Bioinformatics Assignment Worksheet

Exercise 3

**MSA Calculation Template**

**Example MSA:**

Seq A MGDVEKGKKIFVMKCSQCHTVEKGG

Seq B MVDVEKGMKIFVMKCSQCHTVEAGG

Seq C MVDVEKGMLIFVMKCSQCHTVEAGG

Seq D MGDIEKGKLIFVMSCSQCHTVYAGY

Seq E MIDTEKGYIIFVMTCSQCHTVLMGT

From the aligned sequences above we can individually calculate the fraction of differing amino acid residues between the 25 amino acids that make up Seq A and B (identical residues represented by “|”):

Seq A MGDVEKGKKIFVMKCSQCHTVEKGG

| ||||| |||||||||||||| ||

Seq B MVDVEKGMKIFVMKCSQCHTVEAGG

Alignment value 3/25 = 0.12 this value goes into the **Distance Matrix**

**Directions**: Calculate the following alignment identity values and place them into the Distance Matrix below.

Seq A MGDVEKGKKIFVMKCSQCHTVEKGG

Seq C MVDVEKGMLIFVMKCSQCHTVEAGG

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq A MGDVEKGKKIFVMKCSQCHTVEKGG

Seq D MGDIEKGKLIFVMSCSQCHTVYAGY

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq A MGDVEKGKKIFVMKCSQCHTVEKGG

Seq E MIDTEKGYIIFVMTCSQCHTVLMGT

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq B MVDVEKGMKIFVMKCSQCHTVEAGG

Seq C MVDVEKGMLIFVMKCSQCHTVEAGG

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq B MVDVEKGMKIFVMKCSQCHTVEAGG

Seq D MGDIEKGKLIFVMSCSQCHTVYAGY

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq B MVDVEKGMKIFVMKCSQCHTVEAGG

Seq E MIDTEKGYIIFVMTCSQCHTVLMGT

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq C MVDVEKGMLIFVMKCSQCHTVEAGG

Seq D MGDIEKGKLIFVMSCSQCHTVYAGY

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq C MVDVEKGMLIFVMKCSQCHTVEAGG

Seq E MIDTEKGYIIFVMTCSQCHTVLMGT

Alignment Value \_\_\_\_\_\_\_\_\_\_

Seq E MIDTEKGYIIFVMTCSQCHTVLMGT

Seq D MGDIEKGKLIFVMSCSQCHTVYAGY

Alignment Value \_\_\_\_\_\_\_\_\_\_

**Distance Matrix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Seq A | Seq B | Seq C | Seq D | Seq E |
| Seq A | - | - | - | - | - |
| Seq B | 0.12 | - | - | - | - |
| Seq C |  |  | - | - | - |
| Seq D |  |  |  | - | - |
| Seq E |  |  |  |  | - |

**Question 1**

What species in your list of sequences derived from HomoloGene could you utilize as an “outgroup” if you wanted to build a rooted tree?

**Cytochrome C MSA**

**Question 2**

Using the percent identity matrix, which two organisms have the least evolutionary time separating them based on the molecular data analyzed?

**Cytochrome C Phylogenetic Tree**