JUST AGRICULTURE
multidisciplinary e-Newsletter

Bioinformatics and its Techniques in Plant Pathology

Gaurav Rakhonde[1], Shalaka Ahale[2], Kumari Surbhi[3], Himani Jeena[3]
[1]Department of Plant Pathology, University if Agricultural Sciences, GKBK, Bangalore
[2]Department of Plant Pathology, Punjab Agricultural University, Ludhiana
[3]Department of Plant Pathology, College of Agriculture, GBPUAT, Pantnagar

ARTICLE ID: 13

Introduction

Improving crops with a focus on quality and quantity with disease-free crops are in demand right now. It is now possible to improve crops through the development of sequence markers founded on single nucleotide polymorphism and simple sequence repeat identification. To completely comprehend and evaluate the biological system, numerous methodologies, databases, tools and software have been created. Bioinformatics plays crucial role in agricultural crop improvement. Bioinformatics is the use of computing and analysis tools to collect and understand biological data. With the help of computer science, mathematics, physics, and biology, it is an interdisciplinary discipline.

Bioinformatics Techniques

A. **Comparative Analysis:-** In the subject of comparative analysis, genomic sequence characteristics of various organisms including their DNA sequence, regulatory region sequence genes and gene order are compared. Finding the characteristics that are shared by homologous sequences and are encoded in DNA to demonstrate how these characteristics have changed little during evolution or how different regions contribute to variety

B. **Sequence Analysis:-** Using sequence alignment and searches against biological sequence databases like reference genes, proteins, UniProtKB, PDB, etc., it is the process of submitting a DNA, RNA, or protein homologous gene sequence to understand its evolution, function, structure, or properties. To obtain a genetic marker, it is helps to find similarity and dissimilarity, active sites, post-translational modification sites, gene structures, reading frames, distributions of introns and exons, and regulatory elements, point mutations, single nucleotide variants (SNVs), and single nucleotide polymorphisms (SNPs), revealing evolution and genetic diversity of

JUST AGRICULTURE
multidisciplinary e-Newsletter

sequences and organisms and enable molecular structure to be determined solely from sequence. The greatest method for displaying evolutionary and genetic diversity of sequences and species with identifying molecular structure from sequence is a simple local alignment technique (Bawono *et al.*, 2017).

C. **Gene Identification:-** Finding the genomic DNA areas encoding genes is the process at hand. Once an organism genes and genome have been sequenced and made public, one of the first and most crucial stages in comprehending them is gene identification. Following genome sequence assembly, filtering of coding (exonic) and non-coding (intronic) sections and gene discovery are the main phases in genome annotation (Alioto, 2012)

D. **Phylogenetic Analysis:-** It is the process of investigation of the evolutionary connections between collections of similar genes from different species. Based on phylogenetic inference techniques, these phylogenetic connections are found. A phylogenetic tree is a branching tree diagram that shows the relationships that have developed through time among a group of particular biological organisms or species. Based on similarities and variations in their genetic or physical traits, phylogenetic assumptions are made. Understanding genomes, diversity, evolution, and ecology has become largely dependent on phylogenetic analysis (Thompson et al., 2002).

E. **Protein–Protein Interaction:-** Protein-protein interactions are highly specialized physical connections between two or more protein molecules triggered by electrostatic and hydrophobic biochemical processes. In the STRING database, interactions are predicted based on gene neighbourhood, gene fusions, or gene co-occurrence as well as known interactions based on curated databases, empirically determined interactions, and additional interactions based on text mining, co-expression, or protein homology (Szklarczyk et al.2019)

F. **Microarray Data Analysis:-** Researchers can examine the expression level of a huge number of genes from the entire genome of an organism in a single experiment by using microarray data analysis, used to infer information from the data generated from DNA, RNA and protein microarray experiments. A public database called Gene Expression Omnibus (GEO) accepts data submissions adhering to the MIAME (Minimum Information About a Microarray Experiment) standard. The repository

JUST AGRICULTURE
multidisciplinary e-Newsletter

accepts data based on sequences and arrays. Researchers may download and query experimental datasets and gene expression profiles using a variety of methods and tools. The scientific community has provided GEO with a collection of publicly downloadable microarray data, next-generation sequencing data, and other high-throughput functional genomics data (Clough and Barrett 2016).

Structure Prediction and Refinement

Protein structure estimation is the process of inferring a protein's three-dimensional structure from its amino acid composition. The hypothesis has folds and secondary and tertiary structures from its original sequence in three dimensions. It is crucial for both the development of new 3D enzymes and the design of pharmaceuticals (França, 2015).

Molecular Docking Calculation

The interaction of two or more molecules to create a stable complex structure is known as molecular docking. It creates a three-dimensional structural complex based on the binding characteristics of the ligand and target. It is a strategy to forecast how one molecule will align with another molecule in the bound structure, which creates a stable complex. When utilizing scoring functions to estimate the binding potency or affinity between receptor ligand molecules, knowledge of the active site orientation is helpful. The logical design of new pesticides, herbicides, insecticides, and fungicides depends in large part on the characterization of the active binding behaviour (Trosset and Cave, 2019).

Conclusion

Incredible success and the public availability of vast amounts of host-pathogen next-generation sequencing data present us with prospects to comprehend the disease system physiologically and other processes. It is now possible to access, annotate, analyse, and determine the functional elements for characterization at the gene and genome levels for application development because to the availability of host-pathogen genome data in well-known repository systems. The main objective of bioinformatics is to improve our knowledge of biological processes through the use of biological data mining, sequence pattern recognition, machine learning techniques for biological datasets, and molecular and data visualisation.In order to find prospective genes and target proteins for host-pathogen interactions and to collect biological data from plant genomes and their genes, bioinformatics

JUST AGRICULTURE
multidisciplinary e-Newsletter

tools, methodologies, and databases can be employed. Applications of bioinformatics can be highly useful in crop enhancement and the creation of improved crops.

Reference

Alioto T (2012) Gene prediction. Methods Mol Biol 855:175–201

Bawono P, Dijkstra M, Pirovano W, Feenstra A, Abeln S, Heringa J (2017) Multiple sequence alignment. Methods Mol Biol 1525:167–189

Clough, E., and Barrett, T. (2016). The gene expression omnibus database. In *Statistical genomics* (pp. 93-110). Humana Press, New York, NY.

França, T. C. C. (2015). Homology modeling: an important tool for the drug discovery. *Journal of Biomolecular Structure and Dynamics*, *33*(8), 1780-1793.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Mering, C. V. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, *47*(D1), D607-D613.

Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, (1), 2-3.

Trosset, J. Y., and Cavé, C. (2019). In silico drug–target profiling. Target Identification and Validation in Drug Discovery, 89-103.