**Problem 2 –**blastp and blastn BRCA1 against genes in different organisms

**Bos taurus**

blastn top scores: 5781, 5750

blastp top scores: 2592, 2587

**Rattus norvegicus**

blastn top scores: 3386, 3384

blastp top scores: 1854, 1849

**Drosophila melanogaster**

blastn top scores: 41, 41

blastp top scores: 52.4, 52.4

**Mus musculus**

blastn top scores: 3283, 3279

blastp top scores: 1847, 1846

**Xenopus laevis**blastn top scores: 167, 167

blastp top scores: 293, 293

**Problem 3**

Two letters will only be included in the subsequence if they match exactly. The score of the subsequence should increase linearly with the length of the subsequence. The “unit-score” scheme, where matching letters receive a score of 1 and non-matching letters (including 2 gaps) receive a score of 0, will work for this.

**Problem 4**

Letters that are not amino acids: b,j,o,u,x,z

1. a) “cmpterscience” versus “ilgy”

b) “cmpterscience” has the highest score of 30.8

c) “cmpterscience” has a match with a longer string length

1. a) “prtein” versus “aminacid”

b) “aminacid” has a highest score 28.2

c) “aminacid” has a match with a longer string length

1. a) “dynamicprgramming” versus “divideandcnqer”

b) “divideandcnqer” has the highest score of 32

c) “dynamicprgramming” has a match with a longer string length

**Problem 5**

1. Yes. Shakespeare likely used certain phrases or constructs more frequently than expected by random chance. Aligning the books would highlight these phrases if they exist.
2. No. Alignment highlights similarities. A better approach would be to look for k-phrases that occur more frequently than expected in all texts.
3. Yes. As we saw in class with the translation of “massa” alignment against a dictionary could highlight possible translation.
4. No. A substitution cipher would not be helped by alignment. It’s a different story for a Caesar cipher, though.
5. Yes. Aligning the papers could highlight phrases or larger constructs that are similar.

**Problem 6**

I used the NCBI COBALT tool with default settings to do multiple alignment. The results:

10001 1 TPNVSVVDLTVRLGKG--------- 16

10002 1 ------------LEKPAKYDDIK-- 11

10003 1 ------------LDDDVTESDVNAA 13

10004 1 --------------KGASYEDVKAA 11

10005 1 --DVSVVDLTV-------------- 9

10006 1 --------LTCRLEKPAKY------ 11

10007 1 -------------NKETTYDEIKKV 12