No. of the Experiment: 01

Name of the Experiment: Primer Designing of *Homo sapiens* Beta-globin (HBB) gene for PCR and molecular diagnosis.

Principle: A primer is a short strand of RNA or DNA (generally about 18-22 bases) that serves as a starting point for DNA synthesis. It is required for DNA replication because the enzymes that catalyze this process, DNA polymerases, can only add new nucleotides to an existing strand of DNA. The polymerase starts replication at the 3'-end of the primer, and copies the opposite strand. Many in vitro laboratory techniques that involve DNA polymerase in biochemistry and molecular biology (such as DNA sequencing and the polymerase chain reaction), use DNA primers because they are more temperature stable. In experiments, it is often important to use a primer with a similar Tm (melting temperature) to the template strand it will be hybridizing to. A primer with a Tm significantly higher than the reaction's annealing temperature may mis-hybridize and extend at an incorrect location along the DNA sequence, while one with a Tm significantly lower than the annealing temperature may fail to anneal and extend at all. These primers are usually short, chemically synthesized oligonucleotides, with a length of about twenty bases. They are hybridized to a target DNA, which is then copied by the polymerase.

Materials:

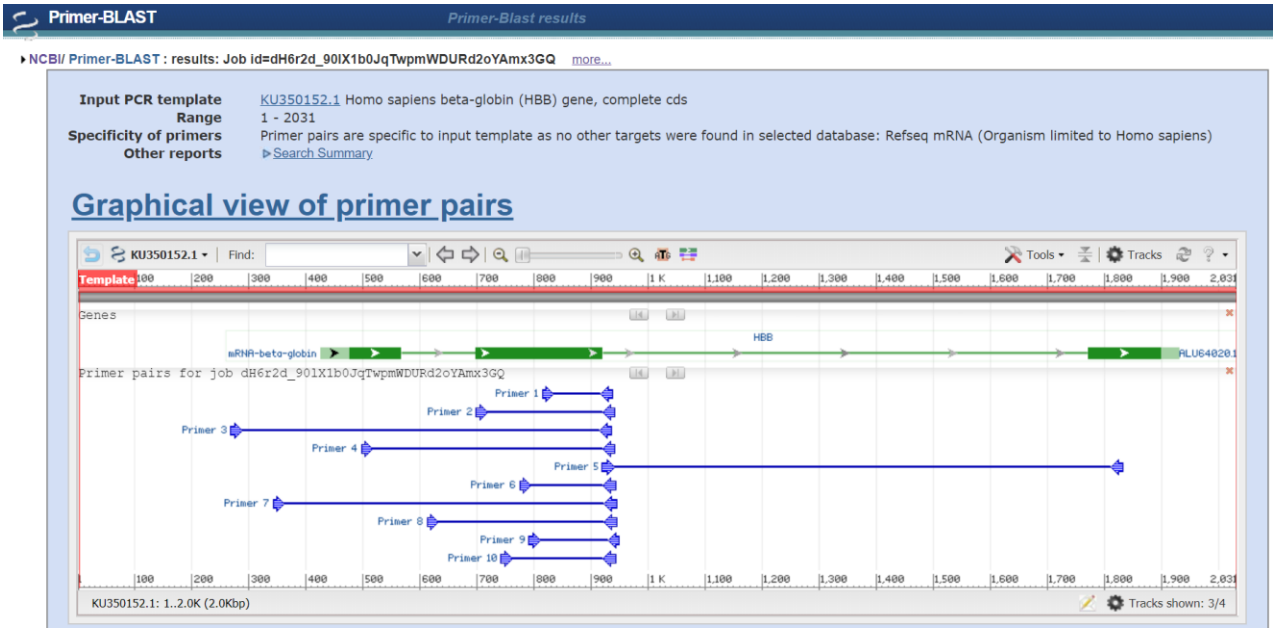
1. Computer
2. Internet access

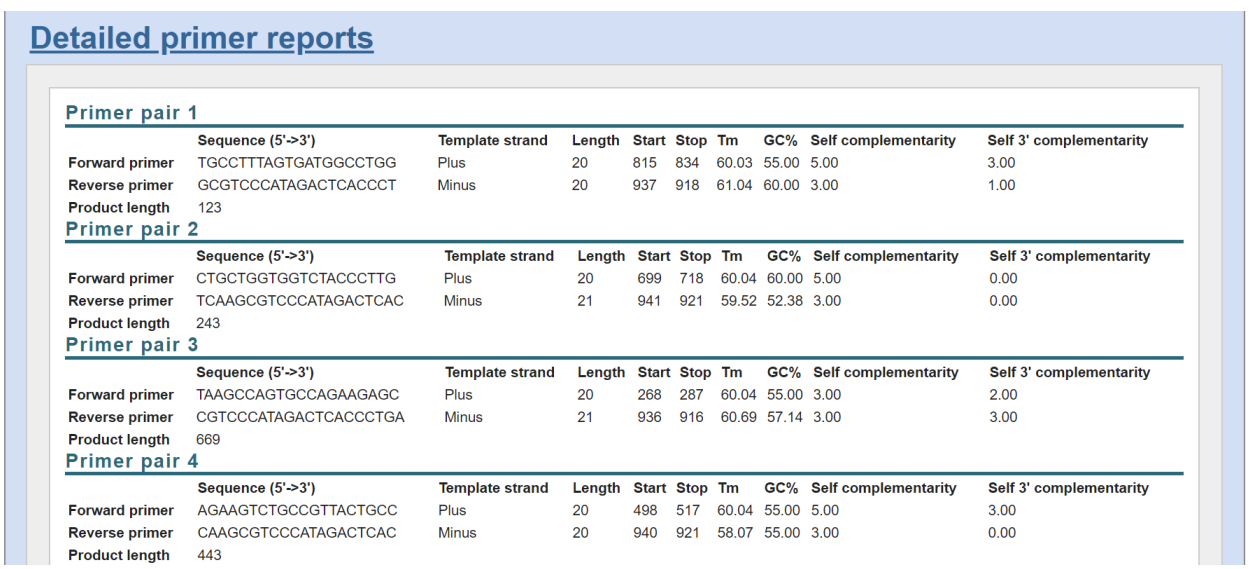
Procedure:

1. The DNA sequence of *Homo sapiens* Beta-globin (HBB) gene is downloaded as FASTA format from NCBI and selected "pick primer".

2. Then the webpage https://www.ncbi.nlm.nih.gov/tools/primer-blast/ is visited and FASTA format is pasted.
3. Then “Get primers” is clicked.
4. Primers specific to our PCR template is processing and we have to wait for several minutes.
5. Then the amplicon sequence is digested with specific restriction enzyme.

Result:





Detailed primer reports

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGCCTTTAGTGATGGCCTGG	Plus	20	815	834	60.03	55.00	5.00	3.00
Reverse primer	GCGTCCCATAGACTCACCCT	Minus	20	937	918	61.04	60.00	3.00	1.00
Product length	123								

Primer pair 2

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CTGCTGGTGGTCTACCCTTG	Plus	20	699	718	60.04	60.00	5.00	0.00
Reverse primer	TCAAGCGTCCCATAGACTCAC	Minus	21	941	921	59.52	52.38	3.00	0.00
Product length	243								

Primer pair 3

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TAAGCCAGTGCCAGAAGAGC	Plus	20	268	287	60.04	55.00	3.00	2.00
Reverse primer	CGTCCCATAGACTCACCCTGA	Minus	21	936	916	60.69	57.14	3.00	3.00
Product length	669								

Primer pair 4

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AGAAGTCTGCCGTTACTGCC	Plus	20	498	517	60.04	55.00	5.00	3.00
Reverse primer	CAAGCGTCCCATAGACTCAC	Minus	20	940	921	58.07	55.00	3.00	0.00
Product length	443								

Primer pair 5

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGTGAGTCTATGGGACGCTTG	Plus	21	920	940	60.47	57.14	3.00	0.00
Reverse primer	GGTGGGGTGAATTCTTTGCC	Minus	20	1832	1813	59.39	55.00	6.00	1.00
Product length	913								

Primer pair 6

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GCAACCCTAAGGTGAAGGCT	Plus	20	775	794	59.96	55.00	6.00	3.00
Reverse primer	CATCAAGCGTCCCATAGACTCA	Minus	22	943	922	59.90	50.00	3.00	1.00
Product length	169								

Primer pair 7

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGGTTGGCCAATCTACTCCC	Plus	20	343	362	60.11	60.00	8.00	0.00
Reverse primer	ACATCAAGCGTCCCATAGACTC	Minus	22	944	923	59.90	50.00	3.00	3.00
Product length	602								

Primer pair 8

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CTGGGCATGTGGAGACAGAG	Plus	20	613	632	60.11	60.00	4.00	0.00
Reverse primer	ACATCAAGCGTCCCATAGACT	Minus	21	944	924	58.89	47.62	3.00	3.00
Product length	332								

Primer pair 9

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGCTCATGGCAAGAAAGTGC	Plus	20	791	810	60.11	55.00	5.00	2.00
Reverse primer	AGAAAACATCAAGCGTCCCA	Minus	20	949	930	57.73	45.00	2.00	0.00
Product length	159								

Primer pair 10

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TTGGGGATCTGTCCACTCCT	Plus	20	742	761	59.88	55.00	4.00	0.00
Reverse primer	ATCAAGCGTCCCATAGACTCA	Minus	21	942	922	58.61	47.62	3.00	1.00
Product length	201								

For primer pair 3:

Forward primer: TAAGCCAGTGCCAGAAGAGC
Reverse primer: CGTCCCATAGACTCACCCTGA
Reverse complementary of reverse primer: TCAGGGTGAGTCTATGGGACG

Our query sequence:

ATTCTGGAGACGCAGGAAGAGATCCATCTACATATCCCAAAGCTGAATTATGGTAGACAAAACTCTTCCA
CTTTTAGTGCATCAACTTCTTATTTGTGTAATAAGAAAATTGGGAAAACGATCTTCAATATGCTTACCAA
GCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGAGGATGTTTTTAGTAGCAATTTGTACTGATG
GTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACTCC**TAAGCCAGTGCCA**
GAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCCTGTGGAGCCACACCCTAGGGTTGGC
CAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCGAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTAT
TGCTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACT
CCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCC
TGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAG
AGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCT
GCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTT
ATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACC
TGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGA
GAACT**TCAGGGTGAGTCTATGGGACG**CTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGT
CATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTGGA

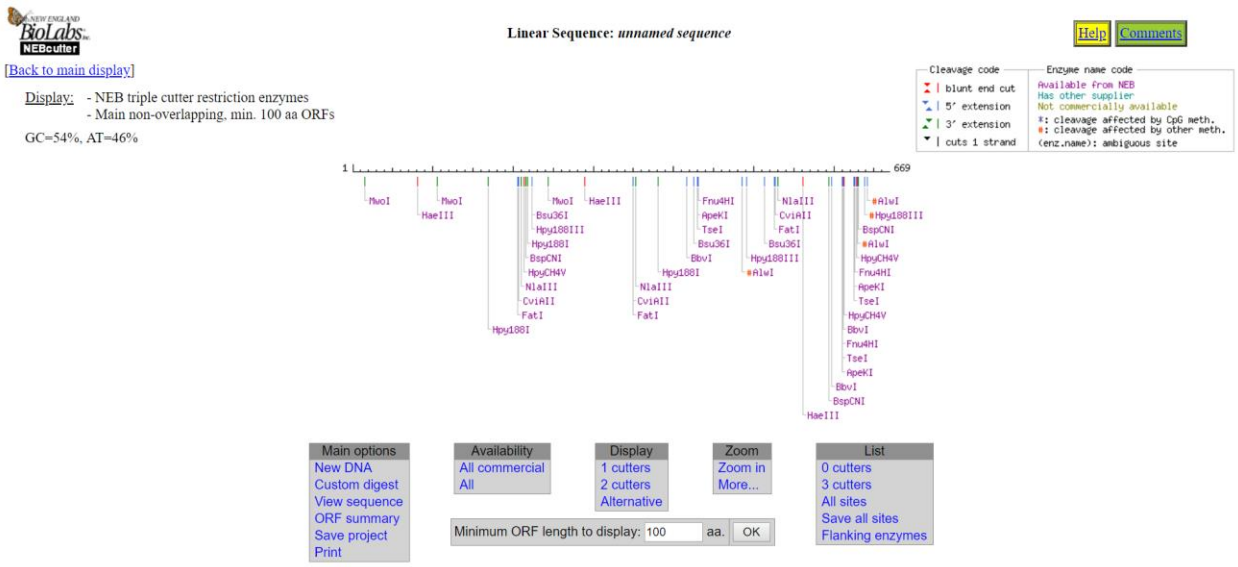
Forward primer

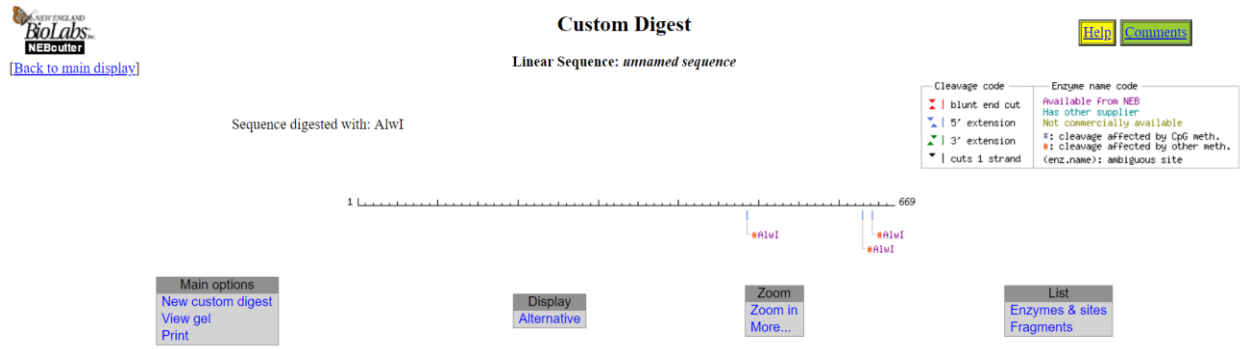
Reverse primer

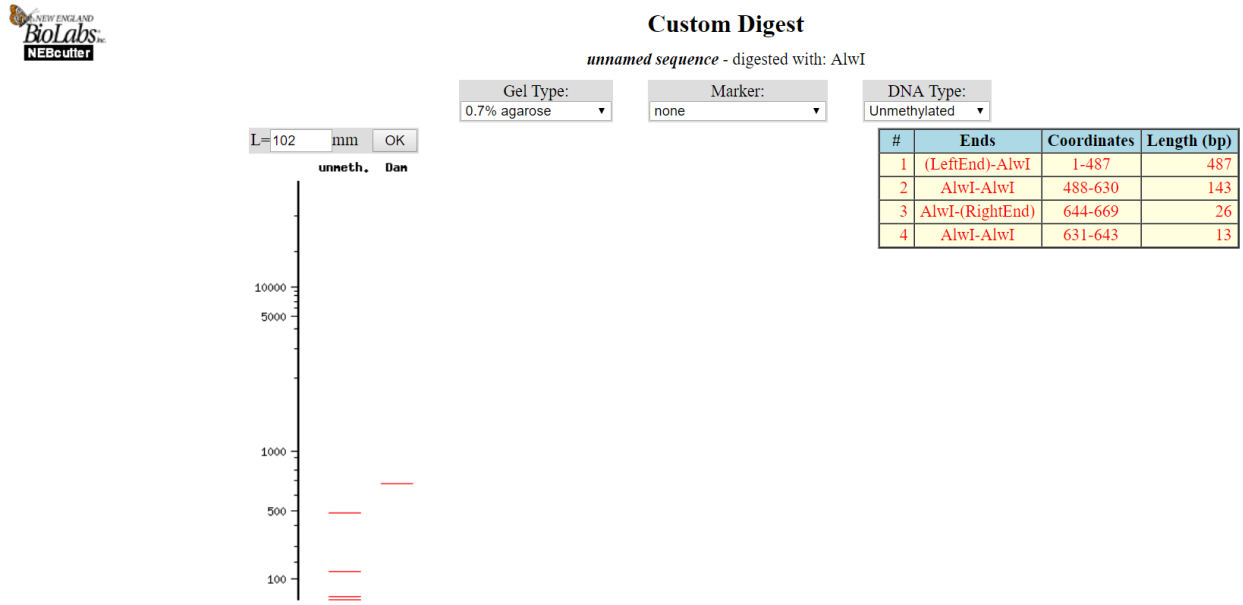
AGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTCATAACAATTGTTTTCTTTTGTTTAATTCTTGC
TTTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAAAAGG
AAATATCTCTGAGATACATTAAGTAACTTAAAAAAAAACTTTACACAGTCTGCCTAGTACATTACTATTT
GGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATTTTTAATTGATAC
ATAATCATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACATATTGACCAAATCAG
GGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTATCTTATTTC
TAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGTATCATGCCTCTTTGCACCATTC
TAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCAATATCTCTGCATATAAATATTTCTGCATA
TAAATTGTAACTGATGTAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTT
ATTTTATGGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCATAC
CTCTTATCTTCCTCCCACAGCTCCTGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGA
ATTCACCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAG
TATCACTAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACT
AAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTTCATTG
C

Amplicon of primer pair 3: Wild Type

5′TAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCCTGTGGAGCCACACCCTA
GGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCGAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTA
TTGCTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAG
GAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGT
ATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTG
ATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAG
GTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGA
AAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTG
CACTGTGACAAGCTGCACGTGGATCCTGAGAACT**TCAGGGTGAGTCTATGGGACG3′**





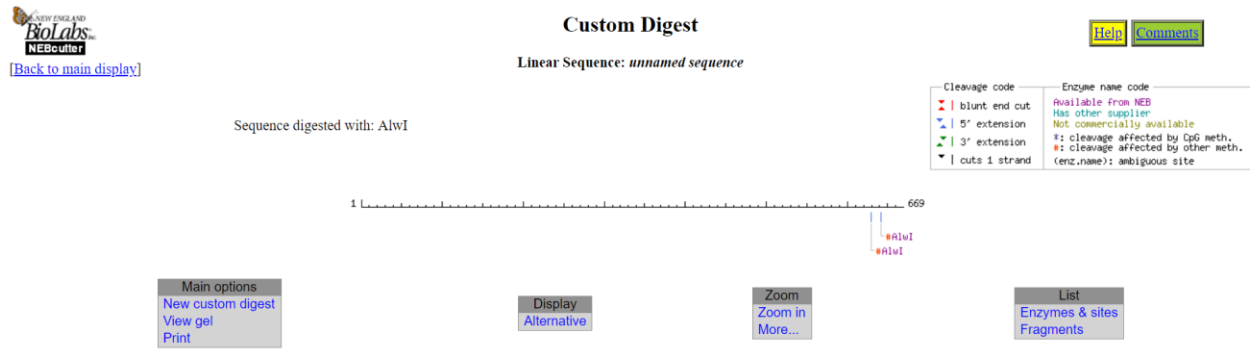


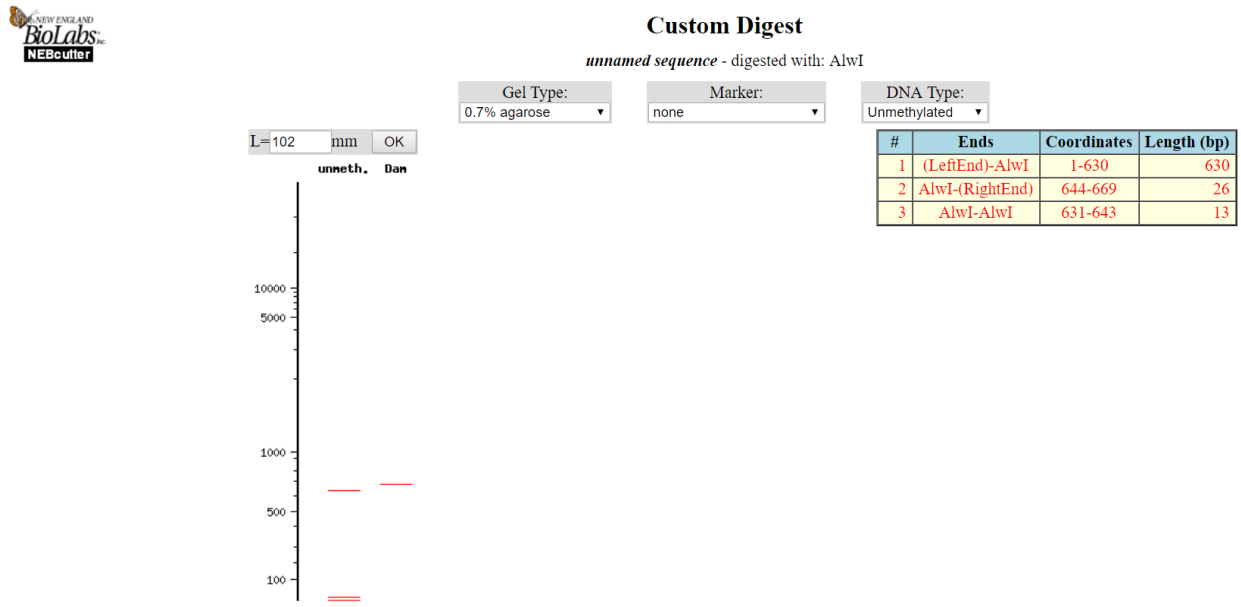
#	Ends	Coordinates	Length (bp)
1	(LeftEnd)-AlwI	1-487	487
2	AlwI-AlwI	488-630	143
3	AlwI-(RightEnd)	644-669	26
4	AlwI-AlwI	631-643	13

Interpretation: The wild type sequence have all the restriction sites and after digested with AlwI, 4 fragments are identified during gel run. The highest fragment length is 487bp and lowest is 13bp.

Amplicon of primer pair 3: Mutate Type

5'TAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCCTGTGGAGCCACACCCTA
GGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCGAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTA
TTGCTTACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAG
GAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGT
ATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTG
ATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAG
GTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGA
AAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTG
CACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACG3'





#	Ends	Coordinates	Length (bp)
1	(LeftEnd)-AlwI	1-630	630
2	AlwI-(RightEnd)	644-669	26
3	AlwI-AlwI	631-643	13

Interpretation: The mutate type sequence contains one mutation and if cut by restriction enzyme AlwI, then 3 fragments is found instead of 4. The highest fragment length is now 630 and lowest is 13.

No. of the Experiment: 02

Name of the experiment: Identification of primer homology sequence by BLAST.

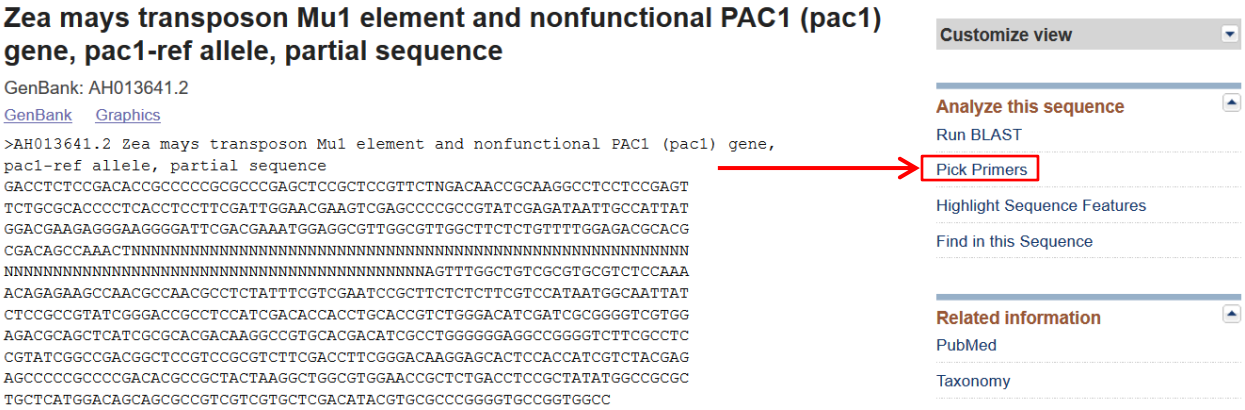
Principle: A primer is a short strand of RNA or DNA (generally about 18-22 bases) that serves as a starting point for DNA synthesis. Many in vitro laboratory techniques that involve DNA polymerase in biochemistry and molecular biology (such as DNA sequencing and the polymerase chain reaction), use DNA primers because they are more temperature stable. BLAST finds regions of local similarities between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST of a primer, shows most disimilar sequences which are specific.

Material:

1. Computer
2. Internet access

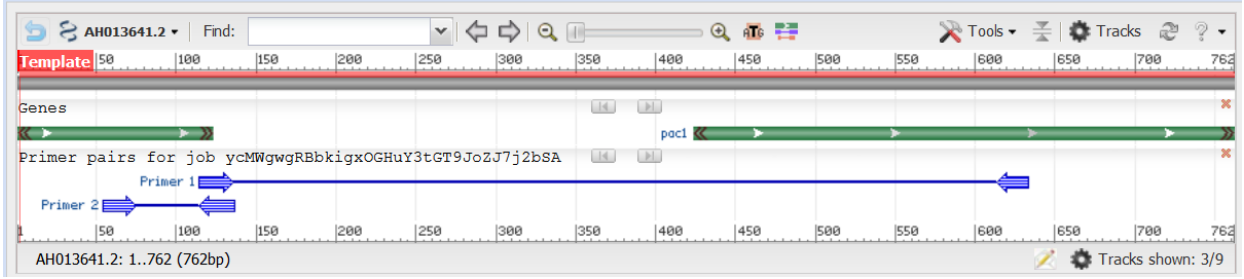
Procedure:

6. The DNA sequence of Zea mays transposon Mu1 element and nonfunctional PAC1 (pac1) gene of FASTA format is searched from NCBI and selected "pick primer".



7. Then the webpage https://www.ncbi.nlm.nih.gov/tools/primer-blast/ is visited and FASTA format is pasted.
8. Then "Get primers" is clicked.
9. Primers specific to our PCR template is processing and we have to wait for several minutes.

Graphical view of primer pairs



Detailed primer reports

Primer pair 1

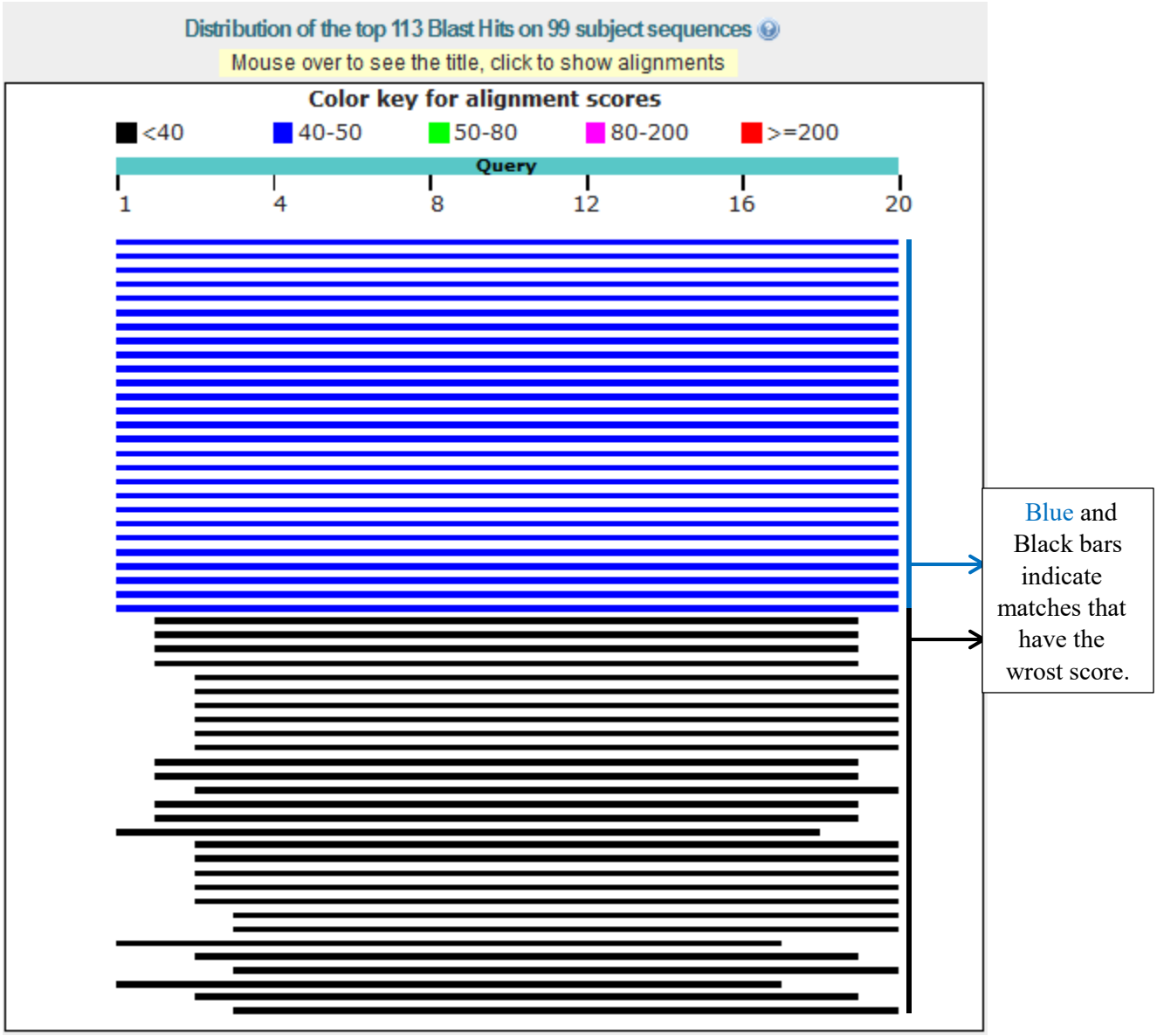
	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CGCCGTATCGAGATAATTGCC	Plus	21	115	135	59.35	52.38	6.00	2.00
Reverse primer	GCTCTCGTAGACGATGGTGG	Minus	20	633	614	59.97	60.00	6.00	0.00
Product length	519								

Primer pair 2

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AAGGCCTCCTCCGAGTTCTG	Plus	20	55	74	61.26	60.00	6.00	2.00
Reverse primer	TGGCAATTATCTCGATACGGCG	Minus	22	136	115	60.92	50.00	6.00	3.00
Product length	82								

10. After the generation of primers, reverse primer is copied from primer pair 1.
11. Then browsed https://blast.ncbi.nlm.nih.gov/BLAST.cgi and 'Nucleotide BLAST (blastn)' is selected.
7. The copied reverse primer is pasted into box where it says "Enter Query Sequence" and "run blast" is selected.

Result: The graphical display shows the similar sequences.



Sequences producing significant alignments:

Select: All None Selected:7

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☑	Hordeum vulgare cultivar OWB-REC WD40 transcription factor (WD40) gene, complete cds	40.1	40.1	100%	0.29	100%	MF679152.1
☐	Hordeum vulgare cultivar OWB-DOM WD40 transcription factor (WD40) gene, complete cds	40.1	40.1	100%	0.29	100%	MF679151.1
☐	Hordeum vulgare cultivar NGB22812 WD40 transcription factor (WD40) gene, complete cds	40.1	40.1	100%	0.29	100%	MF679150.1
☐	Hordeum vulgare cultivar NGB20651 WD40 transcription factor (WD40) gene, complete cds	40.1	40.1	100%	0.29	100%	MF679149.1
☑	Triticum aestivum mRNA, clone: tplb0043l24, cultivar Chinese Spring	40.1	40.1	100%	0.29	100%	AK451018.1
☑	PREDICTED: Sorghum bicolor protein TRANSPARENT TESTA GLABRA 1 (LOC8076027), mRNA	40.1	40.1	100%	0.29	100%	XM_002452704.2
☑	PREDICTED: Aegilops tauschii subsp. tauschii protein TRANSPARENT TESTA GLABRA 1 (LOC109740733), transcript variant	40.1	40.1	100%	0.29	100%	XM_020299784.1
☐	PREDICTED: Aegilops tauschii subsp. tauschii protein TRANSPARENT TESTA GLABRA 1 (LOC109740733), transcript variant	40.1	40.1	100%	0.29	100%	XM_020299783.1
☐	PREDICTED: Elaeis guineensis protein TRANSPARENT TESTA GLABRA 1-like (LOC105043285), transcript variant X2, mRNA	40.1	40.1	100%	0.29	100%	XM_010920775.2
☐	PREDICTED: Elaeis guineensis protein TRANSPARENT TESTA GLABRA 1-like (LOC105043285), transcript variant X1, mRNA	40.1	40.1	100%	0.29	100%	XM_010920773.2
☑	Zea mays transposon Mu1 element and nonfunctional PAC1 (pac1) gene, pac1-ref allele, partial sequence	40.1	40.1	100%	0.29	100%	AH013641.2
☐	PREDICTED: Phoenix dactylifera protein TRANSPARENT TESTA GLABRA 1-like (LOC103712797), transcript variant X2, mRN	40.1	40.1	100%	0.29	100%	XM_008799422.2
☑	Clostridiales sp. SS3/4 draft genome	36.2	36.2	90%	4.5	100%	FP929062.1
☐	Chaetomium globosum CBS 148.51 hypothetical protein (CHGG_02097) partial mRNA	36.2	36.2	90%	4.5	100%	XM_001221317.1
☑	Bosea sp. PAMC 26642, complete genome	34.2	64.4	90%	18	100%	CP014301.1

Alignments:

Download GenBank Graphics

Hordeum vulgare cultivar OWB-REC WD40 transcription factor (WD40) gene, complete cds

Sequence ID: MF679152.1 **Length:** 1071 **Number of Matches:** 1

Range 1: 662 to 681 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
40.1 bits(20)	0.29	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  1    GCTCTCGTAGACGATGGTGG  20
            ||||||||||||||||||||
Sbjct  681  GCTCTCGTAGACGATGGTGG  662
```

Download GenBank Graphics

PREDICTED: Acanthaster planci unhealthy ribosome biogenesis protein 2 homolog (LOC110986494),

Sequence ID: XM_022248396.1 **Length:** 4586 **Number of Matches:** 1

Range 1: 626 to 643 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
36.2 bits(18)	4.5	18/18(100%)	0/18(0%)	Plus/Minus

```
Query  2    CTCTCGTAGACGATGGTG  19
            ||||||||||||||||||
Sbjct  643  CTCTCGTAGACGATGGTG  626
```

Download GenBank Graphics Sort by: E value

Rhodobacter sp. CZR27 chromosome, complete genome

Sequence ID: CP023548.1 **Length:** 3335106 **Number of Matches:** 2

Range 1: 2748913 to 2748929 GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
34.2 bits(17)	18	17/17(100%)	0/17(0%)	Plus/Minus

```
Query  4        CTCGTAGACGATGGTGG  20
                |||||||||||||||||
Sbjct  2748929  CTCGTAGACGATGGTGG  2748913
```

Range 2: 2103457 to 2103471 GenBank Graphics Next Match Previous Match First Match

Score	Expect	Identities	Gaps	Strand
30.2 bits(15)	277	15/15(100%)	0/15(0%)	Plus/Minus

```
Query  6        CGTAGACGATGGTGG  20
                |||||||||||||||
Sbjct  2103471  CGTAGACGATGGTGG  2103457
```

Interpretation: The primer is less similar to other sequences as the highest score is 40.1 which means our primer is highly specific for only particular sequences. Primer which are specific, used for amplification. Less specific primer may binds to any part of the sequences but the highly specific primer will bind only the specific region of the sequence.

No. of the Experiment: 03

Name of the Experiment: Restriction digestion of DNA sequence by NEBcutter

Principle: *NEBcutter* is a program to cleave DNA with restriction enzymes. It is a database where a user can paste a DNA sequence they have obtained (via sequencing techniques) and see what, if any, restriction enzyme cut sites. NEBcutter, version 2.0, is a program available via a web server (http://www.tools.neb.com/NEBcutter) that will accept an input DNA sequence and produce a comprehensive report of the restriction enzymes that will cleave the sequence. It produces a variety of outputs including restriction enzyme maps, theoretical digests and links into the restriction enzyme database.

Material:

1. Computer
2. Internet access

Procedure:

1. Nucleotide sequence of *Triticum aestivum* has been downloaded as FASTA format from NCBI.

FASTA Send to:

Triticum aestivum strain CRB-INRA-CFD-15658 zeta carotene desaturase (ZdsB) gene, partial cds

Accession number GenBank: EF115224 1 **Version**

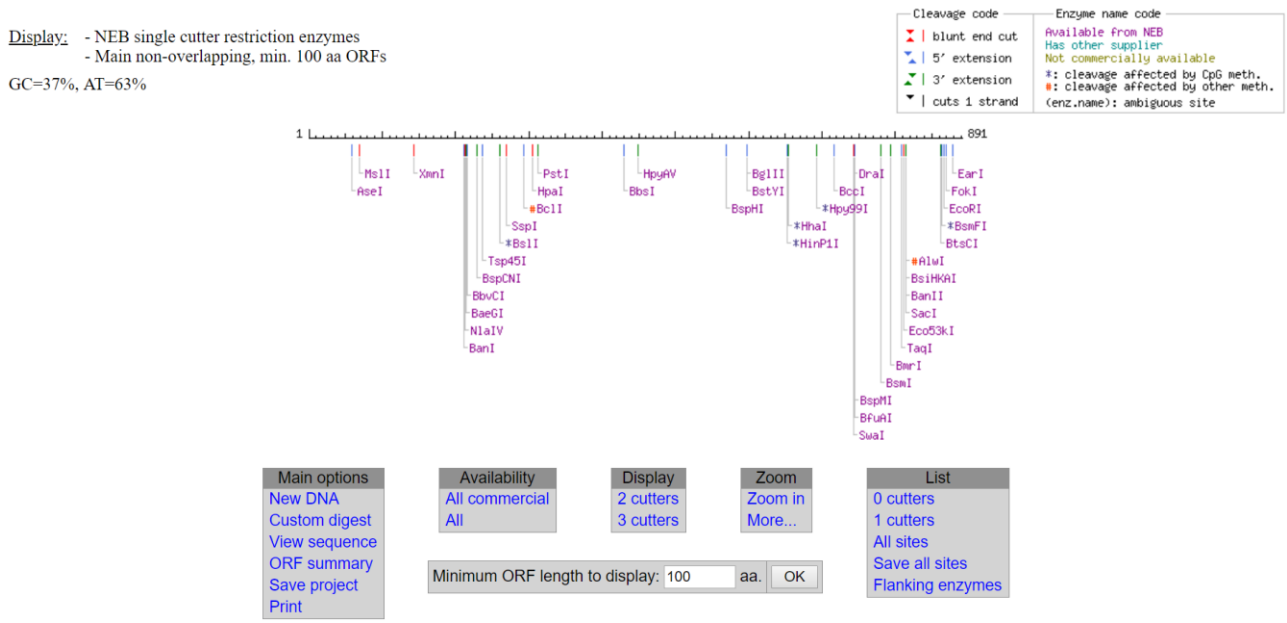
GenBank Graphics PopSet

```
>EF115224.1 Triticum aestivum strain CRB-INRA-CFD-15658 zeta carotene desaturase (ZdsB)
gene, partial cds
TCTTACAGTTTGTAATTACACATTGCAACTTATTGTAACATTTTTGCCTTGCACTTTATTAATATCATGC
CCATGATTCTTGCAGTATCTCTCTATTTGAATCTGCAACTTAACTGTAGTTTTGTATGTTATTCCCTCTG
AACATCTTCTGAAGTTCATCTGAAAGCATATCTTTTCAGCCCATGCCACAATTTTGTGAAAGTAGTAATC
GGGCACCTCAGCACTCCCTTTTCATTAGGTCACAATGCAATCAGCCGAACTATGGGCAAATATTTCTTGA
AAGCAGTTACAGTTGATCACTCTGTTAACTGCAGCAAGCTATTCCCTTTGAAAATAAATAGCTTGCAGCT
CATGCTAATTTCAGATACAGTACAGTTTGTTACTCCCTCCGTAACATAATATAAGACGTTTTTTGACACT
ATGCCAGTGTCAAAAACGTCTTCTATTAAGTTACAGAAGGAGTACAATTTGTGTGTTCTGTTATTCTTCT
GTCTTATGGATAATCAAAGCATTGCAAAGTCCTCACAACTGCCTATTTGTGCAAATACCTTAAATAGTAT
ATACCTAATTCATGACCACCCTATCTTCAGGTTTATTGAGATCTGAATTTTCTGTCTGAACTTGTAACAA
GTGAAAGATTCTATAAGCGTGAGCGCAAAATAAGCACTGGAAGTCTGGTCAACTTTGCCCGTCGCTTTAG
TTTGCTGCATATTGCTTTGTCAGATGGTTTAATTGTATATATTTAAATTGTGTTGCAGGTTTATGATAAG
GCAAGGAATGCAGTTGCTCTTGCCCTAAGCCCAGTTGTTCGAGCTCTTGTTGATCCAGACGGTGCATTGC
AACAAGTACGGGACTTGGATGATGTAAGAATTCCTCTTCTTTTTGTCAGTG
```

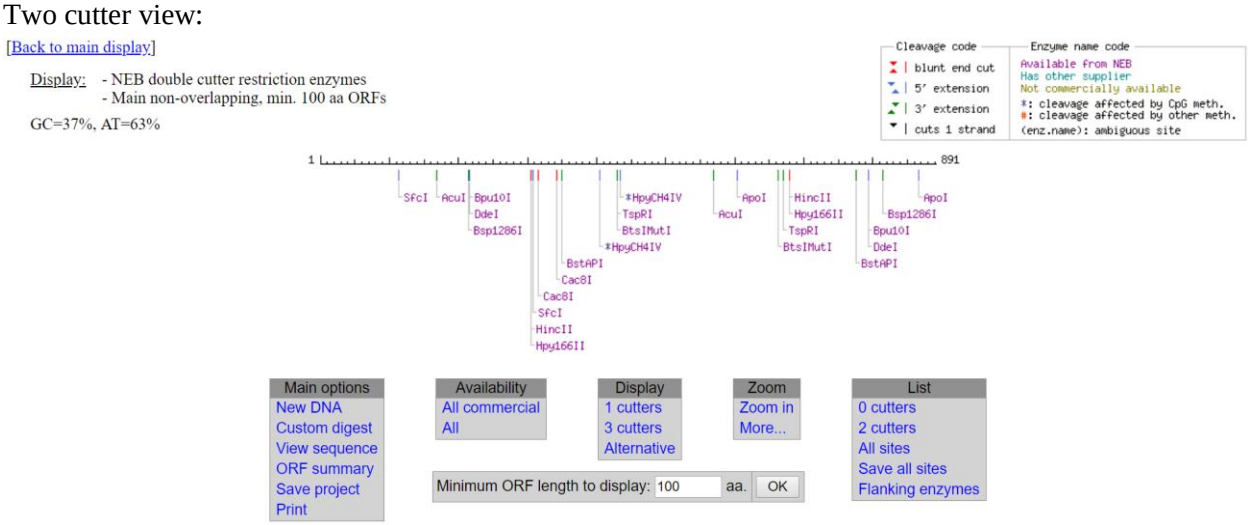
2. Then the website http://www.tools.neb.com/NEBcutter is visited.
3. FASTA format of nucleotide sequence of *Triticum aestivum* is pasted in the box and submitted.

Results:

One cutter view:



Two cutter view:



NEW ENGLAND BioLabs Inc. NEBcutter

[Back to main display]

Enzymes for Custom Digest

EF115224.1 Triticum

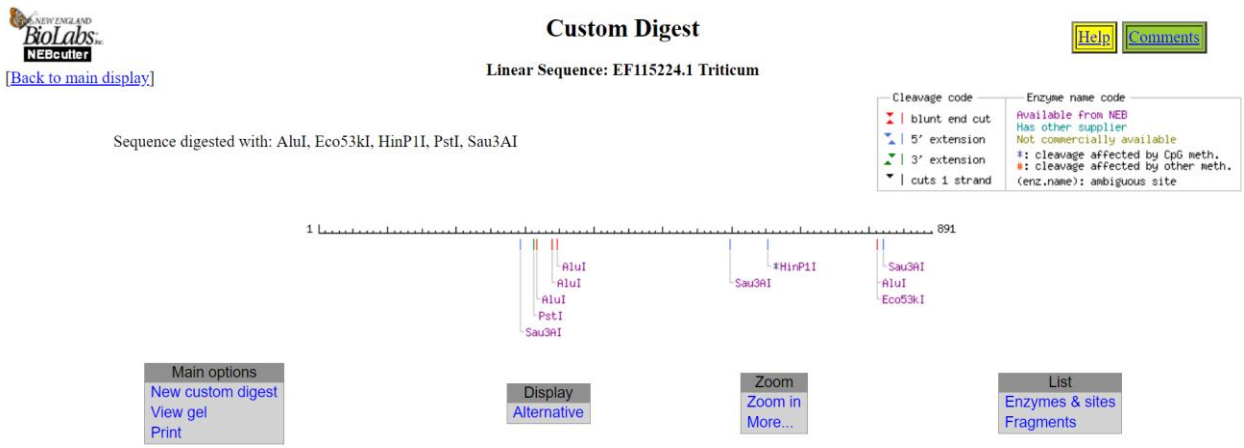
Help Comments

Pick your enzyme or enter a name in the box. Only the enzymes that cut the sequence and match your selection of availability (NEB enzymes) are shown. Non-NEB neoschizomers are not listed, but you can enter any enzyme name in the "Pick this enzyme" box and its prototype will be selected.

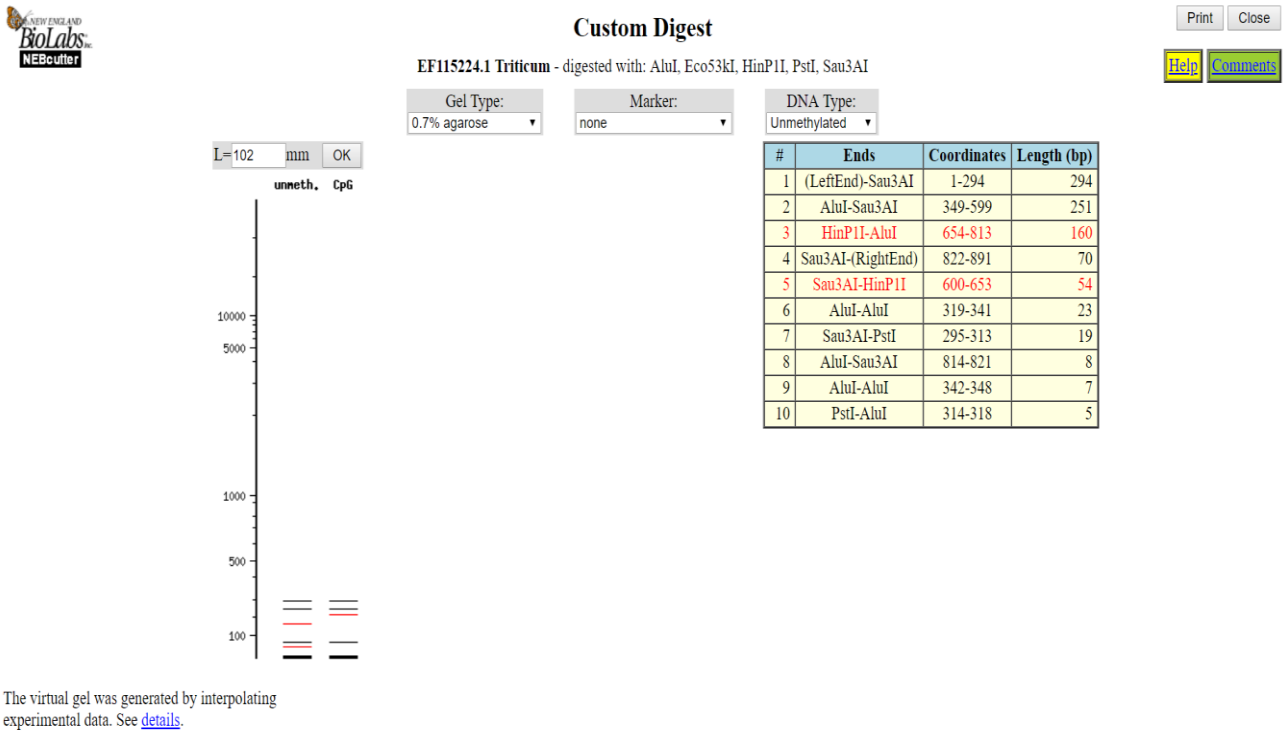
[**All enzymes**] - [Enzymes with compatible buffers] - [Enzymes producing blunt ends] - [Enzymes producing 5' overhangs] - [Enzymes producing 3' overhangs] - [Enzymes with a particular site length] - [Enzymes cutting N times] - [Non-palindromic enzymes]

Pick all	Enzyme	Specificity	Cuts	% activity in			
				1.1	2.1	3.1	CS
	AcuI	CTGAAG(N)$_{14}$▲NN▼	2	50	100	50	100
Picked	AluI	AG▼▲CT	4	25	100	50	100
	AlwI	GGATCNNNN▼N▲	1	50	50	10	100
	ApeKI	G▼CWG▲C	3	25	50	100	10
	ApoI	R▼AATT▲Y	2	10	75	100	75
	AseI	AT▼TA▲AT	1	10	50*	100	10
	BaeGI	G▲KGCM▼C	1	75	75	100	25
	BanI	G▼GYRC▲C	1	10	25	10	100
	BanII	G▲RGCY▼C	1	100	100	50	100
	BbsI	GAAGACNN▼NNNN▲	1	100	100	25	75
	BbvCI	CC▼TCA▲GC	1	10	100	50	100
	BbvI	GCAGC(N)$_{8}$▼NNNN▲	3	100	100	25	100
	BccI	CCATCNNNN▼N▲	1	100	50	10	100
	BclI	T▼GATC▲A	1	50	100	100	75
	BfuAI	ACCTGCNNNN▼NNNN▲	1	10	25	100	10
	BglII	A▼GATC▲T	1	10	10	100	10
	BmrI	ACTGGGNNNN▲N▼	1	75	100	75	100*
	Bpu10I	CC▼TNA▲GC	2	10	25	100	25
	BsiHKAI	G▲WGCW▼C	1	25	100	100	100
	BslI	CCNN▲NNN▼NNGG	1	50	75	100	100
	BsmFI	GGGAC(N)$_{10}$▼NNNN▲	1	25	50	50	100
	BsmI	GAATG▲CN▼	1	25	100	10	100
	Bsp1286I	G▲DGCH▼C	2	25	25	25	100
	BspCNI	CTCAG(N)$_{7}$▲NN▼	1	100	75	10	100
	BspHI	T▼CATG▲A	1	10	50	25	100
	BspMI	ACCTGCNNNN▼NNNN▲	1	10	50*	100	10
	BsrDI	GCAATG▲NN▼	3	10	100	75	25
	BsrI	ACTG▲GN▼	3	10	50	100	10
	BstAPI	GCAN▲NNN▼NTGC	2	50	100	25	100
	BstYI	R▼GATC▲Y	1	25	100	75	100
	BtsCI	GGATG▲NN▼	1	10	100	25	100
	BtsIMutI	CAGTG▲NN▼	2	100	50	10	100
	Cac8I	GCN▼▲NGC	2	50	75	100	100
	CviAII	C▼AT▲G	5	50	50	10	100
	CviKI-1	RG▼▲CY	7	25	100	100	100
	CviQI	G▼TA▲C	3	75	100*	100	75*
	DdeI	C▼TNA▲G	2	75	100	100	100
	DpnI	GA▼▲TC	3	100	100	75	100
	DpnII	▼GATC▲	3	25	25	100*	25
	DraI	TTT▼▲AAA	1	75	75	50	100
	EarI	CTCTTCN▼NNN▲	1	50	10	10	100
Picked	Eco53kI	GAG▼▲CTC	1	100	100	10	100

	FatI	ˇCATG▲	5	10	100	50	50
	Fnu4HI	GCˇN▲GC	3	10	10	10	100
	FokI	GGATG(N)$_9$ˇNNNN▲	1	100	100	75	100
	FspEI	CC(N)$_{12}$ˇNNNN▲	51	10	10	10	100
	HhaI	G▲CGˇC	1	25	100	100	100
Picked	HinP1I	GˇCG▲C	1	100	100	100	100
	HincII	GTYˇ▲RAC	2	25	100	100	100
	HinfI	GˇANT▲C	3	50	100	100	100
	HpaI	GTTˇ▲AAC	1	10	75*	25	100
	Hpy166II	GTNˇ▲NAC	2	100	100	50	100
	Hpy188I	TC▲NˇGA	7	25	100	50	100
	Hpy188III	TCˇNN▲GA	3	100	100	10	100
	Hpy99I	▲CGWCGˇ	1	50	10	10	100
	HpyAV	CCTTC(N)$_5$▲Nˇ	1	100	100	25	100
	HpyCH4III	AC▲NˇGT	6	100	25	10	100
	HpyCH4IV	AˇCG▲T	2	100	50	25	100
	HpyCH4V	TG▲ˇCA	14	50	50	25	100
	LpnPI	CCDG(N)$_{10}$ˇNNNN▲	7	10	10	10	100
	MboI	ˇGATC▲	3	75	100	100	100
	MboII	GAAGA(N)$_7$▲Nˇ	5	100*	100	50	100
	MluCI	ˇAATT▲	9	100	10	10	100
	MnlI	CCTC(N)$_6$▲Nˇ	5	75	100	50	100
	MseI	TˇTA▲A	7	75	100	75	100
	MslI	CAYNNˇ▲NNRTG	1	50	50	10	100
	MspJI	CNNR(N)$_9$ˇNNNN▲	132	10	10	10	100
	MwoI	GCNN▲NNNˇNNGC	3	10	100	100	100
	NlaIII	▲CATGˇ	5	10	10	10	100
	NlaIV	GGNˇ▲NCC	1	10	10	10	100
Picked	PstI	C▲TGCAˇG	1	75	75	100	50*
	RsaI	GTˇ▲AC	3	25	50	10	100
	SacI	G▲AGCTˇC	1	100	50	10	100
Picked	Sau3AI	ˇGATC▲	3	100	50	10	100
	SfcI	CˇTRYA▲G	2	75	50	25	100
	SspI	AATˇ▲ATT	1	50	100	50	50
	SwaI	ATTTˇ▲AAAT	1	10	10	100	10
	TaqI	TˇCG▲A	1	50	75	100	100
	TfiI	GˇAWT▲C	3	50	100	100	100
	TseI	GˇCWG▲C	3	75	100	100	100
	Tsp45I	ˇGTSAC▲	1	100	50	10	100
	TspRI	▲NNCASTGNNˇ	2	25	50	25	100
	XmnI	GAANNˇ▲NNTTC	1	50	75	10	100



Gel view:



#	Ends	Coordinates	Length (bp)
1	(LeftEnd)-Sau3AI	1-294	294
2	AluI-Sau3AI	349-599	251
3	HinP1I-AluI	654-813	160
4	Sau3AI-(RightEnd)	822-891	70
5	Sau3AI-HinP1I	600-653	54
6	AluI-AluI	319-341	23
7	Sau3AI-PstI	295-313	19
8	AluI-Sau3AI	814-821	8
9	AluI-AluI	342-348	7
10	PstI-AluI	314-318	5

Interpretation: DNA bands are found according to their length (bp). The large fragment size: 294bp and short fragment size: 5bp

No. of the Experiment: 04

Name of the experiment: Protein (Hemoglobin A) conserved domain determination.

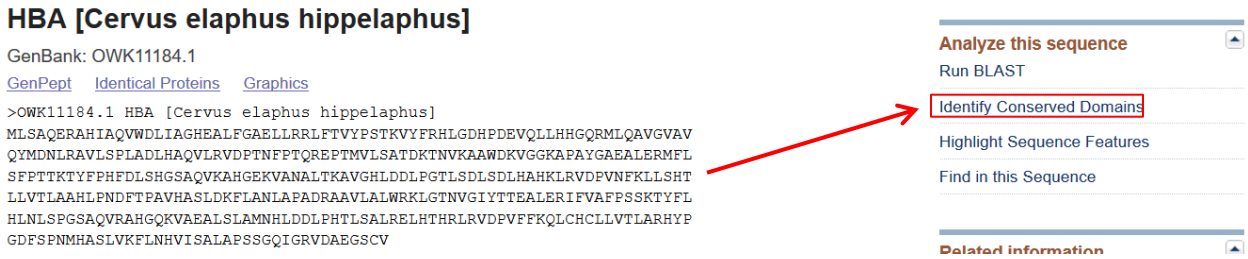
Principle: Conserved domain is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM).

Material:

1. Computer
2. Internet access

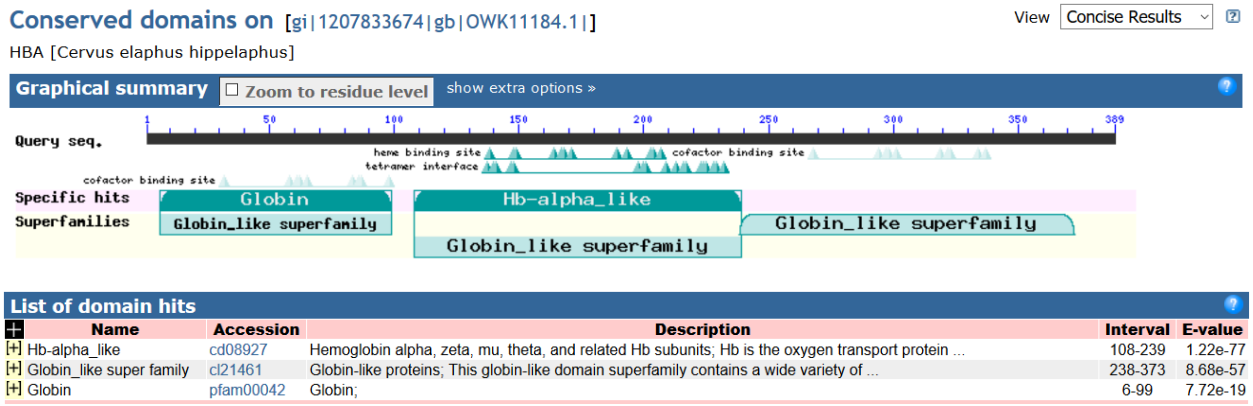
Procedure:

1. NCBI website is browsed and FASTA format of HBA [Cervus elaphus hippelaphus] is searched.

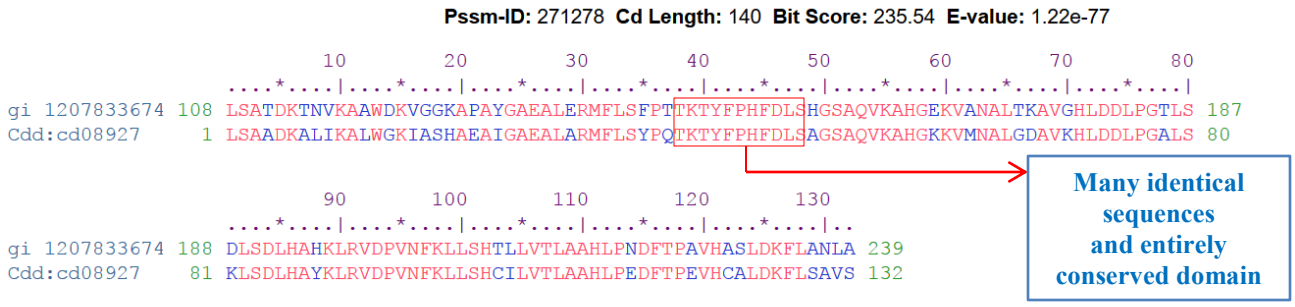


2. Then "identify conserved domains" is clicked.

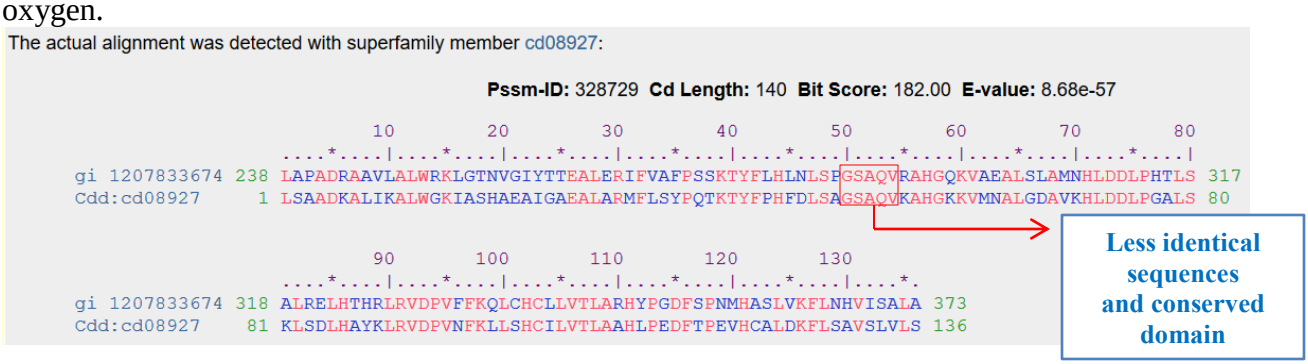
Result: The graphical summary and list of the domain hits are displayed.



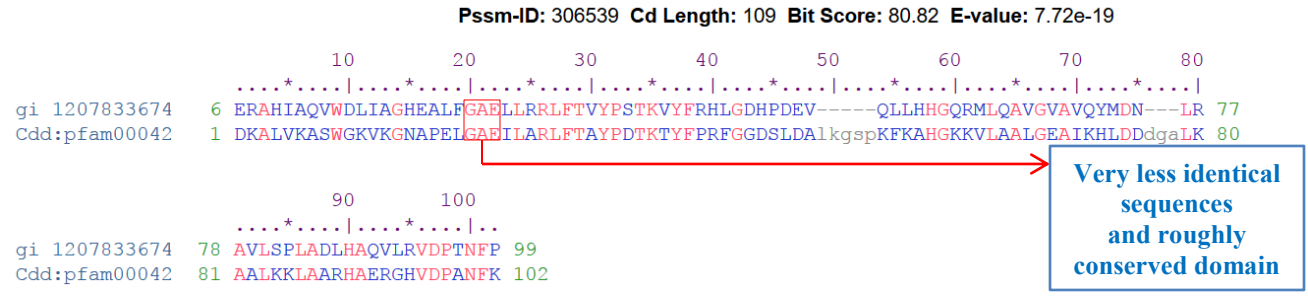
Hemoglobin alpha like: Hemoglobin alpha, zeta, mu, theta, and related Hb subunits; Hb is the oxygen transport protein of erythrocytes. It is an allosterically modulated heterotetramer. Hemoglobin A (HbA) is the most common Hb in adult humans, and is formed from two alpha-chains and two beta-chains (alpha2beta2).



Globin like superfamily: Globin-like proteins; This globin-like domain superfamily contains a wide variety of all-helical proteins that bind porphyrins, phycobilins, and other non-heme cofactors, and play various roles in all three kingdoms of life, including sensors or transporters of oxygen.



Globin:



Interpretation:

- The bit score of hemoglobin alpha subunit is 235.54, which is very higher and E-value is also less. So, In this alignment, the sequences are highly conserved.
- The bit score of Globin like superfamily is 182, which is also higher but less than that of hemoglobin alpha like subunits and E-value is also less. So, In this alignment, the sequences are also conserved but less than that of hemoglobin alpha like subunits.
- The bit score of globin is 80.82, which much less than others and E-value is also higher than others. So, In this alignment, the sequences are roughly conserved.

No. of the Experiment: 05

Name of the Experiment: Protein (P50) secondary structure prediction**.**

Principle: A secondary structure protein refers to the folding of a polypeptide chain due to hydrogen bonding, resulting in an alpha helix, beta sheet or a random coli structure.

Materials:

1. Computer
2. Internet access.

Procedure:

1. The FASTA format of protein sequence (*P50)* has been copied from NCBI or FASTA format is downloaded from NCBI for later uses.

FASTA

P50 [Babesia gibsoni]

GenBank: ACM48522.1

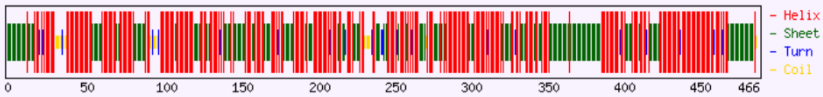
GenPept Identical Proteins Graphics

```
>ACM48522.1 P50 [Babesia gibsoni]
MNVVRSFLFFPIAFSLVRANGEGKTAEATPAGTSSPTEPKAAEAAPKAVDAAAVTFKQYLDFAMKLNEAV
TLREEDTRKKLLVHFPLFGAPPFDGAWGDLKDLLKKVTELRALLLKGHTFGLPAATTTDKQQQDANQTVG
ALFDFLVGVATDAVTVADKATRAVTGMDPDKAVGFHVTPATADALFEFVPDLYEKLKDLHSKVGEWVEIK
STFDDTKLVTQAGDHRPKHWLRQGGFTDQEVKGDTTLETLKTKLGELVGPTKPCEKVLCTLASYALMKTP
QDAAGKQAWLFLLASAMNNNAMKAKLEVAVNAVTPGKGETFVNQLKEVGKSLQLPKEQVPKQYRFPGVYA
NLDVQHFWTVLTGVFGTILTDLEVDEKDAQGKAGQVATRVAELVKVEGPLHSLTVQVAEMTKAGAGAGGE
APAQAAAGTAGARAEAPAKEGQGEDGAHFCGIGMTVFFVSVVIAVF
```

2. Then the copied FASTA format is pasted in http://www.biogem.org/tool/chou-fasman/ and submitted.
3. The results of secondary structure is accessed.
4. Then browsed https://swissmodel.expasy.org/interactive
5. The copied FASTA format is pasted again in the box and “Build model” is clicked.
6. After waited for a while the model is generated.

Target Sequence:

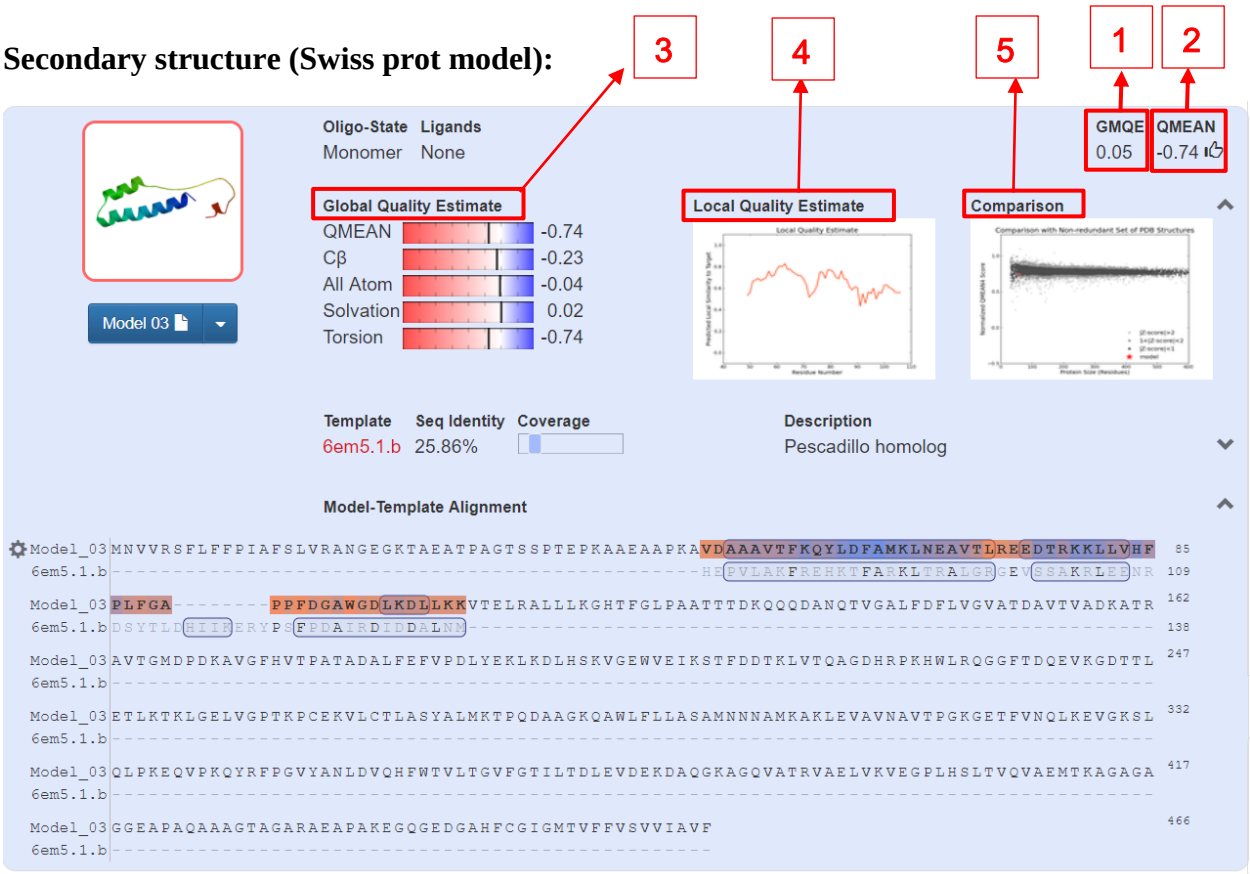
```
        10         20         30         40         50         60         70
MNVVRSFLFF PIAFSLVRAN GEGKTAEATP AGTSSPTEPK AAEAAPKAVD AAAVTFKQYL DFAMKLNEAV

        80         90        100        110        120        130        140
TLREEDTRKK LLVHFPLFGA PPFDGAWGDL KDLLKKVTEL RALLLKGHTF GLPAATTTDK QQQDANQTVG

       150        160        170        180        190        200        210
ALFDFLVGVA TDAVTVADKA TRAVTGMDPD KAVGFHVTPA TADALFEFVP DLYEKLKDLH SKVGEWVEIK

       220        230        240        250        260        270        280
STFDDTKLVT QAGDHRPKHW LRQGGFTDQE VKGDTTLETL KTKLGELVGP TKPCEKVLCT LASYALMKTP

       290        300        310        320        330        340        350
QDAAGKQAWL FLLASAMNNN AMKAKLEVAV NAVTPGKGET FVNQLKEVGK SLQLPKEQVP KQYRFPGVYA

       360        370        380        390        400        410        420
NLDVQHFWTV LTGVFGTILT DLEVDEKDAQ GKAGQVATRV AELVKVEGPL HSLTVQVAEM TKAGAGAGGE

       430        440        450        460
APAQAAAGTA GARAEAPAKE GQGEDGAHFC GIGMTVFFVS VVIAVF
```



Secondary Structure:

```
                    *         *         *         *         *         *
Query 1   MNVVRSFLFFPIAFSLVRANGEGKTAEATPAGTSSPTEPKAAEAAPKAVDAAAVTFKQYLDFAMKLNEAV 70
Helix 1     HHHHHHHHHHHHHHHHHHHHHHHHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 70
Sheet 1   EEEEEEEEEEEEEEEEE                                EEEEEEEEEEEEEEEEE 70
Turns 1                          T T            T   T                       T   70
Struc 1   EEEEEEEEEEEEHEEEEHHHTHTHHHHHHHCCCCTCCHHHHHHHHHHHHHHHHHEEEEEHHHHHHHHHEE 70

                    *         *         *         *         *         *
Query 71  TLREEDTRKKLLVHFPLFGAPPFDGAWGDLKDLLKKVTELRALLLKGHTFGLPAATTTDKQQQDANQTVG 140
Helix 71  HHHHHHHHHHHHHHHHHHH      HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 140
Sheet 71  E     EEEEEEEEEEE            EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 140
Turns 71     TT   T            T   T                    T            T  T   T    140
Struc 71  HHHHHHHHHEEEEEEEHHHCTCCCTHHHHHHHHHEHHHHHHHHHHEEEEHHHHEEHHHHHHTHHEEEEH 140

                    *         *         *         *         *         *
Query 141 ALFDFLVGVATDAVTVADKATRAVTGMDPDKAVGFHVTPATADALFEFVPDLYEKLKDLHSKVGEWVEIK 210
Helix 141 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 210
Sheet 141 EEEEEEEEEEEEEEE     EEEEEE      EEEEEEEEEEEEEEEEEEEEEE       EEEEEEEE 210
Turns 141                                TT                            T          210
Struc 141 HEEEEEEEHHHHEEHHHHHHEEEEHHHHTHHHEEEEEEHHHHHHHEHEEEHHHHHHHHHTHHHHEHHEE 210

                    *         *         *         *         *         *
Query 211 STFDDTKLVTQAGDHRPKHWLRQGGFTDQEVKGDTTLETLKTKLGELVGPTKPCEKVLCTLASYALMKTP 280
Helix 211 HHHHHHHHHHHH     HHHHHHHHHHHHHHHHHHHHHHHHHHHHH    HHHHHHHHHHHHHHHHHHH 280
Sheet 211 EEEEEEEEEE         EEEEEEEEEEEEEEEEEEEEEEEEEEEEEE   EEEEEEEEEEEEEEEEE 280
Turns 211 T    T           TT    TT   TT   T        T                           280
Struc 211 THHHHEEEEHHHCCCCTHHEEETTEEHHHHHHTEHHHHEHHTHHHHEEEECHHHHEEEEHEHEHHHHHHH 280

                    *         *         *         *         *         *
Query 281 QDAAGKQAWLFLLASAMNNNAMKAKLEVAVNAVTPGKGETFVNQLKEVGKSLQLPKEQVPKQYRFPGVYA 350
Helix 281 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH        H 350
Sheet 281       EEEEEE             EEEEEEEE      EEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 350
Turns 281 T    TT           T    T           T TT          T     TT   T          350
Struc 281 HHHHHHHEEEHHHHHHHTHHHHHHHHEHEEEEHHTHHHEEEHHHHHEHHEEHHHHHEEEEEEEEEEEEH 350

                    *         *         *         *         *         *
Query 351 NLDVQHFWTVLTGVFGTILTDLEVDEKDAQGKAGQVATRVAELVKVEGPLHSLTVQVAEMTKAGAGAGGE 420
Helix 351 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 420
Sheet 351 EEEEEEEEEEEEEEEEEEEEEE           EEEEEEEEEEE    EEEEEEEEEEE           420
Turns 351                          TT   T                T                 T  420
Struc 351 EEEEEEEEEEEEEEEEEEEHHHHHHHHHHTHHHEEEEHHHHEHHHHTHEEEEEEHHHHHHHHHHHHHTH 420
                    *         *         *         *
Query 421 APAQAAAGTAGARAEAPAKEGQGEDGAHFCGIGMTVFFVSVVIAVF 466
Helix 421 HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  466
Sheet 421                            EEEEEEEEEEEEEEE   466
Turns 421                    T    T                    466
Struc 421 HHHHHHHHHHHHHHHHHHHTHHHHTHHHEEEEEEEEEEEEEEEEHC 466


Total Residues: H: 433   E: 303   T: 46
       Percent: H: 92.9  E: 65.0  T: 9.9
```

Secondary structure (Swiss prot model):



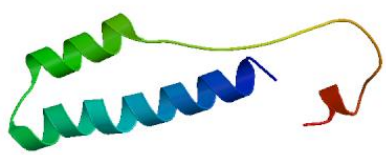


Fig: Secondary structure

Interpretation:

1. **GMQE**

GMQE (Global Model Quality Estimation) is a quality estimation which combines properties from the target–template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Higher numbers indicate higher reliability.

2. QMEAN

QMEAN is a composite estimator based on different geometrical properties and provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates on the basis of one single model.

The QMEAN Z-score ((2) in the figure above) provides an estimate of the 'degree of nativeness' of the structural features observed in the model on a global scale. It indicates whether the QMEAN score of the model is comparable to what one would expect from experimental structures of similar size. QMEAN Z-scores around zero indicate good agreement between the model structure and experimental structures of similar size. Scores of -4.0 or below are an indication of models with low quality. This is also highlighted by a change of the "thumbs-up" symbol to a "thumbs-down" symbol next to the score.

3. QMEAN consists of four individual terms.

The four individual terms of the global QMEAN quality scores are also listed ((3) in the figure above).

i. The white area in the bar plots (numerical values close to zero) indicates that the property is similar to what one would expect from experimental structures of similar size.
ii. Positive values indicate that the model scores higher than experimental structures on average.
iii. Negative numbers indicate that the model scores lower than experimental structures on average.
iv. The QMEAN Z-score itself is shown on top. The individual Z-scores compare the interaction potential between Cβ atoms only, all atoms, the solvation potential and the torsion angle potential

4. The "**Local Quality**" plot ((4) in the figure above) shows, for each residue of the model (reported on the x-axis), the expected similarity to the native structure (y-axis). Typically, residues showing a score below 0.6 are expected to be of low quality.

5. In the **"Comparison**" plot ((5) in the figure above), model quality scores of individual models are related to scores obtained for experimental structures of similar size. The x-axis shows protein length (number of residues). The y-axis is the normalized QMEAN score. Every dot represents one experimental protein structure. Black dots are experimental structures with a normalized QMEAN score within 1 standard devation of the mean (|Z-score| between 0 and 1), experimental structures with a |Z-score| between 1 and 2 are grey. Experimental structure that are even further from the mean are light grey. The actual model is represented as a red star. The mean and standard deviation of the experimental structures around the x-location of the star are the basis to calculate the QMEAN Z-score of the model ((2) in the figure above), i.e. how many standard deviations from the mean scores my model.

No. of the Experiment: 06

Name of the Experiment: Protein (P50) secondary structure predication by ramachandran plot.

Principle: A Ramachandran plot is a way to visualize energetically allowed regions for backbone dihedral angles ψ against φ of amino acid residues in protein structure. Peptide conformations are defined by the values of φ and ψ. Conformation deemed possible are those that involve little or no steric interference, based on calculation using known van der waals radii and dihedral angels.

Materials:

1. Computer
2. Internet access.

Procedure:

1. The FASTA format of protein sequence (*P50)* has been copied from NCBI or FASTA format is downloaded from NCBI for later uses.

FASTA

P50 [Babesia gibsoni]

GenBank: ACM48522.1

GenPept Identical Proteins Graphics

```
>ACM48522.1 P50 [Babesia gibsoni]
MNVVRSFLFFPIAFSLVRANGEGKTAEATPAGTSSPTEPKAAEAAPKAVDAAAVTFKQYLDFAMKLNEAV
TLREEDTRKKLLVHFPLFGAPPFDGAWGDLKDLLKKVTELRALLLKGHTFGLPAATTTDKQQQDANQTVG
ALFDFLVGVATDAVTVADKATRAVTGMDPDKAVGFHVTPATADALFEFVPDLYEKLKDLHSKVGEWVEIK
STFDDTKLVTQAGDHRPKHWLRQGGFTDQEVKGDTTLETLKTKLGELVGPTKPCEKVLCTLASYALMKTP
QDAAGKQAWLFLLASAMNNNAMKAKLEVAVNAVTPGKGETFVNQLKEVGKSLQLPKEQVPKQYRFPGVYA
NLDVQHFWTVLTGVFGTILTDLEVDEKDAQGKAGQVATRVAELVKVEGPLHSLTVQVAEMTKAGAGAGGE
APAQAAAGTAGARAEAPAKEGQGEDGAHFCGIGMTVFFVSVVIAVF
```

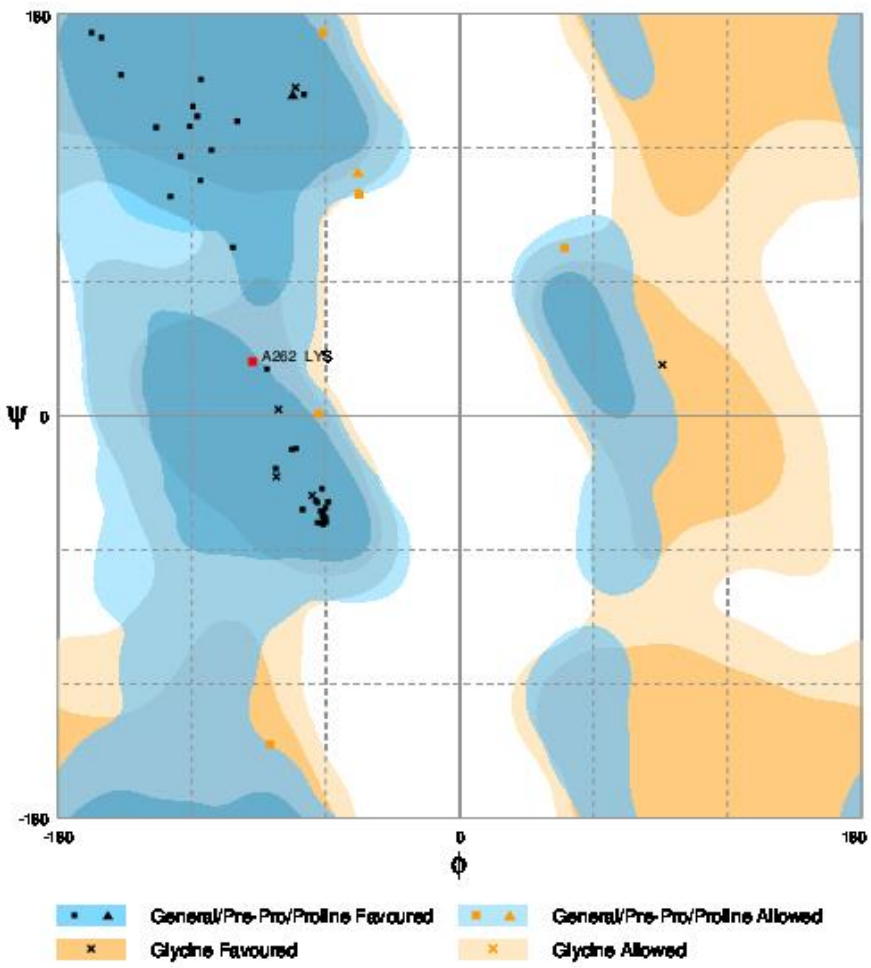
2. Then the copied FASTA format is pasted in **https://swissmodel.expasy.org/interactive** and "build model" is clicked.
3. After waited for a while the model is generated.
4. Then the model is downloaded as PDB format.

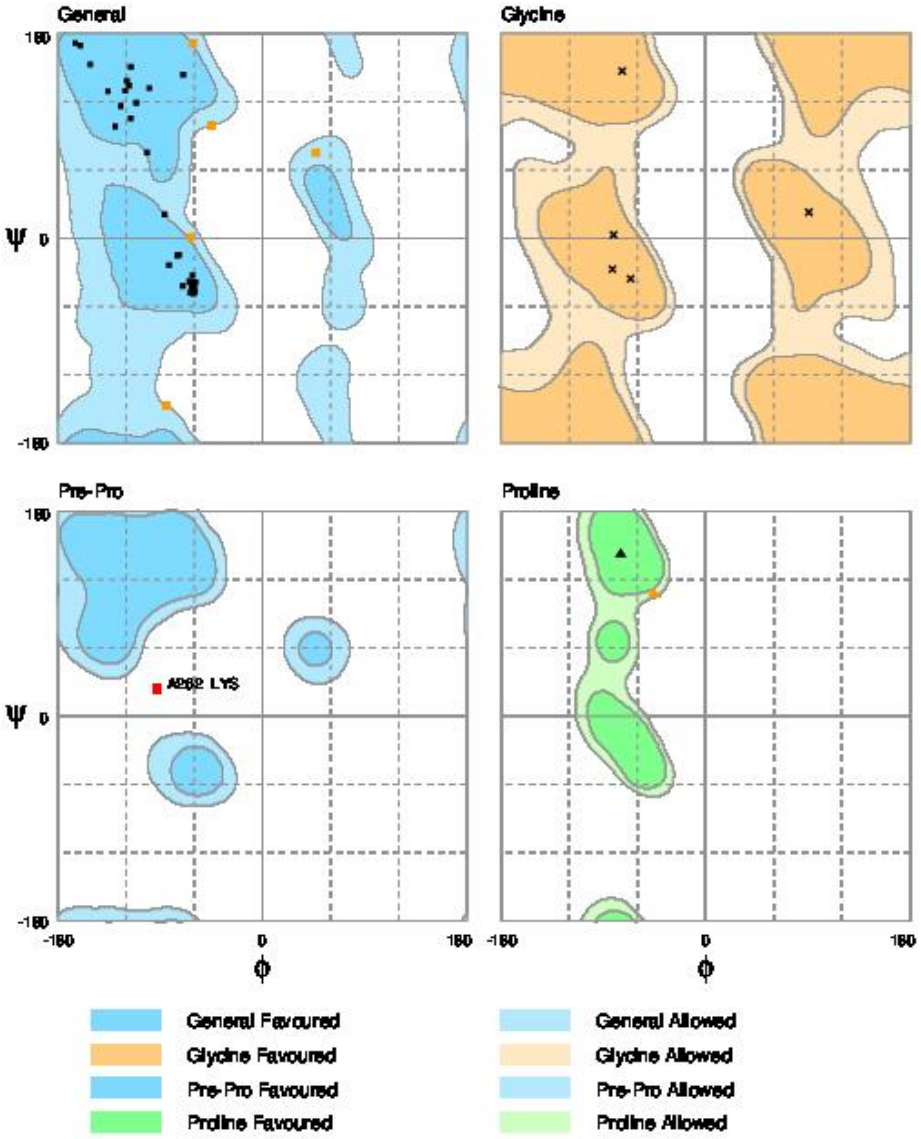
5. Browsed the webpage **http://mordred.bioc.cam.ac.uk/~rapper/rampage.php**
6. The PDB file is uploaded and submitted to rampage.

Result: The colored graphical display shows the favored and allowed regions of different amino acids.

RAMPAGE | Information MolProbity | Crystallography and Bioinformatics Group

RAMPAGE: Assessment of the Ramachandran Plot





Evaluation of residues:

```
Residue [A 238 :ASP] ( -63.14,   0.76) in Allowed region
Residue [A 239 :GLN] ( -44.99,  99.64) in Allowed region
Residue [A 244 :ASP] (  47.11,  75.17) in Allowed region
Residue [A 246 :THR] ( -61.22, 171.50) in Allowed region
Residue [A 260 :PRO] ( -45.55, 109.22) in Allowed region
Residue [A 264 :CYS] ( -84.76,-147.02) in Allowed region
Residue [A 262 :LYS] ( -92.65,  24.31) in Outlier region
Number of residues in favored region     (~98.0% expected):   41 (85.4%)
Number of residues in allowed region     (~02.0% expected):    6 (12.5%)
Number of residues in outlier region                      :    1 (02.1%)
```

Interpretation of Ramachandran plot:

1. **Favored regions:** are rich of amino acids. Glycine, pre-pro, proline are rich in this regions. The yellow region is glycine rich region, for proline residue. The areas shaded dark blue represent conformations that involve no steric overlap and thus are fully allowed and are pre-pro rich region and green region is proline rich region.

2. **Allowed regions:** are just allowed for some amino acids. Glycine, pre-pro, proline is allowed in this regions. The light yellow region is disallowed for all amino acids except glycine residue, The medium blue indicates conformation at the extreme limits for unfavorable atomic contacts which allowed for pre-pro and light green region is allowed for Proline residue. Allowed ranges for branched residues such as Val, Ile and Thr. The Gly residue, which is less sterically hindered, has a much broader range of allowed conformations.

No. of the Experiment: 07

Name of the Experiment: Protein (P50) secondary structure prediction by hydropathy plot

Principle: Hydropathy plots allow for the visualization of hydrophobicity over the length of a peptide sequence. A hydropathy scale which is based on the hydrophobic and hydrophilic properties of the 20 amino acids is used. A moving "window" determines the summed hydropathy at each point in the sequence (Y coordinate). These sums are then plotted against their respective positions (X coordinate). Such plots are useful in determining the hydrophobic interior portions of globular proteins as well as determining membrane spanning regions of membrane bound proteins.

Materials:

1. Computer
2. Internet access.

Procedure:

1. The FASTA format of protein sequence (*P50)* has been copied from NCBI or FASTA format is downloaded from NCBI for later uses.

G-protein [Nephila clavipes]

GenBank: PRD34794.1

GenPept Identical Proteins Graphics

```
>PRD34794.1 G-protein [Nephila clavipes]
MPVADKGFRRGAYRCYCKPNFYLPSGEVHRRNKDYFDGELIERAYVESILHGEDKFLSNFQCLPCRVGCD
MCVGDQPCYFEYGLLLRSLVLCFQSFCSTLVIILIVVVFVLRKKRIFSSSMWVMLETILIGALFLYQTVI
IRYFEPSVVTCMLEPWFRGLGFAICYGAITLKIYRILAEFQTRKAHRVCVRDKDLLKYLMGIITVVIGYL
LAWTAVIIDNMNEGYSVVSHEQTPMGLKYVICRALWWDYVTEVGELLFLLFGVFLTYKIRNARSDCFHEK
ETLAWSIYVEALVSLLMYTIRHVVWSNLQPDYIYLLYFLRCQCTVTIILLFIFAPKLWYHARPPEIMDYQ
SGRQLSSCGDVRDHVPEAMKLHEAITSNGEVDFSDINLGDMDPEDIRVSKSDILLTYR
```

2. Then the copied FASTA format is pasted in **https://web.expasy.org/protscale/** and submitted.
3. After waited for a while the result is accessed.

Protscale result:

User-provided sequence:

```
        10         20         30         40         50         60
GPROTEINNE PHILACLAVI PESGENBANK PRDGENPEPT IDENTICALP ROTEINSGRA

        70         80         90        100        110        120
PHICSPRDGP ROTEINNEPH ILACLAVIPE SMPVADKGFR RGAYRCYCKP NFYLPSGEVH

       130        140        150        160        170        180
RRNKDYFDGE LIERAYVESI LHGEDKFLSN FQCLPCRVGC DMCVGDQPCY FEYGLLLRSL

       190        200        210        220        230        240
VLCFQSFCST LVIILIVVVF VLRKKRIFSS SMWVMLETIL IGALFLYQTV IIRYFEPSVV

       250        260        270        280        290        300
TCMLEPWFRG LGFAICYGAI TLKIYRILAE FQTRKAHRVC VRDKDLLKYL MGIITVVIGY

       310        320        330        340        350        360
LLAWTAVIID NMNEGYSVVS HEQTPMGLKY VICRALWWDY VTEVGELLFL LFGVFLTYKI

       370        380        390        400        410        420
RNARSDCFHE KETLAWSIYV EALVSLLMYT IRHVVWSNLQ PDYIYLLYFL RCQCTVTIIL

       430        440        450        460        470        480
LFIFAPKLWY HARPPEIMDY QSGRQLSSCG DVRDHVPEAM KLHEAITSNG EVDFSDINLG

       490
DMDPEDIRVS KSDILLTYR
```

SEQUENCE LENGTH: 499

Using the scale **Hphob. / Kyte & Doolittle**, the individual values for the 20 amino acids are:

```
Ala:  1.800  Arg: -4.500  Asn: -3.500  Asp: -3.500  Cys:  2.500  Gln: -3.500
Glu: -3.500  Gly: -0.400  His: -3.200  Ile:  4.500  Leu:  3.800  Lys: -3.900
Met:  1.900  Phe:  2.800  Pro: -1.600  Ser: -0.800  Thr: -0.700  Trp: -0.900
Tyr: -1.300  Val:  4.200  : -3.500  : -3.500  : -0.490
```

Weights for window positions 1,..,9, using linear weight variation model:

```
   1     2     3     4     5     6     7     8     9
 1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
 edge                    center                   edge
```

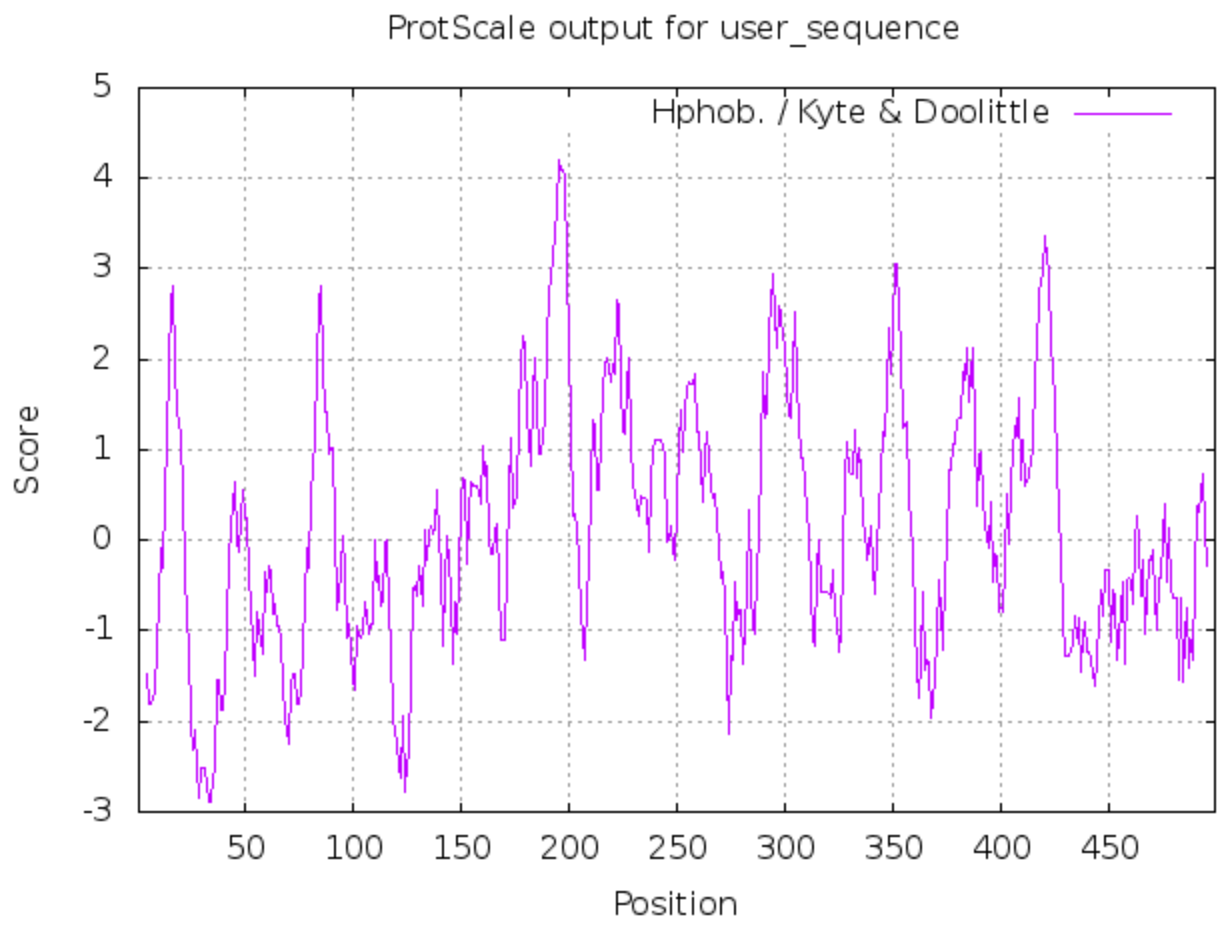


Fig: Hydropathy plot

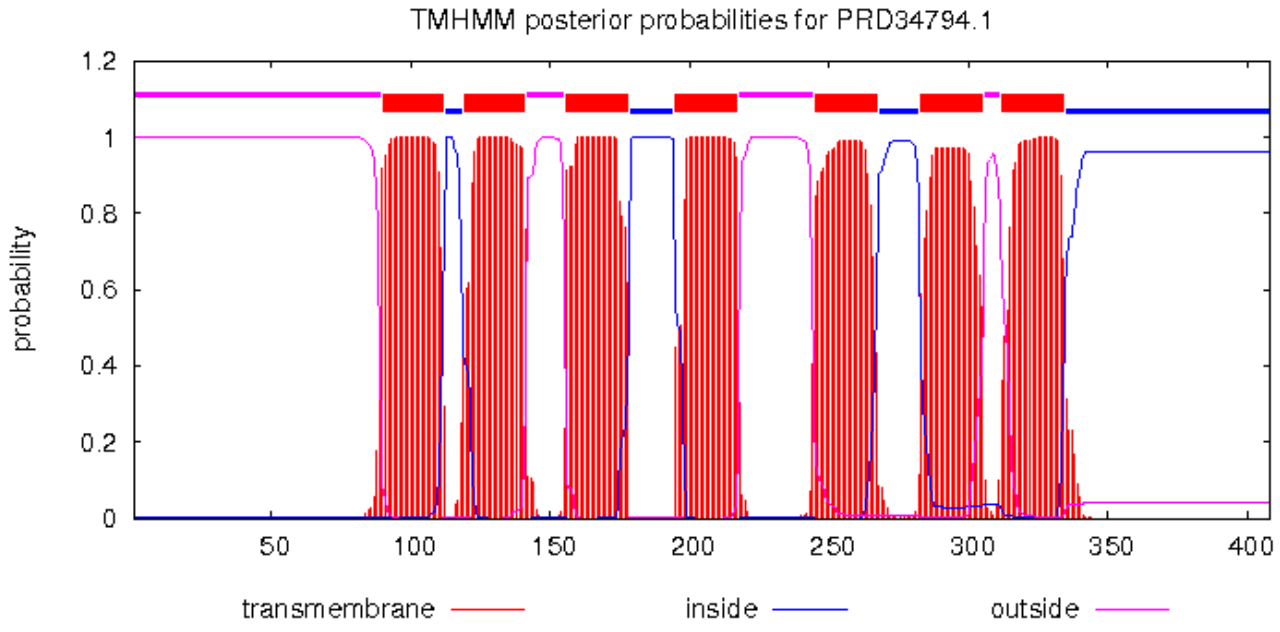
Interpretation of Hydropathy plot:

1. The position of residues is plotted on the x-axis and the hydropathy is plotted on the y-axis. The x-axis cannot be resized and is set to display 450 residues. When analysis plots of proteins with more than 450 residues are calculated, the axis will rescale automatically to display the entire sequence.

TMHMM result:

HELP with output formats

```
# PRD34794.1 Length: 408
# PRD34794.1 Number of predicted TMHs:  7
# PRD34794.1 Exp number of AAs in TMHs: 152.21654
# PRD34794.1 Exp number, first 60 AAs:  0.00062
# PRD34794.1 Total prob of N-in:        0.00061
PRD34794.1      TMHMM2.0        outside      1    89
PRD34794.1      TMHMM2.0        TMhelix     90   112
PRD34794.1      TMHMM2.0        inside     113   118
PRD34794.1      TMHMM2.0        TMhelix    119   141
PRD34794.1      TMHMM2.0        outside    142   155
PRD34794.1      TMHMM2.0        TMhelix    156   178
PRD34794.1      TMHMM2.0        inside     179   194
PRD34794.1      TMHMM2.0        TMhelix    195   217
PRD34794.1      TMHMM2.0        outside    218   244
PRD34794.1      TMHMM2.0        TMhelix    245   267
PRD34794.1      TMHMM2.0        inside     268   282
PRD34794.1      TMHMM2.0        TMhelix    283   305
PRD34794.1      TMHMM2.0        outside    306   311
PRD34794.1      TMHMM2.0        TMhelix    312   334
PRD34794.1      TMHMM2.0        inside     335   408
```



Interpretation: G-protein [Nephila clavipes] is a transmembrane protein **b**ecause most of the domain of this protein are anchored in the membrane. Some portions of this proteins are inside and some portions are outside of the membrane.

No. of the Experiment: 08

Name of the Experiment: Using multiple sequence alignment to construct phylogenetic trees.

Principle: A multiple sequence alignment (MSA) is a sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. In many cases, the input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor. From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins. Visual depictions of the alignment as in the image at right illustrate mutation events such as point mutations (single amino acid or nucleotide changes) that appear as differing characters in a single alignment column, and insertion or deletion mutations (gaps) that appear as hyphens in one or more of the sequences in the alignment. Multiple sequence alignment is often used to assess sequence conservation of protein domains, tertiary and secondary structures, and even individual amino acids or nucleotides.

Materials:

1. Computer
2. Internet access

Procedure:

1. Total (11) Protein sequences (**Accession numbers:** 2GS0_A, 2LOX_A, 2K2U_A, AJV11978.1, AJU65094.1, AJU90190.1, AJU72136.1, NP_010597.3, AAU09707.1, SCU98141.1, SCV02814.1) has been downloaded from NCBI as FASTA format.
2. Then browsed the webpage EMBL
3. Clustal omega is selected and FASTA format of downloaded file is uploaded.
4. After a while the result is generated.
5. Then browsed BLAST homepage https://blast.ncbi.nlm.nih.gov/Blast.cgi and 'Protein BLAST' (blastp) is clicked.
6. The downloaded file is then uploaded and the button "BLAST" was clicked.
7. Then "Multiple alignments" is selected.
8. After getting the alignments, phylogenetic tree is selected.

Result: We found color indicated alignment of multiple sequences and then branch shaped phylogenetic tree.

Results for job clustalo-I20180405-055857-0055-29401969-p1m

Alignments | Result Summary | Phylogenetic Tree | Submission Details

Download Alignment File | Hide Colors | Send to Simple Phylogeny | Send to MView

```
CLUSTAL O(1.2.4) multiple sequence alignment


SCV02814.1       MSHSGAAVFKKVSGILAINEEVSPVEIVWRSTDGDKTHKIALDTIDKLQATPNTSEKMML   60
SCU98141.1       MSHSGAAIFKKVSGILNVDEDASPANLLWRSTDGDKSHKVVLSMIDKLQATPASSEKMML   60
AAU09707.1       MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
AJU72136.1       MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
AJU90190.1       MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
AJU65094.1       MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
AJV11978.1       MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
NP_010597.3      MSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
2GS0_A           XSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
2LOX_A           PSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
2K2U_A           PSHSGAAIFEKVSGIIAINEDVSPAELTWRSTDGDKVHTVVLSTIDKLQATPASSEKMML   60
                  ******:*:*****: ::*:.**.:: ******** *.:.* ******** :******
                                                            Period

SCV02814.1         RLVGKSETTKKVKDNESNDVKPAPPSHVFSFNNRTVMDNIKETLQTIIARYKDEEIYEEK   120
SCU98141.1         RLIEKMDESKKRKDSEGNEIKPMVPSHVFSFNNRTVMDNIKETLQLIIARYKDQEAYEEK   120
AAU09707.1         RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
AJU72136.1         RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
AJU90190.1         RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
AJU65094.1         RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
AJV11978.1         RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
NP_010597.3        RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADIYEEK   120
2GS0_A             RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDAD-----   115
2LOX_A             RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDADGNSS-   119
2K2U_A             RLIGKVDESKKRKDNEGNEVVPKPQRHMFSFNNRTVMDNIKMTLQQIISRYKDAD-----   115
                   **: * : :** **.*.*:: *    *:************* *** **:****
                                                                  Colon
                                                                  Star

SCV02814.1         KRKESSMG----------------------PALSETAPLINTSTLDDSLSREKLLTNLKL   158
SCU98141.1         RRKESSLA----------------------PSPADTAPLINTLTLDDSLSREKLLTNLKL   158
AAU09707.1         RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDDSLSKEKLLTNLKL   180
AJU72136.1         RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDDSLSKEKLLTNLKL   180
AJU90190.1         RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDNSLSKEKLLTNLKL   180
AJU65094.1         RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDDSLSKEKLLTNLKL   180
AJV11978.1         RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDDSLSKEKLLTNLKL   180
NP_010597.3        RRREESAQHTETPMSSSSVTAGTPTPHLDTPQLNNGAPLINTAKLDDSLSKEKLLTNLKL   180
2GS0_A             ------------------------------------------------------------   115
2LOX_A             ------------------------------------------------------------   119
2K2U_A             ------------------------------------------------------------   115


SCV02814.1         QQSLLRENKKLMKVFQETVINSGLPPHEFWSTRVPLLRAFALTTSQRTGPYNVLSTIKPV   218
SCU98141.1         QQSLLKDNKSLMKTFQDAVIYSGLPSDEFWSTRVPLLRAFALTTSQRTGPYNVLSTIKPV   218
AAU09707.1         QQSLLKGNKVLMKVFQETVINAGLPPSELWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
AJU72136.1         QQSLLKGNKVLMKVFQETVINAGLPPSEFWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
AJU90190.1         QQSLLKGNKVLMKVFQETVINAGLPPSEFWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
AJU65094.1         QQSLLKGNKVLMKVFQETVINAGLPPSEFWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
AJV11978.1         QQSLLKGNKVLMKVFQETVINAGLPPSEFWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
NP_010597.3        QQSLLKGNKVLMKVFQETVINAGLPPSEFWSTRIPLLRAFALSTSQKVGPYNVLSTIKPV   240
2GS0_A             ------------------------------------------------------------   115
2LOX_A             ------------------------------------------------------------   119
2K2U_A             ------------------------------------------------------------   115
```

```
SCV02814.1       ASSDNKVNVSVSREKILTIFQTYPIVKKAYDDNVPKNFREQEFWARFFSSKLFRKLRGER   278
SCU98141.1       ASSDNKVNVSVSREKILTIFQTYPIVKKAYDDSVPKNFKEQEFWARFFSSKLFRKLRGER   278
AAU09707.1       ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
AJU72136.1       ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
AJU90190.1       ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
AJU65094.1       ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
AJV11978.1       ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
NP_010597.3      ASSENKVNVNLSREKILNIFENYPIVKKAYTDNVPKNFKEPEFWARFFSSKLFRKLRGEK   300
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115

SCV02814.1       IMQNDRGDMIIDRYLTLDQEFDRKDDELLLHPVKKLIDLEGNIQDDPQRRGNRPDFTMIP   338
SCU98141.1       IMQSDRGDMIIDRYLTLDQEFDRKDDTRLQHPVKKLLDLEGNDQDDPARKGNRPDFTMMA   338
AAU09707.1       IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
AJU72136.1       IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
AJU90190.1       IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
AJU65094.1       IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
AJV11978.1       IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
NP_010597.3      IMQNDRGDVIIDRYLTLDQEFDRKDDDMLLHPVKKIIDLDGNIQDDPVVRGNRPDFTMQP   360
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115

SCV02814.1       GIDENGSSDGVIDILKGMNRLSEKMIKSLENEYSRSHLSA--------DDLDREERDELA   390
SCU98141.1       GVDPNGNSDGVVDILKGMNRLSQKMIQSLENEYSRASLLS--------EDVDKQEREELL   390
AAU09707.1       GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNITNDEEDEDNDERNELK   420
AJU72136.1       GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNTTNDEEDEDNDERNELK   420
AJU90190.1       GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNITNDEEDEDNDERNELK   420
AJU65094.1       GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNITNDEEDEDNDERNELK   420
AJV11978.1       GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNITNDEEDEDNDERNELK   420
NP_010597.3      GVDINGNSDGTVDILKGMNRLSEKMIMALKNEYSRTNLQNKSNITNDEEDEDNDERNELK   420
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115

SCV02814.1       YSDLEDDEVASYTELHLKRRIMIERKTNLAQNVD---SNKANGKVNDEDVKKGIDTLVNG   447
SCU98141.1       FTDLAEAQEANYLEIKVKQRMASSKTLPVSVN------GDAKSVVGPQAAQHQVDLLLEG   444
AAU09707.1       IDDLNESYKTNYAIIHLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
AJU72136.1       IDDLNESYKTNYAIIHLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
AJU90190.1       IDDLNESYKTNYAIIHLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
AJU65094.1       IDDLNESYKTNYAIIQLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
AJV11978.1       IDDLNESYKTNYAIIHLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
NP_010597.3      IDDLNESYKTNYAIIHLKRNAHEKTT-DNDAKSSADSIKNADLKVSNQQMLQQLSLVMDN   479
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115

SCV02814.1       LMNHIELSKVTTDG-NANSVANSKIVKAVKTNAKQVRNTNESVFIGGHSGTKVLDNSEEP   506
SCU98141.1       LTKSIDLTQVTNDDKKFNNTVNQNVIRAVKINAKQAKNTDLSAVVGATVGLSSTESIREV   504
AAU09707.1       LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
AJU72136.1       LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
AJU90190.1       LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
AJU65094.1       LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
AJV11978.1       LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
NP_010597.3      LINKLDLNQVVPN-NEVSNKINKRVITAIKINAKQAKHNNVNSALGSFVDNTSQANELEV   538
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115

SCV02814.1       KTPLPPDLLESCRALHGTCCEFVKHFYIHFQSGEPRQAGTVRKLYNYLIKCTEKISDLLS   566
SCU98141.1       NSPLPPELLETSRILHGTCCEFLKHFYILFQSGEIRQAPTVRKLYIHLKKCLDRITKLLD   564
AAU09707.1       KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
AJU72136.1       KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
AJU90190.1       KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
AJU65094.1       KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
AJV11978.1       KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
NP_010597.3      KSTLPIDLLESCRMLHTTCCEFLKHFYIHFQSGEQKQASTVKKLYNHLKDCIEKLNELFQ   598
2GS0_A           ------------------------------------------------------------   115
2LOX_A           ------------------------------------------------------------   119
2K2U_A           ------------------------------------------------------------   115
```

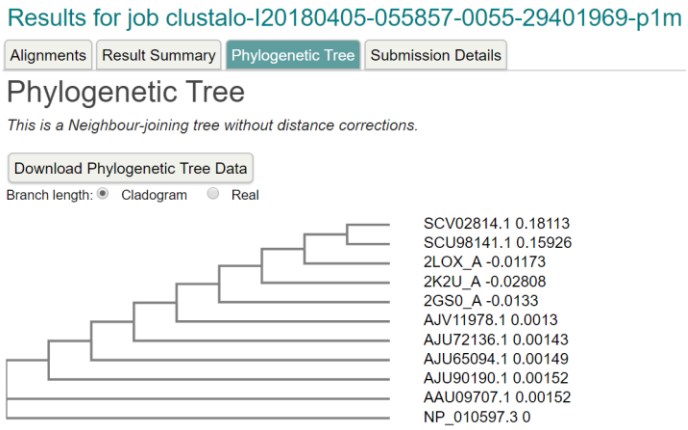
```
SCV02814.1      KVETEDGGKHDALVSTCDALLKPVLGSLHLAKQKYENALLEAGSPAVIDVR    617
SCU98141.1      EVGNDENAERDGLVNNCRAYLKSALDSLNFAIEKYETAVSDAESFRARKAA    615
AAU09707.1      DVLNG---DGESMSNTCTAYLKPVLNSITLATHKYDEYFNEYNNNSN----    642
AJU72136.1      DVLNG---DGESMSNTCTAYLKPVLNSITLATHKYDEYFNEYNNNSN----    642
AJU90190.1      DVLNG---DGESMSNTCTAYLKPVLNSITLATHKYDEYFNEYNNNSN----    642
AJU65094.1      DVLNG---DGESMSNTCTAYLKPVLNSITLATHKYDEYFNEYNNNSN----    642
AJV11978.1      DVLNG---DGESMSNTCTAYLKPVLNSITLATQKYDEYFNEYNNNSN----    642
NP_010597.3     DVLNG---DGESMSNTCTAYLKPVLNSITLATHKYDEYFNEYNNNSN----    642
2GS0_A          --------------------------------------------------    115
2LOX_A          --------------------------------------------------    119
2K2U_A          --------------------------------------------------    115
```

Interpretation of multiple sequence alignment: 20 amino acids have been used to construct multiple sequence alignments for 11 sequences. Gap (-) is present in many sequences.

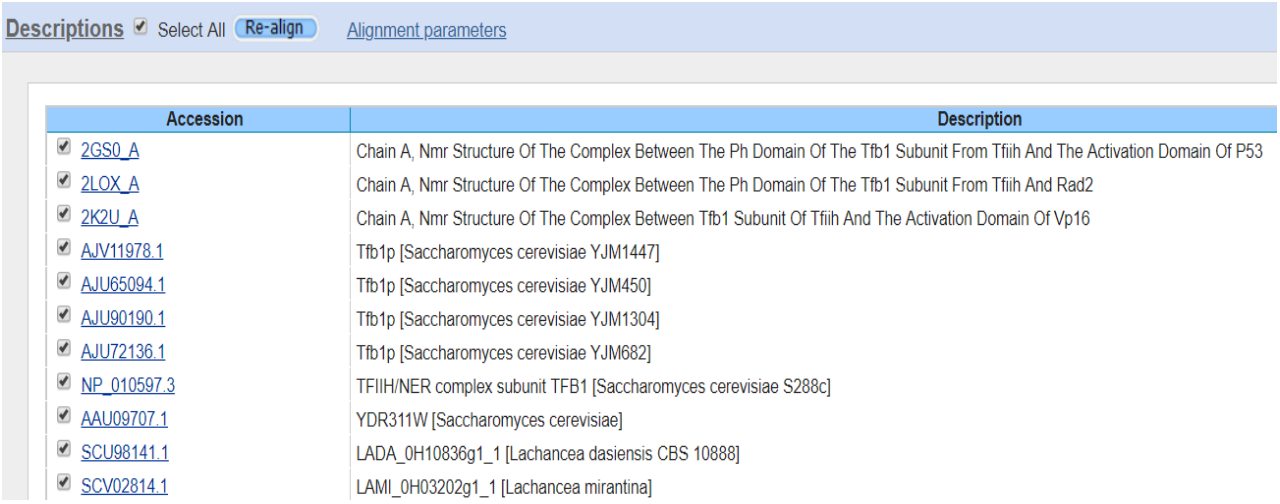
Star (*) - Star sign means those amino acids among protein sequences are identical and entirely conserved. In the first alignment, the 2nd to 7th column of amino acids are identical and entirely conserved

Colon (:) – Colon sign means, have variation among amino acids but include in same group, have same size, same hydrophobicity and those are roughly conserved.

Period (.) – Period sign means those amino acids are evolutionary similar and hydrophobicity of those amino acids has been reserved due to evolutionary linkage.



After BLAST:

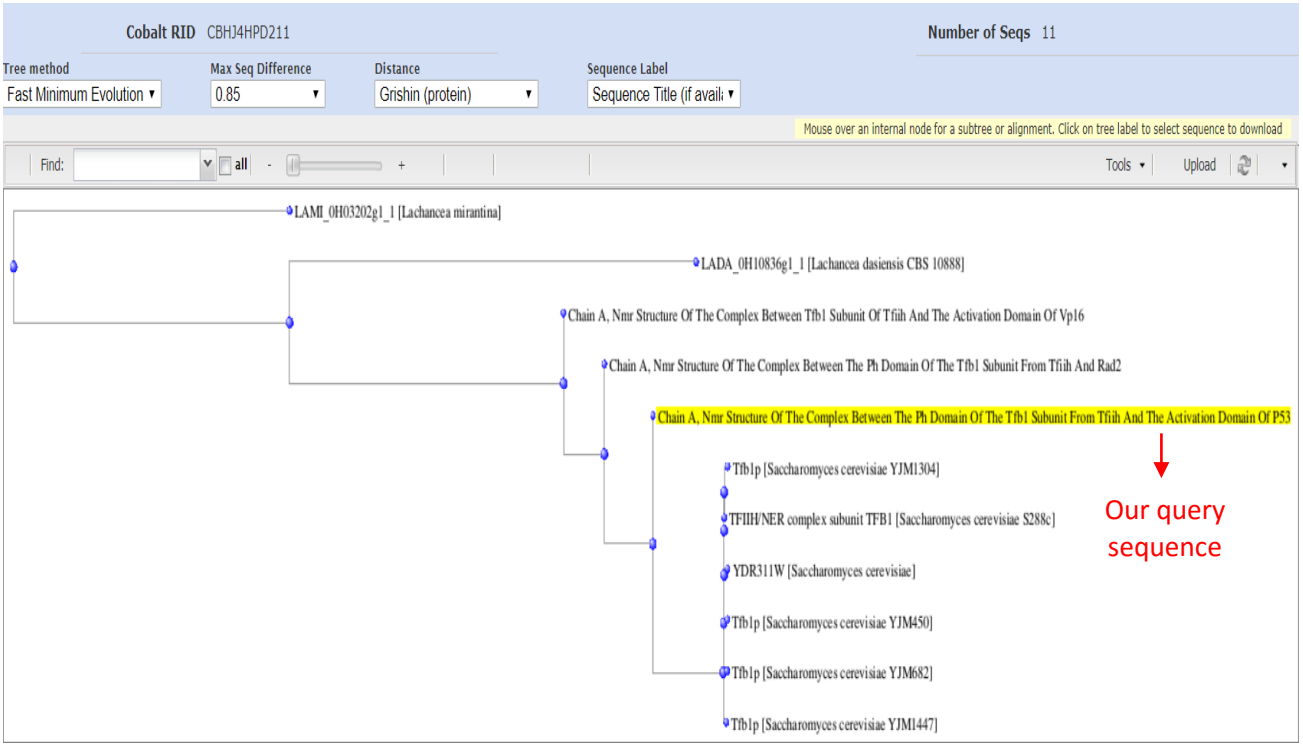


Accession	Description
2GS0_A	Chain A, Nmr Structure Of The Complex Between The Ph Domain Of The Tfb1 Subunit From Tfiih And The Activation Domain Of P53
2LOX_A	Chain A, Nmr Structure Of The Complex Between The Ph Domain Of The Tfb1 Subunit From Tfiih And Rad2
2K2U_A	Chain A, Nmr Structure Of The Complex Between Tfb1 Subunit Of Tfiih And The Activation Domain Of Vp16
AJV11978.1	Tfb1p [Saccharomyces cerevisiae YJM1447]
AJU65094.1	Tfb1p [Saccharomyces cerevisiae YJM450]
AJU90190.1	Tfb1p [Saccharomyces cerevisiae YJM1304]
AJU72136.1	Tfb1p [Saccharomyces cerevisiae YJM682]
NP_010597.3	TFIIH/NER complex subunit TFB1 [Saccharomyces cerevisiae S288c]
AAU09707.1	YDR311W [Saccharomyces cerevisiae]
SCU98141.1	LADA_0H10836g1_1 [Lachancea dasiensis CBS 10888]
SCV02814.1	LAMI_0H03202g1_1 [Lachancea mirantina]

Our selected query sequences:

Description	Max score	Total score	Query cover	E value	Ident	Accession
Chain A, Nmr Structure Of The Complex Between The Ph Domain Of The Tfb1 Subunit From Tfiih And Th	236	236	99%	9e-79	100%	2GS0_A
Chain A, Nmr Structure Of The Complex Between The Ph Domain Of The Tfb1 Subunit From Tfiih And Ra	236	236	99%	1e-78	100%	2LOX_A
Chain A, Nmr Structure Of The Complex Between Tfb1 Subunit Of Tfiih And The Activation Domain Of Vp	235	235	99%	1e-78	100%	2K2U_A
Tfb1p [Saccharomyces cerevisiae YJM1447]	238	238	99%	4e-73	100%	AJV11978.1
Tfb1p [Saccharomyces cerevisiae YJM450]	238	238	99%	4e-73	100%	AJU65094.1
Tfb1p [Saccharomyces cerevisiae YJM1304]	238	238	99%	5e-73	100%	AJU90190.1
Tfb1p [Saccharomyces cerevisiae YJM682]	238	238	99%	5e-73	100%	AJU72136.1
TFIIH/NER complex subunit TFB1 [Saccharomyces cerevisiae S288c]	237	237	99%	5e-73	100%	NP_010597.3
YDR311W [Saccharomyces cerevisiae]	237	237	99%	5e-73	100%	AAU09707.1
Tfb1p [Saccharomyces cerevisiae YJM1273]	237	237	99%	6e-73	100%	AJU89480.1
Tfb1p [Saccharomyces cerevisiae YJM456]	237	237	99%	8e-73	99%	AJU67210.1
LADA_0H10836g1_1 [Lachancea dasiensis CBS 10888]	181	181	99%	6e-52	77%	SCU98141.1
LAMI_0H03202g1_1 [Lachancea mirantina]	176	176	99%	5e-50	72%	SCV02814.1

Phylogenetic tree:



Interpretation of phylogenetic tree: The distance of phylogenetic tree can be determined by max score and E-value. Here we can see, the same max score sequences are comes from the same node or one common node. The differences between their E-values separate them from each other. All **Tfb1p** comes from one common node as their max score (237) is same but distance are differ due to their different E-values. All the chain A and Tfb1p have max score 235-336, 237 respectively, so their distance from each other is less. Lachancea dasiensis and Lachancea mirantina have max score 181 (Identity 77% with chain A) and 176 (Identity 72% with chain A) respectively. For this reason, we can see a large distance between Chain A, Tfbp1 and Lachancea dasiensis, Lachancea mirantina in our evolutionary tree.