

Assignment 4

Step 1


Species	Gene	Architecture	aa
<input type="checkbox"/> <i>Bos taurus</i> domestic cattle	CSN2 casein beta		224
RefSeq transcripts (2)	RefSeq proteins (2)	Architecture	aa
XM_010806178.3	XP_010804480.3		224
XM_015471671.3	XP_015327157.2		216
Genome Viewer			

Figure 1: CSN2 orthologs for vertebrates (*XM_010806178.3*, *cattle.txt*)







<input type="checkbox"/> <i>Mus musculus</i> house mouse	Csn2 casein beta		231
RefSeq transcripts (7)	RefSeq proteins (6)	Architecture	aa
NM_009972.2	NP_034102.1		231
NM_001286020.1	NP_001272949.1		230
NM_001286021.1	NP_001272950.1		230
NM_001286022.1	NP_001272951.1		222
NM_001286023.1	NP_001272952.1		216
NM_001286024.1	NP_001272953.1		215
XR_389259.2			
Genome Viewer InterPro			

Figure 2: CSN2 orthologs for house mouse (*NM_001286022.1*, *housemouse.txt*)


<input type="checkbox"/> <i>Canis lupus familiaris</i> dog	CSN2 casein beta		250
RefSeq transcripts (5)	RefSeq proteins (5)	Architecture	aa
NM_001003086.1	NP_001003086.1		250
XM_038684515.1	XP_038540443.1		250
XM_038684516.1	XP_038540444.1		250
XM_038684517.1	XP_038540445.1		250
XM_038684518.1	XP_038540446.1		212
Genome Viewer InterPro			

Figure 3: CSN2 orthologs for dog (*XM_038684516.1*, *dog.txt*)

<input type="checkbox"/>	<i>Balaenoptera musculus</i> blue whale	CSN2 casein beta	238	⤴
RefSeq transcripts (2)		RefSeq proteins (2)	Architecture	aa
XM_036853046.1		XP_036708941.1		238
XM_036853045.1		XP_036708940.1		224
Genome Viewer				

Figure 4: CSN2 orthologs for blue whale (*XM_036853045.1*, *bluewhale.txt*)

<input type="checkbox"/>	<i>Phocoena phocoena</i> harbor porpoise	CSN2 casein beta	259	⤴
RefSeq transcripts (1)		RefSeq proteins (1)	Architecture	aa
XM_065878181.1		XP_065734253.1		259
Genome Viewer				

Figure 5: CSN2 orthologs for harbor porpoise (*XM_065878181.1*, *harborporpoise.txt*)

Step 2

(a)

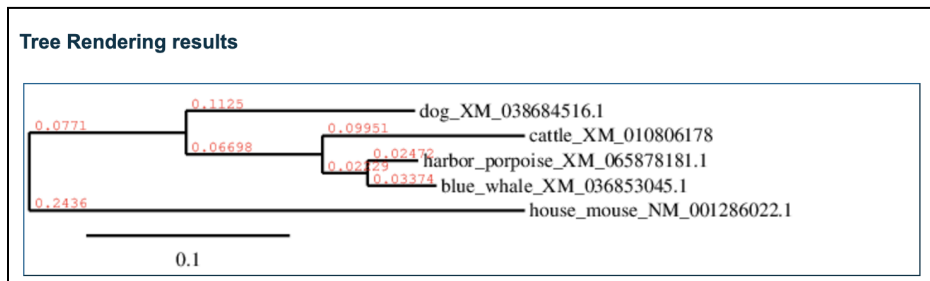


Figure 6: Phylogram with branch lengths (*saha_dibbyo_assignment4_phylogram_step2.png*)

Note: Harbor porpoise and blue whale are more closely related to each other than dog, cattle, and house mouse

(b) The marine mammals (harbor porpoise and blue whale) are more closely related to the dog than other mammals in the tree because of the shorter distance between the branches.

(c) Total distance from blue whale to dog = $0.03374 + 0.02229 + 0.06698 = 0.12$ (rounded to 2 decimal places).

Step 3

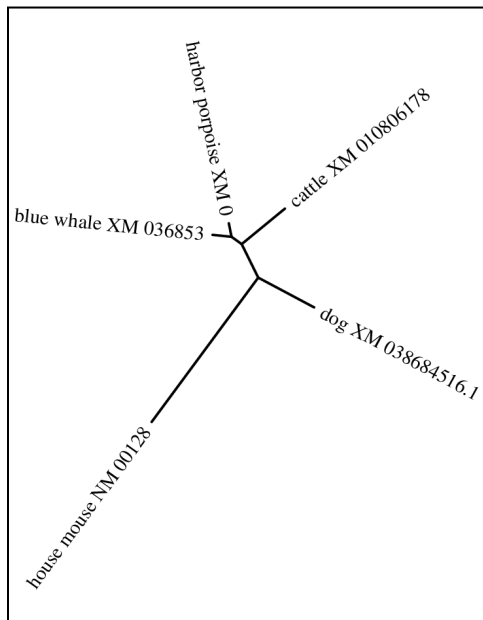


Figure 7: Radial tree (*saha_dibbyo_assignment4_radial_step3.png*)

Step 4

<input type="checkbox"/>	Hippopotamus amphibius kiboko	CSN2 casein beta	311	^
RefSeq transcripts (1)	RefSeq proteins (1)	Architecture	aa	
XM_057726918.1	XP_057582901.1		311	
Genome Viewer				

Figure 8: CSN2 orthologs for hippopotamus (*XM_057726918.1*, *hippopotamus.txt*)

<input type="checkbox"/>	Homo sapiens human	CSN2 casein beta	226	^
RefSeq transcripts (5)	RefSeq proteins (3)	Architecture	aa	
NM_001891.4	NP_001882.1		226	
NM_001302770.2	NP_001289699.1		225	
NM_001385731.1	NP_001372660.1		211	
XR_001756926.1				
XR_001756927.1				
Genome Viewer InterPro				

Figure 9: CSN2 orthologs for human (*NM_001302770.2*, *human.txt*)

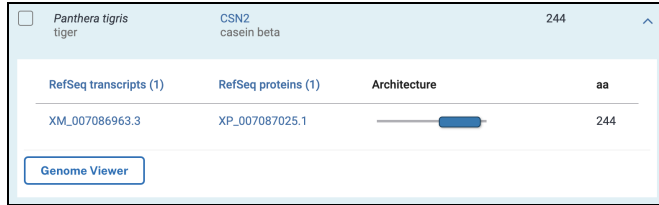


Figure 10: CSN2 orthologs for tiger (NM_001302770.2, tiger.txt)

(a)

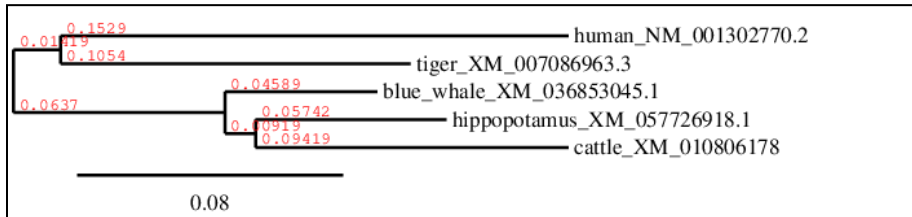


Figure 11: Phylogram with branch lengths (saha_dibbyo_assignment4_phylogram_step4.png)

(b) Hippopotamus is the closest to the whale because of the shortest distance of 0.12 between the branch lengths as opposed to 0.17 for whale to tiger, 0.17 for whale to cattle, and 0.21 for whale to human. Please note that values have been rounded to 2 decimal places.

Step 5

```
>human_NM_001302770.2
ATGAAGTCTCTCATCTTGGCTGCTGGTGGCTCTTGGCTCTTGAAGGAGACATAGAAAGCCTTTCAAGCAGTGAGGA
ATCTATTACAGACATCAAGAAAGTTGAGAAGTTTAAACATGAGGACAGCAGCAAGGAGGATGAACACCGAGGATAAAA
TCCACCCCTTTTCCAGCCACAGCCTCTAGTCTATCTTATGCTGAGACCATCTTCCCTGCTCTCCACAGAACATCTCTG
CCTCTTGGCTCAGCCTGCTGGTGGTGGCTTTCCTTCAAGCTGAAATATGGAAGTCCCTAAAGCTAAAGACATGCTCTA
CACTAAGGAGAGATGATGCTCTTCTGAAATCTCCAGTATGCTCTTCTTGAAGCTCAAAATCCCAAACTCACTGATC
TTGAAAATCTGCACTTCTCTGCTCTGCTCAGCCCTTGTATGACAGAGTCCCTCAGCCTATTCTCAGACTCTTGAC
TTCCCTCTCAGCCCTCTGGTCTGCTCTCAGCCCAAGCTCTGCTATCTCCCAAGAGTGGTGGCTCTACCTCAGAGAG
CTGTCCTGTTCAAGCCTTCTGCTCAACCAAGAACTTCTACTTAACCTGCCCCAGTTTCAATCCCTATTAGTGT

>tiger_XM_007086963.3
ATGAAGTCTCTCATCTTGGCTGCTGGTGGCTCTTGGCTCTTGAAGGAGACATAGAAAGCCTTTCAAGCAGTGAGGA
ATCTATTACAGACATCAAGAAAGTTGAGAAGTTTAAACATGAGGACAGCAGCAAGGAGGATGAACACCGAGGATAAAA
TCCACCCCTTTTCCAGCCACAGCCTCTAGTCTATCTTATGCTGAGACCATCTTCCCTGCTCTCCACAGAACATCTCTG
CCTCTTGGCTCAGCCTGCTGGTGGTGGCTTTCCTTCAAGCTGAAATATGGAAGTCCCTAAAGCTAAAGACATGCTCTA
CACTAAGGAGAGATGATGCTCTTCTGAAATCTCCAGTATGCTCTTCTTGAAGCTCAAAATCCCAAACTCACTGATC
TTGAAAATCTGCACTTCTCTGCTCTGCTCAGCCCTTGTATGACAGAGTCCCTCAGCCTATTCTCAGACTCTTGAC
TTCCCTCTCAGCCCTCTGGTCTGCTCTCAGCCCAAGCTCTGCTATCTCCCAAGAGTGGTGGCTCTACCTCAGAGAG
ACATGCCCGTGCAACCTCTACTGCTGATCCAGGATGCCACCGTGAGGCTGCCCCAGTTTCAATCCCTATTATTGT

>cattle_XM_010806178
ATGAAGTCTCTCATCTTGGCTGCTGGTGGCTCTGCGCTTGAAGGAGAGATGTGGAAGCCTTTCAAGCAGTGAGGA
ATCTATTACAGACATCAAGAAAGTTGAGAAGTTTAAACATGAGGACAGCAGCAAGGAGGATGAACACCGAGGATAAAA
TCCACCCCTTTTCCAGCCACAGCCTCTAGTCTATCTTATGCTGAGACCATCTTCCCTGCTCTCCACAGAACATCTCTG
CCTCTTGGCTCAGCCTGCTGGTGGTGGCTTTCCTTCAAGCTGAAATATGGAAGTCCCTAAAGCTAAAGACATGCTCTA
CACTAAGGAGAGATGATGCTCTTCTGAAATCTCCAGTATGCTCTTCTTGAAGCTCAAAATCCCAAACTCACTGATC
TTGAAAATCTGCACTTCTCTGCTCTGCTCAGCCCTTGTATGACAGAGTCCCTCAGCCTATTCTCAGACTCTTGAC
TTCCCTCTCAGCCCTCTGGTCTGCTCTCAGCCCAAGCTCTGCTATCTCCCAAGAGTGGTGGCTCTACCTCAGAGAG
ATATGCCCATCTCAGGCTTCTTGTCTACCAAGAGCTGTACTGGTCTGCTCGGGGACCTCTCCTATTATTGT

>blue_whale_XM_036853045.1
ATGAAGTCTCTCATCTTGGCTGCTGGTGGCTCTTGGCTTGAAGGAGAGATGTGGAAGCCTTTCAAGCAGTGAGGA
ATCTATTACAGACATCAAGAAAGTTGAGAAGTTTAAACATGAGGACAGCAGCAAGGAGGATGAACACCGAGGATAAAA
TCCACCCCTTTTCCAGCCACAGCCTCTAGTCTATCTTATGCTGAGACCATCTTCCCTGCTCTCCACAGAACATCTCTG
CCTCTTGGCTCAGCCTGCTGGTGGTGGCTTTCCTTCAAGCTGAAATATGGAAGTCCCTAAAGCTAAAGACATGCTCTA
CACTAAGGAGAGATGATGCTCTTCTGAAATCTCCAGTATGCTCTTCTTGAAGCTCAAAATCCCAAACTCACTGATC
TTGAAAATCTGCACTTCTCTGCTCTGCTCAGCCCTTGTATGACAGAGTCCCTCAGCCTATTCTCAGACTCTTGAC
TTCCCTCTCAGCCCTCTGGTCTGCTCTCAGCCCAAGCTCTGCTATCTCCCAAGAGTGGTGGCTCTACCTCAGAGAG
ATATGCCCATCTCAGGCTTCTTGTCTACCAAGAGCTGTACTGGTCTGCTCGGGGACCTCTCCTATTATTGT

>hippopotamus_XM_057726918.1
ATGAAGTCTCTCATCTTGGCTGCTGGTGGCTCTTGGCTTGAAGGAGAGATGTGGAAGCCTTTCAAGCAGTGAGGA
ATCTATTACAGACATCAAGAAAGTTGAGAAGTTTAAACATGAGGACAGCAGCAAGGAGGATGAACACCGAGGATAAAA
TCCACCCCTTTTCCAGCCACAGCCTCTAGTCTATCTTATGCTGAGACCATCTTCCCTGCTCTCCACAGAACATCTCTG
CCTCTTGGCTCAGCCTGCTGGTGGTGGCTTTCCTTCAAGCTGAAATATGGAAGTCCCTAAAGCTAAAGACATGCTCTA
CACTAAGGAGAGATGATGCTCTTCTGAAATCTCCAGTATGCTCTTCTTGAAGCTCAAAATCCCAAACTCACTGATC
TTGAAAATCTGCACTTCTCTGCTCTGCTCAGCCCTTGTATGACAGAGTCCCTCAGCCTATTCTCAGACTCTTGAC
TTCCCTCTCAGCCCTCTGGTCTGCTCTCAGCCCAAGCTCTGCTATCTCCCAAGAGTGGTGGCTCTACCTCAGAGAG
AGATGCCCATCTCAGGCTTCTTGTCTACCAAGAGCTGTACTGGTCTGCTCGGGGACCTCTCCTATTATTGT
```

Figure 11: Cured alignment in FASTA format (saha_dibbyo_assignment4_phylogram_step5.txt)

Step 6

```
In [2]: %runfile /Users/dibbyosaha/Downloads/saha_dibbyo_assignment4/
a4.py --wdir
Matrix:
      Hu      Ti      Ca      BW      Hi
Hu  0.000000 154.534679 193.406857 168.565102 177.185819
Ti 154.534679 0.000000 157.307943 134.233743 157.307943
Ca 193.406857 157.307943 0.000000 94.831860 92.398181
BW 168.565102 134.233743 94.831860 0.000000 67.567297
Hi 177.185819 157.307943 92.398181 67.567297 0.000000
```

Figure 12: Output matrix from terminal produced from *a4.py* and *saha_dibbyo_assignment4_phylogram_step5.txt* (Note: Abbreviations were used for labels like 'Hu' for human, 'Ti' for tiger, 'Ca' for cattle, 'BW' for blue whale, 'Hi' for hippopotamus)

Step 7

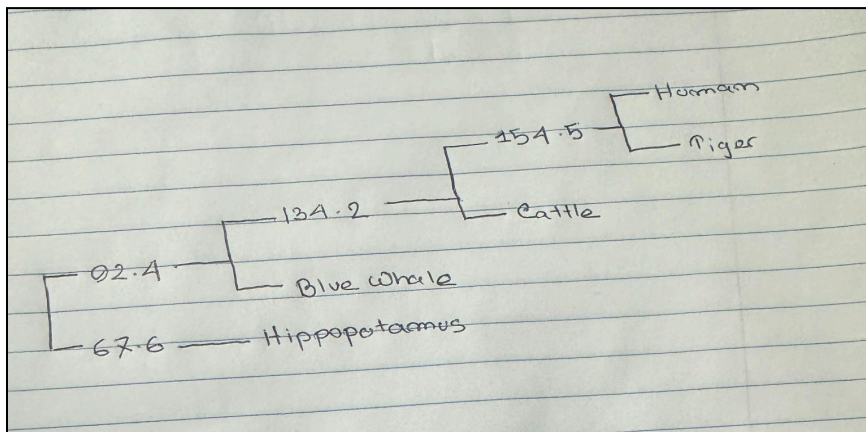


Figure 13: Hand drawn tree from the distance matrix (*saha_dibbyo_assignment4_tree_step7.png*)

Step 8

Similarities: Relationships between species are consistent in both of the trees. Both trees have humans and tigers as closest.

Differences: The hand-drawn tree uses distances from the distance matrix whereas the one from *phylogeny.fr* uses proportional scaling of branches.