanatus (watermelon)
Fragaria vesca (strawberry)
Prunus persica (peach)
Malus domestica (apple)
Linum usitatissimum (common flax)
Populus trichocarpa (poplar)
Ricinus communis (castor oil)
Manihot esculenta (cassava)
Lotus japonicus (lotus)
Medicago truncatula (barrel medic)
Cicer arietinum (chickpea)
Phaseolus vulgaris (kidney bean)
Glycine max (soybean)
Cajanus cajan (pigeon pea)
Eucalyptus grandis (rose gum)
Gossypium raimondii (cotton)
Theobroma cacao (cocoa tree)
Citrus clementina (clementine orange)
Citrus sinensis (sweet orange)
Carica papaya (papaya)
Boechera stricta (rockcress)
Brassica rapa (napa cabbage)
Eutrema parvulum
Eutrema salsugineum (salt cress)
Capsella grandiflora
Capsella rubella (shepherd's-purse)
Arabidopsis halleri
Arabidopsis lyrata (rock-cress)
Arabidopsis thaliana (mouseear cress)
Dicot


NCBI
National Center for Biotechnology Information
Taxonomy
green algae
NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
About the NCBI | Mission | Organization | Research | RSS Feeds
Get Started
Tools: Analyze data using NCBI software
Downloads: Get NCBI data or software
How-To's: Learn how to accomplish specific tasks at NCBI
Submissions: Submit data to GenBank or other NCBI databases
Genetic Testing Registry
A portal to clinical genetics resources with detailed information about genetic tests and laboratories.


NCBI
Resources
How To
Sign in to NCBI
Taxonomy
Taxonomy
green algae
Search
Save search
Limits
Advanced
Help
Display Settings:
Summary
Send to:
Chlorophyta
(green algae), phylum, green algae
Related information
Full text in PMC
GEO DataSets
MeSH
PubChem BioAssay
Conserved Domains
PopSet
Search details
green algae[All Names]

Lineage (full): root; cellular organisms; Eukaryota; Viridiplantae

- **Chlorophyta** (green algae) *Click on organism name to get more information.*
  - **Chlorophyceae**
    - **Chaetopeltidales**
      - **Chaetopeltidaceae**
    - **Chaetophorales**
      - **Aphanochaetaceae**
      - **Chaetophoraceae**
      - **Schizomeridaceae**
      - **Chaetophorales incertae sedis**
    - **Chlamydomonadales**
      - **Asteromonadaceae**
      - **Astrephomenaceae**
      - **Characiochloridaceae**
      - **Characiosiphonaceae**
      - **Chlamydomonadaceae**
      - **Chlorococcaceae**
      - **Dunaliellaceae**
      - **Golenkiniaceae**
      - **Haematococcaceae**
      - **Phacotaceae**
      - **Spondylomoraceae**
      - **Tetrabaenaceae**
      - **Volvocaceae**


Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books
Search for as complete name lock Go Clear
Display 3 levels using filter: none
Chlorophyta
Taxonomy ID: 3041
Genbank common name: green algae
Inherited blast name: green algae
Rank: phylum
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 1 (Standard)
Other names:
synonym: Chlorophyta sensu Bremer 1985
synonym: Chlorophycota
in-part: algae
blast name: green algae
authority: Chlorophyta Pascher, 1914
Lineage(full)
cellular organisms; Eukaryota; Viridiplantae
Entrez records
Database name Subtree links Direct links
Nucleotide 204,915 -
Nucleotide EST 567,333 -
Nucleotide GSS 17,385 -
Protein 177,313 -
Structure 109 -
Genome 43 -
Popset 726 238
Domains 17 4
GEO Datasets 440 1
UniGene 12,899 -
UniSTS 298 -
PubMed Central 8,579 414
Gene 68,148 -
SRA Experiments 305 -
Probe 270 -
Assembly 10 -
Bio Project 75 -
Bio Sample 364 -
Bio Systems 967 -
Protein Clusters 14,466 -
Taxonomy 4,972 1
Comments and References:

SRA

SRA | txid3041[Organism:exp] NOT Chlamydomonas reinhardtii | Search

Save search | Limits | Advanced | Help

Display Settings: Summary, 20 per page

Send to:

Results: 1 to 20 of 346

<< First | < Prev | Page 1 of 18 | Next > | Last >>

1. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 11.8M spots, 2.4G bases, 1.6Gb downloads
   Accession: ERX177569

2. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 11M spots, 2.2G bases, 1.5Gb downloads
   Accession: ERX177568

3. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 9.5M spots, 1.9G bases, 1.3Gb downloads
   Accession: ERX177567

4. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 14M spots, 2.8G bases, 1.9Gb downloads
   Accession: ERX177566

5. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 10.1M spots, 2G bases, 1.4Gb downloads
   Accession: ERX177565

6. Transcriptome Analysis of Chlamydomonas reinhardtii
   1 ILLUMINA (Illumina HiSeq 2000) run: 13.5M spots, 2.7G bases, 1.9Gb downloads
   Accession: ERX177564

Transcriptome Analysis of Chlamydomonas reinhardtii

Filter your results:

- All (346)
- access: Controlled (0)
- access: Public (346)
- aligned data (0)
- source: DNA (38)
- source: metagenomic (1)
- source: RNA (302)
- type: exome (0)
- type: genome (31)

Manage Filters

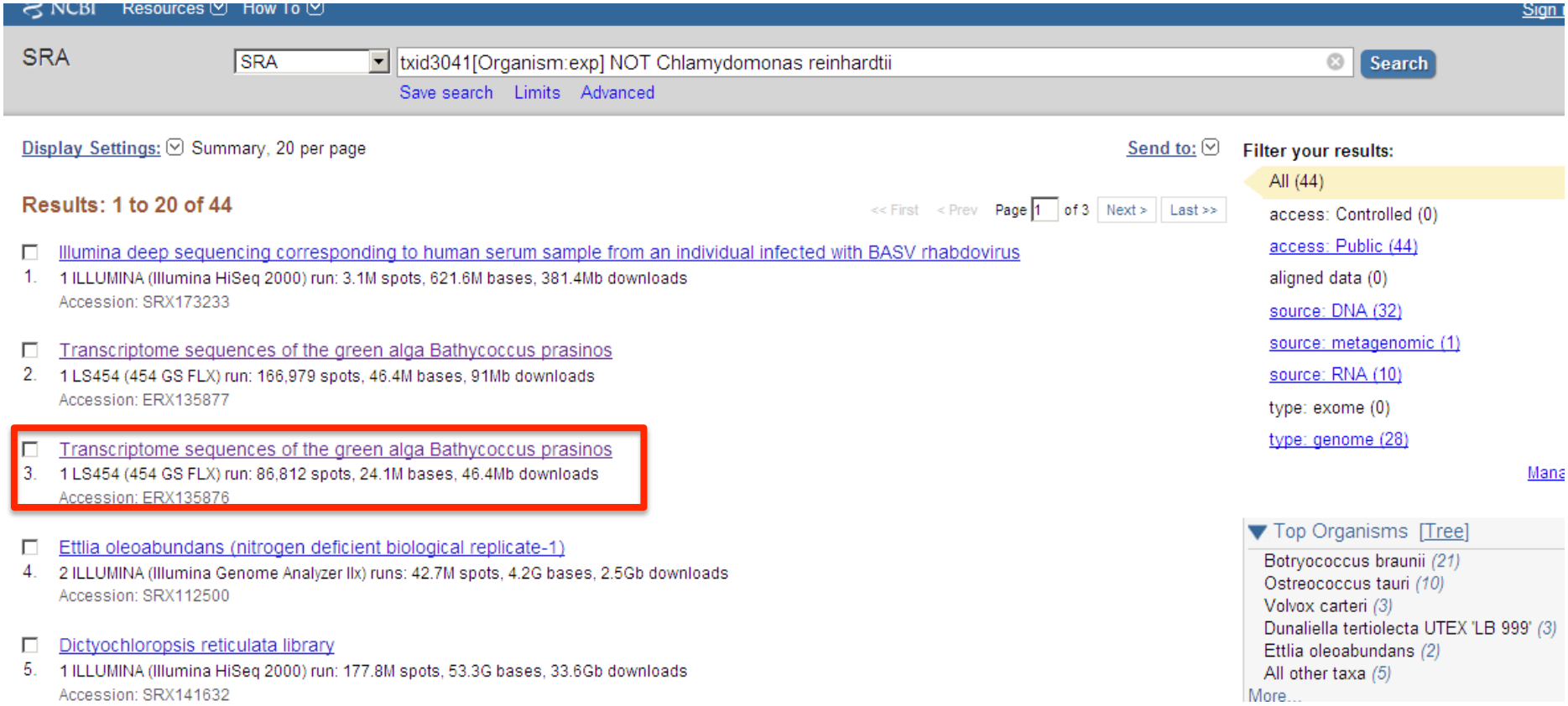
Top Organisms [Tree]

- Chlamydomonas reinhardtii *(302)*
- Botryococcus braunii *(21)*
- Ostreococcus tauri *(10)*
- Volvox carteri *(3)*
- Dunaliella tertiolecta UTEX 'LB 999' *(3)*
- All other taxa *(7)*

More...

Search in related databases

Database	Access	all


NCBI Resources How To
Sign
SRA
SRA
txid3041[Organism:exp] NOT Chlamydomonas reinhardtii
Search
Save search Limits Advanced
Display Settings: Summary, 20 per page
Send to:
Filter your results:
All (44)
access: Controlled (0)
access: Public (44)
aligned data (0)
source: DNA (32)
source: metagenomic (1)
source: RNA (10)
type: exome (0)
type: genome (28)
Mana
Results: 1 to 20 of 44
<< First < Prev Page 1 of 3 Next > Last >>
Illumina deep sequencing corresponding to human serum sample from an individual infected with BASV rhabdovirus
1. 1 ILLUMINA (Illumina HiSeq 2000) run: 3.1M spots, 621.6M bases, 381.4Mb downloads
Accession: SRX173233
Transcriptome sequences of the green alga Bathycoccus prasinos
2. 1 LS454 (454 GS FLX) run: 166,979 spots, 46.4M bases, 91Mb downloads
Accession: ERX135877
Transcriptome sequences of the green alga Bathycoccus prasinos
3. 1 LS454 (454 GS FLX) run: 86,812 spots, 24.1M bases, 46.4Mb downloads
Accession: ERX135876
Ettlia oleoabundans (nitrogen deficient biological replicate-1)
4. 2 ILLUMINA (Illumina Genome Analyzer IIx) runs: 42.7M spots, 4.2G bases, 2.5Gb downloads
Accession: SRX112500
Dictyochloropsis reticulata library
5. 1 ILLUMINA (Illumina HiSeq 2000) run: 177.8M spots, 53.3G bases, 33.6Gb downloads
Accession: SRX141632
Top Organisms [Tree]
Botryococcus braunii (21)
Ostreococcus tauri (10)
Volvox carteri (3)
Dunaliella tertiolecta UTEX 'LB 999' (3)
Ettlia oleoabundans (2)
All other taxa (5)
More...

Display Settings: Full

Send to:

Transcriptome sequences of the green alga Bathycoccus prasinos

Accession: ERX135876
Experiment design: These expression data correspond to a culture in exponential phase of the green alga Bathycoccus prasinos
Submission: ERA148021 by GSC
Study summary: Gene functionalities and genome structure in Bathycoccus prasinos reflect cellular specializations at the base of the green lineage (ERP001629) • Study • All experiments (more...)
Sample: Culture in exponential phase of the green alga Bathycoccus prasinos (ERS158202) (more...)
Library: YKBODS (more...)
Platform: LS454 (more...)
Spot descriptor:

Total: 1 run, 86,812 spots, 24.1M bases, 46.4Mb

#	Run	# of Spots	# of Bases	Size
1	ERR159934	86,812	24.1M	46.4Mb

Related inform
BioProject
BioSample
Taxonomy

Recent activity
txid3041[Or
Chlamydom
txid3041[Or
Chlorophyta
green algae

Next lecture

NCBI resources II: tools and ftp resources