Multiple Sequence Alignment

Introduction:

Multiple sequence alignment (MSA) is a process or the result of sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. Multiple sequence alignment is often used to assess sequence conservation of protein domains, secondary and Tertiary structures, and even individual amino acids or nucleotides. There are two different techniques used in Multiple Sequence alignment.

1. Progressive alignment method :

Widely used and builds up multiple sequence alignment beginning from pairwise alignment. Two stages :-

1st stage- Relationships between the sequences are represented as guide tree,

2nd sage- Add more sequences to that guide tree to form a MSA.

Examples: Clustalw, T-Coffee.

2. Iterative alignment method:

Similar to progressive alignment.

A new sequence is added to the algorithm so the initially aligned sequences are repeatedly realigned in order to obtain the best alignment refinement.

Example MUSCLE, DIALIGN

Phylogenetic tree

Introduction:

Graphical representation of the evolutionary relationship among three or more genes or organisms.

Based on Relatedness of data, divergence times, nature of common ancestors.

1. Rooted Tree:

It is used to make inferences about the most common ancestor of the leaves or branches of the tree. Most commonly the root is referred to as an "outgroup".

Sourav Singh, 198137

2. Unrooted Tree:

Unrooted trees illustrate the relatedness of the leaf nodes without making assumptions about ancestry.

Methodology For MSA

For MSA, Clustalw implements progressive alignment technique and for tree construction , it implements Neighbour joining method.

a. Progressive alignment: It is widely used and also known as hierarchical or tree method. It builds up multiple sequence alignment beginning from pairwise alignment . First, sequences are related by using a guide tree and then sequences are added gradually to the guide tree. Example: Clustalw, T-Coffee.

b. Neighbour joining method: These methods apply general data clustering techniques to sequence analysis using genetic distance as a clustering metric. This method produces unrooted trees, but it does not assume a constant rate of evolution across lineages. It begins with an unresolved star-like tree. Each pair is evaluated for being joined and the sum of all branch's length is calculated of the resultant tree. The pair that yields the smallest sum is considered the closest neighbours and is thus joined . A new branch is inserted between them and the rest of the tree and the branch length is recalculated . This process is repeated until only one terminal is present.

Align any 10 sequences of your choice by using Clustal Omega and explain their relation using a phylogenetic tree.

Here, I have used the prion protein sequences from 10 different species for the multiple sequence alignment and they are:

Accession	number organism
AAH12844	Prion protein [Homo sapiens]
ABY26043	prion protein [Macaca mulatta]
AAB50624	prion protein, partial [Colobus guereza]
AGD99656	prion protein [Bubalus bubalis]
ABL86017	prion protein [Hippotragus niger]

Sourav Singh, 198137

ABU98047   prion protein [Bos taurus]

AAV41893   prion protein, partial [Chiropotes satanas]

AAV41891   prion protein, partial [Plecturocebus moloch]

AAV41879   prion protein, partial [Alouatta belzebul]

AAV41883   prion protein, partial [Sapajus apella]

The CLUSTAL OMEGA result for the above sequences is:

CLUSTAL O1.2.4) multiple sequence alignment

```
AAH12844.1 --MANLGCWMLVLFVATWSDLGLCKKRPKPG-GWNTGGSRYPGQGSPGGNRYPPQGGGGW 57

AAB50624.1 ---------MLVLFVATWSDLGLCKKRPKPG-GWNTGGSRYPGQGSPGGNRYPPQGGGGW 50

ABY26043.1 --MANLGCWMLVLFVATWSDLGLCKKRPKPG-GWNTGGSRYPGQGSPGGNRYPPQGGGGW 57

AAV41883.1 -----------------------------G-GWNTGGSRYPGQGSPGGNRYPPQG-GGW 28

AAV41879.1 -----------------------------G-GWNTGGSRYPGQGSPGGNRYPPQS-GGW 28

AAV41893.1 -----------------------------G-GWNTGGSRYPGQGSPGGNRYPPQGGGGW 29

AAV41891.1 -----------------------------G-GWNTGGSRYPGQGSPGGNRYPPQGGGSW 29

ABU98047.1 MVKSHIGSWILVLFVAMWSDVGLCKKRPKPGGGWNTGGSRYPGQGSPGGNRYPPQGGGXW 60

AGD99656.1 MVKRHIGSWILVLFVVMWSDVGLCKKRPKPGGGWNTGGSRYPGQGSPGGNRYPSQGGGGW 60

ABL86017.1 MVKSHIGRWILVLFVAMWSDVGLCKKRPKPGGGWNTVGSRCPGQGSPGGNRYPPQGGGGW 60

            * **** *** ************ *. * *
```

```
AAH12844.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGTHSQWNKPSKPKTNM 109

AAB50624.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGTHSQWNKPSKPKTSM 102

ABY26043.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGTHNQWHKPSKPKTSM 109

AAV41883.1 GQPHGGGWGQPHGGGWGQPHGGSWGQPHGGGWGQG--------GGTHNQWNKPSKPKTSM 80

AAV41879.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGTHNQWNKPSKPKTNM 80

AAV41893.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGAHSQWNKPSKPKTNM 81

AAV41891.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQG--------GGTHNQWNKPSKPKTNM 81

ABU98047.1 GQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGGWGQGGTHGQWNKPSKPKTNM 120

AGD99656.1 GQPHGGGWGQ--------PHGGGWGQPHGGGWGQPHGGGGWGQGGTHSQWNKPSKPKTNM 112

ABL86017.1 GQPHGGGWGQ--------PHGGGWGQPHGGGWGQPHGGGGWGQGGTHSQWNKPSKPKTNM 112

            ********** ****.*********** **.* **.*******.*
```

```
AAH12844.1 KHMAGAAAAGAVVGGLGGYVLGSAMSRPIIHFGSDYEDRYYRENMHRYPNQVYYRPMDEY 169

AAB50624.1 KHMAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 162

ABY26043.1 KHMAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 169

AAV41883.1 KHVAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 140

AAV41879.1 KHMAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 140

AAV41893.1 KHMAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 141

AAV41891.1 KHVAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 141
```
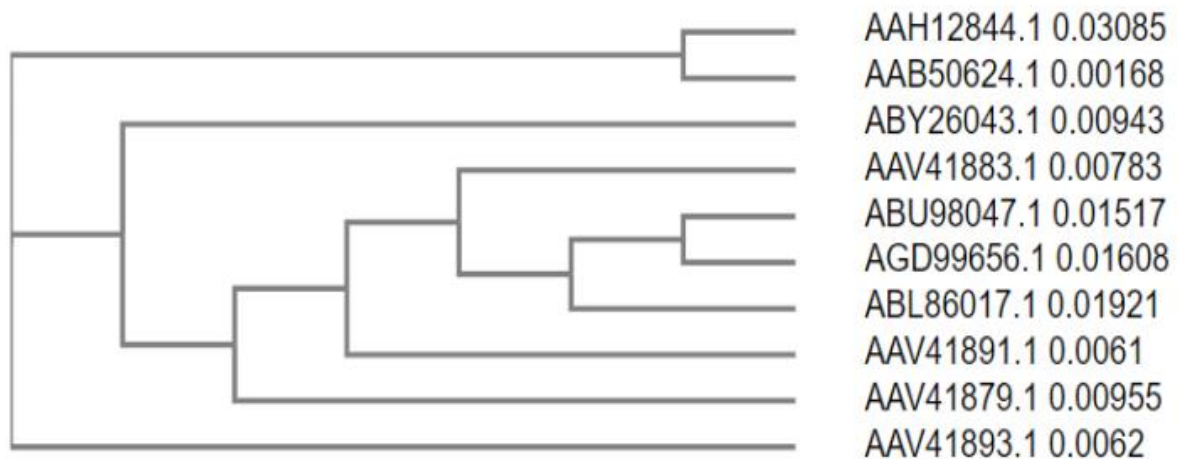
ABU98047.1 KHVAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGSDYEDRYYRENMHRYPNQVYYRPVDQY 180

AGD99656.1 KHMAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMHRYPNQVYYRPVDQY 172

ABL86017.1 KHVAGAAAAGAVVGGLGGYMLGSAMSRPLIHFGNDYEDRYYRENMYRYPNQVYYRPVDQY 172

 :*************:********:****.***********:**********:*:*

AAH12844.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRG 229

AAB50624.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYEKESQAYYQRG 222

ABY26043.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYEKESQAYYQRG 229

AAV41883.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRG 200

AAV41879.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYEKESQAYYQRG 200

AAV41893.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYEKESQAYYQRG 201

AAV41891.1 SNQNNFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYEKESQAYYQRG 201

ABU98047.1 SNQNNFVHDCVNITVKEHTVTTTTKGENFTETDIKMMERVVEQMCITQYQRESQAYYQRG 240

AGD99656.1 SNQNNFVHDCVNITVKEHTVTTTTKGENFTETDIKMMERVVEQMCITQYQRESQAYYQRG 232

ABL86017.1 SNQNNFVHDCVNITVKQHTVTTTTKGENFTETDIKIMERVVEQMCITQYQRESQAYYQRG 232

 **************:*:***************:*:*************::*********

AAH12844.1 SSMVLFSSPPVILLISFLIFLIVG 253

AAB50624.1 SSMVLFSSPPVILLISFLIFLIVG 246

ABY26043.1 SSMVLFSSPPVISLISFLIFLIVG 253

AAV41883.1 SSMVLFSSPP-------------- 210

AAV41879.1 SSIVLFSSPP-------------- 210

AAV41893.1 SSMVLFSSPP-------------- 211

AAV41891.1 SSMVLFSSPP-------------- 211

ABU98047.1 ASVILFSSPPVILLISFLIFLIVG 264

AGD99656.1 ASVILFSSPPVILLISLLIFLIVG 256
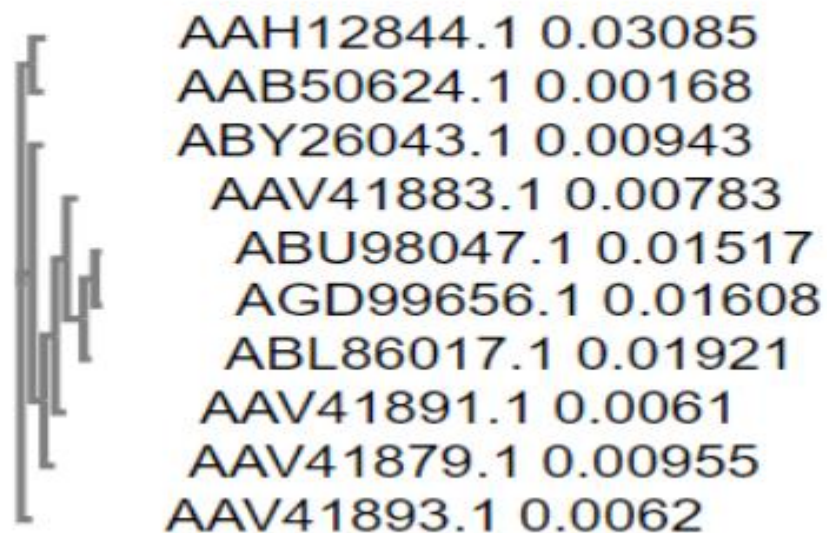
ABL86017.1 ASVILFSSPPVILLISFLIFLIVG 256

 .*..******

PHYLOGENETIC TREE representation is:

AAH12844.1 0.03085
AAB50624.1 0.00168
ABY26043.1 0.00943
AAV41883.1 0.00783
ABU98047.1 0.01517
AGD99656.1 0.01608
ABL86017.1 0.01921
AAV41891.1 0.0061
AAV41879.1 0.00955
AAV41893.1 0.0062

Sourav Singh, 198137

Branch length: ○ Cladogram ● Real

AAH12844.1 0.03085
AAB50624.1 0.00168
ABY26043.1 0.00943
AAV41883.1 0.00783
ABU98047.1 0.01517
AGD99656.1 0.01608
ABL86017.1 0.01921
AAV41891.1 0.0061
AAV41879.1 0.00955
AAV41893.1 0.0062

CONCLUSION:

From the above phylogenetic tree, we can conclude that prion protein from homo

sapiens and the colobus guereza are very closely related. And second pair of species which

have closely related prion protein is Bubalus bubalis and Bos taurus.

Macaca mulatta, Bubalus bubalis, Hippotragus niger, Bos taurus, Plecturocebus moloch,

Alouatta belzebul and Sapajus apella have evolved from the same ancestor.

We can also depict that homo sapiens and the colobus guereza & chiropotes satans;
Macaca mulatta and Alouatta elzebul are distantly related for the prion protein.

Sourav Singh, 198137