

# Phylogenetic Analysis

*In this class, we will talk about ...*

---



## Learning Objectives

- Define the problem and the relevance of Phylogenetic Analysis
- Discuss the Hierarchical Clustering algorithm
- The Unweighted Pair Group Using Arithmetic Averages (UPGMA) approach for Hierarchical Clustering

1. Review the slides “Phylogenetic Analysis”.



### **Task 1 – Phylogenetics Tutorial**

- Go to <https://www.ebi.ac.uk/training/online/courses/introduction-to-phylogenetics/what-is-phylogenetics/>
- 
- *Have a look at the materials*
- *Try to answer the quiz.*

Consider the following multiple sequence alignment from a set of four sequences. Using a MSA with match = 1, mismatch = -1 and gap = -1 calculate:

s1:ATAGC  
s2:ATGAC  
s3:AACG  
s4:AATCG

### Task 1 – Derive the Multiple Sequence Alignment

1) Implement the code from MSA and the above parameters define the multiple alignment for the above sequences. Note: use the code from the last class.

### Task 2 – Build Phylogenetic Tree

2) Calculate the distance matrix. Assume the metric distance as the number of distinct characters in pairwise alignment (assume the pairwise alignment given by the MSA, i.e. use the alignment of every two pairs of sequences from the MSA).

3) Build the tree for the sequences using the UPGMA algorithm.

### Task 3 – Extend the functionalities

4) In the BinaryTree class, write a function called *get\_cluster()* that given a tree returns all the elements (leaves) in the tree as list.

Hint: Traverse tree and collect the elements in the leaves. If it is in an internal node (value == -1) call the function recursively. Note: The method extends a list allows to add the elements of list2 are added to the end of list1: list1.extend(list2).

5) In the BinaryTree class, implement the method, *exists\_leaf()*. For a given input value, returns a boolean indicating if the value exists or not. The value should be found as a leaf in the tree.

6) Fill in the code for method *execute\_clustering()* in class HierarchicalClustering.

7) Fill in the code for methods *create\_mat\_dist()* and *run()* in class UPGMA.