Santhanakrishnan Ramani

SID: 105720585

**Bioinformatics & Genomics**

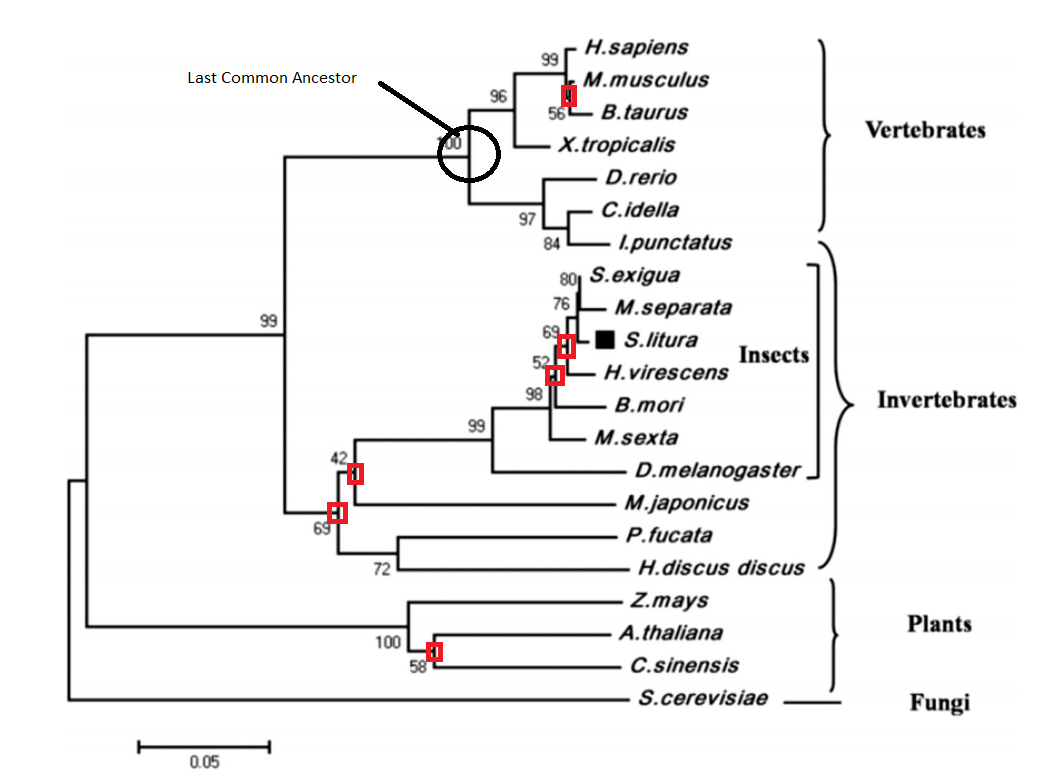
**Assignment – 6**

1. a) Phylogram

b) S.Cerevisiae sequence is outgroup.

c) C.Sinensis is closely related to A.thaliana

d) Have circled the last common ancestor

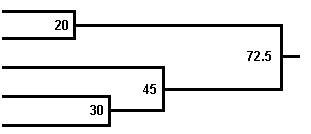


e) The branches marked using a square are the branches that I have least confidence in. The reason

being the minimum recommended critical value while using bootstrapping is 70.

1. Rooted tree using the UPGMA method of tree construction.



A

C

Root

E

D

B

1. a) No of rooted trees: (2n – 3)! / 2(n-2) (n-2)! = (2\*5 – 3)! / 23 3! = 7! / 8 \* 3! = 105

No of unrooted trees: (2n – 5)! / 2(n-3) (n-3)! = (2\*5 – 5)! / 22 2! = 5! / 4 \* 2! = 15

b) Blue

A B E C D

Red

A B E C D

Green

A B E C D

1. In order to explain the difference between an algorithm and scoring scheme, we can take the example of Smith-Waterman algorithm used for local sequence alignment and different Blosum matrices used for scoring the alignment. The algorithm remains the same for local sequence alignment and produces the same result for a given problem, but the value of score obtained changes depending on the scoring scheme employed.
2. a) Molecular Function (MF), Biological Process (BP) and Cellular Component (CC)

b) Directed Acyclic Graph (DAG)

c) The difference between full and filtered GO is filtered GO does not contain any ‘has\_part’ or inter-

Ontology relationships.

d) The Gene Ontology Annotation project.

e) The IEA evidence code is by far the most abundantly used evidence code.

f) Information Content Measure:

Pro: Gives the similarity measure of two GO terms, based on their probability value.

Con: Rare terms have high IC.

g) Criticisms of GO:

* Annotation describe only describe the normal, healthy functioning of genes.
* Data on functional coordination between multi-function genes are not explicitly stored.
* No relationships between the three ontologies were recorded.