BIF-401 Assignment No.2

Bioinformatics



By

Yasin Abbas Chaudhary

❤ [AQUARIUS] ❤

Solution:

Step 1:

How to retrieve protein data from NCBI:

Protein Name: **Chain B. Vinculin**

Organism: **Homosapiens**

Accession No: **5L0F_B**

Sequence Length: **176**

Sequence:

FASTA Format:

9:29 16.0 KB/S 80

Chain B, Vinculin ...
ncbi.nlm.nih.gov

An official website of the United States government Here's how you know

National Library of Medicine
National Center for Biotechnology Information

Log in

Protein Protein Search
Advanced Help

FASTA Send to:

Chain B, Vinculin

PDB: 5L0F_B

GenPept Identical Proteins Graphics

```
>pdb|5L0F|B Chain B, Vinculin
NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGTKRALIQCAKDIAKASDE
VTRLAKEVAKQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQNL
MQSVKETVQEAEAASIKIRTDAGFTLRWVQKTPWYQ
```

Change region shown

Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features

Protein 3D Structure



Human Metavinculin Quadruple Mutant
PDB: 5L0F
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.76 Å

Related information
PubMed
Taxonomy
CDD Search Results
Conserved Domains (Concise)
Conserved Domains (Full)
Full text in PMC
Identical Structures
Structures

Recent activity
Turn Off Clear
Chain B, Vinculin Protein
Humans (4161125) Protein
Taxonomy Links for Protein (Select 1048348310) (1) Taxonomy
peroxiredoxin-6 isoform X2 [Pipistrellus kuhlii] Protein
PREDICTED: peroxiredoxin-6 isoform X1 [Colobus angolens Protein
See more...

You are here: NCBI > Proteins > Protein Database
Support Center

FOLLOW NCBI

Connect with NLM

National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Web Policies
FOIA
HHS Vulnerability Disclosure

Help
Accessibility
Careers

NLM | NIH | HHS | USA.gov

Functions:

The main cause of death globally remains debilitating heart conditions, such as dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM), which are often due to mutations of specific components of adhesion complexes. Vinculin regulates these complexes and plays essential roles in intercalated discs that are necessary for muscle cell function and coordinated movement and in the development and function of the heart. Humans bearing familial or sporadic mutations in vinculin suffer from chronic, progressively debilitating DCM that ultimately leads to cardiac failure and death, whereas autosomal dominant mutations in vinculin can also provoke HCM, causing acute cardiac failure.

Step 2:

How to run Blast for Sequence comparisons and to find orthologs:

9:33 4.00 KB/S 79

NCBI Blast:pdb|5...
blast.ncbi.nlm.nih.gov

An official website of the United States government
Here's how you know

Job Title	pdb|5L0F|B
RID	GBAEYAH7013 *Search expires on 08-25 12:32 pm* Download All
Program	BLASTP Citation
Database	nr See details
Query ID	5L0F_B
Description	Human metavinculin quadruple mutant (residues 959-1134) [Homo sapiens]
Molecule type	amino acid
Query Length	176
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity

to

E value

to

Query Coverage

to

Filter Reset

Compare these results against the new Clustered nr database BLAST

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download

Select column

Show 100

9:33 0.69 KB/S 79

Compare these results against the new Clustered nr database BLAST

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | Select columns | Show 100

select all *100 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
Human metavinculin quadruple mutant ...	Homo sa...	359	359	100%	4e-125	100.00%	176	5L0F_A
Human metavinculin (residues 959-113...	Homo sa...	352	352	100%	2e-122	97.73%	176	5L0C_A
hypothetical protein GHT09_003248 [M...	Marmota...	352	352	100%	4e-122	97.73%	190	KAF7466015.1
vinculin [Ochotona curzoniae]	Ochoton...	352	352	100%	4e-122	97.73%	190	XP_040854131.1
Human metavinculin MVt cardiomyopat...	Homo sa...	350	350	100%	1e-121	97.73%	176	5L0H_A
Raver1 in complex with metavinculin L9...	Homo sa...	354	354	100%	2e-121	97.73%	283	3VF0_A
metavinculin [Sus scrofa]	Sus scrofa	350	350	100%	1e-120	97.16%	224	CAA79600.1
Human metavinculin MVt Q971R, R975D...	Homo sa...	348	348	100%	1e-120	96.59%	176	5L0G_A
hypothetical protein EGK_19717 [Macac...	Macaca ...	356	356	100%	7e-119	97.73%	510	EHH19074.1
PREDICTED: vinculin [Acanthisitta chloris]	Acanthisi...	353	353	100%	1e-118	97.16%	462	XP_009082311.1
Human metavinculin tail domain [Homo ...	Homo sa...	341	341	97%	7e-118	97.67%	172	3MYI_A
PREDICTED: vinculin [Balearica reguloru...	Balearica...	355	355	100%	8e-118	97.16%	555	XP_010308467.1
vinculin-like [Pseudonaja textilis]	Pseudon...	345	345	100%	8e-118	96.02%	288	XP_026579862.1
RecName: Full=Vinculin; AltName: Full=...	Xenopus ...	343	343	100%	8e-118	94.89%	235	Q04615.1
PREDICTED: vinculin-like [Paralichthys o...	Paralicht...	344	344	100%	1e-117	94.89%	272	XP_019968167.1
Human vinculin R903Q, D907R, R910T ...	Homo sa...	338	338	98%	9e-117	94.80%	176	5L0J_A
Cryo-EM reconstruction of the metavinc...	Homo sa...	340	340	97%	3e-116	97.66%	273	3JBK_M
PREDICTED: vinculin-like [Thamnophis s...	Thamno...	342	342	100%	1e-115	95.45%	346	XP_013918592.1
MDFF model of the vinculin tail domain ...	Gallus ga...	337	337	97%	5e-115	97.08%	254	3JBI_V
Human raver1 RRM1, RRM2, and RRM3 ...	Homo sa...	334	334	98%	6e-115	93.64%	188	3H2U_A
Human vinculin head (1-258) in comple...	Homo sa...	333	333	98%	8e-115	93.64%	185	1RKE_B
Human Vinculin (residues 891-1066) in ...	Homo sa...	333	333	98%	1e-114	93.64%	176	4PR9_A
LOW QUALITY PROTEIN: vinculin-like [C...	Castor c...	350	350	100%	4e-114	97.16%	714	XP_020016382.1
vinculin [Homo sapiens]	Homo sa...	335	335	98%	7e-114	93.64%	293	1805212A
hypothetical protein GDO81_004614 [En...	Engysto...	330	330	96%	1e-113	95.27%	169	KAG8552624.1
Vinculin [Liparis tanakae]	Liparis ta...	338	338	100%	2e-113	92.61%	385	TNN56139.1
unnamed protein product [Gulo gulo]	Gulo gulo	354	354	100%	2e-113	97.73%	967	VCW96515.1
vinculin [Phyllostomus discolor]	Phyllosto...	336	336	100%	4e-113	92.05%	367	KAF6112200.1
unnamed protein product [Homo sapiens]	Homo sa...	335	335	98%	4e-113	93.64%	327	BAG62039.1
vinculin isoform X5 [Alligator sinensis]	Alligator ...	355	355	100%	6e-113	97.73%	1053	XP_025068523.1
PREDICTED: vinculin isoform X3 [Cerco...	Cercoce...	355	355	100%	8e-113	97.73%	1041	XP_011940943.1
vinculin [Phyllostomus discolor]	Phyllosto...	335	335	100%	8e-113	92.05%	369	KAF6112201.1
hypothetical protein FD754_011187 [Mu...	Muntiacu...	355	355	100%	9e-113	97.73%	1055	KAB0346330.1
hypothetical protein FD755_005213 [Mu...	Muntiacu...	355	355	100%	9e-113	97.73%	1055	KAB0383296.1
vinculin isoform X3 [Equus asinus]	Equus as...	355	355	100%	1e-112	97.73%	1061	XP_04461338
PREDICTED: vinculin [Equus przewalskii]	Equus pr...	355	355	100%	1e-112	97.73%	1061	XP_00851607
Vinculin [Bos mutus]	Bos mutus	354	354	100%	1e-112	97.73%	1055	ELR53368.1
vinculin isoform X5 [Phyllostomus hasta...	Phyllosto...	355	355	100%	1e-112	97.73%	1062	XP_045698182.1

Feedback

9:34 3.00 KB/S 79

Download GenPept Graphics ▼Next ▲Previous ◀Descriptions

Human metavinculin quadruple mutant (residues 959-1134) [Homo sapiens]

Sequence ID: 5L0F_A Length: 176 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 176 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
359 bits(922)	4e-125	Compositional matrix adjust.	176/176(100%)	176/176(100%)	0/176(0%)

```
Query   1     NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAE
              NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAE
Sbjct   1     NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAE

Query   61    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
              AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
Sbjct   61    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS

Query   121   ISDEESEQATEMLVHNAQNLMQSVKETVQEAEAASIKIRTDA
              ISDEESEQATEMLVHNAQNLMQSVKETVQEAEAASIKIRTDA
Sbjct   121   ISDEESEQATEMLVHNAQNLMQSVKETVQEAEAASIKIRTDA
```

Related Information
Structure 3D structure displays
Identical Proteins - Identical proteins to 5L0F_A

Download GenPept Graphics ▼Next ▲Previous ◀Descriptions

Human metavinculin (residues 959-1134) in complex with PIP2 [Homo sapiens]

Sequence ID: 5L0C_A Length: 176 Number of Matches: 1

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

Range 1: 1 to 176 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
352 bits(904)	2e-122	Compositional matrix adjust.	172/176(98%)	176/176(100%)	0/176(0%)

```
Query   1     NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAE
              NQPVNQPILAAAQSLH+EAT+WSSKGNDIIAAAKRMALLMAE
Sbjct   1     NQPVNQPILAAAQSLHREATKWSSKGNDIIAAAKRMALLMAE

Query   61    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
              AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
Sbjct   61    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS

Query   121   ISDEESEQATEMLVHNAQNLMQSVKETVQEAEAASIKIRTDA
              ISDEESEQATEMLVHNAQNLMQSVKETV+EAEAASIKIRTDA
Sbjct   121   ISDEESEQATEMLVHNAQNLMQSVKETVREAEAASIKIRTDA
```

Related Information
Structure 3D structure displays
Identical Proteins - Identical proteins to 5L0C_A

Download GenPept Graphics ▼Next ▲Previous ◀Descriptions

hypothetical protein GHT09_003248 [Marmota monax]

Sequence ID: KAF7466015.1 Length: 190 Number of Matches: 1

See 1 more title(s) See all Identical Proteins(IPG)

Range 1: 15 to 190 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps
352 bits(904)	4e-122	Compositional matrix adjust.	172/176(98%)	176/176(100%)	0/176(0%)

```
Query   1     NQPVNQPILAAAQSLHQEATQWSSKGNDIIAAAKRMALLMAE
              NQPVNQPILAAAQSLH+EAT+WSSKGNDIIAAAKRMALLMAE
Sbjct   15    NQPVNQPILAAAQSLHREATKWSSKGNDIIAAAKRMALLMAE

Query   61    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
              AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS
Sbjct   75    AKDIAKASDEVTRLAKEVAKQCTDKRIRTNLLQVCERIPTIS

Query   121   ISDEESEQATEMLVHNAQNLMQSVKETVQEAEAASIKIRTDA
              ISDEESEQATEMLVHNAQNLMQSVKETV+EAEAASIKIRTDA
Sbjct   135   ISDEESEQATEMLVHNAQNLMQSVKETVREAEAASIKIRTDA
```

Related Information
Identical Proteins - Identical proteins to KAF7466015.1

Download GenPept Graphics ▼Next ▲Previous ◀Descriptions

vinculin [Ochotona curzoniae]

Sequence ID: XP_040854131.1 Length: 190 Number of Matches: 1

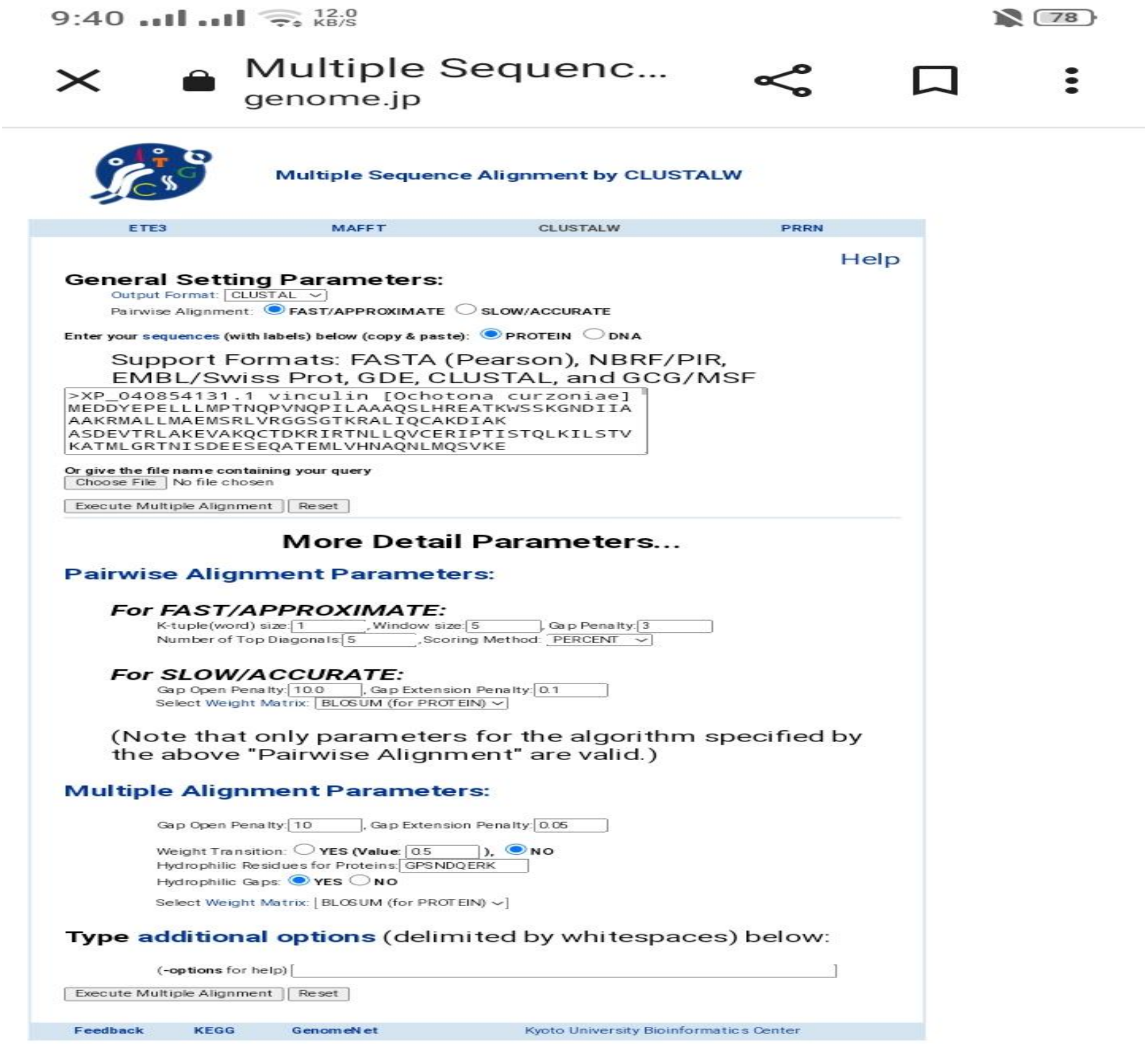
Range 1: 15 to 190 GenPept Graphics ▼Next Match ▲Previous Match

Score	Expect	Method	Identities	Positives	Gaps

Feedback

Step 3:

How to do Multiple Sequence alignments using ClustalW:



9:39 0.46 KB/S 78

CLUSTALW Result

[clustalw.aln][clustalw.dnd][readme]
Select tree menu Exec

```
 CLUSTAL 2.1 Multiple Sequence Alignments

Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: XP_040854131.1    190 aa
Sequence 2: CAA79600.1        224 aa
Sequence 3: EHH19074.1        510 aa
Sequence 4: XP_026579862.1    288 aa
Sequence 5: 3JBI_V            254 aa
Start of Pairwise alignments
Aligning...


Sequences (1:2) Aligned. Score: 98.9474
Sequences (1:3) Aligned. Score: 99.4737
Sequences (1:4) Aligned. Score: 98.4211
Sequences (1:5) Aligned. Score: 96.3158
Sequences (2:3) Aligned. Score: 97.7679
Sequences (2:4) Aligned. Score: 86.1607
Sequences (2:5) Aligned. Score: 86.1607
Sequences (3:4) Aligned. Score: 87.8472
Sequences (3:5) Aligned. Score: 87.7953
Sequences (4:5) Aligned. Score: 88.189
Guide tree file created:   [clustalw.dnd]

There are 4 groups
Start of Multiple Alignment

Aligning...
Group 1: Sequences:   2      Score:3606
Group 2: Sequences:   3      Score:3078
Group 3: Sequences:   4      Score:3750
Group 4: Sequences:   5      Score:3618
Alignment Score 12046

CLUSTAL-Alignment file created  [clustalw.aln]
```

clustalw.aln

```
CLUSTAL 2.1 multiple sequence alignment


CAA79600.1          ------------------------------------------------------------
EHH19074.1          VFDERAANFENHSGKLGATAEKAAAVGTANKSTVEGIQASVKTARELTPQVVSAARILLR
XP_040854131.1      ------------------------------------------------------------
XP_026579862.1      ------------------------------------------------------------
3JBI_V              ------------------------------------------------------------

CAA79600.1          ------------------------------------------------------------
EHH19074.1          NPGNQAAYEHFETMKNQWIDNVEKMTGLVDEAIDTKSLLDASEEAIKKDLDKCKVAMANI
XP_040854131.1      ------------------------------------------------------------
XP_026579862.1      ------------------------------------------------------------
3JBI_V              ------------------------------------------------------------

CAA79600.1          ------------------------------------------------------------
EHH19074.1          QPQMLVAGATSIARRANRILLVAKREVENSEDPKFREAVKAASDELSKTISPMVMDAKAV
XP_040854131.1      ------------------------------------------------------------
XP_026579862.1      ------------------------------------------------------------
3JBI_V              ------------------------------------------------------------

CAA79600.1          ------------------------------------------------------------
EHH19074.1          AGNISDPGLQKSFLDSGYRILGAVAKVREAFQPQEPDFPPPPPDLEQLRLTDELAPPKPP
XP_040854131.1      ------------------------------------------------------------
XP_026579862.1      ----------------------------------------------------LTDELAPPKPP
3JBI_V              ------------------------------------------------------------

CAA79600.1          ---------------------------------------------------KWSSKP-------G
EHH19074.1          LPEGEVPPPRPPPPEEKDEEFPEQKAGEVINQPMMMAARQLHDEARKWSSKP-------G
XP_040854131.1      ------------------------------------------------------------
XP_026579862.1      LPEGEVPPPRPPPPEEKDEEFPEQKAGEVLNQPMMMAARQLHDEARKWSSKPEVTGGGKE
3JBI_V              ------------GSEEKDEEFPEQKAGEAINQPMMMAARQLHDEARKWSSKP------VT

CAA79600.1          NPAAKVGIGVVAEADAADAVGFPVPSDMEDDYEPELLLMPSSQPVNQPILAAAQSLHREA
EHH19074.1          IPAAEVGIGVVAEADAADAAGFPVPPDMEDDYEPELLLMPSNQPVNQPILAAAQSLHREA
XP_040854131.1      ---------------------------MEDDYEPELLLMPTNQPVNQPILAAAQSLHREA
XP_026579862.1      APAETGPGVAREEAEEEAEAEFTLPSDMEDDYEPELLLMPTNQLVNQPILAAAQSLHREA
3JBI_V              VINEAAEAGVDIDEEDDADVEFSLPSDIEDDYEPELLLMPTNQPVNQPILAAAQSLHREA
                                        :************:_* *****************

CAA79600.1          TKWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGTKRALIQCAKDIAKASDEVTRLAKEVA
EHH19074.1          TKWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGTKRALIQCAKDIAKASDEVTRLAKEVA
XP_040854131.1      TKWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGTKRALIQCAKDIAKASDEVTRLAKEVA
XP_026579862.1      TKWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGNKRALIQCAKDIAKASDEVTRLAKEVA
3JBI_V              TKWSSKGNDIIAAAKRMALLMAEMSRLVRGGSGNKRALIQCAKDIAKASDEVTRLAKEVA
                    **********************************_**************************

CAA79600.1          KQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQN
EHH19074.1          KQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQN
XP_040854131.1      KQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQN
XP_026579862.1      KQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTSISDEESEQATEMLVHNAQN
3JBI_V              KQCTDKRIRTNLLQVCERIPTISTQLKILSTVKATMLGRTNISDEESEQATEMLVHNAQN
                    ****************************************_*******************

CAA79600.1          LMQSVKETVREAEAASIKIRTDAGFTLRWVRKTPWYQ
EHH19074.1          LMQSVKETVREAEAASIKIRTDAGFTLRWVRKTPWYQ
XP_040854131.1      LMQSVKETVREAEAASIKIRTDAGFTLRWVRKTPWYQ
XP_026579862.1      LMQSVKETVREAEAASIKIRTDAGFTLRWVRKTPWYQ
3JBI_V              LMQSVKETVREAEAASIKIRTDAGFTLRWVRK-----
                    **********************************
```

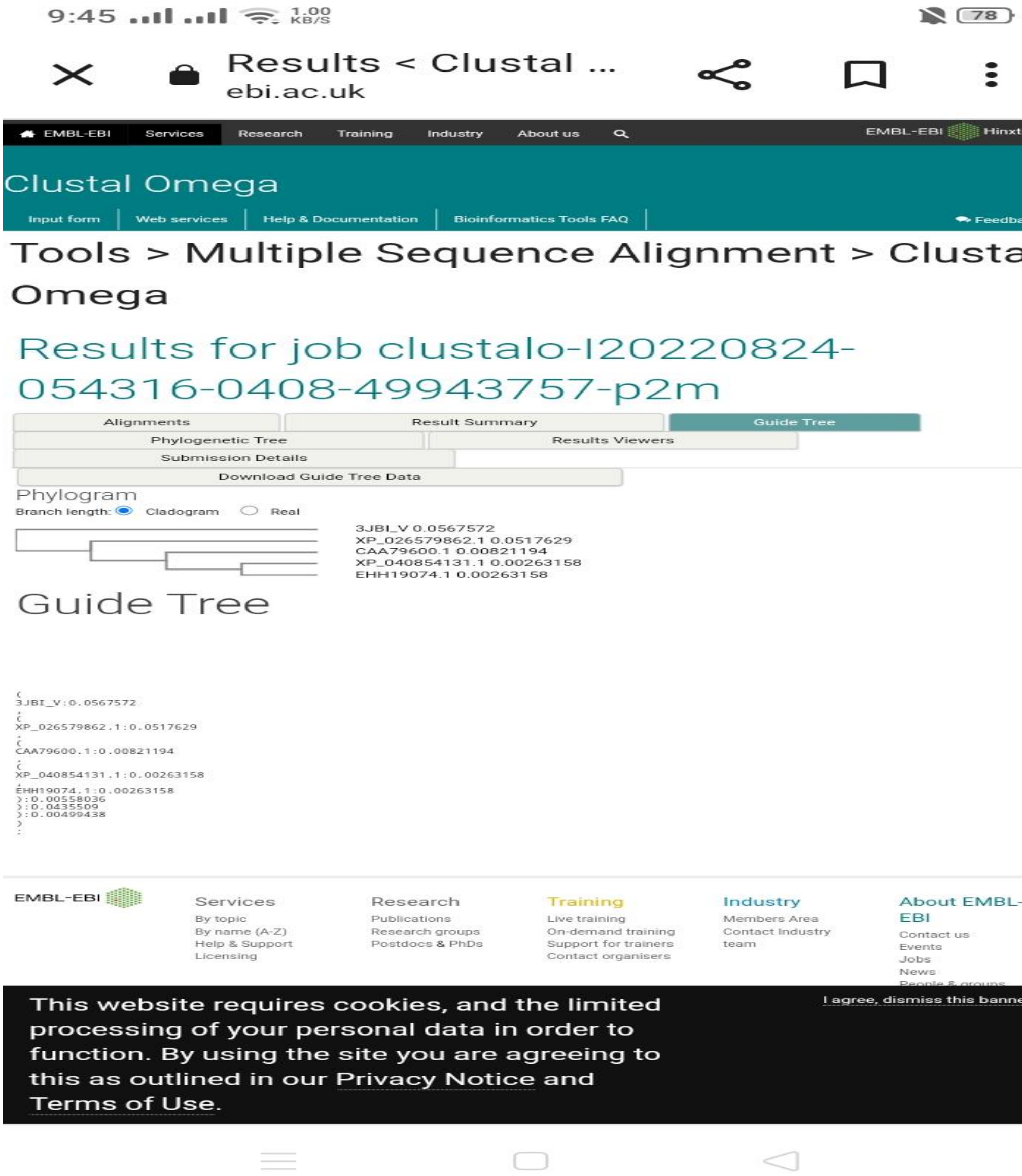
clustalw.dnd

```
(
XP_040854131.1:-0.04794,
(
CAA79600.1:0.01757,
EHH19074.1:0.00475)
:0.04467,
(
XP_026579862.1:0.05373,
3JBI_V:0.06438)
:0.01520);
```

Select tree menu Exec

Step 4:

How to develop Phylogenetic tree using Clustal Omega:



Significance of Chain B. Vinculin Protein:

Humans bearing familial or sporadic mutations in vinculin suffer from chronic, progressively debilitating DCM that ultimately leads to cardiac failure and death, whereas autosomal dominant mutations in vinculin can also provoke HCM, causing acute cardiac failure. The DCM/HCM-associated mutants of vinculin occur in the 68-residue insert unique to the muscle-specific, alternatively spliced isoform of vinculin, termed metavinculin (MV).

What I learn:

By working on this protein in different databases I have learnt how to retrieve data of protein. Now I can easily obtain different data sets of any protein like its i.d accession number its sequence in different formats, how to compare sequences using blast and how to obtain multiple sequence alignment using clustalw or clustal omega and to develop phylogenetic trees. We can retrieve functions by using pumbed data base

Regards:

VU Expert Zoologists: