Bioinformatics Blitz Answers

**Task 1 – Bacteria Blitz**

**Question 1a:** Complete the dotplot below for the following two bacterial sequences, writing one sequence in the top row and the other in the first column:

Sequence 1: VAPSAELYAVKVLG

Sequence 2: VLGVAPSADLYAVK

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | V | A | P | S | A | E | L | Y | A | V | K | V | L | G |
| V | **⋅** |  |  |  |  |  |  |  |  | **⋅** |  | **⋅** |  |  |
| L |  |  |  |  |  |  | **⋅** |  |  |  |  |  | **⋅** |  |
| G |  |  |  |  |  |  |  |  |  |  |  |  |  | **⋅** |
| V | **⋅** |  |  |  |  |  |  |  |  | **⋅** |  | **⋅** |  |  |
| A |  | **⋅** |  |  | **⋅** |  |  |  | **⋅** |  |  |  |  |  |
| P |  |  | **⋅** |  |  |  |  |  |  |  |  |  |  |  |
| S |  |  |  | **⋅** |  |  |  |  |  |  |  |  |  |  |
| A |  | **⋅** |  |  | **⋅** |  |  |  | **⋅** |  |  |  |  |  |
| D |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| L |  |  |  |  |  |  | **⋅** |  |  |  |  |  | **⋅** |  |
| Y |  |  |  |  |  |  |  | **⋅** |  |  |  |  |  |  |
| A |  | **⋅** |  |  | **⋅** |  |  |  | **⋅** |  |  |  |  |  |
| V | **⋅** |  |  |  |  |  |  |  |  | **⋅** |  | **⋅** |  |  |
| K |  |  |  |  |  |  |  |  |  |  | **⋅** |  |  |  |

**Question 1b**: Is there a clear alignment between these sequences? How long are the longest diagonal lines? Write out the matching sequence alignment e.g LGVA……

There are 3 stretches producing diagonal lines that suggest alignment, with 2 of them leading into each other suggesting a clear alignment there, with the following sequence:

VAPSA-LYAVK, where the “-“ is D in one sequence and E in the other.

The longest matching diagonal lines are 5 amino acids long each.

**Question 2:** Calculate sequence alignment scores for the following pairs of sequencealignments**.**

**a:**

S R V Q A P A A H

| | | | | | | | |

S F I N T Q Q A H

|  |
| --- |
| 4 -3 +3 +0 +0 -1 -1 +4 +8 = **14** |

**b:**

G S G V K V A V L

| | | | | | | | |

G S F G T A E M L

6 +4 -3 -3 -1 +0 -1 +1 +4 = **7**

**c:** Which pair of sequences seems more similar at first glance and which has the higher score?

Both have 3 matches and 6 changes, so it’s hard to say which seems more similar at first glance.

The comparison in a) has the higher score, meaning a better similarity match. It seems that conservation of H is important as it gets a high score and more of the substitutions in a) have scores of 0 (e.g. Q-N), suggesting these changes are less likely to affect the whole protein, in contrast with more negative scores in b) (e.g. G-F).

**Question 3:** What is the score for the comparison between A and B?

916.0

**Question 4:** What is the score for the comparison between A and C?

173.0.

**Question 5:** How different are the two scores? Can you notice any obvious differences in the way the sequences are aligned in the A vs B comparison and the A vs C comparison?

A vs C has a much lower score than for A vs B (although it is reasonable to question the scale of scores, i.e. what would be the score for two completely unrelated sequences or for two identical sequences and what impact does sequence length have on score?).

An obvious difference in the alignments is that there are no insertions in the A vs B comparison, whereas there are quite a few in the A vs C alignment.

Vertical line symbols represent identical amino acids in the two sequences, whereas full stops and colons represent different levels of similarity, with colons for more similar amino acids and full stops for less similar amino acids. A vs C has far more full stops and colons than A vs B, meaning that A and B have more identical sequences than A and C.

See screenshots of the alignments on page 3.



**A vs B**



**A vs C**

A few of the insertions are circled in red here. These are sections where it seems that one of the proteins has had a sequence of amino acids inserted into it relative to the other sequence, as they are aligned either side of these sections.

**Question 6:** Which sequences do you think are the two bacterial samples and which is the human sample? Which protein(s) would make a good target for an antibacterial drug?

It looks like A and B are the two bacterial samples and that C is the human sample, as you would expect the human protein to be more different to either bacterial protein than they are to each other. This suggests that any drug targeted against these bacterial proteins (A and B) seems unlikely to affect this human protein, but might be active against both bacterial proteins which could be very useful to treat disease, without causing unwanted side-effects.

# **Task 2: Insertion identifier**

**Question 7:** What organism(s) are these proteins from?

They are both from midges.

**Question 8:** In the middle of the alignment, there is one insertion of 6 amino acids into one of the sequences. What word do the first 5 letters of the insertion spell out?

ALIGN

# **Task 3: Proteins and disease**

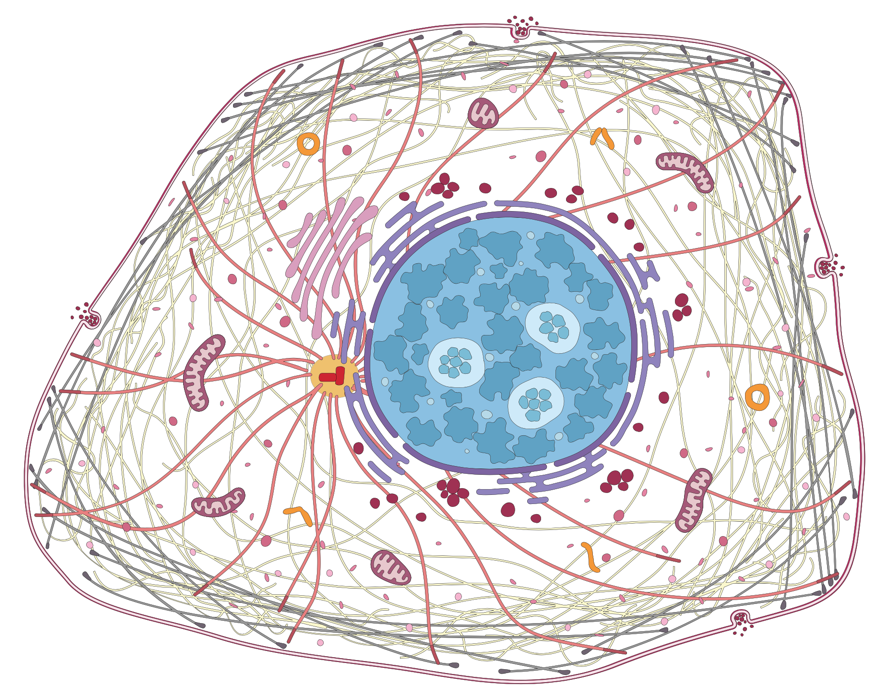
**Question 9:** Complete the table below:

|  |  |  |
| --- | --- | --- |
| **Protein ID** | **Protein Name** | **Disease** |
| **P42858** | Huntingtin | Huntington disease and Lopes-Maciel-Rodan syndrome – both are neurodegenerative diseases |
| **P13569** | Cystic fibrosis transmembrane conductance regulator | Cystic fibrosis – disorder of exocrine glands, characterised by respiratory problems |
| **P35625** | Metalloproteinase inhibitor 3 | Sorsby fundus dystrophy – degenerative disease affecting the back of the eye, resulting in loss of central vision |
| **Q9NPJ1** | McKusick-Kaufman/Bardet-Biedl syndromes putative chaperone | McKusick-Kaufman syndrome – developmental disease |
| **P08100** | rhodopsin | Retinitis pigmentosa – disease involving progressive loss of rod photoreceptors in the eye, leading to loss of night vision and loss of peripheral visual field |
| **Q14204** | Cytoplasmic dynein 1 heavy chain 1 | Charcot-Marie-Tooth disease (and others) – disorder of the peripheral nervous system, involving progressive weakness |

**Question 10:** Highlight the location of each protein on the cell diagram below. (Remember they may be found in more than one location in the cell)

**P13569** Endoplasmic reticulum, endosomes, nucleus, and membrane

**P35625** Extracellular



**P08100**   
Cell membrane and the Golgi apparatus

**P42858**Nucleus

**Q9NPJ1**

Nucleus, cytoplasm, and cytoskeleton

**P42858**   
Cytoskeleton