Group 6:

Tiffany Cook

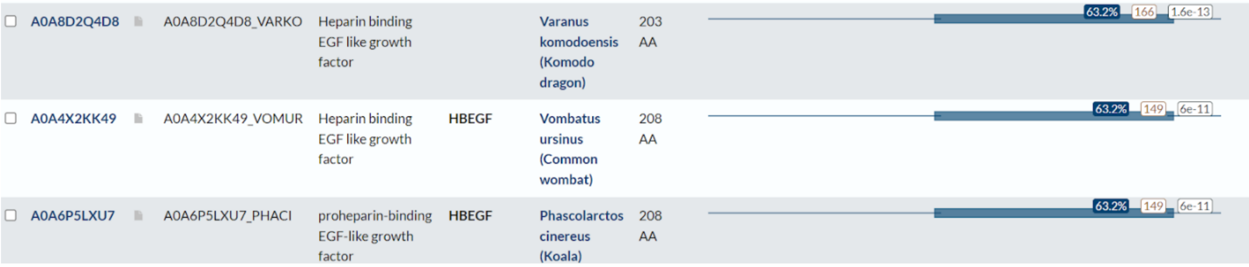
Drashti Mehta

Joshua Mikombo

BINF 6201 Lab Report 3

Part 1: Find Similar Sequences with BLAST

- 1A. What is the best-scoring BLAST result (not including the protein matching against itself)? What is the accession/Uniprot ID #, and what species does this best-scoring protein come from? Is the species with the best hit one that you would expect to be closely-related to an ant species? (1 pt)

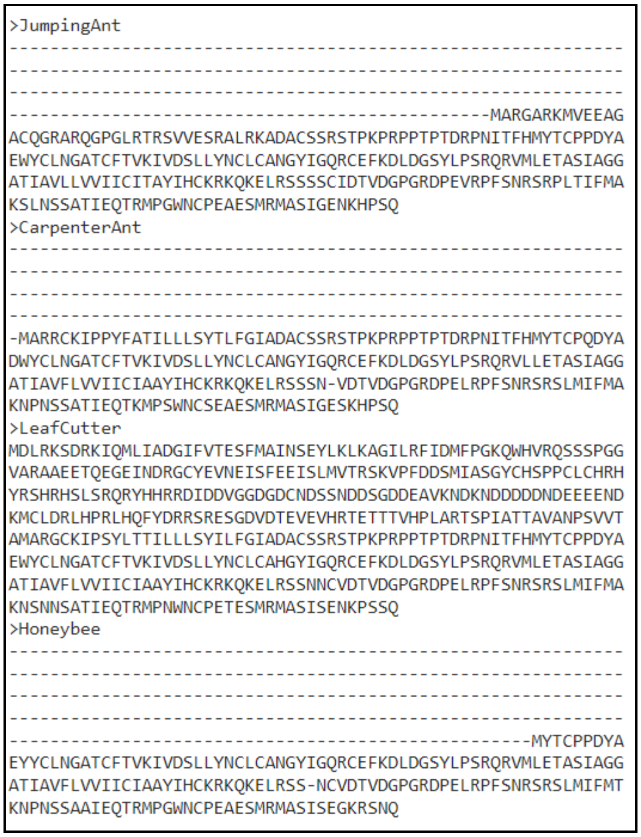


☐	A0A8D2Q4D8	A0A8D2Q4D8_VARKO	Heparin binding EGF like growth factor		Varanus komodoensis (Komodo dragon)	203 AA	63.2%	166	1.6e-13
☐	A0A4X2KK49	A0A4X2KK49_VOMUR	Heparin binding EGF like growth factor	HBEGF	Vombatus ursinus (Common wombat)	208 AA	63.2%	149	6e-11
☐	A0A6P5LXU7	A0A6P5LXU7_PHACI	proheparin-binding EGF-like growth factor	HBEGF	Phascolarctos cinereus (Koala)	208 AA	63.2%	149	6e-11

  - **Refer to screenshot****
  - **The accession number is A0A8D2Q4D8**
  - **It comes from the Komodo dragon.**
  - **We would expect it to be closely related to an ant species because the percent identity and score are of the highest in all 250 blast results, thus indicating the most similarity.**
- 1B. How many proteins from other insect species are in the top 250 BLAST results? (1 pt)
  - **There are 0 other proteins from other insect species**
  - **Next, either download the sequence or copy and paste it into a file, and go to the NCBI website to run PSI-BLAST against the standard databases with default parameters.**
- 1C. Does using the PSI-BLAST method, which should be better for finding more distantly related proteins, change the results you got earlier in any way? (2 pts)
  - **Yes;**
  - **The results of the BLAST with UniProt identified the Komodo Dragon as having the most similar protein sequence to the Red Bulldog Ant. When using the NCBI BLASTp, the same results appear.**
  - **There are 704 hits with BLASTp versus 250 with UniProt. Validating that PSI-BLAST found more distantly related proteins.**

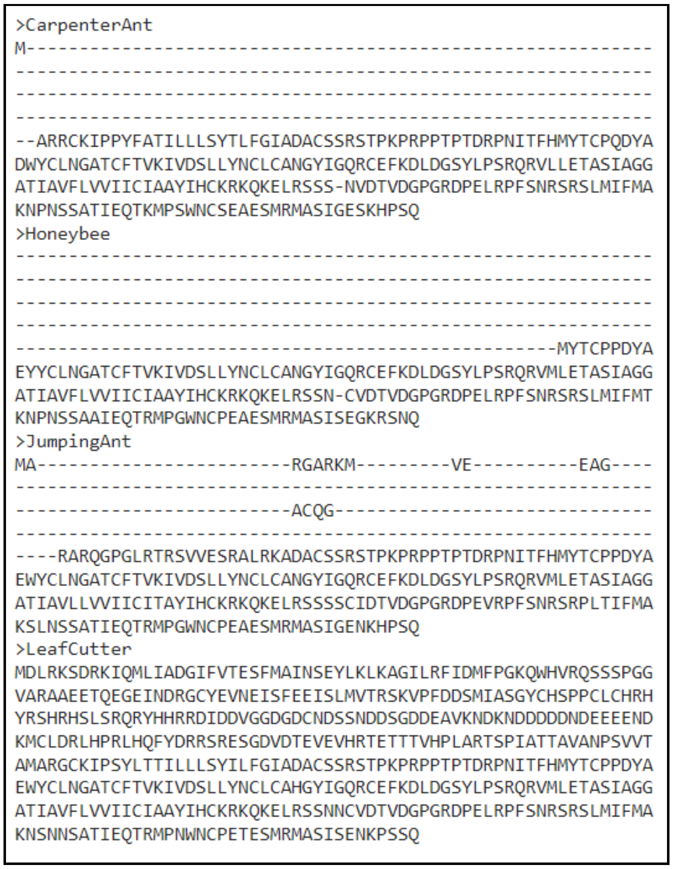
- In the BLASTp, there are 2 other types of ants listed- californicus and ruginodis. These may be insects that are distantly related and why they do not appear in the UniProt BLAST.

Part 2: Compare Multiple Sequence Alignment Results from Different Methods



```
>JumpingAnt
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------------------MARGARKMVEEAG
ACQGRARQGPGLRTRSVVESRALRKADACSSRSTPKPRPPTPTDRPNITFHMYTCPPDYA
EWYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVLLVVIICITAYIHCKRKQKELRSSSSCIDTVDGPGRDPEVRPFSNRSRPLTIFMA
KSLNSSATIEQTRMPGWNCPEAESMRMASIGENKHPSQ
>CarpenterAnt
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-MARRCKIPPYFATILLLSYTLFGIADACSSRSTPKPRPPTPTDRPNITFHMYTCPQDYA
DWYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVLLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSSSN-VDTVDGPGRDPELRPFSNRSRSLMIFMA
KNPNSSATIEQTKMPSWNCSEAESMRMASIGESKHPSQ
>LeafCutter
MDLRKSDRKIQMLIADGIFVTESFMAINSEYLKLKAGILRFIDMFPGKQWHVRQSSSPGG
VARAAEETQEGEINDRGCYEVNEISFEEISLMVTRSKVPFDDSMIASGYCHSPPCLCHRH
YRSHRHSLSRQRYHHRRDIDDVGGDGDCNDSSNDDSGDDEAVKNDKNDDDDDNDEEEEND
KMCLDRLHPRLHQFYDRRSRESGDVDTEVEVHRTETTTVHPLARTSPIATTAVANPSVVT
AMARGCKIPSYLTTILLLSYILFGIADACSSRSTPKPRPPTPTDRPNITFHMYTCPPDYA
EWYCLNGATCFTVKIVDSLLYNCLCAHGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSSNNCVDTVDGPGRDPELRPFSNRSRSLMIFMA
KNSNNSATIEQTRMPNWNCPETESMRMASISENKPSSQ
>Honeybee
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------------------------MYTCPPDYA
EYYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSS-NCVDTVDGPGRDPELRPFSNRSRSLMIFMT
KNPNSSAAIEQTRMPGWNCPEAESMRMASISEGKRSNQ
```

Fig: Muscle Alignment

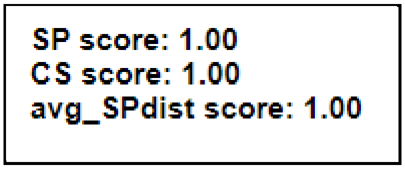


```
>CarpenterAnt
M-----------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
--ARRCKIPPYFATILLLSYTLFGIADACSSRSTPKPRPPTPTDRPNITFHMYTCPQDYA
DWYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVLLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSSS-NVDTVDGPGRDPELRPFSNRSRSLMIFMA
KNPNSSATIEQTKMPSWNCSEAESMRMASIGESKHPSQ
>Honeybee
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------------------------------------MYTCPPDYA
EYYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSSN-CVDTVDGPGRDPELRPFSNRSRSLMIFMT
KNPNSSAAIEQTRMPGWNCPEAESMRMASISEGKRSNQ
>JumpingAnt
MA------------------------RGARKM---------VE----------EAG----
------------------------------------------------------------
--------------------------ACQG------------------------------
------------------------------------------------------------
----RARQGPGLRTRSVVESRALRKADACSSRSTPKPRPPTPTDRPNITFHMYTCPPDYA
EWYCLNGATCFTVKIVDSLLYNCLCANGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVLLVVIICITAYIHCKRKQKELRSSSSCIDTVDGPGRDPEVRPFSNRSRPLTIFMA
KSLNSSATIEQTRMPGWNCPEAESMRMASIGENKHPSQ
>LeafCutter
MDLRKSDRKIQMLIADGIFVTESFMAINSEYLKLKAGILRFIDMFPGKQWHVRQSSSPGG
VARAAEETQEGEINDRGCYEVNEISFEEISLMVTRSKVPFDDSMIASGYCHSPPCLCHRH
YRSHRHSLSRQRYHHRRDIDDVGGDGDCNDSSNDDSGDDEAVKNDKNDDDDDNDEEEEND
KMCLDRLHPRLHQFYDRRSRESGDVDTEVEVHRTETTTVHPLARTSPIATTAVANPSVVT
AMARGCKIPSYLTTILLLSYILFGIADACSSRSTPKPRPPTPTDRPNITFHMYTCPPDYA
EWYCLNGATCFTVKIVDSLLYNCLCAHGYIGQRCEFKDLDGSYLPSRQRVMLETASIAGG
ATIAVFLVVIICIAAYIHCKRKQKELRSSNNCVDTVDGPGRDPELRPFSNRSRSLMIFMA
KNSNNSATIEQTRMPNWNCPETESMRMASISENKPSSQ
```

Fig: T-coffee Alignment

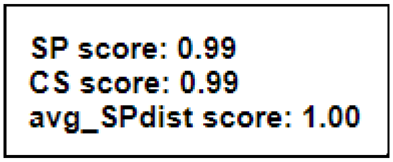
2A. How do the MUSCLE alignment and T-COFFEE alignment results each compare to the reference alignment? *(2 pts)*

- **MUSCLE alignment shows better matching to the reference compared to the T-Coffee alignment. Refer to the images attached below.**
- **Below are the results of VerAlign: Reference with the Muscle and the T-coffee alignment.**



SP score: 1.00
CS score: 1.00
avg_SPdist score: 1.00

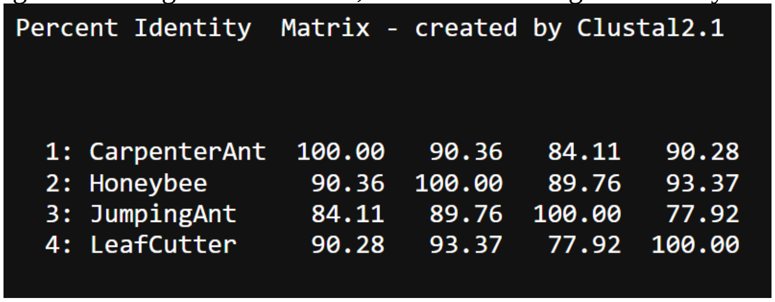
Muscle



SP score: 0.99
CS score: 0.99
avg_SPdist score: 1.00

T-coffee

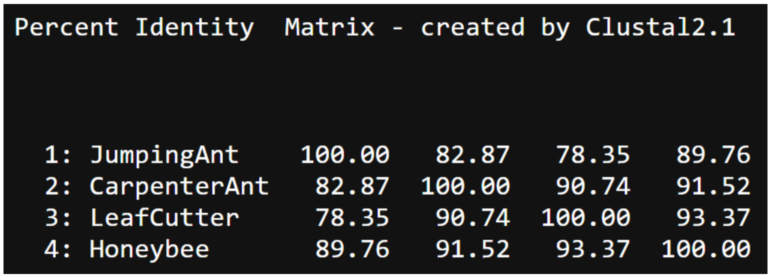
2B. Using the best alignment from 2A, what is the average % identity among the Spitz proteins? *(2 pts)*



```
Percent Identity  Matrix - created by Clustal2.1


  1: CarpenterAnt  100.00    90.36    84.11    90.28
  2: Honeybee       90.36   100.00    89.76    93.37
  3: JumpingAnt     84.11    89.76   100.00    77.92
  4: LeafCutter     90.28    93.37    77.92   100.00
```

T-coffee with an average % identity of 90.725%



```
Percent Identity  Matrix - created by Clustal2.1


  1: JumpingAnt     100.00   82.87   78.35   89.76
  2: CarpenterAnt    82.87  100.00   90.74   91.52
  3: LeafCutter      78.35   90.74  100.00   93.37
  4: Honeybee        89.76   91.52   93.37  100.00
```

Muscle with an average % identity of 90.825%

Part 3: Multiple Sequence Alignments with "Mystery" Venom Protein

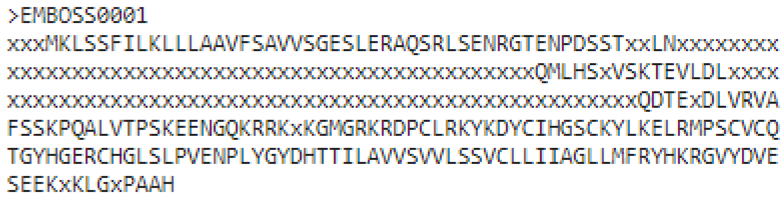
- 3A. On average, what is the percent similarity of the mystery protein from part 1 with the other insect venom proteins? *(1 pt*)
  - **30.07% similarity between the mystery protein and the other insect venom proteins**

3B. What is the average percent similarity of the mystery protein with the human EGF proteins? Is this higher or lower than the average percent identity among the different human proteins? *(2 pts*)

- **34.23% similarity between the mystery protein and the human EGF proteins. This is a higher similarity percentage than the average percent identity between the mystery protein and insect venom proteins.**

Part 4: Find Closest Match to Mystery Protein Among Local Marsupial Species

- 4A. How well does the ant protein align to these marsupial EGF sequences compared to the alignments with the other insects and the other mammals? Which species does it share the closest match with? *(3 pts*)
  - **There is a higher average percentage identity (40.88%) between the mystery protein and the marsupial EGF sequences compared to the human EGF proteins.**
  - **The mystery protein shares the closest match with the wombat (45.45%)**
- 4B. What is the consensus sequence for this alignment? *(1 pt)*?
  - **This is the consensus sequence for the alignment:**



```
>EMBOSS0001
xxxMKLSSFILKLLLAAVFSAVVSGESLERAQSRLSENRGTENPDSSTxxLNxxxxxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxQMLHSxVSKTEVLDLxxxx
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxQDTExDLVRVA
FSSKPQALVTPSKEENGQKRRKxKGMGRKRDPCLRKYKDYCIHGSCKYLKELRMPSCVCQ
TGYHGERCHGLSLPVENPLYGYDHTTILAVVSVVLSSVCLLIIAGLLMFRYHKRGVYDVE
SEEKxKLGxPAAH
```