

2/15/2020

The alignment I selected was for an alignment to the lutra lutra chromosome 13 assembly, otherwise known as the eurasian otter.

<input checked="" type="checkbox"/>	Lutra lutra genome assembly, chromosome: 13	375	1056	54%	4e-99	87.21%	LR738415.1
-------------------------------------	---	-----	------	-----	-------	--------	----------------------------

Lutra lutra genome assembly, chromosome: 13

Sequence ID: [LR738415.1](#) Length: 95726014 Number of Matches: 6

Range 1: 47865220 to 47865524 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
375 bits(415)	4e-99	266/305(87%)	0/305(0%)	Plus/Minus
Query 454		CAGGTCATGATGATGGGCAGCGCCCGAGTGGCGGAGCTGCTGCTGCTCCACGGCGCGGAG		513
Sbjct 47865524		CAGGTCATGATGATGGGCAGCACCCGCGTGGCGGAGCTGCTGCTGCTCCACGGCGCCGAT		47865465
Query 514		CCCAACTGCGCCGACCCCGCCACTCTCACCAGACCCGTGCACGACGCTGCCCGGGAGGGC		573
Sbjct 47865464		CCCAACTGCGCGGACCCCATCACTCTCACCAGACCTGTGCACGACGCGGCCCGGGAGGGC		47865405
Query 574		TTCCTGGACACGCTGGTGGTGTGTCACCGGGCCGGGGCGCGGCTGGACGTGCGCGATGCC		633
Sbjct 47865404		TTCCTGGACACGCTCCTCGTGTGTCACCGAGCCGGGGCGCAGCTGGACGTGCGCGATGCC		47865345
Query 634		TGGGGCCGTCTGCCCGTGGACCTGGCTGAGGAGCTGGGCCATCGCGATGTGCGACGGTAC		693
Sbjct 47865344		TGGGGCCGCCTGCCCGTGGACCTGGCTGAGGAGCGGGGCCACCGCGCTGTGCGCCCGGTAC		47865285
Query 694		CTGCGCGCGGCTGCGGGGGGACCCAGAGGCAGTAACCATGCCCGCATAGATGCCGCGGAA		753
Sbjct 47865284		CTGGGCGCAGCCTCGGGGGCCACCGAAGGTGGTAGCCACTCCCATACGGAGGTGGCGGAA		47865225
Query 754	GGTCC	758		
Sbjct 47865224	GGTCC	47865220		

Q3

Fasta Sequence:

```
LR738415.1:47865220-47865524 Lutra lutra genome assembly, chromosome: 13 GGAC-
CTTCCGCCACCTCCGTATGGGAGTGGCTACCACTTCGGTGGCCCCCGAGGCT-
GCGCCCAGGTACCGGGGCGACAGC GCGGTGGCCCCGCTCCTCAGCCAGGTCCACGGGCAGGCG-
GCCCCAGGCATCGCGCACGTCCAGCTGCGCCCCGGCTCGGT GCAACACGAG-
GAGCGTGTCCAGGAAGCCCTCCCGGGCCGCGTCTGTGCACAGGTGCGGTGAGAGTGATGGGGTCCGCGCA
TTGGGATCGGCGCCGTGGAGCAGCAGCAGCTCCGCCACGCGGGTGTGCCCCATCAT-
CATGACCTG
```

Emboss translations of above fasta

```
EMBOSS_001_1 GPSATSVWEWLPPSVAPEAAPRYRATARWPRSSARSTGRRPQASRTSS-
CAPARCNTRSVS RKPSRAASCTGRVRVMGSAQLGSAPWSSSSSATTRVLPIMTX EM-
BOSS_001_2 DLPPPPYGSYHLRWPPRLRPGTGRQRGGPAPQPGPRAGGPRHRARPAAPRL-
GATRGACP GSPPGPRRAQVGEWGPSRWDRRRGAAAAPPRGCCPSSPX EMBOSS_001_3
TFRHLRMGVATTFGGPRGCAQVPGDSAVAPLLSQVHGQAAPGIAHVQLRPGSVQHEERVQ
EALPGRVVHRSGESDGVRAVGIGAVEQQQLRHAGAAHHDDL EMBOSS_001_4 GHD-
DGQHPRGGAAAAPRRRSQLRGPHHSHTPCARRGPGGLPGHAPRVAPSRGAAGRARCL
GPPARGPGGAGPPRCRPVPGRSLGGHRRW*PLPYGGGGRS EMBOSS_001_5 RS**WAAPAWRSCCSTAPIPT
GAACPWTWLRSGATALSPGTWAQPRGPPKVATPIRRWRKVX EMBOSS_001_6 QVM-
MMGSTRVAELLLHGADPNCADPITLTPVHDAAREGFLDTLLVLHRAGAQLDVRDA
WGRLPVDLAEERGHRAVARYLGAASGATEGGSHSHTTEVAEGP
```

EMBOSS_001_6 had the longest ORF, with the following AA sequence:

Otter_chrom13_putativeCDKN2A MMMGSTRVAELLLLHGADPNCADPITLTRPVH-
 DAAREGFLDTLLVLHRAGAQLDVRDA WGRLPVDLAEERGHRAVARYLGAASGATEGGSHSHT-
 VAEG

Q4

To determine if my protein is novel, I performed blastp on the following AA sequence:

Otter_chrom13_putativeCDKN2A MMMGSTRVAELLLLHGADPNCADPITLTRPVH-
 DAAREGFLDTLLVLHRAGAQLDVRDA WGRLPVDLAEERGHRAVARYLGAASGATEGGSHSHT-
 VAEG

Job Title **Protein Sequence**

RID [4SNXVU9D014](#) Search expires on 02-20 04:31 am [Download All](#) ▼

Program BLASTP [Citation](#) ▼

Database nr [See details](#) ▼

Query ID lcl|Query_17486

Description None

Molecule type amino acid

Query Length 99

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](#)

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments [Download](#) ▼ [Manage Columns](#) ▼ Show ?

☒ select all 100 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	cyclin-dependent kinase inhibitor 2A-like isoform X1 [Enhydra lutris kenyon]	195	195	98%	3e-62	97.96%	XP_022378230.1
<input checked="" type="checkbox"/>	cyclin-dependent kinase inhibitor 2A-like isoform X2 [Enhydra lutris kenyon]	195	195	98%	4e-62	97.96%	XP_022378231.1
<input checked="" type="checkbox"/>	cyclin-dependent kinase inhibitor 2A-like isoform X3 [Mustela erminea]	189	189	100%	2e-60	93.94%	XP_032164685.1
<input checked="" type="checkbox"/>	cyclin-dependent kinase inhibitor 2A-like isoform X4 [Mustela erminea]	190	190	100%	3e-60	93.94%	XP_032164689.1

Figure 2: Screenshot of my blastp results

The top result has a 97.69% similarity to the CDKN2A like gene from the sea otter, so yes we have identified a novel variant of CDKN2A in the eurasian otter genome!

Q5

My original protein:

NP_000068.1 cyclin-dependent kinase inhibitor 2A isoform p16INK4a [Homo sapiens]
 MEPAAGSSMEPSADWLATAAARGRVEEVRLLEAGALPNAPNSYGRRIQVMM-
 MGSARVAELLLLHGAEP NCADPATLTRPVHDAAREGFLDTLLVLHRAGARLDVRDAW-
 GRLPVDLAEELGHRDVARYLRAAAGGTRGS NHAIDAAEGPSDIPD

My novel protein:

Otter_chrom13_putativeCDKN2A MMMGSTRVAELLLLHGADPNCADPITLTRPVH-
 DAAREGFLDTLLVLHRAGAQLDVRDA WGRLPVDLAEERGHRAVARYLGAASGATEGGSHSHT-
 VAEG

Members of this family in other species:

NP_001139762.1 cyclin-dependent kinase inhibitor 2A [Pan troglodytes] MEPAAGSSMEP-SADWLATAAAARGRVEEVRLLEAGALPNAPNSYGRRPIQVMMMGSARVAELLLLHGAEP NCADPATLTRPVHDAAREGFLLDTLVVLHRAGARLDVRDAWGRLPVDLAEELGHRD-VARYLRAAAGGTRGS NHARIDAAEGPLDIPD

NP_001277177.1 cyclin-dependent kinase inhibitor 2A [Felis catus] MEPLADRLATAAAR-GRAEEVRALLAAGAQPAPNRLGRSPIQVMMMGSARVAELLLLHGADPNCADPATL TRPVHDAAREGFLLDTLVVLHRAGARLDVRDAWGRLPVDLAEERGHRRDIVRYLRAATG-GTSGSHTGTDGA EGVADSRT

NP_031696.1 cyclin-dependent kinase 4 inhibitor B [Mus musculus] MLGGSSDAGLATAAARGQVETVRQL-LEAGADPNALNRFGRRIQVMMMGSQAQVAELLLLHGAEPNCADPA TLTRPVHDAAREGFLLDTLVVLHRA-GARLDVCDAWGRLPVDLAEEQGHRDIARYLHAATGD

NP_570825.1 cyclin-dependent kinase 4 inhibitor B [Rattus norvegicus] MLGGGS-DAGLATAAARGQVETVRQLLEAGADPNAVNRFGRRIQVMMMGSQAQVAELLLLH-GAEPNCADPA TLTRPVHDAAREGFLLDTLMVLHKAGARLDVCDAWGRLPVDLAEEQGHRDIARYLHAATGD

NP_001069362.1 cyclin-dependent kinase 4 inhibitor B [Bos taurus] MLSGGGGDAD-LANAAARGQVEAVRQLLEAGVDPNRLNRFGRRIQVMMMGSARVAELLLLHGADPNCADP ATLTRPVHDAAREGFLLDTLVALHRAGGRLDVRDAWGRLPVDLAEERGHRRDVARYL-RATAGD

PNJ81811.1 CDKN2A isoform 6 [Pongo abelii] MEPSADWLATAAARGRVEVRALLEA-GALPNAPNSYGRRPIQVMMMGSARVAELLLLHGAEPNCADPATL TRPVHDAAREGFLLDT-LAVLHRAGARLDVRDAWGRLPVDLAEELGHRDVARYLRAAAGGTRGSNHARIDAA EGPSEMIGNYLWVCRSRHA

NP_001268314.1 cyclin-dependent kinase inhibitor 2A [Mesocricetus auratus] MEP-SADGLARAAAQGREQEVRLLEAGVSPNAPNCFGRTPIQVMMMGNTQVARLLL-LYGAEPNCEDPATL SRPVHDAAREGFLETLAILHQAGARLDVLDARGRLPVDLA-LERGHCDVVQYLRAAGNTPQGSEPAGVTSA QTPPEVSDFADHPLGPY

NP_999289.1 cyclin-dependent kinase 4 inhibitor B [Sus scrofa] MLSGGGGDAGLANAAARGQVETVRQL-LEAGADPNGLNHFGRRIQVMMMGSARVAELLLLHGADPNCADP ATLTRPVHDAAREGFLLDTL-VALRRAGARLDVQDAWGRLPVDLAEERGHRRDVARFLRAAAGD

NP_001270542.1 cyclin-dependent kinase 4 inhibitor B [Macaca fascicularis] MREENKGMPSGGGS-DEGLTSAAARGLVEKVRQLLEAGADPNVNRFGRRAIQVMMMGSARVAELLLLHGA EPNCADPATLTRPVHDAAREGFLLDTLVVLHRAGARLDVRDAWGRLPVILAEERGHRRD-VAGYLRAATGD

I performed multiple sequence alignment with the clustal Omega tool provided by EBI and got the following

CLUSTAL O(1.2.4) multiple sequence alignment

```

Golden_Hamster      -----MEPSADGLARAAQGREQEVRLLEAGVSPNAPNCFGRTPIQVMMMGNTQ  50
Mouse               -----MLGGSSDAGLATAAARGQVETVRQLEAGADPNALNRFGRRP IQVMMMGSAQ  52
Rat                 -----MLGGSSDAGLATAAARGQVETVRQLEAGADPNVNRFGRRP IQVMMMGSAQ  52
Macaque             MREENKMPSPGGGSDGLTSAARGLVKVRQLEAGADPNGVNRFGRRAIQVMMMGSA  60
Cow                 -----MLSGGGGDADLANAAARGQVEAVRQLEAGVDPNRLNRFGRRP IQVMMMGSA  53
Wild_boar           -----MLSGGGGDAGLANAAARGQVETVRQLEAGADPNLNFGRRP IQVMMMGSA  53
Chimpanzee          --MEPAAGSSMEPSADWLATAAARGRVEEVRALLEAGALPNAPNSYGRRP IQVMMMGSA  58
Orangutan           -----MEPSADWLATAAARGRVDEVRLLEAGALPNAPNSYGRRP IQVMMMGSA  50
Otter_chrom13_putativeCDKN2A -----MMMGSTR  7
Cat                 -----MEPLADRLATAAARGRAEEVRALLAAGAQPAPNRLGRSPIQVMMMGSA  50
                                     ****.::

Golden_Hamster      VARLLLLYGAEPNCEDPATLSRPVHDAAREGFLETAILHQAGARLDVLDARGRLPVDLA  110
Mouse               VAELLLLHGAEPNCADPATLTRPVHDAAREGFDTLVVLHRAGARLDVCDAWGRLPVDLA  112
Rat                 VAELLLLHGAEPNCADPATLTRPVHDAAREGFDTLMVLHKAGARLDVCDAWGRLPVDLA  112
Macaque             VAELLLLHGAEPNCADPATLTRPVHDAAREGFDTLVVLHRAGARLDVDRDAWGRLPVILA  120
Cow                 VAELLLLHGADPNCADPATLTRPVHDAAREGFDTLVALHRAGGRLDVDRDAWGRLPVDLA  113
Wild_boar           VAELLLLHGADPNCADPATLTRPVHDAAREGFDTLVALHRAGARLDVQDAWGRLPVDLA  113
Chimpanzee          VAELLLLHGAEPNCADPATLTRPVHDAAREGFDTLVVLHRAGARLDVDRDAWGRLPVDLA  118
Orangutan           VAELLLLHGAEPNCADPATLTRPVHDAAREGFDTLAVLHRAGARLDVDRDAWGRLPVDLA  110
Otter_chrom13_putativeCDKN2A VAELLLLHGADPNCADPITLTRPVHDAAREGFDTLLVLHRAGAQLDVRDAWGRLPVDLA  67
Cat                 VAELLLLHGADPNCADPATLTRPVHDAAREGFDTLVVLHRAGARLDVDRDAWGRLPVDLA  110
                                     **.****:*** ** **:******:*** *:;:.*:*** ** ***** **

Golden_Hamster      LERGHCDVVQYLRAAGNTPQGSEPAGVTSAQTPPEVSDFDHPLGPY-----  157
Mouse               EEQGHARDIARYLHAATGD-----  130
Rat                 EEQGHARDIARYLHAATGD-----  130
Macaque             EERGHARDVAGYLRAATGD-----  138
Cow                 EERGHARDVARYLRATAGD-----  131
Wild_boar           EERGHARDVARFLRAAAGD-----  131
Chimpanzee          EELGHARDVARYLRAAAGGTRGSNHARIDAAEGPLDI-----PD-----  156
Orangutan           EELGHARDVARYLRAAAGGTRGSNHARIDAAEGPSEM-----IGNYLWVCRSRHA  159
Otter_chrom13_putativeCDKN2A EERGHRAVARYLGAASGATEGGSHSHTVAEG-----  99
Cat                 EERGHARDIVRYLRAATGGTSGSHTGTGDAEGVADS-----RT-----  148
                                     * * * :. : * * : .

```

alignment:

Q6

Phylogenetic Tree

This is a Neighbour-joining tree without distance corrections.

Branch length: ☒ Cladogram ☐ Real

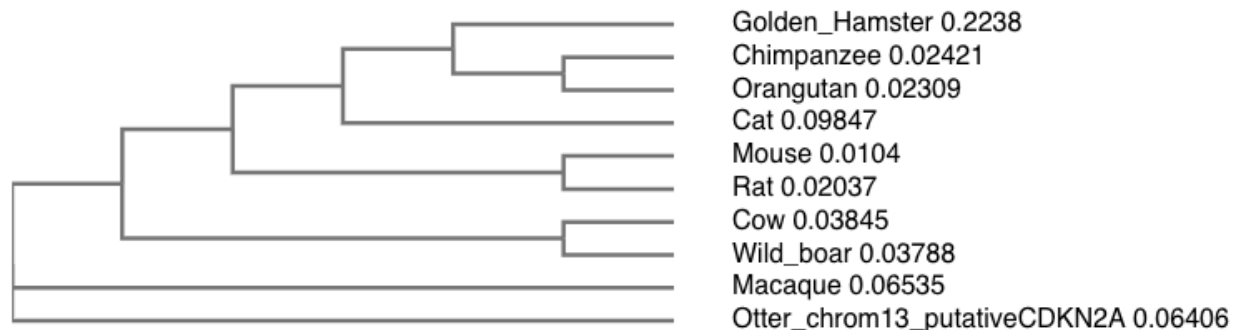


Figure 3: Alignment phylogenetic tree

The clustal alignment I performed in Q5 also provided the above phylogenetic tree

Q7

Loading in input clustal alignment file

```
library(bio3d)
otter_aln <- read.fasta(file = "~/Documents/UCSD_Neuro_Coursework/bggn_213/bio3d_aln_input.fasta")
otter_aln
```

```
## Warning in conserv(ali, method = "entropy10"): non standard residue code: mapped to X
##
```

```
## Warning in conserv(ali, method = "identity"): non standard residue code: mapped to X
##
```

```
##          1          .          .          .          .          50
## [Truncated_Name:1] Golden_Ham -----MEPSADGLARAAAQGREQEVRLLEAGVSPNAPNCFGRTP
## [Truncated_Name:2] Mouse      -----MLGGSSDAGLATAAARGQVETVRQLLEAGADPNALNRFGRRP
## [Truncated_Name:3] Rat        -----MLGGSSDAGLATAAARGQVETVRQLLEAGADPNVNRFGRRP
## [Truncated_Name:4] Macaque    MREENKGMPSGGGSDEGLTSAAGRLVEKVRQLLEAGADPNGVNRFGRRA
## [Truncated_Name:5] Cow        -----MLSGGGGDADLANAAARGQVEAVRQLLEAGVDPNRLNRFGRRP
## [Truncated_Name:6] Wild_boar  -----MLSGGGGDAGLANAAARGQVETVRQLLEAGADPNGLNHFGRRP
## [Truncated_Name:7] Chimpanzee --MEPAAGSSMEPSADWLATAAARGRVEEVRLLEAGALPNAPNSYGRRP
## [Truncated_Name:8] Orangutan  -----MEPSADWLATAAARGRVDEVRLLEAGALPNAPNSYGRRP
## [Truncated_Name:9] Otter_chro -----
## [Truncated_Name:10] Cat       -----MEPLADRLATAAARGRAEEVRALLAAGAQPAPNRLGRSP
##
##          1          .          .          .          .          50
##
##          51          .          .          .          .          100
## [Truncated_Name:1] Golden_Ham IQVMMMGNTQ  VARLLLLYGAEPNCDPATLSRPVHDAAREGFLETLAIL
## [Truncated_Name:2] Mouse      IQVMMMGSAQ  VAELLLLHGAEPNCADPATLTRPVHDAAREGFDLTLVVL
## [Truncated_Name:3] Rat        IQVMMMGSAQ  VAELLLLHGAEPNCADPATLTRPVHDAAREGFDLTMLVL
## [Truncated_Name:4] Macaque    IQVMMMGSAQ  VAELLLLHGAEPNCADPATLTRPVHDAAREGFDLTLVVL
## [Truncated_Name:5] Cow        IQVMMMGSAQ  VAELLLLHGADPNCADPATLTRPVHDAAREGFDLTLVAL
## [Truncated_Name:6] Wild_boar  IQVMMMGSAQ  VAELLLLHGADPNCADPATLTRPVHDAAREGFDLTLVAL
## [Truncated_Name:7] Chimpanzee IQVMMMGSAQ  VAELLLLHGAEPNCADPATLTRPVHDAAREGFDLTLVVL
## [Truncated_Name:8] Orangutan  IQVMMMGSAQ  VAELLLLHGAEPNCADPATLTRPVHDAAREGFDLTLAVL
## [Truncated_Name:9] Otter_chro ---MMMGSTR  VAELLLLHGADPNCADPITLTRPVHDAAREGFDLTLVL
## [Truncated_Name:10] Cat       IQVMMMGSAQ  VAELLLLHGADPNCADPATLTRPVHDAAREGFDLTLVVL
##
##          ****      ***  ****  **^***  **  **^*****^**  *
##          51          .          .          .          .          100
##
##          101         .          .          .          .          150
## [Truncated_Name:1] Golden_Ham HQAGARLDVLDARGRLPVDLA  LERGHCDVVQYLRAAGNTPQGSEPAVGT
## [Truncated_Name:2] Mouse      HRAGARLDVCDAWGRLPVDLA  EEQGHRDIARYLHAATG-----
## [Truncated_Name:3] Rat        HKAGARLDVCDAWGRLPVDLA  EEQGHRDIARYLHAATG-----
## [Truncated_Name:4] Macaque    HRAGARLDVRDAWGRLPVILA  EERGHDRVAGYLRATG-----
## [Truncated_Name:5] Cow        HRAGGRLDVRDAWGRLPVDLA  EERGHDRVARYLRATAG-----
```

```

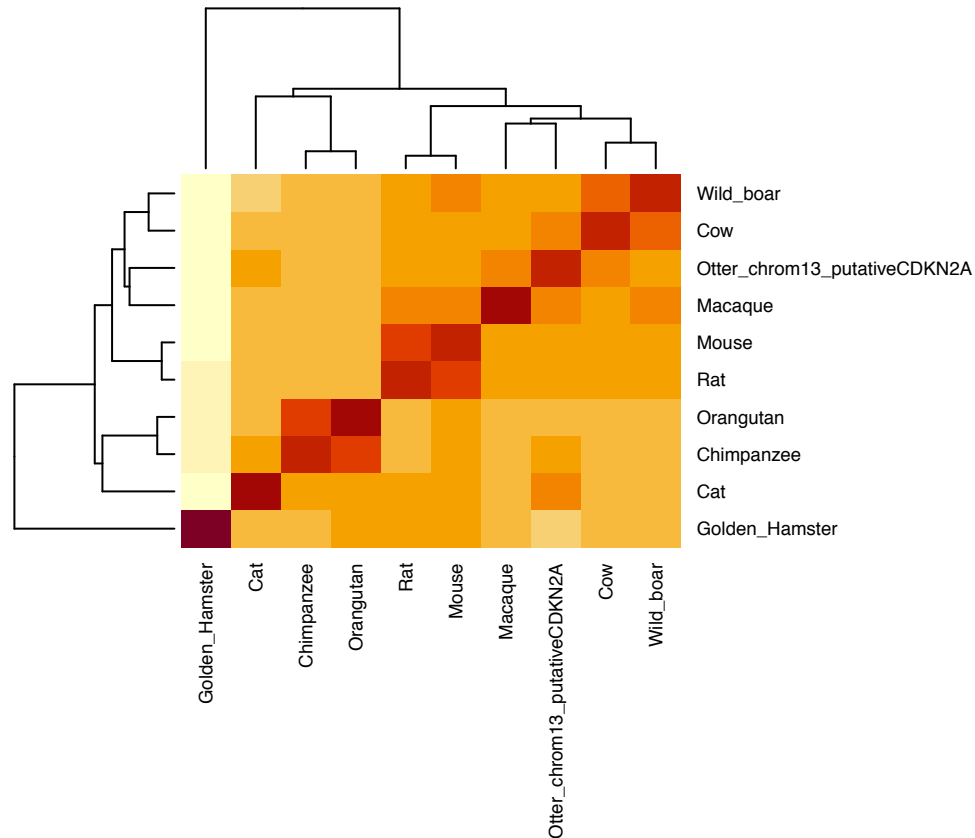
## [Truncated_Name:6]Wild_boar   RRAGARLDVQDAWGRLPVDLA EERGHRDVARFLRAAAGD-----
## [Truncated_Name:7]Chimpanzee  HRAGARLDVRDAWGRLPVDLA EELGHRDVARYLRAAAGGTRGSNHARID
## [Truncated_Name:8]Orangutan   HRAGARLDVRDAWGRLPVDLA EELGHRDVARYLRAAAGGTRGSNHARID
## [Truncated_Name:9]Otter_chro  HRAGAQLDVRDAWGRLPVDLA EERGHRAVARYLGAASGATEGGSHSHT
## [Truncated_Name:10]Cat        HRAGARLDVRDAWGRLPVDLA EERGHRRDIVRYLRAATGGTGSGSHTGTD
##                               ^  **^  ***  **  *****  ***  *  **  ^  ^*  *
##                               101      .      .      .      .      150
##
##                               151      .      .      179
## [Truncated_Name:1]Golden_Ham  SAQTPPEVSDFADHPLGPY-----
## [Truncated_Name:2]Mouse       -----
## [Truncated_Name:3]Rat         -----
## [Truncated_Name:4]Macaque     -----
## [Truncated_Name:5]Cow         -----
## [Truncated_Name:6]Wild_boar   -----
## [Truncated_Name:7]Chimpanzee  AAEGPLDI-----PD-----
## [Truncated_Name:8]Orangutan   AAEGPSEM-----IGNYLWVCRSRHA
## [Truncated_Name:9]Otter_chro  VAEG-----
## [Truncated_Name:10]Cat        GAEGVADS-----RT-----
##                               *
##                               151      .      .      179
##
## Call:
##   read.fasta(file = "~/Documents/UCSD_Neuro_Coursework/bggn_213/bio3d_aln_input.fasta")
##
## Class:
##   fasta
##
## Alignment dimensions:
##   10 sequence rows; 179 position columns (88 non-gap, 91 gap)
##
## + attr: id, ali, call

```

```

seq_iden <- seqidentity(otter_aln)
heatmap(seq_iden, margins = c(10,6), cexCol = .75, cexRow = .75)

```



Q8

Creating consensus sequence alignment

```
otter_consensus <- consensus(otter_aln)
```

```
## Warning in consensus(otter_aln):
## non standard residue code:   mapped to X
```

#uh oh cant seem to connect to the server! Getting the following error when I run blast.pdb(otter_consensus):

[illegible]

Going to do it manually on ncbi blast:

```
x <- otter_consensus$seq
write.fasta(seqs = x, file = "otter.fasta")
```

Top three hits:

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
✓	Chain B, Mechanism Of G1 Cyclin Dependent Kinase Inhibition From The Structure Of The Cdk6-P16ink4a Tumor Suppressor Complex [Homo sapiens]	159	159	97%	2e-51	87.83%	1BI7_B
✓	Chain A, TUMOR SUPPRESSOR P16INK4A [Homo sapiens]	159	159	97%	2e-51	87.83%	1A5E_A
✓	Chain A, Tumor Suppressor P15(Ink4b) Structure By Comparative Modeling And Nmr Data [Mus musculus]	155	155	97%	4e-50	83.48%	1D9S_A
✓	Chain A, Structure Of Cdk Inhibitor P19Ink4d [Homo sapiens]	66.2	66.2	79%	1e-14	52.27%	1BD8_A

Figure 4: Top blastp hits

Picking hits 1,3 and 4 from the above blast, since hit 2 is the same protein from hit 1

```
pdb.1BI7_B <- read.pdb("1BI7")
```

```
## Note: Accessing on-line PDB file
```

```
pdb.1D9S_A <- read.pdb("1D9S")
```

```
## Note: Accessing on-line PDB file
```

```
pdb.1BD8_A <- read.pdb("1BD8")
```

```
## Note: Accessing on-line PDB file
```

```
ano_1BI7 <- pdb.annotate("1BI7")
```

```
ano_1D9S <- pdb.annotate("1D9S")
```

```
ano_1BD8 <- pdb.annotate("1BD8")
```

1BI7_B, id: 1BI7, structure method: X-RAY DIFFRACTION, resolution: 3.4, source: Homo Sapiens

1D9S_A, id: 1D9S, structure method: SOLUTION NMR, resolution: N/A, source: Mus mus

1BD8_A, id: 1BD8, structure method: X-RAY DIFFRACTION , resolution: 1.8, source: Homo Sapiens

Q9

Loading the pdb structure 1BI7 into VMD, and outputting the following structure:

My consensus sequence has 87% identity and 97% query coverage with the above structure, so I assume this is likely to be very similar in structure to my novel otter protein.

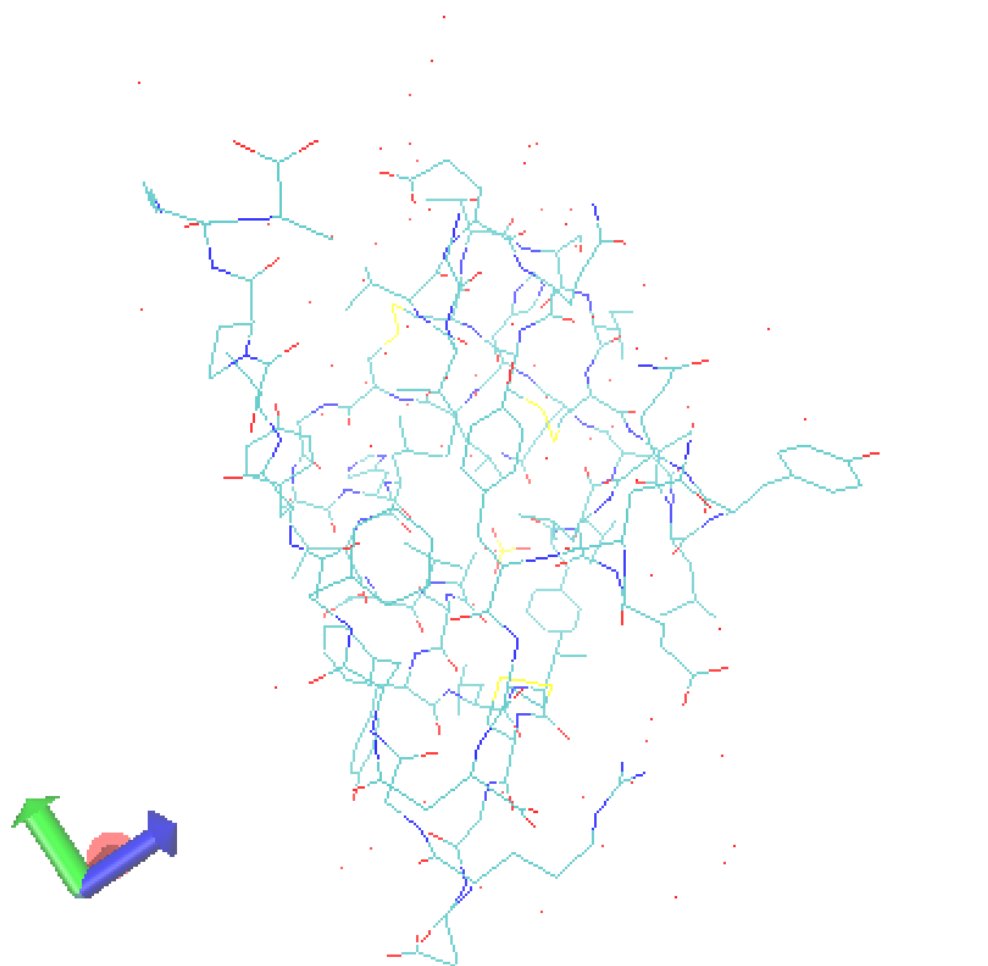


Figure 5: 1B17 Structure

Q10

Ligand Efficiency Plot for novel sequence:

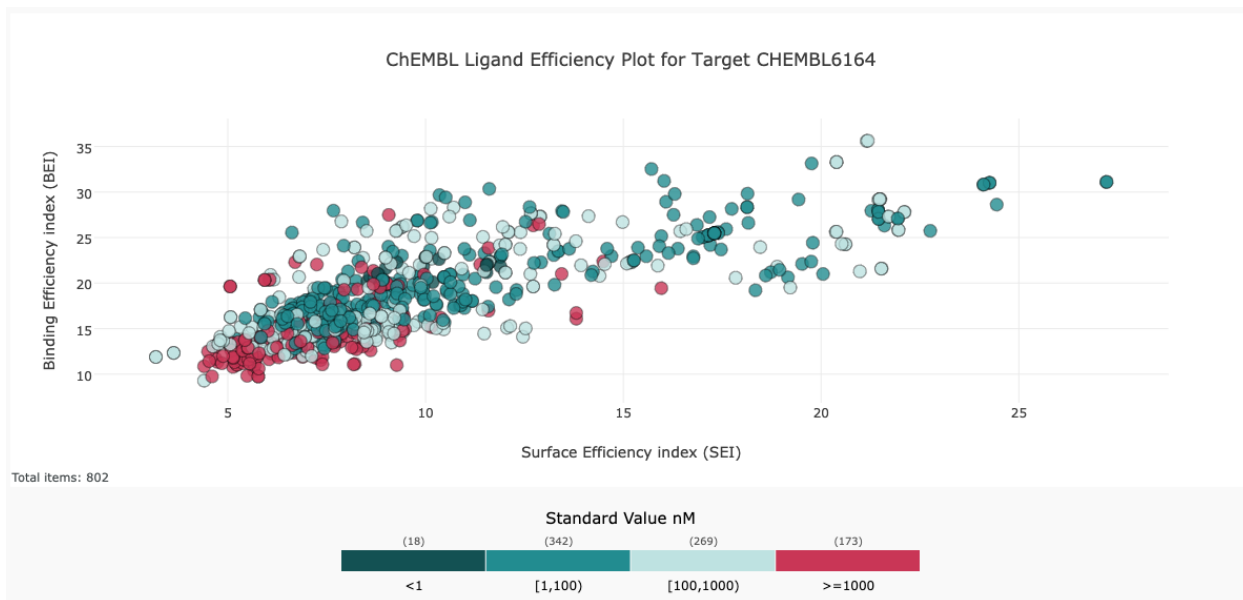


Figure 6: ChEMBL Ligand Efficiency for Novel Otter Protein

One example ligand to try would be: 6(5H)-Phenanthridinone

For target assays we might explore the function of this protein through it's target TNKS1

<input type="checkbox"/> ChEMBL ID	<input type="checkbox"/> Assay Type	<input type="checkbox"/> Description	<input type="checkbox"/> Organism	<input type="checkbox"/> Compounds	<input type="checkbox"/> Document ChEMBL ID	<input type="checkbox"/> BAO Format	<input type="checkbox"/> Source
<input type="checkbox"/> CHEMBL4012895	B	Inhibition of TNKS1 (unknown origin) expressed in HEK293 cells co-expressing TCF/LEF luciferase plasmid assessed as reduction in wnt3a ligand-induced Wnt-dependent increase in beta-catenin preincubated for 1 hr followed by wnt3a ligand stimulation for 24 hrs by luciferase reporter gene assay	Homo sapiens	37	CHEMBL4011625	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4003521	B	Inhibition of recombinant human N-terminal GST-tagged TNKS1 (1001 to 1327 residues) expressed in baculovirus infected sf9 cells using TACS-Sapphire as substrate measured after 30 mins by colorimetric assay	Homo sapiens	7	CHEMBL4002591	cell-based format	Scientific Literature
<input type="checkbox"/> CHEMBL4271456	B	Inhibition of N-terminal GST-tagged human TNKS1 expressed in Sf9 cells incubated for 1 hr by histone ribosylation assay	Homo sapiens	1	CHEMBL4270509	cell-based format	Scientific Literature

Figure 7: Published Potential Target Assays for Novel Otter Protein

And tadaaaaaaaa~~~~~ There's my Find a Gene Project! BGGN213 Winter 2020. Thanks for the good quarter! Learned a lot :)