LEARN PYTHON & R FOR BIOINFORMATICS

Prerequisites:

1. **Multiple sequence alignment:**
2. **MEGA 7 installed on your PC:**
   You can download MEGA 7 on your PC free of cost by visiting to the following link:
   https://mega.software.informer.com/7.2/
3. **Phylogenetic terminologies.**

4. **Maximum likelihood tree.**
5. **Neighbour-Joining tree.**
6. **Minimum-Evolution tree.**
7. **UPGMA tree.**
8. **Maximum Parsimony tree.**
9. **Bootstrap algorithm.**

Introduction:

Phylogenetics:

It is the study of evolutionary relationships among biological entities - often species, individuals or genes/proteins (which may be referred to as taxa). In phylogenetic analysis, we build phylogenetic trees using different methods, in order to find out the evolutionary history between different organisms of different species. In Bioinformatics we use different tools or softwares to create these phylogenetic trees, the most commonly used and efficient tool for this purpose is MEGA.

MEGA:

MEGA is most commonly used tool in Bioinformatics for phylogenetic analysis. It uses different algorithms to build phylogenetic trees based on different methods, i.e, Maximum likelihood, Neighbour-Joining, Minimum-Evolution, UPGMA, and Maximum Parsimony tree generating methods.

Steps:

MSA analysis through MEGA:
[Before entering your query sequences in MEGA, make sure to remove the garbage information,i.e., one-line definition except '>' symbol and the accession numbers, from your FASTA file.]

- To align multiple sequences, downloaded in FASTA file, open MEGA and then click on the 'File' option and then click on 'Open a File' option to load your downloaded FASTA sequence

file in MEGA, and then click on the 'align' option, to analyze the Multiple Sequence Alignment results through MEGA.

- It will open a pop-up window on your screen where it will display you the sequences you input on MEGA. Each row displays a different sequence given against their accession numbers and each residue is represented by a different colour based on their physicochemical properties.

- In the pop-up window, on the top bar it shows different options like File, Edit, Search, Alignment, Web, Display, Help, etc. To align your desired sequence, just select your sequences you want to align and then click on the 'Alignment' option and select the tool (CustalW (subtool of ClustalOmega) or Muscle), you want to align your sequences with.

- Just go with the default options unless you have to change your parameters according to your desired analysis.

- To analyze your results of multiple sequence alignment, kindly watch our video on Clustal Omega from 'Sequence Alignment and Analysis' courses for better understanding of your results.

- You can save your alignment on your PC by clicking on the 'Data' option and then 'Export Alignment' and then select your File Format, you would like to store your alignment in, for MEGA tool MEGA format is preferred.

- Otherwise, if you just want your sequences for phylogenetic tree building analysis, click on 'Data' and then 'Phylogenetic Analysis' and it will load your alignment on MEGA for your further analysis.

To create phylogenetic trees with different methods through MEGA:

- Click on the 'Phylogeny' option present on the menu bar of MEGA and select your required method to generate the phylogenetic tree of your input sequences.

Note: The most reliable method to construct a tree for phylogenetic analysis (for research purposes) is the Neighbour-Joining method with bootstrapping.

1. **Constructing tree with Maximum Likelihood method:**
   - Go to 'Phylogeny' option and then select the 'Construct/Test Maximum Likelihood Tree' option. Then on the pop-up window select 'Remember to reuse the currently active data' and then click YES.
   - It will open up for you another pop-up window where it'll display multiple options to change your parameters according to your research/analysis purposes. You can also go with the default parameters if you don't need to change any.
   - Then, click on the 'Compute' option and it'll display you a phylogenetic tree of your input sequence. Just save your results for further analysis.

2. **Constructing tree with Neighbour-Joining method:**
   - Go to 'Phylogeny' option and then select the 'Construct/Test Neighbour-Joining Tree' option.
   - Go with default parameters (or change according to your research purpose) and then click on 'Compute' option and save your results for further analysis.

3. **Constructing tree with Minimum-Evolution method:**
   - Go to 'Phylogeny' option and then select the 'Construct/Test Minimum-Evolution Tree' option.
   - Go with default parameters (or change according to your research purpose) and then click on 'Compute' option and save your results for further analysis.

4. **Constructing tree with UPGMA method:**
   - Go to 'Phylogeny' option and then select the 'Construct/Test UPGMA Tree' option.
   - Go with default parameters (or change according to your research purpose) and then click on 'Compute' option and save your results for further analysis.

5. **Constructing tree with Maximum-Parsimony method:**
- Go to 'Phylogeny' option and then select the 'Construct/Test Maximum-Parsimony Tree' option.
- Go with default parameters (or change according to your research purpose) and then click on 'Compute' option and save your results for further analysis.

Comparisons of results:

Maximum Likelihood Tree	Neighbour-Joining Tree	Minimum Evolution Tree	UPGMA Tree	Maximum Parsimony Tree
Relatively older method of tree construction, and also doesn't use Bootstrap iterations, hence its phylogram shows relative difference in evolutionary time of the diverging sequences.	It is the most commonly used and relatively efficient method of tree construction. It uses Bootstrap Iterations, hence providing relatively much reliable results.	This method also uses Bootstrap iterations, hence it shows comparatively the same results as Neighbour Joining method. In both these methods, it shows some numbers on the branches of their trees which are basically the no. of iterations performed.	UPGMA simply clusters out the sequences and basically creates a cladogram of the input sequences.	Just like the UPGMA method, it also clusters out our input sequences, yet it shows visible differences from the tree constructed from the UPGMA method.

Note: If you want to do evolutionary analysis on your query sequences, you should use 1st three methods, i.e., Maximum Likelihood, Neighbour-Joining, or Minimum-Evolution methods. And if you just want to cluster out your query sequences, you should go with the other two methods, i.e., UPGMA and Maximum Parsimony methods.

Summary:

In this tutorial, we learned to analyze evolutionary relationships between different sequences of the same protein, present in the organisms of different species, by creating phylogenetic trees, using five different methods. We learned to generate phylogenetic trees using MEGA, we also learned how to align multiple sequences using MEGA. We also compared the results after creating trees in MEGA using all the five different methods.