Alignment Practical Session

Yazdan Asgari

BLAST - DNA

Global Alignment: Amyloid Precursor Protein

HUGO Symbol	Locus	Gene product	Associated disease
APP	21q21	Amyloid precursor protein	Alzheimer's Disease

APP Nucleotide Create alert Advanced uence GI numbers in September 2016. Please use accession.version! Read more... Send to: -See APP amyloid beta precursor protein in the Gene database app reference sequences Genomic (1) Transcript (10) Protein (10) Items: 1 to 20 of 214505 << First < Prev Page 1 of 10726 Next > Last >> Found 638700 nucleotide sequences. Nucleotide (214505) EST (120665) GSS (303530) Neisseria meningitidis strain H44/76 App (app) gene, complete cds 4,374 bp linear DNA Accession: AY150285.1 GI: 25140437 GenBank FASTA Graphics APP=amyloid precursor protein {exon 16/17 junction} [human, familial Alzheimer disease family LIT, Genomic Mutant, 96 nt] 96 bp linear DNA Accession: S45136.1 GI: 257379

GenBank FASTA Graphics

Nucleotide

▼ app mouse

Create alert Advanced

equence GI numbers in September 2016. Please use accession.version! Read more...

Send to: ▼

See <u>App</u> amyloid beta (A4) precursor protein in the Gene database app reference sequences <u>Transcript (7)</u> Protein (7)

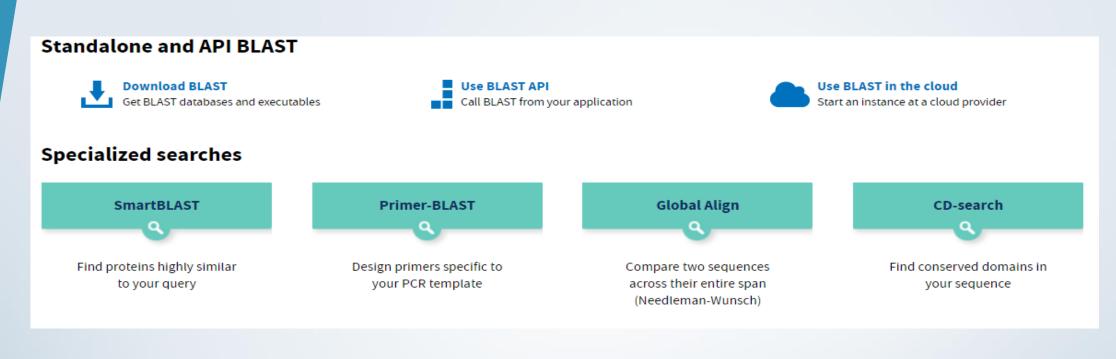
Items: 1 to 20 of 6976

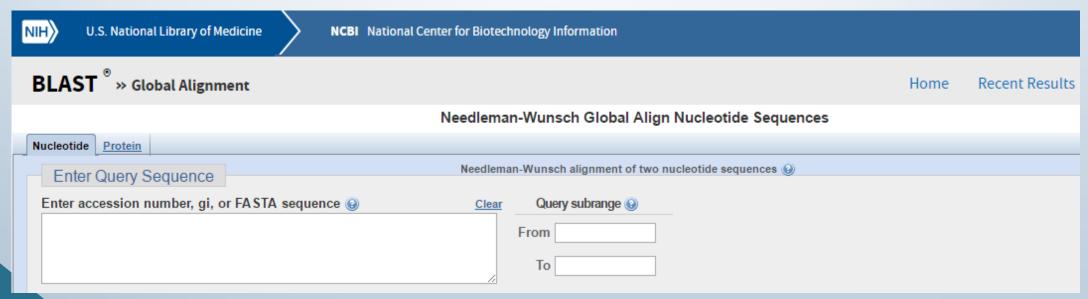
<< First < Prev Page 1 of 349 Next > Last >>

- Found 7307 nucleotide sequences. Nucleotide (6976) EST (331)
- Mus musculus amyloid beta (A4) precursor protein (App), transcript variant 6, mRNA
- 3,323 bp linear mRNA

Accession: NM_001198826.1 GI: 311893407

GenBank FASTA Graphics





BLAST [®] ≫ Global Alignment

Needleman-Wunsch Global Align Nucleotide Sequences Nucleotide Protein Needleman-Wunsch alignment of two nucleotide sequences (a) Enter Query Sequence Enter accession number, gi, or FASTA sequence (a) Query subrange (2) Clear >gi|257379|gb|S45136.1| APP=amyloid precursor protein {exon 16/17 junction} From [human, familial Alzheimer disease family LIT, Genomic Mutant, 96 nt] TTGTCATAGCGACAGTGATCGTCATC To Or, upload file Choose File No file chosen Job Title gi|257379|gb|S45136.1| APP=amyloid precursor... Enter a descriptive title for your BLAST search (a) Enter Subject Sequence Enter accession number, gi, or FASTA sequence (a) Subject subrange (a) Clear >gi|311893407|ref|NM 001198826.1| Mus musculus amyloid beta (A4) From precursor protein (App), transcript variant 6, mRNA GCGGCGCGCGGACACAGCCAGGGCGCGGCGGATCTTCCACTCGCACACGGAGCACTCGGTGGCCCACGC To AGGATCACGATGCTGCCCAGCTTGGCACTGCTCCTGCTGGCCGCCTGGACGGTTCGGGCTCTGGAGGTAC Or, upload file Choose File | No file chosen Align Show results in a new window

Sbjct	1801	TCCCTGCGGTGGCTGAGGAGATTCAAGATGAAGTCGATGAGCTGCTTCAGAAGGAGCAGA	1860
Sbjct	1861	${\tt ACTACTCCGACGATGTCTTGGCCAACATGATCAGTGAGCCCAGAATCAGCTACGGAAACG}$	1920
Sbjct	1921	${\tt ACGCTCTCATGCCTTCGCTGACGGAAACCAAGACCACCGTGGAGCTCCTTCCCGTGAATG}$	1980
Sbjct	1981	GGGAATTCAGCCTGGATGACCTCCAGCCGTGGCACCCTTTTGGGGTGGACTCTGTGCCAG	2040
Query	1	AAAT	4
Sbjct	2041	CCAATACCGAAAATGAAGGTTCTGGGCTGACAAACATCAAGACGGAAGAGATCTCGGAAG	2100
Query	5	TGG	7
Sbjct	2101	TGAAGATGGATGCAGAATTCGGACATGATTCAGGATTTGAAGTCCGCCATCAAAAACTGG	2160
Query	8	TGTTCTTTGCAGAAGATGTGGGTTCAAACAAAGGTGCAATCATTGGACTCATGGTGGGTG	67
Sbjct	2161	TGTTCTTTGCTGAAGATGTGGGTTCGAACAAAGGCGCCATCATCGGACTCATGGTGGGCG	2220
Query	68	GTGTTGTCATAGCGACAGTGATCGTCATC	96
Sbjct	2221	GCGTTGTCATAGCAACCGTGATTGTCATCACCCTGGTGATGTTGAAGAAGAAACAGTACA	2280
Sbjct	2281	CATCCATCCATCATGGCGTGGTGGAGGTCGACGCCGCCGTGACCCCAGAGGAGCGCCATC	2340
Sbjct	2341	TCTCCAAGATGCAGCAGAACGGATATGAGAATCCAACTTACAAGTTCTTTGAGCAAATGC	2400
Sbjct	2401	AGAACTAAGCCCCACCCGCGCCACAGCAGCGGCCTCTGAACTTGGACAGCGAAACCATTG	2460
Sbjct	2461	${\tt CTTCACTACCCATCGGTGTTCATTTATAAAATAACGTGGAAAGAAA$	2520
Sbjct	2521	TATTTACTCACCCTCGGCTTTTGACAGCTGTGCTGTAACACAAGTAGATGCCTGAACTTG	2580
Sbjct	2581	${\tt AATTAATATACAAATCAGTAATGTATTCTCGCTTTCTCTTTTACATTCTGGTCTCTACA}$	2640
Sbjct	2641	TTACATGATTCATGGGTTTTGTGTACTGTAAAAAAAAAA	2700

Blastn: APP

BLAST » blastn suite Home Standard Nucleotide BLAST blastn blastp blastx tblastn tblastx BLASTN programs search nucleotide databases using a nucleotide query, more... **Enter Query Sequence** Enter accession number(s), gi(s), or FASTA sequence(s) (a) Query subrange (a) Clear >gi|257379|gb|S45136.1| APP=amyloid precursor protein {exon 16/17 junction} From [human, familial Alzheimer disease family LIT, Genomic Mutant, 96 nt] TTGTCATAGCGACAGTGATCGTCATC To Or, upload file Choose File No file chosen Job Title gi|257379|gb|S45136.1| APP=amyloid precursor... Enter a descriptive title for your BLAST search (a) Align two or more sequences (a) Choose Search Set Database Human genomic + transcript
 Mouse genomic + transcript
 Others (nr etc.): Nucleotide collection (nr/nt) Organism Enter organism name or id-completions will be suggested Exclude + Optional Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown (a) Exclude ■ Models (XM/XP) ■ Uncultured/environmental sample sequences Optional Limit to Sequences from type material Optional **Entrez Query** You Tube Create custom database Optional Enter an Entrez query to limit search (a)

BLAST [®] ≫ blastn suite ≫ RID-Y2CYBF17013

Home

Recent Resi

BLAST Results

Database Name nr

Description Nucleotide collection (nt)

Program BLASTN 2.5.0+ ▶ Citation

Edit and Resubmit Save Search Strategies > Formatting options > Download

You Tube Hov

gi|257379|gb|S45136.1| APP=amyloid precursor...

RID Y2CYBF17013 (Expires on 09-21 21:04 pm)

Query ID |cl|Query_105649

Description gi|257379|gb|S45136.1| APP=amyloid precursor protein {exon 16/17

junction} [human, familial Alzheimer disease family LIT, Genomic Mutant, 96

nt]

Molecule type nucleic acid

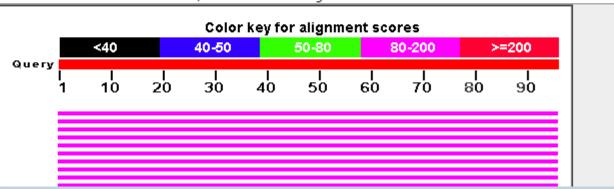
Query Length 96

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

☐ Graphic Summary

Distribution of 100 Blast Hits on the Query Sequence (

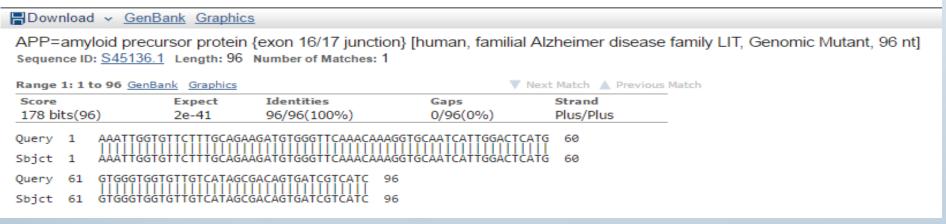
Mouse-over to show defline and scores, click to show alignments



Descriptions

Sequences producing significant alignments: Select: All None Selected:0 Alignments Download GenBank Graphics Distance tree of results Max Total Query E Description Accession score score cover value APP=amyloid precursor protein {exon 16/17 junction} [human, familial Alzheimer disease family LIT, Genomic Mutant, 96 nt] 178 100% 2e-41 100% S45136.1 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X6, mRNA 172 100% 8e-40 99% XM 009448780.2 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X5, mRNA 172 100% 8e-40 99% XM 009453143.2 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X4, mRNA 172 100% 8e-40 99% XM 009452766.2 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X3, mRNA 172 100% 8e-40 99% XM 009452222.2 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X2, mRNA 172 172 100% 8e-40 99% XM 009451769.2 PREDICTED: Pan troglodytes amyloid beta precursor protein (APP), transcript variant X1, mRNA 172 100% 8e-40 99% XM 009451357.2 Homo sapiens clone HTL-T-2 testicular tissue protein Li 2 mRNA, complete cds 172 172 100% 8e-40 99% HM005315.1

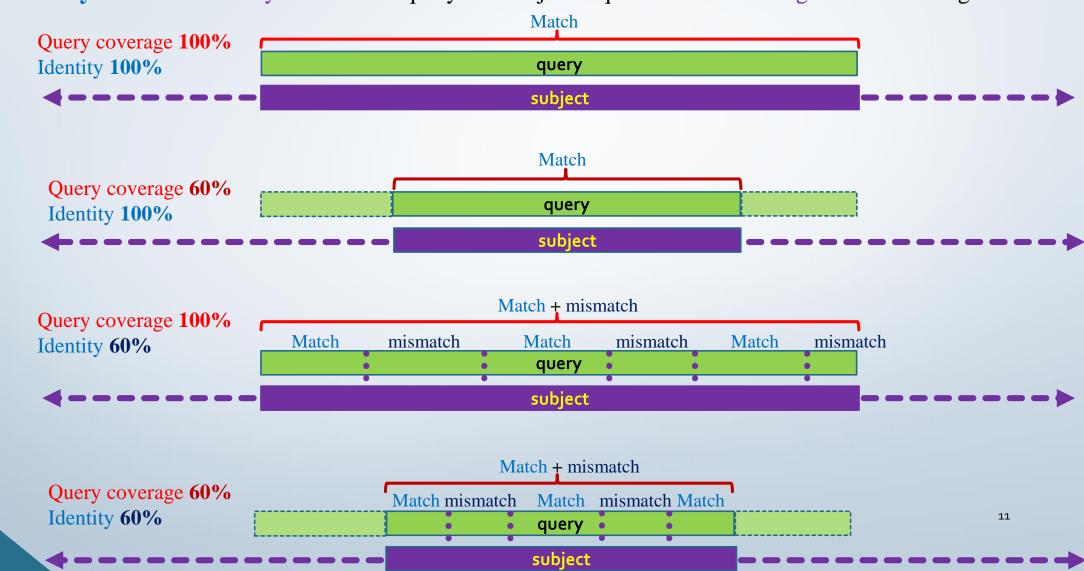
Alignments



Query Coverage vs. Identity

Query Coverage: Percent of the query sequence that overlaps the subject sequence

Identity: Percent similarity between the query and subject sequences over the length of the coverage area



Blastp

Human Hemoglobin Subunit Alpha 4504345

Mouse Hemoglobin Subunit Alpha 145301578

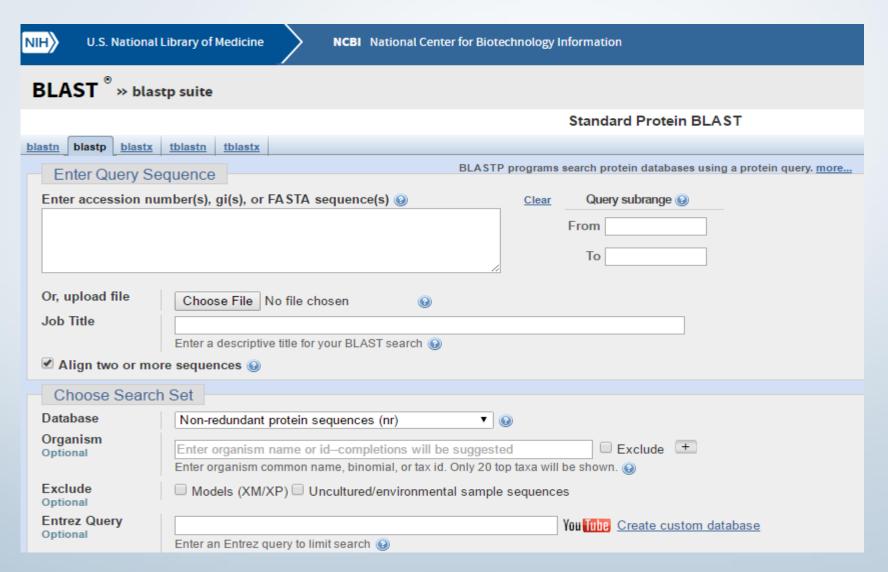
Human Hemoglobin Subunit Beta 4504349

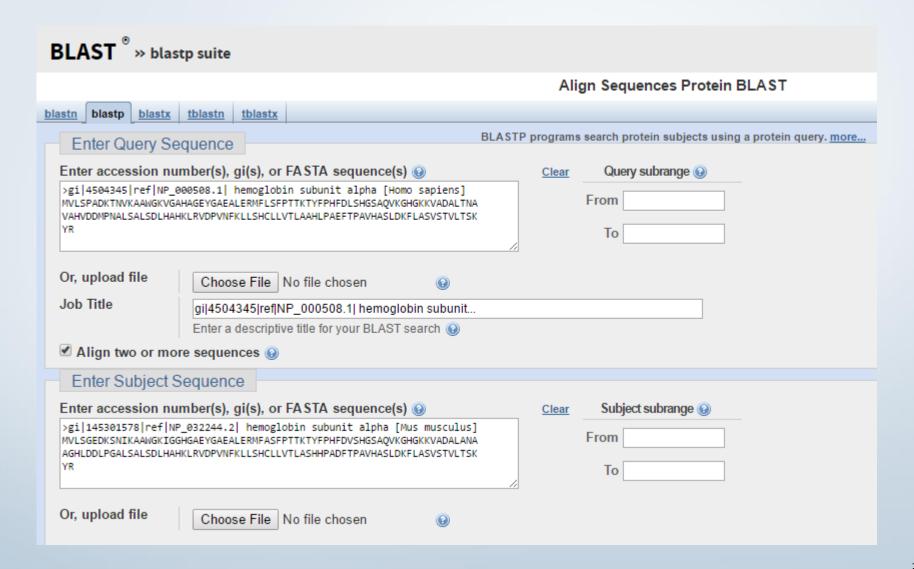
Human Myoglobin 4885477

Orthologs and Paralogs

- Orthologs: Homologous sequences in different species that arose from a common ancestral gene during speciation.
 - Ex. Humans and rats diverged around 80 million years ago ->
 divergence of myoglobin genes occurred.
 - Orthologs frequently have similar biological functions.
 - Human and rat myoglobin (oxygen transport)
 - Human and rat CaM
- Paralogs: Homologous sequences that arose by a mechanism such as gene duplication.
- Within same organism/species
- Ex. Myoglobin and beta globin are paralogs
 - Have distinct but related functions.

Blastp - NCBI





■Download ∨ Graphics

gi|145301578|ref|NP_032244.2| hemoglobin subunit alpha [Mus musculus] Sequence ID: Query_159045 Length: 142 Number of Matches: 1

Range	1: 1 to	142 Graph	nics			▼ Next Mate	ch 🛕 Previous Match
Score		Expect	Method		Identities	Positives	Gaps
255 bi	ts(65	1) 5e-94	Compositional m	natrix adjust.	123/142(87%)	131/142(92%)	0/142(0%)
Query	1		TNVKAAWGKVGAHAG +N+KAAWGK+G H				0
Sbjct	1	MVLSGEDK	SNIKAAWGKIGGHGA	EYGAEALERMFA	SFPTTKTYFPHFDV	SHGSAQVKGHG 6	0
Query	61		NAVAHVDDMPNALSA NA H+DD+P ALSA				20
Sbjct	61		NAAGHLDDLPGALSA				20
Query	121		FLASVSTVLTSKYR FLASVSTVLTSKYR	142			
Sbjct	121			142			

SIMILARITY

IDENTITY: EXTENT TO WHICH TWO SEQUENCES INVARIANT AMINO ACID MUST BE EXACT MATCH

Similarity: includes exact matches AND close matches
Some amino acids are chemically similar
isoleucine (I) and valine (V)
Arginine (R) and lysine (K)

THESE COUNT IN ADDITION TO EXACT MATCHES

LALIGN

Find Internal Duplications (lalign/plalign)	LALIGN/PLALIGN find internal duplications by calculating not sequences. LALIGN shows the alignments and similarity scores,
Search Databases with FASTA	FASTA Program information
Statistical Significance from Shuffles	TASTA Frogram information
Choose: (A) program and (B, C) sequences to compare:	
(A) Program: LALIGN: protein:protein ▼	
(B.1) Enter first (query) sequence: FASTA format ▼ Subset range:	
>gi 4504345 ref NP_000508.1 hemoglobin subunit alpha [Homo	Annotate Query Sequence (SwissProt accessions)
sapiens] MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKK	No annotation
VADALTNA	Upload annotation file: Choose File No file chosen
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLAS VSTVLTSK	
YR	Entrez protein / Entrez DNA sequence browser Uniprot sequence browser
(B.2) Or upload sequence from file: Choose File No file chosen • Protein • DNA (both-strands) • DNA (forward only) • DNA (rev-con	np only)
(C.1) Enter the second sequence: FASTA format ▼ Subset range:	
>gi 4504349 ref NP_000509.1 hemoglobin subunit beta [Homo sapiens]	Annotate Target Sequence (SwissProt accessions) No annotation

Upload annotation file: Choose File No file chosen

```
>>ref|NP 000509.1| hemoglobin subunit beta [Homo sapiens]
                                                                (147 aa)
Waterman-Eggert score: 375; 107.0 bits; E(1) < 1.3e-28
43.4% identity (74.5% similar) in 145 aa overlap (3-141:4-146)
Sequence Lookup Re-search database General re-search
                                                                           [Domains]
[alignment]
                   20
                                                                70
ref NP LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS----HGSAQVKGHGKKVADALTNAVAHVDD
     ref NP LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDN
         10
                   20
                           30
                                            50
                                                    60
                                                            70
       80
               90
                      100
                               110
                                       120
                                               130
                                                       140
ref NP MPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY
     ref|NP_LKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY
           90
                  100
                           110
                                   120
                                           130
                                                   140
```

ALIGNMENT: THREE OUTCOMES

Three possibilities at every position

Матсн

MISMATCH

GAP

MATCH/MISMATCH
WITH PROTEIN, SOME "BETTER" THAN OTHERS

GAP

Penalize for a gap

PENALIZE FOR A LENGTH OF GAP

BLOSUM 62 matrix

	С	S	Т	Р	Α	G	N	D	E	Q	Н	R	K	M	1	L	٧	F	Y	W	
С	9																				С
S	-1	4																			S
T	-1	1	5												4						T
P	-3	-1	-1	7		- 7															P
Α	0	1	90	-1	4																A
G	-3	0	-2	-2	0	6															G
N	-3	1	0	-2	-2	0	6														N
D	-3	0	-1	-1	-2	-1	1	6													D
E	-4	0	-1	-1	-1	-2	0	2	5			0									E
Q	-3	0	-1	-1	-1	-2	0	0	2	5											Q
Н	-3	-1	-2	-2	-2	-2	1	-1	0	0	8										Н
R	-3	-1	-1	-2	-1	-2	0	-2	0	1	0	5									R
K	-3	0	-1	-1	-1	-2	0	-1	1	1	-1	2	5								K
M	-1	-1	-1	-2	-1	-3	-2	-3	-2	0	-2	-1	-1	5							M
1	-1	-2	-1	-3	-1	-4	-3	-3	-3	-3	-3	-3	-3	1	4						1
L	-1	-2	-1	-3	-1	-4	-3	-4	-3	-2	-3	-2	-2	2	2	4					L
V	-1	-2	0	-2	0	-3	-3	-3	-2	-2	-3	-3	-2	1	3	1	4				V
F	-2	-2	-2	-4	-2	-3	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	6			F
Υ	-2	-2	-2	-3	-2	-3	-2	-3	-2	-1	2	-2	-2	-1	-1	-1	-1	3	7		Y
W	-2	-3	-2	-4	-3	-2	-4	-4	-3	-2	-2	-3	-3	-1	-3	-2	-3	1	2	11	W

http://www.acm.org/crossroads/wikifiles/13-1-Graphics/13-1-11/Image7-BLOSUM62.jpg

GLOBAL & LOCAL ALGORITHMS

NEEDLEMAN-WUNSCH (1970)
GLOBAL ALIGNMENT; ALIGNS START TO FINISH
GOOD FOR VERY CLOSE RELATIVES

SMITH-WATERMAN (1981)
LOCAL ALIGNMENT
BETTER FOR MOST SEQUENCES

Global Sequence Alignment	Local Sequence Alignment
In global alignment, an attempt is made to align the	Finds local regions with the highest level of
entire sequence (end to end alignment)	similarity between the two sequences.
A global alignment contains all letters from both the	A local alignment aligns a substring of the
query and target sequences	query sequence to a substring of the target sequence.
If two sequences have approximately the same length	Any two sequences can be locally aligned as
and are quite similar, they are suitable for global	local alignment finds stretches of sequences
alignment.	with high level of matches without
	considering the alignment of rest of the
	sequence regions.
Suitable for aligning two closely related sequences.	Suitable for aligning more divergent
	sequences or distantly related sequences.
Global alignments are usually done for comparing	
homologous genes like comparing two genes with	Used for finding out conserved patterns in
same function (in human vs. mouse) or comparing two	DNA sequences or conserved domains or
proteins with similar function.	motifs in two proteins.
A general global alignment technique is the	A general local alignment method is Smith-
Needleman–Wunsch algorithm.	Waterman algorithm.
Examples of Global alignment tools:	Examples of Local alignment tools:
EMBOSS Needle	BLAST
 Needleman-Wunsch Global Align Nucleotide 	EMBOSS Water
Sequences (Specialized BLAST)	• LALIGN

GLOBAL VS. LOCAL

EBI Pairwise Suite http://www.ebi.ac.uk/tools/psa

SELECT NEEDLE PROGRAM
NEEDLEMAN-WUNSCH FOR GLOBAL

SELECT WATER PROGRAM
SMITH-WATERMAN FOR LOCAL

This website uses cookies. By continuing to browse this site, you are agreeing to the use of our site cookies. To find out more, see our Terms of Use.

EMBL-EBI

Pairwise Sequence Alignment



Tools > Pairwise Sequence Alignment

Pairwise Sequence Alignment is used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences (protein or nucleic acid).

By contrast, <u>Multiple Sequence Alignment</u> (MSA) is the alignment of three or more biological sequences of similar length. From the output of MSA applications, homology can be inferred and the evolutionary relationship between the sequences studied.

Global Alignment

Global alignment tools create an end-to-end alignment of the sequences to be aligned. There are separate forms for protein or nucleotide sequences.

Needle @ (EMBOSS)

EMBOSS Needle creates an optimal global alignment of two sequences using the Needleman-Wunsch algorithm.

Nucleotide 🐧 🖎

Stretcher (EMBOSS)

Genomic Alignment

Genomic alignment tools concentrate on DNA (or to DNA) alignments while accounting for characteristics present in genomic data.

Wise2DBA @

Wise2DBA (DNA Block Aligner) aligns two sequences under the assumption that the sequences share a number of colinear blocks of conservation separated by potentially large and varied lengths of DNA in the two sequences.

Launch Wise2DBA

O---- 11/1--- 2

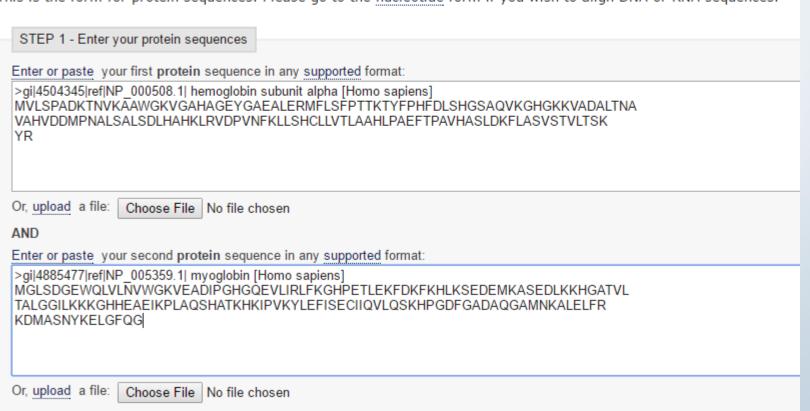
Global Alignment

Tools > Pairwise Sequence Alignment > EMBOSS Needle

Pairwise Sequence Alignment (PROTEIN)

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

This is the form for protein sequences. Please go to the nucleotide form if you wish to align DNA or RNA sequences.



Local Alignment

Tools > Pairwise Sequence Alignment > EMBOSS Water

Pairwise Sequence Alignment (PROTEIN)

EMBOSS Water uses the Smith-Waterman algorithm (modified for speed enhancements) to calculate the local alignment of two sequences.

This is the form for protein sequences. Please go to the nucleotide form if you wish to align DNA or RNA sequences.

STEP 1 - Enter your protein sequences
Enter or paste your first protein sequence in any supported format:
>gi 4504345 ref NP_000508.1 hemoglobin subunit alpha [Homo sapiens] MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK YR
Or, upload a file: Choose File No file chosen
AND
Enter or paste your second protein sequence in any supported format:
>gi 4885477 ref NP_005359.1 myoglobin [Homo sapiens] MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVL TALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFR KDMASNYKELGFQG
Or, upload a file: Choose File No file chosen

Local Alignment

```
#----
# Aligned_sequences: 2
# 1: NP 000508.1
# 2: NP 005359.1
# Matrix: EBLOSUM62
# Gap penalty: 10.0
# Extend penalty: 0.5
# Length: 149
# Identity:
             41/149 (27.5%)
# Similarity:
            62/149 (41.6%)
# Gaps:
             8/149 (5.4%)
# Score: 128.5
#-----
NP 000508.1
              1 MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF---
               NP 005359.1
              1 MGLSDGEWOLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHL
                                                         50
NP 000508.1
             48 --- DLSHGSAOVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRV
                                                         94
                  NP 005359.1
             51 KSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI
                                                         100
NP 000508.1
             95 DPVNF-KLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
                                                        142
                NP_005359.1
             101 -PVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYK
```

Global Alignment

```
#-----
# Aligned_sequences: 2
# 1: NP 000508.1
# 2: NP 005359.1
# Matrix: EBLOSUM62
# Gap penalty: 10.0
# Extend penalty: 0.5
# Length: 155
# Identity:
              41/155 (26.5%)
# Similarity:
             62/155 (40.0%)
# Gaps:
              14/155 ( 9.0%)
# Score: 128.5
#------
               1 MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF---
NP 000508.1
                 NP 005359.1
               1 MGLSDGEWOLVLNVWGKVEADIPGHGOEVLIRLFKGHPETLEKFDKFKHL
NP 000508.1
               48 --- DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRV
                    [....[...:]...]...]...:...........
               51 KSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI
NP 005359.1
                                                             100
               95 DPVNF-KLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR-
NP_000508.1
                                                             142
                  NP 005359.1
              101 -PVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKE
                                                             149
NP 000508.1
              143 -----
NP 005359.1
              150 LGFOG
                        154
```

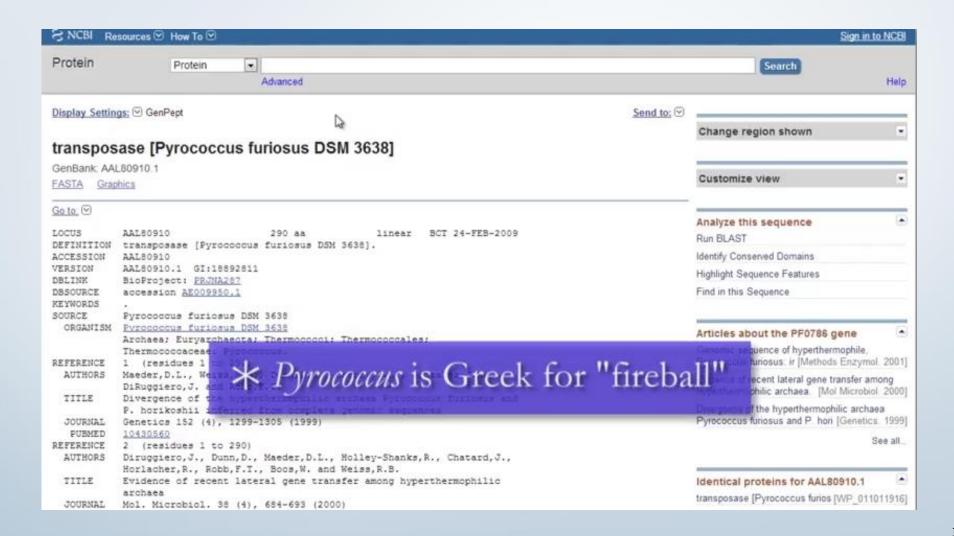
Identity, similarity

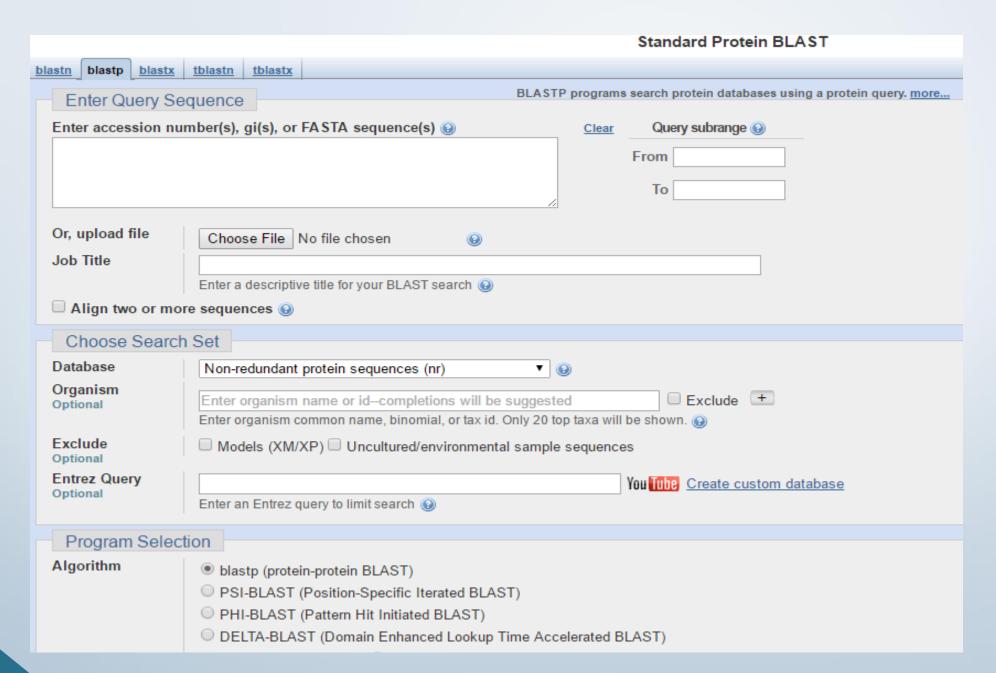
```
1 ...MKCLLLALALTCGAQALIVT..QTMKGLDIQKVAGTWYSLAMAASD. 44 lactoglobulin
                                               97 RBP
51 LFLQDNIVAEFSVDET
                   MSATAKGRVR.LLNNWD..VCADMVGTFT
45 ISLLDAQSAPLRV.YVI
                   LKPTPEGDLEILLQKWENGECAQKKIIAEI
                                              K 93 lactoglobulin
98 DPAKFKMKYWGVASFLO
                                              C 136 RBP
                   GNDDHWIVDTDYDTYAV...
94 IPAVFKIDALNENKVL.
                   .....VLDTDYKKYLLFCMENSAEPEQS 135 lactoglobulin
                                         Very
          Somewhat
                          AOKIVRORO.E
137 RLLNLDG
                                        similar
            similar
136 QCLVRTI
                           IRLSFNPTOLE
                                                      coglobulin
                                     (two dots)
           (one dot)
```

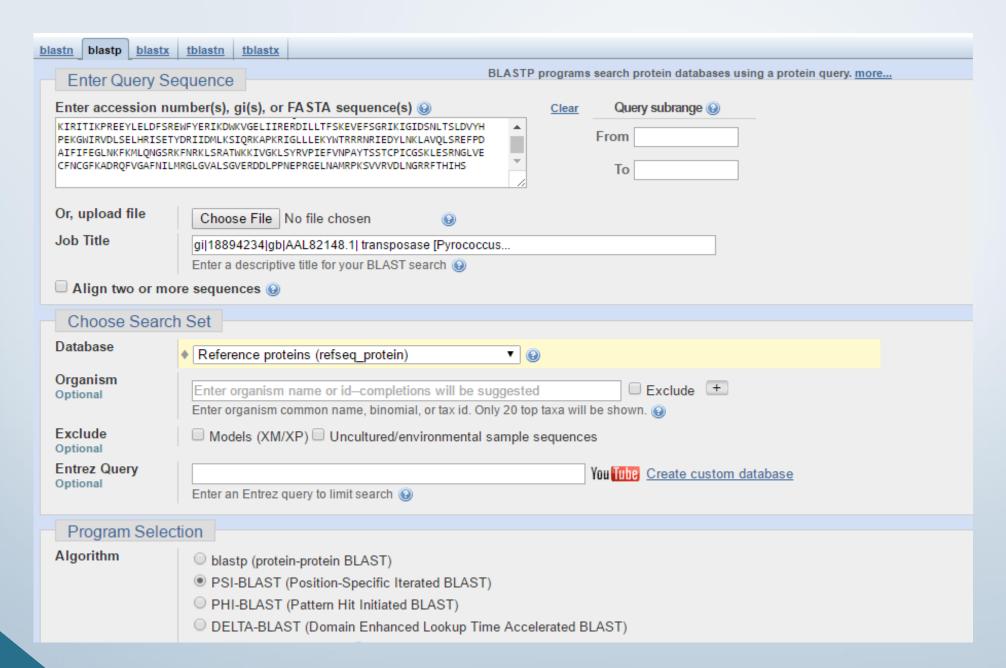
Different algorithms have different output notations

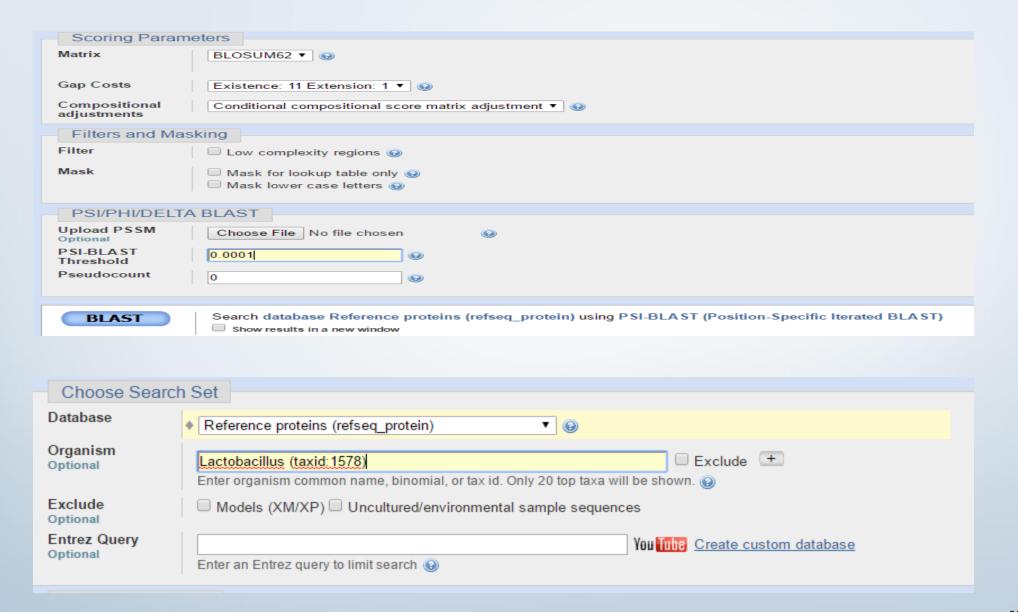
PSI BLAST

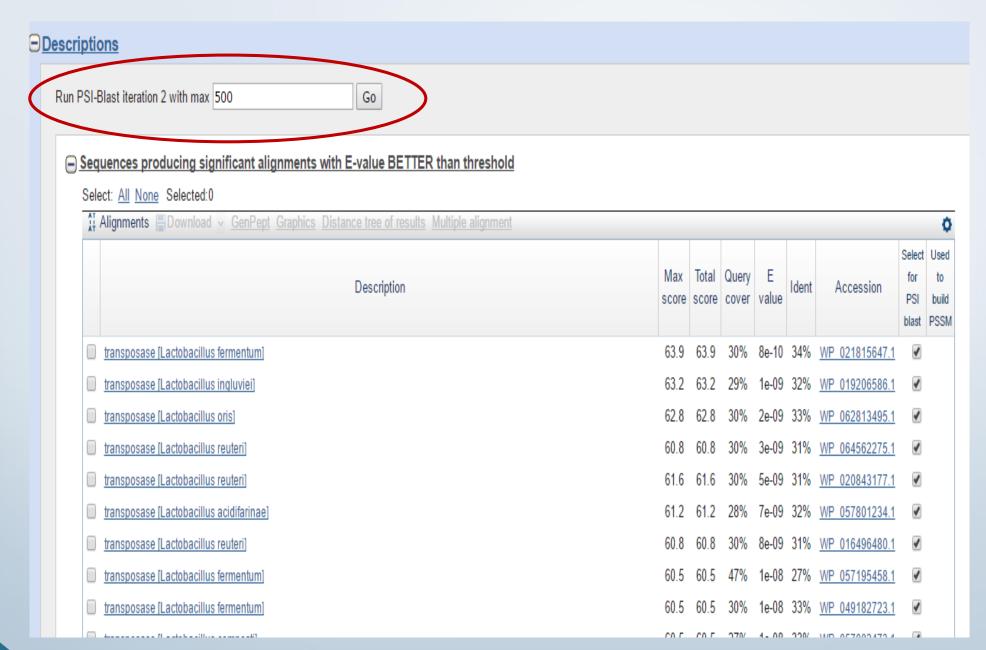
Transposase [Pyrococcus furiosus DSM 3638]











Transposase It.actobacillus helveticus]	79.8	79.8	15%	4e-18	34%	WP 023190705.1	
hypothetical protein LCRIS 00367, partial [Lactobacillus crispatus ST1] > ref[WP 013085846.1] transposase, partial [Lactobacillus crispatus ST1] > ref[WP 013085846.1] transposase, partial [Lactobacillus crispatus ST1]	79.8	79.8	15%	4e-18	36%	YP 003600839.1	7
Transposase (Ladobacillus helveticus)	79.4	79.4	15%	5e-18	34%	WP 023191133.1	V
transposase (Lactobacillus crispatus)	78.2	78.2	15%	1e-17	34%	WP 005726790.1	V
transposase (Ladobacillus heiveticus)	78.2	78.2	19%	1e-17	25%	WP 003624887.1	V
transposase [Lactobacillus helveticus]	78.2	78.2	17%	1e-17	21%	WP 003624884.1	V
[ransposase Lactobacillus helveticus]	77.8	77.8	15%	2e-17	34%	WP 003628277.1	V
possible transposase; partial fl.actobacilius crispatus)	78.2	78.2	29%	2e-17	22%	WP. 005721240.1	V
hypothetical protein LCRIS 00917, partial ILactobacillus crispatus ST11 > reffWP 013086220.11 transposase, partial ILactobacillus cri	77.1	77.1	15%	3e-17	34%	YP 003601389.1	7
transposase (Lactobacillus crispatus)	76.7	76.7	14%	4e-17	40%	WP 005729525.1	V
putative transposase, partial (Lactobacillus helveticus)	77.1	77.1	16%	4e-17	24%	WP 003628351.1	7
hypothetical protein, partial (Lactobacillus crispatus)	77.4	77.4	28%	5e-17	24%	WP 021355801.1	7
transposase, partial it actobacillus helveticus)	77.1	77.1	30%	1e-16	27%	WP 003632589 1	7
transposase (Lactopacillus crispatus)	74.4	74.4	14%	3e-16	36%	WP 005725843.1	V
* Psi-Blast is good for finding pote	<u>en</u>	tia	<u>1</u> ł	or	no	ologs	
that BI AST might have missed							
that BLAST might have missed							
		66.3	20%	1e-12	16%	WP 003616238.1	100
transposase, partial (Lactobacillus delbruecini)	66.3	00.5					18.1
transposase, partial (Lactobacillus delbrueckis) putative transposase, partial (Lactobacillus helveticus)	66.3	65.1	27%	2e-12	24%	WP 003628409.1	

Alignment Algorithms / Softwares

