"MUST HAVE" LIST WITH ESSENTIAL FOR THIS KIND

OF PHYLOGENETIC ANALYSIS

THE MAIN STAGE OF WORK

ON PHYLOGENETIC ANALYSIS

1 DATA COLLECTION:

THE FIRST STEP IS TO GATHER THE NECESSARY DATA FOR ANALYSIS. THIS TYPICALLY INVOLVES OBTAINING DNA OR PROTEIN SEQUENCES FROM THE ORGANISMS OF INTEREST. THE SEQUENCES CAN BE OBTAINED FROM VARIOUS SOURCES, SUCH AS GENOMIC DATABASES, PUBLISHED STUDIES, OR NEWLY GENERATED DATA THROUGH DNA SEQUENCING TECHNIQUES.

2 SEQUENCE ALIGNMENT:

ONCE THE SEQUENCES ARE COLLECTED, THEY NEED TO BE ALIGNED TO IDENTIFY CORRESPONDING POSITIONS IN THE SEQUENCES. SEQUENCE ALIGNMENT HELPS IN IDENTIFYING HOMOLOGOUS REGIONS, WHICH ARE REGIONS THAT SHARE A COMMON ANCESTRY. THERE ARE VARIOUS ALGORITHMS AND SOFTWARE TOOLS AVAILABLE FOR SEQUENCE ALIGNMENT, FOR EXAMPLE, SUCH AS MUSCLE OR MAFFT. YOU CAN COMPARE THE RESULTS USING AN ESTIMATE OF THE ALIGNMENT TIME, ITS LENGTH, AND ITS GRAPHICAL INTERPRETATION.

3 TRIMMING:

IN THIS STEP, REGIONS OF THE SEQUENCE ALIGNMENT THAT ARE POORLY ALIGNED OR CONTAIN AMBIGUOUS REGIONS (E.G., GAPS, MISSING DATA, REPETITIVE REGIONS) ARE REMOVED. TRIMMING HELPS TO IMPROVE THE QUALITY OF THE ALIGNMENT AND REDUCES NOISE IN SUBSEQUENT ANALYSES. SOFTWARE TOOLS TRIMAL ARE COMMONLY USED FOR TRIMMING.

4 MODEL SELECTION:

SELECT THE MOST APPROPRIATE MODEL OF SEQUENCE EVOLUTION THAT BEST FITS THE DATA, TAKING INTO ACCOUNT THE CHARACTERISTICS OF THE SEQUENCE ALIGNMENT AND THE SELECTED PHYLOGENETIC INFERENCE METHOD.

5 DATA CONCATENATION:

IF MULTIPLE GENETIC MARKERS OR GENES ARE AVAILABLE, CONCATENATE THE ALIGNED SEQUENCES FROM DIFFERENT MARKERS OR GENES INTO A SINGLE ALIGNMENT. THIS STEP COMBINES THE INFORMATION FROM MULTIPLE LOCI AND CAN PROVIDE A MORE COMPREHENSIVE REPRESENTATION OF EVOLUTIONARY RELATIONSHIPS. TOOLS LIKE CATFASTA2PHYML OR SEQUENCEMATRIX CAN BE USED FOR DATA CONCATENATION.

6 PHYLOGENETIC TREE CONSTRUCTION:

USE THE TRIMMED SEQUENCE ALIGNMENT TO CONSTRUCT A PHYLOGENETIC TREE REPRESENTING THE EVOLUTIONARY RELATIONSHIPS BETWEEN THE ORGANISMS. SELECT AN APPROPRIATE PHYLOGENETIC INFERENCE METHOD BASED ON THE DATA AND RESEARCH GOALS, SUCH AS DISTANCE-BASED METHODS, MAXIMUM LIKELIHOOD OR BAYESIAN INFERENCE.

7 TREE VISUALIZATION AND INTERPRETATION:

VISUALIZE THE CONSTRUCTED PHYLOGENETIC TREE USING SPECIALIZED SOFTWARE TOOLS. ANNOTATE THE TREE WITH ADDITIONAL INFORMATION AND INTERPRET THE TREE BY ANALYZING BRANCHING PATTERNS, IDENTIFYING CLADES, AND INFERRING EVOLUTIONARY RELATIONSHIPS AND DIVERGENCE TIMES.

ADDITIONAL STEP (STATISTICAL SUPPORT):

ASSESS THE STATISTICAL SUPPORT FOR THE INFERRED RELATIONSHIPS IN THE PHYLOGENETIC TREE, TYPICALLY USING BOOTSTRAPPING OR APPROXIMATE LIKELIHOOD RATIO TESTS (ALRT) TO ESTIMATE THE ROBUSTNESS OF THE TREE.