Use of sequences to determine phylogenetic relationships

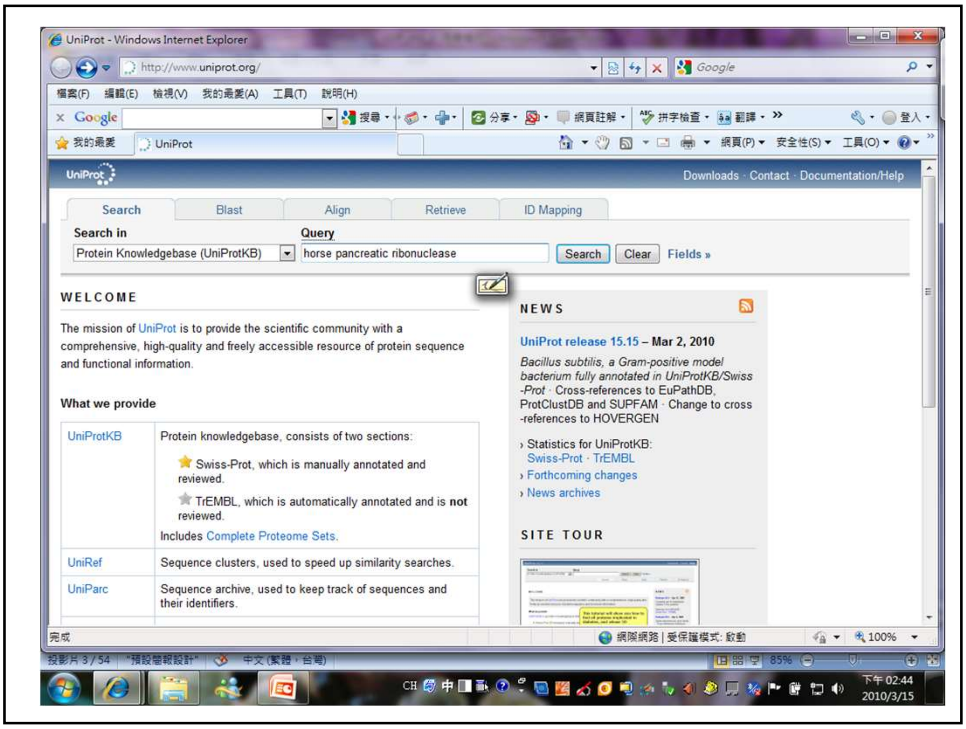
- Example 1.1 Retrieve the amino acid sequence of horse pancreatic ribonuclease
- ExPASy server: http://www.expasy.org/
- NCBI server: http://www.ncbi.nlm.nih.gov/
- Type the keywords: **horse pancreatic ribonuclease (胰臟，核醣核酸脢)**
- Select RNAS1_HORSE and then find FASTA format

Query all databases

horse pancreatic ribonuclease

Detected query type: text

Resource	Hits	Category	Comment
UniProtKB	13 hits	pr	UniProtKB entries
STRING	0 hits	pr	
SWISS-MODEL Repository	0 hits	bi, pr, st	
PROSITE	0 hits	pr	PROSITE documentation entri
ViralZone	0 hits	pr	ViralZone pages (for given viru
OMA	1 hit	ge, ph	
ENZYME	0 hits	pr	ENZYME entries



Search in: Protein Knowledgebase (UniProtKB)

Query: horse pancreatic ribonuclease

WELCOME

The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What we provide

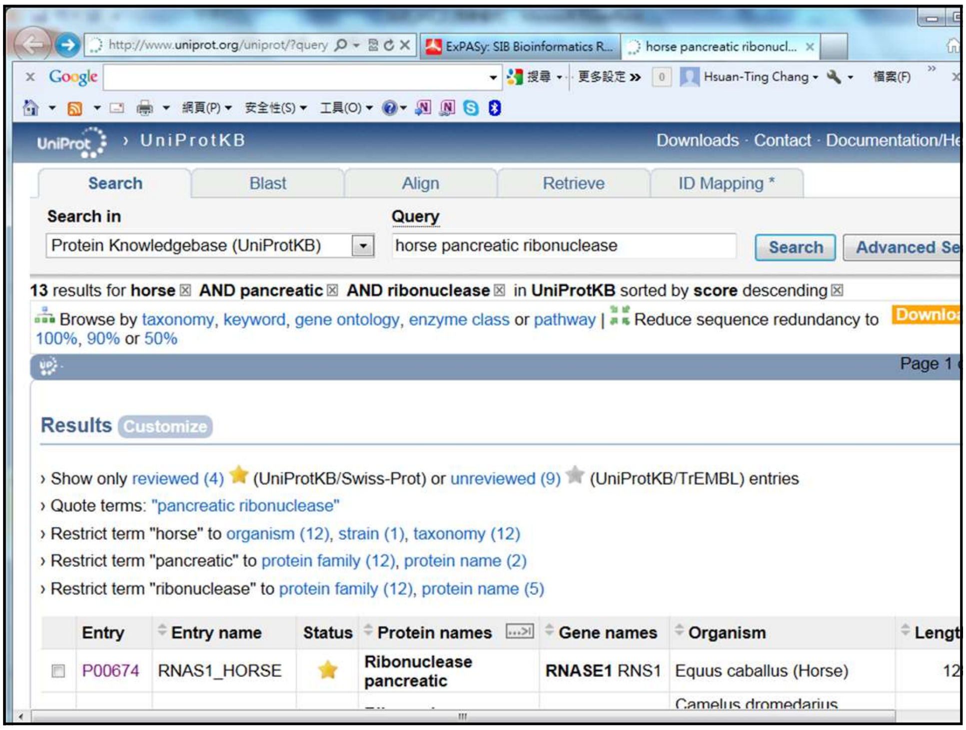
UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is **not** reviewed. Includes Complete Proteome Sets.
UniRef	Sequence clusters, used to speed up similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.

NEWS

UniProt release 15.15 – Mar 2, 2010

Bacillus subtilis, a Gram-positive model bacterium fully annotated in UniProtKB/Swiss-Prot · Cross-references to EuPathDB, ProtClustDB and SUPFAM · Change to cross-references to HOVERGEN

- › Statistics for UniProtKB: Swiss-Prot · TrEMBL
- › Forthcoming changes
- › News archives



Search in: Protein Knowledgebase (UniProtKB)

Query: horse pancreatic ribonuclease

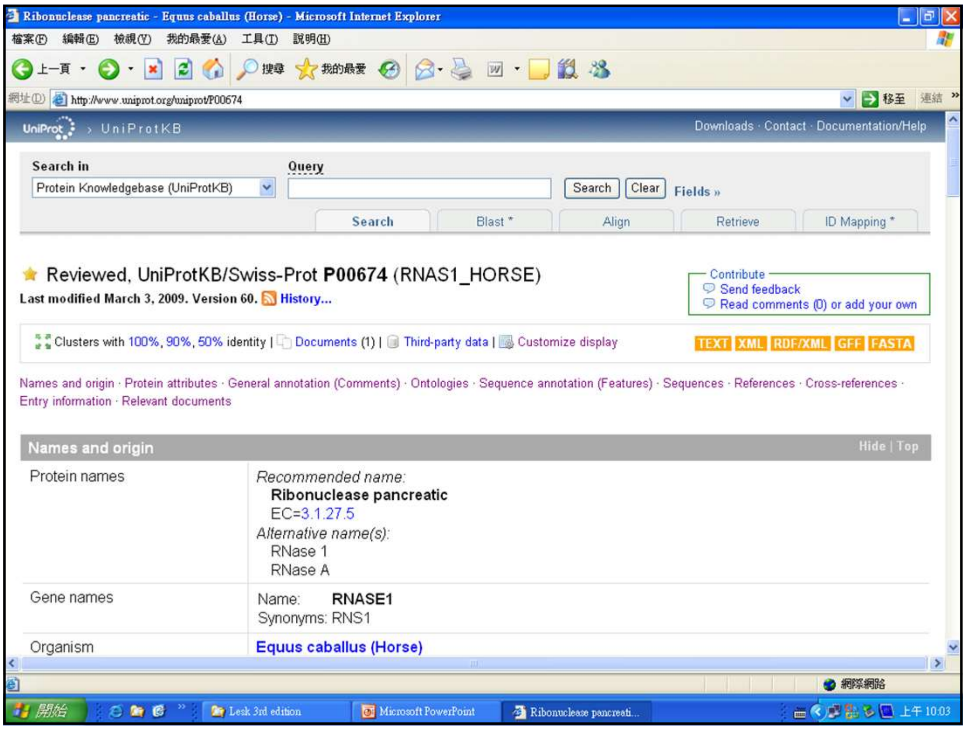
13 results for **horse** ☒ **AND pancreatic** ☒ **AND ribonuclease** ☒ in **UniProtKB** sorted by **score** descending ☒

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50%

Results

- › Show only reviewed (4) ★ (UniProtKB/Swiss-Prot) or unreviewed (9) ★ (UniProtKB/TrEMBL) entries
- › Quote terms: "pancreatic ribonuclease"
- › Restrict term "horse" to organism (12), strain (1), taxonomy (12)
- › Restrict term "pancreatic" to protein family (12), protein name (2)
- › Restrict term "ribonuclease" to protein family (12), protein name (5)

	Entry	Entry name	Status	Protein names	Gene names	Organism	Lengt
☐	P00674	RNAS1_HORSE	★	**Ribonuclease pancreatic**	**RNASE1** RNS1	Equus caballus (Horse)	12
						Camelus dromedarius	



Reviewed, UniProtKB/Swiss-Prot **P00674** (RNAS1_HORSE)

Last modified March 3, 2009. Version 60. History...

Contribute
Send feedback
Read comments (0) or add your own

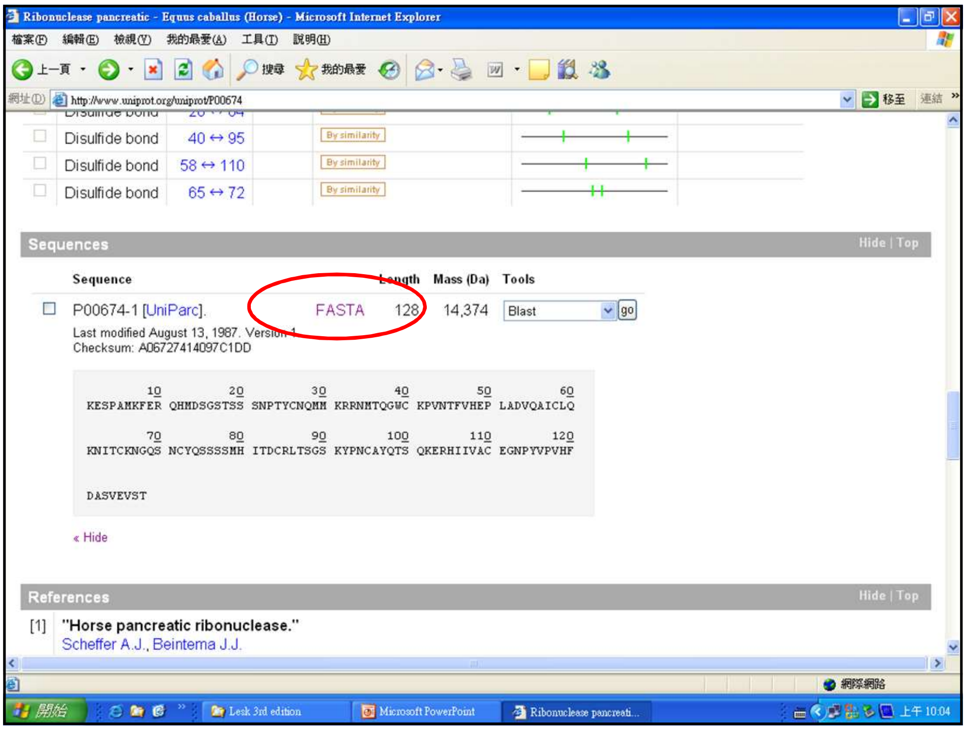
Clusters with 100%, 90%, 50% identity | Documents (1) | Third-party data | Customize display

TEXT XML RDF/XML GFF FASTA

Names and origin · Protein attributes · General annotation (Comments) · Ontologies · Sequence annotation (Features) · Sequences · References · Cross-references · Entry information · Relevant documents

Names and origin Hide | Top

Protein names	*Recommended name:* **Ribonuclease pancreatic** EC=3.1.27.5 *Alternative name(s):* RNase 1 RNase A
Gene names	Name: **RNASE1** Synonyms: RNS1
Organism	**Equus caballus (Horse)**



☐	Disulfide bond	40 ↔ 95	By similarity
☐	Disulfide bond	58 ↔ 110	By similarity
☐	Disulfide bond	65 ↔ 72	By similarity

Sequences Hide | Top

	Sequence		Length	Mass (Da)	Tools
☐	P00674-1 [UniParc]. Last modified August 13, 1987. Version 1 Checksum: A06727414097C1DD	FASTA	128	14,374	Blast go

```
        10         20         30         40         50         60
KESPAMKFER QHMDSGSTSS SNPTYCNQMM KRRNMTQGWC KPVNTFVHEP LADVQAICLQ

        70         80         90        100        110        120
KNITCKNGQS NCYQSSSSMH ITDCRLTSGS KYPNCAYQTS QKERHIIVAC EGNPYVPVHF


DASVEVST
```

« Hide

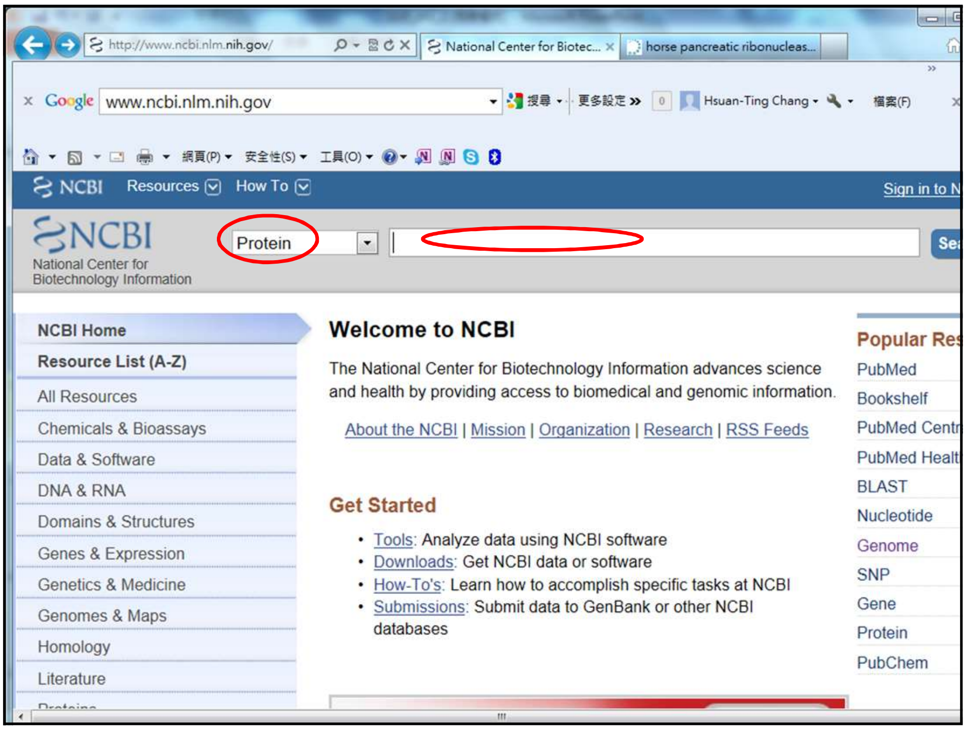
References Hide | Top

[1] **"Horse pancreatic ribonuclease."**
Scheffer A.J., Beintema J.J.

FASTA format

>sp|P00674|RNAS1_HORSE Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1) (RNase A) - Equus caballus (Horse).

KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMT
QGWCKPVNTFVHEPLADVQAICLQKNITCKNGQSN
CYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIV
ACEGNPYVPVHFDASVEVST



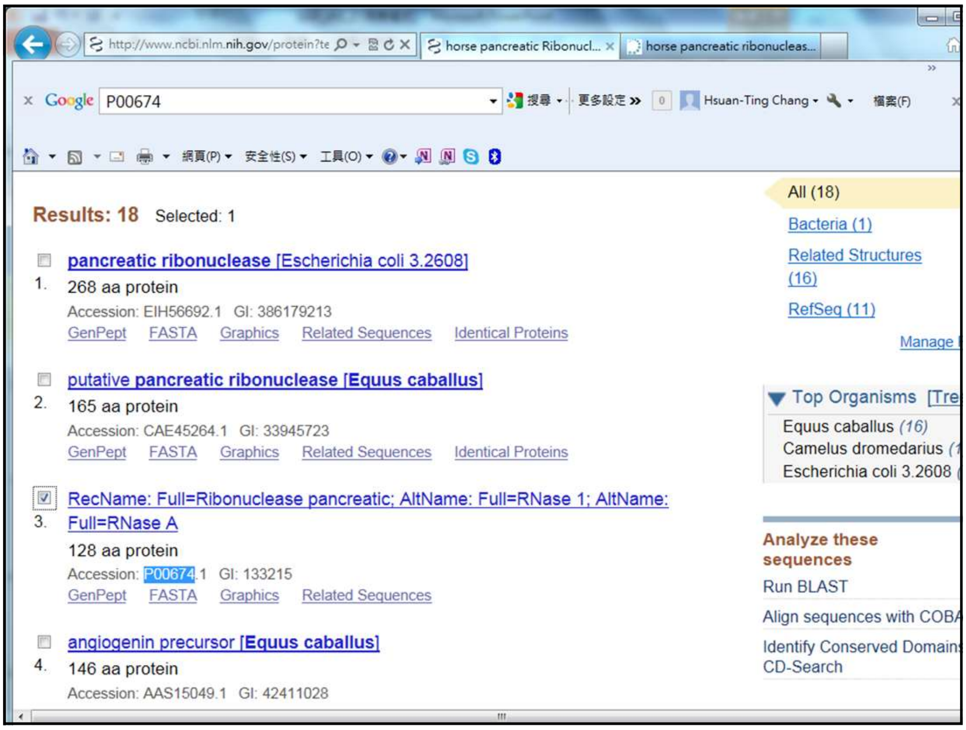
Welcome to NCBI

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



Results: 18 Selected: 1

☐ 1. pancreatic ribonuclease [Escherichia coli 3.2608]
268 aa protein
Accession: EIH56692.1 GI: 386179213
GenPept FASTA Graphics Related Sequences Identical Proteins

☐ 2. putative pancreatic ribonuclease [Equus caballus]
165 aa protein
Accession: CAE45264.1 GI: 33945723
GenPept FASTA Graphics Related Sequences Identical Proteins

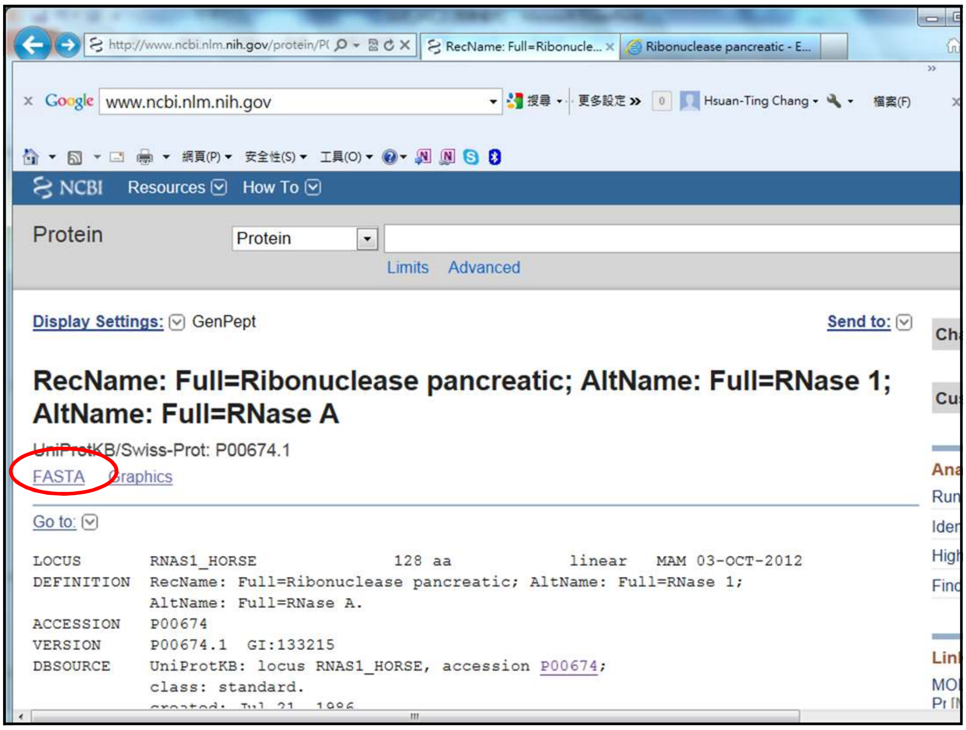
☑ 3. RecName: Full=Ribonuclease pancreatic; AltName: Full=RNase 1; AltName: Full=RNase A
128 aa protein
Accession: P00674.1 GI: 133215
GenPept FASTA Graphics Related Sequences

☐ 4. angiogenin precursor [Equus caballus]
146 aa protein
Accession: AAS15049.1 GI: 42411028

All (18)
Bacteria (1)
Related Structures (16)
RefSeq (11)
Manage

Top Organisms [Tre
Equus caballus (16)
Camelus dromedarius (
Escherichia coli 3.2608 (

Analyze these sequences
Run BLAST
Align sequences with COBA
Identify Conserved Domains CD-Search



Display Settings: GenPept

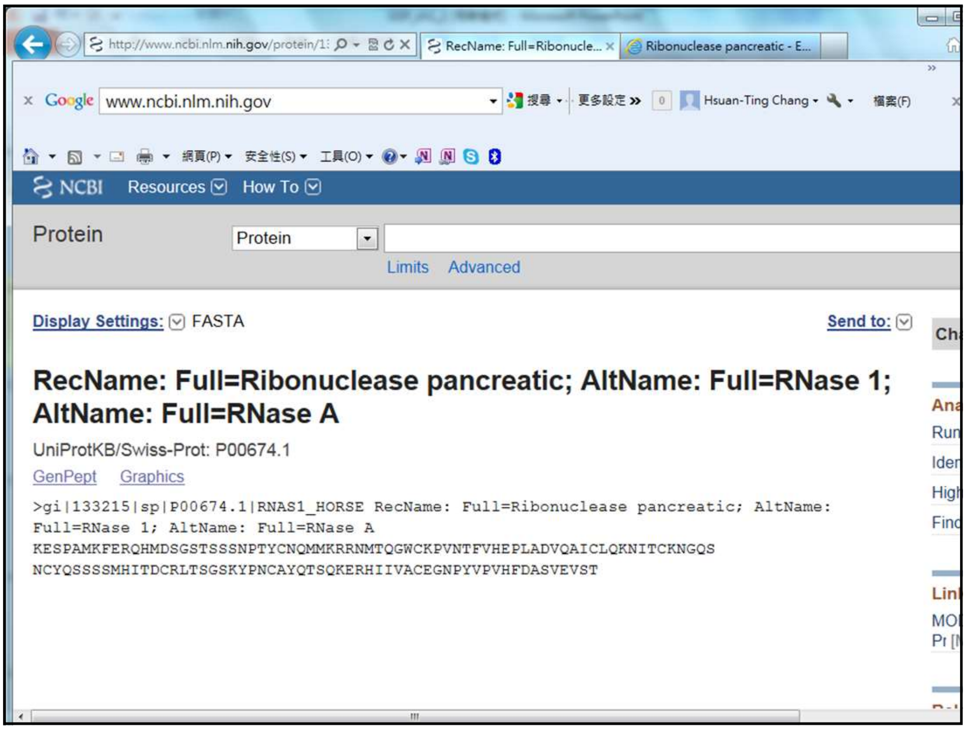
RecName: Full=Ribonuclease pancreatic; AltName: Full=RNase 1; AltName: Full=RNase A

UniProtKB/Swiss-Prot: P00674.1

FASTA Graphics

Go to:

```
LOCUS       RNAS1_HORSE              128 aa            linear   MAM 03-OCT-2012
DEFINITION  RecName: Full=Ribonuclease pancreatic; AltName: Full=RNase 1;
            AltName: Full=RNase A.
ACCESSION   P00674
VERSION     P00674.1  GI:133215
DBSOURCE    UniProtKB: locus RNAS1_HORSE, accession P00674;
            class: standard.
```



RecName: Full=Ribonuclease pancreatic; AltName: Full=RNase 1; AltName: Full=RNase A

UniProtKB/Swiss-Prot: P00674.1

GenPept Graphics

```
>gi|133215|sp|P00674.1|RNAS1_HORSE RecName: Full=Ribonuclease pancreatic; AltName:
Full=RNase 1; AltName: Full=RNase A
KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAICLQKNITCKNGQS
NCYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIVACEGNPYVPVHFDASVEVST
```

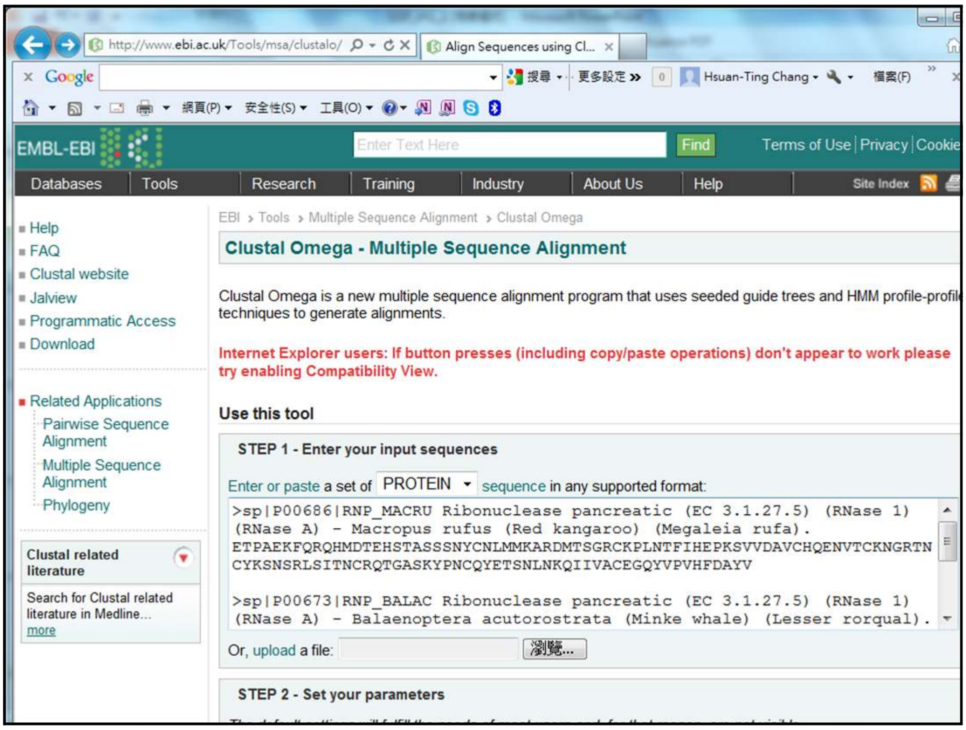
Example 1.2

- Determine, from the sequences of pancreatic ribonuclease from *horse*, *minke whale* (小鬚鯨) and *red kangaroo*, which two of these species are most closely related.
- Retrieve the amino acid sequences from the database and use the ***multiple sequence alignment*** (MSA) tool **CLUSTAL-Omega**
  - http://www.ebi.ac.uk/Tools/msa/
  - New MSA tool that uses seeded guide trees and HMM profile-profile techniques to generate alignments. Suitable for medium-large alignments
- Clustal Omega is a multiple sequence alignment program for proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. volutionary relationships can be seen via viewing Cladograms or Phylograms.

- >sp|P00686|RNP_MACRU ***Ribonuclease pancreatic*** (EC 3.1.27.5) (RNase 1) (RNase A) - Macropus rufus (Red kangaroo) (Megaleia rufa).
- ETPAEKFQRQHMDTEHSTASSSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQENVTCKNGRTNCYKSNSRLSITNCRQTGASKYPNCQYETSNLNKQIIVACEGQYVPVHFDAYV

- >sp|P00673|RNP_BALAC ***Ribonuclease pancreatic*** (EC 3.1.27.5) (RNase 1) (RNase A) - Balaenoptera acutorostrata (Minke whale) (Lesser rorqual).
- RESPAMKFQRQHMDSGNSPGNNPNYCNQMMMRRKMTQGRCKPVNTFVHESLEDVKAVCSQKNVLCKNGRTNCYESNSTMHITDCRQTGSSKYPNCAYKTSQKEKHIIVACEGNPYVPVHFDNSV

- >sp|P00674|RNP_HORSE ***Ribonuclease pancreatic*** (EC 3.1.27.5) (RNase 1) (RNase A) - Equus caballus (Horse).
- KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAICLQKNITCKNGQSNCYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIVACEGNPYVPVHFDASVEVST



Clustal Omega - Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profil techniques to generate alignments.

Internet Explorer users: If button presses (including copy/paste operations) don't appear to work please try enabling Compatibility View.

Use this tool

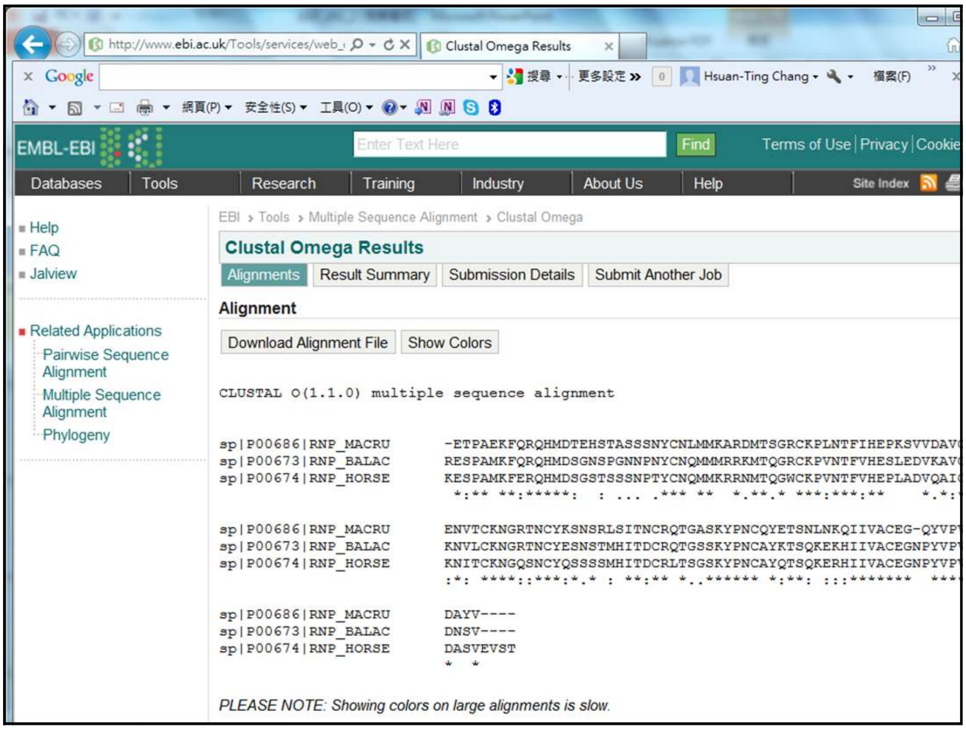
STEP 1 - Enter your input sequences

Enter or paste a set of PROTEIN sequence in any supported format:

```
>sp|P00686|RNP_MACRU Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1)
(RNase A) - Macropus rufus (Red kangaroo) (Megaleia rufa).
ETPAEKFQRQHMDTEHSTASSSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQENVTCKNGRTN
CYKSNSRLSITNCRQTGASKYPNCQYETSNLNKQIIVACEGQYVPVHFDAYV

>sp|P00673|RNP_BALAC Ribonuclease pancreatic (EC 3.1.27.5) (RNase 1)
(RNase A) - Balaenoptera acutorostrata (Minke whale) (Lesser rorqual).
```

Or, upload a file:

STEP 2 - Set your parameters



Clustal Omega Results

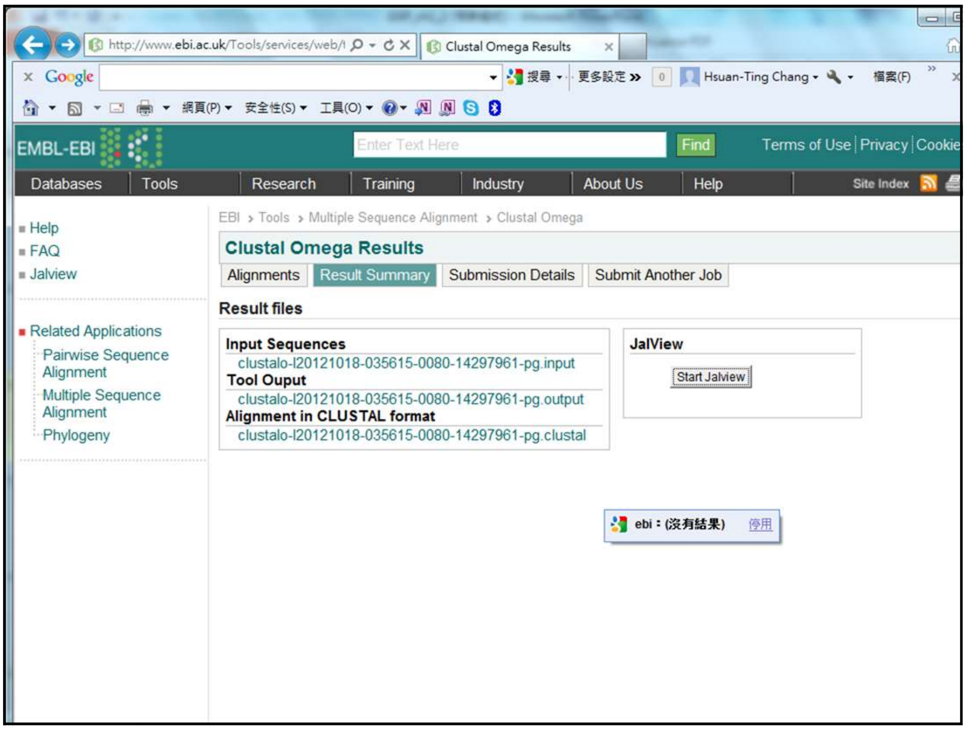
Alignments | Result Summary | Submission Details | Submit Another Job

Alignment

Download Alignment File | Show Colors

```
CLUSTAL O(1.1.0) multiple sequence alignment


sp|P00686|RNP_MACRU      -ETPAEKFQRQHMDTEHSTASSSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAV
sp|P00673|RNP_BALAC      RESPAMKFQRQHMDSGNSPGNNPNYCNQMMMRRKMTQGRCKPVNTFVHESLEDVKAV
sp|P00674|RNP_HORSE      KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAI
                          *:** **:*****:  : ... .*** **  *.**.* ***:***:**    *.*:

sp|P00686|RNP_MACRU      ENVTCKNGRTNCYKSNSRLSITNCRQTGASKYPNCQYETSNLNKQIIVACEG-QYVP
sp|P00673|RNP_BALAC      KNVLCKNGRTNCYESNSTMHITDCRQTGSSKYPNCAYKTSQKEKHIIVACEGNPYVP
sp|P00674|RNP_HORSE      KNITCKNGQSNCYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIVACEGNPYVP
                         :*: ****::***:*.* : **:** *..****** *:**: :::*******  ***

sp|P00686|RNP_MACRU      DAYV----
sp|P00673|RNP_BALAC      DNSV----
sp|P00674|RNP_HORSE      DASVEVST
                         *  *
```

PLEASE NOTE: Showing colors on large alignments is slow.



Clustal Omega Results

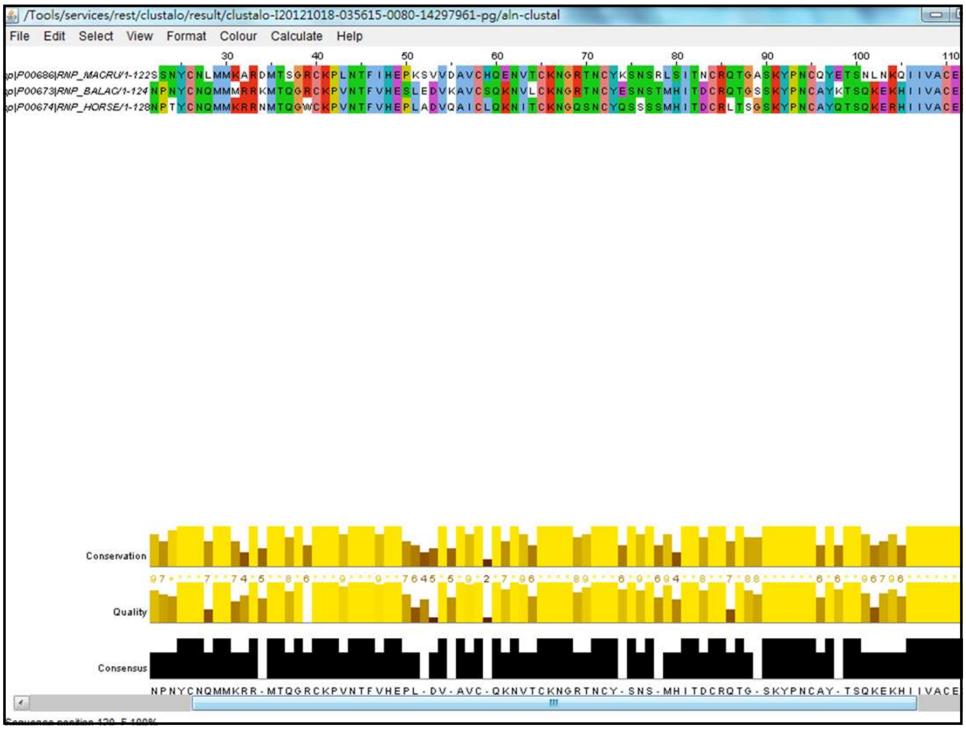
Alignments | Result Summary | Submission Details | Submit Another Job

Result files

Input Sequences
clustalo-I20121018-035615-0080-14297961-pg.input
Tool Ouput
clustalo-I20121018-035615-0080-14297961-pg.output
Alignment in CLUSTAL format
clustalo-I20121018-035615-0080-14297961-pg.clustal

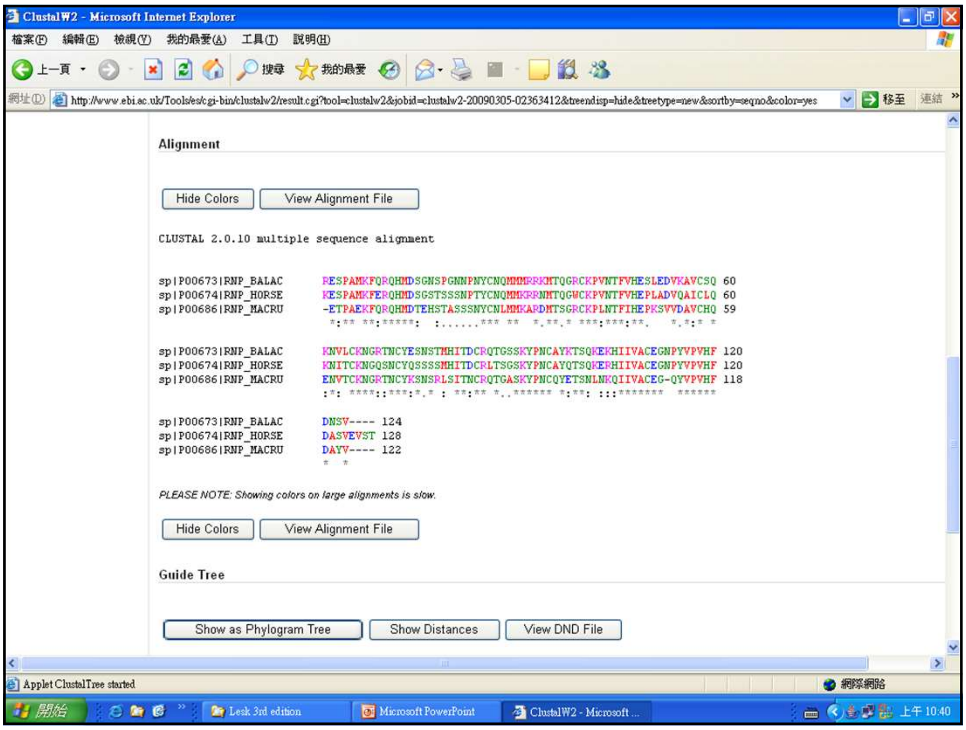
JalView

Start Jalview



```
                         30        40        50        60        70        80        90        100       110
p|P00686|RNP_MACRU/1-122 SSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQENVTCKNGRTNCYKSNSRLSITNCRQTGASKYPNCQYETSNLNKQIIVACE
p|P00673|RNP_BALAC/1-124 NPNYCNQMMMRRKMTQGRCKPVNTFVHESLEDVKAVCSQKNVLCKNGRTNCYESNSTMHITDCRQTGSSKYPNCAYKTSQKEKHIIVACE
p|P00674|RNP_HORSE/1-128 NPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAICLQKNITCKNGQSNCYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIVACE

97+***7**74*5**8*6***9***9**7645*5*9*2*7*96****89***6*9*694**8**7*88******6*6**96796******

NPNYCNQMMKRR-MTQGRCKPVNTFVHEPL-DV-AVC-QKNVTCKNGRTNCY-SNS-MHITDCRQTG-SKYPNCAY-TSQKEKHIIVACE
```

Sequence position 120, F 100%



Alignment

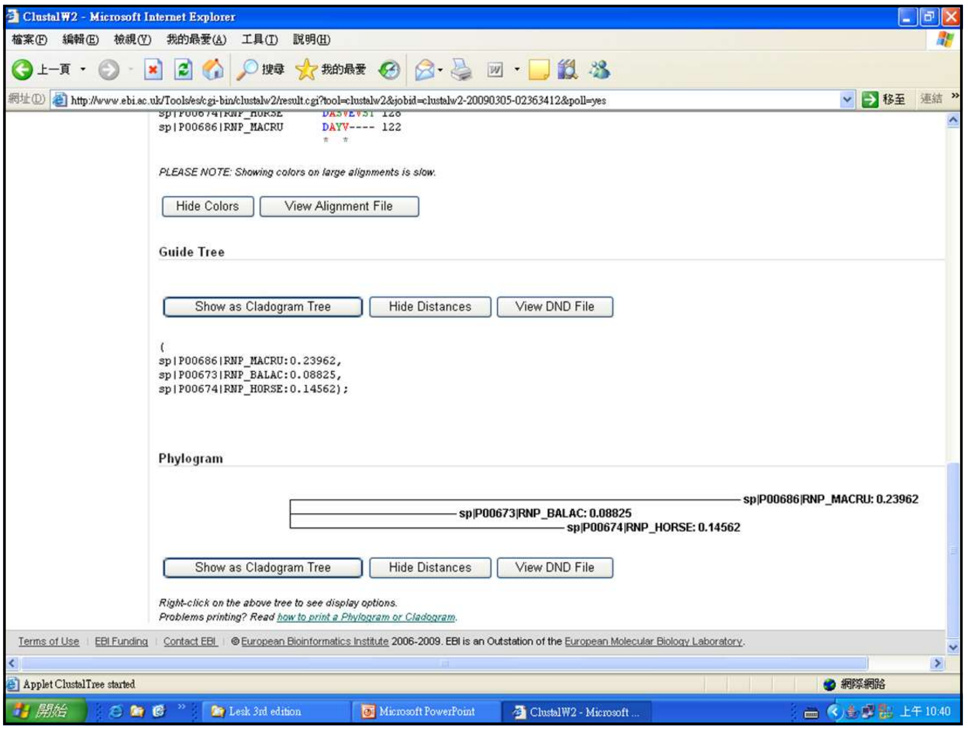
```
CLUSTAL 2.0.10 multiple sequence alignment


sp|P00673|RNP_BALAC      RESPAMKFQRQHMDSGNSPGNNPNYCNQMMMRRKMTQGRCKPVNTFVHESLEDVKAVCSQ 60
sp|P00674|RNP_HORSE      KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAICLQ 60
sp|P00686|RNP_MACRU      -ETPAEKFQRQHMDTEHSTASSSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQ 59
                          *:** **:*****:  :......*** **  *.**.* ***:***:**.   *.*:* *

sp|P00673|RNP_BALAC      KNVLCKNGRTNCYESNSTMHITDCRQTGSSKYPNCAYKTSQKEKHIIVACEGNPYVPVHF 120
sp|P00674|RNP_HORSE      KNITCKNGQSNCYQSSSSMHITDCRLTSGSKYPNCAYQTSQKERHIIVACEGNPYVPVHF 120
sp|P00686|RNP_MACRU      ENVTCKNGRTNCYKSNSRLSITNCRQTGASKYPNCQYETSNLNKQIIVACEG-QYVPVHF 118
                         :*: ****::***:*.* : **:** *..****** *:**: :::*******  ******

sp|P00673|RNP_BALAC      DNSV---- 124
sp|P00674|RNP_HORSE      DASVEVST 128
sp|P00686|RNP_MACRU      DAYV---- 122
                         *  *
```

PLEASE NOTE: Showing colors on large alignments is slow.

Guide Tree

```
p|P00673|RNP_BALAC/1-124 RESPAMKFQRQHMDSGNSPGNNPNYCNQMMMRRKMTQGRCKPVNTFVHESLEDVKAVCSQKNVLCKNGRTNCYESNSTMHITDCRQTGSSKYPNC
p|P00674|RNP_HORSE/1-128 KESPAMKFERQHMDSGSTSSSNPTYCNQMMKRRNMTQGWCKPVNTFVHEPLADVQAICLQKNITCKNGQSNCYQSSSSMHITDCRLTSGSKYPNC
p|P00686|RNP_MACRU/1-122 -ETPAEKFQRQHMDTEHSTASSSNYCNLMMKARDMTSGRCKPLNTFIHEPKSVVDAVCHQENVTCKNGRTNCYKSNSRLSITNCRQTGASKYPNC

-ESPAMKFQRQHMDSG-S--SNPNYCNQMMKRR-MTQGRCKPVNTFVHEPL-DV-AVC-QKNVTCKNGRTNCY-SNS-MHITDCRQTG-SKYPNC
```

Sequence 3 ID: sp|P00686|RNP_MACRU Residue: LEU (78)



```
sp|P00674|RNP_HORSE       DASVEVST 128
sp|P00686|RNP_MACRU       DAYV---- 122
                          *  *
```

PLEASE NOTE: Showing colors on large alignments is slow.

Guide Tree

```
(
sp|P00686|RNP_MACRU:0.23962,
sp|P00673|RNP_BALAC:0.08825,
sp|P00674|RNP_HORSE:0.14562);
```

Phylogram

sp|P00686|RNP_MACRU: 0.23962
sp|P00673|RNP_BALAC: 0.08825
sp|P00674|RNP_HORSE: 0.14562

Right-click on the above tree to see display options.
Problems printing? Read how to print a Phylogram or Cladogram.

© European Bioinformatics Institute 2006-2009. EBI is an Outstation of the European Molecular Biology Laboratory.

Alignment Result

- Horse and whale share the most identical residues. The result appears significant, and therefore confirm our expectation.
  - Knowing that horse and whale are placental (有胎盤的) mammals and kangaroo is a marsupial (有袋動物), we expect horse and whale to be closer pair.

FASTA format description

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. It is recommended that all lines of text be shorter than 80 characters in length. An example sequence in FASTA format is:

```
>gi|532319|pir|TVFV2E|TVFV2E envelope protein
ELRLRYCAPAGFALLKCNDADYDGFKTNCSNVSVVHCTNLMNTTVTTGLLLNGSYSENRT
QIWQKHRTSNDSALILLNKHYNLTVTCKRPGNKTVLPVTIMAGLVFHSQKYNLRLRQAWC
HFPSNWKGAWKEVKEEIVNLPKERYRGTNDPKRIFFQRQWGDPETANLWFNCHGEFFYCK
MDWFLNYLNNLTVDADHNECKNTSGTKSGNKRAPGPCVQRTYVACHIRSVIIWLETISKK
TYAPPREGHLECTSTVTGMTVELNYIPKNRTNVTLSPQIESIWAAELDRYKLVEITPIGF
APTEVRRYTGGHERQKRVPFVXXXXXXXXXXXXXXXXXXXXXXVQSQHLLAGILQQQK
NL LAAVEAQQQMLKLTIWGVK
```

In bioinformatics, a **sequence alignment** is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix. Gaps are inserted between the residues so that identical or similar characters are aligned in successive columns.

Global and local alignments

```
Global  FTFTALILLAVAV
        F--TAL-LLA-AV

Local   FTFTALILL-AVAV
        --FTAL-LLAAV--
```

Homework #1

- Find the pancreatic ribonuclease of any other three species and perform following operations:
- 1. Perform multiple sequence alignment by use of CLUSTAL-W software
- 2. Find the corresponding DNA sequences.
- 3. Perform multiple sequence alignment for retrieved DNA sequences.

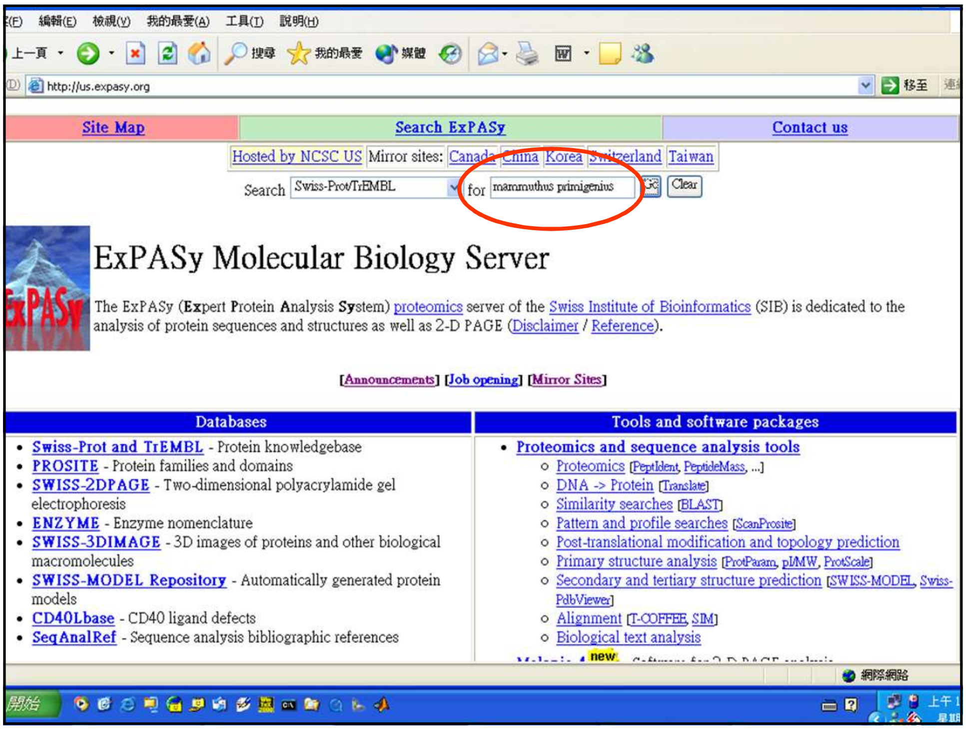
Case Study 1.5 – A harder one

- Two living genera of elephant: the African elephant and the Indian
- 藉由北極長毛象(Mammuthus primigenius)化石之粒腺體細胞色素b (mitochondrial cytochrome b)，界定出哪一種現代大象(Loxodonta africana and Elephas maximus)與遠古之長毛象有較近之血緣關係。
- From the results, it appears that mammoth is more closely related to African elephant. However, there are few differences. Are they significant?
- Questions:
  - Could we tell from these sequences alone that they are from closely related species?
  - Given the differences are small, do they represent evolutionary divergence arising from selection, or merely random noise or drift?

- >sp|P92658|CYB_MAMPR Cytochrome b OS=Mammuthus primigenius GN=MT-CYB PE=3 SV=3
- MTHIRKSHPLLKILNKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM
  TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL
  LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTDLVEWIWGGFSVDKATLNRFFA
  LHFILPFTMIALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILFLL
  LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALLLSILI
  LGIMPLLHTSKHRSMMLRPLSQVLFWTLATDLLMLTWIGSQPVEYPYIIIGQMASILYFS
  IILAFLPIAGMIENYLIK

- >sp|P24958|CYB_LOXAF Cytochrome b OS=Loxodonta africana GN=MT-CYB PE=3 SV=2
- MTHIRKSHPLLKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM
  TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL
  LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLVEWIWGGFSVDKATLNRFFA
  LHFILPFTMIALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILLLL
  LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALLLSILI
  LGLMPLLHTSKHRSMMLRPLSQVLFWTLTMDLLTLTWIGSQPVEYPYIIIGQMASILYFS
  IILAFLPIAGVIENYLIK

- >sp|O47885|CYB_ELEMA Cytochrome b OS=Elephas maximus GN=MT-CYB PE=3 SV=1
- MTHTRKFHPLFKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM
  TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL
  LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLVEWIWGGFSVDKATLNRFFA
  FHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILLLL
  LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSILI
  LGLMPLLHTSKHRSMMLRPLSQVLFWTLTMDLLTLTWIGSQPVEHPYIIIGQMASILYFS
  IILAFLPIAGMIENYLIK



ExPASy Molecular Biology Server

The ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / Reference).

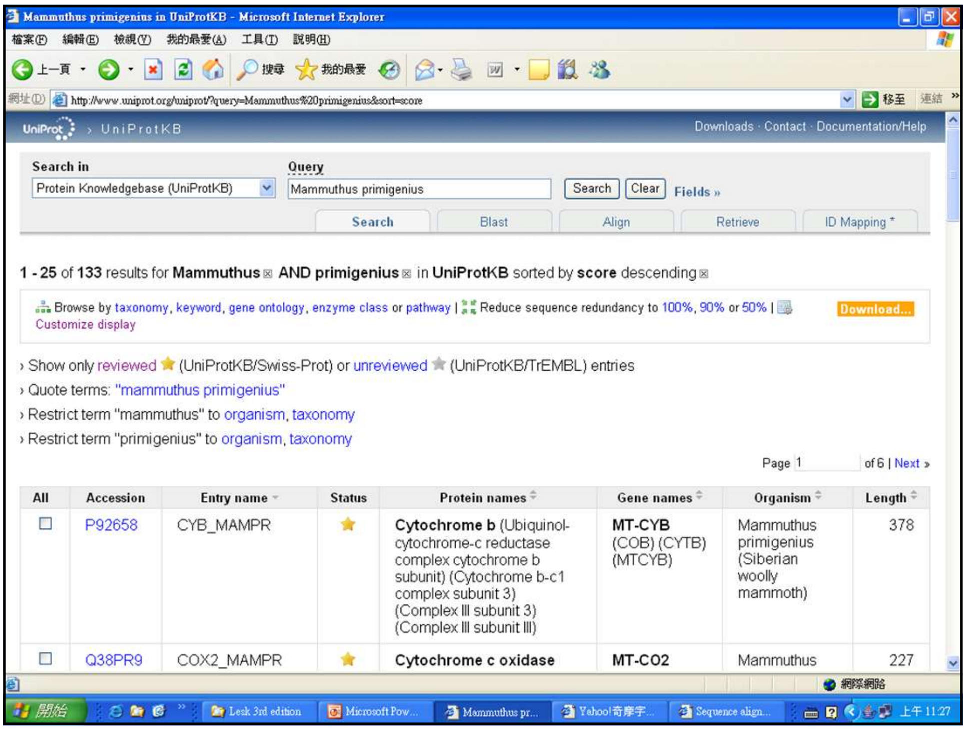
[Announcements] [Job opening] [Mirror Sites]

Databases

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

Tools and software packages

- **Proteomics and sequence analysis tools**
  - Proteomics [PeptIdent, PeptideMass, ...]
  - DNA -> Protein [Translate]
  - Similarity searches [BLAST]
  - Pattern and profile searches [ScanProsite]
  - Post-translational modification and topology prediction
  - Primary structure analysis [ProtParam, pI/MW, ProtScale]
  - Secondary and tertiary structure prediction [SWISS-MODEL, Swiss-PdbViewer]
  - Alignment [T-COFFEE, SIM]
  - Biological text analysis



Search in: Protein Knowledgebase (UniProtKB)

Query: Mammuthus primigenius

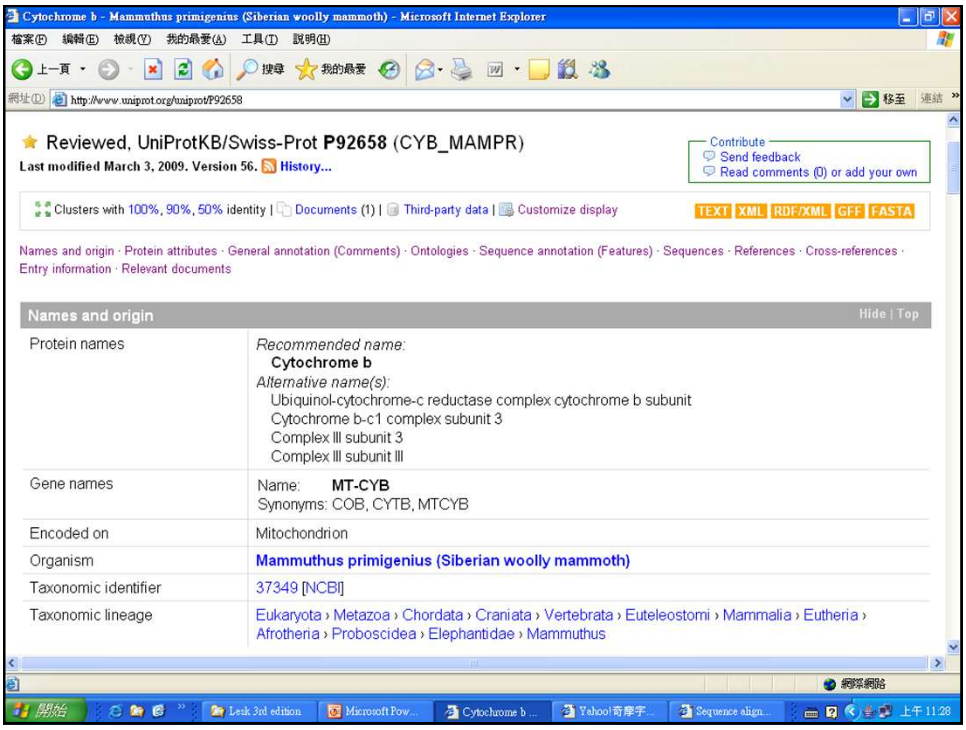
1 - 25 of **133** results for **Mammuthus** AND **primigenius** in **UniProtKB** sorted by **score** descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% |
Customize display

› Show only reviewed ★ (UniProtKB/Swiss-Prot) or unreviewed ★ (UniProtKB/TrEMBL) entries
› Quote terms: "mammuthus primigenius"
› Restrict term "mammuthus" to organism, taxonomy
› Restrict term "primigenius" to organism, taxonomy

Page 1 of 6 | Next »

All	Accession	Entry name	Status	Protein names	Gene names	Organism	Length
☐	P92658	CYB_MAMPR	★	**Cytochrome b** (Ubiquinol-cytochrome-c reductase complex cytochrome b subunit) (Cytochrome b-c1 complex subunit 3) (Complex III subunit 3) (Complex III subunit III)	**MT-CYB** (COB) (CYTB) (MTCYB)	Mammuthus primigenius (Siberian woolly mammoth)	378
☐	Q38PR9	COX2_MAMPR	★	**Cytochrome c oxidase**	**MT-CO2**	Mammuthus	227

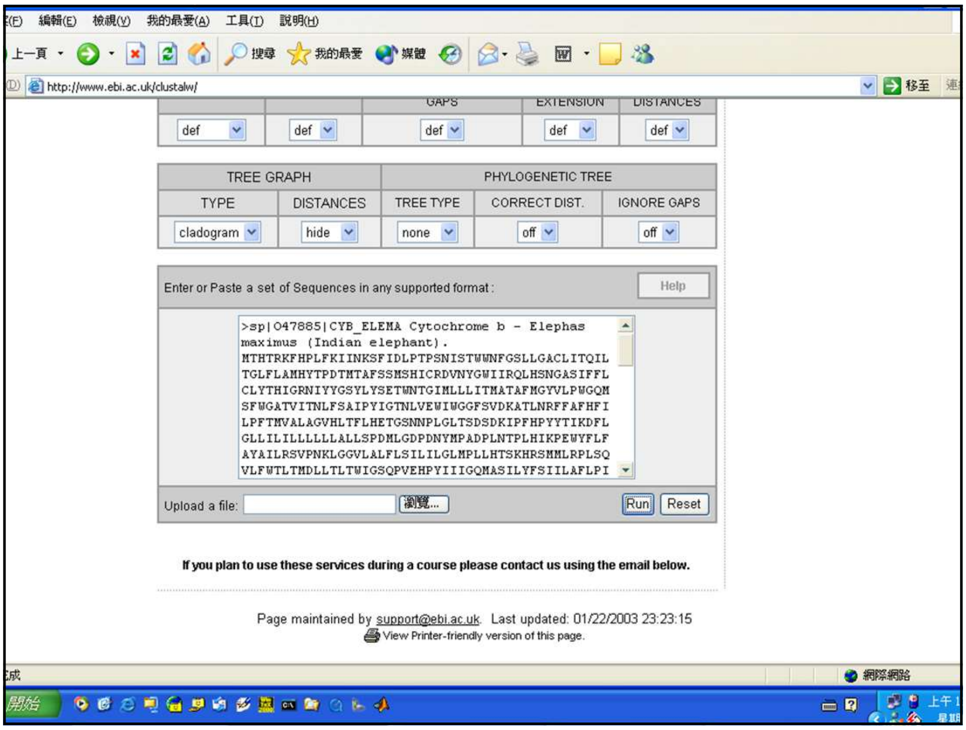


Reviewed, UniProtKB/Swiss-Prot **P92658** (CYB_MAMPR)

Last modified March 3, 2009. Version 56. History...

Names and origin

Protein names	*Recommended name:* **Cytochrome b** *Alternative name(s):* Ubiquinol-cytochrome-c reductase complex cytochrome b subunit Cytochrome b-c1 complex subunit 3 Complex III subunit 3 Complex III subunit III
Gene names	Name: **MT-CYB** Synonyms: COB, CYTB, MTCYB
Encoded on	Mitochondrion
Organism	**Mammuthus primigenius (Siberian woolly mammoth)**
Taxonomic identifier	37349 [NCBI]
Taxonomic lineage	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Mammalia › Eutheria › Afrotheria › Proboscidea › Elephantidae › Mammuthus



		GAPS	EXTENSION	DISTANCES
def	def	def	def	def

TREE GRAPH		PHYLOGENETIC TREE		
TYPE	DISTANCES	TREE TYPE	CORRECT DIST.	IGNORE GAPS
cladogram	hide	none	off	off

Enter or Paste a set of Sequences in any supported format :

```
>sp|O47885|CYB_ELEMA Cytochrome b - Elephas
maximus (Indian elephant).
MTHTRKFHPLFKIINKSFIDLPTPSNISTWWNFGSLLGACLITQIL
TGLFLAMHYTPDTMTAFSSMSHICRDVNYGWIIRQLHSNGASIFFL
CLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQM
SFWGATVITNLFSAIPYIGTNLVEWIWGGFSVDKATLNRFFAFHFI
LPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFL
GLLILILLLLLLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLF
AYAILRSVPNKLGGVLALFLSILILGLMPLLHTSKHRSMMLRPLSQ
VLFWTLTMDLLTLTWIGSQPVEHPYIIIGQMASILYFSIILAFLPI
```

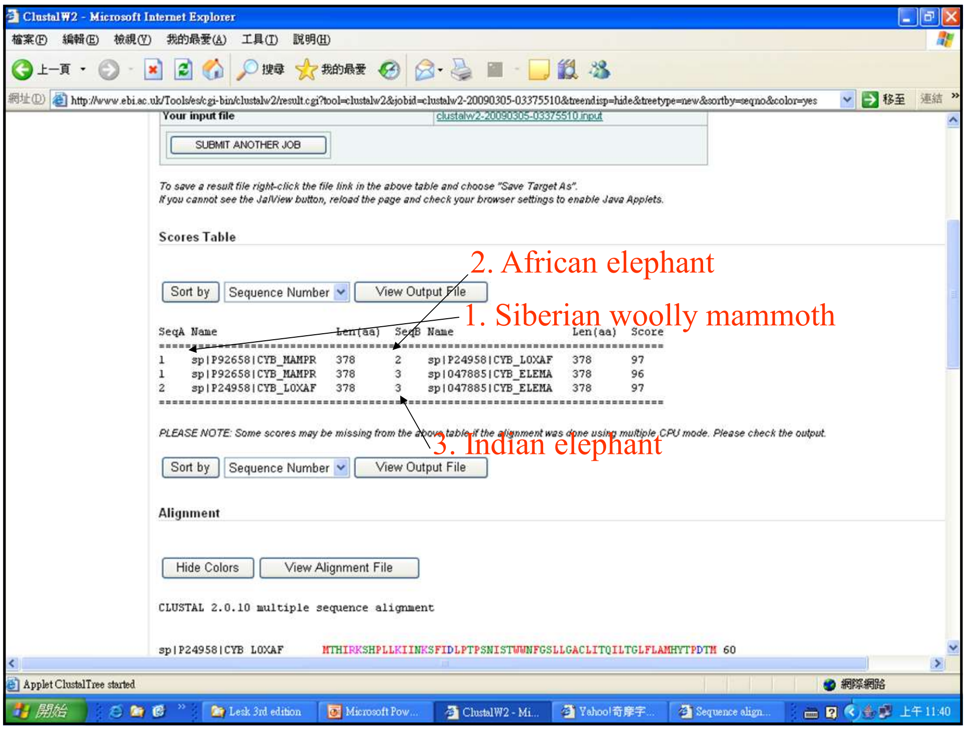
Upload a file:

If you plan to use these services during a course please contact us using the email below.

Page maintained by support@ebi.ac.uk. Last updated: 01/22/2003 23:23:15

View Printer-friendly version of this page.

```
CLUSTAL 2.0.10 multiple sequence alignment


sp|P24958|CYB_LOXAF      MTHIRKSHPLLKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM 60
sp|047885|CYB_ELEMA      MTHTRKFHPLFKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM 60
sp|P92658|CYB_MAMPR      MTHIRKSHPLLKILNKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM 60
                         *** ** ***:**:**********************************************

sp|P24958|CYB_LOXAF      TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL 120
sp|047885|CYB_ELEMA      TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL 120
sp|P92658|CYB_MAMPR      TAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLYTHIGRNIYYGSYLYSETWNTGIMLL 120
                         ************************************************************

sp|P24958|CYB_LOXAF      LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLVEWIWGGFSVDKATLNRFFA 180
sp|047885|CYB_ELEMA      LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLVEWIWGGFSVDKATLNRFFA 180
sp|P92658|CYB_MAMPR      LITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTDLVEWIWGGFSVDKATLNRFFA 180
                         **************************************:*********************

sp|P24958|CYB_LOXAF      LHFILPFTMIALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILLLL 240
sp|047885|CYB_ELEMA      FHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILLLL 240
sp|P92658|CYB_MAMPR      LHFILPFTMIALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLGLLILILFLL 240
                         :********:***********************************************:**

sp|P24958|CYB_LOXAF      LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALLLSILI 300
sp|047885|CYB_ELEMA      LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSILI 300
sp|P92658|CYB_MAMPR      LLALLSPDMLGDPDNYMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALLLSILI 300
                         ******************************************************:*****

sp|P24958|CYB_LOXAF      LGLMPLLHTSKHRSMMLRPLSQVLFWTLTMDLLTLTWIGSQPVEYPYIIIGQMASILYFS 360
sp|047885|CYB_ELEMA      LGLMPLLHTSKHRSMMLRPLSQVLFWTLTMDLLTLTWIGSQPVEHPYIIIGQMASILYFS 360
sp|P92658|CYB_MAMPR      LGIMPLLHTSKHRSMMLRPLSQVLFWTLATDLLMLTWIGSQPVEYPYIIIGQMASILYFS 360
                         **:*************************: *** **********:***************

sp|P24958|CYB_LOXAF      IILAFLPIAGVIENYLIK 378
sp|047885|CYB_ELEMA      IILAFLPIAGMIENYLIK 378
sp|P92658|CYB_MAMPR      IILAFLPIAGMIENYLIK 378
                         **********:*******
```



Your input file clustalw2-20090305-03375510.input

To save a result file right-click the file link in the above table and choose "Save Target As".
If you cannot see the JalView button, reload the page and check your browser settings to enable Java Applets.

Scores Table

2. African elephant

1. Siberian woolly mammoth

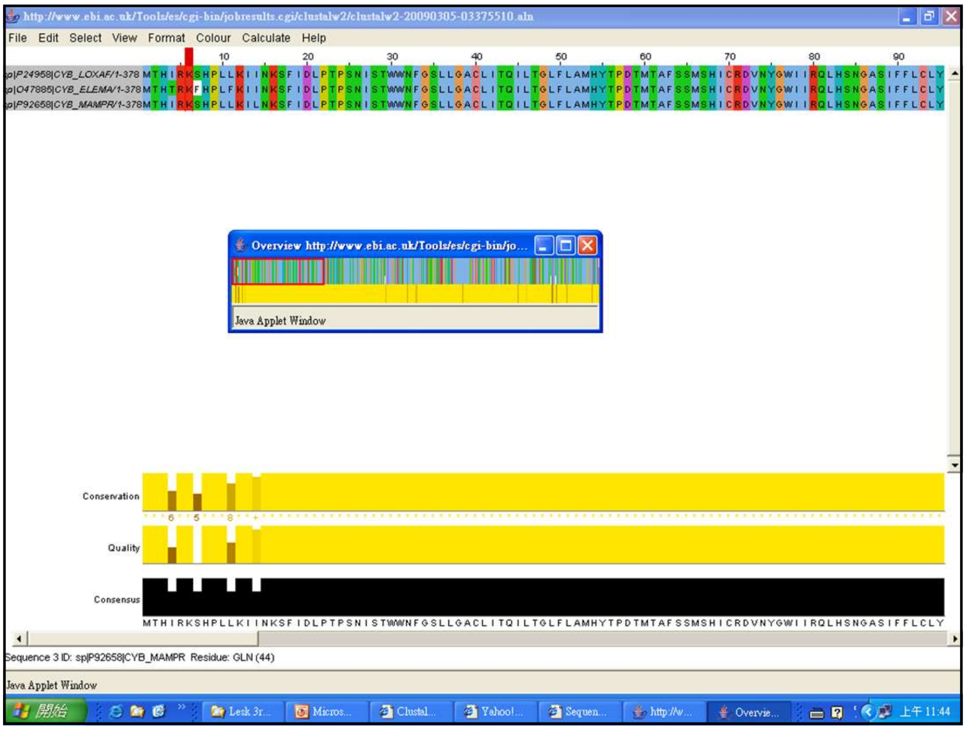
```
SeqA Name                 Len(aa)  SeqB Name                 Len(aa)  Score
===========================================================================
1    sp|P92658|CYB_MAMPR  378      2    sp|P24958|CYB_LOXAF  378      97
1    sp|P92658|CYB_MAMPR  378      3    sp|047885|CYB_ELEMA  378      96
2    sp|P24958|CYB_LOXAF  378      3    sp|047885|CYB_ELEMA  378      97
===========================================================================
```

PLEASE NOTE: Some scores may be missing from the above table if the alignment was done using multiple CPU mode. Please check the output.

3. Indian elephant

Alignment

```
CLUSTAL 2.0.10 multiple sequence alignment


sp|P24958|CYB_LOXAF      MTHIRKSHPLLKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTM 60
```



```
p|P24958|CYB_LOXAF/1-378 MTHIRKSHPLLKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTMTAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLY
p|O47885|CYB_ELEMA/1-378 MTHTRKFHPLFKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTMTAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLY
p|P92658|CYB_MAMPR/1-378 MTHIRKSHPLLKILNKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTMTAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLY

Consensus                MTHIRKSHPLLKIINKSFIDLPTPSNISTWWNFGSLLGACLITQILTGLFLAMHYTPDTMTAFSSMSHICRDVNYGWIIRQLHSNGASIFFLCLY
```

Sequence 3 ID: sp|P92658|CYB_MAMPR Residue: GLN (44)

Background: Similarity & Homology

- ***Similarity***: the observation or measurement of resemblance (相似點) and difference, independent of the source of the resemblance.
- ***Homology***: the sequences and the organisms in which they occur are descended from a common ancestor, with the implication that the similarities are shared ancestral characteristics.
  - Homology must be an ***inference*** from observations of similarity. Only a few special cases is homology directly observables.
- The need for thoughtful scientific judgment!

- Does the very high degree of similarity of the sequences justify the conclusion that they are homologous; or are there other explanations?
- A functional cytochrome b ***requires*** so many conserved residues
- It starts out from independent ancestors and that common selective pressures forced them to become similar??
- Indian elephant has evolved faster than that of the African elephant or the mammoth, accumulating more mutations
- They gained a common gene by transfer from an unrelated organism via a virus.

SINES & LINES

- Short and long interspersed nuclear elements – repetitive non-coding sequences that form large fractions of eukaryotic genomes. 30% of human chromosomal DNA, and over 50% of some higher plant genomes.
- Typically,
  - SINES are ~70 – 500 bps long, up to 10^6 copies may appear.
  - LINES may be up to 7000 bps long, up to 10^5 copies may appear.
- To derive the phylogenetic relationship
- Features of SINES that make them useful for phylogenetic studies include:
  - A SINE is either present or absent.

- SINES are inserted at random in the non-coding portion of a genome.
  - Appearance of similar SINES at the same locus in two species implies that the species share a common ancestor in which the insertion event occurred.
- SINE insertion appears to be irreversible: no mechanism for loss of SINES is known, other than rare large-scale deletions that include SINE.
- Not only do SINES shown relationships, they imply which species came first. The last common ancestor of species containing a common SINE must have come after the last common ancestor linking these species and another that lacks this SINE.

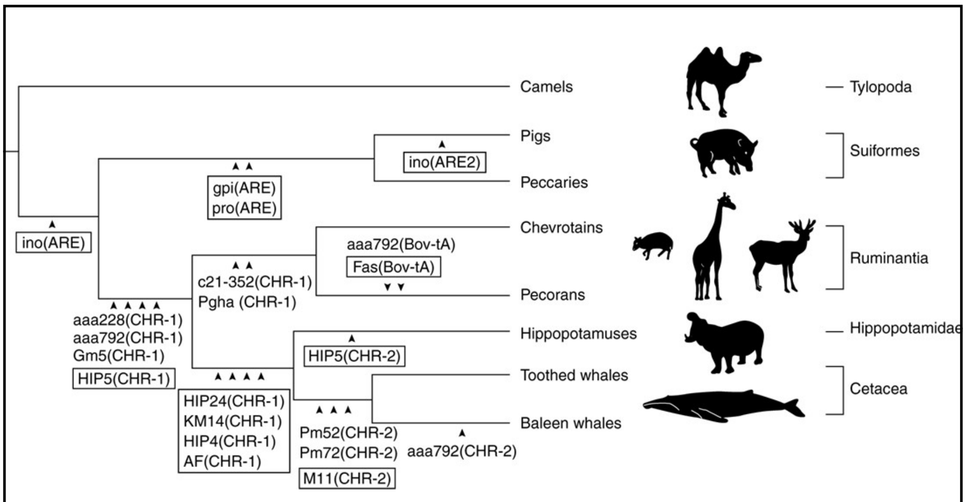
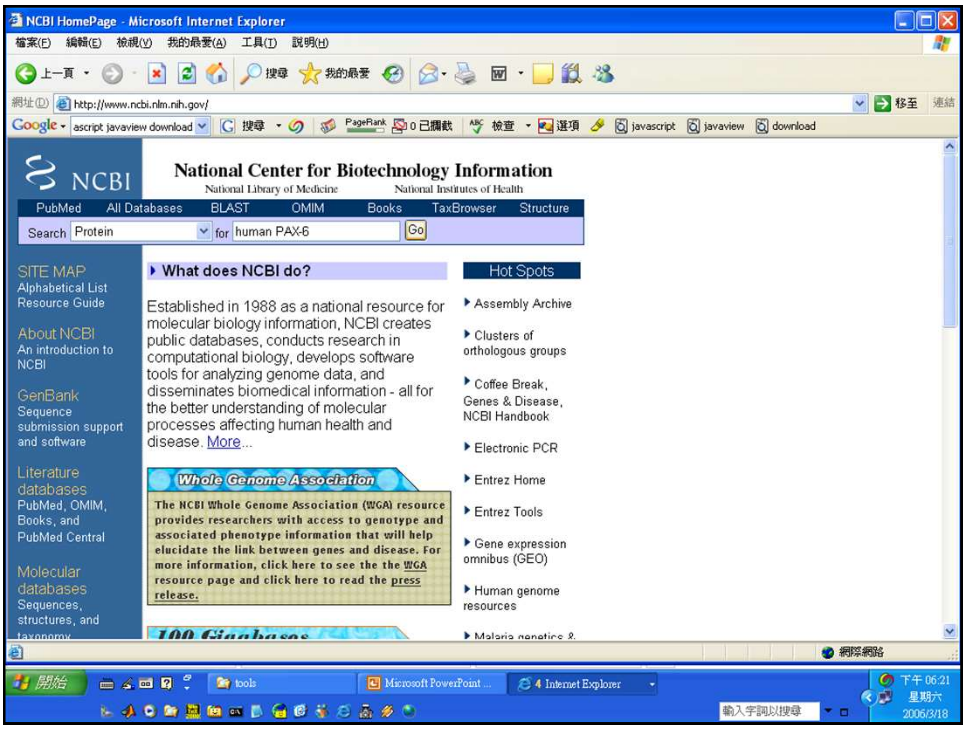


Fig. 1.5 Phylogenetic relationships among cetaceans and other artiodactyl subgroups, derived from analysis of SINE sequences.

Searching for similar sequences in databases: PSI-BLAST

- The search of a database for items similar to a probe.
- The ideal method is both sensitive and selective:
  - Sensitive: it picks up even very distant relationships
  - Selective: all the relationships that it reports are true
- A powerful tool from NCBI USA: PSI-BLAST
  - Position Specific Iterated – basic linear alignment sequence tool
- Example 1.4. Homologues of the human PAX-6 gene.
- Example 1.5. What species contain homologues of human PAX-6 detectable by PSI-BLAST?



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

ascript javaview download

National Center for Biotechnology Information

National Library of Medicine National Institutes of Health

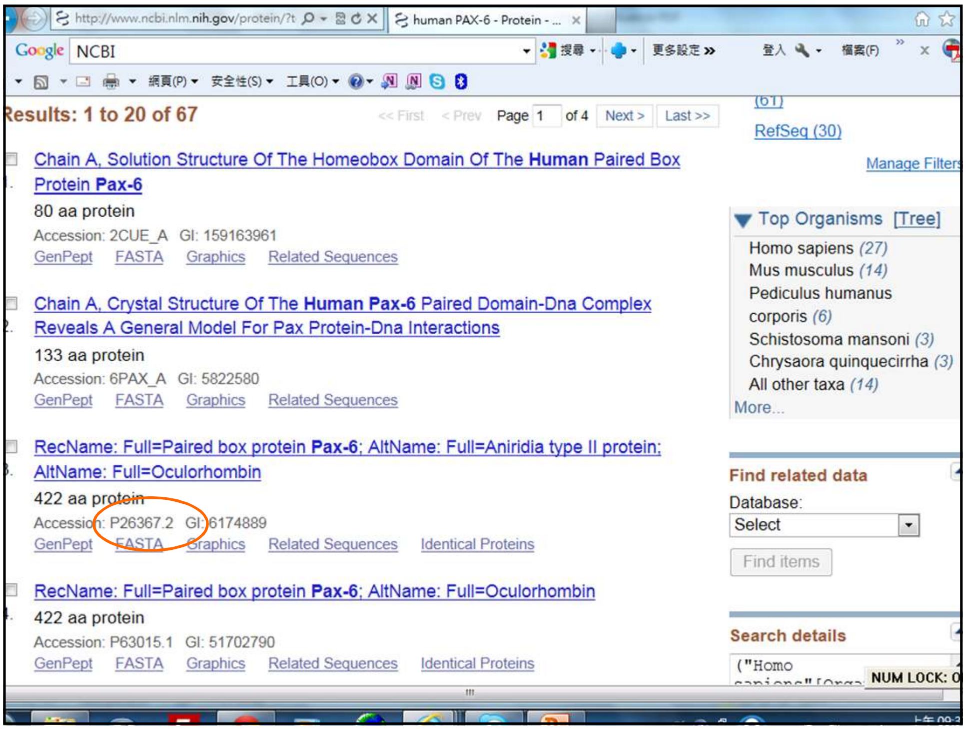
Search Protein for human PAX-6

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...

Whole Genome Association

The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the WGA resource page and click here to read the press release.



Results: 1 to 20 of 67

☐ 1. Chain A, Solution Structure Of The Homeobox Domain Of The **Human** Paired Box Protein **Pax-6**
80 aa protein
Accession: 2CUE_A GI: 159163961
GenPept FASTA Graphics Related Sequences

☐ 2. Chain A, Crystal Structure Of The **Human Pax-6** Paired Domain-Dna Complex Reveals A General Model For Pax Protein-Dna Interactions
133 aa protein
Accession: 6PAX_A GI: 5822580
GenPept FASTA Graphics Related Sequences

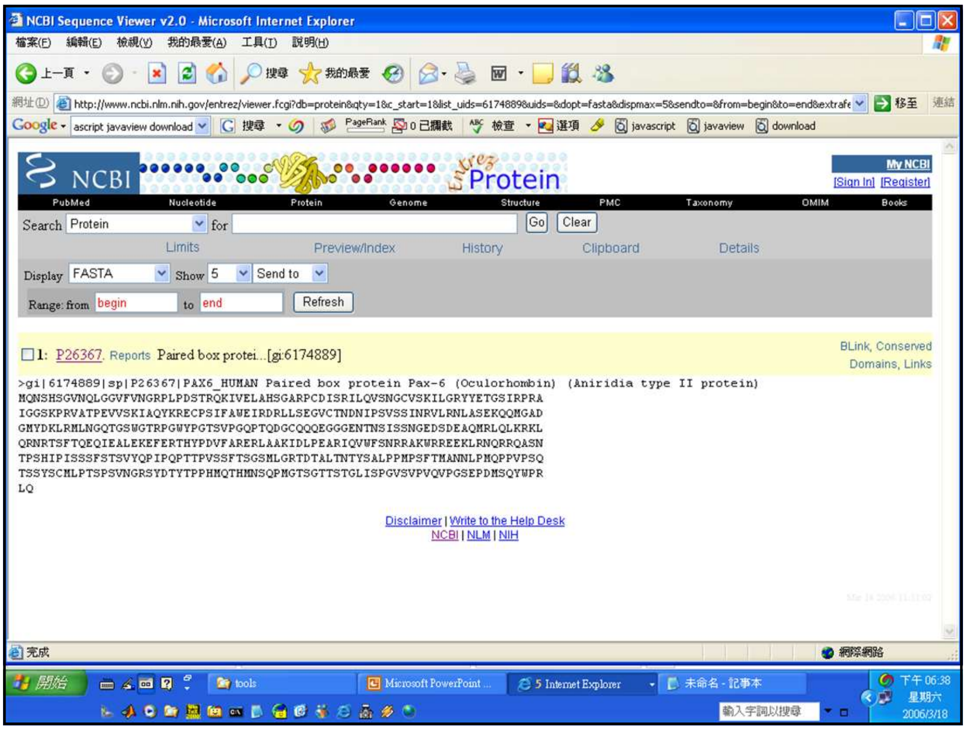
☐ 3. RecName: Full=Paired box protein **Pax-6**; AltName: Full=Aniridia type II protein; AltName: Full=Oculorhombin
422 aa protein
Accession: P26367.2 GI: 6174889
GenPept FASTA Graphics Related Sequences Identical Proteins

☐ 4. RecName: Full=Paired box protein **Pax-6**; AltName: Full=Oculorhombin
422 aa protein
Accession: P63015.1 GI: 51702790
GenPept FASTA Graphics Related Sequences Identical Proteins

Top Organisms [Tree]
Homo sapiens (27)
Mus musculus (14)
Pediculus humanus corporis (6)
Schistosoma mansoni (3)
Chrysaora quinquecirrha (3)
All other taxa (14)
More...

Find related data
Database: Select
Find items

Search details
("Homo sapiens"[Orga



http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&qty=1&c_start=1&list_uids=6174889&uids=&dopt=fasta&dispmax=5&sendto=&from=begin&to=end&extrafe

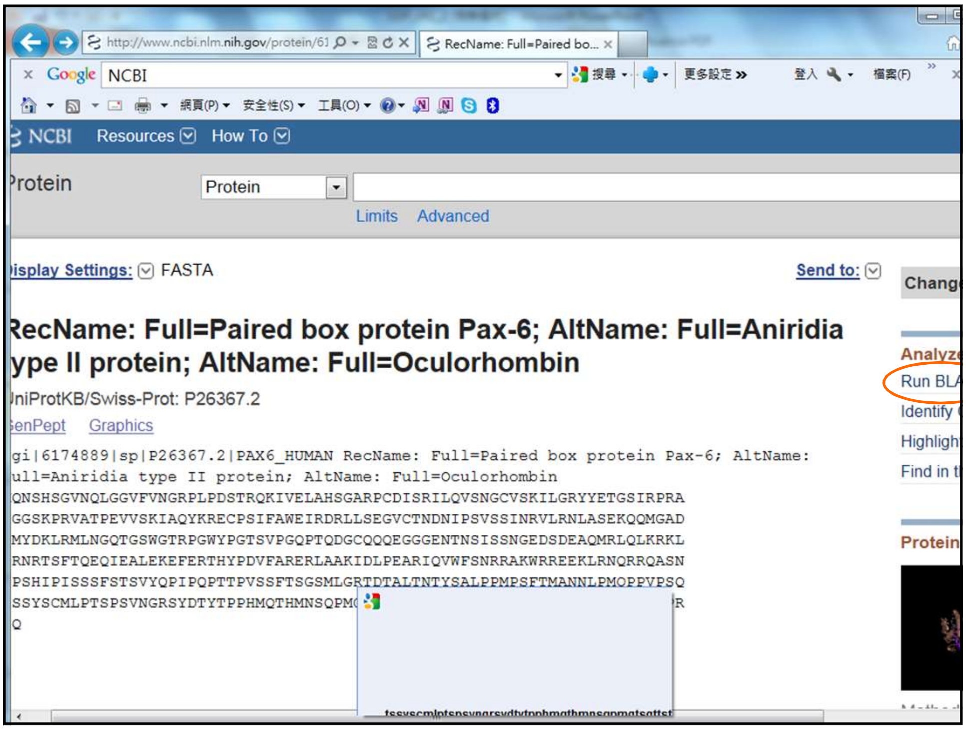
Search: Protein for

Display: FASTA Show: 5 Send to

Range: from begin to end

☐ 1: P26367. Reports Paired box protei...[gi:6174889]

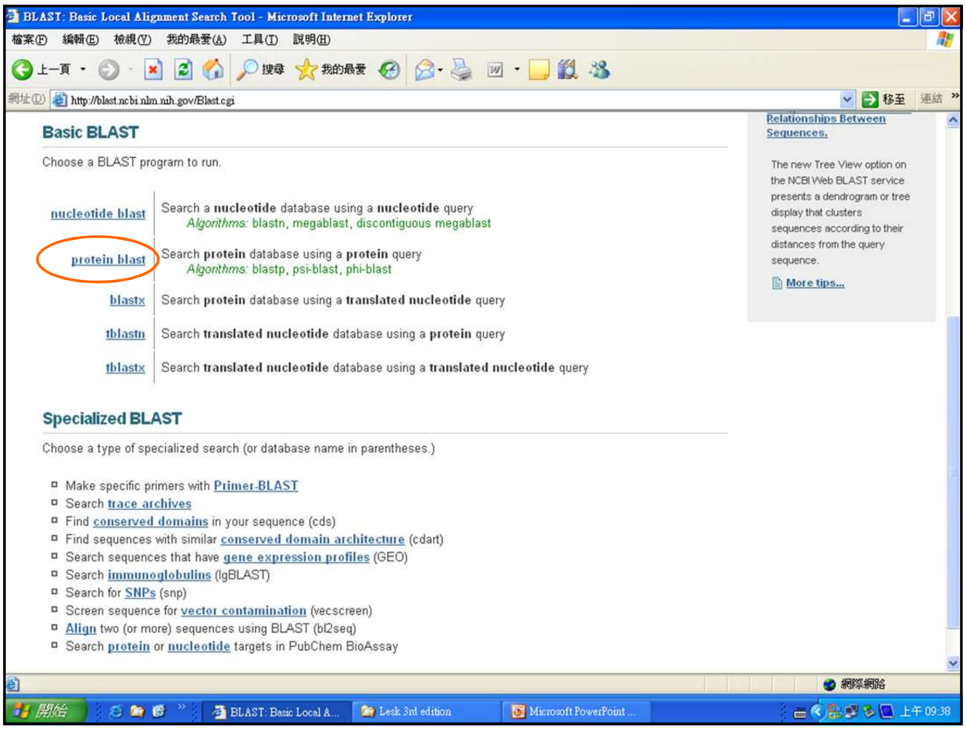
```
>gi|6174889|sp|P26367|PAX6_HUMAN Paired box protein Pax-6 (Oculorhombin) (Aniridia type II protein)
MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRA
IGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMGAD
GMYDKLRMLNGQTGSWGTRPGWYPGTSVPGQPTQDGCQQQEGGGENTNSISSNGEDSDEAQMRLQLKRKL
QRNRTSFTQEQIEALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRRQASN
TPSHIPISSSFSTSVYQPIPQPTTPVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQ
TSSYSCMLPTSPSVNGRSYDTYTPPHMQTHMNSQPMGTSGTTSTGLISPGVSVPVQVPGSEPDMSQYWPR
LQ
```



**RecName: Full=Paired box protein Pax-6; AltName: Full=Aniridia
ype II protein; AltName: Full=Oculorhombin**

JniProtKB/Swiss-Prot: P26367.2

```
gi|6174889|sp|P26367.2|PAX6_HUMAN RecName: Full=Paired box protein Pax-6; AltName:
ull=Aniridia type II protein; AltName: Full=Oculorhombin
QNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRA
GGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMGAD
MYDKLRMLNGQTGSWGTRPGWYPGTSVPGQPTQDGCQQQEGGGENTNSISSNGEDSDEAQMRLQLKRKL
RNRTSFTQEQIEALEKEFERTHYPDVFARERLAAKIDLPEARIQVWFSNRRAKWRREEKLRNQRRQASN
PSHIPISSSFSTSVYQPIPQPTTPVSSFTSGSMLGRTDTALTNTYSALPPMPSFTMANNLPMQPPVPSQ
SSYSCMLPTSPSVNGRSYDTYTPPHMQTHMNSQPM                                 R
Q
```



Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

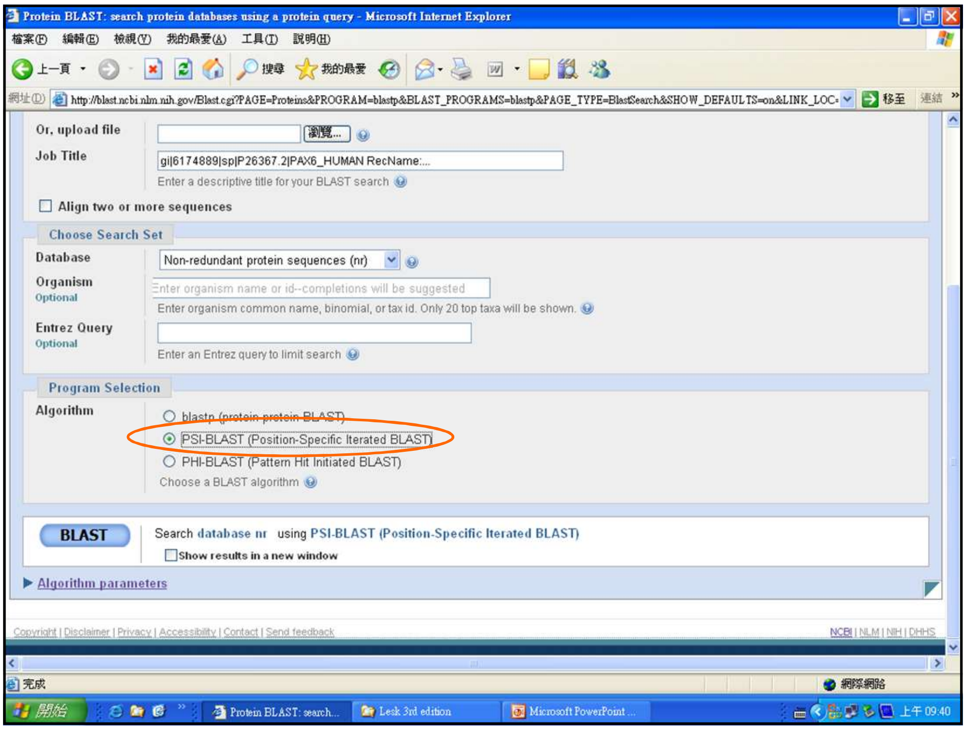
Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search immunoglobulins (IgBLAST)
- Search for SNPs (snp)
- Screen sequence for vector contamination (vecscreen)
- Align two (or more) sequences using BLAST (bl2seq)
- Search protein or nucleotide targets in PubChem BioAssay

Relationships Between Sequences.

The new Tree View option on the NCBI Web BLAST service presents a dendrogram or tree display that clusters sequences according to their distances from the query sequence.

More tips...



Protein BLAST: search protein databases using a protein query - Microsoft Internet Explorer

http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=

Or, upload file

Job Title gi|6174889|sp|P26367.2|PAX6_HUMAN RecName:...

Enter a descriptive title for your BLAST search

☐ **Align two or more sequences**

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional: Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

○ blastp (protein-protein BLAST)

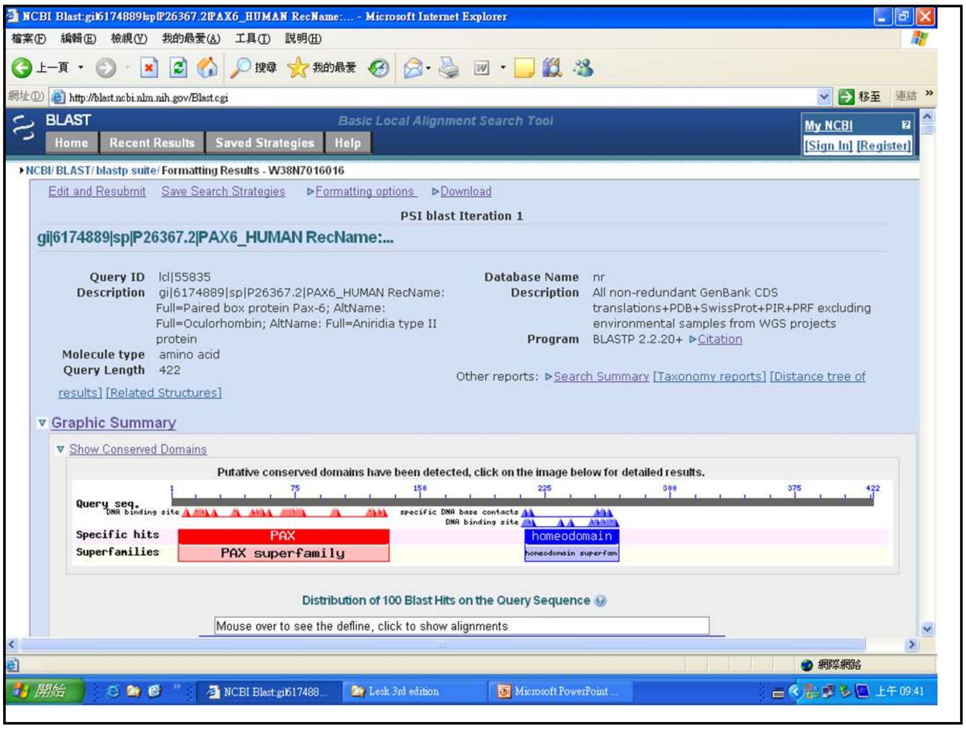
● PSI-BLAST (Position-Specific Iterated BLAST)

○ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

Search database nr using PSI-BLAST (Position-Specific Iterated BLAST)

☐ Show results in a new window



NCBI Blast:gi|6174889|sp|P26367.2|PAX6_HUMAN RecName:... - Microsoft Internet Explorer

http://blast.ncbi.nlm.nih.gov/Blast.cgi

NCBI/ BLAST/ blastp suite/ Formatting Results - W38N7016016

PSI blast Iteration 1

gi|6174889|sp|P26367.2|PAX6_HUMAN RecName:...

Query ID lcl|55835

Description gi|6174889|sp|P26367.2|PAX6_HUMAN RecName: Full=Paired box protein Pax-6; AltName: Full=Oculorhombin; AltName: Full=Aniridia type II protein

Molecule type amino acid

Query Length 422

Database Name nr

Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

Program BLASTP 2.2.20+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Related Structures]

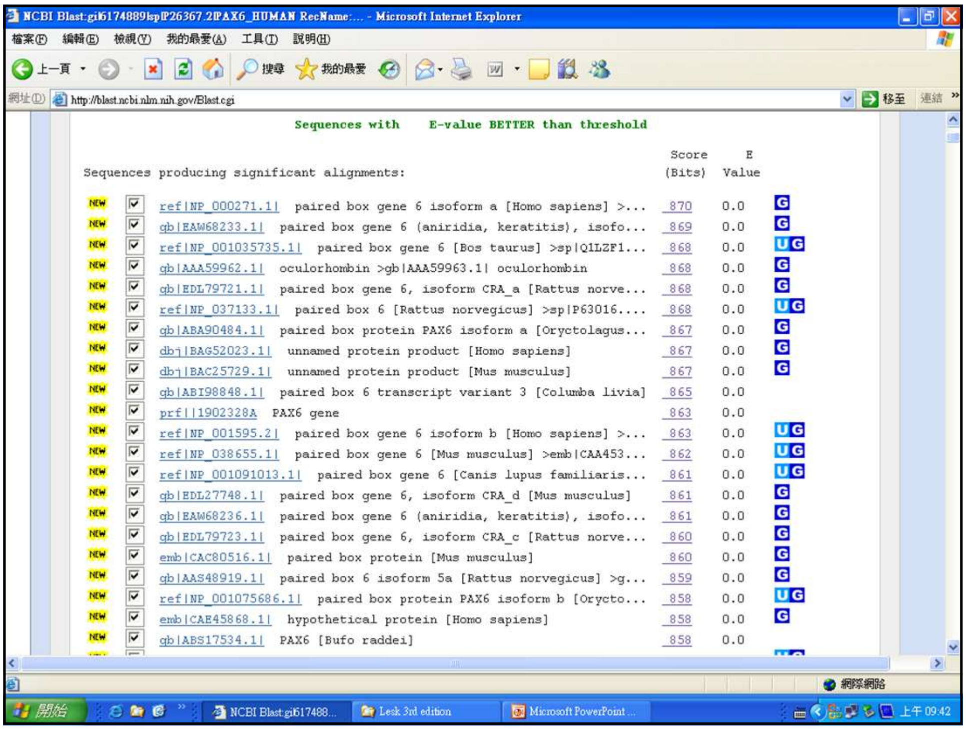
Graphic Summary

Show Conserved Domains

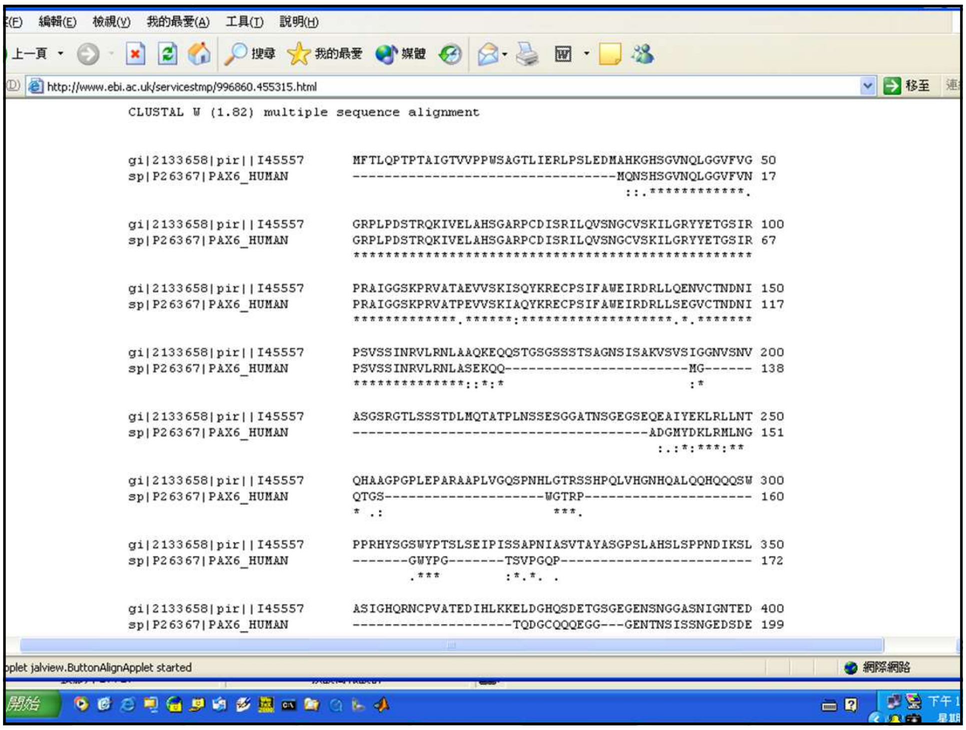
Putative conserved domains have been detected, click on the image below for detailed results.

Distribution of 100 Blast Hits on the Query Sequence

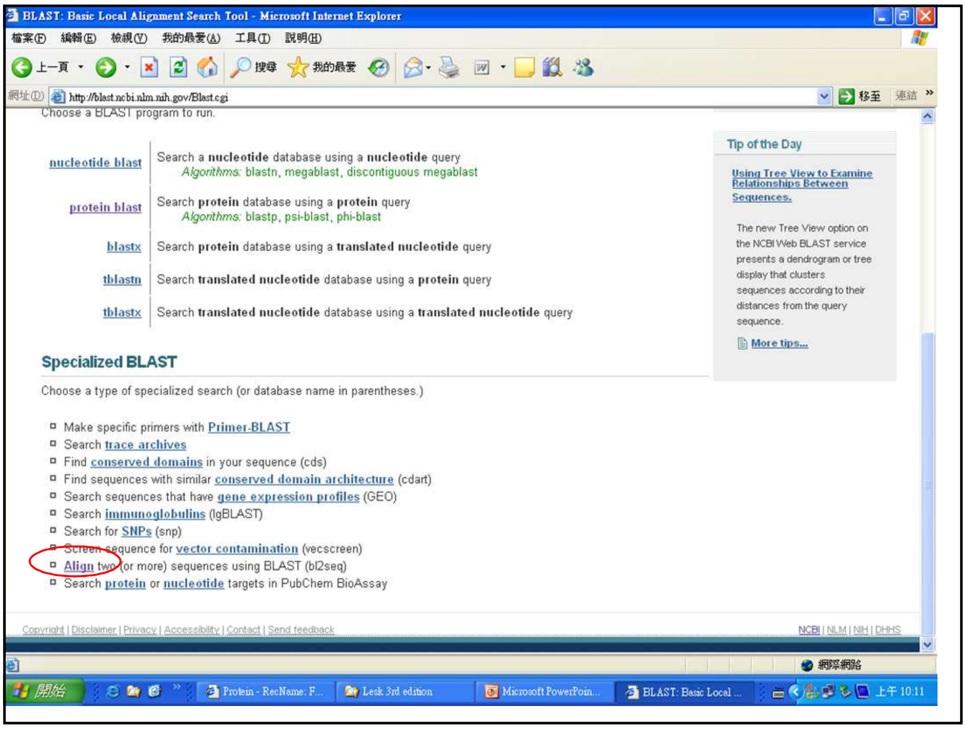
Mouse over to see the defline, click to show alignments



```
                         Sequences with    E-value BETTER than threshold

                                                                        Score     E
Sequences producing significant alignments:                             (Bits)  Value

NEW ☑ ref|NP_000271.1|  paired box gene 6 isoform a [Homo sapiens] >...    870    0.0   G
NEW ☑ gb|EAW68233.1|  paired box gene 6 (aniridia, keratitis), isofo...    869    0.0   G
NEW ☑ ref|NP_001035735.1|  paired box gene 6 [Bos taurus] >sp|Q1LZF1...    868    0.0   UG
NEW ☑ gb|AAA59962.1|  oculorhombin >gb|AAA59963.1| oculorhombin            868    0.0   G
NEW ☑ gb|EDL79721.1|  paired box gene 6, isoform CRA_a [Rattus norve...    868    0.0   G
NEW ☑ ref|NP_037133.1|  paired box 6 [Rattus norvegicus] >sp|P63016....    868    0.0   UG
NEW ☑ gb|ABA90484.1|  paired box protein PAX6 isoform a [Oryctolagus...    867    0.0   G
NEW ☑ dbj|BAG52023.1|  unnamed protein product [Homo sapiens]              867    0.0   G
NEW ☑ dbj|BAC25729.1|  unnamed protein product [Mus musculus]              867    0.0   G
NEW ☑ gb|ABI98848.1|  paired box 6 transcript variant 3 [Columba livia]    865    0.0
NEW ☑ prf||1902328A  PAX6 gene                                             863    0.0
NEW ☑ ref|NP_001595.2|  paired box gene 6 isoform b [Homo sapiens] >...    863    0.0   UG
NEW ☑ ref|NP_038655.1|  paired box gene 6 [Mus musculus] >emb|CAA453...    862    0.0   UG
NEW ☑ ref|NP_001091013.1|  paired box gene 6 [Canis lupus familiaris...    861    0.0   UG
NEW ☑ gb|EDL27748.1|  paired box gene 6, isoform CRA_d [Mus musculus]      861    0.0   G
NEW ☑ gb|EAW68236.1|  paired box gene 6 (aniridia, keratitis), isofo...    861    0.0   G
NEW ☑ gb|EDL79723.1|  paired box gene 6, isoform CRA_c [Rattus norve...    860    0.0   G
NEW ☑ emb|CAC80516.1|  paired box protein [Mus musculus]                   860    0.0   G
NEW ☑ gb|AAS48919.1|  paired box 6 isoform 5a [Rattus norvegicus] >g...    859    0.0   G
NEW ☑ ref|NP_001075686.1|  paired box protein PAX6 isoform b [Orycto...    858    0.0   UG
NEW ☑ emb|CAE45868.1|  hypothetical protein [Homo sapiens]                 858    0.0   G
NEW ☑ gb|ABS17534.1|  PAX6 [Bufo raddei]                                   858    0.0
```



```
CLUSTAL W (1.82) multiple sequence alignment


gi|2133658|pir||I45557       MFTLQPTPTAIGTVVPPWSAGTLIERLPSLEDMAHKGHSGVNQLGGVFVG 50
sp|P26367|PAX6_HUMAN         ---------------------------------MQNSHSGVNQLGGVFVN 17
                                                               ::.************.

gi|2133658|pir||I45557       GRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIR 100
sp|P26367|PAX6_HUMAN         GRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIR 67
                             **************************************************

gi|2133658|pir||I45557       PRAIGGSKPRVATAEVVSKISQYKRECPSIFAWEIRDRLLQENVCTNDNI 150
sp|P26367|PAX6_HUMAN         PRAIGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNI 117
                             *************.******:*******************.*.*******

gi|2133658|pir||I45557       PSVSSINRVLRNLAAQKEQQSTGSGSSSTSAGNSISAKVSVSIGGNVSNV 200
sp|P26367|PAX6_HUMAN         PSVSSINRVLRNLASEKQQ-----------------------MG------ 138
                             **************::*:*                       :*

gi|2133658|pir||I45557       ASGSRGTLSSSTDLMQTATPLNSSESGGATNSGEGSEQEAIYEKLRLLNT 250
sp|P26367|PAX6_HUMAN         -------------------------------------ADGMYDKLRMLNG 151
                                                                   :.:*:***:**

gi|2133658|pir||I45557       QHAAGPGPLEPARAAPLVGQSPNHLGTRSSHPQLVHGNHQALQQHQQQSW 300
sp|P26367|PAX6_HUMAN         QTGS--------------------WGTRP--------------------- 160
                             * .:                     ***.

gi|2133658|pir||I45557       PPRHYSGSWYPTSLSEIPISSAPNIASVTAYASGPSLAHSLSPPNDIKSL 350
sp|P26367|PAX6_HUMAN         -------GWYPG-------TSVPGQP------------------------ 172
                                    .***        :*.*. .

gi|2133658|pir||I45557       ASIGHQRNCPVATEDIHLKKELDGHQSDETGSGEGENSNGGASNIGNTED 400
sp|P26367|PAX6_HUMAN         --------------------TQDGCQQQEGG---GENTNSISSNGEDSDE 199
```



Choose a BLAST program to run.

nucleotide blast	Search a **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous megablast
protein blast	Search **protein** database using a **protein** query *Algorithms:* blastp, psi-blast, phi-blast
blastx	Search **protein** database using a **translated nucleotide** query
tblastn	Search **translated nucleotide** database using a **protein** query
tblastx	Search **translated nucleotide** database using a **translated nucleotide** query

Tip of the Day

Using Tree View to Examine Relationships Between Sequences.

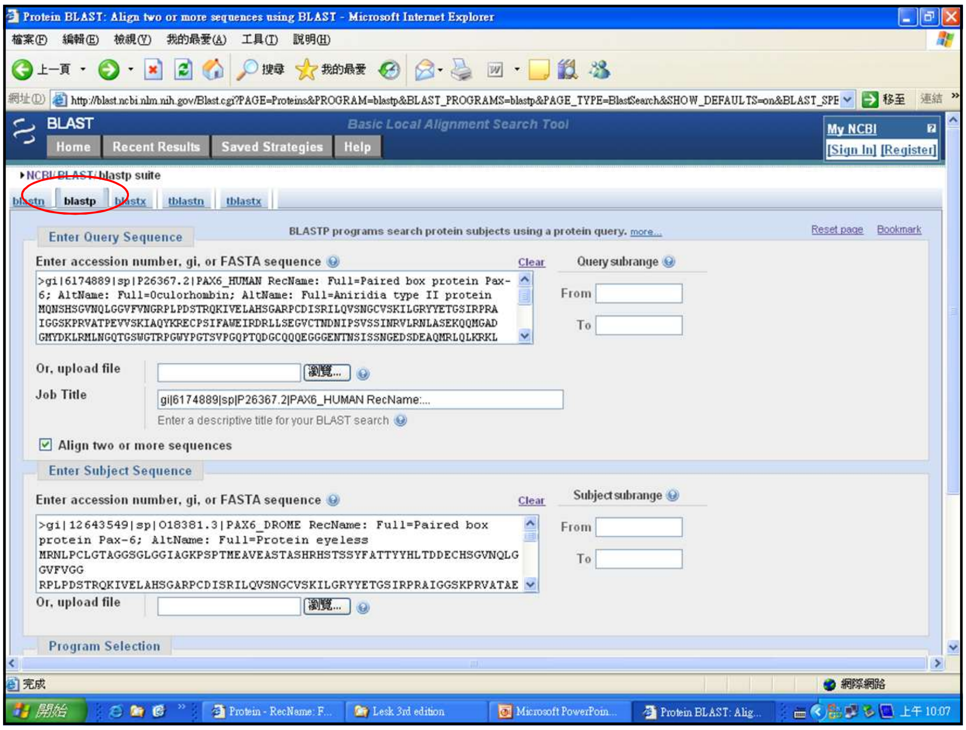
The new Tree View option on the NCBI Web BLAST service presents a dendrogram or tree display that clusters sequences according to their distances from the query sequence.

More tips...

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with Primer-BLAST
- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search immunoglobulins (IgBLAST)
- Search for SNPs (snp)
- Screen sequence for vector contamination (vecscreen)
- Align two (or more) sequences using BLAST (bl2seq)
- Search protein or nucleotide targets in PubChem BioAssay



BLAST Basic Local Alignment Search Tool

NCBI/ BLAST/ blastp suite

BLASTP programs search protein subjects using a protein query. more...

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
>gi|6174889|sp|P26367.2|PAX6_HUMAN RecName: Full=Paired box protein Pax-
6; AltName: Full=Oculorhombin; AltName: Full=Aniridia type II protein
MQNSHSGVNQLGGVFVNGRPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRA
IGGSKPRVATPEVVSKIAQYKRECPSIFAWEIRDRLLSEGVCTNDNIPSVSSINRVLRNLASEKQQMGAD
GMYDKLRMLNGQTGSWGTRPGWYPGTSVPGQPTQDGCQQQEGGGENTNSISSNGEDSDEAQMRLQLKRKL
```

Job Title: gi|6174889|sp|P26367.2|PAX6_HUMAN RecName:...

Enter a descriptive title for your BLAST search

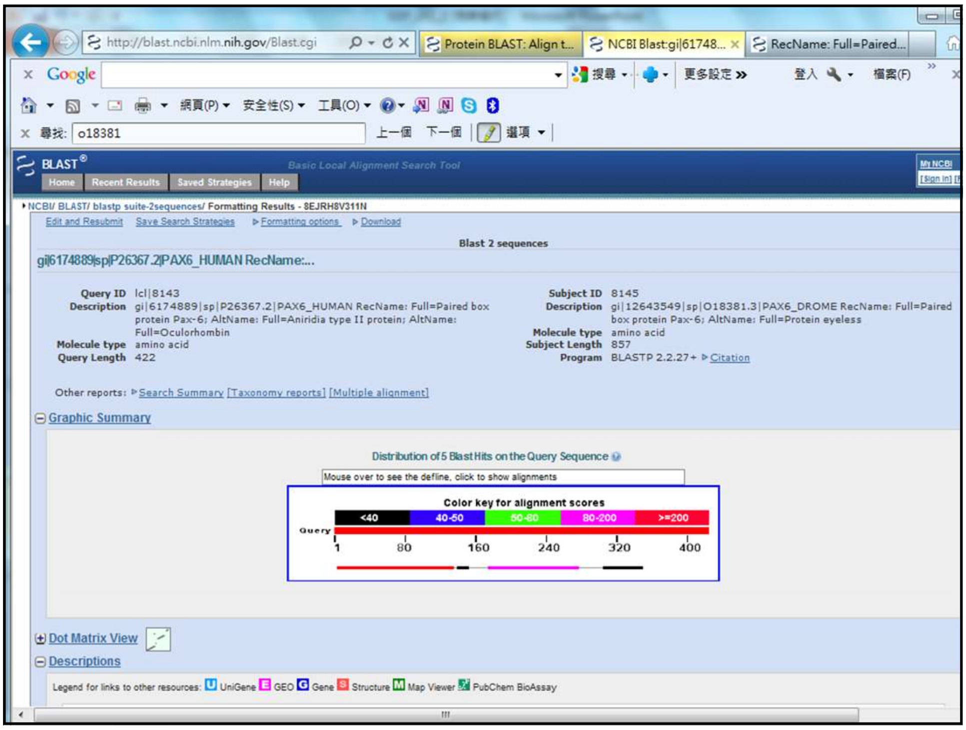
☑ Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

```
>gi|12643549|sp|O18381.3|PAX6_DROME RecName: Full=Paired box
protein Pax-6; AltName: Full=Protein eyeless
MRNLPCLGTAGGSGLGGIAGKPSPTMEAVEASTASHRHSTSSYFATTYYHLTDDECHSGVNQLG
GVFVGG
RPLPDSTRQKIVELAHSGARPCDISRILQVSNGCVSKILGRYYETGSIRPRAIGGSKPRVATAE
```

Program Selection



Blast 2 sequences

gi|6174889|sp|P26367.2|PAX6_HUMAN RecName:...

Query ID lcl|8143
Description gi|6174889|sp|P26367.2|PAX6_HUMAN RecName: Full=Paired box protein Pax-6; AltName: Full=Aniridia type II protein; AltName: Full=Oculorhombin
Molecule type amino acid
Query Length 422

Subject ID 8145
Description gi|12643549|sp|O18381.3|PAX6_DROME RecName: Full=Paired box protein Pax-6; AltName: Full=Protein eyeless
Molecule type amino acid
Subject Length 857
Program BLASTP 2.2.27+ ▷ Citation

Other reports: ▷ Search Summary [Taxonomy reports] [Multiple alignment]

⊖ Graphic Summary

Distribution of 5 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

⊞ Dot Matrix View

⊖ Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer PubChem BioAssay