LEARN PYTHON & R FOR BIOINFORMATICS

Prerequisite Terminologies:

For the complete comprehension of the main topic, you should've the basic concept of the following term:

➢ **BLAST (Basic Local Alignment Search Tool)**

Introduction:

Phytozome is a comparative hub for plant genomes that allow users to access, visualize, analyze or download sequenced plant genomes as well as selected genomes and datasets that have been sequenced in other databases. It is the plant genomics portal of the Department of

Energy's Joint Genome Institute. The point of JGI is that it stores and monitors genomic, proteomic and various other information related to plants. It is the most utilized database for various research on plant genomics such as to retrieve the genes of a particular plant genome, its annotation, enzyme pathways or interpro families.

Steps:
- We'll discuss the detailed introduction of Phytozome, so open the homepage of the phytozome database using the link:https://phytozome.jgi.doe.gov/pz/portal.html.
- As you'll see, the homepage of phytozome is fairly simple. On the right side of the homepage, you'll find *'Help with Phytozome'* where you can see the documentation.
    - In documentation, the option '*view tree representation'* provides you information about how the species dataset of plants is organized within the database. Basically, it shows the classification of plants.
    - From *'Quick Start Guide'.* You can reach out for guidance to use phytozome.
    - From *'FAQs'*, you can check the frequently asked questions about the database.
- The *'News'* option given on the homepage is also useful because it allows you to understand what system status is available.
  [For example, if there is a cross available on BLAST in *system status,* it means that BLAST on JGI is not working properly.**]**
- The most important section on the phytozome homepage is *'Phytozome quick search',* where you'll see every species of plant available on this database along with their pictures. You can select any species you want to retrieve and get started on your analysis.
    - In the *'Phytozome quick search', Flagships* information is that those plants that are frequently utilized for daily research.
    - *Clustered Genomes and Families* means that the species are clustered, as you go along you'll see each species has their own set and they are clustered along.
    - In *Unclustered Genomes,* you'll find genomes organized in alphabetical order and this information is unclustered. Also, you'll find **V** along with the species names, which means what version of that particular species is and how many times it is uploaded.
- On the top left side of the homepage, *'Species'* option is available where you can see every species name that is present in the Phytozome arranged alphabetically in the drop down menu.
- Along with the species, *'Tools'* option is available where in the drop down menu you'll find the following:
    - *Keyword search:* From where you can search for any particular species using keywords then the database will try to figure it out in the entire species list that it has.
    - *BLAST:* You can do blast, for example you would like to see to which particular organism your sequence belongs to, it'll search throughout the database.
      [To have a better understanding of BLAST, watch out for our tutorial on it.]
    - *JBrowse:* It provides graphical visualization of genomes.

- ➔ *PhytoMine:* (InterMine earlier) allows you to retrieve query based data and mine the data from the entire database.
- ➔ *BioMart:* Handled by Ensemble, allows you to do data mining in genomics data. PhytoMine and BioMart both are useful in Bioinformatics analysis.
  [We'll discuss all of these in the other sections of this video.]
- From the download option, you can download all the versions of a particular species.
  [To download the data sets, you must first login or register a new account.]

Summary:

In this video, we've gone through the detailed introduction on the Phytozome database. We also learned its different features, how we can retrieve the data set of our interest and the services it provides.