**Course: Advanced Bioinformatics**

**Module title: MSA for Phylogenetics**

**Module no. : 55**

**Stage 2: Multiple Sequence Alignment**

Multiple sequence alignment is a critical step of phylogenetic analysis. In many cases, the alignment of nucleotide or amino acid residues in a column implies that they share a common ancestor. If you misalign a group of sequences, you will still be able to produce a tree. However, it is not likely that the tree will be biologically meaningful. And if you create a multiple alignment of sequences and include a non-homologous sequence, it may still be incorporated into the phylogenetic tree.

MSA is an extension of a pairwise alignment, where it

* Can be local or global
* The “inputs” are the same
* A set of amino acid or nucleic sequences

– Substitution (scoring) matrices

– Gap penalties

The objectives are similar: find an alignment of more than two sequences

MSA programs: There are several online programs which help us to find MSA. For example following programs provide us options to align several protein sequences

* Clustal: Progressive MSA using phylogenetic tree
* ClustalV:
* ClustalW: command line interface
* ClustalX: This version has a graphical user interface.
* Clustal Omega: Command line-only program.
* T-Coffee: Alignment program that often gives better results, especially when dealing with divergent sequences and long insertions
* Optimal: Extension of dynamic programming to multiple dimensions for performing exhaustive search in sequences.
* Guaranteed to find an optimal score
* Need an n-dimensional matrix for scoring
* Computationally Expensive
* Heuristics: Limits of exhaustive search

• Attempt to rapidly find a good, but not necessarily optimal alignment

• Most popular methods:

- Tree alignments

- Star alignments