Practical 6

## Date: 06-03-2022 Name: Aswin. V

## Day: Sunday Roll Number: BS19B003

**Q1 –**

<https://docs.google.com/document/d/1_wY4mvc4JocC2_cGiF4zzPSb-coUq3nYSZ9Z8VN_0kY/edit?usp=sharing>

\*above doc has all data for the given five conditions; for the given two sets

**Q2 –**

For **Condition 1**

**Highest** Conservation Scores:

|  |  |  |
| --- | --- | --- |
| Position | Residue | Set 1 |
| 117 | K | -1.846 |
| 113 | T | -1.72 |
| 70 | L | -1.673 |
| 73 | S | -1.594 |
| 84 | E | -1.414 |
| 133 | G | -1.414 |
| 36 | V | -1.367 |
| 35 | V | -1.295 |
| 9 | A | -1.169 |
| 6 | K | -1.16 |

|  |  |  |
| --- | --- | --- |
| Position | Residue | Set 2 |
| 85 | T | -1.889 |
| 116 | Q | -1.831 |
| 199 | A | -1.831 |
| 79 | E | -1.735 |
| 75 | V | -1.677 |
| 88 | H | -1.677 |
| 101 | L | -1.677 |
| 145 | S | -1.677 |
| 197 | R | -1.677 |
| 248 | L | -1.677 |

**Lowest** Conservation Scores:

|  |  |  |
| --- | --- | --- |
| Position | Residue | Set 1 |
| 1 | : | 0 |
| 3 | L | 0 |
| 4 | S | 0 |
| 7 | D | 0 |
| 8 | K | 0 |
| 17 | K | 0 |
| 26 | G | 0 |
| 30 | L | 0 |
| 32 | R | 0 |
| 38 | P | 0 |

|  |  |  |
| --- | --- | --- |
| Position | Residue | Set 2 |
| 64 | N | 0 |
| 66 | K | 0 |
| 68 | N | 0 |
| 95 | V | 0 |
| 120 | Q | 0 |
| 121 | N | 0 |
| 128 | G | 0 |
| 129 | A | 0 |
| 130 | F | 0 |
| 131 | T | 0 |

**Q3 & Q4 –**

Q3:

Algorithm:

* Takes list of strings as input (output after MSA).
* Then calculation of unweighted frequency (uwf) of all amino acids at each position in the list.
* Then calculation of conservation score through entropy, variance and sum of pairs.

Q4:

Algorithm:

Part\_1:

Comparing of conservation scores from Clustal Omega, MAFFT, MUSCLE.

Part\_2:

Computes conservation scores from Clustal Omega, MAFFT, MUSCLE (using Q3). And then calls Part\_1.

Code for Q3 & Q4:

<https://drive.google.com/file/d/1meJuirxwVESFkVqcEmmIDsNoBbx5sIw6/view?usp=sharing>

\*code is in the above file

Output:

<https://docs.google.com/document/d/1yyS6VeNnHioyXcg5UT1OKD7movYrJk7ItneE7Okyoe4/edit?usp=sharing>

\*output is in the doc above

**Q5 –**

Manual calculation for Set 1 and Set 2:

<https://drive.google.com/file/d/1MNV7D4VNMnA7a7I4AemnVBiLUamcdsl7/view?usp=sharing>

\*above pdf file has the solution for Q5

**Q6 –**

* Go to ConSurf server
* Select analyse Amino acids, then yes for protein structure, then PDB ID = 1BTM
* Then ‘No’ for MSA.
* Then submit (keep the suggested parameters, can change to dig in more)

Alignment Details:

1. Graphical user interface

   Description automatically generatedThe average number of replacements between any two sequences in the alignment;  
   A distance of 0.01 means that on average, the expected replacement for every 100 positions is 1.  
   Average pairwise distance: 0.99628  
   Lower bound: 0.10971  
   Upper bound: 1.89437