BIOINFORMATICS

GROUP - 12

Protein Sequence Analysis

INTRODUCTION

- **Protein**: Proteins are large biomolecules or macromolecules that are comprised of one or more long chains of amino acid residues.

- **Protein Sequence Analysis:** is the process of subjecting a protein or peptide sequence to one of a wide range of analytical methods to study its features, function, structure, or evolution. Methodologies used include sequence alignment, searches against biological databases, and other methods.

- Now, there are almost 8 million sequences in a nonredundant (NR) database of protein sequences, including the complete genomes of nearly 1,800 different species

- Protein sequencing is used to identify the amino acid sequence and its conformation. The identification of the structure and function of proteins is important to understand cellular processes.

- Some of the **applications** of sequence analysis are - Sequence comparison, Classification of proteins, Comparative genomics and RNA structure prediction.

Applications Of Protein Sequence Analysis

1. Sequencing projects,
2. assembly of sequence data.
3. Identification of functional elements in sequences,
4. gene prediction.
5. Sequence comparison.
6. Classification of proteins.
7. Comparative genomics.
8. RNA structure prediction.
9. Protein structure prediction.
10. It is important for understanding of cellular functions.

SEQUENCE METHODS
N-TERMINAL SEQUENCING

There are two methods in N-Terminal Sequencing. They are:

Sanger's method

1. Treat with DNFB to form a derivative of amino terminal amino acid.
2. Acid hydrolysis.
3. Extraction of DNP-derivative with organic solvent.
4. Identification of DNP-derivative by chromatography and comparison with standards.

Dansyl chloride method

1. Reagent:1-dimethyl aminophthalene-5-sulfonyl chloride (dansyl chloride)
2. Dansyl polypeptide chain is prepared.
3. Acidic hydrolysis liberates all amino acid and N terminal dansyl amino acid.
4. Amino acids are separated. Then, Fluorescence of dansyl amino acid is detected.
5. Types of amino acid is obtained from comparison with standard dansylated amino acids.

TOOLS

BLAST (Basic local alignment search tool):

BLAST is a heuristic search algorithm, it finds the solutions from the all possibilities, which takes input as a protein sequence and compare it with existing databases like NCBI, GenBank etc.

BLAST is one of the pairwise sequence alignment tool which is used to compare different sequences. It finds the local similarity between different sequences and calculates the statistical significance of matches.

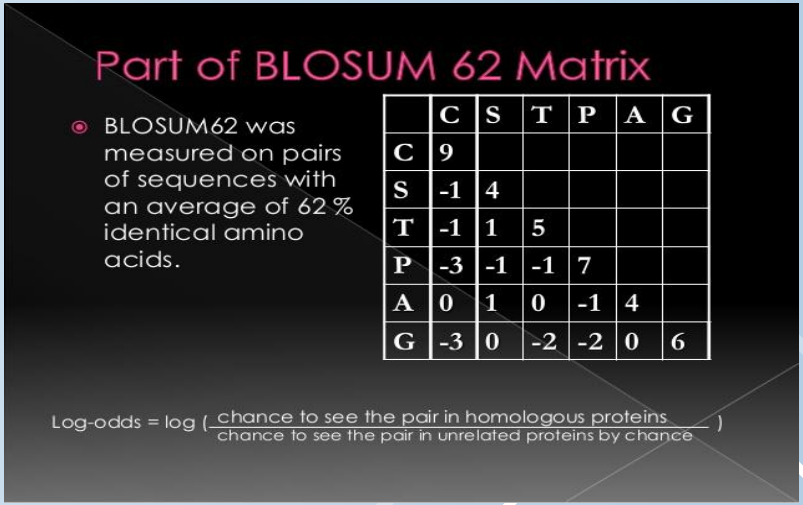
SCORING MATRICES

Mainly used predefined matrices are PAM and BLOSUM.

PAM Matrices:

- Margaret Daihoff was the first to develop the WFP matrix. WFP is an abbreviation for Accepted Dot Moves. The PAM matrix was calculated by observing closely related protein differences.
- One PAM unit (PAM1) shows acceptable point mutations per 100 amino acid residues i.e 1% and 99% changes persist as is.

BLOSUM: Blocks Substitution matrices are actual percentage identity values. Simply to say, they depend on similarity. Blosum 62 means there is 62 % similarity.



	C	S	T	P	A	G
C	9					
S	-1	4				
T	-1	1	5			
P	-3	-1	-1	7		
A	0	1	0	-1	4	
G	-3	0	-2	-2	0	6

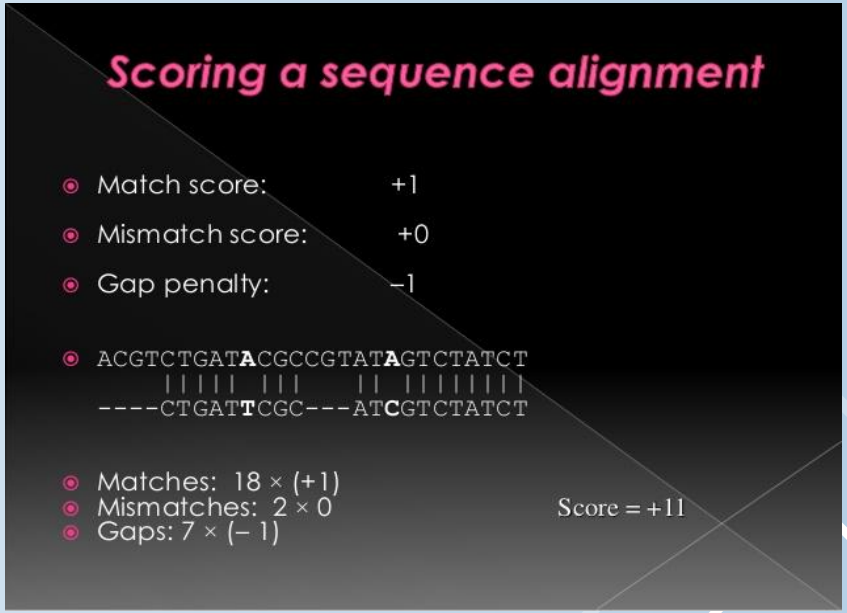
PARAMETERS

Threshold: It is a boundary of minimum or maximum value which can be used to filter out words during comparison.

E-value: It decreases exponentially with the score that is assigned to an alignment between two sequences.

Word size: Whole Search is done by taking the sequence of a certain word size and compares it with the database sequence and scores are assigned for each comparison. Word size is given as 3 for proteins.

Gap score or gap penalty: Dynamic programming algorithms use gap penalties to maximize biological significance. Gap fines will be deducted for each suggested gap. There are different penalties for gaps, such as opening a gap and extending it. The offset score defines the penalty point assigned to the alignment when entered or removed.



BLAST ALGORITHM STEPS

- Query sequence is taken and analyzed for low complex regions. Low complexity regions are regions which contain less information or variations like AAAAAAAA or ATATATAT etc.These low complex regions are marked with alphabets like X or N.
- List of words of a certain word size is made. Usually the word size is 3 or 6 for proteins.
- Scores are calculated for each pair of words using substitution scoring matrices and only the high scoring words i.e. above a threshold value is taken for further alignment. The high-scoring words are organised into efficient search tree and rapidly compared to the database sequence. This is done to find out the exact matches.

BLAST PROCEDURE

• If an exact or good match is found then an alignment is extended in both directions from the position where the exact match occurred.

• High scoring pairs (HSP) which have score greater than a threshold are taken for consideration.

BLAST Procedure
This is the common procedure for any BLAST program.

Step 1: Select the BLAST program
User have to specify the type of BLAST programs from the database like BLASTp, BLASTn, BLASTx, tBLASTn, tBLASTx.

Step 2: Enter a query sequence or upload a file containing sequence
Enter a query sequence by pasting the sequence in the query box or uploading a FASTA file which is having the sequence for similarity search.

STEPS Contd...

Step 3: Select database to search
Sequence similarity search involves searching of similar sequences of the query sequence from the selected databases.

Step 4: Select the algorithm and the parameters of the algorithm for the search
Protein BLAST algorithms like BLASTp, PSI-BLAST, PHI-BLAST, DELTA-BLAST etc need to be selected.

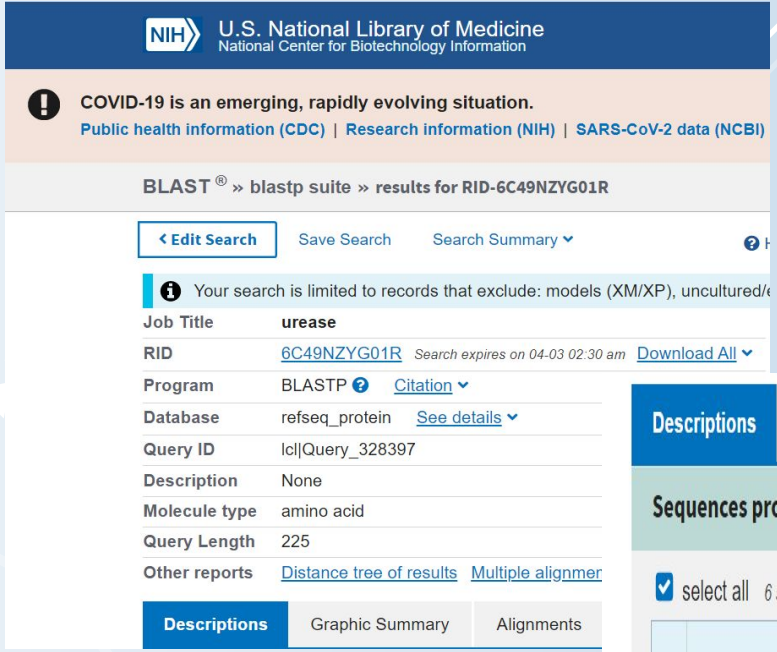
Step 5: Run the BLAST program
Submission of the BLAST program can be done by clicking the BLAST button at the end of the page and you will.

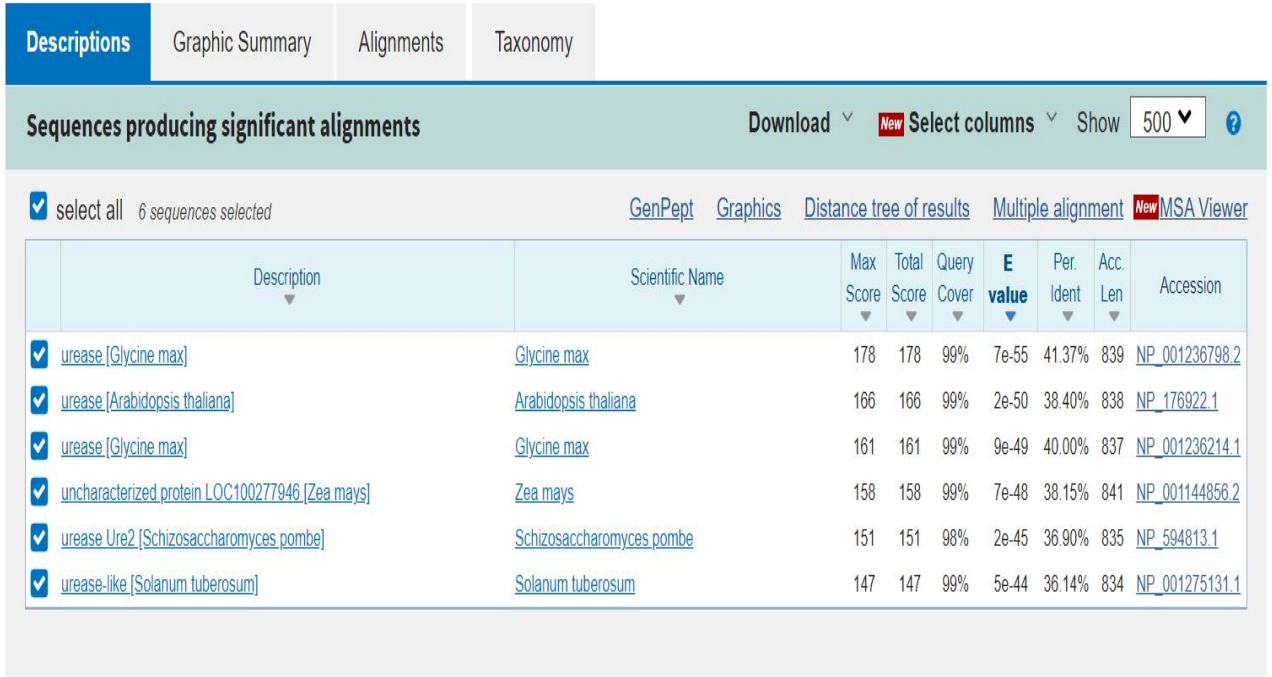
OUR RESULTS

Input Sequence:
MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE
CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDDFKAGEIKFASDKDIEL
NAGKEITELKVTNKGPKSLHVGSHFHFFEANRALEFDREKAYGKRLDIPSGNTLRIGAGE
TKTVHLIPIGGSKKIIGMNGLLNGIADDLHKQKALEKAKHHGFIK

Protein BLAST Program

BLASTp: Finds the similarity between the query protein sequences to a protein sequences available in the protein database. BLASTp also reports for global alignment, which is the preferred result for protein identification. The BLASTp algorithm parses protein sequences into 3 letter “words” the same is done for every sequence in the query database, word matches are being identified from the database.





Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
urease [Glycine max]	Glycine max	178	178	99%	7e-55	41.37%	839	NP_001236798.2
urease [Arabidopsis thaliana]	Arabidopsis thaliana	166	166	99%	2e-50	38.40%	838	NP_176922.1
urease [Glycine max]	Glycine max	161	161	99%	9e-49	40.00%	837	NP_001236214.1
uncharacterized protein LOC100277946 [Zea mays]	Zea mays	158	158	99%	7e-48	38.15%	841	NP_001144856.2
urease Ure2 [Schizosaccharomyces pombe]	Schizosaccharomyces pombe	151	151	98%	2e-45	36.90%	835	NP_594813.1
urease-like [Solanum tuberosum]	Solanum tuberosum	147	147	99%	5e-44	36.14%	834	NP_001275131.1

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view: Pairwise | Restore defaults | Download

6 sequences selected

Download | GenPept | Graphics | ▼ Next ▲ Previous ◀ Descriptions

urease [Glycine max]

Sequence ID: NP_001236798.2 **Length:** 839 **Number of Matches:** 1

Range 1: 1 to 249 GenPept | Graphics | ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
178 bits(451)	7e-55	Compositional matrix adjust.	103/249(41%)	144/249(57%)	26/249(10%)

```
Query  1    MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE  60
            MKL+P+E EK   L+ AG +A+KR    GL+LN   EA+A I+   IM+ AR G+KTVAQLM
Sbjct  1    MKLSPREVEKLGLHNAGYLAQKRLARGLRLNYTEAVALIATQIMEFARDGEKTVAQLMCI  60

Query  61   CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDDFKAGEIKFAS------  114
              H L + +V+P V +++   + VEA FPDGTKLVTV+ PI  +    G+  F S
Sbjct  61   GKHLLGRRQVLPEVQHLLNAVQVEATFPDGTKLVTVHDPISCEHGDLGQALFGSFLPVPS  120

Query  115  --------------------DKDIELNAGKEITELKVTNKGPKSLHVGSHFHFFEANRAL  154
                              D  + LN GK    LKV + G + + VGSH+HF E N  L
Sbjct  121  LDKFAENKEDNRIPGEIIYGDGSLVLNPGKNAVILKVVSNGDRPIQVGSHYHFIEVNPYL  180

Query  155  EFDREKAYGKRLDIPSGNTLRIGAGETKTVHLIPIGGSKKIIGMNGLLNGIADDLHKQKA  214
             FDR KAYG RL+I +G  +R   G++K+V L+ IGG+K I G NG+ +G  ++ + ++A
Sbjct  181  TFDRRKAYGMRLNIAAGTAVRFEPGDSKSVKLVRIGGNKVIRGGNGIADGQVNETNLREA  240

Query  215  LEKAKHHGF  223
            +E     GF
Sbjct  241  MEAVCKRGF  249
```

Related Information

Gene - associated gene details

Genome Data Viewer - aligned genomic context

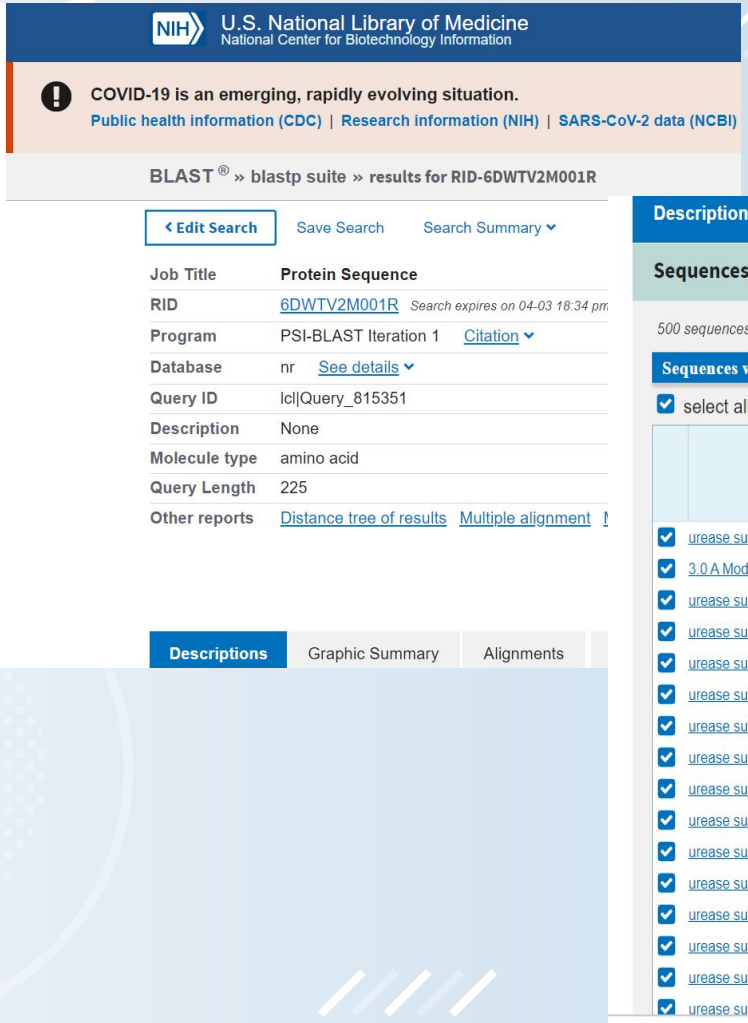
Identical Proteins - Identical proteins to NP_001236798.2

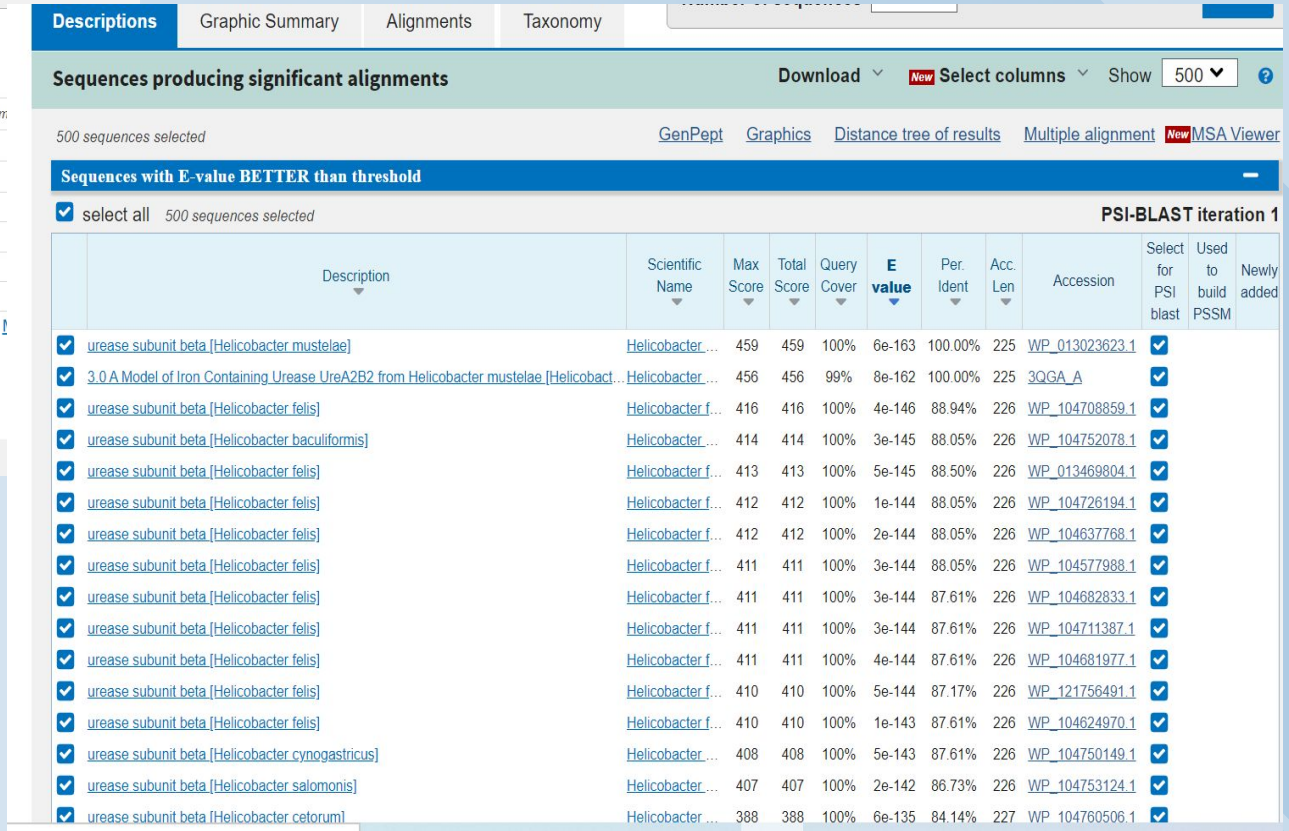
Download | GenPept | Graphics | ▼ Next ▲ Previous ◀ Descriptions

PSI-BLAST

PSI-BLAST Program

Position-Specific Iterated-BLAST is the most sensitive BLAST program. It is used to find very distantly related proteins or new members of the protein family. Algorithm builds a position-specific scoring matrix (PSSM or profile) from an iterative alignment of sequences, returns with E-values and threshold (default=0.005). E-value It decreases exponentially with the score that is assigned to a match between two sequences.





Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
urease subunit beta [Helicobacter mustelae]	Helicobacter…	459	459	100%	6e-163	100.00%	225	WP_013023623.1	☑		
3.0 A Model of Iron Containing Urease UreA2B2 from Helicobacter mustelae [Helicobact…	Helicobacter…	456	456	99%	8e-162	100.00%	225	3QGA_A	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	416	416	100%	4e-146	88.94%	226	WP_104708859.1	☑		
urease subunit beta [Helicobacter baculiformis]	Helicobacter…	414	414	100%	3e-145	88.05%	226	WP_104752078.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	413	413	100%	5e-145	88.50%	226	WP_013469804.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	412	412	100%	1e-144	88.05%	226	WP_104726194.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	412	412	100%	2e-144	88.05%	226	WP_104637768.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	411	411	100%	3e-144	88.05%	226	WP_104577988.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	411	411	100%	3e-144	87.61%	226	WP_104682833.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	411	411	100%	3e-144	87.61%	226	WP_104711387.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	411	411	100%	4e-144	87.61%	226	WP_104681977.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	410	410	100%	5e-144	87.17%	226	WP_121756491.1	☑		
urease subunit beta [Helicobacter felis]	Helicobacter f…	410	410	100%	1e-143	87.61%	226	WP_104624970.1	☑		
urease subunit beta [Helicobacter cynogastricus]	Helicobacter…	408	408	100%	5e-143	87.61%	226	WP_104750149.1	☑		
urease subunit beta [Helicobacter salomonis]	Helicobacter…	407	407	100%	2e-142	86.73%	226	WP_104753124.1	☑		
urease subunit beta [Helicobacter cetorum]	Helicobacter…	388	388	100%	6e-135	84.14%	227	WP_104760506.1	☑		

Download GenPept Graphics ▼Next ▲Previous ◀◀Descriptions

urease subunit beta [Helicobacter mustelae]

Sequence ID: WP_013023623.1 **Length:** 225 **Number of Matches:** 1

See 3 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 225 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
459 bits(1180)	6e-163	Compositional matrix adjust.	225/225(100%)	225/225(100%)	0/225(0%)

```
Query  1    MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE  60
            MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE
Sbjct  1    MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE  60

Query  61   CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDDFKAGEIKFASDKDIEL  120
            CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDDFKAGEIKFASDKDIEL
Sbjct  61   CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDDFKAGEIKFASDKDIEL  120

Query  121  NAGKEITELKVTNKGPKSLHVGSHFHFFEANRALEFDREKAYGKRLDIPSGNTLRIGAGE  180
            NAGKEITELKVTNKGPKSLHVGSHFHFFEANRALEFDREKAYGKRLDIPSGNTLRIGAGE
Sbjct  121  NAGKEITELKVTNKGPKSLHVGSHFHFFEANRALEFDREKAYGKRLDIPSGNTLRIGAGE  180

Query  181  TKTVHLIPIGGSKKIIGMNGLLNGIADDLHKQKALEKAKHHGFIK  225
            TKTVHLIPIGGSKKIIGMNGLLNGIADDLHKQKALEKAKHHGFIK
Sbjct  181  TKTVHLIPIGGSKKIIGMNGLLNGIADDLHKQKALEKAKHHGFIK  225
```

Related Information

Identical Proteins - Identical proteins to WP_013023623.1

Download GenPept Graphics ▼Next ▲Previous ◀◀Descriptions

DELTA BLAST

DELTA BLAST Program

Domain enhanced lookup time accelerated BLAST (DELTA-BLAST), which searches a database of pre-constructed PSSMs(position-specific scoring matrix) before searching a protein-sequence database, to yield better homology detection. For its PSSMs, DELTA-BLAST employs a subset of NCBI's Conserved Domain Database (CDD).

U.S. National Library of Medicine
National Center for Biotechnology Information

COVID-19 is an emerging, rapidly evolving situation.
Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI)

BLAST® » blastp suite » results for RID-6EFY5HF1013

‹ Edit Search | Save Search | Search Summary

Job Title	urease
RID	6EFY5HF1013 *Search expires on 04-04 00:01 am* Download All
Program	DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) Citation
Database	nr See details
Query ID	lcl|Query_74576
Description	None
Molecule type	amino acid
Query Length	225
Other reports	Distance tree of results Multiple alignment MSA viewer

Descriptions | Graphic Summary | Alignments | Taxonomy

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download | New Select columns | Show 500

500 sequences selected GenPept | Graphics | Distance tree of results | Multiple alignment | New MSA Viewer

Sequences with E-value BETTER than threshold

select all *500 sequences selected* **PSI-BLAST iteration 1**

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
urease subunit beta [Proteobacteria bacterium]	Proteobacteri...	434	434	91%	2e-153	45.67%	206	PZN39724.1	☑		
urease subunit beta [Helicobacter anseris]	Helicobacter...	431	431	100%	3e-152	56.64%	225	WP_115578512.1	☑		
urease subunit gamma [Myxococcales bacterium]	Myxococcales...	430	430	92%	1e-151	47.60%	220	RYZ01987.1	☑		
MULTISPECIES: urease subunit beta [unclassified Helicobacter]	unclassified H...	429	429	100%	3e-151	54.87%	225	WP_104697530.1	☑		
urease subunit gamma [Chelatococcus composti]	Chelatococcu...	428	428	91%	4e-151	45.19%	206	WP_183336364.1	☑		
urease subunit beta [Proteobacteria bacterium]	Proteobacteri...	428	428	93%	5e-151	46.23%	211	PZN26908.1	☑		
urease subunit gamma [Chelatococcus daeguensis]	Chelatococcu...	427	427	91%	8e-151	46.15%	206	WP_082831596.1	☑		
MULTISPECIES: urease subunit gamma [unclassified Chelatococcus]	unclassified C...	427	427	91%	1e-150	46.15%	206	WP_019404069.1	☑		
urease subunit gamma [Ruminococcus sp. AF16-50]	Ruminococcu...	427	427	99%	1e-150	51.56%	226	WP_117864124.1	☑		
urease subunit gamma [Ruminococcus sp. AM54-1NS]	Ruminococcu...	427	427	99%	1e-150	51.56%	226	WP_118158883.1	☑		
urease subunit gamma [Filomicrobium insigne]	Filomicrobium...	427	427	91%	1e-150	44.76%	208	WP_090226367.1	☑		
urease subunit gamma [Rhizobium sp.]	Rhizobium sp.	426	426	91%	2e-150	47.57%	204	MBK5655186.1	☑		
urease subunit gamma [Segnochrobactrum spirostomi]	Rhizobiales b...	426	426	91%	2e-150	47.12%	206	MQT12448.1	☑		
urease subunit gamma [Sedimenticola thiotaurini]	Sedimenticola...	426	426	92%	3e-150	47.89%	211	WP_046857994.1	☑		
urease subunit gamma [Ruminococcus sp. AM47-2BH]	Ruminococcu...	426	426	99%	4e-150	51.56%	226	WP_118164004.1	☑		

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view: Pairwise | Restore defaults | Download

500 sequences selected

Download | GenPept | Graphics | ▼Next ▲Previous ◀◀Descriptions

urease subunit beta [Proteobacteria bacterium]

Sequence ID: PZN39724.1 **Length:** 206 **Number of Matches:** 1

Range 1: 1 to 206 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
434 bits(1115)	2e-153	Composition-based stats.	95/208(46%)	132/208(63%)	4/208(1%)

```
Query  1    MKLTPKEQEKFLLYYAGEVARKRKEEGLKLNQPEAIAYISAHIMDEARRGKKTVAQLMEE  60
            M LTP+E++K L+  A  VAR+R E G+KLN PEA+A I+  +++ AR G+ +VA+LME
Sbjct  1    MNLTPREKDKLLIAMAAMVARRRLERGVKLNYPEAVALITDFVVEGARDGR-SVAELMEA  59

Query  61   CVHFLKKDEVMPGVGNMVPDLGVEANFPDGTKLVTVNWPIEPDD--FKAGEIKFASDKDI  118
              H L  D+VM GV  M+ ++ VEA FPDGTKLVTV+ PI       + GE    +  ++
Sbjct  60   GAHVLTPDQVMDGVAEMITEVQVEATFPDGTKLVTVHNPIRGATGKLQPGE-TLPAPGEV  118

Query  119  ELNAGKEITELKVTNKGPKSLHVGSHFHFFEANRALEFDREKAYGKRLDIPSGNTLRIGA  178
             LN G+E   L V N G + + VGSH+HF+E N AL FDREKA G RLDIP+G  +R
Sbjct  119  TLNEGRETVTLTVANTGDRPIQVGSHYHFYETNPALSFDREKARGMRLDIPAGTAVRFEP  178

Query  179  GETKTVHLIPIGGSKKIIGMNGLLNGIA  206
            G+T+ V L+ + G +K+ G    + G
Sbjct  179  GQTREVTLVALAGERKVYGFRQQVMGKL  206
```

Download | GenPept | Graphics | ▼Next ▲Previous ◀◀Descriptions

SMITH WATERMAN ALGORITHM

This algorithm is used for determining the similar regions in nucleic acids and protein sequences. The broad idea is to use Dynamic Programming to optimize the similarity measure and start building from using small segments to tackle larger sequencing problems.

When a new order is found, the structure and function can be easily adjusted by sorting. This is because a sequence that shares a common ancestor is believed to have a similar structure or function. No matter how similar the scenes are, it is possible that they have a similar structure or function.

STEPS OF ALGORITHM

Let A = a1, a2, …, an and B = b1, b2, …, bn are the sequences to be analysed, where n and m are the lengths of A and B respectively.

Step 1. Determine the substitution matrix and define a scheme for gap penalty.

- s(a, b): Similarity score of the elements
- W_k: The penalty of length k

Step 2. Construct a scoring matrix H and initialize its first row and first column.

- $H_{k0} = H_{lo} = 0$ for all k and l in 0 to n and m.

Step 3: Fill the scoring matrix as per the sequence below -

$$H_{ij} = \max \begin{cases} H_{i-1,j-1} + s(a_i, b_j), \\ \max_{k \geq 1} \{H_{i-k,j} - W_k\}, \\ \max_{l \geq 1} \{H_{i,j-l} - W_l\}, \\ 0 \end{cases} \qquad (1 \leq i \leq n, 1 \leq j \leq m)$$

where

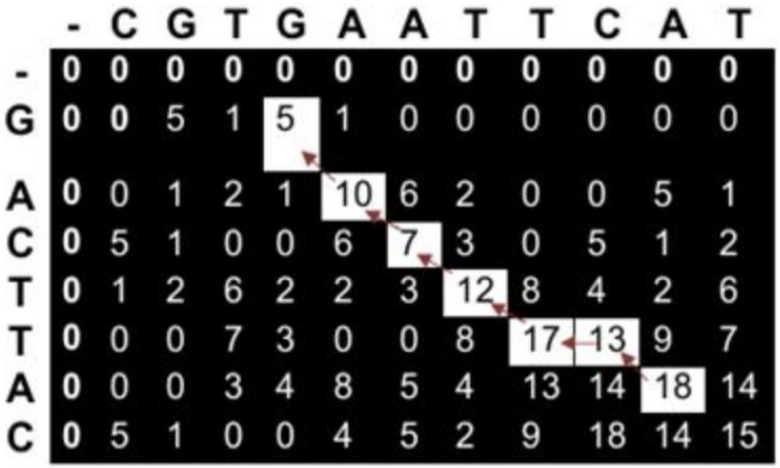
$H_{i-1,j-1} + s(a_i, b_j)$ is the score of aligning a_i and b_j,

$H_{i-k,j} - W_k$ is the score if a_i is at the end of a gap of length k,

$H_{i,j-l} - W_l$ is the score if b_j is at the end of a gap of length l,

0 means there is no similarity up to a_i and b_j.

Step 4: The last step for proper alignment is reversing, before which it is necessary to determine the maximum result obtained in the general matrix for local alignment of the array. You can have maximum results in several cells, in which case two or more alignments and the best alignment is possible by counting.



	-	C	G	T	G	A	A	T	T	C	A	T
-	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	5	1	5	1	0	0	0	0	0	0
A	0	0	1	2	1	10	6	2	0	0	5	1
C	0	5	1	0	0	6	7	3	0	5	1	2
T	0	1	2	6	2	2	3	12	8	4	2	6
T	0	0	0	7	3	0	0	8	17	13	9	7
A	0	0	0	3	4	8	5	4	13	14	18	14
C	0	5	1	0	0	4	5	2	9	18	14	15

Figure 3: Trace back of first possible alignment

ACCELERATED VERSIONS

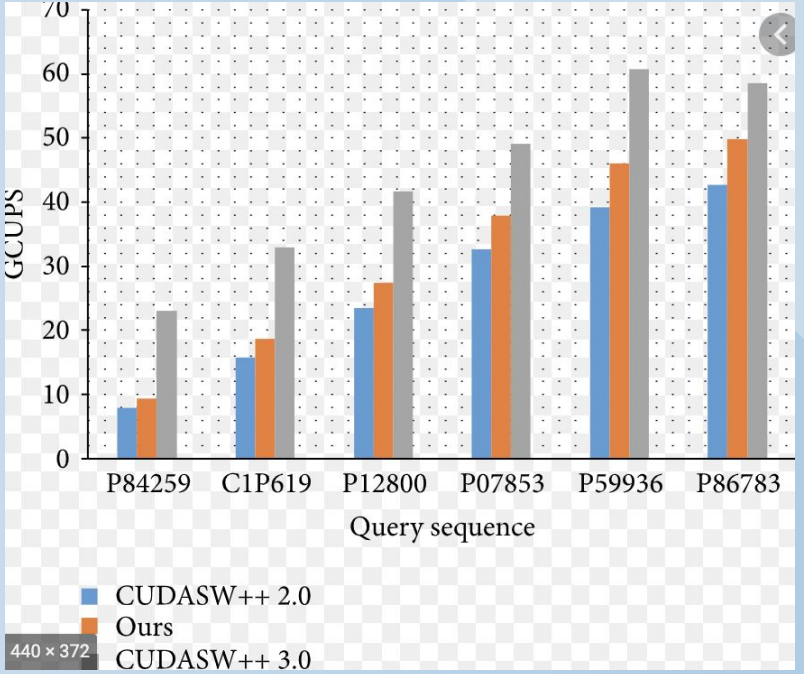
FPGA: Cray shows the acceleration of the algorithm. Smith-Waterman Using a new-tunable compute platform using FPGA chips, with results showing speeds up to 28 times faster than standard microprocessor-based solutions.

Virtex-4 up to 100x on Opteron 2.2 GHz processors TimeLogic DeCypher and CodeQuest also accelerated Smith-Waterman and Framesearch using PCIe FPGA cards.

GPU: The Lawrence Livermore National Laboratory and the United States Department of Energy's Joint Genome Institute used an accelerated version of the local chronological alignment search of the US Department of Energy.

Smith-Waterman Using a graphics processing unit (GPU), with preliminary results showing 2x acceleration compared to software applications. A similar approach has been used in Biofacet software since 1997 with the same acceleration factor.

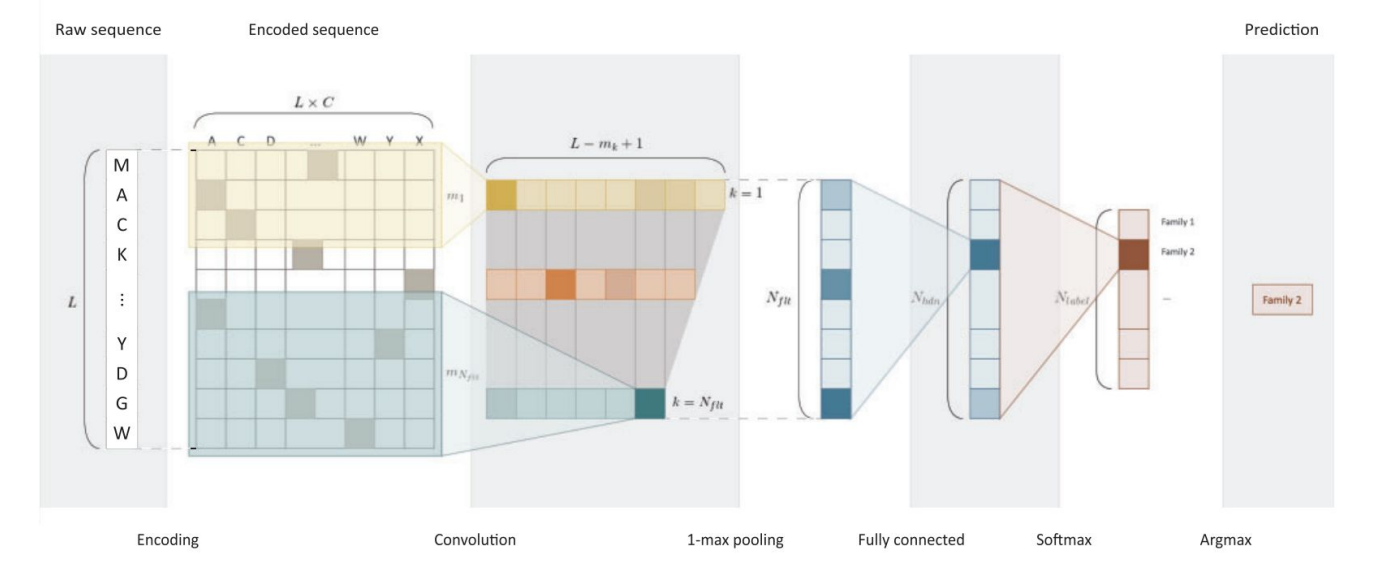
512 × 355



440 × 372

DeepFam - alignment free

GPCR dataset: 14 000 proteins from 3547 species, 7 highly conserved segments

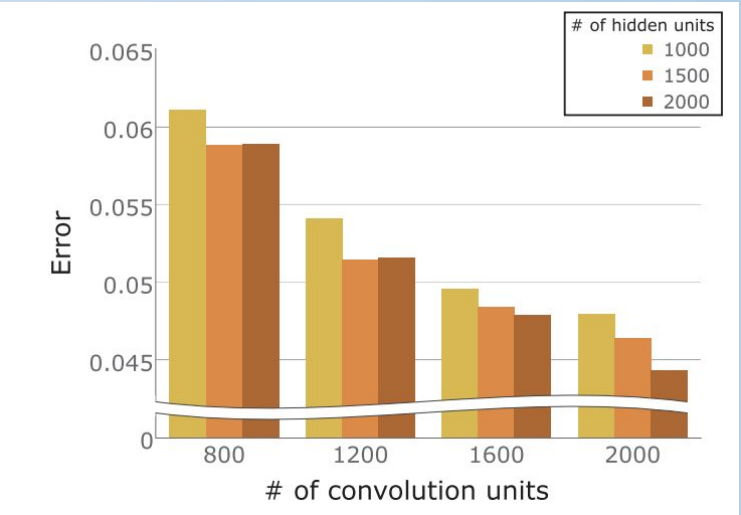


DeepFam - alignment free

Loss used: L1 loss

Results:

- Family basis - 97.17 percent
- Sub family - 86.82
- Sub sub family - 81.17



CONTRIBUTIONS

T. Pavan Kumar - S20180010174
C .Sai kumar - S20180010039
S. Vinay - S20180010169
D.Goutham - S20180010044

Introduction , History, Applications of sequence analysis, Sequence methods (N-Terminal sequencing, C-Terminal sequencing and DNA sequencing)

M. Mani Tej - S20180010104
M. Bhanu Kishore - S20180010098
V. Shankar Sreenu - S20180010186

BLAST(Basic local alignment search tool).
BLAST Algorithm (Understanding)
Defining parameters in BLAST algorithm.
BLAST (Working Procedure).
Showed results of blastp, psi-blast, delta-blast programs using BLAST.

CONTRIBUTIONS

Sayam Kumar - S20180010158
Raahul Singh - S20180010141
Hrishabh Pandey - S20180010064
Sushanth Bondley - S20180010030

→ Smith–Waterman algorithm, Scoring Matrices, all algorithm steps and accelerated versions of the algorithm

THANK YOU

- Group - 12