# INSILICO SYSTEMS BIOLOGY PRACTICALS

# EXPERIMENT- 20

**AIM-** To explore Panther Database

**Theory-** The PANTHER (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System was designed to classify proteins (and their genes) in order to facilitate high-throughput analysis. Proteins have been classified according to:

* Family and subfamily: families are groups of evolutionarily related proteins;

subfamilies are related proteins that also have the same function

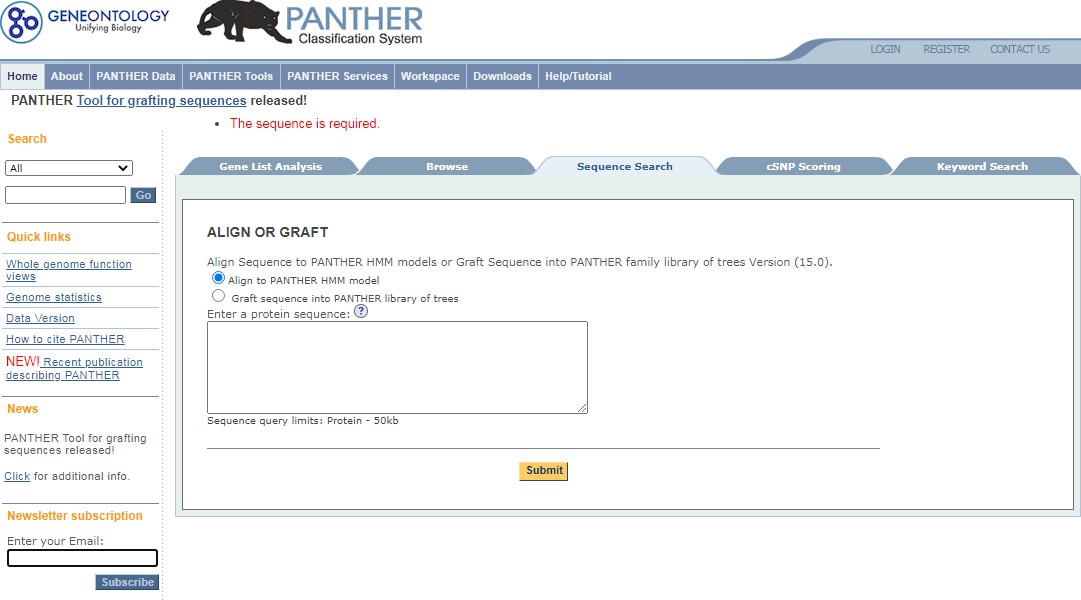
* Molecular function: the function of the protein by itself or with directly interacting proteins at a biochemical level, e.g. a protein kinase
* Biological process: the function of the protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism, e.g. mitosis.
* Pathway: similar to biological process, but a pathway also explicitly specifies the relationships between the interacting molecules.

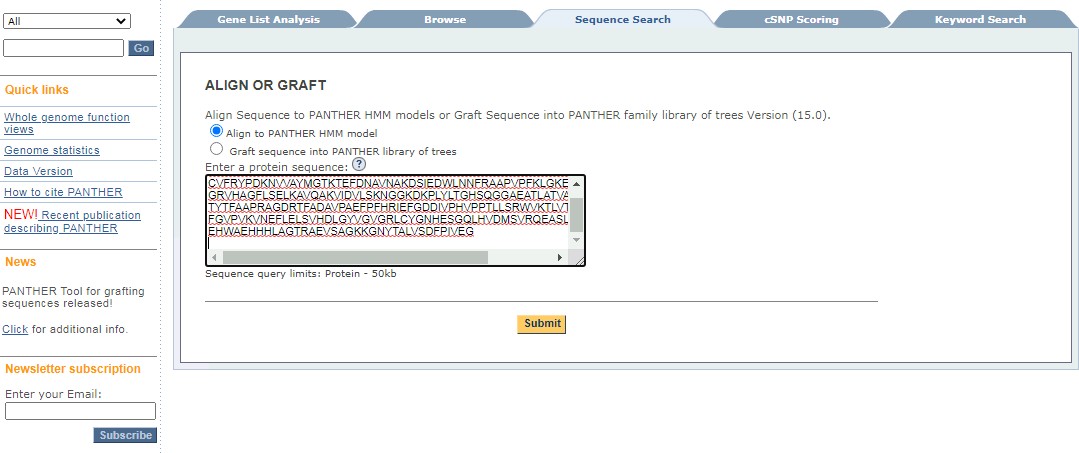
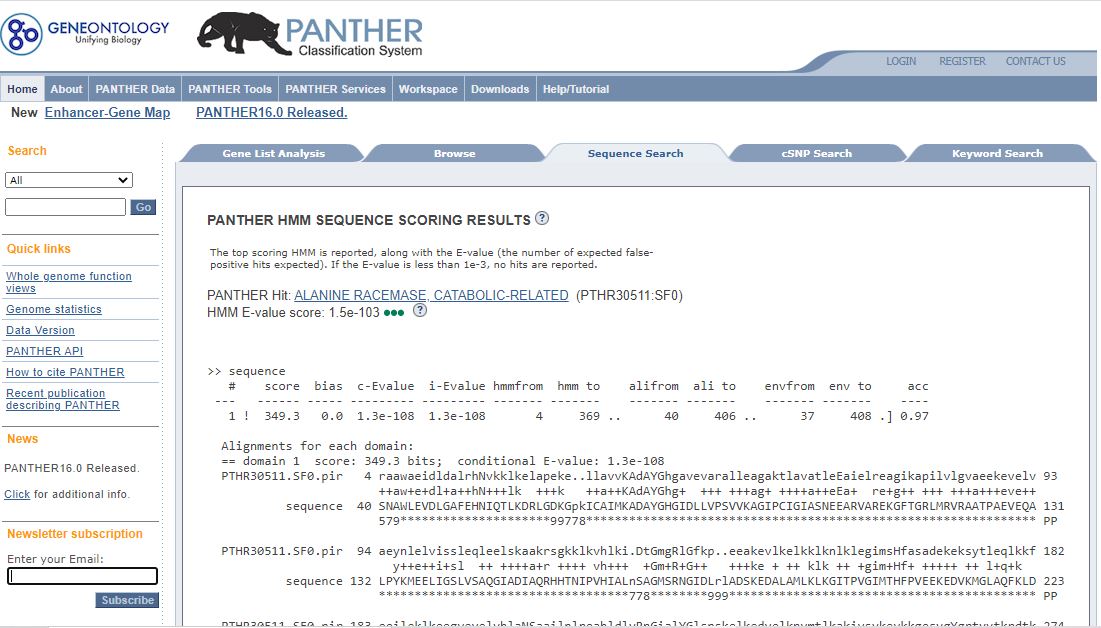
The PANTHER Classifications are the result of human curation as well as

sophisticated bioinformatics algorithms. PANTHER is a comprehensive database of evolutionary and functional information about protein-coding genes. Protein sequences (one representative sequence per gene) from 104 complete genomes (PANTHER uses the UniProt Reference Proteomes data set ) are organized into families of homologous genes. For each family, a phylogenetic tree is constructed, representing the evolutionary relationships between all genes in the family, and the processes by which these genes were first copied and then diverged from each other: speciation, gene duplication and horizontal transfer.

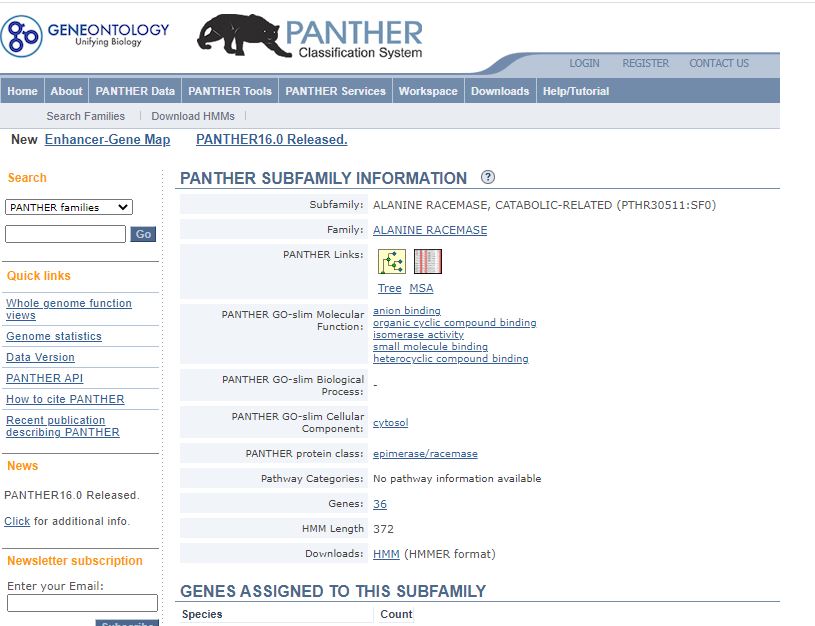
# Protocol-

1. Open Panther Database

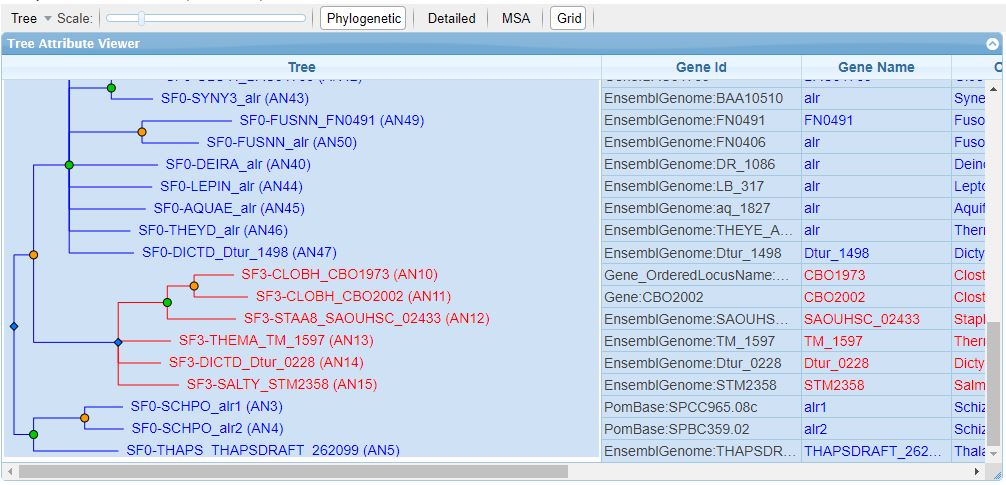


1. Enter a Protein Sequence and press submit button.
2. After submitting, it takes to the window displaying Sequence search results. In this case, the results show that they belong to the Alanine Racemase  family.

4.The hyperlinked protein’s super family takes to the next window showing the information regarding the superfamily and Genes assigned to the subfamily.



1. There are 2 Panther Links. One showing the Phylogenetic tree and the second is the multiple sequence alignment. The Panther Tree viewer looks like-



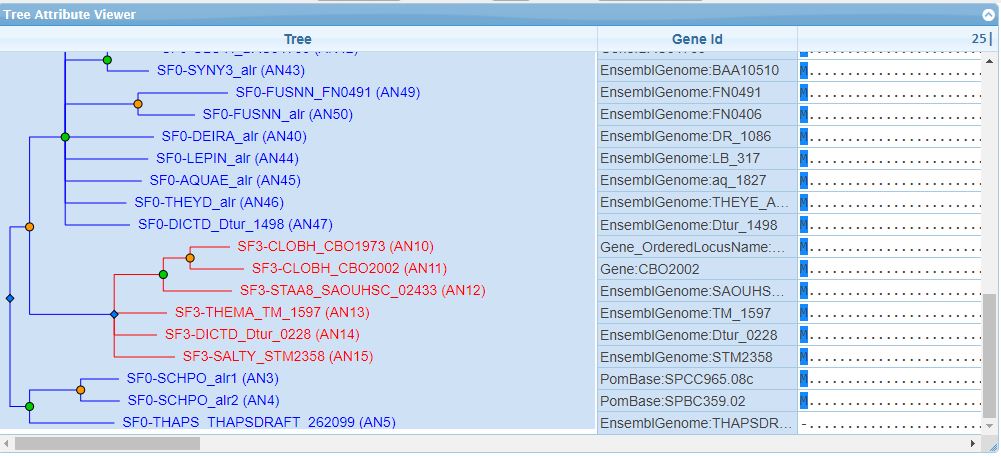
PANTHER version 15.0 contains 15702 protein families, each with a phylogenetic tree relating modern-day genes in 142 organisms. Phylogenetic trees also infer the evolutionary events that occurred, namely speciation (green nodes) and gene duplication events (orange nodes). Horizontal transfer events are not inferred in this version of PANTHER, but will appear in future versions. Expert biologists have divided each family into subfamilies, which are generally orthologous groups but may also contain recently duplicated paralogs. Each family and subfamily is also represented as a hidden Markov model (HMM), which can be used to classify new sequences to an existing subfamily.

1. PANTHER multiple sequence alignments (MSAs) are the basis for the PANTHER distance trees, and therefore of the family/subfamily classification. You can view them in two modes:

**Full MSA**: Includes all (publicly available) sequences in the family that are related closely enough to produce an informative multiple alignment (i.e. the resulting trees and HMMs are useful for function prediction at both a family and subfamily level).

**Partial MSA**: Shows the alignment for only the currently selected subfamilies. In the MSA viewer, you can:

* Change the selection of subfamilies shown by clicking on "Subfamily Selection" just as in the Tree Viewer.
  + Focus on only a part of the sequence alignment ("range").
  + Change the font size of the alignment.
  + Jump to the start or end positions of the HMM alignment (by clicking on the links after the HMM length).



1. The gene statistics

