UTM UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

Semester II 2021/2022

Subject : Bioinformatics I (SECB2103)

Section : 01 – Dr Haslina Hashim

Topic : Class Project – Find a gene

Group : Group 4 - Heong Yi Qing (A20EC0043)

Chong Kah Wei (A20EC0027)

Goh Yitian (A20EC0038)

1. Tell me the name of a protein you are interested in. Include the species and the accession number.

Gag-Pol [Human immunodeficiency virus 1] (AAF13061.1)

2. Perform a BLAST search against a DNA database, such as a database consisting of genomic DNA or ESTs. The BLAST server can be at NCBI or elsewhere. Include the output of that BLAST search in your document. If appropriate, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output (e.g. alt print screen on a PC). It is not necessary to print out all of the blast results if there are many pages.

Figure 1 shows the performance of TBLASTN search against homosapiens ESTs using AAF13061.1 as a query. Some of the results are then shown in Figure 2 and the results are arranged from lower E-value to higher E-value. From the results, we choose a match which is IL3-MT0267-261200-410-C08 MT0267 Homo sapiens cDNA, mRNA sequence. For this match, the E-value is 2e-20 and the score is 96.7bits(239) as shown in Figure 3.

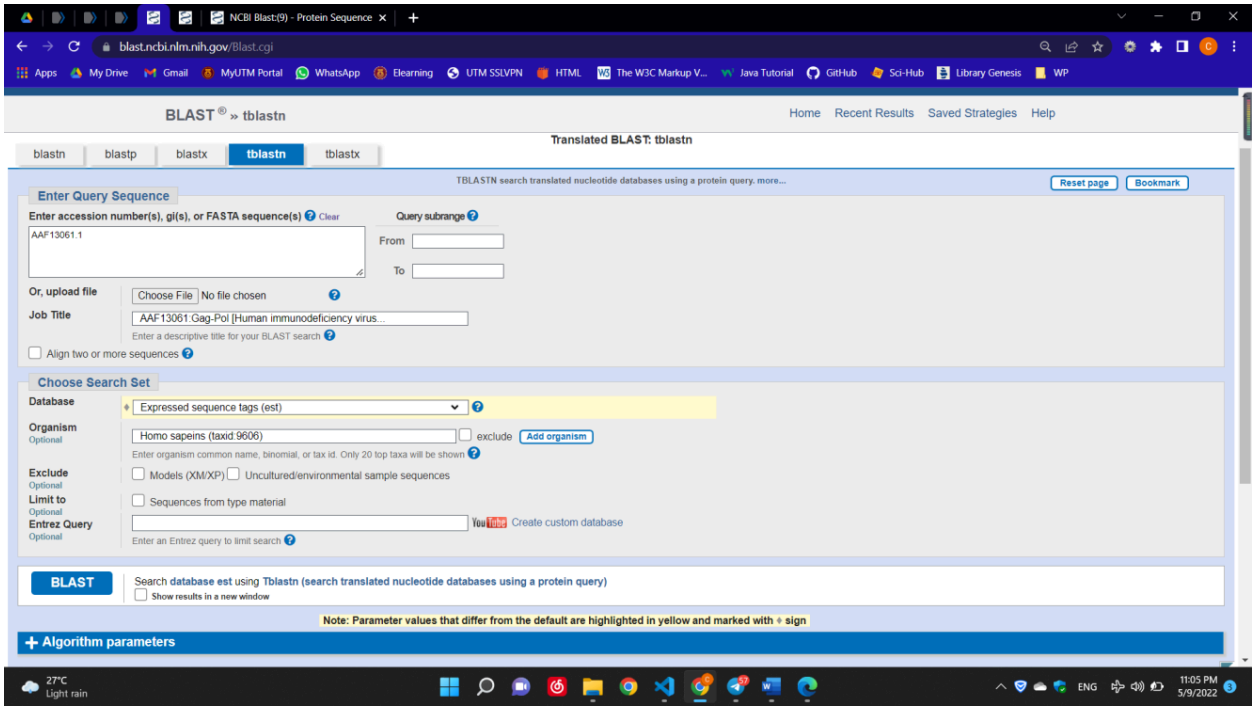


Figure 1: TBLASTN using AAF13061.1 as query

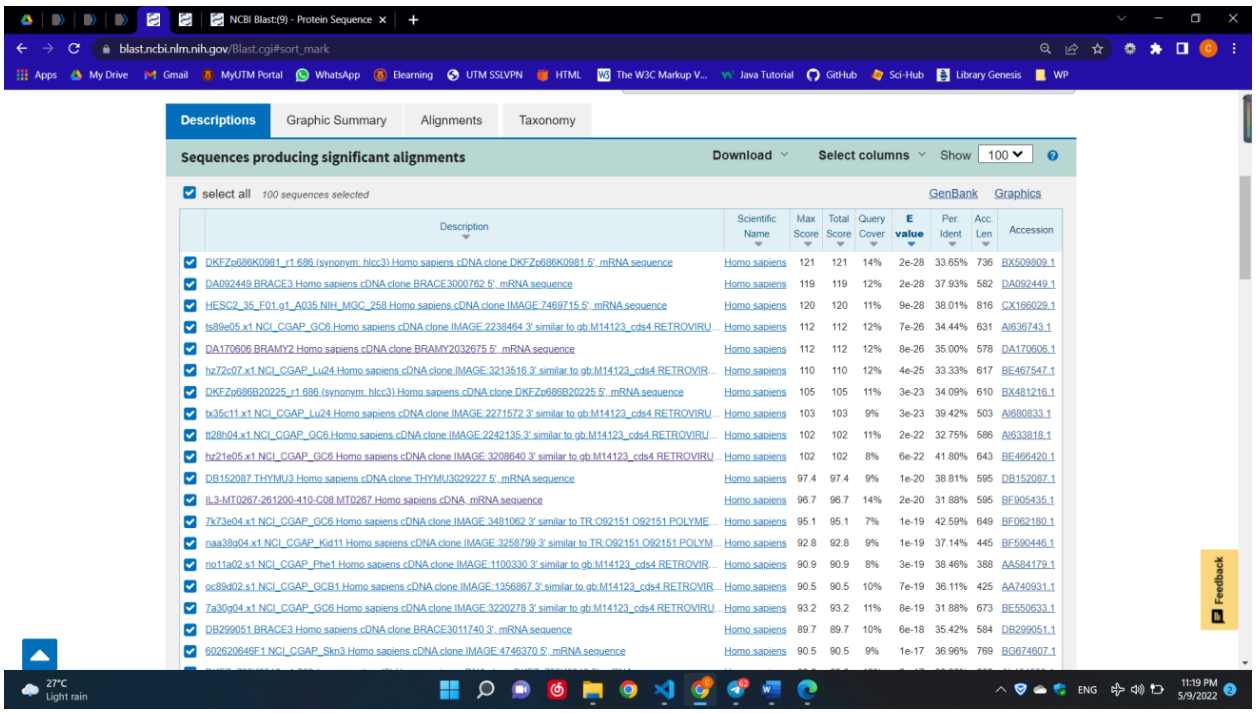


Figure 2: Part of the results

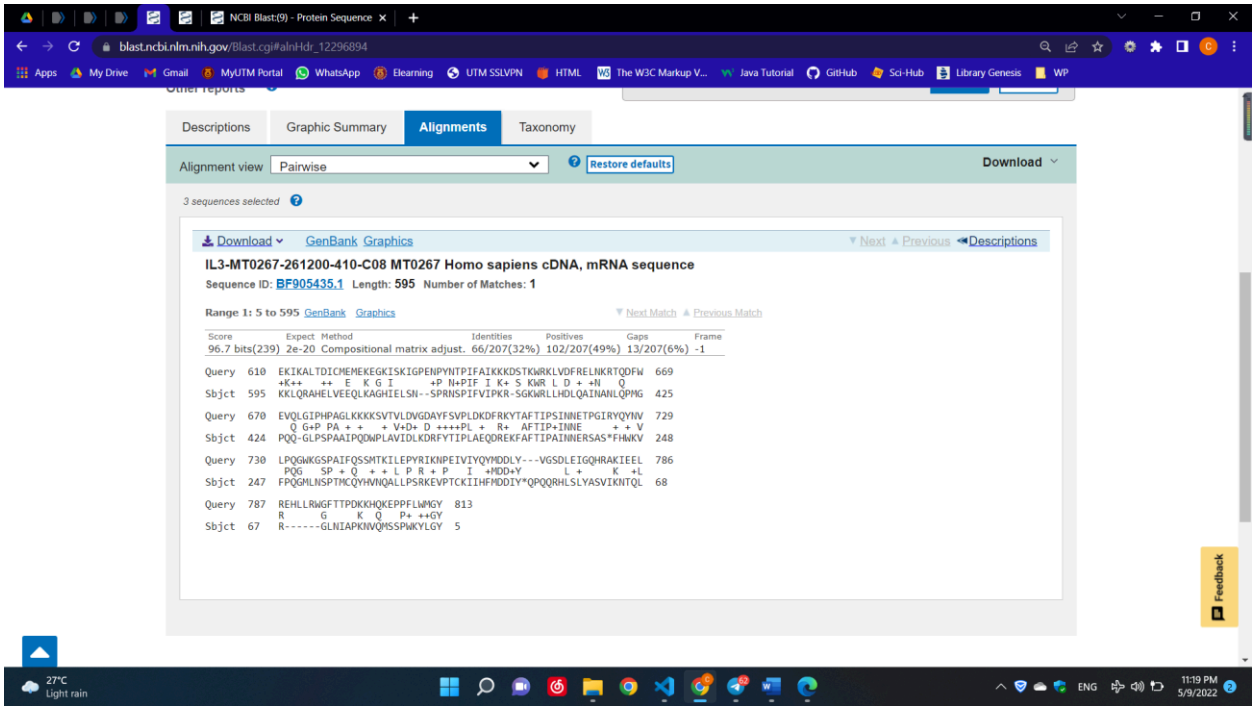


Figure 3: Alignment using IL3-MT0267-261200-410-C08 MT0267 Homo sapiens cDNA, mRNA sequence as subject

>IL3-MT0267-261200-410-C08 MT0267 Homo sapiens cDNA, mRNA sequence

Sequence ID: BF905435.1 Length: 595

Range 1: 5 to 595

Score:96.7 bits(239), Expect:2e-20,

Method:Compositional matrix adjust.,

Identities:66/207(32%), Positives:102/207(49%), Gaps:13/207(6%)

Query 610 EKIKALTDICMEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFW 669

+K++ ++ E K G I +P N+PIF I K+ S KWR L D + +N Q

Sbjct 595 KKLQRAHELVEEQLKAGHIELSN--SPRNSPIFVIPKR-SGKWRLLHDLQAINANLQPMG 425

Query 670 EVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDKDFRKYTAFTIPSINNETPGIRYQYNV 729

Q G+P PA + + + V+D+ D ++++PL + R+ AFTIP+INNE + + V

Sbjct 424 PQQ-GLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAINNERSAS*FHWKV 248

Query 730 LPQGWKGSPAIFQSSMTKILEPYRIKNPEIVIYQYMDDLY---VGSDLEIGQHRAKIEEL 786

PQG SP + Q + + L P R + P I +MDD+Y L + K +L

Sbjct 247 FPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQQRHLSLYASVIKNTQL 68

Query 787 REHLLRWGFTTPDKKHQKEPPFLWMGY 813

R G K Q P+ ++GY

Sbjct 67 R------GLNIAPKNVQMSSPWKYLGY 5

3. Gather information about this "novel" protein. At a minimum, show me the protein sequence of the "novel" protein as displayed in your BLAST results from step [2]. In some cases, you will be able to do further BLAST searches to obtain even more sequence of your novel gene.

Name: Homo sapiens cDNA

Organism: Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;Mammalia; Eutheria; Euarchontoglires;Primates; Haplorrhini;Catarrhini; Hominidae; Homo.

```
>Homo sapiens cDNA, mRNA sequence(Sequence taken from blast result)
KKLQRAHELVEEQLKAGHIELSN--SPRNSPIFVIPKR-SGKWRLLHDLQAINANLQPMG
PQQ-GLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAINNERSAS*FHWKV
FPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQQRHLSLYASVIKNTQL
R------GLNIAPKNVQMSSPWKYLGY
```

4. Prove that this gene, and its corresponding protein, are novel. For the purposes of this project, "novel" is defined as follows. Take the protein sequence (step [3]) and use it as a query in a blastp search of the nr database at NCBI.

Figure 4 shows that the BLASTP search is performed using the protein sequence that we obtained earlier from the TBLASTN result as query, which is:

```
>Homo sapiens cDNA, mRNA sequence(Sequence taken from blast result)
KKLQRAHELVEEQLKAGHIELSN--SPRNSPIFVIPKR-SGKWRLLHDLQAINANLQPMG
PQQ-GLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAINNERSAS*FHWKV
FPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQQRHLSLYASVIKNTQL
R------GLNIAPKNVQMSSPWKYLGY
```

The results are then shown in Figure 5 and the result is arranged from higher percent identity to lower percent identity. Figure 6 shows the top results, which is pol protein from human endogenous retrovirus K.

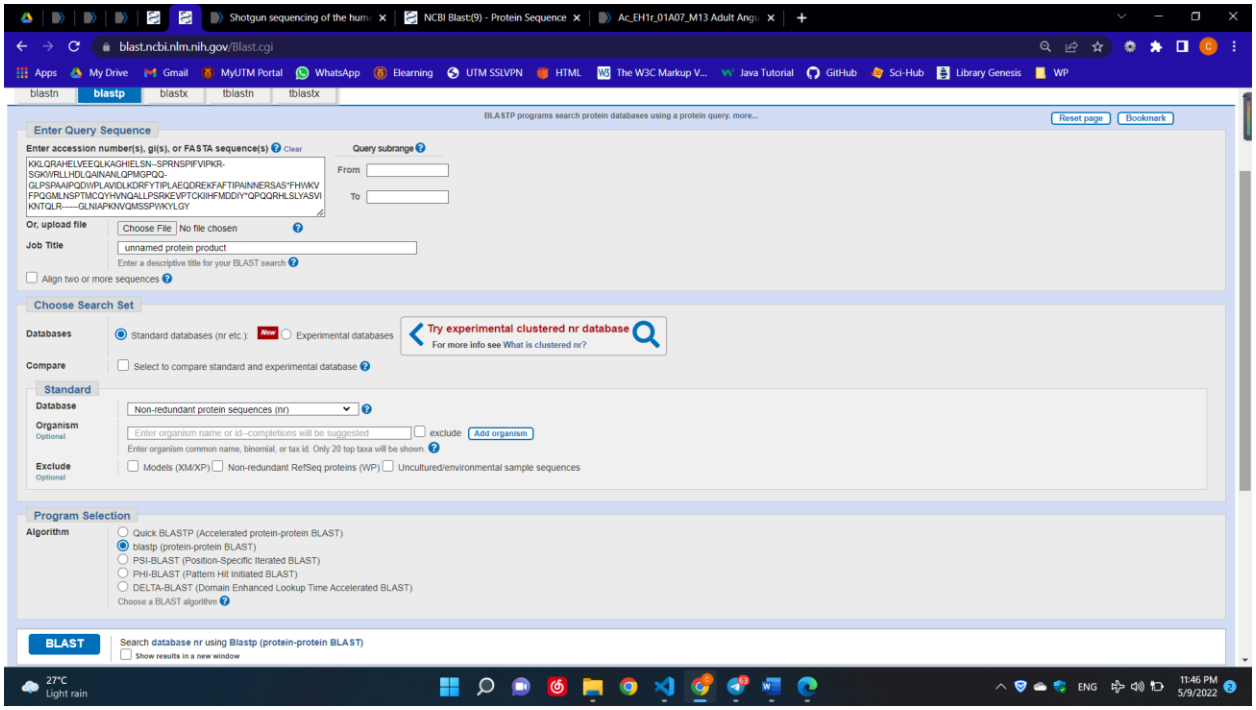


Figure 4: BLASTP search using the protein sequence obtained earlier

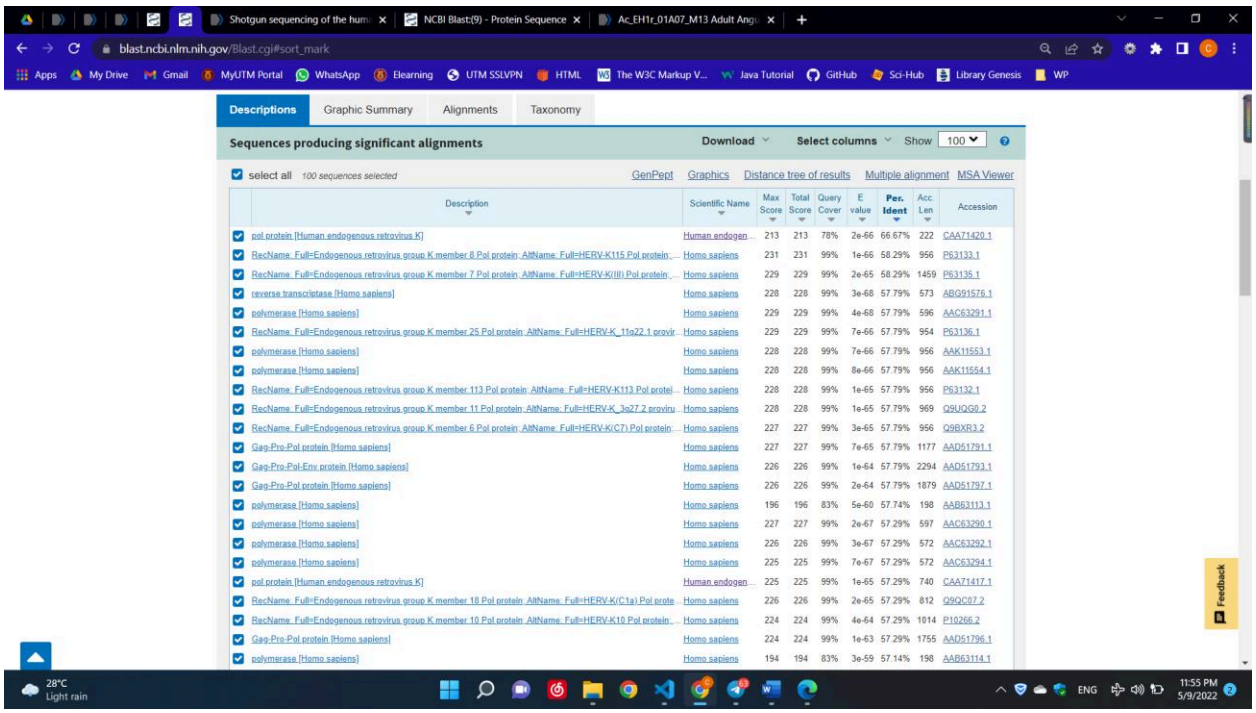


Figure 5: Top of BLASTP results which are arranged according to the percent Identity

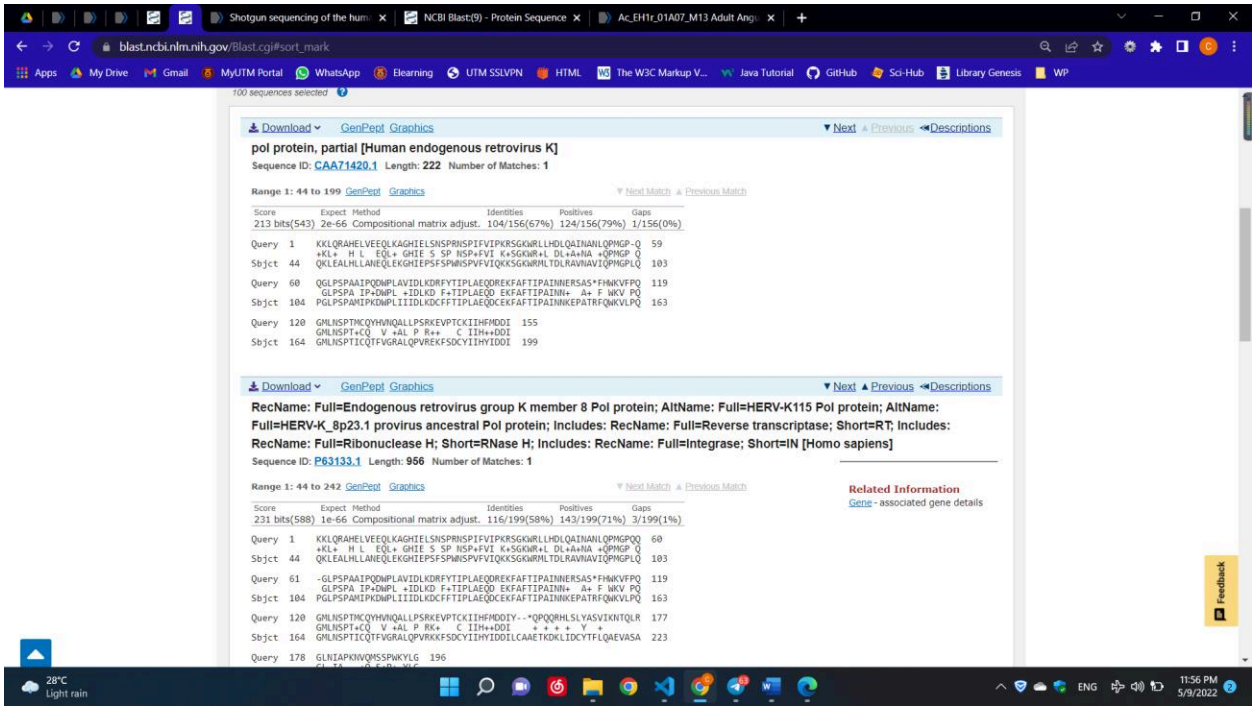


Figure 6: Top result

The top result is only 67% percent identity match to the query and the others are less than that.

Since there is no sequence with 100% percent identity and all sequences are actually less than 100% percent identity matched to the query sequence, there is a possibility that the protein is a novel protein.

5. Generate a multiple sequence alignment with your novel protein, your original query protein, and a group of other members of this family. A typical number of proteins to use in a multiple sequence alignment is a minimum of 5 or 10 and a maximum 30, although the exact number is up to you.

Firstly, we use the sequence found in Q3 to do the BLASTP and set the database to reference protein (refseq protein) to generate the protein sequences for multiple alignments.

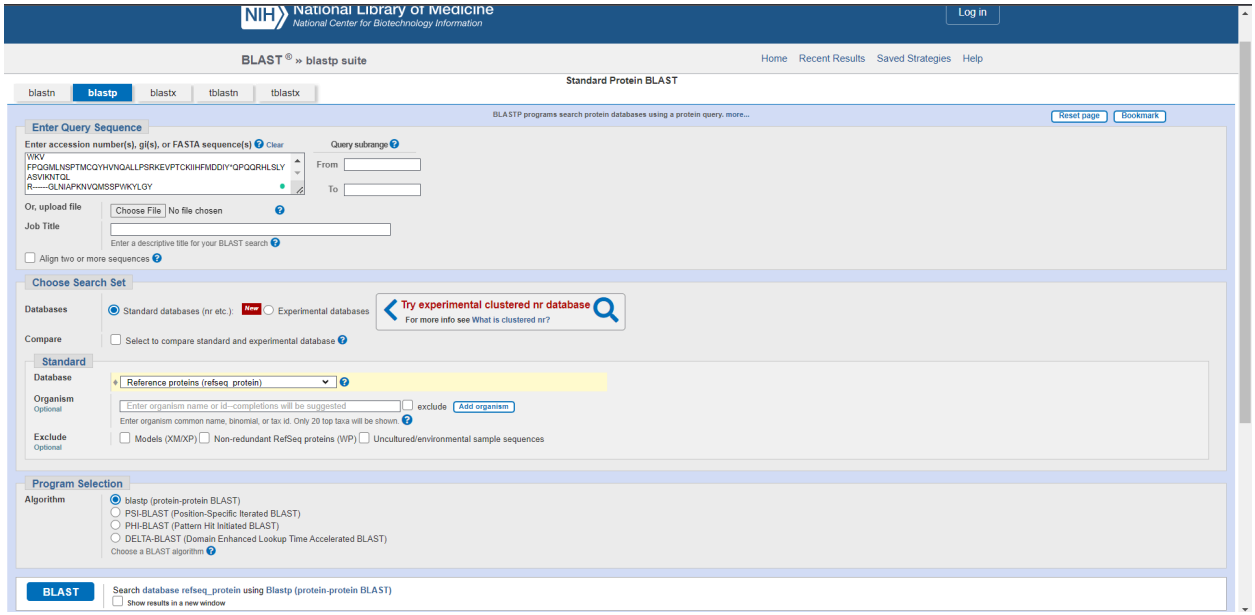


Figure 7: The setting for BLASTP on novel protein

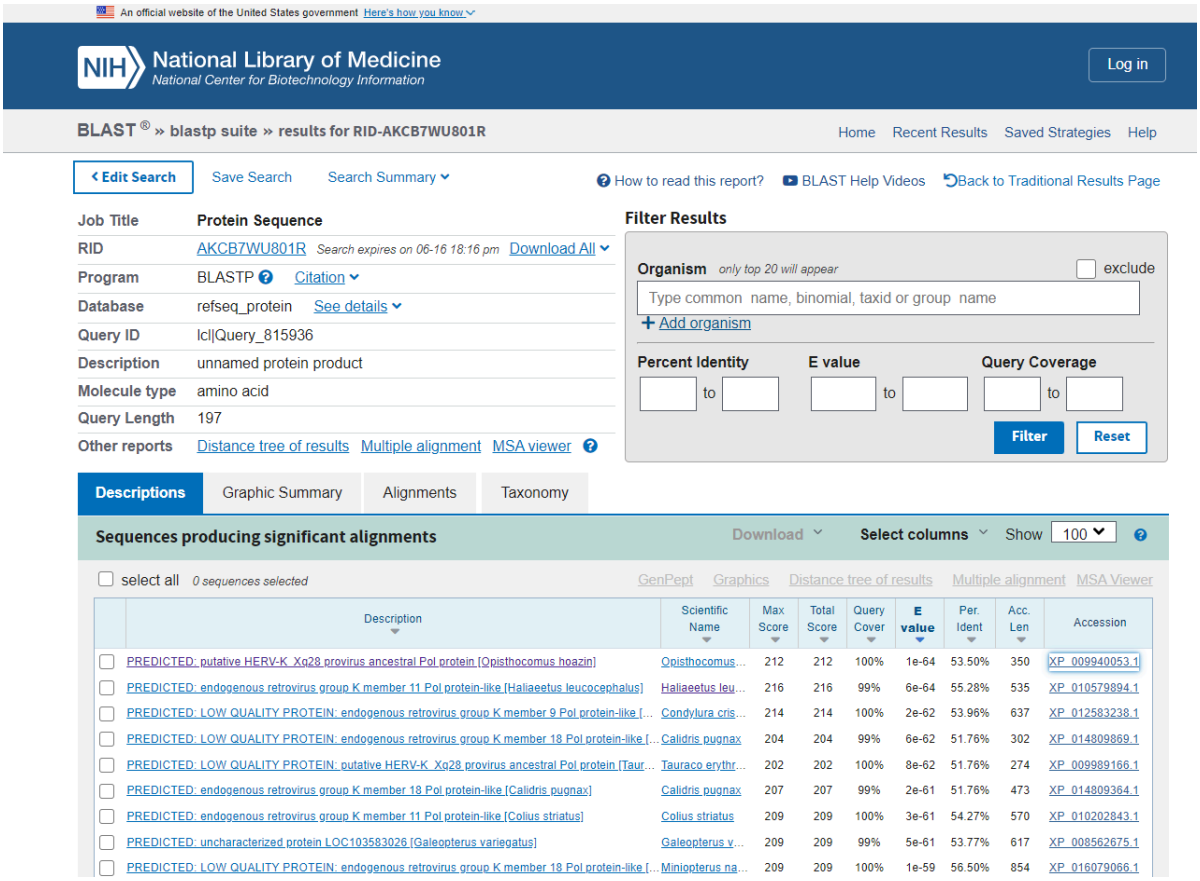
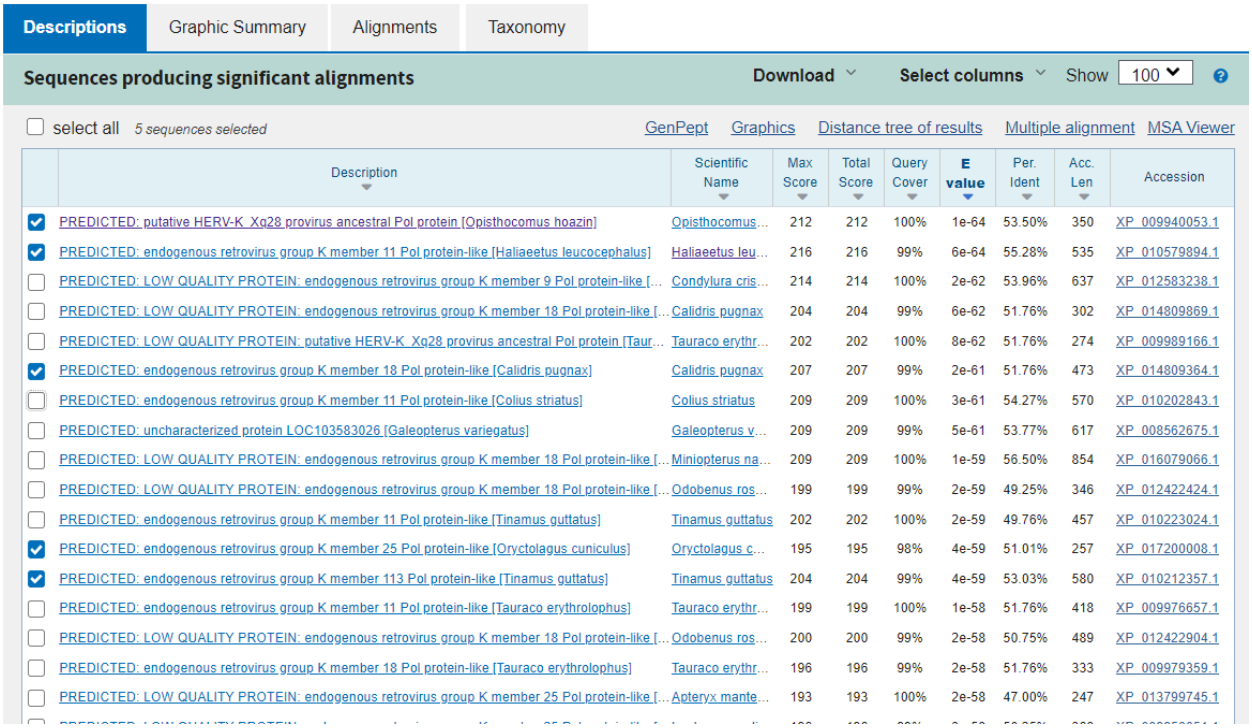


Figure 8: The result of BLASTP on novel protein

From the result, we choose 5 sequences from different family



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

select all 5 sequences selected — GenPept | Graphics | Distance tree of results | Multiple alignment | MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	PREDICTED: putative HERV-K_Xq28 provirus ancestral Pol protein [Opisthocomus hoazin]	Opisthocomus...	212	212	100%	1e-64	53.50%	350	XP_009940053.1
☑	PREDICTED: endogenous retrovirus group K member 11 Pol protein-like [Haliaeetus leucocephalus]	Haliaeetus leu...	216	216	99%	6e-64	55.28%	535	XP_010579894.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 9 Pol protein-like [...	Condylura cris...	214	214	100%	2e-62	53.96%	637	XP_012583238.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 18 Pol protein-like [...	Calidris pugnax	204	204	99%	6e-62	51.76%	302	XP_014809869.1
☐	PREDICTED: LOW QUALITY PROTEIN: putative HERV-K_Xq28 provirus ancestral Pol protein [Taur...	Tauraco erythr...	202	202	100%	8e-62	51.76%	274	XP_009989166.1
☑	PREDICTED: endogenous retrovirus group K member 18 Pol protein-like [Calidris pugnax]	Calidris pugnax	207	207	99%	2e-61	51.76%	473	XP_014809364.1
☐	PREDICTED: endogenous retrovirus group K member 11 Pol protein-like [Colius striatus]	Colius striatus	209	209	100%	3e-61	54.27%	570	XP_010202843.1
☐	PREDICTED: uncharacterized protein LOC103583026 [Galeopterus variegatus]	Galeopterus v...	209	209	99%	5e-61	53.77%	617	XP_008562675.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 18 Pol protein-like [...	Miniopterus na...	209	209	100%	1e-59	56.50%	854	XP_016079066.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 18 Pol protein-like [...	Odobenus ros...	199	199	99%	2e-59	49.25%	346	XP_012422424.1
☐	PREDICTED: endogenous retrovirus group K member 11 Pol protein-like [Tinamus guttatus]	Tinamus guttatus	202	202	100%	2e-59	49.76%	457	XP_010223024.1
☑	PREDICTED: endogenous retrovirus group K member 25 Pol protein-like [Oryctolagus cuniculus]	Oryctolagus c...	195	195	98%	4e-59	51.01%	257	XP_017200008.1
☑	PREDICTED: endogenous retrovirus group K member 113 Pol protein-like [Tinamus guttatus]	Tinamus guttatus	204	204	99%	4e-59	53.03%	580	XP_010212357.1
☐	PREDICTED: endogenous retrovirus group K member 11 Pol protein-like [Tauraco erythrolophus]	Tauraco erythr...	199	199	100%	1e-58	51.76%	418	XP_009976657.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 18 Pol protein-like [...	Odobenus ros...	200	200	99%	2e-58	50.75%	489	XP_012422904.1
☐	PREDICTED: endogenous retrovirus group K member 18 Pol protein-like [Tauraco erythrolophus]	Tauraco erythr...	196	196	99%	2e-58	51.76%	333	XP_009979359.1
☐	PREDICTED: LOW QUALITY PROTEIN: endogenous retrovirus group K member 25 Pol protein-like [...	Apteryx mante...	193	193	100%	2e-58	47.00%	247	XP_013799745.1

Figure 9: Choose 5 BLASTP results from different family

Our original query protein:

```
>AAF13061.1 Gag-Pol [Human immunodeficiency virus 1]

MGARVSVLRGGQLDTWEKIRLRPGGKKKYKMKLLVWASRELERFAVNPGLLETTEGCQQILEQLQPALKA

GSEELKSLYNTVATLYCVHQKIDVRDTKEALDKLEEIQNKSKQKTQQAAANSQVSQNYPIVQNAQGQMVH

QAISPRTLNAWVKVVEEKAFSPEVIPMFTALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDR

THPIHAGPNPPGQMREPRGSDIAGTTSNLQEQIAWMTGNPPIPVGEIYKRWIVLGLNKIVRMYSPVGILD

IRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANPDCKTILRALGPGATLEEMMTACQGVG

GPGHKARVLAEAMSQVQSPNILMQRGNFKGQKRIKCFNCGKEGHLARNCRAPRKKGCWKCGKEGHQMKDC

TERQANFLRENLAFQQREARELSSEQTGAISPTGRELWDKGRNNLLSAAGTEGQGTISSFNFPQITLWQR

PLVTVRIGGQLIEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVRQYDQIPIEICGKKAIGTVLVG
```

PTPVNIIGRNMLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTDICMEMEKEGKISK

IGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVP

LDKDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPYRIKNPEIVIYQYMDDLYV

GSDLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIMLPEKDSWTVNDIQK

LVGKLNWASQIYAGIKVKELCKLLRGAKALTDIVTLTEEAELELAENREILKEPVHGVYYDPTKDLIAEI

QKQGQDQWTFQIYQEPFKNLKTGKYQERVAPYDLSITELTEVVQKVTTESIIIWGKTPKFRLPIQRETWE

AWWMEYWQATWIPEWEFVNTLPLVKLWYQLEKDPIVGAETFYVDGAANRETKLGKAGYVTDRGRQKVVSL

TETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESELVNQIIEKLIGKDKVYLSWVPAH

KGIGGNEQVDNLVSSGFRKILFLDGLDKAQEEHEKFHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEA

MHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFILKLAGRWPVKIIHTDNGS

NFTSAAVRAACWWANVTQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKG

GIGGYSAGERIIDIIASDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKV

VPRRKAKILRDYGKQMAGDDCVAGRQDED

Our novel protein:

>Homo sapiens cDNA, mRNA sequence(Sequence taken from blast result)

KKLQRAHELVEEQLKAGHIELSN--SPRNSPIFVIPKR-SGKWRLLHDLQAINANLQPMG

PQQ-GLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAINNERSAS*FHWKV

FPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQQRHLSLYASVIKNTQL

R------GLNIAPKNVQMSSPWKYLGY

Below is the list of 5 proteins sequences selected by our group for the Multiple Sequence Alignment:

>XP_009940053.1 PREDICTED: putative HERV-K_Xq28 provirus ancestral Pol protein [Opisthocomus hoazin]

MKNKPEIEVMVESQNQEMLKLKMMIDTGADVTIISAPHWPSHWPTVESFTGVYAFLGAAIEQQPVLKIKWKTNNPFWVD
QWPLTAERLQKISELVEEQLQVGHIQPSTSPWNTPIFTIPKNSGKWRLLYDLRAVNNVMEDMGALQPGLPSPVMIPENW
TVLVIDLKDCFFTIPLHPDDAERFAFSVPSVNKEEPARRFHWIVLPQGMKNSPTMCQIFVAWAFQPICKKMPQLLIYHY
MDDILIAGQNMDREFVLQEVVREVESRGLNIAPEKIQKQEPWNYLGWVILQGSIKPPKMQLNPEIKTLNDVQKLVGDIQ
WVRTLCDITNDDLAPLVELLGTTSRADDKRTMEP

>XP_010579894.1 PREDICTED: endogenous retrovirus group K member 11 Pol protein-like [Haliaeetus leucocephalus]

MGGENYHGFLTRATVLQGVRQPTLVLTWLTNNPVWVDQWPLPIEKLKALQELVAEQLAAGHIEPSQSPWNTPVFVIKKK
SGKWRFLHDLQQVNAVMATMGALQPGMPSPAMIPQDWEIIVMDLKDCFFTIPLASQDKEKFAFSVPSINHAEPAKRYQW
RVLPQGMKNSPTICQWFVAQALSPVREKFPTSYCYHYMDDILLASDNKEQLNDMENLARNLLQQYGLVIAPEKVQKIAP
WKYLGMTITSKQVVPQPVKLNLAVKTLNDVQKLMGSLNWIRPYLGLTNSQLQPLLDLLKHSNDPTEPRILNKEALNVIH
MVEQCIYKKFVSRIDLSQLVQFFVLIDKTVPFGALVQWNSEWDDPLHILEWMFLSFRPRKTASGLFELIADVIIKTRKR
CVELIGRDPATIVLPVQNWYFEWCLANNYELQVAMAGFQRQISYHLPSHLLLKFAQEIPFGQKYLSQPEPVKGPTVFTD
GSGKTGKAAVVCLLQRALQRAAGTIFLAQIQKGGIVGGCNGSAECLTVREQRDLEQLVVYP

>XP_014809364.1 PREDICTED: endogenous retrovirus group K member 18 Pol protein-like [Calidris pugnax]

MGLFVLPGIIDADFTGEIKIMAWTPSPPCFVPKGQRIAQLVPLPSVTIPGEGNRKGGFGSTGKPVVLWSKQVSKEQPLL
RCQVHDRHFSGLVDTGADVTIINISDWPPEWPLRDPTSAIVGVGGLQKPKQSAKILTFKGPEGQIAHAAPYILPVPCTL
WGCALYIMGPRLVKPVGNLPENKFSIGAIDKQQTLSLTWKSEKTVWVDQWPLKKDKLLHLHDLVQEQLAAGHIVPTTSP
WNTPVFVIPKKGGRWRLIHDLRAINAVIEPMGALQPGLPSPSMLPQNWPIAIIDLKDCFFAIPLHPKDAPRFAFSVPAV
NQEQPTRRYHWTVLPQGMLNSPTICQLTVANALQPVRNANPHVIIYHYMDDILIAAEKDKDLQIVVCQVRLAVQGAGLQ
IAEEKVQREPPWKYLGWKITTHPIQPQALQLALQIKTLNDLQKLLGTINWLRPFLGITTQDLHPLFQLLPLTNRTLCI

>XP_017200008.1 PREDICTED: endogenous retrovirus group K member 25 Pol protein-like, partial [Oryctolagus cuniculus]

SSAEAFSPRAGFFIGAAEEGIPITWKHEDPVWVPQWPLSSDKLVAAQQLIQEQLNLGHIRPSVSPWNTPIFVIKKKSGK
WRLLHDLRVINMQMQVMGPVQRGLPLLSALPQGWPIIIIDIKDCFFSIPLHTKDCERFAFTLPACNHEQPDQRYEWVVL
PQGMANSPTMCQLFVGQAIGPLRKRFSSLKCIHYMDDILLAAKDEFVLDDGFAYLIQLLESKKLFIAPEKVQKGSIATY
LGSCITSTQLFLKGGIAQDS

>XP_010212357.1 PREDICTED: endogenous retrovirus group K member 113 Pol protein-like [Tinamus guttatus]

MQGVGKREAERGSPPHNDTSTGCLERCSAATRGSAGVDVATAVDVMLTDTRVRVIESELSGPLGQGLSALLLGRSSVSR
QGIFVVPGLIDADYTGVIKIMVYTSAPPVTIPAHSKIAQLIPFKACVPHDALTERCNGEFGSTGPLNMLLAIDIKKGKP

EELVRLQHPGGQTITLTMVVDTGADVSIIPQHMWPRAWPISLAATSVMGVGGAQWTVDYSAFLVAAIGVQSTLKLTWKM
ETPVWVDQWPLKQDRLHIVEQLVQEQLEAGHIVPSQSPWNTPIFTIPKKSGKWRLLHDLRAINAVMQDMGALQPGLPSP
VMLPEGWDLLIIDLKDCFFTIALHPQDAEKFAFIVPSINKAAPAKRYHWVVLPQGMKNSPTICQTFVAWALQPVRAKHP
ELLIYHYMGDILIAGENMCMKSVFQEVGEELGKRGLTIAPEKVQRQGPWNYLGWTIMGSEIRPQKIAIRTEVRTLVDVQ
RLVGDIQWVRGICGITNDDIAPLMPLLGTSVNASEARELSKEQREAVQAIAEKIAGAYASRIILEKKIWFLIVNSGGSH
QHPMGLIAQILQGLRTLHILEWVFLSY

Simplified original query protein, novel protein and 5 proteins of other family.

>Human immunodeficiency virus 1 AAF13061.1 Gag-Pol

MGARVSVLRGGQLDTWEKIRLRPGGKKKYKMKLLVWASRELERFAVNPGLLETTEGCQQILEQLQPALKA
GSEELKSLYNTVATLYCVHQKIDVRDTKEALDKLEEIQNKSKQKTQQAAANSQVSQNYPIVQNAQGQMVH
QAISPRTLNAWVKVVEEKAFSPEVIPMFTALSEGATPQDLNMMLNIVGGHQAAMQMLKDTINEEAAEWDR
THPIHAGPNPPGQMREPRGSDIAGTTSNLQEQIAWMTGNPPIPVGEIYKRWIVLGLNKIVRMYSPVGILD
IRQGPKEPFRDYVDRFFKTLRAEQATQEVKNWMTETLLVQNANPDCKTILRALGPGATLEEMMTACQGVG
GPGHKARVLAEAMSQVQSPNILMQRGNFKGQKRIKCFNCGKEGHLARNCRAPRKKGCWKCGKEGHQMKDC
TERQANFLRENLAFQQREARELSSEQTGAISPTGRELWDKGRNNLLSAAGTEGQGTISSFNFPQITLWQR
PLVTVRIGGQLIEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIKVRQYDQIPIEICGKKAIGTVLVG
PTPVNIIGRNMLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTDICMEMEKEGKISK
IGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVP
LDKDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPYRIKNPEIVIYQYMDDLYV
GSDLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIMLPEKDSWTVNDIQK
LVGKLNWASQIYAGIKVKELCKLLRGAKALTDIVTLTEEAELELAENREILKEPVHGVYYDPTKDLIAEI
QKQGQDQWTFQIYQEPFKNLKTGKYQERVAPYDLSITELTEVVQKVTTESIIIWGKTPKFRLPIQRETWE
AWWMEYWQATWIPEWEFVNTLPLVKLWYQLEKDPIVGAETFYVDGAANRETKLGKAGYVTDRGRQKVVSL
TETTNQKTELHAIHLALQDSGSEVNIVTDSQYALGIIQAQPDRSESELVNQIIEKLIGKDKVYLSWVPAH
KGIGGNEQVDNLVSSGFRKILFLDGLDKAQEEHEKFHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEA
MHGQVDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFILKLAGRWPVKIIHTDNGS
NFTSAAVRAACWWANVTQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKG
GIGGYSAGERIIDIIASDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKV

VPRRKAKILRDYGKQMAGDDCVAGRQDED

>(novel)Homo sapiens cDNA, mRNA sequence(Sequence taken from blast result)

KKLQRAHELVEEQLKAGHIELSN--SPRNSPIFVIPKR-SGKWRLLHDLQAINANLQPMG

PQQ-GLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAINNERSAS*FHWKV

FPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQQRHLSLYASVIKNTQL

R------GLNIAPKNVQMSSPWKYLGY

>Opisthocomus hoazin XP_009940053.1 PREDICTED: putative HERV-K_Xq28 provirus ancestral Pol protein

MKNKPEIEVMVESQNQEMLKLKMMIDTGADVTIISAPHWPSHWPTVESFTGVYAFLGAAIEQQPVLKIKWKTNNPFWVD
QWPLTAERLQKISELVEEQLQVGHIQPSTSPWNTPIFTIPKNSGKWRLLYDLRAVNNVMEDMGALQPGLPSPVMIPENW
TVLVIDLKDCFFTIPLHPDDAERFAFSVPSVNKEEPARRFHWIVLPQGMKNSPTMCQIFVAWAFQPICKKMPQLLIYHY
MDDILIAGQNMDREFVLQEVVREVESRGLNIAPEKIQKQEPWNYLGWVILQGSIKPPKMQLNPEIKTLNDVQKLVGDIQ
WVRTLCDITNDDLAPLVELLGTTSRADDKRTMEP

>Haliaeetus leucocephalus XP_010579894.1 PREDICTED: endogenous retrovirus group K member 11 Pol protein-like

MGGENYHGFLTRATVLQGVRQPTLVLTWLTNNPVWVDQWPLPIEKLKALQELVAEQLAAGHIEPSQSPWNTPVFVIKKK
SGKWRFLHDLQQVNAVMATMGALQPGMPSPAMIPQDWEIIVMDLKDCFFTIPLASQDKEKFAFSVPSINHAEPAKRYQW
RVLPQGMKNSPTICQWFVAQALSPVREKFPTSYCYHYMDDILLASDNKEQLNDMENLARNLLQQYGLVIAPEKVQKIAP
WKYLGMTITSKQVVPQPVKLNLAVKTLNDVQKLMGSLNWIRPYLGLTNSQLQPLLDLLKHSNDPTEPRILNKEALNVIH
MVEQCIYKKFVSRIDLSQLVQFFVLIDKTVPFGALVQWNSEWDDPLHILEWMFLSFRPRKTASGLFELIADVIIKTRKR
CVELIGRDPATIVLPVQNWYFEWCLANNYELQVAMAGFQRQISYHLPSHLLLKFAQEIPFGQKYLSQPEPVKGPTVFTD
GSGKTGKAAVVCLLQRALQRAAGTIFLAQIQKGGIVGGCNGSAECLTVREQRDLEQLVVYP

>Calidris pugnax XP_014809364.1 PREDICTED: endogenous retrovirus group K member 18 Pol protein-like

MGLFVLPGIIDADFTGEIKIMAWTPSPPCFVPKGQRIAQLVPLPSVTIPGEGNRKGGFGSTGKPVVLWSKQVSKEQPLL
RCQVHDRHFSGLVDTGADVTIINISDWPPEWPLRDPTSAIVGVGGLQKPKQSAKILTFKGPEGQIAHAAPYILPVPCTL
WGCALYIMGPRLVKPVGNLPENKFSIGAIDKQQTLSLTWKSEKTVWVDQWPLKKDKLLHLHDLVQEQLAAGHIVPTTSP
WNTPVFVIPKKGGRWRLIHDLRAINAVIEPMGALQPGLPSPSMLPQNWPIAIIDLKDCFFAIPLHPKDAPRFAFSVPAV
NQEQPTRRYHWTVLPQGMLNSPTICQLTVANALQPVRNANPHVIIYHYMDDILIAAEKDKDLQIVVCQVRLAVQGAGLQ
IAEEKVQREPPWKYLGWKITTHPIQPQALQLALQIKTLNDLQKLLGTINWLRPFLGITTQDLHPLFQLLPLTNRTLCI

>Oryctolagus cuniculus XP_017200008.1 PREDICTED: endogenous retrovirus group K member 25 Pol protein-like, partial

SSAEAFSPRAGFFIGAAEEGIPITWKHEDPVWVPQWPLSSDKLVAAQQLIQEQLNLGHIRPSVSPWNTPIFVIKKKSGK
WRLLHDLRVINMQMQVMGPVQRGLPLLSALPQGWPIIIIDIKDCFFSIPLHTKDCERFAFTLPACNHEQPDQRYEWVVL
PQGMANSPTMCQLFVGQAIGPLRKRFSSLKCIHYMDDILLAAKDEFVLDDGFAYLIQLLESKKLFIAPEKVQKGSIATY
LGSCITSTQLFLKGGIAQDS

>Tinamus guttatus XP_010212357.1 PREDICTED: endogenous retrovirus group K member 113 Pol protein-like

MQGVGKREAERGSPPHNDTSTGCLERCSAATRGSAGVDVATAVDVMLTDTRVRVIESELSGPLGQGLSALLLGRSSVSR
QGIFVVPGLIDADYTGVIKIMVYTSAPPVTIPAHSKIAQLIPFKACVPHDALTERCNGEFGSTGPLNMLLAIDIKKGKP
EELVRLQHPGGQTITLTMVVDTGADVSIIPQHMWPRAWPISLAATSVMGVGGAQWTVDYSAFLVAAIGVQSTLKLTWKM
ETPVWVDQWPLKQDRLHIVEQLVQEQLEAGHIVPSQSPWNTPIFTIPKKSGKWRLLHDLRAINAVMQDMGALQPGLPSP
VMLPEGWDLLIIDLKDCFFTIALHPQDAEKFAFIVPSINKAAPAKRYHWVVLPQGMKNSPTICQTFVAWALQPVRAKHP
ELLIYHYMGDILIAGENMCMKSVFQEVGEELGKRGLTIAPEKVQRQGPWNYLGWTIMGSEIRPQKIAIRTEVRTLVDVQ
RLVGDIQWVRGICGITNDDIAPLMPLLGTSVNASEARELSKEQREAVQAIAEKIAGAYASRIILEKKIWFLIVNSGGSH
QHPMGLIAQILQGLRTLHILEWVFLSY

At EMBL-EBI Multiple Sequence Alignment, MAFFT, paste the 7 sequences into the input and set the output format as ClustalW.

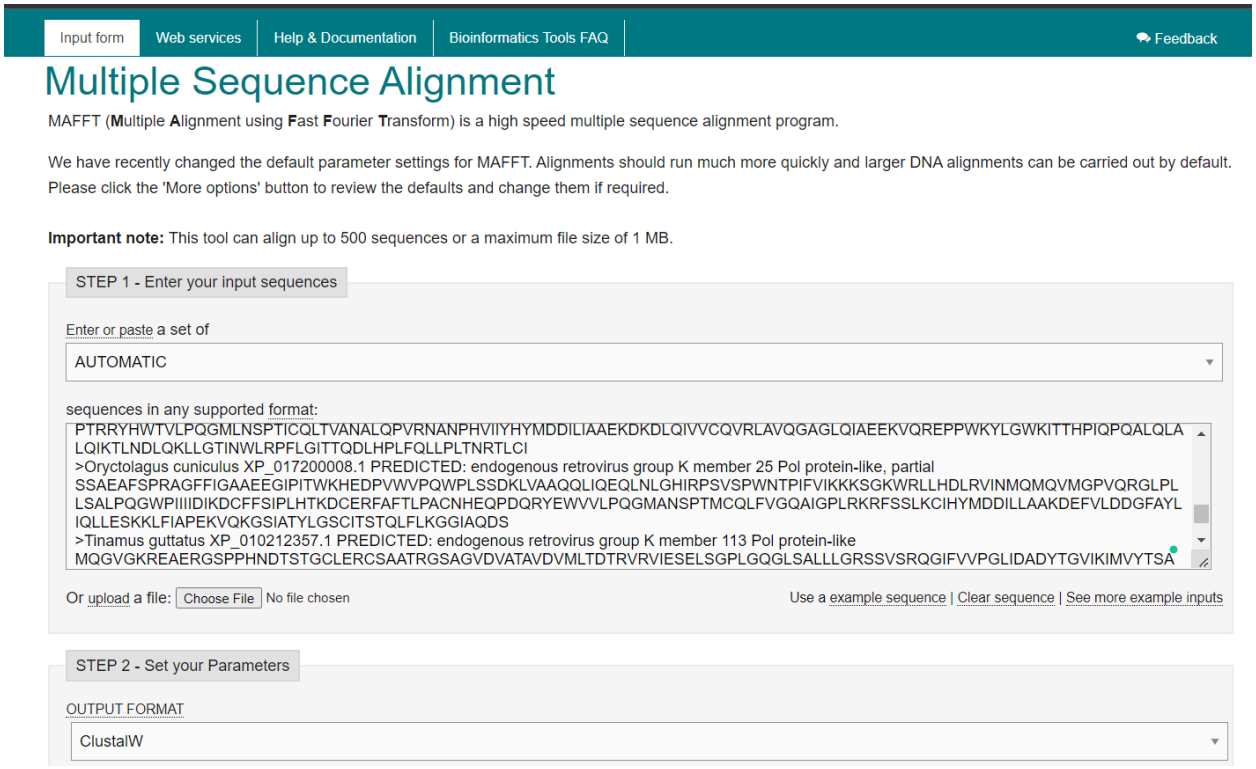


Figure 10: Paste the 7 sequences in the MAFFT, output format set to ClustalW

Figure 11: The colour results of the MAFFT

CLUSTAL format alignment by MAFFT FFT-NS-i (v7.487)

```
Human           MGARVSVLRGGQLDTWEKIRLRPGGKKKYKMKLLVWASRELERFAVNPGLLETTEGCQQI
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    M-----------------------------------------------------------
Tinamus         MQGV----------GK----------REAERGS--PPHNDTSTGC---
Oryctolagus     ------------------------------------------------------------
Calidris        M-----------------------------------------------------------
```

```
Human           LEQLQPALKAGSEELKSLYNTVATLYCVHQKIDVRDTKEALDKLE-EIQNKSKQK-TQQA
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         LERCSAATR-GSAGVD-----VAT------AVDVMLTDTRVRVIESELSGPLGQGLSALL
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------

Human           AANSQVSQNYPIVQNAQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFTALSEGATPQ
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         LGRSSVS--------RQGIFVVPGLIDADYTGVIKIMVYTSAPPVTIP------------
Oryctolagus     ------------------------------------------------------------
Calidris        -----------------GLFVLPGIIDADFTGEIKIMAWTPSPPCFVP------------

Human           DLNMMLNIVGGHQAAMQMLKDTINEEAAEWDRTHPIHAGPNPPGQMREPRGSDIAGTTS-
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         ----------AHSKIAQLI---------------PFKAC-VPHDALTERCNGEFGSTGPL
Oryctolagus     ------------------------------------------------------------
Calidris        ----------KGQRIAQLV---------------PLPSVTIPGEGNRK---GGFGSTGKP
```

```
Human           NLQEQIAWMTGNPPIPVGEIYKRWIVLGLNKIVRMYSPVGILDIRQGPKEPFRDYVDRFF
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ---------------------------------MGGE----------------------
Opisthocomus    ---------KNKPEIEVMVESQNQEMLKLKMMIDTGADVTIISAPHWPSH----------
Tinamus         NMLLAIDIKKGKPEELVRLQHPGGQTITLTMVVDTGADVSIIPQHMWPRA----------
Oryctolagus     ----------------------------------SSAE----------------------
Calidris        VVLWSKQVSKEQPLLRCQVHDRH-----FSGLVDTGADVTIINISDWPPE----------

Human           KTLRAEQATQEVKNWMTETLLVQNANPDCKTILRALGPGATLEEMMTACQGVGGPGHKAR
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    --------------W-------------------------PTVESFTGVY----------
Tinamus         --------------W-------------------------PISLAATSVMGVGG------
Oryctolagus     ------------------------------------------------------------
Calidris        --------------W-------------------------PLRDPTSAIVGVGG------

Human           VLAEAMSQVQSPNILMQRGNFKGQKRIKCFNCGKEGHLARNCRAPRKKGCWKCGKEGHQM
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         --------AQ--------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Calidris        --------LQKP---------KQSAKILTFK-GPEGQIAH--AAPY--------------
```

```
Human           KDCTERQANFLRENLAFQQREARELSSEQTGAISPTGRELWDKGRNNLLSAAGTEGQGTI
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------------------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         ----------------------------------------W-------------------
Oryctolagus     ------------------------------------------------------------
Calidris        --------------------------------ILPVPCTLW----------------GCA

Human           SSFNFPQITLWQRPLVTVRIGGQLIEALLDTGADDTVLEDINLPGKWKPKMIGGIGGFIK
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------NYHGFLTRATVLQGV-----RQPTLVLTWLTN----------
Opisthocomus    ---------------------AFLGAAIEQ---------QPVLKIKWKTN----------
Tinamus         ----------------TVDYSAFLVAAIGV---------QSTLKLTWKME----------
Oryctolagus     ---------------------AFSPRAGFFIGA-----AEEGIPITWKHE----------
Calidris        LYIMGPR---------LVKPVGNLPENKFSIGAID---KQQTLSLTWKSE----------

Human           VRQYDQIPIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPISPIETVPVKLKPGM
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------NPV---------------------------------
Opisthocomus    ------------------------NPF---------------------------------
Tinamus         ------------------------TPV---------------------------------
Oryctolagus     ------------------------DPV---------------------------------
Calidris        ------------------------KTV---------------------------------
```

```
Human           DGPKVKQWPLTEEKIKALTDICMEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVD
(novel)Homo     ------------KKLQRAHELVEEQLKAGHIEL--SNSPRNSPIFVIPKR-SGKWRLLHD
Haliaeetus      ---WVDQWPLPIEKLKALQELVAEQLAAGHIEP--SQSPWNTPVFVIKKK-SGKWRFLHD
Opisthocomus    ---WVDQWPLTAERLQKISELVEEQLQVGHIQP--STSPWNTPIFTIPKN-SGKWRLLYD
Tinamus         ---WVDQWPLKQDRLHIVEQLVQEQLEAGHIVP--SQSPWNTPIFTIPKK-SGKWRLLHD
Oryctolagus     ---WVPQWPLSSDKLVAAQQLIQEQLNLGHIRP--SVSPWNTPIFVIKKK-SGKWRLLHD
Calidris        ---WVDQWPLKKDKLLHLHDLVQEQLAAGHIVP--TTSPWNTPVFVIPKK-GGRWRLIHD
                            .::    ::  *    *:*    . .* *:*:*.* *. . :** : *

Human           FRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDKDFRKYTAFTIPSIN
(novel)Homo     LQAINANLQPMGP-QQGLPSPAAIPQDWPLAVIDLKDRFYTIPLAEQDREKFAFTIPAIN
Haliaeetus      LQQVNAVMATMGALQPGMPSPAMIPQDWEIIVMDLKDCFFTIPLASQDKEKFAFSVPSIN
Opisthocomus    LRAVNNVMEDMGALQPGLPSPVMIPENWTVLVIDLKDCFFTIPLHPDDAERFAFSVPSVN
Tinamus         LRAINAVMQDMGALQPGLPSPVMLPEGWDLLIIDLKDCFFTIALHPQDAEKFAFIVPSIN
Oryctolagus     LRVINMQMQVMGPVQRGLPLLSALPQGWPIIIIDIKDCFFSIPLHTKDCERFAFTLPACN
Calidris        LRAINAVIEPMGALQPGLPSPSMLPQNWPIAIIDLKDCFFAIPLHPKDAPRFAFSVPAVN
                :: :*     :   * *:*    : :   : ::*: * :::::*  .     ** :*: *

Human           NETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPYRIKNPEIVIYQYMDDLYVGSDLEIG
(novel)Homo     NERSAS*FHWKVFPQGMLNSPTMCQYHVNQALLPSRKEVPTCKIIHFMDDIY*QPQ---Q
Haliaeetus      HAEPAKRYQWRVLPQGMKNSPTICQWFVAQALSPVREKFPTSYCYHYMDDILLASD-NKE
Opisthocomus    KEEPARRFHWIVLPQGMKNSPTMCQIFVAWAFQPICKKMPQLLIYHYMDDILIAGQ--NM
Tinamus         KAAPAKRYHWVVLPQGMKNSPTICQTFVAWALQPVRAKHPELLIYHYMGDILIAGE--NM
Oryctolagus     HEQPDQRYEWVVLPQGMANSPTMCQLFVGQAIGPLRKRFSSLKCIHYMDDILLAAK-DEF
Calidris        QEQPTRRYHWTVLPQGMLNSPTICQLTVANALQPVRNANPHVIIYHYMDDILIAAE-KDK
```

```
                :  .   :.: *:***  .**:: *  :   : *     .     ::*.*:    .


Human           QHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIMLPEKDSWTVND
(novel)Homo     RHLSLYASVIKNTQLRGLNIAPKNVQMSSPWKYLGY-----------------------
Haliaeetus      QLNDMENLARNLLQQYGLVIAPEKVQKIAPWKYLGMTITSKQVVPQPVKLNLAVK-TLND
Opisthocomus    DREFVLQEVVREVESRGLNIAPEKIQKQEPWNYLGWVILQGSIKPPKMQLNPEIK-TLND
Tinamus         CMKSVFQEVGEELGKRGLTIAPEKVQRQGPWNYLGWTIMGSEIRPQKIAIRTEVR-TLVD
Oryctolagus     VLDDGFAYLIQLLESKKLFIAPEKVQKGSIATYLGSCITSTQLF----------------
Calidris        DLQIVVCQVRLAVQGAGLQIAEEKVQREPPWKYLGWKITTHPIQPQALQLALQIK-TLND
                               :  . :: *      ::*


Human           IQKLVGKLNWASQIYAGIKVKELCKLLRGAKALTDIVTLTEEAELELAENREILKEPVHG
(novel)Homo     ------------------------------------------------------------
Haliaeetus      VQKLMGSLNWI-RPYLGLTNSQLQPLL-------DLLKHSNDPT----EPRILNKEALNV
Opisthocomus    VQKLVGDIQWV-RTLCDITNDDLAPLV-------ELLGTTSRAD----DKRTMEP-----
Tinamus         VQRLVGDIQWV-RGICGITNDDIAPLM-------PLLGTSVNAS----EARELSKEQREA
Oryctolagus     ---LKG----------GIAQDS-------------------------------------
Calidris        LQKLLGTINWL-RPFLGITTQDLHPLF-------QLLPLTN---------RTL-------


Human           VYYDPTKDLIAEIQKQGQDQWTFQIYQEPFKNLKTGKYQERVAPYDLSITELTEVVQKVT
(novel)Homo     ------------------------------------------------------------
Haliaeetus      IHM--VEQCIY-----------------------KKFVSRI--------DLSQLVQFFV
Opisthocomus    ------------------------------------------------------------
Tinamus         VQA--IAEKIA-----------------------GAYASRI--------ILEKKIWFLI
Oryctolagus     ------------------------------------------------------------
```

```
Calidris        --------CI--------------------------------------------------
```

```
Human           TESIIIWGKTPKFRLPIQRETWEAWWMEYWQATWIPEWEFVNTLPLVKLWYQLEKDPIVG
(novel)Homo     ------------------------------------------------------------
Haliaeetus      LID-----KTVPFGALVQ------WNSEWDDPLHILEWMFLSFRP---------------
Opisthocomus    ------------------------------------------------------------
Tinamus         VNS--GGSHQHPMGLIAQ------I-LQGLRTLHILEWVFLSY-----------------
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------
```

```
Human           AETFYVDGAANRETKLGKAGYVTD---RGRQKVVSLTETTNQKTELHAIHLALQDSGSEV
(novel)Homo     ------------------------------------------------------------
Haliaeetus      -----------RKTASGLFELIADVIIKTRKRCV-------------------------
Opisthocomus    ------------------------------------------------------------
Tinamus         ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------
```

```
Human           NIVTDSQYALGIIQAQPDRSESELVNQIIEKLIGKDKVYLSWVPAHKGIGGNEQVDNLVS
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------------------ELIGRDPATIV-LPV---------------
Opisthocomus    ------------------------------------------------------------
Tinamus         ------------------------------------------------------------
```

```
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------

Human           SGFRKILFLDGLDKAQEEHEKFHSNWRAMASDFNLPPIVAKEIVASCDKCQLKGEAMHGQ
(novel)Homo     ------------------------------------------------------------
Haliaeetus      ------------------QNWYFEW-CLANNYEL--------------QVAMAGFQRQ
Opisthocomus    ------------------------------------------------------------
Tinamus         ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------

Human           VDCSPGIWQLDCTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFILKLAGRWPVK--
(novel)Homo     ------------------------------------------------------------
Haliaeetus      IS----------YHLPSHLLL-----------------KFAQEIPFGQKYLSQPEPVKGP
Opisthocomus    ------------------------------------------------------------
Tinamus         ------------------------------------------------------------
Oryctolagus     ------------------------------------------------------------
Calidris        ------------------------------------------------------------

Human           IIHTDNGSNFTSAAVRAACWWANVTQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEH
(novel)Homo     ------------------------------------------------------------
Haliaeetus      TVFTD-GSGKTGKA----------------------AVVCLLQRALQRAAG--------
Opisthocomus    ------------------------------------------------------------
```

```
Tinamus          ------------------------------------------------------------
Oryctolagus      ------------------------------------------------------------
Calidris         ------------------------------------------------------------

Human            LKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIASDIQTKELQKQITKIQNFRVYYRDSR
(novel)Homo      ------------------------------------------------------------
Haliaeetus       -------TIFLAQIQKGGIVGGCNGSAECLTV-----------------------REQR
Opisthocomus     ------------------------------------------------------------
Tinamus          ------------------------------------------------------------
Oryctolagus      ------------------------------------------------------------
Calidris         ------------------------------------------------------------

Human            DPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKILRDYGKQMAGDDCVAGRQDED
(novel)Homo      ------------------------------------------------------------
Haliaeetus       DL--------------------EQLVVYP---------------------------
Opisthocomus     ------------------------------------------------------------
Tinamus          ------------------------------------------------------------
Oryctolagus      ------------------------------------------------------------
Calidris         ------------------------------------------------------------
```

6. Create a phylogenetic tree, using either parsimony or distance-based approach. Bootstrapping and tree rooting are optional. Use any program such as MEGA, PAUP, or Phylip.

Figure 12: Input the 7 sequence into MEGA software

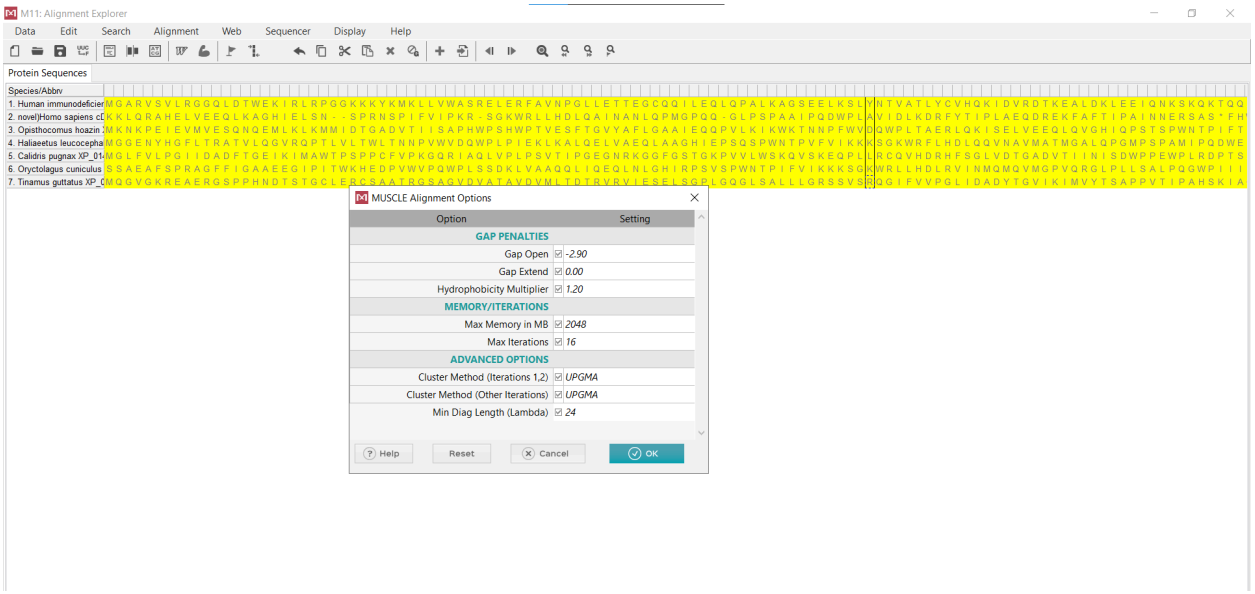


Figure 13: Default MUSCLE Alignment Options

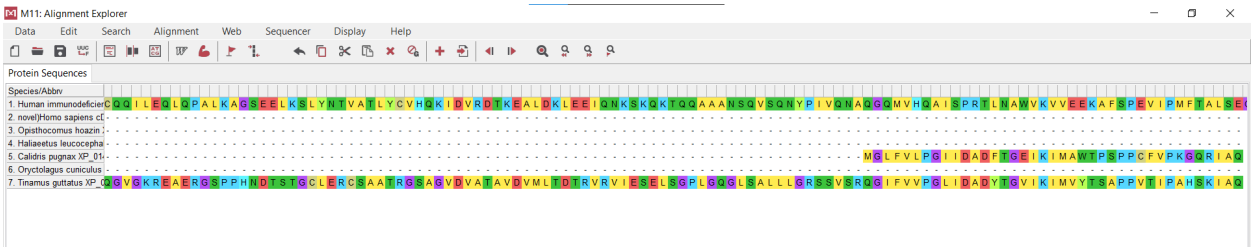


Figure 14: The result of MUSCLE alignments

Figure 15: Export alignment to MEGA format

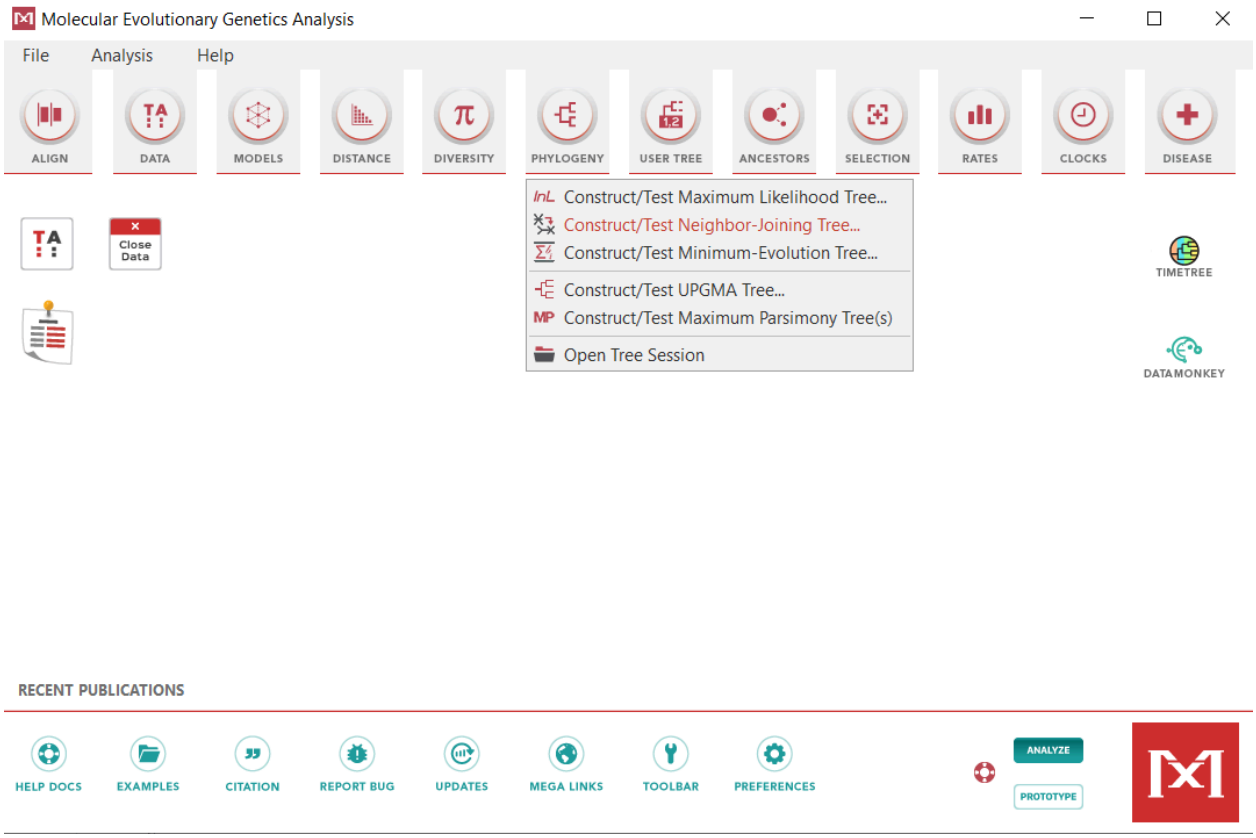


Figure 16: Select Construct/ Test Neighbor-Joining Tree at Phylogeny menu bar

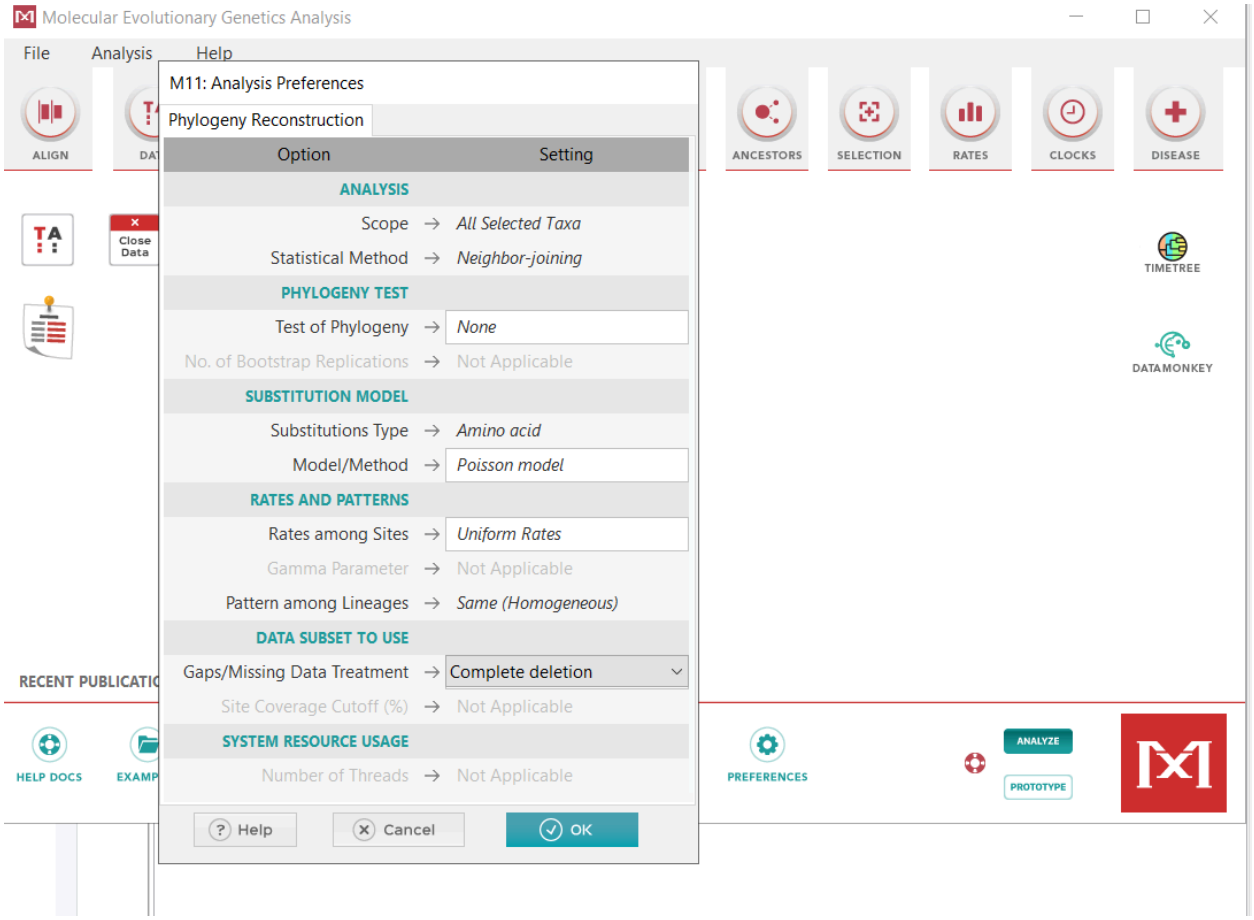


Figure 17: Phylogeny Reconstruction dialog box

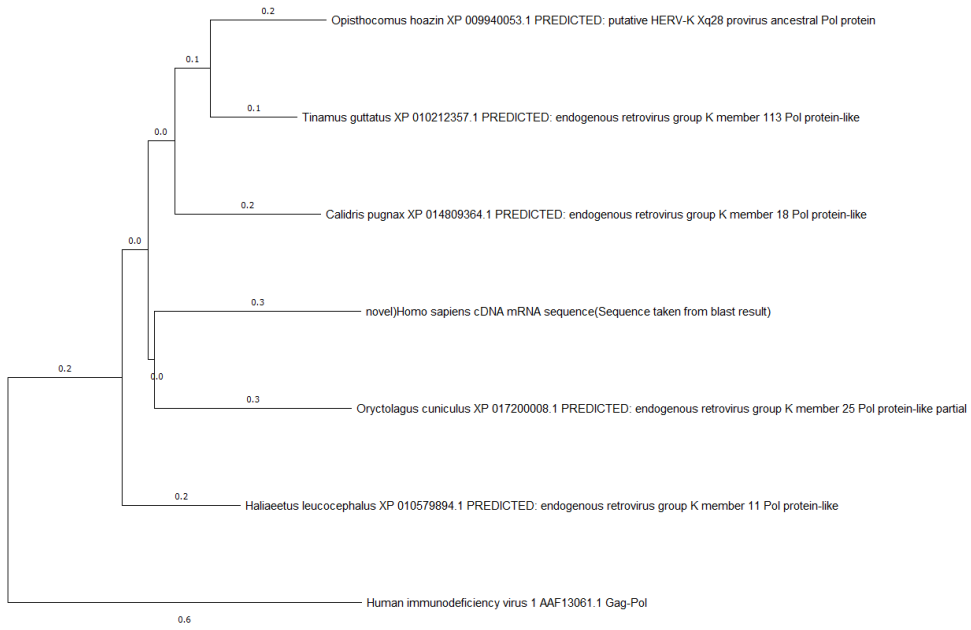


Figure 18: Rectangular Tree Style of Neighbor-Joining Tree with 7 protein sequences.

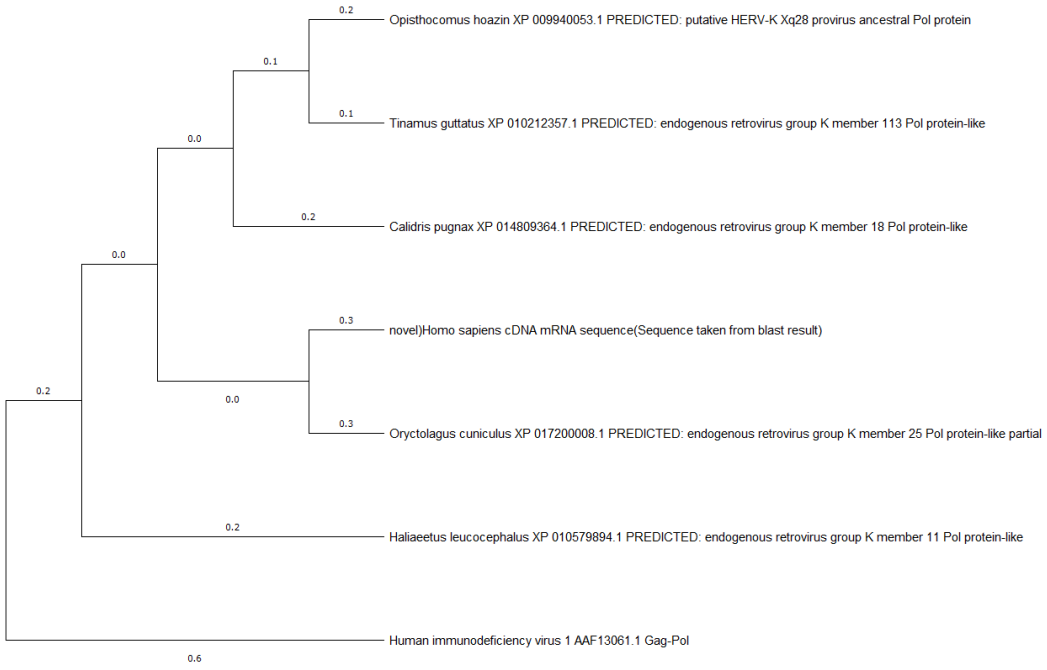


Figure 19 : "Topology Only" Tree Style of Neighbor-Joining Tree with 7 protein sequences.

Bootstrapping with Neighbor-Joining (NJ) Tree

Steps are similar to producing the Neighbor-Joining Tree as mentioned previously, at the dialog box, select Bootstrap method at the Test of Phylogeny and the Number of Bootstrap Replications is set to default which is 500.

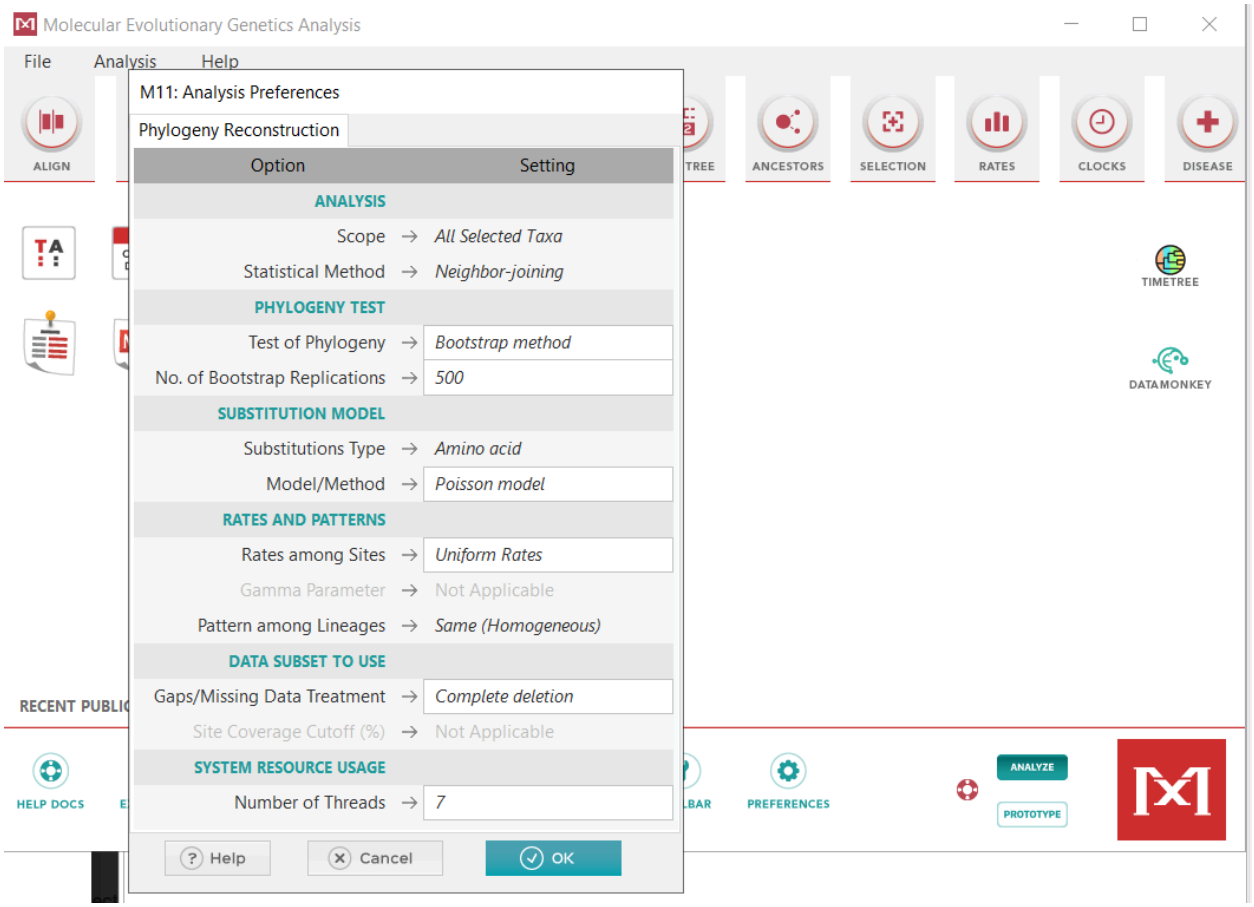


Figure 20: Bootstrap Phylogeny Reconstruction dialog box.

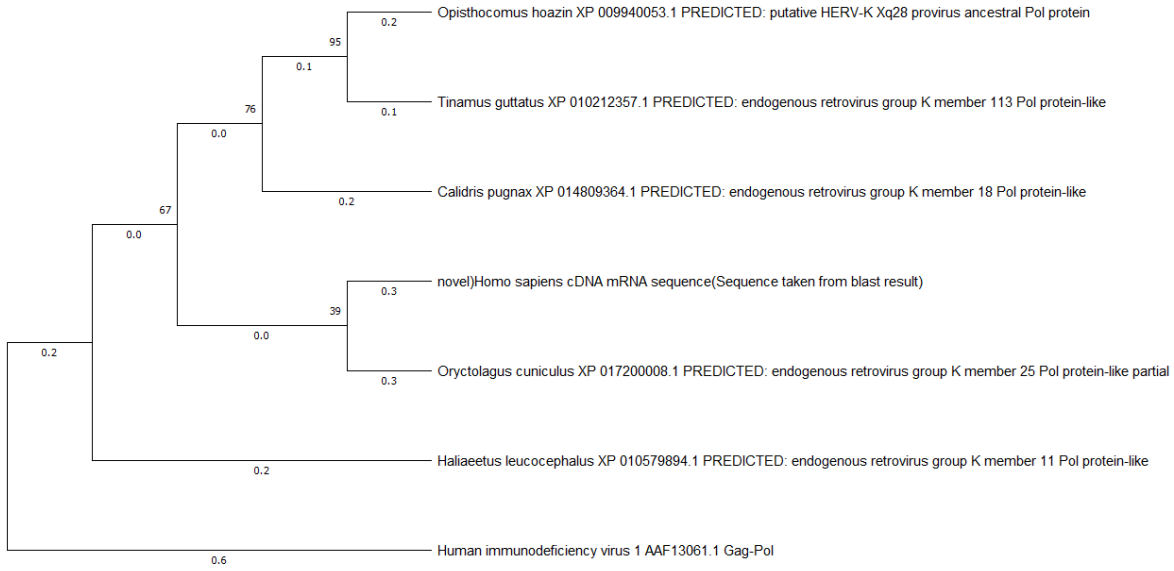


Figure 21: Bootstrap Neighbor-Joining Tree with Poisson Model.

Evolutionary relationships of taxa

The evolutionary history was inferred using the Neighbor-Joining method [1]. The optimal tree is shown. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [2] and are in the units of the number of amino acid substitutions per site. This analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 195 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [3]

1. Saitou N. and Nei M. (**1987**). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.
2. Zuckerkandl E. and Pauling L. (**1965**). Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.
3. Tamura K., Stecher G., and Kumar S. (**2021**). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* https://doi.org/10.1093/molbev/msab120.

7. Discuss the significance of your novel gene. What have you learned about this gene/protein family?

MT0627 Homo sapiens cDNA is a gene sequence from the cloning products that derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. The reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions. In this case, the gene is taken from the database of 10,000 sequences from excised human breast tumors in the research. The particular database has significant contribution to the existing public EST databases which consist of mostly the derived sequences from cDNAs and boost the construction of contigs of transcript sequences. On the other hand, as one of the cloning products of ORESTES PCR, the novel gene can become a reference to early identification of important human genes, the decoding of the draft human genome sequence currently being compiled and the shortgun sequencing of the human transcriptome (Neto et al., 2000).

To complete the project, we have applied the knowledge that we have learned from the Bioinformatics lectures. We started our project by identifying a novel gene by performing TBLASTN search using HIV1 gag-pol with the accession number AAF13061.1. Based on the findings, we discovered that the MT0627 Homo sapiens cDNA is a novel gene since the top result of BLASTP searching of the particular protein is only 67% and there is no sequence with 100% matched to the sequence. Then, we did multiple sequence alignment by using the original query protein, the novel protein and a few other members of the family to identify the evolutionary relationships and common patterns among the sequences. By using the same sequences also, we created a phylogenetic tree of the 7 proteins using MEGA software which shows the path from a common ancestor to different descendants through evolutionary time with various approaches.

References

1. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
2. https://www.ncbi.nlm.nih.gov/protein/AAF13061.1/
3. https://www.ebi.ac.uk/Tools/msa/mafft/
4. https://www.megasoftware.net/
5. Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., Da Silva, W., . . . Simpson, A. J. (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proceedings of the National Academy of Sciences, 97*(7), 3491-3496. doi:10.1073/pnas.97.7.3491