生物信息学：导论与方法

Bioinformatics: Introduction and Methods

https://www.coursera.org/course/pkubioinfo

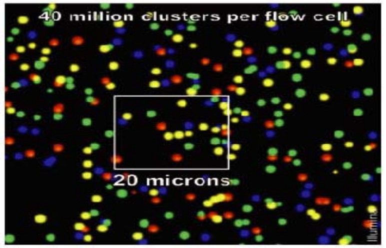
生物信息学：导论与方法
Bioinformatics: Introduction and Methods

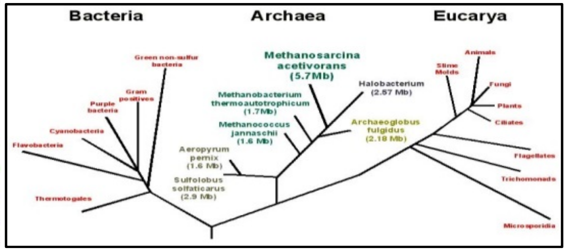
北京大学生物信息学中心 高歌、魏丽萍

Ge Gao & Liping Wei

Center for Bioinformatics, Peking University

```
TAACCCTAACCCTAACCCTAACCCTAACCCTA
CCTAACCCTAACCCTAACCCTAACCCTAACCC
CCCTAACCCCTAACCCTAACCCTAACCCTAAC
AACCCTAACCCTAACCCTAACCCCTAACCCTA
ACCCTAACCCCAACCCCAACCCCAACCCCAAC
CTACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCCTAACCCTAA
```



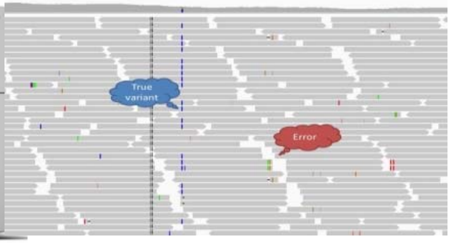

Bacteria
Archaea
Eucarya
Green non-sulfur bacteria
Gram positives
Purple bacteria
Cyanobacteria
Flavobacteria
Thermotogales
Methanosarcina acetivorans (5.7Mb)
Halobacterium (2.57 Mb)
Methanobacterium thermoautotrophicum (1.7Mb)
Methanococcus jannaschii (1.6 Mb)
Archaeoglobus fulgidus (2.18 Mb)
Aeropyrum pernix (1.6 Mb)
Sulfolobus solfaticarus (2.9 Mb)
Slime Molds
Animals
Fungi
Plants
Ciliates
Flagellates
Trichomonads
Microsporidia

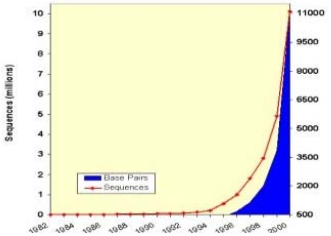
Sequence Database Search

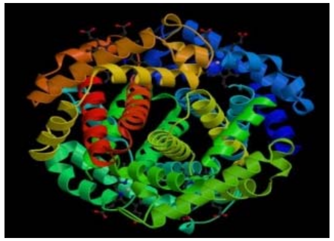
北京大学生物信息学中心 高歌

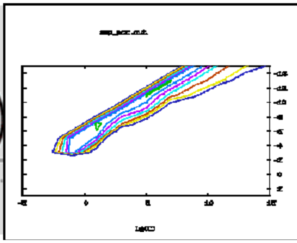
Ge Gao, Ph.D.

Center for Bioinformatics, Peking University

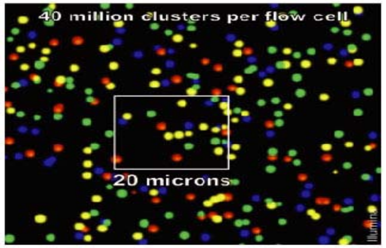


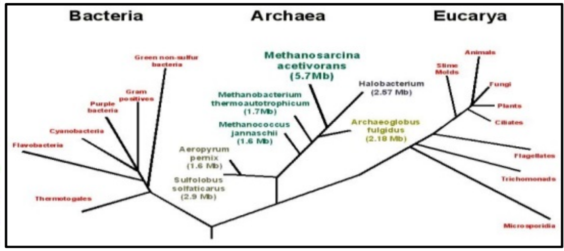






```
TAACCCTAACCCTAACCCTAACCCTAACCCTA
CCTAACCCTAACCCTAACCCTAACCCTAACCC
CCCTAACCCCTAACCCTAACCCTAACCCTAAC
AACCCTAACCCTAACCCTAACCCCTAACCCTA
ACCCTAACCCCAACCCCAACCCCAACCCCAAC
CTACCCTAACCCTAACCCTAACCCTAACCCTA
ACCCTAACCCTAACCCTAACCCCTAACCCTAA
```



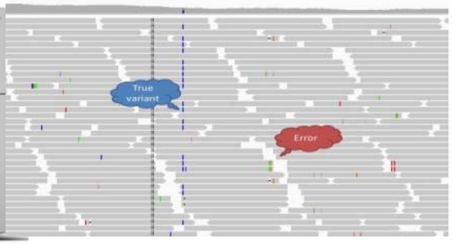

Bacteria
Archaea
Eucarya
Green non-sulfur bacteria
Gram positives
Purple bacteria
Cyanobacteria
Flavobacteria
Thermotogales
Methanosarcina acetivorans (5.7Mb)
Halobacterium (2.57 Mb)
Methanobacterium thermoautotrophicum (1.7Mb)
Methanococcus jannaschii (1.6 Mb)
Archaeoglobus fulgidus (2.18 Mb)
Aeropyrum pernix (1.6 Mb)
Sulfolobus solfaticarus (2.9 Mb)
Slime Molds
Animals
Fungi
Plants
Ciliates
Flagellates
Trichomonads
Microsporidia

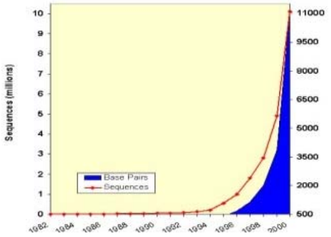
Unit 1: Sequence Databases

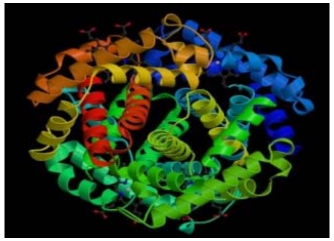
北京大学生物信息学中心 高歌

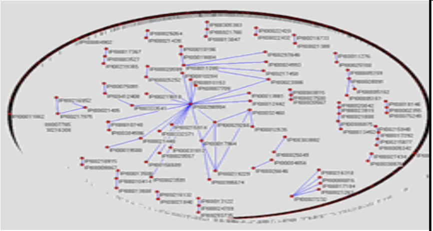
Ge Gao, Ph.D.

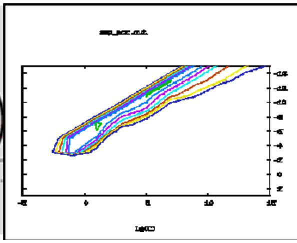
Center for Bioinformatics, Peking University

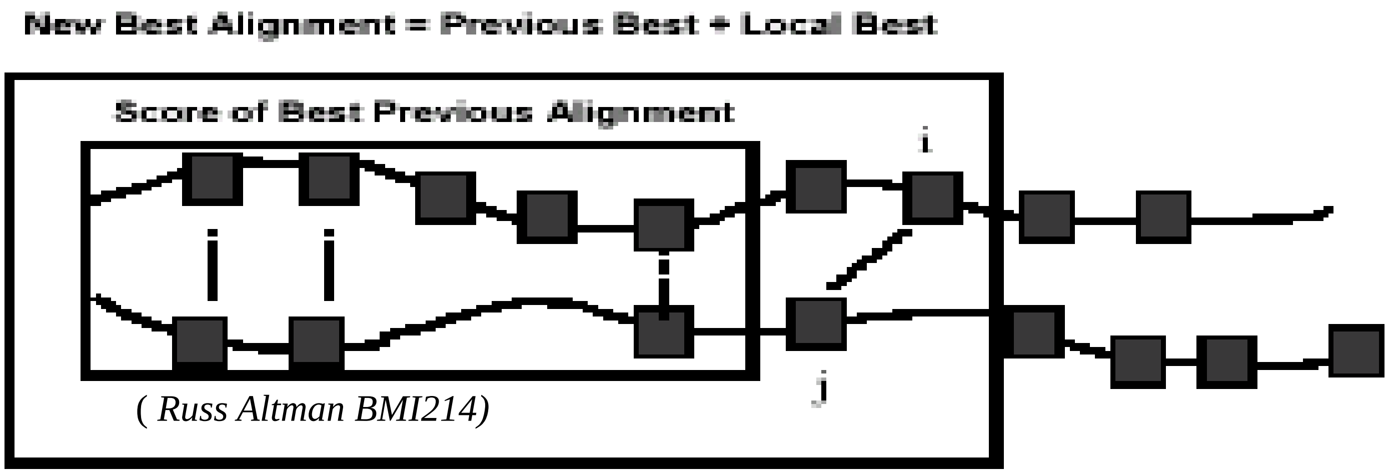











New Best Alignment = Previous Best + Local Best
Score of Best Previous Alignment
i
j

(Russ Altman BMI214)

S	S	-
T	-	T

Global alignment (Needleman-Wunsch)

$$F(0,0) = 0$$

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) + d \\ F(i, j-1) + d \end{cases}$$

Local alignment (Smith-Waterman)

$$F(0,0) = 0$$

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) + d \\ F(i, j-1) + d \\ 0 \end{cases}$$

Sequence Alignment

“How can we determine the similarity between two sequences?”

Why is it important?

- Similar sequence ➔ Similar structure ➔ Similar function (The “*Sequence-to-Structure-to-Function Paradigm*”)
- Similar sequence ➔ Common ancestor (“*Homology*”)

Sequence Database Searching

- Rather than do the alignment pair-wise, it’s more often to search sequence database in a high-throughput style.

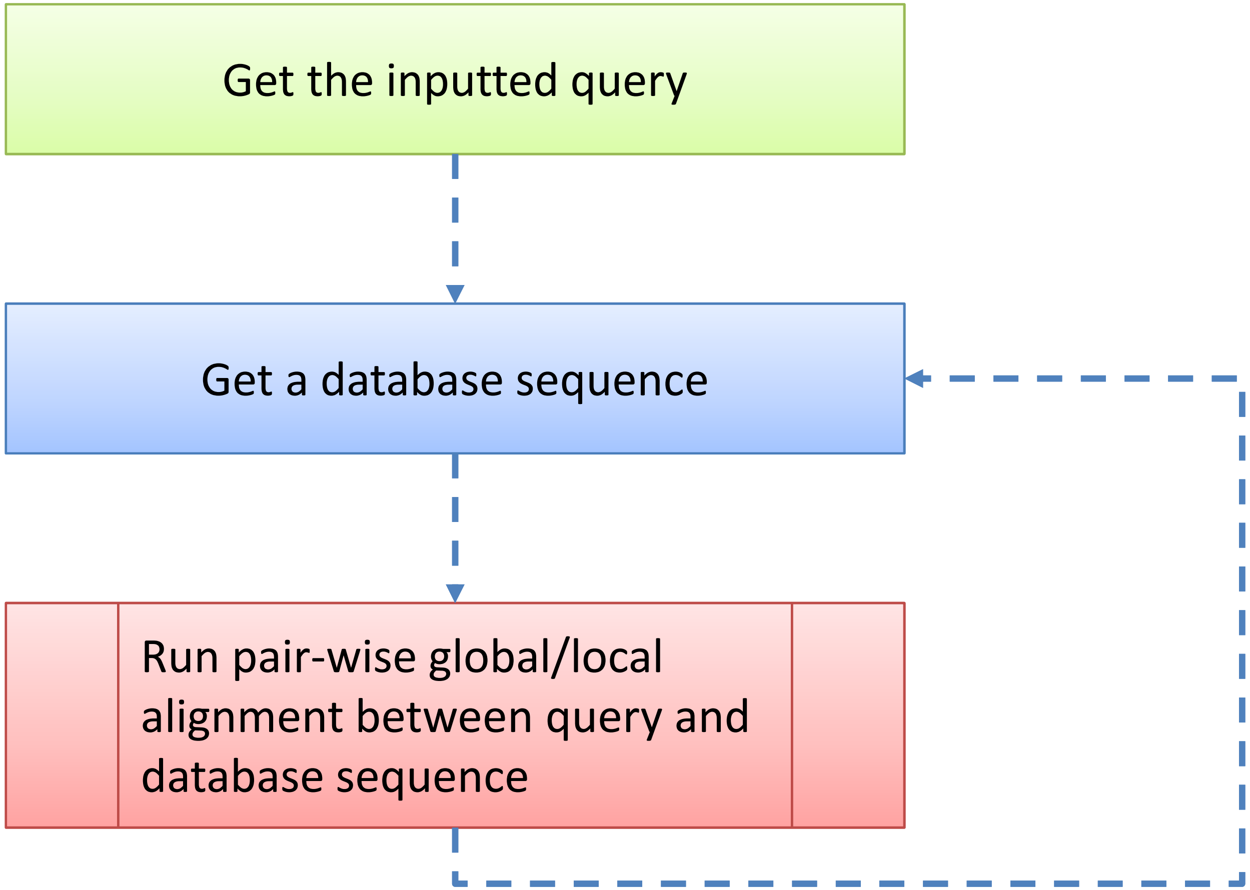
- Or, identify similarities between
  - novel query sequence
    whose structures and functions are usually unknown and/or uncharacterized
  - sequences in (public) databases
    whose structures and functions have been elucidated and annotated.

Sequence Database Searching

- The query sequence is compared/aligned with every sequence in the database

- Statistically significant hits are assumed to be related to the query sequence
  - Similar function/structure
  - Common evolutionary ancestor

A (naïve) algorithm for database searching



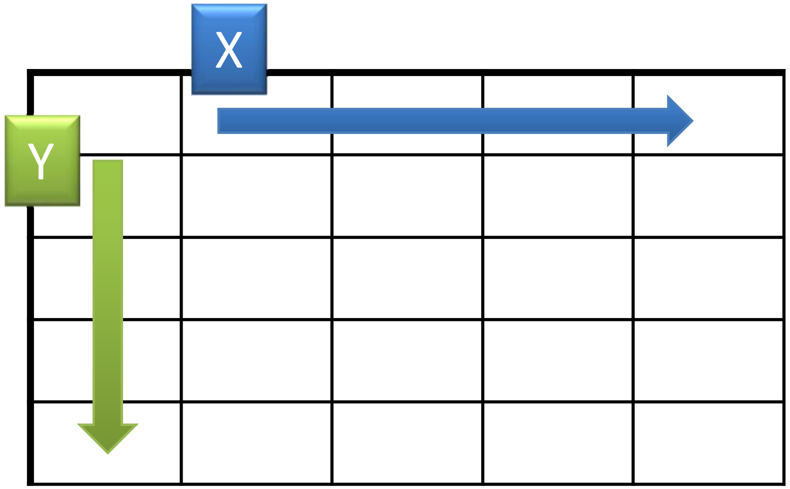
Get the inputted query

Get a database sequence

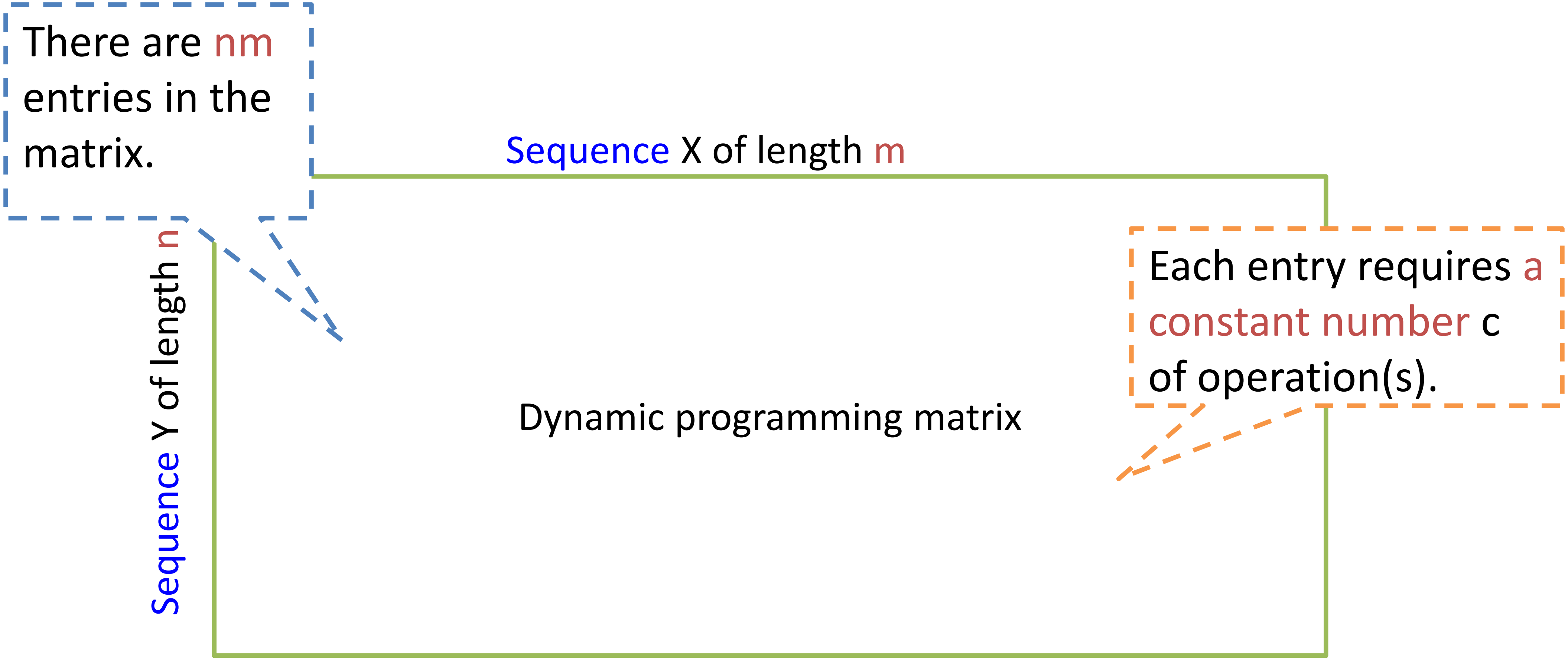
Run pair-wise global/local alignment between query and database sequence

$$F(0,0) = 0$$

$$F(i,\ j) = \max \begin{cases} F(i-1,\ j-1) + s(x_i,\ y_j) & \boldsymbol{x}_{\mathrm{i}} \textbf{ aligned to } \boldsymbol{y}_{\mathrm{j}} \\ F(i-1,\ j) + d & \boldsymbol{x}_{\mathrm{i}} \textbf{ aligned to } \boldsymbol{a\ gap} \\ F(i,\ j-1) + d & \boldsymbol{y}_{\mathrm{j}} \textbf{ aligned to } \boldsymbol{a\ gap} \end{cases}$$



$$\begin{tikzcd}
F(i-1,\ j-1) \arrow[dr] & & F(i,\ j-1) \arrow[d] \\
& s(x_i,\ y_j) \arrow[dr] & d \arrow[d] \\
F(i-1,\ j) \arrow[r] & d \arrow[r] & F(i,\ j)
\end{tikzcd}$$



There are nm entries in the matrix.

Sequence X of length m

Sequence Y of length n

Each entry requires a constant number c of operation(s).

Dynamic programming matrix

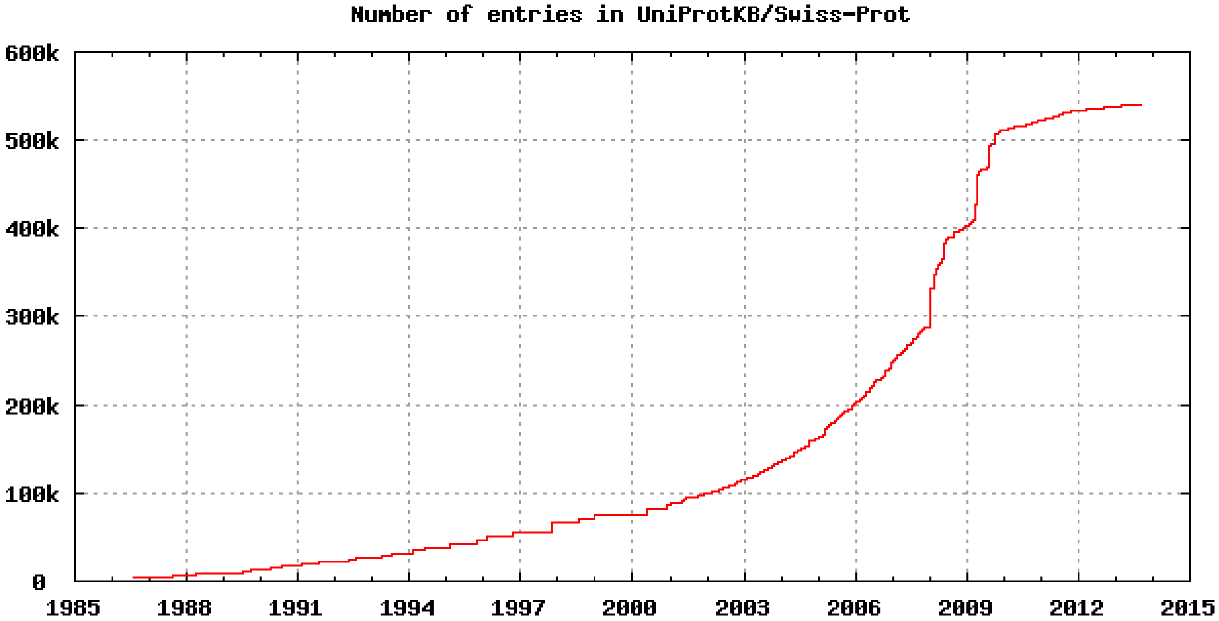
c*m*n operations needed in total, for one pair-wise alignment.

- Say your query sequence (HBA_HUMAN) has 142 amino acids

- Most recent release of human-curated Swiss-Prot protein databases contains 540,958 sequences with 192,206,270 amino acids (Sept 18th, 2013);
  - On average, the sequence length is 192,206,270/540,958 = 355.30 aa

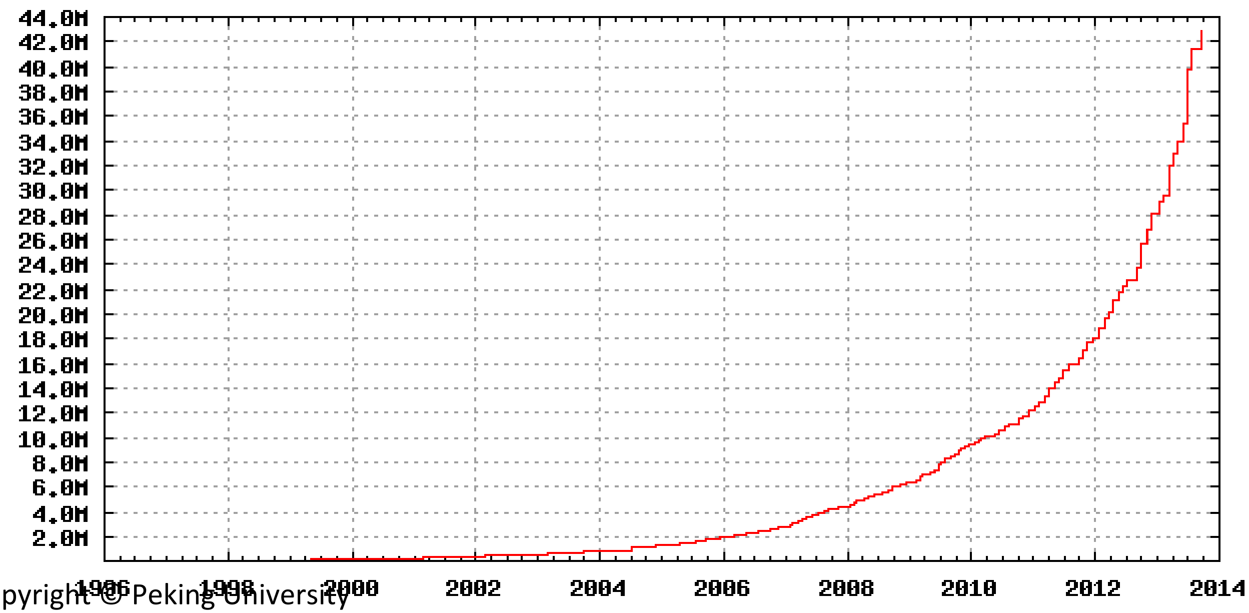
- And assume your super-fast computer can run one operation in 1µs = (0.000001s)

- Then, you will need 7.8 hr for ONE comparison!



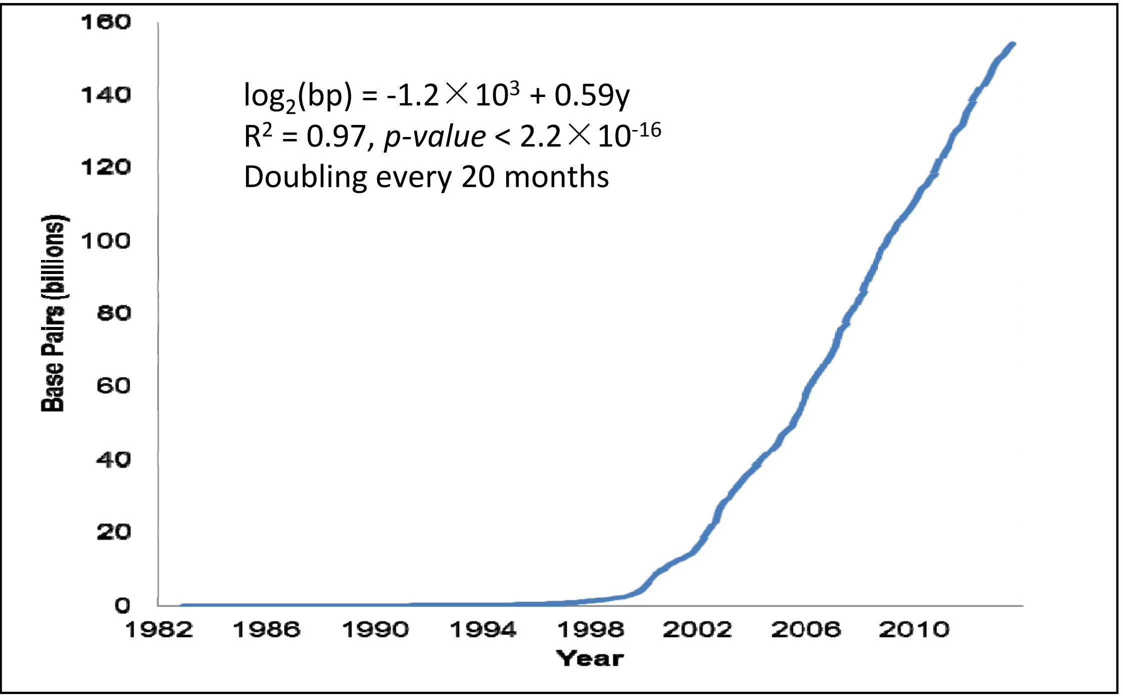
Source: http://web.expasy.org/docs/relnotes/relstat.html

Number of entries in UniProtKB/TrEMBL



Source: http://www.ebi.ac.uk/uniprot/TrEMBLstats

Genbank

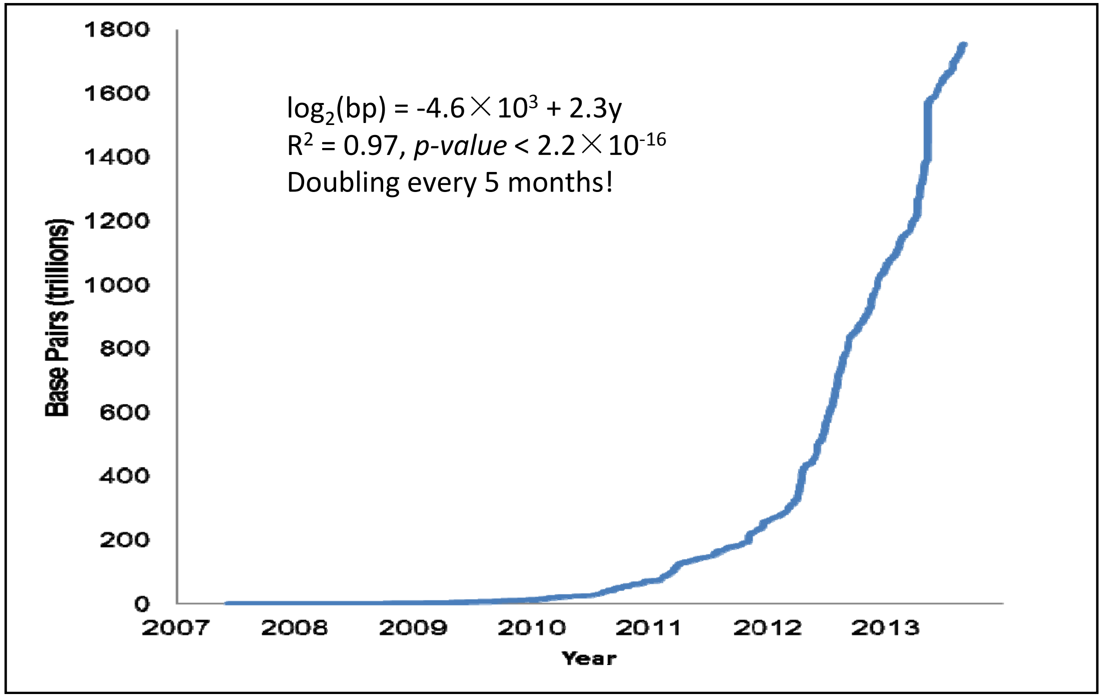


$\log_2(\mathrm{bp}) = -1.2\times10^3 + 0.59y$

$R^2 = 0.97$, *p-value* $< 2.2\times10^{-16}$

Doubling every 20 months

SRA



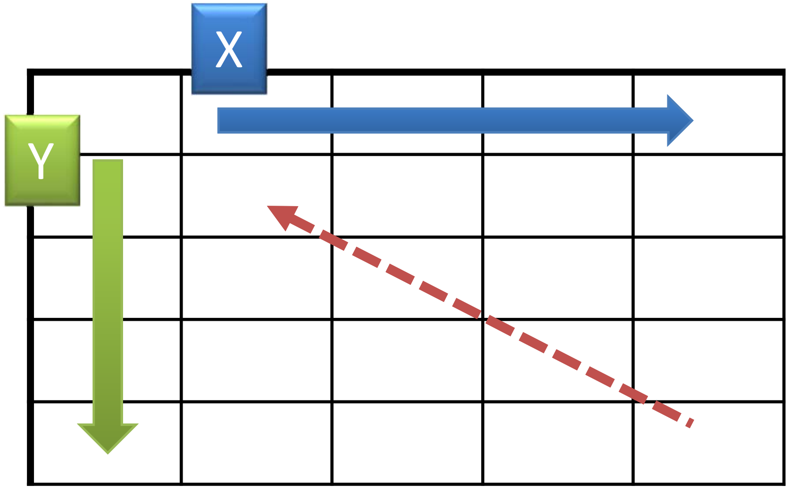
$\log_2(\mathrm{bp}) = -4.6\times10^3 + 2.3y$

$R^2 = 0.97$, *p-value* $< 2.2\times10^{-16}$

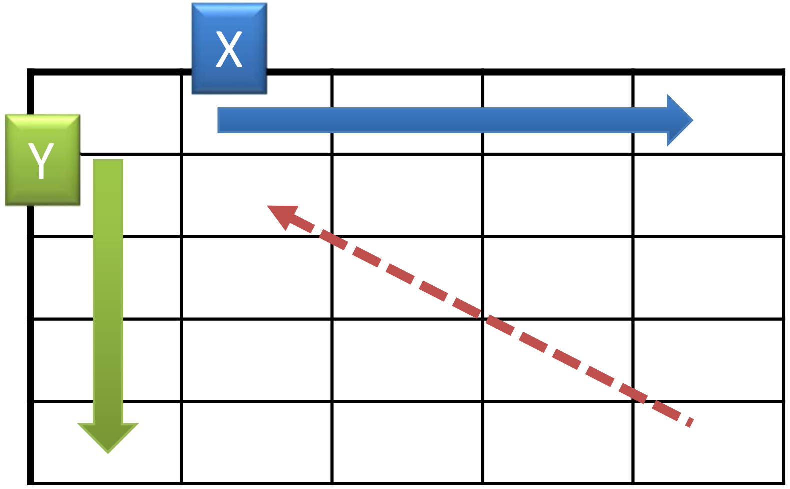
Doubling every 5 months!

$$F(0,0) = 0$$

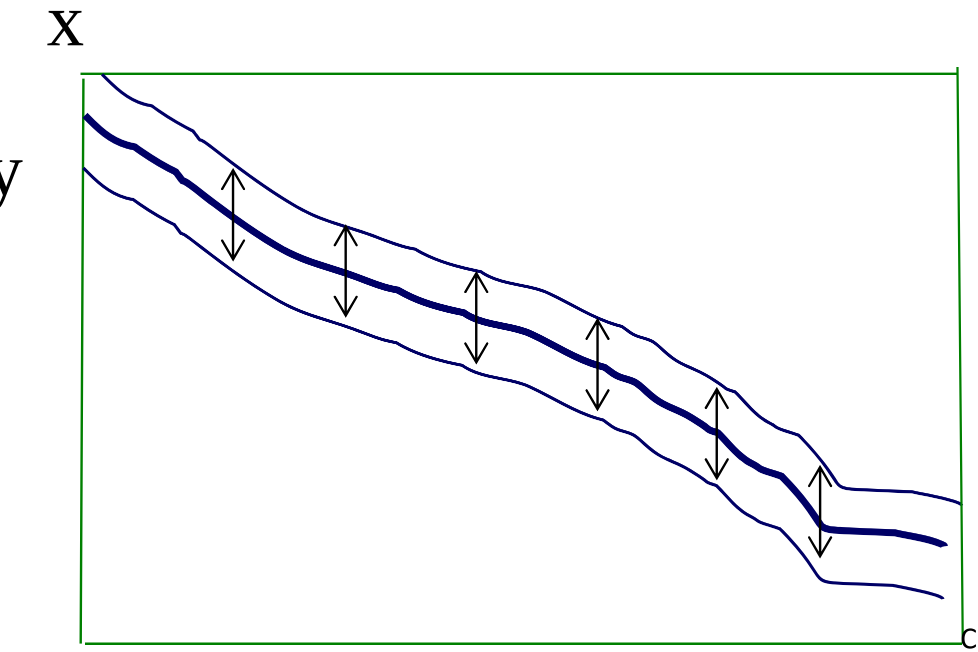
$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) & \textbf{\textit{x}}_\textbf{i} \textbf{ aligned to } \textbf{\textit{y}}_\textbf{j} \\ F(i-1, j) + d & \textbf{\textit{x}}_\textbf{i} \textbf{ aligned to } \textbf{\textit{a gap}} \\ F(i, j-1) + d & \textbf{\textit{y}}_\textbf{j} \textbf{ aligned to } \textbf{\textit{a gap}} \end{cases}$$

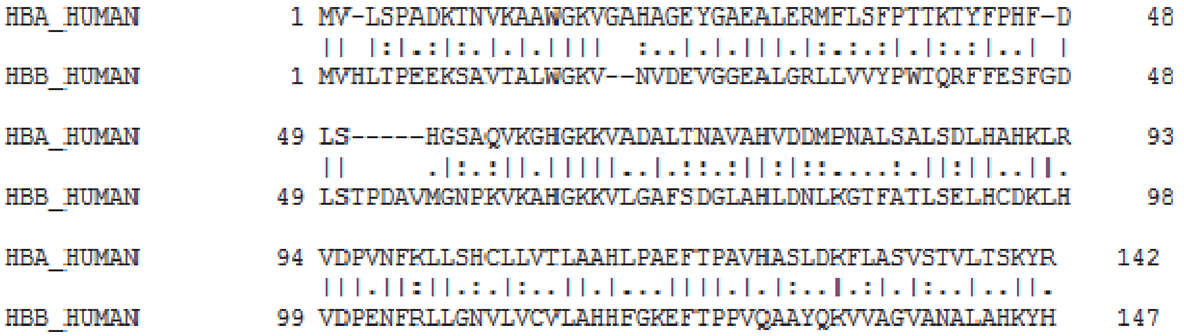


$$\begin{tikzcd}
F(i-1, j-1) \arrow[dr] & & F(i, j-1) \arrow[d] \\
& s(x_i, y_j) \arrow[dr] & d \arrow[d] \\
F(i-1, j) \arrow[r] & d \arrow[r] & F(i, j)
\end{tikzcd}$$



$$\begin{tikzcd}
F\left(i-1,\ j-1\right) \arrow[dr] & & F\left(i,\ j-1\right) \arrow[d] \\
& s\left(x_i,\ y_j\right) \arrow[dr] & d \arrow[d] \\
F\left(i-1,\ j\right) \arrow[r] & d \arrow[r] & F\left(i,\ j\right)
\end{tikzcd}$$



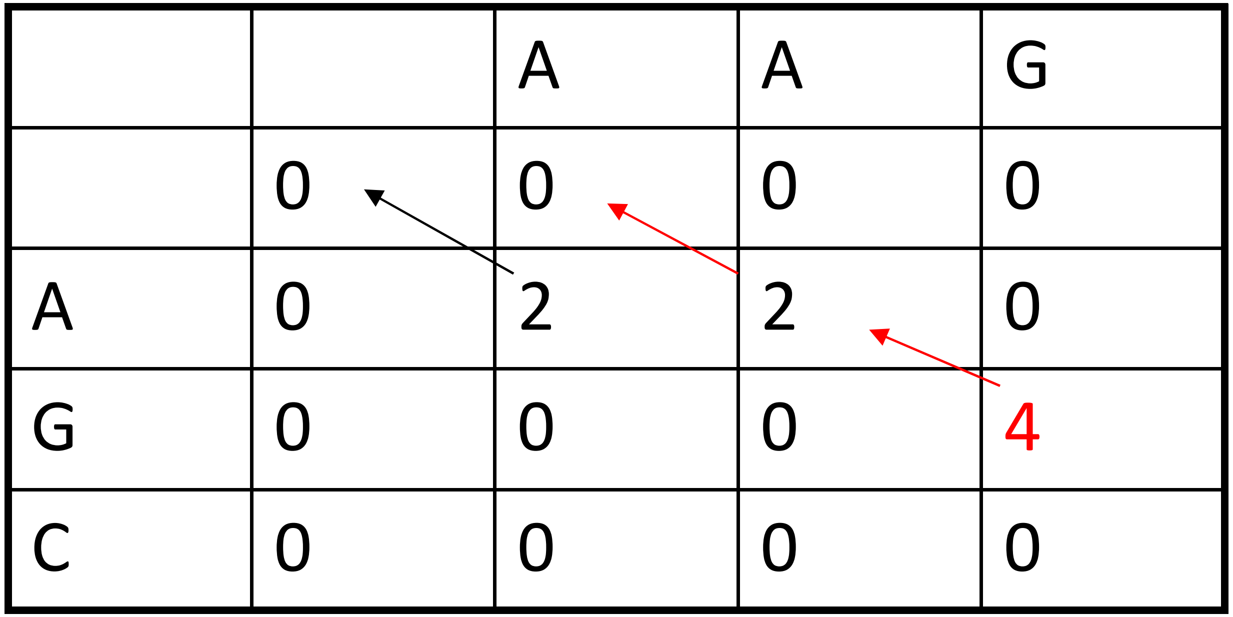


```
HBA_HUMAN         1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D     48
                    || |:|.:|:.|.|.||||  :..|.|.|||.|:.:.:|.|:.:|..| |
HBB_HUMAN         1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

HBA_HUMAN        49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR     93
                    ||     .|:.:||.|||||..|.::.:||:|::....:.||:||..||.
HBB_HUMAN        49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

HBA_HUMAN        94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142
                    |||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||.
HBB_HUMAN        99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

A G
A G



		A	A	G
	0	0	0	0
A	0	2	2	0
G	0	0	0	4
C	0	0	0	0

$$\begin{tikzcd}
F(i-1,\ j-1) \arrow[dr] & & 0 \arrow[ddr] & F(i,\ j-1) \arrow[d] \\
& s(x_i,\ y_j) \arrow[drr] & & d \arrow[d] \\
F(i-1,\ j) \arrow[r] & d \arrow[rr] & & F(i,\ j)
\end{tikzcd}$$

BLAST: Intro

- To make the alignment effectively, a *Heuristic* algorithm BLAST (Basic Local Alignment Search Tool) is proposed by Altschul *et al* in 1990.

- BLAST finds the highest scoring locally optimal alignments between a query sequence and a database.
  - Very fast algorithm
  - Can be used to search extremely large databases
  - Sufficiently sensitive and selective for most purposes
  - Robust – the default parameters just work for most cases

J. Mol. Biol. (1990) 215, 403–410

Basic Local Alignment Search Tool

Stephen F. Altschul[1], Warren Gish[1], Webb Miller[2]
Eugene W. Myers[3] and David J. Lipman[1]

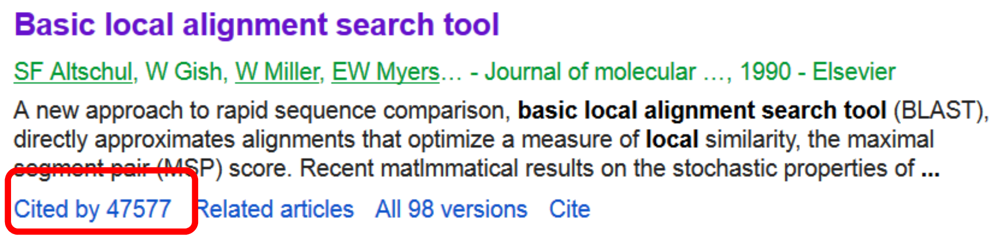
[1]*National Center for Biotechnology Information*
National Library of Medicine, National Institutes of Health
Bethesda, MD 20894, U.S.A.

[2]*Department of Computer Science*
The Pennsylvania State University, University Park, PA 16802, U.S.A.

[3]*Department of Computer Science*
University of Arizona, Tucson, AZ 85721, U.S.A.

(Received 26 February 1990; accepted 15 May 1990)

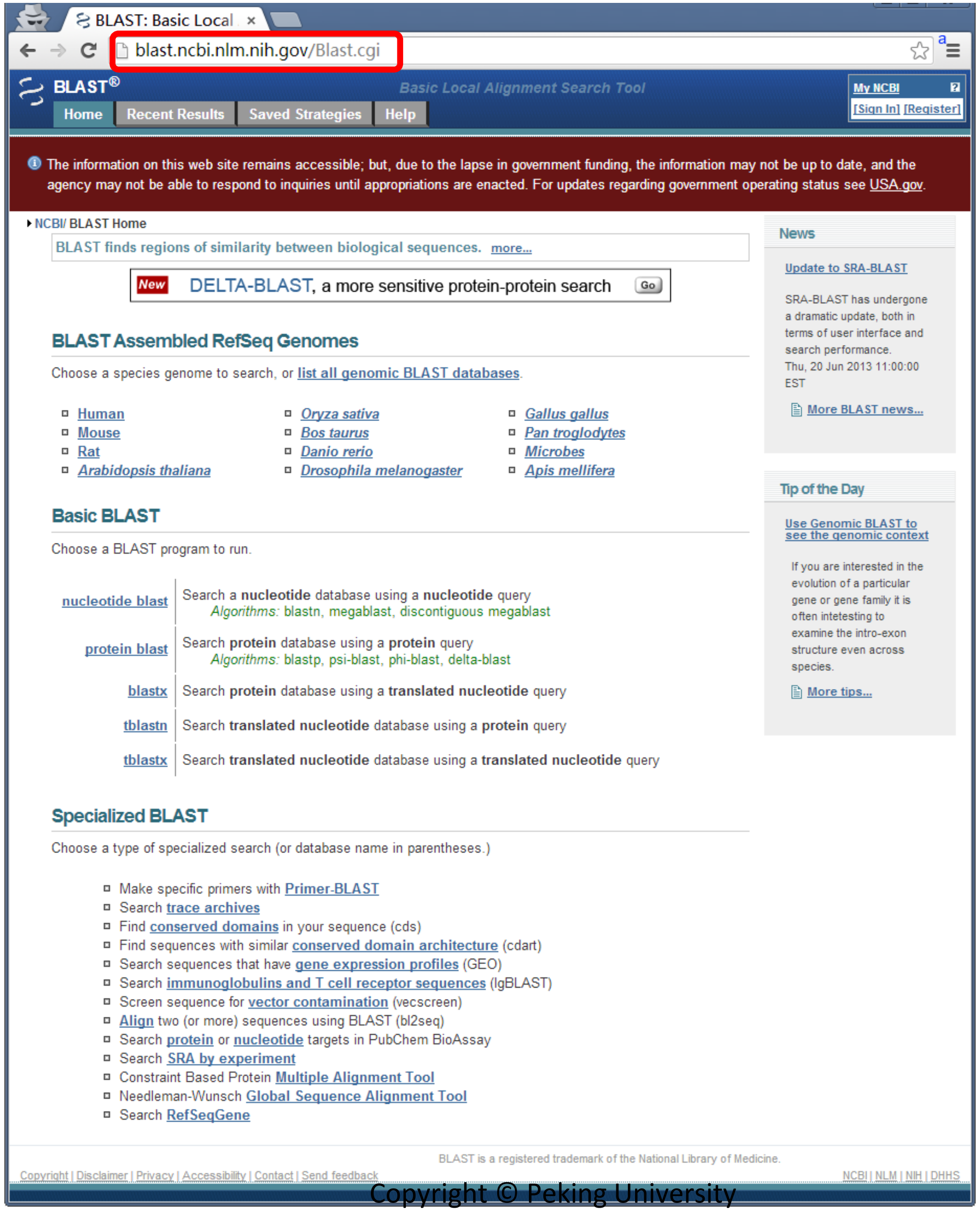
A new approach to rapid sequence comparison, basic local alignment search tool (BLAST), directly approximates alignments that optimize a measure of local similarity, the maximal segment pair (MSP) score. Recent mathematical results on the stochastic properties of MSP scores allow an analysis of the performance of this method as well as the statistical significance of alignments it generates. The basic algorithm is simple and robust; it can be implemented in a number of ways and applied in a variety of contexts including straightforward DNA and protein sequence database searches, motif searches, gene identification searches, and in the analysis of multiple regions of similarity in long DNA sequences. In addition to its flexibility and tractability to mathematical analysis, BLAST is an order of magnitude faster than existing sequence comparison tools of comparable sensitivity.



Basic local alignment search tool

SF Altschul, W Gish, W Miller, EW Myers… - Journal of molecular …, 1990 - Elsevier

A new approach to rapid sequence comparison, **basic local alignment search tool** (BLAST), directly approximates alignments that optimize a measure of **local** similarity, the maximal segment pair (MSP) score. Recent matlmmatical results on the stochastic properties of **...**



The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

New DELTA-BLAST, a more sensitive protein-protein search Go

BLAST Assembled RefSeq Genomes

Choose a species genome to search, or list all genomic BLAST databases.

- Human
- Mouse
- Rat
- *Arabidopsis thaliana*
- *Oryza sativa*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Gallus gallus*
- *Pan troglodytes*
- Microbes
- *Apis mellifera*

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, delta-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 and T cell receptor sequences (IgBLAST)
- Screen sequence for vector contamination (vecscreen)
- Align two (or more) sequences using BLAST (bl2seq)
- Search protein or nucleotide targets in PubChem BioAssay
- Search SRA by experiment
- Constraint Based Protein Multiple Alignment Tool
- Needleman-Wunsch Global Sequence Alignment Tool
- Search RefSeqGene

News

Update to SRA-BLAST

SRA-BLAST has undergone a dramatic update, both in terms of user interface and search performance.
Thu, 20 Jun 2013 11:00:00 EST

More BLAST news...

Tip of the Day

Use Genomic BLAST to see the genomic context

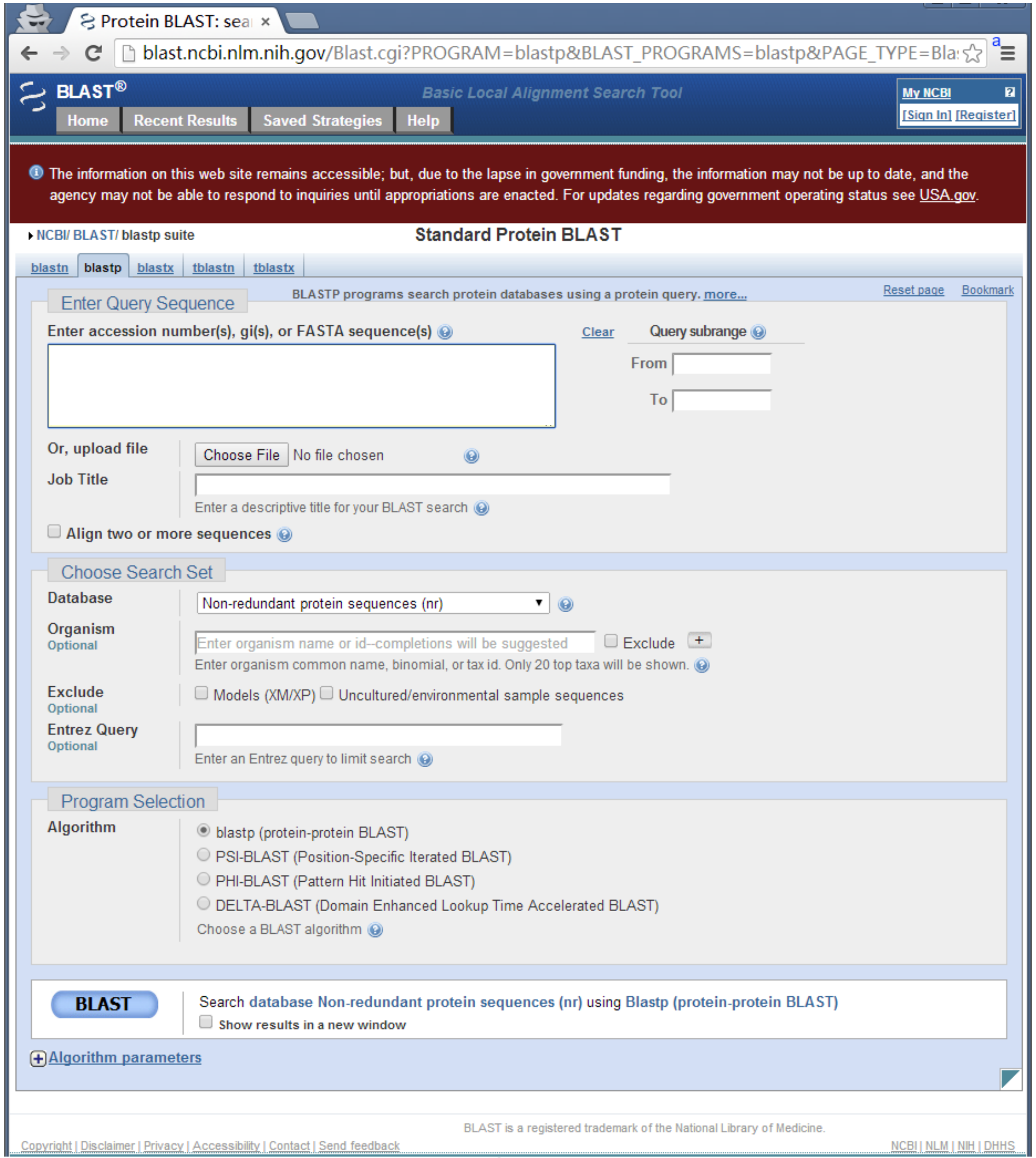
If you are interested in the evolution of a particular gene or gene family it is often intetesting to examine the intro-exon structure even across species.

More tips...

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

NCBI | NLM | NIH | DHHS


Protein BLAST: sea ×
blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=Bla
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]

The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST/ blastp suite

Standard Protein BLAST

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

Query subrange

From

To

Or, upload file: Choose File No file chosen

Job Title

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 □ Exclude +

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

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

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)

○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST

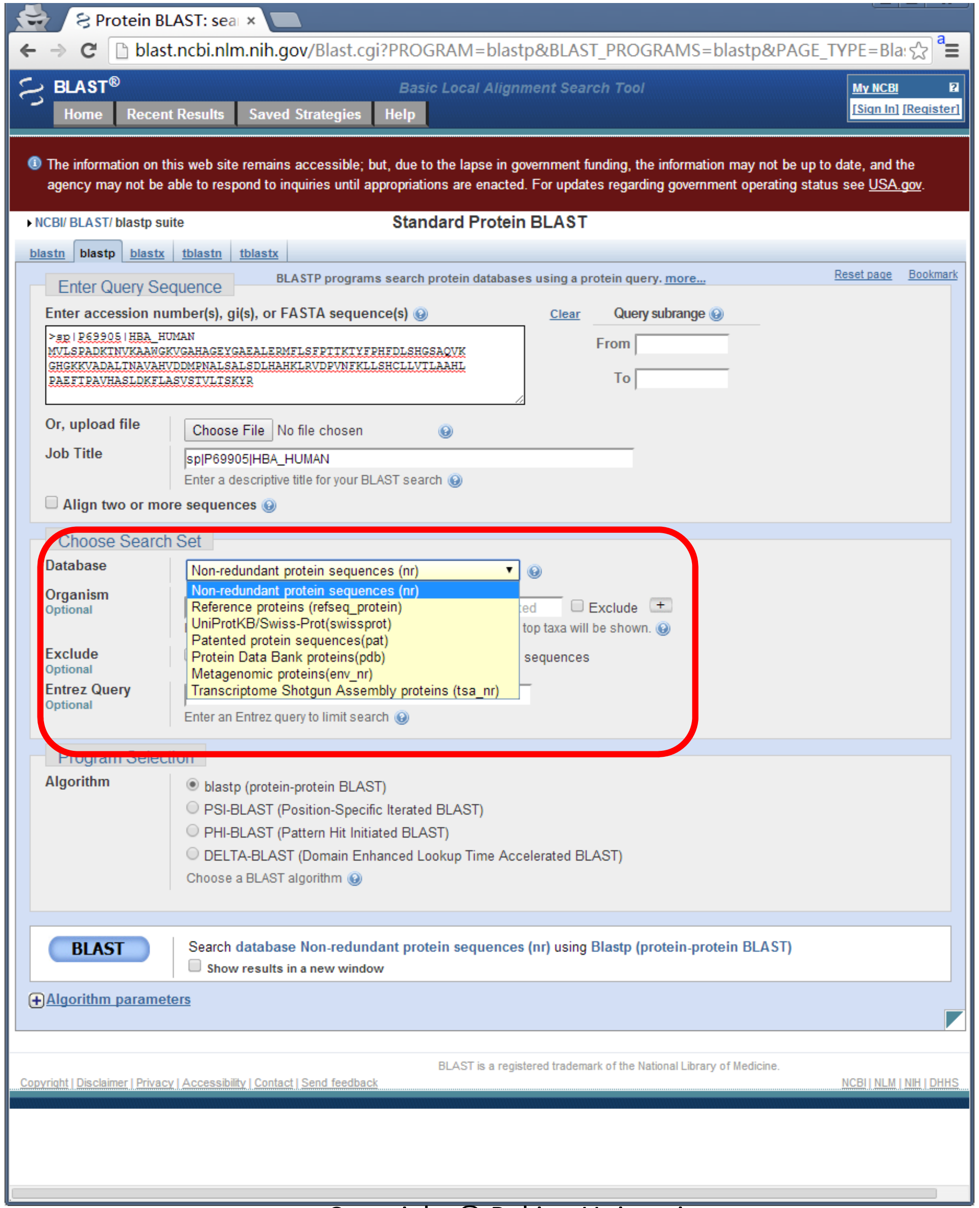
Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

□ Show results in a new window

+ Algorithm parameters

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback
NCBI | NLM | NIH | DHHS



Protein BLAST: sea
blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=Bla
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]

The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST/ blastp suite

Standard Protein BLAST

blastn
blastp
blastx
tblastn
tblastx
Reset page
Bookmark

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>sp|P69905|HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR
```

Query subrange
From
To

Or, upload file: Choose File No file chosen

Job Title: sp|P69905|HBA_HUMAN

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: Non-redundant protein sequences (nr)

Non-redundant protein sequences (nr)
Reference proteins (refseq_protein)
UniProtKB/Swiss-Prot(swissprot)
Patented protein sequences(pat)
Protein Data Bank proteins(pdb)
Metagenomic proteins(env_nr)
Transcriptome Shotgun Assembly proteins (tsa_nr)

Organism
Optional

□ Exclude

top taxa will be shown.

Exclude
Optional

sequences

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)
- ○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

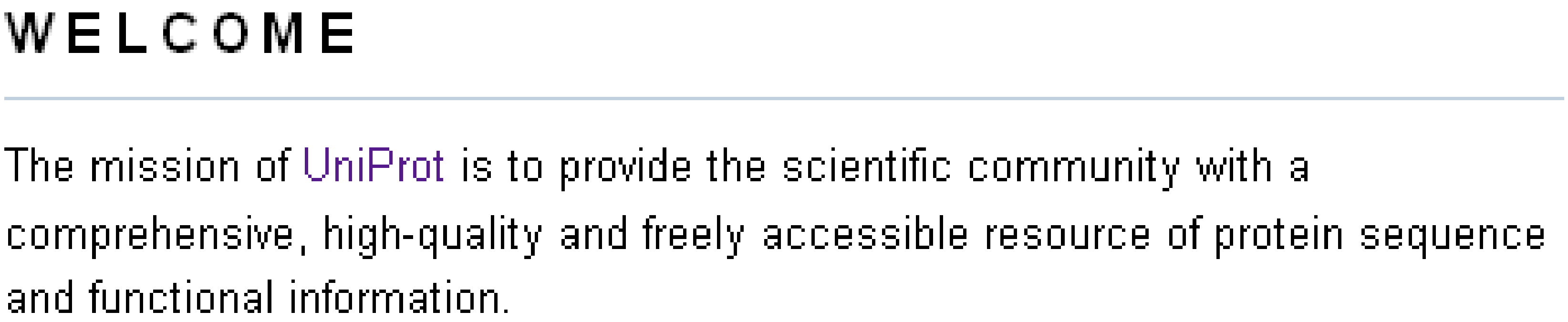
Choose a BLAST algorithm

Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

□ Show results in a new window

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback
NCBI | NLM | NIH | DHHS



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.

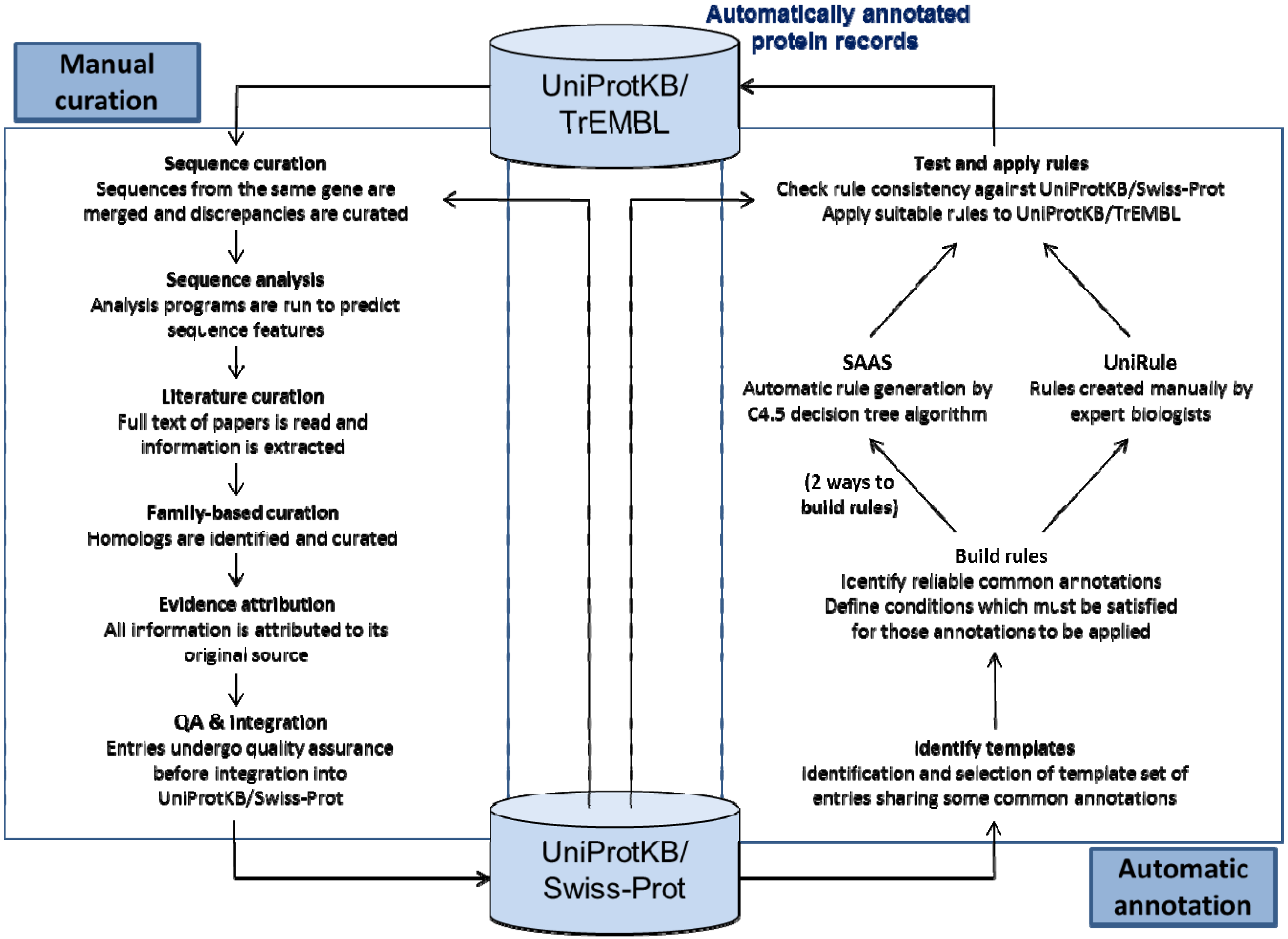
data

protein sequence

knowledge

functional information

(Modified from http://education.expasy.org/cours/Turin/UniProtKB_Turin.ppt)



Manual curation

UniProtKB/ TrEMBL

Automatically annotated protein records

Sequence curation
Sequences from the same gene are merged and discrepancies are curated

Sequence analysis
Analysis programs are run to predict sequence features

Literature curation
Full text of papers is read and information is extracted

Family-based curation
Homologs are identified and curated

Evidence attribution
All information is attributed to its original source

QA & integration
Entries undergo quality assurance before integration into UniProtKB/Swiss-Prot

UniProtKB/ Swiss-Prot

Test and apply rules
Check rule consistency against UniProtKB/Swiss-Prot
Apply suitable rules to UniProtKB/TrEMBL

SAAS
Automatic rule generation by C4.5 decision tree algorithm

UniRule
Rules created manually by expert biologists

(2 ways to build rules)

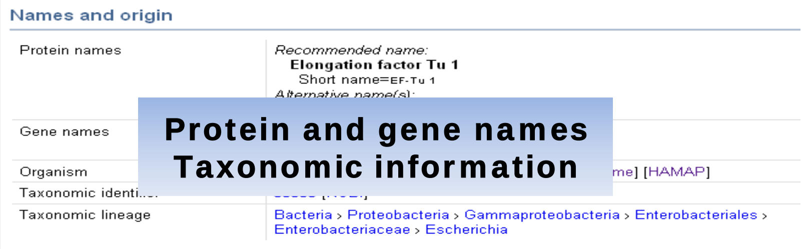
Build rules
Identify reliable common annotations
Define conditions which must be satisfied for those annotations to be applied

Identify templates
Identification and selection of template set of entries sharing some common annotations

Automatic annotation

(Modified from http://education.expasy.org/cours/Turin/UniProtKB_Turin.ppt)

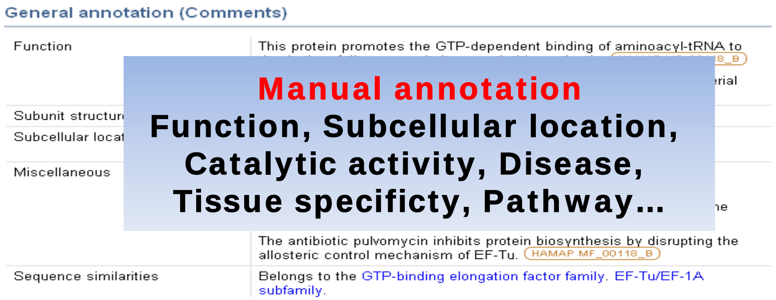
(Modified from http://education.expasy.org/cours/Turin/UniProtKB_Turin.ppt)



Names and origin

Protein names	*Recommended name:* **Elongation factor Tu 1** Short name=EF-Tu 1 *Alternative name(s):*
Gene names	
Organism	me] [HAMAP]
Taxonomic identi...	
Taxonomic lineage	Bacteria › Proteobacteria › Gammaproteobacteria › Enterobacteriales › Enterobacteriaceae › Escherichia

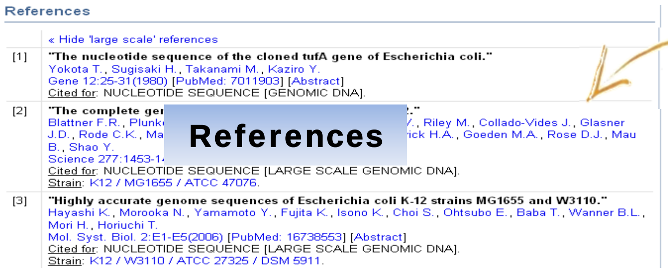
Protein and gene names
Taxonomic information



General annotation (Comments)

Function	This protein promotes the GTP-dependent binding of aminoacyl-tRNA to 8_B erial
Subunit structur	
Subcellular locat	
Miscellaneous	ne The antibiotic pulvomycin inhibits protein biosynthesis by disrupting the allosteric control mechanism of EF-Tu. HAMAP MF_00118_B
Sequence similarities	Belongs to the GTP-binding elongation factor family. EF-Tu/EF-1A subfamily.

Manual annotation
Function, Subcellular location, Catalytic activity, Disease, Tissue specificty, Pathway…



References

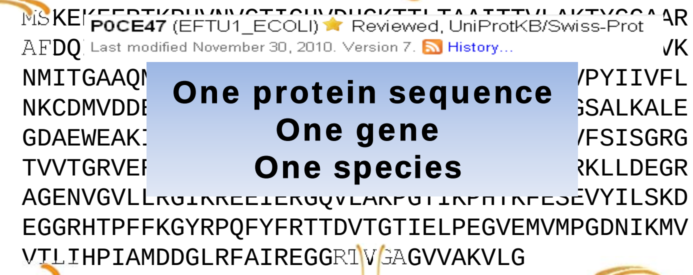
« Hide 'large scale' references

[1] **"The nucleotide sequence of the cloned tufA gene of Escherichia coli."**
Yokota T., Sugisaki H., Takanami M., Kaziro Y.
Gene 12:25-31(1980) [PubMed: 7011903] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA].

[2] **"The complete ge**
Blattner F.R., Plunk ... Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Ma ... ick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.
Science 277:1453-1
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
Strain: K12 / MG1655 / ATCC 47076.

[3] **"Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110."**
Hayashi K., Morooka N., Yamamoto Y., Fujita K., Isono K., Choi S., Ohtsubo E., Baba T., Wanner B.L., Mori H., Horiuchi T.
Mol. Syst. Biol. 2:E1-E5(2006) [PubMed: 16738553] [Abstract]
Cited for: NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
Strain: K12 / W3110 / ATCC 27325 / DSM 5911.

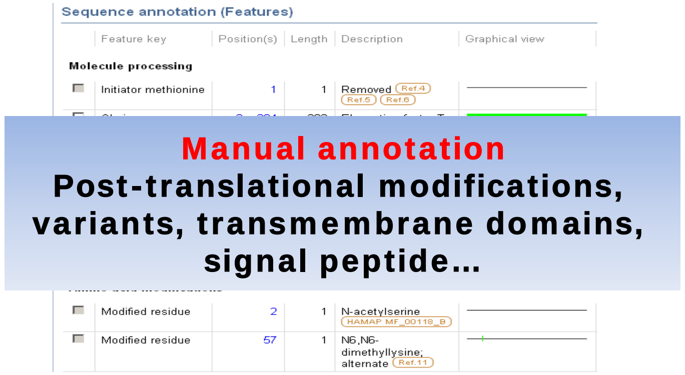
References



P0CE47 (EFTU1_ECOLI) ⭐ Reviewed, UniProtKB/Swiss-Prot
Last modified November 30, 2010. Version 7. History...

```
MSKE                                      AR
AFDQ                                      VK
NMITGAAQ                            VPYIIVFL
NKCDMVDDE                           GSALKALE
GDAEWEAKI                           VFSISGRG
TVVTGRVER                           RKLLDEGR
AGENVGVL                        SEVYILSKD
EGGRHTPFFKGYRPQFYFRTTDVTGTIELPEGVEMVMPGDNIKMV
VTLIHPIAMDDGLRFAIREGGRTVGAGVVAKVLG
```

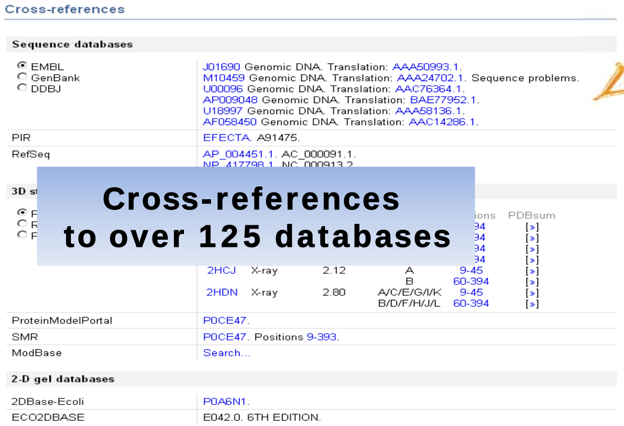
One protein sequence
One gene
One species



Sequence annotation (Features)

	Feature key	Position(s)	Length	Description	Graphical view
Molecule processing					
☐	Initiator methionine	1	1	Removed Ref.4 Ref.5 Ref.6	
☐	Modified residue	2	1	N-acetylserine HAMAP MF_00118_B	
☐	Modified residue	57	1	N6,N6-dimethyllysine; alternate Ref.11	

Manual annotation
Post-translational modifications, variants, transmembrane domains, signal peptide…



Cross-references

Sequence databases

● EMBL ○ GenBank ○ DDBJ	J01690 Genomic DNA. Translation: AAA50993.1. M10459 Genomic DNA. Translation: AAA24702.1. Sequence problems. U00096 Genomic DNA. Translation: AAC76364.1. AP009048 Genomic DNA. Translation: BAE77952.1. U18997 Genomic DNA. Translation: AAA58136.1. AF058450 Genomic DNA. Translation: AAC14286.1.
PIR	EFECTA. A91475.
RefSeq	AP_004451.1. AC_000091.1. NP_417798.1. NC_000913.2.

3D st

					ions	PDBsum
● F ○ R ○ F					94	[»]
					94	[»]
					94	[»]
					94	[»]
	2HCJ	X-ray	2.12	A	9-45	[»]
				B	60-394	[»]
	2HDN	X-ray	2.80	A/C/E/G/I/K	9-45	[»]
				B/D/F/H/J/L	60-394	[»]

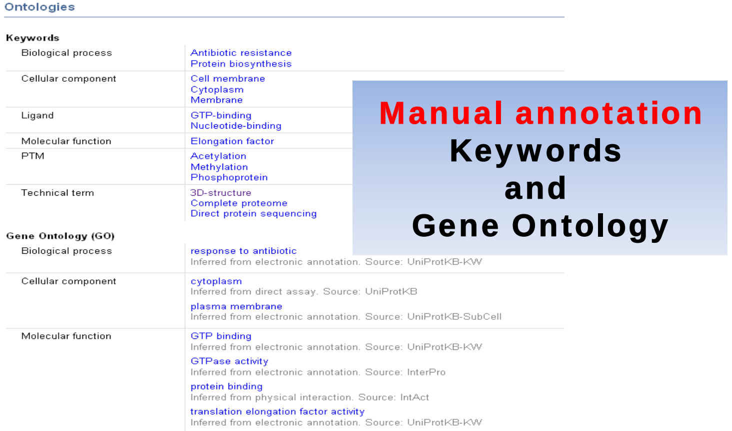
ProteinModelPortal	P0CE47.
SMR	P0CE47. Positions 9-393.
ModBase	Search...

2-D gel databases

2DBase-Ecoli	P0A6N1.
ECO2DBASE	E042.0. 6TH EDITION.

Cross-references
to over 125 databases

Alternative products:
protein sequences produced by alternative splicing, alternative promoter usage, alternative initiation…



Ontologies

Keywords

Biological process	Antibiotic resistance Protein biosynthesis
Cellular component	Cell membrane Cytoplasm Membrane
Ligand	GTP-binding Nucleotide-binding
Molecular function	Elongation factor
PTM	Acetylation Methylation Phosphoprotein
Technical term	3D-structure Complete proteome Direct protein sequencing

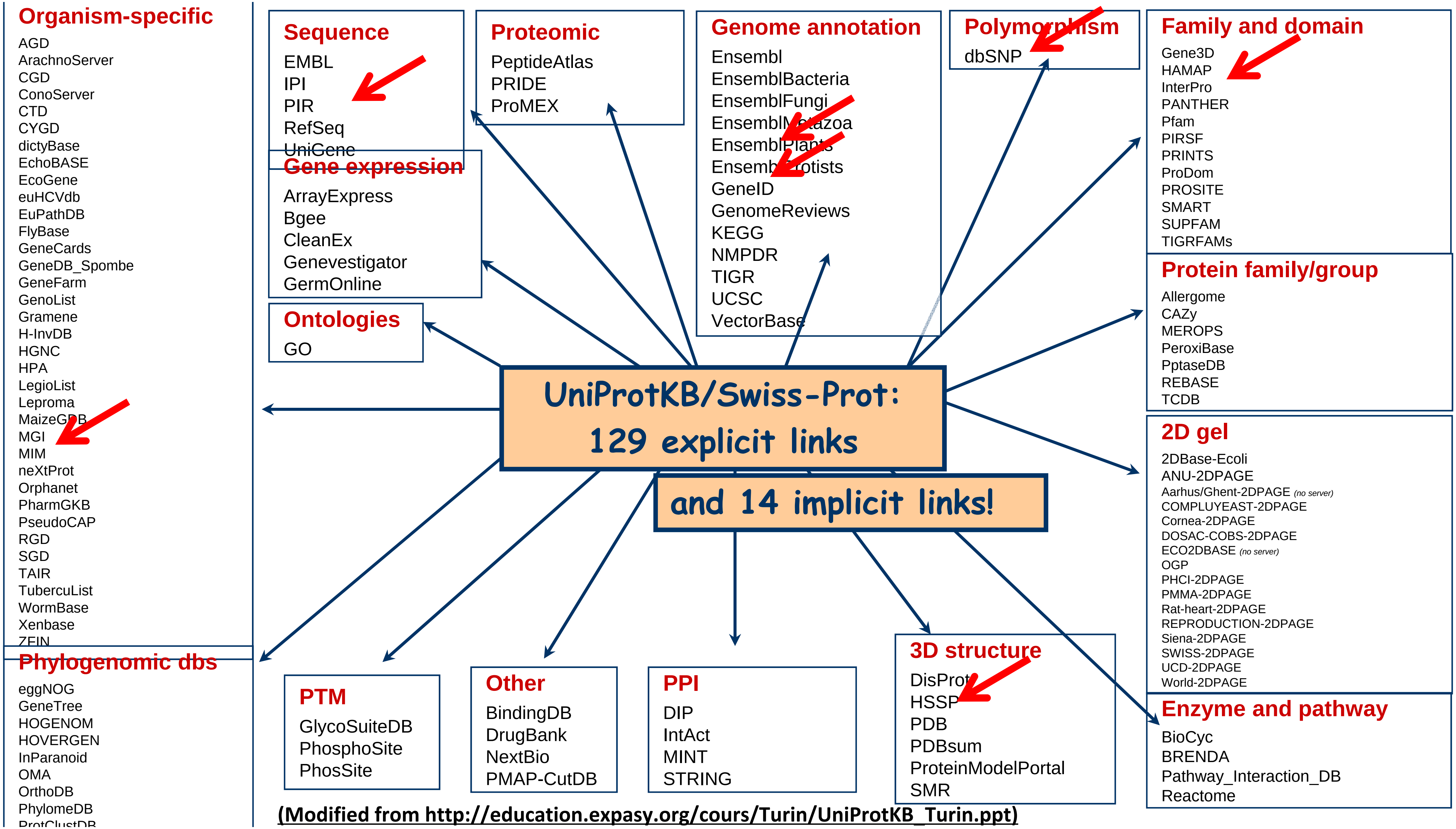
Gene Ontology (GO)

Biological process	response to antibiotic Inferred from electronic annotation. Source: UniProtKB-KW
Cellular component	cytoplasm Inferred from direct assay. Source: UniProtKB plasma membrane Inferred from electronic annotation. Source: UniProtKB-SubCell
Molecular function	GTP binding Inferred from electronic annotation. Source: UniProtKB-KW GTPase activity Inferred from electronic annotation. Source: InterPro protein binding Inferred from physical interaction. Source: IntAct translation elongation factor activity Inferred from electronic annotation. Source: UniProtKB-KW

Manual annotation
Keywords
and
Gene Ontology

UniProtKB/Swiss-Prot

www.uniprot.org



UniProtKB/Swiss-Prot: 129 explicit links

and 14 implicit links!

Organism-specific

AGD
ArachnoServer
CGD
ConoServer
CTD
CYGD
dictyBase
EchoBASE
EcoGene
euHCVdb
EuPathDB
FlyBase
GeneCards
GeneDB_Spombe
GeneFarm
GenoList
Gramene
H-InvDB
HGNC
HPA
LegioList
Leproma
MaizeGDB
MGI
MIM
neXtProt
Orphanet
PharmGKB
PseudoCAP
RGD
SGD
TAIR
TubercuList
WormBase
Xenbase
ZFIN

Phylogenomic dbs

eggNOG
GeneTree
HOGENOM
HOVERGEN
InParanoid
OMA
OrthoDB
PhylomeDB
ProtClustDB

Sequence

EMBL
IPI
PIR
RefSeq
UniGene

Gene expression

ArrayExpress
Bgee
CleanEx
Genevestigator
GermOnline

Ontologies

GO

Proteomic

PeptideAtlas
PRIDE
ProMEX

Genome annotation

Ensembl
EnsemblBacteria
EnsemblFungi
EnsemblMetazoa
EnsemblPlants
EnsemblProtists
GeneID
GenomeReviews
KEGG
NMPDR
TIGR
UCSC
VectorBase

Polymorphism

dbSNP

Family and domain

Gene3D
HAMAP
InterPro
PANTHER
Pfam
PIRSF
PRINTS
ProDom
PROSITE
SMART
SUPFAM
TIGRFAMs

Protein family/group

Allergome
CAZy
MEROPS
PeroxiBase
PptaseDB
REBASE
TCDB

2D gel

2DBase-Ecoli
ANU-2DPAGE
Aarhus/Ghent-2DPAGE *(no server)*
COMPLUYEAST-2DPAGE
Cornea-2DPAGE
DOSAC-COBS-2DPAGE
ECO2DBASE *(no server)*
OGP
PHCI-2DPAGE
PMMA-2DPAGE
Rat-heart-2DPAGE
REPRODUCTION-2DPAGE
Siena-2DPAGE
SWISS-2DPAGE
UCD-2DPAGE
World-2DPAGE

Enzyme and pathway

BioCyc
BRENDA
Pathway_Interaction_DB
Reactome

PTM

GlycoSuiteDB
PhosphoSite
PhosSite

Other

BindingDB
DrugBank
NextBio
PMAP-CutDB

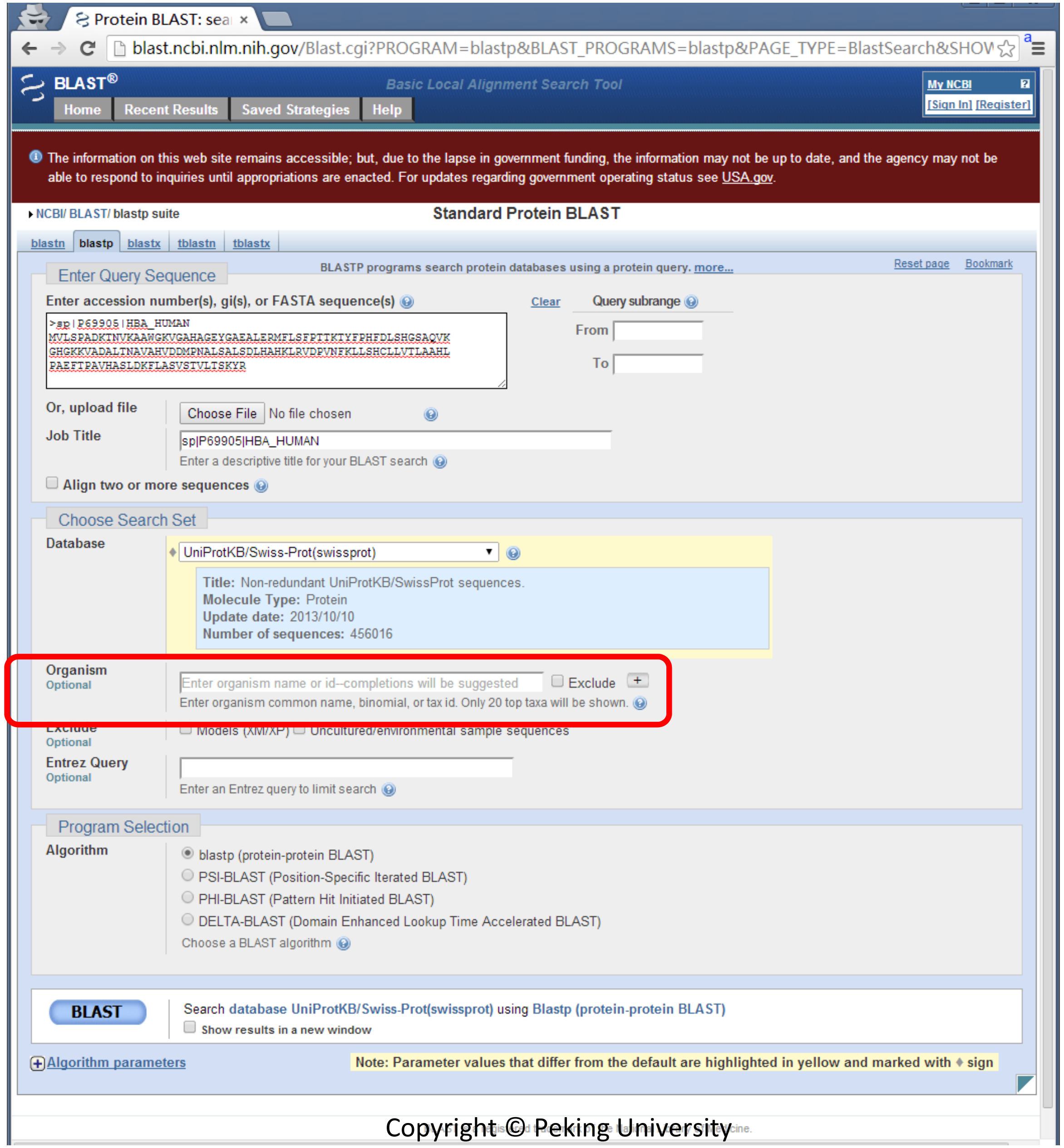
PPI

DIP
IntAct
MINT
STRING

3D structure

DisProt
HSSP
PDB
PDBsum
ProteinModelPortal
SMR

(Modified from http://education.expasy.org/cours/Turin/UniProtKB_Turin.ppt)



Protein BLAST: sea ×
blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]

The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST/ blastp suite
blastn
blastp
blastx
tblastn
tblastx
Reset page
Bookmark

Standard Protein BLAST

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>sp|P69905|HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR
```

Clear

Query subrange

From

To

Or, upload file: Choose File No file chosen

Job Title: sp|P69905|HBA_HUMAN

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: UniProtKB/Swiss-Prot(swissprot)

Title: Non-redundant UniProtKB/SwissProt sequences.
Molecule Type: Protein
Update date: 2013/10/10
Number of sequences: 456016

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

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

Exclude Optional: □ Models (XM/XP) □ Uncultured/environmental sample sequences

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

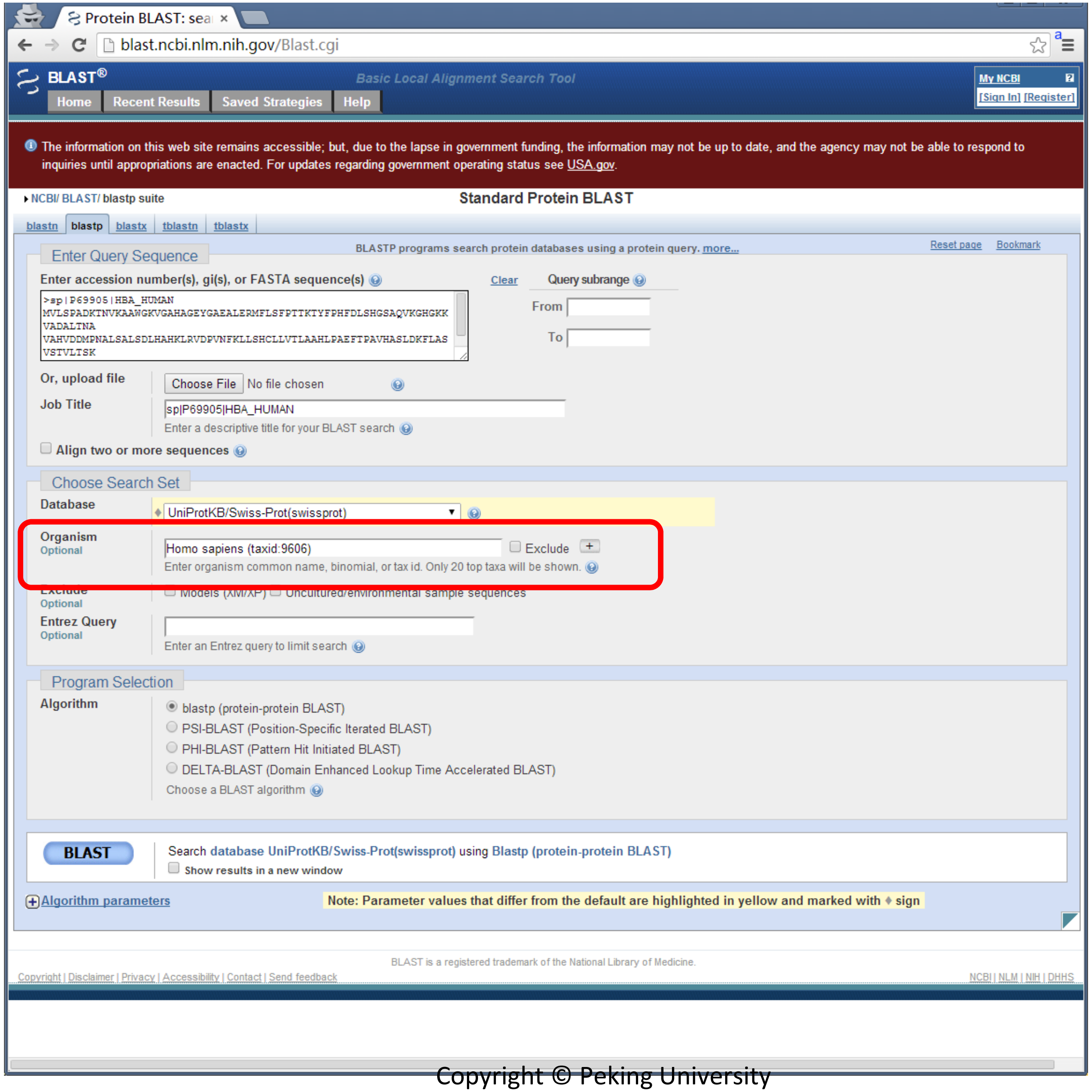
● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

□ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign



The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST/ blastp suite

Standard Protein BLAST

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>sp|P69905|HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKK
VADALTNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLAS
VSTVLTSK
```

Query subrange

From

To

Or, upload file: Choose File No file chosen

Job Title: sp|P69905|HBA_HUMAN

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: UniProtKB/Swiss-Prot(swissprot)

Organism (Optional): Homo sapiens (taxid:9606) □ Exclude +

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

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

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)

○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

□ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

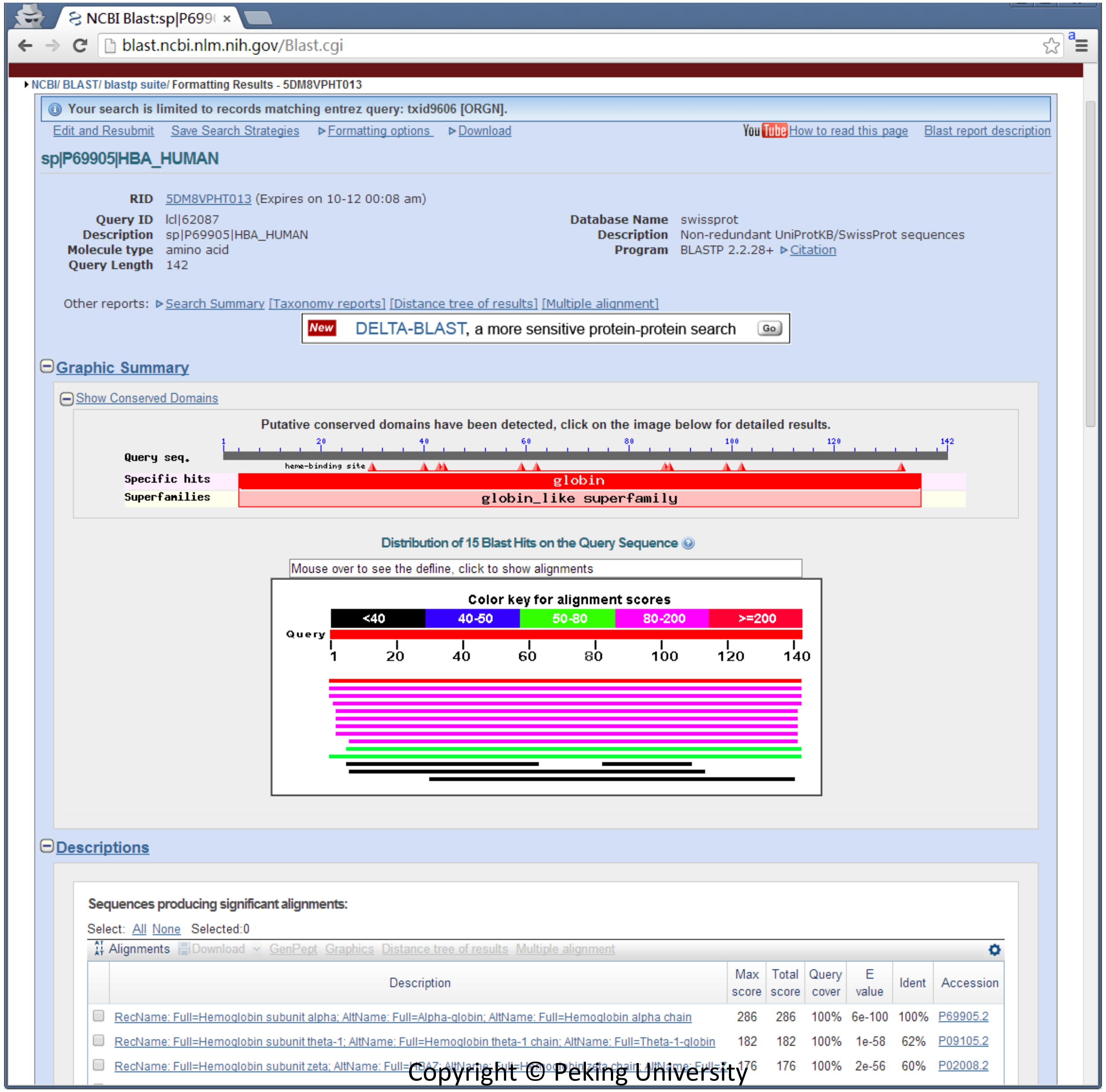
BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback
NCBI | NLM | NIH | DHHS

Summary

Domain(s)

Hits summary



NCBI Blast:sp|P699

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

NCBI/ BLAST/ blastp suite/ Formatting Results - 5DM8VPHT013

Your search is limited to records matching entrez query: txid9606 [ORGN].

Edit and Resubmit   Save Search Strategies   Formatting options   Download   YouTube How to read this page   Blast report description

sp|P69905|HBA_HUMAN

RID	5DM8VPHT013 (Expires on 10-12 00:08 am)		
Query ID	lcl|62087	Database Name	swissprot
Description	sp|P69905|HBA_HUMAN	Description	Non-redundant UniProtKB/SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.2.28+ Citation
Query Length	142		

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]

New DELTA-BLAST, a more sensitive protein-protein search Go

Graphic Summary

Show Conserved Domains

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

Distribution of 15 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments

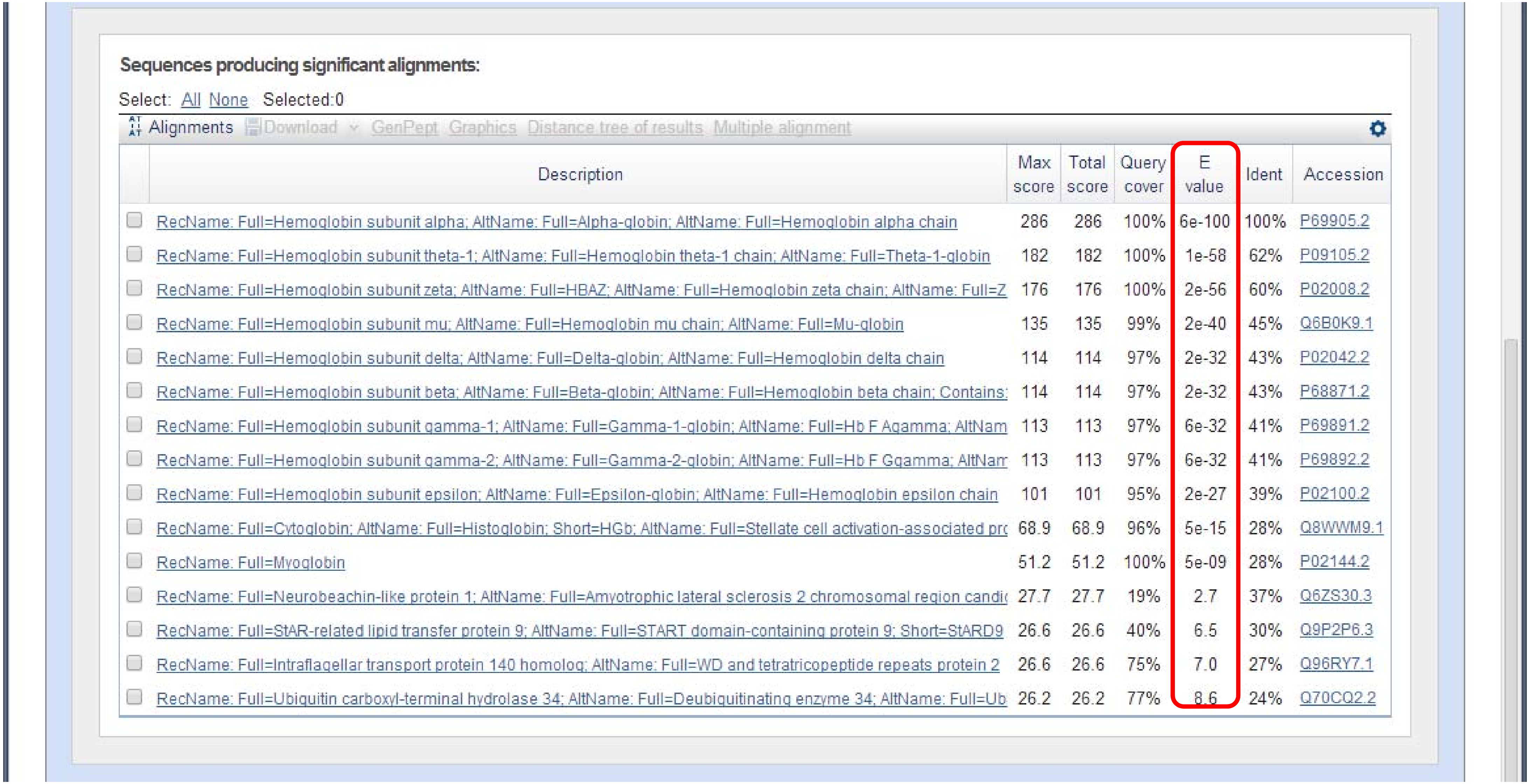
Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
□	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	286	286	100%	6e-100	100%	P69905.2
□	RecName: Full=Hemoglobin subunit theta-1; AltName: Full=Hemoglobin theta-1 chain; AltName: Full=Theta-1-globin	182	182	100%	1e-58	62%	P09105.2
□	RecName: Full=Hemoglobin subunit zeta; AltName: Full=HBAZ; AltName: Full=Hemoglobin zeta chain; AltName: Full=Z	176	176	100%	2e-56	60%	P02008.2

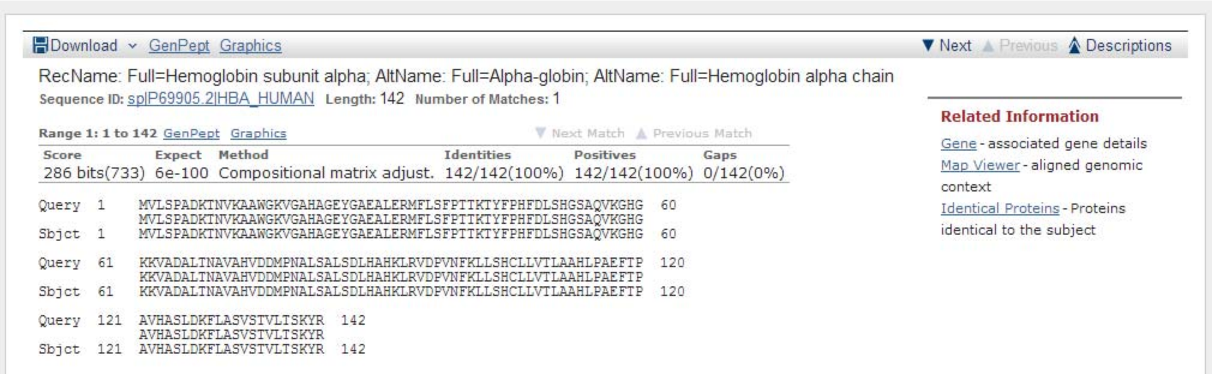


Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
□	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	286	286	100%	6e-100	100%	P69905.2
□	RecName: Full=Hemoglobin subunit theta-1; AltName: Full=Hemoglobin theta-1 chain; AltName: Full=Theta-1-globin	182	182	100%	1e-58	62%	P09105.2
□	RecName: Full=Hemoglobin subunit zeta; AltName: Full=HBAZ; AltName: Full=Hemoglobin zeta chain; AltName: Full=Z	176	176	100%	2e-56	60%	P02008.2
□	RecName: Full=Hemoglobin subunit mu; AltName: Full=Hemoglobin mu chain; AltName: Full=Mu-globin	135	135	99%	2e-40	45%	Q6B0K9.1
□	RecName: Full=Hemoglobin subunit delta; AltName: Full=Delta-globin; AltName: Full=Hemoglobin delta chain	114	114	97%	2e-32	43%	P02042.2
□	RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains:	114	114	97%	2e-32	43%	P68871.2
□	RecName: Full=Hemoglobin subunit gamma-1; AltName: Full=Gamma-1-globin; AltName: Full=Hb F Agamma; AltNam	113	113	97%	6e-32	41%	P69891.2
□	RecName: Full=Hemoglobin subunit gamma-2; AltName: Full=Gamma-2-globin; AltName: Full=Hb F Ggamma; AltNar	113	113	97%	6e-32	41%	P69892.2
□	RecName: Full=Hemoglobin subunit epsilon; AltName: Full=Epsilon-globin; AltName: Full=Hemoglobin epsilon chain	101	101	95%	2e-27	39%	P02100.2
□	RecName: Full=Cytoglobin; AltName: Full=Histoglobin; Short=HGb; AltName: Full=Stellate cell activation-associated prc	68.9	68.9	96%	5e-15	28%	Q8WWM9.1
□	RecName: Full=Myoglobin	51.2	51.2	100%	5e-09	28%	P02144.2
□	RecName: Full=Neurobeachin-like protein 1; AltName: Full=Amyotrophic lateral sclerosis 2 chromosomal region candi	27.7	27.7	19%	2.7	37%	Q6ZS30.3
□	RecName: Full=StAR-related lipid transfer protein 9; AltName: Full=START domain-containing protein 9; Short=StARD9	26.6	26.6	40%	6.5	30%	Q9P2P6.3
□	RecName: Full=Intraflagellar transport protein 140 homolog; AltName: Full=WD and tetratricopeptide repeats protein 2	26.6	26.6	75%	7.0	27%	Q96RY7.1
□	RecName: Full=Ubiquitin carboxyl-terminal hydrolase 34; AltName: Full=Deubiquitinating enzyme 34; AltName: Full=Ub	26.2	26.2	77%	8.6	24%	Q70CQ2.2

HBA_HUMAN



RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain

Sequence ID: sp|P69905.2|HBA_HUMAN **Length:** 142 **Number of Matches:** 1

Range 1: 1 to 142 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
286 bits(733)	6e-100	Compositional matrix adjust.	142/142(100%)	142/142(100%)	0/142(0%)

```
Query  1    MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  60
            MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
Sbjct  1    MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  60

Query  61   KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  120
            KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
Sbjct  61   KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  120

Query  121  AVHASLDKFLASVSTVLTSKYR  142
            AVHASLDKFLASVSTVLTSKYR
Sbjct  121  AVHASLDKFLASVSTVLTSKYR  142
```

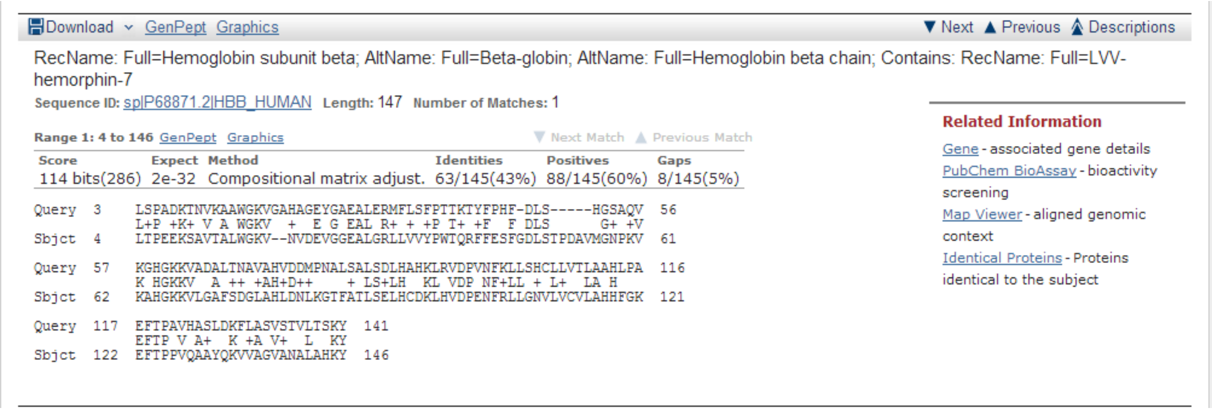
Related Information

Gene - associated gene details

Map Viewer - aligned genomic context

Identical Proteins - Proteins identical to the subject

HBB_HUMAN



RecName: Full=Hemoglobin subunit beta; AltName: Full=Beta-globin; AltName: Full=Hemoglobin beta chain; Contains: RecName: Full=LVV-hemorphin-7

Sequence ID: sp|P68871.2|HBB_HUMAN **Length:** 147 **Number of Matches:** 1

Range 1: 4 to 146 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
114 bits(286)	2e-32	Compositional matrix adjust.	63/145(43%)	88/145(60%)	8/145(5%)

```
Query  3    LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQV  56
            L+P +K+ V A WGKV  +  E G EAL R+ + +P T+ +F  F DLS      G+ +V
Sbjct  4    LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61

Query  57   KGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA  116
            K HGKKV  A ++ +AH+D++     + LS+LH  KL VDP NF+LL + L+  LA H
Sbjct  62   KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK  121

Query  117  EFTPAVHASLDKFLASVSTVLTSKY  141
            EFTP V A+  K +A V+  L  KY
Sbjct  122  EFTPPVQAAYQKVVAGVANALAHKY  146
```

Related Information

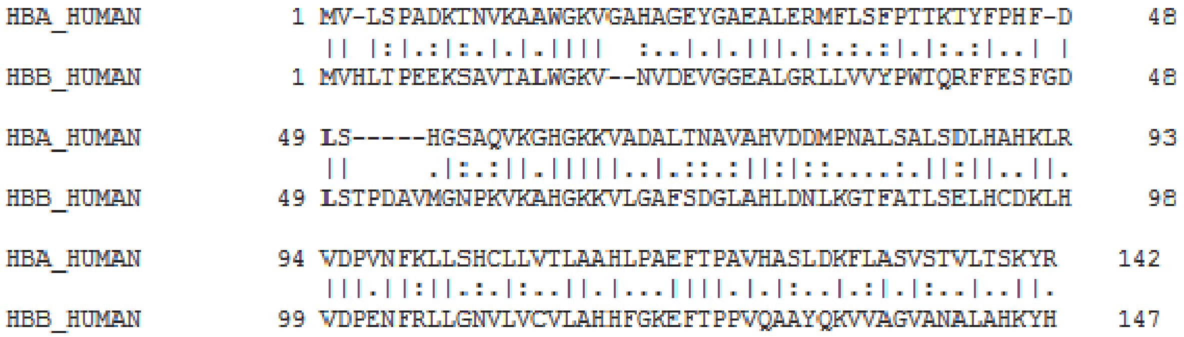
Gene - associated gene details

PubChem BioAssay - bioactivity screening

Map Viewer - aligned genomic context

Identical Proteins - Proteins identical to the subject

Needleman-Wunsch
Global alignment

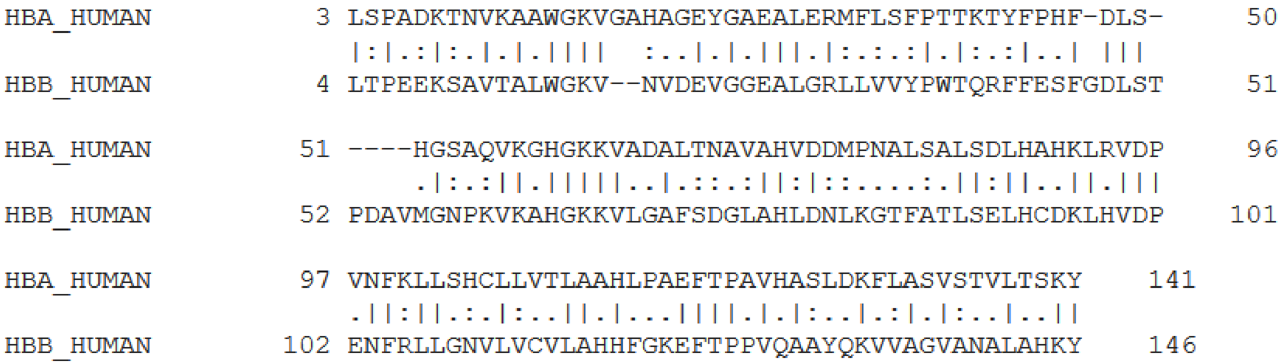


```
HBA_HUMAN           1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D     48
                      || |:|.:|:.|.|.||||  :..|.|.|||.|:.:.:|.|:.:|..| |
HBB_HUMAN           1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

HBA_HUMAN          49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR     93
                      ||     .|:.:||.|||||..|.::.:||:|::....:.||:||..||.
HBB_HUMAN          49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

HBA_HUMAN          94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142
                      |||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||.
HBB_HUMAN          99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

Smith-Waterman
Local alignment

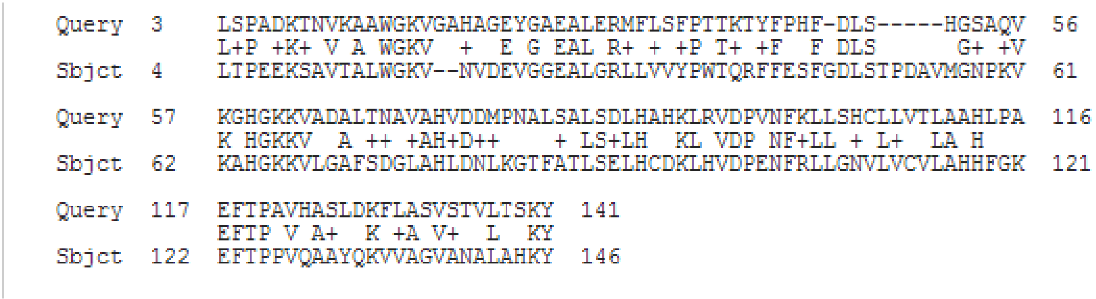


```
HBA_HUMAN           3 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-     50
                      |:|.:|:.|.|.||||  :..|.|.|||.|:.:.:|.|:.:|..| |||
HBB_HUMAN           4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST     51

HBA_HUMAN          51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP     96
                          .|:.:||.|||||..|.::.:||:|::....:.||:||..||.|||
HBB_HUMAN          52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP    101

HBA_HUMAN          97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY    141
                      .||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||
HBB_HUMAN         102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY    146
```

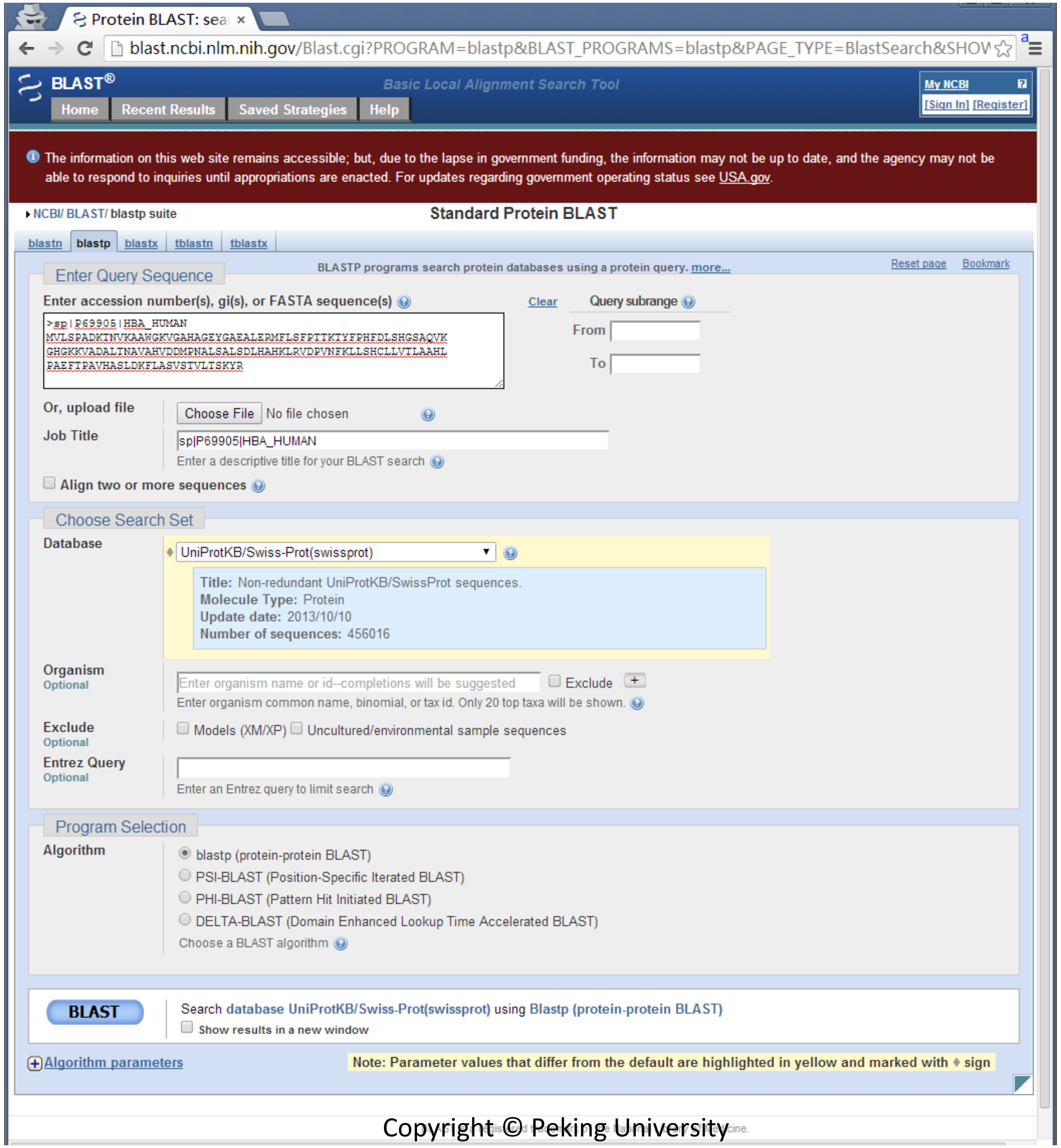
BLAST
Local alignment



```
Query  3     LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQV  56
             L+P +K+ V A WGKV  +  E G EAL R+ + +P T+ +F  F DLS      G+ +V
Sbjct  4     LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61

Query  57    KGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA  116
             K HGKKV  A ++ +AH+D++     + LS+LH  KL VDP NF+LL + L+  LA H
Sbjct  62    KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK  121

Query  117   EFTPAVHASLDKFLASVSTVLTSKY  141
             EFTP V A+  K +A V+  L  KY
Sbjct  122   EFTPPVQAAYQKVVAGVANALAHKY  146
```


Protein BLAST: sea ×
blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW
BLAST®
Basic Local Alignment Search Tool
Home
Recent Results
Saved Strategies
Help
My NCBI
[Sign In] [Register]

The information on this web site remains accessible; but, due to the lapse in government funding, the information may not be up to date, and the agency may not be able to respond to inquiries until appropriations are enacted. For updates regarding government operating status see USA.gov.

NCBI/ BLAST/ blastp suite

Standard Protein BLAST

blastn
blastp
blastx
tblastn
tblastx
Reset page
Bookmark

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

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>sp|P69905|HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVK
GHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHL
PAEFTPAVHASLDKFLASVSTVLTSKYR
```

Query subrange

From

To

Or, upload file: Choose File No file chosen

Job Title: sp|P69905|HBA_HUMAN

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: UniProtKB/Swiss-Prot(swissprot)

Title: Non-redundant UniProtKB/SwissProt sequences.
Molecule Type: Protein
Update date: 2013/10/10
Number of sequences: 456016

Organism (Optional): Enter organism name or id--completions will be suggested □ Exclude

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

Exclude (Optional): □ Models (XM/XP) □ Uncultured/environmental sample sequences

Entrez Query (Optional):

Enter an Entrez query to limit search

Program Selection

Algorithm

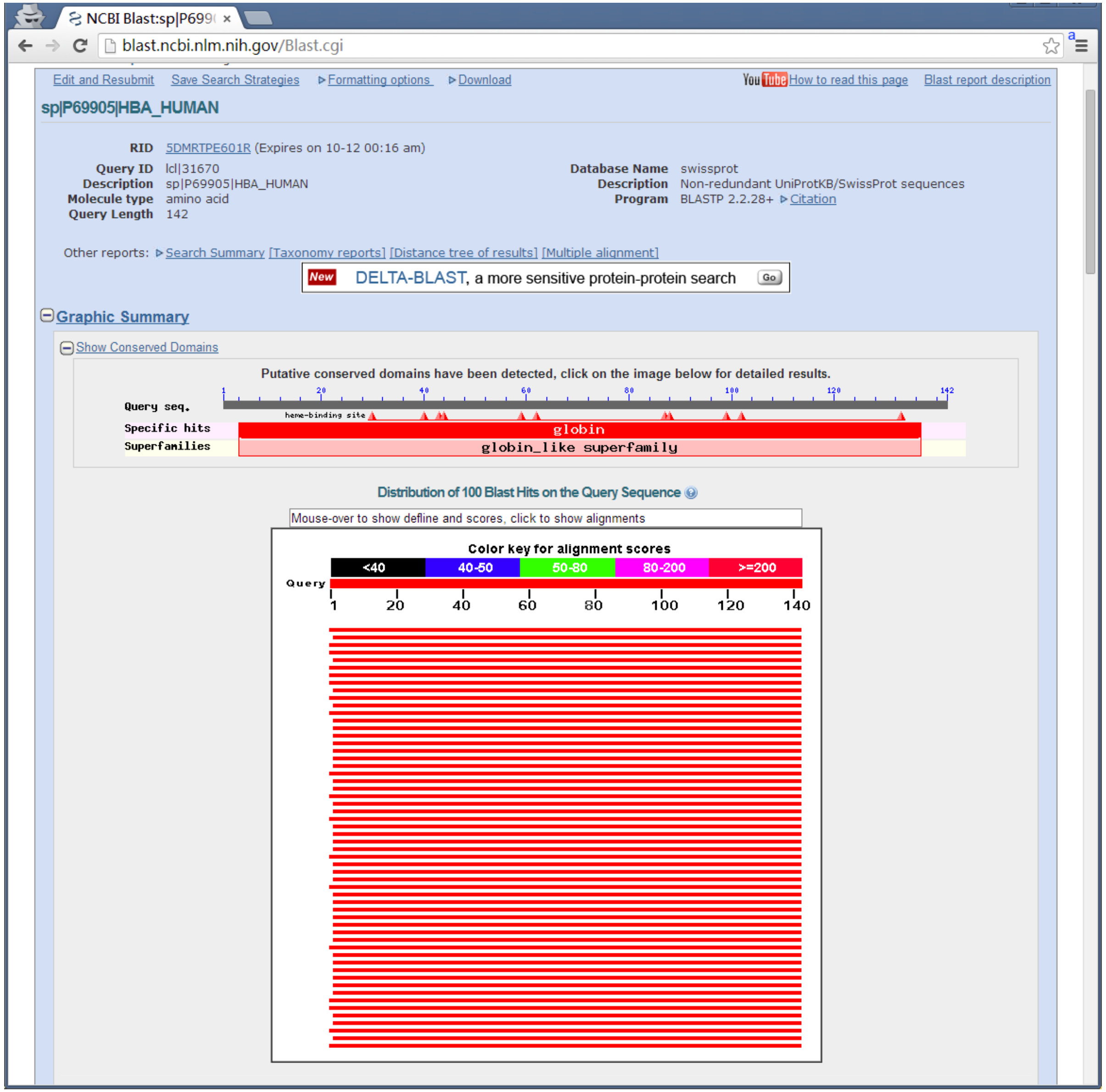
● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

Search database UniProtKB/Swiss-Prot(swissprot) using Blastp (protein-protein BLAST)

□ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign



NCBI Blast:sp|P699(
blast.ncbi.nlm.nih.gov/Blast.cgi
Edit and Resubmit
Save Search Strategies
Formatting options
Download
YouTube How to read this page
Blast report description

sp|P69905|HBA_HUMAN

RID 5DMRTPE601R (Expires on 10-12 00:16 am)

Query ID	lcl|31670	Database Name	swissprot
Description	sp|P69905|HBA_HUMAN	Description	Non-redundant UniProtKB/SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.2.28+ ▷Citation
Query Length	142		

Other reports: ▷Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]

New
DELTA-BLAST, a more sensitive protein-protein search
Go

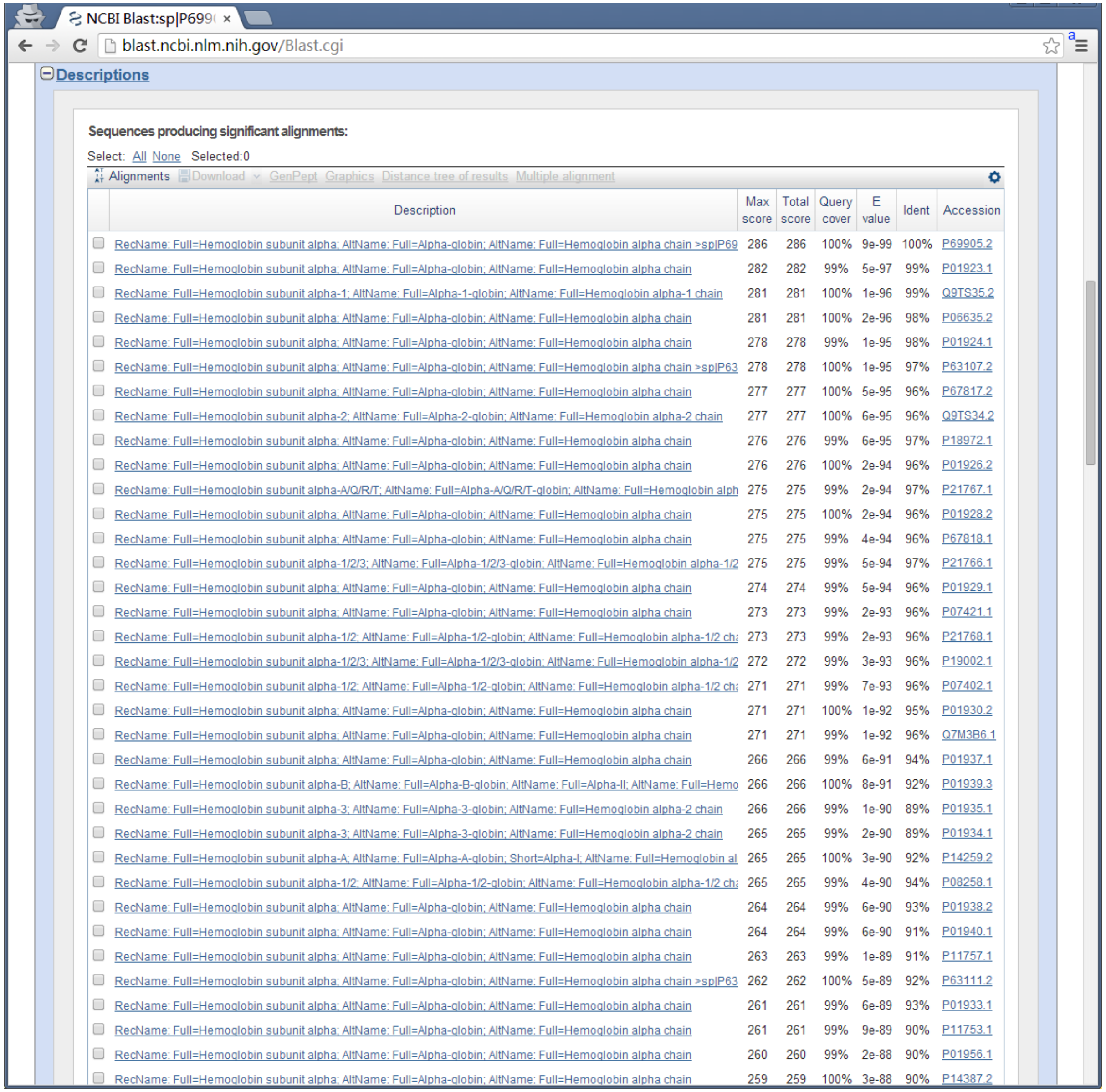
⊟ Graphic Summary

⊟ 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 show defline and scores, click to show alignments



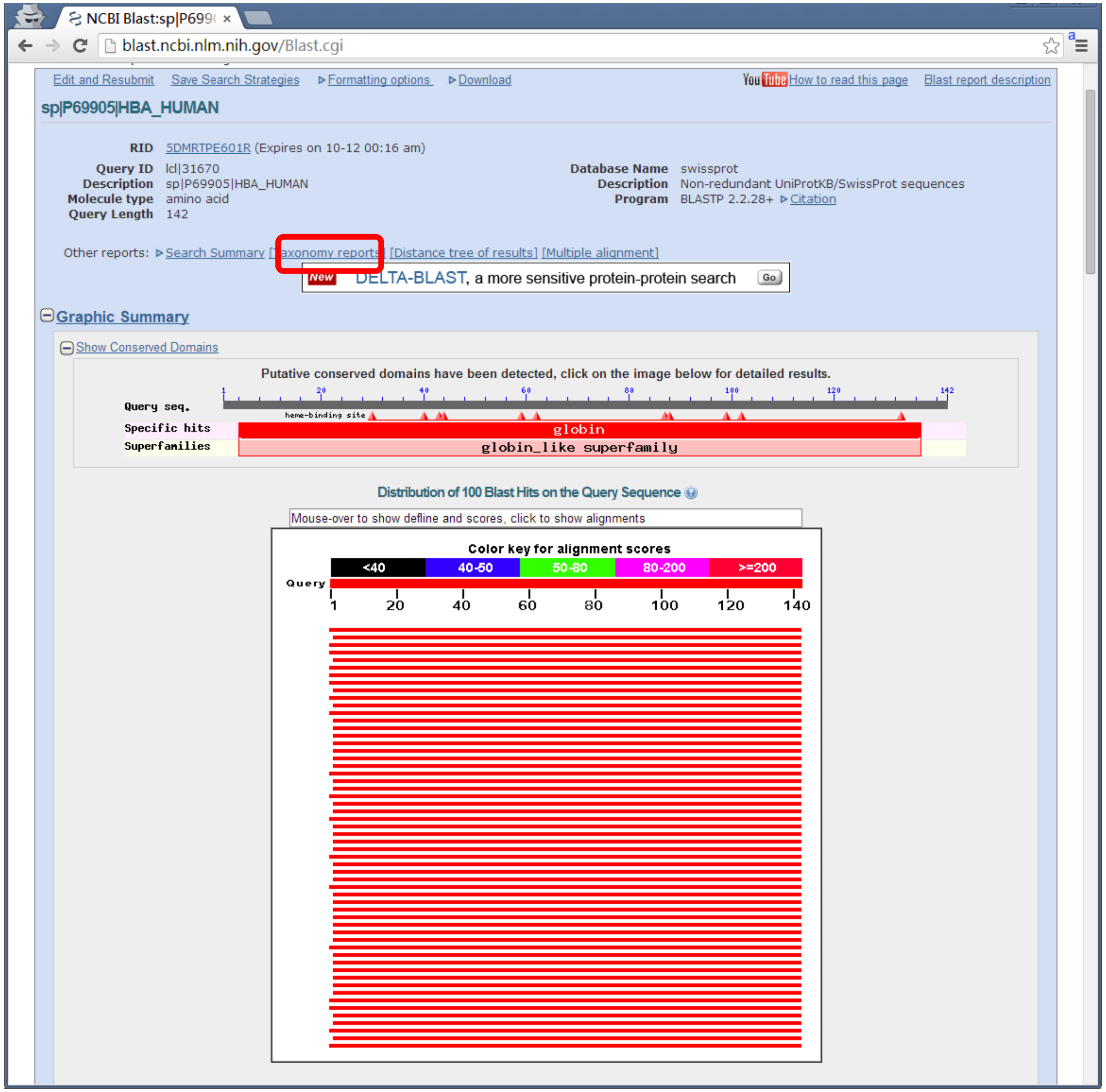
⊟ Descriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain >sp|P69	286	286	100%	9e-99	100%	P69905.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	282	282	99%	5e-97	99%	P01923.1
☐	RecName: Full=Hemoglobin subunit alpha-1; AltName: Full=Alpha-1-globin; AltName: Full=Hemoglobin alpha-1 chain	281	281	100%	1e-96	99%	Q9TS35.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	281	281	100%	2e-96	98%	P06635.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	278	278	99%	1e-95	98%	P01924.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain >sp|P63	278	278	100%	1e-95	97%	P63107.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	277	277	100%	5e-95	96%	P67817.2
☐	RecName: Full=Hemoglobin subunit alpha-2; AltName: Full=Alpha-2-globin; AltName: Full=Hemoglobin alpha-2 chain	277	277	100%	6e-95	96%	Q9TS34.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	276	276	99%	6e-95	97%	P18972.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	276	276	100%	2e-94	96%	P01926.2
☐	RecName: Full=Hemoglobin subunit alpha-A/Q/R/T; AltName: Full=Alpha-A/Q/R/T-globin; AltName: Full=Hemoglobin alph	275	275	99%	2e-94	97%	P21767.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	275	275	100%	2e-94	96%	P01928.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	275	275	99%	4e-94	96%	P67818.1
☐	RecName: Full=Hemoglobin subunit alpha-1/2/3; AltName: Full=Alpha-1/2/3-globin; AltName: Full=Hemoglobin alpha-1/2	275	275	99%	5e-94	97%	P21766.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	274	274	99%	5e-94	96%	P01929.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	273	273	99%	2e-93	96%	P07421.1
☐	RecName: Full=Hemoglobin subunit alpha-1/2; AltName: Full=Alpha-1/2-globin; AltName: Full=Hemoglobin alpha-1/2 ch	273	273	99%	2e-93	96%	P21768.1
☐	RecName: Full=Hemoglobin subunit alpha-1/2/3; AltName: Full=Alpha-1/2/3-globin; AltName: Full=Hemoglobin alpha-1/2	272	272	99%	3e-93	96%	P19002.1
☐	RecName: Full=Hemoglobin subunit alpha-1/2; AltName: Full=Alpha-1/2-globin; AltName: Full=Hemoglobin alpha-1/2 ch	271	271	99%	7e-93	96%	P07402.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	271	271	100%	1e-92	95%	P01930.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	271	271	99%	1e-92	96%	Q7M3B6.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	266	266	99%	6e-91	94%	P01937.1
☐	RecName: Full=Hemoglobin subunit alpha-B; AltName: Full=Alpha-B-globin; AltName: Full=Alpha-II; AltName: Full=Hemo	266	266	100%	8e-91	92%	P01939.3
☐	RecName: Full=Hemoglobin subunit alpha-3; AltName: Full=Alpha-3-globin; AltName: Full=Hemoglobin alpha-2 chain	266	266	99%	1e-90	89%	P01935.1
☐	RecName: Full=Hemoglobin subunit alpha-3; AltName: Full=Alpha-3-globin; AltName: Full=Hemoglobin alpha-2 chain	265	265	99%	2e-90	89%	P01934.1
☐	RecName: Full=Hemoglobin subunit alpha-A; AltName: Full=Alpha-A-globin; Short=Alpha-I; AltName: Full=Hemoglobin al	265	265	100%	3e-90	92%	P14259.2
☐	RecName: Full=Hemoglobin subunit alpha-1/2; AltName: Full=Alpha-1/2-globin; AltName: Full=Hemoglobin alpha-1/2 ch	265	265	99%	4e-90	94%	P08258.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	264	264	99%	6e-90	93%	P01938.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	264	264	99%	6e-90	91%	P01940.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	263	263	99%	1e-89	91%	P11757.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain >sp|P63	262	262	100%	5e-89	92%	P63111.2
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	261	261	99%	6e-89	93%	P01933.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	261	261	99%	9e-89	90%	P11753.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	260	260	99%	2e-88	90%	P01956.1
☐	RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain	259	259	100%	3e-88	90%	P14387.2



sp|P69905|HBA_HUMAN

RID 5DMRTPE601R (Expires on 10-12 00:16 am)

Query ID lcl|31670
Description sp|P69905|HBA_HUMAN
Molecule type amino acid
Query Length 142

Database Name swissprot
Description Non-redundant UniProtKB/SwissProt sequences
Program BLASTP 2.2.28+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results] [Multiple alignment]

New DELTA-BLAST, a more sensitive protein-protein search

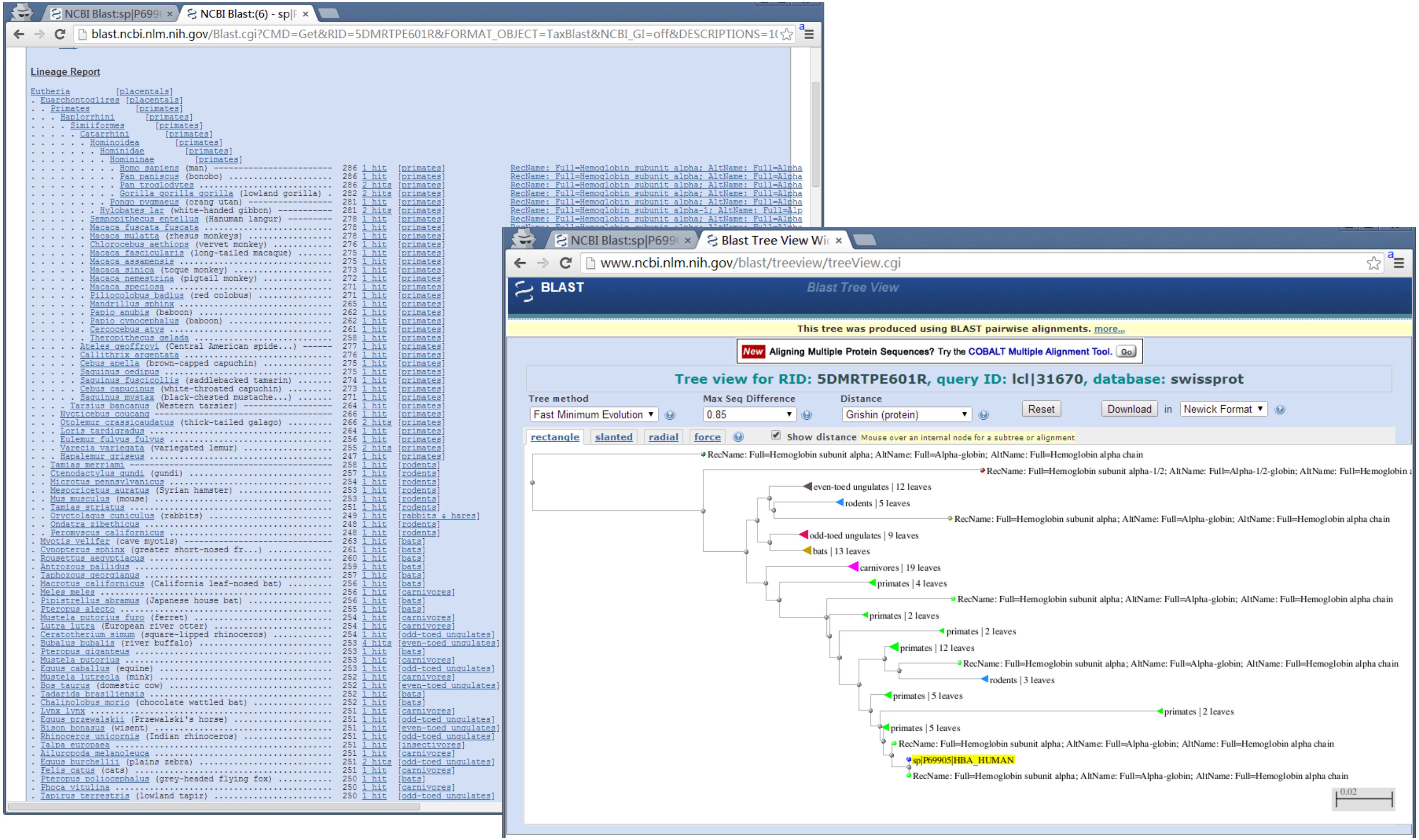
Graphic Summary

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 show defline and scores, click to show alignments



NCBI Blast:sp|P699 × NCBI Blast:(6) - sp|F ×

blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=5DMRTPE601R&FORMAT_OBJECT=TaxBlast&NCBI_GI=off&DESCRIPTIONS=1(

Lineage Report

```
Eutheria           [placentals]
. Euarchontoglires [placentals]
. . Primates           [primates]
. . . Haplorrhini          [primates]
. . . . Simiiformes          [primates]
. . . . . Catarrhini           [primates]
. . . . . . Hominoidea           [primates]
. . . . . . . Hominidae            [primates]
. . . . . . . . Homininae            [primates]
. . . . . . . . . Homo sapiens (man) ----------------------  286 1 hit  [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . . . . Pan paniscus (bonobo) ...................  286 1 hit  [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . . . . Pan troglodytes .........................  286 2 hits [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . . . . Gorilla gorilla gorilla (lowland gorilla) .  282 2 hits [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . . . Pongo pygmaeus (orang utan) ---------------  281 1 hit  [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . . Hylobates lar (white-handed gibbon) ---------  281 2 hits [primates]          RecName: Full=Hemoglobin subunit alpha-1; AltName: Full=Alp
. . . . . . Semnopithecus entellus (Hanuman langur) --------  278 1 hit  [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . Macaca fuscata fuscata ........................  278 1 hit  [primates]          RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha
. . . . . . Macaca mulatta (rhesus monkeys) ...............  278 1 hit  [primates]
. . . . . . Chlorocebus aethiops (vervet monkey) ..........  276 1 hit  [primates]
. . . . . . Macaca fascicularis (long-tailed macaque) .....  275 1 hit  [primates]
. . . . . . Macaca assamensis .............................  275 1 hit  [primates]
. . . . . . Macaca sinica (toque monkey) ..................  273 1 hit  [primates]
. . . . . . Macaca nemestrina (pigtail monkey) ............  272 1 hit  [primates]
. . . . . . Macaca speciosa ...............................  271 1 hit  [primates]
. . . . . . Piliocolobus badius (red colobus) .............  271 1 hit  [primates]
. . . . . . Mandrillus sphinx .............................  265 1 hit  [primates]
. . . . . . Papio anubis (baboon) .........................  262 1 hit  [primates]
. . . . . . Papio cynocephalus (baboon) ...................  262 1 hit  [primates]
. . . . . . Cercocebus atys ...............................  261 1 hit  [primates]
. . . . . . Theropithecus gelada ..........................  258 1 hit  [primates]
. . . . . Ateles geoffroyi (Central American spide...) ------  277 1 hit  [primates]
. . . . . Callithrix argentata ............................  276 1 hit  [primates]
. . . . . Cebus apella (brown-capped capuchin) ............  275 1 hit  [primates]
. . . . . Saguinus oedipus ................................  275 1 hit  [primates]
. . . . . Saguinus fuscicollis (saddlebacked tamarin) ......  274 1 hit  [primates]
. . . . . Cebus capucinus (white-throated capuchin) ........  273 1 hit  [primates]
. . . . . Saguinus mystax (black-chested mustache...) ......  271 1 hit  [primates]
. . . . Tarsius bancanus (Western tarsier) -----------------  264 1 hit  [primates]
. . . Nycticebus coucang ----------------------------------  266 1 hit  [primates]
. . . Otolemur crassicaudatus (thick-tailed galago) ........  266 2 hits [primates]
. . . Loris tardigradus ...................................  264 1 hit  [primates]
. . . Eulemur fulvus fulvus ...............................  256 1 hit  [primates]
. . . Varecia variegata (variegated lemur) ................  255 2 hits [primates]
. . . Hapalemur griseus ...................................  247 1 hit  [primates]
. . Tamias merriami --------------------------------------  258 1 hit  [rodents]
. . Ctenodactylus gundi (gundi) ...........................  257 1 hit  [rodents]
. . Microtus pennsylvanicus ...............................  254 1 hit  [rodents]
. . Mesocricetus auratus (Syrian hamster) .................  253 1 hit  [rodents]
. . Mus musculus (mouse) ..................................  253 1 hit  [rodents]
. . Tamias striatus .......................................  251 1 hit  [rodents]
. . Oryctolagus cuniculus (rabbits) .......................  249 1 hit  [rabbits & hares]
. . Ondatra zibethicus ....................................  248 1 hit  [rodents]
. . Peromyscus californicus ...............................  248 1 hit  [rodents]
. Myotis velifer (cave myotis) ---------------------------  263 1 hit  [bats]
. Cynopterus sphinx (greater short-nosed fr...) ...........  261 1 hit  [bats]
. Rousettus aegyptiacus ...................................  260 1 hit  [bats]
. Antrozous pallidus ......................................  259 1 hit  [bats]
. Taphozous georgianus ....................................  257 1 hit  [bats]
. Macrotus californicus (California leaf-nosed bat) .......  256 1 hit  [bats]
. Meles meles .............................................  256 1 hit  [carnivores]
. Pipistrellus abramus (Japanese house bat) ...............  256 1 hit  [bats]
. Pteropus alecto .........................................  255 1 hit  [bats]
. Mustela putorius furo (ferret) ..........................  254 1 hit  [carnivores]
. Lutra lutra (European river otter) ......................  254 1 hit  [carnivores]
. Ceratotherium simum (square-lipped rhinoceros) ..........  254 1 hit  [odd-toed ungulates]
. Bubalus bubalis (river buffalo) .........................  253 4 hits [even-toed ungulates]
. Pteropus giganteus ......................................  253 1 hit  [bats]
. Mustela putorius ........................................  253 1 hit  [carnivores]
. Equus caballus (equine) .................................  253 1 hit  [odd-toed ungulates]
. Mustela lutreola (mink) .................................  252 1 hit  [carnivores]
. Bos taurus (domestic cow) ...............................  252 1 hit  [even-toed ungulates]
. Tadarida brasiliensis ...................................  252 1 hit  [bats]
. Chalinolobus morio (chocolate wattled bat) ..............  252 1 hit  [bats]
. Lynx lynx ...............................................  251 1 hit  [carnivores]
. Equus przewalskii (Przewalski's horse) ..................  251 1 hit  [odd-toed ungulates]
. Bison bonasus (wisent) ..................................  251 1 hit  [even-toed ungulates]
. Rhinoceros unicornis (Indian rhinoceros) ................  251 1 hit  [odd-toed ungulates]
. Talpa europaea ..........................................  251 1 hit  [insectivores]
. Ailuropoda melanoleuca ..................................  251 1 hit  [carnivores]
. Equus burchellii (plains zebra) .........................  251 2 hits [odd-toed ungulates]
. Felis catus (cats) ......................................  251 1 hit  [carnivores]
. Pteropus poliocephalus (grey-headed flying fox) .........  250 1 hit  [bats]
. Phoca vitulina ..........................................  250 1 hit  [carnivores]
. Tapirus terrestris (lowland tapir) ......................  250 1 hit  [odd-toed ungulates]
```

NCBI Blast:sp|P699 × Blast Tree View Wi ×

www.ncbi.nlm.nih.gov/blast/treeview/treeView.cgi

BLAST — Blast Tree View

This tree was produced using BLAST pairwise alignments. more...

New Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. Go

Tree view for RID: 5DMRTPE601R, query ID: lcl|31670, database: swissprot

Tree method: Fast Minimum Evolution

Max Seq Difference: 0.85

Distance: Grishin (protein)

Reset Download in Newick Format

rectangle slanted radial force

☑ Show distance Mouse over an internal node for a subtree or alignment

RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
RecName: Full=Hemoglobin subunit alpha-1/2; AltName: Full=Alpha-1/2-globin; AltName: Full=Hemoglobin a
even-toed ungulates | 12 leaves
rodents | 5 leaves
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
odd-toed ungulates | 9 leaves
bats | 13 leaves
carnivores | 19 leaves
primates | 4 leaves
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
primates | 2 leaves
primates | 2 leaves
primates | 12 leaves
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
rodents | 3 leaves
primates | 5 leaves
primates | 2 leaves
primates | 5 leaves
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
sp|P69905|HBA_HUMAN
RecName: Full=Hemoglobin subunit alpha; AltName: Full=Alpha-globin; AltName: Full=Hemoglobin alpha chain
0.02

Summary Questions

- Why do we need to perform a database searching?

- What's the major challenge/obstacle when searching sequence database?

生物信息学：导论与方法

Bioinformatics: Introduction and Methods

https://www.coursera.org/course/pkubioinfo