**Practical assignment**

**Module topic: Multiple Sequence Alignment**

**Contact session title: Building a Multiple Sequence Alignment**

**Trainer: Ahmed M. Alzohairy**

**Name:**

**Date:**

**Assignment – Building a Multiple Sequence Alignment**

**Introduction**

***Building multiple sequence alignments (MSA) is far from an exact science.******In fact, it’s more art than science, requiring that you use everything you know in bioinformatics and in biology.*** ***The main idea behind building a multiple sequence alignment is to put similar amino acids or nucleotides in the same column if they contain the same criterion. There are four major criteria most scientists use to build a multiple alignment of sequences that all have different properties.***

**Tools used in this session**

*The most popular programs for building MSA: we will see the differences between Cobalt, ClustalW,* ***MUSCLE****,* ***Tcoffee****.*  For analysis: we will use other online tools.

Please note

* **Hand-in information** If you are formally enrolled in the IBT course,please upload your completed practical assignment to the Vula ‘Practical Assignments’ tab. Take note of the final hand-in date for each practical assignment, which will be indicated on Vula.

*Task1: Reading the slides*

**Task 1: instructions**

*Read and understand the contents of the powerpoint slides*

1. ***State major characterized*** *differences between Cobalt, ClustalW,* ***MUSCLE****,* ***Tcoffee?***
2. ***What are the Main Applications of Multiple Sequence Alignments?***

**Task 2: Retrieve your protein sequences**

**We need to retrieve protein sequences (eg. Heat shock protein “HSP70”) from evolutionary different organisms (*Human, Rat, Mouse, Arabidopsis, Chicken, Pig)***

**Task 2: instructions**

***- Go to*** [***https://www.expasy.org/proteomics***](https://www.expasy.org/proteomics)

***- Search for HSP70 (Heat Shock Protein70)***

***- Click on (***[***UniProtKB***](http://www.uniprot.org/uniprot/?query=HSF1&sort=score)[*http://www.uniprot.org/uniprot/?query=HSF1&sort=score*](http://www.uniprot.org/uniprot/?query=HSF1&sort=score)***)***

***- Retrieve your protein sequences (eg. Heat Shock Protein 70 “HSP70”) from different organisms***

***- Select your organism (Human (P0DMV8), Rat ((P0DMW1)), Mouse (*Q61696*), Arabidopsis (P22953), Bos Taurus*** (***Q27975***)***, Escherichia coli (P0A6Y8))***

***- Click Download (Download Selected) then (Go)***

***- Save it in FASTA format in one text file.***

***- Align the sequences using Clustal Omega***

***- Check the gene-based phylogentics tree***

***Based on the tree:***

1. ***What is the name the MSA program used in UniportKB?***
2. ***How many identical positions are there and what is the identity percentage?***
3. ***How many similar positions are there between those sequences?***
4. ***What is the most different organism in “HSP70” from your list?***
5. ***What is the most similar sequence to Human HSP70 from the set of sequences?***

**Task 3:  *Extracting annotation (identifying conserved domain)***

**Task 3: instructions**

**From the right-hand side panel you can extract some important information**

* ***Identify the first four Amino Acids symbols at the Beta Strand in Human?***
* ***Identify the first four Amino Acids symbols at the Helix in E.Coli?***

***Task4: Extracting amino acid properties***

**Task 4: instructions**

**From the right-hand side panel you can extract some important information**

* ***Identify the first three conserved aromatic amino acids in all specified organisms?***

***- Identify the first three hydrophobic acids in Human?***

**\* The student can choose any three**

***Task5: Interpretation***

**Task 3: instructions**

**Check the information in the last lecture slide**

1. ***The last line contains seemingly ClustalW, MUSCLE, or Tcoffee alignment, cabalistic signs such as (\*), (:), or (.), what do they mean?***
2. ***What could conserved K (Lysine), R (Arginine), D (Aspartic Acid), E (Glutamic Acid) mean in protein MSA?***